

**Barbara Wąsikowska**

## **NIESTANDARDOWE METODY IDENTYFIKACJI CZYNNIKÓW DETERMINUJĄCYCH PRODUKCJĘ SPRZEDANĄ PRZEMYSŁU**

### **Wprowadzenie**

W ostatnich latach, wraz z dynamicznym rozwojem informatyki i dostarczanych przez nią coraz to większych mocy obliczeniowych, obserwuje się ogromny wzrost zainteresowania metodami sztucznej inteligencji w takich dziedzinach, jak: finanse, medycyna, zastosowania inżynierskie, geologia czy fizyka. W obszarze ekonomii metody te znajdują zastosowanie w prognozowaniu bankructw, notowań giełdowych, kursów walut, wskaźników finansowo-ekonomicznych, klasyfikacji przedsiębiorstw, wycenie kontraktów terminowych, spółek, nieruchomości, w ocenie wiarygodności kredytowej oraz identyfikacji klientów według przedstawianej oferty. Obecnie wiele znaczących placówek naukowych i badawczych na całym świecie zajmuje się zastosowaniami sztucznej inteligencji. Stopień zaawansowania prac pozwolił na wydzielenie kilku kierunków badań. Najważniejszymi z nich są sieci neuronowe (ang. *neural networks*), logika rozmyta (ang. *fuzzy logic*), algorytmy genetyczne (ang. *evolutionary algorithms*) i teoria zbiorów przybliżonych (ang. *rough sets*). Metody te, w szczególności sieci neuronowe i algorytmy genetyczne, pozwalają na zautomatyzowanie niektórych, bardziej uciążliwych etapów procesu modelowania, co w dużej mierze ułatwia budowę modelu danego zjawiska ekonomicznego. Z tego względu podjęto się napisania pracy doktorskiej pt. *Niestandardowe metody identyfikacji czynników*

*determinujących produkcję sprzedaną przemysłu*, obronionej 5 kwietnia 2007 roku na Wydziale Nauk Ekonomicznych i Zarządzania Uniwersytetu Szczecińskiego.

Praca doktorska składa się z trzech rozdziałów. W rozdziale pierwszym opisano stany i uwarunkowania systemowe *produkcji sprzedanej przemysłu* w Polsce w latach 1990–2006. Scharakteryzowano dynamikę poziomu produkcji na przestrzeni lat, opisano przyczyny jej gwałtownego spadku oraz przedstawiono stan obecny. Zaprezentowano również zebrany materiał statystyczny i skonstruowano model ekonometryczny *produkcji sprzedanej przemysłu*.

W rozdziale drugim pracy dokonano przeglądu metod sztucznej inteligencji oraz opisano sposób modelowania zjawisk gospodarczych przy użyciu sztucznych sieci neuronowych RBF, algorytmów genetycznych i zbiorów przybliżonych.

Rozdział trzeci został w całości poświęcony badaniom identyfikacyjnym czynników determinujących *produkcję sprzedaną przemysłu* z zastosowaniem wyżej wymienionych metod sztucznej inteligencji. Dokonana została również analiza porównawcza metod niestandardowych (metod sztucznej inteligencji) z metodami standardowymi (metodami ekonometrycznymi).

W zakończeniu pracy zawarto podsumowanie i wnioski wynikające z badań dotyczących możliwości zastosowania metod sztucznej inteligencji do identyfikacji czynników wpływających na dane zjawisko ekonomiczne.

## **1. Cel i hipoteza pracy**

Celem pracy była analiza stosowalności niestandardowych metod modelowania (sieci neuronowych, algorytmów genetycznych, teorii zbiorów przybliżonych) do modelowania *produkcji sprzedanej przemysłu*. Procedura badawcza została oparta na analizie systemowej – metodzie badawczej służącej poszukiwaniu wskazówek do podjęcia ważnych decyzji gospodarczych. Przyjętym obszarem badawczym były uwarunkowania produkcji sprzedanej przemysłu (zwanej w dalszej części pracy produkcją) w Polsce, w okresie od stycznia 1993 roku do grudnia 2000 roku. W pracy zweryfikowano pogląd, że zastosowanie metod sztucznej inteligencji do modelowania *produkcji sprzedanej przemysłu* może stanowić znaczące uzupełnienie przeprowadzonych badań poznawczych przy użyciu metod klasycznych (standardowych).

Rozprawa ma charakter metodyczny i poznawczy. Aspekt metodyczny polega na weryfikacji przyjętych *a priori* metod badawczych. Za miary jakości weryfikowanych metod przyjęto ich użyteczność, wiarygodność i skuteczność w odniesieniu do analizy zjawisk ekonomicznych.

