

Санкт-Петербургский политехнический университет Петра Великого
Институт компьютерных наук и технологий
Кафедра компьютерных систем и программных технологий

Реферат

Теория адаптивного резонанса

(тема работы)

История и методология науки

(наименование дисциплины)

Работу выполнили:

Волкова М.Д.

подпись

Ф.И.О.

Преподаватель:

Мелехин В.Ф.

подпись

Ф.И.О.

Санкт-Петербург
2018

Введение

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. Таким образом, результирующее восприятие должно быть одновременно пластичным (адаптированным к новой информации), и при этом стабильным (не разрушающим память о старых образах). 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. Простые нейронные сети не адаптированы к решению этой задачи. Так, например, многослойный персептрон (рис. 1), 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.

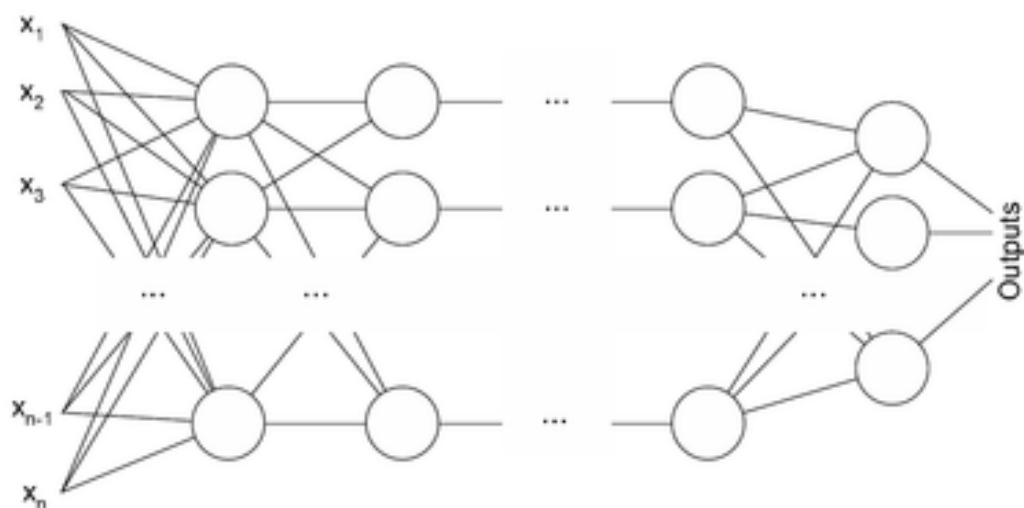


Рис. 1. Многослойный персептрон

Аналогичная ситуация имеет место и в сетях Кохонена и Хемминга (рис. 2),

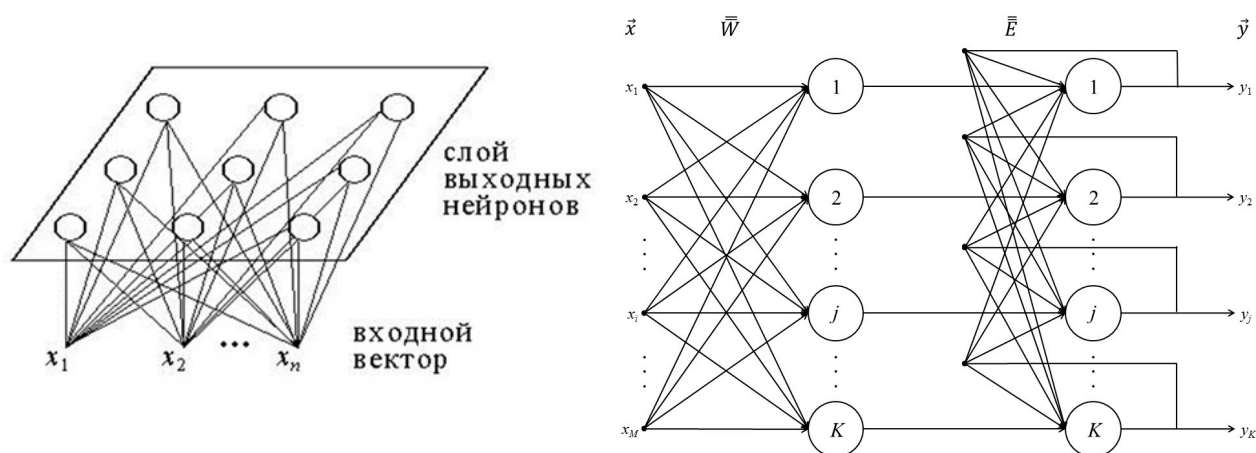


Рис. 2. Сеть Кохонена и Хемминга.

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** (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 специфичных критических особенностей в образе. Рассмотрим это на примере.



Рис. 3

Обе пары картинок на рис. 3 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 (b)), очевидно, является не более чем шумом, и оба образа (b) 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 (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.

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 АРТ предполагает возможность прямого доступа к образу памяти, откликнувшись 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 — АРТ.

Архитектура АРТ

Адаптивная резонансная теория включает две парадигмы, каждая из которых определяется формой входных данных и способом их обработки. АРТ-1 создана для обработки двоичных входных векторов, в то время как АРТ-2, более позднее обобщение АРТ-1, может классифицировать как двоичные, так и непрерывные векторы.

Описание АРТ

Сеть АРТ представляет собой векторный классификатор. Входной вектор классифицируется в зависимости от того, на какой из множества ранее запомненных образов он похож. Свое классификационное решение сеть АРТ выражает в форме возбуждения одного из нейронов распознающего слоя. Если входной вектор не соответствует ни одному из запомненных образов, создается новая категория путем запоминания образа, идентичного новому входному вектору. Если определено, что входной вектор похож на один из ранее запомненных с точки зрения определенного критерия сходства, запомненный вектор будет изменяться (обучаться) под воздействием нового входного вектора таким образом, чтобы стать более похожим на этот входной вектор.

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 заставить измениться существующую память.

Упрощенная архитектура АРТ

На рис. 4 показана упрощенная конфигурация сети АРТ.

