

Искусственные нейронные сети и машинное обучение

Нейронной сетью (НС) называется динамическая система, состоящая из совокупности связанных между собой по типу узлов направленного графа элементарных процессоров, называемых формальными нейронами, и способная генерировать выходную информацию в ответ на входное воздействие.

В последние несколько лет наблюдается взрыв интереса к нейронным сетям, они находят успешное применение в самых различных областях - бизнесе, медицине, технике, геологии, физике. Нейронные сети вошли в практику везде, где нужно решать задачи прогнозирования, классификации или управления, поскольку они применимы практически в любой ситуации, когда имеется связь между переменными-предикторами (входами) и прогнозируемыми переменными (выходами), даже если эта связь имеет очень сложную природу и ее трудно выразить в обычных терминах корреляций или различий между группами.

Нейронные сети нелинейны по своей природе и представляют собой исключительно мощный метод моделирования, позволяющий воспроизводить чрезвычайно сложные зависимости. На протяжении многих лет в качестве основного метода в большинстве областей использовалось линейное моделирование, поскольку для него хорошо разработаны процедуры оптимизации. Там, где линейная аппроксимация неудовлетворительна и линейные модели работают плохо, а таких задач достаточно много, основным инструментом становятся нейросетевые методы. Кроме того, нейронные сети справляются с «проклятием размерности», которое не позволяет моделировать линейные зависимости в случае большого числа переменных.

Нейронные сети учатся на примерах. Пользователь нейронной сети подбирает репрезентативную выборку, а затем запускает алгоритм обучения, который автоматически воспринимает структуру данных. При этом от пользователя, конечно, требуется какой-то набор эвристических знаний о том, как следует отбирать и подготавливать данные, выбирать нужную архитектуру сети и интерпретировать результаты, однако уровень знаний, необходимый для успешного применения нейронных сетей, гораздо скромнее, чем, например, при использовании традиционных методов статистики.

Нейросетевая парадигма

Идея нейронных сетей родилась в ходе исследований в области искусственного интеллекта, а именно, в результате попыток воспроизвести способность биологических нервных систем обучаться и исправлять ошибки, моделируя низкоуровневую структуру мозга. Основной областью исследований по искусственному интеллекту в 60-е - 80-е годы были экспертные системы. Такие системы основывались на высокоуровневом моделировании процесса мышления (в частности, на его представлении как манипуляций с символами). Скоро стало ясно, что подобные системы, хотя и могут принести пользу в некоторых областях, не охватывают некоторые ключевые аспекты работы человеческого мозга. Согласно одной из точек зрения, причина этого состоит в том, что они не в состоянии воспроизвести структуру мозга. Чтобы создать искусственный интеллект, необходимо построить систему с похожей архитектурой.

Мозг состоит из очень большого числа (приблизительно 10000000000) нейронов, соединенных многочисленными связями (в среднем несколько тысяч связей на один нейрон, однако это число может сильно колебаться). Нейроны - это специальные клетки, способные распространять электрохимические сигналы. Нейрон имеет разветвленную структуру ввода информации (дендриты), ядро и разветвляющийся выход (аксон). Аксоны клетки соединяются с дендритами других клеток с помощью синапсов. При активации нейрон посылает электрохимический сигнал по своему аксону. Через синапсы этот сигнал достигает других нейронов, которые могут в свою очередь активироваться. Нейрон активируется тогда, когда суммарный уровень сигналов, пришедших в его ядро из дендритов, превысит определенный уровень (порог активации).

Интенсивность сигнала, получаемого нейроном (а, следовательно, и возможность его активации), сильно зависит от активности синапсов. Каждый синапс имеет протяженность, и специальные химические вещества передают сигнал вдоль него. Один из самых авторитетных исследователей, Дональд Хебб, высказал постулат, что обучение заключается в первую очередь в изменениях «силы» синаптических связей. Например, в классическом опыте Павлова каждый раз непосредственно перед кормлением собаки звонил колокольчик, и собака быстро научилась связывать звонок колокольчика с пищей. Синаптические связи между участками коры головного мозга, ответственными за слух, и слюнными железами усилились, и при возбуждении коры звуком колокольчика у собаки начиналось слюноотделение.

Таким образом, будучи построен из очень большого числа совсем простых элементов (каждый из которых берет взвешенную сумму входных сигналов и в случае, если суммарный вход превышает определенный уровень, передает дальше двоичный сигнал), мозг способен решать чрезвычайно сложные задачи. Разумеется, мы не затронули здесь многие сложные аспекты устройства мозга, однако интересно то, что искусственные нейронные сети способны достичь замечательных результатов, используя модель, которая ненамного сложнее, чем описанная выше.

Искусственный нейрон как основа нейросети

Попытки моделирования работы мозга с помощью построения моделей нейронных сетей (нейрокибернетика) восходят ко времени появления первых ЭВМ и не были напрямую связаны с их появлением. Первые модели нейронных сетей были физическими и основывались на электронных моделях

нейронов. Начало этому моделированию положила идея формального нейрона Мак-Каллока и Питтса (1943 г.), которая была воспринята Розенблаттом и развита им в рамках построенных им и его учениками многочисленных моделей перцептронов (систем для распознавания образов) - одного из классов современных нейронных сетей.

