ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 9(10), DOI: 10.17485/ijst/2016/v9i10/84989, March 2016 ISSN (Online) : 0974-5645*

A Partial Weighted Utility Measure for Fuzzy Association Rule Mining

P. Kayal[1*] and S. Kannan[2]

[1]Research and Development Centre, Bharathiar University, Coimbatore – 641046, Tamil Nadu, India; kayalpaddu@gmail.com
[2]Department of Computer Science, M.K University, Madurai – 625021, Tamil Nadu, India; skannanmku@gmail.com

Abstract

Background/Objectives: Association rules are generated from frequent item set by Association mining. The generation of frequent item set makes a great impact on decision making. The objective of the work is to introduce a new measure called SUF (Skill Utility Factor) to extract meaningful hidden item set and develop a hybrid algorithm FPWUM (Fuzzy Partial Weighted Utility Mining) for decision making. **Methods/Statistical Analysis:** The traditional measures support and confidence is augmented with SUF which can be useful for Human resource personnel to easily predict the work-force calibre in an organization. Using different methods like association rule mining, fuzzy logic and weighted utility mining has improved the prediction of attributes relations efficient and faster. **Findings:** The FPWUM extracts more efficient hidden frequent item sets through which many new and interesting rules are generated. Since the application of attribute's weight are handled wisely and improvising factor is used only for hidden item set the model process time is reduced fairly. The idea of integrating the conventional measures and the SUF is a unique technique. The approach works well on real time dataset compared to the conventional models. The comparative result shows the algorithm's ability. **Improvements/ Applications:** The algorithm uses predefined weighting scheme. It can be enhanced by using dynamic intelligent weighting factor.

Keywords: Fuzzy Association Rules, Hidden Item set, Partial Weight, Skill Utility Factor, Weighted Utility Mining

1. Introduction

1.1 Data mining

Raw data in large volume can be of no use in Data processing techniques or Decision making. The data and their relationship can be discovered through Data mining which is an intelligent task that can turn large raw data volume into Knowledge. For such mining process various database techniques, Statistical concepts, Machine learning tools and Artificial Intelligence algorithms are employed[1].Though the arise of new technique and application are appreciable, other way there is enormous amount of issue related to new discovery of algorithm and knowledge related to performance. Always algorithms which are robust and scalable are noticeable[2].

1.2 Association Rule Mining

A standard association rule[3] is a rule of the form X→Y which says that if X is true of an instance in a database, so is Y true of the same instance, with a certain level of significance as measured by two indicators, support and confidence[4].Here let imagine $\mathbf{I} = \{i_1, i_2, ..., i_n\}$ be Items which is a set of attributes and $\mathbf{D} = \{t_1, t_2, ..., t_n\}$ be the Database which is a set of transactions. Each transaction in the Database D contains unique ID and a subset of the items which is in **I**. A rule can be defined as **X→Y** where $\mathbf{X, Y} \subseteq \mathbf{I}$ and $\mathbf{X} \cap \mathbf{Y} = \emptyset$. The **X** and **Y** are called antecedent and consequent of the rule respectively[4]. From this many rules can be generated. In order to choose the most interesting rules from the group of all possible generated rules, some measures are necessary that depicts the rule

* *Author for correspondence*

interestingness and significance. Such universally proved measures are Support and Confidence[4].

2. Fuzzy Association Rules

2.1 Fuzzy Set and Domain Partitioning

The use of fuzzy approach in our research is well discussed in[5]. The ideology of integration of Association rules with Fuzzy theory is most popular in Big data research for long time. A fuzzy set is a class of objects with grades of membership ranging between [0,1]. The technique called Domain partitioning is used with quantitative association rule mining. This type of partitioning, arise the sharp boundary problem which is explained in[5]. To resolve such problem, a bendable setting of the boundaries interval is enviable, and therefore fuzzy logic is applied[6].

