Student Behavior in a Web-Based Educational System: Exit Intent Prediction

Ondrej Kassak, Michal Kompan, Maria Bielikova[1]
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava
Ilkovicova 2, 842 16 Bratislava, Slovakia
{name.surname}@stuba.sk

Abstract:

The behavior of users over the web is one of the most relevant and research topic nowadays. Not only mining the user's behavior in order to provide better content is popular, but the prediction of the user's behavior is interesting and can increase user experience. Moreover, the business clearly desires such information to improve their services. In this paper we focus to the education domain as it belongs to the most dynamically transforming areas. Web based e-learning systems are nowadays reaching still greater popularity, because of possibilities they offer to students. We analyze various sources of "e-students" feedback and discuss today's challenges from the logging and feedback collecting point of view. Next, we focus on the prediction of student's next action within an e-learning application (in the mean of "stay or leave?" question). Such information can improve students' attrition rate by introducing various personalized approaches. We proposed the classifier based on polynomial regression and stochastic gradient descent to learn the attributes importance. In this way we are able to process a stream of data in one single iteration and thus we are able to reflect dynamic users' behavior changes. Our experiments are based on the log data collected from our web-based education system ALEF during three-year period. We found that there is an extensive heterogeneity in the users' (student) behavior which we were able to handle by using individual weights calculated for every user.

Keywords:

User-feedback, E-Learning, Behavior prediction, Classification, Stochastic gradient descent

Highlights:

-We proposed a classifier predicting if the student ends the session in next action

-Proposed classifier processes the dynamically changing data streams of logs

-List of most important attributes for the session end prediction is identified

-Analysis of students' feedback and behavior in a web-based educational system is provided

1. Introduction

Every one of us is a unique person who responds differently to perceptions obtained from the environment. The task of interaction with a software and various systems can be problematic for this reason, since the systems are mainly designed to operate in a one strictly defined way, regardless of user who interact with them. Nowadays an increasing amount of web-based systems use personalization, because it allows to match the content to specific user's needs and preferences. This process may take many forms – it can be an adaptation of a content presented to the user, a change of a search results order, an arrangement or a change of system interface components appearance etc.

[1] Corresponding author

In order to provide adaptive, personalized or specifically adjusted content or service, the user needs, preferences and often attitudes have to be known and visible to the system. From this point of view, the user's feedback plays crucial role. As the both – users and business as well gains benefits from such a tailored content or service (user access relevant information or products in shorter time, business lowers the adverts cost and raises profits), the "obsession" on collecting user feedback and using it to improve the web increases day by day.

In business sphere, the task of the prediction whether a customer will stay or exit using particular service (e.g., do not prolong the contract) is referred as the attrition or conversion rate. Such a prediction can be computed based on the customer behavior and the feedback he/she leaves during the contract (which is usually long term). However, in this paper, we focus on the task of the learning session end prediction for specific student in the e-learning web-based system (prediction of student exit intent in the session). This represents a novel application of standard long term attrition rate task to the short term behavior, which brings new challenges and also possibilities for user behavior prediction expressed by his/her next action(s).

E-learning systems are currently very popular and millions of students learn using those (Jegatha et al., 2014). Moreover, similarly to e-learning systems the MOOCs (Massive Open Online Course) are often used to promote the universities and allows them to sell certificates to graduates, which brings a huge business potential. E-learning system typically contains various courses containing learning materials divided into logical units called learning objects (LOs), e.g., explanations, questions or the practical exercises to solve presented in various forms combing text and multimedia. This rich information source can be used to improve the e-learning based on specific students' characteristics and behavior.

There are many advantages of e-learning in comparison to traditional education. One of the most important is, that "e-students" can adjust the learning process to their own needs and speed, which fit them the most. Jovanovic et al. proposed the clustering method for grouping students based on their cognitive learning style (Jovanovic et al., 2012). This way are users able to spend their time in e-learning effectively, because system is able to automatically adapt their learning materials with respect to their learning styles.

Another advantage of e-learning systems is the possibility to adapt the course structure, navigation or its content exactly to the needs of every student individually. The concept of the adaptation and personalization of web-based systems for the domain of e-learning was introduced by Brusilovsky (Brusilovsky, 1996) and is still intensively researched nowadays. There were proposed the methods of personalized recommendation, such as hybrid approach from Klasnja-Milevic et al., which is similarly to previous approach based on students' learning styles, but in this case also on frequent sequences of in content learned (Klasnja-Milicevic et al., 2011).

The task of the session end prediction represents an interesting challenge of e-learning. Students sometimes decide to stop learning while they did not understand fully the materials. If the e-learning application would be able to predict that student will probably leave soon, it could motivate him/her to stay longer, remind him/her the learning object he/she has not studied yet or offer him/her some questions to test his/her real knowledge. In this way the system will be able to help the student to learn effectively, e.g., not miss any of topics to learn in order to better prepare for his/her exam.

Our contributions presented in this papers are:

- An analysis of e-learning students' behavior and feedback types and sources.
- Novel approach for student exit intent prediction for actual session designed for highly dynamic data in the form of data stream.

In comparison to the state-of-the-art approaches and challenges in the attrition rate prediction including e-learning domain, our proposed approach focuses on short-term behavior prediction (in the mean of one session). Proposed approach fully takes an advantage of all available user characteristics- including students' performance, their personalities or learning styles. Thanks to the predictor architecture (polynomial binary classifier, using the stochastic gradient descent algorithm), we are able to process students' actions within the system as the data stream and dynamically make predictions for actual sessions. Such a short-term prediction is not used in today's web-based systems, including the e-learning domain.

The rest of the paper is structured as follows. The related work and the state-of-the-art is presented in the next section. The section *"E-Students" Challenges* describes the current trends of e-learning and its advantages in comparison to the traditional learning approaches focusing on the student feedback. We demonstrate the most important features and the ways of collecting the feedback from students' actions considering our e-learning system ALEF (Adaptive LEarning Framework). In the following section we focus on one task of user behavior prediction. We describe proposed method for prediction the next user action in web-based educational system in mean if he/she stay or will leave. The section *Evaluation* shows the results of the proposed method used with various settings. Finally, in the section *Conclusions* we summarize the achieved results and discuss future work.

2. Related Work

As the students' behavior and feedback is collected ex-post (after the action happened), the machine learning and data mining techniques have to be used in order to predict next students' actions. There exist two basic data mining tasks – descriptive and predictive (Kantardzic, 2011). Descriptive tasks are primarily used to discover structure, relations or patterns in mined data. There are used mainly in the unsupervised learning approaches (Grira et al., 2004). On the other hand, predictive approaches use mostly the supervised learning (Kotsiantis et al., 2007) and are used to estimate unknown values or predict future trends in data values.

The quality of user behavior prediction is highly dependent on the quality of user models describing user's previous behavior and preferences. Several artificial intelligence methodologies are used in the domain of web-based learning, e.g., automatized discovering of relations within the content, which are then used for user and domain modelling. A process of automatic relationship discovery in domain model considering learning objects as key elements was researched by (Simko and Bielikova, 2009), an adaptive question selection by (Barla et al., 2010). Other approaches used for improving the e-learning use the predictive data mining analysis (Peña-Ayala, 2014). In addition, the artificial intelligence methods are used for the increase of student experience, typically, for personalization of the e-learning applications. One representative – the adaptive navigation support approaches adaptively select hyperlinks available for individual students from the content of the e-learning application (Brusilovsky and Persin, 1998). Other approaches employ the guidance of relevant content for students by automatically generated ontology-based navigation (Holohan et al., 2005). Outputs of artificial intelligence approaches are in e-learning domain generally used for obtaining quality metadata - used for the description of user preferences or typical behavior and, in the next step, for the prediction of future user actions (Levy, 2007).

In this paper we aim to predict student behavior. We focus on the exit intent prediction within user's session. Our task can be formalized as task of predicting if "*Will the student go from current learning object to another one or will he/she leave the application?"* This task refers to the binary classification problem, which is generally suitable to be solved by supervised learning (Peña-Ayala, 2014). Similar tasks were in the past solved mostly by a Bayesian models (Li et al., 2011), decision trees (Long and Wu, 2012) or a neural networks (Yu et al., 2010). They were however applied mostly in a different scale and also a different context from ours. The problem of user attrition or conversion rate is typically researched in business domains as a retail banking (Li et al., 2011) or telecommunication (Wojewnik et al., 2011), where is the loss of a customer estimated in a long term.

When focusing on the e-learning domain, the task of attrition or conversion rate prediction was in recent years explored strictly on the high abstraction level in the mean of the long term scale. There exist works dealing with students' dropout from the e-learning courses (Tan and Shao, 2015; Halawa et al., 2014, Bayer et al., 2012), dropout from studies (Sangodiah and Balakrishnan, 2014), or freshmen students loss (Delen, 2010). These tasks typically use classification algorithms as the logistic regression (Kotsiantis, 2012; Bukralia, 2010), multilayer perceptron neural networks (Bukralia, 2010), support vector machines (SVM) (Sangodiah and Balakrishnan, 2014; Bukralia, 2010) or rule based prediction (Halawa et al., 2014).

Our work is focused on the classification of user behavior in shorter period of time, i.e., sessions. We aim at predicting the session end, which represent slightly different task. In the case of the students' dropout, there is, according to Halawa et al., possible to notice first signals (of students' dropout intentions) at least two weeks before dropout itself (Halawa et al., 2014). On the contrary, the short term periods as sessions, bring significantly less time to discover and to observe such signals. User short term behavior is also biased by user's current context, mood or intent, thus it is more difficult to predict user's future actions. To our best knowledge no work explored the session end (exit intent) prediction task, especially in the e-learning domain.

The application of users' attrition or a user exit intent prediction to domain of e-learning is a difficult task, because of multiple factors. There exist a lot of data about learning objects visits incoming from the e-learning system as a continuous stream, so standard batch approaches are not usable and it can be only handled as a continuous stream of data. Data also dynamically change in time and there is quite different behavior of students at the various course stages. The classifier has to be able to dynamically adapt to the actual trends (or several prediction models have to be trained for various course stages). Also the users behave very individually and there are no general patterns available. As users typically browse multiple learning objects per session before they actually leave, the classes (continue browse vs. leave the application) are highly unbalanced. Mentioned obstructions made from task of prediction of users' attrition, quite a challenge.

In our work, we deal with a big amount of data that come in the form of a data stream. This fact specifies our task to the single processing of the data, which is typical for working with data streams (PhridviRaj and GuruRao, 2014). Nowadays there can be observed the relatively huge increase of machine leaning methods working with the data streams (Bifet et al., 2010). As the user preferences and behavior change relatively quickly in time (course start, before exam, after exams), it is important also to be able to react to these changes as quick as possible, i.e. in on-line time (Yu et al., 2010).

