Technical Report for WAIC Challenge of Financial QA under Market Volatility

Meng Cao[*1], Ji Jiang[*1], Qichen Ye[*1], Yuexian Zou[†1,2]
[1]School of Electronic and Computer Engineering, Peking University [2]Peng Cheng Laboratory

Abstract

This technical report presents the 1st winning model for Financial Community Question-and-Answering (FCQA), which is a task newly introduced in the Challenge of Financial QA under Marker Volatility in WAIC 2022. FCQA aims to respond to the user's queries in the financial forums with the assistance of heterogeneous knowledge sources. We address this problem by proposing a graph transformer based model for the efficient multi-source information fusion. As a result, we won the first place out of 4278 participating teams and outperformed the second place by 5.07 times on BLUE.

1. Introduction

Community Question Answering [12, 13, 19, 22] (CQA) is a well-defined task that aim to respond to user's queries timely and improve the experience in various platforms including software user communities, interest groups, *etc*. It has potential applications in various downstream tasks, including video understanding [6, 25–27], multi-modal analysis [1, 4, 5, 9, 14], content generation [2, 3, 7, 8, 10, 15, 23], recommendation system [24, 25, 29], *etc*. The Financial Community Question-and-Answering (FCQA) is a new challenge introduced in WAIC[1]. As shown in Figure 1, FCQA focuses on the Q&A tasks in financial scenarios. Compared to the general scenarios, FCQA has the following two difficulties: 1) Marker volatility. The financial data fluctuate over time; 2) High dependence on domain knowledge and expertise. The user questions often require domain expertise for better understanding.

In this challenge, besides the question-and-answer pair, four different types of information sources are also provided including 1) user articles, 2) article comments, 3) related questions and 4) their answers.

To solve FCQA, we propose a novel graph transformer based method to effectively leverage the heterogeneous information source. Benefitting from this, our method achieves the first place among all the competing teams.

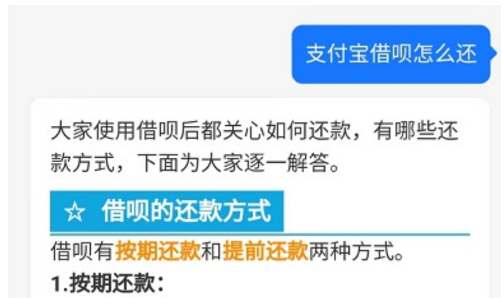

Figure 1. Example of Financial Community Question-and-Answering (FCQA).

Notations	Definition
q	Question
d	Related document
q_s	Related question
a_s	Related answer
c	Related comment

Table 1. The referred notations.

2. Our Model

In this section, we present our graph transformer based approach. As shown in Figure 2, our method has three major components.

2.1. Multi-source Encoder

We employ BART [17] to encode all the source information (*i.e*., questions, source information, and answers). Specifically, we represent each node with the output hidden state of the BART encoder.

Formally, the definitions of some notations referred to in this section are summarized in Table 1.

2.2. Graph Transformer

The Graph Neural Network (GNN) [28] has achieved promising results in QA tasks [11, 16] due to its ability of

∗ denotes the equal contributions. † denotes the corresponding author.
[1]https : / / tianchi . aliyun . com / competition / entrance/532010/introduction

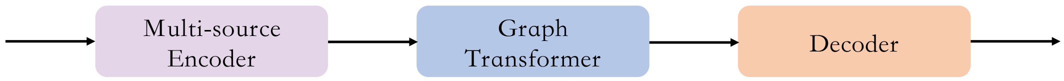

Figure 2. Pipeline of our method.

message passing over nodes. Therefore, we also resort to the graph transformer architecture to aggregate question-related information from different types of input.

Node Representation: We regard the question-answer pair as well as the multiple source information as the specific nodes. All of them are embedded by BART [17] as stated before.

Edge Construction: Intuitively, we build up four types of edges to connect all the nodes: 1) related document to question $\langle d, q \rangle$; 2) related question to question $\langle q_s, q \rangle$; 3) related answer to related question $\langle a_s, q_s \rangle$; 4) related comment to related document $\langle c, d \rangle$.

Question-aware Aggregation: Based on the established edges, we aim to fuse the information in neighbor nodes into the target node representation. Here we employ a vanilla attention mechanism [21]. Firstly, we project the input node representation into two spaces.

$$\begin{aligned} \boldsymbol{p_s} &= \mathrm{MLP}(\mathrm{feat}(\boldsymbol{s})), \\ \boldsymbol{p_t} &= \mathrm{MLP}(\mathrm{feat}(\boldsymbol{t})), \end{aligned} \tag{1}$$

where $\boldsymbol{p_s}$ and $\boldsymbol{p_t}$ denote the source and target nodes, respectively.

