

Глава 4

Лекция 10. Алгоритм AGE

Оглавление

4 Лекция 10. Алгоритм AGE	1
10.1 Алгоритм AGE	2
10.1.1 Кодирование	2
10.1.2 Декодирование устройств	3
10.1.3 Связывание устройств	3
10.1.4 Взаимодействие со средой	4
10.1.5 Операторы	5
10.1.6 Схема алгоритма	6
10.1.7 Результаты экспериментов	7
10.1.8 Приложения	7
10.1.9 Замечания	10

10.1 Алгоритм AGE

Алгоритм AGE (Analog Genetic Encoding) разработан Клаудио Маттиусси и Дарио Флореано [3]¹. Примечательно, что это один из немногих подходов, в которых рассматривается эволюция орграфа в общем виде, абстрактно, привязка к предметной области осуществляется через параметры алгоритма. Т.е. данный алгоритм не является нейроэволюционным «от рождения», т.к. помимо поиска нейронных сетей он может быть применен для синтеза электронных схем, а также реконструкции генных регуляторных сетей.

Применительно к эволюционному поиску ИНС алгоритм AGE может быть использован как для ИНС прямого распространения, так и для рекуррентных ИНС. Ограничений на структуру сети (количество нейронов, связей, слоев и др.) нет, равно как нет и ограничения на направление поиска.

Сам алгоритм ближе к биологии, чем многие подобные алгоритмы как с точки зрения кодирования, так и с точки зрения операторов, применяемых для модификации текущих решений.

10.1.1 Кодирование

Геном особи включает 1 или несколько хромосом (строк символов из конечного алфавита). Пример хромосомы показан на рис. 10.1.

В хромосомах кодируется информация о структуре и весах связей соответствующего орграфа, при этом используется биологически-инспирированный подход. Отдельные подпоследовательности, соответствующие элементам (вершины, метки связей)

¹ Сайт алгоритма с информацией и списком публикаций: <http://lis.epfl.ch/age>.

