

Politechnika Gdańska
Wydział Elektroniki Telekomunikacji i Informatyki

Aproksymacja cyklicznych sygnałów zaszumionych z wykorzystaniem hybrydowej detekcji okresowości

Kacper Doga s198020
Dawid Wołoszyn s198034
Agnieszka Pawłowska s188796

Opis problemu

Celem projektu jest opracowanie metody aproksymacji danych okresowych obciążonych szumem i nieregularnościami. Problem polega na identyfikacji okresowości w sygnale oraz dopasowaniu funkcji aproksymującej, która będzie w stanie odtworzyć podstawową strukturę danych przy jednoczesnym filtrowaniu zakłóceń.

Charakterystyka danych wejściowych

Analizowane dane charakteryzują się następującymi właściwościami:

- **Okresowość:** dane wykazują powtarzający się wzorzec w odstępach o długości mniej więcej równej.
- **Szum:** sygnał zawiera losowe fluktuacje, wynikające z metod prowadzenia pomiaru, niedokładności aparatury lub sztucznego nałożenia na dane.
- **Nieregularności:** występują odchylenia od idealnej okresowości, wynikają one naturalnie z danych, przykładowo zmienność amplitudy, przesunięcia fazowe, lokalne zakłócenia lub niekompletność danych.

Wyzwania problemowe

Detekcja Okresowości w obecności szumu

Standardowe metody analizy częstotliwościowej mogą mieć trudność z identyfikacją prawdziwej okresowości, gdy stosunek sygnał/szum jest niekorzystny. Szum może generować fałszywe piki w analizie autokorelacyjnej lub spektralnej.

Zmienność lokalnych charakterystyk

W długich szeregach czasowych okresowość może ulegać zmianom - amplituda, faza, czy nawet częstotliwość mogą ewoluować w czasie. Globalna analiza może nie wychwycić tych lokalnych zmian.

Optymalizacja stopnia aproksymacji

Wybór odpowiedniego stopnia wielomianu aproksymującego kolejny cykl stanowi klasyczny problem bias-variance tradeoff. Zbyt niski stopień prowadzi do underfittingu, natomiast zbyt wysoki do overfittingu.

Segmentacja danych

Długie szeregi czasowe mogą zawierać okresy o różnych charakterystykach, co wymaga inteligentnego podziału na segmenty umożliwiające lokalną analizę bez utraty globalnej spójności.

Cele projektu

Identyfikacja okresowości: opracowanie metody wykrywania dominujących okresów w danych zaszumionych.

Aproksymacja funkcji: dopasowanie funkcji wielomianowej odtwarzającej strukturę danych.

Optymalizacja parametrów: automatyczne dobieranie stopnia wielomianu i parametrów segmentacji.

Walidacja wyników: ocena jakości aproksymacji z wykorzystaniem MSE oraz wizualnych kryteriów oraz porównanie z danymi oryginalnymi.

Znaczenie praktyczne

Problem aproksymacji danych okresowych z szumem ma szerokie zastosowanie, w ramach opracowania, jak i rozwoju projektu wykorzystywane były medyczne parametry takie jak EKG i dane oddechowe.

Zastosowanie rozwiązania może być jednak szeroko stosowane w różnych dziedzinach, między innymi:

- Prognozowaniu szeregów czasowych ekonomicznych.
- Analizie danych klimatycznych i meteorologicznych.
- Przetwarzaniu sygnałów w telekomunikacji.
- Monitorowaniu procesów przemysłowych.

Opracowanie skutecznej metody może przyczynić się do lepszego zrozumienia ukrytych wzorców w danych rzeczywistych oraz poprawy jakości predykcji w systemach opartych na analizie szeregów czasowych.

Teoretyczny opis wykorzystanej metody

Autokorelacja jako podstawa detekcji okresowości

Autokorelacja sygnału $y(t)$ po odjęciu średniej, zdefiniowana jako:

$R(\zeta) = \text{correlate}(y - \bar{y}, \text{mode} = 'full')$, gdzie:

- $y - \bar{y}$ to sygnał po odjęciu wartości średniej
- ζ to przesunięcie czasowe (lag)
- $\text{mode}='full'$ oznacza pełną autokorelację

Detekcja pików w autokorelacji

Identyfikacja okresów odbywa się poprzez znajdowanie lokalnych maksimum w funkcji autokorelacji, punkt $\zeta = 0$ pomijany, unikając trywialne maksimum.

