

УДК 004.934

Д.А. Сорокоумова, О.Н. Корелин, А.В. Сорокоумов

**ПОСТРОЕНИЕ И ОБУЧЕНИЕ НЕЙРОННОЙ СЕТИ
ДЛЯ РЕШЕНИЯ ЗАДАЧИ РАСПОЗНАВАНИЯ РЕЧИ**

Нижегородский государственный технический университет им. Р.Е.Алексеева

Рассматривается способ решения задачи распознавания речи на примере распознавания отдельных слов ограниченного словаря с использованием нейронной сети прямого распространения, обучаемой методом обратного распространения ошибки. Приведены практические результаты обучения построенной нейронной сети при различных размерах обучающей выборки.

Ключевые слова: распознавание речи, нейронные сети, алгоритм обратного распространения ошибки, обучение, скорость обучения.

Введение

Задача распознавания речи – одна из наиболее актуальных задач современности. Несмотря на то, что на данный момент существует множество готовых систем распознавания речи, основанных на различных технологиях, задача распознавания речи не полностью решена, так как существующие системы имеют определенные недостатки. В частности, зависимость работы системы от доступа к средствам передачи данных и недостаточная точность распознавания.

Одним из перспективных направлений в решении задач распознавания речи является применение искусственных нейронных сетей. Нейронные сети широко применимы в решении различных классов задач распознавания образов ввиду способности к обобщению.

Исходные данные задачи

Рассмотрим аспекты построения и применения нейронных сетей для решения задачи распознавания речи на примере задачи распознавания чисел от 1 до 9, т.е. слов «раз», «два», «три», «четыре», «пять», «шесть», «семь», «восемь» и «девять» соответственно. Поскольку звуки человеческой речи лежат в частотном диапазоне от 100 до 4000 Гц, для решения поставленной задачи достаточно использовать частоту дискретизации 11025 Гц для оцифровки речевых сигналов. Использование данной частоты позволяет уменьшить поток аудиоданных, избежав при этом потери полезных составляющих сигнала. В рамках поставленной задачи звуковые сигналы представлены наборами фреймов, каждый из которых содержит 512 отсчетов.

На основе экспериментального анализа аудиозаписей различных вариантов произношения исследуемых слов была определена максимальная длительность полезного сигнала (рис. 1), составившая 1 с. Соответственно минимальный набор фреймов, охватывающий длительность полезного сигнала, должен состоять из 20 фреймов. Недостающие отсчеты исходного сигнала заполняются нулями.

В качестве входных данных для обучения нейронной сети будем использовать результаты преобразования Фурье, выполненного для каждого анализируемого фрейма. Такой подход позволяет анализировать сигнал как в частотной области (использование спектра фрейма), так и во временной – путем разбиения исходного сигнала на фреймы. Так как значимая информация содержится в действительном частотном спектре, после выполнения преобразования Фурье, мы используем действительный спектр сигнала, отбрасывая информацию о фазе (рис. 2, б).

На выходе нейронной сети ожидаем число, находящееся в диапазоне от 1 до 9 и однозначно соответствующее своему словесному представлению, поданному на вход нейронной сети.

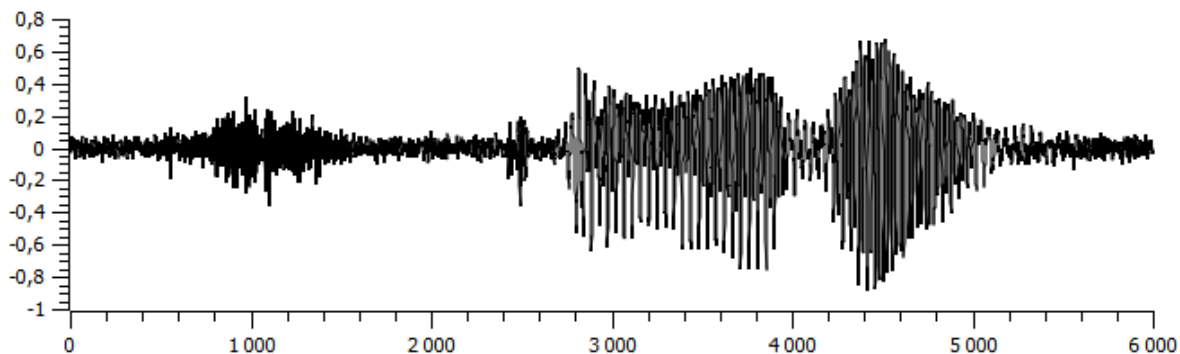


Рис. 1. Временная диаграмма слова «четыре»

