Supplemental Material

Due to the space limit, we cannot include all details in the paper. Therefore, we provide this supplemental material, organized as follows:

- Section 1 introduces event types used in our paper and corresponding preprocessing details when conducting the document-level event labeling.
- Section 2 describes some basic building blocks of our model in details.
- Section 3 presents a complete hyper-parameter setting to enable the reproducible research.
- Section 4 contains more case studies.

Note that, we also open the hand-tagged data and the data-quality evaluation script in the same repository.

1 Event Type Specifications

Table 1 shows detailed illustrations for event types used in our paper, where we mark some key roles that should be non-empty when conducting the document-level event labeling. In addition to requiring non-empty key roles, we empirically set the minimum number of matched roles for EF, ER, EU, EO and EP events as 5, 4, 4, 4 and 5, respectively. While we set these constraints empirically to ensure the labeling quality for our data, practitioners of other domains can adjust these configurations freely to fulfill the task-specific requirements by making the trade-off between precision and recall.

2 Model

In Section 5 of the paper, we frequently employ two basic building blocks: 1) the attentively weighted average (AWA) module [Yang *et al.*, 2016] and 2) the Transformer module [Vaswani *et al.*, 2017]. Next, we present details about these two modules.

2.1 AWA

The AWA module (mentioned in Section 5.1 and 5.2 of the paper) was employed by [Yang *et al.*, 2016] to obtain the sentence embedding from a sequence of word embeddings and produce a final document embedding from a sequence of generated sentence embeddings. In our model, we adopt a similar design to get a single embedding from a sequence of embeddings with the same embedding size. Specifically, given a sequence of embeddings, $\boldsymbol{x} = [\boldsymbol{x}_1; \boldsymbol{x}_2; \cdots; \boldsymbol{x}_{N_x}]$, where each $\boldsymbol{x}_i \in \mathbb{R}^{d_w}$ is an embedding with size d_w and $\boldsymbol{x} \in \mathbb{R}^{d_w \times N_x}$ is the embedding sequence with length N_x, we take the scaled dot-product attention [Vaswani *et al.*, 2017] operations:

$$u_i = \frac{Q^\top \boldsymbol{x}_i}{\sqrt{d_w}}, \tag{1}$$

$$\alpha_i = \frac{\exp(u_i)}{\sum_j \exp(u_j)}, \tag{2}$$

$$\hat{\boldsymbol{x}} = \text{Dropout}\left(\text{LayerNorm}\left(\sum_i \alpha_i \boldsymbol{x}_i\right)\right), \tag{3}$$

to produce a single embedding $\hat{\boldsymbol{x}} \in \mathbb{R}^{d_w}$, where $Q \in \mathbb{R}^{d_w}$ is a trainable parameter, LayerNorm is the layer normalization [Ba *et al.*, 2016] and Dropout is an effective technique to avoid overfitting [Srivastava *et al.*, 2014].

As described in Section 5 of the paper, we employ four different AWA modules for the following purposes:

- getting a single entity mention embedding from a sequence of embeddings of covered tokens;
- getting a single entity mention embedding from a sequence of embeddings of covered tokens;
- inducing a single entity embedding from multiple entity mention embeddings;
- obtaining a document embedding from encoded sentence embeddings to conduct the event-triggering classification.

2.2 Transformer

As for the Transformer module, we mainly follow [Vaswani *et al.*, 2017], but have also referred an excellent guide, "The Annotated Transformer "[1].

While in our setting, we employ the Transformer as a context encoder to exchange information among multiple inputs with the following changes:

- Transformer-1 (mentioned in Section 5 of the paper) looks up a sentence-level trainable position embedding table and add original token embeddings with associated token-positional embeddings, as the multi-headed self-attention mechanism in Transformer is position-agnostic.

