

ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ РОЗШИРЕНОЮ ЗГОРТКОВОЮ НЕЙРОННОЮ МЕРЕЖЕЮ

В статті запропоновано метод прогнозування часових рядів адаптованою розширеною згортковою нейронною мережею за архітектурою, подібною до WaveNet. Використання розширених каузальних згорткових шарів дозволяє збільшити рецептивне поле та врахувати довгострокові залежності в часових рядах. Нейромережева модель є повністю ймовірнісною та авторегресивною, генерує наступне значення часового ряду на основі попередніх, а також на основі додаткової інформації. В нейромережевій архітектурі використовується моделювання спільного розподілу точок як добутку умовних розподілів. Такого типу архітектури здатні моделювати розподіли великої кількості випадкових величин, в тому числі часові ряди. Основною складовою згорткової мережі для прогнозування часових рядів є каузальна згортка. Порядок обчислення згортки передбачає залежність поточного значення часового ряду тільки від попередніх значень, що для 1-D згортки забезпечується зсувом виходу звичайної згортки на відповідну кількість тактів вперед. Таким чином виходи для всіх тактів в процесі навчання розраховуються паралельно. Процес генерації значень часового ряду є послідовним, для кожного такту прогнозується наступне значення, яке використовується для подальшого прогнозування. Для збільшення величини рецептивного поля використовується розширена згортка. Розширена згортка представляє собою згортку, в якій фільтр застосовується до площі, більшої за величину фільтра за рахунок пропуску частини вхідних значень з визначеним кроком. В запропонованій мережі рецептивне поле експоненційно зростає зі збільшенням кількості шарів. Оцінка точності розглянутої нейромережевої моделі проводилася в застосуванні до нелінійних часових рядів, а саме до цін на акції. Були проаналізовані дані трьох компаній, отримані з відкритих джерел, з яких були сформовані навчальна та тестова вибірки. В якості метрик було використано середньоквадратичну похибку та середню абсолютну похибку у відсотках. Розширена каузальна згорткова модель показує кращі точнісні результати і в порівнянні з авторегресивною моделлю, і з рекурентною LSTM моделлю. Нейромережева модель зі згортковими шарами не має рекурентних зв'язків, тому забезпечує вищу швидкість навчання у порівнянні з рекурентними моделями.

Ключові слова: часовий ряд, розширена каузальна згортка, розподіл ймовірностей.

A. KASHTALIAN, O. KASHTALIAN

Khmelnitskyi National University

TIME SERIES PREDICTION WITH DILATED CONVOLUTIONAL NEURAL NETWORK

The method of time series forecasting with the adapted dilated causal convolutional neural network model is proposed in the article. The model architecture is similar to WaveNet. The use of dilated causal convolutional layers allows to increase receptive field and to consider long-term dependencies in time series. The neural network model is full probabilistic and autoregressive, it generates the next time series value on the basis of previous ones and also on the basis of additional information. The modelling joint distribution of points as a multiplication of conditional distributions is used. The such types of architectures is able to model distributions of a big number of random values, including time series. The main part of convolutional neural network for time series forecasting is causal convolution. The order of computation supposes the dependence of current value only from previous values. It provides by shifting output of usual convolution on certain number of time steps forward for 1-D convolution. Thus outputs for all time steps are computed in parallel in training process. The generation process is sequential, the next value is predicted for every time step, and this value is used for further prediction. A dilated convolution is used for increasing receptive field. Dilated convolution is convolution with a filter applied to area bigger than filter size for the account of skipping a part of input values with certain step. A receptive field increases exponentially with increasing layers number in considered network. The estimation of considered neural network model was carried out due to nonlinear time series, in particular to stock prices. The data of three companies was analysed. The data were obtained from open sources and create training and testing samples. Root mean squared error and mean absolute percentage error were used as metrics. The dilated causal convolutional model shows a better accuracy results in comparison both with autoregressive model and recurrent LSTM model. The neural network with convolutional layers does not have recurrent connections, so it provides higher training speed relatively to recurrent models.

Key words: time series, dilated causal convolution, probability distribution.

