2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

METACoref: A Coreference Resolution Approach Based on Meta-information Loss for Document

Ying Mao[1, 2][0009-0006-9649-9898], Yong Peng[1, 2][0009-0006-1316-6642], and Yong Zhong[1, 2][0009-0000-8519-3873]

[1] Chengdu Institute of Computer Application, Chinese Academy of Sciences, Chengdu 610041, China
[2] University of Chinese Academy of Sciences, Beijing 100049, China

{maoying19,pengyong22}@mails.ucas.ac.cn
zhongyong@casit.com.cn

Abstract. Coreference resolution is a key technique in natural language processing, aiming at recognizing different representations pointing to the same entity in a text. However, in order to improve performance, existing methods rely on single semantic features for complex representation and iterative operation on one hand, and introduce multiple complex structures and external knowledge on the other hand, which sacrifices efficiency and generalization performance to a certain extent. Therefore, this study explores the textual meta-information and proposes a meta-information loss-based coreference resolution model, METACoref, which optimizes the task at two levels, i.e., mention recognition and coreference prediction. METACoref first enriches word representation by syntactic information and entity types, and then obtains subword-based word representation based on local and masked attention mechanisms. In mention recognition, METACoref integrates entity type features, speaker features, and belonging sentence position features to compensate for the lack of pure semantic modeling. In coreference prediction, METACoref uses a combination of dynamically balanced semantic loss and structured meta-information loss to complement semantic information. Structured meta-information loss computes a representation of the consistency of speaker information between mentions, and relative distance between mentions. Experiments on the OntoNotes 5.0 dataset show that the method performs superiorly in mention identification and coreference prediction, significantly improving performance of coreference resolution model in terms of efficiency, robustness, and long-distance dependency handling.

Keywords: Mention Identification, Coreference Prediction, Coreference Resolution, Meta-information.

1 Introduction

Coreference resolution [1,2] plays a pivotal role in the field of natural language processing. This technique is widely used in text summarization, machine translation and other fields. Existing end-to-end coreference resolution models generally introduce complex feature representations and structures in order to improve their performance, resulting in complex model architecture, increased computation, and poor generalization and adaptation to new domains and datasets, which sacrifices efficiency and generalization performance to a certain extent. Existing coreference resolution models evolve in two directions. 1. Higher-order semantic representations that rely on a single semantic feature for complex representations and multiple iterative operations. Such as, [3,4] is a class of models that augment the semantic features through interaction or iterative computation. [4-6] is a class of models based on higher-order reasoning. 2. Introducing a variety of complex structures for modeling the task of coreference resolution, including graph structures and external knowledge dependencies, but there are limitations. For example, [7] is based on Graph Convolutional Networks [8] constrains intra-cluster consistency via ternary transmissibility but does not account for candidate mentions that are too far away, and the memory requirement grows linearly with the length of the document. [9] combines graph attention mechanism (GAT) [10] with second-order inference to aggregate multi-hop neighbor information via GAT, alleviating the localization limitation of traditional GCN. [11] combined with Semantic Role Labeling (RL) to filter candidate mentions, but requires additional preprocessing steps and is difficult to adapt to real-time scenarios. [12] introduces external knowledge to assist the representation of mentions, enriches the mention representation, and enhances the coreference representation, but increases the computational complexity and may introduce noise.

In order to improve the generalizability of the coreference resolution model, researchers have proposed some lightweight methods to simplify the model architecture [13,14]. On the other hand, there are some limitations to the utilization of meta-information, and these models tend to utilize meta-information only statically or locally without fully exploiting its potential, like, models [15-17].

Mining deeply into the text's self-contained meta-information features, and fully utilizing these features can improve the model's generalization and adaptability without having to rely on external structures and features, while avoiding performance degradation caused by simplifying the model architecture. Based on these issues, we propose METACoref, a meta-information loss-based coreference resolution model, which integrates meta-information features from multiple aspects of word representation, mention recognition and coreference prediction for simple representation and computation to compensate for the shortcomings of pure semantic models. The main contributions of this paper can be summarized as follows:

1. We propose a coreference resolution model that integrates token-based and word-based mention representation. The word-level representation is enhanced by introducing syntactic and entity type labels, and combined with the implicit generation mechanism of model S2E [18], it effectively reduces the need for explicit memory.
2. We propose several ready-to-use components for mention prediction. Meta-information such as start/stop position, speaker, sentence position and entity type of candidate mention are integrated to enhance mention recognition.

