An Early-Warning Model for Online Learners Based on User Portrait

Ye Sun, Rongqian Chai*

School of Liberal Education, Jincheng College of Sichuan University, Chengdu 611731, China

Corresponding Author Email: dzchrq@163.com

ABSTRACT

In the age of the Internet, online learning is an important learning strategy. At present, a large number of data on learning behavior have been generated on various online education platforms. It is difficult to grasp the learning situation of the numerous learners of these platforms according to the massive data. User portrait offers a possible solution to the problem. This paper firstly classifies the portrait of online learners into three dimensions, and constructs the tag system of learner portrait based on the data fields of online learning platform. Then, the learning behavior data of online learners were analyzed in details. Online learners were divided into multiple groups through data mining, and the learner portrait was generated. From the five dimensions of learner portrait, the learning situation was analyzed to master the learning information of learners. Based on the analysis results, the four-dimensional early-warning of learning situation was realized through sequence analysis and association rule mining. The research results provide a good reference for the improvement of online learning.

1. INTRODUCTION

The development of the Internet directly drives education informatization, giving rise to online education. Similar to traditional education, online education should also be learner centered. The improve the evaluation, resources, and service quality of online education, it is important to understand the various states and behaviors of online learners, explore the evolution of their behaviors and cognition, and track and accurately predict their online learning ability.

In recent years, learning analysis has been developing continuously. The International Conference on Learning Analytics & Knowledge defined learning analysis as the understanding and optimization of learning and its environment through measuring, collecting, analyzing, and reporting the data on learners and learning context. The New Alliance of the United States suggested that a huge amount of data on learning situation might be generated in actual learning, and these data could be examined and explain through learning analysis, using advanced measuring and collection tools. The Massive Open Online Course (MOOC) Alliance held that learning analysis is to measure, collect, analyze and report the data on learning behavior and environment, aiming to facilitate the understanding and optimization of learning process and environment.

According to the existing studies on learning analysis, this paper summarizes the process of learning analysis as using learning analysis technology to track and collect the data on learning process, to discover the laws of education from the data, and to give reasonable explanations for the laws. The learning analysis can improve the learning mode, and provide online teachers with warning about the learning situation, enabling them to promote the learning efficiency by adjusting teaching strategies. Through learning analysis, he learning process can be tracked by collecting the behavior data,

academic data and text data. The analysis on the multi-dimensional data helps to identify the state of online learners, providing support to multi-dimensional teaching.

At present, modern education is calling for better utilization of big data analysis and learning analysis. In the frontier field of online education, it is urgently needed to analyze learning situation and make early-warning based on data, with the aid of novel technologies like user portrait, and to optimize the teaching effect of online education platform.

To improve the participation and effectiveness of online learning, this paper develops an early-warning framework of learning situation on online learning platform based on user portrait, and demonstrates the excellence of the framework through empirical analysis. The research results provide a good reference for the optimization of online learning platforms.

2. LITERATURE REVIEW

Bloom's theory on educational objective and learning outcome is the most influential online learning theory [1, 2]. Fritz et al. [3] defined the analysis on modern learning environment as the extraction of teaching information, knowledge, and thinking modes with various tools of data collection, calculation, and analysis; the extracted data are related to learning process and behavior, and implicit and potentially valuable. Ling et al. [4] performed lag sequence analysis on online learning data, and found that most online learners are motivated by homework. Sun et al. [5] established a multi-dimensional active engagement model for online learning, and measured the degree of the active participation and interaction of learners in online learning. Krotkin et al. [6] constructed evaluation and analysis models of online distance learning, and verified the availability of these models. To sum

up, there is no unified standard for the data analysis of online learning behavior. On the dimension of learning analysis, most researchers have extracted the data on specific learning behavior as per actual needs and purposes of their research, and set up analysis models to mine behavior features.

In terms of data analysis technology, the behavior features of online learners are usually mined from multi-dimensional data on their participation, interaction, and psychological features, through statistical analysis [7], sequence analysis [8], association rule mining [9], social network analysis [10]. Considering the features of online learning, Wang et al. [11] proposed an ant colony optimization (ACO) algorithm, in which the pheromone concentration is adjusted step by step, to optimize the recommendation of learning path and predict the learning performance in future. According to the current situation of online learning, Inès et al. [12] introduced the upper and lower information integration technology into the recommendation process, and adopted the context factor weight to predict the learning situation. Through big data analysis, Li et al. [13] used big investigated the data generated in online learning, and predicted the behavior and performance of online learners, laying the basis for effective intervention in online learning behavior. Ann [14] transferred the concept of mobile learner portrait to the educational field, and defined learner portrait as the objective summary and specific description of learner features.