Вступ. Задача прогнозування часових рядів, в тому числі фінансових, була та є актуальною, оскільки така попередня оцінка є необхідним елементом багатьох видів діяльності, особливо це стосується будь-якої інвестиційної діяльності. Розв'язання цієї задачі є нелегким, оскільки фінансові часові ряди, зокрема такі як ціни на акції, можуть бути нестационарними та нелінійними. Методи аналізу та прогнозування часових рядів можна розділити на чотири категорії: статистичні, розпізнавання образів, машинного навчання та сентиментного аналізу. Ці методи здебільшого відносяться до технічного аналізу з використанням фундаментального аналізу [1].

Статистичні методи аналізу часто передбачають, що часовий ряд відповідає вимогам лінійності та стаціонарності ряду, та використовують дані часового ряду в якості вхідних величин. Одна група статистичних методів включає авторегресивне рухоме середнє (ARMA), авторегресивне інтегроване рухоме середнє (ARIMA), узагальнену авторегресивну умовну гетероскедастичну (GARCH) волатильність, авторегресійну модель з плавним переходом. Інша група статистичних методів включає лінійний дискримінантний аналіз, квадратичний дискримінантний аналіз, регресійні алгоритми.

Розпізнавання образів в загальному є схожим до машинного навчання, однак в аналізі фінансових часових рядів використовується дещо іншим чином. Розпізнавання образів зосереджене на виявленні зразків

та трендів в даних ряду. Технічний аналіз ґрунтується на зразках, виявлених безпосередньо в даних акцій, це включає візуальний аналіз діаграм цін, об'ємів або інших приведених показників. Двома широко використовуваними методами розпізнавання образів в даній області є метод перцепційно важливих точок та порівняння зразків.

Серед методів машинного навчання для прогнозування часових рядів використовуються як більш прості підходи, такі як дерева рішень, дискримінантний аналіз, байєсівський метод. Так само використовуються сучасні більш вдосконалені методи – випадковий ліс, логістична регресія та нейронні мережі. Враховуючи нелінійності в часових рядах використання нейронних мереж стає сучасним трендом в прогнозуванні.

Семантичний аналіз є однією з сучасних складових прогнозування фінансових часових рядів, який дозволяє використати якісну інформацію для розрахунку числових значень. В процесі прогнозування використовується автоматичний аналіз текстової інформації, такої як стрічки новин та публічних повідомлень щодо діяльності компаній. Поширеним підходом до прогнозування часових рядів є поєднання декількох методів з вище згаданих груп.

Аналіз досліджень та публікацій. Оскільки часові ряди в багатьох областях є нестационарними та мають нелінійність, нейронні мережі стають все більш вживаними для прогнозування часових рядів порівняно з іншими підходами. Цьому також сприяє швидкий розвиток науковцями різноманітних архітектур нейронних мереж. Основними архітектурами, адаптованими для прогнозування, є повнозв'язні мережі прямого розповсюдження, рекурентні та згорткові мережі.

Повнозв'язними нейронними мережами прямого розповсюдження є нелінійна авторегресійна мережа (NAR) та нелінійна авторегресійна мережа з зовнішнім входом/входами (NARX) [2, 3]. NAR використовує повнозв'язні шари для прогнозування майбутніх значень на основі попередніх. NARX відрізняється тим, що на входи, крім значень безпосередньо часового ряду, подається також додаткова інформація.

Рекурентні нейронні мережі внаслідок своєї будови органічно підходять для моделювання послідовностей, в тому числі часових рядів. Також рекурентні нейронні мережі мають властивість пам'яті на відміну від мереж прямого розповсюдження. Для прогнозування використовують класичні рекурентні архітектури, такі як мережі Елмана та Джордана [4]. Останні декілька років були розроблені архітектури рекурентних нейронних мереж, які у порівнянні з класичними мають властивості «довгої» та «короткої» пам'яті, зокрема це Long-Short Term Memory (LSTM) та Gated Recurrent Unit (GRU), що дозволяє враховувати при прогнозуванні більш довгострокові залежності [5].

