UNIVERSITAT POLITÈCNICA DE CATALUNYA

Exploration of Customer Churn Routes Using Machine Learning Probabilistic Models

DOCTORAL THESIS

Author:

David L. García

Advisors:

Dra. Àngela Nebot and Dr. Alfredo Vellido

Soft Computing Research Group
Departament de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya. BarcelonaTech

Barcelona, 10th April 2014

A bird doesn't sing because it has an answer,
it sings because it has a song.
Maya Angelou

A Mónica, Olivia, Pol i Jana

Acknowledgments

Abstract

The ongoing processes of globalization and deregulation are changing the competitive framework in the majority of economic sectors. The appearance of new competitors and technologies entails a sharp increase in competition and a growing preoccupation among service providing companies with creating stronger bonds with customers. Many of these companies are shifting resources away from the goal of capturing new customers and are instead focusing on retaining existing ones. In this context, anticipating the customer's intention to abandon, a phenomenon also known as churn, and facilitating the launch of retention-focused actions represent clear elements of competitive advantage.

Data mining, as applied to market surveyed information, can provide assistance to churn management processes. In this thesis, we mine real market data for churn analysis, placing a strong emphasis on the applicability and interpretability of the results. Statistical Machine Learning models for simultaneous data clustering and visualization lay the foundations for the analyses, which yield an interpretable segmentation of the surveyed markets. To achieve interpretability, much attention is paid to the intuitive visualization of the experimental results. Given that the modelling techniques under consideration are nonlinear in nature, this represents a non-trivial challenge. Newly developed techniques for data visualization in nonlinear latent models are presented. They are inspired in geographical representation methods and suited to both static and dynamic data representation.

Contents

List of Figures

7   Distortion visualization in GTM

8   Distortion and Flow Maps visualization for churn analysis

List of Tables

5   Supervised customer loyalty analysis

6   Unsupervised churn analysis in a telecommunications company

7   Distortion visualization in GTM

8   Distortion and Flow Maps visualization for churn analysis

List of Acronyms

ANN:	Artificial Neural Networks.
ARD:	Automatic Relevance Determination.
AUC:	Area Under the Curve.
BI:	Behavioural Intentions.
BMU:	Best Matching Unit.
BSOM:	Batch Self Organizing Maps.
CART:	Classification And Regression Trees.
CCM:	Customer Continuity Management.
CHAMP:	Churn Analysis Modelling and Prediction.
CI:	Computational Intelligence.
CV:	Cartogram Visualization.
CVM:	Customer Value Management.
DB:	Davies-Bouldin index.
DM:	Data Mining.
DMEL:	Data Mining by Evolutionary Learning.
DR:	Dimensionality Reduction.
DT:	Decision Trees.
EANC:	Economic Activities National Classification.
EC:	Evolutionary Computation.
EM:	Expectation-Maximization.
ESANN:	European Symposium on Artificial Neural Networks.
FA:	Factor Analysis.
FL:	Fuzzy Logic.
FRD:	Feature Relevance Determination.
FRD-GTM:	Feature Relevance Determination for Generative Topographic Mapping.
FS:	Feature Selection.
GA:	Genetic Algorithm.
GAM:	Generalized Additive Models.
GMM:	Gaussian Mixture Models.
GTM:	Generative Topographic Mapping.
IDEAL:	Intelligent Data Engineering and Automated Learning.
LET:	Long Echo Time.
LI:	Lift Index.
LTV:	Lifetime Value.
MF:	Magnification Factors.
ML:	Machine Learning.
MRS:	Magnetic Resonance Spectroscopy.
MTD:	Mixture Transition Distribution.

MVD:	Multivariate Data.
NAA:	N-Acetyl Aspartate.
NLDR:	Nonlinear Dimensionality Reduction.
OSQ:	Overall Service Quality.
OSRE:	Orthogonal Search Rule Extraction.
PCA:	Principal Components Analysis.
PMP:	Posterior Mean Projection.
PPM:	Parts Per Million.
PPV:	Positive Predicted Value.
PR:	Pattern Recognition.
RF:	Random Forest.
RFM:	Recency, Frequency, Monetary Value.
ROC:	Receiver Operating Characteristic.
RSM:	Random Subspace Method.
SICO:	Symposium on Computational Intelligence.
SML:	Statistical Machine Learning.
SOM:	Self-Organizing Maps.
SQ:	Service Quality.
SQP:	Service Quality Performance.
SV-1H-MR:	Single-Voxel Proton Magnetic Resonance.
SV-1H-MRS:	Single-Voxel Proton Magnetic Resonance Spectroscopy.
SVM:	Support Vector Machines.
TDL:	Top Decile Lift.
VAS:	Value Added Services.

Chapter 1

Introduction

The ongoing processes of globalization and deregulation are changing the competitive framework in the majority of economic sectors. The appearance of new competitors and technologies entails a sharp increase in competition and a growing preoccupation among service providing companies with creating stronger bonds with customers. Many of these companies are shifting resources away from the goal of capturing new customers and are instead focusing on retaining existing ones. In this context, anticipating the customer's intention to abandon, a phenomenon also known as churn, and facilitating the launch of retention-focused actions represent clear elements of competitive advantage.

Data mining, applied to market surveyed information, can provide assistance to churn management processes. The main aim of this thesis was to mine real market data for churn analysis, placing a strong emphasis on the applicability and interpretability of the results.

One of the constituting stages of most data mining and knowledge discovery methodologies currently in use is data exploration [76, 229]. It should help bringing into focus relevant aspects of the analyzed data, which is a key goal in market analysis. When data are naturally high-dimensional, and this is often the case in such application area, the task of data visualization becomes central to data exploration [154].

For this reason, and with the goal of interpretability at the forefront, much attention is paid in this thesis to the problem of visualization of the experimental results. The visualization of multivariate data (MVD), as used in the pursuit of market knowledge generation, is a problem in between natural and artificial pattern recognition (PR): Natural because information visualization entails complex cognitive processing of visual stimuli [129, 182]; and artificial because, in the face of complex high-dimensional data, researchers are challenged to develop visualization-oriented PR techniques. The natural and artificial aspects of the visualization PR problem are both relevant and inextricable and, as a result, the use of visual metaphors entails the risk of introducing subjectivity in the knowledge generation process [297]. If both aspects are used at their best, they can enhance each other in order to make data exploration a fruitful task [270].

Visualization is a non trivial problem for high-dimensional data, where intuitive insights about inner structure are hardly ever available. Some form of dimensionality reduction is thus required, either through feature selection or through feature transformation and extraction. In recent times, nonlinear dimensionality reduction (NLDR) methods have provided powerful and flexible strategies for high-dimensional data modeling and exploration.

One of the model families belonging to the wide palette of NLDR techniques is that of manifold learning methods, which attempt to represent multivariate data assuming they can be approximated reasonably well by low-dimensional manifolds covering the most densely populated areas where data reside. When data modeling focuses on exploration, these manifolds are often chosen to be 2-dimensional to provide the model with data visualization capabilities.

A drawback limiting the use of NLDR techniques is the difficult interpretation of their resulting data representations. Even manifold learning models, which can represent high-dimensional data in a low-dimensional representation space, are not necessarily straightforward to make sense of. This is because their projected coordinates of representation are a complex nonlinear transformation of the observed ones. Different parts of the original observed data space may undergo different levels of distortion as part of the mapping process, which are not obvious from the data visualization itself.

The potential lack of interpretability is the price paid by these methods for their flexibility to faithfully represent MVD. Linear dimensionality reduction methods are less flexible in the transformation they provide and, as a result, their representation of high-dimensional data can be less faithful. Compensating for this, their subset of representation coordinates can be expressed as a linear combination of the observed data attributes, which often makes these models easy to interpret.

Statistical Machine Learning (SML) models for simultaneous data clustering and visualization lay the foundations for the experimental analyses reported in the following chapters. This type of models replicate the functionality of more traditional computational intelligence clustering methods while providing a sound probabilistic foundation in their definition. This probabilistic framework makes the models easier to compare with traditional multivariate statistics and also provides ground for a motivated choice of model parameters.

Given that the modelling techniques under consideration in this thesis are nonlinear in nature, this represents a non-trivial interpretability challenge. Newly developed techniques for data visualization in nonlinear latent models are introduced in the following chapters. They are inspired in geographical representation methods and suited to both static and dynamic data representation.

One of these methods is inspired from a technique originally designed for the analysis of geographic information, namely *Cartograms* [91]. They are geographic maps in which the sizes of regions are display in sizes that are proportional to underlying quantities such as their population.

The continuity-preservation requirements generated by nonlinear manifold learning techniques are akin to those generated by geographical maps. Thus, one of the conceptual leaps in this thesis consists on extrapolating from geographical maps to the "virtual geographies" of the visualization spaces of manifold learning NLDR models. It also requires the substitution of geography-distorting quantities such as population density by quantities reflecting the mapping distortion introduced by these nonlinear models.

The proposed cartogram-based visualization method reintroduces the distortion, as expressed by readily quantifiable measures, explicitly into the visualization maps. By doing so, these cartogram-distorted maps become more representative and, importantly, more intuitively interpretable for practical purposes such as churn analysis.

This thesis also introduces an important visual interpretation technique for the analysis of the evolution of customer behaviour over time and their propensity to churn, namely the Flow Map. Flow Maps were originally devised to visualize geography related evolution patterns such as, for instance, population migrations, and have become increasingly sophisticated from a computational viewpoint. Given that the analyzed database contains information over time, we use Flow Maps to analyze the customer migrations over the GTM visualization map, aiming to detect foci of potential customer churn. This approach should provide useful information for customer management.

1.1 Main goals of the thesis

The current doctoral thesis has a strong business application component. This is not to say that theoretical developments are not part of it. Quite the opposite, all the applied research reported in this document has its foundations in novel theoretical developments that involve machine learning (ML) techniques. Each of such developments, though, is meant to have a practical application in the context of churn analysis.

Therefore, the goals of the thesis are twofold: Part of them are business problem-related goals, while another part are ML-related theoretical goals. They are both summarily listed in the following paragraphs.

1.1.1 Market analysis goals of the thesis

- The detailed review of existing approaches to customer churn analysis using ML and computational intelligence methods.

- Finding meaningful segments in several customer markets, mostly in the area of telecommunications.

- Providing intuitive methods of visualization of the MVD related to the obtained market segments.

- Relating these visual representations of the obtained market segments to the phenomenon of churn, in such a way that the proposed techniques become a tool for churn management.

- Adapting the previous tools to the analysis of the temporal dynamics of the market segments, in order to achieve the ultimate goal of long-term market profitability.

- Overall, providing a quantitative framework for data-based churn analysis, based on the general constructs of customer service, customer satisfaction, and customer loyalty.

1.1.2 Data modelling goals of the thesis

- Definition of a principled quantitative approach to customer market segmentation oriented towards churn analysis and with an emphasis on MVD visualization.

- Such approach should provide a quantitative method for assessing the relative relevance of individual data features from the point of view of their impact on segment structure and should take advantage of the probabilistic nature of the clustering method at its core for the definition of a suitable segment partition.

- Definition of methods to improve the visualization of MVD when using NLDR techniques for their modelling. The main objective would be reducing the negative impact of local nonlinear distortion on the interpretability of the mapping of MVD to low-dimensional visualization spaces.

- Such methods should aim to explicitly reintroduce the nonlinear mapping distortion locally into the visual representation, so as to improve its faithfulness.

- Such methods should also cater for the need to represent data dynamics over time, in such a way that the evolution of customers and their *migrations* across their *behavioural maps* could be duly tracked.

1.2 Structure of the Thesis

The remaining of the thesis document is structured as follows:

- Chapter 2 introduces, in a self-contained way, all the business-related theoretical background that is necessary to understand the field of application investigated in the present Doctoral Thesis. The general churn problem is first introduced from a business point of view; this is followed by an exhaustive revision of existing explanatory customer loyalty building models proposed in recent literature, with emphasis placed in the concepts of customer continuity management, customer satisfaction and loyalty, and service quality and costs.

- In Chapter 3, the viewpoint of the background research shifts towards churn seen as a data mining problem that is dealt with using PR methods. Together with the previous chapter, the reader is meant to obtain an overall vision of the context of the churn phenomenon prior to the reporting of the developed methods and the corresponding experimental investigation.

- In this thesis, the data mining process mostly concerns the use of unsupervised ML techniques. Within the overall goal of exploring the existence of customer churn routes according to the customers' service consumption patterns, we are interested in methods that are capable of providing simultaneous visualization and clustering of the available data. Chapter 4 provides a self-contained

introduction to general latent models and NLDR techniques. Arguably the best-known and most-used NLDR method is Self-Organizing Maps (SOM); for this reason, this chapter includes a description of its basic forms, which is followed by the introduction of the standard version of its probabilistic counterpart, Generative Topographic Mapping (GTM).

- In chapter 5, we follow a supervised learning approach to analyze the drivers towards customer satisfaction from a survey conducted amongst the customers of several Spanish petrol station brands. Such description is carried out by an artificial neural network (ANN) defined within a Bayesian framework with feature relevance determination. This is complemented with a rule description of the classification performed by the ANN through Orthogonal Search Rule Extraction (OSRE).

- In chapter 6, we focus, from a marketing viewpoint, on proactive bonding. In particular, we propose an indirect and explanatory approach to the prediction of customer abandonment, based on the structured visualization of customer data -consisting of their consumption patterns- on a two-dimensional representation map, to explore the existence of abandonment routes in the Brazilian telecommunications market. This map is obtained with a manifold learning NLDR technique for which a two-tier market segmentation process with embedded feature relevance determination is proposed.

- Chapter 7 provides the main theoretical developments of the thesis. Here, inspired from a technique originally designed for the analysis of geographic information, namely the *-cartogram*, we propose a new method for explicitly reintroducing the geometrical distortion created by NLDR manifold learning models into their low-dimensional representation of the MVD. The proposed cartogram-based method reintroduces the distortion explicitly into the visualization maps. By reintroducing this distortion explicitly, we should now expect to obtain more faithful low-dimensional representations of the data. An extensive set of experiments is carried out, using artificial and real data. With these, we explore the properties of the proposed cartogram method and provide some guidelines for its use.

- The cartogram representation, introduced in the previous chapter, is inspired in a real cartographic technique and it is suited to the visualization of static data as modelled by NLDR methods. In chapter 8 we introduce a second cartography-inspired method: the Flow Map. Flow Maps were originally devised to visualize geography-related evolution patterns such as population migrations and have become increasingly sophisticated from a computational viewpoint. Given that the analyzed databases contain information over time, we use Flow Maps to explore the *customer migrations* over the GTM visualization map, aiming to detect *foci* of potential customer churn.

- Chapter 9 concludes and wraps the thesis up, providing a summary of its contributions and an outlook of possible avenues for future research.

Chapter 2

Customer Continuity Management as a foundation for churn Data Mining

This chapter introduces, in a self-contained way, all the business-related theoretical background that is necessary to understand the application field of the problems investigated in the present thesis. The general customer abandonment -churn- problem is first introduced from a business point of view; this is followed by an exhaustive revision of existing explanatory customer loyalty building models proposed in recent literature. In *Chapter 3* the viewpoint will shift towards churn seen as a data mining problem that is dealt with using PR methods. With this, the reader is meant to obtain an overall vision of the context of the churn phenomenon prior to the reporting of the developed techniques and the corresponding experimental investigation.

2.1 Customer churn: the business case

In the scenario of growing competitive pressure described in the introduction, where all companies fight over their customer portfolios, the possibilities of commercial development and, consequently, of adding value to the company, require prolonging the useful life of customers and their average consumption[1] (see *Figure 2.1*).

Therefore, understanding how customer loyalty construction mechanisms work, anticipating the customer's intention to abandon and facilitating the launch of retention-focused actions, they are all elements of competitive advantage. In this way, a defensive commercial strategy oriented to retain and create loyalty bonds in existing customers is much more effective, and less costly, than an aggressive strategy that tried to expand the overall size of the market, attracting potential customers. Consequently, it is not surprising that the main companies are beginning to modify their commercial paradigm, moving from the massive capture of new customers to the conservation of existing ones.

However, this struggle for achieving customer loyalty collides with the grinding exposure to advertised offers from competitors that customers face every day. The customer's knowledge level and market awareness constantly increases and, as a result, so does his or her exigency. In this environment, the importance of understanding the underlying mechanisms of building loyalty bonds with customers becomes extremely important in order to ensure the continuity of the company in the market.

---

[1] And, additionally, even in a more specialized way, selectively acquiring high-value customers.

Increasing value axes

Guided commercial decisions design
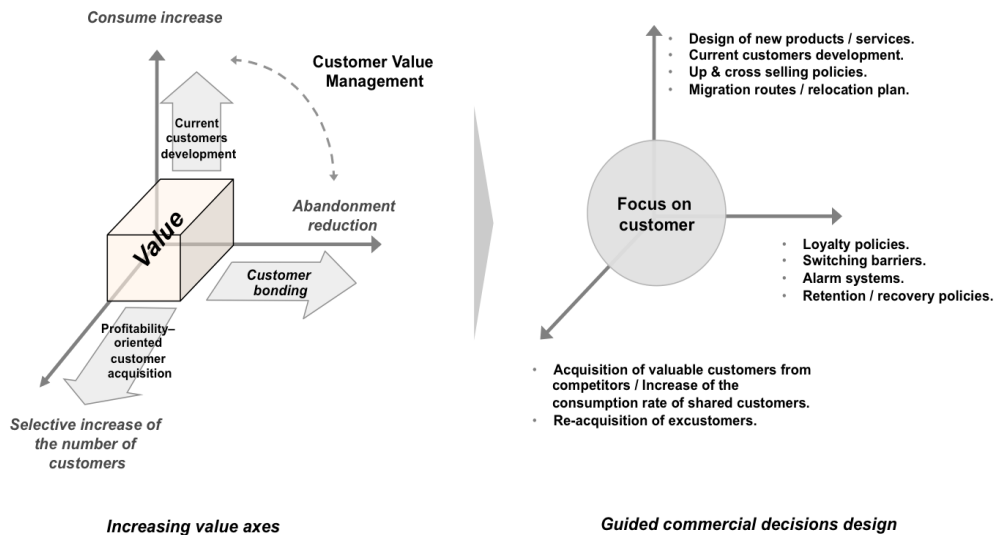
Figure 2.1: Commercial development alternatives in mature markets. Left) the figure shows the main axis of the development of commercial value in a mature market environment: on the one hand, Customer Continuity Management and Customer Development -aspects that complement and configure the so-called Customer Value Management- and, on the other hand, the development of selective strategies of high value customers acquisition. Right) the figure shows the customer-focused policies that can be developed for each one of the defined strategic axes.

Companies have their customers as their main assets and they are responsible for the definition and implementation of policies that allow them to reach and prolong their maximum commercial development potential. In other words, they must prolong as much as possible the life expectancy of their customer portfolio and assure its adequate development in terms of value, through the implementation of suitable commercial actions for each one of the stages of their lifecycle[2] (see *Figure 2.2*).
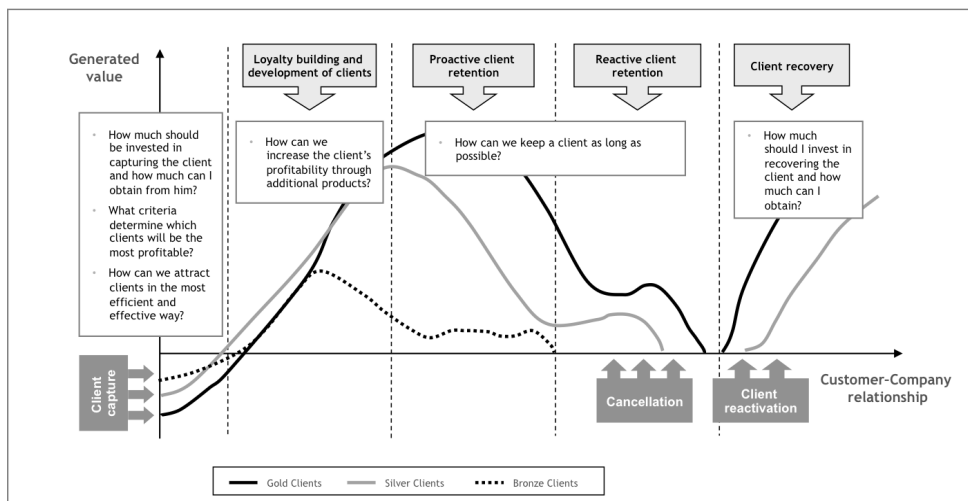
Figure 2.2: Stages in a client's lifecycle. The figure shows the generated value -ordinate axis- of three illustrative customer profiles -*gold*, *silver*, *bronze*- during their time of relationship with the company -abscissa axis-. Moreover, the figure shows the stages of customer-company interactions and the basic commercial aspects to solve in each one of the stages.

---

[2] Through the increase in services used (up-selling); the increase in consumption or wallet share (cross-selling); the construction of stronger loyalty bonds; the proactive retention actions on customers who intend to leave the actual provider; the launch of new products and services (innovation) and/or the adjustment of commercial costs, giving each costumer as expected.

Despite the fact that both dimensions in this figure: generated value and time of customer-company relationship, are strongly related in a unique *Customer Value Management* model, we understand that an appropriate development of customer's commercial value needs to guarantee its continuity[3] first [86]. Thereby, increasing customer's life expectancy should be the primary aim that determines and guides any posterior commercial action of value development. Adopting short-term commercial strategies, focused on a forced upgrade on customer's purchase levels rather than fulfilling their real needs, may lead to bad results in the mid-term, with high abandonment levels and lack of satisfaction that would damage company's image and reputation.

The final objective is self-explanatory: the commercial relationship with customers -the valuable ones-must be kept. For that purpose, companies should build strong schemes to avoid customer deflection. It should be borne in mind that companies and their customers are in a constant evolution, which drives to natural and unpredictable disruptions in the commercial relationship (change of home address, family lifecycle, change in interests, payment type, etc.). Thus, final success will not be based on lengthening our customers' lifecycle in an unnatural fashion, but on ensuring that good customers do not leave prematurely.

2.1.1 Customer continuity management

The complex task of prolonging customers' useful life cannot be improvised. The creation of loyalty bonds in customers requires a systematic approach to its management. Customers require regular check-ups to identify risks and threats, evaluate their evolution over time, and preventively anticipate the possible symptoms that might alert of a possible defection to the competition and that would require therapy policies (more or less aggressive according to the customer value). Therefore, the adoption of a suitable *Customer Continuity Management* model [86] should make it easier for companies to systematically approach a critical review of all the processes and procedures that affect the construction of true loyalty bonds with customers, including policies for everyday management, both for the customer lifecycles and for the predicted and declared cases of customer loss (see *Figure 2.3*).

As expressed in broad terms in the bibliographic review in *Section 2.2*, the level of customer bonding and, as a result, their life expectancy is intimately tied to the level of customer satisfaction relative to the service provided by the company. The higher the level of quality that customers perceive in the performance of the service, the stronger the loyalty bonds that are created [54, 132, 133, 139, 204, 216]. Thus, consumers who experience high levels of satisfaction in the service received usually continue with their current provider. We usually return to places where we have been treated with friendliness and courtesy, where we feel comfortable, where a "differentiated" product that we like is offered, where our demands are quickly dealt with, and so on.

Thus, satisfaction with the service acts as a base for customer loyalty, consolidating permanence and avoiding the substitution by another competitor [54, 132, 133, 139, 148]. In other words, if a company is able to give its customers a level of service that matches their expectations on an ongoing basis, these will not feel the need to change providers. So, continuous maintenance and improvement of the opinion that customers have of the service they are provided becomes the best and most efficient way of making them consistently loyal.

However, in some cases, although customer satisfaction has a positive influence on the level of bonding, it does not suffice. There are numerous situations in which better service quality does not seem to have an equal impact on consumer loyalty: customers that change mobile phone operators in spite of the fact that their current provider offers greater coverage; customers that fill up in slow service stations, with bad accesses and no additional services for the driver; customers who prefer to travel with certain airlines despite the continuous delays to their flights, etc. In consequence, there must be other factors, beyond satisfaction with service, influencing customer loyalty.

---

[3] Although it is equally true that a proactive customer's value development has a positive influence on its relationship with the company: high purchase and high value customers use to have a longer lifetime value.
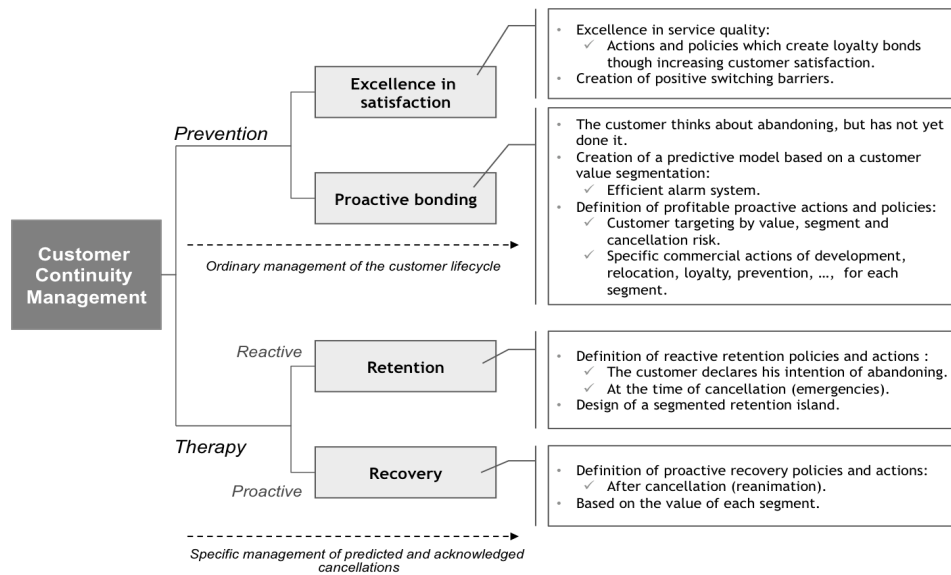
Figure 2.3: Customer lifecycle management model: Customer Continuity Management. The figure shows both a) prevention policies integrated in the ordinary customer management (upper part of the diagram), formed by the excellence in service quality, the creation of positive switching barriers and the development of proactive bonding policies, and b) the specific therapy policies for customer deflection (lower part of the diagram), which requires the creation of reactive retention policies, proactive recovery policies and a segmented retention island.

At this point, we introduce the concept of *barriers to change* as a construct that should mediate the satisfaction-loyalty relationship [54, 132, 133]. When the level of customer satisfaction with respect to different providers is similar, the level of bonding should be expected to depend, to a large extent, on the nature and strength of the *barriers to change* in place. The existing literature usually portrays the *barriers to change* in a negative sense, as difficulties and burdens -emotional, social or financial- that the customer must overcome when taking the decision to change providers[4] [54, 93, 132, 133, 141, 204]. However, in a market place that is becoming increasingly deregulated and competitive, the understanding of the construction of *barriers to change* as bureaucratic, contractual and/or in some cases, as the result of the abuse of a dominant position, is a short-sighted point of view. Barriers such as penalties when you cancel a given service; problems in the portability of mobile phone numbers; delays in the provision of the new service that are the fault of the old provider: all these are actions that are becoming steadily more regulated and penalized by the market and are not sustainable in the medium term.

To be sustainable, barriers to change must be built, like satisfaction, on customer perception. In this way, the active development of barriers to change becomes an excellence factor, in addition to satisfaction with the service, difficult to overcome by competitors in their attempt to attract the best customers. The construction of policies and procedures that maintain and improve excellence in both dimensions (satisfaction and barriers to change) should act as a powerful *vaccination* to protect customers from being lured by competitors (see *Figure 2.4*).

However, not all customers need the same level of service; nor they are all prepared to pay the same for it, or to obtain it in the same way. Common sense tells us that it is not possible to fulfill completely, in an increasingly inhomogeneous environment, the difficult task of developing the loyalty of all customers. For this reason, starting from the certainty that dissatisfied customers will always exist, companies must concentrate their efforts on the development of a broad-spectrum *vaccination* program, maintaining and improving those dimensions of the offer and barriers to change that most and best impact on the overall bonding of customers as a group. The objective is not to protect all customers, but rather as many of them

---

[4]Term more usually linked to the annoyance that a client must endure when changing to an alternative provider, which due to the additional perks that the company can build on its offer, make the other alternatives less attractive to the consumer.
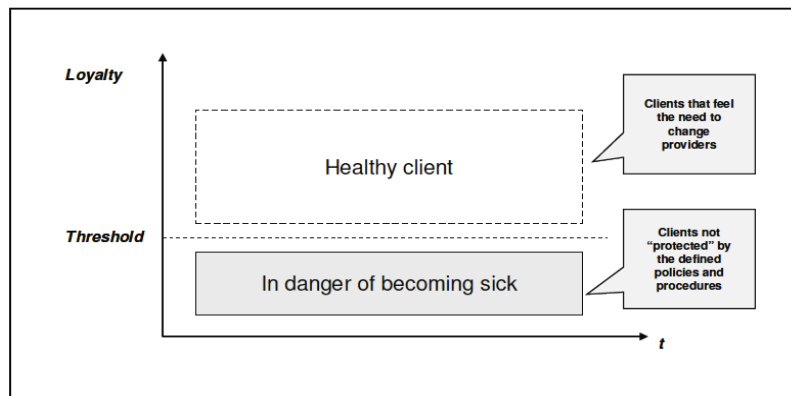
Figure 2.4: Illustration of the effect of developing satisfaction excellence policies and procedures and positive switching barriers.

as possible and, in particular, those who are most valuable to a given company.

It has to be born in mind that the protective effect of one such *vaccination* is never permanent. Over the natural life-span of the customers, it is possible that external changes, such as the appearance of new products, variations in competitors' offers, technological changes; and/or internal changes (improvement in the customer's knowledge level or increase in his exigency, socioeconomic changes, etc.) occur, which might affect customers' expectations and, as a result, their level of satisfaction. Companies must watch out for these changes to adapt their policies and procedures so that they can maintain and improve customers' opinions about the service on offer. The process of analyzing the dimensions with most impact on the satisfaction and the subsequent adjustment in commercial procedures and policies should become an ongoing process over time.

On the other hand, and from an operational perspective, it is not possible, given the high cost involved, to ask all customers from time to time for their opinions[5] on the satisfaction perceived of the service they are being offered and/or their level of bonding. Companies must therefore work with representative enough samples and develop, based on them, appropriate commercial policies.

Customers' evolution must be tracked and the number of customers at risk of churning must be estimated. That is why companies must have a reliable prediction model[6] that allows them to identify -with enough anticipation- those clients that show symptoms of propensity to switch service providers and, thus, launch efficient retention actions. Early diagnosis of the propensity to churn will reduce considerably the aggressiveness of the required loyalty bonding *treatment* and will increase the customer's *recovery* possibilities. In this context, the client's value[7] becomes the fundamental dimension that will determine which type of therapy, proactive and/or reactive, should be applied at any time.

This business effort -measured in the form of discounts, benefits and privileges that are offered to the client so that he will dismiss the idea of changing providers- should be balanced against the customer's expected value. This means that there can be clients that the company will decide not to retain even if their intention to change has identified in advance, since the expected return on the prolongation of their useful life does not justify the cost of the necessary commercial action. Identical criteria can be applied when deciding recovery policies and actions for already lost clients.

---

[5] Even more so in the case of companies with hundreds of thousands or even millions of active clients.

[6] Adapted to the market research and based on internal behaviour variables gathered systematically by the company.

[7] Understanding as client's value the sum of his actual recurrent value and his potential value, in the immediate and future dimensions.

2.2 Customer churn prevention: loyalty construction

The study of customers' loyalty bonds with a certain company or service provider has become one of the main focal points of marketing research in recent years, resulting in the definition of several models attempting to define and explain it. As a preliminary step prior to their review, we consider important to re-examine some of the concepts involved in the construction of these models, describing how different authors have defined and evaluated them. The concepts to be reviewed are as follows:

- Loyalty

- Satisfaction

- Service quality

- Price and sacrifice

- Value of service

- Costs and switching barriers

- Other: non-perceptive variables, socio-demographic and company characteristics, length of time as customer, indifference, inertia and level of expertise.

2.2.1 Basic concepts

Loyalty

The concept of loyalty has been interpreted and defined in many different ways in recent literature. While some authors have approached it solely from the viewpoint of customers' intentions to re-purchase, others have extended the approach to the study of customer repeat purchase intentions and recommendation to others. Let us take closer look at the most prominent approaches:

- Cronin et al. [54], Dabholkar et al. [58] and Kim et al. [139] refer to *loyalty* in terms of intention to re-purchase and tendency to speak well of the company, considering loyalty as the combination of an affective part -speaking well of the company is the objective result of a subjective attachment or feeling towards it- and a part related to behaviour: repeat purchase. With these two aspects of *loyalty* in mind, Lam et al. [148] separated and studied independently the effects of *recommendation* and *repeat purchase*. In a similar approach, Yin et al. [291] described loyalty as a combination of intention of re-purchase and the feeling of preference over competitors. Re-purchase and recommendation have also been used in recent years by Liu et al. [169], Lin and Wang [165] and Deng et al. [65], who added the intention to keep the same provider even if close friends recommended a different service.

- Gounaris and Stathakopoulos [96], working in the context of loyalty towards a brand, consider *loyalty* as a combination of purchasing behaviour, the emotional attachment to the brand and social influences. Within this framework, they define four possible types of loyalty:

    - *No loyalty*: no purchases at all, total lack of attachment to the brand and absence of social influences in favour.

    - *"Covetous" loyalty:* absence of purchase, but attachment and positive predisposition towards the brand. Brand recommendation -the customer does not buy for reasons beyond his reach-.

    - *Inertia loyalty*: purchase through habit or system, but with no attachment -very fragile and easily destroyed type of loyalty-.

    - *"Premium" loyalty:* high level of attachment to brand, high level of repeat purchasing and strong social pressure.

- Other literature provides us with further references from authors who have studied loyalty only from a *repeat purchase intention* viewpoint, such as [132, 133, 204, 216, 257]

- A different focus can be observed in the work of Chen and Hitt [44], who studied which motives led online stock market clients to "disloyalty". In this particular study, "disloyalty" is tackled as two possible behaviours: *change of company* and *stopping of activity*, considering that the causes which lead a client to change broker are different to those which lead them to stop their online stock trading activity.

- On the other hand, some authors, such as Fullerton [81], instead of studying loyalty directly, discuss customer *commitment* to the company as an antecedent to *loyalty*, the latter been understood as recommendation, intentions to change and customer readiness to pay more. The author distinguishes two types of commitment according to their causes: *affective commitment* -affective and attitude-related link of the customer to the company- and *continuance commitment* -commitment acquired due to lack of alternatives and/or due to the cost of change being perceived as too high-. According to the author, *affective compromise* is positive and the customer's relationship with the company should be based on this commitment, while *continuance commitment* is negative, and consequently, if the relationship is based on this commitment, customers remain loyal only because they feel obliged.

- Another author who addressed the commitment aspect is Pura [215]. In his study, he identified commitment concepts such as "lasting desire to continue the relationship" and *behavioural intentions* -intentions to repeat purchase and/or increase frequency of purchase-". In this study, the effect of *commitment* on *behavioural intentions* was investigated.

- More recently, Caceres and Paparoidamis [40] also mentioned commitment as an antecedent to loyalty in his study about business-to-business relationships. They adapted Morgan and Hunt [183] definition and based clients' commitment measurement on three concepts: *feeling involved* with their supplier, *feeling proud* of their supplier and *willingness to defend* it in front of others.

Satisfaction

Due to its influence on loyalty, *satisfaction* has been given much attention in literature of the field. As a starting point, we may consider satisfaction as the evaluation of an emotion, which reflects the extent to which a consumer believes that the purchase and/or use of a service arouses positive sensations [223]. However, recent literature on the subject shows varying nuances both in the definition and in the quantitative measurement of the concept. Let us look at the most significant:

- Cronin et al. [54] distinguish between "emotional" satisfaction and "rational" satisfaction. For this purpose, they use two groups of items to valuate customer satisfaction: one based on the emotions perceived by the customer in buying and/or using a service; and another based on the valuation made by the customer with regard to their choice of purchase and/or use of the service.

- Jones et al. [132] understand satisfaction in terms of the evaluation of the outcome on the basis of all previous experiences with the brand -as a way of distinguishing satisfaction with the service from satisfaction with those who provide it- [5, 26].

- Kim et al. [139] describe satisfaction as the reaction and judgement of the customer with regard to the company's level of compliance, incorporate two operational items in their study: *general satisfaction with the company* and *general satisfaction with the service*.

- Finally, two additional concepts of satisfaction can be identified in literature on the topic: *satisfaction with the specific transaction* and *accumulated or overall satisfaction* [28, 53, 227]. The former provides an immediate and specific vision, in contrast to the latter -which considers the satisfaction accumulated during the customer's entire life cycle- that gives a general vision of the service. Lam et al. [148] focused only on *accumulated satisfaction*. Analogously, Liu et al. [169] described *satisfaction* as the compound of overall *satisfaction with the service provided* and the *relationship with the service provider*.

Service quality

Another of the main targets of marketing studies in recent years has been the concept of *service quality*, which has inspired a large number of proposals and debates concerning its definition, measurement and evaluation:

- Grönroos [99] takes into account the difference between the service expected and the service received, identifying the three components which form *service quality*: functional service quality ("how?"), technical service quality ("what?") and image, based on factors such as tradition, ideology, prices and/or public relations.

- Parasuraman et al. [200, 201, 202] proposed service quality as the difference between the expectations and results of the aspects it is composed of, making an important contribution to the standardization of the measurement of the *service quality* perceived by customers. In 1988, they refined their first investigations publishing the well-known SERVQUAL scale, subsequently used in numerous studies. In this work, the ten original aspects of service quality were reduced to five final aspects: reliability, responsiveness, tangibles, assurance and empathy. In later work (1991 and 1994), although they adjusted the number of measurable items for each aspect, they kept the number of aspects as five. There have been numerous debates, among marketing experts, regarding the dimensionality of the SERVQUAL scale and on the appropriateness, or not, of measuring *service quality* as a distance (gap) between customer expectations and their evaluation of results [52, 202]. The general result of these debates seems to lead us to two important conclusions: on the one hand, a general consensus that is not necessary to measure the expectations of the results of a service in order to measure service quality [53, 295] and, on the other hand, the incapability of experts to resolve the underlying questions regarding the dimensionality of the SERVQUAL scale.

- Cronin and Taylor [52] presented an alternative to the SERVQUAL scale, known as the SERVPERF scale. According to the authors, *service quality* must be measured and conceptualized as an attitude. In their study they explain how the measurement scale based on SERVPERF performance reduces the number of SERVQUAL items by 50% while providing better results.

- In a posterior study, Cronin et al. [54] also consider general service quality as an independent dimension. In this case, two forms of measurement were used to measure the *service quality* of several items:

  - *Service Quality Performance (SQP):* Consisting of 10 questions (see *Table 2.1*) derived from the ten original aspects proposed by Parasuraman et al. [200].

Service Quality Performance (scaling from "very low" to "very high" on a 9-point scale)	
1.	Generally, the employees provide service reliably, consistently, and dependably
2.	Generally, the employees are willing and able to provide service in a timely manner
3.	Generally, the employees are competent (i.e., knowledgeable and skillful)
4.	Generally, the employees are approachable and easy to contact
5.	Generally, the employees listen to me and speak in a language that I can understand
6.	Generally, the employees are courteous, polite and respectful
7.	Generally, the employees are trustworthy, believable, and honest
8.	Generally, this facility provides an environment that is free from danger, risk or doubt
9.	Generally, the employees make the effort to understand my needs
10.	Generally, the physical facilities and employees are neat and clean

Table 2.1: Items of "Service Quality Performance" used by Cronin et al. [54].

– ***Overall Service Quality (OSQ):*** Consisting of three direct and general measures of the *overall service quality:* poor *vs.* excellent; inferior *vs.* superior; and low standards *vs.* high standards (see *Table 2.2*)

Overall Service Quality
"Poor" 1 2 3 4 5 6 7 8 9 "Excellent"
"Inferior" 1 2 3 4 5 6 7 8 9 "Superior"
"Low Standards" 1 2 3 4 5 6 7 8 9 "High Standards"

Table 2.2: Items of "overall service quality" used by Cronin et al. [54].

- Other contributions to the study of service quality may be found in Dabholkar et al. [58], who considers that customers evaluate the different factors related to service: reliability, personalized service, comfort, characteristics, which they hold as antecedents to general *service quality*: a dimension that should be evaluated separately -not as the sum of its components-.

- In their work in the telecommunications industry context, Ranaweera and Neely [216] understand that *perceived service quality* could be broken down into eight basic aspects: courtesy, capacity, ease of contact, reliability, security, service package, understanding and recuperation service. They used a scale of 12 points -adapted from the scale used by Cronin et al. [54]- to value each one of the eight aspects, considering its average as the level of *perceived service quality* (see *Table 2.3*).

SQ perceptions (scaling from "Strongly agree" to "Strongly disagree")	
1.	My phone company always keeps me informed of things that I need to get the best use of the service
2.	My phone company staff make an effort to explain things in a simple way
3.	I am sure that my phone company will suit my needs best in the future
4.	I have no doubts about the future existence of my phone company
5.	My phone company staff are capable
6.	My phone company staff are courteous
7.	Whenever something goes wrong, my phone company takes corrective action without delay
8.	It is easy to contact my phone company whenever necessary
9.	My phone company understands my needs best
10.	My phone company is concerned about my salary
11.	My phone company's service is reliable (service is available whenever I want it)
12.	My phone company offers all the services I expect from a phone company

Table 2.3: Items of "service quality perceptions" used by Ranaweera and Neely [216].

- For their part, Kim et al. [139], also in the telecommunications market, measured *service quality* as: the quality of the call, the price structure -reasonable prices-, the mobile device, the value added services, the convenience of procedures (subscription and change) and customer support. The used items appear in *Table 2.4*.

Variable Service Quality	Operational definition	Measurement items
Call quality	Call quality according to customer perception	Call clarity Coverage
Pricing structure	Pricing and price schedule	Reasonability of price Variety of price schedule Possibility of freely choosing price schedules
Mobile device	Mobile device functionality and design	Quality of mobile device Variety of mobile device types Quality of mobile device design
Value-added services	Type and convenience of value-added services	Variety of value-added services Convenience of use of value-added services Whether value-added services are up-to-date
Convenience in procedures	Subscription and change procedures	Ease of subscribing and changing service Staff friendliness, when subscribing and changing
Customer support	Customer support system and complaint processing	Variety of customer support systems Speed of complaint processing Ease of reporting complaint Friendliness when reporting complaint

Table 2.4: Operational definitions and measurement of "service quality" used by Kim et al. [139].

- Lam et al. [148], through consultations and interviews with agents and managers of a courier firm and based on existing literature on *service quality* measurement [200] selected eight initial attributes which were later reduced to five (see *Table 2.5*).

Item Service Quality Attributes	Description
Q1	Understanding of my business and shipping needs by the staff
Q2	Timeliness of pickup of consignments as promised
Q3	Reliability in delivering shipments (accurately, on time, etc.)
Q4	Ease of booking a shipment with a company
Q5	Promptness in advising about any problems with shipments

Table 2.5: "Service quality attributes" used by Lam et al. [148].

It is important to highlight that they did not use *service quality* as a direct precursor to satisfaction: instead, they took a weighted average -the weighting was given by those interviewed- of the *service quality* and the *perceived price* they calculated the *service value*, which they *did* hold as a precursor to *satisfaction*.

- Further, Fullerton [81] -in line with the argument of Brady and Cronin [32]- concluded that although the SERVQUAL scale is a widely used *service quality* measurement system, one of its serious limitations is that it does not deal with the whole spectrum of questions and attributes by which consumers evaluate service quality. In order to solve this problem, he considers three aspects as antecedents to *overall service quality*: *interaction quality* -encounter between customer and provider -*quality of results* -customer evaluation of the service results, including supplier punctuality- and *environmental quality*- tangible characteristics of the place of service- [32]. According to Fullerton, these antecedents can also have some sub-aspects, which will depend on the characteristics of the company and on the sector to be analyzed. The items used in the study to measure *service quality* are summarised in *Table 2.6*.

Interaction Quality	
1.	I would say the quality of my interaction with Xs employees is high
2.	You can count on the employees of X being friendly
3.	Xs employees respond quickly to my needs
Environment Quality	
1.	Xs physical environment is one of the best in its industry
2.	Xs layout never fails to impress me
3.	At X, I can rely on there being a good atmosphere
Outcome Quality	
1.	I always have an excellent experience when I visit X
2.	I can count on X to keep my waiting time to a minimum
3.	I am consistently pleased with the selection at X
Service Quality	
1.	I believe the general quality of X's services is low (RC)
2.	Overall, I consider Xs service to be excellent
3.	The quality of Xs service is: (1=poor; 7=excellent)

Table 2.6: Dimensions and items of "service quality" used by Fullerton [81].

- The fast-changing environment experienced by internet and online services in recent years forced researchers to adapt their scales to the new needs of customers. Thus, in the field of e-commerce and online services, Parasuraman et al. [203] created the E-S-QUAL scale to measure service quality delivered by online companies, which consisted of four dimensions: *efficiency, fulfillment, system availability* and *privacy*, detailed in *Table 2.7*.

Efficiency	
1.	The site makes it easy to find what I need
2.	It makes it easy to get anywhere on the site
3.	It enables me to complete a transaction quickly
4.	Information at this site is well organized
5.	It loads its pages fast
6.	This site is simple to use
7.	This site enables me to get on to it quickly
8.	This site is well organized
System availability	
1.	This site is always available for business
2.	This site launches and runs right away
3.	This site does not crash
4.	Pages at this site do not freeze after I enter my order information
Fulfillment	
1.	It delivers orders when promised
2.	This site makes items available for delivery within a suitable time frame
3.	It quickly delivers what I order
4.	It sends out the items ordered
5.	It has in stock the items the company claims to have
6.	It is truthful about its offerings
7.	It makes accurate promises about delivery of products
Privacy	
1.	It protects information about my Web-shopping behavior
2.	It does not share my personal information with other sites
3.	This site protects information about my credit card

Table 2.7: Items of E-S-QUAL scale used by Parasuraman et al. [203].

- Bell et al. [16] believe that in mature industries characterized by relatively undifferentiated products, very often it is *service quality* which distinguishes one organization from another. In their study, they analyse the financial sector, highlighting two aspects of quality: *technical service quality* and *functional service quality*.

  - **Technical Service quality.** Aspects related to the service result -the quality and exactness of the advice, achievement or profitability expectations-. A scale of four items, based on Sharma and Patterson [228], developed specifically for the financial services industry was used (see *Table 2.8*).

Technical Service Quality
1. My adviser has assisted me to achieve my financial goals
2. My adviser has performed well in providing the best return on my investments
3. My adviser has helped me to protect my current position by recommending the best investing options
4. My adviser has performed well in investing my money in appropriate investment options

Table 2.8: Items of "technical service quality" used by Bell et al. [16].

  - **Functional Service Quality.** Elements related to the service delivery process –accessibility and empathy of the service provider-. They adapted a scale of five items by Hartline and Ferrell [107], ultimately obtaining a scale of three items (see *Table 2.9*).

Functional Service Quality
1. My adviser gives me personal attention
2. My adviser has my best interests at heart
3. I can share my thoughts with my adviser

Table 2.9: Dimensions and items of "functional service quality" used by Bell et al. [16].

- In line with Bell et al.'s approach, Caceres and Paparoidamis [40], in their study about business-to-business relationship between advertising agencies and their clients, also described *service quality* as the sum of *technical quality* and *functional quality*. In this case, *technical quality* was formed by the attractiveness and adequacy of the advertising campaign, while *Functional quality* -named as commercial service- expands Bell's definition including three antecedents, *communication*, *service delivery* and *administrative service*, as it can be seen in *Table 2.10*.

	Technical Quality
1.	Attractiveness of the advertising campaign proposed
2.	The advertising campaign proposed reflected sufficiently your brand image
	Functional Quality
	Communication
1.	Your supplier informs you sufficiently for the potential internet applications
2.	Your supplier provides clear information concerning the capability of his company concerning internet applications
	Service delivery
1.	Your supplier is aware of your needs concerning distribution of advertising material
2.	The delivery of advertising materials is always on time
	Administrative service
1.	Orders are confirmed on time
2.	The terms of contracts signed are always clear
3.	Invoices sent from your supplier are always clear
4.	Invoices sent from your supplier are always precise

Table 2.10: Items of *service quality* defined by Caceres and Paparoidamis [40].

- For their part, Yin et al. [291] also adapted SERVQUAL [200] scale to the context of hair salons and fast food restaurants, aiming to compare the perception of service quality on two services where the relationship between customers and staff is very different. To that end, they created the 5-item scale shown in *Table 2.11*.

	Service Quality
1.	The staff (hair stylist) always tries to meet your needs
2.	The food (product) quality of this restaurant (hair salon) is good
3.	The staff provides prompt service in taking order and payment / The hair stylist is responsive to your questions and requests
4.	The staff provides accurate service in taking order and payment / The hair stylist provides reliable hair cutting service
5.	The staff (hair stylist) is consistently courteous with you

Table 2.11: Items of *service quality* defined by Yin et al. [291].

- Finally, in a further fast-evolving industry such as telecommunications, Malhotra and Malhotra [176] created the m-SERVQUAL scale to measure service quality perception towards mobile service providers. Based on the original SERVQUAL, preserves the *reliability* and *responsiveness* dimensions and adds: *digital services* offered and *flexibility* (see *Table 2.12*).

Technical reliability	
1.	Allows me to make and receive calls without wait/ interruption
2.	Allows me to make voice calls that are clear
3.	Delivers the service promised
4.	Has excellent connection quality everywhere
5.	Does not drop calls
In-store responsiveness	
1.	Has in-store customer service reps who can offer advice about service plans
2.	Has in-store customer service reps who are knowledgeable
3.	Has physical store locations that are pleasant
4.	Has in-store customer service reps who can resolve my problems and issues
Phone responsiveness	
1.	Has telephone customer reps who can resolve billing problems
2.	Has helpful telephone customer reps
3.	Has telephone customer reps who can solve technical issues
4.	Has telephone customer reps who are knowledgeable
Online service facilitation	
1.	Allows me to easily check my account using a website
2.	Allows me to easily manage my account using a website
Service flexibility	
1.	Lets me change/ upgrade my service plan easily
2.	Lets me change/ upgrade my cell phone easily

Table 2.12: Items of m-SERVQUAL scale used by Malhotra and Malhotra [176].

Price or sacrifice

It is also common to find references from authors who have investigated the fundamental role played by customer perception of _price_ or _sacrifice_ in customer _loyalty_; that is, the effort, time and money necessary to acquire a certain product or service:

- For example, Cronin et al. [54] -in line with the definitions by Heskett et al. [111] and Zeithaml [293]- consider _sacrifice_ as that which is given or sacrificed in order to acquire a service. The monetary price is explicitly measured and the non-monetary price is evaluated using direct measurements of time and effort (see _Table 2.13_).

Sacrifice (scaling from "very low" to "very high" on a 9-point scale)	
1.	The price charge to use this facility is …
2.	The time required to use this facility is …
3.	The effort that I must make to receive the services offered is …

Table 2.13: Items of "sacrifice" used by Cronin et al. [54].

- Ranaweera and Neely [216], after studying the literature related to service provision, highlights in his study that sufficiently tested forms of measuring _price perceptions_ cannot be found. In this regard, and in accordance with the work of Varki and Colgate [256] and Drolet and Morrison [69], he uses only one item to evaluate how reasonable the prices of a company are, compared to competitors: _"the prices charged by my telephone company are reasonable"_ (entirely agree _vs._ entirely disagree).

- More recently, Kim et al. [139] considered _price structure_ a component of _service quality_. The items used to measure this variable were: _reasonableness of prices_, _variety in tariff plans_, and _possibility of choosing freely between tariff plans_.

- Finally, Lam et al. [148], in accordance with Naumann [186], consider that the *sacrifice* or *price* that a client pays can be broken down typically into: transaction costs, costs over the client life cycle and some level of risk. The items used in the study to measure perceptions regarding price were[8] (see *Table 2.14*).

Price attributes	Description
P1	Shipment costs incurred by your company (i.e., rates charged for actual services by the courier firms)
P2	Shipment preparation costs incurred by your company (i.e., printing, labeling, filling shipping forms, etc.)
P3	Delay costs incurred by your company (i.e., costs related to fixing shipment delays, etc.)
P4	Communication costs incurred by your company (i.e, costs of telephone, fax. etc. in dealing with the courier firms)
P5	Costs incurred by your company in fixing problems with the courier forms' invoices and so on

Table 2.14: Description of "price attributes" used by Lam et al. [148].

Service value

Three different viewpoints will help us to understand the *value of a service* [222], although only the last two are relevant to the creation of value for the customer:

- *Value for the company*: understood as the achievement of the maximum profit by the company.

- *Value offered by the company*: as offered to the customers so that they chose the competitive offer of the company in question.

- *Value perceived by the customer*: This completely subjective value depends on the final judgement of the customer.

Excellence in the creation and delivery of *value* for the customer has become a key factor in obtaining sustainable competitive advantages. This "superior" *value* implies, as Weinstein and Johnson [283] state, "the constant creation of business experiences which exceed customer expectations". Authors such as Band [11] and Butz and Goodstein [39] establish the following scale regarding the level of *offered value* which a product or service should provide:

- *Expected product or service*: complies with the minimum characteristics for entry into the market.

- *Perfected product or service*: additional characteristics are added that, although unexpected by the customer, are still welcome.

- *Excellent product or service*: characteristics unimaginable for the customer are added, representing all that is necessary to attract and retain customers, increasing the differential value of the service compared to competitors. Obviously, this level generates a stronger link with the customer.

On the other hand, it is fundamental for a company to know the perception that a customer has of the product or service it offers, given that this may not coincide with what the company believes it is offering. For this reason, the majority of studies pay special attention to the *value perceived by the customer*.

Most investigations highlight the existence of two main dimensions in perceived value. On the one hand, the benefit a customer perceives when they acquire a product or service and, on the other, the sacrifices this acquisition implies.

Investigations differ in the components that make up each of these dimensions, although there is a general tendency to consider the *price* of the product or service within the benefits of the *quality service*

---

[8] Valuing each item on a scale of 1-10 (were 1 = totally dissatisfied, 10 = totally satisfied).

perceived, and within the sacrifices. There are also two conflicting viewpoints which consider these dimensions as an antecedent of perceived *value* -which in this case can therefore be measured directly- and as components of provided *value* -which in this case should be measured as the sum of these components- [222]:

- Zeithaml [293] defined four possible interpretations about how customers base their evaluations on service value. Later, Cronin et al. [54] summarized in one single definition -*perceived value* is the consumers' overall assessment of the utility of a product based on perceptions of what is received and what is given- Zeithaml's work. Cronin included for the purpose two direct systems for measuring service value (see *Table 2.15*).

Service Value (scaling form "very low" to "very high" on a 9-point scale)
1. Overall, the value of this facility's services to me is …
2. Compared to what I had to give up, the overall ability of this facility to satisfy my wants and needs is …

Table 2.15: Items of "service value" used by Cronin et al. [54].

- Lam et al. [148], inspired by the work of Heskett et al. [111], understand customer *value* as a trade-off between the attributes corresponding to "what is obtained" and those corresponding to "what is given in exchange", although in operational terms they use the method for measuring customer value developed by Gale [83]. This method has the advantage of providing a profile of the company in relation to competitors, where service/product attributes and prices are concerned. According to this method, the *value* perceived by the customer is calculated as follows:

Value = (General Relative Quality Level $\times$ Quality Weight) + (Relative Price Level $\times$ Price Weight)

The level of general relative quality is calculated by dividing the level of company *service quality* (valued from 1 to 10) by the average of the competitors' *service quality* level (valued from 1 to 10). The same applies in the case of price. Both the weight of quality and price are provided by the person interviewed.

