ICIC Express Letters
Volume **17**, Number **2**, February **2023**
 ISSN 1881-803X
pp. **217–225**

KNOWLEDGE POINT DIFFICULTY CLUSTERING ALGORITHM BASED ON MULTIDIMENSIONAL TIME SERIES DATA AND MAXIMUM FREQUENT SUBGRAPH

Zhaoyu Shou*, Junli Lai, Hui Wen, Jinghua Liu and Huibing Zhang

School of Information and Communication
Guilin University of Electronic Technology
No. 1, Jinji Road, Guilin 541004, P. R. China
19022304009@mails.guet.edu.cn; {huiwen; zhanghuibing; ddy2003}@guet.edu.cn
*Corresponding author: guilinshou@guet.edu.cn

Received May 2022; accepted August 2022

Abstract. *Aiming at the problem that the existing algorithm for classifying the difficulty of knowledge points fails to consider the learning patterns implicit in learners' interactive behavior, a knowledge point difficulty clustering algorithm based on multidimensional time series data and maximum frequent subgraph is proposed. The algorithm, based on the learner's multiple interactive behaviors in online learning, first constructs individual directed learning path graphs through the learner's time-series behavior, then mines the maximum frequent subgraphs in the directed learning path graph set, and finally measures the similarity of knowledge point difficulty based on student-system interaction by combining the maximum frequent subgraphs and the degree of student-system interaction. Second, a knowledge point difficulty similarity model based on interpersonal interaction is proposed to measure the similarity of knowledge point difficulty based on student-teacher interaction and student-student interaction. Finally, based on the three similarities of knowledge point difficulty, the spectral clustering algorithm is used to classify the difficulty of knowledge points. The experiments on real datasets show that the proposed algorithm has better knowledge point difficulty classification results than the existing methods.*

Keywords: Multidimensional interactive behavior, Directed learning path graph, Spectral clustering, Maximum frequent subgraph, Knowledge point difficulty

1. **Introduction.** With the combination and development of "Internet + education", online teaching has become an important teaching mode at present [1]. However, teachers and students are physically separated in an online learning environment, so teachers cannot intuitively sense students' learning states as they do in offline education and fail to understand the knowledge difficulties encountered by learners in the online learning process. Although teachers may estimate the difficulty of knowledge points based on their own teaching experience, some research has shown that it is difficult for teachers to accurately differentiate the difficulty of knowledge points for learners [2]. In the course of online learning, learners generate a large amount of interactive behavior data stored by online learning systems. [3] summarized online interactive behaviors into three types: student-system interaction, student-teacher interaction, and student-student interaction, and these behaviors reveal the process of learners' online learning [4]. Li et al. [5] analyzed learners' video interaction behaviors and found that actions like frequent pausing, skipping, re-watching, and lower playback speed indicated that the video would be difficult to learn. Brinton et al. [6] analyzed the temporal sequences of learners' video watching behaviors and extracted repetitive subsequences from them to identify repetitive viewing behaviors and found that subsequences were significantly correlated with learning effects.

DOI: 10.24507/icicel.17.02.217

[7] indicated that learners benefit from increased interaction with other learners to understand difficult questions. Therefore, we can use a data-driven mechanism to determine the difficulty level of knowledge points and then provide it to teachers to help them understand the difficulty of knowledge points for learners in a timely manner and improve teaching efficiency. There is little research work on knowledge point difficulty clustering or classification based on learners' interactive activities. Zhang et al. [8] proposed a personalized MOOC video classification method based on cluster analysis and process mining, which clusters students by their test scores, uses process mining techniques to mine the process models of each student cluster based on their learning behavior, and finally measures the difficulty and importance of MOOC videos based on the process structure. Zhang et al. [9] proposed a difficulty-based clustering method for SPOC videos, which uses the SimRank++ algorithm to calculate the difficulty similarity between two videos, and then uses a spectral clustering algorithm to achieve SPOC video difficulty clustering. The above algorithms investigate the mapping model between video difficulty and learning behavior, but they ignore the mechanism of the intrinsic association between multiple interactive behaviors of learners and the difficulty of knowledge points.