Згорткові нейронні мережі набули свого розвитку здебільшого в обробці зображень. Однак з розробкою каузальних згорткових шарів, розширених згорткових шарів та їх поєднання доцільно такі рішення застосовувати в області аналізу та прогнозування часових рядів [6]. Подальший матеріал більш розгорнуто пояснює використання розширеної каузальної згорткової мережі для цієї задачі.

Архітектура нейронної мережі для прогнозування часових рядів. Розширена згорткова нейронна мережа типу WaveNet [7] застосовується для аналізу, обробки та генерації аудіо сигналів. В нейромережевій архітектурі використовується моделювання спільного розподілу пікселів або слів як добутку умовних розподілів. Такого типу архітектури здатні моделювати розподіли понад тисячі випадкових величин [8], в тому числі зображення, аудіо сигнали, часові ряди.

Спільна ймовірність часового ряду $\mathbf{x} = \{x_1, \dots, x_T\}$ розкладається на множники як добуток умовних ймовірностей таким чином

$$p(\mathbf{x}) = \prod_{t=1}^T p(x_t | x_1, \dots, x_{t-1}).$$

Кожна точка x_t залежить від точок усіх попередніх тактів.

Генерація майбутніх точок часового ряду проводиться подібно до нейромережевої архітектури PixelCNN [9]. В PixelCNN порядок залежностей пікселів визначається як при растровому скануванні, генерується рядок за рядком, точка за точкою в кожному рядку.

Основною складовою згорткової мережі для прогнозування часових рядів є каузальна (причинна) згортка. На рис. 1 зображено стек каузальних згорткових шарів. Такий порядок розрахунку забезпечує коректне моделювання даних – для визначення розподілу поточного значення $p(x_{t+1} | x_1, \dots, x_t)$ використовуються тільки попередні значення часового ряду і не враховуються наступні $x_{t+1}, x_{t+2}, \dots, x_T$. Для 1-D згортки такий порядок забезпечується зсувом виходу звичайної згортки на відповідну кількість тактів вперед.

В процесі навчання мережі згорткові виходи для всіх тактів можуть розраховуватися паралельно, оскільки всі такти \mathbf{x} відомі. Процес генерації майбутніх значень є послідовним: для кожного такту прогнозується майбутнє значення, потім воно використовується для прогнозування наступного значення, і цей процес циклічно повторюється.

Нейромережева модель з каузальними згортковими шарами не має рекурентних зв'язків, що зазвичай забезпечує вищу швидкість навчання в порівнянні з рекурентними нейронними мережами, особливо у випадку довгих послідовностей. Однією з проблем використання каузальних згорткових шарів є

необхідність використання великої кількості шарів або фільтрів з значним рецептивним полем. Наприклад, на рис. 1 рецептивне поле дорівнює 5. Тому для збільшення величини рецептивного поля використовується розширена згортка, яка дозволяє збільшити рецептивне поле без значних втрат інформації та часу обчислень.

