

MILAR: Mining Indirect Least Association Rule Algorithm

Zailani Abdullah¹, Aggy Pramana Gusman², Tutut Herawan³, and Mustafa Mat Deris⁴

¹School of Informatics and Applied Mathematics
Universiti Malaysia Terengganu
21030 Kuala Terengganu, Terengganu, Malaysia

²Universitas Putra Indonesia YPTK
Padang, Sumatera Barat, Indonesia

³Faculty of Computer Science & Information Technology
University of Malaya,

50603 Pantai Valley, Kuala Lumpur, Malaysia

⁴Faculty of Computer Science & Information Technology
University Tun Hussein Onn Malaysia,
86400 Batu Pahat, Johor

zailania@umt.edu.my, aggy_gusman@upiyptk.ac.id, tutut@um.edu.my, mmustafa@uthm.edu.my

Received 25 February 2016; Revised 27 March 2016; Accepted 20 June 2016

Abstract

One of the interesting and meaningful information that is hiding in transactional database is indirect association rule. It corresponds to the property of high dependencies between two items that are rarely occurred together but indirectly emerged via another items. Since indirect association rule is nontrivial information, it can implicitly give a new perspective of relationship which cannot be directly observed from the common rule. Therefore, we proposed an algorithm for Mining Indirect Least Association Rule (MILAR) from the real and benchmarked datasets. MILAR is embedded with our scalable least measure namely Critical Relative Support (CRS). The experimental results show that MILAR can generate the desired rules in term of least and indirect least association rules. In addition, the obtained results can also be used by the domain experts to do further analysis and finally reveal more interesting findings.

Keywords: Data mining; Association rule; Indirect; Least; Algorithm.

1. Introduction

The field of data mining is still relatively new and at evolution stage. It stands at the converging of the field of machine learning and statistics. Until this recent, data mining has been successfully employed in variety of domain applications. In simple definition, data mining is the process of extracting some new nontrivial information from large data repository. For more comprehensive definition, it is about making analysis convenient, scaling analysis algorithms to large databases and providing data owners with easy to use tools in helping the user to navigate, visualize, summarize and model the data [1]. In summary, the ultimate goal of data mining is more towards knowledge discovery. The core models in data mining including association rule, prediction, predictive analysis, data reduction, data exploration and data visualization [2]. Association rule was first introduced by Agrawal *et al.* [3] in 1993, and since then several extensively studies have been made [4-12]. It is unsupervised learning and also known as affinity analysis. In marketing, association rule is prevalently known as market basket analysis in an attempt to extract the group of products that tend to be purchased together. Generally, it aims at discovering the interesting relationship between a set of items that are frequently occurred together in transactional database

[13]. Apriori is among one of the most earlier and popular algorithms in mining and judging the strength of the rules. However the drawback of this concept is, infrequent or least items are automatically considered as not important and pruned out during the rules generation. In spite of this and especially in certain domain applications, least items may also provide a new and useful insight about the data such as competitive product analysis [14], text mining [15], web recommendation [16], biomedical analysis [17], and etc. The classical association rule which is derived from the frequent items is also popularly known as direct association rule. The contradiction of this rule namely indirect association rule [18] which refers to a pair of items that are rarely occurred together but their existences are highly depending on the presence of mediator itemsets. It was first proposed by Tan *et al.* [14] for interpreting the value of infrequent patterns and effectively pruning out the uninteresting infrequent patterns. Recently, the problem of indirect association mining has become more and more important because of its unique contributions in various domain applications [18-22]. From the literature, the studies on indirect association mining can be divided into two categories, either focusing on proposing more efficient mining algorithms [15,18,22] or extending the definition of the indirect association for different domain applications [6,18,19].

The process of discovering indirect association rule is a nontrivial and usually relies more on the existing interesting measures that has been detailed in [14]. However, most of the measures are not properly evaluated in term of the least association rule. Therefore, in this paper we propose Mining Indirect Least Association Rule (MILAR) algorithm by utilizing the strength of Least Pattern Tree (LP-Tree) data structure [11]. In addition, Critical Relative Support (CRS) measure [23] is also embedded in the algorithm to mine the indirect least association rules among the least rules. Indeed, CRS has been widely employed in measuring the least association rule [23, 27-35].

The rest of the paper is organized as follows. Section 2 describes the related work. Section 3 explains the proposed method. This is followed by performance analysis through two experiment tests in section 4. Finally, conclusion and future direction are reported in section 5.

