УДК 519.8

РЕШЕНИЕ ЗАДАЧИ ПРЕСЛЕДОВАНИЯ С ПРОСТЫМ ДВИЖЕНИЕМ НА ПЛОСКОСТИ МЕТОДАМИ ТОПОЛОГИЧЕСКОЙ НЕЙРОЭВОЛЮЦИИ

В.В. Шульгин

Таврический национальный университет им. В.И. Вернадского
факультет математики и информатики
пр-т Вернадского, 4, г. Симферополь, 95007, Украина
e-mail: v_i_c@gala.net

Abstract

The pursuit problem with simple motion on a plane is considered in this paper. It is shown that if fitness function depends on capture time and target distance and pursuer is being placed onto Archimedian spiral during learning process, then topology neuroevolution method called "NEAT" is able to efficiently solve the problem. Resultant neural network is being compared with two known theoretical solutions via competition series.

Введение

Рассмотрим следующую математическую игру: два игрока (назовем их для определенности P и E) управляют своими объектами на плоскости. Объекты обозначены точками $P(P_x, P_y)$ и $E(E_x, E_y)$ соответственно. Игрок P старается *осуществить захват*, то есть сократить расстояние $|PE|$ до некоторой наперед заданной величины L, причем за минимально возможное время, игрок E, в свою очередь, прилагает все силы, чтобы избежать захвата или отсрочить его на максимально возможное время. Так как мы закрепили за игроками их роли, далее игрок P будет называться *преследователем*, а игрок E — *преследуемым* или *целью*.

Если объекты движутся так, что их скорость остается постоянной, а направление движения может выбираться произвольно в любой момент времени, то говорят, что объекты обладают *простым движением*. Описанная таким образом игра называется *игрой преследования с простым движением*.

В данной статье *под нейроэволюцией будет пониматься технология развития искусственных нейронных сетей, управляемая генетическими алгоритмами*. Целью работы является применение нейроэволюционных методов для решения задачи преследования, а именно синтезирование искусственной нейронной сети, управляющей объектом-преследователем, которая позволяет эффективно осуществлять захват в рамках проводомиго эксперимента.

1. Постановка задачи

В качестве естественной платы $\mathcal{P}$ игры преследования выбирается время, за которое был осуществлен захват (в случае, если игрок P не в состоянии осуществить захват, полагаем плату равной $+\infty$). Итак, игрок P минимизирует плату, игрок E — максимизирует; цена игры

$$\mathcal{V} = \min_P \max_E \mathcal{P}.$$

Пусть ω — модуль скорости игрока E, τ — модуль скорости игрока P; тогда уравнения движения можно записать в виде:

$$\dot{E}_x = \omega \cos u \quad (1)$$
$$\dot{E}_y = \omega \sin u \quad (2)$$
$$\dot{P}_x = \tau \cos v \quad (3)$$
$$\dot{P}_y = \tau \sin v, \quad (4)$$

$$\tau \geq \omega > 0. \quad (5)$$

Величины u и v, задающие угол наклона вектора скорости над осью абсцисс, называются *управлениями*.

Координаты объекта на плоскости в совокупности с компонентами его вектора скорости образуют фазовые координаты объекта. Предполагается, что в любой момент времени каждый из игроков знает значения своих и чужих фазовых координат, но ему не известны управления противника.

В теории дискретных игр стратегия игрока состоит из множества решений, указывающих, как следует ему вести себя в каждой из ситуаций, которая может возникнуть на протяжении партии. Естественной аналогией в игре преследования является выбор управлений как функций фазовых координат. Обозначим $\dot{E} = (\dot{E}_x, \dot{E}_y), \quad \dot{P} = (\dot{P}_x, \dot{P}_y)$; тогда стратегиями игроков будут функции

$$u = u(E, P, \dot{E}, \dot{P}) \quad (6)$$
$$v = v(E, P, \dot{E}, \dot{P}). \quad (7)$$

Стратегии u' и v' такие, что:

$$\mathcal{P}(u', v') = \min_v \max_u \mathcal{P}(u, v) = \mathcal{V}$$

называются *оптимальными*.

