View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by ZHAW digitalcollection

Churn prediction based on text mining and CRM data analysis

Anders Schatzmann[1], Christoph Heitz[2], Thomas Münch[3]

[1] Zurich University of Applied Sciences, School of Engineering, Winterthur
[2] Zurich University of Applied Sciences, School of Engineering, Winterthur
[3] upc cablecom, Zurich

Abstract

Within quantitative marketing, churn prediction on a single customer level has become a major issue. An extensive body of literature shows that, today, churn prediction is mainly based on structured CRM data. However, in the past years, more and more digitized customer text data has become available, originating from emails, surveys or scripts of phone calls. To date, this data source remains vastly untapped for churn prediction, and corresponding methods are rarely described in literature.

Filling this gap, we present a method for estimating churn probabilities directly from text data, by adopting classical text mining methods and combining them with state-of-the-art statistical prediction modelling. We transform every customer text document into a vector in a high-dimensional word space, after applying text mining pre-processing steps such as removal of stop words, stemming and word selection. The churn probability is then estimated by statistical modelling, using random forest models. We applied these methods to customer text data of a major Swiss telecommunication provider, with data originating from transcripts of phone calls between customers and call-centre agents.

In addition to the analysis of the text data, a similar churn prediction was performed for the same customers, based on structured CRM data. This second approach serves as a benchmark for the text data churn prediction, and is performed by using random forest on the structured CRM data which contains more than 300 variables.

Comparing the churn prediction based on text data to classical churn prediction based on structured CRM data, we found that the churn prediction based on text data performs as well as the prediction using structured CRM data. Furthermore we found that by combining both structured and text data, the prediction accuracy can be increased up to 8%.

These results show clearly that text data contains valuable information and should be considered for churn estimation.

Keywords
churn, churn prediction, text mining, text data, random forest, CRM

1 Introduction

In the context of analytical marketing, customer churn prediction becomes a major issue for firms. As it is known that retaining customers is far more profitable than acquiring new ones, there is considerable focus on retention campaigns (Henning-Thurau & Hansen, 2010). If the churn probabilities on a single customer level are known, marketing can focus its retention programs on the customers with high churn probability, thus increasing the efficiency of those programs.

Personalized customer data can be divided into two major different data types. One is the structured data, which covers information about the customer typically stored in CRM data bases. The other type is text data, also referred to as unstructured data. Text data may originate from emails or transcripts of phone calls with call centres, or other customer contacts over digitised channels.

For churn prediction, the analysis of structured data is the classical approach. Structured data has been the primary information hub for the past decades (Ngai, Xiu, & Chau, 2009). As more and more customer text data is becoming available, one might argue that using this data for churn prediction could improve the churn models. Methods for analysing text data are available from a rich body of literature in the field of computer science. Covering machine-based analysis of texts and different approaches of extracting information, have been developed (Feldman & Sanger, 2006).

As in the past mainly structured data was used for churn prediction, the goal of this paper is to investigate how customer text data can be used for churn prediction, and whether using text data in addition to structured data can influence the churn prediction accuracy. To analyse this problem, we use a test data set with anonymized customer data from a major Swiss telecommunication provider. The data set contains both structured and unstructured (text) data for more than 20'000 randomly selected customers. First we analysed the structured and unstructured customer data separately and compared the results. Second we combined the two data sources to investigate whether the combination leads to an increase in the churn prediction accuracy.

The structured data is taken from the provider's operative systems, such as CRM or ticketing systems. This data has been used previously for churn prediction by the firm. The text data consists of anonymized emails and transcripts of phone calls between customers and call centre agents for matters of questions, complaints and administrative reasons. This data has not yet been used for churn prediction.

After the introduction, this paper provides an overview over the related literature, followed by a description of the theoretical methodology used for the churn prediction. Then a description of the available customer data is given, complemented by the application and the results of the methodology on the data. A final conclusion sums up the paper.

