Journal of the Korea Academia-Industrial Cooperation Society
Vol. 12, No. 11 pp. 5135-5142, 2011

http://dx.doi.org/10.5762/KAIS.2011.12.11.5135

Q 학습을 이용한 교통 제어 시스템

장 정[1], 승지훈[2], 김태영[2], 정길도[2*]
[1]Mechanical Engineering, Xian Jiaotong University, China
[2]전북대학교 전자공학부

Traffic Control using Q-Learning Algorithm

Zhang Zheng[1], Ji Hoon Seung[2], Tae Yeong Kim[2] and Kil To Chong[2*]
[1]**Mechanical Engineering Xian Jiaotong University, China**
[2]**Electronics and Information Department, Chonbuk National University**

요 약 이 논문에서는 도심 지역의 교통 제어 시스템의 동적 응답 성능 향상을 위하여 적응형 Q-Learning 강화 학습 메커니즘을 설계 하였다. 도로, 자동차, 교통 제어 시스템을 지능 시스템으로 모델링 하고, 자동차와 도로 사이는 무선 통신을 이용한 네트워크가 구성된다. 도로와 대로변에 필요한 센터네트워크가 설치되고 Q-Learning 강화 학습은 제안한 메커니즘의 구현을 위해 핵심 알고리즘으로 채택하였다. 교통 신호 제어 규칙은 자동차와 도로에서 매 시간 업데이트된 정보에 따라서 결정되며, 이러한 방법은 기존의 교통 제어 시스템에 비하여 도로를 효율적으로 활용하며 결과적으로 교통 흐름을 개선 한다. 알고리즘을 활용한 최적의 신호 체계는 온라인상에서 자동으로 학습함으로서 구현된다. 시뮬레이션을 통하여 제안한 알고리즘이 기존 시스템에 비하여 효율성 개선과 차량의 대개 시간에 대한 성능 지수가 모두 30% 이상 향상되었다. 실험 결과를 통하여 제안한 시스템이 교통 흐름을 최적화함을 확인하였다.

Abstract A flexible mechanism is proposed in this paper to improve the dynamic response performance of a traffic flow control system in an urban area. The roads, vehicles, and traffic control systems are all modeled as intelligent systems, wherein a wireless communication network is used as the medium of communication between the vehicles and the roads. The necessary sensor networks are installed in the roads and on the roadside upon which reinforcement learning is adopted as the core algorithm for this mechanism. A traffic policy can be planned online according to the updated situations on the roads, based on all the information from the vehicles and the roads. This improves the flexibility of traffic flow and offers a much more efficient use of the roads over a traditional traffic control system. The optimum intersection signals can be learned automatically online. An intersection control system is studied as an example of the mechanism using Q-learning based algorithm, and simulation results showed that the proposed mechanism can improve the traffic efficiency and the waiting time at the signal light by more than 30% in various conditions compare to the traditional signaling system.

Key Words : Intelligent Transportation System; Cooperative Vehicle-Highway Systems; Reinforcement Learning; Traffic Control Mechanism; Intersection Signal Control

1. 서론

지능 수송 체계 시스템에서는 모든 종류의 수송 체계를 개발하기 위해 시스템 공학 개념과 전자통신기술을 도입하여 이용한다[1]. 지능형 자동차 분야는 첨단 전자기술 특히 무선 통신 및 첨단 센서의 출현과 함께 발전하고 있으며, 연구가 활발히 진행되고 있다. 한국을 포함한 세계 여러 나라에서는 교통 체증을 효율적으로 완화하고 안전성을 증진시키기 위해 새로운 교통시스템 개발을 추진하고 있다. 지능형 교통시스템(ITS)은 이러한 필요성에 의하여 개발하고 있는 시스템으로서, 도로, 차량, 신호시스템 등 기존 교통체계의 구성요소에 전자·제어·통신 등

*교신저자 : 정길도(kitchong@jbnu.ac.kr)
접수일 11년 09월 02일 수정일 (1차 11년 10월 11일, 2차 11년 11월 04일) 게재확정일 11년 11월 10일

첨단기술을 접목시켜 구성요소들이 상호 유기적으로 작동하도록 시스템화한 차세대 교통체계를 말한다. 특히 시내 교통 신호 시스템의 최적화는 교통흐름에 직접적인 영향을 미치며 인공지능을 가미한 다양한 연구가 진행되고 있다. 교통체계의 요소들을 에이전트로 설정하고 멀티 에이전트에서 수집한 정보를 활용하며 통계학적 원리를 이용한 Q-learning 강화 학습 등은 매우 중요한 연구 분야로서 활발한 연구가 진행되고 있다.