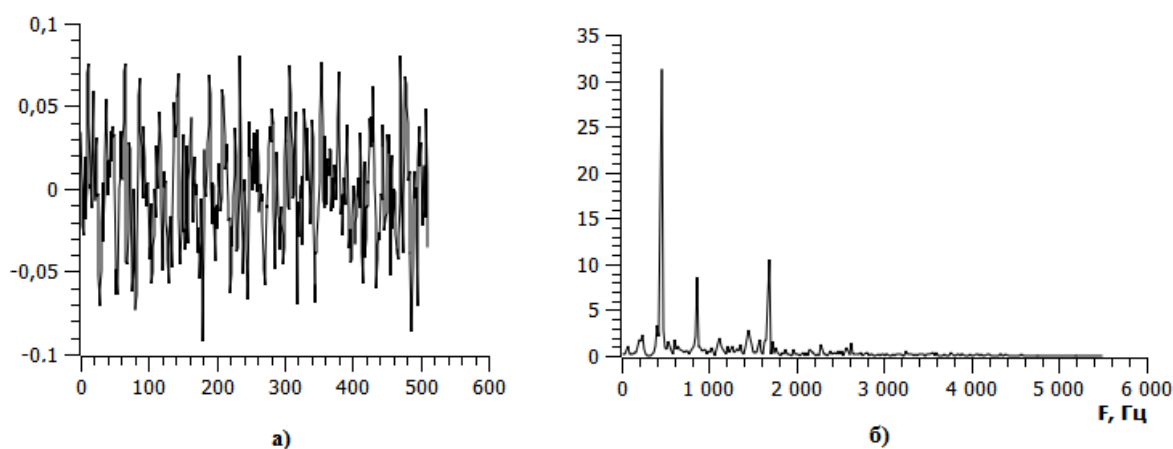


Рис. 2. Первый фрейм сигнала «четыре»:

a – временная диаграмма; *б* – спектр выделенного фрейма

Нейросетевой подход к решению поставленной задачи

Нейронная сеть – это совокупность соединенных и взаимосвязанных между собой искусственных нейронов, аккумулирующих входные значения и генерирующих выходной сигнал при помощи функции активации. Работу одного нейрона можно представить формулой

$$y_j = F\left(\sum w_{ij}x_i\right), \quad (1)$$

где y_j – выходной сигнал j нейрона; w_{ij} – вес связи между i и j нейронами; x_i – выходной сигнал i нейрона; F – функция активации нейрона.

Способы соединения нейронов в искусственных нейронных сетях определяют топологию нейронной сети. По структуре межнейронных связей можно выделить два вида нейронных сетей: нейронные сети прямого распространения и рекуррентные нейронные сети. В нейронных сетях прямого распространения связь между слоями однонаправленная – каждый нейрон связан только с нейронами следующего слоя. Такие сети являются статическими из-за отсутствия обратных связей и динамических элементов. Выход подобной сети зависит только от входных данных. Рекуррентные нейронные сети являются динамическими, благодаря наличию обратных связей. Выход рекуррентной нейронной сети зависит от её предыдущего состояния. Наиболее распространенными архитектурами нейронных сетей являются: сети прямого распространения; сети Хопфилда; карты Кохонена; машины Больцмана; сети на основе Адаптивной Резонансной Теории (АРТ); радиальные базисные функции (РБФ); каскадно-корреляционные сети.

Топология нейронной сети выбирается непосредственно для анализируемой задачи, учитывая особенности и сложность ее решения. Для решения некоторых типов задач уже существуют оптимальные конфигурации. Однако, если задача не может быть сведена ни к одному из известных типов, требуется синтезировать новую конфигурацию нейронной сети непосредственно для решаемой задачи. Поскольку не существует общего метода выбора оптимальной конфигурации нейронной сети, структура нейронной сети подбирается экспериментальным образом.

