

Т.И. ШАРИПОВА, В.Г. РЕДЬКО, Г.А. БЕСХЛЕБНОВА

Научно-исследовательский институт системных исследований РАН,
Москва
gab19@list.ru

МОДЕЛИРОВАНИЕ АВТОНОМНОГО АГЕНТА, ВЕДУЩЕГО ПОИСК В ДВУМЕРНОМ ЛАБИРИНТЕ*

Методом растущего во времени нейронного газа исследуется поисковое поведение автономного агента в двумерном лабиринте. Показана возможность резкого сокращения размеров нейронной сети, в которой достаточно запоминать только те точки пространства, в которых сильно меняется окружающая ситуация.

Ключевые слова: нейронный газ, агенты, поисковое поведение.

1. Введение

В работе [1] была построена модель автономного агента, ведущего поиск в одномерном лабиринте. Использовался метод растущего во времени нейронного газа [2]. Построенная модель продемонстрировала самостоятельное нахождение агентом источника пищи. В настоящей работе моделируется аналогичное поисковое поведение агента в двумерном лабиринте. Также используется метод нейронного газа.

Суть этого метода состоит в следующем.

Имеется сеть связанных между собой нейронов, узлов сети. Каждый нейрон-узел запоминает определенную сенсорную информацию S_i (вектор памяти нейрона), где i – номер нейрона. Нейроны имеют веса, модифицируемые путем обучения. Нейронная сеть управляет поведением агента. Текущая сенсорная информация на входе системы управления агента $S(t)$ сравнивается со всеми векторами памяти нейронов S_i , и ищется k -й нейрон, для которого расстояние между $S(t)$ и S_k минимально (t – дискретное время). Если для этого нейрона расстояние между $S(t)$ и S_k меньше определенного порога, то вектор S_k немного приближается к вектору $S(t)$ и k -й нейрон становится активным. Если это расстояние больше порога, то создается новый нейрон-узел, который в своей памяти запоминает текущую сенсорную информацию $S(t)$.

Связи между нейронами сети формируются при переходе меж-

*Работа частично поддержана РФФИ, проект 16-01-00223.

ду нейронами. Например, когда создается новый нейрон, то формируется связь между нейроном, который был активен в предыдущий такт времени, и новым нейроном.

Если сенсорная информация определяет положение агента в пространстве, то несложно организовать перемещение агента в соответствии с запомненной сенсорной информацией. В этом случае предварительно происходит блуждание агента по пространству. При этом веса нейронов формируются методом обучения с подкреплением [3]. Чем более благоприятно место организма в пространстве, тем больше вес нейрона, вектор S_i которого определяет данную точку в пространстве. После предварительного блуждания по пространству формируется нейронная сеть со связями между нейронами. Движение агента может управляться такой нейронной сетью следующим образом. Каждый такт времени агент находится в точке пространства, соответствующей вектору S_i активного нейрона. Далее анализируются все связи этого нейрона, и выбирается нейрон, который связан с данным активным нейроном и имеет наибольший вес среди таких нейронов. Затем происходит переход от данного нейрона к выбранному нейрону с большим весом. При этом агент перемещается в пространстве в точку с координатами, соответствующими новому выбранному нейрону. Новый нейрон становится активным.

В настоящей модели исследуется поисковое поведение агента в двумерном лабиринте. Особое внимание уделяется процессам формирования растущего нейронного газа в процессе поиска.

2. Описание модели

Описание автономного агента. Автономный агент представляет собой материальную точку с двумерными координатами. Агент может находиться только в пустых областях лабиринта. Агент имеет сенсоры, с помощью которых может измерять расстояние до стенок, находящихся справа, слева, спереди и сзади. Хотя автономный агент может перемещаться в любом направлении, он не теряет своих представлений о нахождении «сторон света».

У автономного агента есть понятие о запасе полезного ресурса R .

Двумерный случай. Перечислим следующие свойства агента, находящегося в двумерном пространстве.

1. Рассматривается агент, который может двигаться в двумерном пространстве x, y .

2. Имеется лабиринт с источником питания. Используемый при моделировании лабиринт показан на рис. 1. Лабиринт представляет собой прямоугольную область, состоящую из квадратов размером 10×10 единиц. Каждый квадрат может быть либо пустым (где может находиться агент), либо полным (представляет из себя стенку, где агент находиться не может). То есть, лабиринт состоит из нескольких прямоугольных участков, «комнат». Более того, рассматривается частный случай лабиринтов, где внешняя граница окружена стеной (заполненными квадратами), а пустые области представляют собой коридоры шириной не более 1 квадрата. Часть коридора (один или несколько рядом стоящих пустых квадратов) также может представлять собой область полезного ресурса, в котором автономный агент может увеличивать свои ресурсы. Данная область представляет собой отдельный коридор. Агент движется внутри лабиринта. Задача агента – исследование лабиринта и поиск источника пищи.