## 2. Prezentacja metod i rozwiązań w badanym obszarze

Obecnie najczęściej stosowanymi metodami do opisu zjawisk ekonomicznych są metody statystyczne i ekonometryczne. Podstawowym narzędziem analizy ekonometrycznej jest opisowy model ekonometryczny. Jest on równaniem (lub układem równań), który w sposób przybliżony przedstawia zasadnicze powiązania ilościowe występujące między rozpatrywanymi zjawiskami ekonomicznymi. Bardzo często zależności występujące między tymi zjawiskami ekonomicznymi są złożone i wielokierunkowe. Zwykle na badane zjawisko oddziałuje wiele innych zjawisk o charakterze zarówno ekonomicznym, jak i pozaekonomicznym. Model ekonometryczny jest sformalizowanym opisem badanego fragmentu rzeczywistości ekonomicznej uwzględniającym tylko istotne jej elementy, a pomijającym mniej istotne, zwane elementami ubocznymi. Zewnętrznym wyrazem tego opisu jest równanie modelu. Modele ekonometryczne można sklasyfikować według pięciu różnych kryteriów<sup>1</sup>:

- liczba równań modelu (modele jednorównaniowe i wielorównaniowe),
- postać analityczna zależności funkcyjnych modelu (modele liniowe i nieliniowe),
- rola czynnika czasu w równaniach modelu (modele statyczne i dynamiczne),
- ogólnopoznawcze cechy modelu (modele przyczynowo-skutkowe i modele symptomatyczne),
- charakter powiązań między nieopóźnionymi zmiennymi endogenicznymi w modelu wielorównaniowym (modele proste, rekurencyjne i modele o równaniach współzależnych).

W praktyce najczęściej wykorzystuje się modele liniowe lub takie, do oszacowania których można wykorzystać klasyczną metodę najmniejszych kwadratów.

---

<sup>1</sup> M. Gruszczyński, M. Podgórska, *Ekonometria*, Oficyna Wydawnicza SGH, Warszawa 2004.

### 3. Charakterystyka problemu w literaturze przedmiotu

Przesłanką do powstania przedstawionej rozprawy była potrzeba wypełnienia luki metodycznej w naukowym warsztacie badawczym, dotyczącym modelowania systemów rzeczywistych. Stosowane metody nie są w stanie w pełni odzwierciedlić otaczającej nas złożonej rzeczywistości ekonomicznej. **Nie istnieje bowiem uniwersalna metoda, która pozwalałaby tworzyć modele różnorodnych zjawisk ekonomicznych** z taką samą, zadawalającą dokładnością (szczególnie dotyczy to identyfikacji nieliniowej). W związku z tym stale poszukuje się nowych metod, które z większą precyzją wyjaśniałyby zjawiska makroekonomiczne. Odpowiedzią na to zapotrzebowanie może być **próba zastosowania niestandardowych metod modelowania procesów makroekonomicznych – metod sztucznej inteligencji, takich jak: sieci neuronowe, algorytmy genetyczne, teoria zbiorów przybliżonych.**

Dokonując przeglądu literatury przedmiotu, można się przekonać, że pojęcia „standardowe (klasyczne) metody” i „niestandardowe (nieklasyczne) metody” są różnie interpretowane<sup>2</sup>. W niniejszej rozprawie słowo „standardowe” oznacza metody wcześniej opracowane, częściej wykorzystywane niż inne, stosowane przy często powtarzających się założeniach. Natomiast pojęcie „metody niestandardowe” odnosi się do metod nowych, rzadziej używanych do opisu zjawisk ekonomicznych – metod sztucznej inteligencji.

### 4. Propozycje autorskich rozwiązań

#### **Model produkcji sprzedanej przemysłu opracowany na podstawie jednoargumentowego rankingu istotności**

Do budowy modelu, opartego o przeprowadzony ranking istotności zmiennych, wybranych zostało osiem zmiennych, których istotność była nie mniejsza niż 20%. W trakcie budowy modelu pojawił się problem ustalenia odpowiedniej ilości neuronów w warstwie ukrytej sieci neuronowej. W celu wyjaśnienia, jaka liczba neuronów byłaby odpowiednia, stworzono po dziesięć modeli dla sieci

<sup>2</sup> Cz. Domański, K. Pruska, *Nieklasyczne metody statystyczne*, PWE, Warszawa 2000.

zawierających od jednego do dwudziestu neuronów w warstwie ukrytej. Każda z sieci była trenowana na zbiorze próbek uczących przez 100 epok, a następnie testowana na zbiorze próbek walidujących.

