IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

IJARSCT
Impact Factor: 6.252

Volume 2, Issue 3, May 2022

An Accommodation Recommendation System for Immigrants using Exploratory Data Analysis on Geolocation Data

Jitechana[1], Lubna Shaikh[2], Shefali Bhattacharjee[3], Vaishali Yenolge[4], Dr. N. P. Kulkarni[5]

Students, Department of Information Technology[1,2,3,4]

Guide, Department of Information Technology[1,2,3,4]

Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Abstract: *In recent years, there has been a noticeable surge in immigration. When these individuals arrive in the target nation, the majority of them are students who require long-term lodging. However, because he is new to the area and does not know many relevant locations, this creates a difficulty. Because of the increasing expansion of network information, individuals may find it challenging to access what they need in the enormous and big data environment using the online approach. People may locate what they want with the help of the recommendation system, which suggests prospective products of interest to them. It frequently makes advantage of existing associations between users and/or things to anticipate people's liking for certain items. The recommendation system is now attracting a lot of interest from the social network engineering and academic research communities. Therefore, this research article defines an effective methodology for Accommodation Recommendation through the use of K Nearest Neighbor Clustering along with Artificial Neural Networks and Decision Making. The experimental evaluation has been performed which has proved the superiority of the presented technique.*

Keywords: K Nearest Neighbors, Artificial Neural Networks, Decision Making, Recommender Systems

I. Introduction

There has been a lot of migration that is being happening across the world which is mainly being done by the students for the purpose of achieving an improved education in another country. This has been one of the oldest practices where the students travel large distances to facilitate expert study or to take advantage of the improved opportunities offered by the host country. This is particularly true for a large number of students that travel to India and also travel from India to various different countries for improved learning experiences across the globe.

The act of travelling far away to a foreign country can be a highly complex and complicated task that can be undertaken by an individual, let alone a student. This task is increasingly more difficult because of the fact that majority of the students travel alone to completely new place that they have never travelled to before. These students usually lack the expertise in the foreign language and do not have any friends or acquaintances for providing them assistance for any task or any problem they might face in that country. The lack of assistance makes it a daunting task to perform regular day to day activities that can be quite trivial in nature.

One of the most problematic tasks in a foreign country is to find a place for an accommodation. This is one of the most important tasks that require local knowledge along with knowing the working language and making connections with the home owners in the area of interest. This type of information is one of the most difficult to collect without being a native or staying in the area for a few days. This is highly problematic for the foreign students to hunt for the accommodation easily and as per their preferences.

The increased immigration puts a lot of individuals at risk of going to an unknown location and facing a lot of problems with accommodation and other problems. This is problematic as a lot of individuals are migrating for the purpose of education and relocating to distant countries for extended periods of time. This time is crucial as the students will have to stay there therefore need accommodation in the vicinity of the University. A large number of individuals and students need accommodation in the foreign country where they have never been in their lives and must be getting to a foreign country in

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 3, May 2022
Impact Factor: 6.252
the first place for the first time. These students might a problem understanding the language or the accent of the people staying in that area.

This puts a lot of hurdles for these individuals that require an accommodation for a long period of time, failing which they can be facing a very dire situation which can also return to be homeless. The realization of an accommodation is necessary which leads to a realization of a recommendation system which can help these individuals find an accommodation surely. Thus, there is a need for an effective approach that can provide a suitable alternative for an accommodation recommendation as per the student's preferences. This type of recommender systems are quite rare and have been shown to achieve very low accuracies.

This research article defines an efficient approach for the purpose of achieving accommodation recommendation through the implementation of the machine learning methodologies. This accommodation recommendation approach utilizes Artificial Neural Networks for performing the recommendation. The Artificial Neural Networks allow for an effective realization of the recommendation of the accommodation based on the requirements of the user. This is due to the fact that the machine learning approach is one of the most effective methodologies for the realization of such complex tasks that can be accomplished. The Decision Making approach is also an integral part of the solution as it allows for the realization of effective classification of the output from the Artificial Neural Networks.

This research study devotes section 3 to an overview of previous work as a literature survey; section 4 elaborates on the suggested methodology; section 5 analyzes the system's performance; and lastly, section 6 finishes the article with hints for further refinement.

