Future Generation Computer Systems 78 (2018) 430–439

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Temporal people-to-people recommendation on social networks with sentiment-based matrix factorization

Davide Feltoni Gurini *, Fabio Gasparetti, Alessandro Micarelli, Giuseppe Sansonetti

Roma Tre University, Via della Vasca Navale 79, 00146 Rome, Italy

HIGHLIGHTS

- A novel people-to-people recommender system on social networks has been proposed.
- The recommender relies on the identification of users' attitudes: sentiment, volume, objectivity.
- A three-dimensional matrix factorization model, one for each attitude, has been devised.
- Temporal dynamics has been included into the factorization model.
- Recommender's accuracy and diversity increases with attitudes and temporal features.

ARTICLE INFO

Article history:
Received 20 June 2016
Received in revised form
21 December 2016
Accepted 14 March 2017
Available online 16 March 2017

Keywords:
People-to-people recommendation
Sentiment analysis
Matrix factorization

ABSTRACT

Nowadays, the exponential advancement of social networks is creating new application areas for recommender systems (RSs). People-to-people RSs aim to exploit user's interests for suggesting relevant people to follow. However, traditional recommenders do not consider that people may share similar interests, but might have different feelings or opinions about them. In this paper, we propose a novel recommendation engine which relies on the identification of semantic attitudes, that is, sentiment, volume, and objectivity, extracted from user-generated content. In order to do this at large-scale on traditional social networks, we devise a three-dimensional matrix factorization, one for each attitude. Potential temporal alterations of users' attitudes are also taken into consideration in the factorization model. Extensive offline experiments on different real world datasets, reveal the benefits of the proposed approach compared with some state-of-the-art techniques.

1. Introduction

Microblogging platforms are one of the most versatile and popular technologies on the Internet today. For instance, Twitter sees over 500 million microposts (or *tweets*) published every day on a huge variety of topics, with spikes of more than 100 thousand tweets per second when particular events occur.[1] With the proliferation of user-generated content such as reviews, discussion forums, blogs, and tweets, detecting sentiments and opinions from the Web is becoming an increasingly widespread form of data interpretation. In particular, sentiment analysis aims to understand subjective information, such as opinions, points of view, and feelings, expressed by users in the content they generate.

People-to-people recommendation is an important application on these platforms. Almost all the services are capable of recommending relevant users to follow. However, this recommendation task is not easy due to huge graphs of social ties and fast changing contents that must be analyzed. In this scenario, simple people recommendation algorithms based on content similarity and popularity paradigms are usually considered, at the expense of the recommendation accuracy.

In this paper, we propose a novel people-to-people recommender system that takes into account the users' attitudes towards discussed topics. The proposed recommender enables us to leverage users' attitudes such as sentiment, volume, and objectivity extracted from the semantics of tweets, define a *sentiment–volume–objectivity (SVO)* function, and exploit such knowledge to suggest relevant people to follow. The rationale behind this work is that people might have similar interests, but different opinions or feelings about them. Therefore, considering the contribution of users' attitudes may yield benefits to people recommendation. For example, two users involved in the discussion

* Corresponding author.
E-mail address: d.feltonigurini@gmail.com (D. Feltoni Gurini).

1 https://blog.twitter.com/2013/new-tweets-per-second-record-and-how (last visited on 20 December 2016).

http://dx.doi.org/10.1016/j.future.2017.03.020

about supporting Hillary Clinton for US President are likely to be friends. However, the two users may exhibit the same (both support or oppose Hillary Clinton) or contradictory sentiments (one supports and the other opposes). Therefore, we suppose that the two users are more likely to follow each other in the former case than in the latter.

To handle large-scale social networks, we model this recommendation task using matrix factorization techniques in four steps: (i) build a three-dimensional matrix in which each dimension is represented by a SVO user feature; (ii) learn a latent embedding space from the user-attitudes matrix; (iii) compute the user–user similarity by taking into account the three matrix dimensions; and (iv) recommend to a target user a list of relevant people to follow.

In this work, we address two research questions that arise when approaching the people-to-people recommendation problem:

1. Does content published by users and, in particular, the inferred attitudes, allows for a better identification of potential relationships that exist between them?
2. How does temporal analysis of these attitudes impact the accuracy of the recommendation?

The scientific contributions coming from this paper are: (i) an algorithm for people-to-people recommendation on microblogging platforms that takes advantage of features that represent the users' attitudes on specific topics; (ii) a comparative experimental results of a set of different evaluation metrics, including a range of non-accuracy measures, such as diversity and novelty; (iii) a proof of how the recommendation accuracy can be improved by taking into account the temporal variations of the attitudes expressed by the user; (iv) an evaluation of the proposed algorithm on real world datasets, showing that the considered users' attitudes have unequal correlation with respect to the accuracy of the recommendation, and strongly depend on the topic under consideration.