- More recently, Pura [215] used the work by Sheth et al. [230], which included the identifications of five aspects of *value*, as a starting point:

  - ***Functional Value***: It represents the value derived from the effective fulfillment of work. It is often related to monetary value or superiority compared to other alternatives [230]. However, other matters must be addressed within functional value apart from the fulfillment of work, such as time and money saving [179] or convenience, understood as the ease of use, speed of acquisition, etc. [3, 42, 43].

  - ***Social Value***: This is related to social approval and self-image improvement in the eyes of others [13]. The work of numerous investigators reinforces the importance of social reputation [20, 116, 230, 242] for self-esteem. Some theories also mention fashion, status and sociability, relating them to aspects of social value, highlighting, for example, that the use of mobile telephone services may serve to express personality, status and image in a public context [159]. Finally, Sweeney and Soutar [242] define social value as the "utility derived from the capacity to improve the concept itself on a social level". Hence, social value is derived from the product or use of the service shared with others [230].

  - ***Emotional Value***: This is achieved when a product/service arouses feelings or emotional states in the consumer [230, 242]. For example, the search for pleasure and fun are reasons related to emotional value [116] that strongly influence the decision to use mobile phone services [159], since the use of technology itself often increases positive sensations, independently of the service used [34].

– **_Epistemic Value_**: This is related to the curiosity felt and the novelty or knowledge acquired. Curiosity [230], novelty and the search for variety [113] are among the main reasons for seeking and purchasing a certain product or service. Customers motivated by epistemic value often return to their habitual consumption patterns after satisfying the need for change and exhausting the effect of novelty [230].

– **_Conditional Value_**: It originally refers to circumstances affecting choice. Such circumstances may be stationary, events which occur once in a lifetime or emergency situations [230]. Holbrook [116] defended that conditional value depends on the context in which the value is judged and exists only in specific situations. In this respect, conditional value will depend on the concept of "context", which is understood as time, location and social environment, available equipment, technological environment and specific user criteria -such as mood, work or free time- [141].

Given that Sheth's work [230] does not provide measurement items to validate his perceived value model in the electronic self-service market context, Pura [215] complements his investigation with the work of others to support the detailed definition of these aspects (see *Table 2.16*).

Constructs	Items and their sources
Monetary value	Adapted from Chen and Dubinsky (2003, Dodds and Monroe (1991) and Sweeney and Soutar (2001) The price of this mobile service is acceptable This mobile service is good value for money This mobile service is better value for money that what I would pay for the same service via internet
Convenience value	Adapted from Anderson and Srinivasan (2003) and Mathwick et al (2001) I value the ease of using this mobile service Using this mobile service is an efficient way to manage my time I value the possibility to use this service instantly via my mobile device I value the convenience of using this mobile service
Social value	Adapted from Soutar and Sweeney (2003) and Sweeney and Soutar (2001) Using this mobile service helps me to feel accepted by others Using this mobile service makes me a good impression on other people Using this mobile service gives me social approval
Emotional value	Adapted from Soutar and Sweeney (2003) and Sweeney and Soutar (2001) using this mobile service gives me pleasure Using this mobile service makes me feel good
Epistemic Value	Adapted from Donthu and Garcia (1999) I used this mobile service to experiment with new ways of doing things I used this mobile service to test the new technologies I used this mobile service out of curiosity
Conditional value	(Created for this study) I value the information this service offers, with the help of which I get what I need in a certain situation

Table 2.16: Constructs, items and their sources of "Service Value" used by Pura [215].

- Likewise, Deng et al. [65] also based their *customer value* measurement on Sheth et al. [230] work, keeping the definitions on *functional value*, *emotional value* and *social value*, and adding *monetary value* as a significant factor. They measured monetary value as the perception of paying an economic price for a valuable service.

Costs or switching barriers

The study of costs or switching barriers arose in the context of investigation in industrial organizations and business strategies. Several authors, pioneers in the field of management such as Day [61], Porter [212] and

Aaker [1] began to develop the concept of customer loyalty through the construction of switching barriers. One of the first definitions was that of Porter [212], who defined the costs of switching as *"those which are associated with the movement from one provider to another"*.

Despite the fact that diverse investigations of an economic, strategic and marketing nature have appeared in recent years with the aim of classifying the different switching costs that customers and companies face, there is no consensus regarding which is the most appropriate categorization. In general, the idea of switching costs is viewed as the "difficulty" associated with changing to a new product, service or system [93], a highly subjective and emotional concept, which is not easily evaluated [284]. In addition, we must take into account the fact that switching costs differ in composition and nature according to the context and sector analyzed, also varying in relation to customer characteristics.

In an early study, Jones et al. [132] described switching barriers as any factor which makes changing providers difficult or expensive for the customer. They examined three barriers in the context of consumer services: interpersonal relationships, switching costs perceived and the attractiveness of alternatives. Such barriers are common in the context studied -banking services and hairdressers- given their high level of personalization and dispersed geographical nature. This classification is very similar to that which Kim et al. [139] later developed, although their work is based on the idea that switching barriers refer to the difficulties of changing to another provider for a customer who is dissatisfied with the current service, or to the financial, social and psychological burdens perceived by a customer on changing companies [78]:

- **Interpersonal relations:** Psychological and social relations such as care, trust, privacy or communication [98]. They refer to the intensity of the links developed between customers and the employees who provide the service [19, 251]. Interpersonal relations are particularly important when we are referring to service provision, given the high level of personal interaction it implies, the intangible nature of the service itself, the heterogeneity of the result and the prominent role played by customers in the service production [31, 57]. Interpersonal relations constructed through recurring interactions between a company and customer build a link between them and ultimately lead to a long term relationship. Investing in a relationship helps to increase customer dependency and therefore increases switching barriers. The items used by Jones et al. [132] to measure the effect of interpersonal relations in the banking sector are compiled in *Table 2.17*.

Interpersonal relationships
1. I feel like there is a "bond" between at least one employee at this bank and myself
2. I have developed a personal friendship with at least one employee at this bank
3. I have somewhat of a personal relationship with at least one employee at this bank
4. I am friends with at least one employee at this bank
5. At least one employee at this bank is familiar with me personally

Table 2.17: Items of "interpersonal relationships" in the banking sector, as used by Jones et al. [132].

- **Perceived switching costs**: These correspond to the consumer's perception regarding the time, money and effort that entails changing service provider [67]. Such costs may be associated with the search for alternatives, such as the learning process -both for the customer and the new service provider- [100]. The items used by Jones et al. [132] to measure switching costs in the banking sector were as follows (see *Table 2.18*):

Switching costs
1. In general it would be a hassle changing banks
2. It would take a lot of time and effort changing banks
3. For me, the costs in time, money, and effort to switch banks are high

Table 2.18: Items of "switching costs" in the banking sector, as used by Jones et al. [132].

Kim et al. [139], in their study in the field of mobile telephone services, subdivided switching costs into three new categories:

- **Loss costs**: Perception of loss of social status or profitability on cancelling the service contract with the service provider.

- **Adaptation costs**: Costs related to the search for and/or process of learning about new alternatives.

- **Installation costs**: Economic costs -such as the purchase of a new device or payment of subscription fees- involved in changing to a new company.

For their part, Liu et al. [169] adapted Kim et al. categories in their study on mobile services, reorganizing them in only two items: *economic loss*, as the sum of monetary costs related to switching provider, and *psychological burden*, as the reluctance to losing social status searching for alternatives.

- **Attractiveness of alternatives**: This refers to customer perceptions regarding the availability of feasible alternatives in the market. It has to do with the *reputation, image* and *service quality* of the alternative companies, which are expected to be superior or more appropriate than those of the current service provider. The attractiveness of alternatives is closely linked to service differentiation and competitive pressure. If a company offers differentiated services that are difficult for a competitor to equal, or if there are few alternatives on the market, customers will tend to stay with the current company [17]. The items used by Jones et al. [132] to measure switching costs in the banking sector were as follows (see *Table 2.19*):

Attractiveness of Alternatives
1. If I needed to change banks, there are other good banks to choose from
2. I would probably be happy with the products and services of another bank
3. Compared to this bank there are other banks with which I would probably be equally or more satisfied
4. Compared to this bank, there are not very many other banks with whom I could be satisfied (Reverse Coded)

Table 2.19: Items of "attractiveness of alternatives" in the banking sector, as used by Jones et al. [132].

In a later study, Jones et al. [133] -as Patterson and Smith [204] a year later- used the switching barrier classification presented by Guiltinan [100], and grouped switching costs into three new categories: continuity, learning and already invested costs:

- **Continuity costs:** This refers to the probability of the loss of benefits and privileges granted by the current service provider. In their investigations, both Jones et al. [133] and Patterson and Smith [204] were of the opinion that continuity costs can be subdivided into:

  - **Costs of losing benefits and privileges:** customers with a repetition pattern known by the company usually accumulate special benefits, preferential service, special favours, etc. These special benefits would be lost if the relationship with the current service provider were to end [177, 251], implying clear disincentives for change [14, 110].

  - **Risk perception:** This refers to the psychological uncertainty or perception of risk regarding whether or not the new provider -not tested- will be on the same level as the current provider [100, 225, 292]. The risk and uncertainty are greater when the quality is difficult to judge or varies considerably between alternatives. Therefore, risk perceptions in services stand out due to their intangibility and heterogeneity [294].

- **Learning costs:** These include the time and effort necessary to acquire information and for the exchange and evaluation of a new provider. Both Jones et al. [133] and Patterson and Smith [204] coincide in the first two aspects into which these costs may be sub-divided:

– **Pre-switch search and evaluation cost:** These represent the consumer's perception regarding the time and effort -prior to changing- necessary to search for information on the available alternatives and evaluate their feasibility. The inclusion of these costs is justified by the service characteristics: geographical spread, limitation of alternatives per region, intangibility of the service and impossibility of separating production and consumption [292].

– **Launch costs**: This refers to the perception of the time necessary and the inconvenience involved in training a new provider. When the level of personalization is high, as is usually the case in services, an additional learning process is necessary for the service to be provided satisfactorily: *the learning process of the service provider*. These costs often fall back on customers [100, 124, 125, 212].

They differ in the third, though: while Jones et al. [133] consider the post-switch knowledge and behaviour costs:

– **Post-switch knowledge and behaviour costs**: These are consumer perceptions regarding the time and effort necessary to adapt to the procedures and routines of the new alternative. This is particularly relevant in the case of services, since consumers generally play a fundamental role in routines and procedures [31, 110].

Patterson and Smith [204] also considers the attractiveness of the alternatives:

– **Attractiveness of alternatives:** Customer evaluation of the likely satisfaction that may be achieved from the alternative relationship [210]. The existence of alternatives is the key factor in defining dependence [72, 244]. In other words, if a customer is unaware of the attractiveness of the alternatives or simply does not perceive them as more attractive than the current relationship, then it is more than likely that they will stay with their current relationship, even when this is perceived as unsatisfactory.

• **Already invested costs:** These include investments, economically irrelevant but psychologically important, prior to changing relationships [68, 100]. More specifically, they represent the customer perception of the time and effort they have already invested in establishing and maintaining a friendly relationship with a certain service provider. Therefore, avoiding the psychological and emotional stress involved in ending such "almost social" relationships will motivate some clients.

The final items used by Patterson and Smith [204] to evaluate the aforementioned aspects are presented in *Table 2.20*. The items used by Jones et al. [133] may be observed in *Table 2.21*.

Special treatment benefits	
1.	Will go out of their way to search for a special deal for me
2.	Will always search for the most reasonably priced solution
3.	Will more likely help me if something goes wrong
4.	Will be more to do what I want
Risk Perceptions	
1.	If I change, there is a risk the new . . . . . . won't be as good
Search costs	
1.	On the whole, I would waste a lot of time searching for another . . . . . . if I changed . . .
Attractiveness of alternatives (four items)	
1.	All . . . are much the same, so it would not matter if I change it
2.	All . . . offer similar range of services
3.	All things considered, most . . . are similar
4.	All . . . give a similar level of service
Need to explain preferences	
1.	If I change, I will need to spend a lot of time to explain my preferences to a new . . .
Loss of interpesonal relationship	
1.	I will lose a friendly and comfortable relationship if I change

Table 2.20: Items of "switching barriers" used by Patterson and Smith [204].

The study of the online stock broking industry by Chen and Hitt [44] presents two sides. On one hand, it provides an explanation of "disloyalty" -switching or cancellation- based on the usage characteristics and demographics of the customer and of the characteristics of the service itself and, on the other hand, the calculation of costs of switching from one of the firms analyzed to another.

Using the classification of Klemperer [141] as a base, three types of switching costs were identified: *transaction costs*, *learning costs* and *contractual (or artificial) costs*. The transaction costs occur when a relationship with a new provider is started and, at times, they also include the costs needed to terminate the existing relationship. The learning costs represent the effort required by customers to find the same level of comfort with the new provider that they had with the old one. The artificial costs are created by the firms themselves through deliberate actions: flyer programmes, repeat purchase discounts, rewards for clicking through, etc. Beyond these explicit costs, the study also identified implicit switching costs associated with decision-making trends and the desire to avoid risk, especially when the customer perceives uncertainty in the quality of other products or brands.

The analysis carried out by Lam et al. [148] considers switching costs -monetary and non-monetary-implied in changing to another provider [109]. In their study, the domain of switching costs also includes the loss of benefits derived from a commercial relationship that is brought to an end [109, 128]. In this respect, the switching costs could conceptually reflect a dependency of the buyer on the seller, which is materialized in the buyer's need to maintain the relationship with a provider in order to reach his or her objectives [80]. The items that their study used to evaluate the different dimensions were the following (see *Table 2.22*):

	Pre-switching search and evaluation costs
1.	It would take a lot of time and effort to locate a new hairstylist/ barber
2.	If I changed hairstylist/ barbers, I would not have to search very much to find a new one
3.	If I stopped going to my current hairstylist/ barber, I would have to search a lot for a new one
4.	It takes a great deal of time to locate a new hairstylist/ barber
5.	If I stopped using my current hairstylist/ barber, I would have to call and look around for a new one to use
	Costs of lost performance
1.	This hairstylist/ barber provides me with particular privileges I would not receive elsewhere
2.	By continuing to use the same hairstylist/ barber, I receive certain benefits that I would not receive if I switched to a new one
3.	There are certain benefits I would not retain if I were to switch hairstylists/ barbers
4.	I would lose preferential treatment if I changed hairstylists/barbers
	Uncertainty costs
1.	i am not sure what the level of service would be if I switched to a new hairstylist/ barber
2.	If I were to change hairstylists/ barbers, the service I might receive at the new place could be worse than the service I now receive
3.	The service from another hairstylist/ barber could be worse that the service I now receive
	Post-switching behavioural and cognitive costs
1.	If I were to switch hairstylists/ barbers, I would have to learn how things work at a new one
2.	I would be unfamiliar with the policies of a new hairstylist/ barber
3.	If I changed hairstylists/ barbers, I would have to learn how the "system works", at a new one
4.	Changing hairstylist/ barber would mean I would have learned about the policies of a new one
	Sunk costs
1.	A lot of energy, time, and effort have gone into building and maintaining the relationship with this hairstylist/ barber
2.	Overall, I have invested a lot in the relationship with this hairstylist/ barber
3.	All the things considered, I have put a lot into previous dealings with this hairstylist/ barber
4.	I have spent a lot of time and money at this hairstylist/ barber
5.	I have not invested much in the relationship with this hairstylist/ barber
	Setup costs
1.	If I changed hairstylist/ barber, it would take a lot of time and effort on my part to explain to the new hairstylist/ barber what I like and what I want
2.	If I changed hairstylists/ barbers, I would have to explain things to my new hairstylist/ barber
3.	There is not much time and effort involved when you start using a new hairstylist/ barber

Table 2.21: Items of "switching barriers" used by Jones et al. [133].

Item	Description
Switching cost	
SW1	It would cost my company a lot of money to switch from DPS to another courier firm
SW2	It would take my company a lot of effort to switch from DPS to another courier firm
SW3	It would take my company a lot of time to switch from DPS to another courier firm
SW4	If my company changes from DPS to another company, some new technological problems would arise
SW5	My company would feel uncertain if we have to choose a new courier firm

Table 2.22: Items of "switching cost" used by Lam et al. [148].

Later, Bell et al. [16] considered that switching costs are a function of time and of the phase of development of the relationship between customer and company. The customers (and the companies) usually make specific investments in the relationship according to its maturity (e.g.: learning of procedures, preferences, own systems, the development of trust in a service provider) and these investments increase customers' perceptions of the costs of switching.

In their study, although it is based on the definition of perceived switching costs given by Jones et al. [133] (perceived economic and psychological costs associated with the change from one provider to another), they greatly simplify their work by proposing, based also on the sub-scales defined by Jones et al. [133], a final scale of three items adapted to the financial services context (see *Table 2.23*):

Perceived Switching Costs
1. If I changed firms, it would take a lot of effort to find a new one
2. If I changed firms, it would take a lot of time and effort on my part to explain to the new financial adviser what I like and what I want
3. If I were to switch firms, I would have to learn how things work at the new one

Table 2.23: Items of "perceived switching costs" used by Bell et al. [16].

Recent studies in the mobile service market have differentiated between positive and negative switching barriers [176, 257]. On the one hand, positive switching barriers, as relational benefits enjoyed by the customer after a continued relationship with the provider, will contribute to loyalty. On the other hand, negative switching barriers, as financial ones, create "spuriously loyal" customers who are not willing to churn just because of the switching costs. Costumers that suffer obligatory bounds with service providers due to unreasonable contractual obligations and sense a lack of ability to exit from the relationship with the company, use to hold a grudge against the provider and tend to switch [36].

In their study, Vázquez-Carrasco and Foxall [257] focussed on the effects of positive and negative switching barriers on loyalty. Those switching barriers created in lieu of satisfaction drive to loyalty, while those which are perceived as forced engagement lead to sabotage, lower acceptance of new products and negative word of mouth. So, they differentiate on the following factors:

- ***Relational benefits***: are the result of having cultivated long-term relationship with a service provider. They include *social benefits* (personal bonds between customer and provider, sense of belonging and empathy), *confidence benefits* (psychological benefits related to comfort and feeling of security) and *special treatment benefits* (combination of economic, such as discounts or better service, and customisation benefits, such as preferential treatment or extra attention). They have strong positive relationship with satisfaction and loyalty.

- ***Switching costs***: are the perception of the incremental costs required to terminate a relationship and secure an alternative. Customer's perception of switching costs leads to loyalty, but when it turns out to a "locked in" feeling, it leads to dissatisfaction.

- ***Availability and attractiveness of alternatives***: refers to customers' perceptions regarding to the extent to which viable competing alternatives are available in the marketplace. When viable alternatives

are lacking, the probability of terminating an existing relationship decreases. The more availability and attractiveness of alternatives, the lower customer satisfaction and loyalty.

For their part, Steyn et al. [237] studied the effect of loyalty cards on customers from a toy retailer. They measured the *perceived benefits* of customers as their valuation -using a 5 point scale- of 10 specific benefit related to the ownership of the loyalty card (see *Table 2.24*).

Perceived Benefits	
B1	Points with every purchase
B2	Your points give you reward coupons every 4 months
B3	Star offers allow exclusive savings for members
B4	Priority session for members at warehouse sale
B5	Fun bonus: buy $350 and get $70 toy coupon
B6	Special offers from Star Card partners
B7	Summer/ Christmas catalogue mailed directly
B8	Email newsletters with latest offers, deals, news
B9	45 days refund period for Star members
B10	Star Card Customer Hotline

Table 2.24: Perceived benefits defined by Steyn et al. [237].

Other Variables

The bibliography in the field of customer loyalty is extensive, and its revision shows that there are also authors who have considered alternative variables in the design of a descriptive model for the construction of loyalty bonds with costumers; variables that have notable effects on the process due to specific context in which the study was performed. We believe it is useful to review them briefly.

"Unappreciated" variables

We use the term "unappreciated" variables to refer to those values that are not connected to the customer's perceived satisfaction with a given service. Thus, for example, Chen and Hitt [44] provide, in the context of the online stock brokering industry, a descriptive model of disloyalty, studying characteristics of the firm/web page: the quality of the system and the information, user friendliness, level of customization, broker costs and variety in the product portfolio; demographic characteristics of the customer: age, sex, income, education, market size, race, household components, marital status and occupation; and customer usage variables: frequency of usage of the website, number of brokers used and change in usage patterns.

Length of relationship with the service provider

This variable was considered in the study by Jones et al. [132]. The length of time for which the customer had maintained a relationship with his current service provider was included in order to control the fact that satisfaction and its behavioural consequences can differ when this is based only on sparse usage rather than when it is built up over years of repeated use. This variable is measured through the following item:

- "Approximately, how long have you used this bank?"

Indifference and Inertia

The literature related to the measurement of *indifference* is scarce, although in occasions it has been used in marketing literature related to a "neither positive nor negative" customer attitude towards advertising. Some research refers to the perceptions of spending and homogeneity in the service provided by a given industry as factors that determine the level of customer *indifference* toward change [149].

Meanwhile, Huang and Yu [121] made use of the concept "I am not prepared to make the effort needed to change". In this way, they defined *inertia* as a type of unconscious human emotion, conceptualizing it

as a unidimensional variable consistent with a pattern of passive service without real loyalty. Ranaweera and Neely [216] included these two new dimensions in their study. They measured indifference using a two-item scale suggested by Lambert [149], which measured perceptions of the offer homogeneity between different companies and the monthly spending level. For their part, they evaluated inertia using a sentence consistent with the work of Huang and Yu [121]: "I can't be bothered changing my phone company".

Expertise level or grade of specialization

Bell et al. [16] included the variable *expertise level* in the model they proposed to explain customer *loyalty* in the finance industry. In their study, specialization of the investment is considered as the extension of customers' prior knowledge of the product, which they use to evaluate the profitability that will result. The concept measures customer expertise in relation to investments in the market, more than their knowledge of one particular provider of financial services. It was estimated using a four-item scale developed by Sharma and Patterson [228], making slight changes to adapt it to the context of the study (see *Table 2.25*):

Investment Expertise
1   I possess good knowledge of financial planning services and products
2   I am quite experienced in this area

Table 2.25: Items of "investment expertise" used by Bell et al. [16].

Trust

Trust has been studied extensively in the existing literature. In terms of services, trust can be defined as the belief by a customer that the service provider will provide the service that meets customer needs [4]. Morgan and Hunt [183] defined trust as the confidence that one part has in the honesty and reliability of his partner. Rauyruen and Miller [218] defined two levels of trust: at the first level, the customer trusts one particular sales representative while at the second level, the customer trusts the institution.

When a customer trusts an organization, she or he has the confidence in service and product quality that leads to a strong loyalty [84]. The positive effect of trust on loyalty has been proved in contexts such as e-commerce [160] and telecommunications [65, 169].

Other studies consider trust as a part of relationship quality [6, 64, 199, 218]. For their part, Yin et al. [291] measured trust, related to restaurants and hair salons, using the four items shown in *Table 2.26*, adapting the previous work of Morgan and Hunt [183]:

Trust
1.   You are confident about the food (product) quality provided at this restaurant (hair salon)
2.   This restaurant provides reliable services / This hair salon provides reliable and professional services
3.   This restaurant (hair salon) has high integrity
4.   Overall, you can confidently rely on this restaurant (hair salon) for service

Table 2.26: Trust definition by Yin et al. [291].

Playfulness

Studies related to online services or telecommunications lately introduced the concept of *playfulness* as one that has a significant effect on customer satisfaction. Even *fun* or *entertainment* features might not be a priority for users: it entails engagement and enjoyment. In electronic commerce, playfulness has been demonstrated to lead to satisfaction [164, 169], exploratory behavior [192, 232] and future intentions to repurchase [144].

Liu et al. [169] considers playfulness as an antecedent of customer satisfaction in his study of mobile service providers. They focus on the individual experiences that result from the use of mobile devices,

which often implies high engagement and concentration levels, leading to enjoyment. Playfulness is measured by 3 items in a 5-point scale, shown in *Table 2.27*:

Playfulness
1.   Using mobile services gives enjoyment to me
2.   Using mobile services is fun for me
3.   Using mobile services keeps me happy

Table 2.27: Playfulness definition by Liu et al. [169].

Need for Variety

Vázquez-Carrasco and Foxall [257] mention in their study the concept of *need for variety* as the fact that switching provider becomes a high priority [136] for some customers. They state that customers need a certain level of stimulation with which they feel comfortable; if it falls below the optimum level, the customer will seek additional variety from the environment in order to increase the stimulation, and can lead to brand switching [279]. Vázquez-Carrasco and Foxall propose that *need for variety* implies a higher attractiveness of existing alternatives and reduces the switching costs perceived by the customer. They measured it using the previous work of Steenkamp and Baumgartner [236].

Expanding this concept, a recent study by Malhotra and Malhotra [176] states that brand innovativeness of products and processes leads to customer loyalty, through fulfilling customers' needs for stimulation and variety and improving brand's image perception.

2.2.2   Models of loyalty

As mentioned in its introduction, one of the objectives of this section is presenting the main descriptive models of loyalty proposed in recent literature from the backgrounds that are described in the preceding section. Organized chronologically, they are:

- The research carried out by Cronin et al. [54] focused on the context of services marketing (spectator sports, participation sports, entertainment, health, long-distance phone calling and fast food) and used four explanatory models of loyalty (see *Figure 2.5*) based on *satisfaction, service quality*, *sacrifice* and the *service value*. While the first three are based on already existing literature [7, 79, 202, 243] the fourth was originally proposed in this study.

Value Model   **Satisfaction Model**

Indirect Model   **Proposed Model**
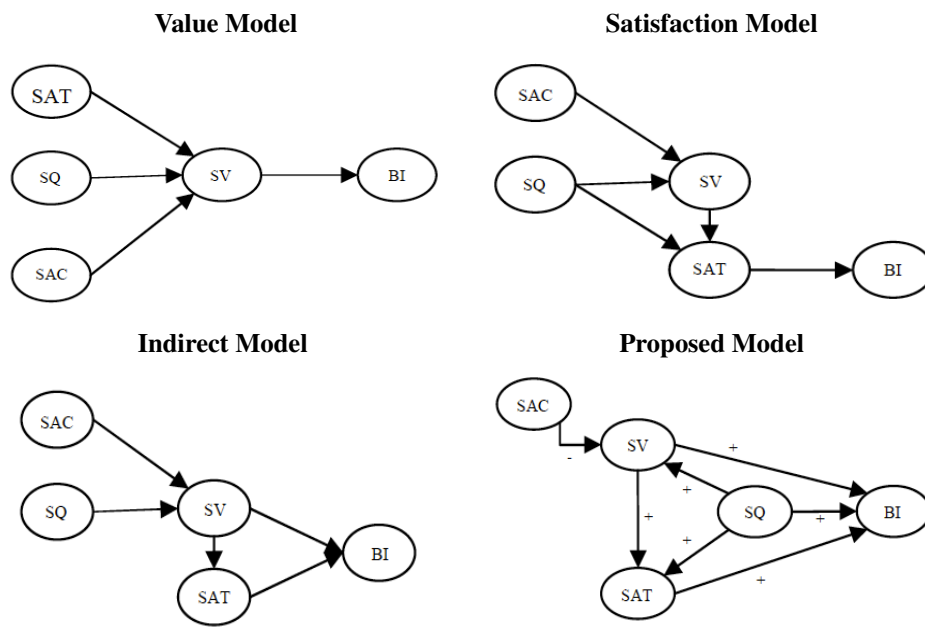
SAT: satisfaction   **SAC**: sacrifice   **SV**: service value   **SQ**: service quality
BI: *behavioural intentions (repeat purchase and recommendation)*

Figure 2.5: Models analyzed by Cronin et al. [54]. On the bottom row, right figure (+) represents a positive effect on described interactions and (-) represents a negative effect.

The statistical analysis carried out by Cronin et al. confirmed empirically that the available data were better adapted to the model proposed, making it possible to explain a greater part of the variation in Behavioural Intentions (BI). The only hypothesis that was not supported by the data was the one referring to *sacrifice* as a factor influencing *service value*.

• For Jones et al. [132], customers' repeat purchase intentions go beyond satisfaction. For this reason, they proposed a model (see *Figure 2.6*) in which the switching barriers appear as factors with a positive influence on customer decisions to remain loyal to a service provider.

Figure 2.6: Model proposed by Jones et al. [132]. (+) represents a positive effect on described interactions and (-) represents a negative effect.

The results of the experiment showed that the influence of satisfaction on repeat purchase intentions decreases under the condition of strong switching barriers. On the other hand, they observed that, although the switching barriers do not influence the repeat purchase intention when satisfaction is high, they did have a positive influence on repeat purchase intention when satisfaction was low.

- The model proposed by Chen and Hitt [44], in the context of the online stock broking industry, explains "disloyalty" -switching or termination- through "unappreciated" variables: customers characteristics (related to demographics and usage patterns) and website characteristics (see *Figure 2.7*).
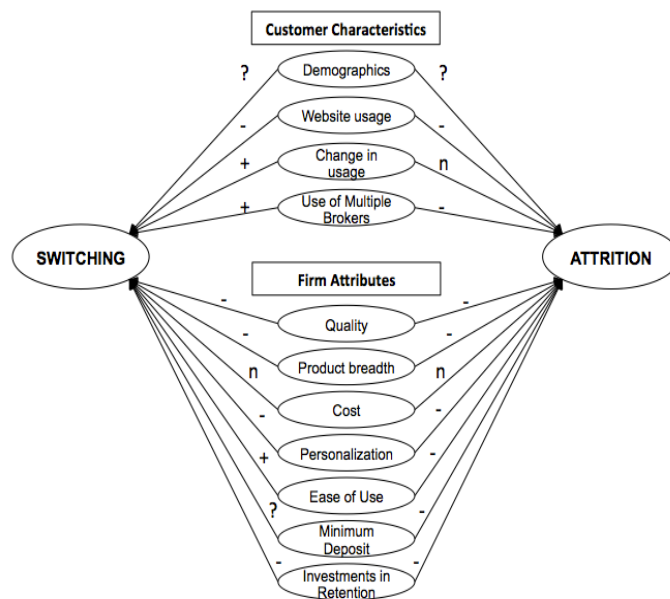
Figure 2.7: Model proposed by Chen and Hitt [44]. (+) represents a positive effect on described interactions, (-) represents a negative effect, (?) represents an unknown relation and (n) represents no relation.

All the hypotheses found empirical support with the exception of the negative relationships between switching and the level of personalization of the website; termination and the level of website customization; termination and the quality of the website; termination and the quantity of offers and, finally, termination and the user-friendliness of the website.

- The descriptive model of loyalty proposed by Ranaweera and Neely [216], in the framework of the UK telecommunications market, analyzed the indirect effects that indifference and price have on the relationship between service quality and customer retention (see *Figure 2.8*). In this study, the authors proposed that the positive effect of the service quality is less intense if indifference is greater and more intense if the perceived price is higher.

  The statistical analysis carried out supported all the hypotheses except for the supposed positive relationship between inertia and customer retention.
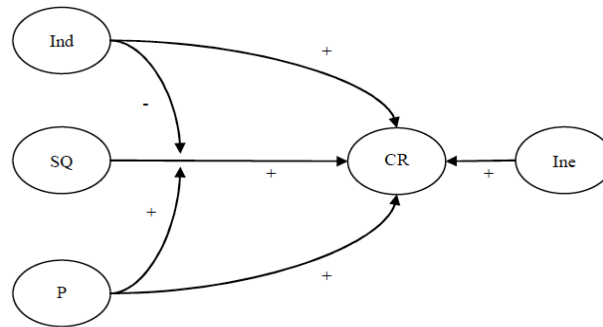
Ind: indifference   **SQ**: service quality   **P**: price perception
Ine: inertia   **CR**: customer retention (repeat purchase intentions)

Figure 2.8: Model proposed by Ranaweera and Neely [216]. (+) represents a positive effect on described interactions and (-) represents a negative effect.

- Patterson and Smith [204] proposed the descriptive model of loyalty illustrated in *Figure 2.9* to explain the degree to which the switching barriers explain the variation in repeat purchase intentions from medical services, travel agencies and hairdressers. After a hierarchical regression analysis -first using only the switching barriers and subsequently adding customer satisfaction- they concluded that switching barriers provide the explanation for most of the variation in repeat purchase intention, with two barriers standing out in these three industries: *loss of treatment* and *loss of good relationship*. Thus, they confirmed that by adding customer satisfaction the impact was greater in hairdressers than in the other two industries analyzed.

  Another of the intended objectives of the study was to analyse the interactions between satisfaction and different switching barriers, and for this reason they added a term of interaction for each barrier (barrier·satisfaction). No significant interactions were found.
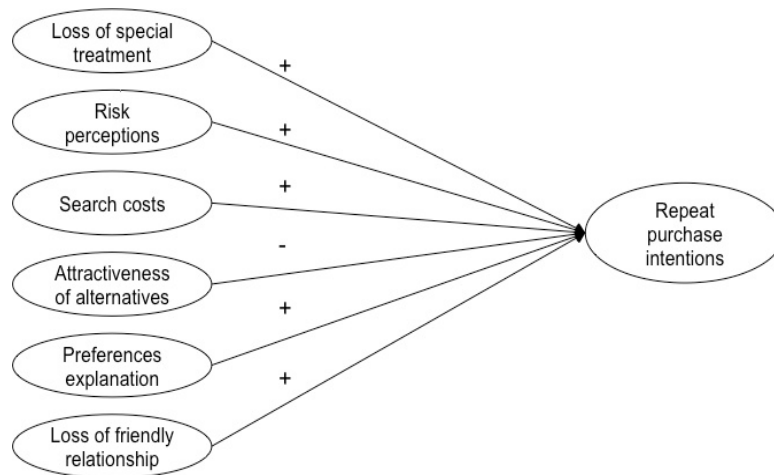
Figure 2.9: Model proposed by Patterson and Smith [204]. (+) represents a positive effect on described interactions and (-) represents a negative effect.

- Gounaris and Stathakopoulos [96] proposed a model (see *Figure 2.10*) to analyse the relationships between the characteristics associated with the consumer: desire to avoid risk and search for variety; brand reputation and availability of substitute products, the social environment and the four types of loyalty defined: "premium" loyalty, inertia loyalty, "covetous" loyalty and non-loyalty. The relationships with the four different types of consumer behaviour identified: "word of mouth" communication, purchase of alternative brands, visits to different shops, and non-purchase, were also analyzed. The context in which the empirical study was carried out was that of whisky consumers.
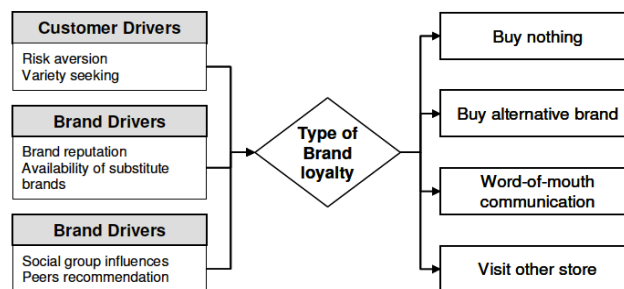
Figure 2.10: Model proposed by Gounaris and Stathakopoulos [96].

The statistical analyses carried out on this data set led to the following conclusions:

- Desire to avoid risk is significantly related to "premium" loyalty and with "non loyalty", but not with the other two.

- Reputation has a positive relationship with "premium" and "covetous" loyalty, a negative relationship with "non loyalty" and no significant relationship with "inertia" loyalty.

- The availability of substitutes has a strong positive relationship with "inertia" loyalty, a positive relationship with "non loyalty", a negative relationship with "covetous" loyalty and no significant relationship with "premium" loyalty.

- Social influences have a positive relationship with "premium" and "covetous" loyalty have positive relationships with "word of mouth" communication and negative ones with "inertia" loyalty.

- Kim et al. [139] proposed a descriptive model of loyalty in the context of Korean mobile phone services, using satisfaction -with *service quality* as the determining factor- and the switching barriers -with switching costs, appeal of the alternatives and interpersonal relationships as the key factors- (see *Figure 2.11*). They did not find any empirical support for hypotheses envisaging a positive relationship between price structures -more reasonable- and satisfaction; between the mobile device and satisfaction; or between convenience of procedures and satisfaction. Nor was there any statistical backing for the predictions of positive relationships between costs of loss and switching barriers; or between the appeal of alternatives and switching barriers.
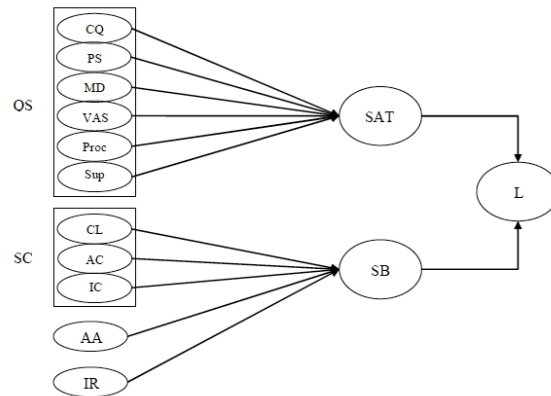
QS: service quality   **CQ**: call quality   **PS**: price structures
MD: mobile device   **VAS**: value added services   **Proc**: convenience of procedures
Sup: customer support   **SC**: switching costs   **CL**: costs of loss
AC: adaptation costs   **IC**: installation costs   **AA**: attractiveness of alternatives
IR: interpersonal relationships   **SAT**: satisfaction   **SB**: switching barriers
L: loyalty

Figure 2.11: Model proposed by Kim et al. [139].

- In a study based on a courier company, Lam et al. [148] proposed a descriptive model of *loyalty* -considering repeat purchase and recommendation separately- based on satisfaction, switching costs and the value perceived by the customer (see *Figure 2.12*). In their study, they gave equal consideration to the moderating effects of switching costs and the quadratic effects resulting from satisfaction.
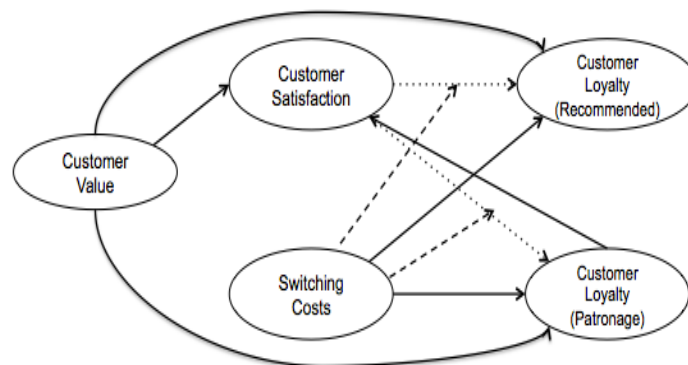
Figure 2.12: Model proposed by Lam et al. [148]. Solid lines represent a lineal effect on described interactions, dashed lines represent a quadratic effect and dotted lines represent a moderating effect.

Initially, they carried out a confirmatory factor analysis including four factors (satisfaction,

switching costs and the two types of loyalty) which did not reach the recommended minimum significance levels. On the bases of these results, they concluded that some of the measurements obtained through this survey could be problematic. A LISREL analysis revealed that the last of the five questions relating to switching costs (*"my company would feel uncertain if we have to choose a new courier firm"*) was the cause of the poor significance, and the results improved notably when this item was eliminated. Meanwhile, there was no support for the quadratic and moderator effects, so that the hypothesis of the quadratic effect of satisfaction on the two types of loyalty could not be sustained (although the lineal effect was), and nor could the effect of loyalty on satisfaction or the moderator effect of switching costs.

- Fullerton [81] tried to predict customers' propensity towards the following behaviours: *intention to recommend*, *intention to change* and *willingness to pay more* depending on their *commitment to continuity*, their *emotional commitment*, the *service quality* -interaction, result and environment- and the *scarcity of alternatives*.

  For this, they focused on the framework of service in retail sales (men's clothing and food products) in Canada. An initial model was proposed (see *Figure 2.13*) with the objective of finding out if service quality could affect customer behaviour, not only through commitment but also directly. It attributed only a mediator role to the general quality of service, expanding it subsequently to a new model that also took the direct effect into account (see *Figure 2.14*).
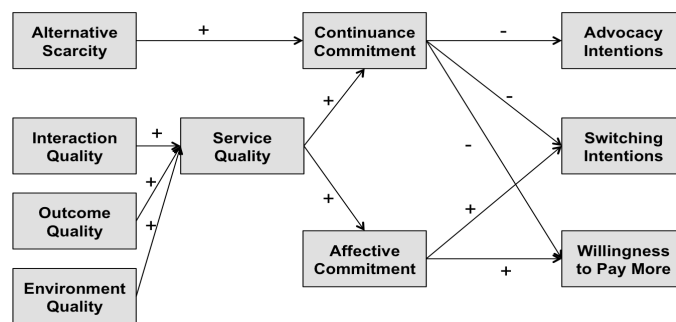
Figure 2.13: Integrated model of retail service relationships proposed by Fullerton [81]. (+) represents a positive effect on described interactions and (-) represents a negative effect.
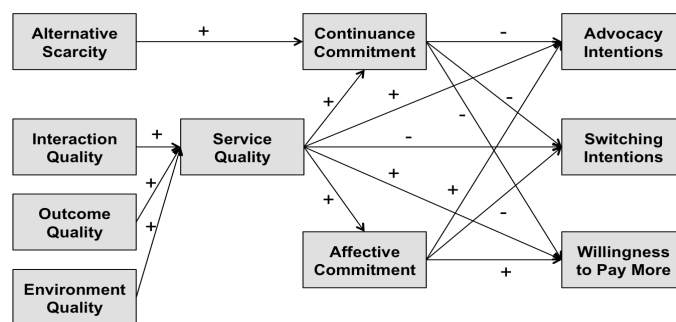
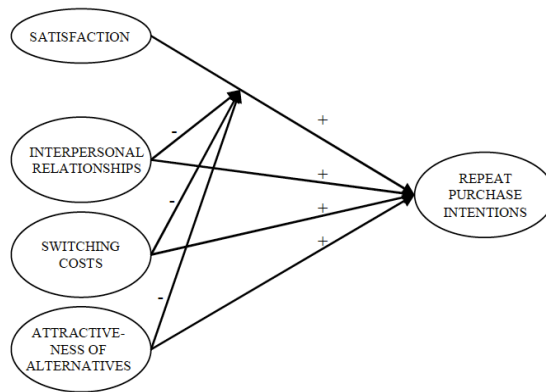*Figure 2.14:* Second variant of integrated model proposed by Fullerton [81]. (+) represents a positive effect on described interactions and (-) represents a negative effect.

All the hypotheses defined for both models were supported, with the exception of the positive relationship between general service quality and willingness to pay more (in the expanded model) which did not find acceptance in the case of food product stores.

- Vázquez-Carrasco and Foxall [257] studied the effects of positive and negative switching barriers on hairdresser costumers' loyalty. They considered that the *naturally created* switching barriers

improve the relationship and contribute to loyalty, but the ones which are perceived as coercive (such as high financial costs) tend to reduce satisfaction and lead to failure in long term relationships. They used three categories of perceived switching barriers used previously by Jones et al. [132]: relational benefits, switching costs and attractiveness of alternatives; adding need for variety as an antecedent, and measured their positive and negative effect on customer satisfaction and loyalty. The proposed model is shown in *Figure 2.15*.
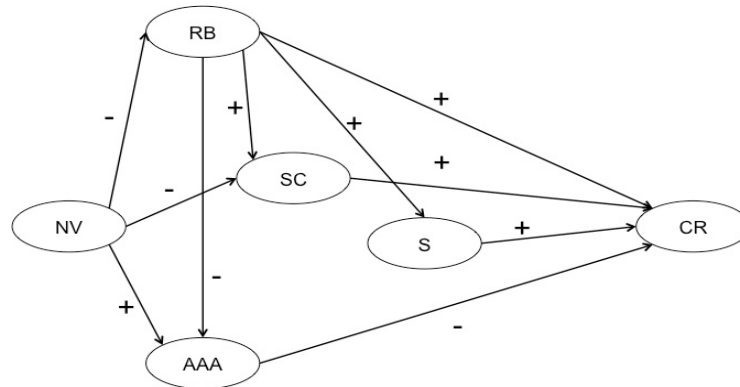
NV: Need for variety   **RB:** Relational Benefits   **SC:** Switching Costs   **S:** Satisfaction
**CR:** Customer Retention   **AAA:** Availability and Attractiveness of Alternatives

Figure 2.15: Model proposed by Vázquez-Carrasco and Foxall [257]. (+) represents a positive effect on described interactions and (-) represents a negative effect.

The obtained results showed a strong positive effect of perceived relational benefits on switching costs and customer retention. Thus, switching costs have a direct and positive effect on customer retention, but also a negative effect on satisfaction. Attractiveness of alternatives is directly and negatively related to customer satisfaction and retention.

Measuring the internal effects of switching barriers, relational benefits contribute positively to higher switching costs, and to a lower attractiveness of alternatives. Additionally, need for variety has a positive and direct effect on attractiveness of alternatives, and affects negatively to relational benefits and switching costs: a customer that seeks constantly for variety will switch more easily to other service providers. The results didn't prove any significant negative effect of switching costs nor attractiveness of alternatives on satisfaction.

- Caceres and Paparoidamis [40] studied business-to-business loyalty in the advertising area measuring the effects of relationship satisfaction, trust and commitment and its antecedents. The variables were measured using an own-built 26-item scale after interviewing experts in the field. The defined model tested the effect of both perceptions of technical quality (measured as the quality of an advertising campaign) and perceptions of functional quality (measured as the aggregate of the quality of commercial service, communication with the supplier, delivery of a service and administrative service) on clients' satisfaction. They also included the effect of customer satisfaction, trust and commitment on loyalty (see *Figure 2.16*).
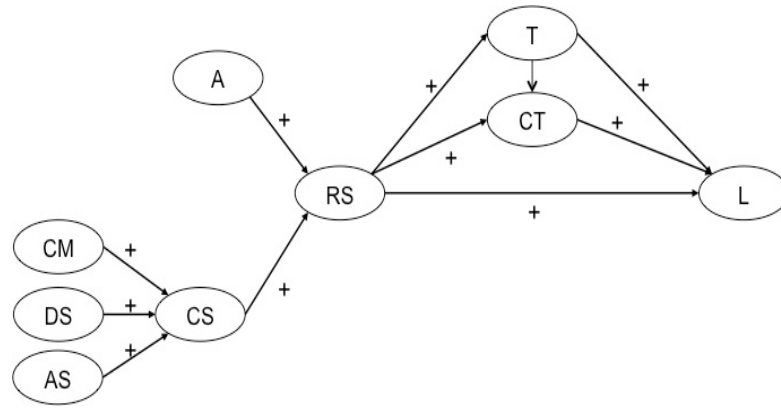
CM: Communication	**A**: Advertising	**T**: Trust
DS: Delivery Service	**CS**: Commercial Service	**CT**: Commitment
AS: Administrative Service	**RS**: Relationship Satisfaction	**L**: Loyalty

Figure 2.16: Model proposed by Caceres and Paparoidamis [40]. (+) represents a positive effect on described interactions.

The results show a significant positive effect of both technical quality (advertising) and functional quality (commercial service) on satisfaction, additionally pointing that the effect of technical quality is stronger. Moreover, the results show an indirect effect of communication, delivery of service and administrative service on satisfaction, mediated through commercial service. Finally, the effect of trust and commitment on loyalty seems to be greater than the effect of customer satisfaction.

- Yin et al. [291] modeled customer loyalty on fast food restaurants and hair salons, focusing on two main areas: customer-staff relationship and customer-firm relationship. Aiming to test the differences between transactional services (where customer-staff interaction is low) and relational services (where customer-staff relationship is meant to be a more influential factor), they compared the significance of the antecedents of loyalty in each case. Thus, the study starts from some generally-accepted assumptions which determine the core of the model: links between quality, satisfaction, social rapport (perception of an enjoyable interaction with a staff member), firm trust and loyalty. Then, the effect of staff trust and loyalty towards firm's is added to the model. Furthermore, they measure the moderating effect of customer-firm affection, which influences the effect of service quality and customer satisfaction on firm trust and firm loyalty intentions.

ST: Staff Trust	**SL:** Staff Loyalty intentions	**SQ:** Service Quality
FT: Firm Trust	**FL:** Firm Loyalty intentions	**SR:** Social Rapport
SAT: Customer Satisfaction	**SPI:** Share of Purchase Intention	
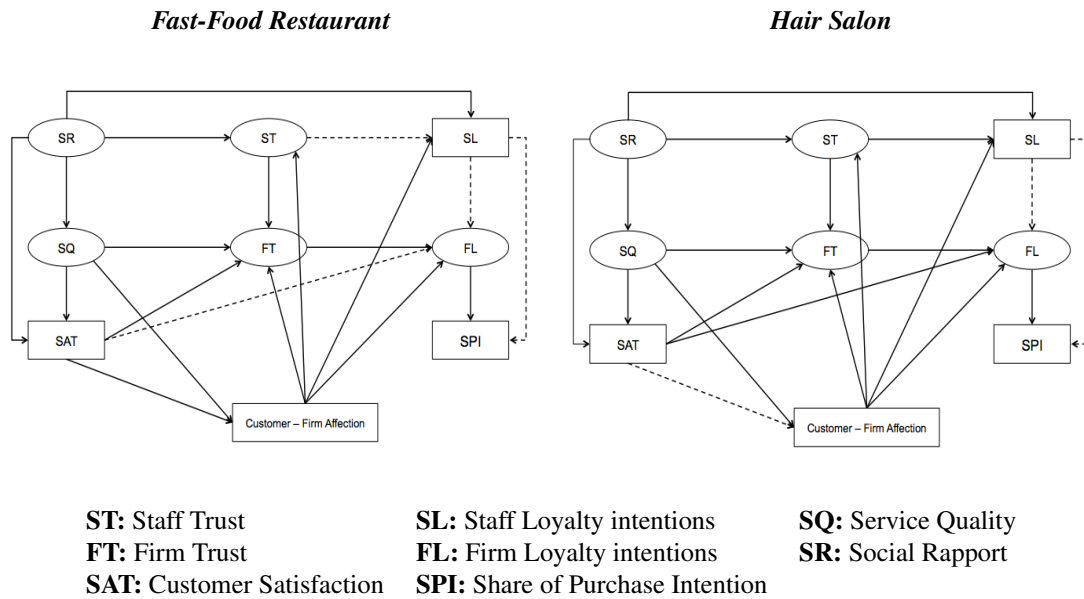
Figure 2.17: Model proposed by Yin et al. [291]. Solid arrows represent significant relationships, dashed arrows represent non-significant relationships previously formulated as hypotheses.

Figure 2.17 reveals the significance of all the assumptions made in the core model: the positive effects of service quality, satisfaction, social rapport and firm trust on firm loyalty. Furthermore, they proved that customers of relational services experience commitment-dominant customer-firm affection, whereas those of transactional services develop passion-dominant customer-firm affection. The difference between transactional and relational services is significant when comparing the effect of staff loyalty on firm loyalty, but it's not in the case of staff trust on firm trust.

- The research of Deng et al. [65] on mobile instant messages customer loyalty introduced some moderating effects (age, gender and usage) on the links between customer satisfaction, trust, perceived switching costs and loyalty. Also, they suppose that the aggregate of functional value, emotional value, social value and monetary value have a positive effect on customer satisfaction. The measurements are based on a 29-item survey assessed by experts in the mobile services field and the moderating effects are calculated after a segmentation of the sample by gender (47.3% male, 52.7% female), age (in two groups: below 24 years old, 47.3%; and older, 52.7%) and usage (have used instant messages for less/more than one year, with approximately the half of the sample in each group). The proposed model is shown in *Figure 2.18* (Again, solid arrows denote significant relationships, while dotted arrows denote non-significant relationships).
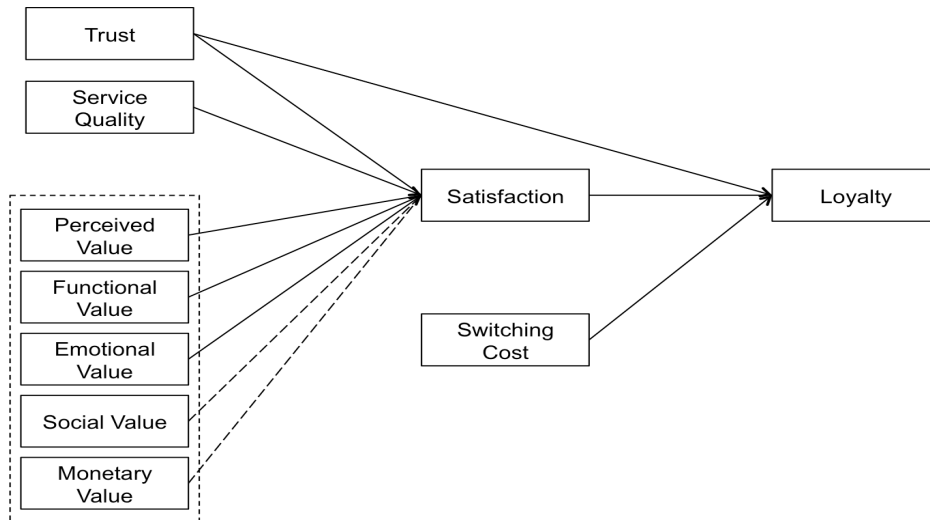
Figure 2.18: Model proposed by Deng et al. [65]. Solid arrows represent significant relationships, dashed arrows represent non-significant relationships previously formulated as hypotheses.

The results proved the significance of all the proposed antecedents of satisfaction, except social value and monetary value. Furthermore, the study of moderating effects (not shown on *Figure 2.18*) reveals that emotional value and trust have a stronger effect on females than males. Likewise, emotional value and trust have a stronger effect on older customers than young ones. Finally, a longer usage of the service links to a stronger effect of customer satisfaction on loyalty.

• Steyn et al. [237] studied the effect of perceived benefits on the feelings of customers participants of a retailer's loyalty program in Asia, surveying customers from five different countries: Malaysia, Singapore, Hong Kong, Taiwan and Thailand. Loyalty is measured thorough three items: use frequency, carry (of the loyalty card) frequency, and recommendation propensity, all of them measured on a 4-point scale. The perceived benefits are measured as financial benefits and information benefits. The whole theoretical model is shown in *Figure 2.19*.
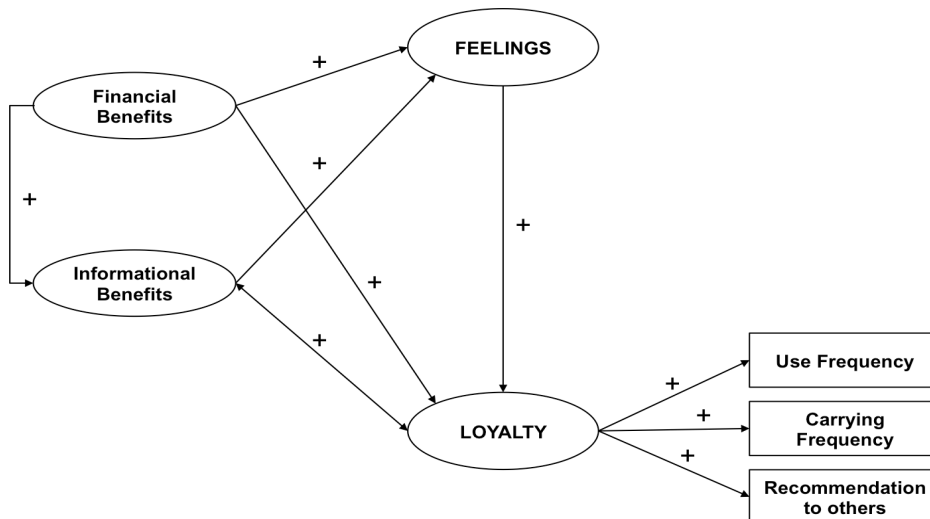
Figure 2.19: Model proposed by Steyn et al. [237]. (+) represents a positive effect on described interactions.

