Using Semantic Roles for Coreference Resolution

Conference Paper · August 2008

DOI: 10.1109/ALPIT.2008.60 · Source: IEEE Xplore

Using Semantic Roles for Coreference Resolution

KONG Fang LIYancui ZHOU Guodong ZHU Qiaoming QIAN Peide*
Jiangsu Provincial Key Lab for Computer Information Processing Technology
School of Computer Science and Technology
Soochow Univ. Suzhou, China 215006
Email: { kongfang, 210513043, gdzhou, qmzhu,pdqian}@suda.edu.cn

Abstract

In this paper, we systematically explore the use of semantic roles in coreference resolution. Here, the semantic roles are automatically determined using a state-of-the-art SRL system, ASSERT, and integrated into the resolution by combining them with detailed pronoun categories. In particular, the anaphoricity of the pronoun "it" in "it" sentence pattern is determined using its semantic role and applied in a filter process. Experimental results on the ACE 2003 corpus show that our method can significantly outperform the baseline system by 3.1% in F-measure.

1. Introduction

Coreference resolution is the process of linking together multiple expressions of a given entity. The key to solve this problem is to determine the antecedent for each referring expression in a document. Now coreference resolution has been proven to be a major obstacle in building robust systems for information extraction, text summarization, question answering and so on.

In the last decade, knowledge-learn approaches have significantly influenced the research of coreference resolution. Current coreference resolution systems are mostly relying on lexical and syntactic attribute of the noun phrases, such as string matching, the distance between the coreferent expressions, name alias, and so on (Soon et al.,2001;Ng and Cardie,2002). However, there is much more knowledge in the small, annotated corpus. And deeper linguistic knowledge needs to be made available to resolver in order to reach the next level of performance. As a result, researchers have re-adopted the once-popular knowledge-rich approach. Some researchers focus on investigating a variety of semantic knowledge source for common noun resolution, such as the semantic relations between two NPs (e.g. Ji et al.,2005), the contextual role played by an NP (e.g. Bean and Riloff, 2004), and statistics-based semantic compatibility information (e.g. Yang,2005). Other researchers try to apply some syntactic structured information into pronoun resolution. Shane Bergsma and Dekang Lin (2006) presented an approach to pronoun resolution based on syntactic paths. Through a simple bootstrapping procedure, the approach achieved the likelihood of coreference between a pronoun and a candidate noun based on the path in the parse tree between the two entities. Yang et al. (2006) proposed a kernel-based method that can automatically mine the syntactic information from the parse trees. They utilized the parse trees directly as a structured feature and apply kernel functions to this feature. Zhou et al. (2008) proposed a context-sensitive convolution tree kernel for pronoun resolution. For a pair of an anaphor and an antecedent candidate, they implemented a dynamic-expansion scheme to automatically determine a proper tree span by taking predicate- and antecedent competitor-related information into consideration. Then they applied a context-sensitive convolution tree kernel for pronoun resolution.

So far, there are still some problems. Whether a given noun phrase is anaphoric or non-anaphoric is very important for subsequent resolution. Consider the following instances extracted from ACE 2003:

(1) *The rebel Sudan*$_1$ *People 's Liberation Army*$_2$ said Thursday that it$_3$ will extend by another three months the current three-month cease-fire when it end Wednesday to allow relief food to reach civilians in war provinces like Bahr el-Ghazal in the south .

(2) It is destiny, he kept repeating to no avail .

In example 1, the pronoun "it3" is anaphoric, but in example 2, the pronoun "It" is non-anaphoric. How to distinguish between these two phenomena is still a hard work. In this paper, we present an approach using semantic role knowledge to resolve this problem. For the example 2, the result of semantic role labeling is:

* QianPeide: Corresponding Author

978-0-7695-3273-8/08 $25.00 © 2008 IEEE
DOI 10.1109/ALPIT.2008.60

It's destiny, [ARG0 he] kept [TARGET repeating] to no avail.

We can see that the pronoun “It” is a non-anaphoric.

