Personalized Recommendation of Travel Itineraries based on Tourist Interests and Preferences

Kwan Hui Lim[*†]
[*]Department of Computing and Information Systems, The University of Melbourne, Australia
[†]Victoria Research Laboratory, National ICT Australia, Australia
limk2@student.unimelb.edu.au

ABSTRACT

Travel itinerary recommendation is an important but challenging problem, due to the need to recommend captivating Places-of-Interest (POI) and construct these POIs as a connected itinerary. Another challenge is to personalize these recommended itineraries based on tourist interests and their preferences for starting/ending POIs and time/distance budgets. Our work aims to address these challenges by proposing algorithms to recommend personalized travel itineraries for both individuals and groups of tourists, based on their interest preferences. To determine these interests, we first construct tourists' past POI visits based on their geo-tagged photos and then build a model of user interests based on their time spent visiting each POI. Experimental evaluation on a Flickr dataset of multiple cities show that our proposed algorithms out-perform various baselines in terms of recall, precision, F1-score and other heuristics-based metrics.

Keywords

Personalization; Tour Recommendations; Trip Planning; Group Recommendations; User Interests

1. INTRODUCTION

Holiday travelling and touring are popular leisure activities around the world, as shown by the 1.1 billion tourists worldwide who travelled in 2014 [22]. Economically, tourism is also an important and lucrative industry with an annual revenue of more than US$1.2 trillion in 2014. The importance of tourism has led to the creation of many tour planning resources such as online travel guides and tour agencies. From a tourist's perspective, their main purpose would be to visit captivating Places-of-Interest (POI) within the duration of their stay in the visited city.

Despite the availability of online travel guides and services provided by tour agencies, tourists still face challenges in tour planning due to the following reasons: (i) online travel guides are effective in recommending popular POIs but these POIs may not cater to the unique interest preferences of individual tourists; (ii) in a foreign city, a tourist would require a customized trip itinerary with personalized POI recommendations, starting/destination points and time constraints (instead of a simple list of popular POIs without an itinerary); (iii) for groups of tourists, tour agencies offer standard group tours which may not cater to the diverse interest preferences of individuals within the tour group.

1.1 Research Goals

Our main research goal is to recommend personalized travel itineraries based on the unique preferences of tourists. This personalization of travel itineraries include the following aspects of preferences, namely: (i) tourist interests; (ii) starting and ending POIs; and (iii) available length of travel duration. More specifically, we aim to investigate the following research questions:

- **R1:** How can we model the interest preferences of *individual tourists* and personalize tour recommendations for these tourists based on their interests, time budgets and preferences for starting/ending points?
- **R2:** Building upon R1, how can we model the interest preferences for *groups of tourists* and make tour recommendations that best satisfy the interest preferences of all tourists in a tour group?

1.2 State-of-the-Art Work

Many works on travel recommendation for individual tourists are based on combinatorial optimization problems such as the Orienteering problem [21, 23] or Generalized Maximum Coverage problem [8]. For example, Choudhury et al. [7] and Brilhante et al. [3, 4] modelled the itinerary recommendation problem based on the Orienteering problem and Generalized Maximum Coverage problem, respectively. In particular, Brilhante et al. [3, 4] optimized the recommended tour itineraries using both POI popularity and user interests, which is based on the (normalized) visit counts to POIs by individual tourists. Others such as Kurashima et al. [12, 13] and Chen et al. [5] also optimized for user interests, in addition to their respective considerations for different transport modes and traffic conditions. Similar to that of [3, 4], Chen et al. also determined user interests based on a similar normalized POI visit count, while Kurashima et al. utilized a probabilistic framework based on a combined topic and Markov model. As part of R1, we extend upon these state-of-the-art by recommending personalized tours using a more fine-grained definition of user interests, which is based on the tourists' past POI visit duration.

Thus far, most travel recommendation research focus on recommending itineraries to a single tourist, whereas tourists frequently travel in groups in real-life. While there are interesting research that aim to recommend top-k POIs to groups [19], these works recommend individual POIs, instead of an itinerary of connected POIs, and constructing individual POIs into an itinerary is not a trivial scheduling problem due to various constraints, e.g., time and distance.

Similarly, there has been several interesting applications [9, 2] that recommend tours to groups of tourists based on user interests and group membership, which are explicitly provided by the tourists. However, it is a challenging task to determine the interest preferences for multiple tourists and cluster these tourists into groups that best align their interests. As part of R2, we aim to explore the problem of group tour recommendation from the perspectives of tourist grouping, itinerary planning, and tour guide assignment.

