FinNLP 2022

The Fourth Workshop on Financial Technology and Natural Language Processing

Proceedings of the Workshop

December 8, 2022

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-959429-10-4

Introduction

Welcome to FinNLP-2022, the 4th Workshop on Financial Technology and Natural Language Processing! Since this year, FinNLP has become a twice-a-year workshop and is colocated with IJCAI and EMNLP. This workshop aims to provide a forum for sharing the latest Interdisciplinary results from either the financial domain's or the NLP field's perspective.

In FinNLP-2022, we have a keynote (Knowledge-Based News Event Analysis and Forecasting) from Dr. Oktie Hassanzadeh, a Senior Research Staff Member at IBM T.J. Watson Research Center, and an overview of recent FinNLP studies from the FinNLP organizer. The accepted papers cover various topics, including emerging trend identification, intent classification, market information prediction, sentiment analysis, digital strategy maturity assessment, and so on. Several kinds of financial documents are explored, such as the transcriptions of earnings calls, social media data, and news articles. The shared task participants share several approaches for evaluating the rationales of amateur investors. We hope the audiences of FinNLP-2022 can learn the latest tendency, and also have a comprehensive understanding of where we are now in financial opinion scoring.

FinNLP-2022 is the result of a collaborative effort of the FinNLP community. We would like to thank our Program Committee - Hiroki Sakaji, Emmanuele Chersoni, Kiyoshi Izumi, Pablo Duboue, Juyeon Kang, Paulo Alves, Luciano Del Corro, Chuan-Ju Wang, Ismail El Maarouf, Damir Cavar, Paul Buitelaar, and Jinhang Jiang - for their help in providing feedback on submissions and selecting the papers. We would also like to thank all authors from 10 countries in both academia (16 institutions) and industry (8 companies) for sharing their insightful results in FinNLP-2022.

Welcome and hope you all enjoy FinNLP-2022.

Chung-Chi Chen (AIST, Japan)
Hen-Hsen Huang (Academia Sinica, Taiwan)
Hiroya Takamura (AIST, Japan)
Hsin-Hsi Chen (National Taiwan University, Taiwan)
FinNLP-2022 Organizers

Acknowledgment FinNLP-2022 is supported by a project JPNP20006, commissioned by the New Energy and Industrial Technology Development Organization (NEDO), and was partially supported by National Science and Technology Council, Taiwan, under grants MOST 110-2221-E-002-128-MY3 and MOST 110-2634-F-002-050-.

Organizing Committee

Organizers

Chung-Chi Chen, Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Japan
Hen-Hsen Huang, Institute of Information Science, Academia Sinica, Taiwan
Hiroya Takamura, Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Japan
Hsin-Hsi Chen, Department of Computer Science and Information Engineering, National Taiwan University, Taiwan

Program Committee

Reviewers

Paulo Alves

Paul Buitelaar

Damir Cavar, Emmanuele Chersoni

Luciano Del Corro, Pablo Duboue

Ismail El Maarouf

Kiyoshi Izumi

Jinhang Jiang

Juyeon Kang

Hiroki Sakaji

Chuan-Ju Wang

Keynote Talk: Knowledge-Based News Event Analysis and Forecasting

Oktie Hassanzadeh
IBM T.J. Watson Research Center

Abstract: In this talk, I will present our ongoing work at IBM Research on building a toolkit for news event analysis and forecasting. The toolkit is powered by a Knowledge Graph (KG) of events curated from structured and textual sources of event-related knowledge. The toolkit provides functions for 1) mapping ongoing news headlines to concepts in the KG, 2) retrieval, reasoning, and visualization for causal analysis and forecasting, and 3) extraction of causal knowledge from text documents to augment the KG with additional domain knowledge. Each function has a number of implementations using state-of-the-art neuro-symbolic techniques. I will go over a number of use cases for the toolkit, including use cases in finance and enterprise risk management.