The rest of the paper is organized as follows: Section 2 introduces the problem formulation. Section 3 describes the recommendation algorithm. Section 4 presents the performed experiments to evaluate the proposed strategy and outlines main results. Section 5 contains a description of some state-of-the-art approaches. Finally, Section 6 reports our conclusions.

2. Problem formulation

In this section, we provide the definition of the people-to-people recommendation problem.

Let $\mathbb{U} = \{u_1, \ldots, u_N\}$ represent the set of users with a valid account on the micro-blogging platform. In our scenario, an adjacency matrix $A^{N\times N}$ represents the explicit ties, where each element $A_{i,j}$ denotes if the user u_i *follows* (or is *friend of*) the user u_j or not, and therefore is usually expressed by a binary value $\{0, 1\}$. Then, let $\overline{U} = \{u_1, \ldots, u_M\}$ represent the set of candidate users $u_j \in \mathbb{U}$ without an explicit tie with the target user u_i, that is,

$$\overline{U} = \{\forall u_j \in \mathbb{U} \mid i \neq j \wedge (A_{i,j} = 0 \wedge A_{j,i} = 0)\}.$$

Under this setting, the problem can be formulated as follows: given the matrix $A^{N\times N}$, which represents a known set of social relations between N users, define the following function r:

$$r : \mathbb{U} \times \overline{U} \rightarrow [0, 1] \tag{1}$$

such that, given a *target* user u and an adjacency matrix, returns a value between 0 and 1, which expresses the relevance degree of the candidate user u_j for the target user u_i. Based on such value, the system provides the target user with a recommendation list of the top relevant candidates.

First attempts to people-to-people recommendation take advantage of global models and collective classification for the definition of the r function. In other words, they operate on the whole graph of related nodes rather than deriving individual structural and content-based attributes. The problem is therefore seen as the optimization of one global objective function.

Since *link prediction problem* [1,2] aims at inferring future interactions and missing links on large graphs, various predictors based on the interpersonal social structure (e.g., common neighbors predictor) are also considered for the ranking task.

Our goal is to define the function r by extending the recommendation analysis to relevant information associated with users that can be retrieved by the micro-blogging platform, namely, the timeline consisting of sequences of microposts. In the rest of the paper, we indicate with $\mathbb{T}$ the set of potential microposts that can be published and with $T_u \subset 2^{\mathbb{T}}$ the most recent microposts published by the user u.

3. The proposed people-to-people recommendation

In this section, we introduce our method for recommendation. A strong correlation exists between the presence of a social tie between two users and the topical similarity of explicit activities of these users in the network [3]. Consequently, it is logical investigating the chance of predicting the presence of a tie based on user profile features. The idea behind the proposed approach is that, by taking into account attitudes, in terms of manifested expressions of favor or disfavor on specific matters, the accuracy of the people-to-people recommender is improved. Multiple steps are demanded to implement the recommendation task, as shown in Fig. 1.

The timeline of users $u_i \in \mathbb{U}$ are first retrieved. A traditional preprocessing of microposts simplifies the identification of relevant features. All characters are converted to lowercase letters and retweet designations (e.g., "RT"), citations, and URLs are removed. Then, text is tokenized into keywords, from which a list of unigram features is created. Traditional stopwords are excluded from the lists.

Micro-blogging services allow users to include metadata tags in the form of keywords followed by the hash symbol #, which are referred to as *hashtags*. By including them in posts, the author is suggesting them as good candidates in quality of search keys. Popular hashtags often refer to topics that most people are interested in, including breaking events and persistent discussions [4]. For this reason, they are often considered for clustering posts related to specific topics [5,6].

Let $\mathbb{C}$ denote the set of all possible concepts. Given a micropost τ, we indicate with $\tau^{(\mathbb{C})}$ the subset of concepts $\mathbb{C}$ that are included in τ, identified by extracting the hashtags in τ. By extension, $T_u^{(\mathbb{C})}$ is the set of concepts that are included in the user u's timeline. The so-obtained representation of microposts is subjected to the SVO analysis (see Sections 3.1 and 3.2), which aims at determining the user's attitude on each topic. Since determining similarities among users who have limited activity on specific topics is a challenging task, the SVO-based analysis is not performed on concepts not appearing in a timeline above a given frequency threshold (i.e., 10 tweets). This procedure is commonly followed when attitudes expressed by large audiences are explored [7].

Each user's timeline is subjected to a text categorization process based on a Support Vector Machine (SVM) algorithm [8], so that one or more categories belonging to the set $\mathbb{K}$ of all possible macro-categories are associated to the user according to the published content. These macro-categories (namely, *world*, *elections*, *business*, *technology*, *entertainment*, *sports*, *science*, and *health*) are similar to the ones of a popular online news aggregator [9]. The training set is built-up by retrieving titles and snippets of each macro-category on the aggregator over a period of one month. We denote with $T_u^{(\mathbb{K})} \subseteq \mathbb{K}$ the macro-categories assigned to the user u.