

Łukasz SOBOLEWSKI

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

Porównanie metod prognozowania zastosowanych w sieci neuronowej GMDH przeznaczonej do wyznaczania prognozy poprawek dla krajowej skali czasu UTC(PL)

Mgr inż. Łukasz SOBOLEWSKI

Jest słuchaczem studiów doktoranckich na Wydziale Elektrotechniki, Informatyki i Telekomunikacji Uniwersytetu Zielonogórskiego. Zainteresowania naukowe są głównie związane z zastosowaniem metod sztucznej inteligencji oraz metod statystycznych w metrologii.



e-mail: l.sobolewski@weit.uz.zgora.pl

Streszczenie

W pracy zostały porównane metody prognozowania poprawek dla krajowej skali czasu UTC(PL). Badania dotyczące prognozowania poprawek prowadzono w oparciu o sieci neuronowe GMDH dwoma metodami, metodą analizy szeregów czasowych i metodą regresji. Prognozowanie poprawek zostało wykonane na 15 dzień miesiąca dla 20 kolejnych miesięcy. Otrzymane wyniki prognoz przy użyciu sieci neuronowej GMDH zestawione z wynikami prognoz otrzymanymi przez Główny Urząd Miar z zastosowaniem metody regresji liniowej pokazują, że lepszą metodą prognozowania poprawek dla krajowej skali czasu okazała się metoda analizy szeregów czasowych.

Słowa kluczowe: Sieci neuronowe GMDH, Polska skala czasu UTC(PL), zegar atomowy.

Comparison of predicting methods used in GMDH neural network for determining the correction prediction for the national timescale UTC(PL)

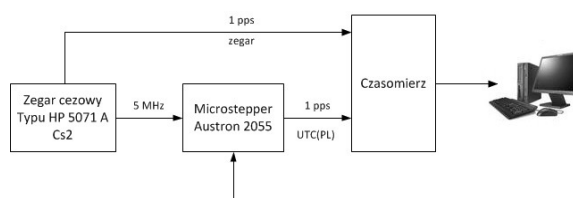
Abstract

The paper discusses the results of predicting the corrections for the national time scale UTC(PL), using GMDH neural networks. The aim of the research was to examine the influence of the GMDH neural network prediction methods on the prediction result. The first section describes the national time scale UTC(PL) and presents the problem of maintaining the best compatibility of the UTC(PL) with UTC. It also presents the method of predicting the corrections used in the Central Office of Measures (GUM), and a new method for predicting the corrections for the UTC(PL) based on GMDH neural network. The second section shows how the input data for the GMDH neural network was prepared. Based on historical measurement data from the cesium atomic clock Cs2 and corrections of the UTC(PL) relative to UTC, two time series (sc1 and sc2) which were the basis for determining the input to GMDH neural network were prepared. The third section describes the predicting methods used in the GMDH neural network and a training data for both methods. The fourth section focuses on the method of predicting the corrections using GMDH neural networks, and contains the research results. The research on predicting the corrections were carried out using two methods, the time series analysis and the regression method. Prediction of the corrections was made on the 15th day of month for 20 consecutive months. The prediction results using the GMDH neural network were compared with the results received by the GUM with use of the linear regression method. The research show that the GMDH neural networks can be used to predict the corrections for the national time scale UTC(PL). A better method of predicting the corrections for the national time scale proved to be the method of time series analysis. The results were better than the prediction results obtained in the GUM for both time series sc1 and sc2. In the case of using the regression method only for times series sc1, the obtained results were better than those obtained in the GUM.

Keywords: GMDH neural network, national timescale UTC(PL), atomic clock.

