Predicting learning styles based on students' learning behaviour using correlation analysis

Li Ling Xiao* and Siti Soraya Binti Abdul Rahman

Department of Artificial Intelligence, Faculty of Computer Science and Information Technology, Universiti of Malaya, Malaysia

Past research has proposed various approaches to automatically detect students' learning styles to address problems associated with traditional research methods (i.e. questionnaire). However, results obtained through traditional research methods have issues in terms of accuracy and precision which need to be addressed. In general, the existing automatic detection approaches are only able to provide satisfactory results for specific learning style models and/or dimensions, or even only work for certain learning management systems. The aim of this study is to propose an automatic detection of learning styles from the analysis of students' learning behaviour by constructing a mathematical model. This study specifically explores the relationship between students' learning behaviour and their learning styles. To investigate this relationship, a pilot experiment was conducted with 33 students. The students used Moodle platform, a learning management system, as supplementary online learning material for Java programing. The students' learning behaviour was tracked and recorded. Thirty students' data (i.e. their learning behaviour and learning styles; measured using the Index of Learning Styles (ILS) instrument) were analysed using the proposed correlation analysis to identify the relationship. The remaining three students' learning behaviour data were used to predict their learning styles. The findings are discussed with regard to accuracy of automatic detection of learning styles using the ILS instrument.

Keywords: Automatic learning style assessment, learning behaviour pattern, student modelling.

E-learning is the use of computer technology to transmit information to individuals[1]. However, many e-learning systems do not take individual differences into consideration. These differences include the ability of learners, background, goal, knowledge foundation and learning style[2,3]. Several studies[4,5] considered learning style as an important factor in determining learning effectiveness during the learning process.

To overcome these constraints, the same learning contents were provided to different learners. The learning contents were also provided to adapt to students' requirements and needs[1]. Detection methods play an important role in adapting to the learning environment. There are several limitations in the detection methods of the current learning style. For instance, learning style detection can only be applied in a specific model of learning style (LS) dimension and cannot be adjusted to adapt to LS preferences in a different learning environment. In this study, a new approach was introduced to automatically detect learning styles.

The key contributions of this paper are as follows: (i) Review past detection methods and approaches in automatically detecting learning styles, especially the accuracy and precision of the detected results, learning style and learning management system (LMS) compatibility; and (ii) propose a new method to detect learners' learning styles based on their learning behaviour patterns. In this study a direct link of a mathematical model is constructed between learners' learning behaviour patterns and their learning style preferences. The rest of the paper is organized as follows: in next section, related work from literature review is discussed. We have then explained the research framework of the proposed approach. Experiment of the proposed approach is then presented along with recommendation for future work and discussion.

Related work

The traditional approach of learning style identification requiring students to fill up questionnaires is rather simple and carried out manually. The drawbacks of using questionnaires are that the questions are fixed and students may tend to answer questions arbitrarily. Other challenges of traditional learning style identification include identifying students' lack of motivation and self-awareness about their learning preferences[6,7]. Therefore, a precise and accurate way of identifying learning styles is needed.

The automatic learning styles detection method is designed to solve problems in the traditional questionnaire in order to avoid intentional or unintentional inaccurate answers and to save students' time in filling up questionnaires. These current detection methods use attributes such as personality factors, behavioural factors and time[8,9].

The automatic detection of learning styles has been gaining significance over time in the e-learning field as it

*For correspondence. (e-mail: litc0766@gmail.com)

has advantages over traditional approaches[10]. Some of the advantages of utilizing the automatic learning style detection approach are: the approach does not require additional effort from students; the approach is free of uncertainties; it uses real data to detect students' learning styles which ensures and enhances accuracy; and it adds convenience in recognizing and updating the changes in students' learning style preferences[9].

