Using formal concept analysis and collaborative filtering to recommendation system in e-commerce

QiPing Zheng, Yi Wang

Abstract

Collaborative filtering is the most studied personalized recommendation technology. Collaborative filtering recommendation systems are used widely in e-commerce website and it is by far the most successful information filtering technologies. Formal concept analysis is a tool for data analysis or approach, especially given the information can be analyzed and processed. This paper presents to construct recommendation system of e-commerce by formal concept analysis and collaborative filtering to make up for these the shortcomings of existing e-commerce recommendation system. The experimental results show that this method can effectively improve the performance of the recommendation system.

Keywords: Collaborative filtering, Formal concept analysis, recommendation system

1. Introduction

In the increasingly fierce competitive environment, e-commerce recommendation system can effectively retain customers, increase sales of e-commerce systems an effective means. Recommended introduction of e-commerce technology and systems research status, analysis of existing technologies and systems exist in the quality, timeliness and scale of the problem [1]. E-commerce based on collaborative filtering recommendation system does not analyze the similarity between the goods, but to learn the history of the target users and the similarity between users buying behavior, and thus according to the user's purchase history is similar to recommendations generated results. Purely based on collaborative filtering recommendation system does not need e-commerce features of the description of goods, it is the user to learn the similarities between buying behavior, rather than commodity-dependent features, so it can be recommended from the looks on the surface characteristics of different but there are actually very relevant goods.

Currently, most e-commerce sites simply try to provide information to the user, without taking into account the needs of users, resulting in e-commerce site provides users the information loss. When users waste a lot of time to download on their own useless information, and user preferences to find the information themselves, their usage of e-commerce site will be reduced. In addition, the current e-commerce model there is also a common problem: lack of interaction between enterprises and customers [2]. From the customer considerations, the face of the website provides a large number of commodities business, the customer if you want to find the goods they want, it takes a lot of time to exclude the information from many of their business information in order to find out with their related information, from business to consider, because they do not know the purpose and each customer buying preferences, we must invest a great price, and risk aversion to risking the customer to provide some product information, even if some of the fixed customer base, we can not provide highly targeted and personalized service.

However, the existing e-commerce recommendation system itself has shortcomings, mainly in two aspects: First, the data sparseness problem is the evaluation of the user when there are fewer goods to buy, recommended quality will be reduced. Second, the cold start problem, into the "new goods problem" and the "new user problem", the performance when adding a new product without any score data, as well as a new user does not have any goods to score on the resulting system can not be recommended to the user recommend.

Knowledge-based recommendation is important and difficult is the need to acquire knowledge, it involves three kinds of knowledge: a catalog of knowledge, that knowledge of recommended products and their characteristics; b. Functional knowledge, the recommended system should be able to meet
customer needs and possible characteristics of the demand for more goods; c. customer knowledge. Knowledge and capabilities which are the customer knowledge acquisition are the study of difficult problems. Knowledge-based reasoning is recommended with the fundamental principles to guide the user through the user interface features clear requirements for the product in order to gain knowledge of customer needs, recommend the system according to user needs access to knowledge and knowledge of the product catalog knowledge, the use of functional knowledge reasoning to identify the products to meet user needs and recommend to the user.

This article presents to construct recommendation system of e-commerce by formal concept analysis and collaborative filtering to make up for these the shortcomings of existing e-commerce recommendation system. Collaborative filtering is the most studied personalized recommendation technology. It generally uses the nearest neighbor technique, the use of information to calculate the user's historical preferences of the distance between the user and then use the target user's nearest neighbor users for goods evaluation to predict the value of the weighted evaluation of the target user's preference for a particular commodity, the system which according to this a preference to the target users of the recommended. The concept of formal concept analysis to the data grid in the form of organically organized, there are a lot of research work explores the concept lattice from the grid nodes and the hierarchy to extract rules. Experimental studies have shown a variety of rules based on concept lattice extraction system is effective, formal concept analysis is very suitable for rule-based knowledge discovery.

