

A Tree Kernel-based Unified Framework for Chinese Zero Anaphora Resolution

Fang Kong Guodong Zhou *

JiangSu Provincial Key Lab for Computer Information Processing Technology
School of Computer Science and Technology Soochow University
{kongfang, gdzhou}@suda.edu.cn

Abstract

This paper proposes a unified framework for zero anaphora resolution, which can be divided into three sub-tasks: zero anaphor detection, anaphoricity determination and antecedent identification. In particular, all the three sub-tasks are addressed using tree kernel-based methods with appropriate syntactic parse tree structures. Experimental results on a Chinese zero anaphora corpus show that the proposed tree kernel-based methods significantly outperform the feature-based ones. This indicates the critical role of the structural information in zero anaphora resolution and the necessity of tree kernel-based methods in modeling such structural information. To our best knowledge, this is the first systematic work dealing with all the three sub-tasks in Chinese zero anaphora resolution via a unified framework. Moreover, we release a Chinese zero anaphora corpus of 100 documents, which adds a layer of annotation to the manually-parsed sentences in the Chinese Treebank (CTB) 6.0.

1 Introduction

As one of the most important techniques in discourse analysis, anaphora resolution has been a focus of research in Natural Language Processing (NLP) for decades and achieved much success in English recently (e.g. Soon et al. 2001; Ng and Cardie 2002; Yang et al. 2003, 2008; Kong et al. 2009).

However, there is little work on anaphora resolution in Chinese. A major reason for this phe-

nomenon is that Chinese, unlike English, is a pro-drop language, whereas in English, definite noun phrases (e.g. *the company*) and overt pronouns (e.g. *he*) are frequently employed as referring expressions, which refer to preceding entities. Kim (2000) compared the use of overt subjects in English and Chinese. He found that overt subjects occupy over 96% in English, while this percentage drops to only 64% in Chinese. This indicates the prevalence of zero anaphors in Chinese and the necessity of zero anaphora resolution in Chinese anaphora resolution. Since zero anaphors give little hints (e.g. number or gender) about their possible antecedents, zero anaphora resolution is much more challenging than traditional anaphora resolution.

Although Chinese zero anaphora has been widely studied in the linguistics research (Li and Thompson 1979; Li 2004), only a small body of prior work in computational linguistics deals with Chinese zero anaphora resolution (Converse 2006; Zhao and Ng 2007). Moreover, zero anaphor detection, as a critical component for real applications of zero anaphora resolution, has been largely ignored.

This paper proposes a unified framework for Chinese zero anaphora resolution, which can be divided into three sub-tasks: zero anaphor detection, which detects zero anaphors from a text, anaphoricity determination, which determines whether a zero anaphor is anaphoric or not, and antecedent identification, which finds the antecedent for an anaphoric zero anaphor. To our best knowledge, this is the first systematic work dealing with all the three sub-tasks via a unified framework. Moreover, we release a Chinese zero anaphora corpus of 100 documents, which adds a layer of annotation to the

* Corresponding author

manually-parsed sentences in the Chinese Treebank (CTB) 6.0. This is done by assigning anaphoric/non-anaphoric zero anaphora labels to the null constituents in a parse tree. Finally, this paper illustrates the critical role of the structural information in zero anaphora resolution and the necessity of tree kernel-based methods in modeling such structural information.

The rest of this paper is organized as follows. Section 2 briefly describes the related work on both zero anaphora resolution and tree kernel-based anaphora resolution. Section 3 introduces the overwhelming problem of zero anaphora in Chinese and our developed Chinese zero anaphora corpus, which is available for research purpose. Section 4 presents our tree kernel-based unified framework in zero anaphora resolution. Section 5 reports the experimental results. Finally, we conclude our work in Section 6.

2 Related Work

This section briefly overviews the related work on both zero anaphora resolution and tree kernel-based anaphora resolution.

2.1 Zero anaphora resolution

Although zero anaphors are prevalent in many languages, such as Chinese, Japanese and Spanish, there only have a few works on zero anaphora resolution.

