TRAFFIC SIGNAL TIMING CONTROL FOR A SMALL-SCALE ROAD NETWORK

Jiann-Shiou Yang
Department of Electrical and Computer Engineering
University of Minnesota
Duluth, MN 55812
USA

Abstract

This paper presents a pilot study of the development and evaluation of a traffic signal timing control for a small-scale road network in downtown Duluth, Minnesota. The study mainly focuses on the alleviation of a sudden traffic flow surge following special events (e.g., concerts, conventions, hockey games, etc.) held at the city's Entertainment and Convention Center. A practical approach for signal timing control that eliminates the need of using traffic flow model is used to optimize the intersection traffic light split times. Our approach is based on neural networks (NNs) with the weight estimation via the Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm. Based on the traffic data collected and the measure-of-effectiveness (MOE) criterion used, the NN weights are adjusted by use of the SPSA algorithm that minimizes the MOE criterion. The performance evaluation using the existing signal timing plan and the one generated by the SPSA algorithm is also compared. The results show the potential to apply this method to control the signal timing over a large-scale road network.

Key Words: Signal Timing Control; Optimization; Stochastic Approximation; Neural Networks; Measure-of-Effectiveness.

1. Introduction

Congestion is one of the most serious problems in today's highway traffic. Improving the timing of the traffic signals at intersections along arterials (or in a network) is generally considered as a very cost-effective means to reduce traffic congestion. The current computer controlled signal systems can be classified into two categories: fixed time control and traffic-responsive control. The fixed time control uses signal timing plans computed off-line using past data (e.g. [1]). One of the quite popular methods is to use TRANSYT for signal timing optimization but the design is based on the average flows with the actual traffic fluctuations not taken into consideration. The traffic-responsive control computes the signal plans according to the prevailing traffic flow [2]. These control methods include, for example, the OPAC system (Optimized Policies for Adaptive Control) [3], SCOOT (Split, Cycle, and Offset Optimization Technique) [4,5], SCAT system (Sydney Coordinated Adaptive Traffic System) [6], etc. The application of using neural networks (NNs) to traffic control has also been reported by many researchers (e.g., [7,8]). These NN-based control strategies require a model for the traffic dynamics, which is usually constructed off-line using past system data. The model is usually a set of differential/difference equations, but it may also be a neural network or non-equation types of models such as fuzzy associative memory or rule-based expert system. Whatever the type of model used, it is serving as a representation of the effect of the signal timings on the traffic flow in the network. However, due to numerous difficult-to-model interactions such as construction blockages, seasonal variations in flow patterns, etc., it is very desirable to develop effective traffic signal timing for an arterial or a road network without using traffic dynamic model. This is also true for small urban traffic flow in downtown Duluth, Minnesota.

Following special events (e.g., conventions, concerts, hockey games, etc.), at the Duluth Entertainment Convention Center (ECC), high volumes of traffic exiting the area create substantial congestion at adjacent intersections. The goal of this study is to provide an effective traffic signal timing control for the high volume traffic movements associated with these special events so that progression through the downtown Duluth and interstate I-35 is as efficient as possible. Our research mainly focuses on the study of after event traffic flow data over a 30-minute time period at key intersections exit the area and then develops an efficient traffic signal timing plan to reduce intersection delay and improve traffic flow. A model-free approach using neural networks (NNs) with the weight estimation via the Simultaneous Perturbation Stochastic Approximation (SPSA) method/algorithm is used [9,10,11]; and a performance evaluation study using the existing signal timing plan and the one developed by the SPSA algorithm is also examined. The signal timing developed here is based on the assumption that all signalized intersections are operated in coordinated mode, and the timing plan only applies to a short period of time (e.g., 30 minutes) immediately following special events. The duration time depends on the traffic volumes related to special events. The primary goal of this study is to provide an improved coordinated signal timing plan at key intersections for the high volume traffic surge associated with special events to reduce congestion.

