

КЕРУВАННЯ В ТЕХНІЧНИХ, ЕКОНОМІЧНИХ ТА БІОЛОГІЧНИХ СИСТЕМАХ

УДК 004.852, 519.246.8

М.Я. Кушнір, К.А. Токарева

ОДНЕ УЗАГАЛЬНЕННЯ LSTM-НЕЙРОННИХ МЕРЕЖ

Кушнір Микола Ярославович

Чернівецький національний університет імені Юрія Федьковича,
orcid: 0000-0001-9480-3856

myk.kushnir@chnu.edu.ua

Токарева Катерина Анатоліївна

Чернівецький національний університет імені Юрія Федьковича,
orcid: 0000-0003-4007-8745

tokarieva.chnu@gmail.com

Основна мета публікації — розробка узагальненої LSTM-нейронної мережі з урахуванням передісторії скінченної глибини. Більшість робіт з даного напрямку присвячено багатопотоковим узагальненням з наявністю кореляції між потоками, тобто розглядається розширення LSTM-моделі в ширину аналогічно до панельних даних у часових рядах. Проводиться розширення моделі в глибину, що дозволить враховувати періодичні компоненти часових рядів та інші детерміновані складові процесу, тобто без передобробки даних. Як показано в прикладі, розробка даної моделі дозволяє розширити використання класичних LSTM-мереж на динамічні системи з чітко вираженою періодичною складовою. Крім самої розширеної моделі, розглядається проблема обчислення параметрів моделі, яке ґрунтується на методі зворотного поширення. Для цього обчислено частинні похідні результуючого вихідного сигналу за невідомими параметрами та показано рекурентність даних похідних за часом аналогічно класичній LSTM-моделі. Додатковою проблемою при створенні розширеної моделі є оцінка гіперпараметра ρ , який вказує на глибину передісторії процесу. В зв'язку з цим розроблено алгоритм оцінки гіперпараметра ρ з подальшою оцінкою всіх параметрів моделі.

Ключові слова: нейронні мережі, оцінка параметрів, метод зворотного поширення, часові ряди.

Вступ

Використання стохастичних моделей як математичного апарата для побудови прогностичних систем набуло бурхливого розвитку у другій половині XX століття. Це пов'язано насамперед зі швидким розвитком теорії стохастичних диференціальних рівнянь [1–3], часових рядів [4–6] та прикладних задач, що зу-

© М.Я. КУШНІР, К.А. ТОКАРЕВА, 2023

*Міжнародний науково-технічний журнал
Проблеми керування та інформатики, 2023, № 4*

мовили розвиток теоретичних здобутків. Особливу увагу автори звертають на дослідження часових рядів, оскільки дані моделі мають низку переваг. По-перше, вони описують дискретні системи, що значно спрощує дослідження реальних систем та побудову прогнозів. По-друге, велика кількість реальних даних є дискретними і описуються рекурентними формулами або послідовностями випадкових величин.

В останні два десятиліття в прикладних задачах широко використовуються методи штучного інтелекту, глибинного інтелекту, машинного навчання. Цей феномен можна пояснити збільшенням обчислювальних потужностей для розв'язання прикладних задач, що сприяло використанню складних систем, таких як нейронні мережі та генетичні алгоритми. У даній роботі увагу звернено на узагальнення однієї з відомих нейронних мереж, а саме Long Short-Term Memory (LSTM) [7–10].

Теорія нейронних мереж є одним з найбільш розвинутих напрямків штучного інтелекту, машинного та глибинного навчання. Безсумнівно, саме нейронні мережі — найбільш точні та якісні моделі в задачах класифікації реальних процесів різної природи. Слід зауважити, що вибір топології нейронної мережі залежить від конкретних даних та важко описується на основі тих чи інших характеристик даних. Тому процес підбору топології нейронної мережі швидше є задачею дослідницького, ніж теоретичного характеру. Це насамперед продиктовано складністю опису характеристик вихідного сигналу (вихідних параметрів) для нейронних мереж зі складною структурою, причому основна проблема полягає у наявності функції активації, яка нелінійно змінює сигнал між шарами нейронної мережі. У роботі розглянуто використання розширення LSTM-нейронної мережі для побудови прогнозу фінансових активів. Слід зауважити, що дослідженню побудови прогнозів присвячена велика кількість робіт, які використовують різні припущення щодо властивостей реальних процесів.

