Survey and a Novel Comprehensive Indicator Interestingness Measures through the Geometric Method: Association Rules Mining Perspective

[1]Niket Bhargava, [2]Dr. Savottam Dixit and [3]Dr. Manoj Shukla,
[1]Department of Comp. Sc. and Engineering, Mewar University Chittaurgarh, India
[2]Professor, Mewar University, Chittorgarh, India
[3]Professor, Bansal Group of Institutes, Bhopal, India

Abstract: In the context of Association Rule mining for finding useful rules, a novel Interestingness Measure is presented. This measure is inspired from the method of Obtaining the Comprehensive Indicator through the Geometric Methods and discussed in this paper. As a statistics based method, association rule mining has certain limitations. First of all, the generation of the association rules is totally based on the fact, data without considering the relationship between the rules. Secondly, affected by data quality and selection of threshold, the generator may produce useless rules or even lose some useful rules. Thirdly, the expression ability of association rules is limited. Thus, evaluating the reliability of the obtained association rules becomes one of the hot spots for researchers. Study on traditional association rules mining is based on support-confidence framework, and the rules are called strong association rules only when they satisfy both thresholds of support and confidence. However, sometimes strong association rules are not what users are interested in and are even misleading. Thus need is to further analyze and evaluate the mined rules in order to find the most valuable association. Generally, since most rules with high support are obvious or are already known by users, low support rules that provide users with some interesting new knowledge may be more novel than high support rules. However, with too low support threshold, it can also produce the combination explosion problem. So the best way to resolve this dilemma is to first set a low support threshold or use dynamic support threshold to complete a series of mining and then employ the new association rules measure framework to screen mining results and extract the most valuable and interesting association rules at the same time.

General Terms: *Data Mining, Big Data, Data Science, Association Rule Mining*

Keywords: *Interestingness Measures, Objective Measures, Subjective Measures, Statistics Based Measures, Comprehensive Measures.*

I. INTRODUCTION

Field of data processing is changing very fast as the volume is increasing at a faster pace and as the more intelligent and automated viewpoint for looking at data are the need of the time. This changing need is from all dimensions of life like Business, Biology Medical Research, Education, Governance, Risk Analysis, Text Analysis and Social Relation Management. These domains putting more and more difficult to manage storage, and computationally complex challenges before scientific community. To summarize or to take decisions, finding interesting and useful patterns in data is must. Association Rule Mining is the branch of data mining that is very helpful in this context.

Association rule mining plays critical and important part in knowledge mining. The difficult task in is discovering hidden knowledge i.e useful rules from the large number of rules generated for reduced support. For pruning of rules or grouping the rules, many techniques are suggested such as rule structure cover methods, informative cover methods, rule clustering, etc. Another way of selecting association rules is based on interestingness measures such as support, confidence, correlation, and so on. In this paper, we study various interestingness measures and their use in selecting interesting and useful rules.

These association rules are of utmost importance in implementing software systems that are based on methods and techniques of Data Science, Big Data and Data Mining. Because, the interestingness evaluation of association rule has the significance for the practical application of association rule mining technology, so it is necessary to study and improve it

So what's getting ubiquitous and cheap? Data. And what is complementary to data? Analysis. All modern computation streams of analytics like Data Science and Big Data Analytics along with older streams Data Mining, Knowledge Discovery, Data Processing, Business Reporting are facing the challenges putted before by the huge amount of data or information produced from various computational and non computational systems(later on digitized using some way).

Computer systems and information communication systems are producing large volume of data. This information is stored in computer database systems. Association rule mining (ARM) is the most popular knowledge discovery technique used in several areas of applications. In ARM, large number of Association rules or patterns or knowledge is generated from the large volume of dataset. But most of the association rules have redundant information and thus all of them can not be used directly for an application. So pruning or grouping rules by some means is necessary to get very important rules or knowledge. One way of selecting very interesting rules is using interestingness measures to rank and select a small set of rules of different characteristics. Another way is forming groups or clusters of rules and selecting very important rules from each cluster.

