Comparing examination results and courses evaluation: a data mining approach DIDAMATICA 2013

R. Campagni, D. Merlini, R. Sprugnoli, M.C. Verri

Università degli Studi di Firenze

Dipartimento di Statistica, Informatica, Applicazioni 'G. Parenti'

Viale Morgagni, 65 - 50134 Firenze

renza.campagni@unifi.it

donatella.merlini@unifi.it

renzo.sprugnoli@unifi.it

mariacecilia.verri@unifi.it

This paper presents an analysis about the comparison between the exam results and the corresponding courses evaluation of university students. The analysis concerns students and courses of the Computer Science program at the University of Florence from 2001/2002 to 2007/2008 academic years. Before the end of each course, students evaluate different aspects of the course, such as the organization and the teaching. We collected and reorganized in an appropriate way evaluation data and the results obtained by students in terms of grades and delays with which they take their exams. Then we used clustering techniques to analyze these data thus showing that there is a correlation between the evaluation of a course and the corresponding average results.

1. Introduction

The evaluation of university education was introduced in Italy at the beginning of the nineties. Successively, in 2006, a National Agency has been constituted for the evaluation of Universities and research Institutes (ANVUR) that defines criteria and methodologies to execute evaluation and to evaluate processes, results and products of education and research. The results of this activity are used to assign resources to the Universities. After a period of experimentation, starting in 2013 the national evaluation system shall be applied to all Universities.

The results of evaluation can be used in the programming and management of the educational activities by monitoring resources (financial, human, structural and others), services (orientation for students and administrative offices), students careers, courses and occupancy rate. In order to evaluate all these aspects, it is important to analyse the opinion of the “users” of university education, i.e. the students.

The evaluation of the educational process falls in the context of the Educational Data Mining (EDM), an emerging and interesting research area that produces useful, previously unknown regularities from educational databases for better understanding and improving the educational performance and assessment of the student learning process [5]. It exploits statistical, machine learning and data mining algorithms over the different types of educational data. Its main objective is to analyse these types of data in order to resolve educational research issues. EDM is concerned with developing methods to explore data in educational settings and, using these methods, to better understand students and the settings in which they learn.

In this paper, we use a data mining approach to compare the evaluation of courses taken from students with their results, in terms of average grade and delay, in the corresponding exams. This study follows an approach similar to that presented in [1, 2, 3]. Our analysis refers to a real case study concerning the University of Florence but it could be applied to different scenarios, except for a possible reorganization of the involved data.

2. Courses evaluation

At the University of Florence, starting from the academic year 2001/2002, a database stores information about evaluation of the courses quality of various degree programs, among which we find the Computer Science degree. The results of this process are available at the address [8], under permission of the involved teacher, and show for each course several information, such as the name of the teacher who took the course and the average rating given by students on various topics. Before the end of each course (at about 2/3 of the course), students compile, anonymously, a module to express their opinion on the course just taken. This form is divided into the following five paragraphs:

- *paragraph 1,* concerns the organization of the degree program;
- *paragraph 2,* concerns the organization of the course;
- *paragraph 3,* concerns the teacher;
- *paragraph 4,* concerns classrooms and equipment;
- *paragraph 5,* concerns the general satisfaction about the course.

Each paragraph is composed by some questions; students could choose among four levels of answers, two negative and two positive levels (“*disagree”, “slightly disagree”, “slightly agree”, “agree”*). These four levels of answers are associated to the numerical values 2, 5, 7 and 10, respectively. For details the interested reader can see the sample of the module in [8]; an excerpt of it is illustrated in Table 1.

For each course and for each paragraph, during the academic years from 2001/2002 to 2007/2008, we computed the percentage of positive answers, that is, of type *"slightly agree" and "agree".* In particular, we grouped together all questions belonging to the same paragraph and their average numerical value.

