An Effective Method for Personalized Image Search Using Query Expansion

C. Anuradha

Department of Computer Science and Engineering,
Bharath University, Chennai-73, India

Abstract: Personalized Web search takes advantage of information about an individual to identify the most relevant results for that person. A challenge for personalization lies in collecting user profiles that are rich enough to do this successfully. One way an individual’s profile can be augmented is by using data from other people. To better understand whether groups of people can be used to benefit personalized search, we explore the similarity of query selection, desktop information and explicit relevance judgments across people grouped in different ways. The groupings we explore fall along two dimensions: the longevity of the group members’ relationship and how explicitly the group is formed. We find that some groupings provide valuable insight into what members consider relevant to queries related to the group focus, but that it can be difficult to identify valuable groups implicitly. Building on these findings, we explore an algorithm to "groupize" (versus "personalize") Web search results that leads to a significant improvement in result ranking on group-relevant queries.

Key words: Personalized image search · Topic model · Social annotation

INTRODUCTION

Web search personalization algorithms improve the Web search experience by using an individual’s data (e.g., topical categories marked interesting, query history, or term vectors of previously viewed content) to identify the results that are the most relevant to that individual. This can be done in several ways. For example, a searcher’s query can be modified to reflect a particular geographic location, occupation, work group, query interest, or results may be re-ranked so that personally relevant results appear higher in the list [1]. Previous research suggests personalization algorithms perform best when there is a large amount of data available about an individual [2]. For this reason, we propose combining an individual’s data with that of other related people to enhance the performance of personalized search. We call the use of group information for personalization “groupization.”

One challenge in the use of group data for personalization lies in the identification of related groups of people. To develop an understanding of what factors are important for building groups for groupization, we conducted two studies of a total of 140 people. The data we collected enabled us to understand whether people grouped by various properties were similar in the queries they selected, the information they had on their desktop, or the relevance judgments they assigned to search results. We explored groupings that varied based on the longevity of the relationship and on whether the group was formed explicitly or implicitly. Specific grouping criteria included task, interests, demographics, geographic location, occupation, work group, query selection and the content on their desktop computers. By correlating group membership with the similarities of the group members’ explicit relevance judgments, we are able to understand what types of groups are most likely to receive value from groupization.

It appears that some attributes are more useful than others for identifying people who find the same results relevant and, in particular, that group membership provides information about what members consider relevant to group-related queries. Using the data we collected to understand group properties, we explore combining information about group members to produce a groupized (versus personalized) result list. We find that it is possible to aggregate personalization
scores from different group members to create a groupized result list that is of higher quality than each individual’s personalized list. Consistent with the understanding we develop of the different attributes, groupization appears most useful for queries related to the group [3]. We begin the paper with a discussion of related work in the areas of personalization, collaborative filtering and collaborative Web search. We then describe our data collection methodology. By analyzing the collected data, we explore the within-group variation of relevance judgments, query selection and user profile information. We then describe a groupization algorithm that extends personalization techniques to include group data [4]. We analyze the value of groupization for the groups represented in our study and conclude with a discussion of practical issues, including techniques for identifying groups and group-related queries outside of experimental settings [5].

**Previous Research:** In the future, we will improve our current work along four directions. 1) In this paper, we only consider the simple case of one word-based queries. Actually, the construction of topic space provides a possible solution to handle the complex multiple words-based queries. We will leave it for our future work. 2) During the user-specific topic modeling process, the obtained user-specific topics represent the user’s distribution on the topic space and can be considered as user’s interest profile. Therefore, this framework can be extended to any applications based on interest profile.

Qiu, F. and J. Cho [1] proposed, automatic identification of user interest for personalized search as more and more topics are being discussed on the web and our vocabulary remains relatively stable, it is increasingly difficult to let the search engine know what we want [6]. Coping with ambiguous queries has long been an important part in the research of Information Retrieval, but still remains to be a challenging task. Personalized search has recently got significant attention to address this challenge in the web search community based on the user preference may help the search engine disambiguate the true intention of a query [7]. We represent how a search engine can learn a user’s preference automatically based on her past click history and it can use the user preference to personalize search result. Personalized search based on user preference significant improvements over the best existing ranking mechanism.