Формальная модель нейрона Мак-Каллока-Питтса, которая и сейчас является наиболее применяемым формализмом для описания отдельного нейрона в нейронной сети, показана на рис. 3.1.

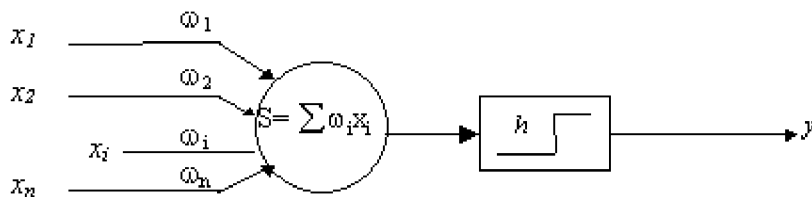


Рис. 3.1. Формальная модель нейрона Мак-Каллока - Питтса.

На рисунке приняты следующие обозначения: x_i - сигнал на i -м входе (синапсе) нейрона; w_i - вес i -го входа (синапса) нейрона; y - выход нейрона; h - порог срабатывания нейрона.

В модели взвешенная сумма сигналов на входах нейрона сравнивается с пороговым значением h , и на выходе есть сигнал, если она превышает порог. В современных моделях нейронов пороговая функция в общем случае заменяется на нелинейную функцию $y=f(S)$, называемую передаточной функцией или функцией активации нейрона. В качестве этой функции может использоваться, одна из сигмоидальных функций, например, рациональная сигмоида

$$f(S) = \frac{S}{S + \alpha}$$

Параметр α обычно называется смещением. Таким образом, иногда говорят, что нейрон состоит из умножителей (на веса), сумматора и нелинейного элемента.

Виды функции активации

Вид функции активации во многом определяет функциональные возможности нейронной сети и метод обучения этой сети. Рассмотрим некоторые виды функций активации, применяемые при конструировании нейронных сетей.

Пороговая функция. В общем случае эта функция активации описывается следующим выражением:

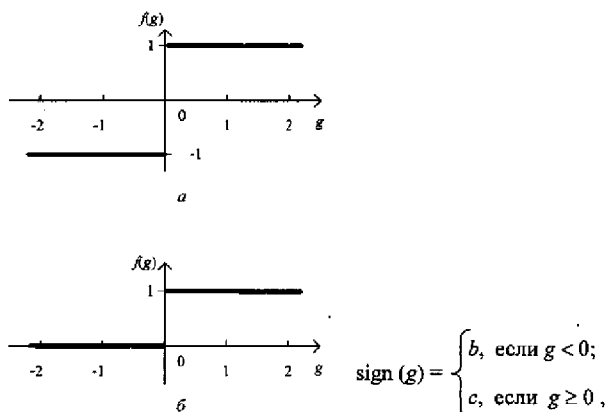


Рис. 3.3. Графики пороговых функций активации: а - симметричная; б - смещенная

Сигмоидная функция активации. Эта функция относится к классу непрерывных. Выбираемая разработчиком для решения конкретной задачи непрерывная функция активации должна удовлетворять следующим условиям: быть непрерывной, монотонно возрастающей и дифференцируемой. Сигмоидная функция полностью удовлетворяет этим требованиям. Как и для пороговой, для этой функции различают смещенный и симметричный вид. Математически оба вида сигмоидной функции представляются следующим образом:

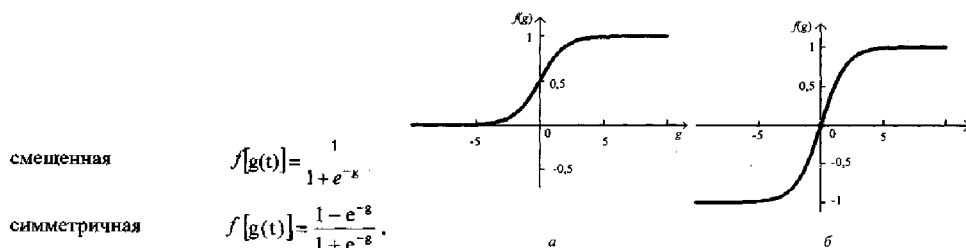


Рис. 3.4. Графики сигмоидных функций на базе экспоненты: а - смещенная; б - симметричная

Линейная функция активации. Эта функция также относится к классу непрерывных. Линейный участок такой функции активации позволяет оперировать с непрерывными сигналами. Зоны нечувствительности определяются физической реализуемостью этих функций.

Степенные функции активации. Функции этого вида используют при проектировании НС, предназначенных для решения задач аппроксимации функций одной и/или многих переменных, для реализации степенных рядов, для реализации сложных вычислительных алгоритмов. Математическая модель таких функций имеет вид

В практике разработки НС применяют функции активации и других видов, отличных от перечисленных выше. Все определяется характером решаемой задачи. В каждом конкретном случае разработчик должен обосновать свой выбор.

