Search shortcuts: Driving Users Towards Their Goals

Ranieri Baraglia
ISTI - CNR
Via G. Moruzzi 1
56124 Pisa, ITALY
r.baraglia@isti.cnr.it

Fidel Cacheda
University of A Coruña
Campus de Elviña s/n
15017 A Coruña SPAIN
fidel@udc.es

Victor Carneiro
University of A Coruña
Campus de Elviña s/n
15017 A Coruña SPAIN
viccar@udc.es

Vreixo Formoso
University of A Coruña
Campus de Elviña s/n
15017 A Coruña SPAIN
vformoso@udc.es

Raffaele Perego
ISTI - CNR
Via G. Moruzzi 1
56124 Pisa, ITALY
r.perego@isti.cnr.it

Fabrizio Silvestri
ISTI - CNR
Via G. Moruzzi 1
56124 Pisa, ITALY
f.silvestri@isti.cnr.it

ABSTRACT

Giving suggestions to users of Web-based services is a common practice aimed at enhancing their navigation experience. Major Web Search Engines usually provide *Suggestions* under the form of queries that are, to some extent, related to the current query typed by the user, and the knowledge learned from the past usage of the system. In this work we introduce *Search Shortcuts* as "*Successful*" queries allowed, in the past, users to satisfy their information needs. Differently from conventional suggestion techniques, our search shortcuts allows to evaluate effectiveness by exploiting a simple train-and-test approach. We have applied several Collaborative Filtering algorithms to this problem, evaluating them on a real query log data. We generate the shortcuts from all user sessions belonging to the testing set, and measure the quality of the shortcuts suggested by considering the similarity between them and the navigational user behavior.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Query formulation, Search process

General Terms

Algorithms, Experimentation, Theory

Keywords

Search shortcut, model, evaluation

1. INTRODUCTION

The main objective of a search engine is to help the user fulfill his information need efficiently. Major search engines usually provide suggestions in the form of queries that are somehow related to the user information need, with the goal to point the user in the right direction. Authors proposed different techniques to address this problem [1, 3]. The design of effective and efficient algorithms for such suggestions is a complex and challenging task, as well as the evaluation of them. In fact, human-based evaluation has been traditionally used. Although very precise, its main inconvenience is the non repeatability of the experiments, which makes difficult an extensive comparison of such techniques.

In this work we introduce the *Search Shortcut Discovery Problem* as a problem related with the use of query suggestions in search engines, and the potential reductions obtained in the user session length. We define an evaluation methodology for this problem based on query logs that will allow a straightforward, yet interesting, comparison of the techniques that could be applied to this problem.

We have focused on the application of Collaborative Filtering algorithms [2] to this problem, a technique based on user preferences, that has been successful in several domains, such as electronic commerce. Collaborative Filtering algorithms can be divided in memory-based and model-based approaches.

In order to apply these techniques to the query shortcuts problem, we have to extract information from the query log data. As the goal in the shortcuts problem is to recommend queries for a given session, it seems reasonable to treat each Query Session as a *user* and each Query as an *item*. Then, the query ratings must be inferred from the information in the query log. As a preliminary approach, in this work we have given to the last query of a session a possitive rating if the user has clicked on, at least, one result, or a negative rating otherwise. All remaining queries are considered neutral.

2. SEARCH SHORTCUTS

The idea of Search Shortcuts can be easily explained with a simple example. Suppose a (sufficiently) high number of users has queried the engine for q_1, q_2, q_3, and finally, after asking for q_4, they clicked on a result presented in the Search Engine Result Page (SERP). We can assume that q_4 is a good query, since it was followed by a click. Therefore, we can also consider q_4 as relevant for users interested in topics related to q_1, q_2, and q_3. Whenever another user starts to search for topics related to q_1, q_2, or q_3, q_4 will be proposed as a *shortcut*. Obviously, the earlier a relevant shortcut is shown during the user session, the more effective it has to be considered. Following this idea, a given algorithm can be easily evaluated using the following function:

$$s\left(h\left(\sigma_{t|}\right), \sigma_{|t}\right) = \frac{\sum_{q\in h\left(\sigma_{t|}\right)} \sum_{m=1}^{n-t} \left[q = \left(\sigma_{|t}\right)_m\right] f\left(m\right)}{|h(\sigma_{t|})|} \qquad (1)$$

Where:

- $\sigma_u =< q_1 \dots q_n >$ is a query session for a given user.
- $\sigma_{t|}$ is the **head** of σ **up to** $t \leq n$ is the sequence of the first t queries in σ, i.e. $\sigma_{t|} =< q_1, \dots, q_t >$
- $\sigma_{|t}$ is the **tail** of σ **from** $t \leq n$ is the sequence of the last $n - t$ queries in σ, i.e. $\sigma_{|t} =< q_{t+1}, \dots, q_n >$
- $h : \mathcal{S} \rightarrow 2^{\mathcal{Q}}$ is the k**-way shortcut** function, taking as argument a session and returning a set of queries of cardinality less then k, i.e. $|h\left(\sigma\right)| \leq k$.
- $f\left(m\right)$ is a monotonic increasing function.
- The function $[q = \sigma_m] = 1$ if and only if the query q is equal to the query σ_m.

