ISSN 2389-8186
E-ISSN 2389-8194
Vol.7, No. 2-1
Julio-diciembre de 2020
doi: https://doi.org/10.16967/23898186.666

ARTÍCULOS ORIGINALES

Towards Association Rule-based Item Selection Strategy in Computerized Adaptive Testing*

pp. 19-30

JOSUÉ PACHECO-ORTIZ**
LISBETH RODRÍGUEZ-MAZAHUA***
JEZREEL MEJÍA-MIRANDA****
ISAAC MACHORRO-CANO*****
GINER ALOR-HERNÁNDEZ******
ULISES JUÁREZ-MARTÍNEZ*******

* The authors are very grateful to Tecnológico Nacional de México for supporting this work. Also, this research paper was sponsored by the CONACYT.
** Master in Administrative Engineering. Tecnológico Nacional de México, Veracruz, México. E-mail: iscjpo@gmail.com. ORCID: 0000-0002-2694-2827. Google Scholar: https://scholar.google.com/citations?user=z2ZZQOoAAAAJ&hl=es
*** PhD in Computer Science. Tecnológico Nacional de México, Veracruz, México. E-mail: lrodriguezm@ito-depi.edu.mx. ORCID: 0000-0002-9861-3993. Google Scholar: https://scholar.google.es/citations?user=2hZw4HAAAAAJ&hl=es.
**** PhD in Computer Science. Centro de Investigación en Matemáticas CIMAT, Zacatecas, México. E-mail: jmejia@cimat.mx. ORCID: 0000-0003-0292-9318. Google Scholar: https://scholar.google.com/citations?user=gdlrfgEAAAAJ&hl=es
***** PhD in Engineering Sciences. Universidad del Papaloapan, Oaxaca, México. E-mail: imachorro@unpa.edu.mx. ORCID: 0000-0002-3822-4478. Google Scholar: https://scholar.google.es/citations?view_op=list_works&hl=es&user=fE8k294AAAAJ.
****** PhD in Electrical Engineering. Tecnológico Nacional de México, Veracruz, México. E-mail: galorh@orizaba.tecnm.mx. ORCID: 0000-0003-3296-0981. Google Scholar: https://scholar.google.es/citations?hl=es&user=8Zgf4KwAAAAJ.
******* PhD in Science in Electrical Engineering. Tecnológico Nacional de México, Veracruz, México. E-mail: ujuarezm@orizaba.tecnm.mx. ORCID: 0000-0002-5911-3136. Google Scholar: https://scholar.google.com/citations?hl=en&user=6Ko_q38AAAAJ

ISSN 2389-8186
E-ISSN 2389-8194

Vol.7, No. 2-1
Julio-diciembre de 2020
doi: https://doi.org/10.16967/23898186.666

COMO CITAR ESTE ARTÍCULO
How to cite this article:

Pacheco-Ortiz, J. et al. (2020). Towards Association Rule-based Item Selection Strategy in Computerized Adaptive Testing. *Revista Perspectiva Empresarial*, *7*(2-1), 19-30.

Recibido: 20 de agosto de 2020
Aceptado: 07 de diciembre de 2020

ABSTRACT One of the most important stages of Computerized Adaptive Testing is the selection of items, in which various methods are used, which have certain weaknesses at the time of implementation. Therefore, in this paper, it is proposed the integration of Association Rule Mining as an item selection criterion in a CAT system. We present the analysis of association rule mining algorithms such as Apriori, FP-Growth, PredictiveApriori and Tertius into two data set with the purpose of knowing the advantages and disadvantages of each algorithm and choose the most suitable. We compare the algorithms considering number of rules discovered, average support and confidence, and velocity. According to the experiments, Apriori found rules with greater confidence, support, in less time.

KEY WORDS Computerized adaptive testing, association rules, e-learning, intelligent systems.

Hacia una estrategia de selección de ítems basada en reglas de asociación en pruebas adaptativas computarizadas

RESUMEN Una de las etapas más importantes de las pruebas adaptativas informatizadas es la selección de ítems, en la cual se utilizan diversos métodos que presentan ciertas debilidades al momento de su aplicación. Así, en este trabajo, se propone la integración de la minería de reglas de asociación como criterio de selección de ítems en un sistema CAT. Se presenta el análisis de algoritmos de minería de reglas de asociación como Apriori, FP-Growth, PredictiveApriori y Tertius en dos conjuntos de datos con el fin de conocer las ventajas y desventajas de cada algoritmo y elegir el más adecuado. Se compararon los algoritmos teniendo en cuenta el número de reglas descubiertas, el soporte y confianza promedios y la velocidad. Según los experimentos, Apriori encontró reglas con mayor confianza y soporte en un menor tiempo.

