Design and Analysis of Web Based Recommendation System Using E-Commerce Data

Priyanka Patel[1], Mehul Barot[2]
[1, 2] LDRP-ITR, Gandhinagar

Abstract- *Recommendation system are used by number of E-Commerce site to help customer find the product to purchase based on their preferences. And also developed recommendation system based on web usage mining pattern. Web usage mining techniques aims to analyze and discover the user's navigational patterns and user's past buying behavior patterns in e-commerce sites. Recommendation system can implemented in different way: Collaboration filtering and content based filtering and other approaches are knowledge base, Demographic. We can also combine above techniques called hybrid recommendation for better recommendation to user. Here, we also compare different algorithm of recommendation System with parameters. In this paper we examine how recommender system help E-Commerce sites to increase sale and improve accuracy of recommendation using refusing of two feature of product and analyze the key challenges for recommendation system.*

Keywords- Recommendation system, E-Commerce, Personalized Recommendation, Web usage mining, Data mining technique, Genetic Algorithm, Content based Recommendation

I. INTRODUCTION

Recommendation System are used for E-Commerce sites to suggest products to their customer's preferences and interest. The product can be recommended based on an analysis of the customer's past buying behaviors[4].RS also predict the Recommendation of product based on user's profile information. The recommender system is a system which is used to generate a collection of products or item which will interest a particular person[1].

This system is mostly implemented in E-commerce website to help user or customer for finding what they are looking for or items those might interest them. These systems have the ability to change with respect to change in user's preferences[1].

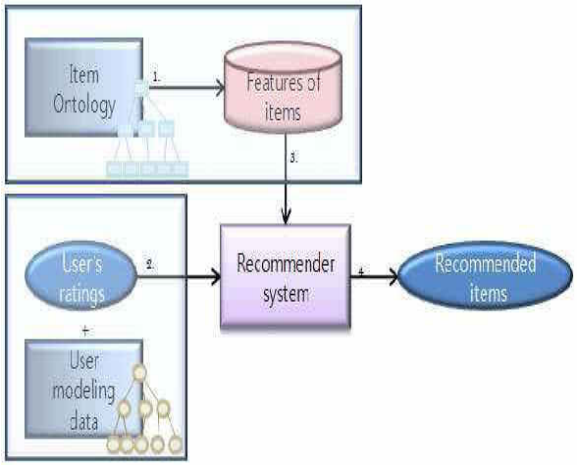

Figure.1: Recommendation System[1]

Web uses mining focus on predicting users' preference and behavior by analyzing web log with help of traditional data mining techniques [12]. Recommender system also known as recommenders are tools that provide important suggestions to users or customers in order to take good decisions while shopping online. The Recommendation systems use old knowledge and some statistical methods to provide these suggestions[8]. The forms of recommendation include suggesting products to the consumer, providing personalized product information, summarizing community opinion, and providing community critiques. Broadly, these recommendation techniques are part of personalization on a site because they help the site adapt itself to each customer[4].

Recommender systems enhance E-commerce sales in three ways[3]:

Converting Browsers into Buyers[3]: Visitors to a Web site often look over the site without purchasing anything. Recommender systems can help consumers find products they wish to purchase.[3]

Increasing Cross-sell [3]: Recommender systems improve cross-sell by suggesting additional products for the customer to purchase. If the recommendations are good, the average order size should increase. For instance, a site might recommend additional products in the checkout process, based on those products already in the shopping cart.[3]

Building Loyalty [3]: In a world where a site's competitors are only a click or two away, gaining consumer loyalty is an essential business strategy. Recommender systems improve loyalty by creating a value-added relationship between the site and the customer. Consumers repay these sites by returning to the ones that best match their needs. The more a customer uses the recommendation system – teaching it what he wants – the more loyal he is to the site. Creating relationships between consumers can also increase loyalty, for consumers will return to the site that recommends people with whom they will like to interact.[3]

