

[DOI: 10.20472/IAC.2019.047.010](https://doi.org/10.20472/IAC.2019.047.010)

ALDA KIKA

University of Tirana, Faculty of Natural Sciences , Albania

LORETA LEKA

University of Tirana, Faculty of Natural Sciences , Albania

SUELA MAXHELAKU

University of Tirana, Faculty of Natural Sciences , Albania

ANA KTONA

University of Tirana, Faculty of Natural Sciences , Albania

USING DATA MINING TECHNIQUES ON MOODLE DATA FOR CLASSIFICATION OF STUDENT'S LEARNING STYLES

Abstract:

Building an adaptive e-learning system based on learning styles is a very challenging task. Two approaches to determine students learning style are mainly used: using questionnaires or data mining techniques on LMS log data. In order to build an adaptive Moodle LMS based on learning styles we aim to construct and use a mixed approach. 63 students from two courses that attended the same subject "User interface" completed the ILS (Index of Learning Styles) questionnaire based on Felder-Silverman model. This learning style model is used to assess preferences on four dimensions (active/reflective, sensing/intuitive, visual/verbal, and sequential/global). Moodle keeps detailed logs of all activities that students perform which can be used to predict the learning style for each dimension. In this paper we have analyzed student's log data from Moodle LMS using data mining techniques for classifying their learning styles focusing on one dimension of Felder-Silverman learning style: visual/verbal. Several classification algorithms provided by WEKA as J48 Decision Tree classifier, Naive Bayes and Part are compared. A 10-fold cross validation was used to evaluate the selected classifiers. The experiments showed that the Naive Bayes reached the best result at 71.18% accuracy.

Keywords:

Learning styles; Felder-Silverman learning style model; Weka; Moodle; data mining

Introduction

Incorporating learning styles in personalized E-learning systems are found to enhance the process of learning for students. Therefore studying techniques for building adaptive systems based on them is a very important topic. A lot of research is made on this aspect. There are various learning style theories given by experts, but the Felder Silverman learning style theory is largely used by researchers on LMS for learning styles' identification on LMS (Rajper et al, 2016).

Felder-Silverman categorizes students in four dimensions as shown in the table below

Table 1: Felder-Silverman learning styles dimensions

Processing dimension: Describes the processing of information	
Active	Reflective
Learners impart by making something with the Information. They need to talk about information to process it.	Learners impart by thinking about Information. They prefer to think and understand before doing anything.
Perception dimension: Describes the perceiving of information	
Sensing	Intuitive
Learners impart by focusing on details, facts, figures and dealing with proven functions that have real applications.	Learners impart by focusing on abstract, initial, and theoretical information. They take a big picture, extract models and connections among ideas and these models.
Input dimension: Describes the presenting of information	
Visual	Verbal
Learners impart by focusing on diagrams, charts or anything can present information.	Learners impart by focusing on explanations by written or spoken sentences.
Understanding dimension: Describes the understanding of information	
Sequential	Global
Learners impart by focusing on managing information in a linear way and dealing with logic and followed steps.	Learners impart by focusing on dealing with information in the historical way, they don't concern about both organizations of information and relationships among it.

Source: (Felder & Silverman, 1988)

As we have proposed in (Leka and Kika, 2018) using an adaptive e-learning system as a pedagogical tool would be very useful to help students learn better by personalizing their learning materials according to their learning styles. The system can use questionnaires at the beginning of the course to define their learning styles. But as their preference can change with time or even during the course the system should also be able to change their learning style by tracking student's behavior in the e-learning environment. By using a data mining technique on the Moodle log data we can predict their learning style.

In this paper we are going to use results from questionnaire based on ILS (Index of Learning Styles) done with 63 students from two courses who attended the same subject "User interface" and the data gathered from Moodle during the development of the course. They were given 44 questions to answer from ILS (Index of learning styles) questionnaire (11 questions from each dimensions), based on Felder-Silverman theory (Felder and Solomon, 2001). Based on the results from this questionnaire, their learning style for each dimension can be calculated. We have used Moodle during the course and collected data about their interaction on this course.

Moodle log data gives information about student's activity like: which material they visited, time of access etc. Based on the results, we will apply three classification algorithms to detect the most accurate data mining technique that can predict student's learning styles from their interaction with the Moodle system.

