

## ZASTOSOWANIE ODCINKOWO-LINIOWEGO MINIMODELU DO MODELOWANIA PRODUKCJI SPRZEDANEJ PRZEMYSŁU

BARBARA WĄSIKOWSKA  
Uniwersytet Szczeciński

### Streszczenie

*W artykule przedstawiono nową metodę modelowania zjawisk ekonomicznych. Metoda odcinkowo-liniowego minimodelu szczególnie dobrze nadaje do modelowania zależności ekonomicznych, przy badaniu, których bardzo często można spotkać się z brakiem informacji na temat czynników wpływających na zmienną objaśnianą lub z istnieniem luk w zebranych do badania danych. Poza wyjaśnieniem konstrukcji minimodelu w artykule przedstawiono również przykład jego zastosowania w oparciu o dane statystyczne dotyczące produkcji sprzedanej przemysłu.*

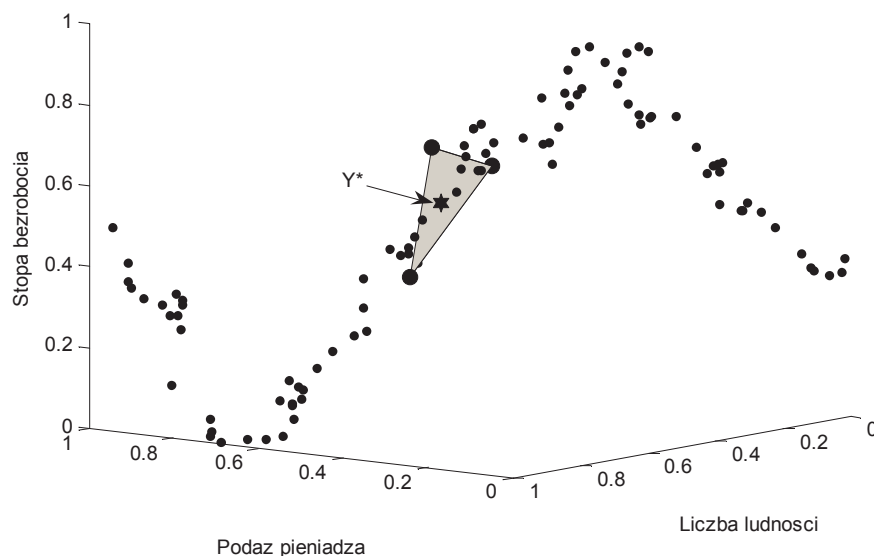
**Słowa kluczowe:** metody modelowania, modelowanie zjawisk oraz zależności ekonomicznych

### 1. Wprowadzenie

Metoda minimodeli została opracowana przez prof. Andrzeja Piegata. Programy do przeprowadzenia eksperymentów zostały napisane przez dra Marcina Korzenia, same zaś eksperymenty zostały przeprowadzone przez dr Barbarę Wąsikowską.

Opracowana metoda służy do obliczania odpowiedzi liczbowych dotyczących zmiennej objaśnianej  $y$  dla danego wektora wartości zmiennych wejściowych  $x_1, x_2, \dots, x_n$ . Minimodel ma więc takie samo zadanie jak typowy model globalny  $y=f(x_1, x_2, \dots, x_n)$  obejmujący całą przestrzeń zmiennych objaśniających  $X_1 * X_2 * \dots * X_n$ . Jednak zakres minimodelu nie jest globalny lecz jedynie lokalny i obejmuje jedynie bliskie sąsiedztwo punktu zapytania  $(x_1, x_2, \dots, x_n)$ . Dzięki temu minimodel może mieć prostą formę matematyczną, a proces jego uczenia jest znacznie krótszy niż modelu globalnego, mającego np. formę sieci neuronowej. Na rysunku 1 pokazano przykładowy, trójkatny minimodel liniowy odpowiadający na zapytanie: *Ile wynosi  $y$  jeśli  $\{x_1, x_2\} = \{0.28, 0.88\}$ ?* Po nastrojeniu modelu próbkami zależności  $y=f(x_1, x_2)$  minimodel oblicz odpowiedź:  $y^* = 0.621$ . Koncepcję oraz pierwsze wyniki badań minimodeli w przestrzeni 3D przedstawiono w [6], stąd ze względu na ograniczenia objętościowe niniejszego artykułu nie będą ponawiane szczegółowo.

