Je-LKS
Journal of e-Learning and Knowledge Society
The Italian e-Learning Association Journal
PEER REVIEWED PAPERS
LEARNING ANALYTICS: FOR A DIALOGUE BETWEEN TEACHING PRACTICES AND EDUCATIONAL RESEARCH
Vol. 15, n.3, 2019
ISSN: 1826-6223 | eISSN: 1971-8829

LEARNING ANALYTICS TO SUPPORT LEARNERS AND TEACHERS: THE NAVIGATION AMONG CONTENTS AS A MODEL TO ADOPT

Sergio Miranda, Rosa Vegliante

Dipartimento di Scienze Umane, Filosofiche e della Formazione,
Università degli Studi di Salerno
semiranda@unisa.it, rvegliante@unisa.it

Keywords: Learning Analytics, Navigation paths, LMS, Moodle, Learning Analytics model

Learners have different needs and abilities; teachers have the ambition to intervene before it is too late. How may e-learning systems support this? Learning Analytics may be the answer but there is not a general-purpose model to adopt. Many learning analytics tools examine data related to the activities of learners in on-line systems. Research efforts in learning analytics tried to examine data coming from LMS tracks in order to define predictive model of students' performances and failure risks and to intervene to improve the learning outcomes. The analytical methods are widely used but no theoretical references are clear.
In this paper, we tried to define a prediction model for learning analytics. In particular, we adopted a Moodle-based LMS in a blended course and collected all data of more than 400 undergraduate students in terms of resource accesses and exam performances. The model we defined was able to identify the learners at risk during their learning processes only by analysing their navigation paths among the contents.

for citations:
Miranda S., Vegliante R. (2019), *Learning Analytics to Support Learners and Teachers: the Navigation among Contents as a Model to Adopt,* Journal of e-Learning and Knowledge Society, v.15, n.3, 101-116. ISSN: 1826-6223, e-ISSN:1971-8829
DOI: 10.20368/1971-8829/1135065

1 Introduction

The information and communications technologies are changing the teaching and learning approaches adopted into the higher education. This is happening mainly because Internet offers the possibility to gather more content that is open and it is transforming traditional courses into richer online experiences (Hoic-Bozic *et al.*, 2009). Moreover, Learning Management Systems (LMSs) easily allow teachers giving their students additional resources and activities as animation, slides, exercises, quizzes, collaborative components (Piña, 2012). Since all the actions tracked by the LMSs are information about the behaviour of the learners, they become the mean to improve learning and teaching. The analysis of this kind of data is what everybody thinks about the learning analytics (Siemens & Baker, 2012).

Learning analytics are defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens & Baker, 2012). There are two main approaches for making decisions based on learning analytics techniques. First approach takes into consideration visual analytics in order to provide students with insights on their own progress and teachers with easy to comprehend information about student's competency and decision making on the education context. The second approach of learning analytics refers to collecting and analysing student data to provide a recommending or adaptive system.

Teachers need better insights from the systems about the interaction among students and technologies. To provide them, systems should be more efficient in processing the large amount of data. Since the traditional approaches analyse structured data to provide feedback to the tutors, learning analytics should examine patterns, correlations and try to transform data in a way to support decision-making, or to give benefits to the Intelligent Tutoring Systems themselves.

Many learning analytics tools examine data related to the activities of learners in on-line systems. Some of them analyse the number of user clicks (Siemens, 2013), others investigate the participation into forums (Agudo-Peregrina *et al.*, 2014), some others consider the time spent and the number of email messages sent (Macfadyen & Dawson, 2010). Another approach (Virvou *et al.*, 2015) analyse logs of user actions and, moreover, collects and analyses feedbacks of learners about level of understanding, satisfactory level, emotion and interaction on each learning object in order to correlate actions done by learners to what they perceive about the adopted content. The system itself process all data to support on-line tutors and give them early warnings about progresses of students.

Research efforts in learning analytics tried to examine data coming from LMS tracks in order to define predictive model of students' performances and failure risks and to intervene by providing personalized injections able to change the learning outcomes (Shum & Ferguson, 2012). The analytical methods are widely used but no mention to theoretical argumentation. Moreover, it is very hard to compare studies and draw overall conclusions because the analysis involve usually few institutions, few courses or only special cases (Gasevic *et al.*, 2016). In fact, many studies have examined similar LMS data, have used similar models and predictors but they have found different results (Gaeta *et al.*, 2016).

In this paper, we describe what we have observed and which kind of prediction model we could apply on learning analytics. In particular, we adopted a Moodle-based LMS in a blended course and collected all data of more than 400 undergraduate students in terms of resource accesses and exam performances in order to define a model able to identify students at risk and eventually to create personalized and target actions.

2 Reference scenario

Since the learning analytics examine data, they are, of course, widely data-driven and they do not refer to some specific theories. Their analysis usually refer to raw data coming from the LMS logs and their interpretation has no direct connection to theoretical or methodological models (Marzano & Notti, 2014).

