



УДК 681.516.52

Калниболотский Ю.М., Арутюнян А.Л.

Национальный технический университет Украины "Киевский политехнический институт".
Украина, Киев

СЕТИ АДАПТИВНОЙ РЕЗОНАНСНОЙ ТЕОРИИ С RBF СЛОЕМ РАСПОЗНАВАНИЯ

Анотація

У статті запропоновано новий тип мереж адаптивного резонансу – мережі ART з мережею радіально-базисної функції у шарі розпізнавання – ART-RBF мереж. Експериментально підтверджені переваги нової ART-RBF парадигми перед аналогами в задачах розпізнавання образів.

Abstract

The description of new type of Adaptive Resonance Networks – ART network with the network of Radial Basis Function in recognition layer – ART-RBF networks is introduced in the article. The advantages of new ART-RBF paradigm over analogs in the tasks of pattern recognition are experimentally confirmed.

Введение

Очевидно, что важным преимуществом нейронных сетей адаптивного резонанса кроме существенного распараллеливания задачи обработки информации является способность самообучаться. Или, другим словами, создавать обобщения: получать обоснованный результат на основании данных, которые сеть (а часто и оператор) не встречали ранее (и в процессе обучения в том числе). Корректное использование обобщающих свойств сети, а также её преимуществ, как: нелинейности, адаптивности, контекстное представление информации, отказоустойчивость, масштабируемость, позволяют решать те задачи, которые на сегодняшний день считаются трудноразрешимыми в рамках классического математического аппарата. Вместе с тем, сеть как таковая, не может обеспечить готовое решение и требует интеграции в сложные системы, которые разбивают глобальную задачу на относительно простые подзадачи, а главное берущие на себя функцию интерпретации полученных результатов и формализации процессов обучения и адаптации.

Сети радиально-базисной функции, благодаря таким свойствам как отсутствие рекурсии и способность к аппроксимации функций, заданных набором шаблонов, являются оптимальным "кандидатом" для построения подобных автоматических самообучающихся систем, но вычислительная сложность расчетов коэффициентов разложения RBF сети, а также высокие требова-

ния, предъявляемые к объему и "качеству" ученика, препятствуют их широкому применению при решении реальных задач.

В данной статье предложен вариант реализации комплексной самообучающейся системы на основе сети адаптивного резонанса (ART) со слоем распознавания на основе нейронов с радиально-базисной функцией ART-RBF, реализующей описанные выше преимущества нейронных сетей и лишённых недостатков RBF сети.

ART-RBF сеть решает задачу поиска субоптимального решения в пространстве меньшей размерности, при этом ответом сети является аппроксимация регуляризованного решения.

1. Использование РБФ сети для поиска регуляризованного решения

В соответствии со своей архитектурой RBF сеть предполагает обучение с "учителем". В задаче аппроксимации функцией $F(x)$, стоящей перед ART-RBF сетью, обучение с учителем является плохо обусловленной задачей, в которой даже большой набор "учебных" примеров может нести малый объем "знаний" о предметной области. Плохо обусловленная задача для области X и диапазона Y , которые связаны фиксированным отображением f , не отвечает хотя бы одному из следующих условий для [1]:

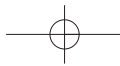
1. Для любого входного вектора $x \in X$ существует выходное значение $y = f(x)$, где $y \in Y$.

2. Для любой пары входных векторов $x, t \in X$ равенство $f(x) = f(t)$ выполняется тогда и только тогда, когда $x = t$.

3. Отображение считается непрерывным, если для любого $\varepsilon > 0$ существует $\delta = \delta(\varepsilon)$, такое, что из условия $\rho_x(x, t) < \delta$ вытекает, что $\rho_y(f(x), f(t)) < \varepsilon$, где $\rho(\dots)$ – расстояние между двумя аргументами в соответствующих пространствах.

Одним из методов решения плохо обусловленных задач является метод регуляризации, предложенный Тихоновым в 1963 году [2]. Метод заключается в использовании при поиске решения дополнительной функции. Предполагается, что данная функция, в дополнение к учебнику (1), будет нести в себе некую априорную информацию о предметной области и о решении.

Метод Тихонова предполагает минимизацию функционала Тихонова - $E(F)$, который, согласно



вышеизложенному, включает в себя два слагаемых:

1. Стандартная ошибка $E_s(F)$, как расстояние между желаемым откликом d_i и фактическим выходным сигналом y_i для примера обучения $i = 1, 2, \dots, N$:

$$E_s(F) = \frac{1}{2} \sum_{i=1}^N [d_i - F(\mathbf{x}_i)]^2, \quad (1)$$

2. Параметр, связанный с регуляризацией $E_c(F)$, который зависит от "геометрических" свойств функции аппроксимации $F(x)$:

$$E_c(F) = \frac{1}{2} \|DF\|^2, \quad (2)$$

здесь D — линейный дифференциальный оператор, несущий априорную информацию о форме решения конкретной задачи. Оператор $\|\cdot\|$ в (2) обозначена норма в функциональном пространстве L_2 из всех действительных функций $f(x)$, $x \in \mathbb{R}^m$ для которых норма $\|f(x)\|^2$ является интегрируемой по Лебегу. $f(x)$ обозначает функцию — генератор пар примеров обучения $\{\{\mathbf{x}_i, d_i\}\}_{i=1}^N$.

