

OPTIMALIZACJA MODELI PROGNOSTYCZNYCH OPARTYCH NA ESTYMATORACH NAJBLIŻSZEGO SĄSIEDZTWA

Grzegorz Dudek

Słowa kluczowe: krótkoterminowe prognozowanie zapotrzebowania na moc elektryczną, estymatory najbliższego sąsiedztwa, metoda k najbliższych sąsiadów, rozmyty estymator funkcji regresji

Streszczenie. Przedstawiono metody optymalizacji nieparametrycznych modeli prognostycznych opartych na estymatorach najbliższego sąsiedztwa. Modele służą do sporządzania krótkoterminowych prognoz zapotrzebowania na moc. Opisano metody doboru wartości parametrów modeli: liczby najbliższych sąsiadów i parametrów funkcji ważącej w modelu k najbliższych sąsiadów oraz parametru szerokości funkcji przynależności w modelu regresji rozmytej. Do selekcji zmiennych wejściowych zastosowano deterministyczne metody sekwencyjne oraz stochastyczną metodę przeszukiwania turniejowego. Zaproponowano metodę łącznej optymalizacji parametrów szerokości oraz selekcji zmiennych wejściowych. Skuteczność metod optymalizacji zweryfikowano na przykładzie aplikacyjnym.

1. WSTĘP

Nieparametryczne metody estymacji funkcji regresji stanowią alternatywę do podejścia parametrycznego. Elastyczność modeli nieparametrycznych jest bardzo przydatna we wstępnej analizie zbioru danych i może być pomocna do utworzenia modeli parametrycznych, zwykle wygodniejszych w użyciu, choć niekoniecznie dokładniejszych.

Ogólny model regresji ma postać [5]:

$$y = m(\mathbf{x}) + \varepsilon, \quad (1)$$

gdzie y jest zmienną objaśnianą, \mathbf{x} jest wektorem zmiennych objaśniających (predyktorów), ε jest błędem z rozkładu normalnego, o zerowej średniej i wariancji σ^2 , a $m(\mathbf{x}) = E(Y | \mathbf{X} = \mathbf{x})$ jest krzywą regresji.

Celem regresji jest estymacja funkcji $m(\cdot)$. To zadanie może być rozwiązane na dwa sposoby. Pierwszy sposób zwany parametrycznym zakłada, że krzywa $m(\cdot)$ ma pewną założoną z góry postać funkcyjną i jest w pełni opisana za pomocą skończonego zbioru parametrów (np. wielomianowa funkcja regresji). W alternatywnej regresji nieparametrycznej krzywa regresji nie jest predefiniowana, lecz konstruowana na podstawie informacji zawartych w danych. Funkcję regresji w tym przypadku estymuje się bezpośrednio, zamiast estymować jej parametry. Większość metod regresji nieparametrycznej zakłada, że funkcja ta jest ciągła i gładka. Do najpopularniejszych metod regresji nieparametrycznej zalicza się [5]: estymację jądrową, estymację k -najbliższych sąsiadów, lokalną regresję liniową i wielomianową (LOESS), estymację szeregiem ortogonalnymi i wygładzanie splajnami.

2. MODELE PROGNOSTYCZNE

Estymator funkcji $m(\cdot)$ typu k -najbliższych sąsiadów (k -NS) zdefiniowany jest jako średnia ważona zmiennych objaśnianych w sąsiedztwie zmiennej objaśniającej x . Sąsiedztwo zdeterminowane jest wartością parametru k i należy do niego k najbliższych (w sensie przyjętej metryki) sąsiadów zmiennej x . Estymator k -NS można uogólnić na przypadek wielowymiarowy, gdzie zmienne objaśniająca i objaśniana są wektorami:

$$\hat{m}(\mathbf{x}) = \frac{\sum_{i \in \Theta_k(\mathbf{x})} w_i(\mathbf{x}) y_i}{\sum_{i \in \Theta_k(\mathbf{x})} w_i(\mathbf{x})}, \quad (2)$$

gdzie: $\Theta_k(x)$ jest zbiorem indeksów obserwacji, które należą do k najbliższych sąsiadów \mathbf{x} , $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_T]^\top$ jest obrazem dobowego przebiegu obciążenia poprzedzającego moment prognozy – $\mathbf{P} = [P_1 \ P_2 \ \dots \ P_T]^\top$, $\mathbf{y}_i = [y_{i,1} \ y_{i,2} \ \dots \ y_{i,T}]^\top$ jest obrazem prognozy kodującym rzeczywiste obciążenia w kolejnych chwilach doby prognozy $i + \tau$: $\mathbf{P}_{i+\tau} = [P_{i+\tau,1} \ P_{i+\tau,2} \ \dots \ P_{i+\tau,T}]^\top$, τ jest horyzontem prognozy.