2 Related literature

In recent years, a large number of machine learning and knowledge discovery techniques have been proposed and applied to the problem of customer retention in the domain of CRM (Berson, Smith, & Thearling, 2000). Originally introduced in the finance sector, customer retention has found its path into other fields, such as telecommunication. Within customer retention, the task of identifying the customers most likely to churn is of crucial importance (Keaveney & Parthasarathy, 2001)

With CRM becoming a critical success factor in today's business environment, academic research produced a vast number of articles covering all areas of CRM. Especially applying data mining methods in order to gain customer knowledge is well-covered in literature. (Ngai, Xiu, & Chau, 2009) present an extensive literature study for data mining techniques in CRM. They revise more than 80 papers, published between 2000 and 2007, many of them covering the domain of customer churn prediction.

Most papers concerning customer churn, as collected by (Ngai, Xiu, & Chau, 2009) focus on structured data and various data mining techniques such as decision trees (Xie, Li, Ngai, & Ying, 2009), logistic regression, support vector machine (Yu, Guo, Guo, & Huang, 2011), artificial neural networks, etc. Most of those methods are based on supervised learning and use a single prediction model. In the past years, those single model approaches have been replaced by hybrid classification models with the goal of increasing the prediction accuracy ((Huang & Kechadi, 2013); (Khashei, Hamadani, & Bijari, 2012); (Tsai & Lu, 2009); (Lee & Lee, 2006); (Coussement, Benoit, & Van den Poel, 2010); (De Bock & Van den Poel, 2011); (De Bock & Van den Poel, 2012)).

Parallel to churn prediction using structured data, approaches of integrating the *Voice of Customers* (VOI) to CRM have been developed. The data for VOI analytics is gained through direct or indirect questioning, with data being either structured data from surveys or unstructured (text) data from emails, transcripts or free text answers in surveys ((Aguwa & Monplaisir, 2012); (Chang, Lin, & Wang, 2009)).

Labour-intensive manual text mining approaches first surfaced in the mid-1980's, but during the past two decades this field has advanced drastically, accompanied by technological advances, especially in computer science. Through these advances, text mining has gained vast attention throughout science and business. Several methods have been developed to classify texts and analyse content for unsupervised mapping of texts (Kao & Poteet, 2006).

Despite the extension of text mining, it has rarely been applied in CRM. Text mining methods for customer churn prediction are mostly non-existent in literature. One example of text mining applied to churn prediction was published by (Coussement & Van den Poel, 2008). They combine text mining with the analysis of structured data. Their research showed that the churn prediction accuracy can be improved by combining the two data sources. With a follow up paper, (Coussement & Van den Poel, 2009) integrate the emotions from client/company interaction emails, in order to improve customer attrition. In this paper the focus is on improving the churn prediction accuracy by combining structured and unstructured data. Our method follows the same approach but uses a different statistical prediction method and a different feature selection.

3 Methodology

In this section, the methods used for getting from the original customer data to the final churn prediction are presented. The methods are described on a theoretical basis, and their application on the real data will then be described in section 0.

Figure 1 shows the complete process from the data sources to the final churn prediction. The first step is to collect the data from the data bases. After having established the data basis, several pre-processing steps are necessary in order to structure the original data, so that it can be used in a prediction model. The last step is to apply a suitable statistical model on the pre-processed data to estimate the churn probabilities for each customer. The methods underlying the steps are described in detail in the following paragraphs.

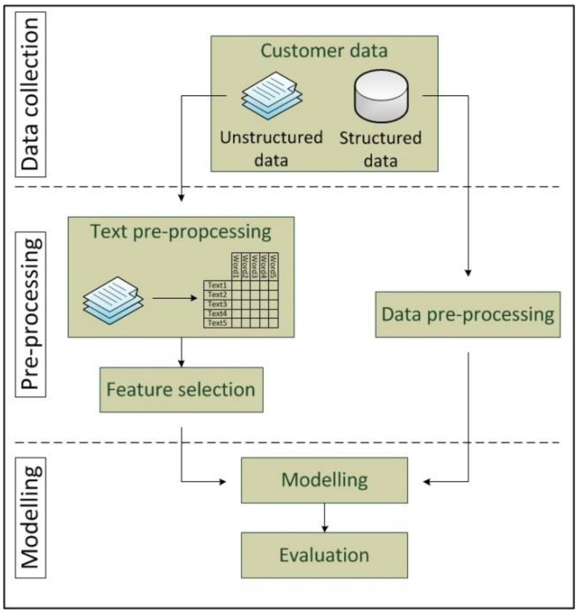

Figure 1: Process model of churn predicting process, which includes structured and unstructured data

3.1 Structured and unstructured Data

Structured data is defined as data that resides in fixed fields within a record or file. Relational databases and spreadsheets are examples of structured data (Enterprise, 2014). Structured customer data is usually stored CRM systems and typically includes personal information, subscribed services and/or products and sociodemographic information. In contrast, unstructured data refers to information that does not reside in traditional row-column database structures. Unstructured data files include text and multimedia content.

