Recommender System for Pharmacy Shop By using Item-Based Collaborative Approach

Nwe Ni Oo : Nan Saing Moon Kham
University of Computer Studies, Yangon
lovelygirl2u@gmail.com

Abstract

Recommender systems use the opinions of a community of users to help individuals in that community more effectively identify content of interest from a potentially overwhelming set of choices. One of the most successful technologies for recommender systems, called collaborative filtering, has been developed and improved over the past decade to the point where a wide variety of algorithms exist for generating recommendations. Item-based collaborative filtering algorithms have been presented to deal with scalability problems associated with user-based collaborative filtering. The computation of item-based collaborative filtering is a large amount items rating by users.

The system provides a solution to the problem of how to choose a pharmacy in the presence of an overwhelming amount of information. This system implements as Recommender System for Pharmacy Shop by using Item-Based Collaborative Approach.

1. Introduction

Recommender systems such as those employed by Amazon and Netflix automate the process of recommending products and services to consumers based on various types of data concerning consumers, products, and previous interactions between consumers and products. Consumer-product interactions can take different forms, such as product/service purchases, ratings, paper citations, and catalog and website browsing activities. Recommender systems are being increasingly adopted in a wide range of applications, especially in e-commerce applications. They have become a standard e-commerce technology that helps increase online and catalog sales and improve customer loyalty.

Recommender systems based on collaborative filtering predict user preferences for product or services by learning past user-item relationships[1]. These systems can be a valuable competitive advantage to retailer companies, especially in e-commerce. One of the most popular and successful techniques that has been in recommender systems is known as collaborative filtering[2]. A system that produces good recommendations can inspire trust in the company and help users find products they truly want. At the same time, an engaging interface for collecting recommendations allows the company to gather preference information from its customers and retailer offerings to each customer. Both the company and its customers stand to benefit. Few researchers have investigated the effect of interfaces on the use of recommendations.

If items and user data are very large, sometimes item-based collaborative filtering can be delay because of calculating large amount of data.

2. Motivation

Personalization and profiling is key to many successful web sites. Consider that there is considerable free content on the web, but comparatively few tools to help us organize or mine such content for specific purposes. rating-based collaborative filtering is to ask users to rate resources so that they can help each other find better content. The system presents a system that approached to item-to-item collaborative filtering which is easy to implement and can support a full range of applications. The objectives of the system are to help customers to quickly find out items that they will probably like, with no tremendous time and effort. User can study recommender system such as personalization and filtering techniques etc. This system provides a way to filter out the excess of information available and to build trust for the items by using this system.

3. Related Work

One of the earliest collaborative filtering recommender systems was implemented as an email filtering system called Tapestry [3]. Later on this technique was extended in several directions and was applied in various domains such as music recommendation and video recommendation. In this section we briefly review the research literature related to collaborative filtering recommender systems. Collaborative filtering algorithms can be classified into 2 categories: One is memory-based, which predicts the vote of a given item for the active user based on the votes from some other neighbor users. Memory based algorithms operate over the entire user voting database to make predictions on the fly. The most frequently used approach in this category is nearest-neighbor collaborative filtering: the prediction is calculated based on the set of nearest-neighbor users for the active user (user-based collaborative filtering approach) or, nearest neighbor items of the given item (item-based collaborative filtering approach).

4. Overview of Recommender System

In this section, the system introduces the background theories used to implement the Recommender System.

4.1 Recommender System

Recommender systems apply data analysis techniques to the problem of helping users find the items they would like to purchase at E-Commerce sites by producing a predicted likeliness score. Recommendation algorithms are best known for their use on e-commerce Web sites, where they use input about a customer's interests to generate a list of recommended items. Many popular e-commerce web sites-Amazon.com for example - have adopted this technique in making their online shopping system more efficient. Many applications use only the items that customers purchase and explicitly rate to represent their interests, but they can also use other attributes, including items viewed, demographic data, subject interests, and favorite artists [4].

