Utilizing Passage Similarity Metric for Passage Ranking

TREMA-UNH at TREC 2019
[1]Sumanta Kashyapi (sk1105@wildcats.unh.edu)
[1]Jordan Ramsdell (jsc57@wildcats.unh.edu)
[1]Shubham Chatterjee (sc1242@wildcats.unh.edu)
[1]Dr. Laura Dietz (dietz@cs.unh.edu)
[1]Department of Computer Science, University of New Hampshire

Objective

- To find out how to improve passage ranking using a Wikipedia passage similarity metric.
- To figure out the best passage similarity metric suitable for ordering passages in a ranking.

Unsupervised Similarity Metric

For a pair of passage p_1,p_2

- **Symmetric BM25 used in UNH-bm25-[Pre-processing]**

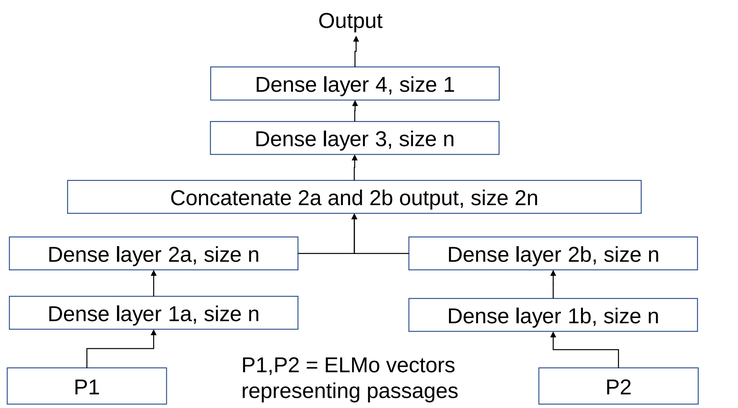
$$score = \frac{BM25(p_1, p_2) + BM25(p_2, p_1)}{2}$$

- **TFIDF similarity used in UNH-tfidf-[Pre-processing]**

$$score = cosine(v_1, v_2)$$

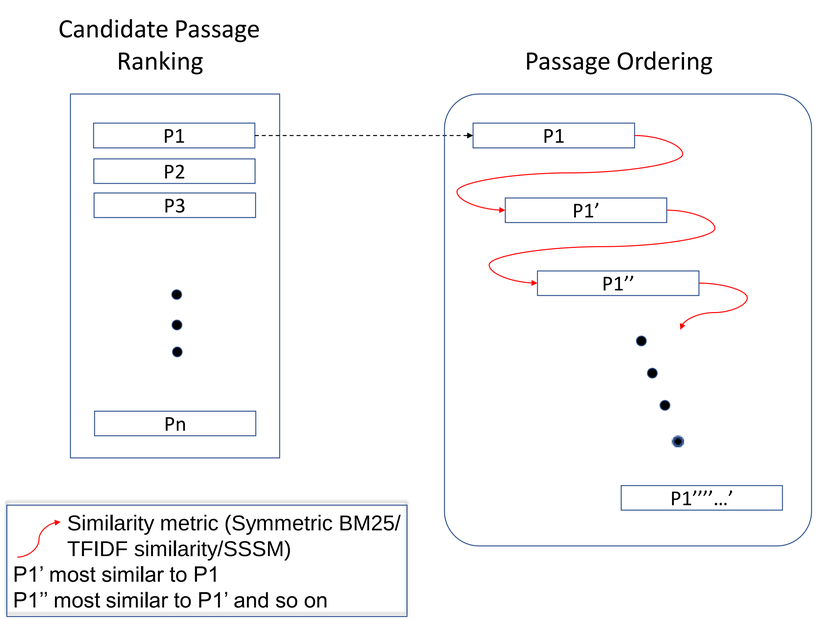
BM25(q,d) = BM25 score between query q and document d
$\mathbf{v_i}$ = TFIDF vector of passage p_i
Pre-processing = stemming (stem)/ lemmatization (lem)/ no pre-processing (pt)

Supervised Siamese Similarity Metric (SSSM)

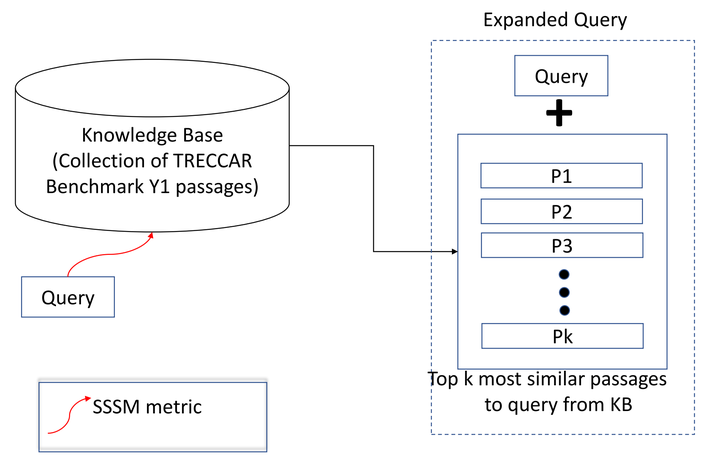


Siamese architecture used to model SSSM used for **UNH-dl[layer size]**

TRECCAR track: Daisy Chain Passage Ordering



Query Expansion using SSSM (UNH-exDL)



Results

TRECCAR track

Method	Facet Overlap	Relevance
UNH-bm25-stem*	0.0622 ± 0.0096	0.1141 ± 0.0165
UNH-tfidf-ptsim*	**0.0756 ± 0.0115**	0.1230 ± 0.0174
UNH-tfidf-lem*	0.0686 ± 0.0105	0.1150 ± 0.0165
UNH-tfidf-stem*	0.0674 ± 0.0105	0.1168 ± 0.0165
UNH-dl100* (dense layer size 100)	0.0403 ± 0.0072	0.1134 ± 0.0165
UNH-dl300* (dense layer size 300)	0.0335 ± 0.0065	0.1093 ± 0.0159
UNH-bm25-ecmpsg	0.0295 ± 0.0065	0.0931 ± 0.0175
UNH-bm25-rm	0.0658 ± 0.0114	**0.1297 ± 0.0170**
UNH-neural	0.0295 ± 0.0065	0.0931 ± 0.0139
UNH-ecn	0.0016 ± 0.0010	0.0188 ± 0.0040
UNH-qee	0.0427 ± 0.0079	0.1201 ± 0.0162

DL track

Method	MAP	Mean NDCG	Mean P@10
UNH-bm25 (BM25 baseline)	0.2565	0.5546	0.3465
UNH-exDL*	0.0364	0.1400	0.0605

* Methods which use Similarity metric

Conclusion

- Current results show that unsupervised similarity metrics outperform SSSM when used for passage ordering task. In future we plan to improve the similarity metric using an ensemble of all the metrics.
- We observe that our query expansion model performs poorly on the re-ranking task. This suggests that our choice of knowledge base for the query expansion (TRECCAR benchmark Y1) is not suitable for the task.