

Moduł HeartClass

MICHAŁ CISZEWSKI, ŁUKASZ DUDEK, KRYSTIAN MUCHA

Akademia Górniczo - Hutnicza w Krakowie

Na przedmiot: Elektroniczne Systemy Diagnostyki Medycznej i Terapii

SPIS TREŚCI

I Wstęp	1
II Koncepcja proponowanego rozwiązania	3
I Klasteryzacja	3
II Klasyfikacja	4
III Rezultaty i wnioski	5
IV Podsumowanie	5
V Bibliografia	5

Streszczenie

Niniejszy artykuł dotyczy części programu do analizy sygnału elektrokardiograficznego, aby wykrywać w nim wszelkie niezgodności z normami. Ten moduł, nazwany "HeartClass", służy do analizy załamków QRS i grupowania ich według odpowiednio zdefiniowanego podobieństwa.

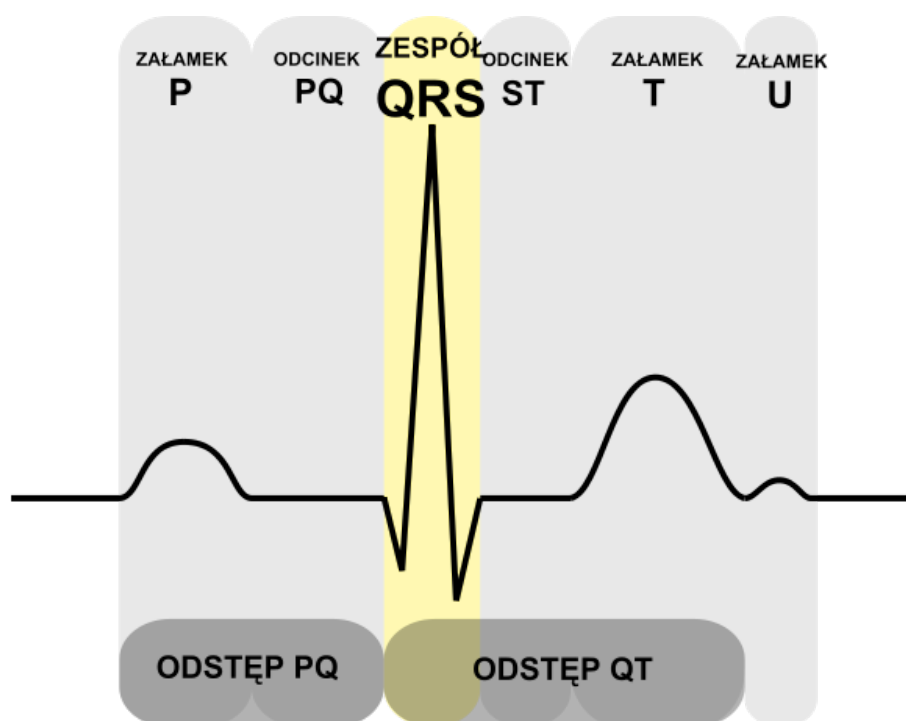
Słowa kluczowe: elektrokardiografia, klasyfikacja załamków QRS, algorytm k-średnich, algorytm G-średnich, metoda wektorów nośnych.

I. WSTĘP

Elektrokardiogram jest jednym z najefektywniejszych narzędzi diagnostycznych do wykrywania chorób serca. ECG dostarcza prawie wszystkich informacji o aktywności elektrycznej serca. Typowy sygnał ECG składa się z załamka P, zespołu QRS oraz załamków T i U. Spośród wszystkich tych elementów sygnału elektrokardiograficznego najbardziej charakterystycznym i zarazem najbardziej znaczącym jest zespół QRS. Na podstawie jego kształtu można zdiagnozować różne dysfunkcje serca, dlatego jego automatyczna klasyfikacja jest ważnym zagadnieniem.

Zespół QRS opisuje pobudzenie mięśni serca i składa się z jednego lub kilku załamków określanych jako Q, R i S.

- Załamek R – każdy załamek dodatni w obrębie zespołu QRS
- Załamek Q – pierwszy ujemny załamek widoczny przed załamkiem R
- Załamek S – pierwszy ujemny załamek widoczny po załamku R



Rysunek 1: Wyidealizowany schemat zapisu EKG z zaznaczonym zespołem QRS. Źródło [1]

Przykładowy (wyidealizowany) zespół QRS widoczny jest na rys. 1, przedstawiającym schematyczny fragment zapisu elektrokardiograficznego.