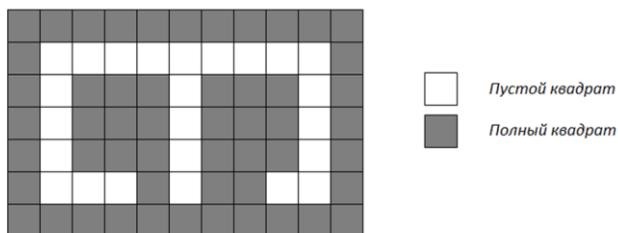


Рис. 1. Схема коридора

3. Источник пищи имеет зону действия – подобласть двумерного лабиринта.

4. Агент имеет ресурс R , который увеличивается при нахождении источника пищи.

5. Агент функционирует в дискретном времени t . Каждый такт времени агент совершает движение, при этом его координаты x , y изменяются на некоторые величины $\Delta x(t)$, $\Delta y(t)$ соответственно.

6. Когда координата агента совпадает с зоной действия источника пищи, ресурс агента за один такт времени увеличивается на величину Δr .

7. Агент имеет свою систему управления, на сенсорный вход которой поступает координаты агента $x(t)$, $y(t)$.

8. Система управления агента задается растущей нейронной сетью. На вход активного нейрона подается текущие координаты агента $x(t)$, $y(t)$.

9. Каждый нейрон имеет память, он запоминает определенные

координаты x_i , y_i в данном случае вектор памяти S_i имеет шесть компонент: четыре расстояния до стенок лабиринта, находящиеся спереди, сзади, слева и справа от агента, а также координаты агента x_i , y_i . Кроме того, каждый нейрон имеет свой вес W_i . Чем больше вес нейрона, тем более предпочтительно то место, которое характеризует нейрон.

10. Имеется два режима динамики агента: 1) режим случайного движения и 2) режим детерминированного перемещения, перемещения в соответствии весами узлов-нейронов нейронного газа. Характерное время перехода из случайного режима в детерминированный равно T_N .

11. Каждый такт времени выбирается первый или второй режим. Причем вероятность выбора первого режима, т.е. режима случайного поиска в начале функционирования агента близка к 1, а дальнейшим эта вероятностью постепенно уменьшается и происходит переход к детерминированному движению в соответствии с весами нейронов. Таким образом, реализуется метод отжига [4]: при малых временах t агент движется случайно, при больших временах – детерминировано, в соответствии с весами нейронов.

12. В режиме случайного поиска после перемещения агента его координаты становятся равными $x(t)$, $y(t)$. Определяются параметры комнаты (длина, ширина), в которой находится агент. Если параметры комнаты изменились, то формируется новый нейрон, в памяти которого записываются расстояния от агента до стенок лабиринта и координаты агента.

13. При появлении нового нейрона формируется связь от предыдущего активного нейрона к новому.

14. В режиме детерминированного перемещения определяются веса всех «контактных» нейронов, которые связаны с текущим активным, и среди этих контактных нейронов находится предпочтительный, имеющий наибольший вес. Этот нейрон становится активным в следующий такт времени. Координата агента становится равной координате, хранящейся в памяти предпочтительного нейрона.

Описание когнитивной карты. В данной работе под когнитивной картой имеется в виду нейронная сеть, отображающая исходный лабиринт. Каждый нейрон имеет свои двумерные координаты. Связь между нейронами подразумевает наличие коридора между этими точками. Когнитивная карта создается автономным агентом по мере изучения лабиринта. Веса нейронов меняются только после нахождения полезного ресурса. Если не найдена об-

ласть полезного ресурса, то веса всех нейронов равны 0. Если область полезного ресурса найдена, то нейроны, находящиеся в данной области, имеют вес, равный единице. Веса остальных нейронов рассчитываются по формуле

$$W_i = 0,5 \max (W_j),$$

где W_j – веса соседних (контактных) нейронов. Под соседями нейрона подразумевается нейроны, находящиеся через одно ребро от рассматриваемого.

Если нейрон находится в зоне действия полезного ресурса, то его вес равняется 1. Веса соседей (контактных нейронов) такого нейрона, если они нулевые, приравниваются половине веса ближайшего ненулевого нейрона. Если вес нейрона ненулевой, то сравнивается, какое из значений больше: текущее или половина от ненулевого соседа (если соседей несколько, то выбирается сосед с наибольшим значением).

Нейроны появляются только в трёх случаях: 1) там, где появился автономный агент (точка старта); 2) в «тупиковых» случаях: подробнее см. ниже; 3) там, где глобально менялись параметры комнаты (у автономного агента есть четыре сенсора, по которым он измеряет длину и ширину комнаты. Если длина или ширина изменились более чем на заданное количество (в компьютерной программе это количество равно 5) единиц, то считается, что автономный агент попал в перекресток или новую комнату.

