

УДК 004.032.26

НЕЙРОСЕТЕВОЙ ПОДХОД К РАСПОЗНАВАНИЮ ОБРАЗОВ

Л.В. СЕРЕБРЯНАЯ, О.А. ШУШИНА

*Белорусский государственный университет информатики и радиоэлектроники
П. Бровки, 6, Минск, 220013, Беларусь*

Поступила в редакцию 27 августа 2009

Даны решения задач распознавания образов на базе искусственных нейронных сетей в виде персептрона и карты Кохонена. Рассмотрены принципы построения, обучения и использования нейросетей. Указаны особенности нейросетевого подхода, а также круг задач, для которых он наиболее предпочтителен. Представлены алгоритмы функционирования, программная реализация и результаты работы искусственной нейронной сети.

Ключевые слова: искусственная нейронная сеть, многослойный персептрон, карта Кохонена, распознавание образов.

Введение

В настоящее время постоянно расширяется круг задач, для решения которых применяются искусственные нейронные сети (ИНС). При построении ИНС были учтены замечательные особенности естественных нейронных сетей. Первая из них состоит в том, что ошибка в действиях отдельного нейрона остается незаметной в общей массе взаимодействующих клеток, и нейронная сеть является устойчивой системой, в которой единичные сбои не оказывают существенного влияния на результаты функционирования сети в целом. Вторая важная особенность — это высокая скорость функционирования. Она достигается благодаря параллельной обработке информации огромным количеством нейронов, соединенных многочисленными связями. Поэтому изучение и копирование естественных сетей позволяет надеяться на то, что удастся создать устройство с высокой степенью надежности и параллельности выполнения независимых операций, чья скорость функционирования может быть приближена к уровню, наблюдаемому в процессах обработки информации биологическими объектами. Кроме того, ИНС демонстрируют ряд свойств, присущих мозгу. Например, они обучаются на основе опыта, обобщают предыдущие прецеденты на новые случаи и извлекают существенные свойства из поступающей информации, содержащей излишние данные [1, 2].

Можно выделить направления прикладных областей науки, где с помощью ИНС были получены существенные результаты:

- распознавание разнообразных образов, отличающихся по природе, сложности и т.д.;
- ассоциативный поиск информации и создание ассоциативных моделей;
- синтез речи и построение естественного языка;
- создание моделей нелинейных и трудно описываемых математически систем, а также прогнозирование их развития во времени.

Целью настоящей работы было решение нескольких актуальных прикладных задач на базе ИНС. Первая из задач посвящена распознаванию вредоносного кода в операционной системе; вторая связана с обнаружением и распознаванием человеческих лиц на изображениях;

а в третьей выполняется классификация информации. Для двух первых задач используется ИНС в виде перцептрона, а для последней — сеть Кохонена.

Общие принципы построения и работы ИНС

В основе ИНС лежит формальный нейрон (ФН), созданный по аналогии с биологическим нейроном. Работа сети состоит в преобразовании входных сигналов во времени, в результате чего меняется ее внутреннее состояние и формируются выходные воздействия. ФН состоит из взвешенного сумматора и нелинейного элемента.

ФН имеет несколько входных сигналов и один выходной сигнал. Параметрами нейрона, определяющими его работу, являются: вектор весовых коэффициентов, пороговый уровень и вид функции активации. Ее выбор определяется решаемой задачей и алгоритмом обучения.

ФН могут объединяться в сети различным образом. Сеть состоит из произвольного количества слоев нейронов, в простейшем случае — слой один. Первый слой называется входным, последний — выходным, а внутренние слои — скрытыми. Количество нейронов в слоях может быть произвольным. Обычно во всех скрытых слоях одинаковое количество нейронов. В каждом слое выполняется нелинейное преобразование линейной комбинации сигналов предыдущего слоя. Следовательно, в тех сетях, где требуется последовательное соединение слоев нейронов друг за другом, необходима нелинейная функция активации. В противном случае многослойность оказывается ненужной, т.к. ее можно заменить эквивалентной однослойной сетью с соответствующими весовыми коэффициентами. Многослойная сеть может формировать на выходе произвольную многомерную функцию при соответствующем выборе количества слоев, диапазона изменения сигналов и параметров нейронов [3].

