

Semantic Chunk Annotation for complex questions using Conditional Random Field

Shixi Fan

Department of computer science
Harbin Institute of Technology
Shenzhen Graduate School,
Shenzhen, 518055, china
fanshixi@hit.edu.cn

Wing W. Y. Ng

Department of computer science
Harbin Institute of Technology
Shenzhen Graduate School,
Shenzhen, 518055, china
wing@hitsz.edu.cn

Xiaolong Wang

Department of computer science
Harbin Institute of Technology
Shenzhen Graduate School,
Shenzhen, 518055, china
wangxl@insun.hit.edu.cn

Yaoyun Zhang

Department of computer science
Harbin Institute of Technology
Shenzhen Graduate School,
Shenzhen, 518055, china
Xiaoni5122@gmail.com

Xuan Wang

Department of computer science
Harbin Institute of Technology
Shenzhen Graduate School,
Shenzhen, 518055, china
wangxuan@insun.hit.edu.cn

Abstract

This paper presents a CRF (Conditional Random Field) model for Semantic Chunk Annotation in a Chinese Question and Answering System (SCACQA). The model was derived from a corpus of real world questions, which are collected from some discussion groups on the Internet. The questions are supposed to be answered by other people, so some of the questions are very complex. Mutual information was adopted for feature selection. The training data collection consists of 14000 sentences and the testing data collection consists of 4000 sentences. The result shows an F-score of 93.07%.

© 2008. Licensed under the *Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported* license (<http://creativecommons.org/licenses/by-nc-sa/3.0/>). Some rights reserved.

1 Introduction

1.1 Introduction of Q&A System

Automated question answering has been a hot topic of research and development since the earliest AI applications (A.M. Turing, 1950). Since then there has been a continual interest in processing knowledge and retrieving it efficiently to users automatically. The end of the 1980s saw a boost in information retrieval technologies and applications, with an unprecedented growth in the amount of digital information available, an explosion of growth in the use of computers for communications, and the increasing number of users that have access to all this information (Diego Moll'and Jose'Luis Vicedo, 2007). Search engines such as Google, Yahoo, Baidu and etc have made a great success for people's information need.

Anyhow, search engines are keywords-based which can only return links of relevant web pages, failing to provide a friendly user-interface with queries expressed in natural language sentences or questions, or to return precise answers to users. Especially from the end of the 1990s, as

information retrieval technologies and methodologies became mature and grew more slowly in pace, automated question answering(Q&A) systems which accept questions in free natural language formations and return exactly the answer or a short paragraph containing relevant information has become an urgent necessity. Major international evaluations such as TREC, CLEF and NTCIR have attracted the participation of many powerful systems.

The architecture of a Q&A system generally includes three modules: question processing, candidate answer/document retrieval, and answer extraction and re-ranking.

1.2 Introduction of Question Analyzing

Question Analyzing, as the premise and foundation of the latter two modules, is of paramount importance to the integrated performance of a Q&A system. The reason is quite intuitive: a question contains all the information to retrieve the corresponding answer. Misinterpretation or too much loss of information during the processing will inevitably lead to poor precision of the system.

The early research efforts and evaluations in Q&A were focused mainly on factoid questions asking for named entities, such as time, numbers, and locations and so on. The questions in the test corpus of TREC and other organizations are also in short and simple form. Complex hierarchy in question types (Dragomir Radev et al, 2001), question templates (Min-Yuh Day et al, 2005), question parsing (Ulf Hermjakob, 2001) and various machine learning methods (Dell Zhang and Wee Sun Lee, 2003) are used for factoid question analysis, aiming to find what named entity is asked in the question. There are some questions which are very complicated or even need domain restricted knowledge and reasoning technique. Automatic Q&A system can not deal with such questions with current technique.

In china, there is a new kind of web based Q&A system which is a special kind of discussion group. Unlike common discussion group, in the web based Q&A system one user posts a question, other users can give answers to it. It is found that at least 50% percent questions (Valentin Jijkoun and Maarten de Rijke, 2005) posted by users are non-factoid and surely more complicated both in question pattern and information need than those questions in the test set of TREC and other FAQ. An example is as follows:

{我想开专卖店工商如何交费手续如何办理} 。

(I want to open a special sell shop, how to pay the tax, and what is the procedure of it)。

This kind of Q&A system can complement the search engines effectively. As the best search engines in china, Baidu open the Baidu Knowledge² Q&A system from 2003, and now it has more than 29 million question-answer pairs.

