Title: Profit Mining

Senqiang Zhou

School of Computing Science, Simon Fraser University,

8888 University Drive, BC, Canada, V5A 1S6

Tel: 1-604-291-5371 Fax: 291-3045

E-mail: szhoua@cs.sfu.ca

Ke Wang

School of Computing Science, Simon Fraser University

8888 University Drive, BC, Canada, V5A 1S6

Tel: 1-604-291-4667 Fax: 291-3045

E-mail: wangk@cs.sfu.ca

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 and 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, competitor's 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 and Han, 2002).

Most statistics-based rule mining, such as association rules (Agrawal, Imilienski and Swami, 1993; Agrawal and 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$\rightarrow$Lipstick" and "Perfume$\rightarrow$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 and Elkan, 2001) extended the error metric by allowing the cost to be example dependent. (Margineantu and Dietterich, 2000) gave two bootstrap methods to estimate the average cost of a classifier. (Pednault, Abe and 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 and Tuzhilin, 1996). (Kleinberg, Papadimitriou and 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 and Wets, 1999; Wong, Fu and Wang, 2003; Wang and 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 and Louie, 2002; Yao, Hamilton and Butz, 2004; Chan, Yang and 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 and 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 and 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 and Li, 1998; Masand and Shapiro, 1996; Wang, Zhou, Yeung and 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*)→(*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 $(10-C)*Q$ 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 domain knowledge that paying a higher price implies the willingness of paying a lower price (for exactly the same item) to search for stronger rules at lower prices. This can be done through multi-level association mining (Srikant and Agrawal, 1995; Han and Fu, 1995), by modeling a lower price as a more general category than a higher price. For example, the sale {<*chicken, $3.8*>} in a transaction would match any of the following more general sales in a rule: <*chicken, $3.8*>, <*chicken, $3.5*>, <*chicken, $3.0*>, *chicken, meat, food.* Note that the last three sales are generalized by climbing up the category hierarchy and dropping the price.

A key issue is how to make a set of individual rules work as a single recommender. Our approach is ranking rules the *recommendation profit*. The recommendation profit of a rule *r* is defined as the average profit of the target item in r among all transactions that match *r*. Note that the rank by average profit implicitly takes into account of both confidence and profit because a high average profit implies that both confidence and profit are high. Given a new customer, we pick up the highest ranked matching rule to make recommendation.

Before making recommendation, however, "over-fitting" rules that work only for observed transactions, but not for new customers, should be pruned because our goal is to maximize profit on new customers. The idea is as follows. Instead of ranking rules by observed profit, we rank rules by *projected profit*, which is based on the *estimated error* of a rule adapted for pruning classifiers (Quinlan, 1993). Intuitively, the estimated error will increase for a rule that matches a small number of transactions. Therefore, over-fitting rules tend to have a larger estimated error, which translates into a lower projected profit, and a lower rank.

For a detailed exposure and experiments on real life and synthetic data sets, the reader is referred to (Wang, Zhou and Han, 2002).

FUTURE TRENDS

The profit mining proposed is only the first, but important, step in addressing the ultimate goal of data mining. To make profit mining more practical, several issues need further study. First, it is quite likely that the recommended item tends to be the item that the customer will buy independently of the recommendation. Obviously, such items need not be recommended, and recommendation should focus on those items that the customer may buy if informed, but may not otherwise. Recommending such items likely brings in *additional* profit. Second, the current model maximizes only the profit of “one-shot selling effect”, therefore, a sale in a large quantity is favored. In reality, a customer may regularly shop the same store over a period of time, in which case a sale in a large quantity will affect the shopping frequency of a customer, thus, profit. In this case, the goal is maximizing the profit for reoccurring customers over a period of time. Another interesting direction is to incorporate the feedback whether a certain recommendation is rejected and accepted to improve future recommendations.

This current work has focused on the information captured in past transactions. As pointed out in Introduction, other things such as competitor’s offers, recommendation by friends or customers, consumer fashion, psychological issues, conveniences, etc. can affect the customer’s decision. Addressing these issues requires additional knowledge, such as competitor’s offers, and computers may not be the most suitable tool. One solution could be suggesting several best recommendations to the domain expert, the store manager or sales person in this case, who makes the final recommendation to the customer after factoring the other considerations.

CONCLUSION

Profit mining is a promising data mining approach because it addresses the ultimate goal of data mining. In this article, we study profit mining in the context of retailing business, but the principles and techniques illustrated should be applicable to other applications. For example, “items” can be general actions and “prices” can be a notion of utility resulted from actions. In addition, “items” can be used to model customer demographic information such as Gender, in which case the price component is unused.

REFERENCES

Agrawal, R., Imilienski, T., & Swami, A. (May, 1993). Mining association rules between sets of items in large databases. *ACM Special Interest Group on Management of Data (SIGMOD),* Washington D.C., USA, 207-216.

Agrawal, R. & Srikant, R. (September, 1994). Fast algorithms for mining association rules. *International Conference on Very Large Data Bases* (*VLDB)*, Santiago de Chile, Chile, 487-499.

Brijs, T., Swinnen, G., Vanhoof, K., & Wets, G. (August, 1999). Using association rules for product assortment decisions: a case study. *International Conference on Knowledge Discovery and Data Mining* (*KDD),* San Diego, USA, 254-260.

Chan, R., Yang, Q., & Shen, Y. (November, 2003). Mining high utility itemsets. *IEEE International Conference on Data Mining* (*ICDM)*, Melbourne, USA, 19-26.