Hybrydowa estymacja okresu podstawowego

Metoda Fast Fourier Transform

Charakteryzuje się szybką implementacją $O(n \cdot \log(n))$, dobrą rozdzielczością dla długich sygnałów. Filtracja Gaussa redukuje wpływ szumu wysokoczęstotliwościowego.

Ograniczeniem tej metody jest jednak problem z sygnałami niestacjonarnymi, oraz rozdzielczość częstotliwościowa ograniczona długością sygnału.

Metoda autokorelacyjna

Metoda działa wobec następujących zasad:

1. Znajdź minimum w pierwszej połowie autokorelacji.
2. Oblicz pochodną od punktu minimum.
3. Pierwszy punkt, dla pochodna większa od 0 wskazuje początek wzrostu.

Fuzja wyników

Końcowy okres jest średnią arytmetyczną estymacji z wykorzystaniem autokorelacji i FFT, co znacząco poprawia stabilność detekcji.

Parametr distance w detekcji pików

Parametr distance w `find_peaks` jest kluczowy dla prawidłowej detekcji okresów, funkcja zapewnia minimalną odległość między wykrywanymi pikami w autokorelacji.

Mechanizm działania opiera się na zagwarantowaniu, że kolejny pik zostanie wykryty dopiero po przejściu odległości równej estymowanemu okresowi, dodatkowo eliminuje redundantne detekcje harmoniczných.

Aproksymacja wielomianowa z normalizacją

Preprocessing danych - normalizacja opiera się na zmapowaniu danych do zakresu $[-1; 1]$, co poprawia stabilność numeryczną. Pozwala to na zmniejszenie overflow/underflow dla wysokich stopni wielomianu, standaryzuje skalę niezależnie od rzeczywistych wartości czasowych.

Problem bias-variance tradeoff

- Niski stopień ($\text{deg} < 3$): wysokie obciążenie (bias), niedopasowanie do struktur nieliniowych.
- Wysoki stopień ($\text{deg} > 8$): wysoka wariancja, overfitting na szumie.
- Optymalny zakres ($\text{deg} = 3-6$): kompromis pomiędzy dopasowaniem a generalizacją.

Odporność na szum addytywny

Jeśli sygnał ma postać $s(t) = \text{signal}(t) + \text{noise}(t)$ to autokorelacja sygnału użytecznego zachowuje strukturę okresową, autokorelacja szumu białego dąży do funkcji delta Diraca. Krzyżowa korelacja sygnał-szum ma wartość oczekiwaną równą zero.

Stabilność detekcji okresów

Kombinacja FFT i autokorelacji zapewnia:

- **Redundancję**: awaryjność jednej metody nie uniemożliwia detekcji.
- **Komplementarność**: FFT lepsze dla sygnałów stacjonarnych, autokorelacja dla niestacjonarnych.
- **Redukcję wariancji**: uśrednienie dwóch niezależnych estymatorów.

Opis realizacji zadania - uzasadnienie podejścia

Architektura rozwiązania

Klasy PeriodicSignal to główna klasa implementująca tryby detekcji okresów, w zależności od parametru `detect=all/segment` następuje analiza całości sygnału lub segmentacja i analiza podzbiorów. Cycle reprezentuje pojedynczy cykl z aproksymacją, służył do porównań kolejnych okresów w zbiorze.

Strategia margiń - rozwiązywanie problemu brzegowego

Problem skoków na brzegach - przy dzieleniu sygnału na okresy występują artefakty interpolacji na granicach segmentów. Przyczyną zjawiska jest brak kontekstu poza granicami przedziału w aproksymacji wielomianowej co prowadzi do niepożądanych oscylacji.

Uzasadnieniem podejścia są teoretyczne podstawy takie jak:

- Wielomiany wysokich stopni wykazują silne oscylacje brzegowe.
- Dodanie kontekstu stabilizuje aproksymację w obszarze badań.
- Po aproksymacji wycinamy fragment odpowiadający rzeczywistemu cyklowi.

Wynikiem jest pozostanie zadowalającej jakości aproksymacji dzięki kontekstowi płynącemu z marginesów, pomimo że w niektórych miejscach okres jest znacząco źle oznaczony.