При построении нейронной сети (когнитивной карты) лабиринта нейроны сети создаются только в особых точках лабиринта. Первая такая точка – это место появления автономного агента. Кроме того, нейроны создаются на перекрестках (по одной точке на один перекресток) и в коридорах (одна точка на один коридор), а также в тупиках (концах коридора). Предложенный принцип построения когнитивной карты позволит существенно уменьшить размер нейронной сети.

Целью автономного агента является составление когнитивной карты существующего лабиринта и нахождение областей полезного ресурса.

Описание схемы передвижения автономного агента. При исследовательском поведении агента сначала рассматриваются возможные варианты выбора общего направления автономного агента. Рассматриваются четыре стороны (вверх, вниз, слева, справа) на возможность прохождения на 10 и более шагов. Если такая возможность есть, то данное направление считается возможным для дальнейшего исследования. Далее в выбранном направлении с от-

клонением в $\pm 90^\circ$ автономный агент пробует произвести шаг. Если произвести шаг нельзя (агент попадает в полный квадрат лабиринта), то выбирается другое отклонение и делается попытка шага в новом направлении. Если агент не может сделать шаг в течение нескольких итераций I_T то считается, что агент зашел в тупик, а направление исследовано.

За каждую единицу времени автономный агент может совершить одно движение или перемещение. Движение представляет собой переход автономного агента на 1 единицу в выбранном направлении. Перемещение представляет собой переход адаптивного агента на 10 единиц в выбранный коридор или перемещение между созданными и запомненными точками-нейронами. Если при выполнении предполагаемого действия агент выходит за пределы лабиринта (натывается на стенку), то он это действие не совершает.

Описание поискового поведения автономного агента. В первоначальный момент времени автономный агент появляется в пустой области лабиринта. С помощью сенсоров он определяет параметры (длину - a и ширину - b) коридора, в котором находится, а также определяет направления, которые он может исследовать. Случайным образом выбирается одно из возможных направлений для исследований. Сначала делается перемещение на 10 единиц строго в выбранном направлении (чтобы, если автономный агент находится на перекрестке, он случайно не ушёл в другой коридор). После этого в выбранном направлении со случайным отклонением автономный агент делает движение. Если в выбранном направлении движение невозможно, то выбирается другое случайное отклонение. Так происходит до тех пор, пока параметры комнаты глобально не изменятся (т.е. длина или ширина, или оба параметра). Тогда в данной точке создаётся новый нейрон. Для него считается исследованным направление, из которого он пришел, и проверяется возможность исследования в других направлениях. При этом в когнитивной карте создается связь между созданным нейроном и тем, из которого пришел автономный агент. Если I_T раз подряд не получается сделать движение в выбранном направлении, то автономный агент считает, что попал в тупик. Тогда в данной точке создаётся нейрон, в котором три направления невозможны для исследования, а оставшееся направления помечается как исследованное. После этого автономный агент возвращается в нейрон, из которого начал движение в данном направлении, помечает коридор, из которого вернулся, как исследованный (создает связь между нейронами) и выбирает другое направление для исследований. Если в данном

месте все направления исследованы, то автономный агент возвращается в тот нейрон, из которого пришел к текущему нейрону, пока не найдет нейрон с неисследованными направлениями. Если весь лабиринт исследован, то дальнейшее поведение автономного агента определяется тем, какой мотив доминирует в автономном агенте.

Параметры модели.

Начальное значение ресурса агента $R=10$. На каждое движение затрачивается часть ресурса $\Delta r=0,01$. Если автономный агент находится в области полезного ресурса, то его запасы увеличиваются на $\Delta r=1$.

Характерное время перехода с режима случайного поиска равно $T_N=1000$ тактов времени.

Количество итерация, после повторения которых автономный агент решает, что попал в тупик $I_T=50$.

Результаты моделирования динамики агента для двумерного случая представлены на рис. 2-5.

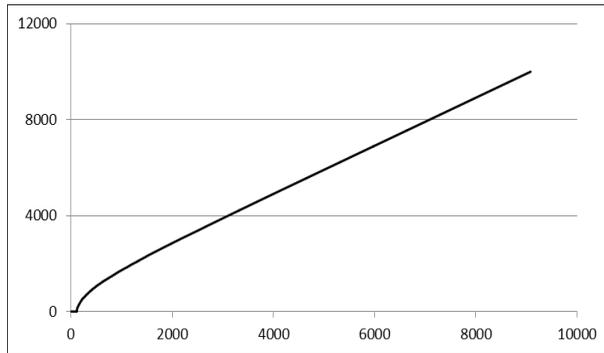


Рис. 2. Зависимость уровня ресурса агента R от времени t

Видно, что, при выбранных параметрах большую часть времени агент находится в области полезного ресурса. Однако, при небольших значениях времен происходит уменьшение полезного ресурса, что соответствует растрате ресурса при совершении движений (рис. 3).