A prediction if a user will leave a page or not, a customer will buy some product or change his/her bank next month is a typical representative of binary classification task. In such a task type, there typically occurs a problem of unbalanced classes, which leads to problematic model learning process (Sun et al., 2009). There are much more observations when a user continues browsing (because he/she typically visits multiple learning objects before the leave) than observations of leaving; e.g., more customers will keep their bank next month than change it. The most often used techniques to reduce the unbalanced classes problem are an oversampling of the rarer class, undersampling the majority class or assigning the different importance to observations (Bottou, 2012).

3. "E-Students" Challenges

One of the most dynamic transformations in recent years can be seen in the education domain. We can see that traditional form of face to face learning is being replaced by e-learning. This transformation creates a new kind of students, so-called "e-students". Not only the students enjoy e-learning education (coursera.org users count reaches 10M[2]) more and more, but the standard education institutions offers a great and still increasing number of courses online (nearly 2.5k courses available

[2] Source: coursera.org (statistics for the end of the 2014)

in coursera.org[2]). Adaptive and especially personalized e-learning systems bring to students new advantages and challenges respectively.

3.1. Mining "e-students"

In standard face to face education, the interaction between the teacher and a student creates a unique environment, which is essential for the learning process. In the opposite, in e-learning usually there is no direct interaction between students and teachers (except indirect forms such as discussion forums), often because of the number of students enrolled for the course. On the other hand, this shortcoming can be reduced by utilizing rich information, which the students leave in the electronic environment, such as accurate time spent on learning objects, a sequence in which student learned them, knowledge learned during studying (continuous test results) or the exact information about which parts of learning objects student read (eye tracking), how long he/she was thinking before answer, etc.

Students interacting with e-learning system provide huge amount of various indications or signals in the form of implicit or explicit feedback, and behavior which can be further analyzed. This we understand broadly as not only student ratings of specific content, but generally the students' knowledge level, their personal characteristics, communication with other students and much more. This rich source of students' feedback brings new challenges and possibilities for adaptation and improving of the learning process in comparison to standard education. By this way, not only the student experience, but the performance as well, can be improved (Horvath and Simko, 2013).

Moreover, the content of adaptive e-learning system itself can be used to provide new stimuli to the students (e.g., personalized suggestions of similar topics, courses). Explanations, interactive animations, videos and much more may improve students understanding user experience.

More and more attention is paid to the collaboration in the e-learning systems today. Various enhancements to utilize the group knowledge have been proposed, e.g., group recommendations. Similarly, new forms of students' feedback are researched as the camera or eye-tracking in order to find interesting parts of study materials or to prevent from cheating etc.

Not only students' implicit data are important from the logging and the feedback point of view. Although they represent the main source of feedback (describing actions of students within the system), other students' characteristics and information sources can be used. To speak more generally, e-learning students' feedback can be observed from several perspectives based on its relationship to the student behavior. These perspectives should be considered when the students' feedback is analyzed and used in further computation (e.g., personalization):

- *Web-based behavior* – obtained directly from the student's feedback, which covers his/her browsing patterns, time spent, session duration and other web-related indicators. It represents the main source of information when modelling students' behavior. As it represents the user's actual short-term preferences - it is generally very biased, influenced by user's actual intent, context and global trends.
- *Content characteristics* - include the learning object content itself, similarities between learning objects, hierarchies, prerequisite or tags. As the content does not represent a kind of feedback, it has an impact to student behavior in system. For this reason, it is an important characteristic, which should be known (at least on the learning objects level).
- *Personal characteristics* – similarly to the content characteristics it is not a feedback source strictly speaking. The description of student's personality traits, learning style etc. provide important information about individual personalities, which provide distinctive information from the adaptation point of view and indirectly influence the student behavior. It represents also the user long term preferences and thus the regular behavior and browsing patterns.
- *User experience behavior* - e.g., web page layout, ergonomics, eye-tracing etc. The user experience directly influences the behavior and feedback received from the students. If there

are any usability issues within the e-learning system, received feedback or behavior patterns can be skewed.

Generally speaking, more relevant data sources we have, the better we are able to describe student past and future behavior. The traditional source of system logs (student activities) are often enhanced in today's modern applications. New opportunities to the data collecting, which describe student behavior, characteristic and preferences are used. There is an intensive research in the eye tracking and related technologies nowadays. Eye tracking represents an alternative source of data to the traditional sources as the system logs, which offers another view on the user activity. Due to the hardware requirements, it is however, not currently widely employed in the standard day to day usage. So there is lack of studies comparing performance of machine learning approaches based also on the eye tracking vs. traditional approaches mostly based on system logs. We believe that a combination of both data sources can bring an improvement to understanding the user behavior.

In the context of prediction in e-learning domain, the eye tracking was successfully used for implicit user intent recognition (Jang et al., 2011), intent inferring (Salvucci, 1999), minimizing user attrition rate (Ribisl et al., 1996), improvement of navigation within e-learning systems (Goldberg et al., 2002), discovering reading patterns (Bielikova et al., 2015) or monitoring users' multitasking activity (Konopka and Navrat, 2015). These approaches extend the tracking of students' actions, e.g., recognizing students' face or gestures, which can indicate students' emotions or even mood (Katsmireou et al., 2014). Additional devices allow also to track student's biometric characteristics as a pulse, heart activity or body temperature. To capture these inputs, the specialized hardware is required, which is usually available in laboratory conditions only. As the history proved, the price of such devices will drop massively, which will result to wide availability. The usage of wearable computers (e.g., smart watch) seems to be a promising area in connection to the user's feedback collection. The reason is that implicit feedback acquired from various devices, which are able to capture user activity (from the biometric point of view) can be used to replace some of the user behavior description (acquired in the standard way). Such an enrichment or replacement is very beneficial in the short-term user behavior prediction, which is naturally very often noisy.

3.2. ALEF – Adaptive LEarning Framework

In order to better illustrate various sources of students' feedback which can be collected within the e-learning system we provide a brief description of ALEF (Simko et al., 2010; Bielikova et al., 2014) – the adaptive learning system based on concepts of Web 2.0, which is developed at our faculty. Since 2009 it offers multiple courses focused on programming (procedural, functional, logical) and software engineering education.

The system idea is based on the standard two core models – domain and user model and the set of framework components which provide adaptive functionality (Bielikova et al., 2014). This represent typical architecture of adaptive e-learning systems, introduced by Brusilovsky (Brusilovsky, 2004). The domain model is based on lightweight semantics in the form of domain relevant terms. The learning content consists of various types of learning objects and their relationships. Three types of learning objects are used:

- *Explanations* – textual studying materials, similar to book chapters in traditional studying materials. They are hierarchically ordered, based on the domain experts and the automated domain model creation as well (Simko and Bielikova, 2009).
- *Questions* – short test questions, where students can quickly test their knowledge. Questions can be answered without need of extensive solving and are evaluated automatically.
- *Exercises* – practical tasks, which have to be solved for a longer time than the questions usually (the implementation task is often used).

The user model is the next core component, which is based on the overlay user modelling. Several layers of user characteristics (or his/her history) are added on the top of the domain model, e.g., which learning object was visited, questions or exercises were solved or other student interactions were made. Another example of the layer refers to the student's knowledge or his/her characteristics mapped to the key domain concepts. Thanks to this, the knowledge is spread among domain concepts using concepts relationships when the student knowledge is updated.

Finally, the framework components provide active-learning, collaboration and adaptation support. The annotation framework allows students to provide, update and share various annotations for educational content, which enriches the learning content. Students can highlight, tag, comment or extend learning content by external sources and definitions. Annotations can be set by users as private (visible only for author) or public (visible for others – anonymously or with author's name). Widgets provide interactive gateway for the adaptation and collaboration features. Thanks to the modular design and implementation, widgets provide great reusable extendable platform for various specific functionalities (recommendations, student's score presentation etc.).

Student Behavior and Feedback

Collection of students' feedback is in the ALEF system realized by the feedback framework and their requests to system are captured by the logging framework. It means, that users' interaction with the system is well recorded. There are both types of feedback, implicit and explicit, collected. Implicit feedback is realized by capturing learning objects visits, the explicit one by users' actions within the learning objects (answering the questions, highlighting texts, commenting, annotating etc.).

Generally, the student behavior can be captured by extensive logging within the system. The explicit feedback however is not the only important source of information for modeling of student's behavior. Ideally, the actual state of the system (in the time of students' interaction) should be also easily able to reconstruct, including the content of the dynamic page components. Such a complex information can help to really understand students' past behavior and based on this knowledge to improve adaptation features of the system in the future. For example, if there was some recommendation presented to the student, which he/she ignored, it indicates that these recommendations are not relevant for particular student and his/her content.

Students' Personal Characteristics

Students' (or users') personal data are important and can be used for the improving the user experience and satisfaction. Within the e-learning system, the student's courses enrolled, study results reached or his/her projects made reveal his/her context and preferences. These data are relatively easy to capture, because they are logged implicitly without need of students' additional effort.

In addition, the data describing students' characteristics from the personality point of view are also important. Student personalities or learning styles can improve the adaptation of learning content by its tailoring the special need of specific student. Moreover, such information is crucial from the collaboration and group construction point of view. The basic assumption for usage such personal characteristics data, is that there is some distinctive value, which can be further used in the automatized processing.

To illustrate the distribution of personal characteristics of students in e-learning system ALEF, we present results (percentile) of NEO-FFI questionnaire for 160 bachelor students of Software engineering course. As we can see (Figure 1), there is huge variety in the students' personalities, while neuroticism and conscientiousness seems to be most diverse. On the opposite the openness and agreeableness are more compact and consistent. As expected the extraversion is moved slightly to lower values as the students of Informatics are generally considered as more introvert. On the other hand, in every dimension both low and very high scores were observed.