Examples are emails, text documents and further forms of texts (Enterprise, 2014). In unstructured data the information content is not stored within a specific field, but is hidden in the content of the text itself. Gaining information from texts requires extensive pre-processing steps in order to reveal its information content.

For training statistical prediction models with supervised learning, the training data has to be labelled. For the case of churn prediction, each customer is assigned to either the class *churn* or the class *no-churn*, based on the observed behaviour in the past.

3.2 Pre-Processing

Pre-processing of the structured data includes checking of consistency and relevance of the variables. The relevance of each data field is to be checked due to the usually large quantity of data available in the CRM system, and a variable selection has to be made. Furthermore there might be the need to combine and convert the raw data into new variables with more valuable information content.

The pre-processing of unstructured data is necessary for structuring the text data in a way that allows the included information content to be analysed by a certain method. According to (Hippner & Rentzmann, 2006), text pre-processing can be split into three main categories. The approach referred to as *Morphologic Approach* is likely to be the most widely spread approach in text mining because of its low complexity, high quality of the results and computational cheapness. The two other approaches called *Syntactic Analysis* and *Semantic Analysis* are of a much higher complexity. The morphologic approach focuses on simply counting the words occurring in a text, whereas the syntactic approach analyses the relationship between words within sentences, with the focus on extracting information on certain syntactic units. The semantic analysis tries to understand the text, comparable to what humans do while reading.

For our investigation, the morphologic approach was chosen because of its simplicity, computational effectiveness and the kind of texts we have. The texts for this project are mainly transcripts of phone calls and emails. While the writing quality of the emails is decent, the transcripts usually consist of many abbreviations, key words and incomplete sentences. Therefore a syntactic or semantic approach would be difficult.

Before applying any pre-processing steps, the texts of the same customer are aggregated. As some customers can have several texts these are merged into one document. Each customer can thus have only one document, but in one document there can be multiple texts. Creating a document is done by simply attaching one text to another. We choose to do so, as the churn prediction is on a customer level rather than on a text level, and therefore each customer should only have one document.

Applying the morphologic approach, all semantic information is neglected. Each document is converted into a high-dimensional vector of weighted frequencies of the occurring words. Thus each document is of the length n, with n being the number of different words occurring in a specific document. The so called *Bag of Words* (BOW) is a representation of all distinct words occurring in all documents and is of size N.

Having disjoined the documents, they are pooled in a *Term Document Matrix* (TDM). In this matrix, each document is represented as a line, and each column corresponds to one word of the bag of words. Each cell of the matrix represents the count of the one specific word in one document. So the TDM is of size DxN with D being the number of documents and N the size of the BOW.

Doc1: The dog walks in the rain
Doc2: The cat walks int he sun
Doc3: The dog walks with the cat

	the	dog	cat	walks	in	sun	rain	with
Doc1	2	1	0	1	1	0	1	0
Doc2	2	0	1	1	1	1	0	0
Doc3	2	1	1	1	0	0	0	1

Figure 2: Example of conversion from text to Term Document Matrix, where Doc stands for document

As a transformation of the raw documents to a TDM typically leads to a very large and only hardly manageable size, several *raw text cleaning* steps are crucial. The aim is to reduce the number of words as much as possible in order to keep the size of the BOW small. Thereafter, further steps of selecting only the relevant words are performed in order to further reduce the size of the TDM.

In the first step of cleaning the documents, special characters and punctuation are removed, followed by the replacement of the acronyms with their radicals, by using a reference list. Next, all the words are replaced with their stem, e.g. *complain* is the stem for *complained*, *complaint*, *complaining*, etc., by using Porters algorithm (Porter, 2006). Stemming drastically reduces the number of words and increases the information retrieval performance (Kraaij & Pohlmann, 1996).

