Intelligent Traffic Control by Multi-agent Cooperative Q Learning (MCQL)

Deepak A. Vidhate and Parag Kulkarni

Abstract Traffic crisis frequently happens because of traffic demands by the large number vehicles on the path. Increasing transportation move and decreasing the average waiting time of each vehicle are the objectives of cooperative intelligent traffic control system. Each signal wishes to catch better travel move. During the course, signals form a strategy of cooperation in addition to restriction for neighboring signals to exploit their individual benefit. A superior traffic signal scheduling strategy is useful to resolve the difficulty. The several parameters may influence the traffic control model. So it is hard to learn the best possible result. The lack of expertise of traffic light controllers to study from previous practice results makes them to be incapable of incorporating uncertain modifications of traffic flow. Defining instantaneous features of the real traffic scenario, reinforcement learning algorithm based traffic control model can be used to obtain fine timing rules. The projected real-time traffic control optimization model is able to continue with the traffic signal scheduling rules successfully. The model expands traffic value of the vehicle, which consists of delay time, the number of vehicles stopped at the signal, and the newly arriving vehicles to learn and establish the optimal actions. The experimentation outcome illustrates a major enhancement in traffic control, demonstrating the projected model is competent of making possible real-time dynamic traffic control.

Keywords Cooperative learning · Multi-agent learning · Q learning

D. A. Vidhate (✉)
Department of Computer Engineering, College of Engineering, Pune, Maharashtra, India
e-mail: dvidhate@yahoo.com

P. Kulkarni
iKnowlation Research Lab. Pvt. Ltd, Pune, Maharashtra, India
e-mail: parag.india@gmail.com

S. Bhalla et al. (eds.), *Intelligent Computing and Information and Communication*,
Advances in Intelligent Systems and Computing 673,
https://doi.org/10.1007/978-981-10-7245-1_47

1 Introduction

Thousands of vehicles distribute in a large and board urban area. It is a difficult and complicated work to effectively take care of such a large scale, dynamic, and distributed system with a high degree of uncertainty [1]. Though the number of vehicles is getting more and more in major cities, most of the current traffic control methods have not taken benefit of an intelligent control of traffic light [2]. It is observed that sensible traffic control and enhancing the deployment effectiveness of roads is an efficient and cost-effective technique to resolve the urban traffic crisis in majority urban areas [3]. Major vital part of the intelligent transportation system is traffic signal lights control strategy becomes necessary [4]. There are so various parameters that have an effect on the traffic lights control. The static control method is not feasible for rapid and irregular traffic flow. The paper suggests a dynamic traffic control framework which is based on reinforcement learning [5]. The reinforcement learning can present a very crucial move to resolve the above cited problems. It is effectively deployed in resolving various problems [6]. The framework defines different traffic signal control types as action selections; the number of vehicles arriving and density of vehicle at a junction is observed as the context of environment and common signal control indicators, including delay time, the number of stopped vehicles and the total vehicle density are described as received rewards. The paper is divided as Sect. 2 gives the insights about the related work done in the area of traffic signal control. Section 3 describes Multi-agent Cooperative Q learning algorithm (MCQL). Section 4 explains about the system model. Experimental results are given in Sect. 5. The conclusion is presented in the Sect. 6.

