

A Clustering Knowledge Tracking Algorithm for Learners Based on Dynamic Key-Value Memory Network

**Yu Lasheng
Zheng Xiaopeng**

School of Computer Science and Engineering
Central South University, Changsha 410083, China

ABSTRACT

In the existing classic deep knowledge tracking model, only the correctness of the student's historical question data is concerned, but the other characteristics of the student are not paid attention to. Human learning is a process of practice. Students gradually improve their knowledge mastery through learning, but the learning ability of each student is not the same. Only a single characteristic of correctness cannot completely distinguish the learning ability of different students. The existing BKT or DKVMN algorithms assume that the learning ability of each student is the same. In order to solve the above mentioned problems, this paper divides students according to their learning ability based on the DKVMN model, and then dynamically divides students into corresponding groups according to the time interval, improves the personalized expression of the DKVMN algorithm, and makes the algorithm more reasonable predictions.

Keywords: Knowledge tracing; Learning Ability; Dynamic Key-Value Memory Network; adaptive learning

INTRODUCTION

The traditional classroom education model is mainly based on teachers and books as the center of education and teaching. With the development of the Internet and artificial intelligence, the personalized adaptive learning model of "Internet + education" has become the new normal. For future education, it is very important to tap the creativity and potential of each student and adjust the teaching strategy according to each student's understanding and mastery of the basic knowledge they have learned. However, most of the current online platforms only carry out offline teaching activities in an online form, and teaching activities are often carried out in an offline way. Although many platforms provide massive learning videos and exercise resources, It is difficult to achieve one-to-one targeted supervision and teaching, and cannot provide individual learning programs for students, and students are often more willing to practice some knowledge points that they have a good degree of mastery, resist or selectively ignore some of the more mastery, which leads to low learning efficiency in students' daily learning. In addition, most platforms provides students with a large number of exercise questions but does not provide corresponding guidance as the questions go deep, which makes the learning process of students usually unplanned, resulting in Students face greater learning pressure. In order to solve the existing online education problems, it is necessary to establish a model that can adjust teaching according to the students' current mastery of knowledge, provide students with a learning path that suits them, and allow the platform to provide targeted learning solutions.

The purpose of this paper is to study the pain points of knowledge tracking in personalized education, study the existing deep knowledge tracking algorithms, design an algorithm with knowledge tracking as the core and solve some problems in the algorithm to achieve better prediction results and performance.

RELATE WORK

With the advancement of modern information technology, the integration of the Internet and education has become closer, and the individualization and intellectualization of education are gradually changing the problem of undifferentiated education in traditional education. Adaptive Learning was first proposed by American scholar Peter Brusilovsky, who believes that adaptive learning systems use machine learning, deep learning and other algorithms to model learners. In the adaptive learning system, it is very important to help students accurately predict the answering of questions. Knowledge tracking is to conduct a cognitive level modeling of the historical performance of the students' questions, which can better grasp the students' mastery of various knowledge, and accurately grasp the students' knowledge status to carry out targeted questions for the students to improve the efficiency of learning. At present, the research methods proposed for knowledge tracking are mainly divided into three algorithms: Bayesian knowledge tracking algorithm, item response theory and deep knowledge tracking algorithm.

This paper mainly studies the problem of knowledge tracking in the field of educational data mining. By comparing the BKT, IRT, DKT and DKVMN knowledge tracking models and their extended models, we found some problems in the models. For example, the BKT model defines two assumptions: one is that students will not forget as long as they learn knowledge points; the other One is that all the questions have the same difficulty, and the order between the questions and the knowledge points is not considered. These problems lead to the low prediction accuracy of the BKT model; the IRT model cannot reflect the students' learning experience from the mistakes to master the knowledge; The DKT model cannot output the mastery of a single knowledge point of students, and there is a long-term forgetting problem; although the DKVMN model has the advantages of the BKT model and the DKT model, it does not consider the group characteristics of students and the cognitive process of students. In order to solve the problems of the DKVMN model, this paper expands and improves the DKVMN algorithm as DKVMN-SC algorithm which combining the group characteristics of students and considering the different learning abilities of students.

