Chinese Frame Identification using T-CRF Model

Ru Li[*], Haijing Liu[+], Shuanghong Li[#]
School of Computer and Information Technology
Shanxi University
[*]liru@sxu.edu.cn
[+]bukaohuaxue@163.com
[#]lishuanghong09@gmail.com

Abstract

As one of the important tasks of SemEval Evaluation, Frame Semantic Structure Extraction based on the FrameNet has received much more attention in NLP field. This task is often divided into three sub-tasks: recognizing target words which are word expressions that evoke semantic frames, assigning the correct frame to them, namely, Frame Identification (FI), and for each target word, detecting and labeling the corresponding frame elements properly. Frame identification is the foundation of this task. Since the existence of links between frame semantics and syntactic features, we attempt to study FI on the basis of dependency syntax. Therefore, we adopt a tree-structured conditional random field (T-CRF) model to solve Chinese frame identification based on Dependency Parsing. 7 typical lexical units which belong to more than one frame in Chinese FrameNet were selected to be researched. 940 human annotated sentences serve as the training data, and evaluation on 128 test data achieved 81.46% precision. Compared with previous works, our result shows obvious improvement.

1 Introduction

In recent years, semantic research has roused great interest in NLP field. With the progress of many semantic lexicons, this research gradually becomes promising and exciting. As one of the tasks of SemEval Evaluation, Frame Semantic Structure Extraction based on the FrameNet grows to be highlighted for special attention.

Given a sentence, the task of Frame Semantic Structure Extraction consists of the following three parts: recognizing the word expressions (target words) that evoke semantic frames; discriminating the word sense (frame) of each evoking expression; for each target word, labeling its syntactic dependents with regard to which roles in that frame they fill (Baker et al., 2006). Among of these three components, frame identification is the fundamental and key problem. However, current research of this task in Chinese is only focused on semantic role labeling based on the given target words and their corresponding frames (Xue, 2008). We insist that whether target words can be assigned correct frames in context is a crucial problem demanding prompt solution in this task.

Chinese FrameNet (CFN) (You and Liu, 2005), developed by Shanxi University, is an ongoing effort of building a semantic lexicon for Chinese based on the theory of Frame Semantics (Fillmore, 1982), referencing the FrameNet(Baker et al., 1998) and supported by corpus evidence. The CFN project currently contains more than 2100 lexical units, more than 300 semantic frames, and has exemplified more than 21600 annotated sentences. The ultimate goal of this project is to generate information about the articulation of the semantic and syntactic requirements of Chinese lexical items and presents this information in a variety of web-based reports and represents the lexical semantics of all the sentences in a Chinese text.

According to statistics, there are 332 lexical units belonging to more than one frame in the current CFN databases. For example, lexical unit "表示"can evoke the following three frames: "表达 (Expressing_publicly) ", "陈述 (Statement) " and "代表(representative) ". In order to extract the semantic structure of a sentence containing ambiguous target words, the first step is to assign the correct frame to the target words in a given context.

This task is similar with the word sense disambiguation (WSD) task to a certain extent (Katrin Erk, 2005). WSD is to resolve the inherent polysemia of words by determining the appropriate sense for each ambiguous word in a given text, while frame identification is assigning a correct frame for the ambiguous target word in the current sentence context. Nevertheless, essential difference exists between them. WSD prefers to disambiguation on static sense, whereas based on the frame semantics, frame identification lays particular emphasis on consistency between sentence scene and the dynamic scene described by the candidate frames.

Since the existence of links between frame semantics and syntactic features, we adopt a tree-structured conditional random field (T-CRF) model to solve Chinese frame identification based on Dependency Parsing. 7 typical lexical units which belong to more than one frame in CFN were selected to be researched. 940 human annotated sentences were collected for the training data, and 128 for test data.

The rest of this paper is organized as follows. Section 2 introduces some related work. Section 3 gives a simple system description. Section 4 describes Chinese frame identification using T-CRF model. Section 5 presents our experimental results and some analysis. Section 6 is the conclusions.

2 Related Work

With the development and improvement of FrameNet, the research based on this lexical resource is increasing gradually. Frame Semantic Structure Extraction based on FrameNet is such hot topics. One sub-tasks of this research is frame identification, which is the research problem in this paper.

