International Journal of Multimedia and Ubiquitous Engineering
Vol.9, No.1 (2014), pp.241-254
http://dx.doi.org/10.14257/ijmue.2014.9.1.23

Mining Least Association Rules of Degree Level Programs Selected by Students

Zailani Abdullah[1], Tutut Herawan[2], A. Noraziah[3] and Mustafa Mat Deris[4]

[1]Department of Computer Science, Universiti Malaysia Terengganu
21030 Kuala Terengganu, Terengganu, Malaysia
[2]Department of Mathematics Education, Universitas Ahmad Dahlan
Jalan 55166, Yogyakarta, Indonesia
[3]Faculty of Computer Systems and Software Engineering, Universiti Malaysia Pahang
Lebuhraya Tun Razak, 26300 Kuantan Pahang, Malaysia
[4]Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat 86400, Johor, Malaysia

zailania@umt.edu.my, tutut81@uad.ac.id, noraziah@ump.edu.my, mmustafa@uthm.edu.my

Abstract

One of the most popular and important studies in data mining is association rules mining. Generally, association rules can be divided into two categories called frequent and least. However, finding the least association rules is more complex and time consuming as compared to the frequent one. These rules are very useful in certain application domain such as determining the exceptional association between university's programs being selected by students. Therefore in this paper, we apply our novel measure called Definite Factors (DF) to determine the significant least association rules from undergraduate's program selection database. The dataset of computer science student for July 2008/2009 intake from Universiti Malaysia Terengganu was employed in the experiment. The result shows that our measurement can mine these rules and it is at par with the existing benchmarked Relative Support Apriori (RSA) measurement.

Keywords: *Data Mining; Association rules; Significant Least, Measure; Educational Data*

1. Introduction

Until this recent, the studies in data mining are kept growing and moved forward. In some literatures, data mining and knowledge discovery in database (KDD) have been widely used in a similar context. However, the more accurate definition is data mining is existed as an integral part of the KDD process. Data mining can be defined as the process of extracting hidden and useful information from large data repositories (Tan, *et al.*, 2006). One of the emerging interdisciplinary research areas in data mining is educational data mining (Baker & Yacef, 2009). By definition, educational data mining is an application of suitable data mining techniques to analyze the educational data (Romero, *et al.*, 2008). It aims at developing new methods that can discover the interesting information from educational settings, and used those methods to better understand the students, and their learning settings

ISSN: 1975-0080 IJMUE

(http://www.educationaldatamining.org). The problem of association rules mining was first coined by (Agrawal, *et al.*, 1993) in an attempt for market-basket analysis. There are two main steps involved in producing the rules. First, is to find all frequent items from transactional database. Second, is to generate association rules from the previous frequent items. The classification of frequent or least items is based on the mechanism of support threshold. A set of items (itemset) is said to be frequent, if it appears more than minimum support count. The item (or itemset) support count is defined as a probability of item (or itemset) appears in the transaction. In addition, confidence is another measure that always used together with support count. The confidence is defined as the probability of the rule's consequent (right side) that also contain the antecedent (left side) in the transaction. The association rule is said to be strong if it meets the minimum confidence threshold.

Least itemset is a set of item that is rarely occurred in the transactional database. It is also known as non-frequent, unusual, exceptional, abnormal, in-balance or sporadic itemset. In some applications domain, these itemsets are very important and in fact it can provide significant information such as air pollution level (Mustafa, *et al.*, 2006), customer relationship management (Au & Chan, 2003), image processing (Aggarwal & Yu, 1998), abnormal learning problems (Romero, *et al.*, 2010), educational data mining (Abdullah, *et al.*, 2011a; Abdullah, *et al.*, 2011b; Herawan, *et al.*, 2011), text mining (Herawan, *et al.*, 2009a; Herawan & Deris, 2011), information visualization (Herawan, *et al.*, 2009b; Abdullah, *et al.*, 2011c), business process management (Huang, *et al.*, 2011) and many more. From the past literature, most of the tradition association rules mining algorithms (Kiran, *et al.*, 2009; Zhou & Yau, 2007; Koh & Rountree, 2005; Yun, *et al.*, 2003; Liu, *et al.*, 1999; Wang, *et al.*, 2003; Tao, *et al.*, 2003; Ding, 2005; Abdullah, *et al.*, 2010b; Abdullah, *et al.*, 2011d) suffer in term of efficiency and evaluating the real datasets. Most of the approaches decrease the minimum support threshold to generate the potential list of the least itemset. However, it accidentally generates the huge number of unnecessary association rules. Indeed, it is tedious works to determine which rules are useful and significant. Besides that, lowering the minimum support will also proportionally increase the computational cost or memory consumption. Since the complexity of study, difficulties in algorithms (Yun, *et al.*, 2003) and it may require excessive computational cost, there are very limited efforts have been paid to discover least association rules.

