ISSN 2007-9737

PIONEER: An Interest-Aware POI Recommendation Engine

Sanjeev K. Cowlessur*,[1], Annappa Basava[2], Bibudhendu Pati[3]

[1] Université des Mascareignes, Pamplemousses, Mauritius

[2] National Institute of Technology Karnataka, Surathkal, India

[3] Rama Devi Women's University, Bhubaneswar, India

scowlessur@adm.ac.mu

Abstract. Over the past decades, tourism has become a key economic industry for many countries. In today's global economy, it is an essential source of employment and revenue. Tourism as a leisure activity is a very popular form of recreation which involves the movement of people to foreign cities to visit new and unfamiliar places of interest (POIs). The task of recommending personalised tours for tourists is very demanding and time-consuming. The recommended tours must satisfy the tourist's interests and must at the same time be completed within a limited time span and within some budget. In existing itinerary recommender systems, if there is no past visit history about a particular POI, then that POI is not included in the recommended itinerary. To address this challenge, we have devised an algorithm called PIONEER which is based on a genetic algorithm for suggesting an itinerary based on tourist interests, POI popularity, and travel costs. Our algorithm recommends itineraries for tourists who want to visit locations which are unfamiliar to them. We have used the publicly available Flickr dataset in our work. The results demonstrate the superiority of our PIONEER algorithm compared to the baseline algorithms with regards to metrics like precision, recall and F1-Score.

Keywords. POI, tour recommendation, NSGA-II, multi-objective optimisation.

1 Introduction

Planning a visit to a foreign city can be a very daunting task [1]. The tourist needs to identify interesting POIs and then plan his/her visits as a connected itinerary while taking into account various spatial and temporal constraints. There are a number of factors which affect a tourist's decision and choice of visiting a particular POI [24].

Some of the factors are internal, that is, personal to the tourist, for example, age, education, occupation, income or his prior travel experiences and some are external meaning that they do not depend on the tourist, for example, climate and reviews from other travellers [9, 19].

In this work, we propose a recommendation engine for tourists which provides the most relevant suggestions for POI visits keeping in mind tourists' interests, popularity of POIs and travelling cost [22]. The remainder of this paper is organised as follows. Section 2 reviews and discusses a few relevant research work undertaken in the area. Then, section 3 provides definitions of necessary concepts.

ISSN 2007-9737

In the following section 4, the problem is defined and the proposed new algorithm is explained in section 5. Then follows section 6 which contains a discussion on the various experiments conducted and the results obtained. Conclusions, ideas and suggestions for future research in the area under study is given in section 7.

2 Review of Related Work

Recently, tour recommendation has become a popular subject of interest among researchers [2]. Several applications [4, 18, 29, 30] have been built to deliver personalised tours.

2.1 The Orienteering Problem

Many tour recommender systems have as their starting point the Orienteering Problem [10, 21]. The idea orienteering problem came from a sports game which consisted of a number of checkpoints each having an associated score. Each player had to start at a given checkpoint, with a view to visit as many checkpoints as possible to accummulate scores.

The player who obtained the largest score in the smallest possible time was declared the winner. One constraint imposed was that each checkpoint had to be visited at most once. However, it was not mandatory for the player to start and end at the same point. In the past years, many researchers have been using the orienteering problem [11, 28] in their tour recommendation works.

2.2 Tour Recommendations based on the Orienteering Problem

In their paper, Choudhury et al. [7] proposed a tour itinerary based on the orienteering approach, in which the tourist begins the tour at some POI and finishes the at some other POI, where the goal was to recommend an itinerary comprising the most popular POIs, all within a given budget. Lim et al. [15] brought modifications to the orienteering problem by ensuring that the tourist visits one POI catergory he/she is interested in. Vansteenwegen et al. [27] proposed an approach for adapting the tour schedule so that it would improve the overall balance between the defined degree of involvement from the starting and end, such as expenditure and all POIs. Lim et al. [17] have identified places to visit which require minimum queuing time. Algorithms have also been developed to recommend tours for groups of tourists which satisfy the different levels of interest of each tourist within the group [1, 17].

2.3 Other Tourism Related Work

The wealth of information available in geo-tagged photos can be used to understand tourists' behaviour and to find out how popular a given POI is. Ji et al. [12] use a graphical model to evaluate the popularity of a POI by using photos which have been uploaded to websites.

The amount of time spent by a tourist at a given POI and in which order he/she visits the POIs can be extracted from the images data [20]. The authors in [6] have used geo-tagged photos to find out the location of clusters where popular and interesting activities are taking place.

In their paper Li et al. [14] were able determine the approximate location of photos [13]. A time aware measurement technique which considers a tourist's current location to recommend the next POI to visit was proposed by Ying et al. [31].

