Evaluating the Rationales of Amateur Investors

Chung-Chi Chen
Department of Computer Science and Information Engineering, National Taiwan University, Taiwan
cjchen@nlg.csie.ntu.edu.tw

Hen-Hsen Huang
Department of Computer Science, National Chengchi University, Taiwan
MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan
hhhuang@nccu.edu.tw

Hsin-Hsi Chen
Department of Computer Science and Information Engineering, National Taiwan University, Taiwan
MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan
hhchen@ntu.edu.tw

Outline

- **INTRODUCTION**
- **TASK SETTING AND DATASET**
- **EXPERIMENTS**
  - Discriminating Expert Rationales
  - Mining High-Quality Opinions
- **DISCUSSION**
  - Comparison with Analysts
  - Comparison between Different Ranking Methods
- **ADVANCED EXPLORATION**
- **FUTURE RESEARCH DIRECTIONS**
- **CONCLUSION**
- **RELATED EVENTS**
  - FinNim-3 in NTCIR
  - FinNLP in IJCAI and FinWeb in TheWebConf
  - EMNLP-2021 Tutorial: Financial Opinion Mining

Goal – Mine High-Quality Opinions by Inspecting the Supporting Rationales

NLP Lab
National Taiwan University

Previous Approaches

- **Opinion Classification Based on Post Content**
  - This type of approach classifies a post as useful or useless. It is **expensive** to annotate a sufficient number of training instances for different domains.

- **Opinion Classification Based on Reader Feedback**
  - This type of approach leverages reader information such as the number of Facebook likes. However, this kind of approach **cannot predict the quality of a post that was just publishe**d, as no reader information is available yet. It is also difficult to evaluate posts from a new account or from accounts with few followers. Furthermore, Twitter, Facebook, and Instagram plan to hide information about likes. The above issues all decrease the feasibility of this kind of approaches in future applications.

Our Approach

NLP Lab
National Taiwan University

- **High-quality posts from the crowd may share characteristics with articles written by experts.**
- We use documents written by experts and the crowd as our dataset, and train models to **discriminate expert rationales from the crowd's rationales.**
- Leveraging the high accuracy of the models, we further use the outcomes of the model to **mine high-quality opinions** from the crowd.
- If rationales written by the crowd are predicted to be expert rationales, we infer that the quality of the opinions in these documents is higher than that of the opinions in documents predicted to be the crowd's rationales
- **We address issues of cold starts and few followers**

Research Questions

(RQ1) To what extent can we use stylistic and semantic features to differentiate between rationales from professional analysts and amateur investors?

(RQ2) If we are able to classify rationales successfully, which kind of features is more useful?

(RQ3) Which approach is better, following high-quality opinions mined by the proposed approach, or following opinions ranked according to the feedback of social media users?

TASK SETTING

Quality Evaluation

- **Quality Evaluation**
  - **The more expert-like sentences in their posts, the higher quality their posts (opinions) are**

- **Expert-like Sentence**
  - We postulate that the rationales of experts are credible rationale
  - We attempt to capture expert-like rationales from the crowd
  - If a rationale from the crowd is classified as an expert's rationale, either the style or the wording of the rationale is similar to that of an expert.
  - We further infer that opinions supported by such expert-like rationales are of high quality.

Dataset

- In Chinese
  - Professional analysts' report from Bloomberg
    - Manually Collected from PDF files
  - Financial social media platform, PTT in Taiwan
    - Rule-based
    - Posts that do not follow the template will be deleted by the administrator of the platform.


1. 標的：2330 台積電

2. 分類：空/盤整

3. 分析/正文：
根據星期六YT上惡補18堂課的技術分析

台積電前低是1/29的591
3/5開584，收601

3/5開盤已經低於前低591，
不考慮周一開盤護盤情況就是下跌

沒有下跌就是進入盤整期

4. 進退場機制：
進場:過623之後，看隔天盤勢再評估
退場:591

Statistics

- Sentence-Level
- Stylistic Features
  - POS Tags
  - Dependency
- Semantic Features
  - Word-Level
  - Char-Level

	Analyst	Crowd
Unique characters	2,737	3,298
Unique tokens	15,696	27,474
Unique POS tags	49	53
Unique tag-tag-arcs	415	622
Unique incoming arcs	25	25
Training set (sentences)	32,000	32,000
Test set (sentences)	812	812

