

УДК 025.4; 004.8.032.26

Г.А. Прокопович

## АДАПТИВНЫЙ НЕЙРОСЕТЕВОЙ КЛАССИФИКАТОР

*Рассматривается проблема затирания памяти в искусственных нейронных сетях при предъявлении новых классов образов. Предлагается адаптивный гибридный классификатор, обладающий способностью к дообучению. Отличительной особенностью описываемого классификатора является обучение с учителем и наличие ассоциативного поиска сохраненных образов в долговременной памяти, который повышает устойчивость классификатора к шумам и уменьшает количество обращений к памяти по сравнению с последовательным способом.*

### Введение

Процесс классификации является неотъемлемой частью нормальной жизнедеятельности любого живого организма [1]. Он также необходим для функционирования множества технических устройств: от простых автоматов включения уличного освещения до самых сложных систем ориентации автономных роботов в пространстве. В теории машинного обучения задача классификации решается, как правило, с помощью методов искусственных нейронных сетей (ИНС).

Однако такие широко используемые ИНС, как многослойные перцептроны, сети Хопфилда, самоорганизующиеся карты, оказались не в состоянии решить проблему стабильного и пластичного дообучения с целью распознавания новых образов [2, 3]. Очень часто обучение новому образу уничтожает или изменяет результаты предшествующего обучения.

### 1. Решение проблемы затирания памяти

В некоторых случаях процесс затирания старой информации является несущественным. Если набор обучающих векторов фиксирован, то процесс обучения многослойного перцептрона по методу обратного распространения ошибки заключается в многократном циклическом предъявлении обучающей выборки. В итоге он запоминает весь пакет обучающей информации. Однако попытки обучить затем перцептрон новому образу приводят к модификации синаптических связей с неконтролируемым разрушением структуры памяти о предыдущих образах. Таким образом, перцептрон не способен к запоминанию новой информации, поэтому необходимо полное переобучение сети [4].

Аналогичная ситуация имеет место и в сетях Кохонена и Хемминга [2, 5], обучающихся на основе принципов самоорганизации. Данные ИНС всегда выдают положительный результат при классификации. Таким образом, эти нейронные сети не в состоянии отделить новые образы от искаженных или зашумленных версий старых образов. В реальной ситуации сеть будет подвергаться постоянно изменяющимся воздействиям, она может никогда не увидеть один и тот же обучающий вектор дважды. При таких обстоятельствах сеть, скорее всего, не будет обучаться; она будет непрерывно изменять свои веса, не достигая удовлетворительных результатов.

В противоположность этому ежесекундно мозг человека выполняет трудную задачу обработки непрерывного потока сенсорной информации, получаемой из окружающего мира. Из огромного объема тривиальных данных ему приходится выделять и обрабатывать жизненно важную информацию и, возможно, регистрировать в долговременной памяти. Однако новые образы запоминаются в такой форме, что ранее запомненные не модифицируются и не забываются [6]. Понимание сути этого процесса представляет собой серьезную задачу для исследователей: решить, каким образом память остается пластичной, способной к восприятию новых образов и в то же время сохраняет стабильность, гарантирующую, что образы не уничтожатся и не разрушатся в процессе функционирования.

Проблема стабильности-пластичности является одной из самых сложных и трудно решаемых задач при построении искусственных систем, моделирующих восприятие и распознавание образов. Способ восприятия внешнего мира живыми организмами, и прежде всего чело-

веком, состоит в постоянном сравнении и оценке поступающих образов. Если некоторый образ является «новой» информацией, то реакция на него должна быть поисково-познавательной, с сохранением этого образа в памяти. Если этот образ является вариантом «старой», уже знакомой картины, то в этом случае реакция организма должна соответствовать ранее накопленному опыту и специальное запоминание образа в последнем случае не требуется. Таким образом, восприятие одновременно пластично, адаптировано к новой информации и при этом стабильно, т. е. не разрушает память о старых образах.

Один из методов, преодолевающих эту дилемму, предлагает адаптивная резонансная теория С. Гроссберга, в которой постулируется необходимость *самоадаптации* алгоритма сравнения новых образов с уже существующими в памяти. В основе этой теории лежит использование слоя распознавания «снизу вверх» и слоя сравнения «сверху вниз». Если входной и изученный образы совпадают, возникает состояние, называемое адаптивным резонансом (т. е. усилением и продлением нейронной активности) [5]. Считается, что резонанс возникает при идентификации какого-либо образа или события. В процессе функционирования классификаторов, основанных на данной теории, происходит циркуляция информации до тех пор, пока не наступит состояние резонанса. Отличительной особенностью таких классификаторов является то, что они сохраняют пластичность при запоминании новых образов и в то же время предотвращают модификацию старой памяти [2, 3].

В настоящей статье используются векторно-матричные обозначения, принятые в монографиях [7, 8].

## 2. Искусственные нейронные сети адаптивного резонанса

Используя предложенную теорию адаптивного резонанса, различными исследователями было предложено множество реализаций указанного метода. Их можно разделить по способу реализации на две группы: с помощью логических операций и преобразований [3] и с помощью ИНС [2, 5].

К последним относится нейроподобная сеть С. Гроссберга АРТ-1 [2, 9]. Она предназначена для обработки бинарных векторов и состоит из двух слоев нейронных элементов: слоя сравнения  $F_1$  и слоя распознавания  $F_2$  (рис. 1). Блоки  $G_1$  и  $G_2$  представляют собой управляющие триггеры, которые разделяют процесс работы сети на три фазы: поиск, сравнение и добавление нового образа.