Zero anaphora resolution in Chinese

Converse (2006) developed a Chinese zero anaphora corpus which only deals with zero anaphora category “-NONE- *pro*” for dropped subjects/objects and ignores other categories, such as “-NONE- *PRO*” for non-overt subjects in non-finite clauses. Besides, Converse (2006) proposed a rule-based method to resolve the anaphoric zero anaphors only. The method did not consider zero anaphor detection and anaphoric identification, and performed zero anaphora resolution using the Hobbs algorithm (Hobbs, 1978), assuming the availability of golden anaphoric zero anaphors and golden parse trees.

Instead, Zhao and Ng (2007) proposed feature-based methods to zero anaphora resolution on the same corpus from Converse (2006). However, they only considered zero anaphors with explicit noun phrase referents and discarded those with split an-

tecedents or referring to events. Moreover, they focused on the sub-tasks of anaphoricity determination and antecedent identification. For zero anaphor detection, a simple heuristic rule was employed. Although this rule can recover almost all the zero anaphors, it suffers from very low precision by introducing too many false zero anaphors and thus leads to low performance in anaphoricity determination, much due to the imbalance between positive and negative training examples.

Zero anaphora resolution in Japanese

Seki et al. (2002) proposed a probabilistic model for the sub-tasks of anaphoric identification and antecedent identification with the help of a verb dictionary. They did not perform zero anaphor detection, assuming the availability of golden zero anaphors. Besides, their model needed a large-scale corpus to estimate the probabilities to prevent them from the data sparseness problem.

Isozaki and Hirao (2003) explored some ranking rules and a machine learning method on zero anaphora resolution. However, they assumed that zero anaphors were already detected and each zero anaphor’s grammatical case was already determined by a zero anaphor detector.

Iida et al. (2006) explored a machine learning method for the sub-task of antecedent identification using rich syntactic pattern features, assuming the availability of golden anaphoric zero anaphors.

Sasano et al. (2008) proposed a fully-lexicalized probabilistic model for zero anaphora resolution, which estimated case assignments for the overt case components and the antecedents of zero anaphors simultaneously. However, this model needed case frames to detect zero anaphors and a large-scale corpus to construct these case frames automatically.

For Japanese zero anaphora, we do not see any reports about zero anaphora categories. Moreover, all the above related works we can find on Japanese zero anaphora resolution ignore zero anaphor detection, focusing on either anaphoricity determination or antecedent identification. Maybe, it is easy to detect zero anaphors in Japanese. However, it is out of the scope of our knowledge and this paper.

Zero anaphora resolution in Spanish

As the only work we can find, Ferrandez and Peral (2000) proposed a hand-engineered rule-based method for both anaphoricity determination and

antecedent identification. That is, they ignored zero anaphor detection. Besides, they only dealt with zero anaphors that were in the subject position.

2.2 Tree kernel-based anaphora resolution

Although there is no research on tree kernel-based zero anaphora resolution in the literature, tree kernel-based methods have been explored in traditional anaphora resolution to certain extent and achieved comparable performance with the dominated feature-based ones. One main advantage of kernel-based methods is that they are very effective at reducing the burden of feature engineering for structured objects. Indeed, the kernel-based methods have been successfully applied to mine structural information in various NLP techniques and applications, such as syntactic parsing (Collins and Duffy 2001; Moschitti 2004), semantic relation extraction (Zelenko et al. 2003; Zhao and Grishman 2005; Zhou et al. 2007; Qian et al. 2008), and semantic role labeling (Moschitti 2004).

Representative works in tree kernel-based anaphora resolution include Yang et al. (2006) and Zhou et al (2008). Yang et al. (2006) employed a convolution tree kernel on anaphora resolution of pronouns. In particular, a document-level syntactic parse tree for an entire text was constructed by attaching the parse trees of all its sentences to a new-added upper node. Examination of three parse tree structures using different construction schemes (Min-Expansion, Simple-Expansion and Full-Expansion) on the ACE 2003 corpus showed promising results. However, among the three constructed parse tree structures, there exists no obvious overwhelming one, which can well cover structured syntactic information. One problem with this tree kernel-based method is that all the constructed parse tree structures are context-free and do not consider the information outside the subtrees. To overcome this problem, Zhou et al. (2008) proposed a dynamic-expansion scheme to automatically construct a proper parse tree structure for anaphora resolution of pronouns by taking predicate- and antecedent competitor-related information into consideration. Besides, they proposed a context-sensitive convolution tree kernel to compute the similarity between the parse tree structures. Evaluation on the ACE 2003 corpus showed that the dynamic-expansion scheme can well cover

necessary structural information in the parse tree for anaphora resolution of pronouns and the context-sensitive convolution tree kernel much outperformed other tree kernels.

