Research on Intelligent Question Answering System Based on College Enrollment

BO SONG¹, XIAO-MEI LI²

¹Software College of Shenyang Normal University, Shenyang, Liaoning, China, 110034,
²Liaoning Basic Education Research and Training Center, Shenyang, Liaoning, China, 110034
songbo63@aliyun.com

Abstract. Due to the increasing number of educated students and the continuous improvement of teaching level, China has formed the largest higher education system in the world, and the information about college enrollment has become the focus of the attention of Chinese students and parents. The intelligent question answering system for college enrollment can understand the user's intention, provide students and parents with the enrollment information of the school more quickly and efficiently, make up for the shortcomings in the enrollment work, and greatly improve the efficiency of enrollment.

1. Introduction

In recent years, the number of college entrance examinations across China has reached record highs. However, because students and parents have different understandings of the school, in the face of external information or unilateral propaganda of some staff members, there are often certain restrictions and not objective in selecting schools, so choose a school that suits you. Professionalism is a big problem for candidates. With the continuous increase in the number of students, how to stand out among many universities and attract more high-quality students has become the focus of attention of all universities. The choice of students and parents for schools and majors depends to a large extent on the information they receive. Therefore, admissions counseling for college entrance examinations is the key to the success of college admissions. In China, millions of candidates take the college entrance examination every year. Faced with overwhelming enrollment information, parents and students use three methods to obtain information. The first is through the promotion of paper platforms such as the "Guide to Admissions for General Senior High Schools". This method often fails to accurately and meticulously display the style and characteristics of the school because of the length of the text, making it difficult for students and parents to choose a school that suits them from many schools and majors. The second is to search for interesting school information through search engines. However, traditional search engines are also difficult to meet people's requirements for accuracy and efficiency. The shortcomings are:

- The search engine returns the user's desired answer by matching the keywords entered by the user. However, users get a large number of web pages related to keywords, so users can't find relevant information quickly and accurately, which wastes time and effort.
The method of keyword matching lacks the ability to analyze and understand long sentences stacked in a large number of words, which makes the search engine unable to accurately understand the meaning of the user, and is not conducive to quickly finding the answers required by the user.

The information found is not accurate enough. Since a large number of advertisers are now advertised on the Internet, users will get a lot of ads with no reference value. At the same time, the information matched by the search engine may only contain a small amount of information that the user needs, not an exact answer.

The third is to obtain school information by telephone consultation or online consultation. However, this method will increase the manpower input and material resources of the school, increase operating costs, and waste resources. Therefore, in order to improve the competitiveness and enrollment efficiency of colleges and universities in enrollment, reduce the operating costs of enrollment consultation in colleges and universities, apply natural language processing technology, understand and analyze user semantics, and develop intelligence for college enrollment that can accurately return information needed by users. The question and answer system have its necessity and practical significance.

2. Intelligent Question Answering System Based on College Enrollment

This paper mainly studies the intelligent question answering system in the limited field. It selects the field of enrollment in colleges and universities and studies the intelligent question answering system based on the FAQ library for college enrollment. Users ask questions about enrollment through natural language, first classify the problem, then analyze the problem through word segmentation technology, then extract and expand the keywords, then match it with the questions in the FAQ library, and finally consult the user. The information is returned to the user. If the match fails, the question is sent to the text library, where it is matched and returned to the user. Its main research contents include:

- Data collection and processing – collect 200 questions that students and parents are most concerned about and most frequently consulted, and answer these questions. The summarized questions and their supporting answers are summarized to form an original one. FAQ library. Afterward, these issues are classified into categories such as basic school information, annual admission scores, enrollment number, professional division, and featured teaching. At the same time, the school's enrollment brochures and other information are collected and entered into the document library. The question and answer in the FAQ library are combined with the documents in the document library to form the system's knowledge base.

- Problem Analysis——Because the system answers the questions for college enrollment, the scope of coverage is small, so the questions raised by users are distinctive and easy to identify. When the user enters a question, the problem type can be classified according to the problem defined in the FAQ library, or the rule-based classification method can be used to identify the keyword to clarify the problem type. In order to improve the accuracy of word segmentation, the word weight is calculated by the traditional TF-IDF technique after the word segmentation is performed using the RMM (reverse maximum length matching method) algorithm.