1. Wprowadzenie

Za realizację polskiej skali czasu UTC(PL) odpowiedzialny jest Główny Urząd Miar (GUM), który w tym celu stosuje komercyjny cezowy zegar atomowy typu 5071A (Cs2) oraz urządzenie (Microstepper Austron 2055), dzięki któremu możliwe jest wprowadzanie poprawek, aby zapewnić jak największą zgodność UTC(PL) z międzynarodową skalą czasu UTC (ang. *Universal Coordinated Time*) [3]. Na rys. 1 zaprezentowano fragment systemu realizujący polską skalę czasu UTC(PL).



Rys. 1. Schemat części systemu realizacji polskiej skali czasu UTC(PL)

Fig. 1. Scheme of a part of the system realizing the Polish timescale UTC(PL)

Poprawki określające rozbieżność poszczególnych krajowych skal czasu UTC(k) względem UTC wyznaczone są w każdym miesiącu z odstępem pięciodniowym przez Międzynarodowe Biuro Wag i Miar (BIPM) i publikowane w biuletynie „*Circular T*” około 10 dnia następnego miesiąca [2]. Zapewnienie utrzymania UTC(PL) w stosunku do UTC na poziomie rozbieżności nie większych niż ± 10 ns jest jednym z głównych zadań GUM. Wobec opóźnienia w publikowaniu biuletynu „*Circular T*” utrzymanie jak najlepszej zgodności UTC(PL) z UTC może być rozwiązane tylko na drodze prognozowania poprawek.

W GUM stosowana jest procedura prognozowania poprawek w oparciu o metodę regresji liniowej. Metoda ta jest dosyć pracochłonna i wymaga dużego doświadczenia metrologicznego od wykonującego analizę [3]. Z literatury znane są również metody prognozowania poprawek dla UTC(k) w oparciu o metodę regresji liniowej ze stochastycznymi równaniami różniczkowymi [9], metodę analityczną wykorzystującą odchylenia Allana [1], czy też metodę opartą na filtrze Kalmana [4].

W procesie prognozowania oprócz metod analitycznych mogą być zastosowane metody wykorzystujące sieci neuronowe. Instytut Metrologii Elektrycznej Uniwersytetu Zielonogórskiego, współpracując z GUM, prowadzi prace nad prognozowaniem poprawek dla UTC(PL) z zastosowaniem sieci neuronowych. Otrzymane dotychczas wyniki badań wskazują na możliwość zastosowania sieci neuronowych MLP i RBF [7] do prognozowania poprawek dla UTC(PL). Dużym utrudnieniem w stosowaniu tego typu sieci neuronowych jest zbyt długi czas otrzymania wyniku prognozy. Wynika on z potrzeby dopasowania właściwej struktury sieci i liczby neuronów do charakteru danych podawanych na jej wejście w procesie uczenia.

Do badań zaproponowano zastosowanie sieci neuronowych GMDH (ang. *Group Method of Data Handling*), które należą do grupy sieci samoorganizujących się. W przypadku tego typu sieci struktura sieci tworzona jest samoczynnie na podstawie przygotowanych zbiorów danych uczących i testujących [5], [8]. Ta zaleta zdecydowała o wyborze sieci neuronowej GMDH do badań pod kątem prognozowania wartości poprawek dla UTC(PL).

2. Przygotowanie danych dla sieci neuronowej GMDH

Podstawą przygotowania zbioru danych wejściowych dla sieci neuronowej GMDH były historyczne wyniki pomiarów czasu fazowego pomiędzy sygnałami 1pps (ang. *pulse per second*) z UTC(PL) i zegara Cs2 wyznaczone na każdy dzień, oraz poprawki UTC(PL) względem UTC wyznaczone przez BIPM. Dane te były dostępne za okres od 1 stycznia 2006 roku (MJD 53736) do 16 sierpnia 2009 roku (MJD 55059) (ang. *Modified Julian Date*).

Stosując dla zbioru danych z BIPM funkcję interpolującą PCHIP (ang. *Piecewise Cubic Hermite Interpolating Polynomial*), dostępną w języku MATLAB, wyznaczono model matematyczny umożliwiający rozszerzenie zbioru danych uczących przez obliczenie wartości poprawek UTC(PL) względem UTC na każdy dzień analizowanego okresu czasu.

