ГИБРИДНАЯ СИСТЕМА ФИЛЬТРАЦИИ СПАМ СООБЩЕНИЙ НА ОСНОВЕ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ

Ломонос Г. В., Захарьев В. А.
Кафедра систем управления,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: lomonosg07@gmail.com, zahariev@bsuir.by

Статья посвящена исследованию современных систем и методов обнаружения вредоносных сообщений в электронной почте. В ней представлен обзор наиболее эффективных подходов и технологий, а также рассмотрены ключевые особенности их реализации.

Введение

В современном мире, где цифровая коммуникация занимает центральное место в нашей повседневной жизни, проблема спам-сообщений становится все более актуальной. Спам не только засоряет наши почтовые ящики, но и может представлять серьезную угрозу безопасности, включая фишинг и распространение вредоносного ПО. В связи с этим разработка эффективных систем фильтрации спама приобретает особую значимость [1]. В данной статье мы предлагаем гибридную систему фильтрации спам-сообщений, основанную на использовании моделей машинного обучения. Эта система сочетает в себе как традиционные алгоритмы машинного обучения, так и методы глубокого обучения, что позволяет повысить точность и надежность классификации сообщений.

I. Архитектура системы

Предлагаемая система фильтрации спама состоит из нескольких ключевых компонентов, которые обеспечивают эффективную обработку и классификацию сообщений. Архитектура системы включает следующие этапы:

– Очистка данных: На первом этапе текстовые сообщения проходят через процесс очистки, где удаляется ненужная информация, такая как HTML-теги и специальные символы.

– Предобработка текста: После очистки данные разделяются на два потока для дальнейшей обработки:

– TF-IDF представление: Сообщения преобразуются в числовые векторы с использованием метода TF-IDF, что позволяет выделить важные слова в тексте. TF-IDF – это статистическая мера, используемая для оценки важности слова в контексте документа, который является частью коллекции или корпуса. Формула для вычисления TF-IDF для термина t в документе d из корпуса D выглядит следующим образом:

$$\text{TF-IDF}(t,d,D) = \text{TF}(t,d) \times \text{IDF}(t,D)$$

где

$$\text{TF}(t,d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}}$$

$$\text{IDF}(t,D) = \log\left(\frac{N}{|\{d \in D{:}t \in d\}|}\right)$$

где $f_{t,d}$ — частота термина t в документе d, N — общее количество документов в корпусе, а $|\{d \in D{:}t \in d\}|$ — количество документов, содержащих термин t.

– Векторизация слов: Используются методы векторизации, такие как Word Embedding, для представления слов в виде векторов, что помогает выявлять семантические связи между словами. Одним из популярных методов является Word2Vec [2], который обучает векторные представления слов, оптимизируя задачу предсказания контекста слова. Формула для представления слова w в виде вектора может быть описана как:

$$v_w = \text{Word2Vec}(w)$$

где v_w – это векторное представление слова w, полученное с помощью модели Word2Vec. Обучение таких векторов обычно основано на нейронной сети, которая минимизирует ошибку предсказания соседних слов в тексте.

– Моделирование с использованием машинного обучения: Традиционные методы: На TF-IDF вектора применяются традиционные алгоритмы машинного обучения, такие как Naive Bayes, SVM и Decision Tree, для классификации сообщений. Глубокое обучение: Векторизированные данные обрабатываются с помощью нейронных сетей, что позволяет выявлять более сложные паттерны в текстах.

– Классификация: Оба подхода завершаются этапом классификации, где сообщения разделяются на спам и не спам. Гибридность предлагаемой системы фильтрации спам-сообщений заключается в сочетании различных подходов и технологий машинного обучения для достижения наилучших результатов в классификации сообщений. Основные аспекты гибридности системы включают:

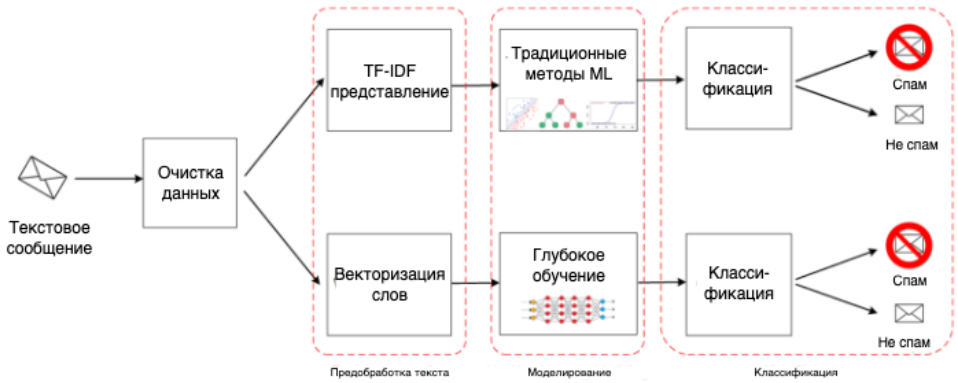

Рис. 1 – Основные аспекты гибридности системы

Комбинация традиционных и глубоких методов обучения: Система использует как традиционные алгоритмы машинного обучения (например, Naive Bayes, SVM, Decision Tree), так и методы глубокого обучения (например, нейронные сети). Это позволяет системе эффективно обрабатывать как простые, так и сложные паттерны в текстах сообщений. Параллельная классификация: После предобработки данные параллельно классифицируются с использованием различных алгоритмов. Это позволяет сравнивать результаты и выбирать наиболее эффективный метод для конкретного набора данных. Интеграция с внешними источниками данных: Гибридная система может интегрироваться с другими источниками данных и методами анализа, такими как анализ сетевого трафика или поведенческий анализ, для более комплексного подхода к фильтрации спама. Таким образом, гибридность системы заключается в интеграции различных технологий и подходов, что позволяет ей быть более гибкой, точной и устойчивой к изменениям в характере спам-сообщений.

II. Тестирование системы

Тестирование системы фильтрации спам-сообщений проводилось с использованием различных моделей машинного обучения, результаты которых представлены в таблице. Основные метрики, такие как точность, полнота и F-мера, были использованы для оценки производительности каждой модели.

Для экспериментов использовались разнообразные наборы данных, такие как SMS Spam Collection, Enron Email Dataset и Twitter Spam Dataset [3]. Эти наборы обеспечивают разнообразие текстовых форматов и контекстов, что позволяет моделям обучаться и тестироваться на различных типах спам-сообщений. SMS Spam Collection предоставляет данные о текстовых сообщениях, Enron Email Dataset – о реальных электронных письмах, а Twitter Spam Dataset – о спаме в социальных сетях. Это разнообразие данных позволяет моделям быть более универсальными и эффективными в реальных условиях.

Таблица 1 – Результаты тестирования системы

Модель	**Точ. (%)**	**Пол. (%)**	**F (%)**
SVM	88.5	86.7	87.6
DT	85.2	83.9	84.5
CNN	91.3	89.8	90.5
LSTM	92.1	90.5	91.3
Hybrid	94.5	93.2	93.8

Гибридная модель, объединяющая различные подходы, показала наилучшие результаты. Это подчеркивает эффективность комбинирования методов для повышения точности классификации.

III. Выводы

Предложенная гибридная система фильтрации спам-сообщений демонстрирует высокую эффективность благодаря сочетанию традиционных и современных методов машинного обучения. Использование различных алгоритмов позволяет системе адаптироваться к изменяющимся условиям и новым угрозам. В результате система обеспечивает надежную защиту от спама и фишинга, минимизируя количество ложных срабатываний и повышая качество пользовательского опыта. В будущем планируется дальнейшее совершенствование системы, включая интеграцию с другими методами анализа данных и расширение функциональности для работы с мультимедийными сообщениями.

IV. Список литературы

1. Y. Zhang, J. Li,&X. Wang. A Comprehensive Survey on Email Spam Filtering: Techniques, Datasets, and Open Challenges. Journal of Information Security and Applications, 2022, vol. 65, 103049.
2. R. Kumar, & P. Singh. Enhancing Email Security: A Hybrid Approach to Spam Detection Using Machine Learning and Deep Learning Techniques. Computers & Security, 2023, vol. 120, 102845.
3. T. T. Ngyueng, & D. M. Tran. A Novel Framework for Social Media Spam Detection Using Advanced Natural Language Processing Techniques. Expert Systems with Applications, 2021, vol. 184, 115502.