International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

Tourism Based Hybrid Recommendation System

Brian Davis Ukken[1], Rhutuja Salunke[2]

[1,2] Student, Department of Computer Engineering, Pillai HOC College of Engineering and Technology, Rasayani, Maharashtra, INDIA

---***---

Abstract - *Recommender system focuses on recommending appropriate packages to users based on their preferences and tastes by analyzing different reviews alongside the given ratings. However, users do not rate enough packages to make the collaborating filtering algorithm, leading to a cold start problem. The following solutions are included in this project to overcome the said problems: 1. the project combines Content-Based filtering along with the above algorithm to solve cold-start user problems and get higher accuracy and better precision. 2. Factors like hotels, destination, cost, and preferences are considered as a piece of additional information for a more personalised recommendation. 3. This model is also integrated with Aspect Based Sentiment Analysis to give better and more accurate results. Multiple tests were conducted on several datasets like TripAdvisor, and the results revealed that the proposed hybrid framework is competitive and superior to conventional approaches. The project also includes various elements such as a Semi-supervised Clustering algorithm which classifies the facets of the given vocabulary into nine pre-defined groups known as tour aspects. The ratings and reviews are stored in our database, which helps and enhances the desired solutions. Hence, the hybrid approach increases the efficiency of the results.*

Key Words: **collaborative filtering, recommender system, content-based filtering, tour package, predictions, ratings, hybrid model, sentiment analysis**

1. INTRODUCTION

The recent boom in the Internet and the broad scope of E-commerce has led to the flow of tremendous information. There is a massive demand for creating very sophisticated and superior systems that can process this massive surge of data. Furthermore, it should aid the users to make choices by proposing products, services, items, etc. that are similar to their respective preferences. Recommendation systems are a promising alternative to deal with these demands and issues.

The primary methods included to develop this model are CF and CB. CB [1] will suggest articles to the users which was preferred by them earlier, and CF [2], will advise things that other individual, identical in tastes, liked before [3]. Individually, each method has its own drawbacks, which include:

- Limited content analysis: Not sufficient content for the algorithm to give the desired results
- New item problem: Not enough interactions with the users.
- New user problem: The newer users cannot be recommended items since the model isn't aware of the user's tastes.
- Cold start problem, etc.

A hybrid recommender system will be implemented where the CB and CF methods will be integrated to anticipate better predictions and get better of the impediments of each approach. This research proposes a hybrid system based on users' information, ratings, and written reviews. This mainly combines collaborative filtering (CF) and content-based (CB) into the recommendation system. Here, we alternate the hybrid direction, where the model will be enhanced by incorporating aspect-based sentiment analysis. This results in the cold start problem being eliminated, which in turn would give high-performance recommendation results. The details of each tour will be stored in an SQL database, which contains general tour information, its assessment and ratings, and an item-based CF technique to predict the unrated features of tours. The lexicon-based approach determines the sentiment orientation towards tour features, and semi-supervised clustering builds the vocabulary of tour aspects. For tour searching, we use context-based information to give a more accurate recommendation.

Our project includes several contributions and improvements:

- A hybrid method to build a better and more efficient model that makes exchange between coverage and accuracy.
- A proposition that resolves the cold start problem.
- Better efficiency and precision are due to the integration of the two different algorithms.

The remaining paper contains the following sections: The next section defines the problem statement. Literature review work is represented in the third section. The fourth section represents our model along with the algorithms. The fifth section reveals the end results of our model. The paper concludes in Section 7 with the possible future of this work.

2. PROBLEM DEFINITION

For the given,

Set U = {u_1, u_2,.....,u_N} which is the user set, and

Set I = {i_1,i_2,....,i_M} which is the item set.

In recommender systems, each user rates the set RI_u = {ui_1,ui_2,...,ui_P}. [19] The user rating u on item *i*, denoted by *ru, i*, can be expressed in more than one way such as integers, expressions, etc. But we prefer to use integer values in the range 1 to 5.

Let *u* belong to the set *U* be a user, and

i belong to the set *I* be an item.

This model predicts the rating *r^u, ,i* of the user u on item *i* such that *ru* and *i* are unknown.

