

УДК 519.688

АНАЛИЗ ЗВУКОВОГО СПЕКТРА ГОЛОСОВЫХ КОМАНД С ПОМОЩЬЮ НЕЙРОННОЙ СЕТИ*

К.А. ЧЕРДАНЦЕВ¹, А.В. КЛАДЬКО²

¹ 630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, студент факультета автоматики и вычислительной техники. E-mail: medmene@yandex.ru

² 630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, студент факультета автоматики и вычислительной техники. E-mail: tonkladko@ya.ru

Задача анализа голоса стала востребованной с широким распространением компьютеров и их периферии. Сама по себе задача анализа речи человека позволяет открыть новые области статистики, психологии и т. д., в которых с помощью машинных вычислений можно анализировать речь реальных людей в таких объемах, в которых это неподвластно обычному человеку или даже группам людей. Слово или фраза, сказанные человеком в реальности, в компьютерном представлении описываются огромным числом параметров, и поэтому только с недавнего времени вычислительные мощности компьютеров позволяют в реальном времени обрабатывать видео- и аудиофайлы. В связи с этим стали появляться такие системы, как «умный дом», в основе которых заложено значительное упрощение взаимодействия машины и человека. В данной системе основным органом управления являются голосовые или даже звуковые команды человека, будь то слово или хлопок. Помимо этой системы широко используется анализ голоса в сети Интернет, где при поиске разного рода информации можно использовать голосовой поиск, к примеру, в случаях когда нет возможности набрать текстовый запрос, и это очень удобно, особенно в эру смартфонов. В будущем при технологиях очень развитого распознавания и анализа речи можно воспроизвести реальный механизм «общения» человека и машины, что позволит развивать и обучать машину как человека.

Ключевые слова: нейронные сети, алгоритмы, эксперимент, сортировка, многослойная нейронная сеть, однослойная нейронная сеть, числовые массивы, оценка алгоритмов, программирование

DOI: 10.17212/2307-6879-2017-4-65-76

* Статья получена 07 августа 2017 г.

ВВЕДЕНИЕ

Для обработки голоса используются разные методы, а также кепстральный анализ, модели Маркова, ЦАП, АЦП и т. д. В нашей работе мы решили использовать для обработки голоса статистический механизм – нейронные сети. Чтобы отправить данные в нейронную сеть, нужно воспользоваться определенными методами, которые позволят как-то уменьшить количество амплитудных показателей исходного звукового сигнала. Обычно в этих случаях и используются методы, перечисленные выше. В нашей задаче для подготовки начальных данных было использовано два метода: кепстральный анализ и метод усреднения (математическое ожидание). После первичной обработки сигналы приводились в более удобный для нейронной сети вид, который позволяет не злоупотреблять количеством нейронов, чтобы не усложнять и не замедлять финальную обработку сигнала.

1. ПОСТАНОВКА ЗАДАЧИ

Для реализации распознавания и обработки голоса необходимо реализовать классы нейронов и нейронную сеть, отвечающие за конечный анализ и получение ответа, а также реализовать предварительную обработку данных в одном случае сведением всех данных к одной числовой константе, в другом – методом кепстрального анализа. Структуры сети будут различаться в зависимости от типа обработки данных.

2. РЕАЛИЗАЦИЯ ПРЯМОГО АНАЛИЗА

Непосредственным или прямым анализом является второй метод. Суть его состоит в уменьшении количества амплитудных значений до фиксированного количества. Таким числом (далее – константа) было выбрано число 10 000, для того чтобы очень незначительно терялась точность обучения. Фактически мы подаем на вход чистые, необработанные данные.

Проект был разделен на следующие части:

- приведение данных к верхнему или нижнему пределу;
- усреднение данных до фиксированного числа;
- создание структуры нейронной сети.

Приведение данных к пределам

Частота дискретизации звукового сигнала очень велика, и если полный звуковой сигнал подавать в нейронную сеть, то появляется две проблемы:

разные структуры сети для разных звуковых сигналов и большое количество нейронов, как следствие – большое время работы.

Чтобы справиться с проблемой разных структур сети, был разработан следующий алгоритм.

