

Exploiting Zero Pronouns to Improve Chinese Coreference Resolution

Fang Kong

Department of Computer Science
National University of Singapore
13 Computing Drive
Singapore 117417
dcskf@nus.edu.sg

Hwee Tou Ng

Department of Computer Science
National University of Singapore
13 Computing Drive
Singapore 117417
nght@comp.nus.edu.sg

Abstract

Coreference resolution plays a critical role in discourse analysis. This paper focuses on exploiting zero pronouns to improve Chinese coreference resolution. In particular, a simplified semantic role labeling framework is proposed to identify clauses and to detect zero pronouns effectively, and two effective methods (refining syntactic parser and refining learning example generation) are employed to exploit zero pronouns for Chinese coreference resolution. Evaluation on the CoNLL-2012 shared task data set shows that zero pronouns can significantly improve Chinese coreference resolution.

1 Introduction

As one of the most important tasks in discourse analysis, coreference resolution aims to link a given mention (i.e., entity or event) to its co-referring expression in a text and has been a focus of research in natural language processing (NLP) for decades.

Over the last decade, various machine learning techniques have been applied to coreference resolution and have performed reasonably well (Soon et al., 2001; Ng and Cardie, 2002; Fernandes et al., 2012). Current techniques rely primarily on surface level features such as string match, syntactic features such as apposition, and shallow semantic features such as number, gender, semantic class, etc.

Despite similarities between Chinese and English, there are differences that have a significant impact on coreference resolution. In this paper, we focus on exploiting one of the key characteristics of Chinese text, zero pronouns (ZPs), to improve Chinese

coreference resolution. In particular, a simplified semantic role labeling (SRL) framework is proposed to identify Chinese clauses and to detect zero pronouns effectively, and two effective methods are employed to exploit zero pronouns for Chinese coreference resolution. Experimental results show the effectiveness of our approach in improving the performance of Chinese coreference resolution. Our work is novel in that it is the first work that incorporates the use of zero pronouns to significantly improve Chinese coreference resolution

The rest of this paper is organized as follows. Section 2 describes our baseline Chinese coreference resolution system. Section 3 motivates how the detection of zero pronouns can improve Chinese coreference resolution, using an illustrating example. Section 4 presents our approach to detect zero pronouns. Section 5 proposes two methods to exploit zero pronouns to improve Chinese coreference resolution, based on a corpus study and preliminary experiments. Section 6 briefly outlines the related work. Finally, we conclude our work in Section 7.

2 Chinese Coreference Resolution

According to Webber (1978), coreference resolution can be decomposed into two complementary subtasks: (1) anaphoricity determination: deciding whether a given noun phrase (NP) is anaphoric or not; and (2) anaphora resolution: linking together multiple mentions of a given entity in the world. Our Chinese coreference resolution system also contains these two components. Using the training data set of CoNLL-2012 shared task, we first train an anaphoricity classifier to determine whether

a mention is anaphoric or not, and then employ an independently-trained coreference resolution system to resolve those mentions which are classified as anaphoric. The lack of gender and number makes both anaphoricity determination and coreference resolution in Chinese more difficult.

2.1 Anaphoricity Determination

Since only the mentions that take part in coreference chains are annotated in the CoNLL-2012 shared task data set, we first generate a high-recall, low-precision mention extraction module to extract as many mentions as possible. The mention extraction module relies mainly on syntactic parse trees. We extract all NP nodes, QP (quantifier phrase, i.e., complex amount/measure phrase) nodes, and all terminals with part-of-speech tags PN (pronoun) and NR (proper noun) in parse trees to form a mention candidate set. Then, we employ some rules to remove unlikely mentions, e.g., those which contain (1) measure words such as ‘一年/one year’ and ‘一顿/one time’; (2) named entities whose categories are PERCENT, MONEY, QUANTITY, and CARDINAL; (3) interrogative pronouns such as ‘什么/what’ and ‘哪儿/where’.

After pruning, we employ a learning-based method to train an independent classifier to determine whether the remaining mentions are anaphoric. Table 1 lists all the features employed in our anaphoricity determination system.

2.2 Coreference Resolution

Our Chinese coreference resolution system adopts the same learning-based model and the same set of 12 features as Soon et al. (2001). Considering the special characteristics of conversation and web texts (i.e., a large proportion of personal pronouns and the organization of a text into several parts¹) and preparing for dealing with zero pronouns, we add some features shown in Table 2.²

¹A text in the CoNLL-2012 data set is broken down into different “parts”.