At present, there are some but not much work on frame identification. Main works are as follows: CL Research participated in the SemEval-2007 task for Frame Semantic Structure Extraction. They integrated the use of FrameNet in the Text Parser component of the CL Research KMS. In particular, they created a FrameNet dictionary from the FrameNet databases with the CL Research DIMAP dictionary software and used this dictionary as a lexical resource. The current FrameNet DIMAP dictionary contains 7575 entries, with many entries having multiple senses. For each sense, the FrameNet part of speech, the definition, the frame name, the ID number, and the definition source (identified as FN or COD) are captured from the FrameNet files. When a lexical unit is recognized in processing the text, the first step is to retrieve the entry for that item in the dictionary and use the frame element realization patterns to disambiguate among the senses. A score is computed for each sense and the score with the highest sense was selected. They evaluated on three texts and the best result is 66.10% precision (Litkowski, 2007).

Adrian Bejan and Hathaway (2007) selected from the FN lexicon 556 target words that evoke at least two semantic frames and have at least five sentences annotated for each frame. And then they assembled a multi-class classifier using two types of models: SVM and Maximum Entropy for each ambiguous target word. They extracted features used in word sense disambiguation (Florain et al., 2002), lexical features of the target word, and NAMED ENTITY FLAGS associated with the root vertex in a syntactic parse tree. For the rest of the ambiguous target words that have less than five sentences annotated, they randomly chose a frame as being the correct frame in a given context. For FI sub-task, they obtained 76.71% accuracy compared to a baseline of 60.72% accuracy that always predicts the most annotated frame for each of the 556 target words.

Johansson and Nugues (2007) firstly used some filtering rules to detect target words, and for the target words left after the filtering, they trained a disambiguating SVM classifier on all ambiguous words listed in FrameNet. The classifier used the following features: target lemma, target word, sub categorization frame, the set of dependencies of the target, the set of words of the child vertexes, and the parent word of the target. Its accuracy was 84% on the ambiguous

words, compared to a first-sense baseline score of 74%.

The above researches focused on English based on FrameNet. To our knowledge, there exists no work for Chinese by far. Most methods mentioned above treat the frame identification as an independent classification problem for each ambiguous target word in a sentence. However, because of neglecting the relations between the candidate frames, the resulting frame assignment may be semantically inconsistent over the sentence.

3   System Description

Our system consists of three stages. The first is corpus construction of our experiments. We selected 7 typical lexical units from the current CFN lexicon which can evoke at least two semantic frames. They are "表示","想","有","叫","倒","下降","装载", respectively. For each of them, we collected sentences containing this word from Sogou Corpus and CCL Contemporary Chinese Corpus of Beijing University. Through a series of refining, 940 sentences annotated correct frame for each target word comprise a standard corpus as the training data. Another 128 sentences serve as the test data.

The second stage is dependency parsing. We used LTP of Information Retrieval Research Center, Harbin Institute of Technology (HIT-CIR) to POS tagging and dependency parsing the training and test sentences. For the obvious lexical and syntax errors in the outputs, manually corrected was conducted.

At last, Chinese frame identification task is regarded as a labeling task on the dependency tree structure. By using T-CRF, we can model this as the maximization of the probability of word sense (frame) trees, given the scores for vertexes and edges. In the training phase, appropriate features of vertex and edge are extracted, and the weight vectors are optimized over the training data.

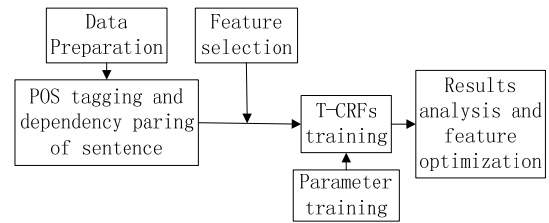Figure 1 gives an illustration of the system.

Figure 1. Framework of the system

4   Chinese Frame Identification

Given a sentence, frame identification is to determine an appropriate frame for each of target words by comparing consistency between sentence context and the dynamic scene described by their candidate frames. Currently, most researchers addressed this task as an independent classification problem for each target word in a sentence. Consequently, the resulting frame assignment for each target word may be semantically inconsistent over the sentence.

