

INTELLIGENT MINING ON ASYNCHRONOUS E-LEARNING INFORMATION SYSTEM

GENADIJUS KULVIETIS, JELENA MAMČENKO, IRMA ŠILEIKIENĖ
Information Technology Department
Vilnius Gediminas Technical University
Saulėtekio al. 11, Vilnius
LITHUANIA

Abstract. Distance education is organised in three main forms: *asynchronous*, *synchronous* and *blended*. Student's data for analysis will be used only the asynchronous learning portion of the information system, since this portion is already used in master's studies for a considerable time, and a sufficient amount of data has been accumulated. Data analysis will be performed on the data of master students and bachelor students as well, that can learn in 11 e-courses. Clustering technique was used for user's data analysis which is fitting best when we know nothing about them. After clustering database we have got nine clusters and that's clusters' interpretation, which let us to except typical student's behaviours.

Keywords: e-learning, asynchronous learning, information system, data mining, intelligent mining

1. Introduction

Distance education is organised in three main forms: *asynchronous*, *synchronous* and *blended*. They differ fundamentally in terms of functionality, so they also naturally differ in terms of technology.

Asynchronous collaboration tools do not require learners and teachers to be online at the same time to communicate with each other. Messages, questions, or assignments can be posted and answered at any moment. Others can read them and respond when they are online.

The architecture of **asynchronous type** is really simple. From a technical standpoint, if there are not many e-courses and students, one server is sufficient with a chosen VLE of asynchronous type installed, in which all required tools of e-learning course creation and provision are integrated. If there are many courses and students, one server might not suffice; in such a case it is recommended to implement a LMS (Learning Management System) with some system for creating courses.

Hence, it can be stated that the architecture of asynchronous type is simple, not hardware-investment intensive, accessible to many, and thus very popular.

Synchronous collaboration provides online communication in real-time. Teachers and learners

come together electronically at a given time, without leaving the workplace or home; they use tools such as chat, audio, video, and whiteboards.

Synchronous learning can be realized in several ways: 1) *video/audio conferences* classified into *point-to-point*, *multipoint* and *classroom – lecture hall* [2]; 2) *workplace videoconferences*.

To organise videoconferences e-learning technologies of synchronous type can be used: Centra Symposium, LearningSpace – Virtual Classroom, ViPS; as well as general purpose technologies for organising virtual meetings: Placeware, WebEx, Interwise, Microsoft NetMeeting, Polycom, CU-SeeMe and the like. **The blended learning method** is [1]:

- a combination of various networked technologies in a single learning package;
- a synthesis of various pedagogic methods that enables to achieve an optimal quality of learning process;
- a combination of various lecturing technologies (video cassettes, compact discs, internet material, etc.) together with direct lecturing by an instructor.

The blended learning method is enable by realising the coupling between virtual classroom and LMS technologies together with course management systems (CMS) Authoring Tool, LearningSpace,

WebCT and others conforming to the SCORM standard.

An information system enabling to realise the blended learning method has been designed and implemented at the Information Technologies Department of VGTU. However, for the analysis of student data only the asynchronous learning portion of the information system will be used, since this portion is already used in master's studies for a considerable time, and a sufficient amount of data has been accumulated. The synchronous learning portion

is still being tested and is not used with students yet for practical purposes.

Asynchronous courses are created with the technology of Lotus IBM LearningSpace, which is implemented on the irma.vtu.lt server. The information on users and their log-in rights is stored on a LDAP server, which is implemented on the gama.vtu.lt server. Shared data on user activities in the e-learning system are recorded and stored in the files of the Domino Web Servers Log. The schema of users' data workflow is presented in Fig. 1.

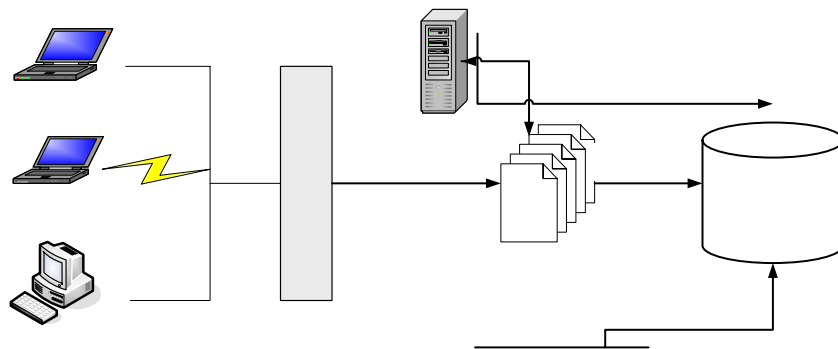


Fig. 1. Users' data workflow schema of e-learning information system

The information on users from the LDAP server, the course information and student results from LearningSpace courses, and the logs from the Domino Web Server are converted and stored in the Data Warehouse. The data from the Data Warehouse will be used in data analysis.

