Моделирование поискового поведения агентов с использованием метода нейронного газа

Т.И. Шарипова

Центр оптико-нейронных технологий НИИСИ РАН

В работе [1] исследовалось поведение модельного организма (агента), имеющего потребности питания, безопасности, накопления знаний. При этом использовался метод растущего во времени нейронного газа.

В настоящей работе строится и анализируется модель поведения агента, использующая 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.

Одномерный случай

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 пищи.
3. Источник пищи имеет небольшой размер d.
4. Агент имеет ресурс $R(t)$, 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 за один такт времени увеличивается на величину Δr.
7. Агент имеет свою систему управления, на сенсорный вход которой поступает координата агента $x(t)$.
8. Система управления агента задается растущей нейронной сетью. На вход активного нейрона подается текущая координата агента $x(t)$.
9. Каждый нейрон имеет память, он запоминает определенную координату x_i, в данном случае вектор памяти $\mathbf{S}_i$ имеет одну компоненту, равную x_i.
10. Имеется два режима динамики агента: 1) режим случайного движения и 2) режим детерминированного перемещения, перемещения в соответствии весами узлов-нейронов нейронного газа.
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 дальнейшем эта вероятность постепенно уменьшается и происходит переход к детерминированному движению в соответствии с весами нейронов. Таким образом, реализуется метод отжига: при малых временах t агент движется случайно, при больших временах – детерминировано.
12. В режиме случайного поиска после перемещения агента его координата становится равной $x(t)$. Определяется нейрон, в памяти которого хранится координата 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 $|x_k - x(t)| > Th$, то формируется новый нейрон, в памяти которого записывается текущая координата $x(t)$.

13. При появлении нового нейрона формируется связь от предыдущего активного нейрона к новому. За счет случайного поиска формируется достаточно большая нейронная сеть, так что в дальнейшем в режиме детерминированного перемещения будут осуществляться переходы между нейронами, которые связаны между собой.

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.

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

$$\Delta W_{t\text{-}1} = \alpha\,[r_{t\text{-}1} + \gamma W_t - W_{t\text{-}1}]\,, \qquad (1)$$

где $W_{t\text{-}1}$ и W_t – веса нейронов, активных в моменты времени t-1 и t, α – скорость обучения, γ – дисконтный фактор, $r_{t\text{-}1}$ – величина подкрепления, полученного в момент времени t-1. $r_{t\text{-}1} = \Delta r$, если в момент t-1 координата агента совпадает с источником пищи, либо $r_{t\text{-}1} = 0$ в противном случае.

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

Результаты моделирования для одномерного случая представлены на рис. 1-4. Рис. 1 показывает зависимость координаты агента от времени. Видно, что сначала агент совершает случайные движения. При больших временах t агент приближается к источнику пищи, а в конце расчета практически остается на месте возле источника. Естественно, что ресурс агента, пополняемый при его питании, с течением времени возрастает (рис. 2).

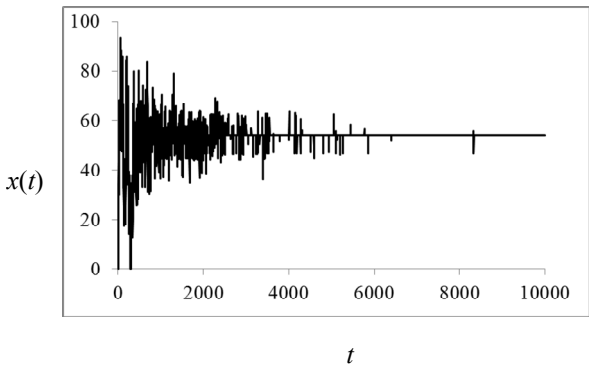

Рис. 1. Зависимость координаты агента от времени

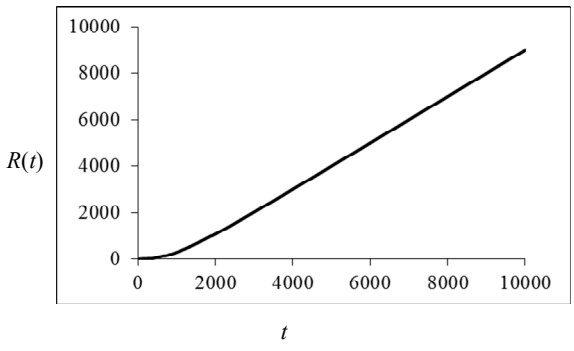

Рис. 2. Зависимость ресурса агента от времени

Динамика веса W_t текущего нейрона, активного в момент времени t, показана на рис. 3.