2. Related Works

Generally, indirect association is closely related to negative association, they are both dealing with itemsets that do not have sufficiently high in term of support. The negative associations' rule was first pointed out by Brin *et al.* [24]. The focused on mining negative associations is better than on finding the itemsets that have a very low probability of occurring together. Indirect associations provide an effective way to detect interesting negative associations by discovering only frequent itempairs that are highly expected to be frequent. Until this recent, the important of indirect association between items has been discussed in many literatures. Tan *et al.* [14] proposed INDIRECT algorithm to extract indirect association between itempairs using the famous Apriori technique. There are two main steps involved. First, extract all frequent items using standard frequent pattern mining algorithm. Second, find the valid indirect associations from the candidate indirect association from candidate itemsets. Wan *et al.* [15] introduced HI-Mine algorithm to mine a complete set of indirect associations. HI-Mine generates indirect itempair set (IIS) and mediator support set (MSS), by recursively building the HI-struct from database. The performance of this algorithm is significantly better than the previously developed algorithm either for synthetic or real datasets. IS measure [25] is used as a dependence measure. Lin *et al.* [26] proposed GIAMS as an algorithm to mine indirect associations over data streams rather than static database environment. GIAMS contains two concurrent processes called PA-Monitoring and IA-Generation. The first process is to set off when the users specify the required window parameters. The second process is activated once the users issues queries about current indirect associations. In term of dependence measure, IS measure [25] is again adopted in the algorithm. Chen *et al.* [18] proposed an indirect association algorithm that was similar to HI-mine, namely MG-Growth. The aim of MG-Growth is to discover indirect association patterns and its extended version is to extract temporal indirect association patterns. The differences between both algorithms are, MG-Growth used the directed graph and bitmap to construct the indirect itempair set. The corresponding

mediator graphs are then generated for deriving a complete set of indirect associations. In this algorithm, temporal support and temporal dependence are used in this algorithm.

Kazienko [16] presented IDARM* algorithm to extract complete indirect association rules. In this algorithm, both direct and indirect rules are joined together to form a useful set of indirect rules. Two types of indirect associations are proposed named partial indirect association and complete ones. The main idea of IDARM* is to capture the transitive path from user-session as part of web recommendation system. A simple measure called Confidence [3] is employed as dependence measure. Lin *et al.* [36] introduced EMIA-LM algorithm for mining indirect association rules over web data stream. EMIA-LM uses a mediator-exploiting search strategy in the process of generating the rule. It also adopts a compact data structure, alleviates unnecessary data transformation processes and minimizes the usage of memory. The preliminary experiments also showed that EMIA-LM is better than HI-mine* for static data in terms of computational speed and memory consumption. Liu *et al.* [37] suggested FIARM (Filtering-Based Indirect Association Rule Mining) algorithm to analyze gene microarray data. It is an Apriori-based algorithm. The algorithm can determine indirect gene associations to assist biologists in finding new insightful knowledge. FIARM-Measure is also introduced to help in discovering indirect association rules from the rules that have a negative correlation. In the analysis, Gene Ontology is employed to verify the accuracy of the relationships. Hajian and Domingo-Ferrer [38] proposed a new technique for mining direct and indirect discrimination prevention of rules. There are two phases involved; discrimination measurement and data transformation. Discrimination measurement determines alpha-discriminatory rules and also redlining rules. In data transformation, the original data will be transformed by removing direct and/or indirect discriminatory biases. This process ensures that no unfair decision rule can be mined from the transformed data.

3. Proposed Method

3.1. Association Rule

Throughout this section the set $I = \{i_1, i_2, \dots, i_{|A|}\}$, for $|A| > 0$ refers to the set of literals called set of items and the set $D = \{t_1, t_2, \dots, t_{|U|}\}$, for $|U| > 0$ refers to the data set of transactions, where each transaction $t \in D$ is a list of distinct items $t = \{i_1, i_2, \dots, i_{|M|}\}$, $1 \leq |M| \leq |A|$ and each transaction can be identified by a distinct identifier TID.

Definition 1. A set $X \subseteq I$ is called an itemset. An itemset with k -items is called a k -itemset.

Definition 2. The support of an itemset $X \subseteq I$, denoted $\text{supp}(X)$ is defined as a number of transactions contain X .

Definition 3. Let $X, Y \subseteq I$ be itemset. An association rule between sets X and Y is an implication of the form $X \Rightarrow Y$, where $X \cap Y = \emptyset$. The sets X and Y are called antecedent and consequent, respectively.

Definition 4. The support for an association rule $X \Rightarrow Y$, denoted $\text{supp}(X \Rightarrow Y)$, is defined as a number of transactions in D contain $X \cup Y$.

