Nat. Volatiles & Essent. Oils, 2021; 8(5): 5071 – 5077

A Survey on Various Techniques for Big Data Recommendation System

Mrs. M. Kavitha[1], Dr.S. Subbaiah[2]

[1]Assistant Professor & Ph. D Part Time Research Scholar, PG & Research Department of Computer Science and Applications, Vivekanandha College of Arts and Sciences for Women, Tiruchengodu, Tamilnadu, India
[2]Associate Professor, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore, Tamilnadu, India
[1]sankavis@gmail.com, [2]subbunaren@yahoo.com

Abstract

Big data is an arising technology that refers the size of the database which is very big. To handle the huge data within a particular time is very challenge one and computational time is more. Lot of benefits provided by the big data to the users but they spent more time to fetch their relevant data. This type of problem is known as information overload. To overcome the information overload problem the recommendation system is used. It is used for finding required information from the large data. This system helps the users to give the required information based on the user's wants and their favorites. It reduces the user's spending time in browsing to fetch their relevant data and also it is providing the communication between users and the products. The different types of recommendation system are proposed by many researchers during the last few years. This paper attempts to detailed survey on recommendation system based on big data. It also discusses merits and demerits and surveyed big data recommendation system and provides a suggestion for further improvement in big data recommendation system effectively

Keywords: Big data, recommendation system, information overload, browsing time.

Introduction

Data is increased drastically because of increasing the number of internet users. The traditional relational database is failed to work in large size of data. So as to work such large size of data, big data [1] technology are now introduced. The most challenging problem in big data is to find and inspect enormous information to find relevant information required for exact aims. Because of the abundance of information in big data, users face the basic problem called information overload. It is extra problem for users to discovery stimulating data. So users wants efficient data filtering technology for finding suitable items or products.

Recommendation system [2] is used to address the information overload problem. Recommendation systems steer the business intelligence cycle and recommend useful items to the users, which normally handle massive amounts of information. Commonly, the recommendation system used two types of methods that are content based filtering and collaborative filtering to produce reference list. The content-based filtering approach make recommendations using the choice of users in the previous purchases. But in this method needs to maintain a user or item profile.

In collaborative filtering approach, predictions are done using gathering the user's likes and their preferences and some other information of many users. In this method, uses the technique that is users with similar behavior have similar buying habits. The merit of this method is it does not necessary to keep any user and product profile. Even though in this method also some problem named as cold start problem [3], sparsity and scalability. In this paper, a summary of prior academics related with the big data recommendation system is discussed. The foremost goal of this paper is reviewing in complete on different big data recommendation system. And also, their drawbacks are talked to additional progress the recommendation system for big data applications.

Survey on Various Techniques for Big Data Recommendation System

The ranking of each candidate service was adjusted for each user, and keywords were used to identify the

needs of customers and the nature of product management [4]. Afterward, a recommendation for personalized service was presented, and the use of user-based Collaborative Filtering (CF) to give the most suitable amenities for the user.

The authors [5] work out an algorithm called alternating direction method-based non-negative latent factor (ANLF) for recommendation system in context of big data. The chosen features were skilled through the single-element-dependent optimization to significantly reduce complexity in storage and computation. Then, the original optimization method was decomposed into number of subtasks and those were trained using NLF to obtain high convergence rate for recommendation system.

The R&D Project recommendation system was developed [6] by the authors. They used the research social network website for implementation. In this paper, an information filtering technique was discovered to regulate appropriate R&D projects as a candidate set first. Then, the information aggregation model was combined with dissimilar restrictions to mention suitable R&D projects for candidates.

The system for recommending proper product to the target user was developed. They proposed the method name called Structural Balance Theory-based Recommendation (SBT-Rec) This approach initially found the friends and enemies of E-commerce target users. Then, the product items chosen by target users based on the rule of structural theory and it was regarded as the recommendation candidates for target users[7].