We regard Chinese frame identification problem as a labeling task on the dependency tree structure due to the links between syntactic features and frame semantics. Our empirical study shows that the frame of target word not only influenced by the adjacent words in position but also its governor and dependents words in syntactic structure. Therefore, we try to solve this problem based on dependency parsing. T-CRF model is a special CRF model, which is different from widely used linear-chain CRFs, in which the random variables are organized in a tree structure. As we can see, it should be feasible and reasonable to adopt a T-CRF model to frame identification after parsing the sentence.

In this section, we firstly introduce the linear-chain CRFs briefly, and then explain the T-CRF model for Chinese frame identification, especially the feature selection and parameter estimation.

4.1   Tree-Structured Conditional Random Field（T-CRF）

Conditional Random Fields (CRFs) are undirected graphical models (Lafferty et al, 2001). For the observation sequence $X = x_1 x_2 x_3 \cdots x_n$ and its corresponding label sequence $Y = y_1 y_2 y_3 \cdots y_n$, CRF defines the conditional probability as:

$$P(Y \mid X)$$

$$= \frac{1}{Z(X)} \{ \ \exp(\sum_i \sum_k \lambda_k f_k(y_{i-1}, y_i, X))$$

$$+ \exp(\sum_i \sum_k \mu_k g_k(y_i, X)) \ \}$$

where X is the observation sequence, and y_i is the label at position i in label sequence Y. $f_k(\cdot)$ and $g_k(\cdot)$ are feature functions. λ_k and μ_k are the weight vectors. $Z(X)$ is the normalization factor. CRFs are state-of-the-art methods for sequence labeling problem in many NLP tasks.

Tree-Structured Conditional Random Field (Tang et al., 2006) is a particular case of CRFs, which can model dependencies across hierarchically laid-out information, such as dependency syntactic relations between words in a sentence.

The graphical structure of T-CRF is a tree, in which three main relations exist for a vertex: parent-child, child-parent and sibling vertexes. In our experiments, we only used parent-child edges and child-parent edges. The sibling-vertexes edges were ignored because of weak dependency syntactic relation between words in a sentence. So the probability distribution in our T-CRF model can be written as below.

$$p(y \mid x)$$

$$= \frac{1}{Z(x)} \exp \sum_{v \in V} \{ F + G + S \}$$

$$F = \sum_j \lambda_j f_j(v, y(v), x)$$

$$G = \sum_k \mu_k g_k(v, y(v), x, v', y(v'))$$

$$S = \sum_l \sigma_l s_l(v, y(v), x, v'', y(v''))$$

where F、G、S represent the feature functions of current vertex, feature functions of parent vertex of current vertex and feature functions of child vertexes of current vertex, respectively. v is a word corresponding to the vertex in the tree, v' is the parent vertex of v and v'' are the child vertexes of v.

In Chinese frame identification, the observation x in T-CRF corresponds to a word in the current sentence. The label y thus corresponds to the frame name for the word. In the experi-

mental corpus, for the target word, y is annotated its correct frame name, while for the other words left, y is annotated tag "null". These target words are the 7 lexical units we selected and their frames come from the current CFN lexicon. At present, only the frame identification of target word was studied, the disambiguation of the other multi-senses words in the sentence was not being processed.

Although T-CRFs are relatively new models, they have already been applied to several NLP tasks, such as semantic role labeling, semantic annotation, word sense disambiguation, image modeling.(Cohn and Blunsom, 2005; Tang et al., 2006; Jun et al., 2009; Awasthi et al., 2007). All these works proved this model to be useful in modeling the semantic structure in a sentence or a text. Our study is the first application of T-CRFs to frame identification.