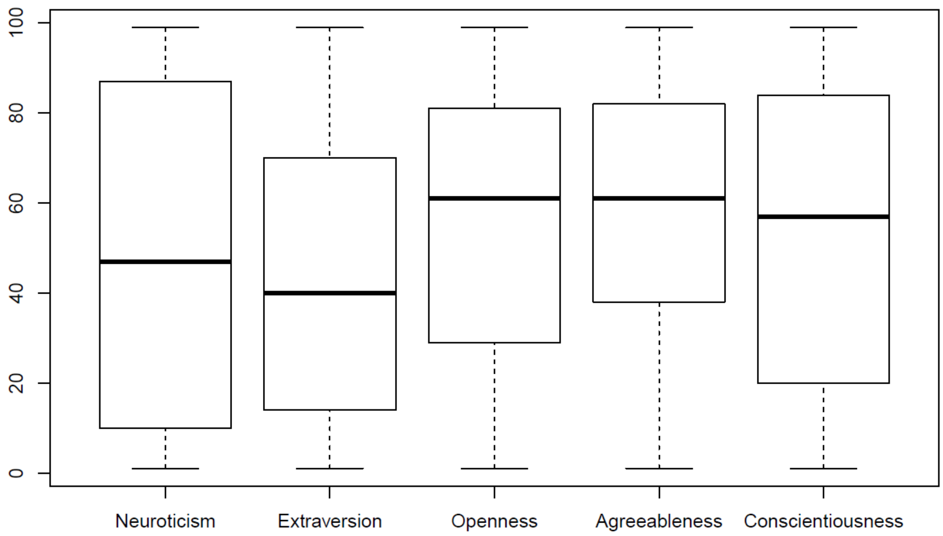

Figure 1. Big5 personality traits for the total of 160 bachelor students of Software engineering course. Results represents percentile for the Slovak population obtained based on the NEO-FFI personality test

Actually there can be observed a massive increase of e-learning systems, which brings to students new possibilities in comparison to the traditional approaches. Thanks to new technologies, there are wide possibilities to acquire students' explicit and implicit feedback. The web-based behavior represents the main source of data, but also the content and students' personal characteristics are able to offer helpful information for building reliable user models (with solid distinctive information). To illustrate the various feedback source and to evaluate proposed approaches we use the Adaptive Learning System ALEF, which represents the state-of-the-art web-based e-learning system. In addition, system contains for part of students also their personal characteristic, which can be used for better behavior prediction.

4. Prediction Approach

In order to be able to predict if a user will leave the application in the next action or he/she will continue the session, we have to deal with several limitations. At first, the data come in a continuous stream, which eliminates the usage of batch approaches. The data stream is represented by users' actions (visits of learning objects realized by users), where every action is described by a set of attributes describing currently visited learning object (LO), user behavior in actual session and also his/her typical behavior (described in section 5.1, Appendix A).

At second, the data are unbalanced, due to the fact that users in a session typically visit multiple LOs, while every session has logically only one leave action. The third limitation refers to the fact that data characteristics dynamically evolve over the time, due to the users' behavior changes (e.g., students behave differently at the beginning of the course or during the night before an exam). For this reason, we need to deal with varying data characteristics in the real time. This is supported by the fact that student behavior intensity is irregular and unexpected, which means that it can't be described by the regular sampling rate.

Based on these limitations, we proposed a polynomial binary classifier, using the stochastic gradient descent algorithm, which is able to process stream of data in real time and dynamically determine the importance of context attributes describing observations. To improve the classifier results we designed it to calculate weights individually for every user. We also devised new attributes describing the observations from the point where it is possible to better differentiate between the classes considered.

The input of the proposed classifier is a stream of user actions representing LO visits. The LO visits came without a regular sampling rate, because students visit educational system irregularly, mostly based on their course deadlines, exams or individual preferences. Every incoming action is at first described by attributes capturing student actual behavior in the session, his/her typical behavior and also by characteristics of visited LOs. These attributes are then used for the prediction if the student will leave the educational application in next *n* actions (or time range), or not. The attributes come from multiple sources. At first, some attributes are acquired directly from educational system logs (the user, LO, timestamp). At second, some attributes are derived or calculated (student behavior in the session, his/her typical behavior, LO characteristics). We provide more detailed description of the attributes in section *5.3 Attributes Importance*.

4.1. Stochastic Gradient Descent Approach

Because the data flow in a continuous stream, which varies over the time (as students behave differently in different time points of the course), we used the stochastic approach instead of the batch one. With this approach all observations are considered only once. This allows us also to handle the large data streams. This is although not absolutely necessary in traditional settings for e-learning domain, but as described above, our proposed classification approach principle is a domain independent, it allows to employ the method in domains with huge amount of observations flowing in. Moreover, considering data signals such as user's gaze or face, even in e-learning domain we are confronted with huge data streams. Our approach also helps to process data coming with irregular intensity and eliminate possible overload, which is a common example in the e-learning domain (e.g., several minutes before deadline, night before exam).

Both, stochastic and batch, approaches are based on a certain hypothesis (Equation 1), which can be in the case of a third degree polynomial classifier, which we use, described as:

$$Hyphotesis: h_{\Theta}(x) = \Theta^T(x, x^2, x^3) = \Theta_1 x_1 + \Theta_2 x_2 + \cdots + \Theta_{n+1} {x_1}^2 + \Theta_{n+2} {x_2}^2 + \cdots + \Theta_{2n+1} {x_1}^3 + \Theta_{2n+2} {x_2}^3 + \cdots + \Theta_{3n} {x_n}^3 \quad (1)$$

where *x* is a set of attributes x_i describing an observation, weights Θ represent a set of importance measures by which are considered observation attributes when observation is being classified. To optimize the weights Θ, is sequentially calculated by Equation 2.

$$\Theta_j := \Theta_j - \lambda \frac{\partial}{\partial \Theta_j} J(\Theta_0, \ldots, \Theta_{3n}) \quad (2)$$

After every iteration (all observation considered), weight Θ_j is reduced by derived cost function *J* multiplied by learning rate λ. The learning rate is used to affect how much attributes weights are changed after incorrect classification. Generally, it is a very small number close to 0 ensuring that weights are decreased by very low value at every time when class is classified incorrectly.

The difference between batch and stochastic approaches is in a way the cost function *J* is calculated. In batch approach (Equation 3) (Robbins and Siegmund, 1971) the cost function *J* is calculated and weights Θ are modified after training the classifier on all data observations (once per iteration).

$$J(\Theta) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_{\Theta}\left(x^{(i)}\right) - y^{(i)} \right)^2 \quad (3)$$

where *m* represents the number of observations considered, $h_{\Theta}\left(x^{(i)}\right)$ is the hypothesis for i^{th} observation $x^{(i)}$ and $y^{(i)}$ is the real observation class. In stochastic gradient descent approach that we used, every observation is considered only one time, so the big amount of data can be processed. In this case, the cost function *J* is calculated after every observation (Bottou and Bousquet, 2008).

This brings the advantage of much faster data processing and effective reaction to dynamically changing data. Single iteration data processing also eliminates the overfitting problem, which can occur in batch approach with too many training iterations. On the other side, stochastic approach requires to be trained on much more observations than the batch one to reach the same classification precision (2).

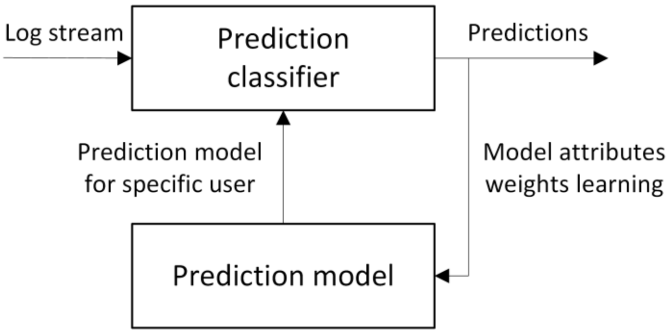

Figure 2. Principle of proposed classifier – log stream is classified using unique prediction model for every user, while model attributes weights are updated with every observation simultaneously.

4.2. Dealing with Unbalanced Data

The binary classification task is a specific type of classification problem, where assessed observation can be classified only into one of two considered classes. In case of classification the web session end, this approach however suffers to unbalanced classes problem. The session classification in e-learning domain is not the exception, because it is predictable that user browses multiple pages (e.g., learning objects) before he/she leaves the application, which bring us to ratio "*1: average session length*" in favor to a class containing observations when users continued to browse content in system. This disproportion causes, that if classifier will classify all observations as belonging to the more numerous class it reaches the very high accuracy despite the zero precision on rarer class.

To reduce this disproportion, there exist three standard techniques - the oversampling the rarer class, undersampling the wider one or assign the different importance to incorrect classification to observations of both classes (Bottou, 2012).

Variant with oversampled and slightly multiplied rarer class (visit of last learning object in a session) brings the more acceptable ratio from the view of balance between classes and concurrently it does not interfere the real situation too much. In case of rarer class observations an overfitting problem can occurs, as these observations have to been multiplied to up to level of the majority class observations occurrence.

Another possibility of dealing with unbalanced classes is to undersample the majority class, which would mean to throw away the huge amount of observations. If we want to use the undersampling technique and to balance the ratio between classes to be similar to above mentioned one, we have to throw away significant portion of all observations, what we consider inappropriate.

The third approach to solve the classes' imbalances problem - assign of the different importance to observations. Based on its idea, we proposed the adjustment of the Equation 2, by assigning weight coefficient w_c as can be seen in Equation 4.

$$\Theta_j \coloneqq \Theta_j - w_c * \lambda \frac{\partial}{\partial \Theta_j} J(\Theta_0, \dots, \Theta_{3n}) \qquad (4)$$

The sense of w_c is to reduce the classes' imbalances problem by changing measure by which the classifier's attributes weights Θ are adjusted after the incorrect observation classification. Its value is calculated by Equations 5 as number of all observations classified divided by number of observations from the class as the actually classified one has.

$$w_c = \frac{\#\ of\ all\ observations}{\#\ of\ actuall\ \ class\ observations} \tag{5}$$

This modification of standard stochastic gradient descent approach results in the approximately equal classification precision for both considered classes, because rarer class is considered with the higher importance in the process of classifier attributes weights learn.

Generally, we try to do not throw any data, as this can lead to incorrectly trained classifier model (by omitting some data and thus students' behavior patterns). The overfitting (introduced by oversampling of rarer class) however may not be problem in the case of stochastic approach. Thus we propose to use the technique of assigning different weights for a penalization of the wrong rarer class prediction, potentially in combination with the oversampling.

5. Evaluation of Proposed Classifier

To evaluate proposed classifier and to tune its features, we used several variants - global classifier (classifier using only the attributes weights trained for all observations) and also personalized "per user" (classifier using attributes weights trained for each user individually) and "per course" (classifier using attributes weights trained for each course separately). Next, we devised optimal settings as the value of learning rate λ, the degree of oversampling and the weights of observations belonging to both classes. We repeat the training process in multiple iterations, in order to compare classification results after different amount of observations processed and to analyze the overfitting of learned model. We also confront our results with similar classification tasks (conversion rate, etc.). As the first step we have to pre-process data coming from the e-learning system to be able to use them in process of evaluation of proposed classifier method.

5.1. Data Pre-processing and Feature Extraction

Evaluation phase was performed on real web-based logs from our educational system ALEF (for more details about system, see Section 3.2). For our task we used the information about students' visits of learning objects (LOs). We had available activity logs about learning object (LO) visits from 882 users within 3 years, which means almost 452,000 observations. These logs contain attributes describing information about the student and LO and also the interaction details such as:

- *Student information* – unique identifier, role
- *Learning object information* – unique identifier, course, type, difficulty, rating, parent, title
- *Interaction details* – user, LO, component, from which user came to LO - menu, some widget or from outside, begin of interaction timestamp and interaction duration

Logged attributes do not offer the sufficient distinctive features for classifier, based on which it would be able to decide between classes. From the classifier performance point of view, the selection of distinctive features plays crucial role (Nagy and Gaspar-Papanek, 2009). For this reason we proposed additional attributes (for the complete list see Appendix A):

- *Session describing attributes*
    - Order of visited LO in session, time spent in session before visit current LO, flag if is current LO first or last in chapter, etc.
    - Average session length in course (month, week, day, hour, month day, week day)
    - Actual difference from average session length in course (month, week, day, hour, month day, week day)
- *Advanced timestamp attributes*
    - Month day, month, week day, week, day, hour
    - Boolean time flags - is holiday? is winter semester? is summer semester? are winter exams? are summer exams?