Наиболее наглядной структурой обладают сети прямого распространения сигнала, названные так ввиду того, что нейроны одного слоя могут быть соединены только с нейронами близлежащих слоев без обратных и рекуррентных связей. Обычно такие сети состоят из входного слоя, одного или нескольких скрытых слоев и выходного слоя. Простейшая структура такой сети изображена на рис. 3. Данная сеть имеет один скрытый слой, входной слой, состоящий из n нейронов и выходной слой, состоящий из m нейронов.

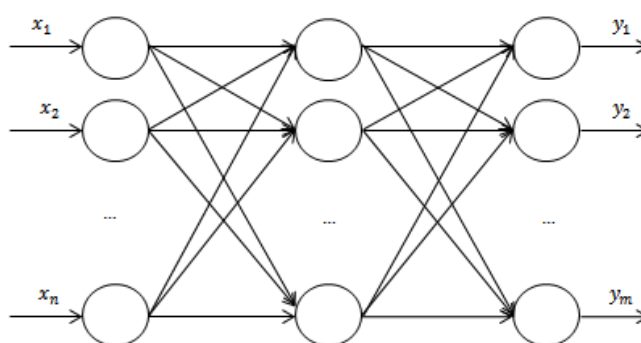


Рис. 3. Типовая структура нейронной сети

При помощи такой нейронной сети данные преобразуются из n -мерного входного пространства в m -мерное выходное. Преимуществом такого типа нейронных сетей является их относительная простота и наглядность, что позволяет анализировать работу используемой нейронной сети.

Исходя из формата входных и выходных данных, для решения поставленной задачи будем использовать нейронную сеть прямого распространения сигнала, входной слой которой содержит такое число нейронов, которое соответствует количеству анализируемых признаков (то есть количеству фреймов, умноженному на число анализируемых спектральных составляющих). Рекуррентные нейронные сети не могут быть использованы при решении поставленной задачи, так как из-за наличия обратных связей выходные значения рекуррентной нейронной сети зависят от предыдущего состояния сети, а поскольку произносимые слова в рамках решаемой задачи не связаны между собой, то предыдущее состояние сети не должно влиять на результат распознавания.

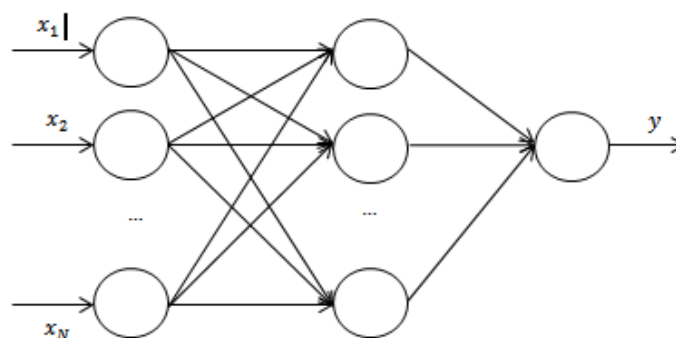


Рис. 4. Используемая структура нейронной сети

Для решения задач распознавания речи наиболее распространенным решением является использование такого числа нейронов в выходном слое, которое соответствует количеству распознаваемых объектов. Однако при решении данной задачи была выбрана архитектура нейронной сети, содержащая в выходном слое один нейрон, выходное значение которого находится в диапазоне от 0 до 1, что соответствует числам от 1 до 9. Используемая нейронная сеть также имеет один скрытый слой (рис. 4).

После определения топологии нейронной сети необходимо выбрать алгоритм обучения. Обучение нейронной сети может производиться двумя способами – с учителем и без учителя. При использовании обучения с учителем нейронной сети представляют пары векторов входных и выходных данных, в результате чего вычисляется ошибка и выполняется настройка весов нейронной сети. Алгоритм повторяется до тех пор, пока ошибка нейронной сети не достигнет требуемого минимального значения. В случае обучения без учителя, выходные данные заранее неизвестны, поэтому используются только входные данные.

При выборе алгоритма обучения необходимо учитывать топологию нейронной сети, модель анализируемых данных и предполагаемый способ обучения нейронной сети. Поскольку нами была выбрана нейронная сеть прямого распространения, будем использовать наиболее известный алгоритм обучения многослойного персептрона – алгоритм обратного распространения ошибки.