Steyn et al. concluded that perceived benefits have a weak direct effect on loyalty behaviors, but have a much stronger effect on feelings, which in turn have a strong effect on loyalty behaviors.

51

Another common fact in all countries is finding recommendation as the strongest item to describe loyalty. Even though, they failed to get further general conclusions from the aggregated data collected from all countries but obtained valuable results when studying each country separately, which can be attributed to cultural differences.

- Recent research by Liu et al. [169] measures the effects of relationship quality -formed by satisfaction and trust- and switching barriers on customers' loyalty, based on a telecommunication customers survey in Taiwan. The study proposes a positive effect of playfulness and service quality on satisfaction; a positive effect of service quality and intimacy on trust; and finally, a positive effect of satisfaction, trust and switching barriers on loyalty (see *Figure 2.20*).
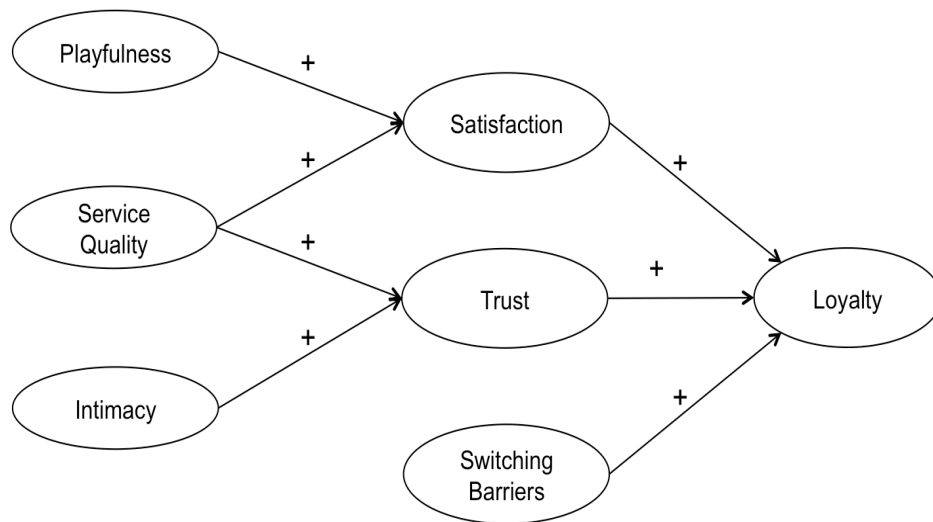
Figure 2.20: Model proposed by Liu et al. [169]. (+) represents a positive effect on described interactions.

The results show a significant effect of all the factors proposed, confirming all the hypotheses. The structural model explains the 48% of customer loyalty variance and shows a stronger effect of satisfaction on loyalty, rather than trust or switching barriers. Service quality shows up to have the greatest effect on customer satisfaction.

Chapter 3

Predictive models in churn management

As explained in *Chapter 2*, the task of prolonging customers' useful life requires a systematic approach to its management. For this task, the design of a suitable Customer Continuity Management model has been suggested.

Anticipating a customer's intention to abandon their current provider company should be considered a key element of any therapeutic strategy in churn management. Early diagnosis of customer abandonment should, at the very least, reduce the aggressiveness of the required therapy, increasing as a result the possibilities of customer recuperation.

In this context, Data Mining techniques, applied to market surveyed information, should play a key role in helping to understand how customer loyalty construction mechanisms work, and how the customers' intention to abandon could be minimized, facilitating the launch of retention-focussed actions. Continuing with the medical analogy used in *Chapter 2*, there is no use in predicting an illness unless it is done in time to administer the appropriate treatment.

However, not all cases of churn -abandonment of the commercial relationship between company and customer- are equally important, nor are they all predictable. According to the reasons behind their abandonment, customers can be classified in different typologies [87]:

- *Involuntary cancellation*: Referred to customers from which the actual company withdraws their service (fraud, arrears...). Generally, companies do not even consider these cancellations as abandonment for their records.

- *Voluntary cancellation*: It corresponds to customers who consciously decide to change provider. Two variants can be considered:

    - *Circumstantial*: Due to changes in the customer's circumstances which do not allow them to continue (change of address, inclusion in the company's social benefit plans, change of marital status, children, ...). This cancellation is intrinsically unpredictable.
    - *Deliberate*: Occurs when the customer voluntarily decides to abandon their current provider for a competitor.

In the present thesis we are mostly interested in this last scenario: voluntary and deliberate churn. Its management requires the design and development of predictive models of churn. Such models stem from different fields of research. Here, we are specially interested in PR approaches to their design.

The current chapter is organized as follows: In *Section 3.1*, the different stages of a standard process of design and development of a predictive model of abandonment will be described. Each of the elements of these stages will be exemplified by recent associated literature. This literature will then be summarily reviewed in *Section 3.2* in the form of tables according to two main grouping criteria: the type of predictive model used in the study and its particular area of application.

3.1 Building predictive models of abandonment

The process of design and development of abandonment predictive models can be divided into four stages [59], as seen in *Figure 3.1*. The last three stages of this process form a cycle that ends when adequate prediction results are achieved. We will now take a closer look at each stage in turn.
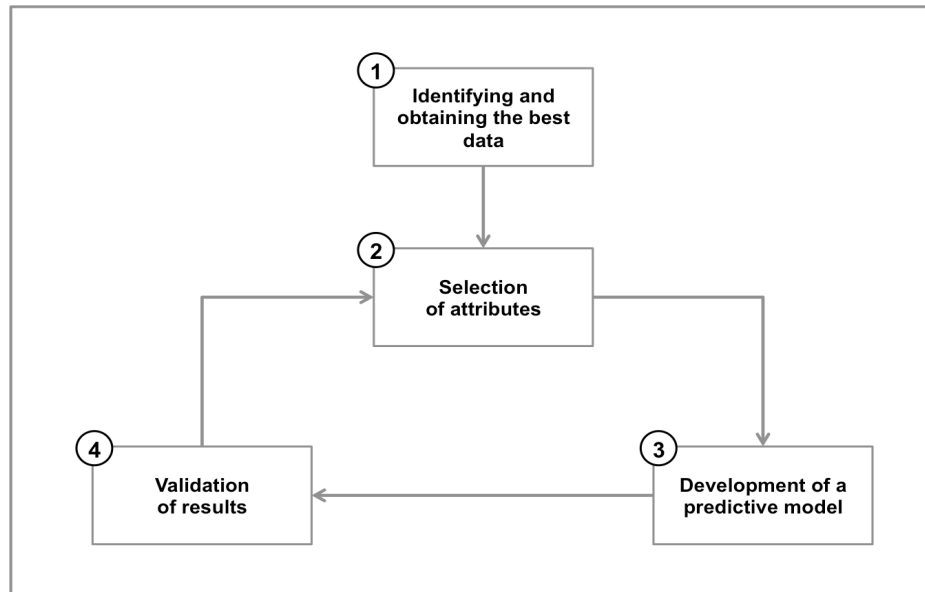
Figure 3.1: Stages of the predictive model building process [59].

3.1.1 Stage 1: Identifying and obtaining the best data

This might arguably be the most relevant stage in the construction of a predictive abandonment model. Experience shows that the quality and suitability of the available data determines the accuracy and predictive power of the resulting model. Different data combinations may be better or worse indicators for different problems and for different sectors. Ultimately, it is a question of identifying the data which best fit the type of analysis being carried out; only in this way will we be able to extract, in subsequent stages, knowledge that is useful and actionable in business terms.

The recent literature presents a selection of different data requirements for the analysis of abandonment. The most relevant are detailed below:

- A large group of studies base their models of abandonment prediction on customer use/consumption variables:
  - Madden et al. [175], in their customer retention model for the Australian ISP (Internet Service Provider) industry, classified and used four categories of variables: economic, use, ISP choice and demographics.
  - Ng and Liu [188] suggested the use of customer consumption for identifying churn in the telecommunications market.
  - In their study, Verhoef and Donkers [275] concluded that the purchase of products and services can be better predicted using historic purchasing data.
  - This last view was backed by Hsieh [117], who proposed that the analysis of transaction data, through historic account and customer data, could provide us with clues to identify the best incentives for a bank to offer its customers and to improve the marketing strategy.

– Data on customer usage have also been used to identify the behaviour of website-using customers [130] and to predict repeat purchasing by mail [254].

- More recently, customer usage/consumption data have been complemented with other variables as key elements in identifying abandonment:

  – In their study of customer deflection in the wireless telecommunications market, Slater and Narver [233] grouped customer data into four types: demographics, usage level, quality of service and marketing features. This method was supported by Zhao et al. [296] in a more recent study in the same sector.

  – Following in the field of telecommunications, Neslin and Gupta [187] classified the selected variables into three main categories: customer behavior (minutes of use, revenue, handset equipment, trends in usage), company interaction data (calls to customer service) and customer household demographics (age, income, geographic location, home ownership).

  – Hung et al. [122], considered that the most significant variables for churn prediction in the mobile telephone industry are: demographic data (age, penetration rate, and gender), payment and account data (monthly quota, billing amount, arrears account), call details (call duration, call type), and customer service data (number of PIN number changes, number of blocks and suspensions).

  – In their research about abandonment for the subscribers of a newspaper publishing company, Coussement and Van den Poel [49] grouped customer data in four groups: subscription data (time since last renewal, monetary value, product), socio-demographics (age, gender), client/company interactions (number of complaints, time since the last complaint, responses to marketing actions) and renewal-related variables (days between subscription renovation and expiry date).

  – More recently, both Nie et al. [190] and Wang et al. [280] used data from credit card holder's in a Chinese bank to predict their abandonment, with a combination of usage variables (daily balance, abnormal usage, limit usage, revoking pays, transactions, . . .) and customer personal information.

- A number of authors [45, 115, 117, 134, 170, 171, 208, 254, 276] coincide in suggesting the use of three groups of variables, globally known as RFM (Recency, Frecuency and Monetary):

  – Length of time since last purchase,

  – Frequency of use,

  – Economic expense effected over a certain time period,

  as a source for predicting the abandonment probability of a particular customer. Bose and Chen [29] stated that RFM variables are amongst the strongest performing variables in explaining future customer behavior.

This stage in the building of predictive models of abandonment would fit, from a Data Mining process point of view, the phases of problem and data understanding and the subsequent one of data pre-processing. Bearing this in mind, and from a practical point of view, it is important to note that the predictive model should ideally be constructed on the basis of the available data gathered routinely by a company from its whole customer base, which can be an extremely costly process. Consequently, those data bearing most of the predictive power may not always be available. We are faced, therefore, with the trade-off problem of identifying the best data from what is available.

The process of understanding and interpreting the data often presents difficulties. Even though the data in each field of a database may seem self-explanatory and unambiguous, interpretation can become difficult because of the use of specific and *ad hoc* company lingo, different numerical formats, or simply because their meaning is different from the apparently obvious. Given the usual lack of standards to facilitate this process at the company level, its success is largely based on good communication between database managers and the data analysts. In fact, these Data Mining stages have not been duly documented in the majority of investigations carried out in recent years [104].

3.1.2 Stage 2: Selection of attributes

This stage consists of the selection of the most appropriate attributes for prediction from those available to us. In a supervised PR setting, that means those that minimize the classification or prediction error. In an unsupervised one, that means those which best reflect the grouping or cluster structure of the data.

This process is paramount as it helps to reduce the dimensionality of the data so that only the important attributes are included for analysis, whereas the redundant, noisy and/or irrelevant ones are excluded [288]. Attribute or feature selection in supervised settings is a problem that has been thoroughly studied throughout the years [101, 172, 239], and providing a survey of selection methods is well beyond the scope of this thesis.

The unsupervised selection of attributes or features, on the other hand, is a theoretical problem to which consistent attention has only been paid in recent years. In an unsupervised learning setting, we are not provided with targets or class membership labels, so that the problem of which features should be retained is a very different one. Some of the features may be redundant; some may be irrelevant and misguide the results of the clustering procedure.

Reducing the number of features would also circumvent the problem that some unsupervised learning algorithms might have with data of high dimensionality. As posed by Dy and Brodley [71] "The goal of feature selection for unsupervised learning is to find the smallest feature subset that best uncovers *interesting natural* groupings (clusters) from data according to the chosen criterion".

Following the categorization that is common for supervised methods, we can talk of wrapper approaches, in which we cluster the data in each candidate feature subspace, and then select the most "interesting" subspace with the minimum number of features. Rather than wrap the search for the best feature subset around a supervised induction algorithm, in this case we wrap the search around a clustering algorithm.

There are different selection criteria for different problems in unsupervised learning. Still, a number of variable ranking criteria are useful across applications, including saliency, entropy, smoothness, density and reliability [101]. Amongst these, saliency and density, in particular, are to be used in this thesis.

3.1.3 Stage 3: Development of a predictive model

Once the data available for analysis have been selected, the next stage entails the choice of the most suitable methods and techniques for building the predictive model. In a simple manner, a predictive model can be defined as one that extracts patterns from the available data in order to make inferences for previously unseen data or future situations [224].

In the area of abandonment prediction, the most commonly used modelling techniques, as reflected in the literature, include decision trees, regression analysis [163] and artificial neural networks (ANN) [50], while in the recent years new methods such as Support Vector Machines (SVM) [45, 49] have proved their adequacy. The following subsections provide an overview of both more traditional and Computational Intelligence (CI) techniques used for predictive churn modelling.

Standard methods

- **Decision Trees (DT)**: The most popular type of predictive model is the DT. In its different forms, it has become an important knowledge extraction method, used for the classification of future events [197]. There are two general distinct phases in their design: building and pruning:

    - **Tree building**: Consists of recursively partitioning the training sets according to the values of the attributes and a given measure of similarity expressed as a type of error.

    - **Tree pruning**: Involves selecting and removing the branches that contain the largest estimated error rate. Tree pruning is known to enhance the predictive accuracy of the decision tree, while reducing their complexity, in an example of bias *versus* variance trade-off [8].

A popular choice of DT, the **C5.0 classification tree** -a variant of the well-known C4.5-, assembles classification trees by recursively splitting the instance space into smaller subgroups, according to an information entropy criterion, until only instances from the same class remain known as a pure node, or a sub-group containing occurrences from different classes known as impure nodes. The tree is allowed to grow to its full potential before it is pruned back in order to increase its power of generalisation on unseen data.

Another frequently used DT is the **classification and regression tree (CART)**, constructed by recursively splitting the instance space into smaller sub-groups until a specified criterion has been met. The decrease in impurity of the parent node against the child nodes defines the goodness of the split. The tree is only allowed to grow until the decrease in impurity falls below a user-defined threshold. At this time the node becomes a terminal, or leaf node [27].

The literature contains some examples of DT being used in the construction of models for abandonment prediction:

- Datta et al. [59] carried out research in the area of churn prediction and developed a model that they called Churn Analysis Modelling and Prediction (CHAMP). CHAMP also uses DTs to predict customer churn in the telecommunications industry.

- Ng and Liu [188] chose C4.5 to automatically generate classification rules for the purpose of identifying potential defectors.

- In the field of wireless telecommunications market, Hwang et al. [123] compared the performance of DT, ANN and logistic regression. They stated that the DT showed slightly better accuracy over the other methods (however, the authors affirmed that these results do not prove DT to be the best choice in all cases). This conclusion supported the studies by Mozer et al. [184], Ferreira et al. [77] and Neslin and Gupta [187].

- DTs have been successfully used in recent years, in fields like churn prediction on email users [189], supplier selection [287], churn on broadband internet [119], telecommunication companies [122, 162], and credit card users [146, 191, 280].

- **Regression analysis**: This is a standard and popular technique used by researchers dealing with the prediction of customer abandonment. Neslin and Gupta [187] state in their study that logistic regression's popularity is due to its quick and robust results as compared to other classification techniques, added to its conceptual simplicity and its closed-form solution available for posterior probabilities. Authors such as Hwang et al. [123] and Lee et al. [156] proved that logistic regression outperformed ANNs, DTs and other methods in their churn prediction studies.

  Mozer et al. [184] and Mihelis et al. [181] used regression analysis to link customer retention with satisfaction and its attributes in the fields of wireless telecommunications and private banking.

  Kim and Yoon [137] used a logistic regression (logit) model to determine subscriber churn in the telecommunications industry, based on discrete choice theory (study of behaviour in situations where decision makers must select from a finite set of alternatives). In the same field, Lemmens and Croux [157] and Lima et al. [162] also used logistic regression to predict abandonment.

  In other studies, Burez and Van den Poel [37] and Coussement and Van den Poel [49] used logit models as a reference of a well-performing method, to predict customer abandonment in a Pay-TV operator and a newspaper publisher, to be compared with more novel methods: Markov chains -in the first one- and SVMs -in the second-. In the first case, logit models showed better accuracy while, in the second case, SVM only showed a better performance when an optimal parameter-selection procedure was applied.

  More recently, Wang et al. [280] used multiple criteria decision models to evaluate the accuracy of 12 different algorithms -including logistic regression, multiple DT algorithms, and different Bayesian networks - when predicting churn on a bank's credit card holders. It was found that logistic regression yielded the highest predictive accuracy. Later, Nie et al. [191] used logistic regression and DTs to forecast customer deflection on credit card holders, and also reported a better performance of logistic regression.

Computational Intelligence methods

CI methods provide, in one form or another, flexible information processing capabilities for handling real life problems. Exploiting the tolerance for imprecision, uncertainty, approximate reasoning and partial truth in order to achieve tractability, robustness, low solution cost, and close resemblance with human-like decision making, is the goal of CI methods [198]. Techniques that fall into that category include evolutionary computation (EC), ANNs and other ML techniques, fuzzy logic (FL), and their combinations, such as neuro-fuzzy systems [104]. We shall briefly review these from the perspective of predictive models of customer abandonment:

- **Artificial Neural Networks**: An ANN is an ML model, loosely based on the biological brain (a natural neural network). It has successfully been used to estimate complex non-linear functions and has been applied to many types of problems with high predictive accuracy, such as classification, control and prediction [15, 95, 123, 139, 187, 255, 296]. One of the features that make ANN different from DT and other classification techniques is that their prediction can be interpreted as a probability. An important factor when considering the practical use of ANN is that they do not necessarily uncover patterns in an easily understandable form [8].

  Datta et al. [59] stated that ANN were still scarcely being used by companies in their day-to-day operations. A possible reason for this could be lack of clear interpretability of the output [10, 282]. In spite of that, many authors have used ANN in an entrepreneurial setting [35, 103, 120, 122, 146, 166, 224, 250] due to its high predictive accuracy. In a recent study, Tiwari et al. [247] described a novel ANN method which predicted customers that were likely to be churn in the future, with less time margin than previous models.

- **Support Vector Machines**: This is an ML method based on statistical learning theory. It is able to optimally separate two class of objects (e.g., churners and retained customers) by the creation of a multivariate hyperplane. Its theoretical basis was established on the work of Boser et al. [30] and Cortes and Vapnik [47]. SVMs have been widely used in recent studies due to its notable advantages such as a lower number of controlling parameters and a good generalization capability [38, 118]. This method, though, remains difficult to interpret in terms of the input attributes [231].

  Lessmann and Voß [158], Suryadi and Gumilang [240], Verbeke et al. [274] used SVM to predict churn in the telecommunication sector. For its part, Coussement and Van den Poel [49] applied this method to newspaper subscribers, concluding that SVMs outperform logistic regression as a predictive method.

- **DMEL (Data Mining by Evolutionary Learning)**: This algorithm aimed to overcome the limitations of interpretation and understanding of the results obtained through some CI techniques -in contrast with the clarity of the if-then-rules obtained through DT, for example-. DMEL uses non-random initial population based on first order rules. Higher order rules are then obtained iteratively using a genetic algorithm (GA) type process. The fitness value of a chromosome uses a function that defines the probability that the attribute value is correctly determined using the rules it encodes. The likelihood of prediction is estimated and the algorithm handles missing values.

  DMEL was used to predict churn in the telecommunications industry by Au et al. [8]. More recently, Yeswanth et al. [290] used a hybrid model to predict churn in mobile networks customers. They combined a pre-processing based on DT algorithms with a GA classification process.

- **Bayesian networks**: Baesens et al. [10] report an attempt to estimate whether a new customer will increase or decrease future spending. A Bayesian network was defined in this work as a probabilistic "white box" that represents a joint probability distribution over a set of discrete stochastic variables. This method has been successfully applied in recent studies of churn forecasting in the field of telecom industry by Kisioglu and Topcu [140] and in banking by Wang et al. [280].

Other alternative methods

- **Semi-Markov processes**: Used by Jenamani et al. [130] to propose a model that considers e-customer behaviour. The discrete-time semi-Markov process was designed as a probabilistic model,

for use in the analysis of complex dynamic systems. It has also been used by Slotnick and Sobel [235] in the study of cutomer retention.

- **Mixture transition distribution (MTD)**: Prinzie and Van den Poel [213] introduced a mixture transition distribution (MTD) to investigate purchase-sequence patterns. The MTD was designed to allow estimations of high order Markov chains, providing a smaller transition matrix facilitating managerial interpretation.

- **Goal-oriented sequential pattern**: Chiang et al. [46] introduced a novel algorithm for identifying potential churners using association rules that identify relationships amongst variables. The authors defined a two-step process for finding out association rules. In the first step, the large item set (attribute-value pairs) is detected, requiring compliance with certain minimum conditions of support and minimum confidence defined by the researcher. In the second stage, an *A Priori* algorithm is used to explore the rules of association.

- **Ensemble Learning**: The predictions yielded by ensemble learners are combinations of the individual predictions of multiple algorithms, and they have been shown strong and robust prediction performance. According to the combination of algorithms, the most popular ensemble learning methods in the field are:

  - **Random Forest (RF)**: It is a combination of Bagging [33], Random Subspace Method [114] and CART Decision Trees [27]. RF solves the high instability that hampers the use of DT and it has been used in several marketing research studies [35, 37, 151] due to its high predictive performance and its robustness to outliers and noise. Coussement and Van den Poel [49] found in their research about newspaper subscribers that RF outperformed SVMs and logistic regression, and recently confirmed its usefulness [48] when predicting abandonment in the online gaming industry.

  - **GAMens**: It's the combination of Bagging and Random Subspace Method (RSM) with Generalized Additive Models (GAM). GAM [18, 49, 147] is a flexible technique for nonparametric regression. The recent work by De Bock et al. [63] has proved that GAMens can be competitive with RF in accuracy. More recently, they proved the predictive ability of this method applied to problems in several industries such as supermarkets, banking and telecommunications [62].

3.1.4   Stage 4: Validation of results

Some of the most commonly used methods for model validation in the literature of the field are:

- **Cross-validation**: Most suitable in those cases in which there is a scarcity of data. Hwang et al. [123] performed validation by creating a 70/30 divide of the data. The 70% divide was used as training set and the 30% divide as validation set. Cross-validation is based on the principle of using the available data for both training and validation. Several cross-validation methods have been proposed in the literature [104], including:

  - **K-fold cross-validation**: The learning set is randomly partitioned into K subsets of equal size. Each individual subset is then used in turn for validation, while the rest of the data are used for training. In the extreme case of choosing single case folds, the procedure is called leave-one-out cross-validation

  - **Monte Carlo (or repeated random sub-sampling) cross-validation**: The learning set is repeatedly divided into two random sets, one of which is used for training and the other for validation. Not all cases are necessarily chosen at any point for validation.

- **Separate validation dataset**: Several authors [27, 59, 213] have successfully used single validation sets separated from the training sets in the validation of their predictive models of abandonment. This method should only be acceptable in those cases in which data availability is not an issue.

When testing the validity of a predicting model, or comparing the results of different methods, a set of indicators are commonly used [62, 163]:

- **Accuracy, Sensitivity and Specificity**: For classification models with a binary target variable. ***Accuracy*** measures the ratio of correctly classified observations to the overall number of cases. ***Sensitivity*** indicates the ratio of correctly predicted events (i.e., churn) to the total number of events, whereas ***specificity*** indicates the ratio of correctly predicted non-events (i.e., not churn) to the total number of non-events . Although accuracy is intuitive and commonly used to compare prediction methods [163, 191, 280], it is not considered to be an optimum figure of merit for churn modelling because it is unreliable in a situation of class imbalance [157].

- **Area under the Receiver Operating Characteristic (ROC) curve**: ROC is a function of the sensitivity versus $1 - specificity$ for all values of the classification threshold. It has been commonly used in churn prediction literature [49, 62, 157, 191, 286]. Its Area Under the Curve (AUC), unlike accuracy, evaluates the ability of a classifier to distinguish between the two classes based on the predicted class membership probabilities and is therefore suitable for imbalance classification problems such as customer churn prediction [150, 214].

- **Lift Chart**: It focuses on the segment of highest-risk customers, arranging them into deciles based on their predicted probability to churn and comparing its results with the rest of the cases. The Lift Chart has also been commonly used in recent studies [48, 62, 191]. It can be found in two different forms:

  - **Top decile Lift** (TDL): Is the churn rate in the top decile of ordered posterior churn probabilities over the churn rate in the total customer population [157].
  - **Lift Index** (LI): Is the weighted index of the correctly predicted churners, ranked by its posterior churn probability [51].

- **Loss function**: Calculated on the basis of customers' Life Time Value (LTV), this method indicates the loss caused by the error of the model, considering the effect of misclassified customers. Some examples can be found in recent work, such as Nie et al. [191] and Glady et al. [95].

3.2   A summarized review of the literature

The following summary Tables 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 3.10, 3.11, 3.12 and 3.13 list the main references in the recent literature that address the problem of building predictive models of customer abandonment. Roughly following the same scheme proposed as the guiding index for the previous section, these table show: the references to the articles; the type of data used in the analysis; the source from which these data have been obtained; the attribute selection technique employed; the possible use of time series data in its definition; the techniques used to develop the predictive models; and, finally, the method used for validation.

The tables are organized according to two main criteria:

- According to the predictive method used: Tables 3.1, 3.2 and 3.3 for standard techniques; tables 3.4, 3.5 and 3.6 for CI methods; and tables 3.7 and 3.8 for alternative ones.

- According to fields of application: Tables 3.9 and 3.10 (telecommunications), 3.11 (banking), and 3.12 and 3.13 (other areas of application).
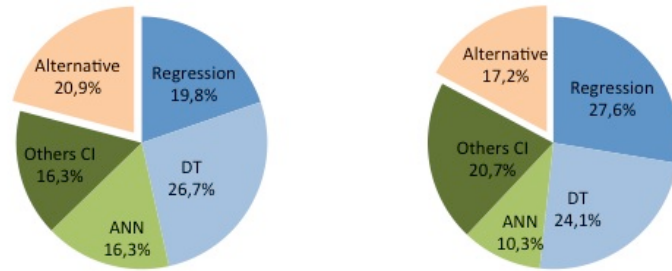
Figure 3.2: Abandonment prediction methods in recent literature. *Left*: overall summary. *Right*: summary of literature published in the last 5 years (2009-2013).

In summary, two standard methods appear as the most popular in recent literature: Decision Trees (DT) and Regression Analysis. Their more than reasonable predictive accuracy is the main reason behind the fact that more than 46% of the referred methods (see *Figure 3.2*, left) belong to these categories. They are mentioned in more than 70% of the analysed literature, acting as a reference to which compare the results of other methods, or as the main predictive method. The popularity of these methods remains high over the years, being mentioned in 51% of the references published in the last 5 years (see *Figure 3.2*, right).

For their part, *Computational Intelligence methods* are also well accepted. Specially, ANN are the more used ones (16% of the analysed literature). Their proficient predictive capacity turns them into very attractive methods for researchers, but their lack of interpretability slows down its usage in entrepreneurial environments.

Finally, the denominated "alternative methods", less specific and constantly improving, have a marginal role ($17,2\%$ of the studied literature in recent years) when referring to its usage as the main method in customer abandonment modelling.
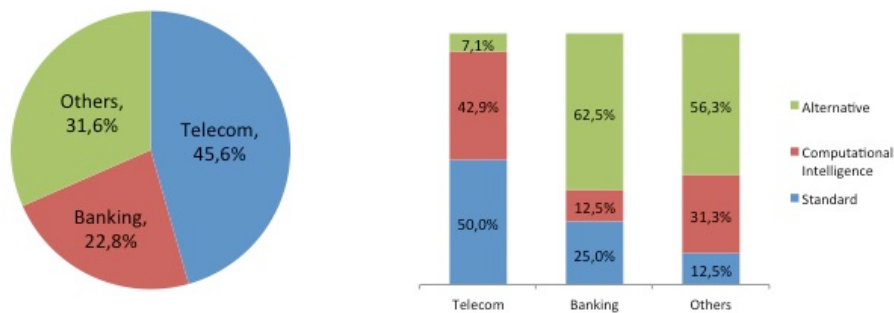
Figure 3.3: Abandonment prediction methods by field of application. *Left*: Distribution by field of application. *Right*: Detail on the type of applied methodology.

The literature review reveals, from an area of application viewpoint (see *Figure 3.3*, left), that telecommunications is the more active sector in terms of churn prediction modelling research (46% of the studied literature refer to this sector), followed by banking, found in 23% of literature. Finally, it is observed that either in banking or e-commerce industries (see *Figure 3.3*, right) *Computational Intelligence methods* and *Alternative methods* have gained influence due to the complexity of the processed data.

Authors	Data type	Data gathering	Attribute selection techniques	Time periods	Modelling techniques	Validation method
Madden et al. [175]	4 categories: use, economical, choice of provider and demographical.	Web Survey.	Not specified.	No.	Logistic Regression.	Accuracy on separate validation sets.
Datta et al. [59]	Billing, service and socio-demographic data on cell phone users.	Database.	Decision Trees and Genetic Algorithms.	Data collected on a monthly basis, updating historical data and predicting future churn.	CHART own method, as a combination of Decision Trees and Artificial Neural Networks.	Accuracy on a separate validation data set.
Mozer et al. [184]	Call details, quality (interferences and signal coverage), financial and service application (contract details, rate plan, handset type and credit report) and demographic information.	Database.	Not specified.	3 months observation and 2 months prediction.	Logistic Regression and Decision Trees (C5.0).	Lift Chart and Cross-validation.
Ng and Liu [188]	Usage data on telecom customers.	Database.	Decision Trees.	Not specified.	Decision Trees (C4.5).	Cross-validation.
Verhoef and Donkers [275]	Historical purchase behavior, socio-demographics and actual purchase of banking customers.	Survey.	Domain Knowledge.	No (survey at the moment of last purchase).	Linear Regression.	Accuracy.
Wei and Chiu [282]	Variables related to the contract (length of service, payment type, contract type) and consumption variables (minutes of use, frequency and sphere of influence).	Database.	Interviews with experts.	3 periods: Observation, retention and prediction.	Decision Trees (C4.5).	Accuracy and sensitivity.
Van den Poel [254]	RFM, behaviourals (specifics of the company, prediction of whether purchase by post will be repeated or not) and non-behaviourals (satisfaction).	Survey and Database.	Sequential Search Algorithm.	4 year historical data plus one survey. 6 months prediction.	Logistic Regression.	Accuracy and AUC.
Au et al. [8]	Customer localisation, customer type, payment method, service plan, monthly use, number of calls made and number of calls abnormally ended (251 variables)	Database.	Interviews with experts.	2 months for training and 1 month for prediction.	Decision Trees (C4.5).	Lift Chart and Accuracy values.
Ferreira et al. [77]	Billing, consumer usage, demographics, customer relationship and market data from a wireless telco company.	Company database.	Domain knowledge.	No (data collected during 9 months).	Decision Trees.	Ten-fold cross validation.
Hwang et al. [123]	Socio-demographic and usage variables.	Database.	R^2 method.	No (6 months data).	Decision Trees and Logistic Regression.	Lift Chart and cross-validation.
Kim and Yoon [137]	Company service, demographic and service use characteristics.	Survey.	Not specified.	No.	Logistic Regression.	Log likelihood.

Table 3.1: Literature on abandonment prediction modelling, listed in chronological order and corresponding to the use of standard methods (1 out of 3, continues in the next table).

Authors	Data type	Data gathering	Attribute selection techniques	Time periods	Modelling techniques	Validation method
Neslin and Gupta [187]	Customer behavior (minutes of use, revenue, handset equipment, trends in usage), company interaction data (calls to customer service) and demographics (age, income, location, house ownership) of wireless telecommunication users.	Data provided by Teradata Center for CRM	Exploratory data analysis, domain knowledge and stepwise.	Data collected on a 3-month period, churn evaluated on the 5th month.	Logistic Regression and Decision Trees.	Top decile lift and Gini coefficient, with two validation data sets.
Zhao et al. [296]	Demographics, usage level (call details, off-peak minutes used, quality of service (dropped calls, poor coverage) and marketing features (email provided, paging, rate plans offered by carrier and its competitors), on wireless telecom customers.	Company database.	All the features contained in the company data set are included.	Training data set on 3 month customer data; churn measured after 5 months.	Decision Trees.	Accuracy.
Buckinx and Van den Poel [35]	Past purchase behaviour, modeled with RFM variables from customers of Fast-Moving Consumer Goods (FMCG) retail company obtained through the use of their loyalty card and complemented with: payment method, length of customer-supplier relationship, shopping behaviour, promotional behaviour and brand purchase.	Company database.	Allusion to previous research.	5-months observation period and 5-months validation period.	Decision Trees.	Accuracy and AUC, using separate validation data sets.
Hung et al. [122]	Transaction and contract data: demographic, payment, call details and customer service data.	Database.	Interviews with experts and customers and z-test meaning.	Data collected on a 6 month period. 1 month prediction.	Decision Trees (C5.0).	5% Top-Decile Lift on separate validation data sets.
Lee et al. [156]	Gender, age, marriage status, educational level, occupation, job, position, annual income, residential status and credit limits, from banking customers.	Database.	Stepwise discriminant analysis.	Not specified.	Classification and Regression Tree (CART) and Logistic Regression.	Accuracy.
Lemmens and Croux [157]	RFM variables, minutes of customer care calls, number of adults in the household and education level, for costumers of a wireless telecommunications company.	Company database, provided by CRM Centre at Duke University.	Allusion to previous research.	Training dada from 4 non-consecutive months; validation data from a future point in time.	Decision Trees.	Top-decile lift and Gini coefficient.
Nie et al. [189]	On a charge email service: customer account data (storage bought, length of using and service), usage (number of payments in the last 3 months, total paid amount, complaints) and personal information (age, phone number provided).	Company database.	Advised by employees familiar with the database.	Historical training data tested with previously known churn data.	Decision Trees.	Accuracy.
Coussement and Van den Poel [49]	Client-company interactions, renewal-related information, socio-demographics and subscription-describing information related to subscribers of a newspaper publishing company.	Database.	Based on Random Forest importance measures.	30 months data collection and 1 year prediction period.	Logistic Regression.	Cross-validation with accuracy and AUC.
Glady et al. [95]	Evolution in time of the RFM variables (number of invoices last month, amount invoiced, number of withdrawals) related to transactions of a financial company customers.	Company database.	Allusion to previous research.	Training set corresponding to 6 consecutive months, validation set to the next 3 months.	Logistic Regression and Decision Trees.	Accuracy, Loss function and AUC.

Table 3.2: (Continues from the previous table) References on abandonment prediction modelling, listed in chronological order and corresponding to the use of standard methods (2 out of 3, continues in the next table).

Authors	Data type	Data gathering	Attribute selection techniques	Time periods	Modelling techniques	Validation method
Kumar and Ravi [146]	Sociodemographic (age, level of studies, income) and behavioural (monthly credit, number of cards, web transactions, margin) data from credit card users.	Dataset from the Business Intelligence Cup, University of Chile in 2004.	Classification and Regression Trees.	Data from 3 consecutive months; results tested during the subsequent year.	Decision trees (J48) and Logistic Regression.	Sensitivity and accuracy values, on Cross-fold validation dataset.
Huang et al. [119]	Henley segments, broadband usage, dial types, spend of dial-up, line-information, billing, payment and account information on broadband internet services customers.	Company database.	Domain knowledge.	No.	Logistic Regression and Decision Trees.	Accuracy.
Lessmann and Voß [158]	Usage data on 9 different datasets corresponding to different industries, where customer classification is needed: three credit companies, two from the mail-order industry, one from the energy industry, direct marketing, fraud detection and e-commerce.	Datasets from UCI Machine Learning Repository and the annual Data Mining Cup.	Recursive Feature Elimination.	No (both training and validation on historical data).	Logistic Regression and Decision Trees.	Ten-Fold Cross validation on AUC values.
Lima [163]	Three independent datasets related to telecom industry containing usage, billing and sociodemographic data.	Provided by CRM Centre at Duke University.	Own method, integrating Domain Knowledge in the feature selection process.	Data collected during 4 months; churn calculated on the 31-60 days period after sampling.	Logistic Regression and Decision Trees.	K-fold cross validation (k=1,5,10,50 and 100), measured on accuracy, sensitivity, specificity and AUC values.
Lima et al. [162]	Usage (average monthly minutes and revenue, days of current equipment, failed voice calls) and sociodemographical (area, ethnicity, presence of child in household) data from a mobile telecommunications company.	Dataset used in the churn tournament by Duke University in 2003.	Domain Knowledge and Stepwise selection.	Data collected during 4 non-consecutive months; churners leave the company 1 to 2 months after being sampled.	Logistic Regression and Decision Trees.	Accuracy, Sensitivity, Specificity and AUC. Cross-fold validation.
Wang et al. [280]	Customer personal information, credit card basic data, transaction and abnormal usage information.	Company database.	Domain knowledge.	12 months observation period and 12 months testing period.	Logistic Regression and Decision Trees.	Accuracy, AUC, Precision, Recall, Mean Absolute Error and computing time.
Nie et al. [191]	Customer personal information, credit card basic data, transaction and abnormal usage information.	Company database.	Delete variables with high multicollinearity.	12 months observation period and 12 months testing period.	Logistic Regression and Decision Trees.	Accuracy and AUC.
Verbeke et al. [274]	Sociodemographics, call behaviour, financial and marketing related variables from eleven wireless telecom operators.	Company and public database.	Based on Fisher score.	No (averages of 1 year data).	Decision Trees.	Accuracy, sensitivity, specificity, AUC.

Table 3.3: (Continues from the previous table) References on abandonment prediction modelling, listed in chronological order. Standard methods (3 out of 3).

Authors	Data type	Data gathering	Attribute selection techniques	Time periods	Modelling techniques	Validation method
Mozer et al. [184]	Call details, quality (interferences and signal coverage), financial and service application (contract details, rate plan, handset type and credit report) and demographic information.	Database.	Not specified.	3 months observation and 2 months prediction.	Artificial Neural Networks.	Lift Chart and cross-validation.
Datta et al. [59]	Billing, service and socio-demographic data on cell phone users.	Database.	Decision Trees and Genetic Algorithms.	Data collected on a monthly basis, updating historical data and predicting future churn.	CHART own method, as a combination of Decision Trees and Artificial Neural Networks.	Accuracy on a separate validation data set.
Behara et al. [15]	41 SERVQUAL-based service quality variables, grouped as tangibles, reliability, responsiveness, assurance and empathy, at an auto-dealership network.	Survey.	Allusion to previous research.	No (study developed at one single point in time).	Artificial Neural Networks.	Least average error, least root mean square error and accuracy values on service quality prediction.
Ho Ha et al. [115]	RFM variables in the online retail industry.	Database.	Not specified.	18 months, without specifying how many periods they are divided into.	SOM networks.	Not Specified.
Au et al. [8]	Customer localisation, customer type, payment method, service plan, monthly use, number of calls made and number of calls abnormally ended (251 variables)	Database.	Interviews with experts.	2 months for training and 1 month for prediction.	Data Mining Evolutionary Learning algorithm and Artificial Neural Networks.	Lift Chart and Accuracy values.
Baesens et al. [10]	Purchasing behaviour: volume of purchases during first 6 months as customer; "broadness" of purchases; "bargaining tendency" and "price sensitivity"; evolution averages during first 6 months	Database.	Not specified.	Yes (8 weekly periods).	Bayesian Networks.	Accuracy and AUC.
Ferreira et al. [77]	Billing, consumer usage, demographics, customer relationship and market data from a wireless telco company.	Company database.	Domain knowledge.	No (data collected during 9 months).	MLP ANN, Neuro-Fuzzy Systems and Genetic Algorithms.	Ten-fold cross validation.
Hsieh [117]	Customer attributes (household income, occupation, sex, age, etc.) and credit card usage (number of purchases, amount of consumption, card type, etc.) for banking customers.	Company database.	A priori association over the different customer segments.	Data collected in a 12 months period; no prediction period.	SOM networks.	Validation based on "goodness of segmentation"; no churn prediction in this study.
Hwang et al. [123]	Socio-demographic and usage variables.	Database.	R^2 method.	No (6 months data).	Artificial Neural Networks.	Lift Chart and cross-validation.
Shin and Sohn [231]	Past transactions carried by the costumer.	Database.	Not specified.	No (3 month: averages and totals).	K-means, SOM networks and Fuzzy K-means.	Intra-class method for cluster compactness.

Table 3.4: References on abandonment prediction modelling, listed in chronological order and for the use of CI methods (1 out of 3, continues in the next table).

Authors	Data type	Data gathering	Attribute selection techniques	Time periods	Modelling techniques	Validation method
Neslin and Gupta [187]	Customer behavior (minutes of use, revenue, handset equipment, trends in usage), company interaction data (calls to customer service) and demographics (age, income, location, house ownership) of wireless telecommunication users.	Data provided by Teradata Center for CRM.	Exploratory data analysis, domain knowledge and stepwise.	Data collected on a 3-month period, churn evaluated on the 5th month.	Artificial Neural Networks.	Top decile lift and Gini coefficient, with two validation data sets.
Zhao et al. [296]	Demographics, usage level (call details, off-peak minutes used), quality of service (dropped calls, poor coverage) and marketing features (email provided, paging, rate plans offered by carrier and its competitors), on wireless telecom customers.	Company database.	All the features contained in the company data set are included.	Training data set on 3 month customer data; churn measured after 5 months.	SVMs, ANNs and Bayesian Networks.	Accuracy.
Buckinx and Van den Poel [35]	Past purchase behaviour, modeled with RFM variables from customers of Fast-Moving Consumer Goods (FMCG) retail company obtained through the use of their loyalty card and complemented with: payment method, length of customer-supplier relationship, shopping behaviour, promotional behaviour and brand purchase.	Company database.	Allusion to previous research.	5-months observation period and 5-months validation period.	Automatic Relevance Determination Neural Networks.	Accuracy and AUC, using separate validation sets.
Lee et al. [156]	Gender, age, marriage status, educational level, occupation, job, position, annual income, residential status and credit limits, from banking customers.	Database.	Stepwise discriminant analysis.	Not specified.	SVMs and ANNs.	Accuracy.
Hung et al. [122]	Transaction and contract data: demographic, payment, call details and customer service data.	Database.	Interviews with experts and customers and z-test meaning.	Data collected on a 6-month period. 1-month prediction.	ANNs.	5% Top-Decile Lift on separate validation data sets.
Coussement and Van den Poel [49]	Client-company interactions, renewal-related information, socio-demographics and subscription-describing information related to subscribers of a newspaper publishing company.	Database.	Based on Random Forest importance measures.	30 months data collection and 1-year prediction period.	SVMs.	Cross-validation with accuracy and AUC.
Hadden [103]	Data related to customer repairs and complaints, applied to three different cases related to telecommunications (residential mobile phone, broadband mobile phone, business landline).	Database	Interviews with company experts.	1 month for training and 12 for prediction.	ANNs	Accuracy.
Suryadi and Gumilang [240]	Usage variables (monthly minutes, overage minutes, roaming calls), relationship (number of customer care calls, months in service, refurbished handset) and demographics (town, suburban) for wireless telecommunications customers.	Database provided by the Center for CRM at Duke University.	Interview with experts.	Not specified.	SVMs.	Accuracy, with a 80% training data set and 20% validation dataset.
Glady et al. [95]	Evolution in time of the RFM variables (number of invoices last month, amount invoiced, number of withdrawals) related to transactions of a financial company customers.	Company database.	Allusion to previous research.	Training set corresponding to 6 consecutive months, validation set to the next 3 months.	ANNs.	Accuracy, loss function and AUC.
Kumar and Ravi [146]	Socio-demographic (age, level of studies, income) and behavioural (monthly credit, number of cards, web transactions, margin) data from credit card users.	Data set from the Business Intelligence Cup, University of Chile in 2004.	SVMs.	Data from 3 consecutive months; results tested during the subsequent year.	SVMs.	Sensitivity and accuracy values, on cross-validation data set.

Table 3.5: (Continues from the previous table) References on abandonment prediction modelling, listed in chronological order and for CI methods (2 out of 3, continues in the next table).

Authors	Data type	Data gathering	Attribute selection techniques	Time periods	Modelling techniques	Validation method
Huang et al. [119]	Henley segments, broadband usage, dial types, spend of dial-up, line-information, billing, payment and account information on broadband internet services customers.	Company database.	Domain knowledge.	No.	SVMs and ANNs.	Accuracy.
Lessmann and Voß [158]	Usage data on 9 different data sets corresponding to different industries, where customer classification is needed: three credit companies, two from the mail-order industry, one from the energy industry, direct marketing, fraud detection and e-commerce.	Data sets from UCI Machine Learning Repository and the annual Data Mining Cup.	Recursive Feature Elimination.	No (both training and validation on historical data).	SVM.	Ten-Fold Cross validation on AUC values.
Tsai and Lu [250]	CRM variables from a telecommunications company.	Company database.	The primary features provided by the company are used.	Data collected during 6 months, churners leave 1 to 2 months after being sampled.	SOMs and ANNs.	Five-fold cross validation.
Lima [163]	Three independent data sets related to telecom industry containing usage, billing and socio-demographic data.	Provided by CRM Centre at Duke University.	Own method, integrating Domain Knowledge in the feature selection process.	Data collected during 4 months; churn calculated on the 31-60 days period after sampling.	ANNs.	K-fold cross validation (k=1,5,10,50 and 100), measured on accuracy, sensitivity, specificity and AUC values.
Wang et al. [280]	Customer personal information, credit card basic data, transaction and abnormal usage information.	Company database.	Domain knowledge.	12 months observation period and 12 months testing period.	Bayesian networks and clustering-based classifiers.	Accuracy, AUC, Precision, Recall, Mean Absolute Error and computing time.
Kisioglu and Topcu [140]	Customer average and trends (grow, maintain or decrease) in call minutes, frequency of calls and billing, added to demographics as place of residence, age and tariff type, for a telecommunications company.	Company Database.	Domain knowledge.	Not specified.	Bayesian Networks.	Accuracy.
Tiwari et al. [247]	35 variables from a real industry retailer, with no specification on the kind of features included.	Company Database.	Not specified.	Data from 13 consecutive months.	ANNs.	Accuracy, sensitivity and specificity.
Verbeke et al. [274]	Sociodemographics, call behaviour, financial and marketing related variables from eleven wireless telecom operators.	Company and public database.	Based on Fisher score.	No (averages of 1 year data).	SVMs, ANNs, Bayesian Networks.	Accuracy, sensitivity, specificity, AUC.
Chen et al. [45]	Purchase history (longitudinal) data and socio-demographic (static) variables on three different fields: supermarkets, leisure and telecom	Database.	Interviews with experts.	1-year data collection and 1-year prediction, with seasonal overlapping.	SVMs.	AUC.

Table 3.6: (Continues from previous table) References on abandonment prediction modelling, listed in chronological order and using CI methods (3 out of 3).

Authors	Data type	Data gathering	Attribute selection techniques	Time periods	Modelling techniques	Validation method
Pfeifer and Carraway [208]	RFM variables to model customer-firm relationship and optimize marketing expenditure, which can be applied to many industries.	Artificial data.	RFM variables adapted to the needs of the different examples provided.	Not Specified.	Markov Chain Modelling.	No validation stage, the study is focussed on the theoretical side of the problem.
Chiang et al. [46]	Transaction data (frequency of transaction of banking customers are analysed).	Database.	Rule extraction from variables; selection of the implied variables.	Data from the last operative month from the last 6 months.	Goal-oriented sequential pattern.	Computational efficiency against existing algorithms.
Jenamani et al. [130]	E-customer behaviour, defining customer usage through the website as a mixture of states, transitions between states, holding time, waiting time and total time spent; using web analytics.	Visitor's navigational data collected from the website.	Domain knowledge.	No (data collected during 3 months with no validation stage).	Semi-Markov Processes.	No validation stage; the model pretends to understand and describe e-customer behaviour as tracked across the website.
Jonker et al. [134]	RFM variables to segment customers from a direct-mail setting for a charity organization: number of non-answered mails, % response in the last 2 years and overall, size of responses in the last 2 years and overall.	Database.	Domain knowledge.	No.	Genetic Algorithms and K-means clustering.	Validation based on the real effects of the improved segmentation on the customer base.
Van den Poel and Laraviére [255]	Customer behaviour information, socio-demographic information, merger and prosperity index.	Database.	Not specified.	77 year database. Length of sub-periods are not specified.	Survival analysis.	Statistical relevance.
Buckinx and Van den Poel [35]	Past purchase behaviour, modeled with RFM variables from customers of Fast-Moving Consumer Goods (FMCG) retail company obtained through the use of their loyalty card and complemented with: payment method, length of customer-supplier relationship, shopping behaviour, promotional behaviour and brand purchase.	Company database.	Allusion to previous research.	5-months observation period and 5-months validation period.	Random Forests.	Accuracy and AUC, using separate validation sets.
Larivière and Van den Poel [151]	Past customer behavior (Specific product ownership, use of e-banking, number of products owned and monetary value, cross-buying), demographics and interaction with intermediaries for banking and insurance customers.	Company database.	Domain knowledge.	Data collected at a specific date, evaluated during the following 9 months.	Random Forests.	AUC with separate validation datasets.
Liu and Shih [170]	RFM variables from customers of a hardware retail company.	Company database.	Analytical Hierarchical Process and interviews with experts.	No (data from the 2 previous years).	Preference-based Collaborative Filtering for customer clustering.	Validation focussed on the goodness of the segmentation.
Slotnick and Sobel [235]	Model for optimizing a manufacturing firm's profit, based on customer specifications and processing time, backlog and due-date quotation.	Statistically modeled with artificial data.	Previous domain knowledge.	Not Specified.	Discrete-time Semi-Markov Decision Processes.	Fractional error, weighted absolute value and weighted sum of differences on the lead times.

Table 3.7: References on abandonment prediction modelling, listed in chronological order and concerning the use of alternative methods (1 out of 2, continues in the following table).

Authors	Data type	Data gathering	Attribute selection techniques	Time periods	Modelling techniques	Validation method
Prinzie and Van den Poel [213]	Purchase behaviour, seen as a sequence of the historical purchased products (grouped in 9 sets of products) of the customers of a Financial Services company.	Company database.	Domain Knowledge.	Not specified.	Mixture Transition Distribution and Markov chains.	Log likelihood for fitness of results and number of parameters. Separate validation data sets.
Burez and Van den Poel [37]	Usage variables on customers from a Pay-TV company.	Company database.	Undersampling.	Data collection at a specific date, evaluation during the subsequent year.	Markov Chains and Random Forests.	Accuracy, AUC and Lift Chart with cross-validation.
Kumar and Ravi [146]	Socio-demographic (age, level of studies, income) and behavioural (monthly credit, number of cards, web transactions, margin) data from credit card users.	Data set from the Business Intelligence Cup, University of Chile in 2004.	SVMs.	Data from 3 consecutive months; results tested during the subsequent year.	Random Forest.	Sensitivity and accuracy values for cross-validation data set.
Coussement and Van den Poel [49]	Client-company interactions, renewal-related information, socio-demographics and subscription-describing information related to subscribers of a newspaper publishing company.	Database.	Based on Random Forest importance measures.	30 months data collection and 1 year prediction period.	Random Forest.	Cross-validation with accuracy and AUC.
Wu [287]	Variables defining the performance of suppliers of a telecommunications company, scored in a 0-1 range. Quality management practices and systems, documentation and self-audit, process and manufacturing capability, management of the firm, design and development capabilities, and cost reduction capability.	Company database, available in previous studies.	Primary features in the data set.	Validation and test data sets correspond to different suppliers on different periods.	Hybrid model combining Data Envelopment Analysis, Decision Trees and ANNs for future performance prediction.	Four-fold cross validation and accuracy values.
De Bock et al. [63]	Demographics, historical transactional information and financial variables on different cases (supermarkets, banking, telecom and mailing services).	Database.	Undersampling.	1 month data collection and 1 year performance period.	Ensemble Classifiers (GAMens).	Accuracy, AUC, Top Decile Lift and Lift Index.
Verbeke et al. [274]	Socio-demographics, call behaviour, financial and marketing related variables from eleven wireless telecom operators.	Company and public database.	Based on Fisher score.	No (averages of 1 year data).	Ensemble methods.	Accuracy, sensitivity, specificity, AUC.
Yeswanth et al. [290]	Usage time, location and customers' underlying social network to predict their churn from a mobile operator.	Company Database.	Research on previous literature.	Training during 3 consecutive months: churn predicted comparing variation between periods.	Hybrid models, combining Decision Trees Genetic Algorithms and Game Theory.	Accuracy.
Coussement and De Bock [48]	RFM variables to define past customer behaviour for an online-gaming website.	Company database.	Variable importance score, related to its predictive accuracy.	17 months for data collection and training, 4 months for churn measurement.	Ensemble models: Random Forests and GAMens.	Top-Decile Lift and Lift Index.

Table 3.8: (Continues from the previous table) References on abandonment prediction modelling, listed in chronological order and concerning alternative methods (2 out of 2).

Authors	Data type	Data gathering	Attribute selection techniques	Time periods	Modelling techniques	Validation method
Madden et al. [175]	4 categories: use, economical, choice of provider and demographical.	Web Survey.	Not specified.	No.	Logistic Regression.	Accuracy on separate validation sets.
Datta et al. [59]	Billing, service and socio-demographic data on cell phone users.	Database.	Decision Trees and Genetic Algorithms.	Data collected on a monthly basis, updating historical data and predicting future churn.	CHART own method, as a combination of Decision trees and ANNs.	Accuracy on a separate validation data set.
Mozer et al. [184]	Call details, quality (interferences and signal coverage) financial and service application (contract details, rate plan, handset type and credit report) and demographic information.	Database.	Not specified.	3 months observation and 2 months prediction.	Logistic Regression, Decision Trees (C5.0) and ANNs.	Lift Chart and Cross-validation.
Ng and Liu [188]	Usage data on telecom customers.	Database.	Decision Trees.	Not specified.	Decision Trees (C4.5).	Cross-validation.
Wei and Chiu [282]	Variables related to the contract (length of service, payment type, contract type) and consumption variables (minutes of use, frequency and sphere of influence).	Database.	Interviews with experts.	3 periods: Observation, retention and prediction.	Decision Trees (C4.5).	Accuracy and sensitivity.
Au et al. [8]	Customer localisation, customer type, payment method, service plan, monthly use, number of calls made and number of calls abnormally ended (251 variables)	Database.	Interviews with experts.	2 months for training and 1 month for prediction.	Data Mining Evolutionary Learning algorithm, Decision Trees (C4.5) and ANNs	Lift Chart and Accuracy values.
Ferreira et al. [77]	Billing, consumer usage, demographics, customer relationship and market data from a wireless telco company.	Company database.	Domain knowledge.	No (data collected during 9 months).	Decision Trees, MLP ANNs, Neuro-Fuzzy Systems and Genetic Algorithms.	Ten-fold cross validation.
Hwang et al. [123]	Socio-demographic and usage variables.	Database.	R^2 method.	No (6 months data).	Neural Networks, Decision Trees and Logistic Regression.	Lift Chart and Cross-fold validation.
Kim and Yoon [137]	Company service, demographic and service use characteristics.	Survey.	Not specified.	No.	Logistic Regression.	Log likelihood.
Neslin and Gupta [187]	Customer behavior (minutes of use, revenue, handset equipment, trends in usage), company interaction data (calls to customer service) and demographics (age, income, location, house ownership) of wireless telecommunication users.	Data provided by Teradata Center for CRM.	Exploratory data analysis, domain knowledge and step-wise.	Data collected on a 3-month period, churn evaluated on the 5th month.	Logistic Regression, Decision Trees and ANNs.	Top decile lift and Gini coefficient, with two validation data sets.
Zhao et al. [296]	Demographics, usage level (call details, off-peak minutes used), quality of service (dropped calls, poor coverage) and marketing features (email provided, paging, rate plans offered by carrier and its competitors), on wireless telecom customers.	Company database.	All the features contained in the company data set are included.	Training data set on 3 month customer data; churn measured after 5 months.	SVMs, ANNs, Decision Trees and Bayesian Networks.	Accuracy.
Hung et al. [122]	Transaction and contract data: demographic, payment, call details and customer service data.	Database.	Interviews with experts and customers and z-test meaning.	Data collected on a 6 month period. 1 month prediction.	Decision Trees (C5.0), ANN.	5% Top-Decile Lift on separate validation data sets.
Lemmens and Croux [157]	RFM variables, minutes of customer care calls, number of adults in the household and education level, for costumers of a wireless telecommunications company.	Company database, provided by CRM Centre at Duke University.	Allusion to previous research.	Training dada from 4 non-consecutive months; validation data from a future point in time.	Decision Trees.	Top-decile lift and Gini coefficient.
Suryadi and Gumilang [240]	Usage variables (monthly minutes, overage minutes, roaming calls), relationship (number of customer care calls, months in service, refurbished handset) and demographics (town, suburban) for wireless telecommunications customers.	Database provided by the Center for CRM at Duke University.	Interview with experts.	Not specified.	SVMs.	Accuracy, with a 80% training data set and 20% validation data set.

