ЗАДАЧА MARL ДЛЯ СВЕТОФОРА НА ПЕРЕКРЁСТКЕ

Тисленко Тимофей Иванович

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

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

Томск, МПОИТЭС 2021

Актуальность

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

Существующие модели адаптивыных систем светофоров

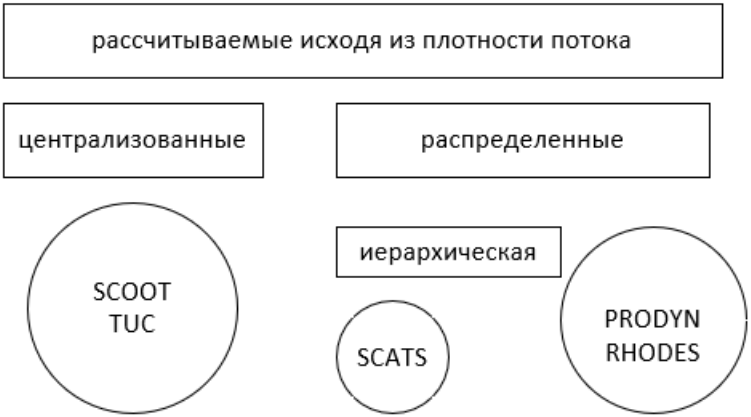

Рисунок:

Цели и задачи

Цель работы

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

Задачи

1. Сделать обзор литературы по соответствующей тематике.
2. Описать математическую модель.
3. Описать алгоритм.
4. Продемонстрировать результаты работы алгоритма.

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

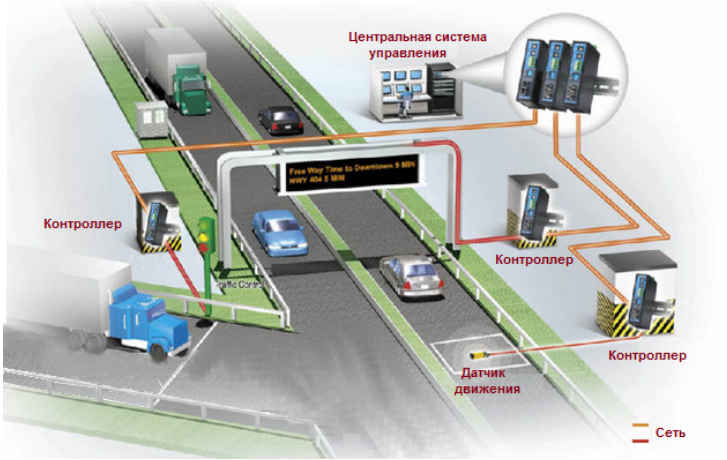

Рисунок:

- Модель среды – Марковский процесс принятия решения-
- агент — светофор
- среда — перекресток, на котором на отрезках дорог за 100м до стоплиний засекается время.
- пространство состояний $S = \{ s_0 = \text{«фаза0»} , s_1 = \text{«фаза1»} \}$
- в момент времени t_k активна фаза светофора S_k, суммарное засеченое всех машин, проходящих через отрезок дороги называется задержкой на фазе S_k
- множество действий A = $\{ a_0 = \text{«оставить фазу»} , a_1 = \text{«сменить фазу»} \}$

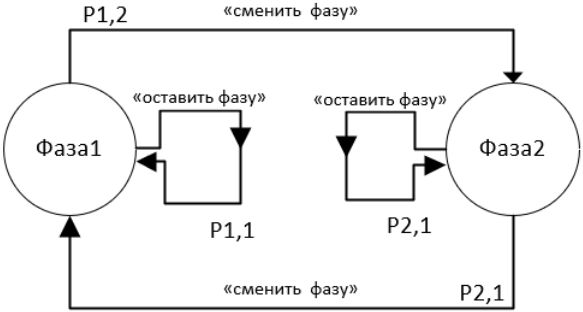

Рисунок:

- $r(s,a)$ = задержка на фазе s после действия a
- $p(i,k;j)$ вероятность того, что система из состояния i при выборе решения k попадает в состояние j, полностью определяется состоянием, в которое переходит процесс.

$V^*(s)$ — функция суммарных внешних доходов от оптимальной политики в состоянии s

$$V^*(s) = \max_{a(\cdot)} \sum_{t=0}^{\infty} \gamma^t r(s_t, a_t). \quad (1)$$

Уравнение Вальда - Беллмана [**?**] для управляемого марковского процесса с конечным числом действий и состояний имеет вид:

$$V^*(s) = \max_{a \in A} \left\{ \sum_{s' \in S} p(s, a; s')(r(s, a; s') + \gamma V^*(s')) \right\}. \quad (2)$$

Т.е.

$$V^*(S_0) = max\{p(S_0, A_0; S_1)(r(S_0, A_0) + \gamma V^*(S_1) + p(S_0, A_0; S_1)(r(S_0, A_0) + \gamma V^*(S_0); \\ p(S_0, A_1; S_1)(r(S_0, A_1) + \gamma V^*(S_1) + p(S_0, A_1; S_1)(r(S_0, A_1) + \gamma V^*(S_0)\}$$

$$V^*(s) = \max_{a \in A} Q(s,a) = \max_{a \in A} Q_t(s,a). \quad (3)$$

$$Q(s,a) = \sum_{s' \in S} p(s,a;s')(r(s,a;s')) + \gamma V^*(s'). \quad (4)$$

Идея Q-обучения заключается в оценке невычислимой правой части:

$$Q_{t+1}(s,a) = Q_t(s,a) + \alpha_t(s,a)\left(r(s,a) + \gamma \max_{a' \in A} Q_t(s',a') - Q_t(s,a)\right) \quad (5)$$

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

Сходимость

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

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

Оказывается, что если используемая стратегия $a(\cdot)$ приводит к тому, что с вероятностью 1 каждая пара (s, a) будет бесконечное число раз встречаться на бесконечном горизонте наблюдения, то из отмеченного выше условия сжимаемости при

$$\sum_{t=0}^{\infty} \alpha_t(s,a) = \infty, \sum_{t=0}^{\infty} \alpha_t(s,a)^2 \leq \infty \tag{6}$$

удет следовать сходимость (с вероятностью 1) процесса 5

$$V^*(s) = \max_{a \in A} \lim_{t \to +\infty} Q_t(s,a). \tag{7}$$

$$a_t(s) = \arg\max_{a' \in A} Q_t(s,a') \tag{8}$$

Основные результаты работы

Целью работы являлось ознакомление с подходами, позволяющими оптимизировать процесс выбора сигнала светофора, с учетом текущей загрузки транспорта, с точки зрения минимизации задержки. В работе получены следующие результаты:

1. 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. Структура мультиагентной системы, включающая в себя единственного агента – светофор, обеспечивающая наиболее эффективное распараллеливание всей задачи на подзадачи, которые будут решены агентом.

- 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. ISBN 978-5-7417-0710-4
- Марковские процессы принятия решений. / Майн Х., Осаки С. Главная редакция физико-математической литературы издательства «Наука», 1977. - 176 с. УДК 519.283

СПАСИБО ЗА ВНИМАНИЕ!!!