Unified Semantic Role Labeling for Verbal and Nominal Predicates in the Chinese Language

JUNHUI LI and GUODONG ZHOU, Soochow University

This article explores unified semantic role labeling (SRL) for both verbal and nominal predicates in the Chinese language. This is done by considering SRL for both verbal and nominal predicates in a unified framework. First, we systematically examine various kinds of features for verbal SRL and nominal SRL, respectively, besides those widely used ones. Then we further improve the performance of nominal SRL with various kinds of verbal evidence, that is, merging the training instances from verbal predicates and integrating various kinds of features derived from SRL for verbal predicates. Finally, we address the issue of automatic predicate recognition, which is essential for nominal SRL. Evaluation on Chinese PropBank and Chinese NomBank shows that our unified approach significantly improves the performance, in particular that of nominal SRL. To the best of our knowledge, this is the first reported work of unified verbal and nominal SRL on Chinese PropBank and NomBank.

13

Categories and Subject Descriptors: I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Language parsing and understanding*

General Terms: Documentation, Languages

Additional Key Words and Phrases: Semantic role labeling, unified approach, verbal predicates, nominal predicates, automatic predicate recognition

ACM Reference Format:
Li, J. and Zhou, G. 2011. Unified semantic role labeling for verbal and nominal predicates in chinese language. ACM Trans. Asian Lang. Inform. Process. 10, 3, Article 13 (September 2011), 21 pages.
DOI = 10.1145/2002980.2002983 http://doi.acm.org/10.1145/2002980.2002983

1. INTRODUCTION

Semantic parsing maps a natural language sentence into a formal representation of its meaning. Due to the difficulty in deep semantic parsing, most of the previous work focuses on shallow semantic parsing, which assigns a simple structure (such as WHO did WHAT to WHOM, WHEN, WHERE, WHY, or HOW) to each predicate in a sentence. In particular, the well-defined semantic role labeling (SRL) task has drawn more and more attention in recent years due to its importance in many natural language processing (NLP) applications and techniques, such as question answering [Moschitti and Quarteroni 2010; Moschitti et al. 2007; Surdeanu et al. 2008], information extraction [Surdeanu et al. 2003], and co-reference resolution [Ponzetto and Strube 2006]. Given a sentence and a predicate (either a verb or a noun) in the sentence, SRL recognizes

This research is supported by Project 60873150, 60970056 and 90920004 under the National Natural Science Foundation of China, and the National Research Foundation for the Doctoral Program of Higher Education of China under Grant No. 20080285006.
Authors' address: J. Li and G. Zhou (corresponding author), School of Computer Science & Technology, Soochow University, 1 Shizi Street, Suzhou, 215006, China; email: {lijunhui, gdzhou}@suda.edu.cn.

DOI 10.1145/2002980.2002983 http://doi.acm.org/10.1145/2002980.2002983

and maps all the constituents in the sentence into their corresponding semantic arguments (roles) of the predicate. According to the predicate types, SRL could be divided into SRL for verbal predicates (verbal SRL, in short) and SRL for nominal predicates (nominal SRL, in short).

During the past few years, verbal SRL has dominated the research on SRL with the availability of the FrameNet project [Baker et al. 1998], the PropBank project [Palmer et al. 2005], and the consecutive CoNLL shared tasks [Carreras and Màrquez 2004, 2005] in English language. As a complement to PropBank on verbal predicates, NomBank [Meyers et al. 2004] annotates nominal predicates and their corresponding semantic roles using similar semantic framework as PropBank. For example, Jiang and Ng [2006] pioneered the exploration of various nominal SRL-specific features besides the traditional verbal SRL-related features on NomBank. Given gold nominal predicates, they achieved the performance of 72.73 and 69.14 in F1-measure on gold and automatic syntactic parse trees, respectively. Instead of relying on manually annotated nominal training data, Padó et al. [2008] presented a data expansion approach to SRL for event nominalizations by harnessing annotated data for verbs to bootstrap a semantic role labeler for nouns.

For SRL in Chinese, Sun and Jurafsky [2004] and Pradhan et al. [2004] pioneered the research on Chinese verbal and nominal SRL, respectively, on small private datasets. Taking advantage of recent release of Chinese PropBank [Xue and Palmer 2003] and Chinese NomBank [Xue 2006a], Xue and his colleagues [Xue 2006b, 2008; Xue and Palmer 2005] pioneered the exploration of Chinese verbal and nominal SRL given gold predicates. Among them, Xue and Palmer [2005] studied Chinese verbal SRL using Chinese PropBank and achieved the performance of 91.3 and 61.3 in F1-measure on gold and automatic syntactic parse trees, respectively. Xue [2006b] extended the study on Chinese nominal SRL and attempted to improve the performance of nominal SRL by simply including the Chinese PropBank training instances into the training data for nominal SRL on Chinese NomBank. However, such integration was empirically proven unsuccessful due to the different nature of certain features for verbal and nominal SRL. Xue [2008] further improved the performance on both verbal and nominal SRL with a better syntactic parser and additional features. Ding and Chang [2008] focused on argument classification for Chinese verbal predicates with a hierarchical feature selection strategy. They achieved the classification precision of 94.68% on gold parse trees on Chinese PropBank.

This article focuses on Chinese SRL for both verbal and nominal predicates. First, we systematically explore a large feature space to achieve a state-of-the-art performance. Then we further improve the performance of nominal SRL with various kinds of verbal evidence, that is, merging the training instances from verbal predicates and integrating various kinds of features derived from SRL for verbal predicates. Finally, we investigate the effect of automatic predicate recognition on the performance of Chinese SRL. Although previous research (e.g., CoNLL'2008) in English SRL reveals the importance of automatic predicate recognition, there has been no reported research on automatic predicate recognition in Chinese SRL for both verbal and nominal predicates.

The rest of this article is organized as follows: Section 2 introduces Chinese PropBank and NomBank while the baseline verbal and nominal SRL systems are described in Section 3 with widely-used standard and additional features. Then, two ways to improve nominal SRL with verbal evidence are explored in Section 4 while automatic predicate recognition is examined in Section 5. Section 6 gives experimental results. Finally, Section 7 concludes the article and presents the future work.

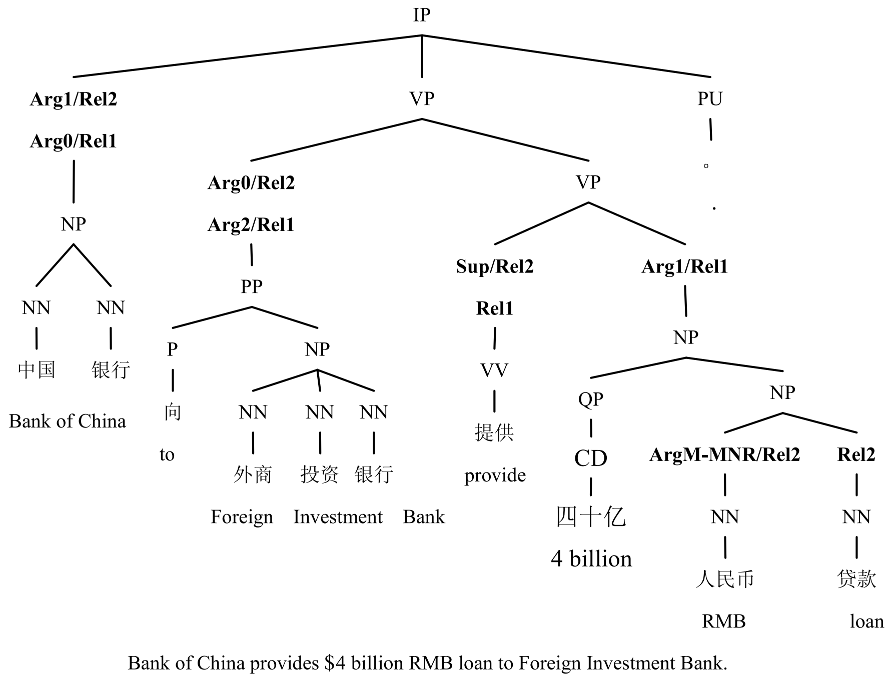

Fig. 1. Two predicates (Rel1 and Rel2) and their arguments in the style of Chinese PropBank and NomBank.