II. Literature Survey

GFP-LORE is a novel hybrid recommendation algorithm framework presented by R. Yue-Qiang [1]. They demonstrate that this technique effectively improves recommendation accuracy by combining social influence, popularity influence, geographic influence, and sequential influence into a unified framework. To begin, they show that user social correlations and POI popularity are sensitive to power law distribution and model them appropriately, using a social component and a popularity factor to recommend a new POI. Then, using the Kernel Density Estimation approach, they calculated the probability of the user's arrival to the new location by assessing the user's individual check-in distribution through the user's checkin history, exploring the customised geographic information in the user's check-in behaviour, and assessing the user's individual check-in distribution through the user's checkin history (KDE). After that, the presented system extracts the sequence pattern from all users' check-in data in the form of dynamic L2TG, which can reflect overall transfer sequence pattern, and uses the Additive Markov Chain (AMC) to derive the probability of the user accessing the new POI, resulting in a recommendation to the user depend on the user's history sequence pattern.

N-PRA, a novel POI recommendation mechanism, was introduced by K. Gao et al. To develop a more exact forecast about users' preferences, several context information from social media networks and IoT have been retrieved and taken into account [2]. The introduced solution outperforms state-of-the-art methods, according to the results. With the rapid growth of current sensor networks and wireless internet, this technology might be used in a variety of scenarios.

S. Ahmad et al. design a unique application that recommends the best travel route depending on the limitations of the user [3]. The maximum time, distance, and popularity of a given location are examples of user limitations. The data is gathered from Wi-Fi routers located at several tourist attractions on Jeju Island, South Korea. They retrieved the visitor's travel routes and patterns and used the Markov Chain model to study the data's stochastic behavior. The authors created multiple transition matrices as part of the Markov Chain model to indicate probability, distance, and time transitions between distinct places. Long-term steady-state forecasts, together with user preferences and constraints, aided in the discovery of optimum routes that met user constraints.

J.-H. Chang et al. describe a method for creating a Twitter-based recommendation system by combining various social media. First, a model is constructed to predict user preferences by augmenting matrix factorization based on user preferences and personal data, using auxiliary data from Yelp. On the other hand, an analytical user posting behavior algorithm is being created for identifying users' posting behavior vectors based on prior tweets and Yelp reviews [4]. The experiments' findings show that, when compared to a Twitter-based recommendation system that does not take into account diverse social media, the recommended technique can improve RECALL accuracy by 30%. Furthermore, it can enhance mean reciprocal rank accuracy by up to 80% and raise precision by up to 80%.

M. Ludewig et al. offered an integrated hybrid strategy for ranking hotels depend on restricted information such as the user's recent browsing history, which is a challenge that arises frequently in search and recommendation situations. They constructed a variety of predictor variables for the provided job, which could be framed as a pair-wise ranking issue, and employed GBDTs as a learning approach, which eventually resulted to competitive performance in the 2019 ACM RecSys Challenge for hotel recommendation [5]. They considered the issue setting to be particularly relevant for the research community in general, such as in the context of session-depend recommendation and search personalization. The challenge's design as a re-ranking assignment, on the other hand, encouraged the inclusion of specific features that were useful for the task at hand.

Q.-H. Le et al. developed DNN DST, a new approach that incorporates a DNN framework with the Dempster-Shafer theory of probability for multi-criteria CF issues. The suggested technique began with a DNN framework that utilized an SVD methodology as its first layer to predict multi-criteria ratings, and then utilized an evidential reasoning approach to describe those ratings as bits of evidence using mass functions, which were then aggregated to forecast the overall rating [6]. Modeling criteria ratings expected by the DNN model as pieces of evidence allows them to investigate the inherent uncertainty connected with these forecasts to be incorporated in multi-criteria rating aggregation.

