A Ranking Approach to Persian Pronoun Resolution

Nafiseh Sadat Moosavi and Gholamreza Ghassem-Sani

Department of Computer Engineering, Sharif University of Technology, Iran
n_moosavi@ce.sharif.edu, sani@sharif.edu

Abstract. Coreference resolution is an essential step toward understanding discourses, and it is needed by many NLP tasks such as machine translation, question answering, and summarization. Pronoun resolution is a major and challenging subpart of coreference resolution, in which only the resolution of pronouns is considered. Classification approaches have been widely used for coreference/pronoun resolution, but it has been shown that ranking approaches outperform classification approaches in a variety of fields such as English pronoun resolution (Denis and Baldridge, 2007), question answering (Ravichandran, 2003), and tagging/parsing (Collins and Duffy, 2002; Charniak and Johnson, 2005). The strength of ranking is in its ability to consider all candidates at once and selecting the best one based on the model, while existing classification methods consider at most two candidate responses at a time. Persian and its varieties are spoken by more than 71 million people, and it has some characteristic that make parsing and other related processing of Persian more difficult than those of English. In this paper, we have evaluated maximum entropy ranker on Persian pronoun resolution and compared the results with that of four base classifiers.

Keywords: Natural Language Processing, Machine Learning, Ranking, Classification, Pronoun Resolution, Persian.

1 Introduction

The final goal of natural language processing (NLP) is that computers understand human languages. Different NLP research areas such as part of speech (POS) tagging, word sense disambiguation (WSD), and grammatical parsing concentrate only on a partial solution of this ultimate goal. All of these are required for a computer to understand a natural language.

NLP tasks can be divided into micro-tasks and macro-tasks. Micro-tasks focus on a word level processing or a sentence level processing such as WSD and parsing. On the other hand, macro-tasks include tasks which do a document level processing such as information retrieval and document classification. Before the introduction of machine learning approaches in NLP, higher level tasks such as semantic processing needed a variety of lower level tasks such as POS tagging and parsing. However, the use of machine learning methods may make it possible to obtain enough statistical in-

© A. Gelbukh (Ed.) *Received 26/11/08*
Advances in Computational Linguistics. *Accepted 15/12/08*
Research in Computing Science 41, 2009, pp. 169-180 *Final version 04/02/09*

formation to make these lower level tasks unnecessary. Although learning methods result in a satisfactory performance in many tasks, they are limited and hardly can provide complete necessary information.

There is a missing link between sentence level processing and document level understanding; and an important example of such a missing link is the resolution of pronouns.

Pronoun resolution is a crucial and difficult subpart of an overall task named coreference resolution. Coreference resolution determines the group of noun phrases that refer to the same real world entity.

In recent years, the ranking (re-ranking) approaches have been successfully applied to a variety of NLP tasks including English pronoun resolution (Denis and Baldridge, 2007), and the achieved results show that the ranker outperforms simple classification methods. In this paper, we presented the evaluation of a ranking model applied to the Persian pronoun resolution.

2 Related Work

Since there is no previous work on Persian coreference/pronoun resolution, we briefly review some of the most recent related works applied to English.

2.1 Classification

The usage of machine learning methods in pronoun resolution began with a simple naïve Bayes approach (Ge et al., 1998), in which the random variable was a candidate reference for a given pronoun. Soon et al. (2001) and Ng and Cardie (2002) used decision tree for coreference decisions; classification is done by pairing each candidate noun phrase and deciding whether they are coreferent or not.

Yang et al. (2003) proposed the use of competition learning in coreference problem. In their proposed method, every training sample is made by one anaphor and a pair of candidate antecedent, one positive antecedent and a negative one. In this way, the classifier has more ability to compare different candidates.

2.2 Clustering

Cardie and Wagstaff (1999) represented every noun phrase with a feature vector and then used a clustering algorithm for partitioning these feature vectors. Every resulting partition represents an entity. Avoiding triangle inconsistency is an advantage of clustering over classification. Classification makes decision about the pairs (“Mr. Green”, “Green”) and (“Green”, “she”) independently, so “she” in the second pair may be recognized as being coreferent with “Green”, while its gender is incompatible with “Mr. Green”, which is coreferent with “Green”. However, in clustering these decisions are made dependently. Before adding a noun phrase to a cluster, its consistency with all existing noun phrases of that cluster is checked.