강화학습은 최적의 행동정책을 구하는 최적화 문제로 주어진 환경과의 상호작용을 통해 보상 값을 최대화하는 것을 목표로 한다. 지능 교통 시스템에서는 에이전트가 현재 교통 상황을 인식하여 적절한 행동을 선택하는 행동 선택 문제에 대한 많은 연구가 진행되어 왔다. 지능형 에이전트의 학습에 있어서 정해진 전략을 사용하는 교사 학습보다는 환경을 감지하여 스스로 최적의 전략을 세울 수 있는 강화학습과 같은 비 교사 학습이 더욱 효과적이다[2].

많은 연구자들은 교통신호체계의 효율을 높이기 위하여 다양한 연구 결과들을 발표하였다. Wei Wu는 멀티 에이전트를 기반으로 한 도시의 교통신호 제어를 연구했으며 논문에서 멀티 에이전트 기술을 적용하여 교통 신호 제어 에이전트의 구조를 구상하였고, 강화 학습 알고리즘을 구현하였다[3]. Min Chee Choy[4]와 Anastasios Kouvelas[5]는 실시간 신호 제어를 위한 하이브리드 에이전트 구조를 소개했으며, [4]에서 매 평가 기간을 위해 제어기 에이전트의 모든 사항을 저장하는 동적 데이터베이스를 제시했다. [5]에서는 새로운 도심지역 감응식 제어(TUC)기법을 소개했다. Dipti Srinivasan[6]는 멀티 에이전트에서 교통 신호 제어 문제의 해결을 위해 신경회로망을 활용하여 현재의 교통 신호 알고리즘과 비교하였고 멀티 에이전트 교통 신호 방법 SPSA(simultaneous perturbation stochastic approximation) 신경회로망을 제안하였다. Balaji Parasumanna Gokulan[7]와 P.G. Balaji[8]는 교통 신호 제어를 위해 멀티 에이전트 기반 제어 방식을 소개하였다. [9]에서 최적의 교통 신호 제어를 위해 멀티 에이전트 구조를 제안했고, 이 멀티 에이전트는 PARAMICS 소프트웨어를 통해 제안한 제어 기법에 대한 성능을 평가하였다. 하지만 위의 기술과 방법들은 시간에 따라 변하는 유동적인 교통 상황에 적응적이지 않다.

Dennis I. Robertson[9]은 SCOOT 시스템을 기존의 Transyt 시스템에 적용하는 연구를 수행하였다. 여기에서 센서들은 순환 흐름 양상(CFPs)을 검출하기 위해 사용되고, SCOOT 최적화에 대한 연구를 수행하였다. Liu Guang-ping[10]은 신호 제어에 있어 교차점 지연의 계산 방법을 연구했다. [10]에서는 교차점 지연의 계산하는 방법과 어떻게 계산된 지연을 신호 제어에 사용할 것에 대해 다뤘으며, Weihua Bao[11]는 독립된 교차로에서 대중교통 우선을 위한 교통 신호 타이밍 계획에 대해 연구했다. LIAO Yongquan[12]은 교차로를 통과하는 차량의 지연을 줄이기 위해 교차로의 신호 타이밍을 Q 학습 기법으로부터 최적화하는 방법을 제안했다. 위 논문들은 교통 신호의 고정된 주기를 사용하는 방법으로 환경의 변화를 적응할 수 없는 방법들이다 [13][14].

본 논문에서는 교통시스템의 모든 요소들을 에이전트로 구성하고 실시간 도로 환경에 적응이 가능한 멀티 에이전트 Q-Learning 강화 학습 알고리즘을 제안하였다. 결과적으로 다음과 같은 특징을 가진 알고리즘을 개발하였다.

1) 새로운 교통 흐름 제어 메커니즘은 자동차, 도로, 교통 관리 시스템을 에이전트로 설정하고, 대로변의 무선 통신 네트워크는 동적인 교통 흐름 제어 방법을 제공한다.
2) 강화 학습은 성능 향상을 위하여 교통 흐름을 동역학적으로 계획하는 알고리즘으로 소개된다.
3) 교차로 교통 신호 제어 시스템으로 구성된 Q-학습은 제안된 메커니즘의 예제로 연구 된다.

이러한 연구 결과의 효율적인 설명을 위하여 논문을 다음과 같이 구성하였다. 2장에서는 Q-학습알고리즘을 소개하고 3장은 교차로 신호 시스템 모델을 기술하였다. 4장에서 앞서 제안한 알고리즘의 성능을 실제 사례에 적용하여 실험하였고, 마지막으로 5장에는 이 연구의 결론에 대해 서술하였다.

2. Q-학습 알고리즘

강화 학습 유형의 Q-학습은 에이전트가 주변 환경에서 행동의 효과에 대한 이전 지식이 없는 경우에도 지연 보상으로부터 최적 제어 전략을 개발할 수 있다[15].

