Exploring Hypergraph of Earnings Call for Risk Prediction (Student Abstract)

Yi He*[1], Wenxin Tai*[1], Fan Zhou[1 2 †], Yi Yang [3]

[1] University of Electronic Science and Technology of China
[2] Kashi Institute of Electronics and Information Industry
[3] The Hong Kong University of Science and Technology
hyjrj@foxmail.com, amperetai@gmail.com, fan.zhou@uestc.edu.cn, imyiyang@ust.hk

Abstract

In financial economics, studies have shown that the textual content in the earnings conference call transcript has predictive power for a firm's future risk. However, the conference call transcript is very long and contains diverse non-relevant content, which poses challenges for the text-based risk forecast. This study investigates the structural dependency within a conference call transcript by explicitly modeling the dialogue between managers and analysts. Specifically, we utilize TextRank to extract information and exploit the semantic correlation within a discussion using hypergraph learning. This novel design can improve the transcript representation performance and reduce the risk of forecast errors. Experimental results on a large-scale dataset show that our approach can significantly improve prediction performance compared to state-of-the-art text-based models.

Introduction

Predicting the risk of public companies is a fundamental task in investment and risk management. Recent studies from financial economics have shown that the information content in a company's quarterly earnings conference call can forecast the company's future risk level. A conference call, however, is a long document and usually contains thousands of words, which poses challenges for the text-based risk forecast. Most of the prior studies (Qin and Yang 2019; Theil, Broscheit, and Stuckenschmidt 2019) consider the long conference call transcript as one single document. However, the structural relationships of the conference call dialogues are ignored, limiting their effectiveness in modeling the transcript-based and forecasting the risk.

In this work, we leverage such structural information to explicitly model the cross-reference correlations. In particular, we decompose the long earnings call into two parts (i.e., *Presentation* and *Q&A*) and extract information separately. For the *Presentation* learning, we calculate the occurrence of words between sentences and utilize TextRank (Mihalcea and Tarau 2004) to find the key sentences. As for the *Q&A*, we split it into several question-answer pairs and exploit hypergraph to model the cross-reference. Experiments conducted on real-world datasets show that our model outperforms the strong baselines.

*These authors contributed equally.

†Corresponding author.

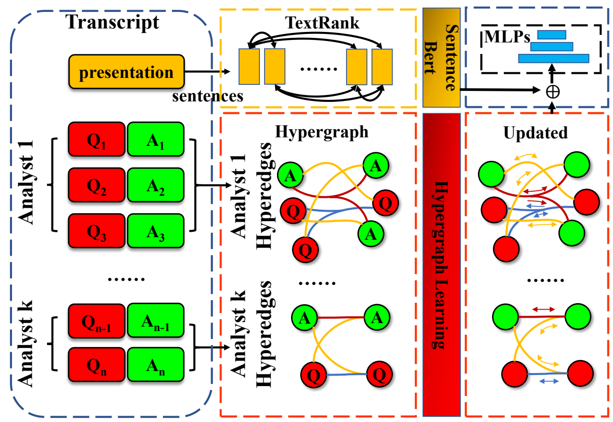

Figure 1: Overview of the proposed model architecture.

Methodology

Earnings call. An earnings call contains two sections: a *Presentation* section and a *Q&A* section. During the Presentation section, managers such as CEOs and CFOs provide information and interpretation of the company performance during the quarter. During the Q&A section, analysts request managers to clarify information and solicit additional information that the management team does not disclose in the Presentation section. There are usually multiple rounds of question-and-answers in a call.

Risk measure. Risk in the financial market is often measured as the volatility of an underlying asset price. We follow (Ye, Qin, and Xu 2020), and define the risk between days t and $t+\tau$ as:

$$v_{[t,t+\tau]} = \ln \sqrt{\frac{1}{\tau - 1} \sum_{i=0}^{\tau} (r_{t+i} - \bar{r})^2}, \quad (1)$$

where $r_t = \frac{p_t}{p_t - 1} - 1$ is the return of a stock with price p_t on day t, and $\bar{r}$ is the average return during the period.

Decompose transcript via structural information. Figure 1 shows the overall architecture of our proposed method.

Firstly, we decompose the long earnings call into two parts (*Presentation* and *Q&A*).

Extract information from *Presentation*. The basic assumption of representation learning in the presentation section is that "a few key sentences in the text store sufficient and necessary information to fulfill the most NLP tasks". More specifically, we determine the key sentence via TextRank. The overlap of two sentences can be determined simply as the number of common tokens between the lexical representations of the two sentences:

$$\mathrm{Sim}(S_i, S_j) = \frac{|\{w_k | w_k \in S_i \bigcap w_k \in S_j\}|}{\log(|S_i|) + \log(|S_j|)}, \quad (2)$$

where S_i and S_j are two different sentences in the presentation section and w_k denotes a word in sentences. After the ranking algorithm, the top-ranked sentences are selected as key sentences. Then, we use Bert to encode them and apply the average pooling to get the final representation of the presentation section.