II. METHODS OF CALCULATING INTERESTINGNESS OF ASSOCIATION RULES

The problem of mining association rules was first introduced in Agrawal et al 1993. The aim of association rule mining is to find interesting and useful patterns in a transaction database. Later on the method was extended to find interesting and useful patterns in relational databases too. The database contains transactions which consist of a set of items and a transaction identifier (e.g., a market basket, due to which also sometimes referred to as Market Basket Analysis). Association rules are implications of the form X -> Y where X and Y are two disjoint subsets of all available items. X is called the antecedent or LHS (left hand side) and Y is called the consequent or RHS (right hand side). Association rules have to satisfy constraints on measures of significance and interestingness. [1]There is a strong need of measures that can tell us about the health of the Association Rule.

Data mining is an area of data analysis that has arisen in response to new data analysis challenges, such as those posed by massive data sets or non-traditional types of data. Association analysis, which seeks to find patterns that describe the relationships of attributes (variables) in a binary dataset, is an area of data mining that has created a unique set of data analysis tools and concepts that have been widely employed in business and science. Data mining receives much attention from artificial intelligence and databases, and the association rule is one of the most important research fields of data mining. In this paper, the advantages and disadvantages of the specific indicators of objective measure, subjective measure, and association rule based on statistical perspective are discussed. Some indicators of statistical perspective are adopted to measure the association rules, which can effectively solve the problems of association rules. This paper first provides a discussion of objective measures for assessing the interestingness of association patterns than cover Subjective Measures.

A. Objective Measures of Intrestingness

The objective measures used to evaluate the interestingness of association patterns are a key aspect of association analysis. Indeed, different objective measures define different association patterns with different properties and applications. Benefit of using objective measures is that they are quantitatively representable, relatively visual and are easy to operate. Objective Measures are support, confidence, lift, improve, validity, influence, Conviction and Bi-lift, Bi-improve, and Bi-confidence, for Lift, Improve, and Confidence, respectively etc.

1) Support

Support [16] means the frequency that the data fields "A" and "B" involved in association rules occur together in the data set. Only the association rules appear frequently in the itemsets, when it gets high accuracy. Support can be used to measure the usefulness of association rules. When the frequency of "A" and "B" occurring at the same time is equal to or greater than the designated minimum support threshold, "A" and "B" meet frequent itemsets. Support can be expressed as

$$s(A \longrightarrow B) = P(AB) = \frac{N(AB)}{|D|},$$

where N(AB) is the record number of "A" and "B" that appeared together, and |D| is the total record number of transactions in data sets.

Support is classic but also has the defects of artificially controlled threshold and rare itemsets. Many infrequent itemsets in the data set may have potential value. Besides, at present in large electronic commerce system, the number of subjects (users) and the amount of projects increase exponentially. Online transaction data and user evaluation data are extremely sparse.

2) Confidence

Confidence [16] is the statistics of probability P(B | A) that subsequent events occur under the condition of occurrence of the precursor events in trading data sets. It is used to measure the reliability of the rules. Formula is

$$c(A \longrightarrow B) = P(B \mid A) = \frac{P(AB)}{P(A)}.$$

It is used to combine confidence with support to form Support-confidence framework for mining association rules [17]. If Support is larger than the designated minimum support threshold and Confidence is larger than the designated minimum confidence threshold, the rules are called strong association rules. But strong association rules are not always effective, some are not what users are interested in, and some are even misleading.

3) Lift

Because of the defects of Support-confidence framework, some scholars analyze the relativity of association rules mined, namely, lift [16]. Lift means the ratio of rule's Confidence to probability of occurrence of the consequent, which reflects positive or negative correlation of antecedent and consequence of rules. It refers to the ratio of the occurrence probability of "B" under the condition "A" to that without considering condition "A" , which reflects the relationship between "A" and "B":

$$lift(A \longrightarrow B) = \frac{c(A \rightarrow B)}{P(B)} = \frac{P(AB)}{P(A)P(B)}.$$

The range of lift values is [0,+∞]. As lift is equal to 1, it shows that A and B appearing at the same time belong to independent random events and have no special significance; namely, A and B are independent of each other with no mutual affection. Call this rule uncorrelated rules; if lift value is less than 1, it shows that the emergence of "A" reduces the emergence of "B," and then we call them negative correlation rules; if Lift value is larger than 1, it shows that the emergence of "A" promotes the emergence of "B," and then we call them positive correlation rules. Problems: Lift takes events A and B in equivalence position. According to the Lift, (A -> B) and (B -> A) are the same; that is to say, if we accept rule (A -> B), (B -> A) should be also accepted, but the fact is not like this[25].