2 Related Work

The traffic control systems can be categorized into offline traffic control systems and online traffic control systems. Offline methods make use of theoretical move toward optimizing the controls. Online methods regulate traffic regulator period dynamically as per instantaneous traffic conditions. Many achievements in collaborative traffic flow guidance and control strategy have been made. The approach of [7] the transportation industry. By means of F-B method approach, the traffic jam difficulty was partially resolved [7]. After that, several enhanced methods based on the F-B approach had developed [8]. Driving reimbursement coefficient and delay time was used to estimate the effectiveness of time distribution system given in [9]. The approach reduces the delay of waiting time, making the method appear to be sharp and sensible. There is a need to discover a new proper technique as this method could hardly solve the heavy traffic problem. Traffic congestion situation has been addressed using intelligent traffic control in [10] but congestion problems among neighboring junctions required better technique. The local synchronization demonstrated a fine result to this problem discussed in [11]. Because of complication and unpredictability, it is of limited opportunity to construct a precise mathematical model for traffic system in advance [12]. It has turned out to be a style to resolve traffic problems by taking benefit of computing expertise and machine learning [13]. Among many machine intelligence methods, reinforcement learning is feasible for the finest control of the transport system [14]. The study using the learning algorithm [15] achieved online traffic control. The approach was able to choose the optimal coordination model under different traffic conditions. Some applications [16] that utilize learning algorithm have received much significant effect. A paper implemented an online traffic control through learning algorithm, yielding good effort in the normal state of traffic congestion [17].

2.1 Traffic Estimation Parameters

Signal lights control has a very crucial responsibility in traffic management. Normally applied traffic estimation parameters [18] comprises of delay time, the number of automobiles stopped at a signal, and a number of newly arriving cars.

Delay Time. The delay between the real time and theoretically calculated time for a vehicle to leave a signal is defined as the delay time. In practice, we can get total delay time during a certain period of time and the average delay time of a cross to evaluate the time difference. The more delay time indicates the slower average speed of a vehicle to leave a signal.

Number of Vehicles Stopped. How many vehicles are waiting behind stop line to leave the road signal gives the number of vehicles stopped. The indicator [18] is used to measure the smooth degree of the road as well as the road traffic flow. It is defined as

$$\text{stop} = \text{stopG} + \text{stopR}, \tag{1}$$

where stopG is the number of automobiles stopped before the green light and stopR is the number of vehicles stopped at the red light.

Number of Vehicles Newly Arrived. The ratio of the actual traffic flow to the maximum available traffic flow gives the signal saturation. Newly arrived vehicle is calculated as

$$S = \frac{\mathit{traffic\,flow}}{(dr \times sf)}, \tag{2}$$

where dr is the ratio of red light duration to green light duration and sf is traffic flow of the signal.

Traffic Flow Capacity. The highest number of vehicles crossing through the signal is shown by traffic flow capacity. The result of the signal control strategy is given by the indicator. Traffic signal duration is associated with traffic flow capacity.

2.2 Reinforcement Learning

Reinforcement learning describes about maximize the numerical reward and mapping the state into actions through different way [18, 19]. Signal agents identify situation and responses from traffic scenarios, learn information depend on learning algorithms. Then it makes action choice with respect to its own accumulated information. Increase the traffic flow and decrease the average delay time is the purpose of traffic light control system. In this traffic arrangement, signals at one intersection coordinate with the signals at other intersection for better transport flow. Throughout the process, signals at each intersection develop a strategy of cooperation to maximize their individual benefit. Cooperation between agents is accomplished by distributing partial information of the states with the adjacent agents. In view of changing scenarios of the real traffic situation, multi-agent cooperative Q learning algorithm (MCQL) is developed for intelligent traffic control approach [19–21]. The Q update equation is given as:

$$Q(s, a) \leftarrow Q(s, a) + \alpha(r + \gamma Q(s', a') - Q(s, a)), \tag{3}$$

where the state, action, immediate reward, and cumulative reward at time t correspondingly stands for s_t, a_t, $r_{t,}$ and Q_t. $Q_t(s, a)$ is called policy function. $\alpha \in [0, 1]$ refers to the learning rate and $\gamma \in [0, 1]$ indicates the discount rate.

