

EXPECTATION CROPS OF CHOSEN AGRICULTURAL FETUSES WITH THE HELP OF NEURAL MODEL BY TIME SERIES

Summary

Prediction becomes a very important stage in many activities. In case of expectation crops of chosen agricultural fetuses we deal with a number of stimuli which consequently transform into the end effect. It is clear that the quality of those predictions has a great influence on subsequent stages in the production and distribution chain of agricultural fetuses.

Neural networks by time series are a sophisticated technique of modeling capable of reflecting very complex functions. In time series problems, the objective is to predict ahead the value of a variable which varies in time, using previous values of that and/or other variables. The time series training data set therefore typically has a single variable, and this has type input/output (i.e., it is used both for network input and network output).

PROGNOZOWANIE PŁONÓW WYBRANYCH PŁODÓW ROLNYCH Z WYKORZYSTANIEM MODELI NEURONOWYCH W POSTACI SZEREGÓW CZASOWYCH

Streszczenie

Jednym z ważnych etapów badania oraz analizy systemów empirycznych jest proces prognozowania, mający praktyczne zastosowanie w szerokim zakresie działalności ludzkiej. W przypadku przewidywania wielkości plodów rolnych mamy do czynienia z szeregiem złożonych bodźców, które w efekcie przekładają się na wynik końcowy, jakim jest plon. Jakość tych prognoz ma ogromne znaczenie dla kolejnych etapów w łańcuchu produkcyjno-dystrybucyjnym plodów rolnych.

Sieci neuronowe w postaci szeregów czasowych są wysublimowaną techniką modelowania, zdolną odwzorować bardzo złożone funkcje. Celem analizy szeregów czasowych jest ustalenie prognozy przyszłych wartości pewnej zmiennej, której wartości zmieniają się w czasie. Najczęściej dąży się do obliczenia prognozy korzystając z wcześniejszych wartości tej samej zmiennej, której wartość ma być przewidywana. Zbiór uczący, wykorzystywany przy neuronowej analizie szeregów czasowych, budowany jest zwykle w oparciu o pojedynczą zmienną, której typ określony jest jako „Wejściowo-Wyjściowy”. Oznacza to, że jest ona wykorzystywana zarówno jako wejście sieci neuronowej jak i jako jej wyjście.

Wstęp

Postęp w rolnictwie możliwy jest dzięki rozwojowi naukowemu m.in. takich dziedzin jak: inżynieria rolnicza, informatyka w rolnictwie, prognozowanie rolnicze itp. Sztuczne sieci neuronowe coraz częściej stanowiące obiekt zainteresowania prowadzących badania w w.w. dziedzinach, umożliwiają rozwiązywanie różnego rodzaju problemów i często z powodzeniem wykorzystywane są w nowoczesnym rolnictwie. W praktyce na ogół stanowią one jądra systemów informatycznych, efektywnie wspomagających procesy podejmowania decyzji w produkcji rolniczej. Dzieje się tak m.in. dlatego, ponieważ modele oparte na sztucznej inteligencji potrafią coraz lepiej prognozować, diagnozować, rozpoznawać oraz klasyfikować.

Celem pracy było wykorzystanie sieci neuronowych w postaci szeregów czasowych w prognozowaniu wybranych plodów rolnych. Przewidywany był plon buraków cukrowych oraz pszenicy. Prognozy zbudowano w oparciu o dane empiryczne reprezentujące wielkości plonu buraków cukrowych i pszenicy (z jednego hektara na obszarze Polski). Dane pochodzą z Roczników Statystycznych z lat 1978-2004. Celem dodatkowym było wytworzenie, przyjaznego użytkownikowi, systemu informatycznego do prognozowania ww. plonów.

Szeregi czasowe w prognozowaniu

Analiza szeregów czasowych opiera się na założeniu, że kolejne wartości w zbiorze danych reprezentują kolejne pomiary wykonane w równych odstępach czasu. Korzystanie z szeregów czasowych ma dwa główne cele: pierwszy to wykrywanie natury zjawiska reprezentowanego przez sekwencję obserwacji a drugi to przewidywanie przyszłych wartości szeregu czasowego. Oba te priorytety wymagają zidentyfikowania i opisanie, w sposób mniej lub bardziej formalny, elementów szeregu czasowego.

