

Intelligent Recommendation System Using Big Data and Hadoop

Harshit Patel¹, Asst. Prof. Pooja Jardosh²

¹M.E Computer Engineer, Silver Oak College of Engineering & Technology, Gujarat, India.

²Assistant Professor, Computer/IT Department, Silver Oak College of Engineering & Technology, Gujarat, India.

ABSTRACT

More E-business Websites give items distinctive costs which made it hard for customers to discover the items and administrations they need. Keeping in mind the end goal to defeat this information over-burden, clever proposal motors are utilized to recommend items and to furnish customers with applicable information to help them choose which items to buy. Suggestion motors are exceedingly computational and subsequently perfect for the Hadoop Platform. Result shows approximately 75 to 80 percent is achieved for register users and 65 to 70 percent is achieved for unregister users. For register users accuracy is comparatively high for propose system then the existing ones E-commerce product to the register as well as the unregister user with increased accuracy by analyzing the interest of the users. This framework goes for building a Web Recommendation motor which utilizes thing or client based suggestion for prescribing Items. It will investigate the information and give recommendations taking into account what comparable clients did and on the past exchange history of the client.

Keyword: - E-Commerce, Recommender System, User-based CF, Item-based CF, Hadoop

I. INTRODUCTION

1. Recommendation System

The items can be prescribed taking into account the top general venders on a site, or on an examination of the past purchasing conduct of the buyer as a forecast for future purchasing conduct. The types of proposal incorporate recommending items to the customer, giving customized item data, outlining group sentiment, and giving group evaluates. Comprehensively, these proposal systems are a piece of personalization on a site since they help the site adjust to every client. Content outline is a chain of a packing a given archive into a condensed variation by removing the most basic data from it.

1.1 Collaborative filtering

These methodologies constructing a model from a client's past conduct (things beforehand acquired or chose and/or numerical appraisals given to those things) and also comparable choices made by different clients. This model is then used to anticipate things (or evaluations for things) that the client may have an enthusiasm for. Communitarian separating approaches regularly experience the ill effects of three issues: chilly begin, adaptability, and sparsity.

- There are two major approaches for collaborative filtering algorithms:

- **User-based approaches:** It certain majority of the customer have same taste then they join into the one group. Recommendation is given to user based on evolution of items by other users from the same group.
- **Item-based approaches:** This method tries to predict user's opinion on different items and then recommends an item based on the users transaction history as well as a numerical value that expresses the predicted likeliness of an item about which the user has not given his opinion.

1.2 Content-based filtering

This approach uses a set of discrete characteristics of an item to recommend more items with similar properties. It is based on item description and user preference.

1.3 Demographic Recommender System

Demographic recommender systems aim to classify the user based on personal attributes and make recommendations based on demographic classes. A demographic recommender system uses the personal information of the user to make predictions regarding their likes. The benefit of a demographic approach is that it may not require a history of user ratings of the type needed by collaborative and content-based techniques.

1.4 Knowledge-Based Recommender System

Knowledge-based recommender systems are a specific type of recommender system that is based on explicit knowledge about the item assortment, user preferences, and ultimately, how the item is useful for the user. Knowledge-based systems tend to work better than others at the beginning of their deployment but if they are not equipped with learning components they may be surpassed by other shallow methods that can exploit the logs of the human/computer interaction.

1.5 Hybrid Recommendation

In some cases, combining the results of any two recommendation technique like collaborative and content-based filtering proves to be more efficient. It can be done by making collaborative and content based recommendations separately and then merging them or by adding content based to collaborative approach or vice versa.

2. Big Data

Big data analysis is one of the upcoming disciplines in data mining where the large unstructured data that is very difficult to store and retrieve in an efficient manner. The three perspectives of big data are volume, velocity and variety. Volume refers to the amount of data that is being processed. Velocity refers to the speed at which the data can be processed with minimal error rate. Variety refers to all types of data starting from unstructured raw data to semi-structured and structured data.

3. Hadoop

Hadoop is an Open Source Framework given by Apache Software Foundation Founded by Doug Cutting. In Big Data large amount (Volume) of data is there may be in form of GB, TB or PB. So, Hadoop is used to manage that massive amount of data. Main aim of Hadoop is the partition a large amount of data set into data fragments. Because efficiency is the big problem for Big Data.

Hadoop includes:

HDFS: Hadoop Distributed File System

Map Reduce: Offline computing engine.

II. Problem with Existing System

This system aims to building a book recommendation engine which uses item or user based recommendation from Mahout for recommending books. After surveying various recommendation systems on

e-commerce web-site we found that most of the recommendation system recommends the product to the register users only it can't recommend the product to the un-register users of e-commerce web-site. Most of the recommendation system have cold star problem that's way accuracy of the system is less.



