

Grzegorz DUDEK*

ANALIZA MODELU KRÓTKOTERMINOWEGO PROGNOZOWANIA OBCIĄŻEŃ SYSTEMÓW ELEKTROENERGETYCZNYCH OPARTEGO NA KLASTERYZACJI ROZMYTEJ**

Zaprezentowano model krótkoterminowego prognozowania obciążeń systemów elektroenergetycznych, oparty na rozmytej klasteryzacji. Opisano klasę modeli prognostycznych, opartych na podobieństwie obrazów sekwencji szeregów czasowych. Poddano analizie różne funkcje przynależności, defuzyfikacji i odległości pomiędzy obrazami w modelu rozmytym. Zbadano empirycznie wrażliwość modelu na zmianę wartości parametru, jego odporność na zaszumione i brakujące dane oraz działanie przy różnych definicjach zbiorów referencyjnych.

Słowa kluczowe: *prognozowanie krótkoterminowe obciążeń systemów elektroenergetycznych, klasteryzacja rozmyta, modele prognostyczne oparte na podobieństwie*

1. Wstęp

Krótkoterminowe prognozowanie obciążeń stanowi integralny proces sterowania i planowania pracy systemów elektroenergetycznych. Na podstawie prognoz krótkoterminowych ustala się harmonogram pracy jednostek wytwórczych, dokonuje ekonomicznego rozdziału obciążeń i alokacji rezerwy wirującej, planuje transfer energii pomiędzy systemami. Warunki rynkowe narzucają konieczność sporządzania prognoz zapotrzebowania także na spółki dystrybucyjne. Dokładne prognozy zapotrzebowania i cen energii elektrycznej pozwalają zoptymalizować kontrakty na zakup energii, oferty kierowane na giełdę energii oraz transakcje na rynku bilansującym.

* Instytut Elektroenergetyki, Wydział Elektryczny, Politechnika Częstochowska, al. Armii Krajowej 17, 42-200 Częstochowa, e-mail: Dudek@el.pcz.czyst.pl

** Praca naukowa finansowana ze środków na naukę w latach 2005–2006 jako projekt badawczy.

Wobec istotności problemu dla funkcjonowania i bezpieczeństwa pracy systemu oraz wymiernych korzyści finansowych związanych ze zwiększaniem precyzji prognoz, w ostatnich 20–30 latach opracowano wiele modeli prognozowania obciążeń. Modele klasyczne wykorzystują metody wygładzania, regresji i analizy statystycznej [14]. Nowe modele oparte są najczęściej na metodach sztucznej inteligencji [24] – sztucznych sieciach neuronowych [10], systemach rozmytych [8, 27] i systemach ekspertowych [28] lub uczeniu maszynowym i teorii rozpoznawania obrazów, np. na drzewach regresyjnych [5], czy hierarchicznych metodach klasteryzacji [12]. Tworzy się także modele hybrydowe łączące najlepsze cechy modeli składowych – systemy neuronowo-rozmyte [18], systemy ekspertowe z logiką rozmytą [29] i łączone z sieciami neuronowymi [17]. Rozwijane są także inne oryginalne podejścia: metoda rozkładu kanonicznego wektora losowego [26], metoda funkcji punktowych [22], metoda wykorzystująca geometrię fraktalną [4].

W tym artykule analizowany jest model prognozowania krótkoterminowego oparty na rozmytej klasteryzacji. Zbadano działanie modelu przyjmując definicje funkcji przynależności i defuzyfikacji, miary odległości między obrazami sekwencji szeregów czasowych obciążeń i różne definicje zbiorów referencyjnych oraz wrażliwość modelu na zmiany wartości jego parametru, dane zaszumione i brakujące.

2. Modele prognostyczne oparte na podobieństwie

Analizowana metoda, wykorzystująca klasteryzację rozmytą, jest metodą opartą na podobieństwie (*similarity-based method*). Można ją zaliczyć do klasy metod analogowych [3], w których wnioskuje się o przyszłości jednych obiektów na podstawie wiedzy o innych, podobnych obiektach lub, w szczególnym przypadku, na podstawie dotychczasowej historii tych samych obiektów. Ten drugi sposób, opisywany w tym artykule, jest uprawniony wówczas, gdy w szeregu występują wahania okresowe lub zbliżone do okresowych.

Metoda wymaga podziału szeregu czasowego na sekwencje, obejmujące zwykle jeden okres. Elementy sekwencji przetwarzają się, otrzymując jej obraz. Zakłada się, że jeśli obraz procesu \mathbf{x}_t w okresie bezpośrednio poprzedzającym prognozę wykazuje podobieństwo do obrazu \mathbf{x}_{t-i} z historii tego procesu, to obraz prognozy \mathbf{y}_{t+j} (skojarzony z obrazem \mathbf{x}_t) jest podobny do obrazu „prognozy” \mathbf{y}_{t-i+j} (skojarzonego z obrazem \mathbf{x}_{t-i}). Przez obraz rozumie się wektor o składowych będących funkcjami rzeczywistych elementów szeregu czasowego (tutaj elementami szeregu są obciążenia systemu elektroenergetycznego w kolejnych godzinach). Obrazy \mathbf{x}_t , \mathbf{x}_{t-i} i \mathbf{y}_{t-i+j} są określane na podstawie historii procesu. Pary \mathbf{x}_t i \mathbf{x}_{t-i} oraz \mathbf{y}_{t-i+j} i \mathbf{y}_{t+j} są zdefiniowane w ten sam sposób

i przesunięte w czasie o tę samą liczbę elementów szeregu i – zwykle jest to wielokrotność okresu wahań.

Sposób definicji obrazów \mathbf{x} i \mathbf{y} zależy od specyfiki szeregu czasowego (składowych periodycznych, trendu) oraz okresu i horyzontu prognozy j . Funkcje przekształcające elementy szeregu w obrazy należy tak dobrać, aby zmaksymalizować jakość modelu prognostycznego i aby obrazy przynosiły jak najwięcej informacji o procesie. Ponadto funkcje przekształcające sekwencje prognoz w obrazy powinny zapewniać możliwość obliczenia rzeczywistych prognoz elementów szeregu. Dyskusja definicji tych funkcji była przedmiotem publikacji [9].

Miary podobieństwa oparte są na miarach odległości, np. na metryce Minkowskiego (w szczególności euklidesowej), Mahalanobisa – uwzględniającej wariancję i kowariancję składowych obrazów, Hamminga – dla obrazów o składowych binarnych lub jakościowych, na miarach korelacji lub na miarach podobieństwa funkcji [3].

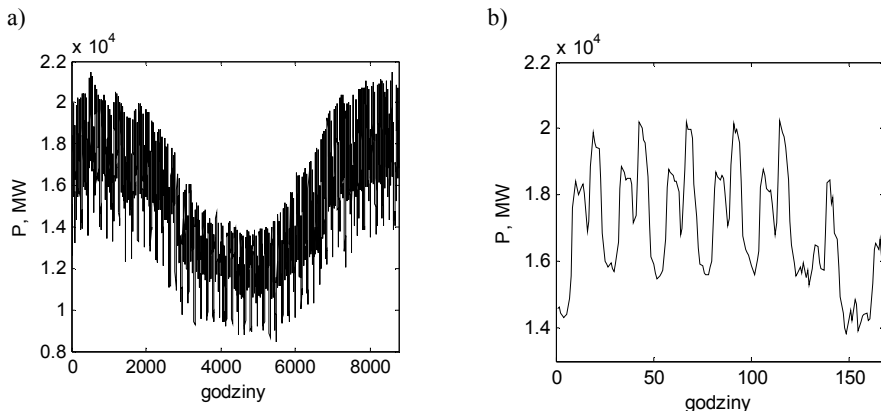
Jedna z klas modeli prognostycznych opartych na podobieństwie korzysta ze zbioru par obrazów referencyjnych (\mathbf{x}, \mathbf{y}) . Dla wektora obrazu wejściowego \mathbf{x} , odnajdywane są najbardziej podobne wektory obrazów \mathbf{x} w zbiorze referencyjnym i ze skojarzonych z nimi wektorów \mathbf{y} konstruowany jest obraz prognozy \mathbf{y}_{t+j} , z którego oblicza się rzeczywistą prognozę szeregu. Taki model zastosowano np. w pracy [23]. W innej klasie modeli wykorzystuje się tylko wektory \mathbf{x} , prognozując także parametry funkcji przekształcającej elementy szeregu w obraz [25, 20]. Model przedstawiony w [12], wykorzystujący jedynie obrazy prognoz \mathbf{y} , pozwala uniknąć prognozowania tych parametrów.

