Learning to Generate Correct Numeric Values in News Headlines

Jui Chu, Chung-Chi Chen
Department of Computer Science and
Information Engineering, National
Taiwan University, Taiwan
{jchu,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

ABSTRACT

Motivated by the significant role of numeric values to convey concise and accurate information in news headlines, we focus the headline generation task on displaying correct numbers. We propose various ways to present the numeric values to the generative model. In the end, we come up with a simple but effective pre-train task to guide the generator to correctly process the values, which outperforms other base models even if the numbers in the headline are newly generated from the article.

CCS CONCEPTS

• **Computing methodologies → Natural language generation**.

KEYWORDS

Summarization, headline generation, numeracy

1 INTRODUCTION

As a subtask of summarization, headline generation has made progress thanks to neural networks. However, more efforts are needed to make the generated headlines reliable for both editors and readers. One of the important challenges could be the concise and accurate information by numeric values to convey the critical points of news. As shown in Figure 1 of three news article-headline pairs, the common processes for values in articles to go through are copy (2.6% in the first pair), round (1.7% from 1.67% in the second pair), and paraphrase (6K from 5981 in the third pair), which might require numerical reasoning [3]. Actually, these values in the headlines bring out the points of the news. Taking the third pair as an example, the count of reports - '6K' tells how serious the issue is. Imagine if the value '6K' in the mentioned headline is misinterpreted as '6' instead of rounded and paraphrased from
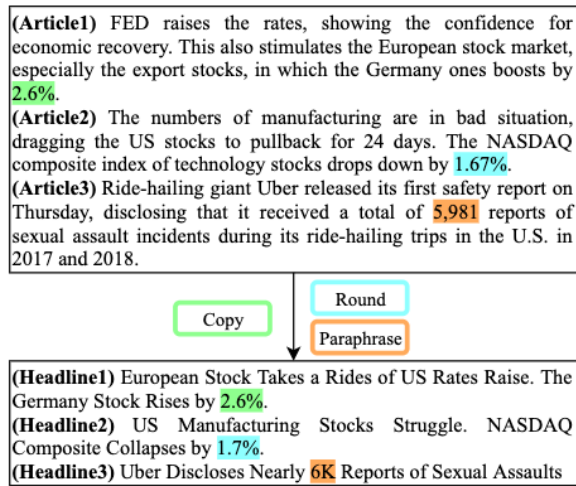
Figure 1: The process of numeric values from news articles to headlines

'5,981' in the article, the main concern of the news could be largely downplayed. However, most training settings treat word tokens and numeric values equally, making the operations on numeric values being limited as they weigh equally to the word tokens in terms of loss function and decoding steps.

The numerals in financial news headlines are one of the focuses of recent researches [1, 2]. Motivated by the mentioned significant role of numbers, we aim to design methods for the model to reliably convey the news points by displaying accurate numbers. In response, we experiment with various forms of numeric values and propose a pre-train task to guide the generator to first learn the representation of numbers before generating the headline. This simple but efficient task helps the generator outperform other base methods even in the extreme case of all the target numbers in headlines are newly generated from the articles.

2 METHODOLOGY

Baseline Models: For this Chinese headline generation task, we choose character-based Transformer as our base architecture due to its self-attention mechanism to capture the long-range dependency. The availability of processing values by either copying the whole value or rounding the value bit by bit depends on the form of numbers presented to the generator. Thus, we experiment with two
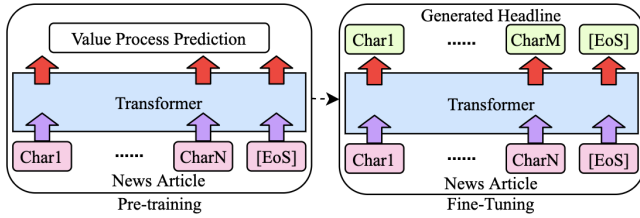
Figure 2: Pre-train by Value Process Prediction

different forms: (1) Token: whole value as a token and (2) Digit: represent the value digit by digit. We also categorize values into common types in news: *year, month, date, percentage, price, and population* to examine if they provide useful information.

