Recommending and Planning Trip Itineraries for Individual Travellers and Groups of Tourists

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
Trip planning is both challenging and tedious for tourists due to their unique interest preferences and various trip constraints. Despite the availability of online resources for tour planning and services provided by tour agencies, there are various challenges such as: (i) selecting POIs that are personalized to the unique interests of individual travellers; (ii) constructing these POIs as an itinerary, with considerations for time availability and starting/ending place preferences (e.g., near a tourist’s hotel); (iii) for tour agencies to group tourists into tour groups such that the recommended tour appeals to the interests of the group as a whole; and (iv) similarly, for tour agencies to assign tour guides with the right expertise to lead each of these tour groups. In our work, we aim to develop algorithms for recommending personalized tours to both individual travellers and groups of tourists, based on their interest preferences, which we automatically determine based on geo-tagged photos posted by these tourists. Using a Flickr dataset of geo-tagged photos as ground-truth for real-life POI visits in multiple cities, we evaluate our proposed algorithms using various metrics such as precision, recall, F1-score, user interest scores and POI popularity, among others.

1 Introduction
1.1 Motivations
Tourism is an important industry to the world economy, contributing more than US$1.2 trillion in revenue and accounting for more than 1.1 billion international tourists (UNWTO 2015). Despite the importance of tourism, planning a tour or trip itinerary is still a challenging task for any visitor in a foreign city, due to unfamiliarity with the various Points of Interest (POI) in the city. Although there are many online resources available for tour planning, there still exist challenges such as: (i) many travel guides simply recommend popular POIs that do not reflect the tourist’s interest preferences or consider various trip constraints, such as the available time for touring and preferred starting/ending location, e.g., starting and ending near the tourist’s accommodation; and (ii) even after obtaining a list of POIs, it is a tedious task to construct an itinerary of sequential POI visits with the considerations of travelling time, visiting time, and specific starting/ending points.

One possible solution is to engage the services of tour operators to organize such tour itineraries. However, tour operators may not be aware of the unique interests of individual tourists and face the same challenge of recommending tours that are personalized to the tourist’s interest preferences. Furthermore, tour operators typically offer group tours to multiple tourists and face the additional challenges of: (i) optimizing for an appropriate tour group size, e.g., large groups to minimize cost overheads or small groups to maximize tourist experience; (ii) constructing tours with POIs that are appealing to multiple tourists in a group; and (iii) assigning tour guides with the appropriate expertise to best lead each tour group.

Our work aims to address the challenges of recommend-
ing tours that are suitable for individual travellers\footnote{We use the terms “traveller” and “tourist” interchangeably.} and groups of tourists, in particular, considering the diverse set of interests among these tourists. To achieve these goals, we implemented a tour recommendation framework (Figure 1) that utilizes geo-tagged photos (Flickr) and crowd-sourced information (Wikipedia), and proposed various algorithms based on variants of the Orienteering problem and various clustering algorithms. In the following section, we describe the three main research questions that we aim to address as part of this work.

### 1.2 Research Questions

Our PhD research aims to develop tour recommendation algorithms that are meaningful at multiple levels, namely for individual tourists, groups of tourists, and the entire tourist population. Our work is further motivated by the following research questions (RQ):

- **RQ 1:** At the individual level, how can we recommend personalized tours that consider the interest preferences and trip constraints (e.g., time/distance budget and preferred starting/destination POIs) of individual tourists?

- **RQ 2:** At the group level, how can we recommend group tours that consider appropriate tour group sizes, interest preferences for multiple tourists in a group and assignment of tour guides based on their expertise?

- **RQ 3:** At the global level, how can we recommend tours that benefit the tourist population as a whole? I.e., how do we plan tours that minimizes undesirable effects at POIs, such as over-crowdedness and long queuing times?

### 1.3 Related Work

As our work aims to recommend tours for individual travellers and groups of tourists, we first discuss some state-of-the-art works in the respective areas of tour recommendation for individual travellers and tour recommendation for groups of tourists.