²AN denotes anaphor, CA denotes antecedent candidate, IP denotes a simple clause, and CP denotes a clause headed by a complementizer. For the feature ANPronounRanking, the relative ranking of a given pronoun is based on its semantic role and surface position, and we assign the highest rank to zero pronouns, similar to Kong et al. (2009).

	R	P	F
GS	76.32	87.14	81.37
Auto	64.87	78.42	71.00

Table 3: Performance of anaphoricity determination on the CoNLL-2012 test set

		R	P	F
AM	Mention Detection	65.26	67.20	66.22
	MUC	51.64	61.82	56.27
	BCUBED	73.40	80.38	76.73
	CEAF	53.16	45.66	49.13
	Average			60.71
GMB	Mention Detection	82.01	69.58	75.29
	MUC	76.21	66.18	70.84
	BCUBED	76.15	86.59	81.04
	CEAF	59.75	50.52	54.75
	Average			68.88
GM	Mention Detection	79.80	100.00	88.77
	MUC	80.86	85.48	83.11
	BCUBED	73.66	91.94	81.79
	CEAF	67.54	64.87	66.18
	Average			77.02

Table 4: Performance of our Chinese coreference resolution system on the CoNLL-2012 test set

2.3 Results and Analysis

All experiments in this section are conducted on the CoNLL-2012 shared task data set. The SVM-light toolkit (Joachims, 1999) with radial basis kernel and default learning parameters is employed in both anaphoricity determination and coreference resolution.

Table 3 reports the performance of anaphoricity determination on the CoNLL-2012 test set using gold-standard parse trees (GS) and automatic parse trees (Auto). All performance figures in this paper are given in percentages. The results show that using both gold parse trees and automatic parse trees, our anaphoricity determination system achieves higher precision than recall. In comparison with using gold parse trees, precision decreases by about 9% and recall 11% on automatic parse trees.

Table 4 reports the performance of our Chinese coreference resolution system on the CoNLL-2012 test set under three different experimental settings: with automatic mentions (AM), with gold mention

Feature	Description
NPType	Type of the current mention (pronoun, demonstrative, proper NP).
NPNumber	Number of the current mention (singular, plural).
NPGender	Gender of the current mention (male, female).
IsHeadWord	Whether the current mention is the same as its headword.
StrMatch	Whether there is a string match between the current mention and another phrase in the previous context.
AliasMatch	Whether the current mention is a name alias or abbreviation of another phrase in the previous context.
Appositive	Whether the current mention and another phrase in the previous context are in an appositive relation.
NestIn	Whether another NP is nested in the current mention.
NestOut	Whether the current mention is nested in another NP.
FirstNP	Whether the current mention is the first NP of the sentence.
FrontDistance	The number of words between the current mention and the nearest previous clause.
BackDistance	The number of words between the current mention and the nearest following clause.
WordSense	Whether the current mention and another phrase in the previous context have the same word sense. Word sense annotation is provided in the CoNLL-2012 data set, based on the IMS software (Zhong and Ng, 2010).

Table 1: Features employed in our anaphoricity determination system

Feature	Description
AN/CAPronounType	Whether the anaphor or the antecedent candidate is a zero pronoun, first person, second person, third person, neutral pronoun, or others. In our coreference resolution system, a zero pronoun is viewed as a kind of special pronoun.
AN/CAGrammaticalRole	Whether the anaphor or the antecedent candidate is a subject, object, or others.
AN/CAOwnerClauseType	Whether the anaphor or the antecedent candidate is in a matrix clause, an independent clause, a subordinate clause, or none of the above.
AN/CARootPath	Whether the path of nodes from the anaphor (or the antecedent candidate) to the root of the parse tree contains NP, IP, CP, or VP.
ANPronounRanking	Whether the anaphor is a pronoun and is ranked highest among the pronouns (including zero pronouns) of the sentence.
AN/CAClosestNP	Whether the antecedent candidate is the closest preceding NP of the anaphor.
AN/CAPartDistance	This feature captures the distance (in parts) between the antecedent candidate and the anaphor. If they are in the same part, the value is 0; if they are one part apart, the value is 1; and so on.
AN/CASameSpeaker	Whether the antecedent candidate and the anaphor appear in sentences spoken by the same person.

