An Association Rule Modeling For An Intelligent Product Recommendation System for Retail Business

Michael Watkins[1*], Kanokwan Watkins[2], Nutsana Na phayap[3], Chanwut Thongkamkaew[4]

[1] Officer, Health Department, Western Australia, Australia

[2,3] Lecturer, Digital Media Design, Didyasarin International College, Hatyai University, Thailand

[4] Lecturer, International Business Management, Didyasarin International College, Hatyai University, Thailand

***Corresponding author, E-mail**: michael.watkins@health.wa.gov.au

Abstract

In order to enhance the number of sales as well as to improve customers' satisfaction, many retail businesses invest the application systems and implement strategies to improve customers' satisfaction and management. Technological supports are a strategy that many businesses provide to service and assist customers, staff and managers. This study proposed an Intelligent Product Recommendation model for retail business in Thailand. The prediction model called Electronic product recommendation (e-Product recommendation). This model is used to predict customer requirement on buying products at the stores and able to assist the staff to offer the products for customers individually during their transaction at the point of sale (POS) or send product offer via email for membership. The e-Product recommendation model is one of the models in the intelligent product recommendation system. This system could assist the business to perceive more sales opportunities. In the experiment, the dataset of customer transaction was used. The results are interpreted as a new model of research conducted on Association Rules and clustering techniques. The experiment results found that the proposed model enhances the accuracy of Association Rules in comparison to the benchmark model. In the experiment, over 35,000 records of transactions were used. The results provide the support for staff to offer possible products for customers in the retail business to predict product for each customer individually.

Keywords: Association Rules, Intelligent System, Customer Satisfaction, Retail Business

Introduction

With the growing of online business and the high competition of the retail market, many retail businesses are working hard to improve the services and find strategies to increase sales. Most retail businesses are focusing on how to increase sales rates and improve services for customer satisfaction. In addition, a retail's performance is also increasingly being used to measure its ranking and reputation (Ou & Abratt, 2006). Human factors and marketing have a good relationship in term of profitability of retail (Resnick, Pompa, & Korn, 2004). It can be said that more retailers develop customer-centric and segment-based business. By this, business intelligence (BI) and customer relationship management (CRM) systems are more significant to evolve retail business in achieving the business goals (Phan & Vogel, 2010).

In addition, many businesses have invested application systems in retail business to assist with management. For example, using a system and methods to satisfy a consumer with a discount and a vender (United States Patent No. US8355948B2, 2011). Another example, Wong et. al. (2012) have combined the use of RFID Technology and Product Cross-selling System (IPCS) to perform cross and up selling in retail business. These systems could assist the business to perceived sales opportunities, achieve customer satisfaction.

One of the initiatives designed to help staff in offering the products for customers is the product recommendation. Such system could be used to provide ranked product for customers to achieve better sales in each transaction. With the high competition in retail business, Rego et. al. (2013) mentioned that the small retailers' issues may arise if the business does not improve their services, which in turn can improve market shares In the point of sale where a buyer come for the purpose of conducting a transaction, more sale opportunities can be applied. For example, the cashier offers a couple interesting products for customer's choices. Hence, it is desirable that some forms of intelligent recommendation tools could be developed to assist staff in offering likely required product in sale transaction. This form is the motivation of this research. With increasing number of online business and expanded number of choices for customers, retailers are dealing with high competitions. It becomes apparent that some forms of intelligent system will be useful in assisting the staff to increase sales and to satisfy customers in retail business.

Objectives

This study aims to:

1. find the ranked product recommendation based on past records from the transaction database from a retail store. This is intended to increase sales in the retail business.

2. investigate and develop the ranked product prediction model in the proposed intelligent recommendation system. ARs are employed to identify the relationship between the data

3. improve the performance of the recommendation model using clustering techniques.

4. propose the integrated techniques and improve the accuracy of the recommendation model in the proposed product intelligent recommendation system.