Задача преследования заключается в нахождении оптимальных стратегий для обоих игроков.

Известно, что при простом движении оптимальной стратегией для обоих игроков является движение вдоль луча PE [1]. В дальнейшем, мы будем ссылаться на эту стратегию при помощи термина *О-стратегия*. Игрок P, следующий О-стратегии, будет называться *О-преследователем*.

По определению оптимальности, если игрок E отклоняется от оптимальной стратегии, преследователь P имеет возможность осуществить захват за более короткое время. В дальнейшем будет развиваться ситуация, в которой E не следует своей оптимальной стратегии.

В [2] рассмотрена так называемая *стратегия параллельного сближения (П-стратегия)*, изначально разработанная для задачи преследования, в которой игроки применяют свою функцию управления дискретно. Она заключается в следующем: пусть в начальный момент времени объекты расположены в точках $E^0(E_x^0, E_y^0)$ и $P^0(P_x^0, P_y^0)$; применим такое преобразование координат на плоскости, что точки

перейдут в $E^0(0,0)$ и $P^0(0,-b)$, $b \geq 0$ соответственно. Тогда П-стратегия предписывает игроку P в любой момент времени выбирать свой направляющий вектор $\dot{P}$ следующим образом:

$$\dot{P}_x = \dot{E}_x \tag{8}$$

$$\dot{P}_y = \sqrt{\tau^2 - \dot{E}_y^2} \tag{9}$$

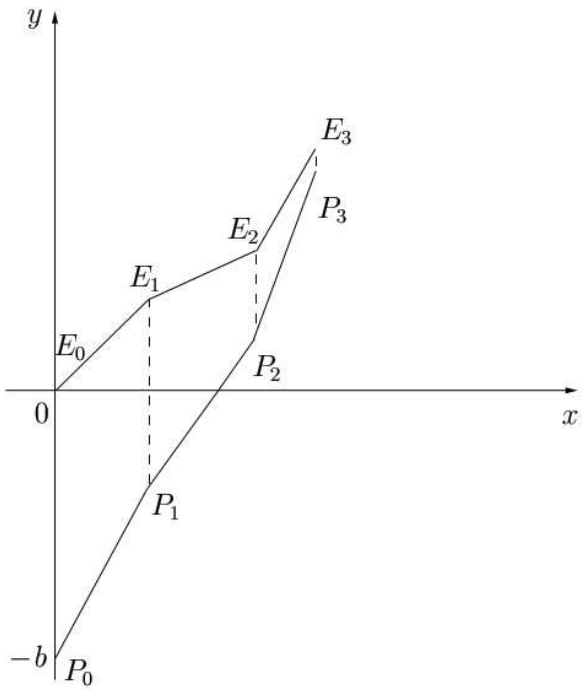

Рис. 3. Поведение игрока P, пользующегося П-стратегией, при $\tau = \frac{3}{2}\omega$.

На рис. 3 показано, как сокращается расстояние $|PE|$ с течением времени, если игрок P следует П-стратегии. Такого игрока будем называть *П-преследователем.*

С точки зрения нейронных сетей, под решением задачи преследования будем понимать *создание нейронной сети, синтезирующей функцию управления для объекта P, которая позволяет осуществлять захват объекта E при любом допустимом его поведении.*

2. Метод наращиваемых топологий

Многие подходы в нейроэволюции подразумевают использование генетических алгоритмов в качестве универсального аппарата многомерной оптимизации для настройки весов нейронных сетей фиксированной топологии, однако, существует и другое направление — TWEANN-методы (Topology and Weight Evolving Artificial Neural

Networks — искусственные нейронные сети, в которых при помощи нейроэволюции настраиваются не только веса, но и топология).

