International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 10
June 2016

Study of Various Association Rule Mining Techniques with Positive And Negative Integration

PRATEEK KUMAR SINGH[1]
M. tech Student
CSE Department, RGPV University, Madhya Pradesh
Lakshmi Narain College of Technology, Jabalpur, MP, India

NAAZISH RAHIM[2]
Assistant Professor
HOD, CSE Department, RGPV University, Madhya Pradesh
Lakshmi Narain College of Technology, Jabalpur, MP, India

SUJEET TIWARI[3]
Assistant Professor
CSE Department, RGPV University, Madhya Pradesh
Lakshmi Narain College of Technology, Jabalpur, MP, India

NEELU SAHU[4]
Assistant Professor
IT Department, CSVTU University, Chhattisgarh
Government Engineering College, Bilaspur, Chhattisgarh, India

ABSTRACT

Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets. Association Rule Mining (AM) is one of the most popular data mining techniques. Association rule mining generates a large number of rules based on support and confidence. However, post analysis is required to obtain interesting rules as many of the generated rules are useless. In this paper, we provide some fundamental concepts related to association rule mining and survey the record of existing association rule mining methods with positive and negative integration. Obviously, a single article cannot be a entire review of the entire algorithms, yet we wish that the references cited will cover up the major theoretical issues, guiding the researcher in motivating research information that have yet to be explored.

I. INTRODUCTION

Data mining is used to deal with very large amount of data which are stored in the data ware houses and databases, to find out desired interesting knowledge and information. Many data mining techniques have been proposed such as, association rules, decision trees, neural networks, etc. It has become the point of attention from many years.

One of the most well-known techniques is association rule mining. It is the most efficient data mining technique. It discovers the hidden patterns from the large databases. It is responsible to find the relationship between the different attributes of data. Association rules are extracted by Agrawal et al in 1993 [1], it is one of the most important research field in data mining, it reveals the potential useful relationship in large-scale affairs among each item sets. From the celebrated Apriori algorithm [2] there have been a remarkable number of variants and improvements of association rule mining algorithms [3]. A typical example of association rule mining application is the market basket analysis. In this process, the behavior of the customers is studied with reference to buying different products in a shopping store. The discovery of interesting patterns in this collection of data can lead to important marketing and management strategic decisions. For instance, if a customer buys bread, what are chances that customer buys milk as well? Depending on some measure to represent the said chances of such an association, marketing personnel can develop better planning of the shelf space in the store or can base their discount strategies on such associations/correlations found in the data. All the traditional association rule mining algorithms were developed to find positive associations between items. By positive associations, we refer to associations between items exist in transactions containing the items bought together.

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 10
June 2016

What about associations of the type: "customers that buy Coke do not buy Pepsi" or "customers that buy juice do not buy bottled water"? In addition to the positive associations, the negative association can provide valuable information, in devising marketing strategies.

Positive and Negative FP Rule Mining:

Author of [4] cleverly explain the concept of positive and negative association rules. According to the [4] two indicators are used to decide the positive and negative of the measure:

i. Firstly find out the correlation according to the value of Corr(P,Q) = sup (P U Q) / sup (P) sup (Q)
which is used to delete the contradictory association rules emerged in mining process.
There are three measurements possible of corr(P,Q) [5]:
a. If >1, Then P and Q are related;
b. If =1, Then P and Q are independent of each other;
c. If <1, Then P and Q negative correlation;

ii. Support and confidence is the positive and negative association rules in two important indicators of the measure. The support given by the user to meet the minimum support (minsupport) a collection of item sets called frequent item sets, association rules mining to find frequent item sets is concentrating on the needs of the user to set the minimum confidence level (minconf) association rules.

II. BASIC CONCEPTS AND TERMINOLOGIES

This section introduces association rules terminology and some related work on negative association rules.

