A recommender system for tourism industry using cluster ensemble and prediction machine learning techniques

Mehrbakhsh Nilashi [a,b,*], Karamollah Bagherifard [c], Mohsen Rahmani [c,*], Vahid Rafe [c]

[a] Faculty of Computing, Universiti Teknologi Malaysia (UTM), 81310 Skudai, Johor, Malaysia
[b] Department of Computer Engineering, Lahijan Branch, Islamic Azad University, Lahijan, Iran
[c] Department of Computer Engineering, Faculty of Engineering, Arak University, Arak 38156-8-8349, Iran

ARTICLE INFO

ABSTRACT

Recommender systems have emerged in the e-commerce domain and are developed to actively recommend the right items to online users. Traditional Collaborative Filtering (CF) recommender systems recommend the items to users based on their single-rating feedback which are used to match similar users. In multi-criteria CF recommender systems, however, multi-criteria ratings are used instead of single-rating feedback which can significantly improve the accuracy of traditional CF algorithms. These systems have been successfully implemented in Tourism domain. In this paper, we propose a new recommendation method based on multi-criteria CF to enhance the predictive accuracy of recommender systems in tourism domain using clustering, dimensionality reduction and prediction methods. We use Adaptive Neuro-Fuzzy Inference Systems (ANFIS) and Support Vector Regression (SVR) as prediction techniques, Principal Component Analysis (PCA) as a dimensionality reduction technique and Self-Organizing Map (SOM) and Expectation Maximization (EM) as two well-known clustering techniques. To improve the recommendation accuracy of proposed multi-criteria CF, a cluster ensembles approach, Hypergraph Partitioning Algorithm (HGPA), is applied on SOM and EM clustering results. We evaluate the accuracy of recommendation method on TripAdvisor dataset. Our experiments confirm that cluster ensembles can provide better predictive accuracy for the proposed recommendation method in relation to the methods which solely rely on single clustering techniques.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Artificial Intelligence (AI) approaches are appearing at the forefront of research in information retrieval and information filtering systems. Recommender systems are a good example of one such AI approach. Such systems have been developed to actively recommend relevant information to users, typically without the need for an explicit search query. They have emerged in the e-commerce domain and are one way to address this issue. Based on the needs of individuals, recommenders assist them in finding the right items (Bordogna & Pasi, 2010; Mettouris & Papadopoulos, 2013; Nilashi, Jannach, bin Ibrahim, & Ithnin, 2015; Shambour, Hourani, & Fraihat, 2016). Collaborative Filtering (CF) techniques in recommender systems are particularly popular and have been applied in many online shopping websites (Bar

ragáns-Martínez et al., 2010; Davoodi, Kianmehr, & Afsharchi, 2013).

The key to successful collaborative recommendation lies in the ability to make meaningful associations between people and their product preferences, in order to assist the end-user in future transactions. Similarities between past experiences and preferences are exploited to form neighborhood of like-minded people from which to draw recommendations or predictions for a given individual user (Nilashi, bin Ibrahim, Ithnin, & Sarmin, 2015).

Recent studies have showed that using multi-criteria ratings instead of single-rating feedback can significantly improve the accuracy of traditional CF algorithms in hotel booking platform (Jannach, Karakaya, & Gedikli, 2012). Traditional CF recommender systems recommend the items to users based on their single-ratings feedback which are used to match similar users. In multi-criteria CF recommender systems, however, multi-criteria ratings are used instead of single ratings which can significantly improve the accuracy of traditional CF algorithms. These systems have been successfully implemented in Tourism domain (Fuchs & Zanker, 2012; Nilashi, bin Ibrahim et al., 2015).

* Corresponding authors at: Faculty of Computing, Universiti Teknologi Malaysia (UTM), 81310 Skudai, Johor, Malaysia.
E-mail addresses: nilashidotnet@hotmail.com (M. Nilashi), m-rahmani@araku.ac.ir (M. Rahmani).

Classification is the problem of identifying a set of observations into several categories, basing on the training result of a subset of observations whose belonging category is known. The unsupervised learning is defined as cluster analysis. Clustering is defined as a process of putting a set of objects into several reasonable groups according to the similarity among them. Single clustering techniques have been useful and effective in implementing the multi-criteria CF algorithms for improving the recommendation predictive accuracy (Liu, Mehandjiev, & Xu, 2011; Nilashi, bin Ibrahim, & Ithnin, 2014a, 2014b). However, due to the nature of combining the strengths of multiple clustering solutions, clustering ensembles generally outperform the single clustering from several aspects including robustness, novelty, stability and confidence estimation, and parallelization and scalability (Ghaemi, Sulaiman, Ibrahim, & Mustapha, 2009).

