LEARNING PERSONALISATION IN VIRTUAL LEARNING ENVIRONMENTS APPLYING LEARNING ANALYTICS

Eugenijus Kurilovas¹,², Jelena Mamcenko¹, Irina Krikun²

¹ Vilnius Gediminas Technical University (LITHUANIA)
² Vilnius University Institute of Mathematics and Informatics (LITHUANIA)

Abstract
The paper aims to analyse application of learning analytics / educational data mining (LA / EDM) to support learning personalisation and optimisation in virtual learning environment Moodle. LA / EDM are known as the measurement, collection, analysis, and reporting of data about learners and their contexts to understand and optimise learning and environments in which it occurs. In the paper, appropriate literature review is performed on LA / EDM methods and techniques that could be applied to personalise students’ learning in Moodle. After that, the authors’ original methodology to personalise learning is presented. First of all, existing Moodle-based learning activities and tools are analysed to be further interlinked with appropriate students’ learning styles. For this purpose, Felder-Silverman learning styles model (FSLSM) is applied in the research. Students’ learning styles according to FSLSM are interlinked with the most suitable Moodle-based learning activities and tools using expert evaluation method. After that, a group of students is analysed in terms of identifying their individual learner profiles according to Soloman-Felder index of learning styles questionnaire. After identifying individual learner profiles, probabilistic suitability indexes are calculated for each analysed student and each Moodle-based learning activity to identify which learning activities or tools are the most suitable for particular student. From theoretical point of view, the higher is probabilistic suitability index the better learning activity or tool fits particular student’s needs. On the other hand, students practically used some learning activities or tools in real learning practice in Moodle before identifying the aforementioned probabilistic suitability indexes. Here we could hypothesise that students preferred to practically use particular Moodle-based learning activities or tools that fit their learning needs mostly. Thus, using appropriate LA / EMD methods and techniques, it would be helpful to analyse what particular learning activities or tools were practically used by these students in Moodle, and to what extent. After that, the data on practical use of Moodle-based learning activities or tools should be compared with students’ probabilistic suitability indexes. In the case of any noticeable discrepancies, students’ profiles and accompanied suitability indexes should be identified more precisely, and students’ personal leaning paths in Moodle should be corrected according to new identified data. In this way, after several iterations, we could noticeably enhance students’ learning quality and effectiveness.

Keywords: learning analytics, educational data mining, learning personalisation, virtual learning environments, students’ learning styles.

1 INTRODUCTION
The paper aims to analyse application of learning analytics / educational data mining (LA / EDM) to support learning personalisation and optimisation in virtual learning environment (VLE) e.g. Moodle. LA is the analysis of electronic learning data which allows teachers, course designers and administrators of VLEs to search for unobserved patterns and underlying information in learning processes.

Learning personalisation is helpful to enhance learning quality and effectiveness.

Learning personalisation by applying learning styles and intelligent technologies became very popular topic in scientific literature during last few years [1], [2], [3], [4], [5], [6], [7], [8], [9], [10].

Personalisation can be seen from two different perspectives, namely, while only one learning object [11], [12], [13], [14] or a learning unit / scenario [15], [16], [17] is selected, and while a set of them is composed, i.e. personalisation of a learning unit / scenario by finding a learning path [7]. The former perspective formulates learning objects selection problem, and the latter one solves curriculum sequencing problem [18].
Personalised learning units / scenarios are referred here as learning units / scenarios composed of the learning components having the highest probabilistic suitability indexes [19] to particular students according to Felder-Silverman Learning Styles Model [20].

In the paper, first of all, systematic review was performed in Clarivate Analytics (formerly Thomson Reuters) Web of Science database. The following research questions have been raised to perform systematic literature review: “What are existing LA / EDM methods, tools, and techniques applied to support personalised learning in VLEs / Learning Management Systems (LMSs)?”

After that, the author’s original learning personalisation methodology based on identifying students’ learning styles and other needs is presented in more detail. At the end, some insights on possible application of LA / EDM to support personalised learning in VLE Moodle are provided.