EXPERIMENTS

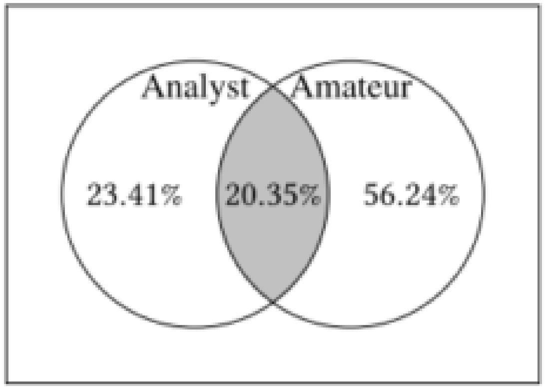
Discriminating Analysts’ and Amateurs’ Rationales

NLP Lab
National Taiwan University

(RQ1) To what extent can we use stylistic and semantic features to differentiate between rationales from professional analysts and amateur investors?

- Models discriminate expert rationales via both stylistic and semantic features with high F1-scores.

Features	Model	Macro-F1
Stylistic		
Dep. - arc	CNN	62.07
	BiGRU	61.54
Dep. - TTA	CNN	61.04
	BiGRU	62.91
POS	CNN	70.16
	BiGRU	73.34
Semantic		
Word-level	CNN	85.24
	BiGRU	85.74
Character-level	BERT-CNN	87.87
	BERT-BiGRU	88.59
Fusion Models		
BERT-BiGRU + BiGRU (POS) + BiGRU (TTA)		90.32
BERT-BiGRU + BiGRU (POS) + CNN (arc)		**90.81***



Venn diagram of wordings.

Details of Mining High-Quality Opinions

NLP Lab
National Taiwan University

- Posts: 2019/05/13 to 2019/06/18
- Price: 2019/05/13 to 2019/09/11
  - The global market was influenced by the China-United States trade war
- No overlaps between new dataset and the dataset used to train the discriminating models
- Maximum possible profit (MPP)
  - Potential profit
  - Potential of the selected opinions
- Maximum loss (ML)
  - Downside risk

$$MPP_{bullish} = (\max(H_{(t+1,T)}) - O_{t+1})/O_{t+1} \qquad MPP_{bearish} = (O_{t+1} - \min(L_{(t+1,T)}))/O_{t+1}$$

$$ML_{bullish} = (\min(L_{(t+1,T)}) - O_{t+1})/O_{t+1} \qquad ML_{bearish} = (O_{t+1} - \max(H_{(t+1,T)}))/O_{t+1}.$$

Results of Mining High-Quality Opinions

NLP Lab
National Taiwan University

Method	Ranking	Average MPP	Average ML
Random		11.94%	-17.28%
Feedback	First decile	8.88%	-8.69%
	Second decile	7.14%	-10.73%
	Top 2 deciles	8.53%	-9.10%
Best FM	First decile	**17.61%**	-3.72%
	Second decile	8.80%	-8.67%
	Top 2 deciles	13.09%	-6.26%
BiGRU(POS)	First decile	15.78%	**-2.46%**
	Second decile	10.52%	-8.72%
	Top 2 deciles	12.71%	-6.11%

DISCUSSION

Comparison with Analysts

NLP Lab
National Taiwan University

Readability comparison.

	Analyst	Crowd
Average hard words	**31.61**	24.60
Sentences with complex semantics	**6.86**	2.66
Noun phrase modifier ratio	**0.27**	0.16
Content word density	**0.87**	0.86
Positive transition words	**3.32**	1.98
Negative transition words	**0.99**	0.93
Number of personal pronouns	0.22	**0.98**
Number of negative words	0.11	**1.24**

Selected words in expert-like lexicon.

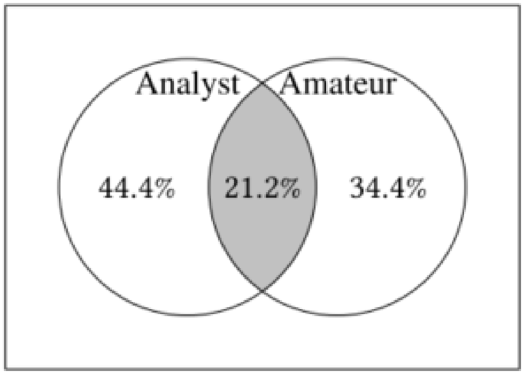
Word	ELScore	Word	ELScore
Price target	2.10	Short	-1.72
Estimate	2.06	Guess	-1.75
We	2.04	I	-1.75
Gross profit ratio	1.91	Pattern	-1.71

Comparison of MPP and ML.

