Neurocomputing 181 (2016) 4–18

Contents lists available at ScienceDirect

Neurocomputing

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

Identifying mismatch between urban travel demand and transport network services using GPS data: A case study in the fast growing Chinese city of Harbin

JianXun Cui [a], Feng Liu [b,*], Jia Hu [c], Davy Janssens [b], Geert Wets [b], Mario Cools [d]

[a] School of Transportation Science and Engineering, Harbin Institute of Technology, #73, Huanghe Road, Nangang District, Harbin 150090, China
[b] Transportation Research Institute (IMOB), Hasselt University, Wetenschapspark 5, bus 6, B-3590 Diepenbeek, Belgium
[c] Turner-Fairbank Highway Research Center, Federal Highway Administration, 6300 Georgetown Pike, McLean, VA 22101, USA
[d] LEMA, University of Liège, Chemin des Chevreuils 1, Bât B.52/3, 4000 Liège, Belgium

ARTICLE INFO

Article history:
Received 27 February 2015
Received in revised form
10 June 2015
Accepted 27 August 2015
Available online 27 November 2015

Keywords:
Travel demand modeling
Road network analysis
Circuity
GPS data
Big data

ABSTRACT

The world's urban population growth and economic development have led to the reshaping of metropolitan space layouts among residential, employment and shopping locations, generating growing mismatch between travel demand and transport services. A reliable method to accurately analyze mobility demand and underlying transport network systems and to identify areas with serious mismatch problems is important for the design of effective policy measures. In this paper, we make use of the wide deployment of GPS devices in vehicles in many cities today, to develop such a method. This approach is developed using GPS data collected from all taxis operating in the Chinese city of Harbin between July and September in 2013. It consists of four major steps. First, city-wide mobility patterns are modeled based on GPS trajectories. This model captures a set of key traffic characteristics for each pair of regions in the entire urban network, including travel demand, travel speed and route directness of travel paths. From this model, a set of indicators is then built to measure the road transport performance between the regions, and the areas with serious mismatch problems are subsequently pinpointed. Finally, the identified problematic regions are further examined and specific transport problems are analyzed. By applying the proposed method to the city of Harbin, the potential and effectiveness of this method are demonstrated. Moreover, with more and more urban vehicles being equipped with GPS devices, the designed method can be easily transferred to other cities, thus paving a way for the adoption of the presented approach for an up-to-date and spatial-temporal sensitive road network analysis approach that supports the establishment of a more sustainable urban transport system.

1. Introduction

With the continuing urbanization of the world's population and the rapid growth of cities, transportation managers and urban planners are faced with many challenges, such as heavily congested roads and increasing levels of pollution [1–3]. Therefore, there is an urgent need to further understand the growing urban travel demand and changing land use structure as well as to accurately assess the conditions of the current transport network, in order to efficiently address the urban transport challenges and improve transportation management and urban planning for a long sustainable future [4,5].

The congestion and environmental problems are even more imminent in the context of developing countries [6,7]. On the one hand, many large cities in these countries are experiencing an unprecedented period of economic growth, accompanied by an influx of foreign investment and domestic rural workers as well as an increasing rate of motorization. Take as an example the Chinese city of Harbin, where our case study is conducted. The urban population in this city has increased by 12.99% over the last decade, and an average of 400 new vehicles joins the urban traffic flow every day [8]. This growth has become a serious challenge to a city where the urban layouts, residential patterns and road infrastructure were designed dozens of years ago, at a time when virtually no citizens owned private cars and when people lived in areas within minutes traveling of their work. On the other hand, although advanced land use and transportation modeling techniques as well as transport network analysis methods are currently available, they are mainly used for transport management and

* Corresponding author. Tel.: +32 11269125; fax: +32 11269199.
E-mail addresses: cuijianxun@hit.edu.cn (J. Cui), feng.liu@uhasselt.be (F. Liu), Jia.Hu@dot.gov (J. Hu), davy.janssens@uhasselt.be (D. Janssens), geert.wets@uhasselt.be (G. Wets), mario.cools@ulg.ac.be (M. Cools).

http://dx.doi.org/10.1016/j.neucom.2015.08.100
0925-2312/

urban planning in the context of the developed world. In contrast, the existing approaches have so far been very limitedly used in developing countries, mainly due to the high-cost as well as other limitations inherent to traditional travel data collecting methods [9,10]. The lack of planning tools in many cities in developing countries leads to a wide gap between the current urban and transport planning practices and what is needed to support the rapid growth and expansion of city development. As a result, transportation managers and urban planners find themselves still ill-adapted to the rapid development of the cities [11], despite the numerous public policy measures that have been introduced to improve land use structure and alleviate the pressure on the transport system.

