Aggregating User-Centered Rankings to Improve Web Search

Lin Li and **Zhenglu Yang** and **Masaru Kitsuregawa**
Institute of Industrial Science, The University of Tokyo,
4-6-1 Komaba, Meguro-Ku, Tokyo 153-8305, Japan
{lilin, yangzl, kitsure}@tkl.iis.u-tokyo.ac.jp

Abstract

This paper is to investigate rank aggregation based on multiple user-centered measures in the context of the web search. We introduce a set of techniques to combine ranking lists in order of user interests termed as a user profile. Moreover, based on the click-history data, a kind of taxonomic hierarchy automatically models the user profile which can include a variety of attributes of user interests. We mainly focus on the topics a user is interested in and the degrees of user interests in these topics. The primary goal of our work is to form a broadly acceptable ranking list, rather than that determined by an individual ranking measure. Experiment results on a real click-history data set show the effectiveness of our aggregation techniques to improve the web search.

Introduction

Nowadays in the context of the web search, it becomes more difficult to obtain desired results than ever due to the ambiguity of user's needs. Chirita et al. (2005) re-ranked search results based on the similarities between search results and the user profiles (the descriptions of user interests). In this way, users with different interests may obtain different ranking lists even with the same query, thus the accuracy of the web search can be improved.

A user profile, however, may contain a number of attributes which describe user interests from their respective viewpoints. In most cases, any individual attribute is inadequate in defining user interests accurately. We regard the user profile as a consensus-based combination which brings the following problems: (1) how to extract various attributes of user interests from information sources (i.e., click-history data here), and represent them in a proper way is not a trivial job, for they are usually heterogeneous objects; (2) in order to leverage the different ranking lists produced by the different attributes, the rank aggregation should intend to form a single ranking list supported by a broad consensus among these attributes. Chirita et al. (2005) just merged the values of the attributes in a linear combination, neglecting the respective characteristics of them. Furthermore, it is important to observe that if the ranking measure is value-based, the ordering implied by the values makes more meaning than the actual values themselves (Dwork *et al.* 2001).

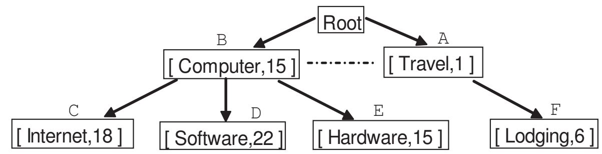

Figure 1: Scheme of user profile

In this paper, we propose techniques to overcome the above difficulties. (1) A hierarchical model for the user profile is designed and contains two attributes of user interests, the topics a user is interested in and the degrees of user interests in these topics. Chirita et al. (2005) introduced a hierarchy-based similarity measure for the web search , but they requires users to select topics which best fit their interests explicitly. Our user profile is modelled automatically without any efforts from users. (2) The attribute-based ranking lists plus the original ranking list of a search engine (i.e., Google here) are aggregated by a set of methods. Dwork et al. (2001) developed the theoretical ground work for describing and evaluating rank aggregation methods. Their main work is to effectively combat "spam". We study the rank aggregation of the attributes of user interests learned from the click-history data to improve the web search.

Aggregation of User-Centered Ranking Lists

Hierarchical User Profile

Our user profile is composed of topics which come from the top four levels of the Google Directory[1](the preliminary analysis to select the number of levels is not described here due to the page limit[2]). Each topic (node) has a value of the number of times the topic has been visited. This value is denoted by the "$TopicCount$" that represents the degree of user interests. Figure 1 illustrates the scheme of a user profile. For example, node C is represented by the $[Internet, 18]$ which means one user has clicked a search result associated with the topic "$Internet$" and the user has visited the "$Internet$" 18 times before this search. When learning the user profile, topics associated with the clicked search results by users, are added into the user profile click by click. The value of "$TopicCount$" also increases.

[1]http://directory.google.com
[2]Refer to http://www.tkl.iis.u-tokyo.ac.jp/ lilin/dews07.pdf

User-Centered Ranking Lists

Hierarchy-Based Semantic Similarity Li et al. (2003) define the similarity as

$$S(i,j) = e^{-\alpha \cdot l} \cdot \frac{e^{\beta \cdot h} - e^{-\beta \cdot h}}{e^{\beta \cdot h} + e^{-\beta \cdot h}}\,,\ \ \alpha \geq 0\,, \beta > 0\,, \qquad (1)$$

Their experiment results show that the optimized values of the two parameters are, α=0.2 and β=0.6. h means the depth of the subsumer (the deepest node common to two nodes), and l is the naïve distance (the number of edges between two nodes). One user profile includes a number of nodes. We further define the semantic similarity between one search result denoted by i and one user profile denoted by j as the maximum value among all the values computed by Equation (1). The re-ranked search results by our semantic similarity form a ranking list in order of one attribute of user interests (i.e., the topics a user is interested in).

