ISSN 2395-8618

Warnings and Recommendation System for an E-Learning Platform

Camilo Peñuela, Elizabeth León, and Jonatan Gómez

Abstract—**A warning messages and recommendation system for an E-Learning system is proposed, the goal is to identify which students are likely to have a poor academic performance, and give them timely feedback by showing alerts and recommended material. The proposed system uses a set of profiles previously identified by a student profiling model, using socio-economic (age and gender) and web navigation data on the system (number of accesses to resources, percentage of accesses in class, average absence time and average session length). Each profile is analyzed and a warning message is assigned to each one; also, the sequences of consultations performed by students with a high academic performance are recognized and used to choose which resources are recommended. Based on the sequence performed by a student in a current session, the platform may recommend access specific resources.**

Index Terms—**Learning management system, web log, student profile, educational data mining.**

I. Introduction

THE traditional educational system is built upon a long-established set of customs: the professor transmits the knowledge by giving a talk to his/her passive students, in a well-established place (class room), linear built (topics are sorted from start to end), standardized tests, and usually with a high student/teacher ratio. This system supposedly guarantees the knowledge transmission, but the standardized tests determine in a general way if the students are acquiring the knowledge, regardless if memorizing or appropriating it.

Thanks to computing and communication technologies, educational models have transformed from monolithic linear models to flexible non-linear models centered in the skills and abilities of each student, they allow the students to learn in their own way and at their own rhythm, guided by an online tool independent of location or student's time [1], [2]. Several educational processes, including both classroom attendance and distance education (radio, TV, internet) have been included [3].

E-learning is a distance educational model which integrates information technologies with pedagogical elements in order to teach online [1], [2], [4], [5]. An e-learning system is comprised of several subsystems, it may include [6]: Learning Management System, Content Management System, Intelligent Tutoring System, Computer Adaptive Test, and Topic interaction and simulation tools. In particular if contents, tests, and tutors are able to adapt according to the skills of each student, the e-learning system is called adaptive e-learning system [7].

These E-learning systems are producing a huge amount of data that can be used for evaluating and improving the learning strategies [8], and there are several research works trying to link data related to a student with his/her academic performance, e.g., trying to discover some student profiles [9]. Profiles are formal structures of pieces of information related to users, usually defined to represent categories of shared common features [10]. By identifying student profiles, an e-learning system can, among others: present personalized content [10], customize the activities to encourage participation [11], predict and prevent academic failure [12], and gather information that allows to improve the course for reducing the dropout ratio and increase the students' performance [1]. Several factors that affect students' academic performance have been found using data mining techniques: familiar, personal, economic, and geographic [13], [14], [15], [16]. However, results of these studies depend on the kind of data used to identify the profiles, e.g., socio-economic, number of accesses to the course material [17], time spent by students in consulting the material [18], and web server logs [19]. Web logs can be used to identify users navigation profiles.

Web mining is the application of data mining to data collected from the web [20]. It has three main categories: content mining (which analyses the web content), structure mining (which analyses the links between pages), and usage mining (which extracts useful information from server logs) [20]. Regarding usage mining, several factors must be taken into account: not all pages across the web are of equal importance to students [21], many pages and documents are consulted just for location rather than for importance [21], and not every access to a page is registered (due to cache loading), such lost accesses can be recovered through the identification of possible paths based on the website structure tree [22].

In this paper, a warnings and recommendation system to give timely feedback to students who are likely to have a poor academic performance, is proposed, this system uses the profiles found by a student profiling model, which is further described in [23], this model uses the data registered by a LMS called "Virtual Intelligent Learning Platform": socio-economic (gender and birth date) and behavioral (accesses to the course

Manuscript received on June 17, 2015, accepted for publication on August 19, 2015, published on October 15, 2015.

Camilo Peñuela and Elizabeth León are with the MIDAS research group at the Universidad Nacional, Bogota, Colombia (web: http://www.midas.unal.edu.co).

Jonatan Gómez is with the ALIFE research group at the Universidad Nacional, Bogota, Colombia (web: http://www.alife.unal.edu.co).

ISSN 2395-8618

TABLE I
AMOUNT OF DATA BY TERM/TEST

Test	Students	Sessions	Accesses
First. 2014-1	544	4675	25205
Second. 2014-1	469	3313	10734
Third. 2014-1	360	2272	5549
First. 2014-2	469	5022	24721
Second. 2014-2	412	3593	12190
Third. 2014-2	312	1698	3795

material). Each cluster is a potential profile, it is analyzed to find a connection between its properties and the academic performance, if the performance was low, a warning message based on the profile properties is assigned to the profile. Once the profiles, each assignment of student to a profile and the warning messages were stored on the knowledge base, the warnings system assigns one profile to each active student (who is currently taking the course), if the student is assigned to a profile with an associated warning message, that message is shown in the Platform.

This paper is organized as follows: In Section 2, a brief introduction to the student profiling model is presented, including the preprocessing, and clustering. In Section 3, the warnings and recommendation system proposed in this paper is presented. In Section 4, conclusions and future work are presented.

II. Student Profiling and Common Sequences Identification Model

Student profiles and common sequences are determined according to web navigation through the Virtual Intelligent Learning Platform [6], age and gender. The model is shown in Figure 1 [23].