Nevertheless, some recent studies tried to orient the learning analytics approaches versus the interaction theory of Moore (1989), the self-regulated learning (Agudo-Peregrina *et al.*, 2014) or the constructivist theory (Gasevic *et al.*, 2016). Students during their activities may reach different performance although they use the same resources and follow the same suggestions. These theories serve to explain these differences. The measurements adopted in learning analytics do not reflect exactly these theories, thus, other theories such as the situated learning (Brown *et al.*, 1989) or the connectivism (Siemens, 2005) should be considered.

Predicting the students' performance seems to be the principal target of the learning analytics approaches. They should be able to forecast whether a student pass the exam and receive a good final grade. The main models adopt data related to the student features. Recent studies abandoned these features to apply predictive analytical techniques only to data coming from the LMSs. The main problem is that there is a wide variety of systems, variables to consider and techniques, thus, it is hard to point out the best approaches, or in particular, the most effective predictors (Tempelaar *et al.*, 2015).

Rafaeli and Ravid (1997) were the first to use LMS data for learning analytics. They analysed the amount of pages read by the users and compared with their results. They are able to explain only the 22% of the variance of final grades.

Morris, Finnegan, and Wu (2005) found that the number of content pages viewed was a significant predictor, but they examined also posts and time spent on viewing discussions. They reach the 31% in the estimation.

Macfadyen and Dawson (2010) correlated the number of links and files viewed with the final grade. Their researches reached a 33% in the prediction. Moreover, they provided better predictions for "at risk students" by analysing posts, messages and assessments. Their predictions were around the 74% of this kind of students.

In following studies (Nandi *et al.*, 2011), the analysis of the participation of students to forum discussions gave only a 40% of accuracy in the predictions of final grades.

Yu and Jo (2014) examined logs, times, regularities of study intervals, downloads, interactions with peers and with the instructors. They found that only the total time online and the interaction with peers has significant correlation with final grades, but the accuracy of their predictions, however, is around the 34% of the variance.

Zacharis (2015) analysed 29 variables. He found that only 14 of which correlated significantly with final grades. In particular, he found that total time online and the amount of files and links viewed has a significant correlation with the final grades.

On the contrary, Macfadyen and Dawson (2010) did not considered these data in their final prediction model of students' performance, but only the number of viewed files, the interactions and the contributions to content. They reached the accuracy of 52% of the variance of the final grades of their students.

All these researches reached not so much significant percentages of estimation and considered few numbers of learners (around some hundreds). Some wider studies conducted on different platforms considering a more significant number of students (around some thousands) (Beer *et al.*, 2010), showed as their main result, that the higher correlation with the final grades is on the number of clicks.

Recent studies underlined that is still unclear how to use data coming from LMSs for predictive modelling and, in case there is a model, its portability is a crucial step because its effectiveness depends on the learning design approach used to create the course (Miranda *et al.*, 2017). By collecting all the variables considered by the most effective learning analytics approaches, the researchers recommended to consider the following parameters: total number of clicks, number of online sessions, total time online (min), number of course page

LMS

metrics

views, irregularity of study time, irregularity of study interval, largest period of inactivity (min), time until first activity (min), average time per session (min), number of resources viewed, numbers of links viewed, number of content page views, number of discussion posts views, total number of discussion posts, number of quizzes started, number of attempts per quiz, number of quizzes passed, number of quiz views, number of assignments submitted, number of assignment views, number of wiki edits, number of wiki views, average assessment grade.

low adoption

As alternative and holistic methodologies are starting to get interest the Multimodal Learning Analytics (MMLA) systems. They include multiple data sources and the data organization and the data processing they need is very complex. For this reason, the creation of MMLA software architectures is quite complicated and their adoption is not so wide.

We hope the issues raised in this paper are useful for the growing community of MMLA researchers. We believe that the short- and medium-term MMLA agenda should encourage researchers to pay special attention to the design of flexible infrastructures that support the whole data value chain, enabling the scalable adoption of MMLA, in real scenarios, and in a sustainable way (Shankar *et al.*, 2018).

objective LMS

Nowadays, the LMSs are widely adopted by institutions and they are generating large amounts of data that are indicative of the interactions of learners with the systems. The goal of the Learning Analytics is being proactive in order to mitigate the risks, to improve the engagement and to increase the performances of learners. In fact, Learning Analytics allow institutions improving the quality of their e-learning courses, fine tuning learning strategies and ensuring better interventions (Mothukuri *et al.*, 2017).

problem with Mooc

In particular, researches in Learning Analytics are going on to face the well-known problems in MOOC (Massive Open Online Course) environments, such as reducing the high dropout ratios, predicting the student performance or gauging the effectiveness of educational resources and activities on learners (Munoz-Merino *et al.*, 2015).