At other online recommendation systems [4] use recommendation algorithms to personalize the online store for each customer. The store radically changes based on customer interests. E-commerce recommendation algorithms often operate in a challenging environment. For example:

- A large retailer might have huge amounts of data, tens of millions of customers and millions of distinct catalog items.
- Many applications require the results set to be returned in real time, in no more than half a second, while still producing high-quality recommendations.
- New customers typically have extremely limited information, based on only a few purchases or product ratings.
- Older customers can have a glut of information, based on thousands of purchases and ratings.
- Customer data is volatile: Each interaction provides valuable customer data, and the algorithm must respond immediately to new information.

Most recommendation algorithms start by finding a set of customers whose purchased and rated items overlap the user's purchased and rated items. The algorithm aggregates items from these similar customers, eliminates items the user has already purchased or rated, and recommends the remaining items to the user. For each of the user's purchased and rated items, the algorithm attempts to find similar items. It then aggregates the similar items and recommends them.

4.2 Personalization Techniques

The most common personalization techniques are Content-Based Filtering, Collaborative Filtering, Rule-Based Filtering and Web Usage Mining

4.2.1 Collaborative Filtering

Collaborative filtering systems predict a user's interest in new items based on the recommendations of other people with similar interests [5]. Collaborative filtering compares a user's tastes with those of other users in order to build up a picture of like-minded people. The choice of content is then based on the assumption that this particular use will value what like-minded people also enjoyed. The user's tastes are either inferred from previous actions are else measured directly by asking the user to rate products.

The user's interests are compared with those of other customers to generate titles that are then recommended during interaction. This is a method echoed by a number of online retailers, and it is also use to power recommendation engines for entertainment and television viewing.

The collaborative approach to recommendation is very different: Rather than recommend items because they are similar to items a user has liked in the past, this system recommends items other similar users have liked. Rather than compute the similarity of the items, this system computes the similarity of the users. Typically, for each user a set of "nearest neighbor" users are found with whose past ratings there is the strongest correlation. Scores for unseen items are predicted based on a combination of the scores known from the nearest neighbors [6]. To predict the rating value of a given item for an active user, a subset of neighbor users are chosen based on their similarity to the active user – called nearest-neighbor users – and their ratings of the given item are aggregated to generate the prediction value for it.

There are two kinds of collaborative filtering. They are:

User-Based collaborative filtering (also known as traditional collaborative filtering or user-to use collaborative filtering or memory-based method) and Item-based collaborative filtering (also known as item-to-item collaborative filtering or model-based method).

4.2.1.1 Challenges of user-based Collaborative filtering

Although the user-based collaborative filtering systems have been very successful in the past, their widespread use has revealed some potential challenges such as sparsity and scalability[7].

Sparsity : In most real-world cases, users rate only a very small percentage of items. This causes data sets to become sparse. In such cases, the recommendation engine cannot provide precise proposals, due to lack of information.

Scalability: One of the major drawbacks of the user-based Collaborative Filtering systems in general is that they do not scale well. The user-based method does little or no offline computation, and its online computation scales with the number of users and items. The computational complexity of these methods grows linearly with the number of users and items, which in commercial applications can each grow to be several million.

The weakness of the user-based Collaborative Filtering algorithm led to explore alternative recommender system algorithms. The item-based collaborative filtering addresses these challenges, especially the scalability challenge.

5. Proposed System

This paper is intended to approach recommender system for pharmacy shop by using item based collaborative filtering.

5.1 Item-Based Collaborative Filtering (IBCF)

This method analyzes the relationships between items rather than between users, because item relationships are relatively static. Similarity relations for items are computed offline. Item-Based methods are also known as model-based methods.

The item-based collaborative filtering method makes recommendations according to the following simple step by step procedure:

1. Users are requested to give numeric ratings to the items.
2. A recommender system correlates the ratings in order to determine which item's ratings are the most similar to other item's ratings.
3. The system predicts ratings of new items for the target user, based on the ratings of similar items already rated by the users.
4. Then, if these new items seem to be preferred, the system recommended them to the user.