- Проверка на кратность константе. Случается очень редко и позволяет сразу перейти к следующему пункту.
- Отбрасывание лишних амплитудных значений. Количество начальных амплитудных значений округляется до кратности 100, значения убираются с конца.
- Проверяем остаток от деления на константу, меньше ли он 5000. Эта проверка позволяет определить несколько незначущих амплитудных значений (добавить их или отбросить).
- В случае остатка меньше 5000 убираем половину остатка с начала и с конца звукового сигнала.
- В противоположном случае вычисляем необходимое количество добавочных амплитудных значений по формуле $a = 10\,000 - o$, где a – количество амплитудных значений, o – остаток. Добавочные амплитудные значения распределяются поровну в начале и в конце звукового сигнала путем дублирования имеющихся амплитудных значений.
- После всех проведенных манипуляций получаем данные, кратные константе. Полученные данные используются в следующих пунктах.

Усреднение данных

После первоначальной обработки количества амплитудных значений переходим непосредственно к изменению самой амплитуды. При прочтении wav файла мы получаем только положительные семплы, хотя в реальном звуковом сигнале они как положительные, так и отрицательные. Приводим семплы к их естественному виду по правилу: если семпл больше $32\,768 (2^{15})$, то отнимаем от него $65\,536 (2^{16})$. Это правило позволяет получить изначальный рисунок сигнала.

Следующим шагом приводим все амплитудные значения к константе. Для начала вычислим свертку путем деления количества амплитудных значений на константу. Далее каждые n амплитудных значений (n – число, соответствующее свертке) усредняются и помещаются во входной массив.

Создание структуры нейронной сети

Нейронная сеть в данной реализации будет многослойной. Входной слой будет принимать значения, полученные после обработки. Первый и второй

слои скрытые, они и будут анализировать поступающие данные. В выходном слое будут находиться нейроны в количестве, равном количеству команд. Каждый из нейронов выходного слоя отвечает за свою команду.

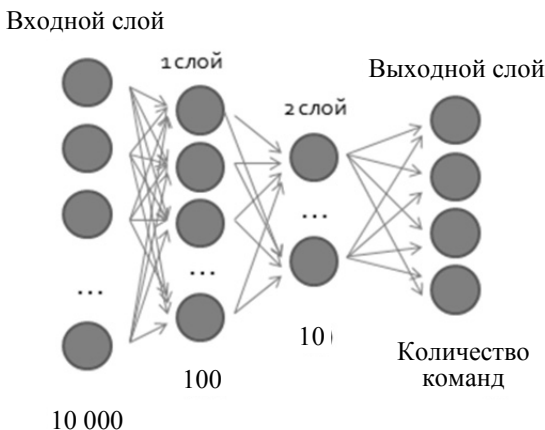


Рис. 1. Структура нейронной сети прямого анализа

3. РЕАЛИЗАЦИЯ КЕПСТРАЛЬНОГО АНАЛИЗА

Метод реализации с использованием кепстрального анализа является первым методом в рамках данной статьи. Его суть – в уменьшении количества входных амплитудных значений до фиксированного количества. Входные данные разделяются на фреймы – небольшие временные промежутки, для каждого фрейма производится кепстральный анализ, и полученные результаты передаются на вход нейронной сети.

Проект был разделен на следующие части:

- разделение исходных данных на фреймы;
- вычисление для каждого фрейма мел-частотных кепстральных коэффициентов;
- усреднение полученных коэффициентов между фреймами;
- создание структуры нейронной сети.

Разделение исходных данных на фреймы

В конкретной точке данные малоинформативны, звуковой сигнал гораздо информативнее, если рассматривать его на некотором промежутке – фрейме.

Исходные данные разделяются на фреймы, которые располагаются внахлест, что позволяет сгладить результаты анализа фреймов, превращая идею фреймов в некоторое «окно», движущееся вдоль значений сигнала.

Вычисление для каждого фрейма мел-частотных кепстральных коэффициентов

Вычисления начинаются с умножения каждого фрейма на оконную функцию, что позволяет устранить разрывы на границах периодов, так как анализируемый сигнал не является периодическим. В качестве оконной функции мы выбрали функцию Хэмминга, которая плавно сводит на нет сигнал вблизи краев анализируемого участка. Далее выполняется быстрое преобразование сигнала Фурье.