[1] http://nlp.seas.harvard.edu/2018/04/03/attention.html

Event Type	Event Role	Detailed Explanations
Equity Freeze (EF)	**Equity Holder** (key)	the equity holder whose shares are froze
	Froze Shares (key)	the number of shares being froze
	Legal Institution (key)	the legal institution that executes this freeze
	Start Date	the start date of this freeze
	End Date	the end date of this freeze
	Unfroze Date	the date in which these shares are unfroze
	Total Holding Shares	the total number of shares being hold at disclosing time
	Total Holding Ratio	the total ratio of shares being hold at disclosing time
Equity Repurchase (ER)	**Company Name** (key)	the name of the company
	Highest Trading Price	the highest trading price
	Lowest Trading Price	the lowest trading price
	Closing Date	the closing date of this disclosed repurchase
	Repurchased Shares	the number of shares being repurchased before the closing date
	Repurchase Amount	the repurchase amount before the closing date
Equity Underweight (EU)	**Equity Holder** (key)	the equity holder who conducts this underweight
	Traded Shares (key)	the number of shares being traded
	Start Date	the start date of this underweight
	End Date	the end date of this underweight
	Average Price	the average price during this underweight
	Later Holding Shares	the number of shares being hold after this underweight
Equity Overweight (EO)	**Equity Holder** (key)	the equity holder who conducts this overweight
	Traded Shares (key)	the number of shares being traded
	Start Date	the start date of this overweight
	End Date	the end date of this overweight
	Average Price	the average price during this overweight
	Later Holding Shares	the number of shares being hold after this overweight
Equity Pledge (EP)	**Pledger** (key)	the equity holder who pledges some shares to an institution
	Pledged Shares (key)	the number of shares being pledged
	Pledgee (key)	the institution who accepts the pledged shares
	Start Date	the start date of this pledge
	End Date	the end date of this pledge
	Released Date	the date in which these pledged shares are released
	Total Pledged Shares	the total number of shares being pledged at disclosing time
	Total Holding Shares	the total number of shares being hold at disclosing time
	Total Holding Ratio	the total ratio of shares being hold at disclosing time

Table 1: Event type specifications.

- Before feeding into Transformer-2 (mentioned in Section 5.2) and Transformer-3 (mentioned in Section 5.3), we add the input embeddings with trainable sentence-positional and role-indicator embeddings, respectively, to enable the awareness of specific encoding tasks.

Moreover, for the entity recognition part, we refer readers to [Huang *et al.*, 2015] for details about stacking the conditional random field (CRF) layer over encoded representations.

3 Hyper-parameters Setting

We summarize all hyper-parameters in Table 2 towards the reproducible research.

4 Case Studies

In addition to the *Equity Pledge* example included by the paper, we show another three cases with comprehensive analyses in Figure 1, 2 and 3 for the *Equity Overweight*, *Equity Underweight* and *Equity Freeze* events, respectively, where we color the wrong predicted arguments as red.

References

[Ba *et al.*, 2016] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

[Huang *et al.*, 2015] Zhiheng Huang, Wei Xu, and Kai Yu. Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*, 2015.

[Kingma and Ba, 2015] Diederick P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.

[Srivastava *et al.*, 2014] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *JMLR*, 2014.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NIPS*, 2017.

Module	**Hyper-parameter**	**Value**
Input Representation	the maximum sentence number	64
	the maximum sentence length	128
	d_w (the embedding size)	768
Entity Recognition	the tagging scheme	BIO (Begin, Inside, Other)
	the hidden size	768 (same to d_w)
Transformer-1 Transformer-2 Transformer-3	the number of layers	4
	the size of the hidden layer	768 (same to d_w)
	the size of the feed-forward layer	$1,024$
Optimization	the optimizer	Adam [Kingma and Ba, 2015]
	the learning rate	$1e^{-4}$
	the batch size	64 (with 32 GPUs)
	the training epoch	100
	the loss reduction type	sum
	λ_1	0.05
	λ_2, λ_3	0.95
	γ	3
	the dropout probability	0.1
	the scheduled-sampling beginning	10^{th} epoch
	the scheduled-sampling ending	20^{th} epoch
	the scheduled probability of employing gold entity mentions	decreasing from 1.0 to 0.1 linearly during the scheduled epochs

Table 2: The hyper-parameter setting.