Definition 5. The confidence for an association rule $X \Rightarrow Y$, denoted $\text{conf}(X \Rightarrow Y)$ is defined as a ratio of the numbers of transactions in D contain $X \cup Y$ to the number of transactions in D contain X . Thus

$$\text{conf}(X \Rightarrow Y) = \frac{\text{supp}(X \Rightarrow Y)}{\text{supp}(X)}$$

Definition 6. (Least Items). An itemset X is called least item if $\text{supp}(X) < \alpha$, where α is the minimum support (minsupp)

The set of least item will be denoted as Least Items and

$$\text{Least Items} = \{X \subset I \mid \text{supp}(X) < \alpha\}$$

Definition 7. (Frequent Items). An itemset X is called frequent item if $\text{supp}(X) \geq \alpha$, where α is the minimum support.

The set of frequent item will be denoted as Frequent Items and

$$\text{Frequent Items} = \{X \subset I \mid \text{supp}(X) \geq \alpha\}$$

3.2. Indirect Association Rule

Definition 8. An item-pair $\{X, Y\}$ is indirectly associated via a mediator M , if the following conditions are fulfilled:

- $\text{supp}(\{X, Y\}) < t_s$ (itempair support condition)
- There exists a non-empty set M such that:
- $\text{supp}(\{X\} \cup M) \geq t_m$ and $\text{supp}(\{Y\} \cup M) \geq t_m$ (mediator support condition)
- $\text{dep}(\{X\}, M) \geq t_d$ and $\text{dep}(\{Y\}, M) \geq t_d$, where $\text{dep}(A, M)$ is a measure of dependence between itemset A and M (mediator dependence measure)

The user-defined thresholds above are known as itempair support threshold (t_s), mediator support threshold (t_m) and mediator dependence threshold (t_d), respectively. The itempair support threshold is equivalent to $\text{minsupp}(\alpha)$. Normally, the mediator support condition is set to equal or more than the itempair support condition ($t_m \geq t_s$)

The first condition is to ensure that (X, Y) is rarely occurred together and also known as least or infrequent items. In the second condition, the first-sub-condition is to capture the mediator M and for the second-sub-condition is to make sure that X and Y are highly dependence to form a set of mediator.

Definition 9. (Critical Relative Support). A *Critical Relative Support (CRS)* is a formulation of maximizing relative frequency between itemset and their Jaccard similarity coefficient.

The value of Critical Relative Support denoted as CRS and

$$\text{CRS}(A, B) = \max\left(\left(\frac{\text{supp}(A)}{\text{supp}(B)}\right), \left(\frac{\text{supp}(B)}{\text{supp}(A)}\right)\right) \times \left(\frac{\text{supp}(A \Rightarrow B)}{\text{supp}(A) + \text{supp}(B) - \text{supp}(A \Rightarrow B)}\right)$$

CRS value is between 0 and 1, and is determined by multiplying the highest value either supports of antecedent divide by consequence or in another way around with their Jaccard similarity coefficient. It is a measurement to show the level of CRS between combination of the both Least Items and Frequent Items either as antecedent or consequence, respectively. Here, Critical Relative Support (CRS) is employed as a dependence measure for 2(a) in order to mine the desired Indirect Association Rule.

3.3. Algorithm Development

Determine Minimum Support. Let I is a non-empty set such that $I = \{i_1, i_2, \dots, i_n\}$, and D is a database of transactions where each T is a set of items such that $T \subset I$. An itemset is a set of item. A k -itemset is an itemset that contains k items. From Definition 6, an itemset is said to be least (infrequent) if it has a support count less than α .

Construct LP-Tree. A Least Pattern Tree (LP-Tree) is a compressed representation of the least itemset. It is constructed by scanning the dataset of single transaction at a time and then mapping onto a new or existing path in the LP-Tree. Items that satisfy the α (Definition 6 and 7) are only captured and used in constructing the LP-Tree.

Mining LP-Tree. Once the LP-Tree is fully constructed, the mining process will begin using bottom-up strategy. Hybrid ‘Divide and conquer’ method is employed to decompose the tasks of mining desired pattern. LP-Tree utilizes the strength of hash-based method during constructing itemset in support descending order.

Construct Indirect Patterns. The pattern is classified as indirect association pattern if it fulfilled the two conditions. The first condition is elaborated in Definition 8 where there are three sub-conditions. One of them is mediator dependence measure. CRS from Definition 9 is employed as mediator dependence measure between itemset in discovering the indirect patterns.

