

Dense Passage Retrieval for Open-Domain Question Answering

...

20 de abril de 2023

Conceitos Importantes

- “Open-domain question answering (QA)” - Tarefa que responde a perguntas usando uma grande coleção de documentos.
- Modelos de compreensão de leitura sugerem uma estrutura de dois estágios muito simplificada:
 - Um contexto O recuperador primeiro seleciona um pequeno subconjunto de passagens onde algumas delas contêm a resposta para a pergunta
 - Um leitor de máquina pode examinar minuciosamente os contextos recuperados e identificar a resposta correta

Conceitos Importantes

- As representações são otimizadas para maximizar os produtos internos da pergunta e os vetores de passagem relevantes, com o objetivo de comparar todos os pares de perguntas e passagens em um lote.
- Dense vector: vetor com valores contínuos, o qual carrega informações semânticas latente e que podem ser usados para busca por similaridade.

Contribuições do Artigo

- Foi demonstrado que, com a configuração de treinamento adequada, o simples ajuste fino dos codificadores de perguntas e passagens em pares de perguntas e passagens existentes é suficiente para superar o desempenho do BM25 (dentro do domínio).
- Foi verificado que, no contexto de resposta a perguntas de domínio aberto, uma maior precisão de recuperação realmente se traduz em uma maior precisão de controle de qualidade de ponta a ponta.
- Com o estudo de ablação foram comparadas funções de similaridade e o produto interno de vetores foi o que gerou melhores resultados.

Dúvida?

- Com a ajuda do índice de memória FAISS para vetores de valor real, o DPR pode se tornar incrivelmente eficiente, processando 995,0 perguntas por segundo, retornando as 100 melhores passagens por pergunta. Em contraste, o BM25/Lucene (implementado em Java, usando índice de arquivo) processa 23,7 questões por segundo por thread da CPU.
 - Isso de fato foi verificado?
 - Para buscas densas não deveria ser mais lento o processo?

Obrigado

Manoel Veríssimo