Składowe i -tych obrazów wyznacza się wg wzorów:

$$x_{i,t} = \frac{P_{i,t} - \bar{P}_i}{\sqrt{\sum_{j=1}^T (P_{i,j} - \bar{P}_i)^2}}, \quad y_{i,t} = \frac{P_{i+\tau,t} - \bar{P}_i}{\sqrt{\sum_{j=1}^T (P_{i,j} - \bar{P}_i)^2}}. \quad (3)$$

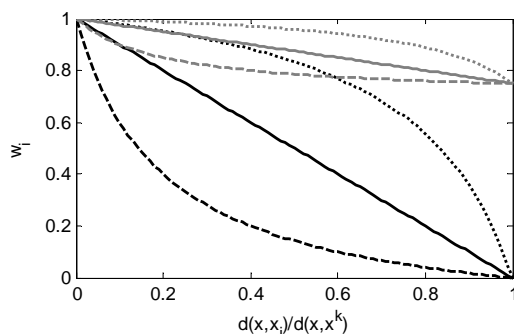
Wagi obserwacji w_i w (2) różnicuje się w zależności od ich odległości od obrazu \mathbf{x} , tak aby obserwacje bliższe miały większy udział w średniej (2). Funkcje ważące zwykle są monotonicznie malejące, osiągają maksymalną wartość w zerze, a minimalną

(nieujemną) – dla k -tego najbliższego sąsiada. Generalnie im gładza jest funkcja wagowa, tym gładza jest funkcja estymowana. Kilka postaci funkcji ważących podano w [1]. Tutaj stosuje się następującą definicję:

$$w_i(\mathbf{x}) = p \left(\frac{1 - \frac{d(\mathbf{x}, \mathbf{x}_i)}{d(\mathbf{x}, \mathbf{x}^k)}}{1 + \lambda \frac{d(\mathbf{x}, \mathbf{x}_i)}{d(\mathbf{x}, \mathbf{x}^k)}} - 1 \right) + 1, \quad (4)$$

gdzie: $i \in \Theta_k(\mathbf{x})$, \mathbf{x}^k – k -ty najbliższy sąsiad obrazu \mathbf{x} w zbiorze uczącym, $d(\mathbf{x}, \mathbf{x}_i)$ – odległość pomiędzy obrazami \mathbf{x} , a \mathbf{x}_i , $p \in [0, 1]$ – parametr sterujący stopniem różnicowania wag, $\lambda \geq -1$ – parametr sterujący wypukłością funkcji.

Dla $p = 1$ wagi są najbardziej zróżnicowane, dla $p = 0$ wagi wszystkich najbliższych sąsiadów są jednakowe, równe 1. Dla $\lambda = 0$ funkcja ważąca (4) jest liniowa, dla $\lambda > 0$ maleje szybciej, a dla $\lambda < 0$ – wolniej niż funkcja liniowa. Wykresy tej funkcji pokazano na rys. 1.



Rys. 1. Wykresy funkcji ważącej (4) dla $\lambda = 0$ – linia ciągła, $\lambda = 5$ – linia kreskowa, $\lambda = -0,8$ – linia kropkowa, $p = 1$ – kolor czarny i $p = 0,25$ – kolor szary.

Wagi można uzależnić nie od odległości pomiędzy obrazami lecz od pozycji (rangi) $r = 1, 2, \dots, k$ obrazu \mathbf{x}_i w rankingu najbliższych sąsiadów. W takim przypadku odległość względną $d(\mathbf{x}, \mathbf{x}_i)/d(\mathbf{x}, \mathbf{x}^k)$ w funkcji ważącej (4) należy zastąpić ilorazem r_i/k .

