

# Association Rule Mining Approach for Customer Relationship Management

Tamanna Kachwala<sup>\*1</sup>, L. K. Sharma<sup>2</sup>

<sup>\*1</sup>Ph. D. Scholar, Rai University, Saroda, Ahmedabad, Gujarat, India

<sup>2</sup>National Institute of Occupational Health, Ahmedabad, Gujarat, India

## ABSTRACT

The customer relationship management is the main goal of every business organization. In this competitive business scenario, every activity starts and ends with the customer. The increasing competition and dynamic environment, every firms needs to identify, anticipate and satisfy consumers to maximize profit. The association rule mining algorithms are using to work out data mining problem in a trendy way. Along with the large variety of existing approaches, it is constantly challenging to select the best possible algorithm for rule based mining task. Generally, empirical methods for evaluating algorithms of association rule mining are based completely on quantitative measures such as correlation between minimum support, a number of rules or frequent item sets and data processing time. In this paper, we are evaluating best rules found by applying the Apriori algorithm of association rule sets. We show that observing rule overlapping, support and confidence in two compared rule sets help evaluate algorithm quality or the measure uniformity of source datasets.

**Keywords:** Customer relationship management, Association Rule Mining, Metrics in Association, Apriori.

## I. INTRODUCTION

Customer Relationship Management (CRM) is a business values involving analysing, understanding and better providing for your customers while building a relationship with each customer to improve customer fulfilment and maximize profits. It is about considerably, forecasting and reacting to customers' needs. Day by day, as the economy is growing, businesses are growing rapidly, for any organization; their customers play the key role to get the remarkable success of the business. Thus, by realizing the customer's importance in the success of any business many organizations shifted their focus from product, process centric to customer centric approach. This changed focus and scenario created the demand of integrated approach with which organization can provide better product-service, maintain relationship with customers, better communication with customers and to maintain the

loyalty of customers. Such requirement was fulfilled by CRM [1][2].

Data mining refers as a growing technology that has made an innovative change in the information world [1][2]. Association rule mining is a broadly used approach in data mining. Association rules are able of declaring every exciting relationship in a large database. The huge amount of information taking in the set of association rules can be used not only for relating the relationships in the database, but also for between other kinds or classes of database instances. The goal of this thought is to discover frequent patterns, interesting correlations, and associations between sets of items in the transaction databases or other data repositories.

The association rule mining can be used to reveal the association among different customer attribute which

will guide to improve the customer satisfaction and retain the customer [2].

The rest of the paper is organized as follows. Next section the related studies on association rule mining is reported. Section 3 will describe the methodology. Experiment and result analysis are provided in the section 4. Finally, the work will be concluded in the section 5.

## II. RELATED STUDY

The most important theme of data mining is the Association Rules. Association Rules can be found that relationship among the different attributes of the dataset. Association rules are first proposed by Agrawal and Srikant [3]. Association rules are necessary to Customer Relationship Management [4]. Association rules can be used to establish customer classification model, which may be used for facilitating sales in many industries, such as mobile telecom market and so on [5]. Apriori algorithm is an important algorithm for customer analysis. It can be used to bank customer segmentation [6], Market based analysis [7] and so on. Apriori algorithm is a process of analysing a huge data from variable perspectives and summarizing it into information and knowledge. It can be helping Researcher to establish a model to identify frequency of customer relationship management in particular geographical area with the aid of association rules analyzed by [8]. The amount of open information available online from heterogeneous sources and domains is growing very quickly, and constitutes an important body of knowledge to support Social CRM. These data sources may disclose significant business opportunities and competitive advantage to those who are able to understand and leverage their value [9].

Moreover, Social CRM benefits from Data because it facilitates more accurate decision-making and a more efficient distribution of knowledge among the customers and the company [10]. It is noteworthy that companies are harnessing the power of Data and

analytics to apply it in customer relationship management [11][12].

In general, most of the existing proposals for mining patterns and associations between patterns follow an optimization based on objective quality measures. This huge attraction for this type of metrics lies in the fact that pattern mining mainly aims at discovering hidden and previously unknown knowledge from datasets, so there is no possibility to compare the extracted knowledge with the subjective knowledge provided by the expert. Besides, the knowledge of two different users into a specific field can differ greatly, which causes inaccuracy in the metrics. Thus, the values obtained for different subjective metrics cannot be properly compared [13]. In this study, popular four objective metrics are considered and experimented. Based on these four metrics best rules are identified.

## III. METHODS AND MATERIAL

### A. Data

In this study, Bank marketing data set is used [14]. The dataset was acquired from UCI Machine Learning Repository [15]. The data is related with direct marketing campaigns. The marketing campaigns were based on phone calls for the accessing the bank product such as bank term deposit would be or not subscribed. The dataset contains different 45211 instances and 17 attributes such as age, job, marital, education, balance, personal loan, housing loan, credit information, outcome etc.