Carmel, D. and N. Zwerdling proposed personalized social search based on the user’s social network, personalized social search based on the user’s social relations, search result are re-ranked according to their relations with individuals in the user’s social network [8]. Personalized the search process by considering the searcher’s personal attributes and preference while evaluating a query, is a great challenge that has been extensively studied in information retrieval [9]. User queries are in generally short and provide an incomplete specification of individual users information need. Search personalization requires the capability of modeling the user’s preference and interest. This is usually done by tracking and aggregating user’s interaction with the system. User’s interaction are structured into a user profile that can utilized during search. A user profile is usually employed in two main methods, first personalized query expansion i.e., adding new term to query and re-weighting original query terms based on the user profile or through re-ranking and filtering the search results while incorporating user’s interest.

Chirita, P.A. et al. [3] proposed personalized query expansion for the web. The inherent ambiguity of short keyword queries demands for enhanced methods for web retrieval [10]. To improve such web queries by expanding them with terms collected from each user’s Personal Information Repository, thus implicitly personalizing the search output. Query Expansion assists the user in formulating a better query, by appending additional keywords to the initial search request in order to encapsulate her interest, as well as to focus the web search output accordingly. Several advantages arises when moving web search personalization down to the desktop level [11]. The algorithm expand web queries with keyword extracted from user’s Personal Information.

Teevan, J. et al. proposed discovering and using groups to improve personalized search. Personalized web search takes advantages of information about an individual to identify the most relevant results for that person [12]. Web search personalization algorithm improve the web search experience by using an individual’s data to identify the results that are the most relevant to that individual. One challenge in the use of group data for personalization lies the identification of related groups of people. The data collected enabled to understand whether people grouped by various properties were similar in the queries they selected. Specific grouping criteria included task, interest, demographics, geographic location, occupation, work group, query selection and the content on their desktop computer [13]. It appears that some attributes are most useful than others for identifying people who find out the same relevant result. The methodology used is data collection. By analyzing collected data, we explore the
within-group variation of relevant judgements, query selection and user profile information.

**Repository (PIR)**

**Hypotheses**

H1: User Specific Topic
H2: Personalized Image Search
H3: Ranking Based Multi-correlation Tensor Factorization

**Research method:** The personalized search model firstly constructs users’ interest profiles dimensions are extracted to obtain users’ interest distributions. Finally, it rearranges the returned photos for the searcher by predicting his preference on these photos.

User Interest Profile Construction: Uploading photos with attached tags and joining in interested groups are two main manners for users to share personal photos, which reflect their implicit interests. Both tag and group are thus used to construct a user’s interest profile. We refer to $U = \{u_1, u_2, \ldots, u_M\}$ as the set of users including the searcher and his friends and $T$ as the bag-of-tags utilized by the users and $G$ as the bag-of-groups the users join in. As illustrated in Fig. 1 (a), the relations of user-tag and user-group construct two correlated bipartite graphs, denoted as $G(U, T, G, W_{UT}, W_{UG})$, where $W_{UT}$ denotes the edge weights between $U$ and $T$ and $W_{UG}$ denotes the edge weights between $U$ and $G$. The simple yet widely used tf-idf is employed to set the edge weights. Therefore, a user’s interest profile is simultaneously represented by a weighted tag-vector and a weighted group-vector.

Latent Interest Dimension Extraction: With users’ interest profiles represented as two correlated bipartite graphs, a soft clustering method is performed to extract the latent interest dimensions. We utilize the clustering method, which aims to co-cluster high-order correlated bipartite graphs based on spectral graph partitioning (as illustrated in Fig. 1 (b)). The basic idea of spectral graph partitioning focuses on minimizing a cost function. In this study, we adopt Ncut which simultaneously minimizes the between-cluster similarities and maximizes the within-cluster similarities. By introducing a parameter to balance the costs on both graphs, the co-clustering task can be turned into the optimization problem of minimizing the following generalized cost function:

**Summary and Concluding Remarks:** We propose a personalized search model to assist users in getting access to their interested photos by predicting the searcher’s preference on returned photos. Preliminary experiments have proved the validity of the proposed model. In the future, we would like to further improve the efficiency of the proposed mechanism. The high quality social data that is available today for individuals in the enterprise, allows the identification of social relations that can be utilized for search personalization and for other applications. The question whether social data out of the firewall, typically with lower quality, can be used effectively for search personalization remains open for further research.

**REFERENCES**

