

Towards Distributed Adaptive Control for Road Traffic Junction Signals using Learning Classifier Systems

L Bull¹, J Sha'Aban², A Tomlinson¹, JD Addison² and BG Heydecker²

¹Faculty of Computing, Engineering, and Mathematical Sciences
The University of the West of England, Bristol, UK
larry.bull@uwe.ac.uk

²Centre for Transport Studies, University College London, London, UK
ben@transport.ucl.ac.uk

1. Introduction

This chapter considers an approach to distributed traffic responsive signal control using Learning Classifier Systems (Holland, 1976). The intention is to accommodate realistic kinds of detector data and wide ranges of candidate performance criteria for traffic management in a fully flexible manner. The approach to achieving this is to use evolutionary computing (eg Holland, 1975) and reinforcement learning (eg Sutton and Barto, 1998) with performance fed back from microscopic traffic simulations: this approach has the advantage that it is not specific to any particular objective or form of primary data. The purpose of this work is to develop an approach to distributed optimisation that can achieve good traffic performance flexibly according to any on a range of possible criteria using data from existing traffic detectors. Here each junction in a road network is controlled by a Learning Classifier System using only locally available input and performance data; a multi-agent approach is proposed.

Learning Classifier Systems (LCS) can be used for optimisation in a way that offers substantial promise for application in traffic-responsive signal control systems where the way in which the control responds to variations in traffic flows can be adapted according to measured conditions. This is important in order to achieve traffic control that is sufficiently flexible to respond rapidly when traffic conditions change in a fundamental way, as occurs at the start of a peak period, without being unduly sensitive to short-term variations in flow. The expectation is that this will be possible by their use of both reinforcement learning and evolutionary computing techniques. Furthermore, they offer the automated rule development of neural networks together with the transparency of production system rules.

The importance of this approach for traffic control is that it offers a means by which signal control strategies can be developed directly according to their performance, evaluated using detailed microscopic simulation as opposed to that estimated from formulae that have been adopted on grounds of analytical convenience. This closed-loop approach to development of control strategies offers several advantages over the use of traditional explicit optimisation formulations. These include flexibility in respect of objectives so that multiple and varying needs can be

accommodated, ability to use various different kinds of detector data according to their availability, and freedom from dependence on a single explicit evaluation formula that is intended to embody the whole of a traffic model. This final point has been found to be especially important in recent research work where certain fine details of the models used have been found to have an unexpectedly strong influence on performance.

2. Road Traffic Junction Control

Strategies for signal control of road traffic have developed from fixed-time, with the possibility of selection between various precalculated plans, to responsive, in which the signal timings that are implemented vary according to traffic flows at the time of implementation. Substantial benefits have been achieved by progressing on each of the two fronts of enhancing the responsiveness of traffic control systems and extending advanced optimisation approaches from isolated road junctions to road networks with a high density of signal-controlled junctions. The approaches that have been developed successfully for responsive control at isolated junctions include heuristics such as rule-based approaches (Van Zuylen, 1976), optimisation approaches such as MOVA (Vincent and Peirce, 1988), and more flexible approaches that respect the uncertainties inherent in the data and models of traffic behaviour by using fuzzy logic (Chiu, 1992; Mieden et al, 1996). Mikami and Kakazu (1993) have successfully developed a hierarchical combination of junction-based learning using a reinforcement method with network-wide search by a genetic algorithm (GA)(Holland, 1975). Here the reinforcement learning is used to generate a single set of rules to control traffic at each of the individual junctions, while the genetic algorithm optimises rules across the network using these identical junction controllers. Sayers et al (1999) have used a genetic algorithm to develop fuzzy logic rules to operate signals at individual junctions in accordance with multiple objectives. Montana and Czerwinski (1996) used genetic programming (Koza, 1991) to evolve a mobile computational agent/program for the control of a simple network. The evolved solution is a LISP tree that indicates whether or not an alteration is appropriate at each traffic signal.

Early investigations of traffic responsive systems for road networks (Holroyd and Hillier, 1969; 1971) demonstrated the difficulty in improving on the performance that could be achieved by good fixed-time control systems such as TRANSYT (Robertson, 1969; Vincent et al, 1980). Since then, research has led to the development of various successful responsive systems that have now become established. These include the SCOOT system (Hunt et al, 1981) which uses a feed-forward approach to plan for the arrival at a junction of traffic that is detected as it leaves an upstream junction, SCATS (Lowrie, 1982) which uses feed-back from stop-line detectors to inform on queue exhaustion, and OPAC (Gartner et al, 1983, 1991) which uses a rolling horizon approach to dynamic optimisation. In this contribution the utility of LCS in controlling junctions in a distributed framework, using only local performance metrics, will be examined.

3. Learning Classifier Systems

A number of investigators have examined the use of LCS in multi-agent environments. Bull et al (1995) describe the use of classifier systems for the control of simulated robots, where each wheel/leg is represented by a separate system. Carse et al (1995) have used fuzzy classifier systems for routing at each node of a telecommunications network. In Potter et al (1995) an agent is represented by a number of classifier systems and a speciation-like process is included to improve performance. Multiple classifier systems have been used by Dorigo and Schnepf (eg 1992) to control an autonomous robot. Seredynski et al (1995) examined the use of local reward sharing in a simple iterated game, and Wyatt and Bull (2002) incorporated rule sharing between controllers for simulated mobile robot control. After (Arthur, 1990) and (Holland and Miller, 1991), a number of researchers have used classifier systems to represent traders in artificial markets. Marimon et al (1990) use classifier agents exchanging and consuming goods to examine the emergence of equilibria in a well-known triangular market. Palmer et al (1994) describe the use of classifier systems to simulate agents creating portfolios, by predicting the value of a stock. Dwormann (1994) has investigated coalition formation in a three-player game/market and Morengo and Tordjman (1996) used a classifier-based system to model belief formation in a market place. Bull (1999) represented traders in an artificial continuous double-auction market using LCS.

In this work we begin by using the ZCS (Wilson, 1994) system. The default parameters presented for ZCS, and unless otherwise stated for this paper, are: $N = 400$, $S_0 = 20$, $\beta = 0.2$, $\gamma = 0.71$, $\tau = 0.1$, $\chi = 0.5$, $\mu = 0.01$, $p = 0.25$.

4. Simple Road Traffic Network Simulator

The initial simulator that was used in the present study represents a small network of four crossroads junctions, each controlled by a separate set of traffic signals (Figure 1). For convenient reference, we suppose that the roads are oriented north-south and east-west. Traffic arrives stochastically with specified mean rate at the points of entry to the network, whilst traffic flows within the network are profiled by upstream signals. This enables a number of development tests to be conducted on the LCS for this application area. Vehicle behaviour is specified according to certain simple rules:

- All vehicles are standard cars.
- All cars have the same desired speed.
- All cars travel straight ahead at junctions.

