Optimized Model of Recommendation System for E-Commerce Website

Fares Aqlan[1] and Abdullah Alqwbani[2]

[1] Central South University

Received: 7 December 2013 Accepted: 2 January 2014 Published: 15 January 2014

Abstract

The purpose of this work is to optimize the recommendation system by creating a new model of recommender system with different services in a global e-commerce website.In this model the most effective data sources are integrated to increase the accuracy of recommendations system, which provides the client more intuitive browsing categories interface.The sources used for this model are the user's searching log on the global website, and data referred extracted from search engines, more clicked URLs, highly rated items, and the recommendation algorithms of new users and new items. In additions, user's interests based on locations, and the hot releases items recommended by the admin or shop owner of the e-commerce website according to the website marketing strategy.When the users browse the website, the data sources will automatically combine to incorporate the derived structure and associate items for each category into a new browsing recommendation interface.

Index terms — ecommerce, data mining, recommendation system, clustering algorithm.

1 Introduction

he global systems internet with World Wide Web has revolutionized the human life like nothing before. Since 1997, the web has progress into a true economy and a new frontier for business [1]. The WWW became more important as a source for the basic data and a place for trading, which called Electronic Commerce (EC).

Electronic commerce includes the use of all kinds of information and communication technology in the business processes among the trade. Moreover, it helps to get a share in the market and improve customer service by creating a Web page and supporting the investors' relations or communicating electronically with customers [2]. Electronic commerce is more than ordering goods from an on-line catalog. It involves all aspects of an organization's electronic interactions with its stakeholders, the people who determine the future of the organization. Such stakeholders include customers, suppliers, government regulators, financial institutions, mangers, employees, and the public at large [3].

Nowadays many sites have a good business and become well known ecommerce sites, such as ebay.com, Amazon.com, taobao.com and others. Business is evenhanded to the process of shopping on the web site. It becomes the way of shopping in wide field including personal need, house need or business need.

Fast growing of Internet technologies presents complicated challenges and opportunities to organizations and guiding them to develop new managerial roles and practices [4]. These explosive developments of the internet and E-commerce technology have led to the daily growth of recommendation systems.

Recommendation systems typically suggest commodities (information, items or services) that are of interest to users based on customer demographics, features of items, and/or user preferences (e.g., ratings or purchasing history) [5]. Recommendation services are used by E-commerce websites to suggest items to their consumers.

Along with EC areas, the B2B (Business to Business) Recommendation system is being spotlighted as an interesting research area considering its size and the potential impact it has overall. Now various recommender systems are being used in seller-centric E-marketplaces, intermediary-centric E-marketplaces, and buyer-centric E-marketplaces etc [6].

However, in many global e-commerce websites, well-defined recommendation systems are not available; moreover, in some other e-commerce sites, the recommendation systems are too coarse and less intuitive to distinguish properties according users interests, which will lead to very bad user experience [7].

To address these problems, in this project we propose building a new model of recommendation system that depends on hierarchical structure for emerging ecommerce products according to users' behavior preference, which can be derived from searching logs and data referred extracted from search engines, highly clicked URLs, top rated items, users interests based on the same area customers, recommendation algorithms for the new items and also the new users. We also create a personalized recommendation strategy managed by the admin of the website.

2 II.

3 Motivation

The E-commerce environment includes all online activities and business operations achieved between multiple parties using electronic techniques.

With the huge development of internet and E-commerce websites; when consumers choose their needs of items and commodities, they confront some serious problems of data overloading. Therefore; many website researches and projects have focused on recommendation system development, in order to provide users more individual recommendation services.

Recommendation system has become serious business tools used by many of the largest commerce websites, in order to provide the users more effective and efficient way to find their interested products. The recommender systems work like salesman who provides users advices and services to help them find the commodities and items they are interested in. However, with the wide use of recommendation services, many common challenges and problems come out, such as real-time, sparsely of information, cold start problem and recommendation quality.

In addition, with the rapid development of web and e-commerce business, a large number of growing user interaction to the application provides a number of very valuable data and information. This interaction forms include users of e-commerce sites click browse, clinch a deal to buy goods online sales and online collection of goods. This increasing interaction behavior leads to the emergence of the information overload problem. In additions, most recommender systems still meet some serious problems and challenges, such as sparsely data, real-time, cold start and the quality of recommendation results ??8].

Therefore need a system that provides services which provide solutions to overcome these common problems by using the interactive information to find user interests and preferred orientation with high quality and real-time techniques.

The goal of this project is to build a new model of personalized recommender system. We have proposed and applied some data mining ideas and clustering algorithms that optimize the recommendation services on a global E-commerce websites.

Our optimized recommender system helps consumers to find their needs and save their efforts and time in complicated operations. For e-commerce sites, our ideal personalized recommendation system will directly increase online sales of commodities brought in, increase the orders-size by turning browsers into buyers.

4 III.

5 Purposed Recommender System

The traditional recommendation technologies have their own advantages and also many shortcoming points [9]. So to solve these issues we have build a new model of recommender system which based on hybrid recommendation techniques and combined with data mining clustering technology to overcome the shortcoming points and provide the best recommendation results which meets all kind of users' interests and needs.

Our system belongs to a complete personalized recommender system, using data mining combined with hybrid recommendation methods. The new model of recommender system provides more adaptive and scalable services; as it is highly considering the recommendation quality, real-time recommendation, and proposed solutions for problems such as cold start and other issues. In the following we introduce the architecture structure, project algorithms and technologies of our E-commerce recommendation model.

6 a) Basic Architecture of Recommender System

The tremendous development of the Internet has led millions of companies to set up shop on the Internet and over 100 million consumers are eagerly participating in the global online marketplace [10].

From this quick development of e-commerce websites; we start to get destruct with the recommendation systems methods and advantages to meet users' need and interests. The enhancements of this project are designed to meet such needs including the recommendation functions and site features.

The recommendation functions are designed to provide the users the ability to discover their real interest ed items with flexibility and high efficiency, which will sav e users' own efforts and time.

Our recommendation system include five parts of functions, first part recommend the items which will be derived from user's searching logs on our website and data referred extracted from search engines, through the searching log and search engines are considered to discover user's attributes and interests.

The second part of our model functions include the most rated items; the convenience of this function is to compare products through the multi-products website, which save more time and effort during all customers' visits.

The third part of our system propose and apply some algorithms which will recommend the new items of the website, as well as, some algorithm to recommend items for our new users. The advantage of this part is to solve the cold start problem of recommendation system.

The forth part of our new model propose the algorithms which do the recommendation according to the user's interests based on locations, our system collect the interest data of same location users, as different location users have different interests; since each location has its own habits, needs and life traditions. But through this function, the users easily can find the most interested items by his location users on our global website.

The last part of our recommendation system model includes the items which can be recommended by the admin or shop owner of the website according to the website marketing strategy.

7 Enhancement

To enhance our new model of recommender system, first we have to enhance the relational database. The enhance operation is divided into two parts: The process model step which focuses on the operation process of database stored data. Second step is the online recommender which analyzes the recommendation type of system, as well as, the recommendation algorithms used and proposed for this project.

