



Dense Passage Retrieval for Multi-Hop Open-Domain Question Answering

Open-domain Question Answering is a task that answers factoid questions using an extensive text collection. It relies on a two-stage framework: 1) A context retriever retrieves the most relevant passages to a question from a large corpus. 2) A machine reader examines the retrieved context and identifies the position of the correct answer.

In this thesis, we would like to extend DPR to multi-hop question answering, evaluating on a benchmark called BeerQA, which combines the questions from the single-hop

SQuAD Open and the two-hop HotpotQA, consisting of 530 human-annotated questions that require information from at least three Wikipedia pages to answer. Furthermore, we need to determine the number of reasoning steps necessary for answering the question. We attempt to tackle this problem with different approaches: 1) We may test and determine a threshold, compare it with the similarity score we computed during inference, and decide if relevant passages still need to be retrieved. 2) We may use classification algorithms to classify the number of hops.

Requirements: Strong motivation, knowledge in deep learning, or a solid background in machine learning. Experience with PyTorch is an advantage. We will have weekly meetings to address questions, discuss progress and think about future ideas.

Interested? Please contact us for more details!

Contact

- Zhao Meng: zhmeng@ethz.ch, ETZ G60.1

