Computers & Operations Research 67 (2016) 34–47

Contents lists available at ScienceDirect

Computers & Operations Research

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

Data-driven approaches for emissions-minimized paths in urban areas

Jan Fabian Ehmke [a,*], Ann Melissa Campbell [b], Barrett W. Thomas [b]

[a] Freie Universität Berlin, School of Business and Economics, Department of Business Information Systems, Garystr. 21, D-14195 Berlin, Germany
[b] University of Iowa, Tippie College of Business, Department of Management Sciences, Iowa City, Iowa, 52242, United States

ARTICLE INFO

Available online 25 September 2015

Keywords:
Emissions
Stochastic shortest path
Green logistics
Floating car data
Big data

ABSTRACT

Concerns about air quality and global warming have led to numerous initiatives to reduce emissions. In general, emissions are proportional to the amount of fuel consumed, and the amount of fuel consumed is a function of speed, distance, acceleration, and weight of the vehicle. In urban areas, vehicles must often travel at the speed of traffic, and congestion can impact this speed particularly at certain times of day. Further, for any given time of day, the observations of speeds on an arc can exhibit significant variability. Because of the nonlinearity of emissions curves, optimizing emissions in an urban area requires explicit consideration of the variability in the speed of traffic on arcs in the network. We introduce a shortest path algorithm that incorporates sampling to both account for variability in travel speeds and to estimate arrival time distributions at nodes on a path. We also suggest a method for transforming speed data into time-dependent emissions values thus converting the problem into a time-dependent, but deterministic shortest path problem. Our results demonstrate the effectiveness of the proposed approaches in reducing emissions relative to the use of minimum distance and time-dependent paths. In this paper, we also identify some of the challenges associated with using large data sets.

1. Introduction

Concerns about air quality and global warming have led to numerous initiatives to reduce emissions. As a significant source of CO_2 emissions, transportation has been an important focus of these efforts. Urban areas, where emissions are exacerbated by congestion, have been particularly active in looking at ways to reduce the environmental impact of transportation, especially freight transportation. For example, Amsterdam and London have introduced low-emission zones that restrict truck traffic near the city centers (see http://www.milieuzones.nl and http://www.tfl.gov.uk/roadusers/lez/default.aspx). Also, many companies have begun to focus on the CO_2 emissions resulting from the movement of goods, highlighting their efforts in annual reports [35] and press releases [36]. Generally, the focus is on minimizing the traveled distance. While minimizing the traveled distance contributes to minimizing emissions, there are other factors that impact emissions.

In general, CO_2 emissions are proportional to the amount of fuel consumed, and the amount of fuel consumed is a function of speed, distance, acceleration, and weight of the vehicle [10]. Optimizing for these factors, notably average speed, as is often the case in the academic literature, may not be feasible, particularly in urban areas. In urban areas, vehicles must often travel at the speed of traffic, and congestion can impact this speed particularly at certain times of day. These changes in speed have a significant impact on emissions, as noted by van Woensel et al. [37]. One of the more well-known models of emissions is the MEET model [23], which is illustrated in Fig. 1. This graph reveals the highly nonlinear relationship between speed and emissions, with particularly high emissions at slower speeds. At certain times of the day when traffic is at its peak and vehicles' speeds are reduced, a path that may yield low emissions at other times may suddenly yield high emissions. Thus, to optimize expected emissions in an urban area, it is critical to consider the speed of traffic on arcs in the network at different times of the day. This data is now increasingly available thanks to GPS equipment on taxicabs in nearly all major cities around the world [15]. The collection of speed observations leads to large historical databases containing speeds for a very detailed road network, which allow for a detailed analysis by route planners. It is well-known that congestion has recurring patterns for different times of the day, so we can use this data to estimate time-dependent speeds on different arcs on the network [16].

Experimentation with these data sets reveals that even at the same time of day, the observations of speeds on a given arc can

* Corresponding author. Tel.: +49 30 838 58731; fax: +49 30 838 458731.
E-mail addresses: janfabian.ehmke@fu-berlin.de (J.F. Ehmke), ann-campbell@uiowa.edu (A.M. Campbell), barrett-thomas@uiowa.edu (B.W. Thomas).

http://dx.doi.org/10.1016/j.cor.2015.08.013
0305-0548/

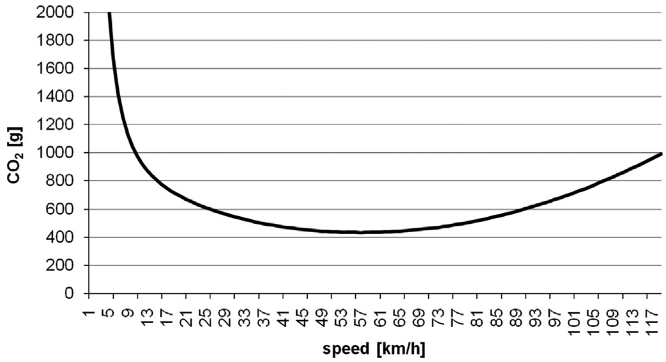

Fig. 1. Evolution of emissions for 7.5–16 ton vehicles (MEET model) for a distance of 1 km.

exhibit significant variability [3]. Due to the nonlinear relationship between speed and emissions, we cannot simply use an average speed on an arc at different times of day to choose the expected emissions-minimizing path between origin and destination. The average speed may not capture the average emissions that occur on a particular arc. For example, if the travel speed on an arc often drops far below the average speed, the actual emissions will be much higher than if travel occurs consistently at the average speed. Thus, to optimize emissions in an urban area, we must explicitly consider the variability in the speed of traffic on arcs in the network at different times of the day.