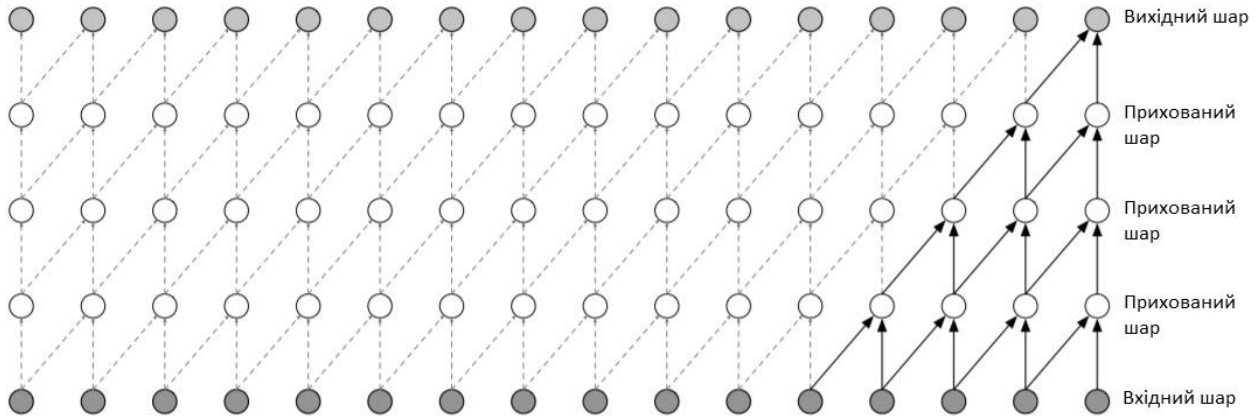


Рис. 1. Стек каузальних згорткових шарів

Розширена згортка представляє собою згортку, в якій фільтр застосовується до площі, більшої за величину фільтра за рахунок пропуску частини вхідних значень з визначеним кроком. Це еквівалентно згортці з фільтром більшої величини, отриманим зі звичайного фільтра розширенням його нульовими значеннями. Такий підхід дає ефект, схожий на використання пулінг шару або згортки з кроком, більшим за одиницю, але при цьому вихід згорткового шару має той же розмір, що і вхід. Згортка з розширенням 1 перетворюється до звичайної згортки. На рис. 2 зображено архітектуру нейронної мережі з розширеними каузальними шарами з розширеннями 1, 2, 4, 8. Стек розширених згорткових шарів дозволяє отримати достатньо велике рецептивне поле порівняно невеликою кількістю шарів, при цьому час обчислення не збільшується. В нейронній мережі WaveNet використовується стек згорткових шарів з розширеннями 1, 2, 4, ..., 512, 1, 2, 4, ..., 512, 1, 2, 4, ..., 512. Таким чином рецептивне поле експоненційно зростає зі збільшенням кількості шарів.

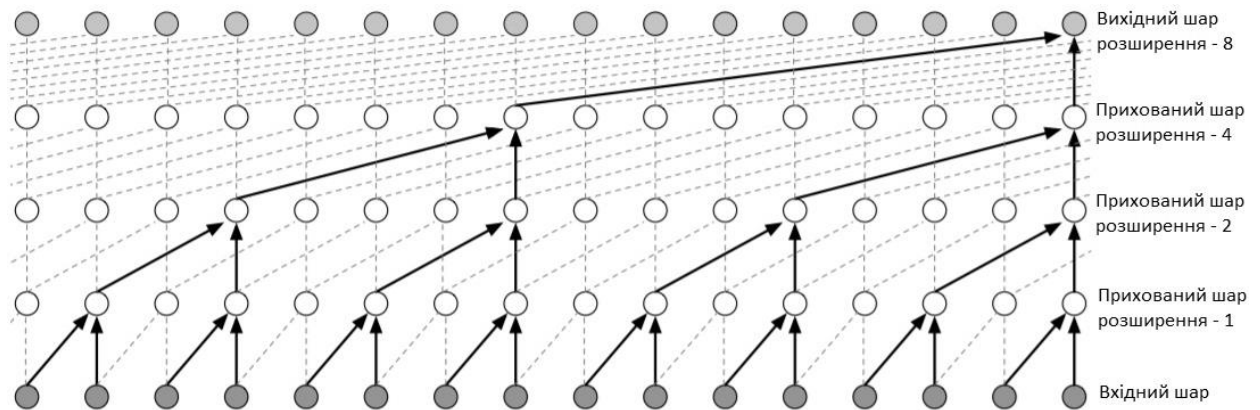


Рис. 2. Стек розширених каузальних згорткових шарів

Модифікована нейронна мережа на основі WaveNet складається з умовної мережі $c = C(v)$, яка перетворює низькочастотні ознаки v до вищих частот, та авторегресивного процесу $P(y_i | c, y_{i-1}, \dots, y_{i-R})$, який прогнозує наступну точку процесу на основі наявних умов поточного такту та попередніх R тактів [10]. R – це розмір рецептивного поля і ϵ параметром, який визначається архітектурою нейронної мережі. Структурна схема нейронної мережі для генерування значень часового ряду наведена на рис. 3.

Структура авторегресивної мережі параметризована числом шарів ℓ , кількістю пропущених каналів s та кількістю залишкових каналів r .

