Information Sciences xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Information Sciences

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

Predicting the popularity of micro-reviews: A Foursquare case study

Marisa Vasconcelos*, Jussara M. Almeida, Marcos André Gonçalves

Department of Computer Science, Universidade Federal de Minas Gerais, Av. Antônio Carlos 6627, CEP 31270-010 Belo Horizonte MG, Brazil

A R T I C L E I N F O

Article history:
Received 24 October 2014
Revised 27 March 2015
Accepted 3 July 2015
Available online xxx

Keywords:
Popularity prediction
Content filtering
Location-based social networks
Micro-reviews

A B S T R A C T

We tackle the problem of predicting the future popularity level of micro-reviews, focusing on Foursquare tips, whose high degree of informality and briefness offer extra difficulties to the design of effective popularity prediction methods. Such predictions can greatly benefit the future design of content filtering and recommendation methods. Towards our goal, we first propose a rich set of features related to the user who posted the tip, the venue where it was posted, and the tip's content to capture factors that may impact popularity of a tip. We evaluate different regression and classification based models using this rich set of proposed features as predictors in various scenarios. As fas as we know, this is the first work to investigate the predictability of micro-review popularity (or helpfulness) exploiting spatial, temporal, topical and, social aspects that are rarely exploited conjointly in this domain.

1. Introduction

Nowadays, the social Web allows people to interact and freely express opinions on products, services or companies in real-time and in large-scale. More and more people base their buying decisions on online reviews written by others [8]. Yet, with the diffusion of smartphones, new services were created targeting mainly social networking users, who spend most of their time accessing information through mobile applications. In this environment, the communication is usually briefer mainly because of the limited amount of information that can be displayed on the mobile screen. This limitation also influenced the creation of new review services (Foursquare, Google+ Local) and the expansion of traditional desktop services to the mobile environment (Yelp, Trip Advisor). In these services, users write *micro-reviews* or *tips*, which are typically much more concise (up to 200 characters), often written while the information is still fresh in the user's mind, and may contain much more subjective and informal content. We here refer to such micro-reviews as simply *tips*.

Accurate tip popularity predictions can drive the design of automatic tip filtering and recommendation schemes, which in turn can help users find tips that are potentially more valuable more easily. Business owners may also benefit from such predictions as they are able to more quickly identify (and fix) aspects of their services or products that may affect revenues most.

However, as in longer review systems, the number of tips on a single product or service may be large and vary greatly in quality [22,26], which makes it hard for users to find helpful reviews. To support that task, many websites allow users to evaluate reviews. Tips, in particular, are often rated by other users clicking on a "like" mark. The number of "likes" received by a tip can then be seen as an estimate of its helpfulness. It can also be used as an estimate of the tip popularity, as it provides a lower bound

* Corresponding author. Tel.: +55 31 3409 5860; fax: +55 31 3409 5858.
E-mail addresses: marisav@dcc.ufmg.br, marisavas@gmail.com (M. Vasconcelos), jussara@dcc.ufmg.br (J.M. Almeida), mgoncalv@dcc.ufmg.br (M.A. Gonçalves).

http://dx.doi.org/10.1016/j.ins.2015.07.001
0020-0255/© 2015 Published by Elsevier Inc.

on the number of people who actually read the tip. Yet, this feedback is usually very sparse. Moreover, ranking tips based only on popularity votes is not useful for promoting recently posted reviews, with a few or no votes, which are doomed to be outranked by older reviews that have already received more votes, and thus lose visibility.

This problem has already inspired a series of studies attempting to automatically predict the quality and helpfulness of online reviews [21,27,41]. However, these efforts focused mostly on textual or content related features (e.g., review length, readability) [21,27,41], which are more suitable for verbose and more formally structured reviews. Moreover, the lack of a "like" does not imply that a tip was not helpful or interesting, as it may not have been seen by any user. This further contributes to make the automatic popularity prediction much harder than in systems that offer a rating scale (e.g., 1–5).