Answer search and return - the user's question and the question and answer pair in the FAQ library, through the innovative vector space model algorithm to compare the similarity, find the information that the user wants, and return it to the user; when in the FAQ library When the search is not available,
use the same method to search in the text library, and return the answer to the user while updating the FAQ library.

### 2.1 Overall design of the system

As shown in Figure 1, the main function of the intelligent question answering system for college enrollment is to answer the questions raised by users about enrollment consultation, and to summarize the questions and answers frequently asked by college enrollment, and to form them into a set of frequently asked questions (FAQ). At the same time, all the information in the enrollment guide is entered into the document library, which combines to form the knowledge base of the system. After the user enters the question, the system analyzes the problem entered by the user. Because the system is a question-and-answer system oriented to a limited domain, the problem characteristics are more conspicuous and easy to discriminate. According to the problem type and rule-based classification method defined in the FAQ library, the problem is classified, and after the problem type is determined, the problem is divided into words, stop words, keyword extraction and expansion. Match the problem obtained after the analysis with the problem in the FAQ library. If the similarity obtained by the matching is greater than or equal to the threshold, the corresponding answer is directly returned to the user; if the similarity obtained by the matching is less than the threshold, then go to The text library searches, calculates the similarity between the problem and the problem in the text library, returns the answer with the highest similarity to the user, and updates the question and answer together with the FAQ library.

![Flow chart of the intelligent question answering system for college admissions](image)

Figure 1. Flow chart of the intelligent question answering system for college admissions

### 2.2 Design of the problem analysis module

Problem analysis is the core module for understanding the purpose of user questions. The core of the algorithm is natural language processing technology. This system mainly applies natural language processing technology to Chinese word segmentation, keyword extraction, and keyword expansion. It is to understand the problems raised by users in natural language as much as possible and return the most accurate answers to users. The overall flow and structure of the Chinese word segmentation are shown in Figure 2.
In terms of Chinese word segmentation, the inverse maximum length matching algorithm (RMM) is used. In terms of keyword extraction, it mainly extracts proper nouns (including nouns and verbs) and definite adverbs as important keywords, extracts adjectives as ordinary keywords and uses TF-IDF technology to assign different keywords; Because it is a system for college admissions, most of the questions asked by users are more concentrated. The types of questions can be determined according to the keywords in the sentences, which facilitates the docking with the knowledge base. In order to improve the accuracy and comprehensiveness of the answers, it is also necessary to extend the extracted keywords; finally, the optimized spatial vector model (VSM) is used to calculate the similarity of sentences.

2.2.1. Reverse maximum length matching algorithm
The algorithm advances the last character in the string to be matched, and the word with the longest suffix matches the word in the dictionary. If a matching word is found, the word is segmented and deleted in the string to be matched. Go to the word, if there is no word in the dictionary to match it, cut it into a word. The inverse maximum length matching algorithm applies to the Chinese Suffix tree. In the dictionary tree, the last character is placed on the first layer of the tree. When using the method of maximal length matching, a method of returning the maximum length matching a word from the specified position (offset) from the string to be sliced is required. From the back to the front, add a word to the suffix tree verbatim.

2.2.2. Binary model (Bigram)
In many cases, two separate words can also be combined into a new word that is completely different from the two words. This situation is called combination ambiguity, such as "Shenyang/University" and "Shenyang University." The maximum length matching method can't correctly distinguish the combination ambiguity, so we need to use the binary model when cutting the word, and cut the word according to the context. Only the more succinct the word sequence is, the more likely it is. The correct segmentation scheme.

To estimate the probability of \( w_1 \) occurrence after \( w_2 \), we usually use the definition of conditional probability, which is:

\[
P(w_2 | w_1) = \frac{P(w_1, w_2)}{P(w_1)}
\]

(1)

According to formula (1), you can get:
\[ P(w_1, w_2) = P(w_1)P(w_2 | w_1) \]  \hspace{1cm} (2)

The same reason:
\[ P(w_1, w_2, w_3) = P(w_1, w_2)P(w_3 | w_1, w_2) \]  \hspace{1cm} (3)

So, have:
\[ P(w_1, w_2, w_3) = P(w_1)P(w_2 | w_1)P(w_3 | w_1, w_2) \]  \hspace{1cm} (4)