Обучение нейронной сети методом обратного распространения ошибки

Алгоритм обратного распространения ошибки предполагает вычисление ошибки, как выходного слоя, так и каждого нейрона обучаемой сети, а также коррекцию весов нейронов в соответствии с их текущими значениями. На первом шаге данного алгоритма веса всех межнейронных связей инициализируются небольшими случайными значениями (от 0 до 1). После инициализации весов в процессе обучения нейронной сети выполняются следующие шаги:

- прямое распространение сигнала;
- вычисление ошибки нейронов последнего слоя;
- обратное распространение ошибки.

Прямое распространение сигнала производится послойно, начиная со входного слоя, при этом рассчитывается сумма входных сигналов для каждого нейрона и при помощи функции активации генерируется отклик нейрона, который распространяется в следующий слой с учетом веса межнейронной связи по формуле (1). В результате выполнения данного этапа мы получаем вектор выходных значений нейронной сети.

Следующий этап обучения – вычисление ошибки нейронной сети как разницы между ожидаемым и действительным выходными значениями. Вычисление ошибки производится для каждого нейрона выходного слоя в соответствии с формулой

$$\delta_k = (EXP_k - y_k)F'(y_k), \quad (2)$$

где δ_k – полученная ошибка k -го нейрона выходного слоя; EXP_k - ожидаемое значение для k выходного нейрона; y_k - фактическое выходное значение k нейрона; $F'(y_k)$ – производная функции активации k нейрона.

Для последующих слоев нейронной сети ошибка нейрона вычисляется с использованием формулы

$$\delta_k = F'(y_k) \cdot \sum_{i=1}^M \delta_i w_{ki}, \quad (3)$$

где δ_k – полученная ошибка для k нейрона; δ_i - ошибка i -го нейрона предыдущего слоя; w_{ki} - вес связи между k нейроном текущего слоя и i нейроном предыдущего слоя; y_k - фактиче-

ское выходное значение k нейрона; $F'(y_k)$ – производная функции активации k нейрона; M – количество нейронов предыдущего слоя.

Полученные значения ошибок распространяются от последнего, выходного слоя нейронной сети, к первому. При этом вычисляются величины коррекции весов нейронов в зависимости от текущего значения веса связи, скорости обучения и ошибки, внесенной данным нейроном. Данный процесс описывается формулой

$$w_{ji} = w_{ji} + h\delta_i y_j \quad (4)$$

где w_{ji} – вес связи между j -м нейроном анализируемого слоя и i -м нейроном последующего слоя; h – параметр, определяющий скорость обучения; δ_i – ошибка i -го нейрона последующего слоя; y_j – выходное значение j -го нейрона текущего слоя.

После завершения данного этапа шаги описанного алгоритма повторяются до тех пор, пока ошибка выходного слоя не достигнет требуемого значения.

При коррекции весов межнейронных связей используется понятие скорости обучения. Скорость обучения нейронной сети – один из наиболее важных параметров, контролирующих процесс обучения. Данный параметр определяет величину изменения весовых коэффициентов межнейронных связей. Для идеального приближения к минимуму ошибки нейронной сети скорость обучения должна стремиться к бесконечно малому значению для обеспечения наилучшей сходимости алгоритма обучения. Однако чем меньше выбранное значение шага обучения, тем дольше происходит обучение сети. Таким образом, при стремлении шага обучения к бесконечно малому значению время, требуемое для обучения нейронной сети, многократно возрастает. В то же время, если выбрать слишком большое значение скорости обучения, то минимум ошибки не будет достигнут нейронной сетью – величина ошибки будет колебаться около минимального значения ввиду слишком большой коррекции весов межнейронных связей.

С целью преодоления указанных проблем используется так называемая динамическая скорость обучения. При использовании данного метода шаг обучения не является постоянной величиной, а зависит от других параметров процесса обучения (времени, номера итерации или ошибки нейрона на предыдущем шаге). Динамическая скорость обучения может быть введена как для каждого нейрона сети в отдельности, так и для всей сети в целом.

Функции, используемые для вычисления скорости обучения, должны обладать следующими свойствами:

- 1) $Y(x) = 0$ при $x = 0$;
- 2) $Y(x) = MAX$ при $x \rightarrow \pm\infty$;
- 3) $Y(x) \rightarrow 0$ при $x \rightarrow 0$.

