

# Research on Recommendation of Online Materials Course Resources Based on Text Similarity

Ziyu Liu\*, Mengying Yao

Hebei University of Science and Technology, Shijiazhuang, Hebei, China

\*Corresponding Author.

## Abstract

*In order to solve the problem that it is difficult for college students to find learning resources related to the courses they are learning quickly and accurately in blended learning. This paper proposes an online materials course resources recommendation method based on text similarity. Firstly, collecting the data of course resources on the online learning platform through web crawler technology. Secondly, preprocessing the data which contend deleting noise data, the Chinese word segmentation and calculating the course similarity based on cosine similarity then getting the course recommendation results according to the similarity ranking. Thirdly, evaluating the recommendation results and optimizing the similarity calculation method according to the evaluation results. Finally, the learners are recommended curriculum resources according to the similarity ranking results. According to the courses learned on the Superstar platform, the experiment recommends similar course resources on the XueYin Online platform. The results show that the online materials course resources recommendation method based on text similarity can recommend relevant online materials course resources for learners quickly and accurately, which has certain reference significance and application value for online course resources recommendation.*

**Keywords:** Mixed teaching, online learning, text similarity, recommended learning resources, materials courses

## 1. Introduction

With the continuous development of information technology, the learning model with "Internet plus" as its basic feature has caused a fundamental change in the way of acquiring knowledge. Teaching and learning are no longer restricted by time, place, personnel and other conditions. The education model is undergoing profound changes [1]. The blended education model, which combines the advantages of online teaching and offline teaching has become the product of the new era in the education field. The blended education model is getting more and more attention. The blended education model formed with the development of Internet technology combines online and offline teaching well [2]. In the context of the blended teaching model. Superstar learning platform used by many colleges and universities because of its complete functions and easy operation. Superstar learning platform is a smart teaching platform that integrates online course platform and interactive teaching platform [3]. Superstar learning APP includes three modules: course interaction, learning resources, and learning management. The course interactive module includes activities such as randomly selecting students, answering questions, questionnaires, topic discussions, online tests, homework, examinations, sign-in and etc. The earning management module includes interactive course scores, number of course visits, video learning time and etc. The learning resource module includes course content, course materials and etc.

On the basis of learning in class, many college students will learn on the MOOC platform. On the one hand, learning on the MOOC platform makes up for the unintelligible content of the course, and on the other hand expands the content of the course learning. XueYin Online is an online learning platform of ChaoXing Company, which contains more than 2,100 course resources. The course introduction, faculty team, and course chapters of each course are explained in detail. When learners study on the learning platform, due to the large number of learning resources and the uneven quality of learning resources [4-5], it is difficult for them to find the target learning resources quickly and accurately. They waste a lot of time for screening. Therefore, finding a faster and more accurate method for recommending online course resources is a problem that needs to be solved urgently.

## II. Related Work

Text similarity refers to obtaining a value through a certain method to describe the similarity between text objects such as words, short sentences, documents and etc. Text similarity algorithm is an important algorithm in text mining. It is also the link between basic research such as text modeling and representation and upper-level application research of text potential information [6]. There are many ways to calculate text similarity. The classification of calculation methods recognized by most scholars is as follows: String-Based Method, Corpus-Based Method, Knowledge-Based Method and Hybrid Method [7-9]. Wang [6] summarized the calculation methods of text similarity into two cases: Surface Text Similarity and Semantic Similarity [10]. The main difference is whether to consider tone, emotion and other semantic factors when calculating text similarity. The research data of this paper is the course text which includes the course name and course chapters. The requirement for semantic analysis is low. Therefore, this paper studies the surface text similarity. In terms of surface text similarity research, Zhao et al [11] made flexible adjustments based on the characteristics of data in different fields in the power grid asset model, and designed a short text similarity calculation method with a degree of adjustability. Ai [12] based on text mining proposed a patent recommendation method to provide users with the patent information needed in a large number of patents. Ye [13] designed a doctor recommendation model using data from the "Good Doctor Online" platform. When designing the algorithm, the text similarity between the patient's decision factors and the doctor's information is considered, so that the recommended doctor is more in line with the actual needs of the patient. Yang [14] established a model to recommend cases and key information related to the handling of cases for the court by calculating text similarity. It can be seen from the above research that the text similarity algorithm is practical in the application of resource recommendation. At the same time, it can be assumed that the higher the similarity value between texts, the greater the commonality, and the lower the similarity value, the greater the difference. In actual application, the calculation of text similarity should be improved according to the different attributes of the recommended resources.