3 Task Definition

This section introduces the phenomenon of zero anaphora in Chinese and our developed Chinese zero anaphora corpus.

3.1 Zero anaphora in Chinese

A zero anaphor is a gap in a sentence, which refers to an entity that supplies the necessary information for interpreting the gap. Figure 1 illustrates an example sentence from Chinese TreeBank (CTB) 6.0 (File ID=001, Sentence ID=8). In this example, there are four zero anaphors denoted as Φ_i ($i=1, 2, \dots, 4$). Generally, zero anaphors can be understood from the context and do not need to be specified.

A zero anaphor can be classified into either anaphoric or non-anaphoric, depending on whether it has an antecedent in the discourse. Typically, a zero anaphor is non-anaphoric when it refers to an extra linguistic entity (e.g. the first or second person in a conversation) or its referent is unspecified in the context. Among the four anaphors in Figure 1, zero anaphors Φ_1 and Φ_4 are non-anaphoric while zero anaphors Φ_2 and Φ_3 are anaphoric, referring to noun phrase “建筑行为/building action” and noun phrase “新区管委会/new district managing committee” respectively.

Chinese zero anaphora resolution is very difficult due to following reasons: 1) Zero anaphors give little hints (e.g. number or gender) about their possible antecedents. This makes antecedent identification much more difficult than traditional anaphora resolution. 2) A zero anaphor can be either anaphoric or non-anaphoric. In our corpus described in Section 3.2, about 60% of zero anaphors are non-anaphoric. This indicates the importance of anaphoricity determination. 3) Zero anaphors are not explicitly marked in a text. This indicates the necessity of zero anaphor detection, which has been largely ignored in previous research and has proved to be difficult in our later experiments.

Figure 2 illustrates an example sentence annotated in CTB 6.0, where the special tag “-NONE-” represents a null constituent and thus the occurrence of a zero anaphor. In our developed corpus, we need to annotate anaphoric zero anaphors using those null constituents with the special tag of “-NONE-”.

Table 1 gives the statistics on all the six categories of zero anaphora. Since we do not consider zero cataphora in the current version, we simply redeem them non-anaphoric. It shows that among 1720 zero anaphors, only 712 (about 40%) are anaphoric. This suggests the importance of anaphoricity determination in zero anaphora resolution. Table 3 further shows that, among 712 anaphoric zero anaphors, 598 (84%) are intra-sentential and no anaphoric zero anaphors have their antecedents occurring two sentences before.

Sentence distance	AZAs
0	598
1	114
>=2	0

Table 3 Distribution of anaphoric zero anaphors over sentence distances

Figure 3 shows an example in our corpus corresponding to Figure 2. For a non-anaphoric zero anaphor, we replace the null constituent with “E-*i* NZA”, where *i* indicates the category of zero anaphora, with “1” referring to “-NONE *T*” etc. For an anaphoric zero anaphor, we replace it with “E-*x-y-z-i* AZA”, where *x* indicates the sentence id of its antecedent, *y* indicates the position of the first word of its antecedent in the sentence, *z* indicates the position of the last word of its antecedent in the sentence, and *i* indicates the category id of the null constituent.

```
(IP (PP-PRP (P 为)
  (IP (NP-SBJ (E-3 NZP))
    (VP (VP (VV 规范)
      (NP-OBJ (NN 建筑)
        (NN 行为))))
    (PU , )
    (VP (VV 防止)
      (IP (NP-SBJ (E-7-1-1-4 AZP))
        (VP (VV 出现)
          (NP-OBJ (ADJP (JJ 无序)
            (NP (NN 现象))))))))))
```

Figure 3: an example sentence annotated in our corpus

4 Tree Kernel-based Framework

This section presents the tree kernel-based unified framework for all the three sub-tasks in zero anaphora resolution. For each sub-task, different parse tree structures are constructed. In particular, the context-sensitive convolution tree kernel, as proposed in Zhou et al. (2008), is employed to compute the similarity between two parse trees via the SVM toolkit SVMLight.