More general form:
\[ P(S) = P(w_1, w_2, ..., w_n) = P(w_1)P(w_2 | w_1)P(w_3 | w_1, w_2)...P(w_n | w_1, w_2, ..., w_{n-1}) \]  \hspace{1cm} (5)

This is called the chain rule of probability. Among them, \( P(w_2 | w_1) \) indicates the probability of \( w_2 \) appear after \( w_1 \). But this formula has the problem of excessive parameter space and sparse data. It is based on our introduction of Markov's hypothesis: the appearance of a word is only determined by the limited one or several words that appear before it. If the simplification into a word depends only on a word in front of it, it is called a binary model (Rigram), that is:
\[ P(S) = P(w_1, w_2, ..., w_n) = P(w_1)P(w_2 | w_1)P(w_3 | w_1, w_2)...P(w_n | w_1, w_2, ..., w_{n-1}) \approx P(w_1)P(w_2 | w_1)P(w_3 | w_1, w_2)...P(w_n | w_{n-1}) \]  \hspace{1cm} (6)

2.3. Design of Information Retrieval Module
The function of the information retrieval module is to retrieve relevant answers from the knowledge base after analyzing the questions raised by the users. In this system, after the user enters the question, first search in the FAQ library to retrieve relevant questions and answers; if the corresponding problem is not found in the FAQ library, continue to search in the text library, through the similarity of sentences Sort, select the one with the highest similarity, and extract the answer. Faced with the FAQ library and the text library, different methods are used to extract the answers, so as to get the most accurate answer. The structure of the question and answer system implemented by the retrieval method is shown in the following figure 3.

2.3.1. Searching the FAQ Library
Because of the FAQ library stores the answers to the existing questions, after the user enters the question, the problem type is divided according to the problem type table. After the problem type is obtained, the same type of problem already exists in the FAQ library. Make up the candidate set and search for the standard question directly in the candidate set. Such as the user's question is: "What type of school is Shenyang Normal University?", can get the standard answer "National unified independent undergraduate colleges, belonging to the national public full-time institutions of higher learning." Because the type of problem has been ore-existing Classification, so when the user searches, the range of candidate sets formed by the problem type is relatively small, which improves the efficiency and correctness of the search. When searching for problems in the FAQ library, you need to set a threshold and use the improved vector space model algorithm to calculate the similarity between the problem of the candidate problem set and the user's question. If the similarity is greater than the threshold, then you want to The answer to the question to be searched is to sort the similarities in the
candidate problem set, and the answer corresponding to the question with the highest similarity is the answer that needs to be returned to the user.

### 2.3.2 Improved algorithm for space vector model

The traditional space vector model is more suitable for sentences containing more vocabulary. In this case, the advantages are obvious, while the advantages in other situations are not obvious. Therefore, when calculating the similarity of sentences, the system improves the traditional space vector model. When calculating the similarity of sentences, the similarity is weighted and summed from multiple aspects, so that two more accurately estimate the degree of similarity between sentences. Algorithm for calculating the similarity of sentences:

- **Morphological similarity:** Starting from the two aspects of the morphological structure and sentence structure of words, the similarity of their forms is calculated. The similarity formula of the question \( Q_1 \) and the question \( Q_2 \) is as follows:

\[
WordSim(Q_1, Q_2) = \frac{\text{SameWord}(Q_1, Q_2)}{\text{Word}(Q_1) + \text{Word}(Q_2) - \text{SameWord}(Q_1, Q_2)}
\]  

(7)

In the formula (7), \( \text{WordSim}(Q_1, Q_2) \) means the word form similarity with the two questions is expressed; \( \text{SameWord}(Q_1, Q_2) \) means the number of the same keywords contained in the sentence; \( \text{Word}(Q_2) \) means the number of keywords in the sentence \( Q_2 \).

- **Sentence similarity:** Start with the length of the sentence and calculate the similarity of the length of the sentence. The sentence length similarity formula of the question \( Q_1 \) and the question \( Q_2 \) is as follows:

\[
LengthSim(Q_1, Q_2) = 1 - \frac{\text{Length}(Q_1) - \text{Length}(Q_2)}{\text{Length}(Q_1) + \text{Length}(Q_2)}
\]  

(8)

In formula (8), \( \text{LengthSim}(Q_1, Q_2) \) means the sentence length similarity is expressed in the question, and \( \text{Length}(Q_2) \) means the length of the sentence in the sentence is indicated.

