

Analysis of Online Learning Style Model Based on K-means Algorithm

Rumei Li

Sino-French Engineer School
Beihang University
Beijing, China 100191

Chuantao Yin

Sino-French Engineer School
Beihang University
Beijing, China 100191

Abstract—Adaptive learning system is an ontology-based system that will provide relevant learning resources as a personalized bespoke online learning package for the learner based on their pedagogical needs. Thus, learner modeling and resources modeling need to be done in order to implementing the adaptive mechanism. Learning styles which refer to learners' preferred ways to learn can play an important role in learner modeling. This paper, by reviewing existing learning style theories and comparing the differences between traditional learning and online learning, proposes a new learning style model for online learning environment. What's more, an online learning style survey is designed for investigating learners' preferences and learning behavior. Then a classification rule for online learning style is discussed after mining the collected data based on k-means clustering algorithm. The results offer insights into a different classification mechanism for learning styles and prove the feasibility of the new online learning style model.

Keywords—adaptive learning system; online learning style; clustering

I. INTRODUCTION

Online learning is an important means of constructing continuing education and lifelong learning system. The emerging online learning platforms have provided high-quality educational resources. However, there is an important defect: how should learners choose the most suitable learning resources and paths in the face of massive learning resources? In this context, a new way of learning, namely, Smart Learning (S-learning) [1] has been developed. Smart learning follows a human-centered learning approach, which provides learners with diverse learning supports (such as intelligent search, resource feedback, etc.) with the support of big data, aiming at optimizing learners' experience and improving learners' efficiency.

In fact, as early as in 1980s, people started to research the Intelligent Tutoring System (ITS), also known as Adaptive Learning System, enabling learning resources to adapt to changes automatically according to the characteristics and learning styles of different students [2]. The core components of the adaptive learning system include learner model, content (learning resource) model and instruction model [3]. According to the relevant information in the learner model and the retrieval rules in the system, the system takes learning

Project supported by the National Natural Science Foundation of China (Grant No.61402028).

resources from the domain knowledge model and provides customized personalized courses with different content and forms, to meet the individual differences of learners to the maximum extent.

The adaptive learning system has a distinct feature from other systems. That is the modeling of user. The five most commonly used features are knowledge state, learning interest, learning goal, background and individual traits [4]. In recent years, researchers have paid more and more attention to the differences of learners' cognitive styles. Learning style is used as a personal trait to describe learners' learning preferences and characteristics.

Learning style is a relatively stable way of learning that learners perceive information and interact with the learning environment and respond to them [5]. For learners, understanding their own learning style can make them more confident in learning, and have the ability to optimize their own learning path [6]. For teachers, understanding the learning styles of students can provide more appropriate feedback and guidance for students, and provide different learning materials for different types of students [7].

Therefore, the development of adaptive learning system cannot be separated from the improvement of learners' model. As learner model is the key component to realize the adaptive function, it shall be supported by the theory of learning style. Most of the adaptive learning systems applying the learning style model are still using the classic model [8]. However, the classical learning style theory has been no longer applicable to online learning. In online learning, learners have a higher degree of freedom in geography, time and resource selection, as well as self-restraint, and emphasize initiative and active learning. Not only that, online learning is different from traditional learning in forms of learning activities, the roles of teachers and students and the forms of interaction, which will lead to different individual preferences associated with them. That is to say learning styles are also different from each other. Therefore, it is necessary to redefine the online learning style and model learners' consistent preferences and ways in the acceptance of digital information, processing and memory and the process of network operation in online learning platforms.

Building an online learning style model can not only help to master the behavior rules and preferences of learners, and provide a reference and basis for the recommendation of

learning resources, but also can help improve the function of learning feedback and learning evaluation. It is one of the cores in the construction of personalized learning platform. This paper proposes a new online learning style model based on the classical learning style theory, and designs an online learning style scale to collect online learners' behavior data. Then, the data mining algorithm is used to analyze the data and verify the feasibility of the model.

II. LEARNING STYLE MODEL

The study of learning style basically revolves around three aspects: theory, teaching and business. Through summary of 71 learning style models for the past 30 years, Coffield has classified the theory of learning style and the style models into 5 types[9]: thinking that the learning styles and learning preferences are based on some essence; thinking that learning style deeply reflects the cognitive structure; thinking that learning style is part of a stable personality; thinking that learning style is relatively stable learning preferences; transforming from learning styles to learning methods, learning strategies, learning direction and learning concept.