PALABRAS CLAVE pruebas adaptativas informatizadas, reglas de asociación, *e-learning*, sistemas inteligentes.

A uma estratégia de seleção de itens baseada em regras de associação em provas adaptativas informatizadas

RESUMO Uma das etapas mais importantes das provas adaptativas informatizadas é a seleção de itens, na qual se utilizam diversos métodos que apresentam certas debilidades no momento da sua aplicação. Assim, neste trabalho, se propõe a integração da mineração de regras de associação como critério de seleção de itens num sistema CAT. Se apresenta a análise de algoritmos de mineração de regras de associação como Apriori, FP-Growth, PredictiveApriori e Tertius em dois conjuntos de dados com o fim de conhecer as vantagens e desvantagens de cada algoritmo e eleger o mais adequado. Se compararam os algoritmos tendo em conta o número de regras descobertas, o suporte e confiança em média e a velocidade. Segundo os experimentos, Apriori encontrou regras com maior confiança e suporte num menor tempo.

PALAVRAS CHAVE provas adaptativas informatizadas, regras de associação, *e-learning*, sistemas inteligentes.

Introduction

Computer Adaptive Testing —CAT— (Chen, Chao and Chen, 2019) has revolutionized the traditional way of evaluating, since it dynamically selects and manages the most appropriate questions depending on the previous answers given by the examinees. One of the central components of a CAT is the item selection criterion (Miyazahua and Ueno, 2019), although the most widely used criterion is Fisher's Maximum Information (Albano et al., 2019), it presents several weaknesses that generate a certain degree of mistrust, for example, bias in the item selection, estimation errors at the start of the exam, or the same question being displayed repeatedly to the tested one (Sheng, Bingwei and Jiecheng, 2018; Du, Li and Chang, 2018; Lin and Chang, 2019; Yigit, Sorrel and de la Torre, 2019; Ye and Sun, 2018). Therefore, in this paper the development of a CAT system that uses association rules for the selection of items is proposed, focusing on using the potential advantages of association rules to find relationships between the questions answered correctly or incorrectly and the questions answered correctly, and thus present the most appropriate questions (most likely to answer correctly) in the tests, according to the responses of the evaluated, considering the best rules (stored in the database of students who submitted the same test previously) with greater support and confidence.

Several research projects have used association rule mining —ARM— with different algorithms in their development, for example, in Rubio Delgado et al. (2018), authors applied Apriori, FP-Growth, PredictiveApriori and Tertius, grouping them according to their configuration characteristics to compare them, so Apriori and FP-Growth were contrasted using different support and confidence values, whereas for PredictiveApriori and Tertius once specified the number of rules, the execution time, the number of rules generated, the support and confidence values were taken into account in all cases. In contrast, Wang et al. (2018) worked with the Apriori algorithm, occupying for the comparison process the minimum and confidence support, whose set of generated rules were debugged based on the minimum Lift, Chi-squared test and minimum improvement. While in Prajapati, Garg and Chauhan (2017) apart from support and confidence, all-confidence, cosine, interestingness of a rule, lift, execution time and conviction were used in the process of comparing the Distributed Frequent Pattern Mining —DFMP—, Count Distribution Algorithm —CDA— and Fast Distributed Mining —FDM— algorithms.

The objective of this paper is to present a comparative analysis of various ARM algorithms that allows to select the most suitable for the implementation in the CAT system that is proposed. This paper consists of five sections: (i) the introduction; (ii) background and some of the works related to this research; (iii) the integration of ARM in the CAT process and the comparison method that was followed; (iv) the results and the analysis of them; finally (v) conclusions and future work.

Background and related works

Over the years, in different projects, various tools have been applied in the development of the phases that make up the CATs, for example: three-parameter logistic model for item calibration (Lee et al., 2018); maximum likelihood estimation for the evaluator's skill estimation (Albano et al., 2019); and root mean square differences as an evaluation criterion (Stafford et al., 2019), among others. Specifically, for the item selection stage, work has been done to solve the problems presented by Fisher's Maximum Information, using other selection strategies, for example, Bayesian networks (Tokusada and Hirose, 2016), Greedy algorithm (Bengs, Brefeld and Krohne, 2018), Kullback-Leibler Information (Chen et al., 2017), Minimum Expected Subsequent Variance (Rodríguez-Cuadrado et al., 2020), to mention a few which, while they have achieved favorable results, most have only been in studies of simulation and not in real application.

