Coreference Resolution using Markov Logic Network

Shujian HUANG[1], Yabing ZHANG[1], Junsheng ZHOU[12] and Jiajun CHEN[1]

[1] State Key Laboratory for Novel Software Technology,
Nanjing University, Nanjing 210093, China
[2] Department of Computer Science, Nanjing Normal University,
Nanjing, Jiangsu, 210097, China
{huangsj, zhangyb, zhoujs, chenjj}@nlp.nju.edu.cn

Abstract. Most previous work treats the solution for pronouns and noun phrases either in two separate processes or in a single process. We argue that resolving them in two processes may result in the loss of potential useful information for each process. However, resolving them in a single process is also problematic. These two types of mentions have very different characteristics in some commonly used features. Current models cannot catch those differences and thus the two types may interfere with each other. In this paper, we propose a modeling strategy using Markov logic networks (MLNs) which can explicitly discriminate the two types in one single process. Experiments on ACE2005 Chinese dataset show that our modeling using MLNs, together with the correlation clustering technique, brings significant improvements to the task.

Key words: Coreference Resolution, Markov Logic Networks

1 Introduction

Coreference resolution (CR) has drawn a lot of attentions over the past decade, especially since McCarthy[1], Cardie and Wagstaff[2] introduced machine learning techniques into this field. It plays an important role in understanding complex texts and is widely used in a lot of applications such as question answering[3], summarization[4], etc. The strong relation with other popular topics such as entity resolution in database and citation analysis[5] makes it more attractive.

Pronoun and noun phrase are two major types of mentions in CR. There are two strategies for the resolution of them. One strategy tends to split the resolution of pronouns and noun phrases into two separate processes (*Separate Strategy*). Some works focus on just pronoun resolution, aiming to find the right antecedent for each pronoun[6–9]. Denis and Baldridge subdivide mentions into five categories such as third person pronouns, speech pronouns, etc. Then, specialized models are proposed for each individual type[10]. However, we argue that just considering pairwise relation between pronoun and each of its antecedent candidates does not make full use of the information among those candidate

Received 10/11/08
Accepted 15/12/08
Final version 04/02/09

phrases. On the other hand, performing noun phrase resolution without considering pronouns may also lead to the loss of potential useful information.

In a more popular branch of researches, these two types of mentions are treated almost synchronously in a single process and only differs in some indicative features (*Uniform Strategy*)[1, 2, 11–15]. In this way, the interaction between noun and pronoun phrases can be captured by building up links between them. Recent works of Yang et al.[12] and Culotta et al.[16] proposed to solve this problem in a set-wise mode, which could capture more complex dependency relations.

However, some characteristic differences between the two types may bring conflicts to this kind of single process solution. We take two examples to informally explore these conflicts. String similarity is an important feature when judging the coreferential relation between noun phrases. Two noun phrases tend to refer to the same entity if their strings are similar to each other. For example, if the phrases "George W. Bush" and "President Bush" occur in the same text (as shown in Figure 1), they are very likely to refer to the same person. On the other hand, pronouns are not so sensitive to string similarity. Even two pronouns are identical, they can refer to different entities as well.

> *George W. Bush* is the 43rd President of the United States. ... Prior to his Presidency, *President Bush* served for 6 years as the 46th Governor of the State of Texas, where *he* earned a reputation for bipartisanship ...

Fig. 1. An example of coreference resolution. (Three phrases in italic refer to the same person.)

Similar conflict can be found in distance features. As we know, pronouns seldom refer to entities far away from them. Thus, long distance may have a strong negative impact on pronoun anaphora resolution. However, noun phrases have a much free characteristic of distance. Two noun phrases that are far apart, for example, occurring at the beginning and the end of an article, respectively, may both refer to the same entity. If pronouns and noun phrases share the same distance feature, the negative impact of long distance for pronoun anaphora will be interfered with by noun phrases. Thus, some pronouns may be linked to phrases that are far away from them, which is against our intuition. On the other hand, long distanced noun phrases co-refer will also be limited.

