Big Data Analytics for Healthcare Recommendation Systems

Muhib Anwar Lambay
Research Scholar
Department of Computer Science & Engineering

B. S. Abdur Rahman
Crescent Institute of Science & Technology,
Chennai, India

Dr. S. Pakkir Mohideen
Associate Professor
Department of Computer Applications

B. S. Abdur Rahman
Crescent Institute of Science & Technology,
Chennai, India

Abstract - **Healthcare industry is an indispensable entity in the real world where large volumes of data is accumulated from time to time. Such data assumes characteristics of big data and it is desirable to analyze it and bring about latent relationships among variables in the healthcare data. Data in healthcare industry is rich in useful information. However, a comprehensive big data approach is essential to mine the data and acquire business intelligence. There are many use cases of big data analytics. However, in healthcare industry it is imperative to have knowledge-driven recommendations that help all stakeholders. With the emergence of cloud computing, big data analytics has become a reality. Distributed programming frameworks like Hadoop and Spark, to mention few, are available with associated Distributed File System (DFS) to manage big data. Many researchers contributed towards developing algorithms based on machine learning which is part of Artificial Intelligence (AI). Since healthcare industry is one of the sources of big data, it needs distributed environments for processing. Big data analytics is essential to analyze healthcare data in a comprehensive manner. The cloud computing and big data ecosystem is playing favorable role in realizing big data analytics for healthcare recommendations. A typical recommender system in healthcare industry is supposed to produce recommendations in various aspects of the domain. This paper throws light into different recommenders in healthcare domain that use big data analytics to generate recommendations. It not only provides useful insights but also discussed research gaps that can be used to investigate further to improve the state of the art.**

Keywords – Big data, big data analytics, healthcare recommender system, business intelligence

I. INTRODUCTION

The emergence of cloud computing and virtualization paved way for handling large voluminous data known as big data. Big data analytics has become an important area of research that contributes to have huge impact on different industries where business intelligence is to be derived from massive amounts of data. There are many challenges related to big data usage and value creation in healthcare industry [1]. Big data analytics is associated with machine learning and Artificial Intelligence (AI) that is part of data science. In [2] a big data analytics based recommender system is proposed based on collaborative filtering technique. They explored different filtering techniques like content based filtering, collaborative based filtering and hybrid filtering techniques. Their health recommendation system is based on classification and feature selection. It provides privacy preserving recommender system. Wang and Hajli [3] emphasized the big data analytics and its capabilities in rendering high quality services in healthcare systems. Recommendations in healthcare domain play vital role in making well informed decisions. A typical healthcare recommender system is associated with the components as illustrated in Fig. 1.

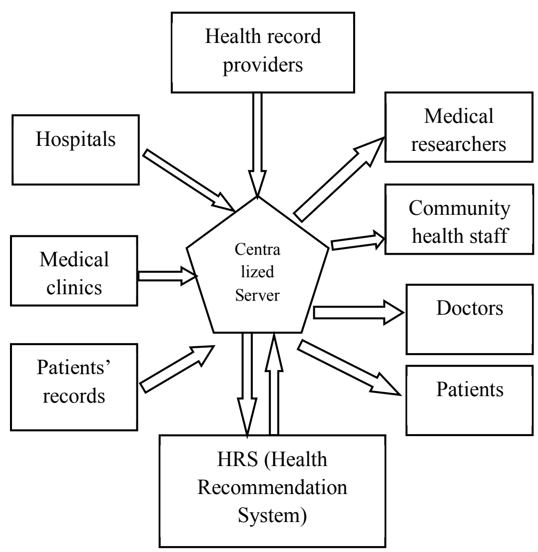

Fig. 1: *A typical healthcare recommender system and its stakeholders*

As presented in Fig.1 there are many stakeholders associated with healthcare recommender system. They include hospitals, medical clinics, patients' records, medical researchers, patients, doctors and

ISBN 978-1-7281-6202-7

community health staff. A centralized server maintains healthcare data that is used for generating recommendations. Kankanhalli et al. [4] emphasized the usage of big data analytics in order to gain good understanding about discovering hidden trends or patterns. Mehta and Pandit [5] show the close relationship between health care recommendations and big data analytics.

Our contributions in this paper are as follows:-

- We investigated into the facts pertaining to big data analytics in healthcare domain.
- We reviewed literature and some of the important recommender systems are discussed.
- We made an empirical study based on the previous observations of the researchers in order to ascertain the utility of recommender systems in healthcare domain.