Формальный нейрон фактически представляет собой процессор с очень ограниченной специальной системой команд (в литературе принято называть нейросетевой базис). Формальные нейроны по способу представления информации бывают аналоговые и цифровые. И те и другие выполняют единообразные вычислительные действия и не требуют внешнего управления.

Построение и применение нейросетей

Построение нейронной сети

Возникает следующий вопрос: как соединять нейроны друг с другом? Если сеть предполагается для чего-то использовать, то у нее должны быть входы (принимающие значения интересующих нас переменных из внешнего мира) и выходы (прогнозы или управляющие сигналы). Входы и выходы соответствуют сенсорным и двигательным нервам, например, идущим соответственно от глаз и к рукам. Кроме этого, однако, в сети может быть еще много промежуточных (скрытых) нейронов, выполняющих внутренние функции. Входные, скрытые и выходные нейроны должны быть связаны между собой.

Ключевой вопрос здесь - обратная связь. Простейшая сеть имеет структуру прямой передачи сигнала; сигналы проходят от входов через скрытые элементы и, в конце концов, приходят на выходные элементы. Такая структура имеет устойчивое поведение. Если же сеть рекуррентная (т.е. содержит связи, ведущие назад от более дальних к более ближним нейронам), то она может быть неустойчива и иметь очень сложную динамику поведения. Рекуррентные сети представляют большой интерес для исследователей в области нейронных сетей, однако, при решении практических задач, но крайней мере до сих пор, наиболее полезными оказались структуры прямой передачи.

Типичный пример сети с прямой передачей сигнала показан на следующем рисунке. Нейроны регулярным образом организованы в слои. Обычно различают три типа узлов (нейронов) - входные (входной слой нейронов или Input layer), выходные (выходной слой или Output layer) и скрытые слои нейронов (Hidden layers) (рис. 3.5). Входной слой служит просто для ввода значений входных переменных. Каждый из скрытых и выходных нейронов соединен со всеми элементами предыдущего слоя.

При работе сети на входные элементы подаются значения входных переменных, затем последовательно отрабатывают нейроны промежуточных и выходного слоев. Каждый из них вычисляет свое значение активации, беря взвешенную сумму выходов элементов предыдущего слоя и вычитая из нее пороговое значение. Затем значение активации преобразуется с помощью функции активации, и в результате получается выход нейрона. После того, как вся сеть отработает, выходные значения элементов выходного слоя принимаются за выход всей сети в целом.

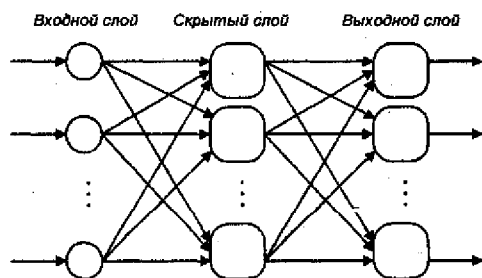


Рис. 3.5. Нейронная сеть с прямыми связями.

Функционирование нейронной сети состоит из двух этапов: обучения сети “правильному” или адекватному реагированию на входную информацию (входной вектор) и использования обученной сети для распознавания входных векторов. Последний этап часто называют тестированием. Другими словами, сеть учится распознаванию входных векторов, то есть формированию выходных векторов, соответствующих распознанному классу входных векторов.

Применение нейронных сетей

Класс задач, которые можно решить с помощью нейронной сети, определяется тем, как сеть работает и как она обучается. Таким образом, сеть можно применять в ситуации, когда имеется определенная известная информация, и вы хотите из нее получить некоторую пока не известную информацию. Вот некоторые важные примеры задач, в которых были успешно реализованы нейросетевые методы [9,с.42].

Распознавание состояния больного. Применение классических статистических методов для решения этой задачи описано еще в работах Неймана. С помощью медицинской аппаратуры можно наблюдать за различными показателями состояния здоровья человека (например, частотой пульса, содержанием различных веществ в крови, частотой дыхания). Стадии возникновения некоторой болезни может соответствовать определенная и весьма сложная (например, нелинейная и взаимозависимая) комбинация изменений наблюдаемых переменных, которая может быть обнаружена с помощью нейросетевой модели.

Прогнозирование на фондовом рынке. Колебания цен на акции и фондовых индексов - еще один пример сложного, многомерного, но, в определенных ситуациях, частично прогнозируемого явления. Многие финансовые аналитики используют нейронные сети для прогнозирования цен акций на основе многочисленных факторов, например, прошлого поведения цен этих и других акций в совокупности с различными другими экономическими показателями. В качестве альтернативных вариантов часто применяются модели авторегрессии и технический анализ.

