I J C T A, 9(19) 2016, pp. 9091-9099

Discovering Interesting Weighted Temporal Relationship Rules

Pankaj Gupta* **and Bharat Bhushan Sagar***

ABSTRACT

Mining time-varying association-rules is a significant task in the meadow of data-mining owing to its extensive purposes. Weighted-Relationship Rule mining in recent times have been anticipated, where transactions are involved with biased values on either some pre-assigned user criterion or on basis of calculation of item-involvement in transactions, more-over in conjunction to this utilizing temporal relationship rule mining one may determine unusual relationship rules regarding different time-intervals. Some relationship rules may hold, through some intervals while not others and this may lead to subsidiary information.

Here we discussed a novel algorithm for discovering weighted time varying relationship rules of high user interest, where weighted temporal relationship rule mining allocates weights to time-varying item's based on qualitative aspects of attributes (e.g. significance, whereas the interestingness of an item can be computed as potentially useful for finding rules.

This script propose a narrative method to find relationship rule on time-variant-weighted data along with its interestingness, which give us a significant benefit in expressions of time and memory recollection utilization though including time, weight and interesting factor.

Our comparison analysis illustrates that projected weighted-relationship rule mining of interest do better than the presented algorithm in terms of efficiency, time and valued rules.

Keywords: Data-Mining-Task, Temporal-Relationship-Rule-Mining, Interestingness, Temporal-Data-Mining, Weight-carrying-Transaction, Weight-carry-Relationship-Rule Mining, Interesting-weighted time-varying Relationship Rules.

I. INTRODUCTION

Data pulling out is a technique intended for pointing the hidden information and unspoken rules in huge volume based data. Relationship-Rules mining are connection between the data values that are recorded over the period of time and finally discover some hidden & interesting information from huge history transaction database. Apriori's Algorithm which is deliberate by Srikant and Agrawal is the well established recurrent item sets mining algorithm [1]. But in modern time, a lot of advancement of Apriori's Algorithm has been observed. Though this advancement in algorithms reduces the total-number of candidate item-sets but still few issues which can be move up in relation to data-mining like–

- Relationship among different item-sets of a definite kind during a categorical time gap.

Or

- Relationship among different item-sets of a definite kind with a categorical periodicity.

In addition to this, if we associate weight i.e. utility or worth of an item-set during its life periodicity and finally look this, in-terms of interestingness i.e. how much this surprise the user or concern, in above process of mining then what will happen?

* Department of Computer Science & Engg, BIT, Mesra: Ranchi, off-Campus: Noida, *E-mails: pgupta@bitmesra.ac.in; drbbsagar@gmail.com*

This designate that data in large volume based dataset is not independent of its time-span, its worthiness' weight and its hidden appearance i.e. interestingness, instead these factors are the principle dimensions which can play an important role to generate reduced-surprising relationship-rule in time-varying authentic dataset.

By means of expansion of data-mining practices and tools, a number of algorithms have been framed to treat with the accepted & computationally exclusive job of relationship-rule mining, also a lot of effort in recent times focused on the finding the reduced worthy and valuable information from large volume-based datasets.

Our anticipated approach is to propose a well-organized procedure for mining Relationship-Ruling in time-varying-weighted dataset which is finally optimized by its interestingness.

Rest in this article is as per the following. Related-work on at relationship-rule mining, time-varying information mining, weighted rule-mining in databases is given in Part 2. Part 3 verbalize about proposed approach and algorithm. Part 5 discuss about performance studies and comparison of proposed algorithm lastly, in Part 5 we finish up and conclude the work.

II. CO-RELATED WORKS

The perception of involvement & relationship ruling was given as Apriori's algorithm [1]. Later on it trailed by upgrades [2], [3], [4], and [5]. Algorithm performance was amended by deploy recurrent pattern magnification approach [6]. In manuscript [4] the oversight of the time dimension in relationship ruling was mention. A temporal feature of relationship-ruling was proposed by [12]. As per this transactions which belong to records be time imprinted and has time span period.