4) Improve

Literature [21] proposed a new interestingness measure method of association rules based on the description of the defects of the traditional interestingness measurement method. This is called "Improve." It means that the difference of the conditional probability P(B|A) and the probability of "B"

$$Improve(A \longrightarrow B) = [P(B \mid A) - P(B)]$$

But shortcomings of Improve (Imp.) are obvious. Firstly, how much improvement of probability can be called improvement? Secondly, the probability of former pieces' occurrence will seriously affect Improve evaluation in such a way that when it is high, the improve value will be very small all the time.

5) Validity

Literature [18] introduces a new measure method of association rules, known as validity. Validity is defined as the difference between the probability of "A" and "B" occurring together and the occurrence probability of "B" without "A" occurring in database D. Because the value range of P(AB) and P(ÀB) are [0,1] , the value range of validity is obviously [-1,1] :

$$Validity(A \longrightarrow B) = P(AB) - P\left(\overline{A}B\right)$$

6) Influence

The interestingness measure standard based on T verification is put forward, namely influence[3]. The statistics T verification method is used to analyze the difference between the association confidence P(B|A) and the expected confidence P(B). If the difference is large, it indicates that the occurrence of A has a relatively large influence on the occurrence of B. The rule (A -> B) is interesting, and formula is as shown as :

Influence(A->B) = [P(B/A) – P(B)] / σ

$$\sigma = \sqrt{\frac{P(B)(1-P(B))}{n}}$$

7) Bi-Lift

Related researches show that lift method has good evaluation results. But obviously lift take “A” and “B” in equivalence position, and it shows that rules (A -> B) and (B -> A) are the same; if we accept rule (A -> B) , rule (B -> A) should be also accepted. But the fact is not like this. For this problem, the paper proposes a Bi-lift measure method; finding that you want to evaluate the relationship of (A -> B) by lift(A -> B), you should also study on the relationship of (À -> B), so we introduce lift(À -> B) to adjust lift(A -> B). The higher the lift(A -> B) is, the better the rule (A -> B) is, while the higher the lift(À -> B) is, the worse the rule (A -> B) is. So based on correction proposed the Bi-lift measure method, as denominator, and as numerator, namely, ratio of to ; Bi-lift formula is as follows:

The premise is P((ÀB) ≠ 0, and “A” and “B” are not certain event or impossible event. Its value range is [0,∞). The Bi-lift method takes deduction of negative premise as a constraint, to form a bi-deduction comparing algorithm so as to improve the reliability of the mutual influence between premise and follow-up.

$$\textit{Bi-lift}(A \longrightarrow B) = \frac{\textit{lift}(A \rightarrow B)}{\textit{lift}(\overline{A} \rightarrow B)} = \frac{P(AB)/P(A)P(B)}{P(\overline{A}B)/P(\overline{A})P(B)} = \frac{P(AB)P(\overline{A})}{P(\overline{A}B)P(A)}.$$

8) Bi-Improve

Because of the defects of improve, the paper[25] put forward Bi-improve. Because the probability of former pieces’ occurrence will seriously affect Improve evaluation in such a way that when it is high, the improve value will be very small all the time. In order to eliminate the influence, made a correction by multiplying the ratio of the occurrence possibility of antecedent to the no occurrence probability of antecedent. Bi-improve formula is as follows:

$$\textit{Bi-improve}(A \longrightarrow B) = [P(B \mid A) - P(B)] * \frac{P(A)}{P(\overline{A})} = \frac{P(AB) - P(A)P(B)}{P(\overline{A})}.$$