В целом работа с ИНС состоит из двух процедур: обучения и непосредственного решения поставленной задачи. В результате первой процедуры сеть должна приобрести способность одинаково реагировать на все объекты одного образа и по-разному — на объекты различных образов. За обучением следует процесс распознавания "незнакомых" объектов, характеризующий действия уже обученной системы.

Для двух первых из вышеназванных прикладных задач применяются ИНС с контролируемым обучением. В этом случае исходные данные делятся на обучающую и тестовую выборки, принцип разделения может быть произвольным. Обучающие данные подаются сети для обучения, а проверочные используются для расчета ошибки сети. Таким образом, если на проверочных данных ошибка уменьшается, то сеть действительно выполняет обобщение (обучение). Если ошибка на обучающих данных продолжает уменьшаться, а ошибка на тестовых данных увеличивается, значит, сеть перестала выполнять обобщение и просто "запоминает" обучающие данные. В подобных случаях обучение обычно прекращают. В результате этапа обучения на основе эталонных наборов данных вход-выход ИНС настраивается таким образом, чтобы в дальнейшем для произвольного входного сигнала выдать достаточно точный результат. Перед началом обучения весовые коэффициенты устанавливаются равными некоторым случайным значениям. В процессе обучения сеть должна корректировать весовые коэффициенты так, чтобы максимально уменьшить значение общей ошибки.

Для корректировки весов многослойной сети в процессе управляемого обучения используется алгоритм обратного распространения ошибок. Его идея заключается в том, что каждый элемент вносит свой "вклад" в ошибку следующего элемента. Поэтому весовые коэффициенты скрытого элемента следует скорректировать пропорционально его "вкладу" в величину ошибки следующего слоя. При распространении сигналов ошибки в обратном направлении ошибка каждого выходного элемента вносит свой "вклад" в ошибку каждого элемента скрытого слоя. Этот "вклад" зависит от величины ошибки выходного элемента и весового коэффициента связи, соединяющей элементы. Следовательно, выходной элемент с большей ошибкой делает больший "вклад" в ошибку того элемента скрытого слоя, который связан с ним большим по величине весом [1, 2].

С каждым входным образцом из обучающей выборки связывается целевой выходной образец. Полный цикл подачи всех пар обучающей выборки называется эпохой. Обучение

продолжается до тех пор, пока изменение усредненной квадратичной ошибки при переходе от одной эпохи к другой не окажется меньше некоторого допустимого значения. Другим критерием окончания обучения можно считать момент, когда выход для каждого учебного образца оказывается в рамках допустимого отклонения от соответствующего целевого выходного образца.

Задачу классификации текстовой информации наиболее удачно удастся решить с помощью самоорганизующейся сети (карты) Кохонена, поскольку она может самостоятельно разбивать данные на классы и организовывать новые рубрики в случае появления текстов, не относящихся ни к одной из уже имеющихся групп.

Самоорганизующаяся карта Кохонена представляет собой ИНС в виде двумерной решетки, в которой каждый нейрон имеет количество входов, равное размерности входного вектора. Входные данные могут быть как двоичными, так и непрерывными. Чем больше размерность нейронной решетки, тем детальнее происходит анализ данных. Каждый нейрон представляет собой сумматор, сеть работает по принципу "победитель забирает все", т.е. в результате работы выбирается нейрон, который характеризует какой-либо класс. Таким образом, данные, относящиеся к одной рубрике, будут активизировать один и тот же нейрон. Поскольку речь идет об обучении без учителя, заранее не известно, какому именно классу соответствует выделенный нейрон, гарантируется лишь разделение различных наборов данных на классы.

Применение ИНС для нахождения вредоносного кода в операционной системе

Если рассматривать задачу обнаружения вируса с точки зрения алгоритмов принятия решений, то она сводится к задаче классификации объектов на два класса: *вирус* и *не вирус*. Аналогично тому, как ИНС распознает оптический образ, она может распознать некий *образ вируса*. В качестве входных сигналов используются команды процессора или системные вызовы. Под образом вируса здесь подразумевается совокупность некоторых характеристик конкретного файла или процесса, например, последовательность действий, выполняемых программой.

