A Survey of Chinese Anaphora Resolution

Shengnan Li[1], Weiguang Qu[1,2](✉), Tingxin Wei[2,3], Junsheng Zhou[1], Yanhui Gu[1], and Bin Li[2]

[1] School of Computer and Electronic Information/School of Artificial Intelligence, Nanjing Normal University, Nanjing 210023, China
[2] School of Chinese Language and Literature, Nanjing Normal University, Nanjing 210097, China
[3] International College for Chinese Studies, Nanjing Normal University, Nanjing 210097, China

Abstract. Chinese anaphora resolution technology has been widely used in many natural language processing tasks, such as machine translation, information extraction and automatic text summarization. In this paper, we first introduce the resources for anaphora resolution, and then present the existing works on Chinese noun phrase resolution based on machine learning, deep learning and reinforcement learning techniques by analyzing the similarities and differences among them. Finally, we discuss the future development trend of Chinese anaphora resolution.

Keywords: Anaphora resolution · Chinese · Noun phrase

1 Introduction

Anaphora [1] is a substituted expression to refer to the denotation of a preceding word or phrase, which is a very common language phenomenon in language. There are several types of anaphora, pronoun anaphora, zero anaphora, and noun phrase anaphora. In many nlp tasks such as information extraction and machine translation, it is crucial to identify which word or phrase the anaphoric expression refer to for text understanding, therefore anaphora resolution is one of the fundamental tasks in natural language processing. Many researchers have put efforts on zero anaphora resolution and pronoun anaphora resolution, and have achieved outstanding improvements on these two subtasks. However, few studies focus on noun phrase resolution which is more complicated yet more urgent to be solved. In this paper, we present the development of noun phrase anaphora resolution from the perspectives of resource, method and feature attribute, especially on Chinese anaphora resolution. Finally, we give a brief discussion about the future trend of anaphora resolution.

2 Resources for Anaphora Resolution

For the task of anaphora resolution, three datasets are widely used to test and evaluate, which are MUC, ACE and OntoNotes. In this section, we give a detailed introduction to these datasets.

X. Sun et al. (Eds.): ICAIS 2021, LNCS 12736, pp. 180–192, 2021.
https://doi.org/10.1007/978-3-030-78609-0_16

2.1 MUC Dataset

The Message Understanding Conference (MUC) proposed two semantic evaluation tasks in 1995 and 1998—anaphora resolution and named entity disambiguation, and provided datasets for task evaluation, namely MUC-6 [2] and MUC-7 [3]. MUC-6 is composed of 25 articles from *The Wall Street Journal*, totaling approximately 30,000 words. It is the first corpus that can be used for the evaluation of anaphora resolution in a real sense. MUC-7 consists of a similar number of articles on aircraft crashes and rocket or missile launch. However, due to the emergence of some larger-scale datasets, MUC-6 and MUC-7 are no longer widely used. Despite of this, they have a profound impact on the study of anaphora resolution of the MUC corpus. Table 1 details the information of the MUC corpus.

Table 1. The statistics of the MUC corpus

Corpus	Language	Size of words			Size of files		
		Training	Dev&Test	Total	Training	Dev&Test	Total
MUC-6	English	12k	13k	25k	30	30	60
MUC-7	English	19k	21k	40k	30	37	67

2.2 ACE Dataset

Automatic Content Extraction (ACE) is an evaluation conference hosted by the National Institute of Standards and Technology. This event began since July 1997 and has been held for 7 sessions so far. The anaphora information in ACE is labeled in the form of an entity chain. The entity chain of each article is independently recorded in a corresponding XML file in the corpus. Therefore, the ACE corpus [4] replaces the MUC corpus as the new evaluation resource for anaphora resolution. The initial version of the ACE corpus only contains news. In the later versions, more types of articles, such as broadcast conversations, web logs and telephone conversations are added, which greatly increases the coverage of the evaluation. In 2003, ACE evaluation began to carry out the evaluation of anaphora resolution for the Chinese corpus. Table 2 details the information of the ACE corpus.

