DocChat: An Information Retrieval Approach for Chatbot Engines Using Unstructured Documents

Author(s): Zhao Yan, Nan Duan, Junwei Bao, Peng Chen, Ming Zhou, Zhoujun Li, Jianshe Zhou

Reference: Zhao Yan et al.2016. DocChat: An Information Retrieval Approach for Chatbot Engines Using Unstructured Documents. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*(2016). DOI:http://dx.doi.org/10.18653/v1/p16-1049

**What they are trying to do?**

Some chatbots used retrieval and generation based methods. With the former, the system retrieves the most probably question-response (Q-R) pairs that best matches with what the user’s Q is to generate a response R for the user. The latter uses an encoder-decoder framework that first encodes a Q, feeds to the decoder to generate a response R. These have costs to fluency and naturality. They built DocChat as a novel response retrieval approach to find responses from unstructured documents. DocChat selects a sentence from given documents directly by ranking all possible sentences based on designed features. It does this instead of looking for the best Q-R pairs or performing language generation.

DocChat retrieves responses in a three step process:

First it retrieves response candidates from a document based on the question Q.

Here BM25 weighting formula is used to retrieve response candidates from docs. For each document, for each sentence selected, the next and previous sentence is also selected. <BOD> and <FOD> are added to the beginning and end.

Ranks responses candidates retrieved and selects the most possible response candidate

Given a question and statements retrieved, it ranks according to a function using features at various levels i.e.Word-level feature, Phrase-level, sentence-level, document-level, relation-level, type-level and topic-level.

They employ a learning-based learning to rank method to train response ranking.

Each sentence in a set of candidate sentence are given a denotion of + or -.

Question and answers are obtained from Q and A sites and the correct answers to questions are given + and wrong answers -

Performs response triggering where it checks if it is confident enough to respond to the question using the statement it found. The trigger is either true or false.

There are chit chat utterances and informative utterances. The latter is more suitable to work. There is no guarantee that responses will return a candidate set with at least one suitable response. It returns yes if two conditions are satisfied: the length of the sentence is less than a pre-defined threshold and the sentence does not start with phrases like but, also, besides as these phrases depend on their context sentences hence are not suitable.

**Evaluation**

Experiment

They used the WikiQA corpus which is a publicly available set of question and sentence pairs.

Each sentence is denoted by 1 or 0. 1 means sentence is a correct answer sentence and 0 is the opposite.

When given a question it selects answer sentences from all the answers in the corresponding document

**Results**

Performance is evaluated using Mean Average Precision and Mean Reciprocal Rank. Lexical features had the highest impact on the results

They compare DocChat to other baseline methods from previous papers and they found that without using WikiQA’s training set DocChat can achieve comparable performance. When it is combined with one of the baseline models, it achieves the best result.

This is because:

Training data comes from open domain sources so they can expect high adaptation capability

For response triggering, Precision, Recall and F1 was used. DocChat performed better than Yang’s which used rich lexical semantic features.