Digital Media Laboratory

Weekly Report November 7, 2021

Ali J. Alaee Report No. 15

alialaee98@gmail.com 00/08/10- 00/08/16

Abstract

I got into the details of [1] to understand it deeply for implementation.

Description

The work of this paper [1] can get divided into upcoming parts. Moreover, I compare the approaches introduced in the paper with our previous embedding-based IR setup. At some point, the authors claim that these methods rank the relevant retrieved documents for a given query. Consequently, we could retrieve k-relevant items with our global embedding system and then rank them with the help of our personalized embedding space.

1. The paper asserts that we should prepare two types of embedding space; a global one and a personalized one specialized for each user. Due to the shortage of training data for fine-tuning the global embeddings to get personalized forms, we can utilize top-k similar users' search sessions instead of only one user's search sessions. The formula for obtaining top-k similar users to one particular user is illustrated below. In addition, to make the implementation of personalized embeddings feasible in terms of memory management, vocabulary for each of them should obey some shrinking policies.
   In contrast to our previous implementation based on Search2Vec, the embedding space represents the semantics of "words" in this paper, so each multi-word query or document is a set of vectors in the embedding space rather than one single vector (as in our model).

2. In the second step, four matrices are obtained for each query and document pair, a global and a personalized version of each. After some contextual extraction operations (a multi-head self-attention layer), these matrices go through a similarity ranking component called KNRM. This component is elaborately discussed in this [2] paper. It has to find context similarity between a query and a document, given their word-by-word embedding representation. Compared to our current model, in which a cosine distance can simply compute the relevance of each query and item, KNRM might be unnecessary. However, the full implementation of KNRM is available at this link.

3. The available results of five KNRM components, each indicating similarities between a given query and document from a particular aspect, as inputs of a multi-layer perceptron, the final decision is made, and a number illustrating the similarity of the given pair is obtained. However, in our model, only two similarities might be obtained; one from global embeddings and one from personalized. If according to the first paragraph, we decide to utilize the first embedding as the first filter to obtain relevant results and then rank them based on the personalized embeddings, this step would not be necessary.

In the experiment section, this method has been tested on the available AOL dataset. One question that remains in my head is that documents in that dataset are domain URLs, so they are not multi-word descriptions. Metrics used for evaluation are MAP, MRR, and P@1.

Next Week

1. Discuss how much detail we want to include in our model.

2. Learn about ranking metrics described in the last paragraph.

References

[1] Jing Yao et al. Employing personal word embeddings for personalized search. *SIGIR '20: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020.

[2] Xiong et al. End-to-end neural ad-hoc ranking with kernel pooling. *SIGIR '17: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2017.