A C4.5 Decision Tree Algorithm with MRMR Features Selection Based Recommendation System for Tourists

B.Santhosh Kumar[1], M.Raghavendra Reddy[2]

[1]Associate Professor, Dept. Of CSE, G.Pulla Reddy Engineering College (Autonomous), Kurnool, AP, India
[2]Assistant Professor, Dept. Of CSE, G.Pulla Reddy Engineering College (Autonomous), Kurnool, AP, India

ABSTRACT

One of the most difficult tasks for tourists when preparing travel, both before and during travel, is selecting a tourist destination from the information that is accessible on the Internet and from other outlets. Previous Travel Recommendation Systems (TRSs) have tried to resolve this issue. We are applying the C4.5 decision tree algorithm in this paper with the collection of MRMR features to propose tourist travel areas by using datasets from previous tourist encounters. Both existing algorithms, such as interactive or content filtering algorithms, use data from current users' previous history to suggest new locations to them. If this current user has no data from previous encounters, these algorithms won't work. To solve the above problem, we use C4.5 decision tree algorithms that take previous user interactions and then generate a model and if new users enter their criteria, the decision tree will predict the best position based on its feedback. Decision Tree does not require previous history data from new users. The framework is built using a two-step process of feature selection to minimize the number of inputs to the system and Decision Tree C4.5 makes recommendations.

Introduction

Recommender Systems (RSs) are digital instruments and techniques that offer recommendations for a user's use of products. The guidelines refer to different processes of decision-making, such as where to go, what things to purchase, what music to listen to or what internet content to read. The general term used to denote what the device suggests to users is "item". A recommendation framework typically focuses on a particular category of item (e.g., CDs, or news) and, accordingly, its architecture, its graphical user interface, and the central recommendation process used to produce the recommendations are all tailored to include useful and efficient advice for that particular type of item. The immense growth of the web and its user base has become the source of vast amounts of online content. For consumers, this information can be helpful in recommending products or services according to their expectations. By gathering user data such as tastes, desires, and places, the recommendation system plays the role of producing recommendations. With the advent of collaborative screening, the study on recommendation systems gained prominence, leading to the introduction of recommendation systems in diverse fields such as e-commerce, advertisement and tourism. A dynamic challenge for recommendation systems is to produce recommendations according to user tastes. Semantic web tools help to quickly recommend schemes to tackle certain tasks. To make forecasts and to propose an object to a customer, the Recommender method uses data from multiple sources. In the produced recommendations, factors such as novelty, stability, and accuracy are balanced. In the suggestion process, filtering mechanisms play an important role. Collaborative filtering, content-based filtering, knowledge-based filtering, and social filtering are the most widely used filtering strategies. Many studies have already led to the growth of numerous recommendation systems, such as video, music, books, e-commerce e-learning, online search, and tourism. The Internet is powerful today.

Sources of tourist experience for research on goods and facilities. Owing to the enormous amount of heterogeneous information accessible on the internet, searching for destinations known as travel plans would overtake travelers. The role of travel planning is dynamic and competitive, requiring many variables, such as the quality of the decision Attractions, itineraries, restaurants, number of travelers, outdoor activities, weather, etc. Tourism has recently made considerable gains from ICT and in particular, from Internet ICT technologies. Visitors and tourism suppliers are scanned, selected, evaluated and determined more effectively than ever for the production of Decision Support Instruments, also known as Recommendation Systems (RS).

Tourist decision-making destinations are to improve the process, one of the key challenges in TRS offering personalized tourist feedback. Uh, method. A deep one is required to do this. Understanding visitor behavior and creating new models for discovering the Approach Mechanism for your data. Uncertainties are also present in the information. The scanning stage of the tourist decision process must be eliminated. Reduction of specifications for more devices. The complexity of the model may be reduced. Both can maximize system satisfaction in return for the consistency of advice and use speed.

