

СИСТЕМА ДИАГНОСТИРОВАНИЯ ПАТОЛОГИИ ГОЛОСОВЫХ СКЛАДКОВ НА ОСНОВЕ АКУСТИЧЕСКОГО АНАЛИЗА ГОЛОСОВОГО СИГНАЛА И МЕТОДА ОПОРНЫХ ВЕКТОРОВ

Коледа Е.О., студент гр.750701

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

Вашкевич М.И. – канд. тех. наук, доцент

Аннотация. В работе рассмотрена задача классификации патологии голосовых складок на основе акустического анализа голосового сигнала и метода опорных векторов с целью выявления патологии голосовых складок. В качестве признаков использовались мел частотные кепстральные коэффициенты, которые рассчитывались в ходе анализа и обработки голосового сигнала. Для тестирования метода использовалась база голосов, содержащая записи пациентов с дисфонией, парезами и людей с нормальным состоянием голосовых складок. В результате достигнута правильность около 94 процента.

Ключевые слова. Метод опорных векторов, патологии голоса, парезы, дисфония, мел частотные кепстральные коэффициенты, один против всех, акустический сигнал, голосовые складки.

Введение. В ходе работы ставилась задача построить систему, которая будет диагностировать патологии голосовых складок. Работа системы основывается на акустическом анализе сигнала и методе опорных векторов. Данная задача актуальна, так как современные системы диагностирования этих патологий не совершенны. Задача усложняется тем, что нам необходимо провести не просто бинарную классификацию определив есть ли во входном сигнале патологии или он подходит под критерии нормального, а создать систему, которая будет относить голос к одному из классов: «норма», «дисфония» или «парез».

Метод опорный векторов. В системе используется метод опорных векторов (МОВ) (англ. SVM, support vector machine) — алгоритм обучения с учителем, который часто используется для задач классификации и регрессионного анализа. Так, например, в работе [1], его использовали для классификации неврологических, функциональных и ЛОР патологий. Разработчики в статье описали извлечение различных признаков из записей голосов пациентов на разных языках. Далее полученные признаки использовались для классификации на основе МОВ. В результате использования данного подхода получена высокая правильность классификации. Одни из самых лучших результатов дало использование МЧКК и характеристик шума в качестве входных признаков для классификатора.

МОВ принадлежит к семейству линейных классификаторов. Особым свойством МОВ является непрерывное уменьшение эмпирической ошибки классификации и увеличение зазора, расстояния между опорными векторами двух классов, поэтому метод также известен как метод классификатора с максимальным зазором [2]. Целевая функция МОВ имеет следующий вид:

$$\mathbf{w}^*, t^*, \xi_i^* = \operatorname{argmin}_{\mathbf{w}, t, \xi_i} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i, \quad (1)$$

где \mathbf{w} – вектор, описывающий нормаль к границе решения; ξ_i – ослабляющие переменные, благодаря которым некоторые примеры могут оказаться внутри зазора или даже с неправильной стороны от решающей границы, C – регуляризирующий параметр, определяющий компромисс между максимизацией зазора и минимизацией ослабляющей переменной. Если мы допускаем больше ошибок зазора, то понадобится меньше опорных векторов, поэтому параметр C в некотором смысле управляет «сложностью» МОВ [3].

Основной задачей алгоритма является найти наиболее правильную линию, или гиперплоскость, разделяющую данные на два класса в пространстве признаков [4]. График функции представлен на рисунке 1.

МОВ использует математическую функцию ядра, чтобы преобразовывать входные признаки в требуемую форму. Различные варианты МОВ используют различные типы функций ядра. Например, линейная, полиномиальная, радиальная базисная функция, сигмоидная и проч. В данной работе использовалась радиальная базисная функция:

$$K(x, y) = e^{-\gamma \|x-y\|^2}, \quad (2)$$

где x, y – характеристические вектора из обучающей выборки, γ – настроечный параметр.

