

Kryteria selekcji modelu w eksperymentalnym rozpoznawaniu sygnałów zdekomponowanych w bazach falkowych

Urszula Libal¹

Streszczenie: Artykuł przedstawia wyniki eksperymentalnego badania zależności między wielkością modelu a wartością ryzyka błędnego rozpoznawania sygnałów. Eksperyment został przeprowadzony dla zaszumionych sygnałów sinusoidalnego i trójkątnego, zdekomponowanych w bazach falkowych Haara. Falkowa reprezentacja sygnału w postaci współczynników falkowych stanowi pełny model. Do selekcji współczynników zredukowanego modelu użyto algorytmu LASSO (Least Absolute Shrinkage and Selection Operator). Zredukowane modele zostały poddane automatycznej klasyfikacji za pomocą algorytmu 5NN (5 Nearest Neighbours). Celem eksperymentu było wyznaczenie optymalnej wielkości modelu wyselekcjonowanej za pomocą algorytmu LASSO, dla której eksperymentalne ryzyko błędnej klasyfikacji osiągnie minimum. Otrzymany rozmiar modelu dla zaproponowanego kryterium (MIN) porównano z modelami opartymi o kryteria Akaike (AIC) i Bayesa (BIC) dla kilku poziomów zaszumienia sygnałów.

Słowa kluczowe: klasyfikacja, falki, selekcja modelu, dekompozycja sygnału

1. Wprowadzenie

Wybór modelu ma kluczowe znaczenie w wielu zagadnieniach opartych o empiryczne dane. Odpowiednio dobrany model nie powinien być ani zbyt uproszczony, ani zbyt skomplikowany (redundantny). W niniejszym artykule zbadano wpływ zmniejszenia modelu sygnału dzięki zastosowaniu nowego kryterium MIN na jakość rozpoznawania mierzoną ryzykiem błędnej klasyfikacji.

1.1. Zaszumiony sygnał

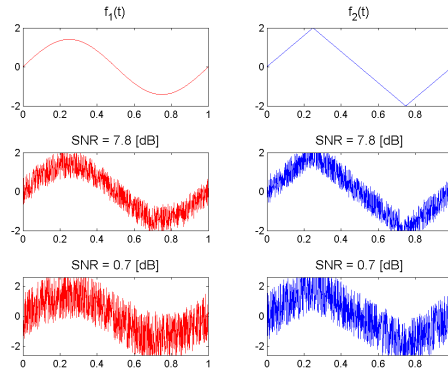
Rozważano problem rozpoznawania sygnałów pochodzących z dwóch klas. W każdej klasie sygnały zawierają zaszumiony wzorec $f(t)$:

$$s(t_i) = f(t_i) + cU_i + \epsilon Z_i, \quad (1)$$

gdzie $t_i = \frac{i}{p_0}$, $i = 0, 1, \dots, p_0 - 1$, a zmienne losowe $\{Z_i\}$ i $\{U_i\}$ są i.i.d. z rozkładu normalnego $Z_i \sim \mathcal{N}(0; 1)$ oraz jednostajnego $U_i \sim \mathcal{U}(-1; 1)$.

¹ Instytut Informatyki, Automatyki i Robotyki, Politechnika Wrocławska, ul. Janiszewskiego 11/17, 50-372 Wrocław, urszula.libal@pwr.wroc.pl

W eksperymencie użyto następujących wzorców: fala sinusoidalna $f_1(t)$ w *klasie 1* oraz fala trójkątna $f_2(t)$ w *klasie 2*, co zostało pokazane na rysunku Rys. 1. Poziom szumu gaussowskiego ustalono na $\epsilon = 0.05$, natomiast dla szumu jednostajnego przyjęto wartości między $c = 0.7$ a $c = 1.6$.



Rys. 1. Deterministyczne wzorce klas: sygnał sinusoidalny $f_1(t)$ i trójkątny $f_2(t)$ oraz sygnały zaszumione $s(t_i) = f(t_i) + cU_i + \epsilon Z_i$ dla $\epsilon = 0.05$, $c = 0.7$ i $c = 1.6$, co odpowiada stosunkowi sygnału do łącznego szumu na poziomach SNR=7.8 [dB] i 0.7 [dB].

