УДК 621.382.2

А.С. КАШТАЛЬЯН
Хмельницький національний університет

ОСОБЛИВОСТІ ЗАСТОВУВАННЯ НЕЙРОННИХ МЕРЕЖ ПРЯМОГО РОЗПОВСЮДЖЕННЯ ДЛЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

Стаття присвячена нейромережевому підходу прогнозування часових рядів. Розглянуто структуру нейронної мережі прямого розповсюдження, призначеної для прогнозування часових рядів, що являє собою нелінійну авторегресійну мережу з зовнішніми входами. Розглянуто вплив параметрів даної нейронної мережі та розмірів навчальної вибірки на здатність мережі до навчання. Основна увага приділена глибині занурення в ряд. Показано, що при недостатній глибині занурення нейронна мережа не здатна навчатися, в прогнозованих значеннях спостерігається відставання на горизонт прогнозування; натомість велика глибина занурення призводить до перенавчання мережі. В статті запропоновано підхід до побудови навчальної вибірки з оптимальними значеннями глибини занурення, достатньої для того, щоб навчити нейронну мережу прогнозувати часовий ряд і одночасно уникнути перенавчання.

Ключові слова: часовий ряд, прогнозування, нейронні мережі прямого розповсюдження, глибина занурення.

A.S. KASHTALIAN
Khmelnitsky National University, Khmelnitsky, Ukraine

THE FEATURES OF FEEDFORWARD NEURAL NETWORK USE FOR TIME SERIES FORECASTING

Abstract. The article is devoted to neural network technique for time series forecasting. Time series forecasting is an important task in physics, biology, economics, etc. A structure of feedforward neural network for time series prediction is considered, it represents nonlinear autoregressive network with external inputs. An influence of given neural network parameters and training set size on network ability to training is regarded. The main attention is paid to an immersion depth into a series. It is pointed that a neural network cannot be trained with an insufficient immersion depth, it observes predicted values lag behind real values on the prediction horizon. Instead a large immersion depth leads to network overfitting. The approach to creating of the training set with optimal values of immersion depth which is sufficient to train neural network to predict a time series and at the same time to avoid overfitting is suggested.

Key words: time series, forecasting, feedforward neural networks, immersion depth.

Постановка задачі. Багато процесів та явищ, що відбуваються в області фізики, техніки, біології, економіки можна описати часовими рядами. До основних задач аналізу часових рядів відносяться визначення природи ряду даних та прогнозування. Прогнозування майбутніх значень часових рядів необхідно для вибору стратегії контролю системи або оптимізації діяльності, виробництва тощо.

Формально часовий ряд є послідовністю векторів, що залежать від часу: $X(t), t = 0,1,2,\ldots,T$ Позначимо часовий ряд як $X(t) = X(1), X(2), \cdots, X(T)$. В момент часу T необхідно визначити значення процесу $X(T)$ в моменти часу $T+1,\ldots,T+P$. Момент часу T називається моментом прогнозу, а величина P - горизонтом прогнозування [1].

Для обчислення значень часового ряду в майбутні моменти часу необхідно визначити функціональну залежність, відображену між минулими та майбутніми значеннями цього ряду

$$X(\ ) = F(X(t-1), X(t-2), X(t-3), \ldots) + \varepsilon_t . \tag{1}$$

Залежність (1) є моделлю прогнозування. Необхідно розробити таку модель прогнозування, для якої середнє абсолютне відхилення істинного значення від прогнозованого прямуватиме до мінімального для заданого горизонту $P : \overline{E} = \frac{1}{P} \sum_{t=T+1}^{T+P} |\varepsilon_t| \to \min$.

Вираз (1) можна переписати у вигляді

$$\widehat{X}(t) = F(X(t-1), X(t-2), X(t-3), \ldots),$$

де $\widehat{X}(t)$ - прогнозовані (розрахункові) значення часового ряду $X(t)$. Крім отримання майбутніх значень $\widehat{X}(T+1), \ldots, \widehat{X}(T+P)$ необхідно визначити довірчий інтервал можливих відхилень цих значень.

