CFEE

Chinese Financial Event Extraction (CFEE) dataset is a financial-domain Chinese corpus where a document contains one or more event templates. The golden structured data comes from a financial event knowledge database and the unstructured text data come from announcements released by the companies which are stored in an unstructured form on the web. These data can be used for training and validation of document-level event extraction. There are three event types on CFEE as shown in follows:

FREEZE: Equity Freezing refers to a compulsory measure adopted by the people's court to restrict the owner of equity to withdraw or transfer his own equity. The main purpose of such measures is to prevent the improper loss of equity returns.

PLEDGE: Shareholder's Equity Pledge is that the shareholders of listed companies use their stocks (equity) as collateral to apply for loans from banks or provide a guarantee for loans from third parties.

OW&UW: Overeighting\nderweighting Shares refers to the shareholders of a joint-stock company resell part of the stocks they hold and reduce or increase their controlling rights over the company.

For each event type, there are different event roles in it as shown in table 1

Event type	Freeze	Pledge	Overweight/Underweight
NAME	Shareholder Name	Shareholder Name	Shareholder Name
ORG	Freeze Organization	Pledge Institution	—
NUM	Number of Frozen Stock	Number of Pledged Stock	Number of Stock Increase or Decrease
BEG	Freezing Start Date	Pledging Start Date	Start Date
END	Freezing End Date	Pledging End Date	End Date

Table 1: Event roles in the three financial event types.

And finally, a total of 7144 document have been labeled. Table 2 shows the statistics of CFEE dataset.

Freeze	Pledge	Overweight	Total
589/150/300	3602/300/300	1303/300/300	5494/750/900

Table 2: Statistics of CFEE Dataste train/dev/test.

There are three JSON files in the folder as shown in follows: FREEZE_Data.json, PLEDGE_Date.json and OW&UW_Data.json. They correspond to three event types: FREEZE\ PLEDGE\ OW&UW.

Each annotated documents consists of the following parts:

1) ID : Event type + TRAIN\DEV\TEST + number. For example, "FREEZE_TEST_0007" indicate this data acts as a TEST set for FREEZE event and its number is 0007.
2) Titile: The title of the document.
3) Doc: Text of the document.
4) Event: Structured event templates in the document.