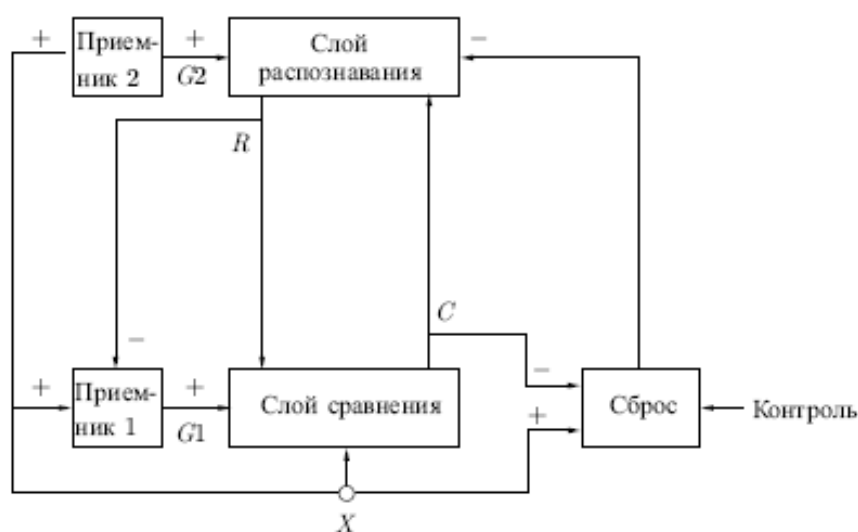


Рис. 4

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.

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. 5) получает три двоичных входа:

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).

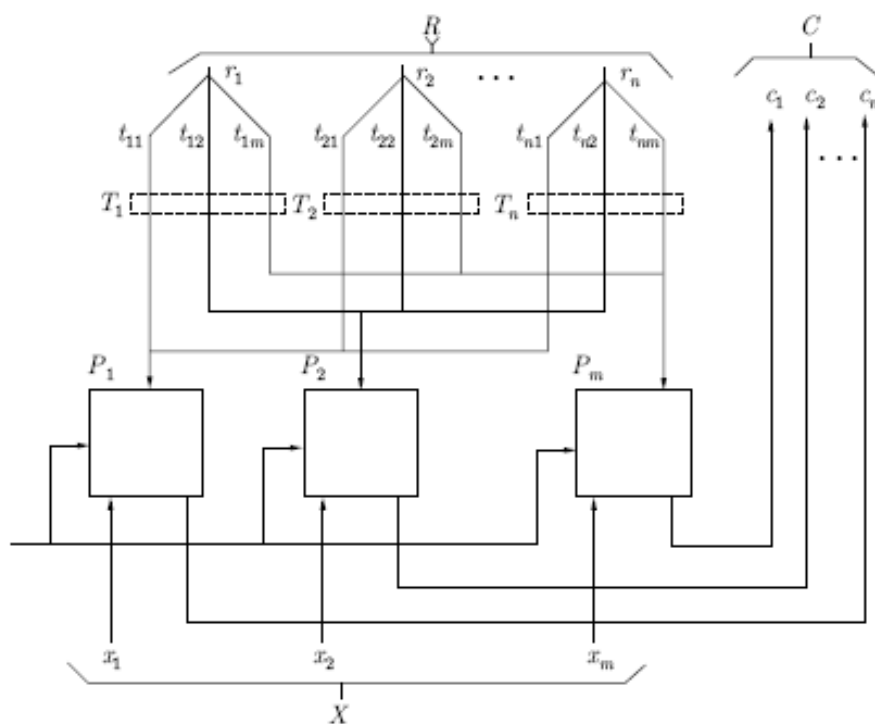


Рис. 5

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 G_i 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. 6, 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. 6); этот набор состоит из весов связей, соединяющих определенные нейроны слоя распознавания, по одному весу на каждый нейрон слоя сравнения.

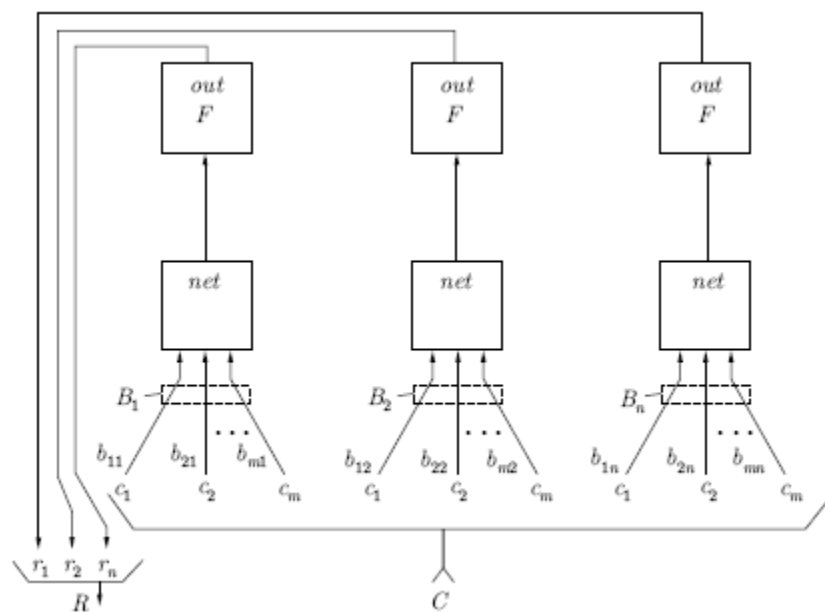


Рис. 6.

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 рисунке, 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. G_2 , выход Приемника 2, равен единице, если входной вектор X имеет хотя бы одну единичную компоненту. Более точно, G_2 является логическим ИЛИ от компонента вектора X .

Приемник 1. Как и сигнал G_2 , выходной сигнал G_1 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, G_1 устанавливается в нуль. Таблица, определяющая эти соотношения:

ИЛИ от компонента вектора X	ИЛИ от компонента вектора R	G_1
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.

