Association Rules-based Tool for Educational Data-Mining (FERA)

Elsa Ferreira Gomes
Instituto Superior de Engenharia do Porto & INESC TEC, Portugal
efg@isep.ipp.pt

Diogo Manuel Vieira
Instituto Superior de Engenharia do Porto,

Paula Correia Tavares
Instituto Superior de Engenharia do Porto,
pct@isep.ipp.pt

Pedro R. Henriques
Centro Algoritmi & Departamento de Informática, Universidade do Minho, Braga, Portugal
prh@di.uminho.pt

Abstract - **Learn to program is an arduous and complex task that raises many challenges both to teachers and students. Motivation has a direct impact on students' ability to acquire knowledge and consequently affects the results of the teaching / learning process. It is important to understand the causes of lack motivation and their disinterest, to combat them. In this paper we present a tool to support the analysis of student's surveys by enabling the discovering and explorations of frequent patterns (association rules). This tool helps to identify interesting rules, visualize the rules generated using graphics and navigate the space of rules. It is easily extensible, having several types of visualizations, allowing the user to generate reports.**

Keywords – Data Mining, FERA, Motivation, Students.

INTRODUCTION

Several theories have been developed to explain the motivation from the beginning of the history of psychology as a science. Because it is a complex phenomenon, the subject has been studied under different prisms (Williams et al., 2011; Almeida, 2012). It is important to understand the causes of this lack of interest. In fact, there are several reasons why students fail to learn programming (Proulx, 2000). To motivate students is one of the biggest challenges that teachers have to face (Tavares et al., 2017). In programming courses (in higher education) this task is particularly difficult due to the skills needed and the complexity of the teaching topic. Programming demands a new thinking paradigm and a high level of abstraction capabilities. In order to assess the real difficulties that students have in the teaching/learning process of programming, we have designed and conducted in Engineering schools a survey to 1.st year students attending Programming, or Algorithms & Data Structures courses. This was applied to 237 students anonymously from different institutions (Tavares et al., 2018). Our work aims to understand the reasons for the real difficulties that arise in the teaching / learning process of Computer Programming. We use Data Mining techniques to infer association rules between the different factors and the motivation. All data items collected were analyzed using the proposed tool FERA that enables the exploration and visualization of a very large set of association rules.

ASSOCIATION RULES

Data mining has been applied in educational research (Karkhanis et al., 2015; Mohamad et al., 2013). Association Rules mining is a widely used data mining technique (García et.al, 2011). In this paper we use association rule mining (the Apriori algorithm) in a educational data mining.

In a dataset composed by binary variables, it is often useful to find frequent associations between sets of variables. These associations allow us to understand the interactions between any variables and can take the form of frequent set of positive co-occurrences of variables (items) or association rule. An association rule has the form A => B where A and B are frequent sets of items and expresses a tendency to observe B whenever we observe A. The confidence measure of a rule reflects the probability of observing B in a transaction of the dataset knowing that A is also observed in the same transaction. Confidence is calculated empirically from the dataset itself and estimates the aposteriori probability of B. Another important measure that characterizes an association rule A => B is the support, the proportion of transactions that simultaneously contain all items in A and B. Association rules are automatically discovered from a dataset using algorithms such as APRIORI (Agrawal et al., 1994) or FP-Growth (Han et al., 2004). These algorithms easily produce a large number of rules, so other measures of interest of the rules, such as lift. The lift measure calculates the quotient between the aposteriori probability of the consequent of the rule and its apriori probability (Bayardo et al., 1999).

FERA OVERVIEW

Data mining techniques and algorithms have been used on a large scale and in many areas from commerce and services, health and industry. More recently, academic analysis work has emerged using data mining algorithms on student surveys providing important information such as expected performance and retention rates. The proposed

FERA tool intends to help the teacher to better understand the students' learning process and identify suitable improvement actions. This tool supports the analysis of students' surveys and it is compatible with google forms. It applies association rules and aims to help predict student behavior and improve learning performance in computer education. It allows the user to discover association rules, explore them using intuitive menus and visualize them with graphics and charts. FERA is an application developed in R using Shiny and RStudio IDE (Rstudio, 2017). The user starts by loading a survey data file on the platform, which contains data to be analyzed in order to infer rules. The file must be in CSV format. Then, the tool generates rules using the apriori algorithm from arules package, with the parameters selected by the user (support, confidence and maximum length). The user can select views of subsets of the generated rules, produce graphic representations of the variables under analysis and plot the rules in a support/confidence space. The discovered rules and other results can be exported as an HTML file. FERA interface has tabs for the different functionalities with buttons to choose the parameters as shown in Figure 1.