The rest of the paper is organised as follows: systematic review on LA / EDM application in VLEs is provided in Section 2, the author’s original learning personalisation methodology applying LA / EDM and based on identifying students’ learning styles and other needs is presented in Section 3, and Section 4 concludes the paper.

2 SYSTEMATIC REVIEW

During XXI century (2001-2017), 82 publications (from which – 35 articles) in English were found on March 26, 2017, in Web of Science database on the topic “TS=(virtual learning environment* AND learning analytics)”, and 604 publications (from which – 264 articles) – on the topic “TS=(learning management system* AND data mining)” (Fig. 1).

2.1 Search History: 26 MAR

<table>
<thead>
<tr>
<th>Results</th>
<th>Save History / Create Alert</th>
<th>Open Saved History</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>(TS=(virtual learning environment* AND learning analytics)) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article) Indexes=SCI-EXPANDED, SSCI, A&amp;HCI, CPCI-S, CPCI-SSH, ESCI Timespan=2001-2017</td>
<td></td>
</tr>
<tr>
<td>264</td>
<td>(TS=(learning management system* AND data mining)) AND LANGUAGE: (English) AND DOCUMENT TYPES: (Article) Indexes=SCI-EXPANDED, SSCI, A&amp;HCI, CPCI-S, CPCI-SSH, ESCI Timespan=2001-2017</td>
<td></td>
</tr>
<tr>
<td>604</td>
<td>(TS=(learning management system* AND data mining)) AND LANGUAGE: (English) Indexes=SCI-EXPANDED, SSCI, A&amp;HCI, CPCI-S, CPCI-SSH, ESCI Timespan=2001-2017</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>(TS=(virtual learning environment* AND learning analytics)) AND LANGUAGE: (English) Indexes=SCI-EXPANDED, SSCI, A&amp;HCI, CPCI-S, CPCI-SSH, ESCI Timespan=2001-2017</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Search history.

After applying B. Kitchenham’s systematic review methodology [21], on the last stage 10 newest most suitable articles were identified to further detailed analysis on possible application of LA / EDM to support learning in VLEs.
According to [22], the heterogeneity of external systems that can be connected in an e-learning environment can impose interoperability and performance requirements for recording and storing the learning data. Web-based protocols have been developed to improve e-learning systems’ interoperability and capability to perform meaningful analytics. The paper [22] describes a web-based learning environment aimed at training how to command and control unmanned autonomous vehicles, provided with analytic capabilities. It integrates an external web content management system and a simulation engine that present different performance requirements for recording all significant events that occur during the learning process. Its record store construction, based on standard interoperability protocols, is explored in the paper from the performance viewpoint. The tests that were conducted to assess regular data stores used for learning analytics show that performance should not be overlooked when constructing and deploying learning analytics systems.

The paper [23] claims that despite the great potential of social network analysis methods and visualisations for learning analytics in computer-supported collaborative learning, these approaches have not been fully explored due to two important barriers: the scarcity and limited functionality of built-in tools in LMSs, and the difficulty to import educational data from formal VLEs into social network analysis programs. The study [22] aims to cover that gap by introducing GraphFES, an application and web service for extraction of interaction data from Moodle message boards and generation of the corresponding social graphs for later analysis using Gephi, a general purpose SNA software. In addition, this paper briefly illustrates the potential of the combination of the three systems (Moodle, GraphFES and Gephi) for social LA using real data from a computer-supported collaborative learning course with strong focus on teamwork and intensive use of forums.

The main objective of the paper [24] is to analyse the effect of the affordances of a VLE and a personal learning environment (PLE) in the configuration of the students’ personal networks in a higher education context. The results are discussed in light of the adaptation of the students to the learning network made up by two undergraduate, inter-university and online courses. Besides, the authors also examine the influence of this effect in the learning process. The findings reflect the effectiveness of a PLE for facilitating student participation and for assisting students in the creation of larger and more balanced personal networks with richer social capital. However, the findings do not provide evidences about a difference in the learning performance between the two environments. From a methodological point of view, this paper serves as an illustration of the analysis of personal networks on digital data collected from technology-enhanced learning environments.

