

## Regular Paper

# Detecting Learning Styles in Learning Management Systems Using Data Mining

MADURA PRABHANI PITIGALA LIYANAGE<sup>1,a)</sup> K.S. LASITH GUNAWARDENA<sup>1,b)</sup> MASAHIRO HIRAKAWA<sup>1,c)</sup>

Received: April 20, 2015, Revised: August 24, 2015,  
Accepted: December 12, 2015

**Abstract:** The use of data mining in the education sector has increased in the recent past. One reason for this is the wide use of learning management systems (LMS), which store data related to learning activities. The goal of this research is to predict individual learning styles using the Moodle LMS by analyzing log data using a data mining technique. We use the Waikato environment for knowledge analysis (Weka), as the data mining tool and compare the differences in the performance of several data mining techniques using course log data. Our experimental results show that the J48 decision tree classification algorithm works best with our dataset. We also propose a group learning map that visualizes the learning styles in a class, which can help instructors and learners achieve learning outcomes more effectively.

**Keywords:** learning management systems, learning styles, moodle, data mining

## 1. Introduction

The past decade has witnessed exponential growth in the number of individuals with regular access to the Internet. Technological advances have led to the falling cost of Internet-ready devices. The availability of high-speed network connections has also allowed people to engage in multimedia-rich experiences. This has contributed to an increasing number of people seeking e-learning to supplement regular classroom education. The recent popularity of massive open online courses (MOOC) is one such case.

Learning management systems (LMS) are a key category of the software platforms currently used to provide e-learning. An LMS can be defined as a framework that handles all aspects of the learning process and is an infrastructure that delivers and manages instructional content, identifies and assesses individual and organizational learning or training goals, tracks progress toward those goals, and collects and presents data for supervision of the learning process of an organization as a whole [1].

Among the LMS software currently in use, the modular object-oriented developmental learning environment (Moodle) is the most common with more than 53,000 sites serving over 69 million users in 230 countries [2]. The ability to run on different platforms and the cost of ownership are two reasons for its wide acceptance. In addition, the ability to introduce third-party enhancements to Moodle via modules allows developers to improve functionality to suit individual requirements.

Though LMSs are a powerful tool, Graf and List [3] suggested that they still have some issues such as the inability to personal-

ize learning, i.e., digital content is offered in the same format to all learners in a particular course. Furthermore, LMSs tend to be course centric rather than learner centric [3]. The issues that must be addressed in LMSs include learner expectations, motivation, learning style, security, and interoperability [4]. Note that diverse efforts to enhance the LMS experience are beginning to appear. The ability to personalize LMSs with style templates and various languages is one such approach. The use of sharable content object reference model (SCORM) standards, which enable interoperability, accessibility, and the reusability of web-based content is another approach.

Meanwhile, several models for classifying the learning styles of learners have been presented by educational theorists. Such models share some common features and can be applied to different scenarios. Among these learning style classification models, the Felder-Silverman learning style model (FSLSM) has been recognized and applied to e-learning environments. In this model, a learning style is explained using four dimensions, each formed by a pair of distinct preferences: active-reflective, sensing-intuitive, sequential-global, and visual-verbal.

Numerous recent studies [5], [6], [7], [8], [9], [10], [11], [12], [13], including one study by the authors [14], attempt to address the issue of identifying learning styles to enable personalization of the learning experiences. These studies have adopted statistical and simple rule-based approaches. An important factor to consider is that an individual learning style may vary because of many factors within the course or LMS. For example, different course content, subjects, threshold data for the course, and learner behavior and experience of online learning may affect an individual's learning style. Thus, such systems must be able to respond dynamically to such divergence.

In our previous study [14], we presented a framework to per-

<sup>1</sup> Interdisciplinary Graduate School of Science and Engineering, Shimane University, Matsue, Shimane 690-8504, Japan

a) research@prabhani.com

b) research@lasith.com

c) hirakawa@cis.shimane-u.ac.jp

sonalize the Moodle LMS by identifying the learning styles of learners and then recommended suitable learning materials for each learner. As a part of this framework we introduced three basic agents, i.e., a learning style monitoring and learning profile creation agent (LLA), an expert recommendation agent (ERA), and an adaptive content presentation and interface enhancement agent (AIA). One of the functions in LLA, which is a key part of the framework, is to detect the learning style of the learner based on the activities performed in Moodle. Here we used a simple rule-based method for the detection. LLA also provides a function for generating a graphical representation of each learner's learning style. A set of threshold values related to each course are introduced in the ERA module and these are used for the estimation of learners' learning styles in the LLA module by referring to Moodle log data. The AIA, which is not yet implemented, aims to provide links for recommended learning materials to the learner.

In this study, we present a trial to predict student learning styles automatically using a data mining technique within LLA. We present a comparison of four data mining algorithms for the analysis of Moodle LMS log data: J48, Bayesian network, naive Bayes, and random forest. The Weka, a powerful data mining toolkit, is applied here. Using a dataset from a course administered at a higher education institute in Sri Lanka, it was determined that the J48 decision tree algorithm would be the best method for our purpose. We then implemented an automatic learning style prediction module based on the J48 algorithm. In addition, we extended a module in LLA that presents the learner's learning style graphically. This extension enables an instructor to compare a group of learners against a targeted individual.

The remainder of this paper is organized as follows. Section 2 describes related studies on learning style approaches, educational data mining, and data mining tools for LMS. Section 3 explains our approach—LMS with learning style-based support facilities. Preparation for the integration of Weka, experiments with four data mining techniques, comparison with existing trials and integration of the J48 classifier are discussed in Section 4. Section 5 discusses the learning style visualization—the group learning map. Limitations of our research are discussed in Section 6. We conclude the paper and elaborate on future research directions in Section 7.

## 2. Related Studies

### 2.1 Learning Styles

The term “learning style” has been defined by Honey and Mumford [15] as “a description of the attitudes and behaviors which determine an individual's preferred way of learning.” Several studies have proposed different models to explain possible learning styles. Of these models, the FSLSM proposed by Richard Felder and Linda Silverman is well known. It is defined by four dimensions, each formed by a pair of distinct characteristics (learning styles).