The signals at each of the four junctions in this test network are arranged with two stages, one permitting north-south and the other east-west movements. This means that for each junction, the controller decisions are required to determine the duration of each of the two stages. Each junction is controlled by an LCS that receives as stimulus a 10-bit binary string, which is split into two components of equal length.

Each of these 5-bit components is used to represent the longest queue \bar{L}_i associated with a stage i . For the present tests, the maximum detectable queue length was taken to be 31; longer queues may occur but the detectors will report them as being of length 31. The motivation for this is that in practice, the distance between a sensor and the junction is limited: in a single lane, a queue of 31 stationary vehicles would extend over about 200m. Four bits are used to represent actions so that there are four possible durations for each of the two stages. After some experimentation, these were set at 5, 18, 31 and 45 seconds; further work reported later in the present paper explored the use of other combinations. The reward R to an individual junction controller at the end of each cycle was then determined as follows:

$$\begin{aligned} \bar{L} &= \min(16, \max_i \bar{L}_i) \\ R &= 4(16 - \bar{L})^2 \end{aligned} \quad (1)$$

Thus rewards are scaled to lie between 0 and 1024, in keeping with much of the work using ZCS (eg Wilson, 1994). A large number of test cases have been investigated but two representative ones are considered here, each characterised by the mean arrival rate of traffic on each approach to the junction. Both of these cases have balanced flow on all approaches: 720 vehicles/hour, which can be considered to be moderate, and 1440 vehicles/hour, which is quite high. Each junction is connected to its neighbours by roads of length 250m and to the perimeter of the network by roads of length 125m; vehicles travel at a maximum speed of 54km/h. Therefore the undelayed network traversal time for a vehicle is about 33 seconds.

We now present the results from using independent identical instances of ZCS to control each of the four junctions in these two test cases.

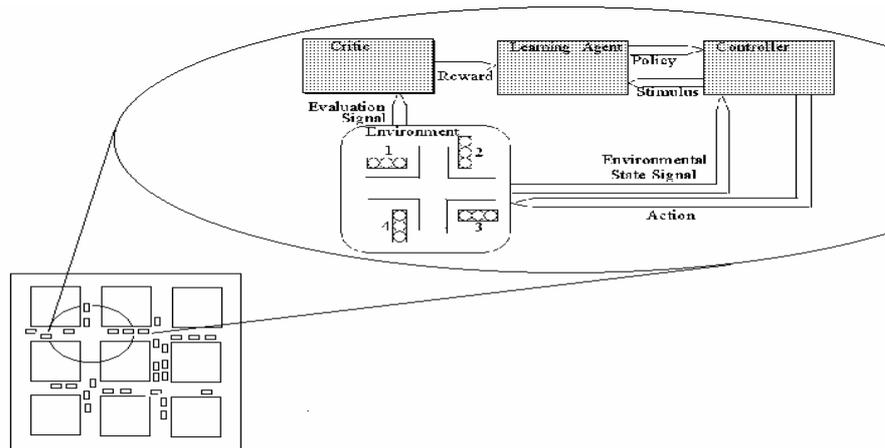


Figure 1: 4-junction network simulation (inset is blow-up of local junction controller signal flow)

5. ZCS Results

Bull and Hurst (2002) showed how ZCS performance can change significantly with different parameters to control size of the rule-base, rule discovery component and reinforcement component. We now report the effects of varying these main parameters of ZCS on the junction control task.

5.1 Rule-base Size

As presented in the original literature (Wilson, 1994) the standard rule-base size for ZCS is 400 rules. The present environmental interface provides over 4000 possible state/action pairs, so we investigated whether an increased population size was beneficial to performance. The ZCS controller performance, measured as mean network traversal time in seconds, was evaluated with rule-bases of 400, 800 and 1600 rules. All other parameters were kept as Section 3. Results were averaged over 10 tests of 10,000 trials during which the last 2000 trials are evaluation trials with rule discovery mechanisms disabled after Bull and Hurst (2002). The figures in the tables, which represent resultant performance at the end of the simulation, are the average delays experienced by the cars of the last 2000 evaluations. It is important to note that this is a global metric that is often used as a measure of performance in transport studies but that in practice is not available to the junction controllers.

Table 1: Mean travel time under ZCS with rule-bases of varying sizes (seconds).

	Size of rule base N (rules)		
Mean arrival rate (cars / hour)	400	800	1600
720	54.0	55.1	53.8
1440	55.8	54.9	56.2

These results suggest that there is no benefit in this application in increasing the rule-base size of the ZCS. In the remainder of the present tests, the rule-base size was therefore maintained at 400 rules.

5.2 Rule Discovery Component

The rate of activation of the GA determines the rate at which new genetic material is introduced to the system. In a multi-agent environment such as the present one, it is possible that adjustments to the GA rate or mutation rate may yield some benefits (Bull, 1998). Hence the ZCS was tested with a higher GA firing rate p and a higher mutation rate μ (Table 2).

Table 2: Performance of ZCS with increased GA/mutation rates (seconds).

Mean arrival rate (cars / hour)	Parameter values		
	Standard ($p = 0.25$, $\mu = 0.01$)	$p = 0.5$	$\mu = 0.04$
720	54.0	54.0	52.1
1440	55.8	58.4	54.5

Although increasing the GA rate p from 0.25 to 0.50 does not appear to yield, increasing the mutation rate μ from 0.01 to 0.04 made slight performance improvements in both test cases. Further increases did not yield any further benefits (not shown) so the ZCS controller was set to operate with the standard GA rate but an increased mutation probability of $\mu = 0.04$ per gene hereafter.

5.3 Reinforcement Component

Bull and Hurst (2002) showed the importance of the learning rate β in ZCS. An increased learning rate allows the system to adjust rule fitness more rapidly which is important in ZCS because fitness varies due to the occupancy of the choice set [A] as well as the external reward received. Bull (1998) also showed that an increased learning rate can allow individuals in multi-agent systems to respond more rapidly to the dynamics of their environment.

Table 3: Performance of ZCS with increased learning rates (seconds).

Mean arrival rate (cars / hour)	Learning rate β		
	0.2	0.5	0.8
720	52.1	54.0	52.1
1440	54.5	56.2	53.5

The results in Table 3 suggest that the performance of ZCS benefits little if at all from an increased learning rate in these single-step tasks. This result contrasts with experience in the delayed reward tasks examined in (Bull, 1998). However, this may be explained by the more direct effect of external reward on rule fitness in the present application.

In summary, ZCS appears able to learn coherent strategies when operating as part of a decentralised control architecture for road traffic junctions. Cao et al (2001) have

also examined the use of ZCS under a similar application, reporting improved performance over a fuzzy logic approach. Use of the more sophisticated XCS is now presented.