Table 2. The statistics of the ACE corpus

Corpus	Language	Training Size	Dev&Test Size	Total
ACE (2000–2004)	English	745k	215k	960k
	Chinese	455k	150k	615k
	Arabic	350k	150k	500k

2.3 OntoNotes Dataset

In 2011, the CoNLL conducted an evaluation of anaphora resolution for English [5]. In 2012, the evaluation for Chinese and Arabic was added [6] to study the differences of anaphora resolution in different languages. There are 5 released versions of the OntoNotes corpus. The latest two versions, OntoNotes 4.0 [7] and OntoNotes 5.0 [8] are used for the CoNLL-2011 shared task and CoNLL-2012 shared task respectively. OntoNotes collects a large number of articles on newswires, broadcast news and conversations, web texts and telephone conversations, and it integrates annotations in multiple levels, including part-of-speech tagging, component syntax analysis, named entity recognition and semantic role tagging, and so on. Its Chinese and English parts contain approximately 1 million and 1.6 million words respectively. At present, it is the largest anaphora resolution corpus. Table 3 details the information of OntoNotes.

Table 3. The statistics of the OntoNotes corpus

Corpus	Language	Size of words				Size of files			
		Train	Dev	Test	Total	Train	Dev	Test	Total
OntoNotes	English	1.3M	160k	170k	1.6M	1940	222	222	2384
	Chinese	750k	110k	90k	950k	1391	172	166	1729
	Arabic	240k	30k	30k	300k	359	44	44	447

3 Anaphora Resolution Methods

In this section, we present existing works on anaphora resolution implemented on the three datasets: MUC, ACE and OntoNotes. In general, the existing anaphora resolution methods can be categorized as rule-based, machine learning-based, deep learning-based and reinforcement learning-based methods. We will review these methods in English anaphora resolution and Chinese anaphora resolution respectively.

3.1 English Anaphora Resolution

The rule-based method [9–16] has been extensively studied. One of the advantages of the rule-based method is its simplicity for designing. However, this method has low flexibility. In the 1990s, the significant attention of anaphora resolution gradually evolved from the rule-based method to the machine learning method. This change was mainly due to the open access of the MUC-6 and MUC-7 datasets to the public. The machine learning method is classified into supervised and unsupervised methods [17–24]. The former requires lots of labeled training data, while the latter does not.

Recently, deep neural networks are used for anaphora resolution in many works [25–29]. One of the advantages of using deep neural networks is that it can extract useful features from the raw texts automatically without human intervention. In addition, these methods utilize word embedding to represent words and describe semantic relationships between them. However, these methods have many defects such as huge dimension size and inappropriate architecture. Hourali et al. [30] have tried addressing these issues by using contextual, semantic and syntactic information for a better representation of spans. They used the neural MCDM method for the accurate ranking of candidate antecedent to increase the detection rate of coreference mentions. Clark et al. [31] proposed a reinforcement learning method for anaphora resolution by only developing a co-reference chain model. Inspired by Clark, Fei et al. [32] proposed a method of anaphora resolution based on end-to-end deep reinforcement learning to avoid the cascading errors in the pipeline system. This method directly considers all text spans, and jointly recognizes entity mentions and performs clustering operating. Table 4 shows the performances of various methods in English anaphora resolution.

Table 4. Experiment results in English anaphora resolution with various methods

Method	Dataset	Model	F1%
Machine learning-based	MUC-6	Cardie et al. [23]	54
		Soon et al. [17]	62.60
		Ng and Cardie et al. [18]	66.30
		Yang et al. [19]	71.30
		Haghighi et al. [24]	70
		Yang et al. [21]	68.70
		Li et al. [22]	68.60
Rule-based		Raghunathan et al. [11]	77.70
	OntoNotes4.0	Lee et al. [12]	61.40
Reinforcement learning-based	OntoNotes5.0	Clark et al. [31]	65.73
		Fei et al. [32]	73.80
Deep learning-based		Lee et al. [26]	68.80
		Lee et al. [25]	73
		Zhang et al. [27]	76.50
		Kantor et al. [28]	76.61
		Joshi et al. [29]	76.90
		Hourali et al. [30]	80

3.2 Chinese Anaphora Resolution

In the task of anaphora resolution, there are many different types of anaphoric expression to be identified, such as common pronoun, demonstrative noun phrase, proper noun phrase, event anaphora and zero anaphora. In this section, we focus on the anaphora resolution of noun phrase in Chinese, which refer to pronouns, proper nouns, indicative noun phrases and common noun phrases. The existing methods of Chinese anaphora resolution can be categorized as rule-based methods and data-driven methods, which are detailed in the following parts of this section.