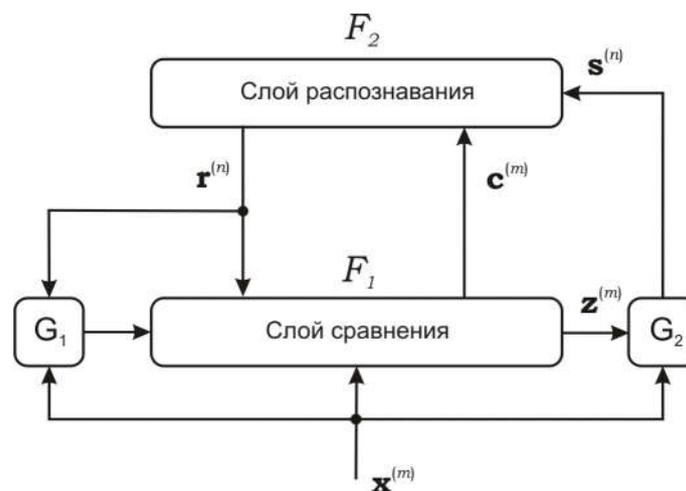


Рис. 1. Архитектура сети АРТ-1

Входной двоичный вектор  $\mathbf{x}^{(m)} = [x_1, x_2, \dots, x_m]$ , где  $m$  – размерность входного вектора, классифицируется в зависимости от того, на какой образ из множества запомненных ранее, он похож. Свое классификационное решение сеть АРТ-1 выражает в форме возбуждения одного из нейронов распознающего слоя  $F_2$ .

Вначале, когда сеть находится в невозбужденном состоянии, блок  $G_1$  подает на слой  $F_1$  единицу (которая удерживает его в ожидающем состоянии), а блок  $G_2$  подает на слой  $F_2$  нуль (тем самым затормаживает весь слой). Если на слой сравнения поступит ненулевой входной вектор  $\mathbf{x}^{(m)}$ , то первоначально он пропустит его неизменным, формируя выходной вектор  $\mathbf{c}^{(m)} = [c_1, c_2, \dots, c_m]$ , который непосредственно идет на входы слоя распознавания  $F_2$  (рис. 2). Слой  $F_2$  содержит  $n$  нейронных элементов, где число  $n$  соответствует количеству образов, хранящихся в долговременной памяти. Каждый  $j$ -й нейрон слоя  $F_2$ , характеризующий определенный кластер образов, через соответствующие компоненты входного вектора  $\mathbf{c}^{(m)}$  имеет прямые синаптические связи со всеми нейронами слоя  $F_1$ , которые характеризуются весовыми коэффициентами вектора  $\mathbf{w}_j^{(m)} = [w_{j,1}, w_{j,2}, \dots, w_{j,m}]^T$ . Коэффициенты весовой матрицы  $W_{m \times n} = [\mathbf{w}_1^{(m)}, \mathbf{w}_2^{(m)}, \dots, \mathbf{w}_n^{(m)}]$  являются носителями долговременной памяти.

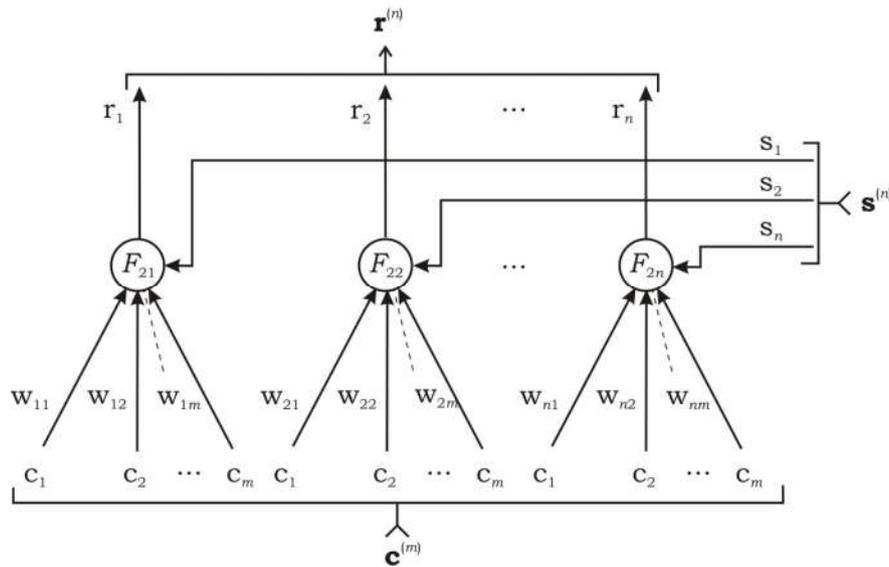


Рис. 2. Структура слоя распознавания  $F_2$

В начальном состоянии все компоненты вектора  $\mathbf{s}^{(n)} = [s_1, s_2, \dots, s_n]$ , который является выходом управляющего блока  $G_2$ , равны 1, тогда в процессе классификации входного вектора  $\mathbf{x}^{(m)}$  участвуют все нейроны распознающего слоя. Нейронные элементы данного слоя функционируют по принципу «победитель забирает все». Победителем является нейрон с индексом  $k$ , который соответствует максимальному значению свертки вектора  $\mathbf{c}^{(m)}$  и матрицы весов  $W_{m \times n}$ :

$$k = \max \left\{ f \left( \mathbf{c}^{(m)} W_{m \times n} \right) \right\},$$

где  $f(x)$  – функция активации. Таким образом, нейрон  $F_{2k}$  возбуждается, а все остальные затормаживаются. При этом соответствующий элемент  $r_k$  выходного вектора  $\mathbf{r}^{(n)} = [r_1, r_2, \dots, r_n]$  становится равным единице, а все остальные элементы по-прежнему остаются равными нулю. Однако этого достаточно, чтобы выходной сигнал управляющего блока  $G_1$  принял нулевое значение.

Соответствующий сигнал  $r_k = 1$  выходного нейрона-победителя через обратные синаптические связи  $V_{n \times m} = [v_1^{(n)}, v_2^{(n)}, \dots, v_m^{(n)}]$  поступает на нейроны слоя  $F_1$ . В результате формируется интегрированный образ активности  $\mathbf{v}_k^{(n)} = [v_{k,1}, v_{k,2}, \dots, v_{k,n}]^T$ , который соответствует кластеру нейрона-победителя (рис. 3).

Каждый из логических элементов  $P_1, P_2, \dots, P_m$  в слое сравнения получает три двоичных входа: значение элемента  $x_i$  входного вектора  $\mathbf{x}^{(m)}$ , сигнал обратной связи  $r_i$  и управляющий сигнал от блока  $G_1$ , которые принимают значения 0 или 1. Чтобы получить на выходе нейрона  $F_{1i}$  слоя сравнения 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» [2].