Fig. 1 Unstructured Data

id	ISBN	BookName	BookAuthor	Publisher	I..ip	date	session
9403	140232990	Sister Stories: Tak	Brenda Peterso	Penguin Bo	66.249.79.88	22/Jan/2016	014c5afb-8a0c-41d3-9dc0-1262b3f21877
9404	1582430349	East into Upper Eas	Ruth Prawer Jh	Counterpoi	66.249.79.88	22/Jan/2016	014c5afb-8a0c-41d3-9dc0-1262b3f21877
9405	1582430152	Heat and Dust	Ruth Prawer Jh	Counterpoi	66.249.79.88	22/Jan/2016	014c5afb-8a0c-41d3-9dc0-1262b3f21877
9406	345430808	Manifold: Origin	Stephen Baxter	Del Rey Bo	66.249.79.88	22/Jan/2016	014c5afb-8a0c-41d3-9dc0-1262b3f21877
9407	142001821	The Passion of Arte	Susan Vreeland	Penguin Bo	66.249.79.88	22/Jan/2016	014c5afb-8a0c-41d3-9dc0-1262b3f21877
9408	806511532	The Short Happy Lif	Philip K. Dick	Citadel Pr	66.249.79.88	22/Jan/2016	014c5afb-8a0c-41d3-9dc0-1262b3f21877
9409	809001608	The Treasure of the	B. Traven	Hill &	66.249.79.88	22/Jan/2016	014c5afb-8a0c-41d3-9dc0-1262b3f21877
9410	380719185	Princess: A True St	Jean P. Sasson	Avon Books	66.249.79.88	22/Jan/2016	014c5afb-8a0c-41d3-9dc0-1262b3f21877
9411	452281903	A Man Named Dave: A	David J. Pelze	Plume Book	66.249.79.88	22/Jan/2016	014c5afb-8a0c-41d3-9dc0-1262b3f21877
9412	813511895	Pinelands Folklife	Rita Zorn Moon	Rutgers Un	66.249.79.88	22/Jan/2016	014c5afb-8a0c-41d3-9dc0-1262b3f21877
16896	451159896	Dead Ringers	Bari Wood	New Amer L	66.249.79.88	23/Jan/2016	016c1407-b9f9-4cf8-b54b-f3f33b399a07
16897	3426015463	Der Rabbi	Noah Gordon	Distribook	66.249.79.88	23/Jan/2016	016c1407-b9f9-4cf8-b54b-f3f33b399a07
16898	3423085304	Gef??hrliche Praxi	Amanda Cross	Dtv	66.249.79.88	23/Jan/2016	016c1407-b9f9-4cf8-b54b-f3f33b399a07
16899	3455024947	Die Kammer.	John Grisham	Hoffmann &	66.249.79.88	23/Jan/2016	016c1407-b9f9-4cf8-b54b-f3f33b399a07
16900	345387319X	Naked.	David Sedaris	Heyne	66.249.79.88	23/Jan/2016	016c1407-b9f9-4cf8-b54b-f3f33b399a07
16901	3499146983	Tod Im Weissen Haub	P.D. James	Rowohlt Ta	66.249.79.88	23/Jan/2016	016c1407-b9f9-4cf8-b54b-f3f33b399a07
16902	3426013347	Ihres Vaters Haus	P.D. James	Droemersch	66.249.79.88	23/Jan/2016	016c1407-b9f9-4cf8-b54b-f3f33b399a07
16903	843396710X	Yo Amo a Mi Mami	Jaime Bayly	Distribook	66.249.79.88	23/Jan/2016	016c1407-b9f9-4cf8-b54b-f3f33b399a07
16904	60080841	The Stenford Wivea	Ira Levin	Perennial	66.249.79.88	23/Jan/2016	016c1407-b9f9-4cf8-b54b-f3f33b399a07

Fig.2 Structured Data

III. Propose System

Propose system is accurately recommend the product to the registered as well as the unregistered users with help of web-logs files.

