Learning Analytics at UWB – First Approach		Petr Grolmus Lucie Rohlíková Faculty of Pedagogy University of West Bohemia, Czech Republic

e-mail of corresponding author: indy@cvz.zcu.cz

Key words: Learning Innovation, Higher Education, Educational Data Mining, Learning Analytics, Student Interaction Analysis, Learning Management System, Prediction Study Success

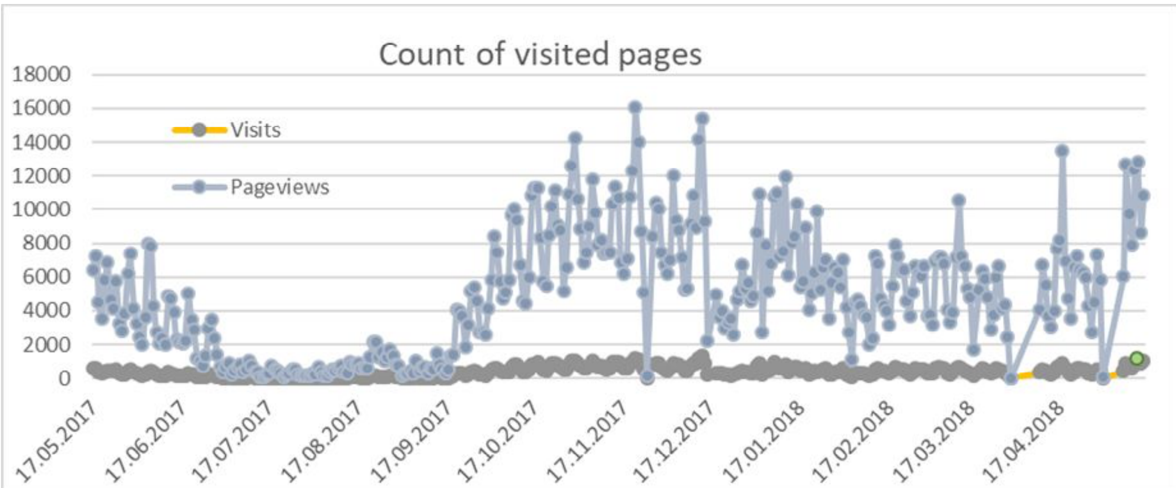
Abstract: The focus of this paper is the first look and interpretation of learning analytics data from learning management system (LMS) at the University of West Bohemia in Pilsen (UWB). We claim that there are three types of granularity of LMS data. The first type is top-level, which describes approaches and usage of LMS as a whole. The second one is course-level, which deals with the behaviour and activities of all users as a whole on a specific course. And the last user-type, which interprets the activities of users in the course, and looks for common patterns of behaviour.
This paper presents the first two types of granularity, based on real data from the university LMS. We are inspired by many previous studies focusing on learning systems of the LMS that often pay attention especially to academic success prediction or at-risk student identification (e.g. Smith et al. 2012, Jayaprakash et al., 2014, Baker et al., 2015).
These findings form the basis for further research on identifying user behaviour on the course and identifying students at risk of learning failure.

Introduction

At the University of West Bohemia in Pilsen (UWB), we have had a great deal of experience in running the various Learning Management Systems (LMS) over the last two decades. It can be said that Moodle, which had originally "enthusiastically" been running since 2006 (the year of the first course), has won the internal e-learning system battle. In 2010, Moodle became an official part of the university computing environment, and was fully associated with both Single Sign-On and Student Agenda STAG, which included all the students' learning outcomes.

Currently (May 2018), we have around 1900 courses in Moodle. In order to maintain persistence, LMS does not delete old courses, and the results or assignments of the students may at any time be backward traced. The fact that Moodle at UWB has become an integral part of the teaching is clearly illustrated by the graph of the number of visitors in the last year Figure 1.