We propose using ARM as an item selection criterion because we can exploit its ability to find associations or correlations between the elements or objects (in this case, test answers given by other students in the past) of a database, it has many advantages, among which are: (i) associations can occur between correct/incorrect answers and correct ones; (ii) they will determine the suitable item according to the answer of the evaluated; and (iii) the items presented to the examinees will be selected, considering interesting metrics widely used in related works. ARM has been used in various areas, among which are: recommendation systems (Dahdouh et al., 2019) and online learning (Gu, Zhou and Yan, 2018), offering positive results in each case; however, to the best of our knowledge, its use as a selection strategy for CAT is not currently reported, therefore, this project contemplates the integration of ARM in the stage of selecting items in the CAT. The expected outcome at the end of the project is a system that uses the benefits of both CAT and association rules in the educational evaluation process, looking as a final product for a system that is not only adaptive, but also learns and evolves according to the experiences that it accumulates over time.

Methodology

The following subsections specify the integration of ARM in the CAT process and the method performed for comparing ARM algorithms. The first subsection shows the proposed CAT process. The second subsection contemplates the data bank used for comparison. The third subsection includes the specificity of the algorithms used

CAT process with ARM as item selection criterion

The process followed by the proposed CAT is shown in Figure 1, which begins with an initial estimate of student knowledge to select and present the first item, once the student's answer is obtained, a new knowledge estimate is made. While the stop criterion is not met, and if the answer to the previous question was correct, then a question with a higher level of complexity is chosen using association rules, else one with less complexity is selected according to association rules, then the item is presented to the student to recalculate his/her level of knowledge estimate. This cycle is repeated until the stop criterion is met, when this happens, all information is saved, a final estimate of knowledge is made, the grade is displayed to the student, which logs out and new association rules are obtained and saved automatically that will serve the next time a student submits the exam.

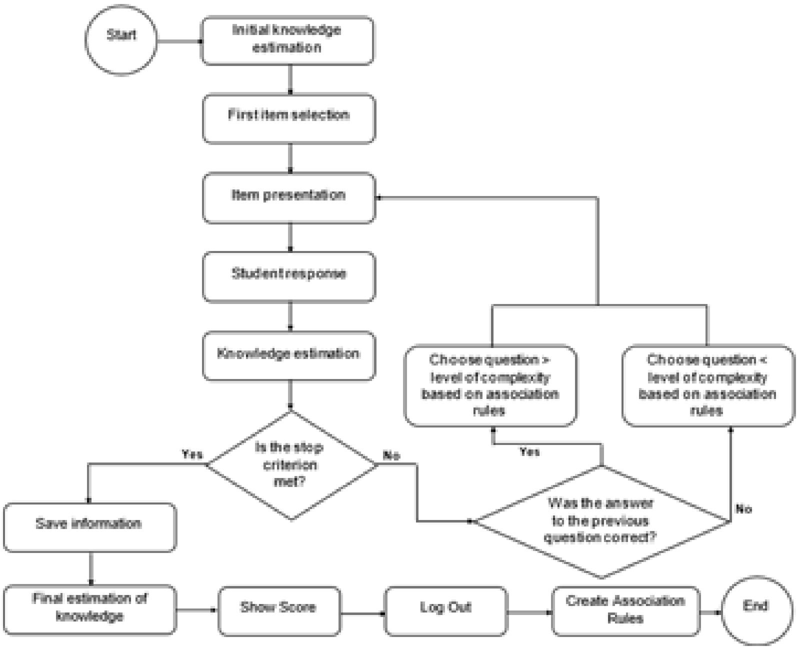

Figure 1. Integration of association rule mining in the item selection phase of the CAT process. Source: author own elaboration.

