Improving Opinion-based Entity Ranking

Christos Makris[1] , Panagiotis Panagopoulos[1]

[1]Department of Computer Engineering and Informatics, University of Patras
{makri, panagoppa}@ceid.upatras.gr

Abstract

There is a massive development of opinions and reviews on the web, which includes reviews of products and services, and opinions about events and persons. For products especially, there are thousands of users' reviews, that consumers usually consult before proceeding in a purchase. In this study we examine the problem of entity ranking using opinions expressed in users' reviews, and we are following the idea of turning the entity ranking problem into a matching preferences problem.

The entity ranking problem as information retrieval problem

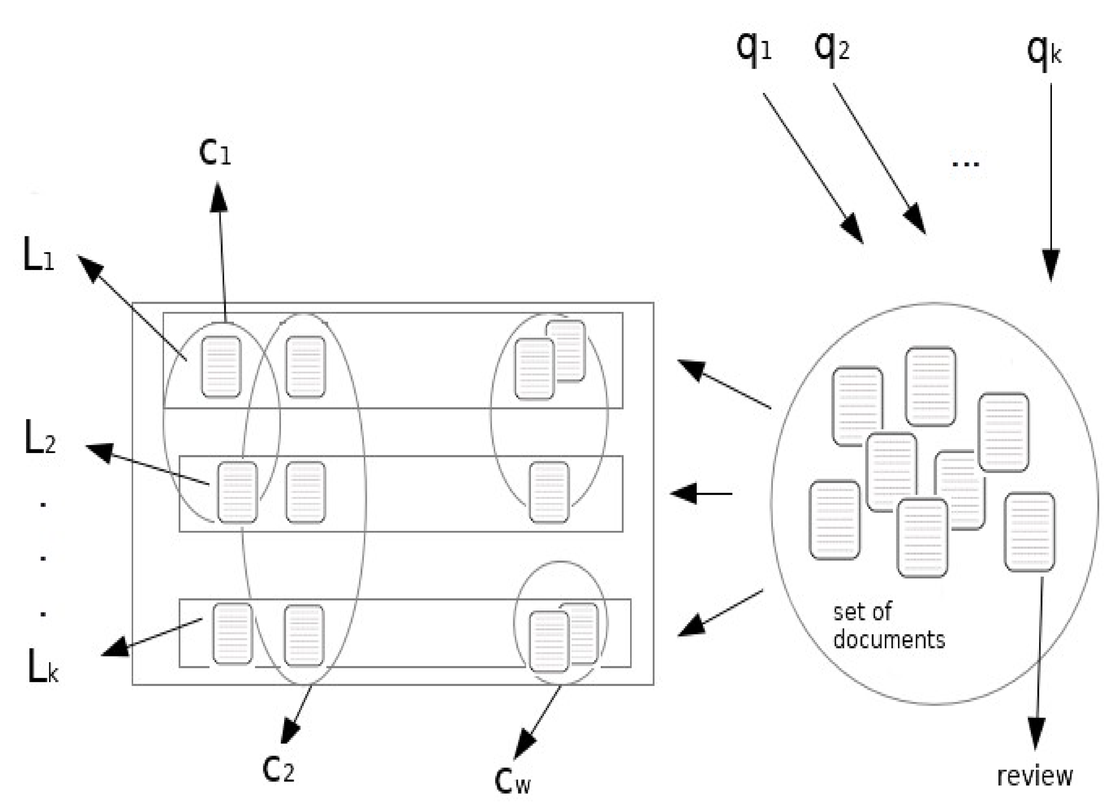
The idea to assess the entities, is to represent each entity with the text of all the reviews referred to that entity. Given a keyword query by a user, which expresses the desirable aspects that an entity should have, we can evaluate the entities based on how well the review texts match the user's preferences. So the entity ranking problem becomes an information retrieval problem. Thus we can employ some of the known information retrieval models, such as BM25.

Opinion-based Aspect Ratings

Instead of using the significance of the query's aspects for a review text to create a ranking for the reviews, we attempt to use the sentiment information that exists in the opinions. We apply two simple, well known, aspect-based sentiment analysis techniques, a lexicon-based and a syntactic patterns based, to extract the sentiment information about the aspects from the sentences, and to calculate the aspect's scores throughout the text.

