Abstract

Objective: The main objective of this research is to generate and provide an optimized itinerary planning for the users with flight and hotel booking as per the user requirement to reach their satisfaction level. Methods: In this research, a novel fuzzy c means clustering is introduced for grouping and generating an optimal itinerary planning for the group of users. This work intends to satisfy the needs of all the users who are requesting for the itinerary plans with the satisfaction of their QoS requirement. To achieve this goal, the point of interest (POI) graphs will be constructed with the help of mapReduce application which will simplify the processing of large amount of data’s. FCM algorithm is proved to be better clustering algorithm which considers the membership values among the various data points. So that the same data points can be clustered in the different cluster to which its membership value belongs. In our work we make sue use of this algorithm to cluster the users with the similar requirements. So that the same user request may clustered in different groups. Among that the optimized cluster of users with unique requirements can be selected for the itinerary planning construction. Results: FCM clustering used in the proposed approach for effective construction of the itinerary planning with the consideration of the users with similar requirement whereas in the existing work user similarity requirements are not considered. The experimental tests conducted were proves that the proposed methodology provides the optimized itinerary with the less time complexity and burden than the existing methodology. This performance analysis is done with the two performance metrics namely processing time and the weight ratio from which it can be proved that the proposed methodology provides better result. Conclusion: The findings demonstrate that the itinerary planning construction using FCM is presented and this method has high clustering accuracy than the previous methodologies.

Keywords: Fuzzy C-means Clustering, Itinerary Plan, MapReduce, Point of Interest

1. Introduction

Itinerary planning is a most time consuming process in the real world which need to taken under consideration more. There are many online services are available which provide only the packaged itinerary planning to the users with their fixed cost and price. But it is having more limitations like, it cannot satisfy the user’s needs and their budget and also it will not give timely help to the users.

The another drawback present in this traditional itinerary planning construction is the involvement of the agents who will acts as an intermediate between the users and the travelling agency to build the itinerary planning. The security need to be eliminated in order to provide the privacy to the users. The automatic construction of the itinerary planning will eliminate the need of users to reveal their personal interest to the unknown users.

Transportation handling is the one of the biggest issue in the today world which may cause the Indian economics\(^1\). The transportation systems need to be flexible in the real time environment and also the configurations should change as per the user
requirements dynamically. The transportation is the main object which needs to be concerned more in our work for obtaining the better optimized k-day itinerary plan. The passenger segmentation is also take important place in the transportation which needs to be done effectively for obtaining effective itinerary plan. The passenger segmentation should be done based on the type of transportation arranged and the user requirements. Electronic vehicles need to be considered for the optimized itinerary planning through which the many places can be covered effectively because of the time savings. The transportation management system for the electronic vehicles is the most complex process that needs to be considered in the future work for the better optimized itinerary plan construction.

The techniques proposed in this work mainly concentrate on providing the better itinerary planning for the users with the satisfaction of the constraints mentioned by the users. To achieve this goal, the points of interest to be visited by the user and the time and budget details will be gathered from them before preparing the plan. The budget and the time constraint will also collected from the users in order to generate the user friendly plan. The generation of itinerary planning is automated in this work by introducing the novel concepts. And also to improve the accuracy of the travelling plan, in this work, Fuzzy C means clustering algorithm is introduced in this work. A novel approach introduced in this work will leads to an optimized construction of an itinerary planning with the customer Satisfaction level. The main contributions of this works are,

- Construction of the better itinerary planning with the consideration of user limited budget and the time constraints.
- Construct the itinerary planning for the multi users in single time based on the similar behavior of users instead of constructing the multiple same itinerary plans.
- Grouping the user’s point of interests with similar behavior.
- Provide a better itinerary planning with the facility of booking the flights and hotels also as per the user requirement of budget and time limitation.