Jako obrazy referencyjne wykorzystuje się obrazy oryginalne lub definiuje się nowe obrazy przez grupowanie obrazów oryginalnych. W tym drugim przypadku stosuje się różne metody klasteryzacji, np. w [12] zastosowano metodę klasteryzacji hierarchicznej, w [1] samoorganizującą się mapę cech Kohonena, w [13] sieć z kontrpropagacją, w [6] sieć o radialnych funkcjach bazowych, a w [21] hybrydę sieci z rezonansem adaptacyjnym ART i kontrpropagacją. Grupowanie danych pozwala zredukować liczbę obrazów referencyjnych, zmniejszyć wpływ szumów, którymi obciążone są dane i poprawić generalizację. Przykładem modeli bazujących na pierwszym podejściu jest metoda \mathbf{k} najbliższych sąsiadów opisana w [9] lub prezentowany tutaj model klasteryzacji rozmytej, w którym klastery tworzone są dynamicznie wokół obrazu wejściowego [8].

3. Prognozowanie obciążeń systemów elektroenergetycznych za pomocą klasteryzacji rozmytej

Szeregi czasowe zapotrzebowania na moc elektryczną wykazują cykliczność roczną, tygodniową i dobową (rys. 1), a także trend obserwowany w okresie kilkuletnim. Różne metody analizy tych szeregów – analizę dyspersji, podobieństwa, rozkła-

dów obciążeń godzinowych, nieliniowości przebiegów dobowych i analizę harmoniczną wahań okresowych, autor przedstawił w pracy [11].



Rys. 1. Szereg czasowy obciążenia systemu krajowego w interwale rocznym (a) i tygodniowym (b)

W prognozowaniu przebiegów dobowych obciążenia systemu obrazy prognoz \mathbf{y} kodują rzeczywiste obciążenia w kolejnych godzinach doby prognozy $t + 1$: $\mathbf{P}_{t+1} = [P_{t+1,1} P_{t+1,2} \dots P_{t+1,24}]$, a obrazy wejściowe \mathbf{x} odwzorowują obciążenia poprzedzające moment prognozy – tutaj przyjęto, że są to obciążenia doby t – $\mathbf{P}_t = [P_{t,1} P_{t,2} \dots P_{t,24}]$, chociaż w ogólnym przypadku interwał ten może być dłuższy niż doba, ziarno dyskretyzacji inne niż godzina (kwadrans, pół godziny), a elementy szeregu zawarte w tym interwale mogą podlegać selekcji [7]. Wektory \mathbf{y} koduje się aktualnymi wartościami parametrów procesu (wyznaczonymi z najbliższej historii), co pozwala uwzględnić bieżącą zmienność procesu i zapewnia możliwość dekodowania. Różne definicje obrazów analizowano w [9], np. obraz $\mathbf{x}_t = [x_{t,1} x_{t,2} \dots x_{t,24}]$ o składowych będących wskaźnikami obciążeń w kolejnych godzinach doby t w stosunku do średniej mocy dobowej \bar{P}_t i podobnie określony obraz prognozy $\mathbf{y}_{t+1} = [y_{t+1,1} y_{t+1,2} \dots y_{t+1,24}]$:

$$x_{t,i} = \frac{P_{t,i}}{\bar{P}_t}, \quad i = 1, 2, \dots, 24, \quad (1)$$

$$y_{t+1,i} = \frac{P_{t+1,i}}{\bar{P}_t}, \quad i = 1, 2, \dots, 24. \quad (2)$$

Model rozmyty odwzorowuje $\mathbf{x}_t \rightarrow \mathbf{y}_{t+1}$. Po znalezieniu optymalnego modelu, określa się wyjście $\hat{\mathbf{y}}$ dla konkretnego wektora wejściowego \mathbf{x}^* , a prognozę $\hat{\mathbf{P}}$ przebiegu obciążenia doby $t+1$ wyznacza się ze wzoru definiującego obraz \mathbf{y} . W procesie optymalizacji poszukuje się postaci funkcji przynależności i jej para-

metrów (szerokości, kształtu, stopnia rozmywania), postaci funkcji defuzyfikacji, postaci funkcji odległości pomiędzy obrazami, a także postaci funkcji definiujących obrazy.

Model opiera się na rozmytej klasteryzacji. Wektorom \mathbf{x}_t odpowiadają punkty w przestrzeni n -wymiarowej. Klaster formowany jest wokół punktu wejściowego \mathbf{x}^* ; do klastra przyporządkowuje się te punkty \mathbf{x}_t ze zbioru referencyjnego, które reprezentują ten sam typ dnia $f = \{pn, wt, śr, cz, pt, sb, nd\}$ co obraz \mathbf{x}^* (w dalszej części pracy symbol \mathbf{x}_t oznacza obraz ze zbioru referencyjnego, reprezentujący ten sam typ dnia co obraz \mathbf{x}^*). To rozbitcie na typy dni spowodowane jest zróżnicowaniem charakterystyk przebiegów dobowych dla poszczególnych dni tygodnia. Przyporządkowanie do klastra nie jest ostre, binarne, lecz rozmyte. Określa się funkcję przynależności $\mu(\mathbf{x}_t, \mathbf{x}^*)$ punktów \mathbf{x}_t do klastra, zależną od ich odległości od punktu \mathbf{x}^* [15]

$$\mu(\mathbf{x}_t, \mathbf{x}^*) = \exp \left[- \left(\frac{d(\mathbf{x}_t, \mathbf{x}^*)}{\sigma} \right)^\alpha \right], \quad (3)$$

gdzie: $d(\mathbf{x}_t, \mathbf{x}^*)$ – odległość punktu \mathbf{x}_t od \mathbf{x}^* ; σ – parametr szerokości; α – parametr kształtu.

Funkcja przynależności (3) jest typu gaussowskiego z centrum w punkcie \mathbf{x}^* .

Obraz prognozy $\hat{\mathbf{y}} = [\hat{y}_1 \hat{y}_2 \dots \hat{y}_m]$ oblicza się ze wzoru (defuzyfikacja)

$$\hat{y}_i = \frac{\sum_{t=1}^N [\mu(\mathbf{x}_t, \mathbf{x}^*) y_{t+1,i}]}{\sum_{t=1}^N \mu(\mathbf{x}_t, \mathbf{x}^*)}, \quad i = 1, 2, \dots, m, \quad (4)$$

gdzie: N – liczba obrazów \mathbf{x}_t w zbiorze referencyjnym, które reprezentują ten sam typ dnia co obraz \mathbf{x}^* .

Punkty \mathbf{x}_t bliższe punktowi \mathbf{x}^* mają większy wpływ na formowanie wyjścia $\hat{\mathbf{y}}$, wyliczanego jako średnia ważona stopniami przynależności $\mu(\mathbf{x}_t, \mathbf{x}^*)$ z wektorów \mathbf{y}_{t+1} skojarzonych z \mathbf{x}_t .

Alternatywna funkcja przynależności Cauchy'ego charakteryzuje się „grubszym ogonem” w stosunku do funkcji gaussowskiej, co zapewnia większy wpływ na kształtowanie prognozy punktom bardziej odległym. Funkcję tę definiuje wzór [15]

$$\mu(\mathbf{x}_t, \mathbf{x}^*) = \left[1 + \left(\frac{d(\mathbf{x}_t, \mathbf{x}^*)}{\sigma} \right)^\alpha \right]^{-1}. \quad (5)$$

Inna popularna funkcja przynależności wyraża się wzorem [15]

$$\mu(\mathbf{x}_t, \mathbf{x}^*) = \begin{cases} 1 - \left(\frac{d(\mathbf{x}_t, \mathbf{x}^*)}{r} \right)^\alpha & \text{dla } d(\mathbf{x}_t, \mathbf{x}^*) < r, \\ 0 & \text{dla } d(\mathbf{x}_t, \mathbf{x}^*) \geq r, \end{cases} \quad (6)$$

gdzie r – parametr szerokości.

Parametr r określa promień klastrera. W formowaniu prognozy biorą w tym przypadku udział tylko obrazy prognoz skojarzone z punktami \mathbf{x}_t leżącymi wewnątrz hiperkuli o środku w \mathbf{x}^* i promieniu r . Jeśli $\alpha = 1$, to funkcja (6) jest liniowa.

Stopnie przynależności i wyjście \hat{y} można obliczać ze wzorów inspirowanych rozmytą wersją metody C średnich (*fuzzy C-means*) [2], która uchodzi za jedną z efektywniejszych metod grupowania:

$$\mu(\mathbf{x}_t, \mathbf{x}^*) = \left[\sum_{\tau=1}^N \left(\frac{d(\mathbf{x}_t, \mathbf{x}^*)}{d(\mathbf{x}_\tau, \mathbf{x}^*)} \right)^{\frac{2}{q-1}} \right]^{-1}, \quad (7)$$

$$\hat{y}_i = \frac{\sum_{t=1}^N [\mu^q(\mathbf{x}_t, \mathbf{x}^*) y_{t+1,i}]}{\sum_{t=1}^N \mu^q(\mathbf{x}_t, \mathbf{x}^*)}, \quad i = 1, 2, \dots, m, \quad (8)$$

gdzie: q – parametr określający stopień rozmywania, $q > 1$; $d(\mathbf{x}_t, \mathbf{x}^*) > 0$.

Jeśli dla M punktów ze zbioru referencyjnego zachodzi $d(\mathbf{x}_t, \mathbf{x}^*) = 0$, przyjmuje się, że stopień przynależności tych punktów jest równy $1/M$, natomiast pozostałych jest równy 0. Funkcja przynależności (7) jest zależna nie tylko od odległości punktu \mathbf{x}_t od \mathbf{x}^* , ale także od odległości pozostałych punktów \mathbf{x}_τ w zbiorze referencyjnym od \mathbf{x}^* i od liczności tego zbioru N . Suma obliczanych z funkcji (7) stopni przynależności punktów referencyjnych jest równa 1, niezależnie od parametru q . Parametr q podlega optymalizacji.

Funkcje przynależności oznaczono symbolami: F1 – (3), F2 – (5), F3 – (6), F4 – (7).

Przykładowe wykresy funkcji przynależności przedstawiono na rysunku 2.

4. Dobór optymalnej funkcji przynależności

Wyboru optymalnej funkcji przynależności oraz jej parametrów dokonuje się minimalizując błąd prognozy wygasłej (*MAPE* obliczany na zbiorze uczącym), którego

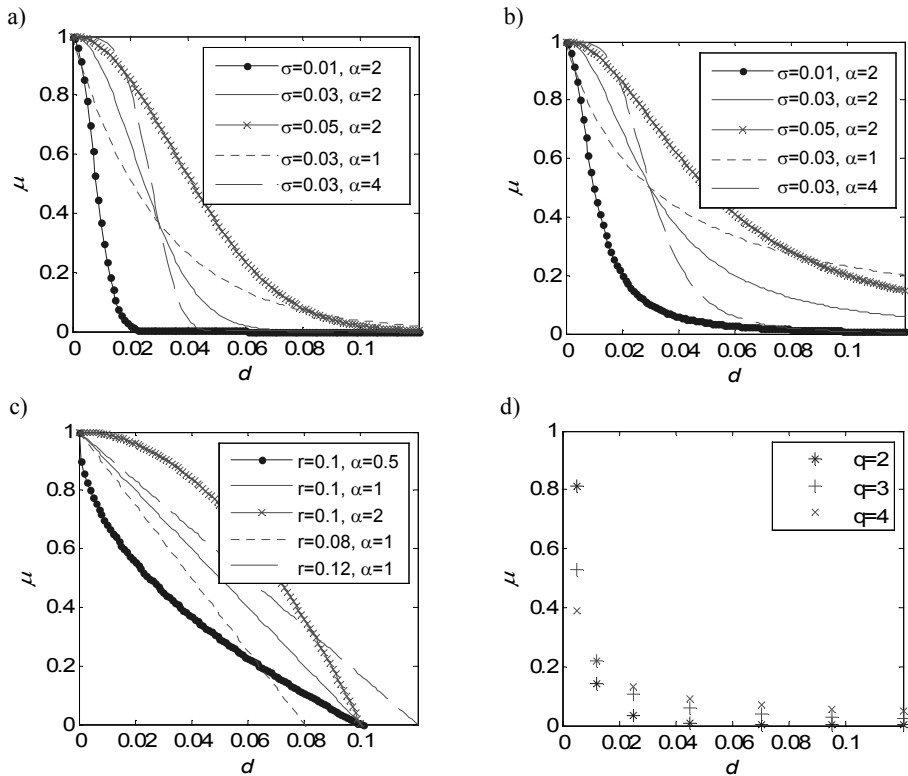
ogólna postać wyrażona jest wzorem

$$\Psi(f_\mu, f_d, f_o, f_x, f_y, \Pi) = \frac{1}{Nm} \sum_{t=1}^N \sum_{i=1}^m |\psi_{t+1,i}(f_\mu, f_d, f_o, f_x, f_y, \Pi)|, \quad (9)$$

gdzie

$$\psi_{t+1,i}(f_\mu, f_d, f_o, f_x, f_y, \Pi) = \frac{P_{t+1,i} - \hat{P}_{t+1,i}(f_\mu, f_d, f_o, f_x, f_y, \Pi)}{P_{t+1,i}} \cdot 100, \quad (10)$$

$f_\mu, f_d, f_o, f_x, f_y$ – postaci funkcji, odpowiednio, przynależności, defuzyfikacji, odległości pomiędzy obrazami \mathbf{x} , definiującej obrazu \mathbf{x} , definiującej obrazu \mathbf{y} ; Π – zbiór parametrów wymienionych funkcji: szerokość σ lub r , parametr kształtu α , stopień rozmywania q funkcji przynależności, parametry funkcji definiujących obrazu.



Rys. 2. Przykładowe wykresy funkcji przynależności:
 a) F1, b) F2, c) F3, d) F4

Badania eksperymentalne wykonano na siedmiu zbiorach danych [9] – sześć zbiorów obejmowało godzinowe przebiegi obciążeń systemów elektroenergetycznych o różnym stopniu regularności (krajowy system elektroenergetyczny i systemy lokalne), jeden zbiór obejmował godzinowy przebieg zapotrzebowania na moc dużego zakładu przemysłowego – oraz, kontrolnie, na trzech szeregach czasowych cen energii na rynku bilansującym (cena rozliczeniowa odchyłeń CRO, zakupu CROz i sprzedaży CROs). Oryginalne dane poddano przekształceniom, otrzymując zbiory ich obrazów \mathbf{x} i \mathbf{y} (obrazy również podlegały optymalizacji – szczegóły w [9]). Przyjęto, że zbiór uczący (do estymacji parametrów modelu) obejmuje 2/3 elementów szeregów chronologicznie najstarszych, a zbiór testowy (do weryfikacji modelu) – 1/3 elementów szeregu chronologicznie najnowszych. Błędy (9) obliczane dla zbioru uczącego oznaczono w dalszej części pracy indeksem *trn*, a błędy obliczane do zbioru testowego – indeksem *tst*. W tabelach i na wykresach zestawiono szczegółowe wyniki analiz dla jednego zbioru danych – godzinowych obciążeń krajowego systemu elektroenergetycznego z lat 1997–2000 (KSE). Elementy tego zbioru zostały przetworzone za pomocą wzorów (1) i (2).