The profiling model has several input data, coming from the platform database [23]:

- Web usage log: Any access performed by each student to each resource of the course material.
- Students' birth date and gender.
- Test attempts: Any attempt performed by each student to each test, including the grade.

The web usage log contains data of around 1.000 students who have taken the "Computer Programming" course in 2014 (around 20 groups per term) and have taken three tests (three tests are applied each term). Each test of a term (e.g. First test) has covered different topics, and the amount of data differs among tests (Table I), because of this, three independent profiling processes were performed, obtaining a set of profiles per test.

The data are pre-processed by removing students who have never taken a test, and multiple accounts of the same student are mixed into only one [23].

TABLE II
FEATURES USED FOR THE CLUSTERING PROCESS

Feature	Description
ag	Student's age
gd	Student's gender
nD	Number of accesses to documents
nE	Number of accesses to exercises
nV	Number of accesses to videos
dC	Percentage of accesses to documents in class time
eC	Percentage of accesses to exercises in class time
vC	Percentage of accesses to videos in class time
sl	Average time spent by session
ta	Average absence time between sessions

A. Data Aggregation

Its goal is to count the number of accesses, carried out by each student for each test taken. The input data of *Web usage log* were aggregated one row by student. By each student, all of his/her accesses are sorted and split according to test attempts dates, any access carried out between the first day of the academic term and the day when the first test was taken, was assigned to the first test, any access carried out between the day when the first test was taken and the day when the second test was taken, was assigned to the second test, and so forth. The values presented in Table II are calculated for each set of accesses. *Absence time* is the time between two sessions.

If a value could not be calculated because of lack of data, which implies a division by zero (e.g. calculate the average absence time with only one session), the null value was changed by the average of another students' values, of the same term.

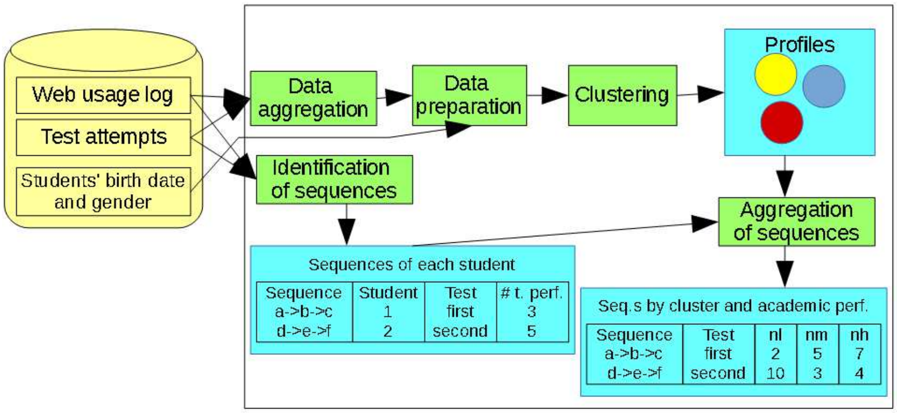
B. Data Preparation

This process transforms each set of tables returned by *Data aggregation* process, of a given test (e.g. all tables of first test), into one unique minable table, which is then used by *Clustering* process to find student profiles.

For each table returned by *Data aggregation* process, all rows are normalized using range transformation to [0.0 1.0], then a density based algorithm [24] is used to identify a discard outliers, any point which is farther away than 0.7 compared with at least 70% of all other points, is marked as outlier and discarded. Then, the grades of test attempts of the remaining points (i.e. all points that were not considered as outliers) are curved by subtracting the mean from the grade of all terms, and finally, the fields are normalized again to [0.0 1.0], to avoid a concentration of points to few values, done by a outlier which was too far away. The students' gender may have the value of 0.0 (Female) or 0.3 (Male).

Once all tables have been processed, each set of tables of a given test (e.g. First test) are mixed into only one, by concatenating all tables of that set. As a result, a minable table for each test is obtained.

ISSN 2395-8618


Web usage log
Test attempts
Students' birth date and gender
Data aggregation
Data preparation
Clustering
Profiles
Identification of sequences
Aggregation of sequences
Sequences of each student
Seq.s by cluster and academic perf.

Sequence	Student	Test	# t. perf.
a->b->c	1	first	3
d->e->f	2	second	5

Sequence	Test	nl	nm	nh
a->b->c	first	2	5	7
d->e->f	second	10	3	4

Fig. 1. Profiling and common sequences identification model

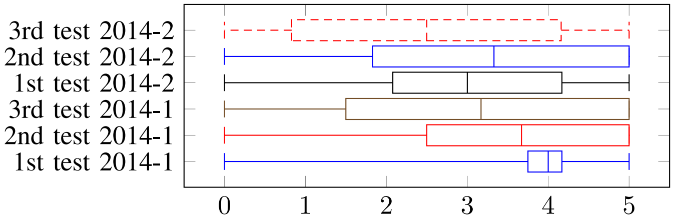


Fig. 2. Box plot of grades gotten in attempts of each test

Once the set of clusters for each test was obtained, all the fields are discretized by frequency in order to improve the cluster descriptions, all the legend shown in Figures 4 to 6 were discretized in this way.