2.1 Association rules

Formally, association rules are defined as follows: Let I = $\{i_1, i_2, \ldots i_n\}$ be a set of items. Let D be a set of transactions, where each transaction T is a set of items such that T I. Each transaction is associated with a unique identifier TID. A transaction T is said to contain X, a set of items in I, if X T. An association rule is an implication of the form "Xà Y", where X I; Y I, and X ∩ Y = Φ. The rule X à Y has support s in the transaction set D if s% of the transactions in D contain X U Y. In other words, the support of the rule is the probability that X and Y hold together among all the possible presented cases. It is said that the rule X à Y holds in the transaction set D with confidence c if c% of transactions in D that contain X also contain Y. In other words, the confidence of the rule is the conditional probability that the consequent Y is true under the condition of the antecedent X. The problem of discovering all association rules from a set of transactions D consists of generating the rules that have a support and confidence greater than given thresholds. These rules are called strong rules, and the framework is known as the support-confidence framework for association rule mining. Definition of Negative Association Rule A negative association rule is an implication of the form X à ¬ Y (or ¬ X à Y or ¬ X à ¬ Y), where X I, Y I and X ∩ Y = Φ (Note that although rule in the form of ¬ X à ¬ Y contains negative elements, it is equivalent to a positive association rule in the form of Y àX. Therefore it is not considered as a negative association rule.) In contrast to positive rules, a negative rule encapsulates relationship between the occurrences of one set of items with the absence of the other set of items. The rule X à ¬ Y has support s % in the data sets, if s % of transactions in T contain itemset X while do not contain itemset Y. The support of a negative association rule, supp(X à ¬ Y), is the frequency of occurrence of transactions with item set X in the absence of item set Y . Let U be the set of transactions that contain all items in X. The rule X à ¬ Y holds in the given data set (database) with confidence c %, if c% of transactions in U do not contain item set Y. Confidence of negative association rule, conf (X à ¬ Y), can be calculated with P(X ¬ Y)/P(X), where P(.) is the probability function. The support and confidence of itemsets are calculated during iterations. However, it is difficult to count the support and confidence of non-existing items in transactions. To avoid counting them directly, we can compute the measures through those of positive rules.

III. RELTED WORK IN NEGATIVE ASSOCIATION RULE MINING

We give a short description of the existing algorithms that can generate positive and negative association rules.

The concept of negative relationships mentioned for the first time in the literature by Brin et.al [6]. To verify the independence between two variables, they use the statistical test. To verify the positive or negative relationship, a correlation metric was used. Their model is chi-squared based. The chi-squared test rests on the normal approximation to the binomial distribution (more precisely, to the hyper geometric distribution). This approximation breaks down when the expected values are small.

A new idea to mine strong negative rules presented in [7]. They combine positive frequent itemsets with domain knowledge in the form of taxonomy to mine negative associations. However, their algorithm is hard to generalize since it is domain dependent and requires a predefined taxonomy. Finding negative itemsets involve following steps: (1) first find all the generalized large itemsets in the data (i.e., itemsets at all levels in the taxonomy whose support is greater than the user specified minimum support) (2) next identify the candidate negative itemsets based on the large itemsets and the taxonomy and assign them expected support. (3) in the last step, count the actual

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 10
June 2016

support for the candidate itemsets and retain only the negative itemsets .The interest measure RI of negative association rule X à ┐Y, as follows RI=(E[support(X U Y)]-support(X U Y))/support(X) Where E[support(X)] is the expected support of an itemset X.

A new measure called mininterest (the argument is that a rule A à B is of interest only if supp(A U B)-supp(A) supp(B) ≥ mininterest) added on top of the support-confidence framework[8]. They consider the itemsets (positive or negative) that exceed minimum support and minimum interest thresholds as itemsets of interest. Although, [9] introduces the "mininterest" parameter, the authors do not discuss how to set it and what would be the impact on the results when changing this parameter.