Secondly, we calculate the relevance between each node pair as the attention scores.

$$\alpha(s, e) = \mathrm{Softmax}\left(\boldsymbol{p_e} \cdot W \cdot \boldsymbol{p_s}\right), \tag{2}$$

where W is the learnable parameter.

Thirdly, we also incorporate the edge type into the attention score calculation.

$$M(s, e) = \mathrm{MLP}\left(\mathrm{feat}(\boldsymbol{s}) \cdot W^{msg}\right), \tag{3}$$

where W^{msg} is the edge type-specific parameter matrix.

Lastly, we apply a weighted sum over all the messages passed from all the source nodes.

$$\mathrm{feat}(t) = \sum \alpha(s, e) \cdot M(s, e). \tag{4}$$

2.3. Decoder

We employ the pre-trained BART decoder as our answer generator. To incorporate both the input question and the multi-source knowledge simultaneously, we compute the attentive scores as follows.

$$\begin{aligned} \boldsymbol{p_{out}} = &\ \mathrm{MHA}\left(\mathrm{feat(i)}, \boldsymbol{q_s}\right) + \mathrm{MHA}\left(\mathrm{feat(i)}, \boldsymbol{a_s}\right) \\ &\ \mathrm{MHA}\left(\mathrm{feat(i)}, \boldsymbol{d}\right) + \mathrm{MHA}\left(\mathrm{feat(i)}, \boldsymbol{c}\right). \end{aligned} \tag{5}$$

where MHA denotes the multi-head attention mechanism.

We employ the text generation cross-entropy loss as our final loss function.

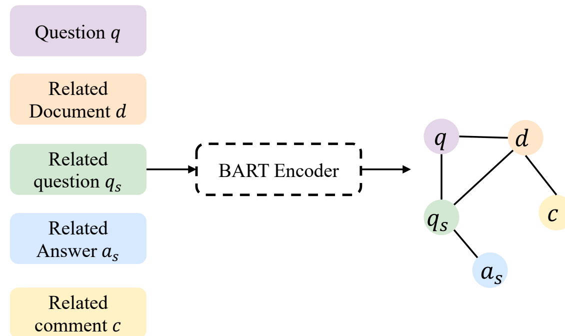

Figure 3. The design of our graph transformer.

Method	BLEU	BLEU@1	@2	@3	@4
Ours	12.20	23.02	12.82	9.45	7.95
Runner-up	2.01	6.66	2.03	1.20	1.00

Table 2. Comparison results (%).

3. Experiment

3.1. Dataset and Metrics

The provided dataset contains 376,948, 1,920, and 2,409 samples for the training, validation, and testing split, respectively.

We report five common automatic metrics: BLEU and BLEU@k ($k = 1, 2, 3, 4$) [18].

3.2. Implementation Details

All of our experiments are conducted in eight NVIDIA V100 GPUs. The pre-trained BART-Chinese-base [20] model is employed to initialize the encoder and decoder.

3.3. Results

The comparison results in the challenge are listed in Table 2. All results are presented on a percentage scale. As shown, our method outperforms the runner-up by 5.07 times on BLUE (12.20 *v.s.* 2.01).

4. Conclusion

In this paper, we address the Financial Community Question-and-Answering (FCQA) task. Accordingly, we propose a graph transformer based model to extract and align multi-source information. We achieved the inspiring performance and won first place out of 4278 participating teams.

5. Acknowledgement

This paper was partially supported by NSFC (No: 62176008) and Shenzhen Science & Technology Research Program (No: GXWD20201231165807007-20200814115301001).

References

[1] Meng Cao, Long Chen, Mike Zheng Shou, Can Zhang, and Yuexian Zou. On pursuit of designing multi-modal transformer for video grounding. *arXiv preprint arXiv:2109.06085*, 2021. 1

[2] Meng Cao, Haozhi Huang, Hao Wang, Xuan Wang, Li Shen, Sheng Wang, Linchao Bao, Zhifeng Li, and Jiebo Luo. Task-agnostic temporally consistent facial video editing. *arXiv preprint arXiv:2007.01466*, 2020. 1

[3] Meng Cao, Haozhi Huang, Hao Wang, Xuan Wang, Li Shen, Sheng Wang, Linchao Bao, Zhifeng Li, and Jiebo Luo. Unifacegan: A unified framework for temporally consistent facial video editing. *IEEE Transactions on Image Processing*, 30:6107–6116, 2021. 1

[4] Meng Cao, Ji Jiang, Long Chen, and Yuexian Zou. Correspondence matters for video referring expression comprehension. *arXiv preprint arXiv:2207.10400*, 2022. 1