II. RELATED WORK

Recommender systems represent consumer preferences for the purpose of suggesting items to buy or examine. E-commerce websites can predict a consumer's future purchasing behaviour through the information collected from a consumer's past purchasing behaviour and demographic data by that Consumer. Those consumers have the similar shopping behaviours and interests are grouped together in order to recommend products with surprises to target consumers. To predict the consumer future purchase the system recommends all those products that are matching consumers' needs, but it's unable to completely satisfy the consumers. Then Marginal utility concept added. Marginal utility is an economic concept because economists and marketing research use it to determine how much of an item a consumer will purchase. According to the Law of Diminishing Marginal Utility, some item has the decreasing marginal utility with the increase of shopping count, such as tablet, computers or other electronic gadgets so on Consumers are not feasible to purchase the same type item again in a short period if they have it already before. Apart from, some products, they used in daily life, it would be preferred to purchase in short time like as eating food. Then the marginal utility will be increase[7].

III. RECOMMENDATION SYSTEM TECHNIQUES

Recommendation System can be defined as the type of information filtering system that is used to recommend the products.[1]

A. Content based (CB) recommendation

Content based filtering also referred as cognitive filtering. In this type of filtering, items are recommended based on comparisons between items and user profiles [1].

Figure.2 : Example of Content based filtering [8]

In content-based (CB) system Ratings expressed by a single consumer have no role in recommendations provided to other consumers. Content Based approach learns a profile of the consumer interests based on some features of the objects the consumer rated. After the system exploits the consumer profile to suggest suitable items by matching the profile representation against that of items to be recommended. content-based techniques are limited by the features that are associated either automatically or manually with the items. No CB system can provide good suggestions if the content does not contain enough information to distinguish items the consumer likes from items the consumer does not like. Enough ratings have to be collected before a CB system can really understand consumer preferences and provide accurate recommendations. Therefore, when few ratings are available, such as for a new consumer, the system would not be able to provide reliable recommendations[7] .

B. Collaborative Filtering

Collaboration Filtering is most successful recommendation technique. The base idea of the Collaboration Filtering is to recommend items based on the opinion of other people. It recommends items for a user which depends on the items rated previously by same mind set [1].

Figure.3: Example of Collaborative filtering [8]

If a new item appears in the database there is no way it can be recommended to a consumer until more information about it is obtained through another consumer either rating it or specifying which other items it is similar to[7]. Collaborative filtering technique takes both user‟s past behavior and similar decisions made by other users. In simple, a collaborative filtering focuses on a customer‟s behavior based on other customer‟s behavior. The effectiveness of collaborative filtering is based on user rating. But, the problem with collaborative filtering is that, user rate only limited number of products. This problem is considered as Data sparsity. It is also difficult to design model based on the little rating information. This problem is known as cold start [8].

C. Hybrid Recommendation

Hybrid approach combines two or more techniques in different ways to improve recommendation performance in order to tackle the shortcoming of underlying approaches including cold-start or data sparsity problem. [8]

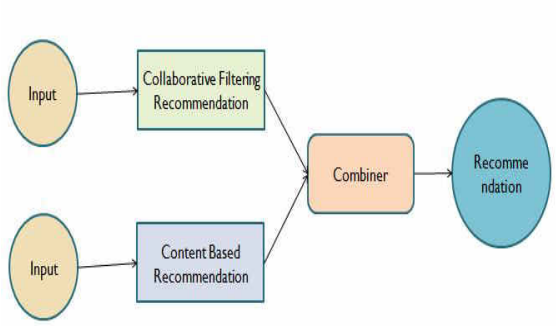

Figure.4 : Example of Hybrid Recommendation

Current hybrid approaches still suffer from a few drawbacks. First, there is insufficient contextual information to model consumers and items and therefore weaknesses to predict consumer taste in domains with complex objects such as education [8].

IV. APPLICATION OF RECOMMENDATION SYSTEM

There are different types of area where use recommendation:

Entertainment

Recommendations for movies, music,etc.[20]

Recommender system for music or movies, in this user find the music and provides result with items resulting in own user preferences. In this first of all extracts the unique properties of music/movies and stored in database. After that stored property analyzed using content based filtering and interactive genetic algorithm. After acquiring records system recommends items appropriate to user's own favorite.

