WUST at the NTCIR-14 STC-3 CECG Subtask

Zhanzhao Zhou, Maofu Liu, Zhenlian Zhang
School of Computer Science and Technology
Wuhan University of Science and Technology
liumaofu@wust.edu.cn

Contents

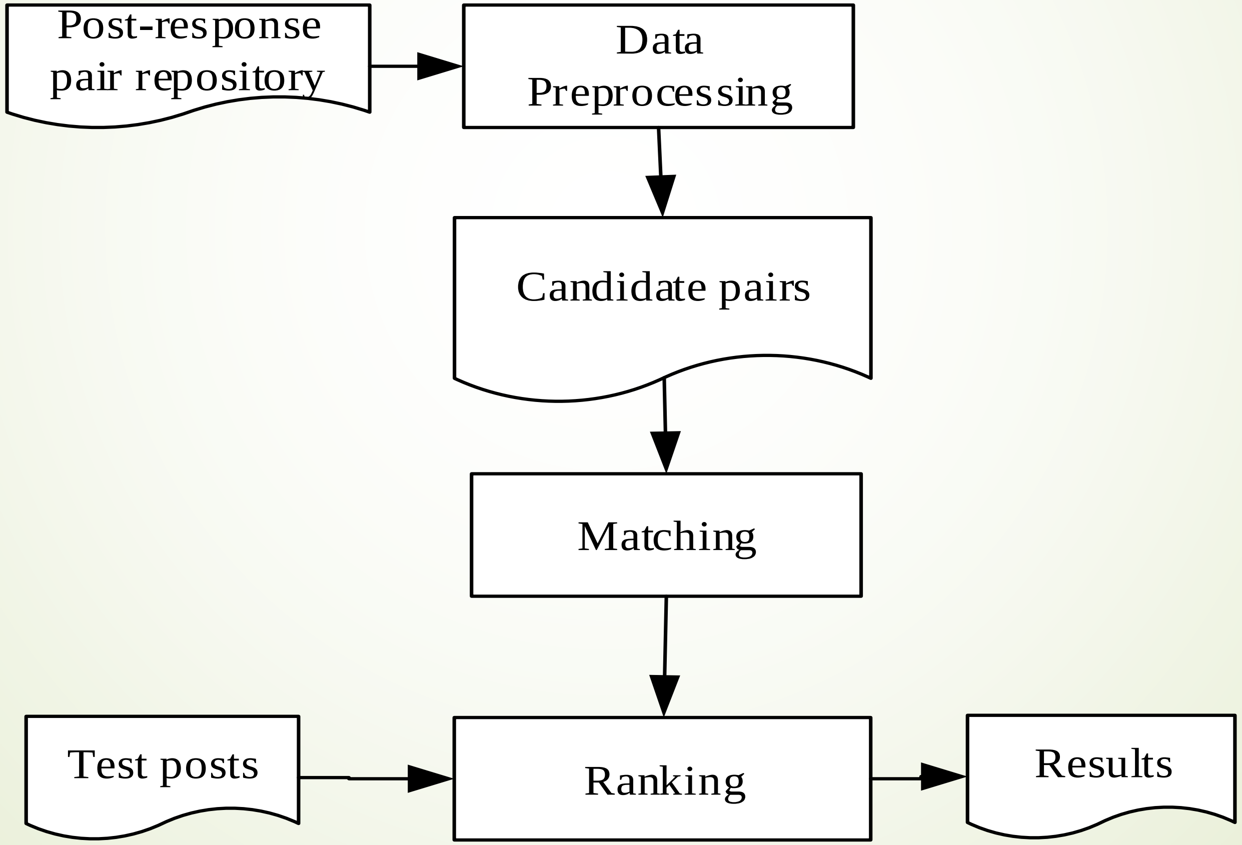
Introduction

- In human-computer interaction, a large quantity of researchers only focuses on the logical coherence and topic relevance of dialogue but pay little attention to the emotions contained in the dialogue. In fact, the understanding of the emotional expression is not only the significant cognitive behavior of human being but also the key to enhancing human-computer interaction.

- We regard the NTCIR-14 STC-3 CECG Subtask as an information retrieval problem. WUST system can retrieve proper responses both in content and emotion. The original training dataset is too large, and it takes too much time to calculate the text similarity.

System Architecture

- The WUST system includes three modules, i.e. candidate generation, candidate matching, and candidate ranking.



candidate generation

- Firstly, we construct the inverted index table for the training dataset.
- Secondly, According to the emotion categories of responses, we divide the training dataset into six subsets and mark the corresponding emotion categories respectively, so the six training subsets are obtained.
- Thirdly, for each testing instance, we search for appropriate responses to construct the candidate dataset in the preceding corresponding training subset whose emotion class is the same as the testing instance emotion class by the inverted index table.

candidate matching

- The vector space model (VSM) is used to represent the dialogue. The VSM expresses the query and each candidate dialogue into a vector in the same vector space and the text similarity is computed by the cosine similarity.

$$sim_{Q2P}(q,p) = \frac{\vec{q}^{T}\vec{p}}{||\vec{q}||||\vec{p}||}$$

$$sim_{Q2R}(q,r) = \frac{\vec{q}^{T}\vec{r}}{||\vec{q}||||\vec{r}||}$$

candidate ranking

- This paper uses a simple linear function to compute the scores of the candidate dialogues and rank them.

- $rank(q,c) = \alpha * sim_{Q2P}(q,p) + (1-\alpha) * sim_{Q2R}(q,r)$

- where the range of α is from 0 to 1. We adjusted the parameter α for several times and found that when α is 0.85, our system has achieved the best performance.

Experiments

- Emotion Consistency: whether the emotion class of a generated response is the same as the pre-specified class.
- Coherence: whether the response is appropriate in terms of both logically coherent and topic relevant content.
- Fluency: whether the response is fluent in grammar and acceptable as a natural language response.

$$OverallScore = \sum_{i=0}^{2} i * num_i$$

$$AverageScore = \frac{1}{N_t} \sum_{i=0}^{2} i * num_i$$

Experimental results

Table 1. Performances of the WUST model

Submissions/Emotions	Label0	Label1	Label2	Total	Overall Score	Average Score
WUST	601	211	188	1000	587	0.587
Like	117	36	47	200	130	0.65
Sadness	124	31	45	200	121	0.605
Disgust	111	69	20	200	109	0.545
Anger	137	48	15	200	78	0.39
Happiness	112	27	61	200	149	0.745

Conclusions

- This WUST system can generate appropriate and reliable responses both in content and emotion. But the method of text similarity calculation is too simple to capture the semantic information. In addition, the aforementioned lexical gap phenomenon also has a bad effect on the content of the responses.

- In the future, we would like to propose a novel model using the generated-based approach. And how to capture the semantic information and alleviate the lexical gap phenomenon is our chief work.

END

Thanks for listening！

2019.6.5