Предоставление кредита. Как правило, у банка имеется большой набор сведений о человеке, обратившемся с просьбой о предоставлении кредита. Это могут быть его возраст, образование, род занятий и многие другие данные. Обучив нейронную сеть на уже имеющихся данных, аналитик может определить наиболее существенные характеристики, и на их основе отнести данного клиента к категории с высоким или низким кредитным риском. Заметим, что для решения подобных задач можно параллельно использовать и классические статистические методы, такие как дискриминантный анализ и деревья классификации.

Системы слежения за состоянием оборудования. Нейронные сети оказались полезны как средство контроля состояния механизмов. Нейронная сеть может быть обучена так, чтобы отличать звук, который издает машина при нормальной работе («ложная тревога») от того, который является предвестником неполадок. После такого обучения нейронная сеть может предупреждать инженеров об угрозе поломки до того, как она случится, и тем самым исключать неожиданные и дорогостоящие простои.

Управление работой двигателя. Нейронные сети используются для анализа сигналов от датчиков, установленных на двигателях. С помощью нейронной сети можно управлять различными параметрами работы двигателя, чтобы достичь определенной цели, например, уменьшить потребление горючего.

Разумеется, вовсе не любую задачу можно решить с помощью нейронной сети. Важное условие применения нейронных сетей: вы должны знать (или хотя бы иметь серьезные подозрения), что между известными входными значениями и неизвестными выходами имеется связь. Эта связь может быть искажена шумом, но она должна существовать (так, едва ли можно ожидать, что по данным из примера с прогнозированием цен акций можно построить абсолютно точный прогноз, поскольку на цену влияют и другие факторы, не представленные во входном наборе данных, и, кроме того, в задаче присутствует элемент случайности).

Как правило, нейронная сеть используется тогда, когда неизвестен точный вид связей между входами и выходами, - если бы он был известен, то связь можно было бы моделировать непосредственно. Другая существенная особенность нейронных сетей состоит в том, что зависимость между входом и выходом находится в процессе обучения сети. Для обучения нейронных сетей применяются алгоритмы двух типов (разные типы сетей используют разные типы обучения): управляемое («обучение с учителем») и неуправляемое («без учителя»).

Для управляемого обучения сети пользователь должен подготовить набор обучающих данных. Эти данные представляют собой примеры входных данных и соответствующих им выходов. Сеть учится устанавливать связь между первыми и вторыми. Обычно обучающие данные берутся из исторических сведений. В рассмотренных выше примерах это могут быть предыдущие значения цен акций и индекса FTSE, сведения о прошлых заемщиках - их анкетные данные и то, успешно ли они выполнили свои обязательства, примеры положений робота и его правильной реакции.

Затем нейронная сеть обучается с помощью того или иного алгоритма управляемого обучения (наиболее известным из них является метод обратного распространения), при котором имеющиеся данные используются для корректировки весов и пороговых значений сети таким образом, чтобы минимизировать ошибку прогноза на обучающем множестве. Если сеть обучена хорошо, она приобретает способность моделировать (неизвестную) функцию, связывающую значения входных и выходных переменных, и впоследствии такую сеть можно использовать для прогнозирования в ситуации, когда выходные значения неизвестны.

Классификация и выбор структуры нейросетей **Классификация нейросетей**

Возможности и функции нейронной сети определяются не только характеристиками нейроподобных элементов, но и особенностями ее архитектуры, а именно топологией межнейронных связей, выбором определенных подмножеств нейроподобных элементов для ввода и вывода информации, способами обучения сети, наличием или отсутствием конкуренции между нейронами, направлением и способами управления и синхронизации передачи информации между нейронами.

С точки зрения топологии можно выделить три основных типа нейронных сетей:

полносвязные (рис.3.6, а);

многослойные или слоистые (рис. 3.6, б);

слабосвязные (с локальными связями) (рис. 3.6, в).

В полносвязных нейронных сетях каждый нейрон передает свой выходной сигнал остальным нейронам, в том числе и самому себе. Все входные сигналы подаются всем нейронам. Выходными сигналами сети могут быть все или некоторые выходные сигналы нейронов после нескольких тактов функционирования сети.

В многослойных нейронных сетях нейроны объединяются в слои. Слой содержит совокупность нейронов с едиными входными сигналами. Число нейронов в слое может быть любым и не зависит от количества нейронов в других слоях. В общем случае сеть состоит из Q слоев, пронумерованных слева направо. Внешние входные сигналы подаются на входы нейронов входного слоя (его часто нумеруют как нулевой), а выходами сети являются выходные сигналы последнего слоя. Кроме входного и выходного слоев в многослойной нейронной сети есть один или несколько скрытых слоев. Связи от выходов нейронов некоторого слоя q к входам нейронов следующего слоя $(q+1)$ называются последовательными.