The first dimension considers the learner's preferred method of processing information—active (ACT) or reflective (REF). The second dimension considers the type of information that the learner preferentially perceives—sensory (SEN) or intu-

itive (INT). The third dimension considers the sensory channel through which the learner most effectively perceives external information—visual (VIS) or verbal (VER). The fourth dimension considers how the learner progresses toward understanding—sequentially (SEQ) or globally (GLO). Interestingly, the ILS developed by Felder and Soloman [16] can be used to assess preferences in four FSLSM dimensions. The ILS comprises 44 questions with 11 questions for each dimension.

Note that the FSLSM is the learning style model that is most frequently cited with respect to computer-based education systems [7], [8], [9], [10], [11], [12], [13], [14], [17], [18], [19]. One approach in applying the FSLSM is to use the ILS as an online questionnaire to evaluate learning preferences and recommend appropriate learning material [8], [9], [10].

### 2.2 Educational Data Mining

Data mining or knowledge discovery in databases is the automatic extraction of implicit and interesting patterns from large data collections [20]. The educational data mining community website [21] defines educational data mining as “an emerging discipline concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students and the settings that they learn in.”

Chang et al. [22] introduced a mechanism that uses k-nearest neighbor classification and genetic algorithms to classify and identify learning styles in a generic model. Kotsiantis et al. [23] presented a comparison of several data mining algorithms (naive Bayes, Bayes network, support vector machine, logistic regression, and decision tree) to detect student mental models in intelligent tutoring systems. Garcia et al. [5] used a data mining method that employs Bayesian network learning styles in a web-based education system. Graf et al. [24] and Dung and Florea [13] applied a simple rule-based method (SRBM) to detect learning styles. Cha et al. [7] used decision tree and a hidden Markov model to detect learning styles according to the FSLSM.

Another approach, used particularly for grouping learners, is the automatic discovery of behavioral patterns such as learning styles. This approach has been performed using clustering algorithms such as bisection K-means and Kohonen's self-organizing map algorithm.

Romero et al. [25] proposed that the contribution provided by educational data mining activities can be classified into several categories:

- I. Analysis and visualization of data
- II. Providing feedback for supporting instructors
- III. Recommendations for students
- IV. Predicting student performance
- V. Student modeling
- VI. Detecting undesirable student behaviors
- VII. Grouping students
- VIII. Social network analysis
- IX. Constructing courseware
- X. Developing concept maps
- XI. Planning and scheduling

Using LMS data in an educational data mining approach to de-

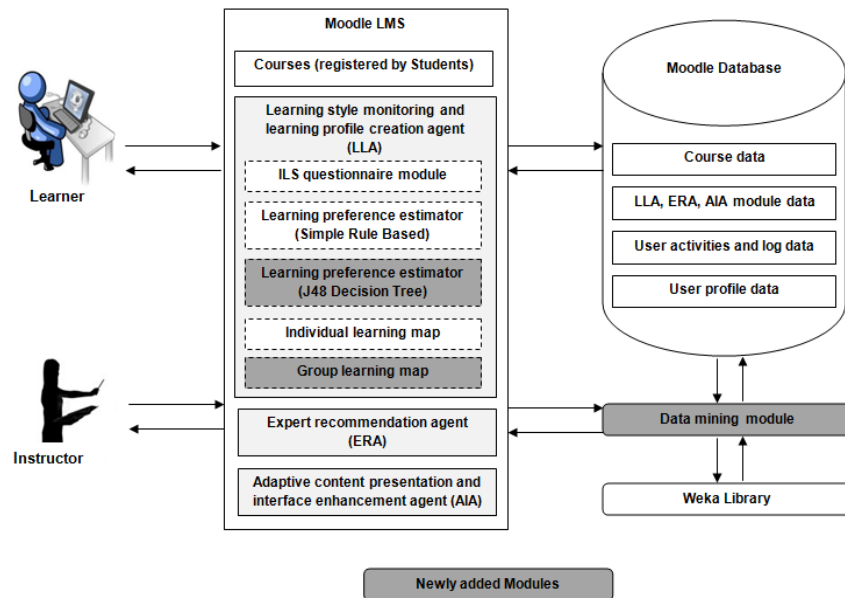


Fig. 1 Framework of the proposed system.

fect learning styles can contribute to categories II, V, VI, and VII. Further content recommendation (III) based on identified learner behaviors may also be promising. Most of the above data mining studies contributed to categories III, V, and VII, because they considered only learner aspects. However, because education is a collaborative process between two parties we focus on instructors as well as learners. Note that we presented ideas for categories II and V in a previous study [14]. Here, we attempt to contribute to categories II and VII through a comparison of educational data mining techniques and the implementation of a Moodle module, called the group learning map and described in Section 5.

### 2.3 Data Mining Tools for LMS

Recently, there has been a surge in the number of studies performed in the educational data mining domain. One reason for this surge is the appearance of powerful data mining tools such as DBMiner [26]. Another important reason has been the emergence of numerous open source public domain data mining tools such as Keel [27], Weka [28], RapidMiner [29], R [30], and KN-IME [31]. Evaluations of such tools have concluded that there is no single best tool and that each has advantages and disadvantages [32], [33].

Among these data mining tools, Weka—a Java-based open-source platform developed at the University of Waikato, New Zealand—is one of the most common. Weka has a graphical user interface and a command line interface. Because of the characteristics of its API, Weka can be embedded into other systems. Weka currently supports a large collection of machine learning and data mining algorithms for data preprocessing, classification, regression, clustering, association rules, and visualization.

## 3. Learning Management System with Learning Style-based Support Facilities

From the developer perspective, the ability to extend the functionality of the Moodle LMS by adding modules is advantageous. Those modules can be developed using PHP.

Fig. 2 ILS questionnaire.

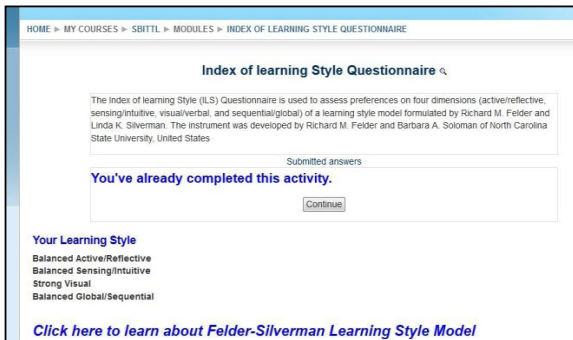
We investigated enhancing the capability of Moodle by detecting learning style and providing meaningful feedback to the learner based on the detected learning style. A framework of the system is illustrated in Fig. 1, which includes three basic agents LLA, ERA, and AIA. This is an extension of our previously proposed framework in Ref. [14], and the new functionalities are highlighted in the figure.