The remaining part of the paper is structured as follows. Section 2 presents need for big data and possible recommender systems in healthcare domain. Different recommender systems with big data analytics are explored in Section 3. Section 4 presents results while Section 5 presents the research gaps identified. Section 6 concludes the paper besides providing directions of future work.

II. NEED FOR BIG DATA AND RECOMMENDER SYSTEMS IN HELAHTCARE INDUSTRY

In the distributed computing environments, it is possible that big data is streamed and such data needs to be processed for acquiring business intelligence. In [6], different distributed computing tools such as Kafka system, Zookeeper and so on. The recommender system showed producers, Kafka broker, stream computing and data analytics. The data analytics architecture contains data analytics with visualization, prediction, decision making, recommendations, prediction and OLAP (Online Analytical Processing). As discussed in [7], there is a huge demand for big data analytics in healthcare domain. Bellle et al. [8] explored big data analytics in healthcare domain. Their research suggests the need for big data analytics for obtaining comprehensive BI. According to Wang et al. [13] there are different phases on the data analytics. They include capturing data, data transformation and consumption.

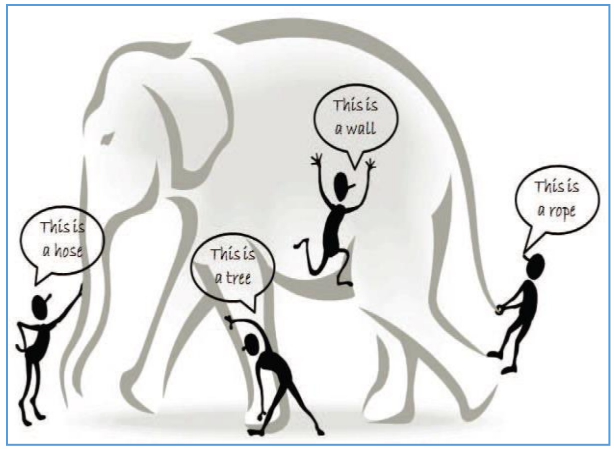

Fig. 2: *Illustrates need for unbiased Intelligence*

As presented in Fig.2, the big data analytics needs a comprehensive approach that considers all kinds of data such as structured, semi-structured and unstructured. When complete data is considered for mining or analytics, it is possible that the machine learning models produce required information for making well informed decisions. If big data is not considered, the results may not reflect full truth. Since healthcare industry is largely associated with big data, it is indispensable to have distributed programming frameworks used to generate recommendations for stakeholders of healthcare domain.

III. RECOMMENDER SYSTEMS WITH BIG DATA ANALYTICS

There are many recommender systems found in the literature. They used different techniques for providing healthcare recommendations. Here are some of the important systems reviewed to have useful insights.

A. Intelligence – Based Health Recommender System

Recommender systems are built with machine learning. When big data is considered, it is known as big data analytics. Intelligence based approach is followed in [2] for healthcare recommendations. They explored different phases found in recommender systems. The phases include information collection, learning and prediction. The last phase gives feedback to the first phase to have a life cycle of phases. There are different techniques used for recommender systems. They are known as content-based filtering, collaborative filtering and hybrid filtering. The collaborative filtering may be user based or item based or both of them. At the same time, the collaborative filtering is model based

ISBN 978-1-7281-6202-7

depending on different machine learning techniques like neural networks, Bayesian networks, association techniques and clustering techniques. The recommendation system proposed in [2] is based on the architecture presented in Fig. 3.

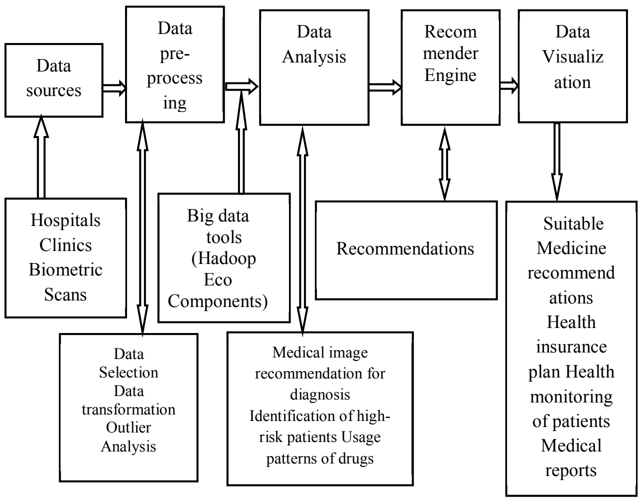

***Fig. 3:** Overview of the recommender system proposed in [2]*

The recommender system presented in Fig.3 is based on the big data analytics. It has different components involved. They include data sources such as hospitals, biometric scans, patient records and clinics; data pre-processing such as data selection, data transformation and outlier analysis; big data tools such as Hadoop, data analysis for different purposes that could generate lead to recommender engine which results in recommendations. The data visitation component takes care of showing different kinds of recommendations in graphical and intuitive fashion. They used collaborative filtering techniques. It has many advantages. They include quick access healthcare related information, cloud based information search that leverages availability and scalability, efficient service model, patient centered approach and so on. Its disadvantages include scalability problem, cold-start problem and problems with synonymy.

