Customer Care Services System (CCSS): The data-mining approach for customer care services in grocery shopping

Liliana Quyen Tang

COMP 602 – Fall 2018

Instructor: Jon Dron

Abstract

Data mining is an intriguing research field that has attracted the attention of both academic and corporate worlds. Especially in retail context, data mining is used to study patterns in customer behaviours for managers to make appropriate prediction and adjustment, or machine learning techniques to make product recommendation for customers. This paper therefore proposes a Customer Care Services System (CCSS) that applies data mining techniques to help grocery stores increase their customer satisfaction, financial performance, and reduce waste in resources. CCSS is a desktop application that directly connects to Point-of-sale system to automatically pull out data from each real-time transaction as well as to the web backend technologies from ecommerce system. This approach is a combination of existing rule-mining based, event-prediction, and web personalization systems.

Keywords: Data mining, rule-mining based systems, event-prediction systems, web personalization systems

Introduction

Over the past decades, with the increasing demand in artificial intelligence, the tremendous need to save time from shoppers, and the paramount call for attention and improvement in technology to attract more customers from retail industry, mining customer behaviours has been proving its crucial importance. Supermarkets start saving costs for hiring cashiers and associates by moving to ecommerce. Two largest retailers, Costco and Walmart, both have websites for online shopping across North America countries, including the United States and Canada while still running physical stores. Customers move toward hybrid shopping style – purchasing products both online and in-store - to save time and money while still being able to main a healthy lifestyle for meals made from fresh and carefully chosen ingredients in-store. However, there is always a gap in demand and supply in the market. Customers have become more well-informed and knowledgeable shoppers. Entrepreneurs have developed more interest in retail, so the question of attracting those knowledgeable shoppers comes into focus. To seek patterns in customer behaviours, the most effective way is through data mining, like David Ciancio, Senior Customer Strategist at Dunnhumby, once states, "Grocery companies need to think of themselves more like tech companies" (Haddon 2017).

Data mining is a growing yet challenging field due to limited access to real-life data for experimental purposes. Big companies, such as Amazon, tend to refuse to share their sales and marketing strategies (Haddon 2017). Therefore, through academic research, this paper aims to propose an approach in data mining to implement a system that acts as a placeholder for three

well-known and existing rule-mining based, event-prediction, and web-personalization systems. Beginning with a concept walk-through of three data mining algorithms employed, the paper then uses a diagram to describe business processes of the proposed system. Later, it makes clear the concept of data mining in CCSS and how data mining stands out from other statistical techniques, such as hypothesis testing. Lastly, the author ends the discussion by briefly discussing strengths and weaknesses of the proposed system.

Literature Review

Griva, A., Bardaki, C., Pramatari, K. & Papakiriakopoulos, D. (2018). Retail Business Analytics: Customer Visit Segmentation Using Market Basket Data. Expert Systems with Applications, 100 (2018), 1-16

Griva, Bardaki, Pramatari, and Papakiriakopoulos (2018) examine an effective business analytics approach in retail through a variety of studies that have a strong emphasis in the application of data mining techniques on customer behaviours. Griva et al. (2018) further categorize those approaches into two streams, corresponding to the main focus of each study, including customer segmentation and market basket. While customer segmentation mainly focuses on the whole picture of the total purchases a shopper makes, regardless of the number of their visits, market basket analysis tries to seek out how each item during a single visit relates to one another in one single purchase. They then point out the research gap with respect to the two above approaches by analyzing each individual's scope of the analysis and product taxonomy. Basket analysis is not a practical approach since it does not cover full scope of interested area and retailers need to understand a bigger picture of each visit. Customer segmentation, on the other hand, depicts a much broader extent, though is still not sufficient to provide a detailed set of information that fulfills business needs. Since each approach has their own pros and cons, the authors later propose another approach that employs clustering technique to achieve the most out of strengths from above mentioned systems while eliminating their weaknesses as much as possible. The paper wraps up with a step-by-step set of processes of the proposed framework, first covering business and data understanding process; then modelling process, such as product taxonomy adjustment, clustering sampling, and input data adjustment and clustering; and lastly evaluation process which takes into consideration both technical and business perspective. This resource greatly enriches the author's knowledge in clustering technique which is one of the three main data mining algorithms she employs in her proposed system. She completely agrees with the arguments regarding the strengths and weaknesses of two common approaches in data mining, and how clustering sample stands out from the crowd by its ability to accomplish both "segmentation" and "characterization" in studying customer behaviour through their visits.