According to [25], one of the main challenges in teaching and learning activities is the assessment: it allows teachers and learners to improve the future activities on the basis of the previous ones. It allows a deep analysis and understanding of the whole learning process. This is particularly difficult in VLEs where a general overview is not always available. In the latest years, LA are becoming the most popular methods to analyse the data collected in the learning environments in order to support teachers and learners in the complex process of learning. If they are properly integrated in learning activities, indeed, they can supply useful information to adapt the activities on the basis of student’s needs. In this context, the paper presents a solution for the digitally enhanced assessment. Two different Learning Dashboards have been designed in order to represent the most interesting LA aiming at providing teachers and learners with easy understandable view of learning data in VLEs.

The authors of [26] consider that the future of educational technology has been envisioned to have increasing focus on simulations, game based learning, VLEs and virtual worlds. The technologies aim to provide authentic learning and enable deeper, more complex and contextual understanding for students. To study the impact of VLEs for natural sciences and engineering education, the authors have designed and implemented a virtual laboratory, LabLife3D, in Second Life. The authors have designed six virtual laboratory exercises in the biological sciences and chemistry and additionally created a system to gather behaviouristic data during laboratory simulations for the purpose of LA. This paper presents the design process of laboratory exercises and discusses the content-specific learning goals and outcomes. Additionally, this paper discusses the use of heuristic usability review used to improve the VLE. Lastly, the results from student and teacher interviews are presented in [26], together with results of the LA study. The discussion also includes student identified affordances and barriers for learning. The authors conclude that authentic and deep learning is possible within virtual worlds. Furthermore, the results of this study are not only limited to virtual worlds, but could also apply to other areas of digital educational technology.

According to [27], LA is the analysis of electronic learning data which allows teachers, course designers and administrators of VLEs to search for unobserved patterns and underlying information in learning processes. The main aim of LA is to improve learning outcomes and the overall learning
process in electronic learning virtual classrooms and computer-supported education. The most basic unit of learning data in VLEs for LA is the interaction, but there is no consensus yet on which interactions are relevant for effective learning. Drawing upon extant literature, this research defines three system-independent classifications of interactions and evaluates the relation of their components with academic performance across two different learning modalities: VLE-supported face-to-face (F2F) and online learning. In order to do so, the authors performed an empirical study with data from six online and two VLE-supported F2F courses. Data extraction and analysis required the development of an ad hoc tool based on the proposed interaction classification. The main finding from this research is that, for each classification, there is a relation between some type of interactions and academic performance in online courses, whereas this relation is non-significant in the case of VLE-supported F2F courses. Implications for theory and practice are discussed next.

The paper [28] claims that the interest in developing LA tools that can be integrated into the well-known Moodle course management systems is increasing nowadays. These tools generally provide some type of basic analytics and graphs about users’ interaction in the course. However, they do not enable a varied set of Data Mining techniques to be applied, such as approaches for classification, clustering, or association. To address this issue, a new and freely available Moodle Data Mining tool, named MDM, has been proposed in this paper. The proposed tool eases the whole knowledge discovery process, including tasks such as selection, data pre-processing, and data mining from Moodle courses. The proposed MDM tool has been developed in PHP programming language, so it can be easily integrated into Moodle as a module for a specific course. Its main features and architecture are described in depth, and a tutorial is also provided as a practical way of using the MDM interface. Finally, some experimental results using a real-life sample dataset of mechanical engineering students are analysed.

