Personalizing Web Search using User's Behavior

Tsoni Sofia
4747941
s.tsoni@student.tudelft.nl

ABSTRACT

Personalized search has been recognized as an important aspect of the recent web search engines. The main reason for using personalization is because of the enormously increasing amount of data available in the WWW. Thus the need of adaptation of the results according to the user's interests is crucial, for overcoming information overload. The most common approach to cope with this problem is to extract information about the user's search and use this information to re-rank the returned results. This survey focuses on the personalization based on user's profile and the corresponding method for re-ranking the results based on the information collected about the user.

1 INTRODUCTION

The last years, great effort has been given in order to develop systems that will take advantage of the user preferences and use them to adjust the ranking of results in a web search. This field of research is extremely important because of the tremendous amount of data that exist in the web. Thus the retrieval systems should adapt user's individual needs and interests, so that the most related results according to their needs will be presented first. Different researchers focus on different sources from which they can mine information about the user. Some examples of such sources are browsing history, location, social network interactions, clicks etc. In this review the user's web history will be used for the creation of the profile.

There are two different kinds of personalization based on the amount of data we have available for the users. The first approach takes into consideration the user's search context of the current session, which called short-term context based personalization. This approach is particularly helpful for new users, where the long term search history is sparse or doesn't exist at all [1] [2]. The problem is that the session data are often too sparse to personalize ideally and personalization can not be done before the second query.

There is also the long-term history based methods, where a long-term browsing log is needed or other long-term sources. The user's profile can also be constructed using information about the documents or the emails a user sees or creates, which leads to a better understanding about the topics that the user may be interested in. On the other hand the user's interactions refer to past queries, returned documents and clickthrought rates. In [3] they model users, classifying previous web visits into a topic hierarchy and then they use this model to re-rank the already returned search results.

Recently there has been research done on the effects of combining short with long term data, since each case has some advantages and some drawbacks. For example in the short term data we may have sparse data and for sure lack users' long term interests. On the other hand in long term data there is not much attention given to the current or new interests of the users. For these reasons we will consider a new framework analyzed in [10], [11] which combines the short and long information about the user.

As far as how information about the user is mined, there are two ways, namely, the implicit and the explicit. The explicit way refers to a questionnaire about the user's preferences, while the implicit is receiving information without explicitly asking the user [12]. Getting information directly from users is not the best solution, since usually the users are not wiling to spend time on clarifying their intentions. As mentioned in [4], even if they are willing, their feedback may not always be helpful. Thus this review will focus on inferring intentions implicitly rather than requiring them explicitly from the users.

Section 2 contains some background information about the field of personalized search and some terms that will be used through the whole suervey. Section 3 is about the different possibilities there exist, for constructing the user profile. Various alternatives will be investigated about the sources of the gathered information and the way the profile is structured. In Section 4 the problem of re-ranking will be presented, but also strategies on how to cope with it. The fifth section is about the evaluation used for each of the aforementioned methods. The strong and weak points of each approach will be investigated and also the metrics that were used. At the end (section 5) the extensions and the further improvements needed will be discussed. Also some conclusions that can be drawn from this survey.

2 BACKGROUND

In this section we will discuss more generally how the field of personalization has been separated for the different needs. As mentioned in section 1 the personalization has become necessary for all the search engines since a single ranking can not be optimal for all the users. The search personalization can be done in two ways or a combination of them [8]. Firstly there is the query expansion, which modifies or augments the user query. And secondly the re-ranking, on which we will focus in this survey. *Re-ranking is the procedure of issuing the same query, fetching the same results, but change the final ranking based on a user profile.*

2.1 Weighting Schemes

For the approaches investigated in the following section we will need the most well known term weighting schemes like TF, TF-IDF or BM25 [7].

2.1.1 TF weighting. This is the simplest possible term weighting, which is just the raw count of the appearances of each word. Thus for the representation of that weight measure, a vector F is needed which at each position i contains the frequency of the term t_i.

2.1.2 TF-IDF Weighting. The TF-IDF is the product of two statistics, term frequency (TF) and inverse document frequency (IDF).

The TF is defined above, as far as the IDF is concerned it is how common or rare the term is across all documents. Usually the rare terms are more informative than the common ones. $idf(t_i) = \frac{N}{log(DF_{t_i})}$, where N is the number of documents.

2.1.3 BM25 Weighting. Is a ranking algorithms, which ranks the matching documents based on their probability of relevance to the issued query [9]. To calculate the weights based on this measure, we investigate which query terms appear in each document. In practice we calculate the probability of each query term to appear in relevant or irrelevant documents and finally we sum over all query terms the logged previous quantity.

For this review a modification of the BM25 will be used, specifically for the personalization, proposed by [5]. The modified formula is the following: $w_{BM25(t_i)} = log(\frac{(r_{t_i}+0.5)(N-n_{t_i}+0.5)}{(n_{t_i}+0.5)(R-r_{t_i}+0.5)})$, where N is the number of documents in the corpus, n_{t_i} the number of documents in the corpus that contain term t_i, R number of documents in the user's history and r_{t_i} the number of these documents that contain the term t_i.

3 USER PROFILE GENERATION

Extensive research for the user profile construction using long-term interests was conducted the timeframe around 2000. The most common way is to keep track of the queries issued by the user and the pages visited from the corresponding results [13], [14], [15], [16]. From these data, terms are extracted which should be weighted in a way. The most common weighting scheme is TF, as mentioned in 2.1, but also others can be used to improve performance. The most important terms for the user, extracted by the previous procedure, constitute the "interests" of the user and will affect the future results.

In the rest of this section we will investigate in depth specific methods that were of great interest and important for the field of Information Retrieval.

One of the methods, is from Teevan et al. [5], which not only keeps track of the search history of the user, but also of his/her local activity (in computer). For the representation of the user an index was used, whose content was web pages that the user visited, emails that were viewed or sent, documents that were created or copied and calendar items.

They not only used the index as a whole to re-rank the results, but also examined how the different subsets of the index affect the re-ranking, subsets like restrict the document type to only emails or Websites, and different dimensions, for example limiting documents to the most recent ones. More emphasis was given on investigating how the date of the data influences the results, for this purpose a subset of the index with data just from the last month was used to be compared with the full index of documents. The users don't want their personal data to be stored in a server and thus mainly for security reasons, but also for ease, this method is usually implemented client-side. For the scoring of the documents the BM25 measure was used, which is analyzed in 2.1.

Teevan et al. [5], considered also two other representations for collecting user's interests. One is based on the queries issued by the user in the past, and the other is based on domains visited in the past. In these two cases, search results, that match a URL, are boosted and come higher in the ranking. These two methods can be collected on server hosting search services, since they don't contain that important personal data for most of the users.

Teevan et al. [5] using the aforementioned methods, they construct a rich representation of the user's interests, with information by the user's hard drives, emails and searches. However as a drawback we could consider that they do not take advantage of the encapsulated structure and the characteristics of the web documents, and they use them as any other document of the user. On the other hand Matthijs and Radlinski [6] focus on web documents specifically and use the underlying structure to extract information.

In more detail, their method has four steps, data capture, data extraction, term list filtering, term weighting. All this steps will be implemented in order to create three lists, a weighted list of terms, a list of previously visited URLs and the frequency of visits in each and a list with previously issued queries and pages visited for these queries. For the data collection they have developed an Firefox add-on which transmits to the server a unique identifier of each user, the URL, the duration of the visit, the length of the HTML and the time and date of the visit. From these data they used different methods to handle the information contained, like full text unigrams, title unigrams, metadata description unigrams, metadata keywords unigrams, extracted terms and noun phrases. The next step was term list filtering, to avoid the redundant terms, which unfortunately did not improve the results. The last step is the weighting of the terms, for which all the three methods mentioned in 2.1 were used.

The last method [10] creates different profiles for the current session or the past history or a combination of both. In this way they aim to investigate not only the users' "permanent" interests, but also the current activities which in some cases may be of great importance.

4 RESULT ADAPTION

After the identification of the users' interests, the next step is to adapt the return results based on them. This procedure can be done in three different ways, the result scoring, the result re-ranking and the result filtering. Re-ranking is applied to the set of documents returned by the search engine and results in the reordering of the return documents based on the data collected for the user. The case of filtering is similar to re-ranking, since the results are reordered in the same way as before and after an threshold the next results are not shown to the user. Last but not least, result scoring adds more parameters (some different features) in the main scoring function of the system.

4.1 Problem Formulation

Assume we have a session with T queries of a user u, where each query consists of some terms. We also have a list of top N documents that have been returned by the search engine in response to a query t_i and need to be re-ranked. The goal of re-ranking is to return a re-ranked list of the aforementioned N documents specifically based on the intensions of user u, using his/her previous short or long term history.

