Knowledge-Based Systems 75 (2015) 192–223

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Promoting the performance of vertical recommendation systems by applying new classification techniques

Ahmed I. Saleh *, Ali I. El Desouky, Shereen H. Ali

Dept. of Computer Eng. & Systems, Faculty of Engineering, Mansoura University, Egypt

ARTICLE INFO

Article history:
Received 28 April 2014
Received in revised form 30 November 2014
Accepted 1 December 2014
Available online 7 December 2014

Keywords:
Recommendation system
Classification
Naïve Bayes
KNN
Association rules
Fuzzy logic

ABSTRACT

Recommender systems (RSs) have proven to be valuable means for online users to cope with the information overload and have become one of the most powerful and popular tools in electronic commerce. RSs are software tools providing suggestions for items of interest to users; hence, they typically apply techniques and methodologies from Data Mining. The most frequently used technique is the classification as it matches the aims of RSs that basically classify items based on user's preferences. The main contribution of this paper is in the area of applying classification techniques to enhance the performance of RSs. In this paper, an Intelligent Adaptive Vertical Recommendation (IAVR) system will be introduced. IAVR recommends text documents related to a specific domain. Basically, the paper concentrates on the first phase of IAVR, which contains two modules; the first is a distiller, while the second is a multi-class classifier. The proposed distiller is employed as a binary classifier that elects documents related to the domain of interest. It is built upon a novel neuro-fuzzy system as well as a modified K Nearest Neighbors (KNN) classifier. On the other hand, the proposed multi-class classifier merges a new instance of Naïve Bayes (NB) classifier, that depends on a proposed learning technique called "accumulative learning", with association rules. Experimental results have proven the effectiveness of the proposed classifiers, which accordingly promote the overall system's recommendation accuracy.

1. Introduction

In recent years, the on-going advances of Internet and Web technologies have promoted the development of electronic commerce, online shopping, social networking and more [1]. However, searching for relevant information on the Web is still a challenge. Even if indexing methods get more efficient, search engines stay passive agents as they do not take into account either the customer's context or his aims. Corporations, on the other hand, view the Web as a potential business accelerator [2]. In order to expand their markets, enterprises have been developing new business portals to provide large amounts of products information. As a result, customers have more opportunities to choose various products that meet their needs. Those new emerging portals are known as Recommendation Systems (RSs) [3].

RSs are (online) computer-based software capable of automatically identifying appropriate choices from several alternatives, on the basis of some specified criteria. They are. RSs try to reduce customers' burden of searching for their preferred items by suggesting a ranked list of items in which customers may interest. The suggestions provided by RSs support users in various decision-making processes, such as what items to buy, what music to listen, or what news to read. Correspondingly, various recommendation techniques have been proposed and during the last decade, many of them have also been successfully deployed in commercial environments [1]. Suggestions for documents on Amazon [4], or movies on Netflix [5], are real world examples of RSs.

RSs constitute a relatively new area of research compared with other classical information system tools and techniques (e.g., databases or search engines). They emerged as an independent research area in the mid-1990s. In recent years, the interest in RSs has dramatically increased, as their development is a multidisciplinary effort, which involves experts from various fields such as; artificial intelligence [2], human computer interaction, information technology, data mining, statistics, adaptive user interfaces, decision support systems and marketing. Several RSs have been introduced; however, most of them bear in their core an algorithm that can be understood as a particular instance of a Data Mining (DM) technique. The most frequently used technique is the classification, which matches the nature and targets of RSs that basically classify a set of items based on users' preferences [6].

* Corresponding author.

http://dx.doi.org/10.1016/j.knosys.2014.12.002
0950-7051/

RSs search in collections of items and suggest the most relevant ones to users [3]. The relevancy of an item to a user can be merely expressed with a decision of whether the user likes it or not. Providing recommendations through such decisions can be approached from a machine learning perspective where classifiers are trained to offer effective solutions to this binary classification problem (i.e., like/dislike). Classifiers have played an active role in the field of recommenders. Probabilistic methods have been used in both content-based [8] and collaborative-filtering [9] recommenders as well as in hybrid solutions [10]. Furthermore, other types of classifiers such as Decision Trees (DT) [11] and Support Vector Machines (SVM) [12] have also been utilized in recommendation methods.

Due to today's information overload, modern RSs rely on automatic retrieval systems, such as search engines' crawler, to collect items that may interest their users over the web [7]. However, such automatic retrieval systems usually suffer from low precision. Many collected items are out of RS's scope. To narrate confidently, consider a recommendation system that recommends research papers in the area of "computer science", which relies on an automatic focused crawler to collect those research papers, and then stores them in the system's database. In spite of its high retrieval rate, focused crawlers still suffer from low precision. Many collected papers may be irrelevant to RS's Domain of Interest (DOI), which certainly harms the recommendation accuracy. To turn around such hurdle, it is becoming necessary to propose a recommendation distiller that elects only items in the scope of the RS's DOI before storing them in the system's database. Moreover, modern RSs should rely on efficient classification techniques to promote the recommendation process [3]. Distillation and classification; (i) simplify the recommendation task, (ii) solve the information overload as only those related items are indexed in the system's database, and accordingly, and (iii) accelerate the recommendation process.

