

Determination of Optimal Successor Function in Phase-based Control Using Neural Network

S.C. Wong, W.H. Law and C.O. Tong
Department of Civil and Structural Engineering
The University of Hong Kong
Pokfulam Road, Hong Kong

Phone +852-2859-1964; Fax +852-2559-5337; E-mail hhecwsc@hkucc.hku.hk

Abstract

A phase-based method for fixed-time signal control of traffic improves significantly the junction performance over conventional stage-based method of control due to the higher flexibility in specification of signal timings, where the control variables comprise the start and duration of green phases and the cycle time at which the junction is operated. The cycle-structure is specified by a successor function, a combination of 0 and 1 for all incompatible pairs of phases, which indicates the order of phases in a cycle. Normal procedure optimises the timings for each of these successor functions to determine the best timing plan. The computing time is found to be approximately proportional the number of such functions. To reduce the computational requirement, and hence enhance its applicability to real-time actuated control, a neural network is employed in this paper to help identify the optimal successor function for further optimisation of timings. Encouraging results are obtained.

1. Introduction

A phase-based method for fixed-time signal control of traffic at individual junction has recently received much attention and has been shown to improve significantly the junction performance over conventional stage-based method of control (Gallivan and Heydecker, 1988; Heydecker and Dudgeon, 1987). The control variables in phase-based method comprise the start and duration of green phases and the cycle time at which the junction is operated. Heydecker (1992) used a successor function to specify the cycle-structure in phase-based control for a signal-controlled junction, and proposed a procedure to identify the set of distinct successor functions for the junction. Optimisation is then

carried out for all successor functions to determine the optimal phase-based signal timings for the junction. The computing time is approximately proportional to the number of successor functions, which could be tremendous for complicated junctions. For real-time operation, accelerated procedure to identify the optimal successor function is highly desirable to reduce the computing time requirement, and optimisation is then only needed to apply to the chosen successor function.

In this paper, a neural network approach is employed to determine the optimal successor function given the traffic flow pattern at the junction as input. The artificial neural networks have been widely used in many areas of research (Wasserman, 1989, 1993), and more recently applied to solve various transportation problems (Faghri and Hua, 1992; Yang et al, 1992; Chin et al, 1992; Bullock et al, 1992; Kaseko et al, 1992; Bullock et al, 1992; Kaseko et al, 1992; Ritchie et al, 1992; Dougherty and Joint, 1992; Yang et al, 1992). Unlike conventional approaches, the neural network identifies certain relationship between the dependent and independent variables in a complex system, without the need to pre-specify the form of relationship. The process of establishing such relationship is called training in neural networks. Plenty of methods for training a neural network are available in the literature, among them the method of back-propagation is employed in the present study. It was found that this method of training provided a simple but yet effective way of producing the desirable results.

An example junction is used to demonstrate the effectiveness of the proposed methodology. Two sets of traffic flow patterns, one set for training of the neural network and the other for validation purpose,

are randomly generated for an example junction and optimisation results are obtained through exhaustive optimisation on all flow patterns using a standard phase-based optimisation package known as SIGSIGN (Sang and Silcock, 1989). In this example, the junction performance is measured by the junction delay incurred by all vehicles approaching the junction. The optimal successor function for delay-minimisation is then identified for each flow pattern. A neural network for this problem is formed with one hidden layer of neural nets, where the traffic flow pattern is the input vector and the optimal successor function as a combination of 0 and 1 is the output.

After trained by the first set of data using the back-propagation algorithm, the second set is used to test the performance of the neural network. Encouraging results are obtained. For over 60% of the cases the network is able to reproduce the optimal successor functions, the rest reproduce sub-optimal functions, but none of the cases it fails to reproduce a valid successor function.