There are also many other systems of this kind such as Google Groups, Yahoo Answers and Sina Knowledge³. This kind of system is a big question-answer pair database which can be treated as a FAQ database. How to search from the database and how to analyze the questions in the database needs new methods and techniques. More deeper and precise capture of the semantics in those complex questions is required. This phenomenon has also been noticed by some researchers and organizations. The spotlight gradually shifted to the processing and semantic understanding of complex questions. From 2006, TREC launched a new annually evaluation CIQ&A (complex, interactive Question Answering), aiming to promote the development of interactive systems capable of addressing complex information needs. The targets of national programs AQUAINT and QUETAL are all at new interface and new enhancements to current state-of-the-art Q&A systems to handle more complex inputs and situations.

A few researchers and institutions serve as pioneers in complex questions study. Different technologies, such as definitions of different sets of question types, templates and sentence patterns (Noriko Tomuro, 2003) (Hyo-Jung Oh et al, 2005) machine learning methods (Radu Soricut and Eric Brill, 2004), language translation model (Jiwoon Jeon, W et al, 2005), composition of information needs of the complex question (Sanda Harabagiu et al, 2006) and so on, have been experimented on the processing of complex question, gearing the acquired information to the facility of other Q&A modules.

Several major problems faced now by researcher of complex questions are stated as follow:

First: Unlike factoid questions, it is very difficult to define a comprehensive type hierarchy for complex questions. Different domains under research may require definitions of different sets of question types, as shown in (Hyo-Jung Oh et al, 2005). Especially, the types of certain ques-

² <http://zhidao.baidu.com/>

³ <http://iask.sina.com.cn/>

tions are ambiguous and hard to identify. For example:

网上炒股的方法是什么？

(how to play stock market ?)

This question type can be treated as definition, procedure or entity.

Second: Lack of recognition of different semantic chunks and the relations between them. FAQFinder (Radu Soricut and Eric Brill, 2004) also used semantic measure to credit the similarity between different questions. Nevertheless, the question similarity is only a simple summation of the semantic similarity between words from the two question sentences. Question pattern are very useful and easy to implement, as justified by previous work. However, just like the problem with question types, question patterns have limitation on the coverage of all the variations of complex question formation. Currently, after the question processing step in most systems, the semantic meaning of large part of complex questions still remain vague. Besides, confining user's input only within the selection of provided pattern may lead to unfriendly and unwelcome user interface. (Ingrid Zukerman and Eric Horvitz, 2001) used decision tree to model and recognize the information need, question and answer coverage, topic, focus and restrictions of a question. Although features employed in the experiments were described in detail, no selection process of those feature, or comparison between them was mentioned.

This paper presents a general method for Chinese question analyzing. Our goal is to annotate the semantic chunks for the question automatically.

2 Semantic Chunk Annotation

Chinese language differs a lot from English in many aspects. Mature methodologies and features well-justified in English Q&A systems are valuable sources of reference, but no direct copy is possible.

The Ask-Answer system⁴ is a Chinese online Q&A system where people can ask and answer questions like other web based Q&A system. The characteristic of this system is that it can give the answer automatically by searching from the asked question database when a new question is presented by people. The architecture of the automatically answer system is shown in figure 1. The system contains a list of question-answer pairs on particular subject. When users input a

question from the web pages, the question is submitted to the system and then question-answer pair is returned by searching from the questions asked before. The system includes four main parts: question pre-processing, question analyzing, searching and answer getting.

The question pre-processing part will segment the input questions into words, label POS tags for every word. Sometimes people ask two or more questions at one time, the questions should be made into simple forms by conjunctive structure detection. The question analyzing program will find out the question type, topic, focus and etc. The answer getting part will get the answer by computing the similarity between the input question and the questions asked before. The question analyzing part annotates the semantic chunks for the question. So that the question can be mapped into semantic space and the question similarity can be computed semantically. The Semantic chunk annotation is the most important part of the system.