	Average MPP	Average ML
Analyst	**22.30%**	-6.52%
Stylistic + Semantic	17.61%	-3.72%
Stylistic	15.78%	**-2.46%**

Statistics of mentioned stocks.

	Crowd	Analyst
Stock Exchange Market	67.72%	66.89%
Over-the-Counter Market	32.28%	33.11%
Average Market Capitalization (Million)	12,556	12,846



Venn diagram of mentioned stocks.

Comparison with Analysts

NLP Lab
National Taiwan University

Readability comparison.

	Analyst	Crowd
Average hard words	**31.61**	24.60
Sentences with complex semantics	**6.86**	2.66
Noun phrase modifier ratio	**0.27**	0.16
Content word density	**0.87**	0.86
Positive transition words	**3.32**	1.98
Negative transition words	**0.99**	0.93
Number of personal pronouns	0.22	**0.98**
Number of negative words	0.11	**1.24**

Selected words in expert-like lexicon.

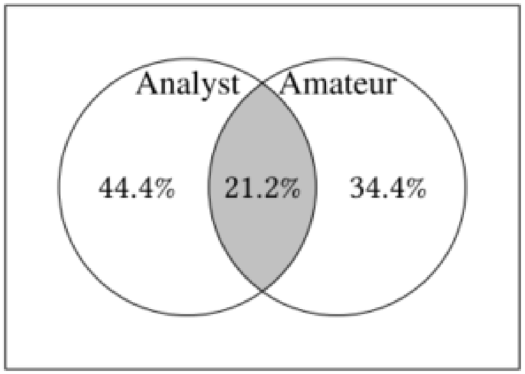
Word	ELScore	Word	ELScore
Price target	2.10	Short	-1.72
Estimate	2.06	Guess	-1.75
We	2.04	I	-1.75
Gross profit ratio	1.91	Pattern	-1.71

Comparison of MPP and ML.

	Average MPP	Average ML
Analyst	**22.30%**	-6.52%
Stylistic + Semantic	17.61%	-3.72%
Stylistic	15.78%	**-2.46%**

Statistics of mentioned stocks.

	Crowd	Analyst
Stock Exchange Market	67.72%	66.89%
Over-the-Counter Market	32.28%	33.11%
Average Market Capitalization (Million)	12,556	12,846

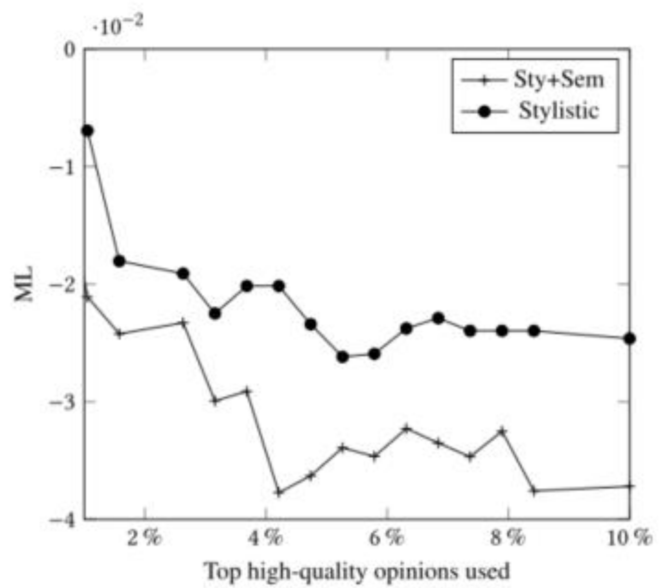


Venn diagram of mentioned stocks.

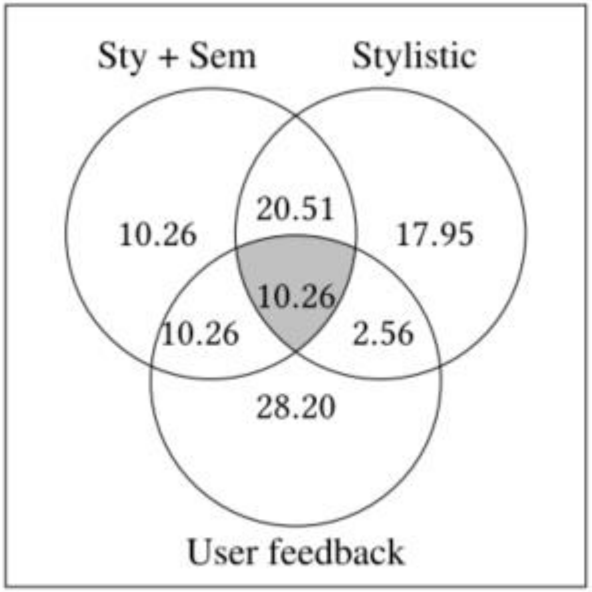
Comparison between Different Ranking Methods

(RQ2) If we are able to classify rationales successfully, which kind of features is more useful?