In the tree kernel-based framework, we perform the three sub-tasks, zero anaphor detection, anaphoricity determination and antecedent identification in a pipeline manner. That is, given a zero anaphor candidate *Z*, the zero anaphor detector is first called to determine whether *Z* is a zero anaphor or not. If yes, the anaphoricity determiner is then invoked to determine whether *Z* is an anaphoric zero anaphor. If yes, the antecedent identifier is finally awakened to determine its antecedent. In the future work, we will explore better ways of integrating the three sub-tasks (e.g. joint learning).

4.1 Zero anaphor detection

At the first glance, it seems that a zero anaphor can occur between any two constituents in a parse tree. Fortunately, an exploration of our corpus shows that a zero anaphor always occurs just before a predicate¹ phrase node (e.g. VP). This phenomenon has also been employed in Zhao and Ng (2007) in generating zero anaphor candidates. In particular, if the predicate phrase node occurs in a coordinate structure or is modified by an adverbial node, we only need to consider its parent. As shown in Figure 1, zero anaphors may occur immediately to the left of 规范/guide, 防止/avoid, 出现/appear, 根据/according to, 结合/combine, 出台/promulgate, which cover the four true zero anaphors. Therefore, it is simple but reliable in applying above heuristic rules to generate zero anaphor candidates.

Given a zero anaphor candidate, it is critical to construct a proper parse tree structure for tree kernel-based zero anaphor detection. The intuition behind our parser tree structure for zero anaphor detection is to keep the competitive information

¹ The predicate in Chinese can be categorized into verb predicate, noun predicate and preposition predicate. In our corpus, about 93% of the zero anaphors are driven by verb predicates. In this paper, we only explore zero anaphors driven by verb predicates.

about the predicate phrase node and the zero anaphor candidate as much as possible. In particular, the parse tree structure is constructed by first keeping the path from the root node to the predicate phrase node and then attaching all the immediate verbal phrase nodes and nominal phrase nodes. Besides, for the sub-tree rooted by the predicate phrase node, we only keep those paths ended with verbal leaf nodes and the immediate verbal and nominal nodes attached to these paths. Figure 4 shows an example of the parse tree structure corresponding to Figure 1 with the zero anaphor candidate $\Phi 2$ in consideration.

During training, if a zero anaphor candidate has a counterpart in the same position in the golden standard corpus (either anaphoric or non-anaphoric), a positive instance is generated. Otherwise, a negative instance is generated. During testing, each zero anaphor candidate is presented to the learned zero anaphor detector to determine whether it is a zero anaphor or not. Besides, since a zero anaphor candidate is generated when a predicate phrase node appears, there may be two or more zero anaphor candidates in the same position. However, there is normally one zero anaphor in the same position. Therefore, we just select the one with maximal confidence as the zero anaphor in the position and ignore others, if multiple zero anaphor candidates occur in the same position.

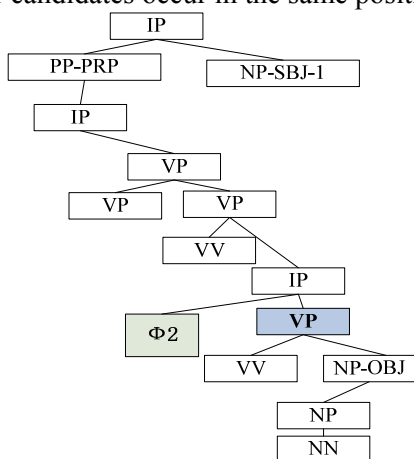


Figure 4: An example parse tree structure for zero anaphor detection with the predicate phrase node and the zero anaphor candidate $\Phi 2$ in black

4.2 Anaphoricity determination

To determine whether a zero anaphor is anaphoric or not, we limit the parse tree structure between the

previous predicate phrase node and the following predicate phrase node. Besides, we only keep those verbal phrase nodes and nominal phrase nodes. Figure 5 illustrates an example of the parse tree structure for anaphoricity determination, corresponding to Figure 1 with the zero anaphor $\Phi 2$ in consideration.

