ЗАДАЧА MARL ДЛЯ СЕТИ СВЕТОФОРОВ

Тисленко Тимофей Иванович

ФГАОУ ВО «СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт математики и фундаментальной информатики

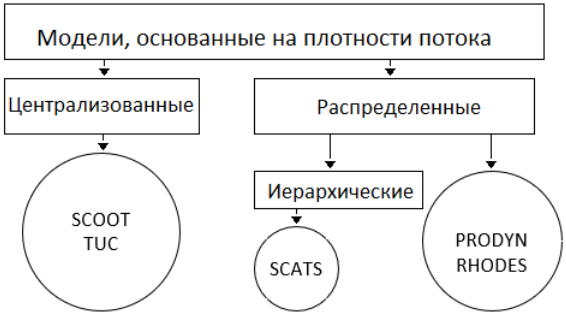
Научный руководитель — к.ф.-м.н., доцент Д.В. Семенова

Томск, ITMM 2021

Актуальность

В Красноярске стремительно с каждым годом растет количество желающих стать автомобилистом. По данным агентства «Автостат» на 1 января 2020 Красноярский край оказался на 12 месте в топ-20 регионов по объему автомобильного парка в России. С количеством автомобилистов увеличивается и время, которое проводится в пробках.

Существующие модели адаптивных систем светофоров



Модели, основанные на марковских процессах принятия решений

Рисунок: 1. Модели адаптивных систем светофоров

Цели и задачи

Цель работы

Разработка и исследование математической модели мультиагентной системы для задачи оптимизации движения на перекрестке.

Задачи

1. Сделать обзор литературы по соответствующей тематике.
2. Построить математическую модель MARL.
3. Разработать алгоритм решения задачи MARL.
4. Провести вычислительные эксперименты.

Основные определения

Определения

1. **Интеллектуальным агентом** называется метаобъект, наделенный долей субъектности, взаимодействующий с другими агентами и средой, выполняющий определенные функции для достижения поставленных целей.
2. **Средой** называется множество объектов, не принадлежащих агенту.
3. **Задачами/ресурсами** называются объекты, распределяемые агентами в ходе достижения целей.
4. **Мультиагентная система** – совокупность взаимосвязанных агентов.
5. **RL(Reinforcement Learning)** — обучение с подкреплением, где в роли учителя выступает среда.

Постановка задачи

- Модель среды – марковская однородная цепь с конечным числом действий **A** и состояний **S** и дискретным временем t;
- среда — участок дорожной сети, где рассчитывается время проезда машин через перекрестки; отсчет времени начинается за 100м до стоп-линий светофоров;
- множество агентов K — все светофоры находящиеся в на участке дорожной сети;
- задержка — суммарное засеченое время машин, проходящих через отрезки дороги, в момент времени t;
- $p_{ss'}$ — вероятность того, что система из состояния s при выборе решения a попадает в состояние s', полностью определяется состоянием, в которое переходит процесс.
- фазы светофора меняются последовательно;

Постановка задачи

- пространство состояний агента k S^k — фазы светофора, определяются классом вычетов по модулю $n^k = |S^k|$: $S^k = n^k\mathbb{Z} = \left\{s^{(0)} = 0, s^{(1)} = 1, \ldots, s^{(n^k-1)} = n^k - 1\right\}$, где $s^{(i)}$ интерпретируется как «активна фаза i»;
- Множеством решений определяется характеристикой кольца $n^k\mathbb{Z}$, т.е. не превыщает ее, $A = \left\{a^{(0)} = 0, a^{(1)} = 1, \ldots, a^{(n^k-1)} = n^k - 1\right\}$, где $a^{(j)}$ интерпретируется как «сменить фазу j раз».

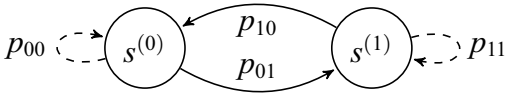
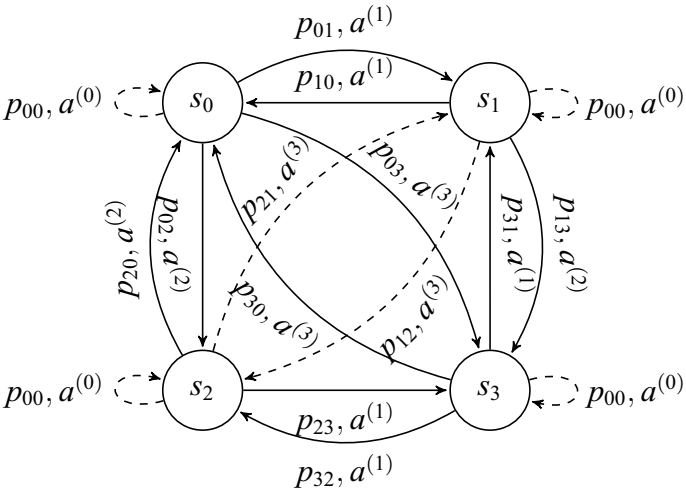