4.2 Re-ranking Methods

In [1] the predicted values for the most likely to be visited documents, computed in the previous steps, are used in order to produce the new re-ranking. It is assumed that the higher the predicted value p_i the more likely it is for the page to be clicked. However this naive approach seems to have the drawback that there are unwanted re-rankings that produce worse results than the initial. To address this problem they decided to introduce a smoothing parameter, μ, so that they will control the influence of personalization in the results. The score is calculated then, by using the Borda's ranking fusion: $score(s_i) = (1-\mu)\frac{1}{r(s_i)} + \mu\frac{1}{i}$, where s_i is the i-th result page, i is the original ranking of the page and $r(s_i)$ the reranked position. Last values of μ reduce the influence of personalization in the score, thus we have a, almost non-personalized, ranking.

Borda's fusion ranking has also been used in other cases for the combination of the initial results we the personalized ones. Another case of usage was [17], where they used it in order to combine the results return by the search engine with the ones from their system (RankSVM). However the performance was worse than just using the initial position of the documents as an extra feature in the model.

Another different approach is provided by [6]. In that research, scoring methods are used in order to assign scores to the result documents. They decided to use scoring methods, because it was shown by [5] that assigning scoring values produced the best re-ordering of the documents. They used four different methods for the calculation of the scores. Finally they consider two different approaches for the calculation of the final score. First, since they do not take into consideration the initial ranking for the score methods, they add it in the final score by multiplying the documents weight by the inverse log of the documents initial rank. The second approach is to give extra weight to the previously visited web pages, where the score is multiplied by the number of visits times another factor u.

5 EVALUATION

The evaluation process is split in two different ways, online and offline. Online tests reflect the actual usage of the system, since real users participate to use the system for their everyday needs. Offline tests use a standard dataset.

Most of the, presented in this survey, papers evaluate their systems in the offline mode because it is usually easier. Ustinovskiy and Serdyukov [1] claim that online approach has the great drawback of having to present both rankings for the user to evaluate. That means that despite the fact that some results will be inferior of the other, actions have to be considered, in order to prevent this issue

For the offline evaluation [5] and [6] use a group of people which is asked to evaluate how relevant the first k documents the results of a query q are. Since the results are evaluated the *Normalized Discounted Cumulative Gain* (NDCG) metric is used to calculate the precision. This metric is specifically for experiments where the relevance is not binary. This metric gives higher scores for the relevant documents which are ranked high in the re-ranked list and vice versa.

For the evaluation of their system, Bennett et al. [10] also used offline evaluation using *relevance judgment*, but since it is difficult to obtain a large number of *relevance judgment* by users they used a log-based approach which was firstly developed by [18] and later on was modified and used by other researchers.

For the online evaluation the most common approach is the *Clickthrough-based Evaluation*. This approach was used both in [1] and [6]. In this approach positive judgments are assigned to the results that were clicked by a user. Then some metric is computed before the re-ranking and after it. The change of the metric before and after the re-ranking shows improvement or worsening of the results.

6 CONCLUSION AND FURTHER IMPROVEMENTS

In this review we investigated different methods for search personalization based on user's behavior. The features that we concentrated to were user clicks and browsing history. The goal was to get a broad idea of what is going on in this field and how the most commonly used algorithms work. From this review we realized the importance of the correct creation of the user profile and how much the selected source of data about the user affected its creation. The more data collected for the user the more detailed the profile can become, however the issue of privacy arises. It is not wanted by the users to share their data, thus all of the methods investigated in this review try to preserve the anonymity of the users when data have to be transmitted to the server. Many researchers chose to keep implement their algorithm to work client-side. However one of the best methods which is by Teevan et al. [5] used also data from the machine of the user, like emails or documents.

In this era, when everything is personalized, we also face the issue of filter bubble. Many users get stuck in a filter bubble, where they can not discover new information since the search results are biased towards his/her constructed user profile. For this problem to be addressed, more methods that consider with greater weight the short-term behavior of the user, should be investigated. A step towards this direction is done by [10]. Nonetheless, this is an issue that many users face today and is still open.