Таким образом, необходимо минимизировать $E(F)$:

$$E(F) = E_s(F) + \lambda E_c(F) = \frac{1}{2} \sum_{i=1}^N [d_i - F(\mathbf{x}_i)]^2 + \frac{1}{2} \lambda \|DF\|^2, \quad (3)$$

где λ — параметр регуляризации (при $\lambda \rightarrow 0$, задача является безусловной и имеет решение $F_\lambda(x)$, целиком зависящее от примеров, при $\lambda \rightarrow \infty$, априорного знания заключенного в D достаточно для определения решения).

Решением задачи минимизации функционала Тихонова будет поиск функции $F_\lambda(x)$. Так как первая производная является касательной к кривой, которая дает наилучшую аппроксимацию, для оценки дифференциала предлагается использовать дифференциал Ферше $E(F)$ в роли оптимального локального линейного аппроксиматора в виде:

$$dE(F, h) = \left[\frac{d}{d\beta} E(F + \beta h) \right]_{\beta=0}, \quad (4)$$

где $h(x)$ — фиксированная функция вектора x .

Соответственно равенство 0 выражения (4) по $F(x)$ для всех x позволит отыскать $F(x)$, который будет экстремумом функционала $E(F)$:

$$dE(F, h) = dE_s(F, h) + \lambda dE_c(F, h) = 0, \quad (5)$$

где $dE_s(F, h)$ и $dE_c(F, h)$ — соответствующие дифференциалы Ферше.

Используя ряд математических преобразований представим решение задачи регуляризации в виде:

$$F_\lambda(\mathbf{x}) = \frac{1}{\lambda} \sum_{i=1}^N [d_i - F(\mathbf{x}_i)] G(\mathbf{x}, \mathbf{x}_i), \quad (6)$$

где функция G называется *функцией Грина* для оператора дифференцирования L . $G(x, y)$, в которой параметр вектор x , аргумент вектор y такой, что для оператора линейного дифференцирования L выполняются следующие условия.

1. Для фиксированного y $G(x, y)$ является функцией от x и удовлетворяет граничным условиям.

2. Во всех точках, за исключением $x = y$, все производные $G(x, y)$ по x являются непрерывными. Количество производных определяется порядком оператора L .

3. Если рассматривать $G(x, y)$ как функцию от x , везде, за исключением $x = y$, она удовлетворяет уравнению в частных производных

$$LG(x, y) = 0, \quad (7)$$

или

$$LG(x, y) = \delta(x-y), \quad (8)$$

где $\delta(x-y)$ — дельта-функция Дирака с центром в точке $x = y$.

Исходя из этого, можно утверждать, что решение задачи регуляризации лежит в N -мерном подпространстве пространства гладких функций, а множество функций Грина $\{G(x, x_i)\}$, с центром в x_i , $i = 1, 2, \dots, N$, определяет базис этого подпространства [5].

Выражение (6) можно интерпретировать следующим образом: субоптимальное минимизирующее решение $F_\lambda(x)$ задачи регуляризации является линейной суперпозицией N функций Грина. При этом векторы x_i являются центрами разложения, а веса $[d_i - F(x_i)]/\lambda$ — соответствующими коэффициентами.

Важным моментом является то, что коэффициенты разложения в (6) являются линейными относительно ошибки оценки (разность между желаемым откликом d_i и соответствующим выходным сигналом $F(x_i)$), и обратно пропорциональными параметру регуляризации λ .

Таким образом, для решения плохо-обусловленной задачи регуляризации необходимо построить многослойную нейронную сеть, в которой один из слоев будет осуществлять нейросетевую реализацию функций Грина, а нейроны другого выполнять вычисление линейной функции коэффициентов разложения, соответственно. При такой реализации, веса нейронов будут:

$$w_i = \frac{1}{\lambda} [d_i - F(\mathbf{x}_i)], \quad i = 1, 2, \dots, N. \quad (9)$$



Соответственно (6) в точках $x_j, j = 1, 2, \dots, N$:

$$F_\lambda(x_j) = \sum_{i=1}^N w_i G(x_j, x_i), \quad j = 1, 2, \dots, N. \quad (10)$$

Так как все составляющие (9) и (10) векторы, то перепишем (6) в соответствующей форме,

$$F_\lambda = Gw, \quad \text{где } w = \frac{1}{\lambda}(d - F_\lambda), \quad \text{следовательно} \quad (11)$$

$$(G + \lambda I)w = d, \quad (12)$$

где I — единичная матрица размерности $N \times N$. Матрица G — матрица Грина.