Min, H. (2006). Developing the Profiles of Supermarket Customers through Data Mining. The Service Industries Journal, 26)7), 747-763

Min (2006) conduct an experiment to study customers behaviour by selectively distributing questionnaires to regular grocery shoppers. The author introduces and employs the concept of decision tree through "IF-THEN" statements to predict customer behaviours, so management team can have a clear target marketing or promotional actions. Furthermore, Min describes three major steps in constructing decision trees, including data collection, data formatting, and rule induction. The outcome points out some crucial implication. First, the volume of grocery purchases negatively depends on the number of stores customers visit on a regular basis and their shopping frequency. For example, if a customer often goes to store A to buy vegetables, store B

to buy meat, and store C to buy canned food, meaning they have to visit three stores, so in comparison with a customer who is able to shop everything at one store, the customer visiting more stores will have lower volume of purchases in each store. Similarly, if a customer goes shopping every week, their order will be much smaller than a customer going shopping monthly. Management team can use this finding to comfortably accommodate their customers. In particular, if the majority of customers shop with low volume of purchases, the store should have more self-checkout than cash registers run by cashiers, and vice versa. Furthermore, a store should offer a wide range of products, so customers do not need to shop at multiple stores for their needs. These accommodations in turn contribute as a remarkable win in customer satisfactory level. Secondly, Min figures that marital status also affects either a customer prefers self-checkout. In fact, married couples often go shopping for grocery together while single customers go on their own and often prefer fast checkout services, such as self-checkout. Therefore, depending if the store has more unmarried shoppers or married customers, management team should develop an innovative approach accordingly to increase customer satisfaction. The study also shows that very young (under 20) or older (over 50) often more appreciate employee courtesy which is in turn the first reason for them to become a loyal customer. To explain, due to their lack of experience or physical limitation, they often need more assistance from employees. Fourth, the outcome of the classification technique from the survey points out that senior citizens who rely on their limited retirement savings are often more sensitive to price, so supermarkets should offer more discount or promotional programs for these customers. This study strengthens the author's previous knowledge in the application of classification techniques through a real-life example that perfectly fits her scenario. Although the article does not provide much advanced technical details, in fact, data formatting phase was conducted in Excel rather than other common programming languages in data science, such as R, Python, or SQL, it effectively provides a set of specific recommendations for each outcome of data mining process.

Witten, I. & Frank, E. (2005). Data mining: practical machine learning tools and techniques. 2nd ed. San Francisco, CA: Elsevler

The book comprehensively covers machine learning – a broader term that include data mining - tools and techniques. Due to limited time and background information in data mining, the author of this paper was not able to fully understand each chapter of the book. However, the book is ranked among one of the most helpful books for a beginner and recommended by instructor of COMP 682 – Data Mining course at Athabasca University, so the author refers and circles back to this book whenever coming across a difficult term during her literature review. The book clearly explains the concept in common language with real-life examples, as such, it is very straightforward to understand. In addition, the content is clearly divided into many independent sections, so the author can skip chapters that are not particularly relevant to her research study.

Yang, Y., Liu, H. & Cai, Y. (2013). Discovery of Online Shopping Patterns Across Websites. INFORMS Journal on Computing, 25(1), 161 – 176

Yang, Liu, and Cai (2013) employ market basket analysis which is an application of association rule technique in data mining to define online shopping patterns. The authors employ OSP-Tree and OSP-Level methods to mine the complete set of frequent online shopping patterns: OSP-Tree is the most common approach due to its efficiency, though when memory is limited, OSP-Level comes into play. This paper provides a significant insight as for how data mining is used in

practice and with the real-life dataset from comScore Inc. The author of this paper particularly enjoys how the article first introduces the theory, then jumps to step-by-step processes including input, output, major steps/ algorithm for each process, and lastly displays the outcome graphically. This article adds to the author's knowledge of how association rule is used to discover online shopping patterns.

The literature review processes include other articles as well, as per references page; however, the author decides not to analyze them in depth since she does not completely agree with their techniques and opinions, but still aims to cite them to demonstrate a few of her arguments throughout the research paper.

