Information Processing and Management 51 (2015) 1–15

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

# On planning sightseeing tours with TripBuilder

Igo Ramalho Brilhante [a], Jose Antonio Macedo [a], Franco Maria Nardini [b,*,1], Raffaele Perego [b], Chiara Renso [b]

[a] *Federal University of Cearà, Fortaleza, Brazil*
[b] *ISTI–CNR, Pisa, Italy*

ARTICLE INFO

*Article history:*
Received 12 May 2014
Received in revised form 3 October 2014
Accepted 8 October 2014
Available online 16 November 2014

*Keywords:*
Recommender systems
Trajectory mining
Sightseeing tours

ABSTRACT

We propose TripBuilder, an unsupervised framework for planning personalized sightseeing tours in cities. We collect categorized Points of Interests (PoIs) from Wikipedia and albums of geo-referenced photos from Flickr. By considering the photos as traces revealing the behaviors of tourists during their sightseeing tours, we extract from photo albums spatio-temporal information about the itineraries made by tourists, and we match these itineraries to the Points of Interest (PoIs) of the city. The task of recommending a personalized sightseeing tour is modeled as an instance of the Generalized Maximum Coverage (GMC) problem, where a measure of personal interest for the user given her preferences and visiting time-budget is maximized. The set of actual trajectories resulting from the GMC solution is scheduled on the tourist's agenda by exploiting a particular instance of the Traveling Salesman Problem (TSP). Experimental results on three different cities show that our approach is effective, efficient and outperforms competitive baselines.



## 1. Introduction

Tourists approaching their destination for the first time have to deal with the problem of planning a sightseeing itinerary that covers the most subjectively interesting attractions, and fits the time available for their visit. Precious information can be nowadays gathered from many digital sources, e.g., travel guides, maps, institutional sites, travel blogs. Nevertheless, the users still need to choose the preferred PoIs, guess how much time is needed to visit them and to move from one attraction to the next one. In this paper we discuss TripBuilder, an unsupervised system helping tourists to build their own personalized sightseeing tour. Given the target destination, the time available for the visit, and the user's profile, our system recommends a time-budgeted tour that maximizes user's interests and takes into account both the time needed to enjoy the attractions and to move from one PoI to the next one. Moreover, the knowledge base feeding TripBuilder recommendation model is entirely and automatically extracted from publicly available Web services, namely, Wikipedia, Flickr and Google maps.

We observe that an increasing number of tourists share on social networks their travel photos. Unofficial estimates state that Flickr, one of the most popular photo-sharing platforms, collected about 518 million of public photos in 2012.[2] Each photo comes with very useful information such as: tags, comments and likes from Flickr social network, number of views,

* Corresponding author.
*E-mail addresses:* igobrilhante@lia.ufc.br (I.R. Brilhante), jose.macedo@lia.ufc.br (J.A. Macedo), francomaria.nardini@isti.cnr.it (F.M. Nardini), raffaele.perego@isti.cnr.it (R. Perego), chiara.renso@isti.cnr.it (C. Renso).

[1] Part of this research has been conducted while visiting Federal University of Cearà, Fortaleza, Brazil.
[2] http://www.flickr.com/photos/franckmichel/6855169886/.

http://dx.doi.org/10.1016/j.ipm.2014.10.003
0306-4573/

information about the user, timestamp, GPS coordinates of the place where the photo was taken. This allows us to roughly reconstruct the movements of users and their interests by analyzing the time-ordered sequence of their photos. However, the process of recognizing relevant PoIs given such set of photos is not trivial due to the noise present in the data. User tags are in many cases missing, wrong, or irrelevant for our purposes (e.g., *me and Ann, travel to Europe, Easter 2012*). Moreover, information available may be sparse and characterized by a skewed distribution.

Fortunately, in Wikipedia,[3] we can find that most entities of interest for tourism are described in a dedicated page from which we can extract: the (multilingual) name of the PoI, its precise geographic coordinates, the categories to which the PoI belongs according to a weak but robust ontology (i.e., the PoI is a church, a square, a museum, a historical building, a bridge, etc.). By spatially joining and re-conciliating tourists' photo albums and related information from Flickr with relevant PoIs data extracted from Wikipedia pages, we can derive a knowledge base that represents the behavior of people visiting a given city.[4] In this knowledge base the popularity of a PoI is estimated from the number of visitors that shot photos there, while from the timestamps of the first and last photos taken in a PoI we estimate the average time spent for the visit. Finally, we exploit the Wikipedia categories of the PoIs visited by a given tourist to build her user profile. For example, when a user takes many pictures of churches and museums, we can infer a preference for cultural/historical attractions. Analogously, we can aggregate this information at the level of itinerary to build a profile for each frequent visiting pattern.

