A Study on the Classification of Chinese Interrogatives Based on Pragmatic Functions

Li Jiangtao[(✉)] and Rao Gaoqi

Research Institute of International Chinese Language Education, Beijing Language and Culture University, Beijing, China
`raogaoqi@blcu.edu.cn`

Abstract. This paper emphasizes the importance of regular questions in question classification, analyzes and summarizes four types of interrogative sentences and their corresponding pragmatic functions. On this basis, we have manually annotated interrogative sentences in selected novels and BaiduQA and have created an interrogative sentences corpus. A formal feature set for question classification is constructed based on corpus, and classification experiments based on rules and statistics are conducted respectively to check the classification effect of each feature. The set of feature rules is formed by optimizing feature combinations iteratively through multiple rounds of experiments, which provides a formal classification basis for the existing Question-answering systems. Through the comparison of various classification models, it is found that the method of finite state automaton can achieve a macro average F1-score of 0.94, whereas the random forest model achieving a macro average F1-score of 0.98, has better performance in classification.

1 Literature Review

Question-answering systems are generally composed of modules performing the following tasks: question classification, query expansion, search engine, answer extraction and answer sorting and selection [1]. Question classification is an important sub-module of a question answering system. Its main task is to classify questions into corresponding semantic categories according to the types of answers, producing a good guiding effect on subsequent answer extraction and answer selection modules of the question answering system [2]. The effectiveness of question classification directly affects the understanding of questions.

There is no unified standard for the current question classification systems. The most authoritative question classification system so far is UIUC, which is a hierarchical classification system based on answer types. In this system, questions are divided into 6 major categories and 50 minor categories, and each major category contains non-repeating minor categories [3, 4]. Previously, many Chinese question classification systems were developed using relevant methods adopted for English question classification. However, the process of Chinese question classification is more complicated than that of English question classification. Because of such discrepancy, Chinese question classification has developed on its own path.

Traditional Chinese question classification is developed mainly based on the types of things asked and answered about, such as people, places, time, numbers, etc. Cao [5] added classification of question phrases, standard question types and feature word segmentation to enhance the computer capability to identify questions. Liu [6] further mapped interrogative word patterns to question types and conducted a question understanding study based on interrogative sentence pattern recognition. In those tasks, the forms of the questions are only used as an auxiliary feature for classification. Yet, in fact, the categories of certain question forms can correspond to certain question functions. Such relationship between forms and function has not been paid much attention to in the studies on question understanding practices.

With the increase of datasets and the more extensive coverage of questions, features of complex question forms have been observed as patches to solve new problems. Yet, in this way, question classification standards become more and more complex. If questions can be formally classified first and then further classified according to the corresponding question functions expressed by different question forms, all questions can be formally covered, and at the same time the specific function information of questions can be obtained in the classification process. Therefore, it is of profound significance to advocate the formal classification of questions on the basis of the existing research (Fig. 1).
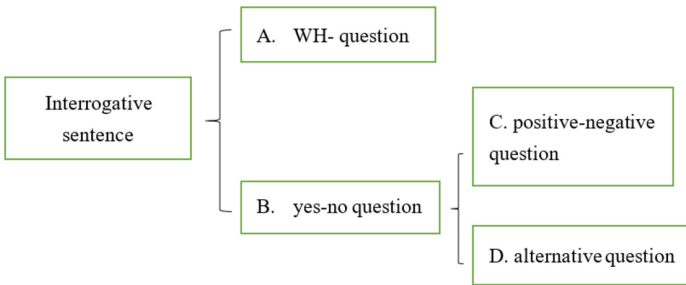

Fig. 1. Lv's question classification system [7]

Lv [7] put forward a "derivation system-based" classification method according to the derivation relation of subclasses of interrogative sentences. He suggested that WH-questions and yes-no questions were two basic types, while positive-negative questions and alternative questions were derived from yes-no questions, and both positive-negative and alternative questions could construct yes-no questions. For example:

你去? (positive) 你不去? (negative) → 你去不去? (yes-no question).
[Ni qu? ni bu qu? → ni qu bu qu?].
'Are you going? You're not going? → Will you go?'

