A Literature Survey on Different Techniques Based Recommendation Systems

[1]Jagadish Kalava, [2]Dr. Pramod Pandurang Jadhav
[1]Research Scholar, Dept. of C.S.E, Dr.A.P.J. Abdul Kalam University, Indore- Dewas Bypass Road, Indore, M.P, India
[2]Research Guide, Dept. of C.S.E, Dr.A.P.J. Abdul Kalam University, Indore- Dewas Bypass Road, Indore, M.P, India

ABSTRACT: From the past decade, there are many recommended system are being used for determining various problem in different field like entertainment, social media and e-commerce. Recommendation systems are like intelligent filters which process a huge amount of information with respect to the preferences of the user and then provide suggestions to them in the form of recommendations such as recommending movies to watch, items to purchase on an e-commerce store, places to visit, etc. This paper presents A literature survey on different techniques based Recommendation systems. This survey paper explains benefits, their use-cases, and metrics used to compare various algorithms have been discussed. In addition to giving a thorough overview of recommendation systems, this study sheds light on a variety of technologies and trends in the service industry, where recommendation systems are used, for the benefit of several academics who are interested in recommendation systems.

KEYWORDS: Recommendation systems, social media, e-commerce.

I. INTRODUCTION

The variety and number of products and services provided by companies have increased dramatically during the last decade. Companies produce a large number of products to meet the needs of customers. Although this gives more options to customers, it makes it harder for them to process the large amount of information provided by companies [1]. All the promotions and discounts are depending on analytics and business research done by professionals inside and outside of different firms. Consumer reviews and product rating are the main parameters that companies used in the e-commerce sites in order to strategize the analysis [2]. These reviews provide a crucial role in users to decide about buying an item or not. Thus, examining consumer feedbacks help shopping companies and manufacturers who can identify specific areas of improvement in their products [3].

It is very difficult to read through each individual review of different items and make a good decision for an individual customer [4]. Recommender systems help customers by presenting products or services that are likely of interest to them based on their preferences, needs, and past buying behaviors. Nowadays, many people use recommender systems in their daily life such as online shopping, reading articles, and watching movies [5].

A recommender system has to identify above feedbacks and predict or estimate the ratings of items that a particular user has not yet interacted with [6]. Therefore, a recommender system is more like a function that matches a user-item pair to a rating which is the prediction itself of the recommender system. Then can be used to evaluate the accuracy of the recommender system considering whether the item will be actually rated by the user.

A major approach to the task of recommendation is called collaborative filtering which uses the user's past interaction with the item to predict the most relevant content. Collaborative filtering technology is currently the most widely used personalized recommendation technology, The core idea has two parts: The first, we calculate the similarity between users according to the user's personalized information; The second, we predict the target user preferences in relation to other products using his neighbors with higher similarity and recommend products for the target user according to his preferences. Collaborative filtering technology and content-based recommendation technology is different. It has no special requirements for recommendation target and could handle music, movies and other unstructured objects [7].

Another common approach is content-based recommendation, which uses features between items and/or users to recommend new items to the users based on the similarity between features. However, amongst the various approaches for collaborative filtering, matrix factorization is the most popular one, which projects users and items into a shared latent space, using a vector of latent features to represent a user or an item [8]. Thereafter, a user's interaction with an item is modeled as the inner product of their latent vectors. Collaborative filtering needs a considerable amount of previous history of interaction before it can provide high quality recommendation. This problem is known as the typical cold start problem [9]. The cold start problem, also known as the bootstrapping problem, is one of the issues to be handled when developing recommender systems. It happens when there are not enough, or no previous history related to a user's interaction with items within the recommendation system. In a situation like this, it is not possible to provide meaningful recommendations to the user [10].

Each recommendation scenario has its own issues which creates the need for different approaches for building recommendation systems. For example, news recommendation may put more focus on the freshness of the content while other systems like that of movie recommendation may emphasize more on content relatedness. Adding to this, specifically in the case of news, user interests keep evolving/changing over time. It might be possible that a user who reads news articles only pertaining to politics may suddenly develop interest in sports due to various reasons.

