AUEB at BioASQ 6: Document and Snippet Retrieval

George Brokos, Polyvios Liosis, Ryan McDonald, Dimitris Pappas and Ion Androutsopoulos
https://github.com/nlpaueb
Document and Snippet Ranking

BioASQ Task 6B Phase A
For each question
1. Select and rank 10 documents
2. Select and rank 10 snippets from these docs
   a. No restriction on length
   b. A document can yield multiple snippets
   c. A document can yield no snippets

RELEVANT DOCUMENTS

SUPPORTING SNIPPETS
Overall System Architecture
Document Ranking

Deep Relevance Ranking Using Enhanced Document-Query Interactions
McDonald, Brokos and Androutsopoulos
EMNLP 2018
Saturday November 3rd at 9:36
Session: 5C: IR / Text Mining
(Deep) Ad-hoc Retrieval / Relevance Ranking

- **Interaction-based**
  - DeepMatch (Lu and Li 2013)
  - ARC-II (Hu et al. 2014)
  - MatchPyramid (Pang et al. 2016)
  - **DRMM** (Guo et al. 2016)
  - **PACRR** (Hui et al. 2017)
  - DeepRank (Pang et al. 2017)

- **Relevance-based**

![Query-Doc term similarity matrices](image)

![Relevance Score Diagram](image)
Deep Relevance Matching Model (DRMM)

Guo et al 2016

Term Score Aggregation ⊕
Term Gates
Term Score
Dense Layers
Doc-aware
Query-term Encodings
Doc-Query Interaction ⊘

Relevance Score

Query-terms: $q_1, q_2, q_3$
Document-terms

Fixed-width
Deep Relevance Matching Model (DRMM)

Guo et al 2016

Relevance Score

Cosine Similarity Histograms
Attention-Based ELEment-wise (ABEL-DRMM)

Element-wise Function (e.g., Hadamard)

Differentiable End-to-end Training

Attention-based doc-encoding
Attention-Based
ELement-wise
(ABEL-DRMM)

Differentiable
End-to-end
Training

Context-sensitive
Encoder
TCN (Bai et al. 2018)

Element-wise Function
(e.g., Hadamard)

Attention-based
doc-encoding
ABEL-DRMM+: Add density

Max ABEL-DRMM score over windows
For illustration: Score = number of term matches

Final Score
document-ABEL-DRMM
+ max-window-ABEL-DRMM
Document MAP from BioASQ 6B Phase A (Average over Runs 1-5)
Devil’s in the details … (read the paper)

- Re-ranked top 100 documents from pubmed index using BM25 score
- System 1-3 are combinations of multiple runs (8)
- All models use final linear layer over
  - Severyn and Moschitti 2015, Mohan et al. 2017
- Years 1-5 used to train models
- Pairwise training with neg. sampling
Snippet Ranking
Basic CNN Text Matching

Yin et al. 2016
Snippet MAP from BioASQ 6B Phase A (Average over Runs 1-5)

BCNN
Snippet MAP from BioASQ 6B Phase A (Average over Runs 2,3,5)

BCNN
Training-testing mismatch

- Snippet extractor trained only on gold documents
- Learns to extract good snippets from good documents
- Not calibrated for non-relevant documents seen at test time

- Heuristic -- only extract snippets from high scoring docs
  - Often returns less than 10 documents
BCNN

High conf docs
Conclusions

● Deep IR models are effective for ranking docs and snippets from Biomed data
  ○ Propose in paper: End-to-end training with extensions to DRMM
  ○ Need to incorporate deep IR scores with traditional features

● Snippet retrieval: incorporate document score
  ○ Heuristics go a long way
  ○ More principled methods likely to improve more
Thanks!

https://github.com/nlpaueb/aueb-bioasq6