In fact, Lv's system merely reflected the current state of question comprehension. So far, analyses of questions mostly focused on WH-questions. However, yes-no questions have more complex and various formal characteristics compared with WH-questions. The language information to answer yes-no questions thus involves more details, which

are rarely mentioned in question comprehension research. If we can distinguish the four different types of questions and classify them formally at the initial stage of question recognition, then research on question recognition and understanding will be more on target, and the subsequent semantic classification can also be carried out under such formal classification system.

Section 2 will start from identifying words that related to interrogatives and the specific sentence patterns of questions, without involving the semantic understanding of questions. On this basis, this article emphasizes that forms should have a higher priority in question classification. Section 3 introduces the construction of corpus of interrogative and the research on question classification based on formal features is to be carried out in Sect. 4. Section 5 concludes this paper.

2 Classification Features Based on Language Forms

At present, there is a question classification system for academic use, which is established based on answer type and semantic information of questions. This classification is oriented towards problem solving and is beneficial to understanding and answering questions. While the system classifies questions according to their semantics, it also uses formal features as reference for classification. In the existing classification methods, formal features are used as supplementary features. However, this kind of processing methods often makes the classification system imperfect since it cannot give full play to the screening and classification functions of formal features, and it cannot cover all questions. Besides, different forms of questions have different pragmatic functions, and the corresponding answers are also different. For example, the four questions of "Is the weather hot or not?", "Is it hot?", "Is the weather hot or not hot?" and "How is the weather?" all belong to the weather category in terms of answer types, but they belong to positive-negative questions, yes-no questions, alternative questions and WH-questions in linguistic classification. When a conversation participant is to answer the above questions, s/he needs to pay attention to the differences in the focus of the questions and also to the amount of information needed. If we only look at the above four questions, in Chinese, the answers could be the same, i.e. 'hot', despite the degree of naturalness in the answers. This drives us to give the formal classification of questions a higher priority before considering other classifications of questions.

In question-and-answer systems developed in the past, the questions that can usually be identified and answered mostly are WH-questions. However, there are three more types of questions if classified in terms of form, namely yes-no questions, alternative questions and positive-negative questions. If the coverage of question recognition is to be expanded, these three types of questions must also be included in the research objective [6]. The categories of questions are defined as follows:

Yes-no questions. The structure of this kind of questions is similar to that of declarative sentences, in the sense that the question is formed with a declarative sentence which sometimes ends with an interrogative particle "吗[ma]", "啊[a] or 哇[wa]" (but not the particle "呢[ne]"). Even if sometimes the sentence does not end with these particles, the sentence is uttered with a rising intonation (while declarative sentences are not uttered with such intonation).. There is usually a "?" at the end of the sentence. For example:

21 世纪人类将要开发月球吗?

[er shi yi shi ji ren lei jiang yao kai fa yue qiu ma?].

Will mankind explore the moon in the 21st century?

WH-questions. Interrogative pronouns are used in-situ to replace the questioned element. Commonly used interrogative pronouns are "谁[*shui*](*who*), 什么[*shen me*](*what*), 哪儿[*na er*](where), 怎么[*zen me*](*how*), 多少[*duo shao*](*how many*)" and so on. The question is ended with "呢[*ne*]" or "啊[*a*]" (but not "吗[*ma*]"). For example:

你有什么要求呢?

[ni you shen me yao qiu ne?].

What request do you have?

Alternative questions. There are several parallel clauses in alternative questions. Two clauses are often connected with "是[*shi*]" and "还是[*hai shi*]". Sometimes the mood particle "呢[*ne*]" or "啊[*a*]" is used, but "吗[*ma*]" is not used. In addition, alternative questions contain both the mood particles and the conjunctions. For example:

他是去了北京还是去了天津啊?

