

## 7. ЛЕКЦИЯ. Адаптивная резонансная теория.

Мозг человека выполняет трудную задачу обработки непрерывного потока сенсорной информации, получаемой из окружающего мира. Из моря тривиальной информации он должен выделить жизненно важную, обработать ее и, возможно, зарегистрировать в «долговременном регистре». Однако новые образы запоминаются в такой форме, что ранее запомненные не модифицируются и не забываются. Понимание сути этого процесса представляет собой серьезную задачу для исследователей: 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. Способ восприятия внешнего мира живыми организмами (и, прежде всего, человеком) состоит в постоянной оценке: 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 картины и в этом случае реакция организма должна соответствовать ранее накопленному опыту, а специальное запоминание 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.

Традиционные искусственные нейронные сети оказались не в состоянии решить проблему стабильности - пластичности. Очень часто обучение новому образу уничтожает или изменяет результаты предшествующего обучения. В некоторых случаях это не существенно. Если имеется только фиксированный набор обучающих векторов, они могут предъявляться при обучении циклически. Рассмотренные на предыдущих лекциях нейронные системы не адаптированы к решению этой задачи. Так, например, многослойный персептрон, обучающийся по методу обратного распространения, запоминает весь пакет обучающей информации, при этом образы обучающей выборки предъявляются в процессе обучения многократно. Попытки затем обучить персептрон новому образу приведут к модификации синаптических связей с неконтролируемым разрушением структуры памяти о предыдущих образах. Таким образом, персептрон не способен к запоминанию новой информации, и необходимо полное переобучение сети.

Аналогичная ситуация имеет место и в сетях Кохонена и 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. 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.

Более того, приведены примеры сети, в которой только четыре обучающих вектора, предъявляемых циклически, заставляют веса сети изменяться непрерывно, никогда не сходясь. Такая временная нестабильность явилась одним из главных факторов, заставивших Гроссберга и его сотрудников исследовать радикально отличные конфигурации. Адаптивная резонансная теория (АРТ) является одним из результатов исследования этой проблемы.

Сети и алгоритмы АРТ сохраняют пластичность, необходимую для изучения новых образов и предотвращения изменений ранее запомненных образов. Открытие этой способности  $A'_1$  вызвало большой интерес к АРТ, но многие исследователи нашли теорию трудной для понимания. Математическое описание АРТ является сложным, но основные идеи и принципы реализации достаточно просты для понимания. Мы сконцентрируемся далее на общем описании АРТ. Нашей целью является изложение конкретной информации, чтобы слушатель мог понять основные идеи и возможности этого важного вида сетей.

### ***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 предъявленного образа с содержимым памяти. При удачном поиске в памяти предъявленный образ классифицируется с одновременной уточняющей модификацией синаптических весов нейрона, выполнившего классификацию. Такую ситуацию называют возникновением адаптивного резонанса в сети в ответ на предъявление образа. Если резонанс не возникает в пределах некоторого заданного порогового уровня, то тест новизны считается успешным 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 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.1).

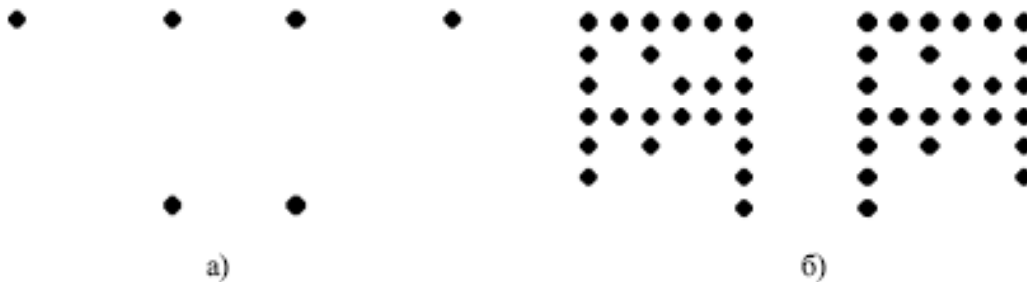


Рис. 1. Общие свойства

Обе пары картинок на рис. 1 имеют общее свойство: в каждой из пар черная точка в правом нижнем углу заменена на белую, а белая точка в левом нижнем углу — на черную. Такое изменение для правой пары картинок (на рисунке — пара (б)), очевидно, является не более чем шумом, и оба образа (б) есть искаженные версии одного и того же изображения. Тем самым, измененные точки не являются для этого образа критическими.