Одним из наиболее перспективных методов является метод NEAT (NeuroEvolution of Augmenting Topologies — нейроэволюция наращиваемых топологий) [3]. Его ключевая особенность заключается в том, что все особи (здесь и далее под термином "особь" будем понимать конкретную нейронную сеть с точки зрения генетических алгоритмов) начинают эволюционировать с минимальной конфигурацией — а именно, имея лишь входной и выходной слои, полностью связанные между собой. Эта особенность позволяет находить максимально простые решения задачи и, при правильной организации процесса, требует меньших временных затрат, чем подходы с фиксированной топологией или TWEANN-методы, работающие со случайной конфигурацией сети. Это объясняется тем, что чем меньше сеть содержит связей, тем меньше размерность пространства поиска оптимальных весов.

Как генетический алгоритм метод NEAT имеет три вида оператора мутации:

- изменение веса связи,
- добавление новой связи между нейронами,
- добавление нового скрытого нейрона.

В соответствии с представлением функции управления (6), архитектура сети выбирается следующим образом. На вход подается девять сигналов: по четыре фазовые координаты обоих игроков, плюс сигнал смещения — постоянный единичный ввод; сеть выдает два выходных значения x и y — компоненты направляющего вектора, по которым вычисляется управление

$$u = \operatorname{arctg}\frac{y}{x}, \quad x^2+y^2 > \varepsilon, \quad 0 < \varepsilon \lll 1, \tag{1}$$

$$u = 0, \quad x^2+y^2 \le \varepsilon. \tag{2}$$

В качестве активационного элемента сети используется функция $\mathrm{th}(x)$ — гиперболический тангенс.

Обучающий эксперимент проводится следующим образом: каждой особи в популяции на протяжении одного поколения предоставляется ряд попыток осуществить захват цели, движущейся по элементарным траекториям, за некоторый отрезок времени. При этом все особи ставятся в равные условия, в том смысле, что имеют сходные начальные условия. Усредненное значение фитнес-функции по всем попыткам считается фитнес-значением данной особи в текущем поколении. В качестве обучающих элементарных траекторий использовались луч и дуга окружности.

3. Выбор фитнес-функции

Как и в любой задаче, решаемой генетическими алгоритмами, основной проблемой является выбор адекватной фитнес-функции. Применимо к задаче преследования, ее необходимым свойством должно быть поощрение захвата, ведь именно это является конечной целью эксперимента. Пусть особи c из N попыток удается осуществить M захватов, тогда простейшая фитнес-функция $f_1 = \frac{M}{N}$; будем предполагать, что фитнес-функция принимает значения из промежутка $[0,1]$. Эта функция обладает

тем недостатком, что она не дает количественной оценки успешности особи в одном опыте, таким образом нельзя решить, обусловлена ли она направленным преследованием или случайным столкновением с целью.

Сконструируем более сложную функцию f_2, обеспечивающую такую оценку; пусть за захват цели в одном опыте особь c получает "призовое" фитнес-значение $\frac{1}{2} \leq p \leq 1$ такое, что: $f_2(c) \geq p$, если захват был произведен, и $f_2(c) \leq p$, иначе. Среди особей, не сумевших произвести захват, можно наградить те особи, которые в конечный момент времени приблизились к цели на меньшее расстояние. А среди особей, удачно завершивших опыт, выделим тех, которые сделали это за меньшее время. Пусть T — время, отведенное на опыт, T_f — время окончания опыта, $T_f \leq T$; положим $D(t) = ((E_x(t) - P_x(t))^2 + (E_y(t) - P_y(t))^2)^{\frac{1}{2}}$ — расстояние между объектами в момент времени t; тогда условием захвата является выполнение $D_f < L$, где $D_f = D(t_f)$. Определим семейство функций

$$f_2^p = \begin{cases} p + (1-p)(1 - \left(\frac{T_f}{T}\right)^2), & D_f < L \\ (1-p)\left(\frac{L}{D_f}\right)^2, & D_f \geq L \end{cases} \quad (1)$$

Чем меньше значение параметра p, тем слабее проявляется корреляция между значением фитнес-функции и числом осуществленных захватов. Заметим, что использование f_1 равносильно использованию f_2^1: это упрощает исследование того, как эффективность обучения зависит от выбранной функции.