The $\mathrm{PIONEER}$ algorithm suggested in this paper differs significantly from the current POI and tour recommendation schemes in that this algorithm uses geo-tagged images to categorise the interests of visitors dynamically depending on time spent at a POI and its popularity.

3 Background

A tourist travelling to a any city across the globe will certainly be looking to visit Points Of Interest (POIs). We can think of a POI as a place which a person finds useful or interesting. Suppose there are n POIs in a given city denoted by: $p_1, p_2, p_3, \ldots, p_n$. Suppose each POI p_i belongs to a category $c_{p_i} \in \mathbb{C}$ associated with it, where $\mathbb{C}$ is the set of all categories of POIs (some examples of POI categories are: parks, museums, shops, restaurants).

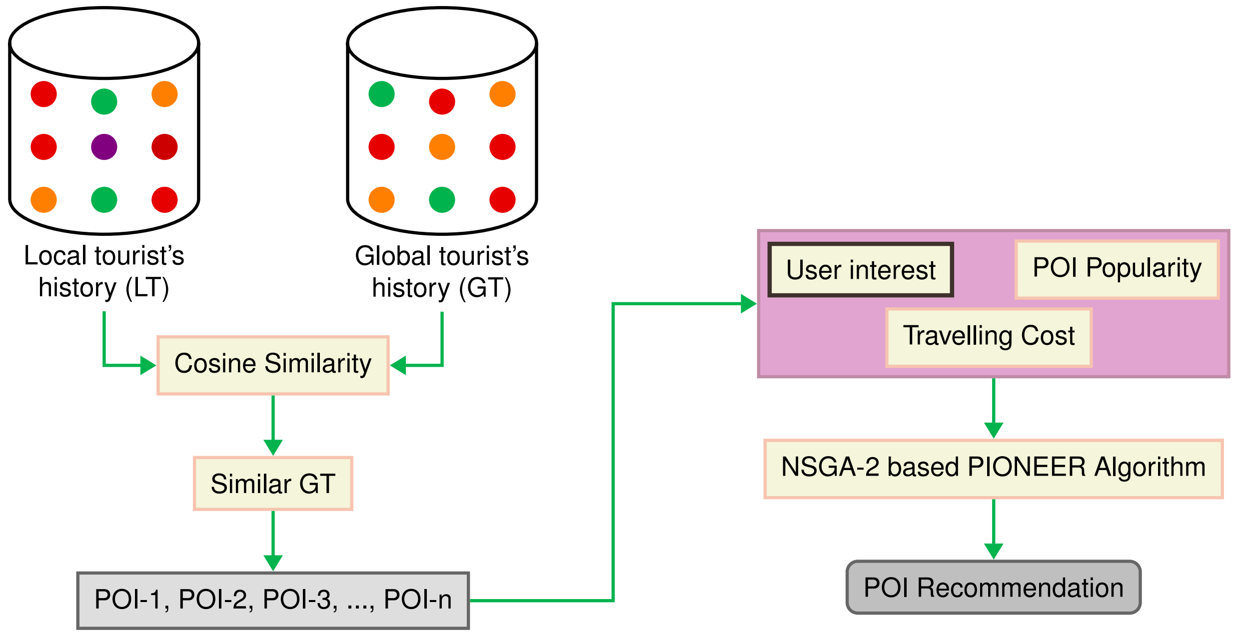

Fig. 1. System framework

3.1 Local and Global Tourists

In this study, tourists are classified into two different categories - local tourists and global tourists. A tourist who has already visited a certain city is referred to as a local tourist with respect to that city. If this tourist now travels to a different city which he has not previously visited, the local tourist of that city becomes a global tourist with respect to the travelling tourist.

To make the concepts clearer, consider two tourists T_1 and T_2 where T_1 has visited Mauritius but not Bangalore and T_2 who has visited Bangalore but not Mauritius. Then T_1 is a local tourist for Mauritius and T_2 is a local tourist for Bangalore. Suppose now that tourist T_1 wants to visit Bangalore. Then T_2 becomes the global tourist for T_1. Similarly, if T_2 flies to Mauritius then T_1 becomes the global tourist for T_2.

3.2 Travelling History of a Tourist

Let $\mathbb{U}$ be a set of tourists. Suppose there is some tourist $u \in \mathbb{U}$ who has so far visited k POIs. Then, the travelling history of u is given by a sequence of triples $\mathbb{H}_u = \big((p_1, t^a_{p_1}, t^d_{p_1}), \dots, (p_k, t^a_{p_k}, t^d_{p_k})\big)$.

In the triple $(p_\iota, t^a_{p_\iota}, t^d_{p_\iota})$, p_ι is the POI visited by the tourist, $t^a_{p_\iota}$ is the time of arrival at POI p_ι and $t^d_{p_\iota}$ is the time of departure from p_ι. The difference between $t^a_{p_\iota}$ and $t^d_{p_\iota}$ gives the amount of time spent at p_ι. To make the notation simpler, we will write $\mathbb{H}_u = (p_1, \dots, p_k)$ instead of $\mathbb{H}_u = \big((p_1, t^a_{p_1}, t^d_{p_1}), \dots, (p_k, t^a_{p_k}, t^d_{p_k})\big)$.