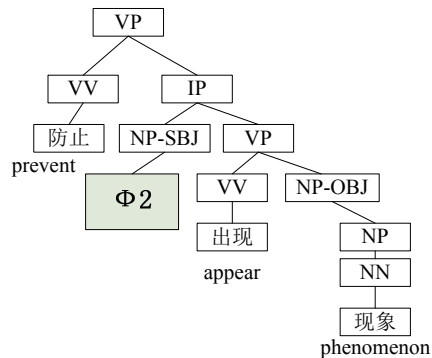


Figure 5: An example parse tree structure for anaphoricity determination with the zero anaphor $\Phi 2$ in consideration

4.3 Antecedent identification

To identify an antecedent for an anaphoric zero anaphor, we adopt the Dynamic Expansion Tree, as proposed in Zhou et al. (2008), which takes predicate- and antecedent competitor-related information into consideration. Figure 6 illustrates an example parse tree structure for antecedent identification, corresponding to Figure 1 with the anaphoric zero anaphor $\Phi 2$ and the antecedent candidate “建筑行为/building action” in consideration.

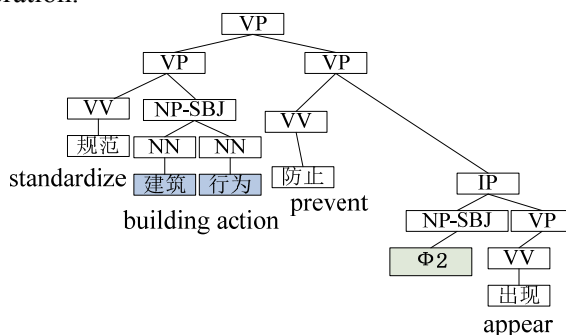


Figure 6: An example parse tree structure for antecedent identification with the anaphoric zero anaphor $\Phi 2$ and the antecedent candidate “建筑行为/building action” in consideration

In this paper, we adopt a similar procedure as Soon et al. (2001) in antecedent identification. Be-

sides, since all the anaphoric zero anaphors have their antecedents at most one sentence away, we only consider antecedent candidates which are at most one sentence away. In particular, a document-level parse tree for an entire document is constructed by attaching the parse trees of all its sentences to a new-added upper node, as done in Yang et al. (2006), to deal with inter-sentential ones.

5 Experimentation and Discussion

We have systematically evaluated our tree kernel-based unified framework on our developed Chinese zero anaphora corpus, as described in Section 3.2. Besides, in order to focus on zero anaphor resolution itself and compare with related work, all the experiments are done on golden parse trees provided by CTB 6.0. Finally, all the performances are achieved using 5-fold cross validation.

5.1 Experimental results

Zero anaphor detection

Table 4 gives the performance of zero anaphor detection, which achieves 70.05%, 83.24% and 76.08 in precision, recall and F-measure, respectively. Here, the lower precision is much due to the simple heuristic rules used to generate zero anaphors candidates. In fact, the ratio of positive and negative instances reaches about 1:12. However, this ratio is much better than that (1:30) using the heuristic rule as described in Zhao and Ng (2007). It is also worth to point out that lower precision higher recall is much beneficial than higher precision lower recall as higher recall means less filtering of true zero anaphors and we can still rely on anaphoricity determination to filter out those false zero anaphors introduced by lower precision in zero anaphor detection.

P%	R%	F
70.05	83.24	76.08

Table 4: Performance of zero anaphor detection

Anaphoricity determination

Table 5 gives the performance of anaphoricity determination. It shows that anaphoricity determination on golden zero anaphors achieves very good performance of 89.83%, 84.21% and 86.93 in precision, recall and F-measure, respectively, although useful information, such as gender and number, is not available in anaphoricity determination. This

indicates the critical role of the structural information in anaphoricity determination of zero anaphors. It also shows that anaphoricity determination on automatic zero anaphor detection achieves 77.96%, 53.97% and 63.78 in precision, recall and F-measure, respectively. In comparison with anaphoricity determination on golden zero anaphors, anaphoricity determination on automatic zero anaphor detection lowers the performance by about 23 in F-measure. This indicates the importance and the necessity for further research in zero anaphor detection.