Content

Personalizednewspapers,recommendationfordocuments,recommendations of Web pages, e-learning applications, and e-mail filters.[20]

In this recommender system extracts the user's basic requirement for newspaper or any document and then according to that requirement recommend the newspaper or any document or anything to user as per preferences.

E-commerce

Recommendations for consumers of products to buy such as books, cameras, PCs etc.[20]

In this recommender extracts the basic features like what type of book user like? Which types of basic specification they wants for camera?etc. After collecting the information recommender system apply any filtering technique and recommend that particular item to user as per their preferences.

Services

Recommendations of travel services, recommendation of experts for consultation, recommendation of houses to rent, or matchmaking services.[20]

V. FRAMEWORK OF E-COMMERCE SITES

Now, here we can used recommendation in different application that explain above but we focus on e-commerce site. we apply recommendation for E-commerce data.

Captured data from e-commerce sites:[6][19]

The proposed recommender system is developed based on the customers' navigational and behavioral patterns in e-commerce sites.[6] Navigational patterns include browsing, searching, product click, basket placement, and actual purchase, while behavioral patterns consist of the click ratio for a certain type of product, length of reading time spent on a specific product, number of visits to a specific product, printing, and bookmarking.[6]

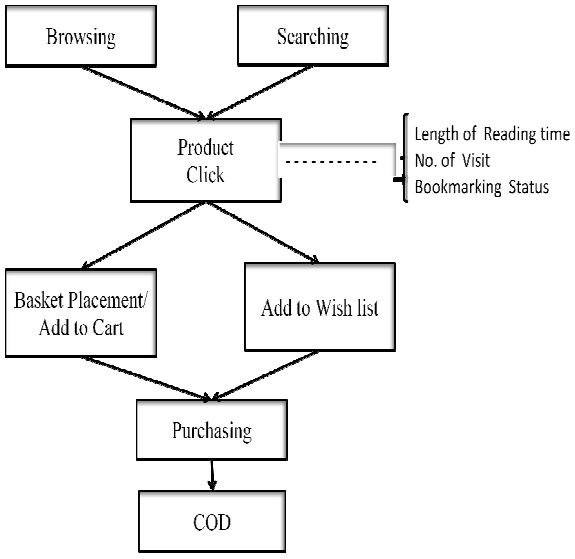

Figure. 5: Framework of E-Commerce sites [6]

Although the proposed system is developed using an experimental e-commerce site as an example, it can be applied to a variety of e-commerce sites as long as the above navigational and behavioral patterns can be captured.[6][19]After logging-in to the web site, a customer can either browse through the site just to check whether there are interesting products or intentionally search for a specific product to purchase.[6] When the customer clicks a product, he or she will be provided with specific information. Then, the customer can either print or bookmark the page as a reference for a future purchase or compare the details of the product with other available goods.[6][19]

Other important information that can be obtained from the customer's actions within the site include[6]: (i) the time it takes for the customer to read about a specific product (length of reading time); (ii) the number of visits to a specific product (number of visits); and (iii) the category to which the product belongs. A product that is frequently viewed and read can be surmised as a popular product. Furthermore, products in a certain category with a high click ratio can also be considered popular [6].

VI. RESERCH CHALLENGES

This section outlines some of the more important technical challenges which is faced by recommendation system in E-Commerce sites.