3.3 Travelling Sequences of a Tourist

Given some tourist, his/her travel history is broken down into several distinct travel sequences if the time difference between two consecutive POI visits is t_{seq} hours or more. In our work we use $t_{\text{seq}} = 8$ as proposed by Lim in [16]. So, the travel history $\mathbb{H}^1_u$ can be written as $\mathbb{H}^1_u, \mathbb{H}^2_u, \dots, \mathbb{H}^k_u$, where k is the number of travel sequences.

3.4 Average Time Spent at $a = \mathrm{POI}$

For every tourist, the history of his/her past travels is known. Given this information, the equation 1 can be used to calculate the mean time spent by all tourists who have visited a specific POI p [3, 5]. This value is denoted by $A(p)$:

ISSN 2007-9737

Algorithm 1: PIONEER Algorithm

Data: u
Result: $(p_1, p_2, \ldots, p_n)$
$\mathcal{I} \leftarrow \emptyset$
for u in $\mathrm{matching\_global\_tourist\_list}$ **do**
  $\mathcal{I} \leftarrow \mathcal{I} \cup \mathcal{I}_u$
$\mathcal{P}^{\text{initial}} \leftarrow \mathcal{I}$
$\mathcal{F}_1 \leftarrow \mathrm{Objective\_1}(\mathcal{P}^{\text{initial}})$
$\mathcal{F}_2 \leftarrow \mathrm{Objective\_2}(\mathcal{P}^{\text{initial}})$
$\mathcal{R} \leftarrow \mathrm{non\_dominated\_sorting}(\mathcal{F}_1, \mathcal{F}_2, \mathcal{P}^{\text{initial}})$
$\mathcal{D}_{\text{crowd}} \leftarrow \mathrm{find\_crowding\_distance}()$
$\mathcal{P}^{\text{sol}} \leftarrow \mathrm{select\_initial\_pop}(\mathcal{R}, \mathcal{D}_{\text{crowd}})$
while termination_condition_not_reached **do**
  $\mathcal{C}^{\text{sol}} \leftarrow \mathrm{gen\_child\_pop}(\mathcal{P}^{\text{sol}})$
  $\mathcal{F}_1 \leftarrow \mathrm{Objective\_1}(\mathcal{C}^{\text{sol}})$
  $\mathcal{F}_2 \leftarrow \mathrm{Objective\_2}(\mathcal{C}^{\text{sol}})$
  $\mathcal{R} \leftarrow \mathrm{non\_dominated\_sorting}(\mathcal{F}_1, \mathcal{F}_2, \mathcal{P}^{\text{sol}} \cup \mathcal{C}^{\text{sol}})$
  $\mathcal{D}_{\text{crowd}} \leftarrow \mathrm{find\_crowding\_distance}()$
  $\mathcal{P}^{\text{sol}} \leftarrow \mathrm{select\_next\_gen}(\mathcal{R}, \mathcal{D}_{\text{crowd}})$
return $\mathcal{P}^{\text{sol}}$

$$A(p) = \frac{\sum_{u=1}^{q}\sum_{\imath=1}^{r}(t^{d}_{p_\imath} - t^{a}_{p_\imath})\delta(p_\imath = p)}{\sum_{u=1}^{q} V_u \, \delta(p_\imath = p)}, \quad \forall p \in \mathcal{P}, \quad (1)$$

where $u = \{1, 2, \ldots, q\}$, $\jmath = \{1, 2, \ldots, r\}$, V_u is the frequency of tourist's u visit to POI p and $\delta(p_\imath = p) = 1$ if $p_\imath = p$ and 0 otherwise.

3.5 Tourist Interest for POI Category

Recall that the symbol $\mathbb{C}$ has been used to denote the set of categories of POIs and $c_p \in \mathbb{C}$ to denote the category of a POI p. Then, the interest a particular tourist u has for a particular category c of POI can be calculated using equation 2:

$$\mathrm{Int}_u(c) = \sum_{\jmath=1}^{n} \frac{(t^{d}_{p_\jmath} - t^{a}_{p_\jmath})}{A(p_\jmath)} \delta(c_{p_\jmath} = c), \quad \forall c_p \in \mathbb{C}, \quad (2)$$

where $\delta(c_{p_\jmath}) = 1$ if $c_{p_\jmath} = c$ and 0 otherwise. The tourist interest for the POI category c is obtained from equation 1 by calculating the time spent by a tourist u at POI category c relative to the total time spent by all the tourists. It makes sense that a tourist will stay for a longer period at a POI category in which he/she is most interested in.