Table 3.9: References on abandonment prediction modelling, listed in chronological order for the telecommunications application area (1 out of 2, continues in the next table).

Authors	Data type	Data gathering	Attribute selection techniques	Time periods	Modelling techniques	Validation method
Hadden [103]	Data related to customer repairs and complaints, applied to three different cases related to telecommunications (residential mobile phone, broadband mobile phone, business landline).	Database	Interviews with company experts.	1 month for training and 12 for prediction.	ANNs	Accuracy.
Huang et al. [119]	Henley segments, broadband usage, dial types, spend of dial-up, line-information, billing, payment and account information on broadband internet services customers.	Company database.	Domain knowledge.	No.	Logistic Regression, Decision Trees, ANNs and Support Vector Machines.	Accuracy.
Lessmann and Voß [158]	Usage data on 9 different datasets corresponding to different industries, where customer classification is needed: three credit companies, two from the mail-order industry, one from the energy industry, direct marketing, fraud detection and e-commerce.	Datasets from the UCI Machine Learning Repository and the annual Data Mining Cup.	Recursive Feature Elimination.	No (both training and validation on historical data).	SVM model, Logistic Regression and Decision Trees.	Ten-Fold Cross validation on AUC values.
Tsai and Lu [250]	CRM variables from a telecommunications company.	Company database.	The primary features provided by the company are used.	Data collected during 6 months, churners leave 1 to 2 months after being sampled.	ANNs and SOM.	Five-fold cross validation.
Wu [287]	Variables defining the performance of suppliers of a telecommunications company, scored in a 0-1 range. Quality management practices and systems, documentation and self-audit, process and manufacturing capability, management of the firm, design and development capabilities, and cost reduction capability.	Company database, available in previous studies.	Primary features in the data set.	Validation and test data sets correspond to different suppliers on different periods.	Hybrid model combining Data Envelopment Analysis, Decision Trees and ANNs for future performance prediction.	Four-fold cross validation and accuracy values.
Lima [163]	Three independent data sets related to telecom industry containing usage, billing and sociodemographic data.	Provided by CRM Centre at Duke University.	Own method, integrating Domain Knowledge in the feature selection process.	Data collected over 4 months; churn calculated on the 31-60 days period after sampling.	Logistic Regression, Decision Trees and ANNs.	K-fold cross validation (k=1,5,10,50 and 100), measured on accuracy, sensitivity, specificity and AUC values.
De Bock et al. [63]	Demographics, historical transactional information and financial variables on different cases (supermarkets, banking, telecom and mailing services).	Database.	Undersampling.	1 month data collection and 1 year performance period.	Ensemble Classifiers (GAMens)	Accuracy, AUC, Top Decile Lift and Lift Index.
Kisioglu and Topcu [140]	Customer average and trends (grow, maintain or decrease) in call minutes, frequency of calls and billing, added to demographics as place of residence, age and tariff type, for a telecommunications company.	Company Database.	Domain knowledge.	Not specified.	Bayesian Networks.	Accuracy.
Lima et al. [162]	Usage (average monthly minutes and revenue, days of current equipment, failed voice calls) and socio-demographic (area, ethnicity, presence of childs in household) data from a mobile telecommunications company.	Data set used in the churn tournament by Duke University in 2003.	Domain Knowledge and Stepwise selection.	Data collected during 4 non-consecutive months; churners leave the company 1 to 2 months after being sampled.	Logistic Regression and Decision Trees.	Accuracy, Sensitivity, Specificity and AUC. Cross-validation.
Verbeke et al. [274]	Socio-demographics, call behaviour, financial and marketing related variables from eleven wireless telecom operators.	Company and public database.	Based on Fisher score.	No (averages of 1 year data).	Decision Trees, Ensemble methods, ANNs, Bayesian Networks, SVMs.	Accuracy, sensitivity, specificity, AUC.
Yeswanth et al. [290]	Usage time, location and customers' underlying social network to predict their churn from a mobile operator.	Company Database.	Research on previous literature.	Training during 3 consecutive months: churn predicted comparing variation between periods.	Hybrid models, combining Decision Trees Genetic Algorithms and Game Theory.	Accuracy.
Chen et al. [45]	Purchase history (longitudinal) data and socio-demographic (static) variables on three different fields: supermarkets, leisure and telecom	Database.	Interviews with experts.	1-year data collection and 1-year prediction, with seasonal overlapping.	SVMs.	AUC.

Table 3.10: (Continues from previous table) References on abandonment prediction modelling, listed in chronological order for the telecommunications application area (2 out of 2).

Authors	Data type	Data gathering	Attribute selection techniques	Time periods	Modelling techniques	Validation method
Verhoef and Donkers [275]	Historical purchase behavior, socio-demographics and actual purchase on banking customers.	Survey.	Domain Knowledge.	No (survey at the moment of last purchase).	Regression.	Accuracy.
Chiang et al. [46]	Transaction data (frequency of transaction of banking customers are analysed).	Database.	Rule extraction between variables; selection of the implied variables.	Data from the last operative month from the last 6 months.	Goal oriented sequential pattern.	Computational efficiency against existing algorithms.
Shin and Sohn [231]	Past transactions carried by the costumer.	Database.	Not specified.	No (3 month: averages and totals).	K-means, SOM networks and Fuzzy K-means.	*Intraclass* method for cluster compactness.
Van den Poel and Laravière [255]	Customer behaviour information, socio-demographic information, merger and prosperity index.	Database.	Not specified.	77 year database. Length of sub-periods are not specified.	Survival analysis.	Statistical relevance.
Hsieh [117]	Customer attributes (household income, occupation, sex, age, …) and credit card usage (number of purchases, amount of consumption, card type, …) for banking customers.	Company database.	A priori association over the different customer segments.	Data collected in a 12 months period, no prediction period.	SOM.	Validation based on goodness of segmentation, no churn prediction in this research.
Larivière and Van den Poel [151]	Past customer behavior (Specific product ownership, use of e-banking, number of products owned and monetary value, cross-buying), demographics and interaction with intermediaries for banking and insurance customers.	Company database.	Domain knowledge.	Data collected at a specific date, evaluated during the next 9 months.	Random Forests.	AUC with separate validation datasets.
Prinzie and Van den Poel [213]	Purchase behaviour, seen as a sequence of the historical purchased products (grouped in 9 sets of products), of a Financial Services company customers.	Company database.	Domain Knowledge.	Not specified.	Mixture transition distribution and Markov chains.	Log likelihood for fitness of results and number of parameters. Separate validation data sets.
Lee et al. [156]	Gender, age, marriage status, educational level, occupation, job, position, annual income, residential status and credit limits, from banking customers.	Database.	Stepwise discriminant analysis.	Not specified.	Classification and Regression Tree (CART), Logistic regression, ANNs and SVMs.	Accuracy.
Glady et al. [95]	Evolution in time of the RFM variables (number of invoices last month, amount invoiced, number of withdrawals) related to transactions of a financial company customers.	Company database.	Allusion to previous research.	Training set (66%) corresponding to 6 consecutive months, validation set (33%) to the next 3 months.	Logistic Regression, Decision Trees and ANNs.	Accuracy, Loss function and AUC.
Kumar and Ravi [146]	Socio-demographic (age, level of studies, income) and behavioural (monthly credit, number of cards, web transactions, margin) data from credit card users.	Dataset from the Business Intelligence Cup, University of Chile in 2004.	Classification and Regression Trees.	Data from 3 consecutive months; results tested during the subsequent year.	Decision trees (J48), Logistic Regression, Random Forest and SVMs.	Sensitivity and accuracy values, on Cross-validation data set.
De Bock et al. [63]	Demographics, historical transactional information and financial variables on different cases (supermarkets, banking, telecom and mailing services).	Database.	Under-sampling.	1-month data collection and 1-year performance period.	Ensemble Classifiers (GAMens).	Accuracy, AUC, Top Decile Lift and Lift Index.
Wang et al. [280]	Customer personal information, credit card basic data, transaction and abnormal usage information.	Company database.	Domain knowledge.	12 months observation period and 12 months testing period.	Decision Trees, Logistic regression, Bayesian Networks and clustering-based classifiers.	Accuracy, AUC, Precision, Recall, Mean Absolute Error and computing time.
Nie et al. [191]	Customer personal information, credit card basic data, transaction and abnormal usage information.	Company database.	Delete variables with high multicollinearity.	12 months observation period and 12 months testing period.	Logistic Regression and Decision Trees.	Accuracy and AUC.

Table 3.11: References on abandonment prediction modelling, listed in chronological order for the Banking and Financial Services field.

Authors	Data type	Data gathering	Attribute selection techniques	Time periods	Modelling techniques	Validation method
Retail						
Ho Ha et al. [115]	RFM variables in the online retail industry.	Database.	Not specified.	18 months, without specifying how many periods they are divided into.	SOM networks.	Not Specified.
Baesens et al. [10]	Purchasing behaviour: volume of purchases during first 6 months as customer; "broadness" of purchases; "bargaining tendency" and "price sensitivity"; evolution averages during first 6 months	Database.	Not specified.	Yes (8 weekly periods).	Bayesian Networks.	Accuracy and AUC.
Buckinx and Van den Poel [35]	Past purchase behaviour, modeled with RFM variables from customers of Fast-Moving Consumer Goods (FMCG) retail company obtained through the use of their loyalty card and complemented with: payment method, length of customer-supplier relationship, shopping behaviour, promotional behaviour and brand purchase.	Company database.	Allusion to previous research.	5-months observation period and 5-months validation period.	Random Forests, Automatic Relevance Determination ANNs and Logistic Regression.	Accuracy and AUC, using separate validation sets.
Liu and Shih [170]	RFM variables from customers of a hardware retail company.	Company database.	Analytical Hierarchical Process and interviews with experts.	No (data from the 2 previous years).	Preference-based Collaborative Filtering for customer clustering.	Validation focussed on the goodness of the segmentation.
De Bock et al. [63]	Demographics, historical transactional information and financial variables on different cases (supermarkets, banking, telecom and mailing services).	Database.	Undersampling.	1 month data collection and 1 year performance period.	Ensemble Classifiers (GAMens)	Accuracy, AUC, Top Decile Lift and Lift Index.
Tiwari et al. [247]	35 variables from a real industry retailer, with no specification on the kind of features included.	Company Database.	Not specified.	Data from 13 consecutive months.	ANNs.	Accuracy, sensitivity and specificity.
Mail and delivery services						
Behara et al. [15]	41 SERVQUAL-based service quality variables, grouped as tangibles, reliability, responsiveness, assurance and empathy, at an auto-dealership network.	Survey.	Allusion to previous research.	No (study developed at one single point in time).	ANNs.	Least average error, least root mean square error and Accuracy values on service quality prediction.
Van den Poel [254]	RFM, behavioural (specifics of the company, prediction of whether purchase by post will be repeated or not) and non-behavioural (satisfaction).	Survey and Database.	Sequential Search Algorithm.	4 year historical data plus one survey. 6 months prediction.	Logistic Regression.	Accuracy and AUC.
Jonker et al. [134]	RFM variables to segment customers from a direct-mail setting for a charity organization: number of non-answered mails, % response in the last 2 years and overall, size of responses in the last 2 years and overall.	Database.	Domain knowledge.	No.	Genetic Algorithms and K-means clustering.	Validation based on the real effects of the improved segmentation on the customer base.

Table 3.12: References on abandonment prediction modelling, listed in chronological order for the Retail, Online Purchasing and Other fields of application (1 out of 2).

Mail and delivery services

Authors	Data type	Data gathering	Attribute selection techniques	Time periods	Modelling techniques	Validation method
Nie et al. [189]	On a charge email service, customer account data (storage bought, length of using and service), usage (number of payments in the last 3 months, total paid amount, complaints) and personal information (age, phone number provided).	Company database.	Advised by employees familiar to the database.	Historical training data tested with previously known churn data.	Decision Trees.	Accuracy.
Coussement and Van den Poel [49]	Client-company interactions, renewal-related information, socio-demographics and subscription-describing information related to subscribers of a newspaper publishing company.	Database.	Based on Random Forest importance measures.	30 months data collection and 1 year prediction period.	SVMs, Logistic Regression, Random Forest.	Cross-validation with accuracy and AUC.
De Bock et al. [63]	Demographics, historical transactional information and financial variables on different cases (supermarkets, banking, telecom and mailing services).	Database.	Undersampling.	1 month data collection and 1 year performance period.	Ensemble Classifiers (GAMens)	Accuracy, AUC, Top Decile Lift and Lift Index.

Others

Authors	Data type	Data gathering	Attribute selection techniques	Time periods	Modelling techniques	Validation method
Jenamani et al. [130]	E-customer behaviour, defining its customer usage through the website as a mixture of states, transitions between states, holding time, waiting time and total time spent; using web analytics.	Visitor's navigational data collected from the website.	Domain knowledge.	No (data collected during 3 months with no validation stage).	Semi-Markov Processes.	No validation stage; the model pretends to understand and describe e-customer behaviour as its track along the website.
Pfeifer and Carraway [208]	RFM variables to model customer-firm relationship and optimize marketing expenditure, which can be applied to many industries.	Artificial data.	RFM variables adapted to the needs of the different examples provided.	Not Specified.	Markov Chain Modelling.	No validation stage, the study is focussed on the theoretical basis.
. Slotnick and Sobel [235]	Model for optimizing a manufacturing firm's profit, based on customer specifications and processing time, backlog and due-date quotation.	Statistically modeled with artificial data.	Previous domain knowledge.	Not Specified.	Discrete-time Semi-Markov Decision Processes.	Fractional error, weighted absolute value and weighted sum of differences on the lead times.
Burez and Van den Poel [37]	Usage variables on customers from a Pay-TV company.	Company database.	Undersampling.	Data collection at a specific date, evaluation during the subsequent year.	Markov Chains, Random Forests and Logistic Regression.	Accuracy, AUC and Lift Chart with Cross-fold validation.
Chen et al. [45]	Purchase history (longitudinal) data and sociodemographic (static) variables on three different fields: supermarkets, leisure and telecom	Database.	Interviews with experts.	1-year data collection and 1-year prediction, with seasonal overlapping.	SVMs.	AUC.
Coussement and De Bock [48]	RFM variables to define past customer behaviour for an online-gaming website.	Company database.	Variable importance score, related to its predictive accuracy.	17 months for data collection and training, 4 months for churn measurement.	Ensemble models as Random Forests and GAMens.	Top-Decile Lift and Lift Index.

Table 3.13: (Continues from previous table) References on abandonment prediction modelling, listed in chronological order for Retail, Online Purchasing and other application areas (2 out of 2).

Chapter 4

Manifold learning: visualizing and clustering data

In this Thesis, the DM framework mostly concerns the use of unsupervised machine learning techniques. Within the overall goal of exploring the existence of customer churn routes according to the customers' service consumption patterns, we are interested in methods that are capable of providing simultaneous visualization and clustering of the available data.

To this end, in *Section 4.1*, general latent models are first introduced, given that they are the basis of the methods used in following chapters. The need to find less constrained and, thus, more flexible methods for data modeling has led research towards the exploration and definition of NLDR techniques, which are becoming increasingly popular [155]. The most interesting contributions to this area range from spectral-based methods to manifold learning techniques. The best-known and most-used NLDR method is SOM Kohonen [142], in its many existing variants; for this reason, and also because it inspired generative models, we include a brief description of its basic forms in *Section 4.2*. This is followed, in *Section 4.3*, by the introduction of the standard version of GTM, the probabilistic counterpart of SOM.

4.1   Latent variable models

Data visualization methods must deal with the problem of finding low-dimensional representations of multivariate data residing on high-dimensional data spaces. This problem can be summarized as follows:

Given N sample vectors $\{y_n\}_{n=1}^{N} \subseteq \Re^D$ drawn from the random vector $\overrightarrow{Y}$, find $\mathcal{G} : \Re^D \to \Re^L$ and $\mathcal{F} : \Re^L \to \Re^D$ such that $\forall n = 1, ..., N$

$$\mathcal{G}(y_n) = x_n$$
$$\mathcal{F}(x_n) = \hat{y}_n \approx y_n$$

where $\{x_n\}_{n=1}^{N} \subseteq \Re^L$ denotes the corresponding set of reduced sample vectors drawn from the random vector and D, L denote the dimensionality of the original data and reduced latent spaces, respectively.

Latent variable models address this problem by representing information from an observable, usually high dimensional data space, on an unobservable or latent low-dimensional space. These models can be typified as belonging to different, but overlapping categories: *projection models, generative models* and *other related models* [241].

Projection models aim for the projection of data points residing in $\Re^D$, onto a hyperplane, $\Re^L$, with $\Re^L \subset \Re^D$, where $L \leq D$ in such a way that the distortion introduced by the projection is minimal. The most popular, and widely used, projection model is the Principal Components Analysis (PCA) [131], although other models, such as principal curves and surfaces [108], auto-associative feed-forward neural networks [145] and kernel based PCA [47, 226] are also commonly used.

Generative models are defined stochastically and try to estimate the distribution of data under a set of constraints that restrict the set of possible models to those with a low intrinsic dimensionality. The key assumption of generative models is that the data variables are conditionally independent given the latent variables. This imposes the constraint that the latent variables should carry the dependency structure of the data. Assuming a particular model structure M, the data distribution is obtained by marginalizing over the latent variables. Possibly, Factor Analysis (FA) [12, 153] is the most widely used generative model. GTM [23], described in the following sections, is in fact defined like a generative model: non-linear in nature and a probabilistic alternative to the SOM [143].

Most of the interest in generative models stems from the fact that they fit naturally into the much wider framework of probability theory and statistics. Furthermore, generative models can resort to well-founded techniques for fitting them to data, combining different models, missing data imputation, outlier detection, etc.

4.2 Self Organizing Maps

A well-know and widely used NLDR method for data visualization in low-dimensional spaces is SOM. Despite the fact that it is neither a latent model nor a manifold learning model as such, it provides functionalities that are akin to both. This is an unsupervised bio-inspired artificial neural network introduced by Kohonen [142]. From its origins as a computational neuroscience neural network simulation technique, it has advanced in a few decades to become a considerable success in a wide range of applications, including business [166, 263].

SOM combines Vector Quantization (leading to clustering) and low-dimensional topographically-ordered data representation (leading to data visualization). Its nonlinearity has not prevented SOM to achieve mainstream status, even in very practical application fields.

4.2.1 Basic SOM

Let X be a data space with N samples $\mathbf{x}$ of dimension d. A SOM consists of a discrete layer (map) of prototypes (also called units or neurons due to its computational neuroscience origins) arranged in a low dimensional regular grid (usually 2-D to allow visualization). Each of these neurons k ($k = 1, \ldots, K$) is related, through an embedding function, with a d-dimensional vector $\mathbf{y}$, usually called prototype or weight vector. The weight $\mathbf{y}_j$ refers to the data sample $\mathbf{x}_j$ according to a distance $d(\cdot, \cdot)$ defined in the data space X.

The iterative algorithm initializes the weight vectors $\mathbf{y}_k$ randomly, chosen from the dataset X (or according to some basic pre-projection of the data). For each data sample $\mathbf{x}_j$ ($j = 1, \ldots, N$), it finds the best matching unit (BMU) $\mathbf{y}_{k_j}$ of index k_j, computed as $k_j = argmin_k \left\{ d(\mathbf{x}_j, \mathbf{y}_k) \right\}$, where $d(\cdot, \cdot)$ is usually defined as the Euclidean distance:

$$L_2(\mathbf{x}_j, \mathbf{y}_k) = \left\| \mathbf{x}_j - \mathbf{y}_k \right\| \tag{4.1}$$

although L_1 or L_∞ distances, for instance, can also be considered.

Let us now define a neighbourhood function $h(\cdot, \cdot)$, which can be chosen from options such as:

$$h(\mathbf{x}, \mathbf{y}_j) = e^{-\frac{d^2(\mathbf{x}, \mathbf{y}_j)}{2\sigma^2}} \quad (gaussian)$$

$$h(\mathbf{x}, \mathbf{y}_j) = \begin{cases} 0 & if\, d(\mathbf{x}, \mathbf{y}_j) > \lambda \\ 1 & if\, d(\mathbf{x}, \mathbf{y}_j) \leq \lambda \end{cases} (bubble)$$

The weight vector $\mathbf{y}_i$ is updated according the following rule:

$$y_i^{(t+1)} = y_i^{(t)} + \alpha^{(t)} h^{(t)}(\mathbf{x}_i, \mathbf{y}_c)\left(\mathbf{x}^{(t)} - \mathbf{y}_i^{(t)}\right) \tag{4.2}$$

where t is time (or iterations), $\mathbf{x}^{(t)} \in X$ is randomly selected at time t, and $0 \leq \alpha^{(t)} \leq 1$ denotes the learning rate.

4.2.2 The batch-SOM algorithm

The original version of the SOM algorithm makes a separate update of the model parameters for each data point, taken one at a time, whereas its batch version, called BSOM, makes the update on the basis of all data points.

Each data point is assigned to the region of the map to which is closest, according to the neighborhood function $h(\cdot, \cdot)$.

The update of prototypes follows the rule:

$$y_k^{(t+1)} = \sum_{j=1}^{N} \frac{h^{(t)}(\mathbf{u}_k, \mathbf{u}_{k_j})}{\sum_{j'=1}^{N} h^{(t)}(\mathbf{u}_k, \mathbf{u}_{k_{j'}})} \mathbf{x}_j \tag{4.3}$$

where $\mathbf{u}_{k_j}$ is the node corresponding to the best matching unit (BMU) for $\mathbf{x}_j$. To improve the method, the data set is partitioned in each training step according to the m Voronoi regions G_j of weight vectors $\mathbf{y}_j$, each one containing n_{V_j} samples. This update equation can be rewritten in a kernel regression form [185], for a given iteration, as:

$$\mathbf{y}_k = \sum_{k'} \left(F(\mathbf{u}_k, \mathbf{u}_{k'})\bar{\mathbf{x}}_{k'}\right) \tag{4.4}$$

where $\bar{\mathbf{x}}_{k'} = \frac{1}{n_{V_k'}} \sum_{j \in G_k'} \mathbf{x}_j$ is the mean of the group $G_{k'}$ of $n_{V_k'}$ data points assigned to a given node k', and

$$F(\mathbf{u}, \mathbf{u}_k) = \frac{N_k h(\mathbf{u}, \mathbf{u}_k)}{\sum_{k'} N_{k'} h(\mathbf{u}, \mathbf{u}_{k'})} \tag{4.5}$$

4.3 Generative Topographic Mapping

GTM [23] is a non-linear latent variable model defined as a probabilistic alternative to the heuristic SOM for clustering and visualization. Unlike SOM, GTM defines an explicit probability density model of the data that can be optimized by Maximum Likelihood using standard techniques such as the Expectation-Maximization (EM) algorithm. Its probabilistic foundations allow its expansion to tackle problems such as missing data imputation [194, 238], outlier detection [259], time series analysis [21, 195] and feature relevance determination [260, 268], amongst others. Furthermore, it has been applied in practice in several areas, such as medicine [261, 265, 266], web mining [289], speech analysis [41, 196] and business [264].

4.3.1 The GTM Standard Model

The GTM is defined as a mapping from a low dimensional latent space onto the observed data space (see *Figure 4.1*). The mapping is carried through by a set of basis functions generating a (mixture) density distribution. The functional form of this mapping is the generalized linear regression model:

$$y = \Phi(u)W \tag{4.6}$$

where y is a vector in a D-dimensional data space, Φ is a set of M basis functions $\Phi(u) = (\varphi_1(u), ..., \varphi_M(u))^1$, u is a point in latent space and W is the matrix of adaptive weights w_{md} that defines the mapping.
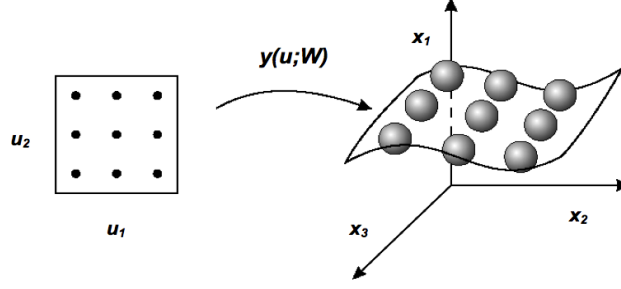
Figure 4.1: Illustration of the mapping from latent to a manifold in data space provided by the standard GTM model.

The probability distribution for data point x in a data space $X = \{x_1, ..., x_N\}$ with $x \in \Re^D$, being generated by a latent point u, is defined as an isotropic Gaussian noise distribution, assuming a single common inverse variance β:

$$
\begin{aligned}
p(x|u, W, \beta) &= N(y(u, W), \beta) \\
&= \left(\frac{\beta}{2\pi}\right)^{D/2} \exp\left\{-\frac{\beta}{2}\|x - y(u, W)\|^2\right\}
\end{aligned}
\tag{4.7}
$$

Integrating out the latent variables u in Eq. (4.7), we obtain the probability distribution in the data space, $p(x)$, expressed as a function of the parameters W and β:

$$
p(x|W, \beta) = \int p(x|u, W, \beta)\, p(u)\, du
\tag{4.8}
$$

This integral can be approximated defining $p(u)$ as

$$
p(u) = \frac{1}{K} \sum_{k=1}^{K} \delta(u - u_k)
\tag{4.9}
$$

where the K latent points u_k are sampled, forming a regular grid, from the latent space of the GTM. This way, Eq. (4.8) becomes

$$
p(x|W, \beta) = \frac{1}{K} \sum_{k=1}^{K} p(x|u_k, W, \beta)
\tag{4.10}
$$

This equation represents a constrained mixture of Gaussians [112, 285], since the centers of the mixture components cannot move independently of each other, as all depend on the mapping $y(u; W)$, and they share the same variance β^{-1} and the mixing coefficient $1/K$.

Assuming an i.i.d. data set, the likelihood of the model can be defined in the form:

$$
L(W, \beta) = \prod_{n=1}^{N} p(x_n|W, \beta) = \prod_{n=1}^{N} \left[\frac{1}{K} \sum_{k=1}^{K} p(x_n|u_k, W)\right]
\tag{4.11}
$$

Thus, the goal is maximizing L with respect to the adaptive parameters, W and β. However, maximizing the log-likelihood is equivalent and more simple. The complete log-likelihood can be defined as

$$
L_C(W, \beta|X) = \log \prod_{n=1}^{N} p(x_n) = \sum_{n=1}^{N} \log \sum_{k=1}^{K} p(x_n|u_k; W, \beta)\, p(u_k)
\tag{4.12}
$$

i.e.

---

[1] In the standard GTM model, the basis functions are defined as spherically symmetric Gaussians.

$$L_C(W, \beta | X) = \sum_{n=1}^{N} \log \left\{ \frac{1}{K} \sum_{k=1}^{K} \left(\frac{\beta}{2\pi} \right)^{D/2} \exp \left\{ -\frac{\beta}{2} \|x_n - y_k\|^2 \right\} \right\} \tag{4.13}$$

4.3.2 The Expectation-Maximization (EM) algorithm

The well-known EM algorithm can be used to obtain the Maximum Likelihood estimates of the adaptive parameters of the model. We can introduce the binary indicator variables $Z = \{z_k\}_{k=1}^{K}$, with $z_k = (z_{k1}, ..., z_{kN})$, which reflect our lack of knowledge about which mixture component k is responsible for the generation of data observation n. The incorporation of these indicators converts the complete log-likelihood into

$$L_C(W, \beta | X, Z) = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{kn} \log \left[\left(\frac{\beta}{2\pi} \right)^{D/2} \exp \left\{ -\frac{\beta}{2} \|x_n - y_k\|^2 \right\} \right] \tag{4.14}$$

The indicators Z are treated as missing data and the re-estimation of the adaptive parameters W and β, in the iterative EM procedure, requires the maximization of the expected log-likelihood $E[L_C(W, \beta | X, Z) | X, W, \beta]$. The expectation of each of the indicators in Z, also known as *responsibility* r_{kn}, can be written as

$$r_{kn} = P(k | x_n, W, \beta) = \frac{\exp \left\{ -\frac{\beta}{2} \|x_n - y_k\|^2 \right\}}{\sum_{k'=1}^{K} \exp \left\{ -\frac{\beta}{2} \|x_n - y_{k'}\|^2 \right\}} \tag{4.15}$$

In the maximization step (M-step), the parameters W and β are adaptively estimated trying to move each component of the mixture towards data points for which it is most responsible. This can be carried out by derivation of Eq. (4.13) respect to W and β. In this way, the weight matrix, W, is obtained from:

$$\Phi^T G \Phi W^T = \Phi^T R X \tag{4.16}$$

where Φ is the $K \times M$ matrix of basis function with elements $\varphi_{km} = \varphi_m(u_k)$, R is the $K \times N$ responsibility matrix with elements r_{kn} (Eq. (4.15)), X is the $N \times D$ matrix containing the data set, and G is a $K \times K$ diagonal matrix with elements:

$$g_{kk'} = \begin{cases} \sum_{n=1}^{N} r_{kn} & k = k' \\ 0 & k \neq k' \end{cases}$$

The parameter β is updated through the expression:

$$\beta^{-1} = \frac{1}{ND} \sum_{n=1}^{N} \sum_{k=1}^{K} r_{kn} \|y(u_k, \tilde{W}) - x_n\|^2 \tag{4.17}$$

where $\tilde{W}$ corresponds to the updated weights.

The initial values of the parameter W can be chosen using a standard PCA-based procedure for its initialization. β can be initialized so that its inverse equals the largest of either the length of the $(L+1)^{th}$ principal component or the half of the average minimal distance between the mixture components. Details can be found in [241].

4.3.3 Data visualization and clustering through GTM

Gaussian Mixture Models (GMM) -a specific case of Finite Mixtures of Distributions- are a *soft* clustering method where each cluster is described in terms of Gaussian density, with has a *centroid*[2] or *prototype* and a covariance matrix. Since GTM is a constrained Gaussian Mixture Model, it can be used for clustering with minimal modifications.

---

[2] As in k-means, for instance.

The non-linear function $y(u; W)$ of GTM defines a manifold embedded in data space given by the image of the latent space under the mapping u → y. In order to use GTM for visualization, it is necessary to assess the relation between each data point and each point of the latent space. In GTM, this relation is explicitly estimated as a responsibility r_{kn}. The *responsibilities* can be seen as a *soft* counterpart to the *winner-takes-all* SOM approach. Thus, if each of the latent space points u_k is considered by itself as a representative cluster, the cluster assignment method is akin to that of SOM, which is based on a winner-takes-all strategy: each data observation (in the data analyzed in this thesis, usually a client in the available database) is assigned to the location in the latent space (cluster) where the mode of the corresponding posterior distribution is highest. i.e. adapting Eq. (4.18)

$$u_n^{mode} = \arg_{u_k} \max r_{kn} \tag{4.18}$$

where r_{kn} can be understood as the probability of client n belonging to cluster (*micro-segment*) k, and it is obtained as part of the EM algorithm. This quantity, also known as *posterior mode*, can be used for visualization purposes.

An alternative method for data visualization is, for each data point x_n, to plot the mean of the posterior distribution in latent space, also known as *posterior-mean* projection

$$u_n^{mean} = \sum_{k=1}^{K} r_{kn} u_k \tag{4.19}$$

The distribution of the *responsibility* over the latent space of states can also be directly visualized.

4.3.4 Magnification Factors for the GTM

One of the most interesting consequences of the probabilistic definition of GTM is that the distortion caused by the nonlinear mapping can be explicitly quantified. Not only that: despite the fact that GTM, as SOM, is a discrete *projection* technique [9] in the sense that only a finite number of latent space points are considered, this distortion, known as Magnification Factors (MF) [22], can be calculated for any point in the latent space continuum.

As remarked in [241], the concept of MF has its origin in the field of computational neuroscience, where it refers to the mapping distortion between the spatial density of biological sensors and the two-dimensional spatial density of the corresponding topographic maps in the visual and somatosensory areas of the cortex. More specifically, the cortical MF would indicate the linear distance along the primary visual cortex concerned with each degree of visual field [211], although controversy remains on whether the cortical magnification of the central visual field reflects its selective amplification, or merely reflects the ganglion cell density of the retina [281]. As expressed in the context of vector quantization models [106], local magnification is the result of a specific connection of the density of model prototypes and stimuli (data).

For GTM, it is shown in [22] that the relationship between a differential area dA (for a 2-D representation) in latent space and the corresponding area element in the generated manifold, dA', can be expressed as $dA = J dA'$, where J is the Jacobian of the mapping transformation. This Jacobian can be written in terms of the derivatives of the basis functions ϕ_m as:

$$dA/dA' = J = \det^{\frac{1}{2}}(\Psi^T W^T W \Psi), \tag{4.20}$$

where Ψ is a $M \times 2$ matrix with elements $\varphi_{mi} = \partial \phi_m / \partial u^i$ and u^i is the ith coordinate ($i = 1, 2$) of a latent point. Note that the MF as expressed by the Jacobian in *equation* (4.20) can be calculated for any value of **u** over the continuum.

From a practical viewpoint, the MFs for GTM were introduced as the geometrical functional equivalent to the distance matrix or U-matrix of the SOM [252]. These factors can provide useful information, such as areas of stretch in the manifold that separate different regions in the data space.

The MFs "add a dimension" to the visual representation of data by providing hints about their global cluster structure. It is easy to see why this should be so if we consider the Gaussian mixture model on the

GTM manifold. The EM algorithm will attempt to place the mixture components in regions of high data density and will move the components away from the regions with low data density. It can do this because the non-linear mapping from latent space to data space enables the manifold to stretch across regions of low data density. This stretching (or magnification) can be measured using techniques of differential geometry and plotting the MFs in latent space may allow the user to discover separation between clusters, if this exists.

Chapter 5

Supervised customer loyalty analysis

In this chapter we describe, following a supervised ML approach, the drivers towards customer satisfaction on the basis of a survey conducted amongst the customers of several Spanish petrol station brands. Such description is carried out through reasonably simple and actionable rules that could be applied in a real business environment. With this, we aim to achieve the necessary level of interpretability of the solutions that is often required from the application of ML methods [271].

A survey of several thousand customers was used to classify them according to satisfaction levels, using an artificial neural network (ANN) defined within a Bayesian framework [173]. An Automatic Relevance Determination (ARD) technique embedded in this model was used for supervised feature relevance quantification, leading to feature selection. The subset of selected features was used, in turn, to obtain a rule description of the classification performed by the ANN through the recently developed OSRE method [73]. OSRE was able to describe the factors driving customer satisfaction in a reasonably simple and interpretable manner that could be swiftly integrated in service management processes.

This brief chapter is structured as follows: the case study including the available data is briefly described in *Section 5.1*. This is followed by a summary technical description of the Bayesian ANN with ARD and the OSRE techniques in *Section 5.2*. Finally, the developed experiments and the obtained results are reported in *Section 5.3*, while some conclusions are drawn in *Section 5.4*.

Results of this study were presented at the 7th Intelligent Data Engineering and Automated Learning International Conference (IDEAL 2006) [267]. This work provided, as envisaged, a first preliminary approximation to the prevention side of the customer retention *vs.* churn problem.

5.1   Petrol station customer satisfaction, loyalty and switching barriers

As detailed in *Chapter 2*, efficient churn management requires a model of both preventive and treatment actions: preventing dissatisfaction before it occurs and treating it when it has already set in. In this chapter we focus on the prevention side of customer loyalty management and, in particular, on customer satisfaction (see *Figure 5.1*). Satisfaction with the received service is likely to act as an antecedent to loyalty, consolidating customer permanence and avoiding substitution by a competitor. It might be a necessary condition for loyalty, but perhaps not sufficient. Therefore, the development by the company of active barriers should also be explored as an antecedent to customer loyalty.

The data analysed in this chapter correspond to a survey carried out among customers of Spanish petrol station main brands. This is a different and smaller data set as compared to the ones investigated in the following chapters. A total of 350 service stations of the Spanish market, sampled by location (urban *vs.* road) and type of service (with attendant *vs.* self-service), were selected for the exercise. The survey questionnaire was answered by over 3,500 clients during the last quarter of 2005.
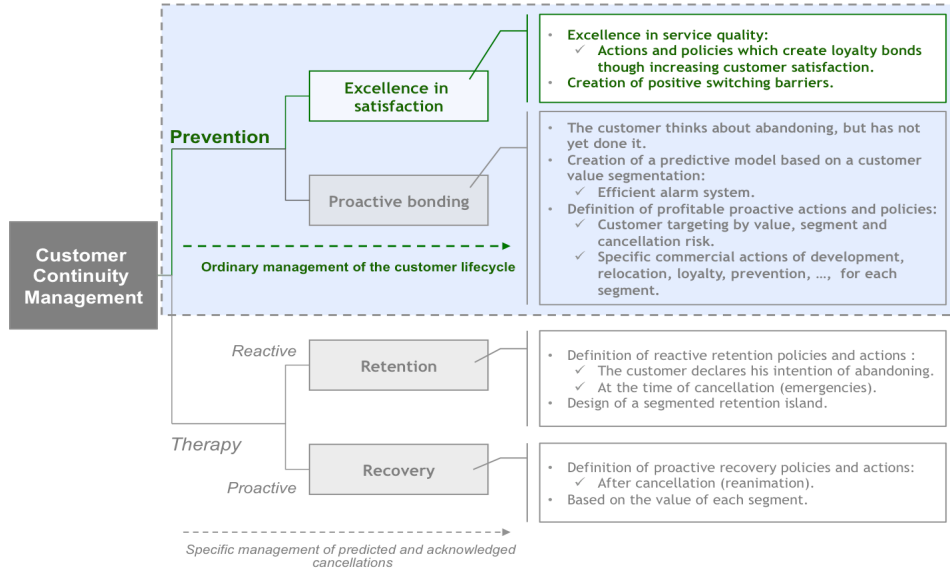
Figure 5.1: Conceptual model of customer loyalty management. Customer satisfaction in prevention side of Customer Continuity Management model (CCM).

The classification and rule extraction analyses described in the next section considered one binary dependent variable: *overall satisfaction* with two possible answers: *satisfied / dissatisfied*.

Overall satisfaction would measure the customer satisfaction construct. The questionnaire included 20 variables, listed in *Table 5.1*, measured in a Likert scale (values range from 1: very good to 5: very bad; value 6 means not answered (NA)). They fit into two qualitative categories: attributes of satisfaction with service and switching barriers.

5.2 Methods

In this section, we provide a summary description of the Bayesian approach to the training of an ANN with ARD embedded for feature selection, as well as of OSRE to obtain a rule description of the classification performed by ANN.

5.2.1 Bayesian ANN with ARD

The Bayesian approach to the training of a multi-layer perception ANN differs from the standard Maximum Likelihood approach in that it does not simply attempt to find a single optimal set of weights; instead, it considers a probability distribution over the weights, which reflects the uncertainty resulting from the use of finite data sets. In that way, the outputs of the ANN in a classification problem can be interpreted as posterior probabilities of class membership given the data and the weights,

$$P(C_i|x,w) = y(x;w) \tag{5.1}$$

where y is the network function, x is an input vector, w is the vector of weights and biases, and C_i is class i. The probability of class membership for a test vector can be obtained by integrating Eq. (5.1) over the weights:

$$P(C_i|x,D) = \int y(x;w)p(w|D)\mathrm{d}w \tag{5.2}$$

where D are the target data for the training set. This conditional probability can be adequately approximated [173].

ANNs are frequently considered as *black boxes* due to, amongst other things, their supposed incapacity to identify the relevance of independent data variables in nonlinear terms. ARD [174] for supervised Bayesian ANNs, is a technique that addresses that shortcoming: in ARD, the weight decay or regularization term can be interpreted as a Gaussian prior over the network parameters of the form $p(w) = A \exp(-\alpha_c \sum_c^C n_{w(c)} \sum_i^{N_w} w_i^2/2)$ where $w = \{w_i\}$ is the vector of network weights and biases, so that individual regularization terms with coefficients α_c are associated with each group c of fan-out weights from each input to the hidden layer (i.e. with each input variable). Therefore $C = $ (number of ANN inputs + 2), $n_{w(c)}$ is the number of parameters in group c so that $\sum_c^C n_{w(c)} = N_w$, where N_w is the total number of network parameters.

The adaptive hyperparameters α_c associated with irrelevant variables will be inferred to be large, and the corresponding weights will become small through training. Therefore, ARD performs soft feature selection of sorts, and, as a result, a direct inspection of the final $\{\alpha_c\}$ values provides and indication of the relative relevance of each variable. ARD has shown to be a useful feature selection method for classification problems [207, 262].

Independent variables	Conceptually linked to:		
	Satisfaction	**Switching barriers**	**Loyalty**
1. Personal attention form staff	+		+
2. Speed and efficiency of staff	+		+
3. Additional services	+		+
4. Ease of access to installations, well indicated	+		+
5. Signs inside installations	+		+
6. Modern and attractive installations	+		+
7. Hygiene and maintenance of the installations	+		+
8. Basic services well-maintained and always in working order	+		+
9. Extended opening hours	+		+
10. Cleanliness of toilets	+		+
11. Exact and reliable pumps	+		+
12. Feeling of security and absence of risk	+		+
13. Top quality fuel	+		+
14. Attractive and stocked shop	+		+
15. Price		+	+
16. Payment cards with discounts		+	+
17. Cards to collect points for gifts		+	+
18. Brand which inspires trust		+	+
19. Brand with an extensive network of service stations		+	+
20. Environmental awareness		+	+

Table 5.1: Description of the 20 variables used in this study and their adscription to the marketing constructs of *satisfaction, switching barriers* and *loyalty*.

5.2.2 Orthogonal Search-based Rule Extraction

OSRE [73] is an algorithm that efficiently extracts comprehensible rules from smooth models, such as those created by the Bayesian ANN in this study. This is a principled approach underpinned by a theoretical

framework of continuous valued logic.

In essence, the algorithm extracts rules by taking each data item, which the model predicts to be in particular class, and searching in the direction of each variable to find the limits of the space regions for which the model prediction is in that class. These regions form hyper-boxes that capture in-class data and they are converted to conjunctive rules n terms of the variables and their values.

The obtained set of rules is subjected to a number of refinement steps: removing repetitions; filtering rules of poor specificity and sensitivity; and removing rules that are subsets of other rules. Specificity is defined as one minus the ratio of the number of out-of-class data records that the rule identifies to the total number of out-of-class data. Sensitivity is the ratio of the number of in-class data that the rule identifies to the total number of in-class data. The rules are then ranked in terms of their sensitivity values to form a hierarchy describing the in-class data. Testing against benchmark datasets [74, 75] has showed OSRE to be an accurate and efficient rule extraction algorithm. Details of the algorithm are beyond the goals of the thesis and the interested reader can find them in [73].

5.3   Experiments

As mentioned in the introduction to the chapter, we aim to describe, through reasonably simple and actionable rules, the drivers towards customer satisfaction on the basis of a survey conducted amongst the customers of several Spanish petrol station brands. The method underlying the experiments can be summarily described as follows:

- The survey data described in *Section 5.1* were used to predict customer satisfaction with the service. A Bayesian ANN [173] with embedded ARD [174] was used to classify the data.

- One of the potential drawbacks affecting the application of ANN models to classification problems is that of the limited interpretability of the results they yield. One way to overcome this limitation is by pairing the ANN model with a rule extraction method. The ARD technique allowed ranking the variables according to their relevance in the classification task. This naturally leads to a process of selection on the basis of this ranking.

- The variables selected in the previous step of the method were used to describe the classification performed by the ANN through rules, using the OSRE technique [73]. Thus, the interpretability of the classification results could be greatly improved by their description in terms of reasonably simple and actionable rules.

5.3.1   Results and discussion

Automatic Relevance Determination and Feature Selection

The application of the ARD described in *Section 5.2.1* for the classification of *overall satisfaction* using the 20 variables of *Table 5.1*, yielded sensible and interesting results: As shown in *Figure 5.2*, two variables turn out to be the most relevant for the classification of *overall satisfaction*, namely numbers 1 (Personal attention form staff) and 7 (Hygiene and maintenance of the installations), followed by a subset of variables with similar relevance, namely numbers 3 (Additional services), 5 (Signs inside installations), 6 (Modern and attractive installations), 14 (Attractive and stocked shop) and 16 (Payment cards with discounts). The relevance of the rest of variables falls clearly behind.

Most, if not all, of these variables are easily actionable (easy to act upon) in terms of service management, which comes to validate the practical of usefulness of the ARD process. It is worth noting that all but one (number 16) of the most relevant variables were *a priori* considered as elements of the satisfaction construct (see *Table 5.1*), proving that the recurrent use of switching barriers by all competitors turns rapidly into a general attribute of the supply.

Overall satisfaction
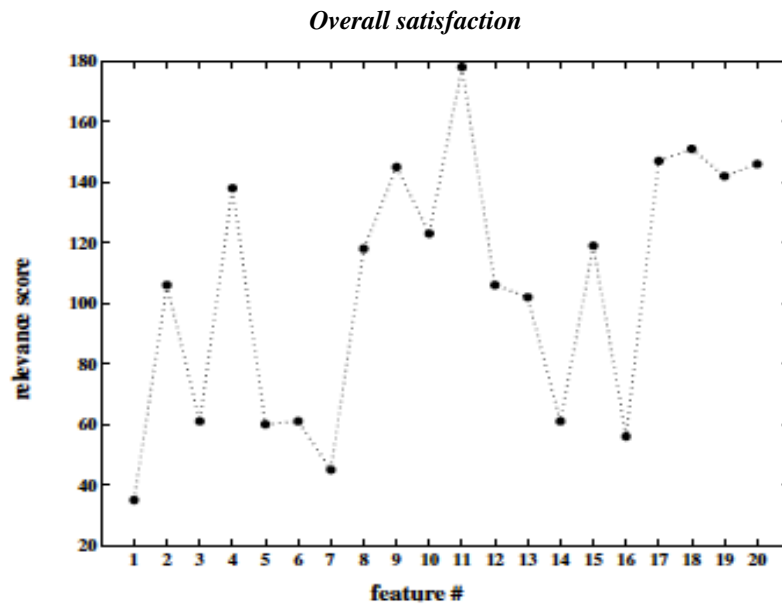
Figure 5.2: Cumulative ranking of relevance calculated according to the ranking of variables on each of the 10 ANN runs. The lower the score, the more relevant the variable (A value of 10 would indicate that the variable was ranked 1st in all ANN runs, whereas the value of 200 would mean that the variable has been ranked as the least relevant in all runs).

Rule Extraction Using OSRE

The available survey data described in *Section 5.1* are strongly unbalanced in terms of class prevalence with only small percentages of customers declaring themselves unsatisfied. This makes rule extraction a challenging task. The Bayesian ANN training was adapted to compensate for this unbalance, using a strategy described in [167] that entailed modifying the log-likelihood and the network output.

OSRE was first tested using 20 variables. For *customer satisfaction* (see *Table 5.2*), whilst the overall specificity was poor, the specificity of each individual extracted rule was quite good; this means that each rule identified different groups of customers who are satisfied for different reasons. The overall coverage of the extracted rules was rather weak, in the area of 71% and concerned 14 out of 20 variables, a far too complex description for marketing purposes. Each rule is a conjunction of the variables and their possible range of values.

Therefore, we decided to use directly the selection of variables obtained by ARD in order to extract the rules. For *overall satisfaction* this meant selecting variables: 1.- *Personal attention from staff*; 3.- *Additional services*; 5.- *Signs inside installations*; 6.- *Modern and attractive installations*; 7.- *Hygiene and maintenance of the installations*; 14.- *Attractive and stocked shop* and 16.- *Payment cards with discounts*. Rules were extracted for two classes: *satisfied / dissatisfied* (see *Table 5.3*). Only *satisfied* is shown.

Hence (rule number 1), when a customer declares an excellent opinion about "personal attention from staff" (v1) and the "hygiene and maintenance of the installations" (v7), there's a high probability for him to be satisfied with the service station (PPV: 85%) and to build loyalty bonds with it. Likewise (rule number 2), when a customer has an excellent opinion about "personal attention from staff" (v1), but not an excellent opinion (2: good) about "hygiene and maintenance of the installations" (v7), he will also have a high probability of building loyalty bonds (PPV: 79%) if the station has an "attractive and stoked" shop (v14).

	CLASS: *satisfied*	**For this rule only**			**For all rules up to row n**		
n	**RULE**	*Spec*	*Sens*	*PPV*	*Spec*	*Sens*	*PPV*
1	$v1 = 1 \land v2 = \{1,2,3\} \land v3 = \{1,2\} \land v7 = 1 \land$ $v9 = \{1,2\} \land v14 = \{1,2\} \land v15 = \{1,2\}$	0.92	0.21	0.81	0.92	0.21	0.81
2	$v1 = 1 \land v2 = \{1,2\} \land v5 = 2 \land v6 = \{2,3,4,5\} \land$ $v9 = \{1,2\} \land v12 = \{1,2\} \land v14 = \{1,2,3,4\}$	0.93	0.16	0.77	0.85	0.34	0.79
3	$v12 = \{1,2,3\} \land v13 = \{1,2\} \land v14 = \{1,2\} \land$ $v15 = \{1,2,3,4\} \land v17 = 1 \land v18 = \{1,2,3\}$	0.92	0.11	0.69	0.79	0.42	0.76
4	$v1 = \{1,2\} \land v2 = \{1,2,3\} \land v3 = \{1,2,3\} \land v5 =$ $\{1,2,3\} \land v7 = 1 \land v15 = \{1,2\}$	0.92	0.18	0.79	0.74	0.49	0.75
5	$v1 = 1 \land v4 = \{2,3,4,5\} \land v5 = \{2,3\} \land v9 =$ $\{1,2\} \land v15 = \{1,2\}$	0.93	0.14	0.77	0.71	0.54	0.75
6	$v1 = \{1,2\} \land v2 = 1 \land v3 = \{1,2,3,4\} \land v6 =$ $\{2,3,4,5\} \land v9 = \{1,2\} \land v15 = \{1,2\}$	0.92	0.14	0.73	0.68	0.58	0.74
7	$v1 = \{1,2,3\} \land v2 = \{1,2,3\} \land v3 = \{1,2,3\} \land$ $v5 = \{2,3\} \land v7 = \{1,2\} \land v9 = 1 \land v14 = \{1,2\}$	0.93	0.11	0.71	0.64	0.62	0.73
8	$v1 = \{1,2\} \land v2 = \{1,2,3\} \land v5 = 1 \land v7 =$ $\{1,2\} \land v9 = \{1,2\} \land v12 = \{1,2\} \land v14 = \{1,2\}$	0.94	0.10	0.73	0.62	0.64	0.73
9	$v1 = 1 \land v2 = \{1,2\} \land v3 = \{1,2,3,4,5\} \land v4 =$ $\{1,2,3,4\} \land v5 = \{1,2,3\} \land v7 = \{1,2,3,4\} \land$ $v9 = 1$	0.93	0.18	0.80	0.61	0.66	0.73
10	$v1 = \{1,2\} \land v2 = \{1,2\} \land v4 = \{1,2,3,4,5\} \land$ $v7 = \{1,2,3,4\} \land v9 = \{1,2\} \land v13 = \{1,2\} \land$ $v14 = \{1,2\} \land v15 = \{1,2\} \land v17 = \{1,2,3,4,5\} \land$ $v18 = \{1,2,3,4\}$	0.92	0.18	0.79	0.60	0.67	0.73
11	$v1 = 1 \land v2 = \{1,2\} \land v5 = \{1,2,3\} \land v7 =$ $\{1,2,3\} \land v9 = 1 \land v12 = \{1,2\} \land v15 = \{1,2\}$	0.92	0.20	0.81	0.59	0.68	0.72
12	$v1 = 1 \land v2 = 1 \land v5 = \{1,2,3\} \land v9 = \{1,2\} \land$ $v14 = \{1,2\} \land v15 = \{1,2,3\}$	0.92	0.15	0.76	0.58	0.69	0.72
13	$v1 = \{1,2\} \land v3 = \{1,2,3\} \land v5 = \{1,2,3\} \land v6 =$ $\{1,2,3\} \land v7 = 1 \land v9 = \{1,2\}$	0.92	0.11	0.70	0.56	0.70	0.71

Table 5.2: OSRE rules for *overall satisfaction* (Class: *satisfied*) and the 20 variables listed in *Table 5.1*. *Spec* stands for Specificity; *Sens* for Sensitivity; *PPV* is the Positive Predictive Value: the ratio of the number of in-class data that the rule predicts to the total number of data the rule predicts. The expression *vn* stands for variable *n*, following the numbering in *Table 5.1*. The possible variable values, in a Likert scale, range from 1: very good to 5: very bad. Value 6: NA.

	CLASS: *satisfied*	For this rule only			For all rules up to row n		
n	**RULE**	*Spec*	*Sens*	*PPV*	*Spec*	*Sens*	*PPV*
1	$v1 = 1 \wedge v7 = 1$	0.95	0.18	0.85	0.95	0.18	0.85
2	$v1 = 1 \wedge v7 = 2 \wedge v14 = \{1,2\}$	0.94	0.14	0.79	0.89	0.29	0.81
3	$v1 = 1 \wedge v6 = \{2,3\} \wedge v14 = 2$	0.95	0.14	0.81	0.85	0.37	0.80
4	$v1 = \{1,2\} \wedge v7 = \{1,2\} \wedge v14 = \{1,2\}$	0.95	0.11	0.75	0.75	0.53	0.77
5	$v1 = 1 \wedge v3 = 1 \wedge v7 = \{1,2\} \wedge v14 = \{1,2,3\}$	0.94	0.18	0.82	0.73	0.56	0.77
6	$v1 = 1 \wedge v14 = \{1,2\}$	0.95	0.16	0.83	0.71	0.58	0.76
7	$v1 = 1 \wedge v7 = \{1,2\} \wedge v14 = \{1,2,3\}$	0.95	0.15	0.82	0.69	0.61	0.76
8	$v1 = \{1,2\} \wedge v3 = \{1,2,3\} \wedge v7 = \{1,2\} \wedge v14 = \{1,2,3\}$	0.95	0.16	0.83	0.68	0.62	0.76
9	$v1 = \{1,2\} \wedge v6 = \{2,3\} \wedge v14 = \{1,2\}$	0.96	0.11	0.80	0.67	0.63	0.75

Table 5.3: OSRE rules for *overall satisfaction* (Class: *satisfied*) and the 7 variables selected by ARD. *Spec* stands for Specificity; *Sens* for Sensitivity; *PPV* is the Positive Predictive Value: the ratio of the number of in-class data that the rule predicts to the total number of data the rule predicts. The expression *vn* stands for variable *n*. The possible variable values, in a Likert scale, range from 1: very good to 5: very bad. Value 6: NA.