Another contribution of our work lies in combining semantic role information with detailed pronoun type knowledge in coreference resolution. This combination can use semantic role knowledge effectively.

The rest of this paper is organized as follows. Section 2 briefly describes some related work that utilizes the semantic knowledge for resolution. Section 3 introduces the framework, as well as the baseline feature space. Section 4 presents in detail the semantic role features. Section 5 reports and discusses the experimental results, and finally, section 6 gives the conclusion.

2. Related Work

Semantic information is an important factor. Related work on exploring semantic information in coreference resolution can be typically classified into three categories: computing semantic similarity between two NPs, contextual role played by NP, semantic class of an NP.

So far, there are several coreference resolution systems using WordNet, HowNet etc. to compute the semantic similarity between two NPs. Vieira & Poesic(2000), Harabagiu et al.(2001), Poesio et al.(2004), Markert & Nissim(2005) and Wikipedia(2006) explore the use of WordNet for coreference resolution. By means of a WordNet search, the results in terms of semantic similarity measures can be used as features for a learner.

As a representative for using contextual role played by NP in coreference resolution, Yang (2005) considered three relationships: possessive-noun, subject-verb and verb-object. Before resolution a large corpus was prepared. Documents in the corpus were processed by a shallow parser that could generate predicate-argument tuples. During resolution, for an encountered anaphor, each of its antecedent candidates was substituted with the anaphor. According to the role and type of the anaphor in its context, a predicate-argument tuple was extracted. Each extracted tuple was searched in the prepared tuples set of the corpus, and the times the tuple occurs was calculated. Because corpus-based knowledge usually suffers from data sparseness problem, Yang also obtained the predicate-argument statistics via a web search engine like Google and Altavista. Then, the statistics from corpus and web combined with twin-candidate model to compute and apply the semantic information. Yang’s study shows that the semantic information can significantly improve the resolution of neutral pronouns. However, Yang does not consider other relations outside the above three. Another problem is that there is no approach, which can be used to extract semantic relationships automatically.

As for semantic class of an NP, some approaches simply assign to a common noun the first (i.e., most frequent) WordNet sense as its semantic class (e.g., Soon et al.(2001), Markert and Nissim (2005)). Ng (2007) acquired semantic class knowledge from a version of the Penn Treebank in which the noun phrases were labeled with their semantic class. Ng adopted a corpus-based approach to semantic class determination. He classified an NP as belonging to one of the ACE semantic classes, so the task of semantic class determination can be known as a six-class classification task.

The above discussion suggests that semantic information can significantly improve the resolution. So far, there are still much semantic information may not be well utilized. So the task of this paper is to find out whether a learning based coreference resolver can be improved by using semantic role knowledge that is automatically acquired by some semantic role labeling tools, and how to use semantic role knowledge effectively.

3. The Resolution Framework

Our correference resolution system adopts the common learning-based framework similar to the one by Soon et al.(2001).

Table 1: Feature set of baseline system

ANPronoun	1 if anaphor is a pronoun; else 0
ANDefiniteNP	1 if anaphor is a definite NP; else 0
ANDemonstrativeNP	1 if anaphor is a demonstrative NP; else 0
CAPronoun	if candidate is a pronoun; else 0
ANCAGenderAgreement	1 if candidate and anaphor agree in gender; else 0
ANCANumberAgreement	1 if candidate and anaphor agree in number; else 0
ANCAAppositive	1 if candidate and anaphor are in an appositive structure; else 0
ANCAHeadStringMatch	1 if candidate and anaphor match in headword; else 0
ANCASentDistance	Distance between candidate and anaphor in sentences
ANCAWORDSENSE	1 if candidate and anaphor agree in the sense achieved from WordNet; else 0
ANCABothProperName	1 if candidate and anaphor are both proper name; else 0
ANCANameAlias	1 if candidate and anaphor are in an alias of the other; else 0

During training, all possible markables in a training document are determined by a pipeline of NLP components, and training examples in the form of feature vectors which are generated for appropriate pairs of markables. These training examples are then given to SVM to build a classifier. During testing, all markables are determined by the same pipeline, and potential pairs of noun phrase's markables are presented to the classifier, which decides whether the two noun phrase corefer.

