

Дилемма стабильности-пластичности восприятия.

Проблема стабильности-пластичности является одной из самых сложных и трудно решаемых задач при построении искусственных систем, моделирующих восприятие. Характер восприятия внешнего мира живыми организмами (и, прежде всего, человеком) постоянно связан с решением дилеммы, является ли некоторый образ "новой" информацией, и следовательно реакция на него должна быть поисково-познавательной, с сохранением этого образа в памяти, либо этот образ является вариантом "старой", уже знакомой картиной, и в этом случае реакция организма должна соответствовать ранее накопленному опыту. Специальное запоминание этого образа в последнем случае не требуется. Таким образом, восприятие одновременно *пластично*, адаптировано к новой информации, и при этом оно *стабильно*, то есть не разрушает память о старых образах. Рассмотренные на предыдущих лекциях нейронные системы не приспособлены к решению этой задачи.

Так например, многослойный персептрон, обучающийся по методу обратного распространения, запоминает весь пакет обучающей информации, при этом образы обучающей выборки пред'являются в процессе обучения многократно. Попытки затем обучить персептрон новому образу приведут к модификации синаптических связей с неконтролируемым, вообще говоря, разрушением структуры памяти о предыдущих образах. Таким образом, персептрон не способен к запоминанию новой информации, необходимо полное переобучение сети.

Аналогичная ситуация имеет место и в сетях Кохонена и Липпмана-Хемминга, обучающихся на основе самоорганизации. Данные сети всегда выдают положительный результат при классификации. Тем самым, эти нейронные сети не в состоянии отделить новые образы от искаженных или зашумленных версий старых образов.

Принцип адаптивного резонанса.

Привлекательной особенностью нейронных сетей с адаптивным резонансом является то, что они сохраняют пластичность при запоминании новых образов, и, в то же время, предотвращают модификацию старой памяти. Нейросеть имеет внутренний детектор новизны - тест на сравнение пред'явленного образа с содержимым памяти. При удачном поиске в памяти пред'явленный образ классифицируется с одновременной уточняющей модификацией синаптических весов нейрона, выполнившего классификацию. О такой ситуации говорят, как о возникновении адаптивного резонанса в сети в ответ на пред'явление образа. Если резонанс не возникает в пределах некоторого заданного порогового уровня, то успешным считается тест новизны, и образ воспринимается сетью, как новый. Модификация весов нейронов, не испытавших резонанса, при этом не производится.

Важным понятием в теории адаптивного резонанса является так называемый **шаблон критических черт** (critical feature pattern) информации. Этот термин показывает, что не все черты (детали), представленные в некотором образе, являются существенными для системы восприятия. Результат распознавания определяется присутствием специфических критических особенностей в образе. Рассмотрим это на примере.

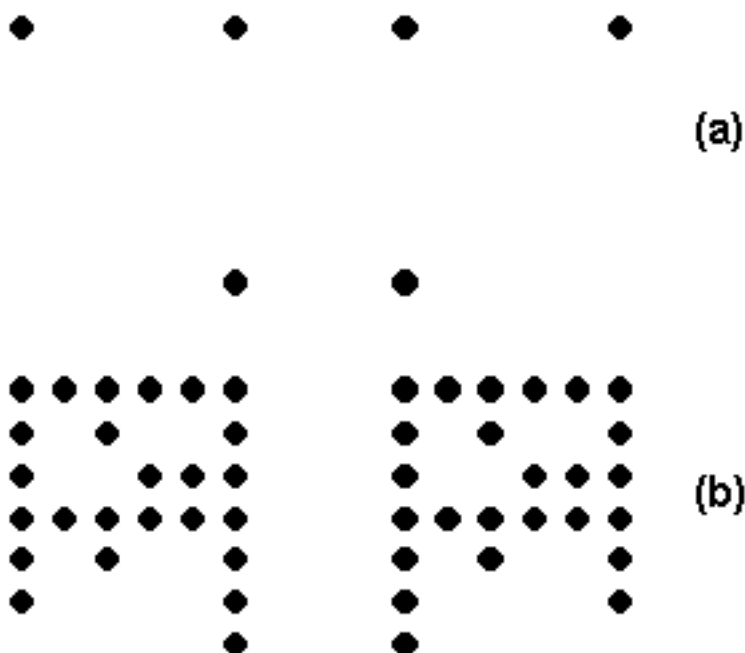


Рис. 11.1. Иллюстрация к понятию критических черт образа.