[ta shi qu le bei jing hai shi qu le tian jin a?].

Did he go to Beijing or Tianjin?

Positive-negative questions. It usually contains a negative word, e.g. "不[*bu*]" or "没有[*mei you*]", and does not take the form of complex sentences. The affirmation and negation of the verb or adjective are coordinated in the question and the focus of the question can be the verb/adjective or the complement [8]. The patterns and the examples are shown in the following table (Table 1):

Table 1. Patterns and examples of positive-negative questions

Pattern	Example
V/Adj + 不[*bu*] + V/Adj	你饿不饿? (*Are you hungry?*)
V + 不[*bu*] + V + X	你吃不吃饭? (*Do you eat or not?*)
V + X + 不 + V	你吃饭不吃? (*Do you eat or not?*)
V + X + 不[*bu*] + V + X	你吃饭不吃饭? (*Do you eat or not?*)
V + 不[*bu*]/没有[*mei you*]/否[*fou*]	你吃饭不? (*Do you eat?*)
V + C + V + Neg + C	这饭你吃得了吃不了? (*Can you eat this meal?*)

The choice of question features depends on the form of question classification. Lin [9] once put forward five formal markers of questions: interrogative pronouns, "是[*shi*]…还是[*hai shi*]…" (the form of alternative questions) and "X不[*bu*]X" (the form of positive-negative questions), modal particles and sentence intonation. However, Huang and Liao

[10] believed that questions can take the form of intonation, interrogative words, modal adverbs and interrogative forms. Li [11] suggested that interrogative markers include interrogative intonation, interrogative modal particles, interrogative pronouns of WH-questions and interrogative formats. He also divided interrogative markers into upper, middle and lower levels according to the distribution of interrogative markers. It was explicated that the upper level included interrogative intonation, the middle level modal particles e.g. "吗[*ma*], 吧[*ba*]" and "呢[*ne*]", and the lower level interrogative pronouns of WH-questions and interrogative formats.

Sentence intonation is a phonological issue. The intonation of questions is mostly marked by a question at the end of the interrogative. The data used in this article all contain the question mark "?", and so sentence intonation is not taken as an interrogative feature.

Instead, what could be taken as significant interrogative feature include i) mood particles, ii) interrogative pronouns, iii) interrogative format and iv) modal adverbs. To be specific: a particle "呢[*ne*]" can distinguish yes-no questions from non-yes-no questions and questions ending with particles such as "吧[*ba*], 吗[*ma*], 么[*me*]" must be yes-no questions. So mood particles can be listed as a feature of questions. Interrogative pronouns are the most prominent markers of WH-questions. All WH-questions contain interrogative pronouns. So interrogative pronouns can also be listed as an interrogative feature. In addition, each question category also has its own interrogative formats, which belong to the lower level in the classification system of Li [11], and happen to be the feature type that carries the most question information. Questions containing specific interrogative formats can often directly and accurately locate their question classification. Therefore, we regard the interrogative formats of yes-no questions, WH-questions, alternative questions, and positive-negative questions as a question feature. As for modal adverbs, they do not have obvious correlation with question classification. Yet, considering that redundant features can be filtered in the later stage with it, modal adverbs are also taken as a question feature here. These four features will be adopted in this paper for question classification.

3   Construction of Interrogative Sentence Corpus

3.1   Corpus Construction

In this paper, question marks are used as markers to search for questions in corpora. As a result, 4300 questions were randomly selected from a batch of modern novel corpora and a BaiduZhidao's question data set. The datasets are then used to constitute the corpus of interrogative sentences, which will be open to academia. The questions were annotated with question categories manually. The annotating standard of questions mainly follows Huang and Liao's definitions of questions.

'YN' refers to yes-no questions; 'WH' refers to WH- questions; 'AL' refers to alternative questions; 'PN' refers to positive-negative questions; 'BD-data' refers to the dataset from the website "BaiduZhidao".

