

Mining predicate rules without minimum support threshold

Hafiz I. Ahmad^{1,*}, Alex T. H. Sim¹, Roliana Ibrahim¹, Mohammad Abrar², Asma Gul³

¹*School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, 81300, Johor,*

²*Dept. of Computer Science, Mohi-ud-Din Islamic University, AJK*

³*Dept. of Statistics, Shaheed Benazir Bhutto Women University, Peshawar*

**Corresponding author: eshfaqahmad@gmail.com*

Abstract

Association rule mining (ARM) is used for discovering frequent itemsets for interesting relationships of associative and correlative behaviors within the data. This gives new insights of great value, both commercial and academic. The traditional ARM techniques discover interesting association rules based on a predefined minimum support threshold. However, there is no known standard of an exact definition of minimum support and providing an inappropriate minimum support value may result in missing important rules. In addition, most of the rules discovered by these traditional ARM techniques refer to already known knowledge. To address these limitations of the minimum support threshold in ARM techniques, this study proposes an algorithm to mine interesting association rules without minimum support using predicate logic and a property of a proposed interestingness measure (g measure). The algorithm scans the database and uses g measure's property to search for interesting combinations. The selected combinations are mapped to pseudo-implications and inference rules of logic are used on the pseudo-implications to produce and validate the predicate rules. Experimental results of the proposed technique show better performance against state-of-the-art classification techniques, and reliable predicate rules are discovered based on the reliability differences of the presence and absence of the rule's consequence.

Keywords: Association rule mining; coherent rules; interestingness measure; minimum support threshold; predicate logic

1. Introduction

Association rule mining (ARM) is a technique widely used for the discovery of relationships among frequently appearing itemsets in large databases. The primary focus of ARM is to find rules enabling users to predict the co-occurrence of items and to discover more interesting relations. A widely popular ARM algorithm is Apriori proposed by Agrawal *et al.* (1993) that requires the presetting of minimum support (*minsupp*) and minimum conf (*minconf*) in a support-confidence framework. Any rule that resides in the limits of the predefined *minsupp* and *minconf* threshold are considered interesting. The *minsupp* threshold derives the frequent itemsets, and then ARs are discovered from these frequent itemsets. This threshold controls the number of ARs discovered, but there is no

clear way for the selection of the threshold value in different datasets (Borgelt, 2012; Fournier-Viger *et al.*, 2019; Salam & Khayal, 2012). These thresholds are estimated intuitively by users, and depending on the choice of these parameters, a huge number of ARs can be found in a dataset that may have redundant rules (Fournier-Viger *et al.*, 2017; Hasan & Mishu, 2018). If the *minsupp* threshold is set to a low value, an exponential number of ARs are discovered (Mustafa *et al.*, 2006). On the other hand, if the *minsupp* threshold value is too high then very few ARs may be generated, and useful knowledge might be left out.

To eliminate the requirement of presetting *minsupp* threshold for mining ARs, Sim *et al.* (2010) proposed a logic-based rule mining technique using

propositional logic without presetting the *minsupp* threshold. The rules discovered in their technique are called coherent rules, and any rule that satisfies the equivalence property of propositional logic is a coherent rule. If a rule $X \rightarrow Y$ satisfies the logic equivalence, then the rule $\neg X \rightarrow \neg Y$ must also satisfy the equivalence where X and Y are two item-sets. On the other hand, if $\neg X \rightarrow Y$ satisfies the logic equivalence, then $X \rightarrow \neg Y$ must also satisfy the equivalence property under certain conditions.

The motivation of this research is to overcome the limitations of the traditional ARM techniques in presetting the *minsupp* threshold based on the discovery of coherent rules. Therefore, this research makes two key contributions i.e. extension of coherent rule mining technique with predicate properties to discover predicate rules without *minsupp* and *minconf* threshold and introducing a new interesting measure to capture interesting predicate rules based on the reliability differences between presence and absence of rule's consequences.

