International Journal of Research in Engineering, Science and Management
Volume-2, Issue-6, June-2019
www.ijresm.com | ISSN (Online): 2581-5792

Book Recommendation System using KNN Algorithm

Y. N. Bhagirathi[1], P. Kiran[2]
[1]*Student, Department of Computer Science and Engineering, RNS Institute of Technology, Bangalore, India*
[2]*Associate Professor, Dept. of Computer Science and Engg., RNS Institute of Technology, Bangalore, India*

***Abstract*: In this paper, it is described how books are recommended using KNN (K- Nearest Neighbor) algorithm. Recommendation systems are used for the purpose of suggesting items to purchase or to see. They direct users towards those items which can meet their needs through cutting down large database of Information. A various techniques have been introduced for recommending items i.e. content, collaborative and association mining techniques are used. This project helps in solving the problem of data sparcity by the use of KNN algorithm and association rule mining to achieve better performance. The system also uses the matrix factorization process in order to find the missing values and then the factorized values are passed to through the KNN algorithm.**

***Keywords*: Recommendation System, KNN Algorithm, associative rule mining**

1. Introduction

Recommendation systems are used in hundreds of different services-everywhere from online shopping to music/movies and many. Recommender systems typically produce a list of recommendations in one of two ways – through collaborative filtering or through content based filtering (also known as the personality based approach). Collaborative filtering approaches build a model from a user's past behavior (items previously purchased or selected and/or numerical ratings given to those items) as well as similar decisions made by other users. This model is then used to predict items (or ratings for items) that the user may have an interest in [8]. Content-based filtering approaches utilize a series of discrete characteristics of an item in order to recommend additional items with similar properties. Recommender systems are a useful alternative to search algorithms since they help users discover items they might not have found otherwise. Of note, recommender systems are often implemented using search engines indexing non-traditional data. There are many approaches for recommender systems used in the development of the machines through–Content-based approach and Collaborative approach or Hybrid approach (combining the above two approaches). Collaborative Filtering Approach: Collaborative filtering approach uses collecting the data and analyzing data based on the user' behaviors, preferences or activities and predicting what users will like based on their similarities with other users. It is based on the assumption that people who agreed in the past will have the same likes in the future.

Content-based Approach: This approach is based on the item description and user profile. In a content-based recommender system, keywords are used to describe the items and a user profile is built to indicate the type of item this user likes. In other words, these algorithms try to recommend items that are similar to those that a user liked in the past (or is examining in the present). In particular, various candidate items are compared with items previously rated by the user and the best matching items are recommended. This approach has its roots in information retrieval and information filtering research.

Hybrid Recommender System: This system is combination of content based and collaborative filtering approach on the data. Hybrid approaches can be implemented in several ways: by making content-based and collaborative-based predictions separately and then combining them by adding

content-based capabilities to a collaborative-based approach (and vice versa); or by unifying the approaches into one model. The recommendation systems are used in almost everywhere be it in a commercial site or at educational site or in the marketing area. These systems help in giving the new users a brief idea of how the system works and the user gets the rough sketch of what is happening in the area of the his/her interest and makes him comfortable to draw ideas and interests accordingly.

Technically a recommendation system can be defined as a software evolved for a class of data analysis (i.e. over a large data) where knowledge discovery techniques to make recommendations for the customers live.

2. Previous work

[4] Proposed a general framework for content-boosted collaborative filtering. This work improved on recommendation by boosting collaborative filtering algorithms with contents. Hence with their results, it was clear that a naïve hybrid method tends to perform better than a pure content-based or collaborative filtering algorithm.

[5] Carried out a survey of collaborative filtering techniques by first identifying the collaborative filtering tasks and the challenges facing them such as data sparsest, scalability, gray sheep, shilling attacks, privacy protections etc. and possible solutions to these challenges. The different collaborative filtering techniques such as memory-based, model-based and

hybrid collaborative filtering algorithms were represented and analyzed to determine their predictive performance and their ability to address the previously stated challenges.

[6] Proposed Amazon.com recommendations: an item-to-item collaborative filtering. This approaches solved recommendation problems with the use of an algorithm different from the traditional collaborative filtering algorithm, cluster models algorithm and search-based algorithm methods. The item-to-item collaborative filtering algorithm applied to Amazon's online shop computations was able to scale independently of the number of customer's and items in the product catalog.

[8] This paper work was carried out by Abhishek Saxena and Navneet Gaur over an e-commerce application development recommender system that defines a personalized information retrieval technique used to identify set(s) of items that will be of user interests. Apriori is mainly used to find frequently purchased items/products. The key idea behind the recommendation is that any item set that occurs frequently together must have each item (or any subset) occur at least as frequently.