8 a) Process Model

The core of recommender system is the recommendation algorithms models, as a different algorithm requires different data, so the system needs to manage the input data to provide a high quality of output results. The main data as shown in logical schemas figure above include: User, Item, and Rating. Due to E-commerce website deals with a huge amount of data which growing rapidly, it makes the algorithms model take a long time, and a big consumption for system resource. That seriously affects the real-time recommendation.

As a result, the recommender system using offline process model to output results. And online recommendation model then uses the output results with the system input data to recommend items for the user.

The process model based on the incremental updates of input data, so when the new ratings data of users reach a certain limit value, it needs to deal with process model again.

9 ? Data preprocessing

According to different algorithms' required data, the system deals with insert data using input data model.

10 ? Model calculation

The recommender system according to data amount updates, regular operates models, calculates the update data, modify the model output results, to ensure the quality of recommendation.

The process model of our recommender system can be displayed as it shown in the following figure 3 The personalized E-commerce recommender system mainly used to recommend items for users based on their interests. The main functions of online recommender are to analyze the recommendation type, and choose the related input and output data of algorithm model, to predict recommendation results, and provide it for users. The main process of online recommender is as shown in the following figure 4: The online recommender uses a real-time recommendation model to provide a high quality recommendation. When a user login the E-commerce website, and browse items, the recommender system reads his/her profile data, user rating data and purchased log to predict interested items, and feedback direct to the user the Top 10 items that user most likely interested in. techniques. Specifically, we use the STC algorithm to analyze the data mining of search engine and search log data. We have also applied neighbor clustering algorithm to complete data mining work as a clustering technique for ratings data. For the classification algorithm, we have applied support vector machine (SVM). We have also proposed some matrixes that determine the users' locations in order to provide recommendation results based on location.

In the following, we introduce the project algorithms and its applications technique.

11 i. STC Algorithm

The STC algorithm clustering that has been applied in previous work **??**12] is an efficient method of clustering search results, but because it's clustering process only start from the characteristics of the document itself, and it gets the clustering results based on the document attributes. So for our best knowledge, this is not enough for a personalized recommendation system. In this project, we combine the user personal interests' model with on STC algorithm, which improves the STC algorithm strategy.

Suffix Tree Clustering (STC) has three logical steps: (1) document ”cleaning”, (2) identifying base clusters using a suffix tree, and (3) combining these base clusters into clusters.

After the suffix tree construction, each node on the document can be used as a base cluster. So as to reduce the clusters numbers, we have to combine some base clusters into a big cluster, this process called ”Combine Base Clusters”. In order to better implementation of the personalized recommendation, the clusters should be ordered according to the user’s interests.

To measure the user interest into any document, we use the following formulas which show the steps of our recommender system technique using search data identifications:

12 Basic Data Construction

Using the Google engine to query on a keyword, the results will show many pages which include this keyword inside its contents. For the search results of Web page, we use the data structure to explain the steps of operation:

Struct Step1: Read FileName into memory Step2: remark snippets, if it is ”<HEAD>”, then proceed operation onto the head of Web catalog file. Or

Step3: remark snippets, if it is ”<BODY>”, then apply operation onto the body of Web catalog file.

Step4: Return CatalogSnippetList;

The Web file is semi-structured data, so to facilitate process, we need to structure the data, and clean all the return results. After the data cleaning operation, we get a list that contain all search results, so we move to next step, clustering analyze.

13 Clustering Analysis

The clustering analyze process will return a large number of search engine data, such as catalog snippets, and then divide it into classes or small clusters. Make the most similar objects into one cluster, and different data objects into different clusters. By comparing the cluster methods, we decide to use an improved STC algorithm method as basic clustering algorithm for search data of our personalized recommender system. Specifically, there are three steps to improve STC algorithm: 1. Create suffix tree structure, so we add each complete cleaned catalog snippet into the suffix tree. 2. Determine the base clustering.

14 Combine the base cluster with clustering results.

The improved STC algorithm combines the cluster results with user interest profile data to provide sorted cluster results.

15 Personalized Recommendation Strategy

The clustering analyze of search results will provide better clean and sorted information, as the improved STC algorithm did implement the measurement of similarity on base clustering combined(D D D D) Year 2014 E 1.

16 2.

3.

17 4.

with content-based technology, as well as, they process the cluster results as sorted data.

These kind of results and techniques help to return users more specific recommendation according to his search information collected by our algorithms, it also arrange the results as Top N more interested and searched items to provide it on the recommendation system interface.

ii. Neighbor Algorithm

The clustering analyze used to divide the stored data of database into significant sub classes. This classification operation is based on the similarity and difference between data.

The algorithm function of neighbor clustering can be constructed as follows:

For given finite sample set {U}, that includes n samples, assign a number C of clusters where?K i,j = 1,2, A, C?

For each model, if the sum of sample’s distances to the cluster center achieves the minimum value,

The mathematical model of clustering can be given by:min ? ? ?U ? v j ? U?K j C j=1 v j = 1 ? x ij n i=1 ? x ij n i=1 U

Where C is the number of clusters, v j is the mean vector of sample j.

So if the model sample i assigned into the centre of cluster j ,Then x ij = 1; else x ij = 0; ? x ij n i=1

= 1 means that model sample i only can be assigned into centre of one cluster.

The clustering analysis classifies models according to the closeness degree between samples features. The basic similarity has the following two functions:

1. Distance Function Sample uses 13 d of features variables for description; each sample can be seen as a point in the empty space, using some distances to indicate the similarity between sample points. The closer sample points, the more similar features they have, and far away distance between different sample points.

So the distance function can be displayed using the following formula: For non-negative conditions, ð ??"ð ??"(??, ??) ? 0; ð ??"ð ??"(??, ??) = 0; and for Symmetry we have ð ??"ð ??"(??, ??) = ð ??"ð ??"(??, ??); which meet the triangle inequality ð ??"ð ??"(??, ??) + ð ??"ð ??"(??, ??) ? ð ??"ð ??"(??, ??).

2. Distance measurement method using Euclidean distance:ð ??"ð ??"(??, ??) = ? ??? ??1 ? ?? ?? 1 ? 2 + ??? ??2 ? ?? ?? 2 ? 2 +?? + ??? ???? ? ?? ???? ? 2

Where ?? ?? = (?? ??1 , ?? ??2 , ??, ?? ????) and ?? ?? = (?? ?? 1 , ?? ?? 2 , ??, ?? ????) are two n-dimensional data objects.

If each attribute of data is given a weight, then the weighted Euclidean distance is expressed as:ð ??"ð ??"(??, ??) = ? ?? 1 ??? ??1 ? ?? ?? 1 ? 2 + ?? 2 ??? ??2 ? ?? ?? 2 ? 2 +?? + ?? ?? ??? ???? ? ?? ???? ? 2 3

. Similarity coefficient: The two sample points are more similar, the similarity coefficient is closer to 1; and the simiarity coefficient is closer to 0 when two sample points are more different. Phase angle cybermetrics: Using vector of included Angle cosine formula to measure the angle's similariy degree between samples ??(?? 1 , ?? 2 , ??, ?? ??) and ??(?? 1 , ?? 2 , ??, ?? ??). The angle cosine formula is:??????(??, ??) = ? ?? ?? ? ?? ?? ?? ??=1 ? ? ?? ?? 2 ? ?? ?? 2 ?? ?? =1 ?? ??=1

Pearson correlation coefficient:

The correlation coefficient of sample i and sample j is as the following:??????(??, ??) = ? (?? ???? ? ??? ??)(?? ???? ? ??? ??) ?? ? ? (?? ???? ? ??? ??) 2 ?? ? (?? ???? ? ??? ??) 2

18 ??

