An Answer Extraction Method Based on Discourse Structure and Rank Learning

Huanyun ZONG Zhengtao YU Jianyi GUO Yantuan XIAN
School of Information Engineering and Automation
Kunming University of Science and Technology
Kunming,China
ztyu@hotmail.com

Jian LI
School of Computer
Beijing University of Posts and Telecommunications
Beijing,China
lijian@bupt.edu.cn

Abstract—**For the complex questions of Chinese question answering system such as 'why', 'how' these non-factoid questions, we proposed an answer extraction method using discourse structures features and ranking algorithm. This method takes the judge problem of answers relevance as learning to rank answers. First, the method analyses questions to generate the query string, and then uses rhetorical structure theory and the natural language processing technology of vocabulary, syntax, semantic analysis to analyze the retrieved documents, so as to determine the inherent relationship between paragraphs or sentences and generate the answer candidate paragraphs or sentences. Thirdly, construct the answer ranking model, extract five group features of similarity features, density and frequency features, translation features, discourse structure features and external knowledge features to train ranking model. Finally, re-ranking the answers with the training model and find the optimal answers. Experiments show that the proposed method can effectively improve the accuracy and quality of non-factoid answers.**

Keywords: complex questions; discourse structure; learning to rank; answer extracting

1. INTRODUCTION

The study of Question-Answering system has become a hot issue in the past few years. A substantial amount of work has been done in dealing with factoid QA. However, little research has focused on QA models for non-factoid questions, such as 'how', 'why', or 'manner' complex questions. One reason for this is that the frequency of non-factoid questions posed to QA system is lower than that of other types of questions such as factoid questions [1]. However, these complex questions cannot be neglected: as input for a QA system, they comprise about 5 percent of all why-questions and 4 percent questions asked to QA are 'how' questions, it shows that this kind of questions do have relevance in QA applications[2].A second reason why this type of question has largely been disregarded until now is that the techniques that have proven to be successful in QA for factoid questions have been demonstrated to be not suitable for questions that expect an explanatory answer instead of a noun phrase.TREC2007 introduced ciqa, NTCIR-7 introduced complex question-answering system at the 2008, there were about 10 institutions participated the relevant evaluation around that time, and most of the algorithm is inherited from the factoid questions. For factoid questions with answers of entity property, the answers usually contained in a clause, the dependency parser extracts the part of speech, words or phrases in line with expected answer type by semantic analysis. However, the answers for kinds of 'how' and 'why' questions usually span several clauses, sentences or paragraphs, the answer information is far from keywords of questions. Furthermore, there may be a more associative relationship between sentences or paragraphs, it is evident that the traditional factoid extracting method is not suitable for these complex questions. For the extraction methods inherited from the factoid questions, the characteristics of complex questions leads to the highest ranked relevant answer is not ranked in the top-10. Currently, most of the machine learning process of answer extraction method takes the judging answer relevance as a binary classification problem, and extract the features of questions and answers to train classification model. However, the task of classifier is learning the instance relevant or irrelevant, which ignores the relative position or order relation between answers. For example, answer A is more relevant than answer B, and answer B is more relevant than answer C, then there must exist the order relation that answer A is much more relevant than C, the relative ranking of answers is more important than relevance. As a result, there will be a large room to improve on the training effects [3].

Therefore, extracting the answers of complex questions is a new research topic. Due to the uncertainty and openness of these questions, how to analyze the questions and their answers, fuse the key information of the relevant answers document, construct the ranking model applied to rank the answers, thus to have the most relevant standing in the front is a critical problem. We focus on the extracting method of non-factoid questions, which called complex questions, the linguistic knowledge, structural analysis and ranking strategy, will be used to extract the optimal answers.

978-1-61284-729-0/11/$26.00 ©2011 IEEE

2. DISCOURSE STRUCTURE AND ANSWERS EXTRACTING

2.1 Discourse structure analysis

Discourse usually refers to a series of segments or sentences which constituted an overall language unit with certain hierarchy. The basic constituent units of all levels in discourse include clauses, sentences, paragraphs, sections. The text is a comparatively complete language whole which composed of various units by using a combination of many relations [4].Usually only analyzed the hierarchical structure and semantic relationship between constituent parts of text discourse can determine the importance of paragraphs or sentences. Discourse structure includes two aspects: physical structure and logical structure. Discourse of physical structure is the basic elements of text (such as title, paragraphs, sentences, words and punctuation, etc) which determined by accurate position, and often has evident feature markers that is easy to identify.