При положительном результате поиска в памяти предъявленный вектор  $\mathbf{x}^{(m)}$  классифицируется с одновременной уточняющей модификацией вектора  $\mathbf{w}_k^{(m)}$  синаптических весов нейрона  $F_{2k}$ , выполнившего классификацию. Другими словами, если было установлено, что входной вектор похож на один из ранее запомненных с точки зрения заданного критерия сходства, определяемого элементами вектора  $\mathbf{z}^{(m)}$  в управляющем блоке  $G_2$ , то запомненный в долговременной памяти вектор будет изменяться под воздействием нового вектора  $\mathbf{x}^{(m)}$ . Следовательно, образ в долговременной памяти становится более похожим на этот входной вектор. Такую ситуацию называют возникновением адаптивного резонанса в сети в ответ на предъявление образа.

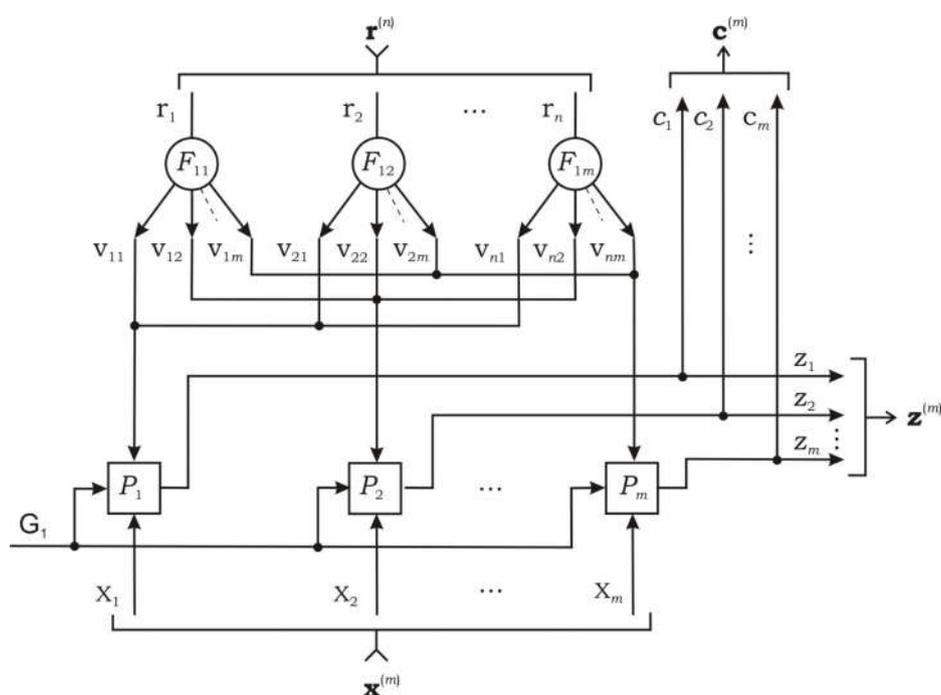


Рис. 3. Структура слоя сравнения  $F_1$

Если результат поиска в пределах некоторого заданного порогового уровня окажется отрицательным, т. е. входной вектор не соответствует ни одному из запомненных образов, тогда резонанс не возникнет и образ воспринимается сетью как новый [2, 7]. В этом случае в слое распознавания  $F_2$  будет создан новый нейрон, в синаптических весах которого будет содержаться образ, идентичный новому входному вектору. Предполагается, что модификация весов нейронов, не испытавших резонанса, при этом не производится.

Запомненный образ не будет изменяться, если текущий входной вектор не окажется достаточно похожим на него. Таким образом, решается дилемма стабильности-пластичности. Новый образ может создавать дополнительные классификационные категории, однако он не может изменить существующую память.

### 3. Принцип работы двунаправленной ассоциативной памяти

Несмотря на то что теория адаптивного резонанса предлагает эффективное решение проблемы стабильно-пластичной организации памяти в ИНС, ее реализация не лишена недостатков.

Нейронные сети адаптивного резонанса характеризуются самоорганизацией в процессе работы и обучаются без учителя. Поэтому без дополнительных изменений в архитектуре невозможно влиять на процесс объединения сетью различных образов в кластеры. В табл. 1 показан процесс классификации входных образов сетью ART-1. Здесь и далее приведены результаты работы адаптивной сети для дискретного случая; сеть реализована на основе алгоритма, описанного в монографии [9].

Из табл. 1 видно, что у одних и тех же букв алфавита, нарисованных разными шрифтами, может не совпадать критерий сходства; следовательно, они распознаются сетью как разные классы. По этой же причине нейронные сети с самоорганизующейся структурой подвержены влиянию шумов, содержащихся во входных образах. Зашумленные образы могут спровоцировать создание новых классов в долговременной памяти сети (табл. 2).

Как видно из приведенных примеров, для нахождения соответствующего кластера в некоторых случаях требуется последовательный поиск среди всех запомненных образов в долговременной памяти, что может привести к значительным затратам машинного времени. Тогда на основе приведенных данных можно сделать вывод, что классическая реализация ART-сетей является неприемлемой для использования в таких системах реального времени, как автономные интеллектуальные роботы и беспилотные летательные аппараты, в которых постоянный процесс классификации является одним из основных показателей эффективности.

Таблица 1  
Классификация набора изображений с помощью сети ART-1

Входной образ	Количество итераций		Присвоенный класс
	при обучении	при распознавании	
	1	1	2
	3	1	1
	4	1	3
	3	7	4
	4	6	5
	3	1	5
	4	1	5
	3	1	6
	4	3	6

Автором была предпринята попытка модернизировать сеть ART-1 таким образом, чтобы уменьшить время поиска сохраненных образов в долговременной памяти. Среди направлений модернизации были выбраны следующие: последовательный поиск сохраненных образов должен быть заменен на ассоциативный, решение о принадлежности образа к тому или иному классу должен принимать не один, а группа нейронов.