Where ??? ?? is mean value,??? ?? = 1 ?? ? ?? ???? ?? ??=1

, and??? ?? = 1 ?? ? ?? ???? ?? ??=1

iii. Support Vector Machine (SVM)

The support vector machine is used to classify data; this task is called machine learning [13]. For given data points which belong to one or more classes, we use SVM to decide which new data point will contain the class.

Suppose x 1 , x 2 , A, x n , where x i ? R d , i = 1, A, m are d-dimensional training samples. The corresponding mark of each sample is y 1 , y 2 , A, y n , where y i ? {1, ?1}, and i = 1, A, m indicating the class to which the vector belongs.

For linear SVM, the hyperplane w ? x + b will classify the training samples, then?? ? ?? ?? + ?? > 0 ??ð ??"ð ??" ?? ?? = 1 ?? ? ?? ?? + ?? > 0 ??ð ??"ð ??" ?? ?? = ?1

This can be rewritten as:?? ? ?? ?? + ?? ? 1 ??ð ??"ð ??" ?? ?? = 1 ?? ? ?? ?? + ?? ? ?1 ??ð ??"ð ??" ?? ?? = ?1 (D D D D D D D D) Year 2014

So according to the theory, hyperplane can classify the samples, and also maximize the distance between the classes. In the following figure, we have three hyperplane (H 1 , H 2 , H 3), we can see that H1 does not separate the classes. H2 does, but only with a small margin. H3 separates them with the maximum margin. The following figure 5 shows the maximum margin hyperplane: problem. This optimization problem has been solved by the saddle point given by (Christopher, 1998):M(w, b, a) = 1 2 (w. w) ? ? a i m i=1 {[(x i . w) ? b]y i ? 1}

Where a i is Lagrangian multiplier. According to above saddle point, we have:w = ? a i m i=1 y i x i

Which declare that only a few a i will be bigger than 0, x i is the support vector that lie on the margin and satisfy condition y i (w? x i + b) = 1.

So by substitute the above formula, we get the following points which show that SVM reduces to the following optimization problem: Maximize (a i) we get:W(a) = ? a i ? 1 2 m i=1

? a i a j y i y j (x i . x j i,j

)

And to the constraint from the minimization:a i ? 0, i = 1, A, m ? a i m i=1 y i = 0, i = 1, A, m

So W can be computed by:w = ? a i sv y i x i , a i ? 0 Where sv is the support vector.

As well as, in the hyperplane function, for constant C, it can be displayed as the following:C = 1 2 ??w. x * (1)? + ?w. x * (?1)??

Where x * (1) declare that belongs to first class of support vector, and x * (?1) declare that belongs to the second class of support vector.

According to above, we get the function of the best classification hyperplane as follows:f(x) = sgn(? a i y i (x i sv . x) ? C) ,

x i is the support vector, a i is Lagrangian multiplier, and C is constant.

Above we have described the training samples classification by using linear SVM, also the support vector and the basic principles of best hyperplane. But if the training samples cannot be classified by linear SVM, then the above principles will be useless. In this situation, we use soft margin to solve problems, soft margin will choose a hyperplane that classify samples as cleanly as possible.

For non-negative slack variables ? 0, i = 1, A, m, so the function becomes:w ? x i + b ? 1 ? if y i = 1 w ? x i + b ? ? 1 if y i = ?1

Then by using Lagrangian multiplier, the optimization problem can be computed by: For, 0 ? a i ? T , ? a i i y i = 1 By minimization we get:? a i ? 1 2 m i=1

? a i a j y i y j (x i . x j i,j

)

Where i = 1, A, m, T > 0 is a constant.

19 iv. User's Location Matrix and Algorithms

There have been some previous works into geolocation technology and software which determine the user's geographic details including country, city, ZIP code, and so on.

The user's location information is effective for recommendation system to provide more specific recommendation results according to the user location interest and preference identification. Since our recommender system builds a preference or interest profile for each user enter our website, so our recommender system will use the user's interest profile to create session-interest matrix to indicate the user's interest based on user's location.

To create the aforementioned session-interest matrix, we need to process the following three steps: ? Session-IP Scope Matrix The system generates all users' session IP address from user session data identification. Then our Support Vector Machine (SVM) will classify all IP addresses in some classes according to the first two segments of session-IP scope list. By creating this matrix, we use value 1 for each session user location in the matrix, so each row contains only one value as 1 and others take 0. ? IP scope-Interest matrix We create this matrix, in which the columns represent users' interests based on aforementioned user profile data, and its rows represent the same IP addresses of session-IP scope matrix created in step 1. The IP scopeinterest matrix indicates the highest interest of website users according to their behavior and experiences on our E-commerce website. To fill the matrix, we use 0 and 1 numbers to make its rows and columns represent user session and his interest value. ? Session-Interest matrix In this part, we create session-interest matrix by multiply the previous obtained matrixes in step 1 (Session-IP scope matrix) and step 2 (IP scope-Interest matrix). The following steps show the method of creating this matrix: } }; According to these steps, we have used values 0 or 1 to fill the elements of session-interest matrix, the output information is a referrer matrix based on user's location and his/her interest profile data. Since every customer who visit our website has an IP address, but not all users have interest profile such as new users who do not have any rated information or purchased data. So to solve this issue, we have process the classification and clustering algorithms on two matrixes, one for users who do not have interest profile. For such users, we use the data integration based on similar users coming from the same location. The system generates the interest items for users who have similar IP addresses with our current user.

For this process, we use k-mean clustering algorithm to generate identification data based on clusters of same location users according to the classification on session-IP scope list.

20 v. Algorithms Integration

The recommendations based on STC algorithm, user profile, neighbor clustering, IP session matrix and support vector machine (SVM) will combine the item's features with user preference. These algorithms will also divide the items according to difference features and catalogs. Then summarize the user's preference value on these different features with measuring their interest into item's lists, until we get the user preference model. According to user's different interest, we use the userprofile data with STC algorithm to measure user's interest by ordering clustering based on their interest model data. This process will integrate the search engine data and user search behavior on our website, in order to generate their interest's information and build an interest model for each user.

For user preference model, we use neighbor clustering, which generate the users who have similar preference level into different features of items. These users became neighbors, to provide real-time recommendation. In addition, combining with contentbased recommendation technology can promote the recommendation of new items.

So considering adjust and analyze the user information; when making recommendation, and choosing a nearest neighbor; it will help to make the user similarity comparison results as weighted to provide a high quality recommendation. The project will uniformly process the user's information, in order to facilitate the comparison.

By combining the neighbor clustering algorithm of content-based technology, we get the clusters units, and then compare the users of each cluster unit to get the similar users. So the hybrid recommendation by combined user information is recommended on the basis of content recommendation technology.