6. XCS

A significant proportion of LCS research is now using Wilson's accuracy-based system – XCS (Wilson, 1995; Butz and Wilson, 2001). A number of studies have found XCS capable of optimal performance in complex domains and hence the system was investigated here. The default parameters presented for XCS, which were used in the present experiments except where stated otherwise, are: $N = 800$, $E_0=10$, $\beta = 0.2$, $\gamma = 0.71$, $\chi = 0.8$, $\mu = 0.01$, $\phi = 0.5$, $p_0=10$, $\alpha=0.1$, $P_{\#}=0.5$, $\delta = 0.1$, $E_I=0$, $F_I=0.01$, $p_I=10$.

6.1 Rule-base Size

The original rule-base size N for XCS was 800 rules (Wilson, 1995) but subsequent research regularly expands this (eg, Lanzi and Wilson, 2001). To observe the effects of increased rule-base size XCS was tested with sizes of 800 and 1600 rules (Table 4) with all other parameter values as specified above.

Table 4: Performance of XCS with different rule-base sizes (seconds).

	Size of rule base N (rules)	
Mean arrival rate (cars / hour)	800	1600
720	46.8	44.5
1440	48.2	46.0

Clearly XCS benefits from an increase in rule-base size and it will be set at 1600 hereafter (larger rule-sets did not appear to help - not shown). It is also important to note that for each of the tasks that we investigated, XCS performs better than ZCS.

6.2 Rule Discovery Component

Since the results with ZCS in Section 5.2 showed a benefit from an increase in the mutation rate, the corresponding parameter μ was examined with XCS. Table 5 shows how an increase in the value of μ appears to help XCS to some degree. Results from further increasing the mutation rate gave no further improvement in performance (not shown), so the XCS controller was set to operate with the increased mutation probability of $\mu = 0.04$ per gene hereafter.

Table 5: Performance of XCS with different mutation rates (seconds).

	Mutation rate μ	
Mean arrival rate (cars / hour)	0.01	0.04
720	44.5	44.2
1440	46.0	45.5

6.3 Reinforcement Component

Table 6 shows the effects of increasing the learning rate β of the XCS. It can be seen that no change in performance is obtained from altering the learning rate from the typical value of 0.2.

Table 6: Performance of XCS with increased learning rates (seconds).

	Learning rate β		
Mean arrival rate (cars / hour)	0.2	0.5	0.8
720	44.2	44.3	44.5
1440	45.5	45.3	45.0

In summary, XCS has been found to perform better than ZCS in all cases tested here. However, by its nature, the task is potentially ambiguous because the payoff received for taking an action in given circumstances depends upon both the behaviour of the traffic and the previous actions taken. Neither ZCS nor XCS contain explicit mechanisms by which to consider the context of the current stimulus. In view of this, we examined the utility of incorporating two mechanisms aimed at enabling XCS to maintain internal memory - rule-linkage and the memory register - in the distributed road traffic junction control task to further improve performance.

7. Internal Memory

7.1 Memory Register

Some environments are not completely observable by the learning entity. That is, with respect to the learner's sensory input, environments can be only partially observable

leading to the same sensory input for different environmental states; this is termed *non-Markov*. Without internal state LCS cannot perform optimally in such cases. Wilson (1994) proposed a simplified version of Holland’s message list (Holland et al, 1986): an internal memory register whose state is considered and altered by the rules of the LCS. Cliff and Ross (1994) implemented Wilson’s mechanism in ZCS, resulting in a system denoted as ZCSM, which finds good but not optimal performance in non-Markov mazes. Tomlinson and Bull (1998) examined the performance of ZCSM in a range of mazes of increasing complexity. They found that for tasks containing numerous ambiguous states benefit could be obtained by adopting a memory register that was bigger than they expected, suggesting that redundancy is required for it to work effectively. Lanzi and Wilson (2001) presented optimal performance with XCS incorporating the memory register in a number of mazes. They also found that for more complex mazes seemingly superfluous internal memory states were beneficial.

The rules of XCS are extended to consider not only the external condition (c bits) but also the state of the internal memory register (m bits). Similarly, rule actions are extended to provide the external action (a bits) and an internal action (m bits). Wildcards are allowed for internal actions that leave the state of the corresponding bit in the memory register unchanged, otherwise the register bits are updated to the value defined in the internal action. Hence rules are of the length $c + 2m + a$ bits and an action set consists of rules proposing the same external and internal actions. Furthermore, Lanzi and Wilson restrict internal updates to occur only when a change in external state results from the chosen external action. The internal action on each cycle is always chosen deterministically from within a given [M] based on the expected payoff of rules, regardless of the external action selection scheme.

Previous work has used the memory register in environments with a restart at which point the memory register is set to all zeros. However, in the road traffic control task, the contents of the memory register are never reset explicitly.

Table 7: Performance of XCS with the memory register (seconds).

Mean arrival rate (cars / hour)	Version		
	XCS	XCSMH2	XCSMH4
720	44.2	58.9	59.0
1440	45.5	59.5	57.5

Table 7 shows the performance of XCS with the memory register applied to the same test cases as before with the parameters determined above. Following (Lanzi and Wilson, 2001), the system with a memory register of two bits is labelled XCSMH2 and that with four bits XCSMH4. It can be seen that the additional mechanism has been detrimental to the performance of the system, reducing the performance of the XCS

down to the levels obtained with the simpler ZCS. That is, the ability to create and consider the context of external stimuli using the memory register does not appear to be beneficial in the present application. In the networks that were investigated here, at least half of the junction approaches have random traffic arrival patterns, and our results suggest that the memory mechanism is ineffective in these circumstances.

7.2 Rule-linkage: Corporations

Wilson and Goldberg (1989) suggested that, by allowing rules to link to others to form new evolutionary units termed rule corporations, inductive chains can be formed without the use of a common memory space. An initial implementation of this idea in ZCS (Tomlinson and Bull, 1998) found this to be the case, with significant improvements in performance in (multiple) ambiguous maze tasks. The system employs linkage between rules in the rule-base to form temporal rule-chains, implemented as doubly linked lists of rules. Each rule in the LCS is equipped with two, initially inactive, link parameters. When activated, either or both of these links may reference another rule in the rule-base symmetrically. The result of such associations is a rule-base of arbitrarily long rule chains, whose members are treated as collective units by the learning mechanisms. It has been noted that this has parallels with a type of symbiosis found in nature (Tomlinson and Bull, 2001).