3. We optimize the coreference clustering prediction that enhances the accuracy of coreference prediction by modeling meta-information such as speaker congruence and relative distance between mentions.
4. We propose a dynamically balanced way of fusing semantic information and structured meta-information.
5. We evaluate our method on the OntoNotes 5.0 dataset and demonstrate that it is an effective approach for coreference resolution.

2 Related Work

2.1 End-To-End Model

E2E [19] encodes text into contextualized vectors, and then based on the specific to the task, these vectors are represented by mentions and scored in pairs. Many subsequent papers are based on the fundamental idea of E2E and improve it, and the main direction of improvement is the span representation of mentions. Improvements on span representation include, one is based on token level [18,20,21], and the other is based on word level [13,22,23].

Token-based mention representation. The token level based one mainly uses vectors including, but not limited to, the start marker, end marker of the span for the mention representation. Where [20] uses the embedding of the first token, the embedding of the last token, and the weighted combination of the embeddings of all the tokens in the mention to obtain the mention representation. [21] uses tandem representation of the first and last token embeddings of the span, as well as the header embeddings of all the tokens, and the representation of the first and last tokens are augmented using morphological information. The main idea of S2E [18] uses the span of the first token and the last token to indicate the mention.

Word-based mention representation. Based on word level, the coreference relationship between individual words is considered, and subsequently the mention representation is constructed. Among them, wl-coref [13] constructs word representation based on subwords, and subsequently constructs mentions based on word expansion. CAW-coref [22] optimizes the processing of mentions containing conjunctions on this basis. MSCAW-coref [23] further extends the support for non-English languages. However, these models typically have high memory requirements, requiring at least 48G of training memory.

In this study, we propose an innovative coreference resolution model, METACoref, which integrates token-based and word-based mention representation. Specifically, we integrate the subword-based word representation method in the wl-coref [13] with the implicit span representation strategy in the S2E [18]. Through this combination, we are able to effectively avoid unreasonable candidate mentions due to improper subwording, thus reducing error propagation. In addition, the method reduces the demand for memory by shortening the input length and avoiding explicit span representation.

3 Our Model

In this paper, we propose a meta-information loss-based coreference resolution model METACoref. METACoref is optimized in terms of word representation, mention recognition, and coreference prediction. In module 3.1, for word representation, syntactic type tagging and entity type tagging are combined to enrich word representation, and subword-based word representation is obtained using local attention and masked attention mechanisms. In module 3.2, for mention recognition, METACoref integrates entity type features, speaker features, and sentence position features to compensate for the lack of purely semantic models, and to improve the model's learning ability and generalizability. Next, in module 3.3, for coreference prediction, METACoref uses a combination of dynamically balanced semantic loss and structured meta-information loss to supplement the semantic information. The architecture diagram for the model is presented in **Fig. 1**.

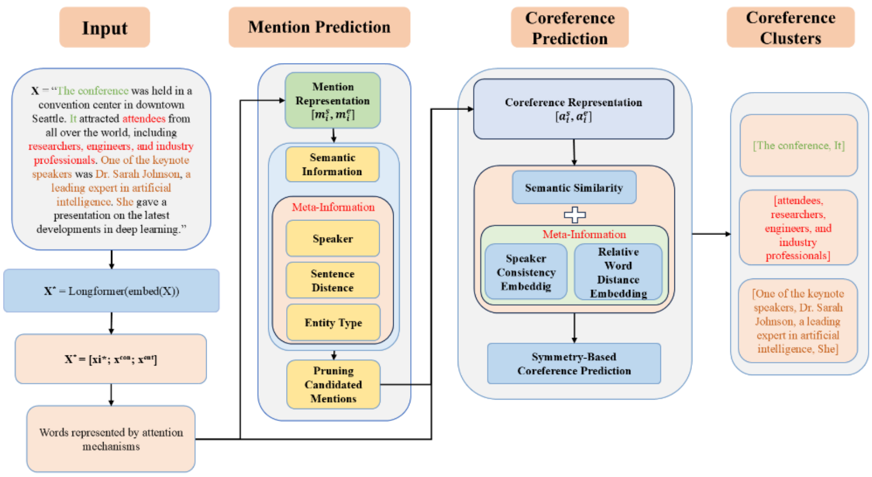

Fig. 1. The overall architecture of our METACoref model.