Обе пары картинок на Рис. 11.1 имеют общее свойство: в каждой из пар черная точка в правом нижнем углу заменена на белую, а белая точка левом нижнем углу - на черную. Такое изменение для нижней пары картинок (на рисунке - пара (b)), очевидно, является не более чем шумом, и оба образа (b) являются искаженными версиями одного и того же изображения. Тем самым, измененные точки не являются для этого образа критическими. Совершенно иная ситуация имеет место для верхней пары картинок (a). Здесь такое же изменение точек оказывается слишком существенным для образа, так что правая и левая картинки являются различными образами. Следовательно, одна и та же черта образа может быть не существенной в одном случае, и критической в другом. Задачей нейронной сети будет формирование правильной реакции в обоих случаях: "пластичное" решение о появлении нового образа для пары (a) и "стабильное" решение о совпадении картинок (b). При этом выделение критической части информации должно получаться автоматически в процессе работы и обучения сети, на основе ее индивидуального опыта.

Отметим, что в общем случае одного лишь перечисления черт (даже если его предварительно выполнит человек, предполагая определенные условия дальнейшей работы сети) может оказаться недостаточно для успешного функционирования искусственной нейронной системы, критическими могут оказаться специфические связи между несколькими отдельными чертами.

Упрощенная архитектура АРТ

На рис. 11.2 показана упрощенная конфигурация сети АРТ, представленная в виде пяти функциональных модулей. Она включает два слоя нейронов — так называемые "слой сравнения" и "слой распознавания". Приемник 1, Приемник 2 и Сброс обеспечивают управляющие функции, необходимые для обучения и классификации. Перед рассмотрением вопросов функционирования сети в целом необходимо рассмотреть отдельно назначения модулей; далее обсуждаются функции каждого из них.

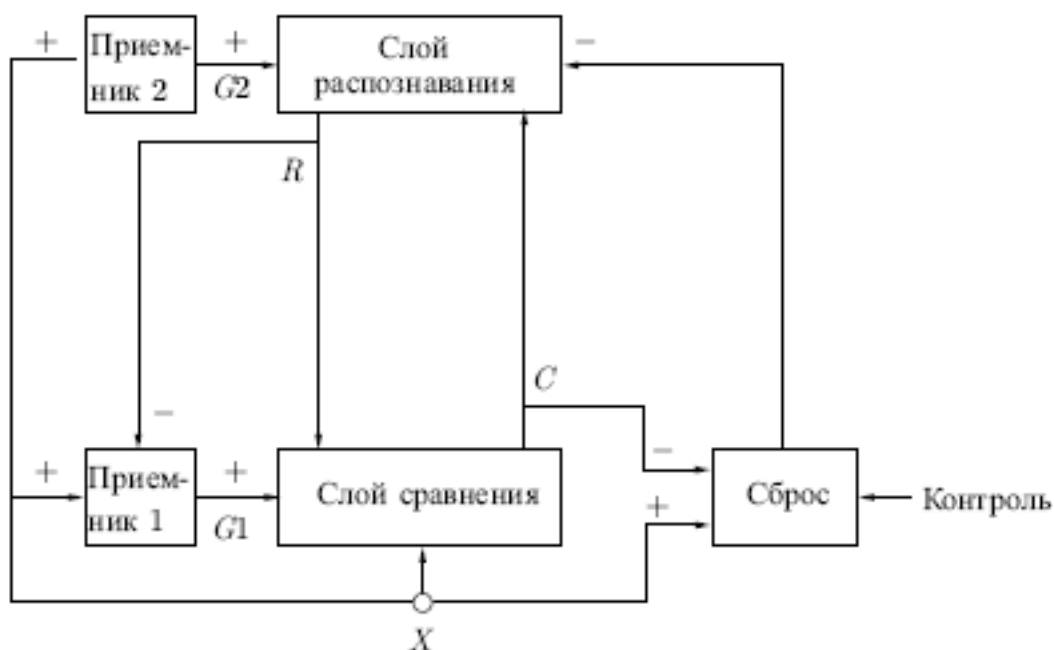


Рис. 11.2.