The rest of the paper is organized as follows. Section 2 presents related work. The proposed technique of discovering predicate rules and interestingness measure is described in Section 3. Section 4 presents the justification of the novel technique proposed through experimental results. A detailed discussion is presented in Section 5. Finally, Section 6 concludes the study.

2. Related work

ARM is an evolving research area and various algorithms have been developed for the generation of strong interesting association rules (ARs) in large datasets. The prominent algorithms include the positive AR (Agrawal *et al.*, 1993; Baralis *et al.*, 2008; Krishnapuram, 2016), negative AR (Balakrishna *et al.*, 2019; Jabbour *et al.*, 2018; Kong *et al.*, 2018), and combined approach for negative and positive rules for large datasets (Bagui & Dhar, 2018; Bemarisika & Totohasina, 2018; Parfait *et al.*, 2018; Zhao *et al.*, 2017). Moreover, application specific ARM algorithms have been developed for various areas including medicine (Bo-rah & Nath, 2018; Harahap *et al.*, 2018; Moses *et al.*, 2015), crime (Buczak & Gifford, 2010; Hassani *et al.*, 2016), agriculture (Bhatia & Gupta, 2014; Bisht & Samantaray, 2015; Geetha, 2015), distributed environments (Qin *et al.*, 2016; Salah *et al.*, 2017), data warehousing (Usman, 2017) etc. Mustafa *et al.* (2006) presented a technique called the Enhanced Apriori Algorithm (EAA) based on second support and confidence for

the discovery of association rules for significant rare data. However, very less attention has been given to address the limitations of the support-confidence framework.

2.1 Supportless ARM

To overcome the limitations of the Apriori and FP-Tree based algorithms, researchers have proposed algorithms for the discovery of ARs without presetting the *minsupp* and *minconf* threshold. An emerging pattern (EP) based approach for discovering all top rules are proposed by (Li *et al.*, 1999) without presetting a *minsupp* threshold. The technique requires the consequence of the expected top rules, based on which the final rules are discovered. In this technique, the dataset is divided into two sub-datasets according to the provided consequent of the rule. Similarly, to replace the requirement of *minsupp* threshold, Koh (2008) proposed a statistically meaningful function called minimum absolute function (MinAbsSup). This function generates a minimum absolute value for each candidate itemset that is used to decide the inclusion or exclusion of a certain rule from the interesting rules (Koh, 2008).

Another supportless ARM technique is presented by Sim *et al.* (2010) for the discovery of a pair of ARs called coherent rules using the concepts of propositional logic. The inclusion of propositional logic extracted strong and interesting rules from datasets without using the *minsupp* threshold. They introduced the "ChSearch" algorithm generates coherent rules from the database instead of frequent itemsets. Moreover, as opposed to the traditional ARM algorithms, ChSearch discovers both the negative and positive ARs (Sim *et al.*, 2010).

2.2 Predicate logic and inference rules

Predicate logic is a mathematical model for reasoning with functions/predicate to map variables to truth values and construct proofs using inference rules for discovering new facts. These rules of inference constitute a set of arguments that are used as part of natural deduction for proving complex arguments. An inference rule is a form of a function in logic that takes premises, analyses these premises, and provides a conclusion. A list of inference rules are given in Table 1.

Predicate logic has been used in the ARM technique by designing an algorithm for mining predicate ARs using gene expression programming (GEP)(Zuo *et al.*, 2002). Predicate ARs discovered

Table 1. List of inference rules Rosen & Krithivasan (2012)

S.No	Rule Name	Formula
1	Modus Ponens	$\frac{P \rightarrow Q}{P} \frac{P}{Q}$
2	Modus Tollens	$\frac{P \rightarrow Q}{\neg Q} \frac{\neg Q}{\neg P}$
3	Hypothetical Syllogism	$\frac{P \rightarrow Q}{Q \rightarrow R} \frac{Q \rightarrow R}{P \rightarrow R}$
4	Disjunctive Syllogism	$\frac{P \vee Q}{\neg P} \frac{\neg P}{Q}$
5	Constructive Dilemma	$\frac{(P \rightarrow Q) \cdot (R \rightarrow S)}{P \vee R} \frac{P \vee R}{Q \vee S}$
6	Simplification	$\frac{P \cdot Q}{P}$
7	Conjunction	$\frac{P}{P \cdot Q} \frac{Q}{P \cdot Q}$
8	Addition	$\frac{P}{P \vee Q}$

