

5. Кустов В.Ф. Основы теории надёжности та функційної безпечності систем залізничної автоматики: [навчальний посібник] / В.Ф. Кустов. – Харків: УкрДАЗТ, 2008. – 212 с.
6. Комплексы технических средств систем керування та регулювання руху поїздів. Функційна безпечність і надійність. Вимоги та методи випробовування: ДСТУ 4178-2003. - [чинний від 2003-07-01]. – К.: Держстандарт України, 2003. – 32с. – (Національний стандарт України).
7. Попов Г.В. Выбор решений и безопасность: [учеб. пособие] / Г.В. Попов. – Иваново: Иван. гос. энерг. ун-т. – 2003. – 92 с.
8. Шутюк С.В. Формирование интегрального показателя априорной состоятельности региона для компании ОАО «РЖД» / С.В. Шутюк // Вестник ВНИИЖТ. 2005. - № 6. С. 5-10.
9. Системный анализ и принятие решений: словарь-справочник / [авт.-сост. Волкова В.Н.]. – М.: Высшая школа, 2004. – 200 с.
10. Смирнов Э.А. Разработка управленческих решений / Смирнов Э.А. – М.: ЮНИТИ-ДАНА, 2002 – 180 с.
11. Корн Г. Справочник по математике (для научных работников и инженеров) / Г. Корн, Т. Корн. - М.: Наука, 1974.- 832 с.
12. Огар О.М. Напрямки удосконалення конструкцій гіркових горловин сортувальних пристроїв з позиції ресурсозбереження / О.М. Огар, О.В. Розсоха // Східно-Європейський журнал передових технологій. Харків, 2007. № 5/2(29). С. 54-58.

УДК 681.3

Запропоновано метод, заснований на використанні системи штучних нейронних мереж з фрактальною структурою, який дозволяє поліпшити якість рішення задач класифікації

Ключові слова: нейронна мережа, метод синтезу, класифікація

Предложен метод, основанный на использовании системы искусственных нейронных сетей с фрактальной структурой, позволяющий улучшить качество решения задач классификации

Ключевые слова: нейронная сеть, метод синтеза, классификация

The method based on the use of system of artificial neuron networks with a fractal structure is offered. It allows to improve the quality of decision of the tasks of classification

Keywords: neuron network, method of synthesis, classification

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ РЕШЕНИЯ ЗАДАЧ КЛАССИФИКАЦИИ НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ

Д. А. Юрков

Кандидат технических наук, доцент
Кафедра автоматизации и компьютерно-интегрированных технологий
Восточноукраинский национальный университет имени
Владимира Даля
ул. Тухачевского, 11, г. Луганск, Украина, 91050
Контактный тел.: 050-475-97-59
E-mail: rusdma@rambler.ru

1. Введение

Многие проблемы управления, с которыми сталкиваются в настоящее время инженеры и исследователи, подразумевают решение задач идентификации для сложных, нелинейных и многосвязных объектов. Получение моделей таких объектов хорошо разработанными и описанными методами синтеза [1,2] сопряжено с рядом принципиальных трудностей, среди которых следует отметить:

1) неопределённость и недостаточность знаний об исследуемой системе, получение которых сопряжено со значительными трудностями экономического характера или же вопросами безопасности;

2) полученные классическими методами модели могут быть чрезмерно сложны в практическом использовании или же не могут быть получены вовсе;

3) значительная часть информации об исследуемой системе может находиться в виде экспертных данных или в виде эвристического описания процессов функ-

ционирования, что затрудняет её выражение математическими зависимостями.

Указанные трудности получения моделей встречаются в практике управления объектов обусловили появление ряда новых подходов к решению подобных трудно формализуемых задач. Одним из таких перспективных и многообещающих подходов является использование специальных структур, названных искусственными нейронными сетями (ИНС) и состоящих из элементов, имитирующих основные свойства своих прототипов – биологических нейронов.

2. Актуальность проблемы

Необходимо отметить, что использование ИНС во многих случаях позволяет обойти или сгладить проблемы, связанные с указанными выше трудностями. Однако, синтез самих нейросетевых моделей изучаемого объекта представляет собой достаточно сложную самостоятельную задачу, описанию и решению которой посвящены многочисленные работы специалистов в области управления.

