Combining Syntactic and Semantic Features by SVM for Unrestricted Coreference Resolution

Huiwei Zhou[1], Yao Li[2], Degen Huang[3], Yan Zhang[4], Chunlong Wu[5], Yuansheng Yang[6]
Dalian University of Technology
Dalian, Liaoning, China
{[1]zhouhuiwei,[3]huangdg,[6]yangys}@dlut.edu.cn
[2]tianshanyao@mail.dlut.edu.cn
[4]zhangyan_zyzy@yeah.net
[5]wuchunlong@gmail.com

Abstract

The paper presents a system for the CoNLL-2011 share task of coreference resolution. The system composes of two components: one for mentions detection and another one for their coreference resolution. For mentions detection, we adopted a number of heuristic rules from syntactic parse tree perspective. For coreference resolution, we apply SVM by exploiting multiple syntactic and semantic features. The experiments on the CoNLL-2011 corpus show that our rule-based mention identification system obtains a recall of 87.69%, and the best result of the SVM-based coreference resolution system is an average F-score 50.92% of the MUC, B-CUBED and CEAFE metrics.

1 Introduction

Coreference resolution, defined as finding the different mentions in a document which refer to the same entity in reality, is an important subject in Natural Language Processing. In particular, coreference resolution is a critical component of information extraction systems (Chinchor and Nancy, 1998; Sundheim and Beth, 1995) and a series of coreference resolution tasks have been introduced and evaluated from MUC (MUC-6, 1995). Some machine learning approaches have been applied to coreference resolution (Soon et al., 2001; Ng and Cardie, 2002; Bengtson and Roth, 2008; Stoyanov et al., 2009). Soon et al.(2001) use a decision tree classifier to decide whether two mentions in a document are coreferent. Bergsma and Lin (2006) exploit an effective feature of gender and number to a pronoun resolution

system and improve the performance significantly, which is also appeared in our feature set. However, automatic coreference resolution is a hard task since it needs both syntactic and semantic knowledge and some intra-document knowledge. To improve the performance further, many deep knowledge resources like shallow syntactic and semantic knowledge are exploited for coreference resolution (Harabagiu et al., 2001; McCallum and Wellner, 2004; Denis and Baldridge, 2007; Ponzetto and Strube, 2005; Versley, 2007; Ng, 2007). In order to make use of more syntactic information, Kong et al. (2010) employ a tree kernel to anaphoricity determination for coreference resolution and show that applying proper tree structure in corefernce resolution can achieve a good performance.

The CoNLL-2011 Share Task (Pradhan et al., 2011) "Modeling Unrestricted Coreference in OntoNotes" proposes a task about unrestricted coreference resolution, which aims to recognize mentions and find coreference chains in one document. We participate in the closed test.

In this paper, we exploit multi-features to a coreference resolution system for the CONLL-2011 Share Task, including flat features and a tree structure feature. The task is divided into two steps in our system. In the first step, we adopt some heuristic rules to recognize mentions which may be in a coreference chain; in the second step, we exploit a number of features to a support vector machine (SVM) classifier to resolute unrestricted coreference. The experiments show that our system gets a reasonable result.

The rest of the paper is organized as follows. In

Section 2, we describe in detail how our system does the work of coreference resolution, including how we recognize mentions and how we mark the coreference chains. The experimental results are discussed in Section 3. Finally in Section 4, we give some conclusion.

2 The Coreference Resolution System

The task of coreference resolution is divided into two steps in our system: mentions detection and coreference resolution. In the first step, we use some heuristic rules to extract mentions which may refer to an entity. In the second step, we make up mention-pairs with the mentions extracted in the first step, and then classify the mention-pairs into two groups with an SVM model: Coreferent or NotCoreferent. Finally we get several coreference chains in a document according to the result of classification. Each coreference chain stands for one entity.