The aim of corporate classifier systems (CCS) is to encourage corporations to encapsulate temporal chains of inference. Hence, rule-linkage acts across subsequent time steps, choosing unlinked candidates from those that were active on the previous time-step to join with (unlinked) active rules on the current time step. It has also been found beneficial to allow a set of linked rules to execute until completion (after Smith, 1992), provided that each rule's conditions remain satisfied: this is termed *corporate persistence*. Thus by evaluating completely a strategy that consists of a linked set of rules, the full worth of the temporal logic they represent can be estimated more effectively.

It can be noted that the structure of CCS is similar to that of the dynamic programming approach introduced by Robertson and Bretherton (1974), which has recently been re-examined (Heydecker and Boardman, 1999). Indeed, dynamic programming with a rolling horizon (Gartner, 1983) can be seen as similar to CCS because each corporation represents a series of actions that are planned for the future, with each rule indicating the worth of that temporal series of actions.

A number of developments have recently been made to the original CCS concept, including the use of XCS as the basis (Tomlinson and Bull, 2001; 2002): link inhibition, which reduces the chances of useful corporations growing too large; different linking strategies, such as using fitness-biased selection between niches; parallel persistence, whereby all active corporations are maintained over successive time-steps; and direct payoff updates, under which the bucket- brigade is executed down a corporation on each system cycle to increase the rate of flow of information down the rule chain. We have applied this system to the present road traffic control task.

Table 8: Performance of XCS with rule-linkage (seconds).

Mean arrival rate (cars / hour)	Version	
	XCS	CXCS
720	44.2	53.4
1440	45.5	54.2

Table 8 shows the performance of the corporate XCS system on the same test cases as above. It can be seen that, just as with the memory register, the extra mechanisms do not aid performance, rather performance drops, although not as much (Table 7). These results have been repeated under a number of test cases and parameter settings.

These investigations make it clear that the standard XCS system, with an appropriate rule-base size, gives the best performance in the distributed traffic control tasks examined. The XCS system was then implemented within a more detailed micro-simulator and comparisons in performance made with a number of well-known junction control approaches, as described below.

8. Evaluation using a Microscopic Traffic Simulation

In order to investigate a range of possibilities for the use of the LCS approach to the signal control of road traffic, a detailed microscopic simulation of road traffic under signal control was required. This development and evaluation work was undertaken using the SIGSIM 2.0 (Sha'Aban, 2003a; b) microscopic traffic simulator, which is based upon the Gipps (1981, 1986) car following and lane-changing models, and includes detailed representation of traffic signals and vehicle detectors. The LCS was integrated into this simulation by extracting stimulus signal from the SIGSIM vehicle detectors, scheduling calls to the LCS optimiser within the simulation framework, and implementing the LCS decisions as signal timings. The resulting combined software provided the main technology for the investigation and development of an LCS approach to signal control of road traffic and for its evaluation. For this investigation, we adopted the XCS form of the LCS system.

Within this broad framework, various choices were available for several key components. These included:

- the choice of stimulus, the choice of reward for the LCS, including both the quantity and its transformation,
- the choice of policy objective for optimisation, and
- the choice of control action for the LCS rules.

We discuss these in turn below, and present the results of experimentation with them.

8.1 Choice of Stimulus for the LCS

Several possibilities exist for the form in which traffic information is extracted from the simulation to provide a stimulus to the LCS for input to its rules: this represents the output of traffic detectors. Many kinds of traffic detector are available in practice, each with its own characteristics and properties. For each of these, the data that they provide can be processed in various ways before being presented to the LCS as stimulus. The kinds of detectors that are currently in use include various kinds of point detectors, microwave based detectors, and video-based detectors.

Point traffic detectors, such as inductive loops, are widely used to provide traffic information for signal control. They are usually placed upstream of the stopline, where they provide information about the imminent arrival of vehicles at the junction. Detectors of this kind can also be placed at or immediately downstream of the stop-line, where they provide information about the flow of traffic through the signals. By integrating over time the difference between detector outputs at these locations, one can estimate the number of vehicles in the vicinity of the junction, though this method is prone to accumulated errors and so requires that the estimates be reinitialised from time to time.

Above-ground detectors such as Doppler effect microwave and video-based systems have several advantages over inductive loops. Some of these arise from reduced installation and maintenance costs, whilst others arise from their operating characteristics. A typical microwave detector will respond to the presence of any traffic in a region of the approach to the junction, often (but not always) starting at the stop-line and extending upstream. The output is binary, indicating one or other of presence of at least one approaching vehicle or absence of any, and is devoid of spatial information within the detection region.

Video image analysis detection systems offer the prospect of more detailed information capture. A suitably mounted video camera in favourable circumstances can acquire a view of several hundred metres of approach to a junction. The level of detail in the data extracted from a video image depends on the sophistication of the image processing that is undertaken, and can range from identification of presence of traffic to reporting of positions and speeds of approaching vehicles.

The possibilities for data extraction from these various kinds of detector system include the times at which vehicles had arrived recently at vehicle detectors, the presence (or absence) of vehicles in a region between a detector and the stop-line, and the number of vehicles in this region. The last of these – the number of vehicles between a detector and the stop-line - provides a practical means for this transfer of information. We consider that this could be achieved in principle by the kinds of detection systems that are either currently in use or will be in the foreseeable future.

8.2 Choice of Reward for the LCS (including transformations)

During the learning phase, the LCS receives a reward R calculated according to the traffic performance resulting from implementation of a set of rules: high values of this reward lead the LCS to favour the rules that give rise to it. Our initial tests used the maximum of the queue lengths \bar{L} on streams at the end of their respective red indications as the basis for this reward. Clearly, this is a positive quantity that would ideally be minimised, so that some decreasing mapping from positive real numbers to positive real numbers is required. We investigated use of each of A linear (2), two piecewise quadratic (3), (4), hyperbolic (5), and logistic (6) transformations and their associated parameters.

$$R = \max(0, 1 - \alpha\bar{L}) \quad (2)$$

$$R = \max(0, \alpha - \bar{L}^2) \quad (3)$$

$$R = (\alpha - \min(\alpha, \bar{L}))^2 \quad (4)$$

$$R = \min\left(1, \left| \frac{\alpha D_m}{(\alpha - 1)D_m + \bar{L}} \right| \right) \quad (5)$$

$$R = \min\left(1, \frac{\beta}{(1 + \exp(\alpha(\bar{L} - D_m)))}\right) \quad (6)$$

This investigation led us to adopt the second quadratic form (4) with the parameter value $\alpha = 16$. We also investigated use of the same quantity (initially mean rate of delay, corresponding to time-averaged queue length) as was our objective for the basis of this reward calculation. However, we found that use of the maximum value of the instantaneous queue lengths at the end of red led to better performance of the LCS.