Ponieważ wraz ze wzrostem liczby neuronów w warstwie ukrytej rośnie liczba parametrów sieci, zwiększono liczbę próbek, dodając do 96 próbek dodatkowych 95 próbek uśrednionych, by uniknąć zbyt szybkiego przeuczenia sieci. W ten sposób powstał zbiór 191 próbek. Spośród 191 próbek każdorazowo losowano 38 próbek stanowiących zbiór walidujący (co stanowi 20% wszystkich próbek), natomiast pozostałe 153 próbki (80% wszystkich próbek) stanowiły zbiór uczący. Próbki w zbiorze walidującym i uczącym nie mogły się powtarzać. Spośród wszystkich otrzymanych modeli, wybrano po jednym, najlepszym (najniższy błąd *MSE*) dla danego rodzaju sieci. Wybranych zostało więc 20 modeli. Dla każdego z nich został policzony błąd uczący i błąd na zbiorze walidującym. Po przeanalizowaniu błędów uczących i błędów na zbiorze walidującym stwierdzono, że dla sieci zawierających od 1 do 8 neuronów w warstwie ukrytej zarówno błąd uczący, jak i błąd na zbiorze walidującym gwałtownie spada. Powyżej jednak dziewięciu neuronów, wartości błędu na zbiorze walidującym zaczynają się „wahać”, tzn. raz rosną, a raz spadają, co może świadczyć o przeuczeniu sieci. By uniknąć niedouczenia, jak i przeuczenia sieci stwierdzono, że najodpowiedniejszą liczbą neuronów w warstwie ukrytej będzie liczba 8. Parametry sieci zostały nastrojone automatycznie przez instrukcję *newrb*, znajdującą się w programie *Matlab*. W celu sprawdzenia czy ustalona liczba neuronów w warstwie ukrytej jest prawidłowa, zbudowano 10 modeli (zwanymi dalej modelami pomocniczymi). Po przeanalizowaniu otrzymanych wyników skonstruowano poniższy model neuronowy.

Uzyskane parametry sieci nie posiadają niestety interpretacji ekonomicznej (w przeciwieństwie do parametrów strukturalnych modelu ekonometrycznego). W przedstawionej rozprawie weryfikację wszystkich otrzymanych modeli oparto na wizualnej ocenie stopnia dopasowania modelu do danych empirycznych oraz obliczono dla każdego z nich *błąd średniokwadratowy MSE*.

Równanie modelu:

$$y(z) = \sum_{i=1}^8 w_i \exp\left(-\frac{\sum_{j=1}^8 (z_j - c_{ij})^2}{2\delta_i^2}\right) + w_0 \quad (1)$$

gdzie:

$$w_0 = -0,1817,$$

$$[w_i] = [6,0868 \quad 4,1600 \quad -2,6441 \quad 0,2480 \quad 1,5589 \quad 0,2852 \quad -9,1775 \quad 1,2257],$$

$$[c_j] = \begin{bmatrix} 0,6829 & 0,8439 & 0,4062 & 0,6672 & 0,8466 & 0,7355 & 0,1073 & 0,6348 \\ 0,7007 & 0,8431 & 0,7937 & 0,8147 & 0,3584 & 0,3914 & 0,3340 & 0,4257 \\ 0,6596 & 0,7468 & 0,8660 & 0,8822 & 0,2454 & 0,5107 & 0,3923 & 0,2434 \\ 1,0000 & 1,0000 & 0,9738 & 0,9733 & 0,7336 & 0,6641 & 0,0034 & 0,4872 \\ 0,6424 & 0,7507 & 0,3213 & 0,4705 & 0,6254 & 0,5506 & 0,4525 & 0,7663 \\ 0,3559 & 0,4601 & 0,7020 & 0,7834 & 0,5165 & 0,7510 & 0,7656 & 0,3329 \\ 0,5779 & 0,8070 & 0,3851 & 0,6678 & 0,7303 & 0,6235 & 0,1841 & 0,6290 \\ 0,2809 & 0,5516 & 0,2522 & 0,3301 & 0,4339 & 0,6636 & 0,1308 & 0,7263 \end{bmatrix},$$

$$[\delta] = \begin{bmatrix} 0,8326 \\ 0,8326 \\ 0,8326 \\ 0,8326 \\ 0,8326 \\ 0,8326 \\ 0,8326 \\ 0,8326 \end{bmatrix}.$$

**Błąd modelu:**  $MSE = 0,0695$

### Model produkcji sprzedanej przemysłu opracowany w oparciu o sieci neuronowe RBF

Konstruując kolejny model, postanowiono sprawdzić wszystkie możliwe kombinacje wejść do modelu za pomocą sieci RBF. W trakcie budowy modelu pominięto całkowicie etap redukcji zbioru zmiennych objaśniających. Najlepszym modelem (tzn. modelem o najmniejszym błędzie) okazał się model zawierający dziesięć wejść:  $X_1, X_4, X_5, X_9, X_{14}, X_{16}, X_{17}, X_{18}, X_{20}, X_{21}$ . Błąd średniokwadratowy modelu produkcji sprzedanej przemysłu otrzymanego przy użyciu sieci neuronowej RBF z ośmioma neuronami w warstwie ukrytej, z pominięciem etapu redukcji zbioru zmiennych wynosi  $MSE = 0,0496$  i jest o 0,02 niższy od błędu modelu otrzymanego w oparciu o ranking istotności. Spadek błędu średniokwadratowego świadczy o jeszcze dokładniejszym dopasowaniu modelu do danych empirycznych.