Liczba najbliższych sąsiadów k jest parametrem modelu, który reguluje stopień wygładzania estymowanej krzywej regresji. Gdy $k = 1$ funkcja regresji ma postać schodkową, dokładnie wpasowaną w punkty uczące. Zwiększenie k prowadzi do wygładzenia funkcji regresji, co implikuje zwiększenie obciążenia i redukcję wariancji modelu. Zwykle k uzależnia się od liczby obserwacji N , np. $k = N^{0,5}$.

Do otoczenia punktu \mathbf{x} w opisanym powyżej podejściu zalicza się jego k najbliższych sąsiadów. W [2, 3] wprowadzono rozmytą przynależność

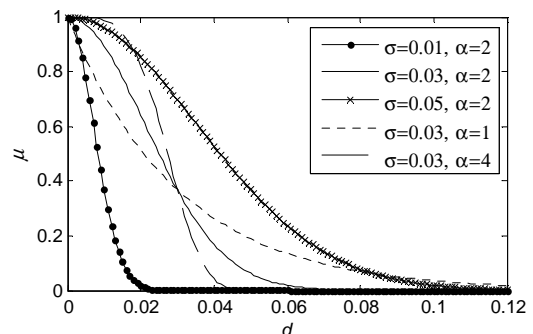
obserwacji do otoczenia punktu \mathbf{x} . W tym przypadku każdy punkt ze zbioru uczącego należy do tego otoczenia w różnym stopniu, zależnym od odległości. Funkcje ważące przyjmują postać funkcji przynależności $\mu(\mathbf{x}, \mathbf{x}_i)$, np.:

$$\mu(\mathbf{x}, \mathbf{x}_i) = \exp \left[- \left(\frac{d(\mathbf{x}, \mathbf{x}_i)}{\sigma} \right)^\alpha \right], \quad (5)$$

gdzie: σ – parametr szerokości, α – parametr kształtu. Funkcja przynależności (5) jest typu gaussowskiego z maksimum w punkcie $d(\mathbf{x}, \mathbf{x}_i) = 0$. Estymator krzywej regresji, zwany dalej rozmytym estymatorem funkcji regresji (REFR), ma postać:

$$\hat{m}(\mathbf{x}) = \frac{\sum_{i=1}^N \mu(\mathbf{x}, \mathbf{x}_i) y_i}{\sum_{i=1}^N \mu(\mathbf{x}, \mathbf{x}_i)}. \quad (6)$$

Wykresy funkcji przynależności (5) przedstawiono na rys. 2.



Rys. 2. Wykresy funkcji przynależności (5).

3. OPTIMALIZACJA MODELI

Dobór wartości parametrów modeli

Opisane powyżej modele oparte na estymatorach najbliższego sąsiedztwa charakteryzują się niewielką liczbą parametrów. W metodzie k -NS należy dobrać liczbę sąsiadów k oraz wartość parametrów funkcji ważącej p i λ . Parametry te estymuje się w procedurach iteracyjnych zmieniając ich wartości w przyjętych granicach. Parametry p i λ należy zdyskretyzować.

W modelach z rozmytą przynależnością występują dwa parametry – σ i α . Do optymalizacji tych modeli w [2] zastosowano metodę pełzającego sympleksu Nelder-Meada oraz metodę quasi-newtonowską. Aby zwiększyć prawdopodobieństwo znalezienia minimum globalnych stosowano multistart

(uruchamiano algorytm z różnych punktów startowych).

Szerokość funkcji przynależności można zindywidualizować dla każdego \mathbf{x} , uzależniając ją np. od zagęszczenia próbek uczących w otoczeniu obrazu \mathbf{x} . Wartość σ w takim przypadku powinna być odwrotnie proporcjonalna do gęstości. Do estymacji gęstości można użyć dowolnej metody z grupy metod parametrycznych lub nieparametrycznych. W najprostszych estymatorach najbliższego sąsiedztwa gęstość jest odwrotnie proporcjonalna do odległości od k -tego najbliższego sąsiada. Stąd przyjęto:

$$\sigma = d(\mathbf{x}, \mathbf{x}^k). \quad (7)$$

Przybliża to model REFR do modelu k -NS. Różnica jest taka, że w wariancie rozmytym funkcja przynależności ma nieograniczony nośnik, co pozwala uwzględniać wpływy obserwacji spoza zbioru k najbliższych sąsiadów. Optymalizacja modelu REFR przy definicji parametru szerokości wg (7) polega na doborze wartości k minimalizującej funkcję błędu.