Analiza segmentowa - ograniczenia i problemy

1. Brak poprawy w detekcji okresów

Obserwacja: Dzielenie na segmenty nie poprawia ani interpolacji ani wyszukiwania okresów. **Przyczyna:** Po kilku iteracjach segmentacji piki zaczynają się przesuwac względem rzeczywistych pozycji okresów.

Mechanizm błędu:

- Każda segmentacja wprowadza małe przesunięcie w detekcji.
- Błędy kumulują się w kolejnych segmentach.
- Końcowy wynik może być gorszy niż analiza globalna.

2. Niewyjaśnione przesunięcie w lewo

Obserwacja: Przy segmentacji interpolacja systematycznie przesuwa się w lewo względem oryginału.

Hipotezy:

- Błąd w indeksowaniu podczas łączenia segmentów.
- Accumulator drift w algorytmie segmentacji.
- Interakcja między margin a algorytmem find_peaks.

Konsekwencje: Mimo teoretycznych korzyści segmentacji, implementacja wprowadza dodatkowe błędy.

Optymalizacja parametrów

Stopień wielomianu

Default: $\text{deg}=5$ w funkcji `interpolate()`

Uzasadnienie wyboru:

- Stopnie 1-2: zbyt proste dla złożonych kształtów biologicznych (underfitting)
- Stopnie 3-5: dobry kompromis pomiędzy dopasowaniem a stabilnością
- Stopnie 6+: ryzyko overfittingu, szczególnie dla krótkich cykli

Mechanizm adaptacji: W przyszłych wersjach można implementować automatyczny wybór na podstawie:

- Długości cyklu (więcej punktów \rightarrow wyższy stopień)
- Poziomu szumu (wyższy szum \rightarrow niższy stopień)
- Validation error na zbiorze testowym

Parametr margin

Default: $\text{margin}=6$ Wpływ na jakość:

- $\text{margin}=0$: skoki na brzegach, niestabilność
- $\text{margin}=3-6$: optymalny zakres dla większości sygnałów
- $\text{margin}>10$: ryzyko włączenia fragmentów z sąsiednich okresów

Krytyczne znaczenie:

- **Za mały distance:** detekcja harmoniczných jako osobnych okresów
- **Za duży distance:** pomijanie prawdziwych okresów w sygnałach o zmiennej częstotliwości
- **Adaptacyjny distance:** bazowany na estymacji z FFT i autokorelacji zapewnia optymalną detekcję

Zalety i ograniczenia implementacji

Zalety

1. **Robustność:** kombinacja FFT i autokorelacji
2. **Stabilność brzegowa:** zastosowanie marginów
3. **Adaptacyjność:** automatyczna estymacja distance
4. **Modularność:** czytelna struktura klas

Ograniczenia

1. **Segmentacja nie działa:** wprowadza dodatkowe błędy
2. **Brak automatycznej optymalizacji stopnia:** stały $\text{deg}=5$
3. **Przesunięcie w segmentacji:** niewyjaśniony błąd systematyczny
4. **Brak walidacji jakości:** implementacja nie ocenia automatycznie jakości aproksymacji

Kierunki rozwoju

1. **Naprawianie segmentacji:** debugowanie przesunięcia w lewo
2. **Adaptacyjny stopień wielomianu:** na podstawie metryk błędu
3. **Cross-validation:** dla optymalnych parametrów
4. **Metryki jakości:** automatyczna ocena MSE, MAE, R^2

