

УДК 004.5

MSC 68T40

DOI: 10.35330/1991-6639-2020-6-98-14-19

ПРИМЕНЕНИЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ ДЛЯ РЕШЕНИЯ ЗАДАЧИ СТРУКТУРИЗАЦИИ ВНЕШНЕЙ СРЕДЫ*

М.И. АНЧЁКОВ

Институт информатики и проблем регионального управления –
филиал ФГБНУ «Федеральный научный центр
«Кабардино-Балкарский научный центр Российской академии наук»
360000, КБР, г. Нальчик, ул. И. Арманд, 37-а
E-mail: iipru@rambler.ru

В работе рассматривается задача структуризации внешней среды. Предложенный подход основывается на методе обучения с подкреплением, обучающим мультиагентную нейронную сеть. Особенностью подхода является то, что структуризация среды производится коллективом роботов, которые могут взаимодействовать друг с другом посредством сообщений. Предложена формализованная постановка задачи обучения.

Ключевые слова: мультиагентная нейронная сеть, мультиагентная система, робототехнические системы, коллективное поведение.

ВВЕДЕНИЕ

В современных реалиях перехода от одиночной робототехники к коллективной актуализируются задачи взаимодействия роботов во внешней среде. Роботы должны функционировать в слабоструктурированных средах без трудоемких операций адаптации их к новым условиям. На наш взгляд, роботы должны научиться не только распознавать объекты среды, но и строить взаимосвязи между ними. Данное исследование основывается на гипотезе о том, что роботы, находящиеся в частично наблюдаемой среде и решающие задачу максимизации внутренней энергии, проведут структуризацию среды посредством коммуницирования друг с другом.

ОБЗОР СУЩЕСТВУЮЩИХ МЕТОДОВ

В [1, 2] решается задача получения семантических знаний о неизвестной или частично неизвестной среде, в которой находится робот.

В [1] авторам удалось собрать информацию о более чем 38000 комнатах, которые расположены в 197 зданиях на 940 этажах. Это данные по двум институтам: Королевскому технологическому институту (Стокгольм, Швеция) и Массачусетскому технологическому институту (Бостон, США). Авторы использовали алгоритм gSpan (graph-based substructure pattern mining), который позволяет находить наиболее часто встречающиеся подграфы. Далее, на основе вероятностных методов делалось предсказание о том, какие именно комнаты могут находиться в неизвестной области. На основе полученных результатов авторы делают выводы о том, что локальные структуры часто повторяются, и этот факт можно использовать для построения гипотез о ранее не исследованных помещениях.

В [2] авторы разработали метод, который позволяет генерировать гипотезы о неизвестных помещениях. Гипотезы представляют собой семантический граф, в котором вершины обозначают комнаты, метки вершин означают название комнат, а ребра графа обозначают

* Работа выполнена при финансовой поддержке грантов РФФИ №№ 18-01-00658, 19-01-00648

дверные проемы. Для поставленной задачи использовался метод машинного обучения graph kernel method. Отличительной особенностью данного метода является то, что он может работать непосредственно с графовыми структурами без необходимости их преобразования в вектор. Авторы отмечают, что хоть предложенный метод не дает высоких результатов предсказаний, их можно использовать как предварительные данные для последующей работы робота.

В [3] авторы предлагают решать задачу построения семантической карты на основе объединения двух иерархических представлений: пространственного в виде *AN-graph* и семантического. Например, стол может быть представлен в виде графа с одной вершиной «Столешница» и четырьмя вершинами «Ножка», которые соединены ребрами с вершиной «Столешница». Семантическая информация была представлена с помощью системы представления знаний NeoClassic [4]. Такое объединение пространственных и семантических данных позволяет роботу, с одной стороны, эффективно решать задачи ориентации и навигации, а с другой – позволяет использовать семантические данные для систем принятия решений. Стоит отметить, что предложенный подход достаточно универсальный, однако это связано с достаточно большой работой, которая необходима для получения семантической информации.

ОПИСАНИЕ КОМПЬЮТЕРНОЙ МОДЕЛИ

С целью проведения компьютерного эксперимента была разработана программная среда RoboSim (рис. 1).