Procedura testowa polega na obliczeniu prognozy wygasłej dla każdej pary $(\mathbf{x}_t, \mathbf{y}_{t+1})$ ze zbioru uczącego, biorąc za zbiór referencyjny pozostałe pary $(\mathbf{x}_t, \mathbf{y}_{t+1})$ z tego zbioru (procedura *leave-one-out*). Optymalizację wykonywano dla różnych funkcji przynależności i dla każdego typu dnia tygodnia indywidualnie. Odległość pomiędzy obrazami $d(\mathbf{x}_t, \mathbf{x}^*)$ wyznaczano w metryce euklidesowej. Do optymalizacji zastosowano metodę sympleksu Nelder–Meada [20], weryfikując w niektórych przypadkach rezultaty metodą quasi-newtonowską. Aby zwiększyć prawdopodobieństwo znajdowania minimów globalnych, uruchamiano algorytm z różnych punktów startowych. Wyniki optymalizacji – błędy (9) na zbiorze uczącym Ψ_{trn} i testowym Ψ_{tst} oraz optymalne parametry funkcji przynależności – zestawiono w tabeli 1.

Najlepsze rezultaty osiągnięto dla funkcji przynależności F1 i F2. W niektórych przypadkach dla funkcji F3 nie można było określić błędów na zbiorze testowym, ponieważ promień r nie obejmował żadnego obrazu \mathbf{x}_t . Przyjęcie stałej wartości wykładnika ($\alpha = 2$) w funkcji F1 i optymalizacja jednego parametru σ spowodowały tylko nieznaczne pogorszenie rezultatów – średnio (dla wszystkich analizowanych szeregów) o ok. 2% na zbiorze uczącym i 0,6% na zbiorze testowym. Stronne funkcje przynależności (mniejsze σ) opisują klaster o mniejszych średnicach, redukując wpływ na prognozę obrazów dalej położonych od centrum klastra \mathbf{x}^* .

Lepsze rezultaty może przynieść optymalizacja szerokości funkcji przynależności σ dla każdego typu dnia tygodnia i dla każdej godziny prognozy $i = 1, 2, \dots, 24$ indywidualnie. Jednak błędy dla tak optymalizowanych modeli (z funkcjami przynależności F1 i $\alpha = 2$) utrzymywały się na podobnym poziomie, jak w modelu z jednakowymi szerokościami σ dla wszystkich godzin [7].

Ostatecznie model z funkcją przynależności F1 i wartością wykładnika $\alpha = 2$ oraz z szerokościami optymalizowanymi dla każdego typu dnia tygodnia (bez optymalizacji dla poszczególnych godzin prognozy) uznaje się za wystarczająco dokładny. Model taki zastosowano w dalszej części pracy.

5. Działanie modelu w zależności od różnych definicji odległości

Obrazy \mathbf{x} są traktowane jako punkty w przestrzeni n -wymiarowej. Do określania odległości pomiędzy punktami stosowano metrykę:

- euklidesową

$$d(\mathbf{x}_t, \mathbf{x}^*) = \sqrt{\sum_{i=1}^n (x_i^* - x_{t,i})^2} = [(\mathbf{x}^* - \mathbf{x}_t)^T (\mathbf{x}^* - \mathbf{x}_t)]^{0,5}, \quad (11)$$

- miejską (*city, Manhattan*)

$$d(\mathbf{x}_t, \mathbf{x}^*) = \sum_{i=1}^n |x_i^* - x_{t,i}|, \quad (12)$$

- Mahalanobisa

$$d(\mathbf{x}_t, \mathbf{x}^*) = [(\mathbf{x}^* - \mathbf{x}_t)^T \Sigma^{-1} (\mathbf{x}^* - \mathbf{x}_t)]^{0,5}, \quad (13)$$

gdzie Σ jest macierzą kowariancji rozkładu wektora losowego \mathbf{x} ,

- korelacyjną

$$d(\mathbf{x}_t, \mathbf{x}^*) = 0,5(1 - \rho), \quad (14)$$

gdzie ρ jest współczynnikiem korelacji liniowej pomiędzy wektorami \mathbf{x}_t i \mathbf{x}^* ,

- kosinusową

$$d(\mathbf{x}_t, \mathbf{x}^*) = 0,5 \left(1 - \frac{\mathbf{x}^{*T} \mathbf{x}_t}{[(\mathbf{x}^{*T} \mathbf{x}^*)(\mathbf{x}_t^T \mathbf{x}_t)]^{0,5}} \right), \quad (15)$$

gdzie drugi składnik jest kosinusem kąta pomiędzy dwoma wektorami, których początki znajdują się w początku układu współrzędnych, a końcami są punkty odpowiadające obrazom \mathbf{x}_t i \mathbf{x}^* .

Odległość euklidesowa i miejska oparte są na normie Minkowskiego. Miary te nie są niezmiennicze względem skali wartości zmiennych [16]. Zmiana skali powoduje

zmianę odległości pomiędzy obserwacjami (zaleca się wcześniejszą normalizację obserwacji lub odpowiednią definicję obrazu). Wady tej pozbawiona jest odległość Mahalanobisa, która pozwala uwzględniać wariancje i kowariancje zmiennych, wchodzących w skład wektora losowego \mathbf{x} .

Odległość korelacyjna (14) jest zerowa przy maksymalnym dodatnim skorelowaniu wektorów obrazów i największa, równa jedności, przy maksymalnym ujemnym skorelowaniu. W przypadku braku korelacji ($\rho = 0$) odległość wynosi 0,5.

Miara odległości kosinusowej opiera się na kosinusie kąta pomiędzy wektorami. Odległość ta wynosi zero, jeśli wektory pokrywają się i jeden, jeśli są skierowane w przeciwne strony (kątem pomiędzy nimi wynosi 2π).

Model z funkcją przynależności F1 ($\alpha = 2$) przetestowano przy różnych miarach odległości. Dla każdego przypadku optymalizowano szerokość funkcji przynależności. Wyniki zestawiono w tabeli 2.

Tabela 2. Optymalne dla zbioru uczącego szerokości funkcji przynależności oraz błędy prognoz przy różnych miarach odległości, wyznaczone na zbiorze uczącym i testowym KSE

Dzień prognozy	Miara odległości														
	euklidesowa			miejska			Mahalanobisa			korelacyjna			kosinusowa		
	σ	Ψ_{trn}	Ψ_{tst}	σ	Ψ_{trn}	Ψ_{tst}	σ	Ψ_{trn}	Ψ_{tst}	σ	Ψ_{trn}	Ψ_{tst}	σ	Ψ_{trn}	Ψ_{tst}
pn	0,0487	1,79	2,87	1,74E-01	1,80	2,88	1,71	2,50	3,84	6,38E-03	1,87	3,06	4,97E-05	1,81	2,91
wt	0,0578	1,43	2,11	2,19E-01	1,46	1,84	1,70	1,97	2,76	3,13E-03	1,47	1,90	7,44E-05	1,42	2,10
śr	0,0407	1,62	2,04	1,75E-01	1,68	1,96	1,76	1,81	2,31	3,69E-03	1,60	1,83	3,89E-05	1,60	2,03
cz	0,0658	2,35	2,19	2,51E-01	2,26	2,23	1,64	2,65	2,51	4,81E-03	2,29	2,20	8,25E-05	2,38	2,19
pt	0,0601	2,13	1,78	2,28E-01	2,11	1,78	1,77	2,38	2,13	6,63E-03	2,22	1,88	7,47E-05	2,12	1,78
sb	0,0479	2,05	2,06	1,86E-01	2,04	2,07	1,70	2,23	2,73	5,38E-03	2,07	2,07	5,25E-05	2,05	2,11
nd	0,0512	1,63	1,95	2,01E-01	1,64	1,97	2,64	2,79	2,89	9,25E-03	1,73	1,98	6,84E-05	1,65	1,96
Błędy średnie	1,86	2,14			1,86	2,11		2,33	2,74		1,89	2,13		1,86	2,15

Najlepsze rezultaty, zarówno na zbiorze treningowym, jak i testowym KSE, także w większości przypadków dla innych zbiorów danych, osiągnięto wykorzystując odległość miejską. Nieznacznie gorzej działał model bazujący na odległości euklidesowej.