Motivated by the need to account for speed variation in the construction of expected emissions-minimizing paths between customers, this paper explores how to determine such paths. We propose two methods. Each method uses speed values from a historical database to determine time-dependent average emissions on each arc in a chosen path. In this way, we can ensure that our solution methods account for the impact of speed variation on emissions.

The first method, referred to as the *path-averaging method*, is a path construction method that uses sampling to evaluate the expected emissions associated with paths under construction. The use of sampling allows us to account for the impact of variability not only on emissions, but also on the arrival times to each node in a path. The expected emissions for a path are estimated based on averaging the emissions across all associated sample paths. To the best of our knowledge, our proposed sampling-based shortest path algorithm is the first to incorporate sampling in a path construction approach. This method can easily be generalized for applications beyond emissions, notably where arc costs are time-dependent and are evaluated via sampling.

Motivated by the computation time associated with the first method, the second approach ignores the variability in arrival times at each node. We refer to this approach as the *arc-averaging method*. In ignoring arrival time variability, we can operate on both expected time-dependent travel time and emissions values for each arc. Importantly, both of these values can be pre-computed. Further, operating on expected values reduces the problem to a deterministic one. Our computational experiments will examine the impact of this change both in the quality of the solutions found and in the runtimes.

We also compare both of our proposed methods with deterministic time-dependent travel time and distance-based shortest path methods commonly used in the literature. Our results demonstrate the value of accounting for speed variability when determining expected emissions-minimizing paths.

This paper makes an important contribution to the literature not only because it addresses how to minimize emissions in a way that makes sense in urban areas, but also because it discusses some of the challenges associated with using large data sets. When working with large data sets, it is important to derive the appropriate level of aggregation that is both computationally tractable and that preserves the most important characteristics of the problem. We present two ways of aggregation, namely the usage of sampling by retaining individual emissions values in path generation, and, alternatively, the usage of pre-computed, time-dependent emissions estimates. In this context, because it is computationally intractable to seamlessly integrate a real-world database of 230 million speed records based on standard database queries, we discuss how to overcome the technical challenges of using large data sets in optimization. Finally, we discuss the issue that, even when considering such an extensive number of speed records, there are some arcs in the network for which there is very little data at all.

The remainder of the paper is outlined as follows. We review related literature in Section 2. We formally define the problem in Section 3 and introduce our solution approaches in Section 4. We describe our implementation and experimental design in Section 5, which includes a discussion of the issues with using large data sets. Our computational experiments are presented in Section 6, and we discuss future work in Section 7.

2. Literature review

In the following, we provide an overview of the related literature. While the authors are unaware of any literature that explicitly addresses the question of computing shortest paths with emissions objectives, shortest paths have a long history in the literature, and we first review this literature. Next, we discuss various models of vehicle emissions. Finally, we discuss vehicle routing problems incorporating emissions objectives. Such problems are the original motivation for this work and represent an application that requires shortest paths as a subproblem.

2.1. Shortest path computation

Given an arc-dependent cost function, the idea of the shortest path problem is to determine the path of minimum cost between an origin and a destination. Arc costs may be based on travel time, distance, toll, energy consumption, or other considerations. Deterministic shortest path computation is a long and well-established field of research with the label-setting algorithm of Dijkstra [11] serving as the basis for nearly all modern shortest path algorithms. Recent computational advances are described by Geisberger et al. [21]. For path planning incorporating real-time information, we refer the reader to Güner et al. [22].

In shortest path computations for vehicles, the availability of detailed traffic information has led many authors to explore time-dependent shortest path problems (TDSPP). In the TDSPP, arc costs depend on the time that the arc is entered, and as a result, the minimum cost path may differ depending on the start time at the origin and the arrival time to the nodes en route. A key to many efficient solutions methods is the FIFO assumption. The assumption basically disallows overtaking, meaning that it is always better to arrive to a node earlier rather than later. With the FIFO assumption, the problem can be solved by modified variants of any deterministic label-setting or label-correcting shortest path algorithm. A discussion of relevant theoretical issues and a review of early literature can be found in Dean [6]. Discussion of the use of traffic data in developing time-dependent networks and a review of relevant literature can be found in Ehmke et al. [17]. Of particular interest to the work presented here is [4]. They demonstrate the effectiveness of the $A^\star$ shortest path algorithm (discussed more fully in Section 4) to the problem of finding shortest paths in