As can be seen in Table 2, the amount and proportions of WH questions are the highest across different datasets, followed by yes-no questions, positive-negative questions and alternative questions, which to some extent reflects the natural distribution of these

Table 2. Datasets of the interrogative sentences

	Novels		BD-data		Total	
	Count	Percent	Count	Percent	Sum	Percent
YN	651	38.8%	527	20.1%	1178	27.4%
WH	749	44.6%	1857	70.9%	2606	60.6%
AL	23	1.4%	40	1.5%	63	1.5%
PN	256	15.2%	197	7.5%	453	10.5%
Sum	1679		2621		4300	

four types of questions in the language. The proportion of alternative questions and positive-negative questions is not high both in novels and in Baidu. This also shows that WH-questions and yes-no questions are more likely to be selected when people ask questions.

Despite the similarities mentioned above, the distributions of the four types of questions are slightly different across the two data sets. In the novels, the proportion of WH-questions is slightly higher than that of yes-no questions. However, in BD-data, WH-questions account for more than 70%, far more than the yes-no questions which account for 20.1%. To some extent, this shows that the distributional characteristics of questions in novels and in BD-data are different despite a small degree of similarity. BD-data is an encyclopedic question dataset, with a large proportion of questions about concepts, causes of events, etc. Therefore, there are more questions containing interrogative pronouns, which explains why more WH-questions are distributed in BD-data. However, there is no such obvious tendency in novels, which makes the distribution of yes-no questions and WH-questions more balanced.

3.2   Feature Selection and Question Feature Set Construction

Syntactic formats and interrogative markers play a major role in question classification. According to the linguistic definitions of yes-no questions, WH-questions, alternative questions and positive- negative questions, the syntactic formats and interrogative markers can be further divided into four categories: interrogative formats, modal particles, modal adverbs and interrogative pronouns. These four categories can be further divided into seven sub-categories according to the actual corpus, namely, i) modal particle "呢[*ne*]", ii) modal particle "吗[*ma*]" and other particles, iii) interrogative pronouns, iv) modal adverbs, v) yes-no question format, vi) positive-negative question format and vii) alternative question format.

One thing worth noting regarding yes-no questions is that some sentences have too few explicit question markers and do not contain any of the features of the seven sub-categories, such as "她走了?(*Has she gone?*)" So, in order to avoid the situations where there is no feature matching for a yes-no question, we need to add a supplementary feature. To elaborate, when there is no interrogative pronoun and the question is not in

the positive-negative question format or the alternative question format, the sentence has the supplementary feature by default, and otherwise there is no such feature.

We use F1-F8 to stand for the above eight types of features. In the corpus, the quantitative statistics of these formal features are as follows:

Table 3. Distribution of question features

Feature	Category	Count/Percent	Explanation
F1	Modal particle	253/5.88%	"呢[*ne*]"
F2	Modal particles	802/18.65%	"吗[*ma*], 么[*me*], 吧[*ba*]", etc
F3	Interrogative pronouns	2790/64.88%	什么(*what*), 啥(*what*), 如何(*how*), 哪(*where*), 几(*how* many), 谁(*who*), 为啥(*why*), 为何(*why*), 为什么(*why*), 咋(*how/why*), 干吗(*why*), 多X(*how* + x), 怎么(*how*), etc
F4	Interrogative format	1292/30.05%	Yes-no questions: modal verb + modal particle
F5	Interrogative format	66/1.53%	Alternative questions: X还是[*hai shi*]X
F6	Interrogative formats	479/11.14%	Positive-negative questions: X 不[*bu*]X, X不[*bu*], X没有[*mei you*], X不成[*bu cheng*], etc
F7	Modal adverbs	96/2.23%	莫非[*mo fei*], 莫不是[*mo bu shi*], 难道[*nan dao*], 难不成[*nan bu cheng*], 到底[*dao di*], 何必[*he bi*], 究竟[*jiu jing*], 岂[*qi*], etc
F8	Supplementary feature	1053/24.49%	supplementary features of F3, F5, F6

