Preface: Is Impact of Scientific Humour Predictable?

Kenichi Iwatsuki
International Kimwipe Table Tennis Association

ABSTRACT

Scientific humour sometimes has an impact on some part of the world, but it was found to be difficult to estimate it.

Introduction

Since 2010, KTTA has made the most of the power of scientific humour to attract attentions of potential Kimwipers. It posts tweets containing jokes and irony that should be understood in scientific backgrounds. They are spread by being retweeted to be found by those who may be interested in our activities.

It is quite difficult to create a lot of tweets that contain the domain-specific humour because KTTA has failed to hire a sufficient number of people who have expertise in scientific disciplines. Thus, we must elaborate each one of them, but in many cases they are not much accepted by the people. If it is possible to estimate an impact of the tweets during polishing them up, the overall performance will improve.

In this preface, we consider the question of how we can predict the number of retweeting humorous tweets. There could be several variables contributing to the counts, but we focus on the tweets as such, which we assume to be the most important factor. The input is the tweets that were composed by KTTA. The output is categorical: not retweeted, retweeted by 1–9 people, and retweeted by more than 9 people. We also considered regression, where the exact counts of retweets were directly predicted, but it failed.

We applied machine-learning to this problem in the supervised manner. We used BERT (Devlin et al., 2018) and the implementation (Wolf et al., 2020) that was quite easy to use for us even though we were unfamiliar with computer science.

Our experimental results show that the classification is difficult. The imbalance between highly-retweeted and unattractive tweets increased the difficulty, which implies that we have to produce more funny tweets to create better training data.

Methods

The data we used were tweets made by @kttajp. All the tweets were written in Japanese. The statistics of the data are listed in Table 1. Figure 1 shows the number of retweets and tweets. Most tweets have not retweeted; only 1.8% of the tweets have attracted 10 or more retweets.

We divide the data into the training and evaluation data. The training data were used for parameter-tuning, and the evaluation data were used for the final evaluation. To mitigate the effect of the imbalance during training, the data were recursively copied after splitting the data into the training and evaluation datasets. This was effective and improved the performance slightly, the details of which are not reported here.

We used the BERT model for the Japanese language with an additional linear layer for the sentence classification. Parameters were to be tuned by cross-validation on the training data, but we trusted our expertise and experience; parameters were determined by intuition. Despite the fact that the average vector of each word embedding has been reported to be better for the sentence classification, the [CLS] vector was used. We conducted the training and evaluation five times with the data randomly shuffled.

Table1: Number of retweets and tweets.

Retweets	Tweets	Ratio
0	29,471	79.3%
1–9	7,055	19.0%
10–	658	1.8%

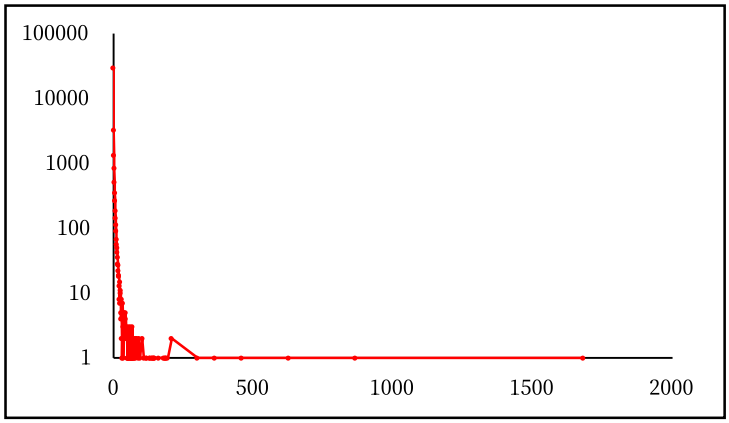

Figure 1: The horizontal axis indicates the number of retweets and the vertical axis indicates the number of tweets. Most tweets have not been retweeted.

Results

The results of the experiment are shown in Table 2. The model seems to fail to detect funny tweets that will be retweeted many times.

Table 2: Results. Standard deviation is also shown in the parentheses.

Retweets	Precision	Recall	F_1-score
0	0.37 (0.04)	0.22 (0.04)	0.27 (0.04)
1–9	0.51 (0.01)	0.53 (0.02)	0.52 (0.01)
10–	0.89 (0.00)	0.89 (0.01)	0.89 (0.00)

Discussion

We do not intend to discuss anything. This section exists so that this preface conforms to the IMRaD structure.

Reference

1. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 4171–4186.
2. Wolf, T., Chaumond, J., Debut, L., Sanh, V., Delangue, C., Moi, A., … & Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45.

Iwatsuki, K. (2021). Preface: is Impact of Scientific Humour Predictable?. *Scientific Sports*, *6*(2), 1–3.