На следующем этапе мы конвертируем частоты, полученные быстрым преобразованием Фурье в мел-частоты.

Затем мы выполняем расчет мел-фильтров, т. е. переходим к психофизической единице высоты звука, основанной на субъективном восприятии среднестатистическими людьми. Теперь нам нужно наложить полученную шкалу на спектр нашего фрейма и рассчитать искомые коэффициенты.

Усреднение полученных коэффициентов между фреймами

Распознаваемое слово состоит из нескольких фреймов, для каждого из которых были рассчитаны мел-частотные кепстральные коэффициенты.

Среди всех фреймов, которые относятся к распознаваемому слову, находим итоговое значение коэффициентов – среднее значение среди всех соответствующих коэффициентов.

Создание структуры нейронной сети

Нейронная сеть в данной реализации будет многослойной. Входной слой будет принимать значения, полученные после обработки. Первый слой – скрытый – будет анализировать поступающие данные. В выходном слое будут находиться нейроны в количестве, равном количеству команд. Каждый из нейронов выходного слоя отвечает за свою команду.

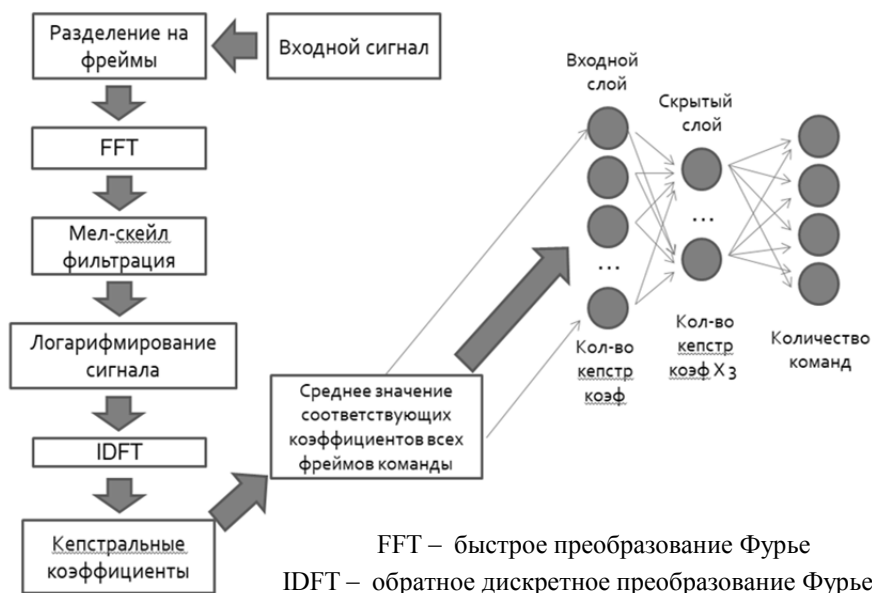


Рис. 2. Структура нейронной сети при кепстральном анализе

4. ОБУЧЕНИЕ НЕЙРОННОЙ СЕТИ ДЛЯ ПРЯМОГО АНАЛИЗА

В данной работе были получены две разные по структуре нейронные сети. У них была разная предварительная обработка данных, однако обучены они по одинаковому принципу, для того чтобы сравнить результаты обучения. Процесс обучения нейронной сети сводится к следующему: на вход подается обработанный голосовой сигнал, на выходном слое в том нейроне, к которому относится эта голосовая команда, стоит единица, на остальные нейроны подается ноль. Так как нейронную сеть можно обучить лишь пошагово (за один шаг один выход), то общий шаг обучения делится на подшаги в количестве нейронов выходного слоя. Обучение нейронной сети было проведено для трех выходных нейронов и соответственно трех голосовых команд. Алгоритм следующий.

- Выбирается количество эпох обучения. В нашем случае было выбрано 20.
- Загружаются сохраненные данные нейронной сети, если они есть.
- Нейронная сеть последовательно обучается каждому из трех слов и каждому из трех вариантов слов (сказанных разными людьми).

- Внутри функции обучения, где каждой голосовой команде соответствует выходной нейрон, вычисляется ошибка.
- После прохождения эпохи обучения (три слова, три вариации слова и три этапа в функции обучения) формируется ошибка сети за эпоху обучения.