- **Word order similarity:** Starting from the context of the keywords, the similarity between the structure and word order of the words is calculated. It is calculated in the reverse order of the adjacent order of synonyms or synonyms of the two sentences. The sentence similarity formula of the question \( Q_1 \) and the question \( Q_2 \) is as follows:

\[
OrderSim(Q_1, Q_2) = 1 - \frac{\text{Re}(Q_1, Q_2)}{\text{Max} \text{Re}(Q_1, Q_2)}
\]  

(9)

In formula (9), \( \text{OrderSim}(Q_1, Q_2) \) means the word order similarity of the question, \( \text{Max} \text{Re}(Q_1, Q_2) \) means the maximum reverse number of the sequence of natural numbers obtained by sorting the number of the same keyword; \( \text{Re}(Q_1, Q_2) \) means the keyword corresponding to the question is corresponding to the inverse number of the sequence of natural numbers formed by the place.

- **Word Spacing Similarity:** Calculate the distance between the same keywords to mark the word spacing similarity of the sentences. The sentence spacing similarity formula of the question \( Q_1 \) and the question \( Q_2 \) is as follows:

\[
SpacingSim(Q_1, Q_2) = 1 - \frac{\text{SameSpa}(Q_1, Q_2)}{\text{Spacing}(Q_1) + \text{Spacing}(Q_2)}
\]  

(10)

In formula (10), \( \text{SpacingSim}(Q_1, Q_2) \) represents the word spacing similarity of the questions \( Q_1, Q_2 \); \( \text{SameSpa}(Q_1, Q_2) \) represents the interval of the same keyword in the question \( Q_2 \); \( \text{Spacing}(Q_1) \) represents the interval of the same keyword in the question \( Q_1 \); if the keyword repeatedly appears multiple times, take its maximum value; \( \text{Spacing}(Q_2) \) represents the interval of the same keyword in the question \( Q_2 \); if the keyword repeatedly appears multiple times, take its minimum value.

- **Sentence similarity:** The first four similarities are weighted and summed to obtain the similarity between the two questions. It is represented by a value between 0 and 1, 0 means not the same, 1
means exactly the same, the larger the value, the more similar the two sentences are. The sentence similarity formula of the question $Q_1$ and the question $Q_2$ is as follows:

$$Sim(Q_1, Q_2) = \lambda_1 \cdot WordSim(Q_1, Q_2) + \lambda_2 \cdot LengthSim(Q_1, Q_2) + \lambda_3 \cdot OrderSim(Q_1, Q_2) + \lambda_4 \cdot SpacingSim(Q_1, Q_2)$$ (11)

In the formula (11), $Sim(Q_1, Q_2)$ represents the sentence similarity of the questions $Q_1$, $Q_2$; where $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1$, and $\lambda_1 \geq 0.5 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 = 0$. After the test of the answer set of the FAQ, take $\lambda_1 = 0.6$, $\lambda_2 = 0.2$, $\lambda_3 = 0.1$, $\lambda_4 = 0.1$, and the selected threshold is 0.65. When searching for the problem, select the problem with the similarity greater than or equal to 0.65, and follow the similarity. Sort the order from largest to smallest, and select the answer with the highest similarity to return to the user.

2.3.3. Retrieval of text library

If you can't find the answer in the FAQ library, you should search in the text library. When searching for answers in the text library, you also need to make a candidate set. The general question and answer system retrieves information in the text library by searching for keywords, finding documents with keywords, and then classifying the problem types to select candidate sets, but this is a waste of time and accuracy is not high. Therefore, the system adopts a method of matching the type of the problem first to narrow the search range, and then searches the keyword to find a candidate set for document formation related to the keyword. This can save time and improve the similarity between the relevant documents in the candidate set and the problem. After selecting the candidate set, it is necessary to calculate the similarity between the document in the candidate set and the problem. The documents in the candidate set contain information about the problem, but they have different abilities to solve the problem. In order to find the most accurate answer, the similarity between the document and the problem needs to be calculated.