Bio: Dr. Oktie Hassanzadeh is a Senior Research Staff Member at IBM T.J. Watson Research Center. He is the recipient of several academic and corporate awards, including a top prize at the FinCausal-2022 Shared Task, a top prize at the Semantic Web Challenge at ISWC conference, and two best-paper awards at ESWC conferences. He has received his M.Sc. and Ph.D. degrees from the University of Toronto, where he received the IBM PhD fellowship and the Yahoo! Key Scientific Challenges awards. He is also a two-time recipient of the first prize at the Triplification Challenge at the SEMANTiCS Conference for his projects in the areas of Semantic Technologies and Linked Data. For more information, refer to his home page: http://researcher.watson.ibm.com/person/us-hassanzadeh

Table of Contents

Contextualizing Emerging Trends in Financial News Articles
Nhu Khoa Nguyen, Thierry Delahaut, Emanuela Boros, Antoine Doucet and Gaël Lejeune 1

AstBERT: Enabling Language Model for Financial Code Understanding with Abstract Syntax Trees
Rong Liang, Tiehua Zhang, Yujie Lu, Yuze Liu, Zhen Huang and Xin Chen 10

Disentangled Variational Topic Inference for Topic-Accurate Financial Report Generation
Sixing Yan 18

Toward Privacy-preserving Text Embedding Similarity with Homomorphic Encryption
Donggyu Kim, Garam Lee and Sungwoo Oh 25

TweetFinSent: A Dataset of Stock Sentiments on Twitter
Yulong Pei, Amarachi Mbakwe, Akshat Gupta, Salwa Alamir, Hanxuan Lin, Xiaomo Liu and Sameena Shah 37

Stock Price Volatility Prediction: A Case Study with AutoML
Hilal Pataci, Yunyao Li, Yannis Katsis, Yada Zhu and Lucian Popa 48

DigiCall: A Benchmark for Measuring the Maturity of Digital Strategy through Company Earning Calls
Hilal Pataci, Kexuan Sun and T. Ravichandran 58

Learning Better Intent Representations for Financial Open Intent Classification
Xianzhi Li, Will Aitken, Xiaodan Zhu and Stephen W. Thomas 68

Exploring Robustness of Prefix Tuning in Noisy Data: A Case Study in Financial Sentiment Analysis
Sudhandar Balakrishnan, Yihao Fang and Xiaodan Zhu 78

A Taxonomical NLP Blueprint to Support Financial Decision Making through Information-Centred Interactions
Siavash Kazemian, Cosmin Munteanu and Gerald Penn 89

Overview of the FinNLP-2022 ERAI Task: Evaluating the Rationales of Amateur Investors
Chung-Chi Chen, Hen-Hsen Huang, Hiroya Takamura and Hsin-Hsi Chen 99

PromptShots at the FinNLP-2022 ERAI Task: Pairwise Comparison and Unsupervised Ranking
Peratham Wiriyathammabhum 104

LIPI at the FinNLP-2022 ERAI Task: Ensembling Sentence Transformers for Assessing Maximum Possible Profit and Loss from Online Financial Posts
Sohom Ghosh and Sudip Kumar Naskar 111

DCU-ML at the FinNLP-2022 ERAI Task: Investigating the Transferability of Sentiment Analysis Data for Evaluating Rationales of Investors
Chenyang Lyu, Tianbo Ji and Liting Zhou 116

UOA at the FinNLP-2022 ERAI Task: Leveraging the Class Label Description for Financial Opinion Mining
Jinan Zou, Haiyao Cao, Yanxi Liu, Lingqiao Liu, Ehsan Abbasnejad and Javen Qinfeng Shi . 122

aiML at the FinNLP-2022 ERAI Task: Combining Classification and Regression Tasks for Financial Opinion Mining
Zhaoxuan Qin, Jinan Zou, Qiaoyang Luo, Haiyao Cao and Yang Jiao 127

Yet at the FinNLP-2022 ERAI Task: Modified models for evaluating the Rationales of Amateur Investors
Yan Zhuang and Fuji Ren 132

LDPP at the FinNLP-2022 ERAI Task: Determinantal Point Processes and Variational Auto-encoders for Identifying High-Quality Opinions from a pool of Social Media Posts
Paul Trust and Rosane Minghim 136