Rule-Based Methods for Chinese Anaphora Resolution. Due to the lack of fine labeled resources, the rule-based method is widely used in the early anaphora resolution system, which mainly focuses on the theoretical exploration and fusion of a large amount of domain knowledge and linguistic knowledge. These rule-based methods implement the anaphora resolution with some linguistic rules concluded by experts.

Zhang et al. [33] proposed a model using a rule-based filtering method for the CoNLL-2012 Shared Task (BCMI). They designed different filtering strategies for different situations. This method achieved an average F1 value of 51.83 in the experiment. Based on the model proposed by Zhang, Zhou et al. [34] added a semantic matching layer. This layer can properly overcome the defects in the Chinese semantic knowledge base by applying Web semantic knowledge. The experimental results on the ACE2005 Chinese corpus reached an average F1 value of 78.2.

In a word, the main idea of the rule-based hierarchical filtering model is to manually set a series of filtering rules for anaphora resolution, which requires a large amount of human labor with low system automation and poor portability. Although these rule bases cannot cover all language phenomena, these methods establish the foundation for future research.

Data-Driven Supervised Methods for Chinese Noun Phrase Resolution. In 2003, ACE initiated the evaluation on Chinese anaphora resolution, and some researchers started to apply the data-driven supervised method on Chinese anaphora resolution.

Hu et al. [35] designed a Chinese system based on a Maximum Entropy model. They extracted 12 features from raw texts to train a maximum entropy-based classifier, then had it tested on ACE news corpus, and achieved 78.87 on F1 value. Liu et al. [36] proposed a supervised correlation clustering algorithm for anaphora resolution. The learning algorithm based on gradient descent is proposed to make the feature parameters trained from the training set fit the objective of the correlation clustering better. With the feature set defined in [17], Liu defined and selected 10 features for Chinese anaphora resolution. Since anaphora resolution requires adequate information to identify the exact expression, and the existing features are still insufficient. Liu proposed a loss function based on minimizing decision errors and optimizing with a bottom-up clustering method [37]. The idea of this method is to use the information in the major opinion from the binary classification to correct wrong decisions, while at the same time, trying to avoid violating the decision made by the original binary classifier in general. The test result on the ACE Chinese corpus achieved 77.98 on F1 value. Li et al. [38] proposed a method based on feature ranking strategy for noun phrase resolution. This method processed personal pronouns and common noun phrases separately when selecting feature vectors, to take full use of the features of different noun phrases for anaphora resolution. The experimental results showed that the F1 value reached 80.72.

Tan et al. [39] proposed a Chinese noun phrases anaphora resolution method based on SVM. The method trained on the ACE2005 corpus of Chinese and resolved all general noun phrases with 13 selected feature vectors. Gao et al. [40] proposed a method based on SVM with 17 selected feature vectors and the experimental results on ACE 2005 achieved 71.59 on F1 value. Zhou et al. [41] considered that the traditional Laplacian support vector machine only uses Euclidean distance to calculate the distance between two samples, which may lead to the false high similarity between two samples from different categories. To address the issue of insufficient Chinese annotated corpus, a data-driven learning method of optimal distance measure is proposed. In this method, the similarity constraints between sample pairs are considered, and the Fisher discriminant criterion is introduced, which increases the similarity within the same category and highlights the discriminant features in the new metric space. Compared with the classical supervised methods on the ACE2005 Chinese corpus, the linear and kernel supervised methods achieve better results with fewer labeled samples. Zhou also proposed a Chinese anaphora resolution model based on an improved hierarchical filtering module [42]. Table 5 are the hierarchical filter modules used in this model.