[Yang *et al.*, 2016] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In *NAACL-HLT*, 2016.

Event Table (DCFEE-O, key sentences: 3)

Equity Holder	Traded Shares	Start Date	End Date	Later Holding Shares	Average Price
[PER1]	[SHARE1]	[DATE1]	NA	[SHARE4]	NA

Event Table (Our Model)

Equity Holder	Traded Shares	Start Date	End Date	Later Holding Shares	Average Price
[PER1]	[SHARE1]	[DATE1]	[DATE1]	NA	NA
[PER2]	[SHARE2]	[DATE1]	[DATE1]	[SHARE2]	NA
[PER3]	[SHARE1]	[DATE1]	[DATE1]	NA	NA

Event Table (DCFEE-M, key sentences: 3)

Equity Holder	Traded Shares	Start Date	End Date	Later Holding Shares	Average Price
[PER1]	[SHARE1]	[DATE1]	NA	[SHARE4]	NA
[PER2]	[SHARE2]	[DATE1]	NA	[SHARE4]	NA
[PER3]	[SHARE3]	[DATE1]	NA	[SHARE4]	NA

Event Table (Ground-truth)

Equity Holder	Traded Shares	Start Date	End Date	Later Holding Shares	Average Price
[PER1]	[SHARE1]	[DATE1]	[DATE1]	[SHARE4]	NA
[PER2]	[SHARE2]	[DATE1]	[DATE1]	[SHARE2]	NA
[PER3]	[SHARE3]	[DATE1]	[DATE1]	[SHARE5]	NA

Entity Mark Table

Mark	Entity	Entity (English)
[PER1]	李兵	Bing Li
[PER2]	毛尊平	Zunping Mao
[PER3]	夏保琪	Baoqi Xia
[SHARE1]	30000股	30000 shares
[SHARE2]	20000股	20000 shares
[SHARE3]	17300股	17300 shares
[SHARE4]	63750股	63750 shares
[SHARE5]	20675股	20675 shares
[DATE1]	2018年12月21日	Dec. 21st, 2018

ID	Sentence
3	公司接到通知，**[DATE1]**，公司董事、总经理**[PER1]**先生通过深圳证券交易所证券交易系统增持公司**[SHARE1]**股份，公司财务负责人**[PER2]**先生通过深圳证券交易所证券交易系统增持公司**[SHARE2]**股份，公司董事会秘书**[PER3]**先生通过深圳证券交易所证券交易系统增持公司**[SHARE3]**股份。
	The company was informed that, in **[DATE1]**, Mr. **[PER1]**, the director and the general manager of the company, bought **[SHARE1]** of the company via the trading system of Shenzhen Stock Exchange, Mr. **[PER2]**, the finance manager, bought **[SHARE2]** of the company via the trading system of Shenzhen Stock Exchange, and Mr. **[PER3]**, the secretary of the board, bought **[SHARE3]** of the company via the trading system of Shenzhen Stock Exchange.
7	本次增持后，**[PER1]**先生共持有公司**[SHARE4]**股份。
	After this overweight, Mr. **[PER1]** hold **[SHARE4]** of the company in total.
8	本次增持后，**[PER2]**先生共持有公司**[SHARE2]**股份。
	After this overweight, Mr. **[PER2]** hold **[SHARE2]** of the company in total.
10	本次增持后，**[PER3]**先生共持有公司**[SHARE5]**股份。
	After this overweight, Mr. **[PER3]** hold **[SHARE5]** of the company in total.