In this paper, we propose the modeling of coreference resolution using Markov logic networks, which can handle pronouns and noun phrases together while discriminate their differences. Specifically, we model the characteristics of pronouns and noun phrases using different formulas in Markov networks while still doing training and inference of them in the same process. A correlation clustering technique is also employed to get the final clustering results from pair-wise coreferential probabilities. Experiments show that our system achieves better results than several baseline systems that use *Separate* or *Uniform* strategies.

The rest of this paper is organized as follows: Section 2 reviews Markov logic networks. Section 3 presents our solution with MLNs. Section 4 reviews correlation clustering technique and presents its application in our work. We show our experimental settings and results in Section 5; and discuss related works in Section 6. Finally, we conclude in Section 7.

2 Markov Logic Networks

Markov logic network, introduced by Domingos and Richardson[17] is a well founded model for Statistical Rational Learning (SRL). Since MLNs are combinations of first order logic and Markov Networks, we firstly review these two parts briefly and then explain how they are used in our framework.

2.1 First Order Logic

First order logic is a formal language which describes the world by means of *constants*, *variables*, *functions*, *predicates* and *formulas.*

Constants are the elements in the world. In the scenario of coreference resolution, constants can be all the mentions in a document, such as "President Bush" and "he" in Figure 1.

A *Variable* is used to represent a set of constants. With typed variables, we can refer to different elements conveniently. For example, if we want to distinguish pronoun and noun phrases in a document, we can define two types of variables: *pronoun* and *noun.* Then we can use a variable p of the type pronoun to stand for phrases like "he", "she" and other pronouns; a variable n of the type noun to stand for phrases like "President Bush".

Functions refer to mappings between elements. For example, function *SemanticClass(Mention n)* can map a mention n to its semantic class. If *n* represents "President Bush", then the value of *SemanticClass(n)* is the constant *human.*

Predicates, which map a number of elements to a truth value, indicate properties of an element or relations between elements. For example, *IsFemale(Mention n)* indicates specify the gender of the mention n. The objective of coreference resolution can also be described by predicate. In this paper, we define the objective as *coreference(Mention n1, Mention n2)*, indicating whether mention n1 and n2 are coreferential.

Formulas are constructed from predicates using logical connectives and quantifiers and represent our knowledge of the world. We can formalize the interactions between the predicate *coreference* and other predicates into a set of formulas, which in first order logic is called a *knowledge base.*

2.2 Markov Networks and MLNs

First order logic uses a set of hard constrains (knowledge base) to describe the world. All the formulas in the knowledge base are treated equally, which means violating any of these constrains will be given an equal penalty. MLNs pack these constrains with weights, thus making the penalties higher for violations of higher weighted constrains. These weights are modeled by Markov Networks.

A Markov network (also known as Markov random field) is a model for the joint distribution of a set of variables $(X_1, X_2, ..., X_n) \in \chi$. Let G be an undirected graph with n nodes, each of which represents a variable. The model has a potential function for each clique in G. The joint distribution is given by

$$P(X = x) = \frac{1}{Z} \prod_k \phi_k(x_k) \quad (1)$$

where (x_k) is the state of the kth clique. And $\phi_k(x_k$ is the potential function on the kth clique. Z, known as the *partition function*, is given by $Z = \sum_{x \in \chi} \prod_k \phi_k(x_k)$. Equation 1 can also be expressed in a log-linear form:

$$P(X = x) = \frac{1}{Z} \exp(\sum_j \omega_j f_j(x)) \quad (2)$$

where $f_j(x)s$ are feature functions indicating the state of cliques.

An MLN defines the probability of variable X in a similar way:

$$P(X = x) = \frac{1}{Z} \exp(\sum_j \omega_j n_j(x)) \quad (3)$$

where $n_j(x)$ is the number of true groundings[3] of $formula_j$ given x; partition function $Z = \sum_{x \in X} \sum_j \omega_j n_j(x)$[17]. Lowd and Domingos[18] propose an effective way of training the weight of MLNs. For more details please refer to their original work.