에이전트의 학습은 다음과 같이 설명된다. 에이전트가 모든 상태(s)에 대하여 $V^{\pi}(s)$를 최대화하는 방법 π를 학습하는 것이 요구된다. 이 방법을 최적 방법(Optimal policy)이라 하고 π^{*}로 표시한다.

$$\pi^{*} \equiv \arg_{\pi}\max V^{\pi}(s),(\forall s) \tag{1}$$

표기법을 단순화하기 위해, $V^{*}(s)$와 같은 최적 방법의 가치 함수 $V^{\pi^{*}}(s)$를 참조한다. $V^{*}(s)$는 에이전트가 상태 s로부터 얻은 초기보상에서 최대 감소된 누적 보상을 부여한다.

이것은 상태 s에서 초기 최적 방법을 따름으로써 감소된 누적 보상이 행해진다. 그러나 이용 가능한 연습 데이터가 $<s,a>$형태의 연습 예제를 제공하지 않기 때문에 직접적으로 함수$\pi^* : S \rightarrow A$를 학습하는 것은 어렵다. 또한, $i = 0, 1, 2, ...$일 경우 학습자에게 제공되는 유일한 연습 정보는 즉각적인 보상 $r(s_i, a_i)$의 연속이다. 따라서 상태와 행동으로부터 정의된 수치적 평가함수를 학습하는 것이 평가 함수의 관점에서 최적 방법을 충족하는 것보다 쉽다.

평가함수를 학습하기 위한 에이전트의 시도는 무엇인가? 하나의 명확한 선택은 V^*이다. $V^*(s_1) > V^*(s_2)$때에 에이전트는 상태s_2 로부터 상태 s_1을 선호한다. 왜냐하면 누적 미래 보상이 s_1에서 더 크기 때문이다. 물론 에이전트의 방법은 반드시 상태들 사이에서가 아니라 행동들 사이에서 선택한다. 그러나 더 좋은 행동들 사이에서 선택하기 위해 특정 설정 안에서 V^*를 사용한다. 상태 s에서 최적 행동은 행동 a이며, a는 즉각적인 보상 $r(s,a)$에 γ를 합하여 감소된 즉각적인 성공 상태를 최대화한다.

$$\pi^* = \arg_a \max [r(s,a) + \gamma V^*(\delta(s,a))] \quad (2)$$

여기에서 $\delta(s,a)$는 적용한 행동 a에서 상태 s까지 상태 결과를 표시한다. 만약 즉각적인 보상 함수 γ와 상태 전달 함수 δ의 완벽한 정보를 가지고 있다면 에이전트는 V^*를 학습함으로써 최적 정책을 획득할 수 있다. 에이전트가 행동에 응답 환경을 사용하는 함수 r과 δ를 알 때, 어떠한 상태 s에 대해서도 최적 행동을 계산하기 위해 식(2)을 사용 할 수 있다. 학습한 V^*는 에이전트가 δ와 γ의 완벽한 정보를 가지고 있을 때에만 최적 정책을 학습하기에 유용한 방법이다.

평가 함수 $Q(s,a)$를 살펴보자, 이것은 최대 감소된 누적 보상이다. 이것은 처음 행동으로써 상태 s와 적용 행동 a로부터 시작하는 것으로 이루어 질 수 있다. 반면에, 아래의 식으로 표현되는 Q의 값은 상태 s로부터 즉각적으로 시행한 행동 a에서 얻어진 보상에 뒤따르는 최적 정책의 값을 더한 것이다.

$$Q(s,a) \equiv r(s,a) + \gamma V^*(\delta(s,a)) \quad (3)$$

$Q(s,a)$는 정확하게 상태 s에서 최적 행동 a를 선택하기 때문에 식(2)은 최대화된 수치가 된다. 그러므로 식(2)은 Q에 관련하여 다시 다음과 같이 쓸 수 있다.

$$\pi^* = \arg_a \max Q(s,a) \quad (4)$$

식(3)를 식(4)로 표현하는 중요한 이유는 만약 에이전트가 V^*대신 Q함수를 학습한다면 함수r과 함수 δ에 대한 지식이 없을 지라도 최적 행동을 선택하는 것이 가능하기 때문이다. 마찬가지로 식(4)도 조금 더 명확해진다. 현재 상태 s에서 각각 가능한 행동 a만을 고려한다.