Logical structure of the text refers to the logical relationship formed by the constituted article topic, structural levels, paragraphs, sentences and keywords of text ideological contents in the conceptual sense. The rhetoric structural theory (RST) is originally developed by Mann and Thompson[5] that commonly used in logical structure analysis of discourse. RST theory states that a variety of clauses in sections and chapters are not chaotically lay together in a heap, but various semantic relations existed between them, namely rhetoric structural relation. Most of these rhetorical structure relations have asymmetrical semantic features, which defined as Nucleus and Satellite relations. Sometimes these relationships with Conjunction taken as formal markers sometimes are completely implied. As a theory of textual structure, RST described and categorized large relations between sentences and that between paragraphs comprehensively. At present, the discourse structure theory has been widely used in automatic text generation, automatic summarization, text analysis, and machine translation field, etc.

2.2 Role of discourse structure in answer extracting

The answers of non-factoid (complex) questions usually across several clauses or paragraphs, the answer information may be far from question keywords, just extracting the answers from the text of sentence-level is not enough, we must be able to find the answers from the perspective of discourse analysis and generate the candidate answers sentence. For example, the following paragraph is the answer of 'Why cats sleep so much?' in TREC11 question sets. The query string is 'cats'+ 'sleeping'+ 'much'+ 'because', these query keywords both appeared in the text and very close to each other(in a sentence),but the text does not contain the answers, which leads to the redundancy and low accuracy of answers.

[text example 1]

Sleep is an important animal behavior it can not go without, which can relieve the fatigue of body and brain, prepares the power and energy for the next activity. The cats sleep so much, ***but*** actually easy to awake. This ***because*** the cats' sleep is divided into two stages of deep sleeps and shallow sleep. The muscle is relaxed in deep sleep state, poor notice in response to the environment sound. It continues for about 6-7 minutes, the following is the 20-30 minutes stage of shallow sleep. At this time, the cats in light sleep and easy to awake. ***Since*** the cats' deep sleep and shallow sleep appear alternately, ***So*** the cats are alert to sleep. If we can correctly segment the sentence:

[cats sleep so much]| turning point association[but easy to awake]causal connection[because the cats' sleep is divided into two stages of deep sleep and shallow sleep.]

It could be found that the sentence composed of the three clauses, the second clause and the third clause constitute a causal connection, the nucleus is clause 3, the turning point association formed between the first clause and span[2,3] and the nucleus is span[2,3].The content is inconsistent with the turning point association, therefore the third clause couldn't explain the content of first clause. It can be concluded that it is not the answer we are looking for. Consequently, we coped with the text that contained answers information by rhetoric structural analysis, which effectively exclude redundant information and find the correct answers. Using the linguistic knowledge and structural analysis plays a very important role and will be helpful for extracting complex answers such as non-factoid questions.

For the characteristics of non-factoid questions, combined the features of physical structure and logical discourse structure to extract the candidate answers. By analysis of section 2.1, the discourse structure feature of text mainly includes the following: title, beginning and end of the text, sentence position, sentence of transition, transitional segments, passage similarity and sentence relevance, etc. From the perspective of physical discourse structure, we consider the text as associated network of linguistic units(words, sentences, paragraphs),and obtain the shallow features of candidate documents by taking the relationship between context into account, and identify the most relevant candidates answers by computing the relevance of query string, text title, paragraphs as well as sentences. From the perspective of logical discourse structure, firstly, use query expansion technique to reconstruct a query string so as to improve the recall rate of system; secondly, carry on a depth analysis for documents to determine the contents of semantic relations between paragraphs, logical relationship between main units and acquire information of each basic unit, create the similar rhetorical structure tree to represent the entire text. The research on rhetorical structure analysis and the method of establishing rhetorical structure tree has been mature. The method established by Liu Ting[6] is used.The difference is that the weight value of each relationship is not specified by the person, but determined by optimization of the whole model parameters. The logical discourse structure features are obtained from this tree: the depth, location and relation features of sentence in the tree (measured by the path from the root node to sentence node in the first four layers).Finally. Extract the features from the questions, question-answers, and the candidate answer documents to train the ranking model, and obtain the optimal ranking model which will predict the unknown answers.