После проведенного анализа среди большого числа известных ИНС была выбрана архитектура двунаправленной ассоциативной памяти (ДАП), которая как нельзя лучше соответствует структуре сетей адаптивного резонанса (может легко в нее интегрироваться), а также обладает следующими дополнительными преимуществами:

- реализация ассоциативного способа обращения к памяти;
- наличие двух ассоциативных слоев;
- обучение с учителем;
- устойчивость к шумам;
- достаточно быстродействующий способ обучения.

Сеть ДАП, как и широко известная сеть Хопфилда, способна к обобщению, вырабатывая правильные реакции, несмотря на искаженные входы. Однако последняя является *автоассоциативной*. Это означает, что образ может быть завершен или исправлен, но не может быть ассоциирован с другим образом [10]. Данный факт является результатом одноуровневой структуры ассоциативной памяти, в которой вектор появляется на выходе тех же нейронов, на которые поступает входной вектор.

Таблица 2  
Классификация набора зашумленных изображений с помощью сети АРТ-1

Шум, %	0	5	10	20	30
Входной образ					
Количество итераций при распознавании	1	1	1	4	5
Присвоенный класс	2	2	2	3	4
Входной образ					
Количество итераций при распознавании	6	1	1	7	8
Присвоенный класс	5	5	4	6	7
Входной образ					
Количество итераций при распознавании	1	1	9	10	11
Присвоенный класс	1	1	8	9	10

ДАП является *гетероассоциативной нейронной сетью*, когда входной вектор поступает на один набор нейронов, а соответствующий выходной вектор вырабатывается на другом наборе нейронов. Поэтому обученная сеть ДАП на поступающие двоичные образы-векторы, обладающие общими характерными признаками, вырабатывает на выходе один и тот же образ, т. е. ассоциирует входные образы с одним выходным (рис. 4).

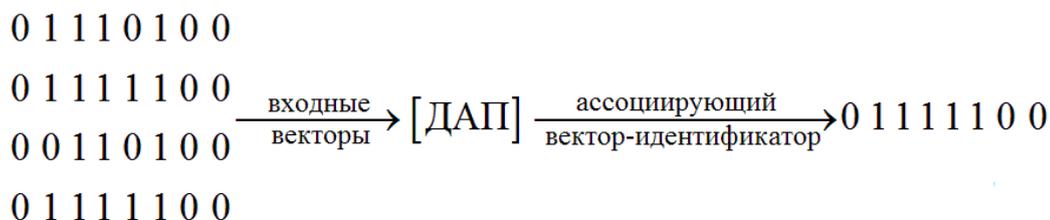


Рис. 4. Пример ассоциации сетью ДАП набора векторов с одним общим вектором-идентификатором

На рис. 5 изображена архитектура сети ДАП, состоящая из двух слоев нейронных элементов  $\mathbf{a}^{(m)} = [a_1, a_2, \dots, a_m]$  и  $\mathbf{b}^{(n)} = [b_1, b_2, \dots, b_n]$ , между которыми имеются двусторонние связи  $W_{n \times m} = [\mathbf{w}_1^{(n)}, \mathbf{w}_2^{(n)}, \dots, \mathbf{w}_m^{(n)}]$  и  $V_{m \times n} = [\mathbf{v}_1^{(m)}, \mathbf{v}_2^{(m)}, \dots, \mathbf{v}_n^{(m)}]$ , соединенные с элементами по принципу «все – со всеми», где  $m$  и  $n$  – количество нейронных элементов слоев 1 и 2 соответственно. Здесь входной вектор поступает на слой нейронов 1, а соответствующий выходной вектор вырабатывается в слое нейронов 2, и наоборот.

Для более детального понимания работы сети ДАП на рис. 6 изображена ее развернутая функциональная схема. Сеть состоит из распределительных нейронов слоя 0, не несущих какой-нибудь информативной нагрузки, а служащих только для распределения сигналов между слоями 2 и 1. Слои 1 и 2 содержат нейроны, выполняющие преобразования входных и выходных векторов  $\mathbf{a}^{(m)}$  и  $\mathbf{b}^{(n)}$ .

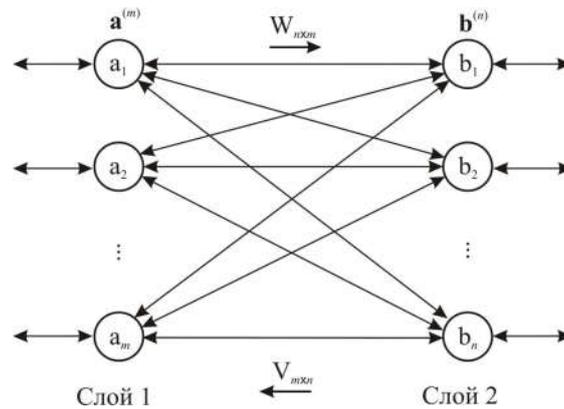


Рис. 5. Архитектура нейронной сети ДАП

Долговременная память (или ассоциации) реализуется в весовых матрицах  $W_{n \times m}$  и  $W_{n \times m}^T$ . Каждый образ состоит из двух векторов: вектора  $\mathbf{a}^{(m)}$ , являющегося выходом слоя 1, и ассоциирующего вектора  $\mathbf{b}^{(n)}$ , являющегося выходом слоя 2. Для восстановления ассоциированного образа значения вектор  $\mathbf{a}^{(m)}$  или его часть кратковременно устанавливаются на выходах слоя 1. Затем вектор  $\mathbf{b}^{(n)}$  воздействует через транспонированную матрицу  $W_{n \times m}^T$ , воспроизводя воздействие исходного вектора  $\mathbf{a}^{(m)}$  на выходе слоя 1. Каждый следующий цикл вызывает уточнение выходных векторов слоя 1 и 2 до тех пор, пока не будет достигнута точка стабильности в сети, при которой ни вектор  $\mathbf{a}^{(m)}$ , ни вектор  $\mathbf{b}^{(n)}$  не будут изменяться. Эта точка в фазовом пространстве может быть рассмотрена как резонансная, так как вектор передается вперед и обратно между слоями сети, всегда вырабатывая текущие выходные сигналы, но больше не изменяя их. Состояние нейронов представляет собой кратковременную память, так как оно может быстро изменяться при появлении нового образа.