At this time there are 11 e-courses in the LearningSpace course catalogue including four courses for master's studies: ODL (Open Distance Learning) Infrastructure, Multimedia, Organising

Information Activities, and Virtual Learning Environment. The e-courses have been used in master's studies for three years; about 70 master students who are identified in the system have studied these courses. Other courses are meant for bachelor students in Informatics as supplementary material or lecture notes. Students are not registered for these courses; they log in using a shared student account. Data analysis will be performed on the data of master students and bachelor students as well.

2. Problem formulation

To find out more about user's was used clustering technique which is fitting best when we know nothing about them. We can confirm that using this technique is the first step in data analysis. Depending on revealed results later will be possible to use on the same data other techniques.

Clustering is widely used in data mining [10]. The main advantage is that we can use it when not trying to predict something specific, when we have no idea

about database structure [3, 4]. It can be very useful since it can reveal some patterns which will be useful applying other techniques for further analysis.

Whereas clustering can be applied to different data format we decided to apply it to web data. Despite that clustering technique is commonly used in a wide range of statistical package it does not mean that data format will be accepted and data will be analyzed. Some of them uses spreadsheet format which is limiting by records, for instance MS Excel is limited by 65536 records, or plain text format. Some other

software simply offer to use smaller data set to perform data analysis.

Clustering technique aim is to group together similar items in a databases. Domlog.nsf database is enough huge in size and it would be better to have a picture of the users being divided into small groups by their characteristics. The problem is that we can not determine characteristics we would like each group to have; the clustering is done first and then examines the clusters to see what their members have in common.

As was mentioned before, analyzed file domlog.nsf is Lotus Notes® document database which has specific structure but ordinary for access log file format.

Whereas Lotus Notes® database data need to be converted/transformed to particular format; there are two ways to export access log file data to particular software for deeper analysis:

- Export data to a flat text format;
- Or create Lotus agent for loading data to a relational database.

Were used both ways for comparing which fits best and uses less time for data preparation for mining.

Data transformation is one of the most important steps in data mining process (Fig. 2) and sometimes it takes much more time comparing with data analysis.

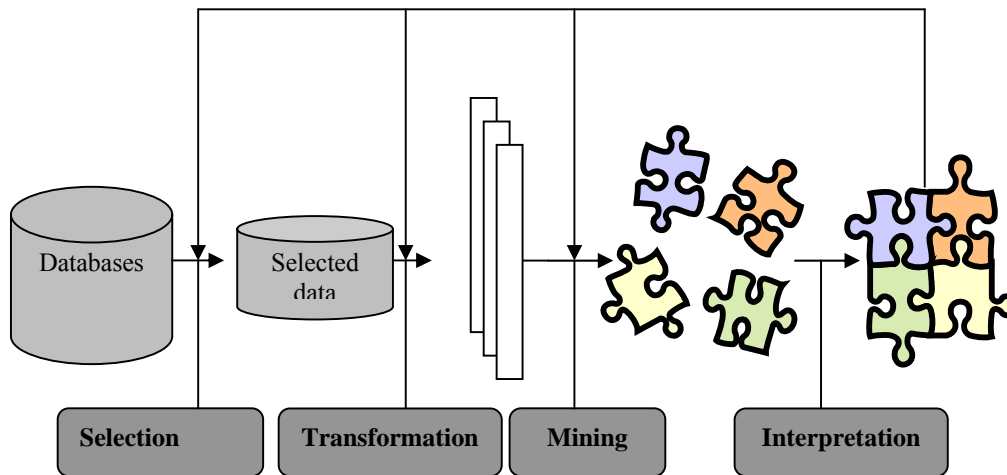


Fig. 2. The process of Data Mining

To convert data from document database to a flat file format some problems with data transformation have occurred. Particular software requires that data in that file should be aligned by columns and every record's end must have enough spaces. It means that any shorter last column record must have the same number of spaces as the longest last column record has symbols. To escape such cross-purposes we had to write C++ script. Missing values were removed manually.

Load data from access log file Lotus Notes® document database to relational DB2 database requires create an automated agent.

We programmed Notes to perform tasks automatically using agents (also known as macros) which can help to perform repetitive tasks, such as managing documents and sending memos, in this work it was used to load data from document

database into relational Domlog database. We described 13 fields (Fig. 3), which will be used for intelligent analysis.

Another concern we should have when analyzing a web data is that most of the existing Web analysis tools provide mechanisms for reporting user activity in servers and various forms of data filtering. Using such tools, for example, it is possible to determine the number of accesses to the server and the individual files within the organization's Web space, the times or time intervals of visits, and domain names and the URLs of users of the Web server. However, in general, these tools are designed to handle low to moderate traffic servers, and furthermore, they usually provide little or no analysis of data relationships between the accessed files and directories within the Web space.