Based on the above analysis, we measure the similarity of the difficulty between any two knowledge points by considering multiple interactive behaviors of learners (student-system interaction, student-teacher interaction, and student-student interaction), and different knowledge point difficulty similarity models are proposed according to different interactive behaviors, which makes the knowledge point difficulty classification better.

The rest of the paper is organized as follows. Section 2 introduces the relevant definitions of the algorithm. Section 3 proposes a knowledge point difficulty clustering algorithm based on multidimensional time series data and maximum frequent subgraph. Section 4 provides a comparative analysis of various experimental results to evaluate the performance of the proposed algorithm. Section 5 concludes the work and looks ahead.

2. Related Definitions.

In this section, the relevant definitions and computational methods of the proposed algorithm are described, and some of the definitions are analyzed and illustrated.

Definition 2.1. ***The degree of student-system interaction.*** *It refers to the degree of engagement of the student in watching the knowledge point videos [3]. Teachers publish knowledge point videos on the online platform for students to study. Each knowledge point video contains one knowledge point in this paper. We extract features from students' behavior when they watch knowledge point videos to portray the degree of student-system interaction as follows:*

$$sc_{u,i} = \lambda_1 \times f_{sc_{u,i}} + \lambda_2 \times t_{sc_{u,i}} + \lambda_3 \times p_{sc_{u,i}} \tag{1}$$

where $f_{sc_{u,i}}$ indicates the frequency of student u studying knowledge point i in student-system interaction. $t_{sc_{u,i}}$ represents the duration of student u studying knowledge point i in student-system interaction. $p_{sc_{u,i}}$ indicates the frequency of pausing and dragging of student u studying knowledge point i in student-system interaction. According to [10], the best portrayal of the degree of student-system interaction is obtained when $(\lambda_1, \lambda_2, \lambda_3) = (1, 5, 4)$. Then the student-system interaction degree matrix $SC = [sc_{u,i}]_{m \times n}$ of students can be obtained.

Definition 2.2. ***The degree of student-teacher interaction.*** *It refers to the degree to which students communicate knowledge points with the teacher by text during the online learning process [3] and is portrayed by the frequency and duration of interaction in student-teacher interactive behaviors. The degree of student-teacher interaction is shown in Formula (2):*

$$\begin{cases} st_{u,i} = \eta_{st_{u,i}} \times f_{st_{u,i}} \\ \eta_{st_{u,i}} = \dfrac{t_{st_{u,i}}}{\max\left\{t_{st_{1,i}}, t_{st_{2,i}}, \ldots, t_{st_{m,i}}\right\}} \end{cases} \tag{2}$$

where $t_{st_{u,i}}$ indicates the duration of student-teacher interaction of student u to knowledge point i. $\eta_{st_{u,i}}$ denotes the normalized $t_{st_{u,i}}$. $f_{st_{u,i}}$ indicates the frequency of student-teacher interaction of student u to knowledge point i. The student-teacher interaction degree matrix $ST = [st_{u,i}]_{m\times n}$ of students can be obtained.

Definition 2.3. ***The degree of student-student interaction.*** *It refers to the degree of engagement between students in communicating about knowledge points through the text [3] and is portrayed by the frequency and duration of interaction in student-student interactive behaviors. The degree of student-student interaction is shown in Formula (3).*

$$\begin{cases} ss_{u,i} = \dfrac{\sum_{v=1}^{m-1} \eta_{ss_{uv,i}} \times f_{ss_{uv,i}}}{m-1} \cdots u \neq v \\ \eta_{ss_{uv,i}} = \dfrac{t_{ss_{uv,i}}}{\max\left\{t_{ss_{u1,i}}, t_{ss_{u2,i}}, \ldots, t_{ss_{um,i}}\right\}} \end{cases} \tag{3}$$

where $f_{ss_{uv,i}}$ represents the frequency of interaction between student u and student v to knowledge point i. m denotes the number of students. $t_{ss_{uv,i}}$ indicates the duration of interaction between student u and student v to knowledge point i. The degree of student-student interaction matrix $SS = [ss_{u,i}]_{m\times n}$ of students can be obtained.