GROUP ANALYSIS OF STUDENTS' LEARNING ABILITY

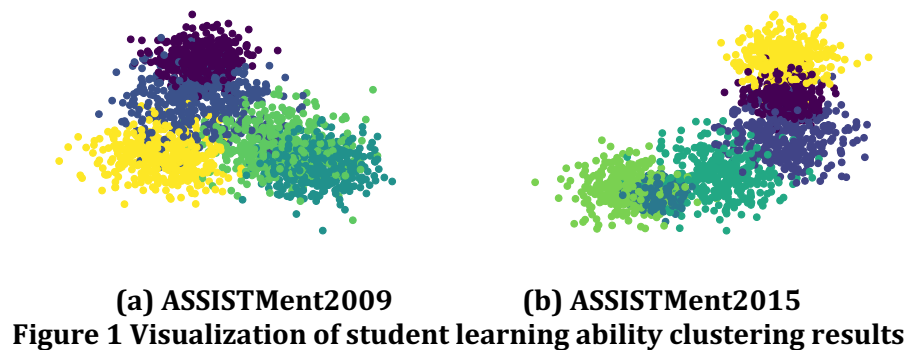
Student learning is a practical process, usually through continuous practice to master knowledge gradually, but the learning situation of students is also affected by personal learning ability, and the learning ability of students will be change dynamically with more or less practice. Based on this point of view, this section proposes a calculation method for the evaluation of students' learning ability and a dynamic division strategy.

Calculation of student learning ability

In the offline learning process, teachers usually group students into groups, and different groups have different learning abilities to facilitate different guidance. In the research in the field of education, researchers divide students into different groups with similar abilities

according to the different performances of students in learning the same content, and provide personalized education guidance for a group of students with similar abilities.

To analyze how students' learning abilities are evaluated, this paper analyzes students' past interactions in the system on two real public data sets, ASSISTMent2009 and ASSISTMent2015. For example, on the ASSISTMent2015 data set, it is finally processed into 19,840 students, 100 question and answer questions, and 614,871 question records after preprocessing. The paper visualizes the student data in the two data sets, and the results are as follows.



Figures 1-a and 1-b represent the ASSISTMent2009 dataset and the ASSISTMent2015 dataset respectively. The visualization results of clustering are based on the correctness of the student's problem-making process and the average request prompt rate. It can be clearly seen from the figure that the learning ability of students has obvious clustering characteristics, and the result is in line with expectations. Students with different learning abilities have different grasping speeds of the knowledge displayed in the same problem. Modeling students with different learning abilities groups can help better track students' knowledge.

In order to obtain the calculation index of students' learning ability, the analysis of the data set ASSISTMents used in this paper shows that the relationship between the request prompt situation and the correct rate of students in a topic exercise can be divided into three types: (a) the students can answer the questions smoothly and correctly without requesting prompts or requesting few prompts during the question. (b) students usually need to use all the prompt information contained in the question, but the final prompt information for each question All are the answers to the questions; and (c) students usually need to request a lot of prompt information, but they can't answer the questions well. This also shows that the student's learning ability has a great relationship with the proportion of students requesting prompt information in the process of doing questions.

Because of the fast convergence speed of k-means algorithm, the clustering effect is more efficient when clustering analysis is performed frequently. Therefore, this paper uses the k-means clustering algorithm (k-means) to divide the student groups, and divide groups with similar learning abilities into a group. The k-means algorithm in this paper determines the similarity between sample points by calculating the Euclidean distance between the sample points. The data set for the model test is the public data set ASSISTments. This data set contains the number of times students have asked for prompts when answering this question and the total number of prompts for this question. It should be noted that when the request reaches the

upper limit of prompts The last hint at the time is the answer to the question. Therefore, the learning ability assessment considers the learner's correct rate of questions and the proportion of the number of prompts requested in the process of doing the questions to the total number of prompts, and converts these two features into a data vector. The formula is as follows:

$$C(x_i) = \frac{x_i == 1}{|N_i|} \quad (1)$$

$$Hint(x_i) = \frac{\sum_{j=1}^m h_{ij}}{|H_i|} \quad (2)$$

$$R(x_i) = (C(x_i), Hint(x_i)) \quad (3)$$

$$d_s = (R(x_1), R(x_2), \dots, R(x_n)) \quad (4)$$

Where $C(x_i)$ represents the correct rate of student s answering x_i knowledge points, $Hint(x_i)$ represents the average prompt request probability of student s answering x_i knowledge points, $|N_i|$ is the sum of the number of practice times of x_i , h_{ij} represents the number of knowledge points request prompts, H_i represents the total number of prompts for this knowledge point. The vector $R(x_i)$ represents the student's ability at the knowledge point x_i , and d_s is the vector representation of the learning ability of the student s .

