

Clustering Moodle Data As a Tool For Profiling Students

Gawali Kavita Bapu
More Bharati Ashok
Salunke Priyanka Shamrao
Sul Gauri Tanaji

Mrs. Pratiksha R. Deshmukh
(Guide)

Department of Computer engineering, Government college of engineering and research,
Awassari(kd), Ambegaon, pune, India

Abstract

We wish to create system that will gather information related to e-learning courses and performing analysis on the data. In this system analysis, clustering and prediction on data is performed to generate results, which makes monitoring data easier. The aim of our project is to use the methods of data mining, machine learning to given students better tools to help their e-learning courses. In this system processing of Moodle data is to be carried out. This project includes online tests, quizzes, displaying notifications, assignments submission, calculating grades, attendance sheet generation etc.

Keywords: Data-mining, Clustering, Prediction, Machine learning

Problem Definition

To make monitoring easier, we will provide reliable statistics, behavior groups and predicted results. In this step of the project, we are clustering students by mining Moodle log data. A first objective is to define relevant clustering features. A second objective is to determine if our students show different learning behaviours. We will experiment whether there is an overall, ideal number of clusters and whether the clusters show mostly qualitative or quantitative differences. Experiments in clustering were carried out using real data obtained from various institute using Moodle platform.

Existing System

These e-learning systems accumulate a vast amount of information which is very valuable for analyzing student's behavior. Learning management systems accumulate a great deal of log data about students activities. They can record whatever student activities are involved, such as reading, writing, taking tests, performing various tasks, and even communicating with peers. They normally also provide a database that stores all the system's information: personal information about the users, academic results, user's interaction data, etc. However, due to the vast quantities of data these systems can generate daily, it is very difficult to manage manually, and authors demand tools which assist them in this task, preferably on a continuous basis.

Although some platforms offer some reporting tools, when there are a great number of students, it becomes hard for an administrator to extract useful information. They do not provide specific tools which allow educators to thoroughly track and assess all the activities performed by their learners and to evaluate the structure and contents of the course and its effectiveness in the learning process.

Proposed System

The proposed system will be generic enough to collect and analyse data from any LMS with a logging system. However, our current implementation connects with only one LMS so far, because in our context, all of the available data

comes from a Moodle [1], [2] platform where the courses are located. We also want to see if we are able to predict their end results at their exams just from their curriculum data, which would mean we could henceforth advise students on how they are doing.

Students basically have their own preference in Moodle. Some students like to visit only one module in Moodle but may be others students tend to visit more than one module at a time. There are a few features in Moodle that students occasionally visit and use.

It is important to analysis this random data in future development of Moodle and also to learn the students learning behavior in Moodle. For example, if the data extracted shows that the user will visit Course module if they visit Assignment module, then the administrator need to figure out the way to make navigation between the modules become easier.

In the last few years, researchers have begun to investigate various data mining methods to help teachers improve e-learning systems. Data mining can be applied to explore, visualize and analyze e-learning data in order to identify useful patterns , to evaluate web activity to get more objective feedback for teacher's instruction and to find out more about how the students learn etc. These methods allow us to discover new, interesting and useful knowledge based on student's usage data.

Product Scope

The scopes of this research are:

- i. The method use is Data Mining.
- ii. Data used is login information, user's information ,login time, action and information of the action.
- iii. Algorithm used for clustering students's learing behavior is K-means Algorithm.

Objectives

The main objectives on why this research is done are:

- i. To study the learning behavior of the students that visiting Moodle.
- ii. To apply the Data Mining for clustering student's learning behavior.
- iii. To develop the application in implementation of Data Mining in grouping Moodle data.
- iv. To develop the application that provides online attendance facility.
- v. To develop application which helps teachers to give advise to the student by analyzing students learning behavior and Moodle log's.

System Architecture

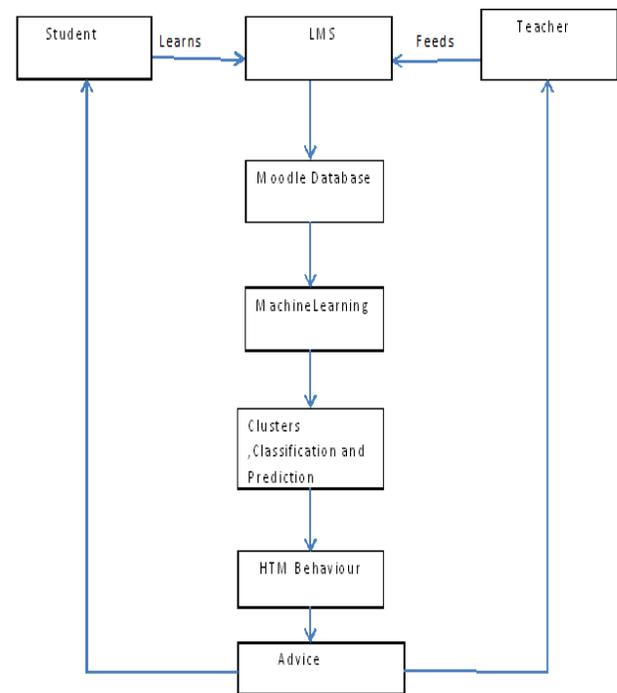


Fig 1.0

Features

The features we have selected are:

