Красовський А.О., здобувач вищої освіти

Національний університет «Києво-Могилянська академія», м.Київ

ORCID: 0009-0005-0164-0557

Кальченко А.С., Technical Trainer

Grid Dynamics, м. Київ

ORCID: 0000-0002-2782-1840

**Дослідження методів машинного навчання для вирішення задачі автоматичного виявлення фейкових новин на основі відкритих джерел даних**

**Ключові слова:** *фейкові новини, дезінформація, NLP, fake detection, dataset, очищення даних, машинне навчання, модель, FNet, FastText, AWD-LSTM.*

**Вступ.** Поява та розвиток інтернету надали людству можливість легкого та швидкого поширення інформації у маси. Останнім часом з’явилася велика кількість онлайн-ресурсів, які створюють нескінченний потік новин. Збільшення обсягів інформації призводить до появи великої кількості фейкових новин, які можуть вводити в оману широке коло людей навіть у світовому масштабі. Часто вони спрямовані на дезінформацію, маніпуляцію, пропаганду, створення конфліктів та поляризацію суспільства, тому так гостро постає питання розпізнавання фейкових новин.

На сучасному етапі розвитку інформаційних технологій стрімко розвиваються методи автоматичного виявлення недостовірних новин на основі статистичних, морфологічних, семантичних ознак і контекстуальних залежностей між елементами тексту. Важливу роль у цьому відіграють методи машинного навчання для виявлення неправдивих новин, які дозволяють автоматично класифікувати новини як правдиві або фейкові, що стало предметом даного дослідження.

**Мета.** Отримання ефективних моделей машинного навчання на основі відкритих джерел даних для задачі виявлення неправдивих новин в умовах обмеженості обчислювальних ресурсів, оперативної пам’яті та часу.

У роботі використано датасет Combined Corpus, створений авторами роботи «A Benchmark Study of Machine Learning Models for Online Fake News Detection», оскільки він зібраний з різних джерел та містить великий обсяг даних на різноманітні теми. На цьому датасеті було проаналізовано різні моделі машинного навчання: від SVM до трансформерів сімейства BERT. Найкращі результати були отримані на попередньо навченому трансформері RoBERTa та ансамблі з 5 Bi-LSTM [1]. З цими результатами було вирішено порівняти результати власних досліджень.

Обидва згаданих вище рішення є доволі вимогливими до обсягу обчислювальних ресурсів, тому для виконання дослідження використовувалися три сучасні моделі, які не вимагають великого обсягу обчислювальних ресурсів, оперативної пам’яті та часу, а саме: FNet, AWD-LSTM та FastText.

**FNet** – архітектура, створена як швидша альтернатива класичним трансформерам без значної втрати точності. Її особливість полягає у заміні механізму самоуваги швидкими перетвореннями Фур’є для «перемішування» токенів, що дозволяє знизити складність обчислень [2]. За основу взято базову імплементацію FNet з офіційного сайту Keras [3].

**AWD-LSTM** – це сучасніша та покращена версія звичайної LSTM мережі з використанням різних технік оптимізації та регуляризації: weight tying, змінної довжини послідовностей у зворотному поширенні, варіативного dropout, тимчасової регуляризації активацій та ін. [4]. Для подальшого навчання було обрано імплементацію моделі від FastAI [5].

**FastText** – модель для створення векторних представлень слів, запропонована командою розробників Facebook AI Research. На відміну від традиційних векторних представлень, FastText не використовує цілі слова для створення векторів, а розбиває їх на n-грами. Завдяки цьому модель добре працює з великою різноманітністю слів, у тому числі з рідковживаними [6].

Для оцінки якості моделей було вирішено використовувати зважену метрику F-1 (тут і надалі просто F-1 метрика чи оцінка), яка враховує незбалансованість у передбаченнях моделей та дозволяє більш повно оцінити результат навчання, на відміну від точності (accuracy).

У ході дослідження отримано такі результати за F-1 метрикою: FNet – 0.97; AWD-LSTM – 0.92; FastText – 0.95. Нижче наведено F-1 метрики моделей, розглянутих під час дослідження, у порівнянні з найкращими моделями у роботі «A Benchmark Study of Machine Learning Models for Online Fake News Detection», де синім кольором виділено дані зі статті, зеленим – результати власних досліджень [1].

Рис. 1: Порівняння F-1 метрики різних моделей на датасеті Combined Corpus.

Як видно із наведеної діаграми, FNet перевершує результат трансформера RoBERTa на 0.01 (див. Рис. 1), але з огляду на розміри цих моделей (4.3 мільйони параметрів та 124.6 мільйонів параметрів відповідно) можна сказати, що перевага FNet навіть більша, адже вона швидше навчається та вимагає менше ресурсів для обчислень. FastText класифікатор досяг однакового результату з Bi-LSTM і лише на 0.01 поступається трансформеру RoBERTa (див. Рис. 1). Варто зазначити, що автори згаданої раніше роботи використовували ансамбль з 5 моделей Bi-LSTM для досягнення таких результатів, але такий підхід вимагає більших обчислювальних потужностей. Крім того, FastText призначений для навчання лише на CPU, і, зважаючи на його високі показники, можна сказати, що навчання такої моделі буде дешевшим, ніж тренування глибинної нейронної мережі.

FNet тренувалася значно швидше (у середньому приблизно 1 хвилина на епоху), ніж AWD-LSTM (17-25 хвилин на епоху) у тому ж середовищі виконання (з ідентичними GPU), тому такі високі результати роблять її особливо привабливою для використання в автоматизованих системах виявлення фейків.

Тренування FastText складалося з двох етапів: етапу автоматичного підбору гіперпараметрів, часом якого можна керувати (у цій роботі тривалість підбору була 20 хвилин), та етапу навчання моделі, який тривав приблизно 20 хвилин. Таким способом, готову модель було отримано за 40 хвилин. Це довше, ніж тренування FNet, але швидше за навчання AWD-LSTM. Проте вартість ресурсів, які потрібні для моделі FastText, значно нижча за вартість GPU, від параметрів яких суттєво залежить швидкість навчання глибинних нейронних мереж.

Нижче у таблиці наведено детальні дані метрик кожної з отриманих моделей. Для більш детального аналізу наводяться не тільки загальні метрики, а й метрики для кожного класу окремо.

|  |
| --- |
| **Розгорнуті дані метрик досліджених моделей машинного навчання** |
| Метрики | FNet | FastText | AWD-LSTM |
| F-1 оцінка | загальна | 0.97 | 0.95 | 0.92 |
| фейки | 0.95 | 0.93 | 0.87 |
| правдиві новини | 0.98 | 0.97 | 0.94 |
| Precision | загальний | 0.97 | 0.95 | 0.93 |
| фейки | 0.96 | 0.97 | 0.98 |
| правдиві новини | 0.97 | 0.95 | 0.90 |
| Recall | загальний | 0.97 | 0.95 | 0.92 |
| фейки | 0.94 | 0.89 | 0.77 |
| правдиві новини | 0.98 | 0.99 | 0.99 |

**Висновки.** Отримано три різні моделі: FNet, FastText та AWD-LSTM, на яких досягнуто високих результатів у класифікації фейкових новин. За допомогою архітектури FNet було досягнуто найвищого результату 0.97 на метриці F-1. Цей показник перевищує найкращу оцінку 0.96, отриману на трансформерах авторами статті «A Benchmark Study of Machine Learning Models for Online Fake News Detection» [1]. Зважаючи на значно менший розмір моделі FNet, отриманий результат можна вважати більш ефективним.

**Перспективи подальших досліджень.** Модель FNet як альтернатива трансформерам потенційно може бути застосована для україномовних новин, як і FastText, яка завдяки специфічному представленню слів також здатна працювати зі складними з погляду морфології та граматики мовами. Це може стати ідеєю для продовження та подальшого розвитку цього дослідження, адже україномовних датасетів загалом мало, а розмічених датасетів новин у відкритому доступі наразі немає.

# Література

1. A benchmark study of machine learning models for online fake news detection / J. Y. Khan et al. *Machine Learning with Applications*. 2021. Vol. 4. P. 100032. URL: <https://doi.org/10.1016/j.mlwa.2021.100032>.

2. FNet: Mixing Tokens with Fourier Transforms / J. Lee-Thorp et al. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Seattle, United States. Stroudsburg, PA, USA, 2022. URL: <https://doi.org/10.18653/v1/2022.naacl-main.319>.

3. Keras documentation: Text Classification using FNet. Keras: Deep Learning for humans. URL: <https://keras.io/examples/nlp/fnet_classification_with_keras_hub/>.

4. Merity S., Keskar N. S., Socher R. Regularizing and Optimizing LSTM Language Models. 2017. URL: <http://arxiv.org/abs/1708.02182>.

5. AWD-LSTM documentation. fastai. URL: <https://docs.fast.ai/text.models.awdlstm.html>.

6. Enriching Word Vectors with Subword Information / P. Bojanowski et al. *Transactions of the Association for Computational Linguistics*. 2017. Vol. 5. P. 135–146. URL: <https://doi.org/10.1162/tacl_a_00051>.