When k -means are clustered on a student, the representation vector of the student's abilities obtained from the above to evaluate the classification to which the student belongs, as shown in Formula 3-5.

$$AbilitySSE = \underset{c=1}{\operatorname{argmin}} \sum_{d_i \in C_c}^k ||d_i - \mu_c||^2 \quad (3-5)$$

Where μ_c is the centrality of each cluster, c represents the number of clustering centers, which are assigned to the corresponding groupings by evaluating the student's abilities.

Answers record dynamic classification strategies

Because students' learning ability changes dynamically with the number of questions, students should be reassigned to a group of students with similar learning ability every once in a while, and the students' ability should be dynamically adjusted and clustered. This paper determines that the time interval is calculated once each time, and that it is dynamically assigned to different groups within each interval. The average accuracy of students over time varies as shown in Figure 2.

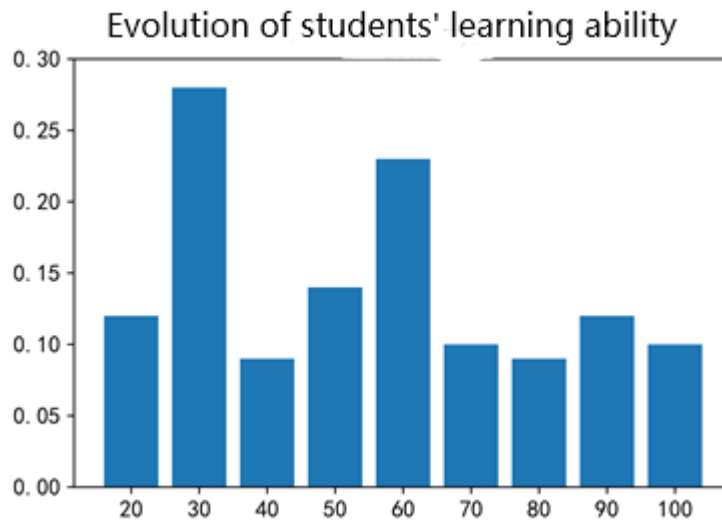


Figure 2 Rate of evolution of students' learning ability

The data in Figure 2 are based on changes in the average accuracy of students over time, which proves that students' learning abilities are dynamic, with the deletion of students whose records are few because they do not make sense for data with very few studies. In the dynamic classification of the student training sequence, the last part with insufficient time interval is filled with -1, for example, a student has 44 questions recorded and the dynamic split interval is 10, then the first four groups each includes 10 records, the last group includes 4 normal records and 6 records with -1.

Knowledge tracking algorithm based on dynamic division of learning ability

Through the analysis of students' learning ability in the previous section, students carry out the same learning process but the effect of learning is not the same, students' learning is limited by factors such as their own ability, however, deep knowledge tracking models such as DKT and DKVMN all assume that participating students have identical learning abilities, which is obviously not in line with the student learning process and affects the predictive effect of the model. This section improves the original DKVMN algorithm and proposes a learner clustering knowledge tracking algorithm(DKVMN-SC)to solve the problem of students' different learning abilities. Based on the historical data of students in the learning system, the DKVMN-SC algorithm divides students into different groups and re-evaluates students' learning ability after each interval, and dynamically assigns students to groups with similar learning abilities based on the results of the evaluation results in the previous interval, each group has its own training parameters, and all groups use the same key value matrix to improve the individual expression of the model.

