

DEVELOPMENT OF RECOMMENDATION SYSTEM FOR FOOTWEAR E-COMMERCE PRODUCTS

Raj Kumar Mahor
Computer Science Engineering
Nagaji Institute of Technology and Science
Gwalior, India
Email:raj.gwalior@gmail.com

Prof. Abhilash Mishra
Computer Science Engineering
Nagaji Institute of Technology and Science
Gwalior, India

ABSTRACT: Past decade has seen a prominent rise in the number of e-commerce applications in the World Wide Web. Dynamic recommendation system for predicting user interests is quite challenging for such systems. Many E-commerce organizations, now a days, with the booming culture of E-commerce, rely on websites, to attract new customers and retain the old customers, and hence to have the growth of the organizations. we proposed a new method of recommending Footwear products by Visitor's behaviour like clicks, ratings, sells, quantities etc. based rational technique that uses Brands, Categories, Discount, Location, Occasion etc. of each visited product for tracking user's dynamic behaviour, uses 'popularity' measures for more accurate recommendation list of items. The results of comparison between proposed system and traditional recommendation system prove that proposed system provide good quality accuracy and diminishes the limitations of traditional system.

Keywords: E-commerce, Recommendation System, Footwear , Sells And Quantities, Web Mining, filtering methods

I INTRODUCTION

Recommendation systems, in the most general sense, are used to give the user of a site/function (whether that is an online shop like Amazon or Flipkart) some added value in the form of recommendations. Some of the recommendations use only behavioural data, e.g. ratings, purchases, clicks and so on, to predict what items a customer would like. There are different kinds of systems and not everyone agrees on what classification should be used, but the most common classification divides all systems into one of three categories: Collaborative filters, Content-based filters and Hybrid filters.

Most sites today use many methods or algorithms using behavioural data in other ways in their recommendation Systems. However, it could be of interest to explore other approaches where other methods, such as hybrid systems using both product data and behavioural data, are used to recommend products, as methods only using behavioural data usually have trouble recommending products when they don't have enough data. The project is also using practical data, which means problems will be encountered that would not be there on manufactured data, and has the advantage of using WAMP test framework (this will be presented in the SIMULATION TOOL chapter). Finally, one of the project goals is to be able to generalise the procedure created during

the project enough to work on different sites without much configuration.

The outline and details of the Paper were decided on by the master's. The project was carried out in three important stages:

1. Complete study of the this research field and the commonly used techniques. From this baseline a few promising methods were chosen for implementation and evaluation.
2. Implementation and comparison both are used method. In that stage the best filtering algorithm was selected for development in the live system.
3. Evaluation of the implemented solution and reporting.

II WEB MINING

One of the filtering applications field of data mining is *World Wide web*, which serves as a widely and huge distributed, global information data service center for every kind of information such as news, advertisements, consumer information, financial management, education, government, e-commerce, health services, and many other information services. With the rapid growth of the WWW, this is more useful to find the useful information from these huge data. The webpages also have a rich and dynamic hyperlink collection and web site usage and access its information for web data mining. The web poses best challenges for effective knowledge filtering and data mining applications. Web site mining is defined as the use of data mining filtering techniques to automatically develop and extract information raw web pages and services. In general, web mining is a important term for three information filtering domains that are related with mining different parts of the website:

Web Structure Mining,
Web Content Mining and
Web Usage Mining

III RECOMMENDATION SYSTEM

Recommendation is a social process which is done by people when they want to describe their degrees of appreciation about someone or something. In computers, recommendation System began to appear in the 90's. Traditional recommendation System is mainly used to recommend products, services or people. According to Resnick and Varian (Resnick and Varian, 1997), in ordinary life, people normally trust recommendations made by others. Those recommendations appear to them as word of mouth reputation, recommendation letters, shoes and book reviews printed in news papers and magazines. Resnick and Varian (Resnick and Varian 1997) define recommendation System as "system where users provide recommendations as inputs data which the system then collect and directs to appropriate recipients." recommendation System attracted public interest during 1990s and it (Resnick and Varian, 1997) has obtain wide scope acceptance now. This sort of systems created the basic ground for new product sales opportunities in e-commerce (Schafer et al., 1999; Sarwar et al., 2000a). Recommendation System applies personalization techniques, considering that different visitors have different preferences and different information needs (Konstan et al., 1997).

The main objective of recommendation System is that of complexity reduction for the human being and selecting those pieces of information that are relevant for the recommendation seeker. From the beginning implementation of the recommendation System focuses on simple database queries. The most popular technologies used, according to (Schafer et al), are: Nearest neighbour: the algorithm computes the distance amongst visitor's preferences or characteristics. Predictions about products (products, services or people) are to be

recommended considering shorter differences amongst the item and the set of the nearest neighbours.

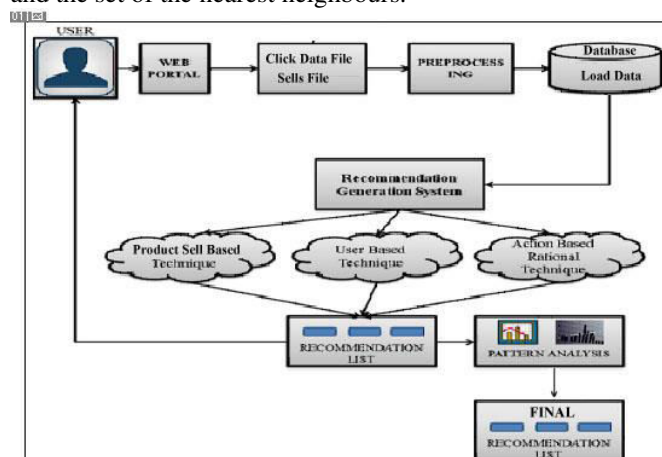


Figure III Recommendation System

IV FILTERING TECHNIQUES

Two principal method for computing recommendations have come out such as content-based and collaborative filtering (Goldberg et al., 1992). These filtering techniques recommend products for the visitor based on the descriptions of previously evaluated products. Each of these filtering techniques will be discussed in this chapter.

A. CONTENT BASED FILTERING

Content-based recommendation filtering technique works with existing profiles of visitors. A profile has information about a visitor and their taste. Taste is based on visitor rating for different products. Generally, whenever a visitor creates his profile, Recommendation engine does a visitor survey to get initial information about the visitor in order to avoid new visitor problem.

B. COLLABORATIVE FILTERING

Collaborative filtering has mainly two approaches:

I. Visitor Based Approach.

In this approach, Products that are recommended to a visitor are based on an evaluation of products by visitors of the same neighborhood visitor, with his shares common preferences. If the article was positively rated by the community, it will be recommended to the visitor. In the visitor based approach, items are already rated by a web visitor, play a most important role in filtering for a group that share the appreciations with him.

II. Item Based Approach.

Mention to the fact that the need of users remains constant or change very slightly, similar recommendation build neighbourhoods based on appreciations of visitors. Afterwards, the system produce recommendations for neighborhood that a visitor might need.

C. KNOWLEDGE BASED FILTERING

Recommendation System using this technique recommends products based on inferences from visitor's preferences and needs. The Visitor Profile consists of the functional knowledge

structure and interpreter according to the guess machine. Some work based on that recommendation technique is: Google (Sergey Brin and Lawrence Page98) recommends the most popular links of web pages that contain the query provided by the visitor. The implementation approach uses probabilistic models. The Entree (Robin Burke.02) recommends restaurants based on visitor's desired restaurant features. The implementation approach is the knowledge-based similarity retrieval based on case-based reasoning.

D. HYBRID FILTERING

Hybrid recommendation aims to avoid some limits of filtering methods by combining two or more filtering methods together. The hybrid method is a combination of two algorithms i.e. Content-based and collaborative-based algorithms. Which is used to make effective recommendation systems? Several techniques are used to combine content-based and collaborative based algorithm. These are:

- I. Separately implementing content-based and collaborative algorithms and combined the predicted result.
- II. Add some content-based characteristics to the collaborative algorithm.
- III. Add some collaborative characteristics to the content-based algorithm.
- IV. Incorporating both algorithms and building a usual framework.

