

Improving entity search over linked data by modeling latent semantics

Zhiltsov N., Agichtein E.

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

Entity ranking has become increasingly important, both for retrieving structured entities and for use in general web search applications. The most common format for linked data, RDF graphs, provide extensive semantic structure via predicate links. While the semantic information is potentially valuable for effective search, the resulting adjacency matrices are often sparse, which introduces challenges for representation and ranking. In this paper, we propose a principled and scalable approach for integrating of latent semantic information into a learning-to-rank model, by combining compact representation of semantic similarity, achieved by using a modified algorithm for tensor factorization, with explicit entity information. Our experiments show that the resulting ranking model scales well to the graphs with millions of entities, and outperforms the state-of-the-art baseline on realistic Yahoo! SemSearch Challenge data sets. Copyright 2013 ACM.

<http://dx.doi.org/10.1145/2505515.2507868>

Keywords

Entity search, Learning to rank, Tensor factorization