Q-학습 알고리즘의 구현은 어떻게 이루어지는가? 중요한 문제는 Q에 대해 연습 값들을 추정하기 위한 확실한 방법을 시간의 흐름에 따라 주어진 즉각적인 보상 r의 신호를 이용하여 구하는 것이다. 이것은 반복적인 근사를 통해 이루어진다. 방법을 보면 Q와 V^*사이의 밀접한 관계($V^*(s) = \max_{a'} Q(s,a')$를 통해 식(3)는 다음과 같이 다시 쓸 수 있다.

$$Q(s,a) = r(s,a) + \gamma \max_{a'} Q(\delta(s,a),a') \quad (5)$$

식(5)은 반복적으로 Q를 근사하는 알고리즘의 기초를 제공한다[16]. 이 알고리즘에서, $\overline{Q}$는 실제 Q-함수의 추정 또는 가정한 값이다. $\overline{Q}$는 각각의 상태 행동 추정에 대해 분리된 입력을 가진 큰 표로 나타내어진다. 표는 임의의 값을 가지고 초기 정렬된다.(만약 하나가 초기 값으로 0을 가진다면 알고리즘을 이해하는 것은 더욱 간단하다). 각각의 변화에 따른 $\overline{Q}(s,a)$에 대해 입력 표는 다음의 규칙에 의해 업데이트 된다.

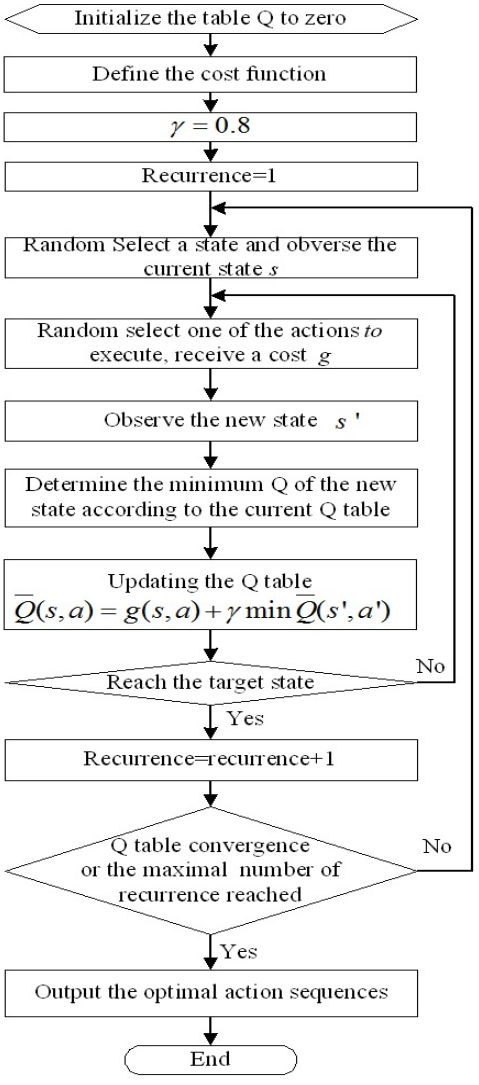
$$\overline{Q}(s,a) \leftarrow r(s,a) + \gamma \max_{a'} \overline{Q}(s',a') \quad (6)$$

위의 연습 규칙은 이전 상태 s에 대해 $\overline{Q}(s,a)$의 추정을 위해 새로운 상태 s'에 대한 에이전트의 현재 $\overline{Q}$값들을 사용하는 것이다. 위의 결정론적인 Markov 해결에 대한 Q-학습 알고리즘은 그림 1에서 설명된다. 그림 1의 흐름도에서 보상함수 $\gamma(s,a)$대신에 손실 함수 $g(s,a)$를 사용한다. 그러므로 반복되는 연습 방법 식(6)은 다음과 같이 대체할 수 있다.

$$\overline{Q}(s,a) \leftarrow g(s,a) + \gamma \max_{a'} \overline{Q}(s',a') \quad (7)$$

이것은 행동이 최적 행동 순서에 기반을 둘 때 학습 목표는 정체 평가치를 최소화함으로써 Q함수를 최소화

하는 것을 의미하며, 이 논문에서 사용된 알고리즘이다.



[그림 1] Q학습 알고리즘의 흐름도
[Fig. 1] Flowchart of Q learning algorithm

3. 교차로 신호 시스템

3.1 신호시스템 모델

교통 시스템은 다양한 요소로 이루어져 있고 교차로는 가장 중요한 것들 중 하나이다[17].

연구된 방법은 각각 몇 개의 차선과 그림 2와 같은 자동차의 흐름에 의존하는 교통신호의 설정을 가지는 두 개의 도로가 교차하는 교차로에 적용된다.

차선 ①, ③, ⑤ 과 ⑦에서 자동차는 교차로에 접근하며 ②, ④, ⑥ 과 ⑧에서 자동차는 교차로에서 빠져나간다. 각각의 접근 차선에 대해, 그림 2와 같이 좌회전, 우회전, 직진의 세 가지 선택을 할 수 있다.

우회전 방향은 다른 방향을 방해하지 않기 때문에 고려하지 않는다. 모델에서 문제를 간단하게 하기 위해서 보행자는 고려하지 않는다. 그러므로 이 문제는 표 1에서 보인 것과 같이 다른 경로에 대해 8개의 대기 열로써 모델링 된다.

