A Report on Movie Recommendations System

Abhyuday Srivastava
Department of Computer Science and Engineering
National Institute of Technology Delhi
Delhi, India
222211004@nitdelhi.ac.in

***Abstract*—Recommender systems have become ubiquitous in our lives. Yet, currently, they are far from optimal. In this project, we attempt to understand the different kinds of recommendation systems and compare their performance on the MovieLens dataset. We attempt to build a scalable model to perform this analysis. We start by preparing and comparing the various models on a smaller dataset of 100,000 ratings. Then, we try to scale the algorithm so that it is able to handle 20 million ratings by using Apache Spark. We find that for the smaller dataset, using user-based collaborative filtering results in the lowest Mean Squared Error on our dataset.**

I. Introduction

A recommendation system is a type of information filtering system which attempts to predict the preferences of a user, and make suggests based on these preferences. There are a wide variety of applications for recommendation systems. These have become increasingly popular over the last few years and are now utilized in most online platforms that we use [1]. The content of such platforms varies from movies, music, books and videos, to friends and stories on social media platforms, to products on e-commerce websites, to people on professional and dating websites, to search results returned on Google. Often, these systems are able to collect information about a users choices, and can use this information to improve their suggestions in the future. For example, Facebook can monitor your interaction with various stories on your feed in order to learn what types of stories appeal to you. Sometimes, the recommender systems can make improvements based on the activities of a large number of people. For example, if Amazon observes that a large number of customers who buy the latest Apple Macbook also buy a USB-C-to USB Adapter, they can recommend the Adapter to a new user who has just added a Macbook to his cart. Due to the advances in recommender systems, users constantly expect good recommendations. They have a low threshold for ser vices that are not able to make appropriate suggestions. If a music streaming app is not able to predict and play music that the user likes, then the user will simply stop using it. This has led to a high emphasis by tech companies on improving their recommendation systems. However, the problem is more complex than it seems. Every user has different preferences and likes. In addition, even the taste of a single user can vary depending on a large number of factors, such as mood, season, or type of activity the user is doing. For example, the type of music one would like to hear while exercising differs greatly from the type of music he'd listen to when cooking dinner.

Another issue that recommendation systems have to solve is the exploration vs exploitation problem. They must explore new domains to discover more about the user, while still making the most of what is already known about of the user. Two main approaches are widely used for recommender systems. One is content-based [2] filtering, where we try to profile the users interests using information collected, and recommend items based [3][4] on that profile. The other is collaborative filtering [5], where we try to group similar users together and use information about the group to make recommendations to the user.

II. Problem Statement

The development of the Internet brought us into an era of big data information and also make people ragged when choosing the required information and recommendation system arises at the historic moment, and get the wide attention and applications [6][7]. Many of existing algorithms work slowly on a huge amount of data. Several techniques such as clustering and parallelization were discovered to overcome the problem. It cannot produce recommendations if there are no ratings available. It demonstrate poor accuracy when there is little data about users ratings. Systems are not content aware meaning that information about items are not considered when they produce recommendations. Serendipity is the ability of the system to give an item surprisingly interesting to a user, but not expected or possibly foreseen by the user. If there are two words spelled differently but having the same meaning pure content based filtering [8] will recognize them as two independent words and will not find similarities.

III. Algorithms

We will use the Cosine Similarity from Sklearn, as the metric to compute the similarity between two movies.

Cosine similarity is a metric used to measure how similar two items are. Mathematically, it measures the cosine of the angle between two vectors projected in a multi-dimensional space.

The output value ranges from 0–1.

0 means no similarity, where as 1 means that both the items are 100% similar.

IV. IMPLEMENTATION

Recommendation System is a system which is used for filtering the information in the system that predicts rating for a given item. Recommended system identifies recommendations for individual users based on past acquisition and searches, and on other users behavior.

Recommended Systems are software tools and techniques providing suggestions for items to be of use to a user. The system helps users to match with the items which they are interested in. It will support and improve the quality of the decisions which the users searching the items in the online.

Various steps involved are ;

1) Combining Relevant Features into a Single Feature

2) Extracting Features

3) Using the Cosine Similarity

4) Generating the Similar Movies Matrix

5) Sorting the Similar Movies List in Descending Order

6) Printing the Similar Movies

CONCLUSION

Movie recommendation and selection is a fundamental issue in service oriented computing. Existing works use either content based or collaborative filtering approaches to recommend movies to the users. Such approaches possess many limitations such as poor recommendation performance and heavily relying on the input from users. This approach exploits a three-way aspect model that systematically combines classic collaborative filtering and content-based recommendation. The proposed hybrid approach simultaneously considers the similarities of user ratings and semantic content of movies data. The experimental results show that our approach outperforms the conventional collaborative and content-based methods in terms of recommendation performance.

FUTURE WORK

Future research will be done in two areas. First,includes exploring more refined/personalized recommendation by considering the specific contexts. This approach apply to other areas such as service clustering. Second, with respect to users, mining their implicit interests from usage records or reviews may be a complement to the explicit interests (ratings). By this means, recommendations can be generated even if there are only few ratings. This will solve the sparsity problem to some extent.

REFERENCES

[1] Liu, Long, Jin Xu, Stephen Shaoyi Liao and Huaping Chen. A Real-Time Personalized Route Recommendation System for Self-Drive Tourists based on Vehicle to Vehicle Communication. IEEE Transactions on Expert Systems with Applications , 41(7):3409-3417, 2014.

[2] Kim J H, Chung K Y, Ryu J K, Rim K W, and Lee J H. A Recommendation Agent System using HMM-based Collaborative Filtering in Smart Home Environment. IEEE Transactions on Datamining, 2(1):214-217, 2008.

[3] Mo, Yijun, Jianwen Chen, Xia Xie, Changqing Luo, and Laurence Tianruo Yang. Cloud-based Mobile Multimedia Recommendation System with User Behavior Information. IEEE Systems Journal, 8(1):334-338, 2014.].

[4] Jiang, Guoyin, Feicheng Ma, Jennifer Shang, and Patrick YK Chau. Evolution of Knowledge Sharing Behavior in Social Commerce: An Agent-based Computational Approach. IEEE Transactions on Information Sciences , 18(11):250-266, 2014.

[5] Yin, Zhimin, Xiangzhan Yu, and Hongli Zhang. Commodity Recommendation Algorithm based on Social Network. IEEE Transactions on Computer Science and its Applications , 34(7):27-33, 2014.

[6] Hannon J, Bennett M, and Smyth B. Recommending Twitter Users to Follow using Content and Collaborative Filtering Approaches. The conference on Recommender systems , pages 199-206, 2010.

[7] Li H, Cai F, and Liao Z. Content-Based Filtering Recommendation Algorithm using HMM. IEEE Transactions on Computational and Information Sciences, 8(1):275-277, 2012.

[8] Lu P Y, Wu X X, and Teng D N. Hybrid Recommendation Algorithm for E-Commerce Website. IEEE Transactions on Computational Intelligence and Design , 23(8):197-200, 2015.

[9] Dong F, Luo J, Zhu X, Wang Y, and Shen J. A Personalized Hybrid Recommendation System Oriented to E-Commerce Mass Data in the Cloud. IEEE Transactions on Systems, Man, and Cybernetics , 16(4):1020-1025, 2013.

[10] Modi H Y and Narvekar M. Enhancement of Online Web Recommendation System using a Hybrid Clustering and Pattern Matching Approach. IEEE conference Nascent Technologies in the Engineering Field, pages 1-6, 2015.