RCSSディスカッションペーパーシリーズ
第21号　2004年9月

ISSN 1347-636X

Discussion Paper Series
No.21　September, 2004

Knowledge Discovery from Consumer Behavior in an Alcohol Market by Using Graph Mining Technique

— An Example of Using an Active Mining Process for a Typical Business Application —

Mami KURODA, Katsutoshi YADA, Hiroshi MOTODA, and Takashi WASHIO

RCSS

文部科学省私立大学学術フロンティア推進拠点
関西大学ソシオネットワーク戦略研究センター

Research Center of Socionetwork Strategies,
The Institute of Economic and Political Studies,
Kansai University
Suita, Osaka 564-8680 Japan
URL : http://www.rcss.kansai-u.ac.jp/
http://www.socionetwork.jp/
http://www.policygrid.jp/
e-mail : keiseiken@jm.kansai-u.ac.jp
tel. 06-6368-1177
fax. 06-6330-3304

Knowledge Discovery from Consumer Behavior in an Alcohol Market by Using Graph Mining Technique

— An Example of Using an Active Mining Process for a Typical Business Application —

Mami KURODA, Katsutoshi YADA, Hiroshi MOTODA, and Takashi WASHIO

RCSS

文部科学省私立大学学術フロンティア推進拠点
関西大学ソシオネットワーク戦略研究センター

Research Center of Socionetwork Strategies,
The Institute of Economic and Political Studies,
Kansai University
Suita, Osaka 564-8680 Japan
URL : http://www.rcss.kansai-u.ac.jp/
http://www.socionetwork.jp/
http://www.policygrid.jp/
e-mail : keiseiken@jm.kansai-u.ac.jp
tel. 06-6368-1177
fax. 06-6330-3304

RCSS Discussion Paper Series

No.21 September, 2004

Knowledge Discovery from Consumer Behavior in an Alcohol Maket by Using Graph Mining Technique

－An Example of Using an Active Mining Process for a Typical Business Application

Mami KURODA*

Katsutoshi YADA**

Hiroshi MOTODA***

Takashi WASHIO****

Abstract

This paper deals with an actual case of the business application of graph mining techniques to purchase history data that employs an active mining process and the verification of the practical utility of the final results. While the current data mining process emphasizes iterative analysis of a given data, active mining is based on user interest levels and both the data and the process evolve dynamically in the form of a spiral configuration. In this paper, we describe the process of how new knowledge concerning consumer behavior in the alcoholic beverage market was discovered and how, based on this, new sales promotion planning was carried out and the effects were verified.

Keywords graph mining, data mining, consumer behavior, alcohol market

* Student, Graduate School of Commerce, Kansai University, mail to da10247@ipcku.kansai-u.ac.jp.

** Associate Professor, Faculty of Commerce, Kansai University, mail to yada@ipcku.kansai-u.ac.jp.

*** Professor, The Institute of Scientific and Industrial Research, Osaka University, mail to motoda@ar.sanken.osaka-u.ac.jp.

**** Associate Professor, The Institute of Scientific and Industrial Research, Osaka University, mail to washio@ar.sanken.osaka-u.ac.jp.

1. Introduction

In recent years, large databases are being accumulated to discover new knowledge and many researchers are interested in this process, called data mining. Likewise, these techniques have attracted the attention of many people in the business world. The key components of the data mining process consist of an understanding of the target data environment, data preprocessing, pattern recognition, use of the results and other important related factors. Up to the present time, there have been many studies of the process of pattern extraction. However, example cases are rare where the data mining processes have dynamically and organically been integrated and useful knowledge has been discovered.

In order to be able to utilize massive amounts of customer purchase data for the discovery of useful knowledge, we have carried out in-depth research on data mining technology and the development of infrastructure platforms to support related mining process [Hamuro 03]. In the course of these studies, we have become aware of many inherent limitations that have existed in most of the studies carried out in the past. For example, it has been unclear as to how to introduce the domain knowledge of experts into the data mining process [Yada 04a]. Another problem has been that although many iterations can be carried out based on specified conditions and given data sets, there has been little guarantee as to whether the experts will be able to attain the type of new knowledge that will satisfy his needs. The basic reason for this is that there has not been a well designed satisfactory, comprehensive and integrated environment to support the data mining process to function effectively in the context of the flexible needs of the analysts and to collect data that is aligned with the objectives and is of basic interest.