На рис. 4 изображены тренды функций $\max f_2^p$ — максимальных фитнес-значений по всем особям в каждом поколении. Видно, что эффективность растет при $0,5 < p < 0,8$ а затем падает при $0,8 < p < 1,0$; можно сделать следующий вывод: наиболее целесообразным является использование f_2^p при $0,7 \leq p \leq 0,8$.

4. Обучающая выборка

Другим немаловажным вопросом с точки зрения обучения нейронной сети является организация учебной выборки. С одной стороны, она должна охватывать как можно больше возможных случаев, с другой стороны, она должна быть подобрана так, чтобы не расстроить веса сети. Обучающая выборка в поколении C_n, $n \geq 0$ состоит из:

- начального положения цели $E^{n,0}$;
- начальных положений преследователя $\{P_i^{n,0}\}_{i=1}^N$ (N — число попыток для одной особи);
- углового смещения траектории цели β_n.

Пусть $\xi(a,b)$ — непрерывная равномерно распределенная на отрезке $[a,b]$ случайная величина; определим $E^{n,0}$ таким образом:

$$E_x^{0,0} = \xi(-1,\ 1); \quad (1)$$

$$E_y^{0,0} = \xi(-1,\ 1); \quad (2)$$

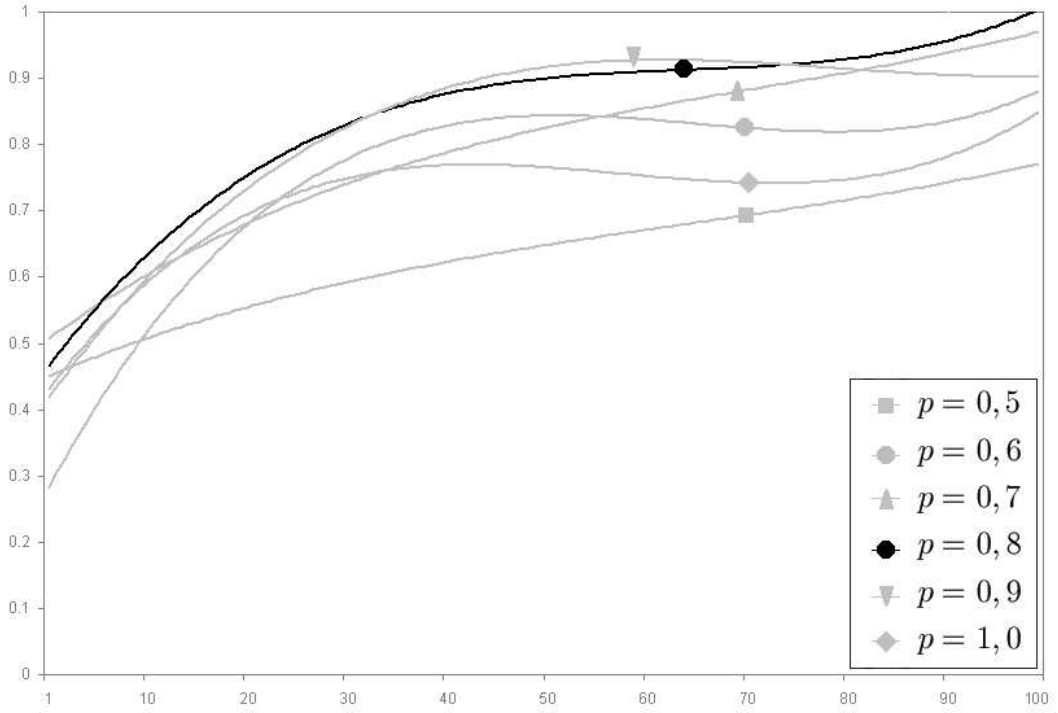

Рис. 4. Тренды $\max f_2^p$ для 100 поколений, построенные по методу наименьших квадратов при помощи кубических полиномов.