Как было указано в (19), оператор дифференцирования L самосопряженный. Следовательно, соответствующая функция Грина $G(x, x_i)$ и матрица Грина — симметричны.

Для решения (12) для всего "учебника" нейронной сети, т.е. для вычисления весов нейронов, реализующих функцию коэффициентов разложения с учетом симметричности матрицы Грина необходимо произвести выбор соответствующих функций Грина, при которых будет обеспечена несингулярность матрицы G . Согласно теореме Мичелли [6], для $\{x_i\}_{i=1}^N$ множества точек \mathcal{R}^{m_0} матрица интерполяции размерности $N \times N$ с элементами $\varphi_{ji} = \varphi(\|x_j - x_i\|)$ является несингулярной, где $\{\varphi(\|x_j - x_i\|) | i = 1, 2, \dots, N\}$ множество произвольных нелинейных радиально-базисных функций (RBF), $\|\cdot\|$ — Евклидова норма.

К классу радиально базисных функций теорема Мичелли относит такие функции как: мультиквадратичная, обратная мультиквадратичная и функция Гаусса. Последние две являются локализованными, т.е. стремятся к 0, когда норма $\|x_j - x_i\|$ стремится к бесконечности. В этом случае матрица Грина кроме всего прочего будет еще и положительно определенной при условии, что точки x_1, \dots, x_N различны.

На практике всегда можно выбрать значение λ , достаточно большое для обеспечения положительной определенности матрицы $G + \lambda I$ и, таким образом, для возможности ее инвертирования. А это, в свою очередь, значит, что система линейных уравнений (29) будет иметь единственное решение:

$$w = (G + \lambda I)^{-1} d. \quad (13)$$

А решение задачи регуляризации задается следующим разложением:

$$F_\lambda(x) = \sum_{i=1}^N w_i G(x, x_i) \quad (14)$$

Таким образом, регуляризованная задача решается:

1. Разложением решений в множестве функций Грина.

2. Количество функций Грина, равно количеству примеров, используемых при обучении. При этом характеристика функции Грина $G(x, x_i)$ для заданного центра x_i зависит только от формы функции, и для обеспечения инвариантности к преобразованиям необходимо обеспечить зависимость функция Грина $G(x, x_i)$ с центром в x_i только от разности между аргументами x и x_i :

$$G_\lambda(x, x_i) = G(x - x_i), \quad (15)$$

а для инвариантности к преобразованиям и повороту — зависимость от Евклидовой нормы вектора разности $(x - x_i)$, т.е.

$$G_\lambda(x, x_i) = G(\|x - x_i\|), \quad (16)$$

что соответствует радиальной базисной функции, и решение задачи регуляризации принимает частную:

$$F_\lambda(x) = \sum_{i=1}^N w_i G(\|x - x_i\|) \quad (17)$$

К недостатку данного подхода необходимо отнести:

1. Высокую вычислительную сложность расчета коэффициентов разложения;
2. необходимость изучения всего "учебника" нейронной сети заранее и одновременно;
3. локальность полученного субоптимального интерполяционного решения в рамках заданного учебника, а как следствие отсутствие таких важных свойств как пластичности и самообучения;
4. отсутствие механизмов контроля и повышения эффективности работы сети, которое при плохой обусловленности задачи, через низкое качество "учебника", может приводить к низкому качеству результата.

Решить данный комплекс проблем при сохранении интерполяционных качеств RBF нейронного слоя позволяет использование при вычислении коэффициентов разложения подхода, реализованного в сетях адаптивного резонанса.

2. ART-RBF сети

Объединение двух сетей ART и RBF в рамках одной парадигмы дает в первую очередь преимущество в практическом использовании данного инструмента решения задачи отображения в пространство меньшей размерности. Вычислительная сложность поиска весов (32) для вычисления регуляризованного решения всего представленного сети на начальном этапе "учебника", заменяется последовательным изучением учебных примеров с использованием инструментария ART сети.

"Скреживая" такие свойствам RBF сетей, как отсутствие рекурсии и способность к аппроксимации функций, которые заданы в неявном виде

набором шаблонов (обучающих образов), и ART сетей — стабильность-пластичность, возможность самообучения и интерпретации полученных результатов, удалось сократить вычислительные затраты традиционного последовательного персонального компьютера, "стоимость" хранения памяти сети, а также ускорить и упростить процесс обучения при сохранении положительных свойств обеих сетей.