Collection and Preparation of Data

Information of tests on pencil and paper corresponding to three units of the Computer Systems Master's Database course were used for the creation of a database in MySQL. From the database records, two binary-matrix were created to serve as the basis for the application of ARM algorithms. In the binary-matrix, questions are represented by the columns and examinees by the rows, where, 1 corresponds to a correct answer and 0 corresponds to an incorrect answer. The first binary-matrix called Exa1 corresponds to the answers of the first unit and includes thirty questions of twenty-five students. The second binary-matrix called Exa2 corresponds to the answers of the second unit and covers thirty questions and twenty-five students. According to the Waikato Environment for Knowledge Analysis —WEKA— tool specifications, the two data sets were analyzed based on their characteristics and it was observed that they did not need any other processing, so they were ready for the next step of the process.

Evaluation of Algorithms

There are several metrics to evaluate association rule mining algorithms, among which are: interest factor, lift, rule interest, conviction, Laplace measure, certainty factor, odds ratio and cosine similarity (Yan, Zhang and Zhang, 2009). There are also some new metrics such as: bi-lift, bi-improve, bi-support (Ju et al., 2015), *Items-based Distance* —ID—, and *Data Rowsbased Distance* (Djenouri et al., 2014). However, the most employees are support and confidence, which are used in this project, adding also the time factor and number of rules.

For the comparative analysis of the association rule algorithms, the following four criteria were evaluated:

1. Confidence: It assesses the degree of certainty of the detected association.

2. Support: It represents the percentage of transactions from the database that the given rule satisfies.

3. Time: Amount of milliseconds that takes the construction of a model.

4. Rules: It represents the number of interesting rules obtained.

The purpose of this comparison process is to identify the algorithm that provides those rules that meet the following search criteria: (i) rules with one antecedent and one consequent, and (ii) rules with a value of consequent equal to 1 (correct answer). For example:

Item5=1 ==> Item6=1 or Item3=0 ==> Item4=1

Where, value 1 means the question had a correct answer and 0 means it had the wrong answer. All of them with the highest levels of confidence and support, found in the shortest possible time.

Four association rule algorithms were applied to the data sets, which are mentioned below:

(i) Apriori (Agrawal, Imielinski and Swam, 1993): It is a classic algorithm for association rule mining. It generates rules through an incremental process that searches for frequent relationships between attributes bounded by a minimum confidence. The algorithm can be configured to run under certain criteria, such as upper and lower coverage limits, and to accept sets of items that meet the constraint, the minimum confidence, and order criteria to display the rules, as well as a parameter to indicate the specific amount of rules we want to show.

(ii) FP-Growth (Han, Kamber and Pei, 2012): It is based on Apriori to perform the first exploration of the data, in which it identifies the sets of frequent items and their support, value that allows us to organize the sets in a descending way. The method proposes good selectivity and substantially reduces the cost of the search, given that it starts by looking for the shortest frequent patterns and then concatenating them with the less frequent ones (suffixes), and thus identifying the longest frequent patterns.

(iii) PredictiveApriori (Scheffer, 2001): The algorithm achieves a favorable computational performance due to its dynamic pruning technique that uses the upper bound of all rules of the supersets of a given set of elements. In addition, through a backward bias of the rules, it manages to eliminate redundant ones that are derived from the more general ones. For this algorithm, it is necessary to specify the number of rules that are required.

(iv) Tertius (Flach and Lachiche, 2001): It performs an optimal search based on finding the most confirmed hypotheses using a no redundant refinement operator to eliminate duplicate results. The algorithm has a series of configuration parameters that allow its application to multiple domains.

For a better understanding in the comparison process, the algorithms were grouped based on their characteristics, so a comparison was carried out between Apriori and FP-Growth, since both allow to set different values for the confidence (*min_conf*) and support (*min_sup*) to obtain 4 different groups of rules (15, 20, 25, and 50) with one antecedent and one consequent, where the value of the latter is equal to 1. Getting in response for each case, the time in milliseconds consumed in the execution of the algorithm, the confidence and the support. While for the comparison of Predictive Apriori and Tertius, it was also necessary to specify the number of rules required, obtaining as a response for each case, the time in milliseconds used, as well as the confidence and support. For comparison between the four algorithms, the number of rules generated, time spent and support are taken into account. Each evaluation was executed 100 times to estimate the average time for the construction of the models. Also, the average values of support and confidence were considered.

Results and Discussion

Tables 1 and 2 show the comparison between Apriori and FP-Growth for the Exa1 and Exa2 data sets.

