An Analytical Review of Web Page Recommendation System Based on Machine Learning

[1]Nikita Raut, [2]Jayant Adhikari

[1]*M.Tech Student, Department of Computer Science & Engineering, Tulsiramji Gaikwad Patil College of Engineering and Technology, Nagpur, Maharashtra, India.*
[2]*Assistant Professor, Department of Computer Science & Engineering, Tulsiramji Gaikwad Patil College of Engineering and Technology, Nagpur, Maharashtra, India.*

Abstract—Recommendation System (RS) are generally used in e-commerce industry to solve the complication of information overloading. Large amount of information is generating now, days due to which user face the difficulty in finding the relevant information of product and services matching to their taste and preferences. Data mining (DM) is the process of mining and extracting useful knowledge from large datasets. The tasks of DM are to do description and prediction of data to retrieve the information. RS is a subfield of information retrieval (IR) and IR is subfield of DM. Recommendation engines basically are data filtering and IR tool that make use of algorithms and data to recommend the most relevant item to particular user. The Various technique and approaches used by RS are content-based (CB) filtering, Collaborative Filtering (CF) and hybrid filtering techniques. This paper illustrates the role of Data Mining in Recommendation System and proposes a workflow of RS. Also describes the review of techniques, challenges of RS & compares recommendation systems of various e-commerce websites.
Keywords—Web Page Recommendation, Web Mining, Sequential Patterns, Knowledge Representation, Domain Ontology

I.   Introduction

Web page recommendation is the technique of web site customization to fulfill the needs of every particular user or group of users. The web has become largest world of knowledge. So it is more crucial task of the webmasters to manage the contents of the particular websites to gather the requirements of the web users.

The web page recommendation systems most part based on the exploitation of the patterns of the sites visitors. Domain ontology's provide shared and regular understanding of a particular domain. Existing system uses preorder linked WAP-tree mining (PLWAP Mine) algorithm that helps web recommendation system to recommend the interested pages but it has some drawbacks, it require more execution time and memory and also it does not work on updated dataset.

To overcome these drawbacks of the system utilizes PREWAP algorithm. The PREWAP algorithm recommends the interested results to web user within less time and also it requires less memory compare to PLWAP Mine algorithm which improves the efficiency of web page recommendation system. In work, various models are presented; the first model is Web Usage Mining which uses the web logs. The second model also utilizes web logs to represent the domain knowledge, here the domain ontology is used to solve the new page problem. Likewise the prediction model, which is a network of domain terms, which is based on the frequently viewed web-pages and represents the integrated web usage. The recommendation results have been successfully verified based on the results which are acquired from a proposed and existing web usage mining (WUM) technique.

Data Mining is cogent tool used by Recommendation Systems that helps e-commerce industry to increase their sales and maintaining the business relationships betweencustomer and organization. It is a discipline of criss-crossing of computer science and statistics used to find hidden pattern into the information banks [1]. The central objective is to dig and extract useful information from large datasets and frame it into knowledgeable structure for future use. This technique can be applied to any kind of unstructured data like multimedia data, network data and transactional data etc. [2]. DM is a core of knowledge discovery in database (KDD) process. The key steps of knowledge discovery are data cleaning, data integration, data selection, data information, data mining, pattern evaluation, knowledge presentation [3]. The various predictive data mining techniques are classification, regression, time series analysis, prediction and descriptive techniques are clustering, summarization, association rule, and sequence discovery [4]. The gigantic growth of products and service in e-commerce industry over the web has turned it strenuous for the user to search for the product which is reasonable to the user. RS is the software that provide the suggestion to the user regarding products and services in the e-commerce industry [5] and suggestion helps the user in decision making like selecting books,

cloths, shoes, watches etc. RS plays a very critical role in highly rated e-commerce sites and applications like Amazon, Flipkart, YouTube, Netflix and many e-learning sites etc. Here are some motivation as to why e-services provider introduce RS like increment in sale, user satisfaction, sell sundry items, user fidelity, more understanding of what user want? Various functions of RS are to express self, recommend a sequence and bundle of products, annotation in context, find some good item, find all good items, just browsing, improve the profile, help others, influence others [6].

Three phases of recommendation process are:
- Information Collection Phase: In this phase RS, initially, have to gather the information regarding user preferences like search history likes/dislikes, rating to an item previously user purchased, current location of the user.
- Learning Phase: In this phase RS are trained using information collected in previous phase for better understanding of user taste.
- Prediction Phase: In this phase RS finally predicts the user relevant items for recommendation [7].

II.   Related Work

In paper [8], recommendation play a very important role in e-commerce websites like Amazon, Flipkart, YouTube, Netflix etc. It helps them in retaining customer and increment in sale of items. The problem of overloaded information is solved by search engines, but the issues of personalization of data remain same. To solve this issue e-commerce industry, introduce RS. It makes the task of online seekers simple by suggesting fine variety and precise recommendation of products. Here author also discusses about different techniques of RS, its advantages and disadvantages.