3.1 Input

The original input of the document is $\mathbf{D} = \{\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_W\}$, and after tokenizing the document is $\mathbf{V} = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N\}$, with the length of the original input being W, and the length of the tokenized input being N. The syntactic constraint type of $\mathbf{t}_i$ is denoted as $\mathbf{x}_i^{con}$. The entity type constraint for the word $\mathbf{t}_i$ is denoted as $\mathbf{x}_i^{ent}$. The speaker information for the word $\mathbf{t}_i$ is denoted as $\mathbf{x}_i^{speaker}$. The number of the sentence to which the word $\mathbf{t}_i$ belongs is denoted as $\mathbf{x}_i^{dsent}$.

Firstly, we encode the input $\mathbf{V}$ according to the Eq. 1 [14]:

$$\mathbf{X} = Encoder(\mathbf{V}) \tag{1}$$

Next, we use Eq. 2 [24] to add syntactic type constraints $\mathbf{x}_i^{con}$, and entity type information $\mathbf{x}_i^{ent}$ to the embedding vector $\mathbf{X}$:

$$\mathbf{X}' = [\mathbf{x}_i; \mathbf{x}_i^{con}; \mathbf{x}_i^{ent}] \tag{2}$$

Finally, based on the local and masked attention mechanism Eq. 3 [13], a representation of the complete word is obtained, where $\mathbf{X}' \in \mathbb{R}^{N\times M}$, ($N$ denotes the text number of sub-tokens in the input, and M is the vector dimension of each sub-token). **mention_attn** denotes the linear layer used to compute the attention score for each subword, yielding a tensor of $\mathbb{R}^{W\times N}$. The $\mathbf{attn\_mask} \in \mathbb{R}^{W\times N}$ denotes the use of masked attention to restrict each word to only its corresponding range of subwords, which is used to label the range of subwords for each word:

$$\mathbf{X}^* = \mathrm{softmax}(\log\,(\mathbf{attn\_mask}) + \mathbf{mention\_attn}(\mathbf{X}')^T, \mathrm{dim} = 1) \tag{3}$$

3.2 Mention Identification

Mention Representation. Based on the Eq. 4 [14], we obtain the mention representation $[\mathrm{m}_i^s, \mathrm{m}_i^e]$ for the mention m_i. $\mathrm{m}_i^s, \mathrm{m}_i^e$ denote the beginning and ending words of the representation of $\mathbf{m}_i$ respectively.

$$\mathbf{m}_i^s = \mathbf{GeLU}(\mathbf{W}_m^s \mathbf{x}_i^*) \qquad \mathbf{m}_i^e = \mathbf{GeLU}(\mathbf{W}_m^e \mathbf{x}_i^*) \tag{4}$$

Mention Prediction. Based on the Eq. 5, Eq. 6, Eq. 7 the non-linear representation of sentence position information, speaker information and entity type information of each word is presented.

$$\mathbf{m}_i^{ent} = \mathbf{GeLU}(\mathbf{W}_{ent}\mathbf{x}_i^{ent}) \tag{5}$$

$$\mathbf{m}_i^{speaker} = \mathbf{GeLU}(\mathbf{W}_{speaker}\mathbf{x}_i^{speaker}) \tag{6}$$

$$\mathbf{m}_i^{dsent} = \mathbf{GeLU}(\mathbf{W}_{dsent}\mathbf{x}_i^{dsent}) \tag{7}$$

Here, $\mathbf{GeLU}$ is an activation function, and the matrices $\mathbf{W}_{ent}$, $\mathbf{W}_{speaker}$, $\mathbf{W}_{dsent}$ are the trainable parameters for the entity-type representation, speaker-information representation, and sentence-location-information representation, respectively.

