Multi-Round Extraction and Dynamic Role Selection Framework For Document-Level Event Extraction

Kaizhou Zhang
Beijing University of Posts and Telecommunications
China

Pengfei Wang
Beijing University of Posts and Telecommunications
China

Lei Zhang
Beijing University of Posts and Telecommunications
China

ABSTRACT

Document-level Event Extraction (DEE) aims to extract structured event information from a document, which is an indispensable downstream task for many NLP applications. Argument-scatter and multi-event are its two main challenges. Recent work decomposes this challenging DEE task into multiple steps such as entity recognition, contextual information modeling, and event arguments extraction. Besides, they all extract event arguments in a predefined fixed role order. Though it is effective, its cumbersome steps and fixed extraction order will bring about the problem of error propagation. To address this issue, we propose a Multi-round Extraction and Dynamic Role Selection (MREDRS) Framework for the DEE task. We model the DEE task in an end-to-end manner to avoid these cumbersome steps. Multi-round extraction can deal with multi-event problem. In order to avoid the error propagation problem caused by the fixed role extraction order, we dynamically select the next role according to the current extraction state. We conducted experiments on the commonly used DEE dataset and extensive experimental results demonstrated the effectiveness of our method.

CCS CONCEPTS

• **Computing methodologies** → **Information extraction**.

KEYWORDS

document-level event extraction, multi-round, dynamic.

ACM Reference Format:
Kaizhou Zhang, Pengfei Wang, and Lei Zhang. 2022. Multi-Round Extraction and Dynamic Role Selection Framework For Document-Level Event Extraction. In *Proceedings of the 5th International Conference on Machine Learning and Natural Language Processing (MLNLP 2022), December 23–25, 2022, Sanya, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3578741.3578768

1 INTRODUCTION

Event Extraction is an important yet challenging task in information extraction (IE) research. As a particular form of information, an event is a real-world occurrence that takes place in a specific location and time that relates to a particular topic [26]. Different topics mean different types of events. There are various event types and then different roles for different types of events. So event extraction first detects the event types from the text, and then finds the entities corresponding to the roles. The ACE conference greatly facilitates the progress of event extraction at the sentence level [2, 5, 9, 14, 16, 20, 24]. But these sentence-level event extraction (SEE) methods are ill-suited in many fields, such as finance, health, etc. And it cannot handle the problem of arguments scattered in different sentences. Therefore, DEE has a wide range of application scenarios and has received more attention recently.

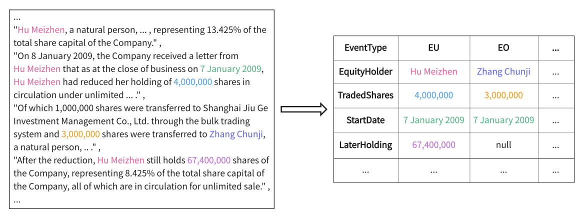

Figure 1: An example of DEE. Some entities are marked in color. We can see that there are multiple events in a document and the argument for each event spans multiple sentences. EU refers to EquityUnderweight, EO refers to EquityOverweight.

In contrast to SEE, DEE (Fig. 1 illustrates an example.) currently has two serious challenges. The first is that the arguments of the event will be scattered in different sentences, which requires the model to have a global understanding of the document; the second is that multiple events will appear in a document at the same time, and there is a dependency among these events. We need to capture the relationship among these events. Recent works [21, 27] decompose it into multiple steps. Firstly, they use a transformer to do a sequence tagging task and get entity, sentence representation. Secondly, contextual information is incorporated into these representations to deal with argument-scatter challenge. Thirdly, the event type is detected. Finally, entity is classified into roles by a tree-expanding structure to deal with multi-event challenge. The two challenge is solved to some extent in their methods. However, as we can see error is propagated in these cumbersome steps, and the fixed role extraction order is not a good choice. It also poses problems of error propagation.

