Control and Intelligent Systems, Vol. 43, No. 2, 2015

THE DISCRIMINATION OF LEARNING STYLES BY BAYES-BASED STATISTICS: AN EXTENDED STUDY ON ILS SYSTEM

Yanping Jing,[*,**] Bo Li,[*,**] Na Chen,[*] Xiaofeng Li,[*] Jie Hu,[*,***] and Feng Zhu[*,***]

Abstract

Educational data mining (DM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from the educational context. As one of the methods, Index of Learning Styles (ILS) system is well designed to identify personalized learning style. However, a remarkable discrepancy between the results of ILS system and participants' self-estimation indicates a relatively constrained applicable range of ILS in learning style identification. In this study, we focused on working out data-mining methods to extend applicable range of ILS, which is achieved by constructing a new questionnaire system and applying novel DM methods to a group of participants. According to our analysis, Bayes-based statistics are found to be effective in distinguishing ILS classes, and a newly constructed classification system – tree map – can help to distinguish learning style for samples from ILS "neutral" class. Therefore, the DM technique applied in this study can be an effective method for enlarging the applicable range of traditional ILS system.

Key Words

Bayes-based statistics, index of learning styles, data mining, machine learning

1. Introduction

Since the 1960s, data mining techniques (DMTs) have formed a branch of applied artificial intelligence. As data mining (DM) allows a search, for valuable information, from large volumes of data [1], it has been applied to different research fields such as e-commerce, bioinformatics [2] and lately, the educational research which commonly known as educational data mining (EDM) [3]. EDM often stresses with the improvement of student models which denote students' knowledge, motivation, meta-cognition and attitudes [4]. It should be noticed that DM is, at this point, not a solution itself, instead it is a tool which supports decision-making process.

In recent years, educational researchers [5]–[7] notice that learning style can be recognized as an indicator of potential learning success. It is considered as the significantly foundational issue in technological education. Learners' learning styles can be captured by a formal questionnaire [8], which are noted as the simplest measure of learning styles, while they are onerous [9]. As a validated tool for Felder–Silverman learning styles detection, the Index of Learning Styles (ILS) [10] is considered as a well-designed system in personalized learning style identification and has been frequently used in many EDM works [11], [12].

However, according to the analysis in previous works [11], [12] and this study, we have found a remarkable discrepancy between the results of ILS system and participants' self-estimation. Moreover, many studied candidates (about half in previous work [11], [12] and above half in this study) have been assigned into the ILS neutral class, which indicate a relatively constrained applicable range of ILS in learning style identification. Therefore, in this study, we focused on extending ILS's applicable range by applying novel data-mining methods: Bayes-based statistics to 156 first-year undergraduates. Our results show that the data-mining technique is able to enlarge the applicable range of traditional ILS system in learning style identification.

2. Literature Review

This section briefly describes how the DM is utilized in educational research and how a student's learning style is detected through the ILS questionnaire.

2.1 Data Mining

DM is a computer-based information system [1], which allows a search for valuable information from large

* Innovative Drug Research Centre & School of Foreign Languages and Cultures, Chongqing University, Chongqing, China; e-mail: {jingyp, libcell, chenn, lixf, hu.jie, zhufeng}@cqu.edu.cn
** Authors contribute equally to this work.
*** The first corresponding author: Asso. Prof. Jie Hu (hu.jie@cqu.edu.cn); The second corresponding author: Prof. Feng Zhu (zhufeng@cqu.edu.cn).
Recommended by Prof. Maoqing Li
(DOI: 10.2316/Journal.201.2015.2.201-2666)

volumes of data. Therefore, it is devoted to store huge data repositories, generate information and discover knowledge. DM aims at finding data patterns, discovering hidden features and structure association rules, estimating sophisticated values for classification and composing clusters of homogenous objects, from the large and complex datasets. In sum, DM's outcomes represent a valuable support for decisions-making.