2. Traffic Data

Based on the field observations at the ECC after special events, we identified six intersections which have more impact on the alleviation of traffic surge in our project study area. In this preliminary study, only the results at the three intersections along the 5th Avenue West section (i.e., Superior Street/5th Avenue West, I-35N/5th Avenue West, I-35S/5th Avenue West) will be reported. Special events of our interest during the period of October 2002-May 2003 included: weekly activities (i.e., University of Minnesota Duluth (UMD) Men's and Women's Hockey games, concerts) and annual events (i.e., Shrine Circus, UMD and College of St. Scholastic graduation ceremonies). Based on the selected intersections, the traffic data was collected using TT-2 and TT-4 road tube counters. The manual counts were also conducted mainly for the purpose of counting the turning movements at these intersections. Except graduation ceremonies, all the special events we studied were held on Fridays and Saturdays from 7 to 9 pm. The data was recorded over a short period of time, roughly 30 minutes, immediately after these events. Similar events data was further averaged and then converted to hourly traffic counts. Among the total 21 times over this period, 12 were for the UMD Men's and Women's Hockey games, 5 for the concerts and circus, and 4 for the graduation ceremonies.

3. The MOE Criterion

The essential part of any control optimization is to determine an adequate measure-of-effectiveness (MOE) calculation that will achieve the specified goal. The MOE used here is the tolerance index which is based on the difference between the nominal throughput during the green phase and the actual traffic data [11]. The nominal throughput is a traffic count calculated from the departure rate (in vehicles per second) on the road segment multiplies the duration time of the green phase. The intent is to use the road capacity and the traffic volume on the road segment in conjunction with the signal setting to minimize the traffic interruption caused by the traffic signal.

Consider at a traffic signal light, let v be the departure rate (in vehicles per second) at speed limit on the road segment facing the signal, t_g be the duration time of the green phase, l be the traffic counts from road tubes (i.e., the number of vehicles, recorded by road tubes, crossing the signal light over the time period t_g), and T be the tolerance index for that signal; then T is defined as $(v\, t_g - l)/l$. That is, the value of the tolerance index associated with the traffic at the selected intersection is mainly calculated from the difference between the nominal throughput and the actual data collected at that road segment. The overall tolerance index is then the summation of the T's for the entire network over the pre-specified time period immediately after special events. Slightly different from the one used in [11], the index value used here actually shows the capacity deviation from the nominal throughput. Obviously, the smaller the index is, the more satisfied the drivers will be. This further implies the lower possibility for vehicles to wait for the red signal.

4. The NN-Based SPSA Algorithm

Generally speaking, an optimization process is a step-by-step procedure for changing the adjustable parameters from some initial guess to a value that offers an improvement in the objective function. Many optimization algorithms have been developed that assume that the gradient associated with the objective function is available. That is, the first derivative of the objective function with respect to the parameters being optimized is available. However, there has been a growing interest in recursive optimization algorithms that do not depend on the gradient information or measurements (e.g., [9,12]). On the contrary, these algorithms are based on an approximation to the gradient formed from noisy measurements of the objective function. A main advantage of the gradient-free algorithms is that they do not require the detailed information of the functional relationship between the parameters being adjusted and the objective function being optimized. Such a relationship can be difficult to develop in some areas, whereas in other areas, there may be large computational savings in calculating an objective function relative to that required in calculating a gradient. It is also quite common that direct measurements of the gradient are not always available or sometime even difficult to obtain in many optimization problems. In contrast, the approaches based on gradient approximations require only conversion of the basic output measurements to sample values of the objective function, which does not require full knowledge of the system input-output relationships. Generally speaking, gradient-free stochastic algorithms exhibit convergence properties similar to the gradient-based stochastic algorithms. The SPSA algorithm is based on forming a succession of highly efficient approximations to the un-computable gradient of the objective function in the process of finding the optimal weights [9,10]. The approximation used only requires observed values of the system, not a model for the system dynamics. Its procedure in the general recursive form can be written as

$$\theta_{k+1} = \theta_k - c_k\, g_k(\theta_k)$$

where c_k is a scalar gain, θ_k is a vector of parameters to be adjusted at iteration k, $g_k(\theta_k)$ is the estimate of the gradient $\partial L/\partial\theta$ at $\theta = \theta_k$, and $L(\theta)$ is the MOE criterion (or the objective function). The above equation simply states that the new estimate (i.e., θ_{k+1}) of θ equals to the previous estimate (i.e., θ_k) plus an adjustment that is proportional to the negative of the gradient estimate. Assume that the parameter vector θ is of dimension p, then the i^{th} component of the gradient estimate $g_k(\theta_k)$ at $\theta = \theta_k$ is calculated as follows:

$$g_{ki}(\theta_k) = [Ł(\theta_k + \varepsilon_k\Delta_k) - Ł(\theta_k - \varepsilon_k\Delta_k)]/2\, \varepsilon_k\Delta_{ki}$$
$$(i = 1, 2, \ldots, p)$$

where $Ł(\bullet)$ represents an observed value of $L(\bullet)$, $\Delta_k = (\Delta_{k1}, \Delta_{k2}, \ldots, \Delta_{kp})$ is a vector of random variables that satisfy certain important regularity conditions [9], and ε_k is a small positive number. The regularity conditions are required to guarantee that the parameter vector θ_k converges to the optimal solution as k becomes large. Obviously, the numerators of the vector $g_k(\theta)$ are all identical; only the denominators are different. Notice that all elements of θ_k are randomly perturbed to obtain two measurements of Ł. This is the difference between the simultaneous perturbation approximation and the finite difference approximation. It is also clear that to find $g_k(\theta)$ only two values of $Ł(\bullet)$ are needed. The following procedure shows how the SPSA algorithm [10] is implemented with some minor modifications:

Given an initial weight vector θ_0, update θ_k to θ_{k+1}, (k = 0, 1, 2, 3, …) as follows:

1. For the current weight estimate θ_k, change it to $\theta_k + \varepsilon_k\Delta_k$.
2. Throughout the given time period, use a NN control $u(\theta, \bullet)$ with weights $\theta_k + \varepsilon_k\Delta_k$. Inputs to $u(\theta, \bullet)$ at any time within the period; include current state information.
3. Form the objective function $Ł(\theta_k + \varepsilon_k\Delta_k)$ based on the simulated system behavior.
4. During the following same time period, i.e., the following date and time of the same special event, repeat steps 1-3 with $\theta_k - \varepsilon_k\Delta_k$ replacing $\theta_k + \varepsilon_k\Delta_k$. Form $Ł(\theta_k - \varepsilon_k\Delta_k)$.
5. With the information from steps 3 and 4 on $Ł(\theta_k + \varepsilon_k\Delta_k)$ and $Ł(\theta_k - \varepsilon_k\Delta_k)$, form the gradient estimate $g_k(\theta_k)$, and then take one iteration to update the value of θ_k to θ_{k+1}.
6. Repeat steps 1-5 with the new θ_{k+1} replacing θ_k until traffic flow is optimized based on the chosen MOE.

5. Implementation

Our approach is based on neural networks (NNs) serving as the basis for the timing control with the weight estimation (i.e., the “training process”) via the SPSA algorithm. The signal timing adjustments is designed to accommodate short-term traffic volume surge exiting the ECC following special events. The NN serves to approximate the true, but unknown, mathematical function representing the optimal signal controller. This controller takes information from the traffic counts and then produces the signal timing to optimize the pre-specified MOE. The SPSA algorithm provides values for the NN weights without using a traffic model. The block diagram showing how they are related is given in Fig. 1. In our study, we used a two-hidden-layer, feed-forward neural network with 20 input nodes and 6 output nodes. The two hidden layers have 6 and 4 nodes, respectively. The 6 output nodes represent the green/red splits of the six key intersections near the ECC. And the 20 input nodes are divided into two groups: (a) the 14 inputs representing the averaged traffic volumes (vehicles per hour divided by cycles per hour) from external nodes to the networks, and (b) the 6 inputs (vectors of timing splits) from the previous green/red splits. Besides the control module which implements the NN controller, our special event signal timing optimization algorithm, written in C++, consists of the SPSA algorithm, traffic flow simulation module, and the MOE evaluation module [13].

