Transportation Research Part C 68 (2016) 532–548

Contents lists available at ScienceDirect

Transportation Research Part C

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

Joint leisure travel optimization with user-generated data via perceived utility maximization

K. Gkiotsalitis *, A. Stathopoulos

National Technical University of Athens, Iroon Polytechniou 5, Zografou 15773, Athens, Greece

ARTICLE INFO

Article history:
Received 30 August 2015
Received in revised form 12 May 2016
Accepted 12 May 2016

Keywords:
Joint travel optimization
Social media
Social networks
Utility-maximization
Stochastic search
Pattern recognition

ABSTRACT

The lack of personalized solutions for managing the demand of joint leisure trips in cities in real time hinders the optimization of transportation system operations. Joint leisure activities can account for up to 60% of trips in cities and unlike fixed trips (i.e., trips to work where the arrival time and the trip destination are predefined), leisure activities offer more optimization flexibility since the activity destination and the arrival times of individuals can vary.

To address this problem, a perceived utility model derived from non-traditional data such as smartphones/social media for representing users' willingness to travel a certain distance for participating in leisure activities at different times of day is presented. Then, a stochastic annealing search method for addressing the exponential complexity optimization problem is introduced. The stochastic annealing method suggests the preferred location of a joint leisure activity and the arrival times of individuals based on the users' preferences derived from the perceived utility model. Test-case implementations of the approach used 14-month social media data from London and showcased an increase of up to 3 times at individuals' satisfaction while the computational complexity is reduced to almost linear time serving the real-time implementation requirements.

1. Introduction

The complexity of organizing joint leisure activities emerges from the need of satisfying simultaneously several activity participants with different preferences. This complexity is increased in urban environments with numerous places of interest due to the abundance of alternatives. For instance, if a group of x individuals with some form of social ties decides to organize a joint leisure activity, the selection of place and time will be based solely on the developed consciousness of individuals from prior experiences of attending leisure activities. Consequently, the solution space of alternative leisure activity places and starting activity times is not fully explored since individuals are not able to comprehend and re-call the whole set of alternatives for searching to an optimal solution. With numerous groups of people organizing joint leisure activities on a daily basis, the problem scales up at a city level and the organization of joint leisure activities based solely on individuals' accumulated experiences leads to sub-optimal selection of activity destination and arrival times.

The importance of optimizing joint leisure activities is higher in urban environments since up to 60% of the conducted trips are related to leisure activities and the complexity of transport, social and activity networks leaves more room for optimization. For instance, TfL (2014) posed that 29.2% of all daily trips are related to leisure activities, while 28% were

* Corresponding author.
E-mail addresses: kgki1987@central.ntua.gr (K. Gkiotsalitis), a.stath@transport.ntua.gr (A. Stathopoulos).

http://dx.doi.org/10.1016/j.trc.2016.05.009
0968-090X/

conducted for shopping and personal business and 10.7% for other activities including escort. Similar results were observed on the New York Regional Travel survey RTS (2014).

Selecting a location and individuals' arrival times for joint activity participation among multiple individuals is not a trivial task due to the problem scalability and the lack of knowledge regarding users' preferences. Nowadays, non-traditional data sources facilitate the collection and analysis of more detailed, user-generated data that can be utilized for improving the level of understanding regarding users' mobility/activity preferences. Social Media data (SM), Floating Car data (FCD), Mobile Phone data (MP) and Smart Card data (SC) are some examples of user-generated data and have the potential to improve the understanding of travelers' patterns and mobility preferences including transport mode selection, their departure and arrival times, frequency and scope of undertaken trips and many more (refer to Gkiotsalitis and Stathopoulos (2015a) for a more comprehensive literature review).