Definition 2.4. ***Directed learning path graph.*** *The directed learning path graph is generated based on the time sequence $\langle v_1, v_2, \ldots, v_n\rangle$ formed by the learner watching the knowledge point video. We define the directed learning path graph of learner u as $G_u = \{V(G_u), E(G_u), L(V(G_u)), L(E(G_u)), L\}$. $V(G_u)$ is the set of knowledge nodes of graph G_u, knowledge nodes denote knowledge point videos that learner u studies online. $E(G_u) = \{(v_i, v_j)|v_i, v_j \in V(G_u)\}$ is the set of directed edges, the direction between two knowledge nodes indicates the time order in which learners watch the two knowledge point videos (v_i to v_j). $L(V(G_u))$ is the set of knowledge node labels. $L(E(G_u)) = \{w_{ij}|\forall(v_i, v_j) \in E(G_u)\}$ is the set of edge labels, w_{ij} is the number of times the learner u watches the knowledge point video from v_i to v_j, L is a function assigning labels to the vertices and the edges. The directed learning path graphs of m learners are combined to generate a graph dataset $DG = \{G_1, G_2, \ldots, G_m\}$.*

3. The Proposed Algorithm.

The block diagram of the knowledge point difficulty clustering algorithm based on multidimensional time series data and maximum frequent subgraph (MFSKPC) is shown in Figure 1.

From Figure 1, firstly, the proposed algorithm extracts the maximum frequent subgraph, SC matrix, ST matrix, and SS matrix from the three interactive behaviors. Secondly, based on the extracted data, two knowledge point difficulty similarity models are proposed to obtain three similarities of behavior-based knowledge point difficulty. Finally, the three similarities are fused to obtain the knowledge point difficulty similarity matrix, and the knowledge point difficulty is classified using a spectral clustering algorithm. So the key parts of the proposed algorithm are knowledge point difficulty similarity model based on student-system interaction, knowledge point difficulty similarity model based on interpersonal interaction, and spectral clustering based on the similarity of knowledge point difficulty. Details of the three parts are explained below.