The Moodle LMS database comprises over 250 tables [34]. To estimate the learning styles we used attributes from 12 existing and four new tables. These tables are listed in Appendix A.1.

The LLA supports three functions. First in the ILS Questionnaire sub-module, it suggests the learners for participation in the ILS questionnaire (Fig. 2) that estimates a learning style based on the FLSM (see Table 1 for a reference example of questionnaire results). The ILS output provides a label each for the four dimensions of the FLSM describing the user's preference (see Fig. 3).

**Table 1** ILS questionnaire summary.

	Number of learners	Percentage
Strong Active	3	4%
Moderate Active	10	13%
Balanced	54	68%
Moderate Reflective	10	13%
Strong Reflective	3	4%
<b>Total</b>	<b>80</b>	<b>100%</b>
Strong Sensing	10	13%
Moderate Sensing	13	16%
Balanced	33	41%
Moderate Intuitive	15	19%
Strong Intuitive	9	11%
<b>Total</b>	<b>80</b>	<b>100%</b>
Strong Global	5	6%
Moderate Global	13	16%
Balanced	38	48%
Moderate Sequential	15	19%
Strong Sequential	9	11%
<b>Total</b>	<b>80</b>	<b>100%</b>
Strong Visual	9	11%
Moderate Visual	9	11%
Balanced	43	54%
Moderate Verbal	11	14%
Strong Verbal	8	10%
<b>Total</b>	<b>80</b>	<b>100%</b>

**Fig. 3** Learning styles estimated using ILS questionnaire.

Each label may take one of five possible values with respect to the selected dimension:

- I. Strong preference for learning style 1
- II. Moderate preference for learning style 1
- III. Balanced (learning style 1—learning style 2)
- IV. Moderate preference for learning style 2
- V. Strong preference for learning style 2

The labels are stored in the mdl\_ILS\_tracking table of the Moodle database. The mdl\_ILS\_value table stores the corresponding scores obtained from the ILS by a user. This data forms part of the user's profile.

Second, in the learning preference estimator sub-modules, the LLA estimates each learner's learning profile based on their activities in the LMS. The system calculates this profile by con-

sidering the mapping provided by Graf et al. [18], [35], [36] to determine the attributes that are relevant for estimating learning styles.

In our previous implementation [14], we adopted a rule-based method because of its simplicity. In that method, we first considered the content offered in a course. The content in each course comprises learning objects (LOs) such as videos, quizzes, exercises, examples, and self-assessments. Using the Moodle log, we can determine the exact number of times each student accessed LOs. Thus, it is possible to calculate the ratio of visits for each type of LOs ( $R_{\text{VisitedLOs}}$ ) by taking the number of visits versus the total number of objects for each LO type. Similarly, it is possible to estimate the time spent visiting each LO. As previously calculated, we determined the ratio of the LO stay time ( $R_{\text{TimeSpentLOs}}$ ) for each learning object type by taking the ratio of this time spent versus an instructor estimated expected time for each LO type.

Graf et al. [18], [35], [36] explain that learners who have different learning style act differently with different LOs. For example, reflective, intuitive, and verbal learners prefer to visit LOs more frequently than active, sensing, and visual learners. Consequently, reflective, intuitive, and verbal learners show positive (+) behavior on a LO visit whereas active, sensing and visual learners show negative (−) behavior.

If the calculated  $R_{\text{VisitedLOs}}$  or  $R_{\text{TimeSpentLOs}}$  ratio lies between a pre-determined upper threshold (UT) and a lower threshold (LT) determined by the instructor for every course, the behavior is considered balanced. On the other hand, if the ratio is less than the lower threshold, then the behavior is considered negative; if the ratio is higher than the upper threshold, the behavior is considered positive. For a certain behavior pattern, when its property (positive or negative) corresponds to the behavior mapping (+ or −) for each learning style, we take the value of the ratio into account for learning style evaluation. For example, assume the ratio of content\_visit,  $R_{\text{content\_visit}}$ , be 0.2, and  $UT_{\text{content\_visit}}$  and  $LT_{\text{content\_visit}}$  be 0.8 and 0.3, respectively. Since  $R_{\text{content\_visit}}$  is less than  $LT_{\text{content\_visit}}$ , the behavior is considered negative. Knowing that active, sensing and visual learners have negative (−) behavior on content visit, the  $R_{\text{content\_visit}}$  value is used for evaluation of those three characteristics only.

We finally calculated the average ratio ( $R_{\text{AVG}}$ ) for each learning style based on the mapping introduced by Graf et al. as:

$$R_{\text{AVG}} = \frac{\sum_{i=1}^n R_i}{n}$$

where  $R_i$  is the ratio of the  $i^{\text{th}}$  behavior pattern and  $n$  is the number of relevant behavior patterns for the selected learning style. See Ref. [14] for more explanation. The eight  $R_{\text{AVG}}$  values pertaining to the learning styles are stored in the mdl\_dimensions table.

In this study, we extended the system functionality by automating the process of learning style extraction in Moodle using data mining. For this purpose, we applied the Weka data mining tool as shown in Fig. 1. This was carried out in the newly developed Learning preference estimator (J48 Decision Tree) sub module (**Fig. 4**).

The third function in the LLA is a visualization of a learning map that presents learner's learning preference. A new feature



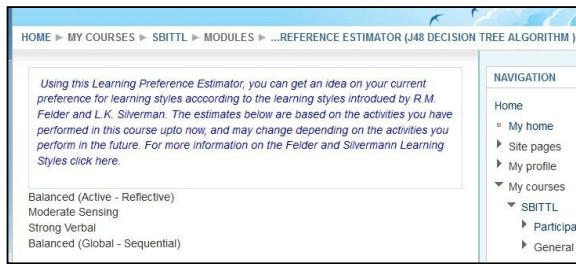


Fig. 4 J48 Decision Tree module output.