2. The Successor Function

The successor function is defined in Heydecker (1992) to specify the cycle-structure of a phase-based signal timing plan in a junction. The function is a collection of the numbers of 0 and 1 for all incompatible pair of phases. A value of 0 implies that the start of green of one phase follows that of the other in a reference cycle and a value of 1 means the opposite. Different combinations of 0 and 1 in the successor function denote different order of phases and hence the cycle-structure at the junction. For example, Figure 1 shows the layout of a Chapel Hill junction (which is extracted from the example junction in Heydecker (1992)), a signal timing plan for the junction consisting of 9 phases, and the successor function for the plan. The function is represented in matrix form with values of 0 or 1 in the position corresponding to the pair of incompatible phases. Note that by definition a value of 1 in a particular position must have a value of 0 in the transpose position, and vice versa, and therefore only half of the matrix (either upper or lower triangular part) is sufficient to define the successor function. In other words, the number of 0 or 1 values in a successor function is equal to the number of incompatible pair of phases in the junction.

For a junction with m pairs of incompatible phases, there is a total of 2^m 0-1 combinations of successor functions. Out of all these possibilities, a large proportion of them are invalid cycle-structures. Tully (1976) developed a efficient generation method to eliminate these invalid combinations. Heydecker (1992) further eliminated the cyclic permutation redundancy to form a set of distinct successor functions for a junction. For the example junction shown in Figure 1, such eliminations reduce the number of successor functions to six combinations. It was also pointed out in Heydecker (1992) that a slight modification of the junction layout could push to the number of functions to over a hundred.

For conventional phase-based calculations, each of these combinations is optimised by solving a linear programming problem for capacity-maximisation and a convex programming problem for delay-minimisation. The computing time is approximately proportional to the number of successor functions, which could be tremendous for a complicated junction due to the large number of combinations. For real-time operation where quicker way of identifying optimal control strategy is highly desirable, a neural network approach to determine the optimal successor function without the need to go through the enumeration process is proposed in the following section. This accelerates the procedure of obtaining optimal signal timing plan, since optimisation is only performed once for the chosen successor function.

3. The Neural Network

Figure 2 shows the connection scheme of a typical multi-layer feed-forward neural network, which consists of three layers: an input layer, a hidden layer and an output layer. Between the input and hidden layers, the neurons are connected by a set of links with a vector of weights u , where the element u_{ij} is the weight of the connection between the neuron i in the input layer and the neuron j in the hidden layer. Let the input vector be q , where the element q_i represents the traffic flow of approach i in the signal-controlled junction used as input value to neuron i in the input layer. The intermediate signal of a neuron j in the hidden layer h_j is obtained by

$$h_j = F\left(\sum_i q_i u_{ij}\right) \quad (1)$$

Figure 1 An example junction.

Figure 1a The layout.

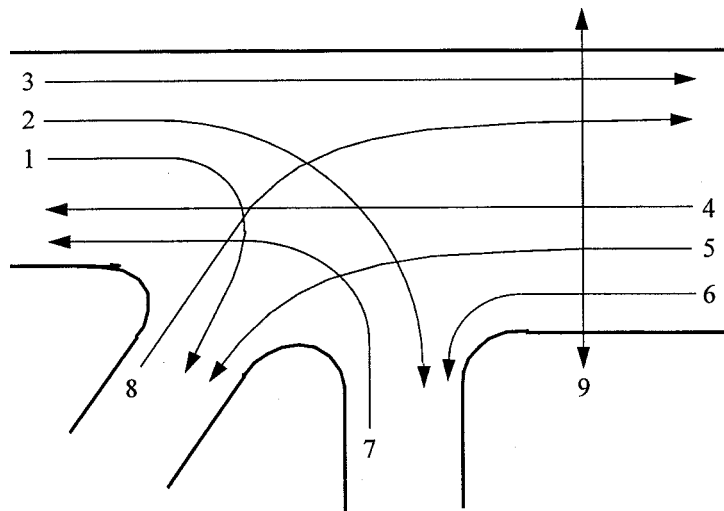


Figure 1b A signal timing plan.