Recently, EDM becomes an emerging discipline, concerned with developing methods for exploring mega data from the educational context. DM has been utilized in different types of educational systems. On the one hand, traditional face-to-face classroom contexts, such as special education and higher education, are utilized [13]; on the other hand, computer-based education and web-based education such as learning management systems [14], web-based adaptive hypermedia systems [15] and intelligent tutoring systems [16] are widely studied. The main difference between these two kinds of utilization is the available data in each context. Traditional classrooms only have information about student attendance, course information, curriculum goals and individualized plan data. However, computer- and web-based educations have been much more informative because these systems can record all the information about learners' actions and interactions onto databases.

2.2 The Index of Learning Styles (ILS)

Learning styles model the different ways that groups of people prefer to learn [10], [17]. A literature review of 70 references of learning style models was conducted from 2000 to 2011 [18]. It turned out that the most preferred learning style model was Felder–Silverman learning style dimensions, which was utilized in 35 studies (50%), followed by Felder–Silverman (17.1%), Kolb [19] (8.6%), VARK (7.1%), Honey and Mumford (5.7%) and other individual models [6], [20]. Therefore, Felder–Silverman learning style model (FSLSM) was selected in this study not only because that it is suitable for basic science issue applications but also because that FSLSM is recognized as the most often used learning styles model in recent times. Some researchers even argue that FSLSM is the most appropriate model, as a detailed measurement of learning styles is crucial to identify relationships between learning styles and student performance [11].

Accordingly, the ILS [10] was selected in this research as the data collection instrument. The ILS is a 44-item forced-choice questionnaire with four dimensions related to learning style preferences. Each learning style dimension contains 11 questions, assigning a style for each dimension. For instance, in the information processing (active/reflective) dimension, there are two polar options in each question presenting a preference for a student with an active or reflective learning style. The ILS represents individual student's learning style in four dimensions, as illustrated in Fig. 1: processing (D1, active/reflective), perception (D2, sensory/intuitive), reception input (D3 visual/verbal) and understanding (sequential/global). The model posits that active students learn by trying things out and working well in groups; reflective students prefer to think things through and work better by themselves. Sensory students are practical and oriented towards facts, data and experimentation; intuitive students are more conceptual and oriented towards principles and theories. Visual students remember best what they see, such as pictures, diagrams or films; verbal students remember much of what they hear or read. Sequential students follow linear and orderly reasoning processes when solving problems; global students prefer to learn in intuitive leaps and may be unable to explain how they came up with solutions.

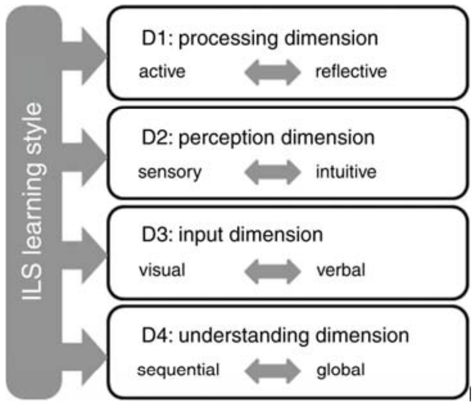


Figure 1. Four dimensions in ILS system.

Based on the analysis in former works [11], [12] and in this study, a remarkable discrepancy between the results of ILS system and participants' self-estimation was found, and many samples in these studies were grouped into ILS neutral class, which constrained the applicable range of ILS to learning style identification. In this study, we applied a new method, namely Bayes-based statistics, to a case study of 156 undergraduates to evaluate its capacity on extending traditional ILS system in learning style identification.

3. Data Collection

One hundred and fifty-six first-year undergraduate students from School of Foreign Languages and Cultures in Chongqing University (Chongqing city, P.R. China) were enrolled in this study. To increase the diversity and representativeness of samples, students collected in this paper major in three different disciplines including English Studies, Japanese Studies and German Studies. The statistics of these three groups of student samples are described as follows:

- Group 1: 94 students major in English Studies (62 females and 32 males).
- Group 2: 36 students major in Japanese Studies (20 females and 16 males).
- Group 3: 26 students major in German Studies (14 females and 12 males).