Wagstaff (2002) offered constrained clustering for coreference resolution. The proposed algorithm accepts the constraints in the forms of “must-link” and “cannot-link”. “must-links” represent noun phrases that should be placed in the same cluster, and “cannot-links” indicate noun phrases that cannot be assigned to the same cluster.

2.3 Bell Tree

Luo et al. (2004) casted the coreference problem as a searching problem and represented the search space as a Bell tree. The root of the search tree is a single noun phrase, the second noun phrase is added to the next level of the tree, and the leaves contain the possible partitioning of all noun phrases

The goal is to find the most probable path from the root to the leaves; the leaf in the most probable path contains the resultant clustering.

2.4 Graph Partitioning

Nicolae and Nicolae (2006) introduced graph partitioning in coreference problem. Graph partitioning can be considered as a clustering algorithm in which the clustering is done by a graph cutting algorithm. Each node of the graph corresponds to a noun phrase. The weight of each edge shows how likely the two connected nodes are coreferent. These weights can be determined by any classification algorithm.

2.5 Co-training

Ng and Cardie (2003) modified the multi-view co-training algorithm (Blum and Mitchell, 1998) to fit the coreference resolution. Rather than separating the feature space into two compatible and uncorrelated views, they used two different learners in a co-training algorithm.

2.6 Conditional Random Fields

McCallum and Wellner (2004) offered three models for the use of Conditional Random Fields (CRF) in coreference resolution. All of the proposed models are conditionally-trained, undirected graphical models which, unlike previous models, are relational. In a relational model, the dependency between the training data and the input features is considered.

2.7 Data Mining

In the method proposed by Harabagiu et al. (2001), the resolution rules from a large corpus are mined and the entropy of each rule is evaluated accordingly. The partitioning of noun phrases is done in a manner in which more rules with a higher confidence accept the specified partitioning.

Bean and Riloff (2004) presented a system which mines the relations between the words and their contexts. In this way, for each noun phrase a kind of semantic rule is achieved, and it helps the improvement of coreference resolution.

Bergsma and Lin (2006) offered a method in which the likelihood of the coreference between a pronoun and its candidate antecedent is learned based on the dependency parse path between the pronoun and its candidate antecedents.

2.8 First-Order Probabilistic Model

Culotta et al. (2007) offered a particular method for doing training and inference in first-order models of coreference, in which the features are over a set of noun phrases rather than a pair of noun phrases. Their method results in a first-order probabilistic model for coreference resolution. First-order probabilistic logic is a first-order logic that associates a real-valued parameter to every predicate.

2.9 Ranking

Denis and Baldridge (2007) proposed a supervised ranking approach for pronoun resolution. The ranking enables all candidate antecedents to be evaluated together; whereas classification methods examine at most two candidate antecedents at a time. They showed that their proposed method outperforms the best classification method.

3 Persian Pronouns

Persian and its varieties are spoken by more than 71 million people in Iran, Afghanistan, and Tajikistan. Normal Persian sentences are structured as " (optional subject) (optional prepositional phrase) (optional object) verb". However, it can also have a free word order in many places, and this characteristic make parsing and other related processing of Persian more difficult than those of English.

Persian is a null-subject language, and nominal pronouns can be omitted from a sentence. It has 18 different pronouns: four first person pronouns, four second person pronouns, nine third person pronouns, and one pronoun which doesn’t have number and can be used in place of every noun phrase.

Persian pronouns have special characteristics that make their resolution more difficult than that of English pronouns. Persian’s third person pronouns do not have any gender information. Besides, some plural pronouns are sometimes used in place of singular human pronouns to show respect, and singular pronouns may also refer to plural non-human noun phrases. These characteristics cause that the gender and number agreement, which are two of the most effective features in pronoun resolution, become either useless or less effective in Persian pronoun resolution.

4 Ranking Approaches

Using ranking (re-ranking) approaches for solving complex natural language processing problems has increasingly received attention in recent years. The main motivation for using the ranking approaches is their ability to directly compare between different responses and pick the most proper response.

Ranking has been successfully applied in a variety of NLP tasks including machine translation (Och, 2003; Shen et al, 2004), question answering (Ravichandran et al, 2003), parsing (Collins, 2000; Charniak and Johnson, 2005), and English pronoun resolution (Denis and Baldridge, 2007).

As it has been mentioned in section 3, Persian pronoun resolution has some characteristics that make it more difficult than that of English. Thus, it can be considered as a new NLP domain for evaluating the strength of a ranking method.

