Personalized knowledge point recommendation system based on course knowledge graph

Yakun Lang\textsuperscript{1*}, Guo zhong Wang\textsuperscript{1}

\textsuperscript{1}School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, 201620, China

\textsuperscript{*}Corresponding author’s e-mail: 1253193643@qq.com

Abstract. Existing class education lacks analysis of students' learning data, which can not reasonably evaluate students' mastery of knowledge points and locate students' learning path in real time, nor effectively recommend the required knowledge points for students. In order to solve the above problems, a personalized knowledge point recommendation system model (KG-PKP) combined with the knowledge graph of the course is proposed. By using accuracy, answer-time and answer-types in answer records, evaluation equations which to judge students' mastery of knowledge points can be constructed. We can extract the knowledge points in answer records and map them into the course knowledge graph. Personalized knowledge points for students can be recommended by using the semantic hierarchical relationship and sequence of knowledge points in the knowledge graph. The comparative experiment proves the validity and interpretability of the model.

1. Introduction
After multimedia and digital education, education has now entered the stage of artificial intelligence education. The recommendation system [1], knowledge graph [2] and data mining [3] are used in the field of education. The personalized knowledge point recommendation for students is an important research direction in the field of intelligent education.

Many scholars compare the recommendation of knowledge points in the field of education to the recommendation of products in e-commerce. The student’s answer score in class is similar to the user’s rating of the product. Collaborative filtering method (CF) [4,5] is the most widely used in the recommendation system. It calculates the similarity between the knowledge points in the student answer records and the knowledge points in the set to be recommended, recommended similar knowledge points to students, or finds similar students based on the students' answer records, and recommend the knowledge points learned by similar students to the students. Probability matrix factorization (PMF) [6] is also a commonly used method in recommendation systems. It can decompose the matrix of students’ answer scores in class, predict the students’ scores and generate recommendations. The DINA model[7,8] in educational data mining can also be used for recommendation of knowledge points, it can model the learning ability of each learner, diagnoses the learner's mastery of knowledge points and recommends the relevant knowledge points.

In class learning, the knowledge points should be ordered to learn. The recommended knowledge points for students should meet the syllabus and the needs of students. Therefore, Students who have not mastered the current knowledge points should recommend similar knowledge points to learn, and students who have mastered the knowledge points should recommend the next knowledge point to
learn. The content and length of the recommended knowledge points should be consistent with the current knowledge level of the learners. However, these mainstream methods can only find similar knowledge points, lack of research on the learning sequence of students [9,10], obviously cannot meet the recommendation requirements of students' personalized knowledge points.

This paper proposes a personalized knowledge point recommendation system model (KG-PKP) based on the course knowledge graph. The main innovations of KG-PKP are as follows:

1. By extracting the sequence of students’ answers for each class, we can mine the hierarchical and sequential relationship between the knowledge points.

2. Construct a course knowledge graph and embed all knowledge points in polar coordinates. The knowledge points of distance can be quantified by using module length and angle of the polar coordinate.

3. Use the student's answer records to construct the evaluation equation. The knowledge points can be trained with the CBOW[11] and can be accurately positioned to the knowledge point polar coordinates. Use the hierarchy and sequence relationship of the knowledge points to make recommendations.

2. The framework of Knowledge point recommendation system

This paper uses the class education system developed by the team-I Class to obtain the answer records of all students in the course of the "Principle of Automatic Control". Using the hierarchical structure of the knowledge points and the characteristics of answering records to design the personalized knowledge point recommendation model. The experiment proves that the student's answer record can not only locate the student's learning sequence, but also recommend knowledge points that are in line with his learning level. The flow chart of personalized knowledge point recommendation system based on the course knowledge graph is shown in Figure 1 and the model is shown in Figure 2.

![Flow chart of personalized knowledge point recommendation system based on course knowledge graph](image-url)
2.1. Construct knowledge graph

The knowledge graph can be understood as a multi-relationship graph. The nodes in the graph represent entities, and the edges represent relationships. The graph usually expressed in the form of triples as <head entity, relationship, tail entity>. Semantic layering is common in the knowledge graph. Taking a triple <Beijing, located, China> as an example, the semantic of entity <China> is more abstract, so it belongs to a higher-level entity in the semantic hierarchy, and semantic of entity <Beijing> is more specific, so it belongs to a lower-level entity in the semantic hierarchy. Course knowledge points also match this hierarchical relationship, such as <open-loop transfer function, belongs, transfer function>. In addition, the knowledge points should be learned in order. For example, the transfer function can be learned only on the basis of learning the differential equations. It is represented by a triple as <differential equation, advanced, transfer function>.