In order to help learners to obtain relevant learning resources quickly and accurately, this paper proposes an online materials course resources recommendation method based on text similarity. According to the courses that learners studied in Superstar Learning APP, this paper recommends the related course on the XueYin online platform. This paper calculates the similarity between courses by calculating text similarity, and recommends learning resources with high similarity to learners, which can reduce the time for learners to search and screen courses resources. In order to obtain a better recommendation result, this paper takes into account the actual situation of the course resources when calculating the text similarity, calculates the similarity for the course name and chapter content separately, and assigns different weight values [15-16].

## III. Research Method and Process

As shown in Fig 1, the research process of this paper is divided into four parts.

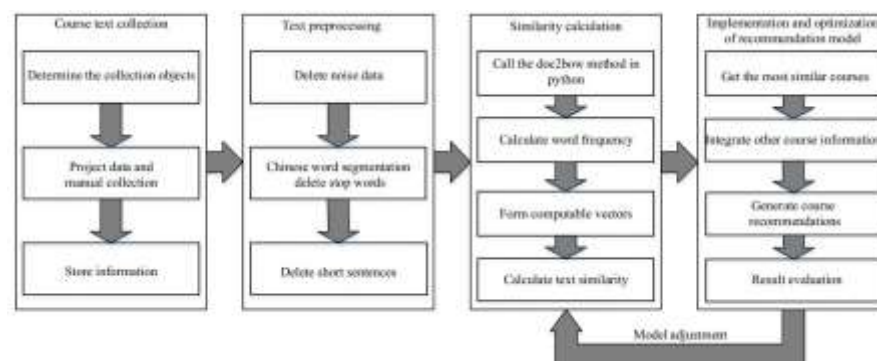


Fig 1 Research process

### 3.1 Data collection

The research data in this paper comes from the Superstar learning APP and the XueYin online platform. The data of the Superstar Learning APP is the course data learned by students in 2020 who majored in materials science and engineering and polymer materials and engineering in Hebei University of Science and Technology. The course data obtained on the Superstar learning APP is used as the recommended reference course and the course data obtained on the XueYin online platform is used as the recommended resource set.

### 3.2 Data preprocessing

Firstly, sort the collected data. Use Excel and Python to merge the related data and delete the noisy data. Secondly, call the jieba word segmentation tool in Python to segment the course text and delete stop words. Then convert the text data into a computable format. In word segmentation, special stop words in different fields and ultra-short sentences have a greater impact on the analysis results. Therefore, such invalid statements need to be deleted. This step is one of the most important steps in the similarity calculation. Failure to complete this step will affect the accuracy of the results.

### 3.3 Similarity calculation

Huang W B [17] and Li L [18] pointed out in the study of text similarity calculation methods that the most commonly used text similarity calculations are the bag-of-words model and the word vector model. The bag-of-words model is more suitable for normative texts and short texts and the word vector model is suitable for large-scale text analysis. The analysis found that the research data of this paper has following characteristics:

The data set used for research is short text;

The text of each course data is less;

The content of the course data is concise.

Based on the above factors, this paper finally chooses the method of combining Doc2bow and TF-IDF to vectorize the text and finally uses the cosine similarity to calculate text similarity.

During the model experiment, call Doc2bow in Genism of Python program. Doc2bow is mainly used to implement the typical bag-of-words model which treats text information as a collection of independent words without interfering with each other. It ignores the word order and grammar of the sentence and uses a set of unordered words to represent text information [12]. Then put all word segmentation into word bag and use Doc2bow to count word frequency. However, only considering Doc2bow will lead to sparse results, so using TF-IDF to calculate the word frequency makes the results more accurate. Finally, text vectors are formed to calculate the similarity of the text [19]. TF-IDF is used to evaluate the importance of a word in the text. The main theory of TF-IDF is that if a word appears in a course text with high frequency (TF) and the word rarely appears in other course texts then it can be considered that this word can well distinguish the course text from others. But if a word appears frequently in many course texts, then the distinguishing ability of the word will decrease as the number of course texts increase. The calculation formula of TF-IDF is:

$$\omega(d, t) = TF(d, t) \times IDF(t) \quad (1)$$

$$IDF(t) = \log\left(\frac{N}{DF(t)}\right) \quad (2)$$

In formula (1),  $TF(d, t)$  represents the word frequency which represents the number of occurrences of the word 't' in the text 'd';  $IDF(t)$  represents the frequency of the inverse text. In formula (2),  $DF(t)$  represents the number of documents in which the word t appears in the text collection; N represents the total number of texts. When calculating text similarity, the most common method is cosine similarity. The calculation formula is:

$$\text{similarity}(\alpha, \beta) = \cos(\theta_{\alpha, \beta}) = \frac{X \cdot Y}{\|X\| \cdot \|Y\|} = \frac{\sum_{i=1}^n (x_i \times y_i)}{\sqrt{\sum_{i=1}^n (x_i)^2} \times \sqrt{\sum_{i=1}^n (y_i)^2}} \quad (3)$$

In formula (3), X and Y represent the word frequency vector corresponding to the course text,  $x_i, y_i$  represent the component values of the vectors X and Y. The similarity result range of the cosine similarity calculation is [0,1] and the larger the value the higher the text similarity.