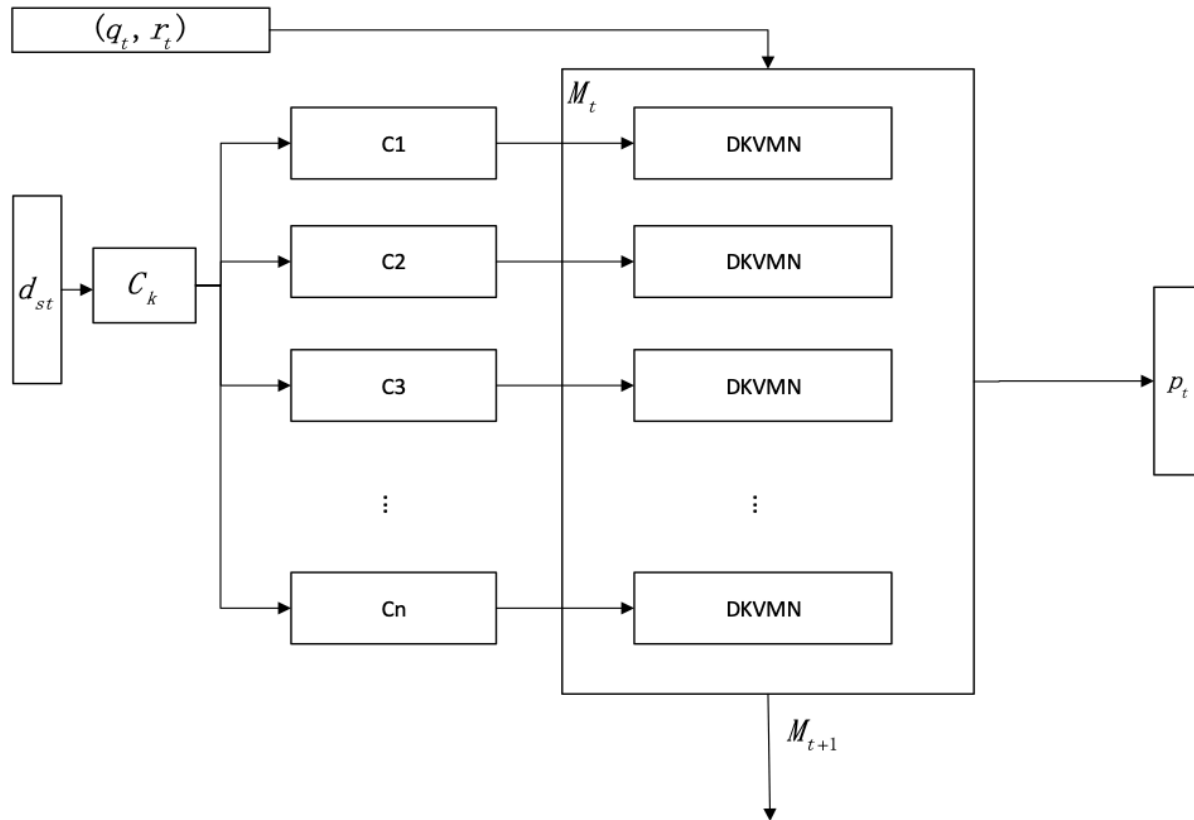


Figure 3 DKVMN-SC algorithm for learners using dynamic key value memory networks

Figure 3 illustrates the overall model structure of the DKVMN-SC, where input d_{st} is the vector representation of the student's ability to learn in the t-period, (q_t, r_t) represents a student's record of a question-and-answer behavior, $C_k = \{C_1, C_2, C_3, \dots, C_n\}$ represents the number of student groups classified, assuming they are divided into n groups, and different training models are used for each group to distinguish the impact of learning ability on algorithms. The algorithm is divided into two main steps, the detailed process is as follows:

- (1) Set time intervals and evaluate students' learning ability. By setting intervals to split data, such as setting an interval of 10, whenever the student has 10 answer records through the model, the student's learning ability needs to be re-evaluated against those 10 records to record the student's learning ability d_{st} , K-means clustering is performed to dynamically divide students into groups with similar abilities.
- (2) Record the students' questions (q_t, r_t) , and enter them into the DKVMN algorithm for each group, train the parameters individually for each category of data, and all students use the same key value matrix, each grouping maintains the corresponding parameters.

