

A New Method for Collaborative Filtering Recommender Systems: The Case of Yahoo! Movies and TripAdvisor Datasets

Mehrbakhsh Nilashi^{a,b,*}, Maryam Salahshour^a, Othman Ibrahim^a, Abbas, Mardani^c, Mohammad Dalvi Esfahani^a,
Norhayati Zakuan^c

^a Faculty of Computing, Universiti Teknologi Malaysia, 81310 Skudai Johor, Malaysia

^b Department of Computer Engineering, Lahijan Branch, Islamic Azad University, Lahijan, Iran

^c Faculty of Management, Universiti Teknologi Malaysia (UTM), 81310 Skudai Johor, Malaysia

* Corresponding author email address: nilashidotnet@hotmail.com

Abstract

Collaborative Filtering (CF) techniques, which attempt to predict what information will meet a user's needs based on data coming from similar users, are becoming increasingly popular as ways to combat information overload. They use a single rating as input. However, the multi-criteria based CF presents a possibility to provide accurate recommendations by considering the user preferences in multiple aspects. This research proposes a new recommendation method using Adaptive Neuro Fuzzy Inference System (ANFIS) and Fuzzy Self-Organizing Map (SOM) for accuracy improvement of multi-criteria recommender systems. We also apply Principal Component Analysis (PCA) for dimensionality reduction and to address multi-collinearity induced from the interdependencies among criteria in multi-criteria CF datasets. Experimental results on Yahoo! Movies and TripAdvisor (users' reviews on hotels) datasets demonstrated that the proposed method significantly improves recommendation accuracy of multi-criteria CF.

Keywords: Multi-criteria recommender systems, Accuracy, Fuzzy SOM, Neuro-Fuzzy, Collaborative Filtering

1. Introduction

Recommender Systems support the online customer in his/her decision making and buying process (Rashidi et al., 2015; Nilashi et al., 2016b; Bagherifard et al., 2013; Nilashi et al., 2013; Nilashi et al., 2014a; Nilashi et al., 2014b; Vahid et al., 2016). Recommender systems based on Collaborative Filtering (CF) are particularly popular and used by large online retailers (Nilashi et al., 2015a; Nilashi et al., 2015b; Farokhi et al., 2016). CF algorithms can be divided into two categories: memory-based algorithms and model based algorithms (Nilashi et al., 2013; Adomavicius and Tuzhilin, 2005).

One of the main problems in the recommender systems especially CF is known as the sparsity problem. Thus, these approaches make poor computation similarity when rating information is insufficient and with considering this problem, system produces the poor recommendation (Park and Chang, 2009). Furthermore, memory based CF approaches suffer from the scalability problem. Therefore, scaling up these system on real datasets is one of the main challenges that many methods have been developed to overcome it (Tsai and Hung, 2012).

Compared with memory based algorithms, model-based algorithms usually scale better in terms of their resource requirements (memory and computing time) and do not

require keeping actual user profiles for predictions (Georgiou and Tsapatsoulis, 2010). Model-based methods, such as Bayesian networks and clustering models (Bilge and Polat, 2013), address the problem from a probabilistic perspective to find the best item for a given user profile, and need only keep the resulting model in memory while the algorithm is running. Model-based CF, for example the work of Breese et al. (1998), can often offer significant advantages over memory-based algorithms in terms of efficiency but have not offered the same level of accuracy. It adopts an eager learning strategy, taking a probabilistic approach to predicting or recommending content, where a model of the data, i.e. the users, items and their ratings for those items, is pre-computed (Rennie and Srebro, 2005). Indeed Breese et al. (1998) found that their model-based algorithms were four times faster than their memory-based algorithms at generating recommendations in terms of runtime.

Multi-criteria based CF presents a possibility to provide accurate recommendations by considering the user preferences in multiple aspects. Adomavicius and Kwon (2007) developed a number of basic strategies to exploit multi-criteria ratings for improving the predictive accuracy of a recommender in terms of typical information retrieval measures. Later on, a number of additional techniques to leverage the detailed ratings in the recommendation process