2. CHINESE PROPBANK AND NOMBANK

Chinese PropBank [Xue and Palmer 2003] and Chinese NomBank [Xue 2006a] adopt similar semantic framework as English, and focus on Chinese verbal and nominal predicates with their arguments in Chinese TreeBank, respectively. The semantic arguments include the following.

(1) Core arguments: Arg0 to Arg5. Generally, Arg0 and Arg1 denote the agent and the patient, respectively, while arguments from Arg2 to Arg5 are predicate-specific.
(2) Adjunct arguments are universal to all predicates, for example, ArgM-LOC for locatives and ArgM-MNR for manners and ArgM-TMP for temporals.

All the arguments are annotated on parse tree nodes with their boundaries aligning with the spans of tree nodes. Figure 1 demonstrates an example in Chinese PropBank and NomBank. In this example, the verbal predicate “提供/provide” is annotated with three core arguments (i.e., “中国银行/Bank of China” as Arg0, “向外商投资银行/to Foreign Investment Bank” as Arg2, and “四十亿 人民币贷款/4 billion RMB loan” as Arg1) while the nominal predicate “贷款/loan” is annotated with two core arguments[1] (i.e., “NP(中国银行/Bank of China)” as Arg1 and “PP(向外商投资银行/to Foreign Investment Bank)” as Arg0), and an adjunct argument (i.e., “NN(人民币/RMB)” as

[1]Please note that the frame of “贷款/loan” in Chinese NomBank is different from that of “loan” in English NomBank. This is an example where semantic definitions of an English predicate and its Chinese counterpart are not always consistent. Due to the fact that all the arguments are annotated on parse tree nodes with their boundaries aligning with the spans of tree nodes, the fragment of “四十亿人民币/4 billion RMB” fails to become an argument of “贷款/loan” since the former does not form a parse tree node.

ArgM-MNR, denoting the manner of loan). It is worth pointing out that there is a (Chinese) NomBank-specific label in Figure 1, Sup (support verb) [Xue 2006a], to help mark the arguments, which occur outside the nominal predicate-headed noun phrase. In (Chinese) NomBank, a verb is considered to be a support verb only if it shares at least an argument with the nominal predicate.

3. BASELINE: CHINESE SRL FOR VERBAL AND NOMINAL PREDICATES

Given a syntactic parse tree and its predicates (verbal and nominal), popular SRL systems cast the SRL problem as a classification task, in which it annotates each constituent in the parse tree either with a numbered core argument label (Arg0, ..., Arg5), or with a adjunct argument label (like ArgM-LOC, ArgM-TMP, and so on), or with the label NULL for non-argument. Since the constituents labeled with NULL are predominant, we divide the system into two consecutive phases as to overcome the imbalance between the training data of the NULL class and that of any other argument class: argument pruning and hierarchical argument classification.

3.1 Argument Pruning

For verbal SRL, we simply adopt the heuristic rules as defined in Xue and Palmer [2005] to filter out constituents that most likely represent non-argument.

For nominal SRL, we categorize arguments into two types, arguments inside NP (called *inside arguments*) and arguments introduced via a support verb (called *outside arguments*) according to the specific argument structures of nominal predicates, and handle them separately. The motivation of grouping arguments of nominal predicates into two types lies in that inside arguments are usually dominated by nominal predicates directly while outside arguments are usually dominated by support verbs directly, rather than nominal predicates. The statistics shows that about 20% and 22% of arguments are introduced via a support verb on (English) NomBank and Chinese NomBank, respectively. Moreover, 54% of outside arguments in Chinese NomBank are core arguments.

For the inside arguments, we adopt the following three heuristic rules to find *inside argument candidates*.

— All the siblings of the predicate are candidates.
— If a CP (clause headed by complementizer) or DNP (phrase formed by “XP + DEG(的)”) node is a candidate, its children are candidates too.
— For any node X, if its parent is an ancestral node of the predicate, and the internal nodes along the path between X and the predicate are all NPs (noun phrases), then X is a candidate.

For outside arguments, we first look for the support verb of the nominal predicate and then adopt the same heuristic rules in argument pruning for verbal SRL to find the candidates for the support verb. The intuition behind it is that outside argument candidates are marked via the support verb, so that the argument candidates of the support verb can be regarded as outside argument candidates of the nominal predicate. However, as support verbs are not annotated explicitly in the testing phase, we identify intervening verbs as alternatives to support verbs in both training and testing phases with the path between the nominal predicate and intervening verb in the form of “VV<VP>[NP>]+NN”, where “[NP>]+” denotes one or more NPs. Our statistics on Chinese NomBank show that 51.96% of nominal predicates have no intervening verb while 48.04% of nominal predicates have only one intervening verb. That is, the number of nominal predicates which have two or more intervening verbs is too few to affect the figures. It may happen to get two or more

Table I. Traditional Features and Their Instantiations for Verbal SRL and Nominal SRL, with "NP(中国银行/Bank of China)" as the Focus Constituent and "提供/provide" as the Predicate, Regarding Figure 1

Feature	Description
b1	Predicate: the predicate itself. (提供/*provide*)
b2	Phrase type: the syntactic category of the constituent. (*NP*)
b3	Head word (b3H) and its POS (b3P). (银行/*bank*, *NN*)
b4	Path: the path from the constituent to the nominal predicate. ($NP<IP>VP>VP>VV$)
b5	Position: the positional relationship of the constituent with the predicate. "left"/"right". (*left*)
Combined features: b51-b52	
b1&b2; b1&b3H	

intervening verbs when conjunct structures (e.g., VP) are involved. For example, in the sentence "警察/police 策划/plan 执行/perform 了/LE 这次/this 调查/investigation", the nominal predicate "调查/investigation" has two intervening verbs "策划/plan" and "执行/perform".

Taking the nominal predicate "贷款/loan" in Figure 1 as an example, "NN(人民币/RMB)" and "QP(四十亿/4 billion)" are identified as inside argument candidates, while "PP(向外资投资银行/to Foreign Investment Bank)" and "NP(中国银行/Bank of China)" are identified as outside argument candidates via the support verb "VV(提供/provide)".

3.2 Hierarchical Argument Classification

Motivated by Moschitti et al. [2005] and Ding and Chang [2008] that the linguistic discrepancy between core arguments and adjunct arguments not only exists but also can be captured, we develop a two-level framework to label each argument candidate with a specific argument label (including the NULL class for non-argument). The first level is a triple classifier (called Triple Classifier) which differentiates non-arguments (NULL), core arguments (ArgN), and adjunct arguments (ArgM). The second level consists of two classifiers: a 5-classes[2] classifier (Core Classifier) for all the five core arguments and a 16-classes classifier (Adjunct Classifier) for all the 16 adjunct arguments.

3.3 Features

A wide range of features have been explored in previous work on Chinese SRL [Ding and Chang 2008; Li et al. 2009; Xue 2006b, 2008; Xue and Palmer 2005]. In this section, we first describe several widely used traditional features and then systematically explore a large space of additional features specially designed for verbal SRL and nominal SRL, respectively.