4.2   Feature Selection

In order to apply T-CRF model, it is necessary to represent the sentence with a hierarchical structure. We used LTP of HIT-CIR to POS tagging and dependency parsing the training and test sentences. To facilitate the description of feature selection based on the dependency tree structure, figure 2 gives the dependency output of an example.
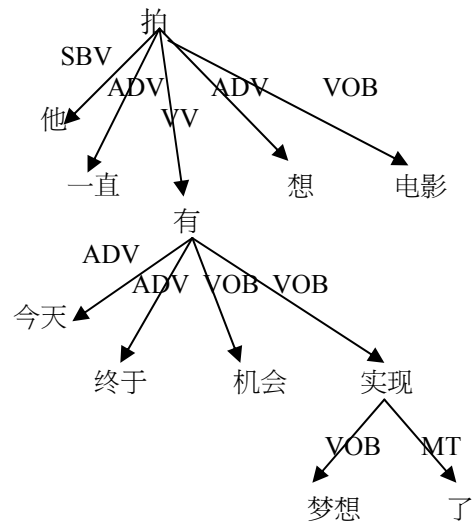

Figure 2. Example of a dependency parsed sentence.

This example sentence is:"他一直想拍电影，今天终于有机会实现梦想了". In English, it reads "He has been want to make films, and finally has the opportunity to realize his dream

today". In the dependency tree structure, arrow points from the parent vertex to child vertex, the label on a arc is the type of dependency relation between the parent and the child vertex.

Feature selection is a core problem in sequence labeling model. In our experiments, 18 template settings were conducted to discover the best features for frame identification. During this process, we considered two main factors: firstly, the number of features should not be too large so as to avoid the over-fitting phenomenon; secondly, the selected features should be able to provide enough information conditioned on tolerated computation, for the purpose of improving the performance of system. With the increasing of the number of features and the cost of the system, if the performance of system can not be improved obviously, we stopped to add features and regard the parameter of current template as the best. At this moment, a good balance between the performance and cost of computation was achieved.

We experimented with two different types of feature settings. One we used was the very basic feature sets based on the words and Part of Speech (POS) and their bigram features. In order to see the effectiveness of dependency features, the other type of feature settings include more informative tree features. These features capture information about a vertex's parent, its children and the relation with its parent and children. These features are semantically and structurally very informative and we expect to improve our performance with them. The base and tree features we used are listed in table 1.

In these features, the setting of basic features is fundamental and meaningful because it can be used to compare T-CRF and linear chain CRF. For the tree features, given the i-th vertex in the observation x_i, $f(y_p, y_c)$ and $f(y_c, y_p)$ represent whether the current vertex has a parent-child dependency with a parent vertex and whether it has a parent-child dependency with a child vertex, respectively. In dependency grammars (Igor' A. Melchuk, 1988), every vertex has only one parent as its governor, and may have more than one child as its dependents. Words in a sentence through certain syntactic relations form the semantic structure of this sentence. Therefore, we argue that the

Category	Features	
Base features	Word and bigram of word, POS and bigram of POS	
Tree features	$f(y_p, y_c)$	Parent vertex of current word
		The edge between current word and its parent
		The dependency relation type between current word and its parent
	$f(y_c, y_p)$	child vertex of current word
		The edge between current word and its child
		The dependency relation type between current word and its child

Table 1. Base Features & Tree Features

words that have syntactic dependency relations with the target word are more important than the ones neighboring with it in position for frame identification. For this reason, we added the parent vertex and children vertexes into the tree features. With respective to the relation type, we used the annotation sets defined by HIT-CIR in LTP, which contain 24 kinds of dependency relation types. One thing should be concerned is that we don't consider all types of children vertexes. This is because that according to our empirical study, not all of the children have strong dependencies with the target word. On the contrary, more features would bring the noise and affect the efficiency seriously. Hence, we chose 4 types of children relation from the linguistic point of view. They are, "SBV(subject-verb)" representing "主谓关系", "VOB(verb-object)" representing "动宾关系", "ADV(adverbial)" representing "状中结构" and "ATT(attribute)" representing "定中关系". From the point of grammars and semantics, these four relations are more influenced on the words in a sentence. As we know, the subject, predicate and object constitute the semantic core of a sentence. The good news is that experimental results proved this hypothesis relatively correct.