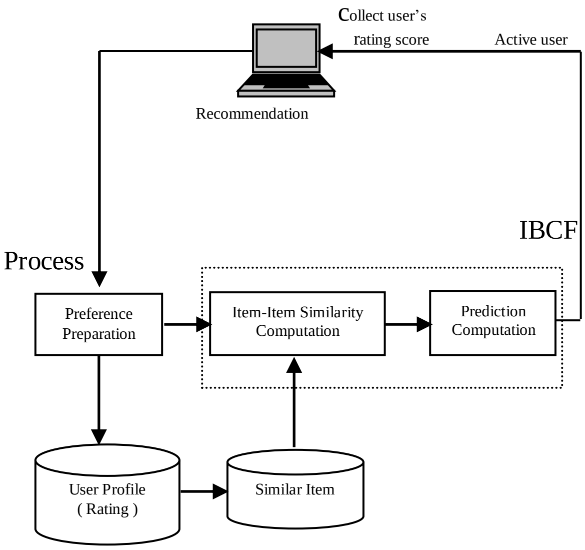

Figure .1 Overview of System Design

In this system, users are requested to give ratings to the items. Rating values are numeric. This system calculate item-item similarity over items. And then this system predicts and recommends items to the active user.

5.2 Item Similarity Computation

One critical step in the item-based collaborative filtering algorithm is to compute the similarity between items and then to select the most similar items. The basic idea in similarity computation between two items i and j is to first isolate the users who have rated both of these items and then to apply a similarity computation technique to determine the similarity si, j .

There are a number of different ways to compute the similarity between items. The most commonly used similarity computation methods are :

- Cosine-based similarity
- Correlation-based similarity
- Adjusted-cosine similarity

5.3 Adjusted Cosine Similarity

One fundamental difference between the similarity computation in user-based Collaborative Filtering and item-based Collaborative Filtering is that in case of user-based CF the similarity is computed along the rows of the matrix but in case of the item-based Collaborative Filtering the similarity is computed along the columns i.e., each pair in the co-rated set corresponds to a different user. Computing similarity using basic cosine measure in item-based case has one important drawback–the difference in rating scale between different users are not taken into account. The adjusted cosine similarity offsets this drawback by subtracting. The corresponding user average from each co-rated pair.

Formally, the similarity between items *i* and *j* using this scheme is given by

$$sim(i,j) = \frac{\sum_{u \in U}(R_{u,i} - \bar{R}_u)(R_{u,j} - \bar{R}_u)}{\sqrt{\sum_{u \in U}(R_{u,i} - \bar{R}_u)^2}\sqrt{\sum_{u \in U}(R_{u,j} - \bar{R}_u)^2}}$$

	I1	I2	I3	I4	I5
Aye Aye	1	4	2	5	4
Mya Mya	4	2	5	2	-
Aung	4	3	-	3	3
Than	2	4	3	?	5

$$sim(i,j) = \frac{(1-3.2)*(4-3.2)+(4-3.25)*(2-3.25)+(4-3.25)*(3-3.25)+(2-3.5)*(4-3.5)}{\sqrt{((1-3.2)^2+(4-3.25)^2+(4-3.25)^2+(2-3.5)^2)}\sqrt{((4-3.2)^2+(2-3.25)^2+(3-3.25)^2+(4-3.5)^2)}} = -0.799$$

5.4 Prediction Computation

The most important step in a collaborative filtering system is to generate the output interface in terms of prediction.

Once we isolate the set of most similar items based on the similarity measures, the next step is to look into the target users ratings and use a technique to obtain predictions.

The following equation (Weighted Sum) is used to compute the prediction on an item *i* for a user *u* by computing the sum of the ratings given by the user on the items similar to *i* .