- *Behavior describing attributes*
    - The number of LO visits (or seconds) the user spends in application before leaving – average for course, month, week, day, hour, month day, week day
    - The difference between number of LO visits (or seconds) the user spends in application before leaving and the average of all users (for described time periods)
    - Global average probability of leaving the application from current LO?
    - Flags – is actual session length (clicks, time) above the average in course (month, week, day, hour, month day, week day)?
    - User's average session length (clicks, time)
- *LO structure describing attributes*
    - LO course
    - Source of LO
    - Flag - is the LO the last one in chapter?

We used 12 directly logged attributes, 14 that originated from the transformation of polynomial attributes into the binominal and we also devised 62 derived attributes, which together result in 88 different attributes. As we used also squared and cubic powers for all attributes, we worked with 264 attributes in total.

All numeric attributes were normalized to interval <-1; 1>. Attributes based on average of other attributes (e.g., average session length, average time spent for LO visit) were calculated only from historical observations trained before. It means that quality of this kind of attributes grew in time.

Observations were classified based on the result of the hypothesis described in Equation 1. Observations with the positive hypothesis result were classified as the visits of last LOs in the session, otherwise as a non-last (a user continues to study the LOs in the e-learning application). The user's activity was split into sessions, based on the rule – two sessions are separated at least by 30 minutes' interval (Liu et al., 2010). This rule was devised especially for sessions in the e-learning domain, where the users could study some LO or a difficult exercise for a long time without leaving the application.

In the pre-processing we removed suspiciously long sessions. These are caused by various robots created due to crawl system educational content or to simulate students' activities. For this reason, we excluded top 10% of longest sessions.

5.2. Classifier Evaluation

At first we compared precision and accuracy of the classifier trained by various ways – *globally* and personalized *per user* and *per course*. The *global* variant means, that the one model for classification was trained for all users. In the opposite, in the personalized "*per user*" variant, one model for each user was trained respectively. Moreover, we also trained a "*per course*" classifier, which analogically means, that one model for every course was trained. As can be seen in Figure 3a, the classifier results increase in all cases logarithmically. After first iteration (452,000 observations considered) reaches the best results the *global* classifier variant (precision: *global* = 0.262, *per course* = 0.256, *per user* = 0.229; accuracy: *global* = 0.812, *per course* = 0.810, *per user* = 0.802). The reason is that this variant was trained on the highest number of observations. In our data there are 5 separate courses, which gives in average 90,400 observations *per course* and 882 users with 512 observations of LO visit in average *per user*. Used dataset contains 30,767 sessions, which gives in average 35 sessions *per user* and consequently the 35 observations of rarer class, which is quite a low amount. For this reason, we ran experiment on dataset in multiple iterations, as a compensation of an observations quantity. We realize the possibility of classifier overfitting, what we verified in later experiments.

From the next iterations point of view, presented in Figure 3a, it is clear, that personalized approaches reach the better results than the global variant. *Per course* variant outperformed the global one after 5th

iteration (2,260,000 observations considered, 452,000 in average per course). *Per user* variant overtook both others after 7th iteration (3,164,000 observations considered, 3,587 in average per user).

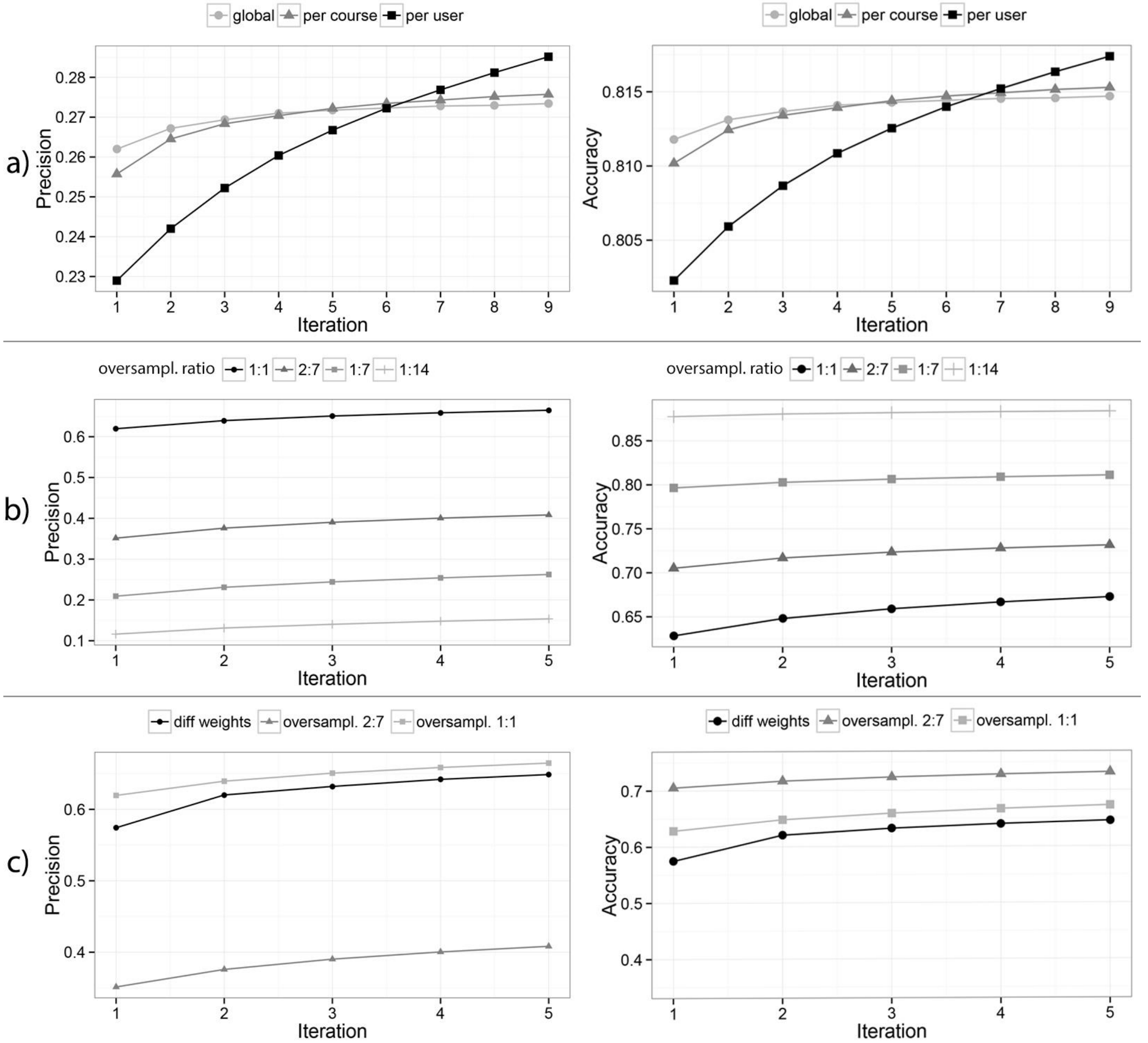

Figure 3. a) Precision and accuracy reached by various classifier variants; b) Per user classifier results for various oversampling degrees of rarer class (user ends the session); c) Per user classifier results for various oversampling degrees of rarer class (user ends the session) in comparison to assigning the different weights to both classes

Personalized variant reached better results as *global* variant, after some amount of iterations considered, because the data contains only few observations, so personalized classifier is not sufficiently trained before. According to collected data, the students of the ALEF system behave very heterogeneously, so it is very difficult for classifier to learn some general rules. This problem occurs mainly for *global* classifier variant, but it persists also in case of *per course* classifier, because there is still a lot of users (typically few hundreds) with different behavior in the course. Our experiment shows (Figure 3a) that the same users behave more similarly across multiple courses than various users in the same course. Based on the results reached in the first experiment we decided to use the classifier with model trained *per user* in the next experiments.

In the second experiment we focused to optimization of the learning rate λ. This parameter is important in a phase of modifying attributes weights as was shown in Equation 2. Results clearly show (Figure 4a) that all tested variants achieve a logarithmical increase of precision. There is also a dominance of variants considering number of rows trained before.

As shown in the Figure 4a, the classification precision increase variously, based on used learning rate. The rate $\lambda = \frac{1}{|rows|+10^{13}}$ which reaches the highest precision after second iteration, was in first two iterations outperformed by 4 another learning rates. The different learning speed based on various λ, is caused by a fact that too small rates slow down the learning process and are unable to dynamically react to changes in data, while too big learning rates do not offer enough sensitivity to fit the classification to data.

Similarly, we compared solutions to classes' imbalances problem. At first we used the technique of oversampling the rarer class (last LO in session). The experiment has been performed with oversampling to between class ratio 1:14 (real ratio between classes), 1:7, 2:7 and 1:1. Results in the Figure 3b show that there is no clear winner, because accuracy and precision create an inverse proportion. When the rarer class is oversampled to 1:7 ratio, there is a relatively high accuracy (0.796) and also the precision of rarer class is more than doubled (0.209 after first iteration, 0.262 after 5 iterations) in comparison to no oversampling. However, it is the most appropriate to use the variant with equal ratio between classes, because in this case the method obtains for both classes similar results (0.628 accuracy, 0.619 precision after first iteration).

In case of oversampling variants, there can easy occur a problem of overfitting of the rarer class (the 1:14 between classes ratio means the rarer class 14 times multiplication, which results to classifier overfitting). In case of variant used in previous experiments, there was rarer class duplicated (2:7 ratio), so the observations occurrence is close to the real situation, but here we can see a poor result of rarer class classification. For this reason, we experimented also the assigning the different weights to observations belonging to individual classes to reduce the classes' imbalances problem (Figure 3c).

When predicting user's future behavior, it is not always important only to identify the user's last action in the session. For this task it is sufficient to find out that the user will probably leave the application in short time or after the next few page views. Such an information can be even more useful than identification of the last action, because it provides more time to make an offer to the student what will make him/her retain in the application.

For this reason, we experimented with various settings of distributing the observations (students' LO visits) into classes in different way. These settings determined which observations are considered as the near to session end. At first, we experimented with zero (as in previous experiments) up to three LOs considered as an end of session. After that we also distributed observations into classes according to time remaining to the session end. In this case we consider as positive class observations the LOs visited under 5, 10, 15 and 30 seconds before the session end.