This paper presents A literature survey on different techniques based Recommendation systems. This survey paper is benefit of several academics who are interested in recommendation systems. Remaining paper is organized as follows: Section II presents Literature survey and Section III presents conclusion of paper.

II. LITERATURE SURVEY

In [11] Machine Learning with Big Data: Challenges and Approaches are explained. traditional machine learning approaches were developed in a different era, and thus are based upon multiple assumptions, such as the data set fitting entirely into memory, what unfortunately no longer holds true in this new context. These broken assumptions, together with the Big Data characteristics, are creating obstacles for the traditional techniques. In the machine learning context, size can be defined either vertically by the number of records or samples in a dataset or horizontally by the number of features or attributes it contains. This is perhaps the easiest dimension of Big Data to define, but at the same time, it is the cause of numerous challenges as Processing Performance, Class Imbalance, Curse of Dimensionality and Non-Linearity. The variety of Big Data describes not only the structural variation of a dataset and of the data types that it contains, but also the variety in what it represents, its semantic interpretation [7] and its sources. The challenges associated with this dimension are Data Locality, Data Heterogeneity and Dirty and Noisy Data. The velocity dimension of Big Data refers not only to the speed at which data are generated, but also the rate at which they must be analyzed. Data Availability, Real-Time Processing/Streaming and Concept Drift are main challenges with velocity dimension. The veracity of Big Data refers not only to the reliability of the data forming a dataset, but also, as IBM has described, to the inherent unreliability of data sources. Data provenance and Data Uncertainty are challenges involved in veracity.

There are two main categories of solutions for challenges. The first category relies on data, processing, and algorithm manipulations to handle Big Data. The second category involves the creation and adaptation of different machine learning paradigms and the modification of existing algorithms for these paradigms. The use of the Big Data definition to categorize the challenges of machine learning enables the creation of causeeffect connections for each of the issues. Furthermore, the creation of explicit relations between approaches and challenges enables a more thorough understanding of ML with Big Data. With the advent of Big Data, many of the assumptions upon which the algorithms rely have now been broken, thereby impeding the performance of analytical tasks. In response to those pitfalls, together with the need to process large datasets fast, a number of new machine learning approaches and paradigms have been developed. However, it remains consistently difficult to find the best tools and techniques to tackle specific challenges.

In [12] presents Logistic recommendation algorithm based on collaborative filtering. Traditional collaborative filtering recommendation technology has lower accuracy and couldn't meet the actual needs. So we convert the recommendation problem into the classification problem. Experimental use MovieLens dataset, which is collected by GroupLens Group. It contains 943 users and 1682 movies, It consists of three documents, The first is user ratings about product file in which user score 1-5 points for movies that had been seen, 1 point represents most dislike, 5 means a favorite, a total of 100,000 records. The second is user personalization profile in which there are 943 user's records and each record includes gender, age, occupation and other property. The third is product personalization profile in which there are 1682 product's records and record includes date, type and so on. We introduce Logistic classification methods based on the collaborative filtering technology and determine whether the recommending units are recommendable according to personal feature of users and products. Experiments use 5-fold cross-validation approach and the training set and test set is ratio of 4: 1, then get the mean of 5 times final results. Experimental results show that logistic recommendation algorithm based on collaborative filtering could get better recommendation effect than traditional collaborative filtering algorithm. We could significantly improve the accuracy of recommendation under the premise of general recall and have higher F1 value.

In [13] presents User Profile Feature-Based Approach to Address the Cold Start Problem in Collaborative Filtering for Personalized Movie Recommendation. Movie recommendation systems assist users to find the next interest or the best recommendation. the number of recommendation techniques used in the system depends on the system requirement. Thus, it is important to understand the features and potentials of different recommendation techniques. In this analysis, MovieLens_100k data set collected by the GroupLens is used. The main goal of this study is to personalize each user preferences based on the feature scores generated out of the previous ratings given to the items. Since all the users are having independent ratings on the items they consumed, the pattern they generate is identical to each user. Thus, identical data from the rating history was extracted and a feature score profile for each user was created. Those feature scores were used to identify similar users. This model contains three major modules as follows; 1) Creating user feature profiles based on past records. 2) Calculate similarity between profiles based on feature scores. 3) Generate recommendations based on created similarity matrix. With the results acquired from Modified Collaborative Filtering (MCF), the accuracy improvement of 8.36% only with maximum of five user record is a significant finding in addressing the cold-start problem which is a major drawback in Collaborative Filtering. Thus, the research implication of this study is introducing MCF as a suitable approach to be used when there are few User-Item Interaction Records (UIIRs).