Modalità di risposta	☹	☹	☺	☺
Organizzazione del corso di studi **1. Il carico di lavoro complessivo degli insegnamenti ufficialmente previsti nel periodo di riferimento** *(bimestre, trimestre, semestre, ecc.)* **è accettabile?** **2. L'organizzazione complessiva** *(orario, esami, intermedi e finali)* **degli insegnamenti ufficialmente previsti nel periodo di riferimento** *(bimestre, trimestre, semestre, ecc.)* **è accettabile?** **3. L'orario delle lezioni è congegnato in modo tale da consentire un'adeguata attività di studio?**				
Organizzazione dell'insegnamento **4. Il carico di studio di questo insegnamento è proporzionato ai crediti assegnati?** **5. Il materiale didattico** *(indicato o fornito)* **è adeguato per lo studio della materia?** **6. Le attività didattiche integrative** *(esercitazioni, laboratori, seminari, ecc.)* **risultano utili ai fini dell'apprendimento?** *(se non sono previste attività didattiche integrative, rispondete non previste)* **7. Le modalità di esame sono state definite in modo chiaro?** **Aspetti relativi alla docenza** **8. Gli orari di svolgimento dell'attività didattica sono rispettati?** **9. Il personale docente è effettivamente reperibile per chiarimenti e spiegazioni?** **10. Il docente stimola / motiva l'interesse verso la disciplina?** **11. Il docente espone gli argomenti in modo chiaro?** **12. Il docente è disponibile ed esauriente in occasione di richieste di chiarimento?** **Aule ed attrezzature** **13. Le aule in cui si svolgono le lezioni sono adeguate** *(si vede, si sente, si trova posto)***?** **14. I locali e le attrezzature per le attività didattiche integrative** *(esercitazioni, laboratori, seminari, ecc.)* **sono adeguati?** *(se non sono previste attività didattiche integrative, rispondete non previste)* **Informazioni aggiuntive e soddisfazione** **15. Le conoscenze preliminari possedute sono risultate sufficienti per la comprensione degli argomenti trattati?** **16. Gli argomenti trattati sono risultati nuovi o integrativi rispetto alle conoscenze già acquisite?** **17. Sei interessato agli argomenti dell'insegnamento?** **18. Sei complessivamente soddisfatto dell'insegnamento?**				

Table 1 – An excerpt of the feedback module (in Italian)

2.1 The career of students

The academic degree under consideration was structured in three years under the D.M. n. 509/1999 and every academic year was organized in two semesters. There were several courses in each of these six semesters and at the end of a semester students could take their examinations. Exams could be taken in different sessions during the same year, after the end of the corresponding courses, or later. Table 2 illustrates an example of students data after a preprocessing phase which allowed us to integrate original attributes, such as the grade and the date of the exam, with both the semester in which the course was given, `Semester1`, and the semester in which the exam was taken, `Semester2`. Finally, we computed the value `Delay` as the difference between the semester of the course and the semester in which the student took the exam.

Student	Exam	Date	Grade	Semester1	Semester2	Delay
100	105035	14/01/01	24	1	1	0
100	105039	20/12/02	27	2	3	1
200	105039	04/06/02	21	2	2	0
30	105035	29/01/01	26	1	3	2
...	...	...		...	...	...

Table 2 – A sample of students data

To relate data of students careers with courses evaluation, for each course, we computed the average grade and the average delay attained by students who took the exam in the same year. An example of this data organization is illustrated in Table 3.

As already observed, the evaluation of courses is anonymous and is done only by students who really take the course. Since we do not know which students evaluated the various courses, we decided to compare the results of courses evaluation in a specific year with the results of students who took the corresponding exams in the same period.

Exam	Year	AvgGrade	AvgDelay
105035	2001	25	1
105039	2002	24	0,5
...	...	...	...

Table 3 – Data of Table 2 aggregated in terms of exam code and year

2.2 Data for analysis

The final dataset for our analysis contains, for each year and for each exam, the following fields:

- **`Exam`**, the code which identifies an exam;
- **`Year`**, the year of the evaluation;
- **`AvgGrade`**, the average grade of the exam;
- **`AvgDelay`**, the average delay, in semesters, of students exams;
- **`Par`*k***, the percentage of positive evaluations of paragraph ***k***;
- **`AvgPar`*k***, the average value of evaluations of paragraph ***k.***

It is important to observe again that we considered information concerning exams of students who may not be the same students who evaluated the courses. Table 4 illustrates a sample of the dataset we analysed.

Exam	Year	AvgGrade	AvgDelay	Par1	AvgPar1		Par5	AvgPar5
105035	2001	25	1	51	4,76		60	5,57
105039	2002	24	0,5	62	6,13		59	6,34
...	...	...	...	...	...	...	...	...