Educational data is one of the potential resources in discovering the significant least association rules. These rules can be very useful for higher authority personnel in assisting them to make right decision. For instance, in every July semester, our university receives approximately 160 students to enroll in computer science program. However, in the first meeting that has been conducted by Department of Computer Science, many of them are not really enjoyed to study. In fact, many of them are already appealed to study at others university with their preferred programs rather than computer science program. In educational system of Malaysia, all matriculation or diploma holder students are required to apply via online application system to choose up to eight (8) preferred bachelor programs that will be offered by Malaysian universities. There are always the cases that the students are uncertain and taken for granted by combining with the various fields of interests. This situation is usually happened to those students that are not so good in their examination result. For instance, the students might choose contradicted field of study in university programs such as business administration, engineering, forestry and computer science. The research question is how to justify the student interests since there is no such field to be specified in the online application system. At the moment, if their choices are not selected, they will be offered to any program in the same or others university according to programs availability.

As a result, if the offered program is not tally with the students' interest, it may demoralize their courage to be excellent students. In many countries and including Malaysia, the chances to work with national or multi-national company are always depend on students' CGPA.

Therefore, this study is very important since it can show the uncommon relationship among the chosen programs. Furthermore, it can also reveal the rules that might contain the combination of both frequent and least university's programs. The output from this study can be used by university representative (higher authority personnel) as a guideline in offering the appropriate programs for that category of students. Indeed, it can assist the university's policy maker to comprehend the issues and also enhance the current educational standards and managements process as a whole (Romero, *et al.*, 2010; Sevindik, *et al.*, 2010; Buldu & Ucgun, 2010; Romero, *et al.*, 2009; Enclieva & Tumin, 2006).

Therefore, in this paper, we apply our novel measure called Definite Factors (DF) to detect the abnormal relationship among university's programs that have been selected by students. Indeed, DF will take into consideration the combination of both frequent and least university's program for generating the desired least association rules. We also employed our LP-Tree and LP-Growth algorithms (Abdullah, *et al.*, 2010a) prior to produce the rules. In this study, the experiment was performed based on the 2008/2009 intake students' that have been offered in Bachelor of Information Technology (Software Engineering) at Universiti Malaysia Terengganu (UMT).

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

2. Related Works

Nowadays, varieties of data mining method methods have been proposed in educational data mining. The proposed methods are slightly different from the standard data mining methods due to the specificities of educational data, such as their multi-level hierarchy and non-independence. Romero and Ventura (2007) suggested two categories of education data mining. The first category contains both statistics and visualization. The second one is web mining which can be divided into three parts. The first part covers clustering, classification, and outlier detection. The second part consists of association rule mining and sequential pattern mining. Finally, the third part is associated with text mining. It can be conclude that, the initial educational data mining is come into sight by analyzing the interaction between student and computer based on detailed logs of all their activities.

Baker and Yacef (2009) proposed educational data mining into five different categories named prediction, clustering, relationship mining, distillation of data for human judge and discovery with model. For the first three categories, it seems to be quite similar to the most standard studies in data mining (More, 2006). Distillation of data for human judge category has become one of the famous research findings (Kay, *et al.*, 2006) and theoretical discussion in educational data mining (Tanimoto, 2007). This category is quite universally and not purely own by data mining research area alone. As compared to educational data mining as proposed by (Romero, *et al.*, 2008), distillation of data for human judge is similar to statistics and visualization category. Recently, discovery with models category is now become one of the most popular methods in educational data mining research. It deals with the sophisticated analysis such as discovering which learning materials sub-categories of students are the most beneficial (Beck & Mostow, 2008), finding how different type of student behavior contribute to student's learning in different way (Cocea, *et al.*, 2009) and revealing how variety of designing the intelligent tutor influence student's behavior over time (Jeong & Biswas, 2008).