3. CONSTRUCTION OF USER PORTRAIT ON ONLINE LEARNING PLATFORM

Based on the real data of each user, user portrait is the modeling of user experience by abstracting the tag information of the user, with the aim to present the original appearance and refine the features of the user. The generation of user portrait requires a massive amount of user data.

The original data were collected a famous MOOC platform in China. There are 14 courses on the platform, attracting over 2 million learners. The original data were sorted out and recorded in Table 1.

As shown in Table 1, the sample data provide relatively complete information of the online learners, showing reasonable field settings and a good structure. Then, the data of four courses were analyzed statistically, and the results were presented in Table 2. It can be seen that only a few of the many learners had finished learning the four courses, and only a few had complete information.

In general, the user portrait is established based on the basic attributes, behavior features, and preference features of users. To provide accurate and personalized recommendation, the key lies in constructing user portrait on a set of tags and the knowledge system of the users. In the field of online education, the user portrait should be created by analyzing the group features and learning situation of online learners. Once

established, the user portrait helps to predict and warn the learning situation, and allows online teachers to make scientific decisions.

Table 1. The data for the construction of user portrait

Information type	Data field
Basic information	User ID
	Country
	Province
	City
	Gender
	IP address
Video learning	Video ID
	Video name
	Start time
	End time
	Viewing times
Homework and test data	Homework submission
	Test submission
	Submission time
	Homework ID
	Test ID
	Final result
Interaction data	Post content
	Reply content
	Post time
	Reply time
	Number of likes of post
	Number of likes of reply
	Post ID
	Poster ID
	Replier ID

This paper constructs learner portrait in five steps:
Step 1. Data collection
Collect the data on the experience of learners in the online learning platform, including platform operation data, homework data, test data, and interaction data; Set up the goal of the portrait, and define the dimension of portrait analysis; Screen the collected data, and eliminate the redundant fields.
Step 2. Data storage
Store the data in the SQL database of the online learning platform, which is convenient for data access and sharing.
Step 3. Portrait modeling
Select a suitable portrait model, and divide the dimensions of the portrait; establish the tag system, and define the mapping relationship between each tag in the system and each dimension of the portrait model.
Step 4. Portrait visualization
Extract the tags as per the goal of the portrait, and analyze the portrait through statistical analysis, cluster analysis, etc.
Step 5. Early-warning of learning situation
Evaluate the learning situation based on learner portrait, make early-warning of future learning situation, and put forward teaching intervention measures.

Table 2. The statistical data of four courses

No.	Course	Number of learners	Number of learners with complete information	Number of graduates	Number of graduates with complete information
1	69002	232,680	9,219	376	185
2	55003	176,890	3,695	167	76
3	33009	241,795	10,185	768	451
4	55008	155,603	4,126	357	167

The data collected from the online learning platform fall into four classes: the basic information of learners, the basic information of courses, the data on learner behavior, and the data on learning result. According to the research needs and data structure, this paper divides the data into three dimensions: basic information, learning behavior, and learning result.

Tag construction is the first step of portrait modeling. Tags, as the identification of learner features, demonstrate the common features of a group of learners. Depending on learner features, the tags can be classified into basic tags and extended tags [15].

The basic tag describes the basic situation and features of the learner, e.g. the basic information. The extended tag describes the learning features of the learners, including but not limited to preference, thinking habit, interests, and hobbies. The extended tag needs to be abstracted through data analysis, and contains more complex features than the basic tag. The dimension and tag division of the portrait are shown in Table 3. It can be seen that the basic information features are the primary basic tags, while the learning behavior features and learning result features are the primary extended tags

The learner portrait should be analyzed based on the learning data of various dimensions. The portrait analysis could reveal the features of the preference and behavior of a learner group hidden behind the massive data, laying the data basis for the effective early-warning service of online learning situation. According to the sources and contents of data, the data fields of portrait tags could be constantly enriched. After sorting out the fields of the original data, the data indices corresponding to the secondary tags of learner portrait were introduced (Table 4).

