

Распознавание речевых сообщений радиобмена методами глубокого обучения с предварительной фильтрацией

Н.А. Андриянов

Финансовый университет при Правительстве Российской Федерации
Москва, Россия
naandriyanov@fa.ru

Д.А. Андриянов

Ульяновский государственный технический университет
Ульяновск, Россия
boss.renome@mail.ru

Аннотация—В статье рассмотрена задача распознавания речевых сообщений радиобмена в условиях действия помеховой обстановки. Сравнение выполняется для рекуррентных нейронных сетей и для сверточных нейронных сетей при обработке автокорреляционных портретов. Применение сверточных нейронных сетей обеспечивает повышение эффективности распознавания на 2-3%, а в условиях фильтрации повышает качество на 10-15% при разном уровне шума. Дополнительная предварительная фильтрация на базе дважды стохастического фильтра улучшает результаты.

Ключевые слова— сверточные нейронные сети, распознавание речи, автокорреляционный портрет, рекуррентные сети, дважды стохастический фильтр

1. ВВЕДЕНИЕ

Задача распознавания речи является одной из самых распространенных задач в глубоком обучении. При этом для решения могут быть использованы как корреляционные алгоритмы [1, 2], так и методы на основе глубокого обучения [3, 4]. Однако, помимо того, что само по себе распознавание речевых команд может выполняться не с максимальной эффективностью, дополнительные сложности создает воздействие шума при передаче по каналу связи.

Зачастую при глубоком обучении в распознавании речи рассматривают сигналы, не подвергшиеся влиянию помех. Однако такой подход справедлив при рассматриваемых задачах близости источника и приемника информации, когда влиянием канала связи можно пренебречь.

В гражданской авиации диспетчерские пункты управления воздушным движением установлены на Земле, в то время как прием осуществляется в воздухе. В связи с этим рассмотрение моделей распознавания речевых сообщений имеет смысл делать в условиях помеховой обстановки. А качественный и оперативный обмен информацией между пилотом и диспетчером – один из признаков безопасности полета.

2. МОДЕЛЬ КАНАЛА СВЯЗИ И ПРЕДВАРИТЕЛЬНАЯ ФИЛЬТРАЦИЯ

Пусть речевой сигнал представлен последовательностью дискретных отсчетов $s(i)$. Рассмотрим модель, когда при передаче сигнал подвергается воздействию белого гауссовского шума $n(i)$. Таким образом, наблюдения примут вид:

$$z(i) = x(i) + n(i), \quad (1)$$

где шум $n(i)$ обладает нулевым средним и дисперсией Dn .

Можно представить исходный речевой сигнал в виде модели с изменяющимися параметрами, так что

$$x(i) = r(i)x(i-1) + e(i), \quad (2)$$

где $r(i)$ – последовательность коэффициентов корреляции, $e(i)$ – случайная добавка с переменной дисперсией и средним значением.

Оценка параметров сигнала может быть выполнена в скользящем окне, либо с использованием дважды стохастического фильтра [5]. На основе тестовой фразы «Предполетная проверка связи» была выполнена оценка параметров модели. При отношении сигнал/шум (дисперсия речевого сигнала к дисперсии шума) $q = 10$ выполнено моделирование передачи сообщения $N = 10000$ раз. Обработка производилась дважды стохастическим фильтром.

В таблице 1 представлены сравнительные результаты по относительной дисперсии ошибки фильтрации для дважды стохастического и авторегрессионного фильтра Калмана.

Таблица 1. Эффективность фильтрации

Фильтр	Дисперсия ошибки
Фильтр Калмана	0,322
Дважды стохастический фильтр	0,164

Видно, что дважды стохастическая модель обеспечивает повышение эффективности в 2 раза. Это связано со сложной природой связей в речевом сообщении.

3. ДАННЫЕ И МОДЕЛИ РАСПОЗНАВАНИЯ

Все записи производились одним спикером, мужчиной. После этого выполнялась их обработка в бесплатной программе Audacity с возможностью частотного анализа. Временные отсчеты импортировались в программу на языке Python для дальнейшей обработки. На рис. 1 представлена запись речевого сигнала «Предполетная проверка связи» и его спектр. На рис. 2 представлена запись речевого сигнала «К взлету готов» и его спектр.



Рис. 1. Пример первого речевого сообщения и его спектра

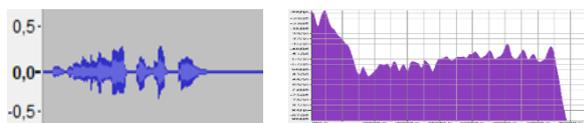


Рис. 2. Пример второго речевого сообщения и его спектра

Затем временные представления с дискретизацией на 1024 отсчета смешивались с белым гауссовским шумом с различными дисперсиями.

