Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/080238

Analysis of Stakeholders' Behaviour Depending on Time in Virtual Learning Environment

Michal Munk[*] *and Martin Drlík*

Constantine the Philosopher University in Nitra, Nitra 949 74, Slovak Republic

Received: 10 Apr. 2013, Revised: 11 Aug. 2013, Accepted: 13 Aug. 2013
Published online: 1 Mar. 2014

Abstract: The aim of the paper is the probability modelling of accesses to the different parts of e-learning course in interactive learning environment depending on time. This problem belongs to the open questions of the contemporary emerging discipline Educational Data Mining. We are concerned with the access probabilities to the individual parts of e-learning course content. For the purpose of modelling of the stakeholders' behaviour dependence on time, we use multinominal logit model which is a special case of Generalized Linear Model. We pay attention to data preparation issues. Data about using e-learning courses are stored in databases. We assume that these data represent time data. Surprisingly, modelling of stakeholders' behaviour dependence on time is missing in Educational Data Mining discipline. The time variable, mostly stored in database in "unixtime" form, integrating date and time, is partially used by sequence rules extraction, where it is only used for the tracking of visited e-learning course parts during each session. Therefore, we describe the applied multinominal logit model and methodology of the modelling in detail. We deal with parameter estimations. Finally, we figure that the multinomial logit model finds its application mainly in the process of restructuring the existing e-learning courses. We discuss about its possible contribution to the improvement of the interactive learning environment as well as in the personalization of the course content and structure.

Keywords: web log mining, data preparation, user behaviour, discovering patterns, virtual learning environment

1 Introduction

One of the first tasks of web mining is the analysis of behaviour of website visitors. Based on this analysis, in educational contexts, it is possible to personalize e-learning courses, adapt educational hypermedia, discover potential browsing problems, automatically identify learning groups in exploratory learning environments or predict student's performance. The analysis of the unique types of data coming from educational systems can help find the most effective structure of e-learning courses, optimize the learning content, recommend the most suitable learning path based on student's behaviour, or provide more personalized environment. The aim of the paper is the probability modelling of accesses to the categories of activities in e-learning course of interactive learning portal at the university. For this purpose, data about accesses to the activities of the e-learning course were collected. We are concerned with the access probabilities to individual activities of e-learning course content depending on the part of the week (workweek and weekend). The probabilities are estimated through multinomial logit model [1,2,3] for students connected from inside and outside of the university network separately.

The multiple logistic regression model that we will use in this paper is fully described in the book [4]. This model is a special type of the generalized linear model [5]. The basic of this theory is obtained in the book [3,6]. We can find its applications mostly in econometrics, genetics and natural language processing. Its usage in the same research area as we describe in our paper is relatively infrequent. If the multiple logistic regression model is used then it is used mainly for choice prediction [7,8].

In general, discovering association and sequence rules, segmentation (cluster analysis, methods based on analogy, etc.), and classification (decision rules, decision trees, Bayes classification, etc.) are the most applied methods in the web log mining (WLM) [9]. Out of the statistic methods we apply e.g. analysis of crosstabulations relative to the categorical character of data [10].

[*] Corresponding author e-mail: mmunk@ukf.sk

Data reading into a data cube is another often used method to carry out OLAP operations for data summarization [11], but some tools combine OLAP and knowledge discovery [12]. Automatically saved data in log files are used as a source of the data, which are, from the point of knowledge discovery, represented by time data. Most methods applied to such data are used to perform segmentation of web visitors, extraction of behaviour patterns of web visitors, and finding associations among visited websites with the aim to personalize or optimize (restructure) websites according to the way they are browsed. The same applies to the web-based educational system [13,14,15,16]. We reviewed other applications of WLM in e-commerce [17, 18] and in recommendation systems development [19, 20]. But neither of the above mentioned methods models the behaviour of the user depending on time. Exactly for this purpose we used multinomial logit model for modelling the probability of accesses depending on time.

The rest of the paper is structured as follows. The next chapter deals with the data pre-processing tasks. We describe used multinomial logit model in third chapter. The fourth chapter describes the determination of the model and parameter estimations. We deal with its evaluation in fifth chapter. Finally, we discuss about results and figure that the multinomial regression model finds its application in various stages of the e-learning course's development life cycle, as well as in the personalization of the course content and learning management.

2 Data pre-processing

The information available on the web is heterogeneous and unstructured [21]. The goal of data preparation is to transform the raw data stored in logs into a set of user profiles [22]. Contemporary interactive learning environments store information about their users not in server log file but mainly in a relational database. Interactive learning environments manage all their services through a relational database. We can find there large log data of students' activities and, usually, interactive learning environments have built-in student monitoring features so they can record all kinds of student activities [23]. Compared to other data mining applications [24], a relational database provides an integrated source of data saving pre-processing effort.

On the contrary, Raju and Satyanarayana [16] point out that the same approach requires an extensive data pre-processing before obtained a single analysis table. They argue that process of database normalization means a problem for mining algorithms that require data to be assembled into a single, integrated and, in short, analysis table. The content of the analysis table is to some extend domain dependent and even, inside a particular domain, task dependent. An important initial decision is concerned with the granularity of the information contained in table.

In the e-learning context, unlike other web based domains, user identification is a straightforward problem as in most cases learners must login using their unique ID [25].

We used log file created from database containing records from the e-learning course with 180 participants. The e-learning course has been created in LMS Moodle. The log file has been dumped from the tables mdl_log and mdl_log_display. Records have been cleaned from irrelevant items. First of all, we have removed entries about users with the role other than student. After performing this task we defined a set WALS (Web Access Log Set) of accesses into the interactive learning environment stored in log file and ordered by time

$$WALS = \{\langle ID, time, userID, IP, module, action\rangle\}.$$

We can see the part of final WALS in Table 1. Finally, 70 553 entries in WALS were accepted to be used in the next task.

Table 1 Part of Web Access Log Set obtained from LMS Moodle tables

ID	time	user ID	IP	module	action
1645357	1222112879	2	88.80.227.157	course	view
1645358	1222112883	2	88.80.227.157	user	view all
1645359	1222112908	2	88.80.227.157	role	assign
1645360	1222112928	2	88.80.227.157	role	assign
1645361	1222112932	2	88.80.227.157	course	view
1645362	1222112939	2	88.80.227.157	resource	view
1645363	1222113015	2	88.80.227.157	course	update mod
1645364	1222113015	2	88.80.227.157	resource	update

2.1 Session Identification

A user session is defined as a sequence of requests made by a single user over a certain navigation period and a user may have a single (or multiple) session(s) during this time period. Session identification is a process of segmenting the log data of each user into individual access sessions [22]. We found many approaches to session identification [26,27,28,29]. The excellent review of user session identification was made in Chitraa and Davamani [22] and Spiliopoulou, Mobasher, Berendt and Nakagawa [30].

First of all, we need to identify a particular user. In the context of interactive learning environments, unlike other web based domains, user identification is a straightforward problem because in the most cases, the learners must login using their unique ID [25].

Therefore, we used user ID for identification of particular user. But, we need to know user's session details to get accurate mining results; therefore we dealt

with methods of session identification in more detail in the next phase of data pre-processing.

A user may have a single (or multiple) session(s) during a period of time. The interactive learning environment stores her/his accesses in log file. Therefore, the aim of session identification phase of data pre-processing is to divide particular accesses of individual user into disjoint sequences. We can define these sequences in several ways.

If we are processing accesses after they are handled by the interactive learning environment, this technique is called ”reactive” while in ”proactive” technique the same (pre)processing occurs during the interactive browsing of web pages of the interactive learning environment by the user.

Reactive session identification uses time and navigation driven heuristics. In regard to the specific features of used interactive learning environment the navigation driven heuristics is out of the question. We turn our attention to time-driven heuristics. Two time-driven heuristic methods are often mentioned in literature [26]:

–session-duration based method,
–timeout threshold based method.

