

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

Выполнил:

Головин Алексей Андреевич, гр. 7381

Руководитель:

Жукова Наталия Александровна, д.т.н., профессор

Консультант:

Куликов Игорь Александрович, инженер каф. МОЭВМ

Цель и задачи

Цель: Разработка и реализация модели бинарной классификации новостных статей в социальной сети на достоверные и недостоверные с использованием технологии графов знаний

Задачи:

1. Анализ существующих методов решения.
2. Выбор и анализ подходящего набора данных.
3. Разработка модели обнаружения недостоверных новостей.
4. Реализация программного комплекса.
5. Исследование полученных результатов.

Актуальность работы

Распространение недостоверных новостей крайне негативно влияет на людей и общество. Способно легко ввести людей в заблуждение и заставить поверить в ложную информацию, что может привести к негативным последствиям.

Внедрение разработанной модели позволит своевременно обнаруживать недостоверные новости в социальной сети

Новизна работы

- Разработанная модель обнаружения недостоверных новостей в социальной сети построена на основе графовой модели с вниманием GAT, что отличает ее от существующих, в которых важность узлов в графе не учитывается. В отличие от традиционных методов обработки естественного языка, использование данной модели позволяет обеспечить эффективное решение задачи при изменении тенденций в мире с течением времени.
- Методика оценки эффективности графовых моделей, обеспечивающая корректное сравнение с другими методами. Исследование проводилось с использованием одного и того же набора данных в двух видах: табличном и предварительно преобразованном к формату графа знаний с помощью технологии RML.

Анализ существующих методов решения

Аналоги:

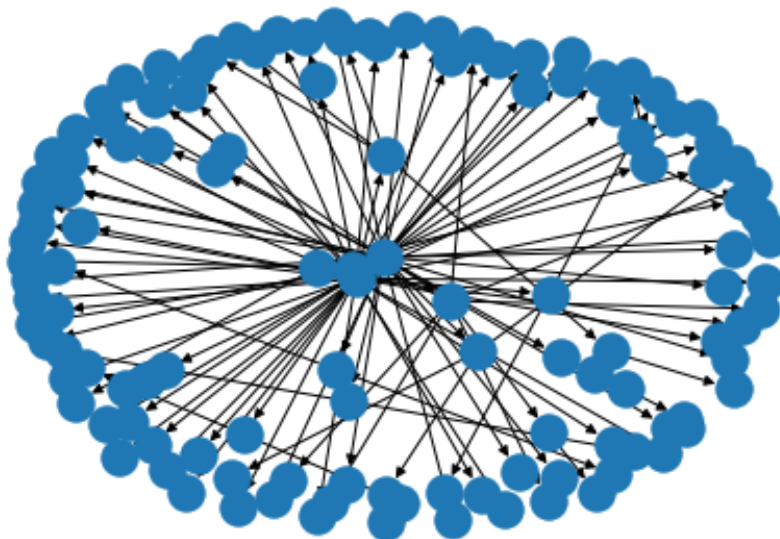
- Методы решения, основанные на классических моделях машинного обучения: наивный байесовский классификатор, логистическая регрессия, SVM. Модели широко используются в современных научных исследованиях в качестве методов, с которыми в дальнейшем сравниваются предлагаемые авторами модели.
Недостатки: устаревшие, нереалистичность допущений (наивный Байес).
- Языковая модель DistilBERT. Выбор объясняется тем, что большинство современных аналогов разрабатываемой модели основаны на методах обработки естественного языка. DistilBERT является мощным инструментом, позволяющим получить достаточно хорошие результаты.
Недостатки: основан на обработке текста, т.е. зависим от содержимого заголовков.

Метод, используемый при разработке модели обнаружения недостоверных новостей:

- Графовая нейросетевая модель GAT. Выбор объясняется высокой эффективностью данного метода, обеспеченной технологией внимания, а также способностью сохранять эффективность с течением времени.

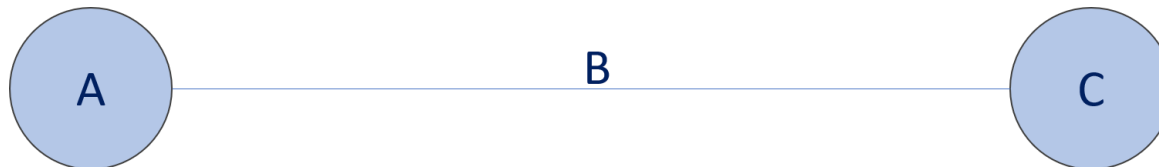
Выбор и анализ подходящего набора данных

- Выбран датасет FakeNewsNet. Набор данных содержит информацию о реальных и недостоверных новостях. Версия датасета, преобразованная к формату графа знаний, использовалась для применения графовой нейронной сети.
- 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.