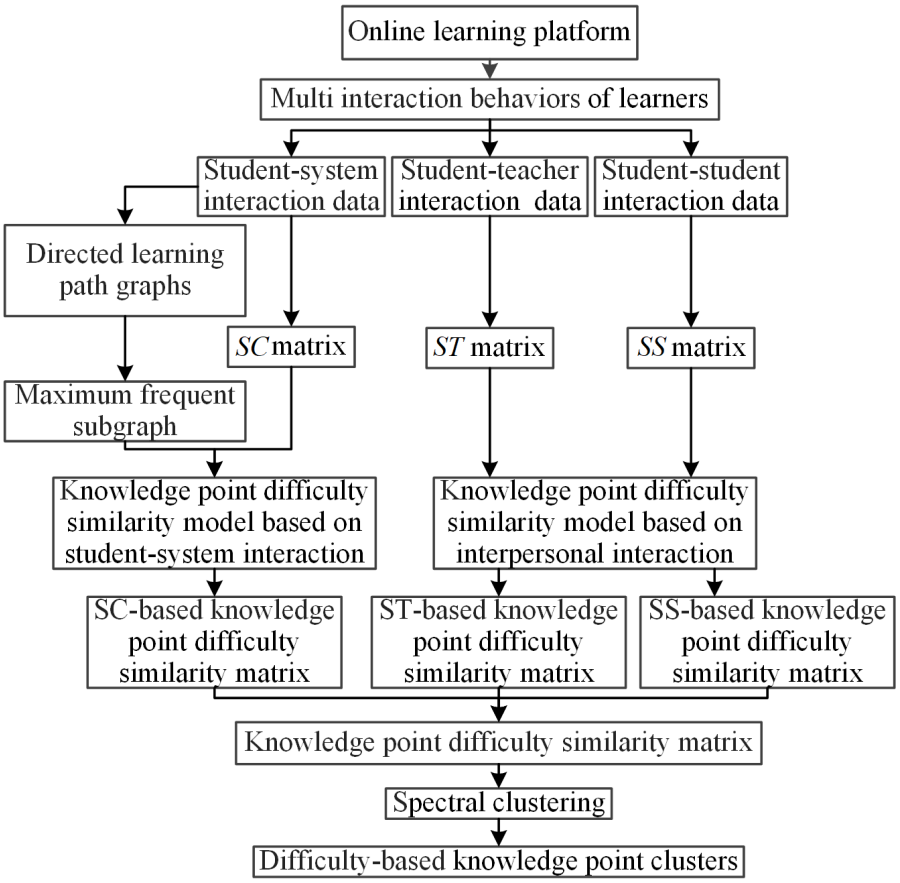

FIGURE 1. Block diagram of MFSKPC

3.1. **Knowledge point difficulty similarity model based on student-system interaction.** There are hubness and concentration problems caused by high dimensionality of the data in measuring the difficulty similarity of knowledge points only from the perspective of the degree of interaction [11]. To solve the above problems, we innovatively constructed a knowledge point difficulty similarity model by combining the degree of student-system interaction and maximum frequent subgraph. The model is shown in Formula (4).

$$Sim_{SC}^{Proposed}(i,j) = Sim_{degree}(i,j) \times Sim_{struct}(i,j) \tag{4}$$

where $Sim_{degree}(i,j)$ denotes the similarity of difficulty based on the degree of student-system interaction between knowledge point i and knowledge point j. $Sim_{struct}(i,j)$ denotes the similarity of difficulty based on maximum frequent subgraph between knowledge point i and knowledge point j.

For $Sim_{degree}(i,j)$, we choose Adjusted Cosine (ACOS) as literature shows that it can better measure the similarity of knowledge point difficulty compared to other similarity measures [11]. The calculation formula is as follows:

$$Sim_{degree}(i,j) = \frac{\sum_{u \in KP_i^{SC} \cap KP_j^{SC}} \left(sc_{u,i} - \overline{sc_u}\right) \times \left(sc_{u,j} - \overline{sc_u}\right)}{\sqrt{\sum_{u \in KP_i^{SC} \cap KP_j^{SC}} \left(sc_{u,i} - \overline{sc_u}\right)^2} \sqrt{\sum_{u \in KP_i^{SC} \cap KP_j^{SC}} \left(sc_{u,j} - \overline{sc_u}\right)^2}} \tag{5}$$

where $sc_{u,i}$ denotes the degree of student-system interaction (Definition 2.1). $\overline{sc_u} = \left(\sum_{i=1}^{n} sc_{u,i}\right)/n$, n is the number of students. KP_i^{SC} and KP_j^{SC} denote the set of students who have interacted with knowledge point i and knowledge point j, respectively.

For $Sim_{struct}(i,j)$, firstly, we build a graph dataset $DG = \{G_1, G_2, \ldots, G_m\}$ according to Definition 2.4, and the gspan algorithm [12] is used to mine the frequent subgraphs from DG. Secondly, $S = \{s_1^{\max}, s_2^{\max}, \ldots, s_n^{\max}\}$ denotes the set of maximum frequent

subgraphs which are filtered out from the frequent subgraphs. To measure more accurately the similarity of knowledge point difficulty, we obtain the value of support threshold when the number of maximum frequent subgraphs which are mined by the algorithm is max. Finally, based on the maximum frequent subgraphs, $Sim_{struct}(i,j)$ is shown in Formula (6).

$$Sim_{struct}(i,j) = \frac{1}{1+\left|\frac{\sum_{s_l^{\max}\in S(v_i)} Sup\left(s_l^{\max},DG\right)}{|S(v_i)|} - \frac{\sum_{s_l^{\max}\in S\left(v_j\right)} Sup\left(s_l^{\max},DG\right)}{|S(v_j)|}\right|} \tag{6}$$

where $S(v_i)$ denotes the set of graphs that is in S and contains knowledge node i. $S(v_j)$ denotes the set of graphs that is in S and contains knowledge node j. $Sup\left(s_l^{\max},DG\right)$ denotes the number of graphs (in DG) in which $s_l^{\max}$ is a subgraph. $|S\left(v_i\right)|$ denotes the number of graphs in the set $S(v_i)$.