After the stemming process, all stop words are removed. Stop words contain either very little or no information content, e.g. *are*, *the*, *at*, *from*, etc. The removal is done by using a pre-defined list of standard stop words extended with application-specific terms.

The remaining words build the basis for the creation of the TDM. In the process of building the TDM, all the words with very low occurrences, e.g. less than three, are removed as these usually contain clerical errors or are artificial words, not recognized by the stemmer or the stop word list. Having cleaned the texts and represented them in the TDM, the next step is to further reduce the number of words by an appropriate feature selection.

3.3 Feature selection

(Do, Hui, & Fong, 2006) state that feature selection aims on removing irrelevant and noisy information from the data, by focusing on relevant and informative features only. Applied to text mining, the goal is to reduce the number of words in the TDM for further statistical modelling, as the words of the TDM can be understood as features. For churn prediction, only those words having significant information content with respect to the churn probability should be selected.

We use labelled data (*churn* or *no-churn*) as training data. This can be taken advantage of by using a supervised method for the feature selection process. One approach to assess the relevance of the features is to measure their distinguishing ability between the two classes *churn* and *no-churn*. We chose the *Discriminating Power Measure* (DPM) (Chen, Lee, & Chang, 2009), a supervised method which focuses on discriminating words in the context of classification problems.

The DPM score is established by using the following notations. Let *w* be any word, its presence or absence in class *i* is defined as follows:

A_i number of documents with word *w* and belonging to class *i*

B_i number of documents with word *w* and not belonging to class *i*

C_i number of documents without word *w* and belonging to class *i*

D_i number of documents without word *w* and not belonging to class *i*

The total number of documents is $N = A_i + B_i + C_i + D_i$, the total number of documents in class *i* is $M_i = A_i + C_i$ and the total number of classes is denoted by *m*. With these notations, the DPM score for word *w* is defined as follows (Chen, Lee, & Chang, 2009):

$$DPM(w) = \sum_{i=1}^{m} \left| \frac{A_i}{M_i} - \frac{B_i}{N - M_i} \right|$$

The fraction A_i/M_i can be interpreted as the probability of word *w* occurring in a document of class i, where $B_i/(N - M_i)$ is the probability of the same word occurring in a document not belonging to class i. The DPM is the absolute difference of these two probabilities, summed up over all classes. The higher the DPM score for a given word *w*, the more discriminate power is contained in it.

Based on the DPM, the words can be ranked according to their discriminative power. For the text analysis, we only use the first *k* words of this list, thus reducing the BOW to one with a higher selectivity.

The issue remains to define the number *k* of features to be selected. In order to find the optimal number of words for the TDM, a graphical approach is chosen. The words are ordered according to their DPM score and then plotted against their rank. From left to right the graph is typically rapidly decreasing, ending in a long flat tail which is cut off.

3.4 Modelling

The goal of the statistical model is to assign each customer a churn probability, based on the available data. Several methods can solve this problem, such as Naive Bayes, logistic regression, support vector machines etc. We decided to use random forest (Breiman, 2001) because of its easy application and ability to handle big data sets. Despite not being widespread in text mining, random forest has become an often-applied method in data mining. It is also suitable for handling a TDM because it can deal with big numbers of variables.

The random forest model is trained with labelled test data. For the training, we used a five-fold cross-validation: The training data is split into five parts. The model is then fitted on four parts, and the fifth part is used for evaluation of the prediction accuracy. This is repeated five times, so that each data point is predicted once (Hastie, Tibshirani, & Friedmann, 2001).

3.5 Evaluation criteria

Validating the performance of our model is a critical step. We decided to use the lift chart, as is an excellent way to show the performance of models. The lift is a measure of the effectiveness of the predictive model. It is calculated as the ratio between the results obtained with and without the predictive model. The lift chart shows the likelihood of responses from customers based on the predictive model and randomly chosen list of customers. The model is performing well if the response within the target segment is much better than the average for the population as a whole ((Jaffery & Liu, 2009); (Wikipedia , 2014)).

As an example, let’s assume that the data has an average churn rate of 25%, and the model has identified a customer segment with an average churn rate of 75%. Then the lift of that segment is 3.

In the lift chart, the customers are ordered decreasingly according to their predicted churn probability. From that, the lift is estimated continuously over all the customers. The lift is represented by the y-axes and the deciles of the ordered customers are on the x-axis.

4 Empirical case study

4.1 Empirical data

In our study, we used anonymized data, obtained from a major Swiss telecommunication provider. Its product portfolio includes internet, digital TV and digital phone. The customers can choose between several product combinations and sub-products with different pricing models. Customers pay a monthly fee, depending on the subscribed products. The customers are in a contractual setting with the firm and can only end their contract with a two month cancellation period after the first 12 months. When an ordinary customer cancels the contract the services by the firm and the payments by the customer continue until the defined cancelation date.

The company has a structured CRM data base where contract related data is stored together with customer data, sociodemographic data, and usage data. Furthermore, all customer care related emails and transcripts of phone calls to the call centre are stored in a separate database. A customer becomes a churner when the provider receives a contract cancellation letter or phone call and it is clear that the customer cannot be held by any taken retention action.

Anonymized customer data of a six month observation period was selected at six corresponding snapshot dates. At each snapshot date, the customer data up to six months prior to the snapshot date was extracted, and a 30 days survey period after the snapshot date was used to define whether a customer is a churner, thus providing the labelling. Figure 2 illustrates this data selection method.

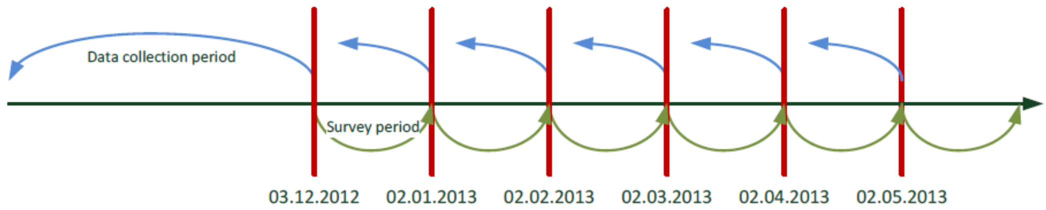

Fig. 2: Data selection process

As the number of churning customers at each snapshot date is small compared to the non-churning customers, an oversampling of the churning customers was applied (Nisbet, Elder, & Miner, 2009). More specifically, all churning customers for each survey period were selected and then complemented with a random sample of non-churning customers, selected at the same snapshot date. This leads to a customer sample on each snapshot date with approximately three times as many non-churning customers. So, the oversampled average churn rate is about 25%.

In order to show the beneficial effect of including unstructured data into the churn prediction model, only customers with existing text data were considered for this case study. Table 1 summarizes the data characteristics.

# of customers	20191
# of churners	5356 (27%)
# of structured variables	305
# of words in TDM	12105

Table 1: Data details

4.2 Data processing

The structured data mainly contains anonymized information about the customer, socio-demographic information, part of usage information, etc. As the provider has used the structured data for churn prediction before, we used the pre-processing and feature selection that has been already done by the provider for building the prediction models.

The unstructured customer data, which consists of anonymized emails and transcripts of phone calls, was pre-processed as described in Section 0. The first step is the raw text cleaning where all punctuation and special characters are removed, followed by the replacement of the acronyms with their radicals. Next the words are reduced to their stem and then the removal of the stop words is done. Following these pre-processing steps, the resulting TDM had 20191 rows (documents) and 12105 columns (words).

Based on the TDM, the DPM score for each word is estimated as a basis for the feature selection. As described in section 0, the determination of the number of features was done by a graphical approach. The DPM score is calculated for each word of the TDM. Then the words are ordered in decreasing order, according to their DPM value, and plotted against their index, where the index is equal to the rank in the order. Figure 3 shows the results for our data set; the x-axis is the rank of the words and the y-axis represents the calculated DPM score. For better readability, the graph is restricted to the 1000 features with the highest DPM value.