It can be seen from the above table that the distribution of features is related to the distribution of different types of questions, and some feature distributions can even directly reflect the overall distribution of questions. For example, the proportions of F3, F4, F5 and F6 are the distribution of four types of questions in the dataset, reflecting that WH-questions and yes-no questions account for a larger proportion of questions, and the number of alternative questions is less than that of positive-negative questions. On the other hand, the total sum of the proportions of these interrogative forms and interrogative pronouns is greater than 100%, which indicates that the result of question classification is not determined only by interrogative forms. That means some questions contain multiple interrogative forms or interrogative pronouns. The complexity of question classification is also reflected by this.

4   Automatic Classification Based on Question Forms

4.1   Finite State Automaton Based on Formal Feature Sets

The contributions of different question features to question classification are different. According to Table 3, we can regard the coverage rate of features as the degree of contribution of question features to question classification. They are in the following order: modal words (吗[*ma*], 么[*me*], 吧[*ba*]) = interrogative formats > interrogative pronouns > others (question adverbs, etc.). Then, based on the contribution ranking of these question features, we can let question features with large contribution values participate in question judgment preferentially, and questions that cannot be covered by question features can be classified as yes-no questions. In this way, the classification of question is carried out within such limited rule. As long as a question is input, the category to which the question belongs can definitely be output. Then the preparation for the construction of a finite state automaton based on formal features is complete.

4.2   Multi-feature Classification Based on Statistical Machine Learning

Based on the question features in Table 3, we carry out feature vectorization on the questions in the corpus. The vector coordinate of the dimension containing the specified feature is marked as 1. Otherwise the vector coordinate is marked as 0. We use 1, 2, 3 and 4 to mark the classification of questions. An example is shown in Table 4.

Table 4.   An example of feature transformation

例:《亮剑》中李云龙求婚对白是第几集啊? e.g. In *which episode of DRAWING SWORD did Li Yunlong propose?*								
F1	F2	F3	F4	F5	F6	F7	F8	CLASS
0	0	1	0	0	0	0	0	2

After obtaining the multi-dimensional vector and its corresponding classification label, the task of classifying questions according to the feature distribution has essentially started. We propose to use six machine learning methods, namely support vector machine, linear classifier, Bayesian classifier, K nearest neighbor, decision tree and random forest to verify the classification effectiveness of question features according to the experience of previous classification tasks.

In addition, the number of selected features will also affect the result of question classification. Features such as F1 to F8 are similar to listing question forms from a linguistic perspective. However, further experiments are needed to prove which combination of features can achieve the best question classification results. Therefore, this paper arranges and combines F1 to F8, totaling 225 combination results.

We used 1679 manually annotated questions in novels as a training corpus, and 2621 subsequently annotated BD-data as a test corpus. After combining machine learning methods with feature combination results, we analyze the classification performance of the model from multiple perspectives in following sub-section.

