

Jędrzej Trajer
Katedra Podstaw Inżynierii
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

SZTUCZNE SIECI NEURONOWE W MODELOWANIU PROCESÓW Z OGRANICZONYM ZBIOREM DANYCH W INŻYNIERII ROLNICZEJ

Streszczenie

Celem pracy jest przedstawienie metody neuronowego modelowania procesów z ograniczonym zbiorem danych. W przykładzie wykorzystano bazę danych zmian cen przechowywanej marchwi. Podano koncepcję budowy modelu neuronowego, który pomimo ograniczonego zbioru danych posiadać może zadowalające własności uogólniające, w sensie rozszerzenia zasięgu jego stosowalności poza zbiór uczący.

Słowa kluczowe: sztuczne sieci neuronowe, modelowanie, przechowalność

Wprowadzenie

Znaczące zainteresowanie w zakresie wykorzystania sztucznych sieci neuronowych wynika z ich charakterystycznych właściwości. Posiadają zdolności aproksymacji dowolnego nieliniowego odwzorowania z zadaną dokładnością, jak również charakteryzują się zdolnościami do uogólniania zdobytej podczas uczenia wiedzy w zakresie modelowanego systemu. W literaturze nie zawsze dostatecznie dużo miejsca poświęca się uwagę tworzeniu neuronowych modeli dla mało licznych zbiorów uczących, dla których niejednokrotnie budowane sieci nie spełniają własności uogólniających. Jakość utworzonego modelu zależy bowiem zarówno od właściwego doboru struktury sieci, przeprowadzenia skutecznego uczenia, jak i liczności użytego zbioru danych w procesie uczenia. Tworzenie modeli neuronowych nie jest zadaniem łatwym, wymaga żmudnych badań i doświadczenia. Wynika to z trudności, jakie stwarza formalna analiza tego efektywnego narzędzia. Przedstawiane w literaturze wytyczne realizacji sieci odnoszą się bądź do wybranych przypadków, bądź są często zbyt ogólne. Poniżej, na przykładzie prognozy cen marchwi, przedstawiono wpływ zbioru danych na charakterystykę tworzonych

modeli. Znajomość występujących zależności decyduje o skuteczności budowy modeli, ułatwia postępowanie i prowadzi do szybkiego znalezienia efektywnych rozwiązań.

Metodyka tworzenia modeli z ograniczonym zbiorem danych

Modele neuronowe zalicza się do modeli indukcyjnych, które wykorzystuje się w celu rozwiązania problemów związanych z dużą złożonością modelowanego systemu i trudnościami w opisie jego dynamiki, w oparciu o uznane prawa teorii. Przydatność modelu określają miary jakości, opisujące zdolność do aproksymacji zbioru danych uczących i zdolność do uogólnień, to znaczy prawidłowego wyznaczenia wartości danych nie uczestniczących w procesie uczenia (zbiór testowy). Na zdolności te ma wpływ zarówno struktura sieci, jak i zbiór danych.

Zdolności sieci neuronowych do aproksymacji określa twierdzenie *Hecht-Nielsen*: Każda funkcja ciągła może być aproksymowana z dowolnie dużą dokładnością przez sieć neuronową posiadającą n neuronów w warstwie wejściowej, $(2n+1)$ neuronów w warstwie ukrytej o odpowiednio dobranych nieliniowych funkcjach aktywacji i jednym neuronie w warstwie wyjściowej o funkcji przejścia.

Natomiast zdolność sieci do uogólnień określa miara *VCdim* (*Vapnika-Chervonenkisa*), zdefiniowana jako liczebność zbioru danych uczących, które mogą zostać bezbłędnie odtworzone we wszystkich możliwych konfiguracjach dla danej topologii sieci. Nie istnieje prosty związek między architekturą sieci wielowarstwowej a miarą *VCdim*, w literaturze [Hush i in. 1993] podaje się oszacowanie jej granic, w formie następującej zależności:

$$2 \cdot \left\lceil \frac{K}{2} \right\rceil \cdot N \leq VC \dim \leq 2 \cdot N_w (1 + \log N_n), \quad (1)$$

gdzie:

N , K , N_n – odpowiednio liczba neuronów w warstwie wejściowej, ukrytej i całkowita ich liczba w sieci,

N_w – całkowita liczba wag w sieci,

$\lceil \]$ – część całkowita.