SCALABILITY

Scalability[5] is challenge facing current ecommerce recommendation systems. Large sites may have millions of users, hundreds of thousands of pages, and tens of millions of page views in a day. These same sites want to maintain the reactivity and responsiveness of smaller sites. Many commercial recommender systems have never handled a database that large, or a load that high. Many AI algorithms that work well for small-scale problems are too inefficient to be used for very large problems. New algorithms are needed that can handle very large-scale problems while maintaining the accuracy and near real time response of smaller scale systems.[5]

Ironically, not enough data is also a challenge for recommender systems, which must collect enough data to make effective recommendations for new customers. Sites can make good recommendations faster if they share information about their customers with other sites. Shared information benefits customers who receive more accurate recommendations in less time. Sites that use shared information benefit because the better recommendations help their customers find more products to purchase and because the customers prefer to return to stores with the better recommendations.[4]

SPARSITY PROBLEM

Sparsity [7] problem is one of the major problems encountered by recommender system and data sparsity has great influence on the quality of recommendation. Generally, data of system like Movie Lens is represented in form of user-item matrix populated by ratings given to movies and as no. of users and items increases the matrix dimensions and sparsity evolves. The main reason behind data sparsity is that most users do not rate most of the items and the available ratings are usually sparse. Collaborative filtering suffers from this problem because it is dependent over the rating matrix in most cases. Many researchers have attempted to reduce this problem; still this area demands more research.[7]

INCORPORATING RICH DATA

Recommender applications currently use a wide range of input data in forming their recommendations, including explicit ratings and simple behavioral data such as purchases and click-through. However, there are many other types of data that can be collected and used. In the future, recommender systems will commonly collect dozens of different types of data and integrate them into effective recommendations. Ongoing research in web usage mining and more general commerce-related data mining may reveal techniques for exploiting complex behavioral data.

Some recommenders, such as those on news sites, will want to provide recommenders that incorporate content

analysis as well as preference analysis. Many machine-learning techniques have been shown to effectively build a model of human preferences based on content. Information filtering systems build models of customer preferences based on keyword vectors. [4]

COLD START PROBLEM

Its difficult to give recommendations to new users as his profile is almost empty and he hasn't rated any items yet so his taste is unknown to the system. This is called the cold start problem. In some recommender systems this problem is solved with survey when creating a profile. Items can also have a cold-start when they are new in the system and haven't been rated before. Both of these problems can be also solved with hybrid approaches.[12] Three kinds of cold start problems are: new user problem, new item problem and new system problem. In such cases, it is really very difficult to provide recommendation as in case of new user, there is very less information about user that is available and also for a new item, no ratings are usually available and thus collaborative filtering cannot make useful recommendations in case of new item as well as new user.[12]

HYBRID DATA

One major challenge is in the area of so-called hybrid systems. The currently dominant systems generate recommendations utilizing only one type of input data about customer preferences for products (for example, explicit ratings data or purchase data). The goal of hybrid system is to take all available preference data simultaneously, and use it in an intelligent way to provide recommendations.[5]

VII. ANALYSIS OF RECOMMENDATION ALGORITHM

Different types of algorithms are used for recommendation:

PSO (Particle Swarm Optimization):

In PSO algorithm is utilized to estimate the features importance and allocate their weights accordingly. PSO is best and fast algorithm compare to other but it gives single result [15].

Although particle swarm optimization[14] is a population- based evolutionary technique like genetic algorithms, it differs in that each particle or solution contains a position, velocity and acceleration. The velocity and acceleration are responsible for changing the position of the particle to explore the space of all possible solutions, instead of using existing solutions to reproduce. As particles move around the space, they sample different locations. Each location has a fitness value according to how good it is at satisfying the objective, in this case, defining the user's preferences. Because of the rules governing the swarming process, particles will eventually swarm around the area in the space where fittest solutions are.[14]

MC (Markov Chain) :

Markov Decision Processes (MDP) a well known stochastic model of sequential decision. MDP-based recommender system must employ a strong initial model, must be solvable quickly, and should not consume too much memory.[11]

In MC require to give initial step before start work, means it work with CF, so for similar item recommendation first check the similar interest of other user.

FUZZY LOGIC:

Fuzzy linguistic approach to efficiently describe the user ratings and weights to precisely rank the relevant items to a user. [13]

Fuzzy recommender system based on collaborative behavior of ants (FARS). FARS works in two phases: modeling and recommendation. First, user's behaviors are modeled offline and the results are used in second phase for online recommendation. Fuzzy techniques provide the possibility of capturing uncertainty among user interests and ant based algorithms provides us with optimal solutions.[10]

GA (Genetic Algorithm):

GAs[17] are stochastic search techniques that guide a population of solutions using the principles of evolution and natural genetics. Extensive research has been performed exploiting the robust properties of genetic algorithms and demonstrating their capabilities across a broad spectrum of optimization problems, including feature selection and weighting tasks. GAs[17] are modeled loosely on the principles of the evolution via natural selection, employing a population of individuals that undergo selection in the presence of variation-inducing operators, such as mutation and crossover. A fitness function is used to evaluate individuals, and reproductive success varies with fitness. A general algorithm is started with a set of solutions (represented by chromosomes) called population. An initial population is created from a random selection of solutions. At every evolutionary step (generation), the solutions in the current

population are evaluated according to some predefined quality criterion, referred to as the fitness or fitness function. Solutions from one population are used to form a new population (next generation). Solutions (parents) are selected according to their fitness - the more suitable they are the more chances they have to reproduce. These solutions then "reproduce" to create one or more new solution (offspring), after which the offspring are randomly produced by crossover or mutation. The process of fitness-dependent selection and application of genetic operators to generate successive generations of solutions is repeated many times until a satisfactory solution is found.[17]

GA contains three phases as data preparation, mining of informative rules and recommendation. The sequences of web pages with best solution were identified by the GA using its three stage Process which is fitness value calculation, crossover and mutation [2].

GA slow compare to PSO but it gives multiple solution and also work with CB filtering.

Based on above analysis Genetic algorithm was best for our project.

VIII. PROPOSED WORK

We use content based approach but it has the limitation such that it focuses on only the accessed items and is not prompt to immediate changes in the potential interest of users[18]. To overcome these limitations, we combine the content based filtering approach and the genetic algorithm in our proposed system.

1. Genetic Algorithm (Proposed Algorithm)

The basic steps of GAs are outlined as follows:

A. Initialization

Generate an initial population using calculating candidate solutions based on product features and evaluate the fitness function of initial population.

B. New population

Create a new population by repeating the following steps until the new population is complete.

a. Selection:

Select two parent solutions from a population according to their better fitness and the greater the chance to be selected.

b. Crossover:

With a crossover probability cross over the parents to form a new offspring. crossover perform randomly between two parent which have highest fitness value.

c. Mutation:

With a random mutation probability, mutate new offspring at each position in the chromosome.

d. Acceptance:

Place a new population.

C. Evaluation

Compute the fitness values for the new population of N solutions.

D. Test

If the stopping criterion is met, stop, and return the best solution in current population.

GA work on a population of candidate solutions; calculate candidate solution based on refusing of two feature each solution has a fitness value indicating its closeness to the optimal solution of the problem. The solutions having higher fitness values than others are selected and also survive to the next generation. GAs then produce better solution by the combination of selected solution.

2. E-Commerce Recommendation System Design using GA

In this paper, first from the construction of the overall design of the system of recommendation, in order to be updated in real time data processing, make electronic commerce website recommendation, this paper proposes the following recommendation system model.