In [14] described User specific product recommendation and rating system by performing sentiment analysis on product reviews. The approach proposed in this paper is novel and serves as a better alternative to rate a product based on its technical specification by analyzing large number of user reviews which are extracted dynamically from several top e-commerce websites. This avoids the need, for the user, to search for opinions and comments online before making a purchase. The Ajax Google API service is used to search the product in the Flipkart and Amazon website. For extraction there are several tools available, some of them are import.io, parsehub, Handy extractor, Helium Scrapper etc. After extracting customer reviews and specifications of the product, text content of the review was scraped out and then it is filtered to remove unwanted symbols, meaningless words, smiley's, stop words. The alchemy API is used for sentiment analysis in order to determine the polarity of individual features in the review. It uses machine learning algorithms to extract semantic meta-data from text content.

After the NLP processing the words are defined with specific tags. So this output can be given to alchemy API. After identifying feature polarity in the review, the classification has to be done in the specification list to calculate an individual feature score. a lexical database is used in the English language called as WordNet and it groups different words in English into sets of synonyms called as Synsets (semantically equivalent data elements). Based on this Synsets, the feature polarity is classified under appropriate specification. Each specification is assigned a score based on polarity i.e. positive/negative feedback. Overall product rate is calculated by aggregating the score specific to individual features. Accuracy of the proposed system is 88.33%. It is also observed that if the user reviews are containing special characters and emoticon texts which express the user opinion, system may not use those reviews for product score calculation. The overall analysis of product will be shown as a graphical output as it will be easy for a layman user to understand better.

In [15] described Deep Neural Architecture for News Recommendation. This model is a hybrid of user-item collaborative filtering (using implicit feedback) which uses the content of the items as well. The key factor in user-item based collaborative filtering is to identify the interaction between user and item features. For this work we use the dataset published by CLEF NewsREEL 2017. The Deep Semantic Structured Model (DSSM) was proposed for the purpose of ranking. Essentially, DSSM can be viewed as a multi-view learning model that often composes of two or more neural networks for each individual view. The resulting model is named as multiview DNN (MV-DNN) since it can incorporate item information from more than one domain and jointly optimize them using the same loss function in DSSM. We modify the way in which inputs are sent to the user view in order to adapt it specifically for the case of news recommendation. One of the major issues in news recommendation is that of changing user interests. Interests of users can be classified into short term as well as long term interests. Hence, it becomes crucial for a news recommender to identify these interests and recommend accordingly. LSTMs have shown to be capable of learning long-term dependencies. Users interests keep changing over time and at the time of recommendation we need to know the current interest and the long term user interest. The performance of a ranked list is judged by Hit Ratio (HR) and Normalized Discounted Cumulative gain (NDCG). Overall we see that as we increase the amount of reading history used, the performance also increases. This shows that a user has multiple interests which slowly get captured as the number of articles used for the user view of Recurrent Attention DSSM is increased. The results promise the efficiency of our model to handle the problem of user cold start as well. This shows the adaptability of our model for other recommendation scenarios which purely rely on implicit feedback.