Degree of User Interests $TopicCount$ weighs the degree of the user interests in a node of the user profile. Thus, the values of $TopicCount$ can also order the search results and produce a ranking list. To keep our rank aggregation from missing the high quality web pages in Google, we also take into consideration the original rankling list of Google.

Methods for Rank Aggregation

Borda's Rule The Borda's rule (Young 1974) is a single winner election method in which votes rank candidates in order of preference. The Borda's rule determines the winner of an election by giving each candidate a certain number of points corresponding to the position in which she is ranked by each voter.

Let $A = a_1, a_2, \cdots, a_m$ be the set of positions in the ranking list, and let the attributes of user interests plus PageRank be named by elements of n. We shall assume for the present that every element of n can be expressed by a linear order in the position set A. We denote a linear order by a sequence $A_i = a_{i_1}, a_{i_2}, \cdots, a_{i_m}$ where for $j < k$, a_{i_j} is preferred to a_{i_k}. We apply a sort of modified Borda's rule here. The voter awards the first-ranked candidate with one point. The second-ranked candidate receives half of the point, the third-ranked candidate receives on third of the point, etc.. When all elements of n have been counted, and each A_i can be thought of a position vector, we sort the search results by the L_1 norm and the L_2 norm of these vectors, the median of the n points, and the geometric mean of the n points.

Spearman's Footrule According to Diaconis et al. (1977), the two measures which we consider are:

$$D(\pi, \sigma) = \sum_{i=1}^{m} \mid \pi(i) - \sigma(i) \mid\,,\ S(\pi, \sigma) = \sum_{i=1}^{m} (\pi(i) - \sigma(i))^2\,. \qquad (2)$$

π and σ are regarded as rankling lists here. Diaconis et al. (1977) also suggest other two measures. One roughly seems similar to D, and the other is unsuitable for general use, having very small variance about a mean very close to its maximum value. Therefore, we choose D and S here.

Inspired by (Dwork *et al.* 2001), we define a weighted balanced bipartite graph $G = (V_1 \cup V_2, W)$. $V_1 = r_1, r_2, \cdots, r_m$ is a set of search results to be ranked. $V_2 = p_1, p_2, \cdots, p_m$ is the m available position in the ranking list. For any two vertices $r \in V_1$ and $p \in V_2$, rp is an edge in G. G is also a complete bipartite graph. The weight $W(r,p)$ is the total distance of a ranking value that places r at position p, given by $\sum_i^n \mid A_i(r) - p \mid$ or $\sum_i^n (A_i(r) - p)^2$ ($A_i(r)$ is the position of r in the ranking list i). Minimizing the total distances to n could be solved by the well-known Hungarian algorithm that finds a minimum cost perfect matching in the bipartite graph.

Table 1: Quality of Rank Aggegration Measured by DCG

Methods	DCG	Relative improvment
L_1 norm	1.93948	7.42%
L_2 norm	1.95184	8.10%
Median	1.97047	9.13%
Geometric mean	1.98762	10.1%
Bipartite_D	1.88534	4.42 %
Bipartite_S	2.07534	14.9%

Experiments and Conclusions

Our experiments were offered the top 20 search results under a query by the Google API (http://code.google.com/apis/). 12 subjects are invited to search the web through our system. In the first seven days, they were asked to query topics closely related to their interests and specialized knowledge for learning their user profiles, and to repeat some queries done before for testing in the last three days. Measured by the DCG (Jarvelin & Kekalainen 2000), we compared the qualities of our techniques and a simply linear combination (SLC) of measures (Chirita *et al.* 2005). The results of the average improvements over all subjects are illustrated in Table 1. The DCG of SLC is 1.80557. Our rank aggregations yield better search results compared with SLC (the largest improvement of ours is 14.9%). We can see that the approaches originated from social choice theory and graph theory produce a broadly acceptable ranking list in terms of various attributes of user interests, thus improve the quality of the web search.

References

Chirita, P. A.; Nejdl, W.; Paiu, R.; and Kohlschütter, C. 2005. Using ODP metadata to personalize search. In *SIGIR*, 178–185.

Diaconis, P., and Graham, R. L. 1977. Spearman's footrule as a measure of disarray. *Journal of the Royal Statistical Society. Series B (Methodological)* 39(2):262–268.

Dwork, C.; Kumar, R.; Naor, M.; and Sivakumar, D. 2001. Rank aggregation methods for the web. In *WWW*, 613–622.

Jarvelin, K., and Kekalainen, J. 2000. IR evaluation methods for retrieving highly relevant documents. In *SIGIR*, 41–48.

Li, Y.; Bandar, Z.; and McLean, D. 2003. An approach for measuring semantic similarity between words using multiple information sources. *IEEE Trans. Knowl. Data Eng.* 15(4):871–882.

Young, H. P. 1974. An axiomatization of borda's rule. *Journal of Economic Theory* 9(1):43–52.