Для создания классификатора использовалась библиотека Scikit-learn[5] для Python. Для получения мультиклассового классификатора применялась стратегия «Один против всех». Согласно данному подходу необходимо обучить N классификаторов, где N – количество классов. Классификатор с самым высоким значением функции выхода присваивает новый объект к определенному классу. Для того, чтобы использовать данную стратегию в нашем классификаторе, необходимо задать параметру `decision_function_shape` значение `ovr(one vs rest)`. Изменяя параметры

γ и C для классификатора, мы влияем на результат работы классификатора. Эти коэффициенты были подобраны с помощью метода «Поиск по сетке» (англ. grid search) и равны $\gamma = 0.001$, $C = 10$.

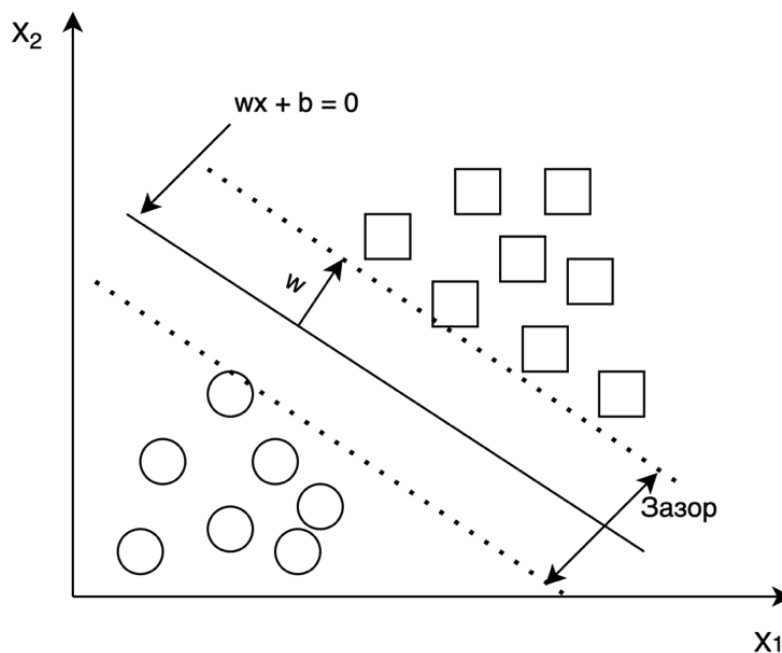


Рисунок 1 – Демонстрация алгоритма SVM для линейного случая, бинарная классификация.

МЧКК. В качестве признаков для данного классификатора были выбраны мел частотные кепстральные коэффициенты (МЧКК)[6]. Алгоритм нахождения признаков из входного сигнала:

1. Разбиение сигнала на отрезки длиной 800мс с перекрытием 400мс, отрезки меньшей длительности мы будем пропускать. Процесс разбиения изображен на рисунке 2.
2. Вычисление МЧКК для каждого 20 мс фрейма.
3. Нахождение среднего арифметического значения коэффициентов.
4. Запись коэффициентов в бинарный файл для дальнейшего обучения классификатора.

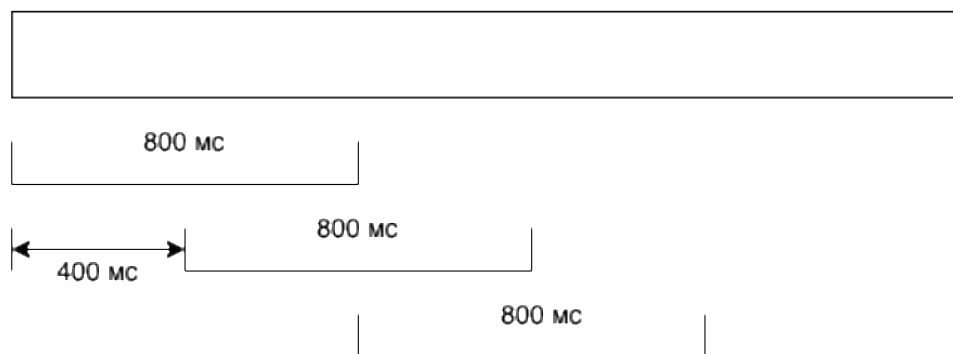


Рисунок 2 – Разбиение сигнала на отрезки по 800мс.

