Improving Chinese Semantic Role Labeling using High-quality Surface and Deep Case Frames

Gongye Jin*, Daisuke Kawahara, Sadao Kurohashi
Graduate School of Informatics, Kyoto University
Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501, Japan
jin@nlp.ist.i.kyoto-u.ac.jp, {dk, kuro}@i.kyoto-u.ac.jp

Abstract

This paper presents a method for improving semantic role labeling (SRL) using a large amount of automatically acquired knowledge. We acquire two varieties of knowledge, which we call surface case frames and deep case frames. Although the surface case frames are compiled from syntactic parses and can be used as rich syntactic knowledge, they have limited capability for resolving semantic ambiguity. To compensate the deficiency of the surface case frames, we compile deep case frames from automatic semantic roles. We also consider quality management for both types of knowledge in order to get rid of the noise brought from the automatic analyses. The experimental results show that Chinese SRL can be improved using automatically acquired knowledge and the quality management shows a positive effect on this task.

1 Introduction

Semantic role labeling (SRL) is regarded as a task that is intermediate between syntactic analysis and semantic analysis in natural language processing (NLP). The main goal of SRL is to extract a proposition from a sentence about *who* does *what* to *whom*, *when, where* and *why*. By using semantic roles, the complex expression of a sentence is then interpreted as an *event* and its *participants* (i.e., a predicate and arguments such as *agent, patient, locative, temporal* and *manner*). Unlike syntactic level surface cases (i.e., dependency labels such as subject and object), semantic roles can be regarded as a deep case representation for predicates. Because of its ability to abstract the meaning of a sentence, SRL has been applied to many NLP applications, including information extraction (Christensen et al., 2010), question answering (Pizzato and Mollá, 2008) and machine translation (Liu and Gildea, 2010).

Semantically annotated corpora, such as FrameNet (Fillmore et al., 2001) and PropBank (Kingsbury and Palmer, 2002), make this type of automatic semantic structure analysis feasible by using supervised machine learning methods. However, supervised SRL methods have the following two major issues. Firstly, as a common issue in almost all the supervised approaches, it is expensive to enlarge manually annotated corpora to learn a more accurate model. Secondly, experiments show that automatic SRL systems strongly depend on syntactic information. In practice, these SRL systems suffer from errors propagated from the lower-level syntactic analyses, such as word segmentation, POS tagging, and dependency parsing. Although some studies use automatic analyses of unlabeled corpora to enrich the training data to solve the first problem (Fürstenau and Lapata, 2009), accumulated errors in such automatic analysis inevitably cause negative effects. Especially, for some hard-to-analyze languages, such as Chinese, which is still difficult to precisely analyze word segmentations, the performance of SRL is always limited due to the above two problems.

In this paper, we focus on Chinese SRL and address the problems mentioned above by using high-quality knowledge automatically acquired from a large-scale raw corpus. We utilize two types of additional knowledge. The first type is compiled using automatic syntactic analysis (specifically, dependency parsing) and is named **surface case frames** which are not expressive in

*The first author is now affiliated with Canon IT Solutions Inc.

Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, pages 568–577,
Valencia, Spain, April 3-7, 2017.

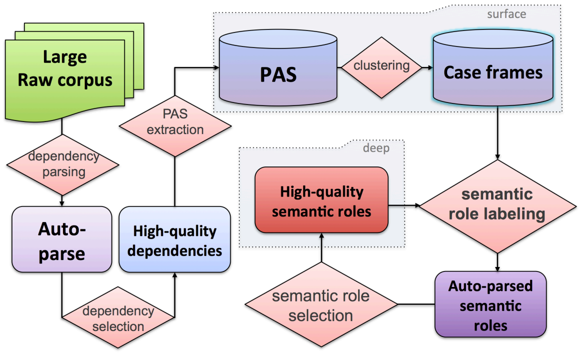

Figure 1: Overview of the framework

semantic level. In order to compensate the drawback of surface case frames, we also compile another type of knowledge using automatic semantic roles. We call this type of knowledge **deep case frames**. We illustrate the whole framework in Figure 1. The additional knowledge can provide not only syntactic information but also semantic information, both of which play crucial roles in SRL. Considering the inevitable noises from automatic analyses, we utilize an automatic selection method to select dependencies and semantic roles of high quality. In order to show that automatically extracted knowledge is beneficial and the quality management is indispensable, we compile both types of knowledge in different quality in our experiments and apply them to Chinese SRL.