The predicates and functions of first order logic have the same expressive power as features in other probability models such as Decision Trees and Maximum Entropy. But the first order formulas give MLNs stronger expressive power in representing knowledge than any other model. The well-founded theories of Markov Networks provide us with an efficient way to perform inference according to the formulas.

3 Solution with MLNs

In this section, we will explain in detail our modeling of coreference resolution using Markov logic networks and why this modeling is able to discriminate different characteristics of pronoun and noun phrases.

3.1 Features

Following the work of Soon et al.[11], we use some lexical features, semantic features and contextual features. All these features are represented as predicates and functions (as shown in Table 1).

[3] A true grounding of formula f is a setting of constants assigned to the variables of f that satisfies f .

Table1.Predicates and functions used in MLNs

Predicates:	**Descriptions:**
Coreference (Mention, Mention)	indicate whether two mentions refer to the same entity
isPronoun (Mention)	indicate whether the mention is a pronoun
isDemostrative (Mention)	indicate whether the mention is demonstrative phrases
Overlap (Mention, Mention)	indicate whether the strings of two mention are overlapping
Apposition (Mention, Mention)	indicate whether the two mentions have an appositive relation
Functions:	**Descriptions:**
SClass (Mention)	the semantic class of the given mention, which is one of person, animal, object, time, space, unknown
Gender (Mention)	the gender of the given mention, which is one of male, female and unknown
Number (Mention)	the number of the given mention, which is one of singular, plural and unknown
SSimilarity (Mention, Mention)	the string similarity ratio between the head word of two mentions; the value is mapping into similar, normal and dissimilar bye threshold 2/3 and 1/3
SDistance (Mention, Mention)	The number of sentences between two mentions; the value is mapping into same, few and many by threshold 0 and 5

3.2 Knowledge Base

Automatically learning formulas from given training data is NP-Hard[19]. However, the MLNs framework provides us with a convenient way to explicitly combine statistical models with human knowledge, which helps a lot in resolving our problem.

In our experiment, we manually construct a few formulas as our first order knowledge base according to previous work and some basic heuristics, and use MLNs to learn the weights. These formulas mainly focus on the following aspects:

- Ordinary Features - Use features described in 3.1 to indicate whether two mentions co-refer. For example:

$$\forall u, v\ Overlap(u, v) \wedge (\neg isPronoun(u)) \wedge (\neg isPronoun(v)) \Rightarrow Coreference(u, v) \tag{4}$$

$$\forall u, v\ SDistance(u, v) \wedge (\neg isPronoun(u)) \wedge (\neg isPronoun(v)) \Rightarrow Coreference(u, v) \tag{5}$$

$$\forall u, v\ Apposition(u, v) \Rightarrow Coreference(u, v) \tag{6}$$

$$\forall u, v\ isDemonstrative(v) \Rightarrow Coreference(u, v) \tag{7}$$

$$\forall u, v\ Gender(u) = Gender(v) \Rightarrow Coreference(u, v) \tag{8}$$

– Agreement Constraints - Prevent mentions that have conflict feature values of number, gender or semantic class from co-refer. For example:

$$\forall u, v \ (Gender(u)! = Gender(v)) \wedge (Gender(u))! = \text{UNK}) \tag{9}$$

$$\wedge((Gender(v)! = \text{UNK}) => !Coreference(u, v)$$

$$\forall u, v \ (Number(u)! = Number(v)) \wedge (Number(u))! = \text{UNK}) \tag{10}$$

$$\wedge((Number(v)! = \text{UNK}) => !Coreference(u, v)$$

– Reflexivity and Transitivity Constraints - Ensure that coreference is a equivalence relation.

$$\forall u, v \quad Coreference(u, v) \Rightarrow Coreference(v, u) \tag{11}$$

$$\forall u, v \quad Coreference(u, w) \wedge Coreference(v, w) \Rightarrow Coreference(u, v) \tag{12}$$

An important advantage of MLNs over previously used models such as decision trees[11, 20], maximum entropy[21] and kernel based models[8] is that MLNs learn the weights of formulas instead of individual features (predicates and functions). As shown in formula 4, we can combine the string similarity feature with the type of the mentions (noun or pronoun) to get a single formula. This formula will only be effective when u and v are both noun phrases. In this way, we can effectively distinguish the similarity of noun phrases from that of pronouns.