In this study, 156 students, between 16 and 21 years old, were enrolled to answer ILS questionnaires [10]. Nineteen out of all 156 questionnaires with missing or inappropriate answer to any question are excluded from the dataset for further analysis. Preliminary analysis on the

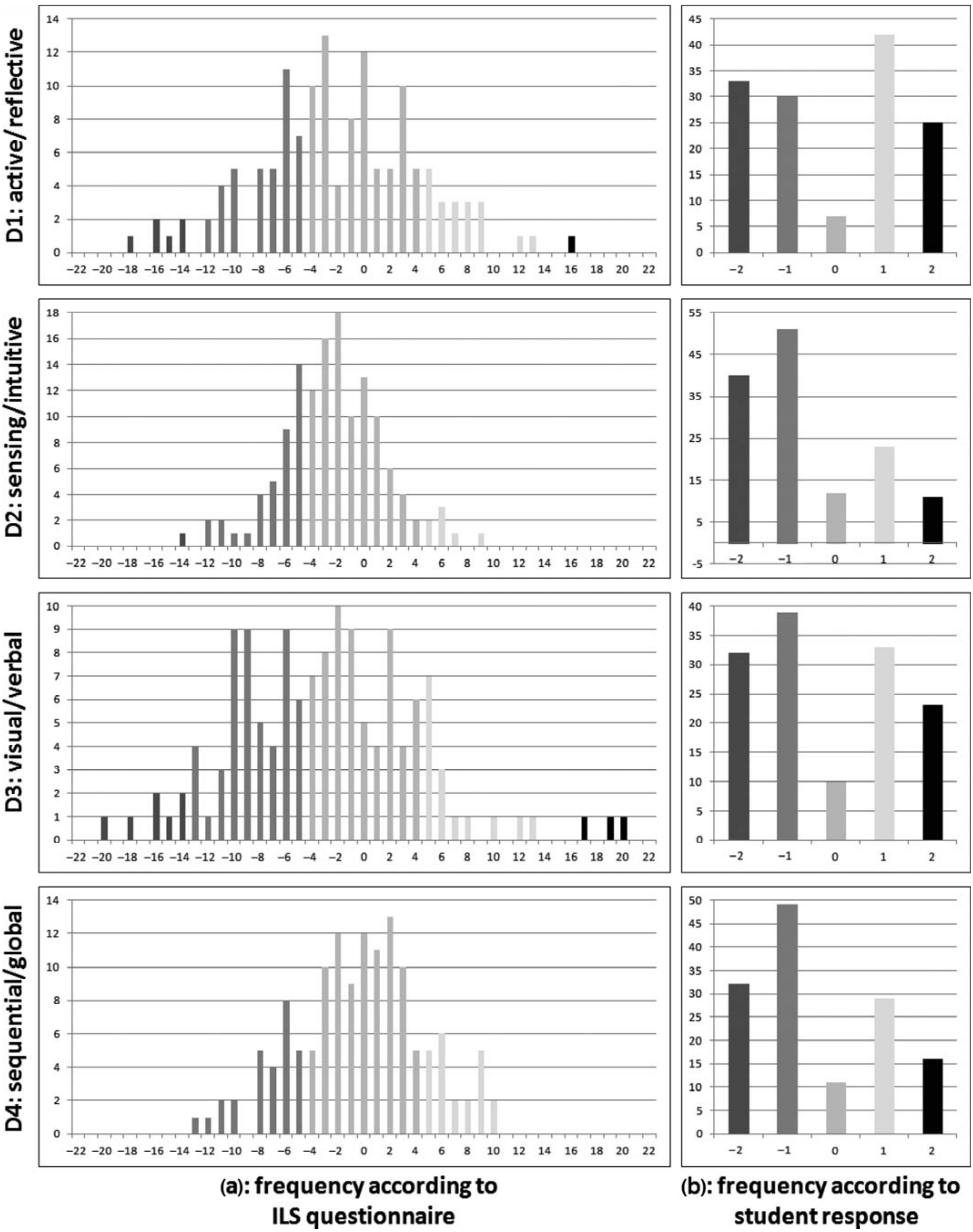


Figure 2. Frequency of students in ILS dimensions: (a) frequency is determined by ILS questionnaire, and (b) frequency is determined by students' self-estimation.

remaining 137 questionnaires show a basic understanding on the nature of data collected. Figure 2 illustrates the frequency of each ILS dimension for the student sample. D1–D4 refer to four different learning styles in ILS system including active/reflective (D1), sensory/intuitive (D2), visual/verbal (D3) and sequential/global (D4) dimensions, respectively. Colours in Fig. 2 distinguish students by five sample classes: negative class in dimgrey, marginal negative class in grey, neutral class in darkgrey, marginal positive class in lightgray and positive class in black. Definitions of each class can be found in Section 4.