B. Privacy Preserving Collaborative Filtering for Health Recommendations

Kaur et al. [10] proposed a recommender system in healthcare domain that preserved privacy. Collaborative filtering is the method employed for generating recommendations. Their method is known as Privacy Preserving Collaborative Filtering (PPCF). This system provides recommendations and also ensure that the privacy is not lost. It supports patient oriented decision making. It employs the concept of Arbitrary Distributed Data (ADD) associated with healthcare services with collaborative filtering. It overcomes different issues like legal, financial and privacy. It has less performance over the method explored in [2]. However, it provides privacy to big data analytics that is missing in other recommenders explored in [11] and [12].

C. Hybrid Recommender System

Collaborative filtering model is employed in [11] for generating recommendations. In [13] a hybrid approach is used for health care recommendations. The hybrid approach is the combination of Multiple Kernel Learning (MKL) and Adaptive Neuro Fuzzy Inference System (ANFIS). The ANFIS system appears as follows. There are different layers of ANFIS model.

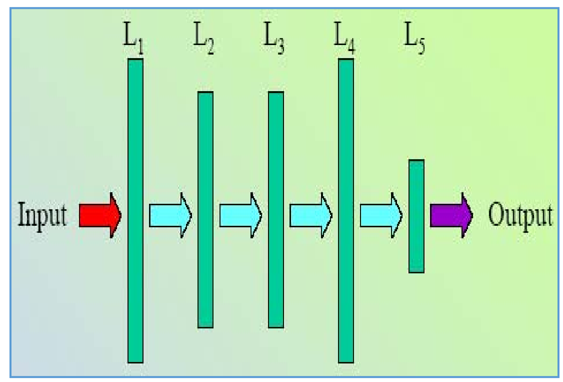

***Fig. 4:** Different layers involved in the ANFIS model*

The layers take care of membership grades, strength of rules, firing strength normalization, function computations and aggregation of outputs.

The flow of the procedures is as presented in Fig. 5

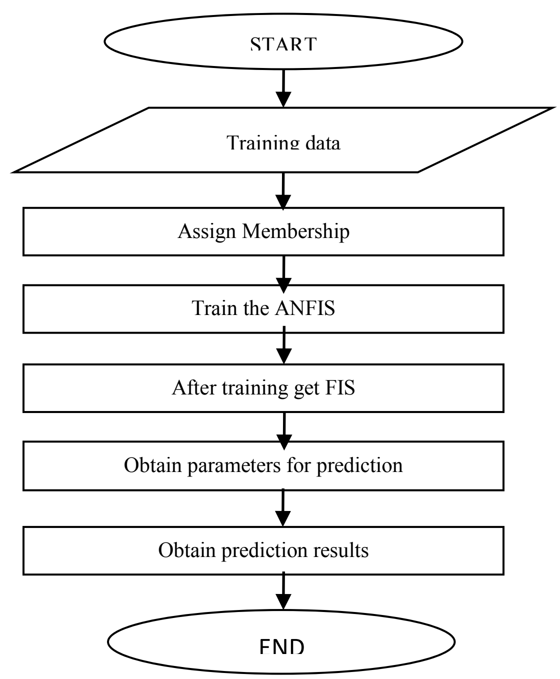

***Fig. 5:** ANFIS model*

ISBN 978-1-7281-6202-7

As presented in ANFIS model, the prediction results are obtained in the form of recommendations in the healthcare domain. It also uses MKL in order to ensure that there is hybrid approach that improves accuracy of healthcare recommendations.

D. Smart Recommender System

Nouh et al. [12] proposed a smart recommender system in healthcare domain. They used multiple learning methods with a hybrid approach. They considered personal life style as the health factor with inclusion of different aspects of personalized wellbeing. They used many learning algorithms to form a filtering approach to generate recommendations. Their methods include usage of collaborative and content-based filtering, usage of context of users and dynamic filtering, usage of individual profiles and generating feedback loops. They provided recommendations related to human wellbeing. The recommender system considers food and user profiles, content functions, inference functions, data transformation and result processing. They observed that combinations of models could improve performance. Their method improved precision. However, it has certain limitations such as cold start problem, limited customization and inability to consider hidden factors.

