

Nders at the NTCIR-13 Short Text Conversation 2 Task

Han Ni

NetDragon WebSoft Inc., China
nihan@nd.com.cn

Liansheng Lin

NetDragon WebSoft Inc., China
Fuzhou University, China
linliansheng@nd.com.cn

Ge Xu

Minjiang University, China
XuGeNLP@nd.com.cn

Abstract

This paper describes our retrieval-based approaches at NTCIR-13 short text conversation 2 (STC-2) task (Chinese). For a new post, our system firstly retrieves similar posts in the repository and gets their corresponding comments, and then finds the related comments directly from the repository. Moreover, we devise two new methods. 1) LSTM-Sen2Vec model to get the vector of sentence. 2) Pattern-IDF to rerank the candidates from above. Our best run achieves 0.4780 for mean nG@1, 0.5497 for mean P+, and 0.5882 for mean nERR@10, and respectively ranks 4th, 5th, 5th among 22 teams.

Result

Run	Mean nG@1	Mean P+	Mean nERR@10
Nders-C-R1	0.4593	0.5394	0.5805
Nders-C-R2	0.4743	0.5497	0.5882
Nders-C-R3	0.4647	0.5317	0.5768
Nders-C-R4	0.4780	0.5338	0.5809
Nders-C-R5	0.4550	0.5495	0.5868

Fig. 2: The official results of five runs for Nders team

Our system

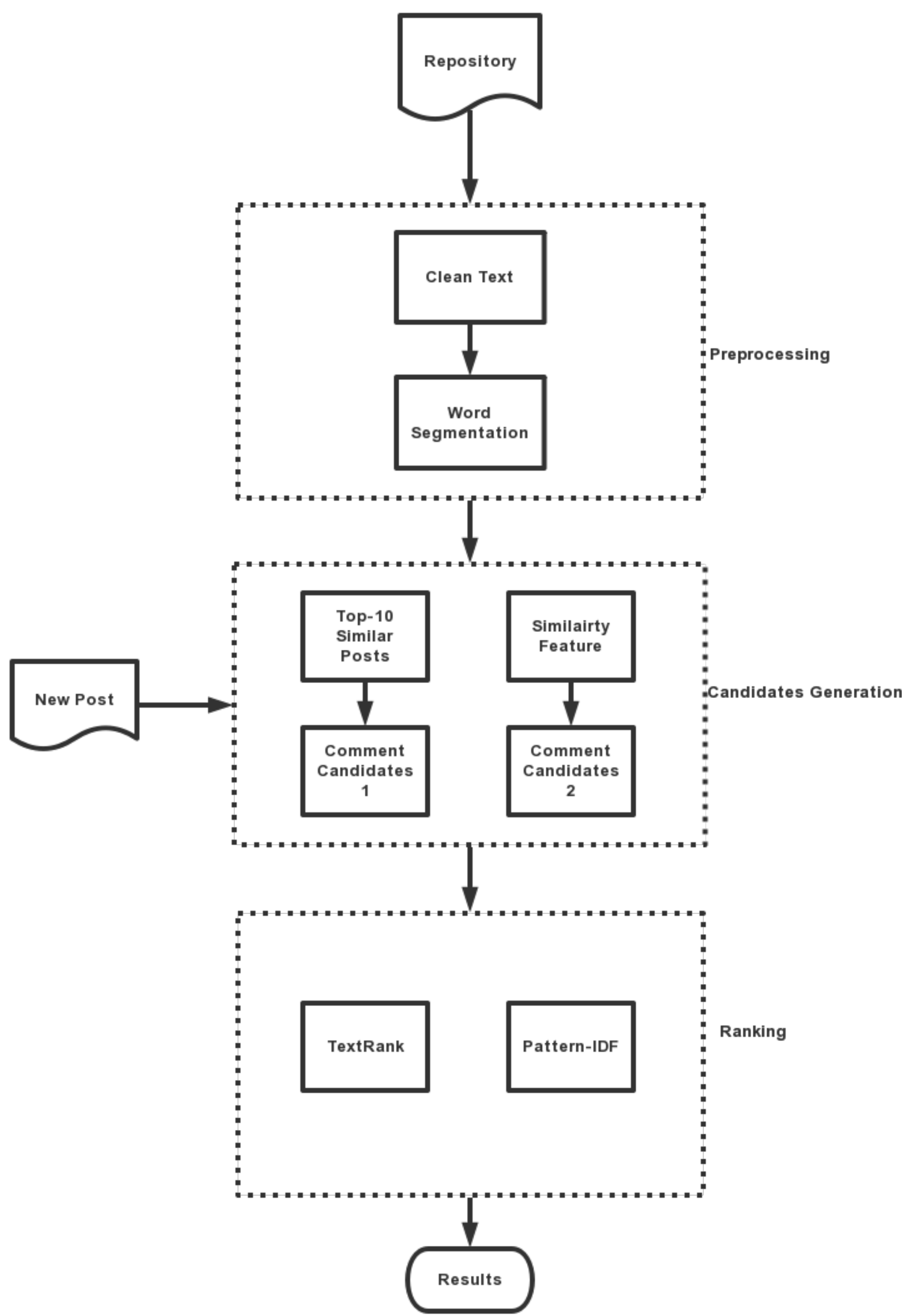


Fig. 1: System Architecture

Conclusions

In this paper, we propose an approach for STC-2 task of NTCIR-13. The LSA, Word2Vec and LSTM-Sen2Vec model are used to find similar posts. The LSA and Word2Vec model are used to retrieve comment candidates. A graph-based algorithm TextRank and the Pattern-IDF we devised are applied to rank the candidates. Results show that the Pattern-IDF we devised is effective for ranking while TextRank not, and LDA model outperforms LSA model in retrieving candidates. Finally, our best run achieves 0.4780(R4) for mean nG@1, 0.5497(R2) for mean P+, and 0.5882(R2) for mean nERR@10, which respectively ranks 4th, 5th, 5th among 22 teams.

Acknowledgements

Thanks for ...