2 Related work

The CoNLL-2009 shared task (Hajič et al., 2009) features a substantial number of studies on SRL that used Propbank as one of the resources. The participating systems can be categorized into two types: joint learning of syntactic parsing and SRL (Tang et al., 2009; Morante et al., 2009), which learns a unique model for syntactic parsing and SRL jointly. This type of framework has the ability to use SRL information in syntactic parsing for improvement, but needs a much larger search space for decoding. The other type is called SRL-only task (Zhao et al., 2009; Björkelund et al., 2009), which uses automatic morphological and syntactic information as the input in order to judge which token plays what kind of semantic role. Our work focuses on the second category of SRL. Our framework is based on those used by Björkelund et al. (2009) and Yang and Zong (2014).

There were several studies using additional knowledge to improve syntactic and semantic tasks. McClosky et al. (2006) used an additional unlabeled corpus to reduce data sparsity. In syntactic level of NLP, rich knowledge, such as predicate-argument structures and case frames, is strong backups for various kinds of tasks. Case frames, which clarify relations between a predicate and its arguments, can support tasks ranging from fundamental analysis, such as dependency parsing and word similarity calculation, to multilingual applications, such as machine translation. Japanese case frames have been successfully compiled (Kawahara and Kurohashi, 2006), where each case slot is represented as its case marker in Japanese such as 'ga', 'wo', and 'ni'. For the case frames of other languages such as English and Chinese, because there are no such case markers that can help clarify syntactic structures, instead of using case markers like in Japanese, syntactic surface cases (i.e., subject, object, prepositional phrase, etc.) are used for argument representation (Jin et al., 2014). Case frames can be automatically acquired using a different method such as Chinese Restaurant Process (CRP) (Kawahara et al., 2014) for different languages. In our work, we employ such syntactic level knowledge, which uses surface cases as argument representation, to help SRL.

One basic idea of semi-supervised SRL is to automatically annotate unlabeled data using a simple classifier trained on original training data (Fürstenau and Lapata, 2009). Since there is a substantial amount of error propagation in the SRL pipeline, the additional automatic semantic roles are not guaranteed to be of good quality. Also, some studies assume that sentences that are syntactically and lexically similar are likely to share the same frame-semantic structure (Fürstenau and Lapata, 2009). This allows them to project semantic role information to unlabeled sentences using alignments. However, the computation of these alignments requires additional information such as word similarity, whose quality is language dependent. Less sparse features capturing lexical information of words can be also used for semi-supervised learning of SRL. Such lexical representation can be learned from unlabeled data (Bengio et al., 2003). Deschacht and Moens (2009) used word similarity learned from unlabeled data as additional features for SRL. Word embeddings have also been used in several NLP tasks including SRL (Collobert et al., 2011). Instead of using word-level lexical knowledge, our work uses syntactic and semantic knowledge, i.e., case frames. Word embeddings can also be incorporated into our method but we leave this to our future work. Zapirain et al. (2009) used selectional preferences to improve SRL. This study is similar to our approaches but the quality of selectional preferences was not concerned at all.

3 SRL task description

In previous studies, SRL pipeline[1] can be divided into three main steps: predicate disambiguation (PD), argument identification (AI), and argument classification (AC). In the PD step, the main goal is to identify the "sense id" of each given predicate. The AI step mainly focuses on judging whether each argument is semantically related to each predicate in a sentence. Based on the results of the AI step, the AC step assigns a semantic role to each semantically related argument. Basically, the PD step and the AC step are regarded as multi-class classification problems while the AI step is a binary classification problem.

In the PD step, because the sense id for a certain predicate is meaningless for other predicates, classifiers for PD are trained separately for each predicate. We basically use the feature set proposed by Björkelund et al. (2009). During the prediction, there are some predicates which have not been seen in the training data. We label the sense of those unseen predicates using the default sense, which is '01' in our work.

