

A FRAMEWORK OF ANALYSIS: APPROACHES IN THE APPLICATIONS OF E-CRM

Dengfeng Hu^{1,2}, Yun Chen¹, Guozheng Zhang¹

1 Shanghai University of Finance & Economics, Shanghai, China, 2. Anhui University of Finance & Economics, Bengbu, China. Email: hdengfeng0212@sohu.com;

Abstract: Data mining techniques provide a popular and powerful tool set to generate various data-driven classification systems. Asim M Ansari and Carl F. Mela(2003) studied a statistical and optimization approach for customization of information on the Internet., and Balaji Padmanabhan (2004) explored the application of optimization and data mining. In this paper, based previous literatures we illustrate how optimization and data mining interact in e-CRM settings, and expound a framework of analysis in e-CRM.

Key words: approaches, data mining, optimization, e-CRM

1. INTRODUCTION

According to the professional and trade literature, more companies are using data mining as the foundation for strategies that help them outsmart competitors, identify new customers and lower cost [1]. In particular, data mining is widely used in marketing, risk management and fraud control [2].

A successful e-business strategy integrates both e-commerce and customer intelligence to provide a complete view "e-customer intelligence". A lucrative opportunity exists for vendor to provide an e-customer intelligence solution that can market, sell, and service customers across e-commerce and customer intelligence technologies such as OLAP, data mining, data warehousing, and campaign management. The Patricia Seybold

This project is supported by the Shanghai Shuguang Project of China under the grant No.05SG38.

Please use the following format when citing this chapter:

Hu, Dengfeng, Chen, Yun, Zhang, Guozheng, 2006, in International Federation for Information Processing (IFIP), Volume 207, Knowledge Enterprise: Intelligent Strategies In Product Design, Manufacturing, and Management, eds. K. Wang, Kovacs G., Wozny M., Fang M., (Boston: Springer), pp. 304-310.

Group has dubbed this opportunistic void between e-commerce and customer intelligence the “e-Gap”.

We systematically explore how optimization and data mining can help one another for certain customer relationship management (CRM) applications in e-commerce, termed analytical e-CRM.

The remainder of the paper is organization into four sections. The first section introduces analytical e-CRM in general. The second section discusses the optimization opportunities in e-CRM, customer interaction in e-CRM and the data mining literature. The third section illustrates possible approaches to solve traditional DM problems applications and examples of data mining in the literatures. Finally, the concluding section highlights the limitations of data mining and suggests other possible.

2. DATA MING AND OPTIMIZATION OPPORTUNITIES IN E-CRM

While the tremendous business value of customer-centric marketing and management strategies is intuitive, implementing CRM initiatives have data storage capabilities, data warehousing applications, and data mining techniques (Berry and Linoff, 1997), it is not just about computer software and hardware. Data mining uses sophisticated statistical processing or artificial intelligence algorithms to discover useful trends patterns from the extracted data. Data mining can yield important insights including predication models and association that can help companies understand their customers better. Swift [3] describes analytically e-CRM as a four-step iterative process consisting of (1) collecting and integrating online customer data, (2) analyzing this data, (3) building interactions with customers based on this analysis such that certain performance metrics such as LTV are optimized, and (4) measuring the effectiveness of these interactions in terms of these performance metrics.

2.1 Maximizing Lifetime Value

Customized products and communications attract customer attention and foster customer loyalty and lock-in. A typical performance metric used in many CRM applications is the LTV of a customer. One of the key questions in CRM is how to develop proactive customer interaction strategies that maximize LTV. Substantial work has been done on modeling LTV, mainly in the marketing literature [4]. The bulk of previous research in CLV has mainly focused on measuring the lifetime value customers or the value of entire customer base.

In contrast, in measuring the value of customers using individual customers of a firm, or limited customer acquisition in the form of switching from some known alternatives. The DM community studied the LTV problem in the presence of large volumes of customer data [5]. Rosset et al. (2002) compute LTV based on large volumes of customer data by focusing on using DM to estimate customer churn and future revenues. Because DM methods are used for computing LTV values, and optimization methods determine incentives resulting in highest LTV values, such applications provide examples of how DM methods can help solve optimization problems.