C. Clustering

For each table returned by *Data preparation* process, the K-Means algorithm (with Euclidean distance) is applied, using all fields presented in Table II, each cluster is a potential profile. In order to determine the k value, experiments varying it between 2 and 20, are executed 1000 times, each process has performed up to 10000 optimization steps.

The values of sl and ta were used in milliseconds, and ag in days, by the clustering algorithm, however, ag is shown in years, sl in minutes, and ta in hours, in the cluster descriptions.

D. Clustering Model Results

The profiling model was applied to the "Computer Programming" course in 2014, around 1000 students have taken the course, and performed around 20500 sessions. The grades use the numeric scale from 0.0 to 5.0, any grade equal to or greater than 3.0 is considered as a passing grade. The clustering model results are also presented in [23].

The Figure 2 shows the box plot of attempt grades per test. It was noticed a high concentration of grades in first test of term 2014-1, namely most students have gotten high grades regardless of their behavior, because of this, they were discarded.

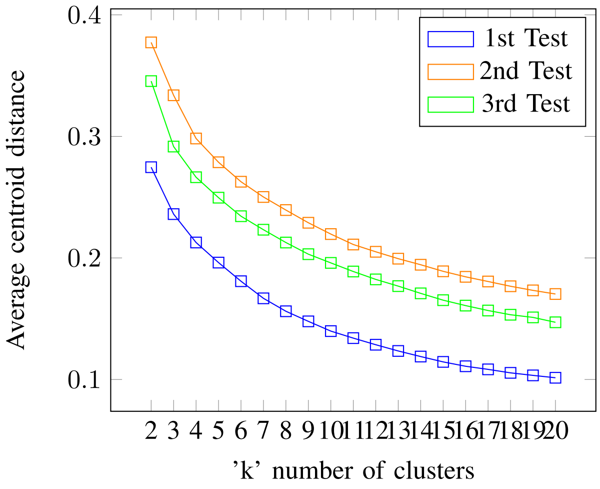


Fig. 3. Average centroid distance vs number of clusters

The Figure 3 shows the convergence of the k value, it indicates that the marginal gain tends to drop at 6, this value was assigned to k.

For the first test, 70 students were assigned to profile 0, 23 to profile 1, 73 to profile 2, 151 to profile 3, 74 to profile 4 and 140 to profile 0. For the second test, 246 students were assigned to profile 0, 33 to profile 1, 204 to profile 2, 257 to profile 3, 123 to profile 4 and 112 to profile 5. For the third test, 89 students were assigned to profile 0, 64 to profile 1, 137 to profile 2, 277 to profile 3, 74 to profile 4 and 187 to profile 5.

The clusters found are described in Figures 4 (First test), 5 (Second test), and 6 (Third test). The values of second and third tests are normalized because the profiles were identified with data coming from both terms.

Based on these descriptions of first test, a connection between the consultation volume of the course material and the academic performance was discovered, this is most noticeable in clusters 2 (many accesses and high academic performance) and 4 (few accesses and low academic performance). Another

ISSN 2395-8618

clusters had a moderated number of accesses and a balanced academic performance. Regarding second and third tests, the academic performance behavior is more balanced among clusters. In second test, the academic performance is slightly better in cluster 3, it is characterized by a moderated number of accesses to all resource types, with preference for documents, which were accessed mostly out of class, namely access documents out of class may improve the academic performance. In third test, the academic performance is slightly better in cluster 2, it is characterized by a lot of accesses to documents, a moderated number of accesses to exercises and only few accesses to videos. Many accesses to documents were performed out of class. Regarding accesses to exercises and videos, the variables describe two behaviors, located at each extreme: a group contains students who have performed most of their accesses (at least 90%) in class, while students of the other group have performed most of their accesses (at least 90%) out of class.

It can be concluded that the platform is an useful tool for developing the course, because every test has a connection between the consultation volume and the academic performance, although the academic performance is more balanced in second and third tests and that connection is not always a direct connection, e.g. in second test the profile 1 has the highest number of accesses and the lowest academic performance.

E. Identification of Sequences

A numeric code was assigned to each unique resource (document, exercise or video). The resources accessed by one student in one session are sorted according to time, obtaining an array, e.g. “1 3 1 1 2” indicates that the student logged in, the he accessed the resource with code “1”, then “3”, and so forth, then he stopped consulting for at least 30 minutes. The *Identification of sequences* process identifies those sequences and counts the number of times that a given sequence was performed by the same person in order to study for a given test (e.g. The first test), that count is assigned to “# t. perf.” (Number of times performed).

F. Aggregation of Sequences

Three grade groups are identified, all the grades gotten by all students of a given test (e.g. the first test) are discretized by frequency in three binds: *nh* (High), *nm* (Medium) and *nl* (Low) academic performance. The number of times that a sequence was performed by students who have gotten a low, medium or high academic performance were counted, and these values are assigned to nl, nm and nh respectively. The sequences with a *nh* value greater than that of *nm* and greater than that of *nl* are identified as of high academic performance, they are selected to choose which resources are going to be recommended.

III. Warnings and Recommendation System Design

The main goal is to give timely feedback to the current student in order to improve his/her academic performance, by suggesting actions based on the profile assigned to him/her. The warnings and recommendation system design is shown in Figure 7.