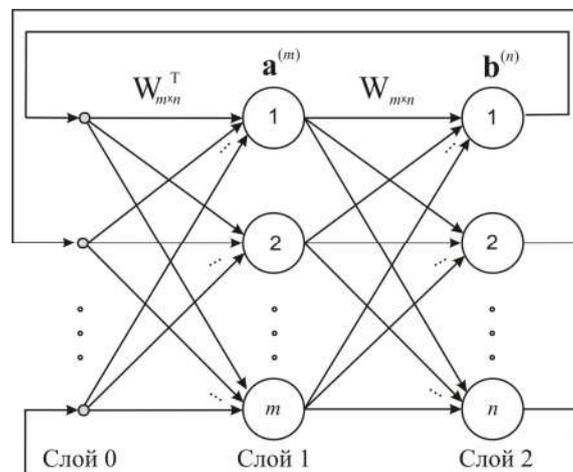


Рис. 6. Функциональная схема работы сети ДАП

Следует заметить, что нейроны в слоях 1 и 2 функционируют, как и в других нейросетевых парадигмах, находя сначала сумму взвешенных входов, а затем вычисляя по ней значение функции активации  $f(x)$ :

$$\mathbf{a}^{(m)} = f(\mathbf{b}^{(n)} W_{n \times m}), \quad (1)$$

где  $\mathbf{a}^{(m)}$  – вектор входных нейронов слоя 1;  $\mathbf{b}^{(n)}$  – вектор выходных нейронов слоя 2;  $W_{n \times m}$  – матрица весов связей между слоями 1 и 2;  $f(x)$  – функция активации нейронов. Аналогично

$$\mathbf{b}^{(n)} = f(\mathbf{a}^{(m)} \mathbf{W}_{n \times m}^T),$$

где  $\mathbf{W}_{n \times m}^T$  – транспонированная матрица весов  $\mathbf{W}_{n \times m}$ .

В качестве активационной функции нейронов часто используются пороговая либо сигмоидальная функции [2]:

$$f(x_{t+1}) = \frac{1}{1 + \exp(-\lambda \cdot x_{t+1})},$$

где  $\lambda \in (0,1)$ , а также функция вида [11]

$$f(x_{t+1}) = \begin{cases} 1, & \text{если } x_{t+1} > 0; \\ f(x_t), & \text{если } x_{t+1} = 0; \\ -1, & \text{если } x_{t+1} < 0. \end{cases} \quad (2)$$

Обучение сети ДАП производится с использованием обучающего набора, состоящего из  $k$  наборов пар соответствующих векторов  $\mathbf{a}_i^{(m)}$  и  $\mathbf{b}_i^{(n)}$  [2, 10]. Процесс обучения реализуется в форме суммирования диад пар этих векторов согласно правилу Хебба [9], т. е. весовая матрица вычисляется как сумма тензорных произведений всех векторных пар обучающего набора:

$$\mathbf{W}_{n \times m} = \sum_{i=1}^k (\mathbf{b}_i^{(n)})^T \otimes \mathbf{a}_i^{(m)}. \quad (3)$$

Обучение сети ДАП, описанное выражением (3), можно рассматривать как обобщение постулата обучения Хебба. Его также называют правилом внешнего (матричного) произведения, поскольку для нахождения матрицы весов  $\mathbf{W}_{n \times m}$  используются матричные операторы [5].

Как показывает опыт, наибольшая производительность достигается при использовании не бинарных, а биполярных векторов [2]. При этом векторная компонента, большая чем 0, принимает значение «+1», а компонента, меньшая или равная 0, принимает значение «-1».

#### 4. Архитектура и принципы работы гибридной сети

Для создания быстродействующего нейроподобного классификатора автором была спроектирована гибридная сеть, в основу которой положена архитектура сети АРТ-1. Отличительной особенностью нового нейросетевого классификатора является то, что слои сравнения и распознавания  $F_1$  и  $F_2$  заменены двухслойной сетью ДАП.

Для сохранения характерных особенностей выбранных нейронных сетей предложенная гибридная сеть изначально ограничивалась следующими требованиями. Для реализации стабильно-пластичной памяти, основанной на сети АРТ-1, необходимым является сохранение очередности фаз поиска, сравнения и выдачи результата. Последняя фаза заключается либо в принятии решения о принадлежности к одному из существующих классов, либо в создании нового. Для реализации ассоциативной памяти и использования преимуществ сети ДАП необходимо:

- наличие двух слоев нейронных элементов;
- взаимодействие посредством матрицы весов нейронов одного слоя со всеми нейронами второго слоя;
- ограничение длительности процесса восстановления ассоциаций по достижении либо точки релаксации сети, либо наперед заданного числа итераций.

Из рис. 7 видно, что предложенная архитектура гибридного нейросетевого классификатора содержит в себе характерные черты сетей АРТ-1 и ДАП. Она состоит из двух слоев  $F_1$  и  $F_2$  идентичных нейронных элементов, поэтому выполнение распознающей или сравнивающей функций определяется не их внутренним строением, а лишь структурным предназначением.

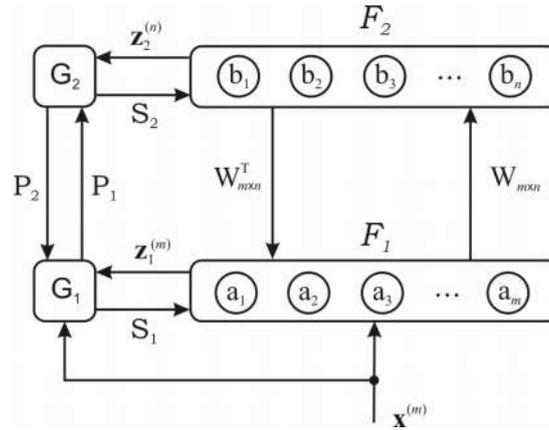


Рис. 7. Архитектура предложенной гибридной сети

Коэффициенты общей весовой матрицы  $W_{m \times n}$ , представляющие собой синаптические веса обоих слоев нейронов  $F_1$  и  $F_2$ , содержат долговременную память. Однако на выходах нейронов слоя  $F_1$  в значениях вектора  $\mathbf{a}^{(m)}$  реализуется кратковременная память, представляющая собой ассоциированные образы именно входного вектора  $\mathbf{x}^{(m)}$ , где  $m$  – размерность входного вектора. Поэтому слой  $F_1$  является входным и выполняет распознающую функцию.