Nowadays, only few attentions have been paid to extract least association rules from educational data. To the best of our knowledge, only one paper (Romero, *et al.*, 2010) is specifically discussed about least association rules. They applied the existing Rare Association Rules Mining (Apriori-like) algorithms to extract association rules from e-learning data. Their objective is to discover the information about infrequent student behavior. Four Apriori-based algorithms were employed to extract these rules named Apriori-Frequent (Agrawal, *et al.*, 1993), Apriori-Infrequent, Apriori-Inverse (Koh & Rountree, 2005) and Apriori-Rare (Szathmary, 2007). From the experiments, Apriori-Inverse and Apriori-Rare are proven more suitable in finding the least association rules.

In term of measurement least association rules, one of the popular measurement is Relative Support Apriori (RSA) proposed by (Yun, *et al.*, 2003). RSA requires three (3) predefined measurements called 1st support, 2nd support and relative support (1st support > 2nd support). An item is said a least item if its support is less than 1st support and greater or more than 2nd support. A frequent item is an item having a support which equal or greater than 1st support. The least association rules are those rules that satisfied all the predefined supports. The main constrain of this algorithm is it increases the computational cost if the minimum relative support is set close to zero. In addition, determination of three predefined measurements is also another issue for this algorithm. Besides RSAA, the others approach to capture least association rules are Multiple Support Apriori (Kiran & Reddy, 2009), Matrix-based Scheme (Zhou & Yau, 2007), Collective Support Apriori (Selvi & Tamilarasi, 2009), *etc*.

3. Proposed Method

Throughout this section the set $I = \{i_1, i_2, \cdots, i_{|A|}\}$, for $|A| > 0$ refers to the set of literals called set of items, $W = \{w_1, w_2, \cdots, w_{|A|}\}$, refers to the set of literals called set of weights with a non-negative real numbers, and the set $D = \{t_1, t_2, \cdots, t_{|U|}\}$, for $|U| > 0$ refers to the data set of transactions, where each transaction $t \in D$ is a list of distinct items $t = \{i_1, i_2, \cdots, i_{|M|}\}$, $1 \le |M| \le |A|$ and each transaction can be identified by a distinct identifier TID.

3.1. Definition

In order to easily comprehend our measurement, some required definitions together with a sample transactional data are presented.

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* $\operatorname{supp}(X)$ *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* $X \bigcap Y = \phi$. *The sets X and Y are called antecedent and consequent, respectively.*

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

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

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

Definition 6. (Definite Factor). *Definite Factor is a formulation of exploiting the support difference between itemsets with the frequency of an itemset against a baseline frequency. The baseline frequency of itemset is presumed as statistically independence.*

The Definite Factor denoted as DF and

$$DF(I) = |P(X) - P(Y)| \times \frac{P(X \cup Y)}{P(X)P(Y)}$$

It also can be expressed as

$$DF(I) = |\mathrm{supp}(X) - \mathrm{supp}(Y)| \times \left(\frac{\mathrm{supp}(X \Rightarrow Y)}{\mathrm{supp}(X) \times \mathrm{supp}(Y)} \right)$$

Example 1. (Definite Factor). Let

$$T = \{\{1,2,5\}, \{2,4\}, \{2,3\}, \{1,2,4\}, \{1,3\}, \{2,3\}, \{1,3\}, \{1,2,3,5\}, \{1,2,3\}\},$$

based on ILSupp $[0.2, 0.3]$ and value of $DF \geq 0.7$, the calculation of DF for itemset $\{1,4\}$ and $\{2,4\}$ are as follows:

$$DF(\{1,4\}) = |6 - 2| \times \frac{(1)}{(6 \times 2)} = 0.33$$

and

$$DF(\{2,4\}) = |7 - 2| \times \frac{(2)}{(7 \times 2)} = 0.71$$

Since $DF(\{1,4\})$ is less than minimum *DF* (Min-*DF*), thus itemset $\{1,4\}$ is rejected and only itemset $\{2,4\}$ is accepted. Therefore, itemset $\{2,4\}$ is accepted as significant least association rules or patterns.

3.2. Construct Definite Least Association Rules

Rule is classified as Definite Least Association Rules (DLAR) if it fulfilled two conditions. First, DF of association rule must be greater than the predefined minimum DF. The range of min-DF is in between 0 and 1. Second, the antecedent and consequence of association rule must represent either Least Items or Frequent Items, respectively. The computation of DF of each association rule is employed from Definition 6. The complete procedure to construct the DLAR algorithm is as follows.