Równanie modelu:

$$y(x) = \sum_{i=1}^8 w_i \exp\left(-\frac{\sum_{j=1}^{10} (x_j - c_{ij})^2}{2\delta_i^2}\right) + w_0, \quad (2)$$

gdzie:

$$w_0 = -0,0260,$$

$$[c_i] = \begin{bmatrix} 0,6738 & 0,9706 & 0,6944 & 0,6284 & 0,5588 & 0,7133 & 0,4800 & 0,1291 & 0,2769 & 0,9031 \\ 0,3062 & 0,1189 & 0,7217 & 0,1451 & 0,4050 & 0,8296 & 0,8514 & 1,0000 & 1,0000 & 0,3256 \\ 0,3559 & 0,7020 & 0,9181 & 0,5165 & 0,7510 & 0,3329 & 0,3879 & 0,0060 & 0,3692 & 0,9186 \\ 0,5546 & 0,2770 & 0,9850 & 0,3926 & 0,5731 & 0,3509 & 0,3951 & 0,1201 & 0,2000 & 0,9496 \\ 0,4788 & 0,1020 & 0,6999 & 0,2552 & 0,3300 & 0,6440 & 0,3823 & 0,2432 & 0,1231 & 0,7984 \\ 0,3622 & 0,1386 & 0,5730 & 0,1914 & 0,4259 & 0,6211 & 0,6841 & 0,7267 & 0,4308 & 0,6201 \\ 0,4787 & 0,2089 & 0,7749 & 0,2147 & 0,4587 & 0,4024 & 0,3763 & 0,2492 & 0,2923 & 0,8566 \\ 0,4217 & 0,2078 & 0,7640 & 0,2184 & 0,3657 & 0,3154 & 0,4242 & 0,2432 & 0,2615 & 0,8372 \end{bmatrix},$$

$$[w_i] = [2,1404 \quad -0,1865 \quad -2,6500 \quad 1,8155 \quad -2,3180 \quad 0,9436 \quad 2,5209 \quad -1,4278],$$

$$[\delta] = \begin{bmatrix} 0,8326 \\ 0,8326 \\ 0,8326 \\ 0,8326 \\ 0,8326 \\ 0,8326 \\ 0,8326 \\ 0,8326 \end{bmatrix}.$$

### Model produkcji sprzedanej przemysłu opracowany w oparciu o metodę hierarchiczną

Znaczną ilość czasu, jakiej wymaga sprawdzenie wszystkich możliwych kombinacji wejść do modelu przy użyciu sztucznych sieci neuronowych, można zredukować, stosując hierarchiczną metodę doboru zmiennych do modelu. Metoda ta polega na stopniowym wprowadzaniu do modelu kolejnych zmiennych objaśniających. Proces budowy modelu produkcji sprzedanej przemysłu rozpoczęto od opracowania 21 modeli jednoargumentowych. Z przeprowadzonych badań wynikało, że jako pierwsza do modelu wynikowego powinna zostać wprowadzona zmienna  $X_1$  – *eksport towarów ogółem*, której model charakteryzował się najmniejszym błędem spośród

wszystkich modeli jednoargumentowych. Kolejne zmienne wprowadzono do modelu, aż do uzyskania modelu z 21 zmiennymi. Po przeanalizowaniu danych zawartych w powyższej tabeli okazało się, że najmniejszym błędem charakteryzował się model zawierający 14 wejść. Błąd średniokwadratowy modelu wyniósł  $MSE = 0,0552$  i był on zaledwie o 0,0056 większy od błędu modelu otrzymanego tylko w oparciu o sieć RBF. Natomiast czas oczekiwania na wyniki uległ znacznemu skróceniu.

### **Model produkcji sprzedanej przemysłu opracowany w oparciu o algorytm genetyczny i sieć RBF**

Do budowy kolejnego modelu produkcji sprzedanej przemysłu postanowiono użyć algorytmu genetycznego, gdyż obok sztucznych sieci neuronowych jest to drugie chętnie stosowane narzędzie do różnego rodzaju analiz ekonomicznych. Badanie wykonane zostało przy pomocy instrukcji napisanej w programie *Matlab*. Poniżej przedstawiono opis słowny poszczególnych kroków zawartych w instrukcji:

1. Ze wszystkich możliwych kombinacji zmiennych wygenerowano drogą losową populację początkową zawierającą 20 osobników.
2. Dla każdego osobnika stworzony został model RBF, na podstawie którego obliczono wartość błędu średniokwadratowego, będącego funkcją dopasowania osobnika. Osobniki posortowano rosnąco (tzn. od „najlepszego” – mającego najniższą wartość błędu, do „najgorszego” – o najwyższej wartości błędu), po czym 1/4 najgorszych usunięto z populacji. W ten sposób przeprowadzono selekcję osobników. Pozostałe osobniki wzięły udział w kolejnym kroku modelowania, jakim jest reprodukcja.
3. Reprodukcję przeprowadzono przy użyciu operatora krzyżowania. O tym, które osobniki mają brać udział w krzyżowaniu, decydowały losowo stosowane funkcje AND, OR i XOR. Proces reprodukcji uznawano za zakończony, gdy liczba osobników w populacji równa była 20.
4. Otrzymałą na drodze reprodukcji populację poddawano mutacji. Z populacji losowano pewną liczbę bitów mającą ulec mutacji. Procesowi mutacji nie ulegał jednak osobnik, którego błąd średniokwadratowy był najniższy. Po wylosowaniu indeksów bitów, czyli miejsc mających ulec mutacji, następowała mutacja polegająca na zamianie 0 na 1 i odwrotnie. Całą procedurę (tzn. od selekcji) powtarzano 100 razy.
5. Na końcu generowano:



- wykres przedstawiający przystosowanie każdego osobnika w danym pokoleniu,
- najlepszego osobnika zapisanego w postaci chromosomu (tabela 1),
- zapis w postaci numerów wejść (wybrane zostały te wejścia, dla których wartość w chromosomie była równa jeden – tabela 1),
- błąd otrzymanego modelu.

