Yavuz UNAL, Recep CAKIR,
Amasya University Education Faculty
Computer Education and Instructional Technology
Amasya Turkey

Using the Data from Learning Management System in Educational Data Mining (Amasya University Moodle Sample)

Abstract

Extracting meaningful results from educational data and using these results in reorganizing education is called educational data mining. In this study, students studying at Amasya University, Faculty of Education, Department of Computer education and Instructional Technology, some of whose undergraduate courses are instructed in accordance with Moodle system, have been analyzed by means of data mining. For this purpose, the log files received from Moodle system and the database files of Moodle system were extracted using SQL queries, and clustering, association rule mining analyses were made with the help of the WEKA program and Microsoft data mining tools software. The obtained results and recommendations have been expressed.

Keywords: *E-Learning, Course management system, Data mining, Moodle system*

INTRODUCTION

With the swift transition from an industrial society to information society, the unbelievable developments in Information Technologies have made human beings come closer to each other thus causing the world to become a global village. The cheap availability of Information Technologies and the Internet, together with personal computers, slowly made it possible for human beings to reach any kind of information anywhere any time (Şeker, 2007).

This was followed by the use of personal computers and the Internet in educational field. One of the most important steps in internet-based education is the use of Teaching Management Systems (Önal, Kaya, & Draman, 2006).

The Learning Management System (LMS) software is a platform which enables the web-based application of teaching activities and various interactions of them that were formed as target-focused teaching activities. With the widespread use of Learning Management Systems (LMS), it has become obvious that these systems can be used in formal education as well as in distant education systems (Önal, Kaya, & Draman, 2006).

Moodle is an internet-based class and website designing software package. Moodle is used in 138 countries and supports 77 languages and has 75000 registered users. Moodle is an online php based education system with open source codes. Classes are established in modules. It provides support for Linux, Unix, Windows and Mac OS X operation systems. The demo version and help system are available on the Internet. It is used with GPL license and managed with portal logic. It provides SSL, TSL support; and since plugins are loaded in modules, it is a very easy system for people who have portal management experience (Elmas, Doğan, Biroğul, & Koç, 2008).

In this study, the data received from the Moodle system are analyzed together with association rules and clusters from data mining algorithms. The results obtained will provide opinions for designing future Moodle activities and for making the learning of students easier. The findings are presented in the conclusion and assessment section .

Datamining

Datamining is defined as the extraction of the potentially useful information which is not previously known and not very clear from among the data in wide databases (Özkan, 2008). In this study, the association rules from data mining algorithms and clustering have been applied to the data received in the study and thus the final results have been obtained (Kaya & Köymen, 2008).

METHOD

Apriori and clustering from data mining algorithms have been applied to the e-learning data received from the Moodle in the study. The purpose was to determine how the class or classes could be better conducted by examining the activities of the students over Moodle. This study has been conducted with the undergraduate students from Amasya University, Education Faculty, Computer Education and Instructional Technology Department in 2012-2013 Academic year. The data received from 51 students in classes over Moodle were used in this study.

Pre-processing Moodle Data

The following steps were used during the datamining application.

- The course usage data of the students were received from the Moodle system. These data were obtained from the database files over the Moodle system.
- Pre-processing were applied to make the data processable. The data were cleared and were converted into the proper format.
- Datamining algorithms were applied.
- The results obtained were used to increase the learning of the students in classes.

MySQL interface was used while the data were being extracted from the Moodle system. The MySQL interface is given in Figure 1. The data in more than one table in the database in a complex structure were extracted with the use of SQL questions and then cleared and converted into the proper format.

