Transportation Research Part A 86 (2016) 35–55

Contents lists available at ScienceDirect

Transportation Research Part A

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

Evaluating the added-value of online bus arrival prediction schemes

Oded Cats [a,b], Gerasimos Loutos [b,*]

[a] Department of Transport and Planning, Delft University of Technology, The Netherlands
[b] Department of Transport Science, Royal Institute of Technology (KTH), Stockholm, Sweden

ARTICLE INFO

Article history:
Received 25 August 2015
Received in revised form 31 January 2016
Accepted 12 February 2016

Keywords:
Real-time information
Reliability
Waiting times
Public transport
Travel time prediction

ABSTRACT

Online predictions of bus arrival times have the potential to reduce the uncertainty associated with bus operations. By better anticipating future conditions, online predictions can reduce perceived and actual passenger travel times as well as facilitate more proactive decision making by service providers. Even though considerable research efforts were devoted to the development of computationally expensive bus arrival prediction schemes, real-world real-time information (RTI) systems are typically based on very simple prediction rules. This paper narrows down the gap between the state-of-the-art and the state-of-the-practice in generating RTI for public transport systems by evaluating the added-value of schemes that integrate instantaneous data and dwell time predictions. The evaluation considers static information and a commonly deployed scheme as a benchmark. The RTI generation algorithms were applied and analyzed for a trunk bus network in Stockholm, Sweden. The schemes are assessed and compared based on their accuracy, reliability, robustness and potential waiting time savings. The impact of RTI on passengers waiting times are compared with those attained by service frequency and regularity improvements. A method which incorporates information on downstream travel conditions outperforms the commonly deployed scheme, leading to a 25% reduction in the mean absolute error. Furthermore, the incorporation of instantaneous travel times improves the prediction accuracy and reliability, and contributes to more robust predictions. The potential waiting time gains associated with the prediction scheme are equivalent to the gains expected when introducing a 60% increase in service frequency, and are not attainable by service regularity improvements.

1. Introduction

Service reliability is one of the main determinants of public transport level-of-service. Public transport services are subject to several sources of uncertainty which results with deviations from the planned service. This is particularly true for bus systems where the difficulty to predict bus arrival times leads to longer waiting times, passenger dissatisfaction and higher operational costs. Public transport systems are increasingly equipped with information and communication technologies such as automatic vehicle location (AVL) and automatic passenger counts (APC). These systems were first deployed to support fleet monitoring and frequency determination but later also facilitated the generation and dissemination of real-time

* Corresponding author.
E-mail addresses: o.cats@tudelft.nl (O. Cats), loutos@kth.se (G. Loutos).

http://dx.doi.org/10.1016/j.tra.2016.02.004
0965-8564/

information (RTI) (TCRP, 2008). Information concerning next bus arrival is considered among the most important source of information and its accuracy is one of the main concerns among bus users (Caulfield and O'Mahony, 2007; Rahman et al., 2013).

Even though considerable research efforts were devoted to the development of bus arrival prediction schemes, which involved the application of computationally-intensive statistical methods, there is lack of knowledge on the performance of real-world RTI systems. While it is often postulated that the accuracy and reliability of RTI are fundamental in realizing its potential benefits, there is limited knowledge on how RTI systems perform in practice. In order to gain a better understanding of the current state of the practice, an empirical analysis of the performance of a commonly deployed timetable-based prediction scheme was conducted by Cats and Loutos (2015). The analysis indicated that the provisioned information deteriorates significantly along the line and becomes increasingly unreliable for prediction horizons longer than several minutes. The results of the empirical study call for the development of alternative schemes for generating RTI. These schemes will address the identified shortcomings, while using the performance of the current system as a benchmark.

The primary objective of this study is to evaluate the added-value of alternative online bus arrival prediction schemes. The added-value is defined in this study in terms of passenger waiting time savings and the capability to foresee downstream vehicle trajectories and thus facilitate the deployment of proactive operations management. The performance of the prediction schemes is assessed in terms of their accuracy, reliability and robustness. The added-value of the proposed schemes is appraised by benchmarking their performance against static information and a commonly deployed prediction scheme. Furthermore, the potential time savings associated with RTI dissemination and improvements, as compared with more expensive measures to improve service frequency and reliability, are examined. We implement and evaluate alternative link-based online prediction schemes.

Two state-of-the-art schemes are designed to enrich the state-of-the-practice by (a) embedding data on downstream travel times and, (b) considering the underlying mechanism that drives the bunching process, namely the impact of headway on dwell times through passenger flows. The schemes were inspired by the mathematical model of Newell and Potts (1964) for bus service reliability, which first established the fundamental relations between bus travel time elements. By analyzing prediction schemes that are directly applicable, rely on existing information sources, account for the prevailing control strategy, and do not involve any estimation or calibration techniques prior to their implementation, this study narrows down the gap between the state-of-the-art and the state-of-the-practice in generating RTI for public transport systems.

Following the literature review (Section 2), the formulation of alternative prediction schemes (Section 3) is presented. The schemes are applied to a trunk bus network in Stockholm, Sweden (Section 4). An analysis and evaluation of their performance, in terms of their accuracy, reliability, robustness and added-value in terms of waiting time savings, as compared with improvements in service provision is provided in Section 5. This paper concludes with a discussion of performance implications (Section 6), practical considerations and suggestions for further research (Section 7).

2. Literature review

There is an extensive literature on public transport prediction models. Since this study is concerned with the evaluation of methods to generate RTI, predictions methods are reviewed in Section 2.1 with a focus on their data requirements and online application considerations. In Section 2.2, approaches and findings related to measuring the impacts of RTI dissemination and its quality on service performance are reviewed.

2.1. Public transport prediction models

Predicting future public transport states, such as vehicle arrival and departure times, and on-board crowding levels, requires the collection, integration and process of instantaneous and historical data. Most research efforts were devoted to the development of bus arrival time predictions because of the importance of waiting times and the uncertainty associated with bus operations. Thus, vehicle arrival time constitutes the primary state to be obtained from prediction schemes. Previous studies applied various statistical and meta-heuristic methods for bus arrival predictions including: regression models (Patnaik et al., 2004; Chang et al., 2010), artificial neural networks (Jeong and Rilett, 2005; Mazloumi et al., 2011), Kalman filter (Cathey and Dailey, 2003; Shalaby and Farhan, 2004; Chen et al., 2004, 2005), support vector machines (Yu et al., 2011), genetic algorithms (Fadaei Oshyani and Cats, 2014) and statistical pattern recognition (Vu and Khan, 2010).

Regression models were used to predict the remaining travel time as function of independent variables. Patnaik et al. (2004) estimated a series of linear regression models for bus arrival times as function of distance, number of stops, passenger volumes and weather conditions. However, in general, variables in public transport operations are highly inter-correlated and exercise complex non-linear relations. These properties can therefore hinder the applicability of regression models. Non-parametric regression models such as the k-nearest neighbors proposed by Chang et al. (2010) can be effective when applied to large datasets which are underlined by non-linear relations. However, their performance is undermined by long computational time because of their reliance on large amounts of historical data.

Machine learning methods can be advantageous over statistical methods in predicting future states of the public transport system due to their capability to utilize large amounts of data, to reveal complex patterns and to address noise in data streams. Artificial neural networks offer an effective algorithm to generate outcomes from complex non-linear systems.