



# Research on recommender system based on ontology and genetic algorithm

Gang Lv, Chunling Hu, Shengbing Chen

Key Laboratory of Network and Intelligent Information Processing, Hefei University, Hefei 230609, China



## ARTICLE INFO

### Article history:

Received 24 April 2015

Received in revised form

27 August 2015

Accepted 25 September 2015

Available online 30 November 2015

### Keywords:

Recommendation

Ontology

Semantic web

Genetic algorithm

Associated data

## ABSTRACT

Allied to the extensive use of online shopping, product recommendation on websites is vitally important for E-commerce. Consequently, collaborative filtering and content-based filtering have been widely used in recommendation systems on E-commerce websites. However, these filtering methods have many problems, such as cold start, prejudiced ratings and inaccurate suggestions. To generate valid and accurate suggestions, researchers have proposed integrating semantic features of the data in an ontology into the recommendation process. However, most existing studies only include the type and features of a product without considering the relational characteristics thereof. Given that the relational characteristics can provide much useful information during the recommendation process, we have designed an easily realizable recommendation system framework based on relational data by integrating the relational data in the domain ontology and applying a genetic algorithm to process the recommendation. Experimental results show that there are obvious improvements in the methods for dealing with sparsity and cold start problems as well as the accuracy and timeliness of recommendations.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

The recent rise of countless online shops and their intense development have forced many E-commerce businesses to provide suitable item suggestions for their customers based on individual preferences and demand. Although the explosive growth in the number of web images has brought about many technical challenges for image search, it has also provided almost unlimited training data and new ideas for solving various computer vision problems [1,2]. To meet this objective, collaborative filtering and content-based filtering have been widely used. However, both methods have their disadvantages [3]. Collaborative filtering is restricted by cold start and data sparsity problems, whereas content-based filtering suffers from limited diversity and invalid recommendations for complicated objects. To mitigate these shortcomings, various researchers have proposed integrated strategies combining content-based filtering and writing-based filtering [4,5].

According to several recent studies, by solving the innate problems of collaborative and content-based filtering using a domain ontology, more accurate suggestions can be obtained. The features of an object can also be used at a much deeper level through an ontology to realize successful recommendations of complicated objects [6,7]. A recommendation system based on an ontology can also solve the cold start problem. Most researchers believe that product samples based only on an ontology have the same structural memory. However, associated data have a more complicated

structure [8,9]. For instance, to ascertain the possibility of an item being purchased by a user, not only the features of the item itself, but also the features of the manufacturer should be considered. In all these data, different objects are associated with different samples in terms of quantity and property [10,11]. As an abstract sample of the domain ontology illustrated in Fig. 1, we define C1 as the recommended product class, which is used in this study as the suggested target class. The class itself has two attributes, A1 and A2, and a subclass C2, which has its own two attributes, A3 and A4. For example, in a book recommendation system, if C1 was the suggested book class, then A1 and A2 could denote the type and publication date, while C2 could be the author of book who has his/her own attributes, A3 and A4, representing former honors and awards he/she has received.

Currently, most research on recommendation systems is based on methods for collaborative and content-based filtering, with the users of the collaborative filtering method indicated by a vector of similarity [12]. Technology used to predict product is based on the views of other users of the user's preferences. The views can be explicitly as a rating score from the user, or by using some implicit measures, from purchase records as the time diary. Content-based information retrieval method is applied first field Recommended text document, its content and user profiles were compared. Right weight extracted from the document out of the word to be added to the user profile corresponding to the word weight, if the user is interested page. The main disadvantage of this approach is the lack of mechanisms to manage web object motion pictures, images,

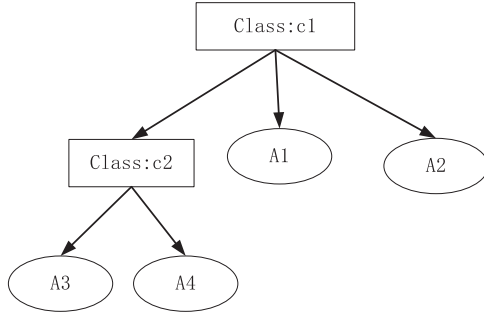


Fig. 1. Abstract example of an ontology.

music and so on. Moreover, it is difficult to deal with large numbers of obtaining product content words. This method is by replacing the word document by product attributes to any kind of product extensions. However, if the number of users is very large, the results of this kind of online recommendation system are not ideal. To mitigate the shortcomings, the relations between products are considered an important factor. As for content-based filtering, it only takes into account the preferences of certain users. A text-based recommendation method where the content, including each theme, must be established, is proposed in [13]. When a user evaluates documents relating to these themes, the system recommends related documents that have been evaluated using a Bayes classifier. Several researchers have suggested that, to resolve the problems of collaborative and content-based filtering, namely, cold start and limited diversity, a combined system must be adopted. The study in [14] advocates a Web recommendation method based on a domain ontology where the time spent by a user browsing a particular page is regarded as the weight of the user's interest in the respective page. Nevertheless, most of these researchers regarded the domain ontology as a simple structure containing only a single class, and neglected the associated features and attributive characters of the ontology class.