	P%	R%	F
golden zero anaphors	89.83	84.21	86.93
zero anaphor detection	77.96	53.97	63.78

Table 5: Performance of anaphoricity determination

Antecedent identification

Table 6 gives the performance of antecedent identification given golden zero anaphors. It shows that antecedent identification on golden anaphoric zero anaphors achieves 88.93%, 68.36% and 77.29 in precision, recall and F-measure, respectively. It also shows that antecedent identification on automatic anaphoricity determination achieves 80.38%, 47.28% and 59.24 in precision, recall and F-measure, respectively, with a decrease of about 8% in precision, about 21% in recall and about 18% in F-measure, in comparison with antecedent identification on golden anaphoric zero anaphors. This indicates the critical role of anaphoricity determination in antecedent identification.

	P%	R%	F
golden anaphoric zero anaphors	88.90	68.36	77.29
anaphoricity determination	80.38	47.28	59.54

Table 6: Performance of antecedent identification given golden zero anaphors

Overall: zero anaphora resolution

Table 7 gives the performance of overall zero anaphora resolution with automatic zero anaphor detection, anaphoricity determination and antecedent identification. It shows that our tree kernel-based framework achieves 77.66%, 31.74% and 45.06 in precision, recall and F-measure. In comparison with Table 6, it shows that the errors caused by automatic zero anaphor detection decrease the performance of overall zero anaphora resolution by about 14 in F-measure, in comparison with golden zero anaphors.

P%	R%	F
77.66	31.74	45.06

Table 7: Performance of zero anaphora resolution

Figure 7 shows the learning curve of zero anaphora resolution with the increase of the number of the documents in experimentation, with the horizontal axis the number of the documents used and the vertical axis the F-measure. It shows that the F-measure is about 42.5 when 20 documents are used in experimentation. This figure increases very fast to about 45 when 50 documents are used while further increase of documents only slightly improves the performance.

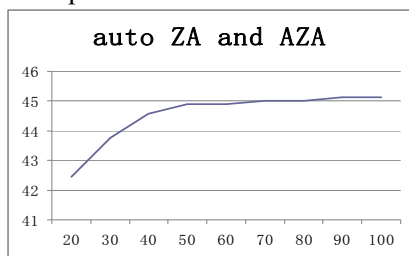


Figure 7: Learning curve of zero anaphora resolution over the number of the documents in experimentation

Table 8 shows the detailed performance of zero anaphora resolution over different sentence distance between a zero anaphor and its antecedent. It is expected that both the precision and the recall of intra-sentential resolution are much higher than those of inter-sentential resolution, largely due to the much more dependency of intra-sentential antecedent identification on the parse tree structures.

Sentence distance	P%	R%	F
0	85.12	33.28	47.85
1	46.55	23.64	31.36
2	-	-	-

Table 8: Performance of zero anaphora resolution over sentence distances

Table 9 shows the detailed performance of zero anaphora resolution over the two major zero anaphora categories, “-NONE- *PRO*” and “-NONE- *pro*”. It shows that our tree kernel-based framework achieves comparable performance on them, both with high precision and low recall. This is in agreement with the overall performance.

ID	Category	P%	R%	F
3	-NONE- *PRO*	79.37	34.23	47.83
4	-NONE- *pro*	77.03	30.82	44.03

Table 9: Performance of zero anaphora resolution over major zero anaphora categories

5.2 Comparison with previous work

As a representative in Chinese zero anaphora resolution, Zhao and Ng (2007) focused on anaphoricity determination and antecedent identification using feature-based methods. In this subsection, we will compare our tree kernel-based framework with theirs in details.

Corpus

Zhao and Ng (2007) used a private corpus from Converse (2006). Although their corpus contains 205 documents from CBT 3.0, it only deals with the zero anaphors under the zero anaphora category of “-NONE- *pro*” for dropped subjects/objects. Furthermore, Zhao and Ng (2007) only considered zero anaphors with explicit noun phrase referents and discarded zero anaphors with split antecedents (i.e. split into two separate noun phrases) or referring to entities. As a result, their corpus is only about half of our corpus in the number of zero anaphors and anaphoric zero anaphors. Besides, our corpus deals with all the types of zero anaphors and all the categories of zero anaphora except zero cataphora.