Задача сводится к разбиению образов на два известных заранее класса [4]. Подобные задачи удобно решать на многослойном персептроне с обратным распространением ошибки, где каждый элемент слоя i связан со всеми элементами слоя $i+1$. Число входных элементов определяется размерностью входного вектора. Число выходных элементов определяется количеством классов, и каждому элементу выходного слоя ставится в соответствие один класс.

Практически в каждом вирусном детекторе присутствуют два компонента: технический и аналитический. Технический компонент отвечает за сбор информации. Аналитический компонент на основе полученной информации принимает решение о том, является ли конкретный процесс вирусом или нет. Выбор способа построения этих двух компонентов определяет общий алгоритм работы программного средства (ПС). В созданном ПС в качестве технического компонента используется мониторинг системных событий, реализованный при помощи перехвата системных API-функций. Аналитический компонент реализован на основе ИНС.

Программная реализация аналитического компонента на основе ИНС

Нейронная сеть реализована с помощью библиотеки классов на языке программирования C#3.0 в среде разработки Microsoft Visual Studio 2008. Диаграмма классов представлена на рис. 1. Библиотека позволяет строить сеть с обратным распространением ошибок для любого количества нейронов и скрытых слоев. Класс *Neuron* реализует основную структурную единицу сети — нейрон. К постоянным свойствам нейрона относятся функция активности и сдвига, *Neuron::activity* и *Neuron::base*, соответственно. Поля *Neuron::totalInput* и *Neuron::out* — это комбинированный ввод и вывод нейрона, соответственно. С этими полями работают методы *SetInput* и *SetExplicitInput*, первый из которых еще используется для вычисления функции активности нейрона. Метод *SetInput* предназначен для нейронов скрытых и выходного слоев, а метод *SetExplicitInput* — для нейронов

входного слоя. В поле *Neuron::_error* записывается ошибка нейрона при обучении сети. Поля *Neuron::_nextLayerNeurons* и *Neuron::_prevLayerNeurons* содержат список указателей на класс *NeuronLink*.

Каждый экземпляр класса *NeuronLink* содержит указатели на два связываемых нейрона, а также вес этой связи. Таким образом, каждый нейрон имеет доступ ко всем нейронам предыдущего и следующего слоев, что соответствует модели персептрона. Для связывания слоев нейронов между собой используется метод *Neuron::SetLinks*. Он создает связи между нейроном этого экземпляра и нейронами, переданными в списке аргументов метода.

Класс *Layer* реализует слой нейронов, он необходим для вычисления отклика слоя. Класс предоставляет два варианта метода *Layer::CalcResponse*. Вариант без параметров вызывается для скрытого или выходного слоя, а вариант с параметром (рис. 1) вызывается для входного слоя нейронов.