Table 4 – Data organization for clustering analysis

2.3 Clustering analysis with Weka

Among the different data mining techniques, clustering is one of the most widely used methods. The goal of cluster analysis is to group together objects that are similar or related and, at the same time, are different or unrelated to the objects in other clusters. The greater the similarity (or homogeneity) is within a group and the greater the differences between groups are the more distinct the clusters are. *K-means* is a very simple and well-known algorithm based on a partitional approach; it was introduced in [4] and a detailed description can be found in [6]. In this algorithm, each cluster is associated with a *centroid* and each point is assigned to the cluster with the closest centroid by using the Euclidean distance. The number K of clusters must be specified.

The evaluation of the clustering model resulting from the application of a cluster algorithm is not a well developed or commonly used part of cluster analysis; nonetheless, cluster evaluation, or cluster validation, is important to measure the goodness of the resulting clusters, for example to compare clustering algorithms or to compare two sets of clusters. In our analysis we measured cluster validity with *correlation*, by using the concept of *proximity matrix* and *incidence matrix*. Specifically, after obtaining the clusters by applying *K-means* to a dataset, we computed the proximity matrix $P=(P_{ij})$ having one row and one column for each element of the dataset. In particular, each element P_{ij} represents the Euclidean distance between elements i and j in the dataset. Then, we computed the incidence matrix $I=(I_{ij})$, where each element I_{ij} is 1 or 0 if the elements i and j belong to the same cluster or not. We finally computed the *Pearson's correlation,* as defined in [6, page 77], between the linear representation by rows of matrices P and I. Correlation is always in the range -1 to 1, where a correlation of 1 (-1) means a perfect positive (negative) linear relationship.

Table 5 illustrates the final grade and the graduation time, expressed in years, of a sample of graduated students. By applying the *K-means* algorithm to this dataset, with K=2 and **`FinalGrade`** and **`Time`** as clustering attributes, we obtain the following two clusters, in terms of the student identifiers: C_1={10, 40, 60, 70} and C_2={20, 30, 50}; the centroids of the clusters have coordinates (107,3.5) and (96,5.33), respectively.

Student	FinalGrade	Time
10	110	3
20	95	5
30	100	5
40	103	4
50	98	6
60	106	4
70	109	3

Table 5 – A sample data set about students

Tables 6 and 7 show the proximity matrix and the incidence matrix corresponding to clusters C_1 and C_2 of the data set illustrated in Table 5.

P	10	20	30	40	50	60	70
10	0						
20	20,12	0					
30	10,25	10	0				
40	7,07	13,08	3,32	0			
50	12,41	8,06	2,24	5,48	0		
60	4,12	16,06	6,16	3	8,31	0	
70	1	19,13	9,27	6,08	11,45	3,16	0

Table 6 - The proximity matrix for data of Table 5

I	10	20	30	40	50	60	70
10	1						
20	0	1					
30	0	1	1				
40	1	0	0	1			
50	0	1	1	0	1		
60	1	0	0	1	0	1	
70	1	0	0	1	0	1	1

Table 7 - The incidence matrix for clustering of data in Table 5

The Pearson's correlation between the linear representation of these two matrices is -0.59, a medium value of correlation.

As already observed, the real dataset we analysed concerns courses and exams during the academic years from 2001/2002 to 2007/2008 at the Computer Science program of the University of Florence. In particular the data set, organized as illustrated in Table 4, refers to the evaluation of 40 courses for a total of 154 records corresponding to different years. We explicitly observe that we did not consider in our analysis those courses evaluated by a small number of students. For clustering, we used the *K-means* implementation of **`Weka`** [7], an open source software for data mining analysis. The aim was to find if there is a relation between the valuation of a course and the results obtained by students in the corresponding exam. We performed several tests with different values of the parameters and we selected different groups of attributes. For each choice of attributes, we applied the *K-means* algorithm to identify the clusters; then, we computed the Pearson's correlation by using the proximity and incidence matrices. The tests we performed pointed out that the exams having good results, in terms of average grade and delay, correspond to courses having also a good evaluation from students.

In particular, we first used **`AvgGrade`**, **`AvgDelay`**, **`Par1`**, **`Par2`**, **`Par3`**, **`Par4`** and **`Par5`** as clustering attributes and *K*=2, obtaining the clusters illustrated in Figures 1, 2, 3 and 4; each figure represents the projection of the clusters along two dimensions corresponding to the following pairs of attributes **`AvgDelay`** and **`Par3`**, **`AvgGrade`** and **`Par3`**, **`AvgDelay`** and **`Par4`** and, finally, **`AvgGrade`** and **`Par4`**. The centroids of the resulting clusters are shown in Table 8, which also contains the average values relative to the full data set.

