Expert Systems with Applications 36 (2009) 11480–11488

Contents lists available at ScienceDirect

Expert Systems with Applications

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

A hybrid recommendation technique based on product category attributes

Amir Albadvi [a], Mohammad Shahbazi [b,*]

[a] Associate Professor of Information Systems, Industrial Engineering Department, Faculty of Engineering, Tarbiat Modares University, Tehran, Iran
[b] Information Technology, Industrial Engineering Department, Faculty of Engineering, Tarbiat Modares University, Tehran, Iran

ARTICLE INFO

Keywords:
Recommender system
Product taxonomy
Product category attribute
Customer preferences
Hybrid recommendation

ABSTRACT

Recommender systems are powerful tools that allow companies to present personalized offers to their customers and defined as a system which recommends an appropriate product or service after learning the customers' preferences and desires. Extracting users' preferences through their buying behavior and history of purchased products is the most important element of such systems. Due to users' unlimited and unpredictable desires, identifying their preferences is very complicated process. In most researches, less attention has been paid to user's preferences varieties in different product categories. This may decrease quality of recommended items. In this paper, we introduced a technique of recommendation in the context of online retail store which extracts user preferences in each product category separately and provides more personalized recommendations through employing product taxonomy, attributes of product categories, web usage mining and combination of two well-known filtering methods: collaborative and content-based filtering. Experimental results show that proposed technique improves quality, as compared to similar approaches.

1. Introduction

Nowadays, with the exponential increase in the amount of information available on the Internet, users are faced with serious information overload problem (Boucher-Ryan & Bridge, 2006; Cho & Kim, 2004; Lihua, Lu, Jing, & Zongyong, 2005; Sarwar, Karypis, Konstan, & Riedl, 2001; Semeraro, Lops, & Degemmis, 2005). In the online retail stores, customers are encountered to so many different and multi-category products and services. Therefore, they have to spend much time and effort on information search and selection before they can find out what they want. To address this problem, many solutions have been proposed. Recommender system is one such case (Cho & Kim, 2004; Karypis, 2001; Lihua et al., 2005; Sarwar et al., 2001; Semeraro et al., 2005). Recommender system represents the recent effort on information filtering and personalized service (Lihua et al., 2005) and is defined as the system, which recommends an appropriate product or service after learning the customers' preferences and desires (Choi, Kang, & Jeon, 2006). These systems also offer several advantages, including increasing the probability of cross-selling, establishing customer loyalty, and fulfilling customer needs by presenting products of possible interest to them (Senecal & Nantel, 2004; Shih & Liu, 2008). The recent commercial success of recommender systems has been demonstrated in many online stores including Amazon.com, CDNow.com, Barnes&Noble.com, and MovieFinder.com. Based on how recommendations are made, recommender systems are usually classified into content-based filtering (CBF) (Adomavicius & Tuzhilin, 2005; Cheung, Kwok, & Law, 2003; Cho & Kim, 2004; Hung, 2005; Leung, Chan, & Chung, 2006; Shih & Liu, 2008; Weng & Liu, 2004), collaborative filtering (CF) (Adomavicius & Tuzhilin, 2005; Boucher-Ryan & Bridge, 2006; Cheung et al., 2003; Cho & Kim, 2004; Karypis, 2001; Leung et al., 2006; Liua et al., 2005; Sarwar et al., 2001; Shih & Liu, 2008; Weng & Liu, 2004) and hybrid approaches (Adomavicius & Tuzhilin, 2005; Burke, 2002; Cho & Kim, 2004; Choi et al., 2006; Kim, Li, Park, Kim, & Kim, 2006; Semeraro et al., 2005; Shih & Liu, 2008).

In real environment, users have unlimited and unpredictable desires and their preferences may vary within different product categories (Choi et al., 2006; Leung et al., 2006; Lihua et al., 2005). For example a user may be interested in buying inexpensive and pocket-size books, while this user may be interested to buy expensive and big toys. In marketing research literature, there are two major families of specialized marketing research methods, perceptual mapping (Diez, Coz, Luaces, & Bahamonde, 2008; Elmore, Heymann, Johnson, & Hewett, 1999; Kleija et al., 2003; Kuhfeld, 2005) and conjoint analysis (Kuhfeld, 2005; Oppewal & Louviere, 2000; Vermeulen, Goos, & Vandebroek, 2008) to evaluate customer preferences on firm's products or services. Perceptual mapping methods produce plots that display product positioning, product preferences, and differences between customers in their product preferences. Conjoint analysis is used to investigate how consumers trade off product attributes when making a purchasing decision.

* Corresponding author. Tel.: +98 09126121921.
E-mail address: shahbazi_mo@yahoo.com (M. Shahbazi).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.03.046

In this study we propose a six-phase recommendation technique in the context of online retail store, called hybrid recommendation based on product category attributes (HRPCA), to handle the customer preferences varieties in different product categories. HRPCA employs theoretical framework of perceptual mapping and conjoint methods to determine the key attributes of product categories and their levels and values. These attributes are used to create customer profile and address the feature selection problem (Cheung et al., 2003; Hung, 2005) in CBF approach. According to sequential combination model in hybrid recommender systems (Kim et al., 2006), customer profile is created by CBF approach and consequently, CF is applied to improve recommendation accuracy. Web usage mining (Adomavicius & Tuzhilin, 2005; Cho & Kim, 2004) is also employed to analyze customers' shopping behaviors on the Web and collects their implicit ratings on the attributes. As dimensionally reduction technique, before applying the algorithms, we employed product taxonomy (Cho & Kim, 2004; Hung, 2005; Leung et al., 2006) to reduce dimensionality reduction in product space.

2. Related works

In this section we review some prerequisite concepts that are used to our research.

2.1. Recommendation methods

There are three major types of recommender systems: content-based, social or collaborative filtering based and hybrid. Content-based filtering (CBF) provides recommendations by matching user profiles (e.g., preferences) with content features (e.g., product attributes) (Shih & Liu, 2008). Each user profile is derived by analyzing the content features of products purchased by the user. The social or collaborative filtering (CF) utilizes preference ratings given by various customers to determine recommendations to a target customer based on the opinions of other similar customers. A typical CF method employs K-nearest neighbors approach to derive top-N recommendations (KNN-based CF method) (Adomavicius & Tuzhilin, 2005; Cheung et al., 2003; Cho & Kim, 2004; Leung et al., 2006; Shih & Liu, 2008).

These approaches have some drawbacks. In CBF approach, when the system can only recommend items that score highly against a user's profile, the user is limited to being recommended items that are similar to those already rated. Hence, unique or different products may not be able to be presented to customers (Adomavicius & Tuzhilin, 2005; Weng & Liu, 2004). Another drawback is related to feature selection problem (Cheung et al., 2003; Hung, 2005). Existing CBF systems rely on preprocessing steps that select a manageable set of "important" features from product attributes. This, however, depends on personal experience and affects quality of recommendation. CF approach has two major limitations. (1) Scalability: the system must perform mass comparisons prior to finding similar communities to target customers (Adomavicius & Tuzhilin, 2005; Cho & Kim, 2004; Sarwar et al., 2001; Weng & Liu, 2004); (2) Sparsity: Most similarity measures used in CF work properly only when there exists an acceptable level of ratings across customers in common. Sparsity in ratings makes the formation of neighborhood inaccurate, thereby resulting in poor recommendation (Adomavicius & Tuzhilin, 2005; Cho & Kim, 2004; Sarwar et al., 2001; Weng & Liu, 2004).

To address sparsity problem in CF algorithms, web usage mining (WUM) can be employed as an implicit ratings approach (Adomavicius & Tuzhilin, 2005; Cho & Kim, 2004). WUM is the process of applying data mining techniques to the discovery of behavior patterns based on web log data (Cho & Kim, 2004). In this technique, customers' behaviors on the web site environment are inspected and their implicit ratings are collected. To handle the scalability problems in CF, several approaches have been developed and the most important one is dimensionality reduction (Cho & Kim, 2004; Karypis, 2001). Singular value decomposition (SVD), Principal Components Analysis (PCA) and Latent Semantic Indexing (LSI) are frequently used in dimensionality reduction techniques (Cho & Kim, 2004; Karypis, 2001); product taxonomy (PT) technique is also used to reduce dimensionality of the customer and product space (Cho & Kim, 2004; Hung, 2005; Leung et al., 2006). As domain specific knowledge provided by marketing managers or domain experts, PT is practically represented as a tree and categorizes a set of products at a low level into a more general product at a higher level (Cho & Kim, 2004).

Several researchers are exploring hybrid methods of combining CF and CBF methods, which helps to avoid certain limitations of CBF and CF systems (Adomavicius & Tuzhilin, 2005; Burke, 2002; Cho & Kim, 2004; Choi et al., 2006; Kim et al., 2006; Shih & Liu, 2008). Different ways to combine collaborative and content-based methods into a hybrid recommender system can be classified as follows: weighted combination model, which implements collaborative and content-based methods separately and combines their predictions with giving adjustable weight to both recommendation (Adomavicius & Tuzhilin, 2005; Burke, 2002; Kim et al., 2006); mixed combination model, in which recommendations from the two techniques are combined together in the final recommendation list (Burke, 2002; Kim et al., 2006). Sequential combination model where initially user profiles are constructed by a CBF algorithm based on the items; then a collaborative algorithm is applied to make predictions based on those user profiles (Kim et al., 2006).

In this study, user profile is created by implicit ratings of user to product attributes. According to sequential combination model in hybrid recommender systems (Kim et al., 2006), customer profile is created by CBF approach and consequently, CF is applied to improve recommendation accuracy. Web usage mining (Adomavicius & Tuzhilin, 2005; Cho & Kim, 2004) is employed to analyze customers' shopping behaviors on the web and collects their implicit ratings on the attributes.

2.2. Customer preferences evaluations

Learning preferences is a useful task in application fields such as collaborative filtering, information retrieval, adaptive assistants or analysis of sensory data provided by expert panels (Diez et al., 2008). In marketing literature, marketing research is an area of applied data analysis whose purpose is to support marketing decision making using both standard data analysis methods, such as descriptive statistics and cross tabulations, and more specialized marketing research methods (Kuhfeld, 2005). One of the major goals in marketing research is to evaluate customer preferences on firm's products or services. There are two major families of specialized marketing research methods, perceptual mapping (Diez et al., 2008; Elmore et al., 1999; Kleija et al., 2003; Kuhfeld, 2005) and conjoint analysis (Kuhfeld, 2005; Oppewal & Louviere, 2000; Vermeulen et al., 2008).

Perceptual mapping methods produce plots that display product positioning, product preferences, and differences between customers in their product preferences and includes correspondence analysis (CA) (Kuhfeld, 2005), multiple correspondence analysis (MCA) (Kleija et al., 2003; Kuhfeld, 2005), preference mapping (PREFMAP) (Kuhfeld, 2005), multidimensional preference analysis (MDPREF) (Diez et al., 2008, 2005), and multidimensional scaling (MDS) (Chen, Tzeng, & Ding, 2008; Kuhfeld, 2005). Preference mapping provides valuable information about each consumer's response in a visual format and provides a clear presentation of the relationship among the products and the individual differences