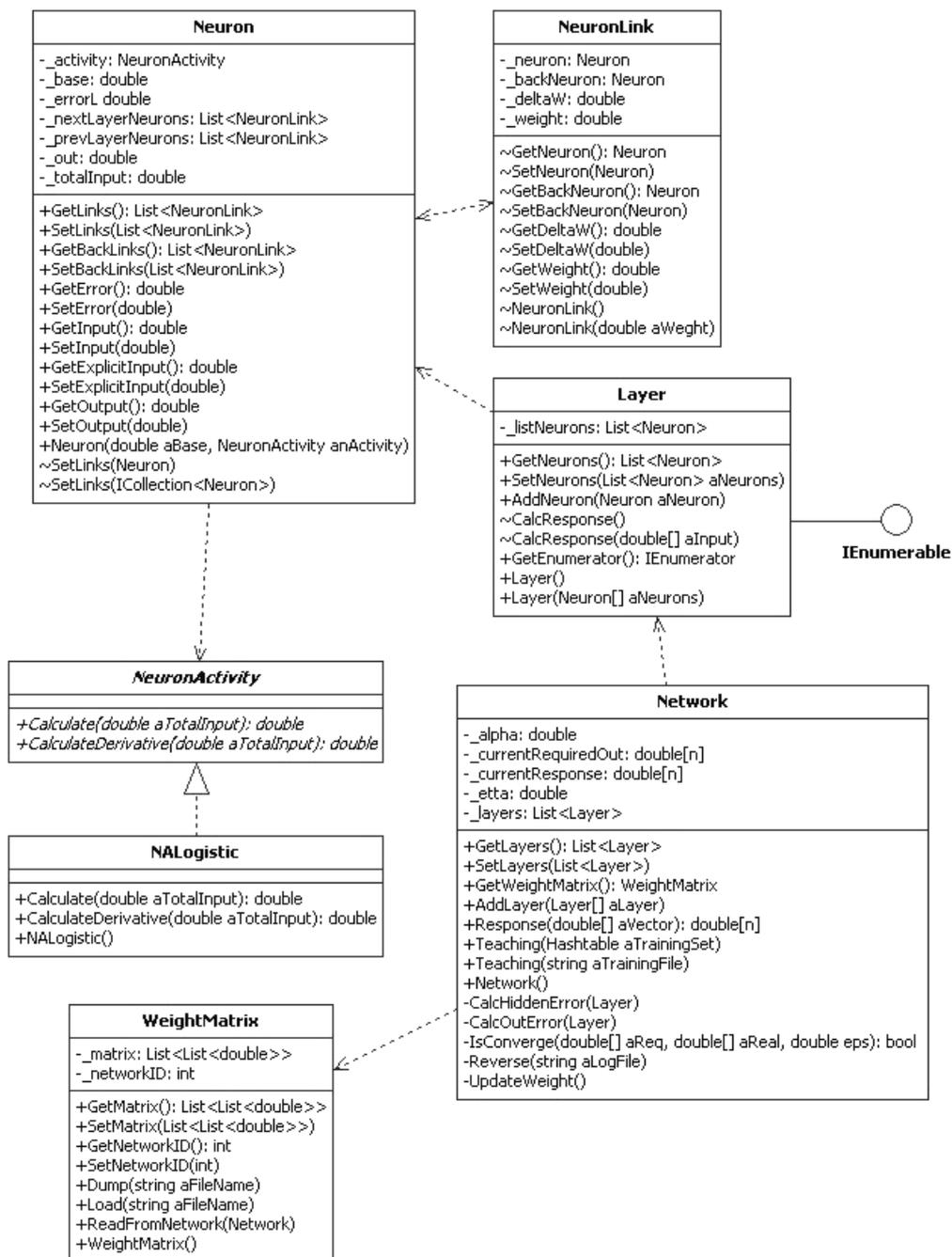


Рис. 1. Диаграмма классов для построения многослойного персептрон

Класс *NeuronActivity* — это абстрактный класс, который описывает функции активности нейрона. Он предоставляет два метода: *NeuronActivity::Calculate* и *NeuronActivity::CalculateDerivative*. Первый метод вычисляет значение функции, второй метод — значение производной этой же функции.

Класс *WeightMatrix* — это матрица весовых коэффициентов сети. Класс содержит идентификатор сети, предотвращающий попытку загрузки матрицы весов одной сети на другую. Матрицу можно загрузить из файла или сохранить в файл, а также создать на основе сети.

Класс *Network* реализует непосредственно нейронную сеть. В интерфейс класса вынесены методы *AddLayer* (добавить слой), *Response* (вычислить отклик сети) и *Teaching* (провести обучение сети). При добавлении нового слоя все связи устанавливаются автоматически. Метод *Response* реализует алгоритм прямого прохода по сети, а метод *Teaching* реализует алгоритм обучения с обратным распространением ошибок. Поля *Network::_alpha* и *Network::_etta* содержат константы обучения: параметр сходимости и норму скорости обучения, соответственно. *Network::_currentRequiredOut* — вектор, ожидаемый на выходе сети (используется при обучении сети). *Network::_currentResponse* — вектор, реально полученный на выходе (используется при обучении сети).

Все перечисленные классы легли в основу создания ИНС. Для проверки корректности ее работы использовалось два набора векторов исходных данных: обучающий и тестовый. Сеть успешно прошла этапы, как обучения, так и распознавания заданных образов.

