

Short Text Event Coreference Resolution

Subjects: Computer Science, Interdisciplinary Applications

Contributor: Xinyou Yong, Chongqing Zeng, Lican Dai, Wanli Liu, Shimin Cai

Event coreference resolution is the task of clustering event mentions that refer to the same entity or situation in text and performing operations like linking, information completion, and validation. Existing methods model this task as a text similarity problem, focusing solely on semantic information, neglecting key features like event trigger words and subject.

Keywords: coreference resolution ; relationship prediction ; universal information extraction

1. Introduction

Event coreference resolution (ECR) is a crucial natural language processing task that involves identifying and clustering together different textual mentions of the same event. The task holds significant importance in enabling a variety of downstream applications, including information extraction, question answering, text summarization, etc. [1][2]. For example, in the application of information extraction, ECR can help to build more coherent and complete knowledge graphs or databases by linking different mentions of the same event and correctly identifying and linking relevant information to answer questions accurately [3]. **Figure 1** shows an example of the ECR task, where the input consists of two different segments of text and the output is the binary confidence of event coreference.

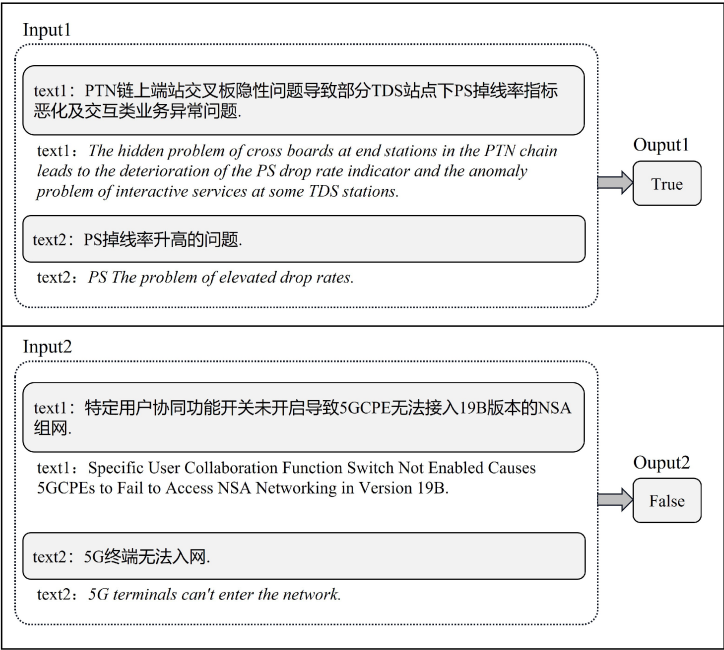


Figure 1. Example illustration of ECR task.

ECR is also an important part of constructing event graphs [4]. The entire process is similar to entity linking in knowledge graphs, where identical event nodes in the real world are clustered together, further improving and supplementing the various components of the event and saving them as a new node in the graph structure. Some researchers have modeled the ECR task as a text similarity calculation problem, using neural networks such as CNN (Convolutional Neural Network) [5] and RNN (Recurrent Neural Network) [6] to represent two pieces of text as vectors and determining whether they are coreferential by calculating the similarity between the vectors. With the emergence of pre-training models such as BERT [7], some researchers have used the output of pre-training models as the vector representation of text and calculated similarity. Others have used Siamese networks [8] to classify event pairs and determine if they are coreferential.

However, training neural networks such as CNN and RNN to represent text as vectors requires training parameters from scratch and is not suitable for small datasets [9]. In contrast, pre-training models only need to be fine-tuned after being

pre-trained on a large corpus, which can achieve faster convergence speed when combined with downstream tasks. Additionally, determining event coreference based on whether the similarity reaches a threshold is biased, as there may not be a clear classification boundary for this task, making it difficult to determine the threshold, or there may not be a clear boundary at all. Using BERT for text representation or Siamese networks for text classification also fails to fully utilize the inherent next sentence prediction (NSP). Furthermore, the above techniques only focus on the semantic information of the event description text, ignoring a variety of key features such as event trigger words and event subjects, resulting in missing features.

To address the limitations of existing techniques, researchers propose a short text event coreference resolution method based on context prediction (referred to ECR-CP). The novelty of ECR-CP is illustrated from three perspectives. In more detail, from a problem-modeling perspective, researchers model the ECR task as a sentence-level relationship prediction issue by utilizing the NSP inherent in BERT. Researchers consider pairs of events that can form a continuous sentence-level relationship to have coreferential relationships. This is consistent with human language habits, where coherent sentences in everyday conversations often describe the same fact. From a feature extraction perspective, researchers extract key information such as trigger words, argument roles, event types, and tense from event extraction and incorporate them as auxiliary features to improve the accuracy. From the algorithm performance perspective, BERT-based ECR-CP has a smaller training cost and achieves better performance in comparison to the other benchmark methods based on neural networks.

