

SLSTC at the NTCIR-13 STC-2 Task

Jun Guan, Tetsuya Sakai

Waseda University

1. Introduction

Goal

- Finding a good way to reuse Weibo comments to respond to new posts

Approach

- **Retrieval-based** method

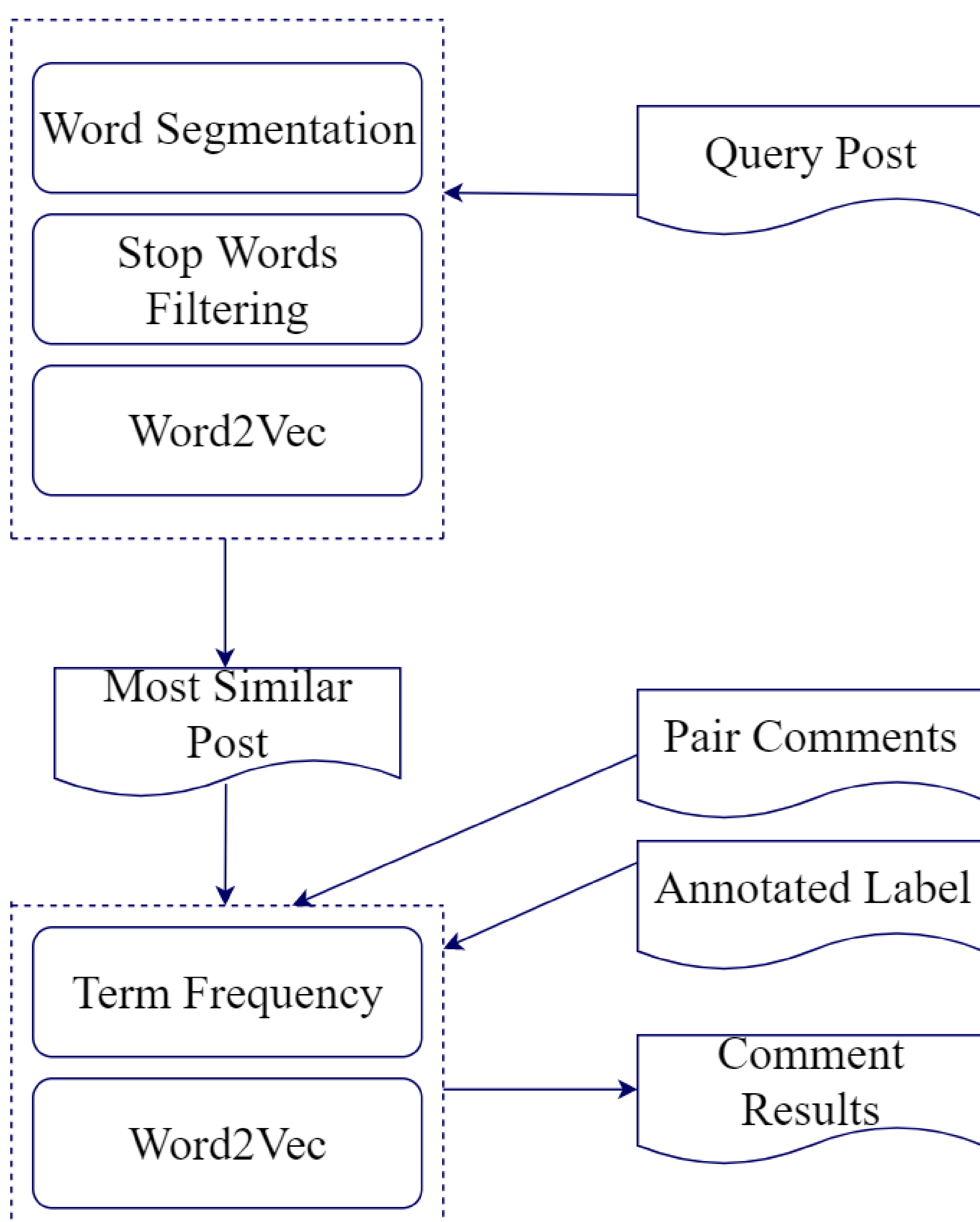
Results

- Our run is statistically indistinguishable from the lowest performing run

2. System

Tools

- Jieba Segmentation
- Baidu Stopwords
- Word2Vec



3. Method

- **Word2Vec** representation
- Select the most similar Post (cosine *sim*)
- Through all pair comments, rank them with their score $S(c)$

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

label score
↓
term frequency

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

4. Conclusion

- The differences between our run and the lowest performer are not statistically significant
- Comment ranking function may need to be improved
- Our run was not good enough

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

5. References

- [1] Z. Chen, R. Song, and X. Xie. at NTCIR-12 STC task
- [5] T. Sakai. Bridging between Information Retrieval and Databases (LNCS 8173) 2014.
- [6] L. Shang, T. Sakai, ... Overview of the NTCIR-13 short text conversation task