Распознавание лиц на изображениях с помощью ИНС

Обнаружение и распознавание лиц на изображениях выполняется с помощью ПС, представленного следующими программными модулями (ПМ): графическим, обработки и хранения данных, детектором образов и анализатором образов. Детектор предназначен для определения наличия нужного образа на изображении, а также для локализации образа. Анализатор делает выводы о различных характеристиках обнаруженных образов [4]. Эти два ПМ реализованы на основе ИНС, построенной в виде многослойного персептрона с обратным распространением ошибок. Правила построения такой сети и работы с ней, алгоритм обучения ИНС, библиотека классов во многом схожи с предложенными в предыдущей задаче. Поэтому перейдем к рассмотрению результатов и правилам работы с созданным ПС.

Исходными данными для ПС является набор изображений. Причем, если использовать заготовки, содержащие только лица, процесс обучения может проводиться полуавтоматически, тогда ПС нужно только указать место расположения изображений. В случае, когда изображения не были подготовлены заранее, программу необходимо обучить в ручном режиме, выделив прямоугольный участок изображения, содержащий лицо. По завершении обучения приступают к поиску (рис. 2) и распознаванию образов (рис. 3).

Классификация текстов с помощью ИНС

Применение нейронных сетей позволяет решить многие задачи и в области обработки текстовой информации: поиск, автореферирование, авторубрикация и др. Тематику текста можно определить, например, на основе ключевых слов. Часто они определяются с помощью алгоритмов вычислительной лингвистики, основанных на законах Зипфа. Затем слова приводят к основной словоформе, т.е. к виду, в котором они представлены в словаре. Первоначально словарь состоит из выделенных терминов, относящихся к определенным темам. Словарь представляет собой вектор, каждой позиции которого соответствует определенное ключевое слово. По аналогии со словарем каждый документ тоже можно представить в виде вектора, поступающего на вход ИНС, в результате ее работы документ относится к той или иной рубрике. Размер словаря равен количеству входов у нейронов сети, и каждому ключевому слову соответствует своя позиция в векторе. Если документ характеризуется определенным ключевым словом, то при построении вектора документа в соответствующую слову позицию заносится 1, иначе — заносится 0.



Рис. 2. Поиск лиц на изображении

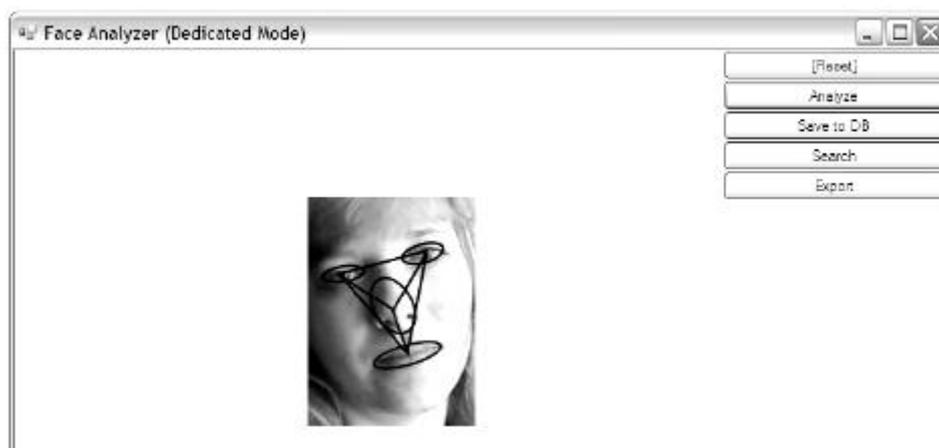


Рис. 3. Анализ и распознавание изображения

Рассмотрим алгоритм решения задачи рубрикации текстов с помощью сети Кохонена [3].

Присваиваются случайные начальные значения весовым коэффициентам нейронов. Затем на вход подается один из набора векторов. Входные векторы нормализуются.

Одним из нижеприведенных способов (шаг 2.1 или 2.2) рассчитывается выход сети Кохонена и выбирается нейрон-"победитель". Определяется скалярное произведение входного вектора и вектора весовых коэффициентов всех нейронов сети. Нейрон, обеспечивший максимальное значение, становится "победителем".

2.2. Вычисляется расстояние между входным вектором и векторами весовых коэффициентов. В этом случае "победителем" назначается нейрон, обеспечивший минимальное расстояние.