2. The research on collaborative filtering recommendation

Collaborative filtering recommendation system can be said from the perspective of the user accordingly recommended, and is automatic, the user access the system from the recommended purchase or browsing behavior patterns obtained implicitly, without user effort to find recommended information to their own interests, such as completing a number of survey forms, etc.. First, the dictionary can rely on ontology learning and knowledge base for ontology construction, following a brief introduction of these two methods. Collaborative filtering has the following advantages [3].

1) Machines can be difficult to automatically filter content analysis of information, such as art, music, etc.

2) Sharing the experiences of others, to avoid the content analysis of incomplete and inaccurate, and can be based on a number of complex, elusive concepts (such as information quality, personal taste) filter.

3) Here has recommended the new information. Content can be found completely similar information, the user of the content of information in advance is recommended unexpected. This is the collaborative filtering and content-based filtering of a large difference between content-based filtering recommendations many of which are already familiar with the contents of the user, and collaborative filtering can be found in the user of potential interest, but their preferences have not yet found.

4) Knowledge-based recommendation is to some extent can be seen as a kind of reasoning (Inference) technology; it is not based on user needs and preferences based on recommendations. Knowledge-based methods because they use different functional knowledge are apparently different [4]. Effectiveness of knowledge (Functional Knowledge) is an entry on how to meet a particular user's knowledge, so they can explain the relationship between needs and recommended, so the user can be any data to support the inference of knowledge structure, it can be a user has been standardized query, it can be a more detailed representation of user needs.

Collaborative filtering systems is to use this idea to the Network Information Service (Information recommended), based on other users of the evaluation of certain information to be recommended to a user. Typically, system selection and the specified users with similar interests of users as a reference object. And how to define users and how to select the reference similarity user base is the focus of collaborative filtering algorithm, as shown by equation 1.

\[ W(Y, k) = \sum_{j \in Y} w_j + \sum_{r=1}^{k-1} w_r \]  \hspace{1cm} (1)

Of course, collaborative filtering has some deficiencies, which is determined by the nature of collaborative filtering: To obtain satisfactory results, must be accurate user information. In general, this
requires the establishment of user information, with its large database, which is difficult to do precisely the same time it also makes applications collaborative filtering technology is relatively narrow (almost concentrated in the entertainment: music, movies), in the broader area (such as content-based filtering has been quite successful in the relevant areas of text) application is still not enough.

User access to information on the main access to user information given (broad sense, narrow sense can refer business to provide goods, etc.) and evaluation. Evaluation is divided into explicit and implicit evaluation of two kinds of evaluation.

In the collaborative filtering algorithm model, the global numerical algorithm in time using the latest information for the user to produce a relatively accurate prediction of the degree of user interest or recommended, but the face of the growing number of users, the dramatic increase in the amount of data, algorithms, scalability issues (expanding the system to adapt to the problem) as a constraint to the implementation of the recommendation system an important factor. Although model-based algorithm, the global numerical algorithms for the development of the model to save the cost of training time, but used to identify the "nearest neighbor" algorithm for computing the amount of users and items as increased greatly increased, for the millions of number, the algorithm will usually experience severe scalability bottleneck. The issues are not directly influence the recommendation based on collaborative filtering technology to provide users with real-time recommendations to solve the problem, but the better recommendation system real-time, higher accuracy, the system will only be accepted by users, as shown by equation 2.

\[
X = \begin{bmatrix}
  x_{11} & x_{12} & \cdots & x_{1M} \\
  x_{21} & x_{22} & \cdots & x_{2M} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{K1} & x_{K2} & \cdots & x_{KM}
\end{bmatrix}
\]

As shown in equation 2, a K × M users - key evaluation matrix X, K-line on behalf of K users, M out on behalf of M items, \(x_{ij}\) is the i-th user on the j-term evaluation of values (or assessed value), and the contents of the evaluation value, and if one is e-commerce products, the user orders the product or not, for example, 1 for order, 0 means no order; or that degree of user interest in how it high, can be different values represent different levels, such as 1-5 and so on. If the item is a Web document, then it can view or not, or degree of user interest in it.

\[
\left( \sum_{j, i \in X} w_j \right) \times \text{Support}(X)
\]