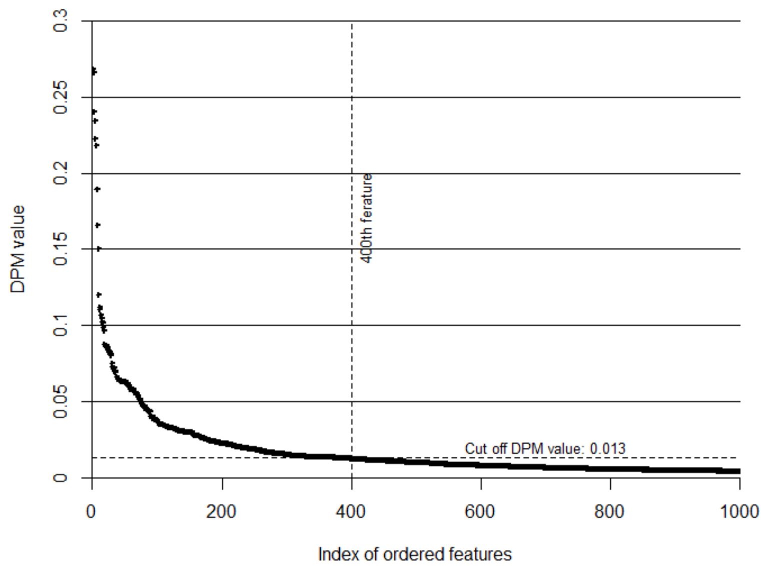

Figure 3: DPM score of the 1000 words with highest DPM value, plotted in decreasing order. The vertical dashed line marks the 400th word and the horizontal dashed lines marks the DPM value at the 400th word.

Figure 3 shows that the most discriminating features have a DPM score of about 0.27 wherefrom the score decreases rapidly. For the features from 1000 to 12000 (not shown in the figure), the drop of the DPM score is less than 0.01. In contrast, the drop for the first 1000 features exceeds 0.25. By using only the first 400 features, 95% of the total DPM range is covered. Based on these considerations, the number of features used for the TDM was set to 400. To further verify this selection, the prediction model was run with several numbers of features and the lift was compared. It turned out that more than 400 features did not increase the lift, while going below 400 features affected the performance negatively.

4.3 Churn prediction

For the churn prediction, two different cases are built. The first case is a comparison of the churn prediction on structured and unstructured data, respectively. The churn prediction on structured data serves as the benchmark as it represents the way churn prediction has been applied to date. The churn prediction on unstructured data illustrates the information content of the text data with respect to predicting churn. The second case is the combination of both structured and unstructured data for churn prediction. All predictions are done by using random forest as a predicting model, using five-fold cross-validation for the performance evaluation.

4.3.1 Churn prediction with structured and unstructured data

Figure 5 shows the lift, using structured and unstructured data separately on the prediction model. The structured data contains 305 variables and the TDM has 400 features, selected by the DPM criteria.

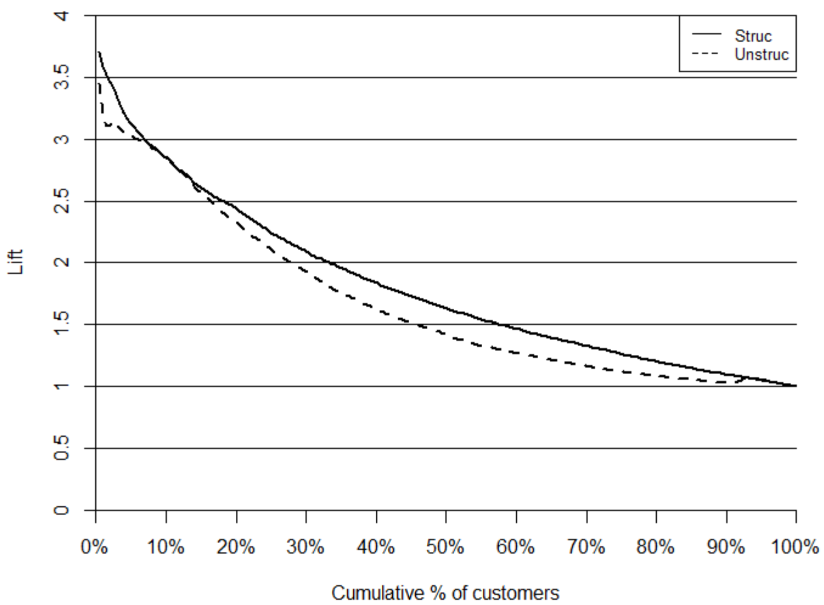

Figure 5: Lift chart of structured and unstructured data, respectively, using random forest. The x-axis shows the cumulative ratio of the ordered model scores and the y-axis shows the lift. The bold line represents the estimated lift of the structured data; the dashed line shows the estimated lift of the unstructured data.

