

Разработка модели обнаружения недостоверных новостей в социальной сети в формате графа знаний

Головин Алексей Андреевич
каф. МОЭВМ
СПбГЭТУ «ЛЭТИ»
Санкт-Петербург, Россия
golovin99@icloud.com

Куликов Игорь Александрович
каф. МОЭВМ
СПбГЭТУ «ЛЭТИ»
Санкт-Петербург, Россия
i.a.kulikov@gmail.com

Жукова Наталия Александровна
каф. МОЭВМ
СПбГЭТУ «ЛЭТИ»
Санкт-Петербург, Россия
nazhukova@mail.ru

Аннотация—Статья посвящена проблеме распространения недостоверных новостей в социальной сети. Исследованы методы определения достоверности новостей. Описан выбранный набор данных, а также его представление в виде графа знаний. Приведено описание и сравнение разработанных алгоритмов, основанных на языковой модели BERT и графовой архитектуре GAT.

Ключевые слова—новости; социальные сети; графы знаний; машинное обучение; BERT; GNN; GAT

I. ВВЕДЕНИЕ

Развитие интернета, а также социальных сетей позволило быстро и эффективно делиться информацией. В настоящее время люди более склонны узнавать новости из социальных сетей, а не из традиционных новостных изданий [1]. Это привело к широкому распространению недостоверных новостей или, как их принято сейчас называть, «fake news» [2]. «fake news» могут представлять собой как просто низкокачественную или неполную информацию, так и заведомо ложные политически замотивированные новости.

Распространение недостоверных новостей крайне негативно влияет на людей и общество. Это особенно актуально в эпоху пандемии Covid19, так как легко ввести людей в заблуждение и заставить поверить в ложную информацию, что способно привести их к фатальному исходу [3]. Поэтому поиск и прогнозирование распространения недостоверных новостей являются актуальными и важными областями исследований [4]. Вовремя обнаруживая и предсказывая распространение «fake news» можно своевременно применять меры по предотвращению их дальнейшего продвижения [5].

Целью данной работы является разработка модели детекции недостоверных новостей в социальной сети в формате графа знаний.

II. МЕТОДЫ РЕШЕНИЯ

На сегодняшний день стало уже классическим использование нейросетевой языковой модели BERT для решения широкого спектра задач обработки естественного

языка. BERT нашел применение и для решения проблемы определения достоверности новости [6].

При этом существует целое семейство моделей, являющихся модифицированными версиями оригинального трансформера BERT. Эти алгоритмы настолько популярны, что существует область исследований BERTology, связанная с изучением их внутренней работы и сравнением производительности для разных задач [7].

Технология графов знаний, в свою очередь, также используется для улучшения результатов моделей, использующих BERT [8] и позволяет улучшить результаты классификации. Сочетание языковой модели BERT и технологии графов знаний нашло свое применение в алгоритме KG-BERT [9]. Данный метод используется для восстановления отсутствующих узлов и связей в графе.

Существуют и активно используются методы, использующие только графовые подходы (например, GNN) для решения задачи определения достоверности новости [10-11].

Таким образом, можно сделать вывод, учитывая перспективность технологии графов знаний для решения схожих задач, что будет полезно рассмотреть графовые подходы в сравнении с уже ставшим классическим и общепринятым алгоритмом BERT, а также сравнить их производительность.

III. НАБОР ДАННЫХ

В качестве набора данных для проведения исследования выбран датасет «FakeNewsNet» [12]. Датасет содержит данные о реальных и недостоверных новостях: источники, заголовки и упоминания в социальной сети Twitter. Эти данные используются для классификации новостных публикаций с целью выявления недостоверных новостей. Данные были предварительно обработаны, очищены и опубликованы на платформе Kaggle [13]. Полученный набор данных представляет собой CSV файл, в котором каждый элемент является публикацией новостной статьи.

В датасете содержится более 20 тысяч записей. Атрибутами являются: заголовок, ссылка на статью, источник новости, количество пользователей, поделившихся новостью, а также метка достоверности статьи.