Klasyfikacja zespołu QRS ma na celu wyodrębnienie grup zespołów podobnych (w zadanym zakresie tolerancji). Odmienny kształt zespołu jest konsekwencją odmiennie przebiegającego pobudzenia. Klasyfikacja polega na stwierdzeniu przynależności klasyfikowanego zespołu do jednej z istniejących klas albo tworzeniu nowych klas, jeżeli przynależności nie stwierdzono [12].

Celem opisywanego modułu jest wyliczenie liczby klas zespołów QRS, określenie reprezentantów każdej z nich oraz oznaczenie klas zespołów QRS na wykresie ECG. Wyodrębnienie klas QRS występujących w sygnale ECG pozwala na określenie prawidłowości rytmu pracy serca. Z reguły nieregularności mają charakter przejściowy, dlatego ich poprawne wyznaczenie wymaga przeprowadzenia 24-godzinnego badania pracy serca, czyli testu Holtera [2].

W literaturze można spotkać się z różnymi podejściami do klasyfikacji zespołów QRS. W [8] autorzy zaproponowali algorytm polegający na wykorzystaniu liniowej analizy dyskryminacyjnej (LDA) w celu zredukowania wymiaru przestrzeni cech. Do klasyfikacji zastosowano maszynę wektorów nośnych (SVM). Ponadto w pracy tej podjęto próbę klasyfikacji przy pomocy MLP (ang. Multilayer Perceptrons) oraz klasyfikatora FIS (ang. Fuzzy Inference System). Najlepsze rezultaty autorzy otrzymali wykorzystując SVM. Podobne podejście zastosowano w [9], gdzie dodatkowo w celu ograniczenia zakłóceń oraz ekstrakcji cech wykorzystano transformację falkową. W [10] autorzy stworzyli adaptacyjny algorytm, działający w czasie rzeczywistym, klasyfikujący zespoły QRS. Zaproponowane rozwiązanie bazuje na modelu funkcji Hermite'a. W innej pracy autorzy wyekstrahowali cztery konkretne cechy z sygnału ECG, aby następnie wykorzystać odległość Mahalanobisa jako kryterium klasyfikacji [11].

Rozwiązanie przyjęte w niniejszej pracy opiera się o ekstrakcję cech z sygnału ECG, klasteryzacji otrzymanych wektorów przy pomocy algorytmu g-means oraz klasyfikacji z wy-

korzystaniem maszyny wektorów nośnych. Podejście takie podyktowane jest ... / wynika z ... ?
 TODO: NAPISAC JAKOS ZE WYKORZYSTUJEMY TO CO NAPISALI ROK TEMU

DO ZROBIENIA WE WSTĘPIE:

1. cel (jest) i założenia projektu
2. badania literaturowe - istniejące rozwiązania (vide linki, które wysłałem)
3. skróconą koncepcję rozwiązania (ekstrakcja - klasteryzacja - klasyfikacja)

II. KONCEPCJA PROPONOWANEGO ROZWIĄZANIA

Algorytm klasyfikacji załamek QRS został podzielony na trzy części. Najpierw dane wejściowe zostają znormalizowane i zkwantyzowane, następnie przeprowadzana jest procedura ekstrakcji cech. W drugiej części następuje klasyfikacja zespołów QRS. Polega on na klasteryzacji danych, czyli grupowaniu ich w klasy, które mają najwięcej wspólnych cech. Warto zaznaczyć, iż każdy współczynnik reprezentuje inną wielkość i z tego powodu wartość tolerancji jest dobierana dla każdego z nich indywidualnie. Do klasteryzacji wykorzystywany jest algorytm g-means clustering. Jego realizacją zajmują się klasy GMeans i SVMClassifier.