Совершенно иная ситуация изображена на левой паре картинок (а). 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 (а) и "стабильное" решение о совпадении картинок (б). 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.

Второй значительный вывод теории — необходимость самоадаптации алгоритма поиска образов в памяти. Нейронная сеть работает в постоянно изменяющихся условиях, так что предопределенная схема поиска, отвечающая некоторой структуре информации, может в дальнейшем оказаться

неэффективной при изменении этой структуры. В теории адаптивного резонанса адекватность достигается введением специализированной ориентирующей системы, которая самосогласованно прекращает дальнейший поиск резонанса в памяти и принимает решение о новизне информации. Ориентирующая система также обучается в процессе работы.

При наличии резонанса теория ART предполагает возможность прямого доступа к образу памяти, откликнувшемуся на резонанс. В этом случае шаблон критических черт выступает ключем-прототипом для прямого доступа.

Эти и другие особенности теории адаптивного резонанса нашли свое отражение в нейросетевых архитектурах, которые получили такое же название – ART.

### **Адаптивные резонансные нейронные сети**

Традиционные ИНС с трудом решают проблему стабильности и одновременной пластичности запоминания, т. е. проблему наращивания числа образов в памяти без искажения тех, которые там уже имеются. Сеть прямого распространения обучается на фиксированном множестве примеров. Если после обучения число примеров увеличивается, то «дообучение» невозможно, требуется полное переобучение. В сетях Кохонена также предполагается, что все входные образы принадлежат к одному из полученных при обучении классов. Выявление нового класса требует полного повторения процесса обучения.

Аналогичная ситуация наблюдается при использовании сетей Хопфилда и Хэмминга.

Сеть ART предназначена для решения проблемы стабильности – пластичности. Существуют так называемые ART1-сети, обрабатывающие двоичные векторы, а также ART2-сети для обработки непрерывных векторов. Кроме того, следует назвать модели ARTMAP и FuzzyART.

Рассмотрим далее работу ART1.

Сеть ART1 представляет собой векторный классификатор. Входной вектор относится к одной из групп ранее запомненных векторов. Решение о классификации выдается в виде возбуждения одного из нейронов распознающего слоя. Отличие ART1 от других ИНС состоит в следующем:

1) если входной вектор похож на один из ранее запомненных векторов, то запомненный вектор будет изменяться, чтобы стать похожим на входной вектор;

2) возможен отказ от классификации, если входной вектор не похож ни на один из запомненных векторов.

В этой ситуации образуется новый класс. Упрощенная конфигурация сети ART1 включает в себя слой сравнения, слой распознавания, сброс, приемник 1 и приемник 2 (рис. 2).

Введем обозначения:  $m$  – размер входного вектора;  $n$  – число запомненных образов. Слой сравнения получает двоичный входной вектор  $X$  и первоначально пропускает его неизменным для формирования выходного вектора  $C$  (на более поздней стадии компоненты вектора  $R$  модифицируют  $C$ )



Рис. 2. Структура нейронной сети ART

Упрощенно *слой сравнения* имеет вид, приведенный на рис. 3.

