Mining Coherent Evolution Patterns in Education through Biclustering

André Vale
Instituto Superior Técnico
Av. Rovisco Pais 1
1049-001 Lisboa, Portugal
andre.vale@tecnico.ulisboa.pt

Sara C. Madeira
Instituto Superior Técnico
Av. Rovisco Pais 1
1049-001 Lisboa, Portugal
sara.madeira@tecnico.ulisboa.pt

Cláudia Antunes
Instituto Superior Técnico
Av. Rovisco Pais 1
1049-001 Lisboa, Portugal
claudia.antunes@tecnico.ulisboa.pt

ABSTRACT
With the spread of information systems and the increased interest in education, the quantity of data about education has exploded along with a new field - Educational Data Mining. Predicting students' performance has been approached by several techniques, but the combination of supervised and non-supervised techniques appeared as a new tool for improving the results. Biclustering algorithms have been successfully applied in areas such as gene expression data and information retrieval, but not used in the educational context. In this paper, we show how to apply biclustering techniques to educational data and to use its results as features to improve the prediction of student's performance.

Keywords
Educational Data Mining, Biclustering, Student's Performance, Coherent Evolution Patterns

1. INTRODUCTION
The prediction of students' performance has deserved a significant attention in Educational Data Mining (EDM) research, with several distinct approaches being proposed, mostly using classification and regression techniques. With the advances and stabilization of these techniques, it is easy to accept that the accuracy results do not depend on the technique used, but on data themselves, both on the training data and on the target variable [8].

While classification tries to find a model to predict an outcome, non-supervised techniques, as pattern mining and clustering, are able to explore the data for identifying frequent behaviors. Previous studies [1, 2] have shown that sequential pattern mining is suited to discover patterns able to model students behaviors, which in turn can be used to enrich training data, improving global classification accuracy on more than 10% [2].

Clustering is perhaps one of the most important tools for both exploratory and confirmatory analysis. Indeed, it is a technique to discern meaningful patterns in unlabeled data by grouping together data points that are similar. Biclustering algorithms [5] are a recent alternative to traditional clustering methods that allows the discovery of local patterns rather than global ones. Besides discovering sequential patterns identified by pattern mining algorithms, biclustering is able to discover other sequential patterns that reveal coherent evolutions [3, 6].

Although both the literature on EDM and biclustering topics are vast, and the results are positive, the combination of these topics is almost nonexistent. Only recently Trivedi et al. [9] applied this technique to education. In their work, they used the idea of co-clustering (namely biclustering) students and their tutor interaction features and interleave it with a bagging strategy which they used previously with clustering [10] for prediction of out-of-tutor performance of students. The results obtained were better than the baseline and also indicated that the dynamic assessment condition returns in a much better prediction of student test scores when compared to the static condition. However, they used one of the most basic techniques of biclustering, using k-means clustering algorithm to cluster students and features (rows and columns). Separately and then combining both clustering results to derive biclusters. This clustering combination is probably the reason why they obtained modest improvements compared with their previous clustering works.

In this paper, we propose to explore biclustering to discover new patterns in educational data and make use of these patterns to enrich training data in order to improve the prediction of students' performance.

2. BICLUSTERING FOR EDM
Biclustering can be applied whenever the data to analyze has the form of a real-valued or symbolic matrix A, where the value a_{ij} represents the relation between row i and column j, and the goal is to identify subsets of rows with certain coherence properties in a subset of the columns. The goal of biclustering algorithms is to identify a set of biclusters. Let A be a matrix defined by its set of rows, R, and its set of columns, C. Then we can define a bicluster $B = (I, J)$ as a submatrix A_{IJ} defined by $I \subseteq R$, a subset of rows, and $J \subseteq C$, a subset of columns [5]. This set of biclusters $B_k = (I_k, J_k)$ satisfies specific characteristics of homogeneity, that can be grouped in four categories: a) Biclusters with constant values; b) Biclusters with constant values on rows or columns; c) Biclusters with coherent values; and d) Biclusters with coherent evolutions. The first three classes analyze directly the numeric values in the data matrix and try to find subsets of rows and subsets of columns with similar behaviors. The fourth class aims to find coherent behaviors regardless of the exact numeric values in the data matrix. The type of patterns in a) and c) can be found using pattern mining, but the ones in b) and d) are not. The work by Madeira and Oliveira [5] presents a deep survey on this topic, describing the most important algorithms.

Studies have demonstrated that sequential pattern mining can be successfully applied for mining students [1] and teachers' frequent behaviours [2], which in turn may enrich training data for improving classification. As explained, biclustering is able to identify more patterns than pattern mining, in particular, patterns that reveal coherent evolutions. In this manner, we propose to explore biclustering algorithms for identifying patterns that may improve the classification task, as previously performed with sequential pattern mining [2].

In the educational context, matrix A targeted by biclustering algorithms can be any matrix relating two distinct entities, whose relation can be measured, and expresses some result. For example, a matrix relating students and subjects through achieved marks, where we might be interested on finding a group of students that shows the same evolution in a particular subset of subjects, but also a matrix for subjects and time, reporting the average performance of students enrolled on the subject in a particular term.

There are a large number of existing approaches for biclustering that finds different types of biclusters and thus different types of frequent patterns. In our case we are interested in biclusters with coherent evolutions, since existing pattern mining approaches are not able to identify them. In this work, we use the algorithms most cited in the literature to find these types of patterns, namely *Bimax* [7], *xMOTIFs* [6], *ISA* [4] and *OPSM* [3].