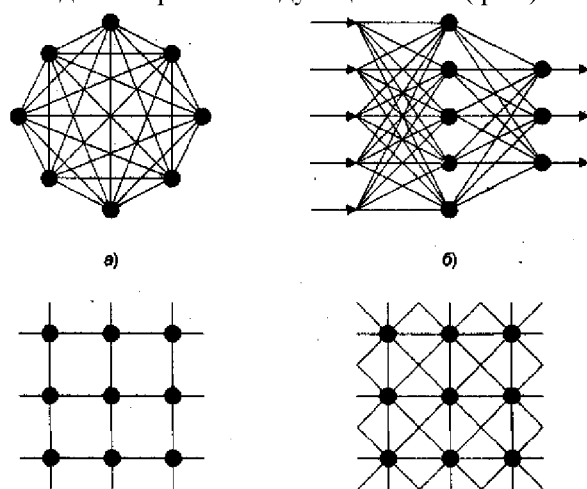


Рис. 3.6. Архитектуры нейронных сетей: а - полносвязная сеть, б - многослойная сеть с последовательными связями, в - слабосвязные сети

В свою очередь, среди многослойных нейронных сетей выделяют следующие типы:

Монотонные. Это частный случай слоистых сетей с дополнительными условиями на связи и нейроны. Каждый слой кроме последнего (выходного) разбит на два блока: возбуждающий и тормозящий. Связи между блоками тоже разделяются на тормозящие и возбуждающие. Если от нейронов блока А к нейронам блока В ведут только возбуждающие связи, то это означает, что любой выходной сигнал блока является монотонной неубывающей функцией любого выходного сигнала блока А. Если же эти связи только тормозящие, то любой выходной сигнал блока В является невозрастающей функцией любого выходного сигнала блока А. Для нейронов монотонных сетей необходима монотонная зависимость выходного сигнала нейрона от параметров входных сигналов.

Сети без обратных связей. В таких сетях нейроны входного слоя получают входные сигналы, преобразуют их и передают нейронам первого скрытого слоя, и так далее вплоть до выходного, который выдает сигналы для интерпретатора и пользователя. Если не оговорено противное, то каждый выходной сигнал g -го слоя подается на вход всех нейронов $(g+1)$ -го слоя; однако возможен вариант соединения g -го слоя с произвольным $(g+n)$ -м слоем.

Среди многослойных сетей без обратных связей различают полносвязные (выход каждого нейрона g -го слоя связан с входом каждого нейрона $(g+1)$ -го слоя) и частично полносвязные. Классическим вариантом слоистых сетей являются полносвязные сети прямого распространения (рис. 3.5).

Сети с обратными связями. В сетях с обратными связями информация с последующих слоев передается на предыдущие. Среди них, в свою очередь, выделяют следующие:

слоисто-циклические, отличающиеся тем, что слои замкнуты в кольцо: последний слой передает свои выходные сигналы первому; все слои равноправны и могут как получать входные сигналы, так и выдавать выходные;

слоисто-полносвязанные состоят из слоев, каждый из которых представляет собой полносвязную сеть, а сигналы передаются как от слоя к слою, так и внутри слоя; в каждом слое цикл работы распадается на три части: прием сигналов с предыдущего слоя, обмен сигналами внутри слоя, выработка выходного сигнала и передача к последующему слою;

полносвязанно-слоистые, по своей структуре аналогичные слоисто-полносвязанным, но функционирующим по-другому: в них не разделяются фазы обмена внутри слоя и передачи следующему, на каждом такте нейроны всех слоев принимают сигналы от нейронов как своего слоя, так и последующих.

В качестве примера сетей с обратными связями на рис. 3.7 представлены частично-рекуррентные сети Элмана и Жордана.

В слабосвязных нейронных сетях нейроны располагаются в узлах прямоугольной или гексагональной решетки. Каждый нейрон связан с четырьмя (окрестность фон Неймана), шестью (окрестность Голея) или восемью (окрестность Мура) своими ближайшими соседями.

Известные нейронные сети можно разделить по типам структур нейронов на гомогенные(однородные) и гетерогенные. Гомогенные сети состоят из нейронов одного типа с единой функцией активации, а в гетерогенную сеть входят нейроны с различными функциями активации.

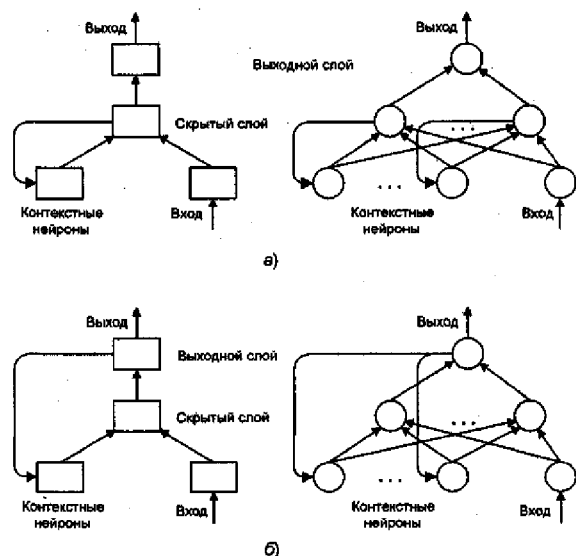


Рис. 3.7. Частично-рекуррентные сети: а - Элмана, б - Жордана

Существуют бинарные и аналоговые сети. Первые из них оперируют только двоичными сигналами, и выход каждого нейрона может принимать значение либо логического нуля (заторможенное состояние) либо логической единицы (возбужденное состояние).