За наявності зовнішніх факторів впливу модель прогнозування ускладняється. Нехай перший зовнішній фактор $Y_1(T_1)$ доступний в дискретні моменти часу $t_1 = 0,1,2,\ldots,T_1$, другий зовнішній фактор $Y_2(T_2)$ доступний в моменти часу $t_2 = 0,1,2,\ldots,T_2$ і так далі. У випадку, якщо дискретність вихідного часового ряду та зовнішніх факторів, а також значення $T, T_1, \ldots, T_S$ різні, то часові ряди зовнішніх факторів $Y_1(t_1), \ldots Y_S(t_S)$ необхідно привести до єдиної шкали часу t. В момент прогнозу T необхідно визначити майбутні значення вихідного процесу $X(t)$ в моменти часу $T+1,\ldots,T+P$, враховуючи вплив зовнішніх факторів $Y_1(t), \ldots Y_S(t)$.

Для визначення майбутніх значень процесу $X(t)$ у вказані моменти часу необхідно визначити функціональну залежність між минулими значеннями $X(t)$ та майбутніми, а також прийняти до уваги вплив зовнішніх факторів на вихідний часовий ряд, що визначає модель прогнозування з врахуванням зовнішніх факторів:

$$X(t) = F\big(X(t-1), X(t-2), \cdots, Y_1(t-1), Y_1(t-2), \ldots, Y_S(t-1), Y_S(t-2), \ldots\big) + \varepsilon_t . \quad (2)$$

Прийнято вважати, що задачі визначення природи ряду та його прогнозування вимагають, щоб модель ряду була ідентифікована та формально описана. Як тільки модель визначена за її допомогою можна інтерпретувати дані та екстраполювати ряд на основі знайденої моделі, тобто прогнозувати його майбутні значення.

Аналіз досліджень та публікацій. Традиційні методи прогнозування часових рядів в більшості випадків в якості вхідних даних використовують тільки інформацію, яка закладена в попередніх значеннях часового ряду. Зовнішні фактори впливу при цьому враховуються неявно, вважається, що попередні значення часового ряду несуть всю необхідну для прогнозування інформацію. На противагу нейронні мережі не мають обмежень на характер вхідної інформації. Поряд з попередніми даними часового ряду, можливо подавати на вхід нейронної мережі безпосередньо інші вхідні величини, які також впливають на майбутні прогнозовані значення ряду. Нейронні мережі в результаті навчання здатні будувати оптимальну модель часового ряду.

Нейромережевий підхід до прогнозування часових рядів є непараметричним в тому сенсі, що не вимагає інформації щодо процесу, який генерує сигнал. Відомо, що рекурентні нейронні мережі є реалізацією нелінійного ARMA (NARMA) процесу [2][3]. Здатність нейронних мереж прогнозувати ґрунтується на властивостях апроксимації та узагальнення. Однак певні обмеження та ускладнення при використанні нейронних мереж залишаються недостатньо дослідженими [4][5][6], в особливості для нових структур нейронних мереж. До таких особливостей відносяться: викиди, що ускладнюють моделювання істинної функціональної залежності; періодичність, наявна в часовому ряді, яку необхідно усувати перед моделюванням; стаціонарність є класичною ознакою для прогнозування стаціонарних часових рядів, тоді як більшість реальних часових рядів є нестаціонарними; число зразків в часовому ряді, дослідниками визначено, що збільшення числа точок не завжди призводить до підвищення точності прогнозування; проблема довгих часових залежностей, що призводить до зникаючого градієнту та «забування» поведінки ряду. Ці проблеми безпосередньо пов’язані з передобробкою даних та формуванням навчальної вибірки і потребують дослідження.

Матеріал та результати досліджень. Для динамічних систем доведена теорема Такенса. Якщо часовий ряд породжується динамічною системою, тобто значення X_t є довільні функції стану системи, то існує така глибина занурення d (приблизно рівна ефективному числу степенів свободи даної динамічної системи), яка забезпечує однозначне передбачення наступного значення часового ряду (рис. 1).