As per Roddick [20] sense, the system foremost looks for the relationship than it is utilized to add in temporal feature, the outcome prove that temporal Apriori's is 4 to 21 times more quick than straight Apriori's, and the finishing time prodigiously reduces with reverence to exact match. Then the objective is to find out all the item-sets whose weight i.e. worthiness' values are beyond a user particular threshold in a time-varying transaction database. The input of weight can efficiently make out the temporal high usefulness item-sets with fewer candidates' item-sets which are finally optimized on the basis of their interestingness of higher use. Related to this present article work, Temporal Relationship-Ruling, Weighted Temporal Relationship-Ruling and Interestingness of use factor are discussed in sub-sections respectively.

2.1. Temporal Relationship Rule Mining [4]

Description: The occurrence of an item-set in excess of a time-period 'T' is the total figure of transactions in which it arise separated by whole figure of transaction in excess of that time-period. Within the identical manner, confidence of an item's set with other item's set is the transaction of together these item's over the time-span separated by initial item' set of that time-span.

$$Temporal\_Support\,(Item-Set\,'X\,') = \frac{Occurance\,of\;\;X\;in\;particular\;time\;gap}{Total\;nos.of\;\;transactions\,in\,that\,time\;gap}$$

Temporal_Confidence (X => Y[T_began,T_finish]) = Support_count(X U Y) over time gap / occurrence of X in time gap.

Where, T_began points out the valid begin time and T_finish point out valid finish time according to temporal data.

2.2. Weighted-Temporal Relationship Rule Mining [11]

Ke Sun & F Bai [9] projected measurement of weighted-support that does not require pre-assigned weights. Our anticipated way for figure-out weighted-support based on this approach, we are computing temporal-weighted-support and temporal-weighted-confidence of items as follows:

$$Temporal - weighted\ support = \frac{\Sigma i[T\_began.T\_finish]\Sigma i(qij)}{\Sigma N[T\_began, T\_finish]}$$

Where, i = 1,2,3,….n and j = 1,2,3,....k and qji stand for a number of an item i ∈ I, in a jth transaction within its valid time-period and ΣN is the sum of all transactions within its valid time-period of all items.

Temporal-weighted Confidence (X => Y[T_began,T_finish]) = Temporal-weighted support(X U Y) over time gap / occurrence of X in time gap.

Here, the primary objective is to keep the closure property in case of weighted-items too and if weights of each entity item are similar than weighted-support should also be similar. So that recurrent item-sets generates through Apriori's algorithm will be more correct, [13] precise and the efficient.

2.3. Interestingness-Correlation Analysis

Several researchers together with S Brin, et al., [17] had appreciated the restriction of the confidence-support structure. Make use of this structure only is pretty not viable to learn the interesting Relationship-Rules. Consequently, the correlation evaluation can be used as an additional assess together with this structure. One of the simplest correlation evaluations is Li-ft.

The lift for a relationship- rule[7], can be expressed as a ratio of the confidence for a relationship-rule X→Y, confidence (X→Y) to the number of transactions in Dataset contains X, . Thus

$$\text{li-ft}(X, Y) = \frac{\text{confidence } (X \rightarrow Y)}{\text{Support}(Y)}$$

More-over,if li-ft(X, Y)=1 then X & Y are independent.

If li-ft(X, Y)>1 then X & Y positively correlated.

If li-ft(X, Y)<1 then X & Y negatively correlated.

Interestingness by means of correlation quantifies how much the rule is of use for the user.

III. PROPOSED APPROACH AND ALGORITHM

In this part, the projected approach and algorithm will be described in details. The main purpose of the projected algorithm is reduction of time-varying dataset by means of weight concept along with optimization of relationship-rules obtained using interestingness feature.

