

# Paper Review

Peitian Zhang

Microsoft Research Asia

July 7, 2021

- 1 Basics
  - Workflow
  - Datasets
  - Metrics
  
- 2 Papers
  - DeepCT

- Create inverted-index
- Use retrieval model such as BM25 and query likelihood to recall several passages according to the query

### Hint

It's not so clear to me that how weighted inverted index works.

- 1

- Recall

# Overview

- Generate the contextualized word weight according to its embeddings (produced by BERT).

$$y = Wx + b, \quad (1)$$

where  $x$  is the BERT embedding of the word,  $W$  and  $b$  are parameters.

- The weights is used to construct inverted index and applied in the first-stage retrieval.

## Insight

Select top  $K$  terms by DeepCT.

# Training

- Fine-tune BERT with ground-truth weight for each term in the document, which is estimated by the cooccurrence of the term in the doc and the query.
- Mean square loss.