Collaborative filtering algorithm can easily provide for the thousands of good recommendations, but for e-commerce sites often need to provide recommendations to millions of users, which on the one hand the need to improve response time requirements, to provide users real-time to be recommended; the other hand should also take into account the storage space requirements, recommended to minimize the burden of system operation. User-based collaborative filtering algorithm first analyzes the user - item matrix, has been the relationship between different users (similarity), and then computing the use of these relationships to provide recommendations to the user [5]. Shown in Figure 1, according to the similarity with the test users sort the results.

\[
P_{ai} = v_a + k \sum_{b \in \text{neighborhood}(a, T)} r(a, b) (v_{b, i} - \overline{v_b})
\]

Values in the global collaborative filtering algorithm, similar to the user through the user focus on the evaluation of a particular value, and the related series of weights to the target user's interest degree to predict. Assuming that the target user. On the predictive value of an item i, \(P_a\), then as shown by equation 5.

\[
\Phi_R = u^T B^T Bu + 2(R - Ax(0))^T Bu + (R - Ax(0))^T (R - Ax(0))
\]
Collaborative filtering algorithm model chart

Global numerical algorithm is mainly based on correlation methods, the full information (user evaluation matrix is only a small amount of vacant value) in the case, have quite good results, however, in reality the user information is often inadequate, can be used to design more for more information, and more effective use of information algorithm[6]. Such algorithm is from the perspective of collaborative filtering probability, given all the information conditions, calculated (estimated) evaluation of the target user expectations of the target item. Suppose the user evaluation value is an integer from 0 to m, for the target user a, forecast evaluation of his term j, as is shown by equation 6.

$$p_{a,j} = E(r_{a,j}) = \sum_{i=0}^{m} p(r_{a,j} = i | A)i$$

Algorithm 1: collaborative filtering algorithm

Input: Expression refers to the conditional probability matrix A is known under the conditions of evaluation, a user evaluation of information j is the probability of i to be recommended item ID: IID, User ID: UID.

Output: Analysis of the similarity between users, the formation of nearest neighbors; finally, have recommended items and results.

1. Define a distance criterion used to measure the similarity between users;
2. $sup_h(X) = \frac{1}{k} \times \sum_{i,h \in X} h \times sup(X)$
3. For (i=1;i<Size; i++)
4. $L1=\{c \in C1 | WSup(c) \geq wminsup\}$
5. For each transaction in D do
6. According to $c=i+j$ the similarity between users similar to user-selected set of
7. (SC)=Counting(D,W);
8. while (|Ck|<k)
9. FOR i ← 0 to r1 + c1 - 3 DO;
10. UpdateNode(L(K1) ; Cp ; C2);
11. Rules (recommended items, L);

The algorithm has the advantage of the following: 1) fully into account the user's evaluation of the value of neighbors is inconsistent between each other, more obvious changes in the user evaluation of the value of that part of the user; 2) to avoid most of the user evaluation value due to the error caused by over-centralized because these values as quantity often than other users, especially the nearest neighbor is a critical nearest neighbor users have a greater impact, leading to bias; 3) For the new model in time according to user preferences to determine focus on adding users to the nearest neighbor.

In this paper, knowledge of e-commerce recommendation system was introduced, the overall e-commerce recommendation system has knowledge and understanding, especially for e-commerce recommendation system, the method used [7]. The recommended system requirements with increasing levels of functionality, and its implementation techniques are also facing serious challenges. In a certain sense, the recommended system faces two challenges there was a contradiction between the
system to improve the algorithm's scalability and response time, quality is bound to lose. Therefore, how to coordinate these two requirements, the recommended system is not only useful and practical, collaborative filtering technology to achieve an important factor to consider.

3. Formal concept analysis technology

Formal concept analysis is a tool for data analysis or approach, especially given the information can be analyzed and processed. The data should be meaningful from the human mind can understand units - taking the concept of the formation of formal elements. Formalization is that the data is handled by formal mathematical entities do not have the concept of human thinking exactly the same, it also pointed out that the formal concept analysis and processing of basic data in the form is the form of background, the background is a form of background knowledge in human small part.