Для работы с нейронной сетью нами была выбрана следующая функция, отражающая зависимость скорости обучения нейрона от величины ошибки:

$$Y(x) = |MAX * (-CST * |x|)|, \quad (5)$$

где MAX – константа, определяющая максимально возможную скорость обучения; x – величина ошибки, внесенной нейроном; CST – константа, определяющая степень крутизны результирующей функции. Функция представлена графиком на рис. 5.

Данная функция отвечает указанным требованиям и обеспечивает наиболее оптимальное изменение скорости обучения. В начале процесса обучения параметр MAX выставляется в максимальное значение скорости обучения (в нашем случае $MAX = 3$), в результате чего при больших значениях ошибки обучения изменения весовых коэффициентов будут значительны. По мере снижения ошибки нейрона скорость обучения будет снижаться, и при стремлении ошибки обучения к нулю скорость обучения будет также стремиться к нулю.

Таким образом, при решении поставленной задачи было реализовано динамическое

управление скоростью обучения, при котором величина шага обучения вычисляется для каждого нейрона в отдельности в зависимости от ошибки, внесенной данным нейроном. Введение данного алгоритма позволило точнее приближаться к минимуму ошибки обучения нейронной сети. При сравнении характера обучения нейронной сети с адаптивной скоростью обучения и нейронной сети с минимальным фиксированным шагом обучения первая демонстрирует более гладкое стремление ошибки к минимальному значению без значительных колебаний.

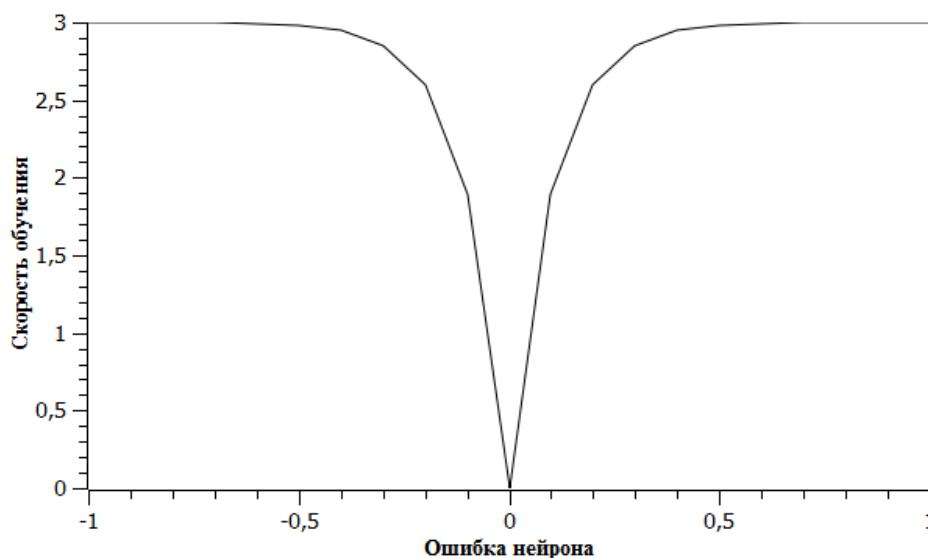


Рис. 5. График скорости обучения отдельного нейрона

Рассмотрим две разновидности ошибок нейронной сети, наиболее полно характеризующих процесс обучения. В процессе обучения нейронной сети различают ошибку обучения и ошибку обобщения. Ошибка обобщения – это ошибка, которую нейронная сеть демонстрирует на примерах, не участвовавших в процессе обучения. Ошибкой обучения, напротив, называется ошибка, которую обучаемая нейронная сеть демонстрирует на примерах обучающей выборки. Теоретическая зависимость данных ошибок от времени обучения проиллюстрирована на рис. 6.