3. Proposed work and methodology

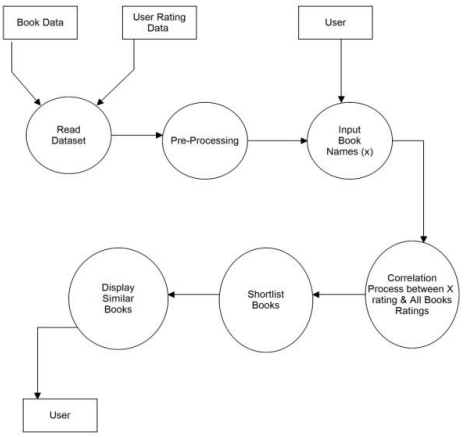

Fig. 1. System Architecture

The proposed framework utilizes blend of community oriented sifting and affiliation mining. Community oriented sifting is utilized for discovering closeness between things which would assist the framework with recommending things and affiliation digging is utilized for filling the empty evaluations where fundamental. At that point it utilizes forecast of target client to the objective thing utilizing thing based shared separating. In this manner the utilization of the two strategies can oversee information meagerly issue and cold begin issue in recommender framework.

In the proposed framework, we are contrasting various techniques utilizing Euclidean Separation, Cosine Comparability and Pearson Relationship of Community way to deal with factorize the information. At that point for expectation we are contrasting and the affiliation mining procedure and continuous thing set strategy.

In the wake of looking at all the strategies which is the best technique which give the outcome in less interim time is analyzed.

The Fig. 1 represents an outline of the proposed suggestion framework. The framework comprises of four units principally, pre-preparing unit, co-connection handling, showing. The framework comprises of two primary useful units: a component extraction unit for dimensional decrease and KNN calculation for the proposal framework.

A. Depiction of the framework

The book information and the client rating information is pre-prepared. For the handling the information i.e., the information should be changed over into the composed information by applying preparing strategies like EDA (Exploratory Information Examination). This makes the client to effectively examine the information and use it for further procedure. Once the EDA is built up, the machine applies the various strategies, similar to connection, cosine comparability, factorization methods to classify them dependent on the EDA investigation. Finally, for the suggestion part the Affiliation mining systems and successive thing set procedures are connected and the client is prescribed the top books dependent on the most astounding evaluations and notoriety of their employments. In this framework, we are attempting to contrast various strategies and various strategies with get the outcomes and pen down the various reactions and analyze their time complexities and dependent on these complexities we can make guarantee what framework is increasingly successful on the what techniques and methods.

Fig. 2 gives the flow diagram gives the detailed flow of the process for the recommendation of the system.

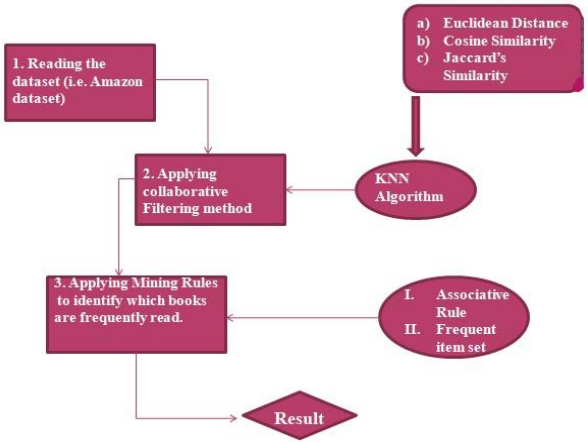

Fig. 2. Flow Diagram

The following gives the description for the above system flow:

Step I: Reading the dataset

We have the Amazon dataset- the dataset consists of all the

book details (over 50 million) book details. The details include the ISBN number of the book, the year of publication, author, ratings (user ratings), image etc.

First, the machine reads the dataset from the device. These datasets are used by the machine as the training sets in-order to complete the project. With the help of these dataset the recommendation is done.

Step II: Applying Collaborative Filtering Methods

In this step, the datasets are converted into training datasets, by classifying them into categories. Since, our project compares different methods of categorizing the datasets we have different methods.

We use KNN algorithm for filtering the datasets. The KNN (K- Nearest Neighbour) algorithm uses many methods in filtering. Here, we are using Euclidean Distance, Cosine Similarity and Pearson Correlation methods and comparing the values.

Before applying the above methods, it is necessary to factorize the dataset in case if we identify the negative values present.

Step III: Applying Mining Rules –Recommendation

In this step, final mining techniques are used for the recommendation process. These mining rules help in identifying the book or items which are frequently used among the users and are recommended to the user in need.

In this project, we are implementing two rules: Association Rules and Frequent Item Set (Apriori Algorithm) for the processed.