Background & Related Work

A Recommendation System (RS), a subset of Decision Support Systems (DSS), is a tool that can recommend an object based on aggregated user preference knowledge. Based on their interests and needs, it assists consumers by

offering useful knowledge to aid them in their decision-making processes. RS plays a major role and is prominent on many popular websites for e-commerce, such as Amazon, Netflix, Pandora, etc. The e-commerce RSs recommend user products that include news, posts, persons, URLs, and so on.

Tourism is a recreational pursuit requiring difficult decision-making procedures, such as the selection of destinations, sites, events and facilities. Therefore, TRS attracts the interest of many experts from the fields of both academia and business. In and on several types of platforms, numerous TRS have been built or deployed (e.g. desktop, browser, mobile). For the purposes of assessing user interest, selecting Points of Interest (POIs), defining resources or paths, rating them in order, or as a holistic trip schedule, TRSs propose results to a user.

Much of the existing TRSs are targeted at helping individual visitors, but there are several programmes that also support travel agencies. They share common concepts but vary in technology selection, personalization enhancement philosophies, data inputs, type of interaction, and techniques of suggestion. The general structure of the new TRSs is seen in Fig 1. Knowledge is combined and processed in the registry from multiple sources (e.g. sensors, GPS coordinates, surveys, ratings, etc) (e.g. database schema, ontology). The recommendation engine can consist of many subsystems, such as a subsystem for optimization, a mathematical subsystem, an intelligent subsystem, etc. This is to recommend, rate, or forecast items based on consumer criteria, tastes, or other hard and soft constraints (i.e. destination, attractions, events, and services) (e.g. user demographic information, number of travel days, travel budgets, travel type, etc.). The TRS will usually take any inputs (implicit, explicit, or both from the tourist before or during the trip to construct a user profile and measure the recommended result that is then sent back to the tourist. The effects of the method can be visualized in many ways by travelers, such as by destination icons on the map interface with a path from point-to-point, timetable, and itinerary. The consequence of using spatial web services, such as the Google Maps API tool, is presented by most TRSs. Lately, by taking into account user background details (e.g. location, weather), certain TRSs are able to tailor the findings to the user. Some TRSs have a feature to allow the user to change the result produced and adjust the outcomes based on user input or user ranking.
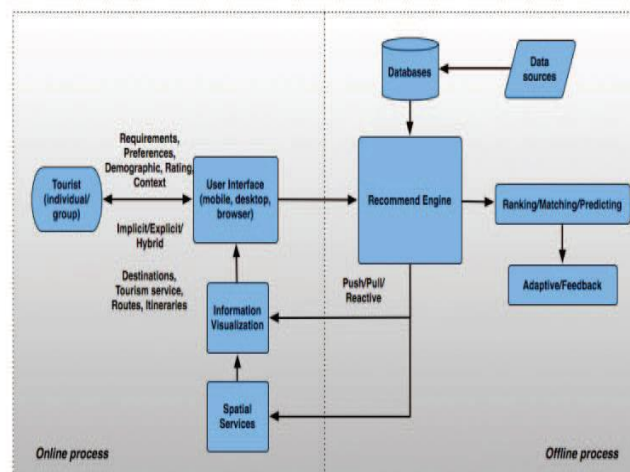
Fig 1.General framework of the travel recommendation systems

Results and Discussion

In this paper author is implementing C4.5 decision tree algorithm with MRMR features selection to recommend travel areas to tourist by using dataset from past tourist experiences. All existing algorithms such as collaborative or content filtering algorithms uses current user past experience data to recommend him new locations. These algorithms will not work if this current user has no past experiences data. To overcome from above problem author is asking to use C4.5 decision tree algorithms which take experiences of previous users and then build a model and if new user enter his requirements then decision tree will predict best location based on his given input. Decision tree don't need new users past experience data.

