Computers and Operations Research 102 (2019) 67–74

Contents lists available at ScienceDirect

Computers and Operations Research

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

Total distance approximations for routing solutions

D. Nicola[a,*], R. Vetschera[a], A. Dragomir[a]

University of Vienna, Faculty of Business, Economics and Statistics Oskar-Morgenstern-Platz 1, Vienna 1090, Austria

ARTICLE INFO

Article history:
Received 16 January 2018
Revised 5 October 2018
Accepted 9 October 2018
Available online 10 October 2018

Keywords:
Distance estimation
Routing problems
Regression

ABSTRACT

In order to make strategic, tactical and operational decisions, carriers and logistic companies need to evaluate scenarios with high levels of accuracy by solving a large number of routing problems. This might require relatively high computational efforts and time. In this paper, we present regression-based estimation models that provide fast predictions for the travel distance in the traveling salesman problem (TSP), the capacitated vehicle routing problem with Time Windows (CVRP-TW), and the multi-region multi-depot pickup and delivery problem (MR-MDPDP). The use of general characteristics such as distances, time windows, capacities and demands, allows us to extend the models and adjust them to different problems and also to different solution methods. The resulting regression models in most cases achieve good approximations of total travel distances except in cases where strong random noise is present, and outperform previous models.

1. Introduction

Many exact and heuristic algorithms have been developed to solve transportation problems in short computational time. However, there are situations in which it is not necessary or possible to apply such algorithms. For example, carriers who need to respond to an inquiry by a potential customer or who are involved in a competitive bidding process need rapid information about costs, but not about the actual routing. This information must often be obtained very quickly for a large number of transportation requests, for example in a combinatorial auction, when bids must be made for many bundles of requests. Combinatorial auctions were frequently proposed as part of collaboration mechanisms to improve the efficiency of the transportation system (Berger and Bierwirth, 2010; Gansterer and Hartl, 2016).

Several approaches for approximating travel distances in transportation problems have been developed, which we will review in Section 2. Previous methods were mostly based on analytically derived approximation formulas (e.g., Beardwood et al., 1959), which were sometimes augmented by empirically estimated parameters (e.g., Christofides and Eilon, 1969). This approach limits the domain of problem classes for which approximations can be developed to comparatively simple problems, in which tour lengths can be approximated analytically. Only few authors (e.g., Çavdar and Sokol, 2015; Hindle and Worthington, 2004) so far considered a more empirical approach, but also limited their work to problems like the traveling salesman problem or vehicle routing problems without complex, real world constraints.

Our work takes an empirical perspective, initially considering a large set of (possibly redundant) potential variables, and then with the help of statistical methods identifying variables that are particularly useful in approximating total costs. Some of these variables, like distances between customers, are present in all routing problems. Other variables, like time windows and capacity constraints, are added to approximate the solution of problems containing those constraints. Previous studies have mostly reported the quality of approximations in terms of goodness of fit to the problem instances for which the model was estimated. We complement this in-sample evaluation with an out-of-sample evaluation, in which we analyze how well our models are able to predict total costs for new problem instances.

Our empirical approach requires actual solutions of some problems to estimate the parameter values. The approach therefore does not necessarily approximate the optimal solution to a problem, but the solution that can be found by the algorithm applied. For many of the applications envisioned, this is exactly the information that is required. For example, in a competitive bidding process, the price charged for performing a set of requests needs to cover the actual costs that will be incurred. If the algorithm used for generating the actual schedule delivers a solution that causes higher than optimal costs, these are the costs that need to be covered. Of course, having a better routing algorithm would make the

* Corresponding author.
E-mail addresses: daniel.nicola.naranjo@univie.ac.at (D. Nicola), rudolf.vetschera@univie.ac.at (R. Vetschera), alina-gabriela.dragomir@univie.ac.at (A. Dragomir).

https://doi.org/10.1016/j.cor.2018.10.008

carrier more competitive, but pricing should be based on the actual planning process.

Although this empirical approach is adaptable to different solution methods, this does not imply that the same empirical model can be applied to approximate solutions found by any algorithm. It is quite possible that the solution quality of different algorithms also depends on different characteristics of the problem, thus different empirical models will be needed to approximate solutions of the algorithms. However, the general method remains the same, even if other variables are included in the models.

In line with much of the existing literature (Daganzo, 1984; Figliozzi, 2009), we consider (variable) costs to be roughly equivalent to travel distance. We develop approximations for travel distance in different types of transportation problems of increasing complexity, ranging from a standard traveling salesman problem (TSP) to the recently introduced (Dragomir et al., 2018) Multi-Region Multi-Depot Pickup and Delivery Problem (MR-MDPDP). This gradual approach allows us to study whether the presence of additional constraints like vehicle capacity or time windows influences the quality of approximations in general. We also study how such factors can be represented in the approximation to make it more accurate even for a complex setting. In our view, the ability to incorporate additional problem characteristics into the approximation in a rather straightforward manner is a considerable advantage of an empirical approach. Thus, we consider it important to determine the potential benefit of adding such additional parameters to the model.

