

A User Preference Tree based Personalized Route Recommendation System for Constraint Tourism and Travel

Kanimozhi U (✉ kanimozhi@crayondata.com)

Crayon Data

Sannasi Ganapathy

VIT University - Chennai Campus

Manjula D

Anna University Chennai

Arputharaj Kannan

VIT University

Research Article

Keywords: Graph Database, Decision Tree, Tourism, Recommendation System, Social Media Mining, Tree Based Recommendation, Trip Advisor, User Centric, Jaccard Coefficients, Latent Dirichlet allocation and Consumer generated content.

Posted Date: August 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-755856/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

A User Preference Tree based Personalized Route Recommendation System for Constraint Tourism and Travel

Abstract: Personalized recommendation systems recommend the target destination based on user-generated data from social media and geo-tagged photos that are currently available as a most pertinent source. This paper proposes a tourism destination recommendation system which uses heterogeneous data sources that interprets both texts posted on social media and images of tourist places visited and shared by tourists. For this purpose, we propose an enhanced user profile that uses User-Location Vector with LDA and Jaccard Coefficients. Moreover, a new Tourist Destination tree is constructed using the posts extracted from TripAdvisor where each node of the destination tree consists of tourist destination data. Finally, we build a personalized recommendation system based on user preferences, A* algorithm and heuristic shortest path algorithm with cost optimization based on the backtracking based Travelling Salesman Problem solution, tourist destination tree and tree-based hybrid recommendations. Here, the 0/1 knapsack algorithm is used for recommending the best Tourist Destination travel route plans according to the travel time and cost constraints of the tourists. The experimental results obtained from this work depict that the proposed User Centric Personalized destination and travel route recommendation system is providing better recommendation of tourist places than the existing systems by handling multiple heterogeneous data sources efficiently for recommending optimal tour plans with minimum cost and time.

Keywords: Graph Database, Decision Tree, Tourism, Recommendation System, Social Media Mining, Tree Based Recommendation, Trip Advisor, User Centric, Jaccard Coefficients, Latent Dirichlet allocation and Consumer generated content.

1. INTRODUCTION

Recently, the recommendation systems are dealing with the issue of information overload that is faced by users online. Initially, the recommendation systems got success on providing personalized recommendations on e-commerce applications. Later on, it was applied to all the

fields including e-learning, e-health services and e-tourism (Staab and Werthner, 2002). Generally, the recommendation systems on tourism maintain a very large volume of data pertaining to the tourism industry (Werthner & Klein, 1999). The Internet has fundamentally transformed the approach of handling the tourist information in a distributed manner through the development of recommendation systems in order to provide an effective way of tour plan for the people's travel (Buhalis & Law, 2008). It is useful for the tourists to plan their trip and activities to an unknown destination since the recommendation systems are able to list the important tourist destinations which match the tourist's interests and preferences and also recommend the tourist packages according to their budget and time constraints (Kabassi, 2010, Berka and Plonig, 2014).

The tourists are considered as responsible for content production called Consumer Generated Content (CGE) by publishing their travel plans, tourist destination information, interests, preferences, comments, likes, dislikes and personal experiences. Moreover, the tour recommendation systems make use of the Web 2.0 technologies including social media, virtual travel communities, Wikipedia, collaborative tagging and travel blogs (MacLaurin, & Crotts, 2007) for providing most suitable recommendations. The travel related data which are available in the web pages related to online tourism domain and it remains as a next step to develop new marketing strategies (Xiang et al 2008). In the current scenario, personalized tourism and travel recommendations are gaining momentum for effectiveness, and success in tourism (Ricci, 2002; Staab and Werthner, 2002). It provides customized information to tourists based on their preferences, likes and tastes (Gao, Liu, & Wu, 2010). This technique is given more importance because the objective of the recommendation system is to filter out irrelevant options and to provide personalized options relevant to the individual tourist (Adomavicius & Tuzhilin, 2005; Ricci, 2002).

In this work, an enhanced user profile-based User-Location Vector has been developed by using the existing LDA and Jaccard Coefficients. Moreover, a new Tourist Destination tree and a heuristic search algorithm based on A* algorithm, Travelling salesman problem solution and 0/1 Knapsack problem based approach are proposed in this paper in which the data are obtained by analyzing the posts that are extracted from TripAdvisor where each node of the destination

tree consists of tourist destination data. Finally, we build a personalized recommendation system based on user preference, tourist destination tree, heuristic search, and LDA and Jaccard coefficient for similarity analysis to make a tree-based hybrid recommendation system. In addition, the 0/1 knapsack algorithm is also applied for recommending the best Tourist Destination with optimal cost and time that provides the travel route plans according to the spatial, temporal, cost and travel constraints of the tourists.

The major contributions of this paper are as follows:

1. The proposal of new a method for user profile maintenance using location constraints and the application of LDA and Jaccard coefficient for providing effective recommendations.
2. A tree based approach based on Trip_Advisor data is proposed in this paper to find the new tourist destinations more efficiently.
3. A new personalized recommendation system is proposed in this work which uses user profile information and the browsing pattern of the users.
4. The Knapsack problem solution approach is used in this work for optimizing the distance and cost of visiting places.
5. Finally, a heuristic search approach is proposed by using the combination of A* algorithm and the Travelling Sales Man problem in order to find the shortest path from the source to destination by visiting all the required places with optimal cost.

The rest of this paper is formulated as below: Section 2 describes in detail about the user profiles that are created and given as insight on various tourism related recommendation systems. Section 3 explains the proposed user preference tree based personalized route recommendation system. Section 4 discusses the data set used and various evaluation parameters which are applied for evaluating the efficiency of the proposed system. Finally, section 5 concludes the work with suggestions for proceeding further in this direction.

2. LITERATURE SURVEY

Many recommendation systems have been proposed by various researchers in the past (Sairamesh et al 2015, Sankar et al 2017, Sankar et al 2019, Kanimozhi et al 2019). Creation of a

user profile is an important component of recommendation systems which is consisting of basic user information, travel interests and travel histories and these are very essential for developing personalized recommendation systems. In the past, Hernandez *et al.* (2013) developed a new approach to construct user profiles from tweets. In their work, the attributes of users are built based on self-description and a rule-based analytics language was used for mining the Boolean features from the collected tweets. Esparza et al (2013) proposed a new user profile creation and maintenance method which performs filtering based on the topics and also categorized the URLs. Ostrowski and Alfred (2015) evaluated the topic modeling approach by using Latent Dirichlet Allocation (LDA) to support a better understanding of trends in Social Media. Yamasaki *et al.* (2013) used a collaborative filtering approach which uses user's travel history as well as the relationship between the user, the location and the interest group of users in order to predict the preference attractions by using the combination of Markov chain modeling and topic modeling. Cao *et al.* (2010) clustered huge amounts of geo-tagged photos using content-based filtering and recommendation techniques to mine the required knowledge from user travel history for providing users with similar attractions.

One such hybrid method of recommendation is based on the principles of fuzzy set theory. In such models, a fuzzy user preference tree is used to represent a user's preferences as stated by Esparza *et al.* (2013) and Wu *et al.* (2015). The application of fuzzy rules in these models help to handle the decision making process with incomplete data. Neriet *al.* (2012) proposed a personalized travel planning system which considers various types of user needs and it also provides the users with a travel schedule planning service simultaneously based on the users' needs using sentiment analysis on the data obtained from the social media. The use of sentiment analysis provides a suitable tourist location with respect to the characteristics of the user. Shenet *al.* (2016) discussed about their recommendation system which provides a user adopted interface and also provides adjustable results dynamically based on the user interests. Sun *et al.* (2013) proposed a new personalized travel recommendation system which recommends places of tourist attraction based on user interaction and collective intelligence in a unified framework. This model is useful to have destinations based on current interest. Chiang and Huang (2015) developed a tourist recommendation system which is capable of identifying and matching visual characteristics of tourist location images pertaining to the user's places of interest, and images

of various tourist attractions. Chen *et al.* (2014) proposed a novel approach to recommend a travel route that helps to visit the popular tourist attractions in a destination without considering the user preferences into account. However, the recommendation systems must consider the user interests as well.

Yamasaki et al (2013) developed a geo-tag recommendation system for recommending travel best routes which considers the cyclic and time constraints. The systems recommends the travel routes based on geo-tags which are attached with Flickr images. Moreover, the Bayesian theory-based matrix manipulation technique was applied by the authors for performing the efficiency calculations without considering the user preferences. Zhang *et al.* (2015) proposed a tour recommendation system that builds a relationship graph between different users based on their behaviors for providing effective recommendations. The tour recommendation system developed by Yeran Sun et al (2015) collects geo-tagged photos and uses this information alone for providing the recommendations. Hsiu-Sen Chiang et al. (2015) developed a tourist recommendation system which collects the information about places from normal travel websites and travel blogs. Their model is able to provide suitable recommendations when complete data are available. Jiangning Hea et al. (2016) discussed about a recommendation system which uses a predefined dataset of places that contains all information about a place so that it is possible to recommend the most relevant places to the users.

In the tour recommender system proposed by Jiangning Hea et al (2016), the user's past travel information is applied for topic modeling and this model is used to recommend tourist locations to new users based on the similarity analysis to find the users with similar interests. Many other researchers including Zhiwen Yu et al. (2016), Hsiu-Sen Chiang et al. (2015) and Jiangning Hea et al. (2016) proposed tourist recommendation systems that collect user's travel history and provide recommendations based on their previous travel patterns. The recommendation system proposed by Zhiwen Yu et al. (2016) analyzes the users data by the application of a-priori algorithm for obtaining the frequent travel patterns of the specific user that are used subsequently for providing travel plan recommendations. Information about the travel roads and routes used by most people is collected and analyzed for providing personalized recommendations in the work developed by Yeran Sun et al. (2015). The major drawbacks of

the existing and surveyed recommendation systems are that the systems proposed by both Zhiwen Yu et al. (2016) and Hsiu-Sen Chiang et al. (2015) collect information (place and time) explicitly from the user and hence these systems may not result as an expert opinion based travel recommendation system. In addition, Hsiu-Sen Chiang et al. (2015) in their system allowed the users to modify the results in order to make further personalization. This unsupervised human entered data may include noise and the recommendations may turn out to be unreliable and thus making the recommender system to be error prone.