2.1 Rule-based Identification of Mentions

The first step for coreference resolution is to identify mentions from a sequence of words. We have tried the machine-learning method detecting the boundary of a mention. But the recall cannot reach a high level, which will lead to bad performance of coreference resolution. So we replace it with a rule-based method. After a comprehensive study, we find that mentions are always relating to pronouns, named entities, definite noun phrases or demonstrative noun phrases. So we adopt the following 5 heuristic rules to extract predicted mentions:

1. If a word is a pronoun, then it is a mention.

2. If a word is a possessive pronoun or a possessive, then the smallest noun phrase containing this word is a mention.

3. If a word string is a named entity, then it is a mention.

4. If a word string is a named entity, then the smallest noun phrase containing it is a mention.

5. If a word is a determiner (a, an, the, this, these, that, etc.), then all the noun phrase beginning with this word is a mention.

2.2 Coreference Resolution with Multi-Features

The second step is to mark the coreference chain using the model trained by an SVM classifier. We extract the marked mentions from the training data and take mention-pairs in one document as instances to train the SVM classifier like Soon et al.(2001) . The mentions with the same coreference id form the positive instances while those between the nearest positive mention-pair form the negative instance with the second mention of the mention-pair.

The following features are commonly used in NLP processes, which are also used in our system:

- i-NamedEntity/j-NamedEntity: the named entity the mention i/j belongs to

- i-SemanticRole/j-SemanticRole: the semantic role the mention i/j belongs to which

- i-POSChain/j-POSChain: the POS chain of the mention i/j

- i-Verb/j-Verb: the verb of the mention i/j

- i-VerbFramesetID/j-VerbFramesetID: the verb frameset ID of the mention i/j, which works together with i/j-Verb

All the 5 kinds of features above belong to a single mention. For mention-pairs, there are another 4 kinds of features as below:

- StringMatch: after cutting the articles, 1 if the two mentions can match completely, 2 if one is a substring of the other, 3 if they partly match, 4 else.

- IsAlias: after cutting the articles, 1 if one mention is the name alias or the abbreviation of the other one, 0 else

- Distance: it is the number of sentences between two mentions, 0 if the two mentions are from one sentenci-Verb/j-Verb: the verb of the mention i/j

- SpeakerAgreement: 1 if both the speakers of the two mentions are unknown, 2 if both the two mentions come from the same speaker, 3 if the mentions comes from different speakers.

All of the 14 simple and effective features above are applied in the baseline system, which use the same method with our system. But coreference resolution needs more features to make full use of the intra-documental knowledge, so we employ the following 3 kinds of features to our system to catch more information about the context.

- i-GenderNumber/j-GenderNumber (GN): 7 values: masculine, feminine, neutral, plural, ?rst-person singular, ?rst-person plural, second-person.

- SemanticRelation (SR): the semantic relation in WordNet between the head words of the two mentions: synonym, hyponym, no relation, unknown.

- MinimumTree (MT): a parse tree represents the syntactic structure of a sentence, but coreference resolution needs the overall context in a document. So we add a super root to the forest of all the parse trees in one document, and then we get a super parse tree. The minimum tree (MT) of a mention-pair in a super parse tree is the minimum sub-tree from the common parent mention to the two mentions, just like the method uesd by Zhou(2009). And the similarity of two trees is calculated using a convolution tree kernel (Collins and Duffy, 2001), which counts the number of common sub-trees.

We try all the features in our system, and get some interesting results which is given in Experiments and Results Section.

3 Experiments and Results

Our experiments are all carried out on CONLL-2011 share task data set (Pradhan et al., 2007).

The result of mention identification in the first step is evaluated through mention recall. And the performance of coreference resolution in the second step is measured using the average F1-measures of MUC, B-CUBED and CEAFE metrics (Recasens et al., 2010). All the evaluations are implemented using the scorer downloaded from the CONLL-2011 share task website [1] .

3.1 Rule-based Identification of Mentions

The mention recall of our system in the mention identification step reaches 87.69%, which can result in a good performance of the coreference resolution step. We also do comparative experiments to investigate the effect of our rule-based mention identification. The result is shown in Table 1. The CRF-based method in Table 1 is to train a conditional random field (CRF) model with 6 basic features, including Word, Pos, Word_ID, Syntactic parse label, Named entity, Semantic role.