Research Background

According to literature, the problem of customer satisfaction and lack of technological support of retail business could be attributed to low number of customers and reduce of the revenue. On the other hand, it was found that the quality and convenience of support services are other factors that influence customers to change the retail stores. Consequently, the concept of CRM has been implemented in various retail businesses so as to assist the improvement of the quality of services and selling (Buttle & Maklan, 2019).

Various research studies have focused on CRM. The definitions of Customer Relationship Management (CRM) is the combination of practices strategies and technologies (Rouse, 2020), which focuses on customers and are aimed to establish effective competition and new strategies in order to improve the performance of a firm. Kuscar et. al (2018) added in the definitions that the relationships and interactions between business and customers are improved. In the system perspective, CRM can also enhance the business competitive advantage. In addition, Ranjan and Bhatnagar's research (2011) showed that Business Intelligence (BI), Knowledge Management (KM) and Analytical CRM (aCRM) could help the businesses in decision making, also could assist customers such as promotional schemes. In terms of Recommendation system to assist the retailers, Lee (2010) employed Collaborative Filtering, one of Data Mining techniques,

to approach the commodity recommendation system or retail business corresponding to customer preferences. This project provided a tool to aid the retail business in offering products for customers at the Point of Sales or online such as via email of members. The CRM strategy also provided the business with CRM practices, including the planned activities to be developed for the staff, as well other relevant participants.

Hence, it can be said that CRM can be utilized as an important means to support and enhance the customer satisfaction. Since understanding the needs of customers is essential for their satisfaction, it is necessary to prepare strategies in both using technological supports and services to support Customer Relationship Management. This paper therefore proposes an innovative information system to assist staff in servicing customers by offering interesting products in retail business in order to support the CRM concept.

Intelligent Technique Used

Data-mining techniques were used in various recommendation systems to determine the relationship between data records (Watkins, Napayap, & Fung, 2015). Classification is one important technique in data mining that can be used to classify data and discover knowledge from databases (Kaur, Singh, & Josan, 2015). In this study, to predict listing status in order to achieve better performance in terms of accuracy measures, the multi-classification model is an important tool in data mining that aimed to extract a model to find the relevant relationships between the attribute set and class labels (Zhou, Tam, & Fujita, 2016). There have been many research reports on the use of AR for classification purposes (Watkins, Napayap, & Fung, 2015). Example applications are product recommendations (Choi, Kang, & Jeon, 2006)], service recommendation system (Al-Hassan, Lu, & Lu, 2015) and business rating predictions (Tiroshi, et al., 2014).

A concept to construct a concise and accurate classifier using an AR was proposed, a novel classification algorithm classification based on Association Rules (MCAR) was presented. MCAR employed Association Rules technique in data mining to discovering frequent items to ensures detailed rules with high confidence. It was claimed that their proposed MCAR classification rule set achieved the high average accuracy with regards to error rate and efficiency (Thabtah, Cowling, & Peng, 2005).

Another study by Paireekreng et al. (Paireekreng, Wong, & Fung, 2011) proposed an integrated method by using classification and association rule techniques to extract knowledge from mobile content in a user profile. This proposed method simplified the

association from outcomes of the classification and clustering processes for the non-interactive recommendation system. Another study by Soliman and Adly (2012) also proposed an algorithm using an AR to find the best subset of rules for all possible ARs to build an efficient classifier. Therefore, many research reports have shown that ARs are an accomplished technique for the classification (Watkins, Napayap, & Fung, 2015). In this study, ARs based on GRI were used to extract the rules for the multiclass-classification problem. Many research reports have shown that the results of ARs based on GRI were of high quality (Al-Hassan, Lu, & Lu, 2015) (Watkins, Napayap, & Fung, 2015) (Kuscar, Draskovic, & Pavicic, 2018).