Еще одна классификация делит нейронные сети на синхронные и асинхронные. В первом случае в каждый момент времени лишь один нейрон меняет свое состояние, во втором - состояние меняется сразу у целой группы нейронов, как правило, у всего слоя. Алгоритмически ход времени в нейронных сетях задается итерационным выполнением однотипных действий над нейронами.

Сети можно классифицировать также по числу слоев. Теоретически число слоев и число нейронов в каждом слое может быть произвольным, однако фактически оно ограничено ресурсами компьютера или специализированных микросхем, на которых обычно реализуется нейронная сеть. Чем сложнее сеть, тем более сложные задачи она может решать.

Кроме того, НС классифицируют по способу решения задач. С точки зрения этого признака сети делят на формируемые, сети с формируемой матрицей связи, обучаемые и комбинированные (смешанные).

Формируемые сети. Сети этого вида проектируют для формализуемых задач, имеющих четко сформулированный в нейросетевом базисе алгоритм решения конкретной задачи.

Сети с формируемой матрицей связей. Сети этого вида применяют для трудноформализуемых задач. Как правило, эти сети имеют одинаковую структуру и различаются лишь матрицей связи (сеть Хопфилда). Достоинством таких сетей является их наглядность в работе.

Обучаемые сети. Этот вид сетей используют для решения неформализуемых задач. В процессе обучения сети автоматически изменяются такие ее параметры, как коэффициенты синаптической связи, а в некоторых случаях и топология. Серьезным препятствием в широком применении ПК все еще остается большое время обучения сети. Поэтому выбор или разработка алгоритма обучения являются ключевой задачей разработчика нейронных систем.

Комбинированные (смешанные) сети. Этот класс сетей сочетает в себе признаки двух, а то и трех основных видов. Как правило, эти сети многослойные, каждый слой которых представляется различной топологией и обучается по определенному алгоритму. В настоящее время этот класс получает наибольшее распространение, так как дает самые широкие возможности разработчику.

Выбор структуры нейросетей

Выбор структуры нейронной сети осуществляется в соответствии с особенностями и сложностью задачи. Для решения отдельных типов задач уже существуют оптимальные конфигурации, описанные в приложении. Если же задача не может быть сведена ни к одному из известных типов, приходится решать сложную проблему синтеза новой конфигурации. При этом необходимо руководствоваться следующими основными правилами:

возможности сети возрастают с увеличением числа нейронов сети, плотности связей между ними и числом слоев;

введение обратных связей наряду с увеличением возможностей сети поднимает вопрос о динамической устойчивости сети;

сложность алгоритмов функционирования сети, введение нескольких типов синапсов способствует усилению мощности нейронной сети.

Вопрос о необходимых и достаточных свойствах сети для решения задач того или иного рода представляет собой целое направление нейрокомпьютерной науки. Так как проблема синтеза нейронной сети сильно зависит от решаемой задачи, дать общие подробные рекомендации затруднительно. В большинстве случаев оптимальный вариант получается на основе интуитивного подбора, хотя в литературе приведены доказательства того, что для любого алгоритма существует нейронная сеть, которая может его реализовать. Остановимся на этом подробнее.

Многие задачи распознавания образов (зрительных, речевых), выполнения функциональных преобразований при обработке сигналов, управления, прогнозирования, идентификации сложных систем, сводятся к следующей математической постановке. Необходимо построить такое отображение $X \rightarrow Y$, чтобы на каждый возможный входной сигнал X формировался правильный выходной сигнал Y . Отображение задается конечным набором пар ($\langle \text{вход} \rangle$, $\langle \text{известный выход} \rangle$). Число этих пар (обучающих примеров) существенно меньше общего числа возможных сочетаний значений входных и выходных сигналов. Совокупность всех обучающих примеров носит название обучающей выборки.

В задачах распознавания образов X - некоторое представление образа (изображение, вектор), Y - номер класса, к которому принадлежит входной образ.

В задачах управления X - набор контролируемых параметров управляемого объекта, Y - код, определяющий управляющее воздействие, соответствующее текущим значениям контролируемых параметров.

В задачах прогнозирования в качестве входных сигналов используются временные ряды, представляющие значения контролируемых переменных на некотором интервале времени. Выходной сигнал - множество переменных, которое является подмножеством переменных входного сигнала.

При идентификации X и Y представляют входные и выходные сигналы системы соответственно.

Вообще говоря, большая часть прикладных задач может быть сведена к реализации некоторого сложного функционального многомерного преобразования. В результате отображения $X \rightarrow Y$ необходимо обеспечить формирование правильных выходных сигналов в соответствии:

со всеми примерами обучающей выборки;

со всеми возможными входными сигналами, которые не вошли в обучающую выборку.

Второе требование в значительной степени усложняет задачу формирования обучающей выборки. В общем виде эта задача еще не решена, однако во всех известных случаях может быть найдено частное решение.

