Transportation Research Part C 55 (2015) 191–202

Contents lists available at ScienceDirect

Transportation Research Part C

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

Optimality versus run time for isolated signalized intersections

David K. Hale [a,*], Byungkyu Brian Park [b], Aleksandar Stevanovic [c], Peng Su [d], Jiaqi Ma [d]

[a] University of Florida, PO Box 116585, Gainesville, FL 32611-6585, United States
[b] University of Virginia, 351 McCormick Road, Charlottesville, VA 22904, United States
[c] Florida Atlantic University, 777 Glades Road, Building 36, Room 225, Boca Raton, FL 33431, United States
[d] Leidos, Inc., 11251 Roger Bacon Drive, Reston, VA 20190, United States

ARTICLE INFO

Article history:
Received 13 October 2014
Received in revised form 2 January 2015
Accepted 17 February 2015
Available online 7 March 2015

Keywords:
Traffic signal timing
Simulation-based optimization
Genetic algorithm
Simulated annealing
Heuristic method
Tabu search

ABSTRACT

Simulation-based optimization of traffic signal timing has become pervasive and popular, in the field of traffic engineering. When the underlying simulation model is well-trusted and/or well-calibrated, it is only natural that typical engineers would want their signal timing optimized using the judgment of that same model. As such, it becomes important that the heuristic search methods typically used by these optimizations are capable of locating global optimum solutions, for a wide range of signal systems. However off-line and real-time solutions alike offer just a subset of the available search methods. The result is that many optimizations are likely converging prematurely on mediocre solutions. In response, this paper compares several search methods from the literature, in terms of both optimality (i.e., solution quality) and computer run times. Simulated annealing and genetic algorithm methods were equally effective in achieving near-global optimum solutions. Two selection methods (roulette wheel and tournament), commonly used within genetic algorithms, exhibited similar effectiveness. Tabu searching did not provide significant benefits. Trajectories of optimality versus run time (OVERT) were similar for each method, except some methods aborted early along the same trajectory. Hill-climbing searches always aborted early, even with a large number of step-sizes. Other methods only aborted early when applied with ineffective parameter settings (e.g. mutation rate, annealing schedule). These findings imply (1) today's products encourage a sub-optimal "one size fits all" approach, (2) heuristic search methods and parameters should be carefully selected based on the system being optimized, (3) weaker searches abort early along the OVERT curve, and (4) improper choice of methods and/or parameters can reduce optimization benefits by 22–33%.

1. Introduction

Simulation-based optimization of traffic signal timing has become pervasive and popular, in the field of traffic engineering. This is mainly due to the robustness and reliability of prominent traffic analysis tools, in terms of their ability to evaluate existing conditions. The traffic analysis tools used for signal optimization usually employ detailed analytical and/or macroscopic simulation models. Despite their complexity it is possible to perform hundreds of model runs within seconds, given the speed of today's computers. This is quite helpful for signal optimization, because each of these model runs can be used to

* Corresponding author at: Leidos, Inc., 11251 Roger Bacon Drive, Reston, VA 20190, United States.
E-mail addresses: david@ce.ufl.edu, david.k.hale@leidos.com (D.K. Hale), bpark@virginia.edu (B.B. Park), astevano@fau.edu (A. Stevanovic), peng.su@leidos.com (P. Su), jiaqi.ma@leidos.com (J. Ma).

http://dx.doi.org/10.1016/j.trc.2015.02.015
0968-090X/

test a different timing plan alternative. Although the microscopic simulation models are also trusted, their computer run time requirements often preclude their direct involvement in the optimization process.

Optimization tools such as PASSER (TTI), TRANSYT (TRL), TRANSYT-7F (McTrans), Synchro (Trafficware), HCS-Streets (McTrans), Vistro (PTV), and ArteryLite (AGA) employ variations of simulation-based optimization. One crucial advantage of simulation-based optimization is the ability to simulate existing conditions, and perform model calibration based on those existing conditions, prior to any design or optimization. When the underlying simulation model is well-trusted and/or well-calibrated, it is only natural that typical engineers would want their signal timing optimized using the judgment of that same model. Only simulation-based optimization can provide this.