The coverage improved to 75%, which indicates that many of the removed variables did not add to the classification, but actually interferred with it. It is worth highlighting that all rules include variable 1 (*Personal attention from staff*), validating its relevance as suggested by ARD. Interestingly, the replacement of staff by self-service was a controversial cost-cutting strategy adopted in recent times by several petrol station brand in Spain. This variable suggests its potential as antecedent of behavioural intentions of petrol station customers.

5.4   Conclusions

This chapter provides a preliminary study of customer satisfaction and loyalty, as key elements of the churn problem, from a supervised perspective. The experiments concerned data from petrol station usage surveys.

The performed analyses focus on classification mostly from the point of view of the achievement of interpretability. This interpretability of the results is paramount for actionable marketing. Feature relevance determination for feature selection and rule extraction were the tools used for achieving such required interpretability. Hence, it's noted how the application of ARD enables the selection of 7 features (v1.- "Personal attention from staf"; v7.- "Hygiene and maintenance of the installations"; v3.- "Additional services", v5.- "Signs inside installations"; v6 "Modern and attractive installations"; v14.- "Attractive and stocked shop" and v16.- "Payment cards with discounts") as those which are more relevant for the classification of overall satisfaction. In this regard, it must be emphasized the non appearance of two feature groups:

1. v15.- "Price", v11.- "Exact and reliable pumps" and v13.- "Top quality fuel". Their non appearance indicates a mature and regulated market, where old procedures of fraud and / or abuse have been surpassed. Currently, petrol station users in Spain have internalized the quality of the replenishment service at competitive prices.

2. v2.- "Speed and efficiency of staff", v4.- "Ease access to installations, well indicated" and v9.- "Extended opening hours". For its part, the non appearance of these features as relevant is highly related with the profile of costumers that used the petrol stations of the analyzed brand. Thus, it was noted that customers were mainly "seniors", expecting an attended service, additional services and personal attention from staff.

The subsequent rule extraction using OSRE enabled to denote that only five of the previously selected features were relevant. Thus:

- v1.- "Personal attention from staff" appeared in all of the 9 obtained rules.

- v14.- "Attractive and stocker shop" appeared in 8 of the 9 obtained rules.

- v7.- "Hygiene and maintenance of the installations" appeared in 6 of the 9 obtained rules.

- v6.- "Modern and attractive installations" appeared in 3 of the 9 obtained rules.

- v3.- "Additional service" appeared in 2 of the 9 obtained rules.

The absence of features v5.- "Signs inside installations" and v16.- "Payment cards with discounts" should be understood, from an interpretability point of view, as the existence of remotely significant differences regarding to these features between the different petrol station competitors.

The obtained results were consistent with recent theory on satisfaction, loyalty and switching barriers models. The attributes and rules obtained in the described experiment enabled the company to define a decalogue of actions from which, currently, they evaluate and reward the performance of the petrol stations (owned or managed) members of the company's network.

Chapter 6

Unsupervised churn analysis in a telecommunications company

One of the major challenges faced today by telecommunications service providers is how to retain their customer base. Immersed in an extremely competitive market, they must engage in strategies to limit customer defection to competitors -a phenomenon also known as churn-. Anticipating the customer's intention to abandon facilitates the launching of retention-focused actions and it represents a clear element of competitive advantage. As we introduced in *Chapter 3*, Data Mining techniques can assist churn (customer attrition) management processes and may provide clues to explain and anticipate churn [104]; and one analytical tool to this purpose is data clustering for market segmentation.

Thus, whereas in *Chapter 5* we focus on customer satisfaction as a key point for churn prevention (see *Figure 5.1*), in the present chapter we focus on proactive bonding (see *Figure 6.1*). In particular, we propose an indirect and explanatory approach to the prediction of customer abandonment, based on the visualization of customer data -consisting of their consumption patterns- on a two-dimensional representation map, to explore the existence of abandonment routes in the Brazilian telecommunications market.
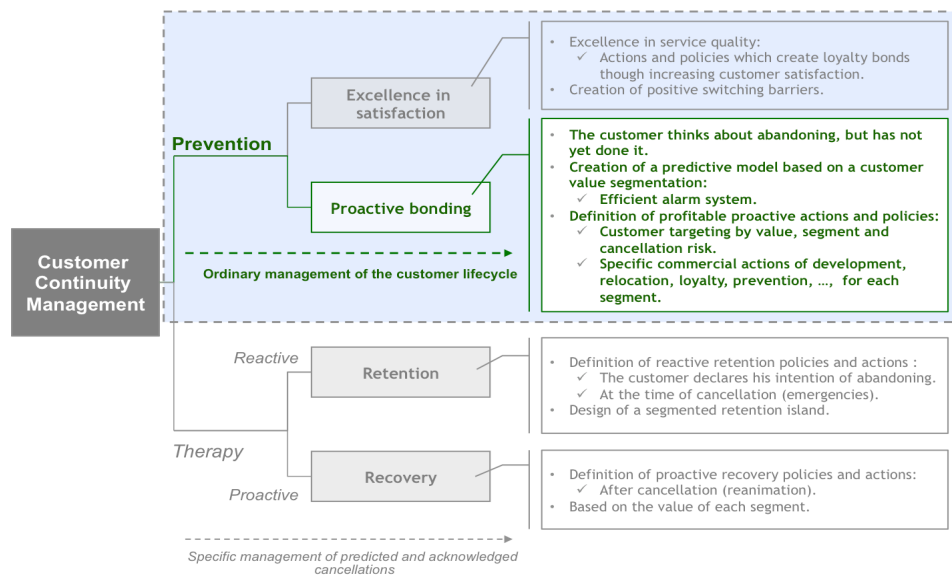

Figure 6.1: Conceptual model of Customer Continuity Management [86].

Our approach is based on two basic hypotheses:

- Different patterns of service consumption, regarding the type of communications established, correspond to different levels of predisposition to abandon;

- Different migration routes between time periods are likely to exist and be identifiable in the representation map, both negative: towards lower customer value and, eventually, service abandonment; and positive: towards higher customer value areas.

Two probabilistic neural network-inspired models of the manifold learning family (GTM and FRD-GTM) are used for the simultaneous visualization and clustering of multivariate data corresponding to customers of a principal Brazilian telecommunications company. These models allow the estimation of the relative relevance of each data feature on the definition of the obtained cluster structure and, in doing so; it eases the interpretability of the segmentation results. From these results, typical customer churn routes are investigated. Several indices of cluster validity for this model are also defined.

Thus, in the present chapter we first introduce the marketing problem and the data features used in the experiments (*Section 6.1* and *Section 6.2*). This is followed by a summary description of the theory behind our approach (*Section 6.3*). Finally, we describe the developed experiments and the obtained results (*Section 6.4*) and the conclusions of this chapter (*Section 6.5*).

Results of this research were presented at the 15th European Symposium on Artificial Neural Networks (ESANN 2007) [85] and at the 2nd Symposium on Computational Intelligence (IEEE SICO 2007) [88].

6.1   Problem Description

Identifying a customer's intention to abandon their current service provider with sufficient anticipation has become one of the main focal points of marketing studies in recent years [104, 122]. The main trends in predicting customer behaviour and, in particular, customer abandonment (churn) are based on the direct identification of the possible churner through the use of historical variables relating to customer behaviour. One way to go about this is to explore the existence of customer groups or segments. Market segmentation is a time-honored goal in marketing studies, commonly accomplished through cluster analysis based on statistical or machine learning methods.

A wide variety of clustering techniques have been applied to market segmentation [104]. Given the behavioral origin of market information, quantitative data are likely to involve uncertainty at different levels, which may be better served by probabilistic clustering methods. The real-world context in which clustering has to be applied here to the segmentation of the Brazilian telecommunications customers makes the interpretability of the results an important requirement. That is the reason GTM and FRD-GTM are used rather than general mixtures. As seen in *Chapter 4*, the constrained definition of GTM endows it with visualization capabilities that are similar, if not superior, to those of Self-Organizing Maps [143, 277] neural networks, and visualization is a main key to interpretability.

Another key to interpretability can be provided by unsupervised feature selection, in the form of an objective method to rank the data covariates by their relative relevance for cluster structure. This is unsupervised relevance determination, a problem that has received scant attention in the past, and on which research is starting to make some inroads. An important advance in feature selection for unsupervised Gaussian mixture models was proposed in [152] and extended to the GTM in [258, 268]. This extension, termed FRD-GTM, was assessed in some detail in [260]. It includes the definition of an unsupervised feature saliency as a measure of relevance, which is estimated using the EM algorithm.

In this chapter, a Brazilian telecommunications market segmentation results are analyzed on the low dimensional visualization space provided by both GTM (*Experiment 1*) and FRD-GTM (*Experiment 2*).

In a trade-off between the visualization and clustering capabilities of GTM, the number of clusters is usually chosen to be large enough to provide a useful data visualization. In terms of market segmentation, though, a solution providing too large a number of cluster/segments makes business actionability extremely

difficult. For that reason, we adopt a two-tier clustering strategy, similar to that proposed in [277], that involves using k-means on top of the GTM and FRD-GTM clustering results to obtain market segments. To that purpose, new variations on established cluster validity indices are defined.

Brazilian customers' routes across the segmentation map are explored, with a focus on departure gates of customer churn. Changes in customer value across periods are also investigated. The identification of customer segment *migration routes* is a novel approach in the field of churn prediction, and could be used to assist churn prediction tasks.

6.2 Telecommunications customer data

For the experiments reported in the next section, a proprietary database containing telephone usage information corresponding to a total of 60,596 small and medium-size Brazilian companies, all of them customers of the main landline telephony telecommunications company in São Paulo (Brazil), was used. The information was measured over two consecutive periods (non-overlapping with holidays): Period 1 (*P*1), from June to December 2003, and Period 2 (*P*2), from March to August, 2004.

The following 14 data features (see *Table 6.1*), which characterize landline usage, were considered for analysis: *v1*.- Percentage of local landline outcoming calls; *v2*.- Percentage of outcoming state landline calls (Brazil is formed by 26 states, each with different telephone tariffs according to call destination); *v3*.-Percentage of outcoming out-of-state landline calls; *v4*.- Percentage of outcoming international landline calls; *v5*.- Percentage of outcoming calls to mobile phones; *v6*.- Percentage of incoming local landline reverse-charge calls; *v7*.- Percentage of incoming state reverse-charge landline calls; *v8*.- Percentage of incoming out-of-state reverse-charge landline calls; *v9*.- Percentage of incoming mobile phone reverse-charge calls; *v10*.- Percentage of calls within standard time slot (8:00-10:00h and 14:00-16:00h); *v11*.- Percentage of calls in differential time slot (10:00-14:00h and 16:00-18:00h); *v12*.- Percentage of calls within mixed time slot (calls that begin and end in different time slots); *v13*.- Percentage of calls within reduced-tariff time slot (18:00-24:00h); *v14*.- Percentage of calls within super reduced-tariff time slot (00:00-06:00h).

p‿ll‿nor	Percentage of normal timetable calls
p‿ll‿dif	Percentage of differentiated calls
p‿ll‿mis	Percentage of mixed calls
p‿ll‿red	Percentage of reduced timetable calls
p‿ll‿sup	Percentage of super reduced timetable calls
p‿ll‿ln	Percentage of local calls or Internet
p‿ll‿ita	Percentage of intrastate calls
p‿ll‿ite	Percentage of interstate calls
p‿ll‿int	Percentage of international calls
p‿ll‿mov	Percentage of calls to mobile
p‿lc‿loc	Percentage of reversed charge local calls
p‿lc‿ita	Percentage of reversed charge intrastate calls
p‿lc‿ite	Percentage of reversed charge interstate calls
p‿lc‿mov	Percentage of reversed charge calls to mobile

Table 6.1: Data features used to describe the consumption of telecom company's customers.

In addition, for profiling the segmentation results, the following information was taken into account: commercial margin, value-added services (VASs) on portfolio, length of time as client, EANC code (Economic Activities National Classification) and number of employees in the company.

6.3 Methods

6.3.1 The FRD-GTM

The interpretability of the clustering results provided by the GTM and even their visualization can be limited for data sets of high dimensionality. Dimensionality can be reduced by methods of feature selection (FS) with or without previous feature relevance determination (FRD). Recently, a method for feature selection in unsupervised model-based clustering with Gaussian mixture models was proposed in [152]. It was extended to GTM (FRD-GTM) in [258] and this extension was assessed in [260]. The FRD method estimates an unsupervised saliency as part of the EM algorithm. Such saliency measures the importance of each feature on the model-defined cluster structure. Formally, the saliency of feature d can be defined as $\rho_d = P(\eta_d = 1)$, where $\eta = (\eta_1, \ldots, \eta_D)$ is a set of binary indicators that are considered missing information by the EM algorithm. A value of $\eta_d = 1(\rho_d = 1)$ corresponds to he maximum relevance attributable to feature d. According to this, the following mixture density for FRD-GTM is defined:

$$p(x|W, \beta, w_0, \bar{\beta}_0, \bar{\rho}) = \sum_{k=1}^{K} \frac{1}{K} \prod_{d=1}^{D} \{\rho_d p(x_d|u_k, w_d, \beta) + (1 - \rho_d) q(x_d|u_0, w_{0,d}, \beta_{0,d})\} \tag{6.1}$$

where w_d is the vector of W corresponding to feature d and $\bar{\rho} = (\rho_1, \ldots, \rho_D)$. The distribution p (or p_{kd}) is a feature -and a component-specific version of *Eq.* (4.7). A feature d will be considered irrelevant if $p(x_d|u_k, w_d, \beta) = q(x_d|u_0, w_{0,d}, \beta_{0,d})$ for all the mixture components k where q (or q_d) is a common density followed by feature d. This is the same as saying that feature d does not contribute to the cluster structure defined by the model. The common component accounts for data points that the GTM constrained mixture components cannot explain well through the model-defined cluster structure. It requires two extra adaptive parameters: $w_0 = (w_{0,1}, \ldots, w_{0,D})$, $\bar{\beta}_0 = (\beta_{0,1}, \ldots, \beta_{0,D})$ and it should reflect any prior knowledge we might have regarding irrelevant features. Maximum Likelihood through EM can again be used to estimate the model parameters, out of which we are especially interested in the estimations of the feature saliencies as:

$$\rho_d = \frac{1}{N} \sum_{n,k} r_{kn} \frac{a_{knd}}{a_{knd} + b_{knd}} \tag{6.2}$$

where $a_{knd} = \rho_d p_{kd}(x_n)$; $b_{knd} = (1 - \rho_d) q_{kd}(x_n)$. Further details of FRD-GTM calculations can be found in [258].

As previously mentioned, each latent space point of the FRD-GTM can be considered by itself as a cluster representative (of a cluster containing the subset of data points assigned to it). The data can be visualized in FRD-GTM 2-dimensional data space using their posterior-mean calculated as $u_n^{mean} = \sum_{k=1}^{K} r_{kn} u_k$. For simplicity, we can also use a simplified cluster assignment method akin to that of SOM, based on a winner-takes-all strategy: each data point x_n (in the case of this study, each telecom client) is assigned to the location in the latent space (to the cluster) where the mode of the corresponding posterior distribution or responsibility is highest, i.e. $u_n^{mode} = \arg\max_{u_k} r_{kn}$.

6.3.2 Two-Tier market segmentation

As stated in the introduction, one of the main roles of GTM and its extensions is providing interpretable visualization of multivariate data in a low dimensional latent space. For that, a sufficiently large number of mixture components, constrained to lie in a manifold, are required. Given that each component can be considered a cluster representative or prototype, GTM usually provides a detailed cluster structure. While this structure may be adequate for visualization, it is likely to be too detailed for practical market segmentation purposes. In order to overcome this limitation, the following two-tier clustering procedure, similar to that proposed in [268], will be used.

In the first tier, FRD-GTM is fitted to the available data, providing an unsupervised feature relevance ranking and a detailed cluster structure, where each cluster is represented by a mixture component (or by a

point in latent space). At this stage, churn behavior can be explored in some detail through the identification of the *departure gates* for customer abandonment of the service provider. These are defined as FRD-GTM clusters for which the probability of churning between the periods defined in *Section 6.2* is high. The probabilistic nature of the GTM allows the definition of a principled Churn Index for each non-empty cluster k as $CI_k = (\sum_n r_{kn})^{-1} \sum_{\{n^*\}} r_{kn}$, where $\{n^*\}$ is the subset of churning clients; r_{kn} is the probability of client n belonging to cluster k, or responsibility, which has been defined in *Section 4.3.3*. Similarly, the Commercial Margin associated to each cluster can be defined as $CM_k = (\sum_n r_{kn})^{-1} \sum_{\{n\}} r_{kn} cm_n$.

In the second tier, the prototypes or component centres of the GTM (in *Experiment 1*) and FRD-GTM (in *Experiment 2*) undergo a second level of clustering using k-means, which favours the obtention of spherical clusters [268]. Even if not the only one, k-means is a sensible second tier choice as GTM also favours spherical clusters, given that its mapping is carried through by Gaussian basis functions generating radially symmetric distributions in data space. In order to avoid misinterpretations, the clusters of prototypes obtained through the k-means procedure will herein be referred to as *macroclusters*. Once interpreted for marketing purposes, the macroclusters will be referred to as *segments*.

The second tier k-means procedure does not restrict, in principle, the number of macroclusters resulting from its application. An adequate number of macroclusters can be inferred from the calculation of suitable cluster validation indices.

Many of these have been defined in the literature [258] but, for this study (*see Experiment 2*), we focus on variations of the Davies-Bouldin (DB) index [268], suitable for the type of spherical macroclusters favoured by k-Means, and on the Gap statistic [245]. The DB index attempts to find clustering solutions that maximize between-cluster distances while minimizing within-cluster distances. In our two-tier approach, the DB index for a partition $Q_C \equiv \{Q_1, \ldots, Q_C\}$ of C macroclusters is defined as:

$$DB_C = \frac{1}{C} \sum_{C=1}^{C} \max_{C' \neq C} \frac{\frac{\sum_{k_c} \|y_{k_c} - c_c\|}{K_c} + \frac{\sum_{k_c} \|y_{k_{c'}} - c_{c'}\|}{K_{c'}}}{\|c_c - c_{c'}\|} \tag{6.3}$$

where an Euclidean distance norm is used; K_c and $K_{c'}$ are, in turn, the number of FRD-GTM clusters assigned to macroclusters c and c'; y_{k_c} and $y_{k_{c'}}$ are the corresponding cluster prototypes given, for each dimension d; and c are the centroids resulting from the k-means procedure.

This formulation of the DB index assumes *hard* cluster assignments according to which a data point is fully assigned to a single cluster. In contrast, the GTM and its extensions provide much richer information on cluster membership, as the responsibility r_{nk} indicates how much the GTM model estimates data point x_n to belong to cluster k. This more detailed information can be incorporated in full to the index if we define a responsibility-weighted DB as:

$$rwDB_C = \frac{1}{C} \sum_{c=1}^{C} \max_{C' \neq C} \frac{1}{\|c_c - c_{c'}\|} \left\{ \frac{1}{K_c} \sum_{k_c} \left\| \frac{\sum_n r_{nk_c} x_n}{\sum_n r_{nk_c}} - c_c \right\| + \frac{1}{K_{c'}} \sum_{k_c} \left\| \frac{\sum_n r_{nk_{c'}} x_n}{\sum_n r_{nk_{c'}}} \right\| \right\} \tag{6.4}$$

where the sums for each macrocluster c now run over the complete set of K clusters, and this over responsibility-weighted sums of all data points, instead of only over those attributed strictly to each macrocluster. In a trained GTM, responsibility usually concentrates in a very small number of mixture components or clusters and, quite often, in just one. This means that big differences between the values of DB_C and $rwDB_c$ would be a clear indication of abundant ambiguous cluster attributions and, possibly, multimodality.

The Gap statistic [245] considers the pooled within-cluster sum of squares around the cluster centroids. In our two-tier approach, this is defined as:

$$W_c = \sum_{c=1}^{C} \frac{1}{2K_c} \sum_{k_c, k_{c'}} \|y_{k_c} - y_{k_{c'}}\|^2 \tag{6.5}$$

where K_c is the number of data points in macrocluster c. The expression $\log(Wc)$ is then compared with its expectation under an appropriate null reference data distribution. According to this, the statistic is defined as the difference:

$$Gap_C = E^* \left[\log(W_C)\right] - \log(W_C) \qquad (6.6)$$

where $E^*\left[\cdot\right]$ denotes the expectation under a sample of the reference distribution (of the same size of the available data). The best cluster solution Q_C will be that which maximizes the Gap statistic. The uninformative features are generated from a uniform distribution over a box aligned with the principal components of the data. Other alternatives might be considered, but this one is specially consistent with the PCA-based initialization of FRD-GTM used in this study. Details can be found in [245]. In practice, the expectation $E^*\left[\log(W_C)\right]$ is estimated as ana average of $\log(W_C^*)$ over a number R of replications, each using a Monte Carlo sample of the reference data. The Gap statistic therefore becomes:

$$Gap_C = \frac{1}{R} \sum_r \log(W_{Cr}^*) - \log(W_C) \qquad (6.7)$$

Not all these cluster validity indices are meant to provide the appropriate number of clusters as a closed solution. Instead, they should be understood as a fair guideline to find it, so that a range of solutions may be considered to be valid. In the current market segmentation application (*see Experiment 2*), these quantitative criteria have to be balanced against practical commercial and marketing considerations.

The previously defined Churn Index and Commercial Margin can also be cumulatively defined for each macrocluster or segment c as, in sum, $CI_c = \sum_{k \subset Q_c} CI_k$, $CM_c = \sum_{k \subset Q_c} CM_k$.

6.4 Experiments

As described in *Section 6.1*, in this chapter we aim to explore the relationship between the telecommunications consumption behaviour of small and medium-sized companies in the Brazilian market and the propensity to abandon the service provider. The two experiments carried out to this end can be summarised as follows:

- **Visualization and clustering using GTM** (*Exp. 1*): In a first stage, the GTM model was fitted to the variables listed in *Table 6.1* for period P1. Each point of the GTM map (corresponding to a *micro-cluster*) was then characterized using the profiling variables described in *Section 6.2*. The resulting *micro-segments* were aggregated into macro-segments using the k-means algorithm in order to make the segmentation results more easily actionable from a business point of view.

  In a second stage, the same steps were followed for the data corresponding to P2, with the exception of clients who abandoned their service provider between periods. Comparing the position of the clients over the GTM maps in each of the periods studied, we aim to identify the micro-segments and the areas on the GTM map associated to higher probabilities of churn (*the departure gates*), as well as those associated to loss or increase of value between periods. Customer movement between areas should provide the basis for the development of a churn warning system.

- **Visualization and clustering using FRD-GTM** (*Exp. 2*): As in *Experiment 1*, in a first stage a FRD-GTM following a PCA-based procedure was fitted to the variables listed in *Table 6.1* for P1. Likewise, each micro-cluster was characterized using profiling variables. The resulting micro-segments were aggregated using k-means algorithm. An adequate number of *macro-clusters* for the k-means approach was selected using the three validation indexes described in *Section 6.3.2*.

  A second stage was implemented in an identical way as in *Experiment 1*.

The existing literature does not show comparable customer churn prediction examples and no other algorithms have been applied in this case.

6.4.1 Visualization and clustering using GTM (*Experiment 1*)

A GTM with an 8×8 cluster grid structure was implemented, following a standard PCA-based procedure for its initialization [263]. The 64 clusters, or *micro-segments*, corresponding to the data of the P1 period, are visualized in *Figure 6.2 (left)*; the relative size of each cluster is directly proportional to the number of clients assigned to it with the criterion described in the previous paragraph. The 64 GTM cluster prototypes were further grouped using k-means in order to find segments of increased market actionability. From a business perspective, four segments were identified this way as: 1.- *Local companies* (83% local calls); 2.- *Commercial companies with mobile employees* (30% calls to mobiles); 3.- *National companies* (43% national calls); and 4.- *Professionals* (15% calls outside working hours). These segments are visualized in *Figure 6.2 (right)*.
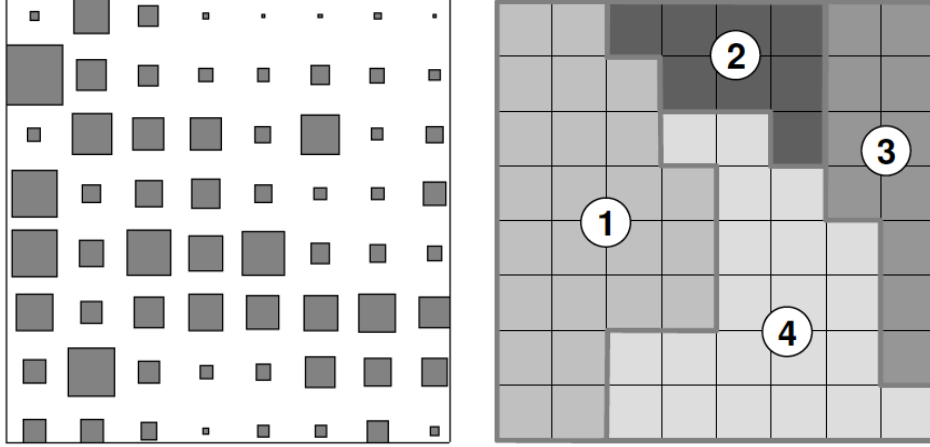
Figure 6.2: GTM maps of clusters and segments for data of the P1 period.

We are interested in the migration routes of clients across segments over the periods P1 and P2. These results are summarized in *Table 6.2*, where the percentages of clients moving (or not) from one segment to another are displayed. Segment 2 (*Commercial companies with mobile employees*) are by far the more volatile (highest value of *relocation* in *Table 6.2*), whereas segment 3 (*National companies*) are the most resilient to change (lowest value of *relocation* in *Table 6.2*). Besides, this volatility is directly proportional to the percentage of churn.

P1/P2	1	2	3	4	Churn	Relocation
1	70.4%	3.9%	0.3%	14.6%	10.8%	**18.8%**
2	9.4%	49.0%	10.9%	16.2%	14.5%	**36.5%**
3	0.8%	5.5%	73.7%	10.2%	9.8%	**16.5%**
4	16.3%	8.2%	8.1%	56.1%	11.3%	**32.7%**

Table 6.2: Segment mobility and percentage of churn over the P1-P2 periods.

We now explore churn in more detail through the identification of the departure gates for abandonment: GTM (micro) segments for which the probability of churning between periods is higher. The probabilistic definition of the GTM allows the definition of a principled churn index $P_{k,churn} = \left(\sum_{\{N\}} r_{kn} \right)^{-1} \sum_{\{N^*\}} r_{kn}$

This index results are colour-coded and displayed in *Figure 6.3* (left): white corresponds to the highest churn index, whereas black corresponds to the lowest. This is accompanied, on the right-hand side map, by a display of the values of commercial margin for each cluster. This last map reveals differentiated areas of commercial margin (clusters of highest margin in white: R$740; and of lowest margin in black: R$124). Using *Figure 6.2*, we see that high margins correspond mostly to small-medium sized micro-segments,

mainly belonging to the *National companies* market segment. Similarly, areas with neatly different churn index can be clearly identified. The high values of the churn index mostly appear associated to small micro-segments (the *departure gates*) belonging to two market segments: *Professionals* and *Commercial companies* with mobile employees. Interestingly, they are also micro-segments with low commercial margin, which is indicative of the commercial health of this operator's client portfolio.
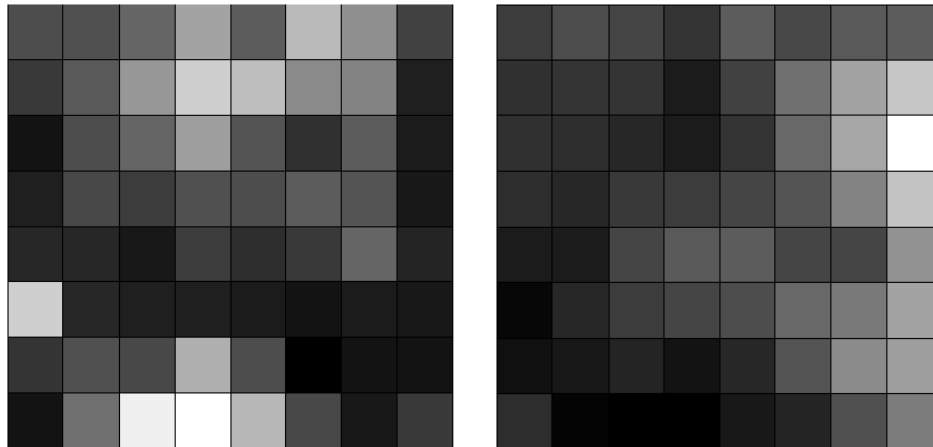
Figure 6.3: Churn index (left) and commercial margin (right) for each micro-segment *(Experiment 1)*.

In general practice, the combination of the churn index, commercial margin and the typical migration routes over periods can provide telecommunication operators with a decision support system that warned about profitable clients moving towards departure gates with higher probability of abandonment.

6.4.2 Visualization and clustering using FRD-GTM (*Experiment 2*)

In a first stage of the experiments, a FRD-GTM model with an 8×8 cluster grid structure, and following a standard PCA-based procedure for its initialization [152], was implemented for period P1 and fitted to the variables listed in *Table 6.1*. The relevance ranking results for these data are shown in *Figure 6.4*. As shall be seen below, the most relevant attributes according to the FRD-GTM ranking (p-ll-nor: *percentage of normal timetable calls*; p-ll-dif: *percentage of differentiated calls*; p-ll-ln: *percentage of local calls*; p-ll-ita: *percentage of intrastate calls*; and p-ll-mov: *percentage of calls to mobile*) will be the ones that better explain the macro-cluster (segment) solution obtained in the second tier of the clustering procedure described in *Section 6.3.2*. This means that the most relevant attributes provide the best explanation of both the micro-cluster structure obtained through the FRD-GTM, and the segments obtained using k-means.

An adequate number of macro-clusters for the k-means procedure had to be selected. To get a sensible indication for this, several suitable cluster validation indices were calculated. They were then challenged by purely market-driven criteria in order to ensure that the results could be used in a practical implementation. In this way, three validation indices (Davies-Bouldin index: *Figure 6.5* responsibility-weighted Davies-Bouldin index: *Figure 6.6*; and Gap statistic: *Figure 6.7*) for 2 up to 10 macro-cluster range (considered adequate from a commercial point of view) were calculated. All ratings indicate that a sensible solution is provided by 6 macro-clusters, which nicely match with an interpretable commercial description of market segments.

Figure 6.4: Ranking of relevance of the attributes.
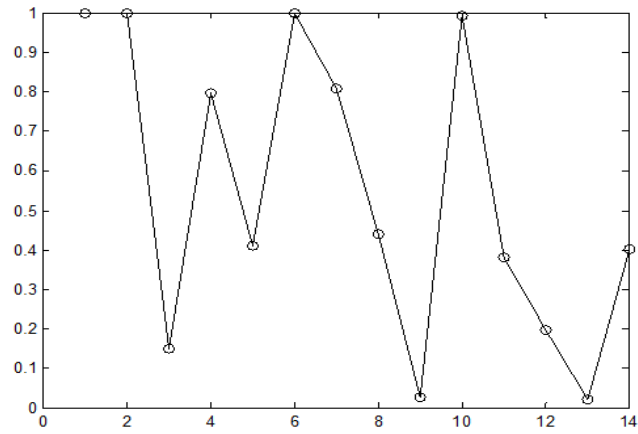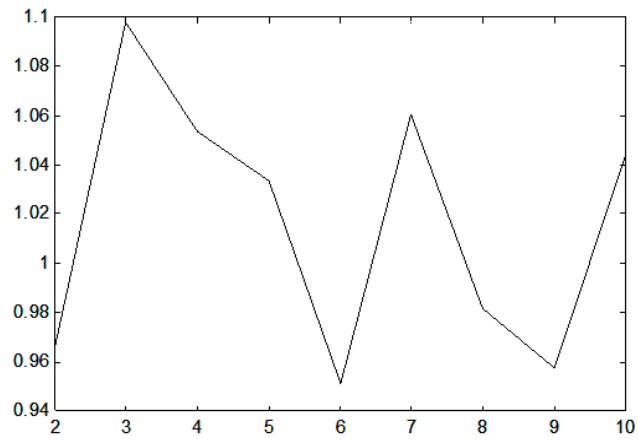

Figure 6.5: Davies-Bouldin index.
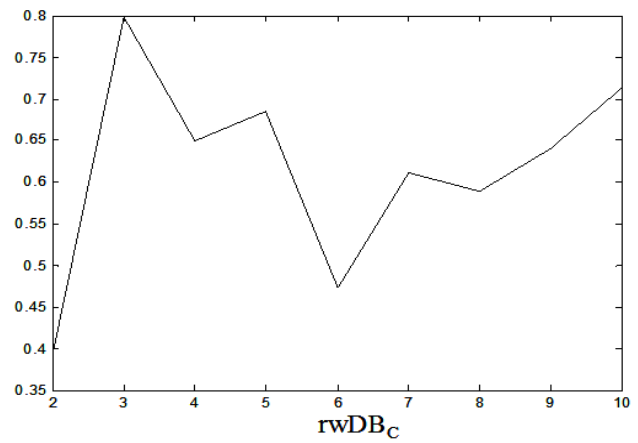
Figure 6.6: Responsibility-weighted Davies-Bouldin index
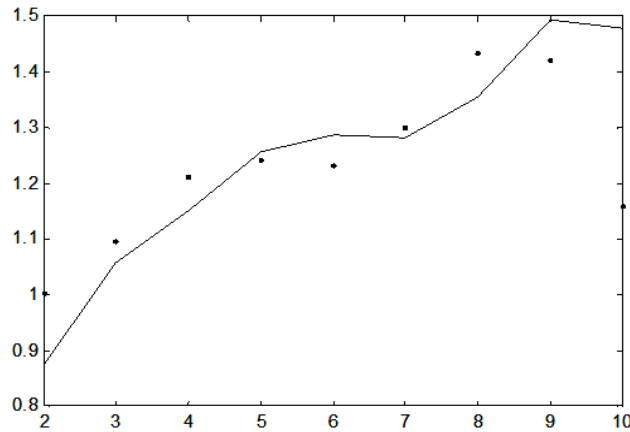
Figure 6.7: Gap statistic. The line shows the $Gap(C)$ whereas the points indicate $Gap(C+1) - S_{C+1}$. If $Gap(C) > Gap(C+1) - S_{C+1}$, then C represents an adequate number of clusters.

The 64 resulting clusters, or *micro-segments*, obtained by FRD-GTM in the first stage of the clustering process are visualized in *Figure 6.8* (left); the relative size of each cluster is directly proportional to the number of clients assigned to it. The 64 FRD-GTM cluster prototypes were further grouped using k-means in order to find segments of increased market actionability for practical purposes.

The 6 segments identified (and some of their defining features) were: 1.- *Regional companies* (44% intrastate calls); 2.- *Professionals* (27% calls outside working hours and 77% local calls); 3.- *Receivers* (52% entrants calls); 4.- *Locals* (82% local calls and 86% calls in normal hours); 5.- *Commercial companies with mobile employees* (28% mobile calls); and 6.- *National companies* (7% interstate calls). These segments are visualized in *Figure 6.8 (right).*
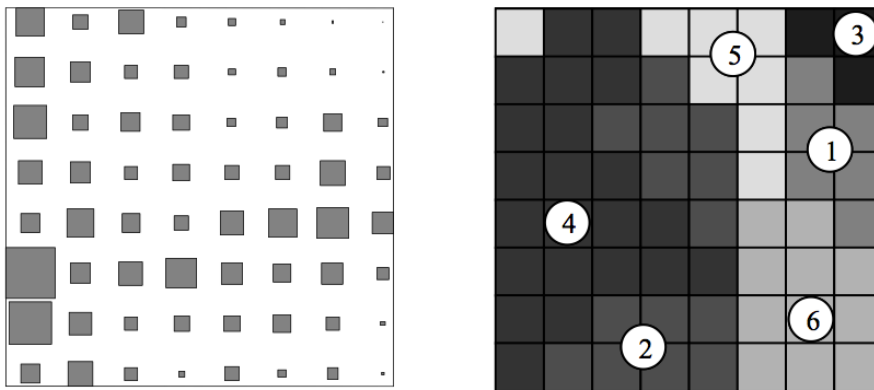
Figure 6.8: FRD-GTM maps of clusters (left) and segments (right) for data of the P1 period.

The same analyses were performed for the data corresponding to P2, with the exception of clients who abandoned the service provider between periods. Comparing the position of the clients over the GTM maps in each of the periods studied, we aim to identify the micro-segments and the areas on the GTM map associated to higher probabilities of churn (the *departure gates*), as well as those associated to loss or increase of value between periods. We are interested in the migration routes of clients across segments over the periods P1 and P2. These results are summarized in *Table 6.3*, where the percentages of clients moving (or not) from one segment to another are displayed. Segment 2 and 5 (*Professionals* and *Commercial companies with mobile employees*) are by far the more volatile (highest value of relocation in *Table 6.3*), whereas segment 4 (*Local companies*) are the most resilient to change (lowest value of relocation in *Table 6.3*). In addition, this volatility is directly proportional to the percentage of churn.

P1/P2	1	2	3	4	5	6	Churn	Relocation
1	51.7%	0.4%	9.2%	0.8%	3.9%	24.5%	9.5%	38.8%
2	0.8%	29.2%	0.1%	32.1%	7.0%	16.1%	14.7%	56.1%
3	15.9%	0.0%	55.2%	0.5%	14.8%	1.8%	11.8%	33.0%
4	0.1%	8.3%	0.0%	72.7%	1.7%	6.6%	10.6%	16.7%
5	14.6%	5.4%	1.1%	12.6%	36.6%	16.3%	13.3%	50.0%
6	4.9%	4.5%	0.2%	9.4%	6.2%	66.5%	8.2%	25.2%

Table 6.3: Segment mobility and percentage of churn over the P1-P2 period.

We now explore churn in more detail through the identification of the *departure gates* for abandonment: FRD-GTM (micro) segments for which the probability of churning between periods is higher. The Churn Index defined in *Section 6.3.2* is colour-coded and displayed in *Figure 6.9 (left)*: white corresponds to the highest Churn Index, whereas black corresponds to the lowest. This is accompanied, on the right-hand side map, by a display of the values of Commercial Margin (also defined in *Section 6.3.2*) for each cluster. This last map reveals differentiated areas of commercial margin (clusters of highest margin in white: R$882; and of lowest margin in black: R$130). Using *Figure 6.9* we see that high margins correspond mostly to small-medium sized micro-segments (compare it with their distribution in *Figure 6.8* mainly belonging to the 1 and 6 market segments (*Regional* and *National Companies*). Similarly, areas with neatly different churn index can be clearly identified. The high values of the churn index mostly appear associated to small micro-segments (the *departure gates*) belonging to two market segments: again 2 and 5 (*Professionals* and *Commercial companies with mobile employees*), which are also micro-segments with low commercial margin. This coincidence of high churn index and low margin, coupled with small segment size is indicative of the commercial health of this operator's client portfolio.
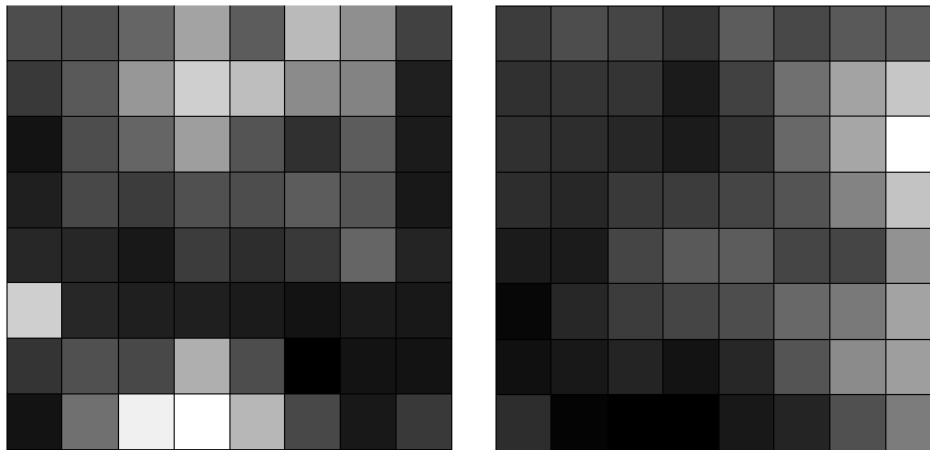
Figure 6.9: Churn index (left) and commercial margin (right) for each micro-segment *(Experiment 2)*.

In such scenario, no urgent marketing campaigns would be required and, instead, finely tuned and targeted strategies might be more suitable.

6.5 Conclusions

In highly evolved and strongly competitive markets, business strategies have to be tailored to customers' needs and requirements. Only this way companies can build their customer loyalty and avoid defection to the competitors, a pervasive phenomenon known as churn. The use of an effective model to explore

customer churn becomes, then, an important task for service providers. In this chapter we have used a novel probabilistic approach of the manifold learning family to cluster and visualize the clients of a major Brazilian telecommunications provider. In this scenario, differentiated customer groups must be identified, and, to this end, market segmentation can be a useful tool.

Quantitative market segmentation is commonly carried out through data clustering methods. We have defined a two-tier clustering procedure whose first tier is based on a probabilistic computational intelligence model of the manifold learning family: Generative Topographic Mapping. On top of the clustering results, GTM also provides intuitive data visualization. This model has been endowed with an in-built unsupervised feature relevance determination method that optimizes clustering by increasing the influence of those features that better describe the natural separation of data groups.

Data on service usage by the clients of a Brazilian telecommunications provider company have been clustered and the corresponding market has been segmented using the two-tier clustering based on GTM and FRD-GTM. The results have been validated with several cluster quality indices, one of them specifically defined for the GTM model. The resulting segmentation solution has also been assessed in business terms and found to be easy to describe according to the features found to be most relevant by the FRD procedure.

Two ad hoc segment solution evaluation metrics: Churn Index and Commercial Margin have also been defined. Different areas where the risk of abandonment are higher, or *departure gates*, have been identified on the basis of service consumption patterns. The migration routes between market segments have also been explored.

We understand that this model should provide the basis for the development of a churn warning system.

Chapter 7

Distortion visualization in GTM

As we introduced in *Chapter 3*, one of the constituting stages of most data mining and knowledge discovery methodologies currently in use is data exploration [76, 229]. It usually helps bringing into focus relevant aspects of the analyzed data. When these data are high-dimensional, and this is often the case, the task of data visualization becomes central to data exploration [154].

In this chapter, inspired from a technique originally designed for the analysis of geographic information -*cartograms* [91]-, we propose a new method for explicitly reintroducing the geometrical distortion created by an NLDR manifold learning model into its low-dimensional representation of the MVD. The proposed cartogram-based method reintroduces the distortion explicitly into the visualization maps. By reintroducing this distortion explicitly, we should now expect the inter-point distances in the low-dimensional representation space to more faithfully reflect those in the observed data space.

Thus, in the present chapter we first provide self-contained and summary descriptions of the antecedents and theory behind cartogram-based geographical representation, with its corresponding visualization of distortion-quantification measures. This is followed by the presentation of results and discussion of an extensive set of experiments, using artificial and real data. With these, we explore the properties of the proposed cartogram method and provide some guidelines for its use.

Results of the research described in this chapter were published by the international *Data Mining and Knowledge Discovery* journal (July 2013) [272] in their special issue "Intelligent Interactive Data Visualization".

7.1   Visualization of multivariate data using cartograms

The visualization of MVD, as used in the pursuit of knowledge generation, is a problem in between natural and artificial PR: Natural because information visualization entails complex cognitive PR processing of visual stimuli [129, 182]; and artificial because, in the face of complex MVD, researchers are compelled to develop visualization-oriented PR techniques, usually stemming from the fields of multivariate statistics and artificial intelligence. The natural and artificial aspects of the visualization PR problem are both relevant and inextricable and, as a result, the use of visual metaphors entails the risk of introducing subjectivity in the knowledge generation process [297]. If both aspects are used at their best, they can enhance each other in order to make data exploration a fruitful task [270].

For the exploratory analysis of MVD using visualization, PR techniques are required to provide scalability [220] in what in fact becomes an extreme form of DR. This is because, at most, human vision can simultaneously make sense of a handful of data attributes in 2-dimensional or 3-dimensional interactive displays. This reduction of dimensionality can be achieved through different approaches, including feature selection [101], feature extraction [102] and clustering [126, 127], amongst others.

A common characteristic of all DR methods for MVD visualization is that they result in information loss, in one way or another. The faithfulness of the low dimensional data representation they provide is

unavoidably limited because it requires a radical simplification of the observed data. At best, these DR methods can aspire to minimize the distortion of the observed data in their representation, according to some objective function.

Some of the most popular DR techniques for visualization are of the feature extraction type and linear in nature. Linear DR methods (a well known and extensively used example of which is PCA [131]) are quite constrained in the MVD transformation they can provide and, as a result, their data representation risks being of limited faithfulness in some cases. Compensating for this, their subset of representation coordinates (or extracted features) can be expressed as a linear combination of the original coordinates (that is, of the observed data attributes), which makes these models easy to interpret for practical purposes, without resorting to often cumbersome post-processing procedures.

Many relevant recent contributions to MVD visualization have stemmed from the field of nonlinear DR [155] and, more in particular, from spectral-based methods [205, 221] and techniques of the manifold learning family. These include methods for the quantification and visualization of the quality of the DR process [273]. Manifold learning attempts to describe (usually high-dimensional) MVD through nonlinear low-dimensional manifolds embedded in the observed data space. These manifolds generate a model by "wrapping around" data while usually preserving their continuity and smoothness properties.

Almost as popular in nonlinear DR for visualization as PCA is in linear DR, the SOM [143] and its many variants attempt to model MVD through a discrete version of a manifold consisting of a topologically-ordered grid of cluster centroids. SOM, as a vector quantization technique, clusters data points according to their proximity to these centroids (The popular k-Means clustering algorithm [126] can, in fact, be seen as an specific instantiation of SOM).

The nonlinearity of these methods entails the existence of different levels of local distortion in the mapping of the data from the observed space into the visualization space. Given that most of these methods rely upon the definition of inter-point distances (Euclidean being the most commonly used) in the metric spaces they deal with, there is no guarantee that the inter-point distances in the observed data space will be uniformly reflected in the visualization space. In other words, points which are distant in the observed data space may end up being represented as closely located in the visualization space and the other way around. These manifold stretching and compression effects can be understood as *geometrical distortions* introduced by the nonlinear mapping [9]. Such effects can also be seen as a local *magnification* process. Recent research has investigated the possibility of actively controlling this magnification as part of the learning of NLDR techniques [106, 278].

The data representation flexibility provided by nonlinear DR methods often makes them more faithful models of the observed MVD than linear ones. The price that these methods must pay for such ability is the usually less straightforward interpretability of the visualizations they provide [271], given that the coordinates of visual representation are no longer linear combinations of the original data attributes. This limitation of NLDR methods makes the definition of approaches to circumvent it a worth-pursuing research goal on its own right.

In this context, we draw inspiration from a technique originally devised for the analysis of geographic information, namely density-equalizing maps, or cartograms [91]. Cartograms are geographic maps in which the sizes of regions such as countries or provinces appear in proportion to underlying quantities such as their population. They have a limitation in that, to scale these regions while not losing their continuity properties, regions' shapes must be distorted in one way or another, potentially resulting in maps that are not obvious to read. The technique proposed in [91] for building cartograms retains the interpretability of the maps while distorting them, but without suffering drawbacks such as the undesired overlapping of regions or a too strong dependence on the choice of coordinate axes.

The continuity-preservation requirements generated by nonlinear manifold learning techniques are akin to those generated by geographical maps. Thus, the conceptual leap in this study consists on extrapolating from geographical maps to the *virtual geographies* of the visualization spaces of manifold learning NLDR models. It also requires the substitution of geography-distorting quantities such as population density by quantities reflecting the mapping distortion introduced by these nonlinear models.

The use of cartograms for the faithful visualization of the nonlinear projections of manifolds generated by DR models falls within the field of computational topology [66]. Here, we illustrate the proposed cartogram-based method with a manifold learning model for which this distortion is readily quantifiable in a continuum over the data visualization space, GTM [23] (see *Chapter 4*). This is a manifold-constrained

mixture model [178] with functional similarities to SOM. It also provides, beyond MVD visualization, vector quantization through the definition of manifold-embedded data prototypes (cluster centroids). In our cartogram-based method, the political borders of geographic maps are replaced by the GTM regular grid of prototype-generating points in the visualization space, while map-underlying quantities such as density of population are replaced by the GTM-induced distortion in the form of Magnification Factors [22].

The proposed cartogram-based method reintroduces the distortion, as expressed by the Magnification Factors, explicitly into the visualization maps. By doing so, we argue that these distorted maps are more representative and, importantly, more intuitively interpretable than the existing implicit method consisting on the joint visualization of the data projection map and the colour-coded Magnification Factors.

Although illustrated with the standard version of GTM in this study, the cartogram visualization of distortion could easily be extended to other variants of the GTM [55, 94, 241, 259] as well as to other NLDR visualization methods, provided a local distortion measure, or some approximation for it, could be calculated.

7.1.1 Methods

In this section, we provide a summary description of the concept of cartogram, its use for the representation of geographic information and the use of physics-inspired techniques for the generation of faithful cartogram representations.

7.1.1.1 Density-equalizing cartograms

Cartograms are cartography maps in which specific areas, often delimited by political borders, are locally distorted (stretched or compressed) to account for locally-varying underlying quantities of interest, such as population density or socio-economic data. The first computer-based cartograms can be traced back to the early work of Waldo R. Tobler [248]. The use of cartograms for the visual representation of socio-economic data in geographical maps has become widely popular of late through public resources such as Worldmapper[1].

The geometrical distortion of cartograms takes (in 2-D) the form of a continuous transformation from an original plane to a transformed one, so that a vector $\mathbf{x} = (x^1, x^2)$ in the former is mapped onto the latter according to $\mathbf{x} \rightarrow T(\mathbf{x})$, in such a way that the Jacobian of the transformation is proportional to an underlying *distorting variable* $\mathbf{d}$:

$$\frac{\partial(T_{x^1}, T_{x^2})}{\partial(x^1, x^2)} \propto \mathbf{d}. \tag{7.1}$$

A computationally-feasible approach to this map distortion process requires the discretization of the plane continuum (and the corresponding distorting variable) to conform a rectangular or hexagonal regular grid. The distorting variable is assumed to take a uniform value over each of the plane fragments defined by the grid. Distorting the map locally in this manner may result in loss of connectivity between the fragment borders.

A method for cartogram building based on the physics principle of linear diffusion processes was recently proposed in [91]. In this method, the distorting variable $\mathbf{d}$ is let to *diffuse* over the map *over time* so that the final result, for $t \rightarrow \infty$, is a map of uniform distortion in which the original locations have displaced according to the process, while preserving the integrity of the existing borders (if $\mathbf{d}$ is population density, the resulting maps are density-equalizing cartograms).

In this instance of the diffusion process, the *current density* $\mathbf{C}$ follows the gradient of the distortion $\nabla \mathbf{d}$ and can be written as product of the current flow velocity $\mathbf{v}$ and the distortion itself, so that $\mathbf{C} = -\nabla \mathbf{d} = \mathbf{v}(\mathbf{x}, t)\mathbf{d}(\mathbf{x}, t)$. The standard diffusion equation takes the form

---

[1] www.worldmapper.org

$$\nabla^2 \mathbf{d} - \frac{\partial \mathbf{d}}{\partial t} = 0, \tag{7.2}$$

which has to be solved for distortion $\mathbf{d}(\mathbf{x}, t)$, assuming that the initial condition corresponds to each map fragment being assigned its value of the distorting variable. Thus, the distortion *diffusion velocity* can be calculated as $\mathbf{v}(\mathbf{x}, t) = -\frac{\nabla \mathbf{d}}{\mathbf{d}}$ and, from it, the map location displacement as a result of which the cartogram is generated can be calculated as:

$$\triangle \mathbf{x} = \int_0^t \mathbf{v}(\mathbf{x}, t') dt'. \tag{7.3}$$

If it were expressing population density instead of any generic distortion, the process could be seen as a flow of population from more to less densely populated areas until density is equalized. It could also be seen [91] as a population Gaussian random walk over the map that, over time, would reach density equalization and in which internal borders would be modified so as to keep a zero net flow through them. To avoid arbitrary diffusion through the overall map boundaries, the map is assumed to be surrounded by an area in which the distortion is set to be the mean distortion of the complete map. This guarantees that the total map area remains constant.

7.1.2 Cartogram visualization of the GTM magnification factors

All the elements of the method are now in place: an NLDR method for MVD visualization with which to illustrate the cartogram representation; a model distortion measure for GTM, the MF, to be used for map equalization (see *Chapter 4*); and a cartogram building procedure. In the following experiments, the GTM latent representation map is transformed into a cartogram using the square regular grid formed by the lattice of latent points $\mathbf{u}_k$ as map internal boundaries and assuming that the level of distortion in the space beyond this square is uniform and equal to the mean distortion over the complete map, that is $1/K \sum_{k=1}^{K} J(\mathbf{u}_k)$, where $J = \det^{\frac{1}{2}}(\mathbf{\Psi}^T \mathbf{W}^T \mathbf{W} \mathbf{\Psi})$. Likewise, we assume that the level of distortion within each of the squares associated to $\mathbf{u}_k$ is itself uniform. As a result, the finer the discrete GTM latent lattice (or, equivalently, the higher the number of points sampled from latent space) that we choose, the more accurately the cartogram will represent the MF local distortion.

The method, as applied in this study, can thus be summarized as the following succession of steps:

- GTM model initialization (see *section 4.3.1*), including:

    - Definition of a latent square grid of K points.

    - Initialization of the model parameters according to a standard procedure described in Bishop et al. [23]: The weight matrix $\mathbf{W}$ is chosen as to minimize the difference between the prototype vectors and the vectors that would be generated in data space by a partial PCA. The inverse variance parameter β is initialized as the inverse of the 3rd PCA eigenvalue. This initialization procedure ensures the replicability of results.

- GTM iterative training: using a maximum likelihood approach, as described in *Section 4.3.2*, and the EM algorithm.

- Calculation, from the model training results, of the posterior mean projections $\mathbf{u}_n^{mean}$ for all data points, as described in *Section 4.3.3*. These will be used for data visualization.

- Cartogram generation, including:

    - Description of the GTM latent grid as a pixelated image in which each node of the latent space is assigned a square of $p \times p$ pixels.

    - Calculation, from the model training results, of the MF for each pixel location in the latent space, as described by *equation* (4.20), in *Section 4.3.4*.

- Assignment of distortion values (average $1/K \sum_{k=1}^{K} J(\mathbf{u}_k)$, where J is given by *equation* (4.20)) for the overall external border of the GTM latent grid.

- Iterative calculation of the MF distortion velocity, as described in *Section 7.1.1.1*, and the corresponding location displacement defined by *equation* (7.3) for each pixel of the map, until obtaining the final cartogram.

- Location displacement calculation for the posterior mean projections of the data points and positioning of these displaced projections in the cartogram.

7.2 Experiments

The cartogram-based representation method described in this chapter is meant to merge the powerful modeling capabilities of NLDR methods and the explicit measurement of the nonlinear distortion they generate. In doing so, we aim to provide an intuitive and compact visualization tool for the exploration of MVD data.

In the case of GTM, used here to illustrate the cartogram method, the direct visualization of the distortion in the form of MF on its latent space can provide insight into the possible existence of dense data areas (or data clusters) and the sparsely populated areas that separate them [241]. This is because they are likely to undergo very different levels of distortion as a result of the nonlinear mapping. Unfortunately, this direct visualization of the MF is not always too intuitive and its direct superposition over the GTM visualization space may become impossible to interpret, especially for large data sets.