W szeregach czasowych wyróżnia się dwie składowe: składową systematyczną, będącą efektem oddziaływań stałego zestawu czynników na zmienną prognozowaną oraz składową przypadkową zwaną często składnikiem losowym lub wahaniami przypadkowymi. Składnik losowy jest obecny we wszystkich zjawiskach ekonomicznych, jest on reprezentowany przez proces stochastyczny. Proces ten reprezentuje ciąg zmiennych losowych o jednakowych rozkładach prawdopodobieństwa zależnych od nielosowego parametru opisującego czas. Wahania w zjawiskach gospodarczych zachodzące w rolnictwie powodowane są przez różne czynniki. Wynikają one z biologiczno-technicznego charakteru produkcji rolniczej oraz wpływu stochastycznych czynników klimatycznych i atmosferycznych. Często

powtarzają się one w określonych porach roku, powodując występowanie sezonowości w procesie produkcji. Sezonowość ta, wywołuje wymierne skutki w sferze rynku rolnego, w szczególności w obszarze dochodów oraz nakładów. Od sezonowości uzależniona jest w dużym sposób podaż produktów rolnych, a także wysokość cen na rynku oraz poziom cen w skupach. Mierzenie poziomu wahań sezonowych oraz ich przewidywanie stwarza warunki do podejmowania trafnych i skutecznych decyzji w gospodarstwie.

Punktem wyjścia do wyboru i zastosowania odpowiedniego modelu, opisującego zmiany w szeregu czasowym, jest wstępna analiza danych empirycznych. Ma to na celu stwierdzenie, które elementy występują w danym szeregu czasowym. O wahaniami sezonowych można wnioskować na podstawie:

- poza statystycznej wiedzy o danym zjawisku,
- analizy graficznej,
- analizy statystycznej.

Najczęściej do wykrycia wahań sezonowych stosuje się analizę graficzną. Polega ona na sporządzeniu dwóch wykresów. Na pierwszym przedstawiamy poziom danego zjawiska dla całego szeregu czasowego, na drugim wykresie umieszczamy poziom danego zjawiska dla jednoimiennych okresów.

Metodyka badawcza

Wiele ważnych praktycznych problemów można zaliczyć do klasy zagadnień związanych z analizą szeregów czasowych. Celem w.w. analizy jest predykcja wartości pewnej zmiennej, na podstawie jej wcześniejszych wartości lub wartości innych zmiennych (Bishop C., 1995).

Zgodnie z def. (Tadeusiewicz R., 1993) szereg czasowy to zbiór wartości danych, na ogół empirycznych, uporządkowany w czasie. W trakcie analizy szeregów czasowych najczęściej dąży się do tego, by kolejne wartości szeregu prognozowane były na podstawie wcześniejszych wartości tej samej zmiennej lub na podstawie wartości innych zmiennych. Fakt ten implikuje charakterystyczną strukturę zbioru uczącego w postaci zmiennej wejściowo/wyjściowej. Efektywnym i wygodnym narzędziem do neuronowej analizy szeregów czasowych jest moduł „*Sieci neuronowe*” zaimplementowany w komercyjnym pakiecie *Statistica v. 7.1* firmy *StatSoft*. Topologie sztucznych sieci neuronowych dostępne w tym programie doskonale nadają się do analizy i predykcji szeregów czasowych. Co więcej, nie dotyczy to jakiegoś konkretnego rodzaju sieci. Można przeprowadzić uczenie praktycznie każdej architektury sieciowej w taki sposób, aby operowała ona na szeregu czasowym. W tym celu należy w trakcie tworzenia sieci wykonać następujące kroki:

- Z natury zadania, jakie jest rozwiązywane w analizie szeregów czasowych wynika, że ta sama zmienna ma być używana jako wejściowa (przeszłe wartości) oraz wyjściowa (przyszłe wartości), należy nadać jej w „*Edytorze danych*” charakter zmiennej *Wejściowo/Wyjściowej*. Warto podkreślić, że w większości problemów związanych z analizą szeregów czasowych występuje pojedyncza zmienna. Łatwo zauważyć, że moduł „*Sieci neuronowe*” pozwala również na korzystanie z sieci neuronowej dokonującej analizy szeregów czasowych dla wielu zmiennych, zarówno wejściowych jak i wyjściowych.