Next, a candidate mention score is calculated by adding three components based on the above feature representation. through Eq. 8, Eq. 9, Eq. 10.

$$f_{speaker}(m) = \mathbf{m}_s^{speaker} \cdot \mathbf{S}_m \cdot \mathbf{m}_e^{speaker} \tag{8}$$

$$f_{dsent}(m) = \mathbf{m}_s^{dsent} \cdot \mathbf{D}_m \cdot \mathbf{m}_e^{dsent} \tag{9}$$

$$f_{ent}(m) = \mathbf{m}_s^{ent} \cdot \mathbf{E}_m \cdot \mathbf{m}_e^{ent} \tag{10}$$

The matrices $\mathbf{S}_m$, $\mathbf{D}_m$, $\mathbf{E}_m$ are the trainable parameters of the mention scoring functions $f_{speaker}(m)$, $f_{dsent}(m)$, and $f_{ent}(m)$, respectively.

Then based on the Eq. 11 [14], the final mention calculation Eq. 12 is obtained:

$$f_m(m) = \mathbf{v}_s \cdot \mathbf{m}_{p_s}^s + \mathbf{v}_e \cdot \mathbf{m}_{p_e}^e + \mathbf{m}_{p_s}^s \cdot \mathbf{H}_m \cdot \mathbf{m}_{p_e}^e \tag{11}$$

$$f(m) = f_m(m) + f_{speaker}(m) + f_{dsent}(m) + f_{ent}(m) \tag{12}$$

The vectors $\mathbf{v}_s$, $\mathbf{v}_e$ and the matrix $\mathbf{H}_m$ are trainable parameters for our mention scoring function $f_m(m)$.

Candidate Mentions. To generate the final list of candidate mentions, based on memory considerations, the top $\alpha * n$ ranked candidate mentions, denoted as M, are obtained based on the computer results of $f_m(m)$, and n is the input text length. Here, α denotes a predetermined hyperparameter.

3.3 Coreference Prediction

Coreference Representation. Firstly, based on the Eq. 13 [14], we use X* to obtain the coreference representation $[\mathbf{a}_i^s, \mathbf{a}_i^e]$ of the coreference $\mathbf{a}_i$. Then, we add $\mathbf{m}_i^{ent}$ to the $\mathbf{a}_i$ using Eq. 14 [25], to refine the coreference representation. $\mathbf{m}_i^{ent}$ is calculated through Eq. 5.

$$\mathbf{a}_i^s = \mathbf{GeLU}(\mathbf{W}_a^s \mathbf{x}_i^*) \qquad \mathbf{a}_i^e = \mathbf{GeLU}(\mathbf{W}_a^e \mathbf{x}_i^*) \tag{13}$$

$$\mathbf{a}_i^s = u * \mathbf{a}_i^s + (1-u) * (\mathbf{a}_i^s + \mathbf{m}_i^{ent}) \qquad \mathbf{a}_i^e = u * \mathbf{a}_i^e + (1-u) * (\mathbf{a}_i^e + \mathbf{m}_i^{ent}) \tag{14}$$

Here $\mathbf{GeLU}$ is an activation function, and the matrices $\mathbf{W}_a^s$, $\mathbf{W}_a^e$ are the trainable parameters of the coreference representation. u denotes a predetermined hyperparameter.

Coreference Prediction Based on Semantic Similarity. Using the Eq. 15 [14], score for mention pointing to the antecedent is calculated based on semantic features.

$$f_a^{sem}(c,q) = \mathbf{a}_{c_s}^s \cdot \mathbf{B}_a^{ss} \cdot \mathbf{a}_{q_s}^s + \mathbf{a}_{c_s}^s \cdot \mathbf{B}_a^{se} \cdot \mathbf{a}_{q_e}^e + \mathbf{a}_{c_e}^e \cdot \mathbf{B}_a^{es} \cdot \mathbf{a}_{q_s}^s + \mathbf{a}_{c_e}^e \cdot \mathbf{B}_a^{ee} \cdot \mathbf{a}_{q_e}^e \tag{15}$$

The matrices $\mathbf{B}_a^{ss}$, $\mathbf{B}_a^{se}$, $\mathbf{B}_a^{es}$, $\mathbf{B}_a^{ee}$ are trainable parameters on computing semantic similarity when mention pointing to the antecedent.