9) Bi-Confidence

Confidence indicates that the appearance of some itemsets will lead to appearance of other itemsets. But the confidence of association rules only thinks about the occurrence possibility of “B” when “A” occurs, but not consider the relationship between “A” and “B” when “A” does not occur. So it makes a lot of association rules mining invalid[25]. For the above problems of association rules, the description of confidence is not perfect and not enough to show the degree of correlation between itemsets. Putting forward the concept of Bi-confidence, and its definition is as follows:

$$\textit{Bi-confidence}(A \longrightarrow B) = \frac{P(AB)}{P(A)} - \frac{P(\overline{A}B)}{P(\overline{A})} = \frac{P(AB) - P(A)P(B)}{P(A) \times [1 - P(A)]}.$$

The value range of Bi-confidence is [-1,1] . If the Bi-confidence value is greater than 0, then P(AB) > P(A)P(B), which shows that “A” and “B” have the positive correlation. If the Bi-confidence is equal to 1, then P(AB) = P(A) = P(B), and it shows that “A” and “B” in record set appear together or not. If the Bi-confidence is equal to 0, and P(AB) = P(A)P(B), which shows that “A” has no relation with “B” If the Bi-confidence is less than 0, then P(AB) < P(A)P(B); it shows that “A” and “B” have the negative correlation, and negative rules also have research value. Bi-confidence’s definition not only contains the correlation factors, but also contains factor. Therefore, Bi-confidence can fully embody the effectiveness of the rules. If we use Support-Bi-confidence framework to replace Support-confidence framework, it not only can mine association rules effectively, but also can reduce the occurrence of the weak correlation rules[25].

10) Conviction

As early as 1997, Brin introduced the concept of conviction (Conv.) [19, 20]:

$$\textit{Conviction}(A \longrightarrow B) = \frac{P(A)P(\overline{B})}{P(A\overline{B})}$$

Its value range is [0,+∞) . When the value of conviction is “1,” it means that “A” has no relation with “B” And the greater the conviction is, the higher the interest in the rule will be. But the conviction constraints are too strict; lots of valuable association rules will be removed.

B. B. Subjective Measures of Interestingness

The subjective interestingness mainly involves the knowledge field, hobbies and other personality characteristics of the main body (users). Compared with the research of objective interestingness, the research of subjective interestingness is relatively rare and immature. The objective measure once evaluated may not be the mode the users of measures are interested in, therefore, need for subjective measures is arises. The subjective evaluation indicator mainly embodies the subjective factors, such as user participation and the integration in the field of knowledge, etc. This evaluation of this level is from the perspective of rules, regardless of the data in the database.

The subjective measures are Novelty, Availability, Simplicity, Trust, Comprehensive Evaluation. And the comprehensive evaluation indicator is the measure indicator including all kinds of indicators, which is obtained through the set weighted average of the objective and subjective measure indicators. Generally, the purpose of association rule mining is to obtain certain utility or benefit through the use of some appropriate association rules. So taking the user’s subjective preference or specific application object into consideration, profit targets (or revenue function) are the real key for the users. So two new subjective measures first, for Utility Function, Incremental Monetary Value(IMV) and second, Interestingness Function Based on Profit (IFBP) including Cost function Measure[25].

1) Novelty

Novelty is a relative concept to the primal knowledge, and its extent reflects on the difference in each item between the discovered rules and the rules based on the knowledge base and the difference is respectively reflected on the difference extent in each item of the antecedent and consequent[26]. Assuming the set made up of the discovered rule is E, and the rules set in the basic knowledge base is K. The number of rules in E is |E|, and the number of rules in K is |K|. Assuming W_i is the novelty of the rule E_i in E relative to K. $W_{(i,j)}$ is the novelty between the rule E_i and rule K_j in the basic knowledge base, namely the difference degree. $W_{(i,j)}$ includes two parts: first, the novelty $L_{(i,j)}$ of the antecedent and the novelty $Z_{(i,j)}$ of the consequent. Assuming J is the set of all the antecedents among the rule K_j in the basic knowledge base, and I is the set of all antecedents among the rule E_i in the E. As for any item I_k in the I, $V_{(i,j),k}$ is the difference degree between this item and the rule K_j, we conclude :

$$V_{(i,j)k} = \begin{cases} 2, I_k \notin J \\ 1 + neg_k, I_k \in J \end{cases}$$

neg$_K$ is the difference degree of the values between the K_{th} iten in I and the same item in J, The novelty of the antecedent is equal to the accumulation for the difference degree on each item of the antecedent, namely:

