

ІВЧЕНКО ІРИНА

Національний університет «Одеська політехніка»

<https://orcid.org/0000-0002-1977-0342>e-mail: i.y.ivchenko@op.edu.ua

ГАЙДАЄНКО ОКСАНА

Національний університет кораблебудування імені адмірала Макарова, Україна

<https://orcid.org/0000-0002-6614-5443>e-mail: oksana.gaidaienko@nuos.edu.ua

КНИРІК НАТАЛІЯ

Національний університет кораблебудування імені адмірала Макарова, Україна

<https://orcid.org/0000-0002-9137-7625>e-mail: natalya.knyrik@nuos.edu.ua

МОРОЗОВА ГАННА

Національний університет кораблебудування імені адмірала Макарова, Україна

<https://orcid.org/0009-0005-9149-7378>e-mail: amorozova711004@gmail.com

ГРИБЕНЮК ДМИТРО

Одеса, Україна, engineering manager

e-mail: uragarage@gmail.com

ПРОГНОЗУВАННЯ ПОВЕДІНКИ ОБ'ЄКТІВ СОЦІАЛЬНИХ МЕРЕЖ

У статті запропоновано підхід до прогнозування популярності контенту в соціальних мережах під впливом зовнішніх факторів на основі рекурентній нейронній мережі. Розроблена модель дозволяє здійснювати прогнозування динаміки репостів у соціальних мережах, що може бути використано для оцінки зростання або падіння популярності контенту.

Ключові слова: економіко-математична модель, прогнозування, соціальна мережа, нейронна мережа.

IVCHENKO IRYNA

Odessa Polytechnic National University

HAIDAIENKO OKSANA

Admiral Makarov National Shipbuilding University, Ukraine

KNYRIK NATALIA

Admiral Makarov National Shipbuilding University, Ukraine

MOROZOVA HANNA

Admiral Makarov National Shipbuilding University, Ukraine

GRYBENIUK DMYTRO

Odessa, Ukraine, engineering manager

FORECASTING THE BEHAVIOR OF SOCIAL MEDIA OBJECTS

The article proposes an approach to formalising qualitatively heterogeneous processes of growth/decline in popularity of social media content under the influence of external factors within the framework of a forecasting model. The author's approach to the use of artificial neural networks in predicting the behaviour of social media objects is substantiated. It is substantiated that the dynamic predictive model will make it possible to refine the predictions on the studied indicators: reposts of content in social networks, video and music trends, and popularity of bloggers at the next modelling step. The main purpose of the model is to mathematically process and analyse stochastic information flows that are formed under dynamic conditions of uncertainty under the influence of external factors. The article uses as input data individual time series reflecting the number of reposts for a specific date for certain objects, such as music trends, bloggers or video content. The values of the number of reposts are random, which leads to uncertainty in the predictions that characterise the behaviour of the system. A system of constraints and predictors is proposed to describe the relationship between input and output. Dynamic control variables set the allocation of resources for optimal use at any given time, which ensures consistency between all the parameters required for analysis. Based on the data obtained, we analyse the status and behaviour of users in social media to develop forecasts of their future activity. In particular, predicting the number of reposts is important for popularising music trends, videos, and bloggers and helps to promote their businesses. The basic element of the neural network developed in this study is a dynamic neuron for calculating the predictive reposting rates of popular trends on a social network. The process of training the neuron takes into account the prediction error for the previous period. The reliability of the results obtained with the developed statistical finite difference model for predicting the graph weight index using large data samples is verified. The overall performance of the model has been tested on large samples of data on music trends in social networks. The practical value lies in the possibility of forecasting time series to predict user behaviour and suggestions for promising content (bloggers, music, videos) for the purpose of further investment, promotion in social networks, advertising and development of new projects.

Keywords: social network, reposts, content, neural network, neuron training, Big Data

Постановка проблеми у загальному вигляді

та її зв'язок із важливими науковими чи практичними завданнями

На сучасному етапі розвитку економіки проблема розробки наукового підходу до управління соціально-економічними процесами стає все більш актуальною. Це зумовлено необхідністю ефективного інвестування у різні види підприємницької діяльності для їх подальшого розвитку, що відкриває нові можливості для підвищення конкурентоспроможності бізнесу, просування товарів, залучення інвестицій та розвитку інновацій.