For the classification results of support vector machine (SVM), it helps to predict the user's nearest neighbor, and proceed a weighted adjustment operation, to further improve the quality of recommendation. Year 2014

21 E

In the E-commerce business, a user buy items not only related to item's features or preference; as the user's basic information (age, occupation, location,? etc) have also a certain relevance.

Because of the reliable results of prediction recommendation based on demographic information for limited data volume; so after the classification operation by support vector machine, we get the similarity degree between users according to the comparison results of users' information. These similarity degrees are used as weighted values for predictive ratings process.

We also have used IP session matrix with support vector machine (SVM) to classify and divide the users according to their locations, and use the identification data based on location to provide user a useful and helpful recommendation including the most interested items by same location users.

V.

22 Implementation a) Identification Based on Search Data

The following steps show our recommender system technique using search data identifications: i. Update User-Interest Profile

The user interest profiles are automatically generated based on the type of content viewed by the user. A system generates user interest profiles by monitoring and analyzing a user's access to a variety of hierarchical levels within a set of structured data.

User's interest is constantly changing, so the update of user interest profile based on user interests must be considered. The retrieve information of user input as the sources information for user interest profile updating process. ii. Improve STC Algorithm

The STC algorithm clustering that has been discussed above is an efficient method of clustering search results, but because it's clustering process only start from the characteristics of the document itself, and it gets the clustering results based on the document attributes. So for our best knowledge, this is not enough for a personalized recommendation system. In this project, we combine the user personal interests' model with on STC algorithm, which improves the STC algorithm strategy.

In order to better implementation of the personalized recommendation, the clusters should be ordered according to the user's interests. To measure the user interest into any document, we use the following formula:

Score(c i)= ? count?l j , c i ?xw j M j=1

Where count ?l j , c i ? as the occurrences number of j keyword l j into i document c i for user interests model. w j is the weight of l j .

To combine base clusters we use Single-Pass algorithm which has a better timeliness compared with Single-Link algorithm.

The basic process of Single-Pass process is as follows:

1. Assign the ?? 1 cluster ?? 1 2. For i=2 to N do (a) Calculate similarity ?? ???? between ?? ?? and ?? ?? for all j. (b) Find the cluster j with largest similarity ?? ???? between ?? ?? and cluster j. (c) If ?? ???? > threshold, then assign ?? ?? to cluster j and recalculate cluster representative for j, else create a new cluster for ?? ??

Cluster representative status such as if the cluster is represented by its centroid.

Here, use the user interest degree to measure the similarity of different documents, similarityS ij . And use the Score value average of document cluster as the cluster centroid. The process steps are as follows: Traverse each base cluster queue, and convert base cluster into one document; Measure the Score value of each document;

Use Single-Pass cluster algorithm to combine all the original base clusters of the same document cluster; order the results according to centroid value;

By comparing the results of user interest's measurement, the Single-Pass algorithm we have used in this project to combine base clusters did improve the algorithm efficiently compared with Single-Link algorithm used by previous works. As an implementation result for the above algorithms on search engine data and search information on our E-commerce website, we could check all the info we need via the management system as we see in following figure 6 items ? ETC 1 1 1 ? 0 2 0 0 ? 1 ? ? ? ? ? K 1 0 ? 0 (D D D D D D D D)

Year 2014

23 E

24 By combining the ratings table and items features table, we get the following table:

According to the data analyze of search data, our recommender system generates all the information need to recommend items based on search engine data, query keywords and clicked URLs on the website.

25 Identification Based on Rated Items

Recommendation system can be defined as a program that predicts a user's preferences using information about the user, other users and the items in the system.

According to our database figure 2, we can see three tables for this section, which are users table, items table and ratings table. The rating table data explain the items rated by users and ratings degree, the following table shows the process of the rating: a) 0 ? 4 User 1 5 0 1 ? 1 User 2 1 1 0 ? 2 ? ? ? ? ? ? User N K 1 0 ? 1

Table 3 shows the rating value of each item and the user who rated the item. As the above info show the user rating and items features, but it doesn't reflect the user interest into different features of items. So we need to convert the user ratings of the items, make each rating value declare the interest degree of each feature, and then explain the user interest into different features of items, as shown in the following steps: ? Initialize relatively matrix of user preference Create user preference matrix CP, the matrix row include eachuser, the column show the Eigen's, and the values on the matrix: This algorithm can be used to recommend the new items of E-commerce website, although the new items don't have ratings or ordered information's, but the features of new items can be used to compare and find the most similar items on our product catalog, then find the interested users on it, and recommend the new items to them.CP ij (i = 1,2, A, N, j ? 1,2, A, M)

First, for the new items features, we create a feature matrix called TestItem form, as a new item to follow up the types, in the form of 0-1.

Second, preferences match of relative users: After the classification of relative preference on behalf of the matrix itemAvg & TestItem, we measure the D-value E, which is integrity, the greater we got, declare the new item classification belongs to the user preference or interests. Then select the classification user to do recommendation. One of the cold start problems is new users who have not any interesting data or purchased items log. So to do the recommendation for these new users, we use the following method which uses the information of users to find their similar neighbors in order to give them high quality recommendation contents.

We compute a feature weight. Each feature weight is calculated separately for each user. ? Users Input Data: as we can see on the database figure 2 , the new users data can be called newuser, and the original users data called UserInfo. ? Feature Weight Calculation, For each user, we assign a weight to each feature in a feature set based on the particular user's past behavior.

26 Comparison between new user and original user

For new users, because we don't have any ratings information or either purchased data, so we can't recommend items according to user interests or content based classification. So we use the new user's data to compare and find the similarity with other original users, and then according to the similarity degree do a prediction rating for the new user. The similarity degree between users can be calculated according to the following formula:?? ?? = |?????????????? ?? ? ????????????ð ??"ð ??"?? ?? | max ?? ? min ??

This formula declares the similarity between users info on the term t.

27 Weight Calculation

Considering all user information terms, according to different extent a , calculate the comprehensive weights of similarity degree between on behalf of the user as the following: The user location identification data based on the input of two clustered matrixes, which are session-IP scope matrix and Interest Scope matrix, as well as, the user session vector.W = ? a i 3 i=1 S i , of which ? a i

The output is a list of recommended items for the user based on his location identification, that represented by session-interest clustered matrix.

When the user views our website, the recommender system algorithms based on location will process the following steps to recommend a list of interest items: First, the system construct session-IP scope matrix for the current user, to determine the location that user belongs to. The matrix row is filled by value 1 for each session user location in the matrix, so each row contains only one value as 1 and others take 0. Second, the system create IP scope-interest matrix, the columns values used to represent the user interest and preference according to his profile data, and its rows represent the same IP addresses of session-IP scope matrix; Since it has one row, so it's also called the user's location vector.

Third, by multiply the two clustered matrixes, we get a new matrix called session-interest matrix, which used to indicate the user location and interest data. Then according to matrix values, the system recommends items for user depend on his location.

Fourth, the user visit our site and he may has an interest profile or not, so considering this point, we use k-mean classification algorithm to find the closest neighbor for user, and recommend items for user based on his neighbor interest. The classification algorithm (KNN) calculates the similarities between users to provide the current user a list of most interest items by his same location users.

