

МОДЕЛИРОВАНИЕ ПОВЕДЕНИЯ АГЕНТОВ, НА ОСНОВЕ ИСПОЛЬЗОВАНИЯ НЕЙРОННОГО ГАЗА¹

Шаринова Т.И.,

Бесхлебнова Г.А., к.т.н.

*Научно-исследовательский институт системных исследований
РАН*

e-mail: gab19@mail.ru

1. ВВЕДЕНИЕ

В работе [1] исследовалось поведение модельного организма, или агента, имеющего потребности питания, безопасности, накопления знаний. При этом использовался метод растущего во времени нейронного газа. Суть этого метода состоит в следующем.

Имеется сеть связанных между собой нейронов, узлов сети. Каждый нейрон-узел запоминает определенную семантическую информацию S_i , 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_i минимально (t – дискретное время). Если для k -го 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 запоминает текущую семантическую информацию $S(t)$.

Связи между нейронами сети формируются при переходе между 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 организуется перемещение агента в соответствии с запомненной сенсорной информацией. В этом случае предварительно происходит блуждание агента по пространству. При этом веса нейронов обучаются методом обучения с подкреплением [2]. Чем более благоприятное место организма в пространстве, тем больше

¹ Работа поддержана программой Президиума РАН № 15, проект 2.10

вес нейрона, вектор S_i которого определяет данную точку в пространстве. После предварительного блуждания по пространству формируется нейронная сеть со связями между нейронами. После этого движение агента может управляться нейронной сетью следующим образом. Каждый такт времени агент находится в точке пространства, соответствующей вектор 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.

В настоящей работе строится и анализируется модель поведения агента, использующего метод растущего нейронного газа. В отличие от работы [1], в которой исследовалось довольно сложное поведение агентов с рядом потребностей и мотиваций, настоящая модель уделяет особое внимание на анализе специфики блуждания агента в одномерном и двумерном пространстве. Также особое внимание уделяется процессам формирования растущего нейронного газа при таком блуждании. Формирование нейронного газа соответствует изложенным выше принципам. Рассматриваем одну потребность – потребность питания.

2. МОДЕЛЬ ДИНАМИКИ АГЕНТА

Одномерный случай. Строим модель динамики агента, считая, что агент должен решать две основные задачи: 1) изучение лабиринта, 2) поиск источника пищи.

При этом используем метод отжига [3]. Т.е. первоначально агент преимущественно совершает случайное блуждание, при этом формируется сеть нейронного газа, а в дальнейшем вероятность случайного блуждания уменьшается и происходит движение между узлами нейронной сети в соответствии с весами нейронов.

Описание модели. Основные предположения одномерной модели состоят в следующем:

1. Рассматривается агент, который может двигаться в одномерном пространстве x .
2. Имеется коридор длиной L с источником питания. Задача агента – исследование коридора и поиск источника питания.
3. Источник пищи имеет небольшой размер d .

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

5. Агент функционирует в дискретном времени t . Каждый такт времени агент совершает движение, при этом его координата x изменяется на некоторую величину $\Delta x(t)$.

6. Когда координата агента совпадает с источником пищи, ресурс агента автоматически увеличивается на величину Δr .

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

7. Система управления задается растущей нейронной сетью. На вход нейрона подается текущая координата агента $x(t)$.

8. Каждый нейрон в качестве памяти запоминает определенную координату x_i .

9. Имеется два режима динамики агента: 1) режим случайного движения и 2) режим детерминированного перемещения, перемещения в соответствии весами узлов-нейронов нейронного газа

10. 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 дальнейшем эта вероятность постепенно уменьшается и происходит детерминированное движение в соответствии с весами нейронов. Таким образом, реализуется метод отжига: при малых временах t агент движется случайно, при больших временах t – детерминировано.

11. В режиме случайного поиска агент перемещается на случайную величину, после чего координата агента становится равной x_k , наиболее близкая к $x(t)$. Если расстояние $|x_k - x(t)|$ меньше порога Th , то величина x_k в памяти нейрона немного сдвигается, приближаясь к $x(t)$. Если координата $|x_k - x(t)| > Th$, то формируется новый нейрон, в памяти которого записывается текущая координата $x(t)$.

12. При появлении нового нейрона формируется связь от предыдущего нейрона к новому. За счет случайного поиска формируется достаточно большая нейронная сеть, так что в дальнейшем можно будет переходить между нейронами, которые связаны между собой.

13. В режиме детерминированного перемещения по координате $x(t)$ определяется текущий нейрон, который соответствует этой координате. Затем определяются веса всех «контактных» нейронов, которые связаны с текущим, и среди этих контактных нейронов находится предпочтительный, имеющий наибольший вес. Координата

агента становится равной координате, хранящейся в памяти предпочтительного нейрона.

14. Тот нейрон, к которому произошел переход в данный момент времени t , в том числе формируемый в данный момент времени t будет называть активным в момент t .

15. При переходе от нейрона к нейрону в обоих режимах производится обучение. При обучении меняются веса нейронов методом обучения с подкреплением [2,3], а именно меняется вес того нейрона, который был активным в момент $t-1$:

$$\Delta W_{t-1} = \alpha[r_t + \gamma W_t - W_{t-1}], \quad (1)$$

где W_{t-1} и W_t – веса нейронов, активных в моменты времени $t-1$ и t , α – скорость обучения, γ – дисконтный фактор, r_t – величина подкрепления в момент времени t , равная Δr , если в данный момент координата агента совпадает с источником пищи, либо равная 0 в противном случае.