In paper [9], presents a framework for understanding RS, its approaches. Also, for major issues: a) how user input is obtained & used b) contribution by people & computations, types of communication involved c) algorithms for linking people and computing suggestions d) demonstration of recommendation to users.

In paper [10], a workflow-based RS model to provide useful knowledge to collaborative team (CT) contexts rather than day to day task, like recommending news, movies etc. The CT contains information about relationship among users, roles, tasks which could combine with CF to obtain user entail. Two workflow-centric methods for mining team members are a) knowledge demand b) determining proper recommendation volume, the limitations for current methods & model are cold start problem.

In paper [11], Rec4LRW is used to facilitate research scholars, it helps in finding research paper for their literature survey. It has three key tasks: a) building a list of papers b) find similar paper c) shortlisting final reading paper.Technique used in proposed work are based on transitional set of protocols that capture the features of a scientific paper.

In paper [12] there are variety of domains where RS are applied like news, books, search queries, social tag and e-commerce sites etc. also discuss various techniques like CF, CB, clustering and classification etc. Used to get better recommendations hence reduction in high precision, MAE & accuracy.

In paper [13], author discusses the top 10 data mining algorithms i.e. SVM, Apriori, EM, k-NN, Naïve Bayes, K-means, PageRank, C4.5, AdaBoost and CART in detail. These algorithms used for clustering, classification, association analysis, link mining and statistical learning.

In paper [14], the author proposes a Hybrid CF recommender algorithm with FCM clustering, slope one and FSUBCF algorithm to solve data sparsity problem. Initially it predicts the rating of items that have note rated yet by the user using slope one algorithm based on FCM cluster. On comparing this algorithm with tradition CF algorithm gave better results.

In paper [15], a radical analysis on research articles of RS from 2011-15, total 61 articles are considered from 434 published in WOS (Web of Science) & Scopus. Also discuss some deficiencies, strengths of various recommendation technique.

In paper [16], author proposes a web-based e-learning systems, focuses on two type of affiliation 1) between user and system 2) system and open web. It is being developed using clustering and CF technique. It finds relevant content on the web, personalize and notify content based on the system consideration. This system is able to spot user's need.

In paper [17], the huge growth of information on web is tough and time-consuming problem. Author purposes a RS for research paper field using DNTC (Dynamic Normalized Tree of Concepts) model. It enhances existing model with tortuous ontology and large number of research paper. This system uses the 20112 version of ACM CCS (Computing Classification System) ontology. The proposed approach is better than previous. Here the user profiling step build a user profile as a DNT (dynamic normalized tree) using dynamic edit tree distance approach to do comparison between unseen research papers and user profile.

In paper [18], the use of RS in academic research is very helpful to research scholars to find papers related to their domain of research. Here author propose a topic analysis of paper using item-based method. Also

considering cold start problem. The proposed model is able to generate recommendations with a smaller number of user ratings.

In paper [19], supposes a method that relate the captured input and clustering algorithms to solve cold start problem. It uses cosine similarity measure to similarities between users and build clusters. It selects top n items for each cluster of users on the base of average rating of item.

III.   Implementation Details

A.   System Overview

Existing system provides recommendation for web pages in a single website on the basis of web usage and domain knowledge. Proposed system shows recommendation for many websites on the basis of browser history and domain ontology. Many researches have been put forth for webpage recommendation but few works on website recommendation. An ample of data is accumulated every day in browser history of every user. This history is volatile and has to be trapped to create more effective recommendations. The application will read and save the browser navigation data from history for every registered user. The application will generate and update the domain ontology created for every user based on his browser data. The application will use integrated search engine to generate website recommendations based on domain knowledge acquired for registered users.
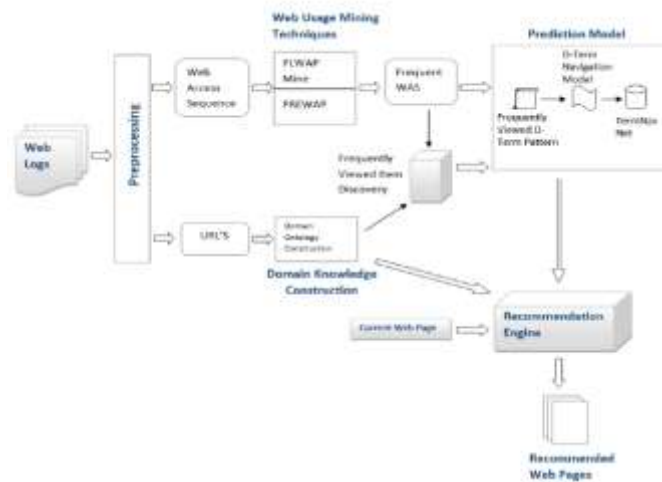

