

IREE: A Fine-Grained Dataset for Chinese Event Extraction in Investment Research

Junxiang Ren, Sibow Wang, Ruilin Song, Yuejiao Wu, Yizhou Gao, Borong An, Zhen Cheng, Guoqiang Xu
China Pacific Insurance (Group) Co Ltd

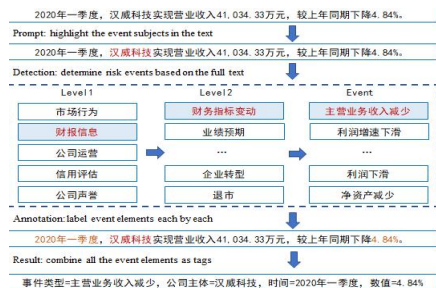
Introduction

In recent years, a large number of datasets for common event extraction as well as financial event extraction have been released. However, they only cover few event types compared with real-world business scenarios. Hence, by combining actual requirements of financial investment risk management with business needs and experience in terms of investment, we construct a dataset for financial event extraction.

IREE construction

Data resource

We use the news corpus from the investment decision support system as the source of data. Original corpus contains various mainstream information platforms, traditional media and new media. We divide all the news into 5 main categories, including 59 kinds of event types.



Level 1	Level 2
市场行为	停牌复牌、戴帽、异常交易、增持、减持、关联交易、股份解禁
财报信息	财务指标变动、业绩预期、管理层变动
公司运营	企业担保、并购重组、发行股份、意外事件、债务、业务重组等
信用评估	债券违约、评级调整
公司声誉	违规处罚、诉讼纠纷、公司丑闻

Data annotation

Data Annotation includes original texts reading, subject prompting, event type detecting, argument annotating and tag integrating. The details are shown in the figure.

Experiment

Event detection Event detection is based on whether the event type is correctly detected. For event detection, we considered three options.

	Micro-F1	Macro-F1
BERT-NER	0.5085	0.4531
BERT-MRC	0.6336	0.5775
BERT-CasRel	0.6649	0.6139

	Micro-F1	Macro-F1
TextCNN	0.6857	0.4952
BERT	0.7091	0.5014
BERT-Prompt	0.6769	0.4663

Argument extraction Argument extraction only evaluates whether the argument type is correct. For event argument extraction, we also consider three options.

Conclusion

This paper presents a Chinese fine-grained dataset for event type extraction, which provides experience for the construction of the datasets for event extraction in the investment field. Annotators identify the significant financial events that reflect investment risks based on their domain expertise in investment. As the field covers widely and investment risks are subject to policies and regulations, we will make continual update as our follow-up work.