The classifier reached in all cases results similar to the classifier variant with different weights considered (described in Figure 3c). In this experiment, we however considered sessions as a units and we observed the percentage of successfully identified session ends (Figure 4b, c).

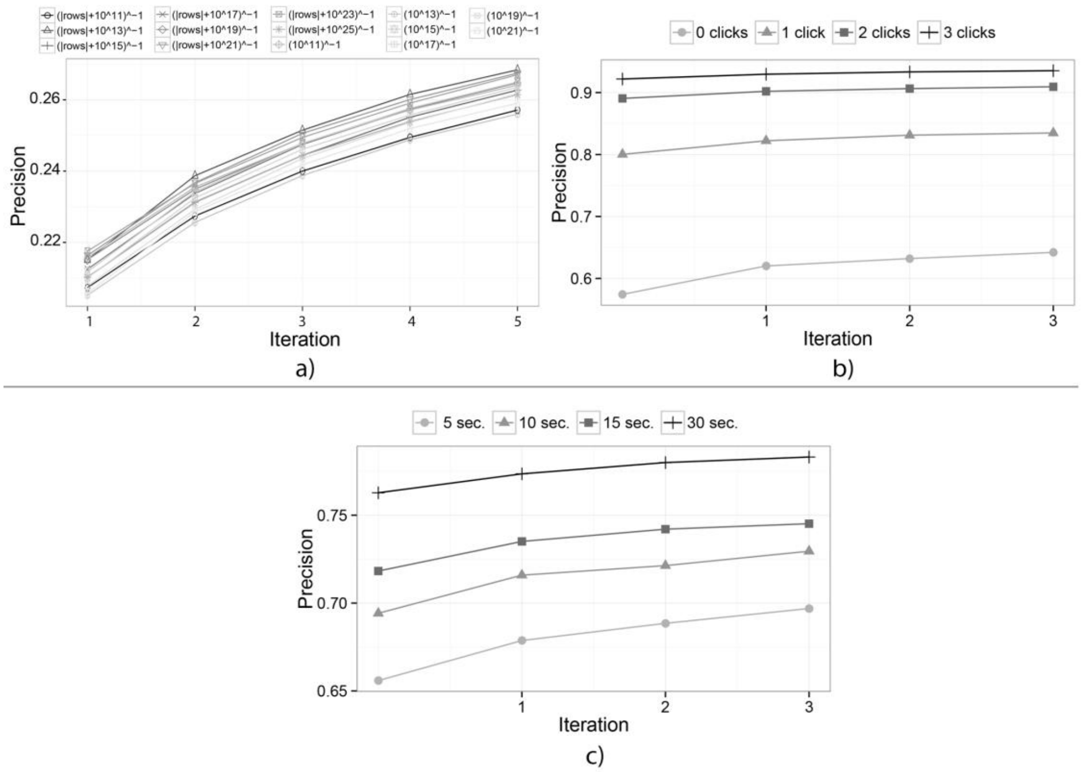

Figure 4. a) Precision of per user classifier reached with various learning rate measure; b), c) Precision of identification the session end. Successful prediction is represented by identification of at least one of observations accomplishing set conditions b) clicks considered, c) time considered.

As we can see (Figure 4b, c), we are able to classify session end for relatively extensive percentage of sessions. The task of identification of session is going to end, i.e. classify at least one of the last *n* LO visits, is very helpful for real word applications.

In previous experiments we continuously trained the classifier on the stream of all available data. Logically, this arises the question, whether the improving classification results (Figure 3c) are not caused in addition to classifier learning by its overfitting. For this reason, we divided sessions of each user into the train and test sets (9:1 ratio). Then we trained the classifier on the only train set in multiple iterations and after each of them we evaluated it only on the observations belonging to the test set. The classification results on the test set reached the initial results very similar to the ones we presented in the Figure 4b, c. In next iterations, however, the increase of precision and also accuracy is still visible, but it is much slower. The result is that the stochastic classifier trained individually per user needs high amount of observations to be trained optimally.

Similar works focusing on exit intent or users' attrition look at the problem typically in a larger scale. In other works, observations describe the user behavior for a longer time period, so they are less variation prone. Also user's decision about e.g., cancelling his/her bank account, buying the insurance or mobile operator program is a long-term decision, while our aim of leaving the session classification is mostly based on the user's current context, which is more difficult to estimate.

For classification of bank customers who are in risk of attrition, Li et al. reached an accuracy from 0.618 (Naïve Bayes), 0.793 (Logistic regression) to 0.847 (if-then rules with RIBBER mechanism) and precision from 0.572 (Naïve Bayes), 0.790 (Logistic regression) to 0.883 (if-then rules with RIBBER mechanism) (Li et al., 2011). In case of classification the student's learning course dropout, Bayer et al. reached with the most successful classier the accuracy 0.688 and precision 0.705 (Bayer et al., 2012). Their solution uses the PART rule classifier based on statistical rules.

From these results is visible that the classification on long term data reach the better results. For this reason we evaluated our classifier also on bank telemarketing dataset (Moro et al., 2014). In this case our approach (*global* variant) reached accuracy 0.876 and precision 0.485 after first iteration through dataset (45,212 observations, 1:9 ratio). Moro et al. reached on the same dataset accuracy 0.810 and precision 0.400 (Moro et al., 2014). Our solution is thus comparable with current state-of-the-art and the computation cost is lower. The lower results in e-learning domain are caused by a difficulty of a task.

5.3. Attributes Importance

It is clear that the attributes were in the classification process considered with different weights (as the result of training phase). In other words, some of the attributes contributed to the results more than other. The top attributes importance for individual users slightly vary from user to user, but there are some attributes, which weights belong consistently to the top important (Appendix AA). These attributes are consistent with the top attributes of the global classifier variant such as (in descending order):

1. Global average probability of leaving the application after visiting the current LO
2. LO visit order in the current session
3. User’s average session length (LO visits) normalized by a static constant
4. The difference of the session length (LO visits) between the current course and average of all user’s courses normalized by a static constant
5. The difference of user’s average session length (LO visits) and the average of all users’ sessions normalized by a static constant

As we can see, the most important attributes (2^{nd} up to 5^{th}) are derived from the number of LOs visited in the current session. Surprisingly, when comparing to the attributes connected to the time aspects of students’ behavior, these proven to be less important as the simple count of LOs visited. Based on this, we conclude that at least in the domain of e-learning, it is more important for students to cover specific topics of materials in comparison to time they spend on it.

The second result is that the important attributes have been normalized by the same normalization method. All attributes were normalized both by division by the maximal value for that attribute measured out up to current observation (maximum was increased progressively) and also by large static constant (estimated once by expert before the training process). Results showed that the attributes normalized by a static constant reached better results. The reason is that the normalization by a dynamically increased maximum can create from the equal attribute values different results in various phases of the classifier training. Other well-known approaches can be used for the normalization of attributes covering the dynamic character of data ranges (e.g., Z-score).

5.4. Prediction Quality Indicators

In previous experiments we found out that users behave heterogeneously as we mentioned before and thus the performance of proposed classifier varies for various users. For this reason, we wanted to know, for which of them we are able to predict the most precise or accurate. At first, we found out there is a correlation between count of user’s activity observations (882 users) and the precision (0.10419, p = 0.00097), respectively the accuracy (-0.19284, p = 0.00001), according to the one tailed t-test on Pearson correlation measure. From these results it is clear that we are able to better classify session ends of users with higher amount of observations, but on the other side the overall classifier accuracy is higher for users with lower number of observations logged.

For 160 users in our dataset, we had also additional information, which were not used in classification process - their study records (midterm exam, activity, final exam, estimation of final exam score based on semester score and also final grade) and the NEO-FFI questionnaire results for Big5 personality traits (neuroticism, extraversion, openness to experience, agreeableness and conscientiousness) (Costa and McCrae, 1989). The NEO-FFI is based on the NEO-PI personality inventory and consist of 60 questions (12 per dimension). Despite the revised inventory have been published (McCrae and Costa, 2004), the basic NEO-FFI is still massively used and considered as reasonable and reliable for most cultures (McCrae and Costa, 2004; Robins et al., 2001). For these users we observed correlation between classification results and these characteristics.

As we found out, the most correlating attributes are classification prediction and final course results as the final course grade, the score reached on final exam and the final course score. We found the statistically significant correlation between the precision and the final exam score (0.1445, $p = 0.03415$) respectively between the precision and the final course grade (0.1283, $p = 0.05434$). In both cases, there was a positive correlation, which means that we were able to classify better the actions of users who reached the better final results in their courses.

In case of NEO-FFI characteristics we found the significant correlation between the conscientiousness and the accuracy (-0.2081, $p = 0.00414$) respective the precision (-0.1853, $p = 0.00949$). This personality trait describes a user's tendency to show self-discipline, act dutifully, and aim for achievement against measures or outside expectations. In this case we can see the negative correlation, which means that we can better classify actions of users who are spontaneous and use instincts in comparison to users who control their behavior. Except the conscientiousness characteristic, we found out also the correlation between precision and extraversion trait (0.1264, $p = 0.05561$).

In addition to NEO-FFI, we experimented also with Felder and Silverman Learning Styles (Felder and Silverman, 1988). In the ALEF system three year history considered for this evaluation, there are learning styles for 233 students available, for which we compared variance between individual model dimensions according to precision and accuracy of the proposed prediction method. Based on one-way analysis of variance (ANOVA), we however found out that the prediction precision (accuracy) results have no statistically significant differences within the individual model dimensions. Based on this result we cannot prove the usability of Felder and Silverman Learning Styles in the task or user session end prediction (when the model is trained per user). Similarly, to the per user variant, the global variant does not seem to perform better, when learning styles are included. However, more evidence is needed to explore learning style usage in various educational systems. In addition to comparison of model dimensions' variance (to the prediction results), we compared also prediction precision with or without learning styles considered. Our experiments however did not bring significant improvements with learning styles included (1st iteration: prec = 0.6018 with LS, prec = 0.598 without LS; 5th iteration: prec = 0.625 with LS, prec = 0.624 without LS).

Results obtained for NEO-FFI characteristics and user activity consideration however indicate the possible improvement of prediction accuracy in comparison to usage of standard user model based on only on the user behavior and content characteristics. Moreover, as the students' characteristics and their study goals vary, we discovered that we are able predict behavior of students receiving higher final grade. This can be explained by the fact that it is hard to predict behavior of randomly browsing students which don't have clearly defined goal when learning. We believe that the usage of more students' characteristics (related to students' personalities), will even increase the precision of presented prediction task.

6. Conclusions

E-learning systems nowadays reach still greater popularity, due to the benefits they bring to the learners. The e-learning became, these days, essential part of the traditional learning, by enhancing it by the variety of offered courses to large amount of students all over the world, personalization of learning materials to individual student's needs, means for collaboration during the learning process and variety of learning content (learning materials, exercises, tests, interactive videos, discuss forums, etc.).

