

Supplementary Materials: Target-Guided Open-domain Conversation

1 Keyword Extraction

Following the previous work (Wang et al., 2018; Yao et al., 2018) on conversation keyword extraction, we develop a TF-IDF based keyword extractor. For a conversation $\{u_1, u_2, \dots, u_n\}$, our topic extractor will obtain a keyword list $\{t_1, t_2, \dots, t_k\}$ from each utterance u_i by the following steps: (1) We regard each utterance in a conversation as a document and each word as a term to calculate the TF-IDF value of each word. (2) We ignore the words appearing less than 10 times in all corpus, or have been mentioned in the former utterance of the current conversation. (3) Considering that the importance of each part of speech in a sentence is different, we set different weights to distinguish them. The part-of-speech weights of nouns, verbs, adjectives and others are 2, 1, 0.5 and 0 respectively. (4) We multiply the TF-IDF value with the part-of-speech weight to obtain the composite score of each word, and delete the words whose score is below the given threshold. We consider the words left as the keywords of an utterance in an conversation.

References

- Wenjie Wang, Minlie Huang, Xin-Shun Xu, Fumin Shen, and Liqiang Nie. 2018. Chat more: Deepening and widening the chatting topic via a deep model.
- Lili Yao, Ruijian Xu, Chao Li, Dongyan Zhao, and Rui Yan. 2018. Chat more if you like: Dynamic cue words planning to flow longer conversations. *arXiv preprint arXiv:1811.07631*.