Developing an Interface Agent Approach to Personalized on Book Shopping Mall

Myat Mon Aye
Computer University (Mandalay)
myatmonaye.j18@gmail.com

Abstract

As the number of items increases dramatically, the problem of information overload becomes more serve when browsing and searching. The agents overcome this problem and help the user interaction on a working environment on the user's interests and desires. The user's interest information is obtained from the user registration and user's activities. Depending on this information the system calculates the collaborative recommendation for user requirements using the agglomerative clustering algorithm and linear regression for the new user and content-based recommendation for existing user using naïve Bayesian classification. This system provides assistance to the user by solving problem through agent user interaction.

Key words

Interface agent, Content based filtering, Collaborative filtering, Agglomerative clustering, and Naïve Bayesian classification

1. Introduction

Today's many web pages provide various types of items to satisfy varying customer's demands and needs. In spite of a large number of items are provided, most users use only a small fraction of the items.

The system uses the idea of interface agent that acts as a personal assistance and the recommender system assists the interface agent to achieve the capabilities of the interface agent. The interface agent stores the user's registration data is regarded as user characteristics and a set of items that are bought by the user is regarded as preferences of the user. The interface agent acts as personal assistance as follows. First, an interface agent is communicated by user. And then, when the user is new, the existing users are clustered based on the users' characteristics and it recommends items based on users who are the same cluster with the new user by using collaborative filtering method. When the user is an existing user, it recommends items based on the past user preferences by using content-based filtering method.

2. Related Work

Interface agent acts as a personal assistance in their working environment and who collaborates with the user to achieve the common goal. And interface agent cooperates with the user in performing its tasks, working as personal user assistance. The agent is pro-active taking the initiative and not passive. A personal Learning apprentice assists a user in managing a meeting calendar and concentrates on the prediction of meeting parameters such as location, duration and day-of-week [10]. Unix consultant (UC) uses natural language and modeling user goals and plans, takes the initiative in offering the user information about certain Unix concepts or commands, correcting, at the same time, any misconceptions [2]. Content based filtering system recommends items to a given user based on the content of items and the preferences of user. Collaborative filtering system recommends items that are liked by other users with similar interests [8]. These filtering systems are used to recommend items such as movies, fashion etc.

In this paper, section 1 describes the introduction, section 2 states the related work, section 3 describes the background theory such as interface agent, recommender system, content-based filtering and collaborative filtering and section 4 describes the system architecture with case study.

3. Background Theory

3.1 Interface Agent

An interface agent is a computerized entity that can perform a particular task that has been assigned to it. It can capable of acting on behalf of the user.

Another aspect of an interface agent is that they are capable of learning. This learning capability allows them to adapt to a new situations they encounter. In addition, they can learn how a user would like them to act [3].

The purposes of the interface agent are: it requires less work from the end-user and application developer, Agent can more easily adapt to the user over time and become customized to individual and organizational preferences and habits and it interacts directly with the users.

In our system, we use the interface agent that assists the user in order to user's interests and desires using recommender system. It supports the old users and new users who are entered the system. So the interface agent solves the cold start problem.

3.2 Recommender System

Recommender systems are agent-based systems that use stored preferences to locate and suggest items of interests to users they serve. Recommender systems have gained increasing popularity on the web, both in research system and online commercial site that offer recommender systems as one way for consumers to find products they might like to purchase.

It can be distinguished between personalized and non-personalized systems. Non-personalized systems offer the same recommendations for all users, e.g. the up-to-date top ten listening in music or film. Personalized systems can be divided in content-based filtering systems, collaborative filtering systems and hybrids [7].

This system uses the collaborative filtering system for new users and the content-based filtering system for old users.

3.2.1 Content Based Filtering

Content based filtering system tries to recommend items similar to those a given user has liked in the past [6]. It recommends items based on the attributes (content) of the item rather than other users rating. It needs to track the user previous traversal history in order to extract the content of the user's interests. And then items are matched either to a user's interest profile or query on the basic of content rather than opinion.

The advantages of the content-based filtering system are no need for data on other users, able to recommend to users with unique tastes and able to recommend new or unpopular items [5]. This system uses Naïve Bayesian classification method in section 3.5 for content based filtering to recommend items.