Elaborating all this data is quite difficult and expensive. Often, LMSs do not allow accessing all of them in order to apply deeper learning analytics (Sergis & Sampson, 2017). Therefore, the goal of this paper is to find a model simpler and effective as well that could be applied in e-learning platforms and support MOOC environments.

3 Methodological approach

In line with the reference scenario, we are trying to define a model that should have a good predictive accuracy and that allows teachers monitoring

they used an algorith that determines success based im following the right order of activities and as data the completion (time/date) of which activity

their students and intervene on their learning processes before it is too late.

We started this research in the cohort 2017-2018 by involving 140 students in the undergraduate courses of Computer science basics for the bachelor degree in Educational Sciences, at the University of Salerno (Miranda *et al.*, 2019). The experiments went on in the following cohort 2018-2019 by involving 267 other students.

We collected data from the Moodle LMS of students that took the final exam on the first round in June immediately after the in-presence activities. Thus, we are going to analyse the data tracked for the total amount of 407 student.

The course has 37 resources: 19 lessons, 11 exercises and 7 formative assessments.

1. Introduction
2. Lesson 1.1: Introduction on computers
3. Lesson 1.2: Basic notions
4. Lesson 1.3: Representation of Information
5. Assessment n.1
6. Lesson 2.1: Sound coding
7. Lesson 2.2: Character encoding standard
8. Lesson 2.3: The coding of numbers
9. Assessment n.2
10. Lesson 3.1: Computer architecture
11. Lesson 3.2: Programming concepts
12. Lesson 3.3: CPU operation
13. Assessment n.3
14. Lesson 4.1: Algorithms
15. Lesson 4.2: Basic programming concepts
16. Assessment n.4
17. Lesson 5.1: Sorting algorithms
18. Lesson 5.2: Animations sorting algorithms
19. Lesson 5.3: Operating system
20. Assessment n.5
21. Lesson 6: Computer networks
22. Assessment n.6
23. Lesson 7.1: Scratch Off line Environment
24. Lesson 7.2: Scratch On Line Environment
25. Lesson 7.3: Scratch script foundations
26. Scratch Exercise 1
27. Scratch Exercise 2
28. Scratch Exercise 3
29. Scratch Exercise 4

30. Scratch Exercise 5
31. Scratch Exercise 6
32. Scratch Exercise 7
33. Scratch Exercise 8
34. Scratch Exercise 9
35. Scratch Exercise 10
36. Scratch Exercise 11
37. Assessment n.7.

The final exam was an online quiz on the same Moodle platform. The students that gained more than 5.75 points out of 10, passed the exam, the others that did not reach this threshold, failed it.

Learning analytics directly available on Moodle may identify students at risk of abandon or may raise warnings for the absence of the teachers and tutors, but they are not able to provide any kind of forecasting about the final learning outcomes.

Since we gave to the students the maximum flexibility in the use of resources, we started analysing the way the learners navigate among them. In particular, we observed that, among the variables considered in the reference scenario, there is no mention of the navigation path. This means that the order the learners navigate among the content has not been considered as a possible predictor of the learning outcomes. In fact, from the analyses conducted in our first experimentation (Miranda *et al.*, 2019), we understood there is no significant correlation between the exam grade and the number of contents seen. Moreover, there is no significant correlation between the exam grade and the visualizations of specific contents or even the time spent online.

Since we observed that chaotic navigation rarely leads to good results, the navigation path itself could be the only indicative element that could represent a variable to consider in the prediction. Consequently, we would show that the students that follow the most orderly paths achieve the better learning outcomes. First, we should define what we mean by an "orderly path". We identified the resources in the course as the previous numbered list from n.1 to n.37. Thus, the orderly path is the following sequence:

"1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37".

The resources and the assessments are in this order because they respect the requirements of the topics they treat. In other words, for instance, we suggest studying the content n.3 before studying the content n.4 because it will be easier to understand the concepts in the content n.4 if you have some knowledge about the content n.3.

Second, we should define some criteria able to measure the distance between the "orderly path" and the student path.

This problem seems a problem occurring in computational biology and in coding theory: comparing and finding common features in sequences or, more in general, measuring distances between two strings.

This theory named "String metric" has a wide variety of solutions including Hamming distance and similar measures. A good criterion for measuring the distances between the two paths is those of Levenshtein (1966). In information theory and language theory, the Levenshtein distance, or edit distance, is a measure for the difference between two strings. It serves to determine how two strings are similar. It is applied, for example, to simple spelling check algorithms and to search for similarities between images, sounds, texts, etc. The Levenshtein distance between two strings X and Y is the minimum number of elementary modifications that allow transforming X into Y. By elementary modification, we mean the deletion of a character, the substitution of one character with another or the insertion of a character. In our case, the characters are the number corresponding to the learning resources and the Levenshtein distance between the "orderly path" and the student path is the similarity between them. This means that the lower the Levenshtein distance, the more the path of the student is close to the right order.