This system predicts the rating *r^u, i* based on the past ratings of the user u on items similar to item *i* (like content-based) or the ratings of the other users which are similar to the user (collaborative filtering) or both like the hybrid model.

3. RELATED WORK

This literature discusses the existing systems which focus on hybrid models that combine collaborative filtering and content-based approaches.

[5] suggests a method that merges materials of items and item-based CF with a clustering approach. The results merge the content and its information into CF to resolve the cold start trouble. But, this method doesn't include the user's demography, which is beneficial to improving the predictions.

[6] proposes HYDRA, a hybrid approach that unites the two algorithms. The authors have incorporated the ratings along with the content information into a federated system. They have shown that mixing features of both the items and users will produce results with better precision. Nonetheless, the authors do not exploit all features, such as the address or time of rating. They haven't explained how their method resolved the issue of cold start in their system.

[7] shows a framework that merges three different algorithms: demographic, CF, CB which filters data and information from the net such as Web pages and news, and other articles required for the model. The model uses HTML pages to collect the user's demographic information. A test group of users beta-tested the system. But tests on fewer users or items can't assure better efficiency of the said model. Also, there isn't any clarification regarding the making of the framework.

Wu Yao et al. [16] proposes a unified probabilistic model FLAME, which addresses the problem of Personalized Latent Aspect Rating Analysis. FLAME stands for Factorized Latent Aspect Model. This paper elaborates the advantages of CF and aspect-based opinion mining. This combination resolves the problem of latent aspect rating. This model sees the personalized preferences on different factsor from their reviews and foretells the aspect ratings on newer items using collective intelligence. In addition, phrase-level sentiment analyses are used to excavate the product's explicit features and the user's corresponding opinions.

Kavinkumar et al. [17] created a recommender system that uses both kinds of CF. In addition to the hybrid technique, the author proposes a framework that includes analysis of the feedback provided to improve the model's efficiency. The first feedback is an external one, here the comments are assembled from public platforms such as automobile websites and social media sites. The next feedback is an internal one, which consists of the input from the users who are provided with recommendations. The main drawback of this Hybrid system is that if technical words occur, then recommendation becomes difficult.

Leung et al. [18] details a rating inference approach which includes textual reviews in CF. A relative-frequency-based method to decide the what is orientation of the sentiment along with opinion word's strength. Then collect this orientation to calculate the user's average sentiment used. Paper resolves the data sparsity issue.

Kai Zhang et al. [19] author proposed a hybrid approach where CB is combined with CF to overcome the sparsity issue. The paper focuses on user preference analysis and the trip's intent while diversity techniques are used to optimise the tour recommendation list. The drawback is that the Preference factor model relies more on feature extraction techniques to give a better recommendation

.In addition, all these approaches lead to remarkable growth in the training time of the model. [11] proposed a model in which the content-based algorithm enhances the existing user data, then collaborative filtering is used for the prediction of ratings. These algorithms increase the result's accuracy while simultaneously covering up each other's drawbacks

Related literature on hybrid model is seen in the references.

4. THE HYBRID MODEL

The primary task of this model will be to increase the precision of the predicted ratings and have improved coverage. The rating prediction model will be discussed in this section.

4.1 Collaborative Filtering approach

It is a conventionally used recommendation approach. This is done by accumulating ratings for items given by the users in a particular domain while finding similarities between users or entities to produce suggestions. This CF technique mainly works by calculating the similarities between the profiles of multiple users based on the provided ratings which then generates newer suggestions or recommendations. The precision has proven to be excellent but sensitive to sparse data [20]. For example, there is a list of *s* users and *t* items. Each user will be able to express their opinions about the catalogue of entities.

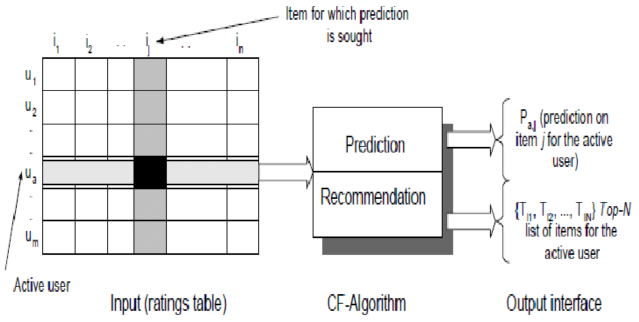

Fig-1: CF Procedure

The given matrix, there is a list of 's' users *U = (U1, U2,..., Us)* in the rows, while the columns represent the items *I = (I1, I2, It).* The user A_U's rating to item I_j is R_a, j, which suggests the user's interest.

