

SLSTC at the NTCIR-13 STC-2 Task

Jun Guan
Waseda University, Japan
gkratos92@gmail.com

Tetsuya Sakai
Waseda University, Japan
tetsuyasakai@acm.org

ABSTRACT

SLSTC participated in the Chinese Subtask of the NTCIR-13 STC Task. We submitted one simple retrieval-based run, SLSTC-C-R1, which was generated by first retrieving a post from the Weibo repository and then selecting one comment for the retrieved post. Unfortunately, our run was not successful.

Team Name

SLSTC

Subtasks

Chinese subtask

Keywords

short text conversation; weibo; word2vec

1. INTRODUCTION

SLSTC participated in the Chinese Subtask of the NTCIR-13 STC Task [6]. We submitted one simple retrieval-based run, SLSTC-C-R1, which was generated by first retrieving a post from the Weibo repository and then selecting one comment for the retrieved post. This general strategy was inspired by the work of Chen, Song and Xie at the NTCIR-12 STC task [3].

2. RUN DESCRIPTION

Our run, SLSTC-C-R1, was generated as follows. Both the input post and the Weibo posts in the repository were converted into bags of words using Baidu stopwords [1] and Jieba Segmentation [2], and then into Word2Vec representations with 300 dimensions [4]. After vector regularisation, we obtained a post from the repository that had the highest cosine similarity with the input post.

The comments associated with the above selected post are the candidates for our output. For each comment c , we computed a score $S(c)$ as follows, and returned the comment with the highest score. Let the cosine similarity between the Word2Vec representations of the input post q and comment c be $sim(q, c)$. As each comment c in the repository has three quality labels, denoted here as $l_1(c), l_2(c), l_3(c)$, we define a simple weight reflecting the quality of c as:

$$w(c) = \frac{l_1(c) + l_2(c) + l_3(c)}{3} + 1. \quad (1)$$

Table 1: Comparison with the official highest/lowest performers at NTCIR-13 STC-2. Each * means “statistically significantly better than SLSTC-C-R1 at $p \approx 0$ ”; Each † means “statistically significantly better than ckip-C-G1 at $p \approx 0$ ”. The differences between SLSTC-C-R1 and ckip-C-G1 are not statistically significant.

| Run | Mean nG@1 | Mean P+ | Mean nERR@10 |
|------------|-----------|----------|--------------|
| SG01-C-G1 | 0.5867*† | 0.6670*† | 0.7095*† |
| SLSTC-C-R1 | 0.0750 | 0.1171 | 0.1148 |
| ckip-C-G1 | 0.0017 | 0.0029 | 0.0015 |

Let t be a term from comment c , and let $ctf(t)$ be the frequency of t in the repository (or *collection*). The overall score for c is computed as:

$$S(c) = w(c)sim(q, c) + \ln \sum_{t \in c} ctf(t). \quad (2)$$

Thus, our simple score takes into account the similarity of the comment with the input post, the quality of the comment with respect to the retrieved post, and the popularity of terms used in the comment.

3. RESULTS AND DISCUSSIONS

Table 1 compares our run with the official highest and lowest performers at NTCIR-13 STC-2 [6]. We ran a randomised Tukey HSD test [5] for each evaluation measure, with $B = 10,000$ trials. With every evaluation measure, SLSTC-C-R1 statistically significantly underperforms the highest performer SG01-C-G1 ($p \approx 0$). Whereas, the differences between SLSTC-C-R1 and the lowest performer ckip-C-G1 are not statistically significant: the p -values for nG@1, P+, nERR@10 are $p = 0.3700, p = 0.1013, p = 0.1295$, respectively. Thus, our run is statistically indistinguishable from the lowest performing run.

4. CONCLUSIONS

SLSTC participated in the Chinese Subtask of the NTCIR-13 STC Task. We submitted one simple retrieval-based run, SLSTC-C-R1, which was generated by first retrieving a post from the Weibo repository and then selecting one comment for the retrieved post. Unfortunately, our run was not successful.

5. REFERENCES

- [1] Baidu stop words.
<http://www.baiduguide.com/baidu-stopwords/>, 2016.
- [2] Jieba segmentation. <https://github.com/fxsjy/jieba>, 2017.
- [3] Z. Chen, R. Song, and X. Xie. Microsoft research asia at NTCIR-12 STC task. In *Proceedings of NTCIR-12*, 2016.
- [4] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. <https://arxiv.org/abs/1301.3781>, 2013.
- [5] T. Sakai. Metrics, statistics, tests. In *PROMISE Winter School 2013: Bridging between Information Retrieval and Databases (LNCS 8173)*, pages 116–163, 2014.
- [6] L. Shang, T. Sakai, H. Li, R. Higashinaka, Y. Miyao, Y. Arase, and M. Nomoto. Overview of the NTCIR-13 short text conversation task. In *Proceedings of NTCIR-13*, 2017.