Каждый весовой коэффициент "победившего" нейрона изменяется пропорционально его разнице с соответствующим входным значением и заданной скоростью обучения на текущем этапе обучения. Изменения коэффициентов осуществляются по следующей формуле

$$w_n = w_c + \alpha(x - w_c),$$

где w_n — новое значение коэффициента; w_c — значение коэффициента на предыдущем шаге; x — значение входа нейрона; α — коэффициент скорости обучения, как правило, изначально берется равным 0,7 и на каждом шаге обучения уменьшается на заданную величину.

На основе ИНС создано ПС, способное выполнять авторубрикации заданных текстов. Возможности ИНС проверены на выборке из 14 текстов. Чтобы легче ориентироваться в результатах, выдаваемых сетью, заранее определены четыре класса, к которым принадлежит каждый текст. Словарь ключевых слов содержит 16 терминов. По вышеописанному правилу составляются векторы входных данных для каждого текста. Они подобраны таким образом, что по некоторым позициям в векторах имеются совпадения.

Первоначально сеть Кохонена задавалась решеткой размерностью 2×2 , т.е. количество нейронов совпадало с предполагаемым количеством классов. В результате сеть не смогла разделить тексты, относящиеся к 3 и 4 классам, они были отнесены к одному нейрону и, следовательно, распознаны как объекты, относящиеся к одному и тому же классу.

Некорректная работа сети могла быть вызвана несколькими причинами. Во-первых, большим сходством входных данных в текстах, относящихся к разным классам. Во-вторых, недостаточным размером сети, состоящей из четырех нейронов при классификации объектов на четыре класса. Поэтому необходим запас нейронов в ИНС. При увеличении решетки до девяти нейронов проблема была решена, и сеть корректно определяла классы для текстов уже после тридцати итераций обучения. Рассмотренный пример показал, что сеть способна выделять классы данных и может быть использована для авторубрикации текстов.

Кроме того, была протестирована выборка из 100 текстов, для чего использовалась сеть из 25 нейронов в идее решетки 5×5 . Из каждого текста было выделено 7–10 ключевых слов, которые после нормализации и исключения повторений вошли в словарь размерностью 80 единиц. После обучения на 300 примерах было выделено 10 тематических рубрик. Способность нейронной сети Кохонена выделять более мелкие кластеры данных зависит от механизма ее обучения. Разные модификации алгоритма работы сети делают карту Кохонена универсальным средством для решения множества задач: прогнозирования, моделирования процессов, классификации, сжатия и восстановления данных, усиления сигналов и др.

Заключение

Таким образом, решены актуальные задачи, связанные с распознаванием образов различной природы. Для этого были созданы ПС, функционирующие на базе искусственных нейронных сетей двух типов: персептрона и карты Кохонена.

Процедура поиска решения задачи с помощью сети, прошедшей обучение, оказывается более гибкой, чем процесс программирования, поскольку ИНС может повышать точность результатов по мере накопления ею опыта и адаптироваться к происходящим изменениям. Это связано с тем, что современные ИНС характеризуются такими свойствами, как: обучаемость, способность к обобщению и способность к абстрагированию. В результате нейронные сети превосходят последовательные машины в решении тех же задач, в которых человек превосходит вычислительную технику. Кроме того, решение задачи распознавания в большой мере зависит от выбора исходного описания объектов, т.е. пространства признаков. При удачном выборе задача распознавания может оказаться тривиальной и, наоборот, неудачно выбранное пространство признаков способно привести либо к очень сложной обработке информации, либо вообще к отсутствию решения.

APPLICATION OF NEURAL NETWORKS TO RECOGNITION OF IMAGES

L.V. SEREBRYANAYA, O.A. SHUSHINA

Abstract

Problems decisions of images recognition on the basis of artificial neural networks are given. Neural networks are presented in a kind of a perception and Kohonen cards. Principles of construction, training and use of neural networks are considered. Features of problems which are successfully solved by means of networks are specified. Algorithms of functioning, program realization and results of work of an artificial neural network are presented.

Литература

1. *Каллан Р.* Основные концепции нейронных сетей.: Пер. с англ. М., 2001.
2. *Хайкин С.* Нейронные сети. Полный курс второе издание.: Пер. с англ. М., 2006.
3. *Тархов Д.А.* Нейронные сети. Модели и алгоритмы. Радиотехника, 2005.
4. *Фор А.* Восприятие и распознавание образов. М., 1989.