3 Multi-agent Cooperative Q Learning (MCQL)

Synchronization in multi-agent reinforcement generates a complex set of presentations achieved from the different agents' actions. A portion of good performing agent group (i.e., a general form) is shared among the different agents via a *specific form*(Q_i) [22]. Such specific forms embrace the limited details about the environment. Such strategies are incorporated to improve the sum of the partial rewards received using satisfactory cooperation prototype. The action plans or forms are created by the way of multi-agent Q learning algorithm by constructing the agents to travel for the most excellent form Q* and accumulating the rewards. When forms $Q_1, \ldots, Q_x$ are incorporated, it is possible to construct new forms that is *General Form* ($GF = \{GF_1, \ldots, GF_x\}$), in which GF_i denotes the **outstanding reinforcement** received by agent i all through the knowledge mode [5]. Algorithm 1 expresses *get_form* algorithm that splits the agents' knowledge. The forms are designed by the Q learning used for all prototypes. Outstanding reinforcements are liable for GF which compiles all outstanding rewards. It will be shared by the way of the added agents [21, 22]. Transforming incomplete rewards as *GF* is considered for outstanding reinforcements to achieve the cooperation between the agents. A *status* utility gives the outstanding form among the opening states and closing state for a known form which approximates *GF* with the outstanding reinforcements. The status utility is calculated by summation of steps the agent needed to get to the destination at the closing state and the sum of the received status in the forms among each opening and the closing state [22].

Algorithm 1 *Cooperative Multi-agent Q Learning Algorithm*

```
Algorithm get_form (I, technique)

1. Initialization $Q_i$(s, a) and $GF_i$(s, a)
2. Coordination of the agents $i \in I$;
3. Agents collaborate till the target state is found; episode ← episode +1
4. Renewal rule which estimates the reward value;

   $Q(s, a) \leftarrow Q(s, a) + \alpha(r + \gamma Q(s', a') - Q(s, a))$

5. Fcooperate (episode, tech, s, a, i);
6. $Q_i \leftarrow GF$ that is $Q_i$ of agent $i \in I$ is updated by means of $GF_i$.
```

Cooperation Models

Various collaboration methods for cooperative reinforcement learning are proposed here:

(i) *Group model*—reinforcements are distributed in a sequence of steps.

(ii) *Dynamic model*—reinforcements are distributed in each action.

(iii) *Goal-oriented model*—distributing the sum of reinforcements when the agent reaches the goal-state (S_{goal}).

Algorithm 2 *Cooperation model*

```
Fcooperate (episode, tech,s,a,i)/*cooperation between agents as four cases*/
q: count of sequence

1. Switch between cases
2. In case of Group method

   if episode mod q = 0 then
   get_Policy($Q_i$, Q*,$GF_i$);

3. In case of Dynamic method

   $r \leftarrow \sum_{j=1}^{x} Qj(s,a)$;
   $Q_i(s, a) \leftarrow r$;
   get_Policy($Q_i$, Q*, $GF_i$);

4. In case of Goal-oriented method

   if $S = S_{goal}$ then
   $r \leftarrow \sum_{j=1}^{x} Qj(s,a)$;
   $Q_i(s,a) \leftarrow r$;
   get_Policy($Q_i$,Q*,$GF_i$);
```

Algorithm 3 *get_Policy*

```
Function get_Policy($Q_i$, Q*, $GF_i$) /*find out universal agent policy */
1. for loop for each agent i ϵ I
2. for loop for each state s ϵ S
3. if value($Q_i$, s) $\leq$ value(Q*, s) then
   $GF_i$(s,a) $\leftarrow$ $Q_i$(s,a);
4. end for loop
```

Group Model: During the learning process each agents receives reinforcements for their actions. At the last part of the series (step q), each agent gives out the cost of Q_j to GF. If reward value is suitable, that is it improves the usefulness of another agents for given state the agents will afterward donate to these rewards [21–23].

4 Model Design

In a practical environment, traffic flows of four signals with eight flow directions are considered for the development. As shown in Fig. 1, there are altogether four junctions at each signal agent, i.e., agent 1, agent 2, agent 3, and agent 4 for Ja, Jb, Jc, and Jd, respectively.