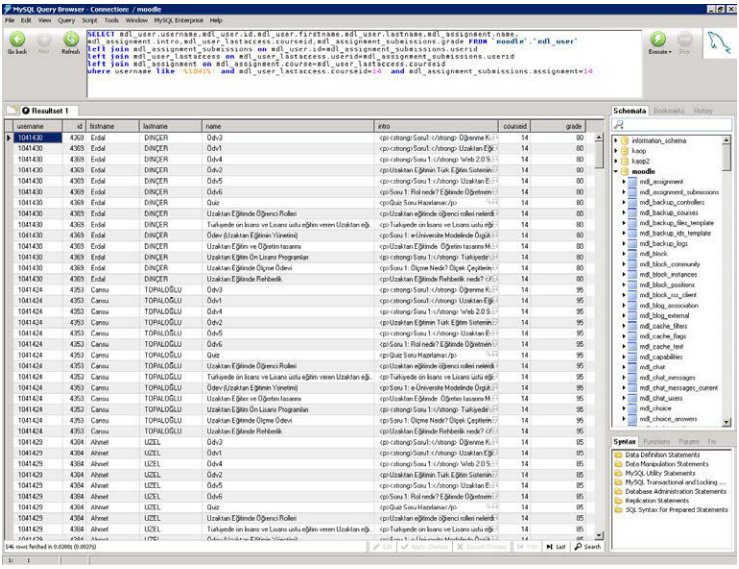

Figure 1. The MySQL screen where the data are chosen by making query from the

One of the sample SQL queries used during the data extraction from the databases is given below.

```
SELECT mdl_user.username,mdl_user.firstname,mdl_user.lastname,mdl_user.id,mdl_files.filename,mdl_user_lastaccess.courseid,mdl_assignment.name,mdl_assignment.intro FROM `moodle`.`mdl_user`

left join mdl_files on mdl_files.userid=mdl_user.id

left join mdl_user_lastaccess on mdl_user_lastaccess.userid=mdl_user.id

left join mdl_assignment on mdl_assignment.course=mdl_user_lastaccess.courseid
```

The data which are extracted from the Moodle database and cleaned with query in MySQL program are converted into .arff format to be analyzed in WEKA program (Figure 2).

```
@relation education

@attribute asm1 {low,medium,high}
@attribute asm2 {low,medium,high}
@attribute asm3 {low,medium,high}
@attribute asm4 {low,medium,high}
@attribute asm5 {low,medium,high}
@attribute asm6 {low,medium,high}
@attribute asm7 {low,medium,high}
@attribute asm8 {low,medium,high}
@attribute asm9 {low,medium,high}
@attribute asm10 {low,medium,high}
@attribute asm11 {low,medium,high}
@attribute asm12 {low,medium,high}
@attribute asm13{low,medium,high}
@attribute quiz {low,medium,high}
@attribute course {low,medium,high}
@attribute blog {low,medium,high}
@attribute forum {low,medium,high}
@attribute quiznumber {low,medium,high}
@attribute resource {low,medium,high}
@attribute midterm {low,medium,high}
@attribute final {low,medium,high}
@attribute pass {fail,middle,success}
```

Figure 2. The arff format used in the study

Applying datamining algorithm to Moodle Data

We analyzed the data from the Moodle system with the help of WEKA datamining program in our study. WEKA (Waikato Environment for Knowledge Analysis) is a data mining and machine learning software developed in New Zealand Waikato University. WEKA software has been developed with Java, which is one of the many programming languages directed to the object. Java provides a smooth platform for various learning algorithms. The most powerful side of WEKA is that it includes many classification techniques. Another property is that it enables the realization of the applications by entering a command. There are Preprocess, Classify, Cluster, Associate, Select Attribute and Visualization panels in WEKA (Önal, Kaya, & Draman, 2006). Attributes used in the study are also shown in Table 1.

Table 1

Attributes used for each student

Course	**Identification number of the course**
Assignment 1	The grade received from the Assignment 1
Assignment 2	The grade received from the Assignment 2
Assignment 3	The grade received from the Assignment 3
Assignment 4	The grade received from the Assignment 4
Assignment 5	The grade received from the Assignment 5
Assignment 6	The grade received from the Assignment 6
Assignment 7	The grade received from the Assignment 7
Assignment 8	The grade received from the Assignment 8
Assignment 9	The grade received from the Assignment 9
Assignment 10	The grade received from the Assignment 10
Assignment 11	The grade received from the Assignment 11
Assignment 12	The grade received from the Assignment 12
Assignment 13	The grade received from the Assignment 13
quiz_number	Number of Quizzes
Quiz	The grade received from the Quiz
Blog	Number of the published items in the blog
Forum	Number of the written messages in the forum
Resource	Number of the reading the added class material
Midterm	The grade of the Midterm
Final	The grade of the Final
passing grade	Pass status according to the Visa and Final exams

Clustering

Clustering analysis is a collection of methods that is helpful in separating the units in the X data matrix whose groupings are not known; and distributing the variables into sub-clusters (group, class) that are similar to each other. Clustering analysis is used to define certain prototypes and to divide the units into homogenous groups by making use of some criteria calculated based on differences or similarities between the variables (Kudyba, 2004).

Another definition for clustering analysis is like this: “Clustering analysis is a group of multi-variable techniques whose basic aim is to group objects (units) according to their characteristic properties (Romero, Ventura, & García, 2008). Clustering analysis clusters objects in similar ways but so as that become distinctive from each other. If the clustering is successful, and if a geometrical drawing is made, the objects will be very close to each other, and the clusters will be far away from each other (Hair et al., 1995). They are used in e-learning to determine the students with similar learning characteristics.