$$L(i,j) = \sum_{k=1}^{|I|} V_{(i,j)k}$$

after the simplification of the rule, the items number of the consequent of all the rules in the basic knowledge base is 1, so is the item number of the consequent of the rules obtained through data mining algorithms. Therefore, the newly discovered rule E_i has only two possible relations with any rule K_j in the knowledge base:(1) the consequent of the two rules belongs to the same attribute. At this time, the difference degree(**neg**) of the corresponding value between the two rules should be calculated first, namely $Z_{(i,j)}$ = 1 + **neg**. (2)the consequent of the two rules does not belong to the same linguistic variable. At this time, make $Z_{(i,f)}$ = 2 and then calculate the sum:

$$W_i = \frac{\sum_{j=1}^{|k|} w(i,j)}{|k|}$$

Finally, the novelty of the new rules can be determined through the calculated novelty W_t . The rules with higher novelty should be kept while the rules with lower novelty should be deleted, and put the leaving rules into the knowledge base.[3]

2) Availability

The aim that clients analyze the data by data mining tools is to utilize the result of the data mining to support the decision. If the clients can improve the work flow and enhance efficiency according to a certain mode of data mining, then the mode is interesting. If the client can utilize the obtained knowledge to take some actions, and thus improve the work efficiency or bringing some economic profit, it is thought to be practical. Availability is the function of the costs needed by the researched system transformed into the associated space(interested state) M_i from the current space(primal state)M_0. Namely:

$$S = \frac{1}{f(M_0 - M_i)}$$

The more it costs to transform, the smaller the availability of the mode has. When the system cannot be transformed into the associated space from the current space, it is s -> 0.

3) Comprehensive Evaluation

In addition, many references also mention the simplicity, trust, comprehensive evaluation, etc.. And the comprehensive evaluation indicator is the measure indicator including all kinds of indicators, which is obtained through the set weighted average of the objective and subjective measure indicators.

4) Obtaining the Comprehensive Measure through the Principal Component Analysis

The fist principal component is obtained through the principal component analysis, namely it is regarded as a new comprehensive measure. The computational formula of the principal component is as shown as follows:

$$F_1 = a_{11}x_1 + a_{21}x_2 + ... + a_{p1}x_p$$

Through the principal component analysis, we can regard the indicator associated with the association rule as the variable X, and not consider the redundancy among the variables[3].

5) Obtaining the Comprehensive Indicator through the Geometric Method:

$$RI = s^{w1} * c^{w2} * lift^{w3} * wi^{w4} * s1^{ws},$$

In this formula, s is support, c is confidence, lift is lift, wi is novelty, and s1 is availability. There are several advantages using the comprehensive indicator RI: firstly, support, confidence and lift respectively represents practicality, credibility, correlation, which can comprehensively reflect all aspects of association with little redundancy; secondly, this comprehensive indicator can not only reflect the objectiveness, but also users' objective interestingness; thirdly, through the weight, this comprehensive indicator can embody the importance of indicator and it is not affected by the inconsistent dimension.

6) Utility Function Incremental Monetary Value(IMV)

The purpose of association rule mining is to translate it into real value and to obtain certain utility. Generally utility function is the concern point of users. Specifically, utility of rules can be incremental monetary value generated by association rule [24]. Incremental monetary value (IMV) is the expected profit (EP) under the guidance of rules minus the profit you would expect to receive without the guidance of rules or due to the natural course[25]. Incremental monetary value is defined as follows in that study:

$$IMV(A \longrightarrow B) = [P(B \mid A) - P(B)] \times \sum \text{Price}(B_i),$$

where Price(B_i) is the unit price of goods B_i and $\sum$ Price(B_i) is the sum price of all the “B” sets.

Incremental monetary value (IMV) has some defects, the probability of former pieces’ occurrence will seriously affect Improve evaluation (Imp.(A -> B = P(B | A) – P(B)), and, in order to eliminate the influence, we make correction by multiplying by the ratio of the occurrence possibility of antecedent to the no occurrence probability of antecedent. In addition, utility or value should be price minus cost. So we put forward antecedent incremental monetary value (AIMV):

$$AIMV(A \longrightarrow B) = \left\{[P(B \mid A) - P(B)] \times \frac{P(A)}{P(\overline{A})}\right\} \cdot \sum [\mathrm{Price}(B_i) - \mathrm{Cost}(B_i)],$$

where Price(B_i) is the unit price of goods B_i and Cost(B_i) is the cost per unit of goods B_i.