Outline of Paper. This paper is structured as follows: Section 2 describes our current progress and contributions; Section 3 highlights our plans for future work; and Section 4 summarizes and concludes this paper.

2. PROGRESS TO DATE

In the following sections, we discuss our progress and contributions to date, which includes: (i) formulating the tour recommendation problem; (ii) modeling of user interests; (iii) developing various algorithms for recommending tours to individuals and groups; and (iv) evaluating our proposed algorithms against various baselines.

2.1 General Problem Formulation

Our basic tour recommendation problem is based on variants of the Orienteering problem [21, 23] and we restate the formal problem definition used in [15]. Consider a particular city with N POIs, and a tourist t with the constraints of a time/distance budget and preferences to start and end at specific POIs p_1 and p_N, respectively. In this case, our main goal is to recommend a travel itinerary $I = (p_1, ..., p_N)$ that optimizes the following:

$$Max \sum_{i=2}^{N-1} \sum_{j=2}^{N} x_{i,j} Utility(i) \tag{1}$$

where $x_{i,j} = 1$ if the travel itinerary includes a travel path from POI i to j, and $x_{i,j} = 0$ otherwise. We then solve for Eq. 1, such that:

$$\sum_{j=2}^{N} x_{1,j} = \sum_{i=1}^{N-1} x_{i,N} = 1 \tag{2}$$

$$\sum_{i=1}^{N-1} x_{i,k} = \sum_{j=2}^{N} x_{k,j} \leq 1, \qquad \forall\ k = 2, ..., N-1 \tag{3}$$

$$\sum_{i=1}^{N-1} \sum_{j=2}^{N} Cost(i,j) x_{i,j} \leq B \tag{4}$$

Eq. 1 is our main objective function, which aims to maximize a certain *utility* that can be obtained from the recommended travel itinerary. This *utility* could be a value unique to individual tourists (e.g., interest preferences) or common to all tourists (e.g., POI popularity). Eq. 2 to 4 are constraints that are applied to the recommended itinerary, namely: (i) Constraint 2 ensures that the travel itinerary starts at p_1 and ends at POI p_N; (ii) Constraint 3 ensures no POIs in the itinerary are disconnected or visited more than once; and (iii) Constraint 4 ensures that the entire itinerary can be completed within the budget B, based on the time or distance cost of travelling between POIs.

2.2 Modeling User Interests using Past Visits

After defining our basic tour recommendation problem, we now describe the approach we use to: (i) obtain the past visit history of tourists; and (ii) determine the interests of these tourists based on their past interest. These approaches were also used in various of our works [15, 14, 16].

2.2.1 Obtaining Past Visits

We use geo-tagged photos as a proxy for tourist real-life visits. In particular, we select geo-tagged photos taken near POIs as these photos imply that the tourist was physically at that POI (hence he/she was able to take that photo). From the series of geo-tagged photos taken by a tourist u, we are then able to determine the *past travel history* of this tourist, which is represented as:

$$H_u = \big((p_1, t^a_{p_1}, t^d_{p_1}), ..., (p_n, t^a_{p_n}, t^d_{p_n})\big) \tag{5}$$

where H_u is an ordered sequence comprising a series of triplets $(p_x, t^a_{p_x}, t^d_{p_x})$. This triplet consists of the visited POI p_x, arrival time $t^a_{p_x}$ and departure time $t^d_{p_x}$ at POI p_x. The visited POI p_x is determined based on geo-tagged photos taken near (e.g., within 100m) that POI. As the geo-tagged photos include their taken time, we can determine the arrival and departure time, $t^a_{p_x}$ and $t^d_{p_x}$ based on the first and last photo consecutively taken at POI p_x.

2.2.2 Modeling of User Interests

Each POI is also tagged with a POI category (e.g., shopping, museum, beach, etc), which we determine using information from Wikipedia. Given that $\bar{D}(p_x)$ is the average amount of time that all tourist spent at POI p_x, we define the interest level of a tourist u in POI category c as follows:

$$Int_u(c) = \sum_{p_x \in H_u} \frac{(t^d_{p_x} - t^a_{p_x})}{\bar{D}(p_x)} \delta(Cat_{p_x}{=}c) \tag{6}$$

where $\delta(Cat_{p_x}{=}c) = 1$ if POI p_x belongs to category c, and $\delta(Cat_{p_x}{=}c) = 0$ otherwise. Eq. 6 determines the interest level of tourist t based on the amount of time he/she spends at POIs of category c, relative to the average amount of time spent by other tourist at the same POIs. Thus, our intuition is that the more (less) time a tourist spends at a POI, the more (less) interested he/she is. This modeling of user interests is discussed further in [15].