Based on these limitations, it is necessary to propose an effective and personalized recommendation system for optimizing the time and cost of tours. Therefore, a new personalized tourism recommendation system is proposed and implemented in this work which overcomes the limitations of the existing tour recommendation systems. In this model, User Profile is created by extracting implicit information available in multiple heterogeneous data sources namely Flickr and Facebook. The geo-tagged photos posted by tourists in Flickr as well as text and the images posted by tourists in Facebook are used in this work to mine the user interest. Each User Profile is stored as a user node in the proposed user preference tree structure that stores this mined explicit and implicit information of users hierarchically. The User Preference tree and Tourist Destination Tree are constructed in this model and they are compared and analyzed to give a personalized recommendation to the tourist. The analysis applies rules, temporal and cost constraints, LDA based topic modeling and Jaccard similarity measure, 0/1 Knapsack approach, A* algorithm and Travelling salesman problem based solution for performing optimal tourist place recommendations.

3. METHODOLOGIES

The flow of the system design is given in Figure 2. The proposed user centric personalized recommendation system collects data from multiple heterogeneous data sources such as posts and photos of users interested in traveling from the social media and interprets the posts and photos to find the topic of interest of users and recommends personalized tourist destinations.

The demographic information, posts and photos of the travelers are collected from Facebook. The photos of tourist destinations from Flickr are collected and clustered based on geo tags and representative images for each cluster is constructed to build a database of images. These tags

are then passed on to a Latent Dirichlet allocation (Blei et al., 2003) model built to obtain probabilities associated with each topic which is the first probability value to be considered in this work. Places travelled by users are collected from Facebook and this information is used in two ways: First the collected information is applied for finding the same users by applying Jaccard similarity. The descriptions of those travelled places are passed to an LDA model to obtain probabilities associated with each topic and this is the second probability value to be considered. The posts that the users post are also passed to an LDA model to obtain associated probabilities, which is the third probability value to be considered. A user-location vector is computed which maps the tourist with various categories of point of interest visited by them. By proposing a new weightage algorithm, each category of point of interest are given different weightage based on likes, comments, posts shared by the tourist for every point of interest. The points of interest the user visited that are extracted with tagged location which is identified from the profile.

All the three probabilities are combined by assigning weights and the resulting combined probabilities with the user-location vector are stored in a newly developed data structure as a user node. As a result of storing user nodes in the graph database, links between users can be easily constructed. Both the user preference nodes and destination tree nodes are used to generate a recommendation list for tourist destinations along with other travel information about the recommended destination. Travel route suggestion is also done using the Knapsack algorithm suggesting the best route for the tourist in the available time. The comprehensive explanation of the research design is explained in the sections below.

3.1 Traveler Information Extraction from Facebook

Initially traveler's data from various Travel Interest Groups available at Facebook is collected which consists of five hundred posts and fifty timeline photos. The ids of these popular travel bloggers are collected using Facebook Graph API, which further allow us to collect demographic information, timeline photos and posts posted by the user. The information about traveler's user id, their Facebook feed, age, gender, photos and tagged location are collected, preprocessed, tokenized and further tagged to build basic user profiles.

3.2 Users' Profile

The photos are grouped by applying their grouped geo-tags for finding the tourist locations where they were collected, and the locations are also identified and labeled with semantic information. The algorithm also creates a user-location matrix and user-user similarity matrix that can be further used for personalized tourist recommendations. To build the basic user profile, the user data are preprocessed, tokenized and POS tagged.

3.2.1 Preprocessing

Users' data such as user's feed, age, gender, photos and their tagged location collection are extracted from Facebook. Here, the user's data is preprocessed by removing redundant terms, unwanted data, irrelevant data, stop words and null values.

3.2.2 Tokenization and POS Tagging

The preprocessed user data are tokenized, and POS tagged in order to identify the tagged location of the places visited by the user. Sequence of tagged location travelled and visited by the user is extracted to identify the travel patterns of the user. The tourist destinations extracted are generally portrayed in the form of either sequences or tree structure as given in Figure 1 to identify the travel patterns and dependability between the extracted destinations (Kavitha et al, 2017).

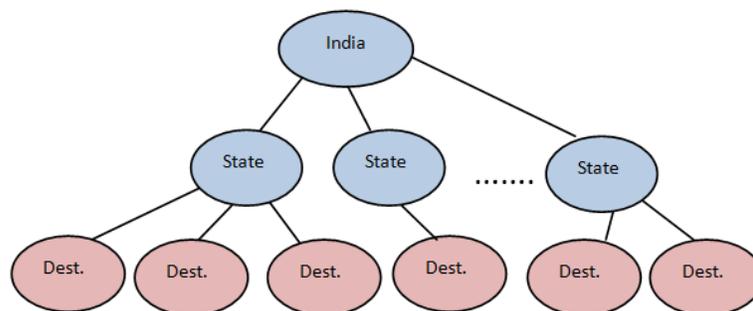


Figure 1. Tourist Destination Tree

3.3 Reference Image Cluster Builder from Flickr Images

The main aim is to share the images in social media that the tags are annotated with. The input to the Reference Image Cluster Builder is the set of photos of tourist destinations from Flickr that are geo-tagged. From the retrieved geo-tagged photos, photo tags are collected, and these tourist destination photos are clustered based on geo-tags.

3.3.1 Retrieve geo-tagged photos

Geo-tagged photos of Tourist destinations are retrieved from Flickr using the Flickr API. Photos that have tags associated with Tourist destinations and have their geo-coordinates enabled are retrieved. Once the source URLs are obtained, the actual photos are extracted.

3.3.2 Photo Tags Collection

The set of photo ids are collected from retrieved geo-tagged photos and these are used to obtain the tags tagged with the photos collected. All tags are collected from states of India and are clustered state-wise basis using geo-tags.

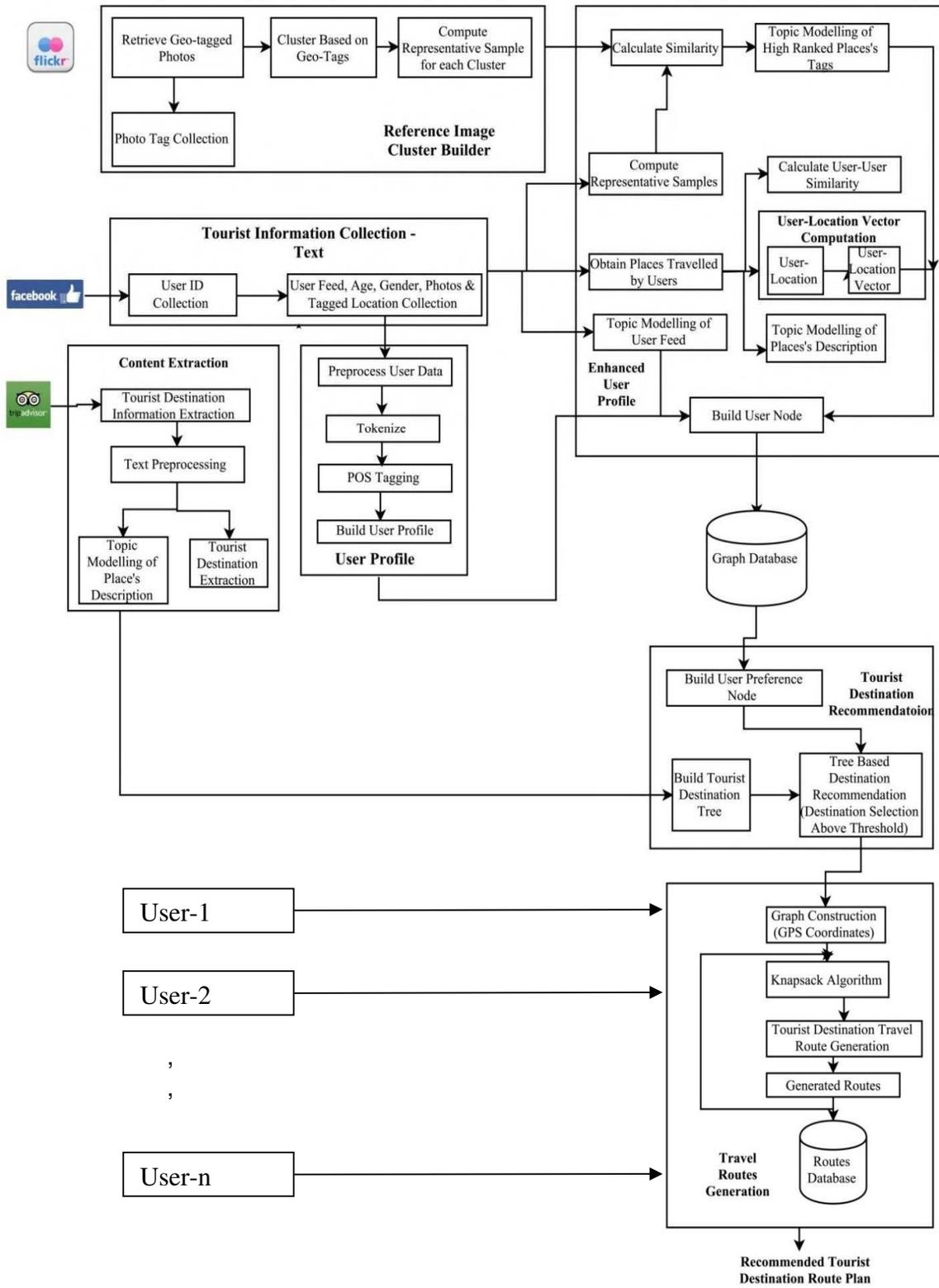


Figure 2. System Architecture

3.3.3 Cluster based on Geo-Tags

The geo-tags are clustered based on Latitudes and longitudes that are retrieved from the photos using the mean shift clustering algorithm to obtain state wise clusters of images. Mean shift clustering is used since it is a non-parametric clustering algorithm.