Coreference Prediction Based on Meta-information Loss. Structured meta-information loss computes relative distance, and the consistency of speaker information between mentions. First, the speaker information of candidate mentions is checked for consistency with the same person, and the embedding representation of inter-mention consistency is obtained, represented as $\mathbf{E}_{speaker}$. Next, we compute the relative distances of words between candidate mentions, and convert them to discrete values and through logarithmic compression (log2) and binning, aiming to obtain the embedding representation of the relative distence between mentions, $\mathbf{E}_{dtoken}$. Finally, concatenation of $\mathbf{E}_{speaker}$ and $\mathbf{E}_{dtoken}$ is used to compute the coreference prediction of structured meta-information based on Eq. 16.

$$f_a^{meta}(c,q) = \mathbf{W}_{meta} \cdot [\mathbf{E}_{speaker} \oplus \mathbf{E}_{dtoken}] + \mathbf{b} \tag{16}$$

The matrix $\mathbf{W}_{meta}$ and the bias vector $\mathbf{b}$ are learnable parameters, and $\oplus$ denotes the vector concatenation operation.

Then, the semantic information is complemented by dynamically balancing the semantic loss and the loss of structured meta-information, based on the Eq. 17, where p is a predetermined dynamic balancing parameter.

$$f_a(c,q) = p \cdot f_a^{sem}(c,q) + (1-p) \cdot f_a^{meta}(c,q) \tag{17}$$

Based on the symmetry [24], the score $f_a(q,c)$ for antecedent pointing to mention is calculated, then based on the Eq. 18 the coreference prediction score is obtained.

$$f_a = (f_a(c,q) + f_a(q,c))/2 \tag{18}$$

Final Scores. The final score for coreference resolution is based on Eq. 19.

$$f = f_m(m) + f_m(a) + f_a \tag{19}$$

4 Experiments and Results

4.1 Tasks

We conducted experiments on two tasks, mention identification and coreference prediction. Mention identification involved identifying the boundaries of mention spans to obtain possible mentions, which were then utilized for coreference prediction.

Dataset. We perform training and evaluation on the document-level English dataset OntoNotes 5.0 [26], which consists of 2802, 343 and 348 documents in the training, development and test data sets. The dataset contains news articles and telephone conversations.

Evaluation Metrics. For coreference resolution, we use coreference scorer [27] to evaluate, which is the unweighted average of the F1 scores of MUC [28], B^3 [29] and CEAF_{ϕ_4} [30]. MUC measures the minimum number of links that need to be inserted or deleted to map the predicted coreference cluster to the gold coreference cluster. The drawback is that there is no way to measure the performance of the predicted singleton entity. B^3 mainly calculates P and R for each candidate mention, judges the accuracy of each candidate mention in coreference prediction, and then takes the average of all candidate mentions as the final result. It can overcome the shortcomings of MUC, which focuses on entities. CEAF_{ϕ_4} focuses more on the similarity between entities. It calculates the common mention ratio after the predicted clusters is aligned with the gold clusters to obtain the similarity of entities between the predicted clusters and the gold clusters.

For mention identification, mention identification results are represented by F1 scores. If a mention exactly matches the gold mention on the boundary, the mention is considered to be correctly detected.

Experimental Detail. We utilize the Longformer-Large [31] as our underlying pretraining model, because of its capability to handle lengthy documents without the need for sliding windows or truncation. The learning rate of the AdamW optimizer is set to 1e-5, α is 0.4, epoch is 129, and adding parameters u, p are 0.5, 0.9 respectively.

4.2 Baseline

In this paper, a meta-information loss-based coreference resolution model METACoref is proposed for the first time, and it is compared with the existing models [3,4] based on higher-order semantic representation, the graph-structure-based models [32,33], and the models [13,14] with a simplified model architecture.

G2GT overlap [32]: a coreference resolution method based on graph structure and multi-level iterative refinement is proposed.

SpanBERT [3]: this work uses SpanBERT to better represent and predict the span of text and for coreference resolution tasks.

SpanBERT + CM [4]: analyzes the impact of higher-order inference on coreference resolution. The model is implemented using four HOI methods: based antecedent, entity equalization, mention clustering and cluster merging.