Второй, выходной слой нейронов  $F_2$  выполняет функцию сравнения. После кратковременной подачи на вход первого слоя  $F_1$  значений исследуемого вектора  $\mathbf{x}^{(m)}$  на выходах нейронов второго слоя  $F_2$  в векторе  $\mathbf{b}^{(n)}$  вырабатываются значения вектора-классификатора, на основании которого делается заключение о принадлежности входного вектора  $\mathbf{x}^{(m)}$  к тому или иному классу, где  $n$  – размерность выходного вектора.

Процесс восстановления ассоциаций, содержащихся в памяти, состоит в следующем. Долговременная память (или ассоциации) реализуются в весовых матрицах  $W_{m \times n}$  и  $W_{n \times m}^T$ , и каждый образ состоит из двух векторов: вектора  $\mathbf{a}^{(m)}$ , являющегося выходом слоя  $F_1$ , и вектора  $\mathbf{b}^{(n)}$ , ассоциированного образа, являющегося выходом слоя  $F_2$ . Для восстановления ассоциированного образа вектор  $\mathbf{a}^{(m)}$  или его часть кратковременно устанавливаются на выходах слоя  $F_1$ . Затем вектор  $\mathbf{a}^{(m)}$  удаляется и сеть приводится в стабильное состояние, вырабатывая ассоциированный вектор  $\mathbf{b}^{(n)}$  на выходе слоя  $F_2$ . Далее полученный вектор воздействует на транспонированную матрицу  $W_{m \times n}^T$ , в результате чего на выходе слоя  $F_1$  воспроизводится исходный входной вектор  $\mathbf{a}^{(m)}$ .

Как правило, размерности векторов  $\mathbf{a}^{(m)}$  и  $\mathbf{b}^{(n)}$  не обязательно должны совпадать. Они зависят только от поставленных задач. Например, в поставленном нами эксперименте ставилась задача спроектировать классификатор двумерных монохромных изображений. Следовательно, входной вектор  $\mathbf{a}^{(m)}$  должен содержать двоичную информацию о двумерном бинарном изображении, а выходной вектор  $\mathbf{b}^{(n)}$  должен содержать значения классифицирующего вектора-идентификатора, где  $n$  – количество классов, хранимых в долговременной памяти сети (при избыточном кодировании). Тогда  $m = a \cdot b$ , где  $m$  – длина входного вектора, а  $a$  и  $b$  – высота и ширина обрабатываемых изображений в пикселях.

Внедрение сети ДАП позволяет ассоциировать большое число входных многомерных векторов с конечным числом малоразмерных векторов-идентификаторов. Следовательно, предложенная гибридная сеть производит сравнение не самих (чаще всего зашумленных) образов, а их ассоциированных векторов-идентификаторов (рис. 8).

Управляющие блоки  $G_1$  и  $G_2$  являются многозадачными. Прежде всего они служат для контроля количества производимых итераций в фазе восстановления ассоциаций. Это достигается путем сравнения двух последних значений на выходах слоев распознавания и сравнения. Выходные значения нейронов слоев  $F_1$  и  $F_2$  поступают на управляющие блоки через каналы  $\mathbf{z}_1^{(m)}$  и  $\mathbf{z}_2^{(n)}$  соответственно.

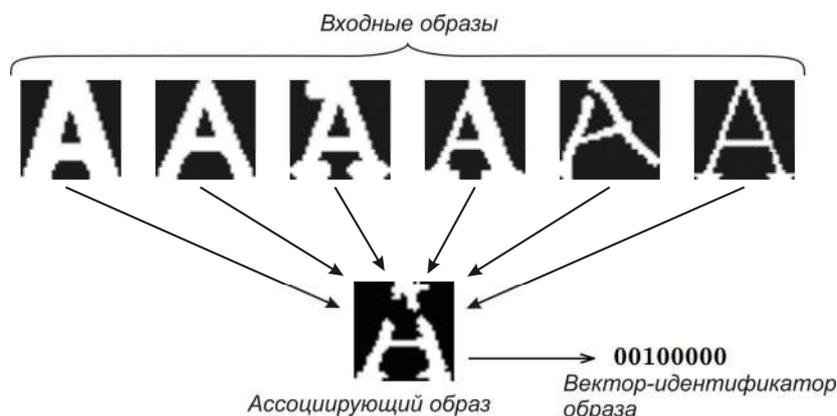


Рис. 8. Процесс ассоциации входных двумерных монохромных изображений с одномерным бинарным вектором

Если предыдущее и действительное значения слоя сравнения  $F_2$  одинаковы, то на блок  $G_1$  от блока  $G_2$  поступает положительный единичный сигнал  $P_2$ , в противном случае он равен нулю. Если подобная ситуация, когда образ воспроизводит самого себя, возникнет и со значениями слоя распознавания  $F_1$  или будет достигнут предел наперед заданного количества итераций, то на слой распознавания поступит положительный сигнал  $S_1$ , на блок  $G_2$  – положительный сигнал  $P_1$ . Таким образом, с помощью положительных сигналов останова  $P_1$  и  $P_2$  процесс распознавания образов будет остановлен и значения на выходах слоев  $F_1$  и  $F_2$  зафиксируются (рис. 9).

