

Finding Similar Experts

Krisztian Balog and Maarten de Rijke

Task

Given a set of **example experts**, return **similar experts**.

Research Questions

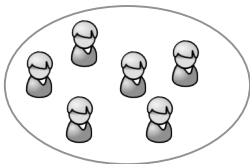
- ❖ What are effective ways of representing experts for this task?
- ❖ How does the size of the sample set affect end-to-end performance?
- ❖ How to evaluate results?

Representing experts and measuring similarity

Representing a candidate expert ca

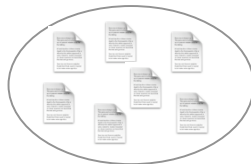
Set of people

ca is working with



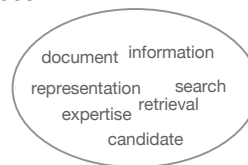
Set of documents

associated with ca



Set of terms

with highest TFIDF,
extracted from associated
docs



TFIDF weighted vector of terms

extracted from associated
docs

association candidate
document example
expertise extract information
precision representation
retrieval search term

Similarity measure

Jaccard coefficient

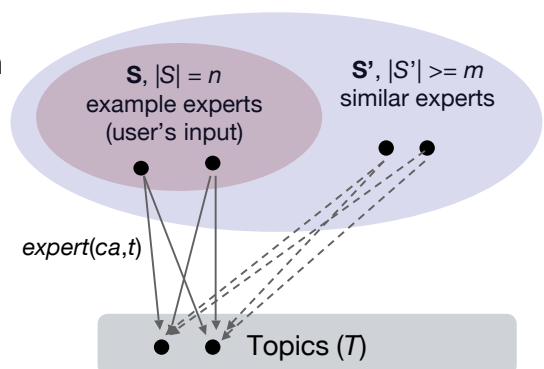
Cosine distance

Experimental design

- ❖ Simulate user's input and generate ground truth

Algorithm

1. Select a set of candidates S at random who are all experts on a set of topics T
 2. S' is the set of additional experts who are experts on T
 3. Sample set S is valid if $|S'| \geq m$
- ❖ TREC 2006 topics and qrels are used to define $expert(ca, t)$



Results

- ❖ More fine-grained representations of candidates consequently result in higher performance
- ❖ Larger input size lead to higher scores
- ❖ Best representation (TERMVECT) delivers excellent performance achieving $MRR=0.853$, $P@5=0.703$ (for $n = 5$)