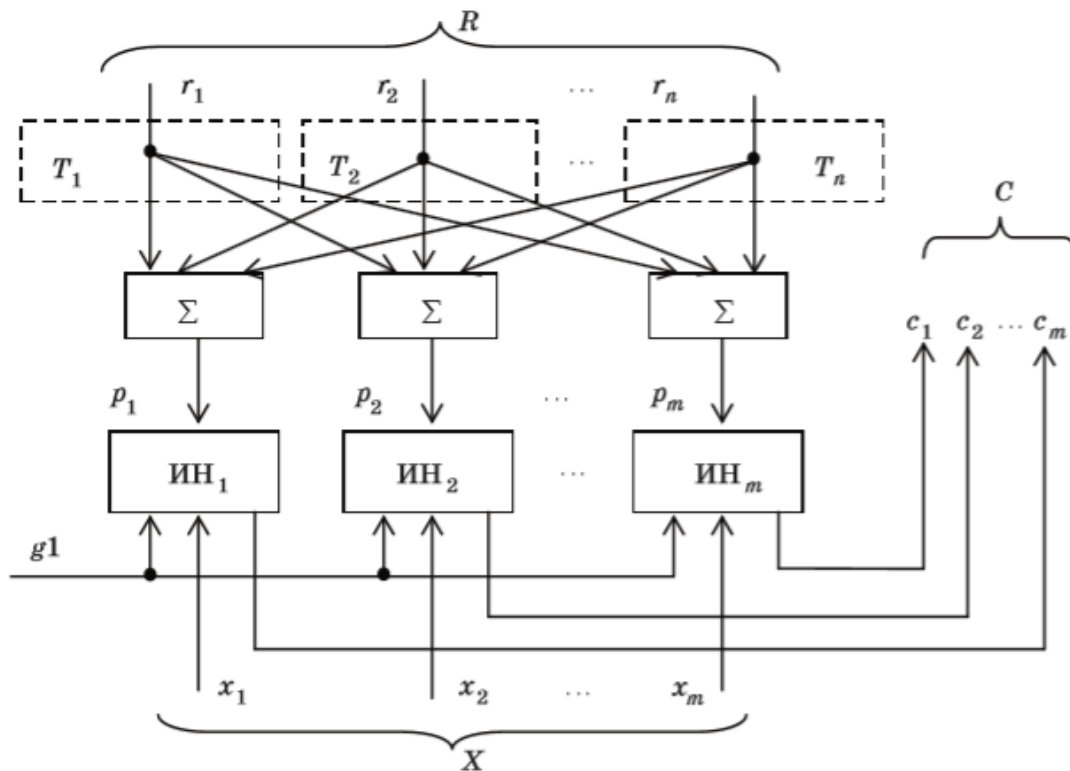


Рис. 3. Структура слоя сравнения

Векторы Т образуются весовыми коэффициентами  $t_{ij}$ , которые являются двоичными числами.

Каждый нейрон слоя сравнения (ИН<sub>1</sub> – ИН<sub>m</sub>) получает три двоичных входа:

- 1) компонента входного вектора  $x_i$ ;
- 2) сигнал обратной связи  $p_j$ ;
- 3) вход от приемника  $g_1$ .

Для того чтобы на выходе нейрона была единица, необходимо, чтобы не менее двух его входов были равны единице. Первоначально сигнал  $g_1$  установлен в «1», а сигналы  $p_j$  – в «0», поэтому исходно вектор С совпадает с X.

**Слой распознавания** служит для классификации входных векторов. Каждый нейрон слоя распознавания имеет вектор весов В. Только один нейрон в этом слое возбуждается, все остальные заторможены. Это достигается благодаря тому, что все нейроны слоя распознавания охвачены механизмом латерального возбуждения/торможения (рис. 4).

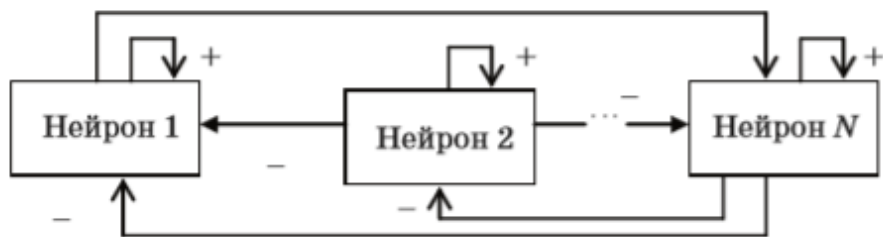


Рис. 4. Взаимодействие нейронов слоя распознавания

Таким образом, чем больше нейрон возбужден, тем больше он тормозит другие нейроны и одновременно поддерживает свой уровень. Упрощенная версия слоя распознавания приведена на рис. 5

