WUST System at NTCIR-13 Short Text Conversation Task

Maofu Liu, Yifan Guo, Yan Yu

School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065,China

liumaofu@wust.edu.cn, guoguoyifan@foxmail.com

Han Ren

Laboratory of Language Engineering and Computing, Guangdong University of Foreign Studies, Guangzhou 510006, China

hanren@whu.edu.cn

Introduction

◆ The STC is a core task of NTCIR-13. STC is a much simplified version of the human-computer conversation: One round of conversation formed by two short texts, with the former being an initial post from users and the latter being a comment given by the computer .

◆ In NTCIR-12, the STC is defined as an IR (Information Retrieval) problem. Besides the retrieval-based method, NTCIR-13 considers the generation-based method to generating new comments.

◆ We take STC as an IR problem. Given the new post, we assume the effectiveness of comments depends on the similarity between the new post and the old comment, or the similarity between the new post and the old post.

System Architecture
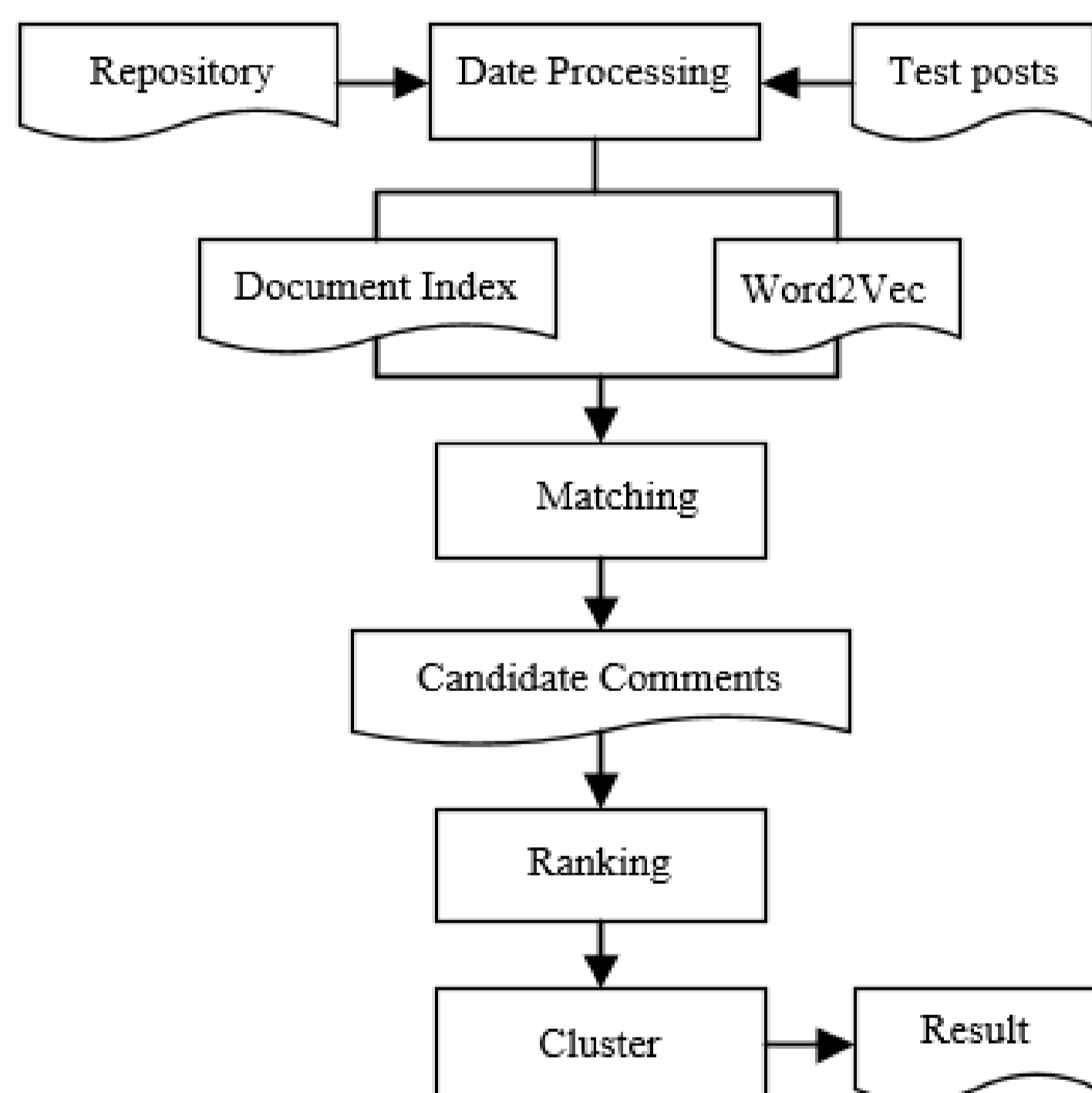
1. Data Preprocessing

◆ First, the system converts traditional Chinese to simplified Chinese. And then we use Ansj to split Chinese text into a sequence of words. Then we get topic words set T by TF-IDF (Term frequency–Inverse Document Frequency).

◆ Second, the system processes the post-comment pair repository and builds the inverted index table of the posts and words separately. In this way we do not need to scan all posts for each input, but instead retrieve a limit number of posts efficiently.

◆ Finally, we trained a Word2Vec model to generate most features. The model is an efficient method for learning high-quality distribution vector representations that capture a large number of precise syntactic and semantic word relationships.

2. Matching

◆ Given the query $Q,$ wo try to find the posts being similar to the query q and use their comments as the candidates .

◆ First, retrieves a number of candidate posts for the given query Q according to the topic words set T of Q.

◆ Second, we use VSM (Vector Space Model) for measuring the similarity between the query q and the candidate post $p.$

◆ Then, we use the VSM for measuring the similarity between the query q and the candidate

3. Ranking

We use some features as follows:

◆ Inverse document frequency (IDF): the sum of the IDF of the common words in the sentence.

◆ Longest Common String (LCS): length of the longest common substring between the query Q and the candidate comment.

◆ Overlapping similarity: The overlapping of some topic words in the two sentences.

◆ Word2Vec similarity: the Word2Vec similarity is the cosine similarity between two word vectors, generated from the sentence word vectors with the additive synthesize method.

◆ Then the system to further evaluate all the candidate comments, and assign a ranking score to each comment. We also use cluster to get diversity answers.

◆ Finally, the system returns the top-10 comments

Experiments

◆ The official evaluation results are listed in Table 1.

Table1. Formal run experiment official results

Run	WUST-C-R1	WUST-C-R2
Mean nDCG@1	0.071	0.094
Mean nDCG@1	0.0984	0.1409
Mean nERR@10	0.0927	0.1349

◆ We submitted two results of our system for STC task in Chinese by retrieval-based method. Table 2 only lists the official evaluation results of our group.

◆ There are only 52 test posts returned appropriate comments by analyzing the results of our system. So the average score is very low, and even many posts have zero score.

◆ One reason is that our model uses the simple VSM rather than semantic similarity for measuring query-posts and query-comments similarities.

Conclusions

◆ In this paper, we have described our model based on VSM for STC task in Chinese. We also analyzed our submitted experimental results and adjusted parameters which outperformed than former.

◆ In the future, We need to enhance the accuracy by matching query and response in terms of semantic relevance, speech act, and entity association. Wse also would like to generate the appropriate and human-like response derived from what we searched from the post-comment pair repository.