4. Methods

In this study, a questionnaire was designed based on FSLSM [10]. Moreover, questionnaires are further distributed to and collected from 156 enrolled student samples; and student samples also indicate their intended learning style by self-estimation. According to answers to each questionnaire, student samples could be divided into sample classes, which would further be applied for selecting distinguishing questions for each dimension. Finally, selected questions for each dimension were applied to construct a novel learning style prediction model. Technical details about methodologies applied in this study are demonstrated in the following sub-sections.

4.1 Design of Questionnaire

Questionnaire used in this study was designed based on FSLSM [10] which is frequently used to distinguish personalized learning styles and help enhance learning efficiency [11], [12]. Our study inherited the sequence of questions used in Ortigosa's reputable study [12] as a starting point, but three major modifications were applied to construct a new questionnaire.

Firstly, comparing to 2 or 3 selectable question options in previous works [10], [12], the number of questions options in the questionnaire of this study was increased to 5. In previous works, only 2 or 3 polar question options were included to indicate whether the learner belongs to the certain learning style or not. However, learning style is a tendency. Instead of a simple "yes or no" issue, five question options were designed to create a gradually changed classification system: negative class (with a score of −2), marginal negative class (with a score of −1), neutral class (with a score of 0), marginal positive class (with a score of 1) and positive class (with a score of 2).

Secondly, to avoid influence by former question, the order of options were rearranged. In the original ILS questions, all the questions were regularly arranged to present positive answers on the one side and negative answers on the other side. There is a concern that participants may be influenced by habitual thinking when answering questions without active thinking. In this case, the questionnaire collected may not be reliable enough to reflect participants' real thoughts. To avoid influence by the former question when answering following questions, options were reordered and no more than two options were recognized as the same positive/negative class.

Thirdly, four dimension-clarification questions were added in, for participants to clarify their own learning style based on their self-estimation. Based on the original 44-question ILS, extra four questions were added into the new questionnaire designed, for participants to clarify their learning style intension by their own judgement. Each of the four questions matched to the certain learning style dimension. For instance, the first added question is that "When considering learning style, do you think you belong to active learner or reflective learner?".

4.2 Enrolment of Student Sample

One hundred and fifty-six first-year undergraduate students from School of Foreign Languages and Cultures in Chongqing University (Chongqing city, P.R. China) were enrolled in this study. All of them were informed that they were participating in a research analysis helping to distinguish their personalized learning style. Moreover, students were told that the analysis results would assist the university to design a more suitable and detailed educational program which might in turn help to improve the quality of questionnaire returned back.

4.3 Definition of Sample Classes

Definitions of each class in Fig. 2 are as follows: in part (a), negative class (in dimgrey from −22 to −14), marginal negative class (in grey from −13 to −5), neutral class (in darkgrey from −4 to 4), marginal positive class (in lightgrey from 5 to 13) and positive class (in black from 14 to 22); in part (b), negative class (in blue equal to −2), marginal negative class (in green equal to −1), neutral class (in yellow equal to 0), marginal positive class (in purple equal to 1) and positive class (in red equal to 2).

4.4 Bayes-based Statistic Analysis

In this study, the analysis begins with the raw data preprocessing, which includes data cleaning (such as missing values filling), integration, transformation and normalization. After data preprocessing, linear models and empirical Bayes-based algorithm [21] were used for feature selection from all 137 questionnaires by R package, limma [22]. This method employs an approach where certain parameter is inferred from the data (hence, empirical). As one of the earliest application cases of empirical Bayes-based linear model, DNA microarray data demonstrated higher prediction performance than the traditional Student *t*-test, as it provided more precise estimates of the statistical significance of each descriptor. As a result, 5, 9, 9 and 1 questions were identified as the most significant ones in D1, D2, D3 and D4, respectively, in distinguishing positive class from negative one.

