Engineering Applications of Artificial Intelligence 51 (2016) 136–149

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Student behavior in a web-based educational system: Exit intent prediction

Ondrej Kassak, Michal Kompan, Maria Bielikova *

Faculty of Informatics and Information Technologies, Slovak University of Technology in Bratislava, Ilkovicova 2, 842 16 Bratislava, Slovakia

ARTICLE INFO

Available online 2 February 2016

Keywords:
User-feedback
E-learning
Behavior prediction
Classification
Stochastic gradient descent

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.

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.

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 Deborah 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

* Corresponding author.
E-mail address: Maria.Bielikova@stuba.sk (M. Bielikova).

http://dx.doi.org/10.1016/j.engappai.2016.01.018
0952-1976/

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. (2012) proposed the clustering method for grouping students based on their cognitive learning style. 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 (1996) and is still intensively researched nowadays. There were proposed the methods of personalized recommendation, such as hybrid approach from Klasnja-Milevic et al. (2011) which is similarly to previous approach based on students' learning styles, but in this case also on frequent sequences of in content learned.

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 modeling. A process of automatic relationship discovery in domain model considering learning objects as key elements was researched by (Simko et al., 2009), an adaptive question selection by (Barla et al., 2010). Other approaches used for improving the e-learning use the predictive data mining analysis (Pena-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 Pesin, 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 (Pena-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 Balamuralithara, 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 Balamuralithara, 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. (2014) possible to notice first signals (of students' dropout intentions) at least two weeks before dropout itself. On the contrary, the short term periods as sessions, bring significantly less time to discover