Recent research in automatic detection of learning styles[10–13] use literature-based approach by designing a deterministic interface system based on pre-defined behavioural patterns in order to deduce students' learning styles. Some restrictions found in this approach are: uncertainty, difficulty and complexity of developing the implementing rules which are not able to infer learning styles effectively from students' actions; and the way it treats students' behaviour as evidences and not as possibilities.

Meanwhile, some studies[14–16] proposed data-driven approaches, including Bayesian networks, neural networks, fuzzy models, etc. The drawbacks of these approaches are computational cost and high complexity. Besides, in general, these approaches are difficult to reuse in other systems as the system and the entire learning process are highly coupled and integrated. In some of these approaches, once the students' learning styles are detected, the results remain static throughout the whole learning process[17].

The approach proposed in this paper is based on correlation analysis, which has the fundamental characteristics of incremental learning and avoids using specific knowledge of the application domain, making the method more generalized and easy to reuse.

Proposed approach

Mathematical model

There are almost 71 different learning style models that state[18] that, no matter how a learning style model is built, it is always perceived differently by students as they can derive their own learning preferences from various learning style models. It is called 'custom model' and incorporates characteristics from one or several traditional learning style models to form a new learning style model[7]. Custom models could cover plenty of learning preferences and are easy to extend with new learning style dimensions. The critical concern is the way to identify the exact nature of students' learning style in order to provide suitable adaptive learning material. Therefore, to resolve these issues, a resolution approach is taken, whereby each dimension is considered as a semantic relation, building relationship with student i and j, and the weight a_{ij} representing similarity of student i and j on this dimension (see Figure 1).

Assuming n students' learning style is already known, Matrix M^t to map the weight of the semantic relation is constructed as follows. M^t is $n * n$ matrix, a_{ij}^t is an element in M^t, then the value of a_{ij}^t

$$a_{ij}^t = (e - e^{\sqrt{|p_i^t - p_j^t|}})/(e-1), \tag{1}$$

$p_i^t \in [0, 1]$ denotes the preference values of student i on dimension t. The above formula indicates the similarity of student i and j on the learning style dimension t. If there are more similarities, the corresponding semantic relation is more relevant, and vice versa. Therefore, if student i is a pure active learner and student j is a pure reflective learner, then $p_i^{\text{active/reflective}} = 0$, $p_j^{\text{active/reflective}} = 1$, and $a_{ij}^{\text{active/reflective}} = 0$, indicating that there is no correlation

$$a_{ij}^t = \frac{e - e^{\sqrt{|p_i^t - p_j^t|}}}{e-1} = \frac{e - e^{\sqrt{|0-1|}}}{e-1} = 0.$$

If student i and j are pure active learners, $p_i^{\text{active/reflective}} = p_j^{\text{active/reflective}} = 0$, then $a_{ij}^{\text{active/reflective}} = 1$

$$a_{ij}^t = \frac{e - e^{\sqrt{|p_i^t - p_j^t|}}}{e-1} = \frac{e - e^{\sqrt{|0-0|}}}{e-1} = 1.$$

If two learners' combined preference values are >0.5, then the correlation similarity will be <0.5.

The correlation matrix table is used to investigate the dependency between multiple variables at a given time. The table contains specific correlation coefficients between each student and others on a specific learning style dimension, assuming that more similar the learning

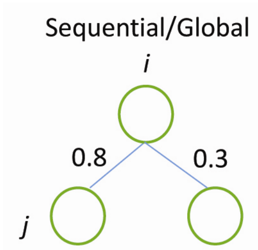

Figure 1. Sematic relation of i and j.