- Due to the wording habits, we cannot glean bearish opinions from amateur investors when we adopt semantic features.
- With stylistic features only, we can remove this restriction.



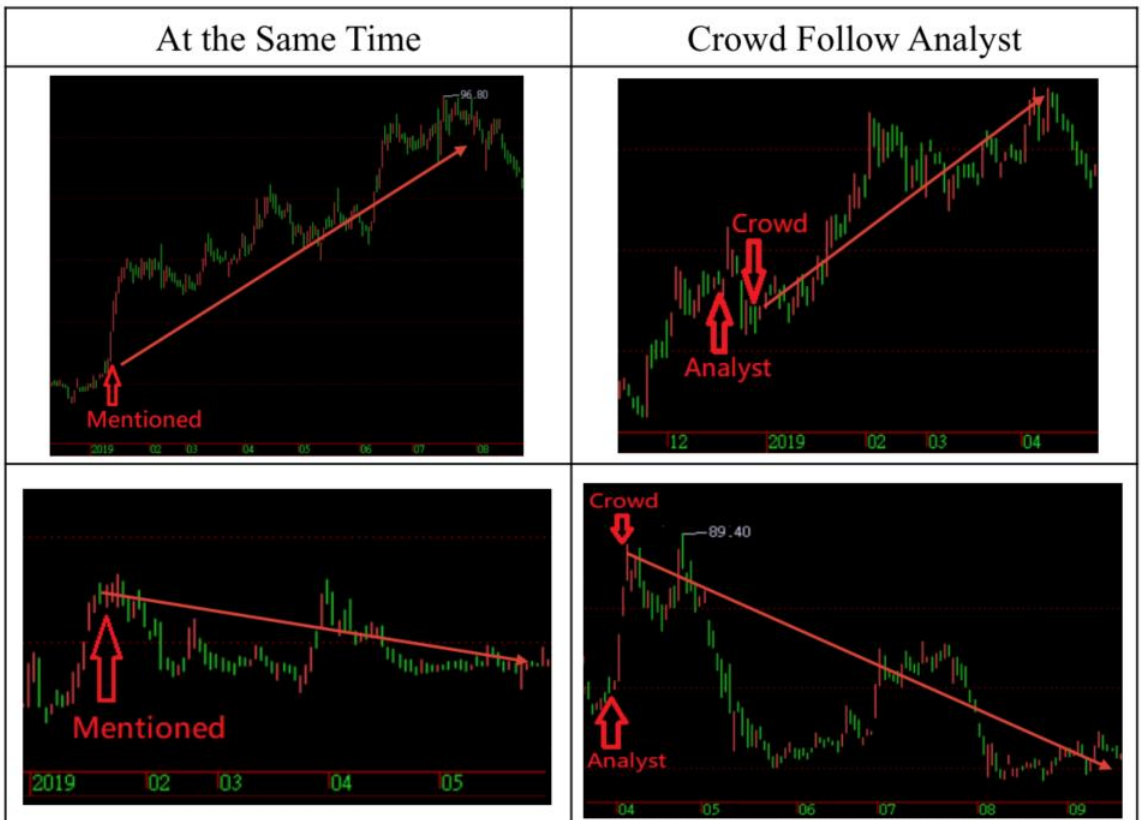
Average ML under different numbers of high-quality opinions. "Sty + Sem" denotes "Stylistic+Semantic".