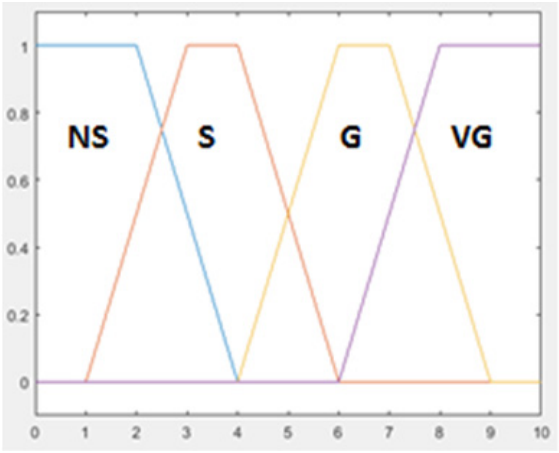

Figure 1. Membership function.

Table 1. Leadership skill level

NS	Not satisfactory(NS)
S	Satisfactory(S)
G	Good(G)
VG	Very Good(VG)

2.2 Fuzzification

The effort of changing a scalar value into a fuzzy value is called fuzzification. A variety of fuzzifiers called as membership functions are used for fuzzification. Some well-known membership functions are Trapezoidal,Triangular ,Gaussian and Bell. For a value **u** , knowning x,y,z (start, peak, end), the membership for **u** in all the assumed partitions can be calculated[5,7]. For example the attribute MG is further defined into four different fuzzy sets on its domain as NS(MG),S(MG),G(MG),VG(MG) with trapezoidal membership function as shown in Figure 1. This mapping of values (quantitative to linguistic) helps in the process of mining most interesting fuzzy association rules. Once the fuzzication process is done the database is populated with multiple sub-attributes for the basic attributes(CS, MG)[5]. The database after fuzzification is shown as in Table 2.

3. Problem Definition

3.1 Utility Mining

Utility mining is an extensive step of mining frequent item set. A mining which not only depends on the frequency of the item set but also the utility involved with the item set is called utility mining. Here Utility is referred to the importance of the item set in the transactions considering in terms of any user specific preferences other than frequency of the items alone[1]. This item set consist of frequent item set and rare item set too. In many domain applications, rare item set play a major role in decision making. The high-profit rare item sets are found to be very useful in many application areas. For example, in medical application, the rare combination of symptoms can provide useful insights for doctors[8-9]. Here in our

application the frequent item-set mined using Fuzzy Partial Weighted Utility Mining (FPWUM) algorithm, helps in finding out different combination of capabilities which helps to understand and decide a better leader without missing even a single skill of a workforce.

Table 2. Fuzzification values

Emp Id	CS				MG				AS				TD			
	NS	S	G	VG	NS	S	G	VG	NS	S	G	VG	NS	S	G	VG
E01	0	0.66	0.33	0	0	0.5	0.5	0	0.5	0.5	0	0	0	0	1	0
E02	0	0.66	0.33	0	0	1	0	0	0	0.5	0.5	0	0	0	0.5	0.5
E03	0	0	0.5	0.5	0	0.25	0.5	0.25	0	0	0.5	0.5	0	0.5	0.5	0
E04	1	0	0	0	0	0.66	0.33	0	0	0	0.5	0.5	1	0	0	0
E05	0.5	0.5	0	0	1	0	0	0	0.5	0.5	0	0	1	0	0	0

3.2 Utility Item Set Mining

Definition 1. Attribute Weight is a non-negative real value given to each attribute ranging [0 -1] with some degree of importance assigned by domain experts.

Definition 2. Skill Utility Factor (SUF) is sum of product of items and its associated weight, of all the transactions containing the item set divided by the no of occurrences of the item set.

Definition 3. Skill Utility Factor Threshold (min_SUF) is the user specified threshold value for the selection of frequent item set which will be between 0 and N, where N is the fraction of number of tuples and the number of attributes.

Definition 4. Fuzzy Item Support Value (FISV) is the support value of the item set specified warily by the expert/user to prevent over-growth or under- growth of rules.

Definition 5. Frequent Item Set (FIS) is the item set which has the support value >min_support (FISV) which is calculated by fractioning the sum of minimum of the item set and the total number of transactions in the database.