Domingos, P. (August, 1999). MetaCost: A General method for making classifiers cost-sensitive. *ACM SIG International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, San Diego, USA, 155-164.

Han, J., & Fu, Y. (September, 1995). Discovery of multiple-level association rules from large databases, *International Conference on Very Large Data Bases* (*VLDB),* Zurich, Switzerland, 420-431.

Kleinberg, J., Papadimitriou, C. & Raghavan, P. (December, 1998). A microeconomic view of data mining. *Data Mining and Knowledge Discovery Journal*, 2(4), 311-324.

Lin, T. Y., Yao, Y.Y., & Louie, E. (May, 2002). Value added association rules. *Advances in Knowledge Discovery and Data Mining, 6th Pacific-Asia Conference PAKDD*, Taipei, Taiwan, 328-333.

Ling, C., & Li, C. (August, 1998) Data mining for direct marketing: problems and solutions. *ACM SIG International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, New York, USA, 73–79.

Margineantu, D. D., & Dietterich, G. T. (June-July, 2000). Bootstrap methods for the cost-sensitive evaluation of classifiers. *International Conference on Machine Learning (ICML)*, San Francisco, USA, 583 – 590.

Masand, B., & Shapiro, G. P. (August, 1996) A comparison of approaches for maximizing business payoff of prediction models. *ACM SIG International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, Portland, USA, 195–201.

Pednault, E., Abe, N., & Zadrozny, B. (July, 2002). Sequential cost-sensitive decision making with reinforcement learning. *ACM SIG International Conference on Knowledge Discovery and Data Mining (SIGKDD)*, Edmonton, Canada, 259 – 268.

Quinlan, J.R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann.

Resnick, P. & Varian, H.R. (1997). CACM special issue on recommender systems. *Communications of the ACM*, 40(3), 56-58.

Schafer, J. B., Konstan, J. A., & Riedl, J. (November, 1999). Recommender systems in E-commerce. *ACM Conference on Electronic Commerce,* Denver, USA, 158-166.

Silberschatz, A. & Tuzhilin, A. (1996). What makes patterns interesting in knowledge discovery systems. *IEEE Transactions on Knowledge and Data Engineering,* 8(6), 970-974.

Srikant, R., & Agrawal, R. (September, 1995). Mining generalized association rules. *International Conference on Very Large Data Bases* (*VLDB),* Zurich, Switzerland, 407-419.

Wang, K., Zhou, S. & Han, J. (March, 2002). Profit mining: from patterns to actions. *International Conference on Extending Database Technology (EDBT),* Prague, Czech Republic, 70-87.

Wang, K., Zhou, S., Yeung, J. M. S., & Yang, Q. (March, 2003). Mining customer value: from association rules to direct marketing. *International Conference on Data Engineering* (*ICDE),* Bangalore, India, 738-740.

Wang, K. & Su, M. Y. (July, 2002). Item selection by hub-authority profit ranking. *ACM SIG International Conference on Knowledge Discovery and Data Mining (SIGKDD),* Edmonton, Canada, 652-657

Wong, R. C. W., Fu, A. W. C., & Wang, K. (November, 2003). MPIS: Maximal-profit item selection with cross-selling considerations. *IEEE International Conference on Data Mining (ICDM),* Melbourne, USA, 371-378.

Yao, H., Hamilton, H. J., & Butz, C. J. (April, 2004). A foundational approach for mining itemset utilities from databases. *SIAM International Conference on Data Mining (SIAMDM),* Florida, USA, 482-486.

Zadrozny, B., & Elkan, C. (August, 2001). Learning and making decisions when costs and probabilities are both unknown. *ACM SIG International Conference on Knowledge Discovery and Data Mining (SIGKDD),* San Francisco, USA, 204 – 213.

TERMS AND THEIR DEFINITION

Transaction: A transaction is some set of items chosen from a fixed alphabet.

Frequent Itemset: The support of an itemset refers to as the percentage of transactions that contain all the items in the itemset. A frequent itemset is an itemset with support above a pre-specified threshold.

Association Rule: An association has the form $I_1 \rightarrow I_2$, where I_1 and I_2 are two itemsets. The support of an association rule is the support of the itemset $I_1 \cup I_2$, and the *confidence* of a rule is the ratio of support of $I_1 \cup I_2$ and the support of I_1.

Over-fitting Rule: A rule has high performance (e.g. high classification accuracy) on observed transaction(s) but performs poorly on future transaction(s). Hence, such rules should be excluded from the decision-making systems (e.g. recommender). In many cases over-fitting rules are generated due to the noise in data set.

Classification: Given a set of training examples in which each example is labeled by a class, build a model, called a classifier, to predict the class label of new examples that follow the same class distribution as training examples. A classifier is accurate if the predicted class label is the same as the actual class label.

Cost Sensitive Classification: The error of a misclassification depends on the type of the misclassification. For example, the error of misclassifying Class 1 as Class 2 may not be the same as the error of misclassifying Class 1 as Class 3.

Profit Mining: In a general sense, profit mining refers to data mining aimed at maximizing a given objective function over decision making for a targeted population (Wang, Zhou and Han, 2002). Finding a set of rules that pass a given threshold on some interestingness measure (such as association rule mining or its variation) is not profit mining because of the lack of a specific objective function to be maximized. Classification is a special case of profit mining where the objective function is the accuracy and the targeted population consists of future cases. This paper examines a specific problem of profit mining, i.e., building a model for recommending target products and prices with the objective of maximizing net profit.