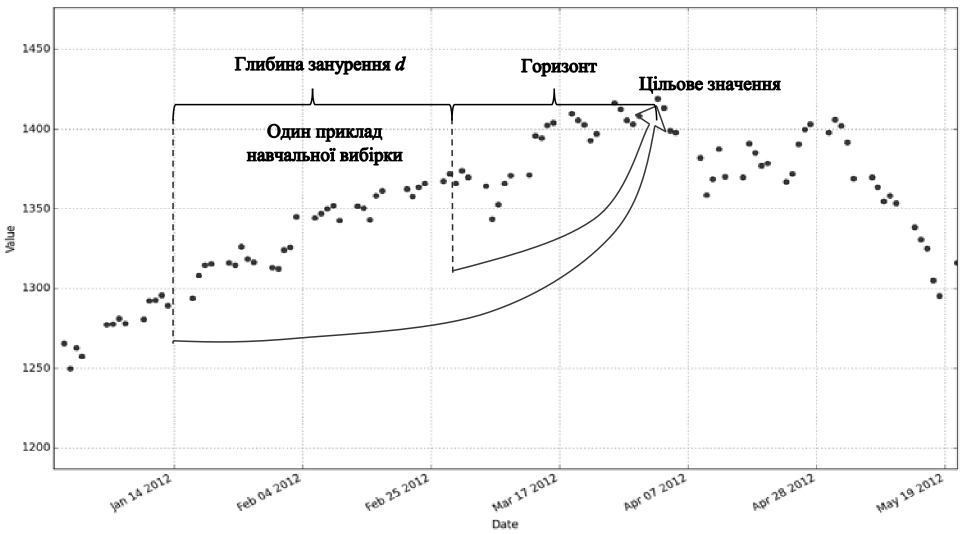


Рис. 1. Формування одного прикладу навчальної вибірки

Таким чином, обравши достатньо велике d, можна гарантувати однозначну залежність майбутнього значення ряду від його d попередніх значень:

$$X_t = f\left(X_{t-d}\right),$$

тобто прогнозування часового ряду зводиться до задачі інтерполяції функції багатьох змінних.

Нейронну мережу далі можна використовувати для відновлення цієї невідомої функції за набором прикладів, заданих історією часового ряду [7].

Для оцінки здатності прогнозування нейронною мережею розглянемо нестаціонарні часові ряди, що генеруються динамічними системами. До таких рядів безумовно можна віднести часові ряди цін на акції, курсів валют, індексів тощо. Для усунення таких явищ як викиди, нестаціонарність, кореляція і т.п. дані часового ряду потребують передобробки, в тому числі попереднього відбору інформативних ознак ряду [8]. В статті в якості нестаціонарного часового ряду використано історичні дані цін закриття на акції компанії Apple Inc.

В статті розглядається нелінійна авторегресійна мережа з зовнішніми входами, необхідними для врахування зовнішніх факторів впливу на значення часового ряду. Нелінійну авторегресійну мережу можна реалізувати в різний спосіб, але найпростішим є використання нейронної мережі прямого розповсюдження з вбудованою пам'яттю, як показано на рис. 2. Таким, чином мережа залежить від d попередніх значень ряду, тобто від глибини занурення. Цей параметр здебільшого асоціюють з часовим вікном, оскільки воно забезпечує обмежений доступ до частини всього ряду.