Końcowy zbiór danych, będący podstawą określenia na każdy dzień danych wejściowych dla sieci neuronowej GMDH obliczono na podstawie zależności:

$$x(t) = x_a(t) + x_b(t), \quad (1)$$

gdzie: $x_a(t)$ – wyniki pomiaru czasu fazowego pomiędzy UTC(PL) a zegarem atomowym realizującym czas UTC(PL), $x_b(t)$ – poprawki określone przez BIPM dla UTC(PL).

Zbiór danych x_i (dla $i = 1 \dots k$), obliczony wg zależności (1) nazywany szeregiem czasowym (sc1), stanowił pierwszą grupę danych, na podstawie których prowadzono proces uczenia sieci neuronowej GMDH i prognozowanie wartości poprawek dla UTC(PL). W szeregu sc1 zawarta jest liniowa składowa trendu $x_r(t)$ i składowa zmienna. Dlatego dla procesu uczenia sieci neuronowej GMDH oraz prognozowania poprawek dla UTC(PL) został przygotowany również drugi szereg czasowy (sc2). Kończąc postać szeregu sc2 otrzymano eliminując z szeregu sc1 trend długoterminowy zmian czasu fazowego $x_r(t)$, który opisany był równaniem regresji liniowej. Poszczególne elementy szeregu czasowego sc2 stanowiące wartości odchylenia od trendu obliczono na każdy dzień wg zależności (4). Szczegółowy opis przygotowania szeregów czasowych sc1 i sc2 jest przedstawiony w: [7] i [8].

$$xd(t) = x(t) - x_r(t) \quad (4)$$

3. Metody prognozowania z zastosowaniem sieci neuronowej GMDH

Prognozowanie z zastosowaniem sieci neuronowych GMDH prowadzono na podstawie analizy szeregu czasowego lub metody regresji. W pierwszym przypadku na wejście sieci był podawany wektor danych stanowiący szereg czasowy sc1 lub sc2 zapisany w postaci wektora o k elementach. Przykładowy zbiór danych wejściowych do prognozowania poprawki na miesiąc styczeń 2008 r. (tab. 1) zawierał w jednym wektorze 730 elementów.

Tab. 1. Przykładowe dane wejściowe dla sieci neuronowej GMDH do prognozowania poprawek na podstawie analizy szeregu czasowego sc1 i sc2, na 54479 MJD
Tab. 1. Sample input data for the GMDH neural network for predicting the corrections based on the time series analysis for time series ts1 and ts2, for 54479 MJD

MJD	x(t) dla sc1 [ns]
53736	663,021
53737	664,269
53738	664,992
53739	665,726
53740	667,304
53741	669,971
53742	673,638
.	.
.	.
.	.
54464	2632,183

MJD	xd(t) dla sc2 [ns]
53736	57,267
53737	55,958
53738	54,124
53739	52,301
53740	51,322
53741	51,432
53742	52,542
.	.
.	.
.	.
54464	164,837

W drugim przypadku na wejście sieci neuronowej GMDH były podawane wektory zawierające kolejne dane szeregu czasowego sc1 lub sc2 za okres 30 dni ($t_0 - 29 \div t_0$) i wartość prognozy na 15-ty dzień ($t_0 + 15$). W przypadku szeregu czasowego sc2 dodatkowo były podawane współczynniki regresji liniowej (a_1 i a_0) składowej trendu $x_r(t)$, obliczane za okres 30 dni (tab. 2).

Przykładowy zbiór danych wejściowych (tab. 2), otrzymany na bazie szeregu czasowego sc2, składał się z 685 wektorów zawierających po 33 elementy każdy.