Statement of the research problem

The author's initial scenario is to design and develop a database for a Point-of-Sale (POS) system in grocery stores, so this research further takes the full advantage of those information stored in the database by applying appropriate data mining techniques. In addition to gain customer satisfaction, the CCSS system plays a crucial role in boosting profit for retailers by incorporating three common and existing systems, including rule-mining based systems acting as a recommendation tool for customers; event-prediction systems contributing as a supply-demand forecast for store managers to order an appropriate amount of items so as to minimize waste from soon expired products; and finally web personalization system that greatly benefits e-grocery shopping model by coming up with a wide-ranging recommendation as advertisement while customer are using Internet for other websites. For instance, the author of this paper often promises to herself that she would stop spending on clothes, and advertisement about fashion deals shows up in one of the websites she was using for school and work, she usually ended up re-visiting the fashion website again and breaking promises to herself. She soon figured out to download the content of the app as a pdf file to her local storage, so she could avoid seeing those advertisement. Therefore, the author thinks web personalization system is a powerful innovation that can dramatically foster sales.

Discussion

Figure 1 depicts business processes in the proposed CCSS system. CCSS directly extracts data from the POS system.

- Data mining techniques employed in CCSS:
    - Association rule aims to associate items that somehow relate to one another; for instance, a customer who buys beer usually also buy chips (Witten & Frank, 2005).
    - Classification rule classifies data by applying decision trees, for example, IF-THEN statements (Pan et al, 2017).
    - Clustering rule groups data that is likely to fall naturally together into product taxonomy (Witten & Frank, 2005).
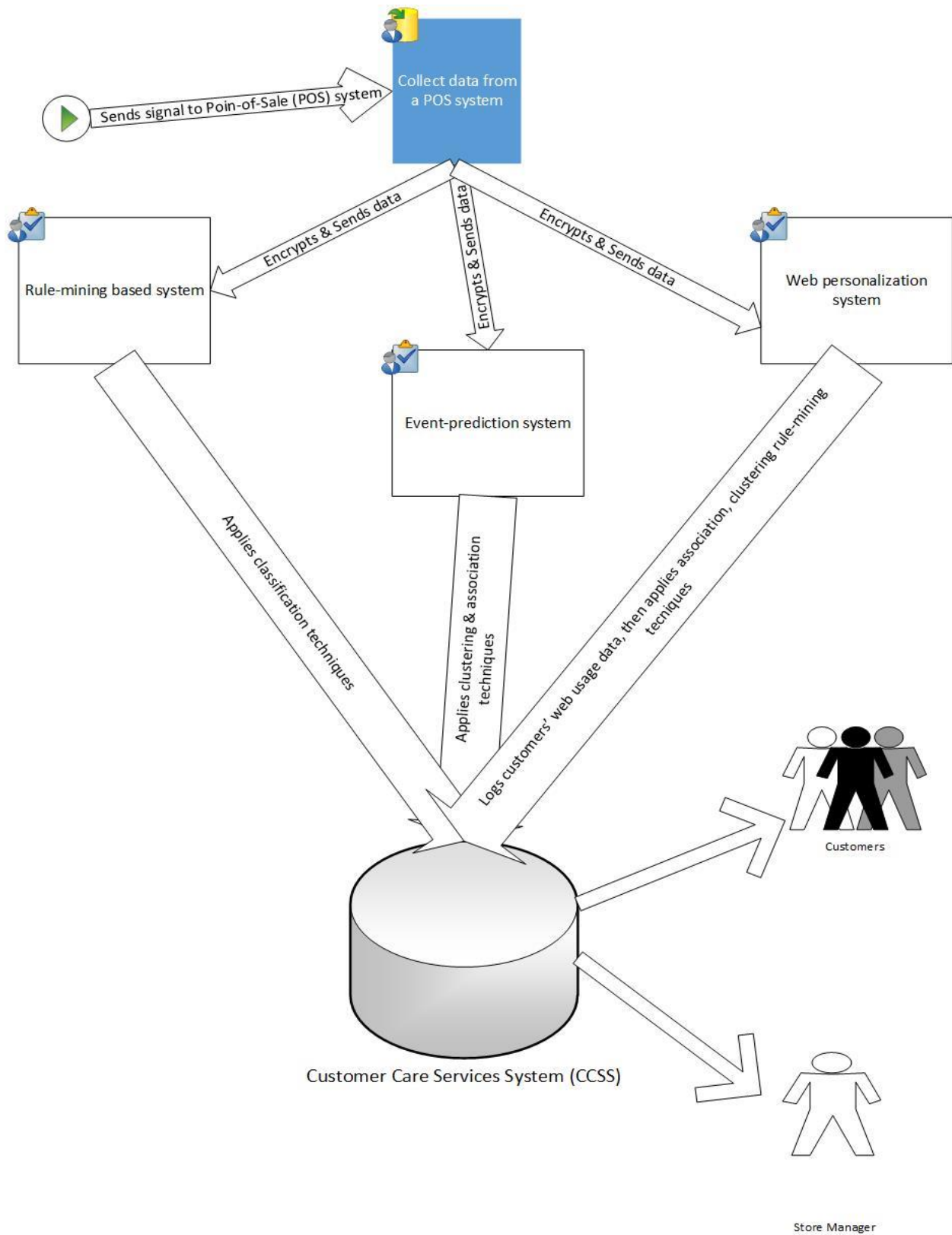
Figure 1 – Customer Services Care System Business Processes