DLAR Algorithm
Specify $DF^{\min}$
for $(DI_a \in DefiniteItemset)$ do
for $(DFI_i \in DI_a \cap FrequentItems)$ do
for $(DLI_i \in DI_a \cap LeastItems)$ do
Compute $DF(DFI_i, DLI_i)$
if $(DF(DFI_i, DLI_i) > DF^{\min})$ do
Insert $DLAR(DFI_i, DLI_i)$
end if
end for loop
end for loop
end for loop

Figure 1. DLAR Algorithm

4. Experimental Results

In this section, we do experiment tests with DF measurements. The weight of all association rules were assigned according to this measurement. These experiments were conducted 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. We evaluate the proposed measurement to 2008/2009 intake students in computer science program. The data was obtained from Division of Academic, Universiti Malaysia Terengganu in a text file and Microsoft excel format. There were 160 students involved and their identities were removed due to the confidentiality agreement. In the original set of data, it consists of 35 attributes and the detail information were explained in 10 tables which provided in Microsoft excel format.

Here, 8 chosen university programs by the students are extracted according to the fix location in the original flat file. The actual location for each programs are based on the fix column. There were in total of 822 bachelors programs offered in Malaysian public universities for July 2008/2009 students' intake. From this figure, 342 bachelor programs were selected by our 160 students and it can be generalized into 47 unique general fields. For simplicity, only 5 bachelor programs were extracted as illustrated in Table 1. In addition, LP-

Tree and LP-Growth algorithm with DF measure (Abdullah, *et al.*, 2010) are employed in the experiment. The total of 4,177 association rules was successfully extracted. Figure 2 reveals the total number of student that had chosen several (or not at all) Computer Science programs offered in the university. It is about 32% of the students are not applied any Computer Science program and approximately 36% of them are chosen 4 Computer Science programs during selecting their preferred university's programs. Figure 3 shows the total number of association rules using different support ranges. The highest number of association rules was produced for the rules that had support below than 1%, while the lowest one was the rules that had support in between 2% to 3%. For further analysis, we are focusing on the rules that had a support more than 3%. The 3% of minimum support is equivalent to the itemset that must appear at least 5 times from the current transactions. Table 2 depicts top 10 of association rules based on the 3% of minimum support.

Table 1. Mapping a part of bachelor programs' offered, its general field and code

Program Description	Field	Code
Bachelor of Business	Business	12
Bachelor of Media Communication	Media Communication	13
Bachelor of Computer Science (Software Engineering)	Computer	14
Bachelor of Counselling	Counselling	15
Bachelor of Dental Surgery	Dentist	16

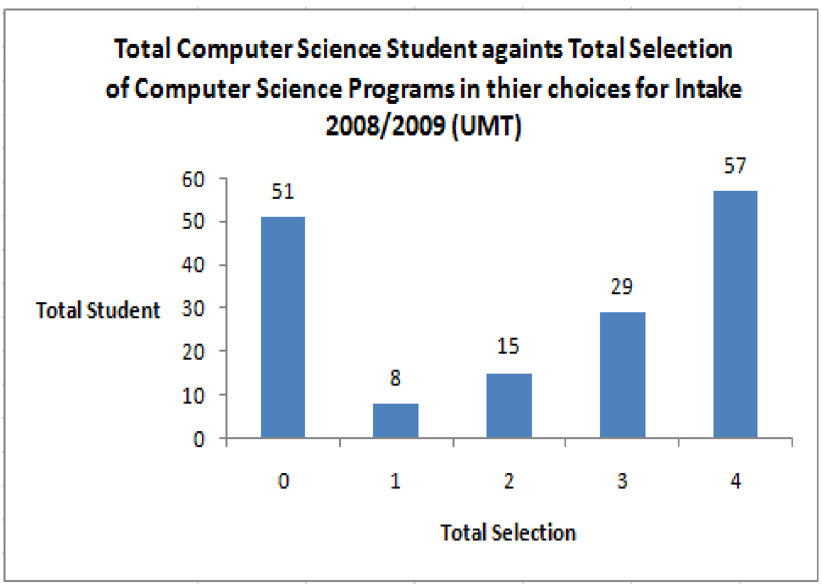

Figure 2. Total number of Computer Science students that had selected the Computer Science programs in their preferred choices