3 Extracting answers based on discourse structure and rank Learning

Section 2 discussed the role of discourse structure in extracting answers. For the uncertainty and openness of such complex questions, this section proposed a new extracting method base on discourse structure and ranking strategy, which combine the features of discourse with the theory of learning to rank to extract the candidate answers. This method adopts the discourse structure to extract initial answers, builds ranking model and regards the judge problem of answers relevance as learning to rank answers, uses the learning to rank theory instead of classification learning which make the answers more accurate. Extract all kinds of features to train ranking model, re-ranking the answers to find the most relevant answers for the questions. We will discuss the construction of ranking model and features extraction below, and the detailed procedure of the new answers extracting method will be given in section 3.3.

3.1 Build the answers ranking model

Learning to rank is a new hot research in the field of information retrieval, between classification learning and regression learning, which focused on building the ranking model by use of the labeled training data and machine learning algorithm, thus to predict the ranking for the unknown queries. It has inspired numerous approaches to resolve many problems such as related work for web image retrieval, online advertisement and collaborative filtering [7][8][9],etc. Ranking-SVM is a typical kind of algorithm solving the problem of learning to rank, its core idea is that covert the pointwise ranking to preference learning of ordered binary classification problem, and use the SVM to solve[10][11].We construct the answers ranking model by labeling a large number of samples.

Given the input feature vector $\vec{x}$ collection of $X = \{\vec{x}_1, \dots, \vec{x}_n\} \subseteq R^n$, and its corresponding label $Y = \{y_1, \dots, y_m\}$,where m denotes the number of training samples, n indicates the dimension of input vectors. The purpose of learning to rank is to find an optimal decision function f^* from the set of decision functions $F = \{f : R^n \mapsto Y\}$, f^* can accurately predict the label y for unknown data $\vec{x}$,satisfies the condition $\vec{x}_i \prec_X \vec{x}_j \Leftrightarrow f(\vec{x}_i) \prec f(\vec{x}_j)$,where $\prec_Y$ and $\prec_Y$ are the ordinal relations defined in X and Y space. For any given two training examples $(\vec{x}_i, y_i)$ and $(\vec{x}_j, y_j)$, Ranking-SVM ingeniously defined a loss function based on ordinal data pairs for the problem of learning to rank:

$$l_{pref}(\vec{x}_1,\vec{x}_2,y_1,y_2,f(\vec{x}_1),f(\vec{x}_2))=\begin{cases}1 & (y_1 \succ y_2)\wedge\neg(f(\vec{x}_1)\succ f(\vec{x}_2))\\ 1 & (y_1 \prec y_2)\wedge\neg(f(\vec{x}_1)\prec f(\vec{x}_2))\\ 0 & else\end{cases}$$

（1）

According with the above definition, for a training set of m elements, samples from $p(\vec{x}_1, \vec{x}_2, y_1, y_2)$, and these samples are no longer independent, define the risk function follow of the above loss function:

$$R_{pref}(f)=\int_{R^n\times R^n\times R\times R} l_{pref}(\vec{x}_1,\vec{x}_2,y_1,y_2,f(\vec{x}_1,f(\vec{x}_2))dp(\vec{x}_1,\vec{x}_2,y_1,y_2)$$

（2）

The risk function based on the principle of minimum experience risk is as follows:

$$R_{EMP\,pref}(f) = \sum_{i=1}^{m}\sum_{j=1}^{m} l_{pref}(\vec{x}_i, \vec{x}_j, y_i, y_j, f(\vec{x}_i), f(\vec{x}_j))$$

（3）

In order to convert minimization of $R_{EMP \quad pref}(f)$ to a binary classification problem, re-defined the training set $S^{'} = (X^{'}, Y^{'})$ as follows:

$$\forall 0 \lhd | y_i^{(1)} - y_i^{(2)} | \qquad S^{'} = (X^{'}, Y^{'}) = \{\vec{x}_i^{(1)}, \vec{x}_i^{(2)}, \Omega(y_i^{(1)}, y_i^{(2)})\}_{i=1}^{\ell}$$
$$\Omega(y_i^{(1)}, y_i^{(2)}) = \operatorname{sgn}(y_i^{(1)} - y_i^{(2)})$$

（4）

Where $(\vec{x}_i^{(1)}, \vec{x}_i^{(2)})$ is ordered pairs represent the first element and second element respectively, its corresponding label are $y_i^{(1)}$, $y_i^{(2)}$. $\Omega(y_i^{(1)}, y_i^{(2)})$ is the indicator function, its value is 1 when $y_i^{(1)} \succ y_i^{(2)}$,and the value is -1 if $y_i^{(2)} \succ y_i^{(1)}$,otherwise is 0 ; l is the size of training samples.

