

АНАЛІЗ НАСТРОЇВ КОРИСТУВАЧІВ TWITTER НА ОСНОВІ ТЕКСТОВИХ ДАНИХ З ВИКОРИСТАННЯМ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ

Гнатушенко В.В.¹, Каштан В.Ю.², Овчаренко М.А.³, Іванько А.М.⁴

¹Національний технічний університет «Дніпровська політехніка»,
д.т.н., проф., Україна

²Національний технічний університет «Дніпровська політехніка»,
к.т.н., доц., Україна

³Національний технічний університет ³«Дніпровська політехніка», Україна

⁴Національний технічний університет ⁴«Дніпровська політехніка», Україна

Анотація. Розглянуто застосування згорткових нейронних мереж (CNN) для аналізу настроїв користувачів на основі текстових даних соціальної мережі Twitter. Актуальність роботи зумовлена зростаючим обсягом неструктурованих текстових даних у соціальних мережах та необхідністю їх ефективної обробки для розуміння громадської думки. Запропонована модель CNN включає представлення текстових даних у векторному просторі, згорткові та пулінгові шари для вилучення релевантних ознак, а також повністю зв'язані шари для класифікації настроїв на позитивні та негативні. Для запобігання перенавантаженню застосовано шари випадкового виключення нейронів (Dropout). Експериментальна оцінка моделі проводилася на спеціально підготовленому наборі даних із Twitter. Результати тестування демонструють перспективність використання CNN для автоматизованого аналізу настроїв користувачів соціальних мереж.

Ключові слова: настрої користувачів, згорткова мережа, неструктуровані дані.

Актуальність дослідження, представленого в даній роботі, зумовлена необхідністю обробки неструктурованої текстової інформації, сформованої в результаті користувацької активності в соціальних мережах, зокрема на платформі Twitter. У цьому контексті особливого значення набуває задача аналізу настроїв користувачів за текстовими повідомленнями, яка дозволяє отримати цінну інформацію про поведінкові тенденції користувачів та їхню реакцію на актуальні події. Велика кількість таких даних ускладнює застосування традиційних методів аналізу та потребує переходу до інтелектуальних методів автоматизованої обробки, таких як глибинне навчання і обробка природної мови [1]. Використання згорткових нейронних мереж (CNN) у задачах аналізу тексту відкриває нові можливості для

високоточної класифікації емоційних станів, що є важливим як для бізнес-аналітики, так і для соціально-гуманітарних досліджень. Однією з переваг CNN є здатність обробляти багатовимірні дані, що робить ці мережі ефективними для роботи з неструктурованими даними.

В роботі використано CNN модель для аналізу настроїв користувачів за текстовими наборами даних. Для реалізації цього завдання в роботі використано текстові набори даних із соціальної мережі Twitter [2]. Модель CNN для класифікації настроїв складається з кількох шарів. Початковий шар векторних представлень перетворює ціле число з словника в векторне представлення, з вхідною розмірністю 5000 (розмір словника) і виходом 50-вимірних векторів. Наступним шаром у моделі застосовується шар випадкового виключення нейронів (Dropout), який є методом регуляризації, спрямованим на запобігання перенавчанню. Під час навчання цей шар випадковим чином встановлює значення 20% вхідних нейронів на нуль. Згортковий шар (Conv1D) використовує 250 фільтрів із розміром ядра 3 для обробки послідовностей векторних представлень, застосовуючи функцію активації ReLU (Rectified Linear Unit). Операція глобального макспулінгу (GlobalMaxPooling1D) вздовж часової осі зменшує розмірність вихідних ознакових карт згорткового шару, обираючи максимальне значення для кожної карти. Наступним компонентом є повністю зв'язаний (Dense) шар, що складається з 250 нейронів з функцією активації ReLU, забезпечуючи повне з'єднання між усіма елементами попереднього шару та нейронами поточного. Для подальшої регуляризації додано шар Dropout з коефіцієнтом 0.2 під час тренування. Вихідним є щільний (Dense) шар із трьома нейронами та функцією активації Softmax для здійснення багатокласової класифікації вхідних даних, що відповідають категоріям настроїв: позитивний та негативний. Для оптимізації процесу навчання обрано оптимізатор Adam [3], що забезпечує стабільне навчання навіть при великій кількості параметрів і складних структурах даних.