Рисунок: Стохастический граф управляемого процесса смены фаз для двухфазного светофора. Пунктиром обозначено действие $a^{(0)}$, а сплошной — $a^{(1)}$.

Постановка задачи

- совокупное состояние среды $\mathbf{s}_t$ в момент времени t $\mathbf{s}_t = \{s_t^1, s_t^2, \ldots, s_t^K\} \in S^1 \times S^2 \times \ldots \times S^K$;
- совокупное управление $\mathbf{a}_t$ в момент времени t $\mathbf{a}_t = \{a_t^1, a_t^2, \ldots, a_t^K\} \in A^1 \times A^2 \times \ldots \times A^K$;
- множество соседних агентов $N(k)$ для светофора k — агенты, с которыми взаимодействует k;
- для пары агентов k и $j \in N(k)$ множество их совместных действий $\mathbf{a}^{kj} \in A^k \times A^j$ и совместных состояний $\mathbf{s}^{kj} \in S^k \times S^j$;
- под задержкой $r(s_t, a_t; s_{t+1})$ на фазе s_t будем понимать суммарное засечённое время всех машин, проходящих через отрезок дороги.

Постановка задачи



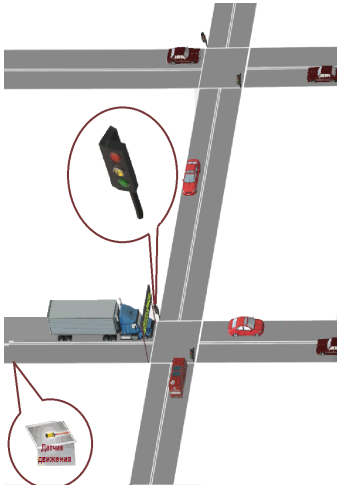


Рисунок: Стохастический граф сети из двух двухфазных светофоров.

Определения и обозначения

Функция суммарных доходов/вознаграждений агента при выбранной стратегии δ примет вид

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

где $0 \leq \gamma \leq 1$ — коэффициент переоценки, $\{S_t\}$ — реализация случайного процесса $\xi(t)$ при выбранной стратегии $\delta = \{a_t, 0 \leq t < \infty\}$.

Задача MARL для одного светофора

Постановка задачи MARL для одного светофора

Требуется найти такое управление δ, которое доставляет максимум функции

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

где функция на каждом шаге t может быть определена для $s = S_t$

$$V^*(s) = \max_a Q(s, a), \tag{2}$$

а $Q(s, a)$, есть функция

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

Критерий сходимости

Идея Q-обучения заключается в оценке невычислимой правой части:

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

где s' — положение процесса на шаге $t+1$, если на шаге t процесс был в состоянии s и было выбрано действие a, α — коэффициент скидки.
Если на шаге t процесс находился в состоянии s и было выбрано действие a, то $0 \leq \alpha_t(s,a) \leq 1$, иначе $\alpha_t(s,a) = 0$.

Критерий сходимости

$Q = \{Q(s,a)\}_{s\in S, a'\in A}$, можно записать итеративно $Q_{t+1} = A(Q_t)$,
где $A\colon \mathbb{R}^1_\infty \to \mathbb{R}^1_\infty$ — сжимающее отображение.

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

Критерий сходимости задачи

Если стратегия $a(\cdot)$ приводит к тому, что с вероятностью 1 каждая пара (s,a) бесконечное число раз встречается, то из условия сжимаемости при

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

следует сходимость процесса (4)

Критерий сходимости

Решение задачи MARL для одного светофора имеет вид:

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

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

Задача MARL для сети светофоров

Функция суммарных доходов/вознаграждений для K агентов примет вид

$$V\Big(\{S_t, \delta\}\Big) = \sum_{t=0}^{\infty} \gamma^t r(\mathbf{s}_t, \mathbf{a}_t), \tag{8}$$

где $0 \leq \gamma \leq 1$ — коэффициент переоценки, $\{S_t\}$ — реализация случайного процесса $\xi(t)$ при выбранной стратегии $\delta = \{\mathbf{a}_t, 0 \leq t < \infty\}$.

