

Learning to Generate Correct Numeric Values in News Headlines



TAIPEI
2020

Jui Chu¹, Chung-Chi Chen¹, Hen-Hsen Huang^{2,3}, Hsin-Hsi Chen^{1,3}

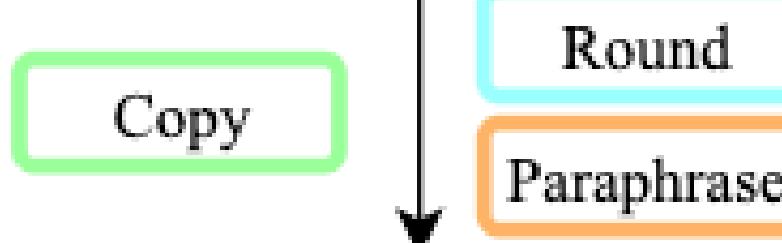
¹Department of Computer Science and Information Engineering, National Taiwan University, Taiwan

²Department of Computer Science, National Chengchi University, Taiwan

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

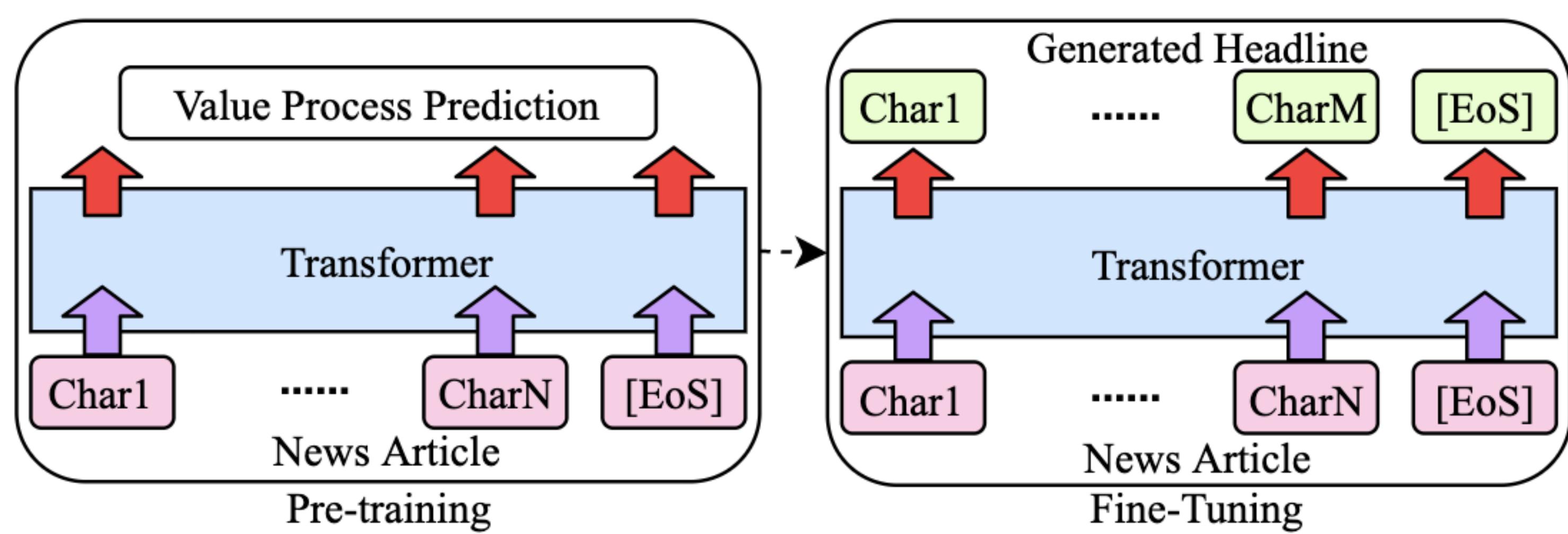
The Process of Numeric Values from News Articles to Headlines

(Article1) FED raises the rates, showing the confidence for economic recovery. This also stimulates the European stock market, especially the export stocks, in which the Germany ones boosts by **2.6%**.
(Article2) The numbers of manufacturing are in bad situation, dragging the US stocks to pullback for 24 days. The NASDAQ composite index of technology stocks drops down by **1.67%**.
(Article3) Ride-hailing giant Uber released its first safety report on Thursday, disclosing that it received a total of **5,981** reports of sexual assault incidents during its ride-hailing trips in the U.S. in 2017 and 2018.



(Headline1) European Stock Takes a Rides of US Rates Raise. The Germany Stock Rises by **2.6%**.
(Headline2) US Manufacturing Stocks Struggle. NASDAQ Composite Collapses by **1.7%**.
(Headline3) Uber Discloses Nearly **6K** Reports of Sexual Assaults

Pre-train by Value Process Prediction



Methodology

Base Architecture: Character-based Transformer

Two Different Forms:

(1) Token: whole value as a token

(2) Digit: represent the value digit by digit.