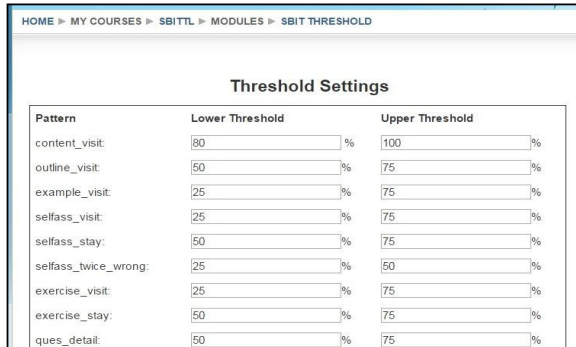


Fig. 5 ERA module threshold settings.

was added so that the learning map can display the distribution of learning preferences for a given class of learners. This helps the instructors determine the type of learning materials they should recommend and provide. In addition, this allows all learners to recognize their learning styles and enables comparison with others in the same class.

The ERA module is prepared to enable the instructor to tune the conditions for the estimation of learning styles. Here, a set of threshold values of UT and LT applied to the LLA contributes to the estimation of learning styles (Fig. 5).

Note that this set of threshold values may differ from course to course. The table mdl Lec.threshold is prepared to store threshold values to determine whether a learner's behavior for a particular dimension is a relevant positive behavior, relevant negative behavior, or irrelevant [14]. Note that these threshold values can be customized by the instructor.

The AIA facilitates adaptive course LO recommendation in Moodle is currently under development and will be reported on at a later date.

## 4. Integration of Weka

### 4.1 Preparation

We adopted SRBM to determine the learning style in a previous study [14]; however, it is worthwhile investigating the capabilities of other sophisticated data mining techniques. Here, we applied Weka to facilitate this investigation. Note that Weka's graphical interface is more commonly used than its command line interface. This graphical user interface is convenient for users; however, the development of a program with an API is mandatory for our purpose. Before we began developing the code to include a data mining technique, we compared the performance of different data mining methods, including J48, Bayesian network, naive Bayes, and random forests.

A comparative experiment was performed using the "Introduc-

**Table 2** Sample data pertaining to a single student used in the training dataset.

Dimensions	Dimension 1		Dimension 2		Dimension 3		Dimension 4	
Learning styles	ACT	REF	SEN	INT	SEQ	GLO	VIS	VER
$R_{AVG}$ values obtained from mdl_dimensions	0.43	0.03	0.75	0.04	0.13	0.58	0.04	0.21
Learning style labels obtained from mdl_ILS_tracking	Moderate active		Strong Sensing		Moderate Global		Balanced	

tion to Information Technology" course at a higher education institution in Sri Lanka. 80 students participated in the course. The course content contained 50 learning material items: 22 content objects, 8 outlines, 2 flash examples, 10 self-assessment quizzes, and 8 exercise quizzes. Note that content object video tutorials were packaged as SCORM material. The course duration was 14 weeks (1 semester).

During the first week of the course, the students participated in the ILS questionnaire, to get an estimate of their learning style. The results are summarized in Table 1.

In our implementation, we did not consider the "content stay" and "outline stay" that have been adopted previously [18], [35], [36], because it is difficult to gather meaningful data for these items from Moodle. For our data analysis, training data were obtained by considering the eight  $R_{AVG}$  values in the mdl\_dimensions table together with the corresponding learning styles labels obtained by the mdl\_ILS\_tracking table. A sample dataset is given in Table 2.

The collected data were transformed into the Weka-specific attribute-relation file format (ARFF). For each student, four instances pertaining to the four dimensions were recorded. Each instance recorded the two  $R_{AVG}$  values obtained for a dimension together with the corresponding ILS label. During preprocessing, we removed data that contained missing values. In addition, we attempted to eliminate bias toward the majority class due to imbalanced data in the dataset [37]. In our analysis, we determined that the classes in the ACT/REF dimension were imbalanced. Note that the synthetic minority oversampling technique (SMOTE) was applied to the imbalanced dataset.

### 4.2 Experiments with Four Data Mining Techniques

We considered sample accuracy rate as the main criterion for determining the most appropriate data mining technique. The results of the performance evaluation are shown in Tables 3, 4, 5, 6 as correctly classified instances generated by Weka. We used the 10-fold cross validation method to estimate the accuracy rates. Two additional criteria (i.e., precision and receiver operating characteristic (ROC) area) given by Weka, are also shown.

As can be seen, the J48 classifier demonstrates reasonably high performance, with the exception of one class (i.e., the active and reflective dimension). For this dimension, the random forest method yielded a sample accuracy of 72.77% compared to 65.26% obtained by J48. Note that sometimes, correctly classified instances can be insensitive to class distribution. Therefore,

**Table 3** Performance in ACT/REF dimension.

		J48	Bayes net	Random Forest	Naïve Bayes
Correctly Classified Instances		65.26%	63.84%	72.77%	59.62%
Precision	Moderate Active	0.865	0.718	0.892	0.857
	Balanced	0.483	0.608	0.6	0.568
	Strong Active	0.667	0.541	0.625	0.541
	Strong Reflective	0.816	0.804	0.868	0.571
	Moderate Reflective	0.487	0.5	0.645	0.5
ROC Area	Moderate Active	0.916	0.932	0.957	0.91
	Balanced	0.742	0.841	0.846	0.772
	Strong Active	0.827	0.887	0.908	0.827
	Strong Reflective	0.904	0.93	0.984	0.851
	Moderate Reflective	0.763	0.89	0.833	0.807

**Table 4** Performance in SEN/INT dimension.

		J48	Bayes net	Random Forest	Naïve Bayes
Correctly Classified Instances		80.00%	66.25%	72.50%	56.25%
Precision	Moderate Sensing	0.917	0.556	0.667	0.692
	Balanced	0.737	0.769	0.65	0.514
	Strong Sensing	0.889	0.667	0.889	0.583
	Strong Intuitive	0.7	0	0.875	0.4
	Moderate Intuitive	0.909	0.55	0.818	0.75
ROC Area	Moderate Sensing	0.921	0.824	0.867	0.885
	Balanced	0.813	0.809	0.77	0.649
	Strong Sensing	0.894	0.846	0.918	0.863
	Strong Intuitive	0.885	0.731	0.844	0.814
	Moderate Intuitive	0.815	0.807	0.87	0.749

