

## ПРОГНОЗУВАННЯ ПОПУЛЯРНОСТІ КОНТЕНТУ В СОЦІАЛЬНИХ МЕРЕЖАХ НА ОСНОВІ ОБРОБКИ ВЕЛИКИХ ДАНИХ

Івченко І.Ю., к.е.н., доц; Михелев І.Л., к.т.н., доц.

<sup>1</sup> Національний університет «Одеська політехніка»; <sup>2</sup> Національний університет кораблебудування імені адмірала Макарова

<sup>1</sup> Україна, Одеса; <sup>2</sup> Україна, Миколаїв

<sup>1</sup> i.y.ivchenko@op.edu.ua , ORCID 0000-0002-1977-0342; <sup>2</sup> igor.michelev@nuos.edu.ua , ORCID 0000-0001-9579-6547

**Анотація.** У статті розглядається задача прогнозування популярності контенту в соціальних мережах за допомогою обробки великих даних. Особлива увага приділяється проблемам роботи з неповними даними та застосуванню нейронних мереж для прогнозування поведінки користувачів. Запропонована прогнозна нейромережева модель поведінки об'єктів на базі часових рядів соціальних мереж. Результати дослідження демонструють практичну цінність моделі для оптимізації стратегій просування та оцінки впливу зовнішніх чинників на популярність контенту.

**Ключові слова:** прогнозування; соціальна мережа; нейронна мережа; часовий ряд.

**Вступ.** Бізнес має бути в курсі соціальних переваг суспільства щодо його потреб і бажань. Ці переваги залежать від цінностей, переконань і звичаїв населення. Компанії часто намагаються впливати на соціальні цінності за допомогою маркетингових, рекламних та цільових стратегій. Тому перед бізнес-аналітиками постає актуальне завдання прогнозування популярності контенту на основі аналізу великих обсягів даних. Джерелами таких даних є соціальні мережі, а також різноманітні платформи для обміну контентом, аналітичні системи веб-трафіку, онлайн-опитування та інші канали взаємодії з аудиторією, що забезпечують доступ до поведінкової та демографічної інформації користувачів.

Прогнозування поведінки об'єктів соціальних мереж має широкий спектр застосувань, що охоплює як комерційні, так і некомерційні цілі. Це сприяє підвищенню ефективності комунікацій компаній і брендів з цільовою аудиторією, забезпеченню актуальності контенту інфлюенсерів, дослідженню суспільних настроїв соціальними та культурними організаціями, а також оптимізації взаємодії користувачів з платформами.

Для якісного аналізу великих обсягів даних, проведення перспективних оцінок, прогнозування та вибору найкращих варіантів розвитку в науковій літературі запропоновано багато математичних інструментів моделювання. Математична теорія прогнозування була закладена у роботах таких вчених як Н. Вінер, А. Колмогоров, А. Ширяєв та інших. Їхні ідеї стали фундаментом для побудови моделей, що ґрунтуються на екстраполяції стаціонарних процесів. Але традиційні моделі прогнозування часто не враховують динамічні зміни, що відбуваються в соціальних мережах та контенті, оскільки популярність останнього підлягає впливу зовнішніх факторів, таких як суспільні події і зміни в поведінці користувачів. Так, використання тільки текстової інформації може не відобразити реальні тенденції популярності, обмежена точність даних може бути спричинена ігноруванням соціальних динамік і реальних подій. Тому інтеграція різноманітних джерел даних залишається складним завданням. У сучасній науці спостерігаються зміни, які обмежують застосування традиційних методів екстраполяції стаціонарних процесів, адже вони недостатньо адаптовані до якісних

змін (Gayo-Avello, 2012). Нестабільність інформації та соціальних впливів робить подібні моделі менш ефективними.

Дослідження вказують на те, що для точного прогнозування популярності контенту важливо враховувати різні типи даних, такі як текстові публікації, зображення та відео, джерелом яких є різноманітні соціальні мережі, телеграм-канали, інтернет-спільноти, чати та форуми. До того ж, соціальні мережі є складними системами, де людська поведінка може бути важко передбачуваною. Наприклад, популярність контенту може значно залежати від випадкових факторів, таких як дії впливових користувачів або несподівані соціальні події. Це ускладнює прогнозування, оскільки сучасні методи часто не можуть врахувати вплив окремих репостів або різких змін у настроях користувачів.