Функционирование сети АРТ в процессе классификации

Процесс классификации в АРТ состоит из трех основных фаз: распознавание, сравнение и поиск.

Фаза распознавания. В начальный момент времени входной вектор отсутствует на входе сети; следовательно, все компоненты входного вектора X можно рассматривать как нулевые. Тем самым сигнал G_2 устанавливается в 0 и, следовательно, в нуль устанавливаются

выходы всех нейронов слоя распознавания. Поскольку все нейроны слоя распознавания начинают работу в одинаковом состоянии, они имеют равные шансы выиграть в последующей конкуренции. Затем на вход сети подается входной вектор X , который должен быть классифицирован. Этот вектор должен иметь одну или более компонент, отличных от нуля, в результате чего и G_1 , и G_2 становятся равными единице. Это "подкачивает" нейроны слоя сравнения, обеспечивая один из двух единичных входов, необходимых для возбуждения нейронов в соответствии с правилом двух третей, и тем самым позволяя нейрону возбуждаться, если соответствующая компонента входного вектора X равна единице. Таким образом, в течение данной фазы вектор C в точности дублирует вектор X .

Далее, для каждого нейрона в слое распознавания вычисляется свертка вектора его весов B_j и вектора C (см. рис.7). Нейрон с максимальным значением свертки имеет веса, наилучшим образом соответствующие входному вектору. Он выигрывает конкуренцию и возбуждается, одновременно затормаживая все остальные нейроны этого слоя. Таким образом, единственная компонента r_j вектора R становится равной единице, а все остальные компоненты становятся равными нулю.

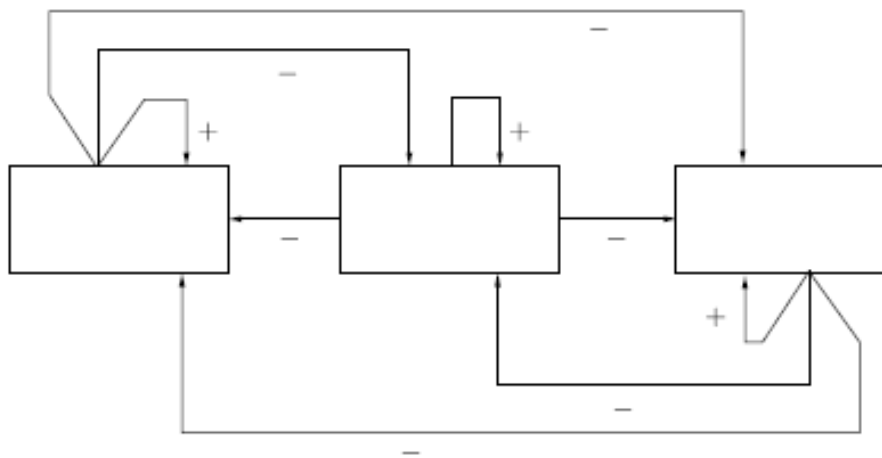


Рис. 7.

В результате сеть АРТ запоминает образы в весах нейронов слоя распознавания — один нейрон для каждой категории классификации. Нейрон слоя распознавания, веса которого наилучшим образом соответствуют входному вектору, возбуждается, его выход устанавливается в единичное значение, а выходы остальных нейронов этого слоя устанавливаются в нуль.

Фаза сравнения. Единственный возбужденный нейрон в слое распознавания возвращает единицу обратно в слой сравнения в виде своего выходного сигнала r_j . Эта единственная единица может быть визуально представлена в виде "верного" выхода, подающегося через отдельную связь с весом t_{ij} на

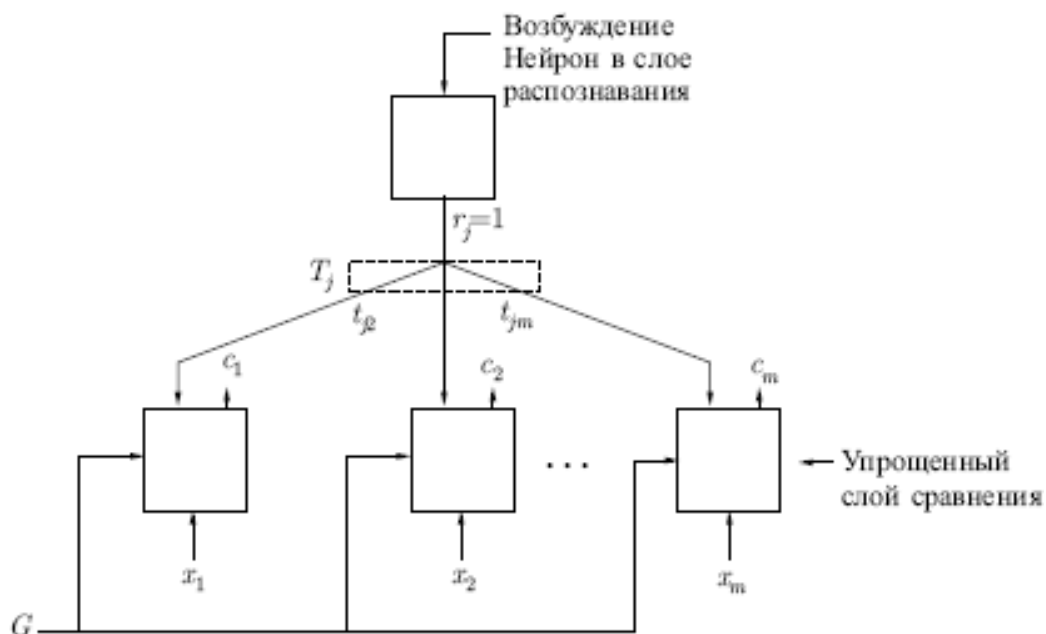


Рис. 8.

каждый нейрон в слое сравнения, обеспечивая каждый нейрон сигналом p_j , равным величине t_{ij} (нулю или единице) (см. рис. 8).

