



Journal Homepage: - www.journalijar.com
**INTERNATIONAL JOURNAL OF
 ADVANCED RESEARCH (IJAR)**

Article DOI: 10.21474/IJAR01/9715
 DOI URL: <http://dx.doi.org/10.21474/IJAR01/9715>



RESEARCH ARTICLE

IMPLEMENTATION OF ONLINE PRODUCT RECOMMENDATION SYSTEM USING COLLABORATIVE FILTERING.

Shivani Patel and Pooja Songadkar.

Thakur College of Engineering and Technology, University of Mumbai.

Manuscript Info

Manuscript History

Received: 08 July 2019

Final Accepted: 10 August 2019

Published: September 2019

Abstract

A recommender system is an Information Retrieval [1] technology that improves access and dynamically recommends relevant items to users by considering their precisely mentioned preferences and objective behavior. Recommender system is one of the major techniques that deal with the problem of information overload by suggesting users with relevant and appropriate items. They basically direct users towards particularly those items which can meet their needs cutting down large database of Information in the process. This paper identifies the issues faced by the current systems and tries to solve them by using a hybrid system of content based and collaborative filtering technique to achieve better and more accurate performance.

Copy Right, IJAR, 2019, All rights reserved.

Introduction:-

Along with the rapid development of the Internet, the number of the servers connected to Internet and the Webs show a rise in the exponential growth. The rapid development of the Internet presents a mass of abundant information to us at the same time, for example, there are tens of thousands movies on Netflix, millions of books on Amazon, more than 10 billion page collection on Delicio.us, so much of information, not to mention find some interesting content in all that, it is impossible to give all of that information a once-over. The traditional search algorithm only presents the same ordered results to all of the users; it fails to provide different service to different users according to their different interests. Personalized recommendation has been thought as one of the most effective tools to resolve the problem of information overload. Radically, the recommendation problem is to substitute user to evaluate the products, which include books, movies, CD, web and so on, it is a process from known to unknown. [2].

Recommendation is the type of information filtering which involves prediction of rating and user preferences, which would help a user to buy items according to their needs and interest. The books suggestion at Amazon or Flipkart is the best example of a recommendation system.

Recommendation system uses various technologies to recommend products viz Content filtering and Collaborative filtering. Content based filtering recommends item which is based on user's profile, which the user has liked in the past. It serves as finding relations/correlation among items in a large database. It works on the condition where X and Y are two sets of items. It finds correlation between X and Y i.e. if we buy an item X, then it finds the probability to buy the item Y as well. Collaborative filtering is a method to analyze the user's behaviour by predicting the user's taste to that of similar to another user.

Corresponding Author:-Shivani Patel.

Address:-Thakur College of Engineering and Technology, University of Mumbai.

Proposed System:-

The proposed system uses a combination of collaborative filtering and content based recommendation system also known as personality based approach. It is a hybrid system and uses two methods in a broad light.

1. Collaborative filtering
2. Content based recommender

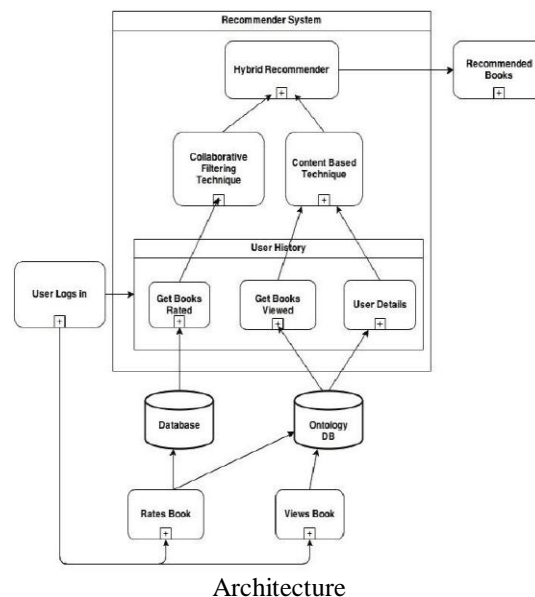
The difference between the two approaches can be highlighted by comparing - last.fm and Pandora, the two popular songs recommender system.

Last.fm – It uses collaborative filtering by creating a ‘station’ that recommend songs on the basis of what band or singer does a user listen to compares that to a similar user with similar interests and recommends the song that does not appear on UI of first user but is often played by second user. To be more specific this is an example of user based collaborative filtering.