Genetic algorithms (GAs), which have been widely used to solve optimization problems, are a kind of heuristic search method that imitates the process of natural selection, including selection, crossover and mutation [15]. The algorithm uses parameters to denote chromosomes with each parameter encoded as a genetic variable. Therefore, a parameter is optimized together with the chromosome containing many genes. The population is thus, a structure established with a certain number of genes together with the variables associated with these genes. A fitness function is used to evaluate the goodness of an individual solution and the next generation is formed by the progeny cells of the chromosomes generated during the crossover process [16]. The chromosome with the highest fitness is selected as the parent to generate offspring in the next generation. This process is repeated until a certain fitness has been achieved or a particular number of offspring have been generated.

## 2. Recommendation method using an ontology and genetic algorithm

To improve the quality of recommendations, we propose applying a traditional GA and ontology to the recommendation process. Concepts and background information required in this research are introduced in this section.

Traditional distance measurement methods, such as Euclidean distance, refer to the calculation of the closeness of two vector spaces and are always used to measure the distance between two vectors in an N-dimensional space [17]. Given that samples in a dataset have different sized association tables, and each character

Table 1  
Table of class "O".

Sample set of class "O"	Attribute set		
	Attribute 1	Attribute 2	Attribute 3
$O_1$	$At_1$	$At_2$	$SC_1$
$O_2$	$At_1$	$At_2$	$SC_2$

Table 2  
Table of class "SC".

Sample set of class "SC"	Attribute set	
	Attribute 1	Attribute 2
$SC_1$	$SAt_1$	$SAt_2$
$SC_2$	$SAt_3$	$SAt_4$

has a different value, traditional methods for measuring distance cannot be applied to such complicated relational data structures. By extending the measurement method of distance and dealing with several tables directly, this study regards each column of a table as an input for the distance measurement. Hence, when calculating related distances, the influence of all the attributes is the same.

In the sample set of Class "O" shown in Table 1, each attribute set has three attribute values, with the third attribute "SC" having its own class (as shown in Table 2). The study in [18] formula for calculating the distance between samples  $O_1$  and  $O_2$  in class O is given as:

$$dist(O_1, O_2) = \frac{\sum_{i=1}^n dist(O_1.At(i), O_2.At(i)) + \sum_{i=1}^k dist(SC_1.At(i), SC_2.At(i))}{n+k-1} \quad (1)$$

where  $n$  and  $k$  denote the numbers of attributes in class "O" and class "SC", respectively.

Another method calculates first the distance between  $SC_1$  and  $SC_2$ , then the attribute distance without referring to other tables, and finally, the average of the values, using recursive computation to obtain the distance between samples  $O_1$  and  $O_2$  until reaching the maximum depth initially set. The actual calculation formula is given as:

$$dist(O_1, O_2) = \frac{\sum_{i=1}^n dist(O_1.At(i), O_2.At(i)) + \frac{\sum_{i=1}^k dist(SC_1.At(i), SC_2.At(i))}{k}}{n} \quad (2)$$

In this study, we propose a new framework for the recommendation system based on the associated semantics by which the features of a recommended product with corresponding weights can be described to transform them into characters in the relationship method. The weight of the related feature is obtained through the GA.

As shown in Fig. 2, the actual framework includes five steps: The first of these splits the log file of an E-commerce website into individual user sessions, containing the specific content or items in the respective user's shopping cart. In the second step, the attribute set of the product is expanded to ensure that it contains the attributes of the product subclass. The third step optimizes the different feature weights in the set using a genetic algorithm. In the fourth step, the attributes provided on the E-commerce website are clustered using the coefficient calculated in step 3 to narrow the search space. The final step recommends the cluster center closest to the product in which the user was interested on

Download English Version:

<https://daneshyari.com/en/article/405898>

Download Persian Version:

<https://daneshyari.com/article/405898>

[Daneshyari.com](https://daneshyari.com)