In information retrieval, the concept lattice support structure can be used as retrieval. Cole and Eklund concept lattice method is applied to the analysis and visualization with 1962 properties and 4000 summary of the medical prescription database [8]. Eklund and Martin demonstrated the concept of levels of Web document indexing and navigation capabilities. Kent and Bowman built the concept lattice-based digital library system for the Nebula and the corresponding interface, the system meets RichardFuruta presented in DigitalLibrary'95 digital library system should meet the four conditions.

A formal context is defined as a triple $\mathbf{K} = (\mathbf{G}, \mathbf{A}, \mathbf{I})$, where $\mathbf{G}$ is a collection of objects, $\mathbf{M}$ is the set of attributes, but $\mathbf{I}$ is between $\mathbf{G}, \mathbf{A}$ binary relation, that $\mathbf{A}, \times \subseteq \mathbf{G}$ elements of $\mathbf{U}$ and $\mathbf{A}$ are called the object domain and the characteristics of the domain (property domain). Let $\mathbf{o} \in \mathbf{O}, \mathbf{d} \in \mathbf{D}$ then the old (that is, $\mathbf{I} = \mathbf{I} = \mathbf{o} \circ \mathbf{d}$) is read as an object $\mathbf{o}$ with characteristic $\mathbf{d}$ (attribute $\mathbf{d}$).

$$\forall \mathbf{O} \subseteq \mathbf{U} : f(\mathbf{O}) = \{ \mathbf{m} \in \mathbf{A} | \forall \mathbf{o} \in \mathbf{O}, \mathbf{old} \}$$

$$\forall \mathbf{D} \subseteq \mathbf{A} : g(\mathbf{D}) = \{ \mathbf{g} \in \mathbf{G} | \forall \mathbf{d} \in \mathbf{D}, \mathbf{old} \}$$

(7)

Functions $f$ and $g$ are usually called the power set of $\mathbf{U}$ and $\mathbf{A}$ the power set $\mathbf{P} (\mathbf{U})$ and $\mathbf{A}$ the power set $\mathbf{P} (\mathbf{A})$ between the Galois connection. Respectively, "$\mathbf{O}$ is all the objects in the set of attributes common to" and "also has all the attributes of the objects in $\mathbf{D}$ set". For a feature, if $\mathbf{C}$ is a concept that contains the characteristics of the largest, the feature's name is attached to the corresponding circle on $\mathbf{C}$. Concept lattice of a label is often used as a communication line graph mode, which makes the concept of a given data structure of the background became clear and easy to understand, enabling visualization of concept lattices show as figure2.

![Concept lattice building diagram](image)

**Figure 2.** Concept lattice building diagram

Concept lattice in the application process, you first need to solve the problem of grid construction. As the concept lattice of completeness, even for the appropriate size of the data, but also will have a large frame structure, its structure is undoubtedly a very time-consuming process. Therefore, the efficient generation of concept lattice algorithm becomes more urgent. From the constructor's point of view, according to concept lattice construction algorithm can be divided into three categories: batch algorithms, incremental construction algorithm and parallel algorithm.
Given formal context $K = (U, D, I)$, if the formal context $K_1 = (U_1, D_1, I_1)$ and $K_2 = (U_2, D_2, I_2)$ satisfy $D_2 \subseteq D$, $D_1 \subseteq D$, $U_2 \subseteq U$, $U_1 \subseteq U$ called $K_1$ and $K_2$ are the same domain in the form background, they are sub-formal context $K$, $K_1$, also known as the formal context of concept lattice $L (K_1)$ and $K_2$ form the background of concept lattice $L (K_2)$ is the same domain concept grid.

$$K_1 + K_2 = (U_1 \cup U_2, A_1 \cup A_2, I_1 \cup I_2)$$

(8)

Currently most of the concept lattice construction algorithm are given in the form for a background concept lattice constructed from scratch, but by the form of background tectonic theory, the worst case the number of background objects or attributes form the number of growth the number of nodes will lead to the concept of exponential growth, form the background to manipulate the grid size of the efficiency of impact is enormous. With the concept lattice theory in the field of data mining applications such as the increasingly widespread, the data also need to deal with almost explosive growth, and how massive amounts of data quickly from concept lattice structure, has become a serious problem. How to reduce from a large-scale, large scale structure in the form of the background complexity of concept lattice becomes the formal concept analysis research in the field of a key and difficult.

