Decision Support Systems 69 (2015) 92–106

Contents lists available at ScienceDirect

Decision Support Systems

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

Alike people, alike interests? Inferring interest similarity in online social networks

Xiao Han [a,*], Leye Wang [a], Noel Crespi [a], Soochang Park [a], Ángel Cuevas [a,b]

[a] Institut-Mines Télécom, Télécom SudParis, 9 rue Charles Fourier, 91011 Evry Cedex France
[b] Universidad Carlos III de Madrid, Av de la Universidad, 30 28911 Legans, Madrid, Spain

ARTICLE INFO

Article history:
Received 16 April 2014
Received in revised form 22 September 2014
Accepted 30 November 2014
Available online 9 December 2014

Keywords:
Social networks
Interest similarity
Homophily
Prediction model

ABSTRACT

Understanding how much two individuals are alike in their interests (i.e., *interest similarity*) has become virtually essential for many applications and services in Online Social Networks (OSNs). Since users do not always explicitly elaborate their interests in OSNs like Facebook, how to determine users' interest similarity without fully knowing their interests is a practical problem. In this paper, we investigate how users' interest similarity relates to various social features (e.g. geographic distance); and accordingly infer whether the interests of two users are alike or unalike where one of the users' interests are unknown. Relying on a large Facebook dataset, which contains 479,048 users and 5,263,351 user-generated interests, we present comprehensive empirical studies and verify the *homophily* of interest similarity across three interest domains (movies, music and TV shows). The homophily reveals that people tend to exhibit more similar tastes if they have similar demographic information (e.g., age, location), or if they are friends. It also shows that the individuals with a higher interest entropy usually share more interests with others. Based on these results, we provide a practical *prediction model* under a real OSN environment. For a given user with no interest information, this model can select some individuals who not only exhibit many interests but also probably achieve high interest similarities with the given user. Eventually, we illustrate a use case to demonstrate that the proposed prediction model could facilitate decision-making for OSN applications and services.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Online Social Networks (OSNs) have boomed and attracted a huge number of people to join them over the last decade. In OSNs, participants publish their profiles, make friends, and produce various contents (photos, answers/questions, videos, etc.). Unlike legacy web systems, OSNs are organized around both people and content, which provide us with unprecedented opportunities to understand human relationships, human communities, human behaviors and human preferences [13, 17, 27].

With the evolution of OSNs, understanding to what extent two individuals are alike in their interests (i.e., interest similarity) has become a basic requirement for the organization and maintenance of vibrant OSNs. On the one hand, such information about users' interest similarity could be leveraged to support friend recommendation and social circle maintenance. For instance, the decision to recommend users who share many interests with each other to be friends could increase users' approval rate of recommendation, because people usually aggregate by their mutual interests [14]. On the other hand, knowing interest similarity between users also facilitates social applications and advertising. For example, instead of randomly hunting for clients, exploring those users with a high interest similarity with existing clients could efficiently enlarge client groups for application providers and businesses.

However, estimating interest similarity between two users is not a straight-forward issue since users do not always explicitly elaborate their interests. In the Facebook data set prepared for this study, 51.6% of users do not present any interests in their profiles; and among nine interest domains in the dataset, except for movies, music and TV shows, less than a quarter of users reveal their interests in any of the other six interest domains (e.g., books, sports or games). Since such lack of users' interests occurs quite often in the real OSN environment, how to infer two users' interest similarity without complete information about their interests poses a challenge.

To deal with this problem, we investigate how two users' interest similarity relates to various social features in depth (e.g. profile overlap, geographic distance, and friend similarity) and further infer whether two users are alike/unalike in interest according to these learned relations. Existing studies have already demonstrated that friends share more interests than strangers [1] and verified that interest similarity strongly correlates to the trust between users [32]. However, the work to date has not address the issue of inferring users' interest similarity without complete information about users' interests. Furthermore, we carry out a comprehensive analysis on the correlations between users'

[*] Corresponding author. Tel.:+33 01 60 76 41 65.
E-mail addresses: han.xiao@telecom-sudparis.eu (X. Han), leye.wang@telecom-sudparis.eu (L. Wang), noel.crespi@telecom-sudparis.eu (N. Crespi), soochang.park@telecom-sudparis.eu (S. Park), acrumin@it.uc3m.es (Á. Cuevas).

http://dx.doi.org/10.1016/j.dss.2014.11.008
0167-9236/© 2014 Elsevier B.V. All rights reserved.

interest similarity and diverse social features, and have unearthed additional relative factors that could enhance interest similarity prediction.

Particularly, we quantify interest similarity over an aggregation of user pairs by two metrics: *probability of sharing interest*, defined as the likelihood that two users have any mutual interests; and *degree of interest similarity*, which captures interest overlaps between two users based on the weighted cosine similarity. In addition, we extract social features (e.g. profile overlap, geographic distance, and friend similarity) from users' social information regarding three aspects: demographic information (age, gender, location, etc.), social relations (i.e., friendship), and obtainable users' interests. Specifically, we conduct the study in three interest domains, namely movies, music and TV shows, over a large dataset of 479,048 users and 5,263,351 user-generated interests crawled from Facebook.