У [11] розглянуто прогноз динаміки захворюваності на COVID-19 у м. Києві та Чехії на основі ланцюгів Маркова для основного процесу

$$S(t) = S(t-1) - i(t)$$

та допоміжного процесу $i(t)$, що описуються нелінійним часовим рядом, де $S(t)$ — кількість сприйнятливих до хвороби осіб в день t , $i(t)$ — кількість нових інфікованих осіб в день t . Слід зауважити, що саме допоміжний процес $i(t)$ будується на основі нелінійного часового ряду, що враховує передісторію процесу. Іншим підходом побудови прогнозів для набору випадкових процесів є використання часових рядів [12]. У даній роботі розглянуто оптимізаційну задачу для оцінки параметрів випадкового поля.

У роботі [13] використовується множинний підхід для оптимізації портфеля цінних паперів. Зауважимо, що основним припущенням даної роботи є велика кількість допустимих стратегій, що унеможливорює пряму перевірку всіх стратегій.

У даній роботі досліджено одне розширення LSTM-нейронної мережі, яке є потужним засобом для аналізу задач прогнозування часових рядів [14, 15], задач класифікації [16, 17], задач розпізнавання мовлення [18], задач розпізнавання рукописних текстів [19] тощо. Крім того, LSTM-нейронні мережі є частинним випадком більш загальних рекурентних нейронних мереж. Останні, в свою чергу, будуються на принципі послідовного наближення вихідного сигналу з використанням ітераційного процесу з тими ж вагами на кожному кроці. Ґрунтуючись на роботі [7], топологію LSTM-нейронної мережі можна описати такими співвідношеннями:

$$\begin{cases} f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f), \\ i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i), \\ o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o), \\ \tilde{c}_t = \sigma_c(W_c x_t + U_c h_{t-1} + b_c), \\ c_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t, \\ h_t = o_t \otimes \sigma_h(c_t), \end{cases} \quad (1)$$

де $x_t \in R^d$ — значення вхідного сигналу, $f_t, i_t, o_t \in (0, 1)^h$ — допоміжні вектори (вектори вентилів), $\tilde{c}_t \in (-1, 1)^h$ — вхідний вектор активації клітинки, $c_t \in R^{h \times d}$ — вихідний вектор комірки (вектор стану комірки), $h_t \in (-1, 1)^h$ — вихідний сигнал, $W_f, W_i, W_o, W_c \in R^{h \times d}$, $U_x, U_i, U_o, U_c \in R^{h \times d}$, $b_f, b_i, b_o, b_c \in R^h$ — ваги нейронної мережі, d — розмірність вхідного сигналу, h — розмірність сигналу з прихованих шарів нейронної мережі, σ_g, σ_c — функції активації нейронної мережі, $\otimes$ — множення Адамара.

У роботах [9, 10] розглянуто розширення LSTM-нейронної мережі на випадок залежності вихідного сигналу від деяких зовнішніх чинників. У цьому разі залежна змінна (вихідний сигнал) залежить від рівня деякої факторної змінної. Наприклад, у [20] автори розглядають декілька дескрипторів, які впливають на залучення дітей під час перегляду відео із супроводом текстової інформації. А саме, цими дескрипторами є положення голови дитини, гістограма орієнтованих градієнтів і гістограма орієнтованого потоку навколо верхньої частини тіла. У цьому разі топологія розширеної нейронної моделі (Multi View LSTM, MV-LSTM) формально описується таким чином:

$$\begin{cases} f_t^{(v)} = \sigma_g(W_f^{(v)} x_t + U_f A h_{t-1}^{(v)} + \sum_{k \neq v} U_f B h_{t-1}^{(k)} + b_f), \\ i_t = \sigma_g(W_i^{(v)} x_t + U_i A h_{t-1}^{(v)} + \sum_{k \neq v} U_i B h_{t-1}^{(k)} + b_i), \\ o_t = \sigma_g(W_o^{(v)} x_t + U_o A h_{t-1}^{(v)} + \sum_{k \neq v} U_o B h_{t-1}^{(k)} + b_o), \\ \tilde{c}_t = \sigma_c(W_c^{(v)} x_t + U_c A h_{t-1}^{(v)} + \sum_{k \neq v} U_c B h_{t-1}^{(k)} + b_c), \\ c_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t, \\ h_t = o_t \otimes \sigma_h(c_t), \end{cases} \quad (2)$$