We used simple Kmeans from clustering algorithms in WEKA program. The WEKA clustering screen is given in Figure 3. The parameters used are as follows: number of cluster: 2 and max. iteration: 500. When the classification algorithm was applied, the student group was divided into two clusters. The characteristics of the students in the clusters are given below:

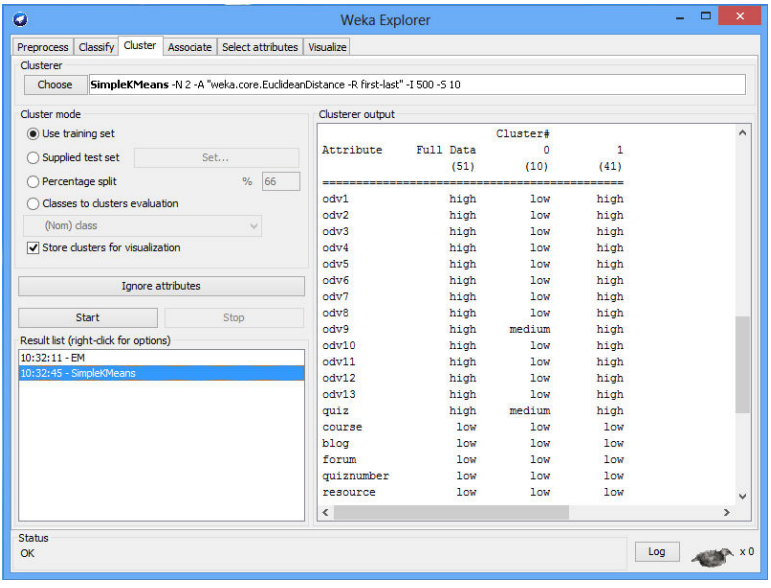

Figure 3. Execution of clustering algorithm in WEKA

The first cluster consists of 10 people and shows that the homework grades, quiz grades, messages sent to the forum and the resources read are relatively lower. Furthermore, the passing status of the students in Cluster 1 is lower. On the other hand, Cluster 2 consists of 41 students and shows that the homework grades, quiz grades, messages sent to the forum and the resources read are higher. Furthermore, the passing status of the students in Cluster 2 is higher. To sum up, it can be suggested that Cluster 1 has used the Moodle system less and become unsuccessful; and Cluster 2 has used the Moodle system more and become successful. After the analyses it is suggested that unsuccessful students should form a separate group and be motivated for the classes more. They should be encouraged to follow the documents in Moodle. The unsuccessful students can be given extra homework and made to learn easier.

Association Rule Mining

The first technique used in datamining is the association rule (Agrawal et al., 1994). The association rule is an approach that enables the analysis of past data and determine the association behaviors in these data thus supporting future studies. “Sunday basket” analysis can be suggested as an example for the association rule datamining application (Frawley et al., 1991). The purpose in association rule is finding the association relationship among the purchased products during shopping and determining the purchase habits of the customers using these relationship data (Kudyba, 2004). Sellers have the opportunity of efficient marketing and selling their products by using the discovered association relationships. For example, customers bought yogurt together with milk and cheese from a market in 70% of the shopping samples. To determine such an association relationship, the products in the pattern should be together in more than one purchase activity. When datamining techniques are applied on millions of data, the algorithms used for association query must be fast (Agrawal and Srikant, 1995).

The name of the Apriori Algorithm is Apriori because it takes the data from the previous step that is the “prior” step (Agrawal and Srikant, 1994). This algorithm has an iterative (repetitive) quality in its basis (Han and Kamber, 2006) and is used in discovering clusters that are frequent in databases which contain movement information. According to the Apriori Algorithm, if the k-elements cluster (the cluster having k number of elements) meets the support criterion, the sub-clusters of this cluster also meet the minimum support criterion.

The association rule mining consists of two steps which are finding all the frequent elements and producing strong association rules from these elements. The Apriori Algorithm used for the first step of the association rule is the most popular and classical algorithm used in frequent elements mining. The characteristics and data in this algorithm are assessed with Boolean association rules (Romero, Ventura, & García, 2008).

The association rules are applied to web-based education systems (Zaiane, 2002).

