Upaya at the FinLLM Challenge Task 1 and 2: DistFin: Distillation based Fine-Tuning for Financial Tasks

Ashvini Kumar Jindal
LinkedIn AI USA

Pawan Kumar Rajpoot
SCB DataX Thailand

Ankur Parikh
UtilizeAI Research India

Abstract

With the advent of Large Language Models (LLM) in finance, financial text analysis and generation tasks have received growing attention. Financial text classification and financial text summarization are some of the very important text analysis and generation tasks, respectively. Adapting LLMs to these tasks is very crucial for domain adaptation. This paper presents a method to fine-tune LLMs to Financial Argument Classification and Financial Abstractive Summarization. The argument classification task focuses on argument unit classification to test the capabilities of LLMs to identify and categorize texts as premises or claims. The summarization task aims to abstract financial texts into concise summaries. The dataset was released along with shared tasks as a part of the 8th Financial Technology and Natural Language Processing (FinNLP), co-located with IJCAI 2024. In this paper, we employed a distillation-based fine-tuning of Llama-3 (8B parameters) to learn the rationale/step generated by Llama-3 (70B parameters) along with labels. In the argument classification task, we achieved an F1-score (evaluation metric) of 0.4166. In the summarization task, we got the 2nd rank with the Rouge-1 score (evaluation metric) of 0.5294.

1 Introduction

Recently, Large Language Models (LLMs) (Brown et al., 2020) such as GPT-2 and GPT-4 (OpenAI et al., 2024), have reshaped the field of natural language processing (NLP) and exhibited remarkable capabilities in specialized domains across mathematics, coding, medicine, law, and finance (Bubeck et al., 2023). Within the financial domain, recent several studies (Xie et al., 2023a; Lopez-Lira and Tang, 2023; Li et al., 2023; Xie et al., 2023b) have shown the great potential of advanced LLMs such as GPT-4 on financial text analysis/prediction and generation tasks. Examples of financial text analysis tasks are sentiment analysis, news headline classification, hawkish-dovish classification, argument unit classification, argument relation classification, ESG issue identification, deal completeness classification, etc. and instances of financial text generation tasks are text summarization, financial report generation, etc. (Xie et al., 2024). This paper focuses on one such analysis task - argument unit classification and one such generation task - abstractive summarization.

The primary objective of the argument unit classification is to categorize argumentative sentences found in earnings conference call text into 'claim' and 'premise' classes (Sy et al., 2023). This classification is a foundational step, enabling a granular breakdown of financial narratives. The precision in isolating these units paves the way for deeper comprehension and subsequent analysis. Recognizing the distinct units of arguments means that investors and stakeholders can better interpret the sentiments conveyed in these financial discussions. (Sy et al., 2023) employed voting-based ensemble of various fine-tuned language models such as BERT (Devlin et al., 2019), ROBERTA (Liu et al., 2019), ETCETRA (Araci, 2019), FINBERT (Clark et al., 2020) etc.

The summarization task aims to abstract financial texts into concise summaries. Summarizing news articles is very useful in trading strategies. By shaping investors' perceptions and assessments of companies, financial news significantly impacts the stock market (Engle and Ng, 1993; Tetlock, 2007). News-based stock prediction models are thus developed to automatically discover signals of stock market movements from the countless news articles that are generated every moment (Kalyani et al., 2016; Shah et al., 2018; Mohan et al., 2019; Zhou et al., 2021). Summaries of company 10-q and 10-k reports also form the working memory of an LLM-based trading agent (Yu et al., 2023). The other summarization task focuses on earnings call transcripts (Mukherjee et al., 2022). (Yang et al.,

Proceedings of the Joint Workshop of the 8th Financial Technology and Natural Language Processing (FinNLP) and the 1st Agent AI for Scenario Planning (AgentScen), Jeju, South Korea, August 3, 2024

2023) has instruction-tuned Llama-65B (Touvron et al., 2023) on various financial tasks including financial summarization task.