### 3.4 Implementation and optimization of recommendation model

After obtaining the similarity results between courses, sort them in descending order of similarity. Then get related courses that are highly similar to the reference courses. At last, recommend the courses which include course title, chapter content and lecturer to the learners. After getting the recommendation result, use the recall, precision and F-score value to evaluate the result and optimize the recommendation model based on the evaluation result.

Recall represents the number of correct predictions of related courses in the course data. There are two results, the first result is that the relevant course is predicted correctly called True Positives which means the recommended courses are relevant courses. The second result predicts related courses as unrelated also known as False Negatives which means a course is not recommended but the learner considers the course related. The Recall calculation formula is formula (4).

$$\text{Recall} = \frac{\text{count}_{TP}}{\text{count}_{TP} + \text{count}_{FN}} \quad (4)$$

Precision counts the number of recommended courses that learners considered relevant based on the recommended results. There are also two results, the first predicts related courses as relevant known as True Positives which means the recommended courses are relevant courses. The other predicts irrelevant courses as relevant called False Positives which means the learners believe the recommended courses are not relevant. The Precision calculation formula is formula (5).

$$\text{Precision} = \frac{\text{count}_{TP}}{\text{count}_{TP} + \text{count}_{FP}} \quad (5)$$

In actual operation, there may be inconsistencies between the changing trends of Precision and Recall. In order to evaluate the results of the course recommendation more accurately, the F-score is used to comprehensively judge. The F-score calculation formula is formula (6).

$$F_{score} = \frac{2 * P * R}{P + R} \quad (6)$$

In formula (6), P is the Precision and R is Recall.

## IV. Data collection and processing

### 4.1 Data collection

#### 4.1.1 Data of Superstar Learning APP

The data of Superstar Learning APP is provided by Chaoxing Company. The data mainly includes two parts: basic information of students and course data. The basic information of students includes name, gender, student ID, major and etc. The data used in this paper is the students' name and student ID which are used to search the course selection when recommending. The student data format is shown in Table 1

Table 1 Learner data

Student ID	Name	Course selection list
16020711	Qin Zheng	Polymer chemistry
		Polymer physics
		Research Methods of Polymer Materials
		Polymer materials
		Polymer matrix composite engineering

The course data contains 120 courses studied by 380 students on the Superstar learning APP in 2020 who are majoring in materials science and engineering and polymer materials and engineering from Hebei University of Science and Technology. Each course data includes course ID, course name, chapter name and chapter content. The data used in this paper are course name, chapter name and chapter content which are used to calculate text similarity. When recommending, public lessons such as mathematics, English that have no professional specificity are manually deleted because of the recommendation utility is low. The course data of Superstar Learning Platform format is shown in Table 2.

Table 2 Course Data of Superstar Learning Platform

Course Title	Course ID	Chapter content
Polymer chemistry	216653529	Introduction; Gradual Convergence; Free Radical Polymerization; Free Radical Copolymerization; Aggregation Method; Ionic Polymerization; Coordination Polymerization...
Polymer physics	214401709	Introduction; polymer chain structure; polymer condensed structure; polymer thermal motion characteristics and polymer mechanical state; the concept of elasticity and viscosity; high polymer elasticity...
Fundamentals of Materials Science	216895524	Introduction; General Plastics; Engineering Plastics; Synthetic Fibers; Rubber; Coatings and Adhesives; Functional Polymer Materials...

#### 4.1.2 Data of XueYin online

XueYin online data was captured through Python web crawler system, including 5507 courses data. Operation process for obtaining web page data is shown in Figure 2.

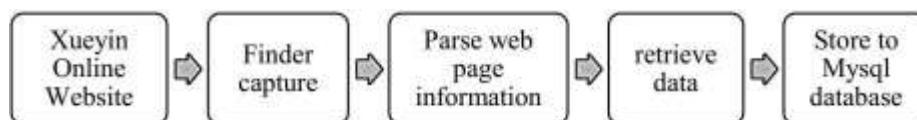


Fig 2: Operation process for obtaining web page data

The data captured on the XueYin online include course ID, course name, chapter name, chapter content, course introduction, picture file, URL, university/college and lecturer. The data used in this paper are course ID, course name, chapter name, chapter content, university/college and lecturer. Course name, chapter name, chapter content is used to calculate text similarity. Course ID, university/college and lecturer are used when generating recommendations. When retrieving the data of XueYin Online, the data contains the subject name to which the course belongs, the name of the course and the chapter content of the course. The subject name can be obtained in the response returned for the first time but the course name can only be obtained by clicking on the link of the subject. The response returned for the first time contains links to the content of the course chapters. Get the chapter content links through data analysis and finally get the relevant course content data through initiating request. The obtained course data contains multiple files, so it needs to be processed before use. Firstly, merge multiple tables based on common fields such as course ID for easy viewing and use. Secondly, delete the noisy data. The course data of XueYin Online Platform format is shown in Table 3.