CF is categorised into the item-based approach and user-based approach. In item-based, a user will receive suggestions of things identical to the ones the user preferred in the past. In contradistinction, user-based will see a user receiving recommendations of items that users with similar tastes liked before. The item-based will first analyse the user-item matrix to find *k* identical items or neighbours that are co-rated by distinct users similarly. For a target item, the model will generate predictions by taking a weighted average of all the active user's item ratings on the neighbour items. The item-based approach will be more systematic as it can pull off an identical or even accurate performance when compared to the user-based approach. Next, we will see the different similarity measures in CF.

4.2 Content-Based filtering

This algorithm uses the user profile which is based on items seen and rated earlier by the user. This generates recommendations based on mapping items identical to the ones the user selected in the past. These models have access to a unique compendium of keywords which the model uses to describe the items. Simultaneously, a profile is created for users, which indicates the type of items they prefer. These methods are crucial to performing exceptionally in text-intensive domains and applications with sufficient information related to the articles. This level of precision is not the desired result as it depends on the quality of extracted features. But this methodology requires better knowledge and engineering efforts to amass the necessary metadata of the items. Furthermore, users' profiles are independent of each other. For example, even if two kinds of items are frequently selected or preferred together by users, the system will never suggest items from a class if the user has not rated any items from that category.

The representation of an item is important in the CB approach since this technique might be based on the model. Specific attributes such as tags or categories can easily be stored in an organised way, so it is easy to determine. On the other hand, the unstructured text is a lot more complicated as it involves counting words and representing their significance. This model can transform a general text into structured data.

4.3 Aspect-based sentiment analysis (ABSA)

This is a text analysis approach used to categorize data by its aspect and identify the sentiment attributed to each. ABSA analyses feedbacks of the users or customers by associating specific emotions with different aspects of products or services. When we say aspects, it is about the characteristics of a product or service. For example, consider the following phrases: "the user experience of a new product," "the response time for a query," "the ease of integration of new software." etc.

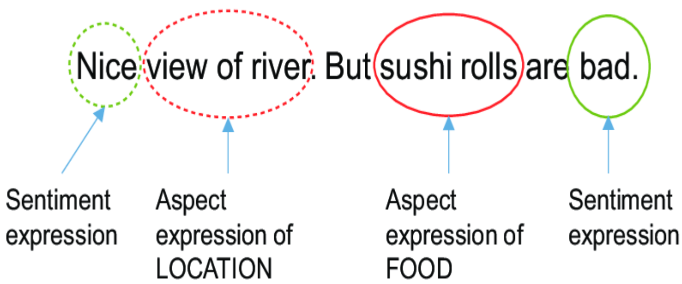

Fig -2: An example of Aspect-based sentiment analysis

Here's an analysis of what ABSA will extricate:

- Sentiments: positive or negative views about a specific aspect or factor
- Aspects: the category, feature, and topic being discussed.

4.4 Prediction of the ratings

One of the most crucial tasks of any recommendation model is prediction of the rating provided. This model obtains the rating prediction by combining the previously mentioned algorithms. Every algorithm predicts a user's rating *u* on item *i* which is independent of the others. We should consider all the contributions of each one in the final prediction in order to amalgamate all the predictions linearly. Suppose the user *u* have already rated a catelogue of items that is similar to the target item *I*. In that case, the content-based algorithm outperforms the other algorithms. Similarly, if the user u has a significant number of neighbours that have rated the target item *i*, then CF is superior to the others. Finally, in the event of a cold start, both algorithms won't be able to perform well.

5. RESULTS

The system is in the form of a web application that includes pre-defined and pre-designed tour packages. The packages consist of the destination, a brief description of the tour, and an itinerary that the user can download and access. The admin is responsible for creating and managing tour packages. Users have to log in and can see and access the packages. The packages also have a facility where the users can rate the packages in the form of stars. They also have to write a review which tells us what they think and feel about the said tour.