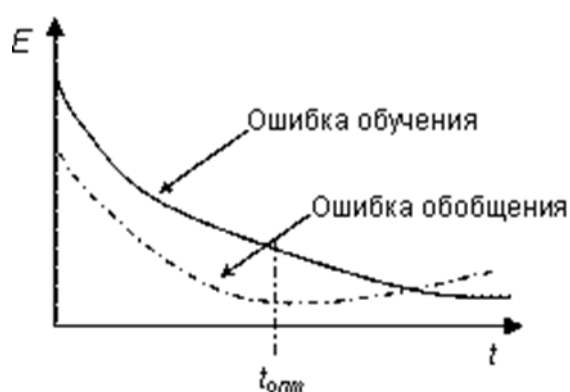


Рис. 6. Зависимость ошибок обобщения и обучения от времени

Из графика на рис. 6 видно, что в процессе обучения ошибка обучения постоянно уменьшается, пока не достигает достаточно малого значения, после чего обучение прекращается. Однако если параллельно отслеживать ошибку обобщения, то можно увидеть, что она сначала также уменьшается, но с некоторой эпохи обучения начинает возрастать из-за эффекта переобучения. Это означает, что, добиваясь большей точности решения на обучающем множестве, исследуемая сеть теряет часть обобщающей способности. Поэтому процесс обу-

чения необходимо останавливать, как только ошибка обобщения начинает монотонно возрастать.

Немаловажным аспектом обучения нейронной сети является обучающая выборка. Обучающая выборка – это набор пар входных и выходных данных (для обучения с учителем), используемых при обучении нейронной сети. Контрольная выборка – часть наборов, не участвующих в обучении нейронной сети – служит для определения ошибки обобщения.

Для корректного обучения нейронной сети обучающая выборка должна обладать свойством репрезентативности. Под репрезентативностью в данном случае следует понимать наличие достаточного количества разнообразных обучающих примеров, отражающих закономерности, которые должны быть обнаружены нейронной сетью в процессе обучения. Репрезентативность обучающей выборки выражается в следующих аспектах:

- достаточность: число обучающих примеров должно быть достаточным для обучения. Для достижения нулевой ошибки необходимо, чтобы число обучающих примеров было в несколько раз больше, чем число весов межнейронных связей, в противном случае модель может не приобрести способности к обобщению;
- разнообразие: обучающая выборка должна содержать большое число разнообразных комбинаций входных и выходных данных в обучающих примерах. Способность нейронной сети к обобщению не будет достигнута, если число примеров достаточное, но все они одинаковые, т.е. представляющие лишь часть классов, характерных для исходного множества;
- равномерность представления классов: примеры различных классов должны быть представлены в обучающей выборке в одинаковых пропорциях. Если один из классов будет преобладать, то данный класс будет определен моделью как наиболее вероятный для любых новых наблюдений.

Увеличение количества примеров в обучающей выборке способствует увеличению времени, необходимого для достижения нейронной сетью заданных показателей по ошибке обобщения.

При обучении построенной нейронной сети были получены результаты, подтверждающие теоретическую зависимость ошибки обобщения от мощности обучающей выборки (рис. 7, а). Также была выявлена зависимость между мощностью обучающей выборки и отклонением ошибки обобщения от установившегося значения (рис. 7, б).