Table 3 Course Data of XueYin Online Platform

Course ID	Course Title	Lecturer	University/Coll	Chapter Content
-----------	--------------	----------	-----------------	-----------------

			<b>ege</b>	
216713829	Principles and Applications of Psychology	Zhuang Guoping	Weifang University	What is psychology; how to study psychological phenomena; the past, present and future of psychology...
216470271	Fault diagnosis and maintenance of CNC machine tools	Zhu Zhanke	Shaanxi Polytechnic Institute	Fundamentals of CNC machine maintenance; the classification of CNC machine tools...
213332222	Property management regulations and case analysis	Zhu Yuan	Sichuan Post and Telecommunication College	Overview of property management laws and regulations: property management laws and regulations...
217228547	Internet thinking	Zhu Liyang	Hunan Arts and Crafts Vocational College	Introduction to Internet Thinking: The Connotation of Internet Thinking; The Development of Internet Thinking...

#### 4.2 Data preprocessing

Word segmentation is an important step in data processing. Cutting continuous sentences into short words can get better results in subsequent natural language processing [20]. This paper uses the jieba tool of the Python to segment the course text. The jieba tool is the most commonly tool when segmenting text which including three modes: precise cutting, full cutting and search engine. This paper uses precision cutting.

Stop words have no actual meaning in the text such as modal particles, quantifiers and etc. The filtering of stop words can effectively reduce the dimensionality of the data. Eliminating useless words and keeping words that can represent the meaning of the text can improve the efficiency of word search [21]. In this paper, after using the Harbin Institute of Technology's common stop words list to delete stop words, the results of the research found that there are stop words that are not sufficiently distinguishable for the course text such as "course", "test", "unit" and etc. To solve this problem, some stop words were manually added. Some stop words are shown in Table 4. When the program is implemented, word segmentation and stop words are performed at the same time and part of the final word segmentation result are "material", "computer", "polymer", "storage", "modeling", "chemistry", "language", "the Internet", "system", "data", "physical", "transaction processing", "information management".

Table 4 Stop words list

<b>Common stop words</b>	<b>Special stop words</b>
if	course
therefore	test
but	unit
firstly	chapter
then	1.1

### V. Implementation and Optimization of the Recommended Model

#### 5.1 Experiment data

This paper obtains the course data required for model realization from the Superstar Learning APP and XueYin Online Platform. The course data is divided into learning course data set and course resource set. The learning course data set contains 120 courses studied by students of Hebei University of Science and Technology on the Superstar Learning APP. Course resource collection is 5507 courses crawled from XueYin online platform through web crawler technology. In order to facilitate the evaluation and reduce the running time of the program, this paper has extracted 200 courses from the course data set and 10 courses from the learning course data set for experiments. Based on the 10 courses learned by the learners, select courses in the course resource set that are highly similar to the

courses studied by calculating course similarity and recommend to the learners.

Taking into account the evaluation criteria of the model results, 30 students, majoring in materials science and engineering and polymer materials and engineering from Hebei University of Science and Technology, were selected to vote for relevant courses for the 10 courses in the resource set containing 200 courses. Select the course with the higher number of votes which is the number of votes more than 15 to be included in the relevant course. Finally, the relevant course resources of 10 learning courses are obtained as standard course resource sets for evaluation. Part of the standard course data is shown in Table 5.

Table 5 Standard Course Data

Course Title	Voting course		
	Course ID	Course title	lecturer
Polymer chemistry	214820470	Polymer chemistry	Zhang Quanyuan
	207149505	Polymer chemistry	Chen Xuegang
	200113724	Polymer chemistry	Wang Dongmei
	206842166	Polymer chemistry	Tan Licheng
Polymer physics	214401709	Polymer physics	Deng Wei
	208987388	Polymer physics	Wen Huiying
	207801465	Polymer physics	Chen Yiwang

## 5.2 Implementation of the model

After preparing the data, use the chapter content of the course as text data for calculating the similarity of the course. Part of the program of the model realization process is shown in Table 6. Part of the program of the model realization process is shown in Table 6. Result evaluation results Table 8.