3.3.1 Traditional Features. Using the feature naming convention as adopted in Jiang and Ng [2006], Table I lists the traditional features [Gildea and Jurafsky 2002; Xue 2008] which are widely used in both verbal and nominal SRL.

[2]Namely Arg0-Arg4. Arg5 doesn't exist in Chinese PropBank and NomBank.

3.3.2 Additional Features. To capture more useful information in the predicate-argument structure, we also study additional features which provide extra information. Since the linguistic discrepancy of argument-predicate structure exists between verbal SRL and nominal SRL, we design different feature spaces by examining their distinction.

Additional features for verbal SRL

For simplicity, let *FC* be the focus constituent, and *P* be the verbal predicate. Table II lists the candidate features for verbal predicates.

In Table II, the candidate features can be grouped into three categories in terms of their relations with *FC* and *P*. Most of these features come from pervious SRL work [Ding and Chang 2008; Li et al. 2009; Pradhan et al. 2004; Sun and Jurafsky 2004; Xue 2008]. Specially, the predicate class (d3) feature was first introduced in Giuglea and Moschitti [2004] for English SRL[3] to overcome the imbalance of the predicate distribution in that some predicates can be only found in the training data while some predicates in the testing data are absent from the training data. In particular, the verb class is classified along three dimensions: the number of arguments, the number of framesets and selected syntactic alternations. For example, the verb class of "C1C2a" means that it has two framesets, with the first frameset having one argument and the second having two arguments. The symbol "a" in the second frameset represents a type of syntactic alternation.

Additional features for nominal SRL

Considering the discrepancy between predicate-argument structures of inside arguments and outside arguments, it is natural to design different feature sets for inside and outside arguments, respectively. For inside arguments, the context information embedded in the highest NP headed by the nominal predicate is expected to be helpful and the context information outside the highest NP is considered much less useful since they usually locate near to the nominal predicate. For example, whether the focus constituent is adjacent to the predicate normally implies whether there is a domination relationship between them. However, the situation reverses with regard to outside arguments, for which the support verbs play a crucial role in labeling their semantic roles. For example, in the sentences "中国银行/Bank of China 提供/provide 贷款/loan" and "中国银行/Bank of China 申请/apply 贷款/loan", the support verb "提供/provide" implies "中国银行/Bank of China" as a lender with semantic role "Arg1" while "申请/apply" implies "中国银行/Bank of China" as a debtor with semantic role "Arg0". The second part in Table III shows the features (ai1-ai7) in better capturing the details between inside arguments and nominal predicates. Specially, features ai6 and ai7 are sibling-related features, inspired by the features related with the neighboring arguments in Jiang and Ng [2006]. For outside arguments, we also identify intervening verbs as alternatives to support verbs since support verbs are not explicitly annotated in the test data. The third part in Table III lists the intervening verb-related features (ao1-ao4, ao51-ao54) employed in this article.

Feature selection

Features proposed above may not be effective in all tasks. We adopt the greedy feature selection algorithm as described in Jiang and Ng [2006] to sift positive features empirically and incrementally according to their contributions on the development data. The

[3]Such feature was further elaborated in Giuglea and Moschitti [2006a, 2006b] for English SRL and in Xue and Palmer [2005] for Chinese SRL.

Table II. Additional Features and Their Instantiations for Verbal SRL, with “NP(中国银行/Bank of China)” as the Focus Constituent and “提供/provide” as the Verbal Predicate, Regarding Figure 1

Type	Feature	Description
FC related	c1	voice position: indicating the positional relation between FC and voice maker, for example, BA/把, BEI/被. (*NULL*)
	c2	voice path: the path from FC to the voice maker. (*NULL*)
	c3	subcat frame1: the rule that expands the parent of FC, for example, IP→NP+NP*+VP, where the later NP is the focus constituent. (*IP→NP*+VP+PU*)
	c4	subcat frame2: the rule that expands FC. (*NP→NN+NN*)
	c5	the first word (c5F)and the last word (c5L) of FC. (中国/*China*, 银行/*bank*)
	c6	phrase type of the siblings to the left (c6L) and the right (c6R). (*NULL*, *VP*)
	c7	phrase type of FC's the parent node. (*IP*)
	c8	indicating whether FC is the head of its parent node. (*No*)
	Combined features: c51-c54	
	c5F&c5L; b2&c8; b2&b3; b2&b4	
P related	d1	subcat frame3: the rule that expands the parent of P. (*VP→VV+NP*)
	d2	subcat frame4: the subcat frame that consists of the NPs surrounding P. (*VV→VV+NP*)
	d3	predicate class: the verb class that P belongs to. (*C3b*)
	d4	POS of P. (*VV*)
	d5	core arguments defined in P's frame file. (*Arg0_Arg1_Arg2*)
	Combined features: d51-d52	
	b1&d2; d4&b2	
FC − P related	e1	compressed path: the brief version of the path feature, where consequent identified labels are replaced by one. (*NP<IP>VP>VV*)
	e2	the amount of VPs in the path feature. (*2*)
	e3	partial path1: the path from FC to their lowest common node. (*NP<IP*)
	e4	partial path2: the path from P to their lowest common node. (*VV<VP<VP<IP*)
	e5	layer of FC: the number of nodes in e3 subtracts the number of nodes in e4. (*-2*)
	e6	whether there is an IP in e4. (*Yes*)
	Combined features: e51-e63	
	b2&d3; b3H&d3; b1&b5; b4&b3H; b4&b5; b1&b4; b1&b4&b5; b4&d2; e1&d3; e1&b5; b1&e1; b1&e1&b5;e1&d2;	

algorithm repeatedly selects one feature each time which contributes most and stops when adding any of the remaining features fails to improve the performance. Taking the Core Classifier in nominal SRL as an example, the feature selection process could be done as follows: run the selection algorithm with the basic set of features (b1-b5,

Table III. Additional Features and Their Instantiations for Nominal SRL, with “NN(人民币/RMB)” as the Inside Argument Candidate, “NP(中国银行/Bank of China)” as the Outside Argument Candidate, and “贷款/loan” as the Nominal Predicate, Regarding Figure 1

Type	Feature	Description
for inside and outside	a1	predicate class: the verb class that the predicate belongs to.
	a2	the first word (a2F)and the last word (a2L) of FC
	a3	phrase type of FC’s the parent node
	Combined features: a51-a53	
	a1&b2; a1&b3H; b4&b5	
for inside only	ai1	Whether FC is adjacent to P. Yes/No. (*Yes*)
	ai2	The headword (ai2H) and pos (ai2P) of P’s nearest right sibling. (*NULL*, *NULL*)
	ai3	Whether P has right siblings. Yes/No. (*No*)
	ai4	Compressed path of b4: compressing sequences of identical labels into one. (*NN<NP>NN*)
	ai5	Whether P has siblings. Yes/No. (*Yes*)
	ai6	For each sibling of FC, combine b2&b3H&b4&b5. (*NULL*)
	ai7	Coarse version of ai6, b2&b5. (*NULL*)
for outside only	ao1	Intervening verb itself. (提供/*provide*)
	ao2	The verb class that the intervening verb belongs to. (*C3b*)
	ao3	The path from the focus constituent to the intervening verb. (*NP<IP>VP>VP>VV*)
	ao4	The compressed path of ao3: compressing sequences of identical labels into one. (*NP<IP>VP>VV*)
	Combined features: ao51-ao54	
	ao1&ao3; ao1&ao4; ao2&ao3; ao2&ao4.	

b51-b52) to pick up effective features from feature space of (a1-a3, a51-a53, ai1-ai7, ao1-ao4, ao51-ao54).