V SIMULATION TOOL

WAMP means "Windows, Apache, MySQL, and PHP." WAMP is a variation of LAMP for Windows OS systems and it is installed as a software bundle (Apache, MySQL, and PHP). It is mostly used for web development and internal testing, but may also be used to serve live websites. The main important part of the WAMP is Apache ("Apache HTTP Server") which is used run the web server within windows platform. Apache web server provide easy and fast testing and execution for web developer in a web browser without publishing them live on the Internet.

WAMP is data evaluation which has been designed web application and web operations. As well as this, WAMP has excellent data capabilities, and its have PHP powerful programming language. Somewhat give an explanation of whole WAMP'S capabilities, we will limit ourselves to only those features apprehensive with handling of Data. We will provoke features, function and procedures as required. WAMP capabilities are the key-word which accepts various parameters, and produces a few type of output: For web operation, a string, a image or determine. There are a few features in WAMP as we have to see, it is easy (and every now and then required) to write down our own. Examples of function might be WAMP is a superior dialect for specialized processing. It incorporates data computation and programming in a simple to-utilize environment where issues and arrangements are communicated in commonplace scientific documentation. Regular uses include:

Data analysis, investigation, and representation Engineering and Web Application advancement, with HTML structure.

Wamp Server refers to a software stack for the Microsoft Windows operating system, created by Romain Bourdon and consisting of the Apache web server, MySQL database and PHP programming language.

USING WAMP SERVER, PHPMYADMIN AND OPENCART

The proposed system is implemented using WAMP, PHPMYADMIN and OPENCART. OPENCART which is an MVC based open source shopping cart system was used to write the testing program. For validating our proposed system, web portal was developed that offered different Footwear products for the customers. The database used was real time data containing 1121 records from web portal "shoes.raj". Figure V.1 Illustrates the GUI of the web portal offering multiple categories of Footwear product. In Figure V.2. A sample of uncleaned log file is shown that contains navigational data. In Figure V.2. Cleaned log file in appropriate structured form is shown.

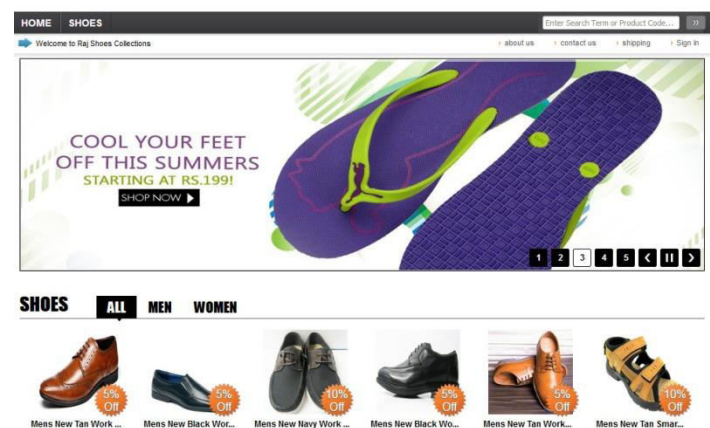


Figure V.1 GUI Proposed System

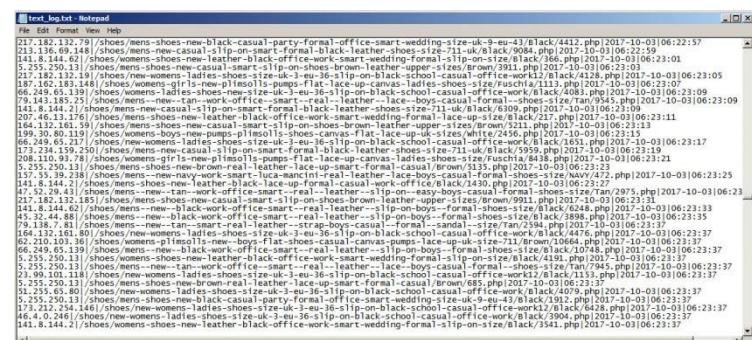


Figure V.2 Uncleaned Log File

VI PROPOSED WORK

On the other hand the visitor's purchasing power needs to calculate. Therefore, the visitor's previous searched products are evaluated. In this part the previously performed search of the data is obtained with their price range (low-high). Similarly the

previous purchased items for with their actual price are used to calculate the coefficient matrix. These three factors- similar score behaviour, sentiment score and popularity score are used to calculate the combined weight for the particular product. Then a coefficient matrix and the computed weights are used to calculate the possible recommendation of the product for the visitor.