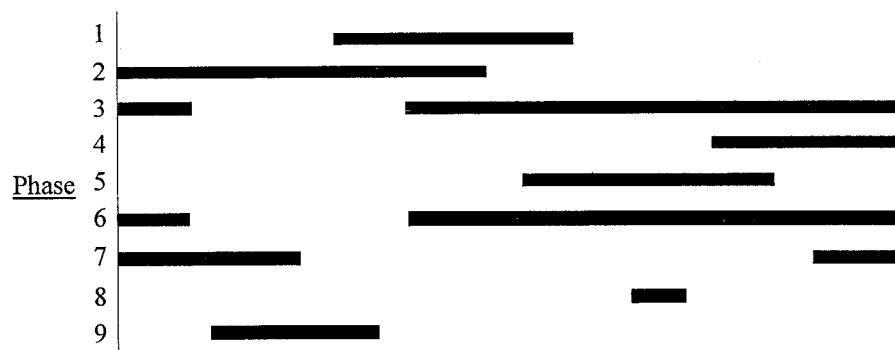
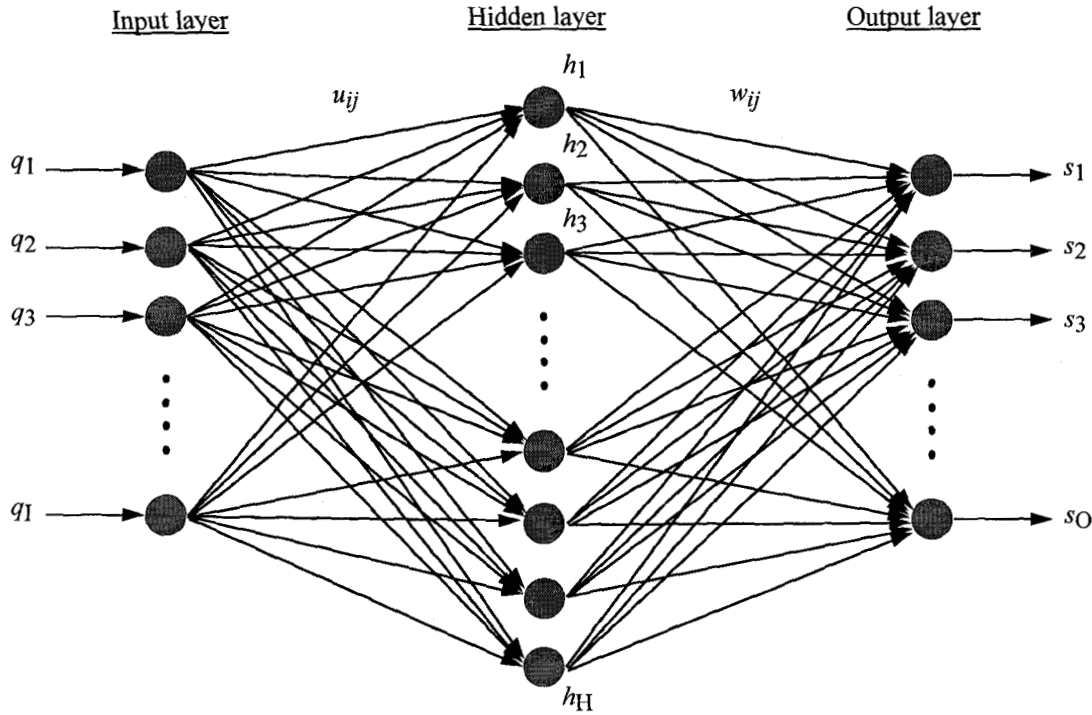


Figure 1c The successor function.

Phase	1	2	3	4	5	6	7	8	9
1				0			0	0	
2				0	0			0	
3									1
4	1	1						1	1
5		1					0		1
6									1
7	1				1			1	
8	1	1		0			0		1
9			0	0	0	0		0	

Figure 2 A typical multi-layer neural network.



where F is an activation function which scales and smoothes the intensity of a signal. In this paper, a sigmoidal activation function is used (Wasserman, 1989) which takes the form:

$$F(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

This sigmoid function transforming the signal to a range (0, 1) has a desirable property that

$$\frac{dF}{dx} = F(1 - F) \quad (3)$$

which is very useful for the training process discussed in the next section. The intermediate signal is the output signal from the connection between input and hidden layers, and forms the input signal to the connection between hidden and output layers. Let the vector of the weights between hidden and output layers be w , where w_{jk} is the weight of the link between neuron j in the hidden layer and neuron k in the output layer. The output signal of the neuron k in the output layer, s_k , is determined similarly by

$$s_k = F\left(\sum_j h_j w_{jk}\right) \quad (4)$$

Since a successor function is a combination of discrete values of 0 and 1, the element in the successor function corresponding to the k th pair of incompatible pair of phases is determined by the value of s_k using a threshold barrier. If s_k exceeds the threshold, the element takes a value of 1; and 0 otherwise. In this paper, the threshold barrier is taken as 0.5 which is a typical choice for sigmoidal activation function.