In hospital, the patient waiting time is high when there is long queue. To avoid waiting time in hospital, first calculate the total number of patients in queue. Then easily calculate the waiting time. The patient can easily get the treatment for their diseases with in short time. So, authors wanted to predict the waiting time in hospital. So easily recommend to the patient whether the particular hospital has high queue or not. If less means it refer that hospital to the patients [8].

The models were introduced along with loop trust model and trust path model for personal recommendation model [9]. Initially, a Total Trust Degree (TTD) was used calculated based on the similarity and the users were arranged based on their TTD. An authority user set was formed using TTD values and the developed system nominated the items. Then the particular user value is selected for the new users.

The user wants to select the particular movie to watch. The appropriate film is not opened for the user at that time. So, to help the user to select the movie, system is developed called movie recommendation system [10]. In this model used the sentiment analysis process to recommend suitable movies since sentiment analysis is more consistent than simply gives rating. Initially, hybrid recommendation method was processed on the movie information and user information to attain an initial approval list and then analysis name called sentiment analysis was discovered to find final recommendation list.

To select the good knowledgeable candidates for particular job is very difficult task. So that the authors [11] wanted to develop the system for recommending candidate for the job. By using this system, the specific knowledge of the consumers was known to provide a framework for a customized recommendation. Also, a tree-based approach was developed for approaching a large volume of data by evaluating them accurately and minimized the calculation time. Taking into account complex values for professional social networks, the online mining and prediction system was adaptable for dataset, so that reliable predictions for the incessant new arrivals was produced in time.

The authors proposed two different types of collaborative filtering algorithm for recommendation bigdata [12]. The first algorithm is known as improved k-means clustering to group the big data. The second one is called principal component analysis with improved k-means clustering method for reducing the bigdata

dimension. It enhanced the exactness of recommendation system for big data.

Another recommendation system developed the group of the authors for treatment for diseases. Initially the system finds out the different types of diseases with their symptoms using density-peaked clustering analysis algorithm [13]. Then, another algorithm was developed called apriori algorithm. It used two types of rules. The two rules named as disease-treatment and disease-diagnosis rules. Based on rules, proper diagnosis and treatments will be provided for the patients.

The authors developed the distributed recommendation system for handling the bigdata. For that they used Apache spark for handling large amount of data and improve the prediction quality [14]. In this system they improved the training process. In this learning process for handling data sparsity and enhanced the prediction quality. It also handled large amount of data and increasing training task. In this system distributed matrix factorization and random forest algorithm was proposed to extra improve the total performance.

For finding the specific product, the authors used random neural network concept [15]. In this process, they uninterruptedly train the network to get the suitable product. Reinforcement Learning (RL) and Gradient Descents (GD) were used to update the network weights rewarding related attributes during penalizing irrelevant ones and to learn the related dimension center respectively. The intelligent recommender system was serving as coherence between the customers which iteratively adjusted expected relevance of the customers. Hence, a relevance criterion was provided that incorporated rating and importance which was used to validate the intelligent recommender system.

Table I illustrates an overview of advantages, disadvantages and performance metrics of above discussed techniques for big data recommendation system.

TABLE Comparison of techniques for big data recommendation system

Ref No.	**Techniques Used**	**Advantages**	**Disadvantages**	**Performance Metrics**
[4]	KASR, CF	High scalability.	Recommendation accuracy will be further improved by distinguishing the positive and negative preferences.	For Top-7 recommendations: Mean Absolute Error (MAE) = 0.4193 Normalized MAE (NMAE) = 0.1059 For 128M data size and 5 number of nodes: Speedup = 2
[5]	ANLF	High accuracy and Faster convergence rate.	Hyper-parameter decides the convergence rate of ANLF.	For Jester 1.1 M dataset: Value of Root Mean Square (RMSE) = 1.0098 Value of Mean Absolute Error (MAE) = 0.7812 Computation time = 23 seconds For MovieLens 10 M