S. G. K. Patro et al., suggested HAR-KNN model which is depend on user behavior data such as the number of views, clicks, and purchased items, as well as purchase histories. The suggested technique creates a user behavior matrix and pre-processes the frequency of product purchases in order to anticipate users depend on a vector of attributes for each neighbor and product. After generating the user behavior matrix, the study makes good use of it to map out other users with similar likes and neighbors. It stands out by calculating the Euclidean distance and properly improvising all of the framework's algorithms to attain a higher similarity level. Furthermore, the HAR-KNN model is successfully compared with IBCF, TRIBCF, TCIBCF, CDIBCF, TWIBCF, CB, and CF in terms of MAE, RMSE, MSE, Recall, Precision, and F1-measure values on the Amazon dataset. The experimental results show that performance measurements enhance predicted accuracy while reducing error when compared to standard techniques.

Q. Xu et al. introduce SSLR, a hybrid recommendation model that takes into account both the spatial and sequence components of a location's geographical influence. According to the author's analysis of two real-world data sets, users are influenced by site spatial and sequence aspects. To describe geographic characteristics and create a unique distribution for each user, the authors adopt a kernel density estimation approach. In addition, a random walk model on a graph integrates user choice, social influence, and geographical influence sequence characteristics [8]. Finally, by merging spatial and sequence impacts with varied weights, the hybrid model is generated.

Himanshu Sharma et al. conducts a two-stage analysis using DEA and ANN to assess the performance of hotels in the Delhi NCR. The first step assessed the efficiency of the 78 hotels that were considered. This stage was carried out utilizing the DEA method. The hotels are divided into two categories in this research: 2-3 star and 4–5-star hotels [9]. The advantage of utilizing DEA over other decision-making techniques is that it can handle several input/output variables at the same time, and these variables can be financial or non-financial. The number of rooms and the ratings relating to various elements of a hotel are the input variables studied here. RevPAR and CS are also considered output variables.

Bahramian, Zahra et al. presents a hybrid interactive context-aware recommender system that has been applied to the tourist area. The technique is unique in that it combines CBR (a knowledge-based recommender system) and ANN (a content-based recommender system) to solve the cold start problem for a new user with few past ratings [10]. The suggested technique may recommend a tour to a user who has limited information about his preferences and takes into consideration the fact that the user's choices may change during the recommendation process. Furthermore, it selects POIs and the path between them on the street network depend only on the preferences and input of the specific user.

III. Proposed Methodology

The presented system for Accommodation Recommendation has been depicted in the system overview given in figure 1 above. The sequential steps to achieve the presented approach have been elaborated in the section below.

Step 1: Data collection and Data feeding – The proposed technique requires the input of the user as well as the accommodation dataset on properties listed on StayZilla startup for providing accommodation services. For this purpose a dataset containing reviews has been extracted from the URL-https://www.kaggle.com/datasets/PromptCloudHQ/properties-on-stayzilla

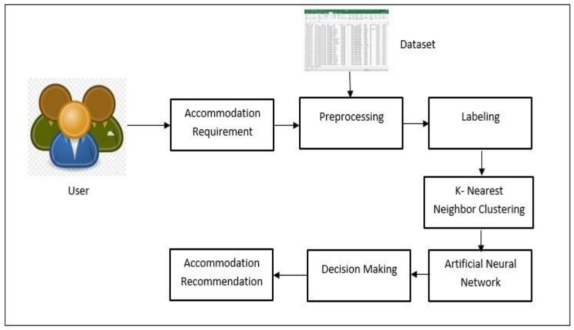

Figure 1: Sentiment Analysis System Overview

This dataset consists of a number of different reviews for smartphones and the requisite information about the reviews. The attributes include the additional_info, amenities, check_in_date, check_out_date, city, country, crawl_date, description, highlight_value, hotel_star_rating, image_count, image_urls, internet, landmark, latitude, longitude, occupancy, pageurl, property_address, property_id, property_name, property_type, qts, query_time_stamp, room_price, room_types, search_term, service_value, similar_hotel, sitename, things_to_do, and things_to_note.

The dataset has been modified extensively for our implementation leading to the creation of a synthetic dataset that is being deployed for the processing in the subsequent steps of this methodology. The reviews in the dataset need to be preprocessed before providing it to the system to reduce the incidences of any error or redundancy that may impact the performance negatively. The preprocessing approach has been elaborated in the next step of the approach.

