Mining User Preferences: Implicit and Explicit

Bhavin M. Desai¹, Hemant D. Metalia² and Maitri Jhaveri³
¹,²,³Institute of Computer Technology, Gujarat Law Society, Ahmedabad, India.
¹bhavin882003@gmail.com, ²hemant_metalia@yahoo.com and ³jmaitri@glsict.org

ABSTRACT
Current information systems deal with a huge amount of content, and deliver in consequence a high number of results in response to user queries. Thus, users are not able to distinguish relevant contents from secondary ones. In this paper we have proposed a hybrid learning approach to provide automated assistance for personalized product recommendation. The novel feature of this work is that the system learns and uses models of both user preferences and the user's intentional context. We propose a model which tries to address these issues in greater detail by adopting a new algorithm for finding user preferences. User preference is learned via proposed algorithm by identifying user behaviour in viewing or selecting the products. This product recommendation scheme is based on an analysis on both the preference and intentional context model. Our approach will support users with different preference structures and intentional contexts. Experiments will show that both the efficiency and performance can be improved compared to the classical techniques.

KEYWORDS
User Preferences, Implicit Preferences, Explicit Preferences, User Context, Preference Mining

1. INTRODUCTION
1.1 INTRODUCTION TO USER PREFERENCES
User Preferences is a set of interest of a user towards an object. [12] [13] [16] A recommender system is a personalization system that helps users to find items of interest based on their preferences. [4] While a substantial amount of research in recommendation systems focus on recommending the most relevant item to the user,[1] a very few focus on additional contextual information like user’s personal information, Hobbies, current location, user’s behaviour while accessing item. The type of user preferences can be classified into two classes:
- Positive Preferences
- Negative Preferences

Positive Preferences are the set of characteristics of objects in which user is interested. In our practical implementation of this algorithm we have used Positive Preferences to suggest other products which are based on those characteristics when user comes to purchase other items. [10]

Negative Preferences are the set of characteristics of objects in which user is not interested. In our practical implementation of this algorithm we have used Negative Preferences to NOT suggest other products which are based on those characteristic when user comes to purchase other items because we assumes that the products based on those characteristics are not of user’s interest.

The process of fetching user’s preferences can be done in two ways:
- Explicit Preferences
- Implicit Preferences

Explicit Preferences are the set of interests towards set of objects that can be assumed directly from the user’s explicit data like user’s Personal Profile Information, Hobbies, Gender, Age Group the user belong to, Location, Income of user etc. [3] Implicit Preferences are the set of interests towards set of objects that can be assumed indirectly from user’s behavior like Time Spent on a particular item, Navigation from one item to another, Repetitive visit, Add or remove from the cart, Purchase of an item etc. while accessing the items.

2. LITERATURE SURVEY OF EXISTING MACHINE LEARNING ALGORITHMS FOR MINING USER’S PREFERENCES
2.1 COMPARATIVE STUDY OF MACHINE LEARNING ALGORITHMS:
Support Vector Machine (SVM) [13]
Advantages:
- SVM Produce very accurate classifiers.
- It is less over fitting.
- It is robust to noise.

Disadvantages:
- SVM is a binary classifier. To do a multi-class classification, pair-wise classifications can be used (one class against all others, for all classes).
- It is computationally expensive.
• For large datasets SVM runs extremely slow.

K Nearest Neighbor (kNN) [2]

Advantages:
• Simple to use
• Effective if the training data is large
• Robust to noisy training data, especially if the inverse square of weighted distance is used as the “distance” measure

Disadvantages:
• Computation cost is quite high because it needs to compute distance of each query instance to all training samples
• The large memory to implement in proportion with size of training set
• Low accuracy rate in multidimensional data sets
• Need to determine the value of parameter K, the number of nearest neighbours
• Distance based learning is not clear which type of distance to use
• Which attributes are better to use producing the best results

Naïve Bayes [19] [13]

Advantages:
• Handles quantitative and discrete data
• Robust to isolated noise points
• Handles missing values by ignoring the instance
• During probability estimate calculations
• Fast and space efficient
• Not sensitive to irrelevant features
• Quadratic decision boundary

Disadvantages:
• If conditional probability is zero, it assumes independence of features
• Naïve Bayesian prediction requires each conditional probability be non-zero

Otherwise, the predicted probability will be zero

\[
P(X | C_i) = \prod_{k=1}^{n} P(x_k | C_i)
\]

Joachims Algorithm [18]

Advantages:
• Simple to use
• Works well on large datasets

Disadvantages:
• It is majorly based on clickthrough items
• This algorithm works on a basic assumption that the user scans search results strictly from top to bottom
• If a user skips one or two data, it may generate wrong results

Spy Naïve Bayes [18]

Advantages:
• It concentrates on positive as well as negative preferences
• It is more efficient than previous algorithms

Disadvantages:
• It has a cold start problem
• It has data sparsity problem
• It do not take into consideration of unlabeled data

3. OUR PROPOSED MODEL:

3.1 INTRODUCTION OF PROPOSED ALGORITHM

Here we have taken an e-commerce application of Online Shopping of an Electronic Items into the consideration and we have developed and implemented algorithm for the same.