Алгоритмы инициализации и обучения построены таким образом, что каждый весовой вектор T_j имеет двоичные значения весов; кроме того, каждый весовой вектор B_j представляет собой масштабированную версию соответствующего вектора T_j . Это означает, что все компоненты P (вектора возбуждения слоя сравнения) также являются двоичными величинами.

Так как вектор R не является больше нулевым, сигнал G_1 устанавливается в нуль. Таким образом, в соответствии с правилом двух третей, возбудиться могут только нейроны, получающие на входе одновременно единицы от входного вектора X и вектора P .

Другими словами, обратная связь от распознающего слоя действует так, чтобы установить компоненты C в нуль в случае, если входной вектор не соответствует входному образу (т. е. если X и P не имеют совпадающих компонент).

Если имеются существенные различия между X и P (малое количество совпадающих компонент векторов), несколько нейронов на фазе сравнения будут возбуждаться и C будет содержать много нулей, в то время как X содержит единицы. Это означает, что возвращенный вектор P не является искомым и возбужденные нейроны в слое распознавания должны быть заторможены. Такое торможение производится блоком сброса, который сравнивает входной вектор X и вектор C и вырабатывает сигнал сброса, если

степень сходства этих векторов меньше некоторого уровня. Влияние сигнала сброса заключается в установке выхода возбужденного нейрона в нуль, отключая его на время текущей классификации.

Фаза поиска. Если не выработан сигнал сброса, сходство является адекватным и процесс классификации завершается. В противном случае, другие запомненные образы должны быть исследованы с целью поиска лучшего соответствия. При этом торможение возбужденного нейрона в распознающем слое приводит к установке всех компонент вектора R в 0, G_1 устанавливается в 1 и входной вектор X опять прикладывается в качестве C . В результате другой нейрон выигрывает соревнование в слое распознавания и другой запомненный образ P возвращается в слой сравнения. Если P не соответствует X , возбужденный нейрон в слое распознавания снова тормозится. Этот процесс повторяется до тех пор, пока не встретится одно из двух событий:

1. Найден запомненный образ, сходство которого с вектором X выше уровня параметра сходства, т. е. $S > P$. Если это происходит, проводится обучающий цикл, в процессе которого модифицируются веса векторов T_j и B_j , связанных с возбужденным нейроном в слое распознавания.
2. Все запомненные образы проверены, определено, что они не соответствуют входному вектору, и все нейроны слоя распознавания заторможены. В этом случае предварительно не распределенный нейрон в распознающем слое выделяется этому образу и его весовые векторы B_j и T_j устанавливаются соответствующими новому входному образу.

Проблема производительности. Описанная сеть должна производить последовательный поиск среди всех запомненных образов. В аналоговых реализациях это будет происходить очень быстро; однако, при моделировании на обычных цифровых компьютерах процесс может оказаться очень длительным. Если же сеть АРТ реализуется на параллельных процессорах, все свертки на распознающем уровне могут вычисляться одновременно. В этом случае поиск может стать очень быстрым.

Время, необходимое для стабилизации сети с латеральным торможением, может быть длительным при моделировании на последовательных цифровых компьютерах. Чтобы выбрать победителя в процессе латерального торможения, все нейроны в слое должны быть вовлечены в одновременные вычисления и передачу. Этот процесс может потребовать проведения большого объема вычислений перед достижением сходимости.

Реализация

АРТ представляет собой нечто большее, чем философия, но намного менее конкретное, чем программа для компьютера. Поэтому возник широкий круг реализаций, сохраняющих идеи АРТ, но сильно отличающихся в деталях. Описываемая далее реализация может рассматриваться в качестве типовой, но необходимо иметь в виду, что другие успешные реализации имеют большие отличия от нее.

Функционирование сетей АРТ

Рассмотрим более детально пять фаз процесса функционирования АРТ: инициализацию, распознавание, сравнение, поиск и обучение.

Инициализация. Перед началом процесса обучения сети все весовые векторы B_j и T_j , а также параметр сходства p , должны быть установлены в начальные значения.

Веса векторов B_j все инициализируются в одинаковые малые значения. Эти значения должны удовлетворять условию $b_{ij} < \frac{L}{L - 1 + m}$, для всех i, j , где m — количество компонент входного вектора, L — константа, большая 1 (обычно $L = 2$).

Эта величина является критической; если она слишком большая, сеть может распределить все нейроны распознающего слоя одному входному вектору.

Веса векторов T_j все инициализируются в единичные значения, так что $t_{ij} = 1$, для всех j, i . Эти значения также являются критическими; показано, что слишком маленькие веса приводят к отсутствию соответствия в слое сравнения и отсутствию обучения.

Параметр сходства p устанавливается в диапазоне от 0 до 1 в зависимости от требуемой степени сходства между запомненным образом и входным вектором. При высоких значениях p сеть относит к одному классу только очень слабо отличающиеся образы. С другой стороны, малое значение p заставляет сеть группировать образы, которые имеют слабое сходство между собой. Для выработки точной классификации полезна возможность изменять коэффициент сходства на протяжении процесса обучения, обеспечивая только грубую классификацию в начале процесса обучения и затем постепенно увеличивая коэффициент сходства.

Распознавание. Появление на входе сети входного вектора X инициализирует фазу распознавания. Так как вначале выходной вектор слоя распознавания отсутствует, сигнал G_1 устанавливается в 1 функцией ИЛИ вектора X , обеспечивая все нейроны слоя сравнения одним из двух входов, необходимых для их возбуждения (как требует правило двух третей). В результате любая компонента вектора X , равная единице, обеспечивает второй единичный вход, заставляя соответствующий нейрон слоя сравнения возбуждаться и устанавливая его выход в единицу. Таким образом, в этот момент времени вектор C идентичен вектору X

Как обсуждалось ранее, распознавание реализуется вычислением свертки для каждого нейрона слоя распознавания, определяемой следующим выражением: $NET_j = (B_j * C)$, где B_j — весовой вектор, соответствующий нейрону j в слое распознавания, C — выходной вектор нейронов слоя сравнения (в этот момент C равно X), NET_j — возбуждение нейрона j в слое распознавания.

