

Wiesław MICZULSKI, Maciej CEPOWSKI

UNIWERSYTET ZIELONOGÓRSKI, INSTYTUT METROLOGII ELEKTRYCZNEJ,
ul. Podgórna 50, 65-246 Zielona Góra

Wpływ typu sieci neuronowej i sposobu przygotowania danych na wynik prognozowania poprawek UTC - UTC(PL)

Dr hab. inż. Wiesław MICZULSKI

Jest zatrudniony na stanowisku profesora nadzwyczajnego w Instytucie Metrologii Elektrycznej Uniwersytetu Zielonogórskiego. Zainteresowania naukowe są głównie związane z pomiarami wielkości elektrycznych i nieelektrycznych, inteligentnymi przyrządami pomiarowymi, analizą systemów pomiarowo-sterujących oraz zastosowaniem metod sztucznej inteligencji do diagnostyki przyrządów pomiarowych i pozyskiwania wiedzy z danych pomiarowych.



e-mail: W.Miczulski@ime.uz.zgora.pl

Mgr inż. Maciej CEPOWSKI

Jest zatrudniony w Instytucie Metrologii Elektrycznej Uniwersytetu Zielonogórskiego na stanowisku kierownika pracowni IME. Zainteresowania naukowe są związane z zastosowaniem metod sztucznej inteligencji i metod statystycznych w metrologii oraz do modelowania złożonych procesów technologicznych.



e-mail: M.Cepowski@ime.uz.zgora.pl

Streszczenie

W pracy omówiono wyniki badań wpływu typu sieci (MLP, RBF), sposobu przygotowania i rozmiaru wektora danych wejściowych na wynik prognozowania poprawek UTC-UTC(PL) dla atomowego wzorca czasu i częstotliwości. Najkorzystniejsze wyniki prognozowania poprawek osiągnięto dla sieci neuronowych typu RBF, w których dane wejściowe stanowiły wektory zawierające wartości odchyżeń od trendu. Otrzymane błędy prognoz nie przekraczają wartości ± 4 ns, co pozwoliło osiągnąć w porównaniu z dotychczasowym sposobem prognozowania opartym na metodzie regresji (prowadzonym w GUM) ponad dwukrotnie lepsze wyniki prognoz.

Słowa kluczowe: sieci neuronowe, prognozowanie, zegar atomowy.

Influence of type of neural network and selection of data preprocessing method on UTC-UTC(PL) prediction result

Abstract

In the paper there are presented the results of investigations on the influence of type of a neural network (MLP, RBF), the way of preparation and size of the input vector on prediction of the UTC-UTC(PL) corrections for the atomic clock realising the national atomic time scale UTC(PL) at the Central Office of Measures (COM). UTC(PL) is the basis for reproducing the time and frequency units and determining the official time in Poland. At the first stage of research, the time series characterising the time instability of the atomic clock in relation to UTC was a basis for proper preparation of the groups of input data for the assumed types of neural networks (Fig. 1). For the process of learning the neural networks as well as further prediction, the input data was assumed to be formed into 30 or 60 element vectors (Fig. 2). At the second stage, the input data were formed into the vectors containing 30 consecutive values of the deviation of the time series from the trend and, additionally, the linear regression coefficients (Fig. 1). The best results of predicting the corrections were achieved for the RBF neural networks in which the input data were vectors of 30 consecutive values of the deviation from the trend and the directional coefficient of linear regression. The obtained errors of the prediction did not exceed the values of ± 4 ns (Tab. 2), which enabled achieving more than two times better results of the prediction compared with the present way of prediction based on the regression method (used in COM).

Keywords: neural networks, prediction, atomic clock.

1. Wprowadzenie

Ogólnie prognozowanie jest najczęściej oparte na metodzie aproksymacji, której modele są funkcjami reprezentowanymi za pomocą trzech elementów:

- s – struktury,
- a – zbioru wag (parametrów),
- X_i – zbioru atrybutów (zmiennych), gdzie $X_i \subseteq X$.