Another advantage of MLNs is that it can perform a global inference instead of just making some local coreferential decisions[8, 11, 20, 21]. In our experiments, formula 11 and 12 are used to ensure the reflexivity and transitivity of coreference, which set up ties among all the coreferential decisions and make those decisions more consistent and reliable.

4 Correlation Clustering

Inference of above MLNs provides us with a probability of coreferential relation between every two mentions. And we use correlation clustering[22] technique to integrate all these pair-wise probabilities into a final clustering result.

Correlation clustering technique aims at providing a global metric for the clustering quality, which is helpful for deciding whether to continue clustering or not. We follow this way and define the global object function (equation 13) as to maximize the agreement within each cluster and the disagreement between clusters:

$$\max \sum_{u,v \in M} \theta(u, v) w(u, v) + \sum_{u,v \in M} (\theta(u, v) - 1) w(u, v) \tag{13}$$

$$s.t. \qquad \theta(u, v) + \theta(w, v) \leq \theta(u, w) + 1 \tag{14}$$

$$\theta(u, v) \in \{0, 1\} \tag{15}$$

$$\theta(u, u) = 0 \tag{16}$$

where M is the set of all mentions; $\theta(u, v)$ is an indicator of whether u and v are in the same cluster; $w(u, v)$ is a similarity measure of u and v; equation 14 ensures the transitivity of clustering result; equation 15 indicates this is a integer programming problem; equation 16 gets rid of the decision making between one mention and itself. In our experiments, we set $w(u, v)$ as probability of *coreference(u, v)* minus the average probability of *coreference(u, v)* for all (u, v) pairs.

As the above constrained integer optimization is NP-Complete[22], we use a greedy based, bottom-up approach to get an approximation. In each step, our approach searches for the best merging of existing clusters that can achieve the largest gain of objective function. It will execute the merging and iterate until no such merging can be found. To avoid misleading merging, we also use a compatibility test which prevents the merging of two clusters that have obviously conflicting features. For example, two clusters will not be merged if mentions of one cluster are identified as women's names, while mentions of the other are men's names.

5 Experiments

5.1 Toolkit and Corpus

We use Alchemy Toolkit[23] for training and testing with Markov logic networks. The corpus we use is the Chinese part of ACE2005 coreference resolution dataset. We skip the process of identifying mentions in the document; and instead, use the annotation of mentions provided in the dataset, which helps us focus on the resolution of coreference itself.

5.2 Systems

Two baseline systems are built following the *Uniform Strategy*. We implement a baseline system following Soon et al.[11], except that a SVM classifier is used instead of a C4.5 decision tree for coreferential relation. All predicates and functions listed in Table 1 are used as binary feature functions. Pairwise decisions are then combined using the best-first strategy[4]. We refer to this system as *SVM-Base*.

Another baseline system uses the same classifier as the first one, but uses correlation clustering for generating final results as described in Section 4. We refer to this system as *SVM-CC*.

We build two systems basing on Markov logic networks and correlation clustering, as described in Section 3. One of them is built according to the *Separate Strategy*. A first round resolution of noun phrases is performed, in which we only consider noun phrases resolution by removing all pronouns from training and testing mentions. Then, in a second round resolution, we add pronouns into the

[4] Each mention is linked to the most confident antecedent according to the output of the classifier[20].

previous result by linking them to their best antecedent. We refer to this system as *MLNs-S*.