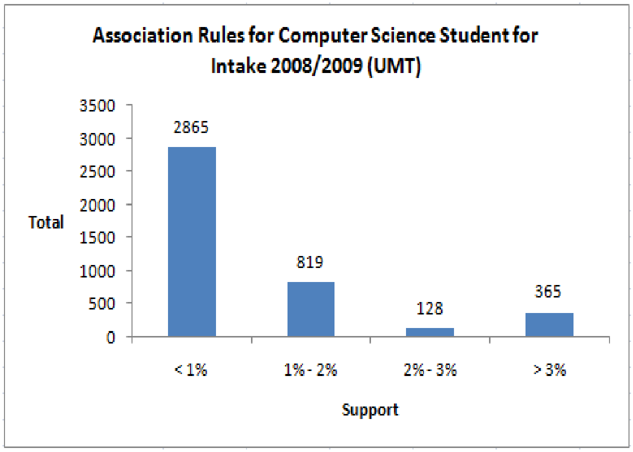

Figure 3. Total number of association rules based on the different support ranges

Table 2. Top 10 of association rules sorted by DF in descending order and 100% confidence

No	Association rules	Supp of Antecedent	Supp of Consequence	Supp of Itemset	Jaccard	Corr	RSA	DF
1	25 --> 9	4.38	90.63	4.38	0.048	22.86	1.00	0.95
2	25 --> 34	6.25	90.63	6.25	0.07	16.00	1.00	0.93
3	25 --> 8	12.50	90.63	12.50	0.14	8.00	1.00	0.86
4	25 --> 43	9.38	90.63	8.75	0.10	10.67	0.93	0.84
5	28 --> 40	3.75	68.75	3.12	0.05	26.67	0.83	0.79
6	25 --> 41	8.13	90.63	6.88	0.07	12.31	0.85	0.77
7	25 --> 38	15.00	90.63	13.75	0.15	6.67	0.92	0.76
8	25 --> 31	5.63	90.63	4.38	0.05	17.78	0.78	0.73
9	28 --> 34	6.25	68.75	5.00	0.07	16.00	0.80	0.73
10	25 28 --> 34	6.25	60.23	5.00	0.08	16.00	0.80	0.72

Table 3. Explanation of top 10 of positive association rules

No	Association rules	Explanation
1	25 --> 9	The student chose Forestry program also chose Banking program
2	25 -> 34	The student chose Forestry program also chose Nursing program
3	25 -> 8	The student chose Forestry program also chose Art Design
4	25 -> 43	The student chose Forestry program also chose Radiotherapy program
5	28 --> 40	The student chose IT program also chose Psychology
6	25 --> 41	The student chose Forestry program also chose Pure Sciences
7	25 -> 38	The student chose Forestry program also chose Physiotherapy
8	25 --> 31	The student chose Forestry program also chose Management
9	28 --> 34	The student chose IT program also chose Nursing
10	25 28 --> 34	The student chose Forestry and IT program also chose Nursing

Table 3 illustrates the meaning of association rules based on the Table 2. The link of interest between the antecedent and consequence for the first rule until fifth rule is quite strange due to the contradiction in the field of study among the respective programs. The sixth rule is very realistic since both programs have a similarity in term of basic requirements, link of interest and nature of study. For the sixth until tenth rules, it is very hard and confused to explains, since there is no link of interest between the programs. From here we can see that the students have mixed up with several interests during choosing their preferred university's programs. Moreover, most of them had chosen Forestry program. In summary, there are existed exceptional association rules in the university's program selection database. This information is very important to give an overall idea about the student interests and how to channel them to a more appropriate university's program.

5. Conclusion

Mining least association rules is very useful to help the organization in making a right decision. In educational context, identifying the suitable program for prospect students is very troublesome and usually ends up with programs availability. The students' interest from the pattern of past chosen university's programs is not taken into account. Therefore, this paper employed the Definite Factors measure to the students' enrolment data of computer science program (intake 2008/2009) at University Malaysia Terengganu. The result shows that the applied measure can discover the significant least association rules. From the generated rules, 32% of the students that have been offered in computer science program are not within their program interests. Thus, effective monitoring process and analysis of these students are very important in helping them to adapt and finally enjoy with the current program.

In a near future, we are going to evaluate the proposed model to others datasets. We also believed that the model is also suitable to others real domain applications.

References

[1] Z. Abdullah, T. Herawan and M. M. Deris, "Mining significant least association rules using fast SLP-Growth algorithm", In T.H. Kim and H. Adeli (Eds.): AST/UCMA/ISA/ACN 2010, Lecture Notes in Computer Science, Springer-Verlag, vol. 6059, **(2010a)**, pp. 324–336.

[2] Z. Abdullah, T. Herawan and M. M. Deris, "Scalable Model for Mining Critical Least Association Rules", In Rongbo Zhu, *et al.*, (Eds.): ICICA 2010, Lecture Notes in Computer Science, Springer-Verlag, vol. 6377, **(2010b)**, pp. 509–516.