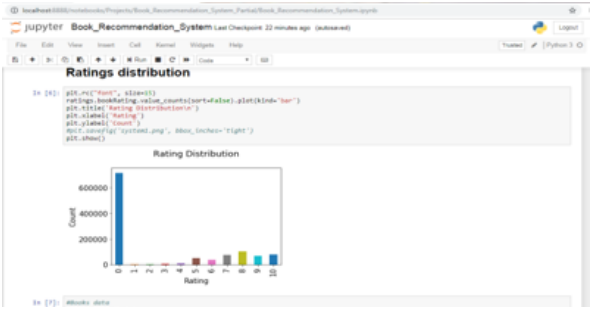

Fig. 3. Rating Distribution

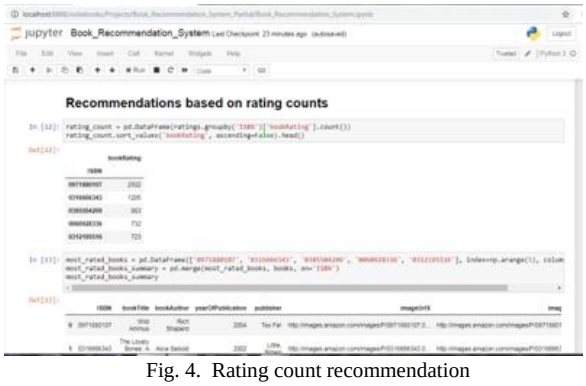

Fig. 4. Rating count recommendation

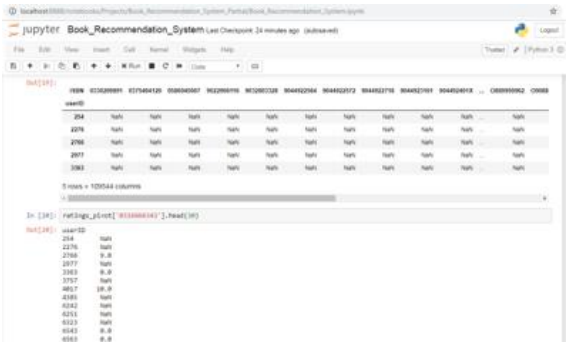

Fig. 5. Correlation value recommendation

4. Conclusion

This recommendation is provided to the users by developing a decision machine which accepts the training datasets and showcasing the resulted outputs to the users as generated by the machine. The expanding requests of online Data have lead to imagine new systems for organizing and showing things of client interests. This venture utilizes thing based community sifting method to create appraisals. The Thing based synergistic separating can evacuate the information sparsity issue and can give great proposal. At last the consequences of comparability figuring give great execution at exactness Since, we are looking at different techniques in this undertaking it encourages us in contrasting the time unpredictability of the considerable number of strategies and aides in discovering which strategy is increasingly responsive in less interim and gives progressively suitable outcomes.

In the future work, there are many different methods which are used in mining the data and can be used for recommendation process. One of the algorithm which can be used in the future work is Sparks ALS (Alternating Least Squares) Algorithm.

Acknowledgment

I would like to thank Dr. Kiran P (Dept. of CSE, RNSIT, Bangalore) for his valuable suggestions and comments that helped improving this work. This support is greatly appreciated.

References

[1] Mahmood, T., and Ricci, F. (2009). Improving recommender systems with adaptive conversational strategies. Hypertext ACM publications, 73–82.
[2] Akshita, J., and Smita, A. (2013). Recommender system: review. International Journal of computer application, 71(24), 38-42.
[3] Burke, R. (2007). Hybrid web recommender systems. The Adaptive Web, Springer Berlin, Heidelberg, 377–408.
[4] Onah, D. and Sinclair, J. (2015). Collaborative filtering recommendation system: a framework in massive online courses. Proceedings of the 9th International Technology, Education and Development Conference, Madrid, Spain, 1249-1257.
[5] Su, X., and Khoshgoftaar T. M. (2009). A survey of collaborative filtering techniques. Advances in artificial intelligence. Hindawi publishing corporation. 1-19.
[6] Linden, G., Smith, B., and York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing, 7(1), 76–80.
[7] Cheng, Z., and Hurley, N. (2010). Effective diverse and obfuscated attacks on model-based recommender systems. RecSys '09: Proceedings

of the Third ACM Conference on Recommender Systems, ACM, New York, NY, USA, 141–148.
[8] Abhishek Saxena and Nanvneet Gaur, “Frequent Item set based Recommendation System using Apriori Algorithm,” International Journal of Science, Engineering and Technology Research, vol. 4, no. 5, pp. 1609-1612, May 2015
[9] J. Han, M. Kamber, Data Mining: Concepts and Techniques, The Morgan Kaufmann Series, 2001.
[10] Agrawal, R., Imielinski, T., Swami, A. N. "Mining association rules between sets of items in large databases". In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, 207216, 1993.
[11] S. S. Sohail, J. Siddiqui and R. Ali, "Book recommendation system using opinion mining technique," *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Mysore, 2013, pp. 1609-1614.
[12] A. Jayasekar, K. C. Akshay Karan, “Collaborative filtering using Euclidean Distance in Recommendation Engine,” 2016.
[13] "Open Source Community". Anaconda.