Firstly, weight/utility factor is applied on the time-varying dataset for dropping the number of attributes and unwanted information from the transactional-dataset. After that, Apriori's Algorithm applicable on the compressed transactional-dataset (weighted) for obtaining the recurrent item-sets and preferred relationship-rules as per required weighted-temporal-support and weighted-temporal-confidence criteria. At last, further optimize the result using interestingness of item for surprising (hidden) rules.

Projected Algorithm

In-put: Time-varying transactional-dataset, minimum weighted-support, minimum weighted-confidence, Interestingness using lift feature.

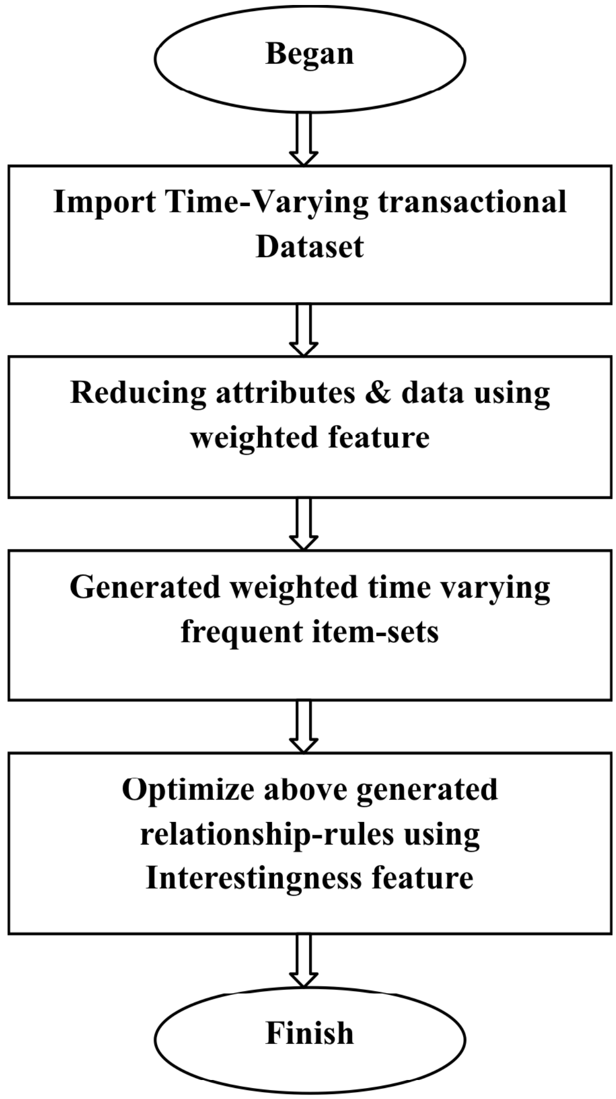

Figure1: Proposed Algorithm Block-Diagram

Out-put: Optimized weighted temporal relationship-rules.

1. Input time-varying transactional dataset.
2. Transform time-varying transactional dataset into assessment table to calculate weight for each item on the basis of above stated principle.
3. Reduce the list of attributes as per required weighted support within their legal time-period.
4. Apply Apriori's algorithm with specified weighted-support and weighted-confidence within their legal time-period and generate frequent item sets.
5. Obtained relationship-rules.
6. Calculate the interestingness for above generated rules using lift as correlation measure and further pruning on this basis.
7. Get the Optimize reduced relationship-rule of interest.

Step 1: Read time-varying transactional dataset.

Example: In a transactional dataset as shown in Table 1, a set of time-variant database indicates the transaction records from January 2015 to March 2015.

Table 1
Transactional Database

Valid Time Legacy	*Transaction id*	*Transactions*
January, 15	T1	B, D
	T2	B, C, D
	T3	B, C
	T4	A, D
February, 15	T5	B, C, E
	T6	D, E
	T7	A, B, C
	T8	C, D, E
March, 15	T9	B, C, E, F
	T10	B, F
	T11	**A, D**

Step 2: Transform transactional-dataset into assessment table to calculate the weight of each item within its valid time-period.