Figure 1: In this case, there are three *equity overweight* (EO) events mentioned by the documents. Although DCFEE-O correctly identifies the key sentence (ID 3), it cannot decide how many events being expressed by this sentence, as its SEE module can only fulfill the sequence tagging task. Therefore, we implement another version, DCFEE-M, which guess possible events by the position closeness, and indeed DCFEE-M produce multiple partially correct events in this case. However, the arguments-completion stage of DCFEE-M is context-agnostic, which is the reason that DCFEE-M does not produce correct arguments for the *End Date* role ("DATE1" is already assigned with the *Start Date* role) and the *Later Holding Shares* role (the closest valid entity is "[SHARE4]"). Moreover, though achieving better results for this case, DCFEE-M is inferior to DCFEE-O in terms of the whole test set (shown in the paper), since the naive multi-event guessing fails on many other cases. As for our model, it also misses two arguments for the *Later Holding Shares* role. After the careful examination, we find that the empty ratio of this role is pretty high during training, and thus our model prefers to be conservative in assigning entities with this role.

Event Table (DCFEE-O, key sentences: 4, 6, 8)					
Equity Holder	Traded Shares	Start Date	End Date	Later Holding Shares	Average Price
[PER1]	[SHARE1]	[DATE2]	[DATE3]	NA	NA
[PER1]	[SHARE1]	[DATE2]	[DATE3]	NA	NA
[PER1]	[SHARE1]	[DATE2]	[DATE3]	NA	NA

Event Table (Our Model)					
Equity Holder	Traded Shares	Start Date	End Date	Later Holding Shares	Average Price
[PER1]	[SHARE1]	[DATE2]	[DATE2]	NA	NA
[PER1]	[SHARE3]	[DATE3]	[DATE3]	NA	NA

Event Table (DCFEE-M, key sentences: 4, 6, 8)					
Equity Holder	Traded Shares	Start Date	End Date	Later Holding Shares	Average Price
[PER1]	[SHARE1]	[DATE2]	[DATE3]	NA	NA
[PER1]	[SHARE1]	[DATE2]	[DATE1]	NA	NA
[PER1]	[SHARE1]	[DATE3]	[DATE1]	NA	NA
[PER1]	[SHARE3]	[DATE2]	[DATE1]	NA	NA
[PER1]	[SHARE1]	[DATE2]	[DATE1]	NA	NA
[PER1]	[SHARE5]	[DATE2]	[DATE1]	NA	NA

Event Table (Ground-truth)					
Equity Holder	Traded Shares	Start Date	End Date	Later Holding Shares	Average Price
[PER1]	[SHARE1]	[DATE2]	[DATE2]	[SHARE5]	NA
[PER1]	[SHARE3]	[DATE3]	[DATE3]	[SHARE6]	NA

Entity Mark Table		
Mark	Entity	Entity (English)
[ORG1]	通裕重工股份有限公司	Tongyu Heavy Industry Co.,Ltd.
[PER1]	朱金枝	Jinzhi Zhu
[DATE1]	2018年10月10日	Oct. 10th, 2018
[DATE2]	2014年6月23日	Jun. 23nd, 2014
[DATE3]	2018年9月28日	Sept. 28th, 2018
[SHARE1]	8000000股	8000000 shares
[SHARE2]	900000000股	900000000 shares
[SHARE3]	12090000股	12090000 shares
[SHARE4]	80075625股	80075625 shares
[SHARE5]	72075625股	72075625 shares
[SHARE6]	204136875股	204136875 shares
[RATIO1]	5%	5%
[RATIO2]	1%	1%
[RATIO3]	0.8889%	0.8889%
[RATIO4]	0.3700%	0.3700%
[RATIO5]	8.8973%	8.8973%