Based on those significant questions in each dimension, a bi-directional hierarchical clustering analysis was conducted for all 137 student samples; clustering results were illustrated via heat map by using gplots package [23]. Finally, the sample classification was predicted by clustering models (tree map) as mentioned above. All data visualization and analysis were carried out in R language environment [23].

5. Results and Discussion

5.1 About Half Students Are Classified into Neutral Class by ILS System

There are two parts in Fig. 2: part (a) indicates the frequency of the student sample by the sum of answers to ILS questionnaire in each dimension, while part (b) indicates the frequency of the student sample by the student's intended learning style on their own judgement. As illustrated in part (a), frequencies of each dimension follow a similar distribution pattern shown in previously published works [12]. However, there is a remarkable discrepancy in patterns reflected by parts (a) and (b), which indicates a great "gap" between questionnaire analysis and students' self-estimation. In questionnaire analysis [part (a)], many students (52.6% for D1, 66.4% for D2, 45.3% for D3 and 63.5% for D4) were classified into neutral class, which limited the applicable arrange of ILS learning system for students. In contrast, students' self-estimation [part (b)] minimized the percentage of neutral class partly because students enrolled tended to group themselves into negative or positive class. As students enrolled in this study are all the first-year undergraduates who have accumulated plenty of experience on their own learning styles during primary- and high-school studies, it is assumed in this study that students estimating themselves as negative or positive class are more confident on their learning style, comparing to those in marginal negative and marginal positive classes. In other words, students with self-estimation number equal to -2 and 2 are more reliable in distinguishing their own learning styles which in turn can be used to construct prediction models for learning style identification.

5.2 Bayes-based Statistics Are Capable of Distinguishing ILS Classes

Table 1 demonstrates the number of student samples classified according to their self-estimation in each ILS dimension. According to students' answers, they were assigned to five different learning style classes as shown in Fig. 2(b): negative (equal to -2), marginal negative (equal to -1), neutral (equal to 0), marginal positive (equal to 1) and positive (equal to 2). With reference to Fig. 1, negative class of ILS processing dimension (D1) refers to "Active", while positive class indicates "Reflective". Similar to D1, negative classes of D2, D3 and D4 are "Sensory", "Visual" and "Sequential", while the corresponding positive classes are "Intuitive", "Verbal" and "Global", respectively. As discussed in the previous section, students with self-estimation number equal to -2 and 2 are more reliable in distinguishing their learning style. Therefore, students' data from negative and positive classes would be further used in this study to construct prediction models for learning style identification.

Bayes-based statistics were applied in this study to identify most significant ones from all 44 questions in ILS questionnaire. According to Bayes analysis, 5, 9, 9 and 1 questions were identified as the most significant ones (p value < 0.05) for D1, D2, D3 and D4, respectively, in distinguishing positive class from the negative one. Based on the identifier of questions in the original ILS questionnaire, those five questions for D1 were question 5, 21, 33, 41 and 2; those nine questions for D2 were question 17, 18, 19, 20, 27, 33, 38, 43 and 44; those nine questions for D3 were question 7, 9, 11, 15, 23, 25, 27, 31 and 43; the only one question for D4 was question 43.

Table 1
The Number of Students Classified by Self-estimation in Each ILS Dimension

Class	D1	D2	D3	D4
Negative	33	40	32	32
Marginal negative	30	51	39	49
Neutral	7	12	10	11
Marginal positive	42	23	33	29
Positive	25	11	23	16

Based on these selected most significant questions, clustering and heat map distribution analysis were conducted to classify 137 student samples into hierarchical clusters. Figure 3 shows the heat map distribution pattern of samples in D3 (input dimension). As indicated in the figure, "Answer score" refers to those five question options designed to create a gradually changed classification system: negative class (with a score of -2), marginal negative class (with a score of -1), neutral class (with a score of 0), marginal positive class (with a score of 1) and positive class (with a score of 2). Therefore, the distribution pattern of answers across all samples and most significant questions can be illustrated by red–green colour. Question numbers on the far right side of Fig. 3 are those most significant questions used in distinguishing "Visual" from "Verbal" class in input dimension (D3).