3.3.4 Compute representative samples for each cluster

The “message passing” concept-based clustering algorithm called Affinity Propagation (AP) for finding the difference between the data points. Affinity propagation algorithm is used to find exemplar for each image in a set of images provided. Messages are iteratively exchanged between data points until a good solution with a set of exemplars is reached. Once the algorithm converges, the exemplar P_i for image i is selected by equation (1).

$$p_i = \arg \max_k [r(i, k) + a(k, i)] \quad (1)$$

Where, $r(i, k)$ is the responsibility, the message sent to candidate exemplars ' k ' from data point ' i ', $a(k, i)$ is the availability, the message sent to data point ' i ' from candidate exemplar ' k '

Similarly $s(i, k)$ i.e. similarity between two images i and k is computed using the following equation as

$$s(i, k) = - \left(\sum_{j=1}^r w_{j1} |E_j^1 - E_j^2| + w_{j2} |\sigma_j^1 - \sigma_j^2| + w_{j3} |s_j^1 - s_j^2| \right) \quad (2)$$

where, i and k are the two images compared.

j indicates the channel index and r represent the total number of channels

E_j^1, E_j^2 are the first order moments calculated for distributing the image.

σ_j^1, σ_j^2 is the second order moments calculated for distributing the image.

s_j^1, s_j^2 is the third order moments calculated for distributing the image.

w_{j1}, w_{j2}, w_{j3} are weights assigned.

3.3.5 *Enhancing User Profile*

The inputs are the user's feed, demographic information like age, gender timeline photos and tagged locations. From the photos collected, representative samples are obtained from the reference image cluster builder module and the output obtained is the enhanced user profile stored as a user node. The steps involved to enhance user profiles are Mine places travelled by users, User-location Vector, Calculate user-user similarity, Compute representative sample of user images, Topic modeling of user posts, places description and high ranked places' tags and Build user node.

3.3.5.1 *Mine places travelled by users*

Popular travel destinations around the world visited by travelers are mined from the posts of the collected users. The list of places visited by each traveler is collected and stored.

3.3.5.2 *User Location Vector Computation*

The user visited the different points on the various user location vector maps. The various weightages are given based on user's likes, comments and posts. The point of interest is also collected according to their likes, comments and posts in the past. The various types of tourist places including historic location, temples, park, beach, mountain etc. are also given different weights based on frequency of visit to various categories of tourist destination by the tourist. Thus a particular user is likely to visit more places with similar characteristics.

Algorithm 1:

n: no. of users

for j=1 to n

 v<-total number of tagged_locations

 l<- likes on locations

 m<- number of category of places

 classifytagged_locations and liked places into predefined categories (point of interest)

 j<-(j1,j2,...,jm) <- value for each category of tagged_location

```

k<-(k1,k2,...,km) <- values for each category of likes location
for p=1 to m
    i[p]<-(2*k[p]+j[p])/(v + 1);
end for
end for
return i [(i1,i2,...i7) <- user-place vector]

```

T	B	P	H	M
0	0	0	0.8	0.2
T	B	P	H	M
0.6	0.2	0.1	0	0.1
T	B	P	H	M
0.7	0.1	0.2	0	0

Figure 3.User Location Vector

Where, T ->Temple, B->Beach, P->Park, H->Historic, M->Museum.

As shown in figure 3, each User is mapped to different categories of point of interest associated with a weightage given based on the user's comments, likes and posts. Weightage is assigned to the user's tagged location. Weights assigned to the tagged location are compared with the weights assigned to point of interest for the likes, comments and posts.

3.3.5.3 Calculate user-user similarity

Similarity between travelers is computed based on the Jaccard similarity coefficient. The similarity coefficient is calculated between traveler pairs for recommending the various locations to various users of the same interest. By using a User-location vector, the same set of users are extracted and used for personalized recommendation based on user interest. Jaccard coefficient is computed by equation as,

$$J(A,B)=\frac{(Places_A \cap Places_B)}{Places_A \cup Places_B} \quad J(A,B) = \frac{(Places_A \cap Places_B)}{(Places_A \cup Places_B)} J(A,B) = \frac{(Places_A \cap Places_B)}{(Places_A \cup Places_B)} \quad (3)$$

Where A, B are users and $Places_A, Places_B$ are places that are visited by A and B respectively, which are extracted from the User-location vector. In this work, based on the Jaccard similarity computation, it is restricted to choose the top three similar users for each user.

3.3.5.4 Compute representative sample of user images

For each traveler a set of uncounted timeline photos posted were collected as part of data collection. These photos are now analyzed to obtain a set of representative samples. Affinity propagation algorithm used while calculating the representative samples of clusters of Flickr photos is used to compute the representative samples of images posted by users. Here, all photos by the user are considered as a single cluster.

3.3.5.5 Topic modeling of user posts, places description and high ranked places' tags

The LDA model developed by applying scraped tourist place data which is used for obtaining the possibilities for a group of predefined topics on tourist destinations, posts posted by the user, description of popular tourist destinations around the world visited by the user and for the set of tags tagged along with the pictures posted in Flickr. Affinity propagation on user images gives a set of representative samples of user's interest which are compared with the representative samples generated from the images collected from Flickr. These images are already clustered with each cluster representing a state. The output is the set of probabilities for each topic, for each user.

3.4 Build User Node

User being the center theme, a data structure (Kavitha *et al.*, 2016) with users as its nodes would best fit this work. Here, each user is represented by the user node built from the preceding modules. Input to the user node is the set of results from topic modeling of posts, places description and high ranked places' tags, the user id and the user-location vector. After obtaining three probabilities (i) probability that correlate with every topic of interest and description of the places travelled (ii) probability that corresponds with each topic of interest and posts posted by user and (iii) probabilities that conform user's interest based on places

travelled by the user, these values are collated using equation (4) to obtain a combined probability value.

$$P_{userk} = P_{userk} = 0.5 * P_k P_k(user posts) + 0.3 * P_k P_k(user photos) + 0.2 * P_k (place visited_{user})$$

$$P_k (place visited_{user}) \quad (4)$$

where, k is k^{th} topic,

$$P_k P_k(user posts) = LDA(user post),$$

$$P_k (user photos) = LDA\{Tags of \{Aff_prop(user images)\}\}$$

$$P_k (user photos) = LDA\{Tags of \{Aff_prop(user images)\}\}$$

$$P_k (places visited_{user}) = \sum_{places\ visited} \frac{1}{N} * P_{topic}(places\ visited) P_k$$

$$P_k (places visited_{user}) = \sum_{places\ visited} \frac{1}{N} * P_{topic}(places\ visited) P_k (user$$

$$posts) = LDA\{Tag\ of\ \{Aff_prop(user\ images)\}\},$$

$$P_k P_k(places\ visited_{user}) =$$

$$\sum_{place\ visited} \frac{1}{N} * P_{Topic}(place\ visited), \sum_{place\ visited} \frac{1}{N} * P_{Topic}(place\ visited),$$

And

$$P_{Topic}(place\ visited) = LDA(place\ descriptions)$$

$$P_{Topic}(place\ visited) = LDA(place\ descriptions)$$

3.5 Storage in Graph Database

The input user nodes built in the enhancing user profile module are successfully stored in graph database neo4j. Links are drawn between a user and three users most like that user, based on the Jaccard coefficients calculated. Each node is identified using the user id, which is unique for each user.

3.6 Tourist Destination Content Extraction from TripAdvisor

In the Tourist Destination content extraction phase document pre-processing and topic modeling are done which are explained below.

3.6.1 Document pre-processing

The user's reviews scraped the tourist data manually that are collected and formatted for selecting the input format needed for building the LDA model proposed by Blei *et al.* (2003). Algorithm 2 is applied to get the formatted text file.

Algorithm: Data set builder (Reviews, DestInfo)

Input: Reviews, DestInfo (destination information collected)

Output: FTD-Formatted Training Dataset

Begin

foreach Review/DestInfo page **do**

 Remove extra spaces, numbers and symbols

 Represent the document as a vector

foreach document **do**

 Add a unique Id and identifier

end

 Output FTD

End

3.6.1 Topic Modeling

Topic modeling module builds a topic model using LDA algorithm. The LDA model is a machine learning algorithm and unsupervised that identifies latent topic data from a huge volume of documents. This method relies on a "bag of words" and treats every document as a vector of word counts. Here, selects a particular topic from a collection of available topics for every document. The selected topic of words is associated with the distributed words. The same kind of activity is repeated for all the available words in the specific document. Finally, the LDA is based on the hypothesis of a person writing a document that has specific topics in mind.