[3] Z. Abdullah, T. Herawan, A. Noraziah and M. M. Deris, "Extracting Highly Positive Association Rules from Students' Enrollment Data", Procedia Social and Behavioral Sciences, vol. 28, **(2011a)**, pp. 107–111.

[4] Z. Abdullah, T. Herawan, A. Noraziah and M. M. Deris, "Mining Significant Association Rules from Educational Data using Critical Relative Support Approach", Procedia Social and Behavioral Sciences, vol. 28, **(2011b)**, pp. 97–101.

[5] Z. Abdullah, T. Herawan and M. M. Deris, "Visualizing the Construction of Incremental Disorder Trie Itemset Data Structure (DOSTrieIT) for Frequent Pattern Tree (FP-Tree)", In H.B. Zaman, *et al.*, (Eds.): IVIC 2011, Lecture Notes in Computer Science, Springer-Verlag, vol. 7066, **(2011c)**, pp. 183–195.

[6] Z. Abdullah, T. Herawan and M. M. Deris, "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, vol. 188, no. II, **(2011d)**, pp. 475–485.

[7] C. C. Aggarwal and P. S. Yu, "A New Framework for Item Set Generation", The Proceedings of the ACM PODS Symposium on Principles of Database Systems, **(1998)**, pp. 18–24.

[8] R. Agrawal, T. Imielinski and A. Swami, "Database Mining: A Performance Perspective", IEEE Transaction on Knowledge and Data Engineering, vol. 5, no. 6, **(1993)**, pp. 914–925.

[9] W. H. Au and K. C. C. Chan, “Mining Fuzzy ARs in a Bank-Account Database”, IEEE Transactions on Fuzzy Systems, vol. 11, no. 2, **(2003)**, pp. 238–248.
[10] R. Baker and K. Yacef, “The state of educational data mining in 2009: a review and future visions”, Journal of Education Data Mining, vol. 1, no. 1, **(2009)**, pp. 3–17.
[11] J. E. Beck and J. Mostow, “How who should practice: Using learning decomposition to evaluate the efficacy of different types of practice for different types of students”, In Proceeding of the 9th International Conference on Intelligent Tutoring Systems, **(2008)**, pp. 353–362.
[12] A. Buldu and K. Ucgun, “Data mining application on students’ data”, Procedia Social and Behavioral Sciences, vol. 2, **(2010)**, pp. 5251–5259.
[13] M. Cocea, A. Hershkovitz and R. S. J. D. Baker, “The impact of off-task and gaming behaviors on learning: immediate or aggregate”, In Proceedings of the 14th International Conference on Artificial Intelligence in Education, **(2009)**, pp. 507–514.
[14] J. Ding, “Efficient Association Rule Mining among Infrequent Items”, Ph.D. Thesis, University of Illinois at Chicago, **(2005)**.
[15] S. Enclieva and S. Tumin, “Application of association rules for efficient learning work-flow”, In Z. Shi, Shimohara K., Feng D. (Eds.), Intelligent Information Processing III, Springer-Verlag, vol. 228, **(2006)**, pp. 499–504.
[15] T. Herawan, I. T. R. Yanto and M. M. Deris, “Soft set approach for maximal association rules mining”, In D. Ślęzak et al. (Eds.): DTA 2009, Communication of Computer and Information Sciences, Springer-Verlag, vol. 64, **(2009a)**, pp. 163–170.
[16] T. Herawan, I. T. R. Yanto and M. M. Deris, “SMARViz: Soft maximal association rules visualization”, In H. Badioze Zaman, *et al.*, (Eds.): IVIC 2009, Lecture Notes in Computer Science, Springer-Verlag, vol. 5857, **(2009b)**, pp. 664–674.
[17] T. Herawan and M. M. Deris, “A soft set approach for association rules mining”, Knowledge Based Systems, vol. 24, no. 1, **(2011)**, pp. 186–195.
[18] T. Herawan, P. Vitasari and Z. Abdullah, “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, vol. 188, no. II, **(2011)**, pp. 495–508.
[19] Z. Huang, X. Lu and H. Duan, “Mining association rules to support resource allocation in business process management”, Expert Systems with Applications, vol. 38, no. 8, **(2011)**, pp. 9483−9490.
[20] H. Jeong and G. Biswas, “Mining student behavior models in learning by teaching environments”, In Proceeding of the 1st International Conference on Educational Data Mining, **(2008)**, pp. 127−136.
[21] J. Kay, N. Maisonneuve, K. Yacef and P. Reimann, “The big five and visualization of team work activity”, In Ikeda, M., Ashley, K.D., and Chan, T-W. (Eds.): Intellegent Tutoring Systems, Springer, **(2006)**, pp. 197–206.
[22] R. U. Kiran and P. K. Reddy, “An Improved Multiple Minimum Support Based Approach to Mine Rare Association Rules”, In Proceeding of IEEE Symposium on Computational Intelligence and Data Mining, **(2009)**, pp. 340−47.
[23] Y. S. Koh and N. Rountree, “Finding Sporadic Rules using Apriori-Inverse”, In: Ho, T.B., Cheung, D., Liu, H. (Eds): PAKDD 2005, Lecture Notes in Computer Science, Springer-Verlag, vol. 3518, **(2005)**, pp. 97−106.
[24] B. Liu, W. Hsu and Y. Ma, “Mining Association Rules with Multiple Minimum Supports”, In Proceeding of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'99), **(1999)**, pp. 337−341.
[25] M. D. Mustafa, N. F. Nabila, D. J. Evans, M. Y. Saman and A. Mamat, “Association rules on significant rare data using second support”, International Journal of Computer Mathematics, vol. 83, no. 1, **(2006)**, pp. 69−80.
[26] A. W. More, “Statistical data mining tutorials”, **(2006)**, http://www.autonlab.org/tutorials/.
[27] C. Romero and S. Ventura, “Educational data mining: A survey from 1995 to 2005”, Expert Systems with Applications, vol. 33, **(2007)**, pp. 135–146.
[28] C. Romero, S. Ventura and E. García, “Data mining in course management systems: Moodle case study and tutorial”, Computers & Education, vol. 51, no. 1, **(2008)**, pp. 368–384.
[29] C. Romero, S. Ventura, A. Zafraa and P. Brab, “Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems”, Computers & Education, vol. 53, no. 3, **(2009)**, pp. 828–840.
[30] C. Romero, J. R. Romero, J. M. Luna and S. Ventura, “Mining rare association rules from e-Learning data”, In Proceeding of The Third International Conference of Education Data Mining, **(2010)**, pp. 171−180.
[31] C. S. K. Selvi and A. Tamilarasi, “Mining association rules with dynamic and collective support thresholds”, Internationall Journal on Open Problems Computational Mathematics, vol. 2, no. 3, **(2009)**, pp. 427−438.