"Table I" shows the commonly used learning styles and Coffield's classification. The more dependent the learning style is, the more convinced the dynamic feature of learning style is. It is an interaction between learners and their experience.

TABLE I. COMMONLY USED LEARNING STYLES IN COFFIELD'S CLASSIFICATION

Type	Model
1	Dunn & Dunn ^[10]
2	Witkin ^[11]
3	Myers-Briggs ^[12]
4	Kolb ^[13] Felder-Silverman ^[14]
5	Entwistle ^[15]

Among them, Dunn and Dunn^[16] are the authority to study the style of learning earlier. They think that the formation of learning styles is influenced by many factors. They are elements of learning style, as shown in "Table II". Learning style elements enable learners to form a habitual preference way of information acceptance, processing and storage. Their researches on the elements of learning style have been revised several times, which lays the foundation for other learning style theories and models.

TABLE II. DUNN & DUNN'S ELEMENTS OF LEARNING STYLE

Stimulation Aspects	Elements of Style
Environment	Sound, light, temperature, place layout
Emotion	Motivation, learning persistence and personality
Society	Collective, independent, companion, group, transformed learning partners
Physiology	Perception, time rhythm, activity and sitting posture
Psychology	Analysis and synthesis, left-brain and right-brain, contemplative and impulsive

In addition, Felder-Silverman's learning style model is the most widely used in adaptive learning system. It draws lessons from predecessors' learning style model, which is more complete. At the same time, it assumes that learning styles are relatively stable and can show learners' learning preferences in different dimensions. According to Shockley's statistics on all kinds of online learning, using Felder-Silverman's learning style model accounted for 50%, while Kolb model accounted for only 8.6%, ranking second^[17].

In addition to the above model, some scholars have improved or integrated the model. However, single model can only reflect the differences of learners in certain aspects. For example, Kolb model only focus on the learning process and the mode of information processing^[18]; Myers-Briggs model only focus on learners' personality. The concepts of learning style they reflect are not complete from the definition. The integrated model is relatively complete, but it should be noted that each theory of learning style has its background and significance. Therefore, we should fully understand and analyze each model in the process of integration, considering the structural validity of each model, rather than a simple make-up of single model.

Due to the differences between traditional learning and online learning, online learning style has been preliminarily studied both at home and abroad. But the results are not very significant. Karin Anna Hummel has analyzed access data of online learner and their log files generated on the online learning platform in the database in order to analyze the learning behavior of learners^[19]. Jia-Jiunn Lo et al. have analyzed and processed the web pages browsed by the online learners on nerve net, so as to study the learners' learning style^[20]. The domestic scholars have also made analysis and research on online learning style analysis and research. According to the theory of information processing, Kolb's experiential learning model and Jung's 3D model of personality characteristics^[21], Chen Li et al. have measured learning style of learners from physiology, experience and psychology, and designed corresponding learning style scales^[22]. Li Suzhen put forward the research model of learning style and learning preference based on cluster analysis and association analysis of learning behavior and other methods^[23]. However, the domestic research is still limited to the stage of written survey report, and the researches on the quantitative learning behavior and learning style of online learners are not thorough.

III. THE TRADITIONAL LEARNING AND ONLINE LEARNING

In the traditional learning environment (especially in changeless teaching forms and constant teaching methods, course contents based on textbooks), knowledge can only be obtained through formal classroom learning and teacher's teaching. So, the traditional education is centered by teachers, classroom and textbooks, and teachers control the whole process of teaching. On the contrary, the online learning is centered by students. The learning tools, learning resources and teaching methods in the network environment provide services to learners, and learners can control the whole learning process according to their own needs. The theory of Dunn & Dunn mentioned in the previous paper suggests that

learners form a particular learning style under the influence of many learning style elements. Therefore, the five aspects of learning style elements can be used to compare the characteristics of traditional learning with characteristics of online learning as shown in "Table III".

