Question Classification with Semantic Tree Kernel

Yan Pan
School of Software
Sun Yat-sen University,
Guangzhou, P. R. China 510275
panyan5@mail.sysu.edu.cn

Yong Tang*, Luxian Lin, Yemin Luo
Department of Computer Science
Sun Yat-sen University,
Guangzhou, P. R. China 510275
issty@mail.sysu.edu.cn,
{linlx,luoyemin}@mail2.sysu.edu.cn
*Corresponding Author, +86 20 84112137

ABSTRACT
Question Classification plays an important role in most Question Answering systems. In this paper, we exploit semantic features in Support Vector Machines (SVMs) for Question Classification. We propose a semantic tree kernel to incorporate semantic similarity information. A diverse set of semantic features is evaluated. Experimental results show that SVMs with semantic features, especially semantic classes, can significantly outperform the state-of-the-art systems.

Categories and Subject Descriptors
H.3.1 [**Content Analysis and Indexing**], H.3.3 [**Information Search and Retrieval**].

General Terms: Algorithms, Experimentation

Keywords
Semantic Class, Tree Kernel, Support Vector Machines, Question Classification, Question Answering, Machine Learning

1. INTRODUCTION
Question Answering (QA), initiated in the Text Retrieval Conference (TREC, http://trec.nist.gov) since 1999, aims to find precise answers to users' questions in a large collection of documents. A typical QA system adopts a pipeline structure that contains "question analysis", "passage retrieval", "answer extraction" and "answer ranking" modules. Question analysis mainly focuses on so-called Question Classification (QC) which divides questions into several predefined semantic categories. As the first processing step of a QA system, QC can help not only to exclude irrelevant answers with types not matching the identified question type, but also to adopt different strategies in the following processing steps. The accuracy of QC has great impact on the overall performance of QA systems.

Kernel based methods such as Support Vector Machines (SVMs) have proven to be powerful in data classification. Here we adopt a new tree kernel incorporating a rich set of semantic features extracted from the questions. Experimental results show that the SVM classifier with the new kernel significantly outperforms the state-of-the-art systems on the community-standard data set.

2. SEMANTIC TREE KERNEL
Tree Kernel (TK) [1] measures the similarity of two trees by counting the number of their common tree fragments. Here we define a new semantic tree kernel by incorporating semantic similarity into SubSet tree kernel [1].

Give two trees T_1 and T_2, suppose $F=\{f_1,f_2,\ldots,f_i,\ldots\}$ be the tree fragment space and $I_i(n)$ be equal to 1 if f_i is rooted by node n, 0 otherwise. The Semantic Tree Kernel (STK) is defined as

$$TK(n_1,n_2)=\sum_{n_1\in N_1}\sum_{n_2\in N_2}\mu^{(d(n_1)+d(n_2))/2}\Delta(n_1,n_2)$$

where N_1 and N_2 are sets of nodes in T_1 and T_2 respectively, $d(n_1)$ and $d(n_2)$ are the depth of n_1 and n_2, and

$$\Delta(n_1,n_2)=\sum_{i=1}^{|F|}I_i(n_1,n_2)$$

which counts the number of common tree fragments rooted at n_1 and n_2.

We define the △ function as:

(1) If n_1 and n_2 are both leaf nodes, then $\triangle(n_1,n_2)=sim(n_1,n_2)$.

$$sim(n_1,n_2)=\begin{cases}\gamma & if\ n_1=n_2\ or\ n_1,n_2\ are\ in \\ & the\ same\ semantic\ class \\ \gamma\times\alpha & if\ n_1\ and\ n_2\ are\ synonyms \\ & in\ WordNet \\ \gamma\times\beta & if,\ for\ n_1\ and\ n_2,\ one\ is\ direct \\ & hypernym\ of\ the\ other\ in\ WordNet \\ 0 & otherwise\end{cases} \quad (1)$$

(2) If n_1 and n_2 are both pre-terminal nodes, then

$$\triangle(n1,n2)=\gamma\times\lambda$$

{3}While n_1 and n_2 are neither leaf nodes nor pre-terminal nodes, if the production rule at n_1 and n_2 are not the same, then $\triangle(n_1,n_2)=0$.Otherwise

$$\Delta(n_1,n_2)=\lambda\prod_{i=1}^{nc(n1)}(\sigma+\Delta(ch(n_1,i),ch(n_2,i)))$$

where nc(n) is the number of node n's children and ch(n,j) represents the j-th child of n.

3. SEMANTIC FEATURES
We use three types of semantic features as evidence to evaluate the similarity of words (leaf nodes in a parse tree).