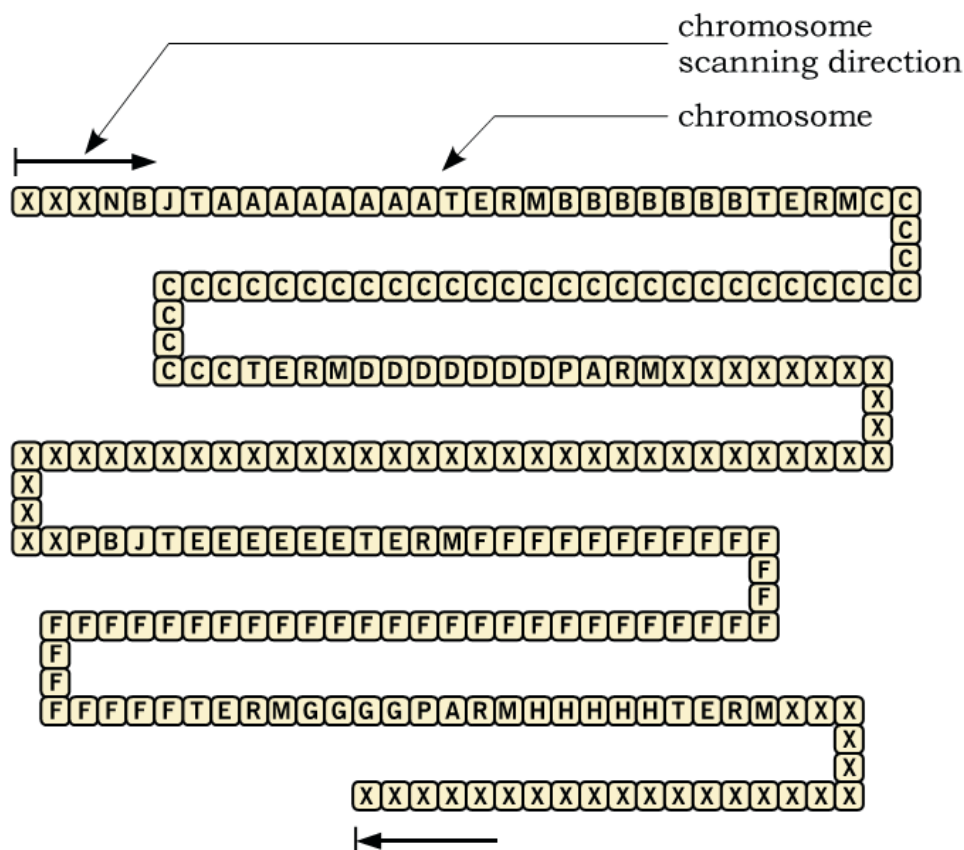


Рис. 10.1: Пример хромосомы, используемой в алгоритме AGE. Использован рисунок из [3].

помечаются (маркируются) специальными последовательностями, токенами, определенными заранее (см. ниже). По имеющимся маркерам можно декодировать хромосому. Пример размеченной хромосомы изображен на рис. 10.2. В хромосоме могут быть и некодирующие участки.

Для всех геномов в популяции задан исходный набор устройств (*devices*), определяющих свойства вершин орграфа. Для каждого устройства задается:

- количество и тип терминалов (входы+выходы)
- начальный токен
- терминальный токен

10.1.2 Декодирование устройств

Кодирование и декодирование информации из хромосомы инспирировано принципами работы генных регуляторных сетей (ГРС) (рис. 10.3).

Пример декодирования описания биполярного транзистора изображен на рис. 10.4.

10.1.3 Связывание устройств

Для связывания используются текстовые метки (последовательности). Фактически сила связи считается с использованием функции $W(L_1, L_2)$, где L_i – метка. Чем боль-

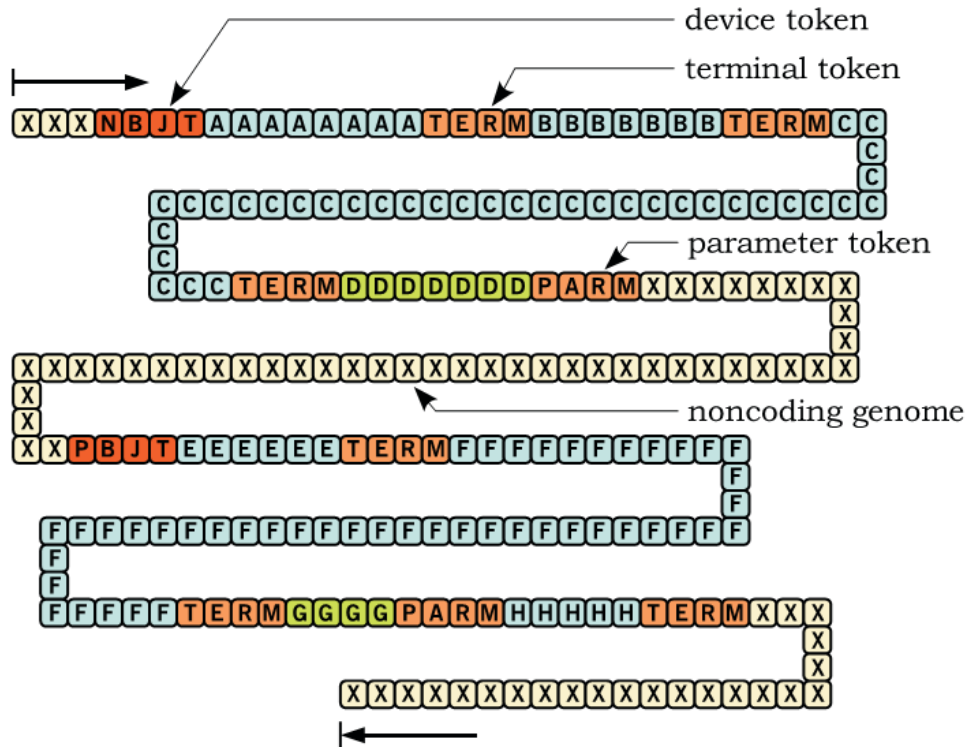


Рис. 10.2: Пример размеченной хромосомы, используемой в алгоритме AGE. NBJT и PBJT являются маркерами начала описания вершин (устройств). TERM и PARM соответствуют окончанию описания меток терминалов (входов/выходов) и параметра. Синим цветом показаны метки терминалов, зеленым – метки параметров. Символы X обозначают некодирующие участки. Использован рисунок из [3].