С математической точки зрения синтез нейронной сети рассматривается как решение некоторой оптимизационной задачи – поиска такого состояния сети, в котором минимизируется некая функция ошибки ε , вычисляемая на основании имеющегося в распоряжении исследователя множества данных (обучающей выборки), которую можно представить в виде [3,4]:

$$\varepsilon = \sum_{j=1}^k (\Omega_j(W^t, z^j) - y^j)^2, \quad (1)$$

где W – состояние сети (все весовые коэффициенты всех нейронов, входящих в сеть); W^t – текущее состояние сети; $\Omega: z \Rightarrow y$ – отображение, реализуемое нейронной сетью; k – количество примеров в обучающей выборке; y^j – требуемое значение выхода для j -го примера; z^j – входной вектор для j -го примера. Таким образом, ошибка обучения сети ε , появляющаяся в некотором состоянии W^t , представлена как сумма квадратов ошибок для всей обучающей выборки. Если сеть работает безошибочно, то $\varepsilon = 0$, а для обученной сети $\varepsilon \rightarrow \min$. Таким образом, задача обучения нейронной сети сводится к задаче поиска экстремума-минимума функции ошибки (1) в пространстве состояний W .

Однако, несмотря на многочисленные и хорошо разработанные методы оптимизации, получаемые в форме ИНС модели изучаемого объекта не всегда адекватны и приемлемы для практического использования. Основная причина такой ситуации очевидна – модель синтезируется на основании обучающего множества данных, которое в подавляющем большинстве случаев, и особенно в условиях малой выборки, не отражает всех характеристик исследуемой системы. Фактически, обучающая выборка не всегда достаточно информативна.

Другой важной причиной, по которой получаемые нейросетевые модели не всегда хорошо отражают действительность, являются недостатки самих алгоритмов синтеза. Например, широко известный алгоритм обратного распространения ошибки (Back Propagation), фактически является итеративным градиентным ал-

горитмом обучения многослойной нейронной сети без обратных связей, т.е. гарантирует поиск локального, а не глобального, экстремума функции ошибки. Существуют также многочисленные модификации данного алгоритма, подробно изложенные в работах [3,4], направленные на увеличение скорости его работы и улучшение некоторых других характеристик. Однако, предлагаемые различными исследователями модификации в основном позволяют, и довольно значительно, улучшить различные характеристики только самого алгоритма, реализующего метод синтеза. Однако, при этом получаемой модели присущи те же недостатки, о которых было сказано ранее.

Ещё одной проблемой синтеза ИНС является выбор структуры сети для решения конкретной задачи, основная суть которой заключается в следующем. Более полувека назад известный русский учёный А.Н. Колмогоров показал возможность представления непрерывных функций нескольких переменных в виде суперпозиции непрерывных функций одного переменного и сложения [5]. На результатах этой работы базируются практически все последующие работы, связанные с теоретическим анализом нейронных сетей. В частности, в своей известной работе [6] Хехт-Нильсен обобщил полученный А.Н. Колмогоровым результат применительно к нейронным сетям. Основным результатом этого обобщения – возможность реализации произвольной функции нейронной сетью [3,4]. В этой работе он показал, что для любой обучающей выборки существует двухслойная однородная (с одинаковыми функциями активации каждого нейрона) нейронная сеть первого порядка с последовательными связями и с конечным числом нейронов, которая реализует отображение Ω , выдавая на каждый входной сигнал правильный выходной сигнал. Фактически, отображение Ω может быть построено уже с помощью двухслойной нейронной сети. Проблема заключается в том, что теорема Хехт-Нильсена не может использоваться в качестве метода синтеза нейронной сети и является только лишь теоремой существования. В ней не заложена методика определения числа нейронов в сети и весов связей между ними для некоторой конкретной обучающей выборки.

Подобная ситуация приводит к тому, что при синтезе сети разработчик должен решить ряд задач, среди которых необходимо выделить следующие:

- 1) выбрать вид соединения нейронов между собой внутри сети, т.е. определить её топологию;
- 2) определить общее число нейронов в сети и распределить их по слоям;
- 3) определить функции активации для каждого нейрона в сети;
- 4) доказать работоспособность полученной ИНС.

