© 2020 | International Journal of Scientific Research in Science and Technology
IJSRST | Volume 7 | Issue 3 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
DOI : https://doi.org/10.32628/IJSRST

A Conceptual Model of Hybrid Recommender using Big Data and Machine Learning Approach

M Venu Gopalachari*[1], A Mohan[2], B S S P Kumar[3]
*[1]Department of IT, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India
[2]Department of CSE, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India
[3]Department of MCA, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India

ABSTRACT

An exponential growth in tourism data had recorded online in the past decade due to the recent developments of web technologies and communication means. At the same time, the information overload incurred on web search engines challenges the quality of recommendations to the users although various recommenders have been developed. The main objective of these recommenders is to attract the tourists in turn promote tourism by means of advanced artificial intelligence and big data technologies. In this paper, a conceptual model is proposed for hybrid recommendation system for tourism data that considers the tourist preferences. Hybrid recommender system is the combination of the content based and collaborative filtering recommenders, which absorbs the benefits of both approaches and leads to the quality recommendations. For this, a deep learning model is developed to study the patterns in the tourism data and recommends the based on the tourist profile.
Keywords : Recommender Systems, User Profiling, Content-Based Filtering, Collaborative Filtering, Hybrid Recommender System; E-Tourism

I. INTRODUCTION

E-tourism is the field that attracted many users by means of recommending appropriate plans for trips and providing useful information through information retrieval technologies [1-5]. The basic input for this is the user profile which is ranked collection of the predicted interests of the user over a period of time [6, 7]. The traditional recommender systems consider the numeric data as key input, which are populated by the ratings given by set of users over set of products as shown in figure 1. The content based approach built the preferences using the past evaluations of the user whereas collaborative filtering approach uses the past evaluations of the other similar users with same kind of preferences [8]. The history of user profiles and what items they like, user's interactions with the items by either purchasing them or spending substantial time with the items play important role in predictions done by the system. Collaborative filtering suffers from data sparsity due to lack of enough rated data from past user interactions, either because many users do not express their preferences, for example through rating or liking an item, or because the users are new and have not interacted with the items before.

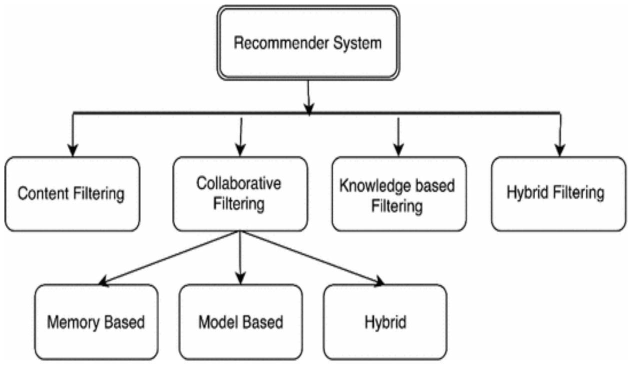

Figure 1 : Types of traditional recommenders

This lack of enough data, weakens the power of prediction thus recommendations suggested become not so accurate. Another weakness with Collaborative filtering is that its recommendations are based on similarity of items and popular items tends to have common features thus putting popular items to be recommended more, and little unknown items are never suggested, not because they are bad choices but because the CFRS are biased on similarity. Another weakness of Collaborative filtering recommender system is that they are not scalable especially with lots of computations of user item matrices that would take a lot of computation resources. In order to overcome the shortcomings of the CFRS, many models have been studied and developed so as to generate personalized recommendation systems. As outlined by some models were those that dealt with data scarcity, modeled using dimensionality reduction methods, neural networks and many other methods concludes that there were no unique models for the real world purview.

Social information also becomes part of the input in addition to rating matrix with the recent advances in the field of social networks to build social recommenders [9]. Similarly, contextual information such as location, weather conditions can also be incorporated in recommenders [10, 11, 12]. However, the tourism data is the combination of variety of information that exhibit heterogeneous nature requires hybrid approach in building recommender system. The data evolution pace in the tourism application domain demands the need of big data technologies in building recommender. The main objective of this work is to designing a framework for the hybrid tourism recommender that aims to integrate the opinion analysis.

II. RELATED WORK

The general scope of a recommender system (RS) is to find the items with maximum utility under a given user's expectations, at a given moment. Usually, user similarities and item characteristics are used to generate an ordered list of items that are most likely to satisfy a user. Based on the filters used on the item list there are various recommenders exist on tourism data with diversified aspects which can be classified in its own ways. The recommender in [14] built the user profile by reviewing the habits of the users, derived by the rating matrix, by which the similar users are defined. Some recommenders also use the text analytics such as sentiment analysis on the review text generated by user.