Zaletą minimodelu (MM) jest nie tylko szybkość uczenia się lecz także fakt, że próbki będące dla MM podstawą obliczeniową mogą być na bieżąco uzupełniane w bazie danych w miarę ich zdobywania. Dzięki temu minimodel może udzielać odpowiedzi obliczeniowych zawsze na podstawie aktualnego zbioru próbek. Duża szybkość uczenia minimodeli w porównaniu z globalnymi modelami neuronowymi wynika z faktu ograniczonego, lokalnego zasięgu MM podczas gdy modele globalne ze względu na swój duży zasięg i często silną nieliniowość powierzchni muszą uczyć się o wiele dłużej.



Rysunek 1. Przykładowy minimodel nastrojony do udzielania odpowiedzi na Ile wynosi  $y^*$  jeśli  $\{x_1^*, x_2^*\} = \{0.28, 0.88\}$ ?. Obliczona odpowiedź:  $y^* = 0.621$ .

Źródło: Opracowanie własne.

## 2. Modelowanie zależności typu $y = f(x)$ na przykładzie produkcji sprzedanej przemysłu (zmienna objaśniana) oraz eksportu towarów ogółem (zmienna objaśniająca)

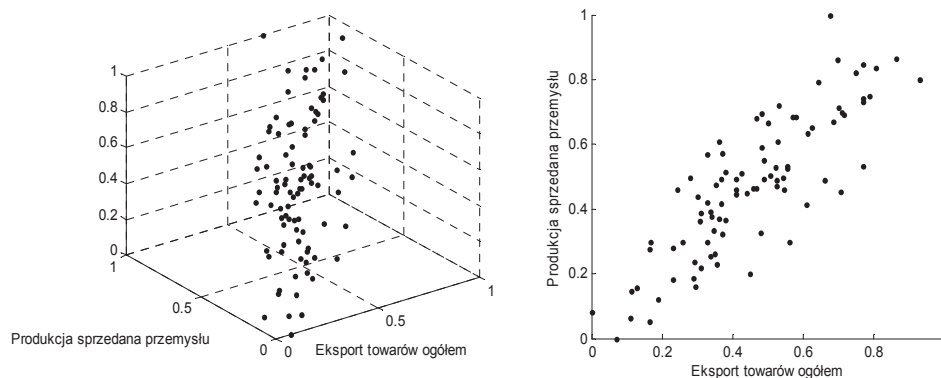
Badania przeprowadzone przez autorkę dotyczące czynników wpływających na poziom produkcji sprzedanej przemysłu w okresie od 1993 roku do 2002 roku wykazały, że produkcja ta zależna jest od dziewięciu czynników  $x_i$ , które podane zostaną według rankingu istotności od najważniejszego do najmniej istotnego (ale istotnego) czynnika [7]:

- $x_1$  – eksport towarów ogółem,
- $x_2$  – przeciętne wynagrodzenie nominalne brutto,
- $x_3$  – pracujący w sektorze przedsiębiorstw ogółem,
- $x_4$  – nakłady inwestycyjne,
- $x_5$  – koszty uzyskania przychodów z całokształtu działalności,
- $x_6$  – dochody budżetu państwa,
- $x_7$  – wydatki budżetu państwa,
- $x_8$  – wskaźnik cen produkcji sprzedanej przemysłu,
- $x_9$  – inflacja miesięczna,

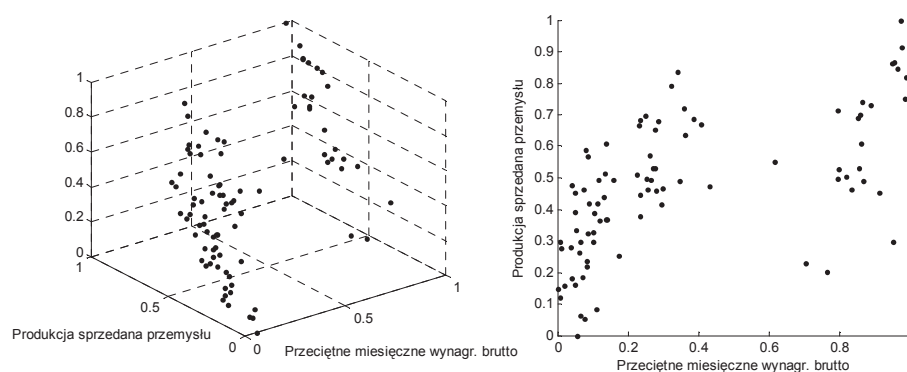
W niniejszym artykule przedstawione zostaną wyniki modelowania zależności  $y = f(x_1)$  to jest zależności pomiędzy produkcją sprzedaną przemysłu a najsilniejszym, najbardziej istotnym czynnikiem – eksportem towarów ogółem (rysunek 2a) oraz zależności  $y = f(x_2)$  czyli zależności pomiędzy produkcją sprzedaną przemysłu a drugim pod względem istotności czynnikiem na nią