For our practical implementation of the algorithm we have taken following parameters of user to mine the explicit preferences of a user:
• Personal Profile
• Age group
• Hobbies
• Gender
• Income of the User

For mining Implicit Preferences of a user we have taken following parameters of the user into consideration:
• Time Spent on a particular Item
• Item Added to a cart
• Item removed from the cart
• Purchase of an Item
• Repeated visits

3.2 Proposed Algorithm

There are major three parts of the algorithm:
1. Mining Explicit Preferences from user’s personal information.
2. Mining Implicit positive preferences by capturing user’s behavior.
3. Mining implicit negative preferences from unlabeled data.

Algorithm for displaying Item
1. Fetch user profile
2. Categorize the user into Age group, Salary, Gender, Hobbies.
3. Apply Weight to all items as per age group
4. Fetch stored user preferences if available
a. whether preferences are available or not
Mining User Preferences: Implicit and Explicit

- Extract items for the weight >= some level (here 2) for age-group for and call it \(I_0\).
- From \(I_0\), subtract the items that do not fit in the cost criteria. Here cost < 3*salary and call it \(I_1\).
- From \(I_1\) subtract the items that do not fit according to user's hobby and call it \(I_2\).
- Add items to \(I_2\) that can be fit into gender criteria (e.g. if item is available in pink color ladies can prefer it) and call it \(I_{profile}\).
- \(I_f = I_{profile}\)

b. if preferences are available

{  
- Fetch the items that fit according to stored positive user preferences and call it \(I_p\).
- Fetch the items that fit according to stored negative user preferences and call it \(I_n\).
- \(I_f = I_p + I_{profile} - I_n\)
}

5. Apply ranking function \(F_r\) on \(I_f\).

6. Display Top Four items.

Algorithm for Explicitly Mining Positive Preferences and Negative Preferences
1. Start the timer when item is selected
2. If time elapsed > 15 seconds
   {  
   - If item is in the collection c
     Increment value of visited counter
   - Add item to the collection c and value of counter visited =1.
   }
3. If item is added to cart
   {  
   - If item is in the collection c
     Set value of variable saved=true.
   - else
     {  
     - add the item in the collection c
     - set visited=1
   - set saved=true.
   - Fetch the preferences for which the item would add to cart.
   }
   }
4. If item is deleted from cart set saved=false.

5. If user buy an item then
- Fetch the parameters that match with age-group, gender, cost and hobbies, and store its positive preferences \((Ip)\).

6. Apply the ranking function on collection and store preferences as positive preferences.

Algorithm for Implicitly Mining Positive Preferences and Negative Preferences
1. Start the timer when item is selected.
2. If user close the page in less than 15 seconds
   {  
   - Fetch the parameters from age-group, gender, hobbies and cost that mismatch and add those preferences to collection of negative preferences \((In)\).

3. for the items that are not clicked
   {  
   - Fetch the positive preferences that are matching with age-group, gender, hobbies and cost and add those preferences to collection of positive preferences \((Ip)\).
   - Fetch the positive preferences that do not matching with age-group, gender, hobbies and cost and add those preferences to collection of negative preferences \((In)\).
   }

4. Apply the ranking function and store preferences as negative preferences

4. RESULTS

Login Page
This is the screen shot of the Login page where the user logins. On successful Login user’s explicit data becomes available to the system. By using explicit data of a user, system fetches explicit preferences of a user which can be useful in recommending items to the user according to user’s interest.
Home Page
Home page displays list of items of a website from which user selects item to purchase. User’s implicit behavior is captured as users open any item, added to a cart, removed from a cart or purchase an item.

Recommendation to the User

CONCLUSION AND FUTURE WORK
We have developed a novel model which addresses the issue of cold start problem and data sparsity problem with reference to recommendation system. Our algorithm works in a simpler way than previous machine learning algorithms. Our algorithm finds user preferences based on profile and actions with respect to products such as clicks, time spent, purchases, save and delete. In our algorithm, no need to follow strict order for accessing the objects for preference mining. Thus, in our proposed approach we have tried to use machine learning algorithms to analyse data and information provided by the user and ultimately helping in improving the Business-to-Business community.
In future we planned to introduce more actions to make the algorithm more efficient.

REFERENCES