TABLE III. COMPARISON OF CHARACTERISTICS OF TRADITIONAL LEARNING AND ONLINE LEARNING

Elements	Traditional Learning	Online Learning
Environment	Limited time and place, with no freedom	Unlimited time and place; freedom is greatly improved
Emotion	Boring and monotonous content; easy to be tired of learning	Learners in the dominant position can play subjective initiative
Society	Get information only from teachers	Can interact with many learners
Physiology	Over single textbook form	Rich media technologies enrich ways to acquire knowledge
Psychology	Changeless teaching method; learners lack understanding on knowledge	Explore and deeply understand knowledge

Online learning is different from traditional learning in the five learning style elements, including environment, emotion, society, physiology and psychology. Therefore, their learning styles are fundamentally different. Its importance to research of online learning style model is such clear.

IV. ONLINE LEARNING STYLE MODEL

In order to apply learning style model in adaptive learning system in the application, the author has learned from other existing classical models based on learning style elements and selected the typical "learning style characteristics", and connected learning style characteristics to corresponding learning behaviors and get the learning style model with four elements and eight dimensions as shown in "Table IV". The purpose is to classify learners through their learning behaviors, and give a reasonable explanation for their behavior classification according the theory of learning style, and finally provide different resources and paths for different online learners according to their learning style and achieve the adaptive learning mechanism.

TABLE IV. ONLINE LEARNING STYLES AND CORRESPONDING LEARNING CHARACTERISTICS AND BEHAVIORS

Elements	Learning Characteristics	Learning Behaviors
Emotion	Motivational	Gross time stayed in learning platform
Society	Communicational	View learning forum and speak in the forum, etc.
Psychology	Visual	Browse videos, pictures of materials and so on
	Verbal	Browse text, audio materials and so on
Physiological	Sensory	Browse materials of stories and reality
	Intuitive	Browse theoretical and highly generalized materials
	Sequential	Follow the prescribed order in learning
	Global	decentralized learning with no logical sequence

This model removes the element of environment. Compared with the traditional classroom, the online learners' learning environment is the network, and the same learning platform is equal to the learners. Considering that the model will eventually be applied to the adaptive learning system, the learning style must be quantified. So, the study style differences caused by learners' online learning behavior should be focused on. The author has referred to the Entwistle model in motivation of emotional element, the Felder-Silverman model in visual and verbal dimension of physiological element and in sequential and global dimension of physiological element, and the Myers-Briggs and Kolb model in sensory and intuitive of psychological element.

The above online learning style model has a choice to summarize the eight characteristics influencing online learning style. In the previous model, learning style in each dimension is opposite (for instance, learners only are of visual type or verbal type in the Felder-Silverman model). However, real learners may take different learning strategies for different learning content, and they both have two characteristics. If a learner chooses the same score for the two seemingly contradictory features of "sequence" and "synthesis", it shows that the learner has the ability to change in learning and to change their habits to learn different courses.

Based on the above model, the online learning style OLSX of learner X in adaptive learning system can be express as:

$$OLS_X = (d_{mot}, d_{com}, d_{vis}, d_{ver}, d_{sen}, d_{int}, d_{seq}, d_{glo}) \quad (1)$$

Among them, d_i is the value of learning style in the above eight dimensions, the value d_i in each dimension is associated and calculated by the learning behaviors a_{i1}, \dots, a_{iN} in this dimension:

$$d_i = \frac{1}{N} \sum_{j=1}^N \omega_{ij} a_{ij} \quad (2)$$

For each learning behavior a_{ij} , there is a corresponding weight ω_{ij} , as some learning behaviors possibly are associated with multiple characteristics. However, the proportions of the influences of characteristic values of different learners are different (for example, a learner likes browsing forums. It not only can increase the value of communication but also affect the value of motivation. However, the influence is relatively smaller. Otherwise, it will lead to too high similarity coefficient of the two dimensions, so that the dimension division lost its meaning).

V. ONLINE LEARNING STYLE SCALE

A. Initial Design

Through online learning style scale, learners' learning style can be targeted and obtained. The paper, based on the characteristics of learning style, has combined with the characteristics of the online learning content, learning activities and other aspects, and referred to the existing scales of classical learning style models and designed to measure the dimensions of learning style. It puts questions to learning behavior of online learners, and finally gets the first scale of Online Learning Style Measurement. It includes 23 measurement topics totally. In order to avoid a wide variety of

topics which may cause testers to be tired, it only set two to three topics for each characteristic. Ask question according to its frequency. In order to avoid neutral answers that cannot be distinguished from the subjects, and to facilitate subsequent comparisons and calculations, the scale only sets four options, including "0 (never)", "1 (rarely)", "2 (general)" and "3 (often)".