Key	Name	Data type	Length	Nullable
	ID	BIGINT	8	No
	DATE1	DATE	4	No
	TIME	TIME	3	No
	IP	VARCHAR	20	No
	AUSER	VARCHAR	40	No
	STATUS	VARCHAR	5	No
	CLENGTH	VARCHAR	15	No
	CTTYPE	VARCHAR	15	No
	METHOD	VARCHAR	35	No
	BROWSER	VARCHAR	35	No
	ETIME	BIGINT	8	No
	UTRANS	VARCHAR	150	No
	COURSELOCATION	CHARACTER	50	Yes

Fig. 3. DB2 table Domlog structure

Another note related to this work is that it's necessary to clean a server log file to eliminate irrelevant items as it is important for any type of Web log analysis, not just data mining. The discovered associations or reported statistics are only useful if the data represented in the server log give an accurate picture of the user accesses of the Web site. Elimination of irrelevant items can be reasonably accomplished by checking the suffix of the URL name. For instance, all log entries with filename suffixes GIF, JPEG, JPG, and map can be removed [5]. Summarizing two ways of data transformation from document database we decided that the best way is to export data to relational database.

5. Practical results

We analyzed a little bit more than about 150000 records from 2005.11.12 04:11 to 2006.02.23 19:25. We decided to include filter condition on filed *ETIME* to receive information about user who spends time more than one minute on site. Was assumed that users who spent less than one minute on site have entered it just by mistake. After clustering database we have got nine clusters which are shown in Table 1. The last and most important step in data mining process (Fig. 2) is results interpretation.

Table 1. Generated results and interpretation

Name	Size	Characteristics
[8] 7	37,29%	88 times user <i>Marija Bilksto</i> was connected to COURSELOCATION <i>struktura</i> to find additional material and papers and spend each time about 3 min.
[6] 5	29,66%	Bachelors who were interesting in new task receiving after one moth of semester began and spent about 2 min on course <i>Organising Information Activities</i> .
[2] 1	8,47%	Users who by mistake or not having passwords tried to connect and spent about 25 sec on site
[4] 3	6,78%	User <i>Lina Zajanckauskiene</i> spent about 4 min on site analyzing results after exam passing
[1] 0	3,81%	Shows user activity in discussion room, which were not authenticated and spent there about 13 min (compared with other cluster spent time shows high user activity)
[3] 2	3,81%	<i>Danguole Kamandulyte</i> who tried to login in to COURSELOCATION multimedia lecture notes. She spent too much time logging about 13 min (may be she forgot password)
[7] 6	3,81%	User <i>Vidas Valkiunas</i> spent 3,5 min in course room attaching home work for sent to teacher

[9] 8	3,39%	<i>Liudmila Terechova</i> was connected too much time (she was interesting more than others) and spent about 2 min each time to make know exam results
[5] 4	2,97%	User <i>Danguole Kamandulyte</i> looked over course catalogues spent about 7 min

6. Conclusions and future work

In this paper we discussed one of the possible applications of Data Mining technology applied to access log file which data were collected from e-learning systems. One of the data mining techniques was realized with IBM Intelligent Miner for Data.

Two data transformation types (flat file and databases) were presented and discussed their advantages and disadvantages were presented. According to received results, which were based on users behaviour they were interpreted and the next step is to implement them to business/educational decisions which should be done by teachers or other experts of this domain.

In future work we will consider on performing analysis of student activity and behaviour as well, will intend to research the relationship between student activity and study results; i.e., to research how intensely students study in e-courses, and whether their study results depend on the intensity. In addition to that it will be interesting to classify them by particular characteristics. Having the university domain, is that once we perform the clustering we might end up observing a cluster being full of students coming from a certain postcode, it might be interesting to investigate it further.

References:

1. Kulvietiene, R., Sileikiene, I., Stankevic, J.: Learning Management System for blended e-learning delivery. In 29th International conference Optimal Teaching and Learning: Achieving higher Education Excellence. Switzerland (2004).
2. Rui, Y., Gupta, A., Grudin, J., He, L.: Automating lecture capture and broadcast: technology and videography. *Multimedia Systems*, Vol. 10. Springer-Verlag (2004) 3-15.
3. Berry Michael J.A. and Linoff Gordon, 1997, *Data Mining techniques: For Marketing, Sales and Customer Support*, John Wiley&Sons Ltd.
4. Mena Jesus, 1999, *Data Mining Your Website*, dpDigital Press.
5. R.Cooley, B.Mobasher, J.Srivastava. *Web Mining: Information and Pattern Discovery on the World Wide Web*.
6. D.Monk, 2005, *Using Data Mining for E-Learning Decision Making*. *The Electronic Journal of e-Learning*, Volume 2, Issue 1, pp. 41-54, is available online at www.ejel.org.
7. K.A.Forcht and K.Cochran, 1999, *Using Data Mining and Data Warehousing Techniques*. *Industrial Management & Data Systems*, Volume 5, pp.189-196.