To address these problems, we propose the Multi-round Extraction and Dynamic Role Selection (MREDRS) Framework for the DEE task. Firstly, we omit these tedious steps and model the DEE

MLNLP 2022, December 23–25, 2022, Sanya, China

ACM ISBN 978-1-4503-9906-7/22/12...$15.00
https://doi.org/10.1145/3578741.3578768

task in an end-to-end manner. We directly map the text to the role of event and avoid sequencing tagging task at the same time. To deal with the multi-event challenge, we model multi-event extraction as multi-round extraction until there is no event information left. Secondly, the extraction order of fixed roles is not a good choice. For example, everyone's life is a different journey, and it is almost impossible to replicate it exactly like everyone else's. So we propose that the current extraction status determines the next role to be extracted. Attention mechanism [1] is used to choose the next role with the highest probability.

Our contributions are summarized as follows:

- Our framework implements one-stop extraction in an end-to-end manner. We directly map the text to the role of an event, and the problem of error propagation caused by cumbersome steps will be avoided.
- Multi-round extraction is used to deal with multi-event problem. In order to avoid the error propagation problem caused by the fixed role extraction order, we flexibly select the next role to be extracted according to the current extraction state.
- We achieve state-of-the-art results on the large-scale and widely used dataset.

2 RELATED WORK

There are two main research directions of Event Extraction: schema induction and specific schema. Schema induction means that researchers use unsupervised method [8, 13] to induce event schema via very large corpus and extract corresponding events. We think it will receive more and more attention in the future. Specific schema event extraction means that we have pre-defined event schema and we only need to extract corresponding events. Recent works can be divided into SEE and DEE.

2.1 Sentence-level Event Extraction

Early works [7, 11] use NLP tools to design many fine features (such as local features and global features) for event extraction. Since 2015, deep learning methods have been gradually applied. [2] first proposes to use dynamic multi-pooling CNN to capture lexical-level and sentence-level features for event extraction in a pipeline manner. [16, 17] use bidirectional RNN to simultaneously extract trigger words and arguments of events to avoid the error propagation problem in the pipeline method. [14] solves the multi-event problem by introducing syntactic shortcut arcs, then uses GCN [10] to aggregate the information of surrounding trigger words and entities, finally extracting trigger words and arguments of multiple events at the same time. Recent works mostly relied on pretrained language models. [24] solves the role overlap problem by maintaining two lists of starting and ending positions for each role. Then using Bert [4] to generate more data to solve the data sparse problem of the ACE dataset. [15] utilizes a pretrained language model to generate structured events in a sequence-to-structure manner. [20] incorporates event schema knowledge into the pre-training model through contrastive pre-training to help event extraction.

2.2 Document-level Event Extraction

DEE has attracted more attention recently because SEE cannot handle the problem of argument scattered in different sentences. MUC-4 [18] proposes template-filling task that aims to extract event role fillers from the document. Recently, [6, 12, 25] follow this direction to extract event role fillers and public corresponding RAMS and WIKIEVENTS datasets. But they are all sub-task of DEE and ignore the challenge of multi-event. [22] proposes a method that first detect key-event and gets other events from surrounding sentences. [27] released the largest DEE dataset to date and proposes a tree-expanding method to extract events by modeling the relation among arguments in the same event type. [21] follows the tree expanding and further captures the contextual information and models the relations between events and arguments. [23] first proposes a parallel prediction paradigm for DEE. They all break down the DEE task into tedious steps to solve and the role order is fixed, which will suffer from an error accumulation problem.

3 METHODOLOGY

Before introducing our MREDRS framework (our framework is shown in Fig. 2) for DEE, we first clarify some important notions. a) **entity mention**: a text span within document that refers to an entity object; b) **event role**: an event role corresponds to a predefined field of the event; c) **event argument**: an event argument is an entity that plays a specific event role; d) **event record**: an entry of a specific event type containing arguments for different roles in the event.

3.1 Encoding

We use transformer encoder [19] to get representation. Given a document D, we need to get character representation and document representation. Let $D = [S_1, S_2, ..., S_n]$ denote N sentence, $S_i = [c^i_1, c^i_2, ..., c^i_m]$ denote m character. So, the character representation is geted by transformer:

$$[g^i_1, g^i_2, ..., g^i_m] = Transformer([c^i_1, c^i_2, ..., c^i_m]),$$

where $c^i_j \in R^d$ Then, we use max-pooling to get sentence representation $S_i = max - pooling([g^i_1, g^i_2, ..., g^i_m])$. After we get all sentence representation, we use mean-pooling to get document representation $D = mean - pooling([S_1, S_2, ..., S_n])$, where $D \in R^d$. It will be used to detect if there are any event information left.