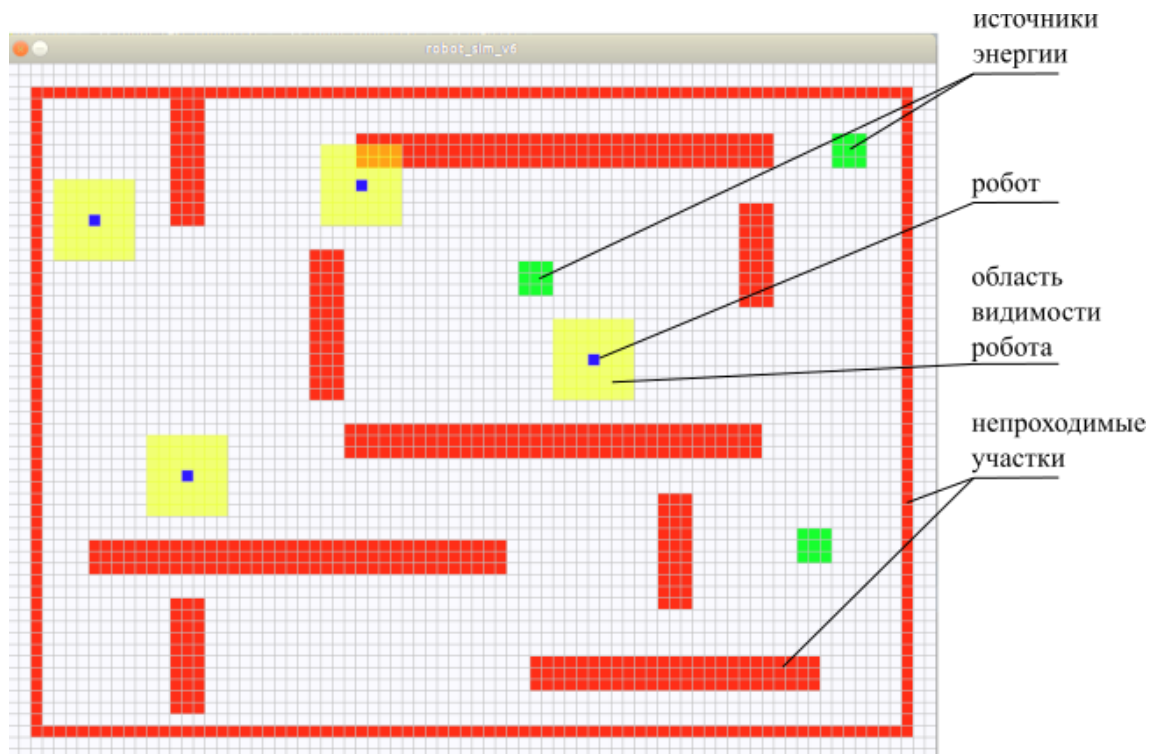


Рис. 1. Главное окно программы RoboSim

Программа позволяет менять конфигурацию непроходимых участков, положение и характеристики роботов, положение источников энергии.

В формализованном виде среду можно описать следующим образом:

$$S = \langle R, E^{ex}, Pos, Im \rangle,$$

где:

R – множество роботов,
 E^{ex} – множество источников энергии,
 Im – множество непроходимых участков,
 Pos – множество проходимых участков.

Каждый робот $r_i \in R (i = \overline{1, n})$, где n – общее количество роботов.

$$r_i = \langle E^{in}, P, M, Ef, S \rangle,$$

где:

E^{in} – внутренняя энергия робота,
 P – положение в пространстве,
 M – мультиагентная нейронная сеть [5],
 Ef – эффекторы,
 S – сенсоры.

На содержательном уровне формулировка задачи структуризации может быть представлена следующим образом: в процессе поиска источников энергии роботы выполняют функцию дать название всем объектам внешней среды. Внешняя среда представляет собой условный лабиринт, ограничивающий область видимости робота. Это приводит к тому, что роботы должны научиться коммуницировать между собой. Например, робот, обнаруживший источник энергии, может разослать информацию о местонахождении источника всем остальным роботам, которые пока не обнаружили этот источник.