In [16] presents Joint Deep Modeling of Users and Items Using Reviews for Recommendation. The proposed model, named Deep Cooperative Neural Networks (DeepCoNN), consists of two parallel neural networks coupled in the last layers. One of the networks focuses on learning user behaviors exploiting reviews written by the user, and the other one learns item properties from the reviews written for the item. DeepCoNN models user behaviors and item properties using reviews. It learns hidden latent factors for users and items by exploiting review text such that the learned factors can estimate the ratings given by users. It is done with a CNN based model consisting of two parallel neural networks, coupled to each other with a shared layer at the top. We have performed extensive experiments on a variety of datasets as Beer dataset, Yelp dataset and Amazon dataset to demonstrate the effectiveness of DeepCoNN compared to other state-of-the-art recommender systems in terms of MSE. In comparison with state-of-the-art baselines, DeepCoNN achieved 8.5% and 7.6% improvements on datasets of Yelp and Beer, respectively. On Amazon, it outperformed all the baselines and gained 8.7% improvement on average. Overall, 8.3% improvement is attained by the proposed model on all three datasets. Experimental results showed that for the users and items with few reviews or ratings, DeepCoNN obtains more reduction in MSE than MF. Especially, when only one review is available, DeepCoNN gains the greatest MSE reduction.

In [17] presents Music recommendation system using emotion triggering low-level features. However, researches on music-related human emotions have much difficulty due to the subjective perception of emotions. We present a personalized music recommendation system, implemented using the proposed low-level features. This recommendation system eliminates scalability problem of tag-based music recommendation systems, by employing automatically extracted low-level features of music which trigger emotions. By analyzing user's listening history, we tried to reduce the semantic gap between low-level features and high level semantic classification information and to effectively reflect dynamic changes of user's behavior of selecting songs depending on listening environments. Regression analysis is a statistical model of predicting dependent variable based on the combinations of given independent variables. The recommendation module retrieves listening history, low-level features, and the context information from the database and creates recommended list of songs to be sent to the client. For this evaluation experiments, 30 users used the given recommendation system for a certain period of time and gave their evaluation on the four different modules of recommendation system. There are 400 songs in the database of the recommendation system and each user listened average of 96 songs using the recommendation system. The users are requested to give evaluation of each module by selecting one of four criteria of very good, good, not bad, and bad. The best prediction is done with pvalue of 0.05. As the p-value gets closer to 0.05, the performance of prediction gets better. Actually, p-value of 0.05 is the value recommended most in statistical experiments.

We notice that the features extracted by the proposed method may not be the best or complete features for representing aural features triggering human emotions. However, by increasing the size of the training set and by getting help from various fields of studies such as psychology, music and emotions, we believe that more accurate and general set of features shall be extracted. As the experimental results shows that the features extracted from the training set of 16 competitions performs better than any of the training set of 15 competitions, we also plan to increase the training set hoping to find a saturation point of feature selection.

III. CONCLUSION

In this paper A literature survey on different techniques based Recommendation systems is described. Recommendation of products to attract customers that meet their requirements is very important for the vendors to survive in the global market. Any environment facilitates the communication and information transmission face this problem. Therefore, the recommender system is applied in this environment to help the users to reach their favorite items, documents and services in faster and easier way. This survey presents advantages and challenges of types of recommender systems. The use of the Big Data definition to categorize the challenges of machine learning enables the creation of causeeffect connections for each of the issues. Furthermore, the creation of explicit relations between approaches and challenges enables a more thorough understanding of ML with Big Data. Also, it will directly support researchers and professionals in their understanding of those types of recommender systems.

IV. REFERENCES

[1] J. Mackiewicz, D. Yeats and T. Thornton, "The Impact of Review Environment on Review Credibility," in IEEE Transactions on Professional

Communication, vol. 59, no. 2, pp. 71-88, June 2016, doi: 10.1109/TPC.2016.2527249.

[2] H. D. N. Harshika, N. Yamada, M. Nishi, K. Sugiura and N. Fukuta, "Toward Extracting and Predicting Instance-Specific Attribute Values from E-Commerce Sites for Used Products," 2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), Kumamoto, Japan, 2016, pp. 704-709, doi: 10.1109/IIAI-AAI.2016.60.

[3] E. E. Manganari, E. Mourelatos and E. Dimara, "Beyond the Lexical Sense of Online Reviews: The Role of Emoticons and Consumer Experience," in Interacting with Computers, vol. 32, no. 1, pp. 475-489, Aug. 2020, doi: 10.1093/iwc/iwab004.