The session-duration based method assumes that if we can estimate the duration of the session θ then we can define the session as a sequence of visited pages in interactive learning environment where each page has its own timestamp for which

$$USS = \{\langle userID, IP, \langle ID_1, time_1\rangle, \ldots, \langle ID_k, time_k\rangle\rangle\},$$
$$time_k - time_1 \leq \theta,$$

where USS means User Session Set, $USS \subset WALS$, $time_k$ is timestamp of the last record in session, $1 \leq k \leq n$ and n is count of records in $WALS$. All other records of $WALS$ with timestamp greater than $time_1 + \theta$ belong to the next session. Contrary, timeout threshold based method assumes that if we can estimate the timeout threshold δ then we can define the session as a sequence of visited pages in interactive learning environment where each page has its own timestamp for which

$$USS = \{\langle userID, IP, \langle ID_1, time_1\rangle, \ldots, \langle ID_k, time_k\rangle\rangle\},$$
$$time_i - time_{i-1} \leq \delta,$$

where $USS \subset WALS$, $time_k$ is timestamp of the last record in session and $1 \leq i \leq k$. If the inequality is not true for two consecutive records from $WALS$ then the records belong to two different sessions.

Against this background we have decided to use reactive time-oriented heuristic method based on time-threshold to identify the users' sessions. We have considered not only user ID, but also the IP address of a computer used by the user. Additionally, in spite of the recommended 30-minute-long timeout threshold [30] we adopted a 15-minute timeout threshold to start a new session with regard to the setting in used interactive learning environment. The correctness of our decision to use this timeout threshold was verified in our previous paper [31].

2.2 Path Reconstruction

Another problem upon searching for the users behavior patterns seems to be the analysis of the backward path, or reconstruction of activities of a visitor of interactive learning environment. The reconstruction of activities is focused on retrograde completion of records on the path went through by the user by means of a back button of web browser, since the use of such button is not automatically recorded into log entries interactive learning environment.

We found and analyzed several approaches mentioned in literature [30, 32]. Finally, we have chosen the same approach as in our previous paper [33].

We used sitemap and special algorithm for pass completion. A sitemap has a great importance for retrograde completion of the path. We can find information on the existence of a link among pages in sitemap, i.e. if a hyperlink from one page to another exists.

We obtained the sitemap for the needs of our analysis by means of Web Crawling application implemented in the used Data Miner. Having lined up the records according to the user ID, IP address and timestamp we searched for some linkages between the consecutive pages. A sequence of pages in interactive learning environment for the selected userID and IP address can look like this: A→ B→ C→ D→ X (Fig.1).

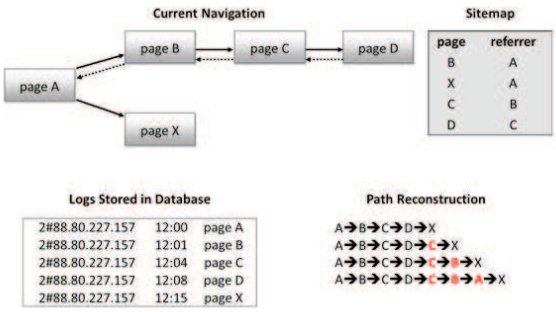

Fig. 1 Example of path reconstruction.

In our example, based on the sitemap the algorithm can find out if there not exists the hyperlink from the page D to our page X. Thus we assume that this page was accessed by the user by means of using a back button in web browser from one of the previous pages. Then, through a backward browsing we can find out, where of the previous pages exists a reference to page X.

In our example we can find out if there no exists a hyperlink to page X from page C, if C page is entered into the sequence, i.e. the sequence will look like this: A→ B→ C→ D→ C→ X. Similarly, we shall find that there

exists any hyperlink from page B to page X and so B can be added into the sequence, i.e. A→ B→ C→ D→ C→ B→ X. Finally algorithm finds out that the page A contains hyperlink to page X. After the termination of the backward path analysis the sequence will look like this: A→ B→ C→ D→ C→ B→ A→ X. Then it means the user used back button in browser in order to cross from page D to C, from C to B and from B to A.

After the application of this method we obtained the dataset with an identification of sessions based on user ID, IP address, time and completing the paths. The count of records increases to 75 372, which means 7% increase of records.

2.3 Derived Variables Calculation

The construction of some derived variables is the last step of data-preprocessing stage. The original log file contains only the variable *time*. We had to calculate other variables, WEEK (weekend and workweek) and TIME (hours in day) (Fig. 2).

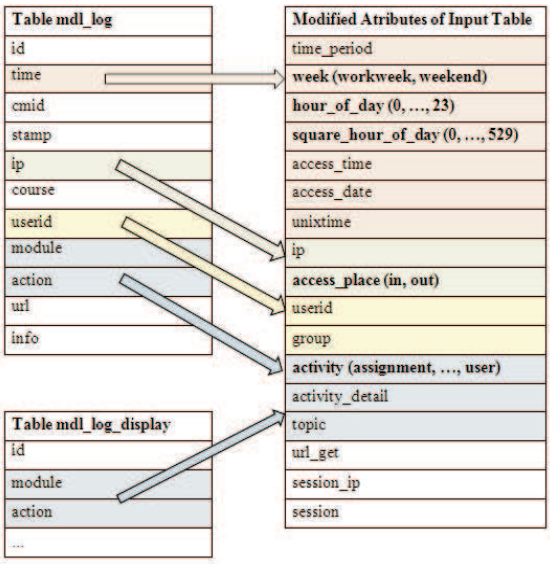

Fig. 2 Mapping attributes of WALS stored in relation tables into input dataset used in experiment. The variables and their ranges used in model are in bold.

3 Model Description

The data describes individual accesses of students into e-learning course during the winter term that was created in LMS Moodle.

First we describe the variables which we need to include in the model. The investigated categorical dependent variable is a variable ACTIVITY with categories: *assignment, book, blog, course, data* (i.e. files, home works and projects uploaded by students), *feedback, forum, glossary, quiz, resource* (i.e. information sources available in the course), *upload* and *user* (student's profile, his/her grades and information about other students).

The variable TIME with values 0-23 is non-dependent variable. We use variable WEEK with categories *weekend* (0) and *workweek* (1) and the variable ACCESS_PLACE with *in* (0) and *out* (1) as dummy variables. Figs. 11-12 show the empirical (and fitted) logits of activities during the workweek for internal accesses and for external accesses (using the last activity user as the reference category) plotted against the time. From the figures it seems to follow that the logits are quadratic functions of time. Similar results were obtained for the weekend. We will therefore include the variable square time TIME^2 in the model.

Let π_{ij} be the probability that the user will choose the web part j, in hour i, while $j = 1,2,\ldots,J$, where J is a number of web parts of portal or system and $i = 0,1,\ldots,23$.

Since $\sum_{j=1}^{J} \pi_{ij} = 1$ is true, there are only $J-1$ parameters.

Let Y_{ij} be the number of accesses into the web part j with observations y_{ij} in hour i, then $\sum_{j=1}^{J} y_{ij} = n_i$ is the number of accesses in hour i.

The probability distribution of the vector $Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iJ})^T$, if the sum n_i is given, is multinomial

$$f_i(y_{i1}, y_{i2}, \ldots, y_{iJ}) = P[Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \ldots, Y_{iJ} = y_{iJ}] = \\ = \frac{n_i!}{y_{i1}! y_{i2}! \ldots y_{iJ}!} \pi_{i1}^{y_{i1}} \pi_{i2}^{y_{i2}} \ldots \pi_{iJ}^{y_{iJ}}.$$

Taking logs, we find that

$$\ln f_i(y_{i1}, y_{i2}, \ldots, y_{iJ}) = \sum_{j=1}^{J} y_{ij} \ln \pi_{ij} + \ln \left(\frac{n_i!}{y_{i1}! y_{i2}! \ldots y_{iJ}!} \right).$$

Since $\sum_{j=1}^{J}\pi_{ij}=1$, we put $\pi_{iJ}=1-\sum_{j=1}^{J-1}\pi_{ij}$ and get

$$\begin{aligned}\ln f_i(y_{i1},y_{i2},\ldots,y_{iJ}) &= \sum_{j=1}^{J-1} y_{ij}\ln\pi_{ij}+y_{ij}\ln\left(1-\sum_{j=1}^{J-1}\pi_{ij}\right)\\ &\quad+\ln\left(\frac{n_i!}{y_{i1}!y_{i2}!\ldots y_{iJ}!}\right)\\ &= \sum_{j=1}^{J-1} y_{ij}\ln\pi_{ij}+\left(n_i-\sum_{j=1}^{J-1}y_{ij}\right)\ln\left(1-\sum_{j=1}^{J-1}\pi_{ij}\right)\\ &\quad+\ln\left(\frac{n_i!}{y_{i1}!y_{i2}!\ldots y_{iJ}!}\right)\\ &= \sum_{j=1}^{J-1} y_{ij}\ln\pi_{ij}+n_i\ln\left(1-\sum_{j=1}^{J-1}\pi_{ij}\right)\\ &\quad-\sum_{j=1}^{J-1} y_{ij}\ln\left(1-\sum_{j=1}^{J-1}\pi_{ij}\right)+\ln\left(\frac{n_i!}{y_{i1}!y_{i2}!\ldots y_{iJ}!}\right)\\ &= \sum_{j=1}^{J-1} y_{ij}\left(\ln\pi_{ij}-\ln\left(1-\sum_{j=1}^{J-1}\pi_{ij}\right)\right)+n_i\ln\left(1-\sum_{j=1}^{J-1}\pi_{ij}\right)\\ &\quad+\ln\left(\frac{n_i!}{y_{i1}!y_{i2}!\ldots y_{iJ}!}\right)\\ &= \sum_{j=1}^{J-1} y_{ij}\ln\frac{\pi_{ij}}{1-\sum_{j=1}^{J-1}\pi_{ij}}+n_i\ln\left(1-\sum_{j=1}^{J-1}\pi_{ij}\right)\\ &\quad+\ln\left(\frac{n_i!}{y_{i1}!y_{i2}!\ldots y_{iJ}!}\right).\end{aligned}$$

From the last modification we obtain

$$\begin{aligned}\ln f_i(y_{i1},y_{i2},\ldots,y_{iJ}) &= \sum_{j=1}^{J-1} y_{ij}\ln\frac{\pi_{ij}}{\pi_{iJ}}+\ln(\pi_{iJ})^{n_i}\\ &\quad+\ln\left(\frac{n_i!}{y_{i1}!y_{i2}!\ldots y_{iJ}!}\right).\end{aligned}$$

We take log

$$\begin{aligned}f_i(y_{i1},y_{i2},\ldots,y_{iJ}) &= \\ exp&\left(\sum_{j=1}^{J-1} y_{ij}\ln\frac{\pi_{ij}}{\pi_{iJ}}\right)\pi_{iJ}^{n_i}\frac{n_i!}{y_{i1}!y_{i2}!\ldots y_{iJ}!}\end{aligned}\quad(1)$$

and put

$$\eta_{ij}=\ln\frac{\pi_{ij}}{\pi_{iJ}},\eta_{ij}=0,\boldsymbol{\eta}_i=(\eta_{i1},\eta_{i2},\ldots,\eta_{iJ-1})^T,\boldsymbol{y}_i=$$
$$=(y_{i1},y_{i2},\ldots,y_{iJ})^T.$$

Then

$$\pi_{ij}=\pi_{iJ}e^{\eta_{ij}},j=1,2,\ldots,J-1$$
$$1=\sum_{j=1}^{J}\pi_{ij}=\pi_{iJ}\sum_{j=1}^{J}e^{\eta_{ij}}=\pi_{iJ}\left(1+\sum_{j=1}^{J-1}e^{\eta_{ij}}\right),$$

of wich we obtain

$$\pi_{iJ}=\frac{1}{1+\sum_{j=1}^{J-1}e^{\eta_{ij}}}.$$

Now the probability distribution function has general exponential form [1].

$$f_i(\boldsymbol{y}_i,\boldsymbol{\eta}_i)=C(\boldsymbol{\eta}_i)exp\left(\sum_{j=1}^{J-1}Q_j(\boldsymbol{\eta_i})T_j(\boldsymbol{y}_i)\right)u(\boldsymbol{y}_i),$$

where

$$\begin{aligned}C(\boldsymbol{\eta}_i) &= \left(1+\sum_{j=1}^{J-1}e^{\eta_{ij}}\right)^{-n_i},Q_j(\boldsymbol{\eta}_i)=\eta_{ij}T_j(\boldsymbol{y}_i)=y_{ij},u(\boldsymbol{y}_i)\\ &= \frac{n_i!}{y_{i1}!y_{i2}!\ldots y_{iJ}!}.\end{aligned}$$

Hence, we can apply generalized linear model with link function logit to estimate the probabilities π_{ij} of selecting web parts j with respect to the hour i [34]. From (Eq.1) we obtain

$$\ln f_i(\boldsymbol{y}_i,\boldsymbol{\eta}_i)=\ln\frac{e^{\sum_{j=1}^{J-1}y_{ij}\eta_{ij}}}{\left(1+\sum_{j=1}^{J-1}e^{\eta_{ij}}\right)}\frac{n_i!}{y_{i1}!y_{i2}!\ldots y_{iJ}!}.$$

The log-likehood function has a form

$$\begin{aligned}\ln\prod_i f_i(\boldsymbol{y}_i,\boldsymbol{\eta}_i) &= \sum_i\left(\sum_{j=1}^{J-1}y_{ij}\eta_{ij}-n_i\ln\left(1+\sum_{j=1}^{J-1}e^{\eta_{ij}}\right)\right)\\ &\quad+\sum_i\ln\frac{n_i!}{y_{i1}!y_{i2}!\ldots y_{iJ}!}.\end{aligned}$$

We assume that the following model is valid

$$n_{ij}=\ln\frac{\pi_{ij}}{\pi_{iJ}}=\boldsymbol{x}_i^T\boldsymbol{\beta}_j,j=1,2,\ldots,J-1,i\in\{0,\ldots,23\},\quad(2)$$

where π_{iJ} is the probability of the last web part, which we will choose as a referential, $\boldsymbol{x}_i^T$ is a line vector, $\boldsymbol{\beta}_j=(\beta_{j1},\beta_{j2},\ldots,\beta_{jk})^T$ is a vector of regression coefficients for $j=1,2,\ldots,J-1$.

There are $J-1$ equations which describe the contrasts between the web part j, for $j=1,2,\ldots,J-1$ and the last web part J (as a reference category, we could choose any other category).

Then there is a log-likehood function without the constants in form

$$\begin{aligned}\ln L(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\beta}) &= \ln\prod_i f_i(\boldsymbol{y}_i,\boldsymbol{x}_i,\boldsymbol{\beta}_j)\\ &= \sum_i\sum_{j=1}^{J-1}y_{ij}(\boldsymbol{x}_i^T\boldsymbol{\beta}_j)\\ &\quad-\sum_i n_i\ln\left(1+\sum_{j=1}^{J-1}e^{(\boldsymbol{x}_i^T\boldsymbol{\beta}_j)}\right).\end{aligned}\quad(3)$$

Maximum likelihood estimation of the parameters of the model (Eq. 2) proceeds by maximization of log of the multinomial likelihood function (without the constants) (Eq. 3).