RGAT [33]: proposes a model combining pre-trained BERT and Syntactic Relation Graph Attention Network (RGAT).

S2E [14]: a lightweight end-to-end coreference resolution model is introduced, which eliminates the reliance on spanning representations, manual features and heuristics.

wl-coref [13]: the model considers associations between individual words rather than relations between spans, reconstructing each word based on subwords. And allows to consider all possible mentions for coreference clustering prediction without pruning.

4.3 Results

Table 1. Coreference resolution performance of different models on the test set of English OntoNotes 5.0 dataset.

Model	MUC			B^3			CEAF_{ϕ_4}			CoNLL
	P	R	F1	P	R	F1	P	R	F1	F1
SpanBert	85.8	84.8	85.3	78.3	77.9	78.1	76.4	74.2	75.3	79.6
SpanBERT+CM	85.9	85.5	85.7	79.0	78.9	79.0	76.7	75.2	75.9	80.2
RGAT	85.7	82.5	84.3	77.6	73.9	76.5	75.0	70.0	72.8	77.7
G2GT overlap	85.8	84.9	85.3	78.7	78.0	78.3	76.4	74.5	75.4	79.7
S2E	86.5	85.1	85.8	80.3	77.9	79.1	76.8	75.4	76.1	80.3
wl-coref	84.6	**87.8**	**86.1**	77.0	**82.2**	**79.5**	75.6	**76.8**	**76.2**	**80.6**
METACoref	**86.6**	84.6	85.6	**80.4**	77.8	79.1	**77.5**	74.8	**76.2**	**80.6**

The results of the experimental comparison between the baseline model and METACoref are shown in **Table 1**. METACoref is based on the word representation of wl-coref and the implicit spanning representation of S2E, with the addition of the loss of structured meta-information, such as speaker information, for mention recognition and coreference prediction. Therefore, the focus is on the comparison of the performance of METACoref, wl-coref, and S2E. Compared with S2E, the precision (P) is elevated regardless of MUC, B^3 or CEAF_{ϕ_4}, indicating that METACoref helps to improve the consistency between mentions within clusters. Compared to wl-coref, the precision (P) is elevated on MUC, B^3 and CEAF_{ϕ_4}, but METACoref is worse on recall (R), which is analyzed based on the fact that wl-coref does not perform pruning. However, compared to wl-coref, METACoref has half requirement of the GPU memory, and the effect of coreference resolution is not reduced. Compared to other coreference resolution models based on high-level semantic representations and graph structures, the accuracy is substantially improved.

4.4 Ablation Experiments

Table 2. Ablation experiments on the test set of the English OntoNotes 5.0 dataset performance on two tasks.

Model	Mention identification			Coreference Prediction		
	P	R	F1	P	R	F1
SP	88.6	79.5	83.8	79.7	70.7	74.9
DP	89.1	87.0	88.1	80.5	78.4	79.4
SP+DP	89.8	86.5	88.1	81.7	78.2	79.9
SP+DP+EP	**89.9**	87.1	**88.5**	**81.8**	78.9	80.4
SP+DP+EP+MP	89.3	**87.5**	88.4	81.6	**79.7**	**80.6**

We conducted ablation experiments to evaluate the impact of each novel component within the METACoref model on the coreference resolution task, and the results are shown in **Table 2**. Specifically, we investigated the speaker information component (SP) in equation Eq. 8, the sentence position information component (DP) in equation Eq. 9, the entity type information component (EP) in equation Eq. 10, and equation Eq. 16 of structured meta-information loss (MP).

The experimental results show that “SP”, “DP”, “EP” and “MP” have different degrees of effect on the between-tasks models with different degrees of contribution to the efficiency improvement between the two tasks. Specifically, the addition of “SP”, “DP”, and “EP” improved the precision (P), recall (R), and F1 values of mention identification, reduced the prediction of invalid mentions, and thereby improving the effectiveness of coreference prediction. “MP” optimizes the model in terms of speaker information consistency and long-tailed distribution optimization, improves the recall (R), makes the overlap between labeled clusters and predicted clusters higher, and improves the results of coreference prediction.