The control coordination between the intersections can be viewed as a Markov process, denoted by $\langle S, R\rangle$, where represents the state of the intersection, stands for the action for traffic control, and indicates the return attained by the control agent.

Definition of State: Agent receives instantaneous traffic state and then returns traffic control decision by the present state of the road. Essential data such as a number of vehicles newly arriving and number of vehicles currently stopped at signal are used to reflect the state of road traffic.

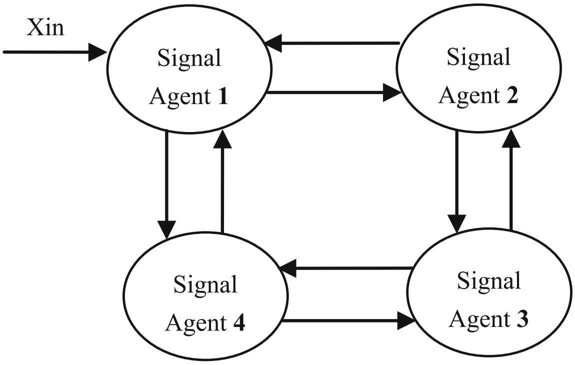

Fig. 1 Traffic flow and control of four intersections with eight flow directions

Number of vehicles newly arriving = X_{max} = x_1, x_2, x_3, x_4 = 10

Number of vehicles currently stopped at junction J = I_{max} = i_1, i_2, i_3, i_4 = 20.

State for agent 1 become (x_1, i_1), e.g., (5,0) that means 5 new vehicles are arriving to agent 1 with 0 vehicles are stopped at junction 1. State for agent 2 become (x_2, i_2), State for agent 3 become (x_3, i_3) and State for agent 4 become (x_4, i_4). State of the system become **Input** as (x_i, i_i). Here, it can get together 200 possible states by combining maximum 10 arriving vehicle and maximum 20 vehicles stopped at signal ($10 * 20 = 200$).

Definition of Action: In reinforcement learning framework, policy denotes the learning agent activities at a given time. Traffic lights control actions can be categorized to three types: no change in signal duration, increasing signal duration, reducing signal duration.

Value	Action
1	No change in signal duration
2	Increase in signal duration
3	Reduce the signal duration

Action set for signal agent 1 is A1 = {1, 2, 3}, action set for signal agent 2 is A2 = {1, 2, 3} and action set for signal agent 3 is A3 = {1, 2, 3}.

Each of them is for one of the following actual traffic scenarios. The strategy of no change in signal duration is used in the case of the normal traffic flow when the lights control rules do not change. The strategy increasing the signal duration is mostly used in the case that in one route traffic flow is stopped and the other route is regular. Increasing the signal duration extends the traffic flow while signal lights are still timing. The strategy decreasing signal duration is mostly used in the case that in one route of traffic flow is little while that of the other route is big. Decreasing signal light duration reduces the waiting time of the other route and lets vehicles of that route pass the junction faster, while signal lights keep timing [23, 24].

Definitions of Reward and Return: Agent makes signal control decisions under diverse traffic circumstances and returns an action sequence. We use traffic value display to estimate the traffic flows as [26].

Reward is calculated in the system as given below:

Assume current state i = (x_i, i_i) and next state j = (x_j, i_j).
current state i → next state j
Case 1: $[x_i, i_i] \rightarrow [x_i, i_{i-1}]$

$$[X_{max} = 10, I_{max} = 20] \rightarrow [X_{max} = 10, I_{max} = 19]$$

That means: one vehicle from currently stopped vehicle is passing the junction

Case 2: $[x_i, i_i] \rightarrow [x_{i+1}, i_{i-1}]$

$$[X_{max} = 9, I_{max} = 20] \rightarrow [X_{max} = 10, I_{max} = 19]$$

That means: one newly arrived vehicle at junction and one vehicle is passing

Case 3: $[x_i, i_i] \rightarrow [x_i, i_{i-3}]$

$$[X_{max} = 10, I_{max} = 20] \rightarrow [X_{max} = 10, I_{max} = 17]$$

That means: More than one stopped vehicles are passing the junction

Case 4: $[x_i, 0] \rightarrow [x_{i+1}, 0]$

$$[X_{max} = 2, I_{max} = 0] \rightarrow [X_{max} = 3, I_{max} = 0]$$

That means: new one new vehicle is arriving and no stopped vehicle at the junction. Depending on above state transitions from current state to next state, reward is calculated as [24]

