MARL FOR TRAFFIC LIGHT NETWORK

Тисленко Тимофей Иванович

ФГАОУ ВО «СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Scool of mathematics and computer science

Научный руководитель — к.ф.-м.н., доцент Д.В. Семенова

Томск, ITMM 2021

THE WORK RELEVANCE

There is a fast motorist comunityfast growing in Krasnoyarsk. According to ”AUTOSTAT” agency’s data our city is on 12 place by such growing. As the number of motorists increases, so does the time spent in traffic jams.

The goals and tasks

The work's goals

The development and study of a mathematical model of a multi-agent system for the problem of optimization of traffic at a crossroads.

Задачи

1. Review literature on relevant topics.
2. Construct the mathematical MARL model .
3. Develop an algorithm for solving the MARL problem.
4. Do a computational experiments.

Main defiunitions and deskriptions

defiunitions

1. **The intellectual agent** is called a meta-object with a partiality of subjectivity, which is interacting with other agents and the environment, performing certain functions to achieve it's objectives.
2. **The environment** is called the set of objects that are not belong to the agent.
3.
4. **The multiagent system** —- set of related agents.
5. **RL(Reinforcement Learning)** is a learning where environment is a teacher.

Problem statement

- The model is a Markov homogeneous chain with finite number of actions **A** and states **S** and discrete time t;
- environment — Section of the road network where the time for cars to pass through intersections is calculated according the rule: the time count begins at 100m before stop lines of traffic lights;
- set of agents K — all traffic lights on the road network;
- reward — total calculated time of vehicles passing through sections of the road at time t;
- $p_{ss'}$ — the probability that the s, when choosing a a solution, falls into the s' state is entirely determined by the state in which the process is going.
- the phases of the traffic lights change sequentially;

Problem statement

- state space of agent k S^k — is a set defined by a residue class modulo $n^k = |S^k|$: $S^k = n^k\mathbb{Z} = \left\{s^{(0)} = 0, s^{(1)} = 1, \ldots, s^{(n^k-1)} = n^k - 1\right\}$, where $s^{(i)}$ interpreted as «the phase i is active»;
- the set of solutions is defined by the main characteristic of the ring $n^k\mathbb{Z}$, as it is not bigger than. . . , $A = \left\{a^{(0)} = 0, a^{(1)} = 1, \ldots, a^{(n^k-1)} = n^k - 1\right\}$, where $a^{(j)}$ interpreted as «change phase j times».

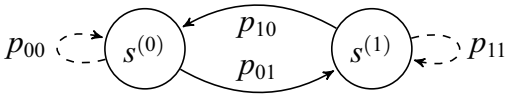
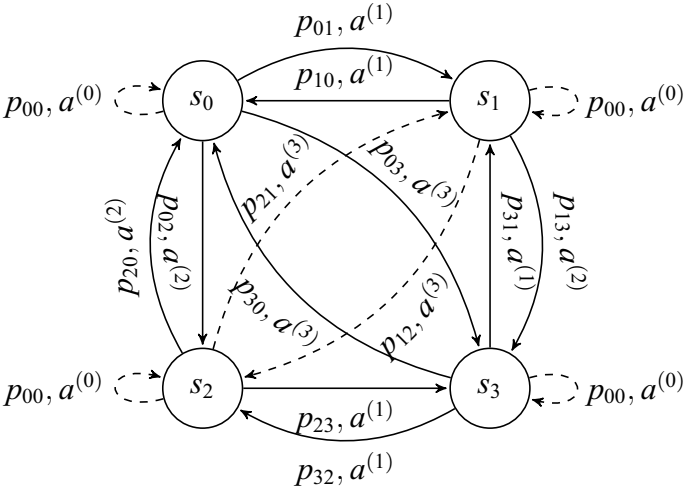


image: The stochastic graph of a controllable phase change process for a two-phase traffic light. Dashed to indicate action $a^{(0)}$, and continious — $a^{(1)}$.