4.1 Maximum Entropy Ranker

The problem of pronoun resolution can be modeled with Maximum Entropy (MaxEnt) in two different ways: classification and ranking. A MaxEnt classifier allocates each pair (consist of a pronoun and a candidate antecedent) to one of "coreferent" or "non-coreferent" classes, while a MaxEnt ranker selects a single candidate as the antecedent of a pronoun. This means that a MaxEnt classifier can select many candidates as antecedents of a single pronoun, as long as the pairs including those antecedents and the pronoun are marked as "coreferent". In contrast, a MaxEnt ranker always selects the most probable candidate antecedent as a pronoun's antecedent.

Suppose we have a set of candidate antecedents $A = \{a_1, a_2, ..., a_n\}$ for a pronoun p, and $f_k(a, p), k = 1,...,K$ are K different feature functions which calculate the features based on the pair (a, p). The maximum entropy classifier can be modeled as equation 1,

$$p(c \mid a, p) = \frac{\exp[\sum_{k=1}^{K} \lambda_{k,c} f_k(a, p)]}{\sum_{c'} \exp[\sum_{k=1}^{K} \lambda_{k,c'} f_k(a, p)]} \quad \textbf{(1)}$$

where, $\lambda_{k,c}$ $k = 1,...,K$ and $c = \{coreferent, non-coreferent\}$ are the model parameters.

Maximum entropy ranker is modeled as equation 2,

$$p(a \mid p) = \frac{\exp[\sum_{k=1}^{K} \lambda_k f_k(a,q)]}{\sum_{a'} \exp[\sum_{k=1}^{K} \lambda_k f_k(a',q)]} \tag{2}$$

where, λ_k $k = 1,...,K$ are the model parameters.

In the classifier model, for each class (coreferent and non-coreferent), the weights of each feature functions are computed separately, whereas in the ranker model, the weights do not depend on class labels and there is the same set of feature weights for all classes. Thus, the ranker model allows all candidate antecedents to be compared together.

5 A Persian Pronoun Resolution System

This section describes a Persian pronoun resolution system which casts the pronoun resolution as a classification and ranking task.

5.1 Data Set

In order to use a learning method for coreference resolution, a suitable coreferentially annotated corpus is needed. The development and evaluation of automatically trained coreference systems is dependent on the existence of such corpora.

In this section, we briefly describe a Persian coreferentially annotated corpus, PCAC-2008, which was developed by appropriate annotation of another Persian corpus named Bijankhan (Bijankhan, 2004).

5.1.1 Bijankhan Corpus

Bijankhan is a corpus containing a huge number of Persian syntactic-semantic annotated documents. It contains wide variety of linguistic data in different subjects. It can be considered as a complete statistical universe of Persian documents. Bijankhan includes syntactic and semantic tagging. It is organized and tagged as a word-level corpus.

Bijankhan contains 3050 different documents and is based on daily news and common texts. One example of the supplied information in Bijankhan is shown in Fig. 1.

Each line contains a word and some syntactic and semantic features such as part of speech (POS), number information of that word. The previous applications of Bijankhan includes morphological analysis (Feyzbakhsh et al., 2008) and unsupervised grammar induction (Mirroshandel et al. 2007; Mirroshandel and Ghassem-Sani, 2008).

```
W × N N.SING.COM.GEN شناخت
W × N N.PL.COM.GEN علایق
W × N N.PL.COM کودکان
W × P P در
W × N N.SING.COM.GEN اشتیاق
W × PRO PRO.DEMO.PL آنها
W × P P به
W × N N.SING.COM کتاب
W × P P از
W × N N.PL.COM.GEN ضرورتهای
W × ADJ ADJ.SIM اجتناب ناپذیر
W × V V.PRE.SIM است
W × DELM DELM .
```

Fig. 1. Annotation of Bijankhan for a sample sentence: Each word of a sentence was tagged in a separate line which contains some basic syntactic-semantic knowledge of that word. Each line begin with "W *" marker and then followed by a POS of each word repeated two times. N, P, PRO, ADJ, V, DELM are the abbreviation of noun, proposition, pronoun, adjective, verb and delimiter respectively. The last word of each line is the tagged word itself.

5.1.2 PCAC-2008 Corpus

PCAC is the abbreviation of Persian Coreferentially Annotated Corpus. It is a partial extension of Bijankhan corpus, in which the coreference information has been added. Different Bijankhan articles in different topics and lengths were annotated in PCAC.