1. *What is data mining?*

Let's remember about the first day at work when you did not know anybody in your office, and you get to know them each and everyday by the way they dress at work, the way they communicate with their coworkers, and the way they treat you, to name a few examples. These patterns form an impression about your coworkers in you. In other words, you try to make sense your colleagues' behaviours by finding patterns in their daily interaction with you and others. To predict is to find common patterns and try to make sense them. Similarly, retailers aim to predict customer behaviour and then properly implement sales and marketing strategies. They obviously should follow the same procedures as we, human-being, try to make sense the world by looking at every single transaction the store has processed. However, a huge amount of data of transactions can potentially exceed human's capacity and resources to process; hence, the need of an automatic approach to seek patterns arises, should data mining emerge. Essentially, data mining is all about "solve problems by analyzing data already present in database" (Witten & Frank, 2005).

2. *Why should we employ data mining in CCSS, but not hypothesis testing for the sake of simplicity and cost saving?*

Hypothesis testing is the most common technique used in statistics to seek trend in a certain interested category because of its simplicity, easy to apply and to learn. However, according to Min (2006), as the sample size increases, which is a must when the business grows, the level of statistical significance also increases; therefore, hypothesis testing could not be consistent. In addition, data miners and decision makers are interested in knowing more than just either to substantiate or disprove a hypothesis which is the only outcome statistical testing methods could provide.

3. *Strengths of CCSS*

First of all, it is a combination of three well-known and existing systems, so it could benefit from their outstanding features that pass the test of time. In addition, because all the information is portable in one system directly synced with POS system, it is convenient to manage and control.

4. *Weaknesses of CCSS*

Due to constraints in time and resources as well as the confidentiality of customers' data, this system was not tested with any real-life examples to ensure its feasibility.

Conclusion

The ultimate goal of this research is to delve deeper into and comprehend existing data mining approaches toward seeking customer behaviours, hence, enable retailers to implement particular sales and marketing strategies accordingly. Towards that end, the author proposes a Customer Care Services System (CCSS) that combines three existing data mining applications, including rule-mining based, event-prediction, and web-personalization systems and is directly synced with the POS system at cash registers. Though the author has gone through the literature review process to learn how those three systems are used and implemented in practice, she has not yet had a chance to execute her proposal in real-world practice, thus, some of her assumption might be subjective and theoretical. Upon given more time and resources, this proposal may be

implemented to become a functional system, hence, can clearly interpret the feasibility of the proposed solution.

References

Griva, A., Bardaki, C., Pramatari, K. & Papakiriakopoulos, D. (2018). Retail Business Analytics: Customer Visit Segmentation Using Market Basket Data. Expert Systems with Applications, 100 (2018), 1-16

Haddon, H. (2017). The future of grocery shopping is all about data. The Wall Street Journal. https://www.marketwatch.com/story/the-future-of-grocery-shopping-is-all-about-data-2017-10-24. Accessed 07.12.2018

Min, H. (2006). Developing the Profiles of Supermarket Customers through Data Mining. The Service Industries Journal, 26)7), 747-763

Pan, S., Giannikas, V., Han, Y., Grover-Silva, E. & Qiao, B. (2017). Using customer-related data to enhance e-grocery home delivery. Industrial Management & Data Systems, 117(9), 1917 – 1933

Rategari, H. & Sap, M. (2008). Data mining and e-commerce: methods, applications, and challenges. Jurnal Teknologi Maklumat, 20(2), 116 – 128

Wang, L, (2016). What SQL Analysts Need to Know About Python. Segment. https://segment.com/blog/sql-vs-python/. Accessed 03.12.2018

Willems, K. (2015). Choosing R or Python for Data Analysis? An Infographic. DataCamp. https://www.datacamp.com/community/tutorials/r-or-python-for-data-analysis. Accessed 03.12.2018

Witten, I. & Frank, E. (2005). Data mining: practical machine learning tools and techniques. 2nd ed. San Francisco, CA: Elsevler

Yang, Y., Liu, H. & Cai, Y. (2013). Discovery of Online Shopping Patterns Across Websites. INFORMS Journal on Computing, 25(1), 161 – 176