Formalny zapis modelu dla zadania aproksymacji funkcji ma postać:

$$y^* = \{s, a, X_i\}. \quad (1)$$

Odkryta zależność funkcyjna ma jak najlepiej przybliżać funkcję docelową zadaną przez wartości rozpatrywanych atrybutów. Warunek ten wynika bezpośrednio z potrzeby, jaką jest prognozowanie wartości wybranego atrybutu zależnego. Struktura funkcji jest zwykle określona poprzez zastosowaną metodę aproksymacji funkcji i jest przyjmowana *a priori*, natomiast wyznaczane są jedynie wartości parametrów charakterystycznych dla tej struktury. W przypadku modeli regresji [1] strukturą będzie sparametryzowane wyrażenie algebraiczne, a dla metod modelowania neuronowego będzie to sieć neuronowa (SN) określonego typu [4, 6]. Wymienione metody aproksymacji są stosowane tam, gdzie duże znaczenie ma dokładność identyfikowanej zależności funkcyjnej. Przykładem zastosowania obu metod aproksymacji jest prognozowanie wartości poprawek dla atomowego wzorca czasu i częstotliwości (zegar) realizującego w Głównym Urzędzie Miar (GUM) krajową atomową skalę czasu (UTC(PL)) [3], która stanowi podstawę do odtwarzania jednostek miar czasu i częstotliwości oraz wyznaczania czasu urzędowego w Polsce. UTC(PL) jest realizowana za pomocą wybranego zegara oraz urządzenia sterującego, które umożliwia wprowadzanie poprawek w celu zapewnienia jak najlepszej zgodności UTC(PL) z UTC. Należy mieć na uwadze, że UTC jest wirtualną atomową skalą czasu, wyliczaną jako średnia ważona na podstawie ciągłych porównań obecnie ponad 300 zegarów znajdujących się w laboratoriach wielu krajów świata. Poprawki czasu fazowego dla poszczególnych lokalnych skal czasu UTC(k), w tym dla UTC(PL), względem UTC są wyznaczane przez Międzynarodowe Biuro Wag i Miar (Bureau International des Poids et Mesures – BIPM) i publikowane w biuletynie serii „CIRT” dla daty MJD (zmodyfikowana data juliańska) kończącej się na cyfry 4 i 9 w danym miesiącu. Problem utrzymania jak najlepszej zgodności UTC(PL) z UTC wynika z dużego opóźnienia wydawania przez BIPM biuletynu „CIRT”, który jest dostępny około połowy następnego miesiąca. Jedynym w tym przypadku możliwym działaniem jest prognozowanie wartości poprawek w kolejnym miesiącu na podstawie danych publikowanych przez BIPM z poprzedniego miesiąca lub kilku poprzednich miesięcy oraz bieżących wyników porównań lokalnych zegara względem UTC(PL). Narodowe laboratoria metrologiczne prowadzące prognozowanie wartości poprawek stosują zwykle metodę regresji liniowej, która jest rozszerzana o stosowanie np. stochastycznych równań różniczkowych [5].

W procesie prognozowania, realizowanym przez SN, bardzo ważnym zadaniem jest dobór struktury modelu neuronowego i właściwe przygotowanie danych wejściowych [6]. Wymaga to przeprowadzenia dla każdego zadania prognostycznego wielu badań pozwalających dokonać wyboru właściwej drogi postępowania. W artykule, na przykładzie prognozowania wartości po-

prawkę dla zegara realizującego UTC(PL), przedstawiono wyniki badań dla przyjętych struktur modelu neuronowego i wybranych sposobów przygotowania danych wejściowych. Badania te przeprowadzono w programie STATISTICA 8.0.