Step 2: Preprocessing – The preprocessing approach is the initial logical step of the approach that is designed to facilitate the processing of the reviews before it can be provided to the system. The extracted dataset achieved in the previous step is provided as an input in this step of the procedure. The dataset is in the workbook format, therefore the JXL library is being used to interface this file with the java code. The dataset is then converted into a double dimension list that can be easily processed by the system.

The process of preprocessing is effective in realization of the conditioning of the input reviews that are provided. This is highly crucial as the performance of the execution depends on the effectively of this step of the procedure. The presence of any unnecessary data can be detrimental to the system as it could result in an error reducing the accuracy. The redundant data can also take longer to process which can be problematic to achieve efficiency. The steps for preprocessing are defined below.

Step 3: K-Nearest Neighbors Clustering – In this step of the technique, the label list developed in the previous step is being used as an input. The k-nearest neighbor algorithm is implemented using the four main components given below.

Distance Evaluation – The distance for the input supplied in the format of a labelled list is obtained in the first step of clustering. The Euclidean distance of every entry in the attribute list or labelled list is used to calculate this distance appropriately. As a result, the distances are inevitably added to the end of the corresponding attributes, together with the other rows. The row distance RD of each particular row is determined using the average of the gathered data and distance. The average distance A_{RD} is calculated using equation 1 below, that includes the row distance R_D into consideration.

$$ED = \sqrt{\sum(ATi - ATj)^2} \quad \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (1)$$

Where, ED=Euclidian Distance

A_{Ti}=Attribute at index i

A_{Tj}= Attribute at index j

Centroid Estimation – The centroids for clustering should be obtained after the row distances have been computed. The use of a row distance list that is then bubble sorted into increasing order of row distances allows these centroids to be implemented successfully. Then, for merging the whole list generated by gathering the indices of the biggest and smallest values from the sorted list, a value of K is allocated. These values are essentially subtracted to obtain the value of k, and the resultant number is halved by two. This value is then used to compute the inner and outer cluster boundary criteria in the next step.

Cluster Formation – The previous stage's boundary values are efficiently employed, and the k value acquired before has been used to split the data points. The attribute values acquired from the label list created in the previous step of the clustering process are used to divide clusters into two groups. The inner clusters created in this stage are then used to determine the most relevant groups. The Artificial Neural Network is then trained using this clustered data. The technique for cluster creation is illustrated in algorithm 1 below.

ALGORITHM 1: KNN Classified Cluster Formation

//Input : Sorted Distance List $SD_{L,}$
//Output:Cluster List K_{CL}
Start
$I_L = \emptyset$ [Inner Layer] $O_L = \emptyset$ [Outer Layer], $K_{CL}=\emptyset$
MIN= 0 , MAX=SDL_{SIZE-1}
K= (MAX-MIN) /2
K=MIN+K
    ***for** i=0 to Size of SD_L*
        $R = SD_{L\,[i]}$
        if**(i<=K), **then
            $I_{L=}\ I_{L+R}$
        else
            $O_{L\,=}\ O_{L\,+R}$
    end for
        $K_{CL[0]} = I_L$
        $K_{CL[1]} = O_L$
    return K_{CL}
Stop

Step 4: Artificial Neural Network – The Obtained cluster list is fed as the input list to the ANN model. Where each of the transaction rows in the frequent item list is subject to evaluate the hidden and output layers for the attributes User ID and purchase value. This is done by considering the same as the target values for the random weights W1,W2,W3,W4,W5,W6,W7,W8,B1,B2. Here B1 and B2 are the bias values which are used to stabilize the neurons. Then by using the Equation 2 and 3 for Hidden layer and Activation function Output layers are estimated. The obtained output layers are aggregated with the target values to achieve the new probability list as ANN probability list.

$$X = (AT1 * W1) + (AT2 * W2) + (AT3 * W3) + (AT4 * W4) + B1 \qquad (2)$$

$$H_{LV} = \frac{1}{(1+\exp(-X))} \qquad (3)$$

Where AT1 is the amenities, AT2 is the property type, AT3 is the room type, and AT4 is the price. Then the sigmoid function is given by Equation 3 of the neural network. H_{LV} – Indicates the hidden layer value.