A whole tourist data extracted which is represented as a mixture of various topics of interest of the user. The probability of occurrence of can be computed using equation (5).

$$\begin{aligned}
P(\text{term}_i | d) &= \sum_{j=1}^n p(\text{term}_i | \text{topic } Z_i = j) p(\text{topic } Z_i = j | \text{document } d) \\
P(d) &= \sum_{j=1}^n p(z_i = j) p(z_i = j | d) P(d) = \sum_{j=1}^n p(z_i = j) p(z_i = j | d)
\end{aligned}
\tag{5}$$

In the equation 5, $P(\text{term}_i | \text{document } d)$ is the probability of occurrence of the i^{th} term related to travel and tourism for a given document 'd' extracted from user profile and details of tourist destination and Z_i is the topic to which the term_i related to travel and tourism present in the document 'd' belongs. Also $P(\text{term} | Z_i = j)$ is the probability of occurrence of i^{th} term related to travel and tourism within topic j related to travel and $P(\text{topic } Z_i = j | \text{document 'd'})$ is the probability of picking a term related to travel and tourism from topic j related to travel and tourism in the document. The number of topics related to travel and tourism Z has to be defined in advance and allows in adjusting the degree of specialization of the topics.

Using equation (5), the posterior probabilities can be estimated with equation (6) and (7)

$$p(z_i = j) = \frac{c_{t,j}^{TZ}}{c_{t,j}^{TZ} + T\beta} p(z_i = j) = \frac{c_{t,j}^{TZ}}{c_{t,j}^{TZ} + T\beta} p(t_i | Z_i = j) = \frac{C_{t,j}^{T,Z}}{\sum C_{t,j}^{T,Z} + T\beta}
\tag{6}$$

$$p(d) = \frac{c_{d,i,z}^{DZ}}{c_{d,i,z}^{DZ} + Z\alpha} p(d) = \frac{c_{d,i,z}^{DZ}}{c_{d,i,z}^{DZ} + Z\alpha} p(Z_i = j | d) = \frac{C_{d,i,z}^{D,Z}}{\sum C_{d,i,z}^{D,Z} + Z\alpha}
\tag{7}$$

where $p(t | Z)$ is the topic term distribution

$p(Z | d)$ is the document topic distribution related to travel and tourism

$C^{T,Z}$ is the count of all topic-term assignments related to travel and tourism

$C^{D,Z}$ Counts the document-topic assignments related to travel and tourism

α and β are the hyper parameters for Dirichlet priors

An LDA model is built using the data set obtained from the preprocessing step.

3.7 Tourist Destination Recommendation

The recommendation of Tourist places is built using the created user preference node and tourist destination tree. A tree-based recommendation system is built which helps in the generation of travel route suggestions based on user interest. Moreover, A* algorithm and Travelling Salesman Problem (TSP) based approach have been used in this work for providing suitable location recommendations.

3.7.1 Build user preference node

Each user node is retrieved a unique user id and parsed for obtaining the probabilities from the database that inputs into a user preference node. Moreover, a user node contains the probabilities of each topic obtained from mining the posts, photos and places visited by the users.

3.7.2 Build destination tree

Tourist Destination trees are a key component in the recommendation system. The text files obtained for each destination are stored inside folders which have the name of states of tourist destinations. The probabilities associated with each file are computed using the LDA model built and the corresponding fields of the database are filled. The completed database is used to create a destination tree. The destination tree consists of two kinds of nodes such as a leaf and intermediate nodes. Here, a leaf node inherits from an intermediate node and the intermediate node contains just a name feature.

3.7.3 Tree based destination recommendation

A fuzzy reference approach is used to recommend. The probabilities in the leaf node are considered as fuzzy values. The algorithm 2 is applied for generating the recommended users list for three tourist destinations of the user.

Algorithm 3: Tree Based Destination Recommendation

Recommender (Node, T)

Input: Node-User preference node, T- destination tree

Output: L_r -Recommendation list

Begin

for each Node of T **do**

 Compare the probabilities and obtain the top three topics for user from the user preference node

 Recursively iterate the tree using A* search with TSP based optimization and obtain the top three places for each topic by comparing the probabilities.

 Output L_t - lists obtained for each topic

for each top destination in each L_t **do**

 Obtain the state of the destination

 Obtain the top five destinations of the state with respect to the user's interesting topics using the probabilities generated

end

end

 Output L_r

End

3.8 Tourist Destination Route Plan

Tourist Destinations that are recommended using the Tree based destination recommendation are passed on to the Knapsack algorithm to generate the best tourist destination route plan using the time constraint of the user as weights.

,

3.8.1 Graph Construction

The graph construction is classified into three such as location-location graph, user-location graph and user-user graph. In the location-location graph visit the edge between two locations that are the correlation between two locations' strengths. In the User-location graph, users and the locations are the two kinds of entities in the user location graph. In the user-user graph, a

node is a user and edge between two nodes represents two relations. The top 5 tourist destinations that fit the user's interests are selected and a weighted graph is constructed. The weights between the nodes being the distance between the places which is calculated using haversine distance. The haversine distance metric finalizes the great-circle distance between two points on a sphere given their longitudes and latitudes.

Algorithm: Graph Construction

Input: selected nodes, their GPS coordinates

Output: weighted Graph of selected places

```

For I =1 to n nodes
    for j =i to n nodes
        draw nodes
        draw edge
        calculatehaversine distance between place[i] and place[j] and
        assign weights
    End for
End for
Return Graph

```

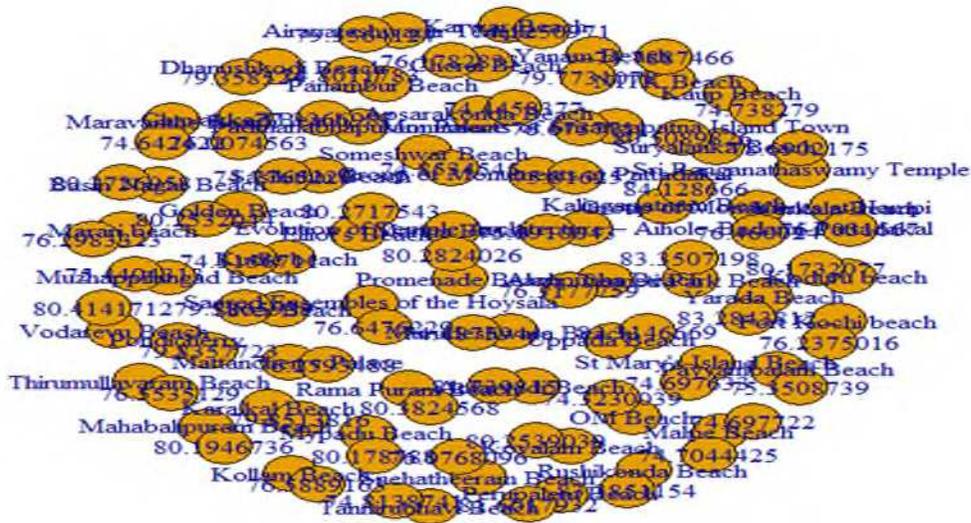


Figure 3.The constructed graph of recommended places

3.8.2 Knapsack Algorithm

The nodes are then subjected to Knapsack algorithm that generates and suggests travel plans of different time periods by changing the capacity of the Knapsack which is the time in this case.

ALGORITHM

Step 1: Input user destination place

Step2: attractions display in the recommended system

Step3: Previous user already done post, share, comment to the location in particular visitors' areas

Step4: Based on user preference the more weightage location will be recommended to the user

Step5: Weightage is based on (existing users likes, comments, shares,)

Step 6: Choose the best location according to the user time and cost and weightage)

Step7: Knapsack algorithm is to find best location according to the like, post, comment

Step8: this algorithm evaluating period of time how many interesting areas to be covered in particular time length;

3.8.2.1 Travel route generation

The recommendation algorithm generates a list of recommended tourist destinations. Top five places to visit are listed out in each state based on the topic of interests of the user and popular destination travel routes in a particular state are also suggested using the Knapsack algorithm. Knapsack algorithm which is also called a rucksack problem is a combinatorial optimization.

Input: weighted Graph of selected places

Output: Travel plans of different time periods

Algorithm:

For W from 24 hrs to 5 days

for j from 0 to W

m[0, j] := 0

for i from 1 to n

```

for j from 0 to W
if w[i] > j then: m[i, j] := m[i-1, j]
           Else: m[i, j] := max(m[i-1, j], m[i-1, j-w[i]] + v[i])
End if
End for
End for
End for
Store route
End for
Return routes of different time periods.

```

The time period considered is from 24 hours to 5 days. Based on the time availability of the user the best travel routes are recommended.

4. RESULTS AND DISCUSSIONS

This section discusses the dataset and evaluation parameters for evaluating the performance of the proposed recommendation system designed and their results are also discussed.

4.1 Dataset

The dataset required for this recommendation system is obtained from multiple heterogeneous sources. Posts, photos posted by users in their Facebook page are obtained by using the Facebook Graph API. Data was collected for about thirty users, five hundred posts and fifty photos per person were collected. For reference cluster builder, fifty photos for eighteen Indian states were collected using the Flickr API. Data about tourist destinations were collected manually and crawled from the TripAdvisor website. We have crawled the tourist information from India. Since Tourism in India is a booming industry. Tourism is the second largest earner in India as foreign exchange. Tourism provides a lot of employment to the large volume of people that are skilled and unskilled. Tourism contributes 6.23% to the national GDP and 8.78% of the total employment in the form of service tax from transport, travel agencies, hostel and

airline. This is compared with the topics of interest generated for each user using the LDA model.

