

Discovering Behavioural Patterns in Internet Log Files: Playing the Devil's Advocate

Alex Büchner* & Maurice Mulvenna*

*Northern Ireland Knowledge Engineering Laboratory, University of Ulster, UK
{ag.buchner, md.mulvenna}@ulst.ac.uk

Abstract

The potential of applying data mining and artificial intelligence techniques to discover behavioural patterns in Internet log files are provocatively outlined. It is shown how national and international regulations have relatively low impact on knowledge discovery, which results in intelligence, hitherto unknown to the marketing community.

1 Introduction

We love our customers. We give them incentives, such as loyalty cards, we provide them with special offers, we put them on our mailing lists to keep them up-to-date, we give them interest-free credits, and we give them their money back if they are not satisfied with what they have purchased. But, do we really love our customers? No, we don't. All we really love about them is their *money*! And now, there is a new customer we love. The innovative shopper on the Internet, the mouse potato who forms part of the customer target group in the electronic commerce scenario. The reason we love our new target group even more is the fact that we can collect every movement in digital format. *Every* movement. The great thing about that is that we can sift through all these log files to discover individual behavioural patterns. Web mining we call this; the synergy between data mining and Internet technology. What is even greater, is the opportunity of cross-matching the footprints and, of course, the knowledge discovered from them, across Internet shops, across Internet shopping malls, across boundaries. One-to-one marketing par excellence. Scared? Huxley and Orwell calling? Shouting for regulations? Forget it! The technology is ready, it works and it's been used already. So, if you are out there, we will get you.

Section 2 briefly recapitulates basic data mining concepts. Section 3 describes how these concepts can be applied in electronic commerce. In Section 4, we discuss some ethical issues, before we conclude, having left the reader, hopefully, with some ideas of how the marketing landscape will change in the very near future.

2 Brief Recapitulation of basic Data Mining Concepts

2.1 Definition

Data Mining is the term given to the automated discovery of non-obvious, potentially useful and previously unknown information from large data sources. It enables industry in different sectors to utilise their most under-utilised resource i.e. data collected by them about various aspects of their businesses, by automatically discovering patterns in their processes which can allow them to gain a competitive edge over their competitors.

2.2 Types of Patterns

Different types of patterns can be detected from given data. The most relevant types for electronic commerce are ([Mul97a]):

- Association rules (relations and dependencies among fields in the data) for basket analysis on the on-line shopping mall,
- Classification rules (mapping data items into one or several predefined categorical classes) for finding potential customers for a certain product or classifying visitors' behaviours,
- Characteristic rules (discovering specifics of one data item) for tackling cross-sales problems,
- Sequential rules (modelling states and patterns of a process) to detect typical paths shoppers tend to chose, and
- Clustering (finding groups of similar entities) to find similar paths of visitors which lead to the same product interest or purchase.

2.3 The Data Mining Process

Data mining is recognised to be a process, rather than a stand alone automated algorithm which discovers knowledge from data without human interference ([Ana98]). A data mining process incorporates data, domain and data mining expertise.

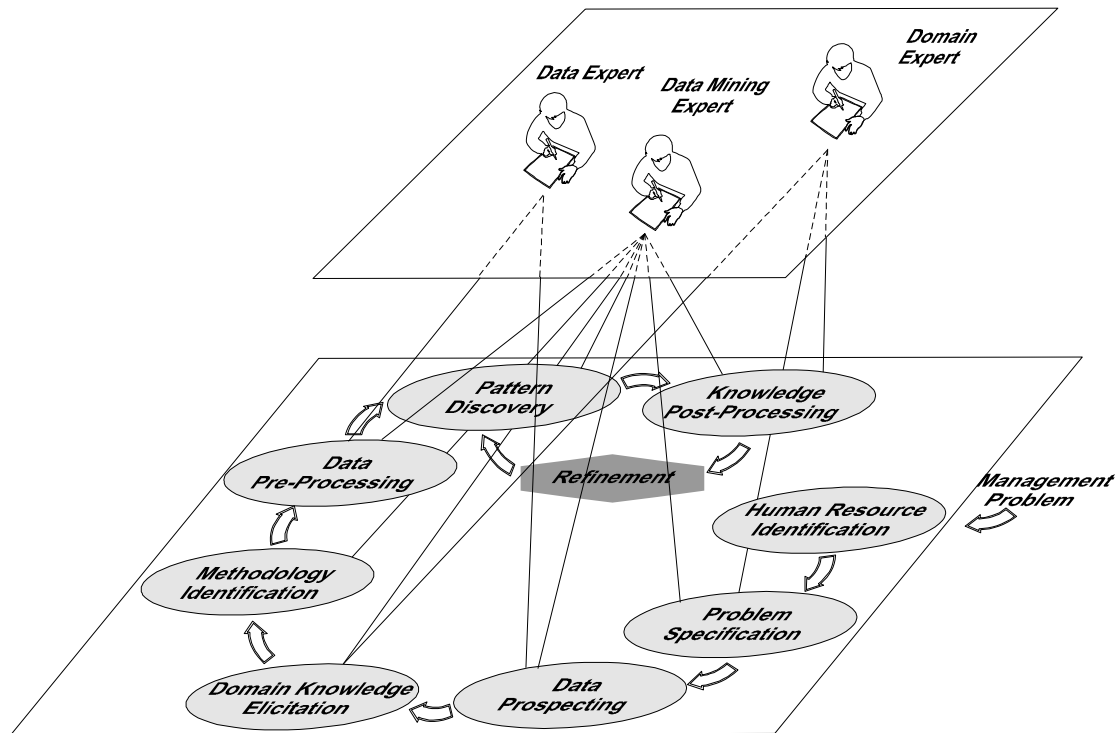


Figure 1: The Data Mining Process

The process is covering the following steps: *human resource identification* is concerned about finding appropriate expertise. In the *problem specification* phase various components of the business solution required as an output from the data mining process are identified. The *data prospecting* stage ensures that all required data (quantitatively and qualitatively) is available and accessible. The *domain knowledge elicitation* incorporates existing knowledge about the tackled problem in the data mining system. The *methodology identification* identifies the methodology utilised to tackle the problem most appropriately. *Pattern discovery* is what is often referred to as the most important data mining step, in which the actual knowledge is discovered. *Knowledge post-processing* is concerned about modifying discovered knowledge so it can be facilitated by the user, e.g. produce natural language like rules. These last two steps usually form a refinement process which has to be iterated through until sufficient results have been achieved.