With the above description of the neural network, the number of neurons in the input layer is equal to the number of approaches in the signal-controlled junction (excluding the pedestrian streams). The number of neurons in the output layer is the number of pairs of incompatible phases in the junction. There is no stringent rule in neural computing on the number of neurons in the hidden layers because it is largely problem dependent. In this application, a

number equal to the total number of neurons in input and output layers works reasonably well.

4. A Back-propagation Algorithm

In this section, a back-propagation algorithm is proposed to train the neural network. Let \mathbf{t} be the target vector for a particular training set. The total error produced by a forward pass with the current sets of weights is specified as

$$E = \frac{1}{2} \sum_k (s_k - t_k)^2 \quad (5)$$

where t_k is the k th element of the vector \mathbf{t} . To train the set of weights \mathbf{w} between hidden and output layers, a descent direction can be identified using the derivative of total error with respect to the weight,

$$\frac{\partial E}{\partial w_{jk}} = h_k s_k (1 - s_k) (s_k - t_k) \quad (6)$$

To reduce the total error, the weight w_{jk} is modified by

$$w_{jk}^{(n+1)} = w_{jk}^{(n)} - \eta h_k s_k (1 - s_k) (s_k - t_k) \quad (7)$$

where $w_{jk}^{(n+1)}$ and $w_{jk}^{(n)}$ are the modified and current weights respectively, and η is the training rate coefficient (typically 0.01 to 1.0). To train the set of weights \mathbf{u} between the input and hidden layers, again the descent direction is identified. The derivative is now determined by

$$\begin{aligned} \frac{\partial E}{\partial u_{ij}} &= \sum_k (s_k - t_k) \frac{\partial s_k}{\partial u_{ij}} \\ &= \sum_k (s_k - t_k) s_k (1 - s_k) \sum_{j'} w_{j'k} \frac{\partial h_{j'}}{\partial u_{ij}} \end{aligned} \quad (8)$$

Since

$$\frac{\partial h_{j'}}{\partial u_{ij}} = \begin{cases} h_j (1 - h_j) q_i & \text{if } j' = j \\ 0 & \text{otherwise} \end{cases}, \quad (9)$$

we have

$$\frac{\partial E}{\partial u_{ij}} = q_i h_j (1 - h_j) \sum_k s_k (1 - s_k) (s_k - t_k) w_{jk} \quad (10)$$

To reduce the total error, the weight u_{ij} is modified by

$$\begin{aligned} u_{ij}^{(n+1)} &= u_{ij}^{(n)} \\ &\quad - \eta q_i h_j (1 - h_j) \sum_k s_k (1 - s_k) (s_k - t_k) w_{jk} \end{aligned} \quad (11)$$

where $u_{ij}^{(n+1)}$ and $u_{ij}^{(n)}$ are the modified and current weights respectively, and η is the training rate coefficient. The procedure for training the weights \mathbf{u} and \mathbf{w} is repeated for all training sets of data.

5. An Example

The signal-controlled junction shown in Figure 1 is used as an example to demonstrate the effectiveness of the proposed method. There is a total of 8 approaches of vehicular traffic to the junction which is controlled by 9 phases (including the pedestrian phase) with 14 pairs of incompatible phases. The numbers of neurons in the input and output layers are therefore 8 and 14 respectively. The number of neurons in the hidden layer is taken as 20 in this example. 400 sets of random flow patterns are generated, and for each of them the delay-minimising optimal successor function is determined using the computer package SIGSIGN (Sang and Silcock, 1989). The first 200 sets of data are used to train the neural network, and the remaining 200 sets of data are used to test the effectiveness of the method in reproducing the optimal successor functions. In this example, encouraging results are obtained. For over 60% of the cases the network is able to reproduce the optimal successor functions, the rest reproduce sub-optimal functions, but none of the cases it fails to reproduce a valid successor function.