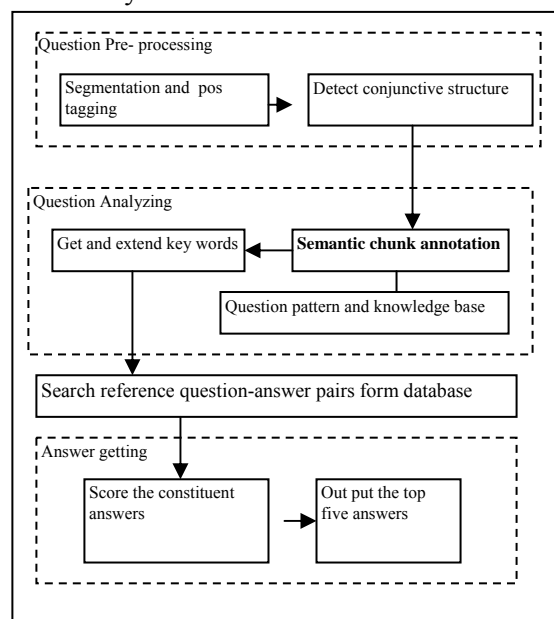


Figure 1 the architecture of the automatically answer system

Currently, no work has been reported yet on the question semantic chunk annotation in Chinese. The prosperity of major on-line discussion groups provides an abundant ready corpus for question answering research. Using questions collected from on-line discussion groups; we make a deep research on semantic meanings and build a question semantic chunk annotation model based on Conditional Random Field.

Five types of semantic chunks were defined: Topic, Focus, Restriction, Rubbish information and Interrogative information. The topic of a

⁴ <http://haitianyuan.com/qa>

question which is the topic or subject asked is the most important semantic chunk. The focus of a question is the asking point of the question. The restriction information can restrict the question's information need and the answers. The rubbish information is those words in the question that has no semantic meanings for the question. Interrogative information is a semantic tag set which corresponds to the question type. The interrogative information includes interrogative words, some special verbs and nouns words and all these words together determine the question type. The semantic chunk information is shown in table 1.

Semantic chunk tag	Abbreviation	Meaning
Topic	T	The question subject
Focus	F	The additional information of topic
Restrict	Re	Such as Time restriction and location restriction
Rubbish information	Ru	Words no meaning for the question
Other	O	other information without semantic meaning
The following is interrogative information		
Quantity	Wqua	
Description	Wdes	The answer need description
Yes/No	Wyes	The answer should be yes or no
List	Wlis	The answer should be a list of entity
Definition	Wdef	The answer is the definition of topic
Location	Wloc	The answer is location
Reason	Wrea	The answer can explain the question
Contrast	Wcon	The answer is the comparison of the items proposed in the question
People	Wwho	The answer is about the people's information
Choice	Wcho	The answer is one of the choice proposed in the question
Time	Wtim	The answer is the data or time length about the event in the question
Entity	Went	The answer is the attribute of the topic.

Table 1: Semantic chunks

An annotation example question is as follows:

银行交易类型 CMD 是什么意思？

What is the meaning of bank transaction CMD?

This question can be annotated as follows:

{银行/nz 交易/nz 类型/n CMD/nx}/T

{是/v 什么/r 意思/n}/Wdef ? /w

This kind of annotation is not convenient for CRF model, so the tags were transfer into the B I O form. (Shown as follows)

银行/nz/B-T 交易/nz/I-T 类型/n/I-T ↓

CMD/nx/I-T 是/v/B-Wdef 什么/r/I-Wdef

意思/n/I-Wdef ? /w/O

Then the Semantic chunk annotation can be treated as a sequence tag problem.