Table 2: Additional features employed in our Chinese coreference resolution system

boundaries (GMB), and with gold mentions (GM). From the results, we find that:

- Using automatic mentions, our system achieves 56.27, 76.73, and 49.13 in F-measure on MUC, BCUBED, and CEAF evaluation metrics, respectively.
- Using gold mention boundaries improves the performance of our system by 14.57, 4.31, and 5.62 in F-measure, due to large gains in both recall and precision. We also find that using gold mention boundaries can boost the recall of mention detection. As described above, our anaphoricity determination model relies mainly on the parser. Using gold mention boundaries can improve the parser performance. Thus our coreference resolution system can benefit much from using gold mention boundaries (especially the recall).
- Employing gold mentions further boosts our system significantly. In comparison with using gold mention boundaries, the performance improvement is attributed more to an increase in precision.

In comparison with the three best systems of CoNLL-2012 in the Chinese closed track (shown in Table 5), considering average F-measure, we find that using automatic mentions, our system is only inferior to that of Chen and Ng (2012); using gold mention boundaries, our system achieves the best performance; and using gold mentions, our system is only a little worse than that of Chen and Ng (2012).

3 Motivation

In order to analyze the impact of zero pronouns on Chinese coreference resolution, we first use the released OntoNotes v5.0 data (i.e., the training and development portions of the CoNLL-2012 shared task) in a corpus study.

Statistics show that anaphoric zero pronouns account for 10.7% of the mentions in coreference chains in the training data, while in the development data, the proportion is 11.3%. The experimental results of our Chinese coreference resolution system (i.e., the baseline) show that using both gold mention boundaries and gold mentions significantly

improves system performance, especially for recall, largely due to improved parser performance. We then analyze the impact of zero pronouns on Chinese syntactic parsing. As a preliminary exploration, we integrate Chinese zero pronouns into the Berkeley parser (Petrov et al., 2006), experimenting with gold-standard or automatically determined zero pronouns kept or stripped off (using gold-standard word segmentation provided in the CoNLL-2012 data). The results indicate that given gold-standard zero pronouns, parsing performance improves by 1.8% in F-measure. Using automatically determined zero pronouns by our zero pronoun detector to be introduced in Section 4, parsing performance also improves by 1.4% in F-measure.

In order to illustrate the impact of zero pronouns on parsing performance, consider the following example:³

Example (1):

将来我们有一个**重建计划**。

#分公园成七个区域，#带来多一些的景点。

...

这个计划我们现在是等到政府的批准我们就可以再进行，预算是#明年可以动工了。

(In future, we have a **reconstruction plan**.

Divide the park into seven regions, and bring some more attractions.

...

Now we wait for approval of the government before implementing **this plan** again. It is expected that work can start next year.)

Without considering zero pronouns, the parse tree of the second sentence output by the Berkeley parser is shown in Figure 1.

Prior to parsing, using our zero pronoun detector to be introduced in Section 4, the presence of zero pronouns (denoted by #) can be detected. Figure 2

³In this paper, zero pronouns are denoted by “#” and mentions in the same coreference chain are shown in bold for all examples.

		MD	MUC	BCUBED	CEAF	Avg
AM	(Chen and Ng, 2012)	71.64	62.21	73.55	50.97	62.24
	(Yuan et al., 2012)	68.15	60.33	72.90	48.83	60.69
	(Björkelund and Farkas, 2012)	66.37	58.61	73.10	48.19	59.97
	Our baseline system (without ZPs)	66.22	56.27	76.73	49.13	60.71
	Our refined system (with auto ZPs)	70.33	59.58	78.15	51.47	63.07
GMB	(Chen and Ng, 2012)	80.45	71.43	77.04	57.17	68.55
	(Yuan et al., 2012)	74.02	66.44	75.02	51.81	64.42
	(Björkelund and Farkas, 2012)	71.02	63.56	74.52	50.20	62.76
	Our baseline system (without ZPs)	75.29	70.84	81.04	54.75	68.88
	Our refined system (with auto ZPs)	75.77	72.62	81.45	58.04	70.70
GM	(Chen and Ng, 2012)	91.73	83.77	81.15	68.38	77.77
	(Yuan et al., 2012)	89.95	82.79	79.79	65.58	76.05
	(Björkelund and Farkas, 2012)	83.47	76.85	76.30	56.61	69.92
	Our baseline system (without ZPs)	88.77	83.11	81.79	66.18	77.02
	Our refined system (with auto ZPs)	91.49	83.46	82.43	65.88	77.26

Table 5: Performance (F-measure) of the three best Chinese coreference resolution systems on the CoNLL-2012 test set

shows the new parse tree, which includes the detected zero pronouns, output by the Berkeley parser on the same sentence. Comparing these two parse trees, we can see that the detected zero pronouns contribute to better division of clauses and improved parsing performance, which in turn leads to improved Chinese coreference resolution.