4. Recommendation system of e-commerce by formal concept analysis and collaborative filtering

For the current status of e-commerce recommendation system, this paper presents the concept of similarity based on formal concept analysis and collaborative filtering and e-commerce recommendation system solves the following two aspects: (1) tap the potential demand for the user, through the concept of ontology reasoning to the user is not recommended there might like to buy goods; (2) reduce the data sparseness.

Collaborative filtering, also known as social filtering (Social Filtering), the basic idea by comparing the user's interests and past behavior of the degree of similarity, to identify and target users with the same or similar interests of user groups, according to their resources assessment to predict the target user's interest, to recommend to the target user of resources. The basic idea is quite popular and now the "word of mouth (word-of-mouth)" a bit similar. I believe we all understand, in real life, the most effective information for their own, often recommended from friends. The system development environment for Windows xp operating system, MS SQL Server2005 database management system and MS Visual Studio.Net2005 Development Kit, and use Matlab6.5 improvements in the ART1 algorithm to perform classification tasks, and products generated by the development of body Protége3.2.1. For virtual store online clothing sales to build the concept of similarity based on formal concept analysis and collaborative filtering and e-commerce recommendation system.

Collaborative filtering systems is to use this idea to the Network Information Service (Information recommended), based on other users of the evaluation of certain information to be recommended to a user. Typically, system selection and the specified users with similar interests of users as a reference object.

E-commerce-based collaborative filtering recommendation method, with the growing number of users in a large number of users within the class, "nearest neighbor search" will become the bottleneck of the algorithm. The item-based collaborative filtering approach is calculating the similarity between items and instead of the similarity between users. For the items concerned, the similarity between them to stabilize, and therefore the greatest amount of work can be done offline similarity calculation step, thereby greatly reducing the online computation and improve the efficiency of recommended.

$$K_1 \leftarrow \text{ProcessByValue(PreProcessByWeight}(K1));$$
$$K_2 \leftarrow \text{ProcessByValue(PreProcessByWeight}(K2));$$

(9)

E-commerce based on FCA and collaborative filtering recommendation system does not analyze the similarity between the goods, but to learn the history of the target users and the similarity between users buying behavior, and thus according to the user's purchase history is similar to recommendations generated results. Purely based on collaborative filtering recommendation system does not need e-commerce features of the description of goods, it is the user to learn the similarities between buying behavior, rather than commodity-dependent features, so it can be recommended from the looks on the surface characteristics of different but there are actually very relevant goods.
Recommendation system for e-commerce knowledge were introduced, the overall e-commerce recommendation system have knowledge and understanding, especially for e-commerce recommendation system, the method used. The recommended system requirements with increasing levels of functionality, and its implementation techniques are also facing serious challenges, the background values for a single algorithm can be used to recommender goods in e-commerce, get Figure 3.

![Figure 3](image.png)

**Figure 3.** Recommendation system of e-commerce by formal concept analysis and collaborative filtering graph

Figure 3 is the currently recommended in e-commerce in China mainly used to find or search technology, which is based on the recommendation of the content to find the target. Relatively narrow recommendation technology, technology exists to find or retrieve a low degree of automation, poor motivation, low level of personalization and real-time defects and poor. In recent years, China began to attach importance to the recommended e-commerce research, Natural Science Foundation has funded the Recommended and automatically recommend the personality of theoretical studies have just begun, then the semantic ontology and e-commerce recommendation technologies, the product knowledge to meet network conditions, user needs and recommend functional knowledge of effective knowledge acquisition, aggregation and intelligent work recommended requirements, to meet large-scale e-commerce recommendation system, high quality and strong real-time requirements.

5. Conclusions

In this study, Formal concept analysis is a tool for data analysis or approach, especially given the information can be analyzed and processed. The data should be meaningful from the human mind can understand units - taking the concept of the formation of formal elements. Collaborative filtering systems is to use this idea to the Network Information Service (Information recommended), based on other users of the evaluation of certain information to be recommended to a user. This paper presents to construct recommendation system of e-commerce based on formal concept analysis and collaborative filtering to make up for these the shortcomings of existing e-commerce recommendation system. Collaborative filtering is the most studied personalized recommendation technology.

6. References