Figure 1. Total number of visitors for the year (May 2017 – May 2018)



Source: Own

From Figure 1 there are also obvious decreases in the activity of users during the main school holidays and during the examination period, respectively just before the beginning of the next semester. Local, regularly recurring drops in attendance correspond to weekends.

Evidence of the steadily increasing popularity of LMS at UWB as a support system for ordinary education also shows visible jumps in the increase in attendance between two semesters, ending in the summer semester of the academic year 2016/2017 (May-June 2017 in the graph) and the winter semester of the academic year 2017/2018 from (from September to December).

The high rate of attendance for e-learning courses generates a large amount of data - logs both in LMS itself, and in the Web server service that provides communication with end users.

For a long time, this information about user behaviour during the learning process and testing has remained untapped, and in fact, only tied up valuable server resources. This has only changed in recent years, with the development of Educational Data Mining (EDM) and Learning Analytics (LA). EDM was on the rise in 2008-2009 (Romero, Ventura 2013), although its origins can be dated back to 2005 (Romero, Ventura 2007). The younger LA field can be dated back to 2010-2011 (Ferguson 2012; Juhaňák Zounek,2016).

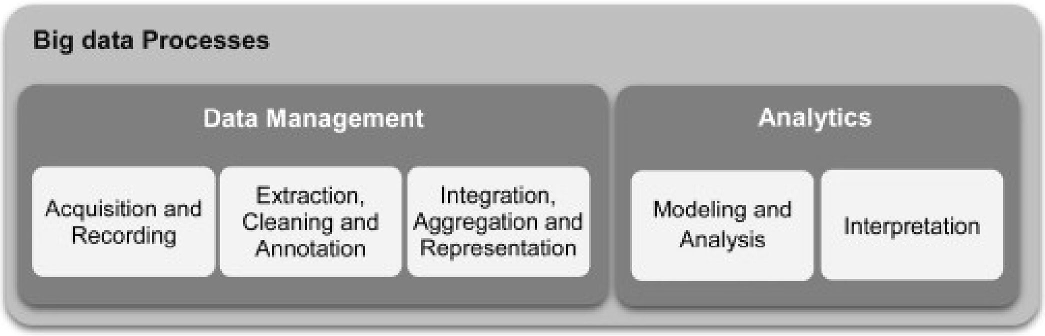
Both EDM and LA research fields differ in many respects, but what they have in common is that they seek to maximize information and context from the data generated and stored within LMS itself. To this end, they use various analytical and data mining methods and procedures to gain important information and knowledge about how students behave, learn, perform tasks, and take tests (Juhaňák et al., 2017).

(Siemens et al., 2011) define the field of LA as:

"Learning analytics is 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. Learning analytics are largely concerned with improving learner success."

The same source states that LA is a special representation of the "Big Data" application and the analysis of education. Activities related to Big Data are generally divided into two main parts: data management and analysis itself (Gandomi, Haider 2015). Data management is further subdivided into data acquisition, logging, data extraction, and data purification, and their subsequent integration, aggregation, and representation. With the data thus obtained, modelling, analysis, and subsequent interpretation of the results are carried out - see Figure 2.