by the proposed algorithm are novel and the experiments showed that the traditional ARM techniques were unable to discover these rules. Predicate logic in combination with fuzzy logic has been studied by Ceruto *et al.* (2014) and proposed an algorithm called FuzzyPred and implemented on multiple medical datasets. The proposed algorithm obtains patterns through fuzzy predicates representing the dependence between items in the datasets. Similarly, Ikram & Qamar (2015) proposed an algorithm and developed a rule-based expert system for the prediction of earthquakes using predicate logic and inferential rules. From their experiments on real-world dataset, they have achieved higher accuracy in predicting earthquakes that happened within 12 hours at most.

3. The proposed rule mining algorithm

3.1 Predicate rule mining algorithm

The proposed predicate rule mining Algorithm is an integration of the coherent rule mining algorithm (Sim *et al.*, 2010) with the concepts of predicate logic and inferential rules. The proposed algorithm for mining predicate rules is described in Figure 1.

3.2 Proposed interestingness measure

We introduced a new interestingness measure i.e. g measure, which selects interesting combinations and calculates the interestingness of the predicate rules discovered. g measure ensures that the interesting rules selected are highly reliable for decision-makers and represents the whole database correctly. Mathematically, g measure is given in Equation 1.

$$g = \frac{P(XY)}{P(Y)} - \frac{P(XY\prime)}{P(Y\prime)} \quad (1)$$

The mathematical proof of the proposed measure g in Equation 1 is given as:

$$\begin{aligned} &= \frac{P(XY)}{P(Y)} - \frac{P(X) - P(XY)}{1 - P(Y)} \\ &= \frac{P(XY)(1 - P(Y))}{P(Y)(1 - P(Y))} - \frac{P(Y)(P(X) - P(XY))}{P(Y)(1 - P(Y))} \\ &= \frac{P(XY) - P(XY)P(Y)}{P(Y)(1 - P(Y))} \\ &= \frac{(Y)(P(X) - P(Y)P(XY))}{P(Y)(1 - P(Y))} \end{aligned}$$

$$\begin{aligned} g &= \frac{P(XY) - P(Y)P(X)}{P(Y)(1 - P(Y))} \\ g &> 0 \text{ iff } P(XY) > P(X)P(Y) \\ &\text{or } |P(Y) - P(Y\prime)| > 0 \quad (2) \end{aligned}$$

Equation 2 shows that the measure compares the reliability differences between the presence and the absence of the rule's consequence, Y . Moreover, the g measure will have a higher value if the consequence of the rule is rare or frequent and will also capture the relation $P(XY) > P(X)P(Y)$. The property $(P(XY) > P(X)P(Y))$ of the g measure is used to search for predicate rules. The value range of g measure is $[-1, 1]$. If the g measure value is greater

Fig. 1. Predicate Rule Mining Algorithm

INPUT: D – a database, n – number of distinct itemsets, m – number of attributes, X – antecedent of rule, Y – consequence of rule, S – support, Q_i – quadrants of contingency table

OUTPUT: A set of predicate rules (PR)

STEP 1: PI = [set of pseudo-implications]
 PR = [set of predicate rules]
 $g = []$ [rule's interestingness measure]
 Y = class [pre-defined]
 U = No of unique values in D
 A = U – Y
 DPC[1] = D_1 itemset [set of possible combinations]
 k = 2
 While DPC[k-1] $\neq \emptyset$
 X = generate all combinations |
 $x \subset DPC \cup w \subset D[k] \wedge x \neq w$
 If $P(XY) > P(X)P(Y)$, then
 DPC[k] = $\bigcup \{X, Y\}$