In this paper, we utilized a distillation-based fine-tuning of Llama-3 (8B parameters) (AI@Meta, 2024) to learn the rationale/step generated by Llama-3 (70B parameters) (AI@Meta, 2024) along with labels. For the argument unit classification task, we prompted Llama-3 (70B) to generate a rationale for the given argumentative sentence and label premise, claim pair. We prompted Llama-3 (70B) to identify the main ideas/sentences given the financial news text and summary pair. We achieved F1-score (evaluation metric) of 0.4166 in the argument classification task. In the summarization task, we got the 2nd rank with the Rouge-1 score (evaluation metric) of 0.5294. Both of our models are available on HuggingFace [1]

2 Preliminary Background

2.1 Argument Unit Classification

2.1.1 Task

Given an input argumentative sentence S, the objective is to adopt an LLM M that accurately categorizes S into either the argument unit A={claim, premise} class.

2.1.2 Data

The dataset released with this task contains 7.75k and 969 data points in training and test data, respectively. These data points represent financial text along with labels - premise and claim.

2.1.3 Evaluation

The prompt template used to evaluate the LLM submission is - Instruction: [task prompt] Text: [input text] Response: [output]. The instruction is - 'Analyze sentences from earnings conference calls and identify their argumentative function. Each sentence is either a premise, offering evidence or reasoning, or a claim, asserting a conclusion or viewpoint. Return only premise or claim'. The evaluation metric is F1-score.

2.2 Abstractive Summarization

2.2.1 Task

Given an input financial news text T, the task is to adapt an LLM M that accurately generates an abstractive summary A.

[1]https://huggingface.co/upaya07/finnlp_task_1, https://huggingface.co/upaya07/finnlp_task_2

Table 1: Prompt for generating rationale for Argument Unit Classification task

```
## Task
We are working on a Text, which is from earnings conference calls and identify their argumentative function. This text can be classified as either a premise or a claim. A premise sentence offers evidence or reasoning, while a claim sentence asserts a conclusion or viewpoint. Analyse following sentence and assume that you secretly know the provided Answer, write a clear one or two line max reasoning that concludes provided with final Answer. Return only the Reasoning part.

## Text
"""{Text}"""

## Reasoning
```

2.2.2 Data

The dataset released with this task contains 8k and 2k data points in training and test data, respectively. These data points represent financial news text along with an abstract summary.

2.2.3 Evaluation

The prompt template used to evaluate the LLM submission is - Instruction: [task prompt] Context: [input context] Response: [output]. The instruction is - 'You are given a text that consists of multiple sentences. Your task is to perform an abstractive summarization of this text. Use your understanding of the content to express the main ideas and crucial details in a shorter, coherent, and natural-sounding text'. The evaluation metric is the ROUGE-1 score.

3 Argument Unit Classification

To add more context to the training data, we prompted the bigger Llama-3 70B model to get rationale behind the gold label. Further, we added this rationale in the training data, reformatted it, and then fine-tuned the smaller Llama-3 8B chat model to generate rationale and an answer/a label based on it.

3.1 Rationale Generation

We used Llama-3-70B-Instruct model to generate rationale. We provide prompt in Table 1.

3.2 Supervised Fine Tuning

We augmented the training data with the generated rationale and LoRA (Hu et al., 2021) fine-tuned the Llama-3 8b Instruct model for generating rationale and the answer in a defined order. We fine-tuned our model for 3 epochs using a Nvidia 4090

Table 2: Example record with reasoning augmented training data for Argument Unit Classification task

Task Analyze sentences from earnings conference calls and identify their argumentative function. Each sentence is either a premise, offering evidence or reasoning, or a claim, asserting a conclusion or viewpoint. Return the Reason first and then Answer that is premise or claim. ## Text So, now that with a SaaS approach, you can reach a much broader base of business customers all over the world, is one opportunity. ## Reason This sentence asserts a conclusion or viewpoint about the opportunity presented by a SaaS approach, rather than providing evidence or reasoning to support a larger argument. Hence claim. ## Answer Claim

Table 3: Results on Argument Unit Classification Task. F1 score is used for final ranking.

Team	Accuracy	F1-Score
Team Barclays	**0.762**	**0.523**
Albatross	<u>0.757</u>	<u>0.517</u>
L3iTC	0.754	0.514
Upaya(ours)	0.709	0.416

GPU system with 40 GB RAM and 24GB VRAM. The hyper-parameters for the fine-tuning are shared along with the model on the huggingface [2]. We provide an example record of reasoning augmented training data in Table 2.

3.3 Results

Overall, we got 7th rank in the task with an accuracy score of around 71 and an f1 score of 41. Table 3 shows results from top-3 teams.