Venn diagram of documents in the first decile under different ranking methods. All numbers are in percentage (%)

Aggregation of Analysts and Top-Ranked Posts

	Average MPP	Average ML
Analyst + Sty	**20.55%**	**-5.43%**
Analyst + Sty + Sem	19.43%	-6.67%
Analyst + User Feedback	19.26%	-7.01%



Investor's Claim-Rationale Dataset (ICRD)

NLP Lab
National Taiwan University

- Professional analysts' report from Bloomberg
  - Manually Collected from PDF files
- Financial social media platform, PTT in Taiwan
  - Rule-based

1. 標的：2330 台積電

2. 分類：空/盤整

3. 分析/正文：
根據星期六YT上惡補18堂課的技術分析

台積電前低是1/29的591
3/5開584，收601

3/5開盤已經低於前低591，
不考慮周一開盤護盤情況就是下跌

沒有下跌就是進入盤整期

4. 進退場機制：
進場:過623之後，看隔天盤勢再評估
退場:591

3. Analysis

Rationales

4. Enter/Exit Strategies

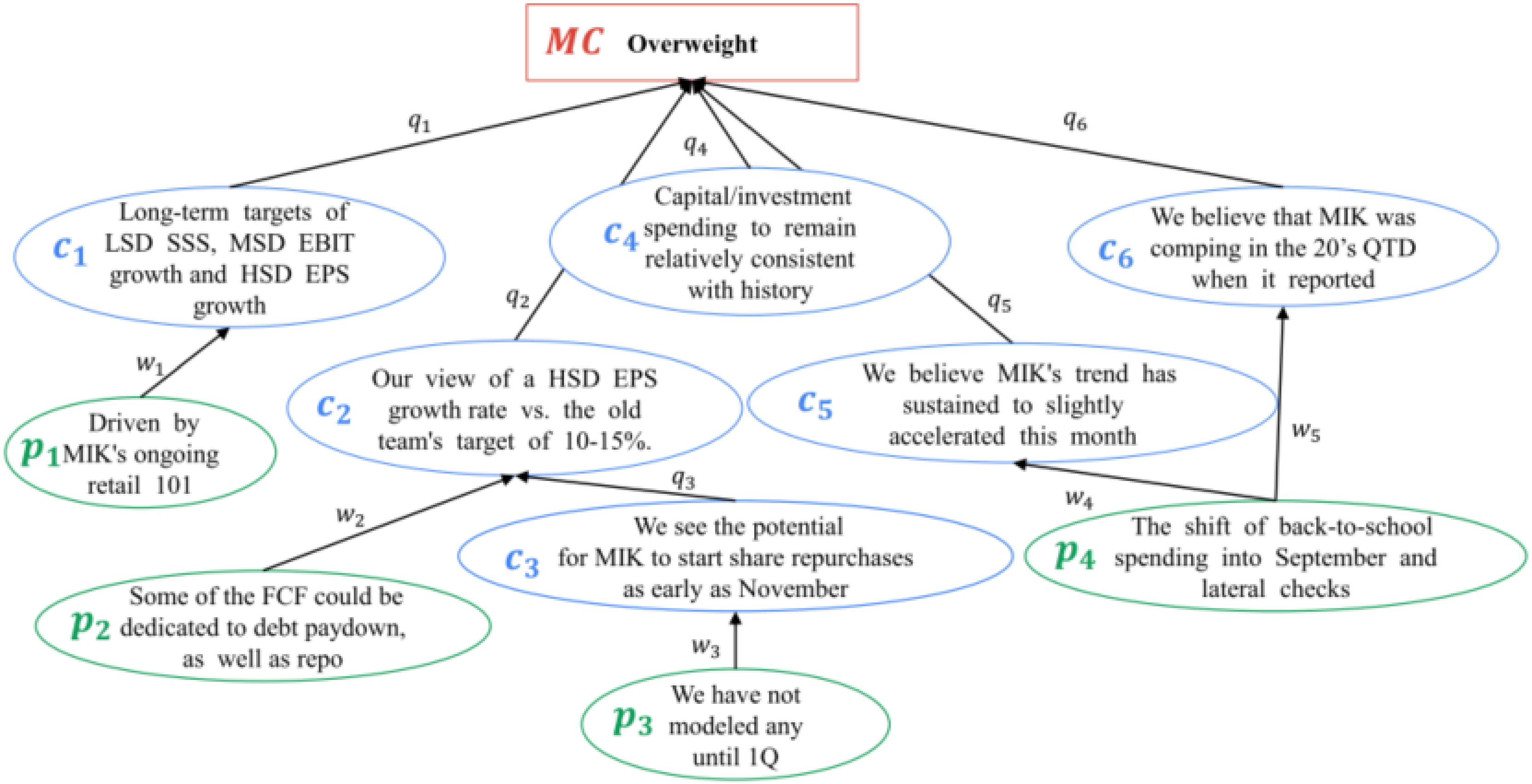
Free-Formed Arguments Extraction

- Claim Detection
  - Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2020. NumClaim: Investor's Fine-grained Claim Detection. In CIKM'20
- **Rationale Detection**
- **Claim-Rationale Inference**