3.2.2 Collaborative Filtering

Collaborative filtering-based system is the most successful recommendation technique to date. Collaborative filtering means that people collaborate to help on another perform filtering by recording their reactions to items they buy.

Collaborative filtering can recommend items based on the users that they have the similar preferences with active user. If the active user has no preference, it applies another type of user data such as demographic data in the case of newly registered user [5]. So, this system uses the Agglomerative clustering algorithm in section 3.3 to group users who have the same characteristics and then recommends items based on that users' preferences. And then this system applies the linear regression in section 3.4 to predict the existing user ratings that causes the accurate recommendation.

3.3 Agglomerative Clustering

Bottom-up merging techniques are called agglomerative clustering. Clusters are merged if the distance metric between the two clusters is less than a certain threshold. The threshold increases at each stage of merging. While merging cluster one by one, we can draw a tree diagram known as dendrogram. Clustering of the data objects is obtained by cutting the dendrogram at the desired level, and then each connected component forms a cluster [9].

Input:
$D = \{t_1, t_2, ..., t_n\}$ // Set of elements
A // Adjacency matrix showing distance between elements.
Output:
DE // Dendrogram represented as a set of ordered triples.
Agglomerative Algorithm:
$d = 0$;
$k = n$;
$K = \{\{t_1\}, ..., \{t_n\}\}$;
$DE = \{< d, k, K >\}$; // Initially dendrogram contains each element in its own cluster.
repeat
$oldk = k$;
$d = d + 1$;
A_d = Vertex adjacency matrix for graph with threshold distance of d;
$< k, K >= NewClusters(A_d, D)$;
if $oldk \neq k$ then
$DE = DE \cup < d, k, K >$; // New set of clusters added to dendrogram.
until $k = 1$

Figure 1. Agglomerative clustering

Where d is the threshold distance, k is the number of clusters and K is the set of clusters.

In the Agglomerative clustering, dissimilarities are most calculated according to the groups' average method such as [9].

$$d_{AB} = \frac{1}{|A||B|} \sum_{i \in A, j \in B} d_{ij} \qquad (1)$$

We use Manhattan distance as

$$d_{ij} = \sum_{k=1}^{K} |X_{ik} - X_{jk}| \quad (2)$$

d_{ij} is the dissimilarity between and d_{AB} is the dissimilarity between the two clusters A and B

3.4 Linear Regression

In Linear regression, data are modeled using a straight line. Bivariate linear regression models a random variable, Y (call a response variable), as a linear function of another random variable, X (called a predictor variable), that is [4]

$$Y = \alpha + \beta X \ , \ \alpha = \bar{y} - \beta \bar{x} \quad (3)$$

$$\beta = \frac{\sum_{i=1}^{s} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{s} (x_i - \bar{x})^2}$$

Where $\bar{x}$ is the average of $x_1, x_2, \ldots, x_s$, y is the average of $y_1, y_2, \ldots, y_s$ and α and β are the coefficient.

3.5 Naïve Bayesian algorithm

The naïve Bayesian classification algorithm is simple classifier, works as follow [4]:

$$P(C_i|X) = \frac{P(C_i)}{P(X)} \prod_{k=1}^{n} P(X_k|C_i) \quad (4)$$

- $X = (x_1, x_2, \ldots, x_n)$ depiction n measurements made on the sample from n attributes
- M classes are C_1, C_2, C_3

4. System Architecture with case study

In order to present capabilities of the interface agent which assistants as well as advisors by modeling their ability to support user. In this system, the interface agent acts as a personal assistance in the application environment, helps the users' interaction with the application and assists the users to get their requirements based on their information and behaviors. And then this system applies the recommender system to assists the interface agent in providing the users' requirements.

The Figure 2 shows the overview of an interface agent approach based on recommender system. In this system, when the new user enters the system, the user must register their information such as name, date-of-birth, gender and password and must select the hobbies such as Pleasure, Learning English, Computer, Biography, History, Religion, General Knowledge and Inspiration. The interface agent stores the user's information and then assists the user, provides user requirements and advices by calculating the similar users based on user characteristics and recommends book shop with appropriate books using collaborative filtering based on similar users' buying information. In calculating the Collaborative filtering, linear regression is used to predict the existing users' ratings as in section 4.1. When the old user enters the system, the user must fill the name and password and then the interface agent checks the user information and behaviors (such as reading and buying) and recommends books based on the user's past preferences using content–based filtering in section 4.2.