The DKVMN-SC algorithm is described as follows :

Algorithm 3.1:DKVMN-SC algorithm

Input: Student answer record data(q_t, r_t) ;

The number of requests for prompt information and the total number of prompts(h_t, H_t) ;

The number k of students grouped, the time interval t between questions;

Output: Prediction results of students answering question;

1. If the student's cumulative record of doing questions exceeds t:

Group through the k-means algorithm to determine which grouping;

Clear the learning ability d_s as empty;

2. Make cumulative updates of the store variables for the knowledge points currently included in the topic $R(x_i)$;

3. Dynamic key value memory network training phase:

4. Select the model parameters corresponding to the grouping(a,b,w1,b1,w2,b2,d,ba) = [A[k], B[k], W1[k], B1[k], W2[k], B2[k], D[k], Ba[k]];

5. Get the embedded matrix similarity_score through full-connection calculations and update parameter a

6. Calculating the weight wt sym.SoftmaxActivation (similarity_score);

7. Based on the calculated relevance w_t Read the value matrix and calculate the summary vector f_t ,
mx.sym.Reshape(mx.sym.batch_dot(read_weight, memory)) ;

8. The resulting aggregate vector is predicted by passing through the full connection layer of the Sigmoid activation function;

9. By embedding the matrix(q_t, r_t), the erasure vector is obtained through the activation function;

10. Update the knowledge point mastery storage value matrix;

11. Calculate the loss value of the iteration and update the array of model parameters;

12. Output the predicted probability obtained during the training of dynamic key value memory network;

1 Experimental design and result analysis

Dataset

To validate the proposed knowledge-tracking model, this paper evaluates the model presented using two data sets, ASSISTMents2009 and ASSISTMents2015. ASSISTMents is an online tutoring system created in 2004 to record math problems for middle and high school students. This paper is compared experimentally on BKT,DKT,DKVMN, and DKVMN-SC algorithms.

Because of the many repetitive data sets in the original ASSISTMents dataset, although the authorities subsequently released a new version of the revised data set skill_builder_data_corrected, the data set had some recorded skill_name empty, and some students had little record of learning in the system and dropped out of school. Therefore, the original dataset is filtered. In all data sets, problems are usually marked as only one skill, but very few may be related to two or three skills, and some researchers divide data records into multiple single skill records, but this way of processing data artificially improves the predicted results, so the data is deleted before training, as shown in Table-1.

Table 1 Introduction to the data set

Datasets	#Students	#Question	#Interactions
ASSISTMents2009	4151	110	315437
ASSISTMents2015	19840	100	614871

The student information that is primarily recorded in the dataset is shown in Table 2. In addition to containing each interaction correctness (correct), student id (user_id), and question id (problem_id), the ASSISTMents dataset contains information such as the skill id (skill_id) associated with the problem, the total number of prompts (hint_total), the number of prompts requested by students when they ask for a question (hint_count), and so on.

Introduction to the Data Set in Table 2

Field	Description
user_id	Student id
problem_id	Question number
correct	The correctness of the answer to the question
skill_id	The skill number that the question contains
hint_count	The number of times a student has made a question request prompt
hint_total	The total number of prompts included in the question
attempt_count	The number of times the student tried this question

Experimental design

The DKVMN-SC algorithm presented in this paper is an extension of the original DKVMN algorithm, which has the same loss function, and each model hidden layer has 200 fully connected hidden nodes. In order to speed up the training process of the model using the mini-batch method for training, the model momentum is set to 0.9, the standard shear threshold setting is set to 50, the learning rate of the model is set to 0.01, the first training is not grouped, the resulting parameters are used as the initial value of all groupings.

In conducting experiments, this paper uses 5x cross-validation to predict two data sets, randomly dividing each dataset into 80% of the training data and 20% of the test data, and averaging each knowledge tracking model by 5 experiments. In the DKVMN-SC algorithm experiment, different cluster centers and different dividing intervals were selected to analyze the effect of the algorithm's predictive performance.

Analysis of experimental results

To verify that the DKVMN-SC algorithm proposed in this paper is influenced by the number of clustering centers and dynamic intervals, the number of clustering centers from 1 to 10 and dynamic division intervals of 10 to 100 are cross-experimented in two data sets, ASSISTMent2009 and ASSISTMent2015, respectively, to determine the impact of these two factors on algorithm accuracy. For example, when you set the number of cluster centers to 5 and the interval to 30, it means that students are divided into 5 groups based on their learning ability, and each time a student performs 30 exercises, the grouping of students needs to be recalculated based on those 30 records, as shown below.

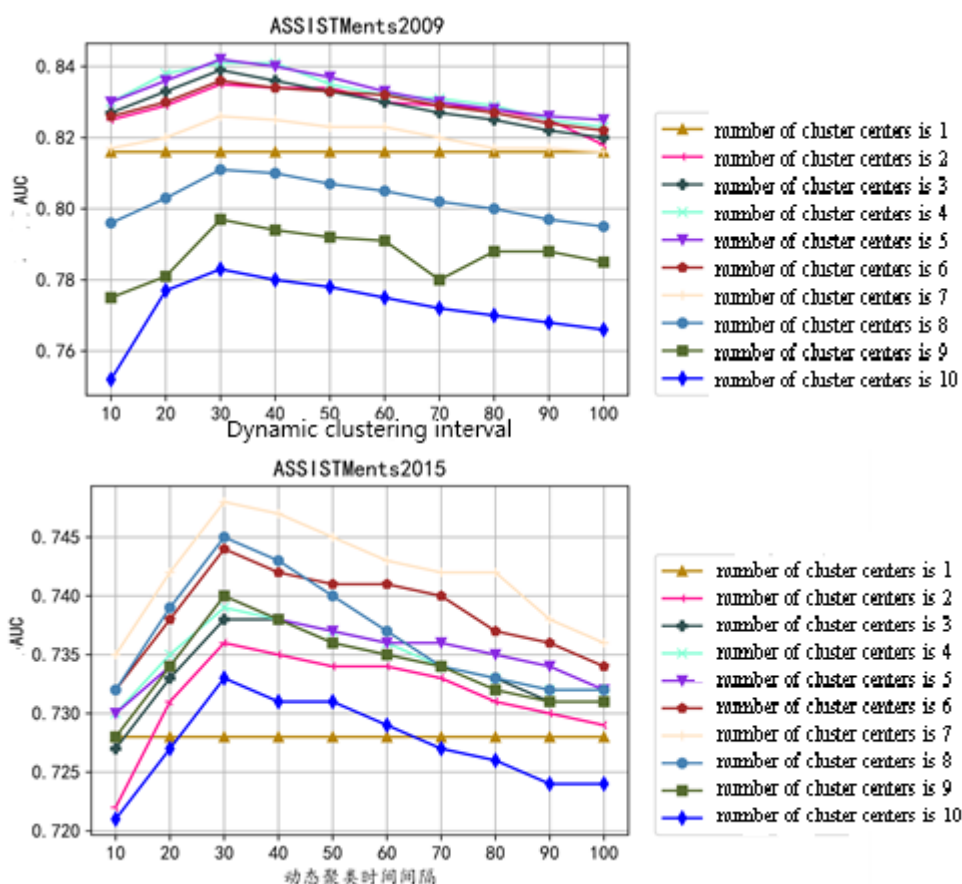


Figure 4 ASSISTMents2009 Cluster Centers and Divide Intervals vs. AUC Relationships
Figure 5 ASSISTMents2015 Cluster Centers and Divide Intervals vs. AUC Relationships

In Figures 4 and 5, the effects of different clustering centers and different time intervals on the DKVMN-SC algorithm AUC were tested, and 10, 20, 30, 40, 50, 60, and dynamic intervals were set, respectively. /b120> 70, 80, 90, and 100. Experimental results show that the algorithm achieves optimal predictive performance when setting dynamic intervals of 30 in both ASSISTMent2009 and ASSISTMent2015 data sets. When setting dynamic division intervals of less than 30, that is, when students answer 10 or 20 questions to regroup students, the predictive performance of the model is reduced due to the influence of "blinding" and "mistake" answers, and when setting dynamic dividing intervals of more than 30, it does not reflect that students' learning ability is affected by the number of exercises, which reduces the dynamics of the model and leads to a decrease in model accuracy. Because the number of students in the ASSISTMent2015 dataset was 19840, compared with 614,871 questions recorded, there were fewer records per student than in ASSISTMent2009, resulting in less stable and volatile algorithm accuracy. It is important to note that when the algorithm sets the number of clustering centers to 1, it degrades into a normal DKVMN algorithm, and is not affected by the time interval, and the prediction accuracy remains the same.