4. IMPROVING NOMINAL SRL WITH VERBAL EVIDENCE

Xue [2008] reported a performance gap of 22.4 in F1-measure (92.0 vs. 69.6) between verbal SRL and nominal SRL on gold parse trees and gold predicates. The lower performance of nominal SRL is partly due to the much smaller amount of the annotation data (about 1/4) in Chinese NomBank than that in Chinese PropBank. Besides, it is due to the inherent difficulty in nominal SRL. According to the annotation criteria of Chinese NomBank [Xue 2006a], even when a noun is a true deverbal noun, not all of its modifiers are legitimate core or adjunct arguments of this predicate. Some modifiers can only co-occur with the nominalized form and cannot co-occur with its corresponding verbal form. Chinese NomBank is only interested in core and adjunct arguments that can co-occur with both the nominal and verbal forms of the predicate. This means that the judgment of arguments is semantic rather than syntactic. Since Chinese PropBank and NomBank are annotated on the same data set with the same lexical guidelines (e.g., frame files), it may be interesting to investigate the contribution of Chinese verbal SRL on the performance of Chinese nominal SRL. Assuming verbal SRL instances are available, this section explores two ways to improve the performance of nominal SRL with verbal evidence.

4.1 Instance Merging: Merging Training Instances from Verbal SRL

A verbal predicate and its nominalized form may share the same frame file where argument labels are defined with regard to their semantic roles of the predicate. For example, in the frame file of predicate "贷款/loan", the debtor is always labeled with Arg0 and the lender labeled with Arg1. This can be demonstrated by the following two sentences: "贷款/loan" is annotated as a nominal and a verbal predicate in S1 and S2, respectively.

— S1 [Arg1 中国银行/Bank of China] [Arg0 向外商投资银行/to Foreign Investment Bank] 提供/provide [Rel 贷款/loan]
— S2 [Arg0 外商投资银行/Foreign Investment Bank] [Arg1 向中国银行/from Bank of China] [Rel 贷款/loan]

Moreover, we learn that verbal and nominal SRL systems share several common features (e.g., features in Table I) which play a dominating role in predicting the semantic role. Finally, one major reason for the low performance of nominal SRL lies in the imbalanced distribution of nominal predicates. Such imbalance could be much alleviated if training instances for verbal predicates are considered in some way. For example, 6.5% of nominal predicates in the test data are absent from the nominal training data while nearly half of them are present in the verbal training data. Therefore, it is straightforward to augment nominal training instances with verbal ones. In order to get rid of noises from non-arguments, we only fetch verbal SRL instances in hierarchical classification stage.

As observed by Jiang and Zhai [2007], some verbal SRL instances may be noisy or misleading and should be excluded from merging into nominal instances. Let Y be the set of class labels, and let XV and XN be the final feature space we choose to represent the observed verbal and nominal SRL instances, respectively. Given a set of labeled verbal and nominal instances, namely $\{(xv_i, y_i)|1 \leq i \leq n, xv_i \in XV, y_i \in Y\}$ and $\{(xn_j, y_j)|1 \leq j \leq m, xn_j \in XN, y_j \in Y\}$, we propose a simple way to determine "misleading" verbal SRL instances. First, we train and get a classifier model M_n with $\{(xn_j, y_j)\}$ (the labeled nominal instances). Then we predict label y_i' for each verbal instance with model M_n. Finally, we filter out misleading instances $\{(xv_i, y_i)|1 \leq i \leq n, y_i' \neq y_i\}$ and keep the remaining ones.

4.2 Feature Integrating: Integrating Features Derived from SRL for Verbal Predicates

Although we have proposed several support verb-related features (ao1-ao4, ao51-ao54 in Table III), one may still ask how large the role is that support verbs can contribute to nominal SRL. It is interesting to note that outside arguments and the highest NP phrase headed by the nominal predicate are annotated as arguments of the support verb in Chinese PropBank. For example, Chinese PropBank marks "中国银行/Bank of China" as Arg0 and "四十亿人民币贷款/4 billion RMB loan" as Arg1 for verb "提供/provide" in Figure 1. Let OA be the outside argument, VV be the support verb, and NP be the highest NP phrase headed by the nominal predicate NN, then there exists a pattern "OA VV NN" in the sentence, where the support verb VV plays a certain role in transferring roles between OA and NN. For example, if OA is the agent of VV, then OA is also the agent of phrase VP(VV NN). Like the example in Figure 1, suppose a NP is the agent of support verb "提供/provide" as well as VP phrase ("提供四十亿人民币贷款/provide 4 billion RMB loan"), we can infer that the NP is the

Table IV. Nominal SRL Features Derived from Verbal SRL. The Feature Instantiations Are Based on "NP(中国银行/Back of China)" as C, "VV(提供/provide)" as V, and "NP(四十亿人民币贷款/4 billion RMB loan)" as NP, with Respect to Figure 1

Feature	Description
ao5	Whether C is an argument for V. Yes/No. (*Yes*)
ao6	The semantic role of C for V. (*Arg0*)
ao7	Whether NP is an argument for V. Yes/No. (*Yes*)
ao8	The semantic role of NP for V. (*Arg1*)
Combined features: ao55-ao61	
ao1&ao5; ao1&ao6; ao1&ao5&b1; ao1&ao6&b1; ao1&ao7; ao1&ao8; ao5&ao7.	

lender of the nominal predicate "贷款/loan" independently on any other information, such as the NP content and the path from the NP to the nominal predicate "贷款/loan".

The above analysis implies the usefulness of semantic role information derived from verbal SRL. Let C be the focus constituent, V be the intervening verb, and NP be the highest NP headed by the nominal predicate. Table IV shows the features (ao5-ao8, ao55-ao61) derived from verbal SRL. Here we adopt the fore-mentioned verbal SRL system to achieve the goal. Nominal SRL should be able to benefit from verbal evidence due to the fact that verbal SRL substantially outperforms nominal SRL, although it may introduce some noise.

5. AUTOMATIC PREDICATE RECOGNITION

Automatic predicate recognition is a prerequisite for the application of SRL systems. For verbal predicates, it is very easy to resolve using some heuristic rules since nearly 99% of verbs are annotated as predicates in the Chinese PropBank.

Unlike verbal predicate recognition, nominal predicate recognition is quite complicated since only 17.5% of nouns are annotated as predicates in Chinese NomBank. It is quite general that a noun is annotated as a predicate in some cases but not in others. Therefore, automatic predicate recognition is vital to nominal SRL. In principle, automatic predicate recognition can be cast as a binary classification (e.g., *Predicate* vs. *Non-Predicate*) problem. For nominal predicates, a binary classifier is trained to predicate whether a noun is a nominal predicate or not. This article employs the convolution tree kernel, as proposed in Collins and Duffy [2001], on automatic recognition of nominal predicates.[4]

Given the convolution tree kernel, the key problem is how to extract a parse tree structure from the parse tree for a nominal predicate candidate. Actually, tree kernel methods have been extensively studied in SRL with different strategies to build structural features [Giuglea and Moschitti 2006b; Moschitti 2004; Moschitti et al. 2006a, 2006b, 2008; Zhou et al. 2011]. In this article, the parse tree structure is constructed as follows: 1) starting from the predicate candidate's POS node, collect all of its sibling nodes (with their headwords); 2) recursively move one level up and collect all of its sibling nodes (with their headwords) till reaching a non-NP node. Specially, in order to explicitly mark the positional relation between a node and the predicate candidate, all nodes on the left side of the candidate are augmented with tags 1 and 2 for nodes

[4]We also tried a feature-based method (with extensive feature engineering) for nominal predicate recognition. For this relatively simple task, the feature-based method actually achieves comparable performance with the tree kernel-based method (e.g., 90.05 vs. 89.79 on gold parse trees). We only report the tree kernel-based method in this article due to its potential in effectively capturing structure information, which we find critical for the predicate recognition task, and that predicate recognition is not our focus of this article.