ше похожи метки, тем сильнее связь. Общая схема приведена на рис. 10.5.

К. Маттиусси называет процесс определения весов более гордо: построение карты взаимодействия устройств (*device interaction map*).

Для определения степени схожести меток применяется локальное выравнивание, суть которого заключается в том, чтобы найти в двух строках максимально похожие подпоследовательности, по которым и будет осуществляться выравнивание.

Степень схожести определяется с использованием матриц счетов (*scoring matrices*), в которой заданы счета за замену, вставку и удаление для всех символов алфавита. Фактически это аналог редакторского расстояния. Счета за вставку и удаление объединены в один indel вектор. Пример матриц счетов, используемых в AGE, показан на рис. 10.6.

После того, как определяется расстояние между двумя подпоследовательностями, используется процедура квантования, которая отображает полученную целочисленную величину в заранее заданный диапазон $[0.001, 1000]$ для ИНС) и таким образом и высчитывается вес.

10.1.4 Взаимодействие со средой

Алгоритм AGE направлен на поиск структуры и весов графа, имеющего входы и выходы, т.е. взаимодействующего с окружающей средой (пусть даже условной). Входы и выходы также определяются как устройства, только они не изменяются, а мет-

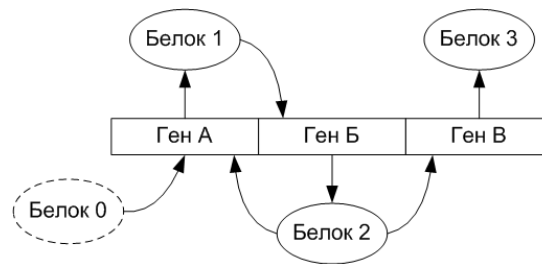


Рис. 10.3: Упрощенная схема кодирования информации в генных регуляторных сетях. Молекула ДНК содержит последовательность генов. Экспрессия гена А может привести к синтезу Белка 1, наличие которого, в свою очередь, может запустить экспрессию гена Б и синтез Белка 2, и так далее. При этом возможно появление цепных реакций, как в данном примере, когда Белок 2 стимулирует синтез Белка 1. Использован рисунок из [4].

ки терминалов этих устройств используются для определения силы взаимодействия полученной сети со средой. За взаимодействие отвечают специальные устройства преобразователи (*transducers*), которые . Пример представлен на рис. 10.7.

10.1.5 Операторы

Одна из особенностей алгоритма AGE заключается в большом количестве используемых операторов мутации. Практически для каждого задается своя вероятность «срабатывания». Операторы условно сгруппированы по «уровню» работы.

Мутации для отдельных нуклеотидов

- Вставка случайного символа в случайную позицию в хромосоме.
- Удаление случайного символа.
- Замена случайного символа на другой случайный символ.

Мутации для фрагментов хромосом

- Дублирование фрагмента. Случайный фрагмент хромосомы копируется в случайную точку случайной хромосомы генома.
- Дублирование фрагмента комплементарной хромосомы. То же, что и предыдущая мутация, только используется комплементарный (???) фрагмент.
- Транспозиция фрагмента. Для каждой хромосомы выбирается случайный фрагмент, который копируется в случайную точку случайно хромосомы особи
- Удаление фрагмента.
- Вставка устройства.