Для зменшення кількості ймовірностей часовий ряд квантується до $a = 256$ з використанням компресії за μ -законом

$$f(x_t) = \text{sign}(x_t) \frac{\ln(1 + \mu|x_t|)}{\ln(1 + \mu)},$$

де $-1 < x_t < 1$ та $\mu=255$. Таке нелінійне квантування дає помітно кращі результати ніж звичайне лінійне квантування.

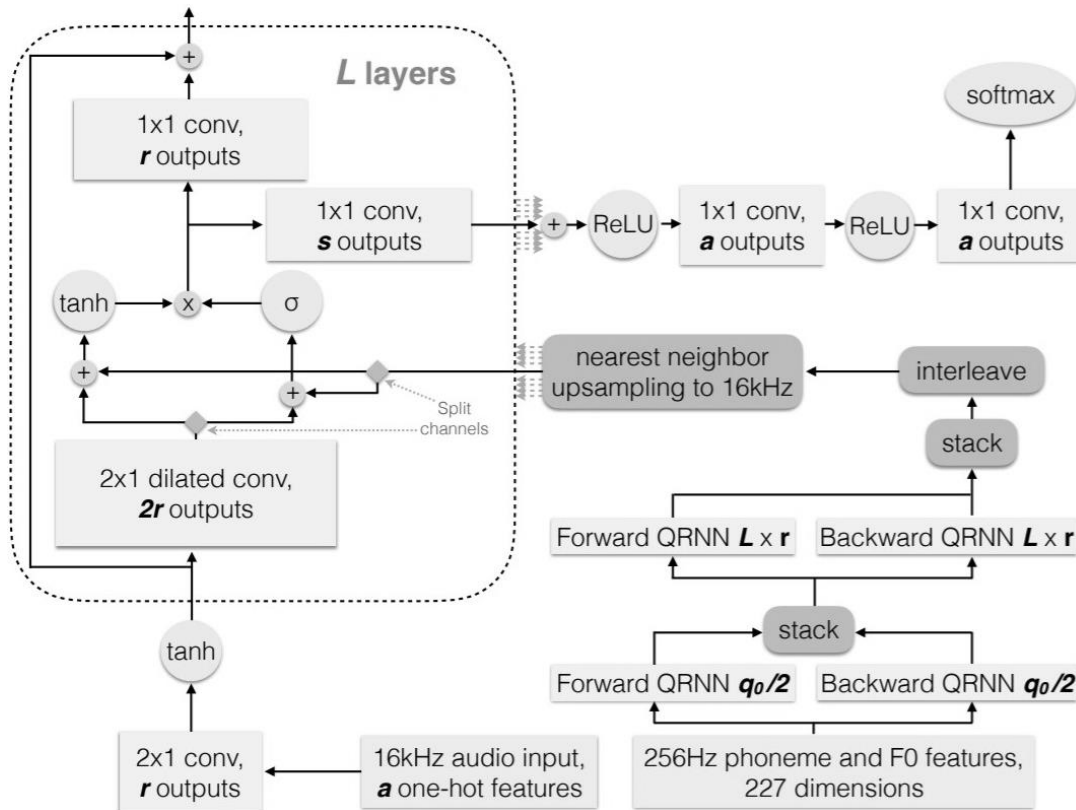


Рис. 3. Структурна схема нейронної мережі WaveNet для генерування значень часового ряду

Кодовані значення проходять через шар згортки 2x1, який генерує входи $x^{(0)} \in \mathbf{R}^r$ для першого шару стеку залишкових шарів:

$$x^{(0)} = W_{embed} * y + B_{embed},$$

де * позначає оператор однорозмірної згортки. Кожний шар послідовності обчислює вектор внутрішніх станів $h^{(i)}$ і потім, відповідно до залишкових зв'язків між шарами, додає до входу $x^{(i-1)}$ для генерації виходів $x^{(i)}$:

$$h^{(i)} = \tanh(W_h^{(i)} * x^{(i-1)} + B_h^{(i)} + L_h^{(i)}) \cdot \sigma(W_g^{(i)} * x^{(i-1)} + B_g^{(i)} + L_g^{(i)}), \quad 1)$$