Графы знаний

- Структура данных, использующаяся для моделирования отношений между понятиями, объектами или событиями в области знаний
- Состоит из трёх основных компонентов: узлов, ребер и меток
- Для интеграции данных широко используется технология RML. Исходные данные могут быть представлены в различных форматах (CSV, JSON, XML, SQL).
- Главное преимущество – возможность применения графовых моделей



Разработка модели обнаружения недостоверных новостей

Рисунок 1 – Архитектура нейросети GAT

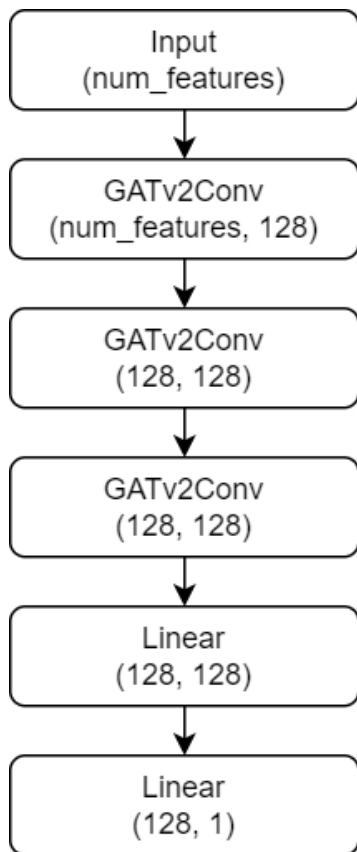
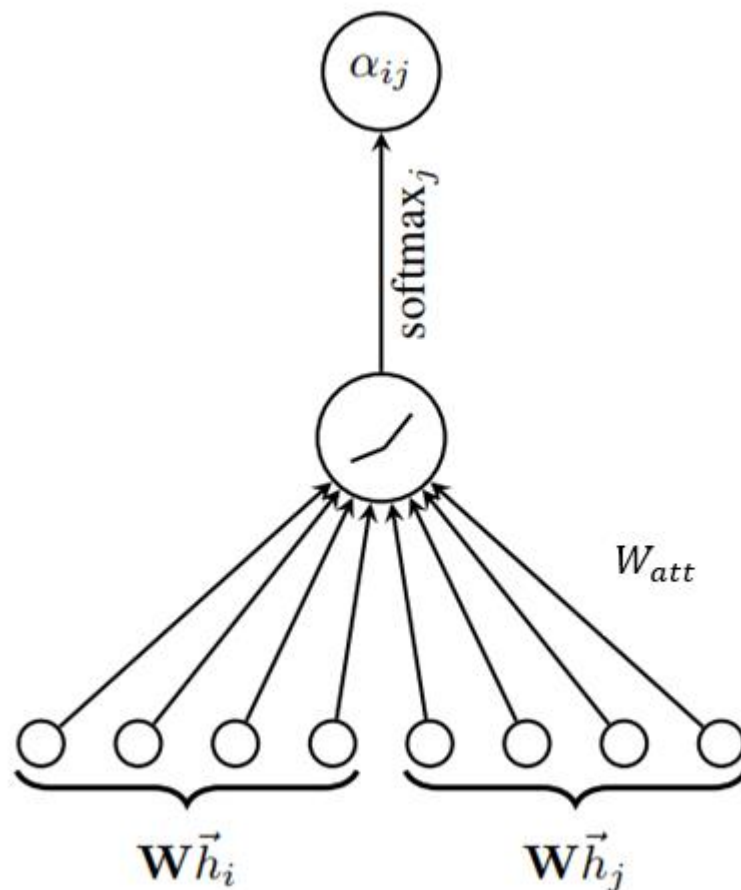


Рисунок 2 – Механизм внимания



Формула коэффициентов внимания для GATv2Conv:

$$e_{ij} = W_{att}^T \text{LeakyReLU}(W[\vec{h}_i || \vec{h}_j]), \quad \alpha_{ij} = \text{softmax}(e_{ij}),$$

где W и W_{att} – весовые матрицы, \vec{h}_k – признаки узла k , $||$ – конкатенация.