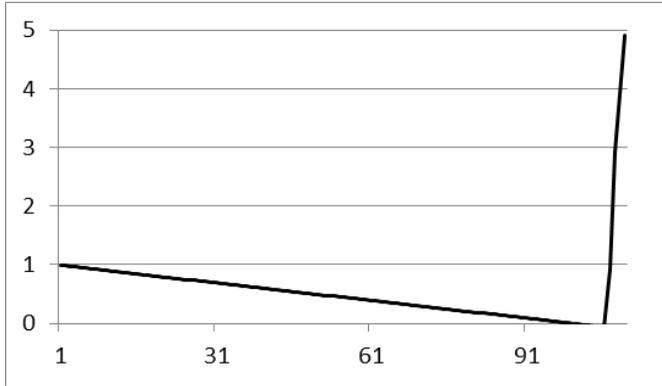


Рис. 3. Зависимость уровня ресурса агента R от времени t (начальный участок графика)

Зависимость числа нейронов от времени представлена на рис. 4. Показано, что, начиная с некоторого отсчета времени размер нейронной сети не изменяется. Это связано с принципами построения нейронной сети (когнитивной карты) лабиринта – нейроны создаются лишь в некоторых подпространствах лабиринта, характеризующих текущую ситуацию нахождения агента в лабиринте.

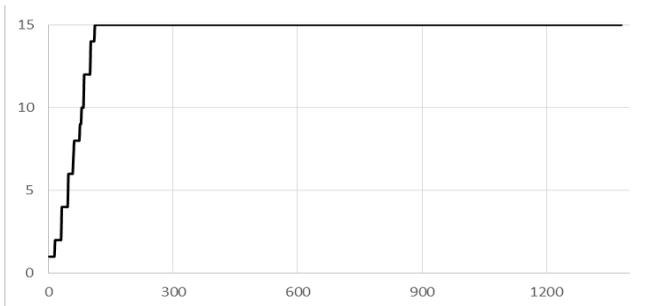


Рис. 4. Зависимость числа нейронов N от времени t

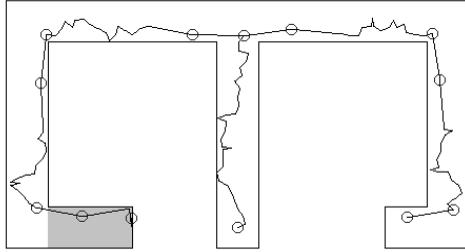


Рис. 5. Движение агента по двумерному лабиринту. Траектория движения агента показана изломанной линией, точки, характеризующие сильное изменение окружающей обстановки, отмечены кружками, источник пищи показан серым фоном

На рис. 5 показана траектория движения агента по двумерному лабиринту. Кружками отмечены создаваемые узлы нейронной сети. Узлы образуются при сильном изменении окружающей среды. Зона полезного ресурса (источник пищи) отмечен на рисунке серым фоном.

Анализ полученных результатов показал, что агент успешно анализирует лабиринт и находит источник пищи, после этого ресурс агента растет. Кроме того, показана возможность резкого сокращения размеров нейронной сети, в которой достаточно запоминать только те точки пространства, в которых сильно меняется окружающая ситуация (в данном случае это соответствует сильному изменению размеров комнаты).

3. Заключение

Таким образом, построена модель поискового поведения агента, система управления которого формируется на основе метода растущего нейронного газа. Разработан метод обучения с подкреплением для растущей нейронной сети; проанализирован этот метод для двумерного лабиринта. Построен вариант модели растущего нейронного газа, в котором радикально сокращается число узлов-нейронов за счет того, что в нейронах запоминаются не все точки, в которых побывал агент, а только те, в которых радикально меняется окружающая среда.

Список литературы

1. Редько В.Г., Шарипова Т.И. Использование метода нейронного газа

для моделирования поискового поведения агентов. // Труды НИИСИ РАН. 2013. Т. 3. №2. С.22-25.

2. Butz M.V., Shirinov E., Reif K. Self-organizing sensorimotor maps plus internal motivations yield animal-like behavior // Adaptive Behavior. 2010. V. 18. No. 3-4. P. 315-337.
3. Саттон Р.С., Барто Э.Г. Обучение с подкреплением. М.: Бином, 2011.
4. Kirkpatrick S., Gelatt C.D., Vecchi M.P. Optimization by simulated annealing // Science. 1983. V. 220. No. 4598. P. 671–680.