Table 1. Correlation matrix M^t

	i	j	k	l	…
i		a_{ij}^t	a_{ik}^t	a_{il}^t	…
j			a_{jk}^t	a_{jl}^t	…
k				a_{kl}^t	…
l					…
…					

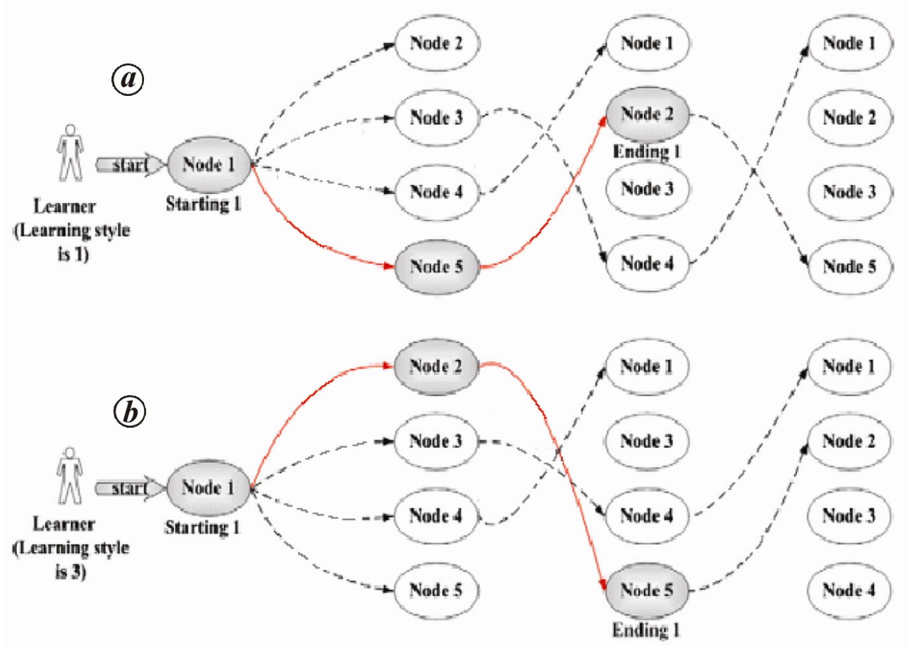

Figure 2. Learning path to avoid spare matrix.

style preference, the higher the correlation of their learning behaviours. Once the learning behaviour matrices of these students are built, a potential relationship or link between their learning styles and behaviour patterns could be discovered. Based on a known link, when new learners join in, their learning style preference could be deduced from behaviour patterns, instead of learning questionnaires.

Detection of learning style

Assuming there are m unknown learning styles for new comers, replacing m students from n known learning styles to build a new student group, its behaviour matrix is M, m_{ij} are the elements of M.

$$m_{ij} = \begin{cases} 1, \text{student } i \text{ and } j \text{ have same behaviour sequence} \\ 0, \text{other.} \end{cases}$$

In behaviour sequence, set $A = \{a|a \in L \cup B\}$, set $L = \{o_1, o_2, \ldots\}$ is denoted for learning objects, and set $B = \{b_1, b_2, \ldots\}$ for behaviours. Examples of learning behaviours are 'participating in a discussion forum', 'doing an exercise', etc.

Naturally, with learning objects existing in large quantities, the behaviour sequence, A, becomes varied, resulting in a sparse matrix. In order to prevent sparse matrices, the approach proactively provides feasible behaviour sequences, which guide learners from different paths as shown in Figure 2; but will not contribute to learning styles and will have no effect on learners' preference.

Consider behaviour matrix M as an image representation matrix. In order to detect m students' learning style, we need to discover the role of learning style preference play in the behaviour matrix. Then, based on the known learning style correlation matrix, near-to-similar combination of the behaviour matrix can be determined. Subsequently, the behaviour matrix is abstracted into a nonlinear programming problem. Setting vector $X = (x_1, x_2 \ldots x_t)$ for newcomers' learning style where t is the dimension of learning style, the objective function is

$$\min f(x), \sum_{i=1}^{t} x_i^2 = 1, \tag{2}$$

$$f(x) = \operatorname{var}\left(M, \sum_{i=1}^{t} x_i \cdot M^i\right)$$
$$= \left(\sum_{i=1}^{n}\sum_{j=1}^{n}\left(m_{ij} - \sum_{h=1}^{t} x_h \cdot a_{ij}^t\right)^2\right)^{1/2}, \sum_{i=1}^{t} x_i^2 = 1. \tag{3}$$