STEP 2: [Map each combination to pseudo-implication based on the conditions]
 if $S(Q_1)$ is greater than $S(Q_2)$, $S(Q_3)$, and $S(Q_4)$ then map Q_1 to $PI(p \rightarrow q)$
 else if $S(Q_2)$ is greater than $S(Q_1), S(Q_3)$, and $S(Q_4)$ then map Q_2 to $PI(\neg p \rightarrow q)$
 else if $S(Q_3)$ is greater than $S(Q_1)$, $S(Q_2)$, and $S(Q_4)$ then map Q_3 to $PI(p \rightarrow \neg q)$
 else if $S(Q_4)$ is greater than $S(Q_1)$, $S(Q_2)$, and $S(Q_3)$ then map Q_4 to $PI(\neg p \rightarrow \neg q)$
 End if

STEP 3: [Validate all pseudo implication in PI against the inference rules of Table 1 over $\{\neg, \wedge, \vee, \rightarrow\}$]
 If PI_I Validate over $\{\neg, \wedge, \vee, \rightarrow\}$ then
 PR = $\cup PI_I$
 End if

STEP 4: [calculate the interestingness of predicate rules]
 $g[i] = \frac{P(XY)}{P(X)} - \frac{P(XY)}{P(Y)}$
 End if
 k=k+1

STEP 5: [sort the set PR] PR = sort (PR, g, descending)
 Return PR

than 0, then X and Y have a positive correlation. If the g measure value is less than 0, then X and Y has a negative relationship among each other. Lastly, if the g measure has a 0 value, then X and Y are independent of each other.

3.3 An example on discovering predicate rules

In this section, a small sample of the zoo dataset is taken as an example to show how the algorithm can mine predicate rules. The dataset includes five attributes where the type represents the class of each record and has integer values (1, 2, 6). The other 4 attributes are binary attributes where 1 represents the presence while 0 represents the absence of the attribute as shown in Table 2.

STEP 1: The process of discovering predicate rules start by generating combinations that satisfy the condition $P(XY) > P(X)P(Y)$. In addition, if an item from the antecedent of the rule does not satisfy the condition all the future combinations containing that item are discarded and not searched. To demonstrate the process of discovering predicate rules, let's take the dataset given in Table 2 where the type attribute (consequence) is set as fixed. A total of 234 different combinations are possible, considering the type attribute has three integer values (1, 2, 6). Out of the possible 234 combinations, only 34 combinations are generated where only 8 combinations satisfied the conditions as shown in Table 3.

For example, the combination between hair (hair = 1) and the type attribute (for type = 1) satisfies the condition because the $P(XY)=0.5$ and $P(X)P(Y)=0.3$, and $0.5 > 0.3$ so the combination (hair, type) is selected and stored in the set DC. The second combination is between the attributes hair, feathers and type where the possible combi-

Table 2. A dataset extracted from zoo dataset

Rid	hair	feathers	milk	breathes	type
1	1	0	1	1	1
2	1	0	1	1	1
3	0	1	0	1	2
4	0	1	0	1	2
5	1	0	1	1	1
6	0	0	0	1	6
7	0	1	0	1	2
8	1	0	1	1	1
9	0	0	1	1	1
10	1	0	1	1	1

Table 3. List of combinations selected

No of At-tributes	Combinations
2	{hair, type(1)}, {not feather, not type(2)}, {type(1), milk}
3	{hair, milk, type(1)}, {not feather, not breathes, not type(2)}, {hair, breathes, type(1)}, {milk, breathes, type(1)}
4	{hair, milk, breathe, type(1)}

nations with type(1) are hair(1) feather(1) type(1), hair(1) feather(0) type(1), hair(0) feather(1) type(1) and hair(0) feather(0) type(1). These do not satisfy the condition and are discarded. For example, for the combination hair(1), feather(1) and type(1) the $P(XY)=0$ and the $P(X)P(Y)=0$, thus the condition $P(XY)>P(X)P(Y)$ is not satisfied and the combination is discarded. Similarly, these combinations with the type (2) and type (6) does not satisfy the condition, so any future combination that has hair and feather will never be generated and will be automatically discarded.