ОПИСАНИЕ МЕТОДА ОБУЧЕНИЯ

В качестве системы принятия решения роботом используется мультиагентная нейронная сеть (МНС) [5].

Мультиагентная нейронная сеть представлена кортежем:

$$M = \langle X, Y, A_{in}, A_{mediator}, A_{out}, D \rangle,$$

где:

X – вектор входных сигналов,
 Y – вектор выходных сигналов,
 A_{in} – множество входных агентов-нейронов,
 $A_{mediator}$ – множество внутренних агентов-нейронов,
 A_{out} – множество выходных агентов-нейронов,
 D – множество дуг, определяющих способ соединения агентов-нейронов.
 Агент-нейрон представляет собой конечный автомат:

$$A = \langle X, Y, S, f_y, f_s, S_0, PP, Pos, D, E \rangle,$$

где:

X – вектор входных сигналов,
 Y – вектор выходных сигналов,
 S – множество состояний,
 f_y – функция выходов,
 f_s – функция переходов между состояниями,
 S_0 – начальное состояние,
 PP – программа позиционирования,
 Pos – координаты агента-нейрона,
 D – программа роста дендритов,
 E – энергия агентов.

Более подробное описание МНС отмечается в [5].

Информацию о внешней среде МНС получает от лидара и от других роботов. Данные от лидара приходят в виде массива, в котором содержатся $\langle \alpha, \delta \rangle$, где: α – угол смещения луча лидара, δ – расстояние до препятствия.

Информация от других роботов приходит в виде сообщения Mes:

$$\text{Mes} = \langle N^r, P^{\text{cur}}, N^T \rangle,$$

где:

N^r – имя робота, который отправил сообщение,

P^{cur} – положение объекта, о котором сообщает робот,

N^T – имя, которое дает отправитель обнаруженному объекту.

Для обучения МНС будет использоваться обучение с подкреплением [6]. Обобщенная схема метода представлена на рисунке 2. В качестве подкрепления используется энергия, полученная роботом из внешней среды.



Рис. 2. Обучение с подкреплением

Выбор данного метода обучения связан с тем, что для обучения не требуется обучающая выборка, а всю необходимую информацию робот получает из окружающей среды. Под обучением с подкреплением понимается не конкретный алгоритм, предложенный в [6], а класс методов обучения, которые в процессе решения задачи взаимодействуют со средой и получают от нее вознаграждение.

Робот, взаимодействуя со средой, каждый такт времени отправляет другим роботам информацию о том, какие объекты он обнаружил. Каждому обнаруженному объекту присваивается имя $n^T \in N^T$.

В таком случае задача структуризации может быть сформулирована следующим образом: необходимо найти такую функцию f , которая

$$f: N^T \rightarrow T, \text{ где } T = (E^{\text{ex}} \cup \text{Im} \cup \text{Pos}), \text{ причем } E^{\text{ex}} \cap \text{Im} \cap \text{Pos} = \emptyset.$$

В общем случае МНС старается максимизировать энергию, получаемую из внешней среды:

$$\sum_{i=1}^n E_i^{\text{in}} \rightarrow \max.$$

Так как робот находится в частично наблюдаемой среде, он не может сразу построить решение задачи, которое максимизирует $\sum_{i=1}^n E_i^{\text{in}}$.

С учетом этого задачу обучения можно сформулировать следующим образом:

$$\sum_{t=0}^k \sum_{i=1}^n \gamma^t E_{t,i}^{in} \rightarrow \max,$$

где:

- t – такты времени,
- k – максимальное время прогноза,
- n – количество роботов,
- γ^t – коэффициент, который указывает на то, что прогноз на более длительное количество тактов приносит меньшее количество энергии.

ЗАКЛЮЧЕНИЕ

В работе предложена математическая постановка задачи структуризации на основе обучения с подкреплением.