4 Learning analytics results

(success in the 70% for full, 74% in the middle of the time of the semester)

The data we got from the Moodle LMS are those related to the completion of the students. The "completion report" includes all dates and times in which the students completed each learning activity: when they read a document, watched a video or submitted a test. This allows us understanding which content has been shown before other ones and which activity has been completed before other ones (Fig.1).

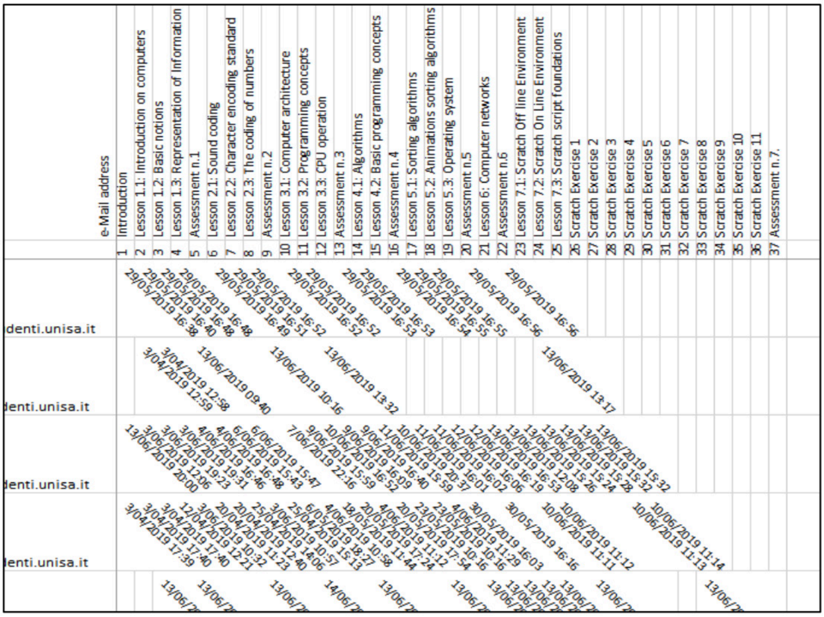

Fig. 1 - The "Completion report" shows all dates and times in which students completed learning activities.

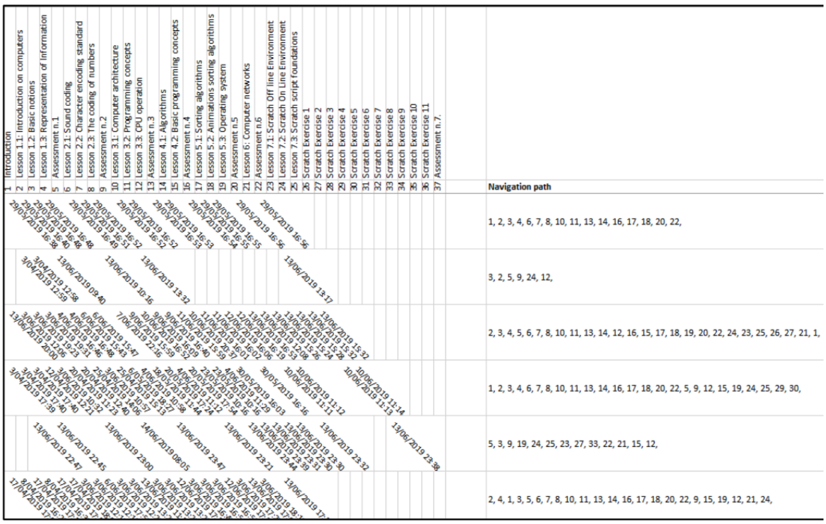

Fig.2 - The navigation paths shows the order the learners followed during their learning experience in the LMS.

We supposed this distance could be indicative of the results. Therefore, we tried to define a sort of prediction of learning outcomes by analysing navigation paths and comparing them with the reference "orderly path".

The easiest way to do it is to define a threshold on the distance and try to do some estimations.

The estimation algorithm we used is very simple and it is in the following:

- IF the Levenshtein Distance is under the Threshold, THEN the Student will pass the final Exam
- ELSE, the Student will not pass the final Exam.

Empirically, we tried all the possible values for the threshold in order to maximize this estimation. The procedure we adopted is in the following:

1. Measure the Levenshtein Distances among the strings relative to the paths of the learners and the string of the "orderly path"
2. Find the maximum M among all the measured distances
3. Try all the possible values of the threshold T between 0 and M
4. For each value of T, estimate by using the pointed out estimation algorithm, which student passes the final Exam and evaluate the total success percentage.

The best results we got are for the threshold equal to 65. In fact, we are able to predict positive results of 313 students out of 407. This means that the percentage of estimation is close to 77% (Table 1, First estimations).

This allowed us forecasting whether a student will pass or will not pass the final exam only by observing his/her navigation path (Fig.3).