$$\begin{aligned} \text{Reward is } r_p(i,\ p,\ j) &= 1 \quad \text{if } x_1' = x_1 + 1\ldots\ldots\ldots\text{Case}\,4 \\ &= 2 \quad \text{if } i_1' = i_1 - 1\ldots\ldots\ldots\text{Case}\,1 \\ &= 3 \quad \text{if } i_1' = i_1 - 3\ldots\ldots\ldots\ldots \mathit{Case}\,2\,\&\,3 \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

5 Experimental Results

The study learns a controller with learning rate = 0.5, discount rate = 0.9, and λ = 0.6. During the learning process, the cost was updated 1000 with 6000 episodes.

Figure 2 shows that delay time versus a number of state given by simple Q learning (without cooperation) and group method (with cooperation). Delay time obtained by cooperative methods, i.e., group method is much less than that of without cooperation method, i.e., simple Q learning for agent 1 in the multi-agent scenario.

Figure 3 shows that delay time versus a number of state given by simple Q learning (without cooperation) and group method (with cooperation). Delay time obtained by cooperative methods, i.e., group method is much less than that of

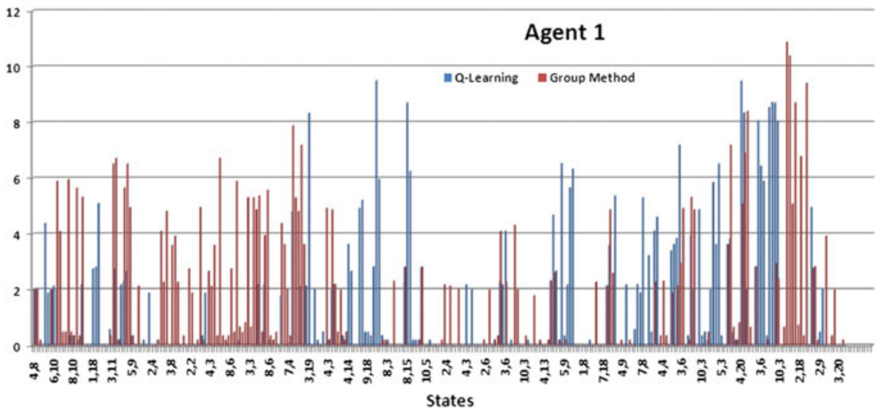

Fig. 2 States versus Delay time for Agent 1 by Q learning and Group Method

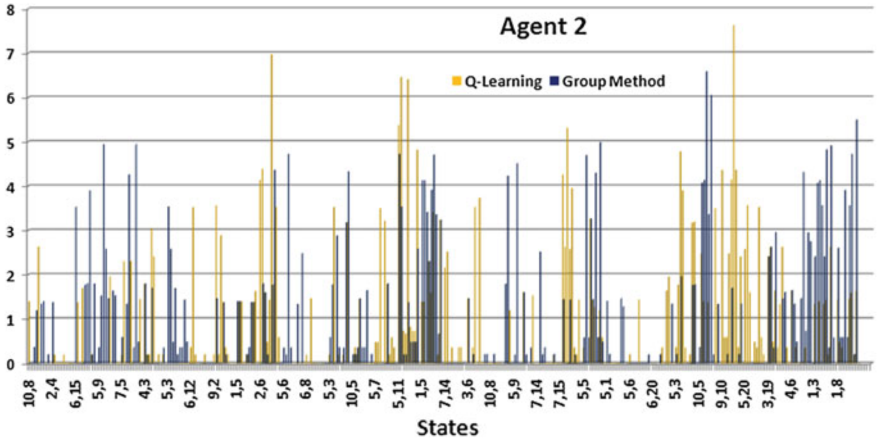

Fig. 3 States versus Delay Time for Agent 2 by Q learning and Group Method

without cooperation method, i.e., simple Q learning for agent 2 in the multi-agent scenario.

Figure 4 shows that delay time vs number of state given by simple Q learning (without cooperation) and group method (with cooperation). Delay time duration obtained by cooperative methods, i.e., group method is much less than that of without cooperation method, i.e., simple Q learning for agent 3 in multi-agent scenario.

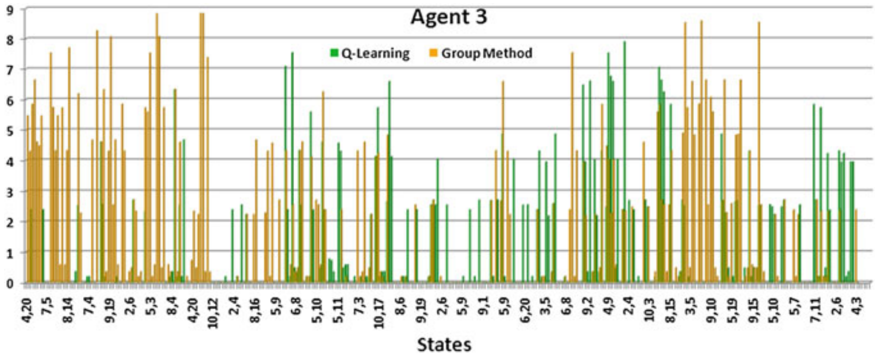

Fig. 4 States versus Delay Time for Agent 3 by Q learning and Group Method

6 Conclusion

Because traffic control system is so complicated and variable that a Q learning model (without cooperation) with defined strategy can rarely manage with the traffic jam and sudden traffic accidents which actually may occur at any time, the demand for combining timely and intelligent traffic control policy with real-time road traffic is getting more and more urgent. Reinforcement learning gathers tests and information by keeping communication with the situation. Although it usually needs a long duration to complete learning, it has good learning ability to a complex system, enabling it to handle unknown complex states well. The application of reinforcement learning in traffic management area is gradually receiving more and more concerns. The paper proposed a cooperative multi-agent reinforcement learning-based models (CMRLM) for traffic control optimization. The actual continuous traffic states are discretized for the purpose of simplification. We design actions for traffic control and define reward and return by mean of traffic cost which combines with multiple traffic capacity indicators.

References

1. F. Zhu, J. Ning, Y. Ren, and J. Peng, "Optimization of image processing in video-based traffic monitoring," *ElektronikairElektrotechnika*, vol.18, no.8, pp. 91–96, 2012.
2. B. de Schutter, "Optimal traffic light control for a single intersection," in *Proceedings of the American Control Conference (ACC '99)*, vol. 3, pp. 2195–2199, June 1999.
3. N. Findler and J. Stapp,"A distributed approach to optimized control of street traffic signals," *Journal of Transportation Engineering*, vol.118, no.1, pp. 99–110, 1992.
4. L. D. Baskar and H. Hellendoorn, "Traffic management for automated highway systems using model-based control,"*IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 2, pp. 838–847, 2012.
5. R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, Mass, USA, 1998.