Selekcja składowych obrazów wejściowych

Składowe obrazów wejściowych x_i są silnie skorelowane, co oznacza, że ta sama informacja o obrazie prognozy przenoszona jest przez wiele składowych. Usunięcie składowych nadmiarowych może podnieść jakość modelu, prowadzi do jego uproszczenia i poprawy generalizacji. Najpopularniejszymi metodami selekcji zmiennych wejściowych, oprócz metody przeglądu zupełnego, która zawodzi przy dużej liczbie zmiennych, są metody sekwencyjnego dodawania i sekwencyjnej eliminacji cech [6].

Metoda sekwencyjnego dodawania cech (SDC) opierając się na prostej zachłannej heurystyce, startuje od pustego zbioru cech istotnych Ω_0 i w każdym kroku dodaje jedną nową cechę do bieżącego zbioru Ω_i . Aby wyłonić tę cechę algorytm dodaje tymczasowo do zbioru Ω_i jedną cechę spośród jeszcze nie wybranych (kandydującą) i testuje dokładność modelu zbudowanego i zoptymalizowanego na cechach zawartych w Ω_i . Ta cecha kandydująca, która zapewnia największą dokładność modelu zostaje ostatecznie włączona do zbioru Ω_i . Algorytm zatrzymuje się, gdy dodanie nowej cechy skutkuje pogorszeniem dokładności modelu. Metoda sekwencyjnej eliminacji cech (SEC) startuje od zbioru zawierającego wszystkie cechy i usuwa sukcesywnie po jednej cesze stosując analogiczny algorytm do SDC.

Skuteczniejszą, globalną optymalizację przestrzeni zmiennych wejściowych x_i zapewniają metody stochastyczne, np. algorytm genetyczny, symulowane wyzarczenie czy metoda przeszukiwania turniejowego [4].

W metodzie turniejowej selekcji cech (TSC) rozwiązanie kodowane jest za pomocą wektora binarnego $\mathbf{b} = [b_1 \ b_2 \ \dots \ b_n]$, gdzie każdy bit reprezentuje cechę, a jego wartość wskazuje czy cecha jest istotna (1) czy nie (0). Przestrzeń cech przeszukiwana jest w sąsiedztwie rozwiązania bazowego \mathbf{b}^* za pomocą operatora przesunięcia. Sąsiedztwo zdefiniowane jest jako $\Phi = \{\mathbf{b} \mid d_H(\mathbf{b}, \mathbf{b}^*) = 1\}$, gdzie d_H jest odległością Hamminga. Operator przesunięcia produkuje $l \in [1, n]$ różnych rozwiązań próbnych $\mathbf{b}_i \in \Phi$, $i = 1, 2, \dots, l$, co oznacza, że rozwiązania próbne losowane są bez powtórzeń z sąsiedztwa rozwiązania bazowego. Startowe rozwiązanie bazowe zainicjowane jest niezależnymi liczbami losowymi o rozkładzie zero-jedynkowym ($P(b_i = 0) = P(b_i = 1) = 0,5$). Rozwiązania próbne są oceniane – wyznacza się błędy prognoz obliczonych za pomocą modelu działającego na wyselekcjonowanych składowych obrazów wejściowych. Rozwiązanie próbne ocenione najwyżej (zwycięzca turnieju) staje się rozwiązaniem bazowym w następnej iteracji, nawet w przypadku gdy jego ocena jest niższa od oceny rozwiązania bazowego. Ten mechanizm, zapewnia możliwość ucieczki z pułapki minimum lokalnego. Jeśli rozmiar turnieju $l = 1$, procedura TSC sprowadza się do przeszukiwania losowego (globalnego). Z drugiej strony, gdy $l = n$ TSC staje się algorytmem wspinaczkowym (lokalnym). Rozmiar turnieju decyduje więc o własnościach eksploracyjno-eksploatacyjnych algorytmu i może być stały lub zwiększać się wraz z licznikiem iteracji.