Если решение последних двух задач во многих случаях не вызывает затруднений, то решение первых двух зависит как от наличия априорной информации о моделируемой системе (что далеко не всегда имеет место), так и от опыта, интуиции и искусства разработчика.

Такая ситуация обуславливает необходимость разработки новых методов синтеза ИНС, которые могут обеспечить требуемое качество обучения и должны иметь преимущества над методами существующими, а также удовлетворять требованиям практического использования – времени обучения, простоте реали-

зации, ограничениям по вычислительным ресурсам и т.п. Это – актуальная проблема, не имеющая на данный момент удовлетворительного для ряда задач решения.

3. Цели и задачи

Принимая во внимание актуальность рассматриваемой проблемы, а также связанные с ней и рассмотренные выше недостатки существующих подходов к получению нейросетевых моделей, определим цель данной работы как разработку новых методов синтеза ИНС, позволяющих избежать указанных трудностей и формализовать процесс синтеза.

Решение ряда задач синтеза, в частности синтез топологии сети рассмотрен ранее [7,8] автором данной работы на примере задачи классификации, в результате чего предложен метод синтеза топологии сети, позволяющий формировать её структуру в процессе итерационного вычисления самоподобной логической функции – метод формирования фрактальной структуры (ФФС). Основным достоинством предложенного метода является практическая возможность обнаружения глобального экстремума-минимума функции ошибки (1) для любой обучающей выборки.

Подход, предложенный в работах [7,8], позволяет снять проблему определения межнейронных соединений внутри сети и выбора числа используемых нейронов, отличается простотой реализации, высокой скоростью работы, наглядностью и очевидной физической интерпретацией получаемого результата. В дополнение к этому, метод представляет собой чётко формализованный на основе математической модели процесс синтеза, в результате которого гарантированно достигается экстремум-минимум функции ошибки (1).

Однако, несмотря на ряд преимуществ, данный метод решает только проблему синтеза топологии сети и обнаружение минимума функции ошибки. К сожалению, для практического использования получаемых методом ФФС сетей этого недостаточно. Дело в том, что получаемая при этом ИНС достаточно грубо аппроксимирует границу области пространства, занимаемую искомым классом, что приводит к значительному числу ошибок при использовании сети для данных, не входящих в обучающую выборку. В связи с этим автором данной работы предложено развитие данного метода [9] на основе объединения всех отдельно полученных методом ФФС сетей в единую структуру, которая названа кооперативной нейросетевой структурой. Каждая входящая в структуру сеть вносит свою равнозначную часть в решение о принадлежности предъявляемого образа искомому классу, а отдельный нейрон, реализующий функцию объединения, вычисляет долю положительных решений и сравнивает её с пороговым значением.

Такой подход обладает рядом достоинств, что позволяет значительно улучшить качество и достоверность классификации. Другим важным достоинством метода является возможность его успешного использования для малой обучающей выборки, что подробно изложено в работе [9].

Однако, в некоторых случаях граница области, занимаемой искомым классом, также получается не-

допустимо изрезанной, что иногда приводит к невозможности использовать полученную ИНС в практических целях, особенно в случаях, когда требуется классификация образов, находящихся близко к этой границе. В связи с этим становится очевидной необходимостью дальнейшее совершенствование предложенных ранее методов и их практическая проверка, что будем рассматривать как основные задачи данной работы.

4. Развитие метода

Результатом использования метода, описанного в работах [7,8], является структура, состоящая из нейронов двух типов.

1. Первый тип нейрона, принадлежащий всегда первому слою нейронной сети, реализует разделение элементов выборки гиперплоскостью.

2. Второй тип нейрона, реализующий вычисление логической функции специального вида:

$$L(x_i) = \gamma(x_i) \cdot \varphi(x_i) + \overline{\gamma(x_i)} \cdot \psi(x_i), \quad (2)$$

где x_i – входной вектор, γ – результат нейрона первого типа, φ и ψ – функции ошибок.

Таким образом, первый слой получаемой сети состоит из нейронов с пороговой функцией активации, все последующие слои состоят из специального «логического» нейрона, который возможно представить в нейросетевом базисе – операций взвешенного суммирования нескольких входных сигналов и последующее преобразование полученной суммы с использованием функции активации.