де $v \in \{1, \dots, V\}$ — рівень факторної змінної, матриці A та B визначаються із співвідношень

$$A_{ij} = \begin{cases} 1, & i = j \leq \alpha d, \\ 0 & \text{в іншому разі;} \end{cases}$$

$$B_{ij} = \begin{cases} 1, & i = j \geq (1 - \beta)d, \\ 0 & \text{в іншому разі;} \end{cases}$$

$\alpha, \beta \in [0, 1]$ — гіперпараметри нейронної мережі MV-LSTM, які вказують на тісний зв'язок між різними рівнями вихідного сигналу $v \in \{1, \dots, V\}$. Слід зауважити, що модель (2) можна інтерпретувати як панельну модель, де одночасно відслідковується декілька реалізацій одного процесу або явища з можливо різними значеннями вихідного сигналу.

Основна модель

Розглянемо переваги та недоліки LSTM- чи MV-LSTM-нейронних мереж, які є класичним прикладом нейронних мереж для одно- та багатопотокових процесів. Переваги даних топологій насамперед — це простота задання та врахування скінченної передісторії у вигляді одного лагу. Основні недоліки моделей (1) та (2) — це використання лише першого лагу в моделі LSTM чи MV-LSTM. При цьому виникає проблема в аналізі часових рядів з чітко визначеною періодичністю. Наприклад, аналіз часового ряду

$$x_t = \varphi x_{t-k} + \sum_{i=1}^{k-1} \varphi_i x_{t-i} + \varepsilon_t, \quad (3)$$

де $|\varphi| < 1$, $\sum_{i=1}^{k-1} |\varphi_i| \ll |\varphi|$, $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$, вказує на низьку точність прогнозу на основі LSTM-нейронної мережі. У задачах побудови прогнозу для часових рядів у класичних моделях припускається, що нинішній стан процесу залежить від його стану в попередні моменти, тобто

$$x_t = f(x_{t-1}, x_{t-2}, \dots).$$

Для подальшого аналізу будемо використовувати топологію розширеної LSTM-нейронної мережі (extended LSTM, E-LSTM), яка визначатиметься такими співвідношеннями:

$$\begin{cases} f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f), \\ i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i), \\ o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o), \\ \tilde{c}_t = \sigma_c(W_c x_t + U_c h_{t-1} + b_c), \\ c_t = \sum_{k=0}^p f_{t-k} \otimes c_{t-1-k} + i_t \otimes \tilde{c}_t, \\ h_t = o_t \otimes \sigma_h(c_t). \end{cases} \quad (4)$$

Як видно з опису, різниця між LSTM та E-LSTM полягає у використанні AR-моделі для випадкового процесу c_t , де коефіцієнти даної AR-моделі виступають як наближення $f_t, t \geq 0$.

Оцінка коефіцієнтів

Розглянемо метод зворотного поширення (backward propagation) для моделі (4), який використовується для оцінки параметрів моделі

$$P_{ar} = (W_f, U_f, b_f, W_i, U_i, b_i, W_o, U_o, b_o, W_c, U_c, b_c)$$

із застосуванням ітераційних методів. Для оцінки параметрів моделі (4) визначимо функціонал якості

$$Q(Par) = Q(Par; Y),$$

де $Y = (y_1, y_2, \dots, y_T)$ визначає значення вихідного сигналу, причому на основі означення моделі (1), $y_t \in (-1, 1)^h$. Зазначимо, що на основі параметрів Par однозначно визначаються значення оцінок

$$\hat{Y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_T) = (h_1, \dots, h_T).$$

Функція Q є цільовою і відображає якість оцінки моделі. Зазвичай як функція Q вибирається функція правдоподібності, функція втрат або сума квадратів залишків.