8.3 Choice of Evaluation Objective (cf policy objective of traffic management)

We investigated use of the LCS approach to minimise each of

- the mean rate of delay and
- the mean rate of fuel usage

associated with a junction. We calculated the mean rate of delay D for the junction according to

$$D = \sum_i q_i d_i$$

where q_i is the mean arrival rate in stream i ,
 d_i is the mean delay in stream i ,

and the summation is over all streams i that are controlled by the signals at that junction. We estimated the mean rate of fuel usage F according to the method proposed by Robertson, Lucas and Baker (1980):

$$F = \sum_i q_i (\alpha_d d_i + \alpha_s s_i)$$

where s_i is the mean rate at which vehicles in stream i stop,
 α_d is the mean amount of fuel used during idling for unit time (about 1.15 litres/pcu-h), and
 α_s is the mean additional amount of fuel used during acceleration from rest to cruising speed (about 0.0063 litres/pcu).

We found that we could not improve the performance in respect of the second objective (fuel usage) beyond that achieved when the mean rate of delay was minimised: this is due in part to the strong interrelationship between these quantities at signal-controlled road junctions, and in part to the effectiveness of using maximum queue length rather than the evaluation objective itself as reward during the learning phase.

8.4 Choice of Control Actions for the Rules (stage durations, incremental greens)

We investigated the way in which the output of the LCS rules is translated into signal control actions. The first approach that we considered in detail for use with this microscopic simulation was to select between predetermined possibilities for the durations of each stage. The calculations for this were performed at the end of the minimum green period for each stage (typically 7s duration), so that the values implemented correspond to extensions beyond this minimum green, which was therefore respected automatically in this process; this implementation was symmetric between stages. An important issue in this case is selection of an appropriate values in the choice set for the green time extensions; we found that by tailoring this choice, performance could be improved. The results of some tests using a simple cross-roads junction with flows of 400 vehicles per hour on each of the four approaches, and also with 600 per hour on each approach are shown in Table 9. This shows that there was little variation in performance between the different choice sets of green time extensions of (0, 2, 4, 8) seconds, (0, 5, 10, 15) seconds, and (0, 10, 20, 30) seconds. For this reason, we investigated both the set (0, 2, 4, 8) s of short and the set (0, 5, 10, 15) s of medium extensions for further testing. Clearly, further investigation of the choice set could lead to an improvement beyond these levels.

The second approach that we developed corresponds to an incremental extension strategy, in which a binary decision of whether or not to continue running the current stage was considered frequently (typically at 0.5 s intervals) during green, starting at the end of the minimum green period. This has the advantage of not requiring a predetermined choice set of stage durations, but the disadvantage of requiring more intensive computation. We found that this extension approach led to rules that gave slightly better (typically in the range 1 – 2 ½ per cent) mean rates of delay. Results of example runs for the LCS with the extension strategy are given in Table 9.

Table 9: Performance of LCS strategies with short-term flow variability (vehicles)

Mean arrival rate on each approach (cars / hour)	Extension choice set (seconds)			Extension strategy
	(0, 2, 4, 8)	(0, 5, 10, 15)	(0, 10, 20, 30)	
400	49.73	49.92	49.33	48.92
600	81.57	81.67	85.10	82.21

9. Results from Detailed Microsimulation

We evaluated the performance of the rule bases that resulted from a wide range of LCS formulations according to the nominated evaluation objective. In these evaluations, our comparators were fixed-time signal control, and System D vehicle-actuated control (DoT, 1991); we optimised the performance of these comparators by direct search over green and (respectively) maximum durations for the signal stages. In each case, we calculated the mean rate of delay at the junction averaged over several (generally 10) mutually independent runs using distinct random number seeds. Each run had a 900 second run-in period followed by a 3600 second evaluation period.

We considered and simulated three different kinds of variability of traffic flows to investigate the capability of the rules generated by the LCS to accommodate them. These were:

- **Short-term variability** due to random fluctuations around a constant mean arrival rate ,
- **Systematic variations** over a timescale of many minutes, as occurs within a morning peak period, and
- **Long-term development** in mean arrival rates of the kind that lead to ageing of signal plans

The results of each of these experiments are given below.

9.1 Short-term Variability

In order to investigate the relative performance of the various control strategies in the presence of short-term random variability in traffic flows, we generated vehicular arrivals at the entries to the simulated road sections according to the shifted exponential distribution with appropriate mean flow. This kind of variability is typical

of road traffic even when the long-term mean flow is constant, and gives rise to possibilities for varying signal indications in a manner that is beneficial to traffic – typically by truncating green indications when they are not being used heavily in favour of advancing the start of other green indications. This kind of variability was included in all the other tests that are reported here.

The results of this experiment are shown in Table 10. We considered a simple cross-roads junction with flows of 400 vehicles per hour on each of the four approaches, and also with 600 per hour on each. We identified by direct search the timings for fixed-time and system D control that provided the best performance. The LCS method was used with a rule-base learnt using the level of flows to be controlled in each case.

Table 10: Performance of LCS strategies with short-term flow variability (vehicles)

Mean arrival rate on each approach (cars / hour)	Control strategy			
	Fixed time	System D	LCS	LCS Extension strategy
400	48.79	48.67	49.92	48.92
600	80.92	80.42	81.67	82.21

We found that the LCS could produce performance that is close to, but slightly worse than, optimised fixed time and optimised System D vehicle-actuated signal control at each of a range of traffic flows: the matched means of the 10 mutually independent runs showed greater mean rate of delay than that achieved by VA control of $\frac{1}{4}$ - $\frac{1}{2}$ per cent for FT and $\frac{1}{2}$ - 1 per cent for LCS control.

9.2 Systematic Variations

We investigated performance of each of the control methods during a synthesised morning peak profile during which the flows increase from an initial value of 200 vehicles per hour up to 600 and then reduced to 400 per hour on each approach. The synthesised profile of mean arrival rate is shown in Figure 2.

In order to make fair comparison, we searched for the best timings for the FT and vehicle-actuated control, and for the best constant learning flows for the LCS. A graph showing variations in the mean rate of delay against green time settings for fixed time and vehicle-actuated control is shown in Figure 3. From this, the settings that achieve minimum mean rate of delay can be seen to be 9s for fixed time, and 11s (maximum green) for vehicle actuated. The LCS rules were learnt using constant mean arrival rates, and we investigated the sensitivity of the test performance to the value of this constant learning flow.