2. Short Text Event Coreference Resolution

ECR is one of the key subtasks in event extraction and fusion. However, early coreference resolution tasks mainly focused on the entity level. In the domain of entity coreference resolution, Kejriwal et al. ^[10] proposed an unsupervised algorithm pipeline for learning Disjunctive Normal Form (DNF) blocking schemes on Knowledge Graphs (KGs), as well as structurally heterogeneous tables that may not share a common schema. This approach aims to address entity resolution problems by mapping entities to blocks. Additionally, Šteflovíč et al. ^[11] aim to enhance classifier performance metrics by incorporating the results of entity coreference analysis into the data preparation process for classification tasks.

Due to the complexity of events themselves, as well as a lack of relevant language resources, research on ECR both domestically and abroad started relatively late and has developed more slowly than event extraction techniques. The ACE2005 corpus ^[12] was the first to add coreference attributes to event information, and in 2015, the Knowledge Base Population (KBP) ^[13] began related evaluation tasks, laying the foundation for subsequent research.

Early research on ECR was mainly based on rule-based methods ^{[14][15][16][17][18]}, but with the widespread application of machine learning methods. Recent research has mainly used traditional machine learning and neural network methods to complete ECR tasks ^[1]. A binary classifier is trained to determine whether two pieces of text refer to the same real-world event, and then coreferential events are optimized by clustering ^[19]. The differences in various methods mainly lie in the problem modeling, text feature extraction methods, and the internal details of the network model. Depending on whether manual annotation of language resources is required, ECR methods can be divided into three categories: supervised, semi-supervised, and unsupervised learning.

Among them, supervised learning is the earliest and most widely used research approach. For example, Fang et al. ^[20] designed a multi-layer CNN to extract event features, obtained deep semantic information, and further improved the performance of the coreference resolution algorithm by using multiple attention mechanisms. Dai et al. ^[21] enhanced the representation of event text features using a Siamese network framework and used the Circle Loss loss function to maximize intra-class event similarity and minimize inter-class event similarity. Liu et al. ^[22] trained a support vector machine (SVM) classifier based on more than 100 event features to determine whether the events refer to the same entity. Another attempt is made to utilize large amounts of out-of-domain text data.

Due to the high cost of manual annotation, many scholars have attempted to use semi-supervised methods to study ECR tasks. The main idea is to use a small amount of labeled data to construct a learning algorithm to learn data distribution and features, thus completing the labeling of unlabeled samples. For example, Sachan et al. ^[23] achieved good results on a small-scale training dataset by using active learning to select information-rich instances. Similarly, Chen et al. ^[24] employed active learning to select informative instances, indicating that only a small number of training sentences need to be annotated to achieve state-of-the-art performance in event coreference. Another attempt is made to utilize large amounts of out-of-domain text data ^[25].

Unsupervised learning methods completely eliminate the dependence on labeled data and are often probabilistic generative models. For example, Bejan et al. [26] constructed a generative, parameter-free Bayesian model based on hierarchical Dirichlet processes and infinite factorial hidden Markov models to achieve unsupervised ECR task learning. Chen et al. [27] addressed the relatively scarce research on an unsupervised Chinese event coreference resolution task by proposing a generative model. When evaluated on the ACE 2005 corpus, the performance of this model was comparable to that of supervised tasks.

In addition, the differences between various methods also lie in the scoring processing steps of coreference relationships, which can be divided into two models: event-pair models and event-ranking models [28][29]. The event-pair model [30][31][32] is a binary classification model that independently determines whether each event pair refers to the same entity and then aggregates events that refer to each other to form coreferential event clusters. The event-ranking model judges the current event's coreference with all other candidate events and obtains a ranking result based on the degree of coreference, then divides the events into clusters according to a set threshold λ .