3. EXPERIMENTS

As data set for our experiments we have used a subset of the AOL query log, consisting of 3,558,412 records. To perform the evaluation, we have divided our query log data in two subsets: training and evaluation, by randomly chosing a percentage of the sessions as the data to be used to train the algorithms. Then, we fed the algorithms with the first two queries from each session in the evaluation set. The evaluation is performed using both traditional metrics and the similarity measure proposed in section 2.

The results, as can be seen in Fig. 1, clearly show the differences between Memory- and Model-Based approaches when applied to this problem. Memory-Based algorithms showed pretty accurate results, but they cover a low fraction of the data, thus resulting useless for the search shortcut problem.

Results on other algorithms are a bit surprising, but we can easily relate such results with both the limitations of traditional metrics, and the way we are implicitly assigning ratings to queries. Regarding this last issue, the simple three-rating schema we have proposed leads to a rating matrix where most queries have a neutral rating, thus biasing the evaluation, specially on the MAE metric. On the other hand, limitations on evaluation methodology are clearly observed on classification and rank accuracy metrics, where all algorithms have obtained modest results. This is a problem related with both the sparsity and high volume of the dataset, and the way these metrics are measured in an offline dataset.

Anyway, by taking into account these limitations, we can note the pretty good results of the Item-Mean, Trends-Based and Personality Diagnosis. This shows that both techniques can obtain enough information from sparse data, so further experiments on these algorithms seem valuable.

4. CONCLUSIONS

In this work we have introduced the *Search Shortcuts* problem, directly related with recommender systems and query suggestion, and we have presented a well-defined model and evaluation framework, that allows the comparison and

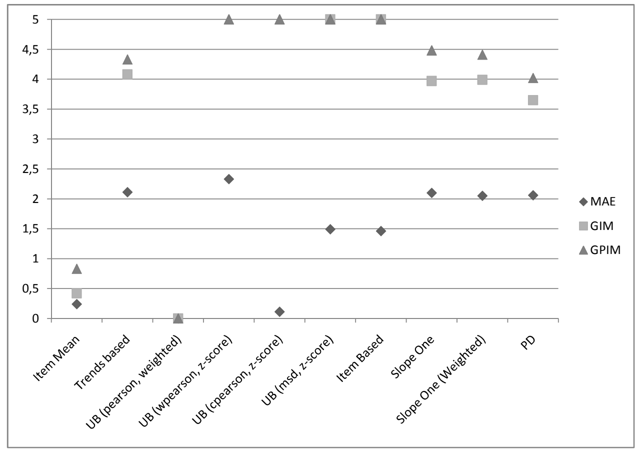

Figure 1: Results in the predictive accuracy metrics for several Collaborative Filtering algorithms.

evaluation of algorithms using an offline dataset. Specifically, we have studied the application of Collaborative Filtering techniques to the search shortcuts problem, evaluating several algorithms on a real query log data.

Our experiments have shown the limitations of traditional algorithms, as the high volume and sparsity of the query log data lead to poor results in most cases. Specifically, traditional memory-based approaches present very low coverage, due to the fact that they are only based on a small part of the information available. Traditional metrics and evaluation methodologies have also shown some important limitations. Classification accuracy metrics obtain valuable results, but only if the details of the evaluation methodology are taken into account. In particular, the usage of offline and sparse datasets impose several limitations in the evaluation, as items considered relevant by the algorithm cannot be compared in many cases with real data.

Several limitations have also been observed in the way used to extract information from the query log data. The three level rating (positive, negative, neutral) does not perform as expected, especially because most queries are in fact neutral. The way we map queries to items can also be improved, by considering the terms that compose a query. Techniques such as stemming or stopwords removal, can effectively reduce the data sparsity, and thus reduce the main reason of the poor performance of most algorithms.

5. REFERENCES

[1] R. A. Baeza-Yates, C. A. Hurtado, and M. Mendoza. Improving search engines by query clustering. *JASIST*, 58(12):1793–1804, 2007.

[2] U. Shardanand and P. Maes. Social information filtering: algorithms for automating "word of mouth". In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 210–217, New York, NY, USA, 1995. ACM Press/Addison-Wesley Publishing Co.

[3] Z. Zhang and O. Nasraoui. Mining search engine query logs for query recommendation. In *WWW '06: Proceedings of the 15th international conference on World Wide Web*, pages 1039–1040, New York, NY, USA, 2006. ACM.