To improve the performance of ARs, K-means clustering was introduced by Tou and Genzalaz in 1974. Liu and He (2004) suggested that clustering can classify data and improve the accuracy of ARs. Plasse et al. (2007) found that the clustered data, which were extracted by ARs, gained more accuracy than normal data. Therefore, in this proposed GPA recommendation model, K-means clustering was used to enhance the performance of the model.

Concept theory/framework

To provide an overview and clear understanding of an intelligent product recommendation system. In this experiment, the sample data were chosen from a retail business database of 36,809 records. transaction records. After the data cleaning process, 35,500 transaction records were used in this study. The distribution of the products, with respect to types, is illustrated in Figure 1.

ใบเสร็จรับเงิน
ร้าน ABC
หมายเลขประจำตัวผู้เสียภาษี 2xxxxxxxxxx3
ใบกำกับภาษี
เลขที่: AOU11052015-0002 วันที่ 11/05/2558
โทร 02 3445 xxxx

เลขที่	รหัสสินค้า	รายการ	จำนวน	ราคา	ยอดรวม
1	112004	ยำยำ จัมโบ้ รสหมูสับ	2	6.00	12.00
2	710045	ปลากระป๋อง โรซ่า	1	17.00	17.00
3	120001	เป็บซี่	1	24.5	24.5
4	650012	ถั่วโก๋แก่ รสกระทิ	1	20.00	20.00
5	650019	ยาสีฟัน คอลเกต	1	15.00	15.00
6	670015	ลิบตัน ทีชู กระป๋อง 245***	1	13.00	13.00
				รวมเป็นเงิน	101.5
				ยอดก่อนคิด VAT	101.5
				ภาษี 7%	7.35
				ยอดสุทธิ	108.85
				รับเงิน	1000.00
				เงินทอน	891.15

Note: *** means the product was successful offered for customer by staff.

Figure 1 example of receipt used in the retail store (2015-2019)

In Figure 1, the product data were obtained from seven academic years of records (2014–2015). Transaction data included records from 2 January 2015 to 30 December 2019. The data comprised of 35,500 records. The transactions performed in this study are only the transaction with the product successfully offered for customers in the store. The data in this study did not indicate any personal information because of privacy issues, and no customer was identified in the research. The data was randomized and all private information was removed in this experiment.

As aforementioned, the process of choosing variables was based on results provided by the retail store. The variables used in these two modules are shown in Table 1.

Table 1 Variables used in the ranked product

No.	ranked programme recommendation	
	Variable name	Type
1	ProductOffer	Target
2	ProductName1	Input
3	ProductName2	Input
4	ProductName3	Input

In the framework, 'ProductOffer is a significant input to discover the product names that should be supported by extracting the successful cases from the transaction database. This ranked product recommendation module provides information on recommended product to the cashier and managers either at the Point of Sales(POS) or emails for membership. The managers are also able to use this ranked products for their promotions. Details of the methodology used in this experiment are described in the next section.

Experiment Methodology and Design

This section describes the methodology and the ranked program and activity recommendation model. Normalization of the data was first carried out as an essential step in pre-processing. To prepare the dataset for the GRI algorithm in the data analysis

process, quantitative data was required. For the training, validation and testing of the model, the dataset was randomized and divided into three sets: 60 per cent, 20 per cent and 20 per cent of data, respectively. The proposed model is illustrated in Figure 2.

