

Article

Data-Driven Personalized Learning Path Planning Based on Cognitive Diagnostic Assessments in MOOCs

Bo Jiang , Xinya Li , Shuhao Yang, Yaqi Kong, Wei Cheng, Chuanyan Hao  and Qiaomin Lin *

School of Educational Science and Technology, Nanjing University of Posts and Telecommunications, Nanjing 210049, China; 1020162909@njupt.edu.cn (X.L.); b18150322@njupt.edu.cn (S.Y.); 1021163408@njupt.edu.cn (Y.K.); chengweiet@njupt.edu.cn (W.C.)

* Correspondence: jiangbo@njupt.edu.cn (B.J.); hcy@njupt.edu.cn (C.H.); qmlin@njupt.edu.cn (Q.L.)

Abstract: Personalized learning paths aim to save learning time and improve learning achievements by providing the most appropriate learning sequence for heterogeneous students. Most existing methods that construct personalized learning paths focus on students' characteristics or knowledge structure, while ignoring the critical roles of learning states. This study describes a dynamic personalized learning path planning algorithm to recommend appropriate knowledge points for online students based on their learning states and the difficulty of each knowledge point. The proposed method first calculates the difficulty of knowledge points automatically and constructs a knowledge difficulty model. Then, a dynamic knowledge mastery model is built based on learning behavior and normalized test scores. Finally, a path that satisfies students' personalized changing states is generated. To achieve the aforementioned goal, a novel method that calculates the difficulty of knowledge points automatically is proposed. Moreover, the personalized learning path planning method proposed in this research is not limited to a particular course. To evaluate the method, we use a series of approaches to verify the impact of the personalized path on student learning. The experimental results demonstrate that the proposed algorithm can effectively generate personalized learning paths. Results demonstrate that the personalized path proposed by the algorithm can improve effective behavior rates, course completion rates and learning efficiency. Results also show that the personalized learning paths based on student states would help students to master knowledge.

Keywords: personalized learning path; mastery learning; learning behavior; personalized instruction scaffolding



Citation: Jiang, B.; Li, X.; Yang, S.; Kong, Y.; Cheng, W.; Hao, C.; Lin, Q. Data-Driven Personalized Learning Path Planning Based on Cognitive Diagnostic Assessments in MOOCs. *Appl. Sci.* **2022**, *12*, 3982. <https://doi.org/10.3390/app12083982>

Academic Editors: Xiao Chen, ASM Shihavuddin and Dan Zheng

Received: 1 March 2022

Accepted: 11 April 2022

Published: 14 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Massive Open Online Courses (MOOCs) are popular among students because of their low registration threshold and liberal learning time. Learners have the freedom and chances to choose excellent courses and quality exercises [1]. However, some obstructions, such as lower completion rates, lower pass rates and lower learning efficiency, hinder MOOCs' development [2,3].

Fixed learning sequences still exist within a particular course in MOOCs. A predetermined learning sequence is unsuitable for all learners' studies, as diverse students have different knowledge structures and learning states. The teachers can easily and expediently support students in face-to-face classes. However, learning in an online scenario weakens the teacher's guidance role. It is challenging for learners to choose appropriate learning materials or sequences based on their actual learning situations. Thus, heterogeneous learners have a growing demand for diversified learning guidelines in an online learning scenario.

The formal definition of personalized learning provided by the U.S. Department of Education in the 2017 National Educational Technology Plan is presented as follows: "Personalized learning is an instructing strategy that tailors the learning speed and tactics to the specific needs of each student. Learning objectives, instructional methodologies, and instructional content (especially its sequencing) can all vary depending on the requirements

of students” [4]. Previous research works show that personalized learning contributes to learning achievement and learning satisfaction [5]. Providing tailored scaffolds helps heterogeneous students to master knowledge better and contributes to improving learners’ subsequent learning [6–8].

This study aims to propose an algorithm that can automatically generate personalized learning paths based on the learners’ learning states in real-time. Unlike previous studies, e.g., [6], this research takes learning behaviors as essential parameters for judging students’ learning states. More advanced than previous studies in which teachers have to manually label the difficulty of knowledge points, this research uses a data-driven scoring model to measure the difficulty of knowledge points for general students automatically. This approach lowers time costs and the stress on teachers to a certain extent. Moreover, the personalized path planning algorithm can popularize all kinds of courses in an online learning scenario. The research questions and highlights/contributions of this study are summarized in Section 2.4.

The following are the article’s main sections: Section 2 provides the theoretical foundation and a review of relevant research, and Section 3 describes the theoretical method for generating a personalized learning path. The critical steps of the dynamic personalized learning path planning algorithm are described in this section. Section 4 contains the experimental results and discussions, and the evaluation and conclusions are provided in Sections 5 and 6.

2. Research Background

Personalized Learning Path (PLP) is a strategy that selects the most suitable learning sequences for learners. Previous studies show that personalized learning paths contribute to students’ academic achievements [9–11]. The personalized learning path as an alternative path aims to replace the predefined learning sequence. While not consistently superior to the original path, alternative paths provide students with unique learning support in a real learning environment. The existing personalized learning path planning methods can be divided into three main categories:

- Personalized Path Planning Based on Student Characteristics;
- Personalized Path Planning Based on Log Data;
- Personalized Path Planning Based on Knowledge Construction.