4.3 Parameter Estimation

The parameter estimation is to optimize the parameters $\theta = \{\lambda1, \lambda2,...; \mu1, \mu2,...\}$ from training data $D\{(x1, y1), (x2, y2),...\}$ with empirical distribution $p(x, y)$. Nowadays, the commonly used method for parameter estimation is maximum likelihood function. That is $L_\theta = \text{argmax} \sum_i \log(p_\theta(y_i / x_i))$ given the observation sequences $\{x_1, x_2,...\}$ and label sequences $\{y_1, y_2,...\}$.

In this paper, the conventional L-BFGS method was used to estimate the optimal parameters $\theta = \{\lambda1, \lambda2,...; \mu1, \mu2,...\}$ (Jorge Nocedal and Stephen J. Wright. 1999).

5 Experiments

5.1 Data preparation

So far, there has no research on Chinese frame identification, thus it is unfeasible to do experiments based on readily available corpus. Accordingly, preparing a good and reasonable training and test data is our fundamental task.

At present, there are 332 lexical units that can evoke at least two frames in the CFN lexicon. In this paper, we selected 7 typical ambiguous lexical units to be researched. They are "表示","想","有","叫","倒","下降","装载". The selection principle is following: first of all, it is time-consuming to construct corpus for all of the 332 lexical units, so currently we just studied part of them to prove the validity of the method we proposed. Secondly, the frames evoked by these lexical units should be distinguished clearly by human annotators. For example, lexical unit "高兴" can evoke these three frames: "心理刺激(Experiencer_obj)", "情感体验 (Experiencer_subj)" and "情感反应 (Emotion_directed)". All these frames describe a tender feeling in psychology, so it is difficult to discriminate among them and thus hard to annotate sentences correctly. Thirdly, these 7 lexical units are high frequency words so it is easier to collect sentences and make the experiments more practical.

For each of 7 lexical units, we collected sentences containing this word from Sogou Corpus and Contemporary Chinese Corpus of Beijing University. After a preliminary screening, about 1000 sentences compose the original and coarse corpus.

Although these sentences were complete and relatively standard, some of them didn't meet the criterion of Chinese frame identification research. Such cases mainly include three aspects. For one thing, the correct frame of ambiguous target word is difficult to decide by human annotator. For the other, the meaning of target word can't correspond to any frame definition in current CFN version. For example, lexical unit "想" can express the meaning of opinion and wish which have the corresponding frames in CFN, while the meaning of thinking and memory did not. Lastly, some words couldn't evoke frames though their word forms are the same as lexical unit. We removed the sentences belonging to the above situations and got a refined corpus containing 940 sentences for training data and 128 for test data. And then, we used LTP to POS tagging and dependency parsing the training and test sentences.

5.2 Experimental Results and Analysis

For the linear-chain CRF, we defined the features based on the words, POS of words and their bigram features as the base features. For T-CRF, we used the base features and tree features. Six different types of template settings on these features are listed in table 2.

template	features
T1	Base features
T2	Add edge between current word and its parent on T1
T3	Add dependency type between current word and its parent on T2
T4	Add edge between current word and its four types children vertexes on T1
T5	Add dependency type between current word and its four types children vertexes on T4
T6	Add all these tree features on T1

Table 2. Template settings on different features

For each of these template settings, we experimented on different observation window size of 1, 2 and 3, which represents one word,

two words and three words previous and next to the current word respectively.

We use the $precision = \dfrac{n}{s}$ to evaluate our system, where n is the number of target words labeled correctly, and s is the total number of target words need to be labeled. In our 128 test sentences, there are 151 target words because there are some sentences containing more than one ambiguous target word. Experimental results on 18 templates are listed in table 3.

From the table 3, we can get four conclusions. Firstly, the best performance 81.46% in T-CRF model increases about 5% over the best performance 76.82% in CRF model. This suggests the dependencies on the tree structure can capture more important characteristics than those on the linear chains do. Secondly, when we added the edge feature between current word and its parent, the performance declined unexpectedly. This can be explained in linguistics: in a dependency parsed sentence, the clique of a governor and its dependents forms "a small world" which can express partial meaning of the sentence, while the parent of current vertex (except the root vertex which has no parent) can not influence much on it because its parent has its own clique, and current word is just a tiny fragment of the clique of its parent, on the contrary, the parent vertex feature will bring negative effect on the current word. For example, the target word "有" in figure 2 can illustrate this case clearly. Thirdly, when we added the children vertexes, the performance increased, that is because current word and its dependents together can form a semantic clique of the sentence. Lastly, when we added the dependency relation type on the features of parent-child edge and child-parent edge, the performance improved slightly because the relation type of edge is coarser than the edge between parent and child. There are only 24 kinds of dependency types but exist hundreds of edge combination possibilities between parent and child. Thus, this feature relived the data sparseness problem to a certain extent.