4.3   Experiments and Analysis

4.3.1   Analysis of Experimental Results and Features

Since there are four types of questions, we mainly analyze the changes of models with the number of features from a macro perspective. That is, we analyze the overall advantages and disadvantages of question classification through the macro-average and micro-average of F1-score of different model classifications. While a certain amount of features involved, different combinations of the features can affect the accuracy of classification results. Consider this, we only select the best result for the certain number of features for comparison.
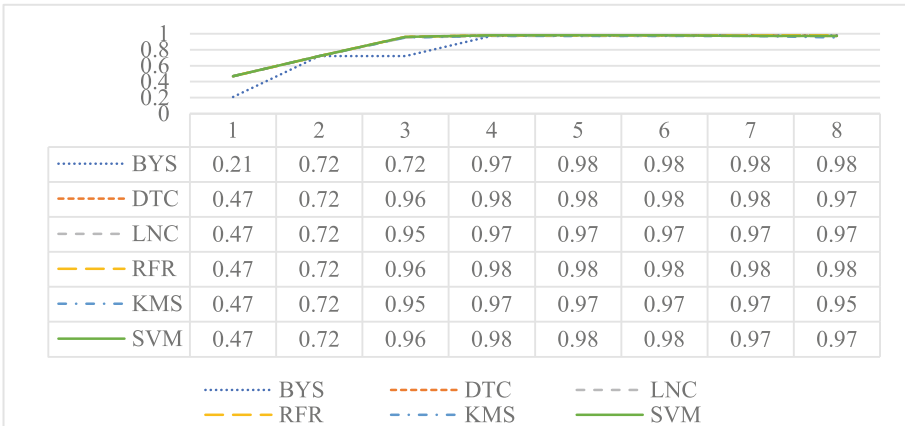
	1	2	3	4	5	6	7	8
·········· BYS	0.21	0.72	0.72	0.97	0.98	0.98	0.98	0.98
------- DTC	0.47	0.72	0.96	0.98	0.98	0.98	0.98	0.97
- - - - LNC	0.47	0.72	0.95	0.97	0.97	0.97	0.97	0.97
— — - RFR	0.47	0.72	0.96	0.98	0.98	0.98	0.98	0.98
- · - · - KMS	0.47	0.72	0.95	0.97	0.97	0.97	0.97	0.95
——— SVM	0.47	0.72	0.96	0.98	0.98	0.98	0.97	0.97

Fig. 2. Macro average of F1-score of each model and the number of features used

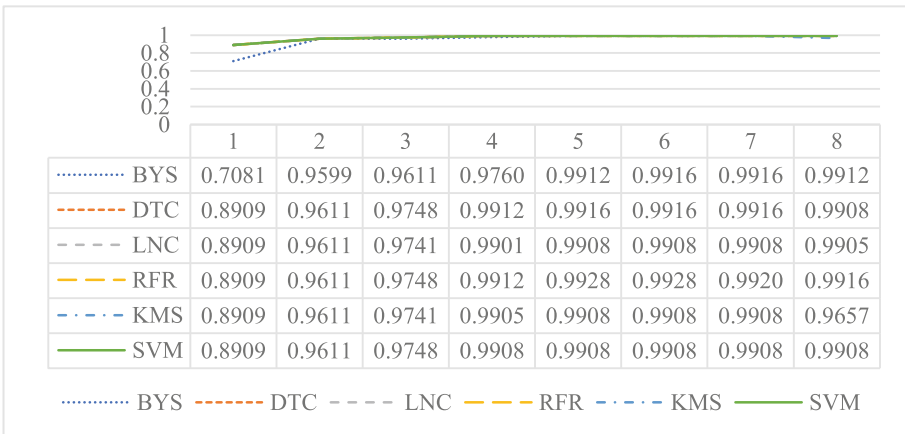|	1	2	3	4	5	6	7	8
·········· BYS	0.7081	0.9599	0.9611	0.9760	0.9912	0.9916	0.9916	0.9912
------- DTC	0.8909	0.9611	0.9748	0.9912	0.9916	0.9916	0.9916	0.9908
- - - - LNC	0.8909	0.9611	0.9741	0.9901	0.9908	0.9908	0.9908	0.9905
— — - RFR	0.8909	0.9611	0.9748	0.9912	0.9928	0.9928	0.9920	0.9916
- · - · - KMS	0.8909	0.9611	0.9741	0.9905	0.9908	0.9908	0.9908	0.9657
——— SVM	0.8909	0.9611	0.9748	0.9908	0.9908	0.9908	0.9908	0.9908

Fig. 3. Micro average of F1-score of each model and the number of features used

In Fig. 2 and 3, the classification effects of all models other than the Bayesian model are similar, and the graphs coincide in the figure. On the whole, the classification

performance of a model becomes better with the increase of features in the early stage, and the classification performance reaches the best when there are 5–7 features. But after that, for most models, the increase of features will make the classification performance worse. This shows that for classification of question forms, the increase in the amount of features has only a certain effect on the classification performance of the model. If new question features are to be added, their contribution to classification must be checked.