4 Abstractive Summarization

The financial text summarization aims to summarize financial news articles into concise summaries. The task provides 8k training data and 2k test data. Metrics such as such as ROUGE (1, 2, and L) and BERTScore are computed for all submission and ROUGE-1 score is used for the final rankings. For this task, we used the similar approach as the task explained in previous section with few modifications outlined below.

[2]https://huggingface.co/upaya07/finnlp_task_1

Table 4: Results on Financial News Summarization task. ROUGE-1 is used for final ranking.

Team	ROUGE-1	BERTScore
LBZ	**0.535**	**0.912**
Finance Wizard	0.521	0.908
Upaya(ours)	<u>0.529</u>	<u>0.911</u>

4.1 Relevant Sentence Extraction

Our approach is based on an intuition that there are a few sentences in original news article that play an important role in writing a coherent summary. Following the intuition, we prompted Llama-3 70B Instruct model to extract maximum of 5 relevant sentences from the original news text that are relevant to the given summary. Along with prompting the model to extract relevant sentences to the summary, we also prompted it to generate a rationale behind importance of each extracted sentence. This scheme helps to extract relevant sentences conditioned on ground truth summary. We provided prompt in Table 5 that we applied to extract relevant sentences from the training data provided in the task.

4.2 Supervised Fine-tuning

Once relevant sentences are extracted using the approach described in the previous section, original 8k training data is augmented and the new output contains extracted sentences along with summary. Next, we fully fine-tuned Llama-3 (8B parameters) model on top of the augmented training data. Specifically, during fine-tuning, the model takes original news text as input and learns to generate both relevant sentences and final summary. In this work, we did not explore adding rationale for model fine-tuning. The hyper-parameters for the fine-tuning are shared along with the model on the huggingface [3]. We provided the prompt in Table 6 that we applied to fine-tune Llama-3-8B-Instruct model.

4.3 Results

We achieved 2nd rank in Financial News Summarization task with ROUGE-1 score of 0.529. Table 4 shows results from top-3 teams.

[3]https://huggingface.co/upaya07/finnlp_task_2

Table 5: Prompt for extracting relevant sentence(s) from financial text

Task
You are given a financial text under "## Financial Text" section. Assuming that you secretly have access to the summary of the financial text under "## Summary" section, you need to extract maximum 5 relevant sentences from original financial text following below instructions:
- Each extracted sentence should be important and contributes to the given summary.
- Rank relevant sentences in order of high to low importance. Each relevant sentence should contain rationale behind its importance on a scale of 1 to 10 where 1 being least important and 10 being most important.
- Do not modify the original sentence and keep rationale limited to one line only.
- Rationale should not contain phrases that directly or indirectly reveal that you have access to the summary.
- There can be less than 5 relevant sentences, hence, you need to only provide relevant ones instead of always providing 5 sentences.

Financial Text
`"""{Financial Text}"""`

Summary
`"""{{ summary }}"""`

[RESPONSE FORMAT]
Generate response as JSON with following schema. Each entry contains extracted sentence, rationale, and importance score on scale of 1-10.

```
[
    {
        "sentence": <sentence>,
        "rationale": <rationale>,
        "importance": <importance>
    },
]
```

[JSON RESPONSE]

Conclusion

This paper explores distillation based fine-tuning of Llama-3 models for two of the financial tasks: 1. Argument Unit Classification 2. Abstractive Summarization. For the first task, we used Llama-3 70B model to distill the rationale behind the label given the financial sentence and the label pair. In the second task, we prompted Llama-3 70B model to distill main ideas behind the summary given the financial text and summary. In both tasks, we augmented training data with this distilled information and performed instruction-tuning to adapt Llama-3 8B model on these tasks. We achieved F1-score of 0.4166 in the argument classification task. In the summarization task, we got the 2nd rank with the Rouge-1 score of 0.5294.

Table 6: Prompt for training model for financial text summarization task

Task
You are given a financial text under "## Financial Text" section and you need to write a summary of the given text.
- First, extract relevant sentences from the given text that you think are important for summary.
- Next, write a summary focusing on extracted sentences and optionally given text.

Financial Text
`"""{Financial Text}"""`

Acknowledgements

We would like to thank Lambda Labs[4] for providing the compute resources required for this work.

References

AI@Meta. 2024. Llama 3 model card.

Dogu Araci. 2019. Finbert: Financial sentiment analysis with pre-trained language models.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners.