Table 3. Part of Transaction table

CS(S)	MG(S)	AS(S)	TD(G)
0.66	0.5	0.5	1
0.6	1	0.5	0.5
0	0.25	0	0.5
0	0.66	0	0
0.5	0	0.5	0

3.3 Fuzzy Partial Weighted Utility Mining (FPWUM)

The k-frequent item set is generated using candidate generation. The minimum Fuzzy Item Set Support Value (FISV) is set as 30%. The k-item frequent set is generated from k-1 item. For frequent set mining not only FISV but SUF is also considered. So there is a possibility of considering in frequent item set but with high skill utility factor in our rule mining**.** When the item set satisfies FISV it is added to the frequent item set and there is no need to calculate SUF. If it does not satisfy, then the SUF of the item set is calculated. If the calculated SUF of that particular item set satisfies the min_SUF then it is added to the frequent item set else it is pruned. The necessity of SUF comes because of setting FISV too high or too low. Both leads to rule over-fitting or rule under- fitting situation respectively. In order to prevent such situation and not to let pass any kind of attributes with higher precedence, FISV combined with SUF works well. Some of the k-item set are given in Table 4.The FISV is set as 30% and min_SUF is set as 0.7 for calculation.

Table 4. K-Item set with FISV and SUF

	Item sets	FISV	SUF
Set 1	CS(S),MG(S),AS(S),TD(G)	0.4	NA
Set 2	CS(VG),TD(G)	0.2	0.55
Set 3	CS(G),AS(VG)	0.2	0.9
Set 4	CS(G),TD(G)	0.6	NA
Set 5	MG(NS),AS(NS)	0.2	0.8

Table 5. Weight associated with item

Attribute	Weight(W)
AS	1.0
LT	0.7
MG	0.6
OP	0.6
TD	0.3
SH	0.4
CS	0.8
CR	0.6
CM	0.7
IG	0.8
DD	1.0
RL	1.0

3.3.1 Skill Utility Factor (SUF) Calculation

Consider Table 3, 4 and 5 for transactions, Item-set and weight respectively. The Item-set 3 {CS (G), AS (VG)} has a FISV of 20% which is less than the specified FISV. But it has a higher SUF.

$$SUF = \sum I_i \times W_i \div N \qquad \text{Eq(1)}$$

3.3.2 Example

SUF {CS (G), AS (VG)} = (CS (G)*W+AS (VG)*W))/ N where N is the no of occurrences of the item set.

= {(0.5*0.8) + (0.5*1)/1

= 0.9

Similarly the set 5 has higher SUF. But the set 2 has less FISV and less SUF. So the set 2 is pruned but not the set 3 and 5. This example illustrates the fact that frequent item-set mining approach may not always satisfy the goal

but things to think with skill utility attributes too. The comparative results are shown in section 4.

3.3.3 Algorithm

- Identify item set ($I_i \in n$)
- If it satisfies the min_support threshold (positive item set), add those item set to frequent item set list ($freq\ (I) = I_i$).
- If it does not satisfy min_support threshold (negative item set) find SUF.
- To find SUF, sum the product of item of the item set, and its corresponding weight for all the transaction where the item set occurred, by the no. of item set occurrences.
- If SUF >=min_SUF add the item set to frequent set list else the item set is pruned.

Algorithmic flow shown in Figure 2.

```
For (I)
    Find (I_i ∈ n)
      For (I_i ∈ n)
            If    I_i >= min_sup
              freq(I) = I_i
            Else
              For (t_i ∈ I_i)
                 SUF  = Σ I_i × W_i ÷ N
                If  SUF >= suf _threshold (SUT)
                  freq(I) = I_i
              end for
            end if
      end for
end for
```

Figure 2. FPWUM algorithmic flow.

3.4 Weighting Scheme

The weight and threshold values specified by the user are from the margin of significance of personnel skills point of view. Applying weights only to item set which does not satisfy the support, has two significant importance.

- Applying weights to item set above the support value has no meaning, because it does not need any improvising factor to be considered in the algorithm. This enhances model's process time and gets rid of unnecessary weighting process.
- The goal of using SUF is to make use of the weight in the mining process and prioritize the selection of different hidden item sets (item set less than support) according to the significant skills required, rather than the frequency alone and now here violating Downward Closure Property (DCP) too.