We address the problem of planning the visit to the city as a two-step process. First, given the profile of the user and the amount of time available for the visit, we formalize and address the TRIPCOVER problem: choosing the set of itineraries across the PoIs that best fits user interest and respects the given time constraint. Then, the selected itineraries are joined in a sightseeing itinerary by means of a heuristic algorithm addressing the Trajectory Scheduling Problem (TRAJSP), a particular instance of Traveling Salesman Problem (TSP). The formalization of TRIPCOVER as an instance of the Generalized Maximum Coverage (GMC) problem can be found in Brilhante, Macedo, Nardini, Perego, and Renso (2013), while (Brilhante, Macedo, Nardini, Perego, & Renso, 2014) demos the capabilities of the TRIPBUILDER application.[5] In this paper, we extend the previous research with the following relevant and original contributions:

- we detail our unsupervised method for mining common patterns of movements of tourists in a given geographic area. To this purpose we use: (i) Flickr, to gather public photos (and their metadata), (ii), Wikipedia to gather information regarding PoIs in the given city, (iii) Google maps to estimate the time needed to move from a PoI to the next one in the sightseeing itinerary. The resulting knowledge base stores PoIs, their popularity, the time needed in average to visit them, the categories for which each PoI is relevant, and the patterns of movement of tourists that visited them in the past. In order to assess our system, we report about the building of a knowledge base covering three Italian cities, which are important for tourism and guarantee variety and diversity in terms of size and the richness of public user-generated content available: Rome, Florence, and Pisa. The resulting knowledge base, available for download to favor the reproducibility of results, is analyzed and its characteristics are here discussed;
- we adopt the result discussed in Brilhante et al. (2013) where TRIPCOVER is defined and an approximate algorithm is proposed to compute the set of trajectories in the knowledge base that best fits user interests given a time budget. In this paper, we address the subsequent step that consists in processing this set of trajectories to build a schedule of the complete sightseeing itinerary over the tourist's agenda. To this end, we define TRAJSP as a variation of the TSP problem. It aims at finding the shortest path crossing all the trajectories in the TRIPCOVER solution. We investigate the TRAJSP problem and its peculiarities, which are exploited to devise an efficient and effective heuristic algorithm;
- finally, we report on several new experiments to evaluate effectiveness and efficiency of all the components of our system and show that our solution outperforms competitive baselines. In particular, we assess TRIPBUILDER performance in providing budgeted sightseeing itineraries made up of actual PoI patterns tailored to the specific preferences of the tourist.

The paper is structured as follows: Section 2 introduces the TRIPCOVER problem and the approximation algorithm used to solve it. Moreover, the TRAJSP problem is defined and addressed in Section 3. Section 4 details the unsupervised method that builds the knowledge base, while Section 5 presents the experiments we perform to assess the effectiveness and the efficiency of our solution. Finally, Section 6 discusses related work and Section 7 draws the conclusions of the work.

## 2. The TripCover problem

Let $\mathcal{P} = \{p_1, \ldots, p_N\}$ be the set of PoIs in our city. Each PoI $p$ is univocally identified by its geographic coordinates, a name, a radius specifying its spatial extent, and a *relevance vector*, $\vec{v}_p \in [0,1]^{|C|}$, measuring the normalized relevance of $p$ w.r.t a set of categories $C$.

Symmetrically, let $u$ be a user from the set $\mathcal{U}$, and $\vec{v}_u \in [0,1]^{|C|}$ the *preference vector* stating the normalized interest of $u$ for the categories in $C$. The preference vector can be explicitly given by the user, or implicitly learned. Without loss of generality,

[3] http://www.wikipedia.org.
[4] Thereinafter, we will consider cities as the destination targets of our users, although our technique is general and scale-independent.
[5] http://tripbuilder.isti.cnr.it/.