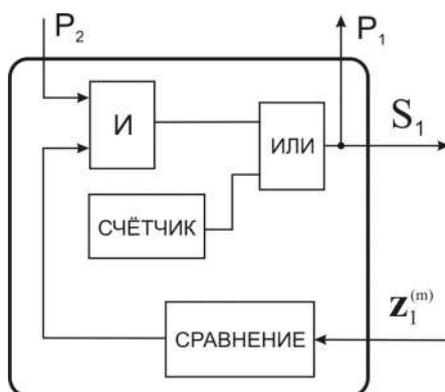


Рис. 9. Структура управляющего блока  $G_1$

Каждый описанный выше цикл вызывает уточнения выходных векторов слоев  $F_1$  и  $F_2$  до тех пор, пока не будет достигнута точка стабильности в пространстве состояний сети. На рис. 10 изображен пример работы предложенной гибридной сети, где происходит «стягивание» значений кратковременной памяти к устойчивому состоянию (аттрактору) в фазовом пространстве. При вводе изображения 1 в конце первого цикла сеть выдает ассоциированное изображение 2. Далее с третьего по пятый цикл идет постепенное уточнение изображения, а начиная с шестого, изображение воспроизводит само себя.



Рис. 10. Значения кратковременной памяти за шесть произведенных итераций

Поэтому в фазовом пространстве состояний входного вектора после пятой итерации блуждающая точка станет стационарной. После поступления на блок  $G_2$  положительного сигнала  $P_1$  происходит подсчет количества единиц в выходном векторе  $\mathbf{b}^{(n)}$ .

Для упрощения и автоматизации режимов работы и обучения гибридной сети была использована следующая кодировка векторов-идентификаторов. Если сделать допущение, что число образов, хранящихся в долговременной памяти, не превосходит число нейронов в слое сравнения  $F_2$ , то порядковый номер записанного в память образа будет равен номеру единственного ненулевого элемента соответствующего вектора-идентификатора. Например, для кодирования четырех образов необходимо, чтобы длина выходного вектора  $\mathbf{b}^{(n)}$  составляла минимум четыре элемента. Тогда при предъявлении нейронной сети образов из обучаемой выборки векторы-идентификаторы будут принимать значения 0001, 0010, 0100 и 1000. Если на входы сети подать неизвестные образы, то на ее выходах появятся векторы, значения элементов которых не будут совпадать с указанными. Например, это могут быть 0110, 1001 или 1110.

Если в конце итерации в значениях вектора  $\mathbf{b}^{(n)}$  будет находиться только одна единица, а все остальные значения элементов будут равны нулю, то будем считать, что сеть классифицировала данный образ правильно. В противном случае с помощью обратных синаптических связей  $W_{m \times n}^T$  значения полученного вектора будут поданы на входы слоя распознавания  $F_1$ , на выходах которого будут наблюдаться изменения в кратковременной памяти.

В свою очередь, полученный образ через прямые связи  $W_{m \times n}$  будет воздействовать на входы слоя сравнения  $F_2$ . После определения нового вектора-идентификатора в блоке  $G_1$  опять будет произведен контроль на наличие и положение единиц в векторе. Если количество итераций превысит наперед заданное число, то на слой сравнения  $F_2$  будет подан положительный сигнал  $S_2$ , который разрешит добавление нового образа.

## 5. Экспериментальные результаты

Для преодоления отмеченных недостатков алгоритма АРТ-1 автор не ограничился только проектированием новой архитектуры. В процессе проведенных экспериментов, направленных на улучшение производительности сети ДАП, был выдвинут и реализован ряд предложений, касающихся как представления выходных значений нейронов, так и модификации процесса обучения:

1. В формуле (2), хорошо себя зарекомендовавшей в процессе экспериментов, был заменен статический порог активации, равный нулю, на динамический  $T_t$ , который вычисляется для каждой итерации отдельно:

$$T_{t+1} = \frac{\max(x_t) + \min(x_t)}{2},$$

где  $\max(x_t)$  и  $\min(x_t)$  – максимальное и минимальное значения кратковременной памяти, хранящейся в значениях элементов векторов  $\mathbf{a}^{(m)}$  и  $\mathbf{b}^{(n)}$  после  $t$ -й итерации.

2. Также в процессе работы хорошие результаты дало использование биполярной кодировки векторов  $\mathbf{a}^{(m)}$  и  $\mathbf{b}^{(n)}$  не только при обучении, но и при распознавании образов, т. е. векторы принимают значения только «+1» или «-1». Учитывая предыдущий пункт, формула (2) приняла следующий вид:

$$f(x_{t+1}) = \begin{cases} +1, & \text{àñèè } x_{t+1} \geq T_{t+1}; \\ -1, & \text{àñèè } x_{t+1} < T_{t+1}. \end{cases}$$

3. После вычисления выходного сигнала слоя 1 по формуле (1) была добавлена двумерная фильтрация в плоскости входного изображения. Фильтрация заключалась в усреднении значений выходов нейронов среди своих четырех соседей (кроме крайних нейронов). Это позволяет избежать аномального роста значений активности отдельных нейронов. На рис. 11 показана графическая интерпретация влияния связей соседних нейронов друг на друга. Если про-

вести свертку вектора  $\mathbf{a}^{(m)}$  в двумерную матрицу  $A_{a \times b}$  с размерами  $a$  и  $b$ , соответствующими входным изображениям, то значение отдельного нейрона  $A_{i,j}$  будет определяться выражением

$$A_{i,j}^{t+1} = A_{i,j}^t + A_{i+1,j}^t + A_{i,j+1}^t + A_{i-1,j}^t + A_{i,j-1}^t.$$

В результате, если среди нейронов с отрицательными значениями находится нейрон с аномальным положительным значением, его значение, т. е. его собственное влияние на соседние нейроны, уменьшается.

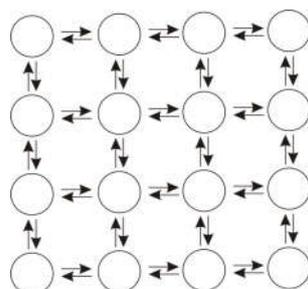


Рис. 11. Направления влияний соседних нейронов друг на друга

4. Была сделана также попытка придать сети ДАП адаптивные свойства. Адаптивная сеть должна изменять свои веса в процессе функционирования с целью более гибкого распознавания. Это означает, что подача на вход сети обучающего набора входных векторов заставляет ее изменять энергетическое состояние до получения резонанса. Постепенно кратковременная память в процессе функционирования сети посредством настройки коэффициентов весовой матрицы должна превратиться в долговременную память.

