ABSTRACT

Most question answering systems feature a step to predict an expected answer type given a question. Li and Roth [7] proposed an oft-cited taxonomy to the categorize the answer types as well as an annotated data set. While offering a framework compatible with supervised learning, this method builds on a fixed and rigid model that has to be updated when the question-answering domain changes. More recently, Pinchak and Lin [10] designed a dynamic method using a syntactic model of the answers that proved more versatile. They used syntactic dependencies to model the question context and evaluated the performance on an English corpus. However, syntactic properties may vary across languages and techniques applicable to English may fail with other languages. In this paper, we present a method for constructing a probability-based answer type model for each different question. We adapted and reproduced the original experiment of Pinchak and Lin [10] on a Chinese corpus and we extended their model to semantic dependencies. Our model evaluates the probability that a candidate answer fits into the semantic context of a given question. We carried out an evaluation on a set of questions either drawn from NTCIR corpus [9] or that we created manually.

Categories and Subject Descriptors

H.3.4 [Systems and Software]: Question-answering (fact retrieval) systems

General Terms

Design, Experimentation, Performance

Keywords

Semantic Role Labeling, Question and Answering, Intelligent Systems

1. INTRODUCTION

Question answering is concerned with building systems that automatically answer questions posed by humans in natural language. Notable applications of it include IBM’s Watson [5] and Apple’s Siri. In question answering, the prediction of the answer type plays a major role to reduce the number of possible answers. Given a question like:

Who is the current president of China?

we can expect from the subject complement the answer to be a person and discard other potential candidates that do not fit this category. Question type prediction helps us narrow down the search space we need to explore to find the correct answer.

Pinchak and Lin [10] introduced a method using grammatical relations and unsupervised learning to construct a probabilistic answer type model for each question. This method proved successful in predicting the question type dynamically [6]. However, as question answering has been applied overwhelmingly to English and syntactic properties may vary across languages, we wanted to verify if a similar technique could be applied to Chinese. This paper investigates the possible applicability to Chinese of a question typing technique designed for English. In addition, syntactic dependencies can easily be extended to semantic ones. Using semantic role labeling, we can concisely extract the predicate, object of the sentence and hence have a more abstract structure of it. We complemented our syntactic model with an extension to semantic roles.

2. CORPUS AND CORPUS PROCESSING

2.1 Corpus

We used the Chinese version of Wikipedia as textual knowledge source, and, as test set, a set of questions from the NTCIR test corpus and 20 questions that we created manually. We tokenized the Chinese Wikipedia corpus and we built a dictionary of all the words. We parsed it by using the Koshik framework [4] and for every word of the corpus, we extracted all its contexts.

2.2 Word Clusters

As Pinchak and Lin [10], we applied a clustering procedure to cope with data sparseness. However, instead of a clustering by committee, we applied the Brown hierarchical word clustering algorithm [2, 8] to the Chinese Wikipedia. We set the number of clusters to 1000. A word cluster often reflects a specific sense as the following one that consists of Chinese city names: 衡阳, 安阳, 丹阳, 汉阳, 淮安, 乌鲁木齐, 常州, 全州, 张家口, 湘潭, 烟台, ...

2.3 Contexts

As noted by Pinchak and Lin [10], the context of a word often constrains its semantic type. Following their idea, we defined the contexts of a word to be the dependency paths starting or ending
with this word and we extended their syntactic formulation to a semantic one: We considered the semantic paths from the predicate to the word with the $A_0$ and $A_1$ labels. This extension sometimes overlaps with Pinchak and Lin [10] as the word with label $A_0$ is often the subject of the predicate and the word with label $A_1$, its object.

Figure 1 and Table 1 show, respectively, an example of a dependency tree and its predicate–argument structure for the sentence: "中国的首都是北京‘Beijing is the capital of China’. From the predicate–argument structure shown in Table 1, 首都 $A_0$ is $A_1 \rightarrow$ 北京, we can extract the context of the word ‘首都’ ‘capital’: $X \leftarrow A_0 \rightarrow$ 北京. Similarly, the context of the word ‘北京‘ ‘Beijing’ is: 首都 $A_0 \rightarrow$ 北京 $\rightarrow$ $A_1 \rightarrow X$.

Table 1: Predicate–argument structure in the CoNLL format.