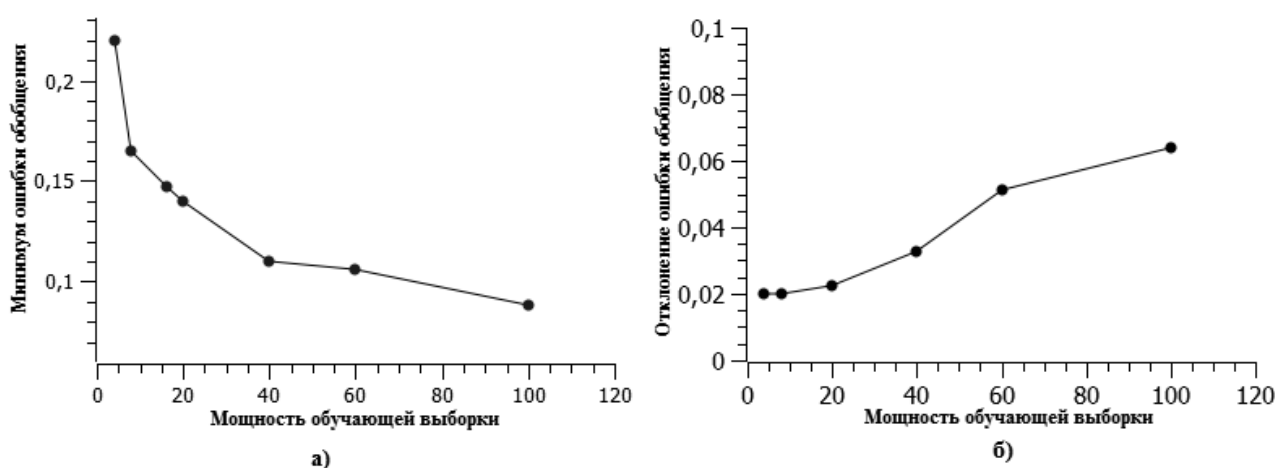


Рис. 7. Поведение ошибки обобщения в зависимости от мощности обучающей выборки:

а – изменение минимального значения ошибки обобщения; *б* – изменение отклонения ошибки обобщения от установившегося значения

На основе полученных результатов можно сделать вывод, что при увеличении объема обучающей выборки минимально возможное значение ошибки обобщения уменьшается, при-

чем характер зависимости ошибки обобщения от мощности обучающей выборки совпадает с теоретическим. Однако при этом увеличиваются временные затраты на обучение нейронной сети, а также увеличивается отклонение ошибки обобщения от установившегося значения.

Таким образом, была реализована модель нейронной сети для решения задачи распознавания слов, соответствующих произношению чисел от 1 до 9. При анализе поведения нейронной сети было определено, что существующего размера обучающей выборки недостаточно для достижения нулевой ошибки обобщения нейронной сети. Однако, построенная сеть показала способность к обучению, подтвержденную экспериментальными данными. При достижении заданной ошибки обобщения программа сохраняет весовые коэффициенты в форме заголовочного файла, что дает возможность восстановить обученную нейронную сеть для последующего использования.

Библиографический список

1. **Сергиенко, А. Б.** Цифровая обработка сигналов / А.Б. Сергиенко. – СПб.: Питер, 2007. – С. 751.
2. **Леонович, А.А.** Современные технологии распознавания речи / А.А. Леонович // Диалог 2005: материалы Междунар. конф. по компьютерной лингвистике. – М., 2005.
3. **Круг, П.Г.** Нейронные сети и нейрокомпьютеры: учеб. пособие / П.Г. Круг. – М.: Изд-во МЭИ, 2002. – 176 с.
4. Элементарное введение в технологию нейронных сетей с примерами программ: [пер. с польск. И.Д. Рудинского] / Р. Тадеусевич [и др.] /. – М.: Горячая линия – Телеком, 2011.
5. **Хайкин, С.** Нейронные сети. Полный курс: [пер. с англ.] / С. Хайкин. – 2-е изд. – М.: Вильямс, 2006. – 1104 с.
6. **Галушка, В.В.** Формирование обучающей выборки при использовании искусственных нейронных сетей в задачах поиска ошибок баз данных / В.В. Галушка, В.А. Фатхи [Электронный ресурс] // Инженерный вестник Дона. 2013. №2.: <http://www.ivdon.ru/magazine/archive/n2y2013/1597>

*Дата поступления
в редакцию 02.07.2015*

D. Sorokoumova, O. Korelin, A. Sorokoumov

DEVELOPMENT AND TRAINING OF THE NEURAL NETWORK FOR VOICE RECOGNITION SOLUTION

Nizhny Novgorod state technical university n.a. R.E. Alexeev

Purpose: Create a neural network mode for recognition of separate words solution, analyze training characteristics and behavior of constructed neural network. Target network had to be portable to DSP platforms that define a format of input data.

Design/methodology/approach: Input signal is divided to 20 frames, each of them contain 512 samples. For every frame FFT gives 255 spectral items those are being provided to the input neurons of the neural network. Based on input data and the output requirements a feed-forward neural network has been chosen. This kind on network is most clear for analysis and development. To train selected neural network model backpropagation algorithm was selected.

Findings: Designed neural network demonstrated expected behavior related to learning and generalization errors. It was found that even if generalization error decreases while increasing training sequence, the errors starts oscillating regardless of introducing dynamic learning speed. The network has been trained enough to match requirements for generalization error, but there is still a possibility to improve generalization error.

Key words: speech recognition, neural networks, backpropagation, learning speed.