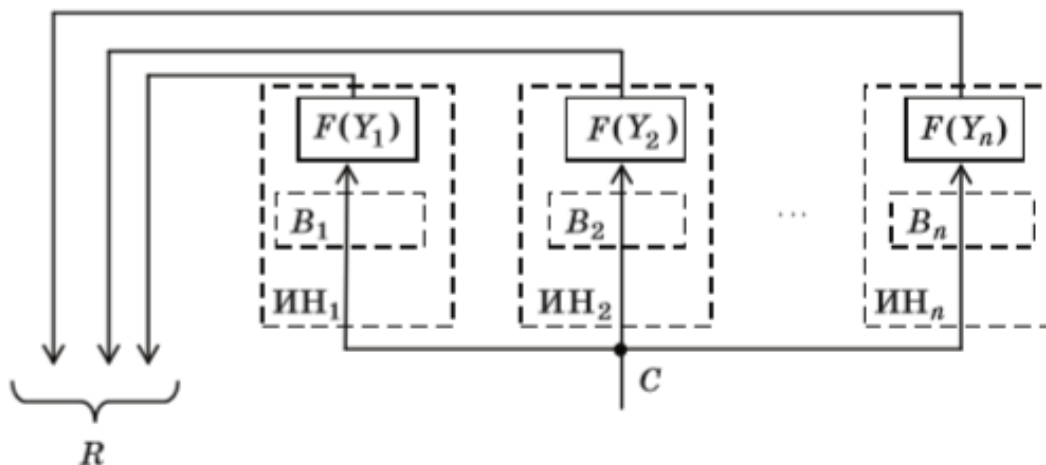


Рис. 5. Взаимодействие нейронов слоя распознавания

На входы всех нейронов слоя распознавания поступают сигналы от всех нейронов слоя сравнения, и наоборот. Каждый нейрон слоя распознавания имеет вектор весов  $V$ . Выход нейрона

$$Y_j = B_j C.$$

Активационная функция  $F$  пороговая:

$$F(Y_j) = \begin{cases} 1, & Y_j \geq T, \\ 0, & Y_j < T. \end{cases}$$

Таким образом, вектор  $R$  двоичный.

Нейрон  $j$  имеет максимальную реакцию, если вектор  $C$  (выход слоя сравнения) максимально похож на его весовой вектор  $B_j$ . Таким образом, веса нейрона представляют собой запомненный образ для группы векторов.

Веса являются действительными числами. Двоичная версия этого образа запоминается в наборе весов слоя сравнения.

В процессе функционирования каждый нейрон слоя распознавания вычисляет свертку вектора собственных весов и входного вектора  $C$ . Чем ближе веса к вектору  $C$ , тем выше выход нейрона. Нейрон-победитель затормаживает остальные нейроны слоя распознавания.

Таким образом, только один нейрон слоя распознавания будет иметь на выходе «1», а остальные – «0». Следовательно, вектор  $R$  будет иметь только одну единичную компоненту.

*Приемник 2.* Его выход равен единице, если входной вектор  $X$  имеет хотя бы одну единичную компоненту. Иначе говоря, сигнал  $g2$  представляет собой логическое ИЛИ компонент входного вектора:

$$g2 = x_1 \vee x_2 \vee x_3 \vee \dots \vee x_m$$

*Приемник 1.* Работает так же, как приемник 2, но приход хотя бы одной единичной компоненты  $R$  сбрасывает  $g1$  в «0»:

$$g1 = (x_1 \vee x_2 \vee \dots \vee x_m) \wedge \overline{(r_1 \vee r_2 \vee \dots \vee r_m)}.$$

*Сброс.* Модуль сброса измеряет сходство между векторами  $X$  и  $C$ . Мерой сходства служит отношение число единиц в  $X$  к их числу в  $C$ :

$$S = N / D,$$

где  $D$  – число единиц в векторе с наибольшим числом единиц.

Например,

$$X = [1011101] (D = 5),$$

$$C = [0011101] (N = 4).$$

Таким образом, параметр сходства изменяется от 0 до 1. Три основные операции АРТ заключаются в распознавании, сравнении и поиске.

*Распознавание* состоит в следующем. Пока входной вектор  $X$  отсутствует, все его компоненты можно считать нулевыми, следовательно,  $g2 = 0$ , и в «0» устанавливаются выходы всех нейронов слоя распознавания (так как  $C = X$ ).

Затем подается входной вектор  $X$ , т. е. одна или больше компонент входного вектора становятся ненулевыми, а  $g1$  и  $g2$  – равными единице. Это создает условия для возбуждения нейронов слоя сравнения, и вектор  $C$  дублирует вектор  $X$ .