According to algorithm calculation, the more much value is gotten, the more similarity of user profile for our current user session. The recommendation weight for current user session is obtained, and the more much weighted value is obtained, the more prioritization of interest items to recommend user. The implementation results of recommendation matrix based on location give us clear information for user's location and a whole picture about their visits log to our website. The following figures 9 and 10 show the results via management system of our application

The recommender system generates the location address of users according to stored data by the visits overview of system. The relational database provide a detailed address including city, province and country for each user browse our website. The following figure 10 indicates the traffic trends of users via management system: e) Prediction Accuracy Evaluation ? Prediction accuracy of rating methods

28 Mean Absolute Error

MAE is a quantity used to measure how close predictions are to the eventual outcomes, it measures the error between new user's predication ratings and the ratings data of real original users. The smaller value of MAE outcome, the better quality of recommendation system we got.

29 ?????? =

? ? (???????????????? ? ????????????????????) ?? ?? =1 ?? ??=1 ?? * ?? Where Testrate is user rating matrix, TestResult is user's predication rating matrix, N is the number of users, and K as the rating terms number.

30 Conclusions

The goal of this work is to enhance and optimize the recommendation system of E-commerce website by providing a new developed and useful model of recommender system. Our system provides some new functions to solve the main serious problems and challenges exist in the current recommendation systems. It provides functions that meet the user and consumer expectations and needs, taking the full consideration of online recommendation system development from the following points: 1. Enhancing recommendation results based on search engine, search data and clicked URLs. Our system use some clustering algorithms to generate and enhance the user search experience in order to build a user interest and preference profile. 2. Enhancing the rating functions by proposing some clustering methods to enhance the functions of rated items which will generate these data as sources to provide a high quality of recommendation results. 3. Proposing solutions for current recommendation system, such as cold start problem. Our system proposed and applied some algorithms which provide solutions for new items and new users recommendation. To measure the recommendation quality standard of the system, there are many methods to use; such as Precision vs. Recall, Clicks, Click through rate and direct user feedback...etc [14]. Here we use two main methods to test the quality and accuracy of recommendation algorithms.

4. Proposing a new function for recommendation system, as our system build a new interface which provide recommendation results according to the data identification based on user's location. ASP.NET is used as a programming language to build this project, and ORACLE is used as a database engine. By the extra services that our E-commerce application site renders; it will be more flexible and efficiency to use comparing with other similar internet (B2B) sites. [1] [2]

1

Figure 1: Figure 1 :

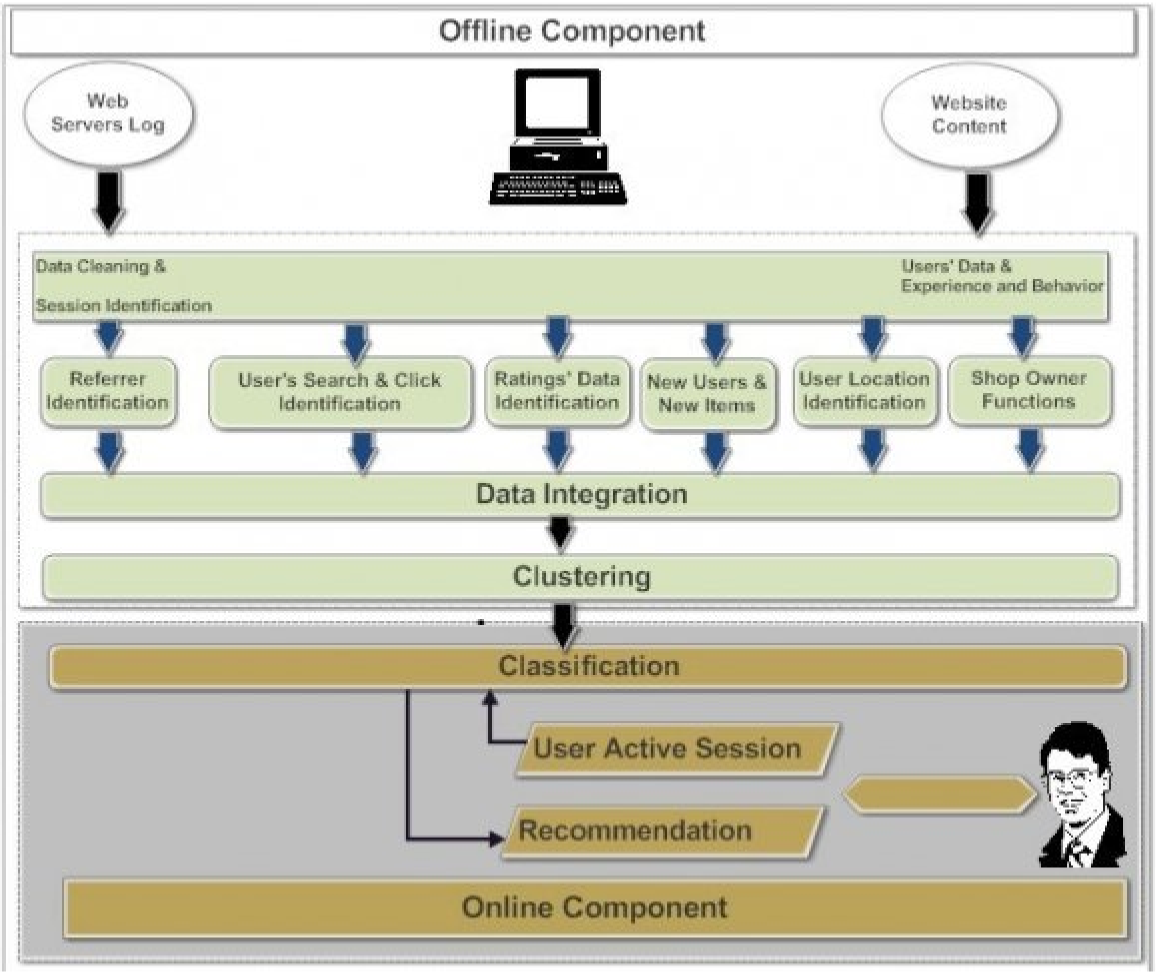
2


Figure 2: Figure 2 :

