

поведение частоты основного тона. Значение -1 соответствует понижению частоты основного тона при озвучивании соответствующего символа, 0 соответствует отсутствию изменения частоты основного тона, а значение 1 соответствует повышению основного тона при озвучивании соответствующих символов.

В качестве базовой модели искусственной нейронной сети для синтеза речи мы использовали модель Tacotron 2 [3]. Её модификация состоит в том, что кроме признаков, представляющих собой синтезируемый текст, ей на отдельный вход подаются так же признаки, соответствующие изменен частоте основного тона. Предложенная архитектура модели синтеза спектрограммы представлена на рисунке (рисунок 1).

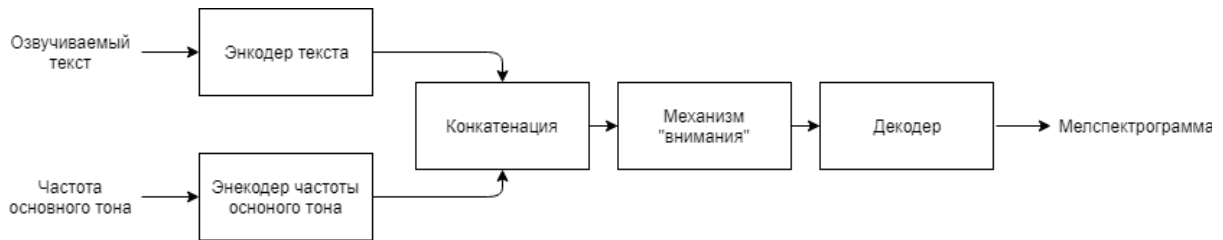


Рисунок 1 – Подпись рисунка отделяется от следующего абзаца пробельной строкой

В качестве вокодера для преобразования мелспектрограммы в аудио использовалась модель FloWaveNet [4], преимуществом которой перед моделью Parallel WaveNet (которая использовалась в [3]), является простота обучения.

Для обучения модели использовался набор данных, объемом около 25 часов русской речи, записанной профессиональным диктором.

Таким образом, основными преимуществами предложенной системы синтеза речи являются: возможность модулирования интонации на уровне отдельных слов, или даже словосочетаний, эффективность обучения вокодера, при примерно одинаковом качестве синтезированной речи.

Список использованных источников:

1. Style Tokens: Unsupervised Style Modeling, Control and Transfer in End-to-End Speech Synthesis / Y. Wang [et al.] // arXiv preprint, arXiv:1803.09017v1. – 2018. – 11p.
2. Towards End-to-End Prosody Transfer for Expressive Speech Synthesis with Tacotron / RJ Skerry-Ryan [et al.] // arXiv preprint, arXiv: 1803.09047. – 2018. – 11p.
3. Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions / J. Shen [et al.] // arXiv preprint, arXiv1712.05884v2. – 2018. – 5p.
4. FloWaveNet : A Generative Flow for Raw Audio / K. Sungwon [et al.] // arXiv preprint, arXiv1811.02155v2. – 2018. – 6p.

АНАЛИЗ КЛАССИФИКАТОРОВ ДЛЯ ДИАГНОСТИРОВАНИЯ НЕВРОЛОГИЧЕСКОГО ЗАБОЛЕВАНИЯ БАС

Гвоздович А.Д.

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Вашкевич М.И. – к.т.н., доцент

Подход к диагностированию БАС на основе анализа речи является одним из перспективных направлений в детектировании данного заболевания, не имеющего характерных биомаркеров. На основе выделенных из речевого сигнала признаков необходимо произвести классификацию на здоровых и больных. В данной работе рассматриваются три различных алгоритма классификации и сравниваются между собой.

Боковой амиотрофический склероз (БАС) является прогрессирующим нейродегенеративным заболеванием, поражающим нервные клетки головного и спинного мозга. Медиана выживаемости при БАС составляет 3 года.

Отсутствие характерных биомаркеров БАС приводит к задержке в постановке правильного диагноза. В последнее время набирает популярность подход к детектированию заболевания основанный на анализе речи пациента. Согласно идее, предложенной в [1], для обнаружения характерных речевых нарушений может быть использована только часть речевого теста, содержащая гласные. В [2] сделан анализ и выбор подходящих гласных. Было показано, что гласные /и/ и /а/

являются наиболее подходящим выбором для обнаружения дизартрии БАС, поскольку их произнесение требует значительных усилий мышц языка.

В предыдущей работе [1] были получены информационные признаки, выделяемые из речевой записи, содержащей счет чисел от 1 до 3. К наиболее удачным признакам относятся: $d1$ – мера схожести спектральных огибающих звука /a/ и /u/, $F1a$, $F2a$, $F1i$, $F2i$ – первые и вторые форманты звуков /a/ и /u/, $F1conv$ и $F2conv$ – абсолютное значение разности между первыми и вторыми формантами звуков /a/ и /u/, соответственно.

Целью данной работы было, опираясь на данный набор признаков, выбрать один из следующих трех методов классификации:

- линейный дискриминантный анализ (LDA);
- квадратичный дискриминантный анализ (QDA);
- алгоритм кластеризации K-средних.

Линейный дискриминантный анализ вычисляет «дискриминантные баллы» для каждого наблюдения, чтобы классифицировать, к какому классу относится входной вектор. Эти оценки получены путем нахождения линейных комбинаций независимых переменных. Для одной переменной предиктора $X = x$ классификатор LDA оценивается как

$$\hat{\delta}_k(x) = x \cdot \frac{\hat{\mu}_k}{\hat{\sigma}^2} - \frac{\hat{\mu}_k^2}{2\hat{\sigma}^2} + \log \log \left(\hat{\pi}_k \right), \quad (1)$$

где $\hat{\delta}_k(x)$ является оценкой дискриминанта означающая, что наблюдение попадет в k -й класс в переменную ответа на основе значения переменной-предиктора x ; $\hat{\mu}_k$ – среднее значение всех тренировочных наблюдений из k -го класса; $\hat{\sigma}^2$ – средневзвешенное значение выборочных дисперсий для каждого из K классов; $\hat{\pi}_k$ – априорная вероятность того, что наблюдение относится к k -му классу

Этот классификатор присваивает наблюдение k -му классу Y_k , для которого оценка дискриминанта является наибольшей.

При работе с более чем одной переменной-предиктором классификатор LDA предполагает, что наблюдения в k -м классе взяты из многомерного гауссовского распределения $N(\mu_k, \Sigma)$, где μ_k – это средний вектор для класса, а Σ – ковариационная матрица. это общее для всех K классов. Включение этого в классификатор LDA приводит к

$$\hat{\delta}_k(x) = x^T \Sigma^{-1} \hat{\mu}_k - \frac{1}{2} \hat{\mu}_k^T \Sigma^{-1} \hat{\mu}_k + \log \log \left(\hat{\pi}_k \right), \quad (2)$$

где наблюдение будет присвоено классу k , когда оценка дискриминанта $\hat{\delta}_k(x)$ является наибольшей.

Квадратичный дискриминантный анализ обеспечивает альтернативный подход. Как и LDA, классификатор QDA предполагает, что наблюдения из каждого класса Y взяты из гауссовского распределения. Однако, в отличие от LDA, QDA предполагает, что у каждого класса есть своя ковариационная матрица. Другими словами, предполагается, что переменные-предикторы не имеют общей дисперсии на каждом из k уровней в Y . Математически предполагается, что наблюдение из k -го класса имеет вид $X \sim N(\mu_k, \Sigma_k)$, где Σ_k есть ковариационная матрица для k -го класса. При этом предположении классификатор назначает наблюдение классу, для которого (3) самый большой.

$$\hat{\delta}_k(x) = -\frac{1}{2} x^T \Sigma_k^{-1} x + x^T \Sigma_k^{-1} \hat{\mu}_k - \frac{1}{2} \hat{\mu}_k^T \Sigma_k^{-1} \hat{\mu}_k - \frac{1}{2} \log |\Sigma_k| + \log \log \left(\hat{\pi}_k \right). \quad (3)$$

Алгоритм **к-средних** делит набор из N выборок X на K непересекающихся кластеров C , каждый из которых описывается средним значением μ_j выборок в кластере. Средства обычно называют кластерными «центроидами»; необходимо обратить внимание, что они, как правило, не являются точками из X , хотя они находятся в одном и том же пространстве. Алгоритм K-средних предназначен для выбора центроидов, которые минимизируют внутрикластерную сумму квадрата критерия:

$$\hat{\mu}_k \sum_{i=0}^n (\|x_i - \mu_j\|^2)$$

Для определения точности предлагаемого БАС-детектора был использован метод перекрестной проверки по k блокам [3], где k равно 4. Вначале тренировочный набор был случайно перемешан, затем процедура обучения была выполнена на основе 75%. По данным, остальные 25% были использованы для тестирования. Эту процедуру повторяли 10000 раз, затем рассчитывали среднее значение и стандартное отклонение для параметров точности.