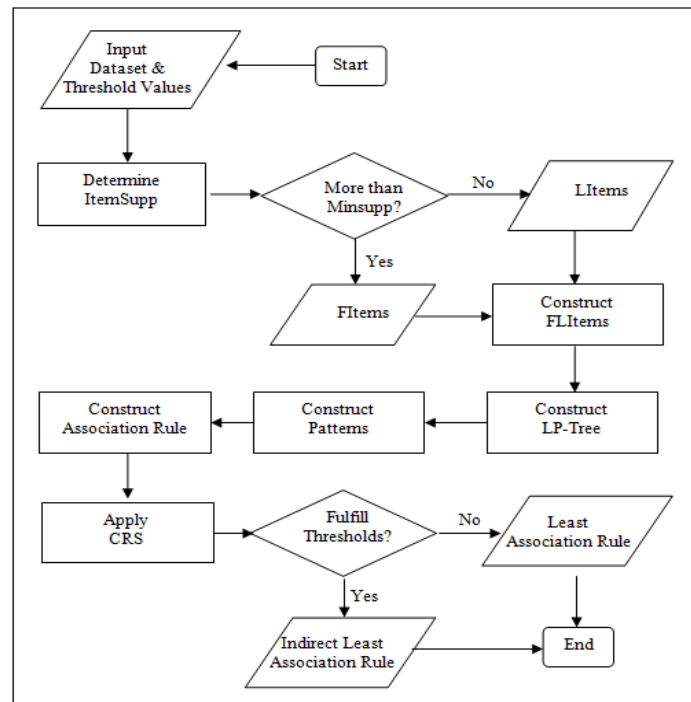


Figure 1: A Complete Graphical Representation of MILAR Algorithm

Figure 1 shows a complete graphical representation to construct the Mining Indirect Least Association Rule Algorithm (MILAR).

4. Experiment Test

In this section, the analysis is made by comparing the total number of association rules being extracted based on the predefined thresholds using our proposed algorithm, MILAR. In the analysis, three items are involved in forming a complete association rule; two items as an

antecedent and one item as a consequence. The mediator is appeared as a part of antecedent. We conducted our experiment using two datasets. The experiment has been performed on Intel® Core™ 2 Quad CPU at 2.33GHz speed with 4GB main memory, running on Microsoft Windows Vista. All algorithms have been developed using C# as a programming language.

The first dataset is language anxiety dataset. The dataset was taken from a survey on exploring language anxiety among engineering students at Universiti Malaysia Pahang (UMP). The respondents were 770 students, consisting of 394 males and 376 females. They are undergraduate students from five engineering based faculties, i.e., 216 students from Faculty of Chemical and Natural Resources Engineering (FCNRE), 105 students from Faculty of Electrical and Electronic Engineering (FEEE), 226 students from Faculty of Mechanical Engineering (FME), 178 students from Faculty of Civil Engineering and Earth Resources (FCEER), and 45 students from Faculty of Manufacturing Engineering and Technology Management (FMETM). To this, we have a dataset comprises the number of transactions (student) is 770 and the number of items (attributes) is 5. Table 1 displays the mapped of original survey dimensions, Likert scale with a new attribute id.

Table 1: The mapping between survey dimensions of language anxiety, Likert scale and a new attribute Id

Survey Dimensions	Likert Scale	Attribute Id
Anxious in the language class	1 – 5	1
Time Speaking	1 – 5	2
Someone correction your mistake	1 – 5	3
Felt the lecturer like native speaker	1 – 5	4
Lack of confidence	1 – 5	5

Item is constructed based on the combination of survey dimension and its likert scale. For simplicity, let consider a survey dimension “Anxious in the language class” with likert scale “1”. Here, an item “11” will be constructed by means of a combination of an attribute id (first characters) and its survey dimension (second character).

Figure 2 shows the performance analysis against Language Anxiety dataset. Minimum Support (*minsupp* or α) and Mediator Support Threshold (t_m) are set to 30% and 10%, respectively.

Varieties of minimum CRS (min-CRS) were employed in the experiment. During the performance analysis, 286 least association rules and 152 indirect least association rules were produced, respectively. The general trend was, the total number of indirect least association rules were kept reducing when the values of min-CRS were kept increasing. However, there are no changes in term of total least association rules and indirect least association rules when the min-CRS values were in the range of 0.15 until 0.20.

