

Personalized Web Search

Biman Gujral (bgujral1, bgujral1@jhu.edu), Rujuta Deshpande (rdeshpa3,rdeshpa3@jhu.edu)

10/31/2014

1 Abstract

This project aims to improve web search results through personalization. It involves re-ranking of results on the basis of user profiles comprising their past searches and actions.

The features used are based upon the short term and long term user search history. Short term history is the user's search queries and actions(dwelling time for URL and clicks on URL) in the same search session. Long term refers to a user's search behavior observed over multiple sessions over a long term period.

This problem lies in the Learning to Rank class of problems.

2 Methods

2.1 Learning and Prediction

The problem of re-ranking web search results to obtain personalized results falls under Learning to Rank which involves use of machine learning for ranking in information retrieval. This approach is achieved usually through Pointwise, Pairwise or Listwise approach. <ADD> We will start by implementing LambdaMART which is a listwise Learning to Rank algorithm as it has been ranked

2.2 Evaluation

The task involves re-ranking of search results in accordance with user's history. Therefore, the evaluation metric used is NDCG (Normalized Discounted Cumulative Gain). It is a metric between 0 and 1 that evaluates the ranking order. It is given by:

$$NDCG_k = \frac{DCG_k}{IDCG_k}$$

where k denotes documents upto rank k and DCG is Discounted Cumulative Gain, given by:

$$DCG_k = \sum_{i=1}^k \frac{2^{rel_i} - 1}{\log_2(i + 1)}$$

where rel is the actual relevance of the document provided by labels and i is the rank given by the algorithm. Thus, a highly relevant document ranked later in the list will result in penalization and a low DCG. The Ideal DCG or IDCG gives the best ranking in accordance with the relevance values. Therefore, $NDCG = 1$ when the ideal ranking is obtained. For our experiment, NDCG is calculated on the newly ranked list of URLs created by the algorithm. Thus, the learned ranking are evaluated with respect to the relevance labels obtained on the basis of clicks and dwell time on a document of a user for a particular query.

3 Resources

Our project is essentially a kaggle competition. Hence, we have obtained the data from kaggle itself. It consists of the web search logs of Yandex, a Russian search engine, collected over a period of 30 days. The first 27 days form the train data and the remaining 3 days form the test data.

The training data is a stream of logs. Each log line includes a SESSION ID, a type of record - which essentially indicates the type of log (query or click or session meta data). Based on the type of record, there is a USER ID, QUERY ID, URL ID. The field TIMEPASSED in the record, indicates the time that has passed uptill that record, from the start of that session.

3.1 Tools

We plan to use Python for our implementation and scikit-learn for any machine learning libraries we might need.

We might use a database to store our learned parameters - SQLite/MongoDB

4 Milestones

4.1 Must achieve

Our aim is to re-rank the returned URLs to obtain a high NDCG score.

4.2 Expected to achieve

4.3 Would like to achieve

Comparison of various machine learning algorithms used for re-ranking. A study of whether long term history alone can be useful for making predictions.

5 Final Writeup

The final write up will consist of the following:

- Abstract
- Related work
- Our approach
- Methodology
- Evaluation Technique
- Graphs
- References

6 Bibliography

- Milad Shokouhi, Ryen W. White, Paul Bennett, and Filip Radlinski. Fighting search engine amnesia: reranking repeated results. In Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval (SIGIR '13).
- David Sontag, Kevyn Collins-Thompson, Paul N. Bennett, Ryen W. White, Susan Dumais, and Bodo Billerbeck. Probabilistic models for personalizing web search. In Proceedings of the fifth ACM international conference on Web search and data mining (WSDM '12).
- Zhicheng Dou, Ruihua Song, and Ji-Rong Wen. A large-scale evaluation and analysis of personalized search strategies. In Proceedings of the 16th international conference on World Wide Web (WWW '07).
- Feng Qiu and Junghoo Cho. Automatic identification of user interest for personalized search. In Proceedings of the 15th international conference on World Wide Web (WWW '06).