The term "active mining [Motoda 02]" refers to a process aiming at organic integration of the entire data mining process and in particular, this area of study is aimed at developing mining and related technologies that can effectively function to collect data from information sources that is

based on user interests. Such technology must be able to integrate existing specialized domain knowledge and from such integrated knowledge discover new knowledge that will be of interest to specialists and, through the integration of this knowledge, will lead to the generation of further new and useful knowledge. The objective of this paper is to clarify how useful knowledge can be discovered and uses an example of the application of the data mining process to an area of actual business. In this paper, by applying graph mining technology to retail purchase data, we arrive at new knowledge and insights. Based on this new knowledge, new sales promotion planning is carried out in the market. By using the new data generated by this process, further new knowledge concerning consumer behavior is generated.

2. Graph Mining and Consumer Behavior

2.1. Graph Mining

Graph mining is a technique used to extract characteristic patterns from a variety of graph structured data [Inokuchi 02]. The graph structure is a nice way of representing and explaining complex data forms and phenomena. In recent years, it has become possible to perform a complete search in extracting directed/undirected and colored/uncolored sub-graphs in a reasonable computation time. Study concerning graph mining is as yet in the development stage and there are many different methods and applications that have been proposed and are being tested to ascertain their practical usefulness.

In the area of medicine, graph mining has been used to analyze time-series hepatitis data, both diagnostic and treatment data, and express the data structure in the form of graph [Geamsakul 03]. Extracted patterns were then used to determine the best treatment. In this way, little by little, the examples of practical applications are beginning to be accumulated. However, there are very few practical examples so far that are related to the area of marketing which is the subject of this paper. The aim of this paper is to apply graph mining on accumulated POS data that is identified by con-

sumer ID to discover new knowledge. Then, using the rules obtained from that data, the aim is to use the data for sales promotion purposes and to carry out such a promotion and verify actual sales against benchmark sales to verify the utility of the method.

2. 2. Graph Structure and Consumer Behavior

Graph structure, with little loss of information, can be used to express the complicated phenomena connected with consumer purchasing behavior. Yada [Yada 04b] took up the subject of knowledge representation using graph data structures based on purchase history data and converted this data to such graph structure data. The data for the example described in this paper has been converted to the graph structures. Table 1 shows the purchase data for a given customer (ID: 1).

The customer (ID: 1) came to the store two times in February and bought various different products. Figure 1 shows the graph structure of the Customer ID: 1 data. Purchases related to a single purchasing occasion are represented as star graphs. The core node is the date and time of these purchases and the satellite nodes indicate the items purchased. Labels indicating price range and other factors are attached to the purchased items. Links are attached to the first and second purchase occasion and a label indicating the time lapse in days between the two purchase occasions is attached. In the same way links and labels are attached to the all purchase occasions.

Table 1: Example of POS Data tagged with Customer ID

ID	Date	Category	goods	Price zone
1	2004Feb02	Beer	Super D	High
1	2004Feb02	Egg	Organic	High
1	2004Feb02	Milk	Umei Milk	High
1	2004Feb02	Mayo	Mayo Q	Normal
1	2004Feb12	Egg	Organic	High
1	2004Feb12	Bread	Genmai bread	High
1	2004Feb12	Milk	Hone Milk	Low
:	:	:	:	:

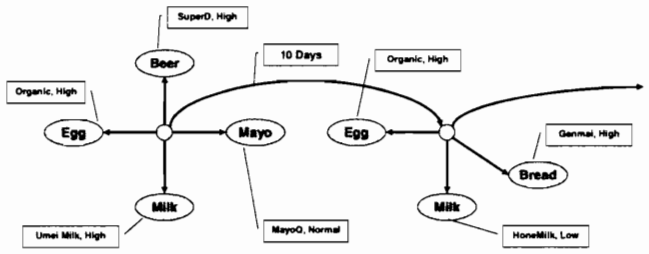

Figure 1: POS Data Converted to Graph Structured Data

The data contained in this graph includes not only the information about simultaneous purchases but also time sequence information about multiple purchases and information on groups of purchased products, making it possible to extract characteristic patterns from this information which can lead to the discovery of new knowledge which were unobtainable by conventional knowledge representations. From these graph structured data that are rich with information that indicates consumer purchase behavior, useful knowledge can be obtained for creating new sales promotion plans if characteristic patterns can be extracted. In this paper, graph mining was used to discover new

knowledge concerning consumer behavior related to the alcoholic beverage market. This new information was then used as the basis for creating new sale promotion plans and these plans were executed in Japanese super markets in order to verify the effectiveness of this data mining process.