У ході проведення експериментальних досліджень порівняно дві моделі згорткових нейронних мереж для класифікації настроїв: стандартну модель

CNN та модель з оптимізатором Adam. Для оцінки ефективності класифікації були використані основні метрики, зокрема точність, повнота та F1-міра [4]. Стандартна модель CNN показала результати з високими значеннями точності (0.84) та повноти (0.85) для позитивного настрою користувача, що свідчить про її здатність правильно класифікувати позитивні настрої. Проте її F1-міра позитивного настрою користувача (0.83) свідчить про певну дисбалансованість між точністю і повнотою. Використання оптимізатора Adam в CNN моделі підвищило точність до 0.85 і повноту до 0.84, що підтверджує покращену здатність моделі правильно класифікувати позитивні настрої.

Висновки. У роботі використано модель згорткової нейронної мережі для аналізу настроїв користувачів на основі текстових даних із Twitter. Проведено перетворення слів на числові вектори, що дозволило передати текстову інформацію у зручний формат для навчання. Модель була оцінена за допомогою метрик: точність, повнота та F1-міра, що продемонструвало її високу ефективність у класифікації настроїв. Використання оптимізатора Adam дозволило досягти більш стабільного навчання та кращої збіжності моделі, навіть при великій кількості параметрів. Це відкриває нові можливості для практичного застосування таких моделей у реальних завданнях, таких як моніторинг суспільних настроїв у соціальних мережах, маркетингові дослідження та політичний аналіз.

ЛІТЕРАТУРА

- 1.Salloum, S.A., et al. Using text mining techniques for extracting information from research articles, in Intelligent natural language processing: Trends and Applications. – 2018, Springer. P. 373-397
- 2.Coronavirus tweets NLP [Електронний ресурс]: <https://www.kaggle.com/datasets/datatattle/covid-19-nlp-text-classification>.
- 3.Saxena P., Sharma S. Systematic Literature Review for Sentiment Analysis Using Big Data social media Streams. In Proceedings of the 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 14–16 December 2022.
- 4.Tiwari, D., Nagpal, B., Bhati, B.S. et al. A systematic review of social network sentiment analysis with comparative study of ensemble-based techniques. Artif Intell Rev 56. – 2023, P. 13407–13461.

ANALYSIS OF TWITTER USERS' SENTIMENT BASED ON TEXT DATA USING CONVOLUTIONAL NEURAL NETWORK

Hnatushenko V., Kashtan V., Ovcharenko M., Ivanko A.

Abstract. *The article considers using convolutional neural networks (CNN) to analyze user sentiment on the Twitter platform based on text data. The relevance of the work is due to the growing amount of unstructured text data in social networks and the need for their effective processing to understand public opinion. The proposed CNN model includes representing text data in vector space, convolutional and pooling layers for extracting relevant features, and fully connected layers for classifying sentiment into positive and negative. To prevent overfitting, we applied the Dropout layers to exclude neurons randomly. An experimental evaluation of the model was conducted on a specially prepared dataset from Twitter. The test results demonstrate the prospects of using CNN for automated analysis of social media users' sentiments.*

Keywords: *user sentiment, convolutional network, unstructured data.*

REFERENCE

1. Salloum, S.A., et al. Using text mining techniques for extracting information from research articles, in *Intelligent natural language processing: Trends and Applications*. – 2018, Springer. P. 373-397
2. Coronavirus tweets NLP [Электронный ресурс]: <https://www.kaggle.com/datasets/datatattle/covid-19-nlp-text-classification>.
3. Saxena P., Sharma S. Systematic Literature Review for Sentiment Analysis Using Big Data social media Streams. In *Proceedings of the 2022 5th International Conference on Contemporary Computing and Informatics (IC3I)*, Uttar Pradesh, India, 14–16 December 2022.
4. Tiwari, D., Nagpal, B., Bhati, B.S. et al. A systematic review of social network sentiment analysis with comparative study of ensemble-based techniques. *Artif Intell Rev* 56. – 2023, P. 13407–13461.