Действительно, представим логическую функцию (2) в виде элементарных логических операций – логического «И», логического «ИЛИ» и логического «НЕ». В этом случае логический нейрон можно представить в виде простейшей нейросетевой структуры, показанной на рис. 1. Структура содержит два нейрона, реализующих логическую функцию «И», один нейрон, реализующий логическую функцию «ИЛИ» и один нейрон, реализующий логическую функцию «НЕ». Таким образом, возможность реализации логического нейрона в нейросетевом базисе сводится к возможности представить указанные элементарные логические операции в этом базисе.

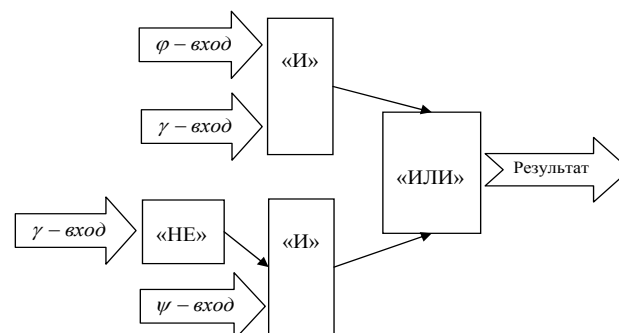


Рис. 1. Представление логического нейрона в виде сети с нейронами, реализующими элементарные логические операции

Рассмотрим способы реализации элементарных логических операций, используя в качестве активационной функции нейрона пороговую функцию $\Phi(x)$ - функцию Хевисайда:

$$\Phi(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases} \quad (3)$$

Представимость логической операции «И» для двух входов в нейросетевом базисе сводится к существованию коэффициентов w_i таких, что выполняется неравенство:

$$w_1\gamma + w_2\varphi + w_0 \geq 0. \quad (4)$$

Очевидно, что для логического «И» можно подобрать бесконечное число решений, однако мы примем в качестве решений следующие значения коэффициентов: $w_0 = -1$; $w_1 = 1/2$; $w_2 = 1/2$. Легко проверить, что при таких значениях w_i для значений $\gamma = 1$ и $\varphi = 1$ активационная функция $\Phi(x)$ также равна единице, что соответствует таблице истинности для логического «И».

Аналогичным образом можно представить и логическую операцию «ИЛИ», для чего значения коэффициентов можно принять следующими: $w_0 = -1$; $w_1 = 1$; $w_2 = 1$. В этом случае активационная функция $\Phi(x)$ равна единице, если $\gamma = 1$ или $\varphi = 1$, что соответствует таблице истинности для логического «ИЛИ».

Для логической операции «НЕ» выражение (4) записывается в виде:

$$w_1\gamma \geq 0. \quad (5)$$

Для реализации логического «НЕ» примем значения коэффициента $w_1 = -1$. Очевидно, что значение активационной функции $\Phi(x)$ противоположно значению входа, т.е. реализуется требуемая логическая операция.

Обобщая выше изложенное, можно записать функцию логического нейрона в нейросетевом базисе:

$$y = \Phi \left[\Phi \left(\frac{\gamma}{2} + \frac{\varphi}{2} - 1 \right) + \Phi \left(\frac{\Phi(-\gamma)}{2} + \frac{\Psi}{2} - 1 \right) - 1 \right]. \quad (6)$$

Таким образом, введённое понятие логического нейрона легко выражается через операции, допустимые в нейросетевом базисе и для каждого нейрона сети используется пороговая функция активации вида (3).

К сожалению, для многочисленных прикладных задач входные данные могут быть неточными и некорректно представленными. Это приводит к трудностям при использовании основанных на них моделях. Например, разделение неточных данных на основе пороговой функции активации в большинстве случаев приведёт к грубому либо неадекватному решению. Учитывая это, желательно вместо пороговой активационной функции использовать функцию, значение которой можно рассматривать как вероятность того или иного решения. В качестве такой функции целесообразно использовать сигмоид, т.е. функцию вида:

$$f(x) = \frac{1}{1 + e^{-bx}}, \quad (7)$$

принимаящую значения в диапазоне от 0 до 1, график которой показан на рис. 2 для различных значений коэффициента b . Коэффициент b определяет, насколько быстро и на каком интервале функция изменяет значение от 0 до 1, и в случае $b \rightarrow \infty$ функцию

(7) можно рассматривать как пороговую функцию. Таким образом, используя коэффициент b , можно управлять свойствами синтезируемой нейросетевой структуры.

