ISSN 2395-1621

Providing Rank in Search Engine Through User Behavior

[#1]Shinde Suchitra, [#2]Bagade Priyanka, [#3]Bhor Gandhali, [#4]Prof. Bharat Burghate

[1]suchishinde92@gmail.com
[2]pyubagade@gmail.com
[3]bhorgandhali@gmail.com
[4]bharatburghate@gmail.com

[#123]Department of Computer Engineering
[#4]Prof. Department of Computer Engineering
JSPM's
Bhivarabai Sawant Institute of Technology & Research,
Wagholi, Pune – 412207

ABSTRACT

In this project, we are trying to improve existing page ranking methods on the basis of user behavior information. In this project we used new factors, which have not been used by Google, to allow relevant Web pages to be ranked higher. We are trying to capture the search history and the preferences of millions of search engine users. At the time search the searching keywords and URL of selected web pages get recorded into our search engine. Moreover, the keyword to the URL relation is also recorded. Based on these recorded data, we define three factors, which are keyword popularity, keyword to Web page popularity, and Web page popularity. Using these popularity factors, our system is able to rank more popular pages higher, which will help most search engine users find the more popular and plausibly the more relevant pages.

ARTICLE INFO

Article History

Received: 24th October 2015

Received in revised form :

24th October 2015

Accepted: 27th October 2015

Published online :

28th October 2015

Keywords: **User behavior, URL, Web pages, Web pages ranking.**

I. INTRODUCTION

Many people interact with search engine. They spend more time on searching actual pages because they click on the different advertisements and different links they will not give the actual information to the user. In this process more reformulation and action takes place and this is very costly and time consuming process.

In our project provide feedback for ranking and user behaviour has become an active area of research. The main idea behind this project is how implicit feedback can be used in a large scale operational environment to improve retrieval. While it is intuitive that user interaction with the web search engine should reveal at least some information that could be used for ranking, estimating user preferences in real web search settings is a challenging problem.

We explore ranking algorithm for ranking web search result using real user behaviour obtained as part of normal interaction with the web search engine. The specific contribution of this project includes incorporating the user behaviour information, implicit feedback, showing the significant improvements derived from incorporating user feedback.

II. LITERATURE SURVEY

"Improving Web Search Ranking by Incorporating User Behavior Information", Eugene Agichtein, Eric Brill, Susan Dumais. He shows that incorporating user behavior data can significantly improve ordering of top results in real web search setting. We examine alternatives for incorporating feedback into the ranking process and explore the contributions of user feedback compared to other common web search features. We report results of a large scale evaluation over 3,000 queries and 12 million user interactions with a popular web search engine. We show that incorporating implicit feedback can augment other features, improving the accuracy of a competitive web search ranking algorithms by as much as 31% relative to the original performance.
"A Relation-Based Page Rank Algorithm for Semantic Web Search Engines",Fabrizio Lamberti, Andrea Sanna, Claudio Demartini. With the tremendous growth of information available to end users through the Web, search engines come to play ever a more critical role. Nevertheless, because

of their general-purpose approach, it is always less uncommon that obtained result sets provide a burden of useless pages. The next-generation Web architecture, represented by the Semantic Web, provides the layered architecture possibly allowing overcoming this limitation. Several search engines have been proposed, which allow increasing information retrieval accuracy by exploiting a key content of Semantic Web resources, that is, relations. However, in order to rank results, most of the existing solutions need to work on the whole annotated knowledge base. In this paper, we propose a relation-based.page rank algorithm to be used in conjunction with Semantic Web search engines that simply relies on information that could be extracted from user queries and on annotated resources. Relevance is measured as the probability that a retrieved resource actually contains those relations whose existence was assumed by the user at the time of query definition.

"Improving web search ranking through user behavior information", Dheerendra Singh Present work proposed new methods for ordering the Web pages returned from search engines. Given a few search keywords, nowadays most search engines could retrieve more than a few thousand Web pages. The problem is how to order the retrieved Web pages and then to present the most relevant Web pages first. We propose new factors to allow relevant Web pages to be ranked higher. The factors include keyword popularity, keyword to Web page popularity, and Web page popularity. The keyword to Web page popularity records which Web pages have been selected corresponding to the search keywords. The Web page popularity determines how often the Web pages have been selected and also how many popular keywords are contained in the pages. Using these popularity factors, our system is able to rank more popular pages higher, which will help most search engine users find the more popular and plausibly the more relevant pages.

"Relevance Ranking for Predicting Web Search Results" Pedro M. Teixeira, in this paper, we focus on how incorporating user behavior data on a web search engine can significantly improve the ranking accuracy of top results in real web search engine. A large click log dataset and a set of query-url relevance ags previously labelled by juries was used to train and evaluate a classifier, as well as to build a re-ranking alternative based on a cascade click model. Each one of these implicit feedback based ranking methods can improve the precision of a web search ranking algorithms by as much as 17% compared to the its original rankings.

III. PROPOSED WORK

One of the most often overlooked factors in Search Engine Optimization is user behavior. We show that incorporating user behavior data can significantly improve ordering of top results in real web search setting. We examine alternatives for incorporating feedback into the ranking process and explore the contributions of user feedback compared to other common web search features We report results of a large scale evaluation and user interactions with a web search engine.