1. Utworzenie startowego rozwiązania bazowego (losowo).
2. Utworzenie l rozwiązań próbnych z rozwiązania bazowego za pomocą operatora przesunięcia.
3. Ocena rozwiązań próbnych przy użyciu modelu prognostycznego.
4. Turniej - selekcja najlepszego rozwiązania w zbiorze rozwiązań próbnych.
5. Przyjęcie zwycięzcy turnieju, jako nowego rozwiązania bazowego.
6. Powtarzaj kroki 2 - 6 do momentu spełnienia warunku zatrzymania.

Algorytm 1. Turniejowa selekcja cech.

Jeśli przyjąć, że składowe wektora \mathbf{b} mogą być liczbami rzeczywistymi (najczęściej z przedziału $[0, 1]$) mamy do czynienia z ważoną selekcją cech. Do przeszukiwania przestrzeni wag $b \in [0, 1]$

zastosowano w [2] strategię ewolucyjną (μ , λ) oraz algorytm ewolucyjny z reprezentacją zmiennopozycyjną.

Optymalizacja łączna: dobór parametrów szerokości i selekcja składowych obrazów wejściowych

Do jednoczesnego przeszukiwania przestrzeni cech i parametrów szerokości (k lub σ) proponuje się algorytm oparty na algorytmie binarnej selekcji cech, gdzie każde rozwiązanie $\mathbf{b} = [b_1 \ b_2 \ \dots \ b_n]$ oceniane jest dla każdej wartości parametru k lub σ z przyjętego zakresu (σ należy zdyskretyzować). Za ostateczną ocenę rozwiązania \mathbf{b} przyjmuje się najlepszą z tych ocen. Algorytm selekcji zwraca najlepiej oceniony wektor \mathbf{b} wraz z optymalną dla tego wektora wartością parametru szerokości. Algorytm ten jest kombinacją algorytmu selekcji cech i przeglądu zupełnego wartości parametru szerokości. Aby uwzględnić przeszukiwanie wartości parametru szerokości, schemat algorytmu TSC (Algorytm 1) należy nieznacznie zmodyfikować zmieniając p. 3 następująco: „Ocena rozwiązań próbnych przy użyciu modelu prognostycznego dla $k = 1, 2, \dots, k_{\max}$ (lub $\sigma = \sigma_{\min}, \sigma_{\min} + \Delta, \dots, \sigma_{\max}$). Ostateczna ocena rozwiązania próbnego jest najlepszą spośród tych ocen.”

3. BADANIA SYMULACYJNE

Wstępne badania [2] wykazały, że przyjęcie stałej wartości wykładnika $\alpha = 2$ w funkcji (5) i dobór wartości tylko jednego parametru (σ) powoduje jedynie nieznaczne pogorszenie rezultatów – średnio o ok. 2% na zbiorach uczących i 0,6% na zbiorach testowych. Proces optymalizacji modelu w tym przypadku jest znacznie mniej kosztowny czasowo. Szerokość funkcji przynależności σ nie jest parametrem krytycznym. Zmiana zoptymalizowanej wartości σ o 20% powodowała nieznaczną zmianę błędu prognozy, średnio o ok. 1% na zbiorze uczącym i 2,3% na zbiorze testowym.

Skuteczność opisanych powyżej metod optymalizacji przetestowano na danych rzeczywistych – szeregu czasowym obciążenia KSE w latach 2002-2004. Zadania prognostyczne polegały na prognozie przebiegu zapotrzebowania w dniu j z wyprzedzeniem jednodobowym. Próba testowa obejmowała obserwacje:

- ze stycznia 2004 r., bez 1-go stycznia, tj. 30 obserwacji,
- z lipca 2004 r., tj. 31 obserwacji.

Modele optymalizowano niezależnie dla każdego zadania prognostycznego na próbach uczących, które zawierały obrazy reprezentujące te same typy dni tygodnia, co obrazy testowe (z wyłączeniem dni nietypowych), z okresu od 01.01.2002 r. do obserwacji poprzedzającej obserwację, którą prognozowano. W trakcie optymalizacji błąd prognozy estymowano na próbkach walidacyjnych w procedurze minus jednego elementu (*leave-one-out cross-validation*) [6].