Tab. 1. Zbiór danych wejściowych szeregu czasowego dla buraków cukrowych i pszenicy

Table 1. Entrance data file of time series for sugar-beets and wheat

| Rok | Buraki cukrowe Plon w dt z 1ha | Rok | Pszenica Plon w dt z 1ha |
|-------|--------------------------------|-------|--------------------------|
| 1978. | 300 | 1978. | 32,6 |
| 1979. | 290 | 1979. | 29 |
| 1980. | 221 | 1980. | 26 |
| 1981. | 338 | 1981. | 29,6 |
| 1982. | 306 | 1982. | 30,7 |
| 1983. | 336 | 1983. | 33,6 |
| 1984. | 340 | 1984. | 35,2 |
| 1985. | 336 | 1985. | 34,3 |
| 1986. | 336 | 1986. | 37 |
| 1987. | 332 | 1987. | 37,2 |
| 1988. | 341 | 1988. | 34,8 |
| 1989. | 341 | 1989. | 38,5 |
| 1990. | 380 | 1990. | 39,6 |
| 1991. | 316 | 1991. | 38 |
| 1992. | 294 | 1992. | 30,6 |
| 1993. | 392 | 1993. | 33,3 |
| 1994. | 292 | 1994. | 31,8 |
| 1995. | 346 | 1995. | 36 |
| 1996. | 394 | 1996. | 34,6 |
| 1997. | 379 | 1997. | 32,1 |
| 1998. | 379 | 1998. | 36,2 |
| 1999. | 338 | 1999. | 35 |
| 2000. | 394 | 2000. | 32,3 |
| 2001. | 358 | 2001. | 35,3 |
| 2002. | 443 | 2002. | 38,5 |
| 2003. | 410 | 2003. | 34 |
| 2004. | 428 | 2004. | 42,8 |

- Kolejnym krokiem jest określenie wartości parametru „*Rzqd*” w oknie „*Tworzenie sieci*”, aby sprecyzować liczbę okresów czasu, z których pobierane będą wartości rozważanej zmiennej, pełniące rolę wielkości wejściowych dla sieci.

- Następnie należy określić wartość parametru „*Horyzont prognozy*”, aby wskazać, o ile okresów czasu w stosunku do ostatniej wartości wejściowej przesunięta jest wartość rozważanej zmiennej traktowana jako wartość wyjściowa z sieci (wartość prognozowana). W znacznej większości przypadków, „*Horyzont prognozy*” przyjmuje wartość 1, ponieważ najczęściej próbuje się prognozować sytuację następującą natychmiast po zbiorze obserwacji stanowiących podstawę do procesu prognozowania (np. pogodę w następnym dniu). Natomiast parametr „*Rzqd*” jest zróżnicowany w zależności od wymagań związanych z badaną dziedziną problemową. Wartość tego

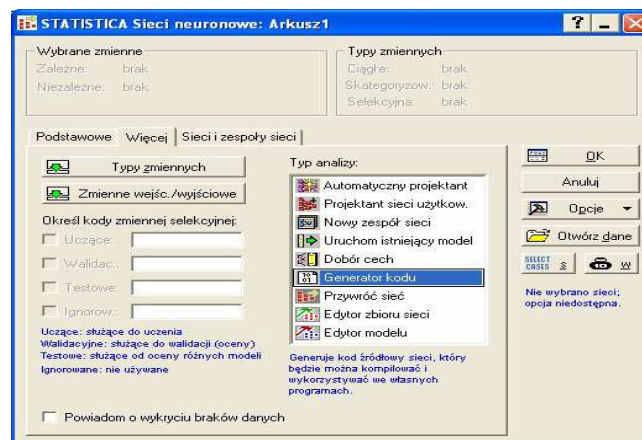
parametru wyraża pogląd na temat tego, jak wiele faktów z przeszłości trzeba znać, by na ich podstawie próbować przewidzieć przyszłość. Wynika z tego pewien praktyczny wniosek. Otóż jeśli prognozowana jest tylko pojedyncza zmienna typu „Wejściowo/Wyjściowego” przy zastosowaniu „Horyzontu prognozy” równego 1, to liczba przesłanek, na których opiera się prognoza jest wtedy dość mała. Można wtedy wprowadzić ponownie prognozowaną wartość na wejście sieci i dokonać kolejnej prognozy. Jest to możliwe właśnie ze względu na specyficzny, wejściowo/wyjściowy charakter używanych zmiennych. Z opisanego mechanizmu rekurencyjnego wykorzystania sieci można niekiedy korzystać w celu realizacji nieograniczonej prognozy szeregu czasowego. Oczywiście trzeba mieć przy tym świadomość, że im dalszy element szeregu czasowego jest przewidywany, tym bardziej niepewna i nieprecyzyjna jest uzyskiwana prognoza (Osowski S., 2000).

Omówienie wyników

W pracy wykonano neuronową predykcję plonów buraków cukrowych oraz pszenicy. Dane wykorzystane do uczenia sieci neuronowej pochodzą ze źródła GUS: banki i bazy danych, bank danych regionalnych oraz z Roczników Statystycznych z lat 1980-2004.