4 Applying high-quality surface case frames to SRL

4.1 High-quality dependency selection

Dependency parsing has been widely employed for knowledge acquisition related to predicate-argument structures. The dependency parsing performance determines the quality of acquired knowledge, regardless of target languages. Reducing dependency parsing errors and selecting high-quality dependencies are of primary importance. Jin et al. (2013) used a single set of dependency labeled corpus (a treebank), a part of which was used to train a base dependency parser. Another part of the labeled corpus was used to apply automatic dependency parsing. By comparing the gold standard data and the automatic parses, correct dependencies were collected as positive examples and incorrect dependencies were collected as negative examples. Then selecting high-quality dependencies was regarded as a binary classification problem. To conduct such binary classification, they employed a set of basic features from Yu et al. (2008). In addition to these basic features, Jin et al. (2013) considered context features that are thought to affect parsing performance. Since the input for high-quality dependency selection method is a dependency tree, tree features are used to identify dependency quality. Also, some dependency parsers output the score of each dependency (i.e., edge confidence value) during the parsing process. They used the real value of the score as an additional feature. We first apply this approach to select high-quality dependencies from automatic parses.

4.2 High-quality surface case frame construction

After applying dependency parsing on a large-scale raw corpus, predicate-argument structures (PASs) are extracted using the high-quality dependencies. Arguments are represented by their dependency labels (i.e., subject, object, etc.) For each predicate, all the PASs are clustered into different case frames to reflect different semantic usages. We show an example of case frames for the verb '谢' in Table 1, which has multiple meanings. '谢(1)' is the case frame used to represent the sense of 'withering of flower'. Similarly, the sense of '谢' which means 'to thank' is represented by case frame '谢(2)'. '谢(3)' is the case frame for the sense of 'curtain call'. In other words, case frames are knowledge that solves word sense disambiguation (WSD) by clustering the PASs. We applied the CRP method described by (Kawahara et al., 2014) for clustering the high-quality PASs to compile high-quality case frames.

4.3 Surface case features for SRL

From the surface case PASs, we extract four types of additional features, for both AI and AC step. These features are described in the upper part of Table 2. We use binned values (i.e., high, middle and low) for all of the feature values calculated from the knowledge. More specifically, for each type of feature, we define the first, second and third tertile of all the feature values as low, middle and high correspondingly. Surface case

[1] Predicate identification (PI) was not concerned in the experiments because we use the data from CoNLL-2009 shared task, in which the target predicates are given.

verb	surface case	instance with frequency in original corpus
谢(1)	nsubj	花儿(flower):14, 花(flower):22
	ad	都(all):16, 也(also):6
谢(2)	nsubj	你们(you):1
	dobj	您(you):8, 我(me):6
	ad	怎么(how):8, 多(very):1
谢(3)	nsubj	大战(battle):1
	dobj	幕(curtain):6
	ad	圆满(successfully):2, 也(also):1, 正式(officially):1
...		

Table 1: Examples of Chinese surface case frames

frames are clustered PASs according to each predicate's semantic usage. Therefore, instead of utilizing all the predicate-argument structures, it is intuitive to use the predicate-argument structures only from the corresponding case frames. So we also create four types of features extracted from case frames. These features are listed in the lower part of Table 2.

Note that a case frame ID and a PropBank sense ID do not correspond to each other. In practice, the number of case frames is always larger than the number of sense in PropBank for each verb. As a result, a mapping process that aligns case frame id(s) to PropBank verb sense is applied. Fisrt, we assign automatic dependency labels to the PropBank corpus using the Stanford parser. We then calculate the similarity between a PropBank sense and a case frame by measuring the PAS similarity. As shown in the left part of Figure 2, for a certain predicate with a sense ID in PropBank, we represent the predicate in each sense by using the collection of all the instances in each syntactic role slot. Each predicate with a sense ID is then transformed into a vector space, which we name PAS vector. The same transformation is applied to case frames. Then the cosine similarity between vectors transformed from a PropBank sense and case frames is calculated. A PAS vector is the concatenation of each syntactic role vector. To form a syntactic role vector, we simply take the average of weighted summation of the word vectors within the case slot. Word vectors are acquired using word2vec[2] from the same raw corpus that we use for knowledge acquisition (see Section 7.1). In our experiments, we only used syntactic role "subj" (subject) and "dobj" (direct object) because these two syntactic roles are considered to be relatively more informative.