**Table 5** Performance in SEQ/GLO dimension.

		J48	Bayes net	Random Forest	Naïve Bayes
Correctly Classified Instances		90.00%	81.25%	88.75%	85.00%
Precision	Moderate Global	1	0.692	0.818	0.727
	Balanced	0.86	0.85	0.857	0.804
	Strong Global	0.8	0	1	1
	Strong Sequential	1	1	1	1
	Moderate Sequential	0.938	0.75	0.933	1
ROC Area	Moderate Global	0.794	0.802	0.773	0.78
	Balanced	0.891	0.814	0.918	0.897
	Strong Global	0.893	0.792	0.997	0.824
	Strong Sequential	0.944	0.915	0.996	0.98
	Moderate Sequential	0.992	0.986	0.997	0.995

when selecting the best technique precision rates for each class and the ROC area values must be considered. An ROC curve was created by plotting the true positive rate against the positive rate for various threshold settings. An optimal classifier should have ROC values that approach 1. By considering the data shown in Tables 3–6, we concluded that J48 was the most appropriate method for our dataset.

#### 4.3 Comparison with Existing Trials

We obtained sample accuracy rates of 65.26%, 80.00%,

**Table 6** Performance in VIS/VER dimension.

		J48	Bayes net	Random Forest	Naïve Bayes
Correctly Classified Instances		81.25%	61.25%	78.75%	60.00%
Precision	Moderate Visual	0.6	0	0.714	1
	Balanced	0.833	0.661	0.833	0.673
	Strong Visual	0.75	0.5	0.667	0.545
	Strong Verbal	1	0	0.714	0.2
	Moderate Verbal	0.875	0.4	0.778	0.455
ROC Area	Moderate Visual	0.846	0.6	0.782	0.618
	Balanced	0.856	0.79	0.746	0.679
	Strong Visual	0.897	0.704	0.891	0.728
	Strong Verbal	0.811	0.707	0.892	0.861
	Moderate Verbal	0.718	0.695	0.828	0.715

**Table 7** Precision rate comparison.

Authors	ACT/REF	SEN/INT	SEQ/GLO	VIS/VER
Garcia et al. [5]	58.00%	77.00%	63.00%	–
Graf et al. [25]	79.33%	77.33%	73.33%	76.67%
Dung and Florea [13]	72.73%	70.15%	65.91%	79.54%
Our previous study SRBM 1st trial [14]	63.64%	77.27%	77.27%	72.73%
Our previous study SRBM 2nd trial [14]	65.00%	75.00%	77.50%	76.25%
This Research	70.89%	84.38%	91.25%	82.50%

90.00%, and 81.25% for the ACT/REF, SEN/INT, SEQ/GLO, and VIS/VER dimensions, respectively. This calculation using Weka was based on the exact matches of actual data and the predicted results. However, the precision measurement proposed by Garcia et al. [5] has been used in many studies that have attempted to predict learning styles. To compare the performance of our trial with that of existing trials, we used the following formula for precision:

$$\text{Precision} = \frac{\sum_{i=1}^n \text{Sim}(\text{LS}_{\text{ILS}}, \text{LS}_{\text{FW}})}{n} \times 100,$$

Here,  $\text{LS}_{\text{ILS}}$  and  $\text{LS}_{\text{FW}}$  are the learning styles obtained by the ILS and that obtained by the proposed method, respectively. The parameter  $n$  is the number of students in the course. The function  $\text{Sim}$  calculates the similarity between  $\text{LS}_{\text{ILS}}$  and  $\text{LS}_{\text{FW}}$ . If the magnitude of  $\text{LS}_{\text{ILS}}$  is equal to that of  $\text{LS}_{\text{FW}}$   $\text{Sim}$  takes 1, 0 if they are opposite, and 0.5 if one is neutral and the other is an extreme value. The accuracy rate given by Weka differs from the precision rate in the above because the weight 0.5 is not considered in the calculation.

**Table 7** compares the precision rates obtained by the proposed method with those of other studies, including our own previous study. Garcia et al. [5] applied Bayesian networks to an artificial intelligence course with 40 students. Graf et al. [24] estimated learning styles using an SRBM for a Web Engineering course with 43 students. Dung and Florea [13] also used an SRBM to estimate learning styles for an artificial intelligence course with 44 students. In our previous study [14], we performed two trials using an SRBM.

For the SEN/INT, SEQ/GLO, VIS/VER dimensions, the proposed method obtains good results, when compared with previ-

ous research. These results could be attributed to the selected J48 decision tree algorithm and the approach of using  $R_{AVG}$  values. The precision rate obtained for the ACT/REF dimension was slightly lower than those for the other dimensions, as well as the corresponding rates obtained by some of the other studies. A possible reason for this is that in our trials, face-to-face content delivery sessions (i.e., traditional classroom lectures) were provided in conjunction with LMS learning sessions. In addition, a printed textbook accompanied the course content. This would have created a situation whereby some students may have no compelling reason to refer to the LMS materials. Furthermore, the ACT/REF dimension dataset was imbalanced. We will continue to improve the performance of learning style prediction in the future.

#### 4.4 Integrating the J48 Classifier

Having determined that the J48 classifier was the best fit for our purpose, we programmed a corresponding module so that the system could realize learning style-based assistance. The system was set up on an Intel Core i5 computer running Windows 7. Moodle 2.3.2 ran on a WAMP Server (Apache 2.2.21, MySQL 5.5.20, and PHP 5.3.10). Weka 3.6.11 was installed on the same server.

Data read from the MySQL server were transformed to ARFF before training a classifier. A new Moodle module was implemented in PHP to invoke program code for the J48 classifier, which was prepared as an executable Java archive (JAR) file. When the system was first executed, ILS data were given to the classifier together with the  $R_{AVG}$  data for training. Once the training of the classifier was completed, the system was ready to perform classification. The result of classification, i.e., a learning style, was then stored in the database. The classification was repeated four times, one each for each learning style dimension. This prediction is accessible via the LMS, and is automatically re-evaluated once per day.