The estimation of the parameters can be done by using an iteratively re-weighted least squares method as the Newton-Raphson technique or Fisher scoring [3]. The

starting values of estimations β_{j0} are computed from empirical logits

$$\eta_{ij0} = \ln \frac{p_{ij}}{p_{iJ}} = \boldsymbol{x}_i^T \boldsymbol{\beta}_j, p_{ij} = \frac{y_{ij}}{n_i}, j = 1,2,\ldots,J-1, \quad i \in \{0,1,\ldots,23\}$$

by linear regression.

The maximum likelihood estimation $\hat{\boldsymbol{\beta}}_j$ has approximately in large samples a multivariate normal distribution with mean equals to the true parameter value and with variance-covariance matrix given by the inverse of the information matrix.

Information matrix is a mean value of the matrix of the second partial derivatives of log-likehood function with respect to its parameters. Standard errors of parameters' estimations are the square roots of diagonal elements of variance-covariance matrix divided by $\sqrt{n}$.

The hypothesis H0: $\boldsymbol{\beta}_j = 0$ can be tested by the Wald test.

Provided, the expected counts $\hat{y}_{ij} = n_i \hat{\pi}_{ij}$ are large enough (that is none are below 1 and no more than 20% of the y $\hat{y}_{ij}$'s are below 5) for comparing current model to a saturated model that estimates the probabilities independently for $i = 0,1,\ldots,23$ we can apply the statistics G^2 (deviance)

$$\begin{aligned} G^2 &= LR(\hat{\pi}) = 2\,(L(p) - L(\hat{\pi})) \\ &= 2\sum_{i=0}^{23}\sum_{j=1}^{J} y_{ij}(\ln p_{ij} - \ln \hat{\pi}_{ij}) \\ &= 2\sum_{i=0}^{23}\sum_{j=1}^{J} y_{ij} \ln \frac{p_{ij}}{\hat{\pi}_{ij}} = 2\sum_{i=0}^{23}\sum_{j=1}^{J} y_{ij} \ln \frac{y_{ij}}{n_i \hat{\pi}_{ij}}. \end{aligned}$$

From the last modification we obtain

$$G^2 = 2\sum_{i=0}^{23}\sum_{j=1}^{J} y_{ij} \ln \frac{y_{ij}}{\hat{y}_{ij}}.$$

Thus, the hypothesis H0: $\pi_i j = \hat{\pi}_i j$ can be tested by the LR test. Through the LR test we could compare our estimations $\hat{y}_{ij}$ to y_{ij}.

The saturated model has $24(J-1)$ free parameters and the current model has $k(J-1)$, then the degrees of freedom df are equal to $(24-k)(J-1)$. The statistics G^2 has approximately $\chi^2(df)$ distribution. To compare the estimations $\hat{y}_{ij}$ to y_{ij} can be also applied the Pearson statistics

$$\chi^2 = 2\sum_{i=0}^{23}\sum_{j=1}^{J} r_{ij}^2,$$

where $r_{ij} = \frac{y_{ij} - \hat{y}_{ij}}{\sqrt{\hat{y}_{ij}}}$ is the Pearson residual, having $\chi^2(df)$ distribution, too.

In the given applied area the usage term of the LR test is often interrupted. The examined variable has usually a considerable number of levels that present web parts of portal or system (pages, contents categories, activities etc.).

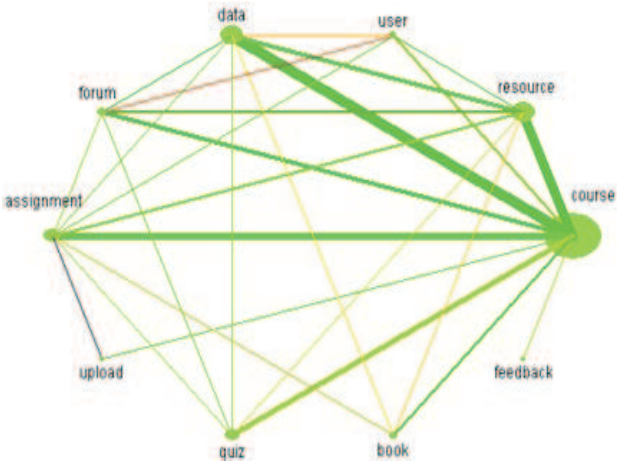

Fig. 3 Web graph - visualization of the found rules.

It has impact on interruption of usage term of LR test that means assumed counts are not large enough. Therefore for the model evaluation we use alternative techniques- visualization of differences of the empirical and theoretical counts, extremes identification, comparison of the distribution of the empirical relative counts of accesses and estimated probabilities of selecting web part j in hour i and the visualization of empirical and theoretical logits for individual web parts except referential [35,36].

4 Model Determination and Parameter Estimation

In the following part the accesses to the course will be described through association rules. The association rules analysis represents a non-sequential approach to the analysed data. Not sequences, but transactions will be analysed, i.e. the analysis will not include time variable. In this case, transaction is a set of course activities accessed by a student during one session.

The web graph (Fig. 3) is a visualization of the determined association rules; in particular the size of the knot shows support of the element, thickness of line shows support of the rule, and brightness of line shows lift of the rule.

From the preceding graph (Fig. 3), which transparently describes selected associations, may be determined that the most frequently accessed course activities include *course* (*support* = 77%), *resource*, *data*, *assignment*, *quiz* (*support*>15%), as well as combinations of the activity course with the other activities (*support* > 10%).

It may also be seen (Fig. 3) that the *assignment* and *upload* occur more frequently together in the sets of the course accessed activities than separately (*lift* = *3.9*). The

same holds true for the activities *forum* and *user* (*lift* = *2.8*). In these cases the highest degree of interest was reached for the lift, determining how more often the accessed activities occur together than if they were statistically independent. If the lift happens to be more than one, the pairs of activities occur more often together than separately in the sets of course activities accessed by students in individual sessions.

On the contrary, the least accessed course activities (Fig. 3) include the *upload* (*support* = 3%), *feedback* (*support* = 2.5%), and the activities *glossary* and *blog* which did not reach even the minimum value of support (*min support* = 1%), as well as combinations *upload*, *assignment* and *feedback*, *course* (*support* = 2%), and the combination *upload*, *course* (*support* = 1%).

Based on the results of the course access analysis, the least accessed activities, i.e. the *upload*, *feedback*, *glossary*, and *blog*, will be excluded from further analyses. We have found out, with respect to contingence

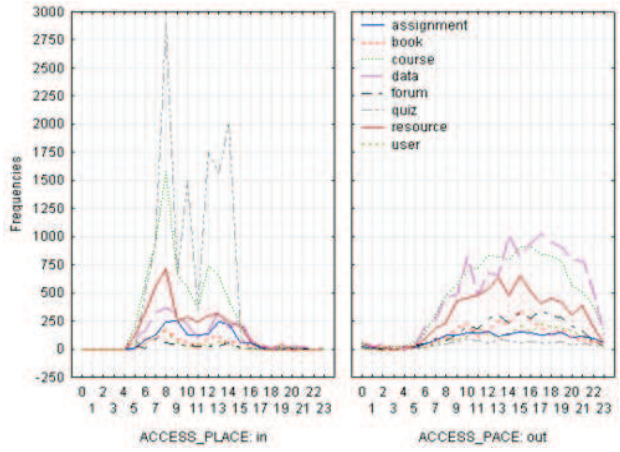

Fig. 4 Interaction plot - ACTIVITY x ACCESS_PLACE x TIME.

tables (Fig. 4), that the significant numbers of accesses for variables ACTIVITY, TIME and ACCESS_PLACE are:

– for internal accesses (*in*) in the hours 5-19, to the activities *assignment*, *book*, *course*, *data*, *forum*, *quiz*, *resource* and *user*,
– for external accesses (*out*) in the hours 0-23, to the activities *assignment*, *book*, *course*, *data*, *forum*, *quiz*, *resource* and *user*.