3.5 Fuzzy Association Rule Generation

Mining fuzzy association rules is the discovery of association rules using fuzzy set concepts such that the quantitative attributes can be handled. Here we view each attribute as a linguistic variable, and the variables divided into various linguistic terms. Fuzzy association rules are expressed in form: If X is A -> Y is B. For example,

if (age is young) → (salary is low)

Given a database D, attribute I with item sets $X \subset I$, $Y \subset I$ and X and Y as X= {AS, LT, MG, OP, TD, SH} Y = {CS, CR, CM, IG, DD, RL} where $X \cap Y = \phi$, we can define sets as A={NS,S,G,VG} and B={NS,S,G,VG} with X,Y respectively. For example (X,A) as (MG, G), (MG,S) and (Y,B) as (CS,NS), (IG,V). The semantics of the fuzzy rule is that when the antecedent **X is A** is satisfied, we can imply that **Y is B** is also satisfied, which means there are sufficient records that contribute their counts to the attribute fuzzy set pairs and the sum of these counts is greater than the specified threshold (FISV), and the fuzzy set pairs formed by FPWUM is greater than the min_SUF.

4. Experimental Observations

4.1 Frequent Item Set Generated with FISV

The frequent item set generated by FISV is shown in Figure 3. The number of frequent item set generated are very less and even it missed many hidden interesting item set compared to the FPWUM method which is shown in Figure 4.

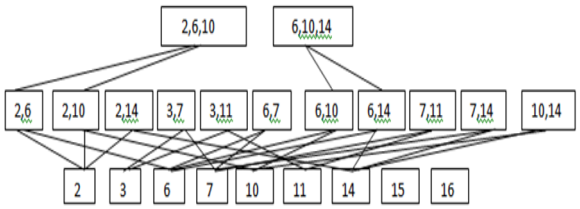

Figure 3. Frequent item generated using FISV.

4.2 Frequent Item Set Generated by FPWUM

In addition to the frequent item sets formed by FISV there is also some other interesting item set generated part of which is shown in Figure 3, plays a vital role in

determining the skill set for the work-force leadership prediction. The greyed box represents the new item set generated by FPWUM and the white box represents the item set that satisfy min_support. This increases the number of interesting fuzzy association rules and we could find more hidden relationships of the attributes.

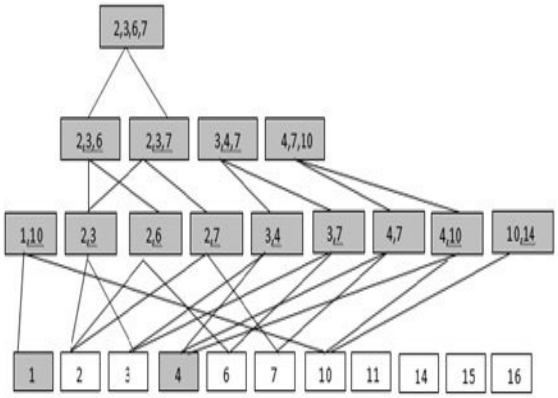

Figure 4. Frequent item set generated using FPWUM.

4.5 Comparative Results of FISV Method and FPWUM

Two sets of experiments were undertaken with two different algorithms namely FISV and FPWUM. Figure 5 explains the Number of frequent Item set generated by both the algorithms. Figure 6 shows the Number of Fuzzy association rules generated. Figure 7 displays the Execution time taken to generate the rules.

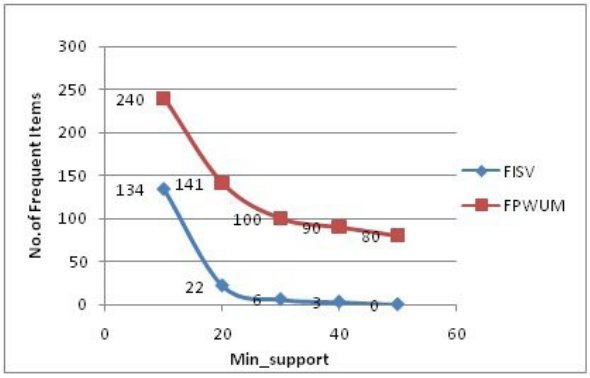

Figure 5. Number of frequent item set generated.