Poniższa tabela przedstawia wynik działania zastosowanego algorytmu.

Tabela 1. Wyniki działania algorytmu genetycznego

|                      |    |   |   |    |    |   |    |   |   |   |     |   |     |     |   |   |     |   |     |   |   |
|----------------------|----|---|---|----|----|---|----|---|---|---|-----|---|-----|-----|---|---|-----|---|-----|---|---|
| Otrzymany chromosom: | 1  | 0 | 0 | 1  | 1  | 0 | 1  | 0 | 1 | 0 | 0   | 1 | 0   | 1   | 1 | 0 | 0   | 1 | 0   | 1 | 0 |
| Wybrane zmienne:     | X1 |   |   | X4 | X5 |   | X7 |   |   |   | X12 |   | X14 | X15 |   |   | X18 |   | X20 |   |   |

Źródło: opracowanie własne.

Błąd średniokwadratowy otrzymanego modelu wynosi  $MSE = 0,0553$  i praktycznie nie różni się od błędu modelu skonstruowanego przy użyciu metody hierarchicznej.

### Model produkcji sprzedanej przemysłu opracowany w oparciu o zbiory przybliżone

Kolejny sposób modelowania zjawisk ekonomicznych, jaki został przedstawiony w rozprawie, oparty był na teorii zbiorów przybliżonych. Jest to zupełnie odmienny sposób modelowania, polegający w pierwszej kolejności na maksymalnym zredukowaniu zbioru zmiennych objaśniających, a następnie na stworzeniu reguł opisujących zależności występujące między tymi zmiennymi a zmienną objaśnianą. W przeciwieństwie do sztucznych sieci neuronowych czy też algorytmów genetycznych modelowanie w oparciu o zbiory przybliżone wymaga znajomości rozległego aparatu pojęciowego. Pojęcia, których znajomość jest niezbędna, przedstawione zostały w rozdziale drugim pracy. Pierwszym krokiem modelowania jest przedstawienie zebranego materiału statystycznego, dotyczącego badanego systemu informacyjnego, w postaci tablicy informacyjnej. Kolejny krok polega na przekształceniu danych zawartych w tablicy informacyjnej do formy zakodowanej. W tym przypadku do przeprowadzenia dyskretyzacji zmiennych posłużono się programem *Analiza danych przy użyciu zbiorów przybliżonych*, działającym w środowisku *Matlab 6.0*.

Zakresy wszystkich zmiennych objaśniających zostały podzielone na dwie klasy w zależności od rozkładu danej zmiennej, zakres zaś zmiennej objaśnianej podzielony został na trzy równe przedziały. W wyniku przeprowadzonego kodowania, otrzymano następującą wtórną tablicę informacyjną. Jak już wspomniano, głównym celem analizy danych metodą zbiorów przybliżonych jest maksymalna redukcja zmiennych oraz określenie algorytmu decyzyjnego. Cel ten osiągnięto, realizując następujące etapy:

1. Wyznaczono elementarne zbiory przykładów.
2. Określono koncepty decyzyjne.
3. Dokonano aproksymacji zbiorów.
4. Dokonano aproksymacji rodziny zbiorów.
5. Wyznaczono reduktów względne i bezwzględne.

Z pośród najkrótszych reduktów względnych wybrano redukt nr 9 i w oparciu o niego zbudowano algorytm decyzyjny zawierający 68 reguł, w tym 14 reguł sprzecznych ze względu na atrybut decyzyjny. Następnie dokonano uproszczenia otrzymanych reguł. W wyniku przeprowadzonych uproszczeń ostatecznie otrzymano 7 reguł. Poniżej przedstawiono przykładowo regułę nr 5 i podano jej interpretację ekonomiczną.

#### Reguła nr 5

Jeżeli  $(q_1=2)$  i  $(q_4=2)$  i  $(q_7=1)$  i  $(q_9=2)$  i  $(q_{11}=1)$  i  $(q_{13}=2)$  i  $(q_{14}=2)$  i  $(q_{16}=2)$  i  $(q_{17}=2)$  i  $(q_{18}=1)$  i  $(q_{20}=1)$  to  $(d=3)$ .