6. Signal Timing Optimization

In general, there are eight possible phases at every intersection, although all of them are not always used. The eight phases are required to accommodate the eight movements (four through and four left turns) at the intersection. In our case, there are two different phase plans (i.e., 5-phase and 3-phase) currently used at the three intersections. Phase numbers are the labels assigned to the individual movements around the intersection. It is common to assign the main street through movements as phase 2 (Φ2) and phase 6 (Φ6), and use odd numbers for left turn signals and even numbers for through signals (i.e., the National Electronics Manufacturing Association (NEMA) standards for signal controller [14]). However, we found that none of the three intersections we studied follow the NEMA phase numbering convention. We contacted the City of Duluth traffic signal engineer regarding the signal timing plans used at the test site and found that the cycle length during the weekdays is 80 seconds, except the time period from 3:30 pm to 7:00 am (with 100 seconds cycle length). The cycle length is the total time to complete one sequence of signalization around an intersection. We let the cycle length remain at 80 second and also kept several data unchanged. These include: the minimum initial, all-red time, vehicle extension time [15], minimum gap time [15], walk and don't walk times (if any phase contains a pedestrian phase). In addition, we set the total lost time to four seconds for all of the intersections we studied. The total lost time is a combination of the recognition time and all-red time. Recognition time is the time it takes an average driver to recognize the signal light has turned green, which is usually about two seconds.

Since the traffic pattern exit the ECC area is very similar, the signal timing presented here can be applied to all ECC related events over a short period of time (e.g., 30 minutes). The time of duration depends on the size (i.e., traffic volume) of the events. This timing plan was presented to the City of Duluth traffic signal engineer. After discussions, it was agreed that a 15% minimum should be placed on the total split times for each of the phases. For example, the split times of both the north and south bound (i.e., Φ4) on Superior Street and 5th Avenue West intersection was re-adjusted (the split time generated by SPSA was 13% due to low traffic volume recorded). When adjusting certain phases (usually left turn phase) to this minimum, adjustments had to be made so that the total split times for all the phases in an intersection does not exceed 100%. The results are summarized in Table 1.

Table 1 shows the original split times used by the City of Duluth (if the coordinated mode is available), the optimized split times generated by both the Synchro software [15] and the SPSA algorithm, and the suggested times after re-adjustments. The software Synchro 5.0 is also used to aid in the signal optimization to compare the results generated by our SPSA algorithm. The Synchro 5.0 software is one of the most used signal timing optimization and traffic analysis software tools [15]. In other words, the column labeled "Original (%)" represents the split times (in percentage of the cycle length) currently used by the city, the third and fourth columns represent the total split times generated by Synchro and SPSA, respectively; and the last column labeled "Suggested (%)" represents the split times we suggested after our consultation with the City of Duluth traffic signal engineer. Note that there are two intersections with split times marked "----" under the column labeled "Original (%)" in Table 1, which simply means that the signal timing only operates in "free" mode (no coordinated mode available). For clarity, the corresponding phase numbering scheme used at each individual intersection is also given. Note that all through traffic also allow right turn movements (in the same direction) unless otherwise specified.

7. Performance Evaluation

The performance evaluation using the original timing plan and those generated by the Synchro software and the NN-based SPSA algorithm is studied. The evaluation is based on two MOE criteria; one is the tolerance index mentioned in Section 3 when using the SPSA algorithm, and the other is to measure the total delay per vehicle exit the ECC after special events. The comparison of the performance index on 5th Avenue West together with the overall system performance measures (i.e., combining both the 5th Avenue West section and the Railroad Street section) is also summarized in Table 2.

Since the tolerance index is used as our MOE, obviously, the SPSA algorithm generates the best results (i.e., the lowest value) in terms of that index measure (see Table 2). The performance measure was also conducted based on the average total time delay incurred for each vehicle exit the network after events. In terms of this index measure (i.e., total delay per vehicle), it seems that Synchro produced the best results. Since we only have the original timing splits for the intersections at Superior Street/5th Ave West, Railroad Street/Canal Park Drive and I-35/ Lake Avenue, the performance index we found by the original splits of these three intersections is 37.21, while the performance index by SPSA is 35.48. A 4.65% decrease in the performance measure shows the new traffic signal timing is superior to the current one used to alleviate the traffic surge following ECC special events. The future work includes a short-term travel time prediction [16,17] using Kalman filtering technique to monitor and gauge the quality of service of the routes exit the ECC area following special events. In other words, we will address the critical issues: how easy is it to exit the ECC area? how much does that "ease of movement" vary after ECC special events? That is, the mobility monitoring and performance measures between the travel conditions during the peak period (i.e., the time period immediately following special events) and off peak period (i.e., no-event daily flow situation) will be further investigated.