As many other learning-based approaches, the knowledge for the reference determination is represented as a set of features associated with the training or test instances. In our baseline system, we only select features that can be obtained with low annotation cost and high reliability. And these features must be generic enough to be used across different domains. All features are listed in Table 1 together with their prospective possible values. They are same with the features of Soon(2001). In our research, we duplicate the Soon system as a baseline system.

4. Incorporating Semantic Role

4.1 Centering Theory

Centering theory is a theory of local discourse structure that models the interaction of referential continuity and salience of discourse entities in the internal organization of a text. The main assumptions of the theory as presented by Grosz, Joshi and Weinstein (1995) and Walker, Joshi and Prince (1998) can be summarized as follows:

- For each utterance in a discourse there is precisely one entity that is the center of attention.
- For consecutive utterances within a discourse, segment should keep the same entity as the center of attention.
- The entity most prominently realized in an utterance should be identified as the center of attention.
- The center of attention is the entity that is most likely to be pronominalized.

The main claims of centering theory are formalized in terms of Cb, the backward-looking center; Cf, a list of forward-looking centers for each utterance Un; and Cp or preferred center, the most salient candidate for subsequent utterances. Cf(Un) is a partial ordering on the entities mentioned (or "realized") in Un, ranked by grammatical role; for example, SUBJ > DIR-OBJ >INDIR-OBJ > COMP(S) > ADJUNCT(S). Cb(Un) is defined as the highest-ranked member of Cf(Un-1) that is realized in Un. Cp(Un) is the highest-ranked member of Cf(Un), and is predicted to be Cb(Un+1).

The ranking of Cf based on Brennan, Friedman, and Pollard (1987) is Subject > Object(s) > Other. That is to say discourse entities realized in subject position is ranked more highly than entities realized in object position, which are both then ranked more highly than entities realized in subordinate clauses or as other grammatical functions. So far, the ranking of Cf by grammatical role has been widely adopted in many NLP applications. However, for English, there is still a problem of the utility of grammatical role. We must consider voice. Using the active voice or using the passive voice, the grammatical role of the same entity should be different. So in this paper, we apply semantic role into resolution.

4.2 Semantic Role

A semantic role is the underlying relationship that a participant has with the main verb in a clause. Semantic role is the actual role a participant plays in some real or imagined situation, apart from the linguistic encoding of those situations.

If, in some real or imagined situation, someone named John purposely hits someone named Bill, then John is the agent and Bill is the patient of the hitting event. Therefore, the semantic role of Bill is the same (patient) in both of the following sentences:

- John hit Bill.
- Bill was hit by John.

In both of the above sentences, John has the semantic role of agent.

During the last few years there has been increasing interest in Semantic Role Labeling (SRL). It is currently a well defined task with a substantial body of work and comparative evaluation. Given a sentence, the task consists of analyzing the propositions expressed by some target verbs and some constituents of the sentence. In particular, for each target verb (predicate) all the constituents in the sentence which fill a semantic role of the verb have to be recognized. Typical semantic roles include Agent, Patient, Instrument, etc. and also adjuncts such as Locative, Temporal, Manner, and Cause, etc. Among semantic roles, agent and patient are steady. Knowledge about other roles (e.g. Locative, Temporal) can be expressed as Word sense or Semantic class.

In this paper, we conjecture that candidate is mostly an antecedent of anaphor if it is the center of one utterance and the ranking is Agent > Patient > Others.

Our approach introduces some semantic role features to describe the semantic role of anaphor and candidate. Considering the stability of semantic role, our research only focused on the most important semantic roles (Arg0 and Arg1), other roles and some information about predicate-argument tuples were not concerned. Semantic role features are listed in table 2. Semantic role were labeled based on the results of a chunk tagger. We used ASSERT as the SRL system,, developed by Sameer Pradhan and others at University of Colorado.