There are several challenges connected to the e-learning and students' feedback mining in particular. The students' behavior prediction is one of them. In this paper we focused on the prediction of the student's session end (generally known as attrition or conversion rate task). Prediction of users' attrition rate within the session can improve quality of user interaction with an application. If we were able to classify that user will probably leave the application very soon, we can offer to him/her some reasons why to stay longer, for example by recommending him/her the interesting content. In e-learning it can be materials he/she did not studied before or test questions to evaluate his/her knowledge, in e-shop it can be a discount coupon or some interesting goods to buy.

From the educational process point of view, the proposed approach can be used to improve not only the student experience (in the system interaction context), but also to the increase of students' performance (from the knowledge point of view). The information indicating students' exit intent – allows us to select, i.e., recommend a content, which will maximize the student's knowledge. According to the actual course roadmap, recommending exercises or content, which provides key concepts, helps the students to learn important concepts more effectively. Similarly, we can motivate the students by stressing various mechanisms often including in e-learning systems (e.g., student's score, badges, discussion). All of these actions should, however, aim to increase the students' knowledge and to increase learning experience.

We explored a classification of dynamically changing data stream of observation about students' interaction with learning objects in various learning courses. We observed that there is a huge heterogeneity over users' behavior during the term. We dealt with unbalanced classification classes also. To be able to classify data with these obstructions we proposed personalized polynomial classifier using attributes' weights calculated for each user individually. The weights are dynamically calculated using stochastic gradient descent approach, which lowers computational cost to standard batch approaches.

We found out that the classifier variant with attributes weights learned for every user reaches after training on sufficient amount of data better results than the variant which learned attributes weights for all users globally. We also experimented with the data oversampling and the attributes types' augmentation which together bring a significant improvement of the classification performance. Due to the chance of overfitting the oversampled class, we assigned the different importance to observations belonging to the individual classified classes. In this way we obtained more robust solution not suffering to overfitting problem. Moreover, we applied proposed method to data from a domain of bank telemarketing. Similarly, to the e-learning domain, we reached better results than other approaches used in both these domains.

For the classification purposes we proposed various attributes describing the students, learning objects, students' visits of learning objects and also individual sessions. Results clearly show that the highest importance in the classification process reached the attributes describing the session (not the learning object content or interaction context). As the most important, showed the attribute carrying the information about global probability that students leave the application on current learning object. The second up to fifth most important attributes however described the session properties. All of them consider the session from the some view of the amount of learning objects visited in session (and not for example of time spent there, which was also considered in another attributes). From this, we can conclude that users in e-learning system (students) care more if they studied sufficient amount of materials, than the overall time spend in session or other characteristic.

We observed also the classifier performance variety over users in the experiments. The classification precision and accuracy correlates with number of observations per user available, which is obvious. But there exists also a correlation between precision and user's course results and also his/her personal characteristics (NEO-FFI personality inventory).

In addition to prediction of last learning object visit in the session, we extend this task to the identification that the session will probably end very soon (in next few learning object visits or the in specified time interval). It gives us more time to influence the student to stay longer in the application, in comparison to the case when the only last action is identified. We are able to predict the session end for more than 92.19% of sessions (at least one from last 4 learning object visits in the session) after 1st data iteration and 93.53% after 5th iteration. Next, we are able to predict session end for 76.27% of sessions after 1st classifier iteration if time was considered (at least one learning object from last 30 seconds of session) and 78.31% after 5th iteration.

We evaluated the proposed approach on the data obtained from standard e-learning system. As our aim is to predict user behavior – from the attrition rate point of view, this is usually not dependent on the specific system structure or content. Thanks to this, our approach is applicable to various, not only e-learning systems. In other words, as presented in Table A2 and A3, all attributes (raw or derived) are usually logged and stored by any e-learning system (including attributes describing a content). Moreover, thanks to the proposed approach variability – more attributes (when available) can be included, while the learning algorithm evaluates its importance.

We have shown that students' feedback and behavior can be in connection to the machine learning and data mining techniques used to improve user experience. Moreover, the prediction of users' behavior is valuable source of information for business as well. In the next years, thanks to popularity of wearable computers or eye tracking technologies, we can expect more and precise sources of users' feedback.

Proposed approach is based on students' actions described (based on the stream data) from various views. We covered information about learning objects, their structure, user typical long-term and short-term behavior in the actual session. Students' actions are however considered separately, independently from previous actions. It is clear that based on students' similarities in standard behavior, some patterns can be discovered and identified. When considering low level actions, patterns in browsing, e.g., tab open, close or switch can be captures. On the contrary, when considering high level actions, e.g., the difficulty or sequences in various difficult LOs' will hopefully bring additional distinctive information for the classifier. Such an information, will be extremely valuable when facing the cold start problem – not enough information about user or student available.

To speak more general, such patterns can be discovered in various domains, and be helpful for the classification task. On the other hand, the source of such pattern have to be chosen for specific domain, e.g., learning object difficulty for e-learning, topic of article content for news or price level for e-shop.

Based on these assumptions, we plan to consider also sequences of previous actions made by student in actual session. This may help to discover typical behavioral patterns and to explore latent dependencies in student's behavior (using Deep belief or Recurrent neural networks) before the session end.

Acknowledgement

This work was partially supported by grants No. VG 1/0774/16, VG 1/0646/15 and KEGA 009STU-4/2014 and it is partial result of the Research and Development Operational Programme for the project International centre of excellence for research of intelligent and secure information-communication technologies and systems, ITMS 26240120039, co-funded by the European Regional Development Fund.

References

1. Barla, M., Bielikova, M., Bou Ezzeddinne, A., Kramár, T., Simko, M., Vozár, O., 2010, On the impact of adaptive test question selection for learning efficiency, in: Computers & Education, 55(2), pp. 846–857.
2. Bayer, J., Bydzovska, H., Geryk, J., Obsívač, T., Popelinsky, L., 2012, Predicting drop-out from social behaviour of students, in: Proc. of the 5th Int. Con. on Educational Data Mining - EDM 2012, 7, pp. 103 - 109.
3. Bielikova, M., Moro, R., Simko, J., Tvarozek, J., 2015, Adaptive Web-Based Textbook Utilizing Gaze Data, in: Journal of Eye Movement Research, 8(4) (Special Issue ECEM 2015). pp. 252.
4. Bielikova, M., Simko, M., Barla, M., Tvarozek, J., Labaj, M., Moro, R., Srba, I., Sevcech, J., 2014, ALEF: from Application to Platform for Adaptive Collaborative Learning, in: Recommender Systems for Technology Enhanced Learning. Manouselis, N. et al. (Eds.), Springer Science+Business Media 2014, New York, pp. 195-225.
5. Bifet, A., Holmes, G., Kirkby, R., Pfahringer, B., 2010, Moa: Massive online analysis, in: The J. of Machine Learn. Res., vol. 11, pp. 1601-1604.
6. Bottou L., 2012, Stochastic gradient descent tricks, Neural Networks: Tricks of the Trade, in: LNCS, vol. 7700, pp. 421-436, Springer.
7. Bottou, L., Bousquet, O., 2008, The Tradeoffs of Large Scale Learning, in: Advances in Neural Inform. Processing Systems, 20, pp. 161–168.
8. Brusilovsky, P., 1996, Methods and techniques of adaptive hypermedia, in: User modeling and user-adapted interaction, Kluwer academic publishers, 6 (2-3), pp. 87-129.
9. Brusilovsky, P., Pesin, L., 1998, Adaptive Navigation Support in Educational Hypermedia: An Evaluation of the ISIS-Tutor, in: Journal of Computing and Information Technology 6, pp. 27—38.
10. Brusilovsky, P., 2004, KnowledgeTree: a distributed architecture for adaptive e-learning, in: Proc of the 13th Int World Wide Web Conf, WWW’04. ACM Press, NY, pp. 104–113.
11. Bukralia, R., 2010, Predicting Dropout in Online Courses: Comparison of Classification Techniques, in: Proceedings of the 5th Midwest Association for Information Systems Conference (MWAIS'10), paper 19, pp. 1-6.
12. Costa Jr., P.T., McCrae, R.R., 1989, The NEO-PI/NEO-FFI manual supplement, in: Psychological Assessment Resources, Odessa, FL.
13. Delen, D., 2010, A comparative analysis of machine learning techniques for student retention management, in: Decision Support Systems 49(4), pp. 498-506, Elsevier.
14. Felder, R. M., Silverman, L. K., 1988, Learning and teaching styles in engineering education, in: Engineering education 78(7), pp. 674-681.
15. Goldberg, J. H., Stimson, M. J., Lewenstein, M., Scott, N., Wichansky, A. M., 2002, Eye tracking in web search tasks: design implications, in: Proceedings of the 2002 symposium on Eye tracking research & applications (ETRA '02), pp. 51-58, ACM.
16. Grira, N., Crucianu, M., Boujemaa, N., 2004, Unsupervised and semi-supervised clustering: a brief survey, in: A rev. of mach. learn. tech. for proc. mult. con., Report of the MUSCLE EU Network of Excellence (FP6).
17. Halawa, S., Greene, D., Mitchell, J., 2014, Dropout prediction in MOOCs using learner activity features, in Experiences and best practices in and around MOOCs (37), pp.7-16.
18. Holohan, E., Melia, M., McMullen, D., Pahl, C., 2005, Adaptive E-learning content generation based on semantic web technology. in: International Workshop on Applications of Semantic Web Technologies for E-Learning (SW-EL 2005) – at the 12th International Conference on Artificial Intelligence in Education AIED 2005, pp 1-8.
19. Horvath, R., Simko, M., 2013. Enriching the Web for Vocabulary Learning, in: Scaling up Learning for Sustained Impact, Lecture Notes in Computer Science 8095, pp. 609-610, Springer Berlin Heidelberg.