Sébastien Bubeck, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee, Yin Tat Lee, Yuanzhi Li, Scott Lundberg, Harsha Nori, Hamid Palangi, Marco Tulio Ribeiro, and Yi Zhang. 2023. Sparks of artificial general intelligence: Early experiments with gpt-4.

Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. Electra: Pre-training text encoders as discriminators rather than generators.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding.

Robert F Engle and Victor K Ng. 1993. Measuring and testing the impact of news on volatility. *The journal of finance*, 48(5):1749–1778.

[4] https://lambdalabs.com/

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models.

Joshi Kalyani, Prof Bharathi, Prof Jyothi, et al. 2016. Stock trend prediction using news sentiment analysis. *arXiv preprint arXiv:1607.01958*.

Xianzhi Li, Samuel Chan, Xiaodan Zhu, Yulong Pei, Zhiqiang Ma, Xiaomo Liu, and Sameena Shah. 2023. Are chatgpt and gpt-4 general-purpose solvers for financial text analytics? a study on several typical tasks.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach.

Alejandro Lopez-Lira and Yuehua Tang. 2023. Can chatgpt forecast stock price movements? return predictability and large language models.

Saloni Mohan, Sahitya Mullapudi, Sudheer Sammeta, Parag Vijayvergia, and David C Anastasiu. 2019. Stock price prediction using news sentiment analysis. In *2019 IEEE fifth international conference on big data computing service and applications (BigDataService)*, pages 205–208. IEEE.

Rajdeep Mukherjee, Abhinav Bohra, Akash Banerjee, Soumya Sharma, Manjunath Hegde, Afreen Shaikh, Shivani Shrivastava, Koustuv Dasgupta, Niloy Ganguly, Saptarshi Ghosh, and Pawan Goyal. 2022. Ectsum: A new benchmark dataset for bullet point summarization of long earnings call transcripts.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. 2024. Gpt-4 technical report.

Dev Shah, Haruna Isah, and Farhana Zulkernine. 2018. Predicting the effects of news sentiments on the stock market. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 4705–4708. IEEE.

Eugene Sy, Tzu-Cheng Peng, Shih-Hsuan Huang, Heng-Yu Lin, and Yung-Chun Chang. 2023. Fine-grained argument understanding with bert ensemble techniques: A deep dive into financial sentiment analysis. In *Proceedings of the 35th Conference on Computational Linguistics and Speech Processing (ROCLING 2023)*, pages 242–249.

Paul C Tetlock. 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3):1139–1168.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models.

Qianqian Xie, Weiguang Han, Zhengyu Chen, Ruoyu Xiang, Xiao Zhang, Yueru He, Mengxi Xiao, Dong Li, Yongfu Dai, Duanyu Feng, Yijing Xu, Haoqiang Kang, Ziyan Kuang, Chenhan Yuan, Kailai Yang, Zheheng Luo, Tianlin Zhang, Zhiwei Liu, Guojun Xiong, Zhiyang Deng, Yuechen Jiang, Zhiyuan Yao, Haohang Li, Yangyang Yu, Gang Hu, Jiajia Huang, Xiao-Yang Liu, Alejandro Lopez-Lira, Benyou Wang, Yanzhao Lai, Hao Wang, Min Peng, Sophia Ananiadou, and Jimin Huang. 2024. The finben: An holistic financial benchmark for large language models.

Qianqian Xie, Weiguang Han, Yanzhao Lai, Min Peng, and Jimin Huang. 2023a. The wall street neophyte: A zero-shot analysis of chatgpt over multimodal stock movement prediction challenges.

Qianqian Xie, Weiguang Han, Xiao Zhang, Yanzhao Lai, Min Peng, Alejandro Lopez-Lira, and Jimin Huang. 2023b. Pixiu: A large language model, instruction data and evaluation benchmark for finance.

Yi Yang, Yixuan Tang, and Kar Yan Tam. 2023. Investlm: A large language model for investment using financial domain instruction tuning.

Yangyang Yu, Haohang Li, Zhi Chen, Yuechen Jiang, Yang Li, Denghui Zhang, Rong Liu, Jordan W. Suchow, and Khaldoun Khashanah. 2023. Finmem: A performance-enhanced llm trading agent with layered memory and character design.

Zhihan Zhou, Liqian Ma, and Han Liu. 2021. Trade the event: Corporate events detection for news-based event-driven trading.