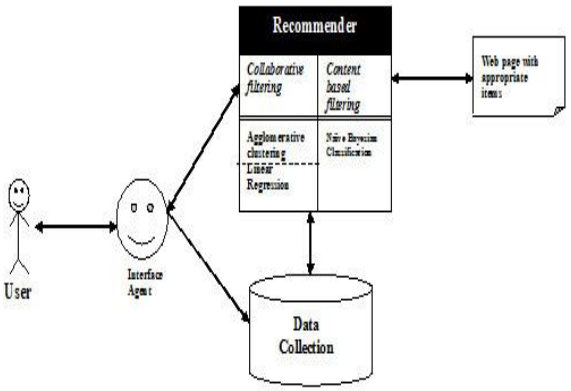

Figure 2. System flow diagram

The Figure 3 describes the step by step procedure of this system. We use five methods such as collaborative filtering, agglomerative clustering algorithm, linear regression, content based filtering and naïve Bayesian classification. This system provides more accurate recommendation by using five methods than one method.

```
begin
    if (user = = new user)
    begin
        Accept userinformation;
        Calculate similar users by using Agglomerative
        clustering Algorithm;
        Preprocess spare data using linear regression;
        Recommend items by using collaborative filtering
        method;
    end
    else
    begin
        Accept name;
        Accept password;
        if (name = = existingusername and password = =
        existinguserpassword)
        begin
                Calculate whether the items are recommended
                or not by using Naïve Bayesian classification;
```

```
            Recommend items by using content based
            filtering method;
        end.
        endif;
    end.
    endif;
    end.
```

Figure 3. Step by step procedure

4.1 Case Study of the Collaborative Filtering

The collaborative filtering does the following. The Table 1 shows the user registration data. For hobbies, 1 represents that the user likes this category and 0 represents the user doesn't like this category. For Gender, 1 represents the female and 0 represents the male.

Table 1. User registration data

	U1	U2	U3	U4	U5	U6	…	U47	U48
Pleasure	1	1	1	1	1	1	…	0	0
English	0	1	1	0	0	1	…	0	0
Computer	0	0	1	1	1	1	…	1	1
Biography	0	0	0	0	0	1	…	0	1
History	0	0	0	0	0	0	…	0	0
Religion	0	0	0	0	0	0	…	1	1
General Knowledge	0	0	0	0	0	0	..	1	0
Inspiration	0	0	0	0	0	0	…	1	0
Age	20	21	12	23	24	30	…	40	35
Gender	1	1	1	1	1	1	…	0	0

In order to recommend items to new user, the system computes the similar users using Agglomerative clustering and Table 2 represents the dissimilarities matrix for the users describe in Table 1.

Table 2. Dissimilarities matrix for users

	U1	U2	U3	U4	U5	U6	…	U47	U48
U1	0	1	2	1	1	4	…	7	6
U2	1	0	1	2	2	3	…	8	7
U3	2	1	0	1	1	2	…	7	6
U4	1	2	1	0	0	3	…	6	5
U5	1	2	1	0	0	3	…	6	5
U6	4	3	2	3	3	0	…	8	5
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
U47	7	8	7	6	6	8	…	0	3
U48	6	7	6	5	5	5	…	3	0

DE={<0,47,{{U1},{U2},{U3},{U4,U5},{U6},…,
{U47},{U48}}>,
<1, 44,{{U1,U2,U3,U4,U5},{U6},…,
{U47},{U48}}>, …,
<10,1,{U1,U2,U3,U4,…,U47,U48}>}

We cut at threshold 1 because there are many similar users and the system recommends items to active user based on similar users. The interface agent records the user buying information. Actual rating based on the buying information is shown in Table 3. In our system, we defines the rating from 1 to 5 and rating 1 and 2 are assumed dislike and 3 to 5 are assumed like. And we assumed that the user3 is the active user.