Detecting the presence of zero pronouns also helps to improve local salience modeling, leading to improved Chinese coreference resolution. Long sentences containing multiple clauses occur more frequently in Chinese compared to English. Furthermore, a coreference chain can span many sentences. Zero pronouns can occur not only within one sentence (e.g., the first and second zero pronouns of Example (1)), but can also be scattered across multiple sentences (e.g., the first and third zero pronouns of Example (1)). The subjects in the second sentence of Example (1) are omitted.⁴ Detection of zero pronouns improves local salience modeling, and leads to the correct identification of all the noun phrases of the coreference chain in Example (1).

4 Zero Pronoun Detection

Empty elements are those nodes in a parse tree that do not have corresponding surface words or phrases. Although empty elements exist in many languages

⁴In Chinese, pro-dropped subjects account for more than 36% of subjects in sentences (Kim, 2000).

and serve different purposes, they are particularly important for some languages, such as Chinese, where subjects and objects are frequently dropped to keep a discourse concise. Among empty elements, type *pro*, namely zero pronoun, is either used for dropped subjects or objects, which can be recovered from the context (anaphoric), or it is of little interest for the reader or listener to know (non-anaphoric). In the Chinese Treebank, type *pro* constitutes about 20% (Yang and Xue, 2010), and more than 85% of them are anaphoric (Kong and Zhou, 2010). Thus, zero pronouns are very important in bridging the information gap in a Chinese text. In this section, we will introduce our zero pronoun detector.

In Chinese, a zero pronoun always occurs just before a predicate phrase node (e.g., VP). 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. A simplified semantic role labeling (SRL) framework (only including predicate recognition, argument pruning, and argument identification) is adopted to identify the predicate phrase subtree (Xue, 2008), i.e., the minimal subtree governed by a predicate and all its arguments.

We carry out zero pronoun detection for every predicate phrase subtree in an iterative manner from a parse tree, i.e., determining whether there is a zero pronoun before the given predicate phrase sub-

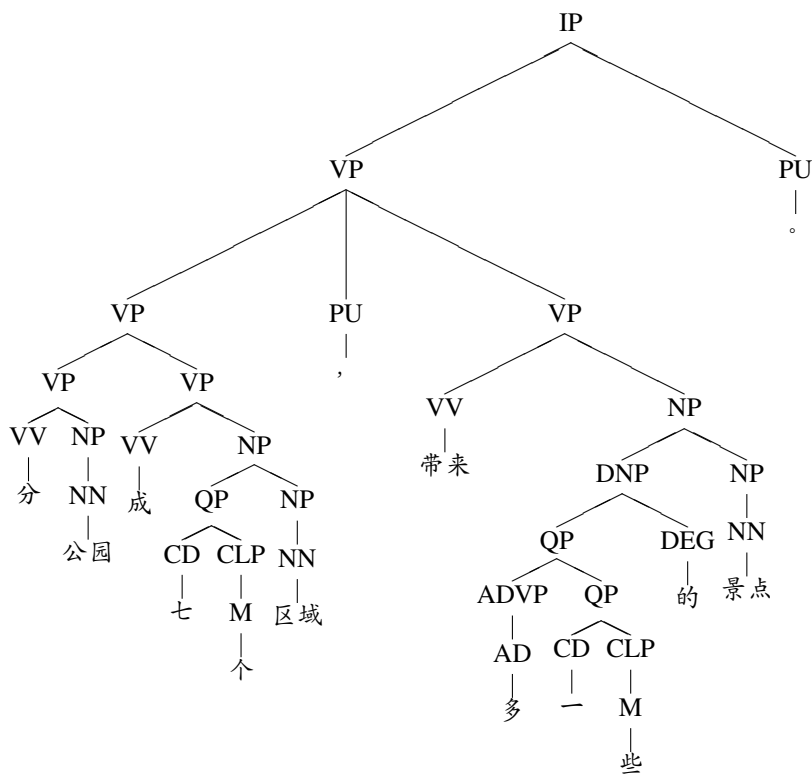


Figure 1: The parse tree without considering zero pronouns

tree. Viewing the position before the given predicate phrase subtree as a zero pronoun candidate, we can perform zero pronoun detection using a machine learning approach.