Method	Recall	Precision	F-score
Rule-based	87.69	32.16	47.06
CRF-based	59.66	50.06	54.44

Table 1: comparative experiments of CRF-based and rule-based methods of mention identification(%)

Table 1 only shows one kind of basic machine-learning methods performs not so well as our rule-based method in recall measure in mention identification, but the F1-measure of the CRF-based method is higher than that of the rule-based method. In our system, the mention identification step should provide as many anaphoricities as possible to the coreference resolution step to avoid losing coreferent mentions, which means that the higher the recall of mention identification is, the better the system performs.

3.2 Coreference Resolution with Multi-Features

In the second step of our system, SVM-LIGHT-TK1.2 implementation is employed to coreference resolution. We apply the polynomial kernel for the flat features and the convolution tree kernel for the minimum tree feature to the SVM classifier, in which the parameter d of the polynomial kernel is set to 3 (polynomial $(a * b + c)^d$) and the combining parameter r is set to 0.2 ($K = tree - forest - kernel * r + vector - kernel$). All the other parameters are set to the default value. All the experiments are done on the broadcast conversations part of CoNLL-2011 corpus as the calculating time of SVM-LIGHT-TK1.2 is so long.

Experimental result using the baseline method with the GenderNumber feature added is shown in

d=?	MUC	B^3	CEAFE	AVE
2	47.49	61.14	36.15	48.26
3	51.37	62.82	38.26	50.82

Table 2: parameter d in polynomial kernel in coreference resolution using the baseline method with the GN feature(%)

Talbe 2. The result shows that the parameter d in polynomial kernel plays an important role in our coreference resolution system. The score when d is 3 is 2.56% higher than when d is 2, but the running time becomes longer, too.

r=?	MUC	B^3	CEAFE	AVE
1	31.41	45.08	22.72	33.07
0.25	34.15	46.87	23.63	34.88
0	51.37	62.82	38.26	50.82

Table 3: combining parameter r ($K = tree - forest - kernel * r + vector - kernel$) in coreference resolution using the baseline with the GN and MT features(%)

In Table 3, we can find that the lower the combining parameter r is, the better the system performs, which indicates that the MT feature plays a negative role in our system. There are 2 possible reasons for that: the MT structure is not proper for our coreference resolution system, or the simple method of adding a super root to the parse forest of a document is not effective.

Method	MUC	B^3	CEAFE	AVE
baseline	42.19	58.12	33.6	44.64
+GN	51.37	62.82	38.26	50.82
+GN+SR	49.61	64.18	38.13	50.64
+GN +SEMCLASS	50.97	62.53	37.96	50.49

Table 4: effect of GN and SR features in coreference resolution using no MT feature (%)

Table 4 shows the effect of GenderNumber feature and SemanticRelation feature, and the last item is the method using the SemanticClassAgreementFeature (SEMCLASS) used by (Soon et al., 2001) instead of the SR feature of our system. The GN feature significantly improves the performance of our system by 6.18% of the average score, which may

be greater if we break up the gender and number feature into two features. As the time limits, we haven't separated them until the deadline of the paper. The effect of the SR feature is not as good as we think. The score is lower than the method without SR feature, but is higher than the method using SEMCLASS feature. The decreasing caused by S-R feature may be due to that the searching depth in WordNet is limited to one to shorten running time.

To investigate the performance of the second step, we do an experiment for the SVM-based coreference resolution using just all the anaphoricities as the mention collection input. The result is shown in Table 5. As the mention collection includes no incorrect anaphoricity, any mistake in coreference resolution step has double effect, which may lead to a relatively lower result than we expect.

MUC	B^3	CEAFE	AVE
65.55	58.77	39.96	54.76

Table 5: using just all the anaphoricities as the mention collection input in coreference resolution step (%)

In the three additional features, only the GN feature significantly improves the performance of the coreference resolution system, the result we finally submitted is to use the baseline method with GN feature added. The official result is shown in Table 6. The average score achieves 50.92%.

MUC	B^3	CEAFE	AVE
48.96	64.07	39.74	50.92

Table 6: official result in CoNLL-2011 Share Task using baseline method with GN feature added (%)

4 Conclusion

This paper proposes a system using multi-features for the CONLL-2011 share task. Some syntactic and semantic information is used in our SVM-based system. The best result (also the official result) achieves an average score of 50.92%. As the MT and S-R features play negative roles in the system, future work will focus on finding a proper tree structure for the intra-documental coreference resolution and combining the parse forest of a document into a tree to make good use of the convolution tree kernel.