Полученные характеристики для детектора БАС сравнивались по таким параметрам, как общая точность (ACC), чувствительность (Sens), положительное прогностическое значение (PPV). Результаты классификации при помощи данных методов приведены в таблице 1.

Таблица 1. Результаты классификации

Характеристики	Acc / Sens / PPV, %		
	LDA	QDA	К средних
d1	78.9 / 69.6 / 75.5	78.6 / 70.6 / 74.2	70.1 / 79.7 / 55.2
F1a	65.9 / 0.4 / 0.8	65.9 / 0.5 / 1.3	56.1 / 57.6 / 38.3
F2a	64.8 / 1.0 / 4.0	64.8 / 0.7 / 2.9	58.1 / 57.0 / 39.9
F1i	65.9 / 0.7 / 4.1	65.6 / 0.6 / 2.1	56.0 / 61.6 / 38.7
F2i	82.2 / 94.2 / 72.7	81.2 / 90.2 / 72.0	68.6 / 99.4 / 52.6
F1conv	65.9 / 0.8 / 1.7	65.9 / 0.4 / 1.0	56.7 / 50.6 / 37.2
F2conv	80.8 / 82.5 / 74.0	80.7 / 81.5 / 74.2	77.6 / 100.0 / 63.4
d1 и F2i	81.9 / 79.7 / 78.3	79.4 / 81.6 / 71.0	68.6 / 99.3 / 52.6
d1 и F2conv	84.8 / 88.1 / 81.3	82.6 / 82.5 / 78.7	77.6 / 100.0 / 63.3
F2i и F2conv	81.9 / 91.3 / 73.1	80.6 / 84.4 / 72.6	72.4 / 99.9 / 56.7
d1, F2i и F2conv	83.3 / 95.5 / 74.5	79.9 / 81.2 / 72.5	72.3 / 99.9 / 56.5
Все характеристики	80.5 / 83.9 / 72.7	64.2 / 90.0 / 48.7	72.8 / 99.8 / 57.0

Исходя из полученных данных можно сделать вывод, что классификатор на основе LDA обладает наилучшей точностью и положительной прогностической значимостью. Метод к средних несмотря на то, что позволяет получить хорошую специфичность имеет недостаточную точность. Небольшие отличия в точности между методами LDA и QDA свидетельствуют о том, что в данной задаче более высокую важность имеют информативные признаки, чем непосредственно метод классификации. Дальнейшая работа должна быть направлена на выработку дополнительных информационных признаков.

Список использованных источников:

1. A. Gvozdoch, M. Vashkevich, Yu. Rushkevich, A. Petrovsky Detection Bulbar Dysfunction in ALS Patients Using Acoustic Analysis of Vowels Extracted from Continuous Speech / accepted to conf. Pattern Recognition and Information Processing (PRIP-2019).
2. M. Vashkevich, E. Azarov, A. Petrovsky and Y. Rushkevich. Features extraction for the automatic detection of ALS disease from acoustic speech signals. Proceedings of the Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA'2018), Poznan, Poland 19-21 Sept. 2018, pp. 321-326.
3. R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proc. of International Joint Conference on Artificial Intelligence, Montreal, Canada 20-25 Aug., 1995, pp. 1137-1143.

СИСТЕМА ОБНАРУЖЕНИЯ ПЕРЕДНЕГО ПЛАНА НА ЕСТЕСТВЕННЫХ ИЗОБРАЖЕНИЯХ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННОЙ СЕТИ

Герасимович Н.Ю.

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Петровский Н.А. – к.т.н., доцент

Рассмотрен один из методов выделения переднего плана на естественных изображениях с использованием нейронной сети Кохонена. Нейронная сеть Кохонена широко используется в задачах: кластеризации данных, прогнозирование свойств, уменьшение размерности данных с минимальной потерей информации.