6. Artificial Intelligence in Transportation *Information for Application*, Transportation Research CIRCULAR, Number E–C 113, Transportation On Research Board *of the National Academies*, January 2007.
7. Deepak A. Vidhate, Parag Kulkarni "New Approach for Advanced Cooperative Learning Algorithms using RL methods (ACLA)" VisionNet'16 Proceedings of the Third International Symposium on Computer Vision and the Internet, ACM DL pp 12–20, 2016.
8. K. Mase and H. Yamamoto, "Advanced traffic control methods for network management," *IEEE Magazine*, vol. 28, no. 10, pp. 82–88, 1990.
9. Deepak A. Vidhate, Parag Kulkarni "Innovative Approach Towards Cooperation Models for Multi-agent Reinforcement Learning (CMMARL)" in Smart Trends in Information Technology and Computer Communications, Springer Nature, Vol 628, pp 468–478, 2016.
10. L. D. Baskar, B. de Schutter, J. Hellendoorn, and Z. Papp, "Traffic control and intelligent vehicle highway systems: a survey," *IET Intelligent Transport Systems*, vol. 5, no. 1, pp. 38–52, 2011.
11. M. Broucke "A theory of traffic flow in automated highway systems," *Transportation Research C*, vol. 4, no. 4, pp. 181–210, 1996.
12. D. Helbing, A. Hennecke, V. Shvetsov, and M. Treiber, "Micro and macro-simulation of freeway traffic," *Mathematical and Computer Modelling*, vol. 35, no. 5–6, pp. 517–547, 2002.
13. S. Zegeye, B. de Schutter, J. Hellendoorn, E. A. Breunesse, and A. Hegyi, "A predictive traffic controller for sustainable mobility using parameterized control policies," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 3, pp. 1420–1429, 2012.
14. Deepak A. Vidhate, Parag Kulkarni "Enhancement in Decision Making with Improved Performance by Multiagent Learning Algorithms" IOSR Journal of Computer Engineering, Volume 1, Issue 18, pp 18–25, 2016.
15. A. Bonarini and M. Restelli, "Reinforcement distribution in fuzzy Q-learning," *Fuzzy Sets and Systems*, vol.160, no.10, pp. 1420–1443, 2009.
16. Y. K. Chin, Y. K. Wei, and K. T. K. Teo, "Qlearning traffic signal optimization within multiple intersections traffic network," in *Proceedings of the 6th UKSim/AMSS European Symposium on Computer Modeling and Simulation (EMS '12)*, pp. 343–348, Nov 2012.
17. L.A. Prashanth and S. Bhatnagar, "Reinforcement learning with function approximation for traffic signal control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 412–421, 2011.
18. Deepak A. Vidhate, Parag Kulkarni "Multilevel Relationship Algorithm for Association Rule Mining used for Cooperative Learning" in International Journal of Computer Applications (IJCA), Volume 86 Number 4- 2014 pp. 20–27.
19. Y. K. Chin, L. K. Lee, N. Bolong, S. S. Yang, and K. T. K. Teo, "Exploring Q-learning optimization in traffic signal timing plan management," in *Proceedings of the 3rd International Conference on Computational Intelligence, Communication Systems and Networks (CICSyN '11)*, pp. 269–274, July 2011.
20. Deepak A. Vidhate, Parag Kulkarni "Multi-agent Cooperation Methods by Reinforcement Learning (MCMRL)", *Elsevier International Conference on Advanced Material Technologies (ICAMT)-2016}No. SS-LTMLBDA*-06-05, 2016.
21. S. Russell and P. Norvi, *Artificial Intelligence: A Modern Approach*, PHI, 2009.
22. Deepak A. Vidhate, Parag Kulkarni "Performance enhancement of cooperative learning algorithms by improved decision making for context based application", International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT) IEEE Xplorer, pp 246–252, 2016.
23. Deepak A. Vidhate, Parag Kulkarni "Improvement In Association Rule Mining By Multilevel Relationship algorithm" in International Journal of Research in Advent Technology (IJRAT), Volume 2 Number 1- 2014 pp. 366–373.
24. Young-Cheol Choi, Student Member, Hyo-Sung Ahn "A Survey on Multi-Agent Reinforcement Learning: Coordination Problems", IEEE/ASME International Conference on Mechatronics Embedded Systems and Applications, pp. 81–86, 2010.