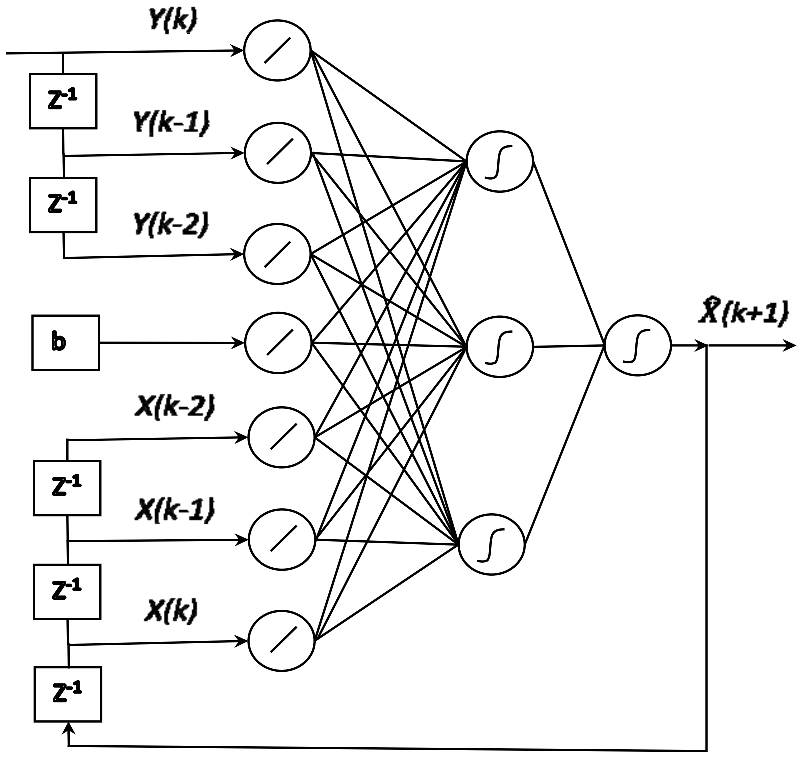


Рис. 2. Багатошаровий персептрон з лінією затримок

Стани даної нелінійної авторегресійної мережі можна представити у вигляді

$$z(k+1)=\begin{cases}\Phi\big(y(k), z_i(k)\big), & i=1\\ z_i(k) & i=2,3,\dots N\end{cases}$$

де вихід $y(k)=z_i(k)$ та $z_i, i=1,2,\dots N$ є змінними стану нейронної мережі.

Нейрони першого шару даної мережі мають лінійну функцію активації. Вихідний шар містить один нейрон з сигмоїдальною функцією активацію. На рис. 2 схематично зображено один прихований шар нейронів з сигмоїдальною функцією активації. В загальному випадку прихованих шарів може бути довільна кількість, однак дослідження показали, що оптимальною кількість є від 1 до 3. Для нейронної мережі, зображеної на рис. 2 загальне рівняння прогнозування наступного значення часового ряду може бути записане як

$$x(k+1)=\Phi_o\left\{w_{b0}+\sum_{h=1}^{n} w_{ho}\cdot\Phi_h\left(w_{ho}+\sum_{i=0}^{d} w_{ih}\cdot y(k-i)+\sum_{j=0}^{d} w_{jh}\cdot x(k-j)\right)\right\}$$

Глибина занурення визначає кількість нейронів вхідного шару перспептрона, якщо у якості вхідних даних використовуються тільки значення часового ряду. Кількість нейронів вхідного шару буде збільшуватись, якщо крім того використовуються додаткові вхідні величини.

На кількість нейронів в прихованому шарі мережі накладаються певні обмеження. Зокрема використовується наступне співвідношення, кількість нейронів в прихованому шарі $N_h=\dfrac{N_S}{\alpha\cdot(N_i\cdot N_o)}$, де N_i - кількість нейронів у вхідному шарі, N_o - кількість нейронів у вихідному шарі, N_S - кількість прикладів навчальної вибірки, тобто її розмір, α - конструктивний коефіцієнт. З цього співвідношення можна зробити висновок, що розмір навчальної вибірки N_S повинен значно перевищувати кількість вхідних нейронів мережі, а отже і глибину занурення d .

Вибір глибини занурення має надзвичайний вплив на результат навчання нейронної мережі, і як наслідок, на точність прогнозування. Для аналізу навчання нейронної мережі необхідно насамперед протестувати нейронну мережу in-sample, тобто на навчальній вибірці, і потім out-of-sample на тестовій вибірці. Якщо глибина занурення буде недостатня, то мережа не навчиться прогнозувати майбутні значення, а натомість в якості прогнозованого даватиме останнє значення, тобто ми отримаємо ефект затримки прогнозованих значень (рис. 3).