REFERENCES

[1] Yury Ustinovskiy and Pavel Serdyukov. 2013. Personalization of web-search using short-term browsing context. In Proceedings of the 22nd ACM international conference on Information & Knowledge Management (CIKM '13). ACM, New York, NY, USA, 1979-1988. DOI=http://dx.doi.org/10.1145/2505515.2505679

[2] Ryen W. White, Paul N. Bennett, and Susan T. Dumais. 2010. Predicting short-term interests using activity-based search context. In Proceedings of the 19th ACM international conference on Information and knowledge management (CIKM '10). ACM, New York, NY, USA, 1009-1018. DOI=http://dx.doi.org/10.1145/1871437.1871565

[3] Feng Qiu and Junghoo Cho. 2006. Automatic identification of user interest for personalized search. In Proceedings of the 15th international conference on World Wide Web (WWW '06). ACM, New York, NY, USA, 727-736. DOI=http://dx.doi.org/10.1145/1135777.1135883

[4] Peter Anick. 2003. Using terminological feedback for web search refinement: a log-based study. In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '03). ACM, New York, NY, USA, 88-95. DOI=http://dx.doi.org/10.1145/860435.860453

[5] Jaime Teevan, Susan T. Dumais, and Eric Horvitz. 2005. Personalizing search via automated analysis of interests and activities. In Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '05). ACM, New York, NY, USA, 449-456. DOI=http://dx.doi.org/10.1145/1076034.1076111

[6] Nicolaas Matthijs and Filip Radlinski. 2011. Personalizing web search using long term browsing history. In Proceedings of the fourth ACM international conference

on Web search and data mining (WSDM '11). ACM, New York, NY, USA, 25-34. DOI: https://doi.org/10.1145/1935826.1935840

[7] Baeza-Yates, R., Ribeiro-Neto, B.: Modern Information Retrieval: The Concepts and Technology Behind Search, 2nd edn. Addison-Wesley, Reading (2011)

[8] James Pitkow, Hinrich SchÃijtze, Todd Cass, Rob Cooley, Don Turnbull, Andy Edmonds, Eytan Adar, and Thomas Breuel. 2002. Personalized search. Commun. ACM 45, 9 (September 2002), 50-55. DOI=http://dx.doi.org/10.1145/567498.567526

[9] Stephen Robertson & Hugo Zaragoza (2009). "The Probabilistic Relevance Framework: BM25 and Beyond". 3 (4). Found. Trends Inf. Retr.: 333-389. doi:10.1561/1500000019

[10] Paul N. Bennett, Ryen W. White, Wei Chu, Susan T. Dumais, Peter Bailey, Fedor Borisyuk, and Xiaoyuan Cui. 2012. Modeling the impact of short- and long-term behavior on search personalization. In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (SIGIR '12). ACM, New York, NY, USA, 185-194. DOI: https://doi.org/10.1145/2348283.2348312

[11] Sugiyama, K., Hatano, K., Yoshikawa, M.: Adaptive Web search based on user profile constructed with- out any effort from users. In: 13th International Conference on World Wide Web (WWW 2004), pp.675-684. ACM, New York (2004)

[12] Chirita, P., Nejdl, W., Paiu, R., and Kohlschutter, C. (2005). Using ODP metadata to personalize search. SIGIR, 178âĂŞ185.

[13] Micro Speretta and Susan Gauch. 2005. Personalized Search Based on User Search Histories. In Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI '05). IEEE Computer Society, Washington, DC, USA, 622-628. DOI=http://dx.doi.org/10.1109/WI.2005.114

[14] Pretschner, A., Gauch, S.: Ontology based personalized search. In: 11th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 1999), pp. 391-398. IEEE, Chicago (1999)

[15] Psarras, I., Jose, J.: A system for adaptive information retrieval. In: Lecture Notes in Computer Science. 4th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (AH 2006), pp. 313-317. Springer, Heidelberg (2006)

[16] Qiu, F., Cho, J.: Automatic identification of user interest for personalized search. In: 15th International Conference on World Wide Web (WWW 2006), pp. 727-736. ACM, Edinburgh (2006)

[17] Biao Xiang, Daxin Jiang, Jian Pei, Xiaohui Sun, Enhong Chen, and Hang Li. 2010. Context-aware ranking in web search. In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval (SIGIR '10). ACM, New York, NY, USA, 451-458. DOI=http://dx.doi.org/10.1145/1835449.1835525

[18] Fox, S., Kuldeep, K., Mydland, M., Dumais, S., and White, T. (2005). Evaluating implicit measures to improve Web search. ACM TOIS, 23(2): 147-168.