It should be noted that, when a particular course commences for the first time, predicting learning styles of learners using data

mining cannot be performed, as there is no log data on learner's behavior history of accessing learning objects. Therefore, all new users are expected to complete the ILS questionnaire. The learning style labels obtained using the ILS are handled together with the  $R_{AVG}$  values as training data, and prediction of a learning style becomes possible after few weeks of classes have passed. Up to that point, the system relies on the ILS result. When the course is re-run with a new set of students, the system does not require learners to run the ILS anymore, it needs to wait until the learner's access the learning objects (at least a week of interaction) before evaluation of relevance to past records. This approach is valid as long as the threshold values for the course in the ERA or LOs do not change.

### 5. Learning Style Map

In our previous study [14], we introduced the visualization of a learning styles map as a tool to represent learning preference. This map comprises four quadrants, each of which corresponding to the learning style dimensions. Note that the  $R_{AVG}$  values for a given learner are highlighted as colored rectangles. The images are generated by the GD graphics library using its PHP interface.

This visualization helps the learner to comprehend his/her own learning style, rather than showing learning style labels as text. The learner can easily get an overall picture and can identify which side of learning style he/she needs to become a balanced learner, and which type of learning material he/she needs to follow to master the course. Knowing social comparison theory [38] this functionality, included in the LLA module, has been improved in our current implementation so that for a certain learner, learning styles of other learners taking the same course can be presented along with his/her own learning style. We realized the visualization of the form in **Fig. 6** (a), where the learning styles of fellow students are plotted with black dots.

This visualization scheme is also beneficial to the instructor. Suppose that a learner has not performed well in an examination compared to his/her colleagues. The visualization tool is used to

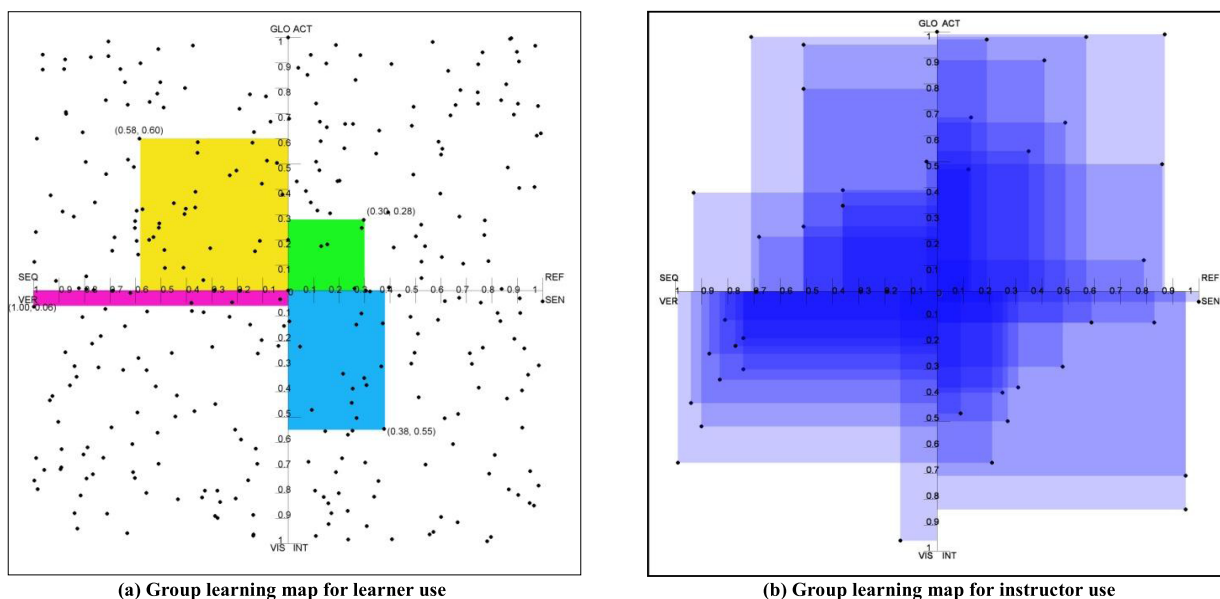


Fig. 6 Group learning maps.

evaluate whether the performance difference could be attributed to the learning style. Here the tool gives an option to select all or an arbitrary subset of learners for comparison. This option is provided only for the instructor, not the student, in order to maintain privacy.

When the instructor needs to determine the diversity of learning styles in a student group or its subset by way of an overall view, the visualization in Fig. 6 (b) can be used. This facility helps the instructor to adjust course materials and/or lecturing styles to achieve the expected learning outcomes more effectively.

For example, in Fig. 6 (b), the SEQ-GLO, ACT-REF, and SEN-INT dimensions exhibit no specific characteristics. However, a rather distinctive feature is found in the VER-VIS dimension; one student exhibits a strong visual learning style but the majority shows a verbal learning style. This indicates that if the course contains slides shows and videos, it would be better to supplement them with audio.

The learning style of a student is based on the actions performed in the LMS; thus, the learning map may change dynamically. This module is used once daily to automatically generate the learning style so that predictions are updated dynamically. This allows learners and instructors to view the most up-to-date learning styles.

## 6. Limitations

There are a few limitations in our study that must be addressed. We summarize these limitations in the following and suggest remedial actions that can be undertaken.

Although different studies, including ours, have come up with different models of learning styles, one common feature is that they consider only a single course in their trials. Considering that a learner may take two or more courses, further examination of multiple course cases is required. In this situation, it is possible to obtain different  $R_{AVG}$  values for each course. There could be many reasons for this. The first is that learner preference may vary for each course due to differing subject matter. Second, learner preference may also vary depending on the type of learning materials used in the LMS (i.e., audio, video, graphics, and text). Third, in the proposed system, the threshold values used to estimate learning styles can be fine-tuned by the course instructor. Finally, the learner's experiences with online/distance learning can affect their interaction with an LMS, which can then affect the  $R_{AVG}$  scores. One simple solution to this is to calculate the  $R_{AVG}$  value as the average values among the multiple courses.

Another limitation is the selection of the algorithm used for prediction. In our experiment, we determined that J48 was the most suitable; however, it is premature to assume that other data mining algorithms are unsuitable for similar data mining-based predictions of learning styles. Depending on the type of course resources and student performance, prediction accuracies may vary. Further testing with different courses is required to obtain a clearer understanding of this issue.