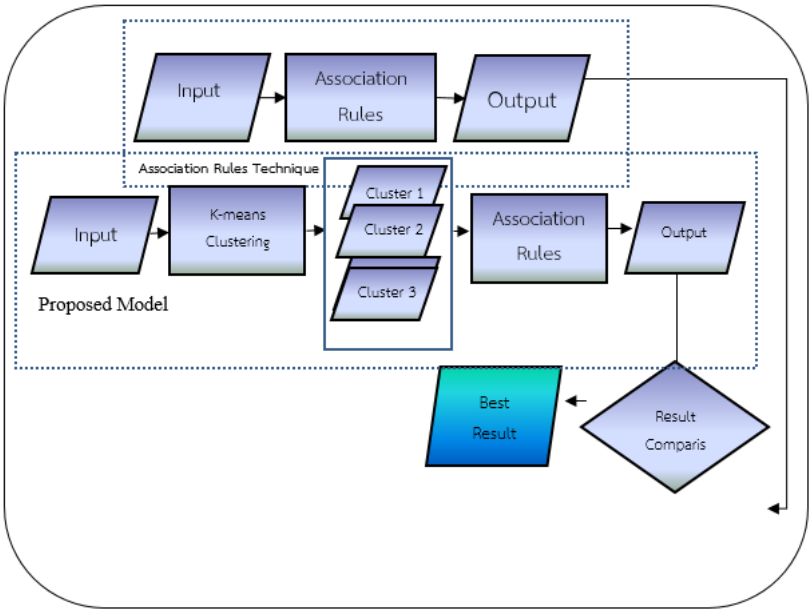

Figure 2 Process to compare performance of AR for ranked product recommendations

The GRI algorithm was used in the first stage. To improve the prediction accuracy, the K-means clustering technique was incorporated with the GRI algorithms, as shown in Figure 2. In this study, 35,500 random records with the aforementioned parameters for the ranked products were used based on the past records, the number of clusters used was three.

After determining the ARs using GRI algorithms to find the correlations between transaction records in the dataset, the confidence levels of the rules from the results in the first stage are sorted according to the ranked products. The extracted rules are filtered and categorized according to the confidence levels 85–100 per cent, 65–84 per cent and 45–64 per cent as the top, second and third ranked products, respectively.

Results

The results in ranked format are provided to the cashiers and the manager to assist them with their product recommendations for the customers.

As aforementioned, after the three rule levels have been set, the next step is matching the rules with product names. Examples of the displayed results are shown in Table 2.

Table 2 Example results of ranked product recommendation

Rank	Recommended product	Product Name
1	Product1 = 0.1	น้ำดื่มตราสิงห์
2	Product2 = 0.9	ปลากระป๋อง ตราสามแม่ครัว
3	Product3 =0.5	ลิบตัน พีช กระป๋อง 245

The results from GRI algorithms with the results from the combination of K-Means clustering and GRI, after the rule extraction process was executed, 207 rules were generated for the ranked product recommendation. The rules were then divided into three rankings according to the confidence levels. It was found that only 149 products out of 351 products were recommended in the rank.

To evaluate the results, 20 percent of the products was used to test the accuracy of the rules. The results from the test, in terms of ranked products are presented in Figure 3 and 4.