3. Analysis of Consumer Behavior Related to the Alcohol Market

3.1. Alcohol Market in Japan

In recent years, the Japanese alcoholic beverage market has gone through drastic changes. Many new products have been launched on the beer market. In addition, low alcohol content beer (subject to a lower tax rate) has appeared on the market. Because of the differences in tax applied to different types of alcoholic beverages, manufacturers have met up with many different opportunities and many different threats. In this type of highly competitive market, it is no longer possible for enterprises to remain in a reactive mode, but rather, it has become necessary to be proactive and to create the changes that lead to new demand. In recent years, Manufacturer "A," that launched the top beer product into the Japanese beer market, does not make only regular beer and low alcohol beer products, but also, in an effort to differentiate itself from the other makers, wanted to provide retailers with new sales plans. In the spring of 2004, the joint research project with marketing staff of Manufacturer "A" and us has started to discover new and useful knowledge from customer purchase history data for new sales promotion to support retailers.

In the analysis of consumer data that is discussed in this paper, one full year's worth of purchase data (08/03 through 07/04) from three stores of supermarket chain in Japan was used. The total data included data for about 50,000 persons/month that visited one of the stores. The purchase records averaged 2 million records per month and came to 400MB in size.

In the case of alcoholic beverage market, sub-categories include beer, low alcohol beer and "chu hai" (a type of canned white liquor high ball). Company "A" has products in all of these sub-categories. We used 15 rough product categories of items such as meat, raw-fish and

fruit-vegetable to understand association between alcohol beverage and other food categories.

3. 2. Characteristics of the Consumer Purchase Behavior extracted from the Data

When graph mining technique were applied to the data described above, it was possible to discover several types of specific consumer behavior concerning the beer and chu-hai product categories. Some representative examples of these findings can be listed as follow:

1) Derived behavior rules related to beer purchasers

The first major feature of consumer behavior was that on the occasions when these customers bought Company "A" products 2 times every 10 days or more frequently, 29% of them also tended to buy "fruit & vegetable" together, and either meat or raw-fish on each occasion. Customers that did not buy beer, but bought meat or raw-fish on each occasion were only 9%. In addition, even among beer-purchasing customers, on occasions when they did not buy beer, meat or fish purchases were 10%. According to marketing experts of Company "A", their beer was a product for which the company emphasized the "high degree of freshness" and thus, they theorized, when customers purchase fruit & vegetable, and meat or fish, for which "freshness" was a primary consideration, they were subconsciously motivated to buy beer that fit this "freshness" mindset.

2) Behavior rules concerning buyers of chu-hai

The second major rule that emerged from the analysis was that in the case of customers that purchased Company "A" chu-hai products every ten days or more frequently, 19% also purchased confectionery products, and fruit & vegetable or meat on each occasion. Figure 2 shows the typical sub-graph extracted from chu-hai purchaser data. It could also be seen that chu-hai purchasers had a continuing tendency to purchase confectionery products on the same occasion that they purchased chu-hai products, a behavior pattern that set them apart from other consumers. We guess that these consumers were consuming chu-hai on occasions other than meals and were buying the confectionary products to serve as snacks to go with chu-hai products on these occasions.

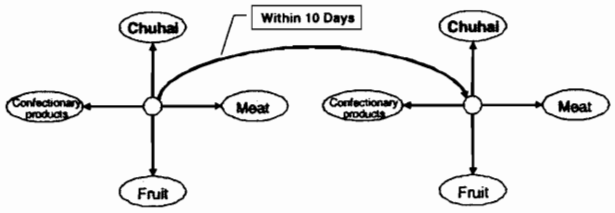

Figure 2: Sub-graph Extracted from Chu-hai Purchaser Data

4. Verification of the effectiveness of the New Sales Promotion

4. 1. The New Sales Promotion Planning

The Company "A" marketing staff studied the finding above and created the new sales promotion plan that is described below. In the past, it has been relatively common that beer and chu-hai products will be put into cardboard packs/cases at room temperature and that these cases will be stacked up at the point of sale. The new sales promotion plan called for chilling the products in cold stockers and placing them as close as possible to the fruit & vegetable and raw-fish areas in the store, thereby appealing to the consumer's tendency to emphasize the factor of freshness when making purchases of such products.