Table 2
Weight Assessment Table

Valid Time Period of Item	*Items*	*T-id*	*Weight Assessment*
January, 15	A	T4	1
	B	T1, T2, T3	3
	C	T2, T3	2
	D	T1, T2, T4	3
February, 15	A	T7	1
	B	T5, T7	2
	C	T5, T7, T8	3
	D	T6, T8	2
	E	T5, T6, T8	3
March, 15	A	T11	1
	B	T9, T10	2
	C	T9, T10	2
	D	T11	1
	E	T9	1
	F	**T9, T10**	**2**

Step 3: Calculation of time-varying weighted support and confidence for items in transactional dataset.

For Time-Period January, 15

Weighted Support(A)=1/9=0.11 Weighted Support(B)=3/9=0.33

Weighted Support(C)=2/9=0.22 Weighted Support(D)=3/9=0.33

For Time-Period February, 15

Weighted Support(A)=1/11=0.09 Weighted Support(B)=2/11=0.18

Weighted Support(C)=3/11=0.27 Weighted Support(D)=2/11=0.18

Weighted Support(E)=3/11=0.27

For Time-Period February, 15

Weighted Support(A)=1/9=0.11 Weighted Support(B)=2/9=0.22

Weighted Support(C)=2/9=0.22 Weighted Support(D)=1/9=0.11

Weighted Support(E)=1/9=0.11 Weighted Support(F)=2/9=0.22

Step 4: Assume minimum_weighted_support = 22%.

Step 5: If the number of items in the list of attributes < minimum support, then remove this list of attributes [15].

Table 3
Modified Table

Valid Time Period of Item	*Items*	*T-id*
January’ 15	B	T1, T2, T3
	C	T2, T3
	D	T1, T2, T4
February’ 15	C	T5, T7, T8
	E	T5, T6, T8
March’ 15	B	T9, T10
	C	T9, T10
	F	**T9, T10**

Step6: Apply Apriori’s Algorithm.

Let Minimum Weighted_Support = 22%, Minimum Weighted_Confidence= 60%.

Generated frequent time-varying item sets with Weighted-Support:

Table 4
Frequent Item sets Table for January’15

Valid Time Period of Item	*Frequent Items*	*Weight Support*
January, 15	B	37.5%
	C	25.0%
	D	37.5%
	BC	25.0%
	BD	25.0%
	CD	12.5%
	BCD	**12.5%**

So, only two frequent item sets (BC & BD) have min_wt._support within its valid time period. Thus the possible rules are:

$$B \longrightarrow C, \quad C \longrightarrow B$$

$$B \longrightarrow D, \quad D \longrightarrow B$$

Now, weighted confidence of above generated rules within their valid time-period is as follows:

For $B \longrightarrow C$ confidence is 66.6%, $C \longrightarrow B$ confidence is 100.0%,

For $B \longrightarrow D$, confidence is 66.6%, $D \longrightarrow B$ confidence is 66.6%, Hence they all qualify.

Step 7 : Find the interestingness using lift for the above generated relationship-rules and optimize the result [16].

Interestingness (B→C)= 66.6/25.5=2.66 Interestingness (C→B)= 100/37.5=2.66

Interestingness (B→D)= 66.6/37.5=1.77 Interestingness (D→B)= 66.6/37.5=1.77

Thus, all of them are interesting rules with user perspective within their valid time-span where the rules (B→C) & (B→C) are having higher interest in comparison to (B→D) & (D→B).

IV. RELATIVE PERFORMANCE STUDIES

To assess the projected algorithm we carry out a number of trials by means of artificial data. This data enclose 10,344 transactions acquired in a time period of three months in a year with different support interval threshold values.