Для вычисления МЧКК была использована библиотека librosa[7] для python. Данная библиотека содержит множество методов. Все эти методы принимают различные параметры для настройки работы, но мы изменим лишь некоторые, так как остальные стандартные параметры соответствуют нашей системе.

Измененные параметры:

1. Частота дискретизации, стандартная величина 22500, но мы будем использовать 44100 Гц.
2. Window, стандартно библиотека использует окно Ханна, но мы изменим его на окно Хэмминга.

3. n_fft – зададим этот параметр равный 882, а $hop_length = \frac{n_fft}{2}$

В качестве результата мы получим матрицу изображенную на рисунке 3. Количество МЧКК будет равно $(79 + 2) \times 20$ для каждого отрезка длительностью 800мс.

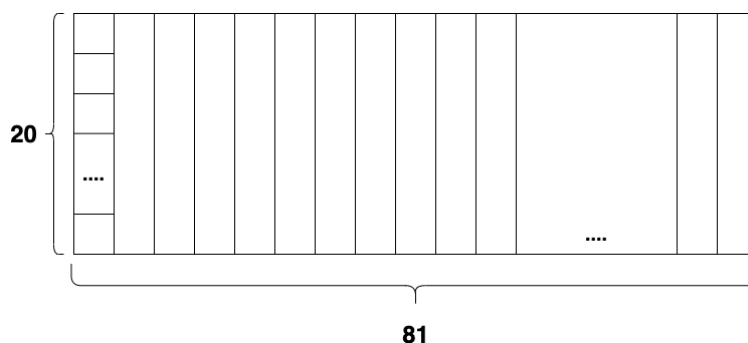


Рисунок 3 – Блок МЧКК, полученный для сегмента длительностью 800 мс

Схема системы диагностирования патологии голосовых складок по звуковому сигналу представлена на рисунке 4.

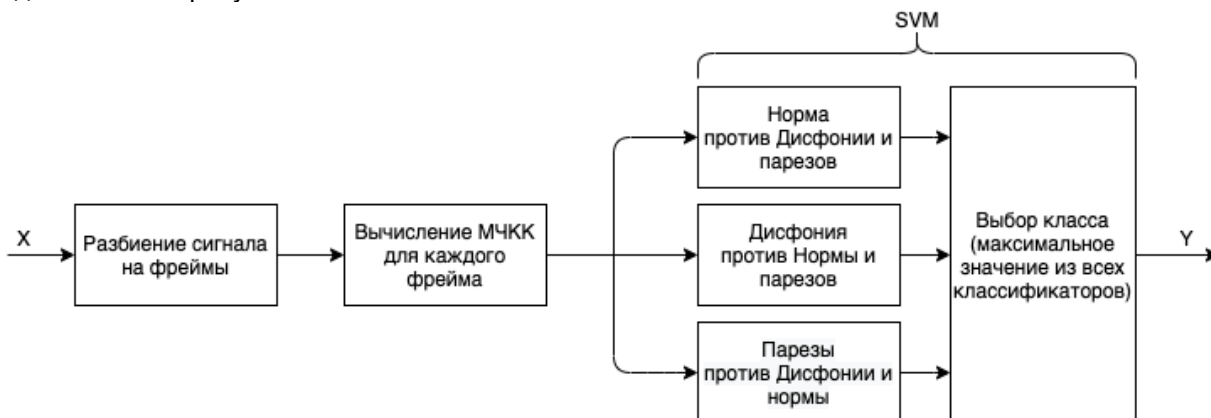


Рисунок 4 – Блок-схема системы диагностирования патологии голосовых складок

Описание базы голосов. В качестве входных данных для обучения нашего классификатора у нас были записи протяжного звука «А» людей с нормальными голосовыми складками, парезами и дисфонией.

Дисфония – расстройство голосообразования, при котором нарушается качество голоса (высота, сила, тембр), вследствие чего он становится неполноценным - слабым, маломодулированным, охриплым и так далее. [8]

Парез гортани (временное, до 12 месяцев, нарушение подвижности одной или обеих половин гортани) — расстройство двигательной функции в виде полного отсутствия произвольных движений вследствие нарушения иннервации соответствующих мышц.[9]

На рисунках 5, 6, 7 представлены графики голосовых сигналов при дисфонии, парезах и отсутствии патологий соответственно.

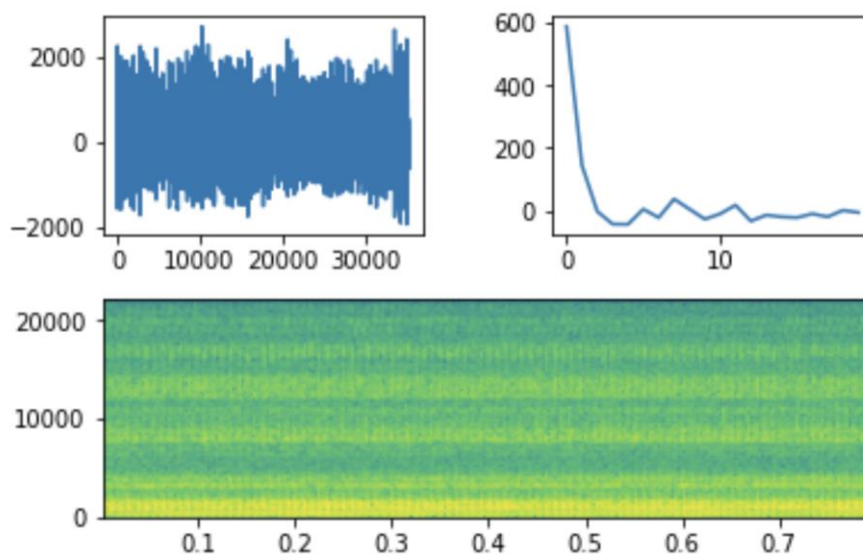


Рисунок 5 – Графики сигнала для дисфонии.

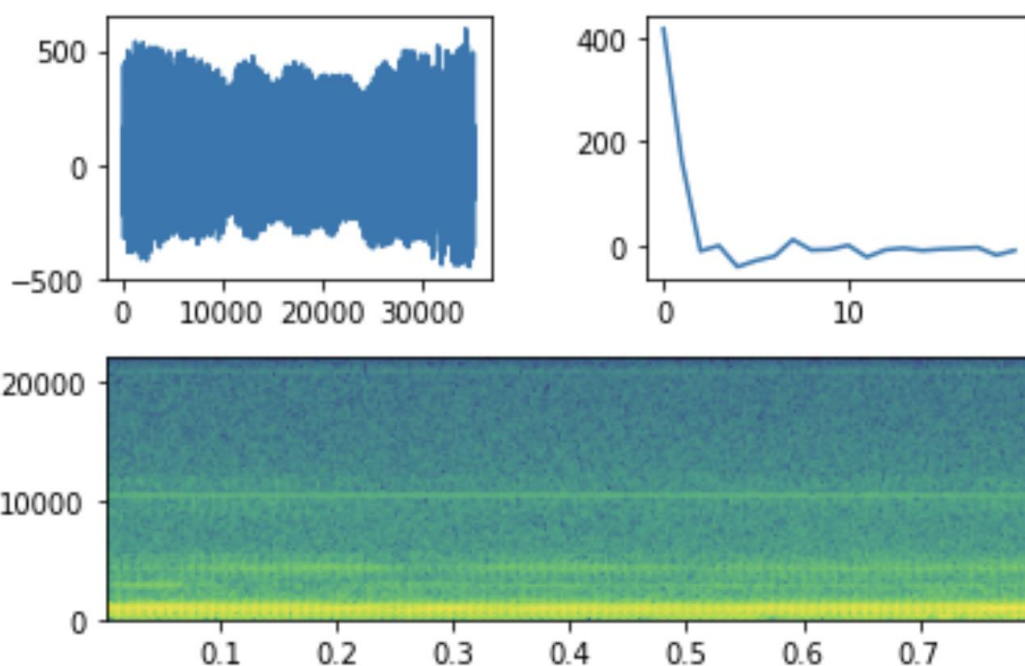


Рисунок 6 – Графики сигнала при парезах.