Table 1. Test results for Apriori and FP-Growth for the Exa1 data set

		15 Rules			20 Rules			25 Rules			50 Rules		
Algorithms	Min_conf/ Min_sup	Conf.	Sup.	Time	Conf.	Sup.	Time	Conf.	Sup.	Time	Conf.	Sup.	Time
Apriori	0.7/0.5	1	0.96	3	1	0.94	5	0.99	0.93	4	0.99	0.87	9
FP-Growth		1	0.96	5	0.98	0.95	5	0.99	0.93	5	-	-	-
Apriori	0.7/0.6	1	0.96	6	1	0.94	6	0.99	0.93	8	0.99	0.87	17
FP-Growth		1	0.96	7	0.98	0.95	4	0.99	0.93	6	-	-	-
Apriori	0.8/0.3	1	0.96	5	1	0.94	5	0.99	0.93	5	0.99	0.87	14
FP-Growth		1	0.96	6	0.98	0.95	6	0.99	0.93	6	-	-	-
Apriori	0.8/0.6	1	0.96	5	1	0.94	6	0.99	0.93	6	0.99	0.87	10
FP-Growth		1	0.96	3	0.98	0.95	4	0.99	0.93	6	-	-	-
Apriori	0.9/0.5 0.9/0.9	1	0.96	6	1	0.94	5	0.99	0.93	4	0.99	0.87	24
FP-Growth		1	0.96	5	0.98	0.95	4	0.99	0.93	6	-	-	-
Apriori		1	0.96	7	0.98	0.95	4	-	-	-	-	-	-
FP-Growth		-	-	-	-	-	-	-	-	-	-	-	-

Source: author own elaboration.

Table 2. Test results for Apriori and FP-Growth for the Exa2 data set

		15 Rules			20 Rules			25 Rules			50 Rules		
Algorithms	Min_conf/ Min_sup	Conf.	Sup.	Time	Conf.	Sup.	Time	Conf.	Sup.	Time	Conf.	Sup.	Time
Apriori	0.7/0.5	0.99	0.90	3	0.97	0.90	3	0.98	0.88	6	0.97	0.85	6
FP-Growth		0.99	0.90	4	0.97	0.90	5	0.99	0.86	7	-	-	-
Apriori	0.7/0.6	0.99	0.90	4	0.97	0.90	4	0.98	0.88	5	0.97	0.85	8
FP-Growth		0.99	0.90	5	0.97	0.90	5	0.99	0.86	7	-	-	-
Apriori	0.8/0.3	0.99	0.90	5	0.97	0.90	3	0.98	0.88	7	0.97	0.85	7
FP-Growth		0.99	0.90	6	0.97	0.90	5	0.99	0.86	6	-	-	-
Apriori	0.8/0.6	0.99	0.90	7	0.97	0.90	4	0.98	0.88	5	0.97	0.85	7
FP-Growth		0.99	0.90	6	0.97	0.90	4	0.99	0.86	6	-	-	-
Apriori	0.9/0.5	0.99	0.90	5	0.97	0.90	4	0.98	0.88	4	0.97	0.85	6
FP-Growth		0.99	0.90	4	0.97	0.90	7	0.99	0.86	7	-	-	-

Source: author own elaboration.

As it is observed in Table 1, Apriori obtained 15 and 25 rules faster than the latter in more cases. While FP-Growth was faster to find 20 rules, the rules found by Apriori had greater confidence. Moreover, Apriori was the only algorithm that obtained 15 and 20 rules, considering a value of 0.9 for *min_conf* and *min_sup*, respectively, and 50 rules five times. Therefore, based on the Time, Confidence and Rules criteria, Apriori is better than FP-Growth for the Exa1 data set.

Likewise, Table 2 shows that Apriori is faster than FP-Growth for 15, 20 and 25 rules. For the group of 25 rules, although FP-Growth has a higher level of confidence in all cases, Apriori has a higher level of support in them. In addition, for the group of 50 rules, Apriori was the only algorithm that obtained rules in five cases. Therefore, based on the Time, Support and Rules criteria, Apriori is also better than FP-Growth for the Exa2 data set.

The comparisons between PredictiveApriori and Tertius for the Exa1 and Exa2 data sets are shown in Table 3 and 4, respectively; as it is observed, although PredictiveApriori's support and confidence is higher, the time that it uses in rule creation is sufficient factor to discard it, because the system should occupy as little time as possible in generating rules that are the basis for selecting the next item.