This risk function can be expressed:

$$R_{EMP\ pref}(f) = \frac{\ell}{m^2} \sum_{i=1}^{\ell} \left((\vec{x}_i^{(1)}, \vec{x}_i^{(2)}), \Omega(\vec{y}_{i_i}^{(1)}, \vec{y}_{i_i}^{(2)}), \Omega(f(\vec{x}_{i_i}^{(1)}), f(\vec{x}_{i_i}^{(2)})) \right)$$

（5）

The above two equations mean that the problem of learning to ranking can be converted into a binary classification problem

Build the Ranking-SVM model:

$$\min_{\vec{w}} M(w) = \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{m} \xi_i$$
$$s.t \quad z_i \langle w, x_i - x_j \rangle \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \dots, m,$$

（6）

Supposed w^* is the optimal weight vector, the ranking functions of Ranking-SVM is as follows:

$$f_{w^{\square}}(x) =< w^{\square}, x > \qquad (7)$$

When applied the Ranking-SVM to solve the problem of answers ranking, first, choose any two different rank of candidate answers to establish a new dataset, then solve the optimal classification function (6), uses the equation (7) to rank finally. The learned predict function will minimize the error prediction on the ranking, and the predicted error is defined as the probability of $f(\vec{x}) \neq y$ .In our approach, the input vector X is composed of various features that extracted from the query and answers.

3.2 Features Extracting

We use questions, question-answers, answers and candidate answer documents as the feature extracting resources, comprehensive utilization of the relevant information about the relationship of questions, answers and answers documents, extract all relevant feature information reflected the correct answers availably besides the above-mentioned discourse features .We analyze the relevance between questions and answers from words surface, syntactic and semantic level ,divided the features into five categories: similarity features, translation features, density and frequency, structure features and external knowledge base features. The experiment will explore which feature can improve the ranking performance effectively. The following sections will introduce each group features.

3.2.1 Similarity Features

In most cases, the candidate answers should be the most similar to query. So we model the queries and answers based on sentence similarity, compute the similarity of each sentence between query and answers as the features value for ranking. It mainly include TF*IDF, BM25, semantic similarity.

TF*IDF: Due to a large number of recalled as well as qualified candidate sentences and paragraphs make a different contribution to the answers, we have to calculate their weight and obtain the sentence and passage with higher score as the candidate answers.

BM25 Score: BM25 retrieval algorithm [12] is presented by Robertson in TREC3 at 1994.It used to calculate the similarity between document D and query Q, we used BM25 to compute the similarity between query Q and candidate answers A. We chose this similarity formula because, out of all the IR models we tried, it provided the best ranking at the output of the answer retrieval component.

Semantic similarity: The above two methods only consider the statistical characteristics of the context without the semantic information of the term itself. With the help of Hownet, we compute the semantic distance between questions and answers as the semantic similarity[13] integrated into the ranking model

Dependency Syntactic Structure Features: In order to get a better understanding of the sentence structure between questions and answers, we meditate on syntactic structure by taking the words relations into account. The candidate answers and query will be expressed as syntax tree by the analysis of the dependencies and relationship types. Extract the syntax tree information and integrate into the ranking model to improve the performance of answer extracting. This feature reflect the semantic modifications between various components of sentence, it can obtain long matches and has nothing to do with the physical location.

By the dependency structure of sentences analysis, calculate the similarity of effective matching words between questions and answers, which referred to the composition of the whole core words and dependency words, the specific computation method referred to literature[14].We adds the type of questions and answers based on that method. The equation as shown in the following:

$$SIM(S_Q, S_A) = \lambda \frac{\sum_{i=1}^{n} W_i}{Max\{PairCount_Q, PairCount_A\}} + (1-\lambda) SIM'(S_Q, S_A) \qquad (8)$$

Where $\sum_{i=1}^{n} W_i$ indicates the total weight of effective matching words between questions S_Q and answers S_A , $PairCount_Q$ is the number of effective matches in questions, $PairCount_A$ is number of the effective matches of candidate answers. $SIM\ (S_Q, S_A)$ is the type function of questions and answers, the value is 1 if they have the same type ,otherwise, the value is 0, λ is the smoothing factor.