Питання автоматизації прийняття рішень на основі математичного аналізу, прогнозування, економіко-математичного моделювання та машинного навчання набувають особливої значущості. Важливим завданням є

вирішення питань збору великої кількості інформації, її аналізу та подальшого використання для моделювання складних соціально-економічних систем. Використання таких даних ускладнене стохастичним характером, а також тим, що вони зазвичай мають неструктурований вигляд (Big Data). Для роботи з Big Data необхідно застосовувати сучасні математичні методи та інформаційні технології: математичну статистику, моделювання, машинне навчання, штучний інтелект, проектування баз даних, а також сучасні мови програмування. Це пояснює зростаючий інтерес до Data Science, яка об'єднує математику, інформатику та системний аналіз.

Особливий інтерес для математичного моделювання соціально-економічних систем становлять інформаційні потоки в інтернет-спільнотах, соціальних мережах, Telegram-каналах та форумах. Вивчення цих потоків дозволяє виявляти закономірності в масивах неструктурованих даних і використовувати їх для складання прогнозів. Використання даних із соціальних мереж дозволяє вирішувати широке коло бізнес-завдань: реклама, популяризація брендів, пошук збуту товарів та послуг та ін. [1, 2] Автоматизація прийняття рішень на основі таких даних сприяє підвищенню маржинальності бізнесу та створює нові можливості для розвитку.

Аналіз досліджень та публікацій

Прогнозування відіграє ключову роль у підготовці та реалізації управлінських рішень. Воно спрямоване на виявлення закономірностей розвитку економічних та соціальних процесів, що дозволяє визначити основні тенденції змін важливих показників. Це дає змогу керівникам та дослідникам ухвалювати обґрунтовані рішення, спрямовані на досягнення стратегічних цілей, оптимізацію ресурсів і зниження ризиків у динамічних умовах невизначеності та впливу зовнішніх факторів. В роботах Балабанова О. С., Калініної І. О., Снитюка В. Є., Cai Q., Zhang D., Zheng та ін. наведено класичні принципи і методи аналізу та прогнозування [3, 4]. До них, в першу чергу, відносять статистичні моделі кластерного та регресійного аналізу. Однак, головний недолік таких моделей – обмеженість лінійними моделями, що знижує їхню ефективність у роботі з даними типу Big Data.

Сучасна наука пропонує нові підходи до вирішення цих викликів, зокрема об'єднання методів прогнозування з нечіткими оцінками експертів або генетичними алгоритмами.

Одним із найкращих інструментів для прогнозування даних в умовах невизначеності є нейронні мережі, що дозволяють ефективно працювати з нелінійними системами і великими обсягами даних [5]. Нейронні мережі, такі як NARX, використовуються для прогнозування складних явищ, наприклад, мережевого трафіку, і показують високу точність [5, 6]. Поєднання декількох нейронних мереж дозволяє покращити якість прогнозів, хоча такі методи можуть бути менш гнучкими.

Актуальність використання нейромережевих технологій для прогнозування в соціально-економічних процесах зростає, оскільки вони дозволяють працювати з динамічною природою соціальних об'єктів і їхньої трансформації, швидкими змінами в інформаційних технологіях і великими обсягами неструктурованих даних. Прогнозні нейронні мережі мають низку переваг порівняно з традиційними методами прогнозування. Вони здатні моделювати нелінійні зв'язки в соціально-економічних системах за умов невизначеності та неповних даних. Нейронні мережі також ефективні для роботи з масивами даних типу Big Data, що характеризуються довільною структурою, а їх здатність до навчання є важливою перевагою над класичними підходами.

Формулювання цілей статті

Метою роботи є обґрунтування доцільності застосування нейронних мереж для аналізу поведінки динамічних об'єктів соціальних мереж та формуванням передбачень репостів користувачів. Основним завданням математичної моделі, що розробляється, є обробка й аналіз багатовимірних стохастичних даних та прогнозування поведінки соціальних об'єктів, що досліджуються, для підтримки управлінських рішень в умовах невизначеності та обмеженої інформації.