In the following sections, each category is reviewed in detail.

2.1. Personalized Path Planning Based on Student Characteristics

Most personalized learning paths help to improve learning efficiency [12]. The personalized learning path planning method based on student characteristics considers characteristics such as learning style, and preference is the essential parameter. Researchers rely on tests or questionnaires to collect learner characteristics [13–15], such as students’ learning goals [16], learning styles [7,17], and preferences [7,18], to construct learner models. For example, Vanitha et al. considered learners’ learning goals and knowledge levels to be significant elements for path planning [19]. Rohloff et al. emphasized the need for path planning and course recommendations based on diverse learning goals in MOOC learning scenarios [16]. Nabizadeh et al. used a depth-first search algorithm to locate as many course sequences as possible by combining the learning goals with a knowledge map [9]. Yang et al. proposed an attribute-based ant colony algorithm to recommend suitable learning objects based on learning styles and a learner’s knowledge level [20].

The aforementioned studies show that constructing a personalized learning path based on the learners’ characteristics is valid and suitable. Nevertheless, constructing personalized learning paths based on learners’ characteristics is prone to overlooking the logic of knowledge. Learners are prone to encountering challenges in grasping the whole knowledge structure during the learning process.

2.2. Personalized Path Planning Based on Log Data

One of the main studies in online learning is how to provide personalized learning paths according to learners' log data [21]. During online learning, these "footprints" can be used as the basic parameters for constructing a learning path [22]. Collecting learning behavior data provides opportunities for researchers or teachers to understand the learners' learning process and to predict their academic achievements [22,23]. For example, Xia et al. proposed a system that can provide suitable questions based on the history data of other students in parallel learning scenarios. The system helped learners to obtain customized and adaptable quiz sequences from a massive question bank [24]. Liu et al. proposed a learning path combination recommendation method based on log data from learners [25].

The log data-based recommendation system made use of students' historical behavior data to grasp their features and recommend the learning objects they need. However, this kind of path generation methods ignored the knowledge structure and learners' unique features. The personalized learning path derived from group data may not be appropriate for all learners. Moreover, it will be challenging to encounter the cold-start problem when data are not sufficient.

2.3. Personalized Path Planning Based on Knowledge Construction

Grasping the prerequisite relationships of a kind of knowledge helps students to master that knowledge. One of the reasons why learners drop out of MOOCs is that they cannot find the right logical sequence of knowledge [26]. Supporting beginners who are lost in materials by mining the prerequisite relationships of a kind of knowledge is thus important.

Fung et al., for example, extracted concept keywords from relevant course materials and calculated the correlation coefficient matrix between concepts [27]. Zhu et al. proposed a novel multi-constraint learning path recommendation algorithm based on a knowledge map to solve the problem that most learners struggle to choose suitable learning materials. Finally, the validity of the algorithm was confirmed by a questionnaire [28]. This personalized learning path, which focuses only on the knowledge structure, ignores learners' unique features and thus is not conducive to their subsequent learning.

2.4. Brief Summary of References

A high dropout rate, low completion rate, and poor learning effects are problems that the MOOC platforms need to solve [29,30]. Forming a customized learning sequence based on learners' characteristics is not a new idea. However, the learning paths generated at once based on learners' feature or learning data are not adapted to the reality of the learning situation [31]. It is necessary to consider the learners' changing state to construct a personalized learning path.

This study proposes a dynamic learning path planning algorithm for the online learning scenario. Following previous research [6], this study also pays attention to the significant role of learners' changing states in learning. In contrast to the strategy of manually labeling the knowledge difficulty used in the previous study, the method proposed in this study can calculate the difficulty of knowledge points automatically based on a data-driven scoring model.

Most current personalized learning path recommending systems are applied to a specific course [32]. Moreover, MOOC platforms, such as XuetangX, cannot provide personalized learning sequences for learners even in a specific course. Thus, this study proposes a novel method that can provide personalized learning path planning based on the timely diagnosis of students' learning states in an MOOC learning scenario. This algorithm is a generic method that can be utilized in various courses.

Based on the above analysis, this study focuses on the following research questions:

RQ1: Is a personalized learning path based on learning states beneficial to MOOC learners' learning efficiency?

RQ2: Is a personalized learning path based on learning status conducive to the continuous learning of MOOC learners?

Moreover, the research highlights and contributions of this work are summarized as follows:

- A personalized learning path planning algorithm based on learners' dynamic learning state and the difficulty of knowledge points.
- A data-driven scoring model that measures the difficulty level of specific knowledge points for general students. A knowledge difficulty model is established based on the scoring model. The knowledge difficulty model is more accurate and convenient when compared to the previous study on manually marking knowledge difficulty levels by subject teachers [6].
- A knowledge mastery model based on learners' learning behavior data and exercise data, such as MOOCCubeX [33], to dynamically evaluate students' learning states.
- A feedback strategy to dynamically arrange learning paths by following a circular learning list based on their real-time state and the knowledge difficulty level. The importance of "mastering learning" is also emphasized.