A novel approach has proposed in [10]. In this, mining both positive and negative association rules of interest can be decomposed into the following two sub problems, (1) generate the set of frequent itemsets of interest (PL) and the set of infrequent itemsets of interest (NL) (2) extract positive rules of the form A=>B in PL, and negative rules of the forms A à ┐ B, ┐ AàB and ┐ A à ┐ B in NL. To generate PL, NL and negative association rules they developed three functions namely, fipi(), iipis() and CPIR().

The most common frame-work in the association rule generation is the "Support-Confidence" one. In [11], authors considered another frame-work called correlation analysis that adds to the support-confidence. In this paper, they combined the two phases (mining frequent itemsets and generating strong association rules) and generated the relevant rules while analyzing the correlations within each candidate itemset. At the end, they keep only those rules generated from item combinations with strong correlation. If the correlation is positive, a positive rule is discovered. If the correlation is negative, two negative rules are discovered. The negative rules produced are of the form X à ┐Y or ┐ X à Y which the authors term as "confined negative association rules". Here the entire antecedent or consequent is either a conjunction of negated attributes or a conjunction of non-negated attributes.

An innovative approach has proposed in [12]. In this generating positive and negative association rules consists of four steps.

(i) Generate all positive frequent item sets L (P1).
(ii) for all itemsets I in L(P_1), generate negative itemsets of the form ┐ (I1 I2).
(iii) Generate all negative frequent itemsets.
(iv) Generate all negative frequent itemsets I1.
(v) Generate all valid positive and negative association additional interesting measure(s) to support-confidence.

VARIOUS PAPERS STUDIED

(i) IMPROVED ASSOCIATION RULE MINING WITH POSITIVE AND NEGATIVE INTEGRATION

Existing work based on Apriori algorithm uses candidate sets for finding frequent pattern to generate association rules, then apply class label association rules where this work uses FP-Tree with growth for finding frequent pattern to generate association rules. Apriori algorithm takes more time for large data set where FP growth is time efficient to find frequent pattern in transaction.

Technique. For this we have integrated the concept of positive and negative association rules into the frequent pattern (FP) method. Negative and positive rules works better than traditional association rule mining and FP cleverly works in large database. Our proposed algorithm has two stages:

a. **Rule Generation**
b. **Classification**

In the first stage, the algorithm calculates the whole set of positive and negative class association rules such that sup(R) support and conf(R) confidence given thresholds. Furthermore, the algorithm prunes some contradictory rules and only selects a subset of high quality rules for classification.

In the second stage i.e. classification, for a given data object, the algorithm extracts a subset of rules found in the first stage matching the data object and predicts the class label of the data object by analyzing this subset of rules.

Rule Generation:

To find rules for classification, the algorithm first mines the training dataset to find the complete set of rules passing certain support and confidence thresholds. This is a typical frequent pattern or association rule mining task. The algorithm adopts FP Growth method to find frequent itemset. FP Growth method is a frequent itemset mining algorithm which is fast. The algorithm also uses the correlation between itemsets to find positive and negative class association rules. The correlation between itemsets can be defined as:

$$Corr(X,Y) = \sup(X \cup Y) / \sup(X)\sup(Y)$$

Where X and Y are itemsets.

When corr(X, Y)>1, X and Y have positive correlation. When corr(X, Y)=1, X and Y are independent.

When corr(X, Y)<1, X and Y have negative correlation.

Also when corr(X, Y)>1, we can deduce that corr(X, -Y)<1 and corr(-X,Y)<1.

So, we can use the correlation between itemset X and class label c_i to judge the class association rules.

When corr(X, ci)> 1, we can deduce that there exists the positive class association rule X → ci

International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 10
June 2016

When corr(X, ci)> 1, we can deduce that there exists the negative class association rule X→-ci

The Rule Generation algorithm works as follow:

Definition FP-tree: A frequent-pattern tree (or FP-tree) is a tree structure defined below:

a. It consists of one root labeled as "null", a set of item-prefix subtrees as the children of the root, and a frequent-item-header table.
b. Each node in the item-prefix subtree consists of three fields: item-name, count, and node-link, where item-name registers which item this node represents, count registers the number of transactions represented by the portion of the path reaching this node, and node-link links to the next node in the FP-tree carrying the same item-name, or null if there is none.
c. Each entry in the frequent-item-header table consists of two fields - (1) item-name and (2) head of node-link (a pointer pointing to the first node in the FP-tree carrying the item-name).