2.3 Personalized Travel Recommendation for Individual Tourist

Building upon our definition of interest in Eq. 6, we developed the PersTour algorithm that aims to recommend personalized tours to individual tourists [15]. This personalization takes place in terms of two aspects, namely: (i) the POIs are recommended based on tourist interest, with a varying emphasis on POI popularity as determined by the tourist; and (ii) the recommended visit duration is determined based on the tourist interest level, i.e., more time spent at POIs that the tourist is more interested in.

For the personalization of recommended POIs, we modified Eq. 1 such that the *utility* is based on both user interest alignment and POI popularity, that is:

$$Utility(i) = \eta Int_u(Cat_i) + (1-\eta) Pop(i) \tag{7}$$

where $Int_u(Cat_i)$ is defined previously in Eq. 6 and $Pop(i)$ is the popularity of POI i, which we define as the number of times POI i has been visited by all tourists. The parameter $\eta = [0, 1]$ allows tourists the flexibility to emphasize on either the user interest or POI popularity components, at varying levels.

For the personalization of POI visit duration, we utilize the interest level of tourist u in POI i (i.e., Eq. 6) and the average visit duration of all tourists at POI i (i.e., $\bar{D}(i)$). Thus, this personalized visit duration/time is defined as:

$$Time_u(i) = Int_u(Cat_i) \times \bar{D}(i) \tag{8}$$

In short, we determine the personalized visit duration for tourist u based on how interested (uninterested) this tourist is in POI i, and accordingly recommend a longer (shorter) visit duration relative to the average visit duration. By factoring in the average visit duration, we are able to adapt to POIs of difference sizes, e.g., less time at a smaller museum but more time at a larger one. We refer readers to [15] for more details on this work.

2.4 Travel Recommendation with Mandatory Category

Extending upon our basic tour recommendation problem (Section 2.1), we proposed the TourRecInt algorithm that aims to recommend tours with a mandatory POI category, which is the POI category that a tourist is most interested in. In this work, we examine a tourist's past POI visit history and define the POI category that this tourist is most interested in based on the most frequently visited POI category. In addition to this mandatory POI category, TourRecInt also personalizes tours based on other tourist preferences such as specific starting and ending points, and any time or distance budgets. Apart from tourism-related applications, TourRecInt can also be extended to consider multiple mandatory POI categories and be generalized to other path planning problems, e.g., John travelling from his office to home but having to drop by a supermarket, restaurant and petrol station to buy some groceries, take-out dinner and top-up petrol, respectively, before heading home. We refer readers to [14] for more details on this work.

2.5 Group Travel Recommendation for Multiple Tourists

Recommending tours for groups of tourists involve additional challenges, compared to recommending tours for individual tourists. Some of these challenges include customizing tours to appeal to the interest preferences of the group as a whole and assigning tour guides with the appropriate expertises to lead each group. We termed this the Group Tour Recommendation (GroupTourRec) problem, which we introduced in [16]. Technically, GroupTourRec is a challenging problem that is NP-hard as it comprises variants of the Orienteering problem and clustering problem, which are also NP-hard [10, 1]. To solve this NP-hard problem, we divide GroupTourRec into more manageable sub-problems, and propose approaches to solve each sub-problem, which include:

- For the sub-problem of recommending tour itineraries to groups, we first determine the group interest preferences based on the average interest among all tourists in a group, then use a variant of the Orienteering problem that considers both POI popularity and group interest preferences to recommend tours.
- For the sub-problem of allocating tour guides to lead each group, we first model the expertises of tour guides based on past tours they have led, then use an Integer programming approach to assign tour guides whose expertises best match the tour recommended to each group.

We refer readers to [16] for more details on this work.

