Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

Personality-aware followee recommendation algorithms: An empirical analysis

Antonela Tommasel [*,1], Alejandro Corbellini [1], Daniela Godoy [1], Silvia Schiaffino [1]

ISISTAN Research Institute, UNICEN University, Campus Universitario, Tandil (B7001BBO), Argentina

A R T I C L E   I N F O

A B S T R A C T

As the popularity of micro-blogging sites, expressed as the number of active users and volume of online activities, increases, the difficulty of deciding who to follow also increases. Such decision might not depend on a unique factor as users usually have several reasons for choosing whom to follow. However, most recommendation systems almost exclusively rely on only two traditional factors: graph topology and user-generated content, disregarding the effect of psychological and behavioural characteristics, such as personality, over the followee selection process. Due to its effect over people's reactions and interactions with other individuals, personality is considered as one of the primary factors that influence human behaviour. This study aims at assessing the impact of personality in the accurate prediction of followees, beyond simple topological and content-based factors. It analyses whether user personality could condition followee selection by combining personality traits with the most commonly used followee predictive factors. Results showed that an accurate appreciation of such predictive factors tied to a quantitative analysis of personality is crucial for guiding the search of potential followees, and thus, enhance recommendations.

1. Introduction

Social networks and micro-blogging sites have increased their popularity in recent years, with hundreds of users joining everyday. Also, users spend more and more time on those sites sharing personal and relevant information and making new friends. In this context, finding high quality social ties becomes a difficult task due to not only the continuous expansion of micro-blogging communities, but also the difficulty of characterising users and their behaviour, which can influence their friend selection patterns. These situations lead to the imperious need of developing both accurate user characterisations and followee recommendation techniques.

In information-oriented social networks like *Twitter*, users might base their decision of starting to follow other users on several and distinctive reasons or characteristics. For example, a user might follow some users because they publish interesting information, others because they have the same interests, others because they are celebrities or popular individuals in the micro-blogging community, or even because they share some common friends, among other possible explanations. Consequently, understanding the reasons why a user selects who to follow becomes crucial for designing accurate and personalised recommendation strategies.

Although personality is considered as one of the primary factors influencing human behaviour, and thus, social relationships, most of the existing recommendation systems only rely on content and topological characteristics as predictive factors for followee recommendation. Thus, they neglect how users' interests and decisions are affected by psychological characteristics. This study aims at assessing the impact of personality in the accurate prediction of followees beyond simple topological and content-based factors. To this end, some of the most common factors influencing the selection of followees in *Twitter* are analysed in relation with each person's own behaviour and characteristics, denoted by their distinguishable personality traits.

The rest of this paper is organised as follows. Section 2 describes strategies found in the literature for personality-based recommendation in the context of traditional content-based and collaborative filtering recommendation systems, and the features commonly observed for predicting a user's personality. Section 3 introduces the problem of recommending who to follow in social networks. Section 4 describes the *Twitter* data used for experimentation. Section 5 presents the proposed strategies for

* Corresponding author. Tel.: +54 249 4439682x35; fax: +54 249 4439681.
E-mail addresses: antonela.tommasel@isistan.unicen.edu.ar (A. Tommasel),
alejandro.corbellini@isistan.unicen.edu.ar (A. Corbellini),
daniela.godoy@isistan.unicen.edu.ar (D. Godoy),
silvia.schiaffino@isistan.unicen.edu.ar (S. Schiaffino).
[1] Also Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.

quantitatively evaluating users' personality and how to combine it with other widely used followee recommendation factors. Section 6 measures the impact of introducing personality traits in a followee recommendation algorithm. Additionally, it discusses the statistical significance of the obtained results. Finally, Section 7 summarises the conclusions obtained from the performed experimental evaluation.

2. Related work

Most approaches for building recommendation systems focus on improving recommendation precision, instead of investigating how users are inherently influenced by their own personality, among other human factors that could influence the quality of suggestions. Wu et al. (2013) aimed at adding personality scores to a content-based movie recommendation system in order to generate more personalised and diverse recommendations. To assess the impact of personality in the recommendation process, the authors developed two systems. The first one used personality to positively adjust the item diversity, whereas the second one aimed at recommending items regardless of users' personality. The purpose of the second system was to analyse whether people would have negative opinions when the recommendation diversity did not match their personality. The study comprised 52 Chinese users who were asked to answer the Big Five test designed by Hellriegel et al. (1987), as well as to specify their movie preferences, and rate the recommended movies in both systems. Then, users were required to answer a questionnaire to express their overall opinions about the recommendation accuracy, system competence and overall satisfaction. The first system obtained significantly higher scores than the second one on every evaluated aspect. Particularly, most users declared that the first system showed recommendations that matched their interests, and that it was more helpful for discovering interesting movies. However, as the approach was not compared with non-personality based systems, it cannot be guaranteed to outperform traditional recommendation systems.