In this paper, we have proposed a new hybrid approach for data mining process. Data mining is the current focus of research since last decade due to enormous amount of data and information in modern day. Association is the hot topic among various data mining technique. In this article we have proposed a hybrid approach to deal with large size data. Proposed system is the enhancement of Frequent pattern (FP) technique of association with positive and negative integration on it. Traditional FP method performs well but generates redundant trees resulting efficiency degrades. To achieve better efficiency in association mining, positive and negative rules generation helps out. Same concept has been applied in the proposed method. Results shows that proposed method perform well and handles very large size of data set.

(ii) MINING POSITIVE AND NEGATIVE ASSOCIATION RULES

Association rule mining is one of the most popular data mining techniques to find associations among items in a set by mining necessary patterns in a large database. Typical association rules consider only items enumerated in transactions. Such rules are referred to as positive association rules. Negative association rules also consider the same items, but in addition consider negated items (i.e. absent from transactions). Negative association rules are useful in market-basket analysis to identify products that conflict with each other or products that complement each other. They are also very useful for constructing associative classifiers. In this paper, we propose an algorithm that mines positive and negative association rules without adding any additional measure and extra database scans.

A new idea to mine strong negative rules presented in [13]. They combine positive frequent itemsets with domain knowledge in the form of taxonomy to mine negative associations. However, their algorithm is hard to generalize since it is domain dependent and requires a predefined taxonomy. Finding negative itemsets involve following steps: (1) first find all the generalized large itemsets in the data (i.e., itemsets at all levels in the taxonomy whose support is greater than the user specified minimum support) (2) next identify the candidate negative itemsets based on the large itemsets and the taxonomy and assign them expected support. (3) in the last step, count the actual support for the candidate itemsets and retain only the negative itemsets .The interest measure RI of negative association rule X ? ┐Y, as follows RI=(E[support(X U Y)]-support(X U Y))/support(X) Where E[support(X)] is the expected support of an itemset X.

A new measure called mininterest (the argument is that a rule A ? B is of interest only if supp(A U B)-supp(A) supp(B) ≥ mininterest) added on top of the support-confidence framework[14]. They consider the itemsets (positive or negative) that exceed minimum support and minimum interest thresholds as itemsets of interest. Although, [15] introduces the "mininterest" parameter, the authors do not discuss how to set it and what would be the impact on the results when changing this parameter.

In this paper, we proposed an algorithm that mines both positive and negative association rules. Our method generates positive and negative rules with existing support-confidence framework and no extra scan is required for mining negative association rules. We conducted experiments on synthetic data set. It is producing larger number of negative rules. In future we wish to conduct experiments on real datasets and compare the performance of our algorithm with other related algorithms.

(iii) EXCEPTION RULES MINING BASED ON NEGATIVE ASSOCIATION RULES.

Exception rules have been previously defined as rules with low interest and high confidence. In this paper a new approach to mine exception rules will be proposed and evaluated. Interconnection between exception and negative association rules will be considered. Based on the knowledge about negative association rules in the database, the candidate exception rules will be generated. A novel exceptionality measure will be proposed to evaluate the candidate exception rules. The candidate exceptions with high exceptionality will form the final set of exception rules. Algorithms for mining exception rules will be developed and evaluated.

The project considers the interconnection between negative association rules and exceptions rules. The exceptions rules

mining algorithm employs the knowledge about the negative association rules in the database and generates candidate exceptional item sets. The candidate item sets exceptionality measure is then verified. If the exceptionality satisfies the minimum exceptionality constraint, the candidate exceptional item sets will be listed in the output of the algorithm as exceptional item sets.