#### Tour Recommendation for Individuals

There are various works that aim to recommend tours for individuals, i.e., a single tourist, and we discuss some key literature from this area. Many of these works approach tour recommendation as an optimization problem, such as the Orienteering problem \cite{Tsiligirides1984, Vansteenwegen2011} or Generalized Maximum Coverage problem \cite{Cohen2008}. For example, \cite{Choudhury2010} was one such work that recommended tours for an individual tourist, with a specific starting and ending POI, while ensuring that the tour can be completed within a certain time. Others like \cite{Gionis2014} extended upon this research area by implementing the constraint of a sequence ordering to the POI visits, e.g., restaurant → shopping → beach → park. Similarly, \cite{Brilhante2013, Brilhante2015} modelled tour recommendation based on the Generalized Maximum Coverage problem, with considerations for both POI popularity and user interests. Other tour recommendation research also included transportation-related considerations, such as \cite{Chen2015} that considered varying travelling times based on traffic conditions, and \cite{Kurashima2010, Kurashima2013} that utilized different modes of transportation in their travel routes. For more information, \cite{Gavalas2014} provides a comprehensive discussion of algorithms that aim to recommend tours to individual tourists. In addition, there have been many web/mobile-based applications developed for the same purpose such as \cite{Brilhante2014, Refanidis2014, Castillo2008}, which are based on variations of the discussed works.

#### Tour Recommendation for Groups

In recent years, group recommendations have been studied in-depth by researchers, such as by \cite{Amer-Yahia2009} and \cite{Hu2014}, who proposed and applied group recommendation algorithms to the retail domain, i.e., recommending top-\(k\) retail items such as movies, books, music.\footnote{While group recommendation research is related to tour recommendation for groups of tourists, the latter involves additional challenges, such as constructing the recommended POIs (items) into a connected itinerary and considerations for specific starting/ending points, and a limited time budget for visiting and travelling between POIs. As such, we focus more on literature regarding tour recommendation for groups of tourists and refer readers to (Boratto and Carta 2011) for a more comprehensive discussion on group recommendation works.} For the tourism domain, there are many interesting works that apply group recommendation algorithms for tourism-related purposes, resulting in applications such as e-Tourism \cite{Garcia2003} and Travel Decision Forum \cite{Jameson2003}. Extending upon \cite{Sebastia2009}, e-Tourism \cite{Garcia2009, Garcia2011} explicitly solicits the interest preferences and group membership details of users, then recommends tours that best satisfy the interest preferences of the entire group based on the user-provided groupings. Other applications like Intrigue \cite{Ardissono2003} and Travel Decision Forum \cite{Jameson2003} aim to fulfill a similar purpose of recommending tours to groups of tourists. The main difference is that Intrigue requires users to provide their POI preferences instead of specific interests, while Travel Decision Forum includes an additional online discussion phase to get its users to mutually agree on proposed changes to the tour itinerary.

### 1.4 Structure and Organization

The rest of the paper is organized as follows. In Section 2, we describe some of our main contributions in the area of tour recommendation. In Section 3, we discuss some future directions that we aim to embark on for the remaining of my PhD. Finally, we summarize and conclude this paper in Section 4.
2 Contributions to Date

In the following sections, we describe some of our main contributions thus far, which include: (i) implementing a general framework for deriving user-POI visit history based on geo-tagged photos (Section 2.1); (ii) formulating the basic tour recommendation problem and various variants (Section 2.2); (iii) proposing tour recommendation algorithms for individual tourists (Section 2.3); and (iv) proposing tour recommendation algorithms for groups of tourists (Section 2.4).

2.1 General Framework

As illustrated in Figure 1, our overall tour recommendation framework makes use of: (i) geo-tagged photos that are tagged with a geographical coordinates (latitude/longitude) and stamped with the time taken; and (ii) a POI list comprising POI names, category and latitude/longitude coordinates. The geo-tagged photos can be obtained from any photo sharing website such as Flickr or Instagram, and the POI list can be obtained from Wikipedia or a specific city’s tourism-related website (e.g., City of Melbourne). This framework comprises the following steps:

1. Map geo-tagged photos to a list of POIs if their coordinates differ by a specific distance, e.g., ≤100m, resulting in a list of POI visits. For calculating this spherical (earth) distance, we make use of the Haversine formula (Sinnott 1984).