In this way, the "Company A" "freshness" product concept would be emphasized and this strategy would fit in the overall product concept. This strategy of the new sales promotion plan was an attempt to influence purchase behavior and this was an opportunity to see to what extent and in what way consumer purchase behavior would be affected.

The new sales promotion plan was executed at about 10 supermarket stores in the Kansai (Osaka) area in Japan during June and July 2004. Figure 3 shows the cool displays (stockers) that were used to display the products in the plan.

Figure 3: Cool Displays Used as Part of the New Sales Promotion Plan

4. 2. Results of the Experiment

The results of this experiment was that during the period of the special promotion, the sales of Company "A" beer products increased more, compared to the products of the other companies. The increase for chu-hai products also rose slightly. However, a matter that was surprising to us was that the store staff and the Company "A" marketing staff members did not seem to pay very much attention to this increase in sales of beer and chu-hai products during the period that the new promotion was going on. Instead, they showed much higher interest in the sales of the other products that the purchasers of beer and chu-hai were buying on the same occasions.

Figure 4 shows results for the customers that purchased Company "A" beer both before the promotion and during the promotion and indicates the average amount spent on items other than beer. It can be seen here that, during the period of the special promotion, the purchase of items other than beer by Company "A" beer purchasers increased. This was a purchasing tendency that was not present in the case of other stores and other company products in which special promotion was implemented. Thus, it is possible to surmise that the special sales promotion had some type of effect on

consumer purchasing behavior.

In addition, concerning chu-hai product related purchases of confectionery items on the occasion of chu-hai purchases, average purchases during the promotion went up to a large 40%. However, there were no noticeable changes in terms of the same occasion purchases of such items as meat, vegetables, fruit, etc.

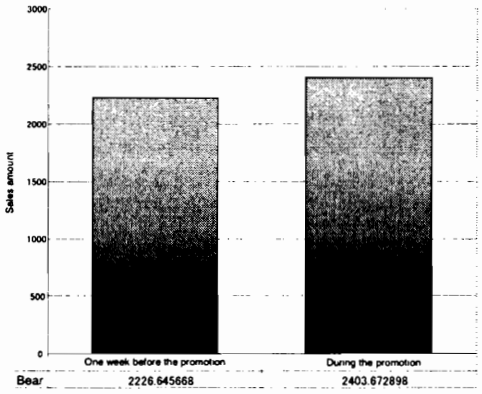

Figure 4: Pre-promotion and Promotion Period Average Purchases of Company "A" Beer and Other Items

Next, we look at the other item product categories during the promotion period for respondents that bought Company "A" beer products. During the promotion period, purchase of other items related to beer by these respondents increased purchase of raw-fish, fruit and vegetables compared with the case of purchase of rival beers. For example, Figure 5 show, for the periods just before and during the promotion period, the purchases of raw-fish by these respondents that had purchased Company "A" beer products and same occasion purchases of raw-fish by purchasers of other company beer for the same periods. The latter other beer purchaser same occasion average purchases of raw-fish were down by 14%. Compared with these results, for the respondents that bought both Company "A" beer products, same occasion average purchases of raw-fish went up by 15%. Same occasion purchases of fruit and vegetables also increased during the promotion more than that of

other company beer and chu-hai purchasers. However, same occasions meat sales increases were not observed to the extent of the Company "A" product purchasers.

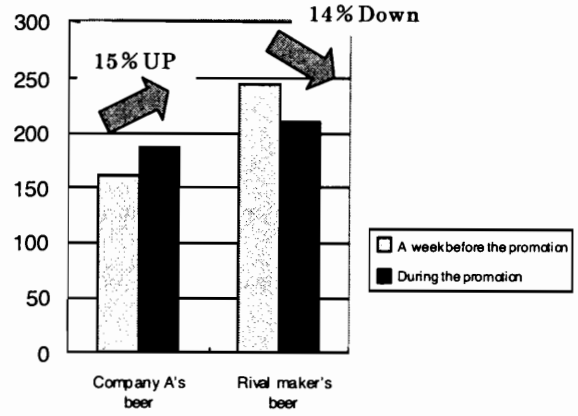

Figure 5: Rival Alcohol Product Purchaser Same Occasion Purchase of Raw-fish during the Promotion Compared to Respondents.

