Research and Implementation of Basic Education Resource Retrieval Strategy Based on the Relationship among Knowledge Points

Zhen Yang
National Engineering Research Center for E-Learning of Central China Normal University, Wuhan, China

Abstract. In order to solve the lack of knowledge relevance and topic drift in traditional keyword search, this paper proposed knowledge extraction strategy based knowledge point association and recognition of user intention based user intention theme map and the mapping from resource to knowledge points strategy. It realized the knowledge systematic and relevant of the search results and improved the retrieval accuracy.

1. INTRODUCTION

In order to solve insufficient or unreasonable knowledge relevance and inaccurate recognition of user intention for traditional keyword search or semantic search, which results in problems of incoherent learning process, the lack of learning content and topic drift in basic education, this study improves the traditional semantic analysis methods and proposes the theme map of user intention as the carrier and query expansion of the knowledge point association, which provides the conditions for the semantic recognition of the retrieved content.

2. Significance and content of the study

Based on the existing technologies of semantic retrieval system for educational resources, this paper focuses on improving the calculation of the correlation of knowledge points and the expansion of query knowledge points. On the basis of Knowledge Association Library and corpus, the knowledge point correlation computation, resource matching and retrieval result processing are implemented to achieve the knowledge systematisms and relevance of retrieval results. The accuracy of query expansion is improved by recognizing the user intention, which ensures the retrieval accuracy and the degree of personalization.

The main work of this paper is listed as follows:

1) To deal with the existing problems in semantic similarity calculation, this paper improves the word source vector method which combines large-scale corpus and local knowledge base in 1. The semantic similarity calculated from the corpus and the structural similarity obtained from the knowledge base are multiplied by a certain proportion to obtain the final similarity, which can improve the accuracy of the calculation results compared with the results in 2.

2) As for the expansion of query knowledge point, this paper recommends to recognize the user intention. The main purpose is to expand the query content knowledge points through the topic map of user intention to make the expansion knowledge points in accordance with the real user intention and to avoid the topic drift and ambiguity. The construction of the topic map is based on the comprehensive analysis of the current learning behaviors. The historical behavior knowledge points of users are treated with the forgetting curve of Hermann Ebbinghaus. Besides, the accuracy of user intention is optimized and the accuracy of system is improved. In the extraction of knowledge point association relation, this paper proposes knowledge topic map as the carrier of knowledge structure, which is used to expand the initial query of knowledge points and improve the recall rate of the system.

3) As for the problem of knowledge point matching of the retrieval resources, the resource matching in this paper is based on the local knowledge point association relation database and the resource and knowledge point association database, which realizes the mapping from resource to knowledge point and improves the retrieval efficiency.

3. Research ideas and related algorithm

The overall idea and process presented in this article are shown in figure 1:
3.1 Knowledge point extraction of retrieved content

The process of extracting knowledge points based on the semantic recognition has been presented as hereunder:

Firstly, both the word segmentation and semantic analysis are required for the extraction of query content knowledge points. The retrieval content segmentation of this paper adopts the ICTCLAS system of Chinese Academy of Sciences and domain dictionary for word segmentation Algorithm design.

The process of semantic analysis is as follows:

List<String> inputList = skNowkeys;//Search content keywords
List<String> knowList = value; // Take out the keywords of each knowledge point in the knowledge base
Word2VEC w = new Word2VEC();
w.loadJavaModel("library/dm20180128.model") ;//
Trained model
Model model = new Model();
similar=model.run(inputList, knowList,w);
similarDegree = new BigDecimal(similar);

3.2 User Intention Recognition

Prior to the retrieval of the resource, it is deemed essential to filter out unrelated preliminary retrieval knowledge points with a topic map that corresponds to the user's behavior feature. The knowledge point structure and associated information in the topic map brings a convenient condition for the semantic recognition of retrieval content.

3.2.1 User Intention Theme Graph Modeling Based on User's Historical Behavior Knowledge Points

The user intention recognition in this paper is based on the analysis of user behavior in the system. The user behavior information includes the historical retrieval information, current learning state, analysis of knowledge points of examination questions, historical information tagging of resources, historical information of learning resources, etc.

3.2.1.1 Construction And Optimization of User Intention Theme Graph Model Based on Forgetting Function

There are numerous factors requiring consideration in the design of user behavior feature model, primarily including the capability of fully reflecting the user's learning situation, and coordinating with the knowledge representation of the curriculum in the system.

Knowledge topic map is regarded as an abstract description of the knowledge tuples in the tree structure of knowledge system. A knowledge point is likely to correspond to multiple topic maps. A topic map is also likely to contain multiple knowledge points (knowledge elements and knowledge tuple both). The user's expertise of the knowledge system in the learning mechanism can also be described by a knowledge topic map.

