The relationship of food items using association rules

Sitthisak Srisawat and Jaruwan Nonkham
Faculty of Technology and Industry Management
King Mongkut's University of Technology North Bangkok
Emails: 5606021610062@fitm.kmutnb.ac.th, 5606021630063@fitm.kmutnb.ac.th

ABSTRACT

This research aims at data mining process and association rule to find a relationship in restaurant's sales data that store in restaurant's database system. Sales data are analyzed by using data mining technique with apriori algorithm. In this process, we also use machine learning software name weka 3.6.9 to extract the rules with confident value above 80%. Generated rules are beneficial for support restaurant's owner decided which food items are interested to make a promotion.

Index Term—Data Mining; Association Rule; Apriori algorithm

1. INTRODUCTION

With a highly saturated and competitive market in Restaurant industry, restaurant operators ought to be more selective in running their business strategies. Some old Restaurant still have a problem. For example, They don't have a promotion to attract a customer. They don't have interested menu presentation. Therefore, they must have a restaurant managing system to improved their sales and compete other restaurant. So we are collected sales order from KB Coffee & Steak restaurant for made a promotion for a customer easy to decided.

To make a promotion, we need to know relationship of food items in sales data. We decide to use data mining technique to find out which food items have relationship with other. For example, if customers ordered noodle, they also ordered iced tea at the same time. Therefore, we choose apriori algorithm to extract the association rules.

In a row with above studies, we have an idea to develop information system using data mining technique with Apriori algorithm. We choose KB Coffee & Steak restaurant as case study because they have a sales data that ready to preprocess and transformation. In a data mining process to create a model, we use a machine-learning software named weka version 3.6.9 to extract the association rules and bring it to help restaurant's owner make a decision to make a promotion to attract a customer and make more profit.

2. SIMILAR RESEARCH

Bayu Adhi Tama[4] have developed "Data mining for prediction customer satisfaction in fast-food restaurant" to predict customer satisfaction in fast-food restaurant. Managing customer satisfaction has become a crucial issue in fast-food industry. This study aims at identifying determinant factor related to customer satisfaction in fast-food restaurant. Customer data are analyzed by using data mining method with two classification techniques such as decision tree and neural network. Classification models are developed using decision tree and neural network to determine
underlying attributes of customer satisfaction. Generated rules are beneficial for managerial and practical implementation in fast-food industry. Decision tree and neural network yield more than 80% of predictive accuracy.

Daniel Simon Sanz and Ankur Agrawa[5] have developed "Automated Menu Recommendation System Based on Past Preferences" to helps a person to choose what food they might want to order in a specific estaurant. The application learns user behavior with each order what they order in each kind of meal and what are the products that they select together. After gathering enough information, the application can suggest the user about the most selected dish in the recent past and since the application started to learn. Applications, such as these, can play a major role in helping make a decision based on past preferences, thereby reducing the
user involvement in decision making.

3. RESEARCH PROCESS

In this section, we depict the data mining process of this study as in Figure 1. Details of each steps is described below.

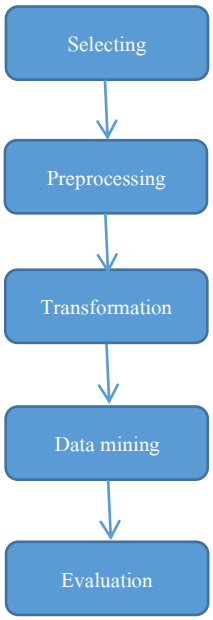

Figure 1. Data mining process diagram

Detail about data mining are shown below.

Data mining[1], the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

Most companies already collect and refine massive quantities of data. Data mining techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line. When implemented on high performance client/server or parallel processing computers, data mining tools can analyze massive databases to deliver answers to questions such as, "Which clients are most likely to respond to my next promotional mailing, and why?"

This white paper provides an introduction to the basic technologies of data mining. Examples of profitable applications illustrate its relevance to today's business environment as well as a basic description of how data warehouse architectures can evolve to deliver the value of data mining to end users.