During training, if a zero pronoun candidate has a counterpart in the same position in the annotated training corpus (either anaphoric or non-anaphoric), a positive example is generated. Otherwise, a negative example is generated. During testing, each zero pronoun candidate is presented to the zero pronoun detector to determine whether it is a zero pronoun.

The features that are employed to detect zero pronouns mainly model the context of the clause itself, the left and right siblings, and the path of the clause to the root node. Table 6 lists the features in detail.

4.1 Results and Analysis

We evaluate our zero pronoun detector using gold parse trees and automatic parse trees produced by the Berkeley parser. The SVM-light toolkit with radial basis kernel and default learning parameters is employed as our learning algorithm.

Table 7 lists the results. From the results, we

	R	P	F
GS	89.32	87.29	88.29
Auto	74.19	77.79	75.95

Table 7: Performance of zero pronoun detection on the test set using gold and automatic parse trees

find that the performance of our zero pronoun detector drops about 12% in F-measure when using automatic parse trees, compared to using gold parse trees. That is, the performance of zero pronoun detection also depends on the performance of the syntactic parser.

5 Exploiting Zero Pronouns to Improve Chinese Coreference Resolution

In this section, we will propose two methods, refining the syntactic parser and refining learning example generation, to exploit zero pronouns to improve Chinese coreference resolution.

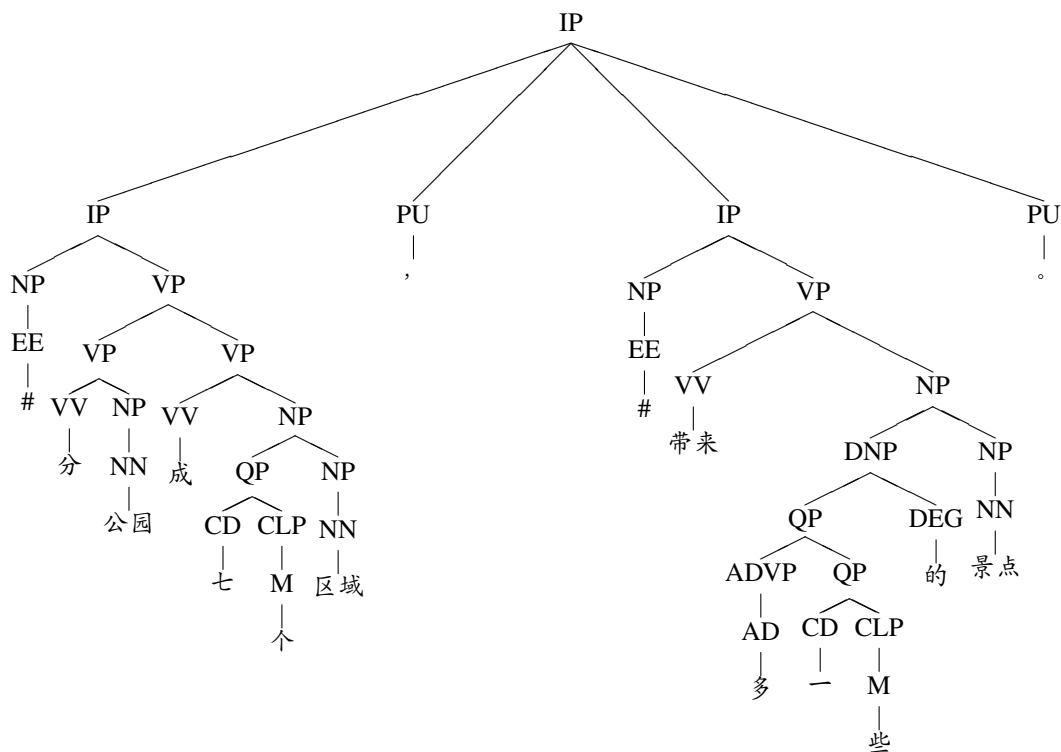


Figure 2: The parse tree with the detected zero pronouns

5.1 Refining the Syntactic Parser

Similar to our preliminary experiments, we retrain the Berkeley parser with explicit, automatically detected zero pronouns in the training set and parse the test set with explicit, automatically detected zero pronouns using the retrained model. In both anaphoricity determination and coreference resolution, the output results of the retrained parser are employed to generate all features.

5.2 Refining Learning Example Generation

In order to model the salience of all entities, we regard all zero pronouns as a special kind of NPs when generating the learning examples. Considering the modest performance of our anaphoricity determination module, we do not determine the anaphoricity of zero pronouns. Instead, in the coreference resolution stage, all zero pronouns will be considered during learning example generation (including both training and test example generation).

For example, consider a coreference chain A1-A2-Z0-A3-A4 containing one zero pronoun found in an annotated training document. A1, A2, A3, and A4 are traditional entity mentions, and Z0 is a

zero pronoun. During training, pairs of mentions in the chain that are immediately adjacent (i.e., A1-A2, A2-Z0, Z0-A3, and A3-A4) are used to generate the positive training examples. Among them, two examples (i.e., A2-Z0 and Z0-A3) are associated with a zero pronoun, which can act as both an anaphor and an antecedent. For each positive pair, e.g., Z0-A3, we find any noun phrase and zero pronoun occurring between the anaphor A3 and the antecedent Z0, and pair each of them with A3 to form a negative example. Similarly, test examples can be generated except that only the preceding mentions and zero pronouns in the current and previous two sentences will be paired with an anaphor.

Incorporating zero pronouns models salience of all entities more accurately. The ratio of positive to negative examples is also less skewed as a result of considering zero pronouns – the ratio changes from 1:7.9 to 1:6.8 after considering zero pronouns.

5.3 Reprocessing

Although in the OntoNotes corpus, dropped subjects and objects (i.e., zero pronouns) are considered during coreference resolution for Chinese, they are not

Feature	Description
ClauseClass	Whether the given clause is a terminal clause or non-terminal clause.
LeftSibling	Whether the given clause has a sibling immediately to its left.
LeftSiblingNP	Whether the left siblings of the given clause contain an NP.
RightSibling	Whether the given clause has a sibling immediately to its right.
RightSiblingVP	Whether the right siblings of the given clause contain a VP.
ParentIP/VP	Whether the syntactic category of the immediate parent of the given clause is an IP or VP.
RootPath	Whether the path from the given clause to the root of the parse tree contains an NP or VP or CP. This feature models how the given clause is syntactically connected to the sentence as a whole, reflecting its function within the sentence.
ClauseType	The given clause is an independent clause, a subordinate clause, or others.
Has-Arg0/Arg1	Whether the given clause has an agent or patient argument.

Table 6: Features employed to detect zero pronouns

used in the CoNLL-2012 shared task (i.e., in the gold evaluation keys, all the links formed by zero pronouns are removed).

As described in Subsection 5.2, during training and testing, all links associated with zero pronouns will be considered in our coreference resolution system. That is, we do not distinguish zero pronoun resolution from traditional coreference resolution, and only view zero pronouns as special pronouns. After generating all the links, zero pronouns are included in coreference chains. For every coreference chain, all zero pronouns will be removed before evaluation.

5.4 Experimental Results and Analysis

For fair comparison, all our experiments in this subsection have been conducted using the same experimental settings as our baseline system. When compared to our baseline system, all improvements are statistically significant ($p < 0.005$).

Table 8 lists the coreference resolution performance incorporating automatically detected zero pronouns. The results show that:

- Using automatically detected zero pronouns achieves better performance under all experimental settings. In particular, using automatic mentions, performance improves by 3.31%, 1.42%, and 2.34% in F-measure on the MUC, BCUBED, and CEAF evaluation metric, respectively. Using gold mention boundaries, automatic zero pronouns contribute 1.82% in average F-measure. Using gold mentions, the

		R	P	F
AM	Mention Detection	71.09	69.58	70.33
	MUC	55.06	64.91	59.58
	BCUBED	76.04	80.38	78.15
	CEAF	53.98	49.19	51.47
	Average			63.07
GMB	Mention Detection	82.44	70.10	75.77
	MUC	75.58	69.89	72.62
	BCUBED	76.35	87.27	81.45
	CEAF	65.17	52.31	58.04
	Average			70.70
GM	Mention Detection	84.31	100.00	91.49
	MUC	80.83	86.27	83.46
	BCUBED	74.18	92.74	82.43
	CEAF	69.91	62.29	65.88
	Average			77.26

Table 8: Performance of our Chinese coreference resolution system incorporating zero pronouns

contribution of zero pronouns is only 0.24% in average F-measure. This is because employing either gold mention boundaries or gold mentions improves parsing performance.