3.2.2 Translation Feature

Berger et al. [15] showed that similarity-based models are doomed to perform poorly for QA because they fail to 'bridge the lexical chasm' between questions and answers. One way to address this problem is to learn question-to-answer transformations using a translation model. Echihabi et al[16] proposed the noise channel model of question-answering system, the idea is that use the noise channel model to interpret the mapping relations between answers and questions as the following process: rewriter an given answer sentence S_A (contains the substring answers A) as query Q by a series of random operations, in which the relationship between terms can be integrated by the probability calculation. On this basis, we will take A as the source language, queries Q as the target language, and the probability of answers transformed Q as the answers relevance, the best answer is the maximum possibilities of answers translated into query. Estimate the conditional probability of the candidate answer sentences translated into query, integrated the translation model as the characteristics of different features into the ranking model. The probability is calculated by IBM's Modle1

$$P(Q \mid A) = \prod_{q \in Q} P(q \mid A) \qquad (9)$$

$$P(q \mid A) = (1-\lambda) P_{ml}(q \mid A) + \lambda P_{ml}(q \mid C) \qquad (10)$$

$$P_{ml}(q \mid A) = \sum_{a \in A} T(q \mid a) P_{ml}(a \mid A)) \qquad (11)$$

Where the equation (9) is the likelihood probability that the question string generated from answer A , smoothed by using the Jelinek-Mercer smoothing method as motioned in equation (10), linear interpolation is used on maximum likelihood model $P_{ml}(q \mid A)$ and the entire collection of answers C , $P_{ml}(q \mid C)$,computed by the maximum likelihood estimator. λ is the smoothing parameter. $P_{ml}(q \mid A)$ as the sum of the probabilities that the question string q is a translation of an answer a , $T(q \mid a)$ is translation probability, which can integrate the relation between words, weighted by the probability that a generated from A ,that is estimating unigram langue model.

3.2.3 Density and Frequency Features

This evaluation indicator measures the density and frequency of question terms in the answer text, mainly includes: numbers, distances, word sequences, etc. Variants of these features were used previously for answer extracting in factoid QA[17]. We will discuss briefly in the following:

Quantity features: The matching nouns (verb, numeral or quantifier) of candidate answers accounts for the proposition of nouns (verb, numeral or quantifier) in query words. The features reflect the similarity of designated part of speech and questions in word matching level.

Distance Features: The distance of candidate's answers from the words or phrase of questions. Compute the distance for each sentence contained the candidate answers and yield a weighted mean value.

$$DS = avg \sum_{i} \frac{1}{abs(pos(keyword_i) - pos(answer))}$$

Sequence Features: Computes the number of nonstop question words that are recognized in the same order in the answer. This feature investigate whether the matching words of sentence, answer sequence and questions sequence in the same order, measured by the words in same sequence percent of the number of question words.

3.2.4 External Knowledge Features

Besides the Hownet mentioned above, on-line knowledge base is an effective resource access to knowledge from large amount of information on the network. This online knowledge base includes on-line biography (for example: biography.com), on-line encyclopedia (for example: Encyclopedia), Wikipedia (wekipedia.com), etc, from which we can get the information about people, organization and other entities. These authoritative dictionaries defined or provided some useful and more authoritative knowledge available. Take the definition question as example, it needs to find a series of segment information related to target words of the questions, which require some auxiliary information to provide the user's interest, on-line knowledge base is highly effective for such questions. Therefore, we get the Boolean features values of relevant answers by validating the online knowledge base, if the answers appears in Wikipedia or Baidu encyclopedia, the corresponding feature value is 1,otherwise,the value is 0.

3.3 Answer Extraction Method Based on Discourse Structure and Ranking Strategy

Question -answering system, which generates query strings on the basis of question analysis and calls the search engine, is to find the answers finally from recalled texts. Therefore, the ability to exclude the irrelevant texts that do not contain the answers from the massive free text of the internet information is extremely important. Similar to the traditional question answering system, the answer extracting method based on discourse structure and ranking technology mainly consists of four modules::(1) question analysis (2) document retrieval (3)generate the answer sentences or paragraph (4)answers ranking. The system architecture shown in Figure 1.Method of answer extracting can be simply summarized as follows:

Firstly, generate a query string by the analysis of question. Secondly, make a profound analysis on the retrieval documents by use of discourse structure theory and the natural language processing technology of vocabulary, grammar, semantic analysis to identify the inherent logic relationship between paragraphs or sentences, and create text rhetorical structure tree to generate candidate answers paragraph or sentence. Finally, we will combine the discourse features of physical structure with logical structure as well as the above several types of features to train ranking model, thus to re-ranking answers. Specific extracting method will be described in the following.