STEP 2: The contingency tables of each of these distinct combinations stored in set DC are calculated. The contingency tables of two combinations are given in Table 4 for type=1 that are between the attributes hair(H), milk(M), breathes(B) with the type(T).

Table 4. Contingency tables for combinations

	M	$\neg M$
T	6	0
$\neg T$	0	4

	T	$\neg T$
H,B	5	0
$\neg (H,B)$	1	4

Each of these combinations are then mapped to pseudo-implications if the support value of a combination is greater than all the other values in the remaining three quadrants. Thus, Q1 in the first contingency table is mapped to a pseudo-implication as its support value is greater than the support value of the other quadrants. Similarly, in the second contingency table, the combination $Q1=H,B \rightarrow T$ with

$S(Q1)=5$ is mapped to a pseudo-implication and are stored in the set PI.

STEP 3: Next, from the set PI containing all the pseudo-implications, two rules are taken, and inference rules are applied for validation and discovery of other possible interesting pseudo-implications. One of the inference rules given in Table 1 is applied as an example on the two pseudo-implications from STEP 2. The premises of these rules follow the template of the hypothetical syllogism rule. It is given as:

$$\frac{H, B \rightarrow T}{\frac{T \rightarrow M}{H, B \rightarrow M}} \quad (3)$$

The implementation of the inference rule resulted in a new pseudo-implication $M,B \rightarrow H$, which is a valid rule based on the conditions for mapping combinations into pseudo-implications of STEP 2. Next, two more pseudo-implications are selected, validated and new rules are discovered until all the pseudo-implications are validated from set PI. The resultant rules referred as predicate rules and stored in the set PR.

STEP 4 and STEP 5: This step calculates the interestingness value (STEP 4) of each predicate rule discovered and arrange them in a descending order (STEP 5) to list the most important and interesting rule at the top while the least interesting rule is placed at the bottom of the list as shown in Table 5.

Table 5. List of rules ordered with g measure

Rule	g value
not feathers \rightarrow not type (2)	1
milk \rightarrow type (1)	1
not feathers, not breathes \rightarrow not type (2)	1
milk, breathes \rightarrow type(1)	1
hair \rightarrow type(1)	0.833
hair, breathes \rightarrow type(1)	0.833
hair, milk \rightarrow type(1)	0.833
hair, milk, breathes \rightarrow type(1)	0.833

4. Experimental results

In this section of the study, the experimental results of the proposed technique are discussed. The experiments were conducted on a personal computer with Intel Core i7, 2.8GHz, and 16GB RAM. Datasets are explained in the next section while the evaluation of the experiments is discussed in Section 4.2.

4.1 Dataset description

The datasets used for the proposed predicate rule mining algorithm are zoo, breast cancer and car evaluation that is commonly cited in the literature for ARM (Alatas, 2012; Dechang & Xiaolin, 2008). The dataset is available at the UCI Machine Learning Repository (Bache & Lichman, 2013). The zoo dataset has a total of 17 attributes and 101 instances with no missing values. The type attribute is the class attribute for each animal in which they are classified, and it has integer values ranging between 1 – 7. The description of the datasets is presented in Table 6.

Table 6. Description of datasets

Dataset Name	Attributes	Instances	Classes
Zoo	17	101	7
Breast Cancer	10	286	2
Car Evaluation	6	1728	4

4.2 Evaluation of experiments conducted

Firstly, the experiments were conducted to show the accuracy of discovering predicate rules with different classification techniques on these three datasets. The accuracy of the proposed algorithm is done by training/testing technique. The proposed framework was trained on 70% of the data and tested against the remaining 30% of the data. The comparison of the proposed technique is made with 11 different classification techniques as shown in Table 7.

The comparison of the proposed technique is carried out against the various classification algorithms because the experiments were conducted to discover classification association rules. The overall results show that our proposed algorithm gives a comparable accuracy against all the classifier in the experimental setup. For zoo dataset, the predicate rules discovered provides significant accuracy of 93.33% and outperforming seven of the classification algorithms. The C 4.5 classifier has performed better (90.35%) than the predicate rule mining algorithm (89.58%) but the proposed technique outperformed all the other techniques with exception to SVM. Lastly, on the breast cancer dataset, the predicate rule mining algorithm has performed significantly better than all the other classification

techniques with an accuracy of 93.02%. The total number of predicate rules discovered from the zoo, breast cancer, and the car evaluation datasets are given in Table 8.