In the literature, Musolesi and Mascolo (2007) utilized cellular data logs for correlating the mobility patterns of an individual with the mobility patterns of his friends and acquaintances and can provide some evidence on the theoretical concepts developed by Carrasco et al. (2008), Arentze and Timmermans (2008) and Chen et al. (2014) on predicting agents' mobility based on their social networks. De Domenico et al. (2013) worked also along the same direction using data from the Nokia Mobile Data Challenge dataset. In addition, Carrel et al. (2015) tested on San Francisco's Muni network a system of integrated methods to reconstruct and track travelers usage of transit by (i) matching location data from smartphones to automatic transit vehicle location (AVL) data and by (ii) identifying all out-of-vehicle and in-vehicle portions of the passengers trips. Those theoretical concepts place the traveler at the center of decisions (ego-centric approaches) and offer a micro-simulation framework where collection of large-scale user-generated data is required.

Calabrese et al. (2011b,a), Gonzàlez et al. (2008), Zhang et al. (2010), Pan et al. (2006) and White and Wells (2002) utilized cellular data for predicting the mobility patterns of individuals in urban scenarios over time and space. Those studies, including studies of White and Wells (2002) and Djuknic and Richton (2001), attempted to exploit the emergence of the mobile positioning technology and the market penetration of smartphones by developing methods for estimating the OD matrices in study areas. In the same way, Sohn and Kim (2008) used cellular communication systems/cell phone towers to transfer information and estimate OD matrices. To give practical examples, Alexander et al. (2015) estimated average daily origin destination trips and the destination types (home, work or other) from triangulated mobile phone records of millions of anonymized users, while Huang and Levinson (2015) developed methodologies for building choice sets for modeling the non-work destination choice based on real-time GPS data travel data. Calabrese et al. (2013) presented techniques to extract useful mobility information from the mobile phone traces of millions of users to investigate individual mobility patterns within a metropolitan area. Then, Calabrese et al. (2013) compared the results to mobility measures computed using odometer readings from the annual safety inspections of all private vehicles for validation and identification of the differences between individual mobility and vehicular mobility. In addition, Iqbal et al. (2014) and Toole et al. (2015) estimated multiple aspects of travel demand using call detail records (CDRs) from mobile phones in conjunction with open- and crowdsourced geospatial data, census records, and surveys. Toole et al. (2015) implemented a travel demand estimation model for generating representative origin destination matrices using big data and route trips through road networks constructed using open and crowd-sourced data repositories. CDR datasets have also used for transport planning and zonal division. Dong et al. (2015) extracted commuters' origins and destinations information from the mobile phone CDR data and then used the extracted data for traffic zone division. A K-means clustering method was used to classify a cell-area (the area covered by a base stations) and tag a certain land use category.

In a similar fashion, social media data from social networks like Facebook, Twitter, and the image sharing service, Flickr, have also been used for capturing the performed users' activities at different locations and day times via advanced spatio-temporal analysis and educated rules (refer to Gkiotsalitis and Stathopoulos (2015b)). In the same work, techniques for estimating individuals' daily schedules and the sequence of activities were developed. Alesiani et al. (2014) focused also on the same topic introducing a probabilistic model for modeling individuals' daily schedules based on input data from several sources (i.e., Social Media, Cellular Data).

However, the works on social media or cellular user-generated data sources focused mainly on the O/D estimation/prediction or the identification of individuals' mobility patterns without exploring the area of joint leisure activity participation and the optimization of the undertaken trips. The real-time optimization of a joint leisure activity includes the selection of the location of the joint leisure activity, the starting time of the activity and the arrival time of all activity participants at that location. This requires the maximization of the perceived utility of all activity participants and depends on the (a) current locations of individuals; (b) the time of the day; (c) the disutility of traveling from a current location to the location of the joint activity; (d) the arrival time of each user to the joint activity location and the waiting time until the activity starts.

The optimization of joint leisure activities requires strong analytics for identifying and modeling users' preferences from historical user-generated data and scalable optimization techniques for the suggestion of the activity location and the arrival times in real-time. In this work, a utility maximization model derived from social media data (analysis of historical data from more than 4 months for each individual with automated pattern recognition techniques) for representing users' willingness to travel a certain distance for participating in leisure activities at different times of day is introduced. Moreover, a scalable optimization technique based on stochastic annealing search for suggesting the preferred location of a joint leisure activity and the arrival times of individuals in real time is presented. In Section 2 the utility maximization model based on pattern recognition of the mobility of users is presented. In Section 3 the stochastic annealing search algorithm is described and in