LEARNING TO RANK SCIENTIFIC PAPERS

Tomáš KRAMÁR, Martin PETLUŠ

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
kramar@fiit.stuba.sk, martinpetlus@gmail.com

Abstract. In this paper, we present a method that automatically adapts ranking of scientific papers to match user's current research interests. We use a standard learning to rank approach with gradient descent to train a model of research interests. We show the features that comprise this model and that are automatically extracted from activity in a digital library. We have shown with an experiment with real users that our approach can quickly adapt to user's preferences and rank the search results in the digital library appropriately.

1. Introduction

Accurately ranking search results is a major problem in Web search and subject of much research. Many approaches have been proposed over the years, "learning to rank" being the most successful among them. Most of the commercial search engines rely on this approach and use it to rank billions of search results on a daily basis.

The basic idea behind learning to rank is that searchers do not click on the search results randomly, but rather evaluate each search result and consider its utility. If we ignore various biases (such as title bias, position bias, trust bias or others [4]), clicks on the search results are objectively based on consideration of various document signals. Searchers might choose to prefer documents in a language they understand, from websites they trust or written by authors they know. There are plenty of signals a searcher may consciously or unconsciously consider before making a decision to click or skip.

A simple, yet effective model of this process is a linear combination of these signals, such that

$$U_d = w_1.s_{d_1} + w_2.s_{d_2} + \ldots + w_n.s_{d_n}$$

where U_d is the utility of the document, s_{d_i} is value of the *i-th* signal of document d and w_i is the preference for the signal s_{d_i}. Thus the utility of the document can be calculated by combining all of the document signals with their preferences. In an optimal case, the ranking of the search results should be consistent with the natural ordering of this expected document utility.

The problem of ranking the documents can be broken into two separate parts:

1. determining and extracting meaningful document signals
2. learning the preferences weights for each of the signals

Naturally, not all document signals are numeric. Many signals can be boolean or nominal, but transforming these into numeric values is simple. Boolean values map naturally to 0 for false and 1 for true, and nominal value can be converted into many boolean values, e.g. a raw signal such as `website: fiit.sk` can be converted to boolean signal: `website_fiit.sk`. This conversion is necessary for each available value of the nominal signal leading to sparse data, but in practice this is not a problem.

There are many approaches for learning the preference weights for the signals. A standard learning to rank approach [3] is to train over document pairs and predict which of the two documents should be ranked higher. This model is based on the limitation of the search engine feedback. It is impossible to tell the real search result utility, but getting the relative ordering information is easily done just by observing the order in which the search results were clicked [2].

In this work, we have focused on determining the document signals in the domain of a digital library and used the standard learning to rank approach to learn the preference weights.

2. Extracting features in the digital library

A closed domain such as the domain of a digital library provides opportunities for very specialized features. We have used the following features:

1. Query dependent features
   - *query_title_cosine*: cosine similarity between the query and paper title
   - *query_abstract_cosine*: cosine similarity between the query and paper abstract
   - *query_keywords_cosine*: cosine similarity between the query and paper keywords
   - *query_top_tags_cosine*: cosine similarity between the query and paper tags
2. Query independent features
   - *authors: nominal feature, authors of the page*
   - *conference*: nominal feature, conference the paper was published at
   - *journal*: nominal feature, journal the paper was published in
   - *publisher*: nominal feature, paper publisher
   - *affiliations*: nominal feature, affiliation of paper authors
   - *num_downloads:* log normalized number of paper downloads
   - *num_citations*: log normalized number of paper citations
   - *num_pages*: log normalized number of paper pages
   - *num_bookmarks*: log normalized number of paper bookmarks

– *acceptance_rate*: log normalized conference acceptance rate
– *2011_paper*: binary feature, is paper newer than 2011?
– *2004_paper*: binary feature, is paper newer than 2004?
– *1999_paper*: binary feature, is paper newer than 1999?
– *1980_paper*: binary feature, is paper newer than 1980?

3. Evaluation

To evaluate how learning to rank can adapt search results in the domain of scientific papers, we have run an online experiment with the users of the Annota[1] system.

Annota is a system designed to manage research papers, annotate and comment them and share them with other users of the system. Annota users can very easily bookmark papers into their collection by using a browser extension. Annota then provides a simple faceted search interface that allows easy filtering and refinding of these bookmarks. In combination with the faceted search, Annota provides fulltext search over the bookmarked papers' abstracts and titles.

We have implemented the feature extraction and learning to rank approach within Annota. To test whether the personalized ranking based on the automatically learned preferences outperforms the standard non-personalized ranking provided by the underlying Elasticsearch (a Lucene-based search tool), we have deployed the personalized ranking for a small subset of Annota users.

These users were not aware that they were taking part in an experiment concerned with search and were using Annota normally. In the first part of the experiment, the personalization part was being passive and the system was just collecting preference data and ranking search results with the standard Elasticsearch ranker. In the second part of the experiment, once enough data was collected, the personalization part was turned on.

In order to minimize position bias and make a fair comparison with the non-personalized version, we have used a standard technique of search results interleaving. For each query, two sets of results were collected and ranked; one with the standard non-personalized ranker and one with the personalized ranker with the learned weight. For every query, one ranker was selected randomly and its top-ranked search result was put at the top position in the interleaved list. The top-ranked search result was taken from the other ranker and put at the second place in the interleaved list. This process continued in the ABAB scheme until all search results were used. In case there was a tie, i.e. both rankers would put the same search result at the same position, the result was not duplicated in the final interleaved list but the tie was noted. The final interleaved list of search results was then presented to the user. The clicks on the search results were automatically logged, together with the source ranker of each clicked result.

To evaluate the relative performance of the two rankers we have used a standard metric for interleaved search and simply compared click counts for each ranker. The results are summarized in Table 1.

[1] Annota, http://annota.fiit.stuba.sk

Table 1. Summary of the experimental results.

User	#1	#2	#3	#4
More clicks on personalized ranker	6	1	2	1
Less clicks on personalized ranker	0	2	1	1
Equal clicks on both rankers	0	0	0	1
No clicks	0	0	0	2

4. Conclusions

The results of the preliminary experiments show that the personalized learning to rank approach is effective in the domain of scientific papers. The closed nature of the domain allows us to extract very specific features that can be used to create better document rankings. Having an accurate model of user's research interest is not only useful for ranking search results, but can also be used in other areas, such as query construction [5] or when searching for related documents [1].

To further optimize the personalized rankings, we could extract more features from the scientific papers. We could also user query-chains to extract preference pairs, instead of using just single query and corresponding clicks.

Acknowledgement: This work was supported by the Scientific Grant Agency of Slovak Republic, grant No. VG1/0675/11 and by the Slovak Research and Development Agency under the contract No. APVV-0208-10.

References

[1] Bielikova, M., Sevcech, J., Holub, M., Moro, R.: Annota – Annotating the Documents in the Domain of Digital Libraries. In *Proc. of Datakon 2013*, Ostrava, Czech Republic, 2013, pp. 143-152 (in Slovak).

[2] Joachims T., Granka L., Pan B., Hembrooke H., and Gay G.. 2005. Accurately interpreting clickthrough data as implicit feedback. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval* (SIGIR '05). ACM, New York, NY, USA, 154-161.

[3] Joachims T.. 2002. Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* (KDD '02). ACM, New York, NY, USA, 133-142.

[4] Keane M. T., O'Brien M., and Smyth B. 2008. Are people biased in their use of search engines? *Commun. ACM* 51, 2 (February 2008), 49-52.

[5] Sevcech, J., Bielikova, M.: Query Construction for Related Document Search Based on User Annotations. In: *Proc. of the 2013 Federated Conference on Computer Science and Information Systems*. Los Alamitos: IEEE Computer Society, 2013. pp. 279-285.