The cartogram-based representation of the GTM visualization space, in which the observed data are mapped according to either the *mode* or the *mean* projections, should instead retain the simplicity of the representation while factoring in the mapping distortion as measured by the MF. This way, we expect this method to provide clearer visual insight into the cluster structure of the data.

Atypical data, or outliers, are known to have a potentially negative impact in data modeling in general, and in NLDR methods in particular [259]. If outliers inhabit sparsely populated areas, we can hypothesize that they should, in general, be mapped onto areas of high distortion of the visualization space. Thus, we would expect cartograms to provide useful visual clues about data outliers.

The following experiments, in which both artificial and real datasets were analyzed, have the objective of assessing these expectations and, as a result, provide the data analyst with some general guidelines about the interpretation of the cartogram-based visual representation.

7.2.1 Experiments with artificial data

The cartogram visual representation of MVD using GTM, described in the previous section, was first investigated in some detail using artificial datasets. Data of simple statistical properties were used before analyzing more involved real-world datasets, so that the impact of their varying characteristics could be straightforwardly interpreted.

The first group of experiments involves 3-D data, which were used to obtain a preliminary but detailed insight into the NLDR mapping process and the visualization of its distortion. This is followed by a more thorough experimental assessment in which several model and data characteristics were varied.

7.2.1.1 Preliminary experiment with 3-D artificial data

We start by providing some initial impressions of the cartogram representation and by investigating the hypothesis that data outliers will be mapped onto areas of high distortion of the visualization space expressed as a cartogram. A simple statistic, described in [206], and extended to the GTM in [259], will be used to characterize to what extent a data point $\mathbf{x}_n$ can be considered to be an outlier. It is defined as

$O_n = \sum_{k=1}^{K} r_{kn} \beta \|\mathbf{y}_k - \mathbf{x}_n\|^2$, where $r_{kn} \equiv p(\mathbf{u}_k|\mathbf{x}_n)$ is the responsibility defined in *equation* (4.15) . An outlier is expected to yield comparatively large values of O_n.

For this experiment, a total of 1,500 3-D points were randomly drawn from three spherical Gaussians (500 points each), all with unit variance, and with centres sitting at the vertices of an equilateral triangle. 3-D data will allow the direct visualization of the model prototypes $\mathbf{y}_k$ (and as a result, the visualization of the generated manifold) in the observed data space. They were modeled using a GTM with a 20×20 grid of latent points.

Nine outliers, in three groups of three each, were first added to the previously described data:

- *Three outliers at the edges of the triangle (type A)*: These three outliers are away from the clusters, over the edges of the imaginary triangle defined by them, and within the plane in which this imaginary triangle would lie.

- *Three outliers near the centroid of the triangle (type B)*: These three outliers are away from the clusters, near the centroid of the imaginary triangle defined by the three clusters, and within the plane in which this imaginary triangle would lie.

- *Three outliers outside the triangle but not far away from the plane in which it lies (type C)*: One of them (*C1*) is located in the direction of one of the cluster centres, at right angles with the plane defined by the three cluster centres, but not too far from the cluster itself; a second one (*C2*) is located near the centroid of the imaginary triangle defined by the three clusters; and a third (*C3*) is located in between two clusters, over one of the edges of the imaginary triangle defined by the three clusters. The three of them are atypical in one way or another with respect to the rest of the data set. The GTM, though, fits these data very differently.

Results and discussion

The original data, together with the nine outliers are superimposed in *Figure 7.1* (top row, left) to the prototypes $\mathbf{y}_k$ and to the approximation of the manifold in which they lie, as generated by the GTM. This smoothly stretching manifold lies near the plane defined by the triangle of clusters. This means that the outliers have not exerted much influence on the GTM data fitting process. The latent space mapping of this data using the posterior mean projection described in *Section 4.3* is also displayed in *Figure 7.1* (top row, right).

The corresponding MF and cartogram can be seen in *Figure 7.1* (center row, left and right, respectively). Areas of high distortion neatly separate the three clusters and the area of highest distortion roughly corresponds to the central area of the imaginary cluster triangle. An interesting effect can be observed: the manifold is less distorted in the directions that join each pair of clusters (compare the MF rope-like features in *Figure 7.1* (center row, left) that link the areas in which the clusters are mapped with the manifold folding at the edges of the imaginary cluster triangle visualized in *Figure 7.1* (top row, left)).

The mapping of the nine outliers is quite telling. Outliers of the *type A* are located in areas of relatively high distortion as measured by the MF, but they are not as well characterized as outliers by the O_n measure, as displayed in *Figure 7.1* (bottom row). This is caused by the aforementioned lower distortion in the directions that join each pair of clusters, which result in a relatively higher concentration of prototypes. Instead, outliers of the *type B* are neatly mapped onto areas of relatively high distortion. This is consistent with their values of MF, but, again, because they lie so close to the manifold, they are not well-characterized by the O_n measure. In fact, the examples of type A and B illustrate a limitation of the own O_n measure: it becomes a poor indicator of atypicality if outliers lie close to the manifold. Finally, outliers of *type C* have a mixed behavior: Those roughly over the triangle centroid and edges behave similarly to their counterparts of types A and B, whereas the one approximately in a perpendicular to the manifold and over one of the clusters is assigned to the prototypes that represent that cluster. Thus, even if all these points show high O_n values, the third one is assigned to a low MF area and will thus not be visualized in the high distortion areas of the cartogram.
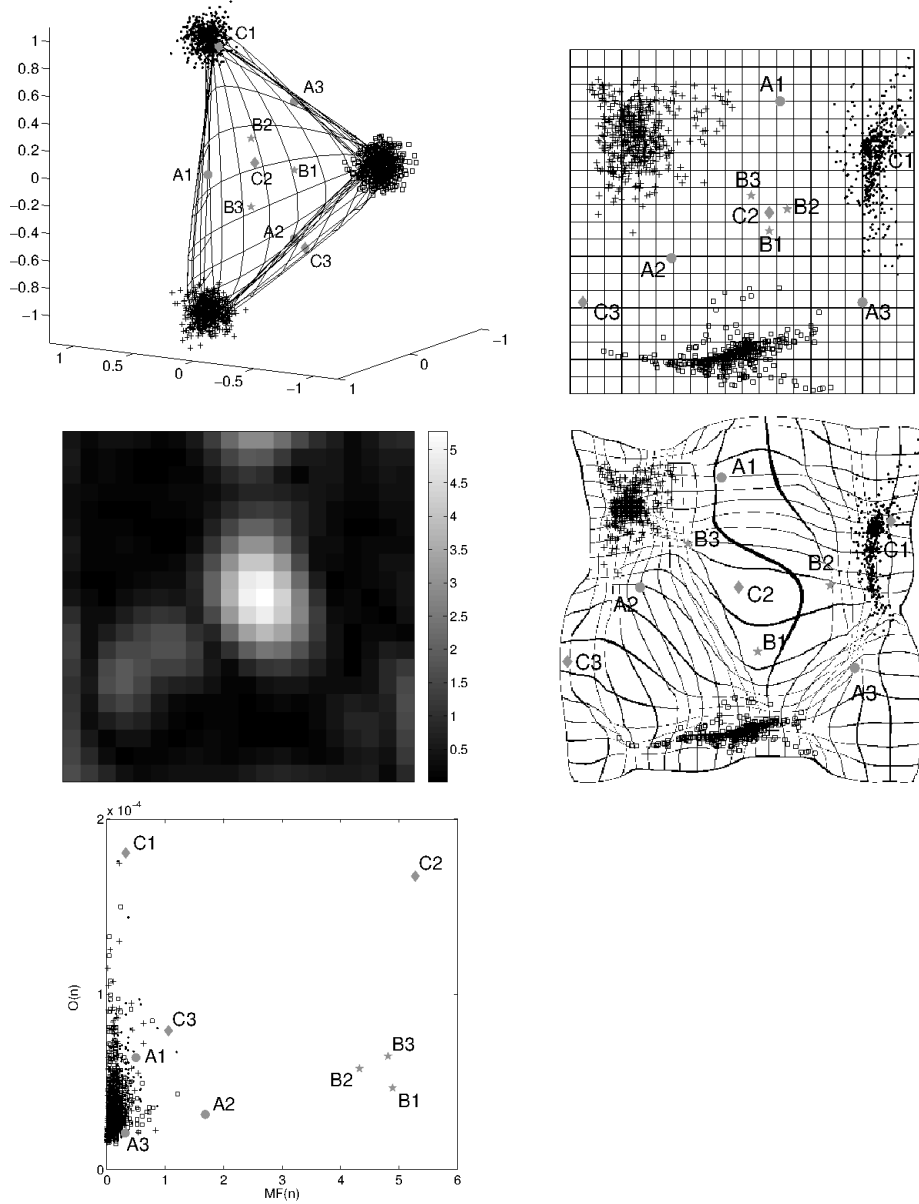
Figure 7.1: Cartogram visualization for the first of the outlier experiments with 3-D data. *Top row:* left) direct visualization of the 3-D observed data together with the model prototypes $\mathbf{y}_k$ linked according to the lattice of corresponding latent points (in an approximation of the GTM-defined manifold). Nine outliers as gray symbols, characterized (A, B, C) as described in the main text; right) GTM visualization map using the posterior mean data projection. Center row: left) The MF, color-coded over the GTM visualization map, with scale column; right) cartogram representation. Bottom row: Values of O_n versus MF for all data points, including the nine outliers.

Guideline 1: *When atypical data are away from the areas of main data density but still near the model manifold, their mapping location can be unexpected and, as a result, they might not always end up in the areas of highest distortion. Rules here are likely to depend on the NLDR method used. For GTM, data points that are located near the manifold and away from the directions linking pairs of clusters are likely to end up in the highly distorted areas of the cartogram. Instead, data points that are located near the manifold and in the general directions linking pairs of clusters might well be mapped away from clusters but not in highly distorted cartogram areas. Finally, data points that are only moderately away from both the clusters and the manifold might not always be mapped either away from clusters or in areas of high*

distortion of the cartogram. In summary, the data analyst might benefit from isolating the data points mapped onto the high-distortion areas of the cartogram to further investigate them as potential outliers, but bearing in mind that some of the outliers might not be amenable to this characterization.

For the next part of this experiment, a different set of three clear outliers (type *D*) were added to the original data, previously described. These points were located further away from the plane defined by the centres of the three clusters, at distances from it that were larger than the inter-cluster distances. The GTM was fitted to this augmented dataset and the results are shown in *Figure 7.2*.
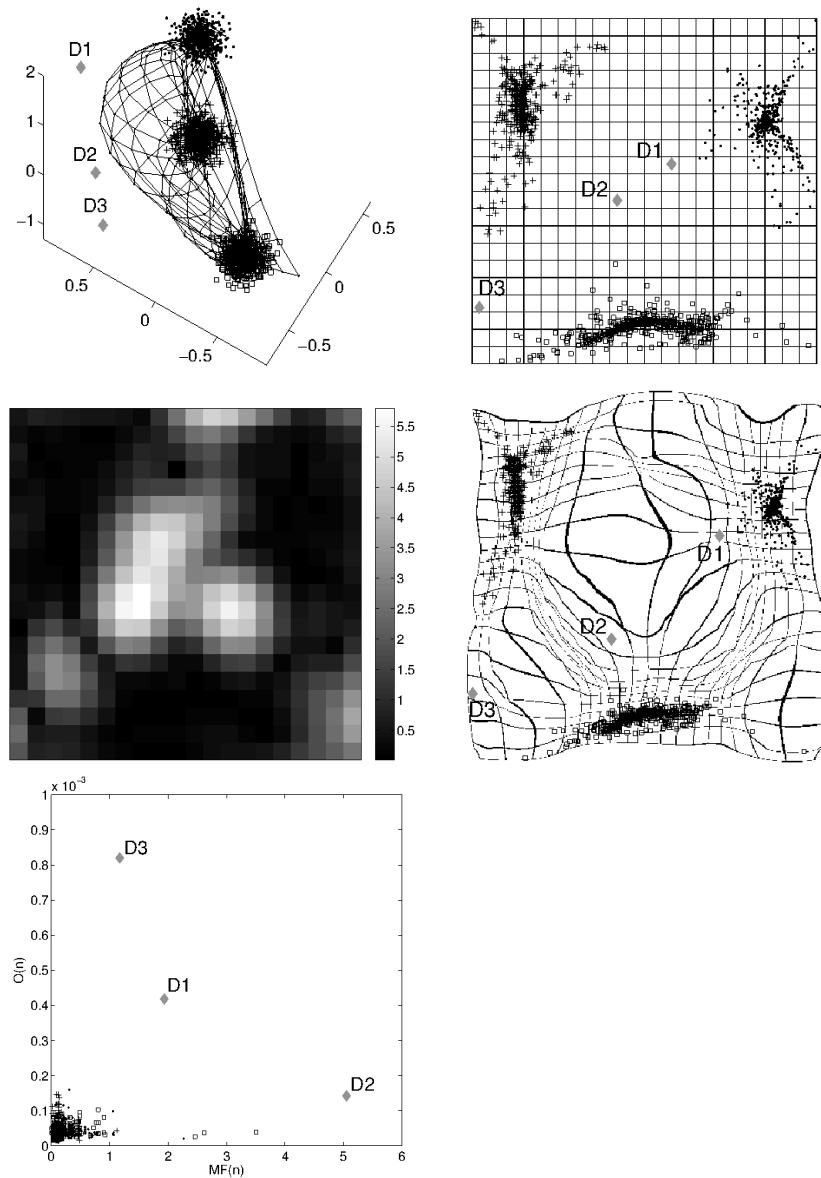
Figure 7.2: Cartogram visualization for the second outlier experiment with 3D data. Representation as in the previous figure.

The direct visualization of the fitted manifold is revealing: just three outliers are enough to exert quite a pull on this manifold, fairly stretching it towards them (see *Figure 7.2*, top row, left). The result is that they are mapped onto latent points that are away from the clusters (see *Figure 7.2*, top row, right) and which correspond to the stretched part of the manifold, as seen in *Figure 7.2* (center row, left). This is despite the

fact that one of them (*D1*) is located in the direction of one of the cluster centres, at right angles with the plane defined by the three cluster centres; a second one (*D2*) is located approximately in the direction of the centroid of the imaginary triangle defined by the three clusters, at right angles with the same plane; and a third (*D3*) is located in the direction of one of the edges of the imaginary triangle defined by the three clusters, at right angles with the same plane.

This is unlike in the previous experiment, where the location of the outliers in relation to the clusters affected their mapping location. Unsurprisingly, they all end up confined into the high-distortion area of the cartogram, displayed in *Figure 7.2* (center row, right). This can again be quantitatively assessed by comparing the values of the statistic O and the MF, as reported in *Figure 7.2* (bottom row): The three outliers show, simultaneously, high values of O_n and MF.

Guideline 2: *Outliers clearly away from the areas of main data density are likely to be mapped into areas of high distortion. This should at least be the case for unregularized NLDR models or density models based on Gaussian distributions (or other distributions that not behave well in the presence of outliers). Therefore, the data analyst might benefit from isolating the data points mapped onto the high-distortion areas of the cartogram to further investigate their atypicality.*

7.2.1.2 Further experiments with artificial data

In the following experiments, data were randomly drawn from radially symmetric normal distributions (with centers located at similar distances). In the experimental setting, several model and data parameters were manipulated as follows:

- *GTM architecture* (exp1): GTM square lattices of different sizes were used, from a basic 10×10 grid of latent points, up to a 30×30 grid. With this, the impact of the representation granularity on the cartogram visualization of the data was assessed. Three Gaussian clusters of 1,000 points each were used.

- *Dimensionality of the data* (exp2): Different data dimensionalities, from three to twenty, were investigated. We hypothesize that the complexity of the manifold embedding generated by GTM in spaces of increasing dimensionality should impact on the low-dimensional visual representation of the data and, as a result, on the distorted cartogram.

- *Number of points per Gaussian cluster* (exp3): By varying the number of data points drawn from each normal distribution, from 250 points per Gaussian (p.p.G.) to 5,000 p.p.G., we intended to assess the impact of data sparsity on the GTM representation, the MF calculation and, as a result, the cartogram visualization. The size of the GTM lattice was fixed to 30×30.

- *Relative density of the Gaussian clusters* (exp4): Three 3-D Gaussians of 250 points each and of different relative densities (by varying the variance σ^2) were used. We aimed to assess the effect of the different degrees of cluster compactness on the MF and its cartogram visualization. This is likely to become an important feature when analyzing the cluster structure of real datasets, which is usually far from homogeneous in terms of compactness.

- *Number of clusters* (exp5): Datasets consisting of three to twelve clusters were created from normal distributions of equal variance. The original illustration [22] of the use of MF to describe the borders between clusters as represented by GTM resorted to examples with a small numbers of clusters. We wanted to investigate the effects of a *cluster-crowded* representation space on the MF-based cartogram visualization.

We expect the explicit reintroduction of the MF into the GTM visualization map, in the form of cartograms, to yield a more compact representation of dense data clusters than the standard GTM posterior mean projection. We also expect it to yield a less compact representation of the less data-populated areas in observed space. This should make cartograms a more intuitive visual representation of the original data structure.

Such effect can be quantified through a variation of the standard Davies-Bouldin (DB) cluster validity index [60, 138]. The DB index is a ratio of intra-cluster inter-point distance variability to inter-cluster variability. Here, we adapt it to measure distances in the 2-D latent visualization space instead of measuring them in the observed space.

The inter-cluster variability D_{ij} for clusters C_i and C_j is described by the Euclidean distance between the centroids of the cluster projections in latent space, that is, $D_{ij} = \|\mu_i - \mu_j\|$.

The intra-cluster variability S_i for cluster C_i is described by the scatter $S_i = ((N_{C_i})^{-1} \sum_{n=1}^{N_{C_i}} \|\mathbf{u}_n - \mu_i\|)^{\frac{1}{2}}$, where N_{C_i} is the number of data points assigned to C_i and $\mathbf{u}_n$ is the location on the latent space representation of the projection of data point $\mathbf{x}_n$, calculated either directly from the posterior mean projection $\mathbf{u}_n^{mean} = \sum_{k=1}^{K} r_{kn}\mathbf{u}_k$ as described in *Section 4.3.3*, or from its corresponding cartogram displacement.

The adapted DB (aDB) for C clusters thus takes the form:

$$aDB = C^{-1} \sum_{i=1}^{C} max_{\{j:i\neq j\}}\left(\frac{S_i + S_j}{D_{ij}}\right). \tag{7.4}$$

Obviously, we do not intend to assess the validity of the cluster solution. Instead, the aDB will allow us to compare the differences in compactness between two alternative visualizations of a controlled clustering experiment. In the previously listed experiments, we would expect the aDB to be smaller for the cartograms than for the posterior mean projection representations, reflecting the fact that the cartogram should visually capture the compactness and separation between the analyzed data clusters better than the posterior mean projection. The only exception could be *exp4*, in which the clusters are built to have different levels of compactness.

Results and discussion

The results of this broad palette of experiments are now reported, discussed, and accompanied by some guidelines that might ease the use and interpretation of the method by its potential users.

GTM architecture (exp1): The results of varying the granularity of the latent subspace lattice are shown in *Figure 7.3*. This type of display will be used in all experiments of this section. It includes, on the left hand side column, the standard GTM visualization map with the *posterior mean projection* of the data, as described in *Section 4.3*. The GTM grid is superimposed. The central column of the figure displays the MF, color-coded over the GTM visualization map; the maximum distortion corresponds to white and the minimum distortion to black. The scale of this color-coding is also displayed by the map for comparative purposes. The cartogram representation in which the GTM map is distorted according to the MF is displayed on the right hand side column of the figure, and, again, the grid is superimposed.

For the sake of brevity, the results corresponding to only three grid sizes are reported. Further experiments with other sizes were consistent with them. No qualitative changes are observed as the grid size increases and, therefore, sampling more latent points only affects the resolution of the display. As expected, areas of higher data density correspond to low values of the MF, whereas the empty space between clusters corresponds to higher MF values, which means that they undergo stronger levels of distortion. This is clearly and intuitively reflected on the cartogram representations. Notice though that the MF direct representation in the square grids of the central column of *Figure 7.3* is far less obvious: The areas in which the data clusters reside show clear low MF values, as we might expect, but they are linked by *rope*-like looking features of low MF values that are nothing but an artifact of the own manifold stretching, similar to the ones reported for the experiments in *Section 7.2.1.1*. They are likely to reflect the concentration of prototypes in the directions that join cluster centres and might thus falsely hint at the existence of cluster sub-structure.
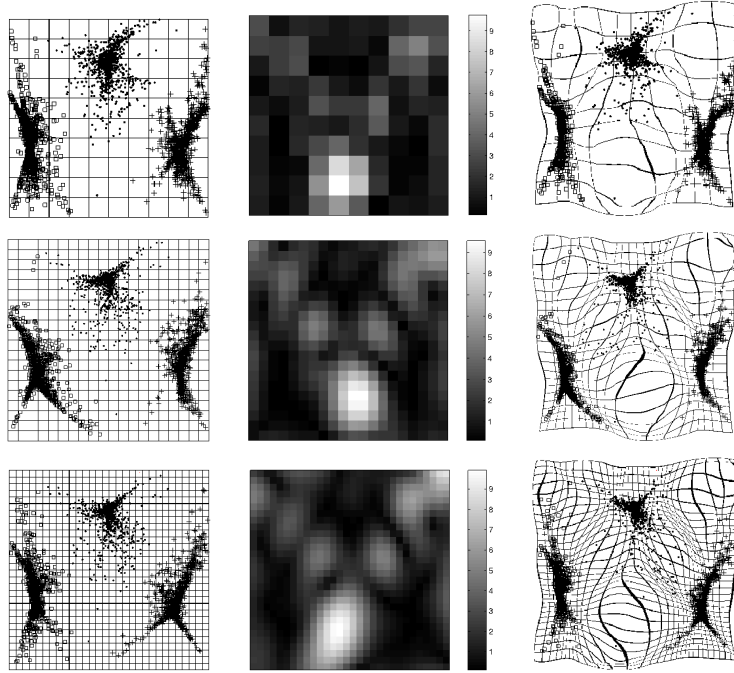
Figure 7.3: Varying the resolution of the GTM grid. By *column*: left) GTM visualization map using the posterior mean data projection; center) The MF, color-coded over the GTM visualization map, with scale column; right) cartogram representation. By *row*: top)10×10 resolution maps; middle) 20×20 resolution maps; bottom) 30×30 resolution maps.

Guideline 3: *The latent grid size only provides different degrees of detail but no qualitative changes in the cartogram representation. Computational burden notwithstanding (for instance, if large datasets are modeled), large grid sizes should be chosen to obtain detailed cartograms (Note that this is valid not only for GTM, but also for models such as SOM and its many variants).*

Dimensionality of the data (exp2): Given that increasing the grid size just provides better resolution, the effect of varying the number of data points in each cluster is now investigated with a fixed square grid of 30×30 size.

The levels of mapping distortion are likely to increase as data dimensionality increases and, with it, data sparseness. Unusual and counter-intuitive effects are expected to be observed in nonlinear manifolds embedded in high-dimensional spaces, due to their inherent *emptiness* (in what is known as "empty space phenomenon" [155]). Different data dimensionalities were investigated. Three of them, namely 3, 10 and 20 are reported in *Figure 7.4*. At first sight, the changes in the GTM map visualization as dimensionality increases are not too dramatic: the vertical alignment of the clusters seemingly increases and a reduced number of data points seem to conform *tail*-like structures sprouting from the cluster denser parts.

The MF maps tell us a different story: The existence of three clusters becomes clearer as the dimension increases, with distortion (see the scale) becoming noticeably higher in the separating spaces between clusters than in the more densely populated areas. The cartograms neatly reflect this in the form of increasingly emptier inter-cluster spaces. Also, the *tail effect* becomes less obvious as most of the *tail*-located data points are shown to occupy areas of higher distortion.
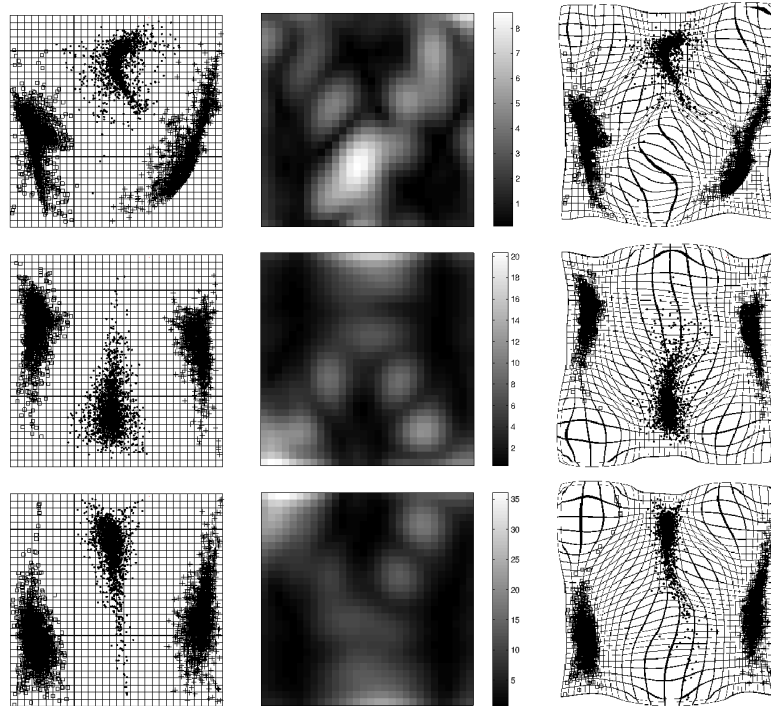
Figure 7.4: Varying the data dimension. By *column*: As in previous figure. By *row*: top) Maps for 3-dimensional data; middle) Maps for 10-dimensional data;; bottom) Maps for 20-dimensional data.

Guideline 4: *As hypothesized, the differences in dimensionality have an impact on the visualization of the data. The increasing values of the MF in inter-clusters spaces reflect that and they are consistent with the empty space phenomenon. The cartogram-based visualization is suitable for the representation of high dimensional data modeled by NLDR methods, as it will factor in the large distortions that are likely to appear in the visual representation space. Importantly, the increase in distortion resulting by the high-dimensionality itself is visually discounted by the cartogram, so that, when interpreting it, the analyst can trust that a large visual distortion is not a byproduct of the high-dimensionality of data, but the result of intrinsic cluster separation.*

Number of points per Gaussian cluster (exp3): According to the results reported in *Figure 7.5*, the increase of p.p.G. does not qualitatively alter the visual representation in any significant manner. More or less the same latent points take responsibility for an increasing number of data points each. If anything, the cluster profiles become more clearly delineated as the number of points increases. The cartogram representation is, again, quite straightforward, and separates the cluster better by reintroducing the distortion on the map. The direct display of the MF suffers from the same problem as in the previous experiment, as the *rope*-like artifacts make it difficult to delineate clear cluster boundaries.

Guideline 5: *The density of data points in the existing clusters appears to have little impact in the quality of the mapping, the level of distortion and, thus, the cartogram representation. The latter should therefore be interpreted in a similar manner regardless the number of points in the modeled dataset.*
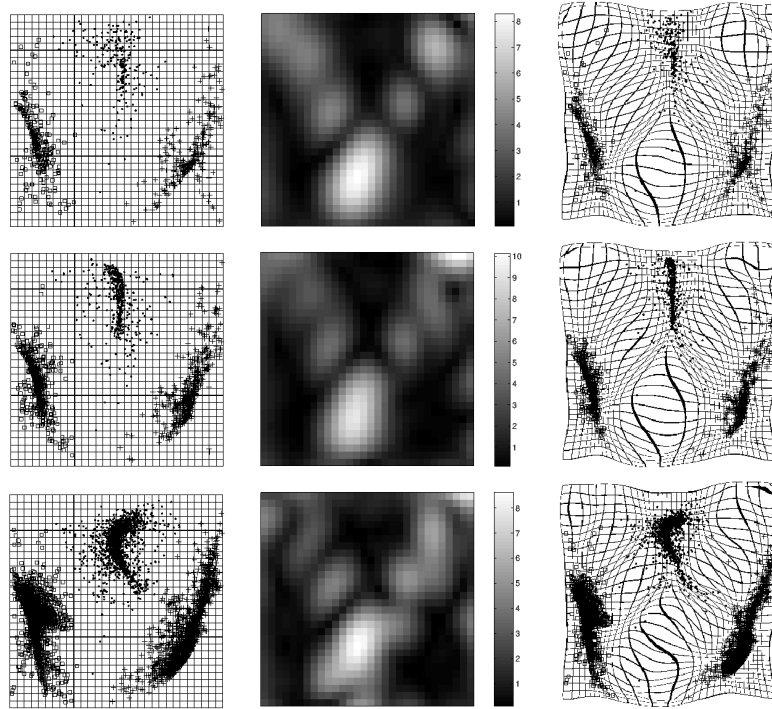
Figure 7.5: Varying the number of points per cluster. By *column*: As in previous figures. By *row*: top row) Visualizations for 250 p.p.G.; middle row) Visualizations for 500 p.p.G.; bottom row) Visualizations for 2000 p.p.G.

Relative density of the Gaussian clusters (exp4): Three 3-D Gaussians of 250 points each and of different variances (in two experiments: the first, with $\sigma_1 = 0.2$, $\sigma_2 = 0.1$, $\sigma_3 = 0.3$; the second, with $\sigma_1 = 0.2$, $\sigma_2 = 0.1$, $\sigma_3 = 0.05$) can be visualized in *Figure 7.6*. Their different compactness is captured by the GTM, as seen in the maps of the left hand column. The MF, by itself, struggles to capture the three-cluster structure, specially in the second experiment (bottom row): Notice the very different MF scale of both experiments. The cartogram representations, instead, capture the diversity of compactness and represent it in such a way that the compression of the most compact cluster and the comparative sparseness of the less compact one are self-evident.
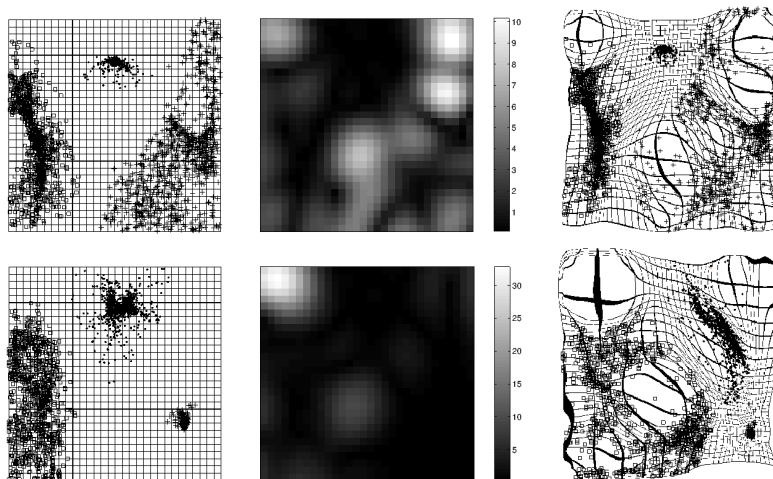
Figure 7.6: Differing levels of cluster compactness. By *column*: As in previous figures. By *row*: top) Gaussians with $\sigma_1 = 0.2$, $\sigma_2 = 0.1$, and $\sigma_3 = 0.3$; bottom) Gaussians with $\sigma_1 = 0.2$, $\sigma_2 = 0.1$, and $\sigma_3 = 0.05$.

The cartograms also reveal that the GTM model struggles to represent the homogeneity of the most sparse cluster.

Guideline 6: *Many real datasets showing grouping structure are likely to consist of clusters of different relative density of points. In most NLDR models for visualization, this should entail different levels of mapping distortion for each cluster. Even though, the emptier spaces between clusters should undergo an even higher distortion. This poses the challenge of selecting a threshold of distortion to differentiate the frontier between an inter-cluster space and a cluster of low density. The visualization of the MF might be an insufficient exploratory tool to face it and cartograms should help by allowing a clearer visualization of cluster separation. Data groups displayed in the cartogram in a roughly continuous but inhomogeneous manner should be further inspected to find possible sub-structure.*

Number of clusters (exp5): In this experiment, we wanted to investigate the effects of a *cluster-crowded* representation space on the MF-based cartogram visualization. The visualization results for an increasing number of clusters, from three to twelve, are shown in *Figure 7.7*. The clusters are correctly separated by GTM throughout the experiments, although, as we might come to expect, the visualization space becomes increasingly crowded and its interpretation increasingly challenging. The MF is again only partially useful to delimit the existing clusters.

Figure 7.7: Varying the number of clusters in the dataset. By *column*: As in previous figures. By *row*: top row) Visualizations for 3 clusters; 2nd row) Visualizations for 6 clusters; 3rd row) Visualizations for 9 clusters; bottom row) Visualizations for 12 clusters.

An interesting effect can be observed as the number of clusters increases: instead of having high-distortion areas surrounding low-distortion ones (clusters), clusters seem to *wrap around* wide empty spaces. This is nicely captured by the cartogram representations.
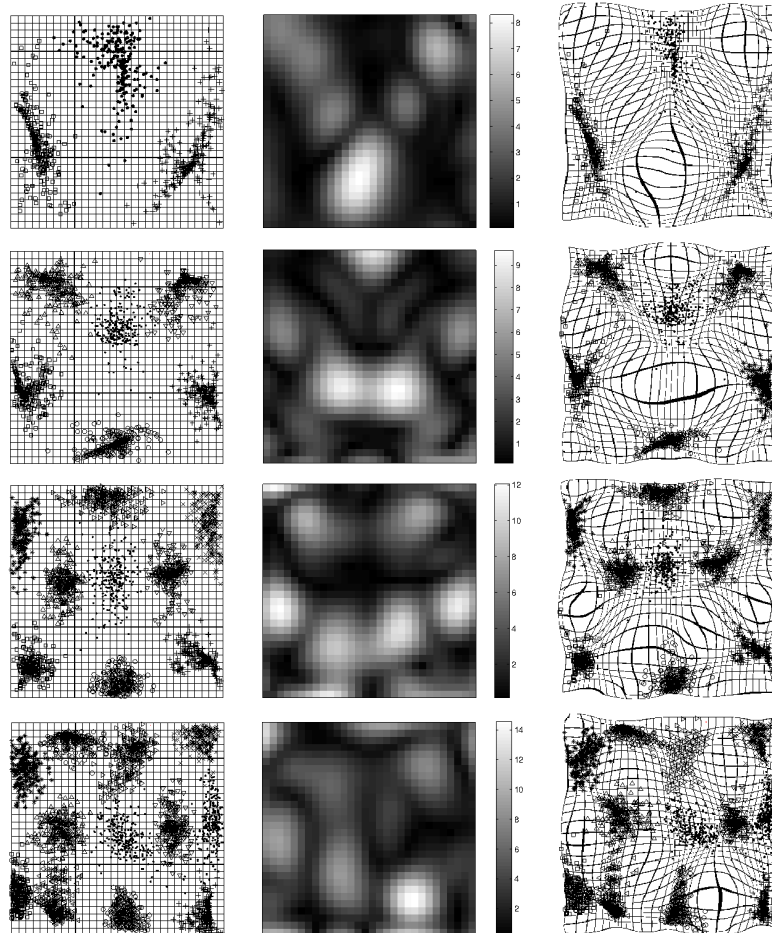
Guideline 7: *The GTM pattern of compression and stretching becomes more complex as the number of clusters increases and, at some point, it fails to reflect some of the boundaries between clusters and, thus, does not fully reflect the rich cluster structure of the experimental data. This is likely to happen in many other NLDR methods as well. In experiments in which rich cluster structure is expected, the analyst should interpret the highly-distorted areas of the cartogram with caution, as they might miss some of the richness of cluster-substructure. In this cases, the analyst might rather use the cartogram representation as part of a hierarchical clustering setting [82, 246].*

The calculations of the *aDB* index described in *equation* (7.4) for the complete set of experiments in this section are compiled in *Table 7.1*. For all variants of experiments *exp*1, *exp*2, *exp*3, and *exp*5, the *aDB* is consistently lower for the cartogram representation than for the standard posterior mean projection representation, as expected. This is a clear indication that the cartograms reflect the compact and separate nature of the observed data clusters better than the alternative representation. The exception to these results, again as hypothesized, is *exp*4. In this case, the *aDB* is bigger for the cartograms. This implies that the cartograms are reflecting the heterogeneous density of the clusters more clearly than the alternative representation, which was the goal of this experiment.

	exp1a	exp1b	exp1c	exp2a	exp2b	exp2c	exp3a	exp3b	exp3c
Cart	0.294	0.289	0.232	0.264	0.280	0.272	0.361	0.349	0.242
PMP+MF	0.359	0.349	0.334	0.383	0.346	0.365	0.486	0.437	0.342

	exp4a	exp4b	exp5a	exp5b	exp5c	exp5d
Cart	0.630	0.683	0.361	0.322	0.385	0.534
PMP+MF	0.556	0.384	0.486	0.406	0.496	0.609

Table 7.1: Values of the *aDB* index for the cartogram-based (Cart) and posterior mean projection-based with MF (PMP+MF) representations. The *exp*1 (GTM architecture) experiments are: a) 10×10 grid b) 20×20 c) 30×30. The *exp*2 (data dimensionality) experiments are: a) 3-D data b) 10-D c) 20-D. The *exp*3 (number of p.p.G.) experiments are: a) 250 points b) 500 points c) 2,000 points. The *exp*4 (relative cluster density) experiments are: a) $\sigma_1 = 0.2$, $\sigma_2 = 0.1$ and $\sigma_3 = 0.3$; b) $\sigma_1 = 0.2$, $\sigma_2 = 0.1$ and $\sigma_3 = 0.05$. The *exp*5 (number of clusters) experiments are: a) 3 clusters, b) 6 clusters, c) 9 clusters, d) 12 clusters.

7.2.2 Experiments with real data

Once the main capabilities and limitations of the cartogram visualization of MVD for NLDR models have been investigated with artificial data, we now proceed to illustrate the method using real data stemming from a neuro oncology problem. It involves the discrimination of human brain tumour types, a problem for which knowledge discovery techniques in general [168], and data visualization in particular [56] can become useful tools.

Neuro oncology data

The available data are single-voxel (spatially localized) proton magnetic resonance spectroscopy (SV-^{1}H-MRS) cases acquired *in vivo* from brain tumor patients. They are part of the multi-center, international web-accessible INTERPRET project database [135]. A total of eight clinical centers from five countries contributed cases to this database.

The spectra provide a metabolic signature of the brain tissue (be it tumour or healthy), as certain metabolites are known to be reflected by resonances at certain frequencies or bands of frequency. The analyzed data were acquired at long echo time (LET). The echo time is an influential parameter in ^{1}H-MRS data acquisition. The use of LET yields relevant information on fewer metabolites, but with clearly resolved amplitude peaks and little baseline distortion, resulting in a more readable spectrum.

The data include 78 glioblastomas, 31 metastases (these are high-grade, malignant tumours of poor prognosis; importantly in our experiments, both of these pathologies are know to be heterogeneous as expressed by their SV-[1]H-MR), 15 normal (healthy) tissue cases (which should have a very homogeneous SV-[1]H-MR signature), and 8 abscesses (abnormal masses that may or may not be a byproduct of tumours, but which are often distinct from the tumours themselves). Clinically-relevant regions of the spectra were sampled to obtain 195 frequency intensity values (data features), spanning approximately from 4.22 down to 0.49 ppm (parts per million) in the frequency range.

Two problems were investigated:

- Glioblastomas *vs.* normal tissue: Both types of brain tissue should be well separated (as they both differ radically in their metabolic composition), but their visualization should reveal that, while normal tissue forms a compact group, glioblastomas lack homogeneity. Atypical cases might be expected [269].

- Metastases *vs.* normal tissue and abscesses: These types should also be separated due to their different metabolic composition. As previously mentioned, normal tissue forms a compact group and most abscesses should be similar. On the contrary, metastases should not show much homogeneity. Some atypical cases might again be expected [269].

Results and discussion

The results for the first of the problems are displayed in *Figure 7.8*. Glioblastomas and normal tissue do indeed occupy separate areas of the visualization map, as seen in *Figure 7.8* (left). Nevertheless, this GTM projection does not clearly suggest clusters of different relative density. The difference in relative density only begins to be revealed by the accompanying MF map on which the data projection is overlaid. There is a clearly distorted space between both groups, which could be reasonably well inferred by this map alone. The corresponding cartogram (as seen in *Figure 7.8* (right)), though, is visually more informative for at least four reasons: the separating space looks increased by the reintroduced distortion; the heterogeneity of glioblastomas is highlighted; some subgroups are revealed (separation between cases mapped in the top-left and top-right areas of the map and neater separation of a few cases on the bottom-left part of the map); and some spectra are more clearly mapped into highly distorted areas of the map. According to *guideline 6*, it might be worth exploring the cluster sub-structure of the less homogeneous data using a hierarchical approach.

Each of these possibilities might merit further investigation on its own, in order to find biomedical explanations of clinical interest, but most of it is beyond the scope of this Doctoral Thesis. Let us focus instead on those spectra that are mapped into highly distorted areas of the cartogram, investigating them according to *guidelines 1* and *2* from the previous experiments with artificial data.

The mean value of the O_n measure for the *glioblastomas* is 9,523.77 ($\pm$ 5,813.08 standard deviation); the corresponding mean and standard deviation for the *normal* tissue is far lower: 7,014.26 ($\pm$ 4,310.08). The MF conforms to the same pattern, with values of 2,094.77 ($\pm$ 1,314.94) for glioblastomas and 1,589.61 ($\pm$ 551.94) for normal tissue. These values confirm that the normal tissue is characterized by comparatively homogeneous spectra, whereas glioblastomas, as described by MRS, are a heterogeneous pathology.

According to *guideline 2*, we would expect cases with comparatively high values of both O_n and MF to be definite outliers, occupying highly distorted areas of the cartogram representation. To illustrate this, we select the four spectra with highest MF values: they are individually identified in *Figure 7.8* and their MF values are 1: 6,086.31, 2: 5,313.23, and 3,4: 4,870.00. Their O_n are, in turn, 1: 27,276.11, 2: 26,411.79, 3: 19,962.87 and 4: 8,944.44. With the exception of the latter case, these values are far higher than the mean for their type, corroborating that they truly are outliers. Case 4 has a lower than average O_n, which,

according to *guideline 1*, might mean that it is an "outlier in disguise", whose low O_n is due to its proximity to the manifold.
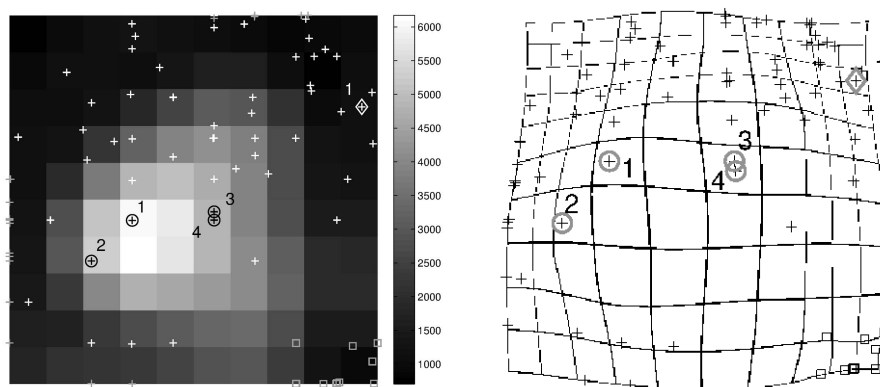

Figure 7.8: Cartogram visualization for the first of the experiments with tumour data (glioblastomas, represented as crosses, *vs.* normal tissue, represented as squares). *Left*: GTM visualization map using the posterior mean data projection overlaid on the MF, color-coded over the GTM visualization map, with scale column; four cases with highest MF are shown encircled, while the case with highest O_n is inscribed in a rhombus: their MF and O_n values are detailed in the main text. *Right*: Cartogram representation, with cases of highest MF and O_n again highlighted.

The actual spectra of these four cases are displayed in *Figure 7.9*, with the median of the spectra of their class superimposed. They are all clearly atypical, including the case with comparatively low O_n, and can be interpreted [97] as follows:

- *Cases with high MF and O_n*: All cases (1, 2 and 3) show an alternative inverted (negative) *alanine* peak *ca.* 1.46 ppm. *Case 1* also shows anomalously high amplitudes from 4 down to 3.3 ppm and absence of *choline* (*ca.* 3.18 ppm) and *creatine* (*ca.* 3.03 ppm) peaks. *Case 2* also shows extremely high *lipids* resonance values (*ca.* 1.3 ppm). *Case 3* shows displaced *lipid* resonances towards the lowest part of the frequency range.

- *Case with high MF and low O_n*: *Case 4* shows an uncharacteristically flat pattern in which barely any metabolite resonance is discernible (with the exception of a weak *choline* resonance (*ca.* 3.18 ppm)).

We now turn to identify the case with highest O_n, valued 29,184.97 (see, again, *Figure 7.8*, inscribed in a rhombus). The display of its spectrum superimposed to the median of its type (glioblastoma) confirms its atypicality (see *Figure 7.9* (bottom row)): it shows an uncharacteristic high ppm range (very negative peaks *ca.* 3.6 and 3.3 ppm) and almost flat profile devoid of resonances from 3 ppm downwards. Interestingly, though, its associated MF is rather low: 1,167.50 (well below the mean for the type) and the case is thus mapped into low distortion area of the cartogram. According to *guideline 1*, this case is likely to be located in such a position with respect to some of the data clusters and the GTM manifold that, even if away from both, still forces it to be mapped onto a low distortion area.

The results for the second problem are displayed in *Figure 7.10*. Metastases, normal tissue, and abscesses do again, with few exceptions, occupy separate areas of the GTM visualization map of *Figure 7.10* (left). This time, the standard GTM projection suggests that metastases have a much lower density than normal tissue. Unfortunately, the accompanying MF map does not suggest a clear three-cluster solution. In fact, this is an instantiation of a phenomenon we have already witnessed in the analysis of artificial data (*Relative density of the Gaussian clusters*). The very different relative cluster densities make it difficult to use the MF as a cluster-separation criterion. Instead, the corresponding cartogram (see *Figure 7.10*, right) becomes once again more informative, as it separates the three types of spectra in a visually intuitive way.

Figure 7.9: Individual spectra (in black) of several cases of high MF or O_n, as described in the text, displayed together with the median spectrum of *glioblastomas* (in gray). They are depicted within a common amplitude range (vertical axis) to ease their comparison. Frequency range measured in ppm, from 4.22 down to 0.49. *Top row*) cases 1 (left) and 2 (right), both of high MF and O_n; *centre row*) case 3 (left) of high MF and O_n and 4 (right), of high MF and low O_n; *bottom row*) case of high O_n and low MF.

We will again focus on spectra mapped into areas of high distortion. The mean value of the O_n measure for the *metastases* is 14,459.27 ($\pm$ 8,806.23); the corresponding mean and standard deviation for the *abscesses* is higher: 25,371.64 ($\pm$ 16,035.91). Again, it is far lower for the *normal* tissue: 8,954.25 ($\pm$ 5,322.97). The MF roughly conforms to a similar pattern, with values of 2,839.54 ($\pm$ 1,455.22) for *metastases*, 2,659.09 ($\pm$ 1,090.47) for the *abscesses*, and 892.64 ($\pm$ 400.74) for *normal* tissue. These values confirm that the normal tissue is characterized by homogeneous spectra, whereas metastases and abscesses, as described by MRS, are much more heterogeneous pathologies.
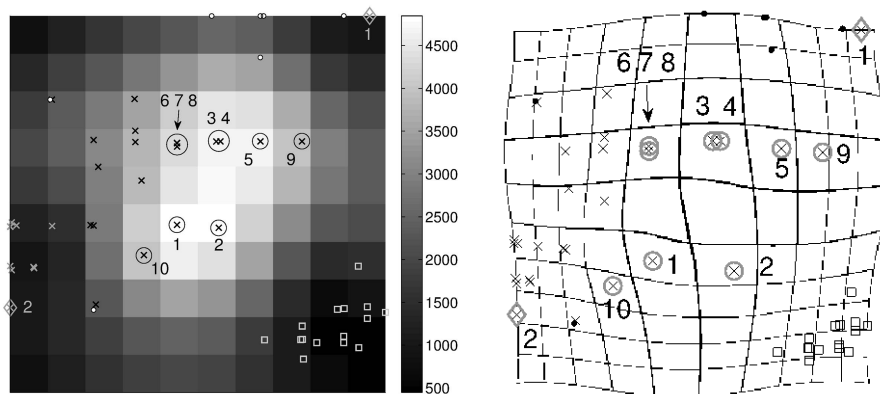
Figure 7.10: Cartogram visualization for the second of the experiments with tumour data (metastases, represented as crosses, *vs.* abscesses, represented as black dots, and normal tissue, represented as squares). *Left*: GTM visualization map using the posterior mean data projection overlaid on the MF, color-coded over the GTM visualization map, with scale column; ten cases with highest MF are shown encircled, while the two cases with highest O_n are inscribed in a rhombus: their MF and O_n values are detailed in the main text. *Right*: Cartogram representation, with cases of highest MF and O_n again highlighted.

This time, we select the ten highest values of MF. They are all metastases mapped into the central area of the GTM map, and their MF values range from 3,958.40 to 4,858.36. Only half of them have corresponding O_n values well over the mean. The values of the rest, according to *guideline 1*, could again be explained by its proximity to the manifold while being located away from data clusters. For illustration, a couple of examples of both are displayed in *Figure 7.11*: they are cases 3 and 6 (both high MF and O_n) on the top row, and cases 4 and 5 (high MF and low O_n) on the center row. They can again be interpreted as follows:

- *Cases with high MF and O_n*: Both *cases 3 and 6* show an extremely high amplitude *ca.* 3.18 ppm (*choline*) and an extremely low one *ca.* 2 ppm (*NAcetyl Aspartate, NAA*). Moreover, *Case 3* seems to be affected by a low baseline artifact and an inverted peak *ca.* 3.6 ppm.

- *Cases with high MF and low O_n*: Both *cases 4 and 5* present an alternative inverted (negative) *alanine* peak *ca.* 1.46 ppm and very high amplitude *ca.* 3.18 ppm (*choline*). *Case 5* seems to be further characterized by near absence of *NAA* (*ca.* 2 ppm) and *lipids* (*ca.* 1.3 ppm) signals.

We now identify (see, again, *Figure 7.10*, inscribed in a rhombus) the two cases with highest O_n, valued 39,275.20 (case 1) and 35,030.40 (case 2). The display of the spectra superimposed (*Figure 7.11* (bottom row)) to the median of their type (metastases) confirms their atypicality. For *case 1*, most signal in the higher range of frequency (down to 2 ppm) seems affected by acquisition noise artifacts. Furthermore, the *NAA* peak (*ca.* 2 ppm) seems displaced, there is an extreme *alanine* negative peak *ca.* 1.46 ppm, and an almost complete absence of lipid signal (*ca.* 1.3 ppm). For *case 2*, most signal in the higher range of frequency (down to 2 ppm) shows uncharacteristically diminished amplitudes. There is almost near absence of *NAA* peak (*ca.* 2 ppm), an extreme lipid resonance (*ca.* 1.3 ppm), and an unusual inverted peak *ca.* 1.7 ppm.
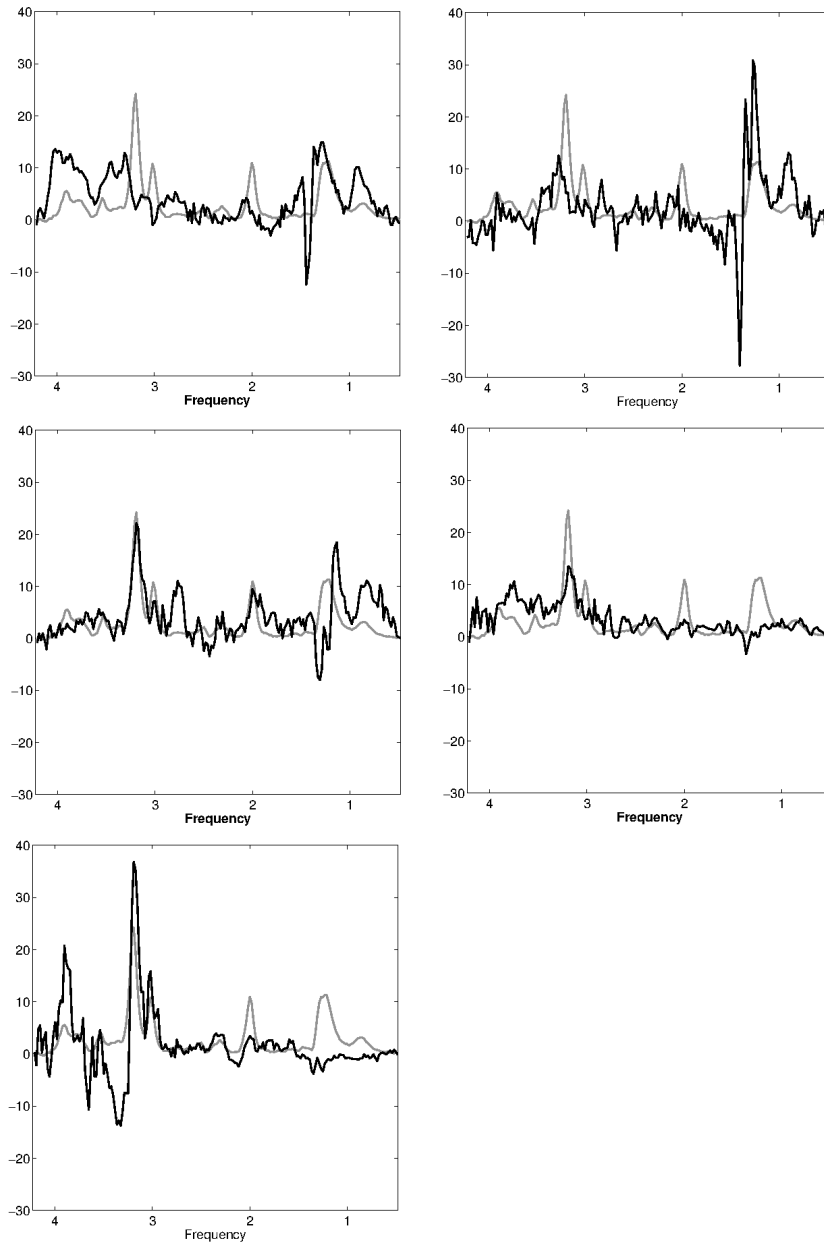
120
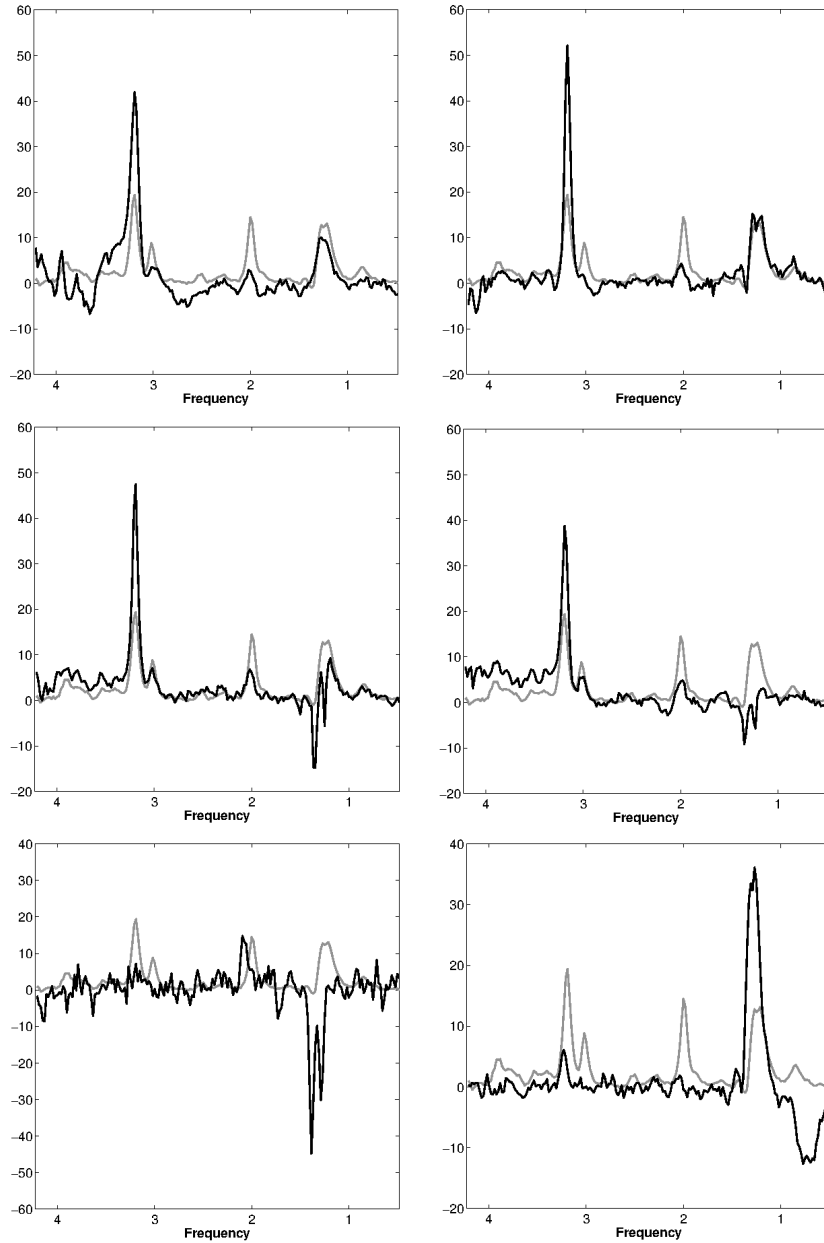
Figure 7.11: Individual spectra (in black) of several cases of high MF or O_n, as described in the text, displayed together with the median spectrum of *glioblastomas* (in gray). Notice that they are not depicted within a common amplitude range (vertical axis). Frequency range measured in ppm, from 4.22 down to 0.49. *Top row*) cases 3 (left) and 6 (right), both of high MF and O_n; *centre row*) cases 4 (left) and 5 (right), of high MF and low O_n; *bottom row*) cases 1 (left) and 2 (right) of high O_n and low MF.