As a hypothesis based on the results of this experiment, it appears that the sales promotion strategy of using cool cases for displaying Company "A" beer and chu-hai products heightened the effects of the "freshness" product concept and this also resulted in increased same occasion purchases of "freshness"-related category products such as raw-fish and fresh fruit and vegetables. However, in the future, there is a need to learn more about the effects of such factors as different groups of customers, differences in reactions by region, differences in effects related to the in-store location of cool case placement and their respective effects on sales so that even more effective promotions can be created.

5. The Active Mining Process in Business World

As was discussed above, the active data mining process can be used to discover new knowledge from various data while integrating domain knowledge of experts. Based on the interest of experts that

this process can stimulate, further new knowledge can be created and new data is collected to integrate currently existing data and, as a result, they create a spiral process that leads to yet further useful knowledge. Compared to the currently existing data mining process that performs analysis iterations on data within pre-defined data, in the case of active mining, using user's interest as a basic in-put, new data can be collected and this changes the process into a dynamic process.

In the framework of business applications of data mining, the active mining approach is especially important and, as was demonstrated by the actual application described in this paper, the process was successful in discovering new and useful knowledge and the utility of this approach was verified. We believe that by carrying out actual applications of active mining to business problems, that in the future, it will be possible to define the remaining necessary technology and as yet unsolved questions.

6. Conclusion

In this paper, we not only analyzed given data sets, but also, in the next steps of the process, discovered new knowledge that was interpreted to be so by experts. Then, additional data based on these new points of interest was collected and integrated with the existing data and this led to the discovery of yet more new knowledge. This spiral process of active mining was conducted for an actual business application. It was demonstrated that this active mining was successful in discovering new and useful knowledge in real business world. In the future, we will continue to develop the technology that supports active mining and will apply it to various fields with the goal of verifying its broad applicability to real world problems.

Acknowledgement

Research of this paper is partly supported by the Grant-in-Aid for Scientific Research on Priority Areas "Active Mining" and RCSS fund by the Ministry of Education, Science, Sports and Culture of

Japan.

References

[Hamuro 03] Y. Hamuro, N. Katoh, E. H. Ip, S. L. Cheung and K. Yada, Combining Information Fusion with String Pattern Analysis: A New Method for Predicting Future Purchase Behavior, V. Torra(ed.), *Information Fusion in Data Mining, Studies in Fuzziness and Soft Computing*, Vol.123, Springer, pp.161-187 (2003).

[Inokuchi 02] A. Inokuchi, T. Washio, Y. Nishimura and H. Motoda, General Framework for Mining Frequent Structures in Graphs, *Proc. of the International Workshop on Active Mining*, pp.23-30 (2002).

[Geamsakul 03] W. Geamsakul, T. Yoshida, K. Ohara, H. Motoda and T. Washio, Extracting Diagnostic Knowledge from Hepatitis Dataset by Decision Tree Graph-based Induction, *Proc. of the International Workshop on Active Mining 2003*, pp.106-117 (2003).

[Mannila 96] H. Mannila, Data Mining: Machine learning, statistics, and databases, *Proc. of the 8th International Conference on Scientific and Statistical Database Management*, pp.1-34 (1996).

[Motoda 02] H. Motoda, M. Numao, T. Yamaguchi and S. Tsumoto, The Active Mining Project, *Journal of the Japanese Society for Artificial Intelligence*, pp.615-621 (2002).

[Yada 04a] K. Yada, Knowledge Discovery Process and Introduction of Domain Knowledge, B. Montano(ed.), *Innovations of Knowledge Management*, pp.86-98, IRM Press (2004).

[Yada 04b] K. Yada, H. Motoda, T. Washio and A. Miyawaki, Consumer Behavior Analysis by Graph Mining Technique, *Proc. of KES 2004*, LNAI3214, Springer, pp.800-806 (2004).