3.1 Selecting

In this process, we export a 3417 sales data 110 foods and drinks list from KB Coffee &Steak's Database. Sales data that we export will keep between 1 January 2015 – 20 November 2015 as shown in table 1.

Table 1. Example of sales data

Date	No.	List	Quantity
1-JAN-2015	1	Chicken steak	1
		Nugget	1
		Water	2
	2	Pork Steak	1
		Nugget	1

3.2 Preprocess

Before extracting rules, we need to delete unnecessary sales data. Unnecessary sales data is shown below.

- Sales data that ordered only 1 item.
- Sales data that ordered more than 1 item but only ordered 1 food list.

Afterward delete unnecessary data, we need to arrange a sales data's format into appropriate format by export a data from database and convert into CSV format. For each sales data, we checked the food items. If customer ordered that food items, checked 1. If customer not ordered that food items, checked 0. The result of this process are shown at Table 2.

Table 2. Example of sales data in appropriate format.

Date	No.	Pork steak	Chicken steak	Nugget	Water
1-JAN-2015	1	0	1	1	1
	2	1	0	1	0

From table 2, we group a data by type of food that make easy to extract the association rules because if we have more food items, we need to put more data to extract the rules. We only have 3000 sales data from KB Coffee & Steak, it not enough to extract the rules.

Table 3. Example of sales data group by type.

Date	No.	Steak	Snack	Drinks
1-JAN-2015	1	1	1	1
	2	1	1	0

3.3 Transformation

In this process, we need to transform data into format that weka version 3.6.9 is compatible. We change a 1 to y and 0 to ?. The results are shown at table 4.

Table 4. Example of sales data in weka format

No.	Steak	Snack	Drinks
1	Y	Y	Y
2	Y	Y	?

Now, we have a sales data that ready to extract the rules in next process.

3.4 Data mining process

To extract the rules, we use machine learining software named weka version 3.6.9 for extract the association rule by using apriori algorithm. The Apriori algorithm [3] was proposed by Agrawal and Srikant in 1994. Apriori is designed to operate on databases containing transactions Other algorithms are designed for finding association rules in data having no transactions (Winepi and Minepi), or having no timestamps (DNA sequencing). Each transaction is seen as a set of items (an itemset). Given a threshold C, the Apriori algorithm identifies the item sets which are subsets of at least C transactions in the database.

Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time (a step known as candidate generation), and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found.

Apriori uses breadth-first search and a Hash tree structure to count candidate item sets efficiently. It generates candidate item sets of length k from item sets of length k-

1. Then it prunes the candidates which have an infrequent sub pattern. According to the downward closure lemma, the candidate set contains all frequent k-length item sets. After that, it scans the transaction database to determine frequent item sets among the candidates.
The pseudo code for the algorithm is given below for a transaction database T, and a support threshold of epsilon . Usual set theoretic notation is employed, though note that T is a multiset. C_{k} is the candidate set for level k. At each step, the algorithm is assumed to generate the candidate sets from the large item sets of the preceding level, heeding the downward closure lemma. count[c] accesses a field of the data structure that represents candidate c, which is initially assumed to be zero. Many details are omitted below, usually the most important part of the implementation is the data structure used for storing the candidate sets, and counting their frequencies.

Figure 2. Weka version 3.6.9

In weka version 3.6.9, we choose Apriori algorithm to extract the rule. We need association rule with confident value above 80%. Therefore, we define minimum support is 0.2 and minimum metric is 0.8 at apriori algorithm parameter setting that shown at figure 3.

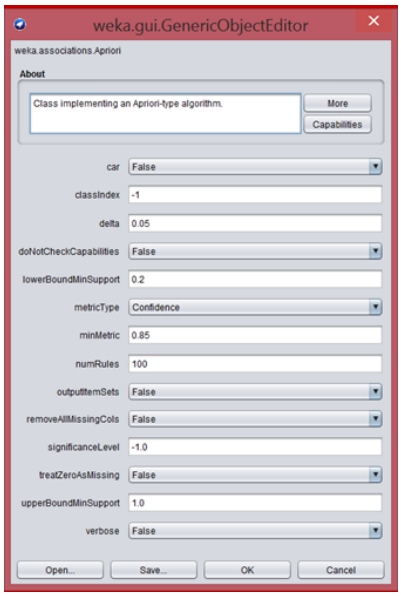

Figure 3. Apriori algorithm parameters in Weka version 3.6.9

4. RESULT AND DISCUSSION

4.1 Evaluation

At the result, we got 7 rules from apriori algorithm by using weka version 3.6.9.