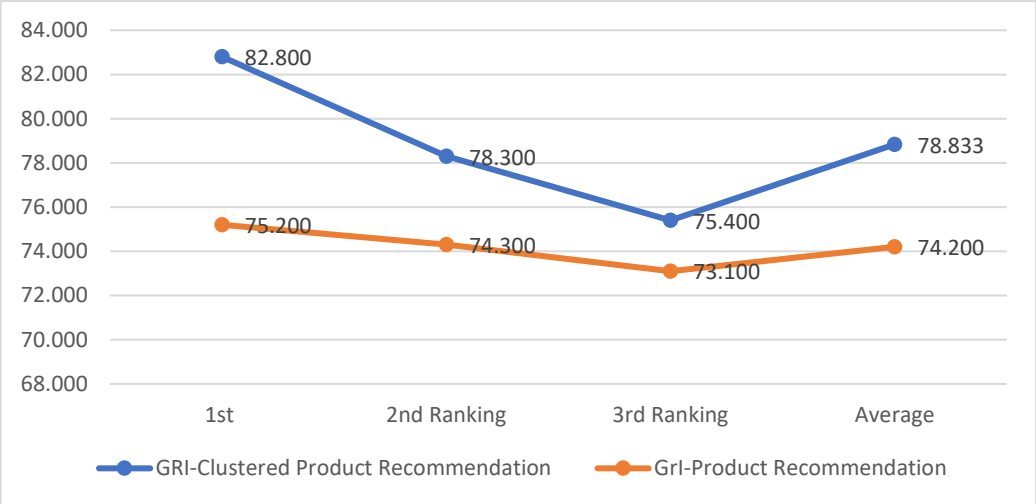

Figure 3 comparison of the accuracy between GRI Clustered and GRI

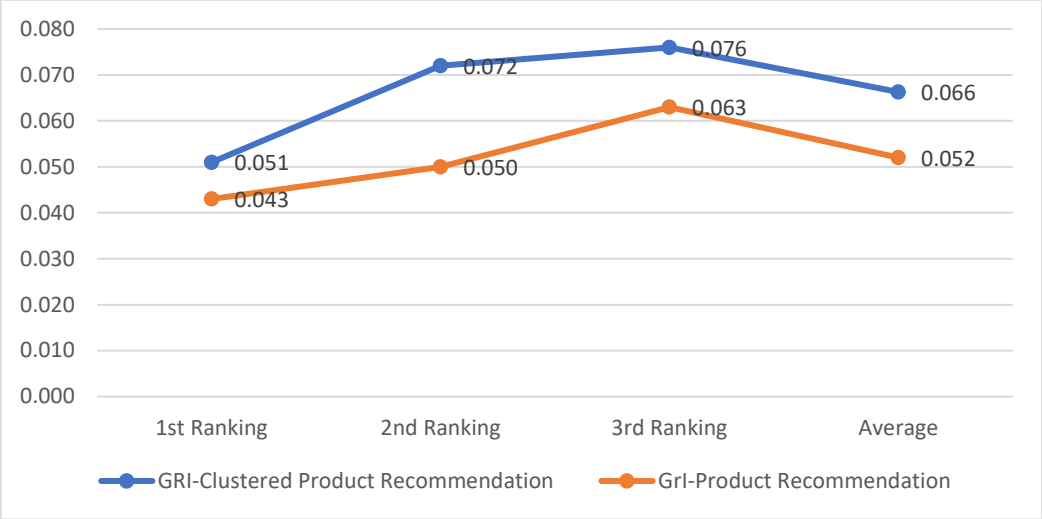

Figure 4 comparison of the means absolute error between GRI Clustered and GRI

The comparison in Figure 3 and 4 shows that the proposed ranked product recommendation using clustered GRI outperformed the GRI technique only. The accuracy

Conclusions and future work

In the retail business, the managers and cashiers could use the system to improve the sale opportunity and enhance the customer satisfaction. In this study, a model for the recommendation of ranked products is proposed to provide three ranked products to the cashiers to offer each customer individually at the point of sales. These offered products have been analyzed from the past 35,500 records which were successful offered during the transaction based on the products that customers were buying at the counter. The use of clustered data could assist to improve the accuracy of the results. In the module (ranked product recommendation), it was found that ARs based on GRI with the incorporation of two sets of clustered data by K-means clustering outperformed the results from the ARs technique based on GRI with uncluttered data.

References

Al-Hassan, M., Lu, H., & Lu, J. (2015, April). A Semantic Enhanced Hybrid Recommenation Approach: A Case Study of e-Goverment tourism service recommendation system. *Decision Support Systems, 72*, 97-109.

Choi, S. H., Kang, S., & Jeon, Y. J. (2006, October). Personalized Recommendation System Based on Product Specification Values. *Expert Systems with Applications, 31*(3), 607-616.

Dabholkar, P. A., Thorpe, D. I., & Rentz, J. O. (1996). A measure of service quality for retail stores: Scale development and validation. *Journal of the Academy of Marketing Science*, 24.

Kumar, V., & Rathee, N. (2011, March). Knowledge Discovery from Database using an Integration of Clustering and Classification. *International Journal of Advanced Computer Science and Applications, 2*(3), 29-33.