We used Apriori algorithm which is among association algorithms in the WEKA program. The WEKA Apriori screen in given in Figure 4. The parameters used are as follows: minimum support: 0.3 and minimum confidence 0.9. We determined many association rules with this algorithm. Some of these rules are explained as follows:

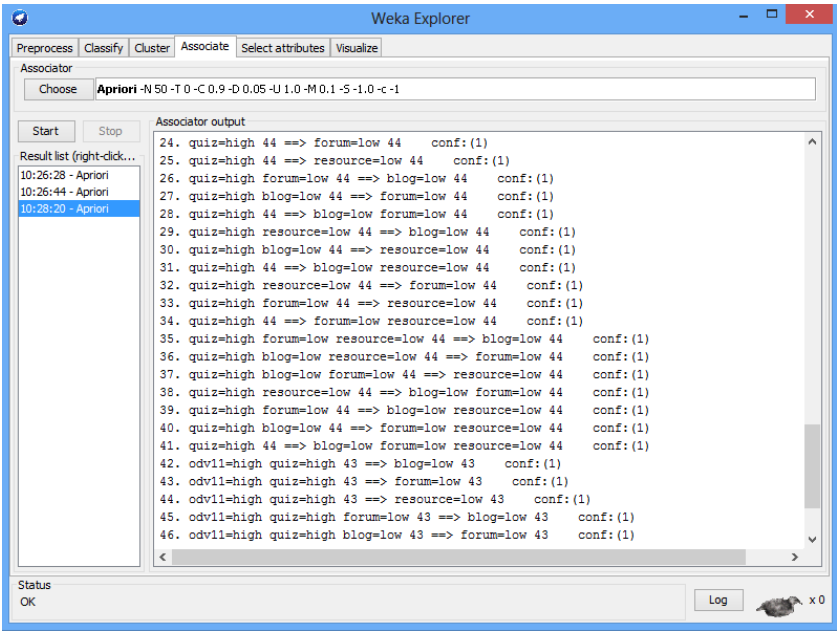

Figure 4. Execution of Apriori algorithm with WEKA

Many association rules are obtained as shown in Figure 4. Some of these rules are interesting and some are not. If we consider some of them which can be used for educational purposes;

If the student has read many resources and has followed class materials and has received high grades from homework, it means that s/he has had a passing grade. In another rule, it is observed that if a student has not fulfilled the assigned practices, s/he has had a failing grade.

By considering these rules, it is concluded that the teacher should make the students follow the resources more; and give extra assignments for unsuccessful students if necessary. Furthermore, the teacher should encourage students to spend more time in the Moodle system. Since students use the forums and blogs less than the expected rates, they should be encouraged to use them more.

Conclusions

The educational data from the Moodle System, which is one of the course management systems (Learning Management Systems), have been received and subjected to datamining application. By using association rules and clustering algorithms, normally invisible results have been revealed. The analysis results of datamining will be used to increase learning skills of students.

Association rules and clustering algorithms of datamining algorithms have been applied to the data received from the Learning Management System Moodle. The other algorithms of datamining can be applied to e-learning data in further studies. Furthermore, similar studies can be conducted with the data received from other Learning management systems. It is obvious that the results of the data mining applications can be used in better teaching activities and in increasing the learning skills of students.

Acknowledgement
This study is supported by the Scientific Research Projects (BAP) of Amasya University.

REFERENCES

Agrawal, R., & Srikant, R. (1994). Fast Algorithms for Mining Association Rules. VLDB.

Agrawal, R., & Srikant, R. (1995). Mining Sequential Patterns. VLDB.

Elmas, Ç., Doğan, N., Biroğul, S., & Koç, M. S. (2008). Moodle Eğitim Yönetim Sistemi İle Örnek Bir Dersin Uzaktan Eğitim Uygulaması. BİLİŞİM TEKNOLOJİLERİ DERGİS, 53-62.

Hair, J. F., Black, W., Babin, B., & Anderson, R. (2009). Multivariate Data Analysis. Macmillan Publishing Company: Newyork.

Han, J., & Kamber, M. (2001). Data Mining Consepts and Tecniques. The Morgan Kaufmann Series in Data Management Systems.

Kaya, H., & Köymen, K. (2008). Veri madenciliği kavramı ve uygulama alanları. Doğu Anadolu Araştırmaları, 159,164.

Kudyba, S. (2004). Managing Data Mining. USA: CyberTech Publishing.

Önal, A., Kaya, A., & Draman, S. (2006). Açık Kaynak Kodlu Çevrimiçi Eğitim Yazılımları. Akademik Bilişim . Denizli.

Özkan, Y. (2008). Veri Madenciliği Yöntemleri. İstanbul: Papatya Yayıncılık.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management-systems: Moodle case study and tutorial. COMPUTERS & EDUCATION , 368-384.

Srikant, R., & Agrawal , R. (1995). Mining Generalized Assocaition Rules. VLBD.

Şeker, T. B. (2007). İnternet ve Bilgi Çağı. İstanbul: Çizgi Kitabevi Yayınları.

Zaiane, O. R. (2002). Building a recommender agent for e-learning systems. International Conference on Computers in Education, (s. 55-59).