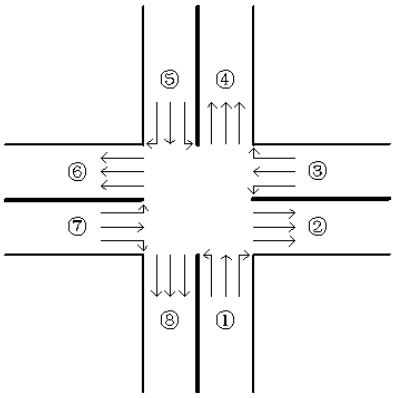
신호가 시작될 때 모든 큐에서 자동차들이 임의로 대기하고 있다고 가정하며, 이것은 환경의 초기 상태가 된다. 그리고 마지막 상태는 초기 상태에서 모든 자동차들은 반드시 교차로를 통과한 것이다. 교차로 신호 제어 시스템은 교차로 주위에서 모든 자동차 에이전트의 행동을 관리하기 위해 주된 에이전트로써 모델링 된다.

따라서 자동차 에이전트의 행동 자료는 A1에서 A8까지 행동을 포함하고 주된 에이전트는 마지막 상태에 도달하기 위한 적절한 행동 또는 그들의 합리적 조합을 선택한다. 만약 A1에서 A8까지 행동 중에 두 가지가 서로 간섭하지 않는다면 그것은 가능한 행동 조합이 되며, 이것을 신호 위상이 다른 조합이라 부른다.

모든 가능한 조합은 표 2와 같다. 그러므로 문제는 마지막 상태에 도달하기 위해 행동 조합들의 최적 순서를 어떻게 찾을 것인지에 달려있다. 이것은 교차로 신호 제어 에이전트의 주된 함수이다. 초기 상태에서 마지막 상태에 이르기까지 각각의 이산 상태에 대해, 최적 정책은 이전 상태에 대해 독립이다.

[표 1] 다른 대기열의 기본 행동 정의
[Table 1] Basic action definition of different queues

Queue	Basic action symbol	Path
Que1	A1	①→④
Que2	A2	①→⑥
Que3	B1	⑤→⑧
Que4	B2	⑤→②
Que5	C1	③→⑥
Que6	C2	③→⑧
Que7	D1	⑦→②
Que8	D2	⑦→④



[그림 2] 교차로
[Fig. 2] Isolated intersection

[표 2] 행동 조합 표시
[Table 2] Action combination symbol

Phase	Action combination symbol	Component
Ph1	Ac1	A1+A2
Ph2	Ac2	A1+B1
Ph3	Ac3	B1+B2
Ph4	Ac4	C1+C2
Ph5	Ac5	C1+D1
Ph6	Ac6	D1+D2

하나의 행동 조합이 이루어진 후에 성공적인 상태는 결정론적이다. 그러므로 이 문제는 결정론적 Markov 해결 과정으로써 모델링 할 수 있다.

3.2 학습과정의 파라미터

결정론적인 Markov 해결 과정에 대한 학습 파라미터를 살펴보면 다음과 같다.

3.2.1 평가 함수

상태 s에서 자동차의 수 n을 가정하고 선택된 행동 a가 완료되면 현재 자동차 수는 n_1이 된다. 이 행동의 평가는 대기시간 t 와 교차로에 남아있는 자동차수 n_1에 의존한다.

$$g(s,a) = n_1 \times (t + t_{transition}) \tag{8}$$

여기서 $t_{transition}$는 [표 3]에서 보여 진 것처럼 세 가지 수{0, 1.5, 3}중에 하나와 같다. 도로를 지나는 각 자동차에 대한 평균 시간은 3초로 가정한다.

[표 3] 단계별 이동 시간 $t_{transition}$
[Table 3] $t_{transition}$ of different phase transition

Phase transition type	Comment	$t_{transition}$(s)
No transition	Current phase is the same as the previous one	0
Half transition	$Ac1 \Leftrightarrow Ac2; Ac2 \Leftrightarrow Ac3;$ $Ac4 \Leftrightarrow Ac5; Ac5 \Leftrightarrow Ac6$	1.5
Full transition	Phase transfer except half transition	3

3.2.2 상태 전달 함수

표 4는 서로 다른 단계 행동을 시행할 때 어떻게 계승 상태를 결정할 것인지를 설명한다. 이 표에서, 각 대기열 자동차 수는 0 보다 크다고 가정한다. 만약 대기 열에서 자동차의 수가 0이 된다면, 대응하는 행동은 현재 상태를 변화시키지 않는다.