wpływającym tj. przeciętnym wynagrodzeniem nominalnym brutto (rysunek 2b).

a)



b)



Rysunek 2. Zależność pomiędzy produkcją sprzedaną przemysłu i eksportem towarów ogółem (rys 2a) oraz pomiędzy produkcją i przeciętnym miesięcznym wynagrodzeniem brutto (rys 2b)

Źródło: Opracowanie własne.

O sile obydwu wybranych czynników świadczą same błędy modelowania, które w obu wypadkach są rzędu kilku procent (tabela 1).

W następnym rozdziale zaprezentowane zostaną trzy fragmenty z przykładowego procesu uczenia minimodelu.

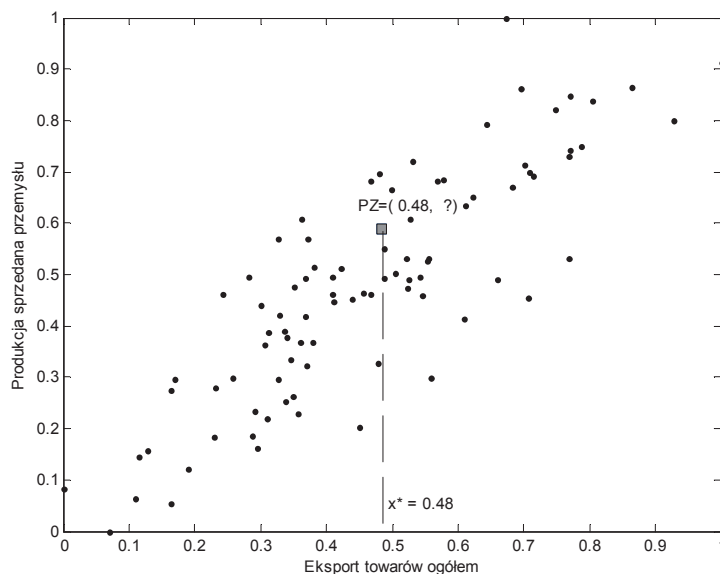
Tabela 1. Wartości błędów MAE modeli uzyskanych podczas stosowania różnych narzędzi

Zastosowane narzędzie	Wartość błędu MAE	
	Eksport towarów ogółem	Przeciętne miesięczne wyn.
Wielomian szóstego stopnia: $y = 31*x^6 - 99*x^5 + 1.2e+002*x^4 - 71*x^3 + 20*x^2 - 1.2*x + 0.068$	0,0872	0,1087
Sieć neuronowa GRNN	0,0557	0,3220
Metoda najbliższych sąsiadów (KNN)	0,0709	0,7512

Źródło: Opracowanie własne.