3. Method

This study aims to construct a personalized learning path planning method based on learners' real-time learning states and knowledge difficulty levels. Unlike the previous study [6], this work proposes a novel method to automatically calculate the difficulty of knowledge points. Moreover, the personalized learning path planning method proposed in this study is not limited to a particular course. This section explains the main method and terminologies of the whole pipeline in details. The overall steps of generating personalized learning paths are as follows:

- (1) Constructing a knowledge difficulty model and calculating the difficulty of knowledge points automatically;
- (2) Constructing a dynamic knowledge mastery model based on students' learning behaviors and normalized exercise scores;
- (3) Generating personalized learning paths for learners based on the knowledge difficulty model and knowledge mastery model.

3.1. Data Preprocessing

The course and learning data stored inside the MOOC platforms mainly comprise two parts, i.e., course resource data and students' learning behavior data. Each course contains many chapters with videos and exercises. Figure 1 illustrates the hierarchical diagram of the course resources.

Each course usually contains a few chapters. The data in each course chapter are listed as follows:

- Course video titles and captions;
- Exercise tests;
- Prerequisite relationships among knowledge points.

The MOOC platforms record all online users' learning behaviors of watching course videos, including repetition, fast-forwarding, and skipping. Students need to complete the chapter exercise after watching the video. The learning data of each student are depicted as follows:

- Video watching behavior;
- Exercise performance;
- Comments and replies in the comment area.

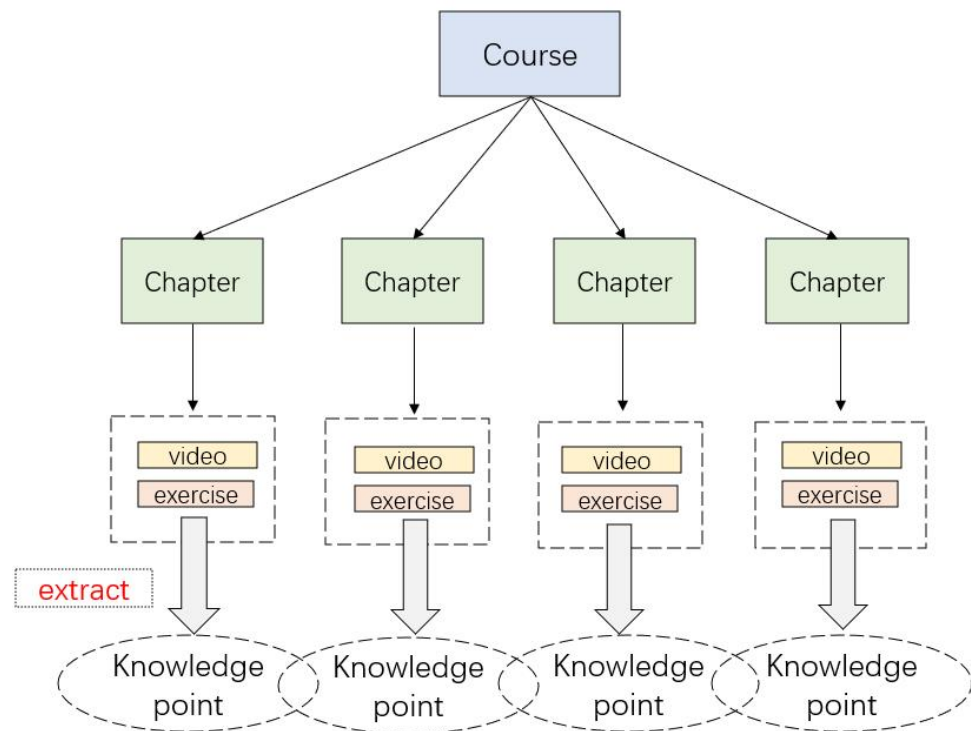


Figure 1. Content hierarchy of course resources.

To develop the dynamic learning path planning algorithm, we processed all the aforementioned data as follows:

- (1) Keyword extraction: Extract keywords from video titles, video subtitles, and chapter exercises.
- (2) Exercise classification: Compare the keywords of the chapter exercise with the keywords of the video titles and subtitles. Each exercise is categorized into the knowledge points with the most occurrences of its keywords.
- (3) Normalization: Normalizing scores of exercise tests.

3.2. Knowledge Difficulty Model

Based on the overall performance of students in MOOC learning scenarios, this section constructs a scoring model to measure the difficulty of specific knowledge points. The input parameters of the knowledge point-based difficulty model are the average exercise test scores of all students who have studied the knowledge point. The output of the knowledge difficulty model is the difficulty level of knowledge.

$$diff(j) = w_1[1 - \overline{sc\bar{o}}(j)] + w_2\overline{rep}(j) + w_3\overline{com}(j) \quad (1)$$

In Equation (1), w_1 , w_2 and w_3 are weights of the input parameters. $\overline{sc\bar{o}}(j)$, $\overline{rep}(j)$ and $\overline{com}(j)$ are the average test score, average number of repeated watching of the video and average number of comments of the j -th knowledge point, respectively. The larger the $diff_j$ value, the more difficult the j -th knowledge is to master.

$$\overline{sc\bar{o}}(j) = \frac{\sum_{i=1}^{N_j^{\text{history}}} sco_{ij}}{N_j^{\text{history}}} \quad (2)$$

$$\overline{rep}(j) = \frac{\sum_{i=1}^{N_j^{\text{history}}} rep_{ij}}{N_j^{\text{history}}} \quad (3)$$

$$\overline{com}(j) = \frac{\sum_{i=1}^{N_j^{\text{history}}} com_{ij}}{N_j^{\text{history}}} \quad (4)$$

In Equations (2)–(4), N_j^{history} is the total number of students who have learned the j -th knowledge point. sco_{ij} , rep_{ij} and com_{ij} are the test score, number of repeated watching instances of the video and the number of comments of the i -th student for the j -th knowledge point, respectively.