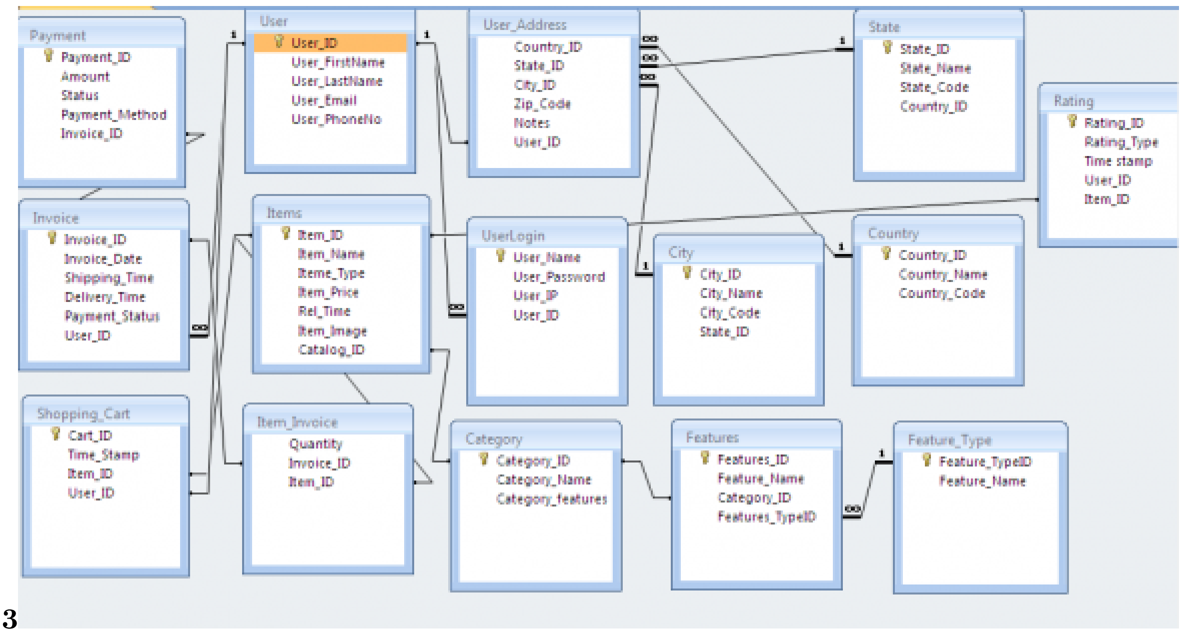
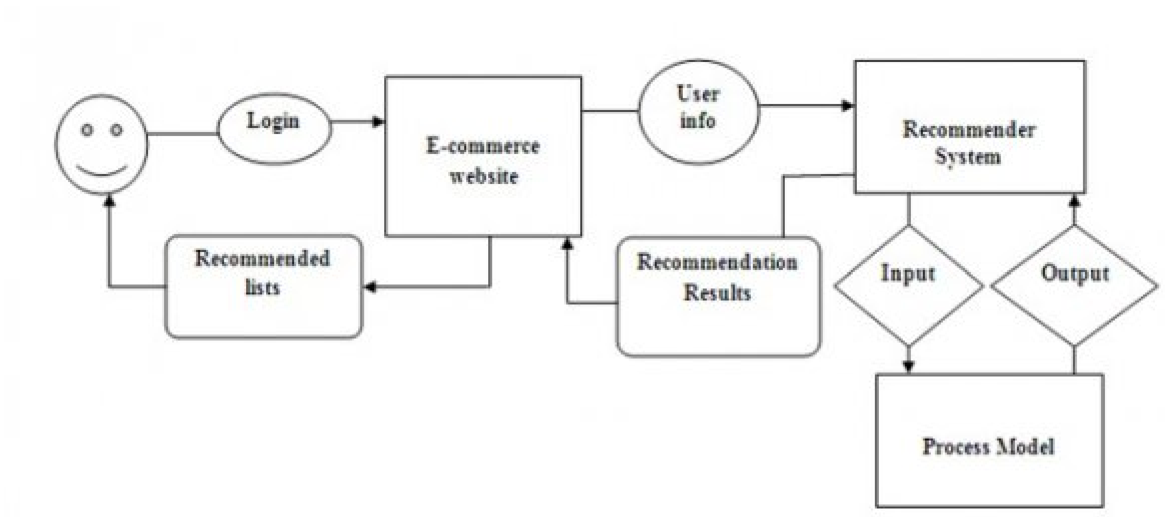
3


Figure 3: Figure 3 :



4

Figure 4: Figure 4 :

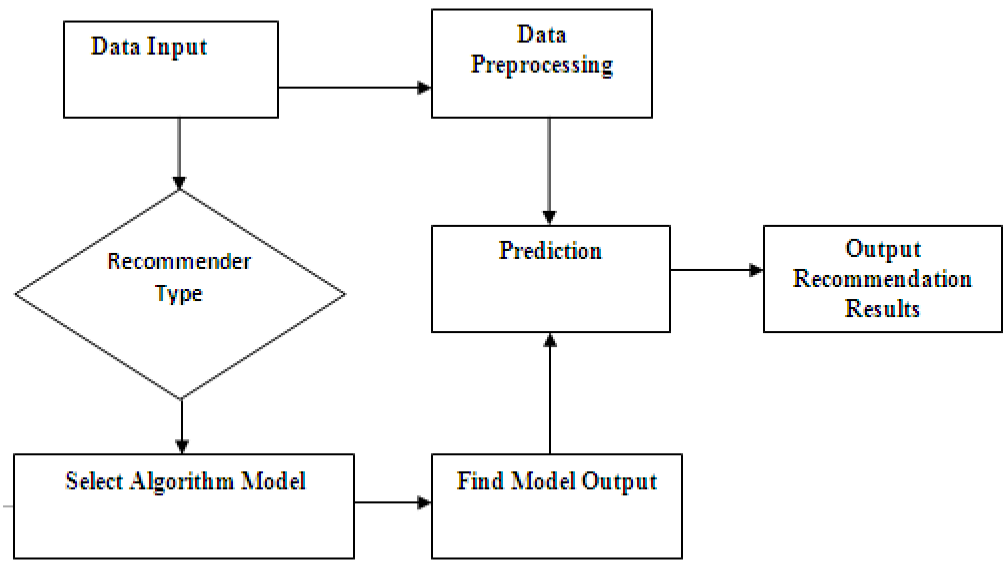
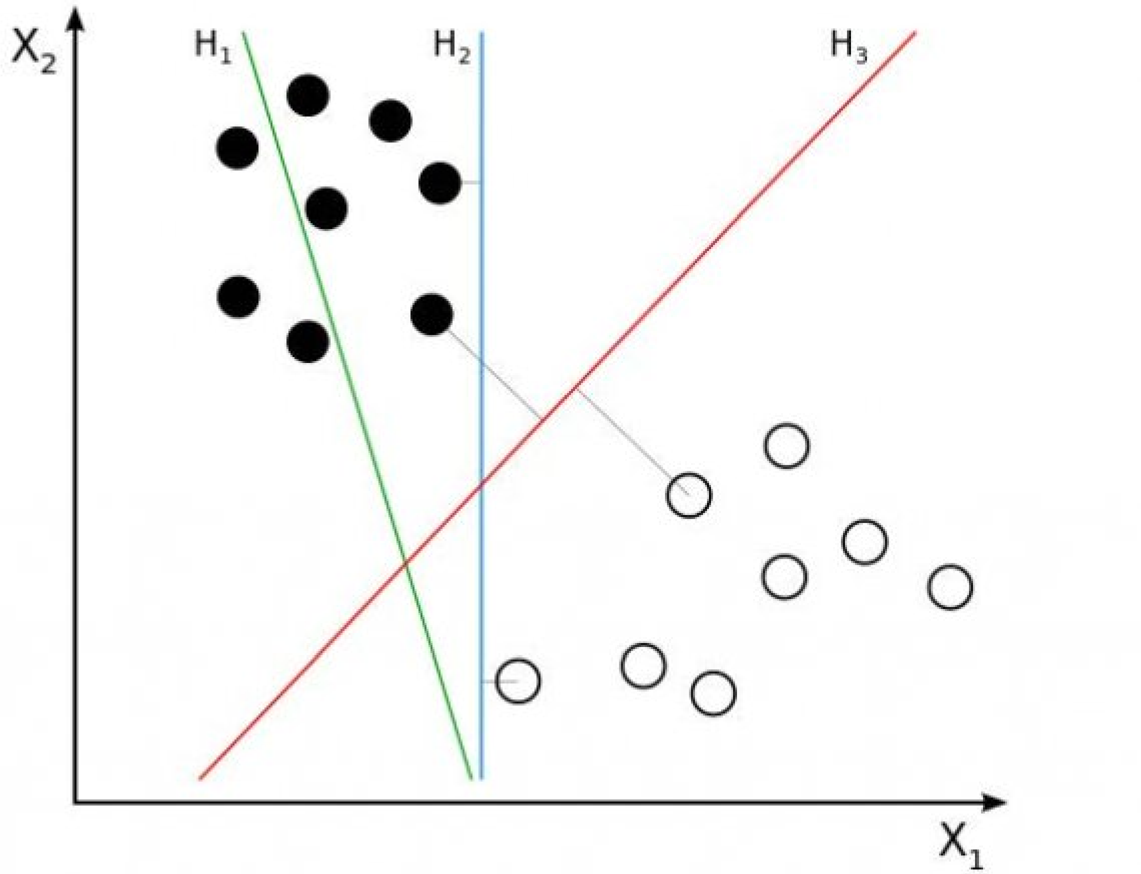