3.2. **Knowledge point difficulty similarity model based on interpersonal interaction.** There is a problem of sparsity in the data of interpersonal interaction (student-teacher interaction and student-student interaction) due to various influencing factors such as learners' study habits and learning time. Traditional similarity models cannot be performed when two knowledge points do not have a co-learner [13]. To solve this problem, we propose a knowledge point difficulty similarity model based on interpersonal interaction by improving the Jaccard Mean Squared Difference (JMSD) similarity model, which is shown in Formula (7).

$$\begin{cases} Sim_{ST}^{Proposed}(i,j) = Sim_{SS}^{Proposed}(i,j) = S_1(i,j) \times S_2(i,j) \\ S_1(i,j) = \dfrac{m}{m+||KP_i|-|KP_j||} \times \dfrac{1}{2}\left(1+\dfrac{|KP_i \cap KP_j|}{|KP_i \cup KP_j|}\right) \\ S_2(i,j) = 1 - \dfrac{\left|\overline{ID_i} - \overline{ID_j}\right| + \sum_{u\in KP_i\cap KP_j}\left(ID_{u,i}-ID_{u,j}\right)^2}{|KP_i \cup KP_j|} \end{cases} \tag{7}$$

where if $|KP_i \cup KP_j| = 0$ then $Sim_{ST}^{Proposed} = Sim_{SS}^{Proposed} = 0$. Since the model applies to both ST and SS, it is next shown when the interaction is ST. $ID_{u,i}$ denotes the degree of student-teacher interaction of learner u to knowledge point i, namely $ID_{u,i} = st_{u,i}$. KP_i and KP_j denote the set of learners who have interacted with knowledge point i and knowledge point j in ST matrix, respectively. $\overline{ID_i} = \overline{st_i} = \left(\sum_{u=1}^{r} st_{u,i}\right)/r$, $r = |KP_i|$. $|KP_i|$ denotes the number of learners in the set KP_i. $KPD_i = |KP_i|/m$, m denotes the number of learners. $KP_i \cap KP_j$ denotes the intersection of KP_i and KP_j. $KP_i \cup KP_j$ denotes the union of KP_i and KP_j. SS is a similar way with ST. The traditional JMSD uses only the co-learner interaction data to measure the difficulty similarity. The proposed similarity model makes full use of the interaction data of the two knowledge points.

3.3. **Spectral clustering based on the similarity of knowledge point difficulty.** The similarity of difficulty between two knowledge points is obtained by weighing the combination of $Sim_{SC}^{Proposed}(i,j)$, $Sim_{ST}^{Proposed}(i,j)$ and $Sim_{SS}^{Proposed}(i,j)$. The calculation process is shown in Formula (8).

$$\begin{cases} Sim_{DKP}^{Proposed}(i,j) = \alpha_1 \times Sim_{SC}^{Proposed}(i,j) + \alpha_2 \times Sim_{ST}^{Proposed}(i,j) \\ \qquad\qquad + \alpha_3 \times Sim_{SS}^{Proposed}(i,j) \\ \alpha_1 + \alpha_2 + \alpha_3 = 1 \end{cases} \tag{8}$$

where α_1, α_2 and α_3 are weighting factors. In this paper, by testing all combinations of the three non-negative variables with a minimum step size of 0.01, we conclude that the clustering precision was the highest when the value was $(\alpha_1, \alpha_2, \alpha_3) = (0.8, 0.1, 0.1)$.

Based on the similarity of knowledge point difficulty $Sim_{DKP}(i,j)$, the spectral clustering algorithm is used to classify the difficulty of knowledge points. Teachers can set the number of clusters (K) based on teaching experience. The specific process is as follows: we construct the knowledge point difficulty similarity matrix $M_{i,j} = Sim_{DKP}^{Proposed}(i,j)$, and then the Laplacian matrix L of M is calculated as $L = D^{-1}(D-M)D^{-1}$. D is the diagonal matrix, $D_{ii} = \sum_{j=1}^{N} Sim_{DKP}^{Proposed}(i,j)$. Then, the eigenvector of the Laplacian matrix L is calculated, and the eigenvectors corresponding to the first K minimum eigenvalues are extracted, forming F of $N \times K$ dimension. The K-means algorithm is used to cluster the feature subspace F with the number of K.

4. **Experimental Evaluation.** To verify the effectiveness of the proposed algorithm (named MFSKPC), we use an external assessment approach to evaluate the clustering results. The specific process is as follows: based on the learners' test data, we analyze the average score of each knowledge point to determine the true difficulty of the knowledge point and compare it with the difficulty of the knowledge point given by the algorithm. We use the precision clustering indicator to measure the accuracy rate of the algorithm [8].