[5] Meng Cao, Tianyu Yang, Junwu Weng, Can Zhang, Jue Wang, and Yuexian Zou. Locvtp: Video-text pre-training for temporal localization. *arXiv preprint arXiv:2207.10362*, 2022. 1

[6] Meng Cao, Can Zhang, Long Chen, Mike Zheng Shou, and Yuexian Zou. Deep motion prior for weakly-supervised temporal action localization. *IEEE Transactions on Image Processing*, 31:5203–5213, 2022. 1

[7] Meng Cao, Can Zhang, Dongming Yang, and Yuexian Zou. All you need is a second look: Towards arbitrary-shaped text detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(2):758–767, 2021. 1

[8] Meng Cao and Yuexian Zou. All you need is a second look: Towards tighter arbitrary shape text detection. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2228–2232. IEEE, 2020. 1

[9] Meng Cao, Yuexian Zou, Dongming Yang, and Chao Liu. Gisca: Gradient-inductive segmentation network with contextual attention for scene text detection. *IEEE Access*, 7:62805–62816, 2019. 1

[10] Dading Chong, Helin Wang, Peilin Zhou, and Qingcheng Zeng. Masked spectrogram prediction for self-supervised audio pre-training. *arXiv preprint arXiv:2204.12768*, 2022. 1

[11] Yuwei Fang, Siqi Sun, Zhe Gan, Rohit Pillai, Shuohang Wang, and Jingjing Liu. Hierarchical graph network for multi-hop question answering. *arXiv preprint arXiv:1911.03631*, 2019. 1

[12] Jinlan Fu, Yi Li, Qi Zhang, Qinzhuo Wu, Renfeng Ma, Xuanjing Huang, and Yu-Gang Jiang. Recurrent memory reasoning network for expert finding in community question answering. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 187–195, 2020. 1

[13] Shen Gao, Yuchi Zhang, Yongliang Wang, Yang Dong, Xiuying Chen, Dongyan Zhao, and Rui Yan. Heteroqa: Learning towards question-and-answering through multiple information sources via heterogeneous graph modeling. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 307–315, 2022. 1

[14] Sixin Hong, Yuexian Zou, Wenwu Wang, and Meng Cao. Weakly labelled audio tagging via convolutional networks with spatial and channel-wise attention. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 296–300. IEEE, 2020. 1

[15] Puzhao Ji, Meng Cao, and Yuexian Zou. Visual relation-aware unsupervised video captioning. In *International Conference on Artificial Neural Networks*, pages 495–507. Springer, 2022. 1

[16] Pin Jiang and Yahong Han. Reasoning with heterogeneous graph alignment for video question answering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 11109–11116, 2020. 1

[17] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*, 2019. 1, 2

[18] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318, 2002. 2

[19] Andreas Rücklé, Krishnkant Swarnkar, and Iryna Gurevych. Improved cross-lingual question retrieval for community question answering. In *The world wide web conference*, pages 3179–3186, 2019. 1

[20] Yunfan Shao, Zhichao Geng, Yitao Liu, Junqi Dai, Fei Yang, Li Zhe, Hujun Bao, and Xipeng Qiu. Cpt: A pre-trained unbalanced transformer for both chinese language understanding and generation. *arXiv preprint arXiv:2109.05729*, 2021. 2

[21] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 2

[22] Wei Wu, Xu Sun, and Houfeng Wang. Question condensing networks for answer selection in community question answering. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1746–1755, 2018. 1

[23] Wangduo Xie, Jian Zhang, Zhisheng Lu, Meng Cao, and Yong Zhao. Non-local nested residual attention network for

stereo image super-resolution. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2643–2647. IEEE, 2020. 1

[24] Yueqi Xie, Peilin Zhou, and Sunghun Kim. Decoupled side information fusion for sequential recommendation. *arXiv preprint arXiv:2204.11046*, 2022. 1

[25] Dongming Yang, Yuexian Zou, Can Zhang, Meng Cao, and Jie Chen. Rr-net: Relation reasoning for end-to-end human-object interaction detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021. 1

[26] Can Zhang, Meng Cao, Dongming Yang, Jie Chen, and Yuexian Zou. Cola: Weakly-supervised temporal action localization with snippet contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16010–16019, 2021. 1

[27] Can Zhang, Meng Cao, Dongming Yang, Ji Jiang, and Yuexian Zou. Synergic learning for noise-insensitive webly-supervised temporal action localization. *Image and Vision Computing*, 113:104247, 2021. 1

[28] Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *AI Open*, 1:57–81, 2020. 1

[29] Peilin Zhou, Zhiqi Huang, Fenglin Liu, and Yuexian Zou. Pin: A novel parallel interactive network for spoken language understanding. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 2950–2957. IEEE, 2021. 1