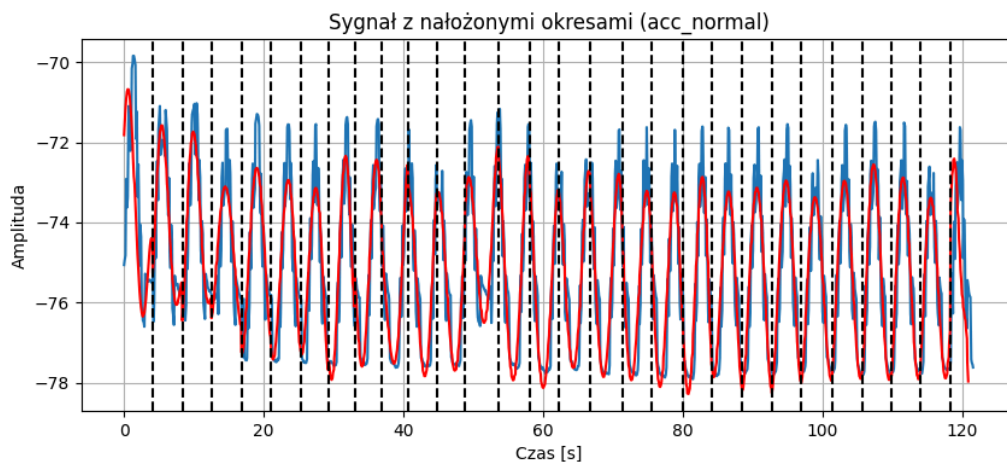
Prezentacja osiągniętych wyników

W wyniku przeprowadzonych prac zrealizowano system umożliwiający aproksymację cyklicznego sygnału wejściowego, charakteryzującego się okresowością, obecnością szumu oraz lokalnymi nieregularnościami. Efektem końcowym jest uzyskanie dopasowanej funkcji wielomianowej dla pojedynczego cyklu oraz rekonstrukcja pełnego przebiegu sygnału przy użyciu tej aproksymacji.

1. Sygnał akcelerometryczny - oddech normalny

Charakterystyka wyniku: Aproksymacja wykazuje bardzo dobre pokrycie względem danych wejściowych. Algorytm skutecznie zidentyfikował dominującą okresowość w sygnale, a dopasowana funkcja wielomianowa odtwarza główne cechy charakterystyczne każdego cyklu. Widoczna jest wyraźna struktura okresowa bez znaczących luk w aproksymacji, co potwierdza

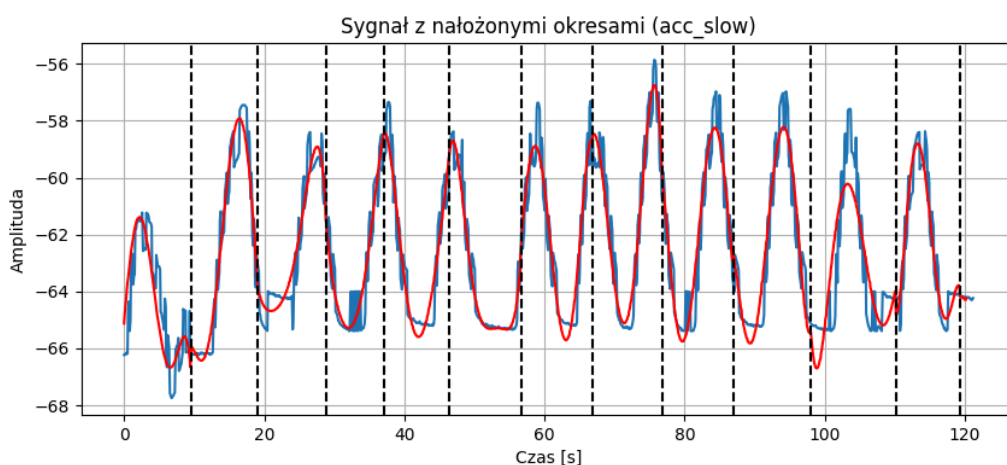
poprawność działania hybrydowej metody detekcji okresów (FFT + autokorelacja).



Wnioski techniczne: Kombinacja estymacji okresów z wykorzystaniem transformaty Fouriera i autokorelacji zapewniła stabilną detekcję podstawowej częstotliwości. Zastosowanie marglin skutecznie wyeliminowało artefakty brzegowe, a stopień wielomianu $\text{deg}=5$ okazał się optymalny dla tego typu sygnału.