Слой сравнения. Слой сравнения получает двоичный входной вектор X и первоначально пропускает его неизменным для формирования выходного вектора C . На более поздней фазе в распознающем слое вырабатывается двоичный вектор R , модифицирующий вектор C , как описано ниже.

Каждый нейрон в слое сравнения (см. рис. 11.3) получает три двоичных входа (0 или 1): (1) компонента x_i входного вектора X ; (2) сигнал обратной связи R_i — взвешенная сумма выходов распознающего слоя; (3) вход от Приемника 1 (один и тот же сигнал подается на все нейроны этого слоя).

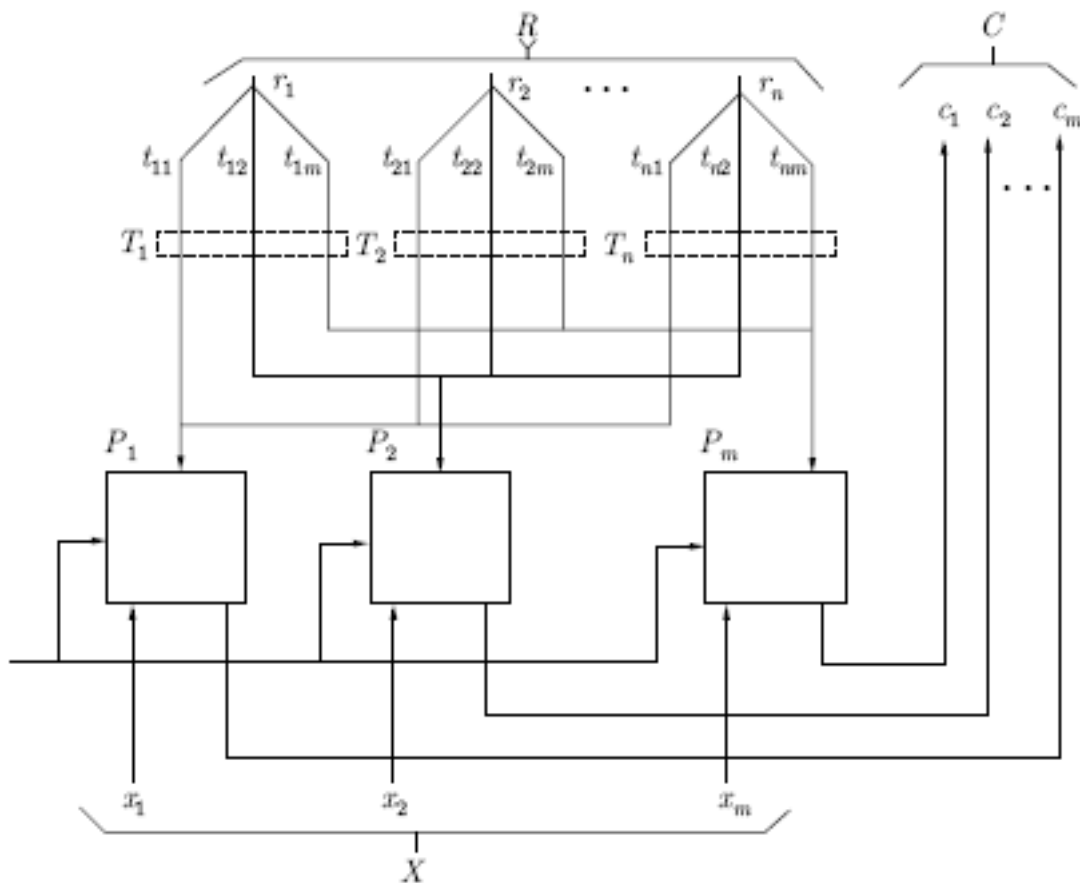


Рис. 11.3.

Чтобы получить на выходе нейрона единичное значение, как минимум два из трех его входов должны равняться единице; в противном случае его выход будет нулевым. Таким образом, реализуется правило двух третей. Первоначально выходной

сигнал $G1$ Приемника 1 установлен в единицу, обеспечивая один из входов, необходимых для возбуждения нейронов, а все компоненты вектора R установлены в 0; следовательно, в этот момент вектор C идентичен двоичному входному вектору X .