Examining macro-averages and micro-averages of F1-scores of the random forest model, it is concluded that the macro-averages and micro-averages of F1-scores of the classification model reach the highest scores of 0.99 and 0.98 respectively when the number of features used with the random forest model is 5.

Table 5. Random forest: classification effect and selected features

Number of features	Micro-average of F1-score	Macro average of F1-score	Features
1	0.8909	0.47	F2
2	0.9611	0.72	F2, F6
3	0.9748	0.96	F2, F6, F5
4	0.9912	0.98	F2, F6, F5, F3
5	0.9928	0.98	F2, F6, F5, F3, F4
6	0.9928	0.98	F2, F6, F5, F3, F4, F8
7	0.9920	0.98	F2, F6, F5, F3, F4, F8, F1
8	0.9916	0.98	F2, F6, F5, F3, F4, F8, F1, F7

Table 6. Random forest: the gain effect of the formal features of questions

Feature	Micro-average gain of F1-score	Macro average gain of F1-score	Intensity of features
F6	+ 0.0702	+ 0.25	Strong
F5	+ 0.0137	+ 0.24	Strong
F3	+ 0.0164	+ 0.02	Sub-strong
F4	+ 0.0016	0	Sub-strong
F8	0	0	Weak
F7	−0.0008	0	Weak
F1	−0.0004	0	Weak

According to Table 5, we can sort the features according to the gain generated by the classification model using the newly added different features. Firstly, the question forms are divided into strong formal features and weak formal features. Strong formal

features are those formal features of questions that are beneficial to the classification of questions, otherwise they are weak formal features. Results are shown in Table 6.

Therefore, we can further divide the eight features of F1–F8 into strong formal features, less-strong formal features and weak formal features. Strong formal features include modal particles (F2), positive-negative question format (F6), and alternative question format (F5); less-strong formal features include interrogative pronouns (F3) and yes-no question format (F4); weak formal features include supplementary features (F8), modal adverbs (F7) and the modal particle "呢[*ne*]" (F1). The greater the strength of the feature, the greater its contribution to question classification. At the same time, this result can be compared with Li's question hierarchy theory mentioned in Sect. 2. Li's interrogative markers correspond to the formal features of questions mentioned in this paper. Li's hierarchy division is based on the distribution range of interrogative markers, so it can be concluded that there is no absolute correlation between the strength of formal features and their distribution range.

4.3.2   A Comparison Between Random Forest Model and Finite State Automaton

With BD-data as the test set, the results of classification performance using finite state automaton and random forest model are obtained, as shown in Table 7:

Table 7. Classification results of random forest model and finite state automaton

Method		Precision	Recall	F1-score
Finite state automaton	Yes-no questions	0.96	0.87	0.91
	WH-questions	0.95	0.99	0.97
	Alternative questions	0.93	0.95	0.94
	Positive-negative questions	0.99	0.88	0.93
	Mac-avg of F1-score	0.94		
	Mic-avg of F1-score	0.96		
Random forest model	Yes-no questions	0.99	0.99	0.99
	WH-questions	1.00	0.99	1.00
	Alternative questions	0.97	0.95	0.96
	Positive-negative questions	0.96	0.99	0.98
	Mac-avg of F1-score	0.98		
	Mic-avg of F1-score	0.99		

Examining the overall performance of the model, it is observed that the macro-average and micro-average of F1-score of random forest are 0.04 and 0.03, which are higher than those of finite state automaton. This shows that the method of finite state automaton classification also has good performance in question classification, and most questions can be effectively covered by specific question rules. However, this method often has a low recall rate and cannot deal with sentences containing some combinations

of features. Therefore, this also suggests that the random forest model actually has better classification performance in question classification.