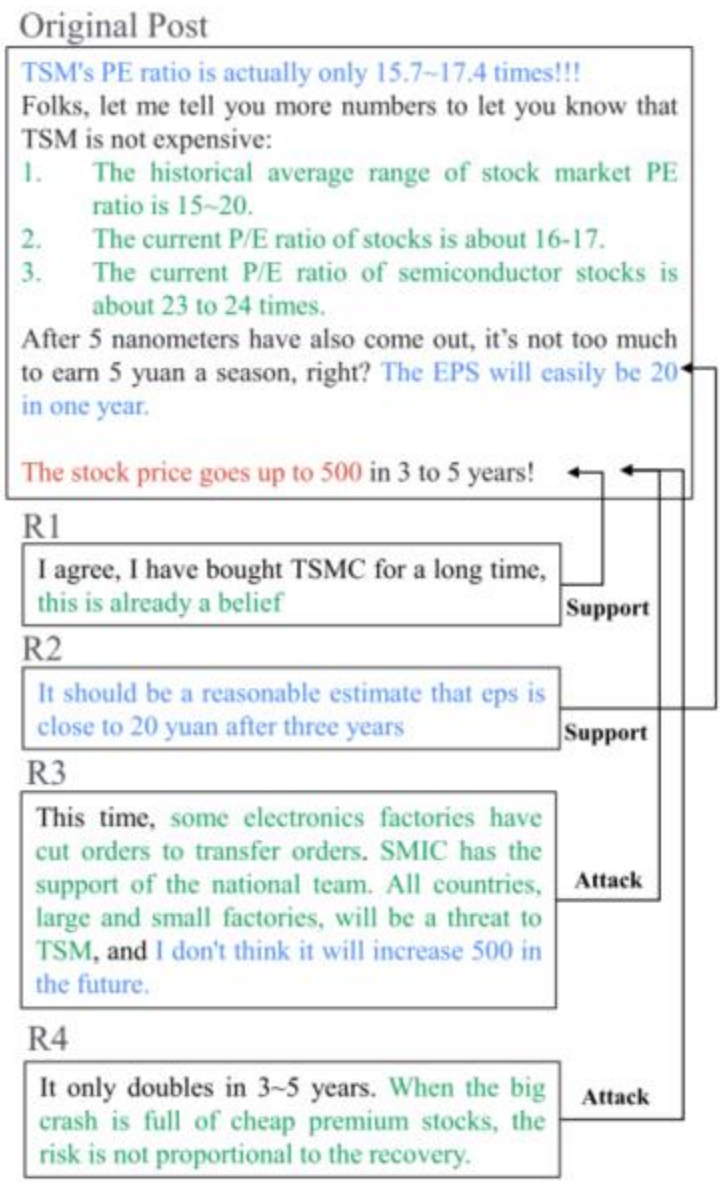
Model	Claim Detection	Rationale Detection	Inference
CNN	76.15	55.25	53.75
BiGRU	77.97	48.62	54.74
CapsNet	77.93	52.47	51.97
BERT	**79.86**	**57.69**	**56.96**

FUTURE RESEARCH DIRECTIONS – Enhancing Opinion Quality Evaluation via Argument Mining Notions

Argumentation Structure in an Opinion



Argumentation Structure of Opinions



CONCLUSION

Contributions

- We present an important task—mining high-quality opinions—for research focusing on extracting or using opinions from user-generated textual data.
- We propose a novel approach to infer opinion quality by how “expert-like” the rationale supporting the opinion is.
- We show that top-ranked crowd opinions mined by our approach are comparable with the opinions of professional analysts in terms of controlling downside risk.
- The future research directions are presented and explored with the proposed pilot dataset, ICRD.

Related Events

- FinNum-3: Investor's and Manager's Fine-grained Argument Detection
  - http://finnum.nlpfin.com/
- The Third Workshop of Financial Technology and Natural Language Processing (FinNLP-2021)
  - http://finnlp.nlpfin.com/
- The 1st Workshop on Financial Technology on the Web (FinWeb)
  - http://finweb.nlpfin.com/

- EMNLP-2021 Tutorial: Financial Opinion Mining
- Springer SpringerBriefs: Financial Opinion Mining (Available in 2021)

https://forms.gle/RB9Qq9ok6z5exu1G6

Feel free to contact us if you have any questions.

Chung-Chi Chen: cjchen@nlg.csie.ntu.edu.tw

NLP Lab
National Taiwan University