Related Work

The process of knowledge discovery from datasets from LMS is also called educational data mining (Prabhani et al, 2016). A lot of research is done on this field. For example (Cha et al, 2006) used decision tree and a hidden Markov model to detect learning styles according to the Felder-Silverman. (Garcia et al, 2007) used Bayesian network learning in a web-based education system. (Graf et al, 2008) applied a simple rule-based method (SRBM) to detect learning styles. (Chang et al, 2009) introduced a mechanism that uses k-nearest neighbor classification and genetic algorithms to classify and identify learning styles in a generic model. (Rajper et al, 2016) used Bayesian network to detect student's learning styles based on Kolb model. It used some attributes from student's behavior and data from a survey done. The data mining software was used to process the data results of survey. The Bayesian network produced the Conditional Probability Tables (CPT) for each learning style using each attribute. These probabilities will be updated with the interaction of the students with LMS when he /she will perform any activities on LMS.

Data mining is the process of efficient discovery of non-obvious valuable patterns from a large collection of data. There are a lot of data mining tools. Some examples of commercial mining tools are: DBMiner, SPSS Clementine and DB2 Intelligent Miner and some examples of public domain mining tools are Weka and Keel (Romero et al, 2007).

In this paper we are going to use Weka (Waikato Environment for Knowledge Analysis). Weka is a free software, based in Java language. Weka contains a compilation of visualization tools and algorithms for data analysis and predictive modeling, mutually with graphical user interfaces for simple contact to these functions (Menaka and Kesavaraj, 2019). It provides a collection of machine learning and data mining algorithms for data pre-processing, classification, regression, clustering, association rules and visualization.

We are going to use classification algorithms in order to predict the students learning style. A classifier is a mapping from a (discrete or continuous) feature X to a discrete set of labels Y. This is supervised classification which provides a collection of labeled (pre classified) patterns; the problem is to label a newly encountered, still unlabeled, pattern. In e-learning, classification has been used for several reasons: discovering students with similar characteristics, predicting students' performance and their final grade, assessing the relevance of the attributes involved, grouping students as hint-driven or failure-driven and finding students' common misconceptions, identifying learners with little motivation and finding remedial actions in order to lower drop-out rates etc. (Romero et al, 2007). In (Ktona et al, 2012) data mining techniques were used to create the appropriate curriculum for the Master of Science "Teacher in Informatics for High Schools" in "Informatics" Department of Tirana University. Data about the teaching and learning Computer Science' Subject, which was collected through online questionnaires and interviews, were analyzed using Decision Tree classification method.

Three algorithms of Weka: J48 Decision Tree classifier, Naive Bayes and Part will be used in the next sessions to classify the learning styles of the students using the data log from Moodle and the results from the questionnaire.

Data preparation

For each student, data are collected from the questionnaire and log data of Moodle. Based on the questionnaire, they are classified on one of five classes for each dimension (information process, information perceived, information presentation and learning process). The information about the classes in each dimension and number of students for each class is presented in the tables from 2 to 5.

Table 2: Information Process Dimension

Classes	Moderate Activist	Strong Activist	Balanced	Moderate Reflector	Strong Reflector
Count	14	4	41	4	0

Table 3: Information Perceived Dimension

Classes	Moderate Sensing	Strong Sensing	Balanced	Moderate Intuitive	Strong Intuitive
Count	31	18	13	1	0

Table 4: Information Presentation Dimension

Classes	Moderate Visual	Strong Visual	Balanced	Moderate Verbal	Strong Verbal
Count	29	8	22	4	0

Table 5: Learning Process Dimension

Classes	Moderate Sequential	Strong Sequential	Balanced	Moderate Global	Strong Global
Count	3	0	54	5	1

We filtered from the log data collected on Moodle the records of the students who completed the questionnaire. These records contain the activities that each student has performed using Moodle during the course. The learner's interaction behavior pattern with the learning objects will be used to predict their learning styles.

In order to use Weka tool to determine the most accurate classification algorithm for the prediction of learning style, we transformed the data to the file format for Weka (arff). For each

student the information related to the dimension of information presentation is recorded. The teaching material which is more appropriate for visual learning contains graphics, tables, images, presentations with slides and videos, while teaching material for verbal learner were text based materials and lessons with objectives and content (Dung and Florea, 2012).