После этого в слое распознавания вычисляется свертка вектора весов  $V_j$  и  $C$ . Нейрон, веса которого наиболее похожи на входной вектор, побеждает и затормаживает остальные нейроны. Единственная компонента  $R$  становится единичной. Таким образом, один нейрон слоя распознавания соответствует одной из категорий классификации.

*Сравнение.* Компоненты вектора  $R$  поступают на все нейроны слоя сравнения через веса  $t_{ij}$ , которые принимают двоичные значения.

Взаимосвязь между векторами весов  $T$  и  $V$  заключается в том, что  $V$  является масштабированной версией вектора  $T$ .

Поскольку  $R$  больше не нулевой, сигнал  $g1$  устанавливается в «0», поэтому в соответствии с правилом 2/3 возбудиться могут только нейроны, получающие на входе одновременно единицы от входного вектора  $X$  и вектора  $R$ . Если  $X$  и  $R$  не имеют совпадающих компонент, то обратная связь от распознающего слоя сбросит компоненты  $C$  в «0» (рис. 6).

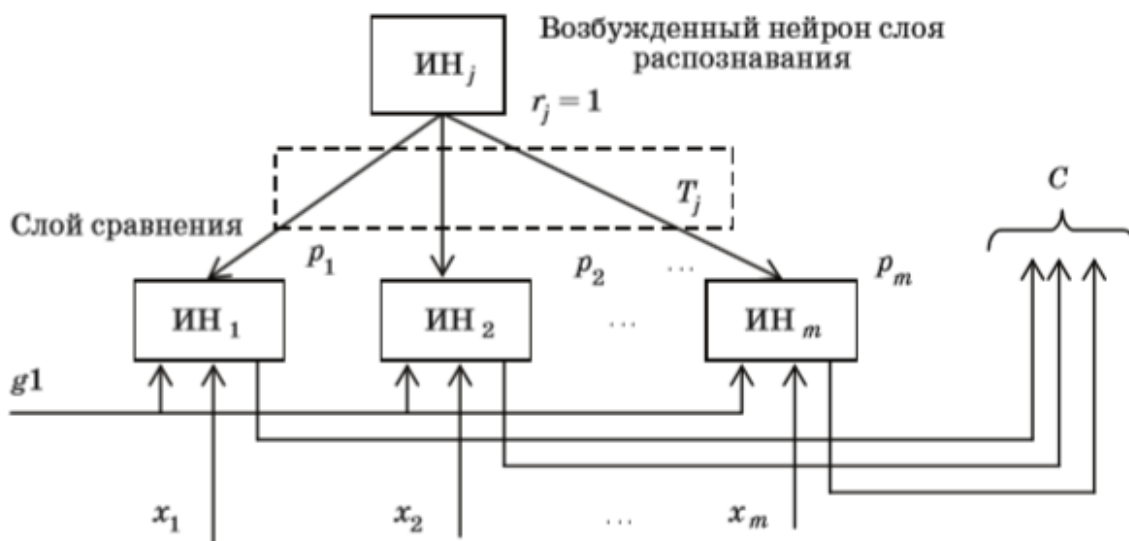


Рис. 6. Взаимодействие слоев распознавания и сравнения



Если различий между векторами  $X$  и  $P$  много, то вектор  $C$  будет иметь нулей значительно больше, чем  $X$ , и вырабатывается сигнал сброса, который отключает возбужденный нейрон на время текущей классификации.

*Поиск.* Если сигнал сброса не вырабатывается, то сходство считается найденным, а классификация завершена. Если же сброс произошел, то все компоненты  $R$  обнуляются,  $g_1$  устанавливается в «1», вектор  $C$  становится равным  $X$ , и происходит новая проверка, но уже без заторможенного нейрона. Эта процедура продолжается до тех пор, пока не будут заторможены все нейроны. В этом случае происходит запоминание нового образа, для чего выделяется новый нейрон, весовые коэффициенты которого  $V$  и  $T$  устанавливаются в соответствии с новым образом.

Если же все-таки будет найден один из нейронов, похожий на входной образ, то производится обучающий цикл с целью коррекции весов  $V$  и  $T$ .

При обучении на вход сети последовательно подаются векторы из обучающего множества и подстраиваются веса сети с тем, чтобы сходные векторы активизировали один и тот же нейрон.

Возможны разные алгоритмы обучения. Рассмотрим так называемый быстрый алгоритм.