This information also helpful in better visualizing the available data for a particular tourist place [11]. This kind of collaborative kind of filters are difficult to built as the data related to similar users is very sparse because of the diversified tour plans of the individuals.

On the other hand, content based filtering recommender is a popular approach that considers the user's history and generates the user profile accordingly [13]. In [15], semantic relationships among data items along with meta data of the same helped in generating recommendations. The Sparsity of the user's data, over specialization of preference prediction are the basic disadvantages of content based filtering techniques. The most commonly used context elements in tourism recommendation systems are location, weather, visit history, and weather. Some recommenders uses the spatio temporal aspects as context in profiling user [16]. The hybrid approach is the combination of all these traditional approaches that try to overcome the demerits of individual approach [17, 18].

The system is a hybridization of three approaches: the content-based approach, the social approach, and the context-based approach. It should be noted that 90% of the current solutions are generally concentrated on a single category of items (hotels, museums, tourist sites, . . .)[13], providing only tourist services information (inserted in the system by the administrator or by experts) to make the trip more pleasant; besides, most of these works use a single approach, with a clear predominance for content-based approaches[13]. For all these reasons, there is a need for a conceptual framework not only to gather the recommendation approaches but also to present the different tourism resources in a single architecture.

III. PROPOSED MODEL

The architecture of the proposed conceptual model is shown in figure 2, in which the input data is three different forms. The first one is user rating data, which is a matrix that contains ratings given by the users over various items defined as R → U X I. The second one is the historical data navigated by user that provides the access data along with temporal constructs. The third one is the demographic data of the user such as location, data of birth, gender etc. The conceptual framework of the proposed architecture is made up of three main sub-processes, which are a process known as user profiling, a process for selecting content (filtering) that best matches user profiles, and a context filtering process.

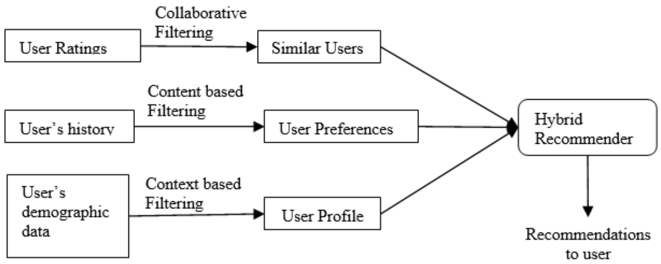

Figure 2 : Conceptual Framework for hybrid recommender system

These processes take place at the intersection of different areas of computer science research, including artificial intelligence and operational research. For example, in artificial intelligence, the profiling process can be expressed as a learning problem that exploits users' past knowledge. Often, the system should learn the user's profile rather than requiring the user to provide it. This usually involves the application of Machine Learning (ML) techniques.

The purpose of the filtering process is to learn how to categorize new information based on previously seen information that has been implicitly or explicitly labeled as interesting or uninteresting by the user. With these labels, ML methods can generate a predictive model that, given a new item, will help to decide the degree of interest the user may have in the item.

In operational research, the context filtering process leads to the formal definition of a combinatorial optimization problem. Although these attributes do not provide information on the ratings, they allow us to refine the user profile and adapt the recommendations. Furthermore, demographic data can be used to calculate recommendations for new users; the demographic approach is first used to solve the cold start problem[19, 20].

Context is based on the integration of contextual information (location, time, physical environment, : : :) for the generation of dynamic and personalized visit itineraries. A special characteristic in tourism, which distinguishes it from other domains in which recommenders have been applied, is the mobility of the users, which may need recommendations in different moments and in different places. For this reason, this particular type of recommender systems has started to incorporate context-aware techniques.The data collected about the user are then selected, analyzed, and saved as independent modules. The conceptual model has some challenges also in generating recommendations. The first challenge is the data Sparsity in the rating matrix which leads to erratic analytics. Cold start problem occurs for new users and new items as the data is very less. The second challenge is the navigation or access data which provides the user's preferences in tourism domain. But most of the tourism data portals have security concerns which do not allow the user's navigation data at full extent. The third challenge is about the demographic data that defines the context of the user's preferences in tourism. On the other hand the demographic data is not reliable most of the users may not show the interest in posting accurate data in public domains. Despite of these challenges the proposed hybrid conceptual model can generate qualitative recommendations by overcoming individual set back in each approach.