В соответствии с алгоритмом кластеризации, мы определяем ART-RBF кластеризующую систему как исполнительную систему, где номер нейрона выходного подслоя РБФ слоя распознавания в каждый момент времени является решением (если таковые вообще имеются), в то время как мера между входным вектором и центром гиперсферы нейрона с максимальным значением функции активации скрытого подслоя в РБФ слое распознавания подвергается ограничению порогом ρ (рис. 1).

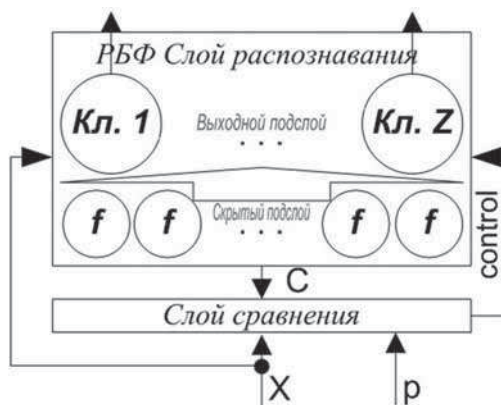


Рис. 1. ART-RBF сеть

В качестве слоя сравнения новой сети будет использован такой слой, что результатом его работы будет λ , равное отношению Евклидова расстояния между входным вектором и центром гиперсферы нейрона победителя слоя сравнения RBF слоя распознавания к N — размерности входного вектора. Если значение $1-\lambda$ выше порога, то считается, что условие распознавания выполнено. Так как значение каждого компонента N мерного входного вектора $x_i \in [0; 1]$, то результат работы слоя сравнения всегда менее или равен 1. Порог задаётся в пределах $0 \leq \rho \leq 1$, где $\rho = 1$ соответствует точному совпадению.

В качестве слоя распознавания выберем такую RBF сеть, в которой ширина окна нейронов скрытого слоя может быть различной для каждого нейрона, а также связь между скрытым и выходным слоем не является полной. Данная

архитектура не в полной мере соответствует описанной в предыдущем разделе, так как не предполагает поиск универсального решения (17) и вычисления (13) для всего "учебника", что существенно снижает вычислительную сложность процесса обучения. При этом поиск универсального регуляризованного решения (17) для всего доступного в данный момент времени учебника заменяется на изучение отдельных, представленных на вход сети в данный момент времени учебных примеров. Такой подход позволяет избежать сложных вычислений, связанных с вычислением симметричной матрицы Грина и приводит решение каждого учебного примера к решению (17) в пределах ограниченных размерности входного пространства сети. Кроме того, именно такой подход, через использование доступного инструментария сетей адаптивного резонанса, позволяет избежать субоптимальности интерполяционного решения в рамках доступного учебника и в полной мере раскрыть способность ART-RBF сети к обучению.

Дополнительно, с целью упрощения реализации параллельной архитектуры нейронных сетей на современных последовательных вычислительных машинах, в архитектуре слоя распознавания ART-RBF сети связи установлены только для тех нейронов скрытого слоя, центры гиперсфер которых принадлежат классу, который обозначает соответствующий нейрон выходного слоя. Нейрон выходного слоя вычисляет взвешенную сумму всех соединенных с ним выходов нейронов скрытого слоя.

В таком случае сама нейронная сеть RBF имеет архитектуру с передачей сигнала от входа к выходу. Каждый нейрон скрытого слоя вычисляет функцию активации от входных данных и характеризуется двумя важными параметрами: 1) центр узла в гиперпространстве и 2) ширина его радиальной функции. В скрытом слое i -й нейрон обеспечивает максимальное значение функции активации, когда входной вектор x близок к вектору его центра c_i . С учетом (16) для оценки расстояние d до центра — Евклидово расстояние, а функция Грина — как функция активации скрытого узла, с учетом теоремы Мичелли выбирается как Гауссова функция.

$$d = \|\bar{x} - \bar{c}_i\| = \sqrt{\sum_{j=1}^N (x_j - c_{ij})^2}, \quad (18)$$

$$f(\bar{x}, \bar{c}) = e^{-\frac{d^2}{2\sigma_i^2}}, \quad (19)$$

где σ_i параметр, описывающий ширину радиальной базисной функции (физический смысл данного параметра — дисперсия шума входного сигнала).

Для обучения скрытого слоя используется набор обучающих примеров, на базе которых строится семейство функций Грина, с центрами в точках пространства которые определены набором тренировочных шаблонов.

$$\bar{c}_j = \bar{x}_j, j = \overline{1, H}, \quad (20)$$

Ширина окон активационных функций нейронов скрытого слоя выбирается:

1) или достаточно большой, но так, чтобы они не накладывались одна на другую в пространстве входных сигналов сети, при этом обеспечивается минимально возможное значение среднеквадратичной ошибки при количестве примеров, стремящемся к бесконечности;

2) или в соответствии с предварительно оцененной дисперсией шума входных сигналов, если есть уверенность, что она не будет меняться в ходе использования сети (например нормальное распределение в задачах контроля качества).