Method

Zhao and Ng (2007) applied feature-based methods on anaphoricity determination and antecedent identification with most of features structural in nature. For zero anaphor detection, they used a very simple heuristic rule to generate zero anaphor candidates. Although this rule can recover almost all the zero anaphors, it suffers from very low precision by introducing too many false zero anaphors and thus may lead to low performance in anaphoricity determination, much due to the imbalance between positive and negative training examples with the ratio up to about 1:30.

In comparison, we propose a tree kernel-based unified framework for all the three sub-tasks in zero anaphora resolution. In particular, different parse tree structures are constructed for different sub-tasks. Besides, a context sensitive convolution tree kernel is employed to directly compute the similarity between the parse trees.

For fair comparison with Zhao and Ng (2007), we duplicate their system and evaluate it on our developed Chinese zero anaphora corpus, using the same J48 decision tree learning algorithm in Weka and the same feature sets for anaphoricity determination and antecedent identification.

Table 10 gives the performance of the feature-based method, as described in Zhao and Ng (2007), in anaphoricity determination on our developed corpus. In comparison with the tree kernel-based method in this paper, the feature-based method performs about 16 lower in F-measure, largely due to the difference in precision (63.61% vs 89.83%), when golden zero anaphors are given. It also shows that, when our tree kernel-based zero anaphor detector is employed², the feature-based method gets much lower precision with a gap of about 31%, although it achieves slightly higher recall.

	P%	R%	F
golden zero anaphors	63.61	79.71	70.76
zero anaphor detection	46.17	57.69	51.29

Table 10: Performance of the feature-based method (Zhao and Ng 2007) in anaphoricity determination on our developed corpus

	P%	R%	F
golden anaphoric zero anaphors	77.45	51.97	62.20
golden zero anaphors and feature-based anaphoricity determination	75.17	29.69	42.57
overall: tree kernel-based zero anaphor detection and feature-based anaphoricity determination	70.67	23.64	35.43

Table 11: Performance of the feature-based method (Zhao and Ng 2007) in antecedent identification on our developed corpus

Table 11 gives the performance of the feature-based method, as described in Zhao and Ng (2007), in antecedent identification on our developed corpus. In comparison with our tree kernel-based method, it shows that 1) when using golden anaphoric zero anaphors, the feature-based method performs about 11%, 17% and 15 lower in precision, recall and F-measure, respectively; 2) when golden zero anaphors are given and feature-based anaphoricity determination is applied, the feature-based method performs about 5%, 18% and 17 lower in precision, recall and F-measure, respectively; and 3) when tree kernel-based zero anaphor detection and feature-based anaphoricity determination are applied, the feature-based method per-

² We do not apply the simple heuristic rule, as adopted in Zhao and Ng (2007), in zero anaphor detection, due to its much lower performance, for fair comparison on the other two sub-tasks..

forms about 7%, 8% and 10 lower in precision, recall and F-measure, respectively.

In summary, above comparison indicates the critical role of the structural information in zero anaphora resolution, given the fact that most of features in the feature-based methods in Zhao and Ng (2007) are also structural, and the necessity of tree kernel methods in modeling such structural information, even if more feature engineering in the feature-based methods may improve the performance to a certain extent.

6 Conclusion and Further Work

This paper proposes a tree kernel-based unified framework for zero anaphora resolution, which can be divided into three sub-tasks: zero anaphor detection, anaphoricity determination and antecedent identification.

The major contributions of this paper include: 1) We release a wide-coverage Chinese zero anaphora corpus of 100 documents, which adds a layer of annotation to the manually-parsed sentences in the Chinese Treebank (CTB) 6.0. 2) To our best knowledge, this is the first systematic work dealing with all the three sub-tasks in Chinese zero anaphora resolution via a unified framework. 3) Employment of tree kernel-based methods indicates the critical role of the structural information in zero anaphora resolution and the necessity of tree kernel methods in modeling such structural information.

In the future work, we will systematically evaluate our framework on automatically-generated parse trees, construct more effective parse tree structures for different sub-tasks of zero anaphora resolution, and explore joint learning among the three sub-tasks.

Besides, we only consider zero anaphors driven by a verb predicate phrase node in this paper. In the future work, we will consider other situations. Actually, among the remaining 7% zero anaphors, about 5% are driven by a preposition phrase (PP) node, and 2% are driven by a noun phrase (NP) node. However, our preliminary experiments show that simple inclusion of those PP-driven and NP-driven zero anaphors will largely increase the imbalance between positive and negative instances, which significantly decrease the performance.