FIGURE 2. BIG DATA PROCESS



Source: Gandomi, Haider, 2015

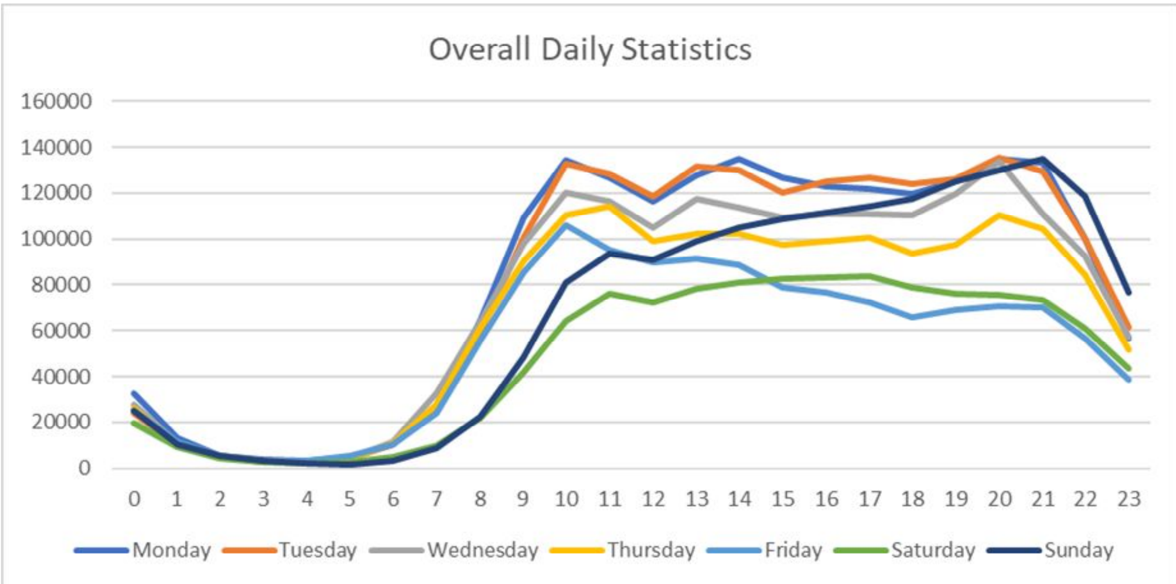
Within the context of increasing needs for quality and evaluation of teaching at UWB, we have primarily focused on eLearning, specifically for reasons previously described - we already own a huge amount of data in LMS and web server logs that can be easily used for EDM research, respectively for LA. This article describes the first experience of UWB in LA, and as such, can be used as a starting point for other interested parties and organizations.

General overview of the data

In our LMS, we have collected data for several years back, and this data can immediately be used to create overall statistics. The first view of this data may be from Figure 1 above, which shows the rate of traffic for the last calendar year.

Another interesting statistic from the overall data is the total number of visitors for each individual day of the week - see Figure 3 - Overall Daily Statistics.

FIGURE 3. OVERALL DAILY STATISTICS



Source: Own

Already in this rough overview of the overall data, some interesting facts can be drawn upon. First of all, the chart's progress for each day of the week is more or less similar, only differing in absolute numbers of accesses. The minimal, almost zero-hour operation was not surprising for us, and only confirmed the correct decision to make regular backups of the LMS system at 4:00 each morning. No surprise was the 2-hour delay in traffic growth during the weekend.

More surprising was the discovery that the students' overall learning habits between 10:00 and 21:00 of all days' operation was evenly stable. A small drop between 11:00 and 12:00 on all days corresponds to lunch times. Another surprise seen in the graph is the all-day slow increase in Sunday traffic, which culminates around the 21.00, hour and equals business hours on business days. Sunday traffic between 21:00 and 23:00, even in total values, exceeds traffic on other days.

Course statistics

In the case of UWB, this is the first piece of work in LA from the data of our long-running Moodle LMS. The first step is therefore the need to define a framework concept and an initial strategy. Our first step has been minimalistically defined following mature thought as follows:

We select only one course to process,

we have complete logs of student actions and/or activities on the course,

the course must have a larger number of students, ideally more than 100, in order to create a statistically significant sample,

the course must have the tasks evaluated by a teacher (feedback),

the course must be repeated every academic year, so that we can compare the results and possibly predict students at risk of learning failure on the basis of earlier or previous data,

know the result of the subject study - i.e. either a credit with a completed/incomplete assessment, or a mark/grade for the subject 1 (best), 2, 3, or 4 (did not fulfil the subject).

From the shortlist, we chose the course "Information Technology in Teaching", which was completed by a credit, taught by the Department of Computer Sciences and Didactic Technology at the Faculty of Education, UWB. In the academic year 2016/2017, a total of 293 students were enrolled, according to the study agenda. Of these, there were a total of 51 combined study and 242 full-time students.