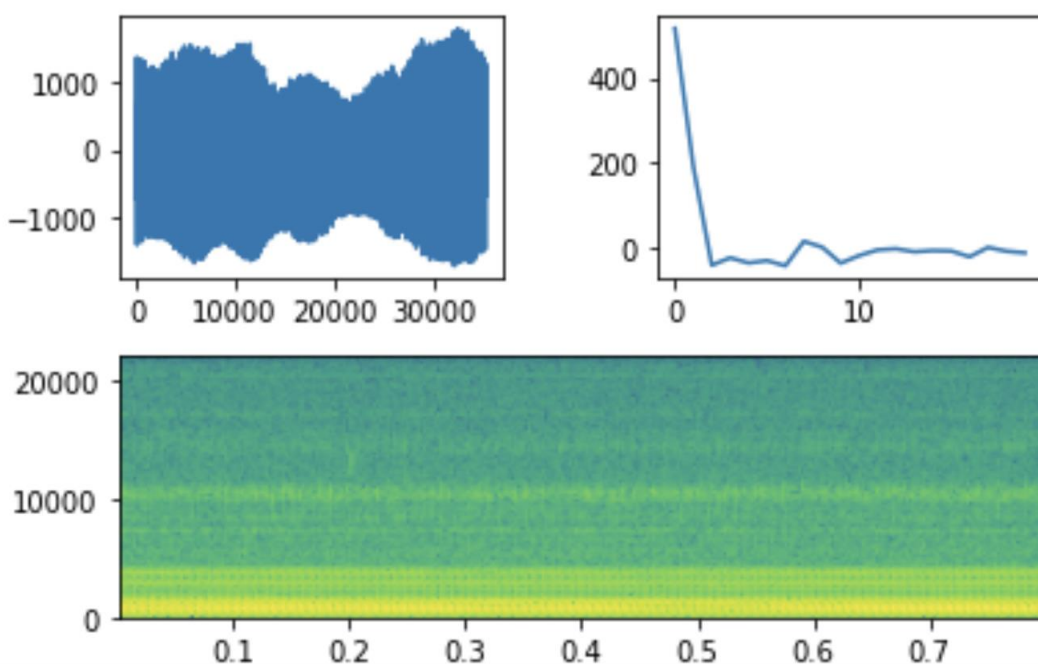


Рисунок 7 – Графики сигнала при отсутствии патологий.

Верхний левый график – это вид входного сигнала во временной области, верхний правый – рассчитанные значения МЧКК этого сигнала, а нижний график – это спектрограмма входного сигнала. Проанализировав данные графики можно заметить, что в случае дисфонии и парезов наблюдаются отличия в области кепстральных коэффициентов и частотной. Также огибающая сигнала во временной области имеет рваную форму в отличие от сигнала без патологий, где огибающая более плавная. Также можно заметить, что при парезах у пациента наблюдается более тихий голос и в следствии этого менее насыщенная спектрограмма. С уверенностью можно сказать, что эти замечания являются отличительными особенностями для данных классов, потому что база голосов записывалась на профессиональном оборудовании, соблюдая правила записи голоса, расстояние до микрофона, громкость записи и так далее.

Описание эксперимента. Результаты проведения эксперимента отображены в таблице 1.