Pre-train by Value Process Prediction: The main challenge for correctly displaying values in the headline is to deal with those have to be newly generated given the context of the article. 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. Thus, we design a pre-train task as shown in Figure 2 to learn how to treat the numeric values by predicting the exact process that each value in the headline has been through from the article. Later, the model can generate the headline based on the learned parameters, which have the process information included.

3 EXPERIMENT AND DISCUSSION

We collected 65K financial news article and headline pairs (all headlines contain at least one numeral) from a well-known newsvendor, MoneyDJ, in Taiwan. There are nearly 30% of numbers in the headlines are new to the articles, i.e., processed via round or paraphrase.

The ROUGE evaluations for the models differ negligibly, ranging from 0.42 to 0.44 of ROUGE-1 and 0.24 to 0.26 of ROUGE-2. That shows the language models for the words are almost the same under different settings to the numerals. Thus, we focus the evaluation on the correctness of values displayed in Table 1. Considering the target numbers that should be included in headlines, the article-headline pairs can be categorized into (1) From Article: all the values in headlines can be copied from articles, and (2) New to Article: there are values in headlines that are not shown in articles. In addition, we have the weighted average according to the proportion of these pairs. The proportions of the two in the data set (65,226 pairs) are 54% and 46%, respectively. We use 80%, 10%, and 10% of data as training, validation, and test sets.

Table 1: F1 scores of numeric correctness (%)

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**

First of all, the scores between the two kinds of pairs reflect the functions of different granularity of units. Using the whole values as token unit help models directly copy values for the From Article Pairs. On the other hand, models with digit-based unit have better scores for the New to Article pairs since they can decode numbers digit by digit through either rounding the values up or paraphrasing the form of digits to other representations.

We also find that representing the tokens or digits with the categories of values only boost the scores for the From Article pairs. It suggests the effect of value types to inform the model correct ways to process them because some specific types of values, e.g., year and month, are usually copied to the headline.

Lastly, extended from the digit-based model, the pre-train task leads to obvious improvement of scores in both kinds of pairs. Especially for the New to Article pairs, it boosts the model to outperform all the other methods. The results show that fine-tuning the generator based on the learned parameters by identifying the process of values effectively helps it include the correct numbers. Examples (E1) and (E2) for the second pair in Figure 1 show that the generator fine-tuned on the pre-trained model successfully rounds the percentage up, while the base model only omits the last few bits.

(E1) **Base Model**: Tech Stocks collapse. The NASDAQ composite drops over **1%**.

(E2) **w/ Digit + pre-train**: Poor US Manufacturing Stocks. The NASDAQ index collapses by **1.7%**.

4 CONCLUSION

To make the headline generative model better fit in real-world scenarios, i.e., conveying news by concise numeric information, we extend the task to focus on having correct numbers in headlines. We experiment with various methods to guide the generator to process values from articles for headlines. Besides different granularity of units, 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. Meanwhile, the results show the challenge of these processes, implying the importance of numerical reasoning. Thus, exploring numerical reasoning methods for generative models is our future work.

ACKNOWLEDGMENTS

REFERENCES

[1] Chung-Chi Chen, Hen-Hsen Huang, Hiroya Takamura, and Hsin-Hsi Chen. 2019. Numeracy-600K: Learning Numeracy for Detecting Exaggerated Information in Market Comments. In *ACL*.
[2] Soichiro Murakami, Akihiko Watanabe, Akira Miyazawa, Keiichi Goshima, Toshihiko Yanase, Hiroya Takamura, and Yusuke Miyao. 2017. Learning to Generate Market Comments from Stock Prices. In *ACL*.
[3] Eric Wallace, Yizhong Wang, Sujian Li, Sameer Singh, and Matt Gardner. 2019. Do NLP Models Know Numbers? Probing Numeracy in Embeddings. In *Proceedings of the 2019 Conference on EMNLP-IJCNLP*.