<table>
<thead>
<tr>
<th>Form</th>
<th>IsPred</th>
<th>Predicate</th>
<th>Arguments: is 01</th>
</tr>
</thead>
<tbody>
<tr>
<td>中国</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>的</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>首都</td>
<td>–</td>
<td>–</td>
<td>A0</td>
</tr>
<tr>
<td>是</td>
<td>Y</td>
<td>is 01</td>
<td>-</td>
</tr>
<tr>
<td>北京</td>
<td>–</td>
<td>–</td>
<td>A1</td>
</tr>
</tbody>
</table>

Following Pinchak and Lin [10], we defined the probability that a candidate word $w$ is the answer to the question as the probability of $w$ appearing in the set of question contexts $\tau_q$, $\text{in}(w, \tau_q)$, given $w$: $P(\text{in}(w, \tau_q) | w)$. Considering the data sparseness, We use a hidden variable $C$ to represent the cluster to which the candidate answer $w$ belongs. Because Brown clusters allow one word to belong to exactly one cluster, we have $P(C | w) = 1$. According to Bayes formula, we have:

$$P(\text{in}(w, \tau_q) | w) = P(\text{in}(w, \tau_q) | C, w)$$

We assume that all the words in a same cluster have the same probability to occur in a context. Therefore:

$$P(\text{in}(w, \tau_q) | C, w) \approx P(\text{in}(C, \tau_q) | C)$$

We can rewrite Eq. 1 as:

$$P(\text{in}(w, \tau_q) | w) \approx P(\text{in}(C, \tau_q) | C) \ldots \ldots (2)$$

When $\tau_q$ consists of multiple contexts, we make the naive Bayes assumption that each individual context $\gamma_q \in \tau_q$ is independent of all the other contexts given cluster $C$.

$$P(\text{in}(w, \tau_q) | w) \approx \prod_{\gamma_q \in \tau_q, \gamma \in \gamma_q} P(\text{in}(\gamma_q, \gamma) | C) \ldots \ldots (3)$$

Given $w$, our probabilistic model needs to estimate the $P(\text{in}(C, \gamma_q) | C)$ term, where word $w$ is member of cluster $C$ and $\gamma$ is a question or candidate context. The word-context dictionary consists of keys and values, where the keys are the words and the context-frequency dictionary, the values. We can derive the joint and marginal frequency counts of the contexts and the words ($\text{in}(w, \gamma_q), |\text{in}(w, \gamma_q)|, |\text{in}(w, \gamma_q)|$) from the dictionary. Using these counts, we compute the probabilities $P(\text{in}(w, \gamma_q)), P(\text{in}(w, \gamma)),$ and $P(\text{in}(w, \gamma))$. We can then compute $P(\text{in}(C, \gamma) | C)$ with a Laplace smoothing; see Eq. 5:

$$P(\text{in}(C, \gamma) | C) = \frac{\sum_{w' \in C} |\text{in}(w', \gamma)| + P(\text{in}(w, \gamma))}{\sum_{w' \in C} |\text{in}(w', \gamma)| + 1} \ldots \ldots (5)$$

4. EXPERIMENTAL SETUP AND RESULTS

4.1 Indexer

We used Coreseek [3] to index the corpus. We retrieved the documents relevant to a question with Coreseek’s $TF \cdot IDF$, BM25 implementation. We applied this system to our test set of 398 NT-CIR questions and we evaluated the results using the median rank, the average rank of the correct answer, the recall rate, and the frequency distribution of correct answers. The Chinese Wikipedia consists of both traditional and simplified Chinese characters. Before we ran the indexer, we converted all characters of the corpus into simplified Chinese. It resulted into a significant performance improvement on correct answer distribution over the same baseline experiment using a corpus that contained traditional Chinese with a cutoff of 120 documents. See Figures 2 and 3.

In the remaining experiments, we only used the converted corpus. Figure 4 shows the average and median rank; Figure 5, the recall rate. To get a balance between precision and recall rate, we set the number of documents retrieved by the search engine to 40. We parsed these documents to extract all the nouns. Finally, we ranked these nouns according to their frequency.

4.2 Reranker

We built a reranker from the probabilistic model described in Sect. 3. Using the results from the indexer, we applied the reranker
to the candidate answers: We multiplied the frequency of each candidate word by the square root of the probability it appears in the question context extracted from a given question; see Eqs. 4 and 5. Answers are reranked by the new score. We used the data set of 398 questions from NTCIR and 20 questions created manually to evaluate our reranker. Table 4.2 shows the median rank of the correct answer in the list of candidates for the question set. In this table, we report the baseline figures without reranker, Pinchak and Lin [10]’s method using syntactic dependencies, and finally the figures resulting from the integration of semantic dependencies. We found that the semantic role method had a result worse than the baseline when we used the NTCIR test set, but it is better when we use the test set created manually. Examining the analysis results, we could find that the cause is the frequent incorrect identification of the predicate by the semantic role labeler for verbs like ‘is in’ as in the sentence: ‘The capital of China is in Beijing.’ To our surprise, it is better than the dependency tree method [10], which defines the contexts of a word to be the undirected paths in dependency trees. We believe, this is due to a better capacity of the semantic role labeler to provide abstract contexts.

5. CONCLUSIONS

We have proposed to use semantic roles to predict answer types. Our probabilistic model uses candidate contexts extracted from corpus and question contexts from questions to calculate the appropriateness of a candidate answer. We showed that dynamic question typing is applicable to Chinese and that semantic role labeling can improve the performance of question typing over grammatical functions.

Acknowledgments

This research was supported by Vetenskapsrådet, the Swedish research council, under grant 621-2010–4800, and the Det digitaliserade samhället and eSSENCE programs.

6. REFERENCES