This seems to be a good results but it has a poor applicability. In particular, if we think in terms of learning analytics, understanding which will be the learning outcomes at the end of a process could not have a high relevance. Thus, we tried to do the same evaluation by analysing just a half of the learning path of each student. It means that we would try if we were able to predict whether the learning outcomes will be good when the learners are in the middle of their learning processes. This should be relevant because there is time to intervene and eventually fill the gap to move the learners on the right way to reach their goals.

To do it, we considered only a half of the orderly path as a reference and we compared it with the half navigation path of each learner, by measuring, the Levenshtein distance.

Once again, empirically, we tried all the possible values for the threshold in order to maximize this estimation. The best results we got are for the threshold equal to 80. In fact, we are able to predict positive results of 301 students out of 407. This means that the percentage of estimation is close to 74% (Table 1, second estimations). Data and predictions are in Fig.4.

Orderly path			
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37,			
Navigation path	**Lev.D.**	**Exam passed**	**Prediction**
2, 3, 1,	130	0	0
1, 2, 3, 4, 6, 7, 8, 10, 11, 13, 14, 16, 17, 18, 5, 9, 20, 22, 24, 25, 23, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 12, 15, 19, 21,	36	1	1
2, 1, 4, 3, 6, 7, 8, 5, 9, 10, 11, 12, 15, 20, 21, 33, 13, 19, 36, 14, 16, 17, 18, 22, 23, 24, 25, 26, 27, 29, 30, 31, 32, 34, 35, 28,	42	1	1
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 25, 24, 22, 19, 20, 21, 23, 27, 29, 33, 26, 28, 30, 31, 32, 34, 35, 36,	23	1	1
25, 22, 2, 3, 10, 23, 11, 1,	114	0	0
1, 2, 3, 4, 6, 7, 8, 10, 11, 13, 14, 16, 17, 18, 20, 22, 23, 24, 26, 27, 30, 31, 34, 25, 5, 28, 29, 32, 36, 21, 12, 15,	44	0	1
3, 2,	133	0	0
2, 1, 3, 6, 7, 8, 9,	118	0	0
1, 2, 3, 4, 6, 7, 8, 10, 11, 13, 14, 16, 17, 18, 20, 22,	82	0	0
13, 14, 11, 2, 3, 4, 5, 24, 22, 25, 26, 36, 33,	94	0	0
2, 4, 3, 6, 7,	124	0	0
2, 4, 16, 17, 3, 5, 9, 15, 21, 24, 36, 12, 19,	99	1	0
2, 4, 6, 7, 8, 10, 11, 3, 13, 14,	105	0	0
2, 3, 4, 6, 7, 8, 10, 11, 13, 14, 16, 17, 18, 20, 22, 5, 9, 24, 15, 25, 29, 30, 28, 34, 31, 35, 36, 19, 1, 21, 26, 33, 32, 23, 27,	49	1	1
1, 2, 3, 4, 5, 25, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 24, 29,	60	1	1
1, 2, 4, 6, 7, 8, 10, 11, 13, 14, 16, 18, 20, 3, 5,	87	0	0
2,	136	0	0
2, 4, 6, 7, 8, 10, 11, 1, 13, 14, 16, 17, 18, 20, 22, 3, 5,	79	0	1
2, 24, 36, 3, 4, 1, 22, 10, 16, 17, 23, 25, 6, 7, 8, 11, 13, 14, 18, 20, 26, 34, 27, 31, 28, 29, 33, 35, 32, 30, 5, 9, 15, 21, 19, 12,	53	0	1
1, 2, 3, 4, 5, 6, 7, 8, 10,	111	0	0
1, 2, 3, 4, 6, 7, 8, 10,	114	0	0
2, 1, 4, 3, 24, 6, 7, 8, 22, 10, 11, 13, 14, 16, 17, 18, 20, 23,	84	0	0
2, 3, 13, 4, 5, 6, 8, 1, 7, 9, 10, 11, 14, 15, 16, 17, 18, 20, 12, 22, 24, 36, 23, 25, 27, 19, 21,	57	1	1
2, 23, 25, 26, 24, 28, 34, 33, 3, 4, 6, 7, 10, 16, 20, 22, 30, 8, 11, 27, 21, 5, 9, 12, 29, 31, 32, 35, 36, 13, 14, 15, 1, 17, 18, 19,	62	0	1
2, 3,	133	0	0
2, 3, 22, 4, 5, 6, 7, 8, 10, 11, 9, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 24, 32,	61	1	1
2, 3, 4, 1, 25, 5, 6, 7, 8, 10, 9, 11, 12, 13, 15, 16, 19, 20, 21, 22, 23, 24, 33, 34, 26, 27, 28, 29, 30, 31, 32, 35, 36,	40	1	1
2, 1, 3, 4, 6, 7, 8,	118	0	0
2, 3, 4,	130	0	0
1, 2,	133	0	0
1, 2, 3, 4, 6, 7, 8, 10, 11, 13, 14, 16, 17, 18, 20, 22,	82	0	0

Fig.3 - In each raw there is the navigation path of the learner, the Levenshtein distance (Lev.D.) between the navigation path and the orderly path, whether the student passed or not the final exam (1 if passed, 0 if not passed) and the prediction of it.