Since it is possible to assume that the behaviour of students of the faculty and the combined form of study will be different, students of the combined form did not use the chosen course, so the further processing only concerns the 242 full-time students.

An interesting and unexpected bonus of the chosen course was the fact that students were not enrolled in the course through the functional Moodle-STAG web service, which normally gives access to all students who have a subject under the study agenda. At the first attendance meeting with the students of the subject, the teacher published an access password for students enrolling in the course themselves. Therefore, we gained another possible indicator for student behaviour.

Only 227 students enrolled in the chosen course. The difference between the 242 students enrolled in the STAG study programme and the 227 enrolled in the course is due to students who stopped studying for various reasons. The successful completion of the course was achieved by 199 students. Unfortunately, 28 students failed this subject.

Another monitored indicator was the number of sessions attended in the Moodle system, according to the individual students enrolled in the course. In total, 35,359 visits by students to this course were registered. Table 1 clearly shows that successful students returned to the course more often than failed students.

TABLE 1 COURSE VISITS

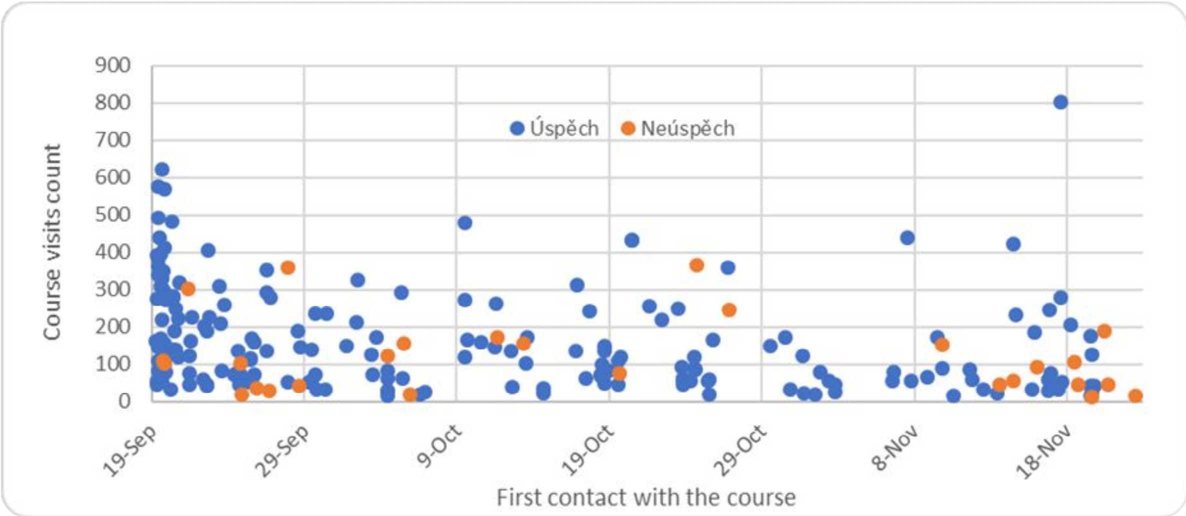
	199 successful students	28 unsuccessful students
Number of visits	32174	3185
Average	161.68	113.75

Source: Own

Since we knew about the success rate for each individual student, the number of visits, and the date of the first contact with the course, we tried to show these figures in the chart, and see if some basic characteristics would be evident - see Figure 4 - Students on the course. The assumption that successful students start studying earlier, and failed students start studying later was not clearly proven. Students spread their first contact with the course throughout the semester. However, in the beginning, there is a predominance of successful students who earned credits. Also, at the end, a relatively frequent occurrence of unsuccessful students can be observed.

It’s also worth mentioning that students who enrolled in the course for the first time on 17th November 2016 - just before the credit test on 21st November 2016, achieved the highest attendance rate - 802 course displays. You can find this in the graph as a point on the right, at the top.