There are two main types of errors in the results: one is that the labeling frames of target words are not correct. For example, in the sentence "白洁异常强硬地表示，"不获全胜决不收兵"。", the correct frame of "表示"should be "表达" instead of "陈述", because it described the attitude of "白洁" not declared a fact or a phenomena. However, this kind of deep semantics of sentence couldn't be capured by T-CRF model based on the dependency syntax. The other is that the labeling frames of some target words are tag "null". The reason is that some lexical units can't evoke a frame sometimes, so in the training data, these words are annotated "null".

5.3 Contrast Experiments

Qu (2008) argues that any words in a sentence has a certain attraction between each other and thus constitute the grammars and semantic structure of the sentence. Based on this cognition, he proposed a generalized collocation theory, which includes fixed collocation, loose collocation and Co-occurrence collocation. According to this theory, a context computing model RFR_SUM was presented to deal with the WSD task.

In essence, frame identification also belongs to context computing, so it should be reasonable to solve this problem with the generalized collocation theory. However, our current corpus is too insufficient to reflect all these three collocations in the statistical sense. Hence, we proposed a method named compatibility of lexical unit based on the Co-occurrence collocation to identify frame for ambiguous target word.

Window size	Precision					
	T1	T2	T3	T4	T5	T6
1	0.7682	0.7219	0.7351	0.8013	0.8146	0.7947
2	0.7682	0.7152	0.6887	0.7881	0.8146	0.7947
3	0.7417	0.6623	0.6689	0.7351	0.8013	0.7616

Table 3. Precisions of different templates based on three types of window size

The connotation of compatibility of lexical unit is as follows. In the CFN frame database, every frame defines a lexical units set, in which each of lexical unit can evoke this frame. When one of these lexical units serves as the target word in a sentence, we can use the compatibilities of other lexical units in this set with the sentence to reflect the consistency between this frame and current sentence. The compatibility of lexical unit with the sentence is computed by the Co-occurrence frequency of lexical unit and the notional words in the sentence in a large corpus. The calculation is as below.

Suppose l_i in the lexical units set $L = \{l_1, l_2, \ldots l_i, \ldots, l_m\}$ serves as the target word in the sentence S. The words in S except the functional words and l_i constitute a word set $W = \{w_1, w_2, \ldots, w_n\}$. And the compatibility of L with S is denoted as C.

$$C = \frac{c(l_1, W) + c(l_2, W) + \ldots c(l_m, W)}{m},$$

where m is the number of lexical units in L.

$$c(l_j, W) = \frac{f(l_j, w_1) + f(l_j, w_2) + \ldots + f(l_j, w_n)}{n},$$

where n is the number of words in W.

$$f(l_j, w_k) = \frac{count(l_j, w_k)}{sum}, \qquad \text{where}$$

$count(l_j, w_k)$ represents the number of sentences, in which l_j and w_k occur together, and these sentences come from the corpus of Peking University People's Daily, January 1998. sum is the total number of sentences in the same People's Daily corpus.

In this way, the consistency between a frame and the current sentence is scored by the compatibility of L belonging to the candidate frame with this sentence, and the one with highest score is regarded as the correct frame. For our test data, 71.73% precision based on this method was obtained.

This model displayed a decline in precision of about 10% over the T-CRF. Analysis of the results found that the compatibility based on Co-occurrence collocation can only reflect a weak correlation between words, neglecting the position and syntactic structure information in a sentence.

In addition, we used the most-frequency-frame experiment as the baseline. In the corpus consisted of 940 training sentences and 128 test sentences, the frequency of each frame was counted for ranking. The result of this method obtained 61.23% precision, which proved that T-CRF model performed obvious improvement.