The course has teaching materials which depending on the above categorization were marked as visual or verbal. For each student we registered the number of visual material and verbal material that he/she has accessed during the course. In order to compare the number of accesses of each student with each other, we calculated the average of the access for each category from all the students.

Let's denote the number of accesses of one student in visual material as no_visual and the average of the accesses in this category by all students as a_visual . We used two parameters K_{LOW} and K_{HIGH} to categorize the number of accesses in one of the categories {LOW, MEDIUM and HIGH} for each student by using the following pseudocode:

```

if  $no\_visual$  less than  $a\_visual * K_{LOW}$  then
    VISUAL_CATEGORY = LOW
else if  $no\_visual$  less than  $a\_visual * K_{HIGH}$ 
    VISUAL_CATEGORY = MEDIUM
else
    VISUAL_CATEGORY = HIGH

```

The same methodology is used for verbal data.

We have used three pairs of values for the parameters K_{LOW} and K_{HIGH} (0.33,0.66); (0.4, 0,7) and (0.5,0.8).

Experiments

As we are interested to improve the process of learning for the students, we applied the data mining techniques for classification: J48, Part and Naive Bayes to the data that contained information about the grade of the course and their learning styles that we identified by the questionnaires. We didn't find any significant results. When we analyzed the information of the students that have failed we noticed that they have different classes for each dimension and didn't follow any pattern.

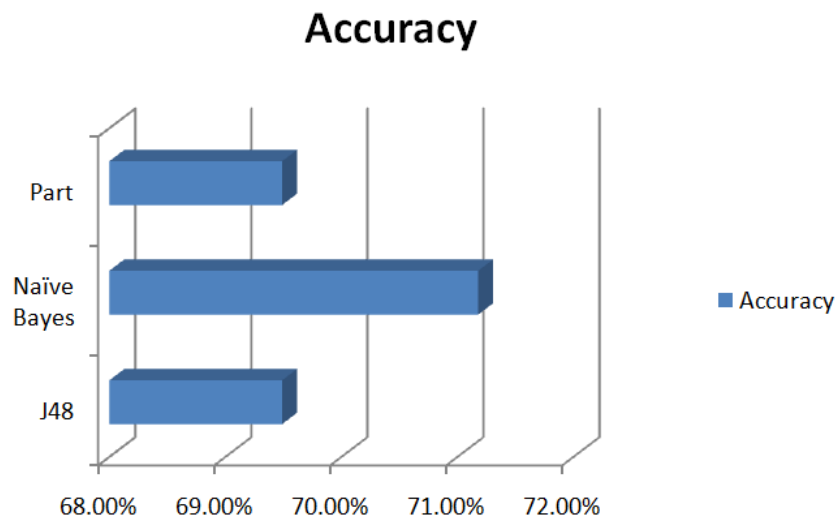
We followed with the next experiments by using the data prepared as in the section above. We added another attribute which is the level of performance of the student based on the average of his grades. We defined the level of the performance of the students by three labels {LOW, MEDIUM, HIGH}.

As table 4 show the data for the dimension Information Presentation are unbalanced. In order to prevent the overfitting problem we applied the filter Resample in Weka which produces a random subsample of a dataset using either sampling with replacement or without replacement increasing the number of instances in the class Verbal which has the lowest number of instances and decreasing the number of instances in the Visual class which has the highest number of instances.

The classification algorithms that we have used are: J48, Naive Bayes and PART. These algorithms are applied in the resembled data with the aim to find the learning style of the students in the dimension of information presentation. We have made several experiments by combining the attributes in order to define the attributes that produce the best result. The best result is reached, for the three algorithms, when the attribute that defines the number of visits for each category of material is produced by applying these coefficients: {0.5,0.8} and the Level of the performance of the students.

The results of the three algorithms are presented in the figure 1.

Figure 1: The accuracy of the three chosen algorithms



As we can see from this graphic the best accuracy is reached by Naive Bayes algorithm with the accuracy 71.18%. In the figure 2 the detailed accuracy by class shows that ROC Area is improved having the average area under the curve greater than in other algorithms, showing that this is the best classifier.

