Research on the Supporting Degree of Multilayer Correlative Rules and Its Minimal Thresholds

CAO Xiaolan

College of Information Science and Technology
Hunan Agricultural University
Changsha, Hunan

PENG Jian

College of Information Science and Technology
Hunan Agricultural University
Changsha, Hunan

PENG Jiahong[*]

College of Information Science and Technology
Hunan Agricultural University
Changsha, Hunan
pjh719@163.com

Abstract—**Supporting degree and its minimal threshold have important impacts on the designment of quick and effective algorithm of multilayer correlative rules. This paper introduces interzone supporting degree for computing supporting degree of item sets of intersectional layer rationally which avoids unmeaningful intersectional layer rules. Further, this paper comes up with a method of adjusting minimal supporting degree thresholds of different layers by the threshold of each abstract layer from experts and the recording of user's feedback. This method overcomes the subjective impact of fixed threshold on the extraction of interesting correlative mode in some degree, making the correlative rules promoted to the user are more closer to the user's interest, shows the character of personalized promotion.**

Keywords-component; data mining; multilayer correlative rule; supporting degree

I. INTRODUCTION

Mining of correlative rules can help to find interesting corealtion between item sets of data, it's an important researching field of data mining technology and its using in the various methods of data mining is the most extensive[1-4]. Because of the decentralization of data distribution in many applications, it's hard to find strong correlative rules in the conception of initial layer or the lowest layer[5]. However, it's more easier to search interesting correlations in the items of multi concept layers than in the initial data, the strong correlative rules found in the higher concept layers maybe provide the knowledge of more universal meaning for a user, but for another, it maybe not. So, mining of multilayer correlative rules is necessary, it basically adopts the framework of supporting degree and believing degree. The spporting degree of correaltive rules describes the usability of correlative rules mined. When come to develop quick and effective multilayer correlative rules algorithm, one consideration is how to keep the consistence between initial database and transaction databasee while computing the supporting degree of item sets of transaction database. On the other hand, the threshold of minimal supporting degree has strong impact on the generation of multilayer correlative rules, for different layer, use the same threshold of minimal supporting degree or multi thresholds constraint of minimal supporting degree to mine frequent item sets? So, it's necessary to study supporting degree of item sets in multilayer transaction database and its minimal threshold.

II. SUPPORTING DEGREE OF ITEM SETS IN MULTILAYER TRANSACTION DATABASE

When come to compute supporting degree of item sets in multilayer transaction database, if ordinary composite method is adopted, the computation of supporting degree of intersectional layer will cause the explosion of composition and huge cost.

We introduce the interzone supporting degree, for some item set, computing its supporting degree according to its layer separately, for item set including intersectional layer, using the interzone supporting degree to express its supporting degree.

Using multidimensional vector$[Xi,Xi+1,\ldots,Xn]$ ($i,n \in [1,h]$,h is the height of tree) to express item set X of different layers, its minimal child is $Xi=[xi1,xi2,\ldots,xik]$, the oldest ancestor is $Xn=[xn1,xn2,\ldots,xns]$，here, $xij(j \in [1,k])$ is the jth item of ith layer of item x, $xnj(j \in [1,s])$ is the jth item of nth layer of item x, k and s express the maximal number of item in this layer, n is the highest layer of item set X.

K-item set of layer m is frequent item set, when and only when the supporting degree of subset Y of any X, $Y_{sup} \geq$ minsup.From the concept of hierachy tree, $Xi \subset X, and, X \subset Xn$, so, $Xn_{sup} \geq X_{sup} \geq Xi_{sup}$, therefore, the supporting degree X_{sup} of item X including different layers lies in the zone $[Xn_{sup},Xi_{sup}]$, that is, $Xn_{sup} \geq X_{sup} \geq Xi_{sup}$. For the given minimal supporting degree, minsup, put item set X into the frequent item set L with the probability p, here, $p = \dfrac{minsup - Xi_{sup}}{Xn_{sup} - Xi_{sup}}$, the introduction of interzone supporting degree not only computes the supporting degree of item sets of intersectional layers rationally, but also avoids a lot of unmeaningful intersectional layer rules, and keep the consistence with the original database.

III. THRESHOLDS OF MINIMAL SUPPORTING DEGREE OF DIFFERENT LAYERS

For the mining of multilayer correlativve rules, the configuration of supporting degree is an important issue.