In Table 6, “Word\_Res” represents the result of segmenting the chapter text. “Func.Cosine()” is a class written in advance, which includes the process of word segmentation, deletion of stop words, and similarity calculation of the course text. R represents the similarity threshold for screening. Courses with similarity higher than R are recommended to learners. In the experiment, R is 0.45 which means recommend courses with similarity higher than 0.45 to learners.

Table 6 Initial algorithm program

Algorithm 1
<Input: Recommended reference courses; recommended course resource collection>
<For Chapter in Zip (Chapter_List) :> <Function=Suanfa (Chapter)> <Word_Res=Function.Spilt_Word ()> <Print (Word_Res)> <Response=Func.Cosine()> <If Response>=R:> <Print ((Title, Response))> <Res.Append((Title2, Response))> <Return Res>

<Output: Recommended results>

In Table 7, "Studied Course" is listed as the courses studied by the learner and "Recommended Course" is listed as the courses with high similarity to the "Studied Course" in the recommended resource. Only one course with the highest similarity is listed in the table.

Table 7 Initial recommendation result

Course ID	Studied Course	Recommended course	Lecturer	University/College	Similarity
216653529	Polymer chemistry	Polymer chemistry	Zhang Quanyuan	Hubei University	0.609
214401709	Polymer physics	Polymer physics	Wen Huiying	Northeast Forestry University	0.6036
216840629	Physical Chemistry	Physical Chemistry	Wang Xinping	Dalian University of Technology	0.5782
216895524	Fundamentals of Materials Science	Fundamentals of Materials Science	Dong Binghai	Hubei University	0.5499

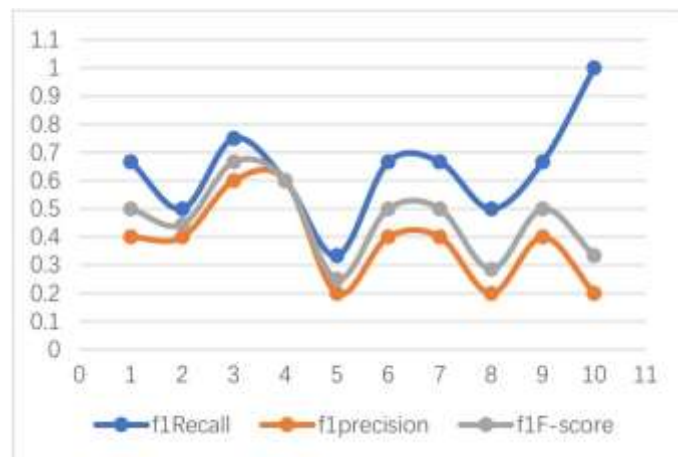


Fig 3: Evaluation of initial recommendation results

As shown in Figure3, the three curves are recall, accuracy and F-score in the evaluation of the recommended results of 10 learning courses. Most of the values are between 0.3 and 0.7. The explanation is good but it can still be improved.

### 5.3 Model optimization

Analyzing the calculation results of the similarity of the courses, there is a problem that the actual correlation between the courses is relatively high, that is, most learners think that the two courses are related, but the calculation result of the similarity is not high. Therefore, the calculation method of similarity is optimized. Studies have found that if only the similarity of chapter content is considered when calculating the similarity of courses, it may cause errors in the results due to differences in chapter structure and chapter naming methods. This shortcoming can be compensated by integrating the similarity of the course name when calculating the similarity of the course. That is to say, when calculating the similarity of courses, the similarity of the course name and chapter content is calculated

separately. Then, assign different weights to the calculation results of the course name and chapter content and get the final similarity result. Part of the program after the optimization of the calculation method is shown in Table 8. The recommendation results obtained according to the similarity ranking are shown in Table 9.

Table 8 Optimized algorithm program

Algorithm 2
<Input: Recommended reference courses; recommended course resource collection>
<pre> &lt;For Titlie2, Chapte2 In Zip (Test_Name, Test_Chapter) :&gt;   &lt;Obj=XYZX.Plus(Titlie2,Chapter,Chapter2,R<sub>1</sub>,R<sub>2</sub>)&gt;   &lt;Res=Obj.Response()&gt;   &lt;If Res&gt;=R:&gt;     &lt;Response_List.Append((Name,Lecturer,School))&gt; &lt;Return Response_List&gt; </pre>
<Output: Recommended results>

In Table 8, "XYZX.Plus" is a class written in advance, which includes the process of word segmentation, deletion of stop words, and similarity calculation of the course text.  $R_1$  and  $R_2$  respectively represent the weighting value of the similarity of the course name and chapter content. R represents the similarity threshold for screening. Because the length of the course name is shorter than that of the chapter content, the information value of the content contained in the course name is less. So, the weight of the course name is given a low weight, namely,  $R_1 < R_2$ . In the experiment,  $R_1$  is 0.3,  $R_2$  is 0.5, R is 0.45.