F является пороговой функцией, определяемой следующим образом:

$$OUT_j = \begin{cases} 1, & \text{если } NET_j > T, \\ 0, & \text{в противном случае,} \end{cases}$$

где T представляет собой порог.

Принято, что латеральное торможение существует, но игнорируется здесь для сохранения простоты выражения. Торможение является причиной того, что только нейрон с максимальным значением NET будет иметь выход, равный единице; все остальные нейроны будут иметь нулевой выход. Можно рассмотреть системы, в которых в распознающем слое возбуждаются несколько нейронов в каждый момент времени, однако это выходит за рамки данной работы.

Сравнение. На этой фазе сигнал обратной связи от слоя распознавания устанавливает G_1 в нуль; правило двух третей позволяет возбуждаться только тем нейронам, которые имеют соответствующие компоненты векторов P и X , равные единице.

Блок сброса сравнивает вектор C и входной вектор X , вырабатывая сигнал сброса, когда их сходство S ниже порога сходства. Вычисление этого сходства упрощается тем, что оба вектора являются двоичными (все элементы либо 0, либо 1). Следующая процедура проводит требуемое вычисление сходства:

1. Вычислить D — количество единиц в векторе X .
2. Вычислить N — количество единиц в векторе C .

Затем вычислить сходство S следующим образом: $S = N/D$.

Например, примем, что
 $X = 1011101$ $D = 5$
 $C = 0011101$ $N = 4$
 $S = N/D = 0,8$.

S может изменяться от 1 (наилучшее соответствие) до 0 (наихудшее соответствие).

Заметим, что правило двух третей делает C логическим произведением входного вектора X и вектора P . Однако P равен T_j , весовому вектору выигравшего соревнования нейрона. Таким образом, D может быть определено как количество единиц в логическом произведении векторов T_j и X .

Поиск. Если сходство S выигравшего нейрона превышает параметр сходства, поиск не требуется. Однако если сеть предварительно была обучена, появление на входе вектора, не идентичного ни одному из предъявленных ранее, может возбудить в слое распознавания нейрон со сходством ниже требуемого уровня. В соответствии с алгоритмом обучения возможно, что другой нейрон в слое распознавания будет обеспечивать более хорошее соответствие, превышая требуемый уровень сходства, несмотря на то, что свертка между его весовым вектором и входным вектором может иметь меньшее значение. Пример такой ситуации показан ниже.

Если сходство ниже требуемого уровня, запомненные образы могут быть просмотрены, чтобы найти образ, наиболее соответствующий входному вектору. Если такой образ отсутствует, вводится новый несвязанный нейрон, который в дальнейшем будет обучен. Чтобы инициализировать поиск, сигнал сброса тормозит возбужденный нейрон в слое распознавания на время проведения поиска, сигнал G_1 устанавливается в единицу и другой нейрон в слое распознавания выигрывает соревнование. Его запомненный образ затем проверяется на сходство, и процесс повторяется до тех пор, пока конкуренцию не выиграет нейрон из слоя распознавания со сходством, большим требуемого уровня (успешный поиск), либо пока все связанные нейроны не будут проверены и заторможены (неудачный поиск).

Неудачный поиск будет автоматически завершаться на несвязанном нейроне, так как его веса все равны единице, своему начальному значению. Поэтому правило двух третей приведет к идентичности вектора C входному вектору X , сходство S примет значение единицы и критерий сходства будет удовлетворен.

Обучение. Обучение представляет собой процесс, в котором набор входных векторов подается последовательно на вход сети, а веса сети изменяются при этом таким образом, чтобы сходные векторы активизировали соответствующие

им нейроны. Заметим, что это - неуправляемое обучение, здесь нет учителя и нет целевого вектора, определяющего требуемый ответ.

Различают два вида обучения: медленное и быстрое. При медленном обучении входной вектор предъявляется настолько кратковременно, что веса сети не успевают достигнуть своих асимптотических значений при единичном предъявлении. В этом случае значения весов будут определяться, скорее, статистическими характеристиками входных векторов, чем характеристиками какого-то одного входного вектора. Динамика сети в процессе медленного обучения описывается дифференциальными уравнениями.

Быстрое обучение является специальным случаем медленного обучения, когда входной вектор прикладывается на достаточно длительный срок, чтобы позволить весам приблизиться к их окончательным значениям. В этом случае процесс обучения описывается только алгебраическими выражениями. Кроме того, компоненты весовых векторов T_j принимают двоичные значения, в отличие от непрерывного диапазона значений, требуемого в случае быстрого обучения. В данной лекции мы опишем только быстрое обучение.

Рассмотренный далее обучающий алгоритм используется как в случае успешного, так и в случае неуспешного поиска.

Пусть вектор весов B_j (связанный с возбужденным нейроном j распознающего слоя) равен нормализованной величине вектора C . Эти веса вычисляются следующим образом:

$$b_{ij} = \frac{Lc_i}{L - I + \sum_k c_k},$$

где c_i — i -я компонента выходного вектора слоя сравнения, j — номер выигравшего нейрона в слое распознавания, b_{ij} — вес связи, соединяющей нейрон i в слое сравнения с нейроном j в слое распознавания, L — константа > 1 (обычно 2).

Компоненты вектора весов T_j , связанного с новым запомненным вектором, изменяются таким образом, что становятся равны соответствующим двоичным величинам вектора C :

$t_{ij} = c_i$, для всех i , где t_{ij} является весом связи между выигравшим нейроном j в слое распознавания и нейроном i в слое сравнения.

Пример обучения сети АРТ

В общих чертах сеть обучается при помощи изменения весов таким образом, что предъявление входного вектора заставляет сеть активизировать нейроны в слое распознавания, связанные со сходным запомненным вектором. Кроме этого, обучение проводится в форме, не разрушающей запомненные ранее образы, и предотвращает тем самым временную нестабильность. Эта задача управляется на уровне выбора критерия сходства. Новый входной образ (который сеть раньше не видела) не будет соответствовать запомненным образам с точки зрения параметра сходства, тем самым формируя новый запоминаемый образ. Входной образ, в достаточной степени соответствующий одному из запомненных образов, не будет формировать нового экземпляра, он просто будет модифицировать тот, на который он похож. В результате при соответствующем выборе критерия сходства предотвращается запоминание ранее изученных образов и временная нестабильность.