Table 1 provides an example of a mark matrix with 10 students and 7 subjects where we draw examples of the types of biclusters these algorithms can find. Bicluster B_1 presents students who had the same marks on all the subjects - a bicluster with constant values that the *Bimax* can get; B_2 has students who have constant marks on different subjects – a bicluster with coherent values on the rows found by *xMOTIFs*; B_3 has students that have all the same notes on the same subjects – a bicluster with coherent values on the columns found by ISA; and finally B_4 shows students that have a coherent evolution between subjects, in this particular case, students' marks satisfy the following evolution pattern: *Subject 3 < Subject 2 < Subject 1 < Subject 4* - a bicluster with coherent evolutions that *OPSM* can find.

Table 1. Example of different types of biclusters in a matrix with marks of ten students at seven subjects.

	Subject 1	*Subject 2*	*Subject 3*	*Subject 4*	*Subject 5*	*Subject 6*	*Subject 7*
Student 1	18 B_2	12	18	10	12	18	12
Student 2	17	16	17	13	15	17	16
Student 3	19	16	19	17	16	19	15
Student 4	17	18	17	B_1 19	19	16	17
Student 5	B_4 18	16	14	19	19	19	18
Student 6	17	15	13	19	14	17	16
Student 7	19	18	16	20	B_3 17	18	19
Student 8	12	16	13	14	17	18	19
Student 9	16	18	15	16	17	18	19
Student 10	13	14	10	11	13	12	10

3. CASE STUDY AND CONCLUSIONS

The data used was gathered from a graduation program (LEIC) at Instituto Superior Técnico (IST), Universidade de Lisboa, considering the student records between 1997 and 2012. The program has the duration of three years (6 semesters) with 30 subjects, and after it, students usually follow to the master program (MEIC). The task was to predict the marks of LEIC students when finishing MEIC. In order to achieve our goal, we analyzed a matrix (*students x subjects*) with 443 students and 20 subjects from LEIC, with students' marks in the cells - numbers between 10 and 20. By applying biclustering algorithms mentioned before to the matrix, we obtained 16 biclusters with *OPSM*, 975 with *xMotifs*, 308 with *ISA* and 39 with *Bimax*. In addition to the data in the matrix, we appended a class label (the mark obtained at the end of the master's program - Fair, Good and Very Good) and obtained our training dataset (baseline). As in [2], a new dataset can be obtained from the previous one, enlarged by *k* Boolean attributes, one for each bicluster. Each bicluster attribute is then filled with the true value whenever the bicluster has the student instance and false otherwise. Classification was performed by *Weka*, with decision trees, using cross-validation with 10 folds for the performance evaluation of decision trees. We then use a feature selection (FS) method, wrapper, to obtain the best attributes.

Without FS and biclusters, we got 52.9% of correctly classified instances. If we add the biclusters, the precision go down to 51.8%, this happens because the model is starting to overfitting. Using FS without biclusters we had a precision of 60.1%, and if we add the biclusters we obtained 65.8% (**Figure 1**).

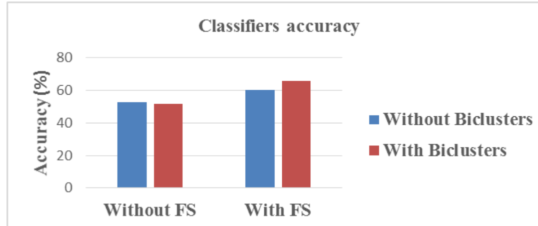

Figure 1. Precision of classifiers accuracy.

With this study we demonstrated that we can improve the accuracy of the decision tree model by more than 5% not having any strict condition to choose the biclusters that are more interesting to use. As such, we believe that after applying more effective metrics to choose the biclusters according to their quality, we can achieve an even better model accuracy. For future work we will develop metrics to apply automated techniques to choose the best biclusters, so we can distinguish the biclusters of interest regardless of what we have in the rows and columns of the matrix.

ACKNOWLEDGMENTS

This work is partially supported by Fundação para a Ciência e Tecnologia under research project *educare* (PTDC/EIA-EIA/110058/2009).

REFERENCES

[1] Antunes, C. 2008. Acquiring Background Knowledge for Intelligent Tutoring Systems. *EDM*. (2008).

[2] Barracosa, J. and Antunes, C. 2011. Anticipating teachers' performance. *KDD 2011 Workshop*. (2011).

[3] Ben-Dor, A. and Chor, B. 2003. Discovering local structure in gene expression data: the order-preserving submatrix problem. *Journal of computational*. (Jan. 2003).

[4] Bergmann, S. et al. 2003. Iterative signature algorithm for the analysis of large-scale gene expression data. *Physical review E*. (2003).

[5] Madeira, S. and Oliveira, A. 2004. Biclustering algorithms for biological data analysis: a survey. *Biology and Bioinformatics, IEEE*. (2004).

[6] Murali, T. and Kasif, S. 2003. Extracting conserved gene expression motifs from gene expression data. *Pacific Symposium on Biocomputing*. (Jan. 2003).

[7] Prelić, A. et al. 2006. A systematic comparison and evaluation of biclustering methods for gene expression data. *Bioinformatics (Oxford, England)*. (2006).

[8] Romero, C. and Ventura, S. 2010. Educational Data Mining: A Review of the State of the Art. *Trans. Sys. Man Cyber Part C*. (2010).

[9] Trivedi, S. et al. 2012. Co-Clustering by Bipartite Spectral Graph Partitioning for Out-of-Tutor Prediction. *International Educational Data Mining Society*. (2012).

[10] Trivedi, S. et al. 2011. Spectral Clustering in Educational Data Mining. *Educational Data Mining* (2011).