Numeracy-600K: Learning Numeracy for Detecting Exaggerated Information in Market Comments

Chung-Chi Chen[1], Hen-Hsen Huang[2,4], Hiroya Takamura[3], Hsin-Hsi Chen[1,4]

[1]Department of Computer Science and Information Engineering, National Taiwan University, Taiwan
[2]Department of Computer Science, National Chengchi University, Taiwan
[3]Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Japan
[4]MOST Joint Research Center for AI Technology and All Vista Healthcare, Taiwan

Numeracy

S&P 500 <.SPX> UP 1.53 POINTS AT _____ AFTER MARKET OPEN

DOW JONES <.DJI> UP 8.70 POINTS AT _____ AFTER MARKET OPEN

U.S. Q3 GDP rises _____.

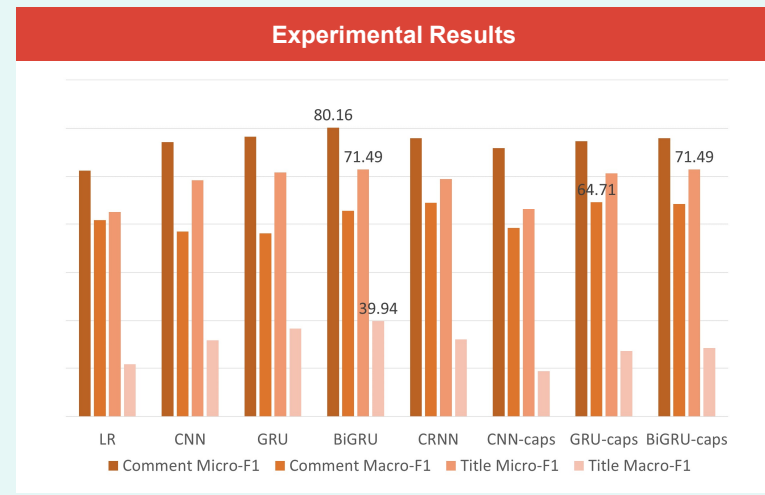## Issues of Understanding Numerals in Financial Documents

Market comment	Issue
CANADA <*TRT*> PCT 2014 BOND AUCTION YIELD LOW 1.110 PCT,HIGH 1.121 PCT	Different contract
DOW JONES <.DJI> UNOFFICIALLY CLOSES UP <*TRT*> POINTS	Market Data
Disney quarterly revenue rises <*TRT*> pct	Past patterns
BILL BARRETT CORP <BBG.N>: BMO CUTS PRICE TARGET TO 8 FROM <*TRT*>	Reference to other numerals
Maersk Drilling wins $ <*TRT*> mln contract from Eni	Main event
OCC SAYS EXCHANGE-LISTED OPTIONS VOLUME REACHED <*TRT*> CONTRACTS IN MAY	Varying amounts

Numeral Understanding
Numerals in financial tweets can be extended into 17 fine-grained meanings.

Numeral Attachment
Detecting the attached target (i.e., cashtag) of the numeral

Intuition of the Magnitude

Magnitude	Range	Comment	Title
Decimal	$0 \le m < 1$	23.24%	0.08%
1	$1 \le m < 10$	37.53%	35.18%
2	$10 \le m < 10^2$	25.36%	30.94%
3	$10^2 \le m < 10^3$	12.21%	8.71%
4	$10^3 \le m < 10^4$	1.12%	24.21%
5	$10^4 \le m < 10^5$	0.29%	0.57%
6	$10^5 \le m < 10^6$	0.23%	0.31%
> 6	$10^6 \le m$	0.01%	0.01%

Experimental Results

Bar chart legend: Comment Micro-F1, Comment Macro-F1, Title Micro-F1, Title Macro-F1. Categories: LR, CNN, GRU, BiGRU, CRNN, CNN-caps, GRU-caps, BiGRU-caps. Labeled values: 80.16, 71.49, 39.94, 64.71, 71.49.

Error Analysis & Cross-source Experiment

BiGRU and GRU-capsule confusion matrices.

Training	Test set	Micro-F1	Macro-F1
Comment	Title	31.38%	11.08%
Title	Comment	25.59%	10.58%

Detecting Exaggerated Information

China Q1 GDP growth 6.4 pct y/y

+10% ➔ China Q1 GDP growth 7.04 pct y/y
+30% ➔ China Q1 GDP growth 8.32 pct y/y
+50% ➔ China Q1 GDP growth 9.60 pct y/y
+70% ➔ China Q1 GDP growth 10.88 pct y/y
+90% ➔ China Q1 GDP growth 12.16 pct y/y
-90% ➔ China Q1 GDP growth 1.22 pct y/y

Bar chart, Micro-F1 and Macro-F1 for ±10%, ±30%, ±50%, ±70%, ±90%:
±10%: 58.54, 57.87
±30%: 56.94, 56.11
±50%: 57.69, 56.85
±70%: 70.92, 70.85
±90%: 79.91, 76.94

FinNum-2 in NTCIR-2020

Semantic Understanding

Numeral Attachment:
$NE OK NE, last time oil was over $65 you were close to $8. Giddy-up…

NTCIR

Conclusion and Future Works

Research Directions
Outside Resource:
Time, Market Data, News…
Fine-grained Numeracy:
Select Proper Number
Domain-specific Numeracy:
Clinical, Geography, …

Numeracy-600K
600K Market Comments from Reuters
600K Article Titles from the Web

Association for Computational Linguistics

Wi 2018 International Conference

SIGIR 2019

國立臺灣大學 National Taiwan University · NATIONAL CHENGCHI UNIVERSITY · AIntu 科技部臺灣大學人工智慧研究中心 MOST JOINT RESEARCH CENTER FOR AI TECHNOLOGY AND ALL VISTA HEALTHCARE · AIRC · NLP Lab National Taiwan University