Figures 2 to 5 show the comparison between the four algorithms with regard to support and time, respectively. The results indicate that the algorithm that generates rules with better support within the Exa1 and Exa2 data sets and in less time is Apriori. Therefore, it is the best algorithm for the data sets.

The product of the two analyses carried out in this section allows to determine that the Apriori algorithm is the one that presents the best results in each of the data sets. For example, in the Exa1 data set, one of the rules generated with a confidence equal to 1 and a support equal to 1 is as follows:

Item13=1 ==> Item7=1

Table 3. Test results for PredictiveApriori and Tertius for the Exa1 data set

Algorithms	Rules	Confidence	Support	Time
PredictiveApriori	15	1	0.90	16959
Tertius	15	0.81	0.41	22
PredictiveApriori	20	1	0.87	48558
Tertius	20	0.79	0.40	29
PredictiveApriori	25	1	0.84	22674
Tertius	25	0.79	0.41	25
PredictiveApriori	50	1	0.73	58364
Tertius	50	0.79	0.43	47

Source: author own elaboration.

Table 4. Test results for PredictiveApriori and Tertius for the Exa2 data set

Algorithms	Rules	Confidence	Support	Time
PredictiveApriori	15	1	0.79	18223
Tertius	15	0.82	0.44	33
PredictiveApriori	20	1	0.73	8539
Tertius	20	0.82	0.45	34
PredictiveApriori	25	1	0.68	8528
Tertius	25	0.79	0.44	26
PredictiveApriori	50	0.99	0.69	13706
Tertius	50	0.79	0.41	52

Source: author own elaboration.

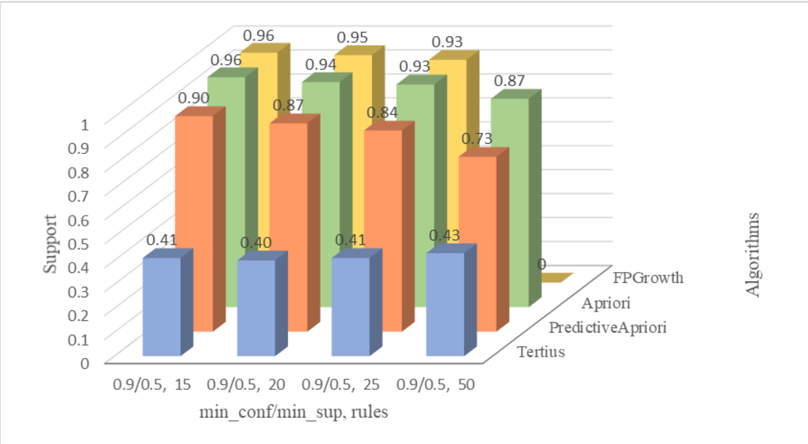

Figure 2. Comparison of Apriori, FP-Growth, PredictiveApriori and Tertius algorithms in terms of support for Exa1 data. Source: author own elaboration.

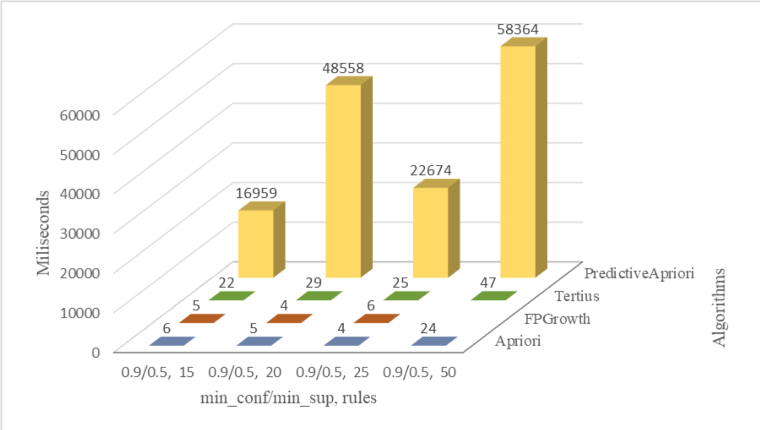

Figure 3. Comparison of Apriori FP-Growth, PredictiveApriori and Tertius algorithms in terms of time for Exa1 data set. Source: author own elaboration.

