

# ПРОЕКТИРОВАНИЕ СИСТЕМЫ ГРУППОВОГО УПРАВЛЕНИЯ НА ПРИМЕРЕ ЛИКВИДАЦИИ ПОЖАРОВ НА ОТКРЫТОЙ МЕСТНОСТИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

К.В. Тюндеров, К.Ю. Усенко, А.Ю. Зарницын  
Томский политехнический университет  
E-mail: kv10@tpu.ru

## Введение

Роботы используются во многих областях науки, техники и промышленности, в первую очередь там, где жизнедеятельность человека либо затруднена, либо вообще невозможна, например, в зонах радиоактивного или химического загрязнения, в условиях боевых действий, при тушении пожаров большой площади и т.п. Одиночный робот может использоваться только для решения некоторых частных задач, либо выполнения довольно простых операций, поскольку он, как правило, обладает сравнительно малыми возможностями для выполнения поставленной задачи (небольшой радиус действия, ограниченный бортовым энергоресурсом, невысокая вероятность выполнения поставленной задачи при функционировании в экстремальных ситуациях, поскольку выход из строя одиночного робота ведет к невыполнимости его миссии и т.п.). Очевидным решением указанных выше проблем является применение сразу нескольких роботов, то есть групп роботов.

Поэтому целью данной работы является создание системы группового управления мобильными роботами способных функционировать и решать задачи в экстремальной среде.

## Описание алгоритма

Поставленная в работе цель может быть декомпозирована на следующие задачи:

- Организовать группу мобильных гомогенных роботов;
- Управляя группой способствовать локализации и дальнейшей ликвидации пожара

Для решения поставленных задач необходимо разработать архитектуру агента, среду для его обучения и определить требуемое поведение агентов при тушении пожара. Предобучение на реальных системах нежелательно из-за длительности обучения, сложности вычислений и излишней стохастичности, предобучение будет проводиться в разработанной эмуляции среды (пожара).

Обучение с подкреплением строится на моделях агента и среды. Сам метод заключается во взаимодействии агента и среды, где агент за совершенные действия получает от среды ее состояние и награду ( $R$ ). Задача агента максимизировать награду.

$$\operatorname{argmax}(R); \quad (1)$$

При разработке агента сначала необходимо выделить параметры среды на основе которых будет обучаться агент, множество действий агента, а также спроектировать структуру поощрений наказаний агента.

Для реализации метода были выделены следующие параметры состояния среды:

- площадь пожара;
- координаты пожара;
- координаты агента и остальных роботов;
- очки существования огня;
- заряд аккумуляторов;
- количество оставшейся тушащей смеси.

Множество действий агента дискретно и состоит из движения вперед-назад на  $\frac{2}{3}$  длины корпуса, поворот влево-вправо дискретизирован по 10 градусам, тушение пожара и бездействие.

Диаграмма сущностей системы управления роботом представлена на рисунке 1.

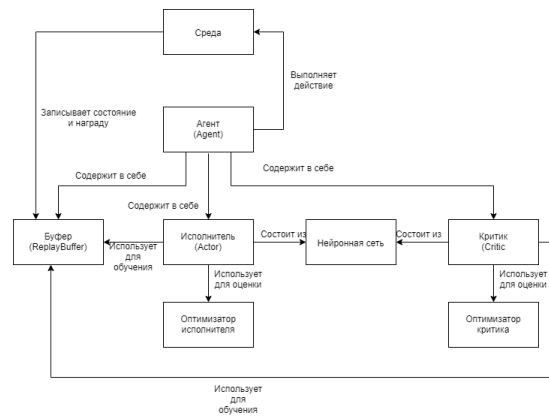


Рис. 1. Диаграмма сущностей системы управления

Агент содержит в себе Исполнителя и Критика, состоящих из нейронных сетей, а также Буфера, куда записываются все состояния среды и награды на протяжении всего обучения за совершенные агентом действия. Исполнитель и Критик используют Буфер и данные записанные в нем для обучения нейронной сети, а оптимизатор для оценки нейронных сетей.

## Моделирование среды

Исходя из поставленной задачи, среда должна состоять из 2 основных частей: пожаров и роботов.