Table 1
DATA RELATED TO BOTH THE FIRST AND THE SECOND ESTIMATIONS

	First estimations	Second estimations
Threshold value	65	80
Number of good prediction	313	301
Number of bad predictions	94	106
Number of students	407	407
Success percentage	76.9%	73.96%

Half orderly path			
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19			
Half navigation path	**Lev.D.**	**Exam passed**	**Prediction**
2, 3, 1,	130	0	0
1, 2, 3, 4, 6, 7, 8, 10, 11, 13, 14, 16, 17, 18, 5, 9, 20, 22, 24, 25	73	1	1
2, 1, 4, 3, 6, 7, 8, 5, 9, 10, 11, 12, 15, 20, 21, 33, 13, 19, 36, 14	80	1	0
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 25, 24	70	1	1
25, 22, 2, 3, 10, 23, 11, 1,	114	0	0
1, 2, 3, 4, 6, 7, 8, 10, 11, 13, 14, 16, 17, 18, 20, 22, 23, 24, 26,	70	0	1
3, 2,	133	0	0
2, 1, 3, 6, 7, 8, 9,	118	0	0
1, 2, 3, 4, 6, 7, 8, 10, 11, 13, 14, 16, 17, 18, 20, 22,	82	0	0
13, 14, 11, 2, 3, 4, 5, 24, 22, 25, 26, 36, 33,	94	0	0
2, 4, 3, 6, 7,	124	0	0
2, 4, 16, 17, 3, 5, 9, 15, 21, 24, 36, 12, 19,	99	1	0
2, 4, 6, 7, 8, 10, 11, 3, 13, 14,	105	0	0
2, 3, 4, 6, 7, 8, 10, 11, 13, 14, 16, 17, 18, 20, 22, 5, 9, 24, 15, 2	72	1	1
1, 2, 3, 4, 5, 25, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19	78	1	1
1, 2, 4, 6, 7, 8, 10, 11, 13, 14, 16, 18, 20, 3, 5,	87	0	0
2,	136	0	0
2, 4, 6, 7, 8, 10, 11, 1, 13, 14, 16, 17, 18, 20, 22, 3, 5,	79	0	1
2, 24, 36, 3, 4, 1, 22, 10, 16, 17, 23, 25, 6, 7, 8, 11, 13, 14, 18,	81	0	0
1, 2, 3, 4, 5, 6, 7, 8, 10,	111	0	0
1, 2, 3, 4, 6, 7, 8, 10,	114	0	0
2, 1, 4, 3, 24, 6, 7, 8, 22, 10, 11, 13, 14, 16, 17, 18, 20, 23,	84	0	0
2, 3, 13, 4, 5, 6, 8, 1, 7, 9, 10, 11, 14, 15, 16, 17, 18, 20, 12, 22	79	1	1
2, 23, 25, 26, 24, 28, 34, 33, 3, 4, 6, 7, 10, 16, 20, 22, 30, 8, 11,	86	0	0
2, 3,	133	0	0
2, 3, 22, 4, 5, 6, 7, 8, 10, 11, 9, 12, 13, 14, 15, 16, 17, 18, 19, 2	80	1	0
2, 3, 4, 1, 25, 5, 6, 7, 8, 10, 9, 11, 12, 13, 15, 16, 19, 20, 21, 22	82	0	0
2, 1, 3, 4, 6, 7, 8	118	0	0

Fig.4 - In each raw there is the half navigation path of the learner, the Levenshtein distance (Lev.D.) between the half navigation path and the half orderly path, whether the student passed or not the final exam (1 or 0) and the prediction of it.

Conclusions

Learning analytics examine data coming from LMS tracks in order to provide feedbacks to learners and change their learning outcomes, but they have not reference to some specific theoretical argumentation. Many researches have examined similar LMS data and used similar models and predictors but they have found different results. Some researches reached significant percentages of estimation but they considered few numbers of learners. Recent studies underlined that is still unclear how to use data coming from LMSs for predictive modelling and, in case there is a model, its portability is a crucial step because its effectiveness depends on the learning design approach or on the technologies used to create and deliver the course. The best approaches uses much kind of data, but LMSs do not allow accessing all of them in order to apply deeper learning analytics and, generally, elaborating all these data is quite difficult and expensive.

Therefore, the goal of this paper is finding a model simpler and effective as well. We described a prediction approach adopted for learning analytics on data coming from a Moodle-based LMS. In line with the reference scenario, we defined a model that have an interesting predictive accuracy and that allows teachers monitoring their students and intervene on their learning processes before it is too late.