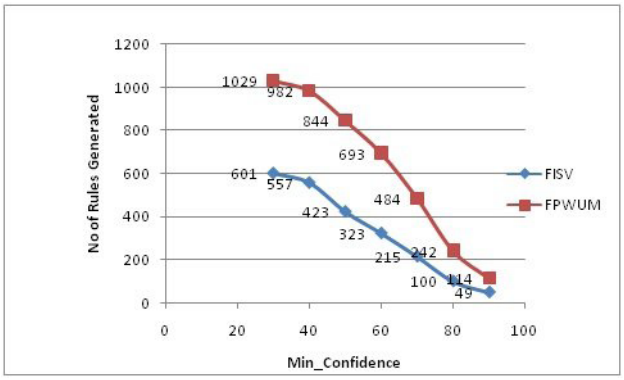

Figure 6. Number of fuzzy association rules generated.

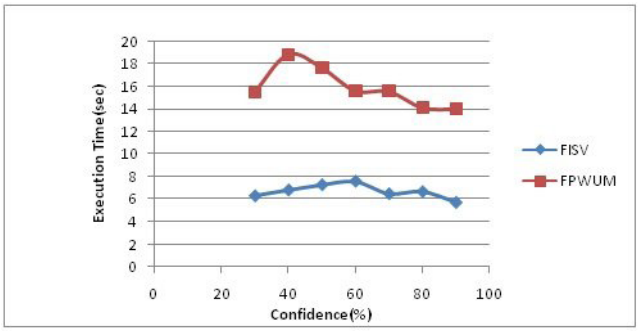

Figure 7. Execution time to generate fuzzy association rules.

The results shows that the proposed algorithm produces better results as it uses all the possible interesting hidden item sets from the so called infrequent item set too and generates Fuzzy Association Rules without violating Downwards Closure Property (DCP).

5. Conclusion

Initially the algorithm takes the entire positive item set considering the universal proven support measure and then it works on the remaining item set for algorithmic utility weighting. We identify the challenge of using weights in the iterative process of generating large meaningful item sets and thereby bringing good fuzzy association rules. Hence the number of hidden correlations of skills set of an individual are found high compared to the traditional FISV method and therefore the leadership capabilities of a work-force can be interpreted clearly and accurately by FPWUM.

6. References

1. Barakare A, Zawar M. A survey on high utility item set mining from transactional databases. International Journal of Advanced Research in Computer Science and Software Engineering. Oct 2015; 5(10):393-6.
2. Razak TA, Ahmed GN. Detecting credit card fraud using data mining techniques- meta-learning. Indian Journal of Science and Technology. Oct 2015; 8(28):1-9.
3. Jafarzadeh H, Torkashv RR, Asgari C, Amiry A. Provide a new approach for mining fuzzy association rules using apriori algorithm. Indian Journal of Science and Technology. Apr 2015; 8(8):707-14.
4. Han J. Data mining: Concepts and techniques. 3rd edn. Morgan Kaufmann Publishers Inc: San Francisco, CA, USA; 2011.
5. Kayal P, Kannan S. Building fuzzy associative classifier using fuzzy values. International Journal of Science and Research. Jul 2014; 3(7):1498-1500.
6. Ross TJ. Fuzzy Logic with Engineering Applications. 3rd-edn. John Wiley and sons Ltd: United Kingdom; 2010.
7. Chen Z, Chen G. Building an associative classifier based on fuzzy association rules. International Journal of Computational Intelligence Systems. Aug 2008; 1(3):262-73.
8. Lan GC, Hong TP, Tseng VS. A novel algorithm for mining rare-utility item sets in a multi-database environment. Proceedings of the 26th Workshop on Combinatorial Mathematics and Computation Theory. 2009. p.293-302.
9. Tseng VS, Chu CJ, Liang T. Efficient mining of temporal high utility item sets from data streams. Proceedings of Second International Workshop on Utility-Based Data Mining, Philadelphia. Aug 2006. p.1-75.