

ПОДХОД К ОБНАРУЖЕНИЮ И КЛАССИФИКАЦИИ СПАМА НА ОСНОВЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Моховиков В.А. (ВКА), Горбачев В.Д. (ВКА), Тельбух В.В. (ВКА)

Научный руководитель – кандидат технических наук, преподаватель Тельбух В. В.
(Военно-космическая академия имени А.Ф.Можайского)

Введение. Спам (англ. spam) – массовая рассылка корреспонденции лицам, не выражавшим желания её получить. Распространителей спама называют спамерами.

Спам остается одной из главных проблем цифрового мира. Согласно отчету Kaspersky Lab, в 2023 году доля спама в мировом почтовом трафике составила 45,60%, что свидетельствует о высокой активности спамеров [1]. В 2024 году сервис Яндекс 360 обработал 48 миллиардов входящих писем, из которых каждое четвертое было классифицировано как спам, что на 23% больше, чем годом ранее [2].

Традиционные методы фильтрации спама включают сигнатурный анализ и эвристические алгоритмы, однако они уступают в эффективности перед динамически изменяющимися тактиками спамеров. Современные антиспам-системы все чаще используют машинное обучение, позволяющее анализировать текстовые паттерны сообщений и выявлять спам с высокой точностью. В данной статье исследуется применение различных алгоритмов машинного обучения для классификации спама, с особым акцентом на сравнение традиционных методов (SVM) и нейросетевого подхода на основе LSTM.

Современный цифровой ландшафт характеризуется стремительным ростом киберугроз, среди которых особое место занимают спам-рассылки. Спам не только создает неудобства для пользователей, но и часто служит инструментом для распространения фишинговых атак, вредоносных программ и других видов киберпреступлений. Традиционные методы защиты, зачастую оказываются неэффективными против современных сложных угроз, которые быстро эволюционируют и адаптируются.

В связи с этим возникает необходимость в разработке новых подходов, способных эффективно противостоять спаму. Одним из перспективных направлений является использование методов машинного обучения, которые позволяют автоматически выявлять сложные закономерности в данных и адаптироваться к новым угрозам.

Основная часть. Целью данного исследования является оценка эффективности методов машинного обучения для классификации спама, с особым акцентом на использование рекуррентных нейронных сетей с долговременной краткосрочной памятью (LSTM) и сравнением их с традиционными алгоритмами, такими как наивный байесовский классификатор и метод опорных векторов (SVM).

Современные методы борьбы со спамом можно разделить на традиционные, основанные на сигнатурном анализе и эвристических алгоритмах, и методы машинного обучения, использующие анализ текстов и статистические закономерности. Традиционные подходы включают в себя сигнатурные фильтры, которые работают на основе заранее составленных списков спамерских адресов, известных шаблонов сообщений и ключевых слов. Они отличаются высокой скоростью работы и простотой реализации, однако их главный недостаток заключается в неэффективности против новых типов спама, что требует постоянного обновления базы данных. Более гибким инструментом являются эвристические алгоритмы, анализирующие структуру писем, частоту встречаемости определенных слов и фраз, длину сообщений и другие параметры. Они позволяют обнаруживать неизвестные образцы спама, но могут давать высокое число ложных срабатываний, а их настройка требует значительных усилий [5].

С развитием технологий машинного обучения появились более продвинутые методы классификации спама. Наиболее современные подходы основаны на применении глубоких нейронных сетей, таких как LSTM, CNN и BERT. В частности, архитектура LSTM разработана для обработки последовательных данных, что делает её особенно эффективной

при анализе текстов. В отличие от традиционных методов, она способна учитывать контекст и сложные зависимости между словами, что значительно повышает точность классификации. Однако использование таких моделей требует больших вычислительных ресурсов и значительных объемов данных для обучения. Несмотря на это, нейросетевые методы остаются одним из наиболее перспективных направлений в области автоматического обнаружения спама [3,4,5].

Таким образом, выдвигается гипотеза исследования: использование рекуррентных нейросетей с долговременной кратковременной памятью (LSTM) для классификации спама позволит достичь более высокой точности по сравнению с традиционными методами, такими как наивный байесовский классификатор и метод опорных векторов (SVM). Предполагается, что благодаря способности LSTM учитывать контекст слов и сложные зависимости в тексте, данная модель сможет лучше выявлять спам-сообщения, минимизируя количество ложных срабатываний и увеличивая общую эффективность классификации.

Результаты эксперимента показали, что LSTM продемонстрировала наивысшую точность классификации – более 95%, опережая как SVM, так и наивный байесовский классификатор, которые показали результаты в диапазоне 85–90%. Анализ F1-меры также подтвердил превосходство LSTM, особенно в части полноты (recall), что свидетельствует о её способности минимизировать количество пропущенного спама. Однако обучение LSTM заняло значительно больше времени по сравнению с традиционными методами: на обучение требовалось в несколько раз больше вычислительных ресурсов, а предсказание занимало больше времени, чем у SVM и наивного байесовского классификатора.

Выводы. Эксперимент подтвердил выдвинутую гипотезу: модель LSTM продемонстрировала наилучшие результаты среди всех протестированных алгоритмов, достигнув точности 96.8%. Это превосходит показатели традиционных методов, таких как метод опорных векторов (SVM) и наивный байесовский классификатор, что подтверждает эффективность глубоких нейросетей в задаче классификации спама. Однако наряду с высокой точностью LSTM требует значительно больше вычислительных ресурсов и времени на обучение, что может стать ограничивающим фактором при её практическом применении в системах реального времени.

Список использованных источников:

1. Kaspersky Lab. Отчет о спае и фишинге за 2023 год [Электронный ресурс]. – URL: (<https://securelist.ru/spam-phishing-report-2023/109104/>) (дата обращения: 25.01.2025).
2. Яндекс 360. Статистика обработки писем за 2024 год [Электронный ресурс]. – URL: (<https://habr.com/ru/news/877498/>) (дата обращения: 25.01.2025).
3. Zhang, Y., Wallace, B. A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification [Электронный ресурс] // arXiv preprint. – 2015. – arXiv:1510.03820. – URL: <https://arxiv.org/abs/1510.03820> (дата обращения: 25.01.2025).
4. Hochreiter, S., Schmidhuber, J. Long Short-Term Memory // Neural Computation. – 1997. – Vol. 9, № 8. – P. 1735–1780.
5. Cormack, G. V. Email Spam Filtering: A Systematic Review // Foundations and Trends in Information Retrieval. – 2008. – Vol. 1, № 4. – P. 335–455.