This paper employs a time-based forgetting function in a bid to update the instantaneous validity of the knowledge points in the optimized user's intent topic graph, in addition to adjusting the correlation weight of different time and repeated knowledge points for the improvement of the reliability of the model, so that the retrieval result could be made more precise.

This paper employs the forgetting function for simulating the attenuating degree of the user's knowledge point of knowledge[3], which has been presented in the formula (1):

\[ F(t) = e^{-\frac{\ln 2}{DT + \Delta t}} \] (1)

In the formula, \( F(t) \) is a forgetting factor, which indicates that the weight of knowledge points related to the history of user behavior attenuates to the original proportion; \( t \) is the current time; \( start \) is the modeling time. The speed of the weight attenuation of the knowledge point is determined and regulated by the parameter \( DT \).

On the reappearance of the knowledge point \( C \), the forgetting half-life \( DT \) of the knowledge point requires appropriate extension. The increment of the half-value period has been presented in the Formula (2):

\[ U_i = \{ K, G, R, T \} \] (2)

The values \( \{ K, G, R, T \} \) in the model can be extracted directly, attaining the relevant knowledge of the user behavior history, user intention topic map and its weight in the respective period of time;

By applying the updated user intent map to the retrieval system to adjust the user's attention of the relevant knowledge points in the retrieval intent topic
map so that the precision of the final retrieval result could be ensured. The user concern adjusts the theme graph of the user's intent that involves using the forgetting function for processing the knowledge element contained in each struct in the Figure 6, suggesting the knowledge tuple. Add the user's attention to all the knowledge elements in each struct, implying the user's attention to the knowledge tuple; accumulate the user concerns of all knowledge elements in each of the knowledge tuples in each theme, which is the user's attention to the topic graph.

The calculation process is designed as shown in the following formula (3)

$$
\text{know\_weight} = \frac{\ln(1+t_{\text{start}})}{DT} + \text{length2} \times e^{-\frac{\ln(1+t_{\text{start}})}{DF+\Delta t}}
$$

(3)

Among them, the learning time $t_{\text{start}}$ is regarded as the first time for the user to learn the knowledge point, and the learning time $t_{\text{end}}$ is the first time for the user to learn the knowledge point. $\Delta t = (t_{\text{end}} - t_{\text{start}}) \times k$, $k$ represents the influence factor, and the $k$ value is employed for controlling the impact of time on the half-life. In this paper, the value of $k$ is correlated with the ratio of the last two learning periods: $k = \text{length2}/\text{length1} = 0.33b/1h = 0.33$.

3.2.2 User Intention Recognition for Retrieving

Content

3.2.2.1 User's query intention recognition strategy

The specific thought flow of query intention identification is shown in Figure 2:

![Figure 2](https://example.com/figure2.png)

**Figure 2.** retrieval of related theme map screening

Step1. According to the knowledge points of the query and the bottom level knowledge elements in the user's personal topic map, the relevance degree of the knowledge points is matched. The relevance degree of the topic map and the query knowledge points is the cumulative degree of integration of each underlying knowledge element and the query content in the topic graph.

Step2. If the correlation degree of the user theme map with the highest correlation degree score in the step1 calculation results is greater than 0.1, it will be employed as the user query topic map, and the knowledge element in the topic map will be employed as the user's intention for querying the knowledge points, together with the corresponding weights in accordance with the degree of relevance of the user and the knowledge element and the query content knowledge point; if the correlation degree of the user theme map with the highest correlation degree score in the step1 calculation results is less than 0.1, then it suggests that the user has not learned related knowledge earlier, implying that it does not require retrieval of intention recognition and direct search of the content related knowledge points and association weights as the final user to retrieve knowledge points.

3.2.2.2 The related algorithms for query intention recognition

The related algorithms are implemented as follows:

```java
for(int i=0;i<num;i++){
    String themeid = jsonObject.getString("themeid");
    String Relateknow=getthemeRelateKnow(themeid);
    JSONArray themeKnowJsonArray = new JSONArray(Relateknow);
    for (int h = 0; h < themeKnowNum; h++) {
        String knowid = jsonObject0.getString("knowid");
        String weight=jsonObject1.getString("weight");
        for (int t = 0; t < inputKnowNum; t++) {
            String kid = jsonObject1.getString("knowid");
            double similar=getSimilar(knowid,kid);
            BScore=BScore+(cor_weight*weight*similar);}
        sum =String.valueOf(inputKnowNum);
        BScore=BScore/sum;
}
```

3.2 Resource Matching Strategy in Retrieval Process

3.3.1 The calculation of correlation degree between each knowledge point and query content in retrieval topic map

The matching of the resources in the search process is the matching of the theme map and resources filtered by the user intent. Owing to the fact that query topic maps contain knowledge points and other knowledge points, it is deemed essential to develop a comprehensive knowledge correlation calculation between the other knowledge points (other than the query content knowledge points) and the query content knowledge points within the query topic map.