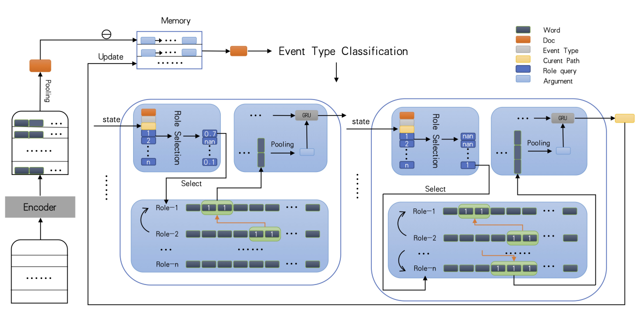

Figure 2: Overview of MREDRS.

3.2 Multi-round Extraction

We perform multi-round extraction based on the document information D. When document information no longer contains event information, the extraction process ends. When an event is extracted, the event information E will be used to update event memory M:

$$M = mean - pooling(MemEnc(M, E)).$$

Then we use subtraction manner to remove the event memory information from document information:

$$D = W(D - M),$$

where W is trainable weights. The new document information is used to detect is there any event information:

$$\hat{y_i} = Sigmoid(W_{di}D + W_{bi}),$$

where i denote the ith event type, W_d and W_b are trainable weights. When all predicted event type labels are 0, it means there is no event information left. Multi-round extraction will be end. Cross entropy is utilized to compute the event type classification loss:

$$L_e = -\sum_{i=1}^{m}\sum_{j=1}^{n}\left[y \log y_{ij} + (1-y)\log\left(1-y_{ij}\right)\right],$$

where m denote the document has m events and n denote the number of event type.

3.3 Dynamic Role Selection

Previous works extract event argument in a fixed order of roles. It is not a good selection. We propose to dynamically select the next role base on current state (current arguments information). The initial state is initialized by a random vector. Then, we use document (D), event type query (E_i), current state (C) and role queries (R_i) to do a role selection. Specifically, we use transformer to get more informative representation. Then attention mechanism is used to select the next role:

$$[\widetilde{D}, \widetilde{E_i}, \widetilde{C}, \widetilde{R_i}] = Transformer([D, E_i, C, R_i]),$$
$$Selected\_role = Softmax(Attention(\widetilde{C}\widetilde{R_i})),$$

where i means ith event type.

After select the next role, we utilize role query to match corresponding character. We add end symbol ([stop]) at the end of character. The symbol is randomly initialized. Attention mechanism is utilized to select the position of event argument:

$$Characters = [c_{11}, c_{12}, ..., c_{21}, c_{22}, ..., ..., [stop]],$$
$$position = Argmax(Attention(R_{ij}^{0}Characters)),$$
$$R_{ij}^{1} = R_{ij}^{0} + c_{position},$$

where i means ith event type, j means the jth role chosen by our role selection module. While the selected position is not the end position, attention mechanism will continue to select the next position. The current state (C) will be updated by GRU [3] when the argument is extracted:

$$C_j = GRU(C_{j-1}, mean - pooling([c_{ij}, ...])).$$

The argument extraction loss is denoted as L_a:

$$L_a = -\sum y \log y_{ij},$$

where y_{ij} denotes the probability of the character c_{ij} being selected by the role query. So the training loss is:

$$\mathcal{L} = \lambda_1 L_e + \lambda_2 L_a,$$

where λ_1 and λ_2 denote the weight of two task.

4 EXPERIMENTS

4.1 Experiments Settings

Dataset. We evaluate our document-level event extraction framework on the widely used ChFinAnn dataset [27]. The ChFinAnn dataset consists of Chinese financial documents. It is the largest document-level event extraction dataset so far which contains 32,040 documents. It focuses on five event types: Equity Freeze (EF, 8 roles), Equity Repurchase (ER, 6 roles), Equity Underweight (EU, 6 roles), Equity Overweight (EO, 6 roles) and Equity Pledge (EP, 9 roles). We follow the standard split of the dataset, 25,632 / 3,204 / 3,204 documents for training/dev/test set. 29% of the documents in the dataset contain multiple events. the maximum number of events contained in one document is 34, the average document contains 1.5 events, so the dataset is very challenging.