Here the stars represent the ratings: 1 and 2 stars mean less favorable, 3 represent slightly optimistic, and 4 and 5 indicate good ratings. Both the ratings and reviews have to be matched to submit the reviews, which means a positive review and a negative rating wouldn't be accepted by the model. The model employs aspect-based sentiment analysis, which stores the ratings and reviews in our SQL database. The 1 and 2 stars convey a negative rating while 3 stars give a neutral rating. 4 and 5 represent a positive rating.

Based on the above ratings and reviews and the hybrid model, unique recommendations are given using existing packages. The ratings and reviews are saved in our database and will be displayed on each package.

Max Rows: 1000

Id	Tour_id	User_Id	Review	Ratings	Sentiment
3	1	3	Good Trip to find pease and relax	5	Possitive
4	1	4	Nice Place	5	Possitive
5	1	5	The staff in tis trip was very carrying and f...	5	Possitive
6	2	1	Pink city was the best part of the trip	5	Possitive
7	2	2	Rajasthan trip was awesome	4	Possitive
8	2	3	Loved it	4	Possitive
9	2	4	Pink city was awesome	4	Possitive
10	2	5	Rajasthani people rocks	4	Possitive
11	3	1	Himachal is full of adventures	5	Possitive
12	3	2	will go again over this place	5	Possitive
13	3	3	Best trek ever	5	Possitive
14	3	4	felt i m at the top of the world	5	Possitive
15	3	5	awesome trip	5	Possitive
16	4	1	jai shiv shambhu	3	Neutral
17	4	2	found peace in Haridwar	5	Possitive
18	4	3	spritiual trip	3	Neutral
19	4	4	Good Trip	4	Possitive
20	4	5	Nice Trip	4	Possitive
21	1002	1	Best Trip Ever	5	Possitive
22	1002	2	Must visit place	5	Possitive
23	1002	3	India is full of places like these	5	Possitive
24	1003	1	leh is heaven	4	Possitive
25	1003	2	Ladhak is beauty	5	Possitive
NULL	NULL	NULL	NULL	NULL	NULL

Fig -3: Ratings and Reviews along with the sentiment stored in the database.

6. CONCLUSION

This research paper introduces a framework for a more efficient hybrid approach. Two algorithms are consolidated into the process for rating prediction.

First, we attributed a confidence measure to each algorithm which is evaluated according to the number of ratings provided for a particular user or individual. Hence, these values are changed dynamically according to the available ratings, allowing solving the cold start problem.

Furthermore, our reports indicate that the suitable integration of CF and CB algorithms will improve the model's performance in the case of rating prediction and coverage precision.

7. FUTURE WORK AND RESEARCH

First, we want to test the efficiency of our model on several other datasets. Then, we would like to extend beyond the existing algorithms to find more efficient methods in this or different domains that overcome the limitations while improving and enhancing the efficiency and accuracy of the current models.

Another way recommendation systems can advance and evolve is through the inclusion of Neural Networks and Deep Learning. Research and studies indicate that this would resolve the recommendation problems. One of its key features is similar to matrix factorization due to its ability to derive latent attributes.

Another feature can be included namely location. Using the geological positions of the users to provide recommendations based on a particular location for example a place near him. With the help of an individual's exact location, we can provide suggestions that are more local and more precise. For tourists places and tastes that are near to their location might be a lot more interesting.

REFERENCES

1. R. J. Mooney and L. Roy, "Content-based book recommending using learning for text categorisation," in DL '00: Proceedings of the fifth ACM conference on Digital libraries. New York, NY, USA: ACM, 2000, pp. 195–204.

2. J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl, "Grouplens: applying collaborative filtering to usenet news," Commun. ACM, vol. 40, no. 3, pp. 77–87, 1997.

3. G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," IEEE

Trans. Knowl. Data Eng., vol. 17, no. 6, pp. 734–749, 2005.

4. G. Huming and L. Weili, "A Hotel Recommendation System Based on Collaborative Filtering and Rankboost Algorithm," 2010 Second International Conference on Multimedia and Information Technology, 2010, pp. 317-320, doi: 10.1109/MMIT.2010.14.