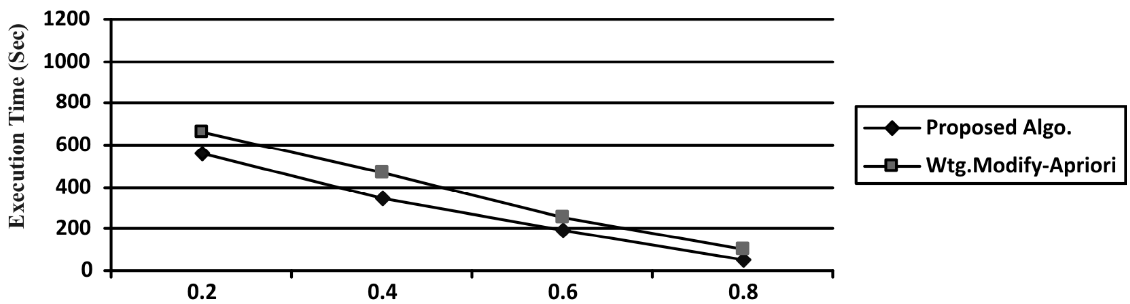

Figure 2: Comparison for Scalability

We as well evaluate the amount of relationship-rules produced by our algorithm against Weighted-Modified-Apriori, as can be seen in Fig. 3 which shows less no. pattern generated as compare to existing and thus reduces the of memory-intake for execution and produce relevant reduct rules.

We also estimate the increase performance of the projected algorithm. In this test we used a transactional datasets with support threshold value from 0.2% to 0.8%. While seeing in Fig.2, the finishing time for volume of the data with our proposed algorithm shows good scalability over weighted-modified-Apriori.

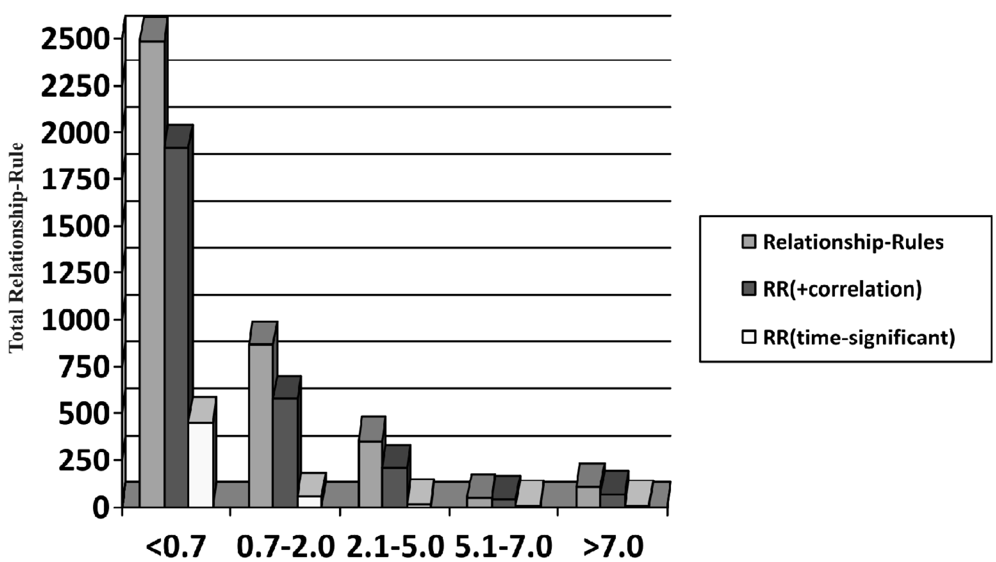

Figure 3: Interesting Time-Varying Weighted Relationship-Rules

V. CONCLUSION

In this script, we have projected an incipient mix approach for data mining procedure. The projected algorithm's provides a competent time responsive approach for mining recurrent items in data-set. Here we propose a new approach for weighted time varying relationship-rule mining with the rules of interest using correlation analysis. With the use of weighted concept along with time varying dataset, proposed approach considers only the reduct of the initial database. Hence, it is clear that the proposed approach works on the reduced dataset which leads to the enhancement in the performance. Furthermore, proposed approach makes use of the correlation analysis to come out with interesting relationship-rules.