[4] J. Mackiewicz and D. Yeats, "Product Review Users' Perceptions of Review Quality: The Role of Credibility, Informativeness, and Readability," in IEEE Transactions on Professional Communication, vol. 57, no. 4, pp. 309-324, Dec. 2014, doi: 10.1109/TPC.2014. 2373891.

[5] Z. Zhao, J. Wang, H. Sun, Y. Liu, Z. Fan and F. Xuan, "What Factors Influence Online Product Sales? Online Reviews, Review System Curation, Online Promotional Marketing and Seller Guarantees Analysis," in IEEE Access, vol. 8, pp. 3920-3931, 2020, doi: 10.1109/ACCESS.2019.2963047.

[6] S. V. Aciar, G. I. Aciar, C. A. Collazos and C. S. González, "User Recommender System Based on Knowledge, Availability, and Reputation From Interactions in Forums," in IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, vol. 11, no. 1, pp. 18-22, Feb. 2016, doi: 10.1109/RITA.2016.2518441.

[7] S. Mensah, C. Hu, X. Li, X. Liu and R. Zhang, "A Probabilistic Model for User Interest Propagation in Recommender Systems," in IEEE Access, vol. 8, pp. 108300-108309, 2020, doi: 10.1109/ACCESS.2020.3001210.

[8] M. Tkalcic, A. Odic, A. Kosir and J. Tasic, "Affective Labeling in a Content-Based Recommender System for Images," in IEEE Transactions on Multimedia, vol. 15, no. 2, pp. 391-400, Feb. 2013, doi: 10.1109/TMM.2012.2229970.

[9] S. -M. Choi, K. Jang, T. -D. Lee, A. Khreishah and W. Noh, "Alleviating Item-Side Cold-Start Problems in Recommender Systems Using Weak Supervision," in IEEE Access, vol. 8, pp. 167747-167756, 2020, doi: 10.1109/ACCESS.2020.3019464.

[10] A. Nanopoulos, "Item Recommendation in Collaborative Tagging Systems," in IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 41, no. 4, pp. 760-771, July 2011, doi: 10.1109/TSMCA.2011.2132708

[11] A. L'Heureux, K. Grolinger, H. F. Elyamany and M. A. M. Capretz, "Machine Learning With Big Data: Challenges and Approaches," in IEEE Access, vol. 5, pp. 7776-7797, 2017, doi: 10.1109/ACCESS.2017.2696365.

[12] Zhang Xiaoyu, Dai Chaofan and Zhao Yanpeng, "Logistic recommendation algorithm based on collaborative filtering", 2nd International Workshop on Materials Engineering and Computer Sciences (IWMECS 2015), 2015, 10.2991/iwmecs-15.2015.169.

[13] L. Uyangoda, S. Ahangama and T. Ranasinghe, "User Profile Feature-Based Approach to Address the Cold Start Problem in Collaborative Filtering for Personalized Movie Recommendation," 2018 Thirteenth International Conference on Digital Information Management (ICDIM), Berlin, Germany, 2018, pp. 24-28, doi: 10.1109/ICDIM.2018.8847002.

[14] M. V. K. Kiran, R. E. Vinodhini, R. Archanaa and K. Vimalkumar, "User specific product recommendation and rating system by performing sentiment analysis on product reviews," 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2017, pp. 1-5, doi: 10.1109/ICACCS.2017.8014640.

[15] Vaibhav Kumar, Dhruv Khattar, Shashank Gupta, Manish Gupta, Vasudeva Varma "Deep Neural Architecture for News Recommendation", Conference and Labs of the Evaluation Forum (Information Access Evaluation meets Multilinguality, Multimodality, and Visualization) (CLEF-2017)

[16] Z. Lei, N. Vahid, S. Philip, "Joint Deep Modeling of Users and Items Using Reviews for Recommendation", Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. Pp. 425-434, 2017, doi.org/10.1145/3018661.301866

[17] K. Yoon, J. Lee and M. -U. Kim, "Music recommendation system using emotion triggering low-level features," in IEEE Transactions on Consumer Electronics, vol. 58, no. 2, pp. 612-618, May 2012, doi: 10.1109/TCE.2012.6227467.