2.2 Customer Analysis in e-CRM

Customer analysis includes two main steps in the e-CRM context: (1) preprocessing data that tracks various online activities of the customers—this involves starting with individual user clicks on a site and constructing logical user “sessions” and summary variables; and (2) building customer profiles from this and other data to learn these patterns from data. DM is used Customer profiles are then built from these results. As argued by Zheng et al.[6], data preprocessing is a critical step of the knowledge discovery process in e-CRM, and the success of most DM methods to a large extent depends on this step.

Some popular session identification methods include session-level characterization [7] that aggregates user clicks into sessions, a fixed-length sliding window method that breaks a session into several sliding windows, and different types of clipping methods that break a session into windows of different sizes using various splitting methods.

2.3 Customer Interaction in e-CRM

Current DM approaches are based on mining Web log files for path traversal patterns and restructuring sites based on the analysis. In Perkwitz and Etzioni (1998), the problem is defined in terms of automatic construction of index pages based on log file data, with the goal of constructing index pages that provide users access to information that they are likely to view. The algorithm proposed has four steps [8]: (1) processing the log files into user visits, (2) computing co-occurrence frequencies between pages and creating a similarity matrix, (3) creating a graph from this matrix and finding cliques in this graph, and (4) creating index pages corresponding to each clique in the graph.

The website is structured as per the solution to the optimization problem. DM can, thus, identify additional constraints using patterns identified by this data. For example, DM can build that most users access the sports pages.

The optimization problem is then solved incorporating the additional constraints. This process can be iteratively done until some user-defined stopping criterion. However, DM will play a key role in the specification of the optimization space, and will also play a part in the specification of the constraints (Padmanabhan, 2004).

2.4 Personalization and Chance

Most of the approaches developed for solving this type of a recommendation problem use statistical and DM approaches, and usually try to determine “good” products to recommend to the customer. Various existing recommendation methods were classified by Balabanovic and Shoham (1997) into content-based, collaborative and hybrid approaches and have been reviewed in surveys, such as the collaborative filtering method — k ‘nearest neighbors’. Many proposals have been made for defining distances and weights to compute the weighted ratings. As argued in Adomavicius and Tuzhilin (2003), the recommendation problem can also be formulated as an optimization problem that selects the best items to users.

3. OPTIMIZATION—THE APPROACH TO SOLVE TRADITIONAL DM PROBLEMS

Optimization can contribute to DM in one of two ways: (1) optimization can be a component of a larger DM process, or (2) new DM techniques can be built using entirely optimization-based methods. Support Vector Machines (SVM) [9] fall into the second category, because selection of an optimal separating a surface constitutes the DM method. So, we consider the following DM problems: feature selection, active learning, DM model optimization, selection of the best DM model or pattern, classification using mathematical programming, rule and constraint discovery, clustering and visualization toolkit.

3.1 Feature Selection and Active Learning

When optimization is used to preprocess click-stream data, as described above, optimization is used before DM. In general, optimization can be useful during the preprocessing stage of the DM process. Good feature selection techniques can significantly improve DM algorithms in two ways. First, these techniques can contribute to dimensionality reduction and faster running times. Second, they can contribute to building more accurate models from data.

Mannino and Mookerjee (1999) address a related problem where they optimize the cost of information acquisition for expert systems. By selecting only the data with high utility to the model, active learning aims to minimize the data needed for model building. The different active learning methods fall into two categories: (1) heuristic based, and (2) optimization based. The data points chosen are those in which there is maximum disagreement among the models. Optimization approaches employ an objective function, and those data points that optimize this objective function are selected.

3.2 DM Model Optimization

Many DM methods can be directly formulated as optimization problems or can have an optimization component that is a part of the DM problem (Hand et al. 2001). As another example of the heuristic search for an optimal solution to a DM problem, consider the problem of selecting the best decision tree using the Minimum Description Length (MDL) principle.

3.3 Selection of the Best DM Model or Pattern

Fu et al. [6] describe how a genetic algorithm selects the best decision tree from a set of decision trees generated by tree induction methods. This selection process is performed in the post processing stage after the DM algorithm generates decision trees. Kennedy et al. (1997) present genetic algorithm encoding decision trees as chromosomes, using a variation of one of the standard tree traversal methods. Fu et al. (2003) build multiple decision trees from the data by subsampling from the data and building individual decision trees for each sample.