There is a long way to go from basic-level transport management practices up to high-level intelligent transport and urban planning techniques that are particularly tailored to developing countries and capable of assisting decision makers in designing more effective measures. To bridge this gap, in this paper, a methodology is developed, which extends the current research on travel demand modeling and road network analysis by means of GPS technologies, and particularly addresses the limitations with respect to the development of reliable methods for the systematic identification of long time persistent transport problems. Current city-wide travel demand and road network systems are analyzed, and areas with serious mismatch problems are identified. Specifically, a set of indicators are developed, which measures both the spatial and temporal travel efficiencies between each pair of regions within the city, based on the travel demand, travel speed, and route directness of travel paths derived from the GPS data. The obtained indicators can then be used to identify serious road transport problems, thus assisting policy makers in seeking optimal solutions that best match the travel demand

Compared to traditional transport network analysis techniques, the new approach offers the following advantages. (i) It is built upon massive GPS data recorded from urban vehicles running in the city. The results derived from the travel data are thus capable of revealing the road conditions in a large part of the transport network. They also provide more objective and up-to-date measures, catching up with the fast pace of urban development and population growth. (ii) Based on the fine-grained GPS data, this method is able to build more spatially and temporally sensitive measures. (iii) Through a long period of data recording, the variations of mobility patterns between weekdays and weekends as well as across different months and seasons are captured. (iv) In many cities around the world today, GPS devices are already installed in taxis or private cars, generating no extra financial-cost in terms of data collection, making it a cost-effective approach and easily transferable to the cities. (v) In this study, the GPS data of all taxis operating in the city of Harbin is explored, providing a unique opportunity for the analysis on a significant share of individuals' trips over the total travel demand of the city. Compared with a number of existing studies on travel demand and road network analysis using GPS data, this study represents a significant novelty in terms of the systematic identification of region pairs with poor transport performance across an extensive area of the city as well as in terms of the in-depth examination into the specific road situations, the source of the traffic, and the particular travel routes of the trips between the detected regions. (vi) The proposed method develops a reality mining approach which places the realized trips of instrumented travelers directly at the center of the analytical process. This is particularly practical in developing countries where, as stated before, the data problems have deterred the much needed development of a new, effective and cheaply realized travel demand modeling and transport network analysis technique. From this perspective, using GPS data to model travel demand, examine transport network conditions and identify potential mismatch problems, will have significant societal impact, as urban mobility planning and services can be improved.

An overview of the existing state-of-the-art in terms of methods and data used for performing urban transport network analysis is provided in Section 2. In Section 3, the GPS data used for developing the analytical framework is described and the context of the case study is delineated. The details of the analytical framework are provided in Section 4, whereas the results of the application of the framework to the case study are presented in Section 5. Finally, Section 6 provides the discussion and conclusions.

2. Literature review

2.1. Travel demand modeling and transport network analysis

Travel demand modeling is a process of analyzing travel patterns in terms of where, when, how and why individuals travel, and estimating travel demand expressed by the number of vehicles or people that move between different regions or that use a specific transport facility like a roadway segment or a railway station. The established models can then be used for the analysis of transport network systems, for the evaluation of different transport and land use policy scenarios, and for the study on a variety of other transportation related issues, e.g. vehicle emission estimates and their impact on population health [12,13].

Travel demand models have been historically dominated by trip-based approaches, e.g. the four step models [14]. To address the behavioral inadequacy of this modeling process [15,16], these trip-based approaches were enhanced by activity-based methods, e.g. Albatross [17], TASHA [18] and Feathers [19]. Despite the advancement in the new modeling process, the activity-based models are still estimated using the same type of data as the traditional trip-based approaches, i.e. household *travel surveys*, which document the full daily activities and travel of a small sample of individuals during a time frame of one or several days. The use of such surveys has the following drawbacks which limit to a certain extent the applicability of the travel demand models calibrated with this data [9,10]. (i) The entire survey is a lengthy process. From the initial data gathering to data cleaning and the exploitation of the first results, it could take months even years, causing a time lag between the data initially obtained and the results that are required for objective and up-to-date activity-travel behavior analysis. (ii) The survey process imposes a significant burden on respondents, resulting in low response rates and under-reporting of short trips. (iii) The collected data has shortcomings in terms of both spatial and temporal resolutions. The granularity of the spatial systems is typically represented by a set of pre-defined zones, e.g. statistical sectors, each of which ranges from a few hundred meters to a few thousand in radius. Similarly, the temporal component is round up to a certain interval, e.g. 5 min, as people are unlikely to recall the exact times of the trips performed throughout the day. (iv) The data is very expensive to collect, leading to a small sample of respondents which report their travel behavior only during a single or a limited number of days. Consequently, this tends to obfuscate the less frequent activities, such as sports or telecommuting activities which are often carried out only once a week or once a month. Questions are also raised about the capability of such a limited sample size in representing activity-travel behavior of a full population.

Apart from travel surveys, travel information has also been gathered from sensors, e.g. loop detectors and video cameras, which are installed in a road network to monitor traffic flow. However, due to the high installation cost, the sensors are usually set up on highways, as it is expensive to instrument a whole city with such static devices. Consequently, traffic data is only limited