2. Przygotowanie danych

Prognozowanie w oparciu o SN wymaga przeprowadzenia procesu ich uczenia, którego jakość zależy od liczby danych uczących [6]. Do dyspozycji były dane historyczne zegara za okres od 1 stycznia 2006 roku (MJD 53736) do końca maja 2008 roku (MJD 54617), określające wyniki pomiarów czasu fazowego pomiędzy UTC(PL) a zegarem na każdy dzień. Dla danych tych dokonano korekty skoków fazy i przestrojenia częstotliwości. Również za ten sam okres czasu były dostępne wyniki porównań UTC i UTC(PL), wyznaczone z odstępem pięciodniowym przez BIPM i opublikowane w biuletynie CIRT. W celu zwiększenia liczby danych w procesie uczenia SN dokonano przygotowania danych, które polegało na określeniu wyników porównań UTC i UTC(PL) na każdy dzień, na podstawie danych określanych przez BIPM. Po analizie wyników interpolacji wykonanej z wykorzystaniem wybranych metod [2], do przetworzenia zbioru danych z BIPM wykorzystano metodę interpolacji wielomianowej (PCHIP) - dostępną w pakiecie Matlab 2008b.

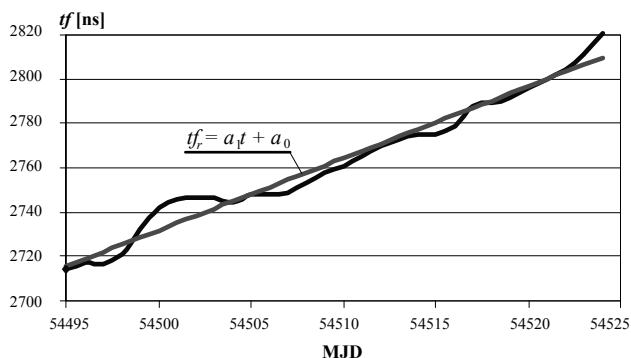
Dostępne dane pomiarowe zegara oraz wyznaczone na każdy dzień wyniki porównań UTC i UTC(PL) były podstawą przygotowania, dla zegara odtwarzającego czas UTC(PL), ostatecznego zbioru danych wejściowych dla SN. Dane te, nazywane czasem fazowym $tf(t)$ i charakteryzujące niestabilność czasową zegara na każdy dzień w odniesieniu do UTC (rys. 1), obliczono z zależności

$$tf(t) = tfa(t) + tfb(t) = \text{UTC} - \text{zegar}, \quad (2)$$

gdzie:

$tfa(t) = \text{UTC(PL)} - \text{zegar}$ – wyniki pomiaru czasu fazowego pomiędzy UTC(PL) a zegarem,

$tfb(t) = \text{UTC} - \text{UTC(PL)}$ – poprawki określone przez BIPM dla zegara realizującego czas UTC(PL).

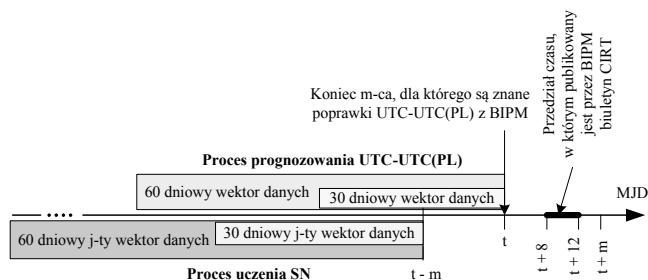


Rys. 1. Przykładowy zbiór czasu fazowego tf oraz trendu tf_r za okres jednego miesiąca
Fig. 1. Exemplary set of the phase time tf and the trend tf_r for one month period

W pierwszej fazie badań otrzymany szereg czasowy wartości $tf(t)$ stanowił podstawę odpowiedniego przygotowania grup danych wejściowych dla przyjętych modeli SN. Dla procesu uczenia SN, jak i późniejszego prognozowania przyjęto, że dane wejściowe będą formowane w wektory 30 lub 60 elementowe zgodnie z zasadą określoną na rys. 2.