Fig. 1: System Architecture

System Module Description
1. Read Web Log / Dataset:
It consist of raw data that is surfing history of web users. Web log contain the web access sequences and the URL's set which are divided in the preprocessing phase to pass the content.
2. Preprocessing:
Used to extract useful information from non-useful data that is raw data. Web log having data in the form of URLs and web access sequence and then passed to further modules.
3. Web Usage Mining:
This technique used to analyze web access pattern that are frequently used by web users. Because of analysis of web access pattern recommendation system will performbetter then usage mining techniques are applied on that is PLWAP-Mine or PREWAPthat will generate frequent view term discovery.
4. Domain Ontology Construction:
Generate Ontology using PORTAGE tool and then extract the important data / terms and relations between terms which will overcome the new page problem improves the efficiency of system.
5. Prediction Model:
From frequent view term discovery prediction model will generate frequently viewed
D-Term Patterns.
6. Recommendation Engine:
On User Input the recommendation engine will recommend the related web pageswhich user likely to visit.

IV.   Conclusion And Future Scope

Recommendation systems are high-priority feature in the success of e-commerce age. In the epoch of online shopping, massive number of products are available on the web. Data mining is the businesslike technique of expressing information from user data, with the correct utilization of DM algorithms we can

amplify the performance of RS and solve its problems. It actually helpful to untangle the problem of RS in order to find homogeneity between users and items. This paper concludes a step by step process of building a recommendation system with the help of various technique, similarity measure matrices and challenges tackled by the various recommenders.

References

[1]. S. Agarwal, ―data mining: data mining concepts and techniques,‖ proc. - 2013 int. Conf. Mach. Intell. Res. Adv. Icmira 2013, pp. 203–207, 2014.

[2]. C. H. Moore, ―data mining based recommendation system using social websites,‖ ieee/wic/acm int. Conf. Web intell. Intell. Agent technol. Data, vol. 19, no. 1, pp. 12–13, 2015.

[3]. Y. Li, ―data mining: concepts, background and methods of integrating uncertainty in data mining,‖ pp. 2–7.

[4]. J. P. J.han, m. Kamber, data mining concepts and techniques, 3rd editio. Morgan kaufmann is an imprint of elsevier, 2011.

[5]. M. Dhanda and v. Verma, ―recommender system for academic literature with incremental dataset,‖ procedia comput. Sci., vol. 89, pp. 483–491, 2016.

[6]. T. Bogers and a. Van den bosch, collaborative and content-based filtering for item recommendation on social bookmarking websites, vol. 532. 2009.

[7]. F. O. Isinkaye, y. O. Folajimi, and b. A. Ojokoh, ―recommendation systems: principles, methods and evaluation,‖ egypt. Informatics j., vol. 16, no. 3, pp. 261–273, 2015.

[8]. N. Vaidya, ―the need of the ecommerce era,‖ ieee 2017 int. Conf. Comput. Methodol. Commun., no. Iccmc, pp. 100–104, 2017.

[9]. L. Terveen and w. Hill, ―beyond recommender systems: helping people help each other recommendation : examples and concepts,‖ hum. Comput. Interact. New millenium, no. 1, pp. 1–21, 2001.

[10]. L. Zhen, g. Q. Huang, and z. Jiang, ―recommender system based on workflow,‖ decis. Support syst., vol. 48, no. 1, pp. 237–245, 2009.

[11]. Sesagiri raamkumar, s. Foo, and n. Pang, ―rec4lrw-scientific paper recommender system for literature review and writing,‖ front. Artif. Intell. Appl., vol. 275, pp. 106–119, 2015.

[12]. P. Khurana and s. Parveen, ―approaches of recommender system: a survey,‖ int. J. Comput. Trends technol., vol. 34, no. 3, pp. 134–138, 2016.

[13]. X. Wu et al., ―top 10 algorithms in data mining,‖ spriger-verlag london ltd. 2, pp. 1–37, 2008.

[14]. Y. Cao, ―collaborative filtering recommendation combining fcm and slope one algorithm,‖ int. Conf. Inf. Cybern. Comput. Soc. Syst. -ieee, pp. 110–115, 2015.

[15]. Anandhan, l. Shuib, m. A. Ismail, and g. Mujtaba, ―social media recommender systems: review and open research issues,‖ ieee access, vol. 6, no. Figure 1, pp. 15608–15628, 2018.

[16]. T. Y. Tang and g. Mccalla, ―smart recommendation for an evolving e-learning system.‖

[17]. M. Al alshaikh, g. Uchyigit, and r. Evans, ―a research paper recommender system using a dynamic normalized tree of concepts model for user modelling,‖ ieee, 2017.

[18]. Pan and w. Li, ―research paper recommendation with topic analysis chenguang,‖ int. Conf. Comput. Des. Appliations, vol. 4, no. Iccda, pp. 264–268, 2010.

[19]. L. Yanxiang, g. Deke, c. Fei, and c. Honghui, ―user-based clustering with top-n recommendation on cold-start problem,‖ int. Conf. Intell. Syst. Des. Eng. Appl., 2013.