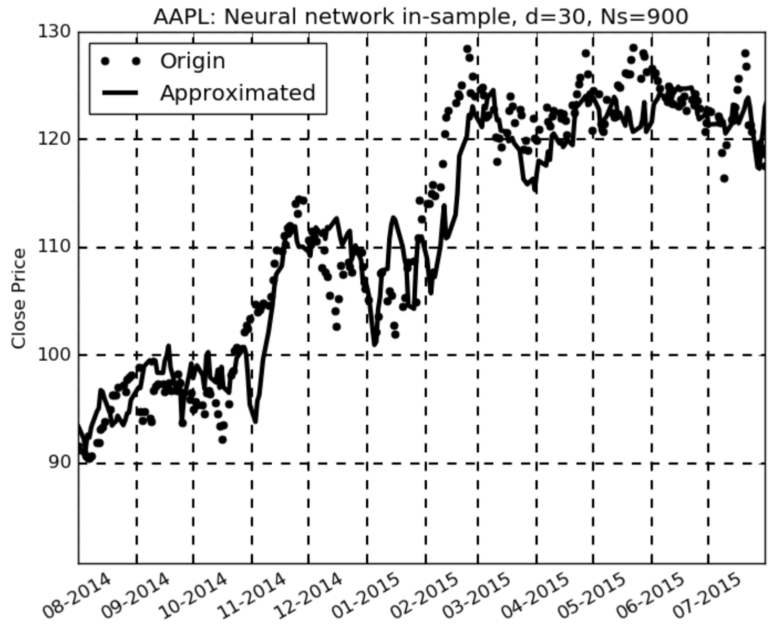
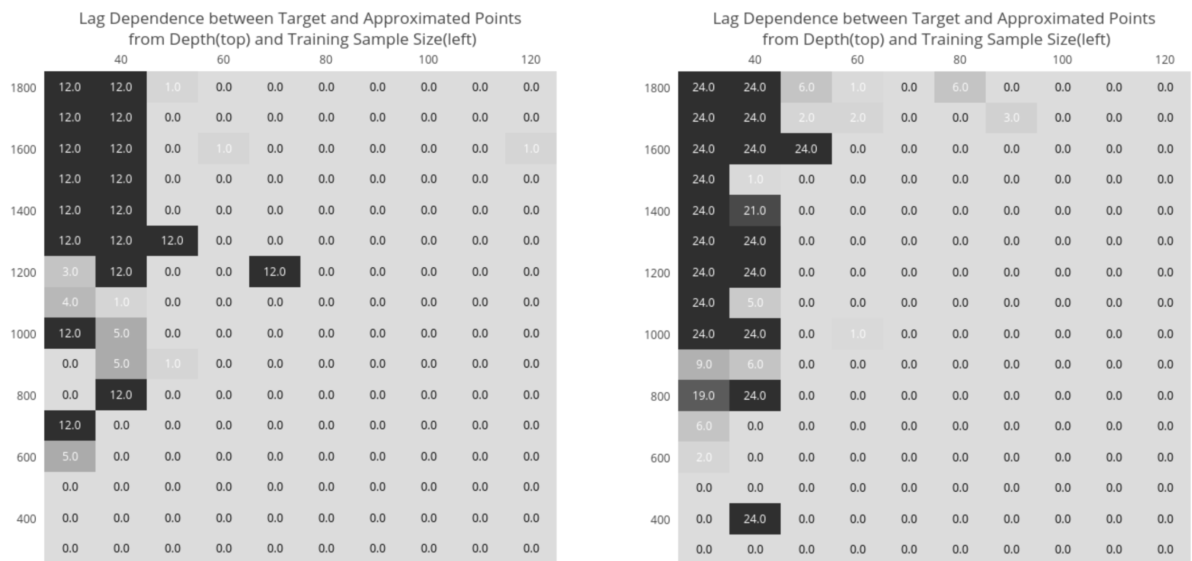


Рис. 3. In-sample нейронної мережі прямого розповсюдження при співвідношенні глибина занурення/розмір навчальної вибірки 30/900

На рис. 3 відображено співвідношення глибини занурення та розміру навчальної вибірки 30/900, такий розмір навчальної вибірки у відношенні до кількості вхідних нейронів повинен забезпечити достатній результат навчання нейронної мережі. Але глибина занурення не забезпечує достатньої вхідної інформації для прогнозування майбутніх значень. На рис. 4 представлено дані аналізу впливу співвідношення глибина занурення/ розмір навчальної вибірки на затримку апроксимованого сигналу, звідки випливає, що затримка (відставання) за недостатньої глибини занурення дорівнює горизонту прогнозування.



а) горизонт прогнозування P=12 **б) горизонт прогнозування P=24**

Рис.4. Затримка між цільовими значеннями та апроксимованими мережею in-sample

Таким чином, щоб мережа відображала модель ряду необхідно збільшувати глибину занурення. З рис. 4 видно, що при збільшенні глибини занурення, затримки між реальними і прогнозованими значеннями in-sample немає. Але якщо якісно проаналізувати in-sample мережі, навченої на вибірці з більшою глибиною занурення (рис. 5), то виникне інша проблема – перенавчання.