Для підвищення точності прогнозів необхідно розвивати адаптивніші методи, здатні враховувати динамічні зміни та інтегрувати різні джерела даних. У дослідженні соціально-економічних процесів, попри значний розвиток адаптивного прогнозування, досі не вирішені питання адаптації моделей для перспективного аналізу соціальних мереж, що доводить актуальність даного дослідження.

Розробка математичних моделей з використанням нейромереж є новим інструментом для вивчення таких соціально-економічних явищ як соціальні мережі та прогнозування популярності контенту на основі обробки великих даних.

**Метою роботи** є вибір та обґрунтування використання моделі, здатної працювати з великими обсягами даних соціальних мереж, аналізувати їх для виявлення закономірностей та передбачати зміни в популярності контенту. Особлива увага приділяється викликам, пов'язаним із роботою з неповними або зашумленими даними, а також обробці та аналізу багатомірних стохастичних даних для побудови прогнозів поведінки соціальних об'єктів у динаміці. Це сприятиме прийняттю управлінських рішень в умовах невизначеності та неповноти інформації, створюючи можливості для соціальних досліджень, вибору оптимальних стратегій просування, розміщення реклами, підвищення ефективності комунікацій з аудиторією та оцінки впливу зовнішніх чинників на популярність контенту, що допоможе забезпечити довгострокову конкурентоспроможність та інноваційний розвиток компаній.

**Основна частина.** Соціальні мережі є значущим джерелом вхідних даних для моделювання, оскільки вони дозволяють відслідковувати динаміку поширення контенту серед великої кількості користувачів.

Об'єктом дослідження є часові ряди, що характеризують репости контенту в соціальних мережах.

Репости контенту є показником соціальної активності користувачів, що відображає динаміку поширення інформації та рівень її популярності серед аудиторії. Ці вхідні дані мають стохастичний характер, належать до типу BigData, формуються за невизначеними правилами і містять значну кількість шуму. Дані про популярність контенту в соціальних мережах можуть не збиратися щодня, і при відсутності достовірних даних на певний період у моделі необхідно передбачити механізм корекції для аналізованих процесів. Це означає, що модель повинна враховувати пропущені дні спостереження, адаптуючи прогноз таким чином, щоб забезпечити коректність і точність результатів, незважаючи на пропуски у вибірці

Використання таких "сирих даних" створює складнощі для традиційних методів аналізу, що вимагає застосування спеціалізованих інструментів, таких як нейронні мережі та методи машинного навчання для обробки та виявлення прихованих закономірностей.

Для аналізу взаємозв'язків між репостами контенту та іншими факторами, що впливають на його популярність, а також для прогнозування динаміки популярності, запропоновано використати рекурентну нейронну мережу (мережа зі зворотними зв'язками). Базовим елементом моделі є динамічний нейрон для обчислення прогнозних показників репостів (рис. 1).