As can be seen from the line chart of the experimental results of the two datasets, when the dynamic interval is set to 30, the optimal AUC value for the cluster center number is obtained regardless of the number of cluster centers. Therefore, the effect of clustering centers on

algorithm performance is presented separately, as shown in Figures 6 and 7. Algorithms for different clustering centers present results at dynamic clustering intervals of 30.

As can be seen from Figure 6, the algorithm achieves the best experimental results when the number of student clusters is 5 in the ASSISTMent2009 dataset. When the number of cluster centers is less than 5, the AUC value of the algorithm increases as the number of cluster centers increases, and the AUC value of the algorithm drops rapidly when the number of cluster centers is greater than 5. This result is mainly due to the fact that there are only 4151 students in the ASSISTMent2009 dataset, and that too many clustering centers result in less data in each grouped model, resulting in inadequate model training, and sometimes large differences in learning ability between students in each group when the number of clustering centers is too small, so that the AUC value of the algorithm increases with the number of clustering centers.

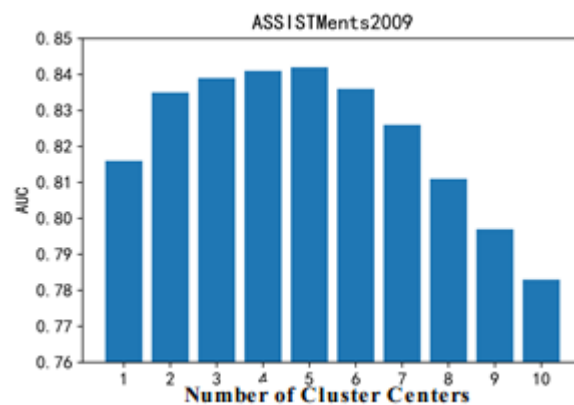


Figure 6 the number of cluster centers for the assistments 2009 dataset relates to auc

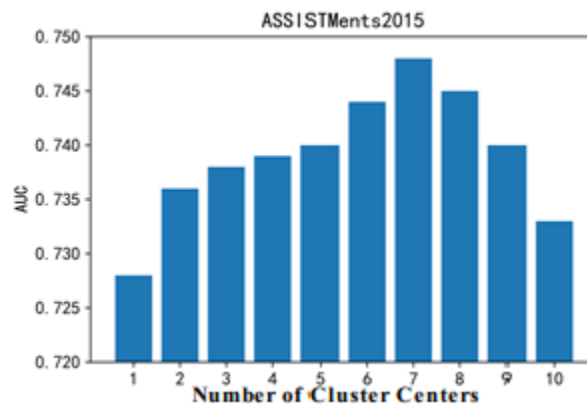


Figure 7 ASSISTMents2015 Data Set ClusterIng Center Number vs. AUC Relationship

Similarly, as can be seen from Figure 7 in the ASSISTMents 2015 dataset, because there are more students in the dataset than in the ASSISTMents 2009 dataset, the model can achieve the optimal AUC value when the cluster center number is set to 7 and has a larger student population.

Table 3-3 AUC on the knowledge-traced model test set

DataSet	BKT	DKT	DKVMN	DKVMN-SC
ASSISTMents2009	0.628	0.802	0.816	0.832
ASSISTMents2015	0.635	0.724	0.728	0.738

The forgotten enhanced learner clustering knowledge tracking algorithm presented in this paper is paired with the AUC results of the BKT, DKT, DKVMN algorithms, as shown in Table 3-3. The DKVMN-SC algorithm presented in this paper performs better than the other models on both data sets, and the results show that in the ASSISTMents 2009 dataset, this paper is compared to the BKT model (0.628), the DKT model (0.802), and the DKVMN model (0.816). The AUC for the DKVMN-FA model is 0.832. The AUC for the DKVMN-FA model in this paper in the ASSISTMents 2015 dataset is 0.738, an increase of 10.3% over the BKT model. The DKT model and the DKVMN model, which are tracked by depth, increased by 1.4% and 1%, respectively.