The results show that the proposed approach can find predicate rules that are reliable and interesting based on the higher average confidence value and their reliability differences between the presence and absence of items.

5. Discussion

This study extends the discovery of coherent rules by proposing an approach based on predicate logic for mining predicate rules without presetting the minsupp and minconf threshold using a property of a novel interestingness measure called the g measure. The proposed approach, firstly, addresses the limitations of the traditional ARM techniques where predefined support threshold should be provided before discovering ARs. This technique finds the ARs without redefining support and confidence value and still has the capability to find rules that are interesting for users to be utilized. If a predicate rule $X \rightarrow Y$ is discovered using the predicate rule framework, then it is ensured that the rules are reliable for decision making. This is because, the searching property of the g measure compares the reliability differences between the presence and absence of consequence of an association rule. In addition, the exponential searching of all the possible combinations are eliminated by searching combinations in a sequence such that, if any item in a combination does not satisfy the property of g measure then all the future combinations having the same item is discarded. Thus, increasing the processing speed for searching of predicate rules. Also, the rule found will be a valid rule as it has been validated by the inference rules of predicate logic.

Secondly, this study proposes an interestingness measure for finding interesting predicate rules and reject/drop those rules that are redundant and non-interesting. The g measure has a searching property for finding predicate rules from the dataset based on the comparison of reliability differences between the presence and absence of the consequence of the rule. Table 8 show the number of rules discovered and from the results it can be seen that the number of rules discovered is higher and the proposed algorithm can discover interesting rules that also covers a large subset of a dataset, providing the generality and coverage property to the tech-

Table 7. Comparison of classification algorithms with the predicate rule mining algorithm

Classification Techniques	Accuracy		
	Zoo	Breast Cancer	Car Evaluation
Predicate Rule Mining Algorithm	93.33%	93.02%	89.58%
Ada Boost Classifier	73.33%	83.53%	87.07%
C4.5	93.33%	80%	90.35%
Gaussian Naive Bayes	93.33%	81.18%	83.40%
Gradient Boost Classifier	93.33%	80%	83.40%
K Nearest Neighbour	76.67%	75.29%	76.83%
Linear Discriminant Analysis	86.67%	80%	81.85%
Logistic Regression	83.33%	77.65%	78.19%
Multinomial Naive Bayes	80%	82.35%	70.66%
Random Forrest Classifier	93.33%	75.29%	81.66%
AODE	86.67%	77.65%	80.89%
Support Vector Machine (SVM)	76.67%	81.18%	89.58%

Table 8. Number of rules discovered using g

Datasets	No of rules	Average Support	Average Confidence
Zoo	8,052	0.029	0.717
Breast Cancer	31,741	0.029	0.857
Car Evaluation	9,575	0.022	0.424

nique applied. Moreover, the interestingness measure shows that the discovered rules are highly reliable such that important decision making can be done on it. For instance, a predicate rule discovered using the proposed predicate rule mining algorithm is given below:

“If an animal is toothed (1), has a backbone (1), breathe (1) with lungs, and has a tail (1), then the animal is a mammal (type=1) (support = 0.38, confidence=0.89 and calculated g measure = 0.83)”. These rules show that if in a traditional ARM technique, a $minsupp$ of 0.5 is predefined, then this rule will be considered as non-interesting and will be discarded resulting in loss of important information. However, the g measure easily discovers this rule even if the support value is low.