### 3. Przykłady uczenia minimodelu w przestrzeni 2D

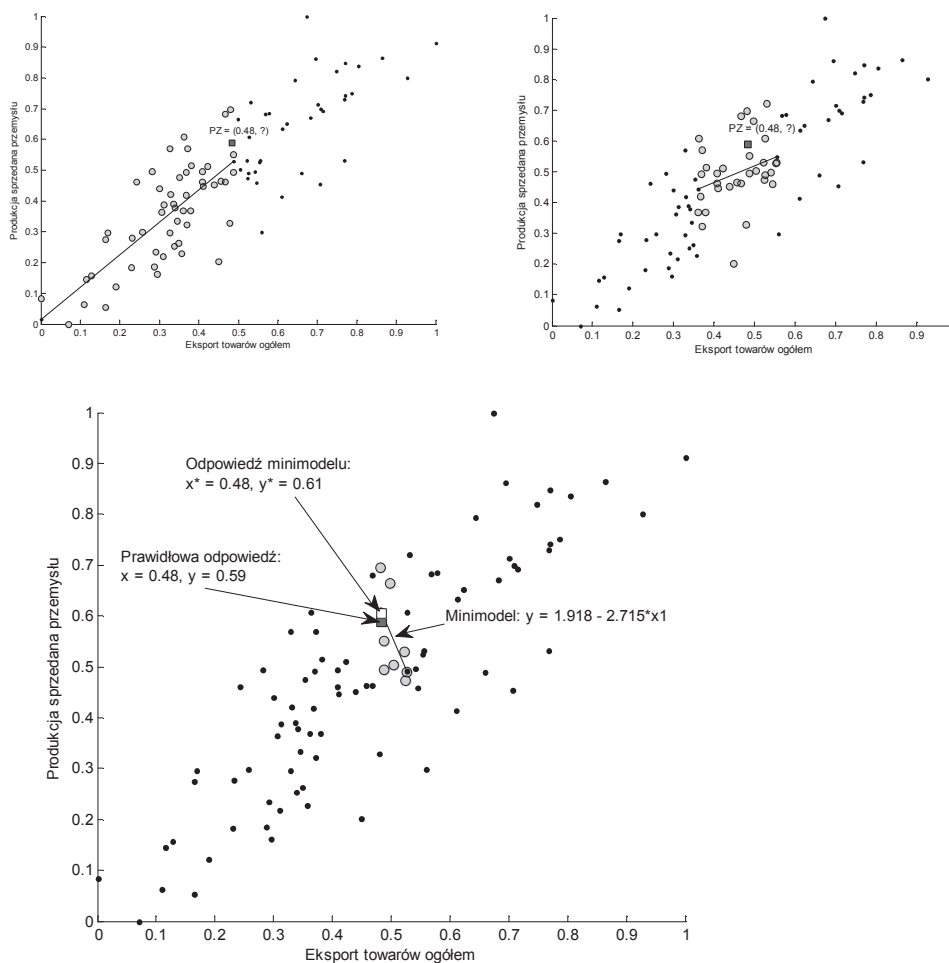
Zadaniem minimodelu w niniejszym eksperymencie jest obliczenie odpowiedzi na zapytanie: *Jaka jest wartość produkcji sprzedanej przemysłu  $y^*$  jeżeli eksport towarów ogółem  $x^* = 0,48$ ?* Minimodel musi obliczyć odpowiedź  $y^*$  na podstawie próbek zależności  $y = f(x_1)$ , które ma do dyspozycji (rysunek 3).



Rysunek 3. Rozkład próbek w przestrzeni 2D wraz z punktem zapytania (PZ): Jaka jest wartość produkcji sprzedanej przemysłu  $y^*$  jeżeli eksport towarów ogółem  $x^* = 0,48$ ?

Źródło: Opracowanie własne.

Wcześniej jednak minimodel musi się nastroić tzn. znaleźć swoje optymalne położenie i dopiero wtedy obliczyć wartość  $y^*$  odpowiedzi. Na rysunku 4a, 4b i 4c pokazane są trzy etapy uczenia minimodelu w przypadku problemu 2D.



Rysunek 4a. Startowa pozycja minimodelu obliczającego odpowiedź na zapytanie: Jaka jest wartość produkcji sprzedanej przemysłu  $y^*$  jeżeli eksport towarów ogółem  $x^* = 0.48$ ?

Rysunek 4b. Jedna z pośrednich pozycji minimodelu w trakcie procesu uczenia,

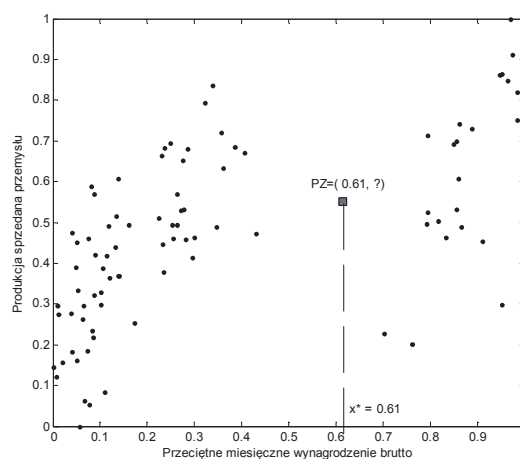
Rysunek 4c. Końcowa pozycja minimodelu  $y = 1.918 - 2.715 \cdot x_1$ ,  $x^* = 0.48$ ,  $y^* = 0.61$ .