Projected system is the improvement of mining algorithm of relationship-rule with weighted constraint of item in the transaction. With the aid of conception of time, weighted constraint and interestingness factor, we consider the frequent patterns of interest that have enough support and confidence in their lifespan period with the required their weighted-utility in transactional dataset.

REFERENCES

[1] Agrawal, R.-Mannila, H.-Srikant, R.- Toivonen, H- Verkano, I.: Fast Discovery of Association Rules. MIT Press, (1996).

[2] Srikent and Agrawal, Fast-Algorithms for Mining Association-Rules. International Business Machine, Almaden (1994)

[3] Agrawal and R Srikant, Mining Sequential Patterns, Proceeding IEEE International Conference on Database Engineering, (1995).

[4] Juan M. Ale and Gustavo Rossi, An approach to discovering temporal Association Rules, Association for Computing Machinery symposium on Applied Computing, (2000).

[5] Book::Jiawei Han, Kamber and Jian Pei, Data Mining: Concepts and Techniques, 3rd Edition, The Morgan Kaufmann Publishers, (2011).

[6] Chang Lee, Cheng.-Ru Lin and Ming-Syan Chen, Sliding-window filtering: An Efficient algorithm for Incremental Mining, Proceeding of Association for Computing Machinery International Conference on Information and Knowledge Management, (2001).

[7] Xiaoxin Yin and Jiawei Han, CPAR: Classification based on predictive association rule, San Francisco, (2003).

[8] Abdelijaber Thabtah, Peter Cowling and Yonghong Peng, Comparison of classification techniques for a personnel scheduling problem, Proceeding of International Business Information Management Conference, (2004).

[9] Ken Sun and F Bai, Mining Weighted Associa-tion Rules without Pre-assigned Weights, Knowledge and Data Engineering, IEEE Transactions, (2008).

[10] Book:: Ian H. Witten, Eibe Frankand mark A. Hall, Data mining: Practical Machine learning tools and techniques, Morgan Kaufmann Publishers, (2000).

[11] Frans Coenen and Paul Leng, An Evaluation of Approaches to Classification Rule Selection, Proceeding of 4th IEEE International Conference on Data Mining, (2004).

[12] Ahmed Ayad, Nagwa El-Makky and Yousry Taha, Incremental Mining of Constrained Association Rules, Proceeding of SIAM Conference on Data Mining, (2001).

[13] Weka:: Data Mining Software, Available:http://www.cs.waikato.ac.nz/ml/weka.

[14] R. Rymon, Search Through Systematic Set Enumeration, Proceeding 3rd International Conference Principles of Knowledge and Reasoning, (1992).

[15] Deepak Kumar and Pankaj Gupta, Fuzzy Software Release Problem with different Types of learning Functions for Fault Dictions and Correction Processes, Advances in Intelligent Systems and Computing, Springer, (2015).

[16] Sergey Brin, Rajeev Motwani and Craig Silverstein, Beyond Market-Basket:: Generalizing Association Rules to Correlations, Spacial Interest Group on Mgmt. of Data (SIGMOD), (1997).

[17] B Ozden , S Ramaswamy and A Silberschatz, Cyclic Association Rule, Proceeding of 14th International conference on Data Engineering, (1998).

[18] Jiawei Han, Jian Pei, Yiwen Yin and Runying Mao, Mining Frequent Pattern without Candidate Generation, Proceeding ACM-SIGMOD International Conference Management of Data, (2000).

[19] J F. Roddick, K Hornsby, M Spiliopoulou, An Updated Bibliography of Temporal, Spatial, and Spatio-temporal Data Mining Research, TSDM, (2000).

[20] T. Herawan, P. Vitasari and Z. Abdullah, Mining interesting association rules of student suffering mathematics anxiety, Communication of Computer and Information Sciences, Springer, (2013).

[21] Pankaj Gupta and B.B.Sagar, Discovering Weighted Time Varying Temporal Relationship, Advances in Intelligent Systems and Computing, Springer, (2015).