Problem statement

- combined state of the environment $\mathbf{s}_t$ at time t is $\mathbf{s}_t = \{s_t^1, s_t^2, \ldots, s_t^K\} \in S^1 \times S^2 \times \ldots \times S^K$ — ;
- combined action of the environment $\mathbf{a}_t$ at time t is $\mathbf{a}_t = \{a_t^1, a_t^2, \ldots, a_t^K\} \in A^1 \times A^2 \times \ldots \times A^K$;
- the set $N(k)$ is set of k neibours — the agents with which it interacts k;
- the pair k and $j \in N(k)$ is a set $\mathbf{a}^{kj} \in A^k \times A^j$ and combined states $\mathbf{s}^{kj} \in S^k \times S^j$;
- reward $r(s_t, a_t; s_{t+1})$ at phase s_t we understand the total recorded time of all cars passing through a section of road.

Problem statement



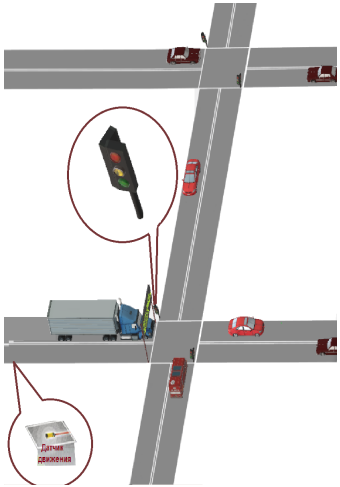


image: A stochastic network graph consisting of two two-phase traffic lights.

Defenitions and deskriptions

The Agent cumulative Earnings/Rewards function of the selected and fixed strategy δ will take the form

$$V\Big(\{S_t, \delta\}\Big) = \sum_{t=0}^{\infty} \gamma^t r(s_t, a_t), \tag{1}$$

where $0 \leq \gamma \leq 1$ — revaluation factor, $\{S_t\}$ — randomizing $\xi(t)$ with the chosen strategy $\delta = \{a_t, 0 \leq t < \infty\}$.

MARL for one traffic light

Setting the MARL problem for one traffic light

Need to find such a control δ, that provides max of the function

$$V\Big(\{S_t, \delta\}\Big),$$

where function

$$V^*\Big(\{S_t\}\Big) = \max_{\delta} Q(s, a), \tag{2}$$

and $Q(s, a)$ is

$$Q(s, a) = \sum_{s' \in S} p(s, a; s')\big(r(s, a; s') + \gamma \max_{a \in A} Q(s', a')\big). \tag{3}$$

Convergence criterion

The main idea of the Q-learning is valuation of the indeterminate right part of equation:

$$Q_{t+1}(s,a) = Q_t(s,a) + \alpha_t(s,a)\left(r(s,a) + \gamma \max_{a' \in A} Q_t(s',a') - Q_t(s,a)\right) \quad (4)$$

where s' — stage situation $t+1$, on step t process was at s and took action a, α — sales factor.
If on step t iteration of process is at state s and action a, is taken then $0 \leq \alpha_t(s,a) \leq 1$, else $\alpha_t(s,a) = 0$.

Convergence criterion

$Q = \{Q(s,a)\}_{s\in S, a'\in A}$, iteratively is $Q_{t+1} = A(Q_t)$,
where $A\colon \mathbb{R}^1_\infty \to \mathbb{R}^1_\infty$ — Contraction mapping.

$$\rho((A\circ Q_1)(s,a),(A\circ Q_2)(s,a)) \le \max|\gamma \max_{a'\in A} Q_1(s',a')) - \gamma \max_{a'\in A} Q_2(s',a'))| =$$
$$= \gamma\rho(Q_1(s,a),Q_2(s,a)), \gamma \in (0;1)$$

Convergence criterion for one traffic light MARL problem

If strategy δ is optimal, then using contraction mapping condition and

$$\sum_{t=0}^{\infty} \alpha_t(s,a) = \infty, \qquad \sum_{t=0}^{\infty} \alpha_t(s,a)^2 \le \infty \tag{5}$$

attracts convergence of Q-process(4)