Equation (2) is an objective function of abstract nonlinear programing, which is to minimize the differences of two matrices. $\sum_{i=1}^{t} x_i^2 = 1$ is the constraint condition of $(\forall x_i \in [0, 1])$. Since M and M^i are symmetric matrices function var() can be set to solve two $n(n-1)/2$ dimension vectors similarity. Objective function is equivalent to the maximum similarity of these two vectors. Thus, eq. (3) computes Euclidean distance of two vectors. Under optimal condition, $M = \sum_{i=1}^{T} x_i * M^i$, if there is an existing solution, then there is a unique solution of X.

Prediction of learning styles

The prediction assumes that once learners perform similar behaviour sequences, they have similar learning style preferences. Therefore, according to behavioural matrix M, and optimal solution vector X (vector components of vector X represent the extent of each learning style dimension's influence on the behavioural matrix), if m students' learning style preference on each dimension is p_i^t, $0 < i \leq m$, then the estimation prediction values are

$$\widehat{p_i^t} = p_{j_{m_{ij}=1}}^{\overline{t}}. \tag{4}$$

Equation (4) represents the average value of student j who has the same behaviour sequence as student i on learning style dimension t. The prediction values are recorded in the form of probability, x_t, indicating that learning style dimension, t plays a decisive role in the behaviour matrix (x_t close to 1), and eventually the learning style value on this dimension is predicted with clustering. These prediction results are closer to the current learning environment. In addition, once the learning environment changes, it is still possible to obtain prediction values on different dimensions of learning style. This proposed adaptive learning style prediction is more accurate than a mere log analysis approach.

Verification and validation of the proposed approach

The proposed student learning style modelling approach was evaluated using a course on JAVA programing. Thirty-three students participated in this pilot experiment. Moodle platform was used as an LMS and its tracking mechanism was used to record students' learning behaviours.

Material and methodology

In this approach, the study concentrates on the Felder–Silverman learning style model (FSLSM) because authors provide a questionnaire and a complete guide to answer it. Moreover, this model has proven to be effective in many adaptive learning systems[19–21]. Additionally, it is easy to benchmark comparisons in the analysis of future results.

To evaluate the proposed approach, a total of 30 from 33 students were required to answer the Index of Learning Styles (ILS) questionnaire before taking the course. The questionnaire was developed by Felder and Silverman for identifying students' learning style. These 30 students interacted with LMS for a JAVA programing course. Their learning behaviours were tracked and recorded. They were considered as 'existing students' in the database. The remaining three students (10% of total amount of students) were replacement students, who acted as newcomers. They were enrolled into the course directly, and only accessed their learning style questionnaires at the end of the experiment to compare the auto detection result and questionnaire consequence for validation purpose.

Labelling learning objects

Each learning object was labelled with one subtype of any elements in the set of 16 types of combinations from four learning style dimensions: sensing/intuitive, visual/verbal, active/reflective, and sequence/global. For example, learning object 1 is labelled as active/sensing/visual/sequential, while learning object 2 is only labelled as visual. Based on the theoretical descriptions about learning style characteristics of Felder–Silverman, and based on the practical research[19,22,23], the learning objects in the pilot experiment are labelled as described in Table 2.

Results

In this pilot experiment, two dimensions of Felder–Silverman model were assessed according to Felder's scale (as shown in Figure 3), including input (visual/verbal) and understanding (sequential/global) dimensions. The scale proposed by Felder ranks students in different levels as ±1, ±3, ±5, ±7, ±9 and ±11 within one dimension. For example, a student categorized as '–9' in understanding dimension has clear sequential behaviour. On the other hand, a student categorized as '+11' in understanding dimension shows a strong global behaviour. Once students' learning style scale are obtained, we must map this scale proposed by Felder to the values of correlation matrix. Thus, we should match this scale to matrix values by calculation and convert to the range of [0, 1] that enable to build the correlation matrix.

