COMPUTER SCIENCE REVIEW 20 (2016) 29–50

Available online at www.sciencedirect.com

ScienceDirect

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

Survey

A survey of active learning in collaborative filtering recommender systems

Mehdi Elahi[a,*], *Francesco Ricci*[b], *Neil Rubens*[c]

[a] *Politecnico di Milano, Milan, Italy*
[b] *Free University of Bozen-Bolzano, Bozen-Bolzano, Italy*
[c] *University of Electro-Communications, Tokyo, Japan*

ARTICLE INFO

Article history:
Received 14 July 2015
Received in revised form
9 May 2016
Accepted 10 May 2016
Published online 2 June 2016

Keywords:
Recommender systems
Collaborative filtering
Active learning
Rating elicitation
Preference elicitation
Cold start
New user
New item

ABSTRACT

In collaborative filtering recommender systems user's preferences are expressed as ratings for items, and each additional rating extends the knowledge of the system and affects the system's recommendation accuracy. In general, the more ratings are elicited from the users, the more effective the recommendations are. However, the usefulness of each rating may vary significantly, i.e., different ratings may bring a different amount and type of information about the user's tastes. Hence, specific techniques, which are defined as "active learning strategies", can be used to selectively choose the items to be presented to the user for rating. In fact, an active learning strategy identifies and adopts criteria for obtaining data that better reflects users' preferences and enables to generate better recommendations.

So far, a variety of active learning strategies have been proposed in the literature. In this article, we survey recent strategies by grouping them with respect to two distinct dimensions: *personalization*, i.e., whether the system selected items are different for different users or not, and, *hybridization*, i.e., whether active learning is guided by a single criterion (heuristic) or by multiple criteria. In addition, we present a comprehensive overview of the evaluation methods and metrics that have been employed by the research community in order to test active learning strategies for collaborative filtering. Finally, we compare the surveyed strategies and provide guidelines for their usage in recommender systems.

Contents

* *Corresponding author.*
E-mail address: mehdi.elahi@polimi.it (M. Elahi).
URLs: http://www.polimi.it (M. Elahi), http://www.unibz.it (F. Ricci), http://www.uec.ac.jp (N. Rubens).

http://dx.doi.org/10.1016/j.cosrev.2016.05.002

1. Introduction

This article surveys the state-of-the-art of active learning for collaborative filtering recommender systems. *Active Learning* in recommender systems tackles the problem of obtaining high quality data that better represents the user's preferences and improves the recommendation quality. This is done by identifying for each user a set of items contained in the system catalogue which have not been rated yet by the user, and by asking the user to rate them. The ultimate goal is to acquire additional ratings that will enable the system to generate better recommendations.

It is worth noting that users are typically not interested and are reluctant to rate items: this activity represents a cognitive cost for the user. For that reason it is important to carefully design an active learning strategy for identifying a small set of items to rate, whose ratings will convey to the system valuable information about users' preference, that is, ratings that will most improve the system generated recommendations.

Active learning is a subfield of machine learning [1–3]. While several machine learning tasks have been studied and a wide range of techniques have been already proposed and applied [4,1,5–8], the application of these techniques often requires a significant amount of high quality data that are not always easily available [9]. Active learning is precisely motivated by these situations where training data is expensive to obtain. In this case, active learning specifies how the data points that could better help the system to perform its task should be selected [10].

So far, several active learning strategies have been proposed and evaluated. We are here interested in those that have been applied to collaborative filtering recommender systems. These strategies have different features, and implement various heuristics for selecting the items to be presented to the users to rate. Instead of directly minimizing the system prediction error, by using heuristic, they try to improve other system properties that influence the system error. For instance, by acquiring ratings for popular items (as the popularity-based strategy does) the system tries to simply acquire more ratings (since many users are familiar with the popular items and can rate them). However, by adopting this heuristic the system may also acquire too many high ratings, as popular items tend to be rated high, hence the system may be also erroneously biased to predict higher ratings [11,12].

In this article, we classify all the important active learning strategies that we have found in the literature with respect to two dimensions that are descriptive and discriminative:

- **Personalization**: that describes to what extent the selection of items for the user to rate is adapted to the user's characteristics. We have *non-personalized*, and *personalized* strategies. In a non-personalized strategy the users are requested to rate the same items. Personalized strategies, on the other hand, ask each user to rate specific items. Personalization is an important aspect in active learning since the users have different tastes, preferences and experiences, hence, the usefulness of the various ratings could vary greatly from user to user. Selecting the items to rate while taking into account the preferences of each user, may provide a more pleasant experience to the user (e.g. by presenting them with items that they can actually rate), and at the same time may be more informative for the system.
- **Hybridization**: this dimension describes whether the strategy takes into account a single heuristic (criterion) for selecting the items for the users to rate or it combines several ones in order to create a more effective strategy. In this regard, the strategies can be classified into *single-heuristic* (or *individual*) and *combined-heuristic* (or *combined*) strategies. Single-heuristic strategies implement a unique item selection rule and select items only based on that. Combined-heuristic strategies hybridize single-heuristic strategies by aggregating and combining some of them, hence utilizing multiple item selection rules in order to better identify the more useful items to rate.

These dimensions where first identified in a previous short article [13]. Here, we illustrate more strategies and we provide