Future Generation Computer Systems 78 (2018) 440–450

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Query personalization using social network information and collaborative filtering techniques

Dionisis Margaris [a], Costas Vassilakis [b,*], Panagiotis Georgiadis [a]

[a] Department of Informatics and Telecommunications, University of Athens, Greece

[b] Department of Informatics and Telecommunications, University of the Peloponnese, Akadimaikou G. K. Vlachou, 22100, Greece

HIGHLIGHTS

- A social network-based query personalization algorithm is presented.
- The algorithm considers influencers from the user's social network.
- Both the user's and the influencers' preferences are used for personalizing queries.
- The algorithm is evaluated both performance- and quality-wise.

ARTICLE INFO

Article history:
Received 20 June 2016
Received in revised form
26 December 2016
Accepted 8 March 2017
Available online 10 March 2017

Keywords:
Social networks
Personalization
Collaborative search
Database query transformation
Presentation of retrieval results

ABSTRACT

Query personalization has emerged as a means to handle the issue of information volume growth, aiming to tailor query answer results to match the goals and interests of each user. Query personalization dynamically enhances queries, based on information regarding user preferences or other contextual information; typically enhancements relate to incorporation of conditions that filter out results that are deemed of low value to the user and/or ordering results so that data of high value are presented first. In the domain of personalization, social network information can prove valuable; users' social networks profiles, including their interests, influence from social friends, etc. can be exploited to personalize queries. In this paper, we present a query personalization algorithm, which employs collaborative filtering techniques and takes into account influence factors between social network users, leading to personalized results that are better-targeted to the user.

1. Introduction

The enormous growth of available content in current applications, and web applications in particular, has created information overload for the application users, necessitating the use of personalization techniques to alleviate this problem [1], by prioritizing or limiting information presented to the users according to its perceived value for them. With the advent of social networks, such as Facebook [2] and Twitter [3], used by millions of people every day, large volumes of data generated by these networks are widely available, and researchers seek methods to exploit these data for personalization purposes [4–8]; these data are deemed of high value in the context of personalization, because of the importance and the intrinsic relationship with people's everyday lives. However, due to the high volume of social network-generated data, identifying the data relevant to each individual user that are highly useful to support personalized queries, still remains a challenge.

A widely-used approach for making recommendations, stemming from user behavior and actions is collaborative filtering. Collaborative filtering (CF) synthesizes the informed opinions of humans (notably these opinions in many cases encompass the aspect of satisfaction), to make personalized and accurate predictions and recommendations. The biggest advantage of collaborative filtering is that explicit content description is not required (as in content-based systems): instead, traditional collaborative filtering relies only on opinions expressed by users on items either explicitly (e.g. a user enters a rating for the item) or implicitly (e.g. a user purchases an item, or clicks an advertisement banner, which indicates a positive assessment). In the context of collaborating filtering, personalization is achieved by considering ratings of "similar users", under the collaborative filtering's fundamental assumption that if users X and Y have similar behaviors (e.g., buying, watching, listening, rating assignment) on some items, they will act similarly

* Corresponding author.
E-mail address: costas@uop.gr (C. Vassilakis).

http://dx.doi.org/10.1016/j.future.2017.03.015
0167-739X/

on other items [9]. Traditional recommender systems though, assume that users are independent and ignore the social interactions among them, hence they fail to incorporate important aspects that denote interaction, tie strength and influence among users, which can substantially enhance recommendation quality [10,7]. Social network data-based recommender systems take into account static data from the user profile, such as location, age or gender, complemented with dynamic aspects stemming from the user behavior and/or social network state such as user preferences, item's general acceptance, and influence from social friends [10,7]. Towards this direction, the metric of *tie strength* between users of social networks has been modeled, quantifying the projected influence of one user to another. This metric can be exploited to further enhance the choice of recommenders, so as to consider the opinions and choices of users that have a high influence on the user for which the recommendation is generated [4,11,12].

