IJIREEICE

ISSN (Online) 2321-2004
ISSN (Print) 2321-5526

International Journal of Innovative Research in
Electrical, Electronics, Instrumentation and Control Engineering

Vol. 7, Issue 3, March 2019

Survey on Recommendation Engines built using Collaborative Filtering Techniques

Manjula HN[1], Nivin Srinivas S[2], Samuel Raj S[3]

Assistant Professor, Department of Information Science Engineering, Atria Institute of Technology, Bangalore, India[1]

Student, Department of Information Science Engineering, Atria Institute of Technology, Bangalore, India [2,3]

Abstract: With the vast amount of data available today, organizations are looking for more accurate ways of using this data for improving productivity and user experience. Recommender system is one such technology that pro-actively suggests items of interest to users based on their objective behavior on their explicitly stated preferences. Recommendation Engine is one of the most important parts of all commercial and social websites. Whenever a user searches for a book, music, movies or any other product, recommender systems play a huge role in suggesting items that are similar. Recommendations in general are of two types, content based and user based. This paper surveys Recommendation Engines using Collaborative filtering techniques.

Keywords: Recommendation Engines, Collaborative Filtering, Content-Based Filtering, Matrix Factorization

I. INTRODUCTION

In today's world data generated is too diverse, fast-changing or massive for conventional technologies, skills and infrastructure to address efficiently. This huge amount of data generated is termed as 'Big Data'. Big data refers to data sets that are enormous or too complex for conventional data- processing application software to deal with. In recent times new technologies have made it possible to realize value from Big Data Organizations are taking up new initiatives and re-evaluating existing strategies to examine how they can transform their businesses using Big Data. Using the big data generated from users from various platforms like movie streaming sites and E-commerce platforms, recommendations can be provided to users to provide personalized content and improved customer experience

II. RELATED SURVEY

The collaborative filtering is an effective technique for predicting the preferences of the user in selecting an item based on the known user ratings. It became popular in late 1990's and the online services that use recommendation engine include Amazon, Yahoo! Music, and Netflix. The scalability of this method plays an important role since collaborative filtering work on large data sets containing millions of users and items. In this work the scalable solutions are demonstrated against the Netflix Prize data set. The Netflix data set contains about 100 million ratings from over 480k users on nearly 18k movies.

Recommender systems typically produce a list of recommendations in one of two ways Content based approaches uses characteristic features, such us demographic data for user profiling, and product information/descriptions for item profiling. Collaborative Filtering (CF) approach makes use of only past user activities. Collaborative filtering algorithms identify relationships between users and items, and make associations using this information to predict user preferences. In this work the users first provides the ratings and reviews for some of the products liked by user and the system recommends similar items based on the ratings and reviews provided by the user.

The main aim of this work is to provide accurate and scalable solution. Scalability is very important since the system deals with millions of users and items. A special Matrix Factorization version is introduced that supports the visualization of user/item features, which can be used to generate explanation for recommendations. Matrix Factorization is the most widely used techniques for Collaborative Filtering problems. Various variants of Matrix Factorization are being validated against the Netflix data set. Neighbour Based (NB) approaches make use of the observation that similar users rate similar items similarly. In the NB scheme a set of similar users is selected for each query from among those who rated the active item. MF and NB approaches complement each other well. The MF approach views the data from a high level perspective. The NB approach is more localized. The combination of the MF approach and the NB approach lead to very high accurate predictions. To improve an existing MF model we add a neighbour based correction term to its answer in the prediction phase. The evaluation of the recommendation system on

DOI 10.17148/IJIREEICE.2019.7309

the Netflix data set can be applied on other datasets as well. Different correction technique, Q-correction and Neighbour based correction, are applied to improve the accuracy of the prediction [1].

Electronic retailers and content providers provide a huge selection of products to match the customers with the appropriate products to enhance user satisfaction and loyalty. Recommendation system adds another dimensionality to the user experience. Recommendation engine has become an essential part of the websites to the e-commerce giants like Amazon.com and Netflix which recommends movies, songs and TV shows to their customers. Customers indicate their level of satisfaction with particular movies through ratings, reviews, etc., so a huge volume of data is available about which movies appeal to which customers. Companies collect and utilize this data to recommend movies to particular customers.

Recommendation systems are based on one of the two approaches: The content filtering approach and the collaborative filtering approach. A known successful realization of content filtering is the Music Genome Project, which is used for the Internet radio service Pandora.com. Collaborative filtering uses previous history of users, analyses relationships between users and interdependencies among products to identify new user-item associations. Collaborative filtering is more accurate than the content filtering approach but it suffers from cold start problem which is its inability to address the new users and new products. Neighbourhood methods and latent factor models are the areas of collaborative filtering. Neighbourhood methods establish the relationship between the items or between the users.

Alternatively the latent factor models try to explain the ratings by characterizing both items and users on, say, 20 to 100 factors inferred from the ratings patterns. Latent factor model are realised using the matrix factorization method. Matrix factorization characterizes both items and users by vectors of factors inferred from item rating patterns. The input data to this recommendation system are placed in a matrix with one dimension representing the users and another dimension representing the item of interest.