W dalszej części pracy przyjmuje się metrykę euklidesową do określania odległości pomiędzy obrazami.

6. Wrażliwość modelu na szerokość funkcji przynależności

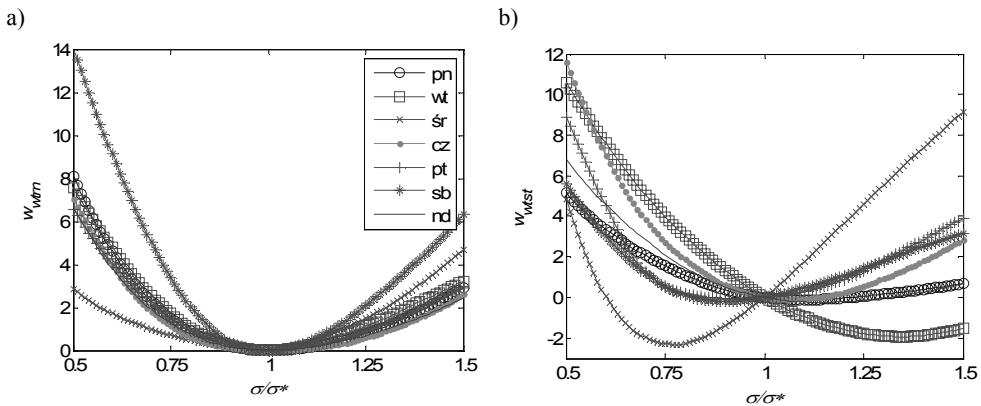
Celem tej analizy jest określenie wpływu szerokości σ funkcji przynależności F1 na błędy prognoz. Definiuje się procentowy wskaźnik wrażliwości, wyrażający

względną różnicę pomiędzy błędem otrzymanym przy szerokości funkcji przynależności σ , a błędem otrzymanym przy szerokości optymalnej σ^* (wyznaczonej na zbiorze uczącym):

$$w_w(\sigma) = \frac{\Psi(\sigma) - \Psi(\sigma^*)}{\Psi(\sigma^*)} \cdot 100, \quad (16)$$

gdzie: σ^* – optymalna wartość szerokości funkcji przynależności dla zbioru uczącego (z tab. 1); $\Psi(\sigma)$, $\Psi(\sigma^*)$ – błąd (9) obliczony przy szerokości funkcji przynależności równej, odpowiednio, σ lub σ^* (dalej błędy te wyznacza się dla zbioru uczącego $\Psi_{trn}(\cdot)$ i testowego $\Psi_{lst}(\cdot)$).

Przebiegi wskaźników w_w , wyznaczonych dla zbioru uczącego ($w_{w\ trn}$) i testowego ($w_{w\ lst}$) w zakresie $0,5\sigma^* \leq \sigma \leq 1,5\sigma^*$, pokazano na rysunku 3.



Rys. 3. Przebiegi wskaźnika wrażliwości w zależności od szerokości funkcji przynależności σ dla zbioru uczącego (a) i testowego (b) KSE

W tabeli 3 zamieszczono średnie wartości modułów wskaźników wrażliwości obliczanych dla zbioru uczącego i testowego przy 20-procentowej zmianie szerokości funkcji przynależności w stosunku do szerokości optymalnej σ^* :

$$w_{w20\%} = \max(|w_w(0,8\sigma^*)|, |w_w(1,2\sigma^*)|). \quad (17)$$

Współzależność wskaźników $w_{w\ trn}(\sigma)$ i $w_{w\ lst}(\sigma)$ zbadano obliczając współczynniki korelacji liniowej Pearsona ρ_w . Wartości współczynników zestawiono w tabeli 3.

W wielu przypadkach minima błędów dla zbioru uczącego i testowego nie pokrywają się. Rozbieżności między optymalną szerokością wyznaczoną na zbiorze uczącym σ^* , a optymalną szerokością wyznaczoną na zbiorze testowym σ^*_{lst} określa procentowy wskaźnik rozbieżności optimów:

$$w_r = \frac{\sigma^*_{tst} - \sigma^*}{\sigma^*} \cdot 100. \tag{18}$$

Procentową względną różnicę między minimalnym błędem dla zbioru testowego $\Psi_{tst}(\sigma^*_{tst})$, a błędem otrzymanym z modelu zoptymalizowanego na zbiorze uczącym $\Psi_{tst}(\sigma^*)$, wyraża wskaźnik $w_{w_{tst}}(\sigma^*_{tst})$

$$w_{w_{tst}}(\sigma^*_{tst}) = \frac{\Psi_{tst}(\sigma^*_{tst}) - \Psi_{tst}(\sigma^*)}{\Psi_{tst}(\sigma^*)} \cdot 100. \tag{19}$$

Tabela 3. Średnie wartości wskaźników opisujących wrażliwość modelu na zmianę szerokości funkcji przynależności dla zbioru KSE

Dzień prognozy	$w_{w_{trn20\%}}$	$w_{w_{tst20\%}}$	w_r	$w_{w_{tst}}(\sigma^*_{tst})$	ρ_w
	%	%	%	%	
pn	0,77	1,45	14	-0,24	0,91
wt	0,89	1,09	11	-0,13	0,93
śr	1,03	3,01	34	-1,93	0,73
cz	1,09	3,66	-22	-2,32	0,83
pt	0,64	1,86	6	-0,10	0,98
sb	0,93	0,90	-8	-0,05	0,96
nd	2,05	1,07	-10	-0,18	0,92

Zmiana szerokości funkcji przynależności σ o 20% powoduje nieznaczną zmianę błędu prognozy, średnio (dla wszystkich zbiorów danych) o ok. 1% na zbiorze uczącym i 2,3% na zbiorze testowym. Nie jest to więc parametr krytyczny.

Rozbieżność między optymalną wartością szerokości σ wyznaczoną na zbiorze uczącym i testowym jest w niektórych przypadkach znaczna, przekraczająca 50%. Rozbieżność ta skutkuje jednak nieznacznym tylko wzrostem błędów na zbiorach testowych – średnio o ok. 1,8%, co wynika z płaskiego przebiegu funkcji błędu w pobliżu ekstremum. W niektórych przypadkach zauważalna jest ujemna korelacja między wskaźnikami wrażliwości na zbiorze uczącym i testowym.

Z wykresów zamieszczonych na rysunku 3 wynika, że większe błędy generuje model z niedoszacowanymi wartościami parametrów σ – o prognozie decydują wtedy obrazy położone blisko obrazu wejściowego (być może zakłócone szumami), wpływ obrazów położonych dalej jest tłumiony. Przeszacowanie skutkuje zwiększeniem wpływu obrazów dalej położonych, co w skrajnym przypadku prowadzi do konstrukcji prognozy jako średniej ze wszystkich obrazów referencyjnych y_{t+1} .

7. Odporność modelu na dane zaszumione

Celem tej analizy jest określenie odporności modelu na szумы w danych wejściowych, wynikające np. z błędów pomiarów. Przyjmuje się, że składowe wektora wejściowego, którego obrazem jest \mathbf{x}^* zakłócone są szumem

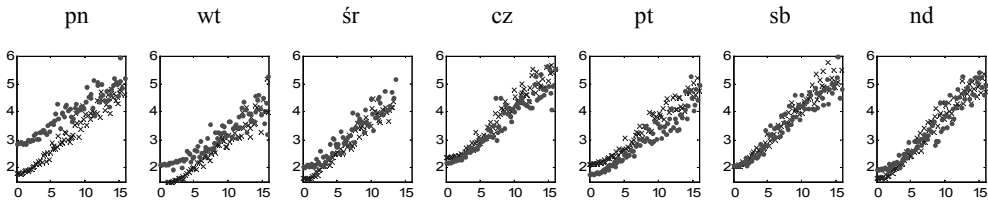
$$P^*_{t,i} = P^*_{t,i} \xi_{t,i}, \quad (20)$$

gdzie $\xi_{t,i}$ – zmienna losowa o wartości oczekiwanej równej 1 i skończonej wariancji.