2.6 Evaluation of Proposed Approaches

Datasets. As mentioned in Section 2.2.1, we use geo-tagged photos to determine a tourist's past visits to POIs. For this purpose, we utilized the Yahoo! Flickr Creative Commons 100M dataset [25, 20], which includes 100M Flickr photos and videos along with their geographical coordinates and date/time taken. Apart from building a model of tourist interest preferences, we are able to use these past POI visits as a ground truth of real-life POI visits, which in turn is used to evaluate our proposed algorithms and various baselines.

Baseline Algorithms. In our research, we compared our proposed algorithms against various baselines, including:

- **StdTour:** Standard tour itineraries that are offered by real-life tour agencies such as www.viator.com and local travel websites in the respective cities.
- **GNear:** A distance-based greedy algorithm that randomly selects the next POI to visit from the three *nearest*, unvisited POIs.
- **GPop:** A popularity-based greedy algorithm that randomly selects the next POI to visit from the three *most popular*, unvisited POIs.
- **Rand:** A baseline that *randomly selects* the the next POI to visit from all unvisited POIs.

We selected these baselines as they reflect real-life tourist behaviours, such as signing up for an organized tour (StdTour) or simply visiting POIs that are nearby (GNear) or popular (GPop). In contrast, Rand shows us the performance of the various algorithm against a random recommendation.

Performance Metrics and Results. Using past POI visits as a ground truth, we utilize various Information Retrieval (IR) metrics such as Recall, Precision and F1-score to compare the performance of our proposed algorithms against the various baselines, in terms of how well the recommend tours reflect the real-life tours taken by tourists. In addition to these IR-based metrics, we also use various heuristics-based metrics such as POI popularity, tourist interest alignment, tour guide expertise and group interest similarity to evaluate the performance of our proposed algorithms in terms of these utility scores. Using a Flickr dataset spanning multiple touristic cities across the world, we evaluated our proposed algorithms against these baselines, with results showing that our proposed algorithms out-perform these baselines for all cities, based on the above-mentioned metrics. Due to limited space, we refer readers to [15, 14, 16] for a more detailed discussion on the results.

3. FUTURE RESEARCH PLAN

Our future research plan includes the following:

1. Utilizing a game-theoretic approach to tour recommendations such that we try to minimize a global utility

of "crowdedness", while trying to personalize tours to individuals. In a museum setting, this would involve recommending an exhibit visit sequence that considers visitor interests but do not over-crowd a particular exhibit by sending all visitors there at the same time.

2. In addition to POI visit duration, we intend to explore other models of user interests using features based on textual contents of social media posts, number of photos posted and user tags.

3. Refining our evaluation methodology by: (i) using Amazon Mechanical Turk for a qualitative study of user opinions on our recommended travel itineraries, such as in [7, 18]; and (ii) using online controlled experiments to better understand user behaviour and their fine-grained actions when deciding between our recommended travel itineraries and other baselines, such as in [11, 17].

4. Other potential ideas for future work include incorporating image recognition techniques [6], considering the current user context (time, location, weather, etc) [24], and modelling the different levels of influence among users in a tour group [26].

4. CONCLUSION

In this paper, we described the problem of travel itinerary recommendation and proposed the PersTour, TourRecInt and GroupTourRec algorithms for recommending itineraries that are personalized based on tourist interests and their preferences for starting/ending POIs and time/distance budgets. We also illustrated our approach of using geo-tagged photos to construct tourists' POI visit history and to build a model of user interests based on these visits. Using a Flickr dataset spanning multiple cities, experimental results show that our proposed algorithms out-perform various baselines in terms of tourist interests, POI popularity, recall, precision, F1-score and other relevant metrics.

Acknowledgments. This work was supported by NICTA and a Google Australia PhD Travel Scholarship. The author thanks Shanika Karunasekera, Christopher Leckie, Jeffrey Chan and the anonymous reviewers for their useful comments.

5. REFERENCES

[1] D. Aloise, A. Deshpande, P. Hansen, and P. Popat. NP-hardness of Euclidean sum-of-squares clustering. *Machine Learning*, 75(2):245–248, 2009.

[2] L. Ardissono, A. Goy, G. Petrone, M. Segnan, and P. Torasso. Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices. *Applied Artificial Intelligence*, 17(8-9):687–714, 2003.

[3] I. Brilhante, J. A. Macedo, F. M. Nardini, R. Perego, and C. Renso. Where shall we go today? Planning touristic tours with TripBuilder. In *Proc. of CIKM'13*, pages 757–762, 2013.