SUMMARY

In this paper, we propose a learner clustering knowledge tracking algorithm based on a dynamic key-value memory network, which solves the problem that the traditional model does not distinguish the learning ability of students. This chapter defines the evaluation method of students' learning ability, transforms students' characteristics into data vectors, and dynamically divides students into groups with similar learning ability according to the time interval through k-means, and improves the DKVMN algorithm through the multi-group model. Personalized expression. Finally, the data set used in this article is analyzed, the experimental design and experimental results are introduced, and the influence of the number of clustering centers and the number of time intervals on the model is verified. The experiment proves that the DKVMN-SC algorithm can make predictions more accurately.

ACKNOWLEDGEMENTS

This work was supported by the first-class curriculum construction project of Hunan Province and the first-class curriculum construction project of Central South University

References

- [1]. Jiao Feng. The application of artificial intelligence + education and the construction of intelligent education platform [J]. *Information and Communication*, 2020, No.215(11):146-149.
- [2]. Yudelson M V, Koedinger K R, Gordon G J. Individualized bayesian knowledge tracing models[C]//International conference on artificial intelligence in education. Springer, Berlin, Heidelberg, 2013: 171-180.
- [3]. Lv Haiyan, Zhou Lijun, Zhang Jie. Research on the application of educational data mining in the analysis of student online learning behavior under the background of big data[J]. *Computing Technology and Automation*, 2017, 036(001):136-140.
- [4]. Xu Y, Wang Z, Shang J S. PAENL: personalized attraction enhanced network learning for recommendation[J]. *Neural Computing and Applications*, 2021: 1-11.
- [5]. Xu Moke, Wu Wenjun, Zhou Xuan, et al. Multi-knowledge point knowledge tracking model and visualization research[J]. *Audio-visual Education Research*, 2018, 39(10):55-61.
- [6]. Fang L, Tuan L A, Hui S C, et al. Personalized question recommendation for English grammar learning[J]. *Expert Systems*, 2018, 35(2): e12244.

- [7]. Vie J J, Kashima H. Knowledge tracing machines: Factorization machines for knowledge tracing[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(01): 750-757.
- [8]. Minn S, Yu Y, Desmarais M C, et al. Deep knowledge tracing and dynamic student classification for knowledge tracing[C]//2018 IEEE International conference on data mining (ICDM). IEEE, 2018: 1182-1187.
- [9]. Zhang J , Shi X , King I , et al. Dynamic Key-Value Memory Networks for Knowledge Tracing[C]// International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2017:765-774.
- [10]. Liu D , Zhang Y , Zhang J , et al. Multiple Features Fusion Attention Mechanism Enhanced Deep Knowledge Tracing for Student Performance Prediction[J]. IEEE Access, 2020, 8:194894-194903.
- [11]. Liu Q, Huang Z, Yin Y, et al. Ekt: Exercise-aware knowledge tracing for student performance prediction[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 33(1): 100-115.
- [12]. Li B , Liu Y , Wang X . Gradient Harmonized Single-stage Detector[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019,33:8577-8584.
- [13]. [1] Kovacic J . Learning parameters of Bayesian networks from datasets with systematically missing data: A meta-analytic approach[J]. Expert Systems with Application, 2020, 141(Mar.):112956.1-112956.11.
- [14]. Lee J, Yeung D Y. Knowledge query network for knowledge tracing: How knowledge interacts with skills[C]//Proceedings of the 9th International Conference on Learning Analytics & Knowledge. 2019: 491-500.
- [15]. Nosofsky RM, Cao R, Harding SM, Shiffrin RM. Modeling short- and long-term memory contributions to recent event recognition. Journal of experimental psychology. Learning, memory, and cognition. 2021;47(2):316-42.