The tree map on the top of Fig. 3 presents a clear pattern of how 137 student samples were distributed and clustered. As illustrated, the tree map divided all samples into two groups: group 1 and group 2. Below the tree map, there was a colour bar which used blue, green, white, purple and red to indicate classes of "Visual", "Marginal visual", "Neutral", "Marginal verbal" and "Verbal", respectively. In this situation, samples were coloured according to the answers to questions in ILS questionnaire. As shown in Fig. 3, all samples in "Visual" and "Verbal" classes were solely assigned into a single group (Visual samples in group 1 and Verbal samples in group 2). Moreover, all "Marginal verbal" samples were assigned to group 2 which was the same group as all "Verbal" samples belonging to. In the meantime, 98% of "Marginal visual" samples were assigned to group 1 which was the same group as all "Visual" samples belonging to.

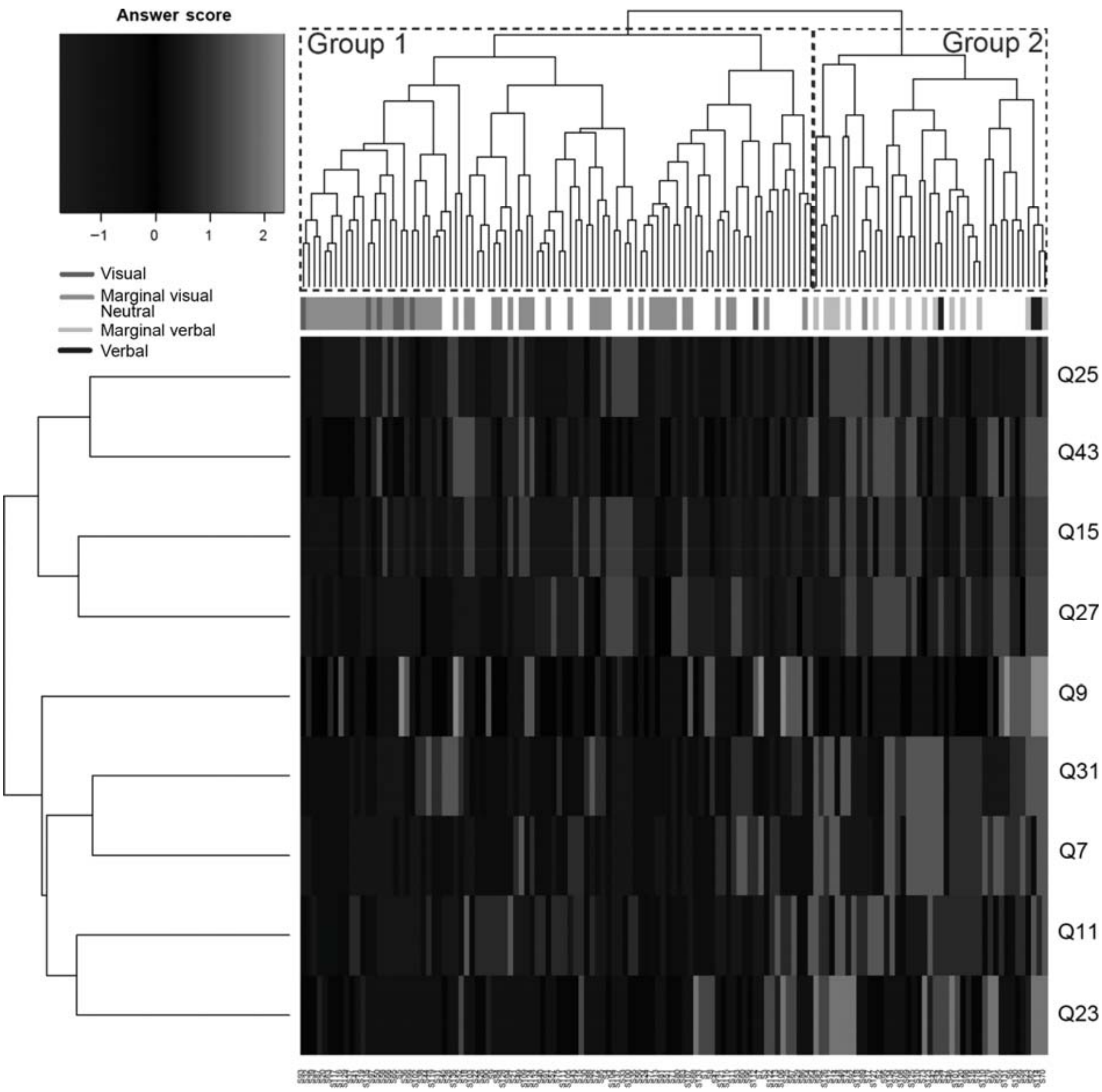


Figure 3. The heat map distribution pattern of samples in D3 (input dimension).

5.3 Tree Map Can Help to Distinguish Learning Style for Samples in ILS Neutral Class

As shown in Fig. 2(a), many student samples (52.6% for D1, 66.4% for D2, 45.3% for D3 and 63.5% for D4) were classified into neutral class, which limited the applicable arrange of ILS system. Based on the tree map, the newly constructed classification system was robust to correctly assign samples into their corresponding classes. Therefore, it is reasonable that the learning style of samples in neutral class could be defined by the learning style of their neighbouring samples. Moreover, because students estimating themselves as marginal negative or marginal positive were less confident on their learning style comparing to those as negative and positive, the newly constructed classification system could also shed light on double checking their true class of learning style.

6. Conclusion

ILS system is well designed to identify personalized learning style. However, a remarkable discrepancy between the results of ILS system and participants' self-estimation indicates that a significant large amount of participants cannot be grouped into any negative/positive class in each ILS dimension. This limits the applicable arrange of ILS in learning style identification. In this study, we focused on working out data-mining methods to