Overcoming Low-Utility Facets for Complex Answer Retrieval

Sean MacAvaney[1], Andrew Yates[2], Arman Cohan[1], Luca Soldaini[1],
Kai Hui[3], Nazli Goharian[1], Ophir Frieder[1]
[1] IRLab, Georgetown University, {firstname}@ir.cs.georgetown.edu
[2] Max Planck Institute for Informatics, ayates@mpi-inf.mpg.de
[3] SAP SE, kai.hui@sap.com

Dual submission to the KG4IR Special Issue.

Complex Answer Retrieval (CAR) is the process of retrieval for questions whose answers require many details or additional context to explain thoroughly [2]. These questions can be formulated as having a topic entity and a facet. For instance, for the question "Is cheese healthy?", the topic is "Cheese" and the facet is "Health effects". We observe that some facets, such as "Health effects", exhibit low utility: answers to questions about the health effects of cheese are unlikely to use the terms directly. Instead, they will include related entities, such as nutrients or related diseases. We call these low-utility facets because the terms in the facet are not used directly in the text, and thus the terms themselves do not provide much value. In contrast, high-utility facets use language that is specific to the topic and can be found directly in relevant answers (e.g., a facet of "Curdling' for the question "Why does cheese curdle?"); it would be difficult (and unlikely) that an answer to this this question does not include the term "curdle" or "curdling". In this talk, we propose a two-pronged approach for CAR by modifying a leading neural information retrieval architecture (PACRR [3]).

First, we introduce two estimators of facet utility. We observe that the CAR query structure (provided as a list of headings) itself implies a notion of utility: the root of the query is the topic (which is likely high-utility), any intermediate facets provide structure (which are likely low-utility), and the leaf (main) facet can be either, depending on the nature of the question. To further disambiguate between high- and low-utility facets, we incorporate facet frequency statistics, with the intuition that high-frequency facets (e.g., "Health effects") are low-utility, and low-frequency facets (e.g., "Curdling") are high-utility.

Second, we incorporate knowledge graph information to overcome low-utility facets. We train knowledge graph embeddings (HolE [4]), and include the top similarity scores during relevance prediction. To avoid bias in our Wikipedia-based evaluation, we propose an approach for building a knowledge graph from the CAR training data. We construct the graph by linking entity mentions to the article topic entity with the containing heading as a label. Since Wikipedia headings are used as facets for CAR, the entity embeddings can be combined with the facet relation embedding to get context-specific scores. Recall that since high-frequency headings correspond to low-utility facets, this approach should naturally provide the most value for the headings that need the most additional context.

We evaluate our approach using TREC CAR version 1.5 [1], and find that our approach performs favorably when compared to other leading approaches. We also provide an empirical evaluation of our assumptions and a detailed analysis of our results. As one of the first comprehensive works with CAR, we expect our findings and observations to shape the directions taken for CAR in the future.

References

[1] Dietz, L., Gamari, B.: TREC CAR: A data set for complex answer retrieval (version 1.5). (2017)

[2] Dietz, L., Verma, M., Radlinski, F., Craswell, N.: TREC complex answer retrieval overview. In: TREC 2017. (2017)

[3] Hui, K., Yates, A., Berberich, K., de Melo, G.: PACRR: A position-aware neural ir model for relevance matching. In: EMNLP 2017. (2017)

[4] Nickel, M., Rosasco, L., Poggio, T.A.: Holographic embeddings of knowledge graphs. In: AAAI. (2016)