W drugiej fazie badań, w zbiorze danych wejściowych dokonano eliminacji trendu długoterminowego zmian czasu fazowego $tf_r(t)$ opisanego równaniem regresji liniowej (rys. 1) [4]. Dla procesu uczenia SN, jak i późniejszego prognozowania przyjęto, że dane wejściowe będą formowane w wektory zawierające 30

kolejnych wartości odchylenia od trendu ($tf - tf_r$) oraz dodatkowo dwa współczynniki regresji a_1 i a_0 lub jeden współczynnik regresji a_1 .



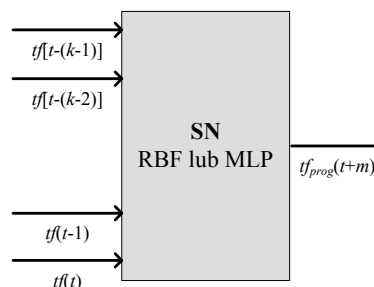
Rys. 2. Zasada przygotowania danych wejściowych dla procesu uczenia SN i procesu prognozowania UTC-UTC(PL)

Fig. 2. The rule of preparation of the input data for the process of learning NN and the process of prediction UTC-UTC(PL)

Ponadto ograniczony zakres możliwych odpowiedzi sieci wymagał zastosowania wstępnego przetwarzania danych wejściowych przed ich podaniem na wejście sieci, a także wymuszał algorytmiczne przekształcenia i odpowiednią interpretację danych wyjściowych otrzymywanych z sieci. Wszystkie wartości wykorzystane do budowy systemu prognostycznego były typu numerycznego, w związku z tym zostały przeskalowane do przedziału odpowiedniego dla danego typu sieci. W programie STATISTICA 8.0 surowe dane wejściowe oraz dane wyjściowe były automatycznie skalowane w zależności od rodzaju i architektury sieci oraz przyjętych funkcji aktywacji.

3. Wyniki badań

Wyniki wstępnych badań wykazały, że do odwzorowania zachowań szeregów czasowych można zastosować sieci radialne (RBF) oraz sieci jednokierunkowe, trójwarstwowe (MLP) (rys. 3) o sigmoidalnych funkcjach aktywacji w warstwie ukrytej, liniowej funkcji aktywacji neuronów wejściowych i neuronu wyjściowego. Przy realizacji badań założono, że prognozy UTC(PL) będą wykonywane na 15 dzień ($m = 15$) pięciu kolejnych miesięcy, tj. na dzień MJD równy 54479, 54509, 54539, 54569 i 54599. Ostatnia prognoza dotyczyła miesiąca, na którym kończył się zbiór danych pomiarowych. Wcześniejsze dane pomiarowe ($\text{MJD} < 54464$) stanowiły zbiór danych umożliwiający przygotowanie $k=30$ lub $k=60$ dniowych wektorów danych dla procesu uczenia SN. Proces uczenia SN dla zadanych liczb neuronów w warstwie ukrytej był prowadzony pięciokrotnie w celu otrzymania 5 wersji sieci dla tej samej liczby neuronów. Otrzymane w ten sposób wersje sieci były zastosowane do prognozowania poprawek.



Rys. 3. Model sieci neuronowej

Fig. 3. Neural network model

W procesie prognozowania wyznaczano wartość prognozy UTC - zegar, na podstawie której określano prognozę UTC(PL) - UTC, stanowiącą wartość poprawki umożliwiającą korygowanie nastaw częstotliwości wzorca atomowego, co teoretycznie powinno wpływać na idealną zgodność UTC(PL) z UTC.