Baselines. We select three competitive DEE baselines and their variants for compare.

- **DCFEE** [22]: DCFEE has two variants: **DCFEE-O** and **DCFEE-M**, where **DCFEE-O** only produces one event record from one key-event sentence, while **DCFEE-M** tries to get multiple possible argument combinations by the closest relative distance from the key-event sentence.
- **Doc2EDAG** [27]: Doc2EDAG first propose tree structure to extract events by modeling the relationship between arguments in the same events in serial prediction paradigm. **GreedyDec** is also a variant of Doc2EDAG which extract one event greedily.
- **GIT** [21]: GIT based on Doc2EDAG, they model the relationship between events and further modeling the relationship between entities and sentences in serial prediction paradigm.

Metrics. We adopt the evaluation metrics standard used in these baselines for a fair comparison. Specifically, We sort the extraction result set in descending order by the number of non-empty roles. Then the most similar groundtruth is selected without replacement to calculate the micro-averaged event-level Precision (P), Recall (R) and F1-score (F1).

4.2 Results and Analysis

Overall Performance. The results of the overall performance on the DEE dataset is illustrated in Table 1. As Table 1 shows, we achieve the SOTA result because we omit tedious steps and model DEE task in an end-to-end manner. Multi-round extraction is used to deal with multi-event extraction and dynamically role selection help us to avoid error propagation problem.

Single-Event vs Multi-Event. To better demonstrate the effectiveness of our method, we conduct experiments on two scenarios: single-event (i.e., documents contain one event) and multi-event (i.e., documents contain multiple events). Table 2 shows the results on single-event and multi-event sets for each event type and the overall F1-score. We can observe that multi-event scenario is extremely challenging as the extraction performance of all models

Table 1: Precision, recall on each event type. EF/ER/EU/EO/EP refer to specific event types. Our method get sota result for each event type.

Models	EF		ER		EU		EO		EP		Overall	
	P.	R.	P.	R.	P.	R.	P.	R.	P.	R.	P.	R.
DCFEE-O	62.9	39.6	81.1	78.3	59.8	31.4	48.3	40.1	61.2	60.1	65.7	57.2
DCFEE-M	48.1	37.2	80.8	77.1	48.7	38.6	40.3	45.9	58.7	63.6	63.2	51.9
Greedy-Dec	76.3	45.1	81.5	72.6	66.7	37.1	65.2	38.5	83.2	46.3	74.9	52.5
Doc2EDAG	75.8	64.8	88.7	80.5	78.7	64.2	80.4	67.1	78.9	72.3	78.1	73.5
GIT	78.3	67.2	90.2	82.1	80.4	65.9	82.7	68.9	82.1	73.9	81.3	74.9
MREDRS	**79.6**	**68.5**	**91.4**	**74.3**	**82.6**	**67.8**	**84.2**	**70.0**	**83.8**	**75.1**	**83.5**	**76.7**

Table 2: F1 scores on single-event (S.) and multi-event (M.) sets. Our method get sota result in the two scenarios.

Models	EF		ER		EU		EO		EP		Overall	
	S.	M.	S.	M.	S.	M.	S.	M.	S.	M.	S.	M.
DCFEE-O	56.0	46.5	86.7	54.1	48.5	41.2	47.7	45.2	68.4	61.1	61.5	49.6
DCFEE-M	48.4	43.1	83.8	53.4	48.1	39.6	47.1	42.0	67.0	60.6	58.9	47.7
Greedy-Dec	76.8	41.0	83.5	51.1	61.3	34.1	62.7	29.8	77.7	36.3	78.0	36.9
Doc2EDAG	75.0	62.8	89.1	67.6	74.7	58.7	75.4	65.1	82.3	68.9	83.7	67.7
GIT	80.7	65.5	91.7	72.5	78.6	66.1	78.0	69.1	84.4	72.8	86.4	71.7
ours	**81.9**	**70.8**	**92.0**	**74.1**	**79.2**	**68.6**	**78.8**	**69.8**	**85.6**	**74.9**	**87.5**	**73.6**

drops significantly compared with single-events. But our framework still gets the highest result on all event types and improves the overall micro F1-score from 71.7% to 73.6% in the multi-event scenario compared to GIT [21]. Because our multi-round extraction manner better solves multi-event challenge. The dynamic role selection can avoid error propagation problem and help our method get improve both in single and multi-event scenario.