IV. RESULTS OF RECOMMENDER SYSTEMS

The recommender systems provided by different researchers are compared to evaluate performance. For instance, the intelligent recommender system in [2] is evaluated and found that with number of parties involved, the MAE (Mean Absolute Error) value in prediction is considered for patients and doctors' recommendations. The results are presented in Table 1.

Table 1: *Performance of the Intelligent Recommender System [2]*

No. of Parties	MAE Value	
	Patients	Doctors
1	0.82	0.935
2	0.79	0.89
3	0.738	0.818
4	0.707	0.75
5	0.69	0.69

The number of parties collaborating is considered. Around 10000 patients available in the dataset that contains 500 doctors with ratings given. And the doctors are grouped in to five different parties. MAE is the measure considered to know performance. When MAE value is lower, its accuracy is high. It is found that the results of the recommender system proposed in [2] showed better performance.

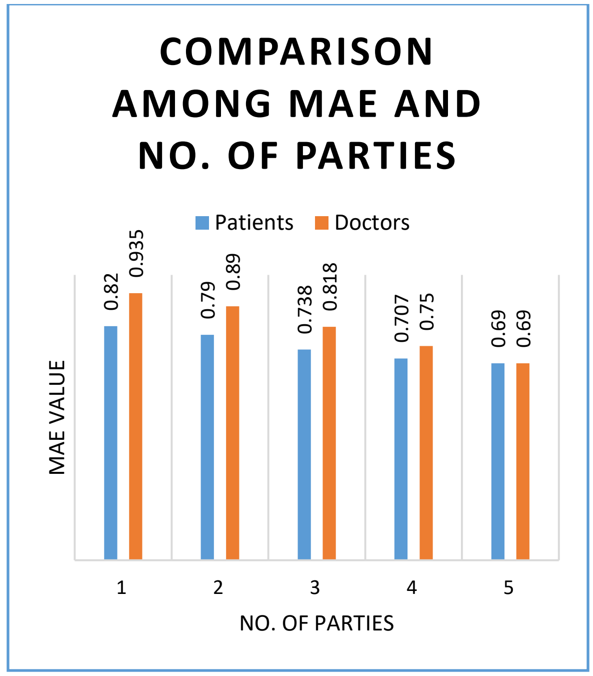

Fig. 6: *Number of parties vs. MAE*

As presented in Fig. 6, the MAE value is provided in vertical axis while the number of parties such as 1, 2, 3, 4 and 5 is presented in horizontal axis. The MAE value is recorded against each number of parties. The results revealed that there is MAE difference for predictions pertaining to doctors and patients. At the same time, the higher in MAE, the lower in accuracy. Another observation is that when more number of parties are in collaboration, the performance is increased.

Table 2: *Performance comparison among Healthcare Recommender Systems*

Recommendation System	Average MAE Values (Patients)	Average MAE Values (Doctors)
Sahoo & Pradhan [11]	0.724	0.795
Kaur et al. [10]	0.739	0.807
Sahoo et al. [2]	0.649	0.717

As presented in Table 2, it is observed that the MAE value of three recommender systems explored in [1], [10] and [11] is compared. The results revealed that the intelligent recommender system presented in [2] showed highest performance as it reveals least MAE.

ISBN 978-1-7281-6202-7

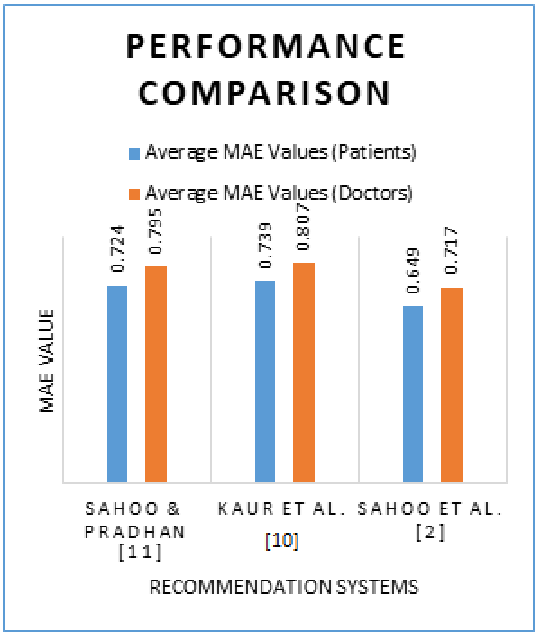

Fig. 7: *Performance comparison among three recommender systems*

The average MAE values for both patients' and doctors' recommendations are shown for three different recommender systems in healthcare domain. When average MAE is decreasing, it reflects higher performance. According to this, the recommender system provided in [2] showed higher performance than that of the ones provided in [10] and [11]. From the study of results and from the review of results, the following section provides important insights of the research that will help in improving state of the art with further investigation.