Ogólnie wymaga się, żeby prognozowane wartości poprawek nie różniły się w stosunku do podawanych przez BIPM, na ten sam dzień MJD, więcej niż ± 10 ns. Prezentowane w artykule wyniki badań przedstawiają ich końcowy efekt - błąd prognozy (Δ_{prog}) określony zależnością (3), czyli różnicę pomiędzy prognozowaną wartością UTC(PL)-UTC a jej wartością odczytaną z biuletynu CIRT dla tego samego dnia prognozy.

$$\Delta_{prog} = (\text{UTC(PL)} - \text{UTC})_{prog} - (\text{UTC(PL)} - \text{UTC})_{\text{CIRT}} \quad (3)$$

W pierwszej fazie badań wykonano prognozowanie poprawek UTC(PL)-UTC na 15 dzień przyjętych kolejnych 5 miesięcy, dla sieci MLP i RBF z 30 i 60 dniovymi wektorami danych wejściowych. W tab. 1 przedstawiono porównanie otrzymanych wyników błędów prognozy dla sieci MLP z 30 dniovym wektorem danych wejściowych i metody regresji stosowanej w GUM. Dla sieci MLP najkorzystniejsze wyniki prognoz otrzymano głównie przy 73 neuronach w warstwie ukrytej sieci.

Tab. 1. Błędy prognozy dla sieci MLP
Tab. 1. Prediction errors for the MLP network

MJD	Δ_{prog} [ns]	
	MLP	GUM
54479	-0,4	-2,0
54509	-1,7	5,1
54539	26,3	1,2
54569	2,1	12,3
54599	-3,7	7

W przypadku sieci MLP z 60 dniovym wektorem danych wejściowych otrzymane wyniki badań wskazały, że nie wystąpiła poprawa wartości błędów prognoz w stosunku do 30 dniovego wektora danych, natomiast znacznie wydłużył się czas uczenia SN.

Dla sieci RBF wszystkie wykonane prognozy i otrzymane wyniki błędów prognoz miały znacznie większe wartości błędów prognoz dla 30 i 60 dniovych wektorów danych w stosunku do sieci MLP. Otrzymane wartości błędów prognoz były różne dla przyjętej liczby neuronów w warstwie ukrytej i zawierały się w przedziale od kilkunastu do kilkuset ns.

W drugiej fazie badań przeprowadzono prognozowanie poprawek UTC - UTC(PL) dla wektora danych wejściowych zawierającego 30 kolejnych wartości odchylenia od trendu oraz dodatkowo dwa współczynniki regresji a_1 i a_0 lub jeden współczynnik regresji a_1 . Otrzymane wyniki błędów prognozowania poprawek przedstawiono w tab. 2.

Tab. 2. Błędy prognozy dla sieci MLP i RBF przy wektorach zawierających 30-ci kolejnych wartości odchylenia od trendu
Tab. 2. Prediction errors for the MLP and RBF networks for vectors containing 30 consecutive values of deviation from the trend

MJD	Δ_{prog} [ns]				
	MLP		RBF		GUM
	a_1, a_0	a_1	a_1, a_0	a_1	
54479	0,0	0,0	-0,4	0,2	-2,0
54509	-0,2	-3,4	0,0	-0,7	5,1
54539	1,4	-1,2	4,9	3,1	1,2
54569	7,9	3,9	4,9	3,3	12,3
54599	3,9	5,4	3,9	0,8	7

Dla sieci MLP najkorzystniejsze wyniki prognoz otrzymano głównie przy 73 neuronach w warstwie ukrytej sieci, natomiast dla sieci RBF przy 150 neuronach.

Zaproponowany w drugiej fazie badań sposób przygotowania danych wejściowych dla SN, polegający na eliminacji trendu szeregu czasowego, pozwolił uzyskać lepszą jakość prognozowania poprawek UTC - UTC(PL) przy zastosowaniu sieci MLP i RBF w stosunku do metody regresji (GUM).

4. Podsumowanie

Krajowa atomowa skala czasu UTC(PL) jest realizowana w Głównym Urzędzie Miar za pomocą wybranego zegara atomowego oraz urządzenia sterującego, umożliwiającego wprowadzanie poprawek w celu zapewnienia jak najlepszej zgodności UTC(PL) z UTC. W GUM prognozowanie poprawek prowadzone jest metodą regresji. Celem prowadzonych badań było sprawdzenie możliwości zastosowania wybranych sieci neuronowych (MLP i RBF) do prognozowania poprawek UTC - UTC(PL).