3.6 Local and Global Tourist Similarity

The degree of similarity between local and global tourists can be determined based on their interests for a given destination. For two distinct tourists u_x and u_y, we can compute their similarity using the cosine similarity measure as shown in equation 3:

$$\mathbb{S}(u_x, u_y) = \frac{\vec{\mathrm{Int}}_{u_x} \cdot \vec{\mathrm{Int}}_{u_y}}{||\vec{\mathrm{Int}}_{u_x} \cdot \vec{\mathrm{Int}}_{u_y}||}. \quad (3)$$

3.7 Tourist Interest for a POI

The interest a particular tourist u has for a particular POI p can be determined using the equation 4:

$$\mathrm{Int}_u(p) = \sum_{i=1}^{n} \frac{(t^{a}_{p_i} - t^{d}_{p_i})\delta(p_i = p)}{A(p_i)\delta(p_i = p)}, \quad (4)$$

where $A(p)$ (see equation 2) is the average time spent by all tourists at POI p and $\delta(p_i = p) = 1$ if $p_i = p$ and 0 otherwise.

3.8 Popularity of a POI

Every POI p has a certain popularity associated with it which is denoted by $\mathrm{Pop}(p)$. The popularity of a POI is taken to be the number of times the POI has been visited by all tourists. More formally, $\mathrm{Pop}(p)$ is defined by:

$$\mathrm{Pop}(p) = \sum_{u \in U} \Phi_{u,\, p}, \quad (5)$$

where U is the set of all tourists and $\Phi_{u,\, p}$ is the number of times tourist u has visited POI p.

3.9 Travelling Cost

There is a cost involved while travelling from one POI to another. In previous studies the cost of travel from one POI p_i to another POI p_j was a measure of the time taken by the tourist to complete the trip from p_i to p_j.

The total cost was considered to be the total time taken for an entire tour. The problem with using time as a measure of travel cost is that travel time depends on the means of transport used.

Table 1. Comparison of $\mathrm{Precision}$ between our proposed method and other baseline algorithms

Algorithms	PIONEER	TRIC	GREEPOP	GREENEAR	RAND
Delhi- Edinburgh	**0.589**$\pm$**0.027**	0.512$\pm$0.018	0.462$\pm$0.029	0.432$\pm$0.014	0.392$\pm$0.019
Osaka-Edinburgh	**0.613**$\pm$**0.025**	0.562$\pm$0.011	0.521$\pm$0.023	0.483$\pm$0.029	0.452$\pm$0.043
Vienna-Edinburgh	**0.692**$\pm$**0.013**	0.610$\pm$0.042	0.582$\pm$0.031	0.554$\pm$0.008	0.535$\pm$0.024
Delhi-Osaka	**0.593**$\pm$**0.029**	0.546$\pm$0.021	0.416$\pm$0.015	0.396$\pm$0.027	0.371$\pm$0.007
Glasgow-Edinburgh	**0.406**$\pm$**0.013**	0.336$\pm$0.029	0.307$\pm$0.006	0.281$\pm$0.035	0.263$\pm$0.014

For example if two POIs are very far apart and the tourist decides to take a flight, then the travel time will be much lower. In this paper, the total distance travelled by the tourist has been used as the travel cost and the aim is to minimise the distance travelled in an itinerary. The total cost of an itinerary $I = (p_1, p_2, \ldots, p_N)$ with N POIs is given by:

$$\mathrm{Cost}(I) = \sum_{\iota=1}^{N-1} \mathrm{Dist}(p_\iota,\ p_{\iota+1}), \tag{6}$$

where $\mathrm{Dist}(p_\iota,\ p_j)$ is the distance between POIs p_ι and p_j which can be calculated using the Haversine formula [25].

4 Problem Definition

The main objective of this work is to suggest an itinerary $I_u = (p_1, \ldots, p_n)$ for a tourist u such that the interests of the tourists and popularity of POIs visited are maximized but at the same time the cost of travel is minimised. This leads to the following optimisation problem [16]:

$$\mathbb{P}(I) = \sum_{i=1}^{n} \alpha\, \mathrm{Pop}(p_i) + (1-\alpha)\, \mathrm{Int}_u(c_{p_i}), \tag{7}$$

$$\mathbb{Q}(I) = \mathrm{Cost}(I), \tag{8}$$

where $\mathrm{Pop}(p_i)$ is the popularity of POI p_i, $\mathrm{Int}_u(c)$ refers to the interest tourist u has for POI category c, $\mathrm{Cost}(I)$ is the total distance between p_1 and p_n and α is a weight parameter which can be adjusted as required. The overall problem is thus:

$$\mathrm{Max}\left(\frac{\mathbb{P}(I)}{\mathbb{Q}(I)}\right). \tag{9}$$

Let $T_{p_i,\,p_j} = 1$, if the tourist travels directly from POI p_i to p_j and 0 otherwise [15]. The aim is to optimise equation 9 taking into consideration the constraints below:

$$\sum_{j=2}^{N} T_{p_1,\ p_j} = \sum_{\iota=1}^{N-1} T_{p_\iota,\ p_N} = 1, \tag{10}$$

$$\sum_{\iota=1}^{N-1} T_{p_\iota,p_m} = \sum_{j=2}^{N} T_{p_m,p_j} \leq 1,\ \forall m = 2, \ldots, N-1, \tag{11}$$

$$2 \leq p_\iota \leq N,\ \forall \iota = 2, \ldots, N, \tag{12}$$

$$p_\iota - p_j + 1 \leq (N-1)(1 - T_{p_\iota,\ p_j}),\ \forall \iota, j = 2, \ldots, N, \tag{13}$$

$$|\operatorname{cost}(I)| \leq B. \tag{14}$$

The limitation set out in eqn. 10 is to ascertain that the recommended itinerary starts at the first POI p_1 and finishes at the last POI p_n. The limitation in eqn. 11 ensures that no POI is visited more than once and that each POI in the itinerary is connected to the other. The restrictions imposed by eqns. 12 and 13 ensure that the proposed itinerary does not include a sub-itinerary 14. Equation 14 ensures that the cost of the itinerary does not exceed some budget B.

5 Proposed PIONEER Algorithm

The PIONEER algorithm proposed in this paper is based on the genetic algorithm called NSGA-II [8] which is a multi objective optimisation algorithm. The algorithm works as follows: Suppose a tourist u is travelling to a city c. A cosine similarity testis conducted between u and global tourists to obtain the top 10 matching global tourists.

Table 2. Comparison of Recall between our proposed method and other baseline algorithms

Algorithms	PIONEER	TRIC	GREEPOP	GREENEAR	RAND
Delhi- Edinburgh	**0.478±0.019**	0.386±0.015	0.359±0.009	0.342±0.038	0.316±0.023
Osaka-Edinburgh	**0.462±0.019**	0.372±0.029	0.346±0.022	0.319±0.036	0.291±0.011
Vienna-Edinburgh	**0.512±0.013**	0.414±0.020	0.388±0.019	0.358±0.031	0.343±0.041
Delhi-Osaka	**0.546±0.015**	0.463±0.013	0.431±0.032	0.407±0.007	0.378±0.025
Glasgow-Edinburgh	**0.372±0.003**	0.287±0.017	0.257±0.029	0.235±0.052	0.206±0.021

The list of POIs visited by the matching tourists and hence their itineraries are obtained from their travel histories. This list becomes the initial population $\mathcal{P}^{\text{initial}}$ (line 5) and is the input to the NSGA-II algorithm.

As defined in equations 7 and 8, popularity and interest of proposed itinerary must be maximised while at the same time minimising the cost. This leads to two objective functions:

$$\mathcal{F}_1 = \mathbb{P}(I), \tag{15}$$

$$\mathcal{F}_2 = \mathbb{Q}(I). \tag{16}$$

The two objective functions for every individual in the initial population $\mathcal{P}^{\text{initial}}$ are evaluated (lines 6-7) and each is assigned a rank and sorted into several fronts using a fast non-domination sorting method as described in [8] (line 8). Individuals belonging to the same front have the same rank. The crowding distance of each individual is determined from their objective values.

The parent population is selected from the initial population based on the rank and crowding distance. The genetic operations of selection, crossover and mutation are applied to the parent population to generate the child population $\mathcal{C}^{\text{sol}}$ (line 12). Fitness values of each itinerary in the child population are calculated (lines 13-14).

The child and parent lists are then combined (line 15) and a sorting algorithm is used to compare each itinerary with other itineraries using the criteria of nondominance and crowding distance. A natural selection is made by selecting all solutions belonging to the first fronts and discarding the others.

The algorithm stops when the maximum number of generations is reached.

6 Experiments

6.1 Dataset

This paper uses the YFCC100M (Yahoo! Flickr Creative Commons 100M) dataset [26]. It is a huge dataset comprising 100 million photos and videos obtained from Flickr.

From the metadata about the dataset information such as the date and time and the latitude and longitude values when the photos were taken and the ids of users who took the photos can be extracted.

6.2 Baseline Algorithms

– **Greedy Nearest (GREENEAR)**: The next POI to be visited is chosen at random from those POIs which are nearest, but which have not yet been visited.

– **Greedy Most Popular (GREEPOP)**: The next POI to be visited is chosen at random from these POIs which are the most popular, but which have not yet been visited.

– **Random Choice (RAND)**: The next POI to be visited is chosen at random from the set of POIs which have not yet been visited.