The original intention of the scale is to put question to each dimension individually. Therefore, it assumes that the weight ω_{ij} of each question is 1, and then the option score is a_{ij} . The learning style of each user can be calculated by formula (2).

B. Antecedent Research

The main purpose was to examine whether the scale was suitable for specific subjects. In the antecedent research, select 10 students from Sino-French Engineer School of Beihang University in a small way, and ask them to fill the scale. Ask them whether the question is clear and the sequence is reasonable, and give precious advice. It mainly reviews whether the subjects can fully understand the questions in the scale and don't impact by other topics when filling certain topic. Consequently, the author found some titles were not clear. Then, the author revised some titles in the first version of scale according to results of antecedent research.

C. Reliability and Validity Test

Reliability refers to the consistency of the results obtained when the same method is used to measure the same object repeatedly. In this study, the retest reliability method was used to randomly select 10 respondents and asked them to fill in the same scale two times with a median interval of 6 weeks. After calculation, the average Pearson correlation coefficient of the two tests was 0.713. According to retest reliability test standard, the above test results are acceptable.

Validity refers to the degree to which a measuring tool or means can accurately measure what is needed to measure. Since the scale is based on other classic model scales, the titles with the same learning characteristics have high similarity. The author only made a little modification for the compound online learning behavior, so it can be considered that the content has certain validity.

VI. DATA ANALYSIS OF ONLINE LEARNING STYLE

After verifying the validity of the scale, the scale was made and released on the <https://www.sojump.com/>. A total of 380 filling results were received within one week. Then, the online learning style data are clustered based on the K-means algorithm, and the common online learning styles are analyzed.

A. Data Preprocessing

Before clustering analysis, normalization of data and removal of outliers are usually performed. Normalization limits the data to be processed within a certain range, which makes the data processing more convenient. The goal of outlier processing is to remove the values deviated over large from the standard, and to reduce the influence of some too large or too small values on the result.

This outlier processing is mainly to remove the suspect data. After screening, remove results of non-online learners, filling within 100 seconds (average 220 seconds) and giving same answer to two opposite measurement questions, and remain 338 valid data.

For data normalization, since each item has four values, including 0, 1, 2 and 3, the normalized data can be obtained by averaging the values of each dimension.

B. Characteristic Selection

Characteristic selection refers to selecting a characteristic subset by removing irrelevant or redundant characteristics, reaching the purpose of reducing the number of characteristics, improving model accuracy and reducing the operating time. Moreover, selecting the really relevant features simplifies the model, making it easy for researchers to understand the process of data generation.

There are eight characteristics in online learning style model. But there are only two to three question for each characteristics in the scale. Through the comparison of variance and correlation coefficient, five questions were removed finally, so that all questions are remained within the weak correlation range.

C. Cluster Analysis

The classification of different learning styles is based on learners. Cluster analysis is to classify data into several categories according to similarity. The data in the same class are similar to each other, and the data in different classes are different. This study uses the most commonly used K-means algorithm.

Due to the unlabeled data set, the selection of evaluation indexes is greatly limited. Uses Silhouette Coefficient^[24] and Calinski-Harabaz index^[25] in this clustering to value the clustering results of unlabeled data sets. Thereinto, the value of Silhouette Coefficient is between -1 to 1; the closer it is to 1, the better the clustering result is; the concentrated the data within class are, the better the separation of inter-class data are. If the result is close to -1, it is on the contrary. The calculation of Calinski-Harabaz is very quick; the value is not certain; the bigger the value is the better the clustering effect is.

In the selection of clustering algorithms and clustering index, set up the following objectives: to discriminate learners carefully with classes as many as possible; to select Silhouette Coefficient and Calinski-Harabaz index as large as possible.