6. Conclusion

In this study, we have proposed an ARM approach for discovering predicate rules without presetting a $minsupp$ and $minconf$ threshold. In addition, an interestingness measure called g measure is also proposed that is used to search for predicate rules from datasets. The algorithm proposed extends the co-

herent rules mining technique to discover predicate rules with the help of predicate logic and inference rules. Experiments performed on the zoo, breast cancer, and car evaluation datasets showed the distinction of the proposed technique by achieving an accuracy of 93.33%, 93.02%, and 89.58%, respectively, in comparison with other classification techniques. The proposed algorithm has the characteristics of finding predicate rules with rare consequences because of the selection of rules based on the comparison on reliability differences between the presence and absence of the rule’s consequence. The results showed that the rules discovered has higher coverage of the dataset and are useful based on their reliability. The main contribution of the proposed approach is the discovery of predicate rules without presetting the $minsupp$ threshold and finding interesting rules. In the future, the research will be focused on improving the completeness of the predicate rules discovered. Also, aspect of high dimensionality with its consequent data sparsity will be explored.

7. ACKNOWLEDGEMENTS

This research is supported by research grants (vot no:4F431, 13H89). The authors also acknowledge the support provided by the Ministry of Higher Education (MOHE), Research Management Centre (RMC) and Universiti Teknologi Malaysia.

References

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993*

- ACM SIGMOD international conference on Management of data*. pages 207–216.
- Alatas, B.** (2012). A novel chemistry based meta-heuristic optimization method for mining of classification rules. *Expert Systems with Applications*, 39(12): pages 11080–11088.
- Bache, K. & Lichman** (2013). Uci: Machine learning repository.
- Bagui, S. & Dhar, P. C.** (2018). Mining positive and negative association rules in hadoop’s mapreduce environment. In *Proceedings of the ACMSE 2018 Conference*. pages 1–1.
- Balakrishna, E., Rama, B., & Nagaraju, A.** (2019). Efficient mining of negative association rules using frequent item set mining. In *First International Conference on Artificial Intelligence and Cognitive Computing*. Springer, pages 709–716.
- Baralis, E., Cerquitelli, T., & Chiusano, S.** (2008). Imine: Index support for item set mining. *IEEE Transactions on Knowledge and data engineering*, 21(4): pages 493–506.
- Bemarisika, P. & Totohasina, A.** (2018). Erapn, an algorithm for extraction positive and negative association rules in big data. In *International Conference on Big Data Analytics and Knowledge Discovery*. Springer, pages 329–344.
- Bhatia, J. & Gupta, A.** (2014). Mining of quantitative association rules in agricultural data warehouse: A road map. *International Journal of Information Science and Intelligent System*, 3(1): pages 187–198.
- Bisht, S. & Samantaray, S.** (2015). Extracting spatial association rules in remotely sensed data of yellow rust disease in wheat crop at udham singh nagar. In *2015 1st International Conference on Next Generation Computing Technologies (NGCT)*. IEEE, pages 847–851.
- Borah, A. & Nath, B.** (2018). Identifying risk factors for adverse diseases using dynamic rare association rule mining. *Expert systems with applications*, 113: pages 233–263.
- Borgelt, C.** (2012). Frequent item set mining. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 2(6): pages 437–456.
- Buczak, A. L. & Gifford, C. M.** (2010). Fuzzy association rule mining for community crime pattern discovery. In *ACM SIGKDD workshop on intelligence and security informatics*. pages 1–10.
- Ceruto, T., Lapeira, O., Tonch, A., Plant, C., Espin, R., & Rosete, A.** (2014). Mining medical data to obtain fuzzy predicates. In *International Conference on Information Technology in Bio-and Medical Informatics*. Springer, pages 103–117.
- Dechang, P. & Xiaolin, Q.** (2008). A new fuzzy clustering algorithm on association rules for knowledge management. *Information Technology Journal*, 7(1): pages 119–124.
- Fournier-Viger, P., Cheng, C., Lin, J. C.-W., Yun, U., & Kiran, R. U.** (2019). Tkg: Efficient mining of top-k frequent subgraphs. In *International Conference on Big Data Analytics*. Springer, pages 209–226.
- Fournier-Viger, P., Lin, J. C.-W., Vo, B., Chi, T. T., Zhang, J., & Le, H. B.** (2017). A survey of itemset mining. *wires interdisciplinary reviews-data mining and knowledge discovery*.
- Geetha, M.** (2015). Implementation of association rule mining for different soil types in agriculture. *International Journal of Advanced Research in Computer and Communication Engineering*, 4(4): pages 520–522.
- Harahap, M., Husein, A., Aisyah, S., Lubis, F., & Wijaya, B.** (2018). Mining association rule based on the diseases population for recommendation of medicine need. In *Journal of Physics: Conference Series*, volume 1007. IOP Publishing, page 012017.
- Hasan, M. M. & Mishu, S. Z.** (2018). An adaptive method for mining frequent itemsets based on apriori and fp growth algorithm. In *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*. IEEE, pages 1–4.
- Hassani, H., Huang, X., Silva, E. S., & Ghodsi, M.** (2016). A review of data mining applications in crime. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 9(3): pages 139–154.
- Ikram, A. & Qamar, U.** (2015). Developing an expert system based on association rules and predicate logic for earthquake prediction. *Knowledge-Based Systems*, 75: pages 87–103.
- Jabbour, S., El Mazouri, F. E., & Sais, L.** (2018). Mining negatives association rules us-

