

NIESTANDARDOWE METODY IDENTYFIKACJI CZYNNIKÓW DETERMINUJĄCYCH PRODUKCJĘ SPRZEDANĄ PRZEMYSŁU

Barbara Wąsikowska

Instytut Informatyki w Zarządzaniu,
WNEiZ Uniwersytetu Szczecińskiego
e-mail: barbara.wasikowska@wneiz.pl

Streszczenie: w artykule przedstawione zostały niestandardowe metody identyfikacji czynników ekonomicznych determinujących *produkcję sprzedaną przemysłu* takie jak: sztuczna sieć neuronowa RBF oraz algorytm genetyczny. Obecnie do identyfikacji czynników mających wpływ na dane zjawisko ekonomiczne stosuje się metody standardowe tj. metody statystyczne i ekonometryczne. Jednak metody te nie są w stanie w pełni odzwierciedlić otaczającą człowieka coraz bardziej złożoną rzeczywistość ekonomiczną. W związku z tym stale poszukuje się nowych metod identyfikacji.

Słowa kluczowe: identyfikacja, sztuczna inteligencja, sieci neuronowe RBF, algorytmy genetyczne.

PROBLEM IDENTYFIKACJI ZMIENNYCH DO MODELU

Rozwój cywilizacji, postępu technicznego i gospodarki światowej prowadzi do powstawania coraz bardziej skomplikowanej rzeczywistości ekonomicznej w otoczeniu człowieka. Zasadniczym celem pracy badacza-ekonomisty jest więc poznanie i w możliwie prosty i najdokładniejszy sposób opisanie i wyjaśnienie zjawisk ekonomicznych. Są one jednak zazwyczaj zjawiskami nieliniowymi, cechującymi się dużą liczbą zmiennych, względnie małą liczbą danych pomiarowych oraz silną, często nieliniową korelacją zmiennych. W zjawiskach tych działają różnorodne prawa ekonomiczne, zależne od warunków ekonomicznych i polityki gospodarczej państwa. Bardzo ważnym problemem jest również wpływający czas. Model opisujący zjawisko ekonomiczne dla danej rzeczywistości może zupełnie nie odpowiadać stanowi przeszłemu, bowiem czynniki mające zasadniczy wpływ na to zjawisko w badanym okresie w innym czasie mogą nie mieć tak istotnego wpływu lub nawet w ogóle nie występować. Wśród innych przyczyn powodujących ogromne trudności w opisanu otaczającej nas rzeczywistości ekonomicznej można wymienić takie jak [Piegat A. 1999]:

- wpływ ludzi (rządów, ministrów, ekspertów itp.) na aktualne zależności ekonomiczne istniejące w kraju,
- istnienie wzajemnych, często bardzo skomplikowanych i trudnych do zdefiniowania powiązań między czynnikami opisującymi badane zjawisko,
- występowanie obok informacji precyzyjnej, również informacji jakościowej i rozmytej,
- możliwość występowania błędów lub luk w zebranych materiale statystycznym,
- brak możliwości zmierzenia pewnych wielkości lub brak informacji na temat kształtowania się tych wielkości w pewnych okresach wcześniejszych,
- wystąpienie procesów, które nie zostały dotychczas zidentyfikowane.

Dokonując przeglądu literatury można zauważyć, że najczęściej stosowanymi metodami służącymi do identyfikacji czynników istotnie wpływających na badane zjawisko są metody klasyczne w postaci metod ilościowych - statystycznych i ekonometrycznych. Używając pojęcia „metody klasyczne” autorka niniejszego artykułu ma na myśli 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. By móc rzetelnie przedstawić możliwości stosowania tych metod do opisu problemów makroekonomicznych autorka niniejszego artykułu zdecydowała się zastosować je do budowy modelu *produkcji sprzedanej przemysłu*, która jest jednym z głównych problemów polskiej gospodarki. Poniżej zaprezentowane zostały dwie metody sztucznej inteligencji mogące służyć do identyfikacji czynników istotnie wpływających na *produkcję sprzedaną przemysłu* tj. sztuczną sieć neuronową RBF oraz algorytm genetyczny.

BADANIA IDENTYFIKACYJNE Z WYKORZYSTANIEM SZTUCZNEJ SIECI NEURONOWEJ RBF

Niezależnie od tego jaka metoda zostanie użyta do opisu badanego zjawiska w pierwszym kroku należy ustalić zbiór kandydatek na zmienne objaśniające. Główną przesłanką, decydującą o wyborze zmiennych, powinna być ich merytoryczna wartość przy jednoczesnym uwzględnieniu celu badania [Nowak. 1984]. Często w doborze kandydatek na zmienne posługuje się również metodą ekspercką, w której wykorzystuje się doświadczenia specjalistów (teoretyków i praktyków) z dziedziny będącej przedmiotem modelowania [Hozer i in. 2005]. Jednak przede wszystkim, należy kierować się istniejącą teorią ekonomii.