The value distribution of these 30 students on sequential/global and visual/verbal dimensions is shown in Figure 4, and the corresponding correlation matrix is depicted in Figure 5.

The scatterplot matrix displays the similarity by Tanimoto coefficient values of m_{ij}. Figure 6 *a* shows the comparison behaviour matrix which included newcomers (10% of the total number of students were replaced). It was discovered that only learning style correlation matrix M_2 is most similar with the behaviour matrix. With regard to the hypothesis 'the more similar the learning style

Table 2. Labels of learning objects in the experiment

Active	Reflective	Sensing	Intuitive	Visual	Verbal	Sequential	Global
Exercises, self-assessment, multiple-choice question exercises	Examples, Outlines, Summaries, result pages	Examples, explana-tion, facts, practical material	Definitions, algorithms	Images, graphics, charts, animations, videos	Text, audio	Step-by-step exercises, constrict link pages	Outlines, summaries, all-link pages

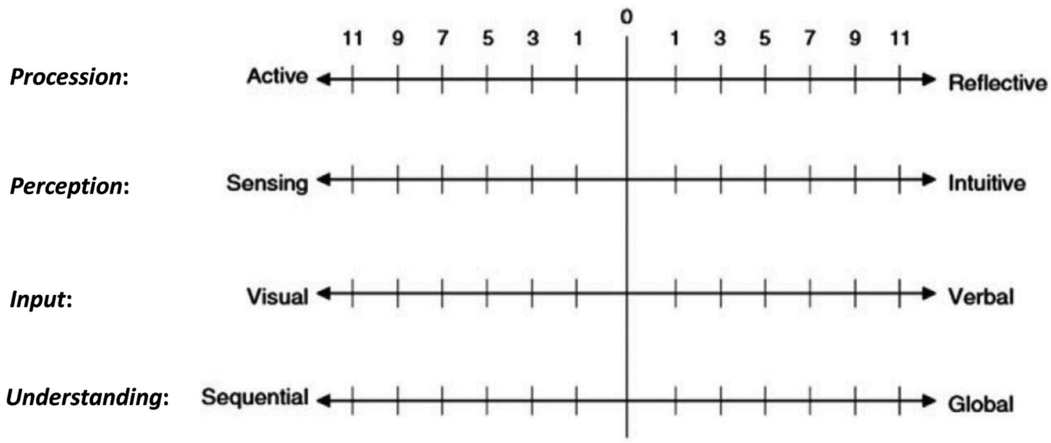

Figure 3. Index of learning style (ILS).

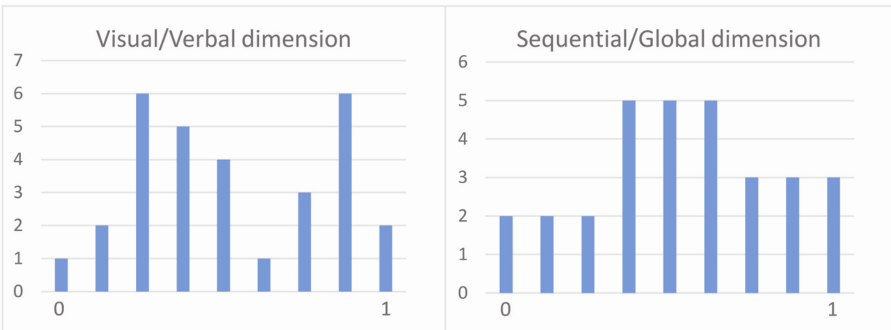

Figure 4. Value distribution of 30 students' learning style.

preference, the higher the similarity of their learning behaviours', it can be deduced that the sequential/global dimension has a more important role in the learning process.

