Int. J. Human-Computer Studies 76 (2015) 50–66

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

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

Collaborative Filtering for people-to-people recommendation in online dating: Data analysis and user trial

A. Krzywicki [a,*], W. Wobcke [a], Y.S. Kim [b], X. Cai [a], M. Bain [a], A. Mahidadia [a], P. Compton [a]

[a] School of Computer Science and Engineering, University of New South Wales, Sydney NSW 2052, Australia
[b] School of Management Information Systems, Keimyung University, 1095, Dalgubeol-daero, Dalseo-gu, Daegu, Republic of Korea

ARTICLE INFO

Article history:
Received 11 March 2013
Received in revised form
11 December 2014
Accepted 14 December 2014
Communicated by M. Zanker
Available online 20 December 2014

Keywords:
Recommender systems
Machine learning
User study

ABSTRACT

A common perception is that online dating systems "match" people on the basis of profiles containing demographic and psychographic information and/or user interests. In contrast, product recommender systems are typically based on Collaborative Filtering, suggesting purchases not based on "content" but on the purchases of "similar" users. In this paper, we study Collaborative Filtering for people-to-people recommendation in online dating, comparing this approach to a baseline Profile Matching method. Initial data analysis highlights the problem of over-recommending popular users, a standard problem for Collaborative Filtering applied to product recommendation, but more acute in people-to-people recommendation. We address this problem with a two-stage recommender process that employs a Decision Tree derived from interactions data as a "critic" to re-rank candidates generated by Collaborative Filtering. Our baseline Profile Matching method dynamically chooses, for each user, attributes that contribute most significantly to successful interactions with candidates having the best matching attribute value. The key evaluation metric is *success rate improvement*, the increase in the chance of a user having a successful interaction when acting on recommendations. Our methods were first evaluated on historical data from a large online dating site and then trialled live over a 9 week period providing recommendations via e-mail to a large number of users. The trial confirmed the consistency of the analysis on historical data and the ability of our Collaborative Filtering method to generate suitable candidates over an extended period. Moreover, the Collaborative Filtering method gives a higher success rate improvement than Profile Matching.

1. Introduction

Recommender systems have become important tools to help users deal with information overload and the abundance of choice. In recent years, due to the proliferation of social networking and online dating sites, people-to-people recommendation has become an active research area (e.g. Terveen and McDonald, 2005; Brŏzovský and Petřiček, 2007; Hitsch et al., 2010; Fiore et al., 2010). The main difference between item-to-people and people-to-people recommender systems is that people to people interactions are two-way (Cai et al., 2010; Krzywicki et al., 2010; Pizzato et al., 2010), so a recommendation is considered successful only if an initial contact is reciprocated – this type of recommender is called a *reciprocal recommender* by Pizzato et al. (2010, 2013). As a consequence, it is critical to avoid giving candidates who are likely to reject a user's contact. A people-to-people recommender system must therefore take into account a user's taste (the people they find desirable) but also their attractiveness (how likely they are to be accepted by a potential candidate), since both of these factors determine the success of an interaction. Another important difference is that, while the same item can be recommended to a large number of users (since an item can be repeatedly reproduced), people can accept only a limited number of contacts. So a people-to-people recommender system should not suggest the same candidate to too many users at the same time.

This work concerns people-to-people recommendation in online dating. While there are various types of online dating sites, the type we have in mind is where users can freely and easily browse a large number of user profiles before deciding whom to contact. In broad terms, the typical "user experience" with such a site involves, after initially setting up a profile, anonymously sending messages (selected from a fixed predefined set) to prospective partners, who may accept, reject or ignore these contacts, then, generally for those accepted, proceeding to (paid) open communication,

* Corresponding author. Tel.: +61 2 9385 7201.
E-mail addresses: alfredk@cse.unsw.edu.au (A. Krzywicki), wobcke@cse.unsw.edu.au (W. Wobcke), YangSok.Kim@kmu.ac.kr (Y.S. Kim), xcai@cse.unsw.edu.au (X. Cai), mike@cse.unsw.edu.au (M. Bain), ashesh@cse.unsw.edu.au (A. Mahidadia), compton@cse.unsw.edu.au (P. Compton).

http://dx.doi.org/10.1016/j.ijhcs.2014.12.003
1071-5819/

though it is also possible for a user to initiate paid communication to anyone, including those who have previously rejected a contact. Moreover, at the same time as contacting other people, the user may receive contacts, which they in turn can either accept, reject or ignore, and may also receive open communications. At some point, communication with a prospective partner is taken offline, and eventually the user reduces activity and leaves the site. There is much potential during this process to provide recommendations. Helping users progress from initial contact to communication can (for the dating site company) increase revenue and user retention, maintain the pool of candidates, and improve user satisfaction, and (for the user) improve their overall experience, and in particular, increase their chance of a successful interaction. In this paper, we focus on the stage in this process where users initiate contact with potential partners, since we have reliable data concerning these interactions, and suitable metrics for evaluating the success of recommendations, whereas we have no information about the success of open communication or feedback from offline meetings.

