Class Association Rule Mining from Students' Test Data

Cristóbal Romero [1], Sebastián Ventura[1], Ekaterina Vasilyeva[2] and Mykola Pechenizkiy[2]
cromero@uco.es, sventura@uco.es, e.vasilyeva@tue.nl, m.pechenizkiy@tue.nl
[1]Department of Computer Science, Córdoba University, Córdoba, Spain
[2] Department of Computer Science, Eindhoven University of Technology, The Netherlands

Abstract. In this paper we propose the use of a special type of association rules mining for discovering interesting relationships from the students' test data collected in our case with Moodle learning management system (LMS). Particularly, we apply Class Association Rule (CAR) mining to different data matrices such as the *score-matrix*, the *relationship-matrix* and the *knowledge-matrix*. These matrices are constructed based on the data relate to students' performance in the test and on the domain knowledge provided by the instructor. We describe how to obtain these matrices and then we have applied a CAR mining algorithm.

1 Introduction

A Class Association Rule (CAR) is a special type of Association Rule (AR) that describes an implicative co-occurring relationship between a set of items and a pre-defined class, expressed in the form of an "IF antecedent (input-attributes) THEN consequent (class)" rule [1]. AR mining finds all rules that satisfy some minimum support and minimum confidence constraints, that is, the target of mining is not predetermined. However, in CAR mining there is one and only one pre-determined target, i.e., the class. So, CAR is a type of target-constraint association rule. Such kind of *focused* rules mining results in a set of independent and comprehensible rules having one (desired) element in the consequent. Such rules usually represent discovered knowledge at a high level of abstraction and can be used directly in the decision making process.

Modern assessment tools and testing systems in particular allow accumulating a lot of useful performance and usage related data (possibly at different levels of granularity). This may include (but is not limited to) actual students' answers and their correctness, final scores, used/execution time (total and for each question) and some statistics about items/questions such as Facility Index or % Correct (F.I.), Standard Deviation (S.D.), Discrimination Index (D.I.), etc. Although in many cases, still only the final scores are used by an instructor to evaluate students' knowledge or performance [2], with recent developments in educational data mining research different ideas for intelligent analysis of assessment data were proposed. In this paper, we show the potential utility of applying CAR mining over the test-related data for providing an instructor interesting relationships discovered from these data presented in the *score-matrix* and *knowledge-matrix*. This information can be turned by an instructor into valuable knowledge for making decision on how to improve both the test and the course.

2 Experimentation

We have used data collected with Moodle's quiz module tool. Starting from these data and some background information provided by the instructor we have created three different data matrices (see Figure 1).

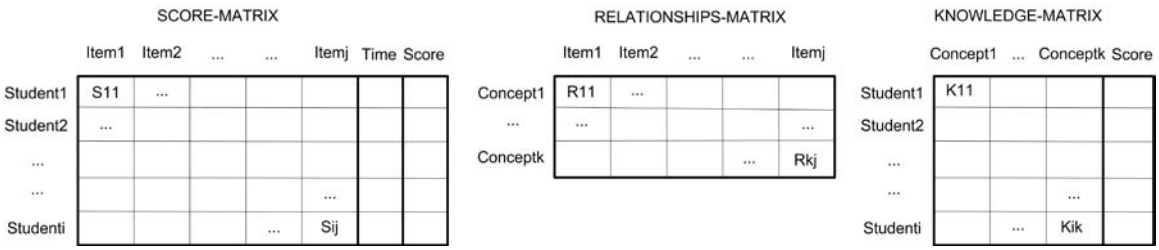

Figure 1. Matrices created from test's data.

We have applied the Apriori-CAR mining algorithm over the previously described data matrices. In the first experiment, we have used the *score-matrix*, and we have selected/filtered as input-attributes (antecedent) only the item answers, and as class the final score. In this way, we can see the relationships between items and how they can predict/determine the final score obtained by students. In the second experiment, we have used the *knowledge-matrix*, and we have selected the knowledge of concepts as input-attributes (antecedent) and the final score as a class attribute. In this way, we can discover the relationships between concepts and between the level of knowledge of these concepts and the final score obtained by students.

3 Conclusions and Future Work

In this paper, we proposed to use a special type of association rules over the assessment data in a particular scenario. We mined different test data matrices rather than only the typical score-matrix. Particularly we used an item-concept relationship matrix created by the instructor and a student-concept knowledge level matrix automatically created based on the information from the other two matrices. Finally, it is important to notice that concepts themselves may need to be presented as a hierarchy rather than a 'flat' set of independent concepts. Mining interesting patterns in such settings is one of the directions of our further work.

References

[1] Liu, B., Hsu, W, Ma, Y. Integrating classification and association rule mining. *Int. Conference on Knowledge Discovery and Data Mining*, KDD'98, pp. 80-86.

[2] Romero, C., Ventura, S., Salcines, E. Data mining in course management systems: Moodle case study and tutorial. *Computer & Education*, 2008, 51(1), pp. 368-384.