[표 4] 상태 전달 규칙
[Table 4] Rules of state transition

Phase	Current state	Successor state
Ph1	{Que1, Que2, Que3, Que4, Que5, Que6, Que7, Que8}	{(Que1-1), (Que2-1), Que3, Que4, Que5, Que6, Que7, Que8}
Ph2	{Que1, Que2, Que3, Que4, Que5, Que6, Que7, Que8}	{(Que1-1), Que2, (Que3-1), Que4, Que5, Que6, Que7, Que8}
Ph3	{Que1, Que2, Que3, Que4, Que5, Que6, Que7, Que8}	{Que1, Que2, (Que3-1), (Que4-1), Que5, Que6, Que7, Que8}
Ph4	{Que1, Que2, Que3, Que4, Que5, Que6, Que7, Que8}	{Que1, Que2, Que3, Que4, (Que5-1), (Que6-1), Que7, Que8}
Ph5	{Que1, Que2, Que3, Que4, Que5, Que6, Que7, Que8}	{Que1, Que2, Que3, Que4, (Que5-1), Que6, (Que7-1), Que8}
Ph6	{Que1, Que2, Que3, Que4, Que5, Que6, Que7, Que8}	{Que1, Que2, Que3, Que4, Que5, Que6, (Que7-1), (Que8-1)}

3.2.3 감쇠인자

시뮬레이션에서 감쇠인자는 $\gamma = 0.8$로 설정한다.

4. 시뮬레이션 및 결과

제안된 알고리즘의 성능을 확인하기 위해, 전통적인 신호 체계를 포함하는 시뮬레이션을 실시했다. 전통적인 신호 체계는 Ph1, Ph2, Ph3, Ph4, Ph5 와 Ph6와 같은 고정된 순차로 구성된다. 그러나 제안된 방법은 개선된 상황을 기초로 자동적으로 최적 단계 순서를 결정한다.

다양한 시간 간격에 대하여 실험을 실시했으나 대부분의 실험 결과가 유사하고, 또한 데이터가 방대하므로 본 논문에서는 90초에 대한 실험 결과에 대하여 살펴보도록 하자.

시뮬레이션 결과를 표 5와 표 6을 이용하여 표시하였으며, 표 5는 시뮬레이션 시간, 랜덤 큐와 큐의 순서를 나타내고 표 6은 기존의 방법과 본 연구에서 제안한 방법의 결과를 나타낸다.

[표 5] 시뮬레이션 결과 1 ($t_{phase}=90s$)
[Table 5] Simulation result 1($t_{phase}=90s$)

P_s	*Random Queues*	O_A
1	{30 30 60 30 30 30 30 30}	{4 5 6 1 2 3}
2	{60 30 60 30 29 29 30 30}	{6 5 4 1 2 3}
3	{27 27 29 29 60 30 29 29}	{3 2 1 6 5 4}
4	{27 27 26 26 26 26 29 29}	{1 2 3 6 5 4}
5	{30 30 52 26 30 30 54 27}	{1 2 3 4 5 6}
6	{56 28 54 27 30 30 29 29}	{4 5 6 1 2 3}
7	{27 27 29 29 60 30 30 30}	{3 2 1 6 5 4}
8	{44 22 27 27 46 23 44 22}	{4 5 6 3 2 1}
9	{42 21 23 23 20 20 30 30}	{6 5 4 3 2 1}
10	{19 19 46 23 27 27 44 22}	{1 2 3 4 5 6}
11	{28 28 58 29 46 23 54 27}	{6 5 4 1 2 3}
12	{19 19 27 27 50 25 44 22}	{3 2 1 4 5 6}
13	{56 28 50 25 44 22 46 23}	{6 5 4 1 2 3}
14	{58 29 13 13 20 10 13 13}	{6 5 4 3 2 1}
15	{24 24 24 12 14 14 34 17}	{1 2 3 4 5 6}
16	{30 30 32 16 24 12 20 10}	{4 5 6 1 2 3}
17	{16 8 5 5 50 25 30 30}	{3 2 1 6 5 4}
18	{26 13 6 3 16 16 29 29}	{3 2 1 6 5 4}
19	{58 29 22 11 24 12 30 30}	{6 5 4 3 2 1}
20	{14 7 6 6 26 13 32 16}	{3 2 1 6 5 4}
21	{18 18 21 21 22 22 29 29}	{3 2 1 6 5 4}
22	{21 21 4 2 18 9 46 23}	{1 2 3 4 5 6}
23	{54 27 10 10 7 7 34 17}	{4 5 6 3 2 1}
24	{16 8 50 25 48 24 30 15}	{1 2 3 4 5 6}
25	{6 6 30 15 7 7 27 27}	{6 5 4 1 2 3}

표에서 P_s는 시뮬레이션 기간, N_{IV} 는 초기 상태에서 자동차의 전체 수, Random Queues는 임의로 생성된 자동차 대기열 수, T_{IQ}는 Q-학습 방법에 대한 초기 상태에 마지막 상태에 이르기 까지 시간 간격, T_{WQ} 는 Q-학습 방법에 대한 전체 대기 시간, T_{IT} 는 전통 적인 방법에 대한 초기 상태에서 마지막 상태에 이르기 까지 시간 간격을 의미한다.