Two Different Setting:

(1) Numeral Category: year, month, date, percentage, price, and population

(2) Pre-train by Value Process Prediction:

While the generator is passively allowed to transform values in the article with digit forms of values given, we expect the model can learn to actively process the values for the headline.

Experimental Results

Method	From Article	New to Article	Weighted Average
Base Model	45.23	21.28	23.81
w/ Token	50.76	18.06	35.71
w/ Token w/o category	49.71	18.74	35.46
w/ Digit	41.37	21.45	32.21
w/ Digit w/o category	39.07	21.63	31.05
w/ Digit + pre-train	45.71	24.25	35.83

Generated Examples

News Title	US Manufacturing Stocks Struggle. NASDAQ Composite Collapses by 1.7%
Base Model	Tech Stocks collapse. The NASDAQ composite drops over 1% .
w/ Number Token	Tech Stocks Lead US Stocks to collapse.
w/ Digit Token	Tech Stocks Lead the US Stocks to drop. The NASDAQ index collapse by 1.6% .
w/ Pre-train	Poor US Manufacturing Stocks. The NASDAQ index collapses by 1.7% .

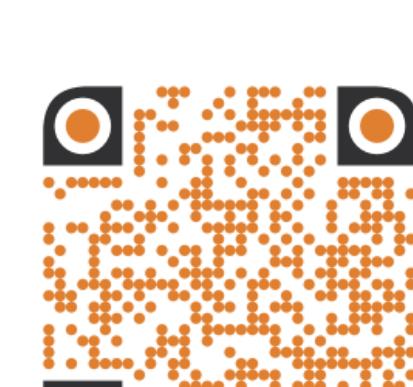
Conclusion & Future Work

- We experiment with various methods to guide the generator to process values from articles for headlines.
- The proposed pre-train task leads the model to correctly generate numbers even in the extreme case of target values in headlines are newly generated by rounding and paraphrasing.
- The results show the challenge of these processes, implying the importance of numerical reasoning.
- Exploring numerical reasoning methods for generative models is our future work.

FinNum-2 in NTCIR-2020

Semantic Understanding

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

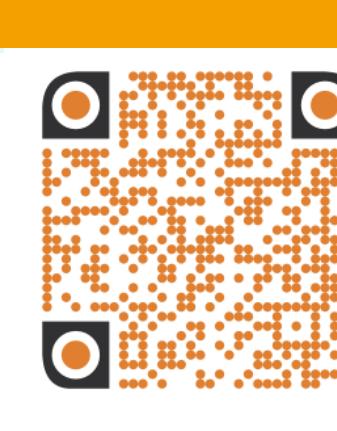


<http://finnum.nlpfin.com>

Related Datasets

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

<http://numeracy600k.nlpfin.com>

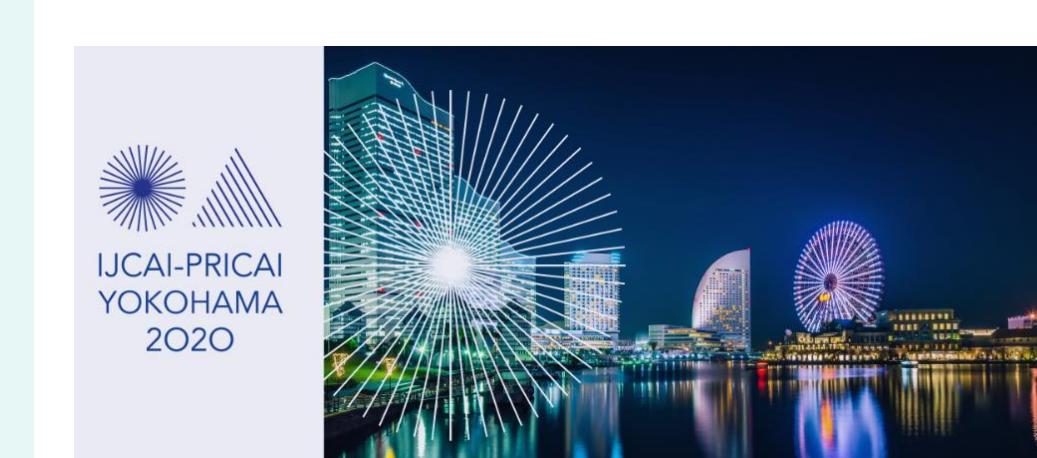


Numeral Understanding:
8,868 Annotations on Financial Tweets
Fine-grained Taxonomy for Numerals

<http://numeralunderstanding.nlpfin.com>

FinNLP Workshop in IJCAI-2020

- Submission Deadline: **April 24, 2020**
- The **Best Paper Award** winner: USD\$500
- Published at **ACL Anthology**



<http://finnlp.nlpfin.com>



國立臺灣大學
National Taiwan University



NATIONAL CHENGCHI UNIVERSITY

科技部人工智慧技術
暨全幅健康照護聯合研究中心
Most Joint Research Center for AI Technology and All Vista Healthcare

NLP
Lab
National Taiwan University