				dataset: RMSE = 0.8006 MAE = 0.6119 Computation time = 385 seconds
[6]	Information filtering technique, Information aggregation model	It helps to meet the needs and requirements of enterprises and researchers.	Other human factors will be considered while recommending project opportunities to enhance its performance.	For Top 5 recommendation: Average Rating (AR) = 3.53 Normalized Discounted Cumulative Gain (NDCG) = 0.72
[7]	SBT-Rec	High feasibility.	Time-aware user ratings will also be considered to further improve the applicability of SBT-Rec.	At 2% density of user-product rating matrix: Mean Absolute Error (MAE) = 1.02 Recall = 32%
[8]	PTTP, HRQ	Minimize waiting time in hospitals.	Accuracy of PTTP is influenced by noisy data.	Execution time (for 5 slave nodes & 200 GB data): PTTP = 1400 s Execution time (for 5 slave nodes & 1000 requests): HQR = 18 s
[9]	Loop trust model and trust path model	Low MAE.	Performance of recommendation algorithm will be improved by including authority trust with the trust model.	For 25 number of neighbors: MAE value = 0.9024 Root Mean Square Error value (RMSE)= 0.9029
[10]	Hybrid recommendation model, sentiment analysis	Achieve high accuracy by using analysis of sentiment.	Dissimilar types of independent thoughts are elaborated inevitably in sentiment analysis. It implements properties on the outcomes.	True Positive (TP) rate = 0.761 False Positive rate = 0.239 Precision = 0.782 F-measure = 0.771
[11]	Online mining and predicting system	Reduces computational load.	User defined parameters greatly influence the performance of online mining.	For 1 round: Cumulative regret = 10000 Average regret = 0.1
[12]	Improved k-means clustering, PCA	High recommendation accuracy.	User specified k value has high impact on recommendation accuracy.	For 50 neighbors: RMSE = 0.895 MAE = 0.892
[13]	DPCA, Apriori algorithm	Derives disease treatment recommendations intelligently	It failed to address security perspectives of the recommender system.	For anemia disease: Accuracy = 88.35% Execution Time (@ 15 nodes) = 700 s

				For Hemorrhagic disorder: Accuracy = 84.2% Execution Time (@ 15 nodes) = 840 s
[14]	Distributed model, learning process, distributed matrix factorization, random forest	Less MAE.	Needs further improvement in terms of computational cost.	For Movielens 10M dataset: MAE = 0.6072 ± 0.0004 RMSE = 0.7934 ± 0.0005
[15]	RNN, RL, GD	High quality of result.	Learning speed is decreased.	Film Database: Quality (For iteration-1): GD = 0.856 RL = 0.8 Improvement (For iteration-1): GD = 0.176% RL = 0.578% Trip Advisor Car Database: Quality (For iteration-1): GD = 0.9353 RL = 0.9450 Improvement (For iteration-1): GD = 0.0003% RL = 0.008%

Conclusion

The different types of recommendation system for different applications are reviewed in this paper. Evidently, it displays some authors used different types of system to improve the performance values in the area of big data platform. Using that analysis and reviews it came know that RNN-based big data recommendation system has better performance than the other big data recommendation system. In this method quality of results is high and also it reduced the learning speed. In future, this problem will be considered and develop an efficient big data recommendation system.

References

1. Jiang, S., Qian, X., Mei, T., & Fu, Y. (2016). Personalized travel sequence recommendation on multi-source big social media. IEEE Transactions on Big Data, 2(1), 43-56.
2. Sahoo, A. K., Mallik, S., Pradhan, C., Mishra, B. S. P., Barik, R. K., & Das, H. (2019). Intelligence-based health recommendation system using big data analytics. In Big data analytics for intelligent healthcare management (pp. 227-246). Academic Press.
3. Volkovs, M., Yu, G., & Poutanen, T. (2017). Dropoutnet: Addressing cold start in recommender systems. In Advances in Neural Information Processing Systems (pp. 4957-4966).