Senjuti Basu Roy et al. discussed the automatic creation of itinerary planning in an interactive way. This work aims to reduce the users burden by providing the previous user feed back visible to the users from which the most visited places is enough to look by the current users instead of reviewing the all the points of interest. This process will be repeated by displaying the previous results to the users until the user satisfaction level is reached. To enable this process, the user feedback model is implemented in this work which intends to gathers the feedback of every user who enters into the network. And also in this work, itinerary scoring semantic is introduced which is used to rank the itineraries utilized by every users in order to generate and provide an optimal itinerary planning to the users.

Barun Chandra et al. discussed a various problems occurred during the set packing algorithm. The set packing algorithm intends to group the objects into one for satisfying the user requirement. However there will be many issues occur when the objects from the different groups are packed together. The main objective of this work is to find the sub collection of more objects with maximum weights from the finite group of available objects. To achieve this goal, this research work introduces a k-set packing problem which intends to pack k objects together. This work starts with initial object taken into the set based on the greedy solution and the further objects are considered based on the local search methodology. This work can be used effectively in all the works which are all based on the itinerary planning construction.

Munmun De Choudhury et al. discussed an automatic construction of an itinerary planning based on social crumbs. In this work, itinerary planning is constructed by leveraging the personal interests of user by analysing the geo temporal bread crumbs. This is done by extracting the details of time stamps and the image contents from the available source information and grouping them to find out the most wanted point of interest of many users. To achieve this one must gather the photos of the popular cities and the time path need to be constructed. Based on this time paths, the optimized itinerary planning can be constructed.

Hyoseok Yoon et al. proposed a novel mechanism for itinerary plan construction based on the user generated GPS trajectories. This work is mainly used to help the unfamiliar people to some places where they like to go for vacation trip. This work mainly aims to find out the start point and the end point of allocation with the minimized cost and the travelling. This is done by gathering the multiple users location details based on the GPS tracking information. When the users are submitting their queries, the query will send to the cloud server where the multiple user location information will be present. The point of interest submitted by the users will be matched with the GPS tracked information and the optimized itinerary planning will be constructed.
Maarten Clements et al. proposed details of discussed an approach for monitoring the user behaviour by predicting the user travelling. In this work, a novel approach is introduced to predict the user interest by analysing the user tagged location based images. In this work, flickr tagged images are downloaded for the purpose of predicting user travelling behaviours. After finding the users point of interest, those extracted places will be ranked further to enable the users to select the particular places that he wants to visit within a particular time period.

Chao Chen et al. discussed a novel way to construct an itinerary planning by using the location based information and the GPS traces. GPS traces are used to identify the crowd sourced information's that is the location which is mostly preferred by the people to visit often. To achieve this, in this work, the heuristic algorithms are effectively utilized which focus to gather the user preferred locations and interest scores.

However all these works described above did not achieve the user satisfaction level. And also the method discussed in the related studies doesn't try to satisfy the user requests in the fine grained manner. The fine grained itinerary planning construction for the group of users are implemented in our work which tries to construct the itinerary planning with the user satisfaction level. The work proposed is discussed in the following sections.

2. Automatic Itinerary Plan Construction

The main goal of this research is to ease the work of travelling agents and improve the user satisfaction level by constructing the itinerary planning as per the user requirements. The construction of optimized itinerary planning with the points of interests that are user likes to visit within the users specified period and cost will be more difficult one. In this work, it is achieved by gathering the user requirement first and then the $k^{th}$ day itinerary planning is constructed.

The multi user environment is created in this work which is used to reduce the computational complexity. In the previous researches, the itinerary planning is constructed by considering only one user requirement whereas in this work, itinerary planning is constructed by considering multiple user requirement. The requirements from the multi users will be gathered and then the itinerary plan will be constructed for group of users who is having most similar requirement. It is achieved by proposing a methodology called fuzzy c means clustering.

The itinerary plan construction consists of the following steps:

- Collecting the Points Of Interest (POI) from the different users
- Group the users with similar POI in single cluster using fuzzy c means clustering approach
- Construct the single day itinerary using the mapReduce programming
- Book the flight and hotels for the group of users

By following these steps, the better itinerary planning can be constructed with the constraints of reduced cost and time as per the user requirements.