Źródło: Opracowanie własne.

Na rysunku 4a pokazano startową formę minimodelu odcinkowo-liniowego. Startowa pozycja dobierana jest przez program komputerowy przypadkowo. Podobnie kąt nachylenia oraz długość. Jedynym warunkiem jest to, aby minimodel obejmował swym zakresem punkt zapytania PZ. Rysunek 4b przedstawia jedną z pośrednich pozycji odcinka, który w trakcie dostrajania zmienia zarówno swoją długość jak i kąt nachylenia. Po zakończeniu procesu uczenia minimodel obliczył odpowiedź  $y^* = 0.605$ . Dokładna wartość odpowiedzi była w tym wypadku znana ( $y = 0.59$ ) bo-

wiem zapytanie dotyczyło jednej z próbek uczących, wyjętej z procesu uczenia. Absolutny błąd MAE otrzymanego modelu wyniósł 0,0167 i jest znacznie niższy od błędów otrzymanych przy zastosowaniu innych narzędzi do modelowania (tabela 1).

W drugim eksperymencie, dotyczącym produkcji sprzedanej przemysłu  $y$  oraz przeciętnego miesięcznego wynagrodzenia brutto  $x_2$ , próbką wyjętą z procesu uczenia była próbka o wartościach (0.61, 0.55). Zadaniem minimodelu było znalezienie odpowiedzi na zapytanie: *Jaka jest wartość produkcji sprzedanej przemysłu  $y^*$  jeżeli przeciętne miesięczne wynagrodzenie brutto wynosi  $x^* = 0,61$ ?* Punkt zapytania  $PZ = (0.61, ?)$  przedstawiony na rysunku 5, usytuowany jest w luce informacyjnej, w której występuje lokalny brak próbek. Takie usytuowanie sprawia, że znalezienie odpowiedzi na zadane wyżej pytanie staje się jeszcze trudniejsze.