Table 9 Optimized recommendation results

Course ID	Studied Course	Recommended Course	Lecturer	University/College	Similarity
216882279	Mechanics of materials	Mechanics of materials	Zhao Zenghui	Shandong University of science and technology	0.7222
216653529	Polymer chemistry	Polymer chemistry	Zhang Quanyuan	Hubei University	0.6694
216840629	Physical Chemistry	Physical Chemistry	Wang Xinping	Dalian University of Technology	0.6503
216527153	Material analysis method	Modern testing and analysis technology of materials	Zhang Aisheng	Nanchang University	0.6437

In Table 9, "Studied Course" is listed as the courses studied by the learner and "Recommended Course" is listed as the courses with high similarity to the "Studied Course" in the recommended resource. Only one course with the highest similarity is listed in the table. It can be seen that after the optimization of the text similarity calculation method, the value of the curriculum similarity has been improved.

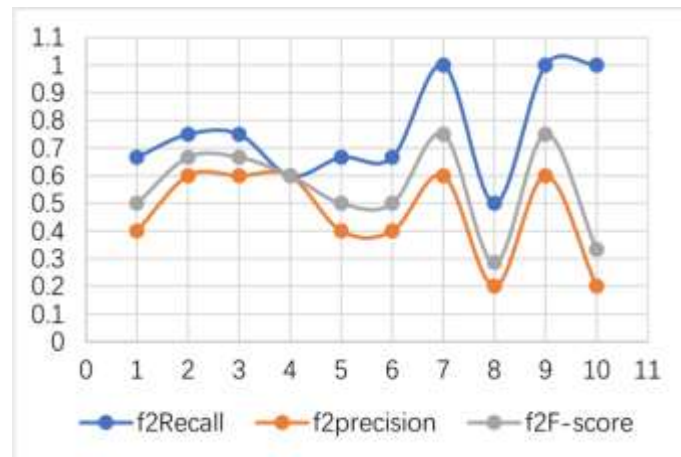


Fig 4 Optimized recommendation result evaluation

As shown in Figure 4, the recall, accuracy and F-score are mostly between 0.5 and 0.8. Each evaluation value has been improved, indicating that the adjustment of the recommendation algorithm is effective.

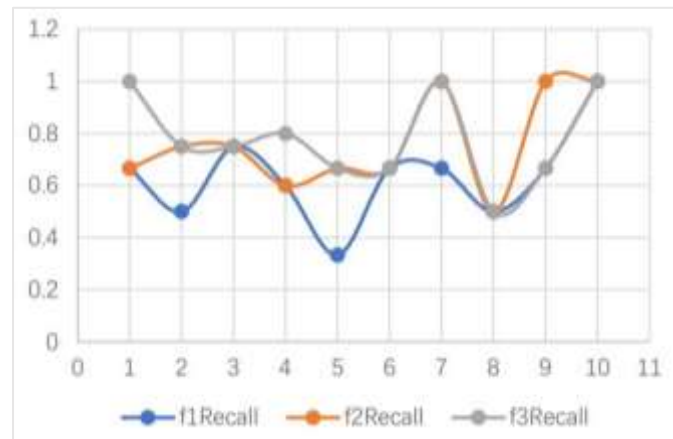
#### 5.4 Weight adjustment and analysis

In the recommendation model after optimizing the text similarity calculation method, the weight assignment of the course name and chapter content similarity is mainly determined according to the length of the text. Because the length of the text of the course name is shorter than that of the chapter content, the information value of the content contained in the course name is less. So, the weight given to the course name is lower than chapter content. Compared with the text similarity calculation method that only considers the content of the chapters, the recommended results of the optimized method are improved slightly. Compared with the text similarity calculation method that only considers the content of the chapters, the recommended results of the optimized method are slightly improved. But the weight assignment can be improved. The improvement idea is that when the similarity of the course name is higher, it should be given a higher weight, otherwise, it should be given a lower weight.