Badania przeprowadzono wg następującego harmonogramu:

1. Dobór liczby sąsiadów k oraz parametrów p i λ w modelu k -NS.
2. Dobór parametru szerokości σ w modelu REFR (przyjęto $\alpha = 2$).
3. Optymalizacja łączna: dobór parametrów szerokości i selekcja składowych obrazów wejściowych dla najlepszego modelu z p. 1 i 2.

Liczbę sąsiadów w p. 1 harmonogramu zmieniano od 1 do 50.

Optymalizację (przeгляд zupełny) przeprowadzono dla trzech wartości parametru λ : 0, -0,8 i 5 (patrz rys. 1), pięciu wartości parametru $p = 0, 0,25, \dots, 1$ i dla dwóch wariantów funkcji ważących: postaci (4) oraz zależnych od rang r_i najbliższych sąsiadów.

Błędy walidacyjne osiągały mniejsze wartości dla większych wartości parametru p . Średnie błędy (MAPE) dla próbek walidacyjnych przy $p = 1$ wynosiły: 1,58 dla modelu liniowego ($\lambda = 0$), 1,60 dla $\lambda = -0,8$ oraz 1,57 dla $\lambda = 5$. Dla rangowych odpowiedników tych modeli błędy były następujące: odpowiednio 1,62, 1,63 oraz 1,61.

W tab. 1 pokazano błędy prognoz dla modelu k -NS z $p = 1$ i $\lambda = 0$, a na rys. 3 optymalne wartości k dla kolejnych próbek testowych.

Parametr szerokości funkcji przynależności (5) (p. 2 harmonogramu badań) zmieniano wg schematu:

- (i) $\sigma = b \cdot d_{\text{med}}$, gdzie d_{med} jest medianą odległości pomiędzy obrazami \mathbf{x} w zbiorze uczącym, $b = 0,02, 0,04, \dots, 1$ lub
- (ii) $\sigma = d(\mathbf{x}, \mathbf{x}^k)$, $k = 1, 2, \dots, 50$.

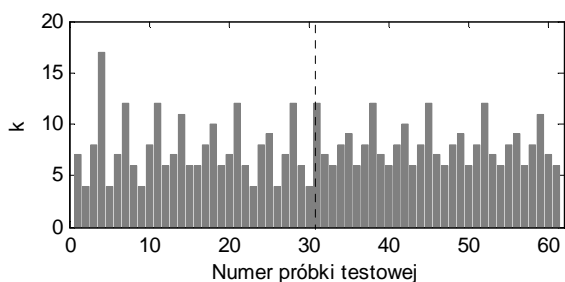
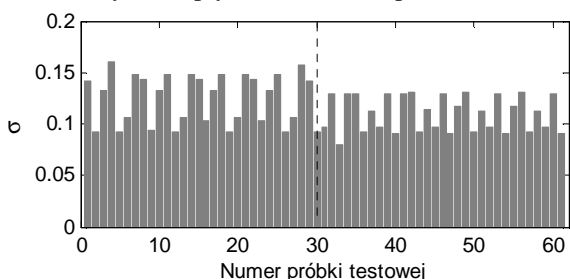
Niższe błędy osiągnęto, gdy σ była dobierana wg schematu (i). Średnie błędy na zbiorach walidacyjnych były następujące: 1,55 dla schematu (i), 1,63 dla schematu (ii). Optymalna wartość parametru σ nie przekraczała $0,32d_{\text{med}}$, przy jego wartości modalnej $0,20d_{\text{med}}$. W tab. 1 zestawiono błędy prognoz, a na rys. 4 pokazano optymalne wartości σ dla kolejnych próbek testowych. Warto odnotować, że przyjęcie jednakowej wartości parametru b dla wszystkich zadań testowych, równej jego wartości modalnej (równej 10), nie pogorszało

jakości modelu mierzonej błędem średnim na próbkach testowych.

Tabela 1

Błędy prognoz dla optymalnych modeli

Model	Styczeń		Lipiec		Średni	
	$MAPE_{tst}$	σ_{tst}	$MAPE_{tst}$	σ_{tst}	$MAPE_{tst}$	σ_{tst}
<i>k</i> -NS	1,47	1,31	0,99	0,84	1,23	1,16
REFR	1,22	1,16	0,96	0,72	1,08	1,02

Rys. 3. Optymalne wartości parametru k .Rys. 4. Optymalne wartości parametru σ .

Dla modelu REFR przeprowadzono optymalizację łączną: selekcję składowych obrazów wejściowych i dobór parametru szerokości. Rozmiar turnieju w TSC wynosił 8, a liczba iteracji $M = 100$. Warunkiem stopu było osiągnięcie zadanej liczby iteracji lub brak postępów przez 0,25M kolejnych iteracji. Wyniki zestawiono w tab. 2.

Tabela 2

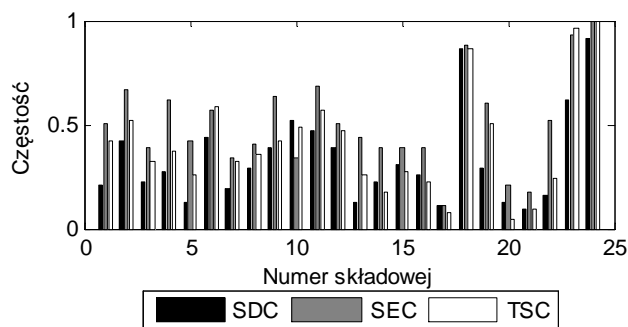
Błędy prognoz dla modelu REFR przy różnych metodach selekcji zmiennych wejściowych

	Styczeń		Lipiec		Średni	
	$MAPE_{wal}$	$MAPE_{tst}$ (σ_{tst})	$MAPE_{wal}$	$MAPE_{tst}$ (σ_{tst})	$MAPE_{wal}$	$MAPE_{tst}$ (σ_{tst})
SDC	1,46	1,13 (1,14)	1,42	0,98 (0,75)	1,44	1,05 (0,96)
SEC	1,44	1,17 (1,11)	1,41	0,92 (0,72)	1,42	1,04 (0,94)
TSC	1,44	1,19 (1,18)	1,41	0,93 (0,68)	1,42	1,06 (0,97)

W wyniku selekcji cech błędy walidacyjne i testowe uległy redukcji. Testy statystyczne na równość median (test Wilcoxon) błędów $MAPE_{tst}$ uzyskanych

w modelu bez selekcji i z selekcją cech w żadnym przypadku nie dały podstaw do odrzucenia hipotezy o równości median błędów testowych (przy $\alpha = 0,05$). Selekcja cech nie poprawiła więc generalizacji modelu.

Średni stopień kompresji liczby składowych wektorów wejściowych w wyniku selekcji cech wyniósł: dla SDC – 66%, dla SEC – 49%, a dla TSC – 59%. Na rys. 6 pokazano wyselekcjonowane składowe wektorów wejściowych dla poszczególnych zadań progностycznych, a na rys. 5 – częstości selekcji składowych. Najczęściej wybieranymi były składowe 18, 23 i 24, a najrzadziej – 17, 20 i 21.



Rys. 5. Częstość selekcji składowych wektorów wejściowych.