5. РЕЗУЛЬТАТЫ

Как видно из диаграммы (рис. 3), I подход позволяет создать НС, имеющую преимущество – меньшее итоговое значение ошибки.

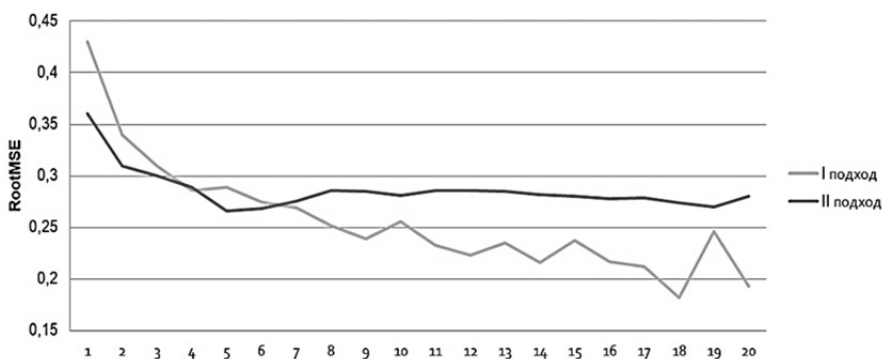


Рис. 3. Зависимость ошибки от количества пройденных эпох обучения

Преимуществом НС, реализованной II подходом, является значительно более высокая скорость обучения.

ЗАКЛЮЧЕНИЕ

По результатам этой работы видно, что ИНС не очень хорошо подходит для сортировки массива чисел. Это было заранее известно, так как сортировки обычно используют четкие, известные и явно реализуемые алгоритмы, которые дают точный результат. В случае с ИНС мы получаем массив, примерно сортированный по возрастанию с редкими значительными отклонениями от полностью отсортированного массива. Точность, достигнутая для данной ИНС, равна 4 % (40 ошибок из 1000 экспериментов) для однослойной ИНС за 24 часа непоследовательного обучения, для многослойной нейронной сети точность достигла 15 % (150 ошибок из 1000) за 2 часа последовательного обучения.

БЛАГОДАРНОСТИ

Авторы выражают свою искреннюю благодарность профессору кафедры автоматики А.А. Воеводе и доценту кафедры вычислительной техники А.А. Малявко за помощь при выполнении работы, а также полезное обсуждение полученных результатов.

СПИСОК ЛИТЕРАТУРЫ

1. *Осовский С.* Нейронные сети для обработки информации / пер. с пол. И.Д. Рудинского. – М.: Финансы и статистика, 2002. – 344 с.
2. *Галушкин А.И.* Нейронные сети: основы теории. – М.: Горячая Линия-Телеком, 2010. – 496 с.
3. Модель бионической нейронной сети и ее применения [Электронный ресурс] / С.С. Ёлкин, С.В. Ёлкин, Э.С. Клышинский, В.Ю. Максимов, Т.Н. Мусаева. – М.: ИПМ им. М.В. Келдыша РАН, 2008. – URL: http://keldysh.ru/papers/2008/prep89/prep2008_89.html (дата обращения: 10.12.2017).
4. *Калан Р.* Основные концепции нейронных сетей. – М.: Вильямс, 2001. – 291 с.
5. *Воевода А.А., Марков А.В., Романников Д.О.* Разработка программного обеспечения: проектирование с использованием UML диаграмм и сетей Петри на примере АСУ ТП водонапорной станции // Труды СПИИРАН. – 2014. – Вып. 3 (34). – С. 218–232.
6. *Марков А.В.* Свойства инверсии сетей Петри // Сборник научных трудов НГТУ. – 2014. – № 4 (78). – С. 139–152.
7. Al-Shayea Q.K. Artificial neural networks in medical diagnosis // International Journal of Computer Science Issues. – 2011. – Vol. 8, iss. 2. – P. 130–154.
8. *Capannini G., Silvestri F., Baraglia R.* Sorting on GPUs for large scale datasets: a thorough comparison // Information Processing and Management. – 2011. – Vol. 48 (5). – P. 903–917.
9. *Барский А.Б.* Нейронные сети: распознавание, управление, принятие решений. – М.: Финансы и статистика, 2004. – 176 с.
10. *Хайкин С.* Нейронные сети. Полный курс. – 2-е изд. – М.: Вильямс, 2006. – 1104 с.
11. *Медведев В.С., Потемкин В.Г.* Нейронные сети. – М.: Диалог-МИФИ, 2002. – 496 с.
12. *Яхьяева Г.Э.* Нечеткие множества и нейронные сети. – М.: Бином, 2006. – 374 с.