We have excluded non-significant accesses. The numbers of accesses to those activities are in Table 2.

We have used the following models. For internal accesses:

$$\eta_{ij} = \beta_{0j} + \beta_{1j} TIME_i + \beta_{2j} TIME_i^2 + \beta_{3j} WEEK_i, \quad (4)$$
$$j = 1,2,\ldots,7, i = 5,6,\ldots,19.$$

For external accesses:

$$\eta_{ij} = \beta_{0j} + \beta_{1j} TIME_i + \beta_{2j} TIME_i^2 + \beta_{3j} WEEK_i, \quad (5)$$
$$j = 1,2,\ldots,7, i = 0,1,\ldots,23.$$

Based on *Test of all effects* and *the Likelihood type I test* and *III* tests in the created logit models (Table 3), results present the week statistically significant sign, which is represented with dummy variable WEEK in the models. Hours, represented by variables TIME and their square $TIME^2$, were shown as statistically significant signs in both cases (for internal accesses and external accesses).

The parameters of the models were estimated in the *STATISTICA Generalized Linear/Nonlinear Models*. They are in Table 4. The significance of parameters was tested through Wald test; significant parameters are coloured.

Table 2 The numbers of accesses to activities.

		Activity							
Acces place	Total count	assign-ment	book	course	data	forum	quiz	re-source	user
in	29289	1678	867	7067	2620	412	12305	3668	672
out	42176	2166	2840	11452	12033	3529	959	6994	2203
total	71465	3844	3707	18519	14653	3941	13264	10662	2875

Table 3 Tests of all effects; Likelihood type I test; Likelihood type III test.

	Model	df	Wald Stat.	p	Log-Likelihood	Chi-Square	p	Log-Likelihood	Chi-Square	p
Intercept	in	7	383.9	0.000	-46807.5					
WEEK	in	7	182.9	0.000	-46587.7	439.5	0.000	-45949.3	407.9	0.000
TIME	in	7	843.7	0.000	-46151.6	872.2	0.000	-46215.9	941.0	0.000
TIME_Q	in	7	733.4	0.000	-45745.4	812.5	0.000	-46151.6	812.5	0.000
Intercept	out	7	512.0	0.000	-75568.4					
WEEK	out	7	301.7	0.000	-75415.5	305.9	0.000	-75314.8	315.1	0.000
TIME	out	7	176.2	0.000	-75248.8	333.3	0.000	-75247.5	180.5	0.000
TIME_Q	out	7	179.2	0.000	-75157.2	183.2	0.000	-75248.8	183.2	0.000

Table 4 shows that logits for individual activities are with external accesses significantly dependent on the hour of access as well as on its square and on the dummy variable WEEK, except for the *forum* activity, which does not depend on the hour of access, and, partially, the *book* activity which does not depend on its square.

On the contrary, in the case of internal accesses (from the university net), except for the forum activity, there is no significant dependence on the hour also for the activities *data* and *resource*. The values of these logits, in the case of external accesses, are significantly influenced by the variable WEEK which distinguishes *workweek* and *weekend*. On the other hand, with internal accesses the

Table 4 The parameter estimations.

	ACTIVITY	ACCESS_PLACE		ACTIVITY	ACCESS_PLACE	
		in	out		in	out
Intercept	assignment	-3.2016	2.4554	forum	1.0485	0.1831
TIME	assignment	0.8246	-0.3028	forum	0.1008	0.0510
TIME_Q	assignment	-0.0281	0.0092	forum	-0.0005	-0.0025
WEEK	assignment	-1.0433	-0.2969	forum	-2.4348	0.1764
Intercept	book	-0.5763	1.6841	quiz	-7.2987	0.7888
TIME	book	0.4277	-0.1149	quiz	1.3443	-0.2302
TIME_Q	book	-0.0130	0.0018	quiz	-0.0538	0.0059
WEEK	book	-1.9493	-0.2602	quiz	2.6482	0.4727
Intercept	course	1.0888	2.9225	resource	3.3272	2.4779
TIME	course	0.4691	-0.1378	resource	0.1246	-0.1173
TIME_Q	course	-0.0176	0.0036	resource	-0.0008	0.0023
WEEK	course	-1.4240	-0.1242	resource	-2.7311	-0.2114
Intercept	data	-0.3836	3.1041	user	-	-
TIME	data	0.0885	-0.1702	user	-	-
TIME_Q	data	0.0047	0.0058	user	-	-
WEEK	data	0.4353	-0.3908	user	-	-

WEEK variable has a significant influence on the value of logits only for the activities *forum*, *quiz* and *resource*.

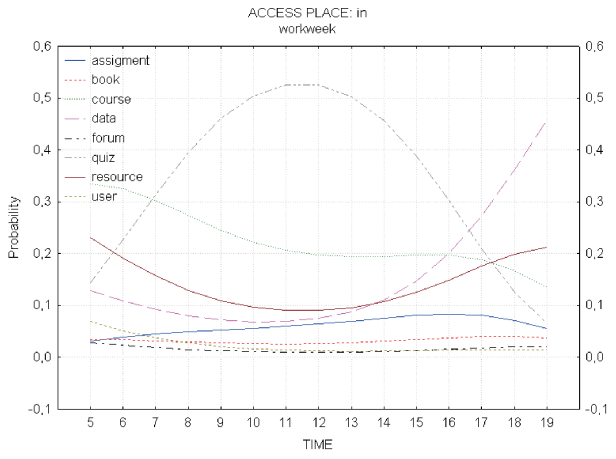

Fig. 5 Probabilities of accesses to the activities during the workweek for internal accesses.

Using the estimated parameters it is possible to enumerate the estimates of logits and the probabilities of selecting individual categories in any given day's hour. Then theoretical numbers of accesses to individual categories can be enumerated. This is proved by some illustrations. Graphs of the probabilities of selection of activities during the workweek (Fig. 5 and Fig. 6) and weekend (Fig. 7 and Fig. 8) by students who study at the university (in the classrooms/study halls) are in the Fig. 5 and Fig. 7 and the Fig. 6 and Fig. 8 obtain these probabilities for students who study outside the university (at work/home).

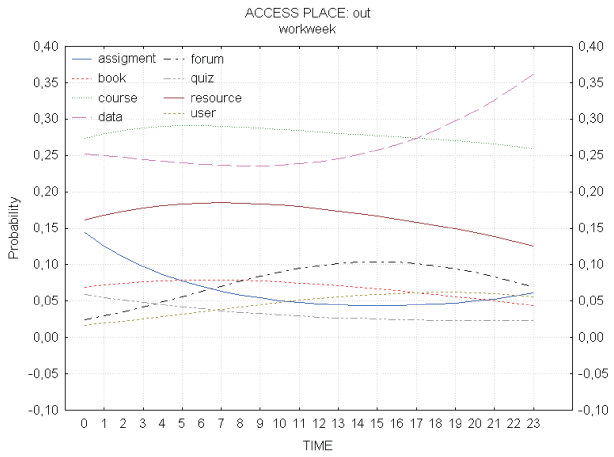

Fig. 6 Probabilities of accesses to the activities during the workweek for external accesses.

As for the accesses from the university net (internal accesses), the higher probability of selection during the workweek (Fig. 5) is reached by the activities *quiz*, *data*, *course* and *resource*. The probability of selecting quiz is small in the morning 0.14 (5.00 a.m.), then it rises, reaching around noon values higher than 0.5, and during the afternoon going down to the value of 0.06 (7.00 p.m.).