Tab. 2. Przykładowe dane wejściowe dla sieci neuronowej GMDH do prognozowania poprawek metodą regresji na 54479 MJD, szereg czasowy sc2

Tab. 2. Sample set of input data for the GMDH neural network for predicting the corrections using the regression method for 54479 MJD, time series ts2

MJD dla $t_0 + 15$	$xd(t_0 - 29)$ [ns]	$xd(t_0 - 28)$ [ns]	$xd(t_0 - 1)$ [ns]	$xd(t_0)$ [ns]	a_1	a_0	$xd(t_0 + 15)$ [ns]
53780	3,259	1,997	-4,603	-4,123	2,51127	-134286	-1,511
53781	2,263	0,470	-3,981	-2,658	2,51572	-134525	-2,528
53782	0,698	-1,092	-2,646	-1,219	2,52343	-134940	-3,773
53783	-0,970	-1,923	-1,318	0,371	2,53133	-135365	-5,083
53784	-1,980	-1,8456	0,285	1,113	2,53239	-135422	-6,166
53785	-2,018	-0,880	1,049	1,451	2,52849	-135212	-7,182
53786	-1,041	-0,112	1,358	1,688	2,52604	-135080	-8,192
.
54464	-1,223	-2,477	-1,454	-0,045	3,38277	-181614	7,139

W sposobie przygotowania danych dla obu metod prognozowania występuje różnica w liczbie danych dostępnych dla procesu uczenia sieci neuronowej GMDH. W prognozowaniu metodą regresji zbiór danych wejściowych kończy się na 15 dni przed końcem miesiąca poprzedzającego wyznaczenie prognozy. Natomiast w przypadku analizy szeregów czasowych zbiór danych wejściowych kończy się w ostatnim dniu miesiąca poprzedzającego prognozę, czyli jest dłuższy o 15 dni w stosunku do zbioru przygotowanego dla metody regresji.

4. Wyniki badań

4.1. Sposób prowadzenia badań

Celem badań była ocena wpływu doboru metody prognozowania sieci neuronowej GMDH na wynik prognozowania poprawki dla krajowej skali czasu UTC(PL). Prognozowanie poprawek dla UTC(PL) wykonywano na 15-ty dzień miesiąca dla 20 kolejnych miesięcy od stycznia 2008 (MJD 54479) do sierpnia 2009 (MJD 55059), w oparciu o dwie metody: metodę analizy szeregów czasowych oraz metodę regresji.

W trakcie badań na wejście sieci neuronowej GMDH były podawane wektory danych przygotowane odpowiednio dla każdej z metod na podstawie szeregów czasowych sc1 i sc2. Na wyjściu sieci otrzymywano prognozę wyznaczaną na ten sam dzień, jak w GUM. W procesie prognozowania wyznaczano dla szeregu czasowego sc1 wartość prognozy x_p , a na jej podstawie w kolejnym kroku obliczono prognozę x_{bp} . Prognoza ta określała wartość poprawki, która umożliwiała korygowanie UTC(PL) tak, aby zapewnić jak najlepszą zgodność UTC(PL) z UTC. W przypadku szeregu czasowego sc2 w pierwszym kroku wyznaczano wartość prognozy odchylenia od trendu xd_p , którą dodawano do prognozy trendu wyliczonej na podstawie równania regresji. Otrzymany wynik stanowił poszukiwaną wartość prognozy x_p . Dalsze obliczenia zmierzające do wyznaczenia prognozy wartości poprawki były prowadzone analogicznie jak dla szeregu czasowego sc1 [8].