Goals and objectives:

Before you redesign or create a web site you need to be clear on what both you and your team want to achive. Without this understanding you can have confusion and waste time and resources creating a website that sucks.

1. Marketing - promoting what you do and products or services that you offer.
2. Sales - selling your products online.
3. Business support - providing support information for an application or business process.
4. Reducing help requests via email, phone and feedback forms.
5. Educating and informing: providing learning resources.
6. Community: building a forum or community.

Outcome

Implicit feedback for web search ranking can be exploited in a number of ways. We compare alternative methods of exploiting implicit feedback, both by re-ranking the top results (i.e., the BM25F-RerankCT and BM25F-RerankAll methods that reorder BM25F results), as well as by integrating the implicit features directly into the ranking process (i.e., the RN+ALL and BM25F+All methods which learn to rank results over the implicit feedback and other features). We compare our methods over strong baselines (BM25F and RN) over the NDCG, Precision at K, and MAP measures defined in Section 5.2. The results were averaged over three random splits of the overall dataset. Each split contained 1500 training, 500 validation, and 1000 test queries, all query sets disjoint. We first present the results over all 1000 test queries (i.e., including queries for which there are no implicit measures so we use the original web rankings). We then drill down to examine the effects on re-ranking for the attempted queries in more detail, analyzing where implicit feedback proved most beneficial.

Applications:

Multi-source information access.
Business process.
Research.
Entertainment.
Shopping.

IV. PROPOSED MODULE

Web page Ranking:

We are facing information overloaded. Finding information relevant to what we are seeking is becoming more important as the Web is growing in explosive speed. Nowadays, most people try to find whatever information on the Web by using search engines. Given a few search keywords, most search engines today will retrieve more than a few thousand

Web pages. The problem now is that we need to scan pages after pages, manually and time consumedly, to find what we need or often give up without getting the needed information. There are several approaches to address the problem. The currently most popular method to address the problem is by ordering the search results and presenting to the users the most relevant pages first. This method is called page ranking, which is one of the important factors that makes Google currently the most successful search engine. Google uses over 100 factors in their methods to rank the search results. Their methods seem to help Web users find the needed information quicker than their competitors. Even with the help of page ranking, we are facing the problem of manually performing sequential search through Web pages after Web pages. Another approach to help Web users to find the information that they need is by presenting the search results in a hierarchical structure much like a directory tree structure. Using the tree structure, the Web users can browse from one group of Web pages to another group, much like browsing the computer files on a directory tree. In this paper, we attempt to improve existing page.

Search Engine:

The second factor to be defined is the keyword to Web page popularity. After the search engine returns the search results to the user, the user will select Web pages for viewing. The relationships between the search keywords and the selected Web pages will be recorded. The relationships capture the preferences of the users. Some search engines, such as Google, currently cannot capture the relationships. Using Google, for example, when a user clicks on a link on the search results, the browser directly goes to retrieve the Web pages based on the given URL. The search engine does not know what link has been clicked. To allow the search engine to know what link clicked, each click needs to be passed through the search engine. The search keywords and the destination URL is embedded on each link provided on the search results. When a user clicks a link, the browser passes these data to the search engine. The search engine records the data and then redirects the browser to go to retrieve the destination Web page.

Web Mining:

Web mining is the Data Mining technique that automatically discovers or extracts the information from web documents. It consists of following tasks:

1. **Resource finding:** It involves the task of retrieving intended web documents. It is the process by which we extract the data either from online or offline text resources available on web.
2. **Information selection and pre-processing:** It involves the automatic selection and preprocessing of specific information from retrieved web resources. This process transforms the original retrieved data into information. The transformation could be renewal of stop words, stemming or it may be aimed for obtaining the desired representation such as finding phrases in training corpus.
3. **Generalization:** It automatically discovers general patterns at individual web sites as well as across multiple sites. Data Mining techniques and machine learning are used in generalization
4. **Analysis:** It involves the validation and interpretation of the mined patterns. It plays an important role in pattern mining. A human plays an important role in information on knowledge discovery process on web.

V. CONCLUSION

After we show that incorporating user behavior data can significantly improve ordering of top results in real web search setting. We examine alternatives for incorporating feedback into the ranking process and explore the contributions of user feedback compared to other common web search features.

REFERENCE

[1].E. Agichtein, E. Brill, and S. Dumais. Improving web search ranking by incorporating user behavior information. In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '06, pages 19{26, New York, NY, USA,.

[2]M. Cui and S. Hu, “Search Engine Optimization Research for Website promotion”, ICM Transp. Manage., China, 4, **(2011),** pp. 100-103.

[3]E. Agichtein, E. Brill, S. Dumais, and R. Ragno Learning user interaction models for predicting web search result preferences. In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '06, pages 3{10, New York, NY, USA, 2006

[4]K. Bharat and M.R. Henzinger, “Improved Algorithms for Topic Distillation in a Hyperlinked Environment,” Proc. ACM-SIGIR, 1998.

[5]Agichtein, E., Brill, E., and Dumais, S. Improving web search ranking by incorpo- rating user behavior information. In SIGIR 2006.

[6]Dheerendra Singh, 2008, **“**Improving web search ranking through user behavior information” http://www.vaughns-1-pagers.com/ internet/googleranking-factors.htm.