In the last system, following the common *Uniform Strategy*, we resolve pronoun and noun phrases in a single process, and use a predicate *isPronoun(Mention)* to distinguish them. Specially designed first order logic formulas, such as formula 4 and 5 are used, so that different weights are learnt for the two mention types. We refer to this system as *MLNs-F*.

5.3 Results

MUC6 scores We use MUC6 metric to get the precision, recall and F-measure of final result. Table 2 shows the MUC6 score of our systems.

Table2. MUC6 scores of all mentions.

	Precision	Recall	F-Measure
SVM-Base	0.78884	0.71501	0.75011
SVM-CC	**0.8374**	0.69477	0.75945
MLNs-S	0.73751	0.78415	0.76011
MLNs-F	0.75972	**0.82378**	**0.79045**

As shown in the table, although we only perform a greedy search in correlation clustering, system *SVM-CC* still improves the result of system *SVM-Base* from 0.75011 to 0.75945. It is mainly because correlation clustering draws decisions according to a global scoring function rather than just local comparison like the best-first clustering. It is also worth noticing that *SVM-CC* achieves a much higher precision over *SVM-Base*, which indicates that our greedy search successfully finds a stop point rather than keeps merging small clusters up.

In *MLNs-based* systems, *MLNs-S* achieves a F-measure of 0.76011, just slightly better than *SVM-CC*. And *MLNs-F* achieves a highest 0.79045 F-measure, which is much better than all the previous systems. For further understanding of the differences between those systems, we compute the noun phrases' MUC6 scores for each system and list them below (in Table 3).

Noun phrases MUC6 scores Noun phrases' MUC6 score is computed with all pronouns removed from both the answers and system output.

Comparison between the results of *MLNs-S* (0.80242) and *SVM-CC* (0.79105) in Table 3 shows that separating the processes of pronoun and noun phrase resolution improves the noun phrase resolution. These two systems got almost the same score in all mentions (in Table 2), which indicates that the resolution of pronouns in *MLNs-S* is not good enough. We can mainly attribute this to *MLNs-S*'s *Separate Strategy* in pronoun resolution. *MLNs-S* only links each pronoun to its best antecedent and unfairly assumes that these decisions are independent of each other.

Table3. MUC6 scores of noun phrases.

	Precision	Recall	F-Measure
SVM-Base	0.79396	0.7541	0.77352
SVM-CC	**0.84597**	0.74283	0.79105
MLNs-S	0.78968	0.81557	0.80242
MLNs-F	0.76121	**0.85246**	**0.80425**

MLNs-F, although using *Uniform Strategy* again like *SVM-CC*, still achieves a comparable (actually slightly better) result (0.80425) on noun phrases resolution as *MLNs-S*. We attribute this to our specially designed formulas such as formula 4 and 5, which prevent the interfere between noun phrases and pronouns. What's more, the *Uniform Strategy* of *MLNs-F* achieves a better results on pronoun resolution than *MLNs-S*, thus bringing *MLNs-F* the highest overall F-measure.

Altogether, as evidenced by the experiment results, our modeling using MLNs successfully takes advantage of *Uniform Strategy* in pronoun resolution while avoiding the interfere between noun phrases and pronouns, and achieves significant better results over baseline systems.

6 Related Work

The exploration of feature conflict has been mentioned by Ng and Cardie[20]. They found that string similarity features were different for pronoun and other types of mentions. As a result, they suggested a split of features for each type of mentions, which did bring some improvements. However, they didn't get good enough results because of the use of much simpler models such as decision trees and an information gain based rule system called RIPPER. The modeling of MLNs is much simpler and more natural than splitting features.

The inspiration of using Markov logic networks comes from[5, 16]. Singla and Domingos[5] used MLNs in entity resolutions of database items, where several simple first order rules brought comparable results with existing methods. Culotta et al.[16] extended the knowledge source of coreferential decision making from mention pairs to mention clusters. Their work motivates us to use probabilistic graph model, such as Markov Networks, for coreference. They also mentioned the conjunction of features, which lead us to the use of logic connectives and quantifiers for building more complex formulas. However, they only reported results using a conjunction of size 2, which did not make full use of the expressive power of first order logic and was not able to capture the complex relations among features.