Таблица 3

Классификация набора изображений с помощью гибридной сети

Входной образ	Количество итераций		Присвоенный класс
	при обучении	при распознавании	
	3	1	1
	1	1	1
	1	3	1
	3	1	2
	3	1	3
	1	1	3
	1	3	1
	1	3	4
	3	1	4

5. В вычислительных экспериментах для поправки весовой матрицы использовалось правило Хэбба [2], в котором изменение веса пропорционально уровню активации его нейрона-источника и уровню активации нейрона-приемника:

$$\Delta\omega_{ij} = \eta a_i b_j,$$

где  $\Delta\omega_{ij}$  – изменение связи  $i$ -го нейрона вектора  $\mathbf{a}^{(m)}$  с  $j$ -м нейроном вектора  $\mathbf{b}^{(n)}$  в матрицах  $W_{m \times n}$  и  $W_{m \times n}^T$ ;  $\eta$  – положительный нормирующий коэффициент обучения, меньший 1.

Данный метод позволяет без добавления нового класса адаптировать сеть и доучить ее распознавать уже существующие классы, когда поступающие образы лишь незначительно отличаются от тех образов, которые записаны в долговременной памяти.

Благодаря новой архитектуре и перечисленным модификациям процесс распознавания нейроподобного классификатора стал более адаптивным. Под адаптивностью подразумеваем не только устойчивость к шуму и выбор критерия сходства для определения характерных признаков, но и наличие обучения с учителем. Последнее позволяет неидентичные образы ассоциировать с одним вектором-идентификатором, что дает возможность при использовании классификатора в системе управления роботом правильно реагировать на предъявляемые образы.

В табл. 3 приведен пример классификации набора изображений, уже использовавшегося при работе алгоритма АРТ-1 (см. табл. 1). При сравнении указанных таблиц видно, что предложенный нейросетевой классификатор не только сузил число распознанных классов с шести до четырех, но и произвел в процессе обучения и распознавания значительно меньше итераций: 29 и 22 против 17 и 15 соответственно. Следовательно, разработанный классификатор тратит меньше времени для поиска подобных образов в долговременной памяти.

Для демонстрации преимуществ обучения с учителем будет уместно привести следующий пример. Разработанный классификатор заранее был обучен распознавать три различных изображения, приведенные в первой строчке табл. 2. После подачи на вход сети зашумленных изображений классификатор выдал результат с распознаванием 100 %.

### Заключение

Благодаря своей гибкой архитектуре предложенная в настоящей статье гибридная сеть сохранила в себе характерные свойства как сети АРТ-1, так и сети ДАП. С одной стороны, она позволяет добавлять в долговременную память новые классы образов, не затирая уже хранящиеся, а с другой стороны, процесс поиска образов в памяти стал ассоциативным. Это положительным образом сказалось на устойчивости классификатора к шумам и его быстродействию. Предложенный гибридный классификатор унаследовал от сети ДАП не только быстрый способ обучения, который сводится (в отличие от более распространенных нейросетевых архитектур) к алгебраическому сложению, но и процесс обучения с учителем. Следует также упомянуть, что благодаря своей простоте предложенная гибридная нейросетевая архитектура является привлекательной с точки зрения как аналоговой, так и цифровой реализации. Полученные изделия могут встраиваться в системы управления сложных технических устройств.

### Список литературы

1. Печерский, Ю.Н. Распознающие системы в природе, науке и технике / Ю.Н. Печерский. – Кишинев : Штиинца, 1986. – 156 с.
2. Уоссермен, Ф. Нейрокомпьютерная техника: теория и практика / Ф. Уоссермен. – М. : Мир, 1990. – 240 с.
3. Джонс, М.Т. Программирование искусственного интеллекта в приложениях / М.Т. Джонс ; пер. с англ. А.И. Осипов. – М. : ДМК Пресс, 2004. – 312 с.
4. Рассел, С. Искусственный интеллект: современный подход / С. Рассел, П. Норвиг. – 2-е изд. – М. : Вильямс, 2007. – 1424 с.
5. Хайкин, С. Нейронные сети: полный курс / С. Хайкин. – 2-е изд., пер. с англ. – М. : Вильямс, 2006. – 1104 с.
6. Симонов, П.В. Лекции о работе головного мозга. Потребностно-информационная теория высшей нервной деятельности / П.В. Симонов. – М. : Институт психологии РАН, 1998. – 98 с.

7. Крот, А.М. Быстрые алгоритмы и программы цифровой спектральной обработки сигналов и изображений / А.М. Крот, Е.Б. Минервина. – Минск : Наука і тэхніка, 1995. – 407 с.
8. Крот, А.М. Дискретные модели динамических систем на основе полиномиальной алгебры / А.М. Крот. – Минск : Наука і тэхніка, 1990. – 312 с.
9. Головкин, В.А. Нейронные сети: обучение, организация и применение. Кн. 4 : учеб. пособие для вузов / А.В. Головкин ; под общ. ред. Л.И. Галушкина. – М. : ИПРЖР, 2001. – 256 с.
10. Kosko, B. Bidirectional associative memories / B. Kosko // IEEE Transactions on Systems, Man, and Cybernetics. – 1988. – Vol. 18, № 1. – P. 49–60.
11. Freeman J.A., Skapura D.M. Neural networks, algorithms, applications, and programming techniques / J.A. Freeman, D.M. Skapura. – N.Y. : Addison-Wesley, 1991. – P. 414.

Поступила 09.02.09

*Объединенный институт проблем  
информатики НАН Беларуси,  
Минск, Сурганова, 6  
e-mail: bigznich@tut.by*

**R.A. Prakapovich**

### **ADAPTIVE NEURONET CLASSIFIER**

The memory overwrite problem in artificial neural networks when presenting new classes of images is considered. The hybrid classifier which capable with retraining is proposed. The distinctive features of the described classifier include the supervised learning and realization of associative search of images in long-term memory, raising noise stability of the classifier and reducing the number of memory access operations in comparison with the sequential method.