The activity *resource* shows -an inverse course; the probability of selecting *resource* culminates in the morning and in the evening ($\pi > 0.2$), and during the day goes down ($\pi < 0.1$). A different development can be observed with the activity *data* where the probability of its selection is small during the morning ($\pi < 0.1$), but it rises during the afternoon, reaching the value of 0.46 (7.00 p.m.). The development exactly opposite to the data activity can be observed with the *course* activity where the probability of its selection culminates in the morning 0.33 (5.00 p.m.) and during the day goes down to the value of 0.14 (7.00 p.m.).

As during the workweek, also during the weekend (Fig. 7) the activities *resource* and *course* reach the highest value of selection probability. With the *course* activity, the probability of selection culminates, like during the workweek, in morning hours ($\pi > 0.3$) and during the day goes down to the value of 0.12 (7.00 p.m.). Approximately the same course as in the workweek can be observed also for the *resource* activity, but with higher values of the probability of selection. The probability of selecting the activity *resource* culminates in the morning ($\pi > 0.61$) and in the evening ($\pi > 0.67$), going down during the day ($\pi < 0.5$).

The accesses from the net outside the university (external accesses) (Fig. 6; Fig. 8) show the highest

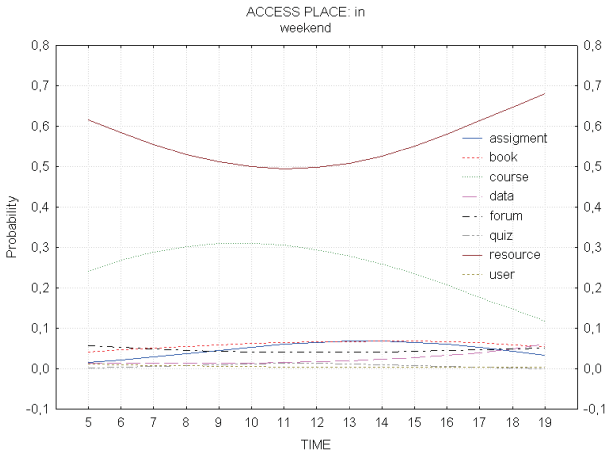

Fig. 7 Probabilities of accesses to the activities during the weekend for internal accesses.

probability of selection for the activities *data*, *course*, *resource* and *assignment*. The data activity shows little probability of selection during the day, but during the evening it rises to the value of 0.36 (11.00 p.m.) and during the weekend even to the value of 0.43 (11.00 p.m.).

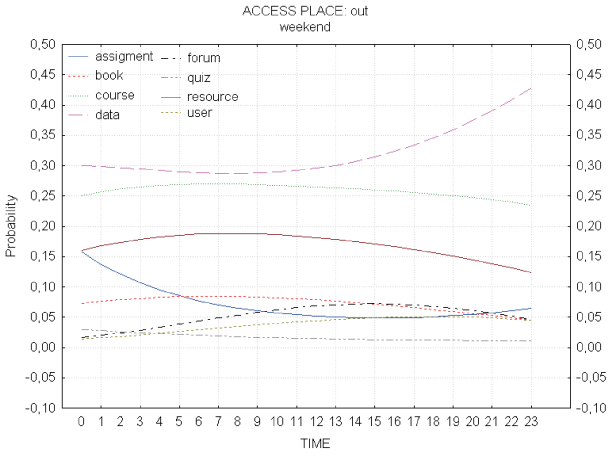

Fig. 8 Probabilities of accesses to the activities during the weekend for external accesses.

An opposite development can be observed in the activity *assignment* where the probability of selection culminates at night 0.15 (0.00) and goes down during the day. Probabilities of selecting the activities *course* ($\pi < 0.27$) and *resource* ($\pi < 0.17$) are more stable, reaching approximately the same values during the whole week.

5 Evaluation of the Model

The fitness of our models cannot be verified by the deviance, because we get many expected counts equal to zero. We have provided this in three ways.

First, we compare the probability distribution of the number of empirical accesses to the probability distribution of number of expected accesses to the activities using the Wilcoxon matched pairs test. The test values were not significant for internal and external accesses for all activities as during the workweek, also during the weekend. We can conclude that there is no significant difference between the empirical and the expected counts.

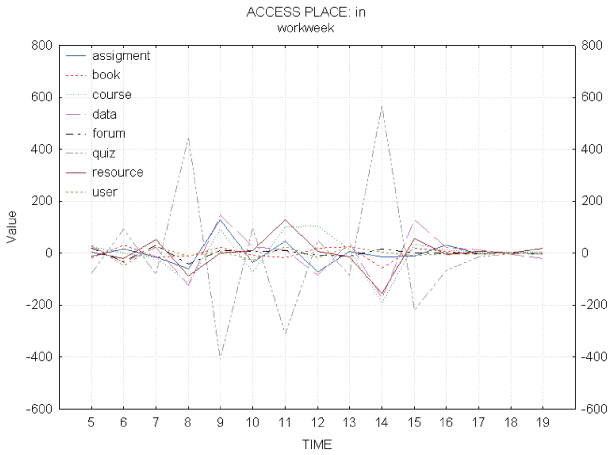

Fig. 9 Differences of empirical and theoretical frequencies of accesses to the activities during the workweek for internal accesses.

Second, we draw the differences of empirical and theoretical frequencies of accesses to the activities during the workweek for internal accesses and for external accesses (Figs. 9-10.). All means of those differences are equal to zero. Some differences are seen to be large, but if we use ”rule of two-times standard deviation” only a few values are outliers (one or two for some activities). Similar results were obtained for the weekend.

Third, we draw the empirical and fitted logits (without the reference category user).

Fig. 11 obtains empirical and theoretical logits during the workweek for selected activities with the highest probabilities for internal accesses and for external accesses. We cannot evaluate the empirical logits in every time i, since some counts equal zero. Similar results were obtained for the weekend (Fig.12).

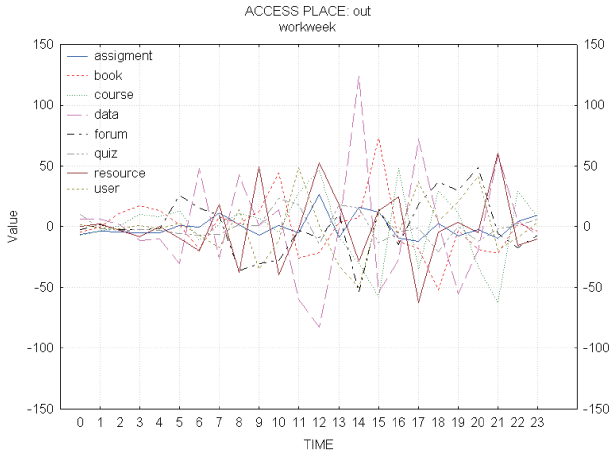

Fig. 10 Differences of empirical and theoretical frequencies of accesses to the activities during the workweek for external accesses.

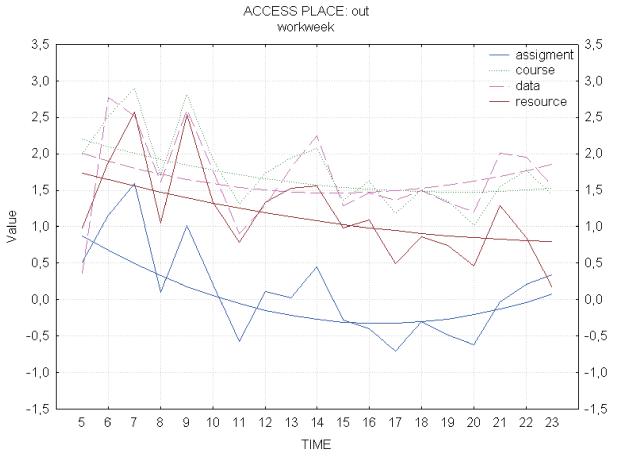

Fig. 12 Empirical and theoretical logits during the workweek for external accesses.