Wytworzono system informatyczny „Prognoza Plonów”, który pokazuje możliwości wykorzystania sieci neuronowych typu perceptron wielowarstwowy MLP (*MultiLayer Perceptron*) w postaci szeregów czasowych do prognozowania wybranych plonów na przykładzie buraków cukrowych oraz pszenicy. System wyznacza przewidywany plon

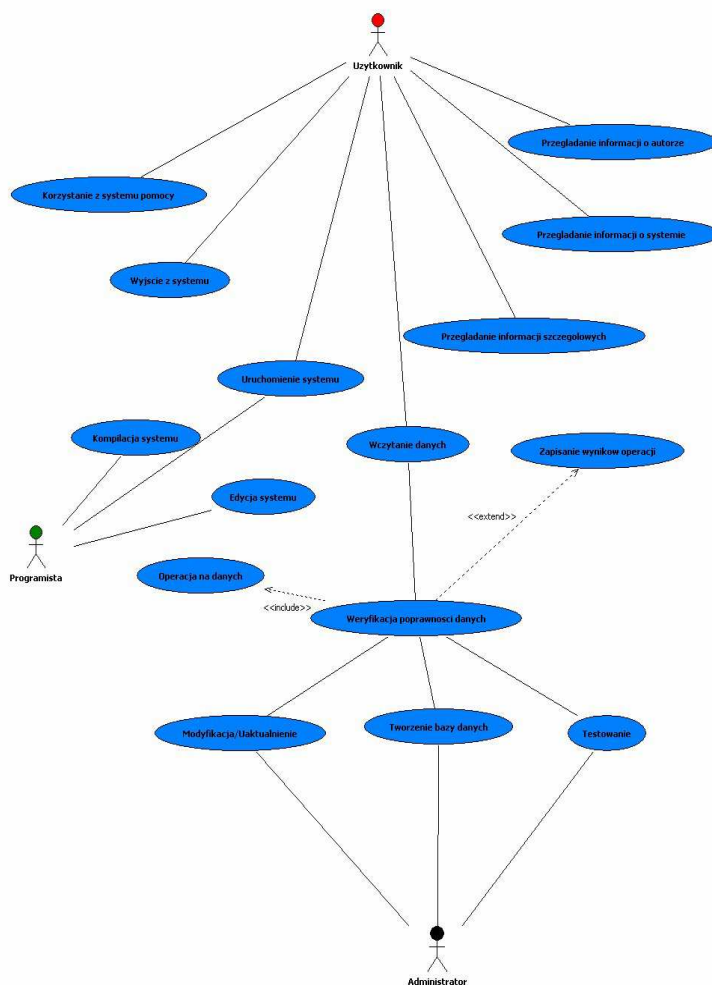
pszenicy oraz buraków cukrowych w określonej perspektywie czasu. Bazą do zbudowania tego programu był kod wygenerowany w programie *Statistica v.7.1.* (moduł „*Sieci neuronowe*”) i zaimplementowany następnie w wytworzonej aplikacji „Prognoza Plonów” z wykorzystaniem platformy *Borland Builder v.6.0.* (rys. 1).



Rys. 1. „Generator Kodu” w programie *Statistica v. 7.1.* (moduł „*Sieci neuronowe*”)

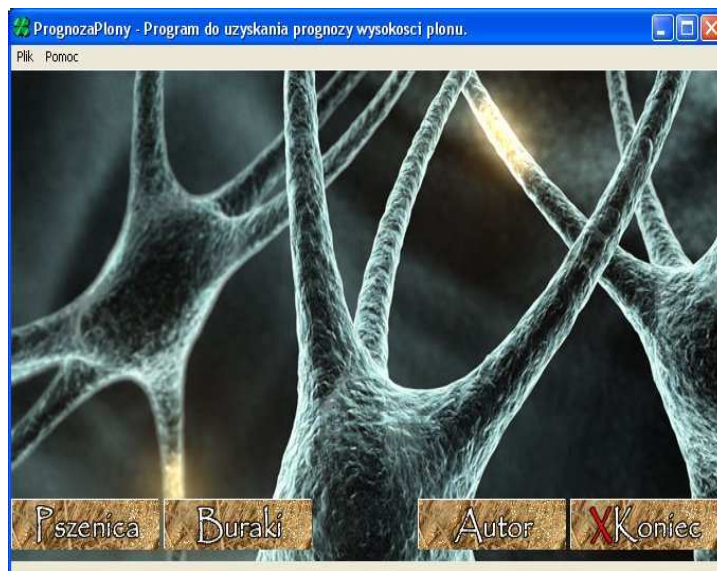
Fig. 1. "Code generator" in programme *Statistica v. 7.1.* (module "Neural networks")

Formularz główny wytworzonego programu „Prognoza Plonów” przedstawia rys. 2.