5. Moshfeghi, Yashar & Piwowarski, Benjamin & Jose, Joemon. (2011). Handling data sparsity in collaborative filtering using emotion and semantic-based features. 625-634. 10.1145/2009916.2010001.

6. Jannach, Dietmar & Zanker, Markus & Felfernig, Alexander & Friedrich, Gerhard. (2010). Recommender Systems.

7. Song, W. W., Wu, Q., Forsman, A., & Yu, Z. (2013). A computational model for trust-based collaborative filtering : an empirical study of hotel recommendations. Web Information Systems Engineering - WISE 2013 : 14th International Conference, Nanjing, China, October 13-15, 2013, Proceedings, Part II, 8182, 266–279. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:du-12033

8. Sharma, Jatin & Sharma, Kartikay & Garg, Kaustubh & Sharma, Avinash. (2021). Product Recommendation System a Comprehensive Review. IOP Conference Series: Materials Science and Engineering. 1022. 012021. 10.1088/1757-899X/1022/1/012021.

9. Khairullah Khan, Baharum Baharudin, Aurnagzeb Khan, and Ashraf Ullah. "Mining opinion components from unstructured reviews: A review" Journal of King Saud University - Computer and Information Sciences, vol. 26, no. 3, 2014. doi:10.1016/j.jksuci.2014.03.009

10. M. Jamali and M. Ester, "Trustwalker: a random walk model for combining trust-based and item-based recommendation," in Proceedings of the SIGKDD conference. New York, NY, USA: ACM, 2009, pp. 397–406.

11. Gang Li, Rob Law, Huy Quan Vu, Jia Rong, Xinyuan (Roy) Zhao, "Identifying emerging hotel preferences using Emerging Pattern Mining technique, Tourism Management", Volume 46,2015,Pages 311-321,ISSN 0261-5177,https://doi.org/10.1016/j.tourman.2014.06.015. (https://www.sciencedirect.com/science/article/pii/S0261517714001174).

12. Thammaboosadee, Sotarat & Muangon, Ananchai & Haruechaiyasak, Choochart. (2014). A Lexiconizing Framework of Feature-based Opinion Mining in Tourism Industry. 2014 4th International Conference on Digital Information and Communication Technology and Its Applications, DICTAP 2014. 10.1109/DICTAP.2014.6821677.

13. Lin Zhang, Kun Hua, Honggang Wang, Guanqun Qian, Li Zhang,"Sentiment Analysis on Reviews of Mobile Users",Procedia Computer Science,Volume 34,2014,Pages 458-465,ISSN 1877-0509, https://doi.org/10.1016/j.procs.2014.07.013. (https://www.sciencedirect.com/science/article/pii/S1877050914008680).

14. Duan, Wenjing & Cao, Ray & Yu, Yang & Levy, Stuart. (2013). Mining Online User-Generated Content: Using Sentiment Analysis Technique to Study Hotel Service Quality. Proceedings of the Annual Hawaii International Conference on System Sciences. 3119-3128. 10.1109/HICSS.2013.400.

15. Dalal, Mita & Zaveri, Mukesh. (2014). Opinion Mining from Online User Reviews Using Fuzzy Linguistic Hedges. Applied Computational Intelligence and Soft Computing. 2014. 1-9. 10.1155/2014/735942.

16. Wu, Yao & Ester, Martin. (2015). FLAME: A Probabilistic Model Combining Aspect Based Opinion Mining and Collaborative Filtering. WSDM 2015 - Proceedings of the 8th ACM International Conference on Web Search and Data Mining. 199-208. 10.1145/2684822.2685291.

17. Kavinkumar, V. & Rachamalla, Rahul & Balasubramanian, Rohit & Sridhar, M. & Sridharan, K. & Venkataraman, D.. (2015). A hybrid approach for recommendation system with added feedback component. 745-752. 10.1109/ICACCI.2015.7275700.

18. Leung, Cane & Chan, Stephen. (2008). Sentiment Analysis of Product Reviews

19. Chikhaoui ,Belkacem ,Mauricio Chiazzaro and Shengrui Wang. "An Improved Recommender System by Combining Predictions", 2011 IEEE Workshops of International Conference on Advanced Information Networking and Applications 2011