Table 5. The sequence of filtering modules for Chinese anaphora resolution

Sequence	Data type	Features
1	Noun phrases	String matching
2	Noun phrases	Entity expression loose string | entity loose string matching
3	Noun phrases	Strict head word matching | loose head word matching
4	Noun phrases	Appositive | abbreviations
5	Noun phrases	HowNet semantic similarity calculation | Web Semantic Similarity Calculation
6	Pronoun	Singular and plural Identity; Gender Identity; Life degree; NER tag

Data-Driven Unsupervised Methods for Chinese Noun Phrase Resolution. The supervised machine learning method requires large scale of labeled training data. However, well-annotated corpus is not easily reachable since not only linguistic knowledge is needed, but also domain knowledge as well. Besides, word segmentation of Chinese also bring obstacles in understanding and annotation. Therefore, it is extremely labor-intensive and time-consuming to construct a large-scale anaphora resolution corpus. In the unsupervised learning model, there is no requirements for a large amount of labeled corpus, which can effectively save the time and cost of manual labeling. In this section, we mainly introduce the unsupervised method for Chinese noun phrase anaphora resolution.

Wang et al. [43] explored the reasons why the Chinese anaphora phenomenon is difficult to resolve and proved that English anaphora resolution technology can be extended to Chinese. It is the first work that attempts to use an unsupervised method to address Chinese noun phrases anaphora resolution.

According to Cardie [23] in English anaphora resolution, Zhou et al. [44] presented a new unsupervised clustering algorithm for noun phrase anaphora resolution. Different from Cardie, this approach firstly converted the issue of anaphora resolution into that of graph clustering, which did not make anaphora decisions for each pair of noun phrases in isolation but fully considered the correlation between items. Li et al. [45] proposed a Chinese anaphora resolution method based on association clustering, and then correlation clustering is used for automatic graph clustering. Compared with the traditional clustering methods such as link-first and link-best, the proposed algorithm takes full use of the relations among the noun phrases sufficiently. In addition, it does not need to specify the desired number of clusters and a distance threshold. It takes each item to be resolved in the test set as a vertex in the graph and regards the confidence value between the two items to be resolved as the weight corresponding to the edge connecting the two vertices, thus forming a graph G (V, E). Then, they use the associated clustering algorithm to automatically divide the graph G, so that the process of referencing resolution is transformed into the division process of the graph G.

However, one of the issues of traditional clustering methods is that it is difficult to judge when to stop. Besides, since it is difficult to obtain or estimate the total number of entities in text, the number of clusters cannot be predicted, which is an important parameter that influence the final performance on clustering. Most of the current methods are based on the experience gained from a large amount of data, without considering the characteristics of the task itself, which affects the performance of clustering in anaphora resolution. In addition, the clustering features used by the existing researches are directly manually selected, and the interference of the weaker distinguishing features on the clustering effect is not considered. To address these issues, Li et al. [46] proposed an unsupervised method (adaptive resonance theory, ART). This method makes full use of the characteristics of noun phrases and dynamically adjusts the number of clusters by changing network parameters. In addition, a feature selection method based on information gain ratio is adopted to reduce the interference of weaker features on cluster results. This method does not rely on manual annotation corpus under the premise of ensuring the correct recognition rate and can be directly applied to real texts cross domains.

Gao et al. [47] designed a Chinese noun phrase anaphora resolution system based on unsupervised clustering, including preprocessing, feature selection and clustering. In the preprocessing stage, the system obtains the feature information of noun phrases. With the feature information and a series of incompatible functions defined, a hierarchical clustering algorithm is used to cluster the noun phrases that may be in a co-reference chain.

Deep Learning-Based Methods for Chinese Noun Phrase Resolution. Most of the existing neural anaphora resolution models only focus on the linear features of text, ignoring the integration of structural information. Fu et al. [48] proposed two measures to address the issue with the constituency parse tree, which is based on the neural network model avoiding structural information loss. Considering the characteristics of Chinese, Fu et al. [49] proposed a Chinese anaphora resolution model with structural information involved. The constituency tree of all sentences in the document is compressed to obtain the leaf node depth. Structural information is vectorized by the Structural Embedding of Component Tree (SECT) method. In addition, the leaf node depth and the SECT information are used as three feature vectors in the model for Chinese anaphora resolution.