Solution of MARL problem for one trafficlight

The MARL solution is

$$V^*(s) = \max_{a \in A} \lim_{t \to +\infty} Q_t(s, a). \tag{6}$$

$$a_t(s) = \arg\max_{a' \in A} Q_t(s, a') \tag{7}$$

Algorithm

```
//количество МАШИН на СТОПЛИНИИ 1,2
int T = (int)time();                                                    // T - time
int[] n = {timecollection1.size(), timecollection2.size()};            // n - number

//суммарное ВРЕМЯ ЗАДЕРЖКИ машин на СТОПЛИНИИ 1,2
double[] t = {T*n[0], T*n[1]};
for (Map.Entry<Agent, Double> entry : timecollection1.entrySet()) t[0] -= entry.getValue();
for (Map.Entry<Agent, Double> entry : timecollection2.entrySet()) t[1] -= entry.getValue();

//НАГРАДЫ ([S_t]->[S_t+1])
reward[0][0] = t[0] - t[1];                                             // r(s,a) = reward[s][a]
reward[0][1] = (n[1]-n[0])*period;
reward[1][0] = (n[0]-n[1])*period;
reward[1][1] = t[1] - t[0];

//Qfactor[номер пред. состояния][номер посл. состояния] (во всех случаях их 2)
int[] maxQnext = {max(Qfactor[0][0],Qfactor[0][1]), max(Qfactor[1][0],Qfactor[1][1])}  ;
for (int a=0;a<=1;a++){
    Qfactor[0][a] = (int)((1-alfa)*Qfactor[0][a] + alfa*(reward[1][a]+ gama*maxQnext[a]));
    Qfactor[1][a] = (int)((1-alfa)*Qfactor[1][a] + alfa*(reward[1][a]+ gama*maxQnext[a]));
}

int p0 = trafficLight.getCurrentPhaseIndex();                           // p0 - phase 0 (текущее состояние)

//СОХРАНЯЕМ ФАЗУ СВЕТОФОРА
if (Qfactor[p0][p0]>Qfactor[p0][1-p0] ) {
    trafficLight.switchToNextPhase();
    trafficLight.switchToNextPhase();
}
```

image: 5. The «Qfactor» code

Computational experiments

goals

Compare the delay time of cars in the model of traffic light control system whose phase duration is obtained by overdrive, and managed by the Markov process, in the Anylogic system.

inputs

- Cars arrive at the intersection from each of three directions at a rate of 1,000 per hour;
- Discount and reassessment coefficients are empirically selected.

outputs

- Acceleration 1.5 times that of the traffic light control system, the phases of which are selected by a step of 1 from 5 seconds to 30 seconds;

Results

The purpose of the work was to introduce the approach to optimize the selection of a traffic light, taking into account the current load of transport, from the standpoint of minimizing delay, as well as creating an algorithm that implements this approach and calculating delays in its navigation. The results are as follows:

1. A mathematical model of the process of selecting the phase of a traffic light is constructed, which differs by taking into account the current position of traffic lights and their loading and allows to formulate optimization tasks, the purpose of which is to minimize the delay of traffic of cars.
2. A multi-agent system structure has been developed, which includes a single agent - a traffic light, which ensures the most efficient parallelization of the entire task on sub-tasks that will be solved by the agent.

References

- El-Tantawy S., Abdulhai B. and Abdelgawad H., Multiagent Reinforcement Learning for Integrated Network of Adaptive Traffic Signal Controllers (MARLIN-ATSC) // IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 3. 2013. -P 1140-1150.
- Лекции по случайным процессам: учебное пособие / А. В. Гасников, Э. А. Горбунов, С. А. Гуз и др.; под ред. А. В. Гасникова. – «Москва»: МФТИ, 2019. – 285 с.
- Марковские процессы принятия решений. / Майн Х., Осаки С. Главная редакция физико-математической литературы издательства «Наука», 1977. – 176 с.
- Sandholm T.W. Contract Types for Satisficing Task Allocation: I Theoretical Results // AAAI Spring Symposium Series: Satisficing Models. 1998. – P. 68-75.

THANKS FOR ATTENTION!!!