PCAC consists of 2006 labeled pronouns drawn from 30 different documents, and each pronoun is marked with its nearest antecedent. Fig. 2 shows how the annotations of Bijankhan have been extended in PCAC-2008.

```
W × N N.SING.COM.GEN شناخت
W × N N.PL.COM.GEN علایق
W × N N.PL.COM.SET9 کودکان
W × P P در
W × N N.SING.COM.GEN اشتیاق
W × PRO PRO.DEMO.PL.SET9 آنها
W × P P به
W × N N.SING.COM کتاب
W × P P از
W × N N.PL.COM.GEN ضرورتهای
W × ADJ ADJ.SIM اجتناب ناپذیر
W × V V.PRE.SIM است
W × DELM DELM .
```

Fig. 2. Annotation of PCAC-2008 for the annotated sentence of Fig. 1: : Coreference information has been added to the Bijankhan annotations by the use of "SET" feature.

5.2 Training Samples

Positive training samples were made by pairing each pronoun and its nearest antecedent, and negative samples were made by pairing pronouns and their negative antecedents. Every noun phrase between each pronoun and its nearest antecedent is considered as a negative antecedent of that pronoun.

5.3 Feature Set

We used the Denis and Baldridge (2007) feature set for evaluating the Persian pronoun resolution system. In addition to the features explained in (Denis and Baldridge, 2007), a genitive feature was also added which determines whether a pronoun is genitive or not. This feature is effective in Persian pronoun resolution and Bijankhan annotation contains this information. Besides, we have not used the gender agreement feature due to the lack of gender information in Persian pronouns.

The used feature set is composed of 25 features that can be divided into three groups: 1) features describing the pronoun, 2) features describing a candidate antecedent, and 3) features describing the relationship between the pronoun and candidate antecedents.

5.4 Learning Methods

5.4.1 Classification Algorithms

Theoretical studies in machine learning such as what was done by Wolpert and Macready (1995), have demonstrated that none of the inductive algorithms is generally superior to others. In order to see which learning algorithm is the most proper for a specific language processing task, it is necessary to compare different machine learning methods experimentally on that particular task. If the bias of a learning algorithm better fits to the properties of a specific task, the resulting model would be more suitable for the new data of the same task.

Daume (2006) mentioned that the most popular choices of the classifier for the coreference resolution task in the literatures are decision trees and MaxEnt models. However, it doesn't mean that these learning methods are also the most appropriate ones for a coreference resolution task.

We used these two classifiers plus Perceptron and SVM classifiers, which are regarded as two of the most effective methods for the binary classification problems.

5.4.2 The Ranking Algorithm

Maximum Entropy modeling has been extremely successful for many ranking tasks (Denis and Baldridge, 2007; Charniak and Johnson, 2005; Elwell, 2008; Ji and Grishman, 2005; Kim and Hovy, 2005; Nguyen and Kim, 2008; Wellner and Pustejovsky, 2007; etc). Thus, we used it for evaluating the effect of ranking in Persian pronoun resolution.

5.5 Testing the Trained System

After training the ranker and classification algorithms, the trained learners are used to guide the resolution of pronouns. First, the learning samples are made by finding candidate antecedent for each pronoun and pairing them. After preparing the learning samples, and in the case of classifiers, each pair is examined by the classifier and is classified as a "coreferent" or "non-coreferent". The candidate antecedent of each coreferent pair is considered as an antecedent of that pair's pronoun. Thus, each pronoun may have several antecedents.

In the case of ranker, all pairs made for the same pronoun are considered at the same time, and the candidate antecedent of the most probable pair is selected as the pronoun's antecedent.

6 Evaluations

The evaluation of the ranker model in contrast with the classification methods is presented in this section.

The same preprocessing modules and feature set were used for evaluation of classification methods and the ranker model; we performed a 10-fold cross validation to evaluate each of the examined methods. The performance is reported in terms of precision, recall and F_1-measure of the referential pronouns.

Regarding setting the learner-specific parameters, we used the default values for all examined learners unless otherwise stated. In the case of MaxEnt, we used 100 iterations of the improved iterative scaling algorithm using Gaussian prior.

The SVM learner was evaluated by RBF, sigmoid, and polynomial kernels and with different degrees for polynomial kernel. The SVM result reported in Table 1 is the best achieved result (i.e., the polynomial kernel of degree 3).

