European Journal of Operational Research 240 (2015) 98–108

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Production, Manufacturing and Logistics

Finding optimal hyperpaths in large transit networks with realistic headway distributions

Qianfei Li, Peng (Will) Chen, Yu (Marco) Nie *

Department of Civil and Environmental Engineering, Northwestern University, United States

ARTICLE INFO

Article history:
Received 17 December 2013
Accepted 29 June 2014
Available online 23 July 2014

Keywords:
Hyperpath
Headway distribution
Greedy method
Enumeration
Erlang distribution

ABSTRACT

This paper implements and tests a label-setting algorithm for finding optimal hyperpaths in large transit networks with realistic headway distributions. It has been commonly assumed in the literature that headway is exponentially distributed. To validate this assumption, the empirical headway data archived by Chicago Transit Agency are fitted into various probabilistic distributions. The results suggest that the headway data fit much better with Loglogistic, Gamma and Erlang distributions than with the exponential distribution. Accordingly, we propose to model headway using the Erlang distribution in the proposed algorithm, because it best balances realism and tractability. When headway is not exponentially distributed, finding optimal hyperpaths may require enumerating all possible line combinations at each transfer stop, which is tractable only for a small number of alternative lines. To overcome this difficulty, a greedy method is implemented as a heuristic and compared to the brute-force enumeration method. The proposed algorithm is tested on a large scale CTA bus network that has over 10,000 stops. The results show that (1) the assumption of exponentially distributed headway may lead to sub-optimal route choices and (2) the heuristic greedy method provides near optimal solutions in all tested cases.

1. Introduction

Thanks to the revolution in information technology, many transit agencies now have the capability to track their entire fleets, make short-term projections, archive the data and distribute passenger information, all in real time. BusTracker of Chicago Transit Authority (CTA), for example, employs GPS-based automatic vehicle location (AVL) data to project the arrival times of the next transit vehicle at any stop on any route. Similar passenger information systems can be found in other major US cities, such as New York and Washington, DC. These new systems not only enable passengers to use transit information in the real time, but also make available a large amount of operational data that can be used to support better routing and planning decisions. For example, several studies have explored the possibility of using transit AVL data for the purpose of probing traffic conditions (Bertini & Tantiyanugulchai, 2004; Chakroborty & Kikuchi, 2004; Pu & Lin, 2008).

This paper will first show that these newly emerging data sources can be used to quantify the irregularities in transit services, in particular headway. Then, a transit routing algorithm that incorporates such empirically observed service irregularities will be implemented and tested to help passengers save travel time and improve travel reliability.

The transit routing algorithm developed in this study is built on the notion of *hyperpath*. A *hyperpath* represents a sequence of routing strategies rather than a simple path consisting of stops. Routing based on hyperpath promises to make better use of availability of alternative lines in the transit systems. It also offers the flexibility to incorporate real-time information, such as the arrival times of all transit vehicles approaching a stop. It is worth noting that the boarding decision at a stop depends on the waiting time as well as the remaining travel time to the destination once the selected transit line is boarded. This remaining travel time, in turn, is affected by future events such as waiting at subsequent transfers and travel between stops. As a result of service irregularities, the remaining travel time may not be reliably estimated based on the published schedule. Accordingly, decisions have to be made according to what are *likely* to happen in the future. In light of this observation, the proposed tool copes with uncertainty by choosing an optimal hyperpath to minimize the *expected* journey time.

Hyperpath based transit routing algorithms have been studied extensively in the literature; see the next section for a brief review. Unlike most existing algorithms, however, the proposed algorithm will operate on realistic headway distributions calibrated from archived trajectory data. It has been commonly assumed in the

* Corresponding author. Tel.: +1 847 467 0502.
E-mail address: y-nie@northwestern.edu (Y. (Marco) Nie).

http://dx.doi.org/10.1016/j.ejor.2014.06.046
0377-2217/

literature that headway is exponentially distributed. With this simplifying assumption all headway distributions can be fully characterized without using any trajectory data, since the standard deviation and mean of the exponential distribution are equal and can be reliably estimated using the scheduled headway. Moreover, exponentially distributed headway reduces the efforts for obtaining expected waiting times and boarding probabilities at transfers to closed form calculation.

To validate this assumption, the empirical headway data archived by CTA's BusTracker Application are fitted into various distributions. The results suggest that the headway data fit much better with Loglogistic, Gamma and Erlang distributions than with the exponential distribution. Accordingly, we propose to model headway using the Erlang distribution in the proposed algorithm. This choice reduces the calculation of expected waiting time and line boarding probabilities to manageable one-dimensional numerical integration. When headway is not exponentially distributed, finding optimal hyperpaths may require enumerating all possible line combinations at each transfer stop, which is tractable only if the number of alternative transit lines is small. To overcome this difficulty, efficient heuristic methods are implemented and compared to the exact method based on enumeration. The algorithms is tested on the CTA bus network, which has over 10,000 stops. We found, among other things, that the assumption of exponentially distributed headway may lead to sub-optimal route choices.

The rest of the paper proceeds as follows. Section 2 briefly reviews the literature on the hyperpath problem in transit routing. Section 3 presents the basic analysis of common-lines problem with general headway distributions, as well as the algorithms for finding optimal paths in a transit network. Section 4 describes the sources of headway data and discusses the fitting procedure and results. Section 5 presents the results of numerical experiments, including hyperpath routing in a large-scale CTA bus network. Section 6 concludes the paper.