8. Conclusion

Following special events at the ECC, high volumes of traffic exiting the area create substantial congestion at adjacent intersections. The purpose of this research is to provide an effective signal timing control for the high volume traffic movements associated with ECC special events to reduce congestion. A practical approach for signal timing adjustments that eliminates the need of using traffic dynamic model is presented to optimize the signal timing. Our approach is based on neural networks with the weight estimation (i.e., the "training process") via the SPSA algorithm. Based on the data collected, the NN-based control operates and makes signal timing adjustments to accommodate the traffic conditions. The NN weights are determined by use of the SPSA parallel estimation algorithm at the six intersections following special events. That is, we set up an SPSA optimization algorithm that allows for updating of the values of weights at each key intersection we studied. The time period chosen depends on the estimated size of the events/activities occurring at the ECC. Convergence was obtained when the MOE in terms of the tolerance index has been optimized subject to intersection capacity, minimum traffic light cycle length, etc. The results from SPSA algorithm are compared with those generated using the Synchro software. After consultation with the City of Duluth traffic signal engineer, our suggested split times (in percentage) following ECC special events are then presented together with the performance evaluation measures. The main goal of this case study is to provide a more practical signal timing control, without using any traffic flow and system model, to address frequent occurrences of congestion and provides efficient traffic progression immediately following ECC events. We believe that the timing plan suggested here should help to move traffic exit the ECC area more efficiently.

9. Acknowledgments

This study was conducted with funding provided by NATSRL. The author would like to thank Paul Scanlan and Mars Cyr, from the City of Duluth Traffic Service Center, for providing us traffic data counters and whose help was also instrumental to the traffic signal timing development in this study. The help from Mr. Nan Zhang, Mr. Feng Qian, and Mr. Nan Yang was also greatly appreciated.

References

[1] Skabardonis and A. D. May, Comparative analysis of computer models for arterial signal timing, *Transportation Research Record,* 1021, Transportation Research Board, National Research Council, Washington, D.C., 1985, 45-52.

[2] Workshop on *Adaptive Traffic Signal Control Systems*, TRB Signal Systems Committee-A3A18, *80th Transportation Research Board Annual Meeting*, Washington, D.C., January 7, 2001.

[3] N. H. Gartner, OPAC: A demand-responsive strategy for traffic signal control, *Transportation Research Record,* 906, Transportation Research Board, National Research Council, Washington, D.C., 1984, 75-81.

[4] D. I. Robertson and R. D. Bretherton, Optimizing networks of traffic signals in real time – the SCOOT method, *IEEE Trans. Vehicular Technology*, 40, February 1991, 11-15.

[5] D. I. Robertson, Research on the TRANSYT and SCOOT methods of signal coordination. *ITE Journal*, January 1986, 36-40.

[6] P. R. Lowrie, The Sydney coordinated adaptive traffic system – principles, methodology, algorithms, Proc. *Int'l Conf. Road Traffic Signaling*, 207, March 1982, 67-70.

[7] M. Dougherty, H. Kirby, and R. Boyle, The use of neural networks to recognize and predict traffic congestion, *Traffic Engineering and Control*, 1993, 311-314.

[8] T. Nataksuji and T. Kaku, Development of a self-organizing traffic control system using neural network models, *Transportation Research Record,* 1324, Transportation Research Board, National Research Council, Washington, D.C., 1991, 137-145.

[9] J. C. Spall, Multivariate stochastic approximation using a simultaneous perturbation gradient approximation, *IEEE Trans. Automatic Control*, 37, March 1992, 332-341.

[10] J. C. Spall and D. C. Chin, A model-free approach to optimal signal light timing for system-wide traffic control, *Proc. 33rd IEEE Conference on Decision and Control*, 1994, 1868-1875.