I. Klasteryzacja

Do grupowania danych w klastry (lub klasy) użyto algorytmu G-średnich (ang. "G-means"). Jest rozszerzeniem popularnego algorytmu k-średnich, który polega na dobraniu k klas w zbiorze danych tak, aby dla każdego punktu był on w klasie, do której środka ciężkości ma najbliżej [3]. Przyjęto, że dane, które należy pogrupować to d -wymiarowe wektory należące do zbioru X o liczności n . S to zbiór klas, a więc $S_j = \{x_i \in X | \text{klasa}(x_i) = j\}$ dla $i = 1, 2, \dots, n$. Zbiór środków ciężkości klas oznaczony został literą C i zdefiniowany jako: $C = \{c_j = \frac{\sum_{x \in S_j} x}{|S_j|}\}$ dla $j = 1, 2, \dots, k$. Cel algorytmu k-średnich to minimalizacja wyrażenia przedstawionego wzorem 1.

$$\sum_{j=1}^k \sum_{x \in S_j} \|x - c_j\| \quad (1)$$

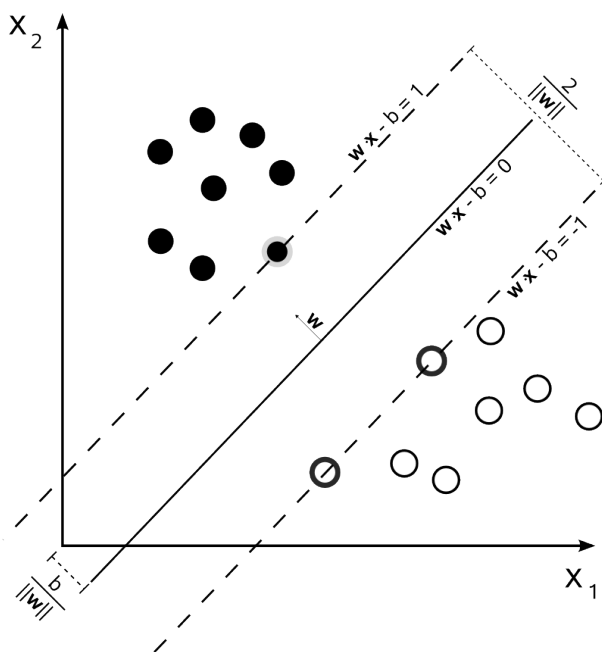
Poważnym problemem tego algorytmu jest fakt, iż liczba klas musi być znana bądź przyjęta z góry, co oznacza jakąś wcześniejszą wiedzę na temat klasteryzowanego zbioru danych [4, 5]. W przypadku, gdy takiej wiedzy się nie posiada, należy zastosować uogólnienie algorytmu k-średnich, które pozwoli dobrać optymalne k względem pewnego wskaźnika jakości. Algorytm, który został zastosowany w opisywanym module dobiera k tak, aby w każdej klasie rozkład punktów był możliwie bliski rozkładowi normalnego. Stąd też wzięta się litera "G" w nazwie - od rozkładu Gaussa [4].

Metoda G-średnich zaczyna od niewielkiej liczby klas, by później odpowiednio zwiększać k - nie jest przewidziana procedura zmniejszania tego parametru. W pierwszym kroku zwykle przyjmowane jest $k = 1$, z czego wynika, że C jest zbiorem jednoelementowym, zawierającym środek ciężkości całego zbioru X [4]. W każdym kroku algorytm sprawdza, czy dana klasa ma rozkład normalny, a jeśli nie, to dodaje jej dodatkowy środek. Między każdym takim dodawaniem środków jest używana procedura k-średnich, aby poprawić jakość rozwiązania.

Sprawdzenie normalności rozkładu wewnątrz klasy odbywa się za pomocą testu Andersona - Darlinga.

II. Klasyfikacja

Aby sklasyfikować powstałe w poprzednim kroku klastry wykorzystano klasyfikator SVM (Support Vector Machine). W najprostszej postaci klasyfikator ten służy do wyznaczenia hiperpłaszczyzny rozdzielającej dwa liniowo separowalne zbiory. Hiperpłaszczyzna ta wyznaczana jest z maksymalnym marginesem, tzn. tak, aby suma jej odległości od najbliższych próbek z obu klas była jak największa (patrz rys. 2).



Rysunek 2: Dwuwymiarowy przypadek hiperpłaszczyzny rozdzielającej dwie klasy z zaznaczonym marginesem. Źródło [6]

W wielu przypadkach nie można zagwarantować liniowej separowalności zbiorów. W takich sytuacjach stosuje się tzw. Kernel trick. Polega on na zwiększeniu wymiaru przestrzeni danych wejściowych, aby w nowej przestrzeni istniała własność liniowej separowalności zbiorów. W tym celu wykorzystuje się różne funkcje jądra (kernel functions). W opisywanym module wykorzystana została funkcja RBF (Radial Basis Function) określona wzorem:

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \quad (2)$$

Aby klasyfikator mógł działać wcześniej należy go wytrenować. Polega to na podaniu mu ciągu wektorów uczących. Opisywany klasyfikator został wytrenowany za pomocą bazy danych MIT-BIH Arrhythmia Database [7]. Gotowy model klasyfikatora wczytywany jest z pliku, w którym zapisane są różne parametry oraz zestaw wektorów nośnych, na których opiera się działanie metody SVM.