Extract information from *Q&A*. In light of the special structure of earnings call, we decompose the *Q&A* into multiple question-answer pairs. Then we consider each question/answer as a vertex and use hyperedge to connect the same analysts' questions and corresponding answers. The vertex representation can be updated via message passing:

$$\mathbf{z}^{t+1} = \sigma\left(\mathbf{D}_v^{-1}\mathbf{H}\mathbf{W}\mathbf{D}_e^{-1}\mathbf{H}^\top \mathbf{z}^t; \mathbf{\Theta}\right), \quad (3)$$

where $\mathbf{H}$ is the incidence matrix, $\mathbf{W}$ is the diagonal matrix of weights, $\mathbf{D}_v$ and $\mathbf{D}_e$ are respectively the diagonal matrices of vertex degrees and hyperedge degrees, and $\mathbf{\Theta}$ denotes the model parameters.

Prediction. We forecast the risk via a three-layer MLP:

$$\hat{y}_m = \mathrm{MLPs}(\overline{\mathbf{z}}_m^p, \overline{\mathbf{z}}_m^{qa}), \quad (4)$$

where $\overline{\mathbf{z}}_{\mathbf{m}}^{\mathbf{p}}$ and $\overline{\mathbf{z}}_{\mathbf{m}}^{\mathbf{qa}}$ denote the representations of *Presentation* and *Q&A*, respectively. Mean Squared Error (MSE) is used as the objective to train the model.

Experiments

Dataset. We collect a large-scale earnings transcripts dataset of U.S. firms in four fiscal years (2015-2018). The transcripts are available at SeekingAlaph website[1] or databases such as Thomson Reuters StreetEvents[2]. The breakdown of the earnings transcripts dataset is presented in Table 1. We choose data from 2015 and 2016 to train the model. Data from 2017 and 2018 are respectively served as the validation and testing set.

Baselines. We select LDA-SVR (Blei, Ng, and Jordan 2003), BERT-SVR (Devlin et al. 2019), ProFET (Theil, Broscheit, and Stuckenschmidt 2019) and MRQA (Ye, Qin, and Xu 2020) as our baselines.

Experimental settings. We consider the regression metrics MSE and Mean Absolute Error (MAE) for evaluation. The top-ranked 20% of sentences in the presentation section will be served as key sentences. For hypergraph learning, we set the hidden size to 384. For risk forecast, the hidden size of each layer is set to 500, 200, and 100. We tune the hyperparameters for all methods systematically with 30 epochs.

[1]https://seekingalpha.com

[2]https://www.streetevents.com

Property	**2015**	**2016**	**2017**	**2018**
# transcripts	10,168	9,765	10,431	11,147
# firms	3,398	3,427	3,451	3,616
avg. # tokens in presentation	3,207	3,172	3,191	3,199
avg. # tokens in QA	4,347	4,197	4,222	4,245

Table 1: Statistic of the Dataset.

Method	**MSE**			**MAE**		
	10d	20d	60d	10d	20d	60d
LDA-SVR	0.3146	0.2436	0.1891	0.4371	0.3830	0.3272
BERT-SVR	0.3070	0.2340	0.1826	0.4295	0.3728	0.3207
ProFET	0.3343	0.2585	0.2130	0.4485	0.3927	0.3574
MRQA	0.3025	0.2311	0.1792	0.4259	0.3699	0.3170
Ours	**0.2968**	**0.2293**	**0.1753**	**0.4225**	**0.3669**	**0.3149**

Table 2: Performance comparisons.

Performance comparison. The main results are presented in Table 2, and we have the following findings. First, ProFET simply considers the dialogue structure of earnings conference calls and models the *Presentation* and *Q&A* by a BiLSTM and an attention module, respectively, but it achieves the worst performance. Compared to ProFET, MRQA captures the interactions between question and answer within a question-answer pair, demonstrating the importance of cross-reference correlations. On the basis of MRQA, our method exploits a hypergraph to encode the complex dialog dependency, allowing us to further investigate the high-order cross-reference between different question-answer pairs. As a result, our approach consistently outperforms the baselines across different time horizons.

Acknowledgments

This work was supported in part by National Natural Science Foundation of China (Grant No.62072077 and No.62176043) and Natural Science Foundation of Sichuan Province (Grant No. 2022NSFSC0505).

References

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *JMLR*, 3: 993–1022.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *NAACL-HLT*, 4171–4186. ACL.

Mihalcea, R.; and Tarau, P. 2004. Textrank: Bringing order into text. In *EMNLP*, 404–411.

Qin, Y.; and Yang, Y. 2019. What you say and how you say it matters: Predicting stock volatility using verbal and vocal cues. In *ACL*, 390–401. ACL.

Theil, C. K.; Broscheit, S.; and Stuckenschmidt, H. 2019. PRoFET: Predicting the Risk of Firms from Event Transcripts. In *IJCAI*, 5211–5217. ijcai.org.

Ye, Z.; Qin, Y.; and Xu, W. 2020. Financial Risk Prediction with Multi-Round Q&A Attention Network. In *IJCAI*, 4576–4582. ijcai.org.