Kuscar, R., Draskovic, N., & Pavicic, J. (2018). Market Challenger Strategic Applications of CRM: A Case of Roofing Tiles Market Re-Invention In Croatia. *International Journal of Sales, Retailing and Marketing*, 73-79.

Lee, S.-L. (2010). Commodity Recommendations of Retail Busienss Based on Decision Tree Induction. *Expert Systems with Applications, 37*(5), 3685-3694.

Liu, X., & He, P. (2004). The research of Improved Association Rules Mining Apriori Algorithm. *International Conference on Machine Learning and Cybernetics*, (pp. 1577-1579).

Mason, A. (2011). *United States Patent No. US8355948B2.* Retrieved from Worldwide application: https://patents.google.com/patent/US8650072B2/en

Ou, W.-M., & Abratt, R. (2006). Diagnosing the Relationship between Corporate Reputation and Retail Patronage. *Coporate Reputation Review*, 243-257.

Paireekreng, W., Wong, K. W., & Fung, C. C. (2011). A model for mobile content filtering on non-interactive recommendation systems. *2011 IEEE International Conference on Systems, Man, and Cybernetics.* AK, USA: IEEE.

Phan, D. D., & Vogel, D. R. (2010). A model of customer relationship management and business intelligence systems for catalogue and online retailers. *Information & Management*, 69-77.

Plasse, M., Niang, N., Saporta, G., Villeminot, A., & Leblond, L. (2007). Combined use of Association Rules Mining and Clustering Methods to find Relevant Links between binary rare attributes in a large dataset. *Computational Statistics & Data Analysis*, *52*, pp. 596-613.

Ranjan, J. (2011, March). Role of Knowledge Managment and Analytical CRM in Business: Data Mining based Framework. *Learning Oranization, 18*(2), 131-148. doi:10.1108/09696471111103731

Rego, L. L., Morgan, N. A., & Fornell, C. (2013, September). Reexamining the Market Share-Customer Satisfaction Relationship. *Journal of Marketing, 77*(5), 1-20.

Resnick, M. L., Pompa, S., & Korn, I. (2004). *Persuasive Design Through Intelligent Recommendation Systems*.

Rouse, M. (2020, June 6). *TechTarget*. Retrieved from SearchCustomerExperience: https://searchcustomerexperience.techtarget.com/definition/CRM-customer-relationship-management

Soliman, O., & Adly, A. (2012). Bio-inspired Algorithm for Classification Association Rules. *The 8th International Conference on INFOrmatics and Systems (INFOS2012).* Giza, Egypt.

Thabtah, F., Cowling, P., & Peng, Y. (2005). MCAR: Multi-class classification based on Association Rule. *The third ACS/IEEE International Conference on Computer Systems and Applications.* Cairo, Egypt: IEEE.

Tiroshi, A., Berkovsky, S., Kaafar, M. A., Vallet, D., Chen, T., & Kuflik, T. (2014). Improving business rating prediction using graph based features. *Proceedings of the 19th International Conference on Intelligent User Interfaces* (pp. 17-26). New York: Association for Computing Mahinery.

Watkins, K., Napayap, N., & Fung, C. C. (2015). An Association Rules Modeling For a Programme and Activity Recommendation System Supporting Student Counseling Management for Universities in Thailand. *International Journal of Computing, Communications and Instrumentation Engg.(IJCCIE)*, (p. 1469).

Wong, W., Leung, S., Gue, Z., Zeng, X., & Mok, P. (2012, January). Intelligent Product Cross-Selling System with radio frequency Identification Technology for retailing. *International Journal of Product Economics, 135*(1), 308-319. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S0925527311003264#!

Zhou, L., Tam, K. P., & Fujita, H. (2016, January). Predicting the Listing status of Chinese listed companies with mulit-class classification models. *Information Sciences, 328*, 222-236.