На рис. 9 показан типичный сеанс обучения сети АРТ. Буквы изображены состоящими из маленьких квадратов, каждая буква размерностью 8x8. Каждый квадрат в левой части представляет компоненту вектора X с единичным значением, не показанные квадраты являются компонентами с нулевыми значениями. Буквы справа представляют запомненные образы, каждый является набором величин компонент вектора T_j

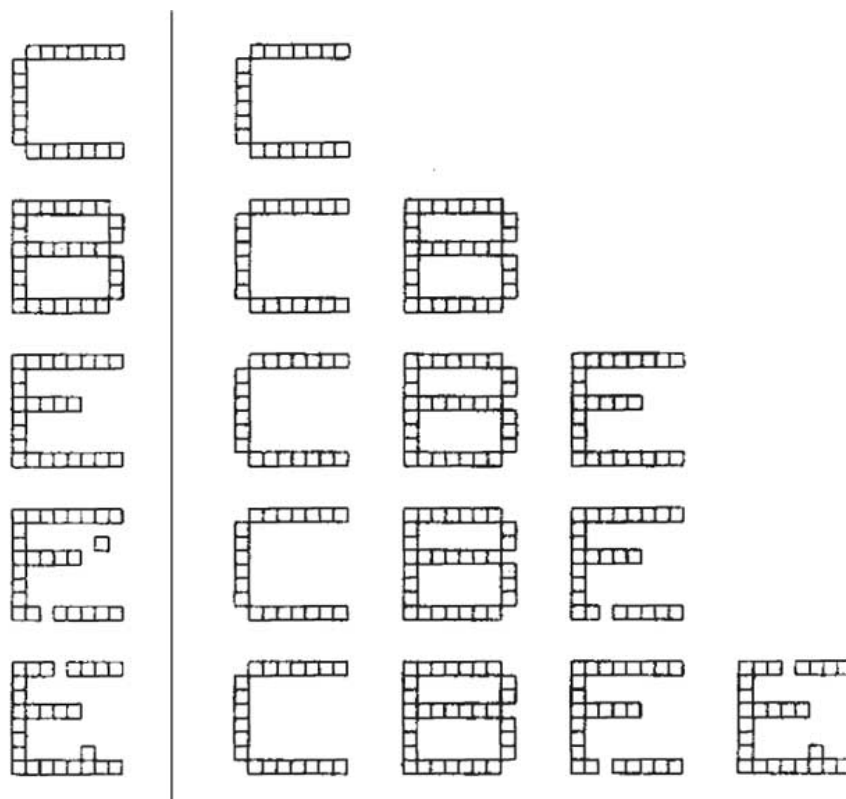


Рис. 9.

Вначале на вход заново проиницированной системы подается буква "С". Так как отсутствуют запомненные образы, фаза поиска заканчивается неуспешно; новый нейрон выделяется в слое распознавания, и веса T_j устанавливаются равными соответствующим компонентам входного вектора, при этом веса B_j представляют масштабированную версию входного вектора.

Далее предъявляется буква "В". Она также вызывает неуспешное окончание фазы поиска и выделение нового нейрона. Аналогичный процесс повторяется для буквы "Е". Затем слабо искаженная версия буквы "Е" подается на вход сети. Она достаточно точно соответствует запомненной букве "Е", чтобы выдержать проверку на сходство, поэтому используется для обучения сети. Отсутствующий пиксель в нижней ножке буквы "Е" устанавливает в 0 соответствующую компоненту вектора C , заставляя обучающий алгоритм установить этот вес запомненного образа в нуль, тем самым воспроизводя искажения в запомненном образе. Дополнительный изолированный квадрат не изменяет запомненного образа, так как не соответствует единице в запомненном образе.

Четвертым символом является буква "Е" с двумя различными искажениями. Она не соответствует ранее запомненному образу (S меньше чем p), поэтому для ее запоминания выделяется новый нейрон.

Этот пример иллюстрирует важность выбора корректного значения критерия сходства. Если значение критерия слишком велико, большинство образов не будут подтверждать сходство с ранее запомненными и сеть будет выделять новый нейрон для каждого из них. Такой процесс приводит к плохому обобщению в сети, в результате даже незначительные изменения одного образа будут создавать отдельные новые категории; далее количество категорий увеличивается, все доступные нейроны распределяются, и способность системы к восприятию новых данных теряется. Наоборот, если критерий сходства слишком мал, сильно различающиеся образы будут группироваться вместе, искажая запомненный образ, до тех пор, пока в результате не получится очень малое сходство с одним из них.

К сожалению, отсутствует теоретическое обоснование выбора критерия сходства, и в каждом конкретном случае необходимо решить "волевым усилием", какая степень сходства должна быть принята для отнесения образов к одной категории. Границы между категориями часто неясны, и решение задачи для большого набора входных векторов может быть чрезмерно трудным.

Гроссберг предложил процедуру с использованием обратной связи для настройки коэффициента сходства, вносящую, однако, некоторые искажения в результаты классификации как "наказание" за внешнее вмешательство с целью

увеличения коэффициента сходства. Такие системы требуют правил оценки корректности для производимой ими классификации.

Характеристики АРТ

Системы АРТ имеют ряд важных характеристик, не являющихся очевидными. Формулы и алгоритмы могут казаться произвольными, в то время как в действительности они были тщательно отобраны и соответствуют требованиям теорем относительно производительности систем АРТ. В данном разделе описываются некоторые алгоритмы АРТ, раскрывающие отдельные вопросы инициализации и обучения.