Figure 2: Detailed accuracy by class of Naive Bayes algorithm

```

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.808	0.303	0.677	0.808	0.737	0.502	0.731	0.679	Visual
	0.600	0.088	0.833	0.600	0.698	0.549	0.726	0.643	Balanced
	0.750	0.078	0.600	0.750	0.667	0.613	0.757	0.782	Verbal
Weighted Avg.	0.712	0.182	0.733	0.712	0.711	0.537	0.732	0.678	

Conclusions and future work

In this paper we compared three classification data mining algorithms for the detection of information presentation dimension (visual/verbal) learning style based on Felder-Silverman Learning Style Model from the behavior of the students in Moodle course. The data used were gathered from the result of the ILS questionnaire and Moodle log data of 63 students from the course User Interface and Moodle log data. The algorithms from Weka that were used in the study were: J48, Naive Bayes and PART. We considered number of visits to visual and verbal materials using a coefficient for determining the level of access of the student and the performance level of the student to predict student's learning style on visual/verbal dimension. Results showed that the best accuracy is reached by Naive Bayes algorithm at 71.18% accuracy. We didn't find any relation between learning style and grades that were taken in the course. This results show that there is not a preferred learning style but the learning materials should be adopted for each style in order to increase the performance of the student.

In the future, we are planning to prepare a variety of learning object for the Moodle LMS in order to classify with data mining techniques the learning styles for other dimensions. We aim to integrate Weka into Moodle in order to use the data mining technique found in these studies on Moodle log data to update student's learning styles and adapt Moodle content.

References

- Cha, H.J., Kim, Y.S., Park, S.H., Yoon, T.B., Jung, Y.M. and Lee, J.H. (2006). Learning style diagnosis based on user interface behavior for the customization of learning interfaces in an intelligent tutoring system, In Proc. 8th International Conference on Intelligent Tutoring Systems (ITS'06), pp.513–524.
- Chang, Y., Kao, W., Chu, C. and Chiu, C.(2009). A learning style classification mechanism for e-learning, In: Computers & Education, Vol.53, No.2, pp.273–285 .
- Dung, P.Q., and Florea, A.M. (2012). An approach for detecting learning styles in learning management systems based on learners' behaviours. International Conference on Education and Management Innovation, IACSIT Press, Singapore, vol. 30, pp. 171–177
- Felder, R., Silverman, L., Learning and Teaching Styles In Engineering Education. Engineering Education 1988, 78(7), 674–681.
- Felder, R.M. & Solomon, B.A., (2001). Index of learning styles questionnaire, North Carolina State University.
- Garcia, P., Amandi, A., Schiaffino, S. and Campo, M. (2007). Evaluating Bayesian networks precision for detecting students' learning styles, Computers & Education, Vol.49, No.3, pp.794–808 .
- Graf, S., Kinshuk and Liu T.(2008) Identifying Learning Styles in Learning Management Systems by Using Indications from Students' Behaviour, Proc. IEEE International Conference on Advanced Learning Technologies (ICALT'08), pp.482–486 .

Ktona A., Collaku D., Kika A., Design and development of the curriculum for master of sciences in teacher of informatics, Albania International Conference on Education, December 2012.

Leka, L. and Kika A.(2018). A study on student's learning styles using Felder-Silverman Model. In: KNOWLEDGE, International Journal Budva, Montenegro .

Menaka, S. and Kesavaraj, G. (2019). Predicting Student Performance Using Data Mining Techniques: A Survey Of The Last 5 Years, International Journal of Advanced Scientific Research and Management , Vol. 4 (1):98-102.

Prabhani P. L., Gunawardena L., and HiraKawa M. (2016). Detecting Learning Styles in Learning Management Systems Using Data Mining, In Journal of Information Processing, 24(4):740-749.

Rajper S., Shaikh N. A., Shaikh Z. A. and Mallah G. A. (2016). Automatic Detection of Learning Styles on Learning Management Systems using Data Mining Technique, Indian Journal of Science and Technology Vol 9(15).

Romero C., Ventura S., Garcí'a E. (2007). Data mining in course management systems: Moodle case study and tutorial. In: Computers & Education, Vol.51, Iss.1, pp.368–384.