

# Heartbeat classification - Linear Discriminant Analysis

Aleksandra Rak, Kamelia Niemczyk

22 grudnia 2016

## 1 Wstęp

Celem projektu jest dokonanie klasyfikacji zespołu QRS sygnału EKG metodą Liniowej analizy dyskryminacyjnej (LDA). Projekt zakłada opracowanie metody klasyfikacji na podstawie cech charakterystycznych zespołu QRS, implementację wersji prototypowej proponowanego rozwiązania, następnie jego implementację w języku C++ oraz porównanie wyników klasyfikacji z wynikami uzyskanymi przez inne zespoły. W pierwszej części projektu zaimplementowano prototyp aplikacji w środowisku Matlab w celu wstępnej oceny zaproponowanego algorytmu. Niniejszy raport zawiera koncepcję proponowanego rozwiązania oraz opis jego implementacji.

## 2 Sygnały wejściowe

Dane niezbędne do przetestowania zaproponowanej metody uzyskano z bazy MIT-BIH Arrhythmia Database. Wybrano 7 półgodzinnych, dwukanałowych zapisów EKG: 105.dat, 106.dat, 114.dat, 118.dat, 210.dat, 223.dat, 233.dat. Za pomocą biblioteki WFDB Toolbox w programie Matlab odczytano adnotacje do poszczególnych sygnałów, zawierające lokalizację wystąpienia załamków R oraz klasę, do której dany zespół QRS należy.

Z posiadanych zapisów EKG wybrano tylko zespoły QRS należące do klas zestawionych w Tabeli 1. Powodem tego była konieczność posiadania klas o odpowiednio dużej liczebności. Ograniczono więc zbiór danych w postaci zespołów QRS (reprezentowanych przez wykryte wystąpienia załamków R) do dwóch klas N oraz VEB utworzonych z podzbiorów N, V oraz E.

Tabela 1: Klasy uderzeń serca z bazy MIT-BIH wykorzystane w analizie

| Klasa                          | Oznaczenie | Podklasa                          |
|--------------------------------|------------|-----------------------------------|
| N - Normal heartbeat           | N          | Normal beat                       |
| VEB - Ventricular ectopic beat | V          | Premature ventriculat contraction |
|                                | E          | Ventricular escape beat           |

Do przeprowadzenia dalszej analizy, oprócz posiadanych lokalizacji załamków R, niezbędne jest posiadanie informacji o przybliżonym wystąpieniu początku oraz końca każdego zespołu QRS, gdyż klasyfikacji nie podlega cały sygnał lecz jedynie fragmenty zespołów QRS.

### 3 Ekstrakcja cech

Celem uzyskania jak najlepszych wyników klasyfikacji, przeprowadzono przegląd literatury w celu wybrania odpowiednich cech do ekstrakcji z sygnału. Cechy te tworzą razem wektor cech, reprezentujący dany zespół QRS, na podstawie którego algorytm przypisze mu odpowiednią klasę. Prawidłowe skonstruowanie wektora cech jest jednym z czynników mających wpływ na maksymalizację wyników klasyfikacji. Wybrane cechy powinny jak najbardziej różnicować obiekty (zespoły QRS) różnych klas.

W modelu prototypowym dokonano ekstrakcji następujących cech:

- Stosunek pola powierzchni zespołu QRS do obwodu.

$$10 * \frac{\sum_{n=0}^N (|s(n)|)}{\sum_{n=1}^N (|s(n) - s(n-1)|)}$$

gdzie:  $s(n)$  - sygnał,  $n$  - numer próbki

- Amplituda załamka R
- Energia zespołu QRS

$$\sum_{n=0}^N (s(n)^2)$$

- Stosunek dodatniej części amplitudy do ujemnej części amplitudy.