– **Tour Recommendation With Interest Category (TRIC)**: The recommended tour must include a compulsory category, which is the most frequently visited POI category in that city 18.

Table 3. Comparison of F1-Score between our proposed method and other baseline algorithms

Algorithms	PIONEER	TRIC	GREEPOP	GREENEAR	RAND
Delhi- Edinburgh	**0.528**±**0.033**	0.440±0.029	0.404±0.081	0.382±0.009	0.350±0.046
Osaka-Edinburgh	**0.527**±**0.010**	0.448±0.013	0.416±0.018	0.384±0.038	0.354±0.041
Vienna-Edinburgh	**0.589**±**0.046**	0.493±0.021	0.466±0.006	0.435±0.030	0.418±0.032
Delhi-Osaka	**0.569**±**0.016**	0.501±0.038	0.423±0.005	0.401±0.018	0.374±0.034
Glasgow-Edinburgh	**0.388**±**0.007**	0.310±0.026	0.280±0.041	0.256±0.027	0.231±0.018

6.3 Real-Life Evaluation

Only those tourists who have completed at least two travel sequences and visited at least two categories of POIs are used to evaluate the proposed algorithm.

The method is applied to both local and global datasets [23], as well as visitors who are comparable. We compare similar visitors in this study by looking at the top 10 associated visitors from global data sets.

For our experiments, categories of real travelling series are chosen based on the history of associated visitors in a given area. The standard evaluation metrics, that is, Precision, Recall and F1-Score have been used to test our algorithm.

- **Tour Recall (TourRec(I))**: Let C_{rec} be the list of POI categories suggested by our algorithm and let C_{real} be the list of all categories of POI which a tourist has visited in reality. Eqn. 17 defines TourRecall, which returns the proportion of POI categories visited by a tourist which were also recommended by the algorithm:

$$\text{TourRec}(I) = \frac{|C_{\text{rec}} \cap C_{\text{real}}|}{|C_{\text{real}}|}. \tag{17}$$

- **Tour Precision (TourPre(I))**: Let C_{rec} be the list of POI categories suggested by the algorithm and let C_{real} be the set of POI categories visited by a tourist in reality. TourPrecision is defined as the ratio of proposed POI categories which are also found in the tourist's actual travel history. TourPrecision is defined as follows:

$$\text{TourPre}(I) = \frac{|C_{\text{rec}} \cap C_{\text{real}}|}{|C_{\text{rec}}|}. \tag{18}$$

- **Tour F1-Score (TourF1-score(I))**: The mean harmonic value of Precision and Recall for the proposed itinerary I is referred to as Tour F1-Score (Eqn. 19):

$$\text{Tour F1-score}(I) = \frac{2 \times \text{TourPre}(I) \times \text{TourRec}(I)}{\text{TourPre}(I) + \text{TourRec}(I)}. \tag{19}$$

6.4 Comparison of Precision, Recall and F1-Score

The proposed PIONEER algorithm performs better when compared to other baseline algorithms such as GREEPOP, TRIC, GREENEAR and RAND. Tables 1, 2 and 3 show how the PIONEER algorithm compares with other baseline approaches in terms of Precision, Recall and $\text{F1} - \text{Score}$ values. The results show that PIONEER fares better than the baseline approaches as far as Precision, Recall and F1-Score metrics are concerned.

Recall measurements depending upon $|C_{\text{rec}}|$ and $|C_{\text{rec}} \cap C_{\text{real}}|$ as per in Eqn. 17. Here the values of $|C_{\text{rec}} \cap C_{\text{real}}|$ is better compared to the various baseline approaches which can be computed utilizing the PIONEER algorithm. Typically, the suggested PIONEER algorithm is based on local as well as global datasets, and ultimately suggests many POIs, resulting in better Recall scores for various baseline approaches.

For the PIONEER algorithm, the Precision scores are more because they are dependent on $|C_{\text{rec}}|$ and $|C_{\text{rec}} \cap C_{\text{real}}|$ as per in Eqn. 18. We found that C_{rec} scores vary for various baseline approaches during the analysis. The values of $|C_{\text{rec}} \cap C_{\text{real}}|$ are higher for the suggested PIONEER algorithm.

ISSN 2007-9737

The F1-Score increase in the suggested PIONEER algorithm based on Precision and Recall compared to other baseline approaches.

7 Conclusion and Future Research

This research has presented a new method called PIONEER which recommends tourist itineraries which maximise tourist interest, POI popularity while at the same time reducing cost. The algorithm uses the actual travel patterns of tourists which are obtained from geo-tagged photos.

From the dataset, tourists' interests, tour popularity and travelling costs are calculated for training the PIONEER algorithm. The suggested method is dependent on the selection of many POIs by taking into account the POI time visiting factor. PIONEER will not depend on the travelling history of a certain individual in new locations.