$$\forall n \geq 1 \quad A_E^n = ((E_x^{n-1,0})^2 + (E_y^{n-1,0})^2)^{\frac{1}{4}} + h_E; \tag{3}$$

$$E_x^{n,0} = E_x^{n-1,0} + \xi(-A_E^n, A_E^n); \tag{4}$$

$$E_y^{n,0} = E_y^{n-1,0} + \xi(-A_E^n, A_E^n). \tag{5}$$

Определения (1), (2) задают начальное положение в окрестности начала координат $O(0,0)$. A_E^n — максимальная величина, на которую могут измениться координаты $E^{n-1,0}$ в следующем поколении ((4), (5)), она рассчитывается как квадратный корень расстояния $|OE^{n-1,0}|$ плюс малая добавка h_E, гарантирующая то, что $E^{n,0}$ не сойдется к O.

В качестве множества начальных расположений преследователя $\{P_i^n\}_{i=1}^N$ целесообразно взять архимедову спираль с центром в $E^{n,0}$, потому что она формирует равномерное распределение расстояний до цели и равномерное распределение углов направлений на цель. Известно уравнение спирали в полярных координатах $\rho(\varphi) = k\varphi$; требуется равномерно расположить точки на спирали так, что: $\varphi_0 \leq \varphi \leq \varphi_0 + 2\Pi s$, $r \leq \rho(\varphi) \leq R$ (r, R, s — параметры обучения, $\varphi_0 = \text{const}$). Тогда $\forall i : 1 \leq i \leq N$

$$r_i = r + \frac{R-r}{N}(i-1); \tag{6}$$

$$\varphi_i = \frac{2\Pi s}{N}(i-1); \tag{7}$$

$$Q_{i,x}^n = E_x^{n,0} + r_i \cos\varphi_i; \tag{8}$$

$$Q^n_{i,y} \quad = \quad E^{n,0}_y + r_i \sin\varphi_i. \tag{9}$$

Точки $P^{n,0}_i$ выбираются как возмущенные Q^n_i с максимальным отклонением $A^n_P = ((E^{n,0}_x - Q^n_{i,x})^2 + (E^{n,0}_y - Q^n_{i,y})^2)^{\frac{1}{4}}$ аналогично (4), (5).

Угол наклона β_n рассчитывается по правилу:

$$\beta_0 = 0;\ \beta_{dev} = \text{const};\ \beta_{mul} = \text{const} < -1; \tag{10}$$

$$U = \{u_1 = 1,\ \forall k > 1 : u_k = u_{k-1} + k\} = \{1, 3, 7, 12, \dots\}; \tag{11}$$

$$\forall k \geq 1,\ \forall j : 1 \leq j \leq u_k; \quad \beta_{u_k+j-1} = \beta_{u_{k-1}} + \beta_{dev} \cdot (\beta_{mul})^k \cdot j. \tag{12}$$

Этот закон задает колебательные изменения β_n с увеличивающейся амплитудой.

За максимальное время, отведенное для захвата в конкретной попытке, берется величина $1,5 \cdot T_1$, где T_1 — время, необходимое для осуществления захвата О-преследователю.

5. Полученные результаты

Задача преследования решалась с параметрами, указанными в таблице 7.

Таблица 7

параметр	значение
число попыток N	37
шаг времени δt	0,1
скорость преследуемого ω	1,0
скорость преследователя τ	1,5
радиус захвата L	0,25
призовое значение p	0,8
нижняя граница радиуса спирали r	4,0
верхняя граница радиуса спирали R	8,0
число витков спирали s	2
величина h_E в (3)	0,1
величина β_{dev} в (10)	0,1
величина β_{mul} в (10)	-1,1
число поколений	300
число особей в популяции	200
вероятность мутации веса	0,75
вероятность мутации-добавления связи	0,05
вероятность мутации-добавления нейрона	0,01
доля неизменных особей	0,20
использование элитизма	да
селекция методом рулетки	да

Пусть, в поколении C_n особь $c \in C_n$ имеет фитнес-значение $f(c)$, тогда процесс нейроэволюции можно охарактеризовать функциями

$$f_{max}(C_n) = \max_{c \in C_n} f(c) \quad \text{и} \quad f_{avg}(C_n) = \sum_{c \in C_n} \frac{f(c)}{|C_n|},$$

их графики изображенными на рис. 5.