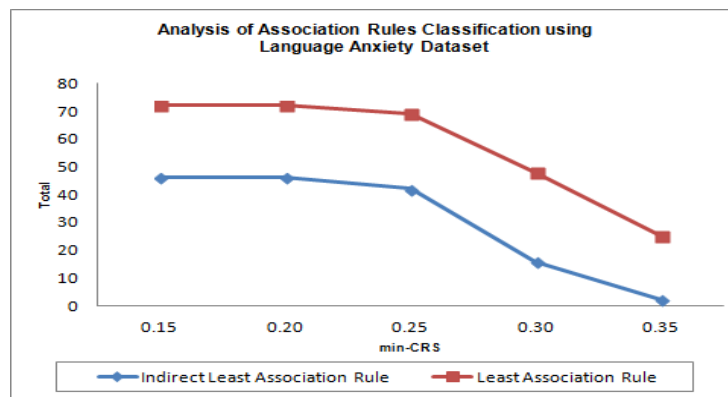


Figure 2: Analysis of the Generated Association Rules against Language Anxiety Dataset with variation of min-CRS

The second experiment was conducted on Breast-Cancer-Wisconsin dataset [39]. The aim of the dataset is to diagnose the breast cancer according to Fine- Needle Aspirates (FNA) test. The dataset was obtained from a repository of a machine-learning database University of California, Irvin. It was compiled by Dr. William H. Wolberg from University of Wisconsin Hospitals, Madison, Madison, WI, United States. It has 11 attributes and 699 records (as of 15 July 1992) with 158 benign and 241 malignant classes, respectively. Table 2 displays the mapped of original attributes with new attributes id. Item is constructed based on the combination of attribute id and its domain. For simplicity, let consider an attribute ‘‘Clump Thickness’’ with domain ‘‘1’’. Here, an item ‘‘101’’ will be constructed by means of a combination of an attribute id (first two characters) and its domain (third character).

Table 2: The mapped breast cancer data

Attributes	Domain	Attributes Id
Sample code number	Id number	Ignored
Clump Thickness	1 - 10	10
Uniformity of Cell Size	1 - 10	20
Uniformity of Cell Shape	1 - 10	30
Marginal Adhesion	1 - 10	40
Single Epithelial Cell Size	1 - 10	50
Bare Nuclei	1 - 10	60
Bland Chromatin	1 - 10	70
Normal Nucleoli	1 - 10	80
Mitoses	1 - 10	90
Class	2,4	111

Figure 3 shows the performance analysis of benchmarked Wisconsin Breast Cancer dataset. The Minimum Support (*minsupp* or α) and Mediator Support Threshold (t_m) are fixed similar as used by the first experiment. Various minimum CRS (min-CRS) were also utilized during the experiment. From the analysis, 416 least association rules and 150 indirect least association rules were generated, respectively. The total number of indirect least association rules were kept reducing when the values of min-CRS were kept increasing.

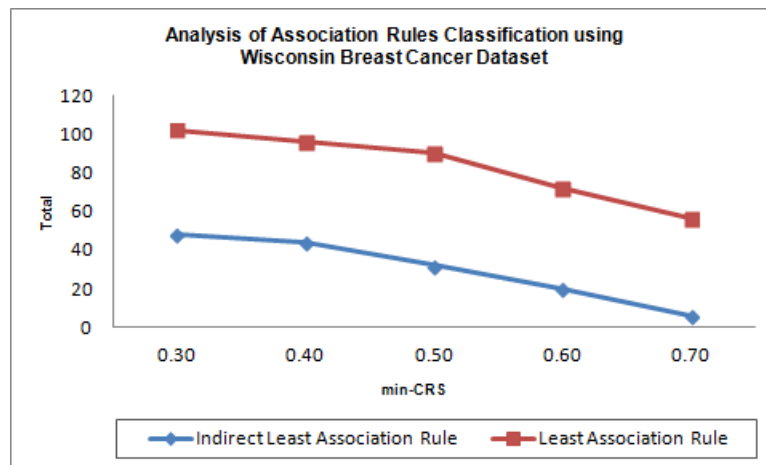


Figure 3: Analysis of the Generated Association Rules against Wisconsin Breast Cancer Dataset with variation of min-CRS

The third experiment was conducted on Mushroom-edible dataset [39]. Originally, Mushroom is a dense dataset and consists of 23 species of gilled mushroom in the *Agaricus* and *Lepiota* Family. Each species is identified as definitely edible, definitely poisonous, or of unknown edibility and not recommended. For simplicity, only two classes’ either edible or poisonous are maintained.

The total number of instances is 8124 data with 22 attributes. From these instances, 4208 (51.8%) is classified as edible and the balance is classified as poisonous. In this experiment, only edible class is considered. Table 3 shows the mapped Mushroom-edible dataset. The detail domain names for each attribute are not shown. For example, attribute “cap-shape” can be classified as either bell or conical or convex or flat or knobbed or sunken. Since the “cap-shape” can be fallen into one of the six categories, thus its domain is in the range of 1 into 6 (1-6) which 1 is represented bell, 2 is correspond to conical and so forth. In Mushroom-edible dataset, item is constructed based on the combination of attribute id and its domain. For simplicity, let consider the attribute “cap-shape”, with domain “1”. Here, an item “11” will be constructed by means of a combination of an attribute id (first characters) and its domain (second character or after first characters).

