

# Модель автономного агента-охранника

Редько В.Г., Сохова З.Б.

НИИ системных исследований РАН, Москва.

## The model of an autonomous agent-guard

Red'ko V.G., Sokhova Z.B.

Scientific Research Institute for System Analysis, Russian Academy of Science, Moscow, Russia

vgredko@gmail.com, zarema\_s@mail.ru

The computer model of an autonomous agent-guard, which is functioning in a circular ring that is divided into sectors, is designed and investigated. Two versions of the model for the agent-guard are analyzed by means of computer simulations.

### 1 Введение

В настоящей работе построена и проанализирована модель автономного агента на примере агента-охранника. Исследованы два типа модели. В первой модели обучение агента производится по факторам потребностей, во второй – обучение производится по общей энергии агента-охранника. Работа развивает модели [1, 2], в которых было начато исследование агентов, обладающих естественными потребностями и мотивациями.

Функция агента-охранника состоит в поиске нарушителей и поддержании своего внутреннего ресурса. Нарушители появляются на охраняемой территории случайно. При обнаружении нарушителя агент ударяет его, и нарушитель покидает территорию. Качество поведения агента-охранника оценивается по числу нарушителей, которых он не пропустил на территорию.

### 2 Модель агента-охранника с двумя потребностями

#### 2.1 Описание модели

Рассматривается один агент-охранник, обладающий внутренним ресурсом  $R(t)$ . Время  $t$  дискретно. Моделью мира агента является круговое кольцо, разбитое на шесть секторов. Кольцо представляет собой границу охраняемой территории. В четных секторах находятся розетки. В каждый такт времени в любом из секторов с вероятностью  $p_1$  появляется нарушитель.

Считаем, что агент-охранник имеет две потребности: *питания* и *охраны территории*. Если ресурс агента  $R$  меньше порога  $r_{th1}$ , то ведущей потребностью является питание, иначе ведущей является потребность охраны.

Потребностям агента соответствуют два фактора: фактор питания  $F_f$  и фактор охраны территории  $F_p$ .

Фактор питания пропорционален ресурсу агента:  $F_f = k_f R(t)$ . Фактор охраны территории увеличивается при выполнении агентом действия «удар» и наличия нарушителя в одной клетке с агентом на  $\Delta F_p$  и уменьшается на 1 в других ситуациях.

Удовлетворение ведущей потребности является положительным подкреплением при обучении агента.

#### 2.2 Система управления агента-охранника с двумя потребностями

Система управления агента основана на наборе правил вида:  $S_k \rightarrow A_k$ , где  $S_k$  – ситуация,  $A_k$  – действие,  $k$  – номер правила. Согласно правилам в ситуации  $S_k$  нужно выполнить действие  $A_k$ . Каждое правило имеет свой вес  $W_k$ . Веса правил изначально случайны, а затем модифицируются методом обучения с подкреплением [3].

Ситуация  $S_k$  определяется 1) наличием или отсутствием розетки в текущем секторе агента и в двух соседних секторах, 2) наличием или отсутствием нарушителя в текущем секторе и в двух соседних секторах, и 3) ведущей потребностью.

В каждый такт времени агент может выполнять одно из следующих действий  $A_k$ : 1) *питание*, 2) *перемещение на один сектор по или против часовой стрелки*, 3) *удар*, 4) *отдых*. Если действие «питание» вырабатывается в секторе, где имеется розетка, то ресурс агента  $R(t)$  увеличивается на  $r_1$ . При выполнении действий «перемещение» (в любом из двух направлений), удар и отдых ресурс агента уменьшается на величины  $r_2, r_4, r_5$ .

Каждый такт времени с вероятностью  $1-\varepsilon$  выполняется то действие, для которого вес  $W_k$  соответствующего ему правила для текущей ситуации максимален, с вероятностью  $\varepsilon$  выполняется случайное действие.

### 2.3 Схема обучения

Используется схема обучения с подкреплением [3]. Подкреплением является изменение фактора ведущей потребности  $F_f$  или  $F_p$ :

$$\Delta W(t-1) = \alpha [F_L(t) - F_L(t-1) + \gamma W(t) - W(t-1)],$$

где  $F_L(t)$  – фактор ведущей в такт времени  $t$  потребности,  $W(t)$  – вес правила, примененного в такт  $t$ ,  $\alpha$  – параметр скорости обучения,  $\gamma$  – дисконтный фактор.