4.1. **Dataset of the experiment.** Use the interactive behavior data of 2019 students studying courses (Data Structure and Algorithm) which is a compulsory course for sophomores at a university as the data source. The dataset consists of 207 knowledge points and 272 learners' multi-dimensional interactive behavior data, including 50,544 student-system interaction data, 7,683 student-teacher interaction data, 1,252 student-student interaction data, and knowledge points test data. The experimental platform is Python 3.6.

4.2. **Experimental results and analysis.**

4.2.1. *Algorithm classification experimental results and analysis.* To evaluate the effectiveness of the proposed algorithm (MFSKPC), we compared MFSKPC with two commonly used classical methods for knowledge point difficulty classification. The first method is MS, which sorts the knowledge points in ascending order according to their interaction degree ikp_i ($ikp_i = \sum_{u=1}^{m} sc_{u,i} + st_{u,i} + ss_{u,i}$) and divides them into K groups on average. The group with the minimum average interaction degree is the easiest knowledge point set, and the group with the maximum average interaction degree is the most difficult knowledge point set. The second method is MC, which uses the K-means clustering algorithm to cluster knowledge points based on the ikp_i. The cluster of the maximum average interaction degree is the most difficult knowledge point set, and the cluster of the minimum average interaction degree is the easiest knowledge point set.

In this experiment, knowledge points were clustered with MFSKPC, MS and MC at $K = 2$, 3, 5, respectively, and the clustering results were evaluated to obtain precision values. The experimental results are shown in Figure 2.

From Figure 2, it can be seen that the clustering precision of the proposed algorithm MFSKPC is higher than that of MS and MC at different K values. When $K = 3$, the precision of MFSKPC is the highest, which is consistent with the actual teaching experience of teachers. The analysis of the experimental data showed that some easy knowledge points have a higher average interaction degree than the difficult knowledge points, so MS cannot distinguish the knowledge point difficulty by only relying on the average interaction degree of knowledge points. MC is easily affected by the individual knowledge points with higher or lower interaction degrees, which leads to inaccurate classification results. Therefore, the MFSKPC algorithm has a better effect of classifying the difficulty of knowledge points.

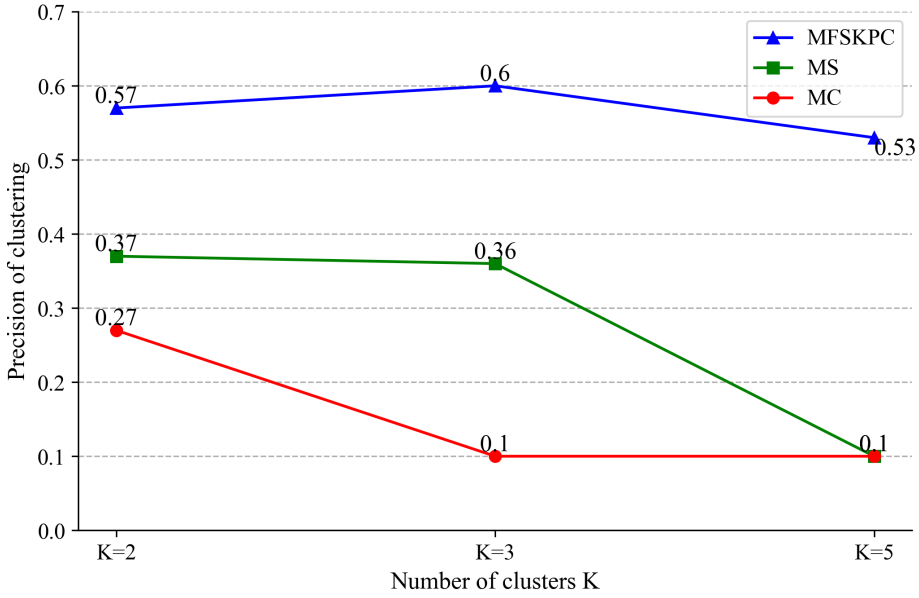

FIGURE 2. Precision of MFSKPC, MS, and MC

4.2.2. *Similarity comparison.* The proposed algorithm MFSKPC mainly measures the similarity of difficulty between any two knowledge points from three similarity models $\left(Sim_{SC}^{Proposed}(i,j),\ Sim_{ST}^{Proposed}(i,j),\ Sim_{SS}^{Proposed}(i,j)\right)$, and then performs spectral clustering to achieve knowledge point difficulty classification. To verify the effectiveness of the proposed algorithm for calculating the similarity of knowledge point difficulty, other similarity methods are used to calculate the similarity of knowledge point difficulty. For example, the calculation is performed using Adjusted Cosine similarity. Considering that the traditional similarity model cannot be measured from the perspective of the maximum frequent subgraph, we set $Sim_{SC}^{ACOS}(i,j) = Sim_{degree}^{ACOS}(i,j)$. The procedure is as follows:

$$\begin{aligned} & Sim_{DKP}^{ACOS}(i,j) \\ &= \alpha_1 \times Sim_{SC}^{ACOS}(i,j) + \alpha_2 \times Sim_{ST}^{ACOS}(i,j) + \alpha_3 \times Sim_{SS}^{ACOS}(i,j) \\ &= \alpha_1 \times Sim_{degree}^{ACOS}(i,j) + \alpha_2 \times Sim_{ST}^{ACOS}(i,j) + \alpha_3 \times Sim_{SS}^{ACOS}(i,j) \end{aligned} \tag{9}$$

where the values of $(\alpha_1, \alpha_2, \alpha_3)$ are the same as MFSKPC. Similarly, MFSKPC is compared with SimRank++ [8], RJMSD [13], JMSD, PCC [14] and ACOS similarity methods in clustering precision. The clustering results are shown in Figure 3.

As shown in Figure 3, under the conditions of different K values, the clustering precision of MFSKPC is better than other similarity models. For the traditional similarity models, JMSD and ACOS have better clustering precision, and PCC is worse. SimRank++ has better clustering precision compared with RJMSD. PCC and RJMSD are generally applicable to student similarity calculations, so they perform worse in measuring the similarity of knowledge point difficulty. Moreover, SimRank++ and JMSD are based on the structural perspective to measure the similarity of knowledge point difficulty, which can have better clustering precision, but there is only co-interaction data that the two algorithms consider. Therefore, the proposed similarity model can more precisely measure the similarity of knowledge point difficulty by making full use of the knowledge point interaction data and considering maximum frequent subgraph and interaction degree.

5. **Conclusions.** To help teachers understand the difficulty of knowledge points for learners in a timely and objective manner, we propose a knowledge point difficulty clustering algorithm based on multidimensional time series data and maximum frequent subgraph. The algorithm quantifies the relationship between three interactive behaviors of learners

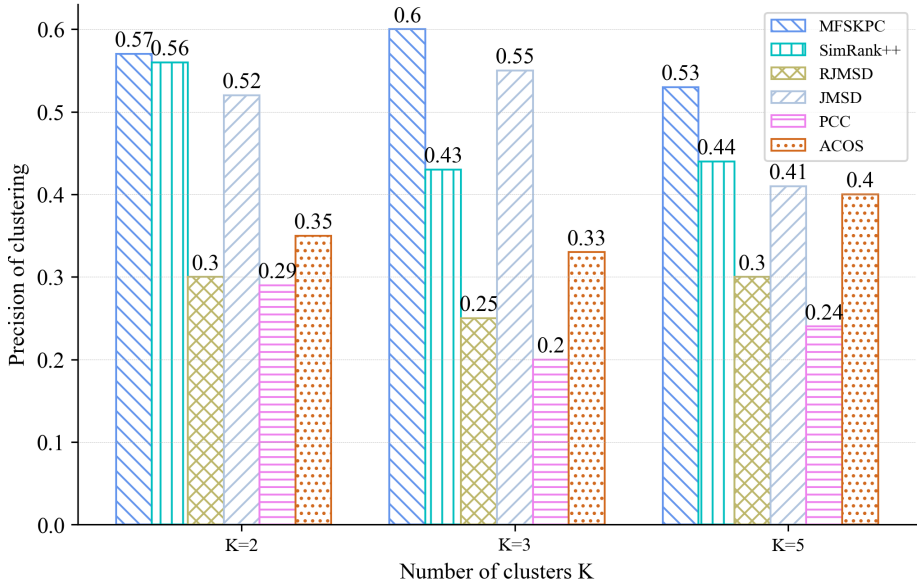

FIGURE 3. Comparison of clustering precision of MFSKPC with other similarity methods