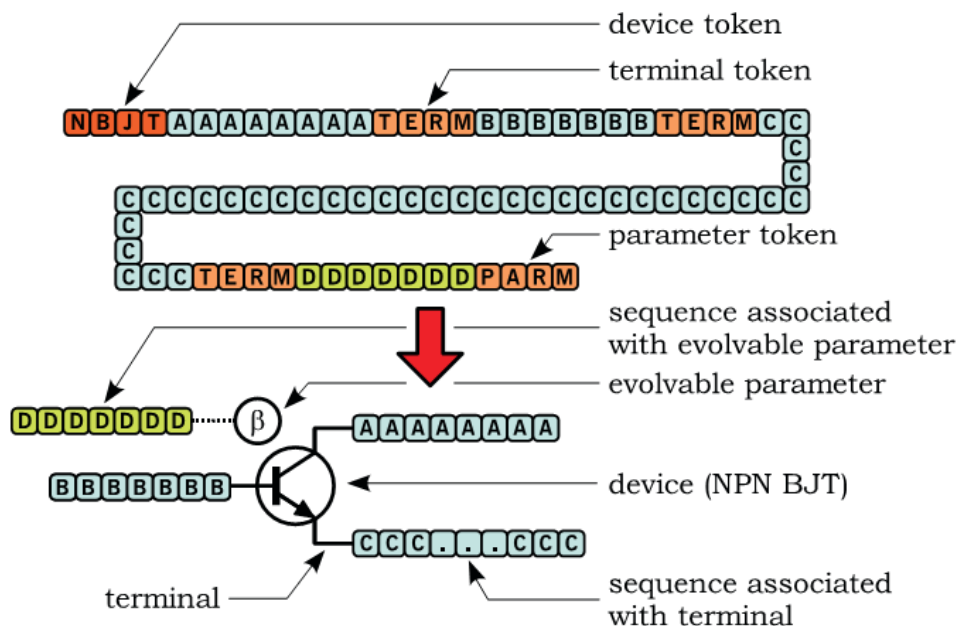


Рис. 10.4: Пример извлечения информации об устройстве из хромосомы в алгоритме AGE. Тип устройства определяется ведущим токеном. Использован рисунок из [3].

Мутации для всей хромосомы

- Дублирование хромосомы. В результате копия пристыковывается к существующей хромосоме, либо создается новая
- Удаление хромосомы.
- Кроссинговер. Если две особи имеют одинаковое количество хромосом, то производится «по-хромосомное» скрещивание. Особенности:
 - В первой хромосоме выбирается точка разрыва и запоминается участок вокруг этой точки
 - Во второй хромосоме ищется похожий участок.
 - Если такой участок найден, то производится обмен участками.
 - Биологически правдоподобная операция.
 - Если хромосомы сильно различаются, то скрещивания не будет.

Мутации для всего генома

- Дублирование.
- Обрезание (*trimming*). Удаление некодирующих участков. Применяется в каждом поколении ко всем геномам с вероятностью p_{gt}

10.1.6 Схема алгоритма

Обычный ГА с турнирной селекцией и элитаризмом :). Настройки тоже типичные, например, 30 особей, размер турнира = 2, используется 1 элитная особь.