At the moment, there is no implementation of cluster ensembles for multi-criteria CF recommendation systems. In addition, although it has been proved that cluster ensembles techniques can significantly outperform many single clustering techniques in CF context (Tsai and Hung, 2012), however, there is no research conducted to show the performance of cluster ensembles for multi-criteria CF recommender systems. Hence, in this work attempts have been made to develop a recommender system based on cluster ensembles. Accordingly, in order to enhance the prediction accuracy of multi-criteria recommender systems, we propose a new recommendation system using prediction and dimensionality reduction and clustering methods. We use Support Vector Regression (SVR) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) as prediction techniques, Principal Component Analysis (PCA) as dimensionality reduction technique and Self-Organizing Map (SOM) and Expectation Maximization (EM) as two well-known clustering techniques. To improve the recommendation accuracy of proposed multi-criteria CF, a cluster ensembles approach, Hypergraph Partitioning Algorithm (HGPA), is then applied on clustering results of SOM and EM. To evaluate the accuracy of the proposed recommendation method, we conduct several experiments on TripAdvisior dataset. In comparison with the previous studies, in our research:

- We develop a recommender system based on multi-criteria CF using cluster ensembles and prediction machine learning techniques.
- We use ANFIS and SVR for prediction task.
- We use PCA to reduce the dimensions of data and solve the multi-collinearity problem.
- We use SOM and EM for data clustering.
- We use HGPA for ensemble clustering task.

Our work is organized as follows: Section 2 presents related work on multi-criteria recommender systems. A brief introduction of the methods incorporated to the recommendation model is presented in Section 3. In Section 4, the methodology is presented. Section 5 presents results and discussion. Finally, we conclude our work in Section 6.

2. Related work

Before giving details of the techniques incorporated in the proposed method and our experimental evaluation, in this section we summarize other existing approaches of multi-criteria recommender systems.

As mentioned in the previous section, our work falls into the category of multi-criteria recommender systems. The optimization of multi-criteria decision making methods and a systematic classification of multi-criteria recommender systems can be found in Manouselis and Costopoulou (2007). In addition, an overview of research in multi-criteria recommender systems and their categories are provided by Adomavicius, Manouselis, and Kwon (2011).

In case of multi-criteria recommender systems, several studies have been conducted. Adomavicius and Kwon (2007) evaluated the regression-based method and similarity-based approach on the dataset of Yahoo!Movies platform. The results of their study showed that multi-criteria rating feedback can significantly improve the accuracy of traditional neighborhood-based approach which uses single-rating feedback. Lee and Teng (2007) extended the concept of single-rating feedback to multi-criteria rating feedback and considered the multi-criteria recommendation problem as a data query problem. Accordingly, they used data query techniques to solve the multi-criteria recommendation problem. Specifically, they applied the techniques of handling skyline queries to solve the conflicts among item ratings in terms of multiple criteria. To evaluate the method, they performed several experimental analyses on Zagat Survey as a data source. Li, Wang, and Geng (2008) used Multilinear Singular Value Decomposition (MSVD) technique in developing a multi-criteria recommender system for a mobile commerce platform. To evaluate the method, they performed several experiments on a restaurant data. The results of their study indicated that higher precision and recall values can be obtained when compared with the methods which solely rely on single-rating feedback. Zhang, Zhuang, Wu, and Zhang (2009) proposed two types of multi-criteria probabilistic latent semantic analysis algorithms. The evaluation results of their method on Yahoo!Movies dataset showed that the method outperforms single-rating item-based algorithms. Shambour and Lu (2010) proposed a hybrid trust-based recommender system based on multi-criteria ratings and the techniques of trust-based filtering to solve sparsity and cold start user problems. The purpose of proposing the recommender system was to handle personalized recommendations in Governments to Business (G2B) e-services. Liu et al. (2011) developed three multi-criteria recommendation methods based on the clustering of users. They used linear least squares regression to determine the importance of the criteria. They evaluated the methods on TripAdvisior dataset and measured the accuracy using the Mean Absolute Error (MAE). Shambour and Lu (2011) developed a hybrid Multi-Criteria Trust-enhanced CF (MC-TeCF) approach to alleviate the standard single-criteria user-based CF limitations. They modified the MovieLens single-rating dataset and added three rating criteria to the dataset to be used as a multi-criteria dataset with multi-criteria ratings. They used MAE and Coverage metrics for evaluating the method. The results of their study demonstrated that the method can effectively alleviate the sparsity and cold start user problems of CF algorithms. Sahoo, Krishnan, Duncan, and Callan (2011) proposed to rely on Flexible Mixture Model (FMM) in developing a method for multi-criteria recommender systems. They evaluated the method on Yahoo!Movies dataset using MAE. The results of their work confirmed that the method which incorporates the criteria ratings achieved better MAE values. Jannach et al. (2012) used SVR and feature selection to further improve the accuracy of multi-criteria recommender systems. Their work was based on user-based and item-based approaches of CF recommender systems. They evaluated the method using fivefold cross validation on the data from the Yahoo!Movies platform and the tourism domain. Premchaiswadi and Poompuang (2013) developed two hybrid multi-criteria recommendation methods, User-Item Matching (UIM) and User-Attribute-Based (UAB), to take the advantage of implicit data in improving the accuracy of multi-criteria recommender systems in the case of the one-neighbor size. They carried out experiments on the MovieLens dataset and evaluated the methods using Precision and MAE metrics. The results of their work showed that precision of the proposed methods was