To implement decision tree model we need to have dataset and this dataset sometime will have empty or garbage values and this values will put bad effect on decision tree model so we can remove such empty or garbage values by applying pre-process techniques.

Sometime to predict or build model no need to use all columns (attributes) values from dataset and this unnecessary attributes can be remove by apply features selection algorithms and here we are using MRMR features selection algorithms to remove unnecessary attributes to reduce execution time of building model and to increase system accuracy. Below are the dataset columns or attributes taken from previous users to build model.

This data set is populated by crawling TripAdvisor.com. Reviews on destinations in 10 categories mentioned across East Europe are considered. Each traveller rating is mapped as Excellent (4), Very Good (3), Average (2), Poor (1), and Terrible (0) and average rating is used against each category per user.

Dataset columns and values
userid,art_galleries,dance_clubs,juice_bars,restaurants,museums,resorts,parks_picnic_spots,beaches,theaters,religious_institutions,location

Above are the column names and below are the column values

User1,0.93,1.8,2.29,0.62,0.8,2.42,3.19,2.79,1.82,2.42,Amsterdam_Heining_2
User2,1.02,2.2,2.66,0.64,1.42,3.18,3.21,2.63,1.86,2.32,Amsterdam_Jachthaven_ijbur
User3,1.22,0.8,0.54,0.53,0.24,1.54,3.18,2.8,1.31,2.5,Amsterdam_Bert_Haanstra_Kad
User4,0.45,1.8,0.29,0.57,0.46,1.52,3.18,2.96,1.57,2.86,Amsterdam_Ruigoord_Ker

In above values first column is USER_ID and second column is ART_GALLERIES and third is DANCE CLUB etc and for each column user had given rating from 4 to 0 and 4 means excellent service. Now using above values we can build C4.5 decision tree and prediction will be done using below test values

'User122',0.93,1.8,2.29,0.62,0.8,2.42,3.19,2.79,1.82,2.42,?
'User222',1.02,2.2,2.66,0.64,1.42,3.18,3.21,2.63,1.86,2.32,?
'User3222',1.22,0.8,0.54,0.53,0.24,1.54,3.18,2.8,1.31,2.5,?
'User4222',0.45,1.8,0.29,0.57,0.46,1.52,3.18,2.96,1.57,2.86,?
'User522',0.51,1.2,1.18,0.57,1.54,2.02,3.18,2.78,1.18,2.54,?

In above test values new user has given values to look for location which has above service rating but new user don't know which location provides such services so he will put question mark and when we upload above test values to decision tree then it will take decision and predict best location and inform to user.

In above screen click on 'Upload Tourist Dataset' button and upload dataset file

After file upload will get below screen with all dataset details. In above screen after uploading test data we can see all values are there in test data but it not has location name and base on test values application predicted or recommend location name. Now click on Features Selection Graph button to get below graph
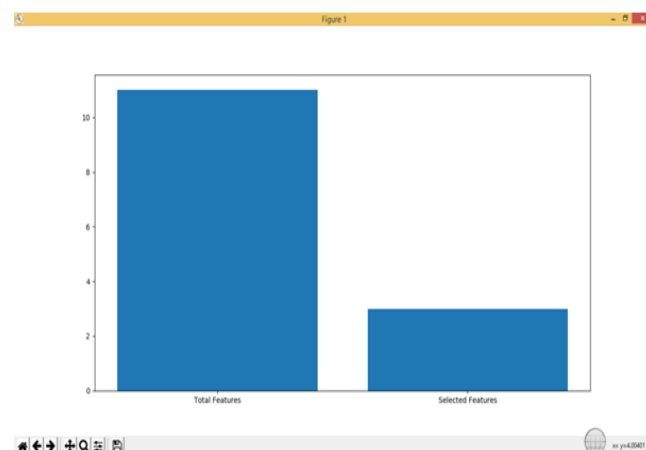
In above graph x-axis contains total features and MRMR selected features and y-axis represents count of features and in above graph we can see after applying MRMR technique features size reduces to 3.