Обученная таким образом сеть гарантирует, что сгенерированная ею гиперповерхность пройдет через тренировочные шаблоны в пространстве входных образов, при этом каждый нейрон выходного слоя обозначает один класс. Соответственно нейрон с максимальным значением — т.е. тот, у которого максимально значение RBF функции или минимально расстояние между центром нейрона и входным вектором, будет указывать на принадлежность входного вектора определенному классу.

Если трансформировать задачу отображения в пространство меньшей размерности как задачу прикладного кластерного анализа, то необходимо отметить, что полное покрытие области пространства входных векторов не требуется, так как при кластеризации группа объектов разбита на множество более или менее гомогенных подгрупп на основе часто субъективно выбранной меры подобия, такой, что подобие между объектами в пределах подгруппы является большим, чем подобие между объектами, принадлежащими различным подгруппам. При этом сама группа объектов в прикладных задачах имеет те или иные пределы и ограничения. Цель кластеризации в данном случае состоит в том, чтобы группировать данные, которые представлены сейчас, а не обеспечить равномерное заполнение всего пространства наблюдаемых и ненаблюдаемых (будущих) выборок, сгенерированных от того же самого распределения вероятности. Вместе с тем при практическом использовании ART-RBF сети с течением времени входная выборка так или иначе будет увеличиваться и преодолевать установленные заранее пределы, при этом важна качественная реализация такого свойства сети ART как стабильность-пластичность, которое обеспечит расшире-

ние "границ" предметной области при сохранении детализированной информации о уже изученных знаниях. Недостатком здесь будет опасность резкого роста объема памяти сети, вместе с ростом изученных примеров, так как в решении задачи регуляризации количество функций Грина соответствует количеству учебных примеров.

Возвращаясь к построению учебника для плохо обусловленной задачи надо отметить, что качественный учебник должен содержать наиболее "яркие" примеры предметной области и отражать её суть и особенности, а не покрывать её полностью. В таком случае, при существенном расширении области знаний нейронной сети разумным предполагается внедрение механизма систематизации знаний и сокращения количества нейронов сети RBF, отвечающих за принадлежность одному классу гиперсферы, или другими словами, механизма пересмотра и систематизации учебных примеров на основании новых знаний сети, полученных в процессе работы. При этом в первую очередь под сокращения должны попасть те учебные примеры, которые отражают одни и те же классы и находятся в пределах естественного шума входного сигнала.

Правило сокращения сформулировано следующим образом. Две близко расположенные гиперсферы, описанные функциями Грина двух нейронов RBF сети, принадлежащие одному кластеру, с центрами, обозначенными N мерными векторами c_1, c_2 и окнами σ_1, σ_2 , могут быть объединены в одну гиперсферу (и замещены одним нейроном) с центром c_3 и окном σ_3 , вычисленным следующим образом:

$$\bar{c}_3 = (c_{3,1}, \dots, c_{3,N}), \quad \bar{c}_{3,i} = (\bar{c}_{1,i} + \bar{c}_{2,i}) / 2 \quad i = \overline{1, N}, \quad (21)$$

$$\sigma_3 = d(\bar{c}_1, \bar{c}_2) / 2 + \max(\sigma_1, \sigma_2), \quad (22)$$

при выполнении ограничений, наложенных на ширину окон активационных функций.

В противовес вычислительно сложному механизму расчета коэффициентов разложения RBF сети, использован интуитивно понятный механизм связи слоя распознавания ART сетей. Связь между нейроном скрытого подслоя и выходного подслоя ART-RBF слоя распознавания устанавливается равным "1", если данная гиперсфера принадлежит классу, и равным "0", если нет. Таким образом, гиперсферы объединяются в гипергрозди, поверхность которых описывает класс. Активационная функция нейрона выходного слоя вычисляется следующим образом:

$$f = \begin{cases} 1, & \sum w f(\bar{x}, \bar{c}) > e^{-\gamma^2} \\ 0, & \sum w f(\bar{x}, \bar{c}) < e^{-\gamma^2} \end{cases} \quad (23)$$

где $1 \leq \gamma \leq 3$ [7]. Отметим, что в прикладных задачах может быть использована и другая мера, например, на основе нормального распределения или априорно установленного дополнительного порога.

3. Сравнение способности к распознаванию образов ART, RBF и ART-RBF сетей

Геометрическая интерпретация внесенных изменений в расчет функций активации в сравнении с парадигмой ART сети приведена на рис. 2. Графики построены с учетом параметров (порог функции сравнения, ширина окон RBF функции) описанного ниже эксперимента для одного элемента входного вектора.