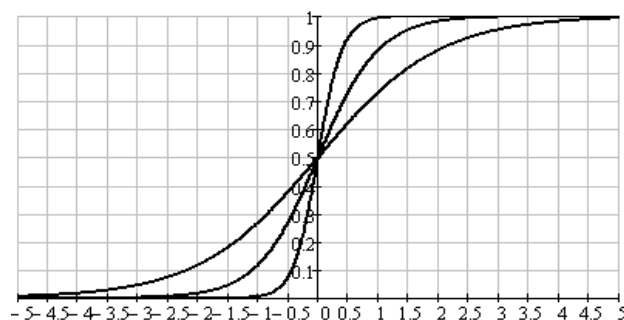


Рис. 2. График функции вида (7) для различных значений коэффициента b

5. Практические результаты

Рассмотрим, каким образом замена пороговой активационной функции входящих в сеть нейронов функцией вида (7) влияет на решения. Как и в работах [7-9], будем рассматривать задачу классификации, а для наглядности результата ограничимся размерностью входного вектора, равной двум ($N=2$) и двумя классами, так как это позволяет легко визуализировать результаты и обобщить их на пространство большей размерности. Таким образом, входной сигнал для нейронной сети рассматривается как точка на плоскости с координатами (x, y) .

Замена пороговой активационной функции на функцию вида (7) вносит отличие в оценку принадлежности предъявляемого ИНС образа к одному из классов. В случае, если используется сеть с кооперативной структурой [9] и пороговой функцией активации входящих в неё нейронов, решение принимается на основе сравнения доли P сетей, принявших положительное решение, с некоторым пороговым значением P_0 :

$$P = \frac{S}{K}, \quad (8)$$

где S - число положительных решений, K - число подсетей, входящих в единую кооперативную структуру.

Если же в качестве активационной функции нейронов используется сигмоид, то каждая из входящих в единую структуру сеть принимает решение о принадлежности образа с некоторой вероятностью. Выходной сигнал всей ИНС следует рассматривать как вероятность принадлежности предъявляемого образа искомого классу, вычисленную на основании результатов входящих в неё подсетей. Это приводит к некоторому изменению границы области, занимаемой элементами искомого класса и определяемой нейронной сетью, что хорошо видно на рис. 3(б).

Обучающая выборка на рисунке показана геометрическими фигурами – прямоугольниками и окружностями, изображающими различные классы. Закра-

шенный элемент означает, что нейронная сеть после обучения отнесла его к искомому классу. Затемнённая область на рисунке соответствует области, занимаемой искомым классом, и найдена путём полного перебора всех возможных точек в границах изображения и их классификации нейронной сетью.

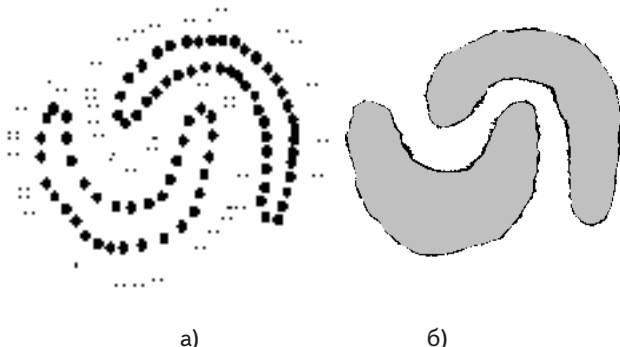


Рис. 3. Пример обучающей выборки (а) и граница области искомого класса (б)

Анализ данного примера позволяет заметить, что область искомого класса, полученная ИНС с пороговыми функциями активации нейронов, показанная тёмным цветом, имеет на некоторых участках изрезанную границу. В случае использования той же ИНС с сигмоидной функцией активации для входящих в неё нейронов граница области, выделенной светлым тоном, гладкая. На рис. 4. для этой же сети показано градиентное изображение, на котором более тёмным цветом выделена меньшая вероятность принадлежности точки искомому классу.