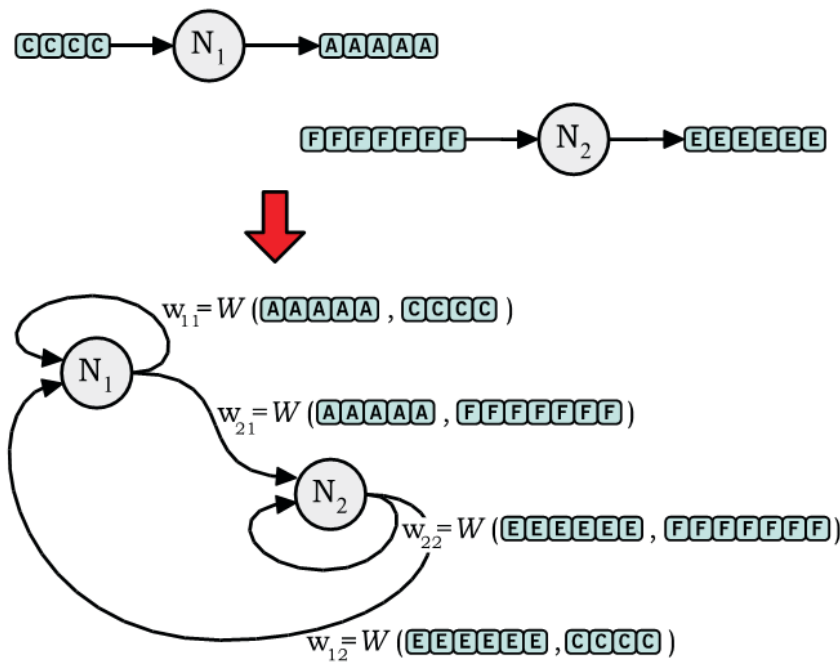


Рис. 10.5: Пример декодирования информации о связях между двумя нейронами. Использован рисунок из [3].

10.1.7 Результаты экспериментов

Маттиусси проведен большой объем исследований, касающихся особенностей работы алгоритма. Здесь будут освещены лишь некоторые результаты.

- Без мутаций, связанных с дублированием, алгоритм работает хуже. Более того, при включении дублирований после 4000 поколений в алгоритме, где дублирований не было, результаты значительно улучшились.
- XOR. Используется смещение в качестве одного из входов. Тоже получена ИНС с 1 скрытым нейроном, как и у Стенли.

10.1.8 Приложения

В силу того, что алгоритм был разработан для эволюционного поиска орграфов, спектр областей приложения довольно широк.

Синтез электрических цепей

В качестве устройств используются биполярный транзистор и полевой транзистор. Вес связи задает параметры резистора.

Рассматривались следующие задачи.

- Стабилизатор напряжения. Результат: 1.97-2.02 В при входном напряжении 4-6 В и диапазоне температур 0..100 С
- Температурный сенсор. Должен изменять напряжение на выходе (0..10В) линейно от входной температуры (0..100 градусов)

substitution matrix

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	5	2	1	0	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1	0	1	2
B	2	5	2	1	0	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1	0	1
C	1	2	5	2	1	0	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1	0
D	0	1	2	5	2	1	0	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1
E	-1	0	1	2	5	2	1	0	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2
F	-2	-1	0	1	2	5	2	1	0	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
G	-5	-2	-1	0	1	2	5	2	1	0	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
H	-5	-5	-2	-1	0	1	2	5	2	1	0	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
I	-5	-5	-5	-2	-1	0	1	2	5	2	1	0	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
J	-5	-5	-5	-5	-2	-1	0	1	2	5	2	1	0	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
K	-5	-5	-5	-5	-5	-2	-1	0	1	2	5	2	1	0	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5
L	-5	-5	-5	-5	-5	-5	-2	-1	0	1	2	5	2	1	0	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5
M	-5	-5	-5	-5	-5	-5	-5	-2	-1	0	1	2	5	2	1	0	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5
N	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1	0	1	2	5	2	1	0	-1	-2	-5	-5	-5	-5	-5	-5	-5
O	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1	0	1	2	5	2	1	0	-1	-2	-5	-5	-5	-5	-5	-5
P	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1	0	1	2	5	2	1	0	-1	-2	-5	-5	-5	-5	-5
Q	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1	0	1	2	5	2	1	0	-1	-2	-5	-5	-5	-5
R	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1	0	1	2	5	2	1	0	-1	-2	-5	-5	-5
S	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1	0	1	2	5	2	1	0	-1	-2	-5	-5
T	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1	0	1	2	5	2	1	0	-1	-2	-5
U	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1	0	1	2	5	2	1	0	-1	-2
V	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1	0	1	2	5	2	1	0	-1
W	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1	0	1	2	5	2	1	0
X	0	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1	0	1	2	5	2	1
Y	1	0	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1	0	1	2	5	2
Z	2	1	0	-1	-2	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-5	-2	-1	0	1	2	5

indel vector

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
-	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3	-3

Рис. 10.6: Пример декодирования информации о связях между двумя нейронами. Использован рисунок из [3].

- Гауссовский генератор тока. Генерация тока на выходе в форме гауссиана, если входное напряжение меняется в диапазоне 2-3 В.