In this article, we study the problem of predicting the popularity *level* a tip will achieve at a future target time, where the popularity level is defined based on the total number of likes the tip receives until that time. More importantly, we focus on predicting the popularity or helpfulness of an individual micro-review exploiting *spatial, temporal, topical and social* aspects that are rarely *conjointly* exploited. To that end, we collected a large and comprehensive dataset of Foursquare tips, covering over 6 million tips and 5 million likes, posted by more than 1.8 million users. Thus, an important contribution of our work is to make this rich dataset available to the research community. Our investigation tackles the following three questions:

Q1: Which are the most important factors for predicting the popularity of Foursquare tips as soon as it is posted? We here identify three key entities related to the Foursquare system that may impact a tip's popularity: the user who posted the tip, the venue where it was posted, and its content. We investigate the potential benefits from exploiting these aspects to predict *at posting time* the popularity level of a given tip at a future time. To that end, we first propose a rich set of features related to these three entities, and then evaluate several regression and classification models as well as various feature subsets as predictor variables.

Q2: To what extent can we improve prediction by monitoring an early period of the tip in the system? How do the prediction models behave as we predict further into the future? Ultimately, how does tip popularity evolve over time? In Q1, we focused on the hardest prediction scenario, i.e., prediction at posting time, when the only information available about the tip consists of its content and historical patterns related to the user and the venue associated with it. In Q2, we assess to what extent prediction accuracy can be improved if, before predicting the popularity of a tip, we monitor the tip for a short period, gathering measures of how its popularity is evolving. Such early popularity measures are then added as predictors to our models. We compare our solutions against three state-of-the-art prediction models that also exploit such early popularity measures [35,38]. We also investigate how far into the future we can predict tip popularity with reasonable accuracy, that is, we analyze how robust our prediction models are when we perform long-term predictions. To motivate these analyses, we first investigate how tip popularity evolves over time.

Q3: Can we improve prediction accuracy by building specialized models? We here investigate whether factors related to a specific geographic region (e.g., city) or a type of venue impact a tip's popularity. To that end, we build specialized models using only tips posted in a specific city or in a specific venue category, and compare such models with the single general model.

In sum, the key contributions of this article are: (1) the identification and effectiveness investigation of a rich set of features that influence tip popularity (Q1), (2) a thorough evaluation of the relative performance of state-of-the-art regression and classification techniques in a domain (tip popularity prediction) where they have not been applied before, jointly with the selected features, in various prediction scenarios (Q1 and Q2), and (3) an investigation of the benefits from building specialized models (Q3). This work is a follow up on our prior study of tip popularity prediction [43], which addressed only Q1. We here build on it by introducing Q2 and Q3. In fact, we are unaware of similar work exploiting the temporality issue in the prediction of the popularity level of micro-reviews (research question Q2). Similarly, model specialization (research question Q3) is rarely performed under the spatial, social and topical dimensions we exploit in our investigation.

Next, we first discuss related work in Section 2. We introduce our prediction models and the features used by them in Section 3. A brief analysis of selected features is presented in Section 4.2. We discuss the results of prediction at posting time (Q1) in Section 5. We then investigate tip popularity temporal dynamics, addressing Q2, in Section 6. The impact of model specialization (Q3) is discussed in Section 7. Section 8 offers conclusions and directions for future work.

2. Related work

The popularity of a tip, estimated by the number of likes received, can also be seen as an estimate of the tip's helpfulness and quality, as it captures the number of people who found the tip useful. Thus, our work is related to two groups of studies: quality assessment of user generated content, and popularity prediction of online content. We review prior efforts in these two directions next.

2.1. Quality assessment of user generated content

Various prior studies focused on analyzing the quality of user generated content, including the quality of Wikipedia articles, video or news comments, and answers on community question answering forums.

For example, Dalip et al. [11] used Support Vector Regression (SVR) to estimate the quality of Wikipedia articles using features related to the text structure, citation network, and article revision history. Siersdorfer et al. [36] proposed a model that uses a term-based representation of YouTube comments (TF-IDF) to automatically classify them as likely to obtain a high overall rating or not. Similarly, Hsu et al. [9] exploited SVR to rank user comments on Digg based on their quality, using various features such as the comment's posting time, number of articles submitted, and comment length. Focusing on user reputation in a comment