Table 2: Semantic Role Features

ANArg0	1 if the semantic role of anaphor is arg0 else 0
ANArg1	1 if the semantic role of anaphor is arg1 else 0
CAArg0	1 if the semantic role of candidate is arg0 else 0
CAArg0	1 if the semantic role of candidate is arg1 else 0

4.3 Detailed Pronoun Type Information

In order to use semantic knowledge effectively, we applied some detailed pronoun type information into resolution. Based on centering theory and SRL, we know that (1) the center of attention is the entity that is most likely to be pronominalized; (2) Arg0 and Arg1 have higher ranking than other semantic roles. So we think semantic role knowledge should be useful for pronoun resolution. Detailed pronoun type feature can describe the pronoun more exactly, and which should be helpful for applying semantic role knowledge. Table 3 presents all the detailed pronoun type features.

Table 3: Detailed Pronoun Type Features

ANFirstPersonPronoun	1 if anaphor is a first personal pronoun; else 0
ANSecondPersonPronoun	1 if anaphor is a second personal pronoun; else 0
ANThirdPersonPronoun	1 if anaphor is a third personal pronoun; else 0
CAFirstPersonPronoun	1 if candidate is a first personal pronoun; else 0
CASecondPersonPronoun	1 if candidate is a second personal pronoun; else 0
CAThirdPersonPronoun	1 if candidate is a third personal pronoun; else 0

5. Experimentation and Discussion

5.1 Experimental Setup

The experiments were done on ACE 2003 NWIRE corpus. For the training set, there are totally 4773 Arg0 and 8351 Arg1. While for the testing set, the number is 1379 and 2416.

An input raw text was preprocessed automatically by a pipeline of NLP components, including tokenization and sentence segmentation, named entity recognition, part-of-speech tagging and noun phrase chunking. Among them, named entity recognition, part-of-speech tagging and text chunking apply the same Hidden Markov Model (HMM) based engine with error-driven learning capability (Zhou and Su, 2000 & 2002) which achieves precision and recall rate of 96.49 and 96.99 for noun phrases of Penn WSJ TreeBank . The recognition of NEs as well as their semantic categories was trained for the MUC NE task and obtained high F-scores of 96.9% (MUC-6) and 94.3%(MUC-7)(Zhou and Su, 2002). Throughout the experiments, default learning parameters were applied to the SVM algorithm. We report performance in terms of recall, precision, and F-measure using the commonly-used model theoretic MUC scoring program (Vilain et al., 1995).

5.2 Results and Discussion

In our study, we want to know whether a learning-based coreference resolver can be improved using semantic role knowledge and how to use the semantic role knowledge effectively. Table 4 summarizes the results using different feature groups.

Table 4: Results of different feature groups

Features	R	P	F
Duplicated Soon Baseline	55.0	53.5	54.2
+Semantic Role Features	52.7	51.9	52.3
+Detailed Pronoun Type Features	54.0	51.9	52.9
+Semantic Role Features & +Detailed Pronoun Type Features	53.4	60.8	56.9
+Filter the non-anaphoric "it"	53.3	58.0	55.6
+Semantic Role Features & +Detailed Pronoun Type Features & +Filter the non-anaphoric "it"	53.4	61.8	57.3

We can see that only introducing semantic role features or detailed pronoun type features, recall and precision of the system are all lowered, and F-measure are also lowered. But the combination of semantic role and detailed pronoun type features can improve the system. The F-measure is 2.7% higher than baseline system. Furthermore, we analyze "it" sentence pattern based on semantic role and apply a non-anaphor filter of "it" into resolution. The last two lines of Table 4 show that the filter of "it" is very useful. The combination of semantic role, detailed pronoun type and filter of non-anaphoric "it" can yield a statistically significant improvement of 3.1% in F-measure.

The second line of Table 4 shows that only applying semantic role features will lower the performance. Semantic role emphasizes the center of one utterance, and the center is the entity that is most likely to be pronominalized. But the baseline system does not distinguish the types of pronoun. Only applying semantic role knowledge can introduce some noise. Figure 1 gives out an example.