[표 6] 시뮬레이션 결과 2 ($t_{phase}=90s$)
[Table 6] Simulation result 2 ($t_{phase}=90s$)

P_s	N_{IV}	T_{IQ} (s)	T_{WQ} (s)	T_{WT} (s)	P_{EI} (%)	P_{WD} (%)	T_L (s)
1	270	462	41850	54900	14.44	23.77	0.8594
2	298	459	54984	57060	15.00	3.64	0.9375
3	260	447	38712	56070	17.22	30.96	0.8594
4	216	336	27198	44730	37.78	39.20	0.9063
5	279	510	50319	62190	5.56	19.09	0.8438
6	283	438	49604	54720	18.89	9.35	0.9063
7	262	450	39330	56970	16.67	30.96	0.9219
8	255	429	40184	54810	20.56	26.69	0.8750
9	209	357	25260	44460	33.89	43.18	0.9375
10	227	420	33804	51300	22.22	34.11	0.9063
11	293	501	52989	66060	7.22	19.79	0.9531
12	233	366	33731	54270	32.22	37.85	0.9531
13	294	459	55809	60210	15.00	7.31	0.9375
14	169	324	16779	25560	40.00	34.35	0.8594
15	163	300	16845	33750	44.44	50.09	0.8906
16	174	300	18441	27360	44.44	32.60	0.9063
17	169	315	16722	45000	41.67	62.84	0.9531
18	138	234	12132	34560	56.67	64.90	0.9375
19	216	381	29658	40680	29.44	27.09	0.9219
20	120	207	10127	30150	61.67	66.41	0.9375
21	180	282	18882	40140	47.78	52.96	0.4844
22	144	252	12222	37350	53.33	67.28	0.4531
23	166	327	16926	30690	39.44	44.85	0.4219
24	216	375	31728	47790	30.56	33.61	0.4688
25	125	222	8277	32580	58.89	74.59	0.468

표 6로부터, 모든 구간에서 Q 학습 프로그램의 총 진행 시간 T_L 은 1 초 보다 작다. 이것은 교차로 신호 제어 시스템의 적용할 수 있는 매우 짧은 시간이다. 표를 통하여 제안한 시스템이 기존의 일반적인 신호체계에 비하여 성능이 개선됨을 알 수 있다. 표에서 교통 효율 개선 확

률은 P_{EI}로 표시되며, 기존 방법에 비하여 최소 5.56%에서 최대 61.67%의 성능향상을 보인다. 그리고 전체 대기시간 감소 확률(P_{WD})은 최소 3.64%에서 최대 67.28%로 성능이 향상됨을 알 수 있다. 그리고 기존 시스템에 비하여 제안한 시스템의 효율성 개선P_{EI}의 평균 확률은 32.2%이다. 그리고 신호등에서 차량의 대기 시간에 대한 성능 지수 P_{WD}의 향상은 32.2%이다.

5. 결론

강화 학습과 멀티 에이전트 모델링 방법의 조합에 기초한 새로운 교통 제어 시스템이 지능 교통 시스템으로 제안 되었다. 제어 시스템, 자동차 그리고 몇 가지 필수적인 도로 센서는 제안된 시스템에서 지능 에이전트로 모델링 하므로 ITS 시스템은 멀티 에이전트 시스템으로 구성됨을 살펴보았다. 또한 지능형 교통시스템이 인공적인 지능 알고리즘으로 효율성을 개선할 수 가능성을 살펴보았다. 특히 독립된 교차로에 대해 Q-Learning 강화학습 방법을 활용한 교통신호의 제어 방법을 제안하였다. 교차로 신호는 제안된 알고리즘에 따라 모델링 되었고, Q-학습을 기반을 둔 새로운 알고리즘을 설계하였다. 제안된 알고리즘이 탑재된 시스템에 대한 시뮬레이션을 수행하였고, 전통적인 교차로 신호 방법의 결과와 비교했다. 시뮬레이션 결과 제안된 알고리즘이 전통적인 방법에 비하여 30%이상 개선되었고 동시에 운전자의 대기시간을 30% 이상 줄일 수 있음을 확인할 수 있었다.

References

[1] http://www.ewh.ieee.org/tc/its/.

[2] Bong Keun Lee, Jae Du Chung, Keun Ho Ryu, "Multi-Agent Reinforcement Learning Model based on Fuzzy Inference", International Journal of Contents Vol. 9 No. 10, pp. 51-58, 2009.