Figure 5:



5

Figure 6: Figure 5 :

ξ_i

Figure 7:

6ξ_i

Figure 8: :Figure 6 :

7ξ_i

Figure 9: ?Figure 7 :

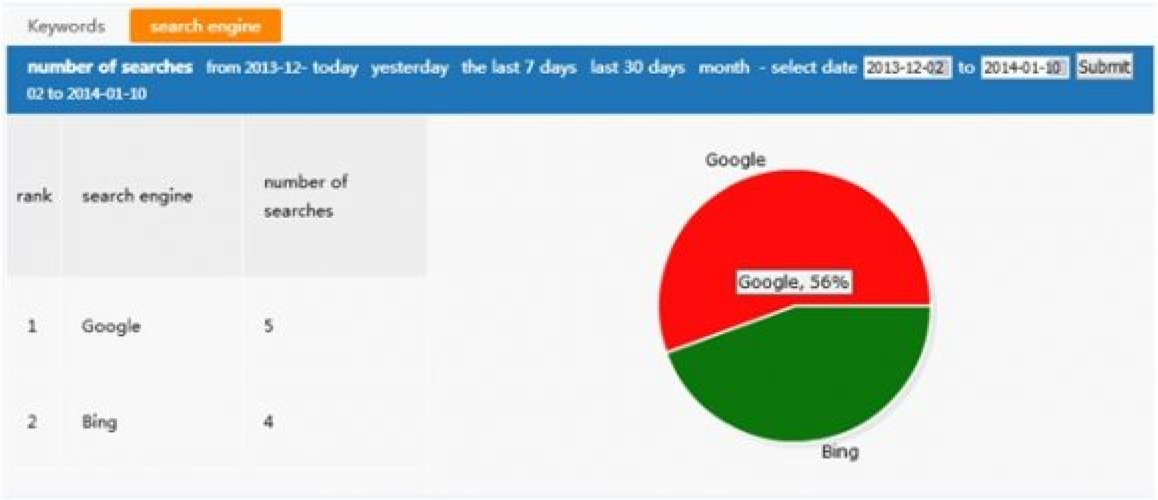
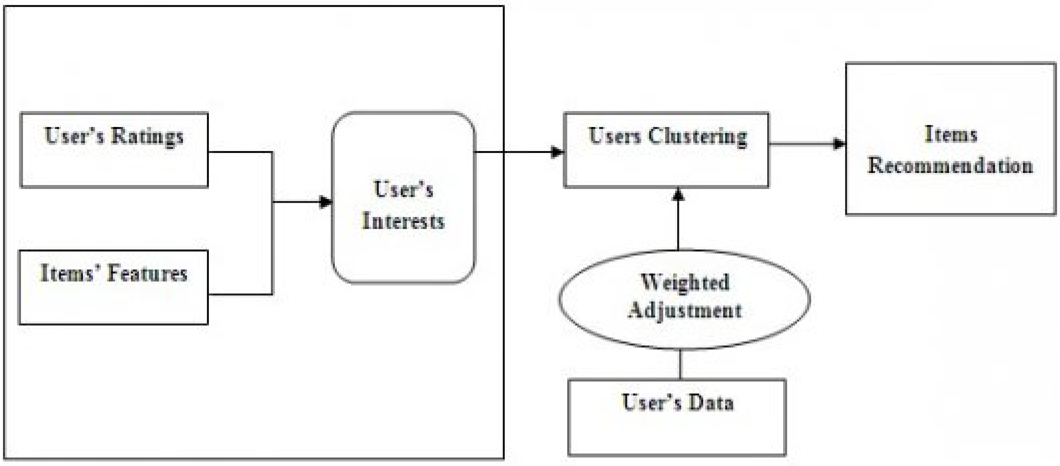
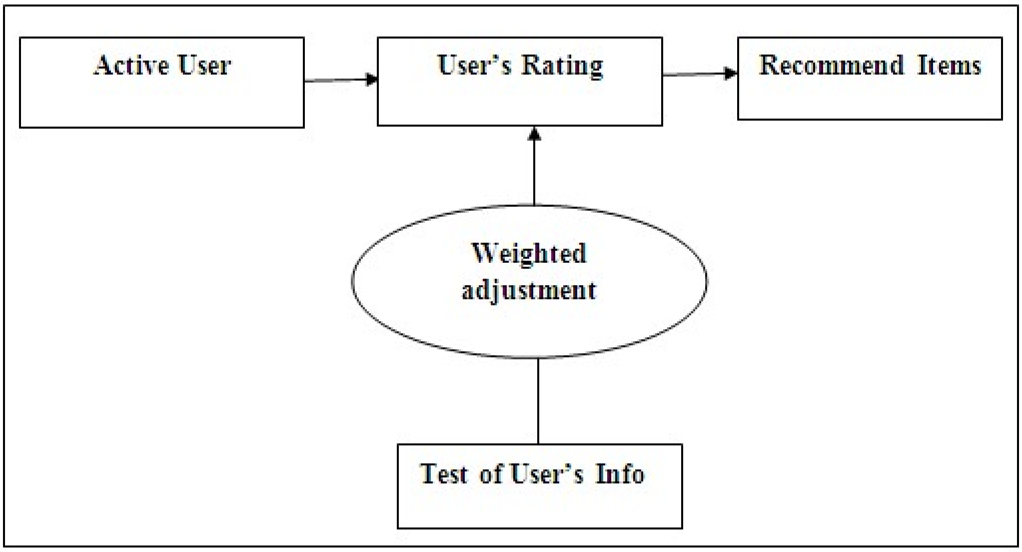


Figure 10:



8

Figure 11: Figure 8 :