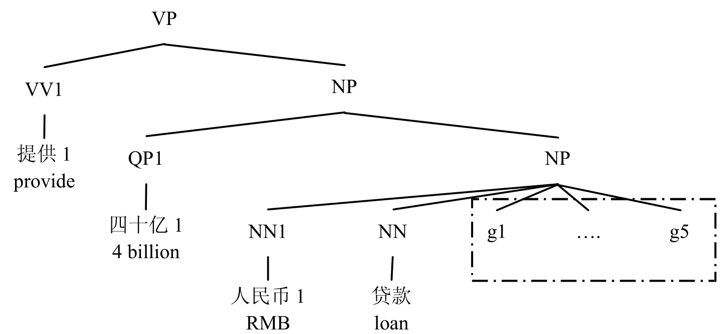

Fig. 2. An example: attaching global features to the parse tree.

on the right side. Figure 2 shows an example of the parse tree structure with regard to the predicate candidate "贷款/loan" as shown in Figure 1.

In addition, we also explore the usefulness of following five global statistic features for kernel-based predicate recognition, given the predicate candidate w_0, its left neighbor word w_{-1} and its right neighbor word w_1.

— g1 Whether w_0 is ever tagged as a verb in the training data? Yes or No.
— g2 Whether w_0 is ever annotated as a nominal predicate in the training data? Yes or No.
— g3 The most likely label for w_0 when it occurs together with w_{-1} and w_1.
— g4 The most likely label for w_0 when it occurs together with w_{-1}.
— g5 The most likely label for w_0 when it occurs together with w_1.

This is done by attaching the five global features as the right siblings of the predicate candidate, as shown in Figure 2. We have explored other ways to include those global features. However, the way as shown in Figure 2 works best.

6. EXPERIMENTATION

We have evaluated our unified approach for Chinese verbal and nominal SRL on Chinese PropBank and Chinese NomBank with Chinese CTB5.1 as its counterpart.

6.1 Experimental Settings

This version of Chinese PropBank and NomBank consists of standoff annotations on the files (chtb001 to 1151.fid) of Chinese Penn TreeBank 5.1. Following the experimental setting in Xue [2008], 648 files (chtb081 to 899.fid) are selected as the training data, 72 files (chtb001 to 040.fid and chtb900 to 931.fid) are held out as the test data, and 40 files (chtb041 to 080.fid) as the development data, with 31,361 (8,642), 3,599 (1,124), and 2,060 (731) verbal (nominal) propositions, respectively. To see whether an improvement in F1-measure is significant, we conducted significance testing using a "stratified shuffling" technique which is actually a "compute-intensive randomized test" [Cohen 1995]. In this article, "$>>>$", "$>>$" and "$>$" denote p-values of an improvement smaller than 0.01, in-between (0.01, 0.05], and bigger than 0.05, which mean significantly better, moderately better, and slightly better, respectively.

As Chinese sentences are not naturally segmented into words, two Chinese automatic parsers are constructed: a word-based parser (using gold word boundaries) and a character-based parser (using automatically recognized word boundaries). Here, Berkeley parser [Petrov and Klein 2007][5] is chosen as the Chinese automatic parser.

[5]Berkeley Parser. See http://code.google.com/p/berkeleyparser/.

With regard to character-based parsing, we employ a Chinese word segmenter, similar to Ng and Low [2004], to obtain the top-best automatic segmentation result for a given sentence, which is then fed into Berkeley parser for further syntactic parsing. Both the word segmenter and Berkeley parser are developed on CTB5.1 with the same training and development data split as our SRL experiments. The word segmenter achieves the performance of 96.1 in F1-measure while the Berkeley parser gives a performance of 82.5 and 85.5 in F1-measure on gold and automatic word segmentation, respectively.[6]

In addition, SVMLight with the tree kernel function [Moschitti 2004][7] is selected as our classifier. In order to handle multi-classification problem in argument classification, we apply the *one* vs. *all* strategy, which builds K classifiers so as to separate one class from all others. For hierarchical classification, we adopt the linear kernel and the training parameter C is fine-tuned to 0.220 on the development data. For automatic recognition of nominal predicates, the training parameter C and the decay factor λ in the convolution tree kernel are fine-tuned to 2.0 and 0.2 on the development data, respectively.

6.2 Results with Gold Parse Trees and Gold Predicates

Effect of additional features

After performing the greedy feature selection algorithm on the development data, Table V lists the selected features. In this table, "*Y*" in the cell indicates this feature has been selected. It is interesting to find out that with the exception of feature ao51, no any other feature is shared by the Core and the Adjunct classifiers. This is not surprising since core arguments and adjunct arguments behave differently. Compared to core arguments, adjunct arguments are more self-explaining, and insensitive to the syntactic structures. Table VI presents the SRL performance on the test data with or without the selected features. It shows that the selected additional features significantly improve the performance of verbal SRL and nominal SRL by 4.29 ($>>>$) and 4.47 ($>>>$), respectively. Specially, the performance of verbal SRL on gold parse trees and gold predicates reaches 92.88 in F1-measure and outperforms the state-of-the-art performance of 92.0 in F1-measure, as reported in Xue [2008]. This suggests the effectiveness of our proposed features and the approach of hierarchical argument classification.

Effect of verbal evidence for nominal SRL

Table VII presents the effect of verbal instances for nominal SRL on the development data. It shows that adding verbal instances enhances the accuracy for both core arguments and adjunct arguments. This suggests the usefulness of verbal SRL instances for nominal SRL. However, to our disappointment, our proposed instance pruning method is empirically not effective. We will leave it in future work.

Table VIII shows the feature selection result for features derived from verbal SRL. It shows that seven additional features are selected for the Triple-classifier while only one additional feature is selected for the Core-classifier and Adjunct-classifier, respectively. This further indicates the linguistic discrepancy of the three tasks (triple-class classification, core argument classification and adjunct classification).

Table IX lists the performance of nominal SRL with our instance merging and feature integrating methods proposed in Section 4. In this table, "baseline" indicates

[6]Parts of Speech are not counted in evaluating the performance of word-based syntactic parser, but they are counted in evaluating the performance of character-based parser. Therefore the F1-measure for the later is higher than that for the former.
[7]SVM-LIGHT-TK. See http://dit.unitn.it/moschitt/.

Table V. The Selected Features for Each Classifier. The First Number Inside Parentheses Indicates the Chronological Order When the Corresponding Feature Is Selected While the Second Number Indicates the Improvement Achieved by the Corresponding Feature

	Feature	Triple	Core	Adjunct
verbal SRL	c1	Y (3, 0.36)		
	c2		Y (2, 0.74)	
	c3	Y (1, 0.72)		
	c4		Y (6, 0.19)	
	c5F			Y (2, 0.30)
	c5L	Y (13, 0.09)		Y (1, 0.31)
	c6R	Y (8, 0.04)		Y (3, 0.18)
	c8	Y (9, 0.06)		
	d1	Y (4, 0.27)		
	d5		Y (7, 0.12)	
	d51	Y (6, 0.18)		
	e1	Y (14, 0.02)		
	e4	Y (11, 0.02)		
	e53		Y (4, 0.40)	
	e54	Y (12, 0.04)	Y (3, 0.71)	
	e56	Y (5, 0.28)	Y (5, 0.31)	
	e58	Y (10, 0.03)		
	e59		Y (1, 1.55)	
	e60	Y (2, 0.58)		
	e61	Y (7, 0.07)		
nominal SRL	a1	Y (6, 0.16)	Y (5, 0.12)	
	a2F	Y (4, 0.46)		Y (2, 0.79)
	a3	Y (9, 0.07)		
	a51			Y (3, 0.79)
	ai1	Y (3, 0.89)		Y (4, 0.53)
	ai2H	Y (7, 0.07)		Y (1, 1.06)
	ai2P	Y (4, 0.33)	Y (4, 0.25)	
	ai4		Y (6, 0.12)	
	ai6	Y (1, 2.70)		
	ai7		Y (3, 0.37)	
	ao1	Y (2, 1.94)	Y (1, 1.86)	
	ao51		Y (2, 0.99)	Y (5, 0.53)
	ao52	Y (8, 0.03)		
	ao53	Y (5, 0.36)		