Слой распознавания. Слой распознавания осуществляет классификацию входных векторов. Каждый нейрон в слое распознавания имеет соответствующий вектор

весов B_j . Только один нейрон с весовым вектором, наиболее соответствующим входному вектору, возбуждается; все остальные заторможены.

Как показано на рис. 11.4, нейрон в распознающем слое имеет максимальную реакцию, если вектор C , являющийся выходом слоя сравнения, соответствует набору его весов; следовательно, веса представляют запомненный образ или экземпляр для категории входных векторов. Такие веса являются действительными числами, а не двоичными величинами. Двоичная версия этого образа также запоминается в соответствующем наборе весов слоя сравнения (рис. 11.3); этот набор состоит из весов связей, соединяющих определенные нейроны слоя распознавания, по одному весу на каждый нейрон слоя сравнения.

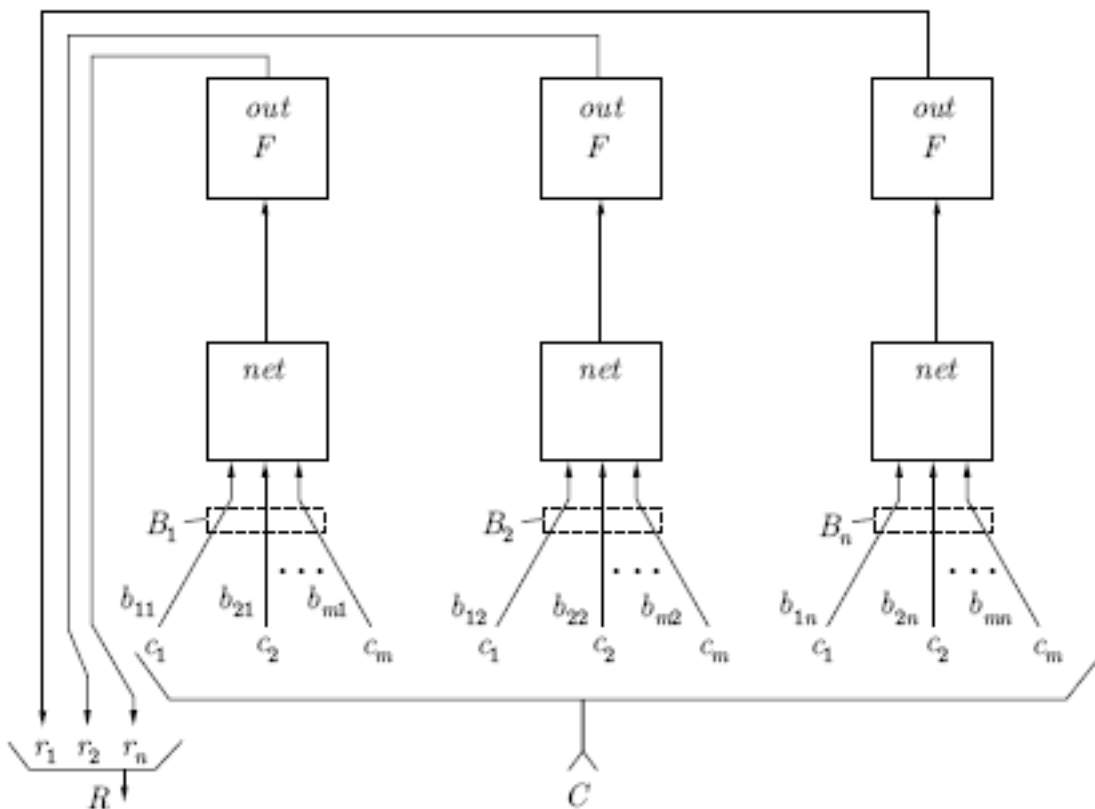


Рис. 11.4.