This system uses the same data that were used by the clustering and common sequences identification model, also the clean data (used to identify the profiles), the profiles and the common sequences (with their *nh*,*nm* and *nl* values), which were stored in the knowledge base.

The warnings system is executed in the background through a scheduled task, which is comprised of several processes shown in Figure 7, this task is responsible for assigning a profile to each active student, based on recent behavioral and socio-economic data. When a student logs in and start consulting, the platform queries the most recent profile assigned, and based on it, the platform may show a warning.

The first thing that must be considered, is that the warnings system works with incomplete data, i.e., the current student still has enough time to carry out more accesses, while the data of the knowledge base were built with complete lapses, because the *Profiles in knowledge base* were filled immediately after the historic students have taken each test. Because of this, the behavioral data of active students must be prepared to allow to compare them with those of historic students.

A. Query Accesses of Previous Sessions

It processes all the active students, one at a time. For each one, it checks if the student has taken all tests with reached deadlines, e.g. define if the student did not take the second test, but now it is too late for taking that, in such case the student has dropped out the course, otherwise the date of the last test attempt indicates the beginning date of the lapse of study for the next scheduled test, all the accesses carried out since that date, were performed in order to study for that future test, they are the data used for assigning a profile to the current student and warn him/her if necessary.

B. Data Aggregation

In a similar way (but using recent activity data of active students) to the *Data aggregation* process of the profiling model, the data returned by the *Query accesses of previous sessions* process is aggregated into one row per student, by calculating the features presented in Table II.

C. Data Preparation

This process prepares the data to allow them be comparable with the data stored in *Profiles in knowledge base*, namely the data must be of equal weight for each field, and the outliers must be discarded. The *Data aggregation* process returns three tables (one per test), for each one, the *Data preparation* process normalizes them using range transformation to [0.0