[11] D C. Chin, J. C. Spall, and R. H. Smith, Evaluation of system-wide traffic signal control using stochastic optimization and neural networks, *Proc. 1999 American Control Conference,* 1999, 2188-2194.

[12] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications* (New York, NY: Springer-Verlag , 1997).

[13] J.-S. Yang, Duluth entertainment convention center special events traffic flow study, *NATSRL Final Report*, Center for Transportation Studies, University of Minnesota, twin Cities, June 2003.

[14] *Mn/DOT traffic Signal Timing and Coordination Manual*, Mn/DOT Office of Traffic Engineering and Intelligent Transportation Systems, June 2002.

[15] *Synchro 5.0 User Guide*. Trafficware Corporation, Albany, CA, 2002.

[16] H. Suzuki, T. Nakatsuji, Y. Tanaboriboon, and K. Takahashi, A neural-Kalman filter for dynamic estimation of origin-destination travel time and flow on a long freeway corridor, *Transportation Research Record*, 1739, Transportation Research Board, National Research Council, Washington, D.C., 2000, 67-75.

[17] W. H. Lin, A. Kulkarni, and P. Mirchandani, Arterial travel time estimation for advanced traveler information systems, *82nd Transportation Research Board Annual Meeting*, Washington, D. C., January 2003.

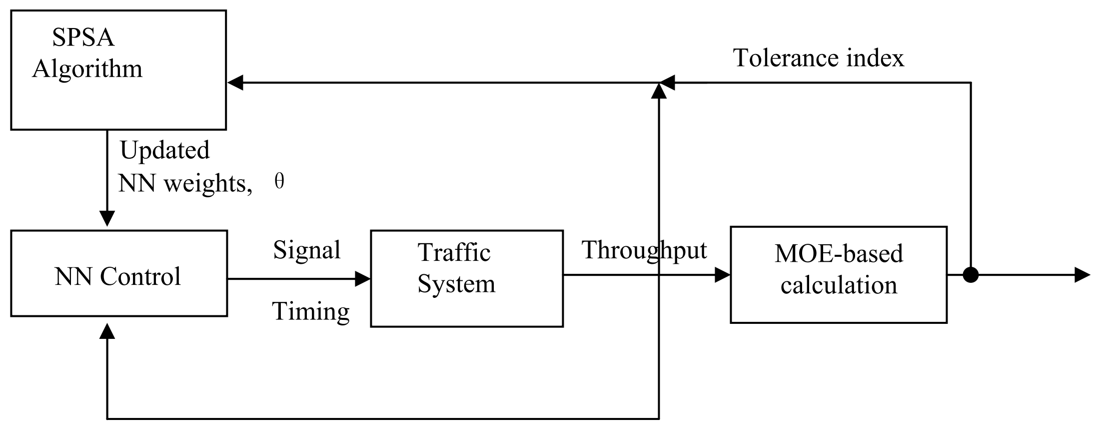

Fig. 1 Block Diagram of the NN-Based SPSA Signal Timing Optimization.

5th Avenue West and I-35 N				
Phase	Original (%)	Synchro (%)	SPSA (%)	Suggested (%)
Φ1	----	60	54	55
Φ2	----	10	16	15
Φ4	----	30	30	30
5th Avenue West and I-35 S				
Φ3	----	33.75	26	25
Φ5	----	18.75	49	50
Φ6	----	47.5	25	25
5th Avenue West and Superior Street				
Φ1	15	35	23	15
Φ2	45	35	64	50
Φ4	40	30	13	35
Φ5	15	35	23	15
Φ6	45	35	64	50

("-----" means signal timing only operates in free mode; no coordinated mode available)

Table 1 Comparison of the Total Split Times at the Signalized Intersections

Performance Measures on the 5th Avenue West section				
MOE	Original	Synchro	SPSA	Suggested
$\rho(vt_g - l)/l$	23.70	38.62	23.23	23.24
Delay/vehicle	47.10	48.40	53.70	43.70
Overall System Performance Measures				
MOE	Original	Synchro	SPSA	Suggested
$\rho(vt_g - l)/l$	50.24	61.17	48.51	48.55
Delay/vehicle	103.10	83.90	94.00	85.60

Table 2 Comparison of the Performance Measures