В ходе проведения эксперимента сформирована обучающая выборка, состоящая из 401 примера. Для такого малого количества данных использование наиболее популярной схемы разбиения 80/20, где 80 процентов – это тренировочная выборка, а 20 – это тестирующая, не целесообразно. По этой причине было принято решение воспользоваться методом перекрестной проверки. При оценке модели имеющиеся в наличии данные разбиваются на k частей. Затем на $k-1$

частях данных производится обучение модели, а оставшаяся часть данных используется для тестирования. Процедура повторяется k раз. В итоге полученные результаты усредняются, и мы получаем правильность нашей модели. В нашем случае мы взяли k равное 10. Для сравнения результатов на каждой итерации мы получали матрицу спутывания. Далее объединили все матрицы в одну и посчитали итоговую правильность, как сумму главной диагонали деленную на количество элементов в обучающей выборке. В итоге правильность классификации (англ. accuracy) составила 94%.

Таблица 1 - результаты проведения эксперимента.

Предсказанный \ Правильный класс	Парезы	Дисфония	Норма
Парезы	146	9	3
Дисфония	8	148	4
Норма	1	0	82

Анализируя таблицу 1 можно сказать, что классификатор достаточно точно детектирует голос в норме. Наибольшие ошибки происходят при классификации голоса при патологии. В 9 случаях вместо патологии «парезы» была определена патология «дисфония». В 8 случаях вместо патологии «дисфония» была определена патология «парезы». Таким образом, можно сделать вывод, что требуется дальнейшая доработка системы классификации.

Заключение. В рамках данной работы была получена система, которая разработана с помощью метода опорных векторов и мел-частотных кепстральных коэффициентов. В итоге была получена система диагностирования, имеющая правильность классификации около 94 процентов, что является хорошим результатом, который, однако, требует дальнейшего улучшения.

Список использованных источников:

- Orozco-Arroyave J. R. et al. Characterization methods for the detection of multiple voice disorders: neurological, functional, and laryngeal diseases //IEEE journal of biomedical and health informatics. – 2015. – Т. 19. – №. 6. – С. 1820-1828.
- Wikipedia [Электронный ресурс]. – Режим доступа: <https://ru.wikipedia.org>. – Дата доступа: 12.04.2021.
- Петер Флах Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных. Учебник / Петер Флах. - М.: ДМК Пресс, 2015. – С. 228-229.
- Habr [Электронный ресурс]. – Режим доступа: <https://habr.com>. – Дата доступа: 12.04.2021.
- Документация Scikit-learn [Электронный ресурс]. – Режим доступа: <https://scikit-learn.org> – Дата доступа: 12.04.2021.
- Habr [Электронный ресурс]. – Режим доступа: <https://habr.com> – Дата доступа: 12.04.2021.
- Документация Librosa [Электронный ресурс]. – Режим доступа: <https://librosa.org> – Дата доступа: 12.04.2021.
- Сайт городской клинической больницы имени С.С.Юдина [Электронный ресурс]. – Режим доступа: <https://gkbyudina.ru/articles/> – Дата доступа: 12.04.2021.
- Сайт ГБУЗ НИКИО имени Л. И. Свержевского [Электронный ресурс]. – Режим доступа: <https://nikio.ru/> – Дата доступа: 12.04.2021.

UDC 616-71 + 612.78

DIAGNOSTIC SYSTEM OF PATHOLOGIES OF VOCAL FOLDS BASED ON THE ACOUSTIC ANALYSIS OF THE VOICE SIGNAL AND SUPPORT VECTOR MACHINE

Koleda E.O.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Vashkevich M.I. – PhD in Physics and Mathematics

Annotation. The paper considers the problem of classification of the pathology of the vocal folds based on the acoustic analysis of the voice signal and support vector machine in order to identify the pathology of the vocal folds. The mel-frequency cepstral coefficients, which were calculated during the analysis and processing of the voice signal, were used as features. To test the method, we used a voice database containing recordings of patients with dysphonia, paresis, and people with normal vocal folds. In the result we have accuracy about 94 percents.

Keywords. Support vector machine, voice pathologies, paresis, dysphonia, mel-frequency cepstral coefficients, one versus rest, acoustic signal, vocal folds.