На графике горизонтальной линией отображен порог функции слоя сравнения ART сети MF на уровне 0,7. При этом функция активации классической ART сети $AF(x)$ превращается в "столбик", покрывающий область, в которой расстояние между центром нейрона и входным образом меньше порогового. При этом демонстрируется явное игнорирование физических процессов реальной предметной области, выраженных через нормальное распределение (линия $ARTRBF1(x)$ описывает нормальное распределение при рассчитанной в эксперименте дисперсии шума и математическом ожидании 0).

Линиями $ARTRBF1(x)$ и $ARTRBF2(x)$ описаны функции активации двух нейронов RBF слоя ART-RBF сети. Линия $ARTRBF3(x)$ соответ-

венно отражает функцию активации ново-созданного нейрона после применения правила сокращения количества нейронов к данным нейронам. Как видно из рис. 2, сеть при этом будет давать устойчивое решение в соответствии с установленным порогом для тех 2-х примеров задачника, на основании которых были построены $ARTRBF1$ и $ARTRBF2$ нейроны при вдвое меньших затратах на хранение памяти, а как следствие и уменьшение вычислительной сложности поиска решения.

Геометрическая интерпретация механизма построения гипергрозды нейронов выходного слоя приведена на рис. 3.

Для двух нейронов $ARTRBF1$ и $ARTRBF2$, принадлежащих одному классу, выходным слоем нейронной сети будет рассчитана следующая функция активации $ARTRBF4$, как суперпозиция функций активации каждого из нейронов. В данном случае сеть проявляет свойство обобщения — предлагая отнести к данному классу область, находящуюся между двумя нейронами, не охваченную функциями активации. При этом, исходя из абсолютного значения функции активации, оператор предупрежден о меньшей вероятности принадлежности входного вектора из данной области к классу, описанному $ARTRBF4$.

Необходимо отметить важность выбора параметра ширины окна RBF функции, как параметра, учитывающего дисперсию шума. Как видно на рис. 2 и рис. 3 этот параметр, управляя шириной окна в паре с высотой порога может быть выбран

таким образом, чтобы сеть повторила классическую ART сеть. Конечно, при этом будет утерян смысл использования RBF слоя распознавания. Экспериментально было установлено, что при наличии большой выборки ширину окна необходимо устанавливать на уровне дисперсии шума учебных примеров принадлежащих одному классу. Если же такой расчет не представляется возможным, необходимо исходить из эмпирически установленного соотношения и выбирать ширину окна в пределах от $1/3$ до 1 расстояния между двумя ближайшими векторами разных классов обучающей выборки.

В дальнейшем после обучения требуется дополнительная настройка сети с учетом анализа реальной предметной области. В случае, когда такой анализ не может быть проведен, оператор должен, управляя уже уровнем

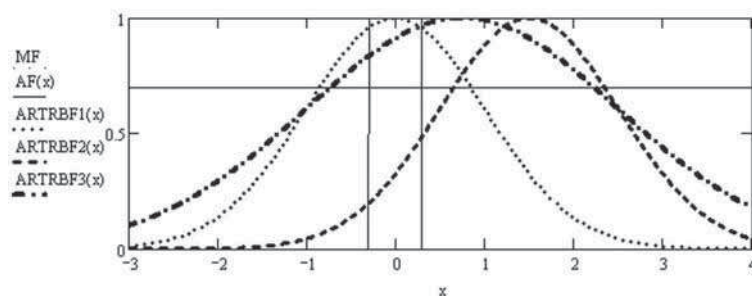


Рис. 2. Функции активации

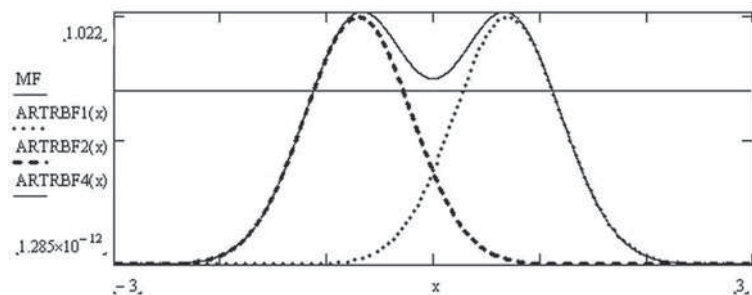


Рис. 3. Функции активации гипергрозды

порога слоя сравнения, добиться корректной работы сети на обучающей выборке.