IV. RESULTS AND DISCUSSION

Initially the list of tourist spots in India is collected. Then the information regarding every tourist spots is fed to the system in order to provide the information to the user. Later some data preprocessing steps applied in which decision tree learning is one of the predictive modeling approaches used in statistics, data mining and machine learning. It uses a decision tree to go from observations about an item to conclusions about the item's target value. Next step is applying text analytics such as sentence Segmentation, Word Tokenization, Predicting Parts of Speech for Each Token, Text Lemmatization, Identifying Stop Words, Dependency Parsing, Finding Noun Phrases and Named Entity Recognition (NER).

We used three distinct data sets; in this article, they are named DS1, DS2, and DS3. Each data set represents a random sample of a larger set of the real-world tourism-related data used by the authors in other studies. The DS1 data set is a sample of 209 surveys of American pleasure travelers regarding their perception of China as a travel destination. The DS2 data set contains 332 TripAdvisor reviews of a historic attraction in the city of St. Augustine, Florida, posted between March 12, 2014, and October 12, 2015, by visitors to the site. Finally, the DS3 data set is a sample of 200 English-language Twitter messages related to the 2014 Winter Olympics held in Sochi, Russia. Together, the data sets represent three frequently used sources of data in tourism analytics: traditional surveys (DS1), travel recommender systems (DS2), and social network data (DS3). The size of the sample data sets was determined to represent a realistic training/testing target for a research team, which prevents deteriorated quality of classification. All data sets have undergone a thorough "data cleansing" process. It has been estimated that the real-world data contain up to 40% of inconsistent data. Each data set was analyzed by four automated sentiment classification programs and four programs were selected to represent choices that researchers face when dealing with sentiment data analysis, primarily in terms of the programs' underlying approach and the amount of effort required on the part of the researcher prior to data analysis. Overall, based on the accuracy, precision, and recall metrics, both SVM and Naïve Bayes classification models exhibited performance.

Table 1: Performance comparison on three data sets for SVM

	SVM		
	Accuracy	Precision	Recall
DS1	0.87	0.52	0.42

DS2	0.74	0.55	0.58
DS3	0.54	0.51	0.52

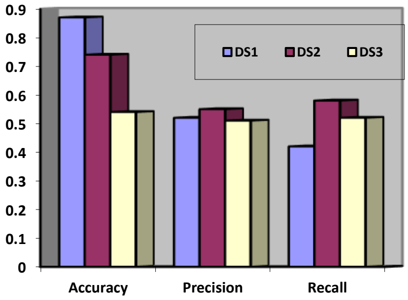

Figure 3: comparison on three data sets for evaluation metrics

The values of accuracy, precision and recall for SVM technique on three data sets are given table1. The results shown fair accuracy of recommendations with the processing model SVM on three data sets. In table2, the evaluation metrics of the accuracy, precision and recall values on three data sets foe naïve bayes technique. However the results for both techniques shown fair performance for hybrid recommenders.

Table 2: Performance comparison on three data sets for Naïve Bayes

	Naïve Bayes		
	Accuracy	Precision	Recall
DS1	0.82	0.72	0.62
DS2	0.71	0.65	0.58
DS3	0.84	0.61	0.64

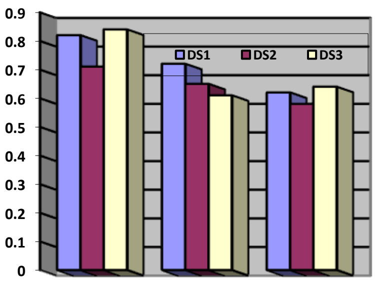

Figure 4: comparison on three data sets for evaluation metrics

V. CONCLUSION

Tourism data in the public domain had wide variety of forms showcasing the heterogeneity in the data set which makes the analysis complicated. This paper describes the current tourism recommender systems and then we have presented a new conceptual framework to implement tourism recommender systems. Our hybrid architecture aims to improve the visitor experience by recommending the most relevant items and helping him to personalize user preferences. This architecture will be implemented, through advanced technologies, such as big data tools, machine learning techniques, and the Internet of things.

VI. REFERENCES

[1]. L. Sebastia, I. Garc´ıa, E. Onaindia, and C. Guzm´an Alvarez, e-Tourism: A tourist recommendation and planning application, International Journal on Artificial Intelligence Tools, vol. 18, no. 5, pp. 717–738, 2009.

[2]. F. Ricci, L. Rokach, and B. Shapira, Introduction to recommender systems handbook, in Recommender Systems Handbook, F. Ricci, L. Rokach, B. Shapira, and P. Kantor, eds. Boston, MA, USA: Springer, 2011, pp. 1–35.

[3]. G. Adomavicius and A. Tuzhilin, Toward the next generation of recommender systems: A survey of the stateof-the-art and possible extensions, IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734–749, 2005.