For these three input parameters, i.e., $\overline{sco}(j)$, $\overline{rep}(j)$ and $\overline{com}(j)$, the Analytic Hierarchy Process (AHP) [34] was used to determine the weights of the score model by professors majoring in education and working in universities in China to quantify the difficulty of knowledge mastery.

Table 1 shows the weights of the knowledge difficulty model generated by AHP. After calculation, the Consistency Ratio (CR) is equal to 0.033, which is less than 0.1. As a consequence, the result passed the consistency test.

Table 1. Weights of the knowledge difficulty model.

w_1	w_2	w_3
0.633	0.260	0.107

To validate the proposed model, data originating from the MOOCCubeX dataset [33] were adopted. Student exercise tests scores, student video watching behavior and student comment data were extracted from this dataset for further testing.

3.3. Knowledge Mastery Model

This study agrees with prior research that learner-specific exercise performance, to some extent, represents learners' level of knowledge acquisition. Researchers paid more attention to learners' exercise performance in previous studies while ignoring the critical impact of learning behaviors.

This work developed a knowledge mastery model based on student performance and learning behavior in MOOCs to dynamically evaluate students' learning state. The students' video watching behaviors of specific knowledge points and the normalized exercise test scores are the input to the knowledge mastery model. The output is the state of the learner. The specific parameters of this model are depicted in Table 2.

Table 2. Specific parameters of the knowledge mastery model.

Variable	Type	Illustration
sco_{ij}	Float	The normalized score of the i -th student for the exercise of the j -th knowledge point.
abn_{ij}	Bool	The i -th student has fast forward or multiple skips behavior when watching the j -th knowledge video.
$state_{ij}$	Int	The mastery of the i -th student to the j -th knowledge point.

Figure 2 illustrates the flow chart of student state judgment. The detailed process are as follows:

- (1) Students study the initial chapters;

- (2) Students complete the video and then perform the chapter exercise tests;
- (3) The learners' states are judged based on the knowledge mastery model;
- (4) Suitable knowledge points are recommended for learners based on their learning states.

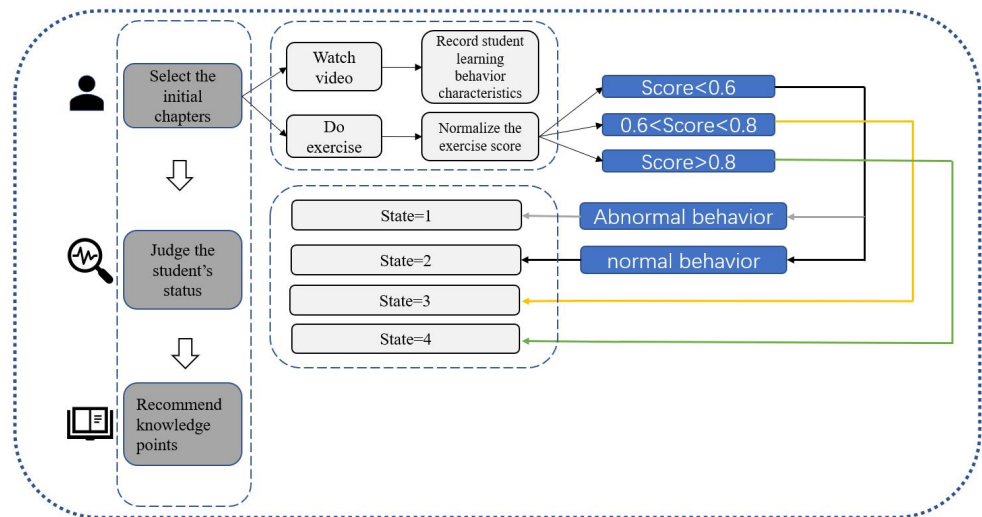


Figure 2. Flow chart of student status judgment. Abnormal behaviors include fast-forwarding or skipping when watching a video, and normal behaviors entail no such operations.

The *state* mentioned above is divided into four levels, from 1 to 4, which represent *unlearned*, *unmastered*, *insufficiently mastered* and *mastered*, respectively.

Unlearned (state = 1) means that a student's normalized test score is lower than 0.6; they fast-forward or skip when watching the video, which is not a normal watching behavior. The student should be assigned corresponding knowledge points for reviewing.

Unmastered (state = 2) means that the student's normalized exercise test score is lower than 0.6, but the entire video is watched without fast-forwarding or skipping. The student should be assigned corresponding knowledge points for review.

Insufficiently mastered (state = 3) means that the student's normalized exercise test score is above 0.6 but less than 0.8. The student should also be assigned corresponding knowledge points for review.