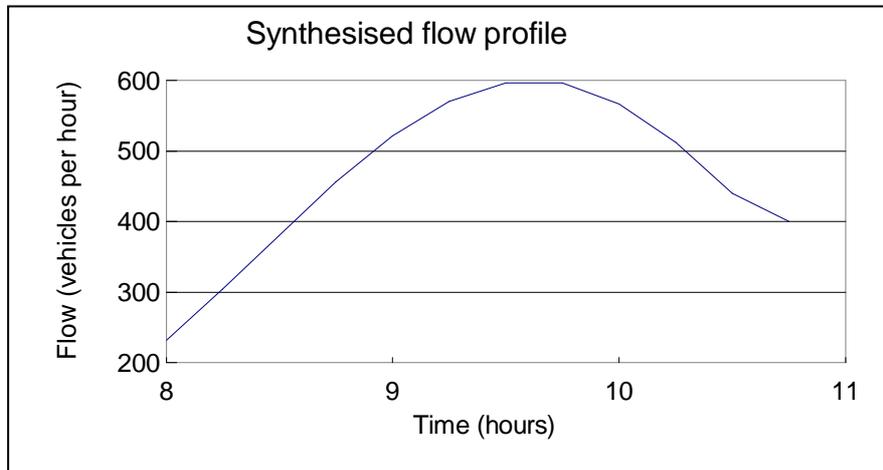


Figure 2: Synthesised peak period flow profile.

The results of this test of the LCS are shown in Table 11. These show that the optimised FT (60.05 vehicles) performed about 0.6 per cent worse than did VA control (59.68 vehicles), and that the extension strategy LCS (60.21 vehicles) performed about 0.9 per cent worse than did VA. However, the LCS with short extensions (0, 2, 4, 8) seconds when the rule base was learnt with a mean arrival rate of 540 vehicles per hour produced mean rate of delay over 10 runs of 57.63 vehicles, which is about 3 per cent better than VA. Investigation of the individual runs showed that the mean rate of delay varied between them, with standard deviation between runs of 2.89 vehicles so that the standard error of estimation of the mean was a little under 1 vehicle. We therefore undertook a series of 100 runs with this rule set, and established a mean rate of delay of 58.84 vehicles with standard deviation of 1.41 vehicles: this is about 1.4 per cent better than VA control, though substantial variability in performance between runs remains. When the longer choice set of green time extensions of (0, 5, 10, 15) s was used, performance was worse at 59.91 vehicles (also achieved with learning flows of 540 vehicles per hour), which is about slightly worse than VA control.

We found that for the extension strategy LCS control method, performance was robust with respect to the value of flows for which the control was optimised, ranging between 60.21 and 60.93 vehicles as the flows used during the learning phase were varied from 300 to 650 vehicles per hour (using increments of 50 vehicles per hour). However, the LCS strategy with choice of extensions (0, 2, 4, 8) seconds produced results that varied substantially with learning flows in the same range, ranging from 57.63 to 61.16 vehicles.

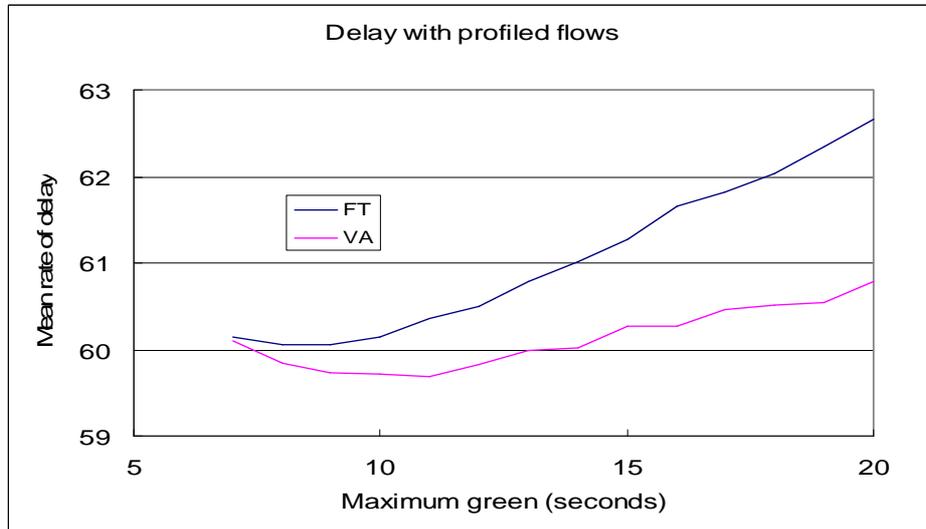


Figure 3: Variation in mean rate of delay for profiled flows.

Table 11: Mean (and standard deviation) of performance of signal control strategies with profiled arrival rates (vehicles).

Control strategy	Control strategy				
	Fixed time (9 second stage duration)	System D (11 second maximum stage duration)	LCS (0, 2, 4, 8)	LCS (0, 5, 10, 15)	LCS Extension strategy
Mean	60.05	59.68	57.63	59.91	60.21
Standard deviation	(0.770)	(0.721)	(2.892)	(3.592)	(0.951)

9.3 Long-term Development in Mean Arrival Rates

We investigated the performance of each of the control strategies as flows vary substantially from those for which the strategies were optimised. This test represents the effect of ageing of signal plans on performance, which arises because the flows that are controlled change over time from those for which the timings were calculated. This phenomenon is known to lead to degradation in performance of fixed-time plans

of about 3 per cent each year (Bell and Bretherton, 1986). We found differences in performance between the strategies that were smaller than the differences between the various flow levels: there was almost no difference between fixed-time and VA control when both used 11s stage durations. At high flows, the LCS with the medium choice set of extensions (0, 5, 10, 15) s performed best, largely because the optimisation of the other strategies for lower flows led them to have more limited traffic capacity. By contrast, the LCS extension strategy performed poorly at high flows. The results of this test are shown graphically in Figure 4.

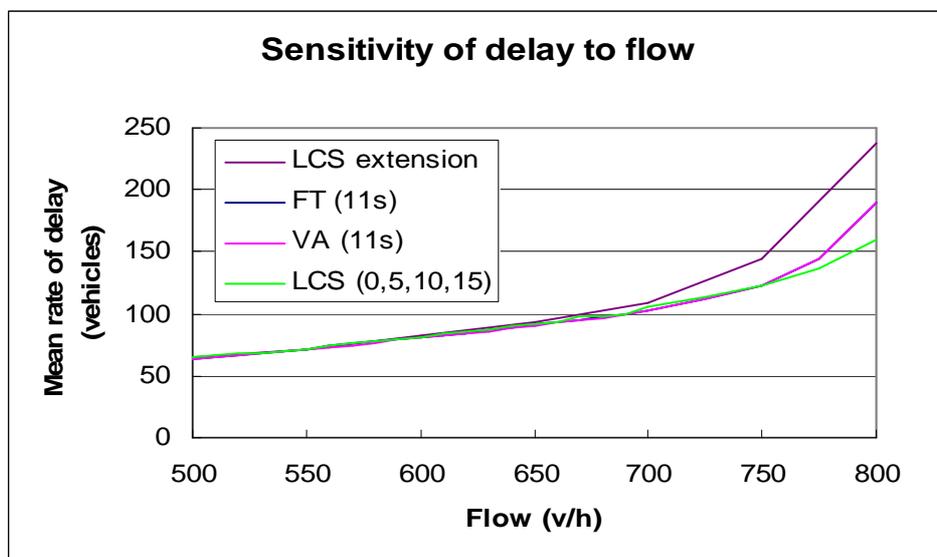


Figure 4: Variation in performance with flows.