Hu and Pu (2011) and Tkalcic et al. (2009) presented approaches to include personality scores as complementary information in traditional rating-based collaborative recommendation systems. Both relied on the explicit assessment of personality through the Big Five test[2] and the IPIP[3] questionnaire respectively. Hu and Pu (2011) based their experimental evaluation on 111 users extracted from the DiscoverMusic dataset (Hu and Pu, 2010). The approach was compared to a traditional rating-based filtering system, showing that the system combining ratings and personality significantly outperformed the systems solely based on either ratings or personality features. Additionally, the approach was reported to help solving the cold-start problem when offering recommendations to new users or in sparse datasets. Tkalcic et al. (2009) proposed to measure the similarity of users in collaborative filtering by computing the Euclidean distance between the personality scores across the five dimensions. Two variants were compared, one considering only the neighbours of a certain user and the other considering all users. Experimental evaluation was based only on 52 users who were asked to rate several items to obtain their item-ratings profile. Results showed that the personality-based recommendation outperformed the rating-based one.

All of the presented approaches share the same drawbacks. First, they included a relatively small number of users, which

prevents the generalisation of results. Second, personality was self-assessed through questionnaires, which not only requires the explicit participation of users but also could result in biased scores. The own view of themselves reported by users could not reflect their actual behaviour and, in turn, their real personality (Selfhout et al., 2009). Finally, the approaches were tested in the context of item recommendation using collaborative filtering techniques, none of the works include personality in the context of user recommendation in social networks. In consequence, the impact of personality in social recommendation systems is yet to be proven.

3. Followee recommendation problem

Social network data grows at an unprecedented rate due to the massive use of social networking sites. Millions of users have started to use micro-blogging sites since their beginning as a tool not only to propagate and share information, but also for finding new friends. Due to such exponentially increasing volume of online activity, effective recommendation systems are needed for guiding users in the search of useful and interesting items. In the context of social networks, recommendation systems can be used to suggest users worth following. This can be seen as a link prediction problem (Liben-Nowell and Kleinberg, 2003), i.e. the problem of inferring which user interactions are likely to occur in a short-time.

Most of the existing followee recommendation systems on micro-blogging platforms rely on either topological or content-based factors (Rowe et al., 2012). Link prediction based on content-based factors aims at suggesting users based on the textual or topical similarity with the target user, i.e. the user receiving the recommendations. In turn, link prediction based on topological factors suggests users to a target user based on a comparison of their neighbourhoods. Since this work is concerned with assessing the influence of user personality over these factors when selecting who to follow, Sections 3.1 and 3.2 describe respectively the content- and topology-based factors considered in this study. Finally, Section 3.3 introduces personality as a followee recommendation factor by stating its influence over social relationships.

3.1. Content-based factors

In micro-blogging platforms users can follow others and subscribe to the content they publish. Thus, content becomes a valuable factor for link prediction, i.e. a user is likely to have a link with other users sharing the same information preferences (Romero and Kleinberg, 2010). The interest of a user can be characterised by means of profiles based on not only the content of the published tweets, but also the tweets a user reads. Whereas the first alternative indicates the interests of users in terms of the information they create and publish, the second one indicates the interests of users in terms of the information they consume, i.e. the information they choose to read and deemed as interesting. These profiles will be referred as *publishing profile* and *reading profile* respectively.

The set of tweets t for a user u_j can be denoted as:

$$tweets(u_j) = \{t_i, \ldots, t_n\} \tag{1}$$

The *publishing profile* of a user is built by considering all user tweets under the assumption that users tend to tweet about things that are relevant to them. Formally, the profile of user u_j can be defined as:

$$pub-profile(u_j) = tweets(u_j) \tag{2}$$

The rationale behind the decision of building a *reading profile* is to adequately capture the preference and interests of users

---