The effect of using K-means algorithm is shown in "Fig. 1". The abscissa is the number of classes (4 to 30), and the ordinate is the computing results of Silhouette Coefficient and Calinski-Harabaz index respectively. It is clear that the distinction between the data itself is not obvious, so the less the number of classes, the better the clustering effect. However, it can be found that the Calinski-Harabaz index has an obvious extremum in $n=5$, and is greater than all subsequent results. Meanwhile, the Silhouette Coefficient here is better than all the results afterwards. So we chose to divide the data into 5 categories.

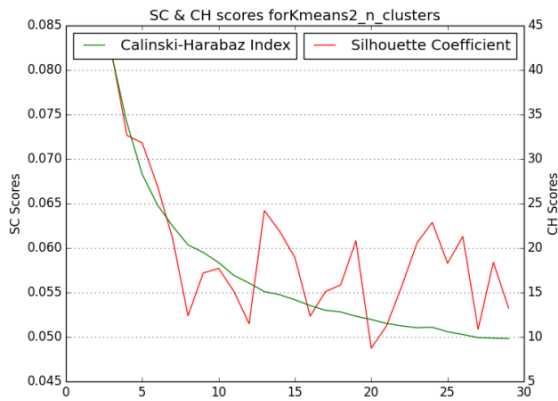


Fig. 1. K-Means Clustering Results.

According to online learning style model, the learners after clustering can be classified as the following types:

1) *Goal-type learners*: In the process of learning they learn jumpily and do not pay attention to the progressive order of learning. Such learners learn knowledge according to their needs, and have their own rhythm in the learning process. They are likely to be learners having knowledge of certain field, aiming at making up deficiencies through the network.

2) *Task-based learners*: It's consistent in all dimensions. But values are smaller than other types'. Learning behaviors of this type of learners is relatively less in the network platform. They maybe are network learning students according to the requirements. Their motivation is weakest. Due to various restrictions of online courses (for example, all learners shall learn certain online courses according to the requirements of the teacher. Learning resources lack diversity), they are not active on learning platform.

3) *Self-learning learners*: The values in each dimension are greater than those in other categories, especially in the communicational dimension and visual dimension. These learners love and are good at online learning. They actively communicate with others, and have higher comprehension on video and text, and strong information-receiving ability. They are learners good at self-learning.

4) *Stable learners*: In all dimensions, they stay on average position of all classes. They are likely to be long-term online learners, and have developed a stable habit of learning.

5) *Traditional learners*: The values in each dimension were not consistent, and the values in the motivation and visual dimensions were prominent, while the values in the sequence dimension were also higher than those in other categories. The learners may often study on online learning platform. They learn video courses according to learning order at a regular time, and are similar with traditional learning.

VII. CONCLUSION

This research mainly completed the following work:

The paper has analyzed online learning characteristics based on learning style elements of Dunn & Dunn and selected eight learning characteristics, including motivation, communication, vision, speech, perception, intuition, sequence, comprehension, as different dimension based on the traditional learning style model of Entwistle, Felder-Silverman, Myers-Briggs and Kolb to construct a online learning style model appropriate for online learners.

With the purpose of quantifying online learning behavior, the author has consulted the scale of traditional model and designed online learning style scale on eight dimensions. The validity of the scale is verified by antecedent research and reliability test and validity test.

According to the survey results on the scale, the author has carried out data processing, characteristic selection and cluster analysis. With the purpose of the optimization of learners' classification, five types of learners are summed up. Their characteristics are analyzed to verify the feasibility of online learning style model for classifying learners.

In a word, the online learning style model not only fills the blank of this research field, but also has four aspects of emotion, society, physiology and psychology in the learning style elements, and is relatively complete in structure. In addition, the classification of learners in the traditional model is mostly based on the researchers' occupation experience in psychology and pedagogy. This paper designed a learning style scale based on reasonable mining and analysis of real data in statistical method, and conversely verified the classification effect of the model on real learners.

Although this study provides a basis for modeling learners' personal traits in adaptive learning systems, the source of the data is only the results of the scale, rather than the actual behavior log data. Next, this paper intends to conduct data mining on user behavior logs on open source data sets, select behaviors corresponding to learning characteristics, and carry out cluster analysis again, which is compared with the classification results in this paper.

In a word, the adaptive mechanism of online learning platform still has a long way to go. For instance, calculate the learning style through user's behavior record and recommend courses for users through their learning style. It is still an expected breakthrough to provide learners with learning resources of varying content and structure by better integrating online learning behavior and learning style.