10. Conclusions

In this chapter, the practical value of Learning Classifier Systems for road traffic junction control has been examined. This has shown that XCS version of LCS outperforms the ZCS in all of the test cases investigated and that no further improvement in performance can be obtained by the inclusion of either rule-linkage or the memory register. The advantages of the LCS approach include its flexibility in respect of the form of data and the objectives of optimisation. When tested with a detailed simulation of signal controlled road traffic, some implementations of the LCS methodology outperformed standard control methods in certain cases. This indicates that the LCS approach has potential for application to road traffic control.

It is possible that both CXCS and XCSMH require further adaptation before they are suitable for this application. It is also possible that none of these systems is able to pick up temporal patterns in the traffic flow that can be exploited so the memory

mechanisms have nothing to work with. In the inner part of an urban road network, most major approaches to controlled junctions will receive flow from neighbouring junctions and so traffic will be platooned. This effect will result in junctions receiving traffic according to the policies and timings at the neighbouring junctions. This could well lead to patterns of traffic flow that could be usefully interpreted by systems such as CXCS and XCSMH. However, in the networks that were investigated in the present paper, at least half of the approaches have random traffic arrival patterns. The present results suggest that these variants of the LCS controller cannot operate effectively with this degree of randomness, resulting in poor performance.

More detailed evaluation of the XCS controller was undertaken at individual road junctions using the SIGSIM traffic simulation, which can model traffic behaviour and signal control in detail. This enabled us to consider various aspects of the implementation of LCS for signal control of road traffic. In particular, we considered the way in which traffic data are presented to the LCS as stimulus for the rules, the form of the LCS reward for performance of its rules, a range of objectives for the control policy, and the choice of control action for the LCS rules to vary the signal control. Each of these was found to affect the performance of the LCS system to some extent.

We undertook a series of tests to investigate the performance of the LCS in the presence of different kinds of variability in traffic flows. These ranged from random variability around a constant mean arrival rate, through profiled flows representing a peak period, to long-term development of mean arrival rates.

The results of this part of the investigation showed that some implementations of the LCS could outperform standard control methods in certain of these tests. Our investigations considered several elements of the way in which the LCS is configured for the signal control task. We found that the performance of the LCS approach varied according to

- the way in which the LCS rules are interpreted as control actions,
- the choice of measure of traffic conditions that is used as stimulus for the LCS rules,
- the corresponding choice for feedback as the basis of the reward, and
- the mathematical function that is adopted to transform traffic measures to rewards.

Because there is substantial choice in each of these respects, further investigation could well lead to further improved performance. Our results have established the potential for LCS using current detector technology to deliver improvements beyond what can be achieved using current control methods. The future availability of further detector technologies and the range of possible developments to the implementation of the LCS to the signal control task offer scope for further improvements.

Acknowledgements

This work was supported under EPSRC Grant no. GR/R06748. The authors are grateful to Professor Richard Allsop at University College London for his contributions throughout this project, and to Neil Hoose of Bittern Partners and Peter Yendall of Oscar Faber for their comments at various stages during this work.

References

- Arthur WB (1990) A Learning Algorithm that Replicates Human Learning. *Technical Report 90-026*, Santa Fe Institute.
- Bell, MC and Bretherton, RD (1986) Ageing of fixed-time traffic signal plans. *Proceedings of the Second IEE Conference on Road Traffic Control*. London: IEE.
- Bull L (1998) On ZCS in Multi-Agent Environments. In AE Eiben, T Baeck, M Schoenauer and H-P Schwefel (eds) *Parallel Problem Solving from Nature - PPSN V*, Springer Verlag, 471-80.
- Bull L (1999) On using ZCS in a Simulated Continuous Double-Auction Market. In W Banzhaf, J Daida, AE Eiben, MH Garzon, V Honavar, M Jakiela and RE Smith (eds) *GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference*. Morgan Kaufmann, 83-90
- Bull L and Hurst J (2002) ZCS Redux. *Evolutionary Computation* **10**(2), 185-205.
- Bull L, Fogarty TC and Snaith M (1995) Evolution in Multi-Agent Systems: Evolving Communicating Classifier Systems for Gait in a Quadrupedal Robot. In LJ Eshelman (ed) *Proceedings of the Sixth International Conference on Genetic Algorithms*, Morgan Kaufmann, 382-88.
- Bull, L and Tomlinson, A (2003) Symbiogenesis in Machine Learning. In R Paton (ed) *Computation in Cells and Tissues: Perspectives and Tools of Thought*. Springer.
- Butz, M and Wilson, SW (2000) An Algorithmic Description of XCS. In P-L Lanzi, W Stolzmann and SW Wilson (eds) *Proceedings of the Third International Workshop on Learning Classifier Systems*. Springer, 253-72.
- Cao YJ, Ireson N, Bull L and Miles R (2001) An Evolutionary Intelligent Agents Approach to Traffic Signal Control. *International Journal of Knowledge-based Intelligent Engineering Systems* **5**(4), 279-89.
- Carse B, Fogarty TC and Munro A (1995) Adaptive Distributed Routing using Evolutionary Fuzzy Control. In LJ Eshelman (ed) *Proceedings of the Sixth International Conference on Genetic Algorithms*, Morgan Kaufmann, 389-97.
- Chiu S (1992) Adaptive traffic signal control using fuzzy logic. *Proceedings of Intelligent Vehicles 1992*, 98-107.
- Cliff D and Ross S (1994) Adding Temporary Memory to ZCS. *Adaptive Behaviour*, **3**, 101-50
- DoT (1991) TR 0141: Microprocessor-based traffic signal controller for isolated, linked and urban traffic control installations. London: Department for Transport.