3 Semantic Chunk Annotation model

3.1 Overview of the CRF model

The conditional random field (CRF) is a discriminative probabilistic model proposed by John Lafferty, et al (2001) to overcome the long-range dependencies problems associated with generative models. CRF was originally designed to label and segment sequences of observations, but can be used more generally. Let X, Y be random variables over observed data sequences and corresponding label sequences, respectively. For simplicity of descriptions, we assume that the random variable sequences X and Y have the same length, and use $x = [x_1, x_2, \dots, x_m]$

and $y = [y_1, y_2, \dots, y_m]$ to represent instances of X and Y , respectively. CRF defines the conditional probability distribution $P(Y|X)$ of label sequences given observation sequences as follows

$$P_\lambda(Y|X) = \frac{1}{Z_\lambda(X)} \exp\left(\sum_{i=1}^n \lambda_i f_i(X, Y)\right) \quad (1)$$

Where $Z_\lambda(X)$ is the normalizing factor that ensures equation 2.

$$\sum_y P_\lambda(y|x) = 1 \quad (2)$$

In equation 2 the λ_i is a model parameter and

$f_i(X, Y)$ is a feature function (often binary-valued) that becomes positive (one for binary-valued feature function) when X contains a certain feature in a certain position and Y takes a certain label, and becomes zero otherwise.

Unlike Maximum Entropy model which use single normalization constant to yield a joint distribution, CRFs use the observation-dependent normalization $Z_\lambda(X)$ for conditional distributions. So CRFs can avoid the label biased problem. Given a set of training data

$$T = \{(x_k, y_k), k = 1, 2, \dots, n\}$$

With an empirical distribution $\tilde{P}(X, Y)$, CRF

determines the model parameters $\lambda = \{\lambda_i\}$ by maximizing the log-likelihood of the training set

$$\Gamma(P_\lambda) = \sum_{k=1}^N \log P_\lambda(y_k | x_k) \quad (3)$$

$$\propto \sum_{x,y} \tilde{P}(x,y) \log P_\lambda(y | x)$$

3.2 Features for the model

The following features, which are used for training the CRF model, are selected according to the empirical observation and some semantic meanings. These features are listed in the following table.

Feature type index	Feature type name
1	Current word
2	Current POS tag
3	Pre-1 word POS tag
4	Pre-2 word POS tag
5	Post -1 word POS tag
6	Post -2 word POS tag
7	Question pattern
8	Question type
9	Is pattern key word
10	Pattern tag

Table 2: the Features for the model

Current word:

The current word should be considered when adding semantic tag for it. But there are too many words in Chinese language and only part of them will contribute to the performance, a set of words was selected. The *word set* includes segment note and some key words such as time key word and rubbish key word. When the current word is in the *word set* the current word feature is the current word itself, and null on the other hand.

Current POS tag:

Current POS tag is the part of speech tag for the current word.

Pre-1 word POS tag:

Pre- 1 word POS tag is the POS tag of the first word before the labeling word in the sentence. If the Pre-1 word does not exit (the current is the first word in the sentence), the Pre- 1 word POS tag is set to null.

Pre-2 word POS tag:

Pre- 2 word POS tag is the POS tag of the second word before the labeling word in the sentence. If the Pre-2 word does not exit, the Pre- 2 word POS tag is set to null.

Post -1 word POS tag:

Post - 1 word POS tag is the POS tag of the first word after the labeling word in the sentence. If the Post -1 word does not exit (the current is the first word in the sentence), the Post - 1 word POS tag is set to null.

Post -2 word POS tag:

Post - 2 word POS tag is the POS tag of the second word after the labeling word in the sentence. If the Post-2 word does not exit, the Pre- 2 word POS tag is set to null.

Question pattern:

Question pattern which is associated with question type, can locate question topic, question focus by surface string matching. For example, (where is <topic>). The patterns are extracted from the training data automatically. When a pattern is matched, it is treated as a feature. There are 1083 question patterns collected manually.

Question type:

Question type is an important feature for question analyzing. The question patterns have the ability of deciding the question type. If there is no question pattern matching the question, the question type is defined by a decision tree algorithm.

Is pattern key word:

For each question pattern, there are some key words. When the current word belongs to the pattern key word this feature is set to “yes”, else it is set to “no”.

Pattern tag:

When a pattern is matched, the topic, focus and restriction can be identified by the pattern. We can give out the tags for the question and the tags are treated as features. If there is no pattern is matched, the feature is set to null.

4 Feature Selection experiment

Feature selection is important in classifying systems such as neural networks (NNs), Maximum Entropy, Conditional Random Field and etc. The problem of feature selection has been tackled by many researchers. Principal component analysis (PCA) method and Rough Set Method are often used for feature selection. Recent years, mutual information has received more attention for feature selection problem.