[32] T. Sevindik, N. Demirkeser and Z. Cömert, "Virtual education environments and web mining", Procedia Social and Behavioral Sciences, vol. 2, **(2010)**, pp. 5120–5124.
[33] L. Szathmary, A. Napoli and P. Valtchev, "Towards rare itmeset mining", In Proceeding of Internationall Conference on Tools with Artificial Intelligence, **(2007)**, pp. 305–312.
[34] P.-N. Tan, M. Steinbach and K. Vipin, "Introduction to Data Mining", Addison-Wesley, **(2006)**.
[35] S. L. Tanimoto, "Improving the prospects for educational data mining", In Proceeding of 11th International conference on user modeling (UM 2007), **(2007)**, pp. 106–110.
[36] F. Tao, F. Murtagh and M. Farid, "Weighted Association Rule Mining using Weighted Support and Significant Framework", In Proceeding of ACM SIGKDD'03, **(2003)**, pp. 661–666.
[37] K. Wang, Y. Hee and J. Han, "Pushing Support Constraints into Association Rules Mining", IEEE Transaction on Knowledge and Data Engineering, vol. 15, no. 3, **(2003)**, pp. 642–658.
[38] H. Yun, D. Ha, B. Hwang and K. H. Ryu, "Mining Association Rules on Significant Rare Data using Relative Support", The Journal of Systems and Software, vol. 67, no. 3, **(2003)**, pp. 181–19.
[39] L. Zhou and S. Yau, "Assocation Rule and Quantative Association Rule Mining among Infrequent Items", In Proceeding of MDM2007 in conjunction with ACM SIGKDD'07, Article No. 9, **(2007)**.