Виклад основного матеріалу

Об'єктом даного дослідження є соціальні мережі, що спеціалізуються на створенні та перегляді коротких відео (наприклад, Facebook, Twitter, Instagram Reels, TikTok, YouTube Shorts). Соціальна мережа — це онлайн-платформа або сервіс, що дозволяє користувачам створювати профілі, взаємодіяти один з одним, ділитися контентом (текстом, зображеннями, відео) та обмінюватися інформацією. Соціальні мережі забезпечують можливість створення та підтримки соціальних зв'язків між людьми, організаціями або спільнотами, за допомогою функцій, таких як коментування, лайки, реакції, репости, чати, групи та інші форми взаємодії. Вони можуть використовуватися для різноманітних цілей, включаючи спілкування, просування брендів, обмін новинами, обговорення тем інтересу та побудову професійних мереж.

Дослідження соціальних мереж передбачає збір первинних даних (кількість репостів контенту) та аналіз отриманих вибірок з подальшою побудовою різноманітних моделей, таких як оптимізаційні, імітаційні, прогнозні тощо. В процесі аналізу перелічених соціальних мереж було встановлено, що вхідні дані в кожен момент часу мають випадковий характер. Поведінка вхідних показників соціальних мереж характеризується невизначеністю та цілеспрямованими змінами стану в часі. Основним чинником невизначеності, який впливає на точність результатів роботи моделі, є випадковість вхідних даних. Тому обов'язковою умовою забезпечення достовірності результатів моделі, що розробляється, є використання великих вибірок даних, які належать до категорії BigData.

На основі отриманих даних здійснюється аналітика стану та поведінки користувачів у соціальних мережах з метою розробки прогнозів щодо їхньої подальшої активності. Зокрема, передбачення кількості репостів має важливе значення для популяризації музичних трендів, відео та блогерів та сприяє просуванню

їхнього бізнесу.

У системному аналізі поширеним є метод формального представлення системи у вигляді "чорного ящика", що має "входи" та "виходи" [7]. І хоча метод "чорного ящика" не розкриває внутрішнього устрою соціальної мережі, але дозволяє аналізувати зміни в системі як залежність вихідних показників від вхідних.

Стан системи на виході залежить не лише від значень вхідних показників у визначений момент часу, але й від значень показників у попередні моменти часу. Математично стан такої системи управління в будь-який момент часу t однозначно визначається залежністю стану виходів $Y_j(t)$ від входів $X_i(t)$ з урахуванням керуючих впливів j :

$$Y(t)=F(X(t),a_j), \quad (1)$$

де

$X(t)=\{x_1(t), \dots, x_n(t)\}$ – вхідні (фактори, які описують систему;

j – предиктори моделі (фактори, які впливають на результат);

F – вигляд функціональної залежності;

$Y(t)=\{y_1(t), \dots, y_m(t)\}$ – вихідні показники.

Вихідні показники не лише відображають зміни в стані системи та її вплив на навколишнє середовище, але й надають можливість прогнозувати поведінку об'єкта дослідження. Прогнозні вихідні показники забезпечать отримання додаткової інформації про майбутню поведінку об'єкта на основі збору структурованих даних про минулі події. Дослідник (зовнішній споживач або власник моделі), маючи в своєму розпорядженні прогнозні дані, зможе ухвалювати ефективні управлінські рішення (наприклад, обирати найперспективніших музикантів, блогерів чи відеоролики) і залишатися на кілька кроків попереду конкурентів.

Одним із ключових показників популярності блогерів, брендів і контент-креаторів у соціальних мережах є метрика, що відображає рівень зацікавленості користувачів — репости. Кількість репостів визначається як сумарна кількість користувачів та сторінок, які здійснили поширення контенту серед своїх друзів і підписників. Збільшення кількості репостів свідчить про ширше охоплення контенту, що, у свою чергу, інтерпретується як підвищення соціальної популярності матеріалів у мережі.