7) Cost Function

The cost or the cost of labor should be concerned when executing the association rules. Let Cost(A -> B) represent the executing cost of rules of (A->B) , such as the unit cost of handling commodity "A" from place A to place B. But the calculation of the so-called executing cost is complicated sometimes. In order to make the study under the maneuverability, the executing cost of rules of (A -> B) can be divided into several levels.

8) Intrestingness Function Based on Profit(IFBP)

On the basis of utility function and the executing cost of rules Cost(A -> B), recently is proposed interestingness function based on profit (IFBP) with subjective preferences and specific application object[25]. The formula for this measure is as follows:

$$IFBP(A \longrightarrow B) = \left\{[P(B \mid A) - P(B)] \times \frac{P(A)}{P(\overline{A})}\right\} \cdot \sum [\mathrm{Price}(B_i) - \mathrm{Cost}(B_i)] - \mathrm{Cost}(A \longrightarrow B).$$

C. Interestingness Measures based on Statistical Perspective

Not discussed in detail in this paper, the measures based on statistical perspective are Contingency Table, Canonical Correlation Analysis, Trust, Comprehensive Measure using Principal Component Analysis, Comprehensive Measure using Geometric Method, Chi-square etc.

III. OUR COMPREHENSIVE EVALUATION INDICATOR

The comprehensive evaluation indicator is the measure indicator including all kinds of indicators, which is obtained through the set weighted average of the objective and subjective measure indicators.

A. Obtaining the Comprehensive Measure through the Principal Component Analysis

The fist principal component is obtained through the principal component analysis, namely it is regarded as a new comprehensive measure. The computational formula of the principal component is as shown as follows:

$$F_1 = a_{11}x_1 + a_{21}x_2 + \ldots + a_{p1}x_p$$

Through the principal component analysis, we can regard the indicator associated with the association rule as the variable X, and not consider the redundancy among the variables[3].

B. Obtaining the Comprehensive Indicator through the Geometric Method

$$RI = s^{w1} * c^{w2} * lift^{w3} * wi^{w4} * s1^{ws},$$

in this formula, s is support, c is confidence, lift is lift, wi is novelty, and s1 is availability. There are several advantages using the comprehensives indicator RI: firstly, support, confidence and lift respectively represents practicality, credibility, correlation, which can comprehensively reflect all aspects of association with little redundancy; secondly, this comprehensive indicator can not only reflect the objectiveness, but also users' objective interestingness; thirdly, through the weight, this comprehensive indicator can embody the importance of indicator and it is not affected by the inconsistent dimension.

C. Our Method for Obtaining the Comprehensive Indicator through the Geometric Method

The component step of our methods are as following:

1. we will use Support-Bi-confidence framework to replace Support-confidence framework, it not only can mine association rules effectively, but also can reduce the occurrence of the weak correlation rules.

2. we will use Bi_lift measure as, the Bi-lift method takes deduction of negative premise as a constraint, to form a bi-deduction comparing algorithm so as to improve the reliability of the mutual influence between premise and follow-up.

3. we will use novelty to evaluate difference among newly discovered rules and what we already have in knowledge base. Large difference represent existence of new knowledge's strong presence which is not id knowledge base.

4. we will use availability, as availability represents capability to transform the mining methodology into real life production system. If user can afford the cost system is usable to user.

5. we will use weights to better capture degree or level of involved measures

6. we will choose both (1) Multiplicative Method and, (2) Additive Method for combining the effect of numerous measures of various classes, here two classes, objective and subjective.