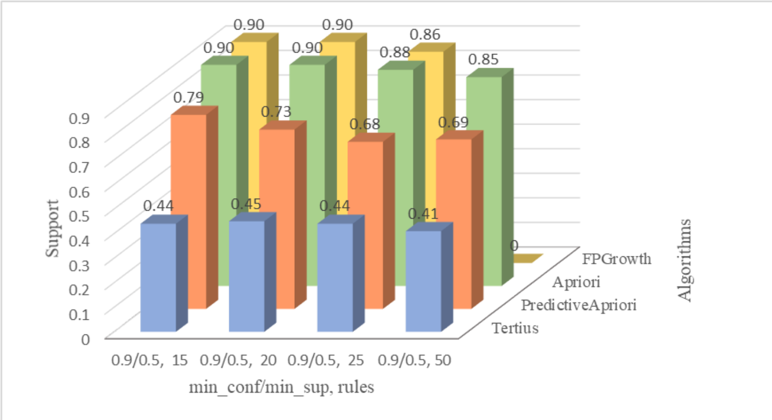

Figure 4. Comparison of Apriori, FP-Growth, PredictiveApriori and Tertius algorithms in terms of support for Exa2 data. Source: author own elaboration.

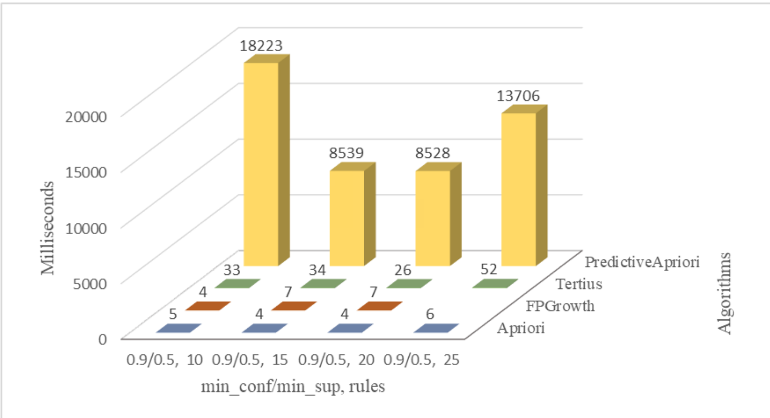

Figure 5. Comparison of Apriori FP-Growth, PredictiveApriori and Tertius algorithms in terms of time for Exa2 data set. Source: author own elaboration.

This indicates that all the twenty-five students who answered question 13 well, they also answered question 7 well.

Another example is found in the Exa2 data set, where the algorithm generated a rule with a confidence of 1 and a support of 0.88:

Item48=1 ==> Item54=1

That means that the pattern appears in 88 % of total transactions, that is, in 22 of the 25 tests. And that every time the students correctly answered question 48, their answer to question 54 was correct. The method of selection of items in the CAT system that is being developed, will use these patterns to determine what items present to students, according to their previous answer, and based on the past experiences of other students who have presented the same exam.

Conclusions

This paper shows the complete process of the proposed CAT system and the comparison of four association rule mining algorithms applied to two data sets in order to find the most suitable one to implement it as a selection method in a CAT system. With the results obtained it can be concluded that the Apriori algorithm has the greatest advantages compared to FP-Growth, PredictiveApriori and Tertius, since it obtained rules with good support, confidence and in less time,although the first two criteria are very important to select interesting rules, the last criterion is critical for this work because the system must occupy as little time as possible in generating rules that will serve for the selection of the next item in the test that is being presented by the evaluated.

In the future, these results will serve to develop and implement a CAT system that uses association rule mining as an item selection criterion, which will be tested in student's master's level in order to compare the estimation of knowledge of the evaluated when presenting physical tests against electronic and adaptive examinations, all in order to verify the effectiveness of the developed system.

References

Agrawal, R., Imielinski, T. and Swam, A. (1993). Mining association rules between sets of items in large databases. In Proceedings of the ACM SIGMOD International Conference on Management of Data.

Albano, A. et al. (2019). Computerized Adaptive Testing in Early Education: Exploring the Impact of Item Position Effects on Ability Estimation. *Journal of Education Measurement*, *56*(2), 437-451.

Bengs, D., Brefeld, U. and Krohne, U. (2018). Adaptive Item Selection Under Matroid Constraints. *Journal of Computerized Adaptive Testing*, *6*(2), 15-36.

Chen, Y. et al. (2017). Research on CAT Unified Model Based on Cognitive Diagnosis Theory. In Proceedings of the 6th International Conference on Information Engineering.