[3] Wei Wu, Gong Shufeng, Liu Hongxiu, "A Coordinated Urban Traffic Signal Control Approach based on Multi Agent," Intelligent Engineering Systems, 2009. INES 2009. International Conference on, pp. 263-267, 2009

[4] Min Chee Choy, Dipti Srinivasan, Ruey Long Cheu. "Cooperative, Hybrid agent Architecture for Real-Time traffic Signal Control." IEEE Transaction on Systems, Man, and Cybernetics - Part A: Systems and Humans, Vol. 33, No. 5, pp. 597-607, Sep. 2003.

[5] Kouvelas, A., Aboudolas, K., Papageorgiou, M., Kosmatopoulos, E. B., "A Hybrid Strategy for Real Time Traffic Signal Control of Urban Road Networks, Intelligent Transportation Systems," IEEE Transactions on, pp. 1-11, 2011.

[6] Srinivasan, D., Min Chee Choy, Cheu, R.L., "Neural Networks for Real-Time Traffic Signal Control, Intelligent Transportation Systems," IEEE Transactions on, pp. 261-272, 2006.

[7] Gokulan, B.P., Srinivasan, D., "Distributed Geometric Fuzzy Multiagent Urban Traffic Signal Control, Intelligent Transportation Systems," IEEE Transactions on, pp. 714-727, 2010.

[8] Balaji, P.G., German, X., Srinivasan, D.,"Urban Traffic Signal Control Using Reinforcement Learning Agents, Intelligent Transport Systems", IET, pp. 177-188, 2010.

[9] Robertson, D.I., Bretherton, R.D., "Optimizing Networks of Traffic Signals in Real Time-The SCOOT Method," Vehicular Technology, IEEE Transactions on, pp. 11-15, 1991.

[10] LIU Guang-ping, ZHAI Run-ping and PEI Yu-long. "A Calculating Method of Intersection Delay under Signal Control." Proceedings of the 2007 IEEE Intelligent Transportation Systems Conference, Seattle, WA, USA, Sept. 30 - Oct. 3, pp. 1114-1119, 2007.

[11] Weihua Bao, Quanlin Chen and Xiaoxia Xu. "An Adaptive Traffic Signal Timing Scheme for Bus Priority at Isolated Intersection." Proceedings of the 6th World Congress on Intelligent Control and Automation, June 21-23, Dalian, China, pp. 8712-8716, 2006,

[12] Liao Yongquan, Cheng Xiangjun, "Study on Traffic Signal Control Based on Q-Learning," Fuzzy Systems and Knowledge Discovery, 2009. FSKD '09. Sixth International Conference on, pp. 581-585, 2009.

[13] D. Srinivasan and M.C. Choy. "Cooperative multi-agent system for coordinated traffic signal control." IEE Proc. Intell. Transp. Syst. Vol. 153, No. 1, pp. 41-50, March 2006.

[14] Lee, J.H., and Lee-Kwang, H. "Distributed and cooperative fuzzy controllers for traffic intersections group." IEEE Trans. Syst. Man Cybern. C, Appl. Rev. 29, (2), 1999.

[15] Michael Wooldridge. "An Introduction to Multi agent Systems." JOHN WILEY & SONS, LTD. Baffins Lane, Chichester, West Sussex PO191UD, England, 2002.

[16] Watkins, C., & Dayan, P. "Q-learning. Machine Learning," 8, pp. 279-292, 1997.
[17] A. D'Ambrogio et al., "Simulation model building of traffic intersections, Simulat. Modell. Pract. Theory", 2008.

장 정(Zhang Zheng) **[정회원]**

- 2005년 : 중국 Xi'an Jiaotong University 기계공학(공학박사)
- 2011년 3월 ~ 현재 : 중국 Xi'an Jiaotong University 기계공학 강의 교수

<관심분야>
Nonlinear time-delay systems, 지능형 교통제어 시스템, 지능형 자동차, 로보틱스

승 지 훈(Ji Hoon Seung) **[준회원]**

- 2010년 2월 : 전북대학교 전자공학과(공학사)
- 2010년 3월 ~ 현재 : 전북대학교 대학원 전자정보공학부 석사과정

<관심분야>
Parameter Estimation, Navigation, Filtering,

김 태 영(Tae-Yeong Kim) **[준회원]**

- 2009년 2월 : 전북대학교 전자공학과(공학사)
- 2009년 3월 ~ 현재 : 전북대학교 대학원 전자정보공학부 석사과정

<관심분야>
Navigation, Robotics, Filtering

정 길 도(Kil To Chong) **[정회원]**

- 1984년 2월 : Oregon State University 기계공학(공학사)
- 1986년 2월 : Georgia Institute of Technology 기계공학(공학석사)
- 1992년 2월 : Texas A&M University 기계공학 (공학박사)
- 2010년 3월 ~ 현재 : 전북대학교 전자정보공학부 교수

<관심분야>
Marine Navigation, Time-Delay, Robotics, 인공지능, 지능형 교통시스템