3.4 Classification, Rule and Constraint Discovery

Using Integer Programming. Logical Analysis of Data (LAD) is method for classification proposed in the optimization literature (Crama et al. 1988). It builds a classifier for a binary target variable based on learning a logical expression that can distinguish between positive and negative examples in a data set. In this sense, the problem is similar to the concept learning studied in machine learning (Mitchell 1997). The solutions to the problem are developed by formulating and solving optimization. Early work in this area was targeted at binary domains. Recent work has extended this target to include numeric predictor attributes by using a process known as “binarization”. In DM, a challenge for optimization researchers building DM algorithms is to explicitly address this issue and to study how optimization over training sets produces optimal models for unseen data.

The opportunity for optimization researchers is to develop new optimization-based search methods that can focus on learning rules faster than traditional DM search methods. Related work in machine learning presents an initial approach to exploring this opportunity.

There are opportunities for DM to help solve problems that are addressed traditionally using an optimization framework. For example, an opportunity for using DM with optimization is to use DM to learn patterns in data that can, in turn, be used as constraints in a traditional optimization framework. This can ensure that the solutions derived from the optimization approach are consistent with the dominant patterns that emerge from the data and, hence, the solution quality can be improved. Further, using better constraints can help solve the optimization problem faster by appropriately restricting the search space. Consider the content management problem of placing advertisements on Web pages, such as done by Double Click. This can be formulated as an optimization problem pointed out by Geoffrion and Krishnan (2001).

3.5 Clustering and Visualization Toolkit

When data are limited to two or three dimensions, the most powerful tool for judging cluster quality is usually the human eye. Clusion, our CLUster visualization toolkit, allows us to convert high-dimensional data into a perceptually more suitable format, and employ the human vision system to explore the relationships in the data, guide the clustering process, and verify the quality of the results.

The visualization toolkit Clusion allows even non-specialists to get an intuitive visual impression of the grouping nature of objects that may be originally defined in it can be extended to other domains, as illustrated by our results on document clustering and grouping web-logs. Alexander Strehl(2003) spent two years in studying the Dell customer data, and pointed out that Clusion effective for getting clusters balanced the number of customers or net dollar (\$) amount, and even more so for conveying the results to marketing management.

4. CONCLUSIONS

The paper has studied the approaches in e-CRM, simultaneity, addressing e-CRM problems by utilizing the strengths of both DM and optimization can, therefore, be a useful area for future research. A better understanding of the problems in e-CRM and of the possible ways in which DM and optimization

can interact, can help exploit the opportunities for synergistically using both approaches in the solution of e-CRM problems.

Specifically, we describe an optimization algorithm based on the assignment algorithm to optimize the design and content of e-CRM. We believe that such a general approach (combining choice models with optimization models) has utility beyond customization and can be used in the design of tailored services, customer designed catalogs, and bundling of goods

Optimization can help DM in e-CRM applications by developing more systematic methods of preprocessing click-stream data, in a manner that optimizes the performance of a DM model subsequently built on the preprocessed data. As (Balaji Padmanabhan, 2004) pointed out there are opportunities for researchers to formulate and solve this optimization problem.

5. REFERENCE

1. Dver, A. (2003), *Customer relationship management: The good, the bad, the future*. Bus. Week (April 28).
2. Kuykendall, L. (1999), The data-mining toolbox, *Credit Card Management* 12(6): 30-40.
3. Swift, R. (2002), *Analytical CRM Powers Portable Relationships: Creating Success by Letting Customers Guide You*. DM Rev.
4. Berger C (1999), *Data Mining to Reduce Churn*, Target Marketing, 22(8).
5. Rosset, S., E. Neumann, U. Eick, N. Vatnik, Y. Idan (2002), Customer lifetime value modeling and its use for customer retention planning, *Proc-ACM Internat Conf. Knowledge Discovery Data Mining Edmonton*.
6. Zheng, Z., B.Padmanabhan, S.Kimbrough(2003), On Data Preprocessing Biases in Web Usage Mining, *INFORMS J. Computer*, 15(2) 148 - 170.
7. Srikant, R., Y. Yang (2001), Mining Web Logs to Improve Website Organization. *Proc. Tenth World Wide Web Conf., Hong Kong*.
8. Balaji Padmanabhan, Lexander Tuzhilin, on the Use of Optimization for Data Mining: Theoretical Interactions and e-CRM Opportunities Management Science. 2003 *INFORMS* Vol. 49, No. 10, October 2003.
9. Balabanovic M., Y. Shoham (1997), Content-based, Collaborative Recommendation. *Comm. ACM* 40(3) 66 - 72.