[4] I. R. Brilhante, J. A. Macedo, F. M. Nardini, R. Perego, and C. Renso. On planning sightseeing tours with TripBuilder. *Inf. Proc. & Manag.*, 51(2):1–15, 2015.

[5] C. Chen, D. Zhang, B. Guo, X. Ma, G. Pan, and Z. Wu. TripPlanner: Personalized trip planning leveraging heterogeneous crowdsourced digital footprints. *IEEE Trans. on Intelligent Transp. Sys.*, 16(3):1259–1273, 2015.

[6] Y.-Y. Chen, A.-J. Cheng, and W. H. Hsu. Travel recommendation by mining people attributes and travel group types from community-contributed photos. *IEEE Transactions on Multimedia*, 15(6):1283–1295, 2013.

[7] M. D. Choudhury, M. Feldman, S. Amer-Yahia, N. Golbandi, R. Lempel, and C. Yu. Automatic construction of travel itineraries using social breadcrumbs. In *Proc. of HT'10*, pages 35–44, 2010.

[8] R. Cohen and L. Katzir. The generalized maximum coverage problem. *Information Processing Letters*, 108(1):15–22, 2008.

[9] I. Garcia, L. Sebastia, and E. Onaindia. On the design of individual and group recommender systems for tourism. *Expert Systems with Applications*, 38(6):7683–7692, 2011.

[10] B. L. Golden, L. Levy, and R. Vohra. The orienteering problem. *Naval Research Logistics*, 34(3):307–318, 1987.

[11] J. Kiseleva, M. J. Mueller, L. Bernardi, C. Davis, I. Kovacek, M. S. Einarsen, J. Kamps, A. Tuzhilin, and D. Hiemstra. Where to go on your next trip?: Optimizing travel destinations based on user preferences. In *Proc. of SIGIR'15*, pages 1097–1100, 2015.

[12] T. Kurashima, T. Iwata, G. Irie, and K. Fujimura. Travel route recommendation using geotags in photo sharing sites. In *Proc. of CIKM'10*, pages 579–588, 2010.

[13] T. Kurashima, T. Iwata, G. Irie, and K. Fujimura. Travel route recommendation using geotagged photos. *Knowledge and Information Systems*, 37(1):37–60, 2013.

[14] K. H. Lim. Recommending tours and places-of-interest based on user interests from geo-tagged photos. In *Proc. of SIGMOD'15 PhD Symposium*, pages 33–38, 2015.

[15] K. H. Lim, J. Chan, C. Leckie, and S. Karunasekera. Personalized tour recommendation based on user interests and points of interest visit durations. In *Proc. of IJCAI'15*, pages 1778–1784, 2015.

[16] K. H. Lim, J. Chan, C. Leckie, and S. Karunasekera. Towards next generation touring: Personalized group tours. In *Proc. of ICAPS'16*, 2016.

[17] K. H. Lim, E.-P. Lim, B. Jiang, and P. Achananuparp. Using online controlled experiments to examine authority effects on user behavior in email campaigns. In *Proc. of HT'16*, 2016.

[18] D. Quercia, R. Schifanella, and L. M. Aiello. The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. In *Proc. of HT'14*, pages 116–125, 2014.

[19] S. B. Roy, L. V. Lakshmanan, and R. Liu. From group recommendations to group formation. In *Proc. of SIGMOD'15*, pages 1603–1616, 2015.

[20] B. Thomee, D. A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L.-J. Li. YFCC100M: The new data in multimedia research. *Communications of the ACM*, 59(2):64–73, 2016.

[21] T. Tsiligirides. Heuristic methods applied to orienteering. *J. of the Operational Research Society*, 35(9):797–809, 1984.

[22] UNWTO. United Nations World Tourism Organization (UNWTO) annual report 2014, 2015. http://www2.unwto.org/annual-reports.

[23] P. Vansteenwegen, W. Souffriau, and D. V. Oudheusden. The orienteering problem: A survey. *European Journal of Operational Research*, 209(1):1–10, 2011.

[24] K. Waga, A. Tabarcea, and P. Franti. Recommendation of points of interest from user generated data collection. In *Proc. of CollaborateCom'12*, pages 550–555, 2012.

[25] Yahoo! Webscope. Yahoo! Flickr Creative Commons 100M dataset (YFCC-100M), 2014.

[26] Q. Yuan, G. Cong, and C.-Y. Lin. COM: a generative model for group recommendation. In *Proc. of KDD'14*, pages 163–172, 2014.