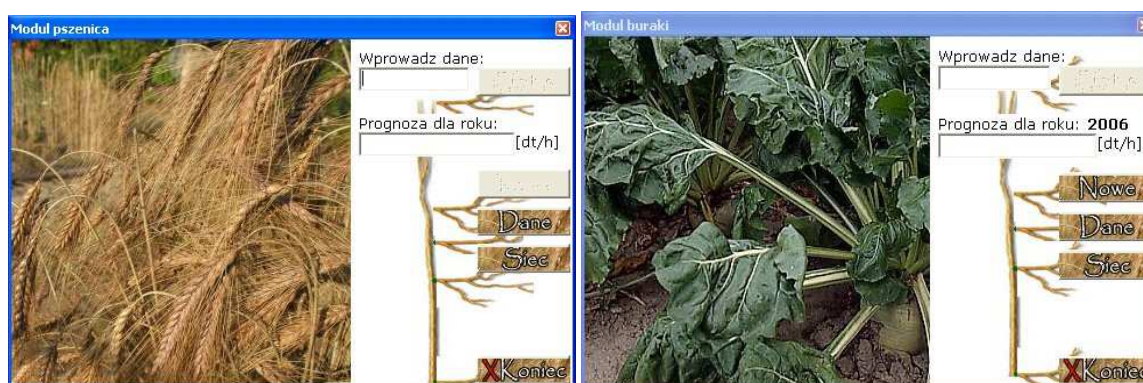
Rys. 2. Diagram przypadków użycia programu „Prognoza Plonów”

Fig. 2. Case diagram of using the programme „Prognoza Plonów”



Rys. 3. Formularz główny programu „Prognoza Plonów”
Fig. 3. Main form of programme "Prognoza Plonów"

Moduły robocze aplikacji „Prognoza Plonów” przedstawia rys. 4.



Rys. 4. Formularze robocze do prognozowania plonu pszenicy oraz buraków
Fig. 4. Working forms to prognose wheat and sugar beet crops

Aplikacja została wytworzona zgodnie z wymogami inżynierii oprogramowania. W związku z tym posłużono się m.in. nowoczesnymi technikami oferowanymi przez język **UML (Unifidet Modeling Language)**, który służy do opracowania specyfikacji, wizualizacji oraz konstruowania i dokumentowania systemów informatycznych. Pozwala on również na modelowanie szerokiej gamy dziedzin, z modelowaniem aplikacji w zakresie wspomagania zarządzania włącznie. Język UML pozwala tworzyć podstawowe w metodyce obiektowej diagramy, które umożliwiają wizualizację, takich jak np. diagram przypadków użycia. Przypadki użycia odwzorowują strukturę systemu tak, jak ją widzą jego użytkownicy. Tworzony model musi być zrozumiały dla przyszłych użytkowników. Diagram przypadków użycia programu „PrognozaPlonów” przedstawia rys. 2.

Uwagi końcowe

Sieci neuronowe typu *MLP* w postaci szeregów czasowych są skutecznym narzędziem do prognozowania wielkości plonu pszenicy i buraków. Wytworzony system informatyczny „PrognozaPlonów” do prognozowania wybranych płodów rolnych za pomocą szeregów czasowych spełnia postawione wymagania jako narzędzie prognostyczne. Aplikacja spełnia również wymagania funkcjonalne postawione w fazie określenia wymagań.

Doświadczenia z badanymi sieciami neuronowymi wykazały, że istotny wpływ na jakość prognozy ma struktura danych wejściowych, a w szczególności liczba przypadków uczących. Im znana jest większa liczba danych „historycznych” tym prognoza obciążona jest mniejszym błędem. Stwierdzono związek jakości predykcji z wielkością horyzontu prognozy. Najlepsze rezultaty uzyskano przy predykcjach krótko terminowych, co należy uznać za własność charakterystyczną dla omawianej techniki prognozowania.

Literatura

- [1] Bishop C., (1995). *Neural Networks for Pattern Recognition*: Oxford University Press.
- [2] Fausett L., (1994). *Fundamentals of Neural Networks*: New York Prentice Hall.
- [3] Kosiński R., (2002). *Sztuczne sieci neuronowe*: WNT, Warszawa.
- [4] Żurada J., Barski M. (1996). *Sztuczne sieci neuronowe*: Wydawnictwo Naukowe PWN, Warszawa.
- [5] Osowski S. (2000). *Sieci neuronowe do przetwarzania informacji*: Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa
- [6] Tadeusiewicz R. (1993). *Sieci neuronowe*: Akademicka Oficyna Wydawnicza, Warszawa