Задача MARL для сети светофоров

Постановка задачи MARL для сети светофоров

Требуется найти такое управление δ, которое доставляет максимум функции

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

где функция на каждом шаге t может быть определена для $\mathbf{s} = S_t$

$$V^*(\mathbf{s}) = \max_{\mathbf{a}} Q(\mathbf{s}, \mathbf{a}), \tag{9}$$

а $Q(\mathbf{s}, \mathbf{a})$, есть функция

$$Q(\mathbf{s}, \mathbf{a}) = \sum_{\mathbf{s}' \in S} p(\mathbf{s}, \mathbf{a}; \mathbf{s}')\big(r(\mathbf{s}, \mathbf{a}; \mathbf{s}') + \gamma \max_{\mathbf{a} \in A} Q(\mathbf{s}', \mathbf{a}')\big). \tag{10}$$

Решение задачи MARL для двух светофоров

Рассмотрим агента k для него определено множество соседних агентов $N(k)$ Для системы из двух светофоров $N(k) = \{j\}$ Искать решение в таком виде неудобно, поэтому перепишем функцию $Q(\mathbf{s}, \mathbf{a})$ как функцию обучения одного агента k, т.е. $Q^k(\mathbf{s}, a)$:

$$Q^k(\mathbf{s}, a_k) = \sum_{a_j \in A^j} p(\mathbf{s}, a_j) p(\mathbf{s}, a_k) Q(\mathbf{s}, \mathbf{a}) \tag{11}$$

$$Q^{kj}(\mathbf{s}, \mathbf{a}) = p(\mathbf{s}, a^j) p(\mathbf{s}, a^k) Q(\mathbf{s}, \mathbf{a}^{kj}) \tag{12}$$

Находим максимизирующее решение для Q-процесса одного агента k на шаге t

$$a_t^k(\mathbf{s}) = \arg\max_{a_k \in A^k} \sum_{a_j \in A^j} Q_t^{kj}(\mathbf{s}, \mathbf{a}^{kj}) \tag{13}$$

Решение задачи MARL для сети светофоров

Имея алгоритм решения задачи для 2ух светофоров, можно получить решение для любого их количества в сети.
Для этого используем алгоритм MARLIN [**?**].
Оптимальное управление для фиксированного агента k будем искать как решение задачи MARL в виде:
$a_k = \arg\max_{a_k \in A^k} \sum_{j \in N(k)} \sum_{a_j \in A^j} Q_t^{kj}(\mathbf{s}, \mathbf{a}^{kj}) p(\mathbf{s}' | (\mathbf{s}, \mathbf{a}^{kj}))$.

Вычислительные эксперименты

Цель

Сравнить время задержки машин в модели системы управления светофоров, длительность фаз которой получена перебором, и управляемой марковским процессом, в системе Anylogic.

Входные данные

- машины прибывают на перекресток с каждого из трех направлений с интенсивностью 1000 в час;
- коэфициенты скидки и переоценки подобраны эмпирическим путём.

Результаты и выводы

- ускорение в 1.5 раза по сравнению с системой управления светофором, длительность фаз которой подобрана перебором от 5 секунд до 30 секунд с шагом 1;

Основные результаты работы

Целью работы являлось ознакомление с подходом, позволяющим оптимизировать процесс выбора сигнала светофора, с учетом текущей загрузки транспорта, с точки зрения минимизации задержки, а также создание алгоритма, реализующего данный подход и вычисление задержек с его пмощью. В работе получены следующие результаты:

1. Построена математическая модель процесса выбора фазы светофора, отличающаяся учетом текущего расположения светофоров и их загрузки и позволяющая сформулировать оптимизационные задачи, целью которых является минимизация задержки трафика автомобилей.
2. Разработана структура мультиагентной системы, включающая в себя единственного агента – светофор, обеспечивающая наиболее эффективное распараллеливание всей задачи на подзадачи, которые будут решены агентом.

Основная литература

- 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.

СПАСИБО ЗА ВНИМАНИЕ!!!