- Our system incorporating zero pronouns outperforms the three best systems in the CoNLL-2012 shared task when using automatic mentions or gold mention boundaries. Using gold mentions, our average F-measure is slightly lower than that of Chen and Ng (2012).⁵

Table 9 presents the contribution of our two methods of exploiting zero pronouns and the impact of gold-standard zero pronouns. We conclude that:

- Both the refined parser and refined example generation improve performance. While the refined parser improves the recall of mention detection and coreference resolution, refined example generation contributes more to precision. Combining these two methods further improves coreference resolution.
- There is a performance gap of 6.01%, 4.08%, and 3.19% in F-measure on the MUC, BCUBED, and CEAF evaluation metric, respectively, between the coreference resolution system with gold-standard zero pronouns and without zero pronouns. This suggests the usefulness of zero pronoun detection in Chinese coreference resolution.
- Our proposed methods incorporating automatic zero pronouns reduce the performance gap by about half. This shows the effectiveness of our proposed methods.

5.5 Discussion

Although the evaluation of the CoNLL-2012 shared task does not consider zero pronouns, we also evaluate the performance of zero pronoun resolution on the development data set (i.e., extracting all the resolved coreference links containing zero pronouns, acting as anaphor or antecedent, to conduct the evaluation independently). The results show that, for the correct anaphoric zero pronouns, the precision

⁵Statistical significance testing cannot be conducted since their output files are not released.

of our system is 94.76%. So viewing zero pronouns as a special kind of NP, zero pronouns can bridge salience and contribute to coreference resolution. In Example (1), the zero pronouns occurring in the second sentence help to bridge the coreferential relation between the mention “这个计划/this plan” in the last sentence and the mention “一个重建计划/a reconstruction plan” in the first sentence.

6 Related Work

In the last decade, both manual rule-based approaches (Lee et al., 2011) and statistical approaches (Soon et al., 2001; Ng and Cardie, 2002; Fernandes et al., 2012) have been proposed for coreference resolution. Besides frequently used syntactic and semantic features, more linguistic features are exploited in recent work (Ponzetto and Strube, 2006; Ng, 2007; Versley, 2007). There is less research on Chinese coreference resolution compared to English.

Although zero pronouns are prevalent in Chinese, there is relatively little work on this topic. For Chinese zero pronoun resolution, representative work includes Converse (2006), Zhao and Ng (2007), and Kong and Zhou (2010).

For the use of zero pronouns, Chung and Gildea (2010) applied some extracted patterns to recover two types of empty elements (*PRO* and *pro*). Although the performance is still not satisfactory (e.g., 63.0 and 44.0 in F-measure for *PRO* and *pro* respectively), it nevertheless improves machine translation performance by 0.96 in BLEU score.

7 Conclusion

In this paper, we focus on exploiting one of the key characteristics of Chinese text, zero pronouns, to improve Chinese coreference resolution. In particular, a simplified semantic role labeling framework is proposed to detect zero pronouns effectively, and two effective methods are employed to incorporate zero pronouns into Chinese coreference resolution. Experiments on the CoNLL-2012 shared task show the effectiveness of our proposed approach. To the best of our knowledge, this is the first attempt at incorporating zero pronouns into Chinese coreference resolution.

	MD			MUC			BCUBED			CEAF			Avg
	R	P	F	R	P	F	R	P	F	R	P	F	
Baseline	65.26	67.20	66.22	51.64	61.82	56.27	73.40	80.38	76.73	53.16	45.66	49.13	60.71
+RP	72.01	66.24	69.00	55.02	61.47	58.07	77.83	78.97	78.40	50.40	49.81	50.10	62.19
+REG	65.92	70.02	67.91	49.98	66.27	56.98	73.64	83.45	78.24	51.12	47.44	49.21	61.48
+AZPs	71.09	69.58	70.33	55.06	64.91	59.58	76.04	80.38	78.15	53.98	49.19	51.47	63.07
+GZPs	72.18	70.59	71.38	58.61	66.45	62.28	78.79	82.94	80.81	54.12	50.63	52.32	65.14

Table 9: Contributions of the two methods of incorporating zero pronouns and the impact of gold zero pronouns (RP: refining parser using auto zero pronouns, REG: refining example generation using auto zero pronouns, AZPs: combining both RP and REG using auto zero pronouns, and GZPs: combining both RP and REG using gold zero pronouns)

Acknowledgments

This research is supported by the Singapore National Research Foundation under its International Research Centre @ Singapore Funding Initiative and administered by the IDM Programme Office.