4.1.1 Accuracy

Accuracy is a direct proportional to the true results, both true positive and negatives among the total number of cases examined. The formula given in equation (8) is used for evaluating the system accuracy.

$$Accuracy = \frac{\{Expected\}\{Obtained\}}{Expected} * 100 \quad Accuracy = \frac{\{Expected\}\{Obtained\}}{Expected} * 100 \quad (8)$$

The recommendation system {expected}, gives the set of destinations selected by user, and {obtained} gives the set of destinations recommended by the algorithm. Thus Accuracy of the recommendation system is calculated using equation 9.

$$Accuracy = \frac{\{Set\ of\ destinations\ Selected\ by\ user\}\{Set\ of\ destinations\ recommended\}}{Set\ of\ destination\ selected\ by\ user} \quad Accuracy = \frac{\{Set\ of\ destinations\ Selected\ by\ user\}\{Set\ of\ destinations\ recommended\}}{Set\ of\ destination\ selected\ by\ user} \quad (9)$$

Five scenarios are used while calculating accuracy to evaluate the system. Five scenarios are used while calculating accuracy to evaluate the system.

- In the first scenario, probabilities obtained from user posts are given 50% weightage, probabilities obtained from user images are given 30% and probabilities obtained from places visited by users is given 20% weightage.
- In the second scenario, probabilities obtained from user posts are given 40% weightage, probabilities obtained from user images are given 30% and probabilities obtained from places visited by users is given 30% weightage.
- For the third scenario, probabilities obtained from user posts are given 20% weightage, probabilities obtained from user images are given 50% and probabilities obtained from places visited by users is given 30% weightage.

- In the fourth scenario, probabilities obtained from user posts are given 30% weightage, probabilities obtained from user images are given 60% and probabilities obtained from places visited by users is given 10% weightage.
- For fifth scenario, probabilities obtained from user posts is given 20% weightage, probabilities obtained from user images is given 30% and probabilities obtained from places visited by user is given 50% weightage

The results obtained are illustrated in Figure 4.

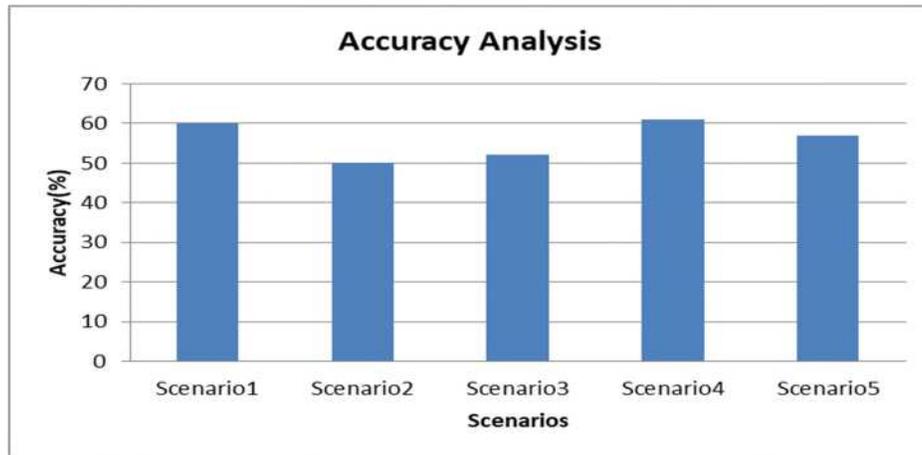


Figure 4 Accuracy Analysis

It is evident from Figure 4 that Scenario 1 and Scenario 4 provide the best accuracy. In Scenario 1 and 4 the user post and user images are given more weightage thus giving more accuracy than other scenarios. The accuracy achieved only 60% and above not 100% because the expected topics were not obtained directly from the user instead of a human classifier.

4.1.2 Recall and Precision

The Recall value is calculated by using the formula which is given in equation (10) which is the number of true positives divided by the sum of true positives and false positives.

$$recall = \frac{\text{true positives}}{\text{true positive} + \text{false positive}} \quad recall = \frac{\text{true positives}}{\text{true positive} + \text{false positive}} \quad (10)$$

According to the recommendation system the recall is given by equation (11).

$$recall = \frac{|(\text{number of destination selected by the user})(\text{recommended destination})|}{|\text{number of destination selected by the user}|}$$

$$recall = \frac{|(\text{number of destination selected by the user})(\text{recommended destination})|}{|\text{number of destination selected by the user}|} \quad (11)$$

Precision provides the recommended items which are preferred by the user given in equation (12).

$$precision = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \quad precision = \frac{\text{true positive}}{\text{true positive} + \text{false negative}}$$

(12)

According to the recommendation system the precision is given by equation (13).

$$precision = \frac{|(\text{number of destination selected by the user})(\text{recommended destination})|}{|\text{recommended destination}|}$$

$$precision = \frac{|(\text{number of destination selected by the user})(\text{recommended destination})|}{|\text{recommended destination}|} \quad (13)$$

Precision and Recall values obtained for different scenarios like Only Text, Only Images etc. are plotted in Figures 4 and Figure 5 respectively.

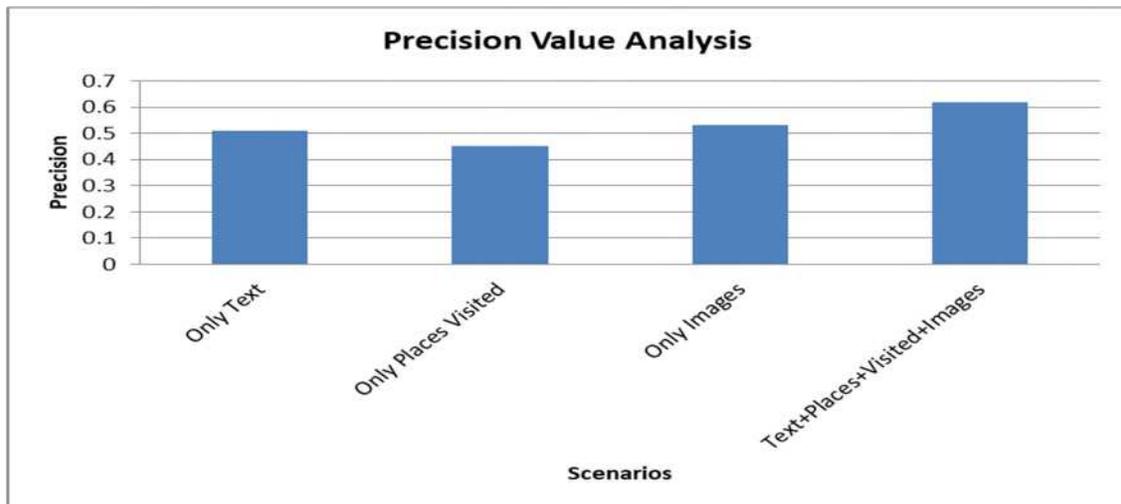


Figure 5.Precision Value Analysis

Figure 5 is a precision graph which infers that, when probabilities obtained from posts, photos and places visited by the user are taken into account, a better value of precision is obtained compared to other recommendations obtained only using text, places visited and images. Posts and photos posted by the users suggest their preferences.

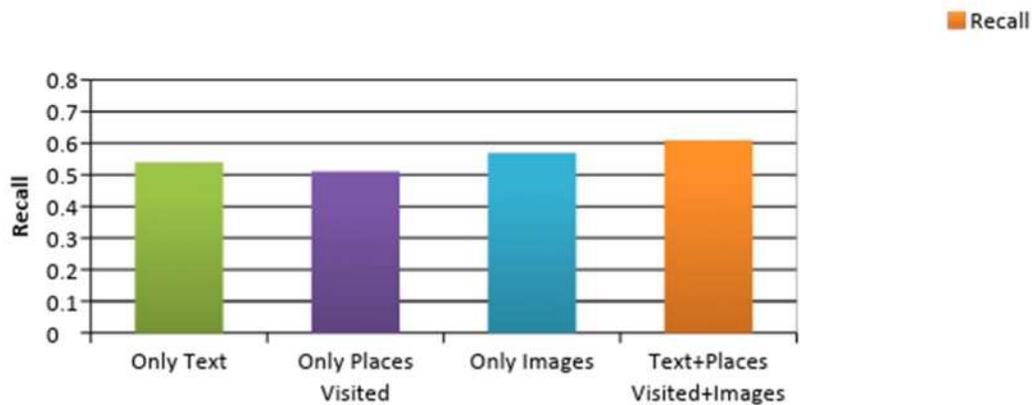


Figure 6.Recall

It is evident from Figure 6 that when probabilities obtained from posts, photos and places visited by the user are taken into account, a better value of recall is obtained. Recall values for a sample of nine user groups are plotted as graph in Figure 6. Values are calculated under two scenarios; First scenario when only posts posted by the user are considered and second when posts, photos and places visited by the user are considered to calculate user probabilities. It is proved that the second method outperforms the first. Figure 6 shows the recommendation system has better recall value when text, places visited and images of tourist destinations are considered. Recommendation using only text and only images is next better than recommending only places visited.

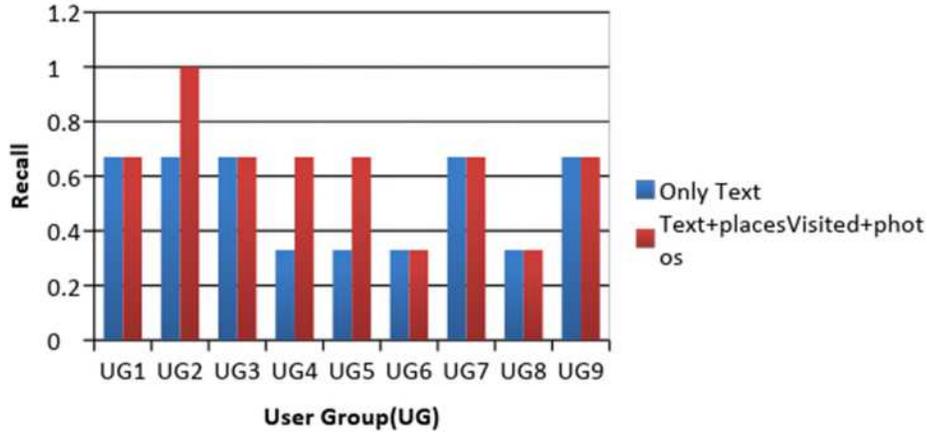


Figure 7.Recalls for Sample User Groups

The high value of precision is preferred for every place which is recommended as an interest of the user. If a place of user's interest has high recall values then, the specific place is recommended. On an average the system gives a recall of 0.67, as shown in Figure 7. The precision and recall obtained are not 100% because of few reasons. First among the reasons is the fact that the expected output is not directly obtained from the user. Second reason is that the built LDA model can identify only a set of six topics, which are not that very accurate in capturing all the interests of users posted. Hence, the bag of words approach used by LDA can only provide approximate results instead of accurate ones. Third reason is because of the limit in the number of interesting topics considered. For the obtained results only three most interesting topics of the user were taken into account.