Table 3. The mapped Mushroom-edible data

Attributes	Domain	Attributes Id
cap-shape	1-6	1
cap-surface	1-8	2
cap-color	1-10	3
bruises	1-2	4
odor	1-9	5
gill-attachment:	1-4	6
gill-spacing	1-3	7
gill-size	1-2	8
gill-color	1-12	9
stalk-shape	1-2	10
stalk-root	1-7	11
stalk-surface-above-ring	1-4	12
stalk-surface-below-ring	1-4	13
stalk-color-above-ring	1-9	14
stalk-color-below-ring	1-9	15
veil-type	1-2	16
veil-color	1-4	17
ring-number	1-3	18
ring-type	1-8	19
spore-print-color	1-9	20
population	1-6	21
habitat	1-7	22

Figure 4 shows the performance analysis of benchmarked Mushroom-edible dataset. The Minimum Support (*minsupp* or α) and Mediator Support Threshold (t_m) are fixed into 40% and 20%, respectively. Various minimum CRS (min-CRS) were also utilized during the experiment. From the analysis, 651 least association rules and 298 indirect least association rules are generated, respectively. The total number of indirect least association rules are kept reducing when the values of min-CRS were kept increasing.

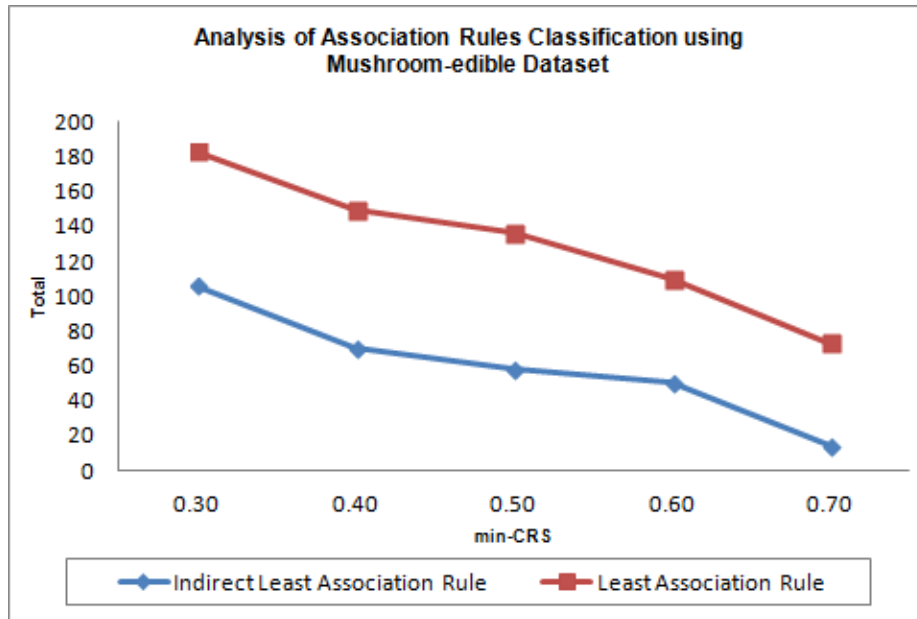


Figure 4. Analysis of the Generated Association Rules against Mushroom-edible Dataset with variation of min-CRS

5. Conclusion

Mining indirect least association rules from data repository is a nontrivial study and very crucial. It has been specially designed to deal with the rarity cases. In fact, it may contribute in discovering a new knowledge which cannot be easily obtained through typical association rules approaches. Indirect least association rule is by definition represents the property of high dependencies between two items that are rarely occurred together in the same transaction but actually appeared indirectly via another itemset. Therefore, in this paper we proposed Mining Indirect Least Association Rule (MILAR) algorithm to extract the hidden and interesting rules called indirect least association rules from the data repository. MILAR algorithm embeds with a scalable measure called Critical Relative Support (CRS) rather than the common interestingness measures in data mining. We conducted the three experiments based on real dataset and benchmarked datasets. The obtained results show that MILAR algorithm can successfully generate the least and indirect least association rules based. It is expected that obtained information can provide a new insight for domain experts to do further investigation and finally discover a new knowledge. In the near future, we plan to apply MILAR algorithm into several benchmarked datasets and real datasets.