Mastered (state = 4) means that the student's normalized exercise score is greater than 0.8 and has mastered most of the knowledge point. The student should be assigned a new chapter to learn.

Students' learning behavior was recorded, and students' exercise test scores were normalized. When a student fast-forwarded or jumped while watching a video and the normalized test score was less than 0.6, the model concluded that the student's state was *unlearned* (state = 1). When the students' video watching behavior was normal, but the score was also less than 0.6, the model concluded that the student's state was *unmastered* (state = 2). If a student's score was greater than 0.6 but less than 0.8, the model concluded that the student's state was *insufficiently mastered* (state = 3). If a student's normalized test score was greater than 0.8, the model concluded that the student's state was *mastered* (state = 4).

4. Experiments and Results

4.1. MOOCCubeX Dataset

The MOOCCubeX [33] is a large scale open-access MOOC dataset originating from the XuetangX MOOC platform. The database is provided by XuetangX and uses a fine-grained approach to reorganize the data from a knowledge perspective. The abundant student learning data and course data strongly supported the completion of this study. For the

detailed description of the dataset, e.g., data type and format, etc., please refer to the original paper [33].

4.2. Personalized Learning Path Generation

Experiments of the dynamic learning path planning algorithm were based on the Application Programming Interfaces (APIs) provided by Tsinghua University, the XuetangX MOOC platform and the MOOCCubeX dataset. Students followed a fundamental logical sequence of knowledge (i.e., the original path sequence) at the beginning of the study, as mastery of advanced knowledge depends on mastering prerequisite knowledge. Consequently, the personalized learning path planning approach proposed in this study used a feedback strategy to form a personalized learning path based on learning states.

The personalized learning path planning algorithm is based on the knowledge difficulty model and knowledge mastery model mentioned in Sections 3.2 and 3.3, respectively. The learning path planning procedure is depicted as follows:

- (1) Students learn in the original learning sequence.
- (2) When students are judged as having completed each chapter, their knowledge mastery status is automatically updated based on the knowledge mastery model. The algorithm will automatically assign the corresponding knowledge point for the student if the algorithm determines that the student has not fully mastered the knowledge.
- (3) Unlearned and unmastered knowledge is added to the review list. Moreover, the knowledge points that are insufficiently mastered for the prerequisite knowledge (unmastered) are also added to the review list.
- (4) Knowledge points at different levels are arranged based on the prerequisite relationship, and knowledge points at the same level are arranged from easy to difficult ranks.
- (5) After reviewing the list of knowledge points in the above order, the student completes the test again, and their state will be updated. If the student's state is still in unlearned (state = 1), unmastered (state = 2) or insufficiently mastered (state = 3) states in this chapter, then the process goes back to Step 3.) If the student fully grasps the knowledge (state = 4), then the process goes back to Step 1 and continues to the next suggested chapter.

Figure 3 illustrates three typical learning paths generated by the proposed method. The learning paths shown in Figure 3a,b significantly differ from the paths planned by the algorithm, but the learning paths shown in Figure 3c coincide with the algorithm's planned paths.

According to Figure 3a, the first typical learning path is in linear style. Students learn all knowledge points sequentially and without retrospection, resulting in a significant number of insufficiently mastered, unlearned and unmastered knowledge points in the learning progress that have not been handled. Therefore, this type of learning progress is ineffective. According to Figure 3b, the second typical learning path is circular. Throughout the learning process, students independently reviewed the knowledge points. Knowledge points that were insufficiently mastered were missed, while some mastered knowledge was repeatedly reviewed. Therefore, this type of learning progress is inefficient as well.

Compared with Figure 3a,b, the third typical learning path automatically generated by the algorithm in Figure 3c precisely covers all unlearned, unmastered and insufficiently mastered knowledge. The knowledge points are also ranked based on prerequisite relationships and difficulty levels. This shows that the learning paths generated by the dynamic learning path planning algorithm are adaptive for students. The personalized learning path based on learning states saves time and increases learning efficiency.

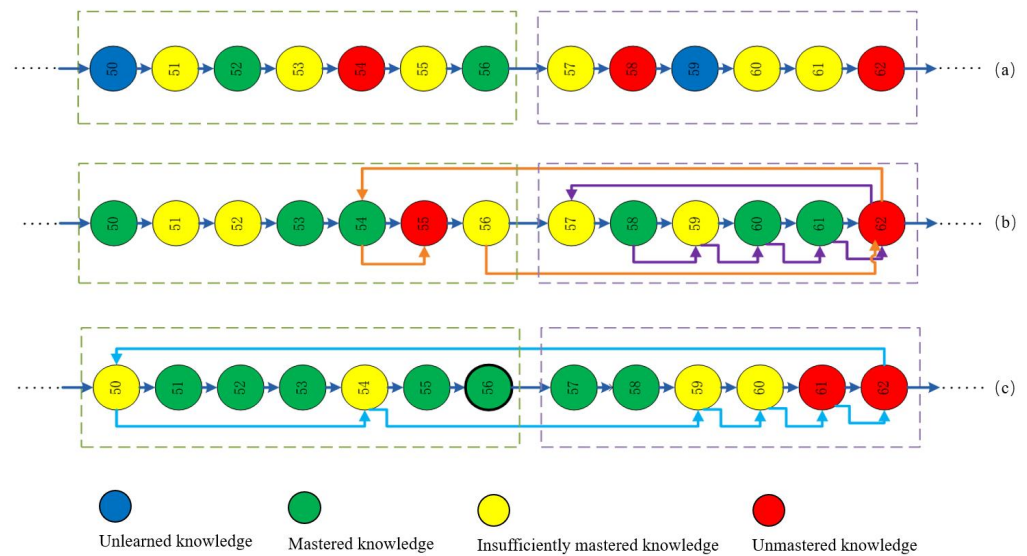


Figure 3. (a–c) is three typical learning paths of MOOC learning. (a) is a typical linear learning path of student in MOOCCubex; (b) is typical circular learning path of student in MOOCCubex; (c) is a learning path automatically generated by the algorithm.