Finally, we will devote more on further developing our corpus, with the ultimate mission of annotating all the documents in CBT 6.0.

Acknowledgments

This research was supported by Projects 60873150, 90920004 and 61003153 under the National Natural Science Foundation of China.

References

- S. Converse. 2006. *Pronominal Anaphora Resolution in Chinese*. Ph.D. Thesis, Department of Computer and Information Science. University of Pennsylvania.
- M. Collins and N. Duffy. 2001. *Convolution kernels for natural language*. NIPS'2001:625-632.
- A. Ferrandez and J. Peral. 2000. *A computational approach to zero-pronouns in Spanish*. ACL'2000:166-172.
- R. Iida, K. Inui, and Y. Matsumoto. 2006. *Exploiting syntactic patterns as clues in zero-anaphora resolution*. COLING-ACL'2006:625-632
- H. Isozaki and T. Hirao. 2003. *Japanese zero pronoun resolution based on ranking rules and machine learning*. EMNLP'2003:184-191
- F. Kong, G.D. Zhou and Q.M. Zhu. 2009 *Employing the Centering Theory in Pronoun Resolution from the Semantic Perspective*. EMNLP'2009: 987-996
- C. N. Li and S. A. Thompson. 1979. *Third-person pronouns and zero-anaphora in Chinese discourse*. Syntax and Semantics, 12:311-335.
- W. Li. 2004. *Topic chains in Chinese discourse*. Discourse Processes, 37(1):25-45.
- A. Moschitti. 2004. *A Study on Convolution Kernels for Shallow Semantic Parsing*, ACL'2004.
- L.H. Qian, G.D. Zhou, F. Kong, Q.M. Zhu and P.D. Qian. 2008. *Exploiting constituent dependencies for tree kernel-based semantic relation extraction*. COLING'2008:697-704
- K. Seki, A. Fujii, and T. Ishikawa. 2002. *A probabilistic method for analyzing Japanese anaphora integrating zero pronoun detection and resolution*. COLING'2002:911-917
- R. Sasano, D. Kawahara and S. Kurohashi. 2008. *A fully-lexicalized probabilistic model for Japanese zero anaphora resolution*. COLING'2008:769-776
- W.M. Soon, H.T. Ng and D. Lim. 2001. *A machine learning approach to coreference resolution of noun phrase*. Computational Linguistics, 2001, 27(4):521-544.
- V. Ng and C. Cardie 2002. *Improving machine learning approaches to coreference resolution*. ACL'2002: 104-111
- X.F. Yang, G.D. Zhou, J. Su and C.L. Chew. 2003. *Coreference Resolution Using Competition Learning Approach*. ACL'2003:177-184
- X.F. Yang, J. Su and C.L. Tan 2008. *A Twin-Candidate Model for Learning-Based Anaphora Resolution*. Computational Linguistics 34(3):327-356
- N. Xue, F. Xia, F.D. Chiou and M. Palmer. 2005. *The Penn Chinese TreeBank: Phrase structure annotation of a large corpus*. Natural Language Engineering, 11(2):207-238.
- X.F. Yang, J. Su and C.L. Tan. 2006. *Kernel-based pronoun resolution with structured syntactic knowledge*. COLING-ACL'2006:41-48.
- D. Zelenko, A. Chinatsu and R. Anthony. 2003. *Kernel methods for relation extraction*. Journal of Machine Learning Research, 3(2003):1083-1106
- S. Zhao and H.T. Ng. 2007. *Identification and Resolution of Chinese Zero Pronouns: A Machine Learning Approach*. EMNLP-CoNLL'2007:541-550.
- S. Zhao and R. Grishman. 2005. *Extracting relations with integrated information using kernel methods*. ACL'2005:419-426
- G.D. Zhou, F. Kong and Q.M. Zhu. 2008. *Context-sensitive convolution tree kernel for pronoun resolution*. IJCNLP'2008:25-31
- G.D. Zhou, M. Zhang, D.H. Ji and Q.M. Zhu. 2007. *Tree kernel-based relation extraction with context-sensitive structured parse tree information*. EMNLP-CoNLL'2007:728-736