V. RESEARCH GAPS

The recommendation system proposed in [2] has certain drawbacks. First, the Collaborative Filtering (CF) algorithm suffers from scalability problem when number items and users are increased. Second, there is cold-start problem when there is less data about a doctor or patient in the system to generate meaningful recommendations. Third, there is synonymy issue when there are many items with very similar or same items with different entries or names. The recommender system proposed in [12] has limitations such as cold-start problem. It could be improved with deep learning methods. It also needs to be improved to consider user experience as a one of the factors.

VI. CONCLUSIONS AND FUTURE WORK

In this paper different recommender systems in healthcare domain based on big data analytics are examined. It is understood that there are different techniques for generating recommendations. They include content based filtering, collaborative filtering and hybrid filtering. The collaborative filtering may be memory based or model based. The memory based approach considers the user and items for collaborative filtering. On the other hand the model based filtering approaches employs machine learning algorithms in order to achieve recommender system. In this paper, we investigated on some of the important recommender systems with underlying merits and demerits. We also performed experiments based on the prior works. The results revealed that MAE determines the performance of recommender system. The more MAE is recorded, the less accuracy of the prediction model. There are some important research gaps identified as presented in Section 5. In future we intend to focus on them for improving the state of the art.

Acknowledgements

The authors would like to thank the editors and an anonymous reviewer for most constructive and developmental review process. The authors also thank Dr. E. Syed Mohamed, Professor & Head (Dept. of CSE, BSARCIST) for his insightful comments and suggestions to improve this article.

References

[1] Galetsi, P., Katsaliaki, K., & Kumar, S. (2019). Values, challenges and future directions of big data analytics in healthcare: A systematic review. Social Science & Medicine, p1-9.

[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. Big Data Analytics for Intelligent Healthcare Management, p227–246.

[3] Wang, Y., & Hajli, N. (2016). Exploring the path to big data analytics success in healthcare. Journal of Business Research, 70, p287–299.

[4] Kankanhalli, A., Hahn, J., Tan, S., & Gao, G. (2016). Big data and analytics in healthcare: Introduction to the special section. Information Systems Frontiers, 18(2), p233–235.

[5] Mehta, N., & Pandit, A. (2018). Concurrence of big data analytics and healthcare: A systematic review. International Journal of Medical Informatics, 114, p57–65.

[6] Van-Dai Ta, Chuan-Ming Liu, & Nkabinde, G. W. (2016). Big data stream computing in healthcare real-time analytics. 2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). p1-6.

[7] Palanisamy, V., & Thirunavukarasu, R. (2017). Implications of big data analytics in developing healthcare frameworks –

 ISBN 978-1-7281-6202-7

A review. Journal of King Saud University - Computer and Information Sciences. p1-30.

[8] Belle, A., Thiagarajan, R., Soroushmehr, S. M. R., Navidi, F., Beard, D. A., &Najarian, K. (2015). Big Data Analytics in Healthcare. BioMed Research International, 2015, p1–16.

[9] Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. Technological Forecasting and Social Change, 126, p3–13.

[10] H. Kaur, N. Kumar, S. Batra, An efficient multi-party scheme for privacy preserving collaborative filtering for healthcare recommender system, Futur.Gener. Comput. Syst. 86 (2018) p297–307.

[11] A.K. Sahoo, C.R. Pradhan, “A novel approach to optimized hybrid item-based collaborative filtering recommendation model using R”, In Proceedings of 9th International Conference on Advanced Computing, 2017, p468–472.

[12] Rayan M. Nouh, Hyun-Ho Lee, Won-Jin Lee and Jae-Dong Lee . (2019). A Smart Recommender Based on Hybrid Learning Methods for Personal Well-Being Services. Sensors, p1-21.

[13] Manogaran, G., Varatharajan, R., & Priyan, M. K. (2017). Hybrid Recommendation System for Heart Disease Diagnosis based on Multiple Kernel Learning with Adaptive Neuro-Fuzzy Inference System. Multimedia Tools and Applications, 77(4), p4379–4399.