2.1 Collecting the Point Of Interest (POI) from the Different users

The optimized itinerary planning with the satisfaction of users can be constructed by gathering the points of interests that are user wants to visit. After gathering user interests, those requests will send to the server for the further proceedings. Along with this information, the budget information will also gather from the users. Based on the cost level, the flight and hotel booking is also done.

In the server, the details of network traffic and road information will be available. These details will be collected from the various sources and that will be stored in the database for the further proceedings. In our work, data set which consists of Road traffic information's are gathered from the Google API. This data set consists of a various location information and as well as the distance among the various locations.

In this work, mapReduce programming concept is used to process the millions of details about the points of interests. In the real environment, it will be more difficult to handle the millions of traffic details about the tourist places. The details gathered from the Google API is used to build the POI graph which make ease of users to

2.2 Group the users with Similar POI in Single Cluster using Fuzzy C Means Clustering (FCM) Approach

Initially, the user with the similar requirement is grouped to achieve the better vacation. Fuzzy C means clustering is introduced in our work to group the users with the similar requirement. To achieve this, the membership
value will be assigned to every user present in the network, so that users can be clustered together based on those values. This clustering mechanism will make ease of booking hotel and flight ticket process in the composite manner.

The FCM clustering algorithm is based on the distance between the data present in the environment. In our work, data is a POI represented by the users. This algorithm intends to form the cluster by fixing the cluster center initially. The data will be added into the corresponding cluster, if the distance between the data and the cluster center is minimal. Hence FCM is a one of the data clustering approach in which the given input data will be clustered in to n level. The data to be added in the particular cluster will have high degree of connection between them. If the data is added to the other cluster then it is said to be data is having low degree of connection with the present cluster.

2.2.1 Algorithm of Fuzzy c-Means Clustering

FCM clustering mechanism is a method which allows one data to be present in the more than two clusters if it is having more similarity. In this research, the similarity levels of users are calculated based on the similar points of interest and their budget level. This is done to reduce the complexity level of constructing individual itinerary plan for every user. The clustering of users with same budget and points of interests can reduce the burden of booking separate flight and hotel booking. The FCM is used to efficiently cluster the users with same requirement. It is based on minimization of the following objective function

\[ J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \| x_i - c_j \|^m, 1 \leq m < \infty \]

Where \( m \) - any real number greater than 1
\( u_{ij} \) - membership degree of \( x_i \) in the cluster \( j \)
\( x_i \) - \( i \)th of \( d \)-dimensional measured data
\( c_j \) - \( d \)-dimension centre of the cluster
\( \| \cdot \|^m \) - any norm expressing the similarity between any measured data and the center.

Through an iterative optimization of the objective function Fuzzy partitioning is carried out, with the update of membership \( u_{ij} \) and the cluster centres \( c_j \) by

\[ u_{ij} = \frac{1}{\sum_{k=1}^{C} \left( \frac{\| x_i - c_k \|^m}{\| x_i - c_j \|^m} \right)^{\frac{1}{m-1}}} \]

\[ c_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} u_{ij}^m} \]

This iteration will stop when

\[ \max_i \{ u_{ij}^{(k+1)} - u_{ij}^{(k)} \} < \varepsilon \]

Where \( \varepsilon \), a termination criterion between 0 and 1 and \( k \) is the iteration steps. This practice converges to a local minimum or a saddle point of \( J_m \).

The algorithm has the following steps:
- Initialize \( U = [u_{ij}] \) matrix, \( U^{(0)} \)
- At \( k \)-step: calculate the centers vectors

\[ U^{(k)} = \left[ u_{ij}^{(k)} \right] \text{ with } U^{(k)} c_j = \sum_{i=1}^{N} u_{ij}^{(k)} x_i \]

- Update, \( U^{(k)} \), \( U^{(k+1)} \) follows.

\[ U_{ij}^{(k+1)} = \frac{1}{\sum_{k=1}^{C} \left( x_i - c_k \right)^{2}} \]

- If \( \| U^{(k)} - U^{(k+1)} \| < \varepsilon \)

Then STOP;

Otherwise return to step 2.