4.1.3 Specificity

The specificity is calculated by using the number true negative and false positive which means that the true negative is divided by the sum of true negative and false positive.

$$Specificity = \frac{true\ negative}{true\ negative + false\ positive} \quad Specificity = \frac{true\ negative}{true\ negative + false\ positive}$$

(14)

Equation 14 is used to calculate specificity of the recommendation system and the values computed are plotted in Figure 8.

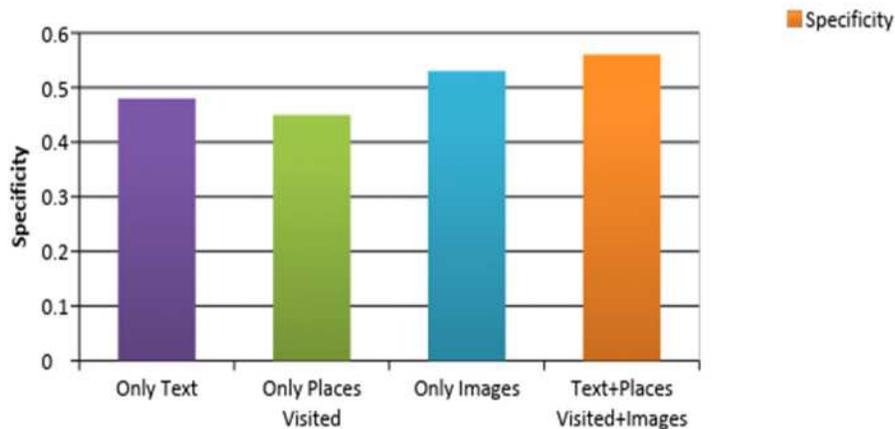


Figure 8.Specificity Analysis

The high value of specificity is preferred for every place which is recommended as interest of the user. If a place of user's interest has high specificity values then, the specific place is recommended. According to Figure 8, when probabilities obtained from posts, photos and places visited by the user are taken into account, a better value of specificity is obtained. Specificity values for a sample of nine users are plotted in Figure 9. Values are calculated under two scenarios; First when only posts posted by the user are considered and second when posts, photos and places visited by the user are considered to calculate user probabilities. It is evident that in almost all cases the second method outperforms the first.

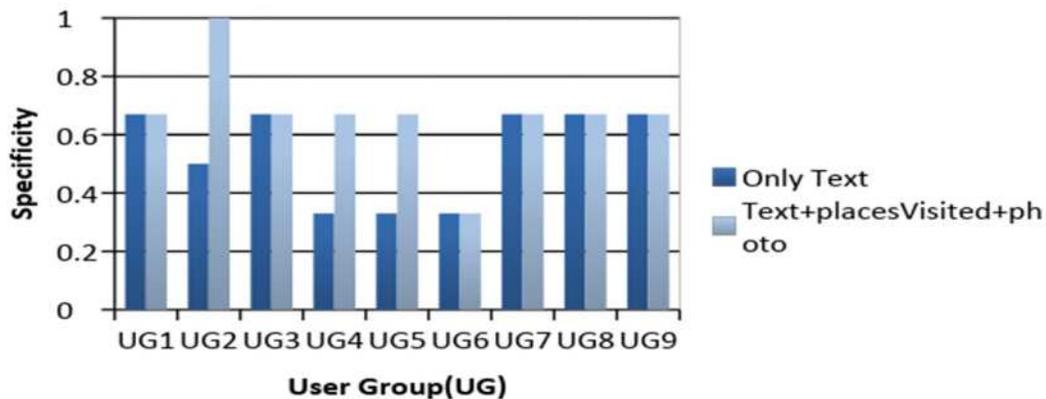


Figure 9.Specificity for sample user groups

Almost all user groups have specificity of about 0.7 in the second scenario, this is because both in the expected and obtained results almost all topics were the same.

4.1.4 F-Score

The F-Score is interpreted as the weighted average of precision and recall. Here, the value 1 is best and 0 is worst.

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (15)$$

The values are computed based on equation (15) and are plotted in Figure 10.

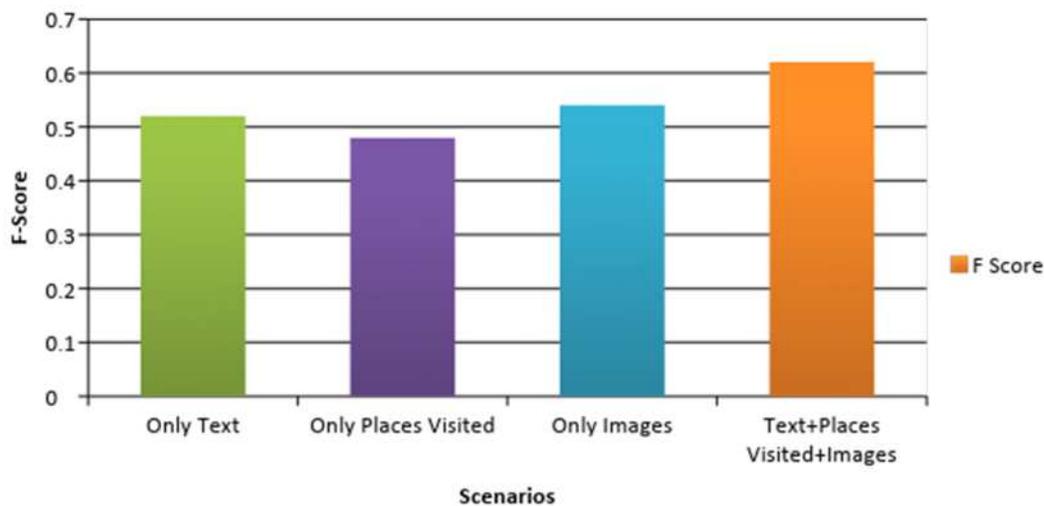


Figure 10. F-Score Analysis

On an average the system gives an F-score of 0.62. F-Score gives the measure of the accuracy of the recommendation system and the score proves that recommending using text, places visited and tourist images gives more accurate recommendation. F-score obtained was not 100% as the expected result was obtained by using a human classifier instead of directly obtaining.

4.1.5 Hit rate

Hit rate can be defined as the intersection of the number of points of interest (POI) recommended and the number of points of interest in the user's list (i.e. the places the users has visited in the past) divided by the number of total POIs recommended. It is used to verify the effectiveness of the personalization of this recommender. This proves how personalized the

recommended places and travel routes are and how it exploits the user's interest to recommend travel routes. The Hit rate is computed as follows:

$$\text{HitRate} = \frac{\text{No. of POIs recommended POIs in the User's list}}{\text{No. of POIs recommended}} * 100$$

$$\frac{\text{No. of POIs recommended POIs in the User's list}}{\text{No. of POIs recommended}} * 100 \quad (16)$$

Table 4.2. Hit rate value table

User Groups	Time Periods(hrs)	Hit rate(%)
User Group 1	24	100
	36	100
	48	100
User Group 2	24	100
	36	75
	48	100
User Group 3	24	100
	36	100
	48	100
User Group 4	36	100
	48	75
	24	100
User Group 5	24	100
	36	100
	48	100
User Group 6	24	100
	36	100
	48	100
User Group 7	24	100
	36	100
	48	100
User Group 8	24	100
	36	75
	48	100
User Group 9	24	100
	36	100
	48	100

User Group 10	24	100
	36	100
	48	100

Figure 11 shows the hit rate value in 24, 36 and 48 hours for the different user groups.

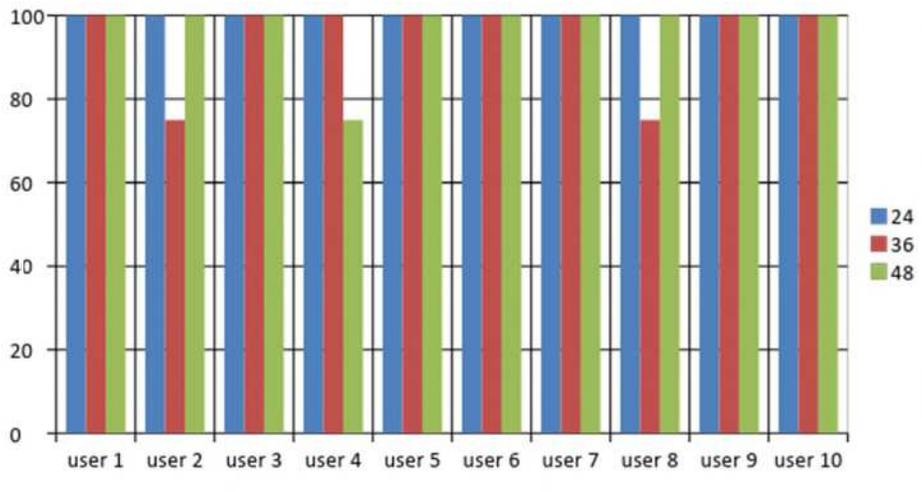


Figure 11.The plotted Hit Rate value in 24 hours, 36 hours and 48 hours

From figure 11, it can be seen that the hit rate value for the three times with 12 hours interval between 24 and 48 hours. Totally, seven user groups achieved 100% hit rate in all timings such as 24, 36 and 48 hours.