Constructing a course knowledge graph requires a strong professional background. The use of word-cutting tools will cause a large deviation in knowledge points. In order to make the knowledge points more accurate. The teacher divides the knowledge combined the learning order and hierarchical relationship of knowledge points to construct knowledge graph.

2.2. Embed knowledge points into polar coordinates

As shown in Figure 2-d, the hierarchy of knowledge points can be regarded as a binary tree. The depth of each node in the tree corresponds to the semantic level of knowledge points, so we can divide the knowledge points in the knowledge graph into two class.

(1) Knowledge points belonging to different semantic levels, such as <transfer function> and <open-loop transfer function>.

(2) Knowledge points that belong to the same semantic level, such as <open-loop transfer function> and <closed-loop transfer function>.

In order to model these two types of entities at the same time, Zhang[12] believes that the polar coordinates can be used to represent entities at different Semantic hierarchy. So we put the word vectors of knowledge points into the polar coordinate system. Any knowledge points in the coordinate can be uniquely determined by the module length and angle. The module length can be used to
distinguish knowledge points of different levels, use angles to distinguish different knowledge points at the same level.

For any triple \((h, r, t)\), use \(h_m\) and \(t_m\) to represent the modulus length of knowledge points, \(r_m\) to represent the relationship between the modulus lengths, use \(h_p\) and \(t_p\) to represent the phase of knowledge points, and \(r_p\) to represent the relation between the phase.

Their relationship is shown in equation (1).

\[
\begin{cases} 
  h_m \times t_m = r_m, \text{where } h_m, t_m \in R^k, r_m \in R^k \\
  (h_p + r_p) \mod 2\pi = t_p, \text{where } h_p, r_p, t_p \in [0,2\pi]^k 
\end{cases}
\]

The \(\circ\) represents the Hadamard product of \(h_m\) and \(t_m\). \(R^k\) and \(R^k\) are used to distinguish between positive and negative samples. If there is a connection between knowledge points, it is defined as a positive sample, otherwise it is a negative sample. The purpose is to minimize the distance of positive samples and maximize the distance of negative samples. Therefore, the distance formula defined by the modulus \(d_{r,m}\) and the phase \(d_{r,p}\) is shown in equation (2):

\[
\begin{align*}
  d_{r,m}(h_m, t_m) &= \|h_m \times t_m - r_m\|_2, \text{where } h_m, t_m \in R^k, r_m \in R^k \\
  d_{r,p}(h_p, t_p) &= \|\sin ((h_p + r_p - t_p)/2)\|_1, \text{where } h_p, r_p, t_p \in [0,2\pi]^k
\end{align*}
\]

For any knowledge points in polar coordinates, the joint distance function can be defined as equation(3):

\[
d_{r}(h, t) = d_{r,m}(h_m, t_m) + \lambda d_{r,p}(h_p, t_p)
\]

\(\lambda\) is a parameter in model training, \(\lambda \in R\).

Using negative sampling algorithm to train the model, the loss function [13] can be defined as in equation (4):

\[
L = -\log \sigma (\gamma - d_{r}(h, t)) - \sum_{i=1}^{n} p(h_i', r, t_i') \log \sigma (d_{r}(h_i', t_i') - \gamma)
\]

The \(\sigma\) is the sigmoid function, and the \(\gamma\) is the adjustment variable to prevent the model from overfitting. \((h_i', r, t_i')\) represents the \(i\)th negative sample. The knowledge points in polar coordinates need to match the following conditions. If the semantic level of the head knowledge point entity is higher than the tail knowledge point entity, then \(r_m = \frac{r_m}{h_m} > 1\), if the semantic level of the head knowledge point entity is lower than the tail knowledge point entity semantic level, then \(r_m = \frac{r_m}{h_m} < 1\), and if the head and tail knowledge point entities are at the same semantic level, then \(r_m = \frac{r_m}{h_m} = 1\).

2.3. Extract the answer sequence

The Alibaba [14] proves that items purchased by users within a fixed time interval can be divided into a sequence of items to ensure the relevance of items. In class scenarios, the students’ answer records for each lesson are also relevant. For example, the knowledge point of series correction and parallel correction in the course of "Principle of Automatic Control" are all correction methods, so we can extract and code the students’ answer records in each lesson as the answer sequence, as shown in Figure 2-a.

2.4. CBOW

The CBOW contains a three-layer network structure, named the input layer, the projection layer and the output layer. The optimization objective function is equation (5):

\[
L = \sum_{w \in C} logp(w|\text{Context}(w))
\]
$C$ represents a corpus of course knowledge points, $w$ represents any knowledge points in $C$, and its learning goal is to maximize the function.

The input layer is a word vector containing $2c$ words in $\text{Context}(w)$. $2c$ is the number of context word vectors. The projection layer is to accumulate $2c$ word vectors of the input layer. The output layer is a binary tree, which uses the words that have appeared in the corpus as leaf nodes, and constructs the Huffman tree based on the weight of the number of times each word appears in the corpus. For any words $w$ in the corpus, there is a unique path from the root node to the word $w$ in the Huffman tree. The branches on the path can be regarded as a process of binary classification. Each classification can generate a probability, so we can obtain the conditional probability is $p(w | \text{Context}(w))$.

Use the knowledge points in the course knowledge graph and the relationship between the knowledge points to train the CBOW. The trained model is used for the knowledge points of class answer records, we can use the knowledge points in the answer records to predict the knowledge points that are connected to them.

2.5. Construct evaluation equation

In class learning, judging the students' mastery of knowledge points depends on the correct rate of answering questions, the difficulty of the questions, the time of the students answer the questions. Taking the course "Principle of Automatic Control" as an example, constructed evaluation equation is shown in equation (6):

$$S = \frac{1}{n} \sum_{i=0}^{n} (ax_i + a(-bx_i + c(dx_i + ex_i + fx_i)))$$

The parameters in the formula are shown in Table 1.

| parameter | Value |
|-----------|-------|
| a         | The correct answer is 1, otherwise it is 0 |
| b         | According to the time interval for answering the questions, with the values of 0.1, 0.3 and 0.5 |
| c         | The difficulty coefficients of questions, with the values of 0.3, 0.5 and 0.7 |
| e         | The answer type is multiple choice, the value is 0.3 |
| f         | The type of answer is blank, the value is 0.5 |
| g         | The answer type is short answer, the value is 0.7 |

In order to effectively assess the mastery of students' knowledge points, a threshold $T$ needs to be defined. After the evaluation of professional teachers and the verification of the experiment, the value of $T$ is 0.65. When $S \leq 0.65$, that the student has not mastered the knowledge Point, otherwise, it is considered that students have mastered this knowledge point.

2.6. Recommended strategy

The knowledge points recommended for students should be in line with the students' current knowledge level. The recommendation strategy can be divided into two methods.

1. If the student has not mastered the knowledge point, that is $S \leq 0.65$, we needs to recommend the similar knowledge points to continue learning. In this case, the ratio $r_m \leq 1$, corresponding to the knowledge graph $<$head knowledge point, include, ? $>$, find the tail knowledge points in the knowledge graph and make recommendations.

2. If the student has mastered this knowledge point, that is $S > 0.65$, the system needs to recommend the next knowledge point for the student according to the learning sequence of the knowledge outline. The ratio $r_m \geq 1$, corresponding to the knowledge graph $<$?, include, tail knowledge points$>$, find the head
knowledge points in the knowledge graph and make recommendations.

3. experiment

3.1. Experimental dataSet and evaluation index

This paper conducted an experiment on the iclassDataSet, as shown in Table 2. The iclassDataSet is the real record of the answers in the course of "Principle of Automatic Control" of Shanghai University of Engineering Science. The dataSet contains 54 students, 59 knowledge points, and 12,539 answer records. The course is 32 hours and each lesson is 90 minutes long.

| project                      | Quantity |
|------------------------------|----------|
| Total hours                  | 32       |
| Number of students           | 54       |
| Number of knowledge points   | 59       |
| Answer records               | 12539    |

There are three types of answer records in the class, which are multiple choice questions, short answer questions, and blank questions. The specific numbers and values are shown in Table 3.