A. DATA ACQUISITION

In this phase the entire navigational data which includes all the web pages visited is collected and stored. The proposed work makes use of common log file format to maintain the data, important attributes namely IP address, timestamp, status code, URL, method (GET and POST), visitor agent and Referrer URL are recorded and used for further analysis. This data obtained is highly unstructured and inconsistent in nature. And therefore has to be preprocessed for further analysis.

B. CREATING THE TEXT FILE

Now we will create file by file handling PHP functions. And dump visitor's behaviours data into text file. We are taking some fields like IP, Product URL, Date, and Time.

C. DATA PREPROCESSING

Good quality input data needs to be served for better analysis. In this phase inconsistent, redundant data is eliminated using following steps. Field separation stage focuses on distinguishing one attribute from another by making use of separator character such as space. In data cleaning stage we filter out outliers data. We check for URL suffixes. Log entries having filename suffixes such as gif, jpeg, tif, jpg are discarded. All records having failed http status code are removed ie. status code greater than 200 and less than 299 are eliminated. In visitor differentiation phase we assign unique visitor ID, to each IP address and registered visitors to differentiate one customer from other. Finally we construct session in session identification phase. In session clustering phase we group together session belonging to unique visitor. Session provide us with complete set of activities done by the visitor in specific time period. Finally in data formatting stage we place the data in tabular form.

BUILDING A RECOMMENDATION ENGINE IN PHP/MYSQL

Recommendation engines are nothing but an automated form of an "online shop". You can buy good product. Not only check that product, but also the related ones which you could buy. So, does this recommendation engines.

The ability of these engines to recommend personalized content, based on past behaviour is incredible. It brings customer pleasure and gives them a reason to keep returning to the same website.

In this paper, I will cover the fundamentals of creating a recommendation system using PHP/MYSQL with LAMP. We will get some study that how recommendation work and create

basic popularity filtering model and a collaborative filtering model.

CONTENT BASED FILTERING MODEL

Now start with developing a popularity based model, i.e. the one where **all the visitors have same recommendation** based on the most popular choices. We'll use the Custom PHP recommender functions popularity recommender for this.

COLLABORATIVE FILTERING MODEL

In this case, an item is a shoes. Once we have the matrix, we use it to determine the best recommendations for a visitor based on the shoes he has already rated. Note that there a some more things to take care in actual development which would require deeper filtering.

HYBRID FILTERING MODEL

Let's start by understanding the basics of a Hybrid filtering method. The Combination of Content based and Collaborative filtering method Called Hybrid filtering method. The core idea works in 3 steps:

- Recommendations List by **High Sells** and **High Quantity** Shoes.
- Recommendations List by **Low Sells** and **High Quantity** Shoes.
- Recommendations List by **Low Sells** and **Low Quantity** Shoes

To give you a high level overview, this is done by making an **item-sells matrix** in which we keep a record of the pair of items which were rated together.

Recommendations List by High Sells and High Quantity Shoes

In this case, Once we have the shoes matrix, we use it to determine the best recommendations by High Sells and High Quantity Shoes for a visitor based on the shoes he has already rated.

Recommendations List by Low Sells and High Quantity Shoes

In this case, Once we have the shoes matrix, we use it to determine the best recommendations by Low Sells and High Quantity Shoes for a visitor based on the shoes he has already rated.

Recommendations List by Low Sells and Low Quantity Shoes

In this case, Once we have the shoes matrix, we use it to determine the best recommendations by Low Sells and Low Quantity Shoes for a visitor based on the shoes he has already rated.

VII RESULTS AND DISCUSSION

In order to test the recommendation models discussed in previous, the raw data collected from server logs are cleaned and converted into user sessions. This chapter discusses the results from experiments run to test the performance and effectiveness of the models.