4. Meng, S., Dou, W., Zhang, X., & Chen, J. (2014). KASR: a keyword-aware service recommendation method on mapreduce for big data applications. IEEE Transactions on Parallel and Distributed Systems, 25(12), 3221-3231.
5. Luo, X., Zhou, M., Li, S., You, Z., Xia, Y., & Zhu, Q. (2015). A nonnegative latent factor model for large-scale sparse matrices in recommender systems via alternating direction method. IEEE transactions on neural networks and learning systems, 27(3), 579-592.
6. Xu, W., Sun, J., Ma, J., & Du, W. (2016). A personalized information recommendation system for R&D project opportunity finding in big data contexts. Journal of Network and Computer Applications, 59, 362-369.
7. Qi, L., Xu, X., Zhang, X., Dou, W., Hu, C., Zhou, Y., & Yu, J. (2016). Structural balance theory-based e-commerce recommendation over big rating data. IEEE Transactions on Big Data, 4(3), 301-312.
8. Chen, J., Li, K., Tang, Z., Bilal, K., & Li, K. (2016). A parallel patient treatment time prediction algorithm and its applications in hospital queuing-recommendation in a big data environment. IEEE Access, 4, 1767-1783.
9. Yin, C., Wang, J., & Park, J. H. (2017). An improved recommendation algorithm for big data cloud service based on the trust in sociology. Neurocomputing, 256, 49-55.
10. Wang, Y., Wang, M., & Xu, W. (2018). A sentiment-enhanced hybrid recommender system for movie recommendation: a big data analytics framework. Wireless Communications and Mobile Computing, 2018.
11. Chen, W., Zhou, P., Dong, S., Gong, S., Hu, M., Wang, K., & Wu, D. (2018). Tree-based contextual learning for online job or candidate recommendation with big data support in professional social networks. IEEE Access, 6, 77725-77739.
12. Zarzour, H., Maazouzi, F., Soltani, M., & Chemam, C. (2018, May). An improved collaborative filtering recommendation algorithm for big data. In IFIP International Conference on Computational Intelligence and Its Applications (pp. 660-668). Springer, Cham.
13. Chen, J., Li, K., Rong, H., Bilal, K., Yang, N., & Li, K. (2018). A disease diagnosis and treatment recommendation system based on big data mining and cloud computing. Information Sciences, 435, 124-149.
14. Hammou, B. A., Lahcen, A. A., & Mouline, S. (2019). An effective distributed predictive model with Matrix factorization and random forest for Big Data recommendation systems. Expert Systems with Applications, 137, 253-265.
15. Serrano, W. (2019). Intelligent Recommender System for Big Data Applications Based on the Random Neural Network. Big Data and Cognitive Computing, 3(1), 15.

Authors Profile

M.Kavitha, received her Bachelor's degree (B.Sc(C.S)) from Gobi Arts & Science College (Autonomous), Bharathiar University in 2003. She received her Master Degree (M.C.A) from Gobi Arts & Science College (Autonomous), Bharathiar University 2006 and she completed M.Phil (CS) in Bharathiar University in the year of 2008. Also, she qualified in SET exam conducted by the Bharathiar University in the year of 2012. Currently, she is pursuing Part time Ph.d(C.S) in Vivekanandha College Of Arts And Sciences for Women (Autonomous) in Periyar University. Her area of research includes data mining and big data. She has published her research paper in various journals.

Dr.S. Subbaiah, received his Bachelor Degree (B.Sc. (C.S)) from University of Madras, Chennai in 1995 and he completed his Master degree M.C.A. in Bharathidasan University, Trichy in 1998 and M.Phil. (C.S) in Bharathidasan University, Trichy in 2005. He did his Ph.D.in Anna

University, Chennai in 2014. He is working as Associate Professor in Computer Science Department at Sri Krishna Arts and Science College (Autonomous), Coimbatore. His Research Area is Data Mining. His Area of Interest is Big Data Analytics, Internet of Things and Computer Networks. His experience is more than 19 years. He organized various International/National Level Conferences, Technical Symposium, Workshops and Seminars. He is a lifetime member in Indian Society of Technical Education. He received fund from EDI to conduct awareness program in Computer Science Subjects and applied for workshop, seminar and awareness program for funding agencies namely UGC, ICMR and EDI. His paper is published in various journals and presented paper in various conferences.