1.2. Falkowa reprezentacja sygnału

W dziedzinie czasowo-częstotliwościowej, po dekompozycji w bazach falkowych, sygnał (1) jest reprezentowany przez ciągi współczynników falkowych

$$W(s(t)) = (c_{j_0}, d_{j_0}, d_{j_0+1}, \dots, d_{j_1-1}) = \underline{x}, \quad (2)$$

gdzie $c_{j_0} = (c_{j_0k})_k$ - ciąg falkowych współczynników aproksymacji dla zgrubnej skali j_0 , a $d_j = (d_{jk})_k$ - ciągi falkowych współczynników detali dla skal $j = j_0, j_0 + 1, \dots, j_1 - 1$. Szczegółowy opis dekompozycji sygnału w bazach falkowych można znaleźć w książkach Daubechies [2] oraz Hasiewiczza i Śliwińskiego [4]. Falkowa aproksymacja sygnału ma następującą postać:

$$s(t) \approx \sum_k c_{j_0k} \phi_{j_0k}(t) + \sum_{j=j_0}^{j_1-1} \sum_k d_{jk} \psi_{jk}(t) \quad (3)$$

gdzie ϕ_{jk} i ψ_{jk} to odpowiednio przeskalowana i przesunięta w czasie funkcja skalująca i falka-matka.

1.3. Selektor LASSO (Least Absolute Shrinkage and Selection Operator)

Algorytm LASSO wyznacza estymator $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$ falkowej reprezentacji sygnału β dla pewnego progu $\lambda > 0$ według następującego schematu:

$$\hat{\beta} = \operatorname{argmin} \|y - X\beta\|_2 = \operatorname{argmin} \sum_{i=1}^n \left(y_i - \sum_{j=1}^p \beta_j x_{ij} \right)^2, \quad (4)$$

$$\text{przy } \|\beta\|_1 = \sum_{j=1}^p |\beta_j| \leq \lambda, \quad (5)$$

gdzie macierz obserwacji $X_{n \times p}$ (n -liczba sygnałów w ciągu uczącym, p -liczba cech) oraz wektor odpowiedzi y są dane. Techniczne szczegóły realizacji algorytmu LASSO można znaleźć we wspólnej pracy Efrona, Hastiego, Johnstone'a i Tibshirani'ego [3], natomiast metoda progowania cech dzięki regularyzacji w normie l_1 została omówiona przez Tibshirani'ego w pracy [7]. Dzięki regularyzacji w l_1 wyselekcjonowany model $\hat{\beta}$ składa się jedynie z $df(\lambda)$ niezerowych współczynników falkowych. Rozmiar modelu $df(\lambda)$ zależy od wyboru proggu λ . W następnym punkcie zaprezentowano dwa znane kryteria wyboru λ oraz propozycję nowego kryterium.

2. Kryteria selekcji modelu

Często stosowane kryteria wyboru λ , także dla selektora LASSO [8] zostały podane poniżej:

1. Kryterium AIC (*Akaike Information Criterion*) [1]

$$AIC(\lambda) = \frac{RSS}{p\hat{\sigma}^2} + \frac{2}{p}df(\lambda), \quad (6)$$

2. Kryterium BIC (*Bayesian Information Criterion*) [6]

$$BIC(\lambda) = \frac{RSS}{p\hat{\sigma}^2} + \frac{\ln(n)}{p}df(\lambda), \quad (7)$$

gdzie:

$RSS = \|y - X\hat{\beta}\|^2 = \|y - \hat{y}\|^2$ (błąd estymacji modelu),

p - liczba współczynników falkowej reprezentacji sygnału (cech),

n - liczba sygnałów w ciągu uczącym,

$df(\lambda)$ - rozmiar modelu dla ustalonego λ (liczba współczynników w zbiorze aktywnym, tzn. liczba niezerowych współrzędnych $\hat{\beta}$).