Table 3. Types of answers

| Type          | Time interval | Quantity |
|---------------|---------------|----------|
| Multiple choice | [0,3]min      | 8526     |
| blank         | [0,5]min      | 2101     |
| Short answer  | [0,7]min      | 1912     |

Precision, Recall and F1 are used to evaluate the effectiveness of KG-PKP, the specific definition is shown in Equation (7).

\[
\text{precision} = \frac{TP}{TP + FP}, \quad \text{recall} = \frac{TP}{TP + FN}, \quad F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (7)
\]

Precision is expressed as the proportion of knowledge points that match the needs of students in the total number of recommended knowledge points. Recall is all the knowledge points that match the students, and these knowledge points are found. F1 is the mean of precision and recall.

3.2. Comparative Experiment

In order to verify the model of KG-PKP, We selected DINA, CF and PMF methods as comparative experiments. The results are shown in Table 4.

Table 4. Algorithm comparison

| algorithm | precision | recall | F1  |
|-----------|-----------|--------|-----|
| DINA      | 0.3       | 0.28   | 0.29|
| CF        | 0.332     | 0.325  | 0.323|
| PMF       | 0.402     | 0.313  | 0.352|
| KG-PKP    | 0.683     | 0.627  | 0.65|

From the comparison of the data in the table, it can be seen that the DINA method, CF method and PMF method perform poorly. This does not mean that their own algorithms are flawed, but this kind of algorithm is not suitable for class learning, these algorithms are recommended based on the similarity of knowledge points. The DINA algorithm can make recommendations for specific knowledge points, and often faces data loss. The CF algorithm ignores the learning order of knowledge points and the degree of students' mastery of knowledge points. The PMF algorithm can
decompose the matrix of the questions answered, but there is no analysis of the degree of mastery of the students' knowledge points.

In class learning scenario, the knowledge points in the learning course should be orderly and in line with the current learning level of the students. The KG-PKP model proposed in this paper considers the similarity and connection order of knowledge points at the same time. So the precision, recall and F1 of the KG-PKP model have been significantly improved. The recommendation results are more applicable to students' actual learning.

4. Conclusion
The KG-PKP model proposed in the paper is compared with the DINA model, CF model, and PMF model on the iclassDataSet. The experiment proves that the KG-PKP model can be better applied to the students' personalized learning scenarios. when the teachers guide their students with no excess energy and the students are constrained by their own abilities, the KG-PKP model can help to improve students’ learning methods.

Acknowledgments
We thank Shanghai University of Engineering and Science Hai Zhu, Chao Su, Jinkai Yang and Xiangsheng Sun for their assistance with the anonymized data.

References
[1] Sarwar B. Item-Based Collaborative Filtering Recommendation Algorithms[C]//Proc International World Wide Web Conference. 2001.
[2] Singhal A., Official Google Blog: Introducing the Knowledge Graph: things, not strings. Official Google Blog (2012) 1-8.
[3] Hope T, Chan J, Kittur A, Shahaf D. Acm. Accelerating Innovation Through Analogy Mining. New York: Assoc Computing Machinery; 2017. 235-43 p.
[4] Linden G, Smith B, York J. Amazon. com recommendations: Item-to-item collaborative filtering[J]. IEEE Internet Computing, 2003, 7(1):76-80.
[5] Chen C M, Wang C J, Tsai M F, et al. Collaborative Similarity Embedding for Recommender Systems[J]. 2019.
[6] Salakhutdinov R, Mnih A. Probabilistic matrix factorization[J]. Advances in neural information processing systems, 2008:1257-1264.
[7] Fernandes E, Holanda M, Victorino M, et al. Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil[J]. Journal of Business Research, 2019, 94(JAN.):335-343.
[8] Ahuja R, Jha A, Maurya R, et al. Analysis of Educational Data Mining[J]. 2019.
[9] Zhu H , Tian F , Wu K , et al. A multi-constraint learning path recommendation algorithm based on knowledge map[J]. Knowledge Based Systems, 2018, 143(MAR.1):102-114.
[10] Dwivedi P , Kant V, Bharadwaj K K . Learning path recommendation based on modified variable length genetic algorithm[J]. Education and information technologies, 2018, 23(2):819-836.
[11] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[J]. Computer science, 2013.
[12] Zhang Z, Cai J, Zhang Y, et al. Learning Hierarchy-Aware Knowledge Graph Embeddings for Link Prediction[J]. 2019.
[13] Sun Z , Deng Z H , Nie J Y , et al. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space[J]. 2019.
[14] Wang J, Huang P, Zhao H, et al. Billion-scale Commodity Embedding for E-commerce Recommendation in Alibaba[J]. 2018.