Therefore, the values of newcomers' learning style are calculated by the proposed approach which is presented in Figure 6 *b*. Figure 6 *b* shows a strong similarity between the behaviour matrix and correlation matrix made by the optimal solution x_i. The comparison matrix $M_x = M - \sum_{i=1}^{3} x_i * M^i$, where x_i is the optimal solution. The learning style values of newcomers x_i are (0.919 0.864 0.096).

When mapping the detected values to the ILS scale, the results indicate that two newcomers were categorized as 'global 9 (+9)' and one newcomer was categorized as 'sequential 9 (–9)' in the understanding dimension. The results were compared with replacement students' learning style questionnaire preference value to verify the validity of results. Table 3 shows positive result, indicating that the proposed approach is appropriate for identifying learning styles.

To assess the precision of the approach, eq. (5) by García *et al.*[21] was used when Sim is 0 if the values calculated with proposed approach and ILS are opposite, 1 if the values are equal, and 0.5 if one is neutral and the other is an extreme value; and n is the number of students

$$\text{Precision} = \frac{\sum_{1}^{n} \text{Sim}(LS_{\text{determined}}, LS_{\text{ILS}})}{n}. \quad (5)$$

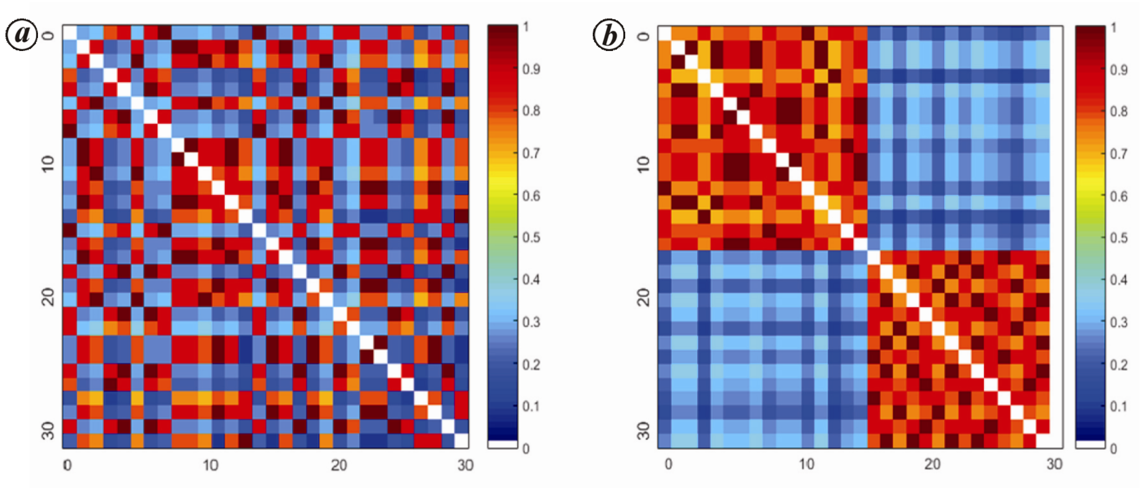

Figure 5. Correlation matrix of visual/verbal (***a***, M_1) and sequential/global (***b***, M_2) dimensions.