Among the classification results of individual question forms, the F1-score of WH-questions is the best. Yet, the F1-score of finite state automaton is much lower than that of the random forest for yes-no questions and positive-negative questions. This reflects the obvious diversity of formal features for identifying yes-no questions and positive-negative questions, and a single formal feature is inadequate to cover most of these questions. As for positive-negative questions, the method of finite state automaton has higher accuracy than the random forest model, which reveals that the formal features play a strong role in identifying positive-negative questions. Yet, the recall rate is lower than that of the random forest model, which reflects the diversity of formal features of positive-negative questions.

5 Conclusions and Prospects

In this paper, the role of question forms in question classification is analyzed in detail. A question corpus has been then constructed, and the distribution of the corpus has been counted. It is noted that the number and proportion of WH-questions are the highest in the data set, followed by yes-no questions, positive-negative questions and alternative questions. This reflects the natural distribution of the four types of questions to a certain extent. Finally, according to the gain generated by the question features in classification performance, the question formal features are divided into strong formal features, less-strong formal features and weak formal features. It is concluded that the question classification can achieve the best results without the weak formal features. The classification results will provide an analytical basis for semantic classification and semantic understanding of interrogatives. In addition, a question classification model is constructed using a machine learning method based on the formal features of questions. Automatic classification experiments show that when the formal feature set is modal particles "吗[*ma*], 么[*me*], 吧[*ba*]", yes-no question format, interrogative pronouns, alternative question format, and positive-negative question format, question classification has high accuracy. This indicates that the modal particles at the end of a sentence not only distinguishes questions, but also is a powerful feature for question classification.

Classification of question forms is a problem with clear characteristics and strong rules. Using a rule-based system can also achieve good results in question classification. Therefore, we believe that a question form classification interface can be added when classifying questions. On the one hand, the accuracy of automatic classification based on question forms is guaranteed; on the other hand, all questions can uniformly be dealt with through the question form classification interface, which provides a basis for further classification of questions.

References

1. Wen, X., Zhang, Y., Liu, T., Ma, J.S.: Syntactic structure parsing based Chinese question classification. Chinese Inf. Process. (02) (2006). (in Chinese)

2. Liu, C.T.: Study on Sentence Type Theory and Its Applications in Chinese Question Answer System. Chongqing University (2010). (in Chinese)
3. Zhen, L.H., Wang, X.L., Yang, S.C.: Overview on question classification in question-answering system. Anhui University of Technology (Nature. Sci) (01) (2015). (in Chinese)
4. Niu, Y.Q., Chen, J.J., Duan, L.G., Zhang, W.: Study on classification features of Chinese interrogatives. Comput. Appl. Softw. (03) (2012). (in Chinese)
5. Cao, Z.J., Li, Z.S., Liu, C.T.: Study of question analysis in question-answering system. Comput. Sci. (11) (2005). (in Chinese)
6. Liu, C.T., Li, Z.S.: Research on question analysis based on sentence type recognition of interrogative sentences. Comput. Sci. (12) (2008). (in Chinese)
7. Lv, S.X.: Question, negation and affirmation. Zhongguo Yuwen (04) (1985). (in Chinese)
8. Yuan, Y.L.: Positive and Negative Questions and Related Typological References. Beijing Language and Culture University Press, Beijing (1994). (in Chinese)
9. Lin, Y.W.: Vocabulary, Grammar and Rhetoric. Shanghai Education Press, Shanghai (1985). (in Chinese)
10. Huang, B.R., Liao, X.D.: Modern Chinese. Higher education Press, Beijing (2017). (in Chinese)
11. Li, Y.M.: Reuse of question markers and decay of marker function. Zhongguo Yuwen (02) (1997). (in Chinese)