Во всех рассматриваемых задачах результаты получились лучше, чем с использованием генетического программирования [2].

Нейронные сети

Для применения алгоритма AGE для поиска ИНС нейрон рассматривался в качестве устройства с входными выходными терминалами. Формирование ИНС производится через карту взаимодействий (т.е. стандартно).

В качестве задачи рассматривалась задача балансирования двух шестов без информации о скорости. Это один из самых сложных вариантов задачи. Результаты приведены в табл. 4.1. Видно, что алгоритм AGE превзошел другие алгоритмы с точки зрения количества вычислений целевой функции.

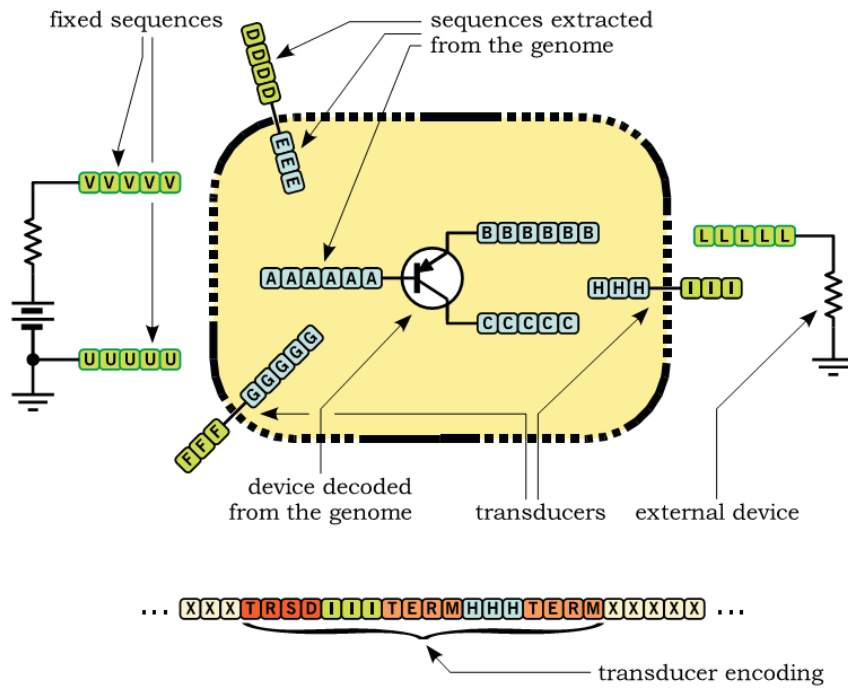


Рис. 10.7: Пример использования преобразователей для взаимодействия сети с входами и выходами. Использован рисунок из [3].

Алгоритм	Кол-во вычислений ЦФ	СКО	Обобщение
CE	840000	N/A	300
ESP	169466	N/A	289
NEAT	33184	21790	286
AGE	25065	19499	317

Таблица 4.1: Результаты решения задачи балансировки двух шестов без информации о скорости. Приведено среднее и СКО количества вычислений целевой функции, а также данные о способности алгоритмов к обобщению. Последняя отражает количество запусков (из 625) для обученной ИНС, в которых шесты удерживались не менее 1000 тактов при случайном начальном положении. Используются данные из [1]

Генные регуляторные сети

Алгоритм AGE стал победителем конкурса DREAM (Dialogue for Reverse Engineering Assessments and Methods), в котором требовалось восстановить структуру ГРС по данным экспрессии генов (обратная инженерия ГРС).

В общем и целом при решении этой задачи были следующие этапы

1. Сбор экспериментальных данных об экспрессии генов, включая информацию с использованием экспериментов (отключение некоторых генов)
2. Поиск сети, которая воспроизводит полученные результаты, с использованием AGE.

10.1.9 Замечания

- Используются суффиксные деревья для повышения быстродействия операций со строками.
- Очень много параметров для настройки.
- Время, которое тратится на декодирование решения и оценку качества для некоторых задач показано на рис. 10.8, 10.9. Использован компьютер на базе процессора Pentium IV 2.4 GHz.