Kryteria AIC i BIC minimalizują błąd RSS między rzeczywistym modelem y a estymowanym obciążonym modelem \hat{y} oraz liczbę niezerowych współczynników df . Kryteria te nie są dedykowane zadaniom rozpoznawania sygnałów, a raczej mają zastosowanie w odsumianiu sygnałów. W eksperymentalnej części pracy pokażemy, że liczba niewyzerowanych współczynników przy użyciu kryteriów AIC i BIC jest relatywnie duża. Dlatego proponujemy nowe kryterium oparte o wyznaczone eksperymentalnie ryzyko dla ciągu uczącego:

3. Kryterium MIN

$$MIN(\lambda) = RISK + g(df(\lambda)), \quad (8)$$

gdzie:

$RISK = R(\hat{\beta})$ - eksperymentalne ryzyko błędnej klasyfikacji modelu $\hat{\beta}$ o $df(\lambda)$ niezerowych współczynnikach falkowych (odpowiadających niezerowym współrzędnym $\hat{\beta}$),

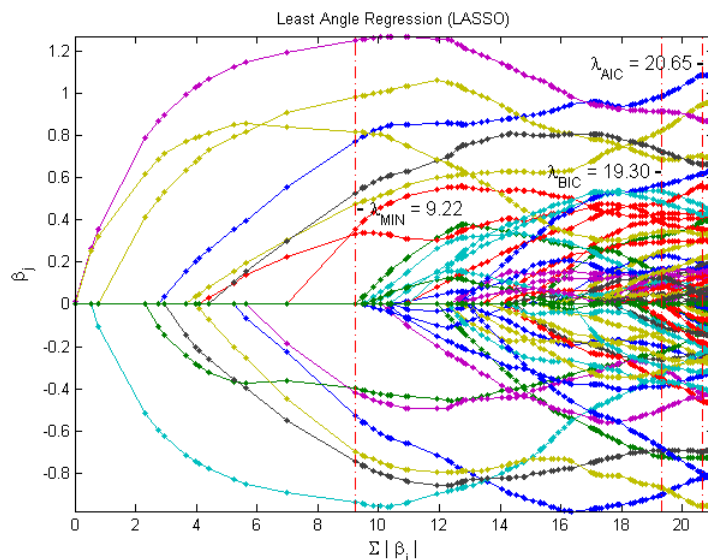
$g(\cdot)$ - niemalejąca funkcja,

$df(\lambda)$ - rozmiar modelu dla ustalonego λ .

Na potrzeby eksperymentu przyjęliśmy, że g jest funkcją identycznościową, tj. $g(x) = x$.

3. Wyniki eksperymentu

Eksperyment jest kontynuacją badań zaprezentowanych w pracy Libal [5]. W niniejszym artykule przyjęto nowe kryterium wyboru progu λ_{MIN} dane wzorem (8). Wygenerowano $n = 100$ zaszumionych sygnałów (po 50 w każdej klasie) dla szumu gaussowskiego na poziomie $\epsilon = 0.05$ oraz jednorodnego na czterech poziomach $c = 0.7, 1.0, 1.3$ i 1.6 . Każdy sygnał zawierał $2^{10} = 1024$ próbek z przedziału czasu $[0, 1]$. W pracy [5] pokazano, że dekompozycja sygnału na większą liczbę poziomów (np. 5, 6 i więcej) skutkuje zmniejszeniem ryzyka. Dlatego zaszumione sygnały zostały zdekomponowane w bazach Haara tylko na dwa ciągi współczynników falkowych $W(s(t)) = (c_{j_0}, d_{j_0})$, aby ułatwić porównanie jakości progowania dzięki wyższym wartościom ryzyka. Otrzymane w ten sposób współczynniki falkowe stanowiły reprezentacje sygnałów. Następnie falkowe współczynniki zostały sprogowane za pomocą techniki LASSO dla wszystkich możliwych wartości progu λ , dla sygnałów z ciągu uczącego. Dla wszystkich sprogowanych modeli $\hat{\beta}$ oraz wszystkich sygnałów z ciągu testowego przeprowadzono rozpoznawanie za pomocą algorytmu 5NN (5 najbliższych sąsiadów). W kolejnym kroku wyznaczono minimalne wartości kryteriów MIN, BIC i AIC oraz obliczono odpowiadające im progi λ_{MIN} , λ_{BIC} and λ_{AIC} .