4.2 Comparative Analysis

The comparative analysis is done by considering the existing tourist recommendation systems and the proposed tourist recommendation system. Here, we have considered the time and cost for performing comparative analysis.

First, figure 12 shows the cost analysis between the proposed tourist recommendation system and the existing tourist recommendation systems that are developed by Keson et al (2017), Tenemaza et al (2020) and Fararni et al (2021). Here, we have conducted five experiments such as E1 (Recommendation 1), E2 (Recommendation 2), E3 (Recommendation 3), E4 (Recommendation 4) and E5 (Recommendation 5).

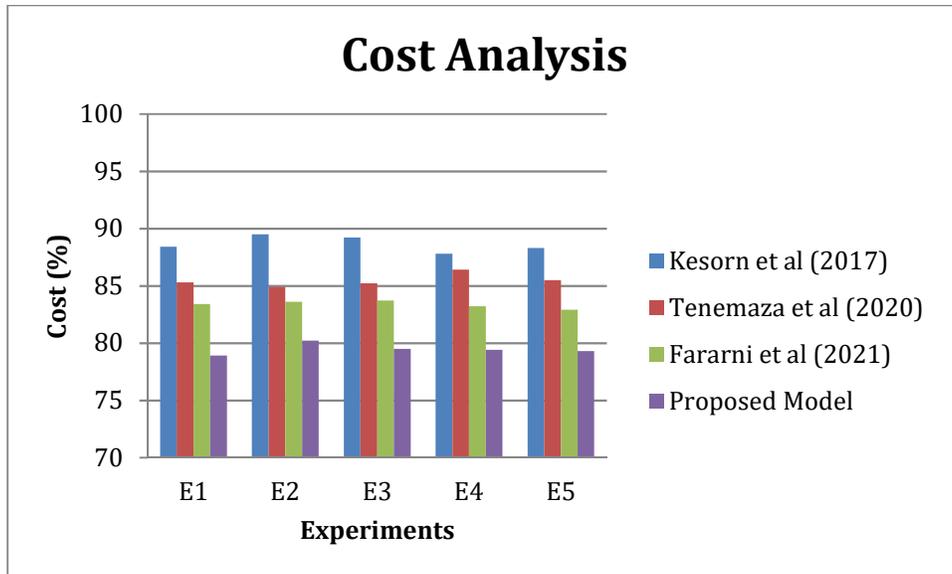


Figure 12. Cost Analysis

From figure 12, it is observed that the cost spent for the recommended plan by the proposed recommendation system is less when it is compared with the cost spent for the existing recommendation systems.

Figure 13 shows the time analysis between the proposed tourist recommendation system and the existing tourist recommendation systems that are developed by Keson et al (2017), Tenemaza et al (2020) and Fararni et al (2021). Here, we have conducted five experiments such as E1 (Recommendation 1), E2 (Recommendation 2), E3 (Recommendation 3), E4 (Recommendation 4) and E5 (Recommendation 5).

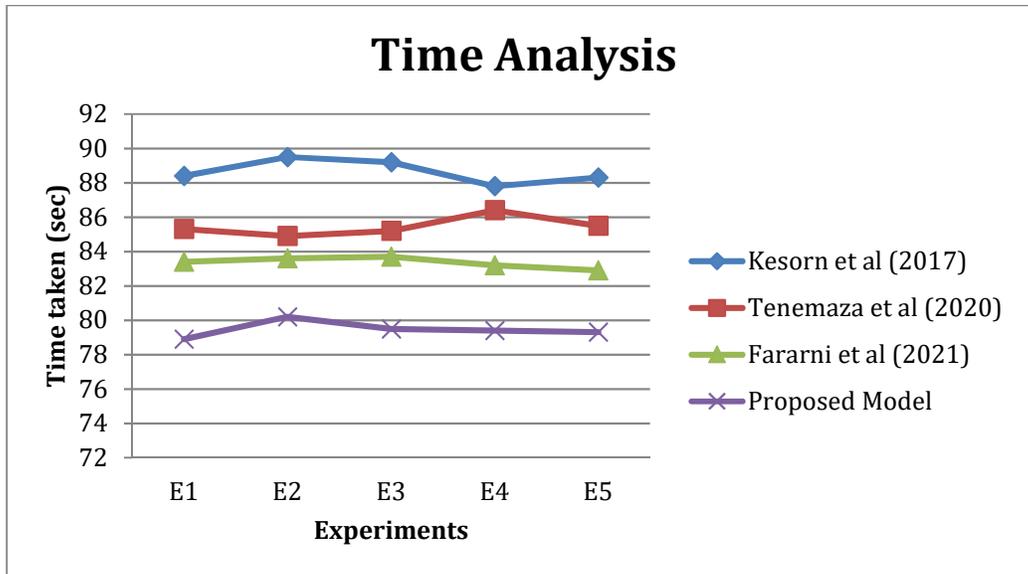


Figure 13. Time Analysis

From figure 13, it is observed that the time taken for a tourist plan by the proposed recommendation system that is less when it is compared with the time taken by the existing recommendation systems that are developed by Keson et al (2017), Tenemaza et al (2020) and Fararni et al (2021).

4.3 Testing of Hypotheses

The following NULL hypotheses have been formulated and tested in this research work.

Hypothesis 1: There is a significant influence between distance of places considered in the tour and the average number of people visiting the cities in a tour package.

Hypothesis 2: There is no relationship between the cost of a tour package and the total number of people opted for the tour package.

To test: There is a significant influence between distance and cost metrics and the number of people opting for a given tour package.

Distance and Cost metrics	Weighted Mean using Historical Data (Observed)	Weighted Mean based on Theoretical Analysis (Expected)	(Observed value (O) – Expected Value) ² / Expected value (E)	Chi-Square value $\chi^2 = \sum \frac{(O - E)^2}{E}$	P-Value (For 6 degrees of freedom)
Knowledge about tourist places and the location of the places	7.69	4.65	0.65	4.12	0.66
Cost of Tour package	6.33	3.55	0.78		
Transport Availability	7.95	4.98	0.60		
Accommodation and food Quality	7.31	3.65	1.00		
Duration of the tour	7.52	4.72	0.59		
Climatic Condition	6.42	4.27	0.50		

Since the p-value is greater than 0.01, the null hypothesis is accepted at 1% level of significance. Therefore, there is a significant influence between the distance and cost metrics and the number of people selecting the tour package for the given places with the prescribed cost.

6. CONCLUSIONS AND FUTURE WORKS

In this paper, a new tourism destination recommendation system has been proposed for recommending optimal routes and locations using User preference trees, A* and TSP based heuristic search, LDA and Jaccard similarity analysis based location modeling, 0/1 knapsack problem based time and cost optimization. In this model, heterogeneous data sources that interpret both texts posted on social media and images of tourist places visited are used for analysis. A user-location vector is designed in this work to represent the relationships between user and places. The proposed user profile tree forms a graph that stores the probability with metadata of the user. This graph is used to retrieve a structure and 0/1 Knapsack algorithm is used to arrive at the optimal travel route for the user. The evaluation results obtained from this work have proved that this proposed tour recommendation system is providing more accurate recommendation of places and routes when it is compared with the other existing tour recommendation systems. This proposed system uses both text and tourist destination images and hence it is effective in both personalization and recommendation of new places. This system works totally based on the user's interests and the travel routes recommended are proved to be optimal compared to other recommendation systems that recommend based on only text or only images and heuristic search approach. As an extension, links can be drawn between users and their friends so that recommendation lists can be shared among friends. The time constraints of places can also be included in the form of intervals as a new factor while recommending travel routes to users.

REFERENCES

1. Adomavicius G, Tuzhilin A. 'Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions'. *IEEE Transactions on Knowledge and Data Engineering* 2005;17(6):734–49.
2. Akay, A., Dragomir, A., and Erlandsson, B. E. (2015), 'Network-Based Modeling and Intelligent Data Mining of Social Media for Improving Care', *IEEE Journal of Biomedical and Health Informatics*, vol. 19, pp. 210-218.

3. Berka T, Plönig M. 'Designing recommender systems for tourism'. In: *Proceedings of the 11th international conference on information technology in travel and tourism (ENTER'04)*; 2004.
4. Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003), 'Latent Dirichlet Allocation', *Journal of Machine Learning Research*, vol. 3, pp. 993–1022.
5. Buhalis, D., & Law, R. (2008). 'Progress in information technology and tourism management: 20 years on and 10 years after the Internet—the state of eTourism research'. *Tourism Management*, 29(4), 609–623.
6. Cao, L., Luo, J., Gallagher, A., Jin, X., Han, J., and Huang, T. (2010), 'A worldwide tourism recommendation system based on geo tagged web photos', *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 2274-2277.
7. Chen, X., Vorvoreanu, M., and Madhavan, K. (2014), 'Mining social media data for understanding students' learning experiences', *IEEE Transactions on Learning Technologies*, vol. 7, no. 3, pp. 246-259.
8. Cheng, Y. (1995). Mean Shift Mode Seeking and Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7), 790-799.
9. Chiang, H. S. and Huang, T. C. (2015), 'User-adapted travel planning system for personalized schedule recommendation', *International journal of Information Fusion*, vol. 21, pp. 3-17.
10. Esparza, S. G., O'Mahony, M. P., and Smyth, B. (2013), 'Catstream: categorising tweets for user profiling and stream filtering', *Proceedings of the International conference on Intelligent user interfaces ACM*, vol. 13, pp. 25-36.
11. Fesenmaier, D. R., Wo'ber, K., & Werthner, H. (2006). 'Introduction: recommendation systems in tourism. In D. R. Fesenmaier, K. Wo'ber, & H. Werthner (Eds.), *Destination recommendation systems: Behavioral foundations and applications*. Wallingford, UK: CABI.
12. Frey, B. J., and Dueck, D. (2007), 'Clustering by passing messages between data points. *Science*', vol. 315, pp. 972–976.