According to the information theory, the uncertainty of a random variable X can be measured by its entropy $H(X)$. For a classifying problem, there are class label set represented by C and feature set represented by F. The conditional entropy $H(C | F)$ measures the uncertainty about

C when F is known, and the Mutual information $I(C, F)$ is defined as:

$$I(C; F) = H(C) - H(C | F) \quad (4)$$

The feature set is known; so that the objective of training the model is to minimize the conditional entropy $H(C | F)$ equally maximize the mutual information $I(C; F)$. In the feature set F, some features are irrelevant or redundant. So that the goal of a feature selection problem is to find a feature S ($S \subset F$), which achieve the higher values of $I(C; F)$. The set S is a subset of F and its size should be as small as possible. There are some algorithms for feature selection problem. The ideal greedy selection algorithm using mutual information is realized as follows (Nojun Kwak and Chong-Ho Choi, 2002):

Input: S- an empty set

F- The selected feature set

Output: a small reduced feature set S which is equivalent to F

Step 1: calculate the MI with the Class set C, $\forall f_i \in F$, compute $I(C; f_i)$

Step 2: select the feature that maximizes $I(C; f_i)$,

set $F \leftarrow F \setminus \{f_j\}, S \leftarrow \{f_j\}$

Step 3: repeat until desired number of features are selected.

1) Calculate the MI with the Class set C and S, $\forall f_i \in F$, compute $I(C; S, f_i)$

2) Select the feature that maximizes $I(C; S, f_i)$, set $F \leftarrow F \setminus \{f_j\}, S \leftarrow \{f_j\}$

Step 4: Output the set S that contains the selected features

To calculate MI the PDFs (Probability Distribution Functions) are required. When features and classing types are dispersing, the probability can be calculated statistically. In our system, the PDFs are got from the training corpus statistically.

The training corpus contains 14000 sentences. The training corpus was divided into 10 parts, with each part 1400 sentences. And each part is divided into working set and checking set. The working set, which contains 90% percent data, was used to select feature by MI algorithm. The checking set, which contains 10% percent data, was used to test the performance of the selected feature sequence. When the feature sequence was selected by the MI algorithm, a sequence of CRF models was trained by adding one feature at each time. The checking data was used to test the performance of these models.

Selected feature sequence	The open test result									
	1	2	3	4	5	6	7	8	9	10
7, 10, 3, 1, 5, 2, 4, 6, 8, 9	0.5104	0.8764	0.8864	0.8918	0.8925	0.8977	0.8992	0.9023	0.9025	0.9018
7, 10, 1, 3, 5, 2, 4, 6, 8, 9	0.5241	0.8775	0.8822	0.8911	0.8926	0.8956	0.8967	0.9010	0.9005	0.9007
7, 10, 1, 3, 5, 2, 4, 6, 8, 9	0.5090	0.8691	0.8748	0.8851	0.8852	0.8914	0.8929	0.8955	0.8955	0.8949
7, 10, 1, 3, 5, 2, 4, 6, 9, 8	0.5157	0.8769	0.8823	0.8913	0.8925	0.8978	0.8985	0.9017	0.9018	0.9010
7, 10, 1, 3, 5, 2, 4, 6, 8, 9	0.5144	0.8821	0.8856	0.8921	0.8931	0.8972	0.8981	0.9010	0.9009	0.9007
7, 10, 3, 1, 5, 2, 4, 6, 8, 9	0.5086	0.8795	0.8876	0.8914	0.8919	0.8960	0.8967	0.9016	0.9013	0.9011
7, 10, 1, 3, 5, 2, 4, 6, 8, 9	0.5202	0.8811	0.8850	0.8920	0.8931	0.8977	0.8980	0.9015	0.9013	0.9009
7, 10, 1, 3, 5, 2, 4, 6, 8, 9	0.5015	0.8858	0.8879	0.8948	0.8942	0.8998	0.8992	0.9033	0.9027	0.9023
7, 10, 1, 3, 5, 2, 4, 6, 8, 9	0.5179	0.8806	0.8805	0.8898	0.8908	0.8954	0.8958	0.8982	0.8982	0.8986
7, 10, 1, 3, 5, 2, 4, 6, 8, 9	0.5153	0.8921	0.8931	0.9006	0.9012	0.9041	0.9039	0.9071	0.9068	0.9067