В процессе функционирования каждый нейрон слоя распознавания вычисляет свертку вектора собственных весов и входного вектора C . Нейрон, веса которого наиболее близки вектору C , будет иметь самый большой выход, тем самым выигрывая соревнование и одновременно затормаживая все остальные нейроны в слое. Как показано на рис. 11.5, нейроны внутри слоя распознавания взаимно соединены в латерально-тормозящую сеть. В простейшем случае (единственном, рассмотренном в данной работе) предусматривается, что только один нейрон в слое возбуждается в каждый момент времени (т. е. только нейрон с наивысшим уровнем активации будет иметь единичный выход; все остальные нейроны будут иметь нулевой выход). Эта конкуренция

реализуется введением связей с отрицательными весами l_{ij} с выхода каждого нейрона r_i на входы остальных нейронов. Таким образом, если нейрон имеет большой выход, он тормозит все остальные нейроны в слое. Кроме того, каждый нейрон имеет связь с положительным весом со своего выхода на свой вход. Если нейрон имеет единичный выходной уровень, эта обратная связь стремится усилить и поддержать его.

Приемник 2. $G2$, выход Приемника 2, равен единице, если входной вектор X имеет хотя бы одну единичную компоненту. Более точно, $G2$ является логическим ИЛИ от компонента вектора X .

Приемник 1. Как и сигнал $G2$, выходной сигнал $G1$ Приемника 1 равен 1, если хотя бы одна компонента двоичного входного вектора X равна единице; однако, если хотя бы одна компонента вектора R равна единице, $G1$ устанавливается в нуль. Таблица, определяющая эти соотношения:

ИЛИ от компонента вектора X	ИЛИ от компонента вектора R	G1
0	0	0
1	0	1
1	1	0
0	1	0

Сброс. Модуль сброса измеряет сходство между векторами X и C . Если они отличаются сильнее, чем требует параметр сходства, вырабатывается сигнал сброса возбужденного нейрона в слое распознавания.

В процессе функционирования модуль сброса вычисляет сходство как отношение количества единиц в векторе X к их количеству в векторе C . Если это отношение ниже значения параметра сходства, вырабатывается сигнал сброса.

Структура нейронной сети АРТ-1

В отличие от большинства структур других нейронных сетей входы АРТ-1 связаны с выходами не одним, а парой синапсов: w_{ij} – синапсы кратковременной памяти; t_{ij} – синапсы долговременной памяти. Назначение кратковременной памяти – установление кластеров, к которым может быть отнесен входной образ. Назначение долговременной памяти – установление степени соответствия входного образа кластерам, определение нейрона-победителя или принятие решения о создании нового кластера.

Алгоритм функционирования АРТ-1 включает 3 стадии жизненного цикла: инициализацию сети, распознавание образа, кластеризацию образа (сравнение). На стадии **инициализации** последовательно выполняются следующие этапы:

1.1. Устанавливается параметр сходства $R_{кр}$ ($0 < R_{кр} < 1$), причем чем он больше, тем выше должно быть сходство образа и кластера для отнесения к последнему.

1.2. Для первого поданного на входы сети образа создается первый нейрон (кластер), значения синаптических связей которого устанавливаются согласно следующим соотношениям:

$$w_{i1} = \frac{\lambda \cdot x_i}{\lambda - 1 + \sum_{p=1}^M x_p}, \quad (1)$$

$$t_{i1} = x_i, \quad (2)$$

где λ – положительная ($> 1,0$) константа, определяющая степень влияния нового входного образа на кратковременную память. Чаще всего λ принимается равной 2,0.

Стадия **распознавания образа** включает следующие этапы:

2.1. На входы сети подается новый входной образ. Для каждого нейрона рассчитывается значение его выхода с учетом коэффициентов кратковременной памяти:

$$y_j = \sum_{i=1}^n w_{ij} x_i. \quad (3)$$

2.2. Положительные выходы нейронов указывают на кластеры, имеющие качественное сходство с входным образом. Если же все выходы оказались нулевыми, входной образ не соответствует ни одному из кластеров, создается новый нейрон с синаптическими коэффициентами, рассчитываемыми по соотношениям (1) и (2), после чего алгоритм продолжает работу с п. 2.1.