The case in which a visitor wants to visit new places is therefore taken into consideration. PIONEER is compared with various baselines using multiple criteria such as Precision, Recall, and F1-Score. The findings of the study demonstrate that the suggested algorithm surpasses baseline approaches.

This research will be extended in the future to cater for tourists who travel in groups (e.g., with family and friends) where the challenge is to cater for the individual interests and preferences of each group member.

References

1. **Anagnostopoulos, A., Atassi, R., Becchetti, L., Fazzone, A., Silvestri, F. (2017).** Tour recommendation for groups. Data Mining and Knowledge Discovery, Vol. 31, No. 5, pp. 1157–1188. DOI: 10.1007/s10618-016-0477-7.

2. **Borràs, J., Moreno, A., Valls, A. (2014).** Intelligent tourism recommender systems: A survey. Expert Systems with Applications, Vol. 41, No. 16, pp. 7370–7389. DOI: 10.1016/j.eswa.2014.06.007.

3. **Brilhante, I., Macedo, J. A., Nardini, F. M., Perego, R., Renso, C. (2014).** Tripbuilder: A tool for recommending sightseeing tours. Proceedings of the European Conference on Information Retrieval, Advances in Information Retrieval, Vol. 8416, pp. 771–774. DOI: 10.1007/978-3-319-06028-6_93.

4. **Castillo, L., Armengol, E., Onaindia, E., Sebastia, L., Gonzalez-Boticario, J., Rodriguez, A., Fernandez, S., Arias, J., Borrajo, D. (2008).** Samap: An user-oriented adaptive system for planning tourist visits. Expert Systems with Applications, Vol. 34, No. 2, pp. 1318–1332. DOI: 10.1016/j.eswa.2006.12.029.

5. **Chen, C., Zhang, D., Guo, B., Ma, X., Pan, G., Wu, Z. (2015).** TripPlanner: Personalized trip planning leveraging heterogeneous crowdsourced digital footprints. IEEE Transactions on Intelligent Transportation Systems, Vol. 16, No. 3, pp. 1259–1273. DOI: 10.1109/tits.2014.2357835.

6. **Crandall, D. J., Backstrom, L., Cosley, D., Suri, S., Huttenlocher, D., Kleinberg, J. (2010).** Inferring social ties from geographic coincidences. Proceedings of the National Academy of Sciences, Vol. 107, No. 52, pp. 22436–22441. DOI: 10.1073/pnas.1006155107.

7. **De-Choudhury, M., Feldman, M., Amer-Yahia, S., Golbandi, N., Lempel, R., Yu, C. (2010).** Automatic construction of travel itineraries using social breadcrumbs. Proceedings of the 21st ACM Conference on Hypertext and Hypermedia, pp. 35–44. DOI: 10.1145/1810617.1810626.

8. **Deb, K., Pratap, A., Agarwal, S., Meyarivan, T. (2002).** A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, Vol. 6, No. 2, pp. 182–197. DOI: 10.1109/4235.996017.

9. **Dumitrescu, L., Fuciu, M. (2015).** Consumer behaviour in the tourist segmentation process – a marketing research. Studies in Business

and Economics, Vol. 10, No. 1, pp. 66–76. DOI: 10.1515/sbe-2015-0005.

10. **Golden, B. L., Levy, L., Vohra, R. (1987).** The orienteering problem. Naval Research Logistics, Vol. 34, No. 3, pp. 307–318. DOI: 10.1002/1520-6750(198706)34:3⟨307::aid-nav3220340302⟩3.0.co;2-d.

11. **Gunawan, A., Lau, H. C., Vansteenwegen, P. (2016).** Orienteering problem: A survey of recent variants, solution approaches and applications. European Journal of Operational Research, Vol. 255, No. 2, pp. 315–332. DOI: 10.1016/j.ejor.2016.04.059.

12. **Ji, R., Xie, X., Yao, H., Ma, W. Y. (2009).** Mining city landmarks from blogs by graph modeling. Proceedings of the 17th Association for Computing Machinery International Conference on Multimedia, pp. 105–114. DOI: 10.1145/1631272.1631289.

13. **Kisilevich, S., Mansmann, F., Keim, D. (2010).** P-DBSCAN: A density based clustering algorithm for exploration and analysis of attractive areas using collections of geo-tagged photos. Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research and Application, pp. 1–4. DOI: 10.1145/1823854.1823897.

14. **Li, J., Qian, X., Tang, Y. Y., Yang, L., Mei, T. (2013).** GPS estimation for places of interest from social users' uploaded photos. IEEE Transactions on Multimedia, Vol. 15, No. 8, pp. 2058–2071. DOI: 10.1109/tmm.2013.2280127.

15. **Lim, K. H. (2015).** Recommending tours and places-of-interest based on user interests from geo-tagged photos. Proceedings of the Association for Computing Machinery Special Interest Group on Management of Data on PhD Symposium, pp. 33–38. DOI: 10.1145/2744680.2744693.