ISSN 2395-8618

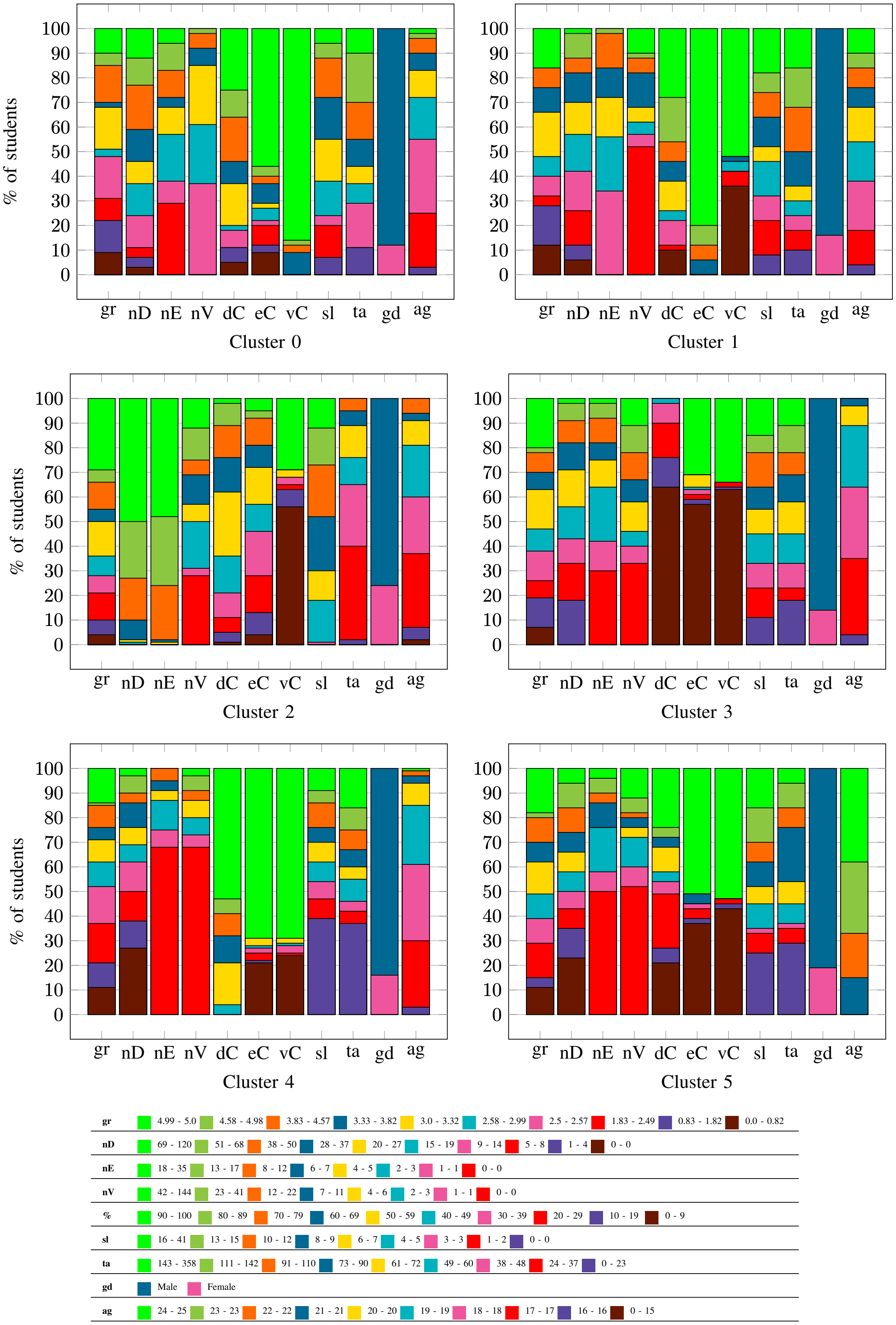


Fig. 4. Variable descriptions of clusters of the first test in 2014-2

ISSN 2395-8618

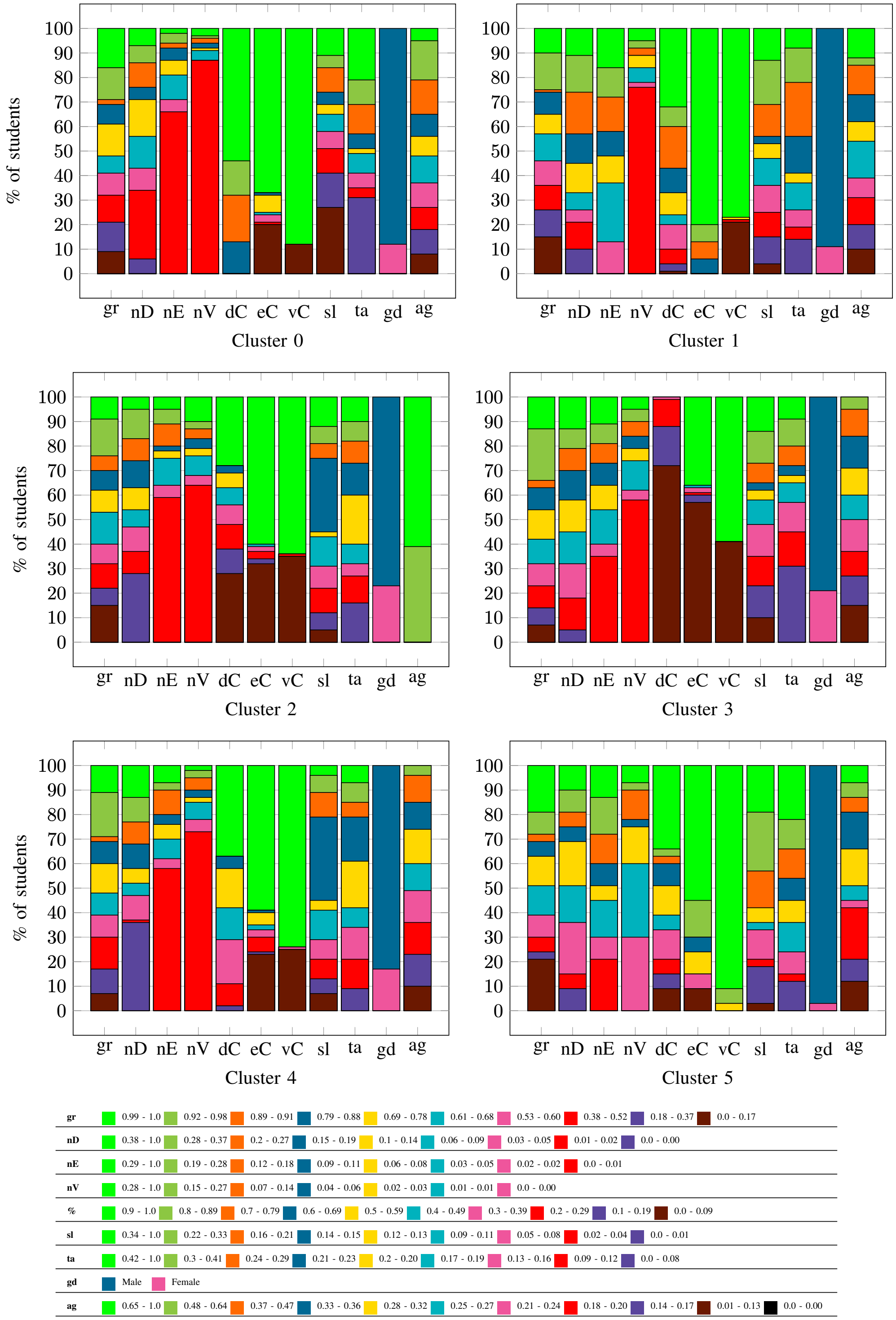


Fig. 5. Variable descriptions of clusters of the second test

ISSN 2395-8618

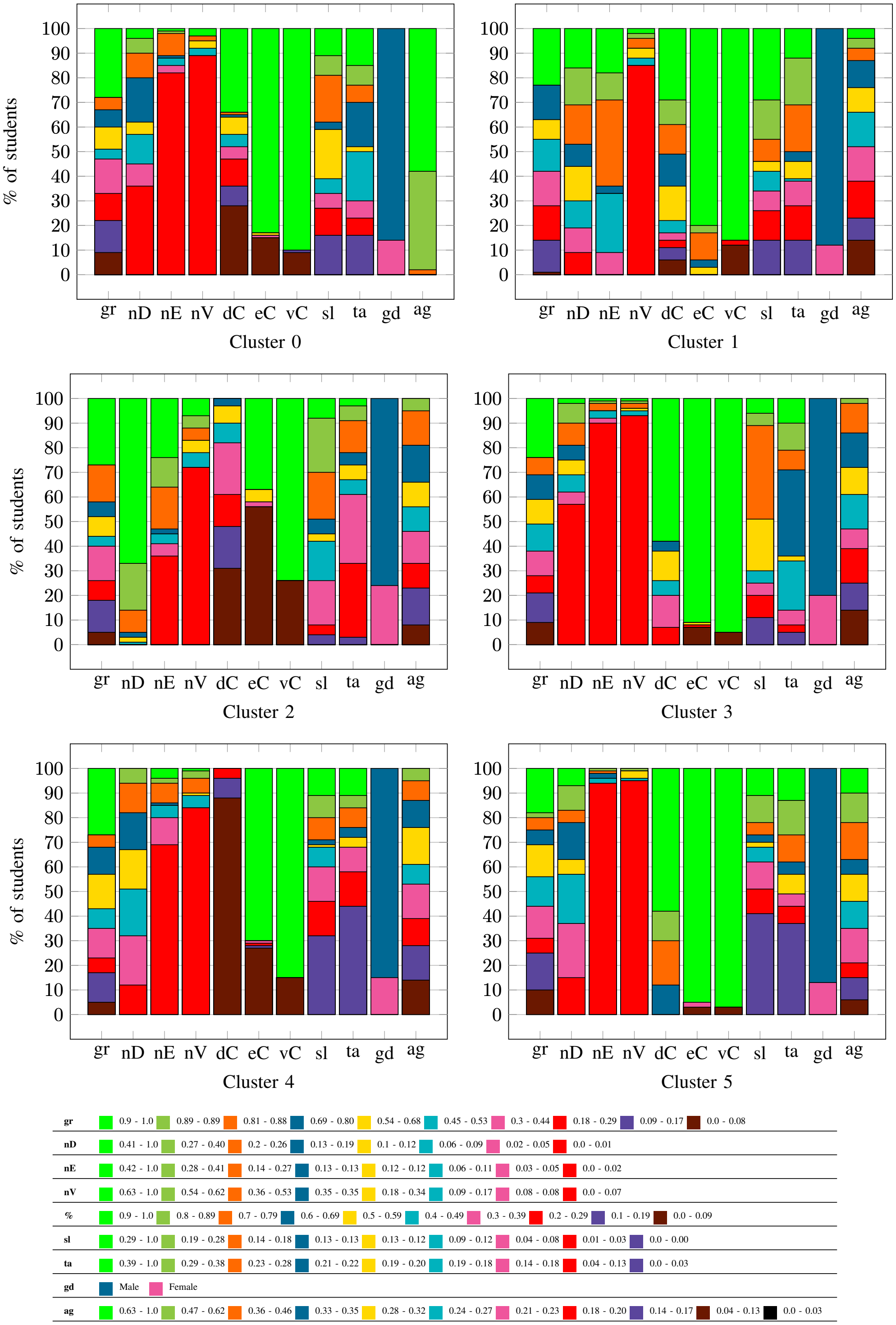


Fig. 6. Variable descriptions of clusters of the third test

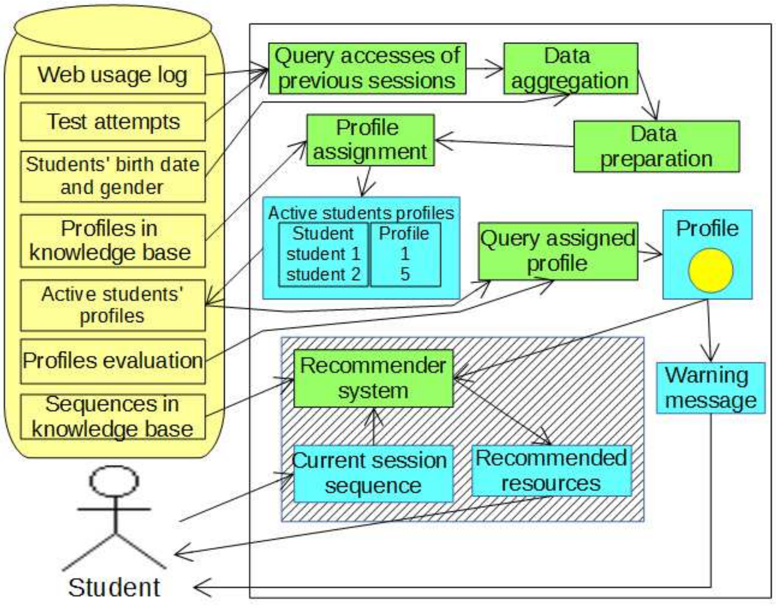


Fig. 7. Warnings and recommendation system

TABLE III
ACADEMIC PERFORMANCE AND WARNING MESSAGE ASSIGNED TO CLUSTER

Warning message	Cluster	Test
You are encouraged to make a stronger use of the course material	3,4	First
You are encouraged to watch more videos	1	Second
You are encouraged to access more exercises and videos	2	Second

1.0], then a density based algorithm was run over the table to identify and discard any row that is farther away than 0.7 to at least 70% of all other rows, after calculating the euclidean distance with the fields: nD, nE, nV, sl and ta. All rows that were not discarded, are normalized using range transformation to [0.0 1.0]. This process is run to avoid concentrations of non discarded rows on few values, due to few outliers that were far away.