Розглянемо спочатку частинні похідні $\hat{Y}$ за параметрами Par . Оскільки схема (4) пропонує покрокове обчислення h_t , то обчислення частинних похідних проводитимемо з кінця до початку:

$$\begin{aligned} \frac{\partial \hat{y}_t}{\partial b_c} &= \frac{\partial h_t}{\partial b_c} = \frac{\partial}{\partial b_c} \left(o_t \otimes \sigma_h \left(\sum_{k=0}^p f_{t-k} \otimes c_{t-1-k} + i_t \otimes \tilde{c}_t \right) \right) = \frac{\partial}{\partial b_c} \sigma_g \times \\ &\times (W_o x_t + U_o h_{t-1} + b_o) \otimes \sigma_h \left(\sum_{k=0}^p f_{t-k} \otimes c_{t-1-k} + i_t \otimes \tilde{c}_t \right) + \sigma_g (W_o x_t + U_o h_{t-1} + \\ &+ b_o) \otimes \frac{\partial}{\partial b_c} \sigma_h \left(\sum_{k=0}^p f_{t-k} \otimes c_{t-1-k} + i_t \otimes \tilde{c}_t \right) = \sigma'_g (W_o x_t + U_o h_{t-1} + b_o) U_o \times \\ &\times \frac{\partial}{\partial b_c} h_{t-1} \otimes \sigma_h \left(\sum_{k=0}^p f_{t-k} \otimes c_{t-1-k} + i_t \otimes \tilde{c}_t \right) + \sigma_g (W_o x_t + U_o h_{t-1} + b_o) \otimes \\ &\otimes \sigma'_h \left(\sum_{k=0}^p f_{t-k} \otimes c_{t-1-k} + i_t \otimes \tilde{c}_t \right) \frac{\partial}{\partial b_c} \left(\sum_{k=0}^p f_{t-k} \otimes c_{t-1-k} + i_t \otimes \tilde{c}_t \right) = \\ &= \sigma'_g (o_t) U_o \frac{\partial}{\partial b_c} h_{t-1} \otimes \sigma_h(c_t) + \sigma_g(o_t) \otimes \sigma'_h(c_t) \frac{\partial}{\partial b_c}(c_t), \end{aligned}$$

де

$$\begin{aligned} \frac{\partial}{\partial b_c}(c_t) &= \frac{\partial}{\partial b_c} \left(\sum_{k=0}^p f_{t-k} \otimes c_{t-1-k} + i_t \otimes \tilde{c}_t \right) = \sum_{k=0}^p \frac{\partial}{\partial b_c} (f_{t-k}) \otimes c_{t-1-k} + \\ &+ \sum_{k=0}^p f_{t-k} \otimes \frac{\partial}{\partial b_c} (c_{t-1-k}) + \frac{\partial}{\partial b_c} (i_t) \otimes \tilde{c}_t + i_t \otimes \frac{\partial}{\partial b_c} (\tilde{c}_t) = \sum_{k=0}^p \sigma'_g (W_f x_t + U_f h_{t-1} + \\ &+ b_f) U_f \frac{\partial}{\partial b_c} (h_{t-1}) \otimes c_{t-1-k} + \sum_{k=0}^p f_{t-k} \otimes \frac{\partial}{\partial b_c} (c_{t-1-k}) + \sigma'_g (W_i x_t + U_i h_{t-1} + b_i) \times \\ &\times U_i \frac{\partial}{\partial b_c} (h_{t-1}) \otimes \tilde{c}_t + i_t \otimes \sigma'_c (W_c x_t + U_c h_{t-1} + b_c) U_c \frac{\partial}{\partial b_c} (h_{t-1}). \end{aligned}$$

Зауважимо, що для $t \leq 0$ всі наближення рівні нульовим векторам відповідних розмірів, тобто

$$\frac{\partial}{\partial b_c}(c_t), \frac{\partial}{\partial b_c}(h_t), t \leq 0,$$