Использовались следующие параметры моделирования: длина коридора $L = 100$, размер источника пищи $d = 10$, центр источника пищи расположен посередине коридора при $x = 50$, увеличение ресурса агента от источника пищи $\Delta r = 1$, порог сравнения координат x_k и $x(t)$ равен $Th = 1$, скорость обучения $\alpha = 0.1$, дисконтный фактор $\gamma = 0.9$, характерное время уменьшения вероятности выбора режима случайного поиска равно 1000 тактов времени, характерное перемещение агента при случайном поиске равно 10.

Результаты моделирования. Результаты моделирования для одномерного случая представлены на рис. 1-3. Рис. 1 показывает зависимость координаты агента от времени. Видно, что сначала агент совершает случайные движения, изучая лабиринт. При больших временах агент приближается к источнику пищи, практически оставаясь на месте. Естественно, что ресурс агента, пополняемый при питании агента, возрастает.

Динамика веса W_t текущего нейрона, активного в момент времени t , показана на рис. 2.

Зависимость весов нейронов W_i от координаты нейрона x_i после обучения, т.е. в конце расчета, показана на рис. 3. Чем дальше x_i находится от источника пищи, тем меньше вес соответствующего нейрона.

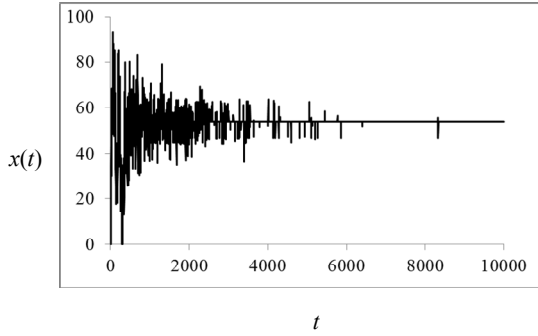


Рис. 1. Зависимость координаты агента от времени $x(t)$

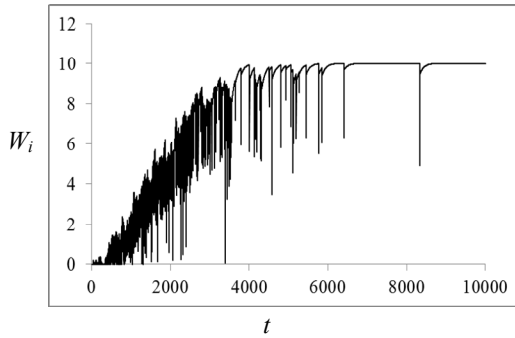


Рис. 2. Вес узла, в котором находится агент, в зависимости от времени t

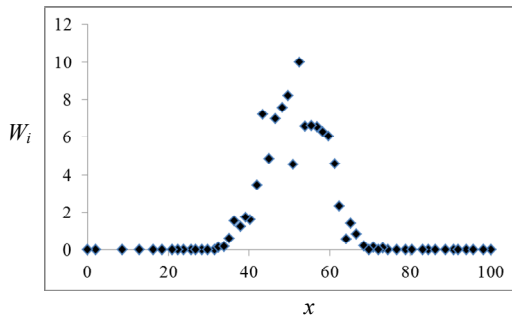


Рис. 4. Зависимость веса нейрона W_i от его координаты x_i

Представленные результаты показывают, что построенная модель обеспечивает нетривиальный вариант режима обучения с подкреплением, который обеспечивает рост весов нейронов и самостоятельное нахождение источника пищи агентом.

Двумерный случай. Для этого случая рассматривалось поведение агента, аналогичное изложенному выше. Основные аспекты модели состоят в следующем.

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

2. Имеется лабиринт с источником питания. Лабиринт представляет собой набор коридоров постоянной ширины (рис. 4), в одном из коридоров расположен источник питания (на рисунке показан серым фоном, находится внизу слева). Задача агента – создание когнитивной карты лабиринта и поиск источника питания.

3. Агент может оценивать изменение текущего своего положения по сравнению с предыдущим тактом времени, когда расстояние от ближайших стенок лабиринта сильно меняется. Позиции, соответствующие сильному изменению окружающей обстановки, отмечены на рис. 4 кружками. Эти позиции специально запоминаются в нейронной сети агента.

Результаты моделирования динамики агента для двумерного случая представлены на рис. 4.

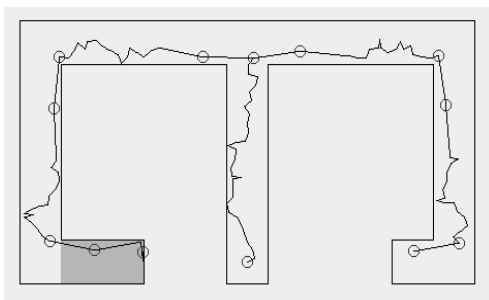


Рис. 4. Движение агента по двумерному лабиринту. Траектория агента показана тонкой изломанной линией, точки, характеризующие сильное изменение окружающей обстановки – кружками, источник питания – серым фоном.

Анализ полученных результатов показал, что агент успешно находит источник питания, после этого ресурс агента растет. Кроме того, показана возможность резкого сокращения размеров нейронной сети, в которой можно запоминать только точки пространства, в которых сильно меняется окружающая ситуация.

3. ЗАКЛЮЧЕНИЕ

Таким образом, построена модель 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 одномерного и двумерного случая. Построен вариант модели растущего нейронного газа, в котором радикально сокращается число узлов-нейронов.

Благодарность

Авторы благодарны Редько В.Г. за полезное и плодотворное обсуждение теоретических подходов и результатов работы

Литература

1. 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.
2. Саттон Р.С., Барто Э.Г. Обучение с подкреплением. М.: Бином, 2011.
3. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P.. Optimization by simulated annealing // *Science*. 1983. Vol. 220. No. 4598. P. 671–680.