FIGURE 4. STUDENTS ON COURSE



Source: Own

Conclusion and future research

Our organization has long endeavoured to provide quality education in order to prepare students for their upcoming professional lives to the fullest extent possible. Every year, the Commission for Quality at UWB processes student feedback on the subjects taught. They anonymously, in an easy-to-read form, "mark", for example, the clarity of the interpretation, the benefit of the subject for the chosen field, its usefulness, and experience gained during the practical exercises. However, filling in the questionnaire is not an obligation, so the feedback depends on the goodwill of the evaluators, who may respond randomly, or even with the aim of damaging the results of the evaluation of a particular subject. According to the latest statistics from the Quality Commission, only 15-20% of students from all nine faculties of UWB provide any feedback at all.

In the form of LA, we now have another usable metric, based on the behaviour and results of the students' subjects. Metrics are obtained independently of the students', from data stored in the LMS. A metric that cannot be easily underestimated by misleading responses to the questionnaires.

So far we have only dealt with data obtained from one subject. Due to the suitably chosen entry condition - the subject must be taught regularly in subsequent academic years - we will be able to easily compare existing results with new results in the future. Moreover, thanks to the knowledge of the behaviour of unsuccessful students from previous years, we are able, with a certain degree of accuracy, to identify students who are at risk of learning failure in the currently running instance of the same course. The teacher of the subject, respectively the tutor of the course, can use this knowledge to focus more on the group of these "at-risk" students, and try to motivate them to successfully master the subject.

An interesting result could be the comparison of the same subject of full-time students with combined students. We assume that the learning style of both groups will vary greatly. The age group of undergraduate students is de facto devoted to study, while the second group of students of combined study already performs their chosen profession. In addition, these students have often already established families, and time to study has to be carved out of their very valuable free time or from their time set aside for sleeping.

In the above example, we have tried the methods of obtaining data relevant to one course in LMS. During this first attempt, it took us a long time to untangle some links within the LMS database, which has more than 300 differently interconnected entities. The experience gained can now be used much faster for larger groups of courses that meet the criteria defined above.

References

Baker, R. S.J.D., Lindrum, D., Lindrum, M. J., & Perkowski, D. 2015. "Analyzing Early At-Risk Factors in Higher Education e-Learning Course" in: Proceedings of the 8th International Conference on Educational Data Mining, Madrid, Spain, Jun 26-29, 2015.

Ferguson, R. 2012. "Learning analytics: drivers, developments and challenges". International Journal of Technology Enhanced Learning, 4 (5/6), pp. 304–317.

Gandomi, A., Haider, M. 2015. "Beyond the hype: Big data concepts, methods, and analytics." International Journal of Information Management, 35 (2), pp.137-144.

Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. 2014. "Early alert of academically at-risk students: An open source analytics initiative." Journal of Learning Analytics, 1(1), pp. 6-47.

Juhaňák, L., Zounek, J. 2016. "Analytika učení: nový přístup ke zkoumání učení (nejen) ve virtuálním prostředí." Pedagogická orientace, 26 (3), pp. 560–583.

Juhaňák, L., Zounek, J. Rohlíková, L. 2017. "Using process mining to analyze students' quiz-taking behavior patterns in a learning management system.", in: *Computers in Human Behavior*, Elsevier, in press.

Romero, C., Ventura, S. 2013. "Data mining in education". WIREs Data Mining and Knowledge Discovery, 3, pp. 12–27.

Romero, C., Ventura, S. 2007. "Educational data mining: a survey from 1995 to 2005". Expert Systems with Applications, 33(1), 135–146.

Siemens, G. et al. 2011 Open Learning Analytics: an integrated & modularized platform, http://solaresearch.org/wp-content/uploads/2011/12/OpenLearningAnalytics.pdf, accessed on 24th July 2018.

Smith, V. C., Lange, A., & Huston, D. R. 2012. “Predictive Modeling to Forecast Student Outcomes and Drive Effective Interventions in Online Community College Courses.” Journal of Asynchronous Learning Networks,16 (3), pp. 51-61.