### B. Association Rule Mining Method

Association rule learning is an important and a well-known method for understanding the relations among variables in huge databases. There are several rules likely even from a so small database, so to select the exciting ones. The constraints on different measures of interest is important. Apriori is the mainly standard and well-known algorithm for mining common patterns. It was introduced by [3]. It

is based on the Apriori property that is useful for trimming unrelated data. It states that any subset of frequent item sets must be frequent. In order to choose attractive rules from the set of all probable rules, constraints on various measures of significance and interest are used. The common metrics to measure the association are minimum thresholds of support and confidence, Leverage, Lift and Conviction. Objective metrics are defined from a statistical point of view and they provide structural properties of data. Some of these metrics have been widely used to evaluate the interest of association patterns, determining that the stronger is the dependence relationship, the more interesting is the pattern. Support, confidence, lift, leverage, and conviction are some of these objective metrics, which are defined in terms of the frequency of occurrence of the patterns [13]. The confidence of a rule is defined as follows:

$$conf(X \rightarrow Y) = \frac{supp(X \cup Y)}{supp(X)}$$

It can also be interpreted as the conditional probability  $P(Y|X)$ , i.e., the probability of finding the item set  $Y$  in transactions given the transaction previously contains  $X$ . It can give some significant insights, but it also has a major disadvantage. It only takes into account the popularity of the item set  $X$  and not the popularity of  $Y$ . If  $Y$  is evenly known, as  $X$  then there will be a superior probability that a transaction containing  $X$  will also contain  $Y$  thus increasing the confidence. To conquer this disadvantage there is a different measure called lift. This measure calculates the degree of dependence between the antecedent  $X$  and the consequent  $Y$  of an association rule, obtaining a value  $< 1$  if they are negative dependent; a value  $> 1$  if they are positive dependent; and 1 in case of independence [13]. Lift can be defined as follows:

$$lift(X \rightarrow Y) = \frac{supp(X \cup Y)}{supp(X) * supp(Y)}$$

It is the ratio of the probability that  $X$  and  $Y$  occur jointly to the many of the two individual probabilities for  $X$  and  $Y$ .

The proportion of further examples covered by both the premise and the consequence above those expected if the premise and consequence were independent of each other. This can be represented by using Leverage. It can be defined as follow:

$$Leverage(X \rightarrow Y) = \frac{P(X, Y)}{P(X)P(Y)}$$

Conviction is similar to lift, but it measures the effect of the right-hand-side not being true. Conviction represents the degree of implication of a rule, and values far from the unity indicate interesting rules. Unlike lift, rules that hold 100% of the time have the highest possible conviction value. The confidence measure also has this property, providing a value of 1 (the maximum value for confidence) to these rules [13].

$$conv(X \rightarrow Y) = \frac{1 - supp(Y)}{1 - conf(X \rightarrow Y)}$$

#### IV. RESULTS AND DISCUSSION

From the dataset, association rules from the frequent item sets are generated using the Apriori algorithm by Weka [16]. The methods were apply by taking four metrics support and confidence, lift, leverage, and conviction and result were compared. From that, some interesting association rules were identified and now we want to evaluate how useful they are. As an example, we have found the following rules with Minimum support: 0.3, Minimum metric  $\langle$ confidence $\rangle$ : 0.6 and Number of cycles performed: 14. The best rules were depicted in the Table I.

**Table 1.** Best Association Rules

Attributes List	Best Rules Found
Age	1. education=tertiary balance=1 loan=no 1715 ==> y=yes 1715 conf:(1)
Job	2. education=tertiary balance=1 loan=no campaign=1 1715 ==> y=yes 1715
Marital	conf:(1)
Education	3. education=tertiary balance=1 loan=no previous=2 1715 ==> y=yes 1715 conf:(1)
Default	4. education=tertiary balance=1 loan=no campaign=1 previous=2 1715 ==> y=yes
Balance	1715 conf:(1)
Housing	5. education=tertiary default=no balance=1 loan=no 1711 ==> y=yes 1711 conf:(1)
Loan	6. education=tertiary default=no balance=1 loan=no campaign=1 1711 ==> y=yes
Contact	1711 conf:(1)
Day	7. education=tertiary default=no balance=1 loan=no previous=2 1711 ==> y=yes
Month	1711 conf:(1)
Duration	8. education=tertiary default=no balance=1 loan=no campaign=1 previous=2 1711
Campaign	==> y=yes 1711 conf:(1)
Pdays	9. education=tertiary balance=1 loan=no duration=1 1708 ==> y=yes 1708 conf:(1)
Previous	10. education=tertiary balance=1 loan=no duration=1 campaign=1 1708 ==> y=yes
Poutcome	1708 conf:(1)
	11. education=tertiary balance=1 loan=no duration=1 previous=2 1708 ==> y=yes
	1708 conf:(1)
	12. education=tertiary balance=1 loan=no duration=1 campaign=1 previous=2 1708
	==> y=yes 1708 conf:(1)
	13. education=tertiary default=no balance=1 loan=no duration=1 1704 ==> y=yes
	1704 conf:(1)
	14. education=tertiary default=no balance=1 loan=no duration=1 campaign=1 1704
	==> y=yes 1704 conf:(1)
	15. education=tertiary default=no balance=1 loan=no duration=1 previous=2 1704
	==> y=yes 1704 conf:(1)
	16. education=tertiary default=no balance=1 loan=no duration=1 campaign=1
	previous=2 1704 ==> y=yes 1704 conf:(1)