In the future work we are going to consider temporal exceptions, which are the temporal patterns in the database related with negative association rules and changing over time. Additional measures will be considered to distinct the temporal exceptions in the database.

(iv) A Study of Negative Association Rules Mining Algorithm Based on Multi-Database

the mutual exclusion relationships among data items are reflected by negative association rules, which is very important on the decision-making analysis. In the last several years, negative association rules are frequently researched, while the study object of it is single mining of database now. With the development of database technology, multi-database mining is more and more important. On the basis of analyzing the related technology, research status and shortage of present negative association rules mining , the selecting rules, weighted synthesis and algorithm are discussed on multi-datobase.

Multi-Database mining are based on decision problem of knowledge discovery under global enterprise distribution data and discover novel useful model process from different databases. It is more difficult excavating negative association rules in multiple databases than that in single database. But we can use a single database mining knowledge to excavate negative association rules in the multiple database, the idea is: Firstly, the multiple databases can be classified according to a certain rules, eliminating ambiguity caused by different database [16]; Secondly, the similar data in each databade

can be pretreatmented, such as removing meaningless redundant noise rules and making database become cleaner; Lastly, making a knowledge synthesis excavated from each same type of databases. Therefore, the multi-database mining generally is divided into three steps: First, classify the database. Second, extract knowledge from the same database. Third, make knowledge synthesis from the same database mining, which generally adopt weighted method for synthesis of all the information in the database.

In this paper, above algorithm is correct and feasible certificated by the experimental data. the multi-databases contain different kinds of databases, the data mining is discussed about the same kind of database only. How to excavate the negative association mining among different kinds of database is still a new research direction, which needs to be further studied. In addition, there are many factors for the mining of multi-database to be considered. And there are many methods for the selection of negative association rules, while which is discussed in this paper with only one method .How to select rules based on different factors still needs to be further researched.

IV. CONCLUSION

Association rule mining has a lot of applications such market basket analysis, medical diagnosis, Website navigation examination, Native soil security and so on. In this paper, we surveyed the list of existing association rule mining methods and discussed about its advantages. We have studied about the various exception rule based on negative association rules, improved association rule mining with positive and negative integration, mining positive and negative association rules and gives an opportunity to analyses them .

REFERENCES

[1] Agrawal R, Imielinski T, Swami A. Mining association rules between sets of items in large databases. Proc of ACM SIGMOD Int Conf Management of Date [C]. Washington D C,

[2] Agrawal , R ; Srikant,R. (1994). *Fast algorithms for mining association rules.* In VLDB, Chile.

[3] Han, J; Pei, J; Yin, Y.(2000). *Mining frequent patterns* without candidate generation. In SIGMOD, dallas,Texas.

[4] Yanguang Shen, Jie Liu and Jing Shen "The Further Development of Weka Base on Positive and Negative Association Rules", IEEE, International Conference on Intelligent Computation Technology and Automation (ICICTA), 2010.

[5] Yanguang Shen,Jie Liu,Fangping Li.Application Research on Positive and Negative Association Rules Oriented Software Defects,2009 International Conference on Computational Intelligence and Software Engineering(CISE 2009)[C].Wuhan,China,December 11-13,2009.

[6] Brin,S; Motwani, R; Silverstein,C. (1997).*Beyond Market* Baskets: Generalizing Association Rules to *Correlations,*" Proc. ACM SIGMOD Conf., pp.265-276.

[7] Pradip Kumar Bala. (2009).A Technique for Mining Negative *Association Rules* . Proceedings of the 2nd Bangalore Annual Compute Conference .