5 Conclusion

This paper proposes a lightweight end-to-end coreference resolution model METACoref based on meta-information loss. Subword-based attention mechanism to reduce text input length and invalid mentions generation. Mention prediction is optimized by designing several ready-to-use components of speaker information, sentence location information, and entity type information to improve the accuracy of candidate mentions. The meta-information loss is constructed using the illuminating strategies “speaker consistency” and “long-tailed distribution optimization”, and the semantic similarity between mentions and meta-information loss are combined in a dynamic balancing approach to improve the overlap between labeled clusters and predicted clusters and optimize the coreference resolution model. In order to build a generalized model, METACoref deeply exploits the meta-information features of text and expresses and utilizes them, which can improve the generalization and adaptability of the model without relying on external structures and features, and avoid the performance degradation due to simplifying the model architecture.

References

1. Kevin Clark and Christopher D. Manning. Improving coreference resolution by learning entity-level distributed representations. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 643–653, Berlin, Germany, August 2016. Association for Computational Linguistics.
2. Pradheep Elango. Coreference resolution: A survey. University of Wisconsin, Madison, WI, page 12, 2005.
3. Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S Weld, Luke Zettlemoyer, and Omer Levy. Spanbert: Improving pre-training by representing and predicting spans. Transactions of the Association for Computational Linguistics, 8:64–77, 2020.
4. Liyan Xu and Jinho D. Choi. Revealing the myth of higher-order inference in coreference resolution. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 8527–8533, Online, November 2020. Association for Computational Linguistics.
5. Kenton Lee, Luheng He, and Luke Zettlemoyer. Higher-order coreference resolution with coarse-to-fine inference. In Marilyn Walker, Heng Ji, and Amanda Stent, editors, Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), pages 687–692, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.
6. Ben Kantor and Amir Globerson. Coreference resolution with entity equalization. In Anna Korhonen, David Traum, and Lluís Màrquez, editors, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 673–677, Florence, Italy, July 2019. Association for Computational Linguistics.
7. Yuanliang Meng and Anna Rumshisky. Triad-based neural network for coreference resolution. In Emily M. Bender, Leon Derczynski, and Pierre Isabelle, editors, Proceedings of the 27th International Conference on Computational Linguistics, pages 35–43, Santa Fe, New Mexico, USA, August 2018. Association for Computational Linguistics.
8. Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907, 2016.
9. Lu Liu, Zhenqiao Song, and Xiaoqing Zheng. Improving coreference resolution by leveraging entity-centric features with graph neural networks and second-order inference. ArXiv, abs/2009.04639, 2020.
10. Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. arXiv preprint arXiv:1710.10903, 2017.

11. Hongming Zhang, Yan Song, and Yangqiu Song. Incorporating context and external knowledge for pronoun coreference resolution. arXiv preprint arXiv:1905.10238, 2019.
12. Hongming Zhang, Yan Song, Yangqiu Song, and Dong Yu. Knowledge-aware pronoun coreference resolution. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 867–876, Florence, Italy, July 2019. Association for Computational Linguistics.
13. Vladimir Dobrovolskii. Word-level coreference resolution. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pages 7670–7675, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
14. Yuval Kirstain, Ori Ram, and Omer Levy. Coreference resolution without span representations. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pages 14–19, 2021.
15. Kong Fang and Fu Jian. Incorporating structural information for better coreference resolution. In Twenty-Eighth International Joint Conference on Artificial Intelligence IJCAI-19, 2019.
16. Michelle Yuan, Patrick Xia, Chandler May, Benjamin Van Durme, and Jordan Boyd-Graber. Adapting coreference resolution models through active learning. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 7533–7549, Dublin, Ireland, May 2022. Association for Computational Linguistics.
17. Klim Zaporojets, Johannes Deleu, Yiwei Jiang, Thomas Demeester, and Chris Develder. Towards consistent document-level entity linking: Joint models for entity linking and coreference resolution. arXiv preprint arXiv:2108.13530, 2021.
18. Yuval Kirstain, Ori Ram, and Omer Levy. Coreference resolution without span representations. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), pages 14–19, Online, August 2021. Association for Computational Linguistics.
19. Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. End-to-end neural coreference resolution. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel, editors, Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 188–197, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.
20. Şeniz Demir. Neural coreference resolution for turkish. Journal of Intelligent Systems: Theory and Applications, 6(1):85–95, 2023.
21. Tuğba Pamay Arslan, Kutay Acar, and Gülşen Eryiğit. Neural end-to-end coreference resolution using morphological information. In Zdeněk Žabokrtský and Maciej Ogrodniczuk, editors, Proceedings of the CRAC 2023 Shared Task on Multilingual Coreference Resolution, pages 34–40, Singapore, December 2023. Association for Computational Linguistics.
22. Karel D'Oosterlinck, Semere Kiros Bitew, Brandon Papineau, Christopher Potts, Thomas Demeester, and Chris Develder. CAW-coref: Conjunction-aware wordlevel coreference resolution. In Maciej Ogrodniczuk, Vincent Ng, Sameer Pradhan,