ID	Sentence
4	**[DATE1]**，**[ORG1]**（以下简称“通裕重工”或“本公司”）收到持股**[RATIO1]**以上股东**[PER1]**先生发来的《关于减持通裕重工股份达到**[RATIO2]**的通知》，自**[DATE2]**至**[DATE3]**，**[PER1]**先生通过大宗交易方式减持公司无限售流通股，......，具体情况如下：
	In **[DATE1]**, **[ORG1]** (abbreviated as “Tongyu Heavy Industries” or “the company”) received the “Notifications on that reduced holding-share ratio of Tongyu Heavy Industries reached **[RATIO2]**” from the shareholder Mr. **[PER1]** , who hold the company equities more than **[RATIO1]**, which stated that from **[DATE2]** to **[DATE3]**, **[PER1]** sold some shares of the company via the block trading, ..., the detailed information is as follows:
5	一、股东减持情况
	First, the information of this equity underweight
6	**[PER1]**先生于**[DATE2]**通过大宗交易方式减持**[SHARE1]**，占公司当时总股本**[SHARE2]**的比例为**[RATIO3]**；**[DATE3]**减持**[SHARE3]**，占公司目前总股本的比例为**[RATIO4]**。
	In **[DATE2]**, Mr. **[PER1]** sold **[SHARE1]** via the block trading, accounting for **[RATIO3]** of the capital stock of the company at that time; in **[DATE3]**, he sold **[SHARE3]** again, accounting for **[RATIO4]** of the capital stock of the company currently.
7	二、本次减持前后持股情况
	Second, the holding information before and after this equity underweight
8	本次减持前，**[PER1]**先生持有本公司股份**[SHARE4]**，占公司总股本**[SHARE2]**的**[RATIO5]**，**[DATE2]**减持**[SHARE1]**后持股**[SHARE5]**，......；**[DATE3]** **[PER1]** 先生减持**[SHARE3]**后持有**[SHARE6]**，......
	Before this underweight, **[PER1]** hold **[SHARE4]** of the company, accounting for **[RATIO5]** of the total capital stock of the company, **[SHARE2]**; while in **[DATE2]**, after selling **[SHARE1]**, he hold **[SHARE5]** of the company; ……; In **[DATE3]**, after selling **[SHARE3]**, Mr. **[PER1]** hold **[SHARE6]** of the company. …

Figure 2: This case shows the typical false positive errors made by DCFEE models. Although the document only contains two distinct *Equity Underweight* events in total, different sentences mention these events multiple times (ID 4, 6 and 8). However, the key-sentence detection module of DCFEE models cannot differentiate duplicated event mentions elegantly. Therefore, both of them produce duplicated event records. Especially, DCFEE-M, guessing multiple event mentions from a single sentence, suffers severe false positive errors in this case. In contrast, our model is naturally robust to such data characteristics, since we conduct the event table filling at the document level. The only missing arguments, belong to the *Later Holding Shares* role, are partially caused by the restriction of the maximum sentence length at the input stage (ID 8).

Event Table (DCFEE-O, key sentences: 14)							
Equity Holder	Froze Shares	Legal Institution	Start Date	End Date	Unfroze Date	Total Holding Shares	Total Holding Ratio
[PER1]	[SHARE2]	[ORG3]	[DATE1]	NA	NA	[SHARE1]	[RATIO2]

Event Table (Our Model)							
Equity Holder	Froze Shares	Legal Institution	Start Date	End Date	Unfroze Date	Total Holding Shares	Total Holding Ratio
[ORG2]	[SHARE1]	[ORG3]	[DATE1]	NA	NA	[SHARE1]	[RATIO2]
[PER1]	[SHARE2]	[ORG3]	[DATE1]	NA	NA	[SHARE2]	NA

Event Table (DCFEE-M, key sentences: 14)							
Equity Holder	Froze Shares	Legal Institution	Start Date	End Date	Unfroze Date	Total Holding Shares	Total Holding Ratio
[PER1]	[SHARE2]	[ORG3]	[DATE1]	NA	NA	[SHARE1]	[RATIO2]

Event Table (Ground-truth)							
Equity Holder	Froze Shares	Legal Institution	Start Date	End Date	Unfroze Date	Total Holding Shares	Total Holding Ratio
[ORG2]	[SHARE1]	[ORG3]	[DATE1]	NA	NA	[SHARE1]	[RATIO1]
[PER1]	[SHARE2]	[ORG3]	[DATE1]	NA	NA	[SHARE2]	[RATIO2]