Мережа показує високу точність in-sample (на навчальній вибірці), але низьку точність на out-of-sample (тестовій вибірці), що є ознакою перенавчання (overfitting).

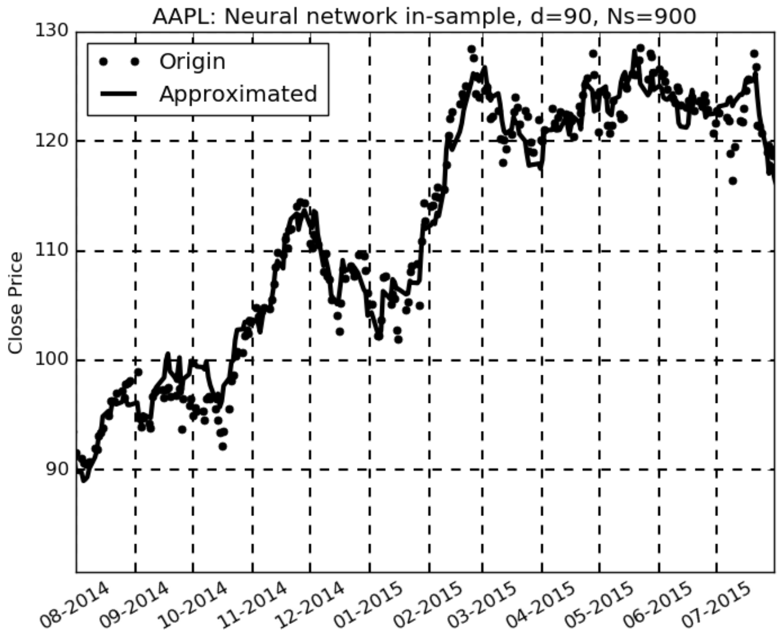


Рис.5. In-sample нейронної мережі прямого розповсюдження при співвідношенні глибина занурення/розмір навчальної вибірки 90/900

Таким чином, необхідно досягти оптимального співвідношення глибина занурення/ розмір навчальної вибірки, яке з одного боку забезпечить достатню кількість вхідної інформації для виключення ефекту запізнення та точності прогнозування, і з другого кількість нейронів мережі, що виключатиме перенавчання.

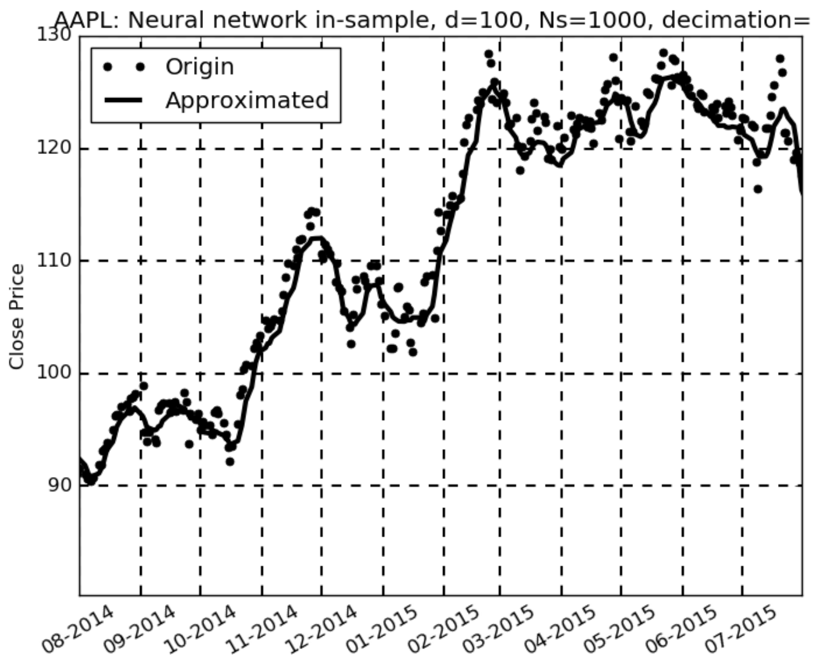


Рис.6. In-sample нейронної мережі прямого розповсюдження при співвідношенні глибина занурення/розмір навчальної вибірки 100/1000, проріджування – 3 значення, попереднє згладжування – фільтр Савіцького-Голея