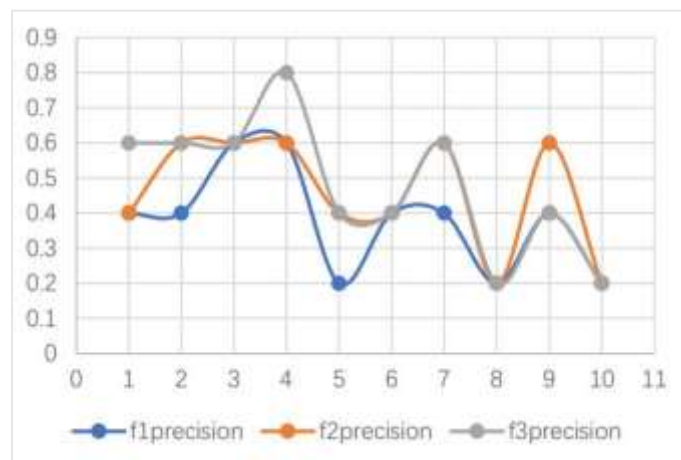
Table 10 Example

Course ID	Studied Course	Recommended Course	Similarity
214540653	Solid state physics	Solid state physics	0.3083

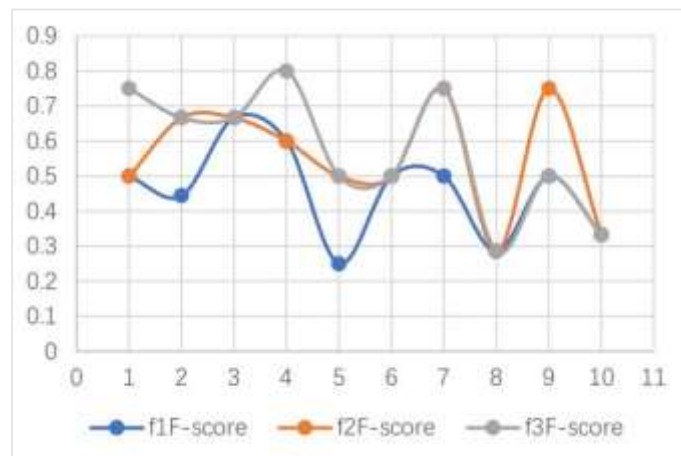
The recommended course in Table 10 was deleted while screening the similarity threshold but it is highly relevant to the studied courses. This is because the differences in the presentation method and chapter structure of the chapter content of different courses lead to low text similarity in the calculated chapter content. Because the weight given to the course name is small, the final similarity is low. Based on this special situation, the similarity of the course name and chapter content are dynamically weighted. Mainly to adjust the weight value of the course name. Comparing the similarity results of the course names, if the similarity is higher, the course name will be given a higher weight, otherwise, the weight will be lower. After optimizing the algorithm, select the top 5 courses in similarity among the courses recommended for the 10 studied courses in different similarity calculation methods. The recall, accuracy, and F-score changes of the recommended results are shown respectively in Figure 5, Figure 6, and Figure 7. In Figure5, Figure 6, Figure 7, f1 is the similarity calculation method that only considers the chapter content, f2 is the similarity calculation method that assigns fixed weights to the course name and chapter content and f3 is the similarity calculation method for dynamically adjusting the weighting of course names.



*Fig 5 Change in Precision*



*Fig.6 Change in Recall*



*Fig.7 Change in F-score*

It can be seen from the three figures that after adjusting the weights, the recall, accuracy and F-score have improved, indicating that the optimization is effective. In the process of model realization, the number of courses with similarity above the threshold is in a decreasing trend, but the accuracy is getting higher and higher. From an application point of view, although the number of courses finally obtained by the recommendation model is decreasing, most of the courses finally recommended are resources that are related to the courses studied. It can be seen that when

recommending learning resources to learners, one should not only consider the number of recommendations, but more importantly, whether the learners need the recommended curriculum resources.

## VI. Conclusion

In the context of blending learning, based on the method of text similarity, this paper uses the course data that students studied in Superstar learning APP who are from Hebei University of Science and Technology of Economics and the course data of XueYin online platform as experimental data to complete the implementation and analysis of the recommendation model. The experimental results show that the online materials course resources recommendation based on text similarity can realize personalized learning resource recommendation for learners well. In the process of continuously optimizing the method, the accuracy of the recommended results has been greatly improved. This recommendation method has certain reference significance and application value in the field of learning resource recommendation. For learners, the quality of the recommendation of learning resources is more important than the number. Recommending personalized learning resources for learners more accurately is the focus of future research.

There are still deficiencies in this paper. On the one hand, the assignment of weights needs more in-depth research. Choose the most appropriate weight value in future research to make the result more accurate. On the other hand, the similarity calculation formula is single, only considering the cosine distance. Other calculation methods can be tried in future research.

## Acknowledgments

This work was supported by the Research and Practice Project of Higher Education Teaching Reform in Hebei Province, Research and Practice on Improving the Learning Effect of Blended Teaching Based on Big Data, under Grant [number 2019GJJG602] and the 100 Outstanding Innovative Talents Support Project of Hebei Province [number SLRC2019005].