The above mentioned FCM algorithm intends to move the cluster centers to the right location within a data. To be specific introducing the fuzzy logic in K-Means clustering algorithm is the Fuzzy C-Means algorithm in general. In fact, FCM clustering techniques are based on fuzzy behaviour and they provide a technique which is natural for producing a clustering where membership weights have a natural interpretation but not probabilistic at all. This algorithm is basically similar in structure to K-Means algorithm and it also behaves in a similar fashion. In the FCM approach, instead, the same given datum does not belong exclusively to a well defined cluster, but it can be placed in a middle way. In this case, the membership function follows a flatter line to designate that every datum might go to numerous clusters with dissimilar standards of the membership constant.

After clustering the multiple users based on their request, novel itinerary planning approach is proposed. In our work, set packing problem is applied to the set of POI and the optimal itinerary planning is constructed. Initially POI graph will be constructed by using the road network and the POI co-ordinates. The details of road traffic information are gathered from the Google API. The dataset gathered from Google API consists of information about the distance among the locations and list of most interested places. And then the approximation
algorithm is used in this work to simplify the concepts where the maximum number of POI is bounded within single itinerary.

2.3 Construct the Single Day Itinerary using the MapReduce Programming

After gathering the points of interests from multiple users, the itinerary planning needs to be constructed. This is done by using the mapReduce programming concept. In this work, the possible single day itinerary will be constructed by using the available mapReduce jobs at the initialization level. And in the next stage, the POIs will be added to the initialized itinerary to construct the optimized single day itinerary. The POI will be added to the available itinerary only if it is feasible to meet the places within a user specified cost and budget. In this work we tried to allow the m mapReduce jobs to bein the single itinerary. That is in the single day itinerary we tried to add m POI to improve the efficiency within the minimal overhead. But in this approach we should also consider the constraint of user who cannot visit many numbers POI in single. Hence the constructed itinerary should be feasible to the multi-users.

The functioning of mapReduce is explained as follows: The mapper will take the input as partial path from the distributed file system and then it will leverage to add the sub additional POI to it. When the new POI is added to the available constructed itinerary the condition of feasibility should be satisfied. That is the new POI to be added should be feasible to visit by the user. This is tested by analyzing the real network traffic data gathered. It will compare the user POI with the POI present in the single day itinerary based on the time and traffic available in the particular route. From those data, the decision will be computed that whether the newly added POI can be included in the itinerary or not.

Like wise many single day itineraries will be generated for the same POI of multi-users. After completion if this process, the single day itinerary will be selected based on the reduced cost and time by the users. The decision of choosing the itinerary will be given to the user hand. For grouping the single day itinerary into the group, the itinerary index concept is introduced. The itinerary index is used to sort the itineraries based on their importance level. The itinerary with more weight will be given more important than other itineraries.

For each and every user, the weight value for the set of POI will be updated and then score will be computed for each single itinerary. This is done to construct the k day itinerary with more score by selecting the single day itinerary with more score. This is achieved in our work by using set packing approach which is based on the polynomial approximate algorithms. The set packing problem makes use of greedy strategy to generate an initial solutions and further refinement is done by adjusting the values. In the adjustment procedure itinerary index is varied to improve the solution.

The work proposed in this research has the benefits of reduced search complexity as well as the time taken to generate the itinerary. The complexity level of the itinerary generation is reduced considerably by grouping the multi-users with similar requirement. In addition to that we are adding the service of flight booking which will also improve the multi-user satisfaction level. Additionally, we are considering the service like flight booking. This will improve the efficiency of our novel system. Our proposed system is well effective than the existing system in terms of search complexity and time complexity. The performance of the system is well improved in this system.