III. REZULTATY I WNIOSKI

Fusce mauris. Vestibulum luctus nibh at lectus. Sed bibendum, nulla a faucibus semper, leo velit ultricies tellus, ac venenatis arcu wisi vel nisl. Vestibulum diam. Aliquam pellentesque, augue quis sagittis posuere, turpis lacus congue quam, in hendrerit risus eros eget felis. Maecenas eget erat in sapien mattis porttitor. Vestibulum porttitor. Nulla facilisi. Sed a turpis eu lacus commodo facilisis. Morbi fringilla, wisi in dignissim interdum, justo lectus sagittis dui, et vehicula libero dui cursus dui. Mauris tempor ligula sed lacus. Duis cursus enim ut augue. Cras ac magna. Cras nulla. Nulla egestas. Curabitur a leo. Quisque egestas wisi eget nunc. Nam feugiat lacus vel est. Curabitur consectetur.

IV. PODSUMOWANIE

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Donec odio elit, dictum in, hendrerit sit amet, egestas sed, leo. Praesent feugiat sapien aliquet odio. Integer vitae justo. Aliquam vestibulum fringilla lorem. Sed neque lectus, consectetur at, consectetur sed, eleifend ac, lectus. Nulla facilisi. Pellentesque eget lectus. Proin eu metus. Sed porttitor. In hac habitasse platea dictumst. Suspendisse eu lectus. Ut mi mi, lacinia sit amet, placerat et, mollis vitae, dui. Sed ante tellus, tristique ut, iaculis eu, malesuada ac, dui. Mauris nibh leo, facilisis non, adipiscing quis, ultrices a, dui.

V. BIBLIOGRAFIA

- [1] QRS Complex, https://en.wikipedia.org/wiki/QRS_complex. Stan na: 05.12.2015 r.
- [2] Studenci Automatyki i Robotyki. *ESDMiT - Raport końcowy*. 04.02.2015 r.
- [3] MacQueen, J. *Some methods for classification and analysis of multivariate observations*. W materiałach: "Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics", strony 281-297, University of California Press, Berkeley, Kalifornia, USA, 1967.
- [4] Hamerly G., Elkan Ch. *Learning the k in k-means*. W: Neural Information Processing Systems, MIT Press, 2003.
- [5] Użytkownik 'ashenfad'. *Divining the 'K' in K-means Clustering*. <http://blog.bigml.com/2015/02/24/divining-the-k-in-k-means-clustering/>. Do-dane: 24.02.2015 r. Stan na: 05.12.2015 r.
- [6] Support vector machine, https://en.wikipedia.org/wiki/Support_vector_machine. Stan na: 05.12.2015 r.
- [7] MIT-BIH Arrhythmia Database, <https://www.physionet.org/physiobank/database/mitdb/>. Stan na: 05.12.2015 r.
- [8] Mi Hye Song, Jeon Lee, Sung Pil Cho, Kyoung Joung Lee, and Sun Kook Yoo, Support Vector Machine Based Arrhythmia Classification Using Reduced Features, International Journal of Control, Automation, and Systems, vol. 3, no. 4, pp. 571-579, December 2005
- [9] Abhishek S. Thakare, QRS Complex Detection and Arrhythmia Classification
- [10] P. Laguna, R. Jane, P. Caminal, Adaptive Feature Extraction for QRS Classification and Ectopic Breat Detection, Institut de Cibernetica, Barcelona, Spain

- [11] JCTB Moraes, MO Seixas, FN Vilani, EV Costa, A Real Time QRS Complex Classification Method using Mahalanobis Distance, Escola Politécnica da Universidade de São Paulo, São Paulo, SP, Brazil
- [12] P. Augustyniak, Przetwarzanie Sygnałów Elektrodiagnostycznych, Uczelniane Wydawnictwo Naukowo-Dydaktyczne, Kraków 2001