W pierwszej fazie badań prognozowanie poprawek odbywało się na podstawie szeregu czasowego charakteryzującego niestabilność czasową zegara atomowego realizującego UTC(PL) w stosunku do UTC. Na pięć wyznaczonych w kolejnych miesiącach prognoz poprawek, cztery były korzystniejsze dla sieci MLP w stosunku do metody regresji stosowanej w GUM. Natomiast w jednym miesiącu błędy prognozy przekroczyły dopuszczalną wartość ± 10 ns. Otrzymane błędy prognoz (odpowiednio 26,3 ns i 12,3 ns) są spowodowane znacznie większą zmianą wartości czasu fazowego tf w stosunku do trendu, w obszarze czasowym (odpowiednio 2m i m - rys. 2), w którym sieć neuronowa i metoda regresji nie zostały objęte procesem uczenia i prognozowania.

W drugiej fazie badań, w zbiorze danych wejściowych dla sieci neuronowych, dokonano eliminacji trendu długoterminowego zmian czasu fazowego. Pozwoliło to ograniczyć rolę sieci neuronowej do przewidywania bardziej subtelnych odchylenia czasu fazowego tf w stosunku do jego trendu. We wskazanym wyżej obszarze czasowym zmiany odchylenia są dwukrotnie większe niż w pozostałym analizowanych obszarze czasowym. Z przeprowadzonych badań wynika, że dla rozważanych typów sieci neuronowych i sposobów przygotowania danych korzystniejsze w prognozowaniu poprawek są sieci RBF z wektorami danych wejściowych zawierającymi 30-ci kolejnych wartości odchylenia od trendu oraz dodatkowo współczynnik kierunkowy regresji liniowej a_1 . Za okres analizowanych pięciu miesięcy błąd prognozy w tym przypadku nie przekraczał wartości ± 4 ns w stosunku do przyjętej dopuszczalnej wartości ± 10 ns. Oznacza to, że w analizowanym okresie stosując sieci RBF można było zapewnić stabilność zegara atomowego na poziomie zmian 1 s na około 700 tys. lat. W przypadku metody regresji stabilność zegara w tym okresie wynosiła 1 s na około 270 tys. lat.

Prowadzone są dalsze badania w zakresie zastosowania hybrydowego systemu prognostycznego do prognozowania UTC(PL) [2] poprzez prognozowanie odchylenia czasu fazowego tf w stosunku do jego trendu, co powinno skutkować dalszym zmniejszeniem błędów prognozy.

Praca naukowa finansowana ze środków na naukę w latach 2007-2009, jako projekt badawczy rozwojowy.

5. Literatura

- [1] Brandt S.: Statistical and Computational Methods in Data Analysis. Springer Verlag, New York, 1997.
- [2] Cepowski M., Miczulski W.: Zastosowanie sieci neuronowych do prognozowania państwowej skali czasu. Materiały XVII Sympozjum „Modelowanie i symulacja systemów pomiarowych”. Wydawnictwo Katedry Metrologii AGH, Kraków 2009, str. 141 – 150.
- [3] Czubla A., Konopka J., Nawrocki J.A.: Realization of atomic SI second definition in context UTC(PL) and TA(PL), Metrology and Measurement Systems 2/2006, Warszawa 2006.
- [4] Masters T.: Sieci neuronowe w praktyce. WNT, Warszawa, 1996.
- [5] Panfilo G., Tavella P.: Atomic clock prediction based on stochastic differential equations, Metrologia 45 (2008).
- [6] Tadeusiewicz R., Lula P.: Neuronowe metody analizy szeregow czasowych i możliwości ich zastosowań w zagadnieniach biomedycznych. W: Biocybernetyka i inżynieria biomedyczna 2000, t.6: Sieci neuronowe. Exit, Warszawa, 2000, str. 521 – 568.