Step 1: Question analysis and query string's generation

We do analysis for questions, such as dependency structure and semantic analysis with fine-grained classification. Such as the 'why' questions can be refined to 'cause', 'reason', 'purpose' and 'motivation', 'explanation'. The type of questions and answers could be determined at that time. Generate the query string to conduct query expansion. For example, causal connection is one of relationship normally present in the answers text of such questions. Add the discourse markers such as 'but', 'because', 'even if' appeared in the text to question keywords to form a query string together, thereby improving the recall rate.

Step 2: document retrieval

According to query expansion, retrieve the candidate answer documents relevant to questions with the help of search engines model.

Step 3: Generate a answer paragraph/sentence

On the level of text, paragraph, or sentence, conduct rhetorical relations analysis for the candidate answer documents, use the 'conjunction structure dictionary' and 'hierarchical structure dictionary' (reflect the feature words or phrase of text rhetorical relations) to extract the relevant information of rhetorical structure, then express as similar rhetorical structure tree, finally generate the candidate answer paragraph or sentence according to the weight value.

Step 4: Build answer ranking model

Extract the discourse structure features of questions and answers and a variety of other features discussed in section 3.2, train the ranking model combined with the external knowledge and then filter as well as rank the new answers, return the most relevant answer finally.

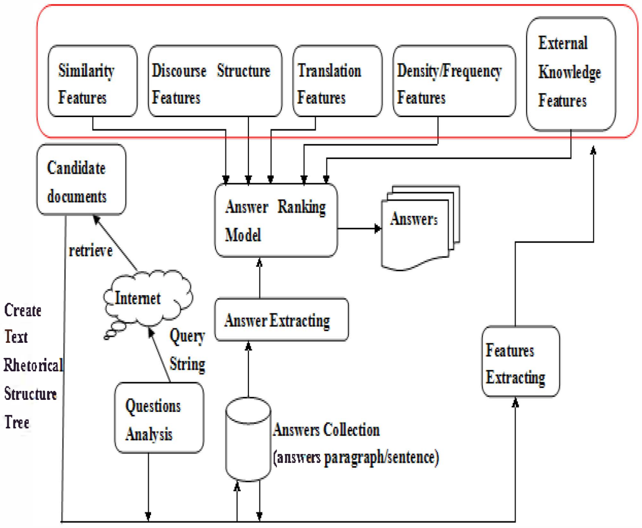

Figure 1: system architecture of complex questions

4 EXPERIMENT AND RESULT ANALYSIS

4.1 *Experiment setup*

In order to evaluate the results of answer extracting method based on discourse structure and ranking strategy, the experiment divides into two parts: The first part extract the features of questions and answers to train the ranking model, and explore which features contribute to increase the ranking performance. In the second part extract the candidate answers based on question analysis, rank the answers using the trained ranking model to examine the effect of answer extracting and compared with other extracting methods.

At present, there are no authoritative corpus resources in Chinese question answering system, so we paraphrase the TREC questions to construct Chinese questions set. Paraphrasing could maintain the consistent distribution of question types, replace those contents closely related cultural background of the familiar contents to Chinese people, thus ensure the answers of questions can be obtained from Chinese documents. The candidate answers will be marked as irrelevant (0), partial relevant(1), relevant (2).In order to avoid over-fitting phenomenon, we divided the data into three parts by use of cross-validation approach, included: training sets, validation sets, testing sets. The trained ranking model is applied to adjust parameters on validation sets and test ranking performance on testing sets. We extract various kinds of features from questions and answers, candidate answer documents to train the ranking model. This part could be called training stage, and we use gradient descent method to optimize the model.

The main task of answer extracting stage is first to analyze the questions, generate query string, retrieve the related documents, next, study on the candidate answer documents deeply to identify the inherent logic relationship between paragraphs or sentences, and then create text rhetorical structure tree to represent the whole text, thus to generate candidate answers paragraph or sentence. Finally, use the trained ranking model to ranking answers, and extract the most relevant answers for users.

4.2 *Experiment and Result analysis*

We adopt MRR(Mean Reciprocal Rank) and P@1(Precision at position 1 for query q) these two evaluation criteria, use vector space model (VSM) and the maximum entropy model as Baselines, and will discuss from the features selection influenced on ranking performance and comparative analysis of answer extracting methods in the following.

