Information Systems 57 (2016) 111–128

Contents lists available at ScienceDirect

Information Systems

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

GeoSRS: A hybrid social recommender system for geolocated data

Joan Capdevila [a,1], Marta Arias [b,2], Argimiro Arratia [b,2,3]

[a] *Barcelona Supercomputing Center, Barcelona Tech/Universitat Politècnica de Catalunya, Spain*
[b] *Computer Science Department, Barcelona Tech/Universitat Politècnica de Catalunya, Spain*

ARTICLE INFO

Article history:
Received 23 November 2014
Received in revised form
5 October 2015
Accepted 13 October 2015
Available online 25 October 2015

Keywords:
Recommender systems
Text mining
Quadtree
Crawling
Social networks
Location-based social network

ABSTRACT

We present *GeoSRS*, a hybrid recommender system for a popular location-based social network (LBSN), in which users are able to write short reviews on the places of interest they visit. Using state-of-the-art text mining techniques, our system recommends locations to users using as source the whole set of text reviews in addition to their geographical location. To evaluate our system, we have collected our own data sets by crawling the social network *Foursquare*. To do this efficiently, we propose the use of a parallel version of the *Quadtree* technique, which may be applicable to crawling/exploring other spatially distributed sources. Finally, we study the performance of *GeoSRS* on our collected data set and conclude that by combining sentiment analysis and text modeling, *GeoSRS* generates more accurate recommendations. The performance of the system improves as more reviews are available, which further motivates the use of large-scale crawling techniques such as the *Quadtree*.

1. Introduction

The proliferation of mobile communication and GPS technologies has allowed users to add geographical identification metadata to various social media, such as photographs, text reviews or video, among many others. Location-based social networks (LBSNs) [54] integrate into a single network user relations (the “social” part) and geo-spatial information (the “location-based” part). By taking into account the physical location of users, LBSNs are bridging the gap between physical world and virtual communities such as *Foursquare*,[4] *Facebook*[5] or *Twitter*.[6]

The extensive use of these social networking sites has made them invaluable sources of information. However, the sheer volume of data flowing through these sites, even for a single user, has made it increasingly difficult for humans to track all this information. Therefore, most social networking sites implement some sort of Social Recommendation System (SRS) [16]: for example, Twitter suggests who to follow, Facebook filters and prioritizes posts in users' walls and Foursquare

E-mail addresses: jc@ac.upc.edu (J. Capdevila), marias@cs.upc.edu (M. Arias), argimiro@lsi.upc.edu (A. Arratia).
[1] Supported by Obra Social “la Caixa”.
[2] Supported by MICINN project TIN2011-27479-C04-03 (BASMATI), MINECO project TIN2014-57226-P (APCOM) and Gen. Cat. project SGR2014-890 (MACDA).
[3] Additional support by MEC project MTM2012-36917-C03-03 (SINGACOM).
[4] https://foursquare.com/
[5] https://facebook.com/
[6] https://twitter.com/