Rys. 2. Wykres współzależności estymatora LASSO modelu $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)$ z zaznaczonymi wartościami progowymi $\lambda = \sum |\beta_i|$ dla kryteriów AIC, BIC i MIN, dla sygnału o parametrach $\epsilon = 0.05$ i $c=1.0$.

Wraz ze wzrostem wariancji sygnału $Var[s(t_i)] = c^2/3 + \epsilon^2$ (tzn. spadkiem SNR), średnia wartość progów λ_{MIN} , λ_{BIC} i λ_{AIC} również rośnie. Tym samym im wyższe zakłócenie sygnału, tym więcej współczynników pozostaje niewyzerowanych. Tabela 1 zawiera średnie wartości λ , natomiast Tabela 2 - wartości eksperymentalnego ryzyka dla tych λ .

W przeprowadzonych eksperymentach średnie wartości $\sum |\beta_i|$ dla kryteriów MIN, BIC i AIC wykazują zależność $\lambda_{MIN} \leq \lambda_{BIC} \leq \lambda_{AIC}$. Na Rys.3 przedstawiono mediany wartości progów wraz z przedziałem (pudełkiem) między 1 i 3 kwantylem. Przedział ten dla λ_{MIN} jest szerszy niż dla λ_{BIC} i λ_{AIC} , zwłaszcza dla wyższych poziomów szumów. Średnia wartość ryzyka dla λ_{MIN} nie przekroczyła w przeprowadzonym eksperymencie 3.6%.

SNR [dB]	7.8	4.7	2.5	0.7
c	0.7	1.0	1.3	1.6
λ_{MIN}	2.7	6.9	13.2	20.1
λ_{BIC}	14.6	19.2	24.2	28.5
λ_{AIC}	16.0	21.2	26.0	30.0

Tab. 1. Średnia wartość progu λ .

SNR [dB]	7.8	4.7	2.5	0.7
c	0.7	1.0	1.3	1.6
R_{MIN}	0.5	0.6	0.7	3.6
R_{BIC}	0.0	0.0	0.2	2.9
R_{AIC}	0.0	0.0	0.2	3.8

Tab. 2. Eksperymentalne ryzyko [%] dla uśrednionych λ .

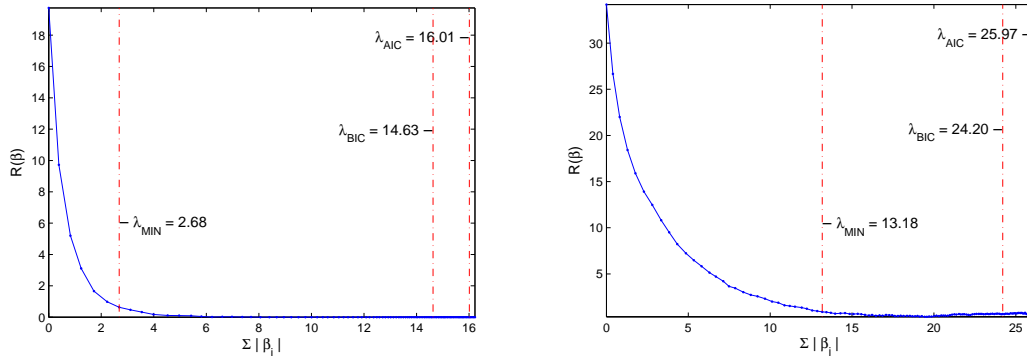
Skutkiem użycia nowego kryterium MIN jest:

- *niski poziom ryzyka* (tylko dla wysokiego poziomu szumu, tzn. dla SNR=0.7 [dB] ryzyko było wyższe niż 1%),
- *mniejszy rozmiar modelu* (zawierający około 10-30 współczynników) w porównaniu do pozostałych dwóch kryteriów (dla BIC około 60-110 pozostawionych cech, a dla AIC około 90-130 cech)
- *i szybsze obliczenia* (potrzeba mniejszej liczby kroków w algorytmie LASSO do wyznaczenia modelu, po przekroczeniu λ_{MIN} procedurę można przerwać).