and Massimo Poesio, editors, Proceedings of The Sixth Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC 2023), pages 8–14, Singapore, December 2023. Association for Computational Linguistics.
23. Houjun Liu, John Bauer, Karel D'Oosterlinck, Christopher Potts, and Christopher D. Manning. MSCAW-coref: Multilingual, singleton and conjunction-aware word-level coreference resolution. In Maciej Ogrodniczuk, Anna Nedoluzhko, Massimo Poesio, Sameer Pradhan, and Vincent Ng, editors, Proceedings of The Seventh Workshop on Computational Models of Reference, Anaphora and Coreference, pages 33–40, Miami, November 2024. Association for Computational Linguistics.
24. Ying Mao, Xinran Xie, Lishun Wang, Zheyu Shi, and Yong Zhong. Symconll: A symmetry-based approach for document coreference resolution. In Fei Liu, Nan Duan, Qingting Xu, and Yu Hong, editors, Natural Language Processing and Chinese Computing, pages 286–297, Cham, 2023. Springer Nature Switzerland.
25. Ying Mao, Xinran Xie, Yong Peng, and Yong Zhong. Cmcoref: A constraintbased approach for document coreference resolution. In De-Shuang Huang, Yijie Pan, and Jiayang Guo, editors, Advanced Intelligent Computing Technology and Applications, pages 478–492, Singapore, 2024. Springer Nature Singapore.
26. Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Hwee Tou Ng, Anders Björkelund, Olga Uryupina, Yuchen Zhang, and Zhi Zhong. Towards robust lin guistic analysis using ontonotes. In Proceedings of the Seventeenth Conference on Computational Natural Language Learning, pages 143–152, 2013.
27. Sameer Pradhan, Xiaoqiang Luo, Marta Recasens, Eduard Hovy, Vincent Ng, and Michael Strube. Scoring coreference partitions of predicted mentions: A reference implementation. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 30–35, Baltimore, Maryland, June 2014. Association for Computational Linguistics.
28. Marc Vilain, John D Burger, John Aberdeen, Dennis Connolly, and Lynette Hirschman. A model-theoretic coreference scoring scheme. In Sixth Message Understanding Conference (MUC-6): Proceedings of a Conference Held in Columbia, Maryland, November 6-8, 1995, 1995.
29. Amit Bagga and Breck Baldwin. Algorithms for scoring coreference chains. In The first international conference on language resources and evaluation workshop on linguistics coreference, volume 1, pages 563–566. Citeseer, 1998.
30. Xiaoqiang Luo. On coreference resolution performance metrics. In Proceedings of human language technology conference and conference on empirical methods in natural language processing, pages 25–32, 2005.
31. Iz Beltagy, Matthew E Peters, and Arman Cohan. Longformer: The long-document transformer. arXiv preprint arXiv:2004.05150, 2020.
32. Lesly Miculicich and James Henderson. Graph refinement for coreference resolution. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, Findings of the Association for Computational Linguistics: ACL 2022, pages 2732–2742, Dublin, Ireland, May 2022. Association for Computational Linguistics.
33. Yuan Meng, Xuhao Pan, Jun Chang, and Yue Wang. Rgat: A deeper look into syntactic dependency information for coreference resolution. 2023 International Joint Conference on Neural Networks (IJCNN), pages 1–8, 2023.