Dolna granica przedziału jest w przybliżeniu równa liczbie wag łączących warstwę wejściową z warstwą ukrytą, natomiast górna jest większa od dwukrotnej liczby wszystkich wag sieci. Nawet szacunkowe przyjęcie tej miary umożliwia ocenę przydatności zbioru danych w stosunku do wybranej topologii sieci. Wytyczne, oparte na badaniach, przedstawiane przez różnych badaczy [Vapnik 1998, Hush i in. 1993] wskazują, że:

liczebność próbek uczących, zapewniających dobre zdolności uogólniające sieci, powinna kształtować się w zakresie od dziesięcio- do dwudziestokrotności miary VCdim.

W zagadnieniach charakteryzujących się małą nieliniowością dopuszcza się też mniej liczne zbiory, jednak o liczebności przypadków nie mniejszej od liczby wszystkich połączeń w sieci. Jak widać, wybór topologii sieci, tzn. jej stopnia złożoności, wymaga kompromisu, jakiego należy dokonać w zastosowanej liczbie neuronów, tak aby model odznaczał się równocześnie dobrymi zdolnościami do aproksymacji i do generalizacji. Kluczowym jest tu, przede wszystkim, określenie liczby neuronów w warstwie ukrytej. Dobrym oszacowaniem, opartym na teorii *Kołomogorowa*, jest liczba neuronów ukrytych K będąca średnią geometryczną liczby wejść N i wyjść M sieci:

$$K = \sqrt{NM} \quad (2)$$

Najczęściej przyjmuje się taką liczbę neuronów ukrytych, a następnie przeprowadza się ich redukcję do osiągnięcia jak najlepszych miar jakości sieci. W przypadku mało licznych zbiorów danych, postępowanie lepiej zacząć od jak najmniejszej ich liczby i stopniowo ją zwiększać.

Osobny problem stanowi wymiar zbioru danych, który determinuje liczbę neuronów w warstwie wejściowej. Redukcja wymiaru przestrzeni danych prowadzi także do uproszczenia architektury sieci. Celem metod redukcji jest zastąpienie pierwotnego zestawu zmiennych nowym, o zmniejszonej liczbie zmiennych wejściowych. Sieci posiadające niewielkie rozmiary mają istotne zalety, do których zaliczyć należy: zwiększoną szybkość uczenia, mniejszą wymaganą liczbę zbiorów danych i większe zdolności generalizujące przy niezmiennych rozmiarach zbiorów danych.

Skuteczność uczenia zależy od właściwego wyboru zbioru walidacyjnego. Przy ograniczonym zbiorze danych jego liczebność jak i wybór nie jest bez znaczenia. Wytyczne co do liczebności tego zbioru, które podawane są w literaturze nie dają dobrych rezultatów uczenia ze względu na ich zbyt duży rozmiar. Dobre wyniki można już uzyskać przy liczebności tego zbioru 10-15% zbioru uczącego, pod warunkiem odpowiedniego doboru tych przypadków. Wybór ten każdorazowo można oceniać poprzez sprawdzenie statystyk regresyjnych odpowiadającej temu przypadkowi sieci liniowej. Do oceny własności uogólniających modelu lepiej jest wybrać zbiór testowy ciągu kolejnych danych (nie biorących udział w uczeniu sieci), dla których obliczamy globalny błąd względny niż sprawdzać statystyki regresyjne danych testowych wybranych przypadkowo. Globalny błąd względny modelu opisuje zależność:

$$\delta(D)_{gl} = \sqrt{\frac{\sum_{i=1}^n (y_i - d_i)^2}{\sum_{i=1}^n y_i^2}}, \quad (3)$$

gdzie:

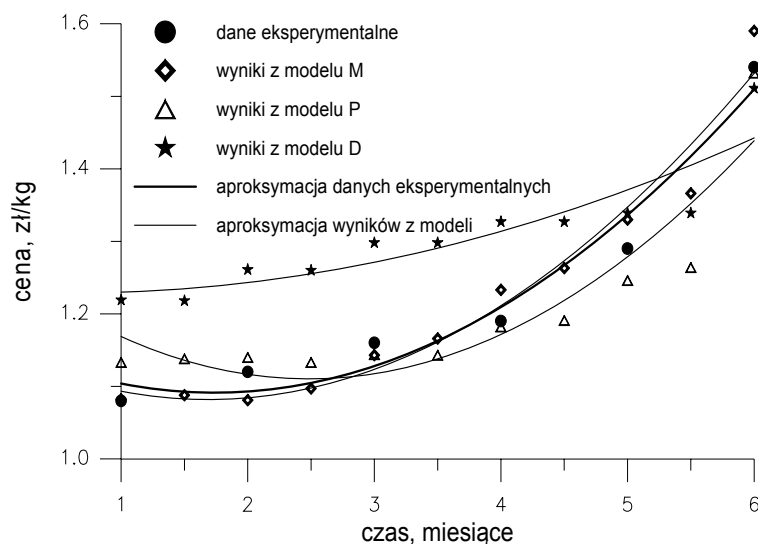
- y_i – wartości zadane danych wyjściowych,
- d_i – wartości wyznaczone za pomocą sieci,
- n – liczba danych.