RECOMMENDATIONS LIST BY HIGH SELLS AND HIGH QUANTITY SHOES RESULTS

A total of 10000 products in the database during the experimental period in the online shopping website. Top 10 most High Sells and High Quantity Shoes recommendation list. This product list is depends on sells and product quantity count so can say that user can buy those products. There is good and most effective recommendations technique. It is very fast and provide best products in these recommendations. See below 10 ten most High Sells and High Quantity Shoes product recommendation list.

Shoes ID	Rating Count	Ratings
11080	3	5,5,5
6915	4	5,5,3,5
8456	3	5,5,2
7463	3	5,5,1
10564	3	5,5,1
300	2	5,5
384	2	5,5
611	2	5,5
777	2	5,5
908	2	5,5

Table 7.1 Top 10 Recommendations List by High Sells and High Quantity Shoes.

RECOMMENDATIONS LIST BY LOW SELLS AND HIGH QUANTITY SHOES RESULTS

A total of 10000 products in the database during the experimental period in the online shopping website. Top 10 most Low Sells and High Quantity Shoes recommendation list. This product list is depends on sells and product quantity count so can say that user can buy those products. There is good and most effective recommendations technique selling pending stocks. It is very fast and provides best products in these recommendations. See below 10 ten most Low Sells and High Quantity Shoes product recommendation list.

Shoes ID	Sells Count	Shoes Quantity
----------	-------------	----------------

1787	1	122
1012	2	122
2992	1	100
777	4	80
8456	4	51
611	3	40
1266	2	33
6915	3	21
10564	3	21
300	3	11

Table 7.2 Top 10 Recommendations List by Low Sells and High Quantity Shoes.

RECOMMENDATIONS LIST BY LOW SELLS AND LOW QUANTITY SHOES RESULTS

We have 10000 products in the database during the experimental period in the online shopping website. Top 10 most Low Sells and Low Quantity Shoes recommendation list. This product list is depends on sells and product quantity count so can't say that user can buy those products. There is not good recommendations technique selling product. It is very fast and provides not good products in these recommendations. See below 10 ten most Low Sells and Low Quantity Shoes product recommendation list.

Shoes ID	Sells Count	Shoes Quantity
1048	2	5
2469	1	10
384	2	10
7463	3	10
908	3	10
300	3	11
11080	7	11
6915	3	21
10564	3	21
1266	2	33

Table 7.3 Top 10 Recommendations List by Low Sells and Low Quantity Shoes.

The experimental results also show that sells product recommendation is more effective than without sells product recommendation. The sells recommendation products are most predicted product rather than most rating or most searches

products. And mostly user can buy that product because they have already rating and comments. And most searchable products have not necessary that all have rating and comments.

VIII CONCLUSION

In this research, we provide a many method of recommending Footwear products to customers by extending the existing collaborative filtering method to reflect the characteristics of Footwear products. First, we considered the fact that Footwear products are sold, and preferences for Footwear products also appear online by using online search data and online purchase data to generate recommendations. Finally, the product which the customer wishes to purchase is a product that replaces or supplements the product that the customer preferred before. We have used product information to make recommendations that reflect this purchase intention. We developed a new recommendation system to reflect these approaches.

In this research work we focus on providing good quality product recommendations to all the users especially unregistered ones of E-commerce site. The beauty of the proposed system is it dynamically provides recommendation as per changing users' behaviour and traversal patterns by making use of web usage mining and constructing patterns from the historical data. The proposed recommendation system minimizes the false positive errors that occur frequently in traditional recommendation system. Also issue of star ratings and cache memory are handled by the system thereby providing good quality recommendations. Results prove that accuracy of approximately 80 to 85 percent is achieved for registered user and 65 to 70 percent for unregistered user in above recommendation technique, which is better than product based technique and almost equivalent to user based approach. The recommendation system has the potential to attract new customers and retain existing ones. This technique can help the E-commerce organization have competitive edge in the market and can be helpful in forecasting demands and sales for a specific product. It would be interesting to evaluate the proposed technique with different conventional recommendation approaches and measure its accuracy. This proposed system can also be tested for other application areas like books recommendation, music recommendation etc. In this research work we focus on providing good quality product recommendations to all the users especially unregistered ones of E-commerce site. The beauty of the proposed system is it dynamically provides recommendation as per changing users' behaviour and traversal patterns by making use of web usage mining and constructing patterns from the historical data. The proposed recommendation system minimizes the false positive errors that occur frequently in traditional recommendation system. Also issue of star ratings and cache memory are handled by the system thereby providing good quality recommendations. Results prove that accuracy of approximately 80 to 85 percent is achieved for registered user and 65 to 70 percent for unregistered user in above recommendation technique, which is better than product based technique and almost equivalent to user based approach. The recommendation system has the potential to attract new customers and retain existing ones. This technique can help the E-commerce organization have