At present, in our ERA module, the instructor must fine-tune the course thresholds since these play an important role in determining how the system classifies individual learning preferences as strong, moderate, or balanced. If we consider the commence-

ment of a course, we can use the default values based on the literature. However, the present system provides no guidance for the instructor when tuning these thresholds. Therefore, we need to propose a new module that graphically presents performance metrics, such as the number of times users engaged with specific resources. This will ensure that during the second run of the same course, the instructor can use performance metrics to determine the most effective threshold values. Alternatively, the system should provide assistance to determine these threshold values most effectively. It would be possible to find the threshold values that result in the best classification performance, by the repetitive execution of training and performance evaluation tasks with changing threshold values.

In addition, dataset used for comparison are different depending on the research organization. There is a difficulty when researchers want to compare their own results with those of others. It would be nice if a common dataset will be available as is seen in, for example, pattern recognition domain.

Finally, based on the observation and analysis of learner behavior, instructors, and the proposed system, it is desirable that meta-rules, which cover high-level and sophisticated learning activities, in general, be determined.

## 7. Conclusions and Future Studies

The aim of this study was to automate the process of learning style extraction from the Moodle LMS using a data mining technique. This study was undertaken using a course with 80 students at a higher education institute in Sri Lanka. We considered four data mining techniques, i.e., J48, Bayesian network, naive Bayes, and random forests, to predict the learning styles. A new Moodle module for the automatic prediction of learning style was developed using the most suitable algorithm for our dataset, i.e., the J48 decision tree algorithm. This study revealed that the proposed system demonstrates better performance than previously proposed systems. We also provided a new feature to visualize and analyze learning styles. This feature can be used by learners and instructors alike.

We are currently developing an adaptive content presentation and interface enhancement agent to customize the content presented to each learner based on their learning style. In addition, we will further extend our research to consider undesirable behaviors, such as not genuine users (i.e., cheating). These efforts will contribute to categories III and VI in the classification system presented by Romero et al. [25], respectively.

## References

- [1] Szabo, M. and Flesher, K.: CMI Theory and Practice: Historical Roots of Learning Management Systems, *Proc. World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2002 Chesapeake, VA: Association for the Advancement of Computing in Education (AACE'02)*, pp.929–936 (2002).
- [2] Moodle Learning Management System (online), available from <http://www.moodle.org/> (accessed 2015-08-24).
- [3] Graf, S. and List, B.: An evaluation of open source e-learning platforms stressing adaptation issues, *Proc. Fifth IEEE International Conference on Advanced Learning Technologies (ICALT'05)*, pp.163–165 (2005).
- [4] De Crook, M., Mofers, F., Veen, M.V., Rosmalen, P.V., Brouns, F., Boticario, J.G., Barrera, C., Santos, O., Ayala, A., Gaudioso, E., Hernández, F., Arana, C., Trueba, I., Fuentes, C. and Rodrigo,



- M.D.M.: *Active learning for adaptive internet—state of the art, project deliverable report* (online), available from [https://adenu.ia.uned.es/alfanet/reports/ALFANET\\_D12.pdf/](https://adenu.ia.uned.es/alfanet/reports/ALFANET_D12.pdf/) (accessed 2015-08-24).
- [5] García, P., Amandi, A., Schiaffino, S.S. and Campo, M.: Evaluating Bayesian networks precision for detecting students' learning styles, *Computers & Education*, Vol.49, No.3, pp.794–808 (2007).
  - [6] Graf, S. and Kinshuk: Enabling Learning Management Systems to Identify Learning Styles, *Proc. Conference ICL (ICL'06)*, pp.1–9 (2006).
  - [7] Cha, H.J., Kim, Y.S., Park, S.H., Yoon, T.B., Jung, Y.M. and Lee, J.H.: Learning style diagnosis based on user interface behavior for the customization of learning interfaces in an intelligent tutoring system, *Proc. 8th International Conference on Intelligent Tutoring Systems (ITS'06)*, pp.513–524 (2006).
  - [8] Surjono, H.D.: The Evaluation of a Moodle Based Adaptive e-Learning System, *International Journal of Information and Education Technology*, Vol.4, No.1, pp.89–92 (2014).
  - [9] Savic, G. and Konjovic, Z.: Learning Style Based Personalization of SCORM E-learning Courses, *Proc. 7th International Symposium on Intelligent Systems and Informatics (SISY'14)*, pp.349–353 (2014).
  - [10] Park, H., Darak-ri, I.G. and Chungbuk, G.C.: Design and development of a mobile learning management system adaptive to learning style of students, *Proc. IEEE International Workshop on Wireless and Mobile Technologies in Education (WMTE'05)*, pp.4–6 (2005).
  - [11] Klasnja-Milicevic, A., Vesin, B., Ivanovic, M. and Budimac, Z.: E-Learning personalization based on hybrid recommendation strategy and learning style identification, *Computers & Education*, Vol.56, No.3, pp.885–899 (2011).
  - [12] Despotović-Zrakić, M., Markoviae, A., Bogdanoviae, A., Baraa, D., Sciences, O. and Iliaea, J.: Providing Adaptivity in Moodle LMS Courses Adaptive e-learning systems, *The Journal of Educational Technology & Society*, Vol.15, No.1, pp.326–338 (2012).
  - [13] Dung, P.Q. and Florea, A.M.: An approach for detecting learning styles in learning management systems based on learners' behaviours, *Proc. International Conference on Education and Management Innovation (ICEMI'12)*, pp.171–177 (2012).
  - [14] Pitigala Liyanage, M.P., Gunawardena, K.S.L. and Hirakawa, M.: Using Learning Styles to Enhance Learning Management Systems, *International Journal on Advances in ICT for Emerging Regions (ICTER'14)*, Vol.7, No.2, pp.1–10 (2014).
  - [15] Honey, P. and Mumford, A.: *The Learning Styles Helper's Guide*, Peter Honey Publications Ltd., Maidenhead, p.1 (1992).
  - [16] Felder, R.M. and Soloman, B.A.: *Felder-Soloman Index of Learning Styles* (online), available from <http://www4.ncsu.edu/unity/lockers/users/f/felder/public/ILSPage.html/> (accessed 2015-08-24).
  - [17] Kanninen, E.S.S.I.: *Learning styles in virtual learning environments: Master of Science Thesis*, p.64 (2009).
  - [18] Graf, S.: *Adaptivity in Learning Management Systems Focussing on Learning Styles: Ph.D. Thesis: Vienna University of Technology*, p.192 (2007).
  - [19] Carver, C.A., Howard, R.A. and Lane, W.D.: Addressing different learning styles through course hypermedia, *IEEE Trans. Education*, Vol.42, No.1, pp.33–38 (1999).
  - [20] Romero, C., Ventura, S. and Pechenizkiy, M.: *Handbook of Educational Data Mining Business & Economics*, pp.94 (2010).
  - [21] The International Educational Data Mining Society (online), available from <http://www.educationaldatamining.org/> (accessed 2015-08-24).
  - [22] Chang, Y., Kao, W., Chu, C. and Chiu, C.: A learning style classification mechanism for e-learning, *Computers & Education*, Vol.53, No.2, pp.273–285 (2009).
  - [23] Kotsiantis, S.B., Pierrakeas, C.J. and Pintelas, P.E.: Preventing Student Dropout in Distance Learning Using Machine Learning Techniques, *Lecture Notes in Computer Science*, Vol.2774, pp.267–274 (2003).
  - [24] Graf, S., Kinshuk and Tzu-Chien, L.: 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 (2008).
  - [25] Romero, C. and Ventura, S.: Educational data mining: A review of the state of the art, *IEEE Trans. Syst. Man, Cybern. C Appl. Rev.*, Vol.40, No.6, pp.601–618 (2010).
  - [26] DBMiner (online), available from <http://www.dbminer.com/> (accessed 2015-08-24).
  - [27] Keel (online), available from <http://www.keel.es/> (accessed 2015-08-24).
  - [28] Weka (online), available from <http://www.cs.waikato.ac.nz/ml/weka/> (accessed 2015-08-24).
  - [29] RapidMiner (online), available from <https://rapidminer.com/> (accessed 2015-08-24).
  - [30] R Data mining (online), available from <http://www.rdatamining.com/> (accessed 2015-08-24).
  - [31] KNIME (online), available from <https://www.knime.org/> (accessed 2015-08-24).
  - [32] Jović, A., Brkić, K. and Bogunović, N.: An overview of free software tools for general data mining, *Proc. International Convention on Information and Communication Technology Electronics and Microelectronics (MIPRO'14)*, pp.26–30 (2014).
  - [33] Wahbeh, A.H., Al-radaideh, Q.A., Al-kabi, M.N. and Al-shawakfa, E.M.: A Comparison Study between Data Mining Tools over some Classification Methods, *International Journal of Advanced Computer Science and Applications*, Vol.2, No.8, pp.18–26 (2010).
  - [34] Moodle architecture (online), available from [https://docs.moodle.org/dev/Moodle\\_architecture/](https://docs.moodle.org/dev/Moodle_architecture/) (accessed 2015-08-24).
  - [35] Graf, S., Viola, S.R. and Kinshuk: Automatic Student Modeling for Detecting Learning Style Preferences in Learning Management Systems, *Proc. International Conference on Cognition and Exploratory Learning in Digital Age (CELDA'07)*, pp.172–179 (2007).
  - [36] Graf, S. and Kinshuk: Analyzing the Behaviour of Students in Learning Management Systems with respect to Learning Styles, Advanced in Semantic Media Adaptation and Personalization, *Springer Series on Studies in Computational Intelligence*, Vol.93, No.1, pp.53–74 (2008).
  - [37] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P.: SMOTE: Synthetic minority over-sampling technique, *Journal of Artificial Intelligence Research*, Vol.16, pp.321–357 (2002).
  - [38] Festinger, L.: A theory of social comparison process, *Human Relations*, Vol.7, pp.114–140 (1954).