D. Profile Assignment

This process, assigns a profile previously identified by the *Clustering* process of the profiling model, to each active student, by processing each minable view returned by *Data preparation* process (it returns one view per test), a row at a time. This process, uses a K-Nearest-Neighbor classifier to assign a profile to the active student, the label is the profile, the labeled data are stored in *Profiles in knowledge base*, and k was set to the 10% of the number of historic students who have taken the test that the current student is going to take. The assignment is stored in *Active students' profiles* overwriting the row registered by the past scheduled task.

E. Giving Feedback to Current Student

Each time the student logs in and watches the topic list (after the selection of the course), the platform queries the last assigned profile by the *Profile assignment* process, and checks if that profile has a warning message, if so, that message is shown in the platform. The process to be run is *Query assigned profile*, it queries an assignment of a cluster to the current student in *Active students' profiles* table, if a profile is found, the process queries a warning message assigned to this profile, in *Profiles evaluation* table, if that message is found, it is shown in the platform. The *Recommender System* process identifies sequences registered on the *Sequences in knowledge base* which are similar to the one performed by the current student in his/her session so far. Those sequences were found for the test that the current student is going to take, and for each sequence, its *nh* value is greater than that of *nm* and than that of *nl*. Those sequences which contain resources not yet accessed by the current student, are selected to choose which resources are going to be recommended.

IV. RESULTS

The warning messages system is aimed to students who are likely to have a low academic performance. Based on the descriptions shown in Figures 4 (First test), 5 (Second test), and 6 (Third test), the profiles characterized by a low academic performance are: 3 and 4 of first test, 60% and 70% of students failed the test, respectively. 1 and 2 of second test, around 60% of students failed the test in both cases. The warning messages assigned to these profiles, are presented in Table III.

A lot of common sequences were identified, consider a student who is going to take the second test, he was assigned to profile 2, he logs in and perform the sequence

TABLE IV
RESOURCE CODES OF CURRENT SESSION SEQUENCE

Code	Resource	Topic
082	Flows exercises	Flows
083	Flows chapter	Flows

TABLE V
RESOURCE CODES OF SEQUENCES REGISTERED ON KNOWLEDGE BASE

Code	Resource	Topic
060	Input and output flows and persistence. Video	Flows
072	Logic exercises	Logic
084	Recursion exercises	Recursive functions
085	Recursion chapter	Recursive functions
087	Loops chapter	Loops
102	Languages presentation	Languages

"082,083", the resources involved in this sequence are shown in Table IV. The platform queries the similar sequences registered on the knowledge base, which were performed in order to study for the second test, and whose nh value was greater than that of nm and than that of nl, for the profile 2. They are "082,083,060,082,083,082,083" and "082,083,083,060,060,085,084,102,072,085,085,087". The resources involved in those sequences are shown in Table V, they are the resources chosen for recommendation.

V. CONCLUSIONS AND FUTURE WORK

A warnings system was proposed, it is aimed to give timely feedback to students who are likely to have a poor academic performance, it uses recent activity data of active students (who are currently taking the course), profiles and normalized data of historic students (who have taken the course in past terms) to assign a profile to the active student, and based on it, show a warning message if necessary.

The profiling model uses students' age and gender and behavioral data (number of accesses to each resource type, percentage of accesses in class, average session length and average absence time), then, it was possible to analyze the clusters in order to notice a connection with the academic performance and assign a warning message to a profile if necessary. The model finds the groups of similar students using the k-means algorithm (to choose the appropriated *k-Value* the average centroid distance was used). Profiles were identified and related to academic performance, it was noticed that some behaviors in navigation through the resources are reflected in the grades of the test attempts.

A recommender system is also proposed, it uses the common sequences found for historic students, and their *nh,nm* and *nl* values, to identify the sequences connected to a high academic performance, and select them to choose which resources are going to be recommended.

As future work, it is necessary to get feedback from students about warning messages and recommendations, to validate the system, however this validation takes at least one term, to obtain a complete set of students' perceptions.

Most common sequences were short, up to three accesses, it was due to the low granularity of topics covered by each document and exercises file, each one covers a whole chapter (e.g. Flows). As future work, the course material could be split into atomic topics, this action will allow identify more specific sequence patterns.

The Virtual Intelligent Learning Platform can be used to support more courses, the profiling and common sequences identification model can be applied to those courses, and their students would receive warning messages and recommendations.

The profiling model could be enriched with more data coming from another systems (e.g. the Academic Information System, which contains more academic data, or the welfare division, which contains more socio economic data), those data will be used by the profiling model as they become available.

REFERENCES

[1] S. Ayesha, T. Mustafa, A. R. Sattar, and M. I. Khan, "Data mining model for higher education system," *European Journal of Scientific Research*, vol. 43, pp. 24–29, 2010.