Do oceny prognoz stosuje się także współczynnik F^2 Theila [Duch i in. 2000], który informuje także o przyczynach rozbieżności pomiędzy wartościami otrzymanymi za pomocą sieci d_i a zadanymi y_i : Wartość jego jest równa kwadratowi globalnego błędu względnego. Składniki: I_1^2 , I_2^2 , I_3^2 tego współczynnika opisują odpowiednio - obciążenie prognozy, błąd spowodowany niedostateczną elastycznością prognozy oraz błąd związany z niedostateczną zgodnością kierunku zmian prognozy z wyznaczaną wartością. Miara ta może jeszcze lepiej charakteryzować własności uogólniające modelu.

Przykłady wpływu topologii sieci na jakość modelu

Ilustrację wpływu ograniczonego zbioru danych na charakterystykę neuronowego modelu przedstawiono na przykładzie prognozowania miesięcznych cen detalicznych marchwi. Zbiór ten składał się z 78 przypadków danych, który dla potrzeb uczenia i weryfikacji sieci podzielono na trzy podzbiory: uczący, walidacyjny i testowy. Otrzymane wartości błędów, pomiędzy zadaną zmienną wyjściową a pochodzącą z modelu, wskazują na jego zdolności aproksymujące (zbiór uczący) i uogólniające (zbiór testowy).

Przebadano trzy neuronowe modele szeregów czasowych o następujących topologiach sieci (odpowiednio: liczba neuronów w warstwach wejściowej, ukrytej, wyjściowej): 6-3-1 model „M”, 6-2-1 model „P” i 6-1-1 model „D”. Ilustrację wpływu liczby neuronów w warstwie ukrytej na zdolności uogólniające sieci neuronowej pokazano na rysunku 1. Wykresy te uzyskane na zbiorze testowym, świadczą o najlepszych własnościach uogólniających modelu z optymalnie dobraną liczbą neuronów ukrytych, model „P”. Właściwości aproksymujące modeli oceniono oparciu o statystyki regresyjne tych modeli, tabela 1. Przyjęto, że *średnia M.B.* - wartość średnia z modułów błędów (zadanych i obliczonego), *iloraz O.S.* - iloraz odchylenia standardowego błędów i odchylenia standardowego (dla bardzo dobrego modelu osiąga wartość poniżej 0,1), *korelacja* - współczynnik korelacji Pearsona, liczony pomiędzy zadanymi i obliczonymi wartościami zmiennej objaśnianej.



Rys. 1. Dane eksperymentalne i wyniki z poszczególnych modeli (M), (P) i (D)
 Fig. 1. Experimental data and results of different models (M), (P) and (D)

Tabela 1. Podstawowe statystyki regresyjne poszczególnych modeli szeregu czasowego

Table 1. Basic statistics of particular models of time series

Miara jakości	Zbiór uczący			Zbiór walidacyjny		
	„D”	„M”	„P”	„D”	„M”	„P”
Śre. M.B.	0,1652	0,0981	0,0897	0,2169	0,1116	0,1399
Ilor. O.S.	0,5181	0,3241	0,2972	0,5198	0,3128	0,3277
Korel.	0,8560	0,9461	0,9549	0,8791	0,9562	0,9503

Przedstawione miary jakości wskazują, że odpowiadające sobie zbliżone wartości poszczególnych statystyk świadczą o reprezentatywności wyboru zbioru walidacyjnego i poprawności przeprowadzenia procesu uczenia. Iloraz O.S. powinien być mniejszy od jedności, dla bardzo dobrych modeli jest z przedziału od 0 do 0,1. Uzyskane miary świadczą o dobrych własnościach aproksymujących modelu rozbudowanego, model „M” i i modelu z właściwie dobraną liczbą neuronów w warstwie ukrytej, model „P”, a tylko dostatecznych modelu o topologii zbyt uproszczonej, model „D”. Ocenę własności uogólniających modeli stanowi wartość globalnego błędu modelu, jaki oblicza się dla danych nie biorących w procesie

tworzenia sieci. Przedstawiono tu jako bardziej reprezentatywny wskaźnik oceny wartości składowe współczynnik I^2 Theila [Duch 2000], tabela 2. Składnikami tego współczynnika są:

$$I^2 = I_1^2 + I_2^2 + I_3^2 \quad (4)$$

odpowiednio reprezentujące: obciążenie prognozy, błąd spowodowany niedostateczną elastycznością prognozy oraz błąd związany z niedostateczną zgodnością kierunku zmian prognozy z kierunkiem zmian wartości prognozowanej. W tabeli podano także wyniki dla modelu opartego na głównych składowych (wybrane zmienne wejściowe) o topologii 4-2-1. Przedstawione wyniki wskazują na dużo lepsze własności modelu głównych składowych niż szeregu czasowego, wartość całkowita współczynnika Theila jest dużo mniejsza dla tego modelu. Wynika to z dwóch powodów: po pierwsze uzyskano strukturalnie prostszy model, a po drugie istnieją o wiele większe jakościowe możliwości opisu dynamiki badanego procesu. Model ten posiada przez to lepsze własności uogólniające w zakresie modelowanego systemu, na co wskazuje mały błąd spowodowany przez składnik I_2^2 charakteryzujący elastyczność prognozy. Dużą zaletą strukturalnie uproszczonego modelu jest możliwość modelowania zjawisk w oparciu o mniejsze zbiory danych, co jest szczególnie cenne w przypadku procesu przechowywania. Charakteryzuje się on bowiem dużą złożonością, a przy tym małą powtarzalnością warunków, w jakich jest przeprowadzany.

Tabela 2. Procentowy udział poszczególnych składowych we współczynniku jakości Theila dla modelu szeregu czasowego i głównych składowych z właściwą liczbą neuronów w warstwie ukrytej

Table 2. Percentage share of individual components of Theil quality index for the model of time series and main components with proper number of neurons in hidden layer

Jakość modelu	I_1^2		I_2^2		I_3^2		I^2
		%		%		%	
Szereg czasowy	0,00436	7,8	0,04836	86,6	0,00311	5,6	0,05583
Główne Składowe	0,00120	78,4	0,00006	3,7	0,00028	17,9	0,00154

Podsumowanie

Wybór struktury modelu jest jednym z głównych problemów w badaniach i zastosowaniach sieci neuronowych. Przy ich wyborze najczęściej stosuje się metodę prób i błędów lub korzysta z różnego rodzaju przesłanek, a nawet wygenerowuje

się je automatycznie. Postępowanie takie może być nieefektywne, zwłaszcza jeśli dysponuje się, do tworzenia modeli, mało licznymi zbiorami danych. W takich przypadkach, poszukiwanie adekwatnej struktury można zawęzić poprzez oszacowanie maksymalnej możliwej liczby neuronów w warstwie ukrytej, na podstawie miary $VCdim$. Oszacowanie to, w fazie analizy wstępnej dla wybranego zbioru danych, umożliwia również zmniejszenie struktury sieci poprzez redukcję wymiaru głównych składowych (zmiennie wejściowe), ograniczając w ten sposób przestrzeń poszukiwań. Na skuteczność uczenia i jakość modelu ma też wpływ odpowiedni wybór zbioru walidacyjnego. Ocenę modelu i jego własności uogólniających najlepiej jest dokonać obliczając błąd względny odpowiednio wybranych danych testowych lub dla prognoz współczynnik Theila.

Bibliografia

- Heht-Nielsen R. 1991. Neurocomputing. Addison Wesley, Amsterdam.
- Hush D., Horne B. 1993. Progress in supervised neural networks. IEEE Signal Processing Magazine, January 93.
- Osowski S. 1996. Sieci neuronowe w ujęciu algorytmicznym. Wydawnictwo Naukowo Techniczne, Warszawa.
- Praca zbiorowa (red. Duch W., Korbicz J., Rutkowski L., Tadeusiewicz R.) 2000. Sieci neuronowe. Biocybernetyka i Inżynieria Biomedyczna, tom 6. Akademicka Oficyna Wydawnicza Exit, Warszawa.
- Tadeusiewicz R. 1998. Elementarne wprowadzenie do techniki sieci neuronowych z przykładowymi programami. Akademicka Oficyna Wydawnicza PLJ, Warszawa.
- Vapnik V. 1998. Statical learning theory. John Wiley and Sons, New York.

**NEURAL NETWORKS
IN MODELING AGRICULTURAL ENGINEERING PROCESSES
WITH LIMITED DATE FILE**

Summary

In this paper the analysis of the neural modeling of the agricultural engineering process was presented. The problems of effectiveness and quality neural networks in these processes was discussed.

Key words: artificial neural networks, modeling, agricultural engineering processes