[2]`https://code.google.com/archive/p/word2vec/source/default/source`

5 Main problem of surface case frames

In previous work (Kawahara and Kurohashi, 2006), case frames for Japanese are composed of all the instances and their corresponding case marker. For example, all the instances in "ga" case are basically the "subject" of the given predicate. Instances in "wo" case are basically the "direct object" of the given predicate. Other cases like "ni" can indicate "location", "time" or "direction". During the automatic PAS extraction for Japanese, there are also ambiguous case makers that can represent multiple cases. The most common one, for example, is "wa" case in Japanese. This case marker always functions as a topic marker. The argument in "wa" case is normally emphasized as the topic of the sentence. It can be equal to either "ga" case or "wo" case, and sometimes "ni" case. To avoid such ambiguous cases, one can simply discard all the instances in "wa" case to make case frames more precise.

For languages that lack such case markers (e.g., English and Chinese), case frames are composed of automatic syntactic roles (Jin et al., 2014). Such syntactic roles include "subject", "direct object", "indirect object" and "prepositional phrases". Such surface cases have limitations on case representation especially for Chinese. Take the following sentences as examples.

(1) 苹果 (apples) / 我 (I) / 吃了 (eaten) / 很多 (a lot).
(2) 我 (I) / 苹果 (apples) / 吃了 (eaten) / 很多 (a lot).
(3) 我 (I) / 吃了 (eaten) / 很多(a lot) / 苹果 (apples).
(4) 我 (I) / 吃了 (eaten) / 很多(a lot).

feature	description
Freq	the co-occurrence frequency of a predicate-argument pair without considering the syntactic role of the argument
Pmi	the point-wise mutual information (PMI) value for each predicate-argument pair without considering the syntactic role of the argument
PAfreq	the frequency of a argument being a certain syntactic role of a predicate
PApmi	the PMI value of an argument with its syntactic role and the predicate
CFFreq	Freq value calculated only from within the corresponding case frames
CFPmi	Pmi value calculated only from within the corresponding case frames
CFPAfreq	PAFreq value calculated only from within the corresponding case frames
CFPApmi	PApmi value calculated only from within the corresponding case frames

Table 2: Surface case features for SRL

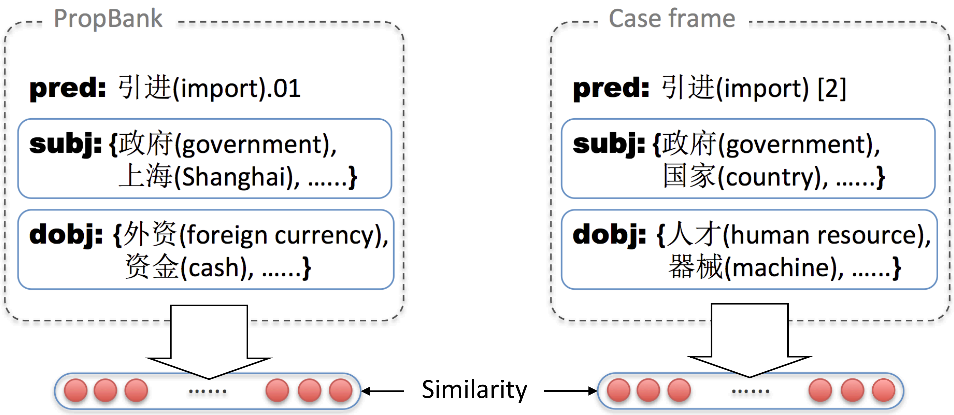

Figure 2: Overview of mapping case frames to PropBank sense

(5) 苹果 (apples) / 吃了 (eaten) / 很多 (a lot).

The first three sentences have the same meaning: "I have eaten a lot of apples." However, as we can see from the sentences, the word "苹果 (apple)" which is a direct object of "吃了 (eaten)", and the word "我 (I)" can be filled in various word orders. Also, because omissions occur frequently in Chinese, sentence 4 and 5 are also commonly used, which mean "I have eaten a lot" and "(I) have eaten a lot of apples", respectively. Without considering the actual meaning of "我 (I)" and "苹果 (apples)", both of them in sentence 4 and 5 are labeled as "subject" in the surface case representation following the syntactic grammar. If one tries to figure out which "subject" is actually in *Nominative Case* (which stands for the person/thing who provides the action) and which "subject" is in *Accusative Case* (which stands for the thing/person who receives/suffers from the action), it is always problematic because of the flexible word order and omission.