4.2.1 *Comparison of feature selection*

Different features set have different impacts on ranking performance, which will, in turn, affect the results of answers extracting. We put more attention on the performance of ranking model, so evaluate the percentage of correct answers ranked top-1 and mean average precision. The kernel function uses linear function; the following table lists the major features that we selected.

TABLE1： FEATURES SELECTION AND THE RESULTS OF ANSWER RANKING

Features		MRR	P@1
Similarity Features	BM25	56.05	41.13%
	+ IF*IDF	61.02	46.23%
	+ WN	62.51	48.24%
	+ D	63.38	48.36%
Translation Features	+ IBM -Model 1	64.42	50.31%
ensity/Frequency Features	+ S1	64.46	50.66%
	+ S2	64.49	50.72%
	+ S3	64.53	50.78%
External Knowledge Features	+ WEB	64.60	50.88%
Discourse Structures Features	+ RD	66.3%	52.88%
	+ RP	68.23%	55.06%
	+ REL	69.53%	56.83%

WN is semantic similarity, D indicates dependency syntax features, T denotes translation features, S_1 is quantity features, S_2 represents distance features, S_3 indicates sequence features, WEB is external knowledge features. RD, RP, REL indicates depths, locations, and relations features, respectively, obtained from rhetorical structure tree. In the experiment, we add the new features continuously to examine the effects on answer extracting. By analysis, we find that add the translation features, the recall rate and accuracy have been improved, which shows that the combination is critical for improvement, but little increase when add dependency syntactic structure features, the introduction of discourse structure features obviously improve the ranking performance and accuracy, thus explain that the analytical method of discourse structure can be effectively improve the results of answer extraction. Furthermore, carry out the combination of various features, repeat the experiment until the small change of accuracy.

4.2.2 Comparison of baseline methods

We choose the vector space model (VSM) and the maximum entropy model as the baseline methods, the recalled top-50 answers as the criterion of comparison. The comparison results of answer extracting method integrated discourse structure and ranking technology with the baseline methods as shown in the following table.

TABLE 2: THE COMPARISON RESULTS OF DIFFERENT ANSWER EXTRACTING METHOD

	Recall	MRR	P@1
VSM	40.2%	55.8%	41.81%
Maximum Entropy Model	44.4%	57.76%	49.90%
St + ranking	60.2%	68.73%	57.99%

Experiments show that the recall rate of answer extracting method based on discourse structure and ranking strategy reaches 60% which is superior to VSM and maximum entropy model , better to 20%, 16.2% accordingly. Obviously, this extracting method effectively use the relevant information of relationship between candidate documents, questions and answers, learn the sequence features of relevance to improve the accuracy of answers. While the maximum entropy model base on classification regards answer extraction as the classification of candidate answers. The task of classifier is learning the answers relevant or irrelevant, ignoring the order relation between answers; therefore, it reduces the accuracy of answers. In conclusion, the contribution of discourse structure and ranking technology is very considerable for answer extracting. Compared with the other methods, the precision of answers at position 1 has improved significantly.

4.2.3 Selection of kernel function

In the training process of ranking model, the ranking performance is different if we use different kernel functions. The experiment investigates the influence of different kernel functions on ranking. All kinds of features involved in the training, we use linear kernel function, polynomial kernel function, RBF kernel function to train the ranking model separately to examine the ranking performance.

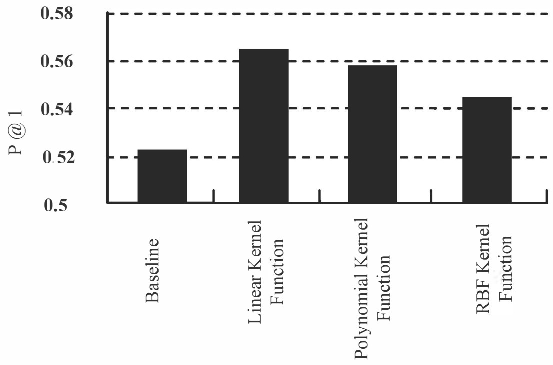

Figure 2. Kernel function selected and the results of answers ranking

From the results in Figure 2, the precision of answers improves relative to baseline methods, the results of linear kernel function is satisfied. It can be concluded that the linear kernel function is more suitable for learning to rank answers than polynomial kernel function and RBF kernel function.

4.2.4 Experimental Evaluation

Experiment results show that the proposed method of answer extracting improved the accuracy and recall rate greatly. The relevant initial answers was not in the front before ranking, which demonstrate that build ranking model to re-rank the answers could make a better use of various information about questions and answers .By analysis, due to the openness and uncertainty of complex questions, it has lower accuracy and recall rate compared with factoid questions. In the future, we will continue to further explore the characteristic of such questions for increasing the answer extracting effect.