Table 3. The dimension and tag division of the portrait

	Primary tag	Secondary tag
Basic tag		Age
	Basic information features	Gender
		Region
Extended tag		Learning time
		Attendance
	Learning behavior features	Course preference
		Interaction level
		Resource preference
	Learning result features	Homework score
		Test score

Table 4. The portrait tags and data indices

Primary tag	Secondary tag	Data indices
Basic information features	Age	Learners' age
	Gender	Learners' gender
	Region	Learners' region
	Learning time	Video watching time, homework time, and test time
Learning behavior features	Attendance	The times of learning course content, the times of learning courseware list, the times of learning homework content, the times of learning homework list, the number of homework submissions, and the number of test submissions
	Course preference	The times of visiting course announcement, the times of visiting courseware list, the times of visiting forum list, the times of visiting post content, the times of watching videos, video watching duration, the times of watching video units
	Interaction level	Number of replies, number of contents, number of posts, and number of likes
	Resource preference	The times of visiting courseware list, the times of visiting class content, the times of visiting forum list, the times of visiting post content, the times of watching weekly video, and the times of watching other videos
Learning results features	Homework score	Homework performance
	Test score	Regular test results, and final test results

Table 5. The classification of tag values of learner portrait

Primary tag	Secondary tag	Data indexes
Basic information features	Age	Teens (≤25), youth (26-40), middle-aged (41-65), senior (>65)
	Gender	Male, Female
	Region	Eastern, central, western
	Learning time	Negative (less time), normal, positive (more time)
	Attendance	Active participation, regular participation, potential dropout, high turnover
Learning behavior features	Course preference	Video type, text type, field-independent type, field-dependent type, meditation type, active type
	Interaction level	Positive interaction, negative interaction
	Resource preference	Direct acquisition, exploratory learning
Learning result features	Homework score	Qualified, unqualified
	Test score	Qualified, unqualified

The tag system should be set up according to the dimensions of learner portrait, and in the light of the features of the learner group. Otherwise, the tags will be unable to reflect the actual features of the learner group.

In this paper, the tag system is divided into three classes: basic information features, learning result features, and learning behavior features. The first two types of features are static tags, and the latter kind of features is relatively dynamic. By refining secondary tags, the data indices that conform to the fields of the original dataset were divided, followed by classification of tag values (Table 5).
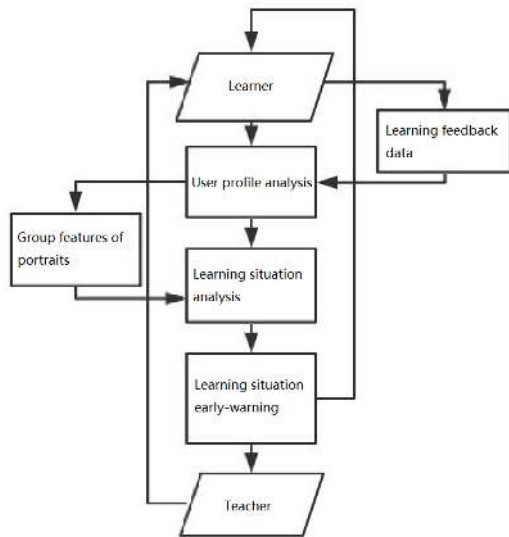
Figure 1. The portrait-based early-warning framework of learning situation

During online education, the state and behavior of learners change significantly with the passage of time. It is necessary to make dynamic early-warning of learning situation. Centering on learner portrait, an early-warning framework was designed for the learning situation of the online learning platform.

As shown in Figure 1, the proposed framework aims to serve specifically online learners. Under the framework, the learning preferences of different learner groups are obtained through the construction and analysis of learner portrait. Then, the learning situation and personalized needs of each leaner are derived from his/her features. Based on the personalized needs, a suitable early-warning strategy is prepared for the learning situation.

Putting learner at the core, the designed early-warning framework makes accurate analysis of the portrait and data of each learner to ascertain the learning situation and implement dynamic early-warning.

4. EMPIRICAL ANALYSIS OF LEARNER PORTRAIT AND EARLY-WARNING OF LEARNING SITUATION

The original data were collected from a famous MOOC platform in China. The total size of the data is about 40GB. After cleaning and preprocessing, the data for empirical analysis are about 2.6GB, covering the behavior records of about 900,000 learners.

The number of learners and the number of learners who completed each of the 14 courses were counted, respectively. The statistical results are shown in Table 6.

Based on the statistical results of the tag values, the learning situation of learners was summarized according to the dimension of portrait analysis. The overall results on learning situation are provided in Table 7.

The following findings were obtained through the analysis of the overall portrait of learners:

On learning activity, very few learners are strongly active. Most of them are not very active in course learning, failing to participate in various learning activities on time.