When calculating the similarity between the document and the question, the vector space model is used, and the TF-IDF algorithm is used to calculate, and the complex document processing process is moved into the vector space, which is converted into a vector operation, and the similarity of the sentence is converted into The similarity in the vector space makes the calculation simpler. The calculation method is as follows:

1) TF-IDF

All the basic language units in the document refer to the index item (term, denoted as T) of the document, and the index indicating the importance of the index item in the document is called the term weight (denoted as W). If a document contains n index entries, then the document can be represented as D (T1, T2, T3, ..., Tn). If the weight of item $kT$ is represented by $kW$, then the document can be written as D (T1, W1, T2, W2, T3, W3, ..., Tn, Wn), abbreviated as D = D (W1, W2, W3, ..., Wn). Then, using the TF-IDF formula, the importance degree $kT$ of the index item to the document $iD$ can be obtained, and the formula is as follows:

$$tf_i k * idf_k$$ (12)

In formula (12), $tf_i k$ represents the frequency at which index item $Ti$ appears in document $D$. $idf_k$ larger value means that $Ti$ is more important for document $D$; $idf_k$ (Inverse Document Frequency) indicates the inverse ratio of index items of the index item. This means that the difference in A document is greater. Where A and N are the total number of all documents, taking $N=|D|$, $m$ is the number of documents containing index item $Ti$, and $n$ is inversely proportional to $Ti$. When $n$ is higher, it means that index item $Ti$ is measuring document. The weight of the aspect of similarity is lower. When an index item $Ti$ appears only in one document, $idf_k = \log(N)$; when an index item $Ti$ appears in all documents, $idf_k = \log(1) = 0$. In order to prevent the occurrence of $idf_k$ being 0, a constant c which is not zero is often introduced to calculate $idf_k = \log(\frac{N}{n_k} + c)$.

In order to reduce the impact of document length on the calculation, the formula (13) is optimized to obtain:
When analyzing a document D (T₁, W₁, T₂, W₂, T₃, W₃, ..., Tₙ, Wₙ), in order to reduce the difficulty of analyzing Ti’s multiple occurrences in the document and its context, the space vector model can be used. Simplify it. The order of Ti in the document is ignored first, so that T₁, T₂, T₃, ..., Tₙ can be simplified into a one-dimensional coordinate system, and W₁, W₂, W₃, ..., Wₙ can be seen as its corresponding coordinate value, so D(W₁, W₂, W₃, ..., Wₙ) can be represented as a vector of n-dimensional space.

- The degree of similarity of sentences
  After converting the document into a vector space model, the similarity can be calculated by calculating the cosine of the angle between the vectors. The formula is:

\[
\text{Sim} (D₁, D₂) = \frac{\sum_{k=1}^{n} W_{1k} * W_{2k}}{\sqrt{\sum_{k=1}^{n} W_{1k}^2} \times \sqrt{\sum_{k=1}^{n} W_{2k}^2}}
\]

(14)

In formula (14), Sim (D₁, D₂) is the cosine of the angle between the two vectors; the keyword in statement D₁ is represented as the vector D₁=(W₁, W₂, W₃, ..., Wₙ), where n represents the number of keywords. The frequency in which the word appears in the document is taken as the value of each component in the vector; the keyword in the statement D₂ is represented as D₂=(W₁, W₂, W₃, ..., Wₙ). According to the formula, the similarity between the question and the documents is calculated, and the descending order is sorted, and the top five documents are taken as the search results.

3. System testing
In order to test whether the system can run accurately, this paper tests some of the main modules of the system.

3.1. Registration and Login Module Testing
- The user name of the pre-registered user is: 16570011, password: 16570011
  - Enter the user name: 16570001, password: 16570001, and the dialog box "Login failed, user name or password is incorrect, or your account is not audited" will pop up.
  - Enter the user name: 16570001, password: 16570011, and the dialog box "Login failed, user name or password is incorrect, or your account is not audited" will pop up;

3.2. Problem accuracy and response time test

3.2.1. Accuracy rate.
In order to test the accuracy of the system, in the school's admissions brochure, select some key terms, and combine them in the form of question words + key names, used to test the accuracy of the system, the combination and correct rate are shown in Table 1. Show. The system retrieval results are compared with the general retrieval method and other question and answer systems, and the results are shown in Figures 4 and 5.
Through the test and comparison of the simple questions in the field of enrollment consultation, it is possible to determine the question of the type of the defined problem. The accuracy of the system can basically reach about 90%, indicating that the accuracy of the system can be guaranteed.

3.2.2. Response time.

We selected 100 questions for enrollment consultation in our school to test more. The response time of the system is less than 1 second. The test results show that the system is running at high efficiency and can quickly respond to user questions. In summary, the intelligent question answering system for college admissions in this paper performs well and has certain practicability.