extend ILS' applicable range, which is achieved by constructing a new questionnaire system and applying DM methods to a group of undergraduate participants. According to the analysis, Bayes-based statistics are found to be effective in distinguishing ILS classes; a newly constructed classification system, the tree map, can help to distinguish learning style for samples from ILS "neutral" class. Therefore, the

DMT applied in this study can be an effective method in enlarging the applicable range of traditional ILS system.

Acknowledgements

We gratefully acknowledge the research support of the National Social Science Foundation of China (12XYY012) and the Fundamental Research Funds for the Central Universities (CDJKXB14005 and CDJZR14468801). We would also like to thank the anonymous reviewers for insightful comments on this paper.

References

[1] S.M. Weiss and N. Indurkhya, *Predictive data mining: A practical guide*, 1st ed. (San Francisco, CA: Morgan Kaufmann, 1998).

[2] F. Zhu, L. Han, C. Zheng, B. Xie, M.T. Tammi, S. Yang, Y. Wei, and Y. Chen, What are next generation innovative therapeutic targets? Clues from genetic, structural, physicochemical, and systems profiles of successful targets, *Journal of Pharmacology and Experimental Therapeutics*, *330*(1), 2009, 304–315.

[3] R. Baker, Data mining for education, *International encyclopedia of education*, *7*(1), 2010, 112–118.

[4] R.S. Baker and K. Yacef, The state of educational data mining in 2009: A review and future visions, *Journal of Educational Data Mining*, *1*(1), 2009, 3–17.

[5] A.F. Gregorc and H.B. Ward, A new definition for individual, *Nassp Bulletin*, *61*(406), 1977, 20–26.

[6] J.W. Keefe, *Learning style theory and practice*, 1st ed. (Washington, USA: ERIC, 1987).

[7] J.C. Tseng, H.-C. Chu, G.-J. Hwang, and C.-C. Tsai, Development of an adaptive learning system with two sources of personalization information, *Computers & Education*, *51*(2), 2008, 776–786.

[8] K.A. Papanikolaou, M. Grigoriadou, H. Kornilakis, and G.D. Magoulas, Personalizing the interaction in a web-based educational hypermedia system: The case of INSPIRE, *User Modeling and User-adapted Interaction*, *13*(3), 2003, 213–267.

[9] V. Yannibelli, D. Godoy, and A. Amandi, A genetic algorithm approach to recognise students' learning styles, *Interactive Learning Environments*, *14*(1), 2006, 55–78.