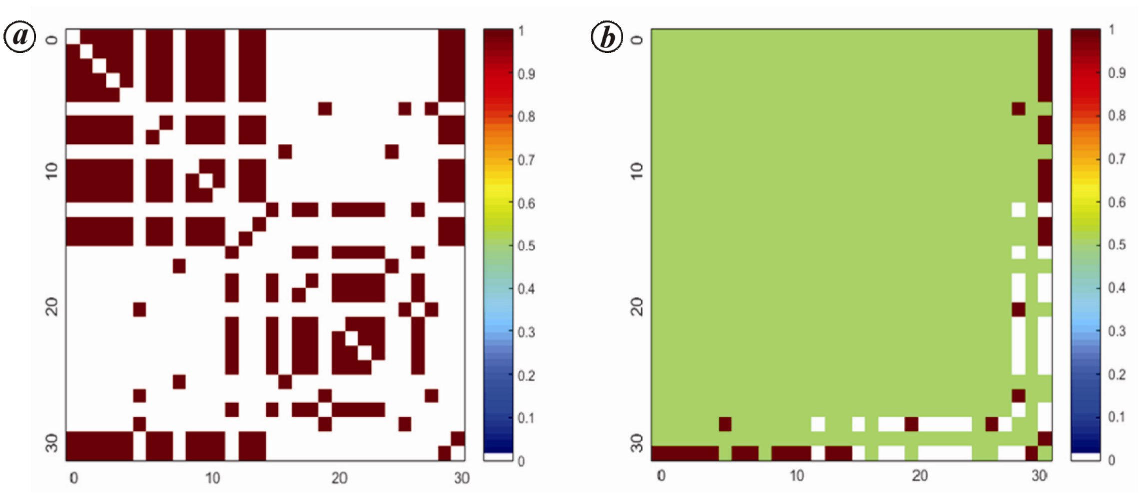

Figure 6. Behaviour matrix M (***a***) and similarity comparison matrix M_X (***b***).

Table 3. Comparison between questionnaire and detection

	1	2	3
ILS questionnaire	Global 9	Global 7	Sequential 11
Detected result	0.969	0.864	0.116

Due to the small scale of the current pilot experiment, the above mentioned procedure will be conducted in future experiments with a larger number of participants and wider range of learning style dimensions.

Conclusion and future work

The proposed approach attempts to detect learners' learning style automatically. The detection method is based only on indications gathered from the learners' behaviour during an online course, more specifically, by constructing dynamic behaviour matrix and learning style correlation matrix that reflect the relationship of current learning style dimensions.

Future research will be carried out to test the approach for precision and accuracy of detected results to firmly validate it. In the current study, the number of replacement students was limited to 10% of the total number of students for getting a better result. If this approach is applied in an open environment with large data, the accuracy of the result can be significantly improved. Further, future research could attempt to study custom learning style models. There will be an experiment on custom models conducted in future. The performance analysis and benchmarking comparison are scheduled for ongoing work.

1. Qutechate, W. A., Almarabeh, T. and Alfayez, R., E-learning system in the University of Jordan: problem solving case study. *J. Theor. Appl. Inf. Technol.*, 2013, **53**(3), 353–358.