The lift of the structured data starts on a very high level, before decreasing. At the first decile, the average lift is around 2.7, which corresponds to around 70% of correctly classified customers. This means that within the 10% of the customers with the highest predicted churn probability, about 70% actually churn. This lift serves as the benchmark.

The lift of the unstructured data starts high as well, despite not quite reaching the benchmark. At the 10% mark, the lifts become identical, before the unstructured lift loses after the 20% mark compared to the benchmark.

Despite these differences, it can be stated that the amount of information about the churning behaviour is similar in both data sets. Additionally and more generally, it can be stated that the unstructured data clearly holds significant information on the churning behaviour.

4.3.2 Churn prediction, combination of data

Having analysed the two data sources separately, the remaining question is whether the combination of both data sources is able to outperform the benchmark. The combination of the two data sources is straightforward, by adding the two data sets together, creating a new data set, now containing 705 variables (305 structured, 400 words), and applying the same random forest approach to this extended data.

Figure 6 shows the lift of the model, using both data sources, and the lift of the benchmark.

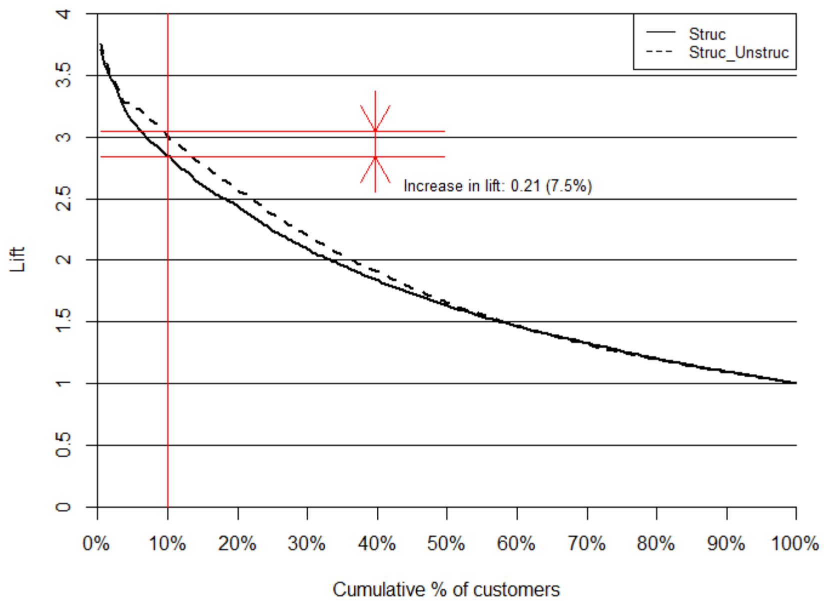

Figure 6: Lift chart of structured data, using random forest. The x-axis shows the cumulative ratio of the ordered model scores and the y-axis shows the lift. The bold line represents the benchmark; the dashed line shows the estimated lift of the combined (structured and unstructured) data. The red lines show the increase of the lift at the first decile

Figure 6 shows clearly that the combined approach exceeds the benchmark significantly. At the 10% level, the combined approach shows an improvement of about 7.5% for the classification precision with respect to the benchmark. This improvement can be attributed to the additional information of the added text data.

5 Conclusions

The main goal of this paper was to investigate whether the combination of structured and unstructured customer data can increase the customer churn prediction precision, compared to using structured data only. We were able to show that the combination of the two data sources does increase the prediction precision of up to 8%. Furthermore we found that the unstructured data itself holds significant information on the churn probability. The information content of the text data is nearly as high as the information content of the structured data.

Based on these results, it can be stated that customer text data does hold information which complements the structured data. Thus, retention marketing campaigns can be directed more exactly by using text data in the churn prediction models, increasing both the efficiency and the effectiveness of the campaigns.

References

Aguwa, C. C., & Monplaisir, L. (2012). Voice of the customer: Customer satisfaction ration based analysis. *Expert Systems with Applicaions*, 10112-10119.

Berson, A., Smith, S. J., & Thearling, K. (2000). *Building Data Mining Applications for CRM.* McGraw-Hill Osborne.