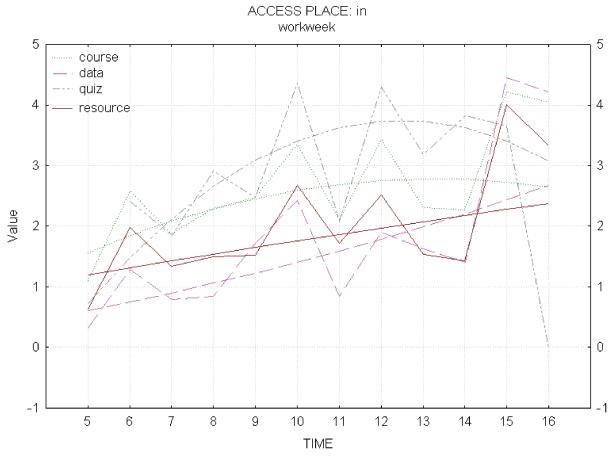

Fig. 11 Empirical and theoretical logits during the workweek for internal accesses.

6 Discussion and Conclusions

The graph (workweek, in; Fig. 5) depicts probabilities of accesses to individual course activities during the day from computers within the university. It confirms several assumptions. At first glance it is evident that the probability of using the quiz activity is highest during the workweek. This fact is a result of requirements put on a test - to ensure objectivity, testing is done in a computer room at the university and under the supervision of an instructor. Further expected development shows the probability of using the book activity, providing on-line study material, as well as discussion forums. Since discussion forums were used for asynchronous communication in the course and for the dealing with student problems, an almost identical probability of their use during the day is not surprise at all.

The graph also shows several interesting facts. First, it is surprising that some part of students accesses e-learning course in early morning hours (course curve), before the first lesson. This probability goes gradually down during the day, from which it may be concluded that after lessons students do not use the university computer rooms to do their assignments, but rather leave the university premises and access e-learning course from their own computers.

Similarly interesting is the highest probability of accessing study resources in the form of off-line materials, downloadable files, etc. (resource) outside direct teaching time, be it in the morning or in the afternoon hours. The last activity with an interesting development during the day is the data. This includes rather passive activities of students, such as browsing through achieved results and the content of the course - checking grades, test results, achieved points and submitted assignments, checking student resources based on their type, etc. From the graph follows that students are concerned with these activities in later hours of the workday. Other categories of activities are used by students during the work day with the probability lesser than 0.1, therefore their analysis, in our opinion, does not have a practical sense either for the teacher or for the author of the e-learning course.

Putting into relation the graphs depicting probabilities of using activities from within the university net and from without it, during a workday, it is possible to acquire a general view of student behaviour in the e-learning course.

It can be noted that the probability of passive browsing through a course content, that is, the activities included in the data category, is highest during evening hours (0.4) as well as late evening hours (0.35) (workweek, out; Fig. 6).

This finding corresponds with the probability of a ”brief look” into the course itself (course).

With regard to this probability we were surprised by the fact that the probability of accessing the e-learning course during the workday from outside the university net is almost the same, being the highest in early morning hours. A probable reason for this is the students way of study, submission of assignment or preparation for continuous tests at the last moment, etc.

Our assertion is partially documented also by the probability of using the assignment activity. As we can see in the graphs, the highest probability of a students submission of assignment is around 7 p.m., if he/she is still on the university premises, or then late at night and in early morning hours.

Taking a more detailed look at students work with course study materials (resource), one can see that it is the use of study resources which is the most frequent activity performed by students accessing the e-learning course from outside the university net in late evening and early morning hours. This is confirmed also by our previous explanation of the unexpectedly high probability of accessing e-learning course in early morning hours.

Relatively surprising is also the following assertion: during the workday there is a higher probability of students entering the e-learning course from the outside than from the university net. This fact suggests that in the time between the in-class teaching students leave premises of the university, or they are not present at lessons, despite the fact that they are mostly obligatory. This is an indirect proof that the University has reserves in the creation of conditions and providing complementary services for study within the obligatory activities of students, therefore students leave its premises.

We will provide a short summary of our findings related to student behavior in the monitored e-learning course during weekends. At the beginning it has to be stressed that teaching rules had a direct influence on final probabilities and on their statistical significance. During weekends there was no teaching on the university premises, which were not freely accessible, and this was reflected in the accesses to the course as well. From the graph (weekend, in; Fig. 7) it may be concluded that during the weekend students signed into the course only to take a look or to download study materials.

More relevant findings may be read from the graph ”weekend, out” (Fig. 8). As with workdays, also during weekends many students access the e-learning course to find out whether there are some complementary materials, tasks or information related to the study (course, data). Their activity is higher in evening hours. From the point of view of the teacher, however, we can claim that in this case it is not an active study.

The probability that during the weekend students look for study materials (resource) is, during the whole day, higher than 0.1, reaching the maximum in the morning hours. A similarity with the curve characterizing the probability of using on-line sources (book) shows that looking for study materials students check this type of study materials as well. Higher values of probability for the category resource (and partly also the book) are connected with the development of the probability of the assignment. As it was said in the discussion regarding the results of study during workdays, it appears that students are willing to work into early morning hours and during weekends if deadline for the submission of assignment is approaching.

The multinomial logit model (MLM) used to assess the probability with which students access individual activities of the e-learning course in individual parts of the day finds its application in various stages of the courses life cycle, as well as for various roles of LMS users.

It is evident that structure of the e-learning course itself predetermines probability of using individual activities in a particular time of the day, or week. Therefore the greatest contribution does not lie in the process of designing a new course, but rather in the process of restructuring the existing e-learning course. Based on the determined probabilities, the author of the course content may change original representation of individual categories of activities in the course according to the student behaviour. If some activity seems to be used only with little probability by students, we can conclude that this way of study was not attractive for them. In such case we should search for alternative activities with the same or similar teaching aim.

The students work with study materials and information sources of the course outside the time of their daily study as it was mentioned in the discussion. Therefore we can say that the results of using MLM may influence not only the structure and content of the e-learning course itself, but also the structure of teaching lessons, e.g. in the combined (blended learning) form of study. The teacher may focus in these lessons on the development of soft skills, use non-traditional forms of work, etc., ensuring the subject content by selecting suitable resources and activities of the course [37, 38, 39, 40].

Further use of results acquired through the application of MLM can be found in the personalization of the course content - not the content for every user, but rather personalization based on stereotyped users. If, for example, the e-learning course happens to be used simultaneously for internal and external forms of study, it is possible to provide students, based on various probabilities with which they access the course activities at various times and from various places, with their combination as well as with obligatory deadlines for the submission of tasks, assignments, or projects.

Drawing on the determined probabilities, the teacher may, provided that it is allowed by the environment of an individual LMS, adjust teaching management to the needs, or rather habits, of the student target group. It is mainly the selection of suitable deadlines, either for the submission of

assignments and tasks, participation in discussion forums, or the realization of tests and self-tests.

This recommendation is also connected with the work of the LMS administrator. From the aspect of the LMS administrator the results acquired through MLM may be used in planning the LMS maintenance. In the discussion we offered a surprising finding that students access the e-learning course with approximately the same probability during the whole day, including the early morning hours. This fact should be taken into consideration during the planning of the LMS maintenance, backing up courses and ensuring an overall accessibility of the system. The system should not be inaccessible when the deadline for the submission of assignments is approaching. Naturally, in searching for a suitable schedule for administrator activities a certain compromise should be attempted and teachers should synchronize deadlines for the submission of assignments in their e-learning courses in order to determine a period with minimum accesses.

Acknowledgements This paper is supported by the project VEGA 1/0392/13 Modelling of Stakeholders' Behaviour in Commercial Bank during the Recent Financial Crisis and Expectations of Basel Regulations under Pillar 3- Market Discipline.