Interpretacja ekonomiczna reguły nr 5: jeżeli eksport towarów ogółem  $\in (0,3982-1>$  i przeciętne wynagrodzenie nominalne brutto  $\in (0,5226-1>$  i zasiłki dla bezrobotnych  $\in <0-0,4159>$  i nakłady inwestycyjne  $\in (0,2850-1>$  i wpływy z podatku dochodowego od osób fizycznych  $\in <0-0,6264>$  i przychody z całokształtu działalności  $\in (0,2625-1>$  i koszty uzyskania przychodów z całokształtu działalności  $\in (0,2280-1>$  i dochody budżetu państwa  $\in (0,4167-1>$  i wydatki budżetu państwa  $\in (0,3125-1>$  i wskaźnik cen produkcji sprzedanej przemysłu  $\in <0-0,4054>$  i inflacja miesięczna  $\in <0-0,3846>$ , to wartość produkcji sprzedanej przemysłu  $\in (0,66-1>$ .

Zbiory przybliżone pozwalają nie tylko na określenie, które ze zmiennych są zmiennymi najbardziej istotnymi, ale pozwalają też określić, jakie reguły rządzą zmiennymi w badanym systemie informacyjnym. Dodatkową zaletą zbiorów przybliżonych jest to, że dopuszczają do badania każdy rodzaj zmiennych (również zmienne lingwistyczne).

## 5. Ocena uzyskanych wyników

W rozprawie doktorskiej opracowano sześć modeli *produkcji sprzedanej przemysłu*. Pierwszy model został skonstruowany przy użyciu standardowych metod ekonometrycznych. Z kolei do konstrukcji pozostałych pięciu modeli użyto różnych metod sztucznej inteligencji. Trzy z nich zbudowano w oparciu o sieci RBF, przy czym różniły się one sposobem doboru zmiennych do modelu, czwarty model – w oparciu o metodę łączącą algorytm genetyczny i sieć RBF, a piąty model – w oparciu o zbiory przybliżone. W celu porównania ze sobą wszystkich skonstruowanych modeli zbudowano dodatkowo jeszcze dwa modele neuronowe RBF. Pierwszy model zawierał zmienne objaśniające, wytypowane w trakcie modelowania ekonometrycznego, drugi model zawierał zmienne objaśniające, wytypowane w czasie modelowania zbiorami przybliżonymi. Parametry sieci oraz liczbę neuronów zachowano na identycznym poziomie, jak przy konstrukcji trzech wcześniejszych modeli neuronowych oraz modelu łączącego algorytm genetyczny i sieć RBF. Dla każdego z nowo zbudowanych modeli wyznaczono błąd *średniokwadratowy MSE*.

Następnie dokonano porównania błędów *MSE* wszystkich skonstruowanych modeli *produkcji sprzedanej przemysłu*. Spośród modeli zbudowanych przy użyciu różnych metod sztucznej inteligencji, modelem mającym najmniejszy błąd okazał się model RBF, sprawdzający wszystkie możliwe kombinacje zmiennych objaśniających. Jednak z powodu zbyt długiego czasu oczekiwania na wyniki modelowania, metoda ta jest metodą mało praktyczną. Metodami pozwalającymi osiągnąć podobne rezultaty modelowania (tzn. niewiele wyższy błąd *MSE* modelu), przy wielokrotnie krótszym czasie pracy komputera, są sieci RBF wspierane hierarchiczną metodą doboru zmiennych oraz metoda łącząca algorytm genetyczny z siecią RBF. W tabeli przedstawiono wady i zalety zastosowanych metod na tle kolejnych etapów modelowania (tabela 2). Na podstawie analizy tabeli 2 należy

stwierdzić, że żadna z przedstawionych metod modelowania nie jest pozbawiona wad. Nie istnieje bowiem „metoda idealna”. Jednak bezsprzeczną zaletą metod sztucznej inteligencji jest możliwość modelowania zjawisk ekonomicznych na podstawie dowolnego zbioru zmiennych zarówno numerycznych, jak i jakościowych, rozmytych itp.

Tabela 2. Wady i zalety stosowania metod klasycznych oraz metod sztucznej inteligencji na poszczególnych etapach modelowania