BiCLmult measure is calculated as follows :

$$\mathbf{RI\_Multiplcative = s^{w1} * Bi\_c^{w2} * Bi\_lift^{w3} * wi^{w4} * s1^{ws},}$$

BiCLadd measure is calculated as follows:

$$\mathbf{RI\_Additive = s^{w1} + Bi\_c^{w2} + Bi\_lift^{w3} + wi^{w4} + s1^{ws},}$$

in this formula, s is support, Bi_c is improved confidence, Bi_lift is improved lift, wi is novelty, and s1 is availability. There are several advantages using the comprehensive indicator RI: firstly, improved support, improved confidence and improved lift respectively represents practicality, better credibility, better correlation interpretation, which can comprehensively reflect all aspects of association with little redundancy; secondly, this comprehensive indicator can not only reflect the objectiveness, but also users' objective interestingness; thirdly, through the weight, this comprehensive indicator can embody the importance of indicator and it is not affected by the inconsistent dimension.

D. Should I use an additive model or a multiplicative model?

Choose the multiplicative model when the magnitude of the patterns in the data depends on the magnitude of the data. In other words, the magnitude of the pattern increases as the data values increase, and decreases as the data values decrease. Choose the additive model when the magnitude of the pattern in the data does not depend on the magnitude of the data. In other words, the magnitude of the pattern does not change as the data volume goes up or down. If the pattern in the data is not very obvious, and you have trouble choosing between the additive and multiplicative procedures, you can try both and choose the one with smaller accuracy measures.

CONCLUSION

Generally, there are two evaluation standards to evaluate whether an association rule is interesting or not: the objective measure and the subjective measure. The method of the objective measure can obtain a quantitative value by the algorithm, and it is relatively visual and easy to operate. However, the rule after the evaluation of the objective measure may be not be the mode users interested in, therefore, the subjective measure is required. In order to ensure that the final mining rule can arouse the interests of the users or the experts in the field, they should be involved in the process and make use of their knowledge to pruning the rule.

References

[1] Michael Hahsler, http://michael.hahsler.net/research/association_rules/

[2] Michael Hahsler, http://michael.hahsler.net/research/association_rules/measures.html last modified 11/16/2015 22:27:02

[3] Gao Yongmei and Bao Fuguang, “The Research on Measure Method of Association Rules Mining,” International Journal of Database Theory and Application, Vol. 8, No.2(2015), pp. 245-258

[4] Liqiang Geng and Howard J. Hamilton. “Interestingness measures for data mining: A survey”. ACM Computing Surveys, 38(3):9, 2006.

[5] P. N. Tan, M. Steinbach, and V. Kumar, “Introduction to Data Mining,” Pearson—Addison-Wesley, Boston, Mass, USA, 1st edition, 2005.

[6] J. H. Fowler and M. Laver, “A tournament of party decision rules,” Journal of Conflict Resolution, vol. 52, no. 1, pp. 68–92, 2008. View at Publisher · View at Google Scholar · View at Scopus

[7] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules,” in Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94), pp. 487–499, 1994.

[8] R. Srikant and R. Agrawal, “Mining generalized association rules,” in Proceedings of the 21st International Conference on Very Large Data Bases, pp. 407–418, Zurich, Switzerland, 1995.

[9] M.-C. Tseng, W.-Y. Lin, and R. Jeng, “Incremental maintenance of generalized association rules under taxonomy evolution,” Journal of Information Science, vol. 34, no. 2, pp. 174–195, 2008. View at Publisher · View at Google Scholar · View at Scopus

[10] M.-C. Tseng and W.-Y. Lin, “Maintenance of generalized association rules with multiple minimum supports,” Intelligent Data Analysis, vol. 8, no. 4, pp. 417–436, 2004. View at Google Scholar · View at Scopus

[11] M.-C. Chen, “Ranking discovered rules from data mining with multiple criteria by data envelopment analysis,” Expert Systems with Applications, vol. 33, no. 4, pp. 1110–1116, 2007. View at Publisher · View at Google Scholar · View at Scopus

[12] M. Toloo, B. Sohrabi, and S. Nalchigar, “A new method for ranking discovered rules from data mining by DEA,” Expert Systems with Applications, vol. 36, no. 4, pp. 8503–8508, 2009. View at Publisher · View at Google Scholar · View at Scopus

[13] L. Geng and H. J. Hamilton, “Interestingness measures for data mining: a survey,” ACM Computing Surveys, vol. 38, no. 3, pp. 1–32, 2006. View at Publisher · View at Google Scholar · View at Scopus