Инициализация весовых векторов Т

В ранее рассмотренном примере обучения сети можно было видеть, что правило двух третей приводит к вычислению вектора C как функции И между входным вектором X и выигравшим соревнование запомненным вектором T_j . Следовательно, любая компонента вектора C будет равна единице в том случае, если соответствующие компоненты обоих векторов равны единице. После обучения эти компоненты вектора T_j остаются единичными; все остальные устанавливаются в нуль.

Это объясняет, почему веса t_{ij} должны инициализироваться единичными значениями. Если бы они были проинициализированы нулевыми значениями, все компоненты вектора C были бы нулевыми независимо от значений компонент входного вектора, и обучающий алгоритм предохранял бы веса от изменения их нулевых значений.

Обучение может рассматриваться как процесс "сокращения" компонент запомненных векторов, которые не соответствуют входным векторам. Процесс необратим, если вес однажды установлен в нуль, — обучающий алгоритм никогда не восстановит его единичное значение.

Это свойство имеет важное отношение к процессу обучения. Предположим, что группа точно соответствующих векторов должна быть классифицирована как одна категория, определяемая возбуждением одного нейрона в слое распознавания. Если эти векторы последовательно предъявляются сети, то при предъявлении первого будет распределяться нейрон распознающего слоя и его веса будут обучены с целью соответствия входному вектору. Обучение при предъявлении остальных векторов будет приводить к обнулению весов в тех позициях, которые имеют нулевые значения в любом из входных векторов. Таким образом, запомненный вектор представляет собой логическое пересечение всех обучающих векторов и может включать существенные характеристики данной категории весов. Новый вектор, включающий только существенные характеристики, будет соответствовать этой категории. Таким

образом, сеть корректно распознает образ, никогда не виденный ранее, т. е. реализуется возможность, напоминающая процесс восприятия в мозге человека.

Настройка весовых векторов V_j

Выражение, описывающее процесс настройки весов, является центральным для описания процесса функционирования сетей АРТ:

$$b_{ij} = \frac{Lc_i}{L - 1 + \sum_k c_k}.$$

Сумма в знаменателе представляет собой количество единиц на выходе слоя сравнения. Заданная величина может быть рассмотрена как "размер" этого вектора. В такой интерпретации "большие" векторы C производят более маленькие величины весов b_{ij} , чем "маленькие" вектора C . Это свойство самомасштабирования делает возможным разделение двух векторов в случае, когда один вектор является поднабором другого, т. е. когда набор единичных компонент одного вектора составляет подмножество единичных компонент другого.

Чтобы проиллюстрировать проблему, которая возникает при отсутствии масштабирования, используемого в данном выражении, предположим, что сеть обучена двум приведенным ниже входным векторам, при этом каждому распределен нейрон в слое распознавания.

Заметим, что X_1 является поднабором X_2 . В отсутствие свойства масштабирования веса b_{ij} и t_{ij} получают значения, идентичные значениям входных векторов. Если начальные значения выбраны равными 1,0, веса образов будут иметь следующие значения: если X прикладывается повторно, оба нейрона в слое распознавания получают одинаковые активации; следовательно, нейрон 2 — ошибочный нейрон — выиграет конкуренцию.

Кроме выполнения некорректной классификации, может быть нарушен процесс обучения. Так как T_2 равно 1 1 1 0 0, только первая единица соответствует единице входного вектора, и C устанавливается в 1 0 0 0 0; критерий сходства удовлетворяется и алгоритм обучения устанавливает вторую и третью единицы векторов T_2 и B_2 в нуль, разрушая запомненный образ.

Масштабирование весов b_{ij} позволяет избежать такого нежелательного течения событий. Предположим, что используется значение $L = 2$, тем самым определяя следующую формулу:

$$b_{ij} = \frac{2c_i}{L - 1 + \sum_k c_k}.$$

Подавая на вход сети вектор X_1 , получим возбуждающее воздействие 1,0 для нейрона 1 в слое распознавания и 1/2 для нейрона 2; таким образом, нейрон 1 (правильный) выиграет соревнование. Аналогично, предъявление вектора X_2 вызовет уровень возбуждения 1,0 для нейрона 1 и 3/2 для нейрона 2, тем самым снова правильно выбирая победителя.

Инициализация весов b_{ij}

Инициализация весов b_{ij} малыми значениями является существенной для корректного функционирования систем АРТ. Если они слишком большие, входной вес вектора, который уже был запомнен, станет скорее активизировать несвязанный нейрон, чем ранее обученный.

Установка этих весов в малые величины гарантирует, что несвязанные нейроны не будут получать возбуждения большего, чем обученные нейроны в слое распознавания. Используя предыдущий пример с $L = 2$, $m = 5$ и $b_{ij} < 1/3$, произвольно установим $b_{ij} = 1/6$. С такими весами предъявление вектора, которому сеть была ранее обучена, приведет к более высокому уровню активации для правильно обученного нейрона в слое распознавания, чем для несвязанного нейрона. Например, для несвязанного нейрона X_1 будет производить возбуждение 1/6, в то время как X_2 будет производить возбуждение 1/2; и то, и другое ниже возбуждения для обученных нейронов.

Поиск. Может показаться, что в описанных алгоритмах отсутствует необходимость фазы поиска, за исключением случая, когда для входного вектора должен быть распределен новый несвязанный нейрон. Это не совсем так: предъявление входного вектора, сходного, но не абсолютно идентичного одному из запомненных образов, может при первом испытании не обеспечить выбор нейрона слоя распознавания с уровнем сходства, большим p , хотя такой нейрон будет существовать; и, тем самым, без поиска не обойтись.

Как и в предыдущем примере, предположим, что сеть обучается следующим двум векторам:

$$X_1 = 1 \ 0 \ 0; \ 0 \ 0$$

$$X_2 = 1 \ 1 \ 1; \ 0 \ 0$$

с векторами весов B_i , обученными следующим образом:

$$B_1 = 1 \ 0 \ 0 \ 0 \ 0$$

$$B_2 = 1/2 \ 1/2 \ 1/2 \ 0 \ 0$$

Теперь приложим входной вектор $X_3 = 11000$. В этом случае возбуждение нейрона 1 в слое распознавания будет 1,0, а нейрона 2 только 2/3. Нейрон 1 выйдет победителем (хотя он не лучшим образом соответствует входному вектору), вектор C получит значение 1 1 0 0 0, S будет равно 1/2. Если уровень сходства установлен в 3/4, нейрон 1 будет заторможен и

нейрон 2 выиграет состязание. C станет равным 1 1 0 0 0, S станет равным 1, критерий сходства будет удовлетворен, и поиск закончится.