6. Conclusion

In this paper, the neural network approach to determine the optimal successor function in phase-based control for an individual signal-controlled junction has been proposed. A neural network is constructed to accept the traffic flow pattern of the junction as input and produce the optimal successor

function as output. A back-propagation algorithm has been used to train the network, and encouraging results have been obtained. The most promising findings are that for the example junction the trained neural network always produces a valid successor function (and hence a valid cycle-structure for phase-based control) and provides a fairly high success rate in reproducing the optimal successor function. This enhances the applicability of the phase-based control method in real-time control of traffic. In response to the traffic flow pattern measured from detectors, better timing plan in association with the optimal successor function obtained from the neural network can be determined in a much quicker way.

Acknowledgement

This research was supported by a research grant (344/061/0007) from the University Research Committee of the University of Hong Kong.

Reference

- Bullock D. Garrett Jr J, Hendrickson C. and Pearce A. (1992) A neural network for image based vehicle detection. In Ritchie S.G. and Hendrickson C. (eds) *Artificial Intelligent Applications in Transportation Engineering*, pp 235-252.
- Chin S.M., Hwang H.L. and Miaou S.P. (1992) Transportation demand forecasting with a computer-simulated neural network model. In Ritchie S.G. and Hendrickson C. (eds) *Artificial Intelligent Applications in Transportation Engineering*, pp 349-361.
- Dougherty M. and Joint M. (1992) A behavioral model of driver route choice using neural networks. In Ritchie S.G. and Hendrickson C. (eds) *Artificial Intelligent Applications in Transportation Engineering*, pp 91-110.
- Faghri A. and Hua J. (1992) Roadway seasonal classification using neural networks. In Ritchie S.G. and Hendrickson C. (eds) *Artificial Intelligent Applications in Transportation Engineering*, pp 363-381.
- Gallivan S. and Heydecker B.G. (1988) Optimising the control performance of traffic signals at a single junction. *Transportation Research*, **22B**, 357-370.
- Heydecker B.G. (1992) Sequencing of traffic signals. In Griffiths J.D. (eds) *Mathematics in Transport and Planning and Control*, pp 57-67, Clarendon Press, Oxford.
- Heydecker B.G. and Dudgeon I.W. (1987) Calculation of signal settings to minimise delay at a junction. *Proceedings of 10th International Symposium on Transportation and Traffic Theory*, MIT, July, pp 159-178, Elsevier, New York.
- Ritchie S.G., Cheu R.L. and Recker W.W. (1992) Freeway incident detection using artificial neural networks. In Ritchie S.G. and Hendrickson C. (eds) *Artificial Intelligent Applications in Transportation Engineering*, pp 215-234.
- Sang A.P. and Silcock J.P. (1989) SIGSIGN user manual. Steer Davies and Gleave Ltd and Transport Studies Group, University College London.
- Tully I.M.S.N.Z. (1976) Synthesis of sequences for traffic signal controllers using techniques of the theory of graphs. PhD Thesis, *OUEL Report*, **1189/77**, University of Oxford.
- Yang H., Akiyama T. and Sasaki T. (1992) A neural network approach to the identification of real-time origin-destination flows from traffic counts. In Ritchie S.G. and Hendrickson C. (eds) *Artificial Intelligent Applications in Transportation Engineering*, pp 253-269.
- Yang H., Kitamura R., Jovanis P.P., Vaughn K.M. and Abdel-Aty M.A. (1993) Exploration of route choice behavior with advanced traveler information using neural network concepts. *Transportation*, **20**, 199-223.
- Wasserman P.D. (1989) *Neural Computing, Theory and Practice*. New York: Van Nostrand Reinhold.
- Wasserman P.D. (1993) *Advanced Methods in Neural Computing*. New York: Van Nostrand Reinhold.