and knowledge point difficulty, and improves the knowledge point difficulty similarity model to achieve more accurate knowledge point difficulty classification. Our future work will focus on natural language processing techniques and improving the precision of the difficulty classification of knowledge points by combining them with learners' text sentiment.

Acknowledgment. This work is partially supported by the National Natural Science Foundation of China (62177012, 61967005), Innovation Project of GUET Graduate Education (2020YCXS022, 2021YCXS033), the Key Laboratory of Cognitive Radio and Information Processing Ministry of Education (CRKL190107). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] G. I. Butnaru, V. Niţă, A. Anichiti and G. Brînză, The effectiveness of online education during COVID-19 pandemic – A comparative analysis between the perceptions of academic students and high school students from Romania, *Sustainability*, vol.13, no.9, 5311, DOI: 10.3390/su13095311, 2021.

[2] E. Verdú, M. J. Verdú, L. M. Regueras, J. P. de Castro and R. García, A genetic fuzzy expert system for automatic question classification in a competitive learning environment, *Expert Systems with Applications*, vol.39, no.8, pp.7471-7478, 2012.

[3] Z. Shou, Y. Wen, P. Chen and H. Zhang, Personalized knowledge map recommendations based on interactive behavior preferences, *International Journal of Performability Engineering*, vol.17, no.1, pp.36-49, 2021.

[4] M. Zuo, Y. Xu, Z. Wang, R. Zhao and X. Li, Analysis of forum interaction behavior based on cloud class, *International Conference on Technology in Education*, vol.1048, pp.131-145, 2019.

[5] N. Li, A. Kidziński, P. Jermann and P. Dillenbourg, MOOC video interaction patterns: What do they tell us?, *Design for Teaching and Learning in a Networked World*, vol.9307, pp.197-210, 2015.

[6] C. G. Brinton, S. Buccapatnam, M. Chiang and H. V. Poor, Mining MOOC clickstreams: Video-watching behavior vs. in-video quiz performance, *IEEE T. Signal Proces*, vol.64, no.14, pp.3677-3692, 2016.

[7] T. Oktavia and S. Sujarwo, A meta-learning recommender system framework for identifying learning partner, *ICIC Express Letters*, vol.15, no.2, pp.117-124, 2021.

[8] F. Zhang, D. Liu and C. Liu, Difficulty-based SPOC video clustering using video-watching data, *IEICE Trans. Information and Systems*, vol.104, no.3, pp.430-440, 2021.

[9] F. Zhang, D. Liu and C. Liu, MOOC video personalized classification based on cluster analysis and process mining, *Sustainability*, vol.12, no.7, 3066, DOI: 10.3390/su12073066, 2020.

[10] H. Zhu, Y. Liu, F. Tian, Y. Ni, K. Wu, Y. Chen and Q. Zheng, A cross-curriculum video recommendation algorithm based on a video-associated knowledge map, *IEEE Access*, vol.6, pp.57562-57571, 2018.

[11] D. Wang, Y. Yih and M. Ventresca, Improving neighbor-based collaborative filtering by using a hybrid similarity measurement, *Expert Systems with Applications*, vol.160, 113651, DOI: 10.1016/j.eswa.2020.113651, 2020.

[12] X. Yan and J. Han, gSpan: Graph-based substructure pattern mining, *2002 IEEE International Conference on Data Mining*, pp.721-724, 2002.

[13] S. Bag, S. K. Kumar and M. K. Tiwari, An efficient recommendation generation using relevant Jaccard similarity, *Information Sciences*, vol.483, pp.53-64, 2019.

[14] Y. Shi, M. Larson and A. Hanjalic, Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges, *ACM Computing Surveys (CSUR)*, vol.47, no.1, pp.1-45, DOI: 10.1145/2556270, 2014.