2. Literature review

The concept of hyperpath appears to originate from the study of common bus lines by Chriqui and Robillard (1975). This seminal paper shows that passengers can select a set of attractive lines and board the first arriving bus in that set in order to minimize the expected total travel time. Spiess and Florian (1989) extends this notion of strategy to a general transit network, namely, the choice of an attractive set of lines is considered at each node where boarding occurs. Nguyen and Pallottino (1988) interpret the above strategy as a hyperpath, which is an acyclic directed graph. They propose both label correcting and label setting algorithms for finding the optimal hyperpaths between a pair of nodes. Volpentesta (2008) proposes a polynomial algorithm to solve one-to-all and one-to-one hyperpath problem. The above hyperpath routing model has been extended and incorporated by many into transit assignment (see e.g. Nguyen & Pallottino, 1988). These studies mostly focus on the interactions between transit route choice and congestion effects (overcrowding), which may be modeled through the effective frequency approach (Spiess & Florian, 1989; de Cea & Fernández, 1993; Wu, Florian, & Marcotte, 1994; Cominetti & Correa, 2001; Cepeda, Cominetti, & Florian, 2006), explicit capacity constraints (Marcotte & Nguyen, 1998; Hamdouch, Marcotte, & Nguyen, 2004; Teklu, 2008), failure-to-board probabilities (Kurauchi, Bell, & Schmöcker, 2003; Schmöcker, Bell, & Kurauchi, 2008; Schmöcker, Fonzone, Shimamoto, Kurauchi, & Bell, 2011), and the queueing theory (Trozzi, Hosseinloo, Gentile, & Bell, 2010; Trozzi, Gentile, Bell, & Kaparias, 2013). Nguyen, Pallottino, and Gendreau (1998) propose a transit assignment model that distributes flows on optimal hyperpaths using a logit model, which may be viewed as an extension of Dial's celebrated STOCH algorithm (Dial, 1979) in transit networks. The focus of their work, however, is to develop an efficient loading procedure to obviate path enumeration, rather than modeling congestion effects. More recently, DAcierno, Gallo, and Montella (2010) apply an ant-colony optimization (ACO) algorithm (see e.g. DAcierno, Montella, & De Lucia, 2006) to solve the hyperpath-based traffic assignment problem. Their study compares the efficiency of ACO algorithm with that of the Method of Successive Averages (MSA). We note that there is another class of models that perform transit assignment based on detailed schedule instead of average frequency. Since this paper is focused on the frequency-based approach, we refer the reader to Tong and Wong (1998) and Wilson and Nuzzolo (2008) for details of this line of work.

Bell (2009) adapts the hyperpath concept to model the reliable routing problem, by interpreting random road travel times as the analogy of the waiting times for transit lines. A hyperpath version of the popular A-star algorithm, known as the hyperstar algorithm, is developed and later extended to the time-dependent case (Bell, Trozzi, Hosseinloo, Gentile, & Fonzone, 2012). Schmöcker, Bell, Kurauchi, and Shimamoto (2009) show that the set of paths generated by a multi-agent, zero-sum game in a network with random road travel time is equivalent to the hyperpath of transit assignment. More recently, Ma, Fukuda, and Schmöcker (2012) explore and compare a variety of implementation issues associated with the hyperstart algorithm, and Noh, Hickman, and Khani (2012) integrate the concept of hyperpath with a schedule-based transit network representation. extend schedule-based transit model to a dynamic transit assignment problem where travel demand is time-dependent.

Central to the hyperpath finding problem is the calculation of expected waiting time at stops. Early studies indicate that the expected waiting time for a single transit line can be estimated from the mean and variance of headway (see e.g. Welding, 1957; Holroyd & Scraggs, 1966; Osuna & Newell, 1972; Seddon & Day, 1974). When multiple lines are present, the expected waiting time depends on the probability of taking each line, i.e. the route choice probability, which in turn is affected by the availability of information, routing strategy and headway distributions. In most literature, the route choice probability is derived from frequencies of each routes (see e.g. Spiess & Florian, 1989; Nguyen & Pallottino, 1988). Among the factors that influence route choice probability, the type of headway distributions affects the computation of expected waiting time the most. Conventionally, headway is assumed as exponentially distributed. Consequently, the expected waiting time can be computed in closed form, which greatly improves the computational efficiency in optimal hyperpath search. However, pure random arrival of transit vehicles implies that the expected waiting time for any given line equals its average headway regardless of passengers' arrival time, which seems overly conservative except for bus services with very small headway (O'Flaherty & Mangan, 1970).

Other types of headway distributions have been considered by several authors. Marguier and Ceder (1984) analyze the waiting time and route choice probabilities in a two-line example, assuming the headways follow either power or Gamma distribution. Hickman and Wilson (1995) model bus headway using lognormal distributions in a simulation study. Gendreau (1984) propose to approximate the line headway distributions by Erlang distribution, which is a special case of Gamma distribution but has better analytical tractability. Erlang distributions are also adopted in Bouzaïene-Ayari, Gendreau, and Nguyen (2001) and Gentile, Nguyen, and Pallottino (2005) to model route choice in the common-lines problem. More recently, Ruan and Lin (2009) fit a sample of observed headway data collected in Chicago to four different