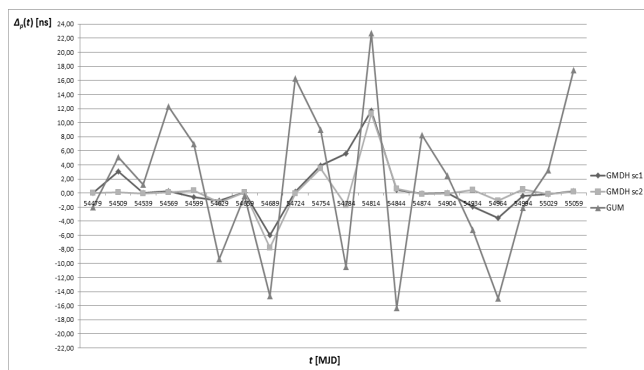
Prezentowane w artykule wyniki badań przedstawiają błąd prognozy (Δ_p) wyznaczony dla sieci GMDH według zależności (3). Jest to różnica pomiędzy prognozowaną wartością poprawki UTC(PL) względem UTC (x_{bp}), a jej wartością odczytaną z biuletynu „Circular T” (x_b) dla dnia wyznaczania prognozy.

$$\Delta_p = x_{b_p} - x_b \quad (5)$$

Dla każdej z metod prognozowania parametry sieci neuronowej GMDH dobierano indywidualnie, dla każdego dnia prognozy tak, aby otrzymany błąd prognozy Δ_p był jak najmniejszy.

4.2. Porównanie metod prognozowania poprawek dla UTC(PL)

Na rys. 2 zaprezentowano końcowe wyniki badań, przedstawiające wartości błędów prognozy otrzymane metodą analizy szeregów czasowych dla szeregów czasowych sc1 i sc2, które zestawiono z wartościami błędów prognozy otrzymanymi w GUM metodą regresji liniowej.



Rys. 2. Błędy prognozy przy prognozowaniu poprawek dla UTC(PL) z zastosowaniem metody analizy szeregów czasowych

Fig. 2. Prediction errors in predicting the corrections for the UTC(PL) using the time series analysis method

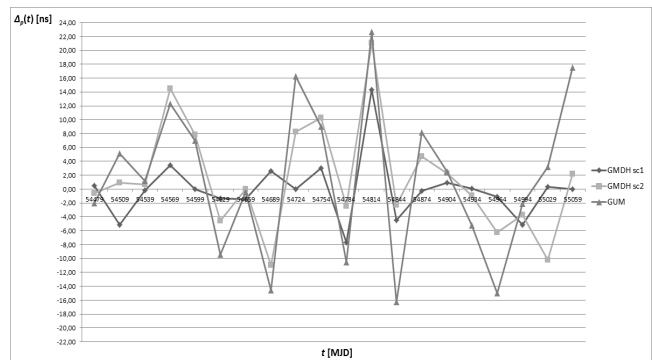
W przypadku wyznaczania poprawek dla UTC(PL) w oparciu o metodę analizy szeregów czasowych można zauważyć, że sposób przygotowania szeregów czasowych (sc1 i sc2) w niewielkim stopniu wpływa na jakość prognozowania. Otrzymane błędy prognozy w analizowanym okresie dwudziestu miesięcy są znacznie mniejsze od błędów otrzymanych w GUM. W najgorszym przypadku, tj. dla dnia prognozy MJD 54814 błąd prognozy przy zastosowaniu metody analizy szeregów czasowych, zarówno dla szeregu czasowego sc1 jak i szeregu czasowego sc2, nie przekracza wartości 12 ns. Dla tego samego dnia prognozy błąd otrzymany w Głównym Urzędzie Miar przy zastosowaniu metody regresji liniowej wynosił 22,7 ns, czyli był praktycznie dwukrotnie większy.

Z rezultatów przedstawionych na rys. 3 wynika, że w przypadku zastosowania metody regresji w sieci neuronowej GMDH, tylko dla szeregu czasowego sc1, otrzymano błędy prognozy korzystniejsze od wartości błędów wyznaczonych w GUM metodą regresji liniowej. W przypadku szeregu czasowego sc2 otrzymane wartości błędów prognozy są zbliżone do wartości błędów otrzymanych w GUM. Dla dnia prognozy MJD 54814, w którym wartości błędów prognozy były największe, błąd prognozy przy zastosowaniu metody regresji dla szeregu czasowego sc1 był równy 14,3 ns, a dla szeregu czasowego sc2 wynosił 21,1 ns. Dla tego samego dnia prognozy błąd otrzymany w Głównym Urzędzie Miar przy zastosowaniu metody regresji liniowej był na poziomie 22,7 ns.