The specific process of the SECT method is as follows. (1) Define a syntactic sequence S(p) for word p which is the leaf node in the syntactic tree, save the path from the leaf node p to the root node of the syntactic tree. For example, for the "NR 李鹏(Li Peng)" node in Fig. 1, the corresponding S(p) is {NR, NP-PN, NP-SBJ, IP-HLN, TOP}. In the experiment, considering the performance and memory utilization, a window is set to limit the length of S(p). Due to the reason that the higher-level nodes in the syntactic tree are more ambiguous, with the help of window size limitation, the higher-level ambiguous nodes in the syntax tree will be removed. (2) A bidirectional LSTM is used to encode the variable-length syntax sequence S(p) into a fixed-length vector representation. Suppose x_t^p represents the t-th node of the word p in the syntactic sequence S(p), and v_t^p represents the hidden layer representation of x_t^p. The formula is as follows:

$$v_t^p = f_{BiLSTM}\left(v_{t-1}^p, x_t^p\right) \tag{1}$$

Then for a given word p, finally select $v_{BiLSTM}^p = v_T^P$ as the final vector representation for S(p).

```
Treebanked sentence:
--------------------
    李鹏 总理 同 阿卡耶夫 总统 举行 会谈
Premier Li Peng and President Akayev hold talks
Tree:
-----
    (TOP (IP-HLN (NP-SBJ (NP-PN (NR 李鹏(Li Peng))
                                (NN 总理(premier)))
                         (CC 同(and))
                         (NP-PN (NR 阿卡耶夫(Akayev))
                                (NN 总统(president))))
                 (VP (VV 举行(hold))
                     (NP-OBJ (NN 会谈(talk))))))
```

Fig. 1. Syntax tree example

Reinforcement Learning-Based Methods for Chinese Noun Phrase Resolution. The coreference resolution systems usually use a heuristic loss function for training, which needs to be adjusted carefully. Clark et al. [31] applied reinforcement learning to directly optimize the neural mention-ranking model used for coreference evaluation metrics, thus avoiding the selection of hyperparameters for each specific language, data set and evaluation metrics. The mention ranking model scores the possibility of mentioning, rather than comparing partial mention clusters. As a result, they operate in a simple environment in which coreference decisions are made independently. The experiments were conducted with two methods, enhanced strategy gradient algorithm and reward rescaling maximum profit target. The results showed that the latter method was more effective, which improved the performance of Chinese and English anaphora resolution respectively in CoNLL-2012 shared task.

3.3 Summary

In this section, we summarize the methods for noun phrase resolution based on machine learning, deep learning and reinforcement learning. Table 6 shows the performance of existing methods on Chinese datasets. It can be seen that the number of feature vectors has great impact on the final performance. Specifically, the performance of the same machine learning method with different number of feature vectors may be quite different. According to the language characteristics and background knowledge, choosing appropriate features can greatly improve the accuracy of the classification model. However, it does not mean that more features bring better performance. That is because with the increase of features, the possibility of data sparseness increases as well. Therefore, the efficiency of the classification model relies on the selection of powerful characteristic attributes.

Table 6. Experimental results on Chinese noun phrase anaphora resolution

Method	Dataset	Model	Feature	F1%
Supervised-based	ACE2005	Tan et al. [39]	13	63.30
		Hu et al. [35]	12	78.87
		Liu et al. [36]	10	78.01
		Zhou et al. [42]	11	77.10
	ACE2007	Li et al. [38]	10	80.72
	OntoNotes3.0 (news)	Gao et al. [40]	17	71.59
Unsupervised-based		Gao et al. [47]	14	59.43
	ACE2005	Zhou et al. [44]	9	62.05
		Li et al. [45]	9	76.45
		Li et al. [46]	8	70.20
Deep learning-based	OntoNotes5.0	Fu et al. [48]	3	62.35
		Fu et al. [49]	3	62.33
Reinforcement learning-based		Clark et al. [31]	—	63.88

4 Future Prospects

With an increasing probability of multiple languages in one text, the multilingual anaphora resolution becomes a hot topic. The main task of multilingual anaphora resolution is to directly apply the anaphora resolution method of a certain language to other languages. Due to the difference among languages in many perspectives, it is a big challenge for researchers to develop novel networks for it.