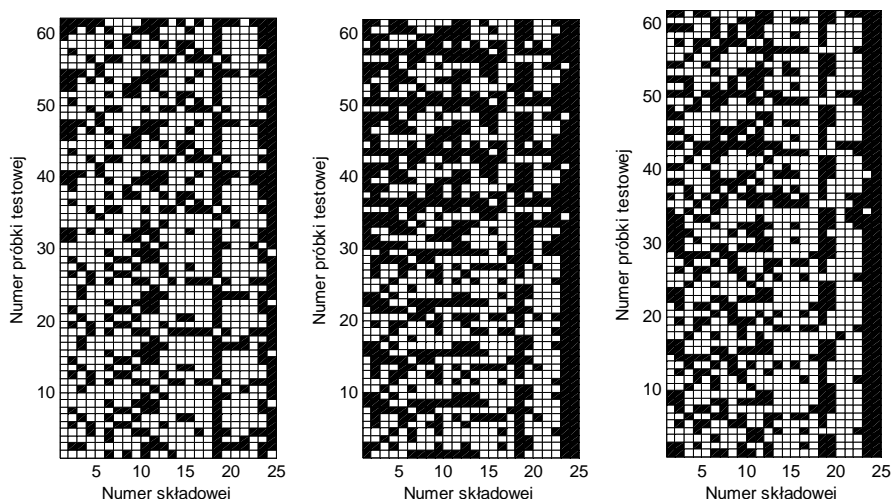
4. PODSUMOWANIE

Prezentowane w tej pracy modele progностyczne oparte na estymatorach najbliższego sąsiedztwa cechuje czytelna struktura i prostota. Liczba parametrów jest tu niewielka, co sprawia, że optymalizacja takich modeli jest stosunkowo prosta.

Parametry szerokości (k i σ) decydują o stopniu wygładzania estymatora, a co za tym idzie o jego obciążeniu i wariancji. W proponowanym podejściu parametry te optymalizuje się lokalnie w estymowanym punkcie. Pozwala to zwiększyć dokładność modelu w konkretnym zadaniu progностycznym, lecz kosztem generalizacji.

Selekcja zmiennych wejściowych, przeprowadzana również lokalnie, pozwala usunąć zmienne nadmiarowe, powielające informację wejściową, co przekłada się na podniesienie jakości modelu.

Praca naukowa finansowana ze środków na naukę w latach 2010-2012 jako projekt badawczy nr N N516 415338.



Rys. 3. Składowe wektorów wejściowych wyselekcjonowane (czarne pola) przez metody selekcji cech, od lewej: SDC, SEC i TSC.

LITERATURA

- [1] Atkenson C.G, Moor A.W., Schaal S.: Locally Weighted Learning. *Artificial Intelligence Review* 11, pp. 75-113, 1997.
- [2] Dudek G.: Krótkoterminowe prognozowanie obciążeń systemów elektroenergetycznych z zastosowaniem klasteryzacji rozmytej i algorytmów genetycznych. Sprawozdanie merytoryczne z projektu badawczego nr 3 T10B 023 29 finansowanego przez Ministerstwo Nauki i Informatyzacji (praca niepublikowana), 2006.
- [3] Dudek G.: Krótkoterminowe prognozowanie zapotrzebowania na moc elektryczną metodą klasteryzacji rozmytej. *Przegląd Elektrotechniczny*, r. 82, nr 9, str. 26-28, 2006.
- [4] Dudek G.: Tournament Searching Method to Feature Selection Problem. *Proc. 10th International Conference on Artificial Intelligence and Soft Computing ICAISC 2010, LNAI 6114*, pp. 437-444, 2010.
- [5] Härdle W.K.: *Applied Nonparametric Regression*. Cambridge University Press 1992.
- [6] Theodoridis S., Koutroumbas K.: *Pattern Recognition*. 4nd ed. Elsevier Academic Press 2009.

OPTIMIZATION OF THE FORECASTING MODELS BASED ON THE NEAREST NEIGHBOR ESTIMATORS

Key words: short-term load forecasting, nearest neighbor estimators, k -nearest neighbor method, fuzzy estimator of the regression function

Summary. The optimization methods of the nonparametric forecasting models based on the nearest neighbor estimators are presented. These models are dedicated to the short-term load forecasting. The selection methods of the model parameter values such as the number of the nearest neighbor and the weighting function parameters in the k -nearest neighbor model and the membership function width in the fuzzy regression model is described. For input variable selection the deterministic sequential methods are used as well as the stochastic tournament searching method. The combined method of the width parameter optimization and variable selection is proposed. The effectiveness of the optimization methods were verified on the application example.

Grzegorz Dudek, (ur. w 1969 r.) tytuł magistra inżyniera (1994 r.) i doktora nauk technicznych (2003 r.) w dziedzinie elektrotechniki uzyskał na Politechnice Częstochowskiej. Od 1994 r. pracuje w Instytucie Elektroenergetyki P.Cz. jako pracownik naukowo-dydaktyczny. Jego zainteresowania naukowe skupiają się wokół metod rozpoznawania obrazów i sztucznej inteligencji oraz ich zastosowań w elektroenergetyce (prognozowanie zapotrzebowania, rozdział obciążeń, problemy optymalizacyjne). Dudek@el.pcz.czyst.pl.