

УДК 004.8

Волколуп Б.А., Пасічник О.А., Скрипник Т.К.

Хмельницький національний університет

МЕТОД КЛАСИФІКАЦІЇ НАСТРОЇВ У ТЕКСТАХ СОЦІАЛЬНИХ МЕРЕЖ НА ОСНОВІ РЕКУРЕНТНИХ НЕЙРОННИХ МЕРЕЖ

Розглянуто метод автоматичної класифікації настроїв у текстах соціальних мереж з використанням рекурентних нейронних мереж типу LSTM. Запропонована архітектура включає шар вкладень з попередньо навченими векторами Word2Vec, LSTM шар для захоплення контекстних залежностей та повнозв'язні шари для класифікації. Розроблено модифікацію для покращення обробки емодзі та сленгу через спеціальні словники. Метод забезпечує точність класифікації 89.4%.

A method for automatic sentiment classification in social media texts using LSTM recurrent neural networks is considered. The proposed architecture includes an embedding layer with pre-trained Word2Vec vectors, an LSTM layer for capturing contextual dependencies, and fully connected layers for classification. A modification for improved processing of emoji and slang through specialized dictionaries has been developed. The method achieves classification accuracy of 89.4%.

Автоматичний аналіз настроїв у соціальних мережах є важливим інструментом для моніторингу громадської думки, аналізу відгуків споживачів та виявлення трендів [1, 2]. Традиційні методи класифікації настроїв базуються на словниках емоційно забарвлених слів та лінійних моделях, які не враховують контекстних залежностей між словами [3, 4]. Використання рекурентних нейронних мереж дозволяє автоматично виявляти складні семантичні патерни та контекстні залежності у текстах.

Метою роботи є розробка методу автоматичної класифікації настроїв у текстах соціальних мереж на основі рекурентних нейронних мереж типу LSTM з модифікацією для покращення обробки емодзі та сленгу, який забезпечує високу точність при прийнятній швидкості обробки.

Запропонований метод базується на використанні глибокого навчання для послідовної обробки текстових даних. Вхідними даними є текстові повідомлення з соціальних мереж максимальною довжиною 128 токенів, а вихідними – клас настрою (позитивний, негативний або нейтральний). Метод складається з чотирьох послідовних етапів (рисунок 1): попередня обробка та токенизація тексту, векторизація токенів з використанням Word2Vec, обробка послідовності LSTM шаром для захоплення контекстних залежностей, класифікація повнозв'язними шарами з функцією активації софтмакс.

Етап попередньої обробки включає видалення URL-адрес та згадок користувачів, обробку хештегів (видалення символу # зі збереженням тексту), видалення спеціальних символів та знаків пунктуації, приведення до нижнього

регістру, токенизацію з використанням бібліотеки NLTK та видалення стоп-слів. Модифікація методу передбачає додатковий етап обробки емодзі та сленгу через спеціальні словники. Словник емодзі містить 100 найпоширеніших емодзі з відповідними текстовими описами емоцій (наприклад, 😄 замінюється на "щасливий"). Словник сленгу нормалізує неформальні вирази до стандартних форм (наприклад, "lol" → "laughing out loud").

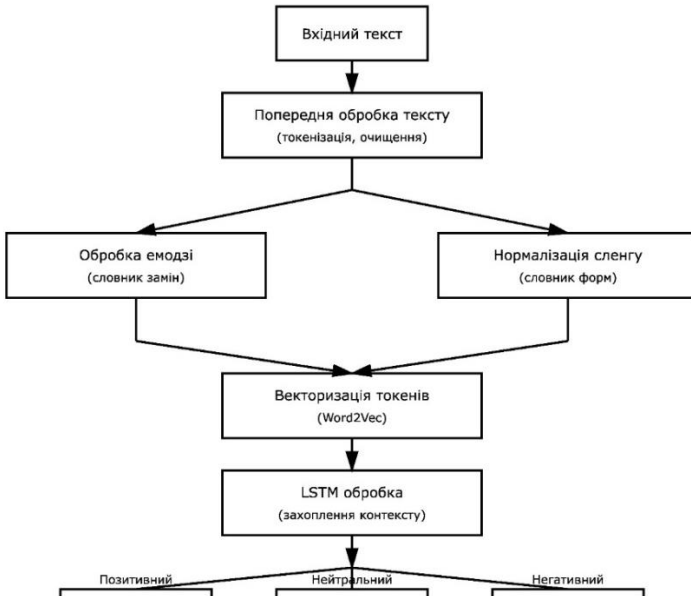


Рисунок 1 – Загальна схема методу класифікації настроїв

Архітектура нейронної мережі. Базова архітектура мережі включає послідовне з'єднання шару вкладень, LSTM шару, шару дропаут та двох повнозв'язних шарів (рисунок 2). Шар вкладень перетворює токени у векторні представлення розмірністю 300, ініціалізовані попередньо навченими векторами Word2Vec. Використовується тонке налаштування вкладень під час навчання для адаптації до специфіки задачі.

LSTM шар є основним компонентом архітектури і має 128 прихованих одиниць. Шар використовує три вентиля для контролю потоку інформації: вентиль забування визначає, яку частину попереднього стану забути; вентиль входу контролює додавання нової інформації; вентиль виходу формує прихований стан.

Після LSTM шару розташований шар дропаут з коефіцієнтом 0.5 для зменшення перенавчання. Перший повнозв'язний шар містить 512 нейронів з активацією ReLU, другий – 256 нейронів також з ReLU. Обидва повнозв'язні шари використовують дропаут 0.5. Вихідний шар має три нейрони з функцією активації софтмакс для формування розподілу ймовірностей по класах настроїв.

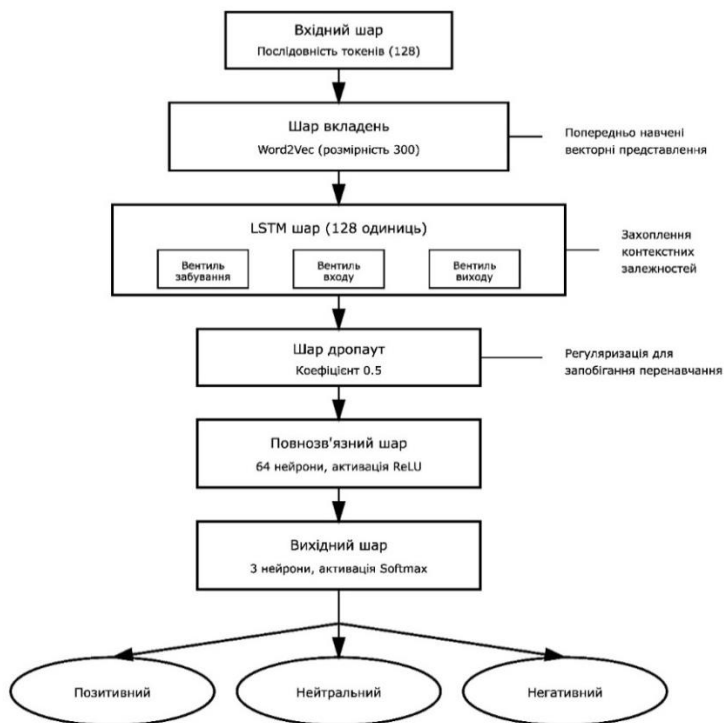


Рисунок 2 – Архітектура моделі класифікації настроїв

Отже, розроблений метод класифікації настроїв на основі рекурентних нейронних мереж LSTM забезпечує точність 89.4%. Модифікація з обробкою емодзі та сленгу через спеціальні словники покращує точність на 1.8% без збільшення обчислювальної складності. Метод є придатним для практичного застосування в системах моніторингу соціальних мереж та аналізу відгуків користувачів.

Перелік посилань

1. Murat Baçal. Natural Language Processing for Sentiment Analysis in Social Media Marketing. Economics World. 2025. Vol. 12, No. 1. URL: <https://doi.org/10.17265/2328-7144/2025.01.004>.
2. Nikhil N., Pratik P., Rutuja P., Rutuja P., Suyash W. Sentiment Analysis Using Natural Language Processing. International Journal of Scientific Development and Research. 2022. Vol. 7, No. 5. Pp. 458–462.
3. Shad R., Potter K., Gracias A. Natural Language Processing (NLP) for Sentiment Analysis: A Comparative Study of Machine Learning Algorithms. International Journal of Artificial Intelligence and Machine Learning. 2025. Vol. 5, No. 1. Pp. 58–69. URL: <https://doi.org/10.51483/IJAIML.5.1.2025.58-69>.
4. Muthukumar P., Ibrahim M. A. S. Multimodal Social Media Sentiment Analysis. Stanford CS224N Final Project. 2024. Pp. 1–12.