At present, another trend in anaphora resolution is to formalize and combine syntactic information, semantic knowledge and background knowledge in anaphora resolution.

Although papers based on the use of background knowledge, grammar and semantic knowledge for anaphora resolution have been published at top natural language processing conferences, the performance is still not very satisfactory, which is mainly caused by the following reasons: 1. the background knowledge extraction and the background knowledge formalization is very difficult. 2. It is difficult to automatically obtain grammatical and semantic knowledge and further formalize it into effective rules. Therefore, obtaining and using syntactic information, semantic knowledge, and background knowledge has been also a new hot topic in the development of anaphora resolution.

Acknowledgments. The authors thank the anonymous reviewers for their constructive suggestions which have resulted in improvement on the presentations. This research is supported by the National Science Foundation of China (grant 61772278, author: Qu, W.; grand number: 61472191, author: Zhou, J. http://www.nsfc.gov.cn/), the National Social Science Foundation of China (grant number: 18BYY127, author: Li B. http://www.cssn.cn) and Jiangsu Higher Institutions' Excellent Innovative Team for Philosophy and Social Science (grand number: 2017STD006, author: Qu, W. http://jyt.jiangsu.gov.cn).

References

1. Wang, H.: Survey: computation models and technologies in anaphora resolution. J. Chin. Inf. Process. **16**(6), 9–17 (2002). (In Chinese)
2. Vilain, M., Burger, J., Aberdeen, J., Connolly, D., Hirschman, L.: A model-theoretic coreference scoring scheme. In: Proceedings of the Sixth Message Understanding Conference (MUC-6), pp. 45–52. Morgan Kaufmann Publishers, San Francisco (1995)
3. Hirschman, L., Robinson, P., Burger, J., Vilain, M.: Automating coreference: the role of annotated training data. In: Proceedings of the AAAI Spring Symposium on Applying Machine Learning to Discourse Processing, pp. 118–121. AAAI, Rhode Island, USA (1997)
4. Doddington, G., Mitchell, A., Przybocki, M., Ramshaw, L., Strassel, S., Weischedel, R.: The automatic content extraction (ACE) program tasks, data, and evaluation. In: Proceedings of the Fourth International Conference on Language Resources and Evaluation, Lisbon, Portugal, pp. 837–840 (2004)
5. Pradhan, S., Ramshaw, L., Marcus, M., Palmer, M., Xue, N.: CoNLL-2011 shared task: modeling unrestricted coreference in ontonotes. In: Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task, pp. 1–27, ACL, Portland, Oregon (2011)
6. Pradhan, S.S., Moschitti, A., Xue, N., Uryupina, O., Zhang, Y.: CoNLL-2012 shared task: modeling multilingual unrestricted coreference in OntoNotes. In: Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning Proceedings-Shared Task, pp. 1–40. ACL, Jeju Island (2012)
7. Weischedel, R., et al.: OntoNotes Release 4.0 LDC2011T03. Web Download. Philadelphia: Linguistic Data Consortium (2011)
8. Weischedel, R., Palmer, M., Marcus, M., Hovy, E., Pradhan, S., Ramshaw, L.: OntoNotes release 5.0 LDC2013T19. Web Download. Philadelphia: Linguistic Data Consortium (2013)
9. Harabagiu, S.: From lexical cohesion to textual coherence: a data driven perspective. Int. J. Pattern Recognit. Artif. Intell. **13**(2), 247–265 (1999)
10. Moosavi, N., Strube, M.: Lexical features in coreference resolution: to be used with caution. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 14–19. ACL, Vancouver, Canada (2017)