Acknowledgement

This research is supported by AMCS Research Center, Indonesia.

References

1. Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P.: From Data Mining to Knowledge Discovery in Databases, pp. 37-54 (1996)
2. Shmueli, G., Patel, N.R., and Bruce, P.C.: Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner. John Wiley & Sons, Inc., Publications (2010)
3. Agrawal, R., & Srikant, R.: Fast Algorithms for Mining Association Rules in Large Databases. In: Proceedings of the 20th International Conference on Very Large Data Bases, pp. 487-499 (1994).
4. Mannila, H., Toivonen, H., and Verkamo, A.I.: Discovery of Frequent Episodes in Event Sequences. Data Mining and Knowledge Discovery, 1, 259–289 (1997)

5. Park, J.S., Chen, M.S., and Yu, P.S.: An Effective Hash-based Algorithm for Mining Association Rules. In: Proceedings of the ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'95), pp. 175–186. ACM Press (1995)
6. Savasere, A., Omiecinski, E., and Navathe, S.: An efficient algorithm for mining association rules in large databases. In: Proceedings of the 21st International. Conference on Very Large Data Bases (VLDB'95), pp. 432–443. ACM Press (1995).
7. Fayyad, U., Patesesky-Shapiro, G., Smyth, P. and Uthurusamy, R.: Advances in Knowledge Discovery and Data Mining. MIT Press, MA (1996).
8. Bayardo, R.J.: Efficiently Mining Long Patterns from Databases. In: Proceedings of the ACM-SIGMOD International Conference on Management of Data (SIGMOD'98), pp. 85–93. ACM Press (1998)
9. Zaki, M.J. and Hsiao, C.J.: CHARM: An efficient algorithm for closed itemset mining. In: Proceedings of the 2002 SIAM Intl. Conf. Data Mining, pp. 457–473. SIAM (2002).
10. Agarwal, R., Aggarwal, C., and Prasad, V.V.V.: A tree projection algorithm for generation of frequent itemsets. *Journal of Parallel and Distributed Computing*, 61, 350–371 (2001)
11. Liu, B., Hsu, W. and Ma, Y.: Mining Association Rules with Multiple Minimum Support. In: Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.337 – 341. ACM Press (1999).
12. Abdullah, Z., Herawan, T. and Deris, M.M.: Scalable Model for Mining Critical Least Association Rules. In Rongbo Zhu *et al.* ICICA 2010, LNCS 6377, pp. 509-516. Springer Heidelberg (2010).
13. Leung, C.W., Chan S.C and Chung, F.: An Empirical Study of a Cross-level Association Rule Mining Approach to Cold-start Recommendations. *Knowledge-Based Systems*, 21(7), October 2008, 515–529 (2008)
14. Tan, P.N., Kumar, V. and Srivastava, J.: Indirect Association: Mining Higher Order Dependences in Data. In: Proceedings of the 4th European Conference on Principles and Practice of Knowledge Discovery in Databases, pp. 632-637. Springer Heidelberg (2000).
15. Wan, Q. and An, A.: An Efficient Approach to Mining Indirect Associations. *Journal Intelligent Information Systems*, 27(2), 135–158 (2006)
16. Kazienko, P.: Mining Indirect Association Rules for Web Recommendation. *International Journal of Applied Mathematics and Computer Science*, 19(1), pp. 165–186 (2009)
17. Tsuruoka, Y., Miwa, M., Hamamoto, K., Tsujii, J. and Ananiadou, S.: Discovering and Visualizing Indirect Associations between Biomedical Concepts. *Bioinformatics*, 27(13), pp. 111-119 (2011).
18. Chen, L., Bhowmick, S.S. and Li, J.: Mining Temporal Indirect Associations. In: PAKDD2006, LNAI 3918, pp. 425-434. Springer Heidelberg (2006)
19. Cornelis, C., Yan, P., Zhang, X. and Chen, G.: Mining Positive and Negative Association from Large Databases. In: Proceedings of the 2006 IEEE International Conference on Cybernetics and Intelligent systems, pp. 1-6. IEEE (2006)
20. Kazienko, P. and Kuzminska, K.: The Influence of Indirect Association Rules on Recommendation Ranking Lists. In: Proceeding of the 5th International Conference on Intelligent Systems Design and Applications, pp. 482-487 (2005)
21. Tseng, V.s, Liu, Y.C and Shin J.W.: Mining Gene Expression Data with Indirect Association Rules. In: Proceeding of the 2007 National Computer Symposium (2007)
22. Wu, X., Zhang, C. and Zhang, S. Efficient Mining of Positive and Negative Association Rules. *ACM Transaction on Information Systems*, 22(3), 381-405 (2004)
23. Abdullah, Z., Herawan, T., Noraziah, A. and Deris, M.M.: Mining Significant Association Rules from Educational Data using Critical Relative Support Approach. *Procedia Social and Behavioral Sciences*, 2011, 28, 97 – 191. Elsevier (2011)
24. Brin, S., Motwani, R., Ullman, J., and Tsur, S.: Dynamic itemset counting and implication rules for market basket data. In: Proceedings of the International ACM SIGMOD Conference, pp. 255–264. ACM Press (1997).
25. Tan, P., Kumar, V., and Srivastava, J.: Selecting the Right Interestingness Measure for Association Patterns. In: Proceedings of the 8th International Conference on Knowledge Discovery and Data Mining, pp.32-41. (2002).