We started this research in the cohort 2017-2018 and went on in the following cohort 2018-2019. We experimented our model on more than 400 students.

Our model refers to navigation path of the learners and is able to provide predictions on the possible learning outcomes when the students are in the middle of their learning processes, so teachers and tutors may have enough time to identify students at risk and, eventually, to create personalized and target actions to improve their performances.

This approach could have some major benefits. From the learner point of view, since it is able to provide predictions on the performances inside an e-learning course, it may be useful to give direct feedbacks and suggest directly to the students some specific order to follow or some particular topic to go in deepening. It could be used also to compare performances among students and creating, by using some gaming approaches, new stimulus for the students themselves. However, it could be able to motivate learners and recommend them resources and activities to reach better results.

In fact, current researches in educational psychology reveals that learners do not use optimal tactics and strategies during their learning processes and, often, they are unaware of the employed study tactics. Just providing them information about the benefits of some of the effective tactics and strategies increases their chances to get better learning outcomes. Moreover, suggestions coming from Learning Analytics are more effective when the learners themselves increase their awareness of the approaches they are following or they should adopt. Therefore, the future work on user-centred learning analytics systems should be on finding the right mechanisms to communicate with the learners by means of complete and effective dashboards able to allow them optimizing their learning processes (Matcha *et al.*, 2019).

From the teacher point of view, the approach we presented in this paper may give an overall picture on the involved learners. It may help teachers monitoring their students' progresses. It may raise warnings on particular contents and activities, on particular students that are at risk of abandon or at risk in reaching their learning goals. It can identify learners needing some helps and provide them support in terms of strategies, learning styles, suggestions or, simply, motivation add-ons in order to increase the quality of teaching and, consequently, the quality of their learning processes. It may also give feedbacks about the instructional design. In fact, the results and, in particular, their relevance even if it is measured at the end of the learning experience, Missing: dashboard with results

allows getting something about the effectiveness of the course itself, about the quality of its structure, about resources and activities it contains.

Our simple-to-use learning analytics approach may be implemented in a communication dashboard for both learners and teachers. It may become the mean to raise alerts and to lead a learning support system able to visualize information and to show feedbacks and suggestions.

As it is, our model represents a quite good predictor, which could be improved by getting suggestions from other models adopted in the literature. Some of them refer to data complex to get and analyse, some other ones refer to data easier to collect.

In this research, although we did not consider any other parameters different from the navigation paths, the results are encouraging. We are confident that by relating these results with the analysis of some features more, we could define a more effective approach for learning analytics, reach better results in the prediction of learning outcomes and provide a model for support systems useful for MOOCs and any other kind of LMSs.

Acknowledgements

Even though the authors have jointly conceived the paper, Sergio Miranda edited the section “2 Reference scenario” and the section “3 Methodological approach”; Rosa Vegliante wrote the section “1 Introduction” and the section “4 Learning analytics results”. Both the authors together wrote the section “5 Conclusions”.

REFERENCES

Agudo-Peregrina, A. F., Iglesias-Pradas, S., Conde-Gonzalez, M.A., Hernandez-Garcıa, A. (2014) “Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning,” Comput. Hum. Behav., vol. 31, pp. 542–550, Feb. 2014.

Beer, C., Clark, K. and Jones, D. (2010) “Indicators of engagement,” in Australian Soc. Comput. Learn. Tertiary Educ. Annu. Conf., 2010, pp. 75–86.

Breiter, A., & Light, D. (2006). Data for School Improvement: Factors for Designing Effective Information Systems to Support Decision-Making in Schools. Educational Technology & Society, 9(3), 206-217.

Brown, J. S. , Collins, A. and Duguid, P. (1989), “Situated cognition and the culture of learning,” Educ. Res., vol. 18, no. 1, pp. 32–42, 1989.

Gaeta, M., Marzano, A., Miranda, S., Sandkuhl, K. (2016), The competence management to improve the learning engagement, Journal of Ambient Intelligence and Humanized Computing, pp. 1-13, Springer.

Gasevic, D., Dawson, S., Rogers, T. (2016) "Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success," Internet High. Educ., vol. 28, pp. 68–84, Jan. 2016.

Hoic-Bozic, N., Mornar, V., Boticki, I. (2009) "A blended learning approach to course design and implementation," IEEE Trans. Educ., vol. 52, no. 1, pp. 19–30, Feb. 2009.

Levenshtein VI (1966), Binary codes capable of correcting deletions, insertions, and reversals, in Soviet Physics Doklady, vol. 10, 1966, pp. 707–10.

Macfadyen, L. and Dawson, S. (2010) "Mining LMS data to develop an 'early warning system' for educators: A proof of concept" Comput. Educ., vol. 54, no. 2, pp. 588–599, Feb. 2010.