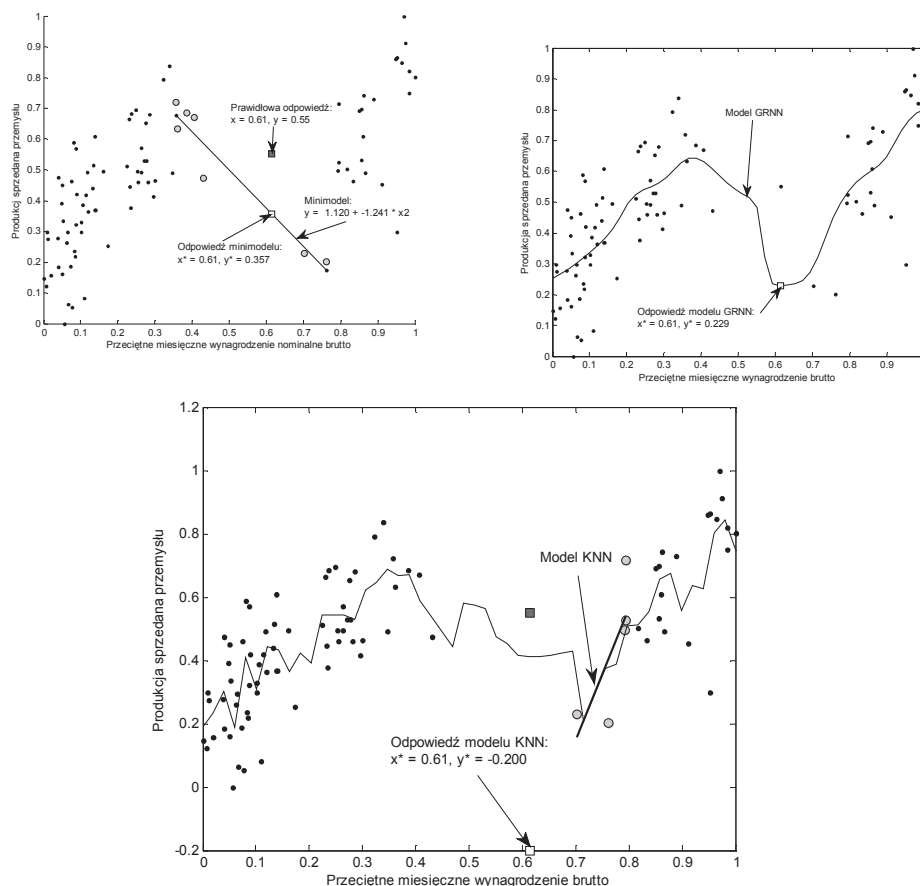


Rysunek 5. Rozkład próbek w przestrzeni 2D wraz z punktem zapytania (PZ): *Jaka jest wartość produkcji sprzedanej przemysłu  $y^*$  jeżeli przeciętne miesięczne wynagrodzenie brutto wynosi  $x^* = 0,61$ ?*

Źródło: Opracowanie własne.

Do znalezienia odpowiedzi na zapytanie, oprócz minimodelu liniowo-odcinkowego użyto w eksperymencie sieci neuronowej GRNN (General Regression Neural Networks) oraz metody najbliższych sąsiadów KNN (rysunek 6a, 6b oraz 6c).

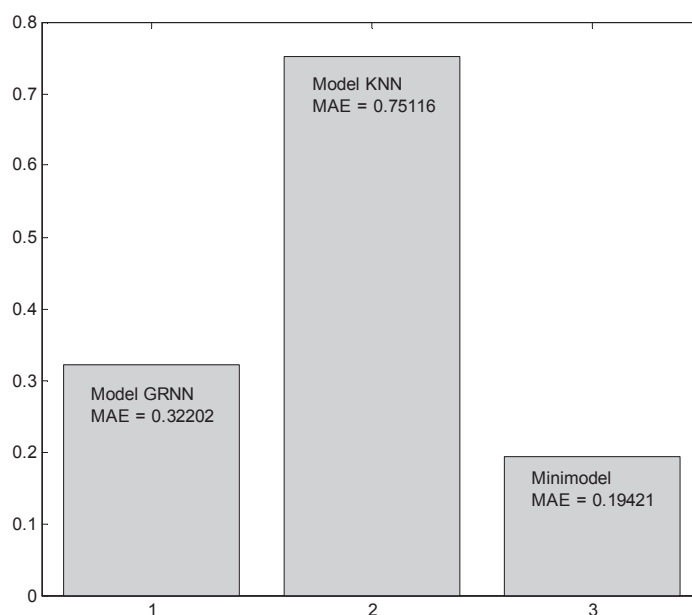
Zastosowanie odcinkowego (2-punktowego) minimodelu  
do modelowania produkcji sprzedanej przemysłu



Rysunek 6a. Minimodel, Rysunek 6b. Globalny model GRNN, Rysunek 6c. Model KNN,  
 $y^* = -0.200$ . Prawidłowa odpowiedź:  $y = 0.55$ .

Źródło: Opracowanie własne.

Rysunek 6a przedstawia uzyskany minimodel  $y = 1.120 - 1.241 \cdot x_2$ . Odpowiedź modelu na zapytanie wynosi  $y^* = 0.357$  (prawidłowa odpowiedź to  $y = 0.55$ ). Na rysunku 6b pokazany jest globalny model neuronowy GRNN. Odpowiedź tego modelu na zapytanie wynosi  $y^* = 0.229$ . Rysunek 6c natomiast, przedstawia otrzymany model KNN. W sytuacji gdy próbka leży w luce informacyjnej metoda KNN oblicza odpowiedź  $y^*$  na podstawie próbek najbliższych, tj próbek leżących z prawej strony punktu zapytania (PZ). Wartość odpowiedzi  $y^* = -0.200$ . Jak widać, metoda KNN nie uwzględnia próbek leżących po lewej stronie zapytania, po lewej stronie luki informacyjnej. Robi to natomiast metoda minimodeli (rysunek 6a), stąd wynik jej odpowiedzi jest lepszy. Dowodem tego może być porównanie absolutnego błędu MAE wszystkich otrzymanych modeli (rysunek 7).



Rysunek 7. Porównanie błędów MAE otrzymanych modeli produkcji sprzedanej przemysłu

Źródło: Opracowanie własne.