У дослідженні як вхідні дані використовуються індивідуальні часові ряди, що відображають кількість репостів за конкретну дату для певних об'єктів, таких як музичні тренди, блогери або відеоконтент. Величини кількості репостів мають випадковий характер, що призводить до виникнення невизначеності в прогнозах. Період моделювання є довільним і не обмежений певним інтервалом, а як крок моделювання обрано добу. Дані про процес можуть збиратися нерегулярно, і в разі відсутності даних за певні періоди в моделі передбачено механізм «відображення» процесів, що дозволяє скорегувати кількість пропущених днів.

Для опису стану об'єкта дослідження на кожен момент часу (добу) використовуються наступні вхідні показники:

– номенклатура об'єктів (моделюється окремо для кожного музичного тренду);

– кількість репостів;

– кількість пропущених днів.

Модель є динамічною та має імовірнісний характер, що дозволяє враховувати випадкові коливання в процесі та точніше прогнозувати поведінку об'єктів у майбутньому. Для реалізації керуючого предиктора на модель на нульовому етапі обчислень задається множина керуючих змінних:

– рівень успішності виконавця (залежить від рівня популярності музичного тренду на початок моделювання);

– можлива кількість репостів на початок моделювання (передбачення, стартова величина);

– виправлення на помилку передбачення.

Припустимо, що керуючі змінні можуть набувати значень у діапазоні від 0 до 1. Це дозволить під час прогнозування, окрім описаних параметрів, точно відображати початкові індивідуальні характеристики виконавців, а також задавати зміни у стані зовнішнього середовища, що безпосередньо впливають на моделювання системи та, відповідно, на її вихідні показники.

У моделі пропонується на кожній ітерації розраховувати такі змінні:

– прогнозована кількість репостів на поточний день;

– похибка передбачення денного зростання кількості репостів;

– прогнозована кількість репостів на наступний крок моделювання;

– якість передбачення.

Врахування динаміки змін показників є критично важливим, оскільки вони змінюються в процесі моделювання через випадковий характер вхідних даних і вплив керуючих змінних на модель.

Для візуалізації результатів прогнозування на основі фактичних і прогнозованих даних про кількість репостів будуть побудовані відповідні графіки, що дозволять оцінити точність моделі та її ефективність у прогнозуванні.

Один із найбільш універсальних підходів до вирішення задач прогнозування полягає у застосуванні нейронних мереж [8]. Основною характеристикою нейронних мереж є їхня здатність до адаптивного навчання. Мережа здатна виявляти складні залежності між вхідними та вихідними параметрами, базуючись на попередньому досвіді, що дозволяє зменшувати похибки в процесі навчання.

У рамках даного дослідження центральним елементом є динамічний нейрон, який використовується

для обчислення прогнозних значень показників репостів. Динамічний нейрон, завдяки своїй архітектурі, яка враховує часові залежності, забезпечує високу точність прогнозування, що є критично важливим для аналізу поведінки користувачів у соціальних мережах. Це дозволяє оптимізувати управлінські рішення та підвищити ефективність стратегій просування контенту. Сукупність нейронів створює нейронну мережу, яку можна представити у класичному розумінні системи, тобто у вигляді «чорної скриньки», яка має входи та виходи.

В рамках даного дослідження була побудована система з рекурентною нейронною мережею (мережа зі зворотними зв'язками). Входи $X = \{R_i(n); dS_i(n)\}$ з'єднують нейрони з ядром за допомогою синаптичних з'єднань. Кожен синапс характеризується вагою $GW_i(n)$, яка визначає ступінь впливу відповідного входу на вихід нейрона. Ця вага є критично важливим параметром, оскільки вона регулює, наскільки сильно сигнал від входу вплине на активацію нейрона. Процес навчання нейронної мережі включає корекцію цих ваг у відповідь на різницю між очікуваним і фактичним виходом, що дозволяє мережі адаптуватися до даних та покращувати точність прогнозів. Ядро нейрона (S_i) обчислює вихідні змінні $Y = \{FR_i(n)\}$ відповідно до визначених алгоритмів та правил. Воно виконує процес активації, в якому вхідні дані зважуються за синаптичними вагами $GW_i(n)$, та підсумовуються для формування загального сигналу. Після цього, застосовуючи функцію активації, ядро генерує вихідний сигнал, який передається далі в мережу.