Зависимость весов нейронов W_i от координаты нейрона x_i после обучения, т.е. в конце расчета, показана на рис. 4. Чем дальше x_i находится от источника пищи, тем меньше вес соответствующего нейрона.

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

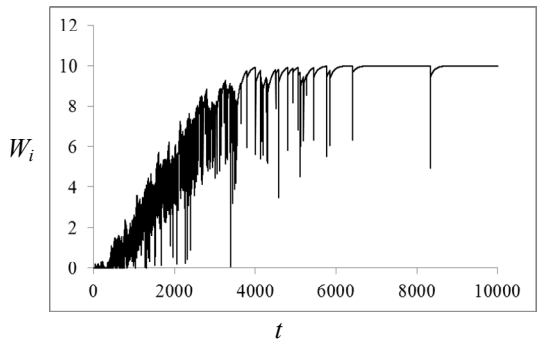

Рис. 3. Вес активного нейрона в зависимости от времени

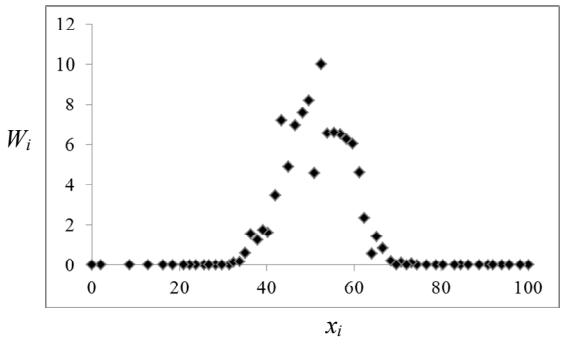

Рис. 4. Зависимость веса нейрона W_i от его координаты x_i для обученного агента

Двумерный случай

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

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

2. Имеется лабиринт с источником питания. Используемый при моделировании лабиринт показан на рис. 5. Лабиринт состоит из нескольких прямоугольных участков, «комнат». Агент движется внутри лабиринта. Задача агента – исследование лабиринта и поиск источника пищи.

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

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 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 $\mathbf{S}_i$ имеет шесть компонент: четыре расстояния до стенок лабиринта, находящихся спереди, сзади, слева и справа от агента, а также координаты агента x_i, y_i.

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

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 дальнейшем эта вероятность постепенно уменьшается и происходит переход к детерминированному движению в соответствии с весами нейронов. Таким образом, реализуется метод отжига: при малых временах *t* агент движется случайно, при больших временах – детерминировано.

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 до стенок лабиринта и координаты.

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.

Использовались следующие основные параметры моделирования: увеличение ресурса агента от источника пищи $\Delta r = 1$, характерное время уменьшения вероятности выбора режима случайного поиска равно 1000 тактов времени.

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

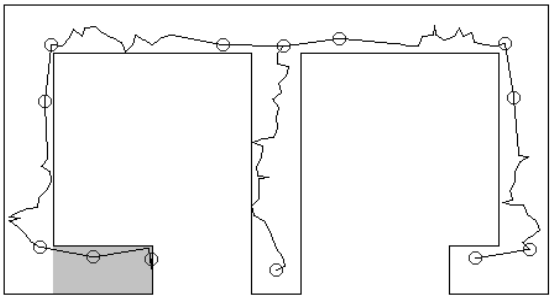

Рис. 5. Движение агента по двумерному лабиринту. Траектория движения агента показана изломанной линией, точки, характеризующие сильное изменение окружающей обстановки отмечены кружками, источник пищи показан серым фоном.

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).

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. Для двумерного случая построен вариант модели растущего нейронного газа, в котором радикально сокращается число узлов-нейронов за счет того, в нейронах запоминаются не все точки, в которых побывал агент, а только те, в которых радикально меняется окружающая среда.

Работа выполнена при поддержке РФФИ, проект 13-01-00399.

Хотелось бы выразить благодарность за содействие и руководство, оказанное Редько В.Г., при выполнении данного доклада.

Литература

1. M.V. Butz, E. Shirinov, K. Reif Self-organizing sensorimotor maps plus internal motivations yield animal-like behavior // *Adaptive Behavior*. – 2010. V. 18. No. 3-4, pp. 315-337.

2. Р.С Саттон, Э.Г. Барто Обучение с подкреплением. М.: Бином, 2011.

3. S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi Optimization by simulated annealing // *Science*. 1983. Vol. 220. No. 459, pp. 671–680.