11. Raghunathan, K., Lee, H., Rangarajan, S., Chambers, N., Manning, C.D.: A multipass sieve for coreference resolution. In: Conference on Empirical Methods in Natural Language Processing, pp. 492–501. DBLP, USA (2010)
12. Lee, H., Peirsman, Y., Chang, A., Chambers, N., Jurafsky, D.: Stanford's multi-pass sieve coreference resolution system at the CoNL-2011 shared task. In: Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task, pp. 28–34. ACL, USA (2011)
13. Kibble, R.: A reformulation of rule 2 of centering theory. Comput. Linguist **27**(4), 579–587 (2001)
14. Zeldes, A., Zhang, S.: When annotation schemes change rules help: a configurable approach to coreference resolution beyond OntoNotes. In: Proceedings of the Workshop on Coreference Resolution Beyond OntoNotes, pp. 92–101. ACL, San Diego, California (2016)
15. Lee, H., Chang, A., Peirsman, Y., Chambers, N., Jurafsky, D.: Deterministic coreference resolution based on entity-centric, precision-ranked rules. Comput. Linguist. **39**(4), 885–916 (2013)
16. Haghighi, A., Klein, D.: Simple coreference resolution with rich syntactic and semantic features. In: Proceedings of 2009 Conference on Empirical Methods in Natural Language Processing, pp. 1152–1161. ACL, Singapore (2009)
17. Soon, W., Ng, H., Lim, D.: Machine learning approach to coreference resolution of noun phrases. Comput. Linguist. **27**(4), 521–544 (2001)
18. Ng, V., Cardie, C.: Improving machine learning approaches to coreference resolution. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp. 104–111. ACL, Philadelphia, USA (2002)
19. Yang, X., Zhou, G., Su, J., Tan, C.: Coreference resolution using competition learning approach. In: Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics, pp. 176–183. ACL, Sapporo, Japan (2003)
20. Qian, W., Guo, Y., Zhou, Y., Wu, L.: English noun phrase coreference resolution via a maximum entropy model. J. Comput. Res. Dev. **40**(9), 1337–1342 (2003). (In Chinese)
21. Yang, Y., Li, Y., Zhou, G., Zhu, Q.: Research on distance information for anaphora resolution. J. Chin. Inf. Process. **22**(5), 39–44 (2008). (In Chinese)
22. Li, Y., Yang, Y., Zhou, G., Zhu, Q.: Anaphora a resolution of noun phrase based on SVM. Comput. Eng. **35**(3), 199–204 (2009). (In Chinese)
23. 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, pp. 82–89 (1999)
24. Haghighi, A., Dan, K.: Unsupervised coreference resolution in a nonparametric Bayesian model. In: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pp. 848–855. ACL, Prague (2007)
25. Lee, K., He, L., Zettlemoyer, L.: Higher-order coreference resolution with coarse-to-fine inference. In: Proceedings of 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 687–692. ACL, New Orleans, Louisiana (2018)
26. Lee, K., He, L., Lewis, M., Zettlemoyer, L.: End-to-end neural coreference resolution. In: Proceedings of 2017 Conference on Empirical Methods in Natural Language Processing, pp. 188–197. ACL, Copenhagen, Denmark (2017)
27. Zhang, R., Santos, C., Yasunaga, M., Xiang, B., Radev, D.: Neural coreference resolution with deep biaffine attention by joint mention detection and mention clustering. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, pp. 102–107. ACL, Melbourne, Australia (2018)