4. Описание эксперимента

Новый тип сетей адаптивного резонанса, объединяя свойства сетей адаптивного резонанса и сетей с радиально базисной функцией, позволяет получить более высокое быстродействие при меньших затратах объемов "памяти" в сравнении с традиционными реализациями своих прототипов. Но главное, исходя из заложенной в основу ART-RBF сети архитектуры, данная сеть существенно повышает качество кластеризации входных данных при равных прочих условиях. С целью подтверждения данного предположения был проведен эксперимент, в котором сравнивались свойства сетей с традиционной архитектурой и новой.

В эксперименте были использованные данные эксперимента, проведенного ранее среди студентов третьего курса НТУУ "КПИ", факультета электроники, кафедры конструирования и производства ЭВА, прошедших курс искусственного интеллекта. Целью эксперимента была проверка способности ART сети к распознаванию зашумленных и искаженных образов. Был разработан учебник, состоящий из массива векторов как эталонов классов, а также массив тестовых примеров, полученный искажением эталонов.

Аналогичный эксперимент был повторен для оценки свойств ART-RBF сети на том же массиве данных. Полученные результаты сравнивались с данными ART сети, полученными в ходе предыдущего эксперимента.

Перед проведением эксперимента были высказаны следующие предположения в отношении уровня функции активации. При равном уровне порога слоя сравнения ART и ART-RBF сети, установленном на уровне $1-e^{-1}$, качество работы ART-RBF сети будет зависеть от точности прогнозирования дисперсии шума обучающей выборки, так для:

1. Ширины окна, в точности соответствующей дисперсии шума, результаты работы ART-RBF сети будут незначительно выше в сравнении с ART сетью, что связано с гладкостью функции аппроксимации ART-RBF сети.

2. Ширины окна, установленного по принципу полного покрытия пространства учебных примеров, т.е. на уровне расстояния между двумя ближайшими центрами RBF функций нейронов, принадлежащих разным классам, ART-RBF сеть будет показывать функции активации

для любых поданных на вход векторов не ниже 0,183. В то же время для образов, устойчиво распознаваемых ART сетью, для заданного выше порога и установленной дисперсии шума учебника, функция активации будет не ниже 0,991704.

3. Независимо от ширины выбранного окна объем памяти, требуемый ART-RBF сетью будет соответствовать количеству изученных классов, в противовес ART сети, объем памяти которой будет соответствовать количеству изученных примеров.

Результаты, полученные в ходе эксперимента, в графическом виде представлены на рис. 4. Значения функции активации для ART-RBF сети для различных уровней зашумления лежат в пределах (0,986;1). Для сравнения, также приведены результаты распознавания зашумленных изображений сетью ART и респондентами.

Эксперимент подтвердил способность ART-RBF сети к распознаванию зашумленных образов. В частности сеть не допустила ни одной ошибки, каждый раз сопоставляя зашумленный образ с его эталоном. Эксперимент показал, что в распознавании малозашумленных образов человек конкурирует с сетью ART и RBF. При сильных шумах сеть выходит на "первое место" и уверенно распознаёт даже те образы, которые были подвергнуты многократному искажению и стали абсолютно непригодными для распознавания человеком.

Эксперимент также полностью подтвердил высказанные предположения для различных выборов ширины окна (табл. 1).

Сравнивая результаты ART, RBF и ART-RBF сетей мы можем видеть убедительное преимущество использования ART-RBF парадигмы, проявляющееся в синергетическом эффекте объединения в одной модели двух высокоэффективных парадигм нейронных сетей.

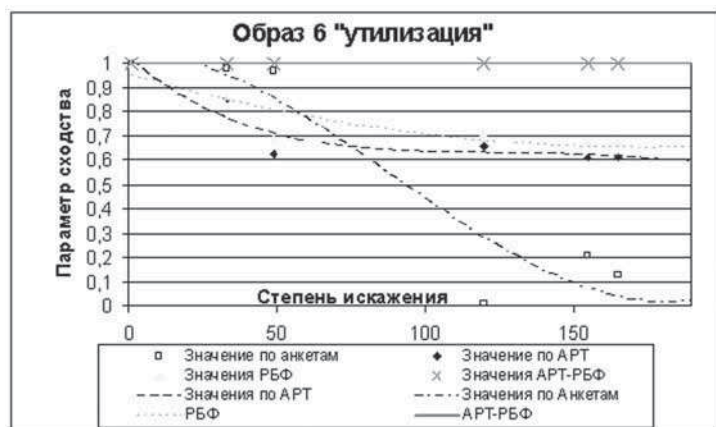


Рис. 4. Зависимость параметра сходимости от степени искажения

Функции активации (образ №6 "Утилизация")