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.

Дальнейшее развитие АРТ: архитектуры АРТ-2 и АРТ-3

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.

Заключение

В ходе исследования были выявлены некоторые основополагающие характеристики теории адаптивного резонанса:

- После стабилизации процесса обучения предъявление одного из обучающих векторов будет активизировать требуемый нейрон слоя распознавания без поиска. Эта характеристика прямого доступа определяет быстрый доступ к предварительно изученным образам.
- Процесс поиска является устойчивым. После определения выигравшего нейрона в сети не будет возбуждений других нейронов в результате изменения векторов выхода слоя сравнения С; только сигнал сброса может вызвать такие изменения.
- Процесс обучения является устойчивым. Обучение не будет вызывать переключения с одного возбужденного нейрона слоя распознавания на другой.
- 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.

Идея адаптивного резонанса позволила объяснить Гроссбергу некоторые особенности человеческого восприятия, например задержку в осознании сенсорной информации по сравнению со временем, требуемым для прохождения сигнала по зрительному или слуховому тракту. Эта задержка есть время, необходимое для установления резонанса и зависящее как от силы ожиданий, так и от степени неопределенности воспринимаемой информации.

С помощью адаптивного резонанса можно объяснить и тот факт, что на то, чтобы в первый раз увидеть объект, спрятанный на изображении-головоломке, уходит значительно больше времени, чем на его восприятие в последующие разы. На таких изображениях присутствует объект, складывающийся из каких-то других случайных объектов, но чтобы его увидеть, необходимы заметные усилия со стороны зрительной системы. Для распознавания спрятанного объекта требуется правильно сгруппировать видимые объекты, что требует большого перебора вариантов. Если бы обработка сенсорной информации шла строго снизу вверх, то зрительной системе пришлось бы каждый раз заново решать эту задачу. Второй раз взглянув на то же изображение, человек мог бы помнить, что на нем изображено, но не видеть этого до тех пор, пока нужная комбинация снова не нашлась. Однако распространение информации сверху вниз позволяет эффективно отсеивать неперспективные гипотезы нижних уровней, эффективно направляя поиск правильной интерпретации изображения.

Трудности в теории адаптивного резонанса возникают не только для верхних уровней восприятия. Не решается и проблема инвариантного распознавания.

Речь здесь идет лишь о разрешении неопределенности для зашумленных образов, тогда как даже простое смещение объекта препятствует распознаванию.

Первоначальная простая архитектура нейронной сети в теории адаптивного резонанса впоследствии была сильно усложнена, и некоторые недостатки были устранены, однако принципиальные проблемы, связанные с отсутствием инвариантности и ограничением на число гипотез, решены не были.

Итак, первичная обработка зрительной информации, распознавание зрительных образов, обработка потоков звуковой информации, распознавание речи, управление движением глаз и представление информации в соматосенсорной коре, все эти задачи могут решаться различными типами сетей ART. Из этих результатов следует, что какая-то разновидность автоматического внимания работает, начиная с низких уровней обработки информации мозгом, например, с уровня латерального коленчатого тела. Но одновременно для работы более высоких уровней необходима система ориентирования, которая позволяет гибко переключать внимание и облегчает волевое управление ожиданиями, передаваемыми сверху вниз.

Развитие теоретических исследований ART продолжается. По высказыванию авторов теории, ART 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 ART с успехом применяются в различных областях. ART сделала также важный шаг вперед в общей проблеме моделирования пластично-стабильного восприятия.

Библиографический список

- Carpenter, G.A. and Grossberg, S. (1987). ART 2: Stable self-organization of pattern recognition codes for analog input patterns. *Applied Optics*, 26 , 4919-4930.
- Carpenter, G.A. and Grossberg, S. (1990). ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural Networks*, 3, 129-152.
- Carpenter, G.A., Grossberg, S., and Rosen, D.B. (1991). Fuzzy ART: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural Networks*, 4, 759-771.
- Carpenter, G.A., Grossberg, S., and Reynolds, J.H. (1991). ARTMAP: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network. *Neural Networks*, 4, 565-588.
- Carpenter, G.A., Grossberg, S., and Rosen, D.B. (1991). ART 2-A: An adaptive resonance algorithm for rapid category learning and recognition. *Neural Networks*, 4, 493-504.
- Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., and Rosen, D.B. (1992). Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. *IEEE Transactions on Neural Networks*, 3, 698-713.
- Asfour, Y.R., Carpenter, G.A., Grossberg, S., and Leshner, G.W. (1993). Fusion ARTMAP: A neural network architecture for multi-channel data fusion and classification. In Proceedings of the world congress on neural networks, Portland, II, 210-215. Hillsdale, NJ: Erlbaum Associates.
- Grossberg, S., Govindarajan, K.K., Wyse, L.L., and Cohen, M.A. (2004). ARTSTREAM: A neural network model of auditory scene analysis and source segregation. Technical Report CAS/CNS TR-2003-017, Boston University. *Neural Networks*, 17, 511-536.
- Carpenter, G.A. and Grossberg, S. (2003). Adaptive resonance theory. In M.A. Arbib (Ed.), *The Handbook of Brain Theory and Neural Networks*, Second Edition, Cambridge, MA: MIT Press, 87-90.
- Grossberg, S. (2003). Adaptive resonance theory. Technical Report CAS/CNS TR-2000-024, Boston University. In *The encyclopedia of cognitive science*. London: Macmillan Reference Ltd.

- <https://web.archive.org/web/20120109162743/http://users.visualserver.org/xhudik/art>