УДК 004

ПРИМЕНЕНИЕ РЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЕЙ LSTM ДЛЯ КЛАССИФИКАЦИИ ТЕКСТОВЫХ КИБЕРУГРОЗ

Ларин С.Э.

студент,

ФГБОУ ВО «Калужский государственный университет

им. К.Э. Циолковского»

Калуга, Россия

Белаш В.Ю.

к.пед.н., доцент,

ФГБОУ ВО «Калужский государственный университет

им. К.Э. Циолковского»

Калