

УДК 004.42

Н.А. Иванова

**АНАЛИЗ ЭФФЕКТИВНОСТИ ПРИМЕНЕНИЯ НЕЙРОННЫХ СЕТЕЙ
В ЗАДАЧАХ РЕЧЕВОЙ ИДЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЯ**

Аннотация. В данной работе рассматриваются возможности нейронных сетей прямого распространения сигнала применительно к задаче текстонезависимого распознавания пользователей по речевому сигналу. В качестве характеристик речевого сигнала использованы логарифмически масштабированные кепстральные коэффициенты. Анализируется точность распознавания разными типами нейронных сетей для разного количества пользователей.

Ключевые слова: речевая идентификация пользователей, искусственные нейронные сети, логарифмически масштабированные кепстральные коэффициенты, точность распознавания.

Введение. Системы определения индивидуальности пользователя по голосу развиваются в последнее время очень активно. Такой интерес обусловлен большим количеством практических задач, которые можно решить с помощью речевого распознавания пользователя. Проверка прав доступа к различным системам (информационным и физическим) по голосу пользователя имеет ряд преимуществ по сравнению с распознаванием по другим показателям. Во-первых, голос не отчуждаем от человека в отличие от ключа, магнитной карты. Во-вторых, он не требует непосредственного контакта с пропускной системой, как это необходимо для отпечатка пальца, ладони (возможно использование телефонного канала).

Речевая идентификация пользователя предполагает следующую ситуацию. Существует ограниченная и строго контролируемая группа пользователей системы. При поступлении речевого сигнала на вход системы определения индивидуальных речевых характеристик, эта система должна идентифицировать, кто из пользователей произносит данную речевую информацию.

Постановка задачи. Целью данной работы является проведение анализа возможностей нейронных сетей прямого распространения при идентификации пользователей по речевому сигналу. При

© Иванова Н.А., 2009

использовании нейронных сетей для речевой идентификации, предполагается, что пользователи могут произносить любые предложения. То есть анализируется текстонезависимый вариант речевого распознавания пользователя. Сравнительный анализ разных архитектур нейронных сетей (с разным количеством слоев) для задачи идентификации пользователя, позволит выбрать оптимальную схему нейронной сети. В данной работе также необходимо оценить результаты процесса распознавания для разного количества дикторов и исследовать результаты работы нейронных сетей с зашумленными входными данными.

Решение задачи. Идентификация пользователей с помощью трехслойной нейронной сети прямого распространения. Для распознавания речевых сигналов и пользователей по речевым сигналам чаще всего в качестве характеристик сигнала используют логарифмически масштабированные кепстральные коэффициенты (MFCC – mel frequency cepstral coefficients). Исходный речевой сигнал разбивается на фреймы длиной 1024 отсчета (5 0,046 с при частоте дискретизации 22050 Гц) с перекрытием, составляющим примерно 1/3~1/2 часть длины фрейма. Для каждого фрейма вычисляются 12 кепстральных коэффициентов [1].

Для проведения исследований использовалась база речевых сигналов, которая содержит 20 предложений, каждое длительностью 5 секунд, для каждого из 20 пользователей. Для хранения информации о пользователях была организована структура данных с такой информацией: имя пользователя, дата регистрации пользователя, массивы кепстральных коэффициентов для всех предложений, произнесенных пользователем. Столбцы данного массива содержат логарифмически масштабированные кепстральные коэффициенты для одного фрейма.

Для идентификации пользователей проанализируем работу трехслойной нейронной сети прямой передачи сигнала [2, 3], для обучения которой используется метод обратного распространения ошибки. Входной слой нейронной сети содержит n нейронов, где n – количество распознаваемых пользователей. Выходной слой содержит один нейрон для определения номера пользователя. Схема нейронной сети представлена на рис. 1.

Нейронная сеть обучалась на сформированных массивах кепстральных коэффициентов. В процессе обучения фиксировалось время обучения. После завершения обучения нейронная сеть использовалась для распознавания всех предложений каждого диктора. Исходная база речевых сигналов была поделена на две половины: первая (условно назовем её внутренней базой речевых сигналов) - используется для обучения нейронной сети и для тестирования работоспособности обученной нейронной сети, вторая (внешняя база речевых сигналов) – будет использоваться только для тестирования нейронной сети. Внутренняя и внешняя база данных содержат по 10 предложений, произнесенных каждым из 20 пользователей.

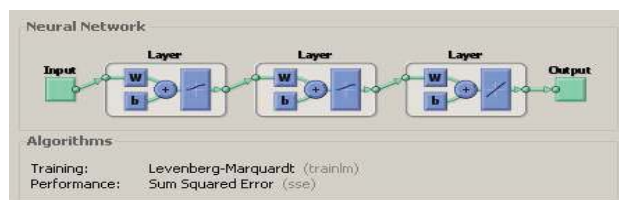


Рисунок 1 - Схема нейронной сети

Циклически перебирая все предложения для каждого пользователя, оценивалась точность распознавания обученной нейронной сети, как отношение количества предложений верно распознанных пользователей к общему количеству предложений. Результаты тестирования обученной нейронной сети представлены в таблице 1.

Для наглядного анализа работы нейронной сети, при идентификации пользователей, построим графики зависимости точности распознавания от количества дикторов (рис. 2).



Рисунок 2 - Зависимость точности распознавания нейронной сетью от количества дикторов

Таблица 1

Зависимость времени обучения трехслойной нейронной сети и точности распознавания от количества пользователей

Количество пользователей	Время обучения сети, сек	Точность распознавания (для внутренней базы речевых сигналов), %	Точность распознавания (для внешней базы речевых сигналов), %
2	78,73	100	100
3	95,02	100	100
4	65,39	100	100
5	123,34	100	100
6	136,18	100	100
7	127,83	100	100
8	379,08	94,17	90,83
10	359,73	87,33	84,67
15	1681,66	65,78	54,22
20	2429,72	35,79	35

Как видно из рис. 2, нейронная сеть обратного распространения ошибки отлично справляется с задачей распознавания пользователей для базы от 2 до 8 пользователей. При таком количестве пользователей время обучения нейронной сети сравнительно невелико (до 6 минут). При увеличении количества дикторов (от 10 и больше) точность распознавания резко уменьшается, а время обучения увеличивается. На двадцати дикторах сеть обучалась более 40 минут, а точность распознавания составляла всего 35%.

Идентификация пользователей с помощью двухслойной нейронной сети прямого распространения. Второй тип сети, который анализируется в данной работе с точки зрения её возможностей по идентификации пользователей, - это двухслойная сеть прямого распространения сигнала, т. е. сеть без скрытого слоя. Входной слой, как и в предыдущей сети, содержит n нейронов (n – количество распознаваемых пользователей). Выходной слой содержит один нейрон, который выдает значение, равное номеру пользователя. Схема двухслойной нейронной сети представлена на рис. 3.

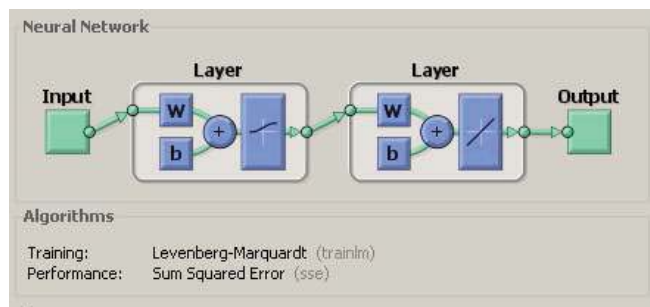


Рисунок 3 - Схема двухслойной нейронной сети

Результаты тестирования данного типа сети для разного количества пользователей приведены в таблице 2.

Таблица 2

Зависимость точности распознавания двухслойной сетью от количества дикторов

Количество пользователей	Время обучения сети, сек	Точность распознавания (для внутренней базы речевых сигналов), %	Точность распознавания (для внешней базы речевых сигналов), %
2	16.92	100	100
3	10.67	100	100
4	18.09	83.33	83.33
5	53.58	88.00	88.00
6	105.53	87.78	83.33
7	58.73	91.43	93.33
8	89.81	80.00	71.67
10	317.04	66.67	63.33
15	470.91	33.33	25.78
20	993.67	13.67	14.67

Для сравнения трехслойной и двухслойной сетей результаты тестирования обеих сетей приведены на рис. 4. Идентификация пользователей проводилась для внешней речевой базы данных.

По результатам, приведенным в таблицах 1 и 2, можно сделать вывод, что двухслойная нейронная сеть (без скрытого слоя) быстрее обучается и требует для обучения меньших аппаратных затрат, но задачу идентификации пользователя по речевому сигналу решает хуже. Двухслойная нейронная сеть смогла распознать с приемлемой

точностью только 7 дикторов, в то же время трехслойная сеть распознает 10 дикторов.



Рисунок 4 - Сравнительный анализ точности распознавания двухслойной и трехслойной нейронными сетями

В данной работе анализируется, при каких уровнях шума нейронная сеть способна распознавать пользователей. На исходные речевые сигналы накладывался белый шум. Трехслойная нейронная сеть прямого распространения содержит 10 нейронов во входном слое, 25 нейронов в скрытом слое, 1 нейрон в выходном слое. Данный тип нейронной сети был обучен на незашумленных речевых сигналах 10 пользователей. Для тестирования использовались зашумленные речевые сигналы и из внешней, и из внутренней речевых баз тех же 10 пользователей. Результаты проведенного исследования приведены на рис. 5.

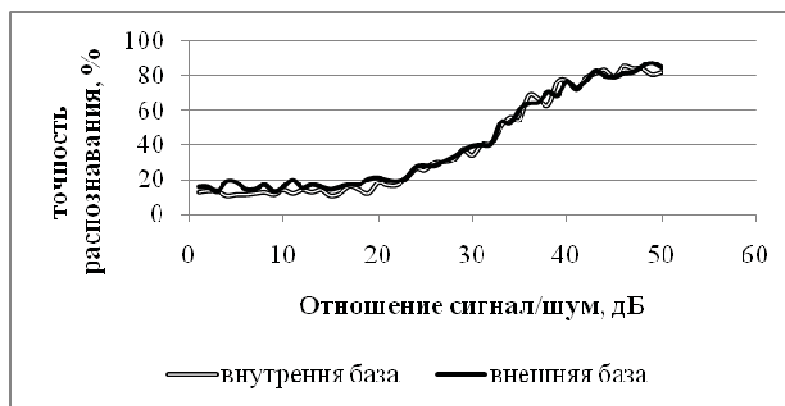


Рисунок 5 - Зависимость точности распознавания 10 пользователей от отношения сигнал/шум

Проанализировав результаты на рис. 5, можно отметить, что трехслойная нейронная сеть распознает пользователей с точностью

более 80% при отношении сигнал/шум более 40 дБ. При увеличении шума (отношении сигнал/шум менее 30дБ) точность распознавания резко уменьшается, и фактически пользователь выбирается нейронной сетью случайным образом. При отношении сигнал/шум менее 20 дБ нейронная сеть не способна распознать входной зашумленный сигнал.

Выводы. Таким образом, использование трехслойной нейронной сети прямого распространения сигнала позволяет создать систему для идентификации не более 10 пользователей. В этом случае время обучения нейронной сети сравнительно невелико при допустимой точности распознавания.

ЛИТЕРАТУРА

1. Аграновский А.В., Леднов Д.А. Теоретические аспекты алгоритмов обработки и классификации речевых сигналов. - Москва: Изд-во «Радио и связь», 2004. - 164 с.
2. Медведев В. С., Потемкин В. Г. Нейронные сети. Matlab6 /Под общ. ред. к. т. н. В. Г. Потемкина. – М.: ДИАЛОГ-МИФИ, 2002. – 496 с.
3. Осовский С. Нейронные сети для обработки информации / пер. с польского И.Д.Рудинского. - М.: Финансы и статистика, 2002. - 344 с.

Получено 15.10.2009г.