References

- Anders Björkelund and Richárd Farkas. 2012. Data-driven multilingual coreference resolution using resolver stacking. In *Proceedings of the Joint Conference on EMNLP and CoNLL – Shared Task*, pages 49–55.
- Chen Chen and Vincent Ng. 2012. Combining the best of two worlds: A hybrid approach to multilingual coreference resolution. In *Proceedings of the Joint Conference on EMNLP and CoNLL – Shared Task*, pages 56–63.
- Tagyoung Chung and Daniel Gildea. 2010. Effects of empty categories on machine translation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 636–645.
- Susan Converse. 2006. *Pronominal Anaphora Resolution in Chinese*. Ph.D. thesis, University of Pennsylvania.
- Eraldo Rezende Fernandes, Cícero Nogueira dos Santos, and Ruy Luiz Milidiú. 2012. Latent structure perceptron with feature induction for unrestricted coreference resolution. In *Proceedings of the Joint Conference on EMNLP and CoNLL – Shared Task*, pages 41–48.
- Thorsten Joachims. 1999. Making large-scale SVM learning practical. In Bernhard Schölkopf, Christopher J. C. Burges, and Alexander J. Smola, editors, *Advances in Kernel Methods: Support Vector Learning*. MIT-Press.
- Young-Joo Kim. 2000. Subject/object drop in the acquisition of Korean: A cross-linguistic comparison. *Journal of East Asian Linguistics*, 9:325–351.
- Fang Kong and Guodong Zhou. 2010. A tree kernel-based unified framework for Chinese zero anaphora resolution. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 882–891.
- Fang Kong, Guodong Zhou, and Qiaoming Zhu. 2009. Employing the centering theory in pronoun resolution from the semantic perspective. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 987–996.
- Heeyoung Lee, Yves Peirsman, Angel Chang, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. 2011. Stanford’s multi-pass sieve coreference resolution system at the CoNLL-2011 shared task. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, pages 28–34.
- Vincent Ng and Claire Cardie. 2002. Improving machine learning approaches to coreference resolution. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 104–111.
- Vincent Ng. 2007. Semantic class induction and coreference resolution. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, pages 536–543.
- Slav Petrov, Leon Barrett, Romain Thibaux, and Dan Klein. 2006. Learning accurate, compact, and interpretable tree annotation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 433–440.
- Simone Paolo Ponzetto and Michael Strube. 2006. Exploiting semantic role labeling, WordNet and Wikipedia for coreference resolution. In *Proceedings of the Human Language Technology Conference of the NAACL*, pages 192–199.
- Wee Meng Soon, Hwee Tou Ng, and Daniel Chung Yong Lim. 2001. A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27(4):521–544.

- Yannick Versley. 2007. Antecedent selection techniques for high-recall coreference resolution. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 496–505.
- Bonnie Lynn Webber. 1978. *A Formal Approach to Discourse Anaphora*. Garland Press.
- Nianwen Xue. 2008. Labeling Chinese predicates with semantic roles. *Computational Linguistics*, 34(2):225–255.
- Yaqin Yang and Nianwen Xue. 2010. Chasing the ghost: Recovering empty categories in the Chinese Treebank. In *Coling 2010: Posters*, pages 1382–1390.
- Bo Yuan, Qingcai Chen, Yang Xiang, Xiaolong Wang, Liping Ge, Zengjian Liu, Meng Liao, and Xianbo Si. 2012. A mixed deterministic model for coreference resolution. In *Proceedings of the Joint Conference on EMNLP and CoNLL – Shared Task*, pages 76–82.
- Shanheng Zhao and Hwee Tou Ng. 2007. Identification and resolution of Chinese zero pronouns: A machine learning approach. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 541–550.
- Zhi Zhong and Hwee Tou Ng. 2010. It Makes Sense: A wide-coverage word sense disambiguation system for free text. In *Proceedings of the ACL 2010 System Demonstrations*, pages 78–83.