In such online dating sites, users typically register to enter a *profile*, which consists of a photo, demographic information (such as age, location, occupation, smoking and drinking habits, etc.), their preferred partner characteristics (of the same type of information), some additional short text about their interests (music, sports, movies, etc.) and longer free text biographical information. This user profile can be viewed and searched by other users, but not all of this information needs to be provided before users can start sending messages. Users expect to find partners quickly and to proceed to exchanging messages with their selected candidates. Search is typically based on profile attributes and/or keywords that match interests or biographical text, however search may return too many results, which then need to be filtered manually, or too few results, so search criteria need to be refined. With many thousands of active users, and thousands more joining daily, the problem of abundant choice is overwhelming. Recommender systems are an ideal way to aid the user in selecting potential candidates for contact.

There are several specific ways in which a recommender system can improve user experience. First, some users receive large amounts of potentially unwanted communication (especially if they are the most attractive people), while others receive very low amounts of communication (e.g. if their profiles are not easily found via common search terms). A message sent to a highly attractive user is likely to be ignored. By directing users' attention towards more suitable candidates, their chance of success is likely to be higher, and at the same time, the number of contacts received by highly popular users will be reduced. Second, search engines typically do not rank results by compatibility with the user (instead ranking the results the same for all users, such as by recency of activity). This requires users to examine many profiles in order to find a suitable partner. By better ranking potential candidates, some of the burden faced by users in sifting through long lists of candidates can be reduced, again increasing the chance of a successful interaction and improving user experience.

A common popular view is that online dating sites provide "matches" based on profile information; such matches may be one-to-one or many-to-many. In terms of recommendation, this is a content-based approach, since matches are determined by comparing profiles of users and candidates, either using certain preferred attributes or some type of learned model. In contrast to such content-based methods, successful item-to-people recommender systems such as that used by Amazon.com (Linden et al., 2003), typically use Collaborative Filtering, where products are recommended to a user based on the purchases of "similar" users – two users are similar to the degree that they have made the same purchases in the past. Surprisingly, the literature on online dating contains little mention of Collaborative Filtering.

The main purpose of this paper is to study Collaborative Filtering techniques for people-to-people recommendation in online dating in comparison to a baseline Profile Matching method. For evaluation, we use both, a historical analysis of interaction data obtained from an online dating company, and a trial of selected methods on the same site. Initial historical data analysis highlights the severity of the problem of over-recommending popular users, an expected problem with Collaborative Filtering. To overcome this problem, we develop a two-stage "cascaded" recommendation method that makes use of a Decision Tree "critic" to re-rank candidates generated by Collaborative Filtering based on prior interactions (Krzywicki et al., 2010, 2012). This has the effect of lowering the rank of candidates who are too popular (attractive) relative to the target user (the user receiving the recommendation). Moreover, the use of the Decision Tree critic, in being derived from interaction data, is general, and candidates are re-ranked based on criteria other than popularity. For each user, the baseline profile matching method dynamically chooses attributes such that the value of the attribute for the user and the corresponding best matching value of the same attribute for the candidates, contribute most significantly to success rate improvement (Kim et al., 2010, 2012).

Our evaluation framework utilizes two metrics, *success rate improvement* and recall. These metrics have different interpretations for historical data analysis and for analysis of trial results. For historical data analysis, evaluation of a recommendation method is with respect to a test set of successful interactions: users have not seen the recommendations, rather the aim is to determine how well a method predicts successful interactions from the interactions that actually occurred. In the trial setting, recommendations are delivered to users via e-mail and the aim is to change user behaviour so that they contact the recommended candidates. One theme of this paper is that, while the values of metrics on historical data do not translate directly to the trial, historical data analysis is sufficient for predicting how methods compare in the live setting.

Success rate improvement measures how much more likely a group of users is to have a successful interaction when following a recommendation as compared to their own behaviour. This is a ratio of success rates: their success rate when following recommendations divided by their baseline success rate. In information retrieval terms, success rate is analogous to precision, which is the proportion of a set of documents returned by a query that are relevant to the user. However, the definition of success rate for a method is more complicated because we have both recommendations and the notion of a successful interaction. Success rate is defined not as the proportion of successful interactions out of all generated recommendations, but rather as the proportion of successful interactions out of those interactions initiated by users from recommendations. The reason for using success rate improvement rather than just success rate is that baseline success rates vary depending on the group of users under consideration. Using success rate improvement enables a fairer comparison of methods since the effect of these variations is factored out, so any improvement in success rate is attributable to the method, rather than due to characteristics of the user group. However, the variations in baseline success rate for the groups we consider are typically very small, and so to verify that the differences are due to the methods, we give all baseline success rates used in the evaluations.

In information retrieval, recall is the proportion of relevant documents that a query returns out of all relevant documents. We define recall to be the proportion of successful interactions generated by a recommendation method. For historical data analysis, recall thus measures the proportion of successful interactions in a test set generated by a method (i.e. interactions with