20. Jang, Y., Lee, S., Mallipeddi, R., Kwak, H.,Lee, M., 2011, Recognition of Human's Implicit Intention Based on an Eyeball Movement Pattern Analysis, in: Neural Information Processing, Lecture Notes in Computer Science 7062, pp. 138-145, Springer Berlin Heidelberg.
21. Jegatha Deborah, L., Baskaran, R., Kannan, A., 2014, Learning styles assessment and theoretical origin in an E-learning scenario: a survey, in: AI Review, vol. 42, 4, pp. 801-819.
22. Jovanovic, M., Vukicevic, M., Milovanovic, M., Minovic, M., 2012, Using data mining on student behavior and cognitive style data for improving e-learning systems: a case study, in: International Journal of Computational Intelligence Systems, 5(3), pp. 597-610.
23. Kantardzic, M., 2011, Data mining: concepts, models, methods, and algorithms (2nd ed.), New Jersey: IEEE Press.
24. Katsimerou, C., Redi, J. A., Heynderickx, I., 2014, A computational model for mood recognition, in: User Modelling, Adaptation and Personalisation, LNCS 8538, pp. 122-133, Springer.
25. Klasnja-Milicevic, A., Vesin, B., Ivanovic, M., Budimac, Z., 2011, E-Learning personalization based on hybrid recommendation strategy and learning style identification, in: Computers & Education, 56(3), pp. 885-899.
26. Konopka, M., Navrat, P., 2015 Untangling Development Tasks with Software Developer's Activity, in: Context for Software Development (CSD 2015), IEEE/ACM 2nd International Workshop, pp.13-14, IEEE/ACM.
27. Kotsiantis, S. B., 2012, Use of machine learning techniques for educational proposes: a decision support system for forecasting students' grades, in: Artificial Intelligence Review 37(4), pp. 331-344, Springer.
28. Kotsiantis, S.B., Zaharakis, I., Pintelas, P., 2007, Supervised machine learning: A review of classification techniques, pp. 3-24.
29. Levy, Y., 2007, Comparing dropouts and persistence in e-learning courses, in: Computers & Education,48 (2), pp. 185–204.
30. Li, F., Lei, J., Tian, Y., Punyapatthanakul, S., Wang, Y. J., 2011, Model selection strategy for customer attrition risk prediction in retail banking, in: Proc. of the 9th Australasian Data Mining Conf., vol. 121, pp. 119-124, Australian Comp. Soc., Darlinghurst, Australia.
31. Liu, C., White, R.W., Dumais S., 2010, Understanding web browsing behaviors through Weibull analysis of dwell time, in: Proc. of the 33rd int. ACM SIGIR conf. on Res. and dev. in inf. Ret., pp. 379-386, ACM, USA.
32. Long, X., Wu, Y., 2012, Application of Decision Tree in Student Achievement Evaluation, in: Int. Conf. on Comuter Science and Electronics Eng., vol.2, pp. 243-247.
33. McCrae, R.R., Costa Jr., P.T., 2004, A contemplated revision of the NEO Five-Factor Inventory, Personality and Individual Differences, Volume 36, Issue 3, February 2004, pp. 587-596, ISSN 0191-8869.
34. Moro, S., Cortez, P., Rita, P., 2014, A Data-Driven Approach to Predict the Success of Bank Telemarketing, in: Decision Support Systems, 2014, vol. 62 pp.22-31, Elsevier.
35. Nagy, I.K., Gaspar-Papanek, C., 2009, User Behaviour Analysis Based on Time Spent on Web Pages, in: Web Mining App. in E-com. and E-services, Studies in Comp. Intell., vol. 172, pp. 117-136, Springer.
36. Pena-Ayala, A., 2014, Review: Educational DM: A survey and a DM-based analysis of recent works, in: Expert Syst. Appl. 41, pp. 1432-1462.
37. PhridviRaj, M.S.B., GuruRao, C.V., 2014, Data mining–past, present and future–a typical survey on data streams, in: Procedia Technology, vol. 12, pp. 255-263.
38. Ribisl, K. M., Walton, M. A., Mowbray, C. T., Luke, D. A., Davidson II, W. S., Bootsmiller, B. J., 1996, Minimizing participant attrition in panel studies through the use of effective retention and tracking strategies: Review and recommendations, in: Evaluation and Program Planning 19(1), pp 1-25, Elsevier.

39. Robbins, H., Siegmund, D.O., 1971, A convergence theorem for non negative almost supermartingales and some applications, in: Optimizing Methods in Statistics.
40. Robins, R.W., Fraley, R.C., Roberts, B.W., Trzesniewski, K.H., 2001, A longitudinal study of personality change in young adulthood, in: Journal of Personality, 69, pp. 617–640.
41. Salvucci, D. D., 1999, Inferring intent in eye-based interfaces: tracing eye movements with process models, in: Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '99). Pp. 254-261, ACM.
42. Sangodiah, A., Balamuralithara, B., 2014, Holistic Prediction of Student Attrition in Higher Learning Institutions in Malaysia Using Support Vector Machine Model, in: International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), 1(1), pp. 29-35, ARC.
43. Sun, Y., Wong, A.K., Kamel, M.S., 2009, Classification of imbalanced data: A review, in: Int. J. of Patt. Rec. and Art. Intell. 23(4), pp. 687-719.
44. Simko, M., Bielikova, M., 2009, Automatic Concept Relationships Discovery for an Adaptive E-Course, in: Barnes T. et al (eds) Proc. of Educational Data Mining 2009: 2nd Int. Conf. on Educational Data Mining, EDM 2009, Cordoba, Spain, pp. 171–179.
45. Simko, M., Barla, M., Bielikova, M., 2010, ALEF: A framework for adaptive web-based learning 2.0, in: Key Competencies in the Knowledge Society, pp. 367-378, Springer Berlin.
46. Tan, M., Shao, P., 2015, Prediction of Student Dropout in E-Learning Program Through the Use of Machine Learning Method, in: International Journal of Emerging Technologies in Learning (iJET), 10(1), pp. 11-17.
47. Wojewnik, P., Kaminski, B., Zawisza, M., Antosiewicz, M., 2011, Social-Network Influence on Telecommunication Customer Attrition, in: Agent and Mul. Agent Sys. Tech, and App., vol. 6682, pp. 64-73, Springer.
48. Yu, C.H., DiGangi, S., Jannasch-Pennell, A., Kaprolet, C., 2010, A data mining approach for identifying predictors of student retention from sophomore to junior year, in: Journal of Data Science, 8, pp. 307-325.

Appendix A

In Table A1 we present list of weights of attributes used in the classification process. Left part refers to the weights of the classifier variant with global attributes' weights; right part: weights of classifier variant with per user attributes' weights (random user). Attributes are marked as w01-w88, detail description can be found in Table A2. Every attribute has listed weights for basic, square and cubic powers. Attributes order is set as sum of absolute values of all three powers. In the Table A3 can be found attributes, which are directly logged by the system ALEF. They are used as the basis for the classification attributes described in Table A2.

Table A1. Example of trained attributes weights. The description of specific weight is presented in Table A2.

Weights trained globally				Weights trained for specific user			
attribute	basic	square	cubic	Attribute	basic	square	cubic
w34	1,65E-17	-2,08E-17	9,90E-18	w71	-2,90E-11	-2,90E-11	-2,90E-11
w48	-1,28E-17	1,23E-17	1,83E-18	w34	5,60E-11	1,63E-11	6,82E-12
w06	7,87E-18	1,88E-18	3,90E-18	w21	3,93E-12	9,98E-12	1,08E-11
w75	1,07E-17	-1,90E-18	5,90E-19	w72	7,47E-12	7,47E-12	7,47E-12
w80	-9,43E-18	-1,86E-18	1,77E-18	w13	-5,96E-12	-7,84E-12	-6,43E-12
w51	-3,36E-18	6,75E-18	-2,16E-18	w73	1,37E-11	3,20E-12	3,20E-12
w74	3,15E-18	-4,72E-18	3,70E-18	w19	5,45E-12	6,84E-12	7,36E-12
w76	3,57E-18	-5,17E-18	2,28E-18	w76	6,23E-12	6,03E-12	5,21E-12
w65	1,17E-18	5,62E-18	3,42E-18	w79	-8,80E-12	-1,67E-12	-6,09E-12

w83	-2,10E-18	-1,96E-18	4,97E-18	w48	3,95E-12	4,94E-12	4,84E-12
w71	-2,90E-18	-2,90E-18	-2,90E-18	w70	4,19E-12	4,19E-12	4,19E-12
w64	-1,97E-18	2,11E-18	3,98E-18	w05	-4,60E-12	3,25E-12	-2,59E-12
w01	-5,76E-18	-5,76E-19	-5,76E-20	w23	3,35E-12	3,35E-12	3,35E-12
w59	-4,44E-18	-9,52E-19	-2,90E-19	w68	3,20E-12	3,20E-12	3,20E-12
w79	-1,25E-18	-2,93E-18	-1,27E-18	w80	-6,49E-12	6,54E-13	-2,37E-12
w07	-1,55E-18	1,87E-18	1,75E-18	w06	-4,40E-12	2,76E-12	-2,04E-12
w54	4,61E-18	3,85E-19	2,45E-20	w87	4,14E-12	3,07E-12	1,83E-12
w62	-2,83E-18	-6,04E-19	-1,42E-18	w26	1,67E-12	4,34E-12	-2,83E-12
w73	1,50E-18	1,50E-18	1,50E-18	w67	5,06E-12	-2,47E-12	-1,25E-12
w04	-3,91E-19	-1,82E-18	-2,11E-18	w42	-2,88E-12	-2,88E-12	-2,88E-12
w82	-1,29E-18	-2,26E-18	2,04E-19	w47	4,06E-12	3,04E-12	1,13E-12
w47	1,67E-20	1,95E-18	-1,55E-18	w22	2,67E-12	2,67E-12	2,67E-12
w87	-4,68E-19	1,80E-18	-1,23E-18	w38	-2,54E-12	-2,54E-12	-2,54E-12
w17	-3,83E-19	1,41E-18	-1,46E-18	w25	2,32E-12	3,49E-12	1,53E-12
w84	5,71E-20	1,31E-18	-1,38E-18	w02	-2,40E-12	-2,40E-12	-2,40E-12
w11	-1,84E-18	-4,96E-19	3,94E-19	w88	-2,38E-12	-2,37E-12	-2,35E-12
w77	1,67E-18	-7,02E-19	1,28E-19	w82	-2,13E-12	-3,64E-12	-1,13E-12
w13	-1,33E-19	-1,52E-18	8,14E-19	w69	-2,00E-12	-2,00E-12	-2,00E-12
w19	3,70E-19	-1,24E-18	8,30E-19	w86	-2,00E-12	-2,00E-12	-2,00E-12
w25	-7,23E-19	1,32E-19	1,55E-18	w39	-1,94E-12	-1,94E-12	-1,94E-12
w14	7,40E-19	1,61E-18	5,49E-20	w09	-3,96E-12	-1,41E-12	-3,40E-13
w02	-8,00E-19	-8,00E-19	-8,00E-19	w41	-2,24E-12	-1,83E-12	-1,50E-12
w72	8,00E-19	8,00E-19	8,00E-19	w33	4,78E-12	-1,68E-13	4,37E-14
w57	2,07E-18	2,09E-19	1,62E-20	w27	1,85E-12	-1,27E-12	1,64E-12
w05	7,95E-19	9,91E-19	4,08E-19	w32	-1,51E-12	-1,51E-12	-1,51E-12
w33	-1,65E-19	-1,78E-18	-2,08E-20	w64	-3,68E-12	7,07E-13	1,38E-14
w52	-1,83E-18	-7,95E-20	4,30E-20	w81	1,39E-12	1,39E-12	1,39E-12
w09	7,32E-20	-1,03E-18	-7,12E-19	w17	1,44E-13	-1,20E-12	-2,55E-12
w21	-2,46E-19	9,28E-19	-4,65E-19	w60	-1,18E-12	-1,18E-12	-1,18E-12
w08	6,62E-19	-9,26E-19	2,82E-20	w78	-1,15E-12	-1,15E-12	-1,15E-12
w10	6,62E-19	-9,26E-19	2,82E-20	w66	3,06E-12	-3,05E-13	3,55E-14
w58	-1,29E-18	-8,65E-20	2,04E-19	w85	1,11E-12	1,11E-12	1,11E-12
w24	-3,20E-21	7,35E-19	-8,30E-19	w37	1,10E-12	1,10E-12	1,10E-12
w70	-5,00E-19	-5,00E-19	-5,00E-19	w18	-1,82E-12	-1,06E-12	3,85E-13
w12	-8,31E-19	-5,16E-19	-1,31E-19	w46	-1,82E-12	-1,06E-12	3,85E-13
w67	1,20E-19	-1,11E-18	-1,86E-19	w28	-1,07E-12	-1,07E-12	-1,07E-12
w03	4,72E-19	-3,53E-19	-5,90E-19	w84	4,76E-13	-7,66E-13	-1,93E-12
w60	-4,00E-19	-4,00E-19	-4,00E-19	w35	-2,10E-13	-1,71E-12	-1,05E-12
w18	-3,23E-19	8,08E-19	-3,75E-20	w40	9,83E-13	9,83E-13	9,83E-13
w46	-3,23E-19	8,08E-19	-3,75E-20	w15	-9,21E-13	-9,21E-13	-9,21E-13
w63	-4,61E-20	-6,01E-19	-2,87E-19	w08	-1,47E-12	-7,70E-13	-3,03E-13
w31	-3,00E-19	-3,00E-19	-3,00E-19	w10	-1,47E-12	-7,70E-13	-3,03E-13
w38	-3,00E-19	-3,00E-19	-3,00E-19	w14	1,30E-12	7,27E-13	2,72E-13
w42	-3,00E-19	-3,00E-19	-3,00E-19	w52	2,02E-12	2,28E-13	1,78E-14