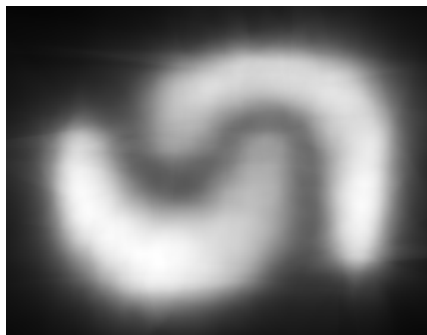


Рис. 4. Градиентное изображение области, занимаемой искомым классом

Таким образом, использование другой активационной функции для составляющих сеть нейронов позволяет сгладить границу области, занимаемой элементами искомого класса, что делает использование полученной сети предпочтительным для некоторых практических приложений.

6. Выводы

На основании изложенного в работе материала, можно утверждать, что для описанного класса задач:

1. Замена пороговой активационной функции каждого входящего в сеть нейрона на сигмоидальную функцию (7) позволяет улучшить качество классификации образов, особенно для случаев, когда они

находятся близко к границе областей, занимаемых различными классами.

2. Показана возможность представления используемых в структуре сети специальных логических нейронов в нейросетевом базисе – операцией взвешенного суммирования и нелинейного преобразования результата, что доказывает эквивалентность синтезируемых структур классическим нейронным сетям.

3. Предложенный метод дополняет ранее изложенные в работах [7-9] методы синтеза нейронных сетей и позволяет на данной основе формализовать получение нейросетевых структур для решения задач классификации.

Литература

1. Справочник по теории автоматического управления [Текст] / А. Г. Александров, В. М. Артемьев, В. Н. Афанасьев и др. ; под ред. А. А. Красовского. — М. : Наука, 1987. — 712 с.
2. Методы классической и современной теории автоматического управления [Текст]. Т. 2. Статистическая динамика и идентификация систем автоматического управления : учеб. / К. А. Пупков, Н. Д. Егупов, Е. М. Воронов и др. ; под ред. К. А. Пупкова, Н. Д. Егупова. — 2-е изд., перераб. и доп. — М. : Издательство МГТУ им. Н. Э. Баумана, 2004. — 640 с.
3. Методы классической и современной теории автоматического управления [Текст]. Т. 5. Методы современной теории автоматического управления : учеб. / К. А. Пупков, Н. Д. Егупов, А. И. Баркин и др. ; под ред. К. А. Пупкова, Н. Д. Егупова. — 2-е изд., перераб. и доп. — М. : Издательство МГТУ им. Н. Э. Баумана, 2004. — 784 с.
4. Методы робастного, нейро-нечеткого и адаптивного управления [Текст] : учеб. / К. А. Пупков, Н. Д. Егупов, А. И. Гаврилов и др. ; под ред. Н. Д. Егупова; изд. 2-ое, стереотипное. — М. : Издательство МГТУ им. Н.Э. Баумана, 2002. — 744 с.
5. Колмогоров, А. Н. О представлении непрерывных функций нескольких переменных в виде суперпозиции непрерывных функций одного переменного и сложения [Текст] / А. Н. Колмогоров // Докл. АН СССР. — 1957. — Т. 114, №5. — С. 953–956.
6. Hecht-Nielsen R. Kolmogorov's Mapping Neural Network Existence Theorem [Text] // IEEE First Annual Int. Conf. on Neural Networks. — 1987. — Vol.3. — P. 11–13.
7. Юрков, Д. А. Метод синтеза нейросетевых модельных структур [Текст] / Д. А. Юрков // Вісн. Східноукр. нац. ун-ту. — 2009. — № 5 (135). — С. 115–122.
8. Юрков, Д. А. Синтез нейронных сетей с фрактальной структурой [Текст] / Д. А. Юрков // Восточно-Европейский журнал передовых технологий. — 2009. — № 4-3 (40). — С. 39–44.
9. Юрков, Д. А. Классификация на основе кооперативных нейросетевых структур [Текст] / Д. А. Юрков // Восточно-Европейский журнал передовых технологий. — 2009. — №5-3 (41). — С. 51–56.