[2] D. R. Garrison, T. Anderson, and R. Garrison, *E-Learning in the 21st Century: A Framework for Research and Practice*, 1st ed. New York, NY, 10001: Routledge, 2003.

[3] X. Zhao, "Adaptive content delivery based on contextual and situational model," Ph.D. dissertation, The University of Electro-Communications, Tokyo, Japan, 2010.

[4] G. Kearsley, *Online education: Learning and teaching in Cyberspace*. Wadsworth Publishing Company, 2000.

[5] G. Salmon, *E-Tivities: The Key to Active Online Learning*. Routledge, 2002.

[6] J. Goméz, E. León, A. Rodriguez, E. C. Cubides, J. Mahecha, J. C. Rubiano, and W. Prado, "A didactic e-learning platform with open content navigation and adaptive exercises," in *2012 International Conference on Education and e-Learning Innovations (ICEELI)*. IEEE, 2012, pp. 1–6.

[7] V. M. García-barrios, F. Mödritscher, and C. Gütl, "Personalisation versus adaptation? a user-centred model approach and its application," 2005.

[8] A. K. Hamada, M. Z. Rashad, and M. G. Darwesh, "Behavior analysis in a learning environment to identify the suitable learning style," *International Journal of Computer Science and Information Technology*, vol. 3, no. 2, pp. 48–59, apr 2011. [Online]. Available: http://dx.doi.org/10.5121/ijcsit.2011.3204

[9] C. Romero, S. Ventura, and E. García, "Data mining in course management systems: Moodle case study and tutorial," *Computers & Education*, vol. 51, no. 1, pp. 368–384, 2008. [Online]. Available: http://dblp.uni-trier.de/db/journals/ce/ce51.html\#RomeroVG08

[10] C. Mencar, C. Castiello, and A. M. Fanelli, "A profile modelling approach for e-learning systems," in *ICCSA (2)*, ser. Lecture Notes in Computer Science, O. Gervasi, B. Murgante, A. Laganà, D. Taniar, Y. Mun, and M. L. Gavrilova, Eds., vol. 5073. Springer, 2008, pp. 275–290. [Online]. Available: http://dblp.uni-trier.de/db/conf/iccsa/iccsa2008-2.html\#MencarCF08

[11] J. Quevedo, E. M. nés, J. Ranilla, and A. Bahamonde, "Automatic choice of topics for seminars by clustering students according to their profile," *International Journal of Social, Behavioral, Educational, Economic and Management Engineering*, vol. 3, no. 6, pp. 473–477, 2009. [Online]. Available: http://waset.org/Publications?p=30

[12] V. P. Bresfelean, M. Bresfelean, N. Ghisoiu, and C.-A. Comes, "Determining students' academic failure profile founded on data mining methods," *ITI 30th Int. Conf. on Information Technology Interfaces*, 2008.

ISSN 2395-8618

[13] C. López, “Data mining model to predict academic performance at the universidad nacional de colombia,” Master’s thesis, Universidad Nacional de Colombia, 2013.

[14] J. P. Vandamme, N. Meskens, and J.-F. Superby, “Predicting academic performance by data mining methods,” *Education Economics*, vol. 15, no. 4, pp. 405–419, 2007. [Online]. Available: http://EconPapers.repec.org/RePEc:taf:edecon:v:15:y:2007:i:4:p:405-419

[15] E. Y. Fethi A. Inan and M. M. Grant, “Profiling potential dropout students by individual characteristics in an online certificate program,” *Int’l J of Instructional Media*, vol. 36, 2009.

[16] B. A. Chansarkar and A. Michaeloudis, “Student profiles and factors affecting performance,” *Int. J. Math. Educ. Sci. Technol*, vol. 32, pp. 97–104, 2001.

[17] S. Valsamidis, S. Kontogiannis, I. Kazanidis, T. Theodosiou, and A. Karakos, “A clustering methodology of web log data for learning management systems,” *Educational Technology & Society*, vol. 15, pp. 154–167, 2011.

[18] I. K. Nagy and C. Gaspar-Papanek, *User Behaviour Analysis Based on Time Spent on Web Pages*. Springer-Verlag Berlin Heidelberg, 2009.

[19] H. Liu and V. Kešelj, “Combined mining of web server logs and web contents for classifying user navigation patterns and predicting users’ future requests,” *Data Knowl. Eng.*, vol. 61, no. 2, pp. 304–330, May 2007. [Online]. Available: http://dx.doi.org/10.1016/j.datak.2006.06.001

[20] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000.

[21] M. Kantardzic, *Data Mining: Concepts, Models, Methods and Algorithms*. New York, NY, USA: John Wiley & Sons, Inc., 2002.

[22] J. L. Ortega Priego and I. F. Aguillo Caño, “Minería del uso de webs: Web usage data mining,” *El Profesional de la Información*, vol. 18, no. 1, pp. 20–26, 2009.

[23] C. Penuela, “Student profiling model for the “Computer Programing” course,” Master’s thesis, Universidad Nacional de Colombia, 2015.

[24] R. I, *Rapid Miner Operator Reference*, 2014. [Online]. Available: http://docs.rapidminer.com/studio/operators/