Breiman, L. (2001). Random forests. *Machine Learning*, 5-32.

Chang, C.-W., Lin, C.-T., & Wang, L.-Q. (2009). Mining the text information to optimizing the customer relationship management. *Expert Systems with Applications*, 1433-1443.

Chen, C.-M., Lee, H.-M., & Chang, Y.-J. (January 2009). Two novel feature selection approaches for web page classification. *Expert Systems with Applications*, S. 260-272.

Coussement, K., & Van den Poel, D. (2008). Integrating the voice of customers through call center emails into a decision support system for churn prediction. *Information & Management*, 164-174.

Coussement, K., & Van den Poel, D. (2009). Improving customer attrition prediction by integrating emotions from client/company interaction emails and evaluating multiple classifiers. *Expert Systems and Applications*, 6127-6134.

Coussement, K., Benoit, D. F., & Van den Poel, D. (2010). Improved marketing decision making in a customer churn prediction context using generalized additive models. *Expert Systems with Applications*, 2132-2143.

De Bock, K. W., & Van den Poel, D. (2011). An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Systems with Applications*, 12293-12301.

De Bock, K. W., & Van den Poel, D. (2012). Reconciling performance and interpretability in customer churn prediction using ensemble learning based on generalized additive models. *Expert Systems with Applications*, 6816-6826.

Do, T. D., Hui, S. C., & Fong, A. C. (2006). Associative feature selection for text mining. *International Journal of Information Technology*.

Enterprise, Q. (2014). *webopedia*. Retrieved 05 15, 2014, from structured data: http://www.webopedia.com/TERM/S/structured_data.html

Feldman, R., & Sanger, J. (2006). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data.* Cambridge University Press.

Hastie, T., Tibshirani, R., & Friedmann, J. (2001). *The Elements of Statistical Learning.* New York: Springer New York Inc.

Henning-Thurau, T., & Hansen, U. (2010). *Relationship Marketing: Gaining Competitive Advantage Through Customer Satisfaction and Customer Retention.* Springer.

Hippner, H., & Rentzmann, R. (2006). Text Mining. *Informatik Spektrum*, 287-290.

Huang, Y., & Kechadi, T. (2013). An effective hybrid learning system for telecommunication churn prediction. *Expert Systems with Applications*, 5635-5647.

Jaffery, T., & Liu, S. X. (2009). Measuring Campaign Performance by Using Cumulative Gain and Lift Chart. *SAS Global Forum 2009*, S. Paper 196-2009.

Kao, A., & Poteet, S. R. (2006). *Natural Language Processing and Text Mining.* Berlin: Springer.

Keaveney, S., & Parthasarathy, M. (2001). Customer swiching behavior in online services: an exploratory study of the role of selected attitudinal, behavioral, and demographic factors. *Journal of the Academy of Marketing Science*, pp. 374-390.

Khashei, M., Hamadani, A. Z., & Bijari, M. (2012). A novel hybrid classification model of artificial neural networks and multiple linear regression models. *Expert Systems with Applications*, 2606-2620.

Kraaij, W., & Pohlmann, R. (1996). Viewing Stemming as Recall Enhancement. *In Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, S. 40-48.

Lee, J., & Lee, J. (2006). Customer churn prediction by hyprid model. *Advanced Data Mining and Applications*, 959-966.

Ngai, E., Xiu, L., & Chau, D. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, pp. 2592-2602.

Nisbet, R., Elder, J., & Miner, G. (2009). *Handbook of statistical analysis and data mining applications.* Amsterdam, Boston, Heidelberg: Academic Press-Elsevier.

Porter, M. (2006). *Stemming algorithms for various European Languages*. Retrieved from http://snowball.tartarus.org/texts/stemmersoverview.html

Tsai, C.-F., & Lu, Y.-H. (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, 12547-12553.

Wikipedia . (30. April 2014). Von Lift (data mining): http://en.wikipedia.org/wiki/Lift_%28data_mining%29 abgerufen

Xie, Y., Li, X., Ngai, E., & Ying, W. (2009). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, 5445-5449.

Yu, X., Guo, S., Guo, J., & Huang, X. (2011). An extended support vector machine forecasting framework for customer churn in e-commerce. *Expert Systems with Applications*, 1425-1430.