Entity Mark Table		
Mark	Entity	Entity (English)
[DATE1]	2018年11月1日	Nov. 1st, 2018
[ORG1]	上海富控互动娱乐股份有限公司	Shanghai Fukong Interactive Entertainment Co.,Ltd.
[ORG2]	上海富控文化传媒有限公司	Shanghai Fukong Culture Media Co., Ltd.
[ORG3]	山东省济南市中级人民法院	Jinan Intermediate People's Court of Shangdong Province
[PER1]	颜静刚	Jinggang Yan
[SHARE1]	157876590股	157876590 shares
[SHARE2]	31825000股	31825000 shares
[RATIO2]	27.42%	27.42%
[RATIO3]	5.53%	5.53%

ID	Sentence
3	**[ORG1]**（以下简称“公司”）于**[DATE1]**获悉公司控股股东**[ORG2]**（以下简称“富控传媒”）和公司实际控制人**[PER1]**先生持有的公司股份被司法轮候冻结，具体情况如下：
	[ORG1] (abbreviated at “the company” below) was informed in **[DATE1]** that the shares hold by **[ORG2]** (abbreviated as “Fukong Media” below), the controlling shareholder, and Mr. **[PER1]**, the actual controller, were froze judicially in turn, the detailed information is listed as follows:
4	一、控股股东股份被冻结情况
	First, the information about the controlling shareholder being froze
5	司法轮候冻结机关：**[ORG3]**；
	The legal institution conducting this freeze: **[ORG3]**;
6	司法轮候冻结数量为：**[SHARE1]**，轮候冻结；
	The number of shares being froze: **[SHARE1]**, froze in turn;
7	冻结起始日：**[DATE1]**；
	The start date of this freeze: **[DATE1]**;
9	二、实际控制人股份被冻结情况
	Second, the information about the actual controller being froze
10	司法轮候冻结机关：**[ORG3]**；
	The legal institution conducting this freeze: **[ORG3]**;
11	司法轮候冻结数量为：**[SHARE2]**，轮候冻结；
	The number of shares being froze: **[SHARE2]**, froze in turn;
12	冻结起始日：**[DATE1]**；
	The start date of this freeze: **[DATE1]**;
14	截至本公告披露日，富控传媒持有公司 **[SHARE1]**，占公司总股本的**[RATIO1]**，已被轮候冻结**[SHARE1]**；**[PER1]**先生持有公司 **[SHARE2]**，占公司总股本的**[RATIO2]**，已被轮候冻结**[SHARE2]**。
	As of the date of this announcement, Fukong Media hold **[SHARE1]** of the company, accounting for **[RATIO1]** of the total share capital, where **[SHARE1]** is froze in turn; Mr. **[PER1]** hold **[SHARE2]** of the company, accounting for **[RATIO2]** of the total share capital, where **[SHARE2]** is froze in turn;

Figure 3: This case, containing two *Equity Freeze* events, is a typical example that violates the key-sentence labeling assumption of DCFEE, which assumes the sentence containing the most arguments as the key-event one. We can observe that the core arguments of the event scatter across multiple sub-sentences, such as ID 5, 6, 7, 10, 11 and 12, but DCFEE-O and DCFEE-M treat the summary sentence (ID 14) as the key-event one. However, the single sentence (ID 14) summarizes these two event records, and DCFEE models cannot address such multi-event sentences elegantly. Note that, each text snippet of ID 5, 6, 7, 10, 11 and 12 is not a complete sentence, but these text snippets are presented in a list manner and some of them even do not have ending punctuations (ID 4, 9). We have tried to merge such short snippets into a single long sentence, but applying this merge on the whole dataset can hurt the performance of DCFEE models on other event types. Thus, we drop this preprocessing option. In contrast, our model is immune to such merging and even benefit with a faster speed due to fewer sentences to be encoded. In terms of the extraction performance, our model correctly identifies these two events and arranges entities into proper table columns with only one missing argument for the *Total Holding Ratio* role. While DCFEE models miss one event and inevitably make mistakes when completing missing arguments for the key-event sentence.