Table 3: the feature selection result and the test result

In table 3, each row contains data corresponding to one part of the training corpus so there are ten rows with data in the table. The third row corresponds to the first part and the last row corresponds to the tenth part. There are eleven columns in the table, the first columns is the fea-

tures sequence selected by the mutual information algorithm for each part. The second column is the open test result with the first feature in the feature sequence. The third column is the open test result with the first two features in the feature sequence and so on. From the table, it is

clear that the feature 7(Question pattern) and 10(Pattern tag) are very important, while the feature 8(Question type) and 9(Is pattern key word) are not necessary. The explanation about this phenomenon is that the “pattern key word” and “Question type” information can be covered by the Question patterns. So feature 8 and 9 are not used in the Conditional Random Field model.

5 Semantic Chunk Annotation Experiment

The test and training data used in our system are collected from the website (Baidu knowledge and the Ask-Answer system), where people proposed questions and answers. The training data consists of 14000 and the test data consists of 4000 sentences. The data set consists of word

tokens, POS and semantic chunk tags. The POS and semantic tags are assigned to each word tokens.

The performance is measured with three rates: precision (Pre), recall (Rec) and F-score (F1).

$$\text{Pre} = \text{Match}/\text{Model} \quad (5)$$

$$\text{Rec} = \text{Match}/\text{Manual} \quad (6)$$

$$\text{F1} = 2 * \text{Pre} * \text{Rec} / (\text{Pre} + \text{Rec}) \quad (7)$$

Match is the count of the tags that was predicted right. Model is the count of the tags that was predicted by the model. Manual is the count of the tags that was labeled manually.

Table 4 shows the performance of annotation of different semantic chunk types. The first column is the semantic chunk tag. The last three columns are precision, recall and F1 value of the semantic chunk performance, respectively.

Label	Manual	Model	Match	Pre.()	Rec.()	F1
B-T, I-T	17061, 78462	16327, 80488	14825, 76461	90.80, 95.00	86.89, 97.45	88.80, 96.21
B-F, I-F	5072, 13029	5079, 13583	4657, 12259	91.69, 90.25	91.82, 94.09	91.75, 92.13
B-Ru, I-Ru	775, 30	11, 0	2, 0	18.18, 0.00	0.26, 0.00	0.51, 0.00
O	8354	8459	6676	78.92	79.91	79.41
B-Wqua, I-Wqua	1363, 934	1327, 1028	1298, 881	97.81, 85.70	95.23, 94.33	96.51, 89.81
B-Wyes, I-Wyes	5669, 1162	5702, 1098	5550, 1083	97.33, 98.63	97.90, 93.20	97.62, 95.84
B-Wdes, I-Wdes	2907, 278	2855, 185	2779, 184	97.34, 99.46	95.60, 66.19	96.46, 79.48
B-Wlis, I-Wlis	603, 257	563, 248	560, 248	99.47, 100	92.87, 96.50	96.05, 98.22
B-Wdef, I-Wdef	1420, 1813	1430, 1878	1280, 1695	89.51, 90.26	90.14, 93.49	89.82, 91.85
B-Wloc, I-Wloc	683, 431	665, 395	661, 392	99.40, 99.24	96.78, 90.95	98.07, 94.92
B-Wrea, I-Wrea	902, 159	873, 83	843, 82	96.56, 98.80	93.46, 51.57	94.99, 67.77
B-Wcon, I-Wcon	552, 317	515, 344	503, 291	97.67, 84.59	91.12, 91.80	94.28, 88.05
B-Wwho, I-Wwho	420, 364	357, 350	348, 336	97.48, 96.00	82.86, 92.31	89.58, 94.12
B-Wcho, I-Wcho	857, 85	738, 0	686, 0	92.95, 0.00	80.05, 0.00	86.02, 0.00
B-Wtim, I-Wtim	408, 427	401, 419	355, 380	88.53, 90.69	87.01, 88.99	87.76, 89.83
B-Went, I-Went	284, 150	95, 81	93, 80	97.89, 98.77	32.75, 53.33	49.08, 69.26
Avg	145577	145577	135488	93.07	93.07	93.07

Table 4: the performance of different semantic chunk

The semantic chunk type of “Topic” and “Focus” can be annotated well. Topic and focus semantic chunks have a large percentage in all the semantic chunks and they are important for question analyzing. So the result is really good for the whole Q&A system.