1.1. Heuristic search methods used for traffic signal timing

As such, it becomes important that the heuristic search methods typically used by these optimizations are mathematically capable of locating global optimum solutions, for a wide range of signal systems. However, what we see in practice is that the popular software solutions only offer a limited subset of the available search methods. Moreover, the software solutions only offer a limited subset of the available parameters, which can often improve the efficiency of those methods. The result is that many signal optimization efforts are likely experiencing premature convergence on mediocre solutions, because they involved ineffective methods and/or parameters. Some traffic signal timing studies (Oda et al., 1996) have shown genetic algorithms to be more effective than hill-climbing methods, and one signal timing study (Agbolosu-Amison et al., 2009) has shown genetic algorithms to be more effective than Harmony search. However no known traffic signal timing studies have examined many other available search methods and parameters.

For example, the effectiveness or viability of heuristic search methods such as simulated annealing and Tabu search in the signal optimization process appears to be unknown. No known traffic signal timing optimization studies have compared the effectiveness of tournament selection and roulette wheel selection, within the genetic algorithm; or the number of steps and step-sizes, within the hill-climbing method. There is a need to perform apples-to-apples comparisons of all these search methods; by applying them to the same underlying evaluation model, across a wide variety of intersection geometries and signal phasing. Finally, no known signal timing studies have focused on the ramifications of widely-varying computer run times required by these various methods.

1.2. Importance of computer run times

The issue of computer run times appears to be particularly underrated. Although many off-line optimizations complete within seconds, these efforts are likely producing sub-optimal solutions for small networks, analyzed across a small number of time periods. Engineers would care more about computer run times if they knew that (a) other available heuristic search methods would often provide better designs for such networks and (b) future optimizations may require larger traffic networks, numerous time periods, and micro-simulation of all candidate timing plans. Regarding real-time optimizations, run times are important because timing plans are changing so frequently. This highlights an incentive to maximize the ratio of optimality (i.e., solution quality) over run time.

The trade-off between optimality and run time becomes clear when working with these heuristic search methods. For example, it is common to observe significant improvement in the first fraction of run time, followed by modest improvements requiring a much longer period of run time. It is common to observe the fastest-converging search methods producing reasonable and effective timing plans, whereas longer-running (without converging) search methods arrive much closer to the global optimum solution. It would be ideal if a new search method could obtain global optimum solutions in the shortest run times, but such a method has been elusive.

Given the trade-off between run times and optimality, it stands to reason that different optimization projects would benefit from different search methods. Faster methods (e.g. hill-climbing search, greedy algorithms, equalizing degree of saturation) are needed for real-time adaptive control, whereas slower methods (e.g. genetic algorithms, simulated annealing, particle swarm optimization) are more acceptable for off-line applications. Faster methods are needed to process larger networks across numerous time periods, whereas slower methods are more acceptable for small datasets. Different engineers have different standards regarding how much time they are willing and able to wait, during an off-line (non-adaptive) optimization run. Faster methods are needed to directly optimize computationally expensive simulation models, whereas slower methods are quite effective for simple capacity analysis (analytical) models.

1.3. Scope of this paper

To shed more light on the trade-offs between solution quality and time needed, this paper compares optimality and computer run times for several top optimization search methods from the literature. Given that no optimization method outperforms all others under all conditions (Ciuffo and Punzo, April 2014), this paper will not try to determine the "best" method for all conditions. Instead the paper will present evidence that some optimization methods, when used with the right parameters, have a strong tendency to avoid premature convergence. To produce this evidence, over a thousand test optimizations of isolated intersections were performed using one objective function, and one underlying traffic model.