## Reference

- [1] Z. Huang, "Exploration and Practice of Engineering Education Based on Mu Class and Hybrid Teaching," *Research in Higher Education of Engineering*, vol. 4, pp. 11-13, 2016.
- [2] Y.P. Tan, "The basic characteristics and implementation strategy of blending teaching mode," *Chinese Vocational and Technical Education*, vol. 32, pp. 5-9, 2018.
- [3] W.J. Wang, C.H. Xiang, L.J. Feng, Z.Y. Sun, T.P. Yin, "Practices and Thoughts About Online Teaching of Analytical Chemistry," *Chinese Journal of Chemical Education*, vol. 42, no. 6, pp. 49-54, 2021.
- [4] J. Wang, H. Xie, O. T. S. Au, D. Zou, F.L. Wang, "Attention-Based CNN for Personalized Course Recommendations for MOOC Learners," *2020 International Symposium on Educational Technology (ISET 2020)*. pp: 180-184, 2020.
- [5] W. Xu, Y.J. Zhou, "Course video recommendation with multimodal information in online learning platforms: A deep learning framework," *British Journal of Educational Technology*, vol. 51, no. 5, pp. 1734-1747, 2020.
- [6] C.L. Wang, Y.H. Yang, F. Deng, "A Review of Text Similarity Approaches," *Information Science*, vol. 37, no. 3, pp. 158- 167, 2019.
- [7] W.H. Gomaa, A.A. Fahmy, "Short answer grading using string similarity and corpus-based similarity," *International Journal of Advanced Computer Science and Applications*, vol. 3, no. 11, pp. 114-121, 2012.
- [8] J. Kadupitiya, S. Ranathunga, G.R. Dias, "Sinhala short sentence similarity calculation using corpus-based and knowledge-based similarity measures," *Proceedings of the 6th Workshop on South and Southeast Asian Natural Language Processing (WSSANLP2016)*, Osaka, Japan. pp 44-53, 2016.
- [9] W.H. Gomaa, A.A. Fahmy, "A Survey of Text Similarity Approaches," *International Journal of*

- Computer Applications, vol. 68, no. 13, pp. 13-1, 2013.
- [10] H.T. Nguyen, P.H. Duong, E. Cambria, "Learning short-text semantic similarity with word embeddings and external knowledge sources," *Knowledge-Based Systems*, vol. 182, no. 3, pp. 104824, 2019.
  - [11] Z.T. Zhao, H. Zhang, Y.L. Yu, "A cross weight product text similarity calculation method for power grid model search," *Water Resources and Hydropower Engineering*, vol. 51, no. S2, pp. 209-214, 2020.
  - [12] C.H. Ai, D. Jiang, J.D. Wu, "Patent recommendation research based on topic model and text similarity calculation," *Information Technology*, vol. 44, no. 4, pp. 65-70, 2020.
  - [13] J.X. Ye, J.X. Xiong, H.X. Xiong, W.X. Jiang, "A Physician Recommendation Algorithm Integrating Inquiries and Decisions of Patients," *Data Analysis and Knowledge Discovery*, vol 4, no. Z1, pp. 153-164, 2020.
  - [14] F. Yang, J. Chen, Y. Huang, "Court Similar Case Recommendation Model Based on Word Embedding and Word Frequency," 2020 12th International Conference on Advanced Computational Intelligence (ICACI2020), August 26 2020, Dali, China. pp: 165-170, 2020.
  - [15] L. Zhang, Z.J. Zhang, J.F. He, Z.Y. Zhang, "UR: A User-Based Collaborative Filtering Recommendation System Based on Trust Mechanism and Time Weighting," 2019 IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS2019), July 14 2019, Tianjin, China. pp: 69-76, 2019.
  - [16] H.J. Wei, D.H. Dai, "Collaboration Filtering Recommendation Algorithm Based on Ratings Difference and Interest Similarity," *Computer Science*, vol. 45, no. S1, pp. 411-414+435, 2018.
  - [17] W.B. Huang, S.K. Che, "Methodological System and Application Scenarios on Text Similarity Calculation," *Information Studies: Theory & Application*, vol 42, no. 11, pp. 128-134, 2019.
  - [18] L. Li, H. Li, "Computing Text Similarity Based on Concept Vector Space," *Data Analysis and Knowledge Discovery*, vol 5, pp. 47-58, 2018.
  - [19] I. Arroyo-Fernandez, C.F. Mendez-Cruz, G. Sierra, "Unsupervised sentence representations as word information series: Revisiting TF-IDF," *Computer Speech & Language*, vol 56, pp. 107-129, 2019.
  - [20] B.L. Hua, "Stop-word Processing Technique in Knowledge Extraction," *Data Analysis and Knowledge Discovery*, vol 8, pp.48-5, 2007.
  - [21] G.H. Feng, W. Zheng, "Review of Chinese Automatic Word Segmentation," *Library and Information Service*, vol. 55, no. 2, pp. 41-45, 2011.