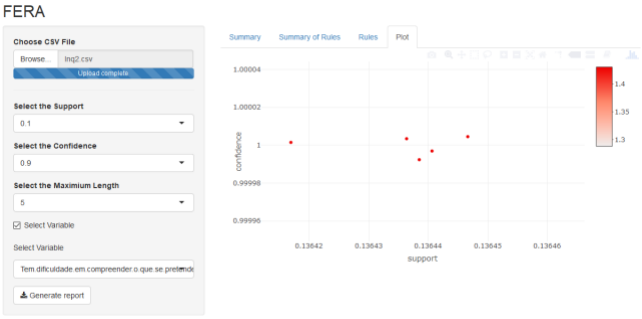

Figure 1 - FERA Association Rules Chart

In summary, the user can: upload a file, define mining parameters, apply filters, plot results, create graphics, select rules and generate files with the final results. The output file is created using the rmarkdown library.

EVALUATION

FERA was evaluated in two ways. First, its ability to support the teacher. Second, its performance. In order to evaluate the capacity of extracting interesting rules, we have used surveys from programming students in engineering schools. The tool was useful to understand the relationship between the student's behavior and motivation. For example, in Figure 2 we can see the rule: "when students don't feel difficulties in programming, they feel motivated for APROG course and prefer to work alone, then they do not find it difficult to understand what is intended in the exercises". In terms of the tool's processing performance of rule extraction we observed that in the current installation (a standard laptop computer) it is able to process more than six million rules in 34 seconds. In a larger set of more than fifteen million rules the tool did not respond in reasonable time.

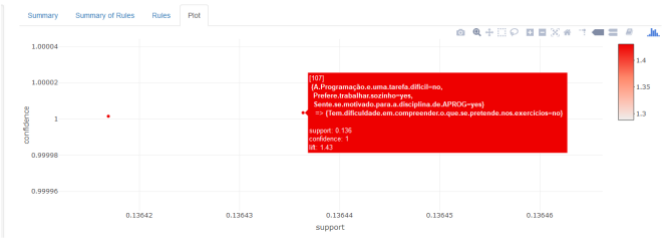

Figure 2 - Rule Representation

CONCLUSIONS

This paper presents a tool that can be useful to process student surveys in general. We have described its application to a case study where the aim was to understand the student's behavior and motivation. We have used surveys from programming students in engineering schools to evaluate the tool capacity to extract interesting rules. FERA was useful to understand the relationship between the student's behavior and motivation.

REFERENCES

Agrawal, R., Srikant, R. (1994). Fast Algorithms for Mining Association Rules. In VLDB-94.

Almeida, D. (2012). A motivação do Aluno no Ensino Superior: um estudo exploratório. Universidade Estadual de Londrina.

Bayardo R., Agrawal R. (1999). Mining the Most Interesting Rules. In Proc. of the 5th International Conference on Knowledge Discovery and Data Mining, KDD-99, pp. 145-153.

García, E., Romero,C., Ventura,S., Castro, C. (2011). A collaborative educational association rule mining tool. In The Int. and Higher Education, 14, Issue 2, 77-88, ISSN 1096-7516.

Han, J., Pei, J., Yin, Y., Mao, R. (2004). Mining Frequent Patterns without Candidate Generation: A Frequent- Pattern Tree Approach. Data Min. Knowl. Discov. 8, 1 (January 2004).

Karkhanis, S. P., Dumbre, S.S. (2015). A Study of Application of Data Mining and Analytics in Education Domain. Int. Journal of Computer Applications, Volume.120, No.22, pp. 23-28.

Mohamad, S. K., Tasir, Z. (2013). Educational Data Mining: A Review. In Procedia - Social and Behavioral Sciences, Volume.97, pp. 320-324, ISSN1877-0428

Proulx, V. (2000). Programming patterns and design patterns in the introductory computer science course. Proceedings of the thirty-first SIGCSE technical symposium on Computer science education, pp. 80-84. New York.

Rstudio, Inc, 2017. Shiny. [Online] Available at: http://shiny.rstudio.com/ [Acedido em 2018].

Tavares, P., Gomes, E., Henriques, P. (2018). Studying Programming Students Motivation using Association Rules. 10th International Conference on Computer Supported Education (CSEDU 2018). Funchal, Portugal.

Tavares, P., Henriques, P., Gomes, E. (2017). A Computer Platform to Increase Motivation in Programming Students – PEP. 9th International Conference on Computer Supported Education (CSEDU 2017). Porto, Portugal.

Williams, K., Williams, C. (2011). Five Key Ingredients for Improving Student Motivation. Research in Higher Education Journal, pp. 104-122.