6.1 Results and Discussion

The results of classification methods in comparison with that of the ranker model are presented in table 1. The results show that the MaxEnt ranker significantly improves the MaxEnt classifier; however, its results are not better than that of the C4.5 and Perceptron base classifiers (while MaxEnt ranker performs better than the best classification method for English pronoun resolution (Denis and Baldridge, 2007)). Thus, the achieved results confirm the Wolpert and Macready (1995) studies and show that the decision tree learner has a better performance than the other examined methods in the Persian pronoun resolution.

Table 1. Results of C4.5, Perceptron, SVM and MaxEnt classifiers in comparison with MaxEnt ranker

Learning method	Recall	Precision	F_1
C4.5	31.70	75.99	44.73
Perceptron	27.47	49.64	35.36
SVM	17.00	79.12	27.98
MaxEnt classifier	4.01	19.92	6.68
MaxEnt ranker	30.34	30.34	30.34

6.2 Error Analysis

Denis and Baldridge MaxEnt ranker (2007) achieved the f_1-measure of 74.0%, while the achieved result for Persian pronoun resolution is 30.34%. Our error analysis reveals that the poor performance of the two examined system (classification and ranker systems) can mainly be attributed to the following reasons:

1. Our non-statistical parser and the Persian free word order grammar that result in a highly unbalanced training data in which positive samples are only 2.8% of the whole data. The parser is used for determination of noun phrases; a non-statistical parser with a free word order grammar finds more noun phrases between a pronoun and its nearest antecedent, and thus results in a highly unbalanced data set.
2. The lack of gender and number agreement features that was addressed in section 3.

7 Conclusions

We have evaluated the maximum entropy ranker and four base classifiers for Persian pronoun resolution in order to evaluate the effect of ranking model in this domain.

The results show that the MaxEnt ranker significantly outperforms the MaxEnt classifier: improving the F_1 measure from 6.68% to 30.34%. However, it does not outperform all the examined classifiers; C4.5 and Perceptron classifiers are more suitable and had better performance in Persian pronoun resolution.

One can suggest several possible types of futures works to improve the performance of the presented system. We used the MaxEnt ranker because it is the most common ranking technique in NLP literatures; however, the use of the probability estimation trees (PETs) is an important avenue to explore further. PETs are trees which estimate the probability of class membership and can be used as a ranker (Breiman et al, 1984, Provost and Domingos, 2000, Margineantu and Dietterich, 2001). As it was shown, the decision tree classifier achieved the best results for the Persian pronoun resolution. On the other hand, the results show that the maximum entropy ranker significantly outperforms the maximum entropy classifier. Thus, the use of PETs for Persian pronoun resolution can be considered as an important extension of this study, which may improve the results considerably.

The use of syntactic parse tree as a structured feature is another important area for future works. Yang et al. (2006) presented a method in which a syntactic parse tree was used as a structured feature, and then proper kernels were applied to such a feature, together with other ordinary feature. Their results showed that the system including the structured feature could increase the success rate significantly. We weren't able to use such a structured feature due to the lack of a non-commercial probabilistic parser. For example, our non-statistical parser finds 2050 different parse trees (and 16 different noun phrases) for a simple sentence with POSs like "N N N N P N N N V". Thus, the use of convolution tree kernels seems impractical with these huge number of parse trees for each sentence.

References

1. Bean, D., Riloff, E.: Unsupervised learning of contextual role knowledge for coreference resolution. In: HLT-NAACL, pp. 297--304 (2004).
2. Bergsma, Sh., Lin, D.: Bootstrapping Path-Based Pronoun Resolution. In: COLING/ACL-06, pp. 33--40 (2006).
3. Bijankhan, M.: The rule of linguistic corpora in writing a grammar: an introduction to a software. Iranian Journal of Linguistics, vol 19 (2004).
4. Blum, A., Mitchell, T.: Combining Labeled and Unlabeled Data with Co-training. In: COLT, pp 92--100 (1998).
5. Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J.: Classification and regression trees. Wadsworth international group (1984).
6. Cardie, C., Wagstaff, K.: Noun Phrase coreference as clustering. In: EMNLP/VLC, pp. 82--89 (1999).
7. Charniak, E., Johnson, M.: Coarse-to-fine n-best parsing and maxent discriminative reranking. In: ACL (2005).
8. Collins, M.: Discriminative reranking for natural language parsing. In: ICML-2000 (2000).
9. Culotta, A., Wick, M., Hall, R., McCallum, A.: First-Order Probabilistic Models for Coreference Resolution. In: NAACL HLT 2007, pp. 81--88 (2007).
10. Daume, H.: Practical Structured Learning Techniques for Natural Language Processing. Ph.D. dissertation, Department of computer science, University of Southern California (2006).
11. Denis, P., Baldridge, J.: A Ranking Approach to Pronoun Resolution. In: IJCAI, pp. 1588--1593 (2007).