Одним з шляхів розв'язку може бути проріджння даних ряду, що подаються на вхід нейронної мережі. Це дає можливість зберегти глибину занурення достатньо великою, зменшивши при цьому кількість вхідних нейронів. Для зменшення втрат інформації, що несеться в проріджених даних, часовий ряд слід попередньо згладити, для чого можна використати різні способи фільтрації, зокрема вейвлет-перетворення або фільтри на зразок Савіцького-Голея. На рис. 6 відображено результати in-sample для ряду, попередньо згладженого фільтром Савіцького-Голея і проріджженого через 3 значення. Це дозволяє отримати кількість нейронів вхідного шару втричі меншою за глибину занурення. В даному випадку глибина занурення складатиме $d = 100$, але кількість нейронів вхідного шару $N_i = 100/3 = 33$, що забезпечую оптимальне співвідношення між кількістю нейронів мережі та розміром навчальної вибірки.

Висновки. В статті проаналізовані можливості використання нейронної мережі прямого розповсюдження для прогнозування часових рядів та певні труднощі, які при цьому виникають. Зокрема, досліджено вплив такого важливого параметру як глибина занурення на результативність навчання нейронної мережі. Підтверджено, що недостатня глибина занурення призводить до появи відставання у прогнозованих значеннях на горизонт прогнозування; тоді як значне збільшення глибини занурення зумовлює перенавчання мережі. Запропоновано шлях подолання цього протиріччя за рахунок передобробки часового ряду, з якого формується навчальна вибірка. Зокрема, пропонується використання згладжуючих фільтрів та проріджування. Це дає можливість зберегти величину глибини занурення такою, що містить достатньо інформації для прогнозування часового ряду.

Дана стаття містить рекомендації щодо вибору структури та параметрів нейронної мережі, подальше дослідження потребує проведення ще більшої кількості чисельних експериментів для уточнення співвідношень. Також потребує аналізу вибір глибини занурення, виходячи з конкретних умов заданого довірчого інтервалу та горизонту прогнозування.

Література

1. Бокс. Дж. Анализ временных рядов, прогноз и управление / Дж. Бокс, Г.М. Дженкинс. – М.: Мир, 1974. – 406с.

2. Haykin S. Neural Networks, Secon Edition/ Simon Haykin. - Pearson Education, 1999. – 823p.

3. Mandic D.P. Recurrent Neural Networks for Prediction/ D.P. Mandic, J.A. Chambers – JohnWiley&Sons, 2001. – 295p.

4. Lin Tsungnan. Learning long-term dependencies in NARX recurrent neural networks, IEEE Transactions on Neural Networks/ Tsungnan Lin, Bill G. Horne, Peter Tino, C. Lee Giles - Vol. 7, No. 6, 1996, pp. 1329-1351.

5. Lin Tsungnan. A Delay Damage Model Selection Algorithm for NARX Neural Networks/ Tsungnan Lin, C. Lee Giles, Bill G. Horne, S.Y. Kung. - IEEE Transactions on Signal Processing, “Special Issue on Neural Networks”, Vol. 45, No. 11, 1997, pp. 2719-2730.

6. Siegelmann H. T. Computational capabilities of recurrent NARX neural networks, IEEE Transactions on Systems/ H. T. Siegelmann, B. G. Horne and C. Lee Giles. - Man and Cybernetics, Part B, Vol. 27, No.2, 1997, pp. 208-215.

7. Ежов А.А. Нейрокомпьютинг и его применения в экономике и бизнесе/ А.А. Ежов, С.А. Шумский. – М.: МИФИ, 1998. – 222с.

8. Pollok D.S.G, A Handbook of Time Series Analysis, Signal Processing and Dynamics/ D.S.G. Pollok. - Academic Press, 1999. – 720p.

Рецензія/Peer review : 10.11.2016 р. Надрукована/Printed : 15.12.2016 р.
Рецензент: д.т.н., проф. Полікаровських О.І.