5. Evaluation

To evaluate the effectiveness of the proposed approach, evaluation methods proposed by Nabizadeh et al. [9] are adopted to demonstrate the effectiveness of the algorithm and the feasibility of its potential applications.

5.1. Offline Evaluation

To evaluate the effectiveness of the proposed personalized learning path algorithm, the offline evaluation is adopted according to the following three steps:

- (1) Path Extraction: Comparing the existing learning paths of students in MOOCCubex with the learning paths generated by the proposed algorithm. The fragment of students whose learning path pattern of the sequence is in accordance with the learning path generated by the algorithm are extracted.
- (2) Student Classification: The students in the database are divided into two categories, one of which is the students mentioned in Step 1 (i.e., training path group), and the rest of the students in the dataset are seen as the control group (i.e., general student group).
- (3) Contraction: A series of evaluation methods were used to compare the effective behavior rate, completion rate and learning effect of the two groups of students. The details are described in the following sections.

5.2. Effective Behavior Rate

We compare the efficiency of learning behaviors in knowledge acquisition progress between the general student group and the training path group. Furthermore, we classified students' online learning behavior into two categories: effective behavior and ineffective behavior. The behavior of learning unlearned knowledge, unmastered knowledge and insufficiently mastered knowledge was defined as effective behavior. The behavior of learning mastered knowledge was defined as ineffective behavior. The learning path's efficiency was defined as follows:

$$\text{Effective behavior rate} = \frac{1}{N} \sum_{i=1}^N \left[1 - \frac{L_i^{rep}}{L_i} \right] \quad (5)$$

In Equation (5), N is the number of students, L_i is the length of the i -th student's learning path, and L_i^{rep} is the number of repeated learning of the mastered knowledge in the i -th

student's learning path. The greater the effective behavior rate value is, the less ineffective behavior on the learning path and the greater the learning efficiency.

After calculation, the average effective behavior rate of the training data group was 93%, while the average effective behavior rate of the general student group was only 71%. This indicates that the proposed dynamic learning path planning algorithm can accurately locate blind spots of knowledge. It can also develop effective learning paths and improve the efficiency of learning in MOOCs.

5.3. Completion Rate

To evaluate whether the proposed personalized learning path planning algorithm contributes to students' continuous learning in the MOOC, we calculate the completion rate of the training data group and the general student group.

The following equation is how the MOOC completion rate was calculated:

$$\text{Completion rate} = \frac{1}{N} \sum_{i=1}^N \left[1 - \frac{L_i^1}{K} \right] \quad (6)$$

In Equation (6), N is the number of students, and L_i^1 is the knowledge point with $state = 1$ in the i -th student's learning path, i.e., the unlearned knowledge.

After calculation, the MOOC completion rate of the training path group was 76%, while the MOOC completion rate of the general student group was only 57%. This indicates that the dynamic learning path planning algorithm can improve MOOC completion by monitoring students' knowledge mastery in real-time during the online learning process and reminding students to review the knowledge points that have not been mastered. This precisely answers the first research question, showing that personalized learning paths are helpful to support the persistent learning of MOOC learners and help to reduce dropout rates.

5.4. Learning Effect

Finally, the total online learning time and average exercise test score of the general student group and training path group were calculated, and the statistics are demonstrated as follows.

Table 3 shows that the average exercise test score of the training path group was significantly higher than that of the general student group. This indicates that the proposed learning path dynamic planning algorithm may improve the students' online learning efficiency (i.e., improve the exercise score and reduce the learning time). Meanwhile, by comparing the average total learning time of students in the same score band, it is found that the average learning time of the training path group is shorter than that of the general student group. The experiments proved that the proposed learning path dynamic planning algorithm may effectively improve students' online learning efficiency. This answers the first research question, showing that personalized learning paths help to improve learners' learning effectiveness.

Table 3. Statistics of the average online learning time and the final test average scores of students.

Statistics	Illustration	General Student Group	Training Path Group
$\overline{SCO}_{final}$	Average score of final test.	54.5	75.2
$num_{final}^{SCO \geq 60}$	Average online learning time of students with a score ≥ 60 .	11.3 h	9.1 h

6. Conclusions

This paper presents a dynamic personalized learning path generation algorithm that can provide suitable knowledge sequences to students based on their learning states and the prerequisite relationships for certain knowledge. We first constructed a knowledge difficulty model to automatically calculate the difficulty of knowledge points. Compared with previous methods, the model achieves the function of automatically calculating the difficulty of knowledge points based on other learners' historical learning behavior data. A knowledge mastery model was also constructed to diagnose learners' states by analyzing students' learning behavior data and normalizing exercise test scores. Different from previous studies, our method creatively takes students' online video watching behaviors and exercise test scores as basic parameters for diagnosing students' states. Finally, by incorporating both the knowledge difficulty model and knowledge mastery model, a personalized learning path was generated. The experiments show that the proposed algorithm can help learners to master knowledge and provides a unique learning sequence. Furthermore, the evaluation results demonstrate that the personalized learning path is capable of improving effective behavior rates, course completion rates and learning efficiency.