Процес навчання нейрона S_i (наприклад, для передбачення репостів контенту заданої соціальної мережі) використовує алгоритм зворотного розповсюдження помилки та враховує поправки на вагові коефіцієнти для отримання більш точного прогнозу репостів. На першому етапі обчислюється вихідний сигнал нейрона, який потім порівнюється з цільовим значенням, що дозволяє визначити помилку. Далі, з використанням обчисленої помилки, алгоритм коригує ваги $GW_i(n)$ через розподіл помилки з виходу до вхідних нейронів, поступово зменшуючи цю помилку на кожній ітерації навчання. В результаті, нейрон S_i адаптується до трендів у даних, забезпечуючи більш точні прогнози щодо кількості репостів.

Значення виходу нейромережі розраховується для кожного нейрона:

$$Fr_i(n+1) = Ri(n) + DFR_i(n), \quad (2)$$

де $Fr_i(n+1)$ – прогноз репостів на наступний день (ForecastingRepost); $Ri(n)$ – кількість репостів контенту (Repost); $DRFi(n)$ – передбачення кількості репостів на сьогодні (DailyRiseForecasting);

Враховуючи минуле, тобто дані попередньої ітерації, вага графа може прогнозувати (обчислювати) показник зросту/падіння репостів на наступну добу (рис. 1).



Рис.1. Процес навчання нейромережі (лютий – березень 2024 р.)
(Джерело - авторська розробка)

На рис. 1 продемонстровано процес навчання нейромережі по даним про кількість репостів за період лютий – березень 2024 р. Лінійка «репости» демонструє фактичну динаміку щоденних змін кількості репостів. Лінійка «прогноз» ілюструє передбачену траєкторію майбутніх репостів. Прогнозування базується на аналізі попередніх даних та корекції помилок попередніх передбачень. На кожному етапі після розрахунку вихідної змінної застосовуються поправки на вагові коефіцієнти, що впливають на наступні ітерації. Таким чином, з кожною новою ітерацією вхідні параметри (кількість репостів) коригуються з урахуванням поточних значень та передбачених виходів, що призводить до поступового покращення точності прогнозу.

Нейронна мережа має здатність до постійного вдосконалення, що дозволяє навчити її створювати високоточні прогнози шляхом інтеграції кількох типів контенту. Можливе об'єднання декількох нейронів у мережу для прогнозування змін у взаємопов'язаних контентах (наприклад, від одного виконавця) або для прогнозування популярності серед представників одного бренду. Такий підхід зводиться до задачі оптимізації прогнозовної моделі з побудовою цільової функції. Цільова функція забезпечує можливість ваговим коефіцієнтам $GW_i(n)$ максимально точно відтворювати вихідні показники (кількість репостів), наближаючи їх до фактичних даних. Процес формування та вибору оптимальних траєкторій управління є предметом подальших досліджень авторів в контексті задач оптимізації на основі розробленої динамічної прогнозовної нейронної моделі, ідеології статистичних експериментів та прийняття рішень в умовах невизначеності зовнішніх впливів.

Висновки з даного дослідження і перспективи подальших розвідок у даному напрямі

У даному дослідженні продемонстровано особливості побудови прогнозової моделі для показників репосту контенту у соціальних мережах. Запропоновано побудувати динамічну рекурентну нейронну мережу. Для забезпечення керуючого впливу введені предиктори, що враховують поведінку зовнішнього середовища. Обмеження моделі базуються на прийнятих припущеннях. Ключовим елементом цієї нейронної мережі є динамічний нейрон, який використовується для обчислення прогнозних показників кількості репостів популярних трендів у соціальних мережах. Навчання нейрона здійснюється з урахуванням помилки прогнозу за попередній період, що дозволяє моделі в режимі реального часу адаптуватися та передбачати, яка музика стане популярною у майбутньому, а які тренди втратять актуальність.