Пожар моделируется областью, ограниченной точками горения (далее очагами). Для того, чтобы пожар имел случайную форму, задаётся начальная координата пожара, куда устанавливается первый очаг, и указывается направление. Очаги генерируются попарно таким образом (3 - 4), чтобы каждая пара с предыдущей образовывала трапецию (это необходимо для упрощения расчетов), определяется точка с определенным шагом распространения очагов в заданном направлении, затем, откладываются случайные величины по направлению оси Y и против. По выставлении заданного количества точек, устанавливается заключительная по ранее заданному направлению.

$$x_{i,i+1} = x_0 + step \cdot i, (3)$$

$$y_{i,i+1} = y_0 + tg(\alpha) \cdot step \pm rand, (4)$$

$$i = 2n - 1,$$

где  $x_0, y_0$  - начальные координаты пожара,  $step$  - шаг распространения очагов,  $\alpha$  - направление генерации пожара,  $rand$  - случайная величина по распределению Гаусса (для  $y_i$  и  $y_{i+1}$  значения  $rand$  разные),  $n$  - натуральное число.

Каждый очаг имеет очки существования, которые уменьшаются при тушении. При истечении очков существования, очаг считается тлеющим. Это значит, что в этой точке огонь больше распространяться не будет, но температура еще высокая и область остается опасной для робота.

В целях упрощения, робот моделируется прямоугольником с габаритами, равными размерам роботов Lego на базе, которого будет проводиться разработка прототипа системы и ограниченным набором действий. В модели робот не имеет массы (инерционности) и перемещается в голономном режиме (5). Он определяется 6 величинами - 4 вершины прямоугольника, точка его центра и направление движения. Робот имеет ресурсы заряда и жидкости для тушения.

$$(x_i \ y_i) = (x_i - x_c \ y_i - y_c) \times$$

$$\times \begin{pmatrix} \cos(\Delta\alpha) & -\sin(\Delta\alpha) \\ \sin(\Delta\alpha) & \cos(\Delta\alpha) \end{pmatrix} + (x_c \ y_c); (6)$$

где  $x_i, y_i$  - координаты вершин корпуса робота,  $x_c, y_c$  - координаты центра робота,  $\Delta\alpha$  - угол изменения направления движения.

В среде регистрируются столкновения роботов, попадания в пожар и попадание очага в область тушения. Это необходимо для определения подкрепления за совершенные действия.

Тушение моделируется возникающей из центра робота треугольной областью, и чем дальше от робота попадает очаг в область, тем слабее он тушится. Усилие рассчитывается по формуле (6):

$$effort = 100 - \frac{(dist - dist_{uncov})}{(dist_{cov} - dist_{uncov})}, (7)$$

где  $dist$  - расстояние до очага от центра робота,  $dist_{uncov}$  - расстояние, где тушение невозможно (от центра робота до переднего его края),  $dist_{cov}$  - максимальное расстояние тушения пожара.

Для наглядности процесса обучения, в среде присутствует опция рендера состояния среды. Пример состояния изображен на рисунке 2.

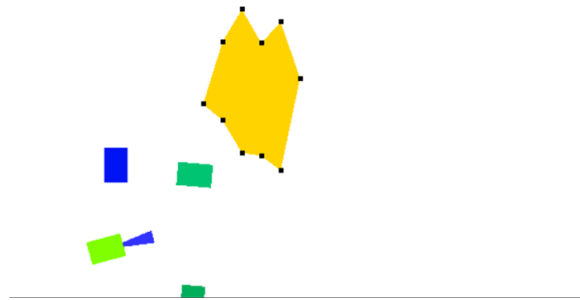


Рис. 2. Рендер среды

### Заключение

В результате была разработана мультиагентная система управления группой роботов с применением одного из подходов искусственного интеллекта - обучение с подкреплением. Разработана система наград агента. Выделены ключевые параметры состояния среды. Разработана эмуляция среды. Спроектирована архитектура агента, выделены основные абстракции системы. Была произведена интеграция алгоритма управления в агентов среды.

В дальнейшем планируется реализовать параллельные вычисления и обучить агентов по алгоритму АЗС. И провести испытание на учебно-исследовательском полигоне с применением данного алгоритма на имеющейся базе мобильных роботов университета.

### Список использованных источников

1. Asynchronous Methods for Deep Reinforcement Learning // arXiv.org URL: <https://arxiv.org/pdf/1602.01783.pdf> (дата обращения: 10.01.2020).
2. Soft Actor-Critic Algorithms and Applications // arXiv.org URL: <https://arxiv.org/pdf/1812.05905.pdf> (дата обращения: 10.01.2020);
3. Wyatt McAllister, Denis Osipychyev, Girish Chowdhary, Adam Davis Multi-Agent Planning for Coordinated Robotic Weed Killing // RSJ International Conference on Intelligent Robots and Systems (IROS). - New York, NY, USA: IEEE, 2018. - С. 7955-7960.