16. **Lim, K. H., Chan, J., Karunasekera, S., Leckie, C. (2017).** Personalized itinerary recommendation with queuing time awareness. Proceedings of the 40th International ACM Special Interest Group on Information Retrieval Conference on Research and Development in Information Retrieval, pp. 325–334. DOI: 10.1145/3077136.3080778.

17. **Lim, K. H., Chan, J., Leckie, C., Karunasekera, S. (2016).** Towards next generation touring: Personalized group tours. Proceedings of the International Conference on Automated Planning and Scheduling, Vol. 26, pp. 412–420. DOI: 10.1609/icaps.v26i1.13775.

18. **Lim, K. H., Chan, J., Leckie, C., Karunasekera, S. (2017).** Personalized trip recommendation for tourists based on user interests, points of interest visit durations and visit recency. Knowledge and Information Systems, Vol. 54, No. 2, pp. 375–406. DOI: 10.1007/s10115-017-1056-y.

19. **Majumder, A., Sarkar, J. L., Pati, B., Panigrahi, C. R., Ramasamy, V., Roy, S., Kumar, V. (2022).** MERIT: Multi-itinerary tourist recommendation engine for industrial internet of things. Proceedings of the IEEE INFOCOM 2022 IEEE Conference on Computer Communications Workshops, pp. 1–6. DOI: 10.1109/INFOCOMWKSHPS54753.2022.9798002.

20. **Popescu, A., Grefenstette, G., Moëllic, P. A. (2009).** Mining tourist information from user-supplied collections. Proceedings of the 18th ACM Conference on Information and Knowledge Management, pp. 1713–1716. DOI: 10.1145/1645953.1646211.

21. **Sarkar, J. L., Majumder, A. (2021).** A new point-of-interest approach based on multi-itinerary recommendation engine. Expert Systems with Applications, Vol. 181, pp. 115026. DOI: 10.1016/j.eswa.2021.115026.

22. **Sarkar, J. L., Majumder, A. (2022).** gTour: Multiple itinerary recommendation engine for group of tourists. Expert Systems with Applications, Vol. 191, pp. 116190. DOI: 10.1016/j.eswa.2021.116190.

23. **Sarkar, J. L., Majumder, A., Panigrahi, C. R., Roy, S. (2020).** MULTITOUR: A multiple itinerary tourists recommendation engine. Electronic Commerce Research and Applications, Vol. 40, pp. 100943. DOI: 10.1016/j.elerap.2020.100943.

24. **Sarkar, J. L., Majumder, A., Panigrahi, C. R., Roy, S., Pati, B. (2022).** Tourism recommendation system: A survey and future research directions. Multimedia Tools and Applications, Vol. 82, No. 6, pp. 8983–9027. DOI: 10.1007/s11042-022-12167-w.

25. **Sinnot, R. W. (1984).** Virtues of the haversine. Sky and Telescope, Vol. 68, No. 2, pp. 159.

26. **Thomee, B., Shamma, D. A., Friedland, G., Elizalde, B., Ni, K., Poland, D., Borth, D., Li, L. J. (2016).** YFCC100M: The new data in multimedia research. Communications of the ACM, Vol. 59, No. 2, pp. 64–73. DOI: 10.1145/2812802.

27. **Vansteenwegen, P., Souffriau, W., Berghe, G. V., Oudheusden, D. V. (2011).** The city trip planner: An expert system for tourists. Expert Systems with Applications, Vol. 38, No. 6, pp. 6540–6546. DOI: 10.1016/j.eswa.2010.11.085.

28. **Vansteenwegen, P., Souffriau, W., Oudheusden, D. V. (2011).** The orienteering problem: A survey. European Journal of Operational Research, Vol. 209, No. 1, pp. 1–10. DOI: 10.1016/j.ejor.2010.03.045.

29. **Vansteenwegen, P., Van Oudheusden, D. (2007).** The mobile tourist guide: An OR opportunity. OR Insight, Vol. 20, No. 3, pp. 21–27. DOI: 10.1057/ori.2007.17.

30. **Wörndl, W., Hefele, A. (2016).** Generating paths through discovered places-of-interests for city trip planning. Information and Communication Technologies in Tourism, pp. 441–453. DOI: 10.1007/978-3-319-28231-2_32.

31. **Ying, H., Wu, J., Xu, G., Liu, Y., Liang, T., Zhang, X., Xiong, H. (2018).** Time-aware metric embedding with asymmetric projection for successive POI recommendation. World Wide Web, Vol. 22, No. 5, pp. 2209–2224. DOI: 10.1007/s11280-018-0596-8.

Article received on 02/01/2023; accepted on 12/11/2023.

** Corresponding author is Sanjeev K. Cowlessur.*