Explainable Earnings Call Representation Learning (Student Abstract)

Yanlong Huang[1], **Yue Lei**[1], **Wenxin Tai**[1,3], **Zhangtao Cheng**[1,3,*], **Ting Zhong**[1,3], **Kunpeng Zhang**[2]

[1]University of Electronic Science and Technology of China, Chengdu, Sichuan 610054, China
[2]University of Maryland, College Park, MD 20742, USA
[3]Kash Institute of Electronics and Information Industry, Kashgar 844000, China
hyloong77@gmail.com, leiyue828@gmail.com, wxtai@std.uestc.edu.cn, zhangtao.cheng@outlook.com, zhongting@uestc.edu.cn, kpzhang@umd.edu

Abstract

Earnings call transcripts hold valuable insights that are vital for investors and analysts when making informed decisions. However, extracting these insights from lengthy and complex transcripts can be a challenging task. The traditional manual examination is not only time-consuming but also prone to errors and biases. Deep learning-based representation learning methods have emerged as promising and automated approaches to tackle this problem. Nevertheless, they may encounter significant challenges, such as the unreliability of the representation encoding process and certain domain-specific requirements in the context of finance. To address these issues, we propose a novel transcript representation learning model. Our model leverages the structural information of transcripts to effectively extract key insights, while endowing model with explainability via variational information bottleneck. Extensive experiments on two downstream financial tasks demonstrate the effectiveness of our approach.

Introduction

An earnings call is a conference call in which the management team of a public firm, including executives, communicates with analysts, investors, and journalists. Due to the extensive length of the transcripts and the specialized knowledge necessary for analysis, many finance professionals find it difficult and time-consuming to understand and extract key information. Moreover, analyzing transcripts manually is susceptible to biases and errors, which can lead to inaccurate identification of crucial information. Nowadays, deep learning-based models have achieved considerable success in obtaining text representations. However, the black-box nature makes the high-dimensional encoding process often opaque and unreliable. In addition, the finance domain may present unique requirements and complexities that may necessitate specialized approaches beyond what standard representation learning models can provide.

Inspired by human cognitive theory (Baddeley 1992), we propose a two-step learning process to tackle these challenges. In the first step, we develop a key sentence extractor that emulates human behavior to extract pivotal insights from lengthy earnings call transcripts. In the second step, we employ these "essential sentences" with any off-the-shelf language model to generate the representation. Consequently, the distilled key sentences can offer succinct yet comprehensive explanations that aid humans in understanding which sentences influence the creation of the final representation by the model, thereby improving the algorithm's explainability. Experiments conducted on risk forecasting and information retrieval tasks show that our model outperforms several strong baselines.

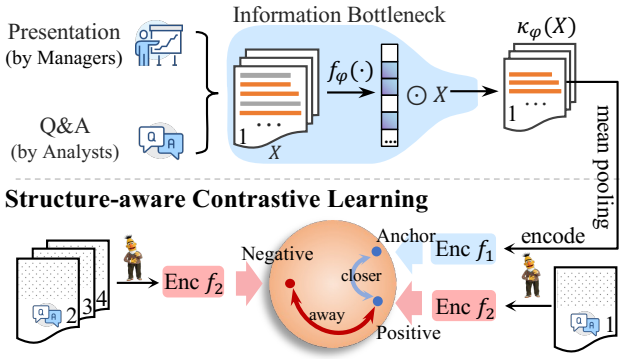

Figure 1: Overview of the proposed model architecture.

Methodology

The primary objective of this work is to develop a neural network that projects each transcript X into a dense d-dimensional vector. A good representation should encapsulate crucial information that can be used for downstream financial tasks. Fig. 1 shows the overview of our method.

Generally, key sentences in earnings call transcripts provide the most important information about the company's financial performance and future prospects, such as revenue growth, earnings per share, cash flow, etc. Therefore, we can devise a neural network as key sentence extractor to find key sentences. In practice, a key sentence extractor can be implemented via $\kappa_\psi(X) = f_\psi(X) \odot X$ where $\kappa_\psi(X)$ is the selected sentence set by neural networks. We use $\odot$ as a key sentence selector, i.e., we put sentences X_i into $\kappa_\psi(X)$ if $f_\psi(X_i)$ equals 1. To guarantee the conciseness of the ex-

*Corresponding Author (zhangtao.cheng@outlook.com).

tracted sentence set, we incorporate the information bottleneck (IB) theory into the training objective:

$$\psi^* = \arg\min_{\psi} \mathcal{L}(g(\kappa_\psi(X)), Y) + \beta\, I(X, \kappa_\psi(X)),$$

where $g(\cdot)$ denotes a function that maps the extracted sentences to the label space, and β is a coefficient that achieves a trade-off between knowledge sufficiency and information compression. We employ a variational approximation (Kim et al. 2021) to the second term:

$$I(X, \kappa_\psi(X)) \leq \mathbb{E}\left[D_{KL}(\mathbb{P}_\psi(M_s|X), r(M_s))\right],$$

where $M_s \in \mathbb{R}^N$ is the output of $f_\psi(X)$, and $r(M_s)$ is the prior distribution of the mask M_s. Given a transcript X with N sentences in total, we select N_s sentences via a predefined ratio α ($N_s = \alpha N$) and use a uniform distribution as the prior of $r(M_s)$.