Data for training:

There are two types of dataset we can use to train and then for prediction purpose namely implicit and explicit data. Explicit data requires effort from the users and they are seldom ready to take in any effort, it can be collected in the form of ratings, information provided in surveys etc. Implicit data is not collected on face value rather the model infers it on based on interaction of user with the system such as click through rate, watch time etc.

The asset of these two can be pooled to feed to a model and get best possible outcome.



Model:

The hybrid model in this paper is a combination of content based filtering and item based collaborative filtering. CB uses principal component analysis for dimensionality reduction similar to UV decomposition for UV matrix. Content based filtering focuses on content of item rather than its similarity with other products. The content of each product is represented as set of descriptors in a vector.

The main ideology behind using a hybrid system is to use all the available data optimally. While using CF we do not consider its content related data such as actors likewise while using CB we don't consider correlation with other users. Since the paper is for developing recommendation engine for small businesses where data is sparse in the start, we don't need to face problems of big organisation like amazon to process abundant data. But when the business grows the system should be robust enough to perform well in those circumstances hence can use clustering or categorization in early stage before using a learning model for recommendation thus dividing the users into several groups and then applying learning model for those groups.

Companies like amazon use multiple GPUs to meet time and space constraint [10] and data generation is often done on CPU along with analysis which is often snubbed. That's one of many ways for tackling big data. Even when the data is not massive enough to be called big data it may still be a huge load for a single processor to process in that we use alternating least square method which we are using in our paper.

Now in hybrid system we use data provided by user implicitly like CTR or explicitly by rating products and content of items which describes it and create a user profile. The weight of descriptors define how important that term is for recommendation. For a movie i for user u on the recommendation is [11]

$$p_{ui_{HF}} = w_{CF} * p_{ui_{CF}} + w_{CB} * p_{ui_{CB}}.$$

For example if the results for CB and CF are,

| p_{CF} | Movie title | p_{CB} | Movie title |
|----------|----------------|----------|----------------|
| 1 | Terminator | 1 | Justice League |
| 2 | The Apartment | 2 | The Prisoner |
| 3 | Justice League | 3 | Spartan |

Then recommendation can be calculated as,

| p_{HF} | Movie title | Calculation |
|----------|----------------|---------------------------|
| 1 | Justice League | $0.4 * 3 + 0.6 * 1 = 1.8$ |
| 2 | Terminator | $0.4 * 1 + 0.6 * 4 = 2.8$ |
| 3 | Spartan | $0.4 * 4 + 0.6 * 3 = 3.4$ |

Literature Survey:-

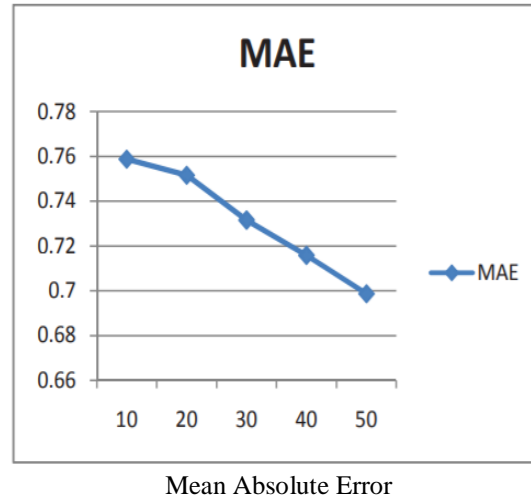
Abundant data [10] is available on the internet today and so in e-commerce. Sites like amazon, flipkart etc. have tons of data to maintain around and process them appropriately.

Sarwar and others [9] have showed the pragmatic results of different similarity algorithms and found out with the help of MAE graph that size of neighbourhood effects the prediction quality. p is the difference between the predicted ratings and actual ratings. High MAE indicates lower accuracy and vice versa for low MAE. Collaborative filtering [11] is the most widely used algorithm for recommending items.

Two types of collaborative filtering methods are proposed i.e. item based CF and user based CF out of which user based is preferred over other.

Content based filtering is used to fill the vacant ratings in the matrix. Hence the proposed amalgamation of collaborative filtering and content based recommendation system will help us to achieve the problem of data sparsity.

Hengsong Tan, et.al, have proposed a new technique to tackle data sparsity by combining item classification and item based CF. This approach classified the item through attributes and then produces prediction for items whose ratings are not available.