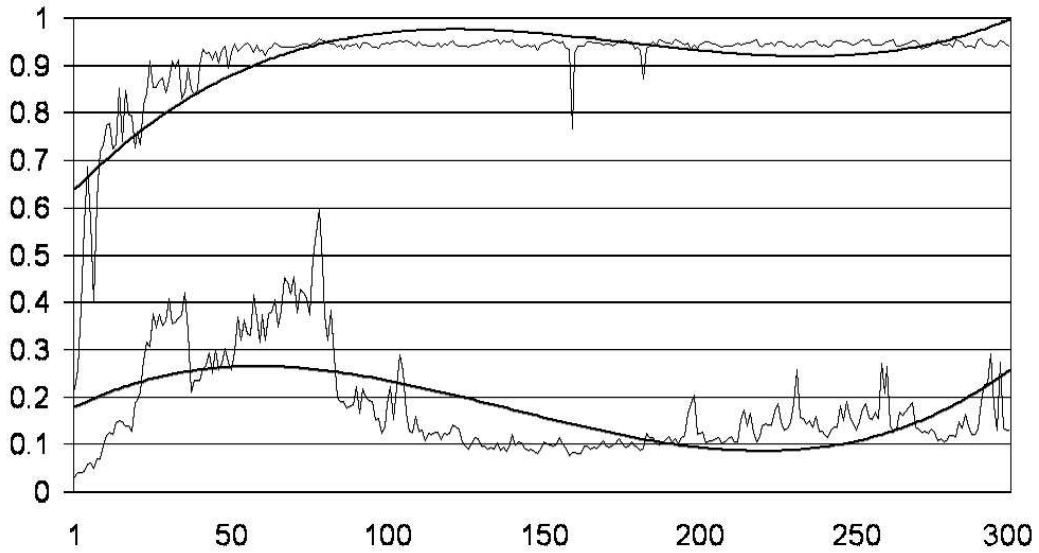

Рис. 5. Графики f_{max}, f_{avg} и их тренды, построенные при помощи кубических полиномов.

Первое решение — особь, осуществившая захват в 100% опытов — было получено в поколении 25, устойчивое решение появилось в поколении 184.

Для сравнительного анализа было проведено 4 серии по 2500 опытов, в которых нейронная сеть соревновалась с П-преследователем и О-преследователем. Так же, как и на этапе обучения, считалось, что нейронная сеть не смогла осуществить захват, если она не смогла это сделать за время $1,5 \cdot T_1$, где T_1 — время, необходимое О-преследователю. Время, затраченное на захват П-преследователем, обозначим T_2.

Вид движения объекта Е зависел от номера серии:

1. по дуге окружности
2. по лучу
3. по дуге окружности с произвольным изменением центра кривизны в моменты времени $t = 1, 2, 3, \ldots$
4. по лучу с произвольным изменением направляющего вектора в моменты времени $t = 1, 2, 3, \ldots$

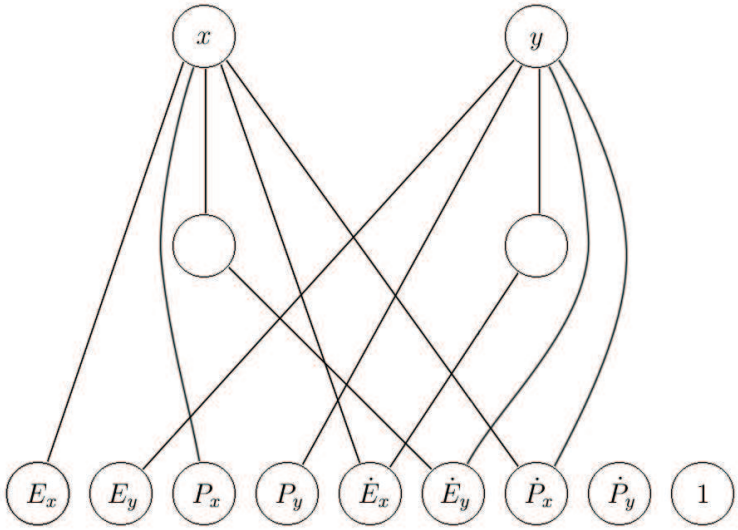

Рис. 6. Топология особи-чемпиона, взятого из поколения 300, проста и имеет всего два скрытых нейрона.