[4]. M. de Gemmis, P. Lops, C. Musto, F. Narducci, and G. Semeraro, Semantics-aware content-based recommender systems, in Recommender Systems Handbook, F. Ricci, L. Rokach, and B. Shapira, eds. Boston, MA, USA: Springer, 2015, pp. 119–159.

[5]. S. Loh, F. Lorenzi, R. Salda˜na, and D. Lichtnow, A tourism recommender system based on collaboration and text analysis, Information Technology & Tourism, vol. 6, no. 3, pp. 157–165, 2003.

[6]. D. Gavalas, C. Konstantopoulos, K. Mastakas, and G. Pantziou, Mobile recommender systems in tourism, Journal of Network and Computer Applications, vol. 39, pp. 319–333, 2014.

[7]. K. N. Rao and V. G. Talwar, Application domain and functional classification of recommender systems—A survey, DESIDOC Journal of Library & Information Technology, vol. 28, no. 3, pp. 17–35, 2008.

[8]. X. Y. Su and T. M. Khoshgoftaar, A survey of collaborative filtering techniques, Adv. Artif. Intell., vol. 2009, p. 421425, 2009.

[9]. I. Cenamor, T. de la Rosa, S. N´u˜nez, and D. Borrajo, Planning for tourism routes using social networks, Expert Syst. Appl., vol. 69, pp. 1–9, 2017.

[10]. G. Fenza, E. Fischetti, D. Furno, and V. Loia, A hybrid context aware system for tourist guidance based on collaborative filtering, in Proc. IEEE Int. Conf. Fuzzy Systems, Taipei, China, 2011, pp. 131–138.

[11]. K. Meehan, T. Lunney, K. Curran, and A. McCaughey, Context-aware intelligent recommendation system for tourism, presented at 2013 IEEE Int. Conf. Pervasive Computing and Communications Workshops (PERCOM Workshops), San Diego, CA, USA, 2013, pp. 328–331.

[12]. O. Boulaalam, B. Aghoutane, D. El Ouadghiri, A. Moumen, and M. L. C. Malinine, Proposal of a big data system based on the recommendation and profiling techniques for an intelligent management of moroccan tourism, Procedia Computer Science, vol. 134, pp. 346–351, 2018.

[13]. J. Borr`as, A. Moreno, and A. Valls, Intelligent tourism recommender systems: A survey, Expert Systems with Applications, vol. 41, no. 16, pp. 7370–7389, 2014.

[14]. L. Ravi and S. Vairavasundaram, A collaborative location based travel recommendation system through enhanced rating prediction for the group of users, Computational Intelligence and Neuroscience, vol. 2016, p. 1291358, 2016.

[15]. P. Di Bitonto, M. Laterza, T. Roselli, and V. Rossano, A recommendation system to promote local cultural heritage, Journal of E-Learning and Knowledge Society, vol. 7, no. 3, pp. 97–107, 2011.

[16]. Z. Aarab, A. Elghazi, R. Saidi, and M. D. Rahmani, Toward a smart tourism recommender system: Applied to tangier city, in Innovations in Smart Cities and Applications, M. Ben Ahmed and A. Boudhir, eds. Cham, Switzerland: Springer, 2018, pp. 643–651.

[17]. J. P. Lucas, N. Luz, M. N. Moreno, R. Anacleto, A. A. Figueiredo, and C. Martins, A hybrid recommendation approach for a tourism system, Expert Systems with Applications, vol. 40, no. 9, pp. 3532–3550, 2013.

[18]. M. Hong, J. J. Jung, F. Piccialli, and A. Chianese, Social recommendation service for cultural heritage, Personal and Ubiquitous Computing, vol. 21, no. 2, pp. 191–201, 2017.

[19]. M. Vozalis and K. G. Margaritis, On the enhancement of collaborative filtering by demographic data, Web Intelligence and Agent Systems, vol. 4, no. 2, pp. 117–138, 2006.

[20]. A. T. Nguyen, N. Denos, and C. Berrut, Exploitation des donn´ees ”disponibles `a froid” pour am´eliorer le d´emarrage a froid dans les syst`emes de filtrage d'information, in Actes du XXIV Congr`es d'INFORSID, Hammamet, Tunisie, 2006,pp. 81–95.

Cite this article as :

M. Venu Gopalachari, A Mohan, B S S P Kumar, "A Conceptual Model of Hybrid Recommender using Big Data and Machine Learning Approach", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 7 Issue 3, pp. 479-484, May-June 2020. Journal URL : https://ijsrst.com/IJSRST218299