Smoothing Ranking with Opinion-based Clusters

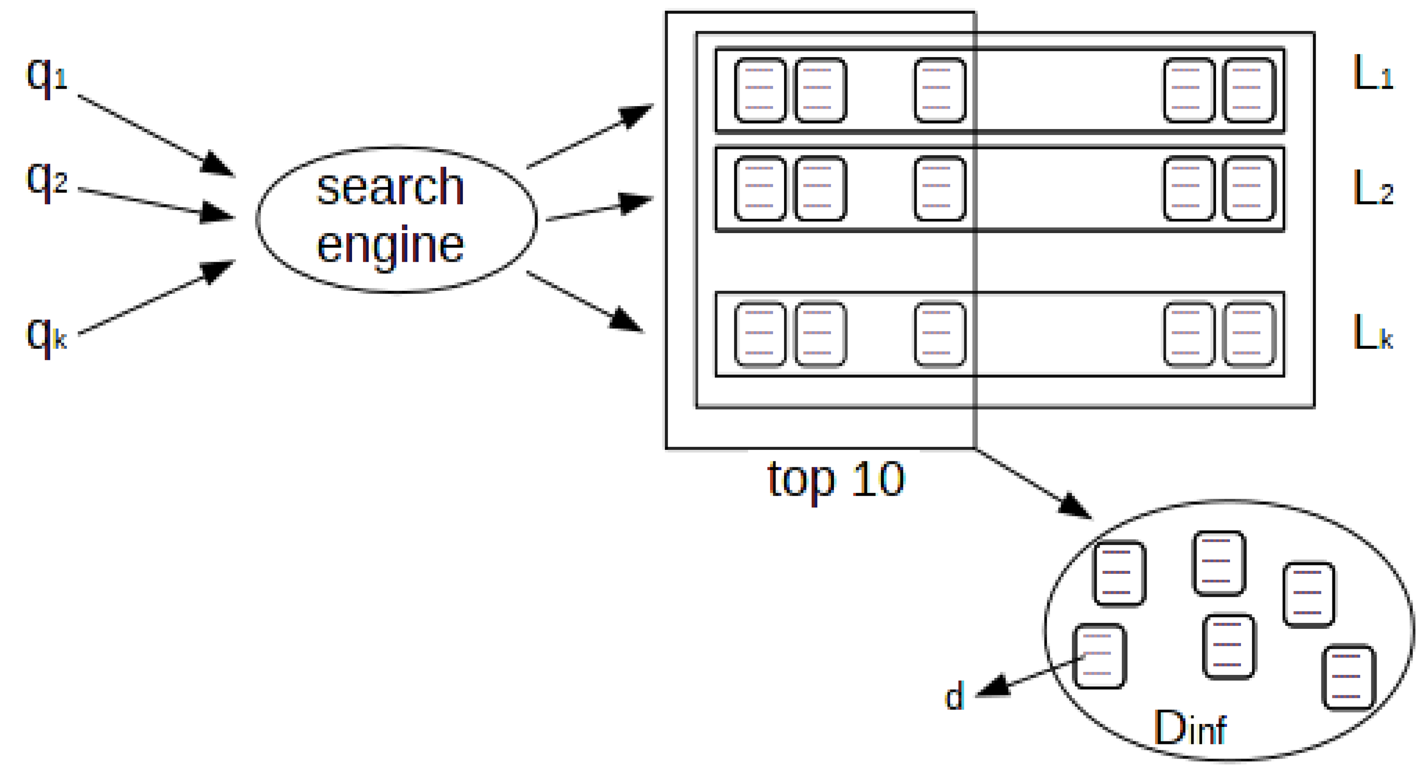
We strive to use clustering information around the reviews to improve the ranking of entities. We use the ClustFuse algorithm, which makes use of two components to provide a score to a document d, the probability's relevance of the text to the query and the assumption that clusters can be used as proxies for the texts, that rewards texts belonging "strongly" in a cluster which is very relevant to the query.



In order to construct opinion-based clusters, we perform the clustering using as features the sentiment ratings of the aspects.

The Naive Consumer Model

We attempt to simulate the behavior of a consumer who is trying to assess entities from a specific domain, but it does not know which features are important to use them as criteria in the assessment. We propose the naive consumer model as an unsupervised schema that utilizes information from the web in order to yield a weight of importance to each of the features used for evaluating the entities. We are attempting to collect the knowledge of the web, on which of the features are more important.



Concerning the significance of a term t in a document d, as part of a text collection, we can say that is calculated from the BM25 score of the term t in d. Having the set of all texts, we are trying to extract how important is each aspect for the entity domain that we examine, by calculating a score of significance.

$$scoreA(a_i)=\frac{\sum_{d\in D_{inf}} BM25(a_i)_d}{\sum_{a'\in A}\left(\sum_{d\in D_{inf}} BM25(a')_d\right)}$$

Having the rate scoreA for each aspect, expressing how important a feature is, when used to evaluate entities of a particular class, we can combine it with the term that expresses how important each aspect for a specific review text is.

$$p(q,ri)=\sum_{t\in q} score(t)_{ri} * scoreA(t)$$

Also we use this in conjunction with the ClustFuse algorithm to take advantage of any cluster of aspects, as they are defined by the users' community.

Experimental Results

method	1st exp. set	2nd exp. set
BM25std	0.87	0.936
BM25std+AvgScoreQAM+ opinExp	0.88	**0.955**
lexicon-based SA	0.865	0.91
syntactic patterns-based SA	0.869	0.94
BM25std+kurland	0.88	0.90
BM25std+kurland+opinion-based clusters	**0.89**	**0.956**
NCM1	**0.891**	0.938
NCM2	**0.893**	**0.96**
m-aspect PRank	-	**0.95**