Решение задач с помощью нейросетей. Общий подход к построению нейросети

Практически любую задачу можно свести к задаче, решаемой нейронной сетью. Допустим, что есть задача распознавания рукописных букв. Дано: растровое черно-белое изображение буквы размером 30x30 пикселей. Необходимо: распознать предъявленную букву (в алфавите 33 буквы). В терминах нейронной сети задача может быть сформулирована так: *Дано, входной вектор из 900 двоичных символов ($900 = 30 \times 30$)* *Необходимо: построить нейронную сеть с 900 входами и 33 выходами, которые помечены буквами. Номер нейрона с максимальным значением выходного сигнала должен соответствовать предъявленному изображению буквы.*

Интерес представляет не аналоговое значение выхода, а номер класса (номер буквы в алфавите). Поэтому каждому классу сопоставляется свой выходной нейрон, а ответом сети считается номер класса, соответствующий нейрону, на чьем выходе уровень сигнала максимален. Значение же уровня сигнала на выходе характеризует достоверность того, что на вход была подана соответствующая рукописная буква.

Нейронная сеть строится в два этапа.

выбор типа (архитектуры) сети.

подбор весов (обучение) сети.

На первом этапе необходимо определить следующее:

какие нейроны использовать (число входов, функции активации);

каким образом следует соединить нейроны между собой;

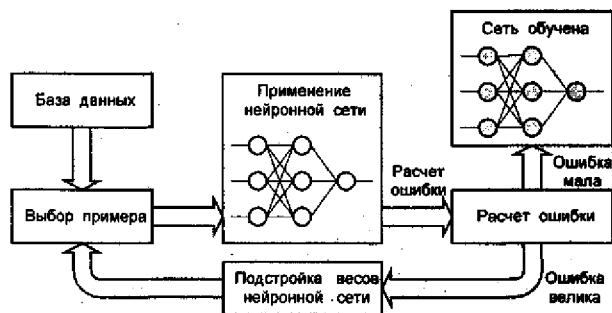
что взять в качестве входов и выходов сети.

Нет необходимости придумывать нейронную сеть «с нуля», так как существуют несколько десятков различных нейросетевых архитектур, причем эффективность многих из них доказана математически. Наиболее популярные и изученные из них - это многослойный персептрон, нейронная сеть с общей регрессией, сети Кохонена и другие. Все основные архитектуры рассмотрены в лабораторном практикуме.

На втором этапе производится обучение выбранной сети посредством настройки ее весов. Количество весов может быть велико, поэтому обучение представляет собой сложный и длительный процесс. Для многих архитектур разработаны специальные алгоритмы обучения, наиболее популярный из которых алгоритм обратного распространения ошибки.

Обучение нейросети

Обучить нейронную сеть это значит, сообщить ей, чего от нее добиваются. Этот процесс похож на обучение ребенка алфавиту. Показав ребенку изображение буквы и получив неверный ответ, ему сообщается тот ответ, который хотят получить. Ребенок запоминает этот пример вместе с верным ответом и в его памяти происходят некоторые изменения в нужном направлении.



Процесс обучения заключается в последовательном предъявлении букв (рис. 3.8).

При предъявлении изображения каждой буквы на входе сеть выдает некоторый ответ, не обязательно верный. Известен и верный (желаемый) ответ. В данном случае желательно, чтобы на выходе соответствующего нейрона уровень сигнала был максимален. Обычно в качестве желаемого в задаче классификации берут набор, где «1» стоит на выходе этого нейрона, а «0» - на выходах всех остальных нейронов. Одну и ту же букву (а также различные изображения одной и той же буквы)

можно предъявлять сети много раз.

После многократного предъявления примеров веса сети стабилизируются, причем сеть дает правильные ответы на все (или почти все) примеры из базы данных. В таком случае говорят, что сеть обучена. В программных реализациях можно видеть, что в процессе обучения величина ошибки (сумма квадратов ошибок по всем выходам) постепенно уменьшается. Когда величина ошибки достигает нуля или приемлемо малого уровня, обучение останавливают, и сеть готова к распознаванию.

Важно отметить, что вся информация, которую сеть приобретает о задаче, содержится в наборе примеров. Поэтому качество обучения сети зависит от количества примеров в обучающей выборке, а также от того, насколько полно эти примеры описывают задачу. Считается, что для полноценной тренировки требуется хотя бы несколько десятков (а лучше сотен) примеров.

Применение обученной нейросети

Важнейшая особенность человеческого мозга состоит в том, что, однажды обучившись определенному процессу, он может верно действовать и в тех ситуациях, которым он не обучался. Так же и обученная нейронная сеть может с большой вероятностью правильно реагировать на новые, не предъявленные ей ранее данные. Например, можно нарисовать букву другим почерком, а затем предложить нейронной сети классифицировать новое изображение. Веса обученной сети хранят достаточно много информации о сходстве и различиях букв, поэтому можно рассчитывать на правильный ответ и для нового варианта изображения.