[10] R.M. Felder and L.K. Silverman, Learning and teaching styles in engineering education, *Engineering Education*, *78*(7), 1988, 674–681.

[11] S. Graf, S.R. Viola, T. Leo, and Kinshuk, In-depth analysis of the Felder–Silverman learning style dimensions, *Journal of Research on Technology in Education*, *40*(1), 2007, 79–93.

[12] A. Ortigosa, P. Paredes, and P. Rodriguez, AH-questionnaire: An adaptive hierarchical questionnaire for learning styles, *Computers & Education*, *54*(4), 2010, 999–1005.

[13] L. Tsantis and J. Castellani, Enhancing learning environments through solution-based knowledge discovery tools: Forecasting for self-perpetuating systemic reform, *Journal of Special Education Technology*, *16*(4), 2001, 39–52.

[14] R.A. Ellis, Minimum indicators to assure quality of LMS-supported blended learning, *Educational Technology & Society*, *10*(2), 2007, 60–70.

[15] S.Y. Chen and G.D. Magoulas, *Adaptable and Adaptive Hypermedia Systems*, 1st ed. (Hershey, USA: IRM Press, 2005).

[16] J. Mostow and J. Beck, Some useful tactics to modify, map and mine data from intelligent tutors, *Natural Language Engineering*, *12*(2), 2006, 195–208.

[17] S.-W. Hsieh, Y.-R. Jang, G.-J. Hwang, and N.-S. Chen, Effects of teaching and learning styles on students' reflection levels for ubiquitous learning, *Computers & Education*, *57*(1), 2011, 1194–1201.

[18] Y. Akbulut and C.S. Cardak, Adaptive educational hypermedia accommodating learning styles: A content analysis of publications from 2000 to 2011, *Computers & Education*, *58*(2), 2012, 835–842.

[19] D.A. Kolb, *Experiential learning: Experience as the source of learning and development*, 1st ed. (New Jersey, USA: Prentice-Hall, 1984).

[20] R. Dunn and K. Dunn, Learning style as a criterion for placement in alternative programs, *The Phi Delta Kappan*, *56*(4), 1974, 275–278.

[21] G.K. Smyth, Linear models and empirical Bayes methods for assessing differential expression in microarray experiments, *Statistical Applications in Genetics and Molecular Biology*, *3*(1), 2004, 1–25.

[22] R. Gentleman, V.J. Carey, W. Huber, R.A. Irizarry, and S. Dudoit, *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (New York, USA: Springer, 2005).

[23] P.M. Valero-Mora and R. Ledesma, Graphical user interfaces for R, *Journal of Statistical Software*, *49*(1), 2012, 1–8.

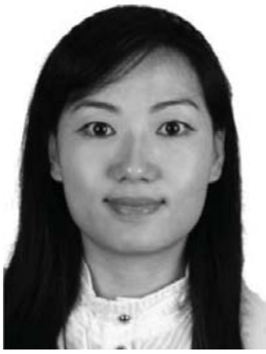
Biographies

Yanping Jing, third-year graduate student, Innovative Drug Research Centre, Chongqing University, China. Her research interests include machine learning method, artificial intelligent, courseware design and development, social psychology and behaviour.

Bo Li, doctoral candidate, Innovative Drug Research Centre, Chongqing University, China. His research interests include DM, feature selection, machine learning methodology design, bioinformatics and drug design.

Na Chen, first-year graduate student, Innovative Drug Research Centre, Chongqing University, Chongqing, China. Her research interests include artificial intelligent and machine learning method.

Xiaofeng Li, doctoral candidate, Innovative Drug Research Centre, Chongqing University, Chongqing, China. Her research interests include machine learning methodology design, bioinformatics and drug design.



Jie Hu, Ph.D. (Warwick, UK), M.A. (York, UK), associate professor, Chongqing University, China. Her research interests include artificial intelligent in education, Bayes-based statistics, DM and machine learning.

Feng Zhu received his Ph.D. from National University of Singapore, Singapore and is a professor and doctoral advisor at Innovative Drug Research Centre, Chongqing University, Chongqing, China. His research interests include artificial intelligent, machine learning, bioinformatics and Web & Database development.