2. Construct the tourist travel history by connecting POI visits (obtained from Step 1) of the same tourist. In particular, we derive a user’s visit duration at a POI based on the time difference between his/her first and last photo (of a consecutive nature) at that POI.

3. Calculate POI popularity and tourist interest preferences based on tourist travel histories from Step 2. POI popularity is based on the number of visits to a specific POI (the more visits, the more popular), while tourist interest is based on variations of POI visit durations (which is discussed later).

While Step 1 typically uses geo-tagged photos, it can be easily extended to other media with a lat/long coordinate and time-stamp, e.g., GPS traces on mobile phones or other location-based social networking services such as geo-tagged tweets on Twitter. As input to this framework, we use Wikipedia and Flickr geo-tagged photos that are publicly available as part of the Yahoo! Flickr Creative Commons 100M dataset (Yahoo! Webscope 2014; Thomee et al. 2016).3 This framework was used in various of our works, such as (Lim et al. 2015b; Lim 2015; Lim et al. 2016), which we describe in more detail in the later sections.

3Our pre-processed dataset (i.e., photos mapped to POI visits and visit sequences) are also made publicly available at https://sites.google.com/site/limkwanhui/datacode.

2.2 Basic Problem Definition

We now restate the basic tour recommendation problem definition that we described in (Lim et al. 2015b). Given the set of POIs, a budget B, starting POI, destination POI, our main goal is to recommend a tour itinerary that maximizes both user interests and POI popularity Pop(i), while adhering to the budget B. Formally, we want to construct a tour itinerary I = (p1, ..., pN) that:

\[
\text{Max} \sum_{i=2}^{N-1} \sum_{j=2}^{N} x_{i,j} \left( \eta \text{Int}(\text{Cat}_i) + (1 - \eta) \text{Pop}(i) \right)
\]  

where \(x_{i,j} = 1\) if we travel directly from POI i to j (i.e., we visit POI i, followed by POI j), and \(x_{i,j} = 0\) otherwise. We then attempt to solve for Eqn. 1, subjected to the following constraints:

\[
\sum_{j=2}^{N} x_{1,j} = \sum_{i=1}^{N} x_{i,N} = 1
\]

\[
\sum_{i=1}^{N-1} x_{i,k} = \sum_{j=2}^{N} x_{k,j} \leq 1, \quad \forall \ k = 2, ..., N-1
\]

\[
\sum_{i=1}^{N-1} \sum_{j=2}^{N} \text{Cost}(i,j) x_{i,j} \leq B
\]

\[
2 \leq p_i \leq N, \quad \forall \ i = 2, ..., N
\]

\[
p_i - p_j + 1 \leq (N - 1)(1 - x_{i,j}), \quad \forall \ i, j = 2, ..., N
\]

Eqn. 1 attempts to maximize a dual-objective of POI popularity and user interests on all POIs in the recommended tour itinerary, and \(\eta\) controls the emphasis given to either POI popularity or user interests. Constraints 2 to 6 ensures that: (i) the itinerary starts and ends at POI 1 and N, respectively (Constraint 2); (ii) all POIs in the itinerary are connected and no POIs are re-visited (Constraint 3); (iii) the total time taken to visit all POIs in the itinerary is within the time budget B, based on a function Cost(p, q) that is computed from both a personalized POI visit duration and travelling time between POIs (Constraint 4); (iv) there are no sub-tours (separate self-looping tours) in the proposed solution, based on the sub-tour elimination constraint proposed in (Miller, Tucker, and Zemlin 1960) for the Travelling Salesman Problem (Constraints 5 and 6). We then proceed to solve this tour recommendation problem as an integer programming problem and for this purpose, we used the Ipsoolve linear programming package (Berkelaar, Eikland, and Notebaert 2004).