In this example, the phrases “the authorities”, “their” and “they” lies on same coreference chain. In baseline system, the pronoun “their” and “they” can be resolved correctly. But using semantic role knowledge separately, the system can not identify the relationship between “they” and “their”. Based on semantic role, for the verb “search”, the role of “they” is Arg0. And for the verb “send”, the role of “they” is also Arg0. At same time, the pronoun “their” has no semantic role. So the relationship between “their” and “they” is lost. In the same way, applying detailed pronoun type information separately is useless, by contraries, the knowledge will introduce some noise.

<s> <COREF ID = " 3053 " REF = " 3055 " > Mascola </COREF> said <COREF ID = " 2441 " REF = " 2445 " > the authorities </COREF> may stop <COREF ID = " 2446 " REF = " 2441 " > their </COREF> practice of detaining <COREF ID = " 3365 " REF = " -1 " > people </COREF> for hours while <COREF ID = " 2447 " REF = " 2446 " > they </COREF> search a threatened building and instead send <COREF ID = " 3366 " REF = " 3365 " > people </COREF> home if there is no clear evidence of danger . </s>

Figure 1: a segment from ACE2003 NWIRE

Incorporating semantic role knowledge with detailed pronoun type information, the center of one utterance is emphasized, and at the same time, the type of pronoun will be matched. The errors of mismatch person pronoun will be eliminated. So the performance is improved. Consider the example described in figure 2.

We are also concerned about how the semantic role knowledge works for different types of noun phrases. Table 5 lists the resolution results for five types of noun phrases. As shown, using semantic role and detailed pronoun type features, the system can significantly boost the performance of the baseline for pronoun resolution.

- Result of Baseline System：

Acting in <COREF ID="32" REF="13" >his</COREF> own defense during <COREF ID="33" REF="13" >his</COREF> pre-sentencing hearing Monday , <COREF ID="34" REF="7" >Akayesu</COREF> said remorsefully , I regret what happened from the bottom of my heart .

- Result using Semantic Role and Detailed Pronoun Type Features

Acting in <COREF ID="32" REF="13" >his</COREF> own defense during <COREF ID="33" REF="13" >his</COREF> pre-sentencing hearing Monday , <COREF ID="34" REF="7" >Akayesu</COREF> said remorsefully , <COREF ID="35" REF="13" >I</COREF> regret what happened from the bottom of <COREF ID="36" REF="13" >my</COREF> heart .

Figure 2: a segment from ACE2003 NWIRE

In English, there are many different use of “it”. E.g. “This is a new car. I bought it yesterday.“ The pronoun “it“ is an anaphor, the antecedent is the noun phrase “a new car“. But in some “it” sentence pattern, “it” is non-anaphoric. E.g. “It is destiny, he kept repeating to no avail.” The neutral pronoun “it” should be a non-anaphor. Because its semantic role is neither arg0 nor arg1. The last two lines show that the filter of non-anaphoric “it” can improve the performance.

Table 5: resolution results for different types of noun phrases

Type of NP	Baseline System			Using Semantic Role and Detailed Pronoun Type Features			Changes
	R	P	F	R	P	F	
Proper Name	80.8	82.5	81.7	80.8	82.5	81.7	0
Pronoun	59.2	59.3	59.2	66.5	59.1	62.6	+3.4
UnDefiniteNP	33.8	40.3	36.8	41.5	31.7	36.0	-0.8
DefiniteNP	34.7	45.6	39.4	35.8	44.5	39.6	+0.2
DemonstrativeNP	90.9	10.0	18.0	90.9	10.0	18.0	0

6 Conclusion

This paper systematically explores the use of semantic role information in coreference resolution. This is done in two aspects. First, semantic role information is determined using an existing SRL system and integrated with detailed pronoun type information. Evaluation on the ACE 2003 corpus shows that such integration improves the performance by 2.7% in F-measure. Second, the semantic role information is used to filter out the non-anaphoric “it”. This further improve the F-measure by 0.4%.

For the future work, we will explore more kinds of semantic information in coreference resolution.