In future research, this approach is expected to be used in the traditional learning scenario to form a blended learning strategy. Similarly, as students' learning statuses change, the time dimension should also be considered to diagnose students' states in future work.

Author Contributions: Conceptualization, B.J., W.C., C.H. and Q.L.; Data curation, X.L., S.Y. and Y.K.; Funding acquisition, B.J.; Investigation, W.C., C.H. and Q.L.; Methodology, B.J., X.L., S.Y. and Y.K.; Project administration, B.J.; Resources, W.C., C.H. and Q.L.; Software, Y.K.; Visualization, Y.K.; Writing—original draft, X.L., S.Y. and Y.K.; Writing—review and Editing, B.J., W.C., C.H. and Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Grant No. 61907025) and the Natural Science Foundation of Jiangsu Higher Education Institutions of China (Grant No. 19KJB520048).

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: <https://github.com/THU-KEG/MOOC CubeX> (accessed on 1 February 2022).

Acknowledgments: The authors would like to thank all the anonymous reviewers for their valuable suggestions to improve this work. Thanks also go to Haoran Xu for his valuable suggestions on the whole pipeline of the algorithm and to Yan Wang, Yuzhou Dai and Feihu Jiang for their great efforts on experimental data processing.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xiao, J.; Jiang, B.; Xu, Z.; Wang, M. The usability research of learning resource design for MOOCs. In Proceedings of the 2014 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE), Wellington, New Zealand, 8–10 December 2014; pp. 277–282.
2. Nawrot, I.; Doucet, A. Building engagement for MOOC students: introducing support for time management on online learning platforms. In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Korea, 7–11 April 2014; pp. 1077–1082.
3. Wang, L.; Wang, H. Learning behavior analysis and dropout rate prediction based on MOOCs data. In Proceedings of the 2019 10th International Conference on Information Technology in Medicine and Education (ITME), Qingdao, China, 23–25 August 2019; pp. 419–423.
4. US Department of Education, Office of Educational Technology. Reimagining the Role of Technology in Education: 2017 National Education Technology Plan Update. In *National Education Technology Plan Update*; US Department of Education: Washington, DC, USA, 2017.
5. Fasihuddin, H.; Skinner, G.; Athauda, R. Towards an adaptive model to personalise open learning environments using learning styles. In Proceedings of the International Conference on Information, Communication Technology and System (ICTS) 2014, Surabaya, Indonesia, 24 September 2014; pp. 183–188.
6. Meng, L.; Zhang, W.; Chu, Y.; Zhang, M. LD-LP Generation of Personalized Learning Path Based on Learning Diagnosis. *IEEE Trans. Learn. Technol.* **2021**, *14*, 122–128. [CrossRef]