Attribute	Full Data	Cluster0	Cluster1
AvgGrade	25,31	25,85	24,58
AvgDelay	2,61	1,8	3,68
Par1	70,86	77,74	61,67
Par2	72,23	82,19	58,94
Par3	84,51	90,25	76,86
Par4	72,03	74,67	68,5
Par5	76,02	80,83	69,61

Table 8 - The centroids of clusters corresponding to Figures 1, 2, 3, 4

The cluster number 0, which correspond to 88 blue stars in the figures, contains the courses which students took with “small” delay and that they evaluated positively. On the other hand, cluster number 1, corresponding to 66 red stars, contains those courses which students took with a large delay and that they evaluated less positively. We observe that the centroids of the two clusters are very close relative to the attribute **`Par4`** which concerns classrooms and equipment. This is also evidenced from Figures 3 and 4, where the blue and red stars are less separated than those in Figures 1 and 2. The Pearson’s correlation corresponding to these clusters is equal to -0.35. We obtained an improvement by excluding the attribute **Par4** from clustering, in fact in this case we find a correlation equal to -0.51.

In general, our tests evidenced that the paragraphs evaluations which are more correlated with students results regard attributes **Par2** and **Par3**, that is, those concerning the course organization and the teacher. Moreover, the clustering results do not change much if we also introduce the attributes **AvgPar1,.., AvgPar5**.

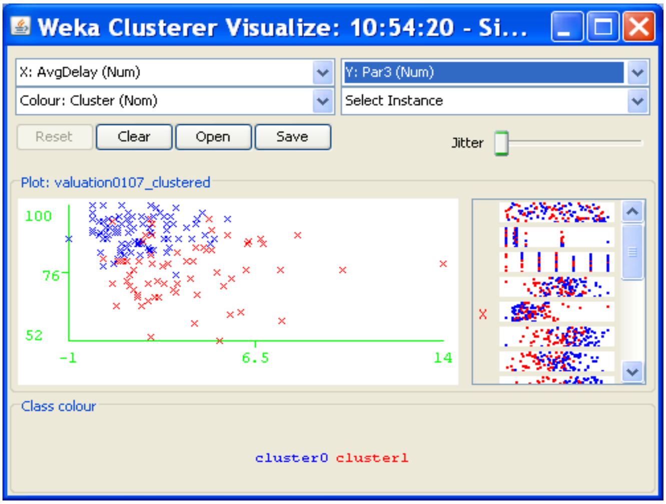

Figure 1 - Clusters of Table 8 with AvgDelay and Par3 in evidence.

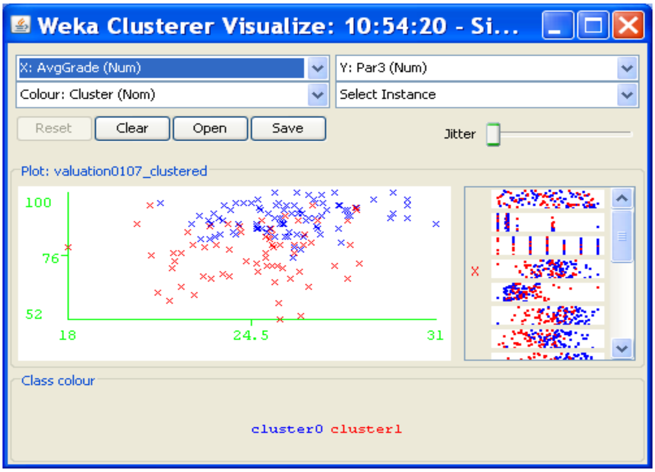

Figure 2 - Clusters of Table 8 with AvgGrade and Par3 in evidence.