Although some studies found that applying simple mapping rules for *Nominative Case* and *Accusative Case* can achieve an overall high baseline for English, we found that this simple mapping cannot work well for Chinese. Here is an example which Chinese people are using for self-deprecating.

(6) 中国 (Chinese) / 乒乓球 (table tennis) / 谁 (who) / 也 (ever) / 赢 (win) / 不了 (not): Nobody can win Chinese table tennis.

(7) 中国 (Chinese) / 足球 (soccer) / 谁 (who) / 也 (ever) / 赢 (win) / 不了 (not): Chinese soccer cannot win anybody.

"Table tennis" and "soccer" should be labeled as *Accusative Case* and *Nominative Case* differently even though the predicate and the syntactic structure for both the sentences are identical.

Similar phenomena also occur in Japanese and make it difficult to analyze as well. However, in case of Japanese, it is possible to make use of the morphemes attached to the predicate. For example, the following sentences are the Japanese translations for sentence 4 and 5.

(8) 私が (I) / たくさん (a lot) / 食べた (eaten).
(9) りんごが (apples) / たくさん (a lot) / 食べられた (eaten).

There is always an additional morpheme (e.g., "られた") attached to the predicate in order to indicate its voice. In the above example, sentence 8 can be regarded as active voice and sentence 9 is in passive voice. Unfortunately, Chinese is a language that lacks morpheme information. There are very few such markers that indicate the transitivity, voice and tense. This makes it almost impossible for a system to automatically recognize the ambiguous syntactic roles. To solve this problem, based on the syntactic analysis, we apply an SRL process to discover a deeper level case representation.

6 Applying high-quality deep case frames to SRL

6.1 High-quality semantic role selection

Similar to the previous work described in Jin et al. (2013), instead of using all the SRL outputs, we propose to use only automatically generated semantic roles of high quality.

In particular, the standard training section of the human-annotated data is used to train a base SRL model (which include three sub-models for predicate sense disambiguation (PD), argument identification (AI) and argument classification (AC)). Then, another part of the human-annotated data is used to apply SRL using the base model. From these results, we acquire training data for semantic role selection by collecting each semantic role. We then judge the correctness of each semantic role according to the gold standard annotations. All correct semantic roles are used as positive examples and the incorrect ones are used as negative examples for semantic role selection. Judging whether an automatic semantic role is reliable can be regarded as a binary classification problem. We use SVMs to solve this problem. We use the feature set for SRL described in Jin et al. (2015) as basic features. It contains predicate features that are extracted from the target predicate; argument features which are extracted from each candidate argument. We also use surface case frames, which have a positive effect on SRL, as additional knowledge.

6.2 High-quality deep case frame construction

Due to the major issues described in Section 5, case frames constructed using surface cases may be problematic. For example, for the predicate "吃 (eat)", both the argument "苹果 (apple)" and "我 (I)" are assigned to the same surface case "subject". If one tries to use this kind of surface case knowledge for tasks that require semantic information, such as machine translation (MT), it may lead to a performance drop. So we propose to construct deep case frames that are relatively more representative than the surface case frames. By the deep case, we mean using the semantic roles for case frame construction.

Compared to syntactic analysis, SRL is mainly used to clarify deeper-level semantic relations (e.g., [*who*] do [*what kind of*] thing to [*whom*] in [*what time*]) in the sentence, which has a better representation for predicate-argument relations. On the other hand, this task is always based on the tasks in preceding levels, such as morphological analysis and syntactic parsing. Especially, the information provided by syntactic parsing is crucial to achieve a good performance in SRL. An SRL system also suffers from the training data size issue as most of the machine learning approaches do. Extensive human efforts are required in order to construct such training data. Sometimes, the requirements for annotators can be higher than those for syntactic analysis. These factors along with the automatic analysis errors propagated from the lower-level analyses make it almost impossible for an SRL system to achieve a high performance.

For predicate identification (PI), we regard every word with a POS tag begining with "V" as a predicate. The PD step in the SRL pipeline assigns a sense ID (frame ID) to each predicate. This is equivalent to the unsupervised clustering for surface case frames and thus no additional clustering process is required. After argument identification and argument classification, we only use these semantic roles with high reliability. For each predicate with different frame IDs, we collect all the high-quality semantic roles to compose the deep case frames.