Authors

Zailani Abdullah

He received his Ph.D from University Tun Hussein Onn Malaysia, (UTHM) in 2012. He has published more than 40 research papers in journal and conference proceedings. He has served as a Co-Chair for the 4th Malaysian Software Engineering Conference 2008 (MySEC 2008), Committee Member for The 16th Asia-Pacific Software Engineering Conference (APSEC 2009), Committee Member for The 2nd KnowledgeGrid Malaysia Forum 2009 and Program Committee Member for The 3rd International Conference on Computer Systems and Software Engineering (ICSECS 2013). He is also Microsoft® Certified Technology Specialist (MCTS), .NET Framework 3.5, ASP.NET Application and Oracle Database 11g Administrator Certified Associate. He has appointed as editorial board member for Journal of Computational Intelligence and Electronic Systems (JCIES) and reviewer for World Applied Science Journal, Aceh International Journal of Science and Technology (AIJST), Global Perspective on Engineering Management (GPEM) and Mosharaka for Researches and Studies, respectively. His research interests include database, data mining and web-based applications.

Tutut Herawan

He received a B.Ed degree in year 2002 and M.Sc degree in year 2006 degree in Mathematics from Universitas Ahmad Dahlan and Universitas Gadjah Mada Yogyakarta Indonesia, respectively. He obtained a PhD in Theoretical Data Mining from Universiti Tun Hussein Onn Malaysia in year 2010. Currently, he is a lecturer with Department of Mathematics Education, Universitas Ahmad Dahlan, Indonesia**.** He currently supervises four PhD and had successfully co-supervised two PhD students and published more than 120 papers in various international journals and conference proceedings. He has appointed as an editorial board member for IJDTA, TELKOMNIKA, IJNCAA, IJDCA and IJDIWC. He is also been appointed as a reviewer of several international journals such as Knowledge-Based Systems, Information Sciences, European Journal of Operational Research, Applied Mathematics Letters, and guest editor for several special issues of international journals. He has served as a program committee member and co-organizer for numerous international conferences/workshops including Soft Computing and Data Engineering (SCDE 2010-2011 at Korea, SCDE 2012 at Brazil), ADMTA 2012 Vietnam, DTA 2011-2012 at Korea, DICTAP 2012 at Thailand, ICDIPC 2012 at Lithuania, DEIS 2012 at Czech Republic, NDT 2012 at Bahrain, ICoCSIM 2012 at Indonesia, ICSDE'2013 at Malaysia, ICSECS 2013 at Malaysia, SCKDD 2013 at Vietnam and many more. His research area includes Knowledge Discovery in Databases, Educational Data Mining, Decision Support in Information System, Rough and Soft Set theory.

Noraziah Ahmad

She received Ph.D in Distributed Database from University Malaysia Terengganu (UMT) in 2007. She has published more than 150 papers in the journals and conference proceedings. Currently, she is currently an associate professor at Faculty of Computer Systems & Software Engineering, University Malaysia Pahang. In addition to serving as international program committee member and reviewers in many conferences, she is currently an editorial board members of the International Journal of Engineering and Technology (IJET), International Journal of Web Application (IJWA) and Journal of Emerging Technologies in Web Intelligence (JETWI); a member of IEEE Computer Society, International Association of Engineers (IAENG), World Academy of Science, Engineering and Technology (WASET), Malaysian National Computer Confederation (MNCC) and Senior member of International Association of Computer Science Information Technology (IACSIT).

Mustafa Mat Deris

He received PhD from University Putra Malaysia in 2002. He is a professor of computer science in the Faculty of Computer Science and Information Technology, UTHM, Malaysia. He has successfully supervised **ten** PhD students and currently he is supervising **six** PhD students and published more than 170 papers in journals and conference proceedings. He has appointed as *editorial board member* for Journal of Next Generation Information Technology, JNIT, Korea, and Encyclopedia on Mobile Computing and Commerce, Idea Group, USA, Guest editor of International Journal of BioMedical Soft Computing and Human Science for Special Issue on "Soft Computing Methodologies and Its Applications” a *reviewer* of several international journals such as IEEE Transaction on Parallel and Distributed Computing, Journal of Parallel and Distributed Databases, Journal of Future Generation on Computer Systems, Journal of Information Sciences, Elsevier, Journal of Cluster Computing, Kluwer, and Journal of Computer Mathematics, Taylor & Francis, UK. He has served as a *program committee member* and *co-organizer* for numerous international conferences/workshops including Grid and Peer-to-Peer Computing, (GP2P 2005, 2006), Autonomic Distributed Data and Storage Systems Management (ADSM 2005, 2006, 2007), and Grid Pervasive Computing Security, organizer for workshops on Rough and Soft Sets Theories and Applications (RSAA 2010), Fukuoka, Japan, and Soft Computing and Data Engineering (SCDE) (2010, 2011, Korea), (2012, Brazil). His research interests include distributed databases, data grid, data mining and soft computing.