We highlight our key findings captured from the wide variety of analysis — the homophily of interest similarity. Generally, homophily shows the level of homogeneity in people's social networks in relation to multiple sociodemographic, behavioral and intrapersonal characteristics [16]. Specifically, in this paper, homophily

- reveals that people tend to be interested in the same movies, music and TV shows when they are similar in their demographic information, such as age, gender and location;
- implies that friends have higher interest similarity than strangers. Furthermore, the interest similarity increases if two users share more common friends;
- indicates that the individuals with a larger interest entropy are likely to share more interests with others. Note that we exploit interest entropy to quantify the characteristics of one user's interests. A user's interest entropy is influenced by two factors: the total number of a user's interests and the popularity of these interests. The more interests a user presents, and the less popular the interests are, the more the user gains in interest entropy.

Based on the empirical studies, we propose a prediction model with a number of features (e.g. geographic distance, friend similarity and interest entropy). This prediction model can determine whether two users are similar or not in interest when one of the users does not provide his interests. The prediction result can be properly applied to various interest similarity based applications (e.g., recommendation system [3, 5], friend prediction [1, 10] and user evaluation system [4]). For instance, the model can help to address the *new user problem* in the typical collaborative recommendations. Normally, a collaborative recommendation system recommends a user some items that are liked by the others with similar interests. Whereas, the recommendation may fail when it comes to a *new user u* not revealing his interests, as the system cannot determine which of its existing users may share interests with *u*. In this case, even without *u*'s interests, the proposed prediction model is able to find some existing users who are predicted being similar to *u* and recommend *u* some items according to their interests.

In summary, the main contributions of this paper include:

- To the best of our knowledge, this is the first work to infer the interest similarity of two users where we do not know one of the user's interests. Owing to the frequent lack of users' interest in OSNs and the common requirement for applications of knowing the interest similarity between users, this research problem has a practical significance.
- We capture various social features depending on users' social information and investigate how interest similarity relates to these social features through a comprehensive perspective at a collective level. We uncover the homophily between these social features and users' interest similarity. Relying on a large dataset crawled from Facebook, the analytical results can advance the collective knowledge of OSNs.
- We devise a practical interest similarity prediction model based on the learned social features, namely *InterestSim* model. We also introduce two baselines referred to *Friend* model and *DemoSim* model. These two baselines depends on users' friendships [12, 29] and demographic similarity [7, 15, 20] respectively. The experiments show that *InterestSim* model outperforms *Friend* and *DemoSim* model by 12%–16% and 3%–4% respectively in terms of AUCs in different interest domains.
- We illustrate a use case where we leverage the proposed *InterestSim* model to practically address the *new user recommendation problem*. Compared with several state-of-the-art approaches, it turns out that our proposed *InterestSim* model can facilitate the *new user recommendation* with a higher precision.

2. Literature review

2.1. Studies on OSNs

Understanding social characters from large-scale OSNs is a hot research topic in recent years. Jure et al. conduct a comprehensive analysis on the MSN message network [13], and Alan et al. examine and compare four social networks (Flickr, YouTube, LiveJournal, Orkut) simultaneously [17]. These early studies mainly shed light on the high-level characteristics and verified many relationship based properties in OSNs, such as power law and small world [27]. Complementary to these studies on basic relationship social graph, some other work aims at users' interactions, such as posts, comments and mentions, and analyzes features on the user interaction graph [28, 30]. Different from the above work, we concentrate on a more specific question — how various social features would affect two users' interest similarity.

In fact, many ways are proposed to model users' similarity. Six similarity measurements are compared in [26] where the authors conclude that cosine distance performs the best for recommending online communities to users. Additionally, users' similarity can be measured by various information, such as profile similarity, connection similarity and interest similarity. Users' similarityis proved to be related to their friendship to some extent. This relation is usually leveraged to estimate the relationship strength between users [1, 31]. Relying on this relation, some other work infers users' missing profile properties, such as age [9] and school [18], via their social relations. In this work, we discuss the users' interest similarity.

Users' interests are normally desirable to know for many applications. When a user's interests cannot be obtained, it is common to infer his interests from the interests of other users who probably are similar to him. For instance, authors deduce a user's interests by considering this user's social neighbors' interests [29]. Also interests are proposed to be inferred from the users who share more demographic attributes [7, 15, 20]. Although [12] evaluates the interest similarity between pairs on CiteUlike and concluded that social connected users exhibit significantly higher interest similarity than the disconnected ones. Unfortunately, to our knowledge, how various social features relate to users' interest similarity has not been discussed in detail in any previous studies. This paper evaluates the interest similarity with multiple social features including demographic characteristics, friend relations as well as interest entropy.

Entropy is widely leveraged in the analysis of OSNs, besides demographic information and friendship. As a lower entropy generally implies a higher predictability, entropy is employed to study the mobility patterns and to infer the predictability of mobile phone users' behavior [21, 25]. Entropy is also used over users' interests and measures to what extent those users focus on topic categories [2, 11]. Our work tries to capture the patterns of users' interests by using interest entropy, where the initial intention is to investigate whether the interest entropy relates to the interest similarity. If the interest entropy does correlate to the interest similarity, then we can introduce it into the prediction as a social feature with other demographic and friendship features.

Download English Version:

https://daneshyari.com/en/article/553101

Download Persian Version:

https://daneshyari.com/article/553101

Daneshyari.com