Profit Mining

Senqiang Zhou
Simon Fraser University, Canada

Ke Wang
Simon Fraser University, Canada

INTRODUCTION

A major obstacle in data mining applications is the gap between the statistic-based pattern extraction and the value-based decision-making. "Profit mining" aims to reduce this gap. In profit mining, given a set of past transactions and pre-determined target items, we like to build a model for recommending target items and promotion strategies to new customers, with the goal of maximizing profit. Though this problem is studied in the context of retailing environment, the concept and techniques are applicable to other applications under a general notion of "utility". In this short article, we review existing techniques and briefly describe the profit mining approach recently proposed by the authors. The reader is referred to (Wang, Zhou & Han, 2002) for the details.

BACKGROUND

It is a very complicated issue whether a customer buys a recommended item. Consideration includes items stocked, prices or promotions, competitors' offers, recommendation by friends or customers, psychological issues, conveniences, etc. For on-line retailing, it also depends on security consideration. It is unrealistic to model all such factors in a single system. In this article, we focus on one type of information available in most retailing applications, namely past transactions. The belief is that shopping behaviors in the past may shed some light on what customers like. We try to use patterns of such behaviors to recommend items and prices.

Consider an on-line store that is promoting a set of *target items*. At the cashier counter, the store likes to recommend one target and a promotion strategy (such as a price) to the customer based on *non-target items* purchased. The challenge is determining an item interesting to the customer at a price affordable to the customer and profitable to the store. We call this problem *profit mining* (Wang, Zhou & Han, 2002).

Most statistics-based rule mining, such as association rules (Agrawal, Imilienski & Swami, 1993; Agrawal & Srikant, 1994), considers a rule as "interesting" if it passes certain statistical tests such as support/confidence. To an enterprise, however, it remains unclear how such rules can be used to maximize a given business object. For example, knowing "Perfume→ Lipstick" and "Perfume→Diamond", a store manager still cannot tell which of Lipstick and Diamond, and what price should be recommended to a customer who buys Perfume. Simply recommending the most profitable item, say Diamond, or the most likely item, say Lipstick, does not maximize the profit because there is often an inverse correlation between the likelihood to buy and the dollar amount to spend. This inverse correlation reflects the general trend that the more dollar amount is involved, the more cautious the buyer is when making a purchase decision.

MAIN THRUST OF THE CHAPTER

Related Work

Profit maximization is different from the "hit" maximization as in classic classification because each hit may generate different profit. Several approaches existed to make classification *cost-sensitive*. (Domingos, 1999) proposed a general method that can serve as a wrapper to make a traditional classifies cost-sensitive. (Zadrozny & Elkan, 2001) extended the error metric by allowing the cost to be example dependent. (Margineantu & Dietterich, 2000) gave two bootstrap methods to estimate the average cost of a classifier. (Pednault, Abe & Zadrozny, 2002) introduced a method to make sequential cost-sensitive decisions, and the goal is to maximize the total benefit over a period of time. These approaches assume a given error metric for each type of misclassification, which is not available in profit mining.

Profit mining is related in motivation to *actionability* (or *utility*) of patterns: a pattern is interesting in the sense that the user can act upon it to her advantage (Silberschatz & Tuzhilin, 1996). (Kleinberg, Papadimitriou & Raghavan, 1998) gave a framework for evaluating data mining operations in terms of utility in decision-making. These works, however, did not propose concrete solutions to the actionability problem. Recently, there were several works applying association rules to address business related problems. (Brijs, Swinnen, Avanhoof & Wets, 1999; Wong, Fu & Wang, 2003; Wang & Su, 2002) studied the problem of selecting a given number of items for stocking. The goal is to maximize the profit generated by selected items or customers. These works present one important step beyond association rue mining, i.e., addressing the issue of converting a set of individual rules into a single actionable model for recommending actions in a given scenario.

There were several attempts to generalize association rules to capture more semantics, e.g., (Lin, Yao & Louie, 2002; Yao, Hamilton & Butz, 2004; Chan, Yang & Shen, 2003). Instead of a uniform weight associated with each occurrence of an item, these works associate a general weight with an item and mine all itemsets that pass some threshold on the aggregated weight of items in an itemset. Like association rule mining, these works did not address the issue of converting a set of rules or itemsets into a model for recommending actions.

Collaborative filtering (Resnick & Varian, 1997) makes recommendation by aggregating the "opinions" (such as rating about movies) of several "advisors" who share the taste with the customer. Built on this technology, many large commerce web sites help their customers to find products. For example, Amazon.com uses "Book Matcher" to recommend books to customers; Moviefinder.com recommends movies to customers using "We Predict" recommender system. For more examples, please refer to (Schafer, Konstan & Riedl, 1999). The goal is to maximize the hit rate of recommendation. For items of varied profit, maximizing profit is quite different from maximizing hit rate. Also, collaborative filtering relies on carefully selected "item endorsements" for similarity computation, and a good set of "advisors" to offer opinions. Such data are not easy to obtain. The ability of recommending prices, in addition to items, is another major difference between profit mining and other recommender systems.

Another application where data mining is heavily used for business targets is *direct marketing*. See (Ling & Li, 1998; Masand & Shapiro, 1996; Wang, Zhou, Yeung & Yang, 2002), for example. The problem is to identify buyers using data collected from previous campaigns, where the product to be promoted is usually fixed and the best guess is about who are likely to buy. The profit mining, on the other hand, is to guess the best item and price for a given customer. Interestingly, these two problems are closely related to each other. We can model the direct marketing problem as profit mining problem by including customer demographic data as part of her transactions and including a special target item NULL representing no recommendation. Now, each recommendation of a non-NULL item (and price) corresponds to identifying a buyer of the item. This modeling is more general than the traditional direct marketing in that it can identify buyers for more than one type of item and promotion strategies.

Profit Mining

We solve the profit mining by extracting patterns from a set of past transactions. A transaction consists of a collection of sales of the form (item, price). A simple price can be substituted by a "promotion strategy", such as "buy one get one free" or "X quantity for Y dollars", that provides sufficient information for derive the price. The transactions were collected over some period of times and there could be several prices even for the same item if sales occurred at different times. Given a collection of transactions, we find *recommendation rules* of the form $\{s_1, \ldots, s_k\} \rightarrow \langle I,P \rangle$, where I is a target item and P is a price of I, and each s_i is a pair of non-target item and price. An example is (*Perfume, price=\$20*)$\rightarrow$(*Lipstick, price=\$10*). This recommendation rule can be used to recommend Lipstick at the price of \$10 to a customer who bought Perfume at the price of \$20. If the recommendation leads to a sale of Lipstick of quantity Q, it generates $(10\text{-}C)*Q$ profit, where C is the cost of Lipstick.

Several practical considerations would make recommendation rules more useful. First, items on the left-hand side in s_i can be item categories instead to capture category-related patterns. Second, a customer may have paid a higher price if a lower price was not available at the shopping time. We can incorporate the

Related Content