13. Gao, M., Liu, K., & Wu, Z. (2010). 'Personalisation in web computing and informatics: Theories, techniques, applications, and future research'. *Information Systems Frontiers*, 12, 607–629.
14. Gabriella Schoier* and Giuseppe Borruso DEAMS(2017), 'A methodology for dealing with spatial big data' *Int. J. Business Intelligence and Data Mining*, Vol. 12, No. 1, 2017
15. Gretzel, U. (2006). 'Consumer generated content – trends and implications for branding'. *e- Review of Tourism Research*, 4(3), 9–11.
16. Hernandez, M. A., Hildrum, K., Jain, P., Wagle, R., Alexe, B., Krishnamurthy, R., ... and Venkatramani, C. (2013), 'Constructing consumer profiles from social media data', *Proceedings of IEEE International Conference on Big Data*, pp. 710-716.
17. JiangningHea, HongyanLiua, HuiXiong,'SocoTraveler: Travel-package recommendations leveraging social influence of different relationship types', *Journal of Information & Management*, May 2016.
18. JonghyunHana ,Hyunju Lee b,'Adaptive landmark recommendations for travel planning: Personalizing and clustering landmarks using geo-tagged social media', *Journal of Pervasive and Mobile Computing*, Vol.18, Issue C,pp. 4–17, April 2015.
19. Kabassi K. 'Personalizing recommendations for tourists'. *Telematics and Informatics* 2010;27(1):51–66.
20. Kavitha, S., Jobi Vijay, and Rajeswari Sridhar. (2016), 'Fuzzy logic based tourism recommendation using social media profiles', *International Conference on Artificial Intelligence and Evolutionary Computations in Engineering Systems (ICAIECES 2016)*, 19-21 May, SRM College, Kattankulathur, Chennai.
21. Kavitha, S & Jobi, Vijay & Sridhar, Rajeswari. (2017). Tourism Recommendation Using Social Media Profiles. 10.1007/978-981-10-3174-8_22.
22. Kurashima, T., Iwata, T., Irie, G. and Fujimura, K. (2010), 'Travel route recommendation using geotags in photo sharing sites', *In Proceedings of the 19th ACM International Conference on Information and Knowledge Management, ACM*, Toronto, Canada, pp. 579–588.
23. Neri, F., Aliprandi, C., Capeci, F., Cuadros, M., and By, T. (2012). 'Sentiment analysis on social media'. *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining*, 919-926.

24. Ostrowski, and Alfred, D. (2015), 'Using latent dirichlet allocation for topic modelling in twitter', *Proceedings of IEEE International Conference on Semantic Computing (ICSC)*, pp. 493-497.
25. Parpinelli, R. S., Lopes, H. S. and Freitas, A. A. (2002), 'Data mining with an ant colony optimization algorithm'. *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 4, pp. 321–332.
26. Pan, B., MacLaurin, T., & Crofts, J. C. (2007). 'Travel blogs and their implications for destination marketing'. *Journal of Travel Research*, 46(1), 35–45.
27. Posner, M. (2015), 'Very basic strategies for interpreting results from the Topic Modeling Tool', available at:
<http://miriamposner.com/blog/very-basic-strategies-for-interpreting-results-from-the-topic-modeling-tool/>.
28. Ricci F. Travel recommender systems. *IEEE Intelligent Systems* 2002;17:55–7.
29. Ricci F, Rokach L, Shapira B, editors. 'Recommender systems handbook'. *Springer Science*; 2011, (<http://www.springer.com/computer/ai/book/978-0-387-85819-7>).
30. Rossetti, M., Stella, F., Cao, L., and Zanker, M. (2015), 'Analyzing user reviews in tourism with topic models', *In Information and Communication Technologies in Tourism, Springer*.
31. Sieg, A., Mobasher, B., & Burke, R. (2007). 'Learning ontology-based user profiles: A semantic approach to personalized web search. *IEEE Intelligent Informatics Bulletin*, 8(1), 7–18.
32. Sheldon, P. J. (1997). 'Tourism information technology'. *Oxon, UK: CAB International*.
33. Shen, J., Deng, C. and Gao, X. (2016), 'Attraction recommendation: Towards personalized tourism via collective intelligence', *International journal of Neurocomputing*, vol.173, pp. 789-798.
34. Staab S, Werthner H (2002) 'Intelligent systems for tourism introduction', *IEEE Intelligent Systems*, vol. 17, no.6, pp. 53–55.
35. Sun, Y., Fan, H., Bakillah, M., & Zipf, A. (2013), 'Road-based travel recommendation using geo-tagged images'. *International journal of Computers, Environment and Urban Systems*, vol. 53, pp. 110-122.

36. United Nations World Tourism Organization. (2014). 'UNWTO tourism highlights'. Madrid. available at: <http://www.e-unwto.org/doi/pdf/10.18111/9789284416226>
37. Werthner, H., & Klein, S. (1999). 'Information technology and tourism: A challenging relationship'. *Vienna: Springer*.
38. Werthner H, Ricci F. 'e-Commerce and tourism'. *Communications of the ACM* 2004;47(12):101–5.
39. Wu, D., Zhang, G., and Lu, J. (2015), 'A fuzzy preference tree-based recommender system for personalized business-to-business e-services', *Proceedings of IEEE Transactions on Fuzzy Systems*, vol. 23, no. 1, pp. 29-43.
40. Xiang, Z., Woßber, K., & Fesenmaier, D. R. (2008). 'Representation of the online tourism domain in search engines'. *Journal of Travel Research*, 47(2), 137–150.
41. YangBo, Peng-fei Zhao. "Survey of Recommendation Algorithm [J]," *Journal of Shanxi University (Nat.Sci.Ed.)*, 2011, 34(3):337-350.
42. Yamasaki, T., Gallagher, A. and Chen, T. (2013), 'Geotag-based travel route recommendation featuring seasonal and temporal popularity', *Proceedings of IEEE 9th International Conference on Information, Communications and Signal Processing (ICICS)*, pp. 1-4.
43. Zhang, M., Li, Y., Zhou, A. and Fang, Z. (2015), 'A Recommendation System for Travel Services Based on Cyber-Anima', *Proceedings of IEEE 12th International Conference on Business Engineering (ICEBE)*, pp. 114-118.
44. Zheng, V. W., Zheng, Y., Xie, X. and Yang, Q. (2010), 'Collaborative location and activity recommendations with GPS history data', *In: Proceedings of the 19th International Conferenc on World Wide Web, ACM*, North Carolina, USA, pp. 1029–1038.
45. Zhiwen Yu, Huang Xu, Zhe Yang, and Bin Guo, 'Personalized Travel Package With Multi- Point-of-Interest Recommendation Based on Crowdsourced User Footprints', *IEEE Transactions On Human-machine Systems*, Vol. 46, No. 1, pp. 151 - 158, February 2016.
46. Liang Di, Yongping Du. "Application of LDA Model in Microblog User Recommendation." *Computer Engineering*, 2014, 40(5): 1-6, 11.(In Chinese)

47. Sankar Pariserum Perumal, SannasiGanapathy, KannanArputharaj, "An Intelligent Fuzzy Rule-Based E-Learning Recommendation System for Dynamic User Interests", *The Journal of Supercomputing*, Vol. 75, No.8, pp. 5145-5160, 2019.
48. SankarPerumal, SannasiGanapathy, ArputharajKannan, "REFERS: refined and effective fuzzy e-commerce recommendation system", *International Journal of Business Intelligence and Data Mining*, Vol.17, No.1, pp. 117-137, 2020.
49. U Kanimozhi, S Ganapathy, D Manjula, A Kannan, "An Intelligent Risk Prediction System for Breast Cancer Using Fuzzy Temporal Rules", *National Academy Science Letters*, Vol. 42, No. 3, pp. 227-232, 2019.
50. L Sai Ramesh, SannasiGanapathy, R Bhuvaneshwari, KanagasabaiKulothungan, V Pandiyaraju, ArputharajKannan, "Prediction of User Interests for Providing Relevant Information Using Relevance Feedback and Re-Ranking", *International Journal of Intelligent Information Technologies (IJIT)*, Vol. 11, No. 4, pp. 55-71, 2015.
51. RV Karthik, SannasiGanapathy, ArputharajKannan, "A Recommendation System for Online Purchase Using Feature and Product Ranking", 2018 Eleventh International Conference on Contemporary Computing (IC3), pp. 1-6, 2018.
52. M. Tenemaza, S. Luján-Mora, A. De Antonio and J. Ramírez, "Improving Itinerary Recommendations for Tourists through Metaheuristic Algorithms: An Optimization Proposal", *IEEE Access*, vol. 8, pp. 79003-79023, 2020.
53. K. A. Fararni, F. Nafis, B. Aghoutane, A. Yahyaouy, J. Riffi and A. Sabri, "Hybrid recommender system for tourism based on big data and AI: A conceptual framework", in *Big Data Mining and Analytics*, vol. 4, no. 1, pp. 47-55, March 2021.
54. K. Kesorn, W. Juraphanthong and A. Salaiwarakul, "Personalized Attraction Recommendation System for Tourists through Check-In Data", *IEEE Access*, Vol. 5, pp. 26703-26721, 2017.