The remainder of the paper is structured as follows: In Section 2, we present a literature review on approximation models for different types of transportation problems. Section 3 introduces the regression based approximation for the different types of logistic problems we study in this paper. Computational experiments and results are presented in Section 4, and Section 5 concludes the paper by summarizing its results and presenting an outlook onto future research.

2. Literature review

Beardwood et al. (1959) are among the first authors who developed a distance approximation. They demonstrated that for a set of n nodes in a compact and convex area A, the length of the TSP tour asymptotically converges to $c\sqrt{nA}$ when $n \to \infty$, c being a constant. Christofides and Eilon (1969) approximated the average TSP route length by $100 \times c\sqrt{n}$. Further distance approximations for the TSP were developed by Chien (1992), Kwon et al. (1995), and Hindle and Worthington (2004). They all recognized the need for additional parameters representing the shape of the area or distances between customers. Chien (1992) modified the area factor A by including the area of the smallest rectangle that covers all customers and also includes distance related measures like the average distance to the depot. Kwon et al. (1995) used regression models and neural networks to improve the TSP approximations. They considered a rectangular length/width ratio as well as a shape factor. Hindle and Worthington (2004) used two different models depending on how customers are distributed over the area. Their first model considered a uniform distribution of locations. The second model used “demand surfaces” to represent different densities of customers across regions. In a recent paper on the approximation of TSP travel distances, Çavdar and Sokol (2015) proposed an approximation based on coordinates of customers. Their approximation was based on the standard deviations of coordinates as well as the standard deviations of distances between customers and the average-point of the area. Stein (1978) considered the bus problem, a TSP with additional constraints in which pickup and delivery nodes are paired. They extended the results of Beardwood et al. (1959) and estimated the length of the optimal tour either for one or for several buses.

The first published approximations for the capacitated vehicle routing problem (CVRP) were developed in the 1960s by Webb (1968), who studied the correlation between route length and customer-depot distance. Eilon et al. (1971) proposed approximations to the length of the CVRP based on the shape and area of delivery, distances between customers and the depot, and capacity of vehicles. Daganzo (1984) approximated CVRP tour length as

$$CVRP(n) \approx 2rn/Q + 0.57\sqrt{nA} \tag{1}$$

where n is the number of customers customers, r is the average distance between the customers and the depot and Q is the maximum number of customers that can be served by a vehicle. Robusté et al. (2004) tested Daganzo’s approximation and proposed adjustments based on the shape of the area, in particular for elliptic zones. Erera (2000) extended Daganzo’s approximation for stochastic versions of the CVRP. Langevin and Soumis (1989) developed an approximate method for planning pickup and delivery zones assigned to vehicles. Their approach included an estimation of the increase in the number of vehicles used if zones are allowed to overlap.

Time window constraints were first considered in the late 1980s by Daganzo (1987) by dividing the time horizon in periods and clustering the customers in rectangles. The problem was simplified by assigning customer time windows to a time period. A more recent contribution for VRP’s with time windows is provided by Figliozzi (2008), who approximated the distance for serving n customers using a known number m of routes as

$$VRP(n) \approx b\frac{n-m}{n}\sqrt{An} + m2r. \tag{2}$$

The parameter b was estimated by linear regression. Figliozzi (2009) studied approximations to the average length of VRP’s when the number of customers, time window constraints and demand levels vary. A detailed literature review of models that use continuous approximation in distribution management can be found in Franceschetti et al. (2017).

3. Regression based approximations

This section introduces the problems that are studied in this paper and describes the variables that are included in the estimation models presented in Section 4. We consider three classes of problems in increasing order of complexity: The traveling salesman problem (TSP), the capacitated vehicle routing problems with time windows (CVRP-TW), and the multi region multi-depot pickup and delivery problem (MR-MDPDP).

These problems are also different with respect to the solution methods that can be applied. For small TSP, an optimal solution can be found via exact methods, for the MR-MDPDP, only one heuristic is available. This allows us to study how well the regression approach can adapt to solutions and methods of different quality. We also study this question in more detail for the CVRP-TW, where we apply the regression approach both to known optimal solutions and to solutions obtained with a simple heuristic.

3.1. Traveling salesman problem

The first models for the TSP approximated the total tour length using the factor $\sqrt{nA}$. Since the total travel distance depends on the distance between nodes, we include different distance measures in the estimation model. To represent different distributions of nodes in the area, we also include dispersion measures such as the variance of distances and maximum distances to the average-node. Based on the approach proposed in Çavdar and Sokol (2015),