Table VI. The SRL Performance on the Test Data with Gold Parse Trees and Gold Predicates

	Feature	Rec. (%)	Pre. (%)	F1
verbal SRL	traditional	88.62	88.56	88.59
	+additional	92.91	92.84	92.88
nominal SRL	traditional	62.03	69.22	65.43
	+additional	66.00	74.30	69.90

Table VII. Performance of Nominal SRL by Merging Verbal SRL Instances

Task	Training instances	Acc. (%)
core arguments	nominal instances	88.85
	+verbal instances	90.00
	+pruned verbal instances	88.38
adjunct arguments	nominal instances	83.97
	+verbal instances	85.12
	+pruned verbal instances	85.29

Table VIII. Feature Selection of Verbal SRL Related Features. The First Number Inside Parentheses Indicates the Chronological Order When the Corresponding Feature Is Selected While the Second Number Indicates the Improvement Achieved by the Corresponding Feature

Feature	Triple	Core	Adjunct
ao5	Y (3, 0.15)		
ao6	Y (6, 0.13)		Y (1, 0.99)
ao55		Y (1, 0.62)	
ao56	Y (4, 0.04)		
ao57	Y (2, 0.15)		
ao58	Y (7, 0.06)		
ao59	Y (5, 0.09)		
ao61	Y (1, 0.92)		

the nominal SRL performance achieved after adding selected features in Section 3. It shows that instance merging and feature integrating improve the performance of nominal SRL by 0.57 ($>>$) and 2.35 ($>>>$), respectively. It also shows that our system outperforms the state-of-the-art [Xue 2008] by about 2.8 in F1-measure.

Table X presents the performance trend over different occurring frequencies of the predicates in the training data. It shows that the predicates with higher occurring frequencies normally achieve better performance than those with lower occurring frequencies. However, as shown in the first row of Table X, verbal predicates unseen in the training data achieves an expected high performance of 90.43 in F1-measure. This is probably due to adjective verbs (i.e., POS tagged as VA) whose arguments are easy to recognize due to their simple syntactic structures. On the test data, the adjective verbs occupy 15.50% of all unseen predicates while they only occupy 8.17% of all seen predicates.

Table IX. The Performance of SRL for Nominal Predicates on the Test Data with Gold Parse Trees and Gold Predicates

System	Rec. (%)	Pre. (%)	F1
baseline	66.00	74.30	69.90
+instance merging	66.54	74.90	70.47
+feature integrating	68.01	77.07	72.25
+both	68.15	77.23	72.41
Xue [2008]	66.1	73.4	69.6

Table X. The Performance of Predicates over Different Occurring Frequencies in the Training Data

	Predicate frequency	Rec. (%)	Pre. (%)	F1
verbal SRL	0	90.53	93.32	90.43
	1–3	89.96	89.55	89.75
	4–5	89.93	89.73	89.83
	>5	93.53	93.52	93.52
nominal SRL	0	60.16	75.51	66.97
	1–3	67.57	77.42	71.01
	4–5	63.38	84.91	72.58
	>5	69.05	76.86	72.74

Table XI. The Performance of SRL on the Test Data with Automatic Parse Trees and Gold Predicates. Note: the Numbers Outside the Parentheses Indicate the Performance Using a Word-Based Parser, While the Numbers Inside Indicate the Performance Using a Character-Based Parser

		Rec. (%)	Pre. (%)	F1
verbal SRL	this paper	73.53 (69.31)	79.14 (78.62)	76.23 (73.67)
	Xue [2008]	62.5 (60.3)	76.8 (74.8)	68.9 (66.8)
nominal SRL	this paper	55.61 (53.31)	66.14 (66.30)	60.42 (59.10)
	Xue [2008]	53.1 (52.9)	62.9 (62.3)	57.6 (57.3)

6.3 Results with Automatic Parse Trees and Gold Predicates

In previous section we assumed the availability of gold parse trees during the testing process. Here we conduct experiments on automatic parse trees, using the Berkeley parser. Table XI presents the SRL performance on the test data by using automatic parse trees. It shows the following.

(1) The performance of verbal (nominal) SRL drops from 92.88 (72.41) to 76.23 (60.42) in F1-measure when replacing gold parse trees with word-based automatic ones. This is mainly due to following two kinds of errors: 1) syntactic parsing errors: 7.45% verbal arguments and 6.91% nominal arguments fail to align with any syntactic constituent and thus are simply discarded; and 2) POS tagging errors: 7.78% verbal predicates and 6.05% nominal predicates are tagged with wrong POS tags and thus it is difficult to recognize their arguments.
(2) The performance of verbal (nominal) SRL drops from 76.23 (60.42) to 73.67 (59.10) in F1-measure when replacing word-based automatic parse trees with character-based ones. This is mainly due to word segmentation errors in that 5.11% verbal

Table XII. The Performance of Automatic Predicate Recognition on the Test Data. Note: the Numbers Outside the Parentheses Indicate the Performance Using a Word-Based Parser, While the Numbers Inside Indicate the Performance Using a Character-Based Parser

Type	Parses	Rec. (%)	Pre. (%)	F1
verbal predicates	gold	100.00	99.26	99.63
	auto	92.22 (87.77)	93.86 (88.22)	93.03 (87.99)
nominal predicates	gold	91.19	88.44	89.79
	auto	85.32 (83.99)	79.52 (79.33)	82.32 (81.59)
all predicates	gold	97.90	96.64	97.27
	auto	94.33 (90.81)	93.95 (89.90)	94.14 (90.35)

and 1.60% nominal predicates on the test data are wrongly segmented in character-based parse trees, respectively.

(3) It also shows that our system substantially outperforms Xue [2008] by 7.33 and 2.82 in F1-measure on verbal and nominal SRL, respectively.

6.4 Results with Automatic Predicates

So far verbal and nominal predicates are assumed to be manually annotated and available. Here we turn to a more realistic scenario where both the parse tree and predicates are automatically obtained. In the following, we first report the results of automatic predicate recognition and then the results of SRL on automatic recognition of predicates.

Table XII lists the predicate recognition results, using the simple rule as described in Section 5 for verbal predicates, and the kernel-based method as described in Section 5 for nominal predicates. For nominal predicates, we have also defined a simple rule that recognizes a noun which is ever a verb or a nominal predicate in the training data as a nominal predicate. Based on gold parse trees, the rule achieves a performance of 81.40 in F1-measure. This suggests that our kernel-based method significantly outperforms the simple rule-based one. It is not surprising that predicate recognition performs much worse on character-based automatic parse trees than on word-based automatic parse trees. Table XII also shows the performance of overall predicate recognition by combining verbal and nominal predicates. It shows that when automatic parse trees are used, the recognition performance of overall predicates is higher than both that of verbal predicates and that of nominal predicates. Taking word-based parsing as an example, 3.17% verbal predicates are wrongly recognized as nominal predicates and 5.60% nominal predicates are recognized as verbal predicates, due to the POS tagging errors.