Zaszumiony obraz \mathbf{x}^* lokuje się w innym miejscu niż obraz oryginalny \mathbf{x}^* , w związku z czym zmienia się środek klastra oraz stopnie przynależności obrazów \mathbf{x}_t ze zbioru referencyjnego $\mu(\mathbf{x}_t, \mathbf{x}^*)$ i w efekcie obraz prognozy $\hat{\mathbf{y}}$.

Odchylenie standardowe rozkładu normalnego, z którego generowany był szum, zmieniano w granicach od 0 do 0,2, odnotowując średnie błędy prognoz. Na rysunku 4 pokazano te błędy w funkcji procentowego udziału szumów w danych, zdefiniowanego następująco:

$$u_\xi = \frac{100}{Nm} \sum_{t=1}^N \sum_{i=1}^m \left| \frac{P^*_{t,i} - P^*_{t,i}}{P^*_{t,i}} \right| = \frac{100}{Nm} \sum_{t=1}^N \sum_{i=1}^m |1 - \xi_{t,i}|. \quad (21)$$



Rys. 4. Błędy prognozy Ψ_ξ (oś Oy) w zależności od procentowego udziału szumów w danych u_ξ (oś Ox) dla zbioru KSE. Krzyżyki – błąd na zbiorze treningowym, kropki – błąd na zbiorze testowym

Jako syntetyczną miarę wrażliwości modelu na szum proponuje się wskaźnik

$$w_\xi = \frac{\Psi_\xi - \Psi}{u_\xi}, \quad (22)$$

gdzie:

Ψ_ξ – średni błąd prognozy obserwowany przy procentowym udziale szumu w danych u_ξ

Ψ – średni błąd prognozy bez szumu w danych.

Wskaźnik ten wyraża stosunek zmiany błędu prognozy, spowodowany szumem w danych do poziomu tego szumu. Zerowa wartość w_ξ oznacza brak wrażliwości modelu na szum, większe wartości oznaczają większą wrażliwość. Wskaźniki w_ξ odpowiadają współczynnikom kierunkowym prostych aproksymujących charakterystyki

przedstawione na rysunku 4. Średnie wartości wskaźników wrażliwości wyznaczone na zbiorze treningowym i testowym dla danych KSE wyniosły: $w_{\xi\ trn} = 0,153$, $w_{\xi\ tst} = 0,140$. Średni procentowy wzrost błędów δ_{ψ} dla $u_{\xi} = 10\%$: $\delta_{\psi\ trn} = 117\%$, $\delta_{\psi\ tst} = 90\%$.

Wrażliwość modelu zależy od regularności danych. Dla danych nieregularnych wrażliwość modelu na szum jest niewielka (w przypadku szeregów cen energii wartości wskaźników w_{ξ} zawierały się w przedziale 0,004–0,020). Wynika to z faktu, że przy słabych zależnościach między podobieństwami obrazów \mathbf{x} i obrazów \mathbf{y} , funkcje przynależności są szerokie i obejmują wiele punktów danych, którym przydzielają porównywalne stopnie przynależności do klastra. Przy zmianie lokalizacji centrum klastra spowodowanej szumem, stopnie przynależności nieznacznie tylko się zmieniają, co ma niewielki wpływ na prognozę. Inaczej w danych, które cechuje większy stopień podobieństwa obrazów i silniejsza zależność między obrazami \mathbf{x} i \mathbf{y} . Dane takie modelowane są dokładniej, wąskimi funkcjami przynależności, których przesunięcie spowodowane szumem zmienia ich obszar recepcyjny, przydzielając inne stopnie przynależności punktom danych. Prognozy obliczone z obrazów prognoz skojarzonych z tymi punktami są obciążone większymi błędami.

8. Odporność modelu na brakujące dane

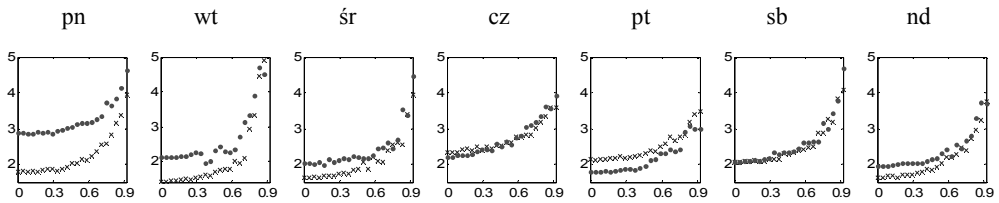
Analizuje się odporność modelu na niekompletną informację wejściową – wektor \mathbf{P}^* , z którego konstruuje się obraz \mathbf{x}^* , ma nieokreślone wartości l składowych. W takim przypadku wartości te są też nieokreślone w obrazach \mathbf{x}^* , ponadto wartości pozostałych składowych obrazów \mathbf{x}^* mogą się różnić od wartości składowych obrazów przy kompletnej informacji wejściowej. Zachodzi to w tych przypadkach, kiedy obraz definiowany jest za pomocą charakterystyki obliczanej ze wszystkich składowych, np. średniej (jak (1) i (2)) lub odchylenia standardowego [9]. W przypadku braku l składowych te wartości oblicza się z pozostałych składowych.

Z powodu braku l składowych odległość $d(\mathbf{x}_t, \mathbf{x}^*)$ wyznacza się z pominięciem tych składowych. Ma to wpływ na rozkład odległości pomiędzy obrazem wejściowym a obrazami ze zbioru referencyjnego (niezmienionymi), a co za tym idzie – na wartości stopni przynależności i w efekcie na obraz prognozy $\hat{\mathbf{y}}$.

Eksperyment polega na usuwaniu składowych z wektorów \mathbf{P}^* i obliczaniu błędów prognoz w takich przypadkach. Składowe w liczbie $l = 1, 2, \dots$ wybierane są do usunięcia w sposób losowy, niezależny dla każdego wektora \mathbf{P}^* . Wyniki – błędy prognoz w zależności od wskaźnika kompletności danych zdefiniowanego następująco:

$$u_l = \frac{l}{n}, \quad (23)$$

gdzie $n = 24$ i $l = 1, 2, \dots, 22$, dla zbioru KSE przedstawiono na rysunku 5.



Rys. 5. Błędy prognozy Ψ_l (oś Oy) w zależności od wskaźnika kompletności danych u_l (oś Ox) dla zbioru KSE. Krzyżyki – błąd na zbiorze treningowym, kropki – błąd na zbiorze testowym

Jako miarę wrażliwości modelu na brakujące dane proponuje się wskaźnik

$$w_l = \frac{\Psi_l - \Psi}{u_l}, \quad (24)$$

gdzie:

Ψ_l – średni błąd prognozy obserwowany przy wskaźniku kompletności u_l ,

Ψ – średni błąd prognozy przy kompletnych danych.

Wartości średnie tych wskaźników dla zbioru KSE wyniosły: $w_{l_{trn}} = 0,52$, $w_{l_{tst}} = 0,54$. Średni procentowy przyrost błędów δ_Ψ przy wektorze wejściowym niekompletnym w 50% ($l = 12$) do błędów przy wektorze kompletnym: $\delta_{\Psi_{trn}} = 15,2\%$, $\delta_{\Psi_{tst}} = 11,6\%$. Dla szeregów czasowych cen przyrost błędów był znacznie mniejszy – od 0,8% do 8,4%.

Błędy wzrastają wraz z liczbą brakujących składowych w sposób nieliniowy. Przy małych l obserwuje się tylko nieznaczny wzrost błędów.

W opisanym eksperymencie obrazy referencyjne \mathbf{x} powstały z kompletnych wektorów \mathbf{P} . Jeśli obrazy referencyjne \mathbf{x} powstają z wektorów \mathbf{P} , z których usuwa się te same składowe, które są nieobecne w \mathbf{P}^* , to wartości wskaźników dla zbioru KSE kształtują się następująco: $w_{l_{trn}} = 0,32$, $w_{l_{tst}} = 0,27$, $\delta_{\Psi_{trn}} = 9,8\%$, $\delta_{\Psi_{tst}} = 5,5\%$. Przyrost błędów δ_Ψ dla pozostałych zbiorów danych również, w większości przypadków, znacznie się zmniejszył. Oznacza to, że taka procedura konstrukcji prognozy (usuwanie brakujących składowych \mathbf{P}^* , także z wektorów referencyjnych \mathbf{P}) jest mniej wrażliwa na brakujące składowe od procedury opisaną wcześniej.