This paper proposes an Intelligent Adaptive Vertical Recommendation (IAVR) system, which recommends text documents related to a specific domain. The recommended documents may be in many forms such as; books, research papers, and reports. IAVR consists of four layers, which are; (i) Content Analyzer Layer (CAL), (ii) Profile Learning Layer (PLL), (iii) Collaborative Layer (CL), and (iv) Matchmaking and Ranking Layer (MRL). The paper presents a detailed description about the first layer of IAVR in which a novel classification strategy has been introduced, which combines evidence from Bayes theorem and association rules. The proposed classification strategy relies on "accumulative learning"; hence, it also solves the traditional overfitting problem. As a result of its high accuracy, the proposed classification strategy promotes the system recommendation accuracy. Moreover, CAL has a domain distiller for outlier rejection. Such distiller is adjusted to elect only those documents related to RS's DOI. It is employed as a binary classifier, which is built upon a novel neuro-fuzzy system as well as a modified KNN classifier.

The remainder of this paper is organized as follows; in Section 2, we discuss the different recommendation techniques. Section 3 introduces a complete review of recent recommendation systems. It also highlights the impact of classification techniques in field of RSs. Moreover, this section provides an overview of the key classification techniques that have been successfully used to promote the recommendation accuracy of recent RSs. In Section 4, the proposed IAVR is introduced in details. Section 5 introduces the core of the paper's work. Details about the first layer of IAVR, which is the Content Analysis Layer (CAL), are illustrated in this section. The new classification techniques introduced in CAL are tested in Section 6. Finally, some conclusions and future directions are drawn in Section 7.

2. Recommendation systems

Generally, there are two main approaches to build a recommendation system, namely; (i) collaborative filtering and (ii) content based. The first, as illustrated in Fig. 1, recommends items via the choice of other similar users. Collaborative filtering approach computes the similarity between users based on their rating profile, and then recommends items that are highly rated by similar users [13]. It assumes that if user U_1 and U_2 rate x items similarly or have similar behavior, they will rate or act on other items similarly. Instead of calculating the similarity between items, a set of "nearest neighbor" users for each user whose past ratings have the strongest correlation are found. Therefore, scores for the unseen items are predicted based on a combination of the scores known from user's nearest neighbors. Content-based RSs, on the other hand, work with profiles of users that are created at the beginning of the system's operation [14]. A profile has information concerning the user's taste, which is based on how he rates the items. Generally, when creating a profile, RSs formulate a survey, to get initial information about a user in order to avoid the new-user problem [15]. During the recommendation process, the system compares items that were already positively rated by the user with items he did not rate and looks for similarities as illustrated in Fig. 2. Those items that are almost similar to the positive rated ones, will be recommended to the user. RSs based in the content are mainly used to recommend documents, Web pages, publications, jokes or news. Some examples are; Syskill & Webert, which recommend Web pages [16] or PTV [17] which recommends TV programs.

However, as researchers have studied different recommendation technologies, many have suggested that no single technology works well for all situations. Thus, hybrid RSs, as illustrated in

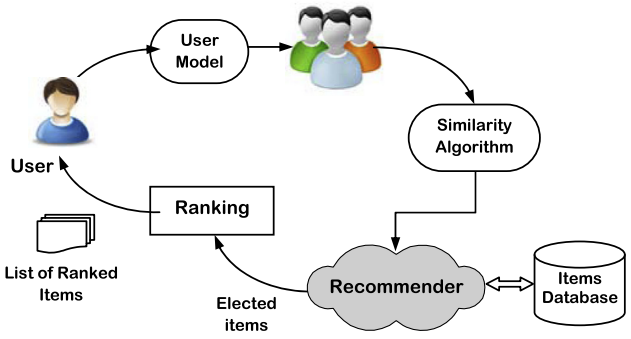

Fig. 1. Execution flow for the collaborative recommendation system.

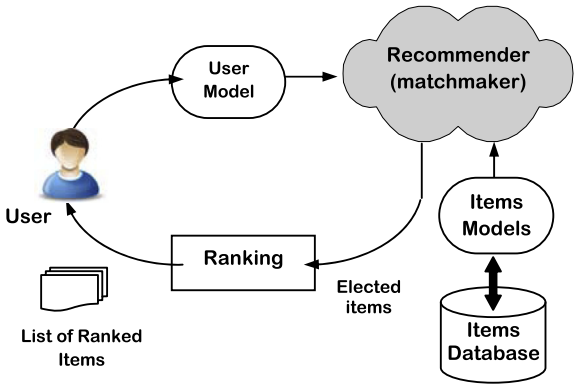

Fig. 2. Execution flow for the content based recommendation system.