Используемый набор данных также преобразован к графовому виду [14], что полезно для дальнейшего использования графовых подходов. В данной версии каждая новость представляется отдельным графом, у которого корневой узел – это сама новость, а последующие узлы – это пользователи Twitter, которые делились друг с другом этой новостью (Fig. 1). Кроме того, узлы содержат текстовую информацию о профилях (предобработанную BERT).

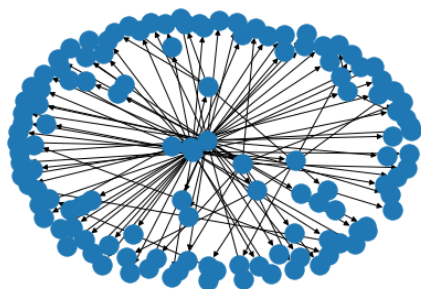


Fig. 1. Граф распространения новостной статьи

Таким образом, на данный момент данные полностью готовы для их дальнейшего использования.

IV. ПОЛУЧЕННЫЕ РЕЗУЛЬТАТЫ

A. «Baseline» алгоритм

В качестве «baseline» модели, то есть используемой как ориентир для оценки качества работы алгоритма, использующего технологию графов знаний, использовалась языковая модель DistilBERT. Это один из алгоритмов ранее упомянутого семейства нейросетевых архитектур BERT [15].

DistilBERT основан на идее дистилляции. Её суть в имитации легковесной моделью поведения более сложной модели-учителя, в данном случае BERT [16]. Языковая модель DistilBERT оптимизирует процесс обучения за счет уменьшения размера и увеличения скорости BERT, при этом сохраняя до 97% производительности.

Модель была использована для получения признаков из новостных заголовков с целью дальнейшей их обработки с помощью алгоритмов логистической регрессии для бинарной классификации новостей на достоверные и не достоверные.

DistilBERT получает на вход заголовки новостных статей, токенизирует их и, далее, на GPU выделяет признаки и подает их в следующую модель для обработки. Преимущество использования алгоритмов семейства BERT в двустороннем анализе контекста, что дает

возможность легко получить достаточно хорошие результаты.

Для задачи определения достоверности новости модель, использующая связку DistilBERT и логистическую регрессию, показала хорошие результаты и достигла точности 82% на тестовом наборе данных (авторы датасета получили результаты порядка 60-70% методами классического машинного обучения [12]). В качестве подтверждения достаточной эффективности DistilBERT относительно классического BERT, было произведено сравнение, которое показало, что модель, которая использует BERT, на тех же данных достигает точности 83,2%. Таким образом, удалось убедиться в способности DistilBERT сохранять эффективность при уменьшении размера и увеличении скорости обучения.

B. Графовый алгоритм

Основной моделью, использующей графовые нейросетевые технологии, выбрана GNN модель. Применялась архитектура графовой сети с механизмом внимания GAT [17] с тремя свёрточными слоями GATConv. Механизм внимания помогает учитывать важность каждого соседнего узла, присваивая ребрам весовые коэффициенты (Fig. 2). Принципиально новым для данной архитектуры является использование технологии «самовнимания», для которой можно провести аналогию с важностью узла. Несмотря на то, что графовые сети внимания обучаются дольше, точность у них существенно выше, чем у обычных графовых сверточных сетей.

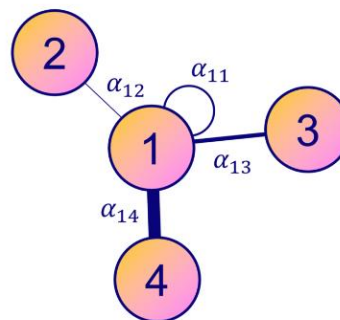


Fig. 2. Иллюстрация механизма внимания

Для задачи обнаружения недостоверных новостей нейронная сеть GAT достигла точности классификации 92% на тестовом наборе данных.

Таким образом, графовый алгоритм позволил улучшить результаты модели, использующей DistilBERT.

V. ЗАКЛЮЧЕНИЕ

В рамках данной работы достигнута поставленная задача – разработана модель детекции недостоверных новостей в социальной сети в формате графа знаний.

Графовая нейронная сеть с механизмом внимания GAT обеспечила точность 92% на датасете «FakeNewsNet». При этом результаты классических алгоритмов машинного обучения на этом наборе данных составляют 60-70%, а использование DistilBERT позволяет достичь точности 82%. Решения, использующие графовые подходы без механизма внимания, на используемом наборе данных обеспечивают точность 80-85%.

Дополнительным преимуществом графового подхода является способность алгоритма сохранять свою эффективность с течением времени по сравнению с алгоритмами, основанными только на обработке естественного языка. Текстовые признаки в заголовках изменяются при появлении новых тенденций в мире. Графовый алгоритм основан на структуре графа распространения новости, а также на информации о пользователях, что позволяет ему эффективно работать в новых условиях.

Таким образом, можно сделать вывод, что использование графовых подходов, а в частности архитектуры GAT способно значительно улучшить качество бинарной классификации новостных статей на достоверные и не достоверные.

ЛИТЕРАТУРА

- [1] Shu K., Wang S., Liu H. Exploiting Tri-Relationship for Fake News Detection // Computer Science and Engineering. 2017. P. 1–10.
- [2] Fake News Detection Using Machine Learning Ensemble Methods / I. Ahmad, M. Yousaf, S. Yousaf, et al. // Hindawi Complexity. 2020. P. 1–11. DOI: <https://doi.org/10.1155/2020/8885861>.
- [3] Evaluating Deep Learning Approaches for Covid19 Fake News Detection / A. Wani, I. Joshi, S. Khandve, et al. // Communications in Computer and Information Science. 2021, Vol. 1402. P. 153–163. DOI: https://doi.org/10.1007/978-3-030-73696-5_15.
- [4] Fake News Detection on Social Media: A Data Mining Perspective / K. Shu, A. Sliva, S. Wang, et al. // ACM SIGKDD Exploration Newsletter. 2017, Vol. 19, № 1. P. 22–36. DOI: <https://doi.org/10.1145/3137597.3137600>.
- [5] Tian L., Zhang X., Peng M. FakeFinder: Twitter Fake News Detection on Mobile // Companion Proceedings of the Web Conference. 2020. P. 79–80. DOI: <https://doi.org/10.1145/3366424.3382706>.
- [6] Kaliyar R.K., Goswami A., Narang P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach // Multimed Tools Appl. 2021, Vol. 80. P. 11765–11788. DOI: <https://doi.org/10.1007/s11042-020-10183-2>.
- [7] BERTology // Hugging Face – The AI community building the future. URL: <https://huggingface.co/docs/transformers/bertology> (дата обращения: 17.12.2022).
- [8] Enriching BERT with Knowledge Graph Embeddings for Document Classification / M. Ostendorff, P. Bourgonje, M. Berger, et al. // Proceedings of the 15th Conference on Natural Language Processing. 2019. P. 1–8.
- [9] Yao L., Mao C., Luo Y. KG-BERT: BERT for knowledge graph completion // arXiv preprint arXiv:1909.03193. 2019.
- [10] Monti F. et al. Fake news detection on social media using geometric deep learning // arXiv preprint arXiv:1902.06673. 2019.
- [11] Hu L. et al. Compare to the knowledge: Graph neural fake news detection with external knowledge // Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2021. – P. 754–763. DOI: <https://doi.org/10.18653/v1/2021.acl-long.62>.
- [12] Shu K. et al. FakeNewsNet: A Data Repository with News Content, Social Context and Spatiotemporal Information for Studying Fake News on Social Media // arXiv preprint arXiv:1809.01286. 2018.
- [13] Fake News // Kaggle. URL: <https://www.kaggle.com/datasets/algord/fake-news> (дата обращения: 17.12.2022).
- [14] Dou Y. et al. User Preference-aware Fake News Detection // Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2021.
- [15] Все, что нужно знать об ALBERT, RoBERTa и DistilBERT // Хабр. URL: <https://habr.com/ru/post/680986/> (дата обращения: 17.12.2022).
- [16] Простое руководство по дистилляции BERT // Хабр. URL: <https://habr.com/ru/company/avito/blog/485290/> (дата обращения: 17.12.2022).
- [17] Velickovic P. et al. Graph Attention Networks // International Conference on Learning Representations. 2018.