6.3 Using high-quality deep case frames for SRL

Syntactic information such as dependencies is essential for SRL. In Section 4, we used surface

feature	description
SRFreq	the co-occurrence frequency of a predicate-argument pair without considering the semantic role of the argument
SRPmi	the PMI value for each predicate-argument pair without considering the semantic role of the argument
SRPAfreq	the frequency of a argument being a certain semantic role of a predicate
SRPApmi	the PMI value of an argument with its semantic role and the predicate
DCFFreq	SRFreq value calculated only from within the corresponding deep case frame
DCFPmi	SRPmi value calculated only from within the corresponding deep case frame
DCFPAfreq	SRPAfreq value calculated only from within the corresponding deep case frame
DCFPApmi	SRPApmi value calculated only from within the corresponding deep case frame

Table 3: Deep case features for SRL

case frames to provide additional knowledge especially syntactic-level knowledge, for an SRL system and gained a slight improvement as shown in Section 7. Deep case frames are compiled using automatic semantic roles that use semantic-level representation. Thus, we consider that using deep case frames as additional knowledge has a more direct impact on the performance on SRL. Similar to the method described in Section 4, we also propose a set of features extracted from deep case frames which are listed in Table 3. The first four features do not concern the predicate sense. These features are similar to the predicate-argument pair features described in Section 4. The rest four features are similar to the case frame features described in Section 4. However, because the deep case frame ID is identical to the PropBank ID, no mapping processes are needed.

7 Experiments

7.1 Experimental settings

For large-scale knowledge acquisition, 40 million sentences from Chinese Gigaword 5.0 (LDC2011T13)[3] were used.

For the high-quality dependency selection approach in the knowledge construction pipeline, the Stanford parser was used to apply dependency parsing. The training section of Chinese Treebank 7.0 was used to train the dependency parser and the official development section was used to train a classifier for high-quality dependency selection. Using the official evaluation section of CTB 7.0, we evaluated the quality of those selected dependencies using unlabeled attachment score (UAS), which calculates the percentage of correctly identified dependency heads.

For SRL, we used the Chinese section of CoNLL-2009 shared task data (we substituted the syntactic dependencies and dependency labels produced by the Stanford parser). Automatically obtained morphological and syntactic information (the columns begin with "P") was used. PD and AI, AC step are regarded as multi-class classification problems. We employed OPAL[4] to solve these problems. We set the options as follows: polynomial kernel with degree 2; passive aggressive I learner; 20 iterations. The base SRL system without using additional knowledge was used as a baseline. To examine the effect of different quality of knowledge, we used different set of PASs which were extracted under different dependency selection thresholds (20%, 50%, and 100%). The official script provided on the CoNLL-2009 shared task website was used for evaluation.

For semantic role selection, similar to dependency selection, the training section of CoNLL-2009 shared task data was used to train the base SRL model. The development section in CoNLL-2009 shared task data was used to apply automatic SRL and obtain training data for the semantic role selector. We evaluated the semantic role selection approach by calculating the percentage of correctly judged semantic roles (predicate senses are not counted). For deep case frame construction, we used the Stanford parser for syntactic analysis.

[3]We only used sentences written in simplified characters in Chinese Gigaword.

[4]http://www.tkl.iis.u-tokyo.ac.jp/~ynaga/opal/

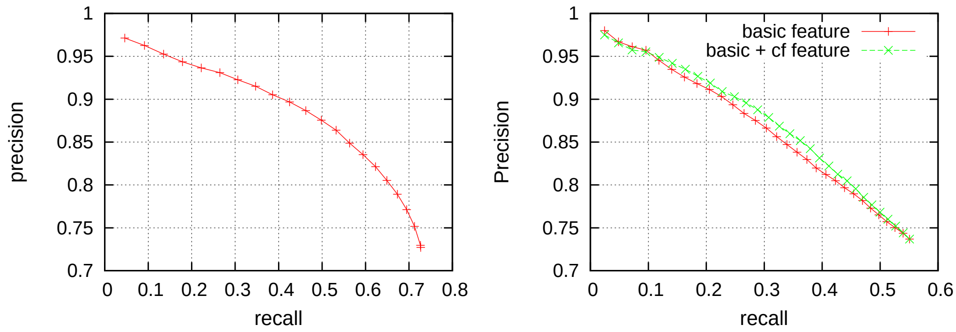

Figure 3: Precision-recall curve of dependency selection & semantic role selection

method	precision	recall	F1
baseline	80.66%	72.98%	76.63
baseline + surface case frames (100%)	79.86%	72.72%	76.12
baseline + surface case frames (50%)	80.40%	73.04%	76.54
baseline + surface case frames (20%)	80.73%	73.32%	****76.85**
baseline + deep case frames (100%)	81.22%	73.55%	**77.19
baseline + deep case frames (50%)	81.30%	73.70%	**77.31
baseline + deep case frames (20%)	81.57%	73.68%	****77.42**

Table 4: Evaluation results of Chinese SRL using surface and deep case frames. The ** mark and * mark mean that the result is regarded as significant (with a p value < 0.01 and a p value < 0.05, respectively) using McNemar's test.

The base SRL system was used to assign semantic roles. We applied the proposed framework to 40 million sentences from Chinese Gigaword 5.0. We constructed deep case frames of different quality (20%, 50%, and 100%) to extract extra features to support the base SRL system.

7.2 Experimental results

Figure 3 shows the precision-recall curves of dependency selection and semantic role selection. For dependency selection, we achieved a precision over 90% when lowering the recall to around 20%. For semantic role selection, using additional surface case frame features gains a slight improvement compared to the basic features.

Table 4 shows the experimental results of SRL using surface and deep case frames as additional features. Knowledge (n%) indicates that the top n% (according to the classifier) of the automatically extracted knowledge was used. '100%' means that the selection step was not applied. It is worth pointing out that when using the baseline method, we achieved an F-value of around 79.4 on CoNLL-2009 shared task original data set (where the dependency labels follow the MaltParser style, which is different from the Stanford dependencies). This result has outperformed the best system for Chinese SRL in CoNLL-2009 shared task, which was 78.60. When applying the baseline system on the substituted version of dataset for dependency label consistency with the additional knowledge, the baseline F-value drops to 76.63. As we can see from the results, using deep case frames gained more improvements than using surface case frames. This is because deep case frames are able to directly provide semantic-level information that is insufficient in the training data of the base SRL system. Furthermore, the results show that the high-quality semantic role selection approach has a positive effect on SRL.

8 Conclusion & future work

We proposed a method for using additional knowledge to improve Chinese SRL. To address the case ambiguity problem in the surface case representation, especially for Chinese, we utilized automatic semantic roles produced by an SRL system for a better representation. The experimental results showed a promising result for high-quality semantic role selection. Also, using high-quality deep case frames that are composed of semantic roles can significantly improve the baseline SRL system.

We plan to make use of other low-level knowledge such as word embeddings (Collobert et al., 2011) and word clusters (Koo et al., 2008) to improve dependency parsing and SRL. The recent SRL approaches are mostly point-wise. Features are extracted from only pairs of the predicate and an argument candidate. We plan to design a higher-order system to capture more global features following the idea of higher-order dependency parsing. Also, reranking is widely utilized in many SRL systems and we plan to combine our surface/deep case knowledge with a reranker in order to further improve Chinese SRL.

References

Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. 2003. A neural probabilistic language model. 3:1137–1155, February.

Anders Björkelund, Love Hafdell, and Pierre Nugues. 2009. Multilingual semantic role labeling. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009): Shared Task*, pages 43–48, Boulder, Colorado, June. Association for Computational Linguistics.

Janara Christensen, Mausam, Stephen Soderland, and Oren Etzioni. 2010. Semantic role labeling for open information extraction. In *Proceedings of the NAACL HLT 2010 First International Workshop on Formalisms and Methodology for Learning by Reading*, pages 52–60, Los Angeles, California, June. Association for Computational Linguistics.

Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. 12:2493–2537, August.

Koen Deschacht and Marie-Francine Moens. 2009. Semi-supervised semantic role labeling using the latent words language model. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 21–29, Singapore, August. Association for Computational Linguistics.

Charles J. Fillmore, Charles Wooters, and Collin F. Baker. 2001. Building a large lexical databank which provides deep semantics. In Benjamin Tsou and Olivia Kwong, editors, *Proceedings of the 15th Pacific Asia Conference on Language, Information and Computation*, Hong Kong.