2.3 Tour Recommendation for Individual Tourist

In the first year of my PhD, we focused our research on personalized tour recommendation for individual tourists (RQ 1). In this research area, there has been various works that aim to recommend interest-based tours based on the Generalized Maximum Coverage problem (Brilhante et al. 2015) and using a combination of topic and Markov models (Kurashima et al. 2013). We built upon these earlier works by exploring an intuitive model of user interests based on POI visit time and recommending tour itineraries with a mandatory visit category. Our contributions include:

Tour Recommendation with Personalized POIs and Visit Duration. In (Lim et al. 2015b), we proposed the PerSTOUR algorithm for recommending personalized tours with POIs and visit duration based on POI popularity and time-based user interests. This algorithm models POI popularity based on POI visit count, and time-based user interests using a tourist’s total visit duration at POIs of a certain category, relative to that of an average tourist. Our intuition is that a tourist is more interested in a POI category if he/she spends more time at POIs of this category. For determining tourist POI visit duration, we utilize the geo-tagged photos taken by a user and calculate their POI visit duration based on the time difference between the first and last photo taken at a specific POI. Based on measures of tour popularity, tourist interest, recall, precision and F1-score, experimental results show that our PerSTOUR algorithm is able to recommend POIs and visit durations that more accurately reflect tourists’ real-life visits, compared to various greedy-based baselines. For more information on this work, please refer to (Lim et al. 2015b).

Customized Tour Recommendation with Mandatory Categories. In (Lim 2015), we proposed the TourReCINT algorithm for recommending customized tours with a mandatory POI category based on tourist interests. This algorithm optimizes a variant of the Orienteering problem (Tsiligirides 1984; Vansteenwegen, Souffriau, and Oudheusden 2011), with a time/distance budget, starting POI, destination POI and mandatory POI category. We defined this mandatory POI category as the most frequently visited POI category based on a tourist’s visit history. Thereafter, we solve this variant of the Orienteering problem as an integer programming problem. Using a ground truth of real-life POI visits by tourists (based on their geo-tagged photos), experimental results show that TourReCINT out-perform various baselines, in terms of precision, recall and F1-score. For more information on this work, please refer to (Lim 2015).

2.4 Tour Recommendation for Groups of Tourists

In the second year of my PhD, we proceeded to investigate customized tour recommendation for groups of tourists (RQ 2). While there is extensive literature on group recommendation of top-k items (Boratto and Carta 2011) and tour recommendation for individual tourist (Gavalas et al. 2014), there is limited work on tour recommendation for groups of tourists. For works that explore tour recommendation for groups (Garcia, Sebastia, and Onaindia 2011; Ardissono et al. 2003; Jameson, Baldes, and Kleinbauer 2003), they focus more on the group recommendation aspect and do not consider the assignment of tour guides to lead these tour groups. Similarly, many of these works assume that the tourist groupings and interest preferences are explicitly provided. As part of RQ 2, we aim to study tour recommendation for groups as a more holistic problem, which includes grouping tourists with diverse interest preferences, recommending tour itineraries and assigning tour guides to these groups. Our contributions include:

Group Tour Recommendation with Tour Guide Assignment.

In (Lim et al. 2016), we introduced the Group Tour Recommendation (GroupTourRec) problem, which involves recommending tours that best satisfy the interest preferences of groups of tourists, where each tour group is subsequently led by a tour guide. To solve this GroupTourRec problem, we proposed an approach for recommending group tours that aims to: (i) determine tourist interests based on past POI visits, and cluster tourists with similar interests into a group; (ii) recommend tours to groups based on a variant of the Orienteering problem that considers both group interests and POI popularity; and (iii) assigns tour guides with the appropriate expertise to lead each tour, using an integer programming approach. In addition, this problem is also technically challenging due to its NP-hard complexity. As such, we use greedy-based approaches and integer programming to solve for smaller subproblems of tourists grouping, POI recommendation and tour guides assignment, as part of the group tour recommendation problem. Based on various measures of group interest similarity, total/maximum/minimum tour interests and total tour guide expertise, results show that our proposed approach out-performs various baselines, including standard tour packages offered by real-life tour agencies. For more information on this work, please refer to (Lim et al. 2016).