Chen, J.-H., Chao, H.-Y. and Chen, S.-Y. (2019). A Dynamic Stratification Method for Improving Trait Estimation in Computerized Adaptive Testing Under Item Exposure Control. *Applied Psychological Measurement*, *44*(3), 182-196.

Dahdouh, K. et al. (2019). Association Rules Mining Method of Big Data for E-Learning Recommendation Engine. *Advanced Intelligent Systems for Sustainable Development*, *5*, 477-491.

Djenouri, Y. et al. (2014). An Efficient Measure for Evaluating Association Rules. In 6th International Conference of Soft Computing and Pattern Recognition (SoCPaR), Tunis, Tunisia.

Du, Y., Li, A. and Chang, H.-H. (2018). Utilizing Response Time in On-the-Fly Multistage Adaptive Testing. *Quantitative Psychology*, 107-117.

Flach, P. and Lachiche, N. (2001). Confirmation-Guided Discovery of First-Order Rules with Tertius. *Machine Learning*, *42*(1), 61-95.

Gu, J., Zhou, X. and Yan, X. (2018). Design and Implementation of Students' Score Correlation

Analysis System. In Proceedings of the 2018 International Conference on Big Data and Education.

Han, J., Kamber, M. and Pei, J. (2012). *Data Mining Concepts and Techniques*. New York, USA: Elsevier.

Ju, C. et al. (2015). A Novel Method of Interestingness Measures for Association Rules Mining Based on Profit. *Discrete Dynamics in Nature and Society*, *4*, 1-10.

Lee, C.-S. et al. (2018). PSO-based Fuzzy Markup Language for Student Learning Performance Evaluation and Educational Application. *Transactions on Fuzzy Systems*, *26*(5), 2618-2633.

Lin, C.-J. and Chang, H.-H. (2019). Item Selection Criteria with Practical Constraints in Cognitive Diagnostic Computerized Adaptive Testing. *Educational and Psychological Measurement*, *79*(2), 335-357.

Miyazahua, Y. and Ueno, M. (2019). Computerized Adaptive Testing Method Using Integer Programming to Minimize Item Exposure. *Advances in Intelligent Systems and Computing*, *11*(28), 105-113.

Prajapati, D., Garg, S. and Chauhan, N. (2017). Interesting association rule mining with consistent and inconsistent rule detection from big sales data in distributed environment. *Future Computing and Informatics Journal*, *2*(1), 19-30.

Rodríguez-Cuadrado, J. et al. (2020). Merged Tree-CAT: A fast method for building precise Computerized Adaptive Tests based on Decision Trees. *Expert Systems with Applications*, *143*, 113066.

Rubio Delgado, E. et al. (2018). Analysis of Medical Opinions about the Nonrealization of Autopsies in a Mexican Hospital Using Association Rules and Bayesian Networks. *Scientific* Programming, *7*, 1-21.

Scheffer, T. (2001). Finding association rules that trade support optimally against confidence. *Principles of Data Mining and Knowledge Discovery*, *9*, 424-435.

Sheng, C., Bingwei, B. and Jiecheng, Z. (2018). An Adaptive Online Learning Testing System. In ICIET 18 Proceedings of the 6th International Conference on Information and Education Technology.

Stafford, R. et al. (2019). Comparing computer adaptive testing stopping rules under the generalized partial-credit model. *Behavior Research Methods*, *51*(3), 1305-1320.

Tokusada, Y. and Hirose, H. (2016). Evaluation of Abilities by Grouping for Small IRT Testing Systems. In 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), Kumamoto, Japan.

Wang, F. et al. (2018). Association Rule Mining Based 1 Quantitative Analysis Approach of Household Characteristics Impacts on Residential Electricity Consumption Patterns. *Energy Conversion and Management*, *171*, 839-854.

Yan, X., Zhang, C. and Zhang, S. (2009). Confidence Metrics for Association Rule Mining. *Applied Artificial Intelligence*, *23*(8), 713-737.

Ye, Z. and Sun, J. (2018). Comparing Item Selection Criteria in Multidimensional Computerized Adaptive Testing for Two Item Response Theory Models. In 3rd International Conference on Computational Intelligence and Applications (ICCIA), Hong Kong, China.

Yigit, H., Sorrel, M. and de la Torre, J. (2019). Computerized Adaptive Testing for Cognitively Based Multiple-Choice Data. *Applied Psychological Measurement*, *43*(5), 388-401.