competitive edge in the market and can be helpful in forecasting demands and sales for a specific product. It would be interesting to evaluate the proposed technique with different conventional recommendation approaches and measure its accuracy. This proposed system can also be tested for other application areas like books recommendation, music recommendation etc.

REFERENCES

- [1] F. O. Isinkaye, Y. O. Folajimi and B. A. Ojokoh, "recommendation system: Principles, method and evaluation", Egyptian Informatics Journal, pp. 261-273, 2015.
- [2] P. Lopes and B. Roy, "dynamic recommendation system using web usages mining for e-commerce users", Procedia Computer Science Volume 45, pp. 60 – 69, 2015.
- [3] Chapter 9- recommendation Systems, available online at: <http://i.stanford.edu/~ullman/mmds/ch9.pdf>
- [4] S. Hernandez, P. Alvarez, J. Fabra, and J. Ezpeleta, "Analysis of Users' Behaviour in Structured e-Commerce Websites", IEEE, VOLUME 5, pp.11941-11958, 2017.
- [5] S. Mittal, A. Goel and R. Jain, "Sentiment Analysis of E commerce and Social Networking Sites", International Conference on Computing for Sustainable Global Development (INDIA Com), IEEE, pp.2300-2305, 2016.
- [6] N. Kumari and S. N. Singh, "Sentiment analysis on Ecommerce application by using opinion mining", International Conference- Cloud System and Big Data Engineering, IEEE, vol 6, pp. 320-325, 2016.
- [7] S. Sivapalan, A. Sadeghian and H. Rahnama, "Recommender System in E-Commerce", World Automatic Congress, IEEE, pp.1-6, 2014.
- [8] M. Aprilianti, R. Mahendra, I. Budi, "Implementation of Weighted Parallel Hybrid recommendation Systems for Ecommerce in Indonesia ", ICACIS IEEE, pp. 321-326, 2016.
- [9] S. Sharma and S. S. Lodhi, Development of Decision Tree Algorithm for Mining Web Data Stream, International Journal of Computer Applications 0975 8887, vol138 2, March 2016
- [10] S. S. Patil and H. P. Khandagale, Enhancing Web Navigation Usability Using Web Usage Mining Techniques, International Research Journal of Engineering and Technology IRJET, vol4 6, June 2016
- [11] CBull, DAsfiandy, AGledson, JMellor, SCouth, Combining data mining and text mining for detection of early stage dementia the SAMS framework, LREC Workshop RaPIDPortorož Slovenia, 2016
- [12] P Ristoski and H Paulheim, Semantic Web in Data Mining and Knowledge Discovery A Comprehensive Survey, Preprint submitted to Journal of Web Semantics, 2015
- [13] P.Sukumar, L.Robert and S.Yuvaraj, "Review on Modern Data Preprocessing Techniques in Web Usage Mining (WUM),"

International Conference on Computational Systems and Information Systems for Sustainable Solutions, 978-1-5090-1022-6/16/IEEE (2016).

[14] B. K. Malviya and J. Agrawal, "A Study on Web Usage Mining: Theory and Applications," International Conference on Communication Systems and Network Technologies, 978-1-4799-1797-6/15/IEEE (2015).

[15] Sunena and K. Kaur, "Web Usage Mining-Current Trends and Future Challenges," International Conference on Electrical, Electronics, and Optimisation Techniques (ICEEOT), 978-1-4673-9939-5/16/IEEE (2016).