- _ Login information
- _ Last login by user
- _ Time spent online by user
- _ Number of PDF downloaded to read later
- _ Number of quizzes, assignments, etc.done
- _ Average grade Calculations
- _ Attendance Sheet generation
- _ Feedback from both student and teachers
- _ Displaying notifications

Clustering

Clustering[5] is the unsupervised grouping of objects into classes of similar objects. In e-learning, clustering can be used for finding clusters of students with similar behavior patterns. In the example of forums, a student can be active or a lurker[3] . These patterns may in turn reflect a difference in learning characteristics, which may be used to give them differentiated guiding. Data mining in general can also be used to better inform teachers about what is going on , or to predict a student's chance of success , which is the final aim of our project.

The pre-processing serves to eliminate useless information, select the data we want to study and, in our case, shape it into features. Our aim with this analysis will be to determine if there is an overall ideal number of clusters and whether the clusters show mostly qualitative or quantitative differences. The clustering will be made by Weka tool, so our experiments will output clusters of students. Hence, we will try to interpret the results in terms of differences of behaviour between students.

Preprocessing Moodle data

Moodle is an open-source learning course management system to help educators create effective online learning communities. Moodle is an alternative to proprietary commercial online learning solutions, and is distributed free under open source licensing. Moodle has been installed at universities and institutions all over the world. An organization has complete access to the source code and can make changes if need be. Moodle can range from a single-teacher site to a 40,000 student university. Moodle's modular design makes it easy to create new courses, adding content that will engage learners. This style of learning believes that students learn best when they interact with the learning material, construct new material for others, and interact with other students about the material. Moodle does not require the use of this style in the courses but this style is what it best supports. Moodle keeps detailed logs of all activities that students perform.

Logging is record keeping that can keep track of what materials students have accessed. Moodle logs every click that students make for navigational purposes and has a modest log viewing system built into it. Log files can be filtered by course, participant, day and activity. The teacher can use these logs to determine who has been active in the course, what they did, and when they did it. For activities such as quizzes, not only the score and elapsed time are available, but also a detailed analysis of each student's with each other.

Teachers can easily get full reports of the activities of individual students, or of all students for a specific activity. Activity reports for each student are available and details about each module as well as a detailed story of each student's involvement.

This can be useful to check to see if everyone has done a certain task, or spent a required amount of time online within certain activities. Moodle does not store logs as text files. Instead, it stores the logs in a relational database. So, data are stored in a single database. MySQL is best supported, but it can also be used with Oracle, Access and

others. We have used MySQL because it is the world's most popular open source database.

Mathematical Model

Set Theory

Our system can be represented as a set

$$\text{System } S = \{I, O, C\}$$

Where,

I = set of inputs

O = set of outputs

C = set of constraints

Input

Input I = {Login, Assignments, Academic Details}

Login = {Username, Password}

Assignments = {Assignments₁, Assignments₂, ..., Assignments_n}

Username = {Username₁, Username₂, ..., Username_n}

Password = {Password₁, Password₂, ..., password_n}

Output

Output O = {Result}

Result = {Reports, Notifications}

Constraint

C = {C₁, C₂}

Where,

C₁ = "User should perform login operation before using the system".

C₂ = "User should have working internet connection".

Conclusion

We propose to create tools that are novel and really needed by training managers. Our application uses data mining and machine learning methods to solve the problem of student monitoring

in e-learning obtained from Moodle courses. We think that these features are comprehensive and generic enough to be reused by othering. These features are then used to conduct a clustering of the data followed by an analysis.

References

- [1] Moodle Trust, \Moodle official site,"2013,<http://moodle.co>
- [2] J. Cole and H. Foster, Using Moodle: Teaching with the popular open source course management system. O'Reilly Media, Inc., 2007.
- [3] J. Taylor, \Teaching and Learning Online: The Workers, The Lurkers and The Shirkers,"*Journal of Chinese Distance Education*, no. 9, pp. 31-37, 2002.
- [4] G. Cobo, D. Garc'a, E. Santamar'a, J. A. Mor'an, J. Melench'on, and C. Monzo, \Modeling student's activity in online discussion forums: a strategy based on time series and agglomerative hierarchical clustering,"in *Educational Data Mining Proceedings*, 2011.
- [5] A. K. Jain, M. N. Murty, and P. J. Flynn, \Data clustering: A review,"1999.