(1) Semantic classes. After analyzing 5500 questions, we manually construct 20 semantic classes and collect a word list of each class. As an example, Table 1 shows 5 semantic classes and sample words in each class.

Table 1. Semantic Classes and Sample Words

Semantic Class	List of Words
Person	man, daughter, actor, resident, robber,……
Creature	animal, plant, flower, tree, tiger, seed,……
Location	airport, bridge, state, city, park, street, ……
Organization	company, bank, club, college, army, ……
Substance	mental, stone, mineral, fuel, gas, ……

(2) WordNet senses. In WordNet, words are organized in a hierarchy of synonyms, hypernyms and hyponyms. As shown in Equation (1), we set appropriate values respectively for the case that two words are synonyms or one is a hypernym of the other.

(3) Named Entities (NEs). We use an external Named Entity Recognition (NER) tool to identify NEs in questions. Here we only consider three types of NEs: person, organization and location.

4. EXPERIMENTS

In our experiments, we use the public available dataset [2] constructed by UIUC, which is frequently used in Question Classification research. The dataset contains about 5,500 questions as training set and 500 questions from TREC 10 QA Track as test set. These manually labeled questions can be divided into 6 coarse-grained categories and 50 fine-grained ones. Here we only focus on coarse-grained classification.

First of all, we process the question in the dataset with a noun phrase chunker. In each noun phrase, we only preserve the head word and ignore other words. After that, we replace the identified Named Entities with their corresponding type names. Given the question "What is the only color Johnny Cash wears on stage?" as an example, after chunking and NE recognition, we find a noun phrase "the only color" and a person NE "Johnny Cash".

What is ***[the only color]*** ***<u>Johnny Cash</u>*** *wears on stage?*

Thus we preserve the head word "color" in the noun phrase and replace "Johnny Cash" with its NE type, finally we get

What is ***color NE_PERSON*** *wears on stage?*

We use an implementation similar to [3] as the baseline. Then we evaluate different combination of the above semantic features. As shown in Table 2, semantic classes can evidently help to improve the accuracy, while the WordNet senses and NE features can only make slightly improvement. Table 3 illustrates the accuracy comparison of our approach with state-of-the-art results in the literatures, in which we can see that our approach observably outperforms the state-of-the-art systems.

5. CONCLUSION

In this paper, we have introduced a new tree kernel to encode textual information in Question Classification task. We have also evaluated the kernel with combination of different semantic feature sets. Our experimental results suggest that semantic information, especially semantic classes, can help to improve accuracy significantly in SVM question classifier.

Since the semantic class information appears to be promising in QC, in our future work, we aim to study on automatically mining semantic classes from the web or external corpus, instead of manual construction like that in this paper. Fine-grained classification also seems to be an interesting research area.

Table 2. Accuracy of the classifier with different combination of semantic features

Semantic Features	Accuracy
baseline	90%
WordNet Senses	90.6%
NEs	90.4%
Semantic Classes	**93.2%**
WorNet Senses + NEs	90.8%
Semantic Classes + WordNet Senses	93.2%
Semantic Classes + NEs	93.6%
ALL	**94%**

Table 3. Comparison with state-of-the-art approaches

Approach		Accuracy
SVM	SubSet TK [3]	90%
	Syntactic and Shallow Semantic TK [5]	91.8%
	Semantic TK (our approach)	**94%**
SNoW with semantic information [4]		92.5%
Log-Linear Model [6]		92.6%

6. ACKNOLEDGEMENT

This work was partially supported by National Natural Science Foundation of China under Grant No. 60673135, 60736020.

7. REFERENCES

[1] M. Collins and N. Duffy. Convolution Kernels for Natural Language. In Proceedings of Neural Information Processing Systems (NIPS14), 2001

[2] X. Li and D. Roth. Learning Question Classifiers. In Proceedings of the 19th International Conference on Computational Linguistics (COLING'02), 2002.

[3] D. Zhang and W. Lee. Question Classification Using Support Vector Machines. In Proceedings of the 26th ACM SIGIR (SIGIR'03). 2003.

[4] X. Li and D. Roth. Learning question classifiers: the role of semantic information. Journal of Natural Language Engineering, 12(3), 229-249, 2005.

[5] A. Moschitti, S. Quarteroni, R. Basili and S. Manandhar, Exploiting Syntactic and Shallow Semantic Kernels for Question/Answer Classification. In Proceedings of the 45th Conference of the Association for Computational Linguistics (ACL'07), 2007.

[6] P. Blunsom, K. Kocik and J. R. Curran. Question Classification with Log-Linear Models. In Proceedings of the 29th ACM SIGIR (SIGIR'06). 2006.