As for “Rubbish” semantic chunk, it only has 0.51 and 0.0 F1 measure for B-Ru and I-Ru. One reason is lacking enough training examples, for there are only 1031 occurrences in the training data. Another reason is sometimes restriction is complex.

6 Conclusion and future work

This paper present a new method for Chinese question analyzing based on CRF. The features are selected by using mutual information algorithm. The selected features work effectively for the CRF model. The experiments on the test data

set achieve 93.07% in F1 measure. In the future, new features should be discovered and new methods will be used.

Acknowledgment

This work is supported by Major Program of National Natural Science Foundation of China (No.60435020 and No. 90612005) and the High Technology Research and Development Program of China (2006AA01Z197).

References

- A.M. Turing. 1950. *Computing Machinery and Intelligence*. Mind, 236 (59): 433~460.
- Diego Moll', Jose' Luis Vicedo. 2007. *Question Answering in Restricted Domains: An Overview*. Computational Linguistics, 33(1),

- Dragomir Radev, WeiGuo Fan, Leila Kosseim. 2001. *The QUANTUM Question Answering System*. TREC.
- Min-Yuh Day, Cheng-Wei Lee, Shih-Hung WU, Chormg-Shyong Ong, Wen-Lian Hsu. 2005. *An Integrated Knowledge-based and Machine Learning Approach for Chinese Question Classification*. Proceedings of the IEEE International Conference on Natural Language Processing and Knowledge Engineering, Wuhan, China,:620~625.
- Ulf Hermjakob. 2001. *Parsing and Question Classification for Question Answering*. Proceedings of the ACL Workshop on Open-Domain Question Answering, Toulouse,:19~25.
- Dell Zhang, Wee Sun Lee. 2003. *Question classification using support vector machines*. Proceedings of the 26th Annual International ACM Conference on Research and Development in Information Retrieval(SIGIR), Toronto, Canada,26 ~ 32.
- Valentin Jijkoun, Maarten de Rijke.2005. *Retrieving Answers from Frequently Asked Questions Pages on the Web*. CIKM'05, Bermen, Germany.
- Noriko Tomuro. 2003. *Interrogative Reformulation Patterns and Acquisition of Question Paraphrases*. Proceeding of the Second International Workshop on Paraphrasing, :33~40.
- Hyo-Jung Oh, Chung-Hee Lee, Hyeon-Jin Kim, Myung-Gil Jang. 2005. *Descriptive Question Answering in Encyclopedia*. Proceedings of the ACL Interactive Poster and Demonstration Sessions, pages 21–24, Ann Arbor.
- Radu Soricut, Eric Brill. 2004, *Automatic Question Answering: Beyond the Factoid*. Proceedings of HLT-NAACL ,:57~64.
- Jiwoon Jeon, W. Bruce Croft and Joon Ho Lee. 2005. *Finding Similar Questions in Large Question and Answer Archives*. CIKM'05, Bremen, Germany.
- Sanda Harabagiu, Finley Lacatusu and Andrew Hickl. 2006 . *Answering Complex Questions with Random Walk Models*. SIGIR'06, Seattle, Washington, USA,pp220-227.
- Ingrid Zukerman, Eric Horvitz. 2001. *Using Machine Learning Techniques to Interpret WH-questions*. ACL.
- John Lafferty, Andrew McCallum, Fernando Pereira. 2001. *Conditional Random Fields: probabilistic Models for Segmenting and Labeling Sequence Data*. Proceedings of the Eighteenth International Conference on Machine Learning, p.282-289.
- Nojun Kwak and Chong-Ho Choi. 2002. *Input feature selection for classification problems*. IEEE Trans on Neural Networks,,13(1):143-159