2. Sygnał akcelerometryczny - oddech wolny

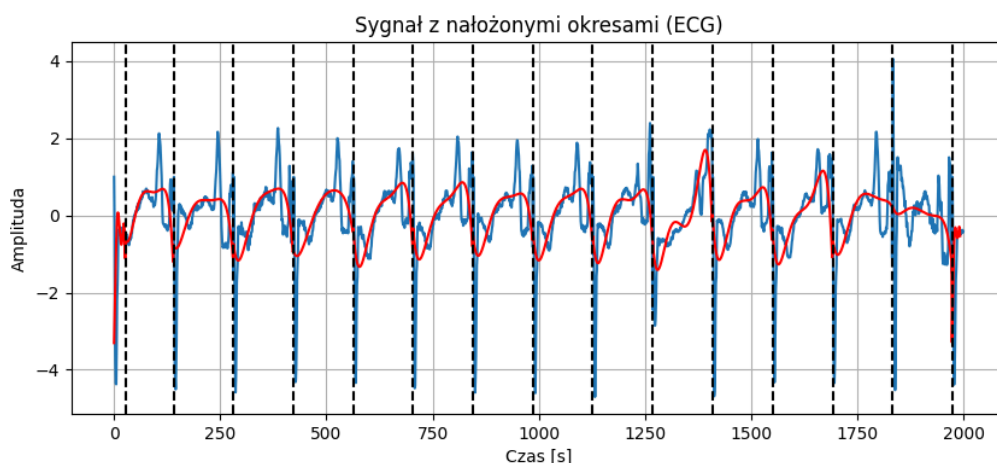
Charakterystyka wyniku: Rezultat przewyższa jakościowo poprzedni przypadek. Aproksymacja wykazuje jeszcze lepsze dopasowanie do struktury danych, a szum wysokoczęstotliwościowy został praktycznie całkowicie wyeliminowany. Algorytm z powodzeniem odfiltrował zakłócenia przy jednoczesnym zachowaniu istotnych cech sygnału użytecznego.



Wnioski techniczne: Wolniejsza zmienność sygnału sprzyjała stabilności detekcji okresów. Niższa częstotliwość podstawowa umożliwiła lepszą rozdzielczość w analizie autokorelacyjnej, a dłuższe cykle zapewniły więcej punktów pomiarowych na pojedynczy okres, co poprawiło jakość aproksymacji wielomianowej.

3. Sygnał EKG - stopień wielomianu w cyklu równy 5

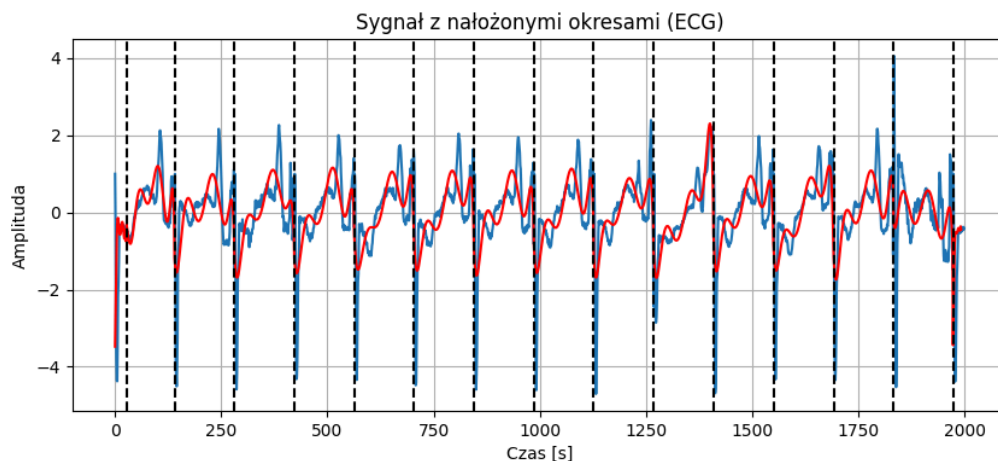
Charakterystyka wyniku: Aproksymacja wykazuje znaczące ograniczenia w odwzorowaniu charakterystycznej morfologii zespołu QRS. Wielomian piątego stopnia okazał się niewystarczający do odtworzenia złożonej struktury sygnału kardiologicznego. Rozkład węzłów interpolacyjnych nie pokrywa adekwatnie ostrych szczytów i stromych zboczy charakterystycznych dla sygnału EKG.



Wnioski techniczne: Sygnał EKG charakteryzuje się wysoką złożonością morfologiczną z ostrymi szczytami (zespół QRS) i stromymi zmianami amplitudy. Wielomian stopnia 5 prowadzi do underfittingu - nie jest w stanie odwzorować nieliniowych charakterystyk sygnału bioelektrycznego serca.