Optimizing ψ^* is impractical due to the difficulties in acquiring the ground-true key sentence set. To this end, we propose a novel self-supervised learning approach that harnesses the inherent structure of transcripts to generate customized supervision signals ideal for financial analysis.

Generally, in the Q&A section, analysts ask follow-up questions and request the executives to clarify information mentioned in the Presentation section. The crucial information is also what investors pay attention to (Chen, Nagar, and Schoenfeld 2018). Therefore, we can turn to optimize:

$$\begin{aligned}\psi^* &:= \arg\max_{\psi} I(\kappa_\psi(X), X_{QA}) \\ &= \mathbb{E}_{\mathbb{P}(X_{QA}, \kappa_\psi(X))}\left[\log \frac{\mathbb{P}\left(X_{QA} \mid \kappa_\psi(X)\right)}{\mathbb{P}\left(X_{QA}\right)}\right], \quad (1)\end{aligned}$$

where X_{QA} is the set of sentences in the Q&A section. We feed each Q&A round's texts into BERT and use the mean representation across rounds as the representation of the whole Q&A section. Then we use infoNCE to estimate the lower bound of Eq. (1):

$$\mathcal{L}_{\text{NCE}} = -\log \frac{\exp(\text{sim}(\kappa_\psi(X), X_{QA})/\tau)}{\sum_{\mathcal{B}} \mathbb{1}_{X'_{QA} \notin X} \exp(\text{sim}(\kappa_\psi(X), X'_{QA})/\tau)},$$

where $\mathcal{B}$ denotes the batch size, $\mathbb{1}$ is a indicator function, and $\text{sim}(\cdot)$ is implemented by the dot product. Finally, our training objective is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{NCE}} + \beta\, D_{KL}(\mathbb{P}_\psi(M_s|X), r(M_s)).$$

Experiments

Dataset. We have collected an extensive dataset of earnings call transcripts from U.S. firms, which is available through sources like the SeekingAlpha website and databases such as Thomson Reuters StreetEvents. We have selected transcripts from four fiscal years (2015-2018) where we designate the years 2015-2016 as our training dataset, 2017 for validation, and 2018 for testing purposes.

Baselines. We select BERT, SimCSE (Gao, Yao, and Chen 2021), Profet (Theil, Broscheit, and Stuckenschmidt 2019) and MR-QA (Ye, Qin, and Xu 2020) for risk forecasting, and LexRank, TextRank, and PMI (Padmakumar and He 2021) for information retrieval.

Metric	BERT	Profet	MR-QA	Ours
MSE(3d)	<u>0.7401</u>	0.8058	0.7868	**0.7371**
MSE(15d)	0.2650	0.3272	**0.2561**	<u>0.2633</u>
MSE(60d)	0.1826	0.2130	<u>0.1792</u>	**0.1789**
MAE(3d)	<u>0.6724</u>	0.6947	0.7022	**0.6711**
MAE(15d)	<u>0.3981</u>	0.4544	**0.3888**	0.3983
MAE(60d)	0.3207	0.3574	**0.3170**	<u>0.3177</u>

Table 1: Performance comparisons on risk forecasting.

Metric	LexRank	TextRank	PMI	Ours
Precision	0.6400	0.8380	<u>0.9300</u>	**0.9400**
MR	4.6800	2.4020	**2.1020**	<u>2.4020</u>
MRR	0.7274	0.8821	<u>0.9595</u>	**0.9602**

Table 2: Performance comparisons on information retrieval. The indicators MR and MRR are Mean Rank and Mean Reciprocal Rank, respectively.

Risk forecasting. We apply a 3-layer MLP to forecast volatility and use Mean Square Error (MSE) and Mean Absolute Error (MAE) as evaluation metrics. The hyperparameter β is set to 0.1 and the learning rate is 0.001. Table 1 shows our model can achieve forecasting performance on par with language models and risk forecasting models even at a high information compression ratio (α=0.6). This highlights the exceptional ability of our model to extract risk-relevant information. We attribute this phenomenon to the use of our self-supervised paradigm and IB mechanism, which allows it to effectively filter out irrelevant noises.

Information retrieval. Information retrieval facilitates a comparative assessment of the performance in extracting pivotal information. We use cosine similarity to evaluate the transcript representations derived from the information extracted by benchmarks, using three metrics: Precision, Mean Rank, and mean reciprocal rank. As presented in Table 2, the outcomes underscore the outstanding performance of our model across all these metrics. This demonstrates the capability of our model to distill the pivotal elements of pertinent information from transcripts, even when subjected to high compression rates (α=0.15).

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant No.62176043 and No.62072077) and Kashgar Science and Technology Bureau (Grant No.KS2023025).

References

Baddeley, A. 1992. Working memory. *Science*, 255(5044): 556–559.

Chen, J. V.; Nagar, V.; and Schoenfeld, J. 2018. Manager-analyst conversations in earnings conference calls. *Review of Accounting Studies*, 23: 1315–1354.

Gao, T.; Yao, X.; and Chen, D. 2021. Simcse: Simple contrastive learning of sentence embeddings. In *EMNLP*, 6894–6910.

Kim, J.; Kim, M.; Woo, D.; and Kim, G. 2021. Drop-bottleneck: Learning discrete compressed representation for noise-robust exploration. *arXiv:2103.12300*.

Padmakumar, V.; and He, H. 2021. Unsupervised extractive summarization using pointwise mutual information. In *EACL*, 2505–2512.

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

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