7 Acknowledgement

This research is supported by Project 60673041 under the National Natural Science Foundation of China, project 2006AA01Z147 under the “863”

National High-Tech Research and Development of China, and project 20060285008 under the National Research Foundation for the Doctoral Program of Higher Education of China.

8 References

[1] Brennan, Susan E, Marilyn W. Friedman, and Carl J. Pollard (1987). “A centering approach to pronouns”. Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics, Cambridge, Mass., 290-2.

[2] Grosz, Barbara J., and Candace L. Sidner (1986). “Attentions, intentions and the structure of discourse”. Computational Linguistics, 12: 175-204.

[3] Brennan, Susan E (1995) “Centering attention in discourse”. Language and Cognitive Process, 10/2: 137-67.

[4] Aone C and Bennett W.W. (1995). Evaluating automated and manual acquisition of anaphora resolution strategies. ACL’1995:122-129.

[5] Grosz, A. JoShi and S. Weinstein. 1995. Centering: a framework for modeling the local coherence of discourse. Computational Linguistics, 21(2):203-225.

[6] Mitkov R. (1998). Robust pronoun resolution with limited knowledge. COLING-ACL’1998:869-875. Montreal, Canada.

[7] Collins M. (1999) Head-driven statistical models for natural language parsing. Ph.D. Thesis. University of Pennsylvania.

[8] Rodger Kibble, Goldsmiths College. (2001). A Reformulation of Rule 2 of Centering. Computational Linguistics, 2001,27(4): 579-587

[9] Soon W.M., Ng H.T. and Lim. (2001). A machine learning approach to coreference resolution of noun phrase. Computational Linguistics, 2001,27(4):521544.

[10] Ng V. and Cardie C. (2002). Improving machine learning approaches to coreference resolution. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, 2002.

[11] Yang X.F.， Zhou G.D.， Su J. and Tan C.L. (2003). Coreference Resolution Using Competition Learning Approach， ACL’2003: 176- 183， Sapporo， Japan， 7- 12 July 2003.

[12] Yang X.F., Su J., Zhou G.D. and Tan C.L. (2004). Improving pronoun resolution by incorporating coreferential information of candidates. ACL’2004:127- 134, Barcelona, Spain, 21- 26 July 2004 .

[13] Zhou G.D. and Su J. (2004). A high- performance coreference resolution system using a multi- agent strategy， COLING’2004:522- 528. 23 - 27 Aug， 2004， Geneva， Switzerland.

[14] Yang X.F.， Su J. and Tan C.L. (2005), Improving Pronoun Resolution Using Statistics - Based Semantic Compatibility Information, ACL’2005:165 -172， Univ of Michgan-Ann Arbor， USA， 25 -30 June 2005

[15] David Westbrook, Ralph Grishman; Using semantic Relations to Refine Coreference Decisions Computational Linguistics, 2005

[16] Zhou G.D., Su J. Zhang J. and Zhang M. (2005). Exploring various knowledge in relation extraction. ACL’2005. 427-434. 25-30 June, Ann Arbor, Michgan, USA.

[17] Bergsma S. and Lin D.K.(2006). Bootstrapping path-based pronoun resolution. COLING-ACL’2006: 33-40.

[18] Yang X.F., Su J. and Tan C.L. (2006). Kernel-based pronoun resolution with structured syntactic knowledge. COLING-ACL’2006: 41-48.

[19] Yang X.F. and Su J. (2007). Coreference Resolution Using Semantic Relatedness Information from Automatically Discovered Patterns ACL’2007 528-535 Prague, Czech Republic, June 2007.

[20] Ng V. (2007). Semantic Class Induction and Coreference Resolution ACL’2007 536-543 Prague, Czech Republic, June 2007.

[21] Ng, V. (2007). Shallow semantics for coreference resolution. In IJCAI 2007, pages 1689–1694.

[22] Zhou G.D., Kong F. and Zhu Q.M. 2008a. Context-sensitive convolution tree kernel for pronoun resolution. IJCNLP’2008