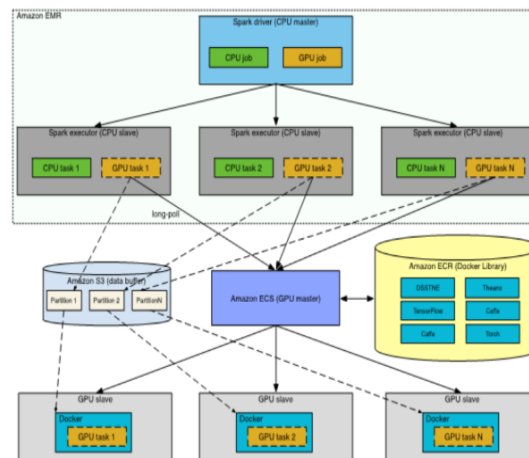


Amazon's Personalization System:

Amazon has a patent for item-to-item collaborative filtering. It is widely used by many giants for developing recommendations for their product. Amazon uses neural networks to generate personalized recommendations for its customer. With millions of users using their product they have distributed NN across multiple GPUs to meet space and time constraint [10]. Deep scalable sparse tensor neural engine is the model they have created and open sourced and is used for training and prediction.

Data generation and analysis the often overlooked area of training is processed on CPU. Now a hybrid of CPU and GPU is tough to manage since cluster managers do not inherently support GPU. They do have a solution where they run both CPU and GPU independently and drive end-to-end modelling process from Apache Spark.

Now amazon has millions of input and output nodes for which they easily reach trillions of weights for a fully connected network even if it is a shallow network and hence a single GPU is not suffice.



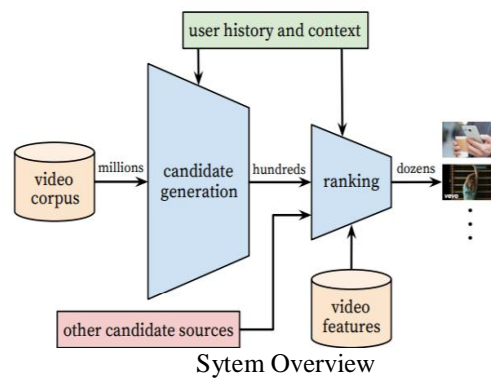
DSSTNE Architecture overview

They open sourced NN i.e. DDSTNE supports model parallel training in which the model is circulated across N-GPUs and the dataset is imitated to all nodes [10]. Each GPU then trains the model like in batches and shares the weight with rest of the nodes.

After training, prediction can be generated for each customer. This model can be scaled up by linearly adding more GPUs.

YouTube:

YouTube's system is built on Google brain open sourced as tensorflow [12]. This system consist two deep NN models, one for candidate generation and another for ranking.

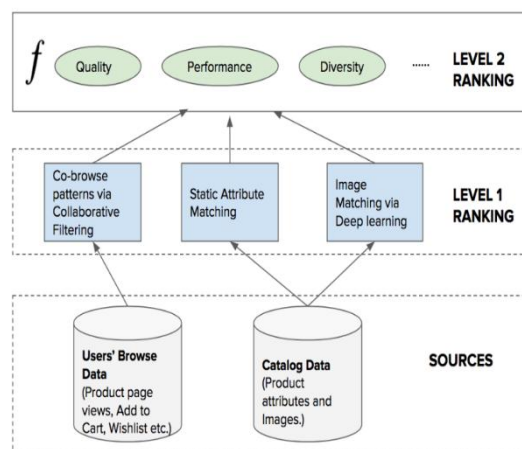


Former is used to select top 100 videos from an inventory of millions which is most relevant to a user. A high precision metric is considered for extracting these top videos. It only provides a wide perspective via collaborative filtering. They treat candidate generation model as a multiclass classification problem of a particular video.

Latter is a bit more personalized for a user and is required to have a more detailed level representation to discriminate among relative importance with high recall. It uses a similar deep neural network architecture to assign a score to each of those 100 videos using logistic regression. Eventual ranking of videos is persistently adjusted to results of live A/B testing and is usually a function of watch time per impression.

Flipkart:

Flipkart uses a hybrid of content based and collaborative filtering for recommending similar products. Users' watch history is used for training and predicting in collaborative filtering to find the most recurrent searched or looked product and product details and image catalogues is used in content based. Ranking of products is obtained by eliciting from these sources.



Flipkart's similar product recommendation ML model overview

System architecture of Flipkart's similar product recommendation consist of two levels.