є нульовими векторами. Аналогічно можемо визначити і похідні за змінними (W_c, U_c)

$$\frac{\partial \hat{y}_t}{\partial U_c} = \sigma'_g(o_t) U_o \frac{\partial}{\partial U_c} h_{t-1} \otimes \sigma_h(c_t) + \sigma_g(o_t) \otimes \sigma'_h(c_t) \frac{\partial}{\partial U_c}(c_t),$$

де

$$\begin{aligned} \frac{\partial}{\partial U_c}(c_t) &= \sum_{k=0}^p \sigma'_g (W_f x_t + U_f h_{t-1} + b_f) U_f \frac{\partial}{\partial U_c} (h_{t-1}) \otimes c_{t-1-k} + \sum_{k=0}^p f_{t-k} \otimes \\ &\otimes \frac{\partial}{\partial U_c} (c_{t-1-k}) + \sigma'_g (W_i x_t + U_i h_{t-1} + b_i) U_i \frac{\partial}{\partial U_c} (h_{t-1}) \otimes \tilde{c}_t + i_t \otimes \sigma'_c (W_c x_t + \\ &+ U_c h_{t-1} + b_c) \end{aligned}$$

$$+U_c h_{t-1} + b_c) U_c \frac{\partial}{\partial U_c} (h_{t-1}) + i_t \otimes \sigma'_c(W_c x_t + U_c h_{t-1} + b_c) h_{t-1},$$

та

$$\frac{\partial \hat{y}_t}{\partial W_c} = \sigma'_g(o_t) U_0 \frac{\partial}{\partial W_c} h_{t-1} \otimes \sigma_h(c_t) + \sigma_g(o_t) \otimes \sigma'_h(c_t) \frac{\partial}{\partial W_c} (c_t),$$

де

$$\begin{aligned} \frac{\partial}{\partial W_c} (c_t) &= \sum_{k=0}^p \sigma'_g(W_f x_t + U_f h_{t-1} + b_f) U_f \frac{\partial}{\partial W_c} (h_{t-1}) \otimes c_{t-1-k} + \\ &+ \sum_{k=0}^p f_{t-k} \otimes \frac{\partial}{\partial W_c} (c_{t-1-k}) + \sigma'_g(W_i x_t + U_i h_{t-1} + b_i) U_i \frac{\partial}{\partial W_c} (h_{t-1}) \otimes \tilde{c}_t + \\ &+ i_t \otimes \sigma'_c(W_c x_t + U_c h_{t-1} + b_c) x_t. \end{aligned}$$

Використовуючи аналогічний підхід, обчислимо частинні похідні:

$$\frac{\partial \hat{y}_t}{\partial b_0}, \frac{\partial \hat{y}_t}{\partial b_i}, \frac{\partial \hat{y}_t}{\partial b_f}, \frac{\partial \hat{y}_t}{\partial U_0}, \frac{\partial \hat{y}_t}{\partial U_i}, \frac{\partial \hat{y}_t}{\partial U_f}, \frac{\partial \hat{y}_t}{\partial W_0}, \frac{\partial \hat{y}_t}{\partial W_i}, \frac{\partial \hat{y}_t}{\partial W_f}.$$

На основі рекурентних формул, визначених для частинних похідних, аналогічно до класичного результату, неважко побудувати оновлення параметрів нейронної мережі на основі градієнтно-орієнтованих методів. Іншим підходом до оцінки параметрів моделі є використання генетичних алгоритмів, в яких оновлення не потребує наявності градієнта. Недоліком використання генетичних алгоритмів є висока ресурсозатратність, зумовлена потребою великої кількості обчислень цільової функції Q для різних значень параметра Par .

Додатковою проблемою, що виникає при дослідженні розширеної моделі (4), є оцінка невідомого параметра p , що вказує на довжину передісторії, яка використовується при аналізі. Для оцінки даного параметра у розширеній моделі буде використано оцінку на основі $AR(p)$ моделі [6], оскільки вихідний сигнал є функцією від того ж сигналу в попередні моменти часу, тобто має місце залежність

$$g(x_t) = f(x_{t-1}, x_{t-2}, \dots, x_{t-p}),$$

де g — відома функція (перетворення), яка становить інтерес при дослідженні. Так, у деяких прикладних задачах функція $g(x)$ визначається співвідношенням

$$g(x) = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0, \end{cases}$$

або

$$g(x) = \begin{cases} 1, & x > h, \\ 0, & x \leq h. \end{cases}$$

Таким чином, підсумовуючи все вищевказане, узагальнимо наступні кроки.

Крок 1. Оцінка гіперпараметра p на основі AR -моделі з відомою функцією g з мінімізацією залишків або інформаційного критерію.

Крок 2. Оцінка параметрів моделі (4) на основі градієнтно-орієнтованих методів або генетичного алгоритму.

Крок 3. Побудова прогнозу для $h_t = g(x_t)$ на основі цільової функції Q .

Висновок

У роботі розглянуто одне узагальнення LSTM-нейронних мереж з урахуванням довжини передісторії нейронної мережі. Основна відмінність від класичної LSTM-моделі або багатовимірного варіанта MV-LSTM-моделі полягає у наявності гіперпараметра p , який відображає глибину залежності вхідного $(x_t, \dots, x_{t+1-p})$ та вихідного сигналу $g(x_{t+1})$ або в більш загальному випадку $g(x_{t+h})$ для $h > 0$. Як вказано вище, часові ряди з явною періодичною залежністю погано оцінюються за допомогою класичної LSTM-нейронної мережі, проте дана проблема вирішується на основі узагальненої моделі (4).

У подальших роботах у напрямку дослідження буде здійснено розробку модуля для реалізації розширеної LSTM-моделі для градієнтно-орієнтованих методів. Обчислення градієнта частини параметрів нейронної мережі здійснено вище, а градієнти за іншими параметрами обчислюються за цим же принципом. Також в модулі планується розробка функцій для оцінки параметрів нейронної мережі на основі генетичних алгоритмів.

M. Kushnir, K. Tokarieva

ONE GENERALIZATION OF LSTM NEURON NETWORKS

Mykola Kushnir

Yuriy Fedkovych Chernivtsi National University, Ukraine,
myk.kushnir@chnu.edu.ua

Kateryna Tokarieva

Yuriy Fedkovych Chernivtsi National University, Ukraine,
tokarieva.chnu@gmail.com

The main goal of the work is the development of a generalized LSTM neural network taking into account the history of finite depth. Most of the works in this direction are devoted to multi-stream generalizations with the presence of correlation between streams (sub processes), that is, the generalization of the LSTM model in width is considered similarly to panel data in time series theory. In this work, the model is generalized in depth, which will allow taking into account the periodic components of time series and other deterministic components, that is, without preprocessing the data. As will be shown in the example, the development of this model allows generalizing the use of classical LSTM networks to dynamic systems with a clearly defined periodic component. In addition to the generalized model itself, the problem of calculating model parameters based on the backpropagation method is considered. For this, the partial derivatives of the resulting output signal with respect to unknown parameters are calculated and the recurrence of these derivatives over time is shown, similar to the classic LSTM model. An additional problem when creating an extended model is the estimation of the hyperparameter p , which indicates the depth of the background history of the process. In this regard, an algorithm for estimating the hyperparameter p with subsequent estimation of all model parameters was developed.

Keywords: neural networks, parameter estimation, backward propagation, time series.