Отметим, что задачи классификации (типа распознавания букв) очень плохо алгоритмизируются. Если в случае распознавания букв верный ответ очевиден заранее, то в более сложных практических задачах обученная нейронная сеть выступает как эксперт, обладающий большим опытом и способный дать ответ на трудный вопрос. Примером такой задачи служит медицинская диагностика, где сеть может учитывать большое количество числовых параметров (энцефалограмма, давление, вес). Конечно, «мнение» сети в этом случае нельзя считать окончательным.

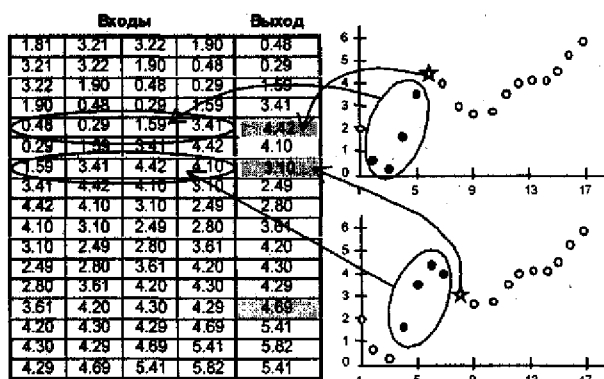
Классификация предприятий по степени их перспективности - это уже привычный способ использования нейронных сетей в практике западных компаний. При этом сеть также использует множество экономических показателей, сложным образом связанных между собой.

Нейросетевой подход особенно эффективен в задачах экспертной оценки по той причине, что он сочетает в себе способность компьютера к обработке чисел и способность мозга к обобщению и распознаванию. Говорят, что у хорошего врача способность к распознаванию в своей области столь велика, что он может провести приблизительную диагностику уже по внешнему виду пациента. Можно согласиться также, что опытный трейдер чувствует направление движения рынка по виду графика. Однако в первом случае все факторы наглядны, т. е. характеристики пациента мгновенно воспринимаются мозгом как «бледное лицо», «блеск в глазах». Во втором же случае учитывается только один фактор - курс за определенный период времени.

Нейронная сеть позволяет обрабатывать огромное количество факторов {до нескольких тысяч}, независимо от их наглядности. Это универсальный «хороший врач», который может поставить свой диагноз в любой области.

Помимо задач классификации, нейронные сети широко используются для поиска зависимостей в данных и кластеризации. Кластеризация - это разбиение набора примеров на несколько компактных областей (кластеров), причем число кластеров заранее неизвестно. Кластеризация позволяет представить неоднородные данные в более наглядном виде и использовать далее для исследования каждого кластера различные методы. Например, нейронная сеть на основе методики использования МГУА (метода группового учета аргументов) позволяет по обучающей выборке построить зависимость одного параметра от других в виде полинома. Такая сеть может не только мгновенно выучить таблицу умножения, но и найти сложные скрытые зависимости в данных (например, финансовых), которые не обнаруживаются стандартными статистическими методами, быстро выявить фальсифицированные страховые случаи или недобросовестные предприятия.

Особенно важны для практики, в частности, для финансовых приложений, задачи прогнозирования, поэтому поясним способы применения нейронных сетей в этой области более подробно.



Рассмотрим задачу прогнозирования курса акций на день вперед. Пусть имеется база данных, содержащая значения курса за последние 300 дней. Построим прогноз завтрашней цены на основе курсов за последние несколько дней. Понятно, что прогнозирующая нейронная сеть должна иметь всего один выход и столько входов, сколько предыдущих значений мы хотим использовать для прогноза, например, четыре последних значения. Составить обучающий пример очень просто, входными значениями будут курсы за четыре последних дня, а желаемым выходом - известный курс в следующий за ними день.

Рис. 3.9. Подготовка данных для нейронной сети в Excel

Если нейронная сеть совместима с какой-либо системой обработки электронных таблиц (например, Excel), то подготовка обучающей выборки состоит из следующих операций:

- занести значения курса акций последовательно в столбец таблицы;
- скопировать значения котировок в 4 соседних столбца;
- сдвинуть второй столбец на 1 ячейку вверх, третий столбец - на 2 и т. д. (рис. 3.9).

Смысл этой подготовки состоит в том, что каждая строка таблицы теперь представляет собой обучающий пример, где первые четыре числа - входные значения сети, а пятое число - желаемое значение выхода. Исключение составляют последние четыре строки, где данных недостаточно. Поэтому эти строки не учитываются при обучении. Заметим, что в четвертой снизу строке заданы все четыре входных значения, но неизвестно значение выхода. Именно при применении к этой строке обученной сети и можно получить прогноз на следующий день.

Как видно из этого примера, объем обучающей выборки зависит от выбранного количества входов. Если сделать 299 входов то такая сеть потенциально могла бы строить лучший прогноз, чем сеть с 4 входами, однако в этом случае имеется всего один обучающий пример, и обучение бессмысленно. Это следует учитывать при выборе числа входов, выбирая разумный компромисс между глубиной предсказания (число входов) и качеством обучения (объем обучающей выборки).