http://dx.doi.org/10.1016/j.is.2015.10.003
0306-4379/

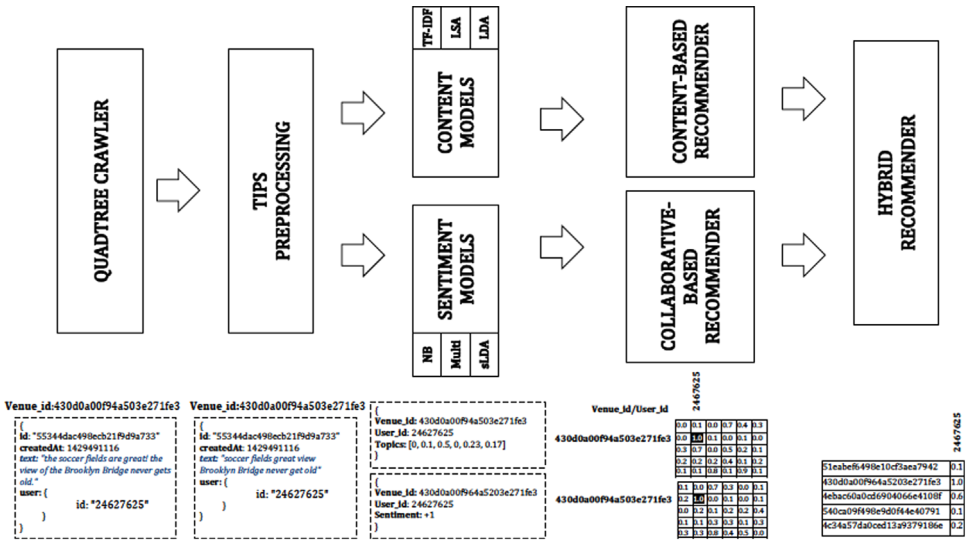

Fig. 1. Fully integrated location-based Information System. It includes modules for crawling, preprocessing crawled data, modeling and recommendation. The bottom part of the figure shows examples of data chunks that each module produces.

recommends locations where to go. When the social content is geotagged, it becomes strictly necessary to consider user and item localization in the recommendation paradigm.

Location-based recommendation constitutes a unique application in LBSNs and it substantially differs from traditional recommender systems in the fact that the latter does not take into account the spatial properties of users and items [34]. Moreover, location-based recommendation on top of LBSNs might also benefit from the interaction between the three layers composing a LBSN, namely the user, the location and the content layer [47].

In this paper, we propose a fully integrated system for *information retrieval* of geolocated data and *end-to-end location-based recommender*, suitable for the popular social network *Foursquare*. The reader should note, however, that our methods are applicable to any other social network that contains geolocated time-referenced reviews and hence, in our presentation we abstract from the fact that we are using this particular site.

We believe that fully operative recommender systems on top of LBSNs require end-to-end designs, capable of performing data retrieval from social networks, cleaning the noisy and duplicated data, extracting relevant features and, not least, performing recommendation.

Our proposed location-based information system is summarized in Fig. 1. It retrieves the short reviews together with their geolocated venues and reviewers identification as its basis for recommendations. In Foursquare, users are able to *check-in* to places of interest (venues), write short reviews (tips) for the venues where they checked into, and share this information with users within their social network. For the task of crawling venues, users and tips, we had to use Foursquare's API, an interface which imposes restrictions to the amount of information one can query and the amount of requests one can make. Therefore, it was imperative to devise a crawling mechanism that would make optimal use of the queries available to us. To achieve this task, we have designed a parallel version of the *Quadtree* algorithm [42], which is very well suited for crawling venues that are spatially distributed, while at the same time gaining considerable throughput. We have found that crawling all venues from large urban areas such as Mexico D.F. or New York in reasonable time was possible by the proposed Quadtree algorithm. We consider the application of the parallel Quadtree algorithm to this problem an important contribution of our paper and we believe that problems that require sweeping spatial devices (sensors) could also benefit from it.

To make recommendations our system makes extensive use of user's reviews (tips). In order to extract meaningful information from these free-form reviews, *GeoSRS* relies upon many state-of-the-art techniques for text mining and sentiment analysis, which are evaluated in terms of recommendation accuracy and the ones that outperform are selected to be used in *GeoSRS*. Another relevant contribution from our paper is the increase of accuracy when mixing the review's sentiment and content into a simple but rather effective weighted hybrid recommender setup [9]. Sentiment refers to the global opinion that is reflected in the review (positive, negative or neutral) while content indicates the topics that the review addresses. This enforces the idea that pure review-based choices are not merely based on the opinion reflected on a short review (The service was too slow), but also on the content relevant to the user (This is a kinda working place rather than a coffee shop).

To evaluate our system, we have collected our own data set of restaurants and tips from the area of Manhattan in New York City. We have chosen Manhattan due to the high density of venues and the number of active users, to validate both the scalability of the Quadtree crawler and the effectiveness and coverage of the recommender system. Recommender system is evaluated in terms of retrieval accuracy (performance) measures rather than statistical accuracy measures since we do not