## Appendix

### A.1 Tables for Analysis of Learning Style in Moodle

Type of Table	Moodle Table	Attributes Extracted
Original	mdl_user	User identification number
	mdl_course	Course identification number
	mdl_log	User-performed activities in Moodle LMS
	mdl_quiz	No. of contents, outlines, examples, exercises, self-assessments available; no. of times content visit, outline visit, example visit, exercise visit, self-assessment visit; no. of correctly answered questions about details, overview knowledge, facts, concepts, graphics, text, interpreting solutions, developing new solutions; time spent on self-assessment tests, exercise, examples
	mdl_resource	No. of contents, outlines, examples, exercises, self-assessments available; no. of times content visit, outline visit,
	mdl_scorm	example visit, exercise visit, self-assessment visit
	mdl_forum	No. of forums available, no. of times forum viewed, time spent on forum
	mdl_question_attempts	No. of times giving wrong answer for the same quiz twice; no. of correctly answered questions about details, overview knowledge, facts, concepts, graphics, text, interpreting solutions, developing new solutions
	mdl_question_attempt_steps	
	mdl_question	No. of correctly answered questions about details, overview knowledge, facts, concepts, graphics, text, interpreting solutions, developing new solutions
	mdl_quiz_question_instances	
	mdl_quiz_attempts	Time spent on self-assessment tests, exercise, examples
Newly Created	mdl_lec_threshold	Instructor's recommended thresholds for course activities.
	mdl_dimensions	Student's average ratio for each learning style ( $R_{AVG}$ )
	mdl_ILS_tracking	ILS questionnaire data pertaining to a student, learning style predicted by ILS
	mdl_ILS_value	



#### Madura Prabhani Pitigala Liyanage

received her B.Sc. from the University of Sri Jayewardenepura, Sri Lanka in 2006 and Masters from the University of Colombo, Sri Lanka in 2011. She is currently a Doctoral student at the Interdisciplinary Graduate School of Science and Engineering, Shimane University, Japan. Her research interests include human-computer interaction and e-learning. She is a member of the Internet Society.



#### K.S. Lasith Gunawardena

received his B.Sc. from the University of Sri Jayewardenepura, Sri Lanka in 1999 and M.Sc. from the University of Colombo, Sri Lanka in 2009. He is currently a Doctoral student at the Interdisciplinary Graduate School of Science and Engineering, Shimane University, Japan. His research interests include human-computer interaction and e-learning. He is a member of the IEEE Computer Society, ISOC, and ACM.



#### Masahito Hirakawa

graduated from Hiroshima Institute of Technology in 1979, and received his M.E. and Ph.D. degrees from Hiroshima University in 1981 and 1984, respectively. He has been a professor at Shimane University since 2002. His research interests include human-computer interaction and e-learning. He is a member of IPSJ, IEICE, and ACM, and a senior member of IEEE.