Level 1 consumes of hundreds of millions of data of all users and products and is optimized to derive correlation between items and relative similarity.

Level 2 ranking [13] is optimized for performance and it uses point-wise logistic regression that can predict CTR and product to purchase conversion rate and rank them as per objective function. This model is used by flipkart to recommend in batch and in real time too.

Feasibility Analysis:-

Economic Feasibility:-

Our application will achieve accurate analysis and faster provision of recommendations, which will reduce human efforts, which in turn will reduce cost.

Technical Feasibility:

Aim to achieve accurate recommendation in various fields using collaborative filtering and content based in python language.

Operational Feasibility:

Data on which recommendation system is based is provided by the user itself and the system will analyze the pattern between the various items in the system.

Applications

Movies Recommendation:

Movies of various genres, languages, ratings etc will be recommended to the user.

News, Video searches:

When a user follows one particular type of news or videos, the further recommendation list will show similar type of news and videos.

Music apps:

The app will compile the songs played by the user and create a playlist according to the user's listening habits.

Online shopping (Clothes, electronics etc):

Depending on the previous purchases of a user, the system will recommend products of various brands, price, colors etc. A new user will have

Future Scope

Web services in particular suffer from producing recommendations of innumerable items to millions of users. The time and computational power can restrain the performance of even the best hybrid systems available. For a larger dataset we can focus on scalability problems of a recommendation system

The Prediction approach can also be tried on numerous datasets to test harmony performance of system scalability problems of recommendation systems.

Result and Discussion: -

Our system will be a web-based application which will provide the user with accurate and unerring recommendations based on the user's preferences and interests.

For the frequent users: The system will show them recommendations according to their previous searches and purchases. Similar items will be displayed to them which will conveniently match their previous choices.

For the new users: Our system will give them the top recommendations which will be based on the data collected from the frequent user.

Conclusion:-

Recommendation system provides user with the most accurate and efficient recommendation according to their interests and previous choices. For building the system we used the algorithm Collaborative filtering and Content based filtering. Both of which helped in attaining the desired results. Content based filtering recommended the item which is based on user's profile, which the user has liked in the past. It served as finding relations/correlation among

items in a large database. And Collaborative filtering was used to analyze the user's behavior by predicting the users' taste to that of similar to another user.

References:-

1. S.Khusro, and Z. Ali "Recommender Systems: Issues, challenges and research opportunities." Information Science and Applications (ICISA) 2016
2. Y. Chen, C. Wu, M. Xie and X. Guo "Solving the Sparsity Problem in Recommender Systems Using Association Retrieval" Journal of Computers, 2011
3. E. Uko Oko, B. O. Eke and P.O Asagba "An Improved Online Book Recommender System using Collaborative Filtering Algorithm" in proceedings of International Journal of Computer Applications, 2018 "Recommender System" [ONLINE]
4. Available: <https://www.studysevt.com/science/article/pii/S1110866515000341>
5. Corina Underwood, "Use Cases of Recommendation Systems in Business – Current Applications and Methods", 2019. [ONLINE] Available: <https://emerj.com/ai-sector-overviews/use-cases-recommendation-systems/>
6. J. Dianshuang and W. Mingsong "Recommender system application developments: A survey" 2015.
7. Gediminas Adomavicius, Jesse Bockstedt, Shawn Curley and Jingjing "Impact of Recommender Systems on Consumer Preferences: A Study of Anchoring Effects", California, 2014, pp 241-261
8. W. Kowalczyk, Z. Szilávik And M. C. Schut "The Impact of Recommender Systems on Item-, User-, and Rating-Diversity", in proceedings of International Workshop on Agents and Data Mining Interaction, Web, Amsterdam, 2011, pp 261-287
9. Sarwar, B., Karypis, G., Konstan, J., And Reidl, J. "Item Based Collaborative Filtering Recommendation Algorithms", in proceedings of 10th International Conference on World Wide Web, ACM, New York, 2001, pp 285-295.
10. K. Chung, "Generating Recommendations at Amazon Scale with Apache Spark and Amazon DSSTNE" 2016. [ONLINE] Available: <https://aws.amazon.com/blogs>
11. Steven Postmus, "Recommender system techniques applied to Netflix movie data". Research Paper Business Analytics, 2018.
12. Paul Covington, Jay Adams, Emre Sargin, "Deep Neural Networks for YouTube Recommendations"
13. <https://tech.flipkart.com/e-commerce-recommendations-using-machine-learning-5002526e>.