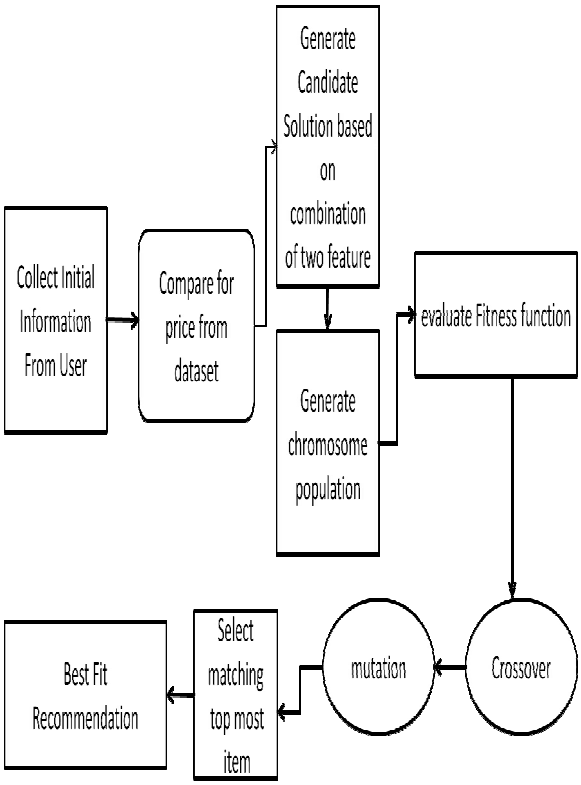

Figure.6: Work flow of proposed System

1. Collect the initial information from user (about price range and product features)
2. Information will be used to recommend the other products,

   a) Features of the input information will be compare with the features of the products in the database
   b) After compare the feature Display the price of product from both dataset
   c) Calculate Candidate solution based on product price range & product brand name
   d) Generate initial population chromosomes
   e) Fitness function will be evaluate (Maximizing the fitness value)
   f) Apply crossover and mutation
   g) And generate new result with best fit value
   h) Select best highest value from above result and it compare with DB and find range of matching
   i) Best fit matching product will be selected to recommend the user and least fit will be discarded .

3. DATASET

We will try to use real time dataset through E-commerce website otherwise referred dataset from UCI Repository which is related to E-Commerce sites and also extract data from Amazon or flipkart using import.io data Extraction tool.

IX. CONCLUSION

In this paper, we have discussed about techniques of recommendation system that are used in E-Commerce and explain framework of this system. In this paper, use Recommendation system by number of E-Commerce site to help customer find the product to purchase based on their preferences. And also developed recommendation system based on web usage mining pattern. Web usage mining techniques aims to analyze and discover the user's navigational patterns and user's past buying behavior patterns in e commerce sites. We used different approach for recommendation in e-commerce site. Here we examine how recommender system help E-Commerce sites to increase sale. We used Genetic approach with content based recommendation technique for recommendation based on two feature price range and brand of that product in e-commerce site rather than single feature of product and find the product with lowest price range using comparison of product from two e-commerce web sites and reduce time.

Here define different challenges of recommendation in e-commerce and its handle by combination of techniques of RS.

X. FUTURE WORK

The future work for the given approach to implement and test the system across several datasets for different product .

We can compare result with single feature recommendation and recommendation with refusing of two or more feature.

REFERENCES

[1] Tuhina Sultana "Recommendation System for Music File" et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (1) , 2015, 275-277

[2] Prince Mary, S. and E. Baburaj "CONSTRAINT INFORMATIVE RULES FOR GENETIC ALGORITHM-BASED WEB PAGE RECOMMENDATION SYSTEM". Journal of Computer Science 9 (11): 1589-1601 ,ISSN: 1549-3636 © 2013 Science Publication

[3] J. Ben Schafer, Joseph Konstan, John Riedl "Recommender Systems in E-Commerce" GroupLens Research Project Department of Computer Science and

Engineering University of Minnesota Minneapolis, MN 55455 1-612-625-4002

[4] J. Ben Schafer, Joseph A. Konstan, John Riedl "E-Commerce Recommendation Applications" J. Ben Schafer, Joseph A. Konstan, John Riedl GroupLens Research Project Department of Computer Science and Engineering University of Minnesota Minneapolis, MN 55455 1-612-625-4002

[5] Robert Driskill and John Riedl "Recommender Systems for E-Commerce: Challenges and Opportunities" From: AAAI Technical Report WS-99-01. Compilation copyright © 1999, AAAI (www.aaai.org). All rights reserved. Net Perceptions 7901 Flying Cloud Drive EdenP rairie, Minnesota5 5344 { rdriskill,riedl } @netperceptiocnosr.