Вектор	Степень искажения	Функция активации		
		ART	ART-RBF Ширина окна соответствует дисперсии	ART-RBF Ширина окна соответствует полному покрытию
<i>Предположение</i>			<i>Незначительно выше чем у ART</i>	<i>Не ниже 0,991704</i>
Utiliz	0	1	1	1
Utiliz 12 3 33	33	0,728333333	0,794030281	0,990817076
Utiliz 6 3 49	49	0,718055556	0,780036479	0,990112621
Utiliz 18 3 155	155	0,660555556	0,697627205	0,985700408
Utiliz 28 3 165	165	0,668055556	0,708691667	0,98632103

Выводы

Описанная в данной статье архитектура ART-RBF сети реализует модель самообучающейся системы на основе сети адаптивного резонанса обладающей свойствам аппроксимации RBF сетей, но лишенной вычислительной сложности и локальности применения присущей последним. Существенным преимуществом является также адаптация модели и алгоритма функционирования требованиям современных последовательных ЭВМ.

Представленная сеть позволяет осуществлять реализацию быстрых, высокоточных и не требовательных к ресурсам систем кластеризации входных данных на базе последовательных ЭВМ для решения широкого спектра задач. В том числе внедрение таких систем было успешно проведено в таких сферах народного хозяйства как: машиностроение (системы входного контроля и контроля качества), мобильная связь (системы учета абонентов), информационные технологии (системы распознавания атак), сертификация продукции и услуг (системы независимого контроля) и другие.

Проведенный эксперимент продемонстрировал эффективность новой модели сети в сравнении с предыдущими аналогами ART и RBF. Кроме способности решения задачи поиска субоптимального решения в пространстве меньшей размерности через аппроксимацию регуляризованного решения, полученного от сети RBF, сеть ART-RBF обладает такими преимуществами:

1. ART-RBF сеть за счет адаптивной ширины окон активации нейронов скрытого подслоя, а также механизма создания "гипергрозди" способна

Таблица 1

к кластеризации гиперпространства областями, описываемыми любой поверхностью (суперпозиция решений нескольких локальных задачи регуляризации), сохраняя при этом низкую "стоимость" хранения памяти и возможность решения проблемы стабильности-пластичности сети.

2. Благодаря автоматической системе формирования связи нейронов скрытого и выходного слоя, а также адаптивным механизмам функционирования, исключается вмешательство оператора и привнесение субъективной оценки. Уменьшается вычислительная сложность.

3. Абсолютные значения функции активации нейронов скрытого подслоя являются реальной мерой подобия принадлежности к определенному классу, а не внутренним параметром сети.

В таблице 2 дана сравнительная характеристика различных типов сетей в сравнении с ART-RBF.

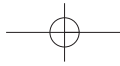
Таблица 2

Характеристики нейронных сетей

Однослойный перцептрон	ART-RBF	Многослойный перцептрон
1 скрытый слой	1 скрытый слой	Несколько скрытых слоев
Нейроны однотипны	В разных слоях разные нейроны	Нейроны однотипны
	Функции активации в нейронах: скрытого слоя — нелинейные; выходного слоя — линейные	Функции активации в нейронах нелинейные
	Аргумент функции активации — Евклидова норма	Аргумент функции активации — Скалярное произведение
Решаемые задачи		
Линейные преобразования	Локальная аппроксимация, нелинейное отображение	Глобальная аппроксимация, нелинейное отображение

Литература

1. *Kirsch A.* An Introduction to the Mathematical Theory of Inverse Problems. — New York: Springer-Verlag, 1996.
2. *Debnath L. and P. Mikusinski.* Introduction to Hilbert Spaces with Applications. — New York: Academic Press, 1990.



3. *Lancoz C.* Linear Differential Operators. —London: Van Nostrand, 1064.
4. *Courant R. And D. Hilbert.* Methods of Mathematical Physics. — 1970, vol.I and II, New York: Wiley Interscience.
5. *Poggio T. and F. Girosi.* "Networks for approximation and learning", Proceedings of the IEEE, 1990, vol. 78, p. 1481–1497.
6. *Micchelli C.A.* "Interpolation of scattered data: Distance matrices and conditionally positive definite functions", Constructive Aproximation, 1986, vol. 2, p. 11–22.
7. *Musavial M.T. et al.*, "On the training of radial basis function classifiers", Neural Networks, vol. 5, pp. 595–603, 1992.
8. *Арутюнян А.Л.* "Электроника и связь" №20, 2003, "Современные реализации нейронных сетей адаптивного резонанса".
9. *Yang F., Paindavoine M.* "Implementation of an RBF Neural Network on Embedded Systems: Real-Time Face Tracking and Identity Verification" IEEE Trans. Neural Networks, vol. 14, no. 5, pp. 1162-1175, September 2003.
10. *Арутюнян А.Л.* "Сети адаптивной резонансной теории с РБФ слоем распознавания", Сборник трудов "Интеллектуальный анализ информации", рос.-укр. науч. семинар, 2004 г. Киев.