Вес связи  $b_{ij}$ , связывающей  $i$ -й нейрон слоя сравнения и  $j$ -й нейрон слоя распознавания, может быть найден по формуле

$$b_{ij} = \frac{2C_i}{\sum_k C_k + 1},$$

где  $j$  – возбужденный нейрон;  $\sum$  – число единиц на выходе слоя сравнения.

Компоненты вектора весов  $T_j$ , связанного с новым запомненным вектором, изменяются таким образом, чтобы они стали равны соответствующим компонентам вектора  $C$ :  $t_{ij} = c_i$  ( $t_{ij}$  – связь между выигравшим нейроном  $j$  в слое распознавания и нейроном  $i$  в слое сравнения).

Сумма может быть рассмотрена как «размер» вектора. Если вектор  $C$  «большой», то  $b_{ij}$  маленькое. Таким образом, оказывается возможным разделение векторов, если один из них является поднабором другого (т. е. входит в него).

Пример. Пусть имеются векторы  $X_1 = [1\ 0\ 0\ 0\ 0]$ ,  $X_2 = [1\ 1\ 1\ 0\ 0]$  (т. е.  $X_1$  – поднабор  $X_2$ ). Если в формуле для  $b_{ij}$  убрать сумму, получим

$$T_1 = V_1 = [1\ 0\ 0\ 0\ 0],$$

$$T2 = B2 = [1\ 1\ 1\ 0\ 0].$$

Если затем подать на вход сети вектор  $X_1$ , то оба нейрона будут возбуждены одинаково. Если же в формуле для  $b_{ij}$  использовать сумму, то получим

$$B1 = [1\ 0\ 0\ 0\ 0],$$

$$B2 = [0,5\ 0,5\ 0,5\ 0\ 0],$$

и при подаче вектора  $X_1$  для нейрона 1 уровень возбуждения будет соответствовать «1», а для нейрона 2 – 0,5 (правильно). При подаче  $X_2$  получим соответственно 1 и 3/2 (тоже правильно). Пусть далее подан вектор

$$X_3 = [1\ 1\ 0\ 0\ 0].$$

Для нейрона 1 уровень возбуждения будет единица, а для нейрона 2 он составит 2/3. Нейрон 1 победит, вектор  $C$  приобретет значение  $[1\ 1\ 0\ 0\ 0]$ . Величина  $S = 1/2$ , и если уровень сходства  $r = 2/3$ , то нейрон 1 будет заторможен, и выиграет нейрон 2 ( $C = [1\ 1\ 0\ 0\ 0]$ ),  $S = 1$ .

Таким образом, сеть АРТ может распознавать входные векторы, сильно похожие на эталонные образы, корректировать эталонные образы под модификации входных векторов или отслеживать медленные изменения эталона, а также наращивать число эталонных образов. Разумеется, при моделировании сетей АРТ на традиционных компьютерах трудно добиться высокого быстродействия, в таких случаях особенно выгодно применять параллельно работающие вычислители.

### ***Теоремы АРТ***

Гроссберг доказал некоторые теоремы, которые описывают характеристики сетей АРТ. Четыре результата, приведенные ниже, являются одними из наиболее важных:

1. После стабилизации процесса обучения предъявление одного из обучающих векторов (или вектора с существенными характеристиками категории) будет активизировать требуемый нейрон слоя распознавания без поиска. Такая характеристика "прямого доступа" обеспечивает быстрый доступ к предварительно изученным образам.

2. Процесс поиска является устойчивым. После определения выигравшего нейрона в сети не будет возбуждений других нейронов из-за изменения векторов выхода слоя сравнения  $C$ ; только сигнал сброса может вызвать такие изменения.

3. Процесс обучения является устойчивым. Обучение не будет вызывать переключения с одного возбужденного нейрона слоя распознавания на другой.

4. Процесс обучения конечен. Любая последовательность произвольных входных векторов будет производить стабильный набор весов после конечного количества обучающих серий. Повторяющиеся последовательности обучающих векторов не будут приводить к циклическому изменению весов.

### ***Нерешенные проблемы и недостатки АРТ-1***

Нейронные сети АРТ, при всех их замечательных свойствах, имеют ряд недостатков. Один из них — большое количество синаптических связей в сети, в расчете на единицу запоминаемой информации. При этом многие из весов этих связей (например, вектора  $T$ ) 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, и 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.