The authors of [29] think that the use of LMSs has grown exponentially in the last several years and has come to have a strong effect on the teaching-learning process, particularly in higher education. The present study intends to examine students’ asynchronous learning processes via an EDM approach using data extracted from the Moodle logs of students who were grouped according to similar behaviours regarding effort, time spent working, and procrastination. The behaviours were then matched with different levels of achievement. First, the different patterns of students’ involvement in the learning process in a LMS were clustered. Second, the different variables selected from the Moodle records were studied to see if they were equally suitable for the configuration of student clusters. Third, the relationships between those patterns to students’ final marks were examined. After analysing the log data gathered from a Moodle 2.0 course in which 140 undergraduate students were enrolled, four different patterns of learning with different final marks were found. Additional results showed that there are variables more related to achievement and more suitable to group the students on the basis of which the different groups were characterised, namely, two Task Oriented Groups (socially or individually focused) and two Non Task Oriented Groups (procrastinators or non-procrastinators). These results have implications in the design of interventions for improving students’ learning processes and achievement in LMSs.

According to [30], blended learning is recognised as one of the major trends in higher education today. To identify how blended learning has been actually adopted, this study employed a data-driven approach instead of model-driven methods. Latent Class Analysis method as a clustering approach of EDM was employed to extract common activity features of 612 courses in a large private university located in South Korea by using online behaviour data tracked from LMS and institution’s course database. Four unique subtypes were identified. Approximately 50% of the courses manifested inactive utilisation of LMS or immature stage of blended learning implementation, which is labelled as Type I. Other subtypes included Type C - Communication or Collaboration (24.3%), Type D - Delivery or Discussion (18.0%), and Type S - Sharing or Submission (7.2%). The authors discussed the implications of blended learning based on data-driven decisions to provide strategic institutional initiatives.

Systematic review has shown that LA / EDM are already quite actively used in VLEs e.g. Moodle to solve different problems e.g. academic assessment, predicting students’ success and dropout, predicting instructional effectiveness of VLEs, etc. At the same time, LA / EDM are still rarely used to personalise learning in VLEs according to students’ needs, and future research is needed in the area.
3 LEARNING PERSONALISATION METHODOLOGY APPLYING LEARNING ANALYTICS IN VLE MOODLE

According to [31], learning software and all learning process should be personalised according to the main characteristics / needs of the learners. Learners have different needs and characteristics, that is, prior knowledge, intellectual level, interests, goals, cognitive traits (working memory capacity, inductive reasoning ability, and associative learning skills), learning behavioural type (according to his / her self-regulation level), and, finally, learning styles.

In personalised learning, first of all, integrated learner profile (model) should be implemented using e.g. Soloman-Felder Index of Learning Styles Questionnaire [32] [19]. After that, interlinking of learning components (learning objects, activities, and environments) with learners’ profiles should be performed, and an ontologies-based personalised recommender system should be created to suggest learning components suitable to particular learners according to their profiles [31].

Interlinking and ontologies creation should be based on the expert evaluation results (e.g. [33]). Experienced experts should evaluate learning components in terms of its suitability to particular learners according to their learning styles and other preferences / needs. A recommender system should form the preference lists of the learning components according to the expert evaluation results.

Probabilistic suitability indexes [19] should be identified for all learning components in terms of its suitability level to particular learners. These suitability indexes could be easily calculated for all learning components and all students if one should multiply learning components’ suitability ratings obtained while the experts evaluate suitability of the learning components to particular learning styles (like in [33]) by probabilities of particular students’ learning styles (like in [19]). These suitability indexes should be included in the recommender system, and all learning components should be linked to particular students according to those suitability indexes. The higher the suitability indexes, the better the learning components fit the needs of particular learners.

Thus, personalised learning units / scenarios (i.e. personalised methodological sequences of learning components) could be created for particular learners. An optimal learning unit / scenario (i.e. learning unit of the highest quality) for particular student means a methodological sequence of learning components having the highest suitability indexes.

A number of intelligent technologies should be applied to implement this methodology, for example, ontologies, recommender systems, intelligent software agents, decision support systems to evaluate quality of learning units / scenarios etc.

This pedagogically sound learning units / scenarios personalisation methodology is aimed at improving learning quality and effectiveness. Learning unit / scenario of the highest quality for particular student means a methodological sequence of learning components with the highest suitability indexes. The level of students’ competences, that is, knowledge / understanding, skills and attitudes / values directly depends on the level of application of high-quality learning units / scenarios in real pedagogical practice.