7. Graf, S.; Kinshuk; 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*, 3–14.
8. Auvinen, T. Harmful study habits in online learning environments with automatic assessment. In Proceedings of the 2015 International Conference on Learning and Teaching in Computing and Engineering, Taipei, Taiwan, 9–12 April 2015; pp. 50–57.
9. Nabizadeh, A.H.; Gonçalves, D.; Gama, S.; Jorge, J.; Rafsanjani, H.N. Adaptive learning path recommender approach using auxiliary learning objects. *Comput. Educ.* **2020**, *147*, 103777. [[CrossRef](#)]
10. Cai, D.; Zhang, Y.; Dai, B. Learning path recommendation based on knowledge tracing model and reinforcement learning. In Proceedings of the 2019 IEEE 5th International Conference on Computer and Communications (ICCC), Chengdu, China, 6–9 December 2019; pp. 1881–1885.
11. Li, W.; Zhang, L. Personalized learning path generation based on network embedding and learning effects. In Proceedings of the 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 18–20 October 2019; pp. 316–319.
12. Wanichsan, D.; Panjaburee, P.; Chookaew, S. Enhancing knowledge integration from multiple experts to guiding personalized learning paths for testing and diagnostic systems. *Comput. Educ. Artif. Intell.* **2021**, *2*, 100013. [[CrossRef](#)]
13. Niknam, M.; Thulasiraman, P. LPR: A bio-inspired intelligent learning path recommendation system based on meaningful learning theory. *Educ. Inf. Technol.* **2020**, *25*, 3797–3819. [[CrossRef](#)]
14. Dwivedi, P.; Kant, V.; Bharadwaj, K.K. Learning path recommendation based on modified variable length genetic algorithm. *Educ. Inf. Technol.* **2018**, *23*, 819–836. [[CrossRef](#)]
15. Adorni, G.; Koceva, F. Educational concept maps for personalized learning path generation. In Proceedings of the Conference of the Italian Association for Artificial Intelligence, Genova, Italy, 29 November–1 December 2016 ; Springer: Cham, Switzerland, 2016; pp. 135–148.
16. Rohloff, T.; Sauer, D.; Meinel, C. On the acceptance and usefulness of personalized learning objectives in MOOCs. In Proceedings of the the Sixth (2019) ACM Conference on Learning@ Scale, Chicago, IL, USA, 24–25 June 2019; pp. 1–10.
17. Christudas, B.C.L.; Kirubakaran, E.; Thangaiah, P.R.J. An evolutionary approach for personalization of content delivery in e-learning systems based on learner behavior forcing compatibility of learning materials. *Telemat. Inform.* **2018**, *35*, 520–533. [[CrossRef](#)]
18. Feng, X.; Xie, H.; Peng, Y.; Chen, W.; Sun, H. Groupized learning path discovery based on member profile. In Proceedings of the International Conference on Web-Based Learning, Shanghai, China, 7–11 December 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 301–310.
19. Vanitha, V.; Krishnan, P.; Elakkiya, R. Collaborative optimization algorithm for learning path construction in E-learning. *Comput. Electr. Eng.* **2019**, *77*, 325–338. [[CrossRef](#)]
20. Yang, Y.J.; Wu, C. An attribute-based ant colony system for adaptive learning object recommendation. *Expert Syst. Appl.* **2009**, *36*, 3034–3047. [[CrossRef](#)]
21. Xie, H.; Chu, H.C.; Hwang, G.J.; Wang, C.C. Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017. *Comput. Educ.* **2019**, *140*, 103599. [[CrossRef](#)]
22. Yu, C.H.; Wu, J.; Liu, A.C. Predicting learning outcomes with MOOC clickstreams. *Educ. Sci.* **2019**, *9*, 104. [[CrossRef](#)]
23. Goulden, M.C.; Gronda, E.; Yang, Y.; Zhang, Z.; Tao, J.; Wang, C.; Duan, X.; Ambrose, G.A.; Abbott, K.; Miller, P. CCVis: Visual analytics of student online learning behaviors using course clickstream data. *Electron. Imaging* **2019**, *2019*, 681-1–681-12. [[CrossRef](#)]
24. Xia, M.; Sun, M.; Wei, H.; Chen, Q.; Wang, Y.; Shi, L.; Qu, H.; Ma, X. Peerlens: Peer-inspired interactive learning path planning in online question pool. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Scotland, UK, 4–9 May 2019; pp. 1–12.
25. Liu, H.; Li, X. Learning path combination recommendation based on the learning networks. *Soft Comput.* **2020**, *24*, 4427–4439. [[CrossRef](#)]
26. Dalipi, F.; Imran, A.S.; Kastrati, Z. MOOC dropout prediction using machine learning techniques: Review and research challenges. In Proceedings of the 2018 IEEE Global Engineering Education Conference (EDUCON), Santa Cruz de Tenerife, Spain, 17–20 April 2018; pp. 1007–1014.
27. Fung, S.; Tam, V.; Lam, E.Y. Enhancing learning paths with concept clustering and rule-based optimization. In Proceedings of the 2011 IEEE 11th International Conference on Advanced Learning Technologies, Athens, GA, USA, 6–8 July 2011; pp. 249–253.
28. Zhu, H.; Tian, F.; Wu, K.; Shah, N.; Chen, Y.; Ni, Y.; Zhang, X.; Chao, K.M.; Zheng, Q. A multi-constraint learning path recommendation algorithm based on knowledge map. *Knowl.-Based Syst.* **2018**, *143*, 102–114. [[CrossRef](#)]
29. Liang, J.; Li, C.; Zheng, L. Machine learning application in MOOCs: Dropout prediction. In Proceedings of the 2016 11th International Conference on Computer Science & Education (ICCSE), Nagoya, Japan, 23–25 August 2016; pp. 52–57.
30. Li, X.; Xie, L.; Wang, H. Grade prediction in MOOCs. In Proceedings of the 2016 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC) and 15th Intl Symposium on Distributed Computing and Applications for Business Engineering (DCABES), Paris, France, 17 July 2016; pp. 386–392.
31. Shi, D.; Wang, T.; Xing, H.; Xu, H. A learning path recommendation model based on a multidimensional knowledge graph framework for e-learning. *Knowl.-Based Syst.* **2020**, *195*, 105618. [[CrossRef](#)]

-
32. Rahayu, N.W.; Ferdiana, R.; Kusumawardani, S.S. A systematic review of ontology use in E-Learning recommender system. *Comput. Educ. Artif. Intell.* **2022**, *13*, 100047. [[CrossRef](#)]
 33. Yu, J.; Wang, Y.; Zhong, Q.; Luo, G.; Mao, Y.; Sun, K.; Feng, W.; Xu, W.; Cao, S.; Zeng, K.; et al. MOOCCubeX: A Large Knowledge-centered Repository for Adaptive Learning in MOOCs. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Online, 1–5 November 2021; pp. 4643–4652.
 34. Saaty, T.L.; Kearns, K.P. Chapter 3—The Analytic Hierarchy Process. In *Analytical Planning*; Pergamon: Oxford, UK, 1985; pp. 19–62. [[CrossRef](#)]