5. Podsumowanie

Z przeprowadzonych badań wynika, że sieci neuronowe GMDH mogą być zastosowane do prognozowania poprawek dla krajowej skali czasu UTC(PL). Zastosowanie komercyjnego narzędzia GMDH Shell w wersji 2.2 pozwala na otrzymanie mniejszych wartości błędów prognozy, niż w przypadku użycia metody regresji liniowej, stosowanej w GUM. Lepszą metodą prognozowania poprawek dla krajowej skali czasu okazała się metoda analizy

szeregów czasowych. Otrzymane wyniki błędów prognozy dla szeregu czasowego sc1 i szeregu sc2, za okres 20 miesięcy, nie przekraczały wartości ± 12 ns i były prawie dwukrotnie mniejsze od rezultatów otrzymanych w GUM.



Rys. 3. Błędy prognozy przy prognozowaniu poprawek dla UTC(PL) z zastosowaniem metody regresji

Fig. 3. Prediction errors in predicting the corrections for the UTC(PL) using the regression method

W przypadku metody regresji tylko dla szeregu sc1 otrzymano rezultaty lepsze od wyników otrzymanych w GUM.

Istotną przewagą metody regresji nad metodą analizy szeregów czasowych jest możliwość wprowadzenia dodatkowych wektorów do zbioru danych uczących, charakteryzujących właściwości zegara Cs2, co może skutkować lepszymi wynikami prognoz poprawek dla UTC(PL). Obecnie trwają badania nad sprawdzeniem wpływu dodatkowych wektorów wejściowych sieci neuronowej GMDH na wynik prognozowania poprawek dla UTC(PL).

Ponadto ważną zaletą sieci GMDH, w stosunku do sieci MLP i RBF, jest automatyczne dopasowanie struktury i liczby neuronów do charakteru danych podawanych na wejście tej sieci w procesie uczenia, co znacznie skraca czas otrzymania wartości poprawki dla UTC(PL).

6. Literatura

- [1] Bernier L. G.: Use of the Allan Deviation and Linear Prediction for the Determination of the Uncertainty on Time Calibrations Against Predicted Timescales, IEEE Transactions on Instrumentation and Measurement, Vol.52, No.2, 2003, pp. 483-486.
- [2] BIPM Annual Report on Time Activities, Vol. 6, Sevres, BIPM 2011.
- [3] Czubla A., Konopka J., Nawrocki J.: Realization of atomic SI second definition in context UTC(PL) and TA(PL); Metrology and Measurement Systems, No. 2, 2006, pp. 149-159.
- [4] Davis J. A., Shemar S. L., Whibberley P. B., A Kalman filter UTC(k) prediction and steering algorithm, NMS Physical Metrology Programme, United Kingdom.
- [5] Duch W., Korbicz J., Rutkowski L., Tadeusiewicz R.: Biocybernetyka i inżynieria biomedyczna - Sieci neuronowe, Akademicka oficyna Wydawnicza EXIT, Warszawa 2000.
- [6] Masters T.: Practical neural networks recipes in C++, Academic Press, Inc., 1993.
- [7] Miczulski W., Cepowski M.: Influence of type of neural network and selection of data pre-processing method on UTC-UTC(PL) prediction result; The Measurements, Automation and Monitoring, No. 11, 2010, pp. 1330-1332.
- [8] Miczulski W., Sobolewski Ł.: Influence of the GMDH neural network data preparation method on UTC(PL) correction prediction results, Metrology and Measurement Systems, Vol. XIX, No. 1, 2012, pp. 123-132.
- [9] Panfilo G. and Tavella P.: Atomic clock prediction based on stochastic differential equations, Metrologia, No 45, 2008, pp. 108-116.