Query personalization is an important area in personalization, in which queries are dynamically enhanced with related preferences stored in a user profile with the purpose of providing more focused answers [13]. Query personalization systems may be used either by end-users who directly pose the queries that are subsequently personalized, or as an underlying infrastructure in any database-supported application, in order to tailor the results returned by queries to the application user's profile. To this end, query personalization systems have been studied in a number of works [13–17]. These works consider user personal and/or collaborative preferences that are stored in a preference repository, they do exploit however opportunities offered by social media data to enhance query adaptation quality.

In this paper, we propose a novel algorithm for query personalization, based on social network information. The algorithm exploits both the users' choices of items (browsing and rating) and the influence information from social networks to suitably adapt queries for each user. Query adaptation is performed via (re)writing the query sorting specification to order the qualifying data according to their projected interest for the user. The proposed algorithm is evaluated both in terms of personalization accuracy and execution performance.

The proposed algorithm extends the state of the art query personalization algorithms by (a) exploiting social graphs for improving personalization quality and (b) arranging for mapping the outcome of the personalization procedure proposing an efficient query rewriting technique.

The rest of the paper is structured as follows: Section 2 presents the proposed algorithm. An evaluation of the algorithm, both in terms of personalization quality and performance is given in Section 3. Section 4 overviews related work, while finally, Section 5 concludes the paper and outlines future work.

2. The query personalization algorithm

In this section, we first describe the concept of influence in social networks, which is central to the proposed algorithm. Subsequently we elaborate on the representation of the information model used by the algorithm: this model includes user preferences, the influence information from social networks and a sample database schema for exemplifying the functionality of the personalization. Finally, we detail the operation of the algorithm and its phases.

2.1. Influence in social networks

Within a social network, "social friends" greatly vary regarding the nature of the relationship holding among them: they may be friends or strangers, with little or nothing in between [18]. Users have friends they consider very close, and know each other in real life, acquaintances they rarely meet or communicate in real life, and persons they have never actually met with, such as singers, actors and athletes. According to Anagnostopoulos et al. [19], three main causes of correlation in social networks exist: *influence* (also known as induction), where the action of a user is triggered by one of her friend's recent actions (e.g. when a user buys a product because one of her friends has recently bought the same product); *homophily*, which means that individuals often establish "social friendship" with others who are similar to them, and hence perform similar actions (e.g. sharing a common interest, such as mountaineering); and *environment* (also referenced as *external influence*), where external factors are correlated both with the event that two individuals become friends and also with their actions (e.g. two inhabitants of the same city posting pictures of the same landmarks in an online photo sharing system can become "social friends").

Bakshy et al. [5] suggest that a social network user responds significantly better to recommendations that originate from friends of the social network to which the user has high *tie strength*; the strength of the directed tie between users i and j is computed as:

$$W_{i,j} = \frac{C_{i,j}}{C_i}$$

where C_i is the total number of communications posted in a certain time period in the social network by user i, whereas $C_{i,j}$ is the total number of communications posted by user i on the social network during the same period and are directed at user j or on posts by user j. In [5], a period of 90 days is considered for computing the tie strength. Note that a post may be directed towards multiple users (individually specified recipients or all members of a group, such as "close friends"), in which case it contributes to the tie strength of all its recipients. Because of this fact, it holds that

$$\sum_j W_{i,j} \geq 1.$$

2.2. The information repository

In order to be able to personalize a user's queries, the information repository must include (a) data concerning the items that users want to retrieve through their queries (b) data regarding the influencers for the particular user, which are computed from the social network graph and (c) the preferences on items for both the user and her influencers. Regarding the items that users want to retrieve through their queries, in this work we consider a typical database schema used in query personalization research (e.g. [14, 16]), representing movies, actors and directors. With respect to the users' preferences on items, we consider a simple model representing preference of users towards movies. This kind of information can be extracted from social networks [20–22] or from specialized site databases (e.g. IMDB [23]) provided that linkage between the social network profiles and specialized site accounts are established by the users. Finally, regarding the influencer information, we consider periodic snapshots computed from the social network graph. The complete information repository schema is illustrated in Fig. 1. The left side of this figure depicts the schema extension used in our work to support the personalization process, which includes the following information:

- *Users*: users are extracted from the social network API or from dataset files (e.g. the MovieLens dataset [24] or the Amazon datasets [25,26]). Since our personalization scheme does not include demographic or other user-profile based information, only the user id is retained in the database.