28. Kantor, B., Globerson, A.: Coreference resolution with entity equalization. In: Proceedings of the 57th Conference of the Association for Computational Linguistics, pp. 673–677. ACL, Italy (2019)
29. Joshi, M., Levy, O., Weld, D.S., Zettlemoyer, L.: BERT for coreference resolution: baselines and analysis. In: Proceedings of 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 5802–5807. ACL, Hong Kong, China (2019)
30. Hourali, S., Zahedi, M., Fateh, M.: Coreference resolution using neural MCDM and fuzzy weighting technique. Int. J. Computat. Intell. Syst. **13**(1), 56–65 (2020)
31. Clark, K., Manning, C.: Deep reinforcement learning for mention-ranking coreference models. In: Proceedings of 2016 Conference on Empirical Methods in Natural Language Processing Texas, pp. 2256–2262. ACL, Austin, Texas (2016)
32. Fei, H., Li, X., Li, D., Li, P.: End-to-end deep reinforcement learning based coreference resolution. In: Proceedings of the 57th Conference of the Association for Computational Linguistics, pp. 660–665. ACL, Florence, Italy (2019)
33. Zhang, X., Wu, C., Zhao, H.: Chinese coreference resolution via ordered filtering. In: Joint Conference on Emnlp and Conll-shared Task, pp. 95–99. ACL, Jeju Island, Korea (2012)
34. Zhou, X., Liu, J., Luo, Y., Han, Y.: Comparison of Chinese anaphora resolution models. Comput. Sci. **43**(2), 31–34 (2016). (In Chinese)
35. Hu, N., Kong, F., Wang, H., Zhou, G., Zhu, Q.: Realization on Chinese coreference resolution system based on maximum entropy model. Appl. Res. Comput. **26**(8), 2948–2951 (2009). (In Chinese)
36. Liu, W., Zhou, J., Huang, S., Chen, J.: Coreference resolution with supervised correlation clustering. Comput. Sci. **36**(9), 182–185 (2009). (In Chinese)
37. Liu, W., Zhou, J., Huang, S.: global optimization based on clustering for coreference resolution. In: Frontier Progress of Chinese Computer Linguistics, pp. 295–301. CIPSC, China (2009). (In Chinese)
38. Li, Y., Gan, R., Yang, Y., Shi, S.: Chinese coreference resolution method based on feature respective selection strategy. Comput. Eng. **37**(18), 180–182 (2011). (In Chinese)
39. Tan, W., Kong, F., Wang, D., Zhou, G.: An SVM-based approach to chinese anaphora resolution. High-Performance Comput. Technol. 0(2), 30–36 (2010). (In Chinese)
40. Gao, J., Kong, F., Zhu, Q., Li, P.: Research of Chinese noun phrase anaphora resolution: an SVM-based approach. Comput. Sci. **39**(10), 231–234 (2012). (In Chinese)
41. Zhou, X., Liu, J., Shao, P., Xiao, L., Luo, F.: Chinese anaphora resolution based on metric-optimized Laplacian SVM. Acta Electron. Sin. **44**(12), 3064–3071 (2016). (In Chinese)
42. Zhou, X., Liu, J., Shao, P., Luo, F., Liu, Y.: Chinese anaphora resolution based on multi-pass sieve model. J. Jilin Univ. (Eng. Technol. Ed.) **46**(4), 1209–1215 (2016). (In Chinese)
43. Wang, C.S., Ngai, G.: A clustering approach for unsupervised Chinese coreference resolution. In: Proceedings of the 5th SIGHAN Workshop on Chinese Language Processing, pp. 40–47. ACL, Sydney, Australia (2006)
44. Zhou, J., Huang, S., Chen, J., Qu, W.: A new graph clustering algorithm for Chinese noun phrase coreference resolution. J. Chin. Inf. Process. **21**(2), 77–82 (2007). (In Chinese)
45. Li, Y., Zhou, J., Chen, J.: Applying correlation clustering to Chinese noun phrase coreference resolution. Comput. Sci. **34**(12), 216–218 (2007). (In Chinese)

46. Li, S., Zhao, T., Chen, C., Liu, P.: An unsupervised approach based on ART network for coreference resolution of Chinese. High-tech Commun. **19**(9), 926–932 (2009). (In Chinese)
47. Gao, J., Kong, F., Zhu, Q., Li, P., Hua, X.: Research of unsupervised Chinese noun phrase coreference resolution. Comput. Eng. **38**(17), 189–191 (2012). (In Chinese)
48. Fu, J., Kong, F.: Coreference resolution incorporating structural information. Comput. Sci. **47**(3), 231–236 (2020). (In Chinese)
49. Fu, J., Kong, F.: End to end Chinese coreference resolution with structural information. Comput. Eng. **46**(1), 45–51 (2020). (In Chinese)