Detecting Location-centric Communities.

In (Lim et al. 2015a), we investigated a complementary problem of detecting communities of users that frequently visit or reside in similar locations. In this work, we proposed the use of Social-Spatial-Temporal (SST) links, which are traditional social/friendship links between users augmented with spatial and temporal information, e.g., visited the same place within a certain time-frame. Using standard community detection algorithms (such as the Louvain (Blondel et al. 2008), Infomap (Rosvall and Bergstrom 2008) and Label-Prop (Raghavan, Albert, and Kumara 2007) algorithms) on these SST links, we were able to detect location-centric communities comprising users who exhibit strong similarities in terms of the places they visit and reside in. In another related work (Lim and Datta 2016), we also observed that user communities with similar interests are more likely to reside in the same locality. In the future, we intend to extend this work to determine if users are travelling alone or as a group, and accordingly recommend tour that are appropriate.

This work (Lim et al. 2016) will also be presented at the 26th International Conference on Automated Planning and Scheduling (ICAPS’16).
ate for individuals and groups. For more information on this work, please refer to (Lim et al. 2015a).

3 Future Research Plan

For the remaining of my PhD, we aim to work on tour recommendation strategies that benefit the tourist population as a whole (RQ 3) and we intend to work on the following:

Game-theoretic Approaches to Tour Recommendation. Traditionally, tour recommendation algorithms aim to propose tours that maximize the personal profit of individual tourists. One limitation of this approach is that while the individual tourist benefits, the entire tourist population could potentially “lose” (e.g., everyone going to the most popular POI but ends up overcrowding and creating long queues at that POI, leading to a poor tour experience for most people). To address this problem, we intend to adopt a game theoretic approach to tour recommendations where we model POI “crowdedness” as a common utility and derive equilibrium strategies to recommend tours that will benefit all tourists as a whole. Potential applications of this work would be in optimizing for queuing times at attractions/rides in theme parks and preventing over-crowding at exhibits within museums. For example, instead of recommending the most popular attraction in a theme park to all visitors and increasing the queuing times, we may want to recommend some less popular attractions that have shorter queuing times to a subset of visitors.

4 Conclusion

In summary, we introduced the general problem of tour recommendation, and discussed in greater detail, the specific problems of recommending tours for individual travellers and groups of tourists, along with the consideration of their unique interest preferences. We then described our various contributions in the general area of tour recommendation, which include the following:

- Proposing the PERS TOUR algorithm for recommending personalized tours with POIs and visit duration based on POI popularity and time-based user interests (Lim et al. 2015b).

- Proposing the TOUR RECI NT algorithm for recommending customized tours with a mandatory POI category based on user interests, i.e., the most frequently visited POI category (Lim 2015).

- Introducing the GROUP TOUR RECI problem and proposing an approach to cluster tourists with similar interests into groups, recommend tours based on group interest preferences and POI popularity, and assign tour guides to lead these groups (Lim et al. 2016).

- Developing an approach for detecting location-centric communities using SST links, which are traditional friendship links augmented with spatial and temporal information (Lim et al. 2015a).

Using a Flickr dataset of tourist visits to POIs in multiple cities, we compare our proposed algorithms against various baselines using evaluation metrics such as precision, recall, F1-score, user interest scores, POI popularity scores, and others. Experimental results show that our proposed algorithms out-perform their respective baselines in terms of these metrics, across all cities. We refer readers to the respective papers (listed above) for a more detailed discussion on these results.

As part of future work, we also described our plans to adopt a game theoretic approach to tour recommendation. For this work, we aim to model POI “crowdedness” as a common utility and implement equilibrium strategies to recommend tours that minimize over-crowding at POIs for the tourist population as a whole.

5 Acknowledgments

National ICT Australia (NICTA) is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre of Excellence Program. The author thanks Shanika Karunasekera, Christopher Leckie and Jeffrey Chan for their useful comments and discussions.

References


