

Projekt z SI

Dominik Lau, Mateusz Kowalczyk, Michał Tarnacki

29 marca 2023

1 Wstęp

Celem projektu było określanie chwil na nagraniu, w których osoba bierze wdech i wydech. Dokonano oceny jakościowej za pomocą detekcji oddechu na żywo jak i ilościowej (przy wykorzystaniu dalej wymienionych metryk).

2 Teoria

2.1 SVM

Pierwszy model, jakim dokonujemy klasyfikacji, jest oparty o SVM. Metoda ta polega na szukaniu optymalnych wag \mathbf{w}, b . Następnie będziemy klasyfikować według funkcji maszyny uczącej

$$f(x) = \text{sgn}(\mathbf{w}^T \mathbf{x} + b)$$

jeśli f zwraca 1 to traktujemy to jako jedną z binarnych decyzji (w naszym przypadku np. wdech) a jeśli zwraca -1 to drugą (czyli np. wydech). \mathbf{x} jest wektorem cech. Znalezienie optymalnych wag będzie polegało na minimalizacji funkcji kosztu

$$J(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 + \frac{C}{N} \sum_x \max\{0, 1 - y(\mathbf{w}^T \mathbf{x} + b)\}$$

gdzie C jest parametrem uregulowania. Funkcję będziemy minimalizować metodą spadku po gradientie (wersję wsadową).

2.2 Ocena modeli

Dokładność (*accuracy*) definiujemy jako

$$\text{accuracy} = \frac{\# \text{correct}}{\# \text{all}}$$

3 Zbieranie i obróbka danych

3.1 Nagrywanie danych oddechowych

Żeby zapewnić dobre oznaczenie danych, etykietujemy je jeszcze w trakcie nagrywania dźwięku. Osoba nagrywana naciska przycisk, żeby zasignalizować, że przestała brać wdech i zaczyna wydychać powietrze lub na odwrót. Pozyskane natężenie przy dalszej obróbce odsumiamy za pomocą dostępnej w Pythonie biblioteki *noisereduce*.

3.2 Średnia częstotliwość w czasie

Pierwotnie przyjętym założeniem było, że podczas wdechu średnia częstotliwość dźwięku jest wyższa niż gdy osoba wydycha. Z pliku w formacie .wav dla bloków próbek generujemy spektrogramy i obliczamy średnią ważoną częstotliwość (czyli znacznie redukujemy rozmiar danych tj. $nowy = stary / rozmiar_{bloku}$). Cechami, na podstawie których dokonywana byłaby predykcja mogłyby być na przykład $\vec{x} = [\bar{f}(t_0), \bar{f}(t_1), \dots, \bar{f}(t_n)]$. Podejście to jednak jak i pierwotne założenie jest mylne - dla niektórych osób dźwięk wydechu jest bowiem wyższy niż wdechu. Ponadto, stosując to podejście, ograniczylibyśmy się tylko do niektórych typów oddechu (np. wdychanie przez nos i wydychanie przez usta). Biorąc pod uwagę wszystkie wady podejście zostało odrzucone.

3.3 Dane wejściowe ze spektrogramu

Innym podejściem jest wzięcie całego spektrogramu (a przynajmniej jego części) jako dane wejściowe do modelu. Podobnie jak ostatnio dzielimy dane na bloki ale tym razem nie liczymy średniej tylko zostawiamy całą taką klatkę. Metoda pochodzi od przypuszczenia, że człowiek rozpoznaje i rozróżnia wdech/wydech na podstawie barwy dźwięku. Wektor cech ma więc postać $\vec{x} = [I_{f_1}, I_{f_2}, \dots, I_{f_n}]$. W dalszej części korzystamy z tego podejścia.

3.4 Ekstrakcja najistotniejszych częstotliwości

3.5 Redukcja szumów

3.6 Standaryzacja

4 Implementacja modeli

4.1 SVM ze spektrogramem

Naszym pierwszym podejściem był SVM, w którym wektor cech stanowił 160 poziomów częstotliwości, tj. $\vec{x} = [I_{f_1}, I_{f_2}, \dots, I_{f_{160}}]$. Na tej podstawie uzyskaliśmy dokładność $\sim 70\%$ dla danych testowych. Problemem było jednak, że często zdarzały się sytuacje, w których klasyfikator w połowie

wdechu/wydechu "przeskakiwał" na chwilę na przeciwną wartość, predykcje te były więc dosyć niestabilne.

4.2 Informacje o poprzednich stanach

Model z poprzedniego podpunktu rozszerzyliśmy o informację o wcześniej przewidzianym stanie, czyli $\vec{x} = [I_{f_1}, I_{f_2}, \dots, I_{f_{160}}, s]$, gdzie $s = 1$ jeśli poprzednio stwierdzono wdech, $s = -1$ jeśli poprzednio stwierdzono wydech. Takim podejściem otrzymaliśmy dokładność $\sim 99\%$ dla danych testowych oraz ograniczyliśmy przeskakowanie przewidywań.