Количественное сходство входного образа с кластерами определяется на стадии **кластеризации образа (сравнения)**:

3.1. Рассчитывается количественная мера сходства входного образа с кластером, имеющим наибольшее значение выхода нейрона в п. 2.1:

$$R_j = \frac{\sum_{i=1}^M t_{ij} x_i}{\sum_{i=1}^M x_i} \quad (4)$$

3.2. Если выполняется условие $R_j > R_{кр}$, j -й нейрон считается нейроном-победителем, а входной образ – соответствующим j -му кластеру. В этом случае для нейрона выполняется пересчет весовых коэффициентов по соотношениям:

$$w_{ij}^{(q+1)} = (1 - v) \cdot w_{ij}^{(q)} + v \frac{\lambda \cdot x_i}{\lambda - 1 + \sum_{p=1}^M x_p} \quad (5)$$

$$t_{ij}^{(q+1)} = (1 - v) \cdot t_{ij}^{(q)} + v \cdot x_i \quad (6)$$

где v – коэффициент скорости адаптации ($0 < v < 1$). Большее значение v соответствует быстрой адаптации, меньшее – медленной.

3.3. Если условие $R_j > R_{кр}$ не выполняется, алгоритм повторяется с п. 3.1 для остальных кластеров, следующие в порядке убывания их выходных значений, рассчитанных в п. 2.1, до тех пор, пока не будет получен нейрон-победитель или не будут использованы все кластеры с положительным выходом нейрона.

3.4. Если нейрон-победитель не найден, создается новый кластер с синаптическими коэффициентами нейрона, рассчитанными по соотношениям (1) и (2), после чего алгоритм продолжает работу с п. 2.1.

Рассмотрим пример работы сети АРТ-1 на протяжении всего жизненного цикла.

Последовательно подадим на ее входы 4 бинарных вектора, состоящих из 9 элементов, кодирующих различные образы. Графическая и цифровая интерпретации образов приведены на рис. 2.

Очевидно, что структура нейронной сети для кластеризации указанных образов будет включать 9 входов. Выберем следующие параметры настройки нейронной сети: $R_{кр} = 0,7$, $\lambda = 2,0$, $v = 0,6$. Проведем расчеты в соответствии с алгоритмом работы АРТ-1.

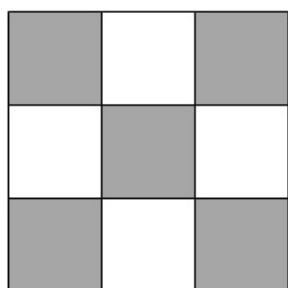
На стадии инициализации подаем на входы сети образ 1. Он формирует первый нейрон, весовые коэффициенты которого рассчитываются по соотношениям (1),

(2): $w_{11} = w_{31} = w_{51} = w_{71} = w_{91} = 0,33$; $w_{21} = w_{41} = w_{61} = w_{81} = 0,00$; $t_{11} = t_{31} = t_{51} = t_{71} = t_{91} = 1,00$; $t_{21} = t_{41} = t_{61} = t_{81} = 0,00$.

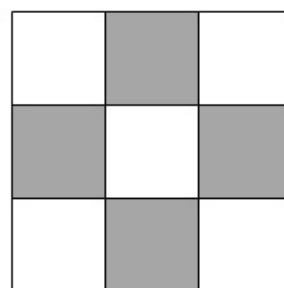
Далее работа с каждым образом начинается со стадии распознавания образа. Подаем на входы сети образ 2. Выходное значение единственного существующего на данный момент нейрона, рассчитанное по соотношению (12.3), составит: $y_1 = 0,00$. Таким образом, образ 2 оказался не соответствующим первому кластеру, и для него создается новый нейрон с весовыми коэффициентами, рассчитанными по соотношениям (12.1),

(12.2): $w_{12} = w_{32} = w_{52} = w_{72} = w_{92} = 0,00$; $w_{22} = w_{42} = w_{62} = w_{82} = 0,40$; $t_{12} = t_{32} = t_{52} = t_{72} = t_{92} = 0,00$; $t_{22} = t_{42} = t_{62} = t_{82} = 1,00$.