Jetsons at the FinNLP-2022 ERAI Task: BERT-Chinese for mining high MPP posts
Alolika Gon, Sihan Zha, Sai Krishna Rallabandi, Parag Pravin Dakle and Preethi Raghavan . 141

No Stock is an Island: Learning Internal and Relational Attributes of Stocks with Contrastive Learning
Shicheng Li, Wei Li, Zhiyuan Zhang, Ruihan Bao, Keiko Harimoto and Keiko Harimoto 147

Prospectus Language and IPO Performance
Jared Sharpe and Keith Decker 154

It's Time to Reason: Annotating Argumentation Structures in Financial Earnings Calls: The FinArg Dataset
Alaa Alhamzeh, Romain Fonck, Erwan Versmée, Elöd Egyed-Zsigmond, Harald Kosch and Lionel Brunie 163

How Can a Teacher Make Learning From Sparse Data Softer? Application to Business Relation Extraction
Hadjer Khaldi, Farah Benamara, Camille Pradel and Nathalie Aussenac-Gilles 170

Astock: A New Dataset and Automated Stock Trading based on Stock-specific News Analyzing Model
Jinan Zou, Haiyao Cao, Lingqiao Liu, Yuhao Lin, Ehsan Abbasnejad and Javen Qinfeng Shi . 178

Next-Year Bankruptcy Prediction from Textual Data: Benchmark and Baselines
Henri Arno, Klaas Mulier, Joke Baeck and Thomas Demeester 187

AdaK-NER: An Adaptive Top-K Approach for Named Entity Recognition with Incomplete Annotations
Hongtao Ruan, Liying Zheng and Peixian Hu 196

A Sentiment and Emotion Annotated Dataset for Bitcoin Price Forecasting Based on Reddit Posts
Pavlo Seroyizhko, Zhanel Zhexenova, Muhammad Zohaib Shafiq, Fabio Merizzi, Andrea Galassi and Federico Ruggeri 203

FinSim4-ESG Shared Task: Learning Semantic Similarities for the Financial Domain. Extended edition to ESG insights
Juyeon Kang and Ismail El Maarouf 211

Using Contextual Sentence Analysis Models to Recognize ESG Concepts
Elvys Linhares Pontes, Mohamed Ben Jannet, Jose G. Moreno and Antoine Doucet 218

Automatic Term and Sentence Classification Via Augmented Term and Pre-trained language model in ESG Taxonomy texts
Ke Tian, Zepeng Zhang and Hua Chen 224

Knowledge informed sustainability detection from short financial texts
Boshko Koloski, Syrielle Montariol, Matthew Purver and Senja Pollak 228

TCS WITM 2022@FinSim4-ESG: Augmenting BERT with Linguistic and Semantic features for ESG data classification
Tushar Goel, Vipul Chauhan, Suyash Sangwan, Ishan Verma, Tirthankar Dasgupta and Lipika Dey 235

Ranking Environment, Social And Governance Related Concepts And Assessing Sustainability Aspect of Financial Texts
Sohom Ghosh and Sudip Kumar Naskar 243

Using Transformer-based Models for Taxonomy Enrichment and Sentence Classification
Parag Pravin Dakle, Shrikumar Patil, Sai Krishna Rallabandi, Chaitra Hegde and Preethi Raghavan 250

Program

Thursday, December 8, 2022

09:00 - 09:10 *Opening Remarks*

09:10 - 09:45 *Keynote - Knowledge-Based News Event Analysis and Forecasting*

09:45 - 10:00 *Overview of Recent FinNLP Studies*

09:00 - 10:30 *Main Track I*

Contextualizing Emerging Trends in Financial News Articles
Nhu Khoa Nguyen, Thierry Delahaut, Emanuela Boros, Antoine Doucet and Gaël Lejeune

Contextualizing Emerging Trends in Financial News Articles
Nhu Khoa Nguyen, Thierry Delahaut, Emanuela Boros, Antoine Doucet and Gaël Lejeune

10:30 - 11:00 *Coffee Break*

11:00 - 12:30 *Main Track II*

TweetFinSent: A Dataset of Stock Sentiments on Twitter
Yulong Pei, Amarachi Mbakwe, Akshat Gupta, Salwa Alamir, Hanxuan Lin, Xiaomo Liu and Sameena Shah