6 Conclusions

In this paper, we investigated the problem of Frame Identification in Chinese which is the first work on Chinese FrameNet. A tree-structured conditional random field (T-CRF) model was applied to this task based on the dependency syntactic structure. This model provides a way to incorporating the long-distance dependencies between target words and the syntactic related words with it. In our experiments, the syntactic dependency features were shown to work effectively for Frame Identification, with 71.73%, 76.82%, and 81.46% precision for compatibility of lexical unit, CRF and T-CRF, respectively.

Although a relatively good performance was achieved on the test data, the small-scale and simplicity of sentence structure in corpus cannot be ignored compared with the Frame-Net corpus. However, the experimental results that we gained is still promising, suggesting that our model is comparatively appropriate to the Frame Identification task and still has a great potential for improvement. The next work will focus on the three aspects: firstly, build a larger corpus containing various sentence structures in Chinese; the other is that more semantic features will be tried to add in the T-CRF model, such as the frame elements and the semantic relations between frames, finally, we will try to identify frames of target words using other machine learning methods which has been proved high performance in this task.

thors would like to thank HIT-CIR for their LTP.

References

Charles J. Fillmore. 1982. Frame Semantics. In *Linguistic in the Morning Calm*, pages 111-137, Seoul, Korea: Hanshin Publishing Company.

Collin Baker, Michael Ellsworth and Katrin Erk. 2007. SemEval'07 Task 19: Frame Semantic Structure Extraction. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 99-104, Prague.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In *Proceedings of the COLING-ACL,* pages 86-90, Montreal, Canada.

Cosmin Adrian Bejan and Hathaway Chris. 2007. UTD-SRL: A Pipeline Architecture for Extracting Frame Semantic Structures. In *45th annual meeting of Association for Computational Linguistics*. pages 460-463, Prague.

Igor A. Mel'ˇcuk. 1988. Dependency Syntax: Theory and Practice. *State University Press of New York*, Albany.

John Lafferty, Andrew McCallum and Fernando Pereira. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *proceedings of the 18th International Conference on Machine Learning,* pages 282-289, San Francisco, CA, USA.

Jorge Nocedal and Stephen J. Wright. 1999. *Numerical Optimization*. Springer, New York.

Jun Hatori, Yusuke Miyao and Jun'ichi Tsujii. 2009. On Contribution of Sense Dependencies to Word Sense Disambiguation. *Natural Language Processing*, 16(5):51-77.

Katrin Erk. 2005. Frame assignment as word sense disambiguation. In *Proceedings of the 6th International Workshop on Computational Semantics (IWCS-6)* .

Ken. Litkowski. 2007. CLR: Integration of FrameNet in a Text Representation System. In *45th annual meeting of Association for Computational Linguistics*. pages 113-116, Prague.

Pranjal Awasthi, Aakanksha Gagrani and Balaraman Ravindran. 2007. Image modeling using tree structured conditional random fields. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI 2007)*. Pages 2060-2065.

Qu Weiguang. 2008. Automatic Disambiguation of Modern Chinese Words in Word-level. *Beijing: Science Press(in Chinese)*.

Richard Johansson and Nugues Pierre. 2007. LTH: Semantic Structure Extraction using Nonprojective Dependency Trees. In *45th annual meeting of Association for Computational Linguistics*. pages 227-230, Prague.

Tang Jie, Mingcai Hong, Juanzi Li, and Bangyong Liang. 2006. Tree-structured Conditional Random Fields for Semantic Annotation. In *Proceedings of 5th International Conference of Semantic Web (ISWC'2006)*, Athens, GA, USA

Trevor Cohn and Philip Blunsom. 2005. Semantic role labeling with tree conditional random fields. In *Proceedings of CoNLL2005*.

Wang Ruiqin and Fansheng-Kong. 2009. The Research of Unsupervised Word Sense Disambiguation. *Journal of Software*, (20)8: pages 2138−2152.

Xue Nianwen and Martha Palmer. 2005. Automatic Semantic Role Labeling for Chinese Verbs. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*. Edinburgh, Scotland.

You Liping, Kaiying Liu. 2005. Building Chinese FrameNet database. In *Proceedings of IEEE NLP-KE'05*.