[6] Yong Soo Kim, Bong-Jin Yum*, Junehwa Song, Su Myeon Kim Korea "Development of a recommender system based on navigational and behavioral patterns of customers in e-commerce sites"Elsevier, Advanced Institute of Science and Technology, 373-1 Gusung-Dong, Yusung-Gu, Daejon 305-701, South Korea. Expert Systems with Applications 28 (2005) 381–393

[7] Mohammad Daoud1, S. K. Naqvi2 "Recommendation System Techniques in E-Commerce System" International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

[8] R.Srinivasa Raju ,I.Kali Pradeep, I.Bhagyasri P. Praneetha ,P.S.S. Teja "Recommender Systems for E-commerce: Novel Parameters and Issues" , International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 9, September 2013

[9] Yi-shin Chen and Cyrus Shahabi "Improving User Profiles for E-Commerce by Genetic Algorithms", Integrated Medi System Center and Computer Science Department University of Southern California,Los Angeles, CA 90089-2564.

[10] Shiva Nadi1, Mohammad. H. Saraee2, Mohammad Davarpanah Jazi3 and Ayoub Bagheri "FARS: Fuzzy Ant based Recommender System for Web Users ", Islamic Azad University, Najafabad Branch, Isfahan, Iran Department of Electrical and Computer Engineering, Isfahan University of Technology Isfahan, Iran Department of Computer Engineering, Foulad Institue of Technology Fouladshahr, Isfahan, Iran, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011 ISSN (Online): 1694-0814 www.IJCSI.org

[11] Guy Shani, David Heckerman , Ronen I. Brafman, "An MDP-Based Recommender System", Journal of Machine Learning Research 6 (2005) 1265–1295 Submitted 10/03; Revised 4/04; Published 9/05.

[12] Daniar Asanov "Algorithms and Methods in Recommender Systems", Berlin Institute of Technology Berlin, Germany.

[13] Saad M. Darwish*, Magda M. Madbouly, Eman Abd-El Reheem "Contribution to Collaborative Filtering Based on Soft Computing to Enhance Recommender System for e-Commerce" Institute of Graduate Studies and Research, Alexandria University, 163Horreya Avenue, El-Shatby, 21526 P.O. Box 832, Alexandria, Egypt. Manuscript submitted June 10, 2014; accepted August 28, 2014. doi: 10.7763/ijeeee.2014.V4.3 41, International Journal of e-Education, e-Business, e-Management and e-Learning

[14] Supiya Ujjin and Peter J. Bentley "Particle Swarm Optimization Recommender System ", University College London Department of Computer Science Gower Street, London WC1E 6BT s.ujjin@cs.ucl.ac.uk

[15] Amira ABDELWAHAB , Hiroo SEKIYA, Ikuo MATSUBA, Yasuo HORIUCHI, Shingo KUROIWA "Feature Optimization Approach for Improving the Collaborative Filtering Performance Using Particle Swarm Optimization", Graduate School of Advanced Integration Science, Chiba University, Japan Journal of Computational Information Systems, Copyright © 2012 Binary Information Press January 2012

[16] "Badrul Sarwar, George Karypis, Joseph Konstan, and John RiedlEC Analysis of Recommendation Algorithms for Ecommerce", October 1720, 2000, Minneapolis, Minnesota. Copyright 2000 ACM 1581132727/ 00/0010.

[17] Chein-Shung Hwang "Genetic Algorithms for Feature Weighting in Multi-criteria Recommender Systems", Journal of Convergence Information Technology Volume 5, Number 8, October 2010.

[18] Hyun tae kim, Eungyeong kim, Jong-HyunLee, Chang Wook Ahn, "Recommender system based on Genetic

Algorithm For Music Data", School of Information & Communication Engineering @ 2010 IEEE.

[19] YONG SOO KIM "Recommender System Based on Product Taxonomy in E-Commerce Sites" Department of Industrial and Management Engineering Kyonggi University Gyeonggi-do, 443-760 Korea-2010.

[20] Francesco Ricci · Lior Rokach · Bracha Shapira,Paul B. Kantor "Recommender Systems Handbook" Springer.