Step 5: Decision Making – The obtained ANN prediction list from the prior step is used as the input for the Decision Classification process to determine the accommodation recommendation. In this process the if-then rules are being used to completely classify the accurate and precise accommodation recommendations from the probability list. The effective realization of the classification allows for the proper accommodation recommendation for the user based on their input which is displayed to the user in the interactive User Interface.

IV. Results And Discussions

The proposed methodology for the intention of attaining accommodation recommendation has indeed been demonstrated through the use of the java programming language. The NetBeans IDE has been utilized to complete the systems integration. The development machine has an Intel Core i5 CPU and 8GB of RAM. It also has 1TB of storage. The MySQL Database server is in charge of database tasks and obligations.

The accuracy of the accommodation recommendation needs to be evaluated to understand the execution accuracy of the methodology. The central aspect of the presented approach that actually performs the recommendation is the Artificial Neural Networks. This neural network needs to be deployed extremely accurately to achieve the accommodation recommendation. Therefore, the Artificial Neural Networks approach needs to be evaluated for its performance in the section given below.

4.1 Performance Evaluation through Root Mean Square Approach

A series of investigations were conducted to establish the error achieved by the suggested approach, which uses Artificial Neural networks to provide accommodation recommendation. The errors of the strategy for correctly recommending the accommodation according to the provided details is being utilized to define the performance criterion.

The Root Mean Square Error (RMSE) is used to calculate the error attained by the approach. The occurrence of error in the suggested strategy for accommodation recommendation utilizing user input and ANN indicates the operational effectiveness of the proposed methodology. The RMSE approach makes evaluating errors within two continuously associated parameters considerably straightforward. In this approach, the expected accommodation recommendations and the obtained accommodation recommendations are ascertained. These values are recorded, and the error is calculated using equation 1.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(x_{1,i} - x_{2,i}\right)^2}{n}}$$

Where, $\sum$ - Summation

$(x_1 - x_2)^2$ - Differences Squared for the summation in between the expected accommodation recommendations and the obtained accommodation recommendations

n - Number of Trails

These two properties were measured for 10 distinct trials machine with varying input parameters. The results of these evaluations is summarized in table 1 below.

User No.	Expected Accomodation Recommendations	Obtained Accomodation Recommendations	MSE
1	10	10	0
2	10	9	1
3	10	10	0
4	10	9	1
5	10	10	0
6	10	10	0
7	10	9	1
8	10	10	0
9	10	8	4
10	10	9	1

Table 1: Mean square Error measurement

The results of the technique's empirical evaluation have made it simpler to visually depict the error rate, as illustrated in Figure 3. Depending on individual input parameters, the graph indicates the system's minimal amount of error in accommodation recommendation. This is owing to the Artificial Neural Network's extremely accurate implementation, which dramatically enhances recommendation performance. The Decision Making technique maximizes the precision, as evidenced by MSE and RMSE values of 0.8 and 0.89, respectively. This evaluation reveals how the accommodation recommendation technique was implemented precisely and accurately.

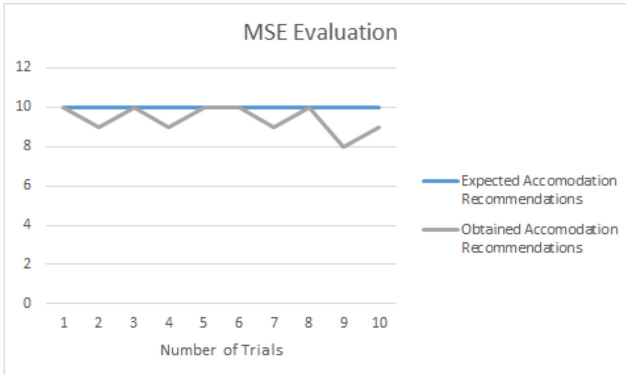

Figure 2: Comparison of MSE in between expected accommodation recommendations V/s obtained accommodation recommendations