3.3.2 User intention topic map and resource
In the mechanism of the user intended theme map and resource matching, primarily, the resources under the intended theme graph are aggregated for forming a collection of candidate resources. Secondly, with the help of the knowledge point and the resource association library, the repository is capable of directly finding the resources associated with the knowledge points in the query topic map. A knowledge point is likely to correspond to the multiple resources having varied relational weights.

4. Experimental Design and Result Analysis

The research of this paper is based on the knowledge base, knowledge point and resource relation database.

4.1 Experimental Objective

The experiment in this paper aims at verifying whether the retrieval methodology based on the knowledge-based resource retrieval strategy is capable of promoting and improving the robustness and retrieval performance and user's intention recognition precision, in comparison with the conventional global retrieval and resource retrieval strategy.

4.2 System Query Performance Test

4.2.1 Index of retrieval quality evaluation

The retrieval impact of an information retrieval system typically requires evaluation by the evaluation index. Among them, the two most important and classical indices in the performance evaluation of the retrieval system include recall rate and precision ratio.

In this chapter, the aggregate number of educational resources existent in the retrieval system has been represented by S; a suggests the related resource results retrieved by the system; b indicates the result of the irrelevant educational resource retrieved; c depicts the undiscovered educational resource result; and, d is a representation of the results of undocumented educational resources, which have not been retrieved. Where $a = b + c + d$; $a + b$ = the retrieved result; $c + d$ = the undetected result; $a + c$ = the related result; and, $b + d$ = the uncorrelated result. The formula is calculated as hereunder:

$$\text{Precision} = \frac{a}{a+b} \quad (1)$$

$$\text{Recall} = \frac{a}{a+c} \quad (2)$$

4.2.2 Experimental Objective

The search theme input by the user is Q, and the theme Q is described as five types of keywords, short sentences, short paragraphs, related knowledge points and topic knowledge points, correspondingly, to be input as a query. The experimental results of the 5 kinds of input situation of the user's 10 query intent system precision and recall rate statistical analysis, to achieve the retrieval system's knowledge associated retrieval performance evaluation.

4.2.3 Experimental data

For the purpose of reflecting the precision of the retrieval system to figure out and extract the associated knowledge points, this paper selects different knowledge levels of no cross-knowledge points as the standard of query content selection.

4.2.4 experimental result

The results of the evaluation are shown in the following figure:

![Fig. 3. Retrieval system precision experiment based on knowledge point correlation](image)

![Fig. 4. Recall experiment of retrieval system based on knowledge point correlation](image)

The above empirical results indicate that, first of all, the average precision of the system subjected to the "keyword" input format is slightly higher than the "topic knowledge point", but the average accuracy difference of the two amounts to be approximately 0.3. This suggests that the system possesses a higher user intention recognition ability for fuzzy queries.

Secondly, the average precision subjected to the input form of the keywords is slightly higher than the short sentence, and the average difference between the two's precision amounts to be approximately 0.5, implying that the system's semantic analysis of the sentence is basically precise. The precision of the "paragraph" input form is the lowest, revealing that the system does not carry out precise analysis of the paragraph semantics.
Ultimately, the system precision rate subjected to the input of “associated knowledge points” is lower than the "keywords", "thematic knowledge points" and "short sentences", but its "average precision and precision" is the "key The word ". The word "is floating approximately 0.1, together with being higher than the average precision rate in the form of "short paragraph" and "traditional keyword"; this suggests that the retrieval system based on the knowledge association relationship in this paper is more precise in the identification and extraction of associated knowledge points, and the precision of the retrieval relative to traditional keywords is improved by approximately 34%.

As regards the recall rate, the recall rate of the academic resource retrieval system based on the knowledge association relationship is approximately 80% on average that is almost 48% higher as compared with the traditional keyword retrieval.

5. Conclusion

This paper throws emphasis on the improvement of the calculation of the correlation degree of knowledge points and the personalized retrieval of users, making use of the domain large-scale corpus and the locally constructed knowledge base, the two-level matching of the knowledge points' semantic level and the knowledge structure level is realized; moreover, the precision of calculation of the knowledge point association weight is also improved. Considering the personalized retrieval aspect, this paper emphasizes the improvement of the diversity of users' historical behavior data sources, in addition to the timeliness and repetition of the knowledge points; moreover, the issue of data sparsity through a single retrieval log is avoided, and the historical behavior knowledge points of the users are treated with the forgetting curve of Hermann Ebbinghaus. In addition, the precision of the user attention is optimized, followed by improvement of the precision of the system. Whole extracting the knowledge point association relation, this paper puts forward knowledge topic map as the carrier of knowledge structure that is employed for expanding the preliminary query of the knowledge points and improving the recall rate of the system.

Acknowledgment

This research was partially supported by the National Science and Technology Support Program of China(No. 2015BAH33F01)

Reference