3 The Data Mining and Electronic Commerce Synergy

So what are the benefits of using data mining techniques to gather intelligence about a market? It allows on-line retailers to 'fine tune' their selling strategy ([Nor97]). This gives them a greatly enhanced insight into the number and types of lines in stock, the best electronic shop front format, and which offers should be directed at which customers and how should they be communicated. The potential for direct marketing is enormous - in theory heavy buyers of a product can be identified and targeted with attractive offers and sales promotions. Targeting does not have to involve discounted offers alone. It could mean making certain categories of customer aware of products or services that might be of interest to them, or inviting them to special on-line events.

The application of data mining techniques to on-line shopping mall data provides high-level knowledge - in the form of patterns - that describes consumer navigational and purchasing behaviour. These rules capture trends and behaviour patterns that may be applied within a marketing strategy. The high-level, descriptive, behavioural rules supply the marketing specialists with the means to implement a *soft-push* marketing strategy ([Mul97b]).

Data mining allows marketers to reveal layers of information about markets or subsets of markets in ever increasing detail, enabling consumer profiles to be built and the identification of profitable and non-profitable segments. The application of data mining techniques to on-line shopping data provides high-level knowledge in the form of rules that describe consumer navigational and purchasing behaviour [Mul98]. These rules capture trends and behaviour patterns that may be applied within a relationship marketing strategy. For example, the high level mined rules may be incorporated as a rule-based system into the architecture of an on-line shopping mall service. When each new customer interacts with the on-line mall to navigate and purchase goods or services, their behavioural patterns are identified by the rule-based system. When this happens, the on-line mall system immediately reacts to change dynamically the information presented to that consumer.

With data mining techniques the potential for on-line direct marketing and hence competitive advantage increases enormously. With behavioural and descriptive data about on-line consumers, it is possible to target: those who spend the most with offers of generous incentives; those who have not purchased for some weeks or months; or perhaps those who have switched to a rival company. The information provided by the database and data mining is related to customers, earning these two technologies an important place within relationship marketing ([Nor97])

4 Ethical Concerns

So far, we have presented some data mining fundamentals and demonstrated how these techniques can be applied in electronic commerce. We now raise a few ethical issues which we believe knowledge engineers, marketing experts as well as customers should be aware of.

Imagine you purchase a mobile phone at your local telecommunication shop. A few days later you find special offers in your private mail about mobile phone equipment. During the following week you receive phone calls from various GSM providers about special pricing schemes. An immediate complaint to your local shop manager about your privacy is the least you would consider as action. But, what about if the same happens over the Internet, the only difference that all contact is being sought by email. Are you still complaining? If so, to whom? Even more drastic, you only browse a shop and every single item you look at is traced. You come back to that mall and you are, through your IP and / or unique cookie, identified immediately. If this information is then used as input for data mining exercises and combined with log files and discovered knowledge from other electronic malls as well as external sociographic and psychographic data, your privacy is everything but respected. Thus, in addition to commonly discussed regulative topics about the Internet, the following issues have to be tackled to protect consumers:

- the types of log file information available in current and upcoming transfer protocol standards
- the exchange of log files among service providers
- the period of time of how long Internet server logs are stored
- the storage and maintenance of behavioural patterns
- the cross-fertilisation of discovered knowledge

5 Conclusions

The impact of applying artificial intelligence techniques to Internet server log files has been outlined. Further, it has been demonstrated how individual consumer profiles can be created and how, by generating dynamic HTML pages, marketing action can be taken instantaneously. Finally, a few ethical issues have been raised.

In information technology, as in most other sciences, there is always a trade-off between possible and ethically justified technologies. We believe that a compromise between the two paradigms - total regulation and full transparency - will be the most promising and satisfactory approach. And, remember, the only thing we really want is your total webisfaction. Well, as long as you pay for it.

References

- [Ana98] S.S. Anand, A.G. Büchner. Decision Support through Data Mining, Financial Times Pitman Publishing, 1998.
- [Mul97a] M.D. Mulvenna, A.G. Büchner. Data Mining and Electronic Commerce, in *Overcoming Barriers to Electronic Commerce (OBEC '97)*, Malaga, Spain, 1997
- [Mul97b] M.D. Mulvenna, A.G. Büchner, M.T. Norwood, C. Grant: The Soft Push: Mining Internet Data for Marketing Intelligence, in *Proc Working Conference Electronic Commerce in the Framework of Mediterranean Countries Development*, Metsovo, Greece, 1997.
- [Mul98] M.D. Mulvenna, M.T. Norwood, A.G. Büchner. Data-Driven Marketing, *International Journal of Electronic Markets*, 8(2), forthcoming, 1998.
- [Nor97] M. Norwood, M.D. Mulvenna, A.G. Büchner, C. Grant: Building Competitive Advantage on the Internet using Relationship Marketing, in *1st Berlin Internet Economics Workshop*, Berlin, 1997.