Information Sciences 221 (2013) 142–169

Contents lists available at SciVerse ScienceDirect

Information Sciences

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

A comparative study of heterogeneous item recommendations in social systems

Alejandro Bellogín, Iván Cantador *, Pablo Castells

Escuela Politécnica Superior, Universidad Autónoma de Madrid, Spain

ARTICLE INFO

Article history:
Received 2 March 2012
Received in revised form 8 September 2012
Accepted 12 September 2012
Available online 3 October 2012

Keywords:
Recommender system
Social Web
Collaborative tagging
Social network
Implicit feedback
Evaluation

ABSTRACT

While recommendation approaches exploiting different input sources have started to proliferate in the literature, an explicit study of the effect of the combination of heterogeneous inputs is still missing. On the other hand, in this context there are sides to recommendation quality requiring further characterisation and methodological research – a gap that is acknowledged in the field. We present a comparative study on the influence that different types of information available in social systems have on item recommendation. Aiming to identify which sources of user interest evidence – tags, social contacts, and user-item interaction data – are more effective to achieve useful recommendations, and in what aspect, we evaluate a number of content-based, collaborative filtering, and social recommenders on three datasets obtained from Delicious, Last.fm, and MovieLens. Aiming to determine whether and how combining such information sources may enhance over individual recommendation approaches, we extend the common accuracy-oriented evaluation practice with various metrics to measure further recommendation quality dimensions, namely coverage, diversity, novelty, overlap, and relative diversity between ranked item recommendations. We report empiric observations showing that exploiting tagging information by content-based recommenders provides high coverage and novelty, and combining social networking and collaborative filtering information by hybrid recommenders results in high accuracy and diversity. This, along with the fact that recommendation lists from the evaluated approaches had low overlap and relative diversity values between them, gives insights that meta-hybrid recommenders combining the above strategies may provide valuable, balanced item suggestions in terms of performance and non-performance metrics.

1. Introduction

The environments in which recommender system technologies are commonly deployed have undergone a remarkable evolution in the last few years in terms of scale, richness, and complexity of the available data. Modern recommendation applications do not just have a user-item ratings matrix available, but complex user interaction data, rich item profiles, and large-scale (owned, public, or third-party) resources of many different types. This has been paralleled by a no less remarkable progress in the development of effective recommendation algorithms, and an evolution in the understanding of the role of recommendation functionalities in different application domains. The availability and convergence of technologies and resources in social systems – personal user data, user interaction records, user-contributed content, social networks, rich item databases, geospatial information, and so forth – have transformed the context in which the recommendation problem is addressed, multiplying the opportunities for enhanced solutions.

* Corresponding author.
E-mail addresses: alejandro.bellogin@uam.es (A. Bellogín), ivan.cantador@uam.es (I. Cantador), pablo.castells@uam.es (P. Castells).

0020-0255/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ins.2012.09.039

In recent years it has been made clear that a single algorithm is generally insufficient to optimise the effectiveness of recommendations – the Netflix contest[1] is a paradigmatic example of the superiority of hybrid recommenders over stand-alone approaches [34]. Likewise, recommendation based on a single source of input data is generally suboptimal, inasmuch as the multiplicity of available hints for good recommendations are missed [23]. At the same time, the purpose and scenarios for recommendation are diverse, and consequently, a single view of recommendation quality becomes insufficient to assess the value and usefulness of a recommendation approach. Evaluation methodologies and metrics need to be extended for this purpose, which is currently an open area of research and development in the field [8,12,51].

While recommendation approaches exploiting different input sources have started to proliferate in the literature, an explicit study of the effect of the combination of heterogeneous sources is still missing. In fact, there is little reported evidence yet on the comparative effectiveness enabled by different types of input data, when used individually, and the additional gain that can be leveraged from their combined effect. Furthermore, such a study requires an extension of the traditional evaluation dimensions and metrics in order to properly capture the effects of such combinations in their different relevant angles – a methodological gap that has been identified in the field [51]. The recommender systems literature has indeed been strongly focused on recommendation accuracy (to be more precise, on the accuracy in rating value prediction) as the main – generally the only – quality criterion. Aspects such as the coverage, diversity and novelty of recommended items, which may be critical in practice for the target users of recommendation – or the business beneath – have been barely addressed yet in the literature by a sound body of shared methodologies and metrics [1]. While progress in accuracy optimisation seems to have somewhat peaked in the field [34], and is getting circumscribed to small increments, we see considerable room for progress in such, to a large extent, unexplored dimensions. The combination of sources and the extension of evaluation methodologies thus present themselves as two interrelated research goals, where we see in the former a direction for improvement in terms of the latter, and the latter is a necessity to evaluate the former.

Motivated by the above considerations, in this paper we address the following research questions:

- **RQ1.** *Which available sources of information in social systems are more effective for recommendation?*
  We study this question in terms of several performance metrics borrowed from the Information Retrieval field, for recommendation approaches that exploit different sources of information, namely ratings, tags and social contacts.
- **RQ2.** *Do recommendation approaches exploiting different sources of information in social systems offer heterogeneous item suggestions, from which hybrid strategies may gain additional benefits?*
  We address this question by considering several recommendation quality metrics beyond accuracy, measuring dimensions as coverage, diversity, novelty, and overlap, on the recommendation approaches studied in RQ1.

In order to support this study we have implemented a set of generic content-based filtering, collaborative filtering, and social recommendation approaches for social systems, and have built three datasets with information of different types obtained from Delicious,[2] Last.fm[3] and MovieLens.[4] By using these recommenders and datasets we conduct a twofold experiment. First, we compare the performance of the recommenders with ranking quality metrics from the Information Retrieval field, namely precision, recall, and nDCG. Second, we compare additional characteristics of the recommenders with a number of novel metrics that measure coverage, diversity, novelty, and overlap of and between ranked lists of recommended items.

The reminder of the paper is organised as follows. Section 2 describes relevant works related to our study. Section 3 presents the evaluated content-based, collaborative filtering, and social recommendation approaches. Section 4 explains the experimental setup of the study, describing the utilised datasets, the proposed performance and non-performance metrics, and the followed evaluation protocol. Section 5 discusses the results obtained in the conducted experiment, and finally, Section 6 depicts some conclusions and future research lines.

2. Related work

With the advent of the Social Web, a variety of new recommendation approaches have been proposed in the literature. Most of these approaches are based on the exploitation of social tagging information and explicit friendship relations between users.

In **social tagging** systems, such as Delicious, Flickr and Last.fm, users annotate/tag resources (Web pages, photos, music tracks, etc.) for the purpose of personal multimedia content management, browsing and search. Interestingly, these personalisation functionalities can be extended to collaborative recommendation functionalities when the whole set of annotations [user-tag-resource] (known as *folksonomy*) are taken into account. A user's preferences are described in terms of her tags and tagged resources. Based on such a profile model, similarities with other users can be found, and item recommendations can be produced. Hotho and colleagues [30] present FolkRank, a PageRank-like algorithm applied to the tripartite graph formed by nodes associated to users, tags and items of a folksonomy, and weighted edges related to co-occurrences between users and tags, items and tags, and users and items. Other approaches like those proposed by Niwa and colleagues [44], Shepitsen

[1] Netflix prize, http://www.netflixprize.com.
[2] Delicious, Social bookmarking, http://delicious.com.
[3] Last.fm, Internet radio and music catalogue, http://www.last.fm.
[4] MovieLens, Movie recommendations, http://movielens.umn.edu.