| Etap I | Metody klasyczne   | Metody sztucznej inteligencji   |
|--------|--|---|
| Opis   | Faza I: Ustalenie zbioru kandydatek na zmienne objaśniające oraz zebranie materiału statystycznego   |   |
| Wady   | Konieczność posiadania danych statystycznych spełniających określone wymogi formalne, w zasadzie tylko zmienne numeryczne (precyzyjne), metody nie radzą sobie z danymi niekompletnymi, jakość danych odgrywa znaczącą rolę.   | Wymagana jest duża liczba danych, jakość danych odgrywa znaczącą rolę.  |
| Zalety | Powszechność opisu ilościowego w czasie i przestrzeni procesu identyfikacji czynników determinujących badane zjawisko ekonomiczne (patrz: roczniki statystyczne i in.).  | Każdy rodzaj zmiennych: numeryczne, jakościowe, rozmyte itp., metody radzą sobie z danymi niekompletnymi, nie są potrzebne założenia dotyczące natury rozkładu danych (choć wskazane jest równomierne pokrycie przestrzeni zmiennych danymi). |
| Etap I | Metody klasyczne   | Metody sztucznej inteligencji   |
| Opis   | Faza II: Wybór najlepszej kombinacji zmiennych objaśniających do modelu  |   |
| Wady   | Pracochłonność obliczeń związana ze sprawdzeniem wszystkich możliwych kombinacji zmiennych.  | Pracochłonność obliczeń związanych z uczeniem sieci neuronowych.  |
| Zalety | Pozwala na zredukowanie zbioru zmiennych objaśniających przez „pozbycie się” zmiennych nieistotnych lub współzależnych, a tym samym pozwala na dokładniejsze oszacowanie parametrów strukturalnych modelu, możliwość sprawdzenia, czy dany zbiór zmiennych jest zależny statystycznie. | Nie wymagają wstępnej analizy zebranych kandydatek na zmienne, aczkolwiek analiza taka może spowodować redukcję liczby zmiennych wejściowych, możliwość sprawdzenia wszystkich kombinacji zmiennych i wybór najlepszej.                       |

| Etap II  | Metody klasyczne   | Metody sztucznej inteligencji   |
|----------|--|---|
| Opis     | Konstrukcja modelu   |   |
| Wady     | Zwykle trudno jest znaleźć postać równania, które równocześnie spełnia kryteria merytoryczne doboru zmiennych z równoczesną formalną estymacją zgodności (np. z rozkładem normalnym).                          | W przypadku stosowania sieci MLP i RBF nie można zapisać matematycznie postaci szacowanej funkcji, istotny jest czynnik czasu, jaki można poświęcić na proces uczenia sieci neuronowej, możliwość przeuczenia lub niedouczenia sieci neuronowej.  |
| Zalety   | Jeżeli rozkład danych rzeczywistych jest zgodny z założonym rozkładem, możliwe jest łatwiejsze dobranie struktury modelu, jak również jego skonstruowanie.   | Nie wymaga się, aby relacja między zmienną zależną a zmiennymi niezależnymi była liniowa, nie musi być również znana natura nieliniowości.  |
| Etap III | Metody klasyczne   | Metody sztucznej inteligencji   |
| Opis     | Estymacja parametrów modelu  |   |
| Wady     | Konieczność przyjmowania założeń, których weryfikacja w praktycznych zastosowaniach jest trudna, a czasem wręcz niemożliwa, pracochłonność obliczeń, skonstruowany model może być lokalnie bardzo niedokładny. | Brak spójnych zasad związanych z procesem projektowania, a później uczenia modelu neuronowego, co powoduje, że prace na tym etapie opierają się na metodzie „prób i błędów” (co wynika z nieznaności nieliniowości badanego systemu), utrudniona analiza działań zachodzących w warstwach ukrytych. |
| Zalety   | Otrzymane na tym etapie wartości parametrów strukturalnych pozwalają na dokonanie wstępnej interpretacji ekonomicznej skonstruowanego modelu.  | Możliwość przetwarzania dużej ilości danych w relatywnie krótkim czasie.  |
| Etap IV  | Metody klasyczne   | Metody sztucznej inteligencji   |
| Opis     | Weryfikacja otrzymanego modelu oraz interpretacja ekonomiczna otrzymanych wyników  |   |
| Wady     | Bardzo czasochłonna, wymaga wyznaczenia wielu wskaźników.  | Jedynie przez analizę dopasowania wygenerowanych wartości do danych pomiarowych lub wyznaczenie błędu modelu (MSE, MAE itp.), parametry nieliniowych sieci neuronowych MLP i RBF nie posiadają interpretacji ekonomicznej, otrzymane wyniki nie mogą być uogólniane na populację.                   |

|        |  |  |
|--------|--|--|
| Zalety | Możliwość badania wpływu pojedynczej zmiennej lub łącznego wpływu wszystkich zmiennych objaśniających na zmienną objaśnianą; wyniki informują o rządzących prawach ekonomicznych badanej zbiorowości; wyniki mogą być uogólniane na populację. | Informują, jakie czynniki wpływają na badane zjawisko ekonomiczne, w sytuacji, gdy trudno jest opisać zależności między zmiennymi za pomocą równań liniowych; generują prognozy lepiej dopasowane do danych empirycznych; możliwość tworzenia reguł decyzyjnych, również w oparciu o dane jakościowe; otrzymane modele charakteryzują się niższym błędem; wyznaczenie prognoz wymaga znacznie mniej czasu. |
|--------|--|--|