4. Sygnał EKG - stopień wielomianu w cyklu równy 8

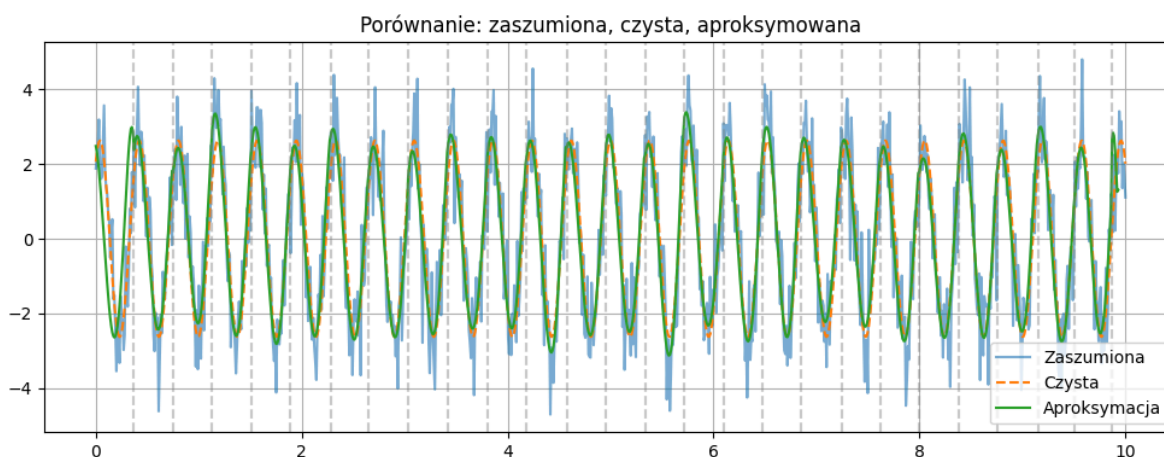
Charakterystyka wyniku: Podwyższenie stopnia wielomianu do 8 skutkuje lepszym odwzorowaniem ogólnego charakteru sygnału EKG. Aproksymacja lepiej oddaje kształt zespołu QRS oraz inne cechy morfologiczne. Jednak nadal obserwuje się znaczący problem z odwzorowaniem początkowej części cyklu, gdzie występuje charakterystyczny spadek amplitudy.



Wnioski techniczne: Wzrost stopnia wielomianu poprawił dopasowanie do złożonych struktur sygnału, ale wprowadził ryzyko overfittingu. Utrzymujący się problem z aproksymacją początku cyklu może wynikać z ograniczeń wielomianowej reprezentacji oscylacyjnych charakterystyk sygnału EKG oraz potencjalnych artefaktów w segmentacji cykli.

5. Sygnał sztuczny z nałożonymi zakłóceniami

Charakterystyka wyniku: Aproksymacja osiąga wyjątkową jakość odwzorowania względem oryginalnych danych bez szumu. Algorytm z powodzeniem odfiltrował wszystkie wprowadzone zakłócenia, w tym: szum addytywny, przeskalowania amplitudy, przesunięcia fazowe oraz inne nieregularności. Rezultat praktycznie idealnie pokrywa się z sygnałem wzorcowym.



Wnioski techniczne: Wynik potwierdza skuteczność opracowanej metody w scenariuszach kontrolowanych. Zastosowanie autokorelacji jako podstawy

detekcji okresowości wykazuje wysoką odporność na szum addytywny, zgodnie z teoretycznymi przewidywaniami. Fuzja estymacji z FFT i autokorelacji zapewnia stabilną detekcję podstawowej częstotliwości nawet przy znacznym poziomie zakłóceń.

Analiza porównawcza i wnioski końcowe

Przeprowadzone testy wykazały, że opracowana metoda osiąga najlepsze rezultaty dla sygnałów o:

- Regularnej strukturze okresowej
- Umiarkowanej złożoności
- Wyraźnie zdefiniowanej częstotliwości podstawowej

Ograniczenia metody ujawniają się w przypadku sygnałów biomedycznych o wysokiej złożoności (EKG), gdzie wielomianowa aproksymacja może być niewystarczająca dla odwzorowania ostrych przejść i nieregularnych kształtów charakterystycznych dla sygnałów elektrofizjologicznych.