In the process of getting multilayer correlative rules, the use of unified minimal thresholds can simplify searching

process.Because user only need to set a minimal supporting threshold, so the entire mining method is simple. But it is impossible for low layer items to present more times than corresponding high layer items, so, there are probably some problems while extracting non-interesting mode of high abstract layer or unfound interesting mode of low abstract layer. If the value of minimal supporting threshold is set too high, meaningful correlative rules presenting in the lower abstract layer maybe lost. On the contrary, unmeaningful rules can present in the hgher abstract layer. Therefore, various minimal supporting degree thresholds are adopted for mulitlayer correlative rules, each abstract layer has its minimal supporting degree threshold. The more lower abstract layer, the corresponding threshold smaller, The more higher abstract layer, the corresponding threshold bigger[6]. So, there is the problem of selecting minimal thresholds for different layers.

Generally, if a user clicks some piece of promoted rule, we can think he affirms this promotion and can evaluate and adjust the minimal supporting degree thresholds of different layers according to this.

First, experts give the minimal supporting threshold of each abstract layer. For the minimal threshold of intersectional layer which acrosses various layers, the impact of single abstract layer on it is very small, so, unified threshold is adopted. Then, adjust threshold according to the user's feedback, that is, recording the promoted rules clicked by user, totaling the number of each layer's rules clicked by user, computing the supporting degree of each layer, adjusting the minimal supporting degree threshold of each layer according its change.

Assuming si is the minimal supporting degree threshold of the ith layer, ni is the number of the ith layer's rules clicked by user, $n_i^{'}$ is last statistic number, $\sum n$ is total number of promoted rules clicked by user, so, the supporting degree ti is $\dfrac{n_i}{\sum n}$. $t_i^{'}$ is last statistic supporting degree, so, the minimal supporting threshold is adjusted as:

$$s_i = s_i + \Delta t_i \times k$$

Here, $\Delta t_i = t_i - t_i^{'}$, represents the variation compared to the last statistic, k is the adjusting value, making the variation of is under an acceptable scope of (0,1), k is some constant between 0 and 1.

If $\Delta t_i < 0$，it shows that user's recognition of promoted rules decreases, so, reducing the value of si , promoting more rules for user.

If $\Delta t_i > 0$，it shows that user's recognition of promoted rules increases, so, enlarging the value of si , on the basis of last promotion, promoting more higher requiring rules for user.

If $\Delta t_i = 0$，it shows that user's recognition of promoted rules doesn't change, there is no need to change the si.

From above, threshold can change with user's interest, the setting of threshold is more objective and reasonable.

IV. CONCLUSION

One of the extreme important functions of data mining is to provide dynamic and personalized promotion for user according to their visit to websites. This paper introduces interzone supporting degree for computing supporting degree of item sets of intersectional layer rationally, avoids unmeaningful intersectional layer rules. The adjustment of minimal supporting degree thresholds of different layers according to the user's feedback overcomes the subjective impact of fixed threshold on the extraction of interesting correlative mode in some degree, making thresholds more objective and reasonable, shows the dynamic and personalized promotion. All of above is necessary complement to the mining algorithm of multilayer correlative rules, it provides a good basis for the the designment of quick and effective algorithm of multilayer correlative rules.

REFERENCES

[1] Han J W, Pei J, Yin Y W. Mining frequent patterns without candidate generation[C]. In: Chen W D, Naughton J, Bernstein P A. 2000 ACM SIGMOD Intl. Conference on Management of Data. USA: ACM Press, 2000, 1~12.

[2] Yugami N., Ohto Y., Okamoto S.. Fast discovery of interesting rules[C]. Proceeding of the 4thPacific –Asia Conference, PAKDD2000.Japan, 2000.17-27.

[3] YANG Hong-ju, LIANG Ji-ye. An Effective Association Rules Mining Method[J]. Computer Applications.2004.24(3):88-89.

[4] Park J. S., Chen M. S., and Yu P. S.. An effective hash-based algorithm for mining association rules[C]. Proceedings of ACM SIGMOD International Conference on Management of Data, San Jose, CA, May 1995,175-186.

[5] Liu J, Pan Y, Wang K et al. Mining Frequent Item Sets by Opportunistic Projection[C]. In: Zaïane O. Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. USA: ACM Press, 2002：229－238.

[6] Mobasher B, Dai H, Luo T et al. Effective personalization based on association rule discovery from Web usage data[D]. Technical Report 01-010, Department of Computer Science, DePaul University.