$$x^{(i)} = x^{(i-1)} + W_r^{(i)} \cdot h^{(i)} + B_r^{(i)},$$

де $L^{(i)}$ – вихід шару умовної мережі. Оскільки кожний шар додає власні виходи до входів, розмірність шарів повинна залишатися рівною числу залишкових каналів r . Цього вдається досягнути згорткою з r вхідними та $2r$ вихідними каналами. В процесі генерації ця згортка замінюється двома добутками матриця-вектор з матрицями W_{prev} та W_{cur} . Таким чином розрахунок $h^{(i)}$ для окремого такту (1) перетворюється до вигляду:

$$h^{(i)} = W_{prev}^{(i)} \cdot x_{t-d}^{(i-1)} + W_{cur}^{(i)} \cdot x_t^{(i-1)} + B^{(i)} + L^{(i)},$$

$$h^{(i)} = \tanh(h_{0r}^{(i)}) \cdot \sigma(h_{r:2r}^{(i)}),$$

де $L^{(i)}$ – конкатенація $L_h^{(i)}$ та $L_g^{(i)}$, $B^{(i)}$ – конкатенація $B_h^{(i)}$ та $B_g^{(i)}$.

Прихований стан $h^{(i)}$ кожного шару з 1-го по ℓ -й об'єднується з W_{skip} до кількості пропущених каналів s :

$$h = \begin{bmatrix} h^{(1)} \\ h^{(2)} \\ \vdots \\ h^{(\ell)} \end{bmatrix}, \quad h \in \mathfrak{R}^{\ell r},$$

$$z_s = \text{relu}(W_{skip} \cdot h + B_{skip}), \quad z_s \in \mathfrak{R}^s,$$

де $\text{relu}(x) = \max(0, x)$.

Потім z_s подається на два повнозв'язаних шари для генерації вихідного розподілу $p \in \mathfrak{R}^a$:

$$z_a = \text{relu}(W_{relu} \cdot z_s + B_{relu}), \quad z_a \in \mathfrak{R}^a,$$

$$p = \text{soft max}(W_{out} \cdot z_a + B_{out}).$$

Модель, навчена без умовної інформації, прогнозує майбутні значення часового ряду тільки на основі попередніх значень. Для підвищення точності прогнозування кожне значення прогнозується не тільки на основі попередніх значень, а також на основі додаткової умовної інформації. Додаткова умовна інформація зміщує кожний шар вектором, генерованим з додаткових вхідних сигналів. Якщо додаткова інформація змінюється з меншою частотою, ніж безпосередньо значення часового ряду, то використовується процедура підвищувальної дискретизації шляхом повторення або розгортки. В даній архітектурі додаткові вхідні сигнали проходять через два двонаправлених квазі-RNN (QRNN) шари з пулінгом та 2x1 згортою. Однонаправлений QRNN шар визначається такими рівняннями:

$$\begin{aligned}\tilde{h} &= \tanh(W_h * x + B_h), \\ o &= \sigma(W_o * x + B_o), \\ f &= \sigma(W_f * x + B_f), \\ h_t &= f_t \cdot h_{t-1} + (1 - f_t) \cdot \tilde{h}_t, \\ z_t &= o_t \cdot h_t.\end{aligned}$$

Результат двонаправленого QRNN шару отримується шляхом об'єднання результатів двох однонаправлених QRNN шарів, одного для вхідної послідовності та одного для реверсивної копії вхідної послідовності. Після двонаправленого QRNN шару додаткові сигнали нижчої частоти повторюються або розгортаються.