Алгоритм работы программного комплекса

Рисунок 1 – Построение моделей аналогов

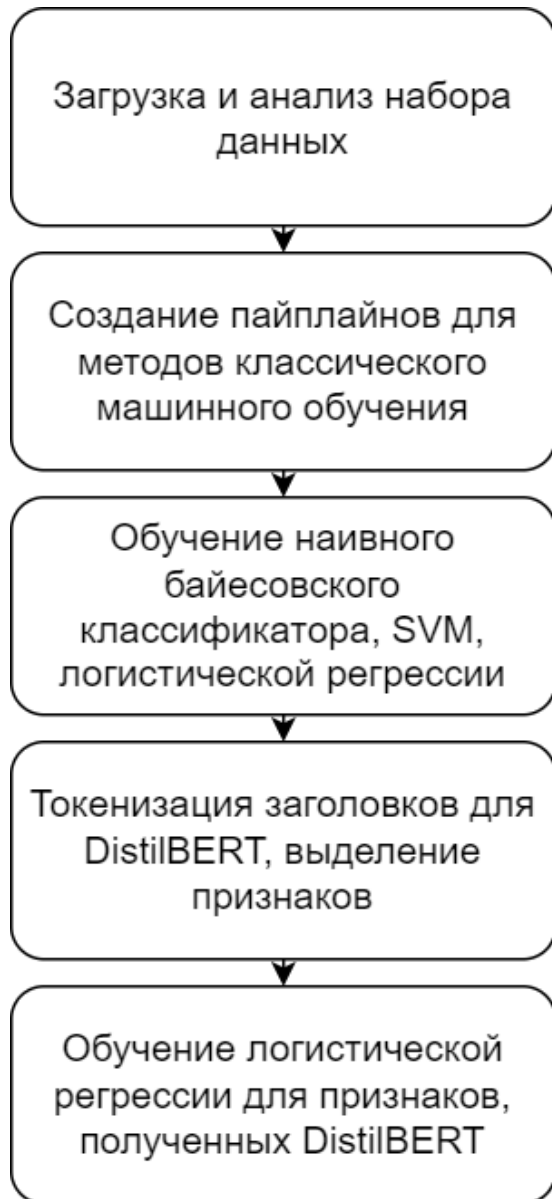


Рисунок 2 – Построение графовой модели



Реализация программного комплекса

- Для программной реализации разрабатываемой модели использован высокоуровневый интерпретируемый язык программирования Python 3.10.
- Основными используемыми в данном программном комплексе библиотеками являются NumPy, Pandas, Matplotlib, PyTorch и PyG
- Итоговая модель с демонстрацией работы и сравнением с другими моделями представлена в формате Jupyter Notebook

Методика оценки эффективности

- Основным показателем сравнения эффективности модели выбрана метрика F1-мера
- F1-мера (F1-score) – это среднее гармоническое precision и recall и является более сбалансированной метрикой, чем каждая из них по отдельности

$$F1 = 2 \times \frac{precision \times recall}{precision + recall},$$

где:

$$precision = \frac{TP}{TP + FP}, \quad recall = \frac{TP}{TP + FN}.$$

- TP – истинно положительные прогнозы
- TN – истинно отрицательные прогнозы
- FP – ложно положительные прогнозы
- FN – ложно отрицательные прогнозы

Обучение нейросети GAT

- Обучение на платформе Google Colab на GPU
- Данные разделены на обучающую, валидационную и тестовую выборки
- Оптимизатор: Adam
- Функция ошибок: BCELoss
- Реализована ранняя остановка
- Наилучшее состояние модели сохраняется с помощью сериализации