Для использования сверточных нейронных сетей, распространенных в распознавании изображений [6], было реализовано построение корреляционных портретов. Была написана специальная функция для получения корреляционных портретов по дискретным отсчетам сигнала

$$R(i, k) = \frac{1}{N} \sum_{i=0}^{N-1} \frac{1}{M * \sigma^2(i)} \sum_{k=0}^{M-1} [S(i * M) - m(i)] * [S(i * M + k) - m(i)] \quad (3)$$

где N – количество строк в корреляционном портрете, M – количество столбцов в корреляционном портрете, $\sigma(i)$ – дисперсия речевого сообщения в i -й строке, $m(i)$ – среднее значение речевого сообщения в i -й строке.

На рис. 3 представлен корреляционный портрет размером 32 на 32 для фразы «Предполетная проверка связи».



Рис. 3. Автокорреляционный портрет речевого сигнала

Отметим, что в случае обработки зашумленных сигналов, развертка выполнялась для речевого сообщения после фильтрации.

Для классификации было использовано десять речевых фраз: «Предполетная проверка связи», «Принимаю вас на пять», «Информация Элфа на борту», «Разрешите предварительный», «К взлету готов», «Взлет разрешаю», «Контроль вторичный», «Захват курсового», «Посадку разрешили», «Разрешите запуск на стоянке». Каждая фраза была записана в количестве 100 примеров, поэтому для оценки качества оправдано использовать метрику Ассигасу. Обучение и тестирование выполнялось на основе разделения выборки в пропорции 80% на 20%. Результаты работы различных алгоритмов при различных отношениях сигнал/шум представлены в таблице 2.

Таблица II. ЭФФЕКТИВНОСТЬ РАСПОЗНАВАНИЯ

Модель	Accuracy (нет шума)	Accuracy (q = 1)	Accuracy (q = 0,1)
Рекуррентная сеть	0,96	0,81	0,64
Сверточная сеть	0,98	0,52	0,42
Рекуррентная сеть с фильтром	0,96	0,88	0,75
Сверточная сеть с фильтром	0,98	0,91	0,78

Из представленной таблицы видно, что фильтрация позволяет существенно повысить качество дальнейшей обработки. При этом при шуме рекуррентные сети сохраняют большую устойчивость, так как способны запоминать формы сигналов. В тоже время методы на базе сверточных нейронных сетей требуют нормализации данных, но большие шумы даже при нормализации приводят к потерям.

4. ЗАКЛЮЧЕНИЕ

Модели CNN превосходят модели RNN на 2% на незашумленных данных и на 3% на зашумленных данных с применением алгоритма фильтрации. Однако в присутствии помех прирост эффективности CNN в результате фильтрации может достигать 35-40%, в то время как для RNN всего лишь порядка 10%. Повысить устойчивость CNN к шуму в будущем можно с помощью дополнительных обучающих примеров.

БЛАГОДАРНОСТИ

Исследование выполнено при поддержке РФФИ, Проект №19-29-09048.

ЛИТЕРАТУРА

- [1] Андриянов, Н.А. Распознавание речевых сообщений радиообмена в авиации на базе корреляционного анализа / Н.А. Андриянов // Известия Самарского научного центра Российской академии наук. – 2021. –Т.23, №1. – С. 91 – 96.
- [2] Армер, А.И. Распознавание речевых команд на фоне интенсивных акустических шумов по кросскорреляционным портретам / А.И. Армер, Е.Ю. Галицкая, Н.А. Крашенинникова // Сборник трудов конференции ИТНТ. – 2017. – С. 840 - 845.
- [3] Mukhamadiyev, A. Automatic Speech Recognition Method Based on Deep Learning Approaches for Uzbek Language / A. Mukhamadiyev, I. Khujayarov, O. Djuraev, J. Cho // Sensors. – 2022. – Vol. 22. – P. 3683. DOI: 10.3390/s22103683
- [4] Andriyanov, N. Intelligent Processing of Voice Messages in Civil Aviation: Message Recognition and the Emotional State of the Speaker Analysis / N. Andriyanov, D. Andriyanov // 2021 International Siberian Conference on Control and Communications (SIBCON). – 2021. – P. 1-5. DOI: 10.1109/SIBCON50419.2021.9438881.
- [5] Васильев, К.К. Дважды стохастическая фильтрация пространственно неоднородных изображений / К.К. Васильев, В.Е. Дементьев // Радиотехника и электроника. – 2020. – Т. 65, №5. – С. 487-494.
- [6] Андриянов, Н.А. Обнаружение объектов на изображении: от критериев Байеса и Неймана–Пирсона к детекторам на базе нейронных сетей EfficientDet / Н.А. Андриянов, В.Е. Дементьев, А.Г. Ташлинский // Компьютерная оптика. – 2022. – Т. 46, № 1. – С. 139-159. – DOI: 10.18287/2412-6179-CO-922.