On learning engagement, most learners spent lots of time watching videos, possibly because watching videos is the most direct means to acquire knowledge points quickly.

On learning interaction, most learners prefer independent learning over interaction.

On learning preference, the learners generally prefer to obtain learning resources directly, learn independently through video, read various pages of the course frequently, and think actively during online learning.

On learning results, the learners are more concerned about the final test than regular homework. The situation of homework completion is not very good. Most learners achieve good test scores, because they care more about outcome evaluation.

Next, the attendance times were calculated through clustering analysis, after the weight of the sum of the times of visiting the class content and visiting the courseware list was set to 0.3, the weight of that of visiting the homework content and visiting the homework list was set to 0.2, and the weight of the total number of homework and test submissions was set to 0.2. The clustering results are presented in Table 8.

Table 6. The statistics on the number of learners who completed each course

Course No.	Number of learners	Number of learners who completed each course
1	233,000	378
2	176,000	168
3	244,000	789
4	157,000	365
5	246,000	327
6	151,000	487
7	166,000	233
8	241,000	272
9	168,000	282
10	179,000	512
11	250,000	126
12	183,000	281
13	153,000	556
14	202,000	647

Table 7. The overall results on learning situation based on learner portrait

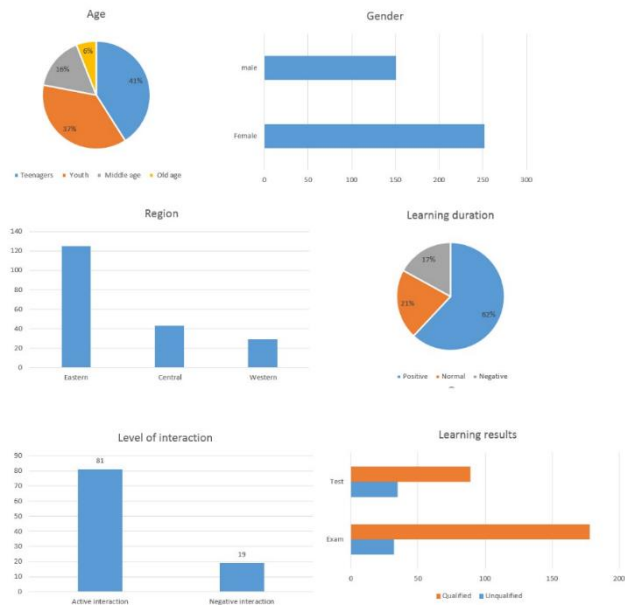| Portrait tag	Dimension	Learning situation	Example of learning situation
Attendance	Learning activity	Strongly active Moderately active Slightly active Strongly inactive	Nearly half of the learners are not strongly active and may drop out.
Learning time	Learning engagement	Study time for homework Study time for video Study time for test	Most learners spend their time watching videos and doing homework. The learning time of most learners falls between 4 and 12 hours.
Interaction level	Learning interaction	Active interaction Negative interaction	Most learners do not like interaction.
Course preference	Learning preference	Prefer watching videos Prefer reading text Prefer direct access to learning resources Prefer indirect access to learning resources	Most learners like to learn directly and independently by watching videos, read various pages of the course frequently, and act as active learners.
Homework and test score	Learning results		A large portion of learners have low homework scores, but high test scores. Most learners prefer sitting test to doing homework.

Table 8. The clustering results of attendance times

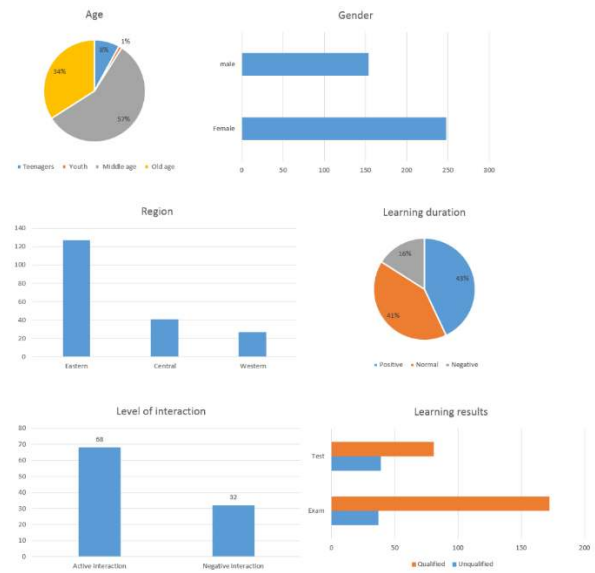| Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4
The number of attendance times	23	50	33	13

Based on the clustering results, the portraits of online learners were split into four types: active participation, regular participation, potential dropout, and high loss. The portraits of active participation, regular participation, potential dropout, and high loss learners are given in Figures 2-6, respectively.