Образ 1

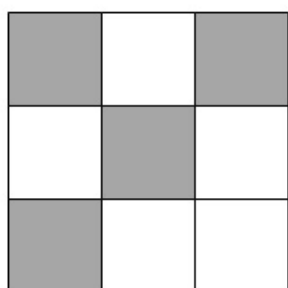

 $[1, 0, 1, 0, 1, 0, 1, 0, 1]$

Образ 2


 $[0, 1, 0, 1, 0, 1, 0, 1, 0]$

Образ 3

(зашумленный образ 1)


 $[1, 0, 1, 0, 1, 0, 1, 0, 0]$

Образ 4

(зашумленный образ 2)

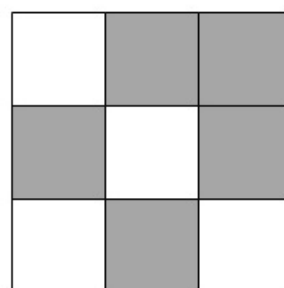

 $[0, 1, 1, 1, 0, 1, 0, 1, 0]$

Рис. 2. Набор входных образов для кластеризации с помощью сети АРТ-1

Подаем на входы нейронной сети образ 3. Выходные значения имеющихся двух нейронов равняются соответственно: $y_1 = 1,33$, $y_2 = 0,00$. Очевидно наличие возможного соответствия только образу, ранее сформировавшему первый кластер. Проверим сходство по критерию (4). В результате расчета $R_1 = 1,0$, что больше заданного значения $R_{кр}$. Таким образом, соответствие образа 3 первому кластеру считаем установленным. В соответствии с алгоритмом стадии кластеризации образа далее по соотношениям (5) и (6) пересчитываются коэффициенты кратковременной и долговременной памяти первого нейрона. Их новые значения составят: $w_{11} = w_{31} = w_{51} = w_{71} = 0,37$; $w_{91} = 0,13$; $w_{21} = w_{41} = w_{61} = w_{81} = 0,00$; $t_{11} = t_{31} = t_{51} = t_{71} = 1,00$; $t_{91} = 0,40$; $t_{21} = t_{41} = t_{61} = t_{81} = 0,00$.

Подаем на входы нейронной сети образ 4. Выходные значения нейронов составят соответственно: $y_1 = 0,37$, $y_2 = 1,60$. Это говорит о том, что в разной степени возможно сходство с обоими образами, однако со вторым оно гораздо вероятнее. Следовательно, по соотношению (4) рассчитываем значение критерия: $R_2 = 0,8$. Оно больше заданного критического значения, а значит образ 4 должен быть отнесен ко второму кластеру. По соотношениям (5) и (6) пересчитываются синаптические коэффициенты второго нейрона, в результате чего

получим: $w_{12} = w_{52} = w_{72} = w_{92} = 0,00$; $w_{32} = 0,20$; $w_{22} = w_{42} = w_{62} = w_{82} = 0,36$; $t_{12} = t_{52} = t_{72} = t_{92} = 0,00$; $t_{32} = 0,60$; $t_{22} = t_{42} = t_{62} = t_{82} = 1,00$.

Теоремы АРТ.

1. По достижении стабильного состояния обучения пред'явление одного из обучающих векторов будет сразу приводить к правильной классификации без фазы поиска, на основе прямого доступа.
2. Процесс поиска устойчив.
3. Процесс обучения устойчив. Обучение весов нейрона-победителя не приведет в дальнейшем к переключению на другой нейрон.
4. Процесс обучения конечен. Обученное состояние для заданного набора образов будет достигнуто за конечное число итерации, при этом дальнейшее пред'явление этих образов не вызовет циклических изменений значений весов.

Нерешенные проблемы и недостатки АРТ-1.

Нейронные сети АРТ, при всех их замечательных свойствах, имеют ряд недостатков.

Одним из них является большое количество синаптических связей в сети, в расчете на единицу запоминаемой информации. При этом многие из весов этих связей (например, ветвора Т) оказываются после обучения нулевыми. Эту особенность следует учитывать при аппаратных реализациях.

Сеть АРТ-1 приспособлена к работе только с битовыми векторами. Это неудобство преодолевается в сетях АРТ-2 и АРТ-3. Однако в этих архитектурах, равно как и в АРТ-1, сохраняется главный недостаток АРТ - локализованность памяти. Память нейросети АРТ не является распределенной, некоторой заданной категории отвечает вполне конкретный нейрон слоя распознавания. При его разрушении теряется память обо всей категории. Эта особенность, увы, не позволяет говорить о сетях адаптивной резонансной теории, как о прямых моделях биологических нейронных сетей. Память последних является распределенной.