w61	-3,00E-19	-3,00E-19	-3,00E-19	w16	6,80E-13	6,80E-13	6,80E-13
w69	3,00E-19	3,00E-19	3,00E-19	w43	6,72E-13	6,72E-13	6,72E-13
w81	-3,00E-19	-3,00E-19	-3,00E-19	w58	-6,93E-13	-8,46E-13	-2,74E-13
w88	-3,87E-19	2,34E-20	4,31E-19	w44	-4,04E-13	-4,04E-13	-4,04E-13
w27	-1,56E-19	-4,15E-19	-2,49E-19	w74	8,12E-13	2,83E-14	-2,12E-13
w41	-3,39E-19	-2,20E-19	-1,63E-19	w61	-3,32E-13	-3,32E-13	-3,32E-13
w26	4,04E-20	-5,17E-19	-9,49E-20	w75	-3,75E-13	-3,89E-13	-1,81E-13
w53	-5,75E-19	2,84E-20	3,37E-20	w07	4,38E-13	3,01E-13	1,25E-13
w55	-5,75E-19	2,84E-20	3,37E-20	w24	-2,35E-13	-5,99E-13	-2,03E-14
w16	2,00E-19	2,00E-19	2,00E-19	w3	-5,31E-13	-1,26E-13	-3,10E-14
w22	-2,00E-19	-2,00E-19	-2,00E-19	w45	2,20E-13	2,20E-13	2,20E-13
w23	-2,00E-19	-2,00E-19	-2,00E-19	w56	5,85E-13	6,54E-14	5,58E-15
w29	-2,00E-19	-2,00E-19	-2,00E-19	w53	4,81E-13	9,49E-14	1,35E-14
w32	-2,00E-19	-2,00E-19	-2,00E-19	w55	4,81E-13	9,49E-14	1,35E-14
w36	-2,00E-19	-2,00E-19	-2,00E-19	w04	-4,50E-13	-8,79E-14	-1,74E-14
w37	-2,00E-19	-2,00E-19	-2,00E-19	w30	1,85E-13	1,85E-13	1,85E-13
w43	-2,00E-19	-2,00E-19	-2,00E-19	w83	2,12E-13	2,32E-13	-6,22E-14
w44	-2,00E-19	-2,00E-19	-2,00E-19	w54	3,81E-13	9,73E-14	1,11E-14
w78	-2,00E-19	-2,00E-19	-2,00E-19	w51	-7,12E-14	2,33E-13	1,17E-13
w66	3,52E-19	2,64E-21	1,85E-21	w65	4,87E-14	2,52E-13	-7,95E-14
w15	1,00E-19	1,00E-19	1,00E-19	w59	3,29E-13	3,90E-14	3,41E-15
w20	-1,00E-19	-1,00E-19	-1,00E-19	w77	2,64E-13	2,50E-14	2,68E-15
w30	1,00E-19	1,00E-19	1,00E-19	w11	-2,38E-14	1,42E-13	7,14E-14
w39	-1,00E-19	-1,00E-19	-1,00E-19	w62	-1,28E-13	-4,52E-15	1,81E-16
w45	-1,00E-19	-1,00E-19	-1,00E-19	w57	-1,05E-13	-2,21E-15	2,59E-16
w50	-1,30E-19	-1,21E-19	-3,37E-20	w12	4,48E-14	3,37E-14	-3,34E-16
w35	-6,64E-20	-1,49E-19	6,57E-20	w63	-2,24E-14	3,04E-15	3,30E-16
w56	-1,90E-19	1,06E-20	3,07E-21	w01	0	0	0
w86	-5,98E-32	-5,98E-32	-5,98E-32	w20	0	0	0
w40	-5,08E-32	-5,08E-32	-5,08E-32	w29	0	0	0
w85	-4,03E-32	-4,03E-32	-4,03E-32	w31	0	0	0
w28	-3,12E-32	-3,12E-32	-3,12E-32	w36	0	0	0
w68	-2,60E-33	-2,60E-33	-2,60E-33	w49	0	0	0
w49	0	0	0	w50	0	0	0

Table A2. Detail description of attributes used in proposed classification method. Shortcut ‘avg’ used in meaning of arithmetical mean.

w01	Type of Question answer; 0 if LO is not a Question	w45	Flag if LO type is s
w02	Flag if LO is approved by LO users	w46	Day in month
w03	Avg session length (LO visits) in current course	w47	Month in year
w04	Avg session length (LO visits) in current course normalized by static constant	w48	LO visit order in current session
w05	Difference of session length (LO visits) between current course and avg of all user’s courses	w49	LO rating from users in course

w06	Difference of session length (LO visits) between current course and avg of all user's courses normalized by static constant	w50	User's role (student, teacher)
w07	Avg session length (LO visits) in current week of year	w51	Avg session length (time)
w08	Avg session length (LO visits) in current day	w52	Avg session length(time) in current week of year
w09	Avg session length (LO visits) in current hour	w53	Avg session length (time) in current day
w10	Avg session length (LO visits) in current day of month	w54	Avg session length (time) in current hour
w11	Avg session length (LO visits) in current month	w55	Avg session length (time) in current day of month
w12	Avg session length (LO visits) in current day of week	w56	Avg session length (time) in current month
w13	Avg session length (LO visits) in current day of year	w57	Avg session length (time) in current day of week
w14	Avg session length (LO visits) in current year	w58	Avg session length (time) in current day of year
w15	Actual session length (LO visits) in course above avg value	w59	Avg session length (time) in current year
w16	Actual session length (time) in course above avg value	w60	Flag if it is summer exams
w17	Week of year	w61	Flag if it is summer semester
w18	Day in month	w62	Avg session length (time) in current course
w19	Difficulty of LO	w63	Avg session length (time) in current course normalized by static constant
w20	Flag if it is holiday	w64	Difference of session length (time) between current course and avg of all user's courses
w21	Hour in day	w65	Difference of session length (time) between current course and avg of all user's courses normalized by static constant
w22	Flag if LO is 1st in chapter	w66	Type of LO relation (did exercise relation, did question relation, rated difficulty of LO relation, Followed LO link relation, rated LO relation, visited LO relation)
w23	Flag if LO is last in chapter	w67	Type of LO (explanation, question, exercise)
w24	Flag if LO is in setup	w68	Flag if user did exercise relation
w25	Element in system from which user comes to LO (menu, widget, etc.)	w69	Flag if user did question relation
w26	LO title	w70	Flag if user rated difficulty of LO relation
w27	LO course	w71	Flag if user followed LO link relation
w28	Flag if LO includes to functional and logical programming course	w72	Flag if user rated LO relation
w29	Flag if LO includes to the course of programming in Lisp language	w73	Flag if user visited LO relation
w30	Flag if LO includes to the course in procedural programming	w74	User's avg session length (LO visits)
w31	Flag if LO includes to the course of programming in Prolog language	w75	User's avg session length (LO visits) normalized by static constant
w32	Flag if LO includes to software engineering course	w76	User's avg session length (time)
w33	Order in course chapters hierarchy	w77	User's avg session length (time) normalized by static constant
w34	Global avg probability of leaving the system after visiting current LO	w78	Length of current session (LO visits) above user's avg session length
w35	Source of LO	w79	Difference of user's avg session length (LO visits) and avg of all users' sessions

w36	Flag if LO source is bok	w80	Difference of user’s avg session length (LO visits) and avg of all users’ sessions normalized by static constant
w37	Flag if LO source is book	w81	Length of current session (time) above user’s avg session length
w38	Flag if LO source is op	w82	Difference of user’s avg session length (time) and avg of all users’ sessions
w39	Flag if LO source is sg	w83	Difference of user’s avg session length (time) and avg of all users’ sessions normalized by static constant
w40	Flag if LO source is tg	w84	Day of week
w41	Type of LO	w85	Flag if it is winter exams
w42	Flag if LO type is e	w86	Flag if it is winter semester
w43	Flag if LO type is p	w87	Day of year
w44	Flag if LO type is q	w88	Year

Table A3. Description of data logged in e-learning system ALEF (Bielikova et al., 2014) and used in proposed classification method. Logged data belong into the three main categories – data describing Users, Learning objects and Interaction itself.

Category	Attribute	Description
User	User_id	Unique identifier of user
	Role	User role [Student | Teacher]
Learning object	Learning_object_id	Unique identifier of LO
	Label	LO title
	Type	Type of LO [Explanation | Question | Exercise]
	Rating	LO popularity
	Difficulty	LO difficulty <0;1>
	Parent	Superior LO in the course hierarchy
	Course	Superior course
Interaction	Type	Type of relation [Visit from menu | Hyperlink follow | Suggestion follow]
	Interaction	Type of source [Menu | Widget | Outside]
	Duration	Visit duration in seconds
	Created_at	Timestamp