At present, the intelligent question answering system is more and more widely used in user consultation, but most of them are aimed at popular research fields. In some unpopular fields, using the intelligent question answering system to answer questions is still a gap. With future research, professional The intelligent question and answer system in the field can also be used. This is a thing that can save more costs, reduce human resource consumption, and have a positive use for society. In this article, there are two shortcomings:

- Because it is a system oriented to a limited field, its information is relatively concentrated, so the algorithm used in this paper is relatively small, and you can explore a wider field in the future and try more new methods.
Although the system proved to be real and effective in the actual measurement, there is room for further improvement for the intelligent question answering system.

In summary, with the continuous improvement of technology in the future, the intelligent question answering system will be better developed, and our system will be continuously improved and improved to adapt to the new environment.

4. Conclusion
At present, the intelligent question answering system is more and more widely used in user consultation, but most of them are aimed at popular research fields. In some unpopular fields, using the intelligent question answering system to answer questions is still a gap. With the development of research, professional intelligent question answering system can also be applied in this field. This is a thing that can save more costs, reduce human resource consumption, and have a positive use for society. In this article, there are two shortcomings:

- Because it is a system oriented to a limited field, its information is relatively concentrated, so the algorithm used in this paper is relatively small, and you can explore a wider field in the future and try more new methods.
- Although the system proved to be real and effective in the actual measurement, there is room for further improvement for the intelligent question answering system.

In summary, with the continuous improvement of technology in the future, the intelligent question answering system will be better developed, and our system will be continuously improved and improved to adapt to the new environment.

References
[1] Morceau V, Sanjuana E, Tinnier X, et al. Overview of the 2009 QA Track: Towards a Common Task for QA, Focused IR and Automatic Summarization Systems[C]// International Conference on Focused Retrieval & Evaluation. 2009.
[2] Hoyt R E, Snider D, Thompson C, et al. IBM Watson Analytics: Automating Visualization, Descriptive, and Predictive Statistics[J]. Jmir Public Health Surveil, 2016, 2(2): e157.
[3] Hoyt R E, Snider D, Thompson C, et al. IBM Watson Analytics: Automating Visualization, Descriptive, and Predictive Statistics[J]. Jmir Public Health Surveil, 2016, 2(2): e157.
[4] Feng G, Du Z, Xing W. A Chinese Question Answering System in Medical Domain[J]. Journal of Shanghai Jiaotong University (Science), 2018, 23(05):88-93.
[5] Ho H, Mawardi V C, Dharmawan A B. Question answering system with Hidden Markov Model speech recognition[C]// International Conference on Science in Information Technology. 2018.
[6] Garg S, Kumar S. JOSN: JAVA oriented question-answering system combining semantic web and natural language processing techniques[C]// India International Conference on Information Processing. 2017.
[7] Yao Q Y, Liu G S, Xiang L I. VSM-based Text Clustering Algorithm[J]. Computer Engineering, 2008, 34(18):39-41.

[8] Saquete E, Martínezbarco P, Muñoz R, et al. Splitting Complex Temporal Questions for Question Answering Systems. [C]// Meeting on Association for Computational Linguistics. 2004.

[9] Mitok R. Oxford Handbook of Computational Linguistics[J]. Literary & Linguistic Computing, 2005, volume 18(3):333-336(4).

[10] Yu L, Hermann K M, Blossom P, et al. Deep Learning for Answer Sentence Selection[J]. Computer Science, 2014.

[11] Pota M, Esposito M, Pietro G D. Learning to rank answers to closed-domain questions by using fuzzy logic[C]// IEEE International Conference on Fuzzy Systems. 2017.

[12] Schwoebel M, Puffiest W, Pedersen R U, et al. Worst-case execution time analysis for a Java processor[J]. Software: Practice and Experience, 2010, 40(6):507-542.

[13] Figueroa A, Neumann G. Category-specific models for ranking effective paraphrases in community Question Answering[J]. Expert Systems with Applications, 2014, 41(10):4730-4742.

[14] Searle J R. The Turing Test: 55 Years Later[M]// Parsing the Turing Test. 2009.

[15] Qu Z, Song X, Zheng S, et al. Improved Bayes Method Based on TF-IDF Feature and Grade Factor Feature for Chinese