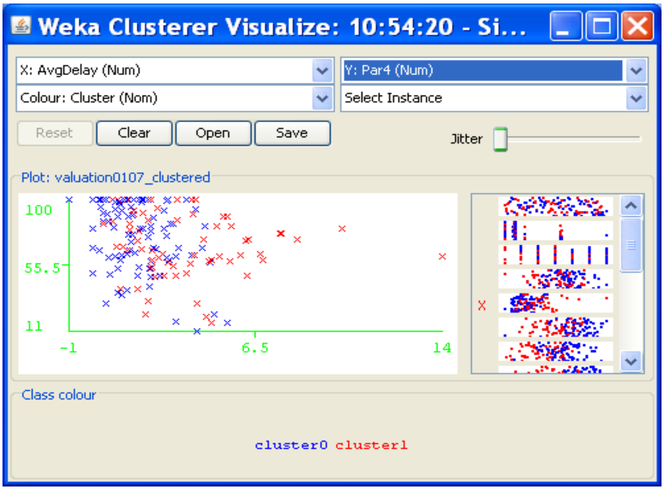

Figure 3 - Clusters of Table 8 with AvgDelay and Par4 in evidence.

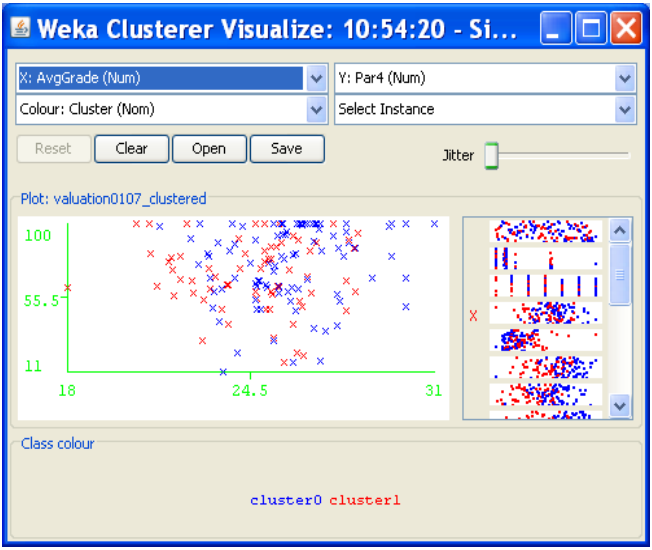

Figure 4 – Clusters of Table 8 with AvgGrade and Par4 in evidence.

3. Conclusions and future works

The results of the previous sections show, in a formal way with data mining techniques, that there is a relationship between the evaluation of the courses from students and the results they obtained in the corresponding examinations. In particular, the analysis performed on data related to the Computer Science degree program at the University of Florence illustrates that the courses which received a positive evaluation correspond to exams in which students obtained a good average mark and that they took with a small delay. Conversely, the worst evaluations were given to those courses which do not match good achievements by students. A result of this type points out a critical issue in the involved courses and can be used to implement improvement strategies. It would be also important to understand how student assessment is influenced by students evaluations of the previous years. The approach used in this work could be refined if it was possible to identify the students involved in the courses evaluation in order to connect properly the results of the evaluation with those of exams. Of course, it should be guaranteed the privacy of results by showing only aggregate information.

Bibliografia

[1] Campagni R., Merlini D., Sprugnoli R.: Analyzing paths in a student database, The 5th International Conference on Educational Data Mining, Chania, Greece, 2012.

[2] Campagni R., Merlini D., Sprugnoli R.: Data mining for a student database, ICTCS 2012, 13th Italian Conference on Theoretical Computer Science, Varese, Italy, 2012.

[3] Campagni R., Merlini D., Sprugnoli R.: Sequential patterns analysis in a student database, ECML-PKDD Workshop: Mining and exploiting interpretable local patterns (I-Pat 2012), Bristol, 2012.

[4] MacQueen J.: Some methods for classifications and analysis of multivariate observations. In Proc. of the 5th Berkeley Symp. on Mathematical Statistics and Probability, 281–297. University of California Press, 1967.

[5] Romero C., Ventura S.: Educational Data Mining: A Review of the State of the Art. IEEE Transactions on Systems, Man, and Cybernetics, 40(6): 601-618, 2010.

[6] Tan P. N., Steinbach M., Kumar V.: Introduction to Data Mining, Addison-Wesley, 2006.

[7] Witten I. H., Frank E., Hall M. A.: Data Mining: Practical Machine LearningTools and Techniques, Third Edition. Morgan Kaufmann, 2011.

[8] https://valmon.ds.unifi.it/sisvaldidat/unifi/index.php