5 CONCLUSION

In this paper, we focus on the challenging document-level Event Extraction task. We propose a Multi-round Extraction and Dynamic Role Selection (MREDRS) Framework for the DEE task. Event is extracted in an end-to-end manner in our framework. We deal with the multi-event challenge in a multi-round extraction manner. In order to avoid the error propagation problem caused by the fixed role order, we propose to select the next role dynamically. It depends on the current extraction state. On the commonly and widely used DEE dataset, we achieve the SOTA result. In future work, we will apply our model on other datasets to demonstrate the generalization of the proposed model.

REFERENCES

[1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).

[2] Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao. 2015. Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*. The Association for Computer Linguistics, 167–176. https://doi.org/10.3115/v1/p15-1017

[3] Kyunghyun Cho, Bart van Merrienboer, Çaglar Gülçehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, Alessandro Moschitti, Bo Pang, and Walter Daelemans (Eds.). ACL, 1724–1734. https://doi.org/10.3115/v1/d14-1179

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, 4171–4186. https://doi.org/10.18653/v1/n19-1423

[5] Xinya Du and Claire Cardie. 2020. Event Extraction by Answering (Almost) Natural Questions. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, 671–683. https://doi.org/10.18653/v1/2020.emnlp-main.49

[6] Xinya Du, Alexander M. Rush, and Claire Cardie. 2021. GRIT: Generative Role-filler Transformers for Document-level Event Entity Extraction. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021, Online, April 19 - 23, 2021*, Paola Merlo, Jörg Tiedemann, and Reut Tsarfaty (Eds.). Association for Computational Linguistics, 634–644. https://doi.org/10.18653/v1/2021.eacl-main.52

[7] Yu Hong, Jianfeng Zhang, Bin Ma, Jianmin Yao, Guodong Zhou, and Qiaoming Zhu. 2011. Using cross-entity inference to improve event extraction. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*. 1127–1136.

[8] Lifu Huang, Taylor Cassidy, Xiaocheng Feng, Heng Ji, Clare R. Voss, Jiawei Han, and Avirup Sil. 2016. Liberal Event Extraction and Event Schema Induction. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*. The Association for Computer Linguistics. https://doi.org/10.18653/v1/p16-1025

[9] Lifu Huang, Heng Ji, Kyunghyun Cho, Ido Dagan, Sebastian Riedel, and Clare R. Voss. 2018. Zero-Shot Transfer Learning for Event Extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, Iryna Gurevych and Yusuke Miyao (Eds.). Association for Computational Linguistics, 2160–2170. https://doi.org/10.18653/v1/P18-1201

[10] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net. https://openreview.net/forum?id=SJU4ayYgl

[11] Qi Li, Heng Ji, and Liang Huang. 2013. Joint Event Extraction via Structured Prediction with Global Features. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, ACL 2013, 4-9 August 2013, Sofia, Bulgaria, Volume 1: Long Papers*. The Association for Computer Linguistics, 73–82. https://aclanthology.org/P13-1008/

[12] Sha Li, Heng Ji, and Jiawei Han. 2021. Document-Level Event Argument Extraction by Conditional Generation. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021*, Kristina Toutanova, Anna Rumshisky, Luke Zettlemoyer, Dilek Hakkani-Tür, Iz Beltagy, Steven Bethard, Ryan Cotterell, Tanmoy Chakraborty, and Yichao Zhou (Eds.). Association for Computational Linguistics, 894–908. https://doi.org/10.18653/v1/2021.naacl-main.69