2. Surjono, H. D., The evaluation of a Moodle based adaptive e-learning system. *Int. J. Inf. Educ. Technol*., 2014, **4**(1), 89–92.
3. Yarandi, M., Jahankhani, H. and Tawil, A.-R. H., Towards adaptive E-learning using decision support systems. *Int. J. Emerg. Technol. Learn*., 2013, **8**(1), 44–51.
4. Bajraktarevic, N., Hall, W. and Fullick, P., Incorporating learning styles in hypermedia environment: Empirical evaluation. In *Proceedings of the Workshop on Adaptive Hypermedia and Adaptive Web-based System* spp., 2003, pp. 41–52.
5. Kinshuk, G. S., Providing Adaptive Courses in Learning Management Systems with Respect to Learning Style, the World Conference on E-Learning in Corporate, Government Healthcare, and Higher Education. AACE, 2007, pp. 2576–2583.
6. Dung, P. Q. and Florea, A. M., An approach for detecting learning styles in learning management systems based on learners' behaviours. *Int. Proc. Econ. Dev. Res*., 2012, **30**, 171–177.
7. Feldman, J., Monteserin, A. and Amandi, A., Automatic detection of learning styles: state of the art. *Artif. Intell. Rev*., 2015, **44**(2), 157–196.
8. Ahmad, N., Tasir, Z., Kasim, J. and Sahat, H., Automatic detection of learning styles in learning management systems by using literature-based method. *Procedia – Soc. Behav. Sci.*, 2013, **103**, 181–189.
9. Dung, P. Q. and Florea, A. M., A Literature-based Method to Automatically Detect Learning Styles in Learning Management Systems. Paper presented at the Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics, 2012, p. 46.
10. Graf, S., Using cognitive traits for improving the detection of learning styles. In Database and Expert Systems Applications (DEXA), 2010 Workshop on IEEE, 2012, pp. 74–78.
11. Graf, S. and Liu, T. C., Supporting teachers in identifying students' learning styles in learning management systems: an automatic student modelling approach. *J. Educ. Technol. Soc*., 2009, **12**(4), 3.
12. Cha, H. J., Kim, Y. S., Park, S. H., Yoon, T. B., Jung, Y. M. and Lee, J. H., Learning styles diagnosis based on user interface behaviors for the customization of learning interfaces in an intelligent tutoring system. In *Intelligent Tutoring Systems*, Springer, Berlin, Heidelberg, 2006, pp. 513–524.
13. Limongelli, C., Sciarrone, F., Temperini, M. and Vaste, G., Adaptive learning with the LS-plan system: a field evaluation. *IEEE Trans. Learn Technol*., 2009, **2**(3), 203–215.
14. Cabada, R. Z., Estrada, M. L. B. and Garcia, C. A. R., A fuzzy-neural network for classifying learning styles in a Web 2.0 and mobile learning environment. In *Web Congress*, LA-WEB'09. Latin American, 2009, pp. 177–182.
15. García, P., Amandi, A., Schiaffino, S. and Campo, M., Evaluating Bayesian networks' precision for detecting students' learning styles. *Comput. Educ*., 2009, **49**(3), 794–808.
16. Zatarain-Cabada, R., Barrón-Estrada, M. L., Zepeda-Sánchez, L., Sandoval, G., Osorio-Velazquez, J. M. and Urias-Barrientos, J. E., A Kohonen network for modeling students' learning styles in web 2.0 collaborative learning systems. In *MICAI 2009: Advances in Artificial Intelligence*, Springer, Berlin, Heidelberg, 2009, pp. 512–520.
17. Castillo, G., Gama, J. and Breda, A. M., An adaptive predictive model for student modeling. *Advances in Web-based Education: Personalized Learning Environments*, Information Science Publishing, Hershey, 2005, p. 7092.
18. Özyurt, Ö., Özyurt, H. and Baki, A., Design and development of an innovative individualized adaptive and intelligent E-learning system for teaching–learning of probability unit: details of UZWEBMAT. *Expert Syst. Appl*., 2013, **40**(8), 2914–2940.
19. Hong, H. and Kinshuk, D., Adaptation to student learning styles in web based educational systems. In *EdMedia: World Conference on Educational Media and Technology*, Association for the Advancement of Computing in Education (AACE), Waynesville, NC, 2004, pp. 491–496.
20. Graf, S. and Lin, T., Analysing the relationship between learning styles and cognitive traits. In *Seventh IEEE International Conference on Advanced Learning Technologies, ICALT*, 2007, pp. 235–239.
21. García, P., Amandi, A., Schiaffino, S. and Campo, M., Evaluating Bayesian networks' precision for detecting students' learning styles. *Comp. Educ*., 2007, **49**(3), 794–808.
22. Kinshuk, G. S. and Tzu-Chien, L., *Identifying Learning Styles in Learning Management Systems by Using Indications from Students' Behaviour*. Paper Presented at the Eighth IEEE International Conference Advanced Learning Technologies, 2008. ICALT'08, 2008, pp. 482–486.
23. Popescu, E., Trigano, P. and Badica, C., Relations between learning style and learner behavior in an educational hypermedia system: An exploratory study. In *Eighth IEEE International Conference on Advanced Learning Technologies, ICALT'08*, 2008, pp. 725–726.

Received 26 October 2016; revised accepted 30 May 2017

doi: 10.18520/cs/v113/i11/2090-2096