Література

9. Балабанов О. С. Аналітика великих даних: принципи, напрямки і задачі (огляд). Проблеми програмування. 2019. № 2. С. 47–68. DOI: <https://doi.org/10.15407/pp2019.02.047>.
10. Вэй Тан, Брайан Блейк, Иман Салех. Аналитика Больших Данных и социальные сети // Открытые системы. СУБД. 2013. № 8. С. 37–41. URL: <http://www.osp.ru/os/2013/08/13037856>
11. Снитюк В. Є. Прогнозування. Моделі. Методи. Алгоритми : навчальний посібник. Київ : «Маклаут», 2008. 364 с.
12. Cai Q., Zhang D., Zheng W., Leung S. C.H. A new fuzzy time series forecasting model combined with ant colony optimization and auto-regression. Knowledge-Based Systems, V. 74, (2015) pp. 61–68
13. Вітлінський В.В. Штучний інтелект у системі прийняття управлінських рішень. Нейро-нечіткі технології моделювання в економіці. 2012. № 1. С. 97–118.
14. Сериков Т., Жетписбаева А., Мирзакулова С., Жетписбаев К., Ибраева З., Соболева Л., Толегенова А. и Жумажанов Б. (2021). Применение нейронной сети NARX для прогнозирования одномерного временного ряда. Восточно-Европейский журнал корпоративных технологий, 5 (4 (113), 12–19. <https://doi.org/10.15587/1729-4061.2021.242442>
15. Соколовська З.М., Андриєнко В. М., Івченко І. Ю., Клепікова О. А., Яценко Н. В. "Математичне та комп'ютерне моделювання економічних процесів: [Монографія]", Одеса: "Астропринт", 2016. 273 с.
16. Дослідження нейромережевих методів у задачах прогнозування [Електронний ресурс]. – Режим доступу: <https://lib.chmnu.edu.ua/pdf/naukpraci/computer/2009/106-93-17.pdf>

References

1. Balabanov O. S. (2019) Analitika velykykh danykh: pryntsyru, napriamky i zadachi. Problemy prohramuвання, no. 2, pp. 47–68. DOI: <https://doi.org/10.15407/pp2019.02.047>.
2. Вэй Тан, Брайан Блейк, Иман Салех. Аналитика Больших Данных и социальные сети // Открытые системы. СУБД. 2013. № 8. С. 37–41. URL: <http://www.osp.ru/os/2013/08/13037856>
3. Snytiuk V. Ye. (2008) Prohnozuvannia. Modeli. Metody. Alhorytmy [Forecasting. Models. Methods. Algorithms]. Kyiv: «Maklout», 364 p.
4. Cai Q., Zhang D., Zheng W., Leung S. C. H. (2015). A new fuzzy time series forecasting model combined with ant colony optimization and auto-regression. Knowledge-Based Systems, vol. 74, pp. 61–68. DOI: <http://doi.org/10.1016/j.knosys.2014.11.003>.
5. Artificial intelligence in the system of managerial decision-making. Neuro-fuzzy modelling technologies in economics. 2012. № 1. С. 97–118.
6. Serikov T., Zhetopisbaeva A., Mirzakulova S., Zhetopisbaev K., Ibraeva Z., Soboleva L., Tolegenova A. and Zhumazhanov B. (2021). Application of the NARX neural network for univariate time series forecasting. East European Journal of Corporate Technology, 5 (4 (113), 12–19. <https://doi.org/10.15587/1729-4061.2021.242442>
7. Sokolovska Z.M., Andrienko V.M., Ivchenko I.Y., Klepikova O.A., Yatsenko N.V. 'Mathematical and Computer Modelling of Economic Processes: [Monograph]', Odesa: 'Astroprint', 2016. 273 c.
8. Research of neural network methods in forecasting tasks [Electronic resource]: <https://lib.chmnu.edu.ua/pdf/naukpraci/computer/2009/106-93-17.pdf>.