9. Działanie modelu przy różnych definicjach zbiorów referencyjnych

Umownie przyjęto, że zbiór referencyjny stanowi 2/3 początkowych elementów szeregu czasowego. W praktyce przy prognozie sekwencji szeregu dla doby $t+1$ dostępny jest szereg o długości aż do doby t . Przyjmując ten szereg za zbiór referencyj-

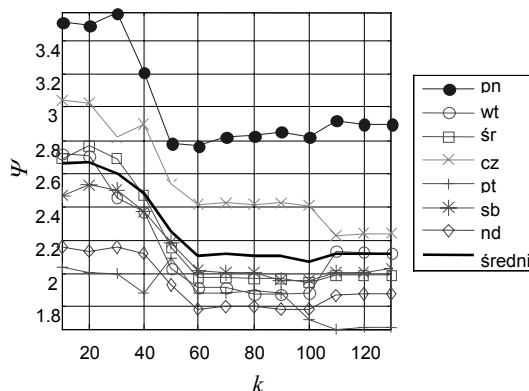
ny, maksymalnie wykorzystujemy dostępne informacje. Błędy prognoz dla zbioru testowego (1/3 końcowych elementów szeregu) i tak zdefiniowanego zbioru referencyjnego przedstawiono w tabeli 4. Symbol w w tabeli oznacza względną procentową redukcję błędu w stosunku do wariantu, w którym zbiór referencyjny stanowi 2/3 początkowych elementów szeregu czasowego.

Tabela 4. Błędy na zbiorze testowym KSE przy rozszerzonym zbiorze referencyjnym oraz procentowa względna poprawa rezultatu (w) w stosunku do wariantu ze zbiorem referencyjnym obejmującym 2/3 początkowych elementów szeregu

Typ dnia	pn	wt	śr	cz	pt	sb	nd	Wartości średnie
Ψ_{est}	2,81	2,12	2,01	2,22	1,68	2,00	1,92	2,11
w	1,99	-0,61	1,23	-1,43	5,52	3,14	1,77	1,44

Dla pozostałych zbiorów danych, poza jednym przypadkiem, również obserwuje się redukcje błędów prognoz.

Można postawić hipotezę, że najwięcej informacji o sekwencji szeregu dla doby $t + 1$ zawierają elementy tego szeregu z najbliższej historii procesu. Elementy dalsze (np. przesunięte o rok wstecz) mogą nawet zakłócać prognozę, jeśli w analizowanym procesie zachodzą jakieś zmiany i zakłócenia deformujące relacje $x \rightarrow y$. W takim przypadku zbiór referencyjny powinien obejmować elementy z najbliższej historii. Hipotezę tę zweryfikowano eksperymentem. Przyjmowano zbiory referencyjne obejmujące $k = 10, 20, 30, \dots$ najbliższych sekwencji szeregu, poprzedzających prognozowaną sekwencję $t+1$. Na przykład dla prognoz zapotrzebowania w poniedziałek i $k = 10$ zbiór referencyjny obejmował obrazy kolejnych 10 niedziel i skojarzonych z nimi poniedziałków, bezpośrednio poprzedzających „prognozowany” poniedziałek. Błędy prognoz wyznaczone na zbiorze testowym dla różnych k zobrazowano na wykresie (rys. 6).



Rys. 6. Błędy prognoz na zbiorze testowym dla KSE w zależności od k

Obserwuje się zmniejszenie błędu wraz ze wzrostem rozmiaru zbioru referencyjnego k . Dotyczy to także pozostałych zbiorów danych z jednym wyjątkiem. Przeczy to postawionej hipotezie i oznacza, że w analizowanych szeregach czasowych relacja $x \rightarrow y$ jest w miarę stabilna.

Podsumowanie

W przedstawionym modelu prognostycznym opartym na podobieństwie, o przebiegu szeregu czasowego w przyszłości, wnioskuje się na podstawie jego zachowania w przeszłości, zakładając, że obrazy sekwencji szeregu następujące po obrazach sekwencji podobnych do obrazu sekwencji, poprzedzającej prognozę, będą podobne do obrazu sekwencji prognozowanej. Jeśli hipotezę, że zależność między podobieństwem obrazów x , a podobieństwem skojarzonych z nimi obrazów y nie wynika z losowego charakteru próby można potwierdzić przy wysokim poziomie ufności, uzasadnia to sens konstrukcji i stosowania modeli prognostycznych opartych na podobieństwie obrazów sekwencji szeregu czasowego. Metodę weryfikacji tej hipotezy, opartą m.in. na teście chi-kwadrat, przedstawiono w pracy [7].

Stopień podobieństwa i siłę relacji między sekwencjami można w pewnym stopniu kształtować za pomocą definicji obrazów tych sekwencji, definicji funkcji przynależności i defuzyfikacji oraz definicji miar odległości pomiędzy obrazami.

Na podstawie przeprowadzonych doświadczeń można sformułować następujące wnioski:

1. Kształt funkcji przynależności wydaje się nie mieć istotnego wpływu na działanie modelu, pod warunkiem jednak, że funkcja ta jest monotonicznie malejąca wraz ze wzrostem odległości pomiędzy obrazami i posiada nieograniczony nośnik.

2. Popularne miary odległości pomiędzy obrazami – miejska i euklidesowa zapewniają najlepsze działanie modelu.

3. Model oparty na klasteryzacji rozmytej, w wersji przyjętej za dostatecznie dokładną, ma tylko jeden parametr – szerokość funkcji przynależności. Parametr ten jest łatwy do estymacji, a wrażliwość modelu na jego odchyłki od wartości optymalnej, szczególnie w górę, jest ograniczona.

4. Wrażliwość modelu na szумы w danych zależy od szerokości funkcji przynależności. Wrażliwość jest większa w przypadku danych wykazujących duży stopień regularności i silne związki pomiędzy obrazami x i y , modelowanych wąskimi funkcjami przynależności.

5. Model nie wymaga szacowania wartości brakujących składowych wejściowych wektorów obciążeń. Wpływ brakujących składowych na działanie modelu jest niewielki, gdy ich liczba jest ograniczona do ok. 30% i uwydatnia się wraz ze wzrostem

tej liczby. Przy konstrukcji obrazów referencyjnych należy wyeliminować te składowe, które są nieobecne w wektorze wejściowym.

6. Jeśli relacja między obrazem wejściowym a obrazem prognozy nie zmienia się znacząco w czasie, zbiór referencyjny powinien obejmować obrazy z całej dostępnej historii procesu, aż do chwili prognozy.