Stock Price Volatility Prediction: A Case Study with AutoML
Hilal Pataci, Yunyao Li, Yannis Katsis, Yada Zhu and Lucian Popa

Learning Better Intent Representations for Financial Open Intent Classification
Xianzhi Li, Will Aitken, Xiaodan Zhu and Stephen W. Thomas

Exploring Robustness of Prefix Tuning in Noisy Data: A Case Study in Financial Sentiment Analysis
Sudhandar Balakrishnan, Yihao Fang and Xiaodan Zhu

A Taxonomical NLP Blueprint to Support Financial Decision Making through Information-Centred Interactions
Siavash Kazemian, Cosmin Munteanu and Gerald Penn

	Toward Privacy-preserving Text Embedding Similarity with Homomorphic Encryption Donggyu Kim, Garam Lee and Sungwoo Oh
12:30 - 14:00	*Lunch Break*
14:00 - 15:30	*Main Track III and Shared Task I*
	DigiCall: A Benchmark for Measuring the Maturity of Digital Strategy through Company Earning Calls Hilal Pataci, Kexuan Sun and T. Ravichandran
	Disentangled Variational Topic Inference for Topic-Accurate Financial Report Generation Sixing Yan
	Overview of the FinNLP-2022 ERAI Task: Evaluating the Rationales of Amateur Investors Chung-Chi Chen, Hen-Hsen Huang, Hiroya Takamura and Hsin-Hsi Chen
	PromptShots at the FinNLP-2022 ERAI Task: Pairwise Comparison and Unsupervised Ranking Peratham Wiriyathammabhum
	LIPI at the FinNLP-2022 ERAI Task: Ensembling Sentence Transformers for Assessing Maximum Possible Profit and Loss from Online Financial Posts Sohom Ghosh and Sudip Kumar Naskar
	DCU-ML at the FinNLP-2022 ERAI Task: Investigating the Transferability of Sentiment Analysis Data for Evaluating Rationales of Investors Chenyang Lyu, Tianbo Ji and Liting Zhou
	UOA at the FinNLP-2022 ERAI Task: Leveraging the Class Label Description for Financial Opinion Mining Jinan Zou, Haiyao Cao, Yanxi Liu, Lingqiao Liu, Ehsan Abbasnejad and Javen Qinfeng Shi
	aiML at the FinNLP-2022 ERAI Task: Combining Classification and Regression Tasks for Financial Opinion Mining Zhaoxuan Qin, Jinan Zou, Qiaoyang Luo, Haiyao Cao and Yang Jiao
15:30 - 16:00	*Coffee Break*

Thursday, December 8, 2022 (continued)

16:00 - 16:30 *Shared Task II and EMNLP Findings*

Yet at the FinNLP-2022 ERAI Task: Modified models for evaluating the Rationales of Amateur Investors
Yan Zhuang and Fuji Ren

LDPP at the FinNLP-2022 ERAI Task: Determinantal Point Processes and Variational Auto-encoders for Identifying High-Quality Opinions from a pool of Social Media Posts
Paul Trust and Rosane Minghim

Jetsons at the FinNLP-2022 ERAI Task: BERT-Chinese for mining high MPP posts
Alolika Gon, Sihan Zha, Sai Krishna Rallabandi, Parag Pravin Dakle and Preethi Raghavan

16:30 - 16:42 *EMNLP Findings - DialogueGAT: A Graph Attention Network for Financial Risk Prediction by Modeling the Dialogues in Earnings Conference Calls*

16:42 - 16:54 *EMNLP Findings - VarMAE: Pre-training of Variational Masked Autoencoder for Domain-adaptive Language Understanding*

16:54 - 17:06 *EMNLP Findings - ASDOT: Any-Shot Data-to-Text Generation with Pretrained Language Models*

17:06 - 17:18 *EMNLP Findings - DocFiNet: Augmenting Text and Speech Transformers with Semi-structured Document Representations for Financial Tasks*

17:18 - 17:30 *EMNLP Findings - Long Text and Multi-Table Summarization: Dataset and Method*

17:30 - 17:40 *Closing*