В каждом новом опыте объекты P и E помещались на плоскость со случайными координатами в круг радиуса $R = 32$ с центром в начале координат так, что $|PE| > L$; время изменялось с шагом $\delta t = 0.001$. Все остальные параметры имели такие же значения, что и на этапе обучения.

В таблице 8 дана информация об успешных исходах опытов и о частичном превосходстве одной из стратегий над двумя другими. Здесь, под $p(T_f \leq T_1)$ понимается доля исходов, в которых T_f не превосходит T_1, выраженная в процентах (аналогично для других комбинаций из T_f, T_1 и T_2).

Во-первых, как видно из второго столбца таблицы, сеть справляется с задачей в 100% случаев, причем на множестве входных данных значительно превосходящим множество, использовавшееся на этапе обучения.

Во-вторых, анализ временных характеристик эксперимента дает интересный факт: ни одна из стратегий не превосходит две другие стратегии абсолютно, можно сказать, что они находятся в состоянии некоторого паритета. Так, хотя П-стратегия имеет общее лидерство, в серии 4 она уступает О-стратегии. В свою очередь, О-стратегия в серии 1 проигрывает нейронной сети. В среднем, приблизительно в $\frac{1}{3}$ случаев, сеть ведет себя лучше, чем О-стратегия и приблизительно в $\frac{1}{7}$ случаев — лучше П-стратегии (случаи $T_f = T_1$ и $T_f = T_2$ маловероятны в виду малости δt и поэтому не рассматриваются).

В таблице 9 приведена числовая оценка того, насколько сеть затрачивает больше времени на захват цели в единственном опыте относительно других стратегий. Здесь, под $avg(x)$ понимается усредненное значение x по всем опытам. Значение -3,2% во втором столбце означает, что нейронная сеть в первой серии опытов, в среднем,

Таблица 8

серия	$D(T_f) < L$, %	$p(T_f \leq T_1)$, %	$p(T_f \leq T_2)$, %	$p(T_1 \leq T_2)$, %
1	100	58,9	10,2	8,0
2	100	28,2	2,9	1,5
3	100	29,7	22,3	4,6
4	100	16,7	21,8	67,4
в среднем	100	33,4	14,3	20,4

осуществляет захват быстрее на 3,2%, чем О-преследователь. По итогам эксперимента, в среднем, сеть для захвата цели затрачивает на 4,7% времени больше, чем О-преследователь и на 14,0% больше, чем П-преследователь.

Таблица 9

серия	$avg(\frac{T_f - T_1}{T_f})$, %	$avg(\frac{T_f - T_2}{T_f})$, %
1	-3,2	13,0
2	1,2	22,2
3	4,5	6,4
4	16,2	14,4
в среднем	4,7	14,0

Заключение

Полученные результаты свидетельствуют о том, что использование метода нейроэволюции наращиваемых топологий NEAT в качестве аппарата автоматической настройки архитектуры нейронной сети позволяет синтезировать решение рассматриваемой задачи преследования, которое в целом не многим уступает, а в некоторых случаях даже превосходит известные теоретически разработанные решения.

В дальнейшем имеет смысл исследовать применимость метода к модели с более сложным видом движения, а также решение задачи преследования с несколькими кооперирующимися объектами-преследователями.

Список литературы

1. *Айзекс Р.* Дифференциальные игры. — М.: Мир, 1967. — С. 32–45.
2. *Петросян Л.А., Рихсеев Б.Б.* Преследование на плоскости. — М.: Наука, 1991. — С. 23.
3. *Stanley K., Miikkulainen R.* Efficient Reinforcement Learning through Evolving Neural Network Topologies. — Genetic and Evolutionary Computation Conference (GECCO-2002) papers.