- ing constraints. *Procedia Computer Science*, 127: pages 481–488.
- Koh, Y. S.** (2008). Mining non-coincidental rules without a user defined support threshold. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, pages 910–915.
- Kong, H., An, D., & Ri, J.** (2018). Itemsets of interest for negative association rules. arXiv preprint arXiv:1806.07084.
- Krishnapuram, B.** (2016). *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. Association for Computing Machinery.
- Li, J., Zhang, X., Dong, G., Ramamohanarao, K., & Sun, Q.** (1999). Efficient mining of high confidence association rules without support thresholds. In *European Conference on Principles of Data Mining and Knowledge Discovery*. Springer, pages 406–411.
- Moses, D. et al.** (2015). A survey of data mining algorithms used in cardiovascular disease diagnosis from multi-lead ecg data. *Kuwait Journal of Science*, 42(2).
- Mustafa, M. D., Nabila, N. F., Evans, D. J., Saman, M. Y., & Mamat, A.** (2006). Association rules on significant rare data using second support. *International Journal of Computer Mathematics*, 83(1): pages 69–80.
- Parfait, B., Harrimann, R., & André, T.** (2018). An efficient approach for extraction positive and negative association rules from big data. In *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. Springer, pages 79–97.
- Qin, Z., Ren, K., Yu, T., & Weng, J.** (2016). Dp-code: Privacy-preserving frequent visual patterns publication on cloud. *IEEE Transactions on Multimedia*, 18(5): pages 929–939.
- Rosen, K. H. & Krithivasan, K.** (2012). *Discrete mathematics and its applications: with combinatorics and graph theory*. Tata McGraw-Hill Education.
- Salah, S., Akbarinia, R., & Masegla, F.** (2017). A highly scalable parallel algorithm for maximally informative k-itemset mining. *Knowledge and Information Systems*, 50(1): pages 1–26.
- Salam, A. & Khayal, M. S. H.** (2012). Mining top- k frequent patterns without minimum support threshold. *Knowledge and information systems*, 30(1): pages 57–86.
- Sim, A. T. H., Indrawan, M., Zutshi, S., & Srinivasan, B.** (2010). Logic-based pattern discovery. *IEEE Transactions on Knowledge and Data Engineering*, 22(6): pages 798–811.
- Usman, M.** (2017). Multi-level mining of association rules from warehouse schema. *Kuwait Journal of Science*, 44(1).
- Zhao, L., Hao, F., Xu, T., & Dong, X.** (2017). Positive and negative association rules mining for mental health analysis of college students. *EURASIA Journal of Mathematics, Science and Technology Education*, 13(8): pages 5577–5587.
- Zuo, J., Tang, C., & Zhang, T.** (2002). Mining predicate association rule by gene expression programming. In *International Conference on Web-Age Information Management*. Springer, pages 92–103.

Submitted: 20/05/2020

Revised: 20/12/2020

Accepted: 20/12/2020

DOI: 10.48129/kjs.v48i4.9782