References

[1] J. Anděl, Základy matematické štatistiky. MATFYZPRESS, Praha, (2007).
[2] M. Munk, Počítačová analýza dát, Nitra, (2011).
[3] G. Rodríguez, Generalized linear models, (2011).
[4] D. W. Hosmer, S. Lemeshow, Applied logistic regression. John Wiley and Sons, Inc., (2000).
[5] Rassoul Abdelaziz, Reduced Bias Estimation of the Reinsurance Premium of Loss Distribution, Journal of Statistics Applications & Probability, **1**, 147-155 (2012).
[6] B. D. Baltagi, Econometrics. Springer Verlag, Berlin Heidelberg, (2008).
[7] L. P. Macfadyen, S. Dawson, Mining LMS data to develop an "early warning system" for educators: A proof of concept. Comput. Educ., **54**, 588-599 (2009).
[8] L. S. Stratton, J. N. Wetzel, A multinomial logit model of college stopout and dropout behavior. Economics of Education Review, **27**, 319-331 (2008).
[9] J. M. Domenech, J. Lorenzo, A tool for web usage mining. Proceedings of the 8th international conference on Intelligent data engineering and automated learning. Springer-Verlag, Birmingham, UK, (2007).
[10] M. Drlík, M. Munk, J., Skalka, Usage analysis of system for theses acquisition and plagiarism detection. Procedia Computer Science, (2010)
[11] V. Chitraa, A. S. Davamani, A survey on preprocessing methods for web usage data. International Journal of Computer Science and Information Security, **7**, (2010).
[12] O. R. Zaine, M. Xin, J. Han, Discovering web access patterns and trends by applying OLAP and data mining technology on web logs. Proceedings of the Advances in Digital Libraries Conference. IEEE Computer Society, (1998).
[13] E. Mor, J. Minguillon, E-learning personalization based on itineraries and long-term navigational behavior. Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters. ACM, New York, NY, USA, (2004).
[14] W. Wei, W. Jui-Feng, S. Jun-Ming, T. Shian-Shyong, Learning portfolio analysis and mining in SCORM compliant environment. Frontiers in Education, 2004. FIE 2004. 34th Annual, **11**, 17-24 (2004).
[15] L. Talavera, E. Gaudioso, Mining student data to characterize similar behavior groups in unstructured collaboration spaces. Workshop on Artificial Intelligence in CSCL, Valencia, Spain, 17-23 (2004).
[16] A. A. Ramli, Web usage mining using apriori algorithm: UUM learning care portal case. International Conference on Knowledge Management, Malaysia, 1-19 (2005).
[17] Khine Khine Su-Myat, Jules J. S. de Tibeiro, Pranesh Kumar, An Integrated Approach to Regression Analysis in Multiple Correspondence Analysis and Copula Based Models, Journal of Statistics Applications & Probability, **1**, 1-21 (2012).
[18] Ilyas Saleem, Hamza Nawaz, Istaqlal Ahmed, S. Muzahir Abbas, Analytical Evaluation of Tri-band Printed Antenna, Information Sciences Letters, **1**, 85-89 (2012).
[19] Mary Iwundu and Polycarp Chigbu, A Hill-Climbing Combinatorial Algorithm for Constructing N-Point D-Optimal Exact Designs, **1**, 133-146 (2012).
[20] Tim Berners-Lee, James Hendler and Ora Lassila, The Semantic Web, Scientific American, (2001).
[21] M. Munk, M. Drlík, J. Kapusta, D. Munková, Methodology Design for Data Preparation in the Process of Discovering Patterns of Web Users Behaviour. Applied Mathematics and Information Sciences, **7**, 27-36 (2013).
[22] G. T. Raju, P. S. Satyanarayana, Knowledge discovery from web usage data: a complete preprocessing methodology. IJCSNS International Journal of Computer Science and Network Security, **8**, (2008).
[23] C. Romero, S. Ventura, Educational data mining: A survey from 1995 to 2005. Expert Systems with Applications, **33**, 135-146 (2007).
[24] E. Gaudioso, L. Talavera, Data mining to support tutoring in virtual learning communities: Experiences and challenges. Data mining in e-learning. Wit Press, Southampton, 207-226 (2006).
[25] H. Ba-Omar, I. Petrounias, F., Anwar, A Framework for using web usage mining to personalise e-learning. Advanced Learning Technologies. ICALT 2007. Seventh IEEE International Conference on, 937-938 (2007).
[26] R. Cooley, B. Mobasher, J. Srivastava, Data preparation for mining world wide web browsing patterns. Knowledge and Information Systems, **1**, 5-32 (1999).
[27] H. Zhang, W. Liang, An intelligent algorithm of data pre-processing in web usage mining. Proceedings of the World Congress on Intelligent Control and Automation, 3119-3123 (2004).
[28] M. A. Bayir, I. H. Toroslu, A. Cosar, A new approach for reactive web usage data processing. Data Engineering Workshops, 2006. Proceedings. 22nd International Conference on. (2006).

[29] L. Yan, F. Boqin, M. Qinjiao, Research on path completion technique in web usage mining. Computer Science and Computational Technology. ISCSCT '08. International Symposium on, **1**, 554-559 (2008).
[30] M. Spiliopoulou, B. Mobasher, B. Berendt, M. Nakagawa, A framework for the evaluation of session reconstruction heuristics in web-usage analysis. INFORMS J. on Computing, **15**, 171-190 (2003).
[31] M. Munk, M. Drlík, Influence of different session timeouts thresholds on results of sequence rule analysis in educational data mining. DICTAP 2011. Communications in Computer and Information Science, **166**, 60-74 (2011).
[32] L. Yan, F. Boqin, The construction of transactions for web usage mining. Computational Intelligence and Natural Computing, CINC '09. International Conference on, **1**, 121-124 (2009).
[33] M. Munk, J. Kapusta, P. Švec, Data preprocessing evaluation for web log mining: reconstruction of activities of a web visitor. Procedia Computer Science, **1**, 2273-2280 (2010).
[34] K. W. Wober, Benchmarking in tourism and hospitality industries: The selection of benchmarking partners. CABI., **2002**.
[35] M. Munk, M. Vrábelová, J. Kapusta, Probability modeling of accesses to the web parts of portal. Procedia Computer Science. Elsevier, **3**, 677-683 (2011).
[36] M. Munk, M. Drlík, M. Vrábelová. Probability modeling of accesses to the course activities in the web-based educational system. Lecture Notes in Computer Science. Springer, **6786**, 485-499 (2011).
[37] D. Klocoková, Integration of heuristics elements in the web-based learning environment: experimental evaluation and usage analysis. Procedia Social and Behavioral Sciences, **15**, (2011).
[38] M. Vozár, P. Kuna, Innovative education with e-learning support in European project. 10th IEEE International Conference on Emerging eLearning Technologies and Applications (ICETA 2012). Stara Lesna, Slovakia, 395-399 (2012).
[39] M. Vozár, Use of computer technique in teaching of optimization. 6th WSEAS International Conference on Education and Educational Technology (EDU'07), Venice, Italy, 279-282 (2007).
[40] Z. Balogh, S. Koprda, Modeling of Control in Educational Process by LMS. 9th International Scientific Conference on Distance Learning in Applied Informatics (DIVAI 2012), Sturovo, Slovakia, 43-51 (2012).

Michal Munk received the PhD degree in Mathematics from the Department of Mathematics at Constantine the Philosopher University in Nitra, in 2007. He is currently an associate professor at the Computer Science Department of Constantine the Philosopher University in Nitra. His research interests are in the areas of web log mining and user behaviour modelling.

Martin Drlík received the PhD degree in Computer Science from the Department of Computer Science at Constantine the Philosopher University in Nitra, in 2009. He is currently an assistant professor at the Computer Science Department of Constantine the Philosopher University in Nitra. His research interests are in the areas of web log mining, educational data mining and learning analytics.