[13] Xiao Liu, Heyan Huang, and Yue Zhang. 2019. Open Domain Event Extraction Using Neural Latent Variable Models. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28-August 2, 2019, Volume 1: Long Papers*, Anna Korhonen, David R. Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, 2860–2871. https://doi.org/10.18653/v1/p19-1276

[14] Xiao Liu, Zhunchen Luo, and Heyan Huang. 2018. Jointly Multiple Events Extraction via Attention-based Graph Information Aggregation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii (Eds.). Association for Computational Linguistics, 1247–1256. https://doi.org/10.18653/v1/d18-1156

[15] Yaojie Lu, Hongyu Lin, Jin Xu, Xianpei Han, Jialong Tang, Annan Li, Le Sun, Meng Liao, and Shaoyi Chen. 2021. Text2Event: Controllable Sequence-to-Structure Generation for End-to-end Event Extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, 2795–2806. https://doi.org/10.18653/v1/2021.acl-long.217

[16] Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. 2016. Joint Event Extraction via Recurrent Neural Networks. In *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, Kevin Knight, Ani Nenkova, and Owen Rambow (Eds.). The Association for Computational Linguistics, 300–309. https://doi.org/10.18653/v1/n16-1034

[17] Lei Sha, Feng Qian, Baobao Chang, and Zhifang Sui. 2018. Jointly Extracting Event Triggers and Arguments by Dependency-Bridge RNN and Tensor-Based Argument Interaction. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, Sheila A. McIlraith and Kilian Q. Weinberger (Eds.). AAAI Press, 5916–5923. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16222

[18] Beth Sundheim. 1992. Overview of the fourth message understanding evaluation and conference. In *Proceedings of the 4th Conference on Message Understanding, MUC 1992, McLean, Virginia, USA, June 16-18, 1992*. ACL, 3–21. https://doi.org/10.3115/1072064.1072066

[19] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). 5998–6008. https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

[20] Ziqi Wang, Xiaozhi Wang, Xu Han, Yankai Lin, Lei Hou, Zhiyuan Liu, Peng Li, Juanzi Li, and Jie Zhou. 2021. CLEVE: Contrastive Pre-training for Event Extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, 6283–6297. https://doi.org/10.18653/v1/2021.acl-long.491

[21] Runxin Xu, Tianyu Liu, Lei Li, and Baobao Chang. 2021. Document-level Event Extraction via Heterogeneous Graph-based Interaction Model with a Tracker. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli (Eds.). Association for Computational Linguistics, 3533–3546. https://doi.org/10.18653/v1/2021.acl-long.274

[22] Hang Yang, Yubo Chen, Kang Liu, Yang Xiao, and Jun Zhao. 2018. Dcfee: A document-level chinese financial event extraction system based on automatically labeled training data. In *Proceedings of ACL 2018, System Demonstrations*. 50–55.

[23] Hang Yang, Dianbo Sui, Yubo Chen, Kang Liu, Jun Zhao, and Taifeng Wang. 2021. Document-level event extraction via parallel prediction networks. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 6298–6308.

[24] Sen Yang, Dawei Feng, Linbo Qiao, Zhigang Kan, and Dongsheng Li. 2019. Exploring Pre-trained Language Models for Event Extraction and Generation. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, Anna Korhonen, David R. Traum, and Lluís Màrquez (Eds.). Association for Computational Linguistics, 5284–5294. https://doi.org/10.18653/v1/p19-1522

[25] Zhisong Zhang, Xiang Kong, Zhengzhong Liu, Xuezhe Ma, and Eduard H. Hovy. 2020. A Two-Step Approach for Implicit Event Argument Detection. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault (Eds.). Association for Computational Linguistics, 7479–7485. https://doi.org/10.18653/v1/2020.acl-main.667

[26] Liang Zhao. 2021. Event Prediction in the Big Data Era: A Systematic Survey. *ACM Comput. Surv.* 54, 5 (2021), 94:1–94:37. https://doi.org/10.1145/3450287

[27] Shun Zheng, Wei Cao, Wei Xu, and Jiang Bian. 2019. Doc2EDAG: An End-to-End Document-level Framework for Chinese Financial Event Extraction. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics, Hong Kong, China, 337–346. https://doi.org/10.18653/v1/D19-1032