3. Experimental Results

To evaluate the performance, the dataset is gathered from the Google which consists of a POI constructed by the agencies with the agency id. And also the data set related to hotel details and the flight ticket availability is gathered. The experimental tests were conducted with the help of hadoop environment. The automatic itinerary plan constructions were by using the hadoop toolkit. Apache Hadoop is an open-source software framework for distributed storage and distributed processing of Big Data on clusters of commodity hardware. The performance evaluation is done by comparing the proposed methodology itinerary plan construction with FCM (ICPFCM) with the existing methodology MR-set ICP (MRICP) based on the performance metrics called the processing time and the weight ratio.

3.1 Processing Time

The processing time is the time taken to construct the k day itinerary plan. The less processing time will lead to amore user satisfaction level and will improve the efficiency. The processing time consumed for the number of POI for both the existing and the proposed methodology in shown in the Table 1.
A Novel Multi-user Itinerary Planning for Traveling Services

Table 1. Processing Time

<table>
<thead>
<tr>
<th>Number of selected POI</th>
<th>Processing Time in ms</th>
<th>MRICP</th>
<th>ICPFDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>22</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>28</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>33</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>37</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>250</td>
<td>41</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>46</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>350</td>
<td>49</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>400</td>
<td>54</td>
<td>42</td>
<td></td>
</tr>
</tbody>
</table>

The graphical result for the comparison of proposed methodology with the existing method is shown in the following Figure 1.

Figure 1. Processing Time.

The above graph proves the proposed methodology constructs an itinerary planning better than the existing methodology. In the x axis Total number of POI is taken and in the y axis processing time is taken. The comparison proves that the existing methodology consumes more processing time and it cannot process well more than particular amount of POI whereas the proposed methodology works efficiently.

3.2 Weight Ratio

The weight ratio is defined as the quality of the generated itineraries based on its efficiency. The weight is calculated by dividing the total weights of both proposed methodology and the existing methodology. The weight ratio will increase if the proposed methodology is better than the existing methodology. The weight ratio is calculated as follows

\[
\max_{i,j} \left\{ \left| \mathbf{u}_{ij}^{(k+1)} - \mathbf{u}_{ij}^{(k)} \right| \right\} < \varepsilon
\]

The real values obtained while executing the proposed method and the existing method is shown in the Table 2.

Table 2. Weight

<table>
<thead>
<tr>
<th>Number of selected POI</th>
<th>MRICP</th>
<th>ICPFDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>20</td>
<td>32</td>
</tr>
<tr>
<td>100</td>
<td>32</td>
<td>54</td>
</tr>
<tr>
<td>150</td>
<td>48</td>
<td>90</td>
</tr>
<tr>
<td>200</td>
<td>65</td>
<td>125</td>
</tr>
<tr>
<td>250</td>
<td>95</td>
<td>188</td>
</tr>
<tr>
<td>300</td>
<td>105</td>
<td>216</td>
</tr>
<tr>
<td>350</td>
<td>110</td>
<td>235</td>
</tr>
<tr>
<td>400</td>
<td>125</td>
<td>256</td>
</tr>
</tbody>
</table>

The graphical representation of the result obtained for the weight ratio is shown in the Figure 2.

Figure 2. Weight Ratio.

From the graph (Figure 2) it can be concluded that the quality of the itinerary plan is increased when grouping of multi-users is done based on similarity level. In the above graph, the total number of POI is taken in the x axis and the weight ratio is taken in the y axis. The comparison proves that the itinerary construction with FCM is better than existing method.

4. Conclusion

In this work, an automatic generation of itinerary planning is done for the user who wants to go to the vacation trip. In our work, the service is done for creating the multiple day itineraries based on the multiple user preferences. In order to handle the multiple users in an effective way, in this work fuzzy c means clustering mechanism is introduced. The fuzzy c means clustering mechanism is used to cluster the users with the similar
requirements. Membership values very effective method for clustering the multi-users. We are well reducing the search complexity as well as time complexity of the itinerary generation system by using the grouping the users according to user’s requirement similarity. This will show the efficiency of our novel system. Our proposed system is well effective than the existing system in terms of search complexity and time complexity. The performance of the system is well improved in this system.

5. References