CYUT at the NTCIR-15 FinNum-2 Task: Tokenization and Fine-tuning Techniques for Numeral Attachment in Financial Tweets

MIKE TIAN-JIAN JIANG
ZEALS CO, LTD.
TOKYO, JAPAN

YI-KUN CHEN SHIH-HUNG WU
CHAOYANG UNIVERSITY OF TECHNOLOGY
TAICHUNG, TAIWAN

ABSTRACT

The paper describes our submissions to the NTCIR-15-FinNum-2 shared task in financial tweets analysis.

The first run is our baseline system, which is based on the BERT model with our preprocessing strategy.

The second run is our fine-tuned system based on the XLM-RoBERTa pretraining model with more tokenization and fine-tuning techniques.

The macro-F1 of run 2 is 95.99% on development set, and 71.90% on formal test which ranked second best.

Proposed Approaches

We fine-tune two Transformer-based models, namely BERT and XLM-RoBERTa.

For the latter one, we further apply techniques developed by fastai for ULMFiT, such as discriminative fine-tuning and a variation of one-cycle policy .

Tokenization Trick

To better represent the structure of a financial tweet, we not only utilize XLM-RoBERTa's special tokens, namely the beginning of a sentence (<s>), the end of a sentence (</s>), and the separator of sentences (</s> </s>), but also customize a couple of tokens in the fastai convention of "xx" prefix that provides context. For example, consider a tokenized tweet below:

<s> _$ xxtag _RAD _about xxnum _9 _million _more
    _share s _than _the _90 _day _average . …
    </s>

The special tokens xxnum and xxtag annotate the numeral (_9 but not _90) and the cashtag (_RAD) in question, respectively. Combining with the actual subwords of number/cashtag right next to xxnum/xxtag, the annotated tokens provide certain features of the token sequence.

Fine-tuning Techniques
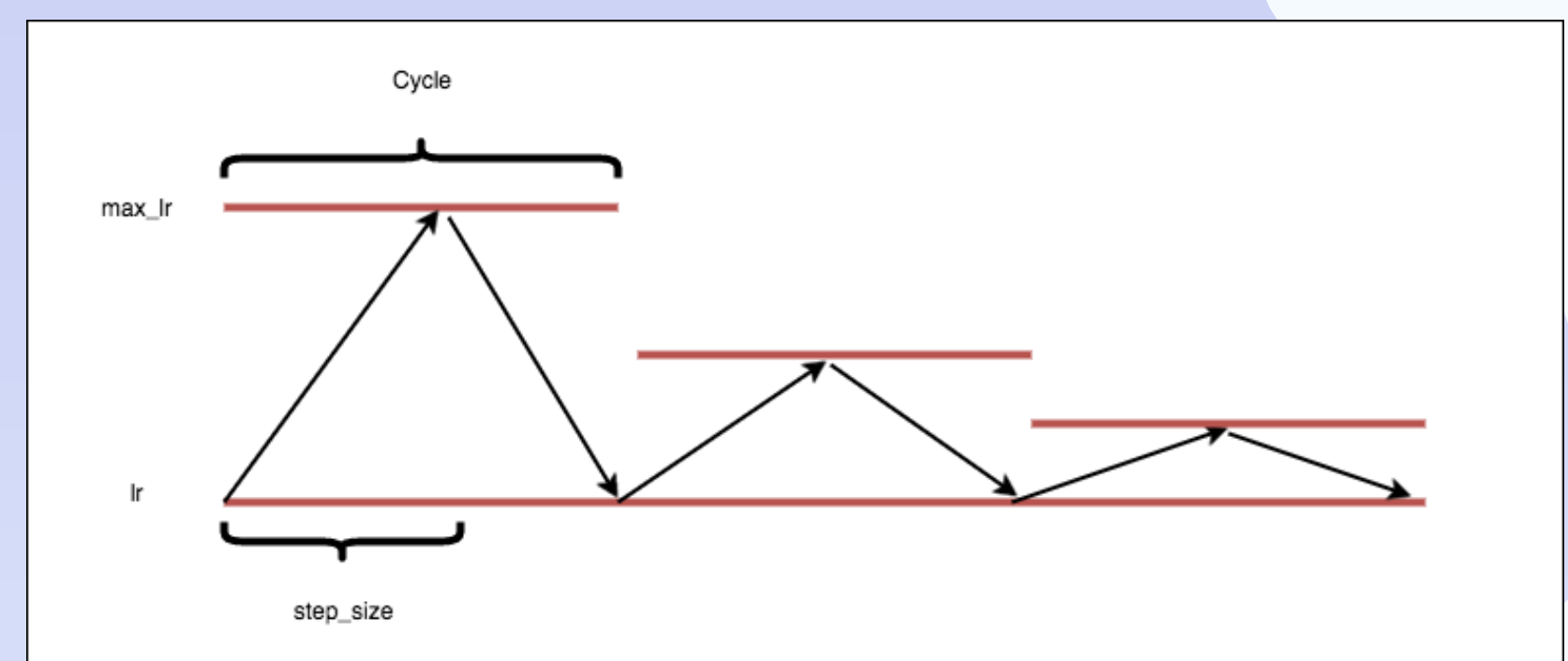
One-cycle Policy with a Max-learning-rate Decay.
Image credit: https://github.com/bckenstler/CLR

As above figure shows, it uses a warm-up and annealing for the learning rate while doing the opposite with the momentum.

Result and Discussion

Table. 1 Official Run Experimental results. (macro F-1 in %)

	Development	Test
Majority	44.88	44.93
CYUT-1	48.64	48.02
CYUT-2	95.99	71.90
Average of 17 runs	88.18	64.11

Table. 2 Additional Run Experimental results. (F-1 in %)

	Development	Test
CYUT-1 w preprocessing	86.6	62.7
CYUT-1 w/o preprocessing	49.9	49.2

The performance in the CYUT-2 is shown in Table 1. The performance of formal test also drops greatly. According to the overview report, the test result ranks second best .

The performance in the CYUT-1 w/o preprocessing in Table 2 can be seen as the performance of BERT in this problem.