5 CONCLUSION

For the uncertainty and dispersion of non-factoid questions, a new answer extracting method based on discourse structure and ranking strategy is prepossessed. This method makes a profound analysis for the candidate answer documents from the perspective of discourse structures, and generates the candidate answer sentence or paragraph according to the size of node weight value of text rhetorical structure tree, by comprehensive utilization of the relevant information about the relationship of questions, answers and answers documents to build the ranking model, thereby improving the accuracy of answers. Experiments show that the presented method can effectively improve the accuracy and quality of non-factoid answers. This method can be used to rank the answers to all kinds of questions by slightly changing the features of questions and answers.

6. ACKNOWLEDGMENT

This paper is supported by National Nature Science Foundation (No. 60863011, 61163022, 61175068, 61100205), The Key Project of Yunnan Nature Science Foundation (No. 2008CC023), National Innovation Fund for Technology based Firms(No. 11C26215305905).

REFERENCES

[1] Hovy, E.H., U. Hermjakob, C.-Y. Lin, and D. Ravichandran. 2002. A Question/Answer Typology with Surface Text Patterns. Notebook of the Human Language Technology conference (HLT). San Diego, CA.[2] M. Maybury, editor. Toward a Question Answering Roadmap, pages 8-11. 2003.

[2] [M. Maybury, editor. Toward a Question Answering Roadmap, The MITRE Corporation. Technical Paper, Bedford, MA 01730, November.pages 8-11. 2003.

[3] D Ravichandran, E Hovy. Statistical QA-Classfier vs.Re-ranker:What's the difference? Proceedings of the ACL 2003 Workshop on Multilingual Summarization and Question Answering, pp. 69-75.

[4] Chengying Chi,zhiyi Ma,tianshun Yao,et,al.Text comprehension and analysis of Chinese text structure. Journal of Chinese Information Processing.1997.1

[5] Mann,W.C.and Thompson,S.A.Rhetorical Structure theory: A theory of text organization. Information Sciences Institute, University of Southern California,1987.

[6] Ting Liu, Kaizhu Wang. Research on automatic abstracting based on text multilevel dependency structure. Journal of computer research & development.1999.36(4):479-488

[7] Pinji,Li and Jun Ma.Use Genetic Programming to rank Web Images. China Communications.2010.1:80-92

[8] Jianyi,Liu,Cong Wang,Wenbin Yao.Keyword Extraction for Contextual Advertising. China Communications .2010.10:51-56

[9] NN Liu and Q. Yang. Eigenrank: a ranking-oriented approach to collaborative filtering. In SIGIR, pages 83–90, 2008

[10] Herbrich, R., Graepel, T., and Obermayer, K. Large Margin Rank Boundaries for Ordinal Regression. Smola, A., Bartlett, P., Scholkopf, B., and Schuurmans, D., eds., Advances in Large Margin Classifiers. MIT Press, 2000, 115~132.

[11] Joachims, T. Optimizing Search Engines Using Click-through Data. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA, 2002, 133~142.

[12] Robertson, S. E., S. Walker, M. Hancock-Beaulieu and M. Gatford. Okapi in TREC3. In Proceedings of Text Retrieval Conference, Gaithersburg, USA. U.S. National Institute of Standards and Technology, NIST Special Publication 500-225: 1994. 109~126.

[13] Qun Liu,Sujian Li.Word Similarity Computing Based on How-net.[C]//Proceedings of 3th Chinese lexical semantics. 2002-05

[14] Bin Li, Ting Liu,Bing Qin.Chinese Sentence Similarity Computation Based on Semantic Dependency Relationship Analysis. Application Research of Computers, 2003.20(12): 15-17

[15] Berger, R. Caruana, D. Cohn, D. Freytag, and V. Mittal.2000. Bridging the Lexical Chasm: Statistical Approaches to Answer Finding. Proc. of SIGIR.

[16] A.Echihabi and D. Marcu. A Noisy-Channel Approach to Question Answering [C]//Proceedings of ACL 2003.Brown, S. Della Pietra, V. Della Pietra, R. Mercer.

[17] Baoshun Hu,Daling Wang.An Answer Extraction Algorithm Based on Syntax Structure Feature Parsing and Classification. Chinese journal of computers.2008.31(4):663-675