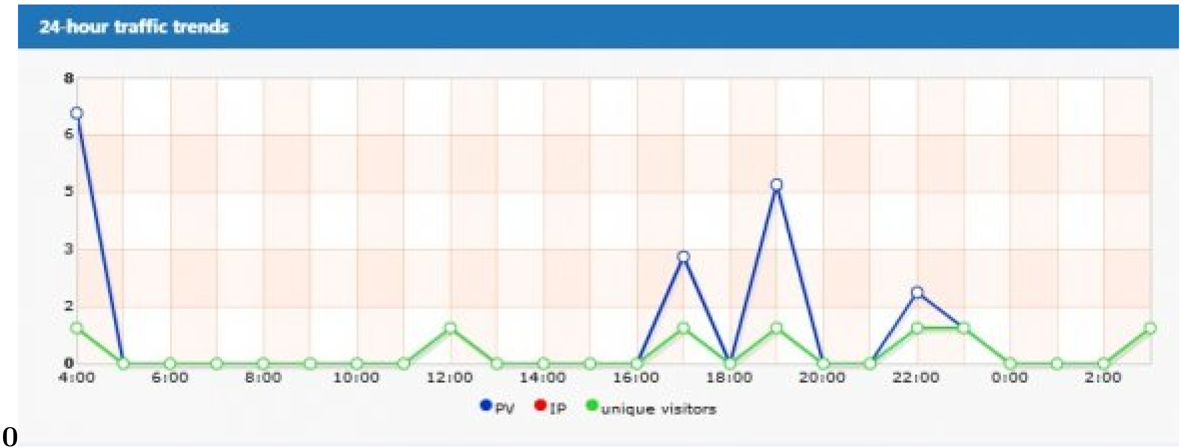
9

Figure 12: Figure 9 :

Visits Overview				view all
	PV	unique visitors	IP	per capita Views
today	1	1	1	1.00
yesterday	22	8	8	2.75
average daily	31	6	6	

2

Figure 13: 2 .



10

Figure 14: Figure 10 :

The profile update algorithm is as
follows:
Input : search query, user interest ?? ?? Output: User interest ?? ?? Method:
Step1: Extract keywords from search query;
Step2: Define Constant c, 0 ? ?? ? 1 ?? ? ;
Step3:
For (Keyword L of query)
{
if (keyword L in group{?? 1 , continue;
}
Else
{
extract node (?? ?? , ?? ?? =min{?? ?? | 1 ? ?? ? ??});
if (c>?? ??)(L,c)? (?? ?? , ?? ??) else continue;
}
}
step4: ?? ?? units;
Step5: Return ?? ?? ;

Figure 15:

1

UserID	ItemID	Rating
User 1	1	5
User 2	2	4
User 3	3	2
?	?	?
User N	K	I
Every item has its own features; the items table data declares the features of items such as item ID, title,	price, item type and so on. The following table shows items Eigen's data:	

Figure 16: Table 1 :

2

ItemID	Home goods	Technology

Figure 17: Table 2 :

3

UserID	ItemID	Home goods	Technology items	?	Rating
User 1	1	1	1	?	5
User 1	2	0			

Figure 18: Table 3 :

Show the interest value of user I into Eigen j, convert into 0 matrix. ? Calculate the interest value of each user into the corresponding Eigen or feature. According to combination between user ratings and items Eigen's shown in table 3; input the data of matrix P, as the following: 1. Generate the rating matrix of user i, e.g. the row 1 equal to sub matrix of i ?? ?? ?? ?? = ?? ?? ?? ?? ?? ?? ?? ?? ?? ð ??"ð ??" ?? ?? ?? ?? ð ??"ð ??" ?? ?? ?? ?? ?? ?? The row 2 of P i shows the user rated items ID 1,2,5, the last row as the user entirely ratings value for each item ID 5,4,1, the other 3-6 rows show the features of item ID 1,2,5 2. Generate last row of ?? ?? , which means the user entirely ratings value of items, and then respectively times the row of items Eigen's value ???? ?? = ?? ?? (: , ??: ?? ? ??) * ?? ?? (??) ð ??"ð ??"ð ??"ð ??"ð ??"ð ??"ð ??"ð ??" ??ð ??"ð ??"; ???? ?? = ?? ?? ?? ?? ð ??"ð ??" ð ??"ð ??" ?? ?? ?? ð ??"ð ??" ?? ?? ?? ?? ð ??"ð ??" ?? ?? ?? ?? ?? ?? ? ?????? ?? = (?? ?? ? ?? ???? ? ? Iterate Steps (1) to (3) until we get the interest data ?? ???? ? ð ??"ð ??" ð ??"ð ??" ?) of every user, we get the user interest data matrix CIP i . Each column data is divided by the sum of all row vectors multiplied by the general users:

Figure 19:

[Wei et al.] , A Survey Of Recommendation Systems In Electronic Commerce Chih-Ping Wei , Michael J Shaw , Robert F Easley .

[Hyndman and Koehler ()] *Another look at measures of forecast accuracy*, R Hyndman , A Koehler . 2005.

[Content-based Recommender Systems: State of the Art and Trends Pasquale Lops, Marco de Gemmis and Giovanni Semeraro] *Content-based Recommender Systems: State of the Art and Trends Pasquale Lops, Marco de Gemmis and Giovanni Semeraro*,

[Nikhilesh Dholakia, Ruby Roy Dholakia, Detlev Zwick, and Martin Laub (ed.)] *Electronic Commerce and the Transformation of Marketing*, Nikhilesh Dholakia, Ruby Roy Dholakia, Detlev Zwick, and Martin Laub (ed.)

[Christopher ()] 'Great Britain: Financial Times / Prentice Hall. 12. Web Document Clustering: A Feasibility Demonstration Oren Zamir and Oren Etzioni Department of'. Martin Christopher . WA 98195-2350 U.S.A. *Logistics and Supply Chain Management: Strategies for Reducing Cost and Improving Service*, 1998. Computer Science and Engineering University of Washington Seattle (2nd edition)

[Fui-Hoon Nah and Davis (2002)] *HCI Research Issues In E-Commerce*, Fiona Fui-Hoon Nah , S Davis . March 2002. p. 98.

[Bian et al.] *Optimizing User Exploring Experience in Emerging E-Commerce Products Xiubo Geng, Xin Fan*, Jiang Bian , Xin Li , Zhaohui Zheng .

[Cortes and Vapnik ()] 'Support-vector networks'. C Cortes , V Vapnik . 10.1007/BF00994018. *Machine Learning* 1995. 20 (3) p. 273.

[Pitt et al. (2002)] 'The Internet And The Birth Of Real Consumer Power'. Leyland F Pitt , Pierre Berthon , Richard T Watson , George M Zinkhan . *Business Horizons*, July-August(2002. p. .

[Today Need of e-Commerce Management to e-Skill Trainings Rashad Yazdanifard and Adnis Zargar] *Today Need of e-Commerce Management to e-Skill Trainings Rashad Yazdanifard and Adnis Zargar*,

[Turban et al.] E Turban , J K Lee , D King , M Chung . *Electronic Commerce: A Managerial Perspective*,

[Watson et al. ()] Richard T Watson , Pierre Berthon , F Leyland , George M Pitt , Zinkhan . *Electronic Commerce: The Strategic Perspective. Global Text*, 2008.