### 2.4 Результаты моделирования

Параметры компьютерного моделирования составляли:  $\Delta F_p = 5$ ,  $k_F = 0.2$ ,  $\varepsilon = 0.05$ ,  $\gamma = 0.9$ ,  $\alpha = 0.1$ ,  $r_{th1} = 50.0$ ,  $r_1 = 50.0$ ,  $r_2 = 1$ ,  $r_4 = 1$ ,  $r_5 = 5$ ,  $p_1 = 0.1$ .

Моделирование показало, что агент-охранник обучается поддерживать внутренний ресурс на уровне не ниже заданного порога. Количество нарушителей также уменьшается. Зависимости ресурса агента и количества нарушителей, остающихся в мире, от номера такта времени представлены на рис. 1, 2.

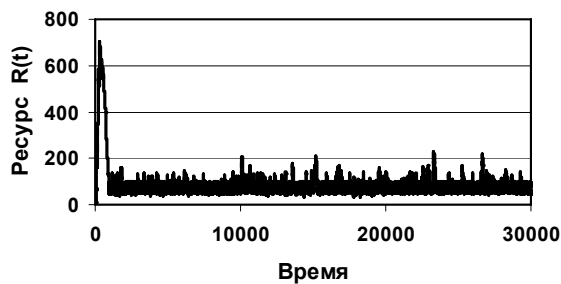


Рис. 1. Динамика ресурса агента-охранника  $R(t)$

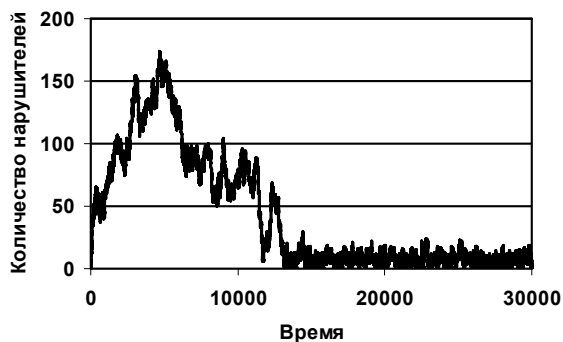


Рис. 2. Динамика количества нарушителей

## 3 Модель агента-охранника без потребностей

### 3.1 Описание модели

Система управления агента аналогична системе управления агента с потребностями, но в отличие от модели, описанной выше, не содержит ведущую потребность.

В данной модели под ресурсом агента  $R(t)$  понимается *общая энергия*, которую агент накапливает при выполнении любых действий. При выполнении действия «питание» и наличии розетки в текущем секторе ресурс агента увеличивается на  $dR_f$ . При выполнении действия «удар» и наличии нарушителя в секторе ресурс агента увеличивается на  $dR_s$ . При питании, перемещении, ударе и отдыхе ресурс агента уменьшается соответственно на  $r_0$ ,  $r_1$ ,  $r_2$ ,  $r_3$ .

### 3.2 Схема обучения

В данном варианте модели подкреплением является изменение ресурса  $R(t)$ :

$$\Delta W(t-1) = \alpha [R(t) - R(t-1) + \gamma W(t) - W(t-1)],$$

где  $W(t)$  – вес правила, примененного в такт  $t$ ,  $\alpha$  – параметр скорости обучения,  $\gamma$  – дисконтный фактор.

### 3.3 Результаты моделирования

Параметры компьютерного моделирования составляли:  $dR_f = 1$ ,  $dR_s = 3$ ,  $\varepsilon = 0.05$ ,  $\gamma = 0.9$ ,  $\alpha = 0.3$ ,  $r_0 = 1$ ,  $r_1 = 1$ ,  $r_2 = 1$ ,  $r_3 = 1$ ,  $p_1 = 0.1$ . Моделирование показало, что обученный агент справляется со своей задачей, количество агентов-нарушителей резко уменьшается, при этом ресурс агента растет

## 4 Заключение

Построена и исследована модель автономного агента-охранника с потребностями и без потребностей. Компьютерные эксперименты показали, что после обучения агента в модели без потребностей количество нарушителей остающихся в мире незначительно меньше, чем в модели с потребностями.

В модели с потребностями агент-охранник показывает более разумное поведение, так как он обучается поддерживать ресурс на определенном уровне.

## Список литературы

- [1] А.Г. Коваль, В.Г. Редько. Поведение модельных организмов, обладающих естественными потребностями и мотивациями. *Математическая биология и биоинформатика*, 2012, Т. 7. № 1, 266-273.
- [2] В.Г. Редько, Г.А. Бесхлебнова. Моделирование адаптивного поведения автономных агентов. *Нейрокомпьютеры: разработка, применение*, 2010. № 3, 33–38.
- [3] Р.С. Саттон, Э.Г. Барто. *Обучение с подкреплением*, М.: Бинум, 2011.