Hagen Fürstenau and Mirella Lapata. 2009. Semi-supervised semantic role labeling. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 220–228, Athens, Greece, March. Association for Computational Linguistics.

Jan Hajič, Massimiliano Ciaramita, Richard Johansson, Daisuke Kawahara, Maria Antònia Martí, Lluís Màrquez, Adam Meyers, Joakim Nivre, Sebastian Padó, Jan Štěpánek, Pavel Straňák, Mihai Surdeanu, Nianwen Xue, and Yi Zhang. 2009. The conll-2009 shared task: Syntactic and semantic dependencies in multiple languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009): Shared Task*, pages 1–18, Boulder, Colorado, June. Association for Computational Linguistics.

Gongye Jin, Daisuke Kawahara, and Sadao Kurohashi. 2013. High quality dependency selection from automatic parses. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 947–951.

Gongye Jin, Daisuke Kawahara, and Sadao Kurohashi. 2014. A framework for compiling high quality knowledge resources from raw corpora. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 109–114.

Gongye Jin, Daisuke Kawahara, and Sadao Kurohashi. 2015. Chinese semantic role labeling using high-quality syntactic knowledge. In *Proceedings of the Eighth SIGHAN Workshop on Chinese Language Processing*, pages 120–127, Beijing, China, July. Association for Computational Linguistics.

Daisuke Kawahara and Sadao Kurohashi. 2006. A fully-lexicalized probabilistic model for Japanese syntactic and case structure analysis. In *Proceedings of HLT-NAACL 2006*, pages 176–183.

Daisuke Kawahara, Daniel Peterson, Octavian Popescu, and Martha Palmer. 2014. Inducing example-based semantic frames from a massive amount of verb uses. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pages 58–67, Gothenburg, Sweden, April. Association for Computational Linguistics.

Paul Kingsbury and Martha Palmer. 2002. From treebank to propbank. In *Language Resources and Evaluation*.

Terry Koo, Xavier Carreras, and Michael Collins. 2008. Simple semi-supervised dependency parsing. In *Proceedings of ACL-08: HLT*, pages 595–603, Columbus, Ohio, June. Association for Computational Linguistics.

Ding Liu and Daniel Gildea. 2010. Semantic role features for machine translation. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 716–724, Beijing, China, August. Coling 2010 Organizing Committee.

David McClosky, Eugene Charniak, and Mark Johnson. 2006. Reranking and self-training for parser adaptation. In *Proceedings of the 21st International Conference on Computational Linguistics and*

44th Annual Meeting of the Association for Computational Linguistics, pages 337–344, Sydney, Australia, July. Association for Computational Linguistics.

Roser Morante, Vincent Van Asch, and Antal van den Bosch. 2009. Joint memory-based learning of syntactic and semantic dependencies in multiple languages. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009): Shared Task*, pages 25–30, Boulder, Colorado, June. Association for Computational Linguistics.

Luiz Augusto Pizzato and Diego Mollá. 2008. Indexing on semantic roles for question answering. In *Coling 2008: Proceedings of the 2nd workshop on Information Retrieval for Question Answering*, pages 74–81, Manchester, UK, August. Coling 2008 Organizing Committee.

Buzhou Tang, Lu Li, Xinxin Li, Xuan Wang, and Xiaolong Wang. 2009. A joint syntactic and semantic dependency parsing system based on maximum entropy models. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009): Shared Task*, pages 109–113, Boulder, Colorado, June. Association for Computational Linguistics.

Haitong Yang and Chengqing Zong. 2014. Multi-predicate semantic role labeling. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 363–373, Doha, Qatar, October. Association for Computational Linguistics.

Kun Yu, Daisuke Kawahara, and Sadao Kurohashi. 2008. Cascaded classification for high quality head-modifier pair selection. In *Proceedings of NLP 2008*, pages 1–8.

Beñat Zapirain, Eneko Agirre, and Lluís Màrquez. 2009. Generalizing over lexical features: Selectional preferences for semantic role classification. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 73–76, Suntec, Singapore, August. Association for Computational Linguistics.

Hai Zhao, Wenliang Chen, Chunyu Kity, and Guodong Zhou. 2009. Multilingual dependency learning: A huge feature engineering method to semantic dependency parsing. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL 2009): Shared Task*, pages 55–60, Boulder, Colorado, June. Association for Computational Linguistics.