Their associated MF values are low: in turn, 952.59 and 1,040.42 (well below the mean for the type) and the cases are thus mapped into low distortion areas of the cartogram. As for the previous problem, and according to *guideline 1*, these cases could be located in a position away from some of the data clusters and the GTM manifold that still forces them onto a low distortion area.

User testing may be a relevant step for the development of real applications of information visualization. In order to assess the capabilities and limitations of the proposed cartogram-based visualization as applied to the neuro-oncology problem described in this section, we conducted a necessarily limited user study. This study involved 14 participants from 3 different universities in Spain, 2 in United Kingdom and

1 in Italy and Colombia. They all had at least some experience in the quantitative analysis of biomedical data (many of them, specifically in the analysis of MRS), and some of them had hands-on experience in oncologic radiology. All participants had at least some knowledge on visualization-oriented data modeling as applied to biomedical problems, and some of them had extensive experience on the use of NLDR techniques. The opinion of this sample of users was thus considered to be sufficiently qualified.

All subjects replied individually to a brief questionnaire that contained five separate questions that had to be answered as Likert-type items [161, 180]. The participants were provided with the corresponding relevant visualizations and with the following text:

"To the best of your knowledge, you have to answer the following five questions according to an agreement level scale in which:

- A value of 1 corresponds to "strongly disagree".

- A value of 2 corresponds to "disagree".

- A value of 3 corresponds to "neither agree nor disagree".

- A value of 4 corresponds to "agree".

- A value of 5 corresponds to "strongly agree".

The questions refer to the visualization results summarized in the images shown in *Figure 7.8* and *Figure 7.10*.

- **Q.1**: The *cartogram visualization* (CV) technique singles out and isolates the possible atypical data or outliers in the data set better than the *posterior mean projection* technique together with the *magnification factors* (PMP+MF).

- **Q.2**: The PMP+MF technique informs the brain tumour MRS data cluster structure better in visual terms than the CV.

- **Q.3**: The CV technique does visually reveal the distinction between different brain tumour pathologies better than the PMP+MF technique.

- **Q.4**: The PMP+MF technique does visually reveal the heterogeneity of some brain tumour pathologies better than the CV technique.

- **Q.5**: The CV technique is better than the PMP+MF technique at helping to understand the nonlinear distortion introduced by the model in its mapping of the data."

The results are summarized in *Table 7.2*.

	♯ score occurrences						
	1	**2**	**3**	**4**	**5**	**mean**	**median**
Q1	1	1	2	8	2	3.64	4
Q2	7	6	1	0	0	1.57	1.5
Q3	0	2	2	9	1	3.64	4
Q4	3	4	6	1	0	2.36	2.5
Q5	0	0	2	3	9	4.50	5

Table 7.2: Number of occurrences of each score (from 1, strongly disagree, to 5, strongly agree, in columns) for each of the five questions (from question 1, Q1, to Q5, in rows) of the user survey. In questions 1, 3 and 5, a 5 score indicated strong agreement in favour of the cartogram-based method. In questions 2 and 4, a 5 score indicated a strong agreement in favour of the alternative method. See appendix for details on the questionnaire. Mean and median values of the scores for each question are reported in the last two columns.

The replies to these questions were mostly favorable to the use of the cartogram-based method, according to the mean and median scores. This method seems to be especially appreciated for helping to

understand the nonlinear distortion introduced by the NLDR method and, to a lesser extent, for informing the MRS data cluster structure better than the alternative method, as well as for its capability to reveal the distinction between tumour pathologies and to isolate atypical data. The comparison between both methods is at its least conclusive when it comes to their relative ability to reveal the internal heterogeneity of some of the pathologies. Overall, the results of this user survey support the adequacy of cartogram-based visualization in a very challenging real problem.

7.3 Conclusions

The visualization of MVD can provide us with inductive reasoning insights that could be difficult to gain from direct deductive reasoning from the raw data. Linear projection methods are now commonplace for performing this task. They are easy to interpret, even though the faithfulness of their representation is limited. Over the last decade, nonlinear dimensionality reduction methods for MVD visualization [155] have provided novel and sophisticated approaches to this problem. Their adoption is hindered, though, by the difficulty of interpreting the visualizations they provide in terms of the original data attributes and also by the non-uniform distortion they generate. This non-uniform distortion or local magnification may be a key impediment for the interpretation of the visualization maps these methods yield [105].

In this thesis, we have adapted the cartogram technique, originally defined for the distortion of geographic maps according to underlying attributes, to MVD visualization using NLDR models. Its use has been illustrated with an NLDR model of the manifold learning family, GTM. The proposed density-equalizing cartogram representation of the GTM visualization maps allows explicitly reintroducing the mapping distortion created by the model, thus providing more faithful data visualizations. The capabilities and limitations of the proposed technique have been assessed through a battery of experiments with both artificial and real data, from which several guidelines of use of practical interest have been extracted.

Although the cartogram representation has been illustrated using GTM, this technique could easily be extended to other nonlinear dimensionality reduction methods, provided the mapping distortion could be calculated or, at least, approximated. The latter is the case, for instance, of the well-known SOM, using the U-Matrix [252] or the U-Maps for Emergent SOM [253] as approximate distortion measures. An explicit measure of MF can be calculated, though, for the batch version of SOM [25]. Preliminary experiments for the cartogram representation of the MF and the U-Matrix in batch-SOM have been carried out in [249].

Taking advantage of the vector quantization nature of GTM, we have used its lattice of latent points to establish the limiting borders of the distortion regions. This is akin to using a centroidal Voronoi tesselation [70] of the latent space. But this is by no means the only possible approach to border definition for the generation of cartograms. In fact, Voronoi diagrams (also know as Voronoi tessellations or decompositions [193]) of the visualization space, based on more or less compact data representations based on their *posterior mean projections*, could also be used to the purpose of creating cartograms. As remarked by Rong and colleagues [219], Voronoi diagrams are widely used in computational science and engineering. This approach would open the application of cartogram techniques to nonlinear methods that do not provide vector quantization.

The extension of the proposed technique to growing architectures of SOM [2], or to related methods such as Neural Gas [278] should be straightforward. Cartograms could also be used as a visual guide for interactive hierarchical models for MVD clustering and visualization [24, 217, 246], for which different levels of the hierarchy could be semi-automatically controlled, allowing user interaction, according to levels of mapping distortion.

Chapter 8

Distortion and Flow Maps visualization for churn analysis

As discussed in the previous chapters of this thesis, in the current global situation of economical crisis, competition becomes fierce, especially in deregulated or loosely regulated markets. Customer management thus becomes a key to gain competitive advantage and avoiding customer defection and ensuring the retention of the most valuable customers should become a central managerial preoccupation.

Our research investigates the churn phenomenon mostly from the point of view of exploratory Data Mining, emphasizing methods that provide simultaneous MVD clustering and visualization. Previous chapters have dealt with the problem of market segmentation using the GTM statistical machine learning technique.

The previous chapter, in particular, has provided a method, based on geographical information representation, to address the difficult problem of improving the interpretability of the low-dimensional visualization of MVD when the mapping from the original observed high-dimensional data space is nonlinear in nature. This method, the Cartogram, has been shown to retain the interpretability of the maps while distorting them, but always retaining the continuity of the map internal and external borders. It has been extrapolated from geographical maps to the GTM visualization maps (although the method is by no means restricted to GTM), replacing geography-related quantities by quantities reflecting the mapping distortion introduced by GTM, which is explicitly quantifiable.

In this chapter, we suggest the combination of Cartograms with a second method of MVD visualization, also inspired in geographical information representation: The Flow Map. Flow Maps were originally devised to visualize geography-related evolution patterns such as, for instance, population migrations [234] and have become increasingly sophisticated from a computational viewpoint.

The standard GTM, including the representation of its mapping distortion using Cartograms, provides us with a static snapshot of the current market segments. But the fact is that markets evolve, slower or faster, over time. Any instance of intelligent customer management should investigate customer evolution over time, trying to prevent individual customers drifting towards churn-risk areas. This time-dependent component should allow the service provider to design and launch customer retention actions oriented towards the retention of the most profitable customers.

Given that the analyzed databases contain information over time, we use Flow Maps to analyze the *customer migrations* over the GTM visualization map, aiming to detect *foci* of potential customer churn. As reflected in the experiments reported in this Thesis, the use of both methods helps increasing the interpretability of the visualization of the analyzed database, thus assisting in the process of useful knowledge extraction that could have a practical impact on customer retention management strategies.

Two databases were analyzed: one corresponding to telephone customers from a Brazilian telecommunications company, and another corresponding to customers of an Spanish pay-per-view television service provider.

8.1 Methods and Materials

8.1.1 Flow Maps for the visualization of customer migrations in GTM

Flow Maps are usually combinations of geographical maps and flow graphs that were originally devised to visualize evolution patterns such as population migrations. Again[1] we propose their use in NLDR-based visualization to display the evolution over time of individual points, in this chapter with GTM. This type of visualization can be specially suitable for tracking the behavioural evolution of individual customers, anticipating the possibility and potential cost of their defection.

A method for the generation of Flow Maps using hierarchical clustering was recently proposed in [209]. In brief, its algorithm operates through six differentiated stages. These stages, as applied to the GTM representation, are as follows:

- 1) *Layout adjustment*, enforcing a minimum separation distance among the nodes (in our case, each of the squares in the GTM lattice corresponding to individual latent points in the visualization space);

- 2) *Primary clustering*: merging of flow edges that share destinations, obtained by agglomerative hierarchical clustering. The resulting binary tree describes the branching structure of the Flow Map;

- 3) *Rooted clustering*, generated such that the root of the Flow Map is the root of the tree;

- 4) *Spatial layout*, which actually defines the flow hierarchical tree from the rooted hierarchical cluster solution;

- 5) *Edge routing*, in which edges are re-routed around the bounding boxes within the same hierarchical cluster to avoid unwanted crosses;

- 6) *Rendering*, in which each flow edge in the visualization map of GTM is rendered as a catmull-rom spline, generating an interpolation between the nodes of the spatial layout hierarchical tree. Their width is proportional to the magnitude of the flow.

8.1.2 Brazilian telecommunication company

For the first set of experiments, a proprietary database containing telephone usage information from customers of the main landline telephony telecommunications company in São Paulo (Brazil) was used. The database has previously been used in the *Chapter 6* of this Thesis and *Section 6.2* provides a detailed description of its data characteristics and features.

8.1.3 Spanish pay-per-view television company

For the second set of experiments, a proprietary database belonging to a Spanish pay-per-view television company was used. It includes monthly data from 33,992 customers, monitored for churn over 7 months, from March to September 2008. Their behaviour is described through 59 variables corresponding to channel usage (36 variables, see *Table 8.1*) and customer-company interaction (23 variables, see *Table 8.2*), including post-sale, customer, and technical service; complaints and billing.

---

[1] As with Cartograms (see *Chapter 7*).

v1	Operator's own channel	v19	Documentaries pack
v2	Operator's second channel	v20	40TV channel (Music)
v3	Operator's own channel with a 30 minutes delay	v21	Music pack
v4	Basic pack	v22	Traveling channel
v5	Movies pack 1	v23	Barça channel (Football team own channel)
v6	Movies pack 2	v24	Classical music channel
v7	Movies pack 3	v25	Playin TV option (basic games)
v8	Classic movies pack	v26	Hunting channel
v9	Series pack	v27	Playboy channel
v10	Cosmopolitan channel	v28	Multiroom option (watch on multiple TV sets)
v11	Odisea channel	v29	Golf channel
v12	Cooking channel	v30	PVR (Power Video Recorder) option
v13	Football channel	v31	Bullfight subscription
v14	Golf channel	v32	Football subscription
v15	Sports pack	v33	PPV Movies
v16	Eurosport channel	v34	PPV Pressing catch
v17	Kids pack 1	v35	PPV Football
v18	Kids pack 2	v36	PPV Adult movies

Table 8.1: Data features used to describe the consumption of pay-per-view customers. All data features are binary: if the customer has contracted the corresponding channel, pack, option or subscription, or has used the described PPV at least one time, its value is 1; otherwise, it's 0. Packs refer to groups of content-related channels, options refer to added-value services, subscriptions allow 1-month channel availability and PPV refers to individual payment for specific programs.

v37	Calls for technical reasons, simple resolution	v49	2nd level customer service
v38	Calls for technical reasons, complex resolution	v50	1st level retention calls
v39	Technical interventions paid by the company	v51	2nd level retention calls
v40	Technical interventions paid by the customer	v52	Calls attended at 3rd level of retention
v41	Changes of address, paid by the company	v53	3rd level retention calls
v42	Technical complaints	v54	*Last chance* retention calls
v43	Economical complaints	v55	Returned receipt due to non-payment
v44	Other complaints	v56	Number of debt condonations
v45	Complaints in process	v57	Bad-debt recovering actions
v46	Transferred calls to retention unit 1	v58	Calls asking for 100% of debt condonation
v47	Transferred calls to retention unit 2	v59	Calls asking for 50% of debt condonation
v48	1st level customer service		

Table 8.2: Data features used to describe customer-company interaction of customers. All data features are binary: 1 if the described event has happened one or more times, 0 if it never happened.

8.2 Experiments

Our approach to the exploratory visualization of the available databases relies on three basic assumptions, supported by previous preliminary research [85], that can be expressed as follows:

1. Different customer service usage patterns determine different levels of churn propensity.

2. The identification of customer migration routes between two consecutive time periods is possible. These routes maybe either negative: towards representation space areas of lower value for the company and, eventually, churn; or positive: towards representation space areas of higher value for the company and higher customer fidelity.

3. In the absence of promotional actions, customers' usage behavior tends to remain stable. This entails lack of migration or migrations towards neighbouring areas in the visual representation space.

The visual exploratory analysis of the reported experiments aims to identify potential customer churn routes through the combination of three processes:

1. The visualization of customer usage patterns through the nonlinear mapping onto a 2-D representation space using GTM.

2. The enhancement of this visualization using Cartogram representation.

3. The visual representation of customers' transitions over periods using Flow Maps, aiming to discover potential churn and customer retention routes over the GTM visual representation map.

The experimental settings corresponding to the GTM models and the Flow Maps are first described. This is followed by a presentation and discussion of the results of the analyses of the databases.

8.2.1 Brazilian telecommunication company

8.2.1.1 Experimental Setup

As described in *Section 7.1.2*, the adaptive parameters of the GTM model were initialized according to a standard procedure described in [23]: The weight matrix $\mathbf{W}$ was defined so as to minimize the difference between the prototype vectors $\mathbf{y}_k$ and the vectors that would be generated in the observed space by a partial PCA process. The inverse variance parameter β was initialized as the inverse of the 3rd PCA eigenvalue. This initialization procedure has been shown to be reliable while avoiding the lack of replicability that might result from the random initialization of parameters.

Different GTM lattice sizes were explored but, in the end, a trade-off between detail (which would be proportional to the size of the lattice) and practical visual interpretability had to be achieved. For the analyzed data, it was found that a suitable layout was a 10×10 grid for the GTM lattice. This was chosen for all the reported experiments.

In the reported experiments, the GTM input to the Flow Map algorithm included: The GTM map layout, in the form of a regular visualization lattice built from the discrete sampling of the latent space; The GTM model for periods $P1$ and $P2$, in the form of the assignment of each data point (customer) to a given lattice node (cluster); the flow from the $P1$ to the $P2$ visual representations, in the form of cumulative customer information for each of the lattice nodes

8.2.1.2 Results

The data described in *Section 8.1.2* were first mapped into the standard GTM model. Data from period $P1$ are represented in *Figure 8.1* and data from period $P2$, in *Figure 8.2*. *Figure 8.1* and *Figure 8.2* (top-left) show all data as mapped into the 2-D GTM visualization space continuum, according to their *posterior mean projection*, which was described in *Section 4.3.3*.

The images in *Figure 8.1* and *Figure 8.2* (top-right) represent the same data over the same space, but this time using the *posterior mode projection*, so that the visualization informs of which of the 100 GTM nodes each of the data points is assigned to. The relative size of each square is proportional to the ratio of data mapped into that node. As a result, areas filled with (relatively) big squares usually correspond to areas of the mapping with high data density.

The local distortion introduced by the nonlinear mapping, as represented by the MFs described in *Section 4.3.4*, is color-coded in *Figure 8.1* and *Figure 8.2* (bottom-left), and this is again represented in the same 10×10 visualization grid. Note that this representation is the same for both periods (both figures) because we are mapping the data from the second period in the model generated by the first one. This quantification of the local mapping distortion in the form of MFs is then explicitly reintroduced in the visualization space of *posterior mean projections* through the Cartograms in *Figure 8.1* and *Figure 8.2* (bottom-right).
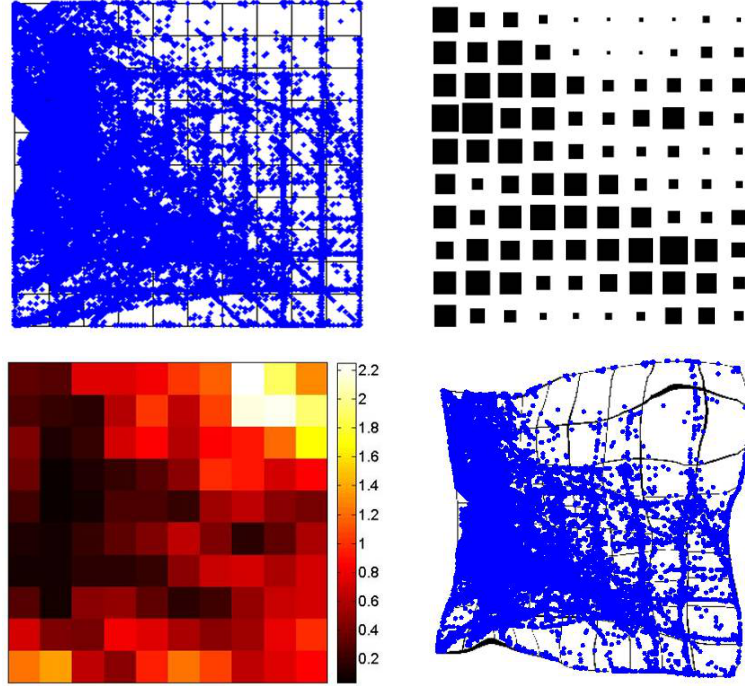
Figure 8.1: Basic MVD visualization over the GTM representation map for the data corresponding to period *P*1. Top left) *Posterior mean projection* of the data. Each dot is a customer represented over the continuum of the latent space. Top right) *Posterior mode projection* of the data. Each customer is assigned to a GTM node (represented as a square) over a discrete representation map. The relative size of each square is proportional to the ratio of customers assigned to that node to the total number of customers. Bottom left) Values of the MF for each GTM node, represented as a color map on the discrete latent space of the model. Bottom right) Cartogram representation of the *posterior mean projection* of the data in which the distortion is proportional to the MF.

Once this basic representation is established, we build on it by adding further customer profiling information. As listed in *Section 8.1.2*, this includes commercial margin, AVS on portfolio, time as a company customer, EANC code and number of employees in the customer company. This helped us to establish a market-meaningful comparison between periods *P*1 and *P*2, in order to identify map areas of commercial interest. The following quantities are visualized in the *posterior mode projection* maps of *Figure 8.3*:

1. *Percentage of churn*, defined as:

$$churn_i = (A_i/\mu_i)100$$

   where A_i is the number of customers mapped into node *i* that abandoned the company between periods *P*1 and *P*2; and μ_i is the average of customers over the two periods in that node[2]. It is visualized in *Figure 8.3* (top-left).

2. *Percentage of stable customers*, defined as

$$stab_i = (S_i/\mu_i)100$$

   where S_i is the number of customers that remained in node *i* between *P*1 and *P*2. It is visualized in *Figure 8.3* (top-right).

3. The previous quantities helped us to identify potential *departure gates* for customers and customer *strongholds*, but did not clarify their value. For that, we calculated and visualized (in *Figure 8.3*,

---

[2] This calculation of churn is common business practice.
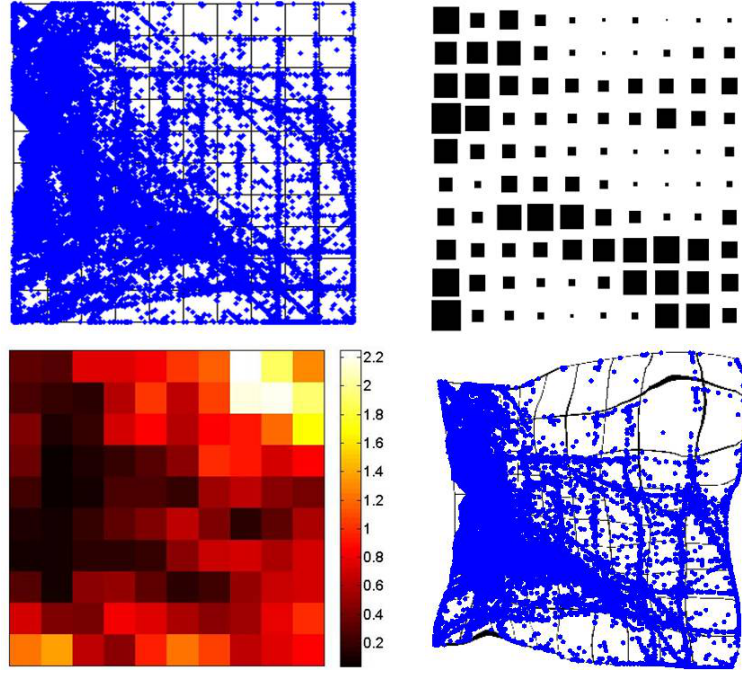
Figure 8.2: Basic MVD visualization over the GTM representation map for the data corresponding to period *P*2, as in *Figure 8.1*.
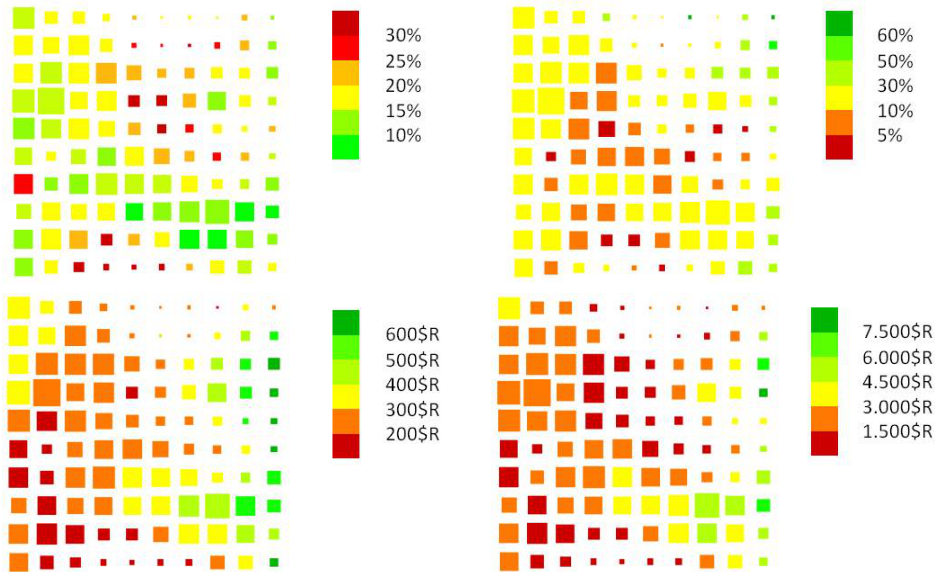
Figure 8.3: Visualization of profiling parameters over the *posterior mode projection* of the data in the GTM representation space, using color maps. Top left) Visualization of the *percentage of churn*. Top right) Visualization of the percentage of stable customers. Bottom left) Visualization of customers' commercial margin. Bottom right) Visualization of customers' LTV.

bottom-left) the *commercial margin* of each GTM node, defined as the average commercial margin of the customers mapped into it.

4. Finally, we visualized in *Figure 8.3* (bottom-right) the *life-time value* (LTV) of a GTM node *i*, cal-

culated as the commercial margin of the node divided by its percentage of churn[3].

The visualization of the percentage of churn per node without direct information of the absolute number of churners may not be intuitive enough. At this point, we suggest using the concept of Cartogram to reintroduce the absolute number of churners into the visualization space. That is, instead of distorting the GTM according to the MF as in *Figure 8.1* and *Figure 8.2* (bottom-right), we suggest distorting it directly according to the absolute number of customers abandoning the service provider company from a given node. The result can be seen in *Figure 8.4*.
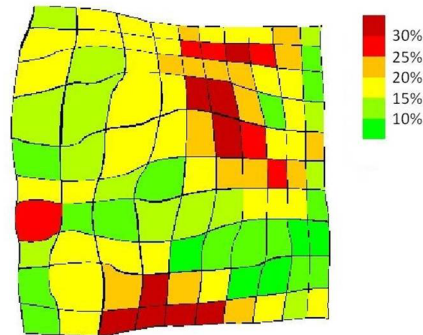
Figure 8.4: Cartogram of the *percentage of churn* of *Figure 8.3* (top-right), where the distortion is proportional to the total number of churning customers in each node.

Each GTM node or micro-cluster is not, by itself, too actionable from a marketing viewpoint. We thus further grouped these micro-clusters into market segments using the well-know K-means algorithm [126]. See details of this procedure in García et al. [85, 88]. The obtained market segments are displayed in *Figure 8.5*.
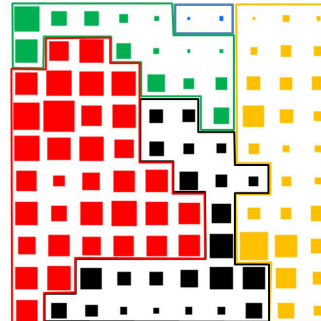
Figure 8.5: Segmentation of the analyzed customers according to a procedure that uses K-means to agglomerate the basic clustering results of GTM. The resulting five segments are color-coded: red for *Locals*, green for *Street Force*, yellow for *Nationals*, blue for *Providers*, and black for *SoHo*.

Once this overall market characterization by segments was achieved, we turned our attention to the customer base transition between periods *P*1 and *P*2. For that, we overlaid the GTM-based visualization with the migration of customers between GTM nodes, as visualized using Flow Maps. For the sake of brevity, this is illustrated in *Figure 8.6* with the migration for just a couple of GTM nodes.

---

[3]This is, again, common business practice.
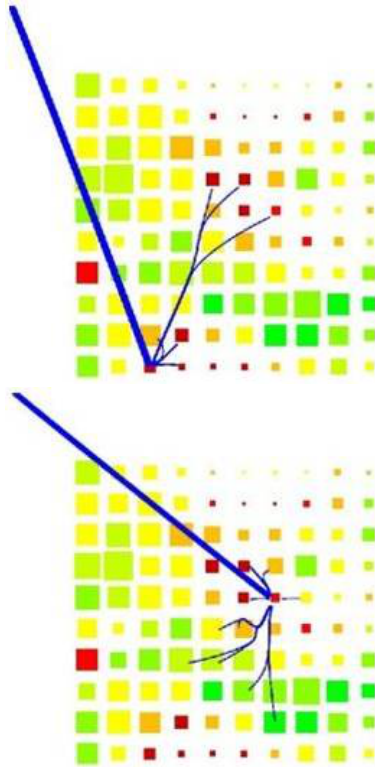
Figure 8.6: Flow Maps for two specific GTM nodes displayed on top of the *posterior mode projection* of the data in the GTM representation space, using a color map to represent *percentage of churn*. The lines moving away of the map represent the churn, whereas the lines between GTM nodes represent the migrations of the remaining customers. The width of the lines is proportional to the ratio of customers migrating to a given arrival node, to the total number of customers in the departure node. Top) a node of the *SoHo* segment in which the migration pattern reflects the failure of a commercial action. Bottom) a node from a different area of the *SoHo* segment in which the migration pattern reflects this time the success of a different commercial action.

8.2.2 Discussion

Figure 8.1 provides different visualizations of the 57,422 analyzed customers from *P*1 in their GTM representation maps. The most detailed one is the *posterior mean projection* in Figure 8.1 (top-left). The big size of the data set makes this representation rather obscure and uninformative. It reflects a common trait to be found in customer usage data, which is an apparent absence of global grouping structure and densely populated representation areas gently and gradually connected to less densely populated ones, without neat borders between them.

Given that these maps represent customer usage, it is perhaps not surprising that the main and rather indistinct data concentration corresponds to a majority of customers showing a very standard service usage, strongly mediated by outgoing local, within-state and mobile calls (which constitute the 95% of all calls).

This visual information becomes much more operational using the *posterior mode projection* map shown in Figure 8.1 (top-right), in which the relative ratios of customer assignment to each GTM node provide insights into a somehow richer cluster structure. The comparison of periods *P*1 and *P*2 in *Figure 8.1* and *Figure 8.2* is illustrative: the mean projection does not show any clear differences, whereas the mode projection at least shows that *P*2 has led to slightly more clearly differentiated groupings than *P*1.

The areas of high-data density usually undergo little distortion in the nonlinear mapping generated by GTM. This effect is clearly reflected in the MF maps of *Figure 8.1* and *Figure 8.2* (bottom left), where densely data populated areas correspond to low magnification (distortion). On the contrary, more sparsely populated areas correspond to high magnifications, suggesting the diversity of the less standard customers

(and, thus, the existence of potentially interesting market segments).

This uneven customer distribution is neatly captured by the Cartograms in *Figure 8.1* and *Figure 8.2* (bottom right), in which the data from standard customers become more concentrated than in the standard mean projection, whereas the less standard ones occupy an expanded visualization area that reflects their original diversity more faithfully.

So far, visualizations have only hinted about the general structure of the data. A richer insight can be obtained from the GTM maps of *Figure 8.3*, describing the significant local variations of *percentage of churn*, *percentage of stable customers*, *commercial margin* and *LTV*. The *percentage of churn* map in *Figure 8.3* (top-left) reveals large variations between different areas of the map, from values close to 0% to values over 30%. These results corroborate the initial hypothesis that different service usage patterns can determine the level of propensity to churn.

Three areas of high churn (dark red nodes) were identified and singled out for further investigation:

- The individual node in the first map column from the left and seventh row from the top is characterized by a very low overall service usage, consisting mostly of companies either close to liquidation for economical reasons, or that were about to replace the telephone service provider by their own mobile call center.

- The second churn area in the low part of the map, sparsely populated and occupying the center of the last two rows, consists of companies for which the reduction of mobile phone tariffs and their landline/mobile calls mix made the transition from landline to mobile specially attractive.

- The third churn area, also sparsely populated and occupying most of the central part of the top half of the map, corresponds to customers attracted by call plans offered by telecommunication companies specialized in long-distance calls.

The cartogram of the churn map distorted according to the absolute number of churners in each node, shown in *Figure 8.4*, provides complementary visualization that reveals that the third churn region described in the previous paragraph includes more churners than the others, which suggests the adequacy of a marketing action that prioritized campaigns to counter the luring effect of those carried out by companies specialized in long-distance services.

Even if the focus of this study is on the analysis of churn and on the detection of *churn gates* of customer departure, market knowledge can also be acquired from the exploration of those customers that do not vary the usage pattern over the studied periods and, thus, do not vary their location over the visualization map. *Figure 8.3* (top right) reveals that the most stable customers (in green) are located at the top and bottom right corners of the GTM map, which means that they are clearly separated from the bulk of the customer sample. These are mostly nationwide operating companies with a varied *communication mix*, that is, companies that have incoming and outgoing calls to all destinations and covering all time bands. Interestingly, telecommunications companies do not have competitive offers that match this usage pattern.

A perhaps more valuable information can be obtained from the similar, but not equivalent, commercial margin and LTV representation maps in *Figure 8.3* (bottom-left and right, respectively). The customer *departure gates*, or GTM nodes with high churn, are important *per se*, but this importance must be weighted by the commercial value of the customers assigned to them. Marketing preventive actions must prioritize big churning areas of high commercial value. Fortunately for the service provider, most of the areas of high churn had relatively low commercial and LTV value for the analyzed data.

Often, service providers require a less detailed market segmentation than the one provided, for instance, by the reported 10×10 GTM representation. The 5-segment solution resulting from the application of K-means as a post-processing of the GTM results, reported in *Figure 8.5*, can be characterized as follows:

- *Locals* (54.4%): Companies that, essentially, perform local tasks in standard working hours.

- *Nationals* (17.4%): Companies with national reach and a mix of local, national and international calls, made during standard working hours.

- *Street Force* (9.8%): Companies with mobile employees (sales force, maintenance services, messengers, etc.), with whom they mostly communicate through incoming and outgoing mobile calls.

- *SoHo* (18.3%): Self-employed workers that use their telephone line both for work-related and personal calls.

- *Providers* (0.2%): Companies with plenty of free-call customer service lines, including services and care providers, public companies, etc.

Useful market insight can be obtained by tracking customers as they evolve, from period $P1$ to period $P2$, through the five obtained segments. More than 50% of total churn had its origin in the *Locals* segment (which decreases by more than 10%, with relevant migrations towards the *SoHo* -9.06%- and *Street Force* -6.37%-, both with high levels of churn). The reason for this is the strong competition between mobile and long-distance providers for this segment. On the opposite side, the *Nationals* and *Providers* segments show the lowest mobility (75.46% and 72.63% of segment permanence, in turn), due to the difficulty for providers other than those specialized in their profiles to offer sustainable competitive plans.

Although this high-level segment vision of the market allows the practical implementation of commercial actions, it still misses the fine grain of the local migration characteristics over the GTM visualization map. This can be fully appreciated through the use of Flow Maps, as in *Figure 8.6*. One was obtained for each of the 100 nodes of the GTM map, but, for brevity, only two of them are shown in this figure to illustrate the interest of this visualization method.

The overall inspection of the Flow Maps corroborated the initial assumption that, in most cases, migrations happen between neighbouring nodes, whereas brisk jumps over distant locations in the GTM map do not abound. This reflects that the changes in customer usage patterns are, in this case, mostly gradual. This does not preclude major changes, such as, for instance, those illustrated by *Figure 8.6* (top), in which transitions are towards GTM nodes that, even if distant, share a rather high churn rate. In this particular case, the abrupt evolution was motivated by inadequate commercial actions (indiscriminate landline-to-mobile call card gifts) that artificially modified the usage profile without modifying the underlying customer behaviour and propensity to churn.

Figure 8.6 (bottom) singles out the opposite case of an adequate commercial action that took part of the customers away from churn regions. In the illustrated example, a *friend numbers* campaign allowed transferring part of the landline-to-mobile usage into landline-to-landline usage, increasing customer usage stability, commercial margin and LTV as a result.

8.2.3 Spanish pay-per-view company

The adaptive parameters of the GTM were initialized according to a standard procedure described in [23] and similar to the one used for the previous database. A 10×10 grid for the GTM lattice was used in all the experiments. The GTM input to the Flow Map algorithm includes: The GTM map layout (visualization lattice); the GTM model for the different months, in the form of the assignment of each customer to a given lattice node; and the flow from month-to-month GTM representations, in the form of cumulative customer information for each of the lattice nodes.

8.2.3.1 Results and discussion

The projection of the 33,992 customers in the GTM map is shown in *Figure 8.7*. The visualization includes the mean projection (top row, left) and the mode projection (top row, right) of the March 2008 data. They are accompanied by a visualization of the MF as a colour map, with white indicating highest MF and black, lowest MF (middle row, left) and the corresponding cartogram of the MF with the mean projection overlaid onto it (middle row, right).

Beyond the distribution of individual customers, we are interested in their commercial TV package usage, which is identified in the segment partition of the GTM map also depicted in *Figure 8.7* (bottom row, left), and in the spread over the map of the customer churn rate, which is again colour-coded with white indicating the highest churn rate and black, the lowest (bottom row, right).

Although similar results are available for the complete analyzed period, they are not shown for the sake of brevity. The same can be said for the corresponding flow maps: there is one for each of the 100
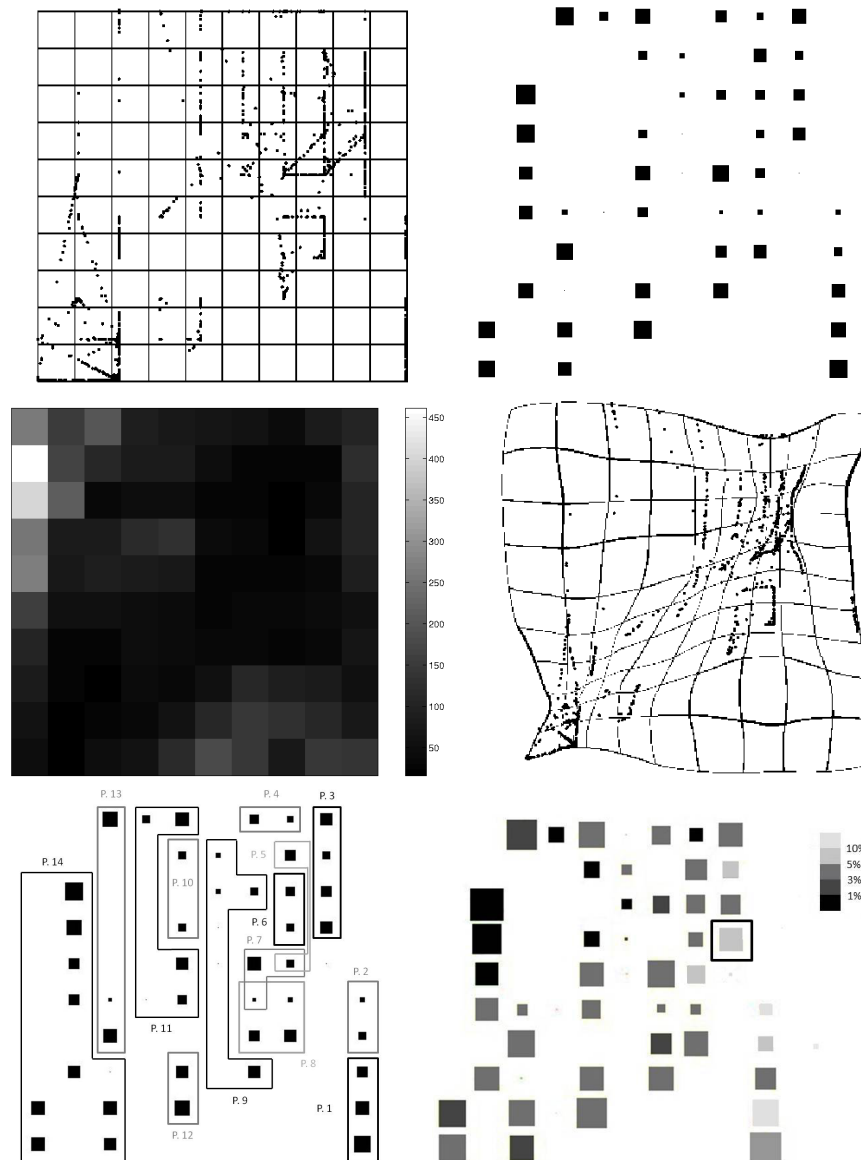
Figure 8.7: Top row: *left) mean projection* of the customers over the GTM visualization lattice; *right)* corresponding *mode projection*, with relative square size in proportion to the ratio of customers assigned to individual nodes. Middle row: *left)* MF colour map, with highest distortion in white; *right)* corresponding cartogram distorting the GTM lattice according to the MF. Bottom row: *left)* Customer segments according to service packages, overlaid on top of the *mode projection*; *right)* churn ratio as a colour map superimposed to the *mode projection* (node 84 single out by an overlaying black square).

GTM map nodes for each month. For illustration, *Figure 8.8* shows the detailed migration routes and the customer flow (*Figure 8.9*) *from* a high-churn node (*number 84*, highlighted in *Figure 8.7*, bottom row, right).

Both data projections in the top row of *Figure 8.7* suggest the existence of data structure, with both densely populated and completely empty spaces in the map. The display of the MF map in *Figure 8.7* (middle row, left) hints the existence of general segments, and this is visually emphasized by the cartogram display in *Figure 8.7* (middle row, right), which enhances the *mean projection* (top row, left) by explicitly reintroducing the MF-quantified distortion into the map. At least to some extent, this structure is likely to be dominated by separation due to customer diversity in their usage of distinct pay-per-view packages.
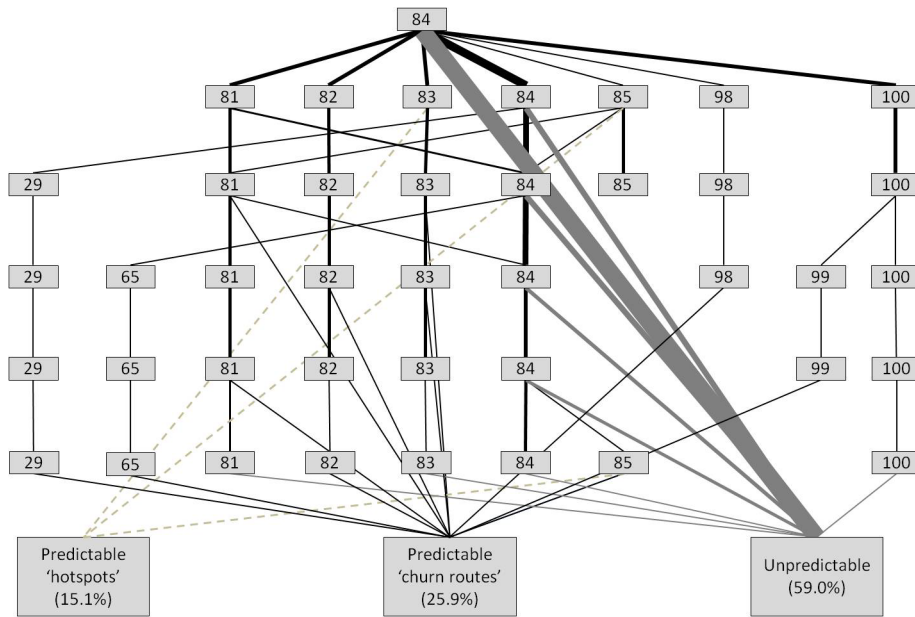
Figure 8.8: Customer migrations from GTM *node 84*. Nodes are numbered according to their column-wise location in the 10×10 GTM map (leftmost column, top-to-bottom: nodes 1-10; rightmost column: 91-100). Nodes are grouped into three categories: predictable *hotspots*, predictable churn routes and unpredictable ones. The thickness of lines is proportional to the volume of migration.
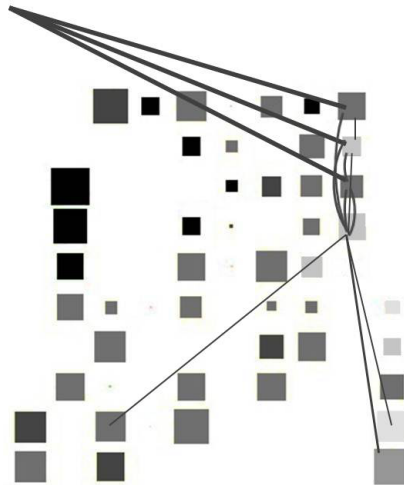
Figure 8.9: Flow Map corresponding to the previous figure, showing the *geography* of the customer migrations outflowing from *node 84* over the GTM visualization map. Lines projected out of the map indicate churn.

This is confirmed by the delimitation of customer segments shown in *Figure 8.7* (bottom row, left), where packages are described as p.$\sharp n$. The right-hand side ones are mostly basic and economic commercial packages, whereas the ones on the center are mid-range thematic TV channels, and the ones on the left are premium packages. This structure reflects the company's commercial strategy.

The visualization of the churn rate by colour coding the GTM *mode projection* (bottom row, right) is most revealing, indicating the high propensity to churn in customers mapped into the bottom, right-hand side corner of the GTM map, which are fairly isolated from the rest of customers. This area corresponds to basic channel packages. The use in isolation of these non-specialized channel bundles thus reveals itself as the main gate to customer churn for the company.

The migration routes (see *Figure 8.8*) for *node 84* allow us to identify *predictable* and *unpredictable* churn routes as well as *hot spots*, which are areas in the GTM map that would require preferential attention. These routes are overlaid in the GTM as Flow Maps in *Figure 8.9*. A total of 25.9% of churners with origin in this node show behaviours (in the form of movements across the visual map) that could be anticipated up to 3 months in advance.

In summary, the combined use of GTM maps, their cartogram representation and Flow Maps allows the data analyst to anticipate and prevent churn. For the analyzed company, it helped to anticipate 42.5% of the 3.795 customer requests for service cancellations over the 7 month period. A total of 75.2% of them were deemed to be suitable for preventive commercial action.

8.3 Conclusions

The analysis of business information often requires the use of exploratory data mining techniques. Amongst them, MVD visualization is likely to provide invaluable insights for knowledge discovery.

The discovery of adequate models for the analysis of customer churn has become paramount for the achievement of competitive advantage. In the current chapter, we have proposed a novel method of MVD visualization that combines the flexibility of the GTM nonlinear manifold learning model with the abilities of two visualization techniques from the field of geographical representation: Cartograms and Flow Maps.

A number of experiments with two large databases of telecommunication and pay-per-view TV customers have illustrated the usefulness and actionability of the proposed MVD visualization method. High churn areas, or customer *departure gates*, have been visually identified in a manner that allows their description in terms of customer usage and, thus, the implementation of commercial campaigns oriented to increase customer retention. Importantly, the method has also provided a detailed visualization of customer migration routes, which should enable preventive marketing actions to avoid churn.

Chapter 9

Conclusions and future research

9.1 Conclusions

Any data mining process aims to extract knowledge from models that are built from the available data. The acquisition of such knowledge is bounded though by the requirement that those models be in fact interpretable.

Model interpretability is by no means a given. In particular, many modeling techniques belonging to the field of ML have been hampered in their adoption and, as a result, are arguably underused precisely because they are (or, sometimes, they have been labeled as) "black boxes".

In practice, the data analyst is often faced by the need of trading off modeling accuracy and flexibility against model simplicity and interpretability. This is one of the main reasons that justify the popularity of simple, even is sometimes suboptimal, linear models.

The use of nonlinear techniques imbues the modeling process with flexibility, as it can adapt to the local characteristics of data, but often at the price of making the interpretation of results more involved and in no way straightforward.

Even if a successful extraction of knowledge from data is achieved, this could not be enough in some application fields. There are problems in which we not only need knowledge, but *usable* knowledge; that is, knowledge that an analyst can act upon in practical terms. Marketing and, in particular customer relationship management is one of those fields in which most problems have that type of requirement attached.

This Thesis was developed in the framework of business cases in which the growing competitive pressure makes companies fight over their customer portfolios; a fight that leads to the common phenomenon of customer attrition or churn. In this scenario, understanding how customer loyalty construction mechanisms work; anticipating the customer's intention to abandon and the proactive implementation of retention-focused actions, are all elements that should lead to competitive advantage.

In this context, from a business point of view, during the development of the work that constructs the present Doctoral Thesis it has been possible to validate some initial hypothesis related to the possibility of configuring a visualization method for abandonment routes (and value generation) which is efficient in terms of commercial actionability. Thus, it has been possible to validate the following hypothesis:

- Different patterns of service consumption determine different levels of predisposition to abandon.

- It is possible to group different prototypes or micro-clusters from the GTM projection in commercially interpretable segments, which are operative and improve their market actionability.

- Different migration routes between time periods are likely to exist and to be identifiable in the representation map (using flowmaps), both:

  - Negative: towards lower customer value areas and, eventually, service abandonment.
  - Positive: towards higher customer value areas.

- It is possible to interact, via commercial actions, with the identified migration routes:

  - Changes in customer behaviour / customer migrations between nodes or prototypes tend to be nearby their original position (except in case there are changes / promotions in the competitive offer).

  - A quarter of churners show behaviours (seen as movements across the visual map) that could be anticipated up to 3 months in advance.

  - The anticipation of movements directed to "departure gate" zones enables to implement customer retention commercial actions.

Summarizing, the application of these methods helps increasing the interpretability of the visualization of the analyzed database, thus assisting in the process of useful knowledge extraction that could have a practical impact on customer retention management strategies:

- For the analyzed companies, it helped to anticipate 42.5% of customer requests for service cancellation. A total of 75.2% of them where deemed to be suitable for preventive commercial actions.

In chapter 2, we have reviewed the concept of customer churn making use of a "customer continuity management" concept that assumes that companies must prolong the life expectancy of their customer base as much as possible, assuring its adequate development in terms of value, through the implementation of suitable commercial actions for each of the stages of their lifecycle.

This concept makes us understand customer churn prevention as a matter of customer loyalty construction, involving customer service and satisfaction, as well as switching costs and barriers. All these business constructs have operational quantitative descriptions that allow the implementation of quantitative management strategies.

Given that one paramount quantitative management strategy is churn prediction, which could activate proactive customer retention strategies, we have devoted Chapter 3 to an in-depth review of current predictive models in churn management, with a strong focus on machine learning and computational intelligence approaches.

From this extensive review, it is clear that the business areas in which churn is analyzed quantitatively are plenty, although some of them predominate, namely banking and finance, and telecommunications. The predominant techniques are, on one side (that of reasonably interpretable models), logistic regression and decision trees, and, on the other (ML techniques) artificial neural networks and support vector machines. One of the key lessons of the review is that data relevance and quality are quite heterogeneous, too often ignoring that the success of the analyses usually depends on these factors.

In this Thesis we have explored the potential of advanced nonlinear techniques (mostly for dimensionality reduction), while trying to keep the models interpretable and actionable. The data mining approach we take mostly concerns the use of unsupervised machine learning techniques. Within the overall goal of exploring the existence of customer churn routes according to the customers' service consumption patterns, we are particularly interested in methods that are capable of providing simultaneous visualization and clustering of the available data. To this end, Chapter 4 summarily introduced this type of models.

In Chapter 5, we focused in nonlinear techniques for classification, assisted by a model for rule extraction from the results, acknowledging that the latter is a viable strategy for conveying results in a way that is often found to be more amenable to business interpretation, as in decision trees. Here, we provided some preliminary experiments concerning customer satisfaction and loyalty, as elements of the churn problem, from a supervised learning perspective. The experiments concerned data from petrol station usage surveys. Feature relevance determination for feature selection and rule extraction were the tools used for achieving interpretability. The obtained results were consistent with recent theory on satisfaction, loyalty and switching barriers models.

We then moved from supervised to unsupervised learning: an approach that we took in the remainder of the Thesis. Whereas in Chapter 5 the focus was placed on customer satisfaction as a key to churn prevention, in Chapter 6 we changed the viewpoint to proactive customer bonding. An indirect and explanatory approach to the prediction of customer abandonment was proposed. It is based on the visualization of customer data -consisting of their consumption patterns- on the two-dimensional representation map of

a principled statistical machine learning method of the NLDR manifold learning family. It was used to explore the existence of regular abandonment routes in the Brazilian telecommunications market.

A two-tier market segmentation process was proposed, which involved a number of analytical novelties. The underlying model is endowed with an in-built unsupervised feature relevance determination method that optimizes clustering by increasing the influence of those features that better describe the natural separation of data groups. The segmentation results were validated with several cluster quality indices, one of them specifically defined for the GTM model. The resulting segmentation solution was also assessed in business terms and found to be easy to describe according to the features found to be most relevant by the FRD-GTM. Two *ad hoc* segment solution evaluation metrics: Churn Index and Commercial Margin were also defined. Different areas where the risk of abandonment are higher, or *departure gates*, were identified on the basis of service consumption patterns. The migration routes between market segments were also explored in preparation for the results reported in Chapter 8. As a whole, this method should provide a solid basis for the development of a *churn warning system*.

Chapter 7 added a more theoretical extension of the models presented in the previous chapter, bearing in mind the overarching goal of model interpretability. It addresses one of the characteristics of NLDR that most affects such interpretability. The fact that these models distort nonlinearly the data projection in a local fashion, limiting the possibility of making sense of the results in terms of the original data variables.

In this chapter, inspired from a technique originally designed for the analysis of geographic information, namely the cartograms, we have proposed a new method for explicitly reintroducing this geometrical distortion in the low-dimensional representation of the MVD. The proposed cartogram-based method reintroduces the distortion explicitly into the visualization maps. By reintroducing this distortion explicitly, we show that the local neighborhood relationships in the low-dimensional representation space reflect more faithfully those in the observed data space. Extensive experimentation with artificial and real data were carried out to assess the capabilities and limitations of the proposed technique. Importantly, several guidelines of use of practical interest were extracted from these experiments.

These advances are then put to test for the analysis of churn in the telecommunications & media market in Chapter 8. Here, cartograms were applied together with a second method of MVD visualization, also inspired in geographical information representation: The Flow Map. Originally devised to visualize geography-related evolution patterns such as, for instance, population migrations, we use Flow Maps to analyze the customer migrations over the GTM visualization map, aiming to detect foci of potential customer churn.

Given that the analyzed databases contain information over time, for the first time in the Thesis, we go beyond a static snapshot of current market segments and investigate customer evolution over time, trying to prevent individual customers drifting towards churn-risk areas. This time-dependent component should allow the service provider to design and launch customer retention actions oriented towards the retention of the most profitable customers.

The combination of cartograms and flow maps was shown to provide very informative results: High churn areas, or customer departure gates, were visually identified in a manner that allows their description in terms of customer usage and, thus, the implementation of commercial campaigns oriented to increase customer retention. Importantly, the method also provided a detailed visualization of customer migration routes, which should enable preventive marketing actions to avoid churn.

9.2 Suggestions for future research

As very often with research thesis, this one has tied a number of novel developments up that are, somehow, self-conclusive. Some other developments, though, probably open as many doors to new research paths as they close.