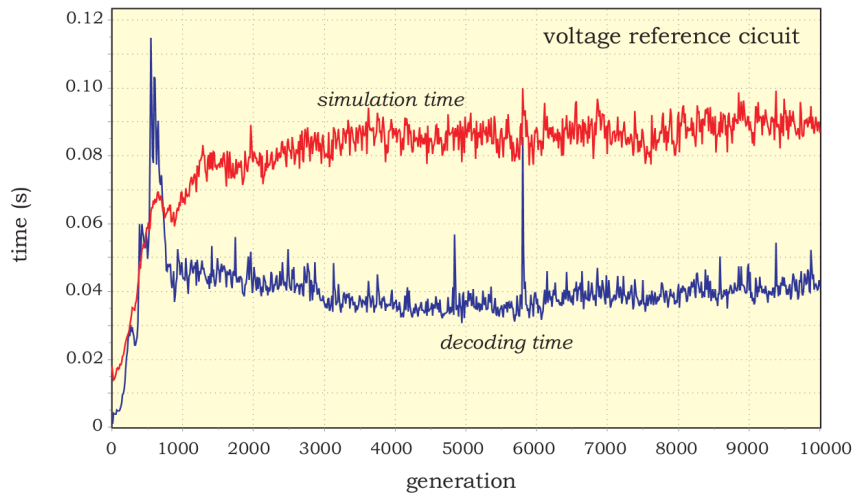


Рис. 10.8: Время декодирования и оценки качества решений для задачи синтеза стабилизатора напряжения. Использован рисунок из [3].

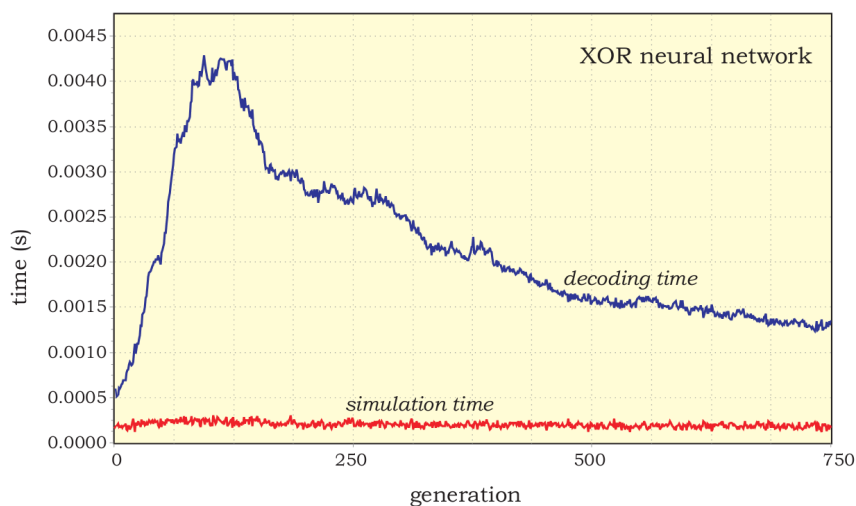


Рис. 10.9: Время декодирования и оценки качества решений для задачи XOR. Использован рисунок из [3].

Литература

- [1] P. Durr, C. Mattiussi, and D. Floreano. Neuroevolution with analog genetic encoding. In *Proceedings of the 9th Conference on Parallel Problem Solving from Nature (PPSN IX)*, pages 671–680, 2006.
- [2] J.R. Koza, F.H. Bennet III, D. Andre, and M.A. Keane. *Genetic Programming III*. Morgan Kaufmann, San Francisco, CA, 1999.
- [3] C. Mattiussi. *Evolutionary synthesis of analog networks*. Ph.d. thesis, Ecole Polytechnique Federale de Lausanne (EPFL), Lausanne, Switzerland, 2005.
- [4] Yury R. Tsoy. Computational regulatory networks and self-adaptive neuroevolutionary algorithm. In *11th Int. Conf. with Int. Participation on Artificial Intelligence (CAI-2008)*, volume 3, pages 50–57. Moscow: LENAND, 2008. In Russian.