[14] N. Hoque, B. Nath, and D. K. Bhattacharyya, “A new approach on rare association rule mining,” International Journal of Computer Applications, vol. 53, no. 3, pp. 1–6, 2012. View at Publisher · View at Google Scholar

[15] J. Zhang, Y. Wang, and J. Feng, “Attribute index and uniform design based multiobjective association rule mining with evolutionary algorithm,” The Scientific World Journal, vol. 2013, Article ID 259347, 16 pages, 2013. View at Publisher · View at Google Scholar · View at Scopus

[16] J. Han and M. Kamber, Data Mining Concept and Technology, China Machine Press, Beijing, China, 2007.

[17] G. Piatetsky-Shapiro, “Discovery, analysis, and presentation of strong rules,” in Knowledge Discovery in Databases, G. Piatetsky-Shapiro and W. Frawley, Eds., pp. 229–238, AAAI/MIT Press, 1991. View at Google Scholar

[18] L. Ma and W. Jie, “Research on judgment criterion of association rules,” Control and Decision, vol. 18, no. 3, pp. 277–280, 2003. View at Google Scholar · View at Scopus

[19] S. Brin, R. Motwani, J. D. Ullman, and S. Tsur, “Dynamic itemset counting and implication rules for market basket data,”in Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '97), pp. 255–264, ACM, Tucson, Ariz, USA, May 1997. View at Publisher · View at Google Scholar

[20] P.-N. Tan, V. Kumar, and J. Srivastava, “Selecting the right objective measure for association analysis,” Information Systems, vol. 29, no. 4, pp. 293–313, 2004. View at Publisher · View at Google Scholar · View at Scopus

[21] Y. Li, C. Wu, and K. Wang, “A new interestingness measures for Ming association rules,” Journal of the China Society for Scientific and Technical Information, vol. 30, no. 5, pp. 503–507, 2011. View at Google Scholar

[22] J. Chen and Y. Gao, “Evaluation criterion for association rules with influence degree,” Computer Engineering and Application, vol. 45, no. 8, pp. 141–143, 2009. View at Google Scholar

[23] F. Berzal, I. Blanco, D. Sanchez, and M.-A. Vila, “Measuring the accuracy and interest of association rules: a new framework,” Intelligent Data Analysis, vol. 6, no. 3, pp. 221–235, 2002. View at Google Scholar · View at Scopus

[24] D. H. Choi, B. S. Ahn, and S. H. Kim, “Prioritization of association rules in data mining: multiple criteria decision approach,” Expert Systems with Applications, vol. 29, no. 4, pp. 867–878, 2005. View at Publisher · View at Google Scholar · View at Scopus

[25] Chunhua Ju, Fuguang Bao, Chonghuan Xu, and Xiaokang Fu, “A Novel Method of Interestingness Measures for Association Rules Mining Based on Profit ,” Discrete Dynamics in Nature and Society, Volume 2015 (2015), Article ID 868634, Received 8 April 2015; Revised 17 June 2015; Accepted 18 June 2015

[26] Q. Yanxie, “Measurement of Novelty: Factor of Evaluation for the Association Rules”, Application Research of Computers,(2004), no. 1, pp. 7-19

Authors Profile

Niket Bhargava, born in state is MP in the India, on July 6, 1978. He graduated from the Oriental Institute of Science And Technology, Bhopal, and after qualifying in GATE, completed his Master in Technology from the RGTU- the technical university of MP, India. His employment experience included the 10 years as Teacher in Top Most Ranking Institutions of state of MP ,and 03 years of Industry experience of BI product development. Presently, he is pursuing his PhD in CSE with focus on business applications of data mining , data science, big data techniques for Business Intelligence.

Prof.(Dr.) Sarvottam Dixit, is working as professor in Mewar University, INDIA. He is having more than 20 years of experience in teaching and hold positions like director, pro vice chancellor, and others in various institutes in India.

Prof.(**Dr.)** M**ANOJ SHUKLA**, is working as professor in BCE, Mandiddeep. He is having more than 15 years of experience in teaching and hold positions like principal, and others in various institutes in Madhya Pradesh, India.