[8] Wu, X; Zhang, C; Zhang, S.(2002). *Mining both positive and* negative association rules.In: Proc. of ICML. 658–665

[9] Teng, W; Hsieh, M; Chen, M.(2002). *On the mining of* substitution rules for statistically dependent items. In: Proc. of ICDM. 442–449

[10] Wu, X; Zhang, C; Zhang, S.(2004).Efficient m*ining both positive and negative association rules*. ACM Transactions on Information Systems, Vol. 22, No.3,Pages 381-405.

[11] Antonie,M.L; Za¨ıane,O.R.(2004). *Mining Positive and* Negative Association Rules: an Approach for Confined *Rules*, Proc. Intl. Conf. on Principles and Practice of Knowledge Discovery in Databases, pp 27–38.

[12] Chris Cornelis; peng Yan; Xing Zhang; Guoqing Chen.(2006).Mining Positive and Negative Association Rules from Large Databases , IEEE conference.

[13] Savasere, A; Omiecinski, E; Navathe, S.(1998). *Mining for* Strong negative associations in a large data base of *customer transactions*. In: Proc. of ICDE. 494- 502

[14] Wu, X; Zhang, C; Zhang, S.(2002). *Mining both positive and* negative association rules.In: Proc. of ICML. 658–665

[15] Teng, W; Hsieh, M; Chen, M.(2002). *On the mining of* substitution rules for statistically dependent items. In: Proc. of ICDM. 442–449

[16] SuYiJuan, YanXiaoWei. An improved algorithm for mining frequent set [J]. Journal of guangxi normal university (natural science edition), 2001, 12 (3) : 22-26.

[17] S. Srivastava et al, 2011 "On Performance Evaluation of Mining Algorithm for Multiple-Level Association Rules based on Scale-up Characteristics", Journal of Advances in Information Technology, VOL. 2, NO. 4.

[18] [Nuntawut et al., 2014] "A Technique to Association Rule Mining on Multiple Datasets", Journal of Advances in Information Technology, vol. 5, no. 2, may 2014.

[19] S. Kotsiantis, D. Kanellopoulos **"**Association Rules Mining: A Recent Overview**",** GESTS International Transactions on Computer Science and Engineering, Vol.32 (1), 2006, pp. 71-82

[20] P. Kandpal, *"* Association Rule Mining In Partitioned Databases: Performance Evaluation and Analysis",(Master Thesis) IIIT-Allahabad,India

[21] T. Siddiqui, M Afshar Aalam, and Sapna Jain, 2012 "Discovery of Scalable Association Rules from Large Set of Multidimensional Quantitative Datasets" journal of advances in information technology, vol. 3, no. 1

[22] R. S. Thakur *et al*., 2006 "Mining Level-Crossing Association Rules from Large Databases" Journal of Computer Science 2 (1): 76-81, 2006 ISSN 1549-3636

[23] N. Kaoungku et al, 2014 " A Technique to Association Rule Mining on Multiple Datasets" Journal of Advances in Information Technology, vol. 5, no. 2,

[24] [S. Dehuri, et al. 2006] "Multi-objective Genetic Algorithm for Association Rule Mining Using a Homogeneous Dedicated Cluster of Workstations" American Journal of Applied Sciences 3 (11): 2086-2095, 2006 ISSN 1546

Prateek kumar Singh Received the B.E degree in information Technology from R.I.T.E.E College, Raipur, India, in 2012. And pursuing M.Tech from the LNCT, Jabalpur, India. Email - Prateek8030@gmail.com

Naazish Rahim Working as an HOD & Assistant Professor, Department of Computer Science & Engineering at Lakshmi Narain College of Technology, Jabalpur, MP, India. Email- naazish.rahim786@gmail.com

Sujeet Tiwari Working as an Assistant Professor, Department of Computer Science & Engineering at Lakshmi Narain College of Technology, Jabalpur, MP, India. Email- sujeet.tiwari08@gmail.com

Neelu Sahu Working as an Assistant Professor, Department of Information Technology Government Engineering College , Bilaspur, Chhattisgarh, India
Email- neelu.sahu.12@gmail.com