На кожному такті модель генерує розподіл точок $P(s)$, визначений попередніми значеннями часового ряду та додатковими умовними сигналами. Є декілька варіантів використання генерованого розподілу:

а) пряма вибірка – випадкова вибірка з $P(y)$;

б) температурна вибірка – випадкова вибірка з розподілу відповідно до температури t :

$$\tilde{P}_t(y) = \frac{1}{Z} P(y)^{1/t}, \text{ де } Z - \text{нормалізована константа};$$

в) вибірка середнього значення розподілу: $E_p[y]$;

г) вибірка найбільш ймовірної точки: $\arg \max P(y)$;

д) вибірка з корегованого розподілу, що дозволяє тільки k верхніх точок:

$$\tilde{P}_k(y) = \begin{cases} 0 & \text{якщо } y < k_{th}(P(y)) \\ P(y)/Z & \text{інакше} \end{cases}, \text{ де } Z - \text{нормалізована константа.}$$

В режимі прогнозування значень часового ряду нелінійні функції апроксимуються з високою точністю відповідними співвідношеннями. Враховуючи, що $\tilde{e}(x)$ є апроксимацією $e^{|x|}$, для функції гіперболічного тангенсу та сигмоїдальної функції використовуються такі вирази:

$$\begin{aligned}\tanh(x) &\approx \text{sign}(x) \frac{\tilde{e}(x) - 1/\tilde{e}(x)}{\tilde{e}(x) + 1/\tilde{e}(x)}, \\ \sigma(x) &= \begin{cases} \frac{\tilde{e}(x)}{1 + \tilde{e}(x)}, & x \geq 0, \\ \frac{1}{1 + \tilde{e}(x)}, & x < 0. \end{cases}\end{aligned}$$

Для представлення $\tilde{e}(x)$ обрано поліном четвертого ступеня.

Для апроксимації e^x використано апроксимацію 2^x і співвідношення $e^x = 2^{x/\ln 2}$:

$$2^x = 2^{[x]} \cdot 2^{x-[x]} = 2^{[x]} \cdot \left(1 + \left(2^{x-[x]} - 1\right)\right),$$

де $0 \leq 2^{x-[x]} - 1 < 1$, оскільки $0 \leq x - [x] < 1$.

Розмірність вихідних значень дорівнює розмірності вхідних за рахунок відсутності pooling шарів. Виходом моделі є категорійний розподіл величини x_t з softmax шару.

Результати експерименту. Оцінка розглянутої нейромережевої архітектури типу WaveNet проводилася в застосуванні до нелінійних часових рядів, якими зокрема є фінансові часові ряди, а саме ціни на акції. Всі дані бралися з відкритих джерел, ресурсу Yahoo Finance (<https://finance.yahoo.com>). Були проаналізовані дані трьох компаній: Xilinx Inc. (XLNX), Walmart Inc. (WMT) та Apple Inc. (AAPL), – в комбінації з об'ємами продажів акцій та додатковими даними індексів: S&P 500, Russel 2000, Dow 30, Nasdaq, VIX.

Навчальна вибірка була сформована з даних за п'ять років (з 2014 р. до 2018 р. включно) для врахування довгострокових трендів у змінах на ціни акцій. В якості значення ціни на акції бралася зважена ціна закриття. Тестова вибірка включила дані за перші шість місяців поточного року (січень – червень 2019 р.), на яких проводилося прогнозування на один день вперед. Дані навчальної та тестової вибірок були нормалізовані за співвідношенням

$$y = \frac{x - \mu}{\sigma},$$

де x – вихідні дані, μ – середнє значення, σ – середньоквадратичне відхилення.

В якості метрик для тестів було використано середньоквадратичну похибку (root mean squared error – RMSE) та середню абсолютну похибку у % (maximum absolute percentage error – MAPE).

Для порівняння взято результати прогнозування часових рядів існуючими моделями: відомою в економетриці авторегресивною моделлю (AR) та рекурентною моделлю на основі LSTM [11]. Результати прогнозування наведено в табл. 1.

Таблиця 1

Результати прогнозування

	Авторегресивна модель (AR)		Рекурентна модель (LSTM)		Згорткова модель (ConvNet)	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
XLNX	0,34	2,05	0,11	1,54	0,08	1,23
WMT	0,32	1,55	0,10	0,91	0,07	0,75
AAPL	0,19	1,64	0,07	1,19	0,05	0,88

Таким чином, розширена каузальна згорткова мережа показує вищі точнісні результати у порівнянні з розглянутими моделями.