ЛИТЕРАТУРА

1. Aydemir A., Jensfelt P., Folkesson J. What can we learn from 38,000 rooms? Reasoning about unexplored space in indoor environments // IEEE/RSJ International Conference on Intelligent Robots and Systems. 2012. Pp. 4675-4682.
2. Luperto M., Amigoni F. Predicting the global structure of indoor environments: A constructive machine learning approach // Autonomous Robots. 2018. V. 43. № 4. Pp. 813-835.
3. Galindo C., Saffiotti A., Coradeschi S., Buschka P., Fernandez-Madrigal J. A., Gonzalez J. Multi-hierarchical semantic maps for mobile robotics // IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Alta., 2005. Pp. 2278-2283.
4. Patel-Schneider P.F., Resnick A.L., McGuinness D.L., Weixelbaum E., Abrahams M., Borgida A. NeoClassic Reference Manual: Version 1.0. AT&T Labs Research, Artificial Intelligence Principles Research Department, 1996.
5. Анчёков М.И., Кривенко М.П. Компьютерная модель возникновения коллективного поведения роботов // Известия КБНЦ РАН. 2019. № 6 (92). С. 21-26
6. Саттон Р.С., Барто Э.Г. Обучение с подкреплением. М.: Бином. Лаборатория знаний, 2017. 399 с.

REFERENCES

1. Aydemir A., Jensfelt P., Folkesson J. What can we learn from 38,000 rooms? Reasoning about unexplored space in indoor environments // IEEE/RSJ International Conference on Intelligent Robots and Systems. 2012. Pp. 4675-4682.
2. Luperto M., Amigoni F. Predicting the global structure of indoor environments: A constructive machine learning approach // Autonomous Robots. 2018. V. 43, № 4. Pp. 813-835.
3. Galindo C., Saffiotti A., Coradeschi S., Buschka P., Fernandez-Madrigal J. A., Gonzalez J. Multi-hierarchical semantic maps for mobile robotics // IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Alta., 2005. Pp. 2278-2283.
4. Patel-Schneider P.F., Resnick A.L., McGuinness D.L., Weixelbaum E., Abrahams M., Borgida A. NeoClassic Reference Manual: Version 1.0. AT&T Labs Research, Artificial Intelligence Principles Research Department, 1996.
5. Anchokov M.I., Krivenko M.P. *Komp'yuternaya model' vznikhoveniya kollektivnogo povedeniya robotov* [Computer model of the emergence of collective behavior of robots] // Izvestiya KBSC RAN/News of the KBSC of RAS. 2019. No. 6 (92). Pp. 21-26.
6. Sutton R.S., Barto E.G. *Obuchenie s podkrepleniem* [Reinforcement learning]. М.: Binom. Knowledge Laboratory, 2017. 399 p.

APPLYING REINFORCED LEARNING FOR SOLVING THE PROBLEM OF STRUCTURING THE EXTERNAL ENVIRONMENT*

M.I. ANCHEKOV

Institute of Computer Science and Problems of Regional Management –
Branch of Federal public budgetary scientific establishment «Federal scientific center
«Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences»
360000, KBR, Nalchik, 37-a, I. Armand St.
E-mail: iipru@rambler.ru

The paper is concerned with the problem of structuring the external environment. The proposed approach is based on a reinforcement learning method that trains a multi-agent neural network. A feature of the approach is that the structuring of environment is carried out by a team of robots that can interact with each other through messages. The formalized formulation of training problem is proposed.

Keywords: multiagent neural network, multiagent system, robotic systems, collective behavior.

Работа поступила 10.12.2020 г.

Сведения об авторе:

Анчеков Мурат Инусович, н.с. отдела «Мультиагентные системы» Института информатики и проблем регионального управления – филиала Кабардино-Балкарского научного центра РАН.
360000, КБР, г. Нальчик, ул. И. Арманд 37-а.
E-mail: murat.antchok@gmail.com

Information about the author:

Anchekov Murat Inusovich, researcher of the Department of the multiagent systems of the Institute of Computer Science and Problems of Regional Management of KBSC of the Russian Academy of Sciences.
360000, KBR, Nalchik, I. Armand street, 37-a
E-mail: murat.antchok@gmail.com

* The work was carried out with the financial support of the RFBR grants No. No. 18-01-00658, 19-01-00648