References

1. Liu, R.; Mao, R.; Luu, A.T.; Cambria, E. A brief survey on recent advances in coreference resolution. *Artif. Intell. Rev.* 2023, 56, 14439–14481.
2. Chen, L.C.; Chang, K.H. An Extended AHP-Based Corpus Assessment Approach for Handling Keyword Ranking of NLP: An Example of COVID-19 Corpus Data. *Axioms* 2023, 12, 740.
3. Yang, Y.; Wu, Z.; Yang, Y.; Lian, S.; Guo, F.; Wang, Z. A survey of information extraction based on deep learning. *Appl. Sci.* 2022, 12, 9691.
4. Hu, Z.; Jin, X.; Chen, J.; Huang, G. Construction, reasoning and applications of event graphs. *Big Data Res.* 2021, 7, 80–96.
5. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* 1998, 86, 2278–2324.
6. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D Nonlinear Phenom.* 2020, 404, 132306.
7. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* 2018, arXiv:1810.04805.
8. Chopra, S.; Hadsell, R.; LeCun, Y. Learning a similarity metric discriminatively, with application to face verification. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, San Diego, CA, USA, 20–26 June 2005; Volume 1, pp. 539–546.
9. Mohammed, A.; Kora, R. A comprehensive review on ensemble deep learning: Opportunities and challenges. *J. King Saud-Univ.-Comput. Inf. Sci.* 2023, 35, 757–774.
10. Šteflovíč, K.; Kapusta, J. Coreference Resolution for Improving Performance Measures of Classification Tasks. *Appl. Sci.* 2023, 13, 9272.
11. Kejriwal, M. Unsupervised DNF Blocking for Efficient Linking of Knowledge Graphs and Tables. *Information* 2021, 12, 134.
12. Walker, C.; Strassel, S.; Medero, J.; Maeda, K. ACE 2005 multilingual training corpus. *Phila. Linguist. Data Consort.* 2006, 57, 45.
13. Mitamura, T.; Liu, Z.; Hovy, E.H. Overview of TAC KBP 2015 Event Nugget Track. In *Proceedings of the TAC*, Gaithersburg, MR, USA, 12–13 November 2015.
14. Hobbs, J.R. Resolving pronoun references. *Lingua* 1978, 44, 311–338.
15. Humphreys, K.; Gaizauskas, R.; Azzam, S. Event coreference for information extraction. In *Proceedings of the Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, Madrid, Spain, 11 July 1997.
16. Glavaš, G.; Šnajder, J. Exploring coreference uncertainty of generically extracted event mentions. In *Proceedings of the International Conference on Intelligent Text Processing and Computational Linguistics*; Springer: Cham, Switzerland, 2013; pp. 408–422.
17. Li, L.; Jin, L.; Jiang, Z.; Zhang, J.; Huang, D. Coreference resolution in biomedical texts. In *Proceedings of the 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Belfast, UK, 2–5 November 2014; pp. 12–14.

18. Lample, G.; Ballesteros, M.; Subramanian, S.; Kawakami, K.; Dyer, C. Neural architectures for named entity recognition. *arXiv* 2016, arXiv:1603.01360.
19. Prasad, K.V.; Vaidya, H.; Rajashekhar, C.; Karekal, K.S.; Sali, R. Automated neural network forecast of PM concentration. *Int. J. Math. Comput. Eng.* 2023, 1, 67–78.
20. Fang, J.; Li, P.; Zhu, Q. Employing Multi-attention Mechanism to Resolve Event Coreference. *Comput. Sci.* 2019, 46, 277–281.
21. Dai, B.; Qian, J.; Cheng, S.; Qiao, L.; Li, D. Event Coreference Resolution based on Convolutional Siamese network and Circle Loss. In *Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN)*, Padua, Italy, 18–23 July 2022; pp. 1–7.
22. Liu, Z.; Araki, J.; Hovy, E.H.; Mitamura, T. Supervised Within-Document Event Coreference using Information Propagation. In *Proceedings of the 2014 the Language Resources and Evaluation Conference*, Reykjavik, Iceland, 26–31 May 2014; pp. 4539–4544.
23. Sachan, M.; Hovy, E.; Xing, E.P. An active learning approach to coreference resolution. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, Buenos Aires, Argentina, 25–31 July 2015.
24. Chen, C.; Ng, V. Joint inference over a lightly supervised information extraction pipeline: Towards event coreference resolution for resource-scarce languages. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Phoenix, AZ, USA, 12–17 February 2016; Volume 30.
25. Peng, H.; Song, Y.; Roth, D. Event detection and co-reference with minimal supervision. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, TX, USA, 1–5 November 2016; pp. 392–402.
26. Bejan, C.A.; Harabagiu, S. Unsupervised event coreference resolution. *Comput. Linguist.* 2014, 40, 311–347.
27. Chen, C.; Ng, V. Chinese event coreference resolution: An unsupervised probabilistic model rivaling supervised resolvers. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Denver, CO, USA, 31 May 2015; pp. 1097–1107.
28. Tran, H.M.; Phung, D.; Nguyen, T.H. Exploiting document structures and cluster consistencies for event coreference resolution. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, Online, 1–6 August 2021; Volume 1.
29. Lu, J.; Ng, V. Constrained multi-task learning for event coreference resolution. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Online, 6–11 June 2021; pp. 4504–4514.
30. Barhom, S.; Shwartz, V.; Eirew, A.; Bugert, M.; Reimers, N.; Dagan, I. Revisiting joint modeling of cross-document entity and event coreference resolution. *arXiv* 2019, arXiv:1906.01753.
31. Zeng, Y.; Jin, X.; Guan, S.; Guo, J.; Cheng, X. Event coreference resolution with their paraphrases and argument-aware embeddings. In *Proceedings of the 28th International Conference on Computational Linguistics*, Online, 8–13 December 2020; pp. 3084–3094.
32. De Langhe, L.; Desot, T.; De Clercq, O.; Hoste, V. A Benchmark for Dutch End-to-End Cross-Document Event Coreference Resolution. *Electronics* 2023, 12, 850.