Висновки і перспективи подальших досліджень. В статті проаналізовано застосування розширеної каузальної згорткової нейронної мережі типу WaveNet для прогнозування часових рядів. Завдяки використанню розширеної згортки збільшено рецептивне поле, що дає можливість враховувати більш довгострокові залежності в часових рядах. Також мережа має додаткові входи, які отримують інформацію з додаткових даних, від яких залежить часовий ряд. Прогнозування часових рядів, особливо нелінійних, є складною задачею, для якої розроблено досить велику кількість підходів. Розглянута розширена каузальна згорткова мережа є порівняно простою, ефективною та зрозумілою моделлю, яка показує результати порівняні та кращі відносно відомих моделей, таких як авторегресійна модель та рекурентна нейромережева LSTM модель. Крім того розглянута згорткова мережа показує кращий час навчання порівняно з рекурентними мережами. Архітектура типу WaveNet є досить новою і має ресурси для подальшого її вдосконалення. Одним з таких напрямків є досягнення компромісу між кількістю шарів та фільтрів, що потрібно для моделювання нелінійності, та уникненням перенавчання.

Література

1. Dev Shah, Haruna Isah, Farhana Zulkernine. Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. *International Journal of Financial Studies*, 2019, 7, 26, P. 1–21.
2. Boussaada Z., Curea O., Remaci A., Camblong H., Bellaaj N.M. A nonlinear autoregressive exogenous (NARX) Neural Network Model for the Prediction of the Daily Direct Solar Radiation. *Energies*, 2018, Issue 11, P. 620–641.
3. Kashtalian A. The features of feedforward neural network use for time series forecasting. *Herald of Khmelnytskyi National University: Technical Science*. Khmelnytsky: KhNU, 2016, Issue 6(243), P. 210–215.
4. Wang J., Wang J., Fang W., Niu H. Financial Time Series Prediction Using Elman Recurrent Random Neural Networks. *Computational Intelligence and Neuroscience*, 2016, Issue 12, P. 1–14.
5. Petnehazi G. Recurrent Neural Networks for Time Series Forecasting. arXiv:1901.00069v1 [cs.LG] 1 Jan 2019. URL: <https://arxiv.org/pdf/1901.00069.pdf>. Accessed 10 Sep 2019.
6. Borovykh A., Bohte S., Oosterlee C.W. Conditional time series forecasting with convolutional neural networks. arXiv:1703.04691v5 [stat.ML] 17 Sep 2018. URL: <https://arxiv.org/pdf/1703.04691.pdf>. Accessed 10 Sep 2019.
7. Van den Oord A., Dieleman S., Zen H., Simonyan K., Vinyals O., Graves A., Kalchbrenner N., Senior A., Kavukcuoglu K. WaveNet: A Generative Model for Raw Audio. arXiv:1609.03499v2 [cs.SD] 19 Sep 2016. URL: <https://arxiv.org/pdf/1609.03499.pdf>. Accessed 10 Sep 2019.
8. Aaron van den Oord, Kalchbrenner N., Kavukcuoglu K. Pixel recurrent neural networks. arXiv:1601.06759v3 [cs.CV] 19 Aug 2016. URL: <https://arxiv.org/pdf/1601.06759.pdf>. Accessed 19 Sep 2019.
9. Van den Oord A., Kalchbrenner N., Vinyals O., Espeholt L., Graves A., Kavukcuoglu K. Conditional Image Generation with PixelCNN Decoders. arXiv:1606.05328v2 [cs.CV] 18 Jun 2016. URL: <https://arxiv.org/pdf/1606.05328.pdf>. Accessed 19 Sep 2019.
10. Arik S. O., Chrzanowski M., Coates A., Diamos G., Gibiansky A., Kang Y., Li X., Miller J., Ng A., Raiman J., Sengupta S., Shoyebi M. Deep Voice: Real-time Neural Text-to-Speech. arXiv:1702.07825v2 [cs.CL] 7 Mar 2017. URL: <https://arxiv.org/pdf/1702.07825.pdf>. Accessed 19 Sep 2019.
11. Fisher T., Krauss C. Deep Learning with Long Short-Term Memory networks for financial market predictions. *FAU Discussion papers in Economics*, 2017, Issue 11, P. 961–970.