### ***Сети АРТ-2 и АРТ-3***

Основной отличительной чертой нейронной сети АРТ-2 является возможность работы с аналоговыми векторами и сигналами. По сравнению с АРТ-1 в архитектуре сети сделаны некоторые изменения, позволяющие отдельным подсистемам функционировать асинхронно, что является принципиальной необходимостью для аппаратных реализаций.

Важное отличие аналоговых сигналов от битовых — принципиальная возможность аналоговых векторов быть сколь угодно близкими друг к другу (в то время как пространство битовых векторов дискретно). Это накладывает дополнительные требования на функционирование нейронов слоя сравнения: требуется более тонкий и чувствительный механизм для выделения областей резонанса. Общим решением здесь является переход к многослойной архитектуре, со все более точной настройкой при переходе от слоя к слою, что и применено в АРТ-2. Функционирование слоя распознавания принципиально не изменяется.

Сети АРТ-2 применялись для распознавания движущихся изображений. Успешные эксперименты проведены в Массачусетском Технологическом

Институте (MIT). Поскольку нейросистемы АРТ не содержат механизма инвариантного распознавания (в отличие от неокогнитрона, см. следующие лекции), то в сочетании с ними применяются специализированные (часто не нейросетевые) системы инвариантного представления образов, например, двумерное преобразование Фурье или более сложные алгоритмы. Более подробное рассмотрение особенностей и применений АРТ-2 требует профессионального изучения и не входит в наши цели.

Следующим шагом в развитии АРТ явилась сеть АРТ-3. Особенности обучения нейронов сетей АРТ-1 и АРТ-2 не позволяют использовать эти сети в качестве элементов более крупных иерархических нейросистем, в частности, компоновать из них многослойные сети. Поэтому представление в АРТ иерархически организованной информации затруднительно, и это весьма отдаляет ее от систем восприятия человека и животных.

Изложенные проблемы решены в сети АРТ-3, которая выступает как многослойная архитектура. При переходе от слоя к слою происходит контрастирование входных образов и запоминание их в виде все более общих категорий. При этом основной задачей каждого отдельного слоя является сжатие входящей информации. Образ входит в адаптирующийся резонанс между некоторой парой слоев, в дальнейшем этот резонанс распространяется на следующие слои иерархии. В АРТ-1 и АРТ-2 недостаточный уровень резонанса приводил к генерации сигнала сброса, что приводило к полному торможению слоя распознавания. В случае многослойной сети АРТ-3 подобное недопустимо, так как при этом разрывается поток информации. Поэтому в АРТ-3 введен специальный механизм — зависимость активности синапсов обратных связей от времени, — аналогичный рефрактерному торможению биологического нейрона после передачи возбуждения. Поэтому вместо полного сброса сигнала происходит торможение синаптических сигналов обратной связи, и слой сравнения получает исходное состояние возбуждения для выполнения фазы поиска нового резонанса.

Интересным предложением является также использование в многослойной иерархии слоев, которые не являются слоями АРТ, а принадлежат некоторой другой архитектуре. В этом случае система получается гибридной, что может привести к возникновению новых полезных свойств.

Развитие теоретических исследований АРТ продолжается. По высказыванию авторов теории, АРТ представляет собой нечто существенно более конкретное, чем философское построение, но намного менее конкретное, чем законченная программа для компьютера. Однако уже в современном виде,

опираясь на свою более чем 20-летнюю историю, сети АРТ с успехом применяются в различных областях. АРТ сделала также важный шаг вперед в общей проблеме моделирования пластично-стабильного восприятия.

### *Другие АРТ-сети*

Приведем краткий обзор сетей ARTMAP и FUZZY-ART.

ARTMAP. Они объединяют элементы обучения и самообучения (или обучения с поощрением и без поощрения, Supervised and Unsupervised Training). Для этого обычно формируется комбинация из двух АРТ-сетей.

FUZZY-ART-сети (нечеткие АРТ-сети). Они представляют собой прямое расширение АРТ-1-сетей средствами нечеткой логики (Fuzzy Logic). В них применяются следующие операторы:

- для определения класса образов в слое распознавания;
- для расчета степени сходства (Reset-критерий);
- для адаптации весов.