- Dorigo M and Schnepf U (1992) Genetics-based Machine Learning and Behaviour-based Robotics: A New Synthesis. *IEEE Trans. on Sys. Man and Cybernetics*, **22**(6), 141-54.
- Dworman G (1994) Games Computers Play: Simulating Characteristic Function Game Playing Agents with Classifier Systems. In *Proceedings of the 1994 IEEE Conference on Evolutionary Computation*, IEEE, 696-701.
- Gartner NH (1983) OPAC: a demand-responsive strategy for traffic signal control. *Transportation Research Record*, **906**, 75-84.
- Gartner NH, Kaltenbach MH and Miyamoto MM (1983) Demand-responsive decentralized urban traffic control, part II: network extensions. *Report DOT/OST/P-34/85/009*.
- Gartner NH, Tarnoff PJ and Andrews CM (1991) Evaluation of optimised policies for adaptive control strategy. *Transportation Research Record*, **1324**, 105-14.
- Gipps, PG (1981) A behavioural car-following model for computer simulation. *Transportation Research*, **24B**(2), 105-11.
- Gipps, PG (1986) A model for the structure of lane-changing decisions. *Transportation Research*, **28B**(4), 269-87.
- Heydecker BG and Boardman RM (1999) Optimisation of Timings for Traffic Signals by Dynamic Programming. *Proceedings of the 31st UTSG Annual Conference*.
- Holland JH (1975) *Adaptation in Natural and Artificial Systems*, University of Michigan Press.
- Holland JH, Holyoak KJ, Nisbett RE and Thagard PR (1986) *Induction: Processes of Inference, Learning and Discovery*, MIT Press.
- Holroyd J and Hillier JA (1969) Area traffic control in Glasgow: a summary of results from four control schemes, *Traffic Engineering and Control*, **11**(5), 220-3.
- Holland JH and Miller J (1991) Artificially Adaptive Agents in Economic Theory. *American Economic Review*, **81**(2), 365-70.
- Holroyd J and Hillier JA (1971) The Glasgow experiment: PLIDENT and after. Transport and Road Research Laboratory *Report LR 384*. Crowthorne: TRL.
- Hunt PB, Robertson DI, Bretherton RD and Winton RI (1981) SCOOT: a traffic-responsive method of coordinating signals. Transport and Road Research Laboratory *Report LR 1014*. Crowthorne: TRL.
- Koza, JR (1991) *Genetic Programming*. MIT Press.
- Lowrie PR (1982) The Sydney co-ordinated adaptive traffic system - principles, methodology, algorithms. *Proceedings of the IEE International Conference on Road Traffic Signalling*, **207**, London: IEE, 67-70.
- Marengo L and Tordjman H (1996), "Speculation, Heterogeneity and Learning: A Model of Exchange Rate Dynamics" *KYKLOS* **49**(3), 407-38.
- Marimon R, McGrattan E and Sargent T (1990) Money as a Medium of Exchange in an Economy with Artificially Intelligent Agents. *Economic Dynamics and Control*, **14**, 329-73.
- Mieden Th, Sayers TM, Busch F and Bell MGH (1996) Verkehrsabhängige Lichtsignalsteuerung mit fuzzy-logik - ein modulares, praxisorientiertes verfahren. *Strassen-verkehrstechnik*, **10/96**, 485-9.

- Mikami S and Kakazu K (1993) Self-organised control of traffic signals through genetic reinforcement learning. *Proceedings IEEE Intelligent Vehicles Symposium*, IEEE, 113-18.
- Montana DJ and Czerwinski S (1996) Evolving control laws for a network of traffic signals. In J Koza, DE Goldberg, DB Fogel and RL Riolo (eds) *Proceedings of the 1st Annual Conference on Genetic Programming*, MIT Press, 333- 38.
- Potter M, De Jong K and Grefenstette J (1995) A Coevolutionary Approach to Learning Sequential Decision Rules. In LJ Eshelman (ed) *Proceedings of the Sixth International Conference on Genetic Algorithms*, Morgan Kaufmann, 366-72.
- Robertson DI (1969) TRANSYT: a traffic network study tool. Transport and Road Research Laboratory *Report LR 253*. Crowthorne: TRL.
- Robertson D and Bretherton RD (1974) Optimum Control of an Intersection for any Known Sequence of Vehicular Arrivals. *Proceedings of the Second IFAC-IFIP-IFORS Symposium on Traffic Control and Transportation Systems*.
- Robertson, DI, Lucas, CF and Baker, RT (1980) Coordinating traffic signals to reduce fuel consumption. Transport and Road Research Laboratory *Report: LR 934*. Crowthorne: TRL.
- Sayers T, Anderson J and Clement S (1999) The multi-objective optimisation of a traffic control system. In: A Ceder (ed) *Transportation and Traffic Theory*, ISBN 965-222-993-8, 153-76.
- Seredynski F, Cichosz P and Klebus G (1995), "Learning Classifier Systems in Multi-Agent Environments." In *Proceedings of the First IEE/IEEE Conference on Genetic Algorithms in Engineering Systems: Innovations and Applications*, IEE, 287-92.
- Sha'Aban, J (2003a) SIGSIM User Guide Part A: SIGSIM theory (version 2.0). *Working Paper* from the Centre for Transport Studies, University College London.
- Sha'Aban, J (2003b) SIGSIM User Guide Part B: Serial SIGSIM user guide (version 2.0). *Working Paper* from the Centre for Transport Studies, University College London.
- Sutton R and Barto A (1998) *Reinforcement Learning*. MIT Press.
- Smith RE (1992) Memory exploitation in learning classifier systems. *Evolutionary Computation*, 2(3), 199-220.
- Tomlinson A and Bull L (1998) A Corporate Classifier System. In: AE Eiben, T Baeck, M Schoenauer and H-P Schwefel (eds) *Parallel Problem Solving from Nature - PPSN V*, Springer Verlag, 471-80.
- Tomlinson A and Bull L (2001) Symbiogenesis in Learning Classifier Systems. *Artificial Life* 7(1), 33-62.
- Tomlinson A and Bull L (2002) An Accuracy-Based Corporate Classifier System. *Soft Computing* 6(3-4), 200-15.
- Van Zuylen HJ (1976) Acyclic traffic controllers. Note from the Centre for Transport Studies, UCL.
- Vincent RA, Mitchell AI and Robertson DI (1980) User guide to TRANSYT version 8. Transport and Road Research Laboratory *Report LR 888*. Crowthorne: TRL.

- Vincent RA and Peirce JR (1988) MOVA: traffic responsive, self-optimising signal control for isolated intersections. Transport and Road Research Laboratory, *Report RR 170*. Crowthorne: TRL.
- Wilson SW (1994) ZCS: A Zeroth-level Classifier System. *Evolutionary Computation* **2**(1), 1-18.
- Wilson SW (1995) Classifier Fitness Based on Accuracy. *Evolutionary Computation*, **3**(2), 149-76.
- Wilson SW and Goldberg DE (1989) A critical review of classifier systems. In: JD Schaffer (ed) *Proceedings of the 3rd International Conference on Genetic Algorithms*, Morgan Kaufman, 244-55.
- Wyatt, D and Bull, L (2002) An Investigation into Island Model Rule Migration for a Number of Mobile Autonomous Learning Classifier System Agents. *Technical Report UWELCSG02-005*. Available from [http:// www.cems.uwe.ac.uk/lcsg](http://www.cems.uwe.ac.uk/lcsg)