## V. CONCLUSION

The association rules perform a key role in various data mining applications, trying to find interesting patterns in data bases. Apriori is the simplest algorithm which is used for mining of frequent patterns from the database. If we look at the above rules we can see that the customers who have tertiary education, none of loan taken and their average bank balance is 1 (upto Rs.50000) are subscribed a term deposit in the bank. The confidence of this rule is 100% so it is very believable.

Using the same logic, we can interpret all the other rules that the algorithm has revealed. Apriori algorithm uses large item set property, easy to implement.

## VI. REFERENCES

- [1]. T. Kachwala, L. K. Sharma (2017), Comparative Study of Supervised Learning in Customer Relationship Management, International Journal of Computer Engineering & Technology, 8(6), pp. 77-82. ISSN 0976-6367

- [2]. K. Tsipstis and A. Chorianopoulos (2009), *Data Mining Techniques in CRM: Inside Customer Segmentation*, John Wiley and Sons, Ltd., ISBN: 978-0-470-74397-3
- [3]. Agrawal, R. and Srikant, R. (1994). Fast algorithms for mining association rules. In Proc. 20th Int. Conf. Very Large Data Bases, 487-499.
- [4]. Gu, G. And Y. He, (2010). Reserch on the applications of data mining to Customer Relationship Management in the mobile communication industry. Proceedings of the 3rd IEEE international conference on computer science and information technology, July 9-11, 2010, Chengdu, China, pp: 597-599.
- [5]. Ahn, H., J. J Ahn, K. J. Oh and D. H. Kim, (2011). Facilitating cross-selling in a mobile telecom market to develop customer classification model based on Hybrid data mining techniques. Expert system and applications. Pp: 5005-5012.
- [6]. Yung, G. X. (2013). The research of improved apriori mining algorithm in bank customer segmentation. Proceeding of the 2nd international conference on optoelectronics engineering and material research. Oct 19-20, 2013. Zhengzhou, China, pp: 2244-2249.
- [7]. Balaji, M. And G. S. V. Rao, (2013). An adaptive implementation case study of apriori algorithm for a retail scenario in a cloud environment. Proceeding of the 13th IEEE/ACM international Symposium on cluster, cloud and Grid computing, May 13-16, 2013, Delft, Netherlands pp: 625-629.
- [8]. Ilayaraja, M. And T. Meyyappan, (2013). Mining medical data to identify frequent diseases Using Apriori algorithm. Proceeding of the international conference on Pattern reorganization, Informatics and Mobile engineering, Feb 21-22, pp: 194-199.
- [9]. Torre-Bastida AI, Villar-Rodriguez E, Gil-Lopez S, Del Ser J (2015) Design and implementation of an extended corporate CRM database system with big data analytical functionalities. *J Univ Comput Sci* 21(6):757–776
- [10]. Anshari M, Alas Y, Yunus N, Sabtu NI, Hamid MH (2015) Social customer relationship management and student empowerment in online learning systems. *Int J Electron Cust Relat Manage* 9(2/3):104–121
- [11]. Marshall A, Mueck S, Shockley R (2015) How leading organizations use big data and analytics to innovate. *Strateg Leadersh* 43(5):32–39
- [12]. A.M. Khattak, A.M. Khan, Sungyoung Lee and Young-Koo Lee (2010) “Analyzing Association Rule Mining and Clustering on Sales Day Data with XLMiner and Weka” In *International Journal of Database Theory and Application*, Volume 3, No. 1.
- [13]. S. Ventura, J.M. Luna (2016), *Pattern Mining with Evolutionary Algorithms*, Springer, pp. 27-44, DOI 10.1007/978-3-319-33858-3\_2
- [14]. S. Moro, P. Cortez and P. Rita (2014). A Data-Driven Approach to Predict the Success of Bank Telemarketing. *Decision Support Systems*, Elsevier, 62:22-31.
- [15]. Dua, D. and Karra Taniskidou, E. (2017). *UCI Machine Learning Repository* <http://archive.ics.uci.edu/ml>. Irvine, CA: University of California, School of Information and Computer Science.
- [16]. Eibe Frank, Mark A. Hall, and Ian H. Witten (2016). *The WEKA Workbench*. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.