Bibliografia

- [1] BAUMANN T., GERMOND A., *Application of the Kohonen network to short-term load forecasting*, Proceedings of the Second International Forum on Application of Neural Networks to Power Systems, 1993, s. 407–412.
- [2] BEZDEK J. C., *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New York 1981.
- [3] CIEŚLAK M. (red.), *Nieklasykne metody prognozowania*, PWN, Warszawa 1983.
- [4] DOBRZAŃSKA I., *Wymiar Hausdorffa jako narzędzie predykcji krokowej*, *Badania Operacyjne i Decyzje*, 1991, nr 2, s. 31–62.
- [5] DUDEK G., *Drzewo regresyjne jako narzędzie prognostyczne*, *Mat. konf. „Prognozowanie w Elektroenergetyce PE’2004”*, Częstochowa 2004, s. 99–105.
- [6] DUDEK G., *Krótkoterminowe prognozowanie obciążenia systemu elektroenergetycznego z zastosowaniem sieci neuronowych o radialnych funkcjach bazowych*, *Mat. konf. „Prognozowanie w Elektroenergetyce PE’2000”*, 2000, s. 59–68.
- [7] DUDEK G., *Krótkoterminowe prognozowanie obciążeń systemów elektroenergetycznych z zastosowaniem klasteryzacji rozmytej i algorytmów genetycznych*, *Sprawozdanie merytoryczne z projektu badawczego nr 3 T10B 023 29*. Politechnika Częstochowska, Częstochowa 2006 (praca niepublikowana).
- [8] DUDEK G., *Krótkoterminowe prognozowanie zapotrzebowania na moc elektryczną metodą klasteryzacji rozmytej*, *Przegląd Elektrotechniczny*, 2006, r. 82, nr 9, s. 26–28.
- [9] DUDEK G., *Przetwarzanie danych w oparciu na podobieństwie metodach prognozowania przebiegów dobowych zapotrzebowania na moc elektryczną*, *Przegląd Elektrotechniczny*, 2006, r. 82, nr 9, s. 15–19.
- [10] DUDEK G., *Sporządzanie dobowych prognoz obciążenia systemu elektroenergetycznego przy użyciu sieci neuronowych – problemy wyboru struktury i uczenia sieci*, *Mat. konf. „Algorytmy Ewolucyjne i Optymalizacja Globalna”*, Łądek Zdrój 2000, s. 59–66.
- [11] DUDEK G., *Wybrane metody analizy szeregów czasowych obciążeń elektroenergetycznych*, *Mat. konf. „Prognozowanie w Elektroenergetyce PE’2004”*, 2004, s. 116–125.
- [12] DUDEK G., *Zastosowanie hierarchicznych metod grupowania do prognozowania dobowych charakterystyk obciążeń elektroenergetycznych*, *Przegląd Elektrotechniczny*, 2006, r. 82, nr 9, s. 9–11.
- [13] DUDEK G., *Zastosowanie sieci neuronowej Hecht–Nielsena do krótkoterminowej predykcji obciążeń systemu elektroenergetycznego*, *Mat. konf. „Aktualne Problemy w Elektroenergetyce APE’97”*, 1997, t. 4, s. 65–72.
- [14] GROSS G., GALLIANA F.D., *Short-term load forecasting*, *Proc. IEEE*, 1987, Vol. 75, No. 12, s. 1558–1573.
- [15] HÖPPNER F., KLAUWONN F., KRUSE R., RUNKLER T., *Fuzzy cluster analysis*, Wiley, Chichester 1999.
- [16] JAJUGA K., *Statystyczna teoria rozpoznawania obrazów*, PWN, Warszawa 1990.
- [17] KIM K. i in., *Implementation of hybrid short-term load forecasting system using artificial neural networks and fuzzy expert systems*, *IEEE Trans Power Syst*, 1995, Vol. 10, No. 3, s. 603–610.
- [18] KODOGIANNIS V.S., ANAGNOSTAKIS E.M., *Soft computing based techniques for short-term load forecasting*, *Fuzzy Sets and Systems*, 2002, No. 128, s. 413–426.

- [19] LAGARIAS J.C., REEDS J.A., WRIGHT M.H., WRIGHT P.E., *Convergence properties of the Nelder-Mead Simplex method in low dimensions*, SIAM Journal of Optimization, 1998, Vol. 9, No. 1, s. 112–147.
- [20] LENDASSE A. i in., *Prediction of electric load using Kohonen maps – application to the Polish electricity consumption*, Proceedings of the American Control Conference, Anchorage 2002, s. 3684–3689.
- [21] LIU T.-C., LI R.-K., *A new ART-counterpropagation neural network for solving a forecasting problem*, Expert Systems with Applications, 2005, Vol. 28, s. 21–27.
- [22] ŁYP J., *Metodyka analizy i prognozy obciążeń elektroenergetycznych systemów lokalnych*, praca doktorska, Politechnika Częstochowska. Częstochowa 2000.
- [23] MALKO J., MIKOŁAJCZYK H., SKORUPSKI W., *Progress in employment of neural network for power system load forecasting*, Mat. konf. „Neural Networks and Their Applications”, 1996, t. 1, s. 317–323.
- [24] METAXIOTIS K. i in., *Artificial intelligence in short term load forecasting: a state-of-the-art survey for the researcher*, Energy Conversion and Management, 2003, Vol. 44, s. 1525–1534.
- [25] OSOWSKI S., SIWEK K., KĄDZIELAWA A., *Neural network approach to load forecasting*, Mat. konf. „Neural Networks and Their Applications”, 1996, t. 1, s. 355–360.
- [26] POPLAWSKI T., DAŚAL K., *Canonical distribution in short-term load forecasting [w:] Technical and economic aspect of modern technology transfer in context of integration with European Union*, Kosice 2004, s. 147–153.
- [27] POPLAWSKI T., *Application of the Takagi-Sugeno (TS) fuzzy logic model for load curves prediction in the local power system*, III-rd International Scientific Symposium “Elektroenergetika”, Stara Leśna, Slovak Republic 2005.
- [28] RAHMAN S., HAZIM O., *A generalized knowledge-based short term load forecasting technique*, IEEE Trans Power Syst, 1993, Vol. 8, No. 2, s. 508–514.
- [29] RANAWEERA D., HUBELE N., KARADY G., *Fuzzy logic for short term load forecasting*, Electr. Power Energ. Syst., 1995, Vol. 18, No. 4, s. 215–222.

The analysis of the short-term power system load forecasting model based on the fuzzy clustering

A model of the short-term power system load forecasting based on fuzzy clustering is presented. It can be classified as similarity-based models relying on the assumption that if patterns of the time series sequences are similar, then the forecast patterns associated with them are also similar. This means that the patterns of the neighboring sequences are in some to each other, which does not change significantly with time. Depending the character and properties of the time series, this relation as well as similarity degree can be shaped with the help of definitions of patterns, membership and defuzzification functions, and the distance measure between patterns.

Four types of the membership functions with optimized parameters were used in the model construction. The model performance with the various distance measures between patterns was empirically examined. The model sensitivity to the membership function width was analyzed. The model resistance to the blurred noisy and missing data and the model performance with the various definitions of the reference sets were analyzed. The tests allow us to formulate some conclusions on the model quality and resistance.

Keywords: *short-term load forecasting, fuzzy clustering, similarity-based forecasting models*

Tabela 1. Optymalne dla zbioru uczącego parametry funkcji przynależności oraz błędy prognoz wyznaczone na zbiorze uczącym i testowym KSE

Dzień prognozy	Funkcja przynależności																	
	F1				F2				F3				F4			F1 z $\alpha = 2$		
	σ	α	Ψ_{trn}	Ψ_{lst}	σ	α	Ψ_{trn}	Ψ_{lst}	r	α	Ψ_{trn}	Ψ_{lst}	q	Ψ_{trn}	Ψ_{lst}	σ	Ψ_{trn}	Ψ_{lst}
pn	6,71E-03	0,77	1,77	2,90	0,0556	7,62	1,79	2,92	0,126	1,26E-04	1,83	–	1,44	1,81	2,96	0,0487	1,79	2,87
wt	7,78E-02	3,37	1,42	2,11	0,0804	9,73	1,41	2,00	0,194	4,66E-05	1,51	–	1,59	1,49	1,87	0,0578	1,43	2,11
śr	6,62E-02	5,13	1,58	2,01	0,0733	79,81	1,57	2,03	0,240	1,02E-04	1,87	–	1,61	1,68	2,07	0,0407	1,62	2,04
cz	3,11E-02	1,22	2,31	2,18	0,0671	6,39	2,35	2,30	0,243	9,12E-05	2,58	–	1,82	2,37	2,35	0,0658	2,35	2,19
pt	8,70E-02	4,25	2,12	1,78	0,0469	4,45	2,14	1,80	0,185	1,30E-05	2,21	1,84	1,98	2,15	1,81	0,0601	2,13	1,78
sb	6,68E-02	3,35	2,04	2,10	0,0606	7,18	2,02	2,03	0,228	6,30E-03	2,27	2,35	1,55	2,10	2,09	0,0479	2,05	2,06
nd	4,89E-02	1,91	1,63	1,95	0,0732	10,01	1,64	1,95	0,239	9,19E-05	2,05	2,16	1,50	1,72	2,03	0,0512	1,63	1,95
Błędy średnie			1,84	2,15			1,85	2,15			2,05	2,12		1,90	2,17		1,86	2,14