An Approach for Weakly-Supervised Deep Information Retrieval

Sean MacAvaney, Kai Hui, Andrew Yates

Motivation

Relevance judgements are expensive to collect. We explore techniques to use alternative data sources for training data-hungry neural IR models.

In this work, we evaluate our methods using the New York Times corpus.

Pseudo-query → **GAS INVENTORIES ~~AT A~~ 5-YEAR LOW**

Pseudo-document → The nation's gasoline inventories fell last week to their lowest levels in almost five years, the government said yesterday, quantifying for the first time Hurricane Katrina's effect on energy supplies and production. Americans bought and used

Source: New York Times, September 5, 2009

Problems

Hard-Negative: Randomly chosen non-relevant documents are trivial to distinguish.

Mismatched-Interaction: Not all pseudo-relevant text pairs interact like relevant query-document pairs (e.g. poetic headlines.)

Figurative headline → **WHEN BIRD FLIES IN**

Larry Bird was in town the other day, and left, but chances are he'll be heard from again. He didn't arrive in disguise, and he didn't leave in disguise. Not quite. His game's the same - fill the hoop, fill the lane, fill the other team with vexation.

Source: New York Times, November 12, 1987

Ranking Filter

Rank pseudo-documents using an unsupervised ranker (e.g. BM25) and…

(1) Disregard pseudo-queries that do not retrieve their pseudo-relevant document in the top n_{rank}.

(2) Select the top n_{neg} retrieved documents are negative training examples.

- Relevant document
- Non-relevant documents
- Discarded documents
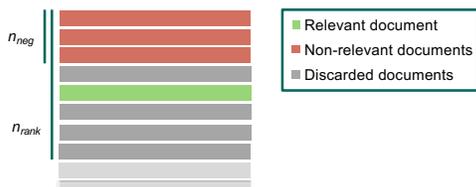
Interaction Filter

General Approach: Generate *mock interaction embeddings* and filter training examples down to those the most nearly match a set of template query-document pairs (given a distance function). Since interaction embeddings specific to what a model "sees," interaction filters are model-specific.

PACRR Interaction Filter: Interaction embedding calculated as maximum value for each query term in similarity matrix. Distance function is the *aligned mean squared error*—the MSE under the alignment that results in the minimum value.

Results

The technique improved results for models that encode ngram interactions (PACRR and MatchPyramid with 5×5 kernel).

Future work:
- Develop interaction filters for other models
- Alternative sources of template pairs for interaction filter
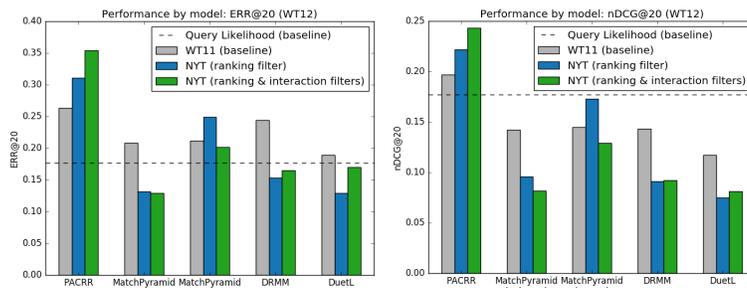
(Similar results when evaluated on WT13 and WT14)