V. Conclusion And Future Scope

The presented methodology for accommodation recommendation based on user's input parameters through the use of Artificial Neural Networks has been outlined in this research paper. The presented system takes the user requirement as well as dataset as input, the dataset is in the form of an excel file which is first preprocessed. The preprocessing involves the removal of redundant and incomplete data before giving it as an input to the system. The preprocessed dataset and the user input parameters are subjected to clustering through the use of K nearest Neighbor clustering mechanism. The clustering is performed using the user input parameters and the achieved clusters are provided to the next step for neuron generation. The Artificial Neural Network is deployed which effectively initiates the hidden layer and output layer measurement through the neurons on an activation function to attain the probability list. This probability list is then provided to the next step for the purpose of classification. The Decision making approach utilizes the if-then rules to achieve the classification of the probability list. The output thus achieved is displayed to the user through the interactive user interface. The approach has been quantified through the evaluation of the error occurred in the recommendation. The results have displayed a decent level of RMSE which indicates an error free deployment of the accommodation recommendation approach.

The objective of future research should be to turn this recommendation technique into an API for universal use and integration. This study should be expanded to enhance suggestions based on a bigger and more diverse dataset.

References

[1]. R. Yue-Qiang, W. Ze, S. Xiao-Na and S. Shi-Min, "A Multi-Element Hybrid Location Recommendation Algorithm for Location Based Social Networks," in IEEE Access, vol. 7, pp. 100416-100427, 2019, DOI: 10.1109/ACCESS.2019.2929313.

[2]. K. Gao et al., "Exploiting Location-Based Context for POI Recommendation When Traveling to a New Region," in IEEE Access, vol. 8, pp. 52404-52412, 2020, DOI: 10.1109/ACCESS.2020.2980982.

[3]. S. Ahmad, I. Ullah, F. Mehmood, M. Fayaz and D. Kim, "A Stochastic Approach Towards Travel Route Optimization and Recommendation Based on Users Constraints Using Markov Chain," in IEEE Access, vol. 7, pp. 90760-90776, 2019, DOI: 10.1109/ACCESS.2019.2926675.

[4]. J. -H. Chang, C. -E. Tsai and J. -H. Chiang, "Using Heterogeneous Social Media as Auxiliary Information to Improve Hotel Recommendation Performance," in IEEE Access, vol. 6, pp. 42647-42660, 2018, DOI: 10.1109/ACCESS.2018.2855690.

[5]. M. Ludewig and D. Jannach, "Learning to Rank Hotels for Search and Recommendation from Session-based Interaction Logs and Meta Data," in RecSys Challenge '19: Proceedings of the Workshop on ACM Recommender Systems Challenge, September 2019 Article No.: 5Pages 1–5https://doi.org/10.1145/3359555.3359561.

[6]. Q. -H. Le, T. N. Mau, R. Tansuchat and V. -N. Huynh, "A Multi-Criteria Collaborative Filtering Approach Using Deep Learning and Dempster-Shafer Theory for Hotel Recommendations," in IEEE Access, vol. 10, pp. 37281-37293, 2022, DOI: 10.1109/ACCESS.2022.3165310.

[7]. S. G. K. Patro et al., "A Hybrid Action-Related K-Nearest Neighbour (HAR-KNN) Approach for Recommendation Systems," in IEEE Access, vol. 8, pp. 90978-90991, 2020, DOI: 10.1109/ACCESS.2020.2994056.

[8]. Q. Xu, J. Wang, and B. Xiao, "Personalized location recommendation for location-based social networks," 2017 IEEE/CIC International Conference on Communications in China (ICCC), 2017, pp. 1-6, DOI: 10.1109/ICCChina.2017.8330459.

[9]. Sharma, Himanshu & Suri, Gunmala & Savara, Vandana. (2022). An Approach Combining DEA and ANN for Hotel Performance Evaluation. 10.4018/978-1-6684-2408-7.ch070.

[10]. Bahramian, Zahra & Abbaspour, Rahim & Claramunt, Christophe. (2017). A Cold Start Context-Aware Recommender System for Tour Planning Using Artificial Neural Network and Case Based Reasoning. Mobile Information Systems. 2017. 1-18. 10.1155/2017/9364903.