Marzano, A., Notti, A.M. (2014) Educational assessment: Semantic representation and ontologies. Proceedings - 2014 International Conference on Intelligent Networking and Collaborative Systems, IEEE INCoS 2014, art. no. 7057172, pp. 695-698..

Matcha, W., Ahmad Uzir, N., Gasevic, D. and Pardo, A. (2019), "A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective," in IEEE Transactions on Learning Technologies. doi: 10.1109/TLT.2019.2916802

Miranda, S., Marzano, A., Lytras, Miltiades D. (2017) A research initiative on the construction of innovative environments for teaching and learning. Montessori and Munari based psycho-pedagogical insights in computers and human behavior for the "new school", Computers in Human Behavior, Volume 66, January 2017, pp. 282-290

Miranda, S., Vegliante, R., De Angelis, M. (2019), "I processi di valutazione nell'e-learning". In: Training actions and evaluation processes. Atti del Convegno Internazionale SIRD Pensa Multimedia Pag.687-699 ISBN:978-88-6760-634-4, Training actions and evaluation processes Salerno 25-26 Oct. 2018

Moore, M. G. (1989) "Editorial: Three types of interaction," Amer. J. Distance Educ., vol. 3, no. 3, pp. 1–6, Jan. 1989

Morris, L. V., Finnegan, C. and Wu, S.-S. (2005) "Tracking student behavior, persistence, and achievement in online courses," Internet High. Educ., vol. 8, no. 3, pp. 221–231, 2005.

Mothukuri U. K. et al. (2017), "Improvisation of learning experience using learning analytics in eLearning," 2017 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH), Hyderabad, 2017, pp. 1-6. doi: 10.1109/ELELTECH.2017.8074995

Munoz-Merino, P. J., Ruiperez-Valiente, J. A., Alario-Hoyos, C. , Perez-Sanagustn, M. and Delgado Kloos, C. (2015) "Precise effectiveness strategy for analyzing the effectiveness of students with educational resources and activities in MOOCs," Comput. Human Behav., vol. 47, pp. 108–118, 2015.

Nandi, D., Hamilton, M., Harland, J. and Warburton, G. (2011) "How active are students in online discussion forums?" in Proc. 13th Australasian Comput. Educ.

Conf., 2011, vol. 114, pp. 125–134.

Piña, A. A. (2012) "An overview of learning management systems," in Virtual Learning Environments: Concepts, Methodologies, Tools and Applications, 1st ed. Louisville, KY, USA: Sullivan Univ. Syst., 2012, pp. 33–51.

Pistilli, M. D., Willis III, J. E., & Campbell, J. P. (2014). Analytics through an Institutional Lens: Definition, Theory, Design, and Impact. In Learning Analytics (pp. 79-102). Springer New York.

Rafaeli, S. and Ravid, G. (1997) "Online, web-based learning environment for an information systems course: Access logs, linearity and performance," in Proc. Inf. Syst. Educ. Conf., 1997, vol. 97, pp. 92–99.

Sergis, S., Sampson, D. (2017) "Teaching and Learning Analytics to Support Teacher Inquiry: A systematic literature review" In Learning Analytics: Fundaments, Applications and Trends, P. Ayala Ed., Springer International Publishing, 2017, 25-63.

Shankar, S. K., Prieto, L. P., Rodríguez-Triana, M. J. and Ruiz-Calleja, A. (2018) "A Review of Multimodal Learning Analytics Architectures," 2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT), Mumbai, 2018, pp. 212-214. doi: 10.1109/ICALT.2018.00057

Shum, S. B. and Ferguson, R. (2012), "Social learning analytics," Educ. Technol. Soc., vol. 15, no. 3, pp. 3–26, 2012.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. International Journal of Instructional Technology and Distance Learning, 2(1), 3-10

Siemens, G. (2013). Learning analytics: The emergence of a discipline. American Behavioral Scientist, 57(10), 1380–1400.

Siemens, G., Baker, R. S. (2012) "Learning analytics and educational data mining: Towards communication and collaboration," in Proc. 2nd Int. Conf. Learn. Analytics Knowl., 2012, pp. 252–254.

T. Yu T. and Jo, I.-H. (2014) "Educational technology approach toward learning analytics: Relationship between student online behavior and learning performance in higher education," in Proc. 4th Int. Conf. Learn. Anal. Knowl., 2014, pp. 269–270.

Tempelaar, D. T., Rienties, B., Giesbers, B. (2015) "In search for the most informative data for feedback generation: Learning analytics in a data-rich context," Comput. Human Behavior, vol. 47, pp. 157–167, Jun. 2015.

Virvou, M., Alepis, Sotirios, Christos Sidiropoulos (2015) "A learning analytics tool for supporting teacher decision", Information, Intelligence, Systems and Applications (IISA), 2015.

Zacharis, N. Z. (2015) "A multivariate approach to predicting student outcomes in web-enabled blended learning courses," Internet High. Educ., vol. 27, pp. 44–53, Oct. 2015.