Yang et al.[12] proposed a twin-candidate model which considered the relation between three mentions instead of two in previous resolution framework. Denis and Baldridge[9] extended it into a candidate ranking model, which took all candidates of antecedent into consideration. Both of the two works solve this problem from a local view that only considers the antecedents for one mention at a time. But our modeling using MLNs is able to capture the global interactions not only between candidates of antecedent but also between any other two mentions in the article. And the use of correlation clustering provides a global arrangement of different coreferential relations, which may lead us to a better solution.

Our approach shares the same motivation with Choi and Cardie[14], namely, the resolving of anaphora needs structured information. They solved this problem in a framework based on conditional random field and achieved convincing results in an English corpus.

Poon and Domingos also proposed the use of Markov logic networks for coreference resolution[24]. However, they designed MLNs in an unsupervised manner, thus made their work quite different from ours. Instead of directly modeling the coreferential possibility between two lexicalized mentions, we are trying to model the latent rules for the CR task, which is more challenging and data dependent. As a large amount of CR data is usually difficult to get, some researchers began to explore unsupervised CR methods and also got promising results[24–26].

7 Conclusions and Future Work

We analyze the two strategies of pronouns and noun phrases coreference resolution, especially the relation and interference between these two mention types. Based on the analysis, we propose the use of Markov logic networks for solving these two types of coreference in a single process. Our model is able to use global information for anaphora resolution while successfully avoid the interference between pronouns and noun phrases. We also employ a correlation clustering technique which gives us a global metric during combining various coreferential relations from MLNs. Experiments show that our strategy improves the performance significantly over the baseline systems.

Future work will focus on integrating other knowledge sources like Centering Theory and syntactic information into our framework and making use of more shallow semantic information as[27].

To improve the performance, we plan to use LP chunking techniques for the solution of correlation clustering in Section 4, instead of currently used greedy based approach. We will also try to extend our pairwise target predicate to a setwise one, which may integrate the coreferential decision making and clustering into one process.

Another interesting direction is to automatically distinguish the features and knowledge bases between pronouns and noun phrases, and to further explore the interactions of pronouns and noun phrases in coreference resolution.

Acknowledgement. This work is supported by the National Natural Science Foundation of China under Grant No. 60673043, the National 863 High-Tech Program under Grant No. 2003AA010109, the National Social Science Foundation of China under

Grant No. 07BYY051, and the Natural Science Foundation of Jiangsu Higher Education Institutions of China under Grant No. 07KJB520057.