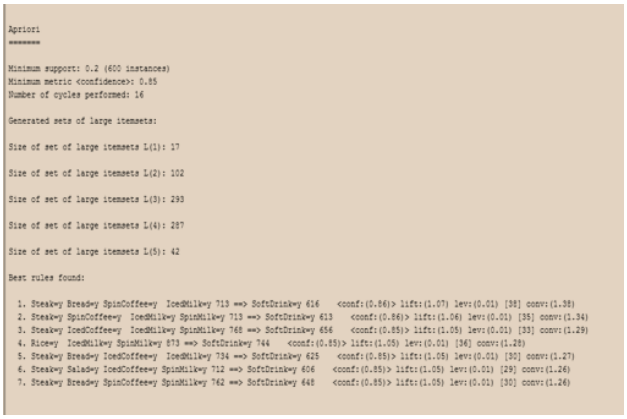

Figure 3. Result of extract rules from weka 3.6.9

Rule 1, if customer ordered bread, steak, iced milk and spin coffee. They also order a soft drink with confidence value = 86%
Rule 2, if customer ordered steak, spin coffee, iced milk and spin milk. They also order a soft drink with confidence value = 86%
Rule 3, if customer ordered steak, iced coffee, iced milk and spin milk. They also order a soft drink with confidence value = 85%
Rule 4, if customer ordered rice, iced milk and spin milk. They also order a soft drink with confidence value = 85%
Rule 5, if customer ordered steak, bread, iced coffee and iced milk. They also order a soft drink with confidence value = 85%
Rule 6, if customer ordered steak, salad, iced coffee and spin milk. They also order a soft drink with confidence value = 85%
Rule 7, if customer ordered steak, bread, spin coffee and spin milk. They also order a soft drink with confidence value = 85%

Otherwise, Weka's result is hard to understand for users. We'll developed a web application and transform this result into a graph and picture that shown at figure 5.

Figure 5. Web application shown a result from Weka 3.6.9

4. CONCLUSION

This research's objective is learning about data mining process and association rule for analyze sales data's associate from KB Steak & Coffee restaurant to help restaurant's owner to make a promotion. We use weka version 3.6.9 and apriori algorithm to extract the rules from sales data. At the result, we got 7 rules at minimum support is 0.2 and minimum metric is 0.8 by using weka version 3.6.9. This result, Restaurant's owner can bring it to make a promotion to attract customer and make more profit. Suggestion, in selection process in data mining ,if we have more sales data to extract the rules, the rules will more accurate. Minimum metric, if you define a minimum metric higher, the result is better. We recommend a minimum metric is 0.8 or above to get a best and excellent association rules.

5. ACKNOWLEDGEMENT

This research was supported by Faculty of technology and industry management, King Mongkut's university of technology North Bangkok.
We would also like to show our gratitude to the
Asst. Prof. Dr.Wanthanee Prachuabsupakij for assistance with Data mining technique.

REFERENCES

[1] Rakesh Agrawal and Ramakrishnan Srikant, Fast algorithms for mining association rules in large databases. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago, Chile, Septemb er 1994.

[2] BAYU ADHI TAMA, "DATA MINING FOR PREDICTING CUSTOMER SATISFACTION IN FAST-FOOD RESTAURANT", Department of Information Systems, Faculty of Computer Science, Sriwijaya University, Indonesia, May 2015.

[3] Daniel Simon Sanz and Ankur Agrawal, "Automated Menu Recommendation System Based on Past Preferences", Department of Computer Science, Manhattan College, Riverdale, NY 10471

[4] An Introduction to Data Mining[online]. Available: http://www.thearling.com/text/dmwhite/dmwhite.htm.

[5]Margaret Rouse. Data mining[online]. Available: http://searchsqlserver.techtarget.com/definition/data-mining.