13. *Комашинский В.И., Смирнов Д.А.* Нейронные сети и их применение в системах управления связи. – М.: Горячая Линия-Телеком, 2003. – 94 с.
14. *Загуменнов А.В.* Компьютерная обработка звука. – М.: ДМК Пресс, 2011. – 580 с.
15. *Лоянич А.Т.* Запись и обработка звука на компьютере. – М.: Эксмо, 2008. – 390 с. – (Просто как 2х2).
16. *Воевода А.А., Романников Д.О.* Асинхронный алгоритм сортировки массива чисел с использованием ингибиторных сетей Петри // Труды СПИИРАН. – 2016. – Вып. 5 (28). – С. 198–213.
17. *Воевода А.А., Полубинский В.Л., Романников Д.О.* Сортировка массива целых чисел с использованием нейронной сети // Научный вестник НГТУ. – 2016. – № 2 (63). – С. 151–157.

Черданцев Константин Артурович, студент факультета автоматизации и вычислительной техники Новосибирского государственного технического университета. Имеет одну публикацию. E-mail: medmene@yandex.ru.

Кладко Антон Владимирович, студент факультета автоматизации и вычислительной техники Новосибирского государственного технического университета. E-mail: tonkladko@ya.ru.

Analysis of the sound spectrum of voice commands with the neuron network*

К.А. Cherdantsev¹, А.В. Kladko²

¹ *Novosibirsk State Technical University, 20 Karl Marks Avenue, Novosibirsk, 630073, Russian Federation, student of the automation department. E-mail: medmene@gmail.ru*

² *Novosibirsk State Technical University, 20 Karl Marks Avenue, Novosibirsk, 630073, Russian Federation, student of the automation department. E-mail: tonkladko@ya.ru*

The task of voice analysis has become popular with the widespread use of computers and their peripherals. By itself, the task of analyzing human speech allows us to discover new areas of statistics, psychology, etc., in which, with the help of computer calculations, it is possible to analyze the speech of real people, in such volumes, in which it is not possible for an ordinary

* Received 07 August 2017.

person or even groups of people. A word or phrase spoken by a person in reality, in computer representation is described by a huge number of parameters and therefore only recently computing power of computers allows real-time processing of video and audio files. In this connection, such systems as "smart house" began to appear, based on the considerable simplification of the interaction of the machine and the person. In this system, the main control is the voice or even the sound commands of a person, be it words or say cotton. In addition to this system, voice analysis on the Internet is widely used. There, when searching for various kinds of information, you can use voice search, for example, when there is no possibility to type a text request, and it is very convenient, especially in the era of smartphones. In the future, with the technologies of highly developed speech recognition and analysis, it is possible to reproduce the real mechanism of "communication" between a person and a machine, which will allow developing and training the machine as a human being.

Keywords: Neural networks, algorithms, experiment, neural networks, speech recognition, cepstral analysis, multilayer neural network, evaluation algorithms, programming

DOI: 10.17212/2307-6879-2017-4-65-76

REFERENCES

1. Osowski S. *Sieci neuronowe do przetwarzania informacji* [Neural network for analysing information]. Warszawa, Oficyna wydawnicza politechniki warszawskiej, 2000 (Russ. ed.: Osovskii S. *Neironnyye seti dlya obrabotki informatsii*. Translated from Polish. Moscow, Finansy i statistika Publ., 2002. 344 p.).
2. Galushkin A.E. *Neironnyye seti: osnovy teorii* [Neural network. Base theory]. Moscow, Hotline-Telecom Publ., 2010. 496 p.
3. Yolkin S.S., Yolkin S.V., Klishinskiy E.S., Maksimov V.Y., Musaeva T.N. *Model' bionicheskoi neironnoi seti i ee primeneniya* [The bionic neural network model and its applications]. Moscow, Keldysh Institute of Applied Mathematics RAS, 2008. Available at: http://keldysh.ru/papers/2008/prep89/prep2008_89.html (accessed 10.12.2017).
4. Callan R. *The Essence of neural networks*. London, Prentice Hall, 1999 (Russ. ed.: Kalan R. *Osnovnye kontseptsii neironnykh setei*. Moscow, Williams Publ., 2001. 291 p.).
5. Voevoda A.A., Markov A.V., Romannikov D.O. Razrabotka programmnogo obespecheniya: proektirovanie s ispol'zovaniem UML diagramm i setei Petri na primere ASU TP vodonapornoj stantsii [Software development: software design using uml diagrams and petri nets for example automated process control system of pumping station]. *Trudy SPIIRAN – SPIIRAS Proceedings*, 2014, iss. 3 (34), pp. 218–232.

6. Markov A.V. Svoistva inversii setei Petri [Properties inversion Petri nets]. *Sbornik nauchnykh trudov Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Transaction of scientific papers of the Novosibirsk state technical university*, 2014, no. 4 (78), pp. 139–152.
7. Al-Shayea Q.K. Artifical neural networks in medical diagnosis. *International Journal of Computer Science Issues*, 2011, vol. 8, iss. 2, pp. 130–154.
8. Capannini G., Silvestri F., Baraglia R. Sorting on GPUs for large scale datasets: a thorough comparison. *Information Processing and Management*, 2011, vol. 48 (5), pp. 903–917.
9. Barskii A.B. *Neironnye seti: raspoznavanie, upravlenie, prinyatie reshenii* [Neural network: recognition, control, accept solutions]. Moscow, Finansy i statistika Publ., 2004. 176 p.
10. Haykin S. *Neural networks*. Upple Saddle River, Prentice Hall, 1999 (Russ. ed.: Khaikin S. *Neironnye seti. Polnyi kurs*. 2nd ed. Moscow, Williams Publ., 2006. 1104 p.).
11. Medvedev V.S., Potemkin V.G. *Neironnye seti* [Neural network]. Moscow, Dialog – MIFI Publ., 2002. 496 p.
12. Yakh"yaeva G.E. *Nechetkie mnozhestva i neironnye seti* [Fuzzy sets and neural networks]. Moscow, Binom Publ., 2006. 374 p.
13. Komashinskii V.I., Smirnov D.A. *Neironnye seti i ikh primeneniye v sistemakh upravleniya svyazi* [Neural networks. Using in communications system]. Moscow, Hotline-Telecom Publ., 2003. 94 p.
14. Zagumennov A.V. *Komp'yuternaya obrabotka zvuka* [Analyzing sounds by computer]. Moscow, DMK Press Publ., 2011. 580 p.
15. Loyanich A.T. *Zapis' i obrabotka zvuka na komp'yutere* [Write and analyzing sounds]. Moscow, Eksmo Publ., 2008. 390 p.
16. Voevoda A.A., Romannikov D.O. Asinkhronnyi algoritm sortirovki massiva chisel s ispol'zovaniem ingibitornykh setei Petri [Asynchronous sorting algorithm for array of numbers with the use of inhibitory Petri nets]. *Trudy SPIIRAN – SPIIRAS Proceedings*, 2014, iss. 3 (34), pp. 218–232.
17. Voevoda A.A., Polubinskii V.L., Romannikov D.O. Sortirovka massiva tselykh chisel s ispol'zovaniem neironnoi seti [Sorting the array of integers using a neural network]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Science bulletin of the Novosibirsk state technical university*, 2016, no. 2 (63), pp. 151–157.

Для цитирования:

Черданцев К.А., Кладько А.В. Анализ звукового спектра голосовых команд с помощью нейронной сети // Сборник научных трудов НГТУ. – 2017. – № 4 (90). – С. 65–76.

For citation:

Cherdantsev K.A., Kladko A.V. Analiz zvukovogo spektra golosovykh komand s pomoshch'yu neironnoi seti [Analysis of the sound spectrum of voice commands with the neuron network]. *Sbornik nauchnykh trudov Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta* – *Transaction of scientific papers of the Novosibirsk state technical university*, 2017, no. 4 (90), pp. 65–76.