REFERENCES

- [1] Hwang G J. Definition, framework and research issues of smart learning environments-a context-aware ubiquitous learning perspective. *Smart Learning Environments*, 2014, 1(1): 1.
- [2] Murray T. Authoring intelligent tutoring systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education (IJAIED)*, 1999, 10: 98-129.
- [3] Oxman S, Wong W, Innovations D V X. White paper: Adaptive learning systems. *Integrated Education Solutions*, 2014.
- [4] Brusilovsky P, Millán E. User models for adaptive hypermedia and adaptive educational systems. *The adaptive web*. Springer-Verlag, 2007: 3-53.

- [5] Keefe J W, School applications of the learning style concept: Student learning styles. Reston, VA: National Associations of Secondary School Principals, 1979.
- [6] Davidson A, Ritchie K L. The early bird catches the worm! The impact of chronotype and learning style on academic success in university students. *Teaching and Learning Innovations*, 2016, 18.
- [7] Stash N. Incorporating cognitive/learning styles in a general-purpose adaptive hypermedia system. *Dissertation Abstracts International*, 2007, 68(04).
- [8] Truong H M. Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities. *Computers in Human Behavior*, 2016, 55: 1185-1193.
- [9] Coffield F, Moseley D, Hall E, et al. Should we be using learning styles? What research has to say to practice. 2004.
- [10] Dunn R. Learning style: State of the science. *Theory into practice*, 1984, 23(1): 10-19.
- [11] Witkin H A, Goodenough D R. Field dependence and interpersonal behavior. *Psychological bulletin*, 1977, 84(4): 661.
- [12] Myers I B M, Quenk M H, Hammer A L. *MBTI Manual: a guide to the development and use of the Myers-Briggs type indicator. Manual: a guide to the development and use of the Myers-Briggs type indicator.* Consulting Psychologists Press, 1998.
- [13] Kolb D A. Experiential learning theory and the learning style inventory: A reply to Freedman and Stumpf. *Academy of Management Review*, 1981, 6(2): 289-296.
- [14] Felder R M, Silverman L K. Learning and teaching styles in engineering education. *Engineering education*, 1988, 78(7): 674-681.
- [15] Marton F, Hounsell D, Entwistle N J. *The experience of learning: Implications for teaching and studying in higher education.* Scottish Academic Press, 1997.
- [16] Dunn R S, Dunn K J. *Teaching students through their individual learning styles: A practical approach.* Prentice Hall, 1978.
- [17] Shockley, Russell D. Learning styles and students' perceptions of satisfaction in community college Web-based learning environments. *The Turkish Online Journal of Distance Education*, 2008(1): 397-415.
- [18] DeCoux V M. Kolb's learning style inventory: A review of its applications in nursing research. *Journal of Nursing Education*, 2016, 29(5): 202-207.
- [19] Jonassen D H, Rohrer-Murphy L. Activity theory as a framework for designing constructivist learning environments. *Educational Technology Research and Development*, 1999, 47(1): 61-79.
- [20] Lo J J, Shu P C. Identification of learning styles online by observing learners' browsing behavior through a neural network. *British Journal of Educational Technology*, 2005, 36(1): 43-55.
- [21] Chen Li, Zhang Weiyuan, Hao Dan. 3D Model of Learning Style Characteristics of Chinese Remote Learners. *Open Education Research*, 2005, 11(2): 48-52.
- [22] Chen Li, Zhang Weiyuan, Hao Dan. Measurement Tools of Learning Style and Learning Style Elements of Chinese Remote Learners. *Open Education Research*, 2005, 11 (4): 65-71.
- [23] Li Suzhen. *Research on Mining Model of Network Learning Style and Learning Preferences Based on Network Learning Behavior Analysis.* Wuhan: Huazhong Normal University, 2009.
- [24] Peter J. Rousseeuw (1987). "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis". *Computational and Applied Mathematics* 20: 53-65. doi:10.1016/0377-0427(87)90125-7.
- [25] Caliński, T., & Harabasz, J. (1974). "A dendrite method for cluster analysis". *Communications in Statistics-theory and Methods* 3: 1-27. doi:10.1080/03610926.2011.560741.