12.Elwell, R. B.: Robust Methods for Automated Discourse Connective Argument Head Identification. Master thesis, University of Texas at Austin (2008).
13.Feyzbakhsh, M., Sadraei, M., Ghassem-Sani, Gh. R.: Unsupervised Morphology of Persian Words. In: CSICC’2008 (2008).
14.Ge, N., Hale, J., Charniak, E.: A statistical approach to anaphora resolution. Sixth Workshop on Very Large Corpora (1998).
15.Harabagiu, S., Bunescu, R., Maiorano, S.: Text and knowledge mining for coreference resolution. In: NAACL, pp. 55--62 (2001).
16.Kim, S. M., Hovy, E.: Identifying Opinion Holders for Question Answering in Opinion Texts. In: AAAI (2005).
17.Luo, X., Ittycheriah, A., Jing, H., Kambhatla, N., Roukos, S.: A mention-synchronous coreference resolution algorithm based on the bell tree. In: ACL, pp. 136--143 (2004).
18.Margineantu, D. D., Dietterich, T. G.: Improved class probability estimation from decision tree models. In: Nonlinear Estimation and Classification (2001).
19.McCallum, A., Wellner, B.: Conditional models of identity uncertainty with application to proper noun coreference. In: NIPS (2004).
20.Mirroshandel, A., Ghassem-Sani, Gh.: Using of the Constituent Context Model to Induce a Grammar for a Free Word Order Language: Persian. In: LTC2007, pp. 443--447 (2007).
21.Mirroshandel, A. Ghassem-Sani, Gh.: Unsupervised Grammar Induction Using a Parent Based Constituent Context Model. In: ECAI (2008).
22.Ng, V., Cardie, C.: Improving machine learning approaches to coreference resolution. In: *ACL*, pp. 104--111 (2002b).
23.Ng, V., Cardie, C.: Bootstrapping Coreference Classifiers with Multiple Machine Learning Algorithms. In: EMNLP, pp. 113--120 (2003).
24.Nguyen, N. L.T., Kim, J. D.: Exploring Domain Differences for the Design of Pronoun Resolution Systems for Biomedical Text. In: Coling 2008, pp. 625--632 (2008).
25.Nicolae, C., Nicolae, G.: BESTCUT: A graph algorithm for coreference resolution. In: EMNLP, pp. 275--283 (2006).
26.Och, F. J.: Minimum error rate training for statistical machine translation. In: ACL (2003).
27.Provost, F., Domingos, P.: Well-trained PETs: Improving probability estimation trees. CeDER working paper (2000).
28.Ravichandran, D., Hovy, E., Och, F. J.: Statistical qa - classifier vs re-ranker: What's the difference? In: ACL Workshop on Multilingual Summarization and Question Answering-Machine Learning and Beyond (2003).
29.Shen, L., Sarkar, A., Och, F. J.: Discriminative reranking for machine translation. In: NAACL/HLT (2004).
30.Soon, W. M., Ng, H. T., Lim, D. Ch.: A machine learning approach to coreference resolution of noun phrases. In: Computational Linguistics 27(4), pp. 521--544 (2001).
31.Wagstaff, K.: Intelligent Clustering with Instance-Level Constraints. Ph.D. dissertation, Department of computer science and engineering, Cornell University (2002).
32.Wellner, B., Pustejovsky, J.: Automatically identifying the arguments of discourse connectives. In: EMNLP-CoNLL, pp. 92--101 (2007).
33.Wolpert, D., Macready: No Free Lunch Theorems for Search. Technical report *SFI-TR-95-02-010*, Santa Fe Institute (1995).
34.Yang, X., Zhou, G., Su, J., Tan, Ch. L.: Coreference resolution using competitive learning approach. In: ACL, pp. 176--183 (2003).
35.Yang, X., Su, J., Tan, Ch. L.: Kernel-Based Pronoun Resolution with Structured Syntactic Knowledge. In: COLING/ACL-06, pp. 41--48 (2006).