ПОСИЛАНИЯ

1. Гихман И.И., Скороход А.В. Стохастические дифференциальные уравнения. К. : Наук. думка, 1968. 355 с.
2. Schilling R.L. Brownian motion: a guide to random processes and stochastic calculus. 3 rd. ed. De Gruyter, Textbook, 2021. 500 p.
3. Rüschendorf L. Stochastic processes and financial mathematics. Berlin : Springer, 2023. 313 p.
4. Mohammadi M., Rezakhah S., Modarresi N., Amindavar H. Continuous-time autoregressive models excited by semi-Lévy process for cyclostationary signal analysis. *Digital Signal Processing*. 2021. Vol. 118. P. 103–195. DOI: <https://doi.org/10.1016/j.dsp.2021.103195>
5. Costacurta J., Duncker L., Sheffer B., Williams A., Gillis W., Weinreb C., Markowitz J., Datta S., Linderman S. Distinguishing discrete and continuous behavioral variability using warped autoregressive HMMs. *bioRxiv*. June 28, 2022. 19 p. DOI: <https://doi.org/10.1101/2022.06.10.495690>
6. Brockwell P.J., Davis R.A. Introduction to time series and forecasting. 3 rd. ed. Bern : Springer, 2016. 439 p.
7. Hochreiter S., Schmidhuber J. Long short-term memory. *Neural Computation*. 1997. Vol. 9, N 8. P. 1735–1780. DOI:10.1162/neco.1997.9.8.1735
8. Monner D.D., Reggia J.A. A generalized LSTM-like training algorithm for second-order recurrent neural networks. *Neural Networks*. 2010. Vol. 25, N 1. P. 70–83. DOI: 10.1016 /j.neunet.2011.07.003
9. Rajagopalan S.S., Morency L.P., Baltrušaitis T., Goecke R. Extending long short-term memory for multi-view structured learning. In: B. Leibe, J. Matas, N. Sebe, M. Welling. *Computer Vision*. 2016. Vol. 9911. P. 338–352. DOI: https://doi.org/10.1007/978-3-319-46478-7_21
10. Zadeh A., Liang P.P., Mazumder N., Poria S., Cambria E., Morency L.-P. Memory fusion network for multi-view sequential learning. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2018. Vol. 32, N 1. P. 5634–5641. DOI: 10.1609/aaai.v32i1.12021
11. Кнопов П.С., Богданов А.В. Использование стохастической модели для прогнозирования длительных эпидемий. *Международный научно-технический журнал «Проблемы управления и информатики»*. 2021. № 3. С. 50–57.
12. Кнопов П.С., Касицька Е.И. Состоятельность и свойства больших уклонений эмпирических оценок в задаче стохастической оптимизации для однородного случайного поля при неоднородных и однородных наблюдениях. *Кибернетика и системный анализ*. 2021. № 1. С. 21–34.
13. Гарашенко Ф.Г., Кулян В.Р., Петрович В.Н., Юнькова Е.А. Моделирование динамики и диверсификация портфеля акций. *Международный научно-технический журнал «Проблемы управления и информатики»*. 2016. № 4. С. 124–135.
14. Liang B., Wang S., Huang Y., Liu Y., Ma L. F-LSTM: FPGA-Based heterogeneous computing framework for deploying lstm-based algorithms. *Electronics*. 2023. Vol. 12. P. 11–39. <https://doi.org/10.3390/electronics12051139>
15. Ismanto E., Vitriani. LSTM network hyperparameter optimization for stock price prediction using the optuna framework. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*. 2023. Vol. 9, N 1. P. 22–35. DOI: 10.26555/jiteki.v9i1.249440
16. Alshingiti Z., Alaqel R., Al-Muhtadi J., Haq Q., Saleem K., Faheem M. A deep learning-based phishing detection system using CNN, LSTM, and LSTM-CNN. *Electronics*. 2023. Vol. 12, N 1. P. 1–18. <https://doi.org/10.3390/electronics12010232>
17. Dou G., Zhao K., Guo M., Mou J. Memristor-based LSTM network for text classification. *Fractals*. 2023. Vol. 31, N 6. 12 p. <https://doi.org/10.1142/S0218348X23400406>
18. Tinnemore A., Montero L., Gordon-Salant S., Goupell M. The recognition of time-compressed speech as a function of age in listeners with cochlear implants or normal hearing. *Frontiers in Aging Neuroscience*. 2022. Vol. 14. 16 p. DOI: 10.3389/fnagi.2022.887581
19. Xu Hesheng, Hu Bin. Legal text recognition using LSTM-CRF deep learning model. *Computational Intelligence and Neuroscience*. 2022. 10 p. <https://doi.org/10.1155/2022/9933929>
20. Snoek C.G., Worring M., Smeulders A.W. Early versus late fusion in semantic video analysis. *Proceedings of the 13th Annual ACM International Conference on Multimedia*. 2005. P. 399–402. DOI : <https://doi.org/10.1145/1101149.1101236>

Отримано 23.05.2023