*Źródło: opracowanie własne na podstawie: A. Goryl, Z. Jędrzejczyk, K. Kukuła, J. Osiewalski, A. Walkosz, Wprowadzenie do ekonometrii w przykładach i zadaniach, PWN, Warszawa 1996; M. Gruszczyński, M. Podgórska, Ekonometria, Oficyna Wydawnicza SGH, Warszawa 2004; J. Hozer [i in.], Ekonometria stosowana z zadaniami, Katedra Ekonometrii i Statystyki Uniwersytetu Szczecińskiego, Szczecin 2005; E. Nowak, Problemy doboru zmiennych do modelu ekonometrycznego, PWN, Warszawa 1994; E. Nowak, Zarys metod ekonometrii – zbiór zadań, PWN, Warszawa 2006; S. Osowski, Sieci neuronowe w ujęciu algorytmicznym, Wydawnictwo Naukowo-Techniczne, Warszawa 1996; A. Piegał, Modelowanie i sterowanie rozmyte, Akademicka Oficyna Wydawnicza Exit, Warszawa 1999; L. Rutkowski, Metody i techniki sztucznej inteligencji, PWN, Warszawa 2005; D. Witkowska, Sztuczne sieci neuronowe w analizach ekonomicznych, Wydawnictwo Menadżer, Łódź 2000.*

## Podsumowanie

Przedstawione w dysertacji badania koncentrowały się na analizie stosowności metod sztucznej inteligencji do identyfikacji czynników ekonomicznych, wpływających na dane zjawisko ekonomiczne. Otrzymane na podstawie przeprowadzonych badań wyniki pozwoliły na potwierdzenie założonej hipotezy badawczej, mówiącej o tym, że zastosowanie metod sztucznej inteligencji do modelowania produkcji sprzedanej przemysłu stanowi znaczące uzupełnienie przeprowadzonych badań poznawczych przy użyciu metod klasycznych (standardowych). Dzięki bowiem zastosowanym metodom sztucznej inteligencji zidentyfikowano dodatkowe czynniki, mające istotny wpływ na produkcję, a które zostały pominięte w trakcie modelowania ekonometrycznego.

Wkład własny autorki rozprawy stanowi:

- zbudowanie trzech modeli produkcji sprzedanej przemysłu w oparciu o sztuczne sieci neuronowe RBF,
- zbudowanie modelu produkcji sprzedanej przemysłu w oparciu o zastosowanie algorytmu genetycznego wraz z siecią RBF,
- zbudowanie modelu produkcji sprzedanej przemysłu przy użyciu zbiorów przybliżonych,
- zbudowanie modelu produkcji sprzedanej przemysłu w oparciu o klasyczne metody modelowania,
- dokonanie analizy porównawczej zastosowanych metod modelowania.

Otrzymane wyniki dostarczyły dostatecznie wielu dowodów, aby stwierdzić, że opisane w rozprawie metody sztucznej inteligencji mogą być z powodzeniem stosowane w praktyce jako skuteczne narzędzie analiz ekonomicznych.

## Literatura

- Domański Cz., Pruska K., *Nieklasyczne metody statystyczne*, PWE, Warszawa 2000.
- Goryl A., Jędrzejczyk Z., Kukuła K., Osiewalski J., Walkosz A., *Wprowadzenie do ekonometrii w przykładach i zadaniach*, PWN, Warszawa 1996.
- Gruszczyński M., Podgórska M., *Ekonometria*, Oficyna Wydawnicza SGH, Warszawa 2004.
- Hozer J. [i in.], *Ekonometria stosowana z zadaniami*, Katedra Ekonometrii i Statystyki Uniwersytetu Szczecińskiego, Szczecin 2005.
- Nowak E., *Problemy doboru zmiennych do modelu ekonometrycznego*, PWN, Warszawa 1984.
- Nowak E., *Zarys metod ekonometrii – zbiór zadań*, PWN, Warszawa 2006.
- Osowski S., *Sieci neuronowe w ujęciu algorytmicznym*, Wydawnictwo Naukowo-Techniczne, Warszawa 1996.
- Piegat A., *Modelowanie i sterowanie rozmyte*, Akademicka Oficyna Wydawnicza Exit, Warszawa 1999.
- Rutkowski L., *Metody i techniki sztucznej inteligencji*, PWN, Warszawa 2005.
- Witkowska D., *Sztuczne sieci neuronowe w analizach ekonomicznych*, Wydawnictwo Menedżer, Łódź 2000.

## **NON-STANDARD METHODS FOR IDENTIFICATION OF FACTORS DETERMINING SOLD PRODUCTION OF INDUSTRY**

### **Summary**

The main aim of the dissertation was the analysis of applicability of non-standard modelling methods, such as neural networks, genetic algorithms and rough sets, for modelling economic phenomena on the example of sold production of industry. As a part of conducted research there were constructed five models of sold production of industry based on the artificial intelligence methods and one econometric model. Conclusion received as a result of carried out research enable to prove the assumed research hypothesis, that applying the artificial intelligence to modelling sold production of industry is a significant supplementation for research conducted on the base of classic (standard) methods. Thanks to applied methods of the artificial intelligence, there were identified additional factors having essential influence on the production, which were omitted during econometric modelling.

Methods of artificial intelligence presented in the dissertation can be used to describe compound economic, technical or medical phenomena. These methods deal very well with identification of the most essential factors influencing the analyzed phenomena and by this the construction of models of these phenomena is considerably facilitated.

*Translated by Barbara Wąsikowska*