Existing Moodle-based learning activities and tools should be analysed to be further interlinked with appropriate students’ learning styles. For this purpose, Felder-Silverman learning styles model (FSLSM) [20] should be applied. Students’ learning styles according to FSLSM should be interlinked with the most suitable Moodle-based learning activities and tools using expert evaluation method [33].

FSLSM classifies students according to where they fit on a number of scales pertaining to the ways they receive and process information:

a) By information type: (1) Sensory (SEN) – concrete, practical, oriented towards facts and procedures vs (2) Intuitive (INT) – conceptual, innovative, oriented towards facts and meaning;

b) By sensory channel: (3) Visual (VIS) – prefer visual representations of presented material e.g. pictures, diagrams, flow charts vs (4) Verbal (VER) – prefer written and spoken explanations;

c) By information processing: (5) Active (ACT) – learn by trying things out, working with others vs (6) Reflective (REF) – learn by thinking things through, working alone; and

d) By understanding: (7) Sequential (SEQ) – linear, orderly, learn in small incremental steps vs (8) Global (GLO) – holistic, systems thinkers, learn in large leaps [20].

In Table 1, the main Moodle tools / activities are interlinked with the most suitable learning styles according to FSLSM.
<table>
<thead>
<tr>
<th>Activity</th>
<th>Description</th>
<th>Most suitable learning styles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignments</td>
<td>Enable teachers to grade and give comments on uploaded files and assignments created on and off line</td>
<td>REF</td>
</tr>
<tr>
<td>Chat</td>
<td>Allows participants to have a real-time synchronous discussion</td>
<td>ACT</td>
</tr>
<tr>
<td>Choice</td>
<td>A teacher asks a question and specifies a choice of multiple responses</td>
<td>INT</td>
</tr>
<tr>
<td>Database</td>
<td>Enables participants to create, maintain and search a bank of record entries</td>
<td>ACT</td>
</tr>
<tr>
<td>External tool</td>
<td>Allows participants to interact with Learning Tools Interoperability compliant learning resources and activities on other web sites</td>
<td>ACT</td>
</tr>
<tr>
<td>Feedback</td>
<td>For creating and conducting surveys to collect feedback</td>
<td>ACT</td>
</tr>
<tr>
<td>Forum</td>
<td>Allows participants to have asynchronous discussions</td>
<td>ACT</td>
</tr>
<tr>
<td>Glossary</td>
<td>Enables participants to create and maintain a list of definitions, like a dictionary</td>
<td>INT, GLO</td>
</tr>
<tr>
<td>Lesson</td>
<td>For delivering content in flexible ways</td>
<td>SEN, SEQ</td>
</tr>
<tr>
<td>Quiz</td>
<td>Allows the teacher to design and set quiz tests, which may be automatically marked and feedback and/or to correct answers shown</td>
<td>REF, SEN, SEQ</td>
</tr>
<tr>
<td>SCORM</td>
<td>Enables SCORM packages to be included as course content</td>
<td>REF, SEN, SEQ</td>
</tr>
<tr>
<td>Survey</td>
<td>For gathering data from students to help teachers learn about their class and reflect on their own teaching</td>
<td>REF, GLO</td>
</tr>
<tr>
<td>Wiki</td>
<td>A collection of web pages that anyone can add to or edit</td>
<td>ACT, GLO</td>
</tr>
<tr>
<td>Workshop</td>
<td>Enables peer assessment</td>
<td>ACT</td>
</tr>
</tbody>
</table>

Next, students should be analysed in terms of identifying their individual learner profiles according to [32]. After identifying individual learner profiles, probabilistic suitability indexes [19] should be calculated for each analysed student and each Moodle-based learning activity to identify which learning activities or tools are the most suitable for particular student. From theoretical point of view, the higher is probabilistic suitability index the better learning activity or tool fits particular student’s needs.