In order to have a clear performance comparison among Chinese SRL on gold/automatic parse trees and gold/automatic predicates, Table XIII lists all the results in those scenarios as well as the overall SRL result by combining verbal SRL and nominal SRL. It shows the following.

(1) The performance of verbal SRL suffers a drop of 22.85 (from 92.88 to 70.03) in F1-measure when using a character-based parser and automatic predicates while the performance of nominal SRL suffers a drop of 18.51 (from 72.41 to 53.90). This suggests the great challenge in Chinese SRL as well as the disappointment of Chinese syntactic parsing.

(2) When using gold parse trees, automatic recognition of verbal predicates has little influence on verbal SRL performance (i.e., 92.88 vs. 92.54 in F1-measure), largely

Table XIII. The Performance of SRL on the Test Data with the Choices of Gold/Automatic Parse Trees and Gold/Automatic Predicates. Note: the Numbers Outside the Parentheses Indicate the Performance Using a Word-Based Parser, While the Numbers Inside Indicate the Performance Using a Character-Based Parser

Type	Parses	Predicate	Rec. (%)	Pre. (%)	F1
verbal SRL	gold	gold	92.91	92.84	92.88
		auto	92.91	92.18	92.54
	auto	gold	73.53 (69.31)	79.14 (78.62)	76.23 (73.67)
		auto	73.53 (69.31)	75.35 (70.76)	74.43 (70.03)
nominal SRL	gold	gold	68.15	77.23	72.41
		auto	64.72	73.96	69.04
	auto	gold	55.61 (53.31)	66.14 (66.30)	60.42 (59.10)
		auto	52.87 (50.91)	57.55 (57.28)	55.11 (53.90)
all	gold	gold	87.65	90.01	88.81
		auto	87.03	89.01	88.01
	auto	gold	70.43 (66.60)	74.94 (74.38)	72.61 (70.28)
		auto	69.78 (66.07)	72.69 (68.97)	71.20 (67.49)

due to the high performance (i.e., 99.63 in F1-measure) of verbal predicate recognition. For nominal SRL on gold parse trees, the drop of 3.37 in F1-measure (i.e., 72.41 vs. 69.04) is not salient too, considering that the performance of automatic recognition of nominal predicates is 89.79 in F1-measure. This is largely due to the fact that the wrongly recognized nominal predicates are usually less frequently occurring ones with lower performance, as shown in Table X.

(3) The performance of overall SRL by combining verbal and nominal SRL compromises the errors in either verbal SRL or nominal SRL. In occasions where predicates are wrongly grouped (e.g., a verbal predicate is incorrectly POS tagged as a noun and in turn is recognized as a nominal predicate), it is possible that their arguments be still recognized correctly (e.g., by nominal SRL). Taking word-based parsing as an example, further examination reveals that although 3.75% of predicates are wrongly grouped, 18.05% of their arguments are still correctly recognized.

6.5 Comparison

Chinese nominal SRL vs. Chinese verbal SRL

Our experiments show that the performance of Chinese nominal SRL is about 20 lower (e.g., 72.41 vs. 92.88 in F1-measure) than that of Chinese verbal SRL, partly due to the smaller amount of annotated data (about 1/4) in Chinese NomBank than that in Chinese PropBank. Moreover, according to Chinese NomBank annotation criteria [Xue 2006a], even when a noun is a true deverbal noun, not all of its modifiers are legitimate core or adjunct arguments of this predicate. Some modifiers can only co-occur with the nominalized form and cannot co-occur with its corresponding verbal form. Chinese NomBank is only interested in core and adjunct arguments that can co-occur with both the nominal and verbal forms of the predicate. This means that the judgment of arguments is semantic rather than syntactic. These facts may also partly explain the lower nominal SRL performance, especially the performance of argument identification. This can be illustrated by the statistics on the development data that 96%

(40%) of verbal (nominal) predicates' siblings are annotated as arguments. Finally, the predicate-argument structure of nominal predicates is more flexible and complicated than that of verbal predicates as illustrated in Xue [2006a].

Chinese verbal SRL vs. English verbal SRL

Given gold parse trees and gold predicates, the performance in F1-measure (e.g., 92.88) of Chinese verbal SRL is fairly high considering that the state-of-the-art of English verbal SRL trained on PropBank is about 92.0 in F1-measure [Toutanova et al. 2005]. This is partly due to: 1) Chinese verbs appear to be less polysemous; 2) Adjectives in Chinese are traditionally counted as verbs in CTB, and they generally have one argument with a much simpler syntactic structure; and 3) Labels of A2-A5 usually are hard to be predicted and they are less frequent in Chinese than in English (about 7% in Chinese vs. 14% in English). However, this apparent advantage is diminishing when automatic parse trees are adopted. The performance gap between Chinese and English SRL is largely due to the lower performance of Chinese syntactic parsing.

Chinese nominal SRL vs. English nominal SRL

Liu and Ng [2007] reported the performance of 77.04 and 72.83 in F1-measure on English NomBank when gold and automatic parse trees are used, respectively. Taking into account that Chinese verbal SRL achieves comparable performance with English verbal SRL on gold parse trees, the performance gap between Chinese and English nominal SRL (e.g., 72.41 vs. 77.04 in F1-measure) presents a great challenge for Chinese nominal SRL. Moreover, while automatic parse trees only decrease the performance of English nominal SRL by about 4.2 in F1-measure, automatic parse trees significantly decrease the performance of Chinese nominal SRL by more than 12 in F1-measure due to the much lower performance of Chinese syntactic parsing. Of nominal predicate recognition considered, Gerber et al. [2009] did a similar study on English NomBank by focusing on finding those nominal predicates that surface without overt arguments. Moreover, they extended their study to nominal SRL by annotating implicit arguments which are either inter-sentential or intra-sentential [Gerber and Chai 2010].

7. CONCLUSION AND FUTURE WORK

In this article we investigate SRL in Chinese language. First, various kinds of features are systematically examined for verbal SRL and nominal SRL, respectively. Then, we further improve the performance of nominal SRL with various kinds of verbal evidence. Finally, we address the issue of automatic recognition of predicates, which is essential in SRL systems. Experiments are carefully designed over gold/automatic parse trees and gold/automatic predicates for both verbal and nominal SRL. To the best of our knowledge, this is the first research on unified semantic role labeling for verbal and nominal predicates on Chinese PropBank and NomBank.

For verbal SRL, the biggest challenge lies in its high dependence on the quality of syntactic parsing. While it may be difficult to further improve syntactic parsing, SRL on *N*-best parse trees, as a natural extension of SRL on the top-best parse tree, would alleviate the severe dependence on the quality of the top-best parse tree to some extent. Moreover, given the close interaction between syntactic parsing and SRL, joint learning on the two tasks will not only allow the uncertainty about syntactic parsing to be carried forward to SRL but will also allow useful information from SRL to be carried backward to syntactic parsing.

The above analysis is also valid for nominal SRL. Besides, nominal SRL also suffers from its inherent difficulties, including automatic nominal predicate recognition. For example, the domination relationship among consecutive nouns is always complicated and thus makes it hard to identify whether a noun phrase is dominated by the given nominal predicates. Moreover, even for the same predicate, the same syntactic structure could result in different semantic roles. For example, although the fragments of "展开/perform 调查/investigation 的/DE 工作人员/employees" and "展开/perform 调查/investigation 的/DE 方法/approach" share the same syntactic structure, "工作人员/employees" is annotated as semantic role Arg0 while "方法/approach" is annotated as non-argument of the predicate "调查/investigation".

In future work, we will explore the above issues systematically.

ACKNOWLEDGMENTS

We would like to thank the reviewers for very constructive and detailed comments.