Poszukując zmiennych potencjalnie wpływających na wielkość *produkcji sprzedanej przemysłu* w Polsce, poza wskazaniami płynącymi z ogólnych teorii

ekonomii, oparto się m.in. na wynikach badań przeprowadzonych przez *Instytut Badań nad Gospodarką Rynkową*, wynikach przedstawionych w poszczególnych numerach *Monitora Ekonomicznego PSE S.A.*, *Biuletynu Ekonomicznego Ministerstwa Spraw Zagranicznych*, *Makro Trendów* wydawanych przez BGŻ S.A., raportów umieszczonych na stronie internetowej *eGospodarka.pl* oraz na podstawie analiz opracowywanych przez *Centrum Analiz Społeczno-Ekonomicznych*. Wytypowanych zostało dwadzieścia jeden zmiennych potencjalnie wpływających na produkcję sprzedaną przemysłu: eksport towarów ogółem, import towarów ogółem, saldo eksport import, przeciętne wynagrodzenie nominalne brutto, pracujący w sektorze przedsiębiorstw ogółem, bezrobotni zarejestrowani ogółem, zasiłki dla bezrobotnych, podaż pieniądza ogółem, nakłady inwestycyjne, wpływy z podatku dochodowego od osób fizycznych, kurs dolara, przychody z całokształtu działalności, koszty uzyskania przychodów z całokształtu działalności, wynik budżetu państwa, dochody budżetu państwa, wydatki budżetu państwa, wskaźnik cen produkcji sprzedanej przemysłu, wskaźnik cen transakcyjnych eksportu, inflacja miesięczna, liczba ludności.

Kolejnym krokiem w prowadzonych badaniach było wybranie spośród kandydatek na zmienne objaśniające tych, które mają istotny wpływ na produkcję oraz skonstruowanie modeli *produkcji sprzedanej przemysłu* przy użyciu niestandardowych metod identyfikacji zmiennych jakimi są metody sztucznej inteligencji.

Jako pierwszy zostanie przedstawiony model neuronowy RBF sprawdzający wszystkie możliwe zestawy zmiennych objaśniających. Liczbę wszystkich możliwych kombinacji zmiennych wyliczono na podstawie wzoru dwumianowego Newtona (z pominięciem oczywiście zbioru pustego, czyli braku jakichkolwiek istotnej zmiennej objaśniającej): $2^n - 1$. Zatem liczba wszystkich możliwych kombinacji, przy dwudziestu jeden kandydatkach na zmienne objaśniające, równa się: $2^{21} - 1 = 2\,097\,151$. W trakcie budowy modelu pojawił się problem ustalenia odpowiedniej ilości neuronów w warstwie ukrytej sieci neuronowej. W celu wyjaśnienia jaka liczba neuronów byłaby odpowiednia, postanowiono stworzyć po dziesięć modeli dla sieci zawierających od jednego do dwudziestu neuronów w warstwie ukrytej. Tak więc, skonstruowano dziesięć modeli dla sieci z jednym neuronem w warstwie ukrytej, dziesięć modeli dla sieci z dwoma neuronami w warstwie ukrytej itd. W kolejnym kroku każda z sieci była trenowana na zbiorze próbek uczących przez 100 epok, a następnie testowana na zbiorze próbek walidujących. 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 błędów na zbiorze walidującym zaczynają się „wahać”, tzn. raz rosną a raz spadają, co może

świadczą 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*.

Ostatecznie do przeprowadzenia badania użyto sieci zawierającej osiem neuronów w warstwie ukrytej. Zadaniem sieci było sprawdzenie wszystkich możliwych kombinacji oraz policzenie błędu dla każdego z otrzymanych modeli według wzoru:

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}, \quad (1)$$

gdzie:

y_i - rzeczywista wartość zmiennej objaśnianej dla próbki i ,

\hat{y}_i - wartość zmiennej objaśnianej dla próbki, wyznaczona podstawie modelu

n - liczba obserwacji.

a następnie, podanie najlepszej kombinacji wejść dla każdego z typu modeli, czyli dla modelu zawierającego jedno wejście, dwa wejścia itd., aż do modelu zawierającego wszystkie możliwe zmienne objaśniające. Najlepszym modelem (tzn. modelem o najmniejszym błędzie) okazał się model zawierający dziesięć wejść: $X_1, X_4, X_5, X_9, X_{14}, X_{16}, X_7, X_{18}, X_{20}, X_{21}$. Wybrane spośród wszystkich kandydatek zmienne są tymi czynnikami, które wywierają największy wpływ na *produkcję sprzedaną przemysłu*.

Błąd średniokwadratowy modelu *produkcji sprzedanej przemysłu*, zbudowanego w oparciu o wybrane zmienne objaśniające, wyniósł: $MSE = 0,0496$. Poniżej przedstawiono 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_i - wagi,

x_j - zmienne objaśniające,

c_{ij} - punkt stanowiący centrum funkcji radialnej,

d - szerokość funkcji.