From a data analysis viewpoint, the following possibilities for future research are highlighted:

- The use of alternative rule extraction methods to OSRE in the supervised framework of analysis presented in chapter 5, as well as the design of rule extraction procedures to obtain commercially

actionable rule descriptions of market segments obtained within an unsupervised framework.

- The extention of cartograms to other NLDR methods for data visualization. As mentioned in chapter 7, nothing precludes us from, beyond GTM, adapting this technique to other NLDR methods, provided some sort of distortion measure could be quantified. For instance, some preliminary experiments for the cartogram representation of the MF and the U-Matrix in batch-SOM have been carried out in [249]. The lattice of latent points in GTM has been used to establish the limiting borders of the distortion regions in our experiments, but this is not the only possible approach to border definition for the generation of cartograms. In fact, Voronoi diagrams [193] of the visualization space, based on more or less compact data representations based on their *posterior mean projections*, could also be used to the purpose of creating cartograms. This could open the application of cartogram techniques to nonlinear methods that do not provide vector quantization. Cartograms could also be used as a visual guide for interactive hierarchical models for MVD clustering and visualization [24, 217, 246], for which different levels of the hierarchy could be semi-automatically controlled, allowing user interaction, according to levels of mapping distortion.

- Design of alternative methods to cartograms for the integration of the distortion in the visualization maps of NLDR methods. A cartogram as a geographical representation-inspired visual metaphor is just one of the possibilities available to the analyst to reintegrate distortion on nonlinear MVD mappings. Other possibilities might well be considered. Some research in that direction has only recently been published [92].

- Although flow maps have provided us with a tool for the visualization of the dynamics of churn over time periods, a more principled procedure for the integration of cartograms and flow maps in proper time series is yet to be defined.

On the commercial area, two main avenues for future work are open for development:

- On one side, the optimization and operative deployment of the proposed working methodology as a tool for customer abandonment prediction. This would entail the design of an appropriate decision support system software.

- On the other side, the study of value generation routes that would allow the identification of customer commercial development policies (cross-selling and up-selling actions), as well as selective customer acquisition policies, focussed on customers with high potential value.

Publications

Here, we list the publications resulting from the research reported in this Doctoral Thesis:

- A. Vellido, T. A. Etchells, D. L. García, and À. Nebot. Describing customer loyalty to Spanish petrol stations through rule extrction. In *Proceedings of the 7th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2006)*, Lecture notes in Computer Science, pages 970–977, Burgos, Spain, 2006.

- D. L. García, A. Vellido, and À. Nebot. Customer continuity management as a foundation for churn data mining. Technical report LSI-07-2-R, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, 2007.

- D. L. García, A. Vellido, and À. Nebot. Predictive models in churn data mining: a review. Technical report LSI-07-4-R, Universitat Politécnica de Catalunya (UPC), Barcelona, Spain, 2007.

- D. L. García, A. Vellido, and À. Nebot. Identification of churn routes in the Brazilian telecommunications market. In 15th *European Symposium on Artificial Neural Networks, (ESANN)*, pages 585–590, 2007.

- D. L. García, A. Vellido, and À. Nebot. Finding relevant features for the churn analysis-oriented segmentation of a telecommunications market. In I. R. Ruiz and H. P. Cintas, editors, *Actas del II Simposio de Inteligencia Computacional (IEEE SICO 2007)*, pages 301–310. Thomson, 2007.

- A. Vellido, D. García, and À. Nebot. Cartogram visualization for nonlinear manifold learning models. *Data Mining and Knowledge Discovery*, 27(1):22–54, 2013.

- D. L. García, À. Nebot, and A. Vellido. Telecommunications customers churn monitoring using flow maps and cartogram visualization. In *GRAPP 2013 / IVAPP 2013 - Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications*, pages 451–460, Barcelona, Spain, 2013. SciTePress.

- D. L. García, À. Nebot, and A. Vellido. Visualizing pay-per-view television customers churn using cartograms and flow maps. In *Proceedings of the 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, Bruges, Belgium, 2013.

References

[1] D. Aaker. *Strategic Market Management*. John Wiley & Sons, New York, 1988.

[2] D. Alahakoon, S. K. Halgamuge, and B. Srinivasan. Dynamic self-organizing maps with controlled growth for knowledge discovery. *IEEE Transactions on Neural Networks*, 11(3):601–614, 2000.

[3] B. Anckar and D. D'Incau. Value creation in mobile commerce: findings from a consumer survey. *Journal of Information Technology Theory and Application*, 4(1):43–64, 2002.

[4] E. Anderson and B. Weitz. Determinants of continuity in conventional industrial channel dyads. *Marketing Science*, 8(3):10–23, 1989.

[5] E. W. Anderson and C. Fornell. A customer satisfaction research prospectus. In *Service Quality: New Directions in Theory and Practice*. Sage Publications, Inc., Thousand Oaks, CA, 1994.

[6] D. Athanasopoulou. Relationship quality: A critical literature review and research agenda. *European Journal of Marketing*, 43(5/6):583–610, 2009.

[7] A. D. Athanassopoulos. Customer satisfaction cues to support market segmentation and explain switching behaviour. *Journal of Business Research*, 47(3):191–207, 2000.

[8] W. H. Au, K. C. Chan, and X. Yao. A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Transactions on Evolutionary Computation*, 7(6):532–545, 2003.

[9] M. Aupetit. Visualizing distortions and recovering topology in continuous projection techniques. *Neurocomputing*, 70(7-9):1304–1330, 2007.

[10] B. Baesens, G. Verstraeten, D. Van den Poel, M. Egmont-Peterson, P. Van Kenhove, and J. Van-thienen. Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers. *European Journal of Operational Research*, 156(2):508–523, 2004.

[11] W. A. Band. *Creación del Valor, la Clave de la Gestión Competitiva*. Diaz de Santos, Madrid, 1994.

[12] D. J. Bartholomew. *Latent Variable Models and Factor Analysis*. Charles Griffin and Co. Ltd, London, 1987.

[13] W. Bearden and R. Netemeyer. *Handbook of Marketing Sales: Multi-item Measures for Marketing and Consumer Behavior Research*. Sage, London, 2nd edition, 1999.

[14] S. E. Beatty, M. Mayer, J. E. Coleman, K. E. Reynolds, and J. Lee. Customer-sales associate retail relationships. *Journal of Retailing*, 72(3):223–47, 1996.

[15] R. S. Behara, W. W. Fisher, and J. G. Lemmink. Modelling and evaluating service quality measurement using neural networks. *International Journal of Operations and Production Management*, 22(10):1162–1185, 2002.

[16] S. J. Bell, S. Auh, and K. Smalley. Customer relationship dynamics: Service quality and customer loyalty in the context of varying levels of customer expertise and switching costs. *Journal of the Academy of Marketing Science*, 33(2):169–183, 2005.

[17] N. Bendapudi and L. L. Berry. Customers' motivations for mantaining relationships with service providers. *Journal of Retailing*, 73(1):15–37, 1997.

[18] D. Berg. Bankruptcy prediction by generalized additive models. *Applied Stochastic Models in Business and Industry*, 23(2):129–143, 2007.

[19] L. L. Berry and A. Parasuraman. *Marketing Services: Competing through Quality*. Free Press, New York, 1991.

[20] S. Bhat, R. Burkhard, K. A. O'Donell, and D. L. Wardlow. Version 6.0.1 anyone? An investigation of customer software upgrading behaviour. *Journal of Marketing Theory and Practice*, 6(2):87–96, 1998.

[21] C. Bishop, G. Hinton, and I. Strachan. GTM trough time. In *IEE Fifth International Conference on Artificial Neural Networks*, pages 111–116, Cambridge, UK, 1997.

[22] C. Bishop, M. Svensén, and C. K. I. Williams. Magnification factors for the GTM algorithm. In *Proceedings IEEE Fifth international conference on artificial neural networks*, pages 64–69, Cambridge, UK, 1997.

[23] C. Bishop, M. Svensén, and C. K. I. Williams. GTM: The generative topographic mapping. *Neural Computing*, 10(1):215–234, 1998.

[24] C. M. Bishop and M. E. Tipping. A hierarchical latent variable model for data visualization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):281–293, 1998.

[25] C. M. Bishop, M. Svensén, and C. K. I. Williams. Magnification factors for the SOM and GTM algorithms. In *1997 Workshop on Self-Organizing Maps*, pages 333–338, Helsinki, Finland, 1997.

[26] M. J. Bitner and A. R. Hubbert. Encounter satisfaction versus overall satisfaction versus quality. In R. T. Rust and R. L. Oliver, editors, *Service Quality: New Directions in Theory and Practice*. Sage Publications, Inc., New York, 1994.

[27] J. M. Bloemer, T. Brijis, K. Vanhoof, and G. Swinnen. Comparing complete and partial classification for identifying customers at risk. *International Journal of Research in Marketing*, 20(2):117–131, 2003.

[28] R. N. Bolton and J. H. Drew. A multistage model of customers' assessments of service quality and value. *Journal of Consumer Research*, 17(4):375–384, 1991.

[29] L. Bose and X. Chen. Quantitative models for direct marketing: A review from systems perspective. *European Journal of Operational Research*, 195(1):1–16, 2009.

[30] B. E. Boser, I. M. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. In *Fifth Annual Workshop on Computational Learning Theory*, pages 114–152, Pittsburg, 1992.

[31] D. E. Bowen. Managing customers as human resources in service organizations. *Human Resource Management*, 25 (3):371–383, 1986.

[32] M. Brady and J. J. Cronin. Some new thoughts on conceptualizing perceived service quality: a hierarchical approach. *Journal of Marketing*, 65(3):34–49, 2001.

[33] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.

[34] A. P. Brief and R. J. Aldag. The intrinsic-extrinsic dichotomy: toward conceptual clarity. *Academy of Management Review*, 2(3):496–500, 1977.

[35] W. Buckinx and D. Van den Poel. Customer base analysis: Partial defection of behaviorally loyal clients in a non-contractual FMCG retail setting. *European Journal of Operational Research*, 164(1): 252–268, 2005.

[36] M. P. Bunker and A. D. Ball. Causes and consequences of grudge-holding in service relationships. *Journal of Services Marketing*, 22(1):37–47, 2008.

[37] J. Burez and D. Van den Poel. CRM at a pay-TV company: using analytical models to reduce customer attrition by targeted marketing for subscription services. *Expert Systems with Applications*, 32(2):277–288, 2007.

[38] C. Burges. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167, 1998.

[39] H. E. Butz and L. D. Goodstein. Measuring customer value: gaining the strategic advantage. *Organizational Dynamics*, 24(3):63–77, 1996.

[40] R. C. Caceres and N. G. Paparoidamis. Service quality, relationship satisfaction, trust, commitment, and business-to-business loyalty. *European Journal of Marketing*, 41(7):836–867, 2007.

[41] M. A. Carreira-Perpiñán and S. Renals. A latent variable modelling approach to the acoustic-to-articulatory mapping problem. In *14th International Congress of Phonetic Sciences*, Cambridge, UK, 1999.

[42] J. Carroll, S. Howard, J. Peck, and J. Murphy. A field study of perceptions and use of mobile telephones by 16 to 22 year olds. *Journal of Information Technology Theory and Application*, 4(2): 49–61, 2002.

[43] J. Carroll, S. Howard, J. Peck, and J. Murphy. Just what do the youth of today want? Technology appropriation by young people. In *35th Hawaii International Conference on System Sciences*. IEEE, 2002.

[44] P. Chen and L. Hitt. Measuring switching costs and the determinants of customer retention in internet-enabled businesses: A study of the online brokerage industry. *Information Systems Research*, 13(3), 2002.

[45] Z. Y. Chen, Z. P. Fan, and M. Sun. A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data. *European Journal of Operational Research*, 223(2):461–472, 2012.

[46] D. Chiang, Y. Wang, S. Lee, and C. Lin. Goal-oriented sequential pattern for network banking and churn analysis. *Expert Systems with Applications*, 25(3):293–302, 2003.

[47] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.

[48] K. Coussement and K. W. De Bock. Customer churn prediction in the online gambling industry: the beneficial effect of ensemble learning. *Journal of Business Research*, 66(9):1629–1636, 2013.

[49] K. Coussement and D. Van den Poel. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, 34(1):313–327, 2008.

[50] F. Crespo and R. Weber. A methodology for dynamic data mining based on fuzzy clustering. *Fuzzy Sets and Systems*, 150(2):267–284, 2005.

[51] S. F. Crone, S. Lessmann, and R. Stahlbock. The impact of preprocessing on data mining. an evaluation of classifier sensitivity in direct marketing. *European Journal of Operational Research*, 173(3):781–800, 2006.

[52] J. J. Cronin and S. A. Taylor. Measuring service quality: A re-examination and extension. *Journal of Marketing*, 56(3):55–68, 1992.

[53] J. J. Cronin and S. A. Taylor. SERVPERF versus SERVQUAL: Reconciling performance based and perceptions-minus-expectations measurement of service quality. *Journal of Marketing*, 58(1): 125–131, 1994.

[54] J. J. Cronin, M. K. Brady, and G. T. Hule. Assessing the effects of quality, value, and customer satisfaction on customer behavioural intentions in service environments. *Journal of Retailing*, 76(2): 193–218, 2000.

[55] R. Cruz and A. Vellido. Semi-supervised geodesic generative topographic mapping. *Pattern Recognition Letters*, 31(3):202–209, 2010.

[56] R. Cruz and A. Vellido. Semi-supervised analysis of human brain tumours from partially labeled MRS information, using manifold learning models. *International Journal of Neural Systems*, 21(1): 17–29, 2011.

[57] J. A. Czepiel. Service encounters and service relationships: Implications for research. *Journal of Business Research*, 20 (1):13–21, 1990.

[58] P. A. Dabholkar, C. D. Shepherd, and D. I. Thorpe. A comprehensive framework for service quality: An investigation of critical conceptual and measurement issues through a longitudinal study. *Journal of Retailing*, 76(2):139–173, 2000.

[59] P. Datta, B. Masand, P. R. Mani, and B. Li. Automated cellular modelling and prediction on a large scale. *Artificial Intelligence Review*, 14(6):485–502, 2000.

[60] D. L. Davies and D. W. Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2):224–227, 1979.

[61] G. Day. *Analysis for Strategic Market Decisions*. West Publishing, St. Paul, 1986.

[62] K. De Bock and D. Van den Poel. Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models. *Expert Systems with Applications*, 39(8):6816–6826, 2012.

[63] K. De Bock, K. Coussement, and D. Van den Poel. Ensemble classification based on generalized additive models. *Computational Statistics & Data Analysis*, 54(6):1535–1546, 2010.

[64] M. H. De Canniere, De Pelsmacker, and M. Geuens. Relationship quality and the theory of planned behavioral intentions and purchase behavior. *Journal of Business Research*, 62(1):82–92, 2009.

[65] Z. Deng, Y. Lu, K. K. Wei, and J. Zhang. Understanding customer satisfaction and loyalty: An empirical study of mobile instant messages in china. *International Journal of Informational Management*, 10(4):289–300, 2009.

[66] T. K. Dey, H. Edelsbrunner, and S. Guha. Computational topology. In *Advances in Discrete and Computational Geomatry*, volume Contemporary Mathematics, 223, pages 109–143. American Mathematical Society, 1999.

[67] A. S. Dick and K. Basu. Customer loyalty: toward an integrated conceptual framework. *Journal of the Academy of Marketing Science*, 22(2):99–113, 1994.

[68] A. S. Dick and K. R. Lord. The impact of membership fees on consumer attitude and choice. *Psychology & Marketing*, 15 (1):41–58, 1998.

[69] A. L. Drolet and D. G. Morrison. Do we really need multiple item measures in service research? *Journal of Service Research*, 3(3):196–204, 2001.

[70] Q. Du, V. Faber, and M. Gunzburger. Centroidal Voronoi tessellations: Applications and algorithms. *SIAM Review*, 41(4):637–676, 1999.

[71] J. Dy and C. Brodley. Feature selection for unsupervised learning. *Machine Learning research*, 5(1): 845–889, 2004.

[72] R. M. Emerson. Social exchange theory. *Annual Review of Sociology*, 2:35–362, 1976.

[73] T. A. Etchells and P. J. Lisboa. Orthogonal search-based rule extraction (OSRE) method for trained neural networks: A practical and efficient approach. *IEEE Transactions on Neural Networks*, 17(2): 374–384, 2006.

[74] T. A. Etchells, I. H. Jarman, and P. J. Lisboa. Empirically derived ruled of adjuvant chemotherapy in breast cancer treatment. In *MEDSIP International Conference*, pages 345–351, Malta, 2004.

[75] T. A. Etchells, A. Nebot, A. Vellido, P. J. Lisboa, and F. Mugica. Learning what is important: feature selection and rule extraction in a virtual course. In *Proceedings of the 14th European Symposium on Artificial Neural Networks (ESANN)*, pages 401–406, Bruges, Belgium, 2006.

[76] U. Fayyad, G. Piatetski-Shapiro, and P. Smith. From data mining to knowledge discovery in databases. *AI Magazine*, 17(3):37–54, 1996.

[77] J. B. Ferreira, M. Vellasco, M. A. Pacheco, and C. H. Barbosa. Data mining techniques on the evaluation of wireless churn. In *Proceedings of the 12th European Symposium on Artificial Neural Networks (ESANN)*, pages 483–488, 2004.

[78] C. Fornell. A national customer satisfaction barometer. the Swedish experience. *Journal of Marketing*, 56(1):6–21, 1992.

[79] C. Fornell, M. D. Johnson, E. W. Anderson, J. Cha, and B. E. Bryant. The American customer satisfaction index: nature, purpose and findings. *Journal of Marketing*, 60(4):7–18, 1996.

[80] G. L. Frazier. On the measurement of interfirm power in channels of distribution. *Journal of Marketing Research*, 20(2):158–166, 1983.

[81] G. Fullerton. The service quality-loyalty relationship in retail services: does commitment matter? *Journal of Retailing and Consumer Services*, 12(2):99–111, 2005.

[82] T. Furukawa. SOM of SOMs. *Neural Networks*, 22(4):463–478, 2009.

[83] B. T. Gale. *Managing Customer Value: Creating Quality and Service that Customers can See*. Free Press, New York, 1994.

[84] E. Garbarino and M. S. Johnson. The different roles of satisfaction, trust, and commitment in customer relationships. *Journal of Marketing*, 63(2):70–87, 1999.

[85] D. L. García, A. Vellido, and À. Nebot. Identification of churn routes in the Brazilian telecommunications market. In *15th European Symposium on Artificial Neural Networks, (ESANN)*, pages 585–590, 2007.

[86] D. L. García, A. Vellido, and À. Nebot. Customer continuity management as a foundation for churn data mining. Technical report LSI-07-2-R, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain, 2007.

[87] D. L. García, A. Vellido, and À. Nebot. Predictive models in churn data mining: a review. Technical report LSI-07-4-R, Universitat Politécnica de Catalunya (UPC), Barcelona, Spain, 2007.

[88] D. L. García, A. Vellido, and À. Nebot. Finding relevant features for the churn analysis-oriented segmentation of a telecommunications market. In I. R. Ruiz and H. P. Cintas, editors, *Actas del II Simposio de Inteligencia Computacional (IEEE SICO 2007)*, pages 301–310. Thomson, 2007.

[89] D. L. García, À. Nebot, and A. Vellido. Visualizing pay-per-view television customers churn using cartograms and flow maps. In *Proceedings of the 21st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, Bruges, Belgium, 2013.

[90] D. L. García, À. Nebot, and A. Vellido. Telecommunications customers churn monitoring using flow maps and cartogram visualization. In *GRAPP 2013 / IVAPP 2013 - Proceedings of the International Conference on Computer Graphics Theory and Applications and International Conference on Information Visualization Theory and Applications*, pages 451–460, Barcelona, Spain, 2013. SciTePress.

[91] M. T. Gastner and M. E. Newman. Diffusion-based method for producing density-equalizing maps. *Proceedings of the National Academy of Sciences USA*, 101(20):7499–7504, 2004.

[92] N. Gianniotis. Interpretable magnification factors for topographic maps of high dimensional and structured data. In *Procs. of the IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, pages 238–245. IEEE, 2013.

[93] R. Gilbert. Mobility barriers and the value of incumbency. *Handbook of Industrial Organisation*, 1: 475–535, 1989.

[94] A. Gisbretch, B. Mokbel, and B. Hammer. Relational generative topographic mapping. *Neurocomputing*, 74(9):1359–1371, 2011.

[95] N. Glady, B. Baesens, and C. Croux. Modeling churn using customer lifetime value. *European Journal of Operational Research*, 197(1):402–411, 2008.

[96] S. Gounaris and V. Stathakopoulos. Antecedents and consequences of brand loyalty: an empirical study. *The Journal of Brand Management*, 11(4):283–306, 2004.

[97] V. Govindaraju, K. Young, and A. A. Maudsley. Proton NMR chemical shifts and coupling constants for brain metabolites. *NMR in Biomedicine*, 13(3):129–153, 2000.

[98] D. D. Gremler. The effect of satisfaction, switching costs, and interpersonal bonds on service loyalty. Unpublished dissertation, Arizona State University, 1995.

[99] C. Grönroos. A service quality model and its marketing implications. *European Journal of Marketing*, 18(4):36–44, 1984.

[100] J. P. Guiltinan. A classification of switching costs with implications for relationship marketing. In R. P. B. T.L. Childers and J. P. Peter, editors, *1989 AMA Winter Educators' Conference: Marketing Theory and Practice*, pages 216–220, Chicago, IL, 1989. American Marketing Association.

[101] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *The Journal of Machine Learning Research*, 3(1):1157–1182, 2003.

[102] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh. Feature extraction: Foundations and applications. In *Studies in Fuzziness and Soft Computing*. Springer, 2006.

[103] J. Hadden. A customer profiling methodology for churn prediction. PhD thesis, Cranfield University, UK, 2008.

[104] J. Hadden, A. Tiwari, R. Roy, and D. Ruta. Computer-assisted customer churn management: State-of-the-art and future trends. *Computers and Operations Research*, 34(10):2902–2917, 2007.

[105] B. Hammer and T. Villmann. Mathematical aspects of neural networks. In *Proceedings of the 11th European Symposium on Artificial Neural Networks (ESANN)*, pages 59–72. d-side pub, Brussels, Belgium, 2003.

[106] B. Hammer, A. Hasenfuss, and T. Villmann. Magnification control for batch neural gas. *Neurocomputing*, 70(7-9):1225–1234, 2007.

[107] M. D. Hartline and O. C. Ferrell. The management of customer-contact service employees: An empirical investigation. *Journal of Marketing*, 60(4):52–70, 1996.

[108] T. Hastie. Principal curves and surfaces. Technical report, Department of Statistics, Stanford University, 1984.

[109] J. B. Heide and A. M. Weiss. Vendor consideration and switching behaviour for buyers in high-technology markets. *Journal of Marketing*, 59 (3):52–70, 1995.

[110] J. L. Heskett, W. E. Sasser, and C. W. Hart. *Service breakthroughs: Changing the rules of the game*. The Free Press, New York, 1990.

[111] J. L. Heskett, T. O. Jones, G. W. Loveman, E. W. Sasser, and L. A. Schlesinger. Putting the service-profit chain to work. *Harvard Business Review*, 72(2):164–74, 1994.

[112] G. Hinton, C. K. I. Williams, and M. D. Revow. Adaptive elastic models for hand-printed character recognition. *Advances in Neural Information Processing Systems*, 4:512–519, 1992.

[113] E. Hirschman. Innovativeness, novelty seeking and consumer creativity. *Journal of Consumer Research*, 7(3):283–95, 1980.

[114] T. K. Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998.

[115] S. Ho Ha, S. Min Bae, and S. Chan Park. Customers' time-variant purchase behaviour and corresponding marketing strategies: an online retailer's case. *Computers and Industrial Engineering*, 43(4):801–820, 2002.

[116] M. B. Holbrook. *The Nature of Customer Value: An Axiology of Services in the Consumption Experience*. Sage, Thousand Oaks, CA, 1994.

[117] N. Hsieh. An integrated data mining and behavioural scoring model for analysing bank customers. *Expert Systems with Applications*, 27:623–633, 2004.

[118] C. Hsu, C. Chang, and C. Lin. A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University, Taiwan, 2008.

[119] B. Q. Huang, M. T. Kechadi, and B. Buckley. Customer churn prediction for broadband internet services. In M. K. M. T. B. Pedersen and M. Tjoa, editors, *DaWaK 2009, LNCS 5691*, pages 229–243. Springer-Verlag, Berlin, 2009.

[120] B. Q. Huang, M. T. Kechadi, and B. Buckley. Customer churn prediction in telecommunications. *Expert Systems with Applications*, 39(1):1414–1425, 2012.

[121] M. H. Huang and S. Yu. Are consumers inherently or situationly brand loyal? A set intercorrelation for conscious brand loyalty and non conscious inertia. *Psychology and Marketing*, 16(6):534–544, 1999.

[122] S. Hung, D. C. Yen, and H. Y. Wang. Applying data mining to telecom churn management. *Expert Systems with Applications*, 31(3):512–524, 2006.

[123] H. Hwang, T. Jung, and E. Suh. An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert Systems with Applications*, 26(2):181–188, 2004.

[124] B. B. Jackson. *Winning and Keeping Industrial Customers*. Lexington Books, Lexington, MA, 1985.

[125] B. B. Jackson. *Build Customer Relationship that Last*. Harvard Business Review, 1985.

[126] A. K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651–666, 2010.

[127] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *Association for Computing Machinery Computer Surveys*, 31(3):264–323, 1999.

[128] S. D. Jap and G. Shankar. Control mechanisms and the relationship life cycle: Implications for safeguarding specific investments and developing commitment. *Journal of Marketing Research*, 37(2):227–245, 2000.

[129] H. Jeanny. Vision: Images, signals and neural networks. In *Models of Neural Processing In Visual Perception*. World Scientific Publishing, 2010.

[130] M. Jenamani, P. K. Mohapatra, and S. Ghose. A stochastic model of e-customer behaviour. *Electronic Commerce Research and Applications*, 2(1):81–94, 2003.

[131] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, 1986.

[132] M. A. Jones, D. L. Mothersbaugh, and S. E. Beatty. Switching barriers and repurchase intentions in services. *Journal of Retailing*, 70(2):259–270, 2000.

[133] M. A. Jones, D. L. Mothersbaugh, and S. E. Beatty. Why customers stay: measuring the underlying dimensions of services switching costs and managing their differential strategic outcomes. *Journal of Business Research*, 55(6):441–450, 2002.

[134] J. Jonker, N. Piersma, and D. Van den Poel. Joint optimization of customer segmentation and marketing policy to maximize long-term profitability. *Expert Systems with Applications*, 27(2):159–168, 2004.

[135] M. Juliá-Sapé, D. Acosta, M. Mier, C. Arús, and D. Watson. A multi-centre, web-accessible and quality control checked database of in vivo MR spectra of brain tumour patients. *Magnetic Resonance Materials in Physics*, 19(1):22–33, 2006.

[136] S. M. Keaveney and M. Parthasarathy. Customer switching bahavior in online services: an exploratory study of the role of selected attitudinal, behavioral and demographic factors. *Journal of the Academy of Marketing Science*, 29(4):374–390, 2001.

[137] H. Kim and C. Yoon. Determinants of subscriber churn and customer loyalty in the Korean mobile telephony market. *Telecommunications Policy*, 28(9-10):751–765, 2004.

[138] M. Kim and R. S. Ramakrishna. New indices for cluster validity assessment. *Pattern Recognition Letter*, 26(15):2353–2363, 2005.

[139] M.-K. Kim, M.-C. Park, and D.-H. Jeong. The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services. *Telecommunications Policy*, 28(2):145–159, 2004.

[140] P. Kisioglu and Y. L. Topcu. Applying Bayesian belief network approach to customer churn analysis: a case study on the telecom industry of Turkey. *Expert Systems with Applications*, 38(6):7151–7157, 2010.

[141] P. Klemperer. Markets with consumer switching cost. *The Quarterly Journal of Economics*, 102(2): 375–394, 1987.

[142] T. Kohonen. Self-organizing formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69, 1982.

[143] T. Kohonen. *Self-Organizing Maps*. Springer-Verlag, Berlin, 3rd edition, 2001.

[144] M. Koufaris. Applying the technology acceptance model and flow theory to online consumer behavior. *Information Systems Research*, 13(2):205–233, 2002.

[145] M. A. Kramer. Nonlinear principal components analysis using autoassociative neural networks. *AICHe Journal*, 37(2):233–243, 1991.

[146] D. Kumar and V. Ravi. Predicting credit card customer churn in banks using data mining. *International Journal of Data Analysis Techniques and Strategies*, 1(1):4–28, 2008.

[147] M. J. Lado, C. Cadarso-Suarez, J. Roca-Pardinas, and P. G. Tahoces. Using generalized additive models for construction of nonlinear classifiers in computer-aided diagnosis systems. *IEEE Transactions on Information Technology in Biomedicine*, 10(2):246–253, 2006.

[148] S. Y. Lam, V. Shankar, M. K. Erramilli, and B. Murthy. Customer value, satisfaction, loyalty, and switching costs: An illustration from a business-to-business service context. *Journal of the Academy of Marketing Science*, 32(3):293–311, 2004.

[149] R. A. Lambert. Customer satisfaction and future financial performance: discussion of - are non financial measures leading indicators of financial performance? An analysis of customer satisfaction. *Journal of Accounting Research*, 36:37–46, 1998.

[150] P. Langley. Crafting papers on machine learning. In P. Langley, editor, *17th International Conference on Machine Learning (ICML 2000)*, pages 1207–1216. Stanford University, 2000.

[151] B. Larivière and D. Van den Poel. Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications*, 29(2):472–484, 2005.

[152] M. Law, M. A. Figuerido, and A. K. Jain. Simultaneous feature selection and clustering using mixture models. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 26(9):1154–1166, 2004.

[153] D. N. Lawley and A. E. Maxwell. *Factor Analysis as a Statistical Method*. Butterworth and Co., London, 2nd edition, 1971.

[154] G. Leban, B. Zupan, G. Vidmar, and I. Bratko. VizRank: Data visualization guided by machine learning. *Data Mining and Knowledge Discovery*, 13(2):119–136, 2006.

[155] J. A. Lee and M. Verleysen. Nonlinear Dimensionality Reduction. In *Information Science and Statistics*. Springer, 2007.

[156] T. S. Lee, C. C. Chiu, Y. C. Chou, and C. J. Lu. Mining the customer credit using classification and regression tree and multivariate adaptive regression splines. *Computational Statistics and Data Analysis*, 50:1113–1130, 2006.

[157] A. Lemmens and C. Croux. Bagging and boosting classification trees to predict churn. *Journal of Marketing Research*, 43(2):276–286, 2006.

[158] S. Lessmann and S. Voß. A reference model for customer-centric data mining with support vector machines. *European Journal of Operational Research*, 199(1):520–530, 2009.

[159] L. Leung and R. Wei. More than just talk on the move: uses and gratifications of the cellullar phone. *Journalism and Mass Communication Quarterly*, 77(2):308–320, 2000.

[160] C. Liao, P. Palvia, and H. N. Lin. The roles of habit and web site quality in e-commerce. *International Journal of Information Management*, 26(6):469–483, 2006.

[161] R. Likert. A technique for the measurement of attitudes. *Archives of Psychology*, 22(140):1–55, 1932.

[162] E. Lima, C. Mues, and B. Baesens. Monitoring and backtesting churn models. *Expert Systems with Applications*, 38(1):975–982, 2010.

[163] E. O. Lima. Domain knowledge integration in data mining for churn and customer lifetime value modelling: New approaches and applications. PhD thesis, University of Southampton, Faculty of Law, Arts and Social Sciences, Southampton, UK, 2009.

[164] C. Lin, S. Wu, and R. J. Tsai. Integrating perceived playfulness into expectation-confirmation model for web portal context. *Information and Management*, 42(5):683–693, 2005.

[165] H. H. Lin and Y. S. Wang. An examination of the determinants of customer loyalty in mobile commerce contexts. *Information and Management*, 43(3):271–282, 2006.

[166] P. J. Lisboa, B. Edisbury, and A. Vellido. *Business Applications of Neural Networks*. World Scientific Publishing Co, 2000.

[167] P. J. Lisboa, A. Vellido, and H. Wong. Bias reduction in skewed binary classification with Bayesian neural networks. *Neural Networks*, 13(4-5):407–410, 2000.

[168] P. J. Lisboa, A. Vellido, R. Tagliaferri, F. Napolitano, M. Ceccarelli, J. D. Martin-Guerrero, and E. Biganzoli. Data mining in cancer research. *IEEE Computational Intelligence Magazine*, 5(1): 14–18, 2010.

[169] C. T. Liu, Y. M. Guo, and C. H. Lee. The effects of relationship quality and switching barriers on customer loyalty. *International Journal of Information Management*, 31(1):71–79, 2011.

[170] D. Liu and Y. Shih. Integrating AHP and data mining for product recommendation based on customer lifetime value. *Information and Management*, 42(3):387–400, 2005.

[171] D. Liu and Y. Shih. Hybrid approaches to product recommendation based on customer lifetime value and purchase references. *The Journal of Systems and Software*, 77(2):181–191, 2005.

[172] H. Liu and H. Motoda. *Computational Methods of Feature Selection*. Chapman and Hall / CRC Data Mining and Knowledge Discovery Series, 2007.

[173] D. J. Mackay. The evidence framework applied to classification networks. *Neural Computation*, 4(5):720–736, 1992.

[174] D. J. Mackay. Bayesian methods for back-propagation networks. In *Models of Neural Networks III*, pages 211–254. Spinger, New York, 1994.

[175] G. Madden, S. J. Savage, and G. Coble-Neal. Subscriber churn in the australian ISP market. *Information Economics and Policy*, 11:195–207, 1999.

[176] A. Malhotra and C. K. Malhotra. Exploring switching behavior of US mobile service customers. *Journal of Service Marketing*, 27(1):13–24, 2013.

[177] M. F. Maute and W. R. Forrester. The structure and determinants of consumer complaint intentions and behaviour. *Journal of Economic Psychology*, 14(2):219–47, 1993.

[178] G. McLachlan and D. Pell. Finite Mixture Models. In *Series in Probability and Statistics*. Wiley-Blackwell, 2000.

[179] M. L. Meuter, A. L. Ostrom, R. I. Roundtree, and M. J. Bitner. Self-service technologies: understanding customer satisfaction with technology-based service encounters. *Journal of Marketing*, 64(3):50–64, 2000.

[180] L. S. Meyers, A. Guarino, and G. Gamst. *Applied Multivariate Research: Design and Interpretation*. Sage Publications, 2005.

[181] G. Mihelis, E. Grigoroudis, Y. Siskos, Y. Politis, and Y. Malandrakis. Customer satisfaction measurement in the private bank sector. *European Journal of Operational Research*, 130(2):347–360, 2001.

[182] R. Miikkulainen, J. A. Bednar, Y. Choe, and J. Sirosh. *Computational Maps in the Visual Cortex*. Springer, 2005.

[183] R. M. Morgan and S. D. Hunt. The commitment-trust theory of relationship marketing. *Journal of Marketing*, 58(3):20–38, 1994.

[184] M. C. Mozer, R. Wolniewicz, D. B. Grimes, E. Johnson, and H. Kaushansky. Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on Neural Networks*, 11(3):690–696, 2000.

[185] F. Mulier and V. Cherkassky. Self-organization as an iterative kernel smoothing process. *Neural Computation*, 7(6):1165–1177, 1995.

[186] E. Naumann. *Creating Customer Value: The Path to Sustainable Competitive Advantage*. Thomson Executive Press, Cincinnati, OH, 1995.

[187] S. A. Neslin and S. Gupta. Defection detection: Improving predictive accuracy of customer churn models. *Journal of Marketing Research*, 43(2):204–211, 2006.

[188] K. Ng and H. Liu. Customer retention via data mining. *Artificial Intelligencie Review*, 16(4):569–590, 2000.

[189] G. Nie, L. Zhang, X. Li, and Y. Shi. The analysis on the customers churn of charge email based on data mining. In *Sixth IEEE International Conference on Data Mining (ICDM)*, pages 843–847. Springer-Verlang, Hong Kong, China, 2006.

[190] G. Nie, G. Wang, P. Zhang, Y. Tian, and Y. Shi. Finding the hidden pattern of credit card holder's churn: a case of China. In *Computational Science - ICCS 2009*, pages 561–569. Springer-Verlag, Berlin, 2009.

[191] G. Nie, W. Rowe, L. Zhang, Y. Tian, and Y. Shi. Credit card churn forecasting by logistic regression and decision tree. *Expert Systems with Applications*, 38(12):15273–15285, 2011.

[192] T. P. Novak, D. L. Hoffman, and Y. F. Yung. Measuring the flow construct in online environment; A structural modeling approach. *Marketing Science*, 19(1):22–44, 2000.

[193] A. Okabe, B. Boots, K. Sugihara, and S. N. Chiu. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Wiley-Blackwell, 2nd edition, 2000.

[194] L. Olier and A. Vellido. Comparative assessment of the robustness of missing data imputation through Generative Topographic Mapping. In *Computational Intelligence and Bioinspired Systems*, Cabestany, J., Prieto, A., and Sandoval, F. (Eds) Proceedings of the 8th International Work-Conference on Artificial Neural Networks (IWANN), Vilanova i la Geltrú, Barcelona, Spain, 2005.

[195] L. Olier and A. Vellido. Capturing the dynamics of multivariate time series through visualization using generative topographic mapping through time. In 1st *IEEE International Conference on Engineering of Intelligent Systems (ICEIS 2006)*, Islamabad, Pakistan, 2006.

[196] C. Orphanidou, S. J. Roberts, and L. M. Moroz. Voice morphing using the Generative Topographic Mapping. In *International Conference on Computer, Communication and Control Technologies*, pages 222–225, 2003.

[197] K.-M. Osei-Bryson. Evaluation of decision trees: a multi criteria approach. *Computers and Operations Research*, 31(11):1933–45, 2004.

[198] S. K. Pal and A. Ghosh. Soft computing data mining. *Information Sciences*, 163(1-3):5–12, 2004.

[199] P. Palvia. The role of trust in e-commerce relational exchange: A unified model. *Information and Management*, 46(4):213–220, 2009.

[200] A. Parasuraman, V. A. Zeithaml, and L. L. Berry. A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49 (4):41–50, 1985.

[201] A. Parasuraman, V. A. Zeithaml, and L. L. Berry. SERVQUAL a multiple item scale for measuring consumer perception of service quality. *Journal of Retailing*, 64 (1):12–37, 1988.

[202] A. Parasuraman, L. L. Berry, and V. A. Zeithaml. Refinement and measurement of the SERVQUAL scale. *Journal of Retailing*, 67(4):420–50, 1991.

[203] A. Parasuraman, V. A. Zeithaml, and A. Malhotra. e-SERVQUAL: a multiple-item scale for assessing electronic service quality. *Journal of Service Research*, 7(3):213–33, 2005.

[204] M. P. Patterson and T. Smith. A cross-cultural study of switching barriers and propensity to stay with service providers. *Journal of Retailing*, 79(2):107–120, 2003.

[205] F. V. Paulovich, D. M. Eler, J. Poco, C. P. Botha, R. Minghim, and L. G. Nonato. Piecewise Laplacian-based projection for interactive data exploration and organization. In *EuroVis 2011 Proceedings of the 13th Eurographics / IEEE - VGTC Conference on Visualization*, pages 1091–1100, 2011.

[206] D. Peel and G. J. McLachlan. Robust mixture modelling using the t-distribution. *Statistics and Computing*, 10(4):339–348, 2000.

[207] W. D. Penny and S. J. Roberts. Bayesian neural networks for classification: How useful is the evidence framework? *Neural Networks*, 12(6):877–892, 1999.

[208] P. E. Pfeifer and R. L. Carraway. Modeling customer relationships as Markov chains. *Journal of Interactive Marketing*, 14(2):43–55, 2000.

[209] D. Phan, L. Xiao, R. Yeh, P. Hanrahan, and T. Winograd. Flow map layout. In *IEEE Symposium of Information Visualization (InfoVis'05)*, pages 219–224, 2005.

[210] R. A. Ping. The effects of satisfaction and structural constraints on retailer exiting, voice, loyalty, opportunism, and neglect. *Journal of Retailing*, 69 (3):321–349, 1993.

[211] J. S. Pointer. The cortical magnification factor and photopic vision. *Biological Review*, 61(2):97–119, 1986.

[212] M. Porter. *Competitive Strategy, Techniques for Analyzing Industries and Competitors*. Macmillan, New York, 1980.

[213] A. Prinzie and D. Van den Poel. Investigating purchasing-sequence patterns for financial services using Markov, MTD and MTDg models. *European Journal of Operational research*, 170(3):710–34, 2006.

[214] P. Provost, T. Fawcett, and R. Kohavi. The case against accuracy estimation for comparing induction algorithms. In *15th International Conference on Machine Learning (ICML 1998)*, pages 445–453. Morgan Kaufman, Madison, Wisconsin, 2000.

[215] M. Pura. Linking perceived value and loyalty in location-based mobile services. *Managing Service Quality*, 15(6):509–538, 2005.

[216] C. Ranaweera and A. Neely. Some moderating effects on the service quality-customer relation link. *International Journal of Operations and Production Management*, 23(2):230–248, 2003.

[217] A. Rauber, D. Merkl, and M. Dittenbach. The growing hierarchical self-organizing map: exploratory analysis of high-dimensional data. *IEEE Transactions on Neural Networks*, 13(6):1331–1341, 2002.

[218] P. Rauyruen and K. E. Miller. Relationship quality as a predictor of B2B customer loyalty. *Journal of Business Research*, 60(1):21–31, 2007.

[219] G. Rong, Y. Liu, W. Wang, X. Yin, X. D. Gu, and X. Guo. GPU-assisted computation of centroidal Voronoi tessellation. *IEEE Transactions on Visualization and Computer Graphics*, 17(3):345–356, 2011.

[220] F. Rossi. Visual data mining and machine learning. In *Proceedings of the 14th European Symposium on Artificial Neural Networks (ESANN)*, pages 251–264, Brussels, Belgium, 2006. d-side publications.

[221] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[222] C. Ruíz Moreno and A. Picón Berjoyo. La importancia del valor percibido y los costes de cambio en el marketing de relaciones. In *La Empresa Familiar en un Mundo Globalizado*, volume 4, pages 53–64. Universidad Santiago de Compostela, Lugo, Spain, 2003.

[223] R. T. Rust and R. L. Oliver. *Service Quality: New Directions in Theory and Practice*. Sage Publications, Inc., New York, 1994.

[224] J. Rygielski, J. Wang, and D. C. Yen. Data mining techniques for customer relationship management. *Technology in Society*, 24:483–502, 2002.

[225] R. Schmalensee. Product differentiation advantages of pioneering brands. *American Economic Review*, 72(3):349–65, 1982.

[226] B. Schoelkopf, A. Smola, and K. Muller. Nonlinear component analysis as a kernel cigenvalue problem. Technical report, Max-Planck Institut fur biolgische Kybernetik, 1996.

[227] V. Shankar, A. K. Smith, and R. A. Customer satisfaction and loyalty in online and offline environments. *International Journal of Research in Marketing*, 20(2):153–175, 2003.

[228] N. Sharma and M. P. Patterson. The impact of communication effectiveness and service quality on relationship commitment in consumer, professional services. *Journal of Services Marketing*, 13(2): 151–70, 1999.

[229] C. Shearer. The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing*, 5(4):13–22, 2000.

[230] J. Sheth, B. Newman, and B. Gross. *Consumption Values and Market Choices, Theory and Applications*. South-Western Publishing, Fort Knox, TX, 1991.

[231] H. W. Shin and S. Y. Sohn. Segmentation of stock trading customers according to potential value. *Expert Systems with Applications*, 27(1):27–33, 2004.

[232] Y. Skadberg and J. R. Kimmel. Visitors' flow experience while browsing a web site: Its measurement, contributing factors and consequences. *Computers in Human Behavior*, 20(3):403–422, 2004.

[233] S. F. Slater and J. C. Narver. Intelligence generation and superior customer value. *Journal of the Academy of Marketing Science*, 28(1):120–127, 2000.

[234] T. A. Slocum. *Thematic Cartography and Visualization*. Prentice Hall, New Jersey, U.S.A, 1998.

[235] S. A. Slotnick and M. J. Sobel. Manufacturing lead - time rules: Customer retention versus tardiness costs. *European Journal of Operational Research*, 163(3):825–856, 2005.

[236] J. Steenkamp and H. Baumgartner. The role of optimum stimulation level in exploratory consumer behavior. *Journal of Consumer Research*, 19 (3):434–448, 1992.

[237] P. Steyn, L. Pitt, A. Strasheim, C. Boshoff, and R. Abratt. A cross-cultural study of the perceived benefits of a retailer loyalty scheme in Asia. *Journal of Retailing and Consumer Services*, 17(5): 355–373, 2010.

[238] Y. Sun, P. Tino, and I. Nabney. GTM-based data visualization with incomplete data. Technical report, Aston University, 2001.

[239] Z. Sun, G. Bebis, and R. Miller. Object detection using feature subset selection. *Pattern Recognition*, 37(11):2165–2176, 2004.

[240] K. Suryadi and S. Gumilang. Actionable decision model in customer churn monitoring based on support vector machines technique. In 9th *Asia Pacific Industrial Engineering and Management Systems Conference*, Bandung, Indonesia, 2008.

[241] M. Svensén. GTM: The Generative Topographic Mapping. PhD thesis, Aston University, 1998.

[242] J. C. Sweeney and G. N. Soutar. Consumer perceived value: the development of a multiple item scale. *Journal of Retailing*, 77(2):203–220, 2001.

[243] J. C. Sweeney, G. N. Soutar, and L. W. Jonson. The role of perceived risk in the quality-value relationship: A study in retail environment. *Journal of Retailing*, 75(1):77–105, 1999.

[244] J. W. Thibault and H. H. Kelly. *The Social Psychology of Groups*. Wiley, New York, 1959.

[245] R. Tibshirani, G. Walther, and T. Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society*, 63(2):411–423, 2001.

[246] P. Tino and I. Nabney. Hierarchical GTM: Constructing localized nonlinear projection manifolds in a principled way. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):639–656, 2002.

[247] A. Tiwari, J. Hadden, and C. Turner. A new neural network based customer profiling methodology for churn prediction. In *ICCSA 2010*, chapter IV, pages 358–369. Springer-Verlag, Berlin, 2010.

[248] W. R. Tobler. Thirty-five years of computer cartograms. *Annals of the Association of American Geographers*, 94(1):58–73, 2004.

[249] A. Tosi and A. Vellido. Cartogram representation of the batch-SOM magnification factor. In *Proceedings of the* 20th *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, pages 203–208, Bruges, Belgium, 2012.

[250] C. F. Tsai and Y. H. Lu. Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, 36(10):12547–12553, 2009.

[251] L. R. Turnbull and D. T. Wilson. Developing and protecting profitable customer relationships. *Industrial Marketing Management*, 18 (3):233–238, 1989.

[252] A. Ultsch. Self-organizing neural networks for visualization and classification. In *Proceedings of the* 16th *Annual Conference of the Geselleschaft fur Klassification e.V.*, Dortmund, Germany, 1992.

[253] A. Ultsch and F. Morchen. ESOM-Maps: tools for clustering, visualization, and classification with emergent SOM. Technical report, CS Department, Philipps-University Marburg, Germany, 2005.

[254] D. Van den Poel. Predicting mail-order repeat buying: which variables matter? *Tijdschrift voor Economie and Management*, 48(3):371–403, 2003.

[255] D. Van den Poel and B. Laraviére. Customer attrition analysis for financiat services using proportional hazard models. *European Journal of Operational research*, 157(1):196–217, 2004.

[256] S. Varki and M. Colgate. The role of price perceptions is an integrated model of behavioral intentions. *Journal of Service Research*, 3(3):232–40, 2001.

[257] R. Vázquez-Carrasco and G. R. Foxall. Positive vs negative switching barriers: the influence of service consumers' need for variety. *Journal of Consumer Behavior*, 5(4):367–79, 2006.

[258] A. Vellido. Preliminary theoretical results on a feature relevance determination method for Generative Topographic Mapping. Technical report LSI-05-13-R, Universitat Politecnica de Catalunya, Barcelona, Spain, 2005.

[259] A. Vellido. Missing data imputation through GTM as a mixture of t-distributions. *Neural Networks*, 19(10):1624–1635, 2006.

[260] A. Vellido. Assessment of an unsupervised feature selection method for Generative Topographic Mapping. In *Proceedings of the 16th International Conference on Artificial Neural Networks (ICANN 2006)*, Lecture Notes in Computer Science, Athens, Greece, 2006.

[261] A. Vellido and P. J. Lisboa. Handling outliers in brain tumour MRS data analysis through robust topographic mapping. *Computers in Biology and Medicine*, 36(10):1049–1063, 2006.

[262] A. Vellido and P. J. G. Lisboa. An electronic commerce application of the Bayesian framework for MLPs: the effect of marginalization and ARD. *Neural Computing & Applications*, 10(1):3–11, 2001.

[263] A. Vellido, P. J. Lisboa, and J. Vaughan. Neural networks in business: a survey of applications (1992-1998). *Expert Systems with Applications*, 17(1):51–70, 1999.

[264] A. Vellido, P. J. Lisboa, and K. Meehan. The Generative Topographic Mapping as a principled model for data visualization and market segmentation: an electronic commerce case study. *International Journal of Computers, Systems and Signals*, 1(2):119–138, 2000.

[265] A. Vellido, W. El-Deredy, and P. J. Lisboa. Studying embedded human EEG dynamics using Generative Topographic Mapping. Technical report LSI-04-8-R, Universitat Politecnica de Catalunya (UPC), Barcelona, Spain, 2004.

[266] A. Vellido, P. J. Lisboa, and D. Vicente. Handling outliers and missing data in brain tumor clinical assessment using t-GTM. In *European Symposium on Artificial Neural Networks (ESANN 2005)*, pages 121–126, Bruges, Belgium, 2005.

[267] A. Vellido, T. A. Etchells, D. L. García, and À. Nebot. Describing customer loyalty to Spanish petrol stations through rule extrction. In *Proceedings of the 7th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2006)*, Lecture notes in Computer Science, pages 970–977, Burgos, Spain, 2006.

[268] A. Vellido, P. J. Lisboa, and D. Vicente. Robust analysis of MRS brain tumor data using t-GTM. *Neurocomputing*, 69(7-9):754–768, 2006.

[269] A. Vellido, E. Romero, F. F. Gonzalez-Navarro, L. Belanche-Munoz, M. Juliá-Sapé, and C. Arús. Outlier exploration and diagnostic classification of a multi-centre 1H-MRS brain tumour database. *Neurocomputing*, 72(13-15):3085–3097, 2009.

[270] A. Vellido, J. D. Martín, F. Rossi, and P. J. Lisboa. Seeing is believing: The importance of visualization in real-world machine learning applications. In *Proceedings of the 19th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*, pages 219–226, Brussels, Belgium, 2011. d-side publications.

[271] A. Vellido, J. D. Martín-Guerrero, and P. J. Lisboa. Making machine learning models interpretable. In *ESANN 2012*, pages 163–172, Brussels, Belgium, 2012. d-side pub.

[272] A. Vellido, D. García, and À. Nebot. Cartogram visualization for nonlinear manifold learning models. *Data Mining and Knowledge Discovery*, 27(1):22–54, 2013.

[273] J. Venna. Dimensionality Reduction for Visual Exploration of Similarity Structures. Doctoral thesis, Helsinki University of Technology, Espoo, Finland, 2007.

[274] W. Verbeke, K. Dejaeger, D. Martens, J. Hur, and B. Baesens. New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, 218:211–219, 2011.

[275] P. C. Verhoef and B. Donkers. Predicting customer potential vualue, an application in the insurance industry. *Decision Support Systems*, 32:189–99, 2001.

[276] P. C. Verhoef, P. N. Spring, J. C. Hoekstra, and P. S. Leeflang. The commercial use of segmentation and predictive modelling techniques for database marketing in The Netherlands. *Decision Support Systems*, 34:471–81, 2002.

[277] J. Vesanto and E. Alhioniemi. Clustering of the Self-Organizing Map. *IEEE Transactions on Neural Networks*, 11(3):586–600, 2000.

[278] T. Villmann and J. C. Claussen. Magnification control in Self-Organizing Maps and neural gas. *Neural Computation*, 18(2):446–469, 2006.

[279] R. G. Wahlers and M. J. Etzel. A structural examination of two optimal stimulation level measurement models. *Advances in Consumer Research*, 17:415–425, 1990.

[280] G. Wang, L. Liu, Y. Peng, G. Nie, G. Kou, and Y. Shi. Predicting credit card holder churn in banks of China using data mining and MCDM. In *IEEE / WIC / ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 215–218, 2010.

[281] H. Wassle, U. Grunert, J. Rohrenbeck, and B. B. Boycott. Retinal ganglion cell density and cortical magnification factor in the primate. *Vision Research*, 30(11):1897–1911, 1990.

[282] C. P. Wei and L. T. Chiu. Turning telecommunications call details to churn preduction: a data mining approach. *Expert Systems with Applications*, 23(2):103–12, 2002.

[283] A. Weinstein and W. C. Johnson. *Designing and Delivering Superior Customer Value*. CRC Press LLC, Florida, 1999.

[284] A. Weiss and E. Anderson. Converting from independent to employee salesforces: The role of perceived switching costs. *Journal of Marketing Research*, 29(1):101–115, 1992.

[285] C. K. I. Williams. Combining deformable models and neural networks for handwrite digit recognition. PhD thesis, Dept. of Computer Science, University of Toronto, 1994.

[286] I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufman, San Francisco, 2nd edition, 2005.

[287] D. Wu. Supplier selection: A hybrid model using DEA, decision tree and neural network. *Expert Systems with Applications*, 36(5):9105–9112, 2009.

[288] L. Yan, R. H. Wolniewicz, and R. Dodier. Predicting customer behaviour in telecommunications. *IEEE Intelligent Systems*, 19(2):50–58, 2004.

[289] J. Yang and B. T. Zhang. Customer data mining and visualization by generative topographic methods. Technical report, School of Computer Science and Engineering, Seoul National University, Seoul, Korea, 2002.

[290] V. Yeswanth, V. Vimal Raj, and M. Saravanan. Evolutionary churn predition in mobile networks using hybrid learning. In *Proceedings of the 24th International Florida Artificial Intelligence Research Society Conference (FLAIRS)*, pages 471–476, Florida, USA, 2011.

[291] C. Yin, D. K. Tse, and K. W. Chan. Strengthening customer loyalty through intimacy and passion: Roles of customer-firm affection and customer-staff relationships in services. *Journal of Marketing Research*, 45(6):741–756, 2008.

[292] V. A. Zeithaml. How Consumer Evaluation Processes Differ between Goods and Services. In *Marketing of Services*, pages 186–90. American Marketing Association, Chicago, IL, 1981.

[293] V. A. Zeithaml. Consumer perceptions of price, quality and value: A means-end model and synthesis of evidence. *Journal of Marketing*, 52(3):2–22, 1988.

[294] V. A. Zeithaml, A. Parasuraman, and L. L. Berry. Problems and strategies in services marketing. *Journal of Marketing*, 49(2):33–46, 1985.

[295] V. A. Zeithaml, L. L. Berry, and A. Parasuraman. The behavioral consequences of service quality. *Journal of Marketing*, 60:31–46, 1996.

[296] Y. Zhao, B. Li, X. Li, W. Liu, and S. Ren. Customer churn prediction using improved one-class support vector machine. In X. Li, S. Wang, and Z. Yang-Dong, editors, *Advanced Data Mining and Applications*, Lecture Notes in Computer Science Volume 3584, pages 300–306. Springer, 2005.

[297] C. Ziemkiewicz and R. Kosara. Preconceptions and individual differences in understanding visual metaphors. *Computer Graphics Forum*, 28(3):911–918, 2009.