References

A. McCallum and B. Wellner. 2004. Conditional models of identity uncertainty with application to noun coreference. *In Advances in Neural Information Processing Systems (NIPS)*, 2004.

Chinchor, Nancy A. 1998. Overview of MUC-7/MET-2. *In Proceedings of the Seventh Message Understanding Conference (MUC-7).*

Eric Bengtson, Dan Roth. 2008. Understanding the Value of Features for Coreference Resolution *Proceedings of the 2008 Conferenceon Empirical Methods in Natural Language Processing*, pages294C303.

Fang Kong, Guodong Zhou, Longhua Qian, Qiaoming Zhu. 2010. Dependency-driven Anaphoricity Determination for Coreference Resolution *Proceedings of the 23rd International Conferenceon Computational Linguistics (Coling2010)*, pages599C607.

Guodong Zhou, Fang Kong. 2009. Global Learning of Noun Phrase Anaphoricity in Coreference Resolution via Label Propagation. *In Proceedings of the 2009 Coreference on Empirical Methods in Natural Language Processing*, pages 978-986, 2009.

M. Collins, N.Duffy. 2001. Convolution Kernels for Natural Language Resolution *NIPS' 2001*.

Marta Recasens, Lluŀs Mrquez, Emili Sapena, M. Antnia Martł, Mariona Taul, Vronique Hoste, Massimo Poesio, Yannick Versley 2010. SemEval-2010 Task 1: Coreference Resolutionin Multiple Languages *In Proceeding SemEval 2010 Proceedings of the 5th International Workshop on Semantic Evaluation*, 2010.

MUC-6. 1995. Coreference task definition (v2.3, 8 Sep 95) *In Proceedings of the Sixth Message Understanding Conference (MUC-6)*, pages 335-344.

P.Denis, J.Baldridge. 2007. Joint determination of anaphoricity and coreference resolution using integer programming. *In Proceedings of HLT/NAACL*, 2007.

V. Ng and C. Cardie. 2002. Improving machine learning approaches to coreference resolution. *In Proceedings of ACL*, 2002.

V. Ng. 2007. Shallow semantics for coreference resolution. *In Proceedings of IJCAI*, 2007.

Veselin Stoyanov, Nathan Gilbert, Claire Cardie, Ellen Riloff. 2009. Conundrums in Noun Phrase Coreference Resolution: Making Sense of the State-of-the-Art *Proceeding ACL '09 Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL.*

W.Soon,H.Ng,and D.Lim. 2001. A machine learning approach to coreference resolution of noun phrase. *Computational Linguistics*, 27(4):521-544,2001.

S. M.Harabagiu,R.C.Bunescu,and S.J. Maiorano. 2001. Text and knowledge mining for coreference resolution. *In Proceedings of NAACL*, 2001.

S.Ponzetto, M.Strube. 2005. Semantic role labeling for coreference resolution. *In Proceedings of EACL*, Italy, April 2005.

Sameer Pradhan, Lance Ramshaw, Mitchell Marcus, Martha Palmer, Ralph Weischedel, Nianwen Xue. 2011. CoNLL-2011 Shared Task: Modeling Unrestricted Coreference in OntoNotes. *Proceedings of the Fifteenth Conference on Computational Natural Language Learning (CoNLL 2011).*

Sameer S. Pradhan, Lance Ramshaw, Ralph Weischedel, Jessica MacBride, Linnea Micciulla. 2007. Unrestricted Coreference: Identifying Entities and Events in OntoNotes. *In International Conference on Semantic Computing*, 2007.

Shane Bergsma, Dekang Lin. 2006. Bootstrapping Path-Based Pronoun Resolution. *In Proceedings of the 21st International Conference on Computational Linguistics*, 2006.

Sundheim, Beth M. 1995. Overview of results of the MUC-6 evaluation. *In Proceedings of the Sixth Message Understanding Conference (MUC-6)*, pages 13-31.

Y.Versley. 2007. Antecedent selection techniques for high-recall coreference resolution. *In Proceedings of EMNLP/CoNLL*, 2007.