References

1. McCarthy, J.F., Lehnert, W.G.: Using decision trees for coreference resolution. In: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence. (1995) 1050–1055
2. Cardie, C., Wagstaff, K.: Noun phrase coreference as clustering. In: Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, University of Maryland, MD, Association for Computational Linguistics (1999) 82–89
3. Morton, T.S.: Coreference for nlp applications. In: Proceedings of the 38th Annual Meeting on Association for Computational Linguistics. (2000)
4. Steinberger, J., Kabadjov, M.A., Poesio, M., Sanchez-Graillet, O.: Improving lsa-based summarization with anaphora resolution. In: Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing. (2005)
5. Singla, P., Domingos, P.: Entity resolution with markov logic. In: Proceedings of the Sixth International Conference on Data Mining, Washington, DC, USA, IEEE Computer Society (2006) 572–582
6. Wang, H., Mei, Z.: An empirical study on pronoun resolution in chinese. In: Computational Linguistics and Intelligent Text Processing, 5th International Conference, CICLing 2004, Seoul, Korea, February 15-21, 2004, Proceedings. (2004) 213–216
7. Iida, R., Inui, K., Matsumoto, Y.: Anaphora resolution by antecedent identification followed by anaphoricity determination. ACM Transactions on Asian Language Information Processing (TALIP) **4**(4) (2005) 417–434
8. Yang, X., Su, J., Tan, C.L.: Kernel-based pronoun resolution with structured syntactic knowledge. In: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL, Morristown, NJ, USA, Association for Computational Linguistics (2006) 41–48
9. Denis, P., Baldridge, J.: A ranking approach to pronoun resolution. In: Proceedings of the 20th International Joint Conference on Artificial Intelligence. (2007) 1588–1593
10. Denis, P., Baldridge, J.: Specialized models and ranking for coreference resolution. In: Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, Honolulu, Hawaii, Association for Computational Linguistics (October 2008) 660–669
11. Soon, W.M., Ng, H.T., Lim, D.C.Y.: A machine learning approach to coreference resolution of noun phrases. Computational linguistics **27**(4) (2001) 521–544
12. Yang, X., Su, J., Tan, C.L.: A twin-candidate model of coreference resolution with non-anaphor identification capability. In: Second International Joint Conference. (2005) 719–730
13. Nicolae, C., Nicolae, G.: Bestcut: A graph algorithm for coreference resolution. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Sydney, Australia, Association for Computational Linguistics (July 2006) 275–283

14. Choi, Y., Cardie, C.: Structured local training and biased potential functions for conditional random fields with application to coreference resolution. In: Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT/NAACL), Rochester, New York, Association for Computational Linguistics (April 2007) 65–72
15. Yang, X., Su, J., Lang, J., Tan, C.L., Liu, T., Li, S.: An entity-mention model for coreference resolution with inductive logic programming. In: Proceedings of ACL-08: HLT, Columbus, Ohio, Association for Computational Linguistics (June 2008) 843–851
16. Culotta, A., Wick, M., Hall, R., McCallum, A.: First-order probabilistic models for coreference resolution. In: Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics (HLT/NAACL). (2007) 81–88
17. Domingos, P., Richardson, M.: Markov logic: A unifying framework for statistical relational learning. In: Proceedings of the ICML-2004 Workshop on Statistical Relational Learning and its Connections to Other Fields, Banff, Canada (2004) 49–54
18. Lowd, D., Domingos, P.: Efficient weight learning for markov logic networks. In: Proceedings of 11th European Conference on Principles and Practice of Knowledge Discovery in Databases. (2007) 200–211
19. Richardson, M., Domingos, P.: Markov logic networks. Machine Learning **62**(1-2) (2006) 107–136
20. Ng, V., Cardie, C.: Improving machine learning approaches to coreference resolution. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Morristown, NJ, USA, Association for Computational Linguistics (2002) 104–111
21. Ng, H.T., Zhou, Y., Dale, R., Gardiner, M.: A machine learning approach to identification and resolution of one-anaphora. In: Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence. (2005) 1105–1110
22. Bansal, N., Blum, A., Chawla, S.: Correlation clustering. Machine Learning **56** (2004) 89–113
23. Kok, S., Singla, P., Richardson, M., Domingos, P.: The alchemy system for statistical relational ai. Technical report, Department of Computer Science and Engineering, University of Washington, Seattle, WA, http://www.cs.washington.edu/ai/alchemy/ (2005)
24. Poon, H., Domingos, P.: Joint unsupervised coreference resolution with Markov Logic. In: Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, Honolulu, Hawaii, Association for Computational Linguistics (October 2008) 650–659
25. Haghighi, A., Klein, D.: Unsupervised coreference resolution in a nonparametric bayesian model. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic, Association for Computational Linguistics (June 2007) 848–855
26. Ng, V.: Unsupervised models for coreference resolution. In: Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, Honolulu, Hawaii, Association for Computational Linguistics (October 2008) 640–649
27. Ng, V.: Shallow semantics for coreference resolution. In: Proceedings of the 20th International Joint Conference on Artificial Intelligence. (2007) 1689–1694