REFERENCES

BAKER, C. F., FILLMORE, C. J., AND LOWE, J. B. 1998. The Berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL'98)*. 86–90.

CARRERAS, X. AND MÀRQUEZ, L. 2004. Introduction to the CoNLL-2004 shared task: Semantic role labeling. In *Proceedings of the 8th Conference on Computational Natural Language Learning (CoNLL'04)*. 89–97.

CARRERAS, X. AND MÀRQUEZ, L. 2005. Introduction to the CoNLL-2005 shared task: Semantic role labeling. In *Proceedings of the 9th Conference on Computational Natural Language Learning (CoNLL'05)*. 152–164.

COHEN, P. R. 1995. In *Empirical Methods for Artificial Intelligence*.

COLLINS, M. AND DUFFY, N. 2001. Convolution kernels for natural language. In *Proceedings of the 14th Advances in Neural Information Processing Systems (NIPS'01)*. 625–632.

DING, W. AND CHANG, B. 2008. Improving Chinese semantic role classification with hierarchical feature selection strategy. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'08)*. 324–333.

GERBER, M. AND CHAI, J. 2010. Beyond nombank: A study of implicit arguments for nominal predicates. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL'10)*. 1583–1592.

GERBER, M., CHAI, J., AND MEYERS, A. 2009. The role of implicit argumentation in nominal srl. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL'09)*. 146–154.

GILDEA, D. AND JURAFSKY, D. 2002. Automatic labeling of semantic roles. *Comput. Linguist. 28*, 3, 245–288.

GIUGLEA, A.-M. AND MOSCHITTI, A. 2004. Knowledge discovery using framenet, verbnet and propbank. In *Proceedings of the Workshop on Knowledge Discovering and Ontology at the 15th European Conference on Machine Learning (ECML'04).*

GIUGLEA, A.-M. AND MOSCHITTI, A. 2006a. Semantic role labeling via framenet, verbnet and propbank. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (ACL-COLING'06)*. 929–936.

GIUGLEA, A.-M. AND MOSCHITTI, A. 2006b. Shallow semantic parsing based on framenet, verbnet and propbank. In *Proceedings of the 17th European Conference on Artificial Intelligence (ECAI'06)*. 563–567.

JIANG, J. AND ZHAI, C. 2007. Instance weighting for domain adaptation in nlp. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL'07)*. 264–271.

JIANG, Z. P. AND NG, H. T. 2006. Semantic role labeling of nombank: A maximum entropy approach. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'06)*. 138–145.

LI, J., ZHOU, G., ZHAO, H., ZHU, Q., AND QIAN, P. 2009. Improving nominal SRL in Chinese language with verbal SRL information and automatic predicate recognition. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'09)*. 1280–1288.

LIU, C. AND NG, H. T. 2007. Learning predictive structures for semantic role labeling of nombank. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL'07)*. 208–215.

MEYERS, A., REEVES, R., MACLEOD, C., SZEKELY, R., ZIELINSKA, V., YOUNG, B., AND GRISHMAN, R. 2004. Annotating noun argument structure for nombank. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04)*.

MOSCHITTI, A. 2004. A study on convolution kernels for shallow semantic parsing. In *Proceedings of the 42nd Annual Meeting of the Association of Computational Linguistics (ACL'04)*. 335–342.

MOSCHITTI, A. AND QUARTERONI, S. 2010. Linguistic kernels for answer re-ranking in question answering systems. *Inf. Proc. Man.* http://www.sciencedirect.com/science/article/B6VC8-50K53XN-1/2/f6dee662afba19e7f1055f74f87f96e4.

MOSCHITTI, A., GIUGLEA, A.-M., COPPOLA, B., AND BASILI, R. 2005. Hierarchical semantic role labeling. In *Proceedings of the 9th Conference on Computational Natural Language Learning (CoNLL'05)*. 201–204.

MOSCHITTI, A., PIGHIN, D., AND BASILI, R. 2006a. Semantic role labeling via tree kernel joint inference. In *Proceedings of the 10th Conference on Computational Natural Language Learning (CoNLL'06)*. 61–68.

MOSCHITTI, A., PIGHIN, D., AND BASILI, R. 2006b. Tree kernel engineering in semantic role labeling systems. In *Proceedings of the European Chapter of the Association for Computational Linguistics Workshop on Learning Structured Information in Natural Language Applications (EACL'06)*. 49–56.

MOSCHITTI, A., QUARTERONI, S., BASILI, R., AND MANANDHAR, S. 2007. Exploiting syntactic and shallow semantic kernels for question/answer classification. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL'07)*. 776–783.

MOSCHITTI, A., PIGHIN, D., AND BASILI, R. 2008. Tree kernels for semantic role labeling. *Comput. Linguist. 34*, 2, 193–224.

NG, H. T. AND LOW, J. K. 2004. Chinese part-of-speech tagging: One-at-a-time or all-at-once? word-based or character-based? In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'04)*. 277–284.

PADÓ, S., PENNACCHIOTTI, M., AND SPORLEDER, C. 2008. Semantic role assignment for event nominalizations by leveraging verbal data. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING'08)*. 665–672.

PALMER, M., GILDEA, D., AND KINGSBURY, P. 2005. The proposition bank: An annotated corpus of semantic roles. *Comput. Linguist. 31*, 1, 71–105.

PETROV, S. AND KLEIN, D. 2007. Improved inference for unlexicalized parsing. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL'07)*. 404–411.

PONZETTO, S. P. AND STRUBE, M. 2006. Semantic role labeling for coreference resolution. In *Proceedings of the Demonstrations Session (EACL'06)*. 143–146.

PRADHAN, S., SUN, H., WARD, W., MARTIN, J. H., AND JURAFSKY, D. 2004. Parsing arguments of nominalizations in English and Chinese. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL'04)*. 141–144.

SUN, H. AND JURAFSKY, D. 2004. Shallow semantic parsing of Chinese. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL'04)*. 249–256.

SURDEANU, M., CIARAMITA, M., AND ZARAGOZA, H. 2008. Learning to rank answers on large online QA collections. In *Proceedings of the 46th Annual Meeting of the Association of Computational Linguistics (ACL'08)*. 719–727.

SURDEANU, M., HARABAGIU, S., WILLIAMS, J., AND AARSETH, P. 2003. Using predicate-argument structures for information extraction. In *Proceedings of the 41st Annual Meeting of the Association of Computational Linguistics (ACL'03)*. 8–15.

TOUTANOVA, K., HAGHIGHI, A., AND MANNING, C. D. 2005. Joint learning improves semantic role labeling. In *Proceedings of the 43rd Annual Meeting of the Association of Computational Linguistics (ACL'05)*. 8–15.

XUE, N. 2006a. Annotating the predicate-argument structure of Chinese nominalizations. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06)*. 1382–1387.

XUE, N. 2006b. Semantic role labeling of nominalized predicates in Chinese. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL'06)*. 431–438.

XUE, N. 2008. Labeling Chinese predicates with semantic roles. *Comput. Linguist. 34*, 2, 225–255.

XUE, N. AND PALMER, M. 2003. Annotating the propositions in the Penn Chinese treebank. In *Proceedings of 2nd SIGHAN Workshop on Chinese Language Processing (SIGHAN'03)*. 47–54.

XUE, N. AND PALMER, M. 2005. Automatic semantic role labeling for Chinese verbs. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI'05)*.

ZHOU, G., LI, J., FAN, J., AND ZHU, Q. 2011. Tree kernel-based semantic role labeling with enriched parse tree structure. *Inf. Proc. Man. 47*, 3, 349–362.

Received November 2010; revised February 2011; accepted April 2011