The two types of input data are explicit feedback and implicit feedback. The most convenient data is high-quality explicit feedback which includes the user ratings to an item. Implicit feedback is used when explicit feedback is not available. Implicit feedback includes the search history, purchase history, browsing history, and mouse movements. The cold start problem is overcome by incorporating additional information about the users. The items popularity may change over time and the users may also change their baseline ratings over time. Matrix factorization accounts for the temporal effects reflecting the dynamic, time-drifting nature of user-item interactions. Temporal dynamics also affect user preferences and therefore the interaction between users and items. Matrix factorization has become more dominant within the collaborative filtering recommendation systems. Experiments with the Netflix Prize dataset proved that the accuracy is superior to the classical nearest - neighbour techniques. Matrix factorization is more convenient than other models because it can integrate many critical aspects of data such as multiple forms of input, temporal dynamics and confidence levels [2].

The way the products are related to each other is very important in the modern recommendation systems. For example, when a user is browsing for a mobile phone it makes sense to recommend other phones, but once the user buy the phone the system has to recommend products like battery, cases, and chargers. These are the two types of recommendations which are referred as substitutes and compliments. Substitutes are products that can be purchased instead of other product while compliments refer to the products that can be purchased in addition to the other product. Here a method to infer networks of substitutable and complementary products is developed. The primary input data for this method is the product reviews, though other data such ratings, specifications, brands are also used. The source of input for this system is taken from Amazon product catalogue which consists of 9 million products, 237 million links, and 144 million reviews. Recommendation systems are widely used in applications ranging from e-commerce to social media, video and online news platform. Making use of the large number of products to recommend the users with new and previously unknown products is the key to enhance the user experience.

The important problem in recommendation system is to understand the relationships between the products to make relevant recommendation to the given context. Therefore a product graph for these relationships is constructed where nodes represent products and the edges represent the product relationship. These graph helps in navigation between related products, discovery of new and previously unknown products, identification of interesting product combinations, and generation of better and more context-relevant recommendations. There are some challenging questions in constructing this graph like: the types of relationships, finding the relationships and how the products are related. So a network of product relationships is inferred.

So a system titled Sceptre (Substitute and Complementary Edges between Products from Topics in Reviews) is designed, that is capable of predicting the relationship between the products using the product reviews. Moreover, Sceptre harnesses the fact that products are arranged in a category hierarchy and allows us to extend this hierarchy to discover ‘micro-categories’ fine-grained categories of closely related products. Sceptre is used to build a product graph where for every product the system recommends the most relevant complimentary and substitutable products. Sceptre obtained accuracies between 91.28% and 93.67% at predicting substitutes and complements [3].

Recommender Systems collect information on the preferences of the users for a set of items. The items can be movies, songs, books, jokes, gadgets, applications, websites, travel destinations, e-learning material and so on. This information can be implicit (users' behaviour such as songs heard, movies searched) or explicit (users' ratings). Recommendation Systems (RS) use different sources of information for providing recommendations to the users. Collaborative Filtering (CF) methods along with other filtering techniques like content-based or knowledge-based are used to provide recommendations to the users.

The process for generating an RS recommendation is based on a combination of the following considerations:

i. The model chosen.
ii. The techniques being employed.
iii. The level of sparsity of the database and the desired scalability.
iv. The system performance.
v. The objective.
vi. The quality of output.

Recommendation researchers require a set of public databases to facilitate investigations on the techniques, methods and algorithms. The scientific community can use these datasets to replicate experiments to validate and improve their techniques and accuracy.

The most commonly used filtering algorithms are:

(a) collaborative filtering,
(b) demographic filtering,
(c) content- based filtering and
(d) hybrid filtering

Content-based filtering provides recommendations to users based on choices made in the past. A similarity can be established between objects that a user has bought, visited, heard, viewed and ranked positively. Demographic filtering is based on the principle that users with certain common personal attributes (sex, age, country, etc.) will also have common preferences.

Collaborative Filtering makes recommendations to each user based on information (users' ratings) provided by users. The k Nearest Neighbours (kNN) is the most widely used algorithm for collaborative filtering. Hybrid filtering commonly uses a combination of CF with demographic filtering or CF with content-based filtering to overcome the disadvantage of each one of these techniques.

CF based on the k Nearest Neighbours algorithm is very simple and generally produces good quality of predictions and recommendations but it suffers from cold-start problem. Another major problem with k Nearest Neighbour algorithm is its low scalability. A metric or a Similarity Measure (SM) is used to determine the similarity between two users or the similarity between two of items. Evaluation measures are required to test and improve the quality of RS recommendations [4].

III. CONCLUSION

Recommendations are provided to users for personalized content and improved customer experience. Recommendation engines have become a great utility to provide recommendations and harness the power of Big Data generated in this era of Data. Hence building Recommendation Engines are crucial for various applications ranging from e-commerce to social media, video and online news platform. We conclude that Collaborative Filtering is the most suitable and is widely implemented for building Recommendation Engines for Big Data and Matrix factorization is more convenient than other models because it can integrate many critical aspects of data.

REFERENCES

[1]. G. Takacs, I. Pil´aszy, B. N´emeth, and D. Tikk, "Scalable collaborative filtering approaches for large recommender systems," Journal of machine learning research, vol. 10, no. Mar, pp. 623–656,2009.

[2]. Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," Computer, vol. 42, no. 8,2009.

[3]. J. McAuley, R. Pandey, and J. Leskovec, "Inferring networks of substitutable and complementary products," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015, pp.785–794.

[4]. J. Bobadilla , F. Ortega, A. Hernando and A. Gutiérrez, "Recommender systems survey", Knowledge-based systems, vol. 46,pp.109-132,2013.

DOI 10.17148/IJIREEICE.2019.7309