### 4 Liniowa Analiza Dyskryminacyjna

Liniowa analiza dyskryminacyjna (ang. Linear Discriminant Analysis, LDA) należy do technik klasyfikacji z nadzorem – wykorzystuje zbiór uczący do stworzenia reguł (klasyfikatorów), które pozwalają na przypisanie nowego obiektu, o nieznanym przynależności, do którejś z klas. Klasyfikatory są liniowymi funkcjami dyskryminacyjnymi, których współczynniki dobrane są tak, by funkcje te jak najlepiej separowały od siebie obiekty zbioru uczącego należące do różnych klas. Reguły klasyfikacyjne dzielą zatem przestrzeń cech na podzbiory odpowiadające klasom (liczba podzbiorów równa jest liczbie klas). W przypadku, gdy liczba klas wynosi dwa, wystarczające jest wyznaczenie jednej funkcji dyskryminacyjnej. W celu przypisania nowego obiektu do klasy należy obliczyć wartość funkcji dyskryminacyjnej dla jego zmiennych.

$$FD = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n$$

gdzie:

$a_0, a_1, a_2, \dots, a_n$  - współczynniki funkcji dyskryminacyjnej

$x_0, x_1, \dots, x_n$  - zmienne

$n$  - liczba zmiennych

Do stworzenia i przetestowania modelu klasyfikacji z wszystkich posiadanych obserwacji wybrano 2000 pochodzących z różnych sygnałów (bazowano na kanale MLII). Zbiór ten jest na tyle liczny by udało się na jego podstawie zbudować model, a zarazem nie jest na tyle duży by doprowadzić do przeuczenia klasyfikatora.

Aby utworzyć zbiory treningowe i testowe, dostępne wektory cech obliczone dla wszystkich zespołów QRS z danej klasy, podzielono na dwie grupy. 70% załamek z każdej klasy utworzyły zbiór uczący, a pozostałe 30% posłużyły do weryfikacji jakości otrzymanego modelu.

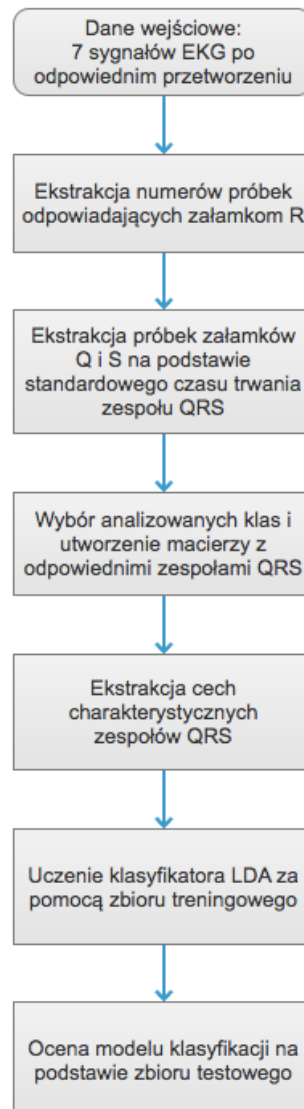
## 5 Opis implementacji

Do wykonania modelu klasyfikacji użyto 7 sygnałów z bazy MIT-BIH Artythmia Databasw, które były już odpowiednio przetworzone i pozwoliły na utworzenie wektorów próbek dla załamek R, a następnie zespołów QRS, którym przyporządkowane były zdefiniowane klasy. Próbkę dla Q oraz S zostały oszacowane na podstawie typowego czasu trwania całego zespołu, który w celu zapewnienia detekcji został zawyżony. Różnicę pomiędzy Q a R ustalono o wartości 63 ms, a R – S: 94 ms. Aby stworzyć macierze klas zawierające załamki jednej klasy, lecz pochodzące z różnych sygnałów, napisano funkcję, która grupowała te dane. Dla grup zespołów QRS obliczono cechy charakterystyczne. W tym celu utworzono 4 funkcje, które opisano w akapicie 3. Tak utworzone wektory cech poddano Liniowej Analizie Dyskryminacyjnej, której celem było uzyskanie odpowiedniego klasyfikatora. Model przetestowano i oceniono na przygotowanym zbiorze testowym.

Tabela 2 przedstawia uzyskane rezultaty klasyfikacji zespołów QRS metodą LDA w założeniu prototypowym. Przyjęta koncepcja rozwiązania zaowocowała poprawną klasyfikacją 540 obserwacji spośród 600, co daje wynik 90% poprawnej klasyfikacji. W kolejnej części projektu, przyjęta metoda zostanie zaimplementowana w języku C++.

Tabela 2: Zestawienie wyników klasyfikacji metodą LDA

| Real/Predicted | N      | VEB    |
|----------------|--------|--------|
| N              | 92,33% | 7,67%  |
| VEB            | 12,33% | 87,67% |



Rysunek 1: Schemat algorytmu