Wei SONG, Ying LIU, Li-zhen LIU, Han-shi WANG, 2018. Semantic composition of distributed representations for query subtopic mining. *Frontiers of Information Technology & Electronic Engineering*, 19(11):1409-1419.
https://doi.org/10.1631/FITEE.1601476

Semantic composition of distributed representations for query subtopic minining

Key words: Subtopic mining; Query intent; Distributed representation; Semantic composition

Corresponding author: Li-zhen LIU
E-mail: Liz_Liu7480@cnu.edu.cn
ORCID: https://orcid.org/0000-0002-9125-4326

Motivations

1. Query subtopic mining is significant for representing and inferring query intent for information retrieval.

2. Query subtopic mining is challenge because of the nature of short query.

3. Recently, distributed representations have been developed and impacted many fields.

Main ideas

1. We build the distributed representations of queries and query intent phrases for query subtopic mining to deal with the term mismatch problem.

2. We compare the impacts of distributed representation learning with traditional representations and different data types.

Methods

1. Two types of strategies are explored:

   (1) **Paragraph vector**: We directly learn the distributed representations of an arbitrary length of word sequences;

   (2) **Word vectors and semantic composition**: We learn the distributed representations of words, and then adopt semantic composition approaches to obtain the distributed representations of phrases.

2. Two types of datasets are used:

   (1) Text data: Baidu Baike;

   (2) Query logs.

Major results and conclusions

1. Word vector and semantic composition performs better than the paragraph vector approach.

2. Multiplicative semantic composition is the most effective method.

3. We use the distributed representation methods with the traditional term based representations.

4. Representations learned from query log data perform slightly better.