Table 3. rating table before linear regression

	I1	I2	I3	I4	I5	
U3	1	4	3	5	1	Test
U1	1	5	3	4	1	Training
U2	2	2	2	-	1	
U4	1	5	3	5	1	
U5	1	-	-	5	2	

Using Agglomerative, the system recommends as

U3 1 3 2 3 1

The system uses linear regression to predict the existing user ratings. Table 4 shows the ratings after applying the linear regression.

Table 4. rating table after linear regression

	I1	I2	I3	I4	I5	
U3	1	4	3	5	1	Test
U1	1	5	3	4	1	Training
U2	2	2	2	0	1	
U4	1	5	3	5	1	
U5	1	5	5	5	2	

It uses both linear regression and Agglomerative clustering and then the system recommends as

U3 1 4 3 4 1

4.2 Case Study of Content-based Filtering

For the old user, the system calculates whether the item is recommended or not using the Naïve Bayesian classification based on the past user preferences. The Table 5 represents the contents of the past user preference items.

Table 5. User×Item rating table

User	I1	I2	I3	I4	I5	I6
U1	3	4	3	2	1	3
	Training					Test

The Table 6 shows the contents of the items.

Table 6. content of items

I1	Min Lu	Pleasure	Myanmar	like
I2	Min Zaw	Pleasure	Myanmar	dislike
I3	Heinemann	English	Cambridge	like
I4	Min Lu	Pleasure	Myanmar	like

I5	M. Wooldridge	Computer	UK	dislike
I6	Min Lu	Pleasure	Myanmar	?

The system computes whether the I6 is recommended or not as follows:

P(X | like) = 2/3 * 2/3 * 2/3 =0.343

P(X | dislike) = 0*1/2*1/2=0

The system recommends this item because like probability is grater than dislike probability.

5. Evaluation of the System

Statistical accuracy matrics evaluate the accuracy of a system by comparing the numerical recommendation scores against the actual user ratings for the user-item pairs in the test dataset. Mean Absolute Error between ratings and predictions is a widely used matric. The MAE is computed by first summing these absolute errors of the N corresponding ratings-prediction pairs and then computing the average [1].

$$MAE = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N} \quad (5)$$

P_i = prediction generated by Recommender system

q_i = rating defined by user

N =user item pairs

The lower the MAE, the more accurately the recommendation engine predicts user ratings [1]. We compute the mean absolute error for section 4.1 as the following.

For collaborative filtering without linear regression,

MAE= (|1-1|+|3-4|+|2-3|+|3-5|+|1-1|)/5=0.8

For collaborative filtering with linear regression,

MAE= (|1-1|+|4-4|+|3-3|+|4-5|+|1-1|)/5=0.2

The MAE without linear regression is 0.8 and the MAE with linear regression is 0.2. So, the system applies both the linear regression and Agglomerative clustering.

6. Conclusion and Further Extension

This system proposes an interface agent approach based on recommender system to act as a personal assistance. It also acts as an advisor that helps user to get their needs and interests with good quality service. And in most web application, personalization techniques are used to tailor information services to personal user needs and wants. And then interface agent based on the recommender system that can achieve significant accurate advices.

This system works especially for file with .doc format (Microsoft word file format) that do not support other files format. In the future, this system is intended to use any file format. Another great-expected extension is to make movie recommendation, Music recommendation and so on.

7. References

[1] Badrul Sarwar et al,"Item-based Collaborative Filtering recommendation algorithms", 2001

[2] Chin, D," Intelligent interfaces as agents, in Intelligent user interfaces", Sullivan, J. and Tyler S. (editors), ACM Press, New York, USA (1991).

[3] Henry Liebernam and Ted Selker, "Agent for the user interface"

[4] Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann publishers, 2001.

[5] Jong-Hwan Seo, Kun-Pyo Lee, "Development of Website Design Personalization Service Using Design Recommender System"

[6] Marko Balabanovic and Yoau Shoham, "Content-based, Collaborative Recommendation", 1997

[7] Michel Hecking, Florian Kleene, Oliver Schmitt and Philipp Westrich, "Recommender Systems"

[8] Mohamed JEMNI, Mohamed Koutheair KHRIBI, "Toward a Hybrid Recommender System for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval"

[9] Patrick Andre Pantel, "Clustering by Committee", University of ALBERTA, 2003

[10] Pattie Maes and Robyn Kozierok, "Learning Interface Agents"