Podany wyżej przykład pokazuje, że metoda KNN nie jest efektywną metodą regresji w lukach informacyjnych (w przeciwieństwie do minimodelu). Rysunek 6 wyjaśnia przyczynę tego stanu rzeczy. Jeżeli zapytanie  $x^*$  leży bliżej jednego brzegu luki niż drugiego, to metoda KNN wybiera do obliczenia odpowiedzi  $y^*$  próbki brzegu najbliższego. W ten sposób metoda ta traci informacje o tendencjach zmian  $dy/dx$  modelowanej zależności. Tymczasem metoda minimodeli wykrywa tendencje  $dy/dx$  i wykorzystuje ją do obliczenia odpowiedzi  $y^*$ . Dodatkowo długość minimodelu jest kolejną, trzecią informacją o zakresie ważności minimodelu i o zakresie utrzymywania się obliczonej tendencji. Jak więc widać informacja dostarczana przez minimodel jest znacznie bogatsza niż ta dostarczana przez metodę KNN.

#### 4. Podsumowanie

W artykule przedstawiono koncepcję odcinkowo-liniowego, lokalnego minimodelu, którego celem jest udzielenie – na podstawie próbek zależności  $y=f(x)$  jaką dysponujemy w bazie wiedzy – odpowiedzi na zapytania obliczeniowe typu *Ile wynosi  $y$  jeśli  $x=...$ ?* Badania wykazały, że dokładność obliczeń realizowanych przez minimodel jest wysoka, a czas obliczeń krótki. Uzyskana wysoka dokładność jest wynikiem lokalności minimodelu. W czasie licznych eksperymentów, przeprowadzonych przez autorkę artykułu, metoda minimodeli okazała się być skutecznym narzędziem predykcji w zagadnieniach regresji. Metoda ta szczególnie nadaje się do modelowania zja-



wisk wielowymiarowych, zawierających znaczne luki informacyjne.

#### **Bibliografia**

- [1] Cherkassky V., Mulier F.: *Learning from data*, Wiley-Interscience, IEEE Press, 2007 r.
- [2] Geissler S.: *Predictive inference*. Chapman and Hall, New York, 1993r.
- [3] Hand D., Mannila H., Smyth P.: *Principles of data mining*. Massachusetts Institute of Technology, USA, 2001 r.
- [4] Klęsk P.: *Construction of a neurofuzzy network capable of extrapolating (and interpolating) with respect to the convex hull of a set of input samples in  $R^n$* . IEEE Transactions on Fuzzy Systems, vol 16, issue 5, pp.1161–1179, 2008 r.
- [5] Kordos M., Blachnik M., Strzempa D.: *Do we need whatever more than  $k$ -NN?* Proceedings of 10-th International Conference on Artificial Intelligence and Soft Computing, Zakopane, Poland, June 13–17, 2010, Springer, Germany, pp.414–421., 2010 r.
- [6] Piegat A., Wąsikowska B., Korzeń M.: *Zastosowania samouczącego się, 3-punktowego minimodelu do modelowania stopy bezrobocia w Polsce*. Artykuł ukaże się w 2010 roku w zeszycie naukowym *Studia Informatica Uniwersytetu Szczecińskiego*.
- [7] Wąsikowska B.: „*Niestandardowe metody identyfikacji czynników determinujących produkcję sprzedaną przemysłu*”, praca doktorska, Uniwersytet Szczeciński, Szczecin, 2007 r.
- [8] Witten I.A., Frank E.: *Data mining*. Morgan Kaufmann Publishers, San Francisco, 2005 r.

**APPLICATION OF THE LINEAR-SEGMENT MINI-MODEL IN MODELLING OF  
SOLD PRODUCTION OF INDUSTRY**

## Summary

*The paper presents new method of economic phenomena modelling. The linear-segment mini-model method suits very well for modelling economic dependencies, where frequently a part of explanatory variables influencing the explained variable is not known because of lack of data. Aside from the construction of the mini-model, the paper presents the example of its application, that is based on statistical data regarding sold production of industry.*

**Keywords:** quantitative methods in data analyses and modeling, modeling of business processes

Barbara Wąsikowska  
Uniwersytet Szczeciński  
ul. Mickiewicza 64, 71-101 Szczecin  
e-mail: barbara.wasikowska@wneiz.pl