Użyta w badaniu metoda sztucznych sieci neuronowych RBF pozwoliła na zidentyfikowanie czynników w sposób istotny wpływających na badaną produkcję. Niestety, poważną wadą tej metody jest jej bardzo duża czasochłonność.

BADANIA IDENTYFIKACYJNE Z WYKORZYSTANIEM ALGORYTMU GENETYCZNEGO

Do budowy drugiego modelu *produkcji sprzedanej przemysłu* postanowiono użyć algorytmu genetycznego, gdyż obok sztucznych sieci neuronowych, jest to drugie pod względem popularności narzędzie sztucznej inteligencji stosowane 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” - którego wartość błędu była najwyższa), po czym ¼ 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 ma drodze reprodukcji populację poddawano mutacji. Z populacji losowano pewną liczbę bitów, która ma ulec mutacji. Procesowi mutacji nie ulegał jednak osobnik, którego błąd średniokwadratowy był najniższy. Po wylosowaniu indeksów bitów (czyli miejsc, które mają ulec mutacji) następowała mutacja polegająca na zamianie 0 na 1 i odwrotnie.
5. Całą procedurę (tzn. od selekcji) powtarzano 100 razy.
6. Na końcu generowano najlepszego osobnika zapisanego w postaci chromosomu oraz błąd otrzymanego modelu.

W wyniku działania algorytmu genetycznego otrzymano chromosom składający się z 21 genów: 100110101001011001010. Geny, których wartość jest równa jedności wskazują na to, które ze zmiennych objaśniających zostały wybrane do modelu *produkcji sprzedanej przemysłu* i w sposób istotny na nią wpływają. Błąd średniokwadratowy otrzymanego modelu wyniósł: $MSE = 0,0553$ i był on większy tylko o 0,0057 od błędu modelu otrzymanego przy użyciu sieci RBF sprawdzającej wszystkie możliwe kombinacje wejść, natomiast czas oczekiwania na wyniki uległ znacznemu skróceniu.

WNIOSKI Z PRZEPROWADZONYCH BADAŃ

Przedstawione w niniejszym artykule badania koncentrowały się na analizie stosowalności sztucznych sieci neuronowych RBF oraz algorytmów genetycznych do identyfikacji czynników ekonomicznych, wpływających na *produkcję sprzedaną przemysłu*. Poniżej zestawiono błędy obu skonstruowanych modeli oraz zestawy zmiennych wybranych w trakcie modelowania przy użyciu poszczególnych metod (Tabela 1).

Tabela 1. Zestawienie wyników efektów identyfikacji.

Rodzaj modelu	Błąd modelu	Wybrane zmienne
RBF	0,0496	X ₁ , X ₄ , X ₅ , X ₉ , X ₁₆ , X ₁₈ , X ₁₉ , X ₂₀ , X ₂₂ , X ₂₄
Alg. genetyczny + RBF	0,0553	X ₁ , X ₄ , X ₅ , X ₇ , X ₉ , X ₁₄ , X ₁₆ , X ₁₇ , X ₂₀ , X ₂₂

Źródło: obliczenia własne

Na podstawie otrzymanych wyników można stwierdzić, że zastosowanie metod sztucznej inteligencji pozwala na zidentyfikowanie czynników mających istotny wpływ na *produkcję sprzedaną przemysłu*. Metody sztucznej inteligencji mogą więc być z powodzeniem stosowane - obok metod klasycznych do identyfikacji czynników mających wpływ na badane zjawisko ekonomiczne.

LITERATURA

- Arabas J., (2001) Wykłady z algorytmów ewolucyjnych, Wydawnictwo Naukowo-Techniczne, Warszawa
- Goldberg D. E., (1998) Algorytmy genetyczne i ich zastosowania, Wydawnictwo Naukowo-Techniczne, Warszawa
- Hozer J. i in., (2005) Ekonometria stosowana z zadaniami, Katedra Ekonometrii i Statystyki Uniwersytetu Szczecińskiego, Szczecin
- Nowak E., (1984) Problemy doboru zmiennych do modelu ekonometrycznego, PWN, Warszawa
- Osowski S., (1996) Sieci neuronowe w ujęciu algorytmicznym, Wydawnictwo Naukowo-Techniczne, Warszawa
- Piegat A., (1999) Modelowanie i sterowanie rozmyte, Akademicka Oficyna Wydawnicza EXIT, Warszawa
- Rutkowski L., (2005) Metody i techniki sztucznej inteligencji, PWN, Warszawa

Non-standard Methods for Identification of Factors Determining Sold Production of Industry

Summary: The paper presents non-standard methods for identification of economic factors determining *sold production of industry*, such as: radial basis network (RBF) and the genetic algorithm. Nowadays for identification of factors having influence on particular economic phenomenon there are used standard methods, i.e. statistical and econometric methods. However, these methods are not able to reflect fully more and more compound reality surrounding the man. Therefore new methods of identification are constantly seeking.

Key words: identification, artificial intelligence, radial basis networks, genetic algorithms