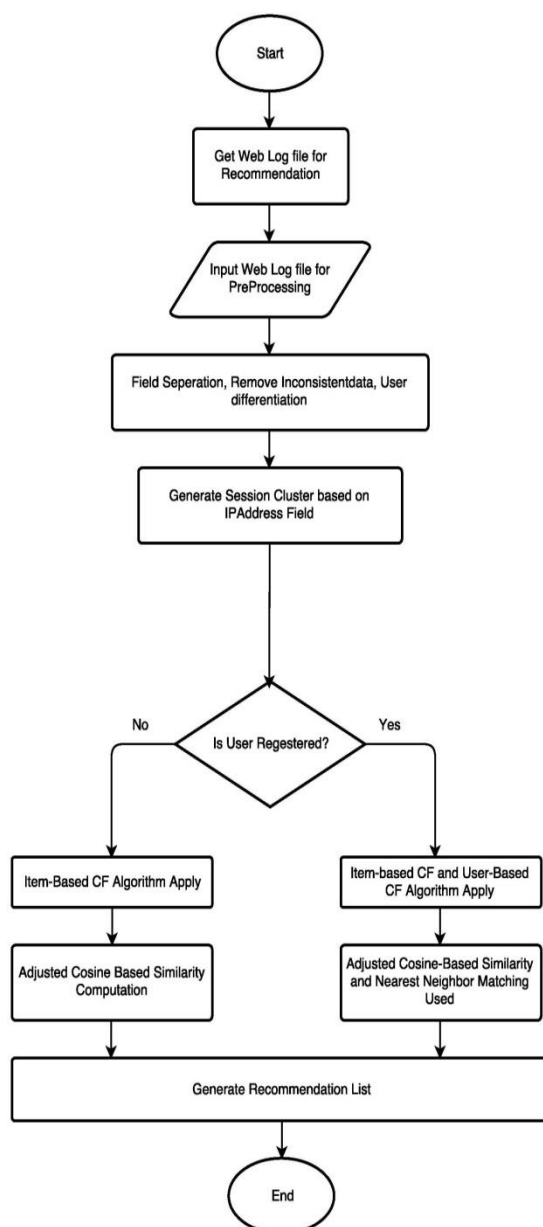


Fig. 1 Flow chart of propose system

1. Data acquisition

In this stage the whole navigational information which incorporates all the website pages went by is gathered and put away. The propose work makes utilization of regular log record configuration to keep up the information. This information got is very unstructured and conflicting in nature so that must be preprocessed for further examination.

2. Data Preprocessing

In this stage conflicting, excess information is wiped out utilizing taking after stride. Field division stage concentrates on recognizing one trait from another by making utilization of separator character, for example, comma. In information wiping stage we sift through exceptions information. We check for URL postfixes. Log sections having filename postfixes, for example, gif, jpeg, jpg are disposed of. All records having fizzled http status code are evacuated i.e status code more noteworthy than 200 and under 299 are killed. In client separation stage we dole out one of a kind ID, to every IP deliver and enlisted clients to separate one client structure other. At long last we build session in session ID stage. In this stage we aggregate together session having a place with special client.

3. Recommendation system

Two different recommendation systems have been proposed. One is for register user which is work based on user profile and another is for unregister user which is based on recent transaction history of the users.

3.1 For Register Users

If users are register then recommend the item based on user profile and transaction history of the users. In this section Item-based CF and User-based CF both approaches are apply so that this system use adjusted cosine algorithm for finding item similarity and nearest neighbor algorithm for finding similar kind of user. After finding the similarity of users and items system will generate the recommendation list for register users.

3.2 For Unregister Users

If users are unregister then recommend the item based on current transaction history of the user based on IP address of the users. In this section Item-based CF approach is apply so that this system use adjusted cosine algorithm for finding item similarity. Based on item similarity system will generate the recommendation list for unregister users.

IV. Result



Fig. 2 Result of Propose System

IV. CONCLUSION

In this research work we focus on providing good quality recommendation to all the users especially unregistered users of e-commerce site. The idea of proposed system is to develop a recommendation engine that can recommend. Results show that accuracy of proposed system is higher than the existing system. Results show that approximately 75 to 80 percent is achieved for registered users and 65 to 70 percent is achieved for unregistered users. For registered users accuracy is comparatively high for proposed system than the existing ones. E-commerce product is recommended to registered as well as the unregistered user with increased accuracy by analyzing the interest of the users.

V. REFERENCES

- [1]. Arekar, Tejal, Mrs RS Sonar, and N. J. Uke. "A Survey on Recommendation System." International Journal of Innovative Research in Advanced Engineering (IJIRAE) ISSN: 2349-2163 Volume 2 Issue 1 (2015)
- [2]. Siddiqui, Ahmad Tasnim, and Sultan Aljhdali. "Web Mining Techniques in E-Commerce Applications." arXiv preprint arXiv:1311.7388 (2013).
- [3]. Personalized recommendation engine using HADOOP Sahu, U.; Tripathy, A.K.; Chitnis, A.; Corda, K.A.; Rodrigues, S. Technologies for Sustainable Development (ICTSD), International Conference on Year: 2015 Pages: 1-6.
- [4]. Building Personalized Recommendation System With Big Data and Hadoop Map reduce S. Vinodhini, V. Rajalakshmi, B. Govindarajulu International Journal of Engineering Research & Technology Vol. 3 - Issue 4 Year: 2014 Pages: 2278-0181
- [5]. Item-based Hybrid Recommender System for newly marketed pharmaceutical drugs Bhat, S.; Aishwarya, K. Advances in Computing, Communications and Informatics (ICACCI), 2013 International Conference on Year: 2013 Pages: 2107 – 2111
- [6]. a formal concept analysis approach for web usage mining link.springer.com/content/pdf/10.1007/0387-23152-8_55.pdf by B Zhou - 2005
- [7]. Comparative Study on Approaches of Recommendation research. ijcaonline.org/volume118/number2/pxc3903059.pdf Comparative Study on Approaches of Recommendation. System. Roshni K. Sorde. Dept of CS & IT. Dr. Babasaheb Ambedkar Marathwada University.
- [8]. An Efficient Web Recommender System based on Approach of Mining Frequent Sequential Pattern from Customized Web Log Preprocessing Manisha Valera, Uttam Chauhan 4th ICCNT 2013 July 4-6, 2013.
- [9]. R. Suguna., D. Sharmila. An Efficient Web Recommendation System using Collaborative Filtering and Pattern Discovery Algorithms. International Journal of Computer Applications (0975 – 8887). Volume 70– No.3, May 2013. 37.