Conclusion

In this article, a tourist suggestion method based on the decision tree was introduced in an effort to address the current problem of the TRS destination. Using specific tourism domain information, the data collection has been decomposed into two sub-data sets. This was achieved to improve the consistency rate of classification and to decrease the difficulty of the decision tree. For destination collection, the optimal decision trees from NMIFS were designed with the highest accuracy rate and simplicity (i.e. less number of leaves and tree size). Decision guidelines have been derived from decision trees. For both data sets, it can be shown that NMIFS is the optimal approach since it uses fewer features than MRMR. Finally, the experimental findings validate the proposed TRS as applicable. The suggested TRS meets the criteria of tourists who plan to visit or visit the city of Chiang Mai during their visit. To improve the classification accuracy rate for the data sets, various types of classifiers may be considered for future work. In addition, they can build and develop a front-end web framework and an immersive and scalable user interface.

References

[1] J. Camhi, "Former Cisco CEO john chambers predicts 500 billion connected devices by 2025," Business Insider, 2015.

[2] K. Wu, P. Lu, and Z. Zhu, "Distributed online scheduling and routing of multicast-oriented tasks for profit-driven cloud computing," IEEE Communications Letters, vol. 20, no. 4, pp. 684–687, Apr. 2016.
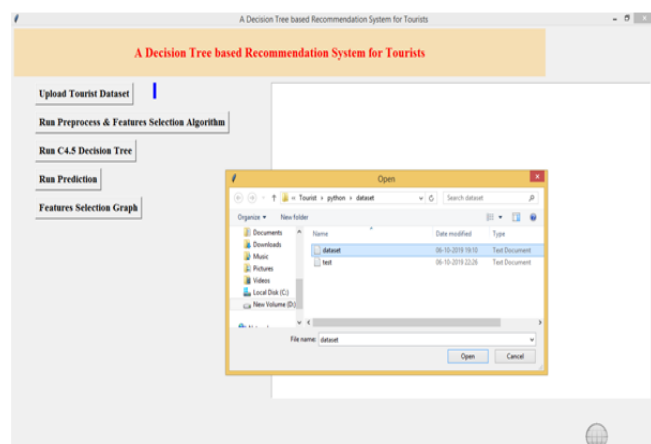
[3]   J. Yao, P. Lu, L. Gong, and Z. Zhu, "On fast and coordinated data backup in geo-distributed optical inter-datacenter networks," Journal of Lightwave Technology, vol. 33, no. 14, pp. 3005–3015, July 2015.

[4]   Z. Zhu, P. Lu, J. J. P. C. Rodrigues, and Y. Wen, "Energy-efficient wideband cable access networks in future smart cities," IEEE Communications Magazine, vol. 51, no. 6, pp. 94–100, Jun. 2013.

[5]   F. Bonomi, R. Milito, P. Natarajan, and J. Zhu, "Fog computing: A platform for internet of things and analytics," in Big Data and Internet of Things: A Roadmap for Smart Environments, N. Bessis and C. Dobre, Eds. Cham: Springer International Publishing, 2014, pp. 169–186.

[6]   X.SunandN.Ansari,"EdgeIoT:Mobileedgecomputingf ortheinternet of things," IEEE Communications Magazine, vol. 54, no. 12, pp. 22–29, Dec. 2016.

[7]   K. Han, S. Li, S. Tang, H. Huang, S. Zhao, G. Fu, and Z. Zhu, "Application-driven end-to-end slicing: When wireless network virtualization orchestrates with NFV-based mobile edge computing," IEEE Access, vol. 6, pp. 26567–26577, 2018.

[8]   J. Yao and N. Ansari, "QoS-aware joint BBU-RRH mapping and user association in Cloud-RANs," IEEE Transactions on Green Communications and Networking, vol. 2, no. 4, pp. 881–889, Dec. 2018.