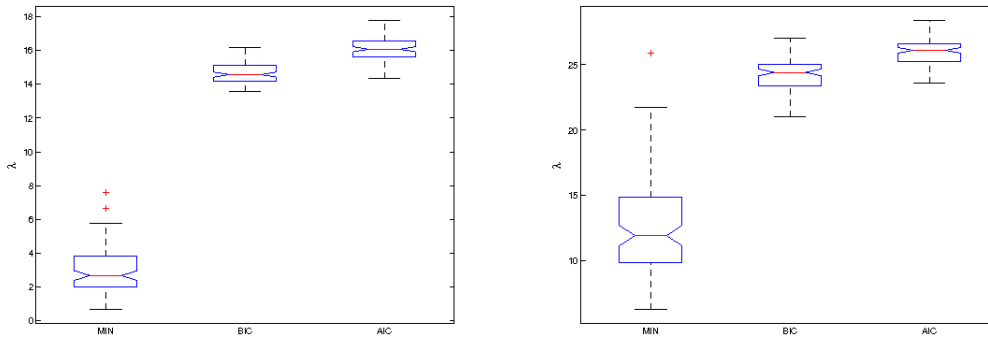
Literatura

- [1] H. Akaike, *Information theory and an extension of maximum likelihood principle*, Proc. 2nd International Symposium on Information Theory, Eds. B.N. Petrov and F. Csaki, 267–281, SciTePress, Budapest, 1973.
- [2] I. Daubechies, *Ten Lectures on Wavelets*, SIAM Edition, Philadelphia, 1992.
- [3] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, *Least Angle Regression*, Annals of Statistics, 32(2), 407–499, 2004.
- [4] Z. Hasiewicz, P. Sliwinski, *Orthogonal wavelets with compact support. Application to non-parametric identification systems (in Polish)*, Exit, Warsaw, 2005.
- [5] U. Libal, *Feature Selection for Pattern Recognition by LASSO and Thresholding Methods – a Comparison*, Proc. 16th IEEE International Conference on Methods and Models in Automation and Robotics - MMAR 2011, 22-25 Aug., 168–173, 2011.
- [6] G. Schwarz, *Estimating the dimension of a model*, Annals of Statistics, 6, 461–464, 1978.
- [7] R. Tibshirani, *Regression Shrinkage and Selection via the Lasso*, Journal of the Royal Statistical Society. Series B, 58(1), 267–288, 1996.

[8] H. Zou, T. Hastie, R. Tibshirani, *On the “degrees of freedom” of the lasso*, *Annals of Statistics*, 35(5), 2173–2192, 2007.



Rys. 3. Eksperymentalne ryzyko R [%] błędnej klasyfikacji z zaznaczonymi przykładowymi wartościami progowymi $\lambda = \sum |\beta_i|$ dla kryteriów AIC, BIC i MIN. Po lewej dla $c = 0.7$, po prawej $c = 1.3$.



Rys. 4. Wykresy pudełkowe progów λ_{MIN} , λ_{BIC} i λ_{AIC} . Po lewej dla $c = 0.7$, po prawej $c = 1.3$.

MODEL SELECTION CRITERIA FOR EXPERIMENTAL PATTERN RECOGNITION OF SIGNAL DECOMPOSED IN WAVELET BASES

There is searched the balance between an increase of pattern recognition risk and a decrease of a model size. The experiments are performed for noisy signals decomposed in Haar wavelet bases. Wavelet representation of signals, by wavelet coefficients called signal features, constitutes the full model. The presented feature selection method is based on the Lasso algorithm (Least Absolute Shrinkage and Selection Operator). The aim of the experiments is to find an optimal model size and investigate the relations between the risk, the number of signal features and the noise level. We propose a new model selection criterion MIN that minimizes both the risk and the number of signal features. We compare the experimental risk of 5NN classification rule for all possible reduced models and for several values of noise variance. Experiments shows that the new criterion MIN allows to reduce the size of model simultaneously with a small increase of misclassification risk in comparison to AIC and BIC criteria.