26. Lin, W-Y., Wei, Y-E., and Chen, C-H. Generic Approach for Mining Indirect Association Rules in Data Streams. LNCS 6703, pp. 95-104. Springer Heidelberg. (2011).
27. Herawan, T., Vitasari, P., and Abdullah, Z.: Mining Interesting Association Rules of Students Suffering Study Anxieties Using SLP-Growth Algorithm. International Journal of Knowledge and Systems Sciences, 3(2), pp. 24-41 (2012)
28. Abdullah, Z., Herawan, T. Noraziah, A. and Deris, M.M.: Detecting Critical Least Association Rules in Medical Databases. International Journal of Modern Physics: Conference Series, World Scientific, 9, pp. v464-479 (2012)
29. Herawan, T. and Abdullah, Z.: CNAR-M: A Model for Mining Critical Negative Association Rules. In Zhihua Cai *et al.* (Eds): ISICA 2012, Communications in Computer and Information Science, Springer-Verlag, 316, pp. 170-179 (2012)
30. Abdullah, Z., Herawan, T. Noraziah, A. and Deris, M.M and Abawajy, J.H.: IPMA: Indirect Patterns Mining Algorithm. In N.T. Nguyen *et al.* (Eds.): ICCCI 2012, Advanced Methods for Computational Collective Intelligence Studies in Computational Intelligence, Springer-Verlag, 457, pp. 187-196 (2012).
31. Herawan, T., P. Vitasari, and Abdullah, Z.: Mining interesting association rules of student suffering mathematics anxiety. In J.M. Zain *et al.* (Eds.): ICSECS 2011, Communication of Computer and Information Sciences, Springer-Verlag, 188, II, 495-508 (2011)
32. Abdullah, Z., Herawan, T. and Deris, M.M.: Efficient and Scalable Model for Mining Critical Least Association Rules. In a special issue from AST/UCMA/ISA/ACN 2010, Journal of The Chinese Institute of Engineer, Taylor and Francis, 35, No. 4, 27 June 2012, pp. 547-554 (2012)
33. Abdullah, Z., Herawan, T. Noraziah, A. and Deris, M.M.: Extracting Highly Positive Association Rules from Students' Enrollment Data. Procedia Social and Behavioral Sciences, 28, 107-111 (2011)
34. Abdullah, Z., Herawan, T. Noraziah, A. and Deris, M.M.: Mining Significant Association Rules from Educational Data using Critical Relative Support Approach. Procedia Social and Behavioral Sciences, 28, pp. 97-101 (2011)
35. Abdullah, Z., Herawan, T. and Deris, M.M.: An Alternative Measure for Mining Weighted Least Association Rule and Its Framework. In J.M. Zain *et al.* (Eds.): ICSECS 2011, Communication of Computer and Information Sciences, Springer-Verlag, 188, II, pp. 475-485 (2011)
36. Lin, W-Y., and Chen, Y-C.: A Mediator Exploiting Approach for Mining Indirect Associations from Web Data Streams. IEEEExplore in the 2nd Intl. Conference on Innovations in Bio-inspired Computing and Applications (IBICA) 2011, pp. 183-186 (2011)
37. Liu, Y-C., Shin, J.W., and Tseng, V.S.: Discovering Indirect Gene Associations by Filtering-Based Indirect Association Rule Mining, International Journal of Innovative Computing, Information and Control, 7(10), pp. 6041-6053 (2011)
38. Hajian, S., and Domingo-Ferrer, J.: A Methodology for Direct and Indirect Discrimination Prevention in Data Mining. IEEE Transactions on Knowledge and Data Engineering, 25(7), pp. 1445-1459 (2013)
39. UCI Machine Learning Repository archive.ics.uci.edu/ml/datasets/