Рисунок 1 – График зависимости значения функции ошибки от номера эпохи

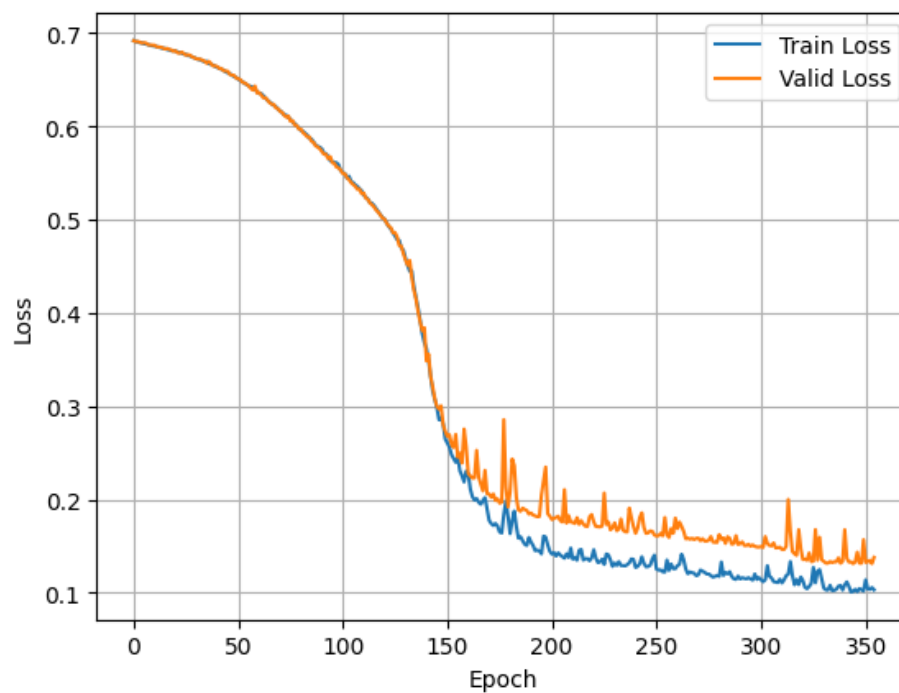
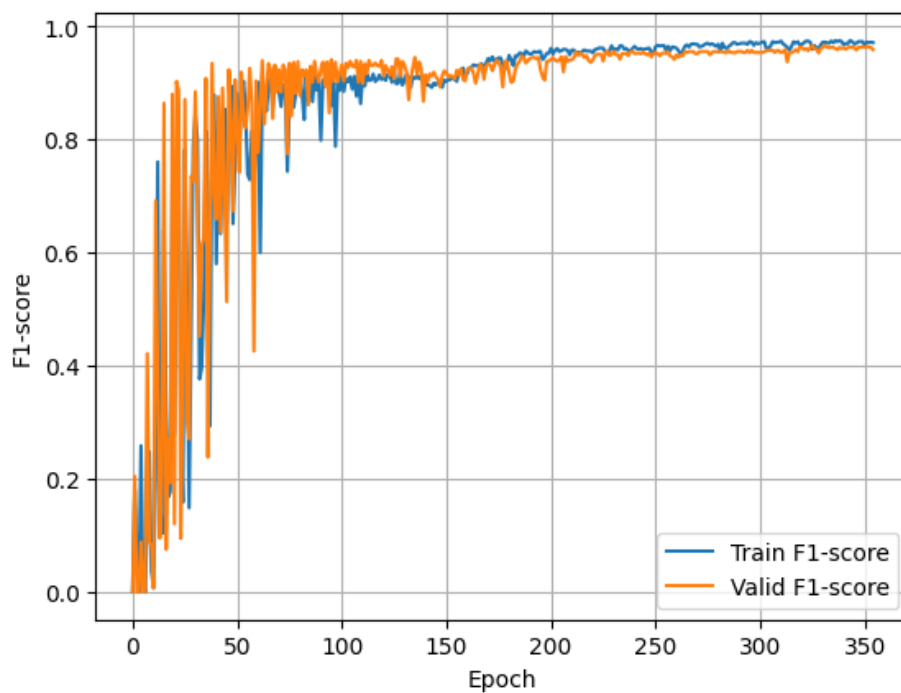


Рисунок 2 – График зависимости значения F1-меры от номера эпохи



Исследование полученных результатов

Таблица 1 – Сравнения моделей по метрикам

Название метода	Точность	F1-мера
Наивный Байес	85%	64%
SVM	85%	65%
Логистическая регрессия	83%	59%
DistilBERT	84%	82%
GAT	96%	96%

Использование графовых подходов, а в частности архитектуры GAT способно значительно улучшить качество бинарной классификации новостных статей на достоверные и не достоверные

Дополнительным преимуществом графового подхода является способность модели сохранять свою эффективность с течением времени по сравнению с моделями, основанными только на обработке естественного языка

Заключение

Все поставленные задачи были успешно выполнены:

- Проанализированы существующие методы решения проблемы обнаружения недостоверных новостей в социальной сети
- Выбран и проанализирован набор данных «FakeNewsNet». С помощью технологии RML был приведен к графовому виду.
- Разработана модель обнаружения недостоверных новостей, основанная на графовой нейронной сети с вниманием GAT
- Реализован программный комплекс для выбранных моделей на языке Python с помощью библиотек PyTorch, PyG.
- Полученные результаты исследованы. Модель, использующая архитектуру GAT, позволила значительно улучшить качество обнаружения недостоверных новостей и обеспечила значение F1-меры 96%.

Направление дальнейшей разработки:

- Внедрение разработанной модели в социальную сеть

Апробация работы

- Головин А.А., Куликов И.А., Жукова Н.А. Разработка модели обнаружения недостоверных новостей в социальной сети в формате графа знаний // Научно-технический семинар кафедры МО ЭВМ. – 2023. – С. 17-19.
- Golovin A.A. Fake News [Набор данных] // Kaggle. – 2022.
DOI: <https://doi.org/10.34740/KAGGLE/DS/2623949>
- Репозиторий проекта: <https://github.com/A1gord/GraduationProject>



Запасные слайды

Социологическое исследование

Было проведено социологическое исследование с использованием двух методик: интервью с экспертами и социологического опроса.

В качестве экспертов были выбраны:

- Журналист регионального СМИ с опытом работы более 6 лет
- Программист в области ИИ с опытом работы более 5 лет

В качестве выборки социологического опроса использовались участники различных сообществ с разной тематикой для обеспечения наибольшей репрезентативности. В ходе опроса своим мнением поделились 167 человек.

Перед интервью и опросом были получены согласия респондентов на публикацию исследования, а также сохранена их анонимность

Исследование показало, что разработанная модель является актуальной для целевой аудитории и решает социально значимую проблему