Each ratings is weighted by the corresponding similarity $s_{i,j}$ between items *i* and *j* . Formally, using the notion we can denote the prediction $P_{u,i}$ as

$$P_{u,i} = \frac{\sum_{\text{all similar items, } N}(s_{i,N} * R_{u,N})}{\sum_{\text{all similar items, } N}(|s_{i,N}|)}$$

	I1	*I2*	*I3*	*I4*	*I5*
I1	*1*	*-0.79*	*0.22*	*0.01*	*-0.6*
I2	*-0.79*	*1*	*-0.42*	*0.20*	*0.95*
I3	*0.22*	*-0.42*	*1*	*-0.11*	*-0.46*
I4	*0.01*	*0.20*	*-0.11*	*1*	*0.02*
I5	*-0.6*	*0.95*	*-0.46*	*0.02*	*1*

$$P(\text{Than Than}, I4) = \frac{(0.20*4)+(0.02*5)}{0.20+0.02} = 4.09 = 4$$

Basically, this approach tries to capture how the active user rates the similar items. The weighted sum is scaled by the sum of the similarity terms to make sure the prediction is within the predefined range.

5.5 Database Design of the System

Table 1 : User Table

UserTable	
UserID	int
UserName	varchar (100)
Password	varchar(50)
AdultChild	int
Email	varchar(50)
RegDate	datetime
Role	varchar(20)

Table 2 : Category Table

Category Table	
CategoryID	int
CategoryName	varchar (50)

Table 3 : Sub Category Table

Sub Category Table	
SubCategoryID	int
SubCategoryName	varchar (50)

Table 4 : Item Table

Item Table	
ItemID	int
CategoryID	int
SubCategoryID	int
ItemName	varchar (100)
Composition	varchar (200)
Actions	varchar (200)
Dosage	varchar (200)
ItemImage	varchar (100)
Description	varchar (200)
AdultChild	int

Table 5 : User Item Rating Table

User Item Rating Table	
UserItemID	int
UserID	int
ItemID	int
Rating	int

In this system, all data are stored in the database. When users rate the items, this system stored user ratings to the rating table. When user wants to know recommendation value, this system calculate item-item similarity and predict the recommendation items.

6. Future Work

This system can extend to improve the accuracy of recommendation to deliver personalized services. The system can be extended by combining the content-based filtering and the collaborative filtering for future work.

7. Conclusion

Recommender Systems have been used in e-commerce sites to make product recommendations and to provide customers with information that helps them decide which product to buy. Recommender systems provide a solution to the problem of how to choose a product in the presence of an overwhelming amount of information.

Recommender Systems enhance e-commerce sales in three ways: converting browser into buyers, increasing cross-sell, and building loyalty between seller and buyer. For large retailers like Amazon.com, a good recommendation algorithm is scalable over very large customer bases and product catalogs, requires only subsecond processing time to generate online recommendations.

The main advantages of an item-based system over a user-based one is scalability. Item-based solutions do not have to search carefully databases containing potentially millions of users in real time in order to find users with similar tastes or interest. Instead, they can pre-score content based on user ratings and/or their attributes and then make recommendations without incurring high computation costs.

References

[1] R. M. Bill and Y. Koren, "Improved Neighborhood-based Collaborative Filtering," *At&T Labs*-Research 180 park Ave,Florham Park,NJ07932.

[2] Jiyong Zhang and Pearl pu. "A Recursive Prediction Algorithm for Collaborative Filteing Recommender Systems," *In Proceedings of the ACM Conference on Recommender Systems*, pages 57-64, 2007.

[3] B. Srwar, G. Karypis, J. Konstan, and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms,"*10th Int'l World Wide Web Conference,ACM Press*, pp.285.295, 2001.

[4] B.Smith, G.Linden and J. York,"Amozon.com recommendations:item-to-item collaborative filtering," *in Proc. IEEE Internet Computing,* Volume7,Issue 1,pages 76-80, January, 2003.

[5] J. Herlocker, J.A. Konstan and, J Riedl, "An Empirical Analysis of Design Choice in Neighborhood.based Collaborative Filtering Algorighms," *Information Retrieval* 5, 287-310, 2002.

[6] M. Balabanvoic and Y. Shoham, "Content-based, collaborative recommendation", *Communications of the ACM.* 40(3): pp-66-72, March 1997.

[7] G. Karypis, "Evaluation of the Item Based Top-N Recommendation Algorithms", *Technical Report Cs-TR-oo 46,* Computer Science Dept, University of Minnesota.