On the other hand, students practically used some learning activities or tools in real learning practice in Moodle before identifying the aforementioned probabilistic suitability indexes. Here we could hypothesise that students preferred to practically use particular Moodle-based learning activities or tools that fit their learning needs mostly.

Thus, using appropriate LA / EMD methods and techniques, it would be helpful to analyse what particular learning activities or tools were practically used by these students in Moodle, and to what extent.

The basic LA / EDM techniques and their application in VLE Moodle examples are shown in Table 2. These techniques can be used together or one after the other, depending on the complexity of the task solved.
Table 2. EDM techniques and application examples

<table>
<thead>
<tr>
<th>LA / EDM techniques</th>
<th>Application examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification</td>
<td>To classify each item in a set of data into one of predefined set of learners group</td>
</tr>
<tr>
<td>Clustering</td>
<td>To determine groups of students that need special course profiling</td>
</tr>
<tr>
<td>Association rules</td>
<td>To discover interesting relations between course elements which were used by particular students</td>
</tr>
<tr>
<td>Prediction</td>
<td>To predict dependencies of using Moodle activities and final student’s learning outcomes</td>
</tr>
</tbody>
</table>

To determine and to set appropriate algorithm to a new data set is a difficult task because there is no single classificatory which equally well suited for all data sets. In practice it is very important to choose the proper classification / clustering or other algorithm to a particular data set.

After that, the data on practical use of Moodle-based learning activities or tools should be compared with students’ probabilistic suitability indexes. In the case of any noticeable discrepancies, students’ profiles and accompanied suitability indexes should be identified more precisely, and students’ personal learning paths (i.e. learning units / scenarios) in Moodle should be corrected according to new identified data. In this way, after several iterations, we could noticeably enhance students’ learning quality and effectiveness.

4 CONCLUSION

Systematic review has shown that LA / EDM are already quite actively used in VLEs e.g. Moodle to solve different problems e.g. academic assessment, predicting students’ success and dropout, predicting instructional effectiveness of VLEs, etc. At the same time, LA / EDM are still rarely used to personalise learning in VLEs according to students’ needs, and future research is needed in the area.

In the paper, original learning personalisation methodology applying LA / EDM in VLEs e.g. Moodle is presented. According to this methodology, first of all, Moodle-based learning activities / tools should be analysed to be further interlinked with appropriate students’ learning styles. Students’ learning styles should be interlinked with the most suitable Moodle-based learning activities / tools using expert evaluation method (Table 1).

Second, groups of students should be analysed in terms of identifying their individual learner profiles. After that, probabilistic suitability indexes should be calculated for each analysed student and each Moodle-based learning activity / tool to identify which learning activities / tools are the most suitable for particular student. From theoretical point of view, the higher is probabilistic suitability index the better learning activity / tool fits particular student’s needs.

On the other hand, students practically use some learning activities / tools in real learning practice in Moodle before identifying appropriate probabilistic suitability indexes. Here we could hypothesise that students preferred to practically use particular Moodle-based learning activities / tools that fit their learning needs mostly. Thus, using appropriate LA / EDM methods and techniques, it would be helpful to analyse what particular learning activities / tools were practically used by these students in Moodle, and to what extent.

In the paper, basic LA / EDM techniques and their application in VLE Moodle examples are presented (Table 2). These techniques are as follows: classification, clustering, association rules, and prediction. These techniques can be used together or one after the other, depending on the complexity of the task solved. To determine and to set appropriate algorithm to a new data set is a difficult task because there is no single classificatory which equally well suited for all data sets. In practice it is very important to choose the proper classification / clustering or other algorithm to a particular data set.

After that, the data on practical use of Moodle-based learning activities / tools should be compared with students’ probabilistic suitability indexes. In the case of any noticeable discrepancies, students’ profiles and accompanied suitability indexes should be identified more precisely, and students’
personal leaning paths in Moodle should be corrected according to new identified data. In this way, after several iterations, we could noticeably enhance students’ learning quality and effectiveness.

REFERENCES