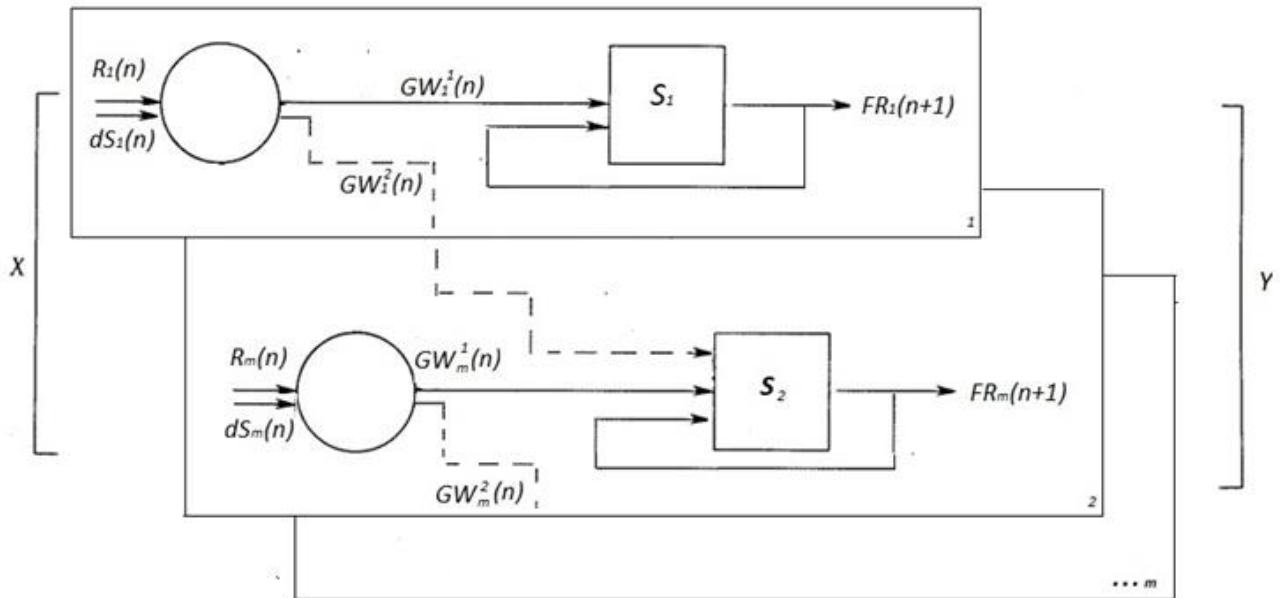


Рисунок 1 – Схема нейронної мережі

Вхідні змінні  $X = \{R_i(n); dS_i(n)\}$  пов'язують нейрони з ядром через синапси, які характеризуються вагою  $GW_i(n)$ . Вага показує ступень впливу кожного входу на вихід нейрона. Таким чином,  $GW_i(n)$  репрезентує надію нейрона на якість передбачення.

В процесі навчання нейронної мережі вагові коефіцієнти  $GW_i(n)$  постійно уточнюються. Ядро нейрона  $S_i$  розраховує значення вихідних змінних  $Y = \{FR_i(n)\}$  та видає вихідний сигнал. Навчання нейрона  $S_1$ , наприклад, для передбачення репостів контенту як показника соціальної активності користувачів, здійснюється за допомогою алгоритму зворотного розповсюдження помилки, що передбачає налаштування вагових коефіцієнтів для підвищення точності прогнозів.

Кожна нова ітерація використовує змінені вхідні параметри (прогноз репостів  $FR_i(n)$ ), які коригуються відповідно до поточних вхідних значень (кількість репостів  $R_i(n-1)$ ) та передбачених виходів (передбачення кількості репостів  $DRF_i(n-1)$ ). Завдяки процесу навчання, якість прогнозу швидко покращується, і нейронна мережа з кожною новою ітерацією стає здатною видавати все більш точні прогнози.

**Висновки.** У дослідженні представлено особливості побудови динамічної рекурентної нейронної мережі, призначеної для прогнозування часових рядів, що характеризують поведінку складних систем, до яких відносяться соціальні мережі. Базовим елементом нейронної мережі, запропонованої в цьому дослідженні, є динамічний нейрон, що виконує обчислення прогнозних показників репостів контенту в соціальній мережі. Нейронна мережа здатна постійно вдосконалюватися, що дозволяє в майбутніх дослідженнях забезпечити високу точність прогнозів шляхом інтеграції даних з різних типів контенту. Практична цінність розробленої моделі полягає в її універсальності та можливості застосування в будь-якому середовищі, яке потребує перспективного аналізу та прогнозування даних.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Wiener, N., (1965). *Cybernetics or control and communication in the animal and the machine*. M.I.T. Press Cambridge, Massachusetts P. 223
- [2] Gayo-Avello, D. (2012). *No, You Cannot Predict Elections with Twitter*. *Internet Computing, IEEE*. 16. 91-94. 10.1109/MIC.137.
- [3] Балабанов, О. С. (2019). *Аналітика великих даних: принципи, напрямки і задачі (огляд)*. *Проблеми програмування*. № 2. С. 47-68. <http://jnas.nbuv.gov.ua/article/UJRN-0001000726>, <https://doi.org/10.15407/pp2019.02.047>

Ina Ivchenko, Ihor Mykheliev.

### **Prediction of content popularity in social networks based on big data processing**

**Abstract.** *The article addresses the task of forecasting content popularity on social media through the processing of big data. Special attention is given to the challenges of working with incomplete data and the application of neural networks for predicting user behavior. A predictive neural network model based on time series from social networks is proposed. The research results demonstrate the practical value of the model for optimizing promotion strategies and assessing the impact of external factors on content popularity.*

**Keywords:** forecasting; social network; neural network; time series.