NumClaim: Investor's Fine-grained Claim Detection

Chung-Chi Chen[1], Hen-Hsen Huang[2,3], Hsin-Hsi Chen[1,3]

[1]Department of Computer Science and Information Engineering, National Taiwan University, Taiwan

[2]Department of Computer Science, National Chengchi University, Taiwan

[3]MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan

Student Travel Grants 2020

ACM SIGIR Student Travel Grants

Overview

NLP Lab
National Taiwan University

- Argument mining issue in finance
- Expert-annotated dataset, NumClaim
- We show that encoding with numeral encoder and co-training with the numeral understanding auxiliary task are helpful for the numeral-oriented task.

Motivation

NLP Lab
National Taiwan University

- Over **58.47%** of sentences in analysis report contain at least one numeral

- Investors always make a claim **with an estimation**
  - **(X) We estimate that the sales may growth**
  - **(O) We estimate that the sales growth rate may exceed 40%**

- The importance of fine-grained claims and the numerals.
  - We estimate that the sales growth rate may exceed 5%
  - We estimate that the sales growth rate may exceed 40%

NumClaim

NLP Lab
National Taiwan University

- Chinese financial analysis reports
- The annotators work in the financial industry (bank's treasury department and hedge fund)
- The Cohen's kappa agreements between the experts are 88.31%
- 5,144 instances: 23.78% "In-claim" and 76.22% "Out-of-claim"

Sentence	Label
We estimate that the sales growth rate may exceed **40**%.	In-claim
Professional audio/visual products account for **20**%.	Out-of-claim

In-claim		Out-of-claim	
estimate	2.86	lower/higher than	-1.37
price target	2.80	cause	-1.37
downgrade	2.58	last year	-1.26
upgrade	1.55	influence	-1.25

Auxiliary Task – Numeral Understanding

NLP Lab
National Taiwan University

- The Cohen's kappa agreements between the experts are 89.55%

Category	Subcategory	In-claim	Out-of-claim	Sum
Monetary	price	42	33	75
	money	506	368	874
	change	3	15	18
Percentage	absolute	208	500	708
	relative	408	402	810
Temporal	date	0	2,134	2,134
	time	0	3	3
Quantity	absolute	55	219	274
	relative	0	4	4
Product Number		1	135	136
Ranking		0	3	3
Other		0	105	105
Total		1,223	3,921	5,144

Chung-Chi Chen, Hen-Hsen Huang, Yow-Ting Shiue, and Hsin-Hsi Chen. 2018. Numeral understanding in financial tweets for fine-grained crowd-based forecasting. In *IEEE/WIC/ACM International Conference on Web Intelligence*

Statistics

NLP Lab
National Taiwan University

Dataset	NumClaim	CRC [13]	PE [12]
Language	Chinese	Chinese	English
Source	Analysis Report	Hotel Review	Persuasive Essay
# Word	42,594	21,848	97,420
# Numeral	5,144	67	111

	NumClaim	CRC
# hard words	**31.95**	18.28
# negative words	0.14	**0.60**
# synonym	0.28	**1.49**
Noun phrase modifier ratio	0.29	**0.38**
Noun phrase ratio	**31.79**	26.62
# transition words	**4.86**	1.62

[12] Steffen Eger, Johannes Daxenberger, and Iryna Gurevych. 2017. Neural End-to-End Learning for Computational Argumentation Mining. In ACL
[13] Steffen Eger, Johannes Daxenberger, Christian Stab, and Iryna Gurevych. 2018.Cross-lingual Argumentation Mining: Machine Translation (and a bit of Projection) is All You Need!. In COLING.

Experimental Results

NLP Lab
National Taiwan University

- Encoding: BERT
- Baseline: CNN, BiGRU, CapsNet
- Metrics: Macro-F1
- Class Weight (CW)
- Numeral Encoder
  - Represent the digit (0--9) and the decimal point as a 11-dimension tensor, and concatenate it with a tensor for the inter-numeral position information.
- Joint Learning with Category Classification Task (CG)

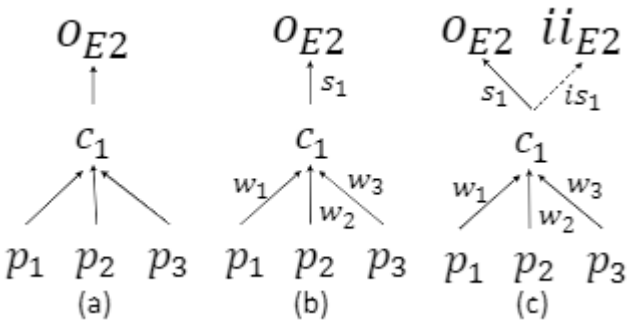
Architecture	CNN	BiGRU	CapsNet
Baseline	76.15%	77.97%	77.93%
+ CW	77.26%	78.29%	78.68%
+ CW & NE (CNN)	78.19%	79.06%	80.91%
+ CW & NE (CNN) & CG	81.35%	81.65%	**82.62%**

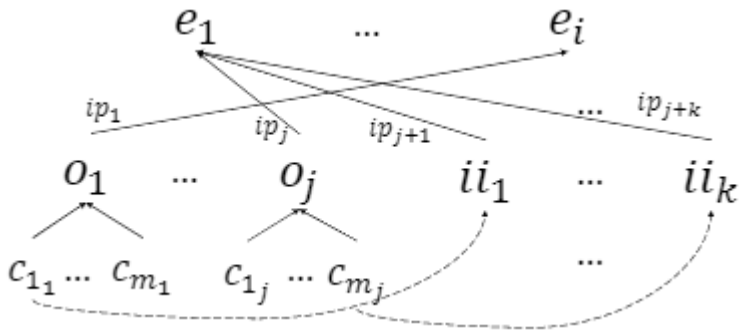
Conclusion & Future Direction

NLP Lab
National Taiwan University

- Our contributions
  - Explore the argument mining issue in finance
  - Provide an expert-annotated dataset – NumClaim
  - Propose helpful methods for solving numeral-oriented task

- Future Directions – Fine-grained Financial Opinion Mining
  - Premise detection and relation linking
  - Rationality assessment





Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2020. Fine-grained Financial Opinion Mining: A Survey and Research Agenda. In *arXiv:2005.01897*

Related Datasets and Events

NLP Lab
National Taiwan University

- FinNum-1: Fine-Grained Numeral Understanding in Financial Tweets (NTCIR-14, 2018-2019)
- FinNum-2: Numeral Attachment in Financial Tweets (NTCIR-15, 2019-2020)
- **FinNum-3: Investor's Fine-grained Argument Detection (Will submit proposal to NTCIR-16)**

- Tutorial in AACL-IJCNLP 2020: **Natural Language Processing in Financial Technology Applications**

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

Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2020. *NLP in FinTech Applications: Past, Present and Future*. In arXiv:2005.01320.

Feel free to contact us if you have any questions.

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

Student Travel Grants 2020

ACM SIGIR Student Travel Grants

NLP Lab

National Taiwan University