As shown in Figure 2, most active participation learners are females, who have high pass rates in academic achievement, and strong willingness to learn. These learners are well motivated and highly active in various learning activities of the platform. Their learning time is generally longer than other learners. In terms of age, the youth takes up a high proportion of active participation learners, featuring strong learning ability and active interaction. With strong self-learning ability, these learners are willing to get involved in the whole learning process.

As shown in Figure 3, most regular participation learners are also females, who have high pass rates in academic achievement, and strong willingness to learn. These learners have strong motivations and can adhere to online learning. Their participation in various learning activities of the platform meets the attendance requirements. In terms of age, the middle-aged and seniors account for a high proportion in regular participation learners, and exhibit relatively strong learning ability. With excellence in autonomous learning, these learners stably partake in the whole learning process, and maintain a high level of interaction.

Figure 3. The portrait of regular participation learners

As shown in Figure 4, the majority of potential dropout learners are males, who have mediocre academic performance, average learning intention, and a high probability of becoming high loss learners. With a weak motivation, these learners do

Figure 2. The portrait of active participation learners

not participate actively in various activities of the learning platform for a long time, and often skip the classes. In terms of age, many of potential dropout learners are youth with poor interactivity.

As shown in Figure 5, there are more males than females among high loss learners. As the name suggests, high loss learners are very likely to get lost, due to their low pass rate and weak learning intention. With no motivation, these learners seldom attend the various activities of the online platform for a long time. In terms of age, the youth takes up a good portion of high loss learners. They are not highly involved in learning or interaction, and perform poorly in homework and test.
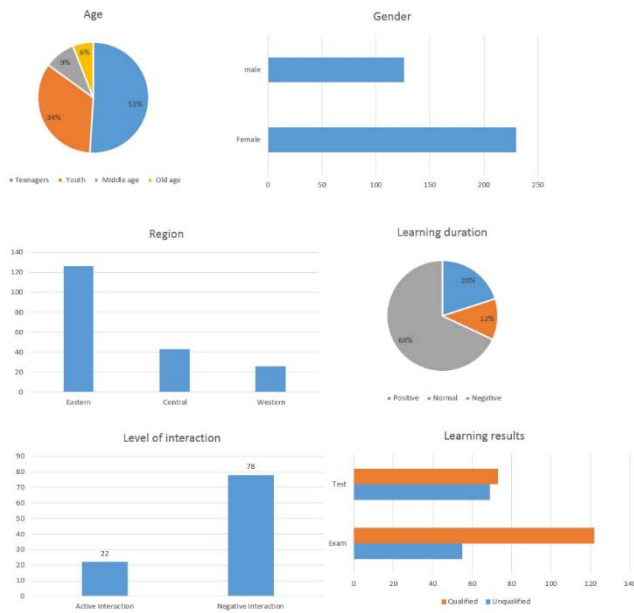

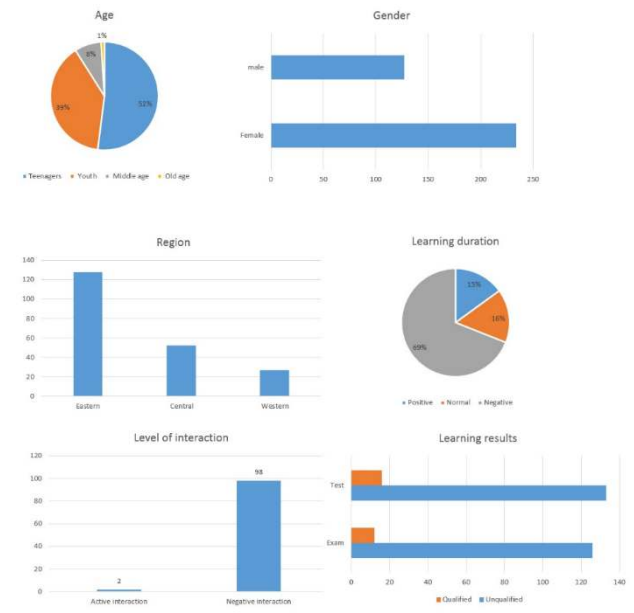**Figure 4.** The portrait of potential dropout learners

Figure 5. The portrait of high loss learners

5. CONCLUSIONS

Online learning faces two serious problems, namely, low participation and poor learning effect. To solve the problems, this paper tries to realize the early-warning of learning situation based on user portrait. Firstly, the portrait dimension and tag system were determined, and used to construct the user portrait. Then, the early-warning framework of learning situation was established based on user portrait. Under the framework, the online learners are divided into different groups, the learning situation of each group is evaluated, and the early-warning is performed through various data mining and analysis methods, e.g. cluster analysis, and association rule mining. Finally, the proposed framework was proved valid through empirical analysis on the data of an actual online learning platform.

REFERENCES

[1] Faria, E.S.J., Yamanaka, K., Tavares, J.A. (2012). A methodology for computer programming teaching based on bloom's taxonomy of educational objectives and apllied through the pair programming. IEEE Latin America Transactions, 10(2): 1589-1594. http://doi.org/10.1109/TLA.2012.6187603

[2] Rohrdantz, C., Mansmann, F., North, C., Keim, D.A. (2014). Augmenting the educational curriculum with the Visual Analytics Science and Technology Challenge: Opportunities and pitfalls. Information Visualization, 13(4): 313-325. https://doi.org/10.1177/1473871613481693

[3] Fritz, C.M. (2003). The learning environment as place: an analysis of the United States Department of Education's six design principles for learning environments. Journal of Antimicrobial Chemotherapy, 54(3): 634-639. https://doi.org/10.1093/jac/dkh395

[4] Wang, L., Hu, G., Zhou, T. (2018). Semantic analysis of learners' emotional tendencies on online MOOC education. Sustainability, 10(6): 1921. https://doi.org/10.3390/su10061921

[5] Sun, G.X., Bin, S. (2018). Construction of learning behavioral engagement model for MOOCs platform based on data analysis. Educational Sciences: Theory & Practice, 18(5): 2206-2216. https://doi.org/10.12738/estp.2018.5.120

[6] Schmidt-Jones, C. (2017). Offering authentic learning activities in the context of open resources and real-world goals: A Study of self-motivated online music learning. European Journal of Open, Distance and E-learning, 20(1): 112-126. https://doi.org/10.1515/eurodl-2017-0007

[7] Destercke, S. (2014). Comments on "A distance-based statistical analysis of fuzzy number-valued data" by the SMIRE research group. International Journal of Approximate Reasoning, 55(7): 1575-1577. https://doi.org/10.1016/j.ijar.2014.04.001

[8] Chum, K., Guy, R.K., Jacobson, Jr, M.J., Mosunov, A.S. (2018). Numerical and statistical analysis of aliquot sequences. Experimental Mathematics, (196): 1-12. https://doi.org/10.1080/10586458.2018.1477077

[9] Rolfsnes, T., Moonen, L., Di Alesio, S., Behjati, R., Binkley, D. (2018). Aggregating association rules to improve change recommendation. Empirical Software Engineering, 23(2): 987-1035. https://doi.org/10.1007/s10664-017-9560-y

[10] Sun, G.X., Bin, S., Jiang, M., Cao, N., Zheng, Z., Zhao,

H., Xu, L. (2019). Research on public opinion propagation model in social network based on blockchain. CMC-Computers Materials & Continua, 60(3): 1015-1027.

[11] Wang, F.H. (2012). On extracting recommendation knowledge for personalized web-based learning based on ant colony optimization with segmented-goal and meta-control strategies. Expert Systems with Applications, 39(7): 6446-6453. https://doi.org/10.1016/j.eswa.2011.12.063

[12] Saâdi, I.B., Hamdani, A. (2019). A semantic approach for situation-aware ubiquitous learner support. International Journal of Smart Technology and Learning, 1(2): 162-187. https://doi.org/10.1504/IJSMARTTL.2019.097971

[13] Li, H.J., Peng, M. (2019). Online course learning outcome evaluation method based on big data analysis. International Journal of Continuing Engineering Education and Life Long Learning, 29(4): 349-361. https://doi.org/10.1504/IJCEELL.2019.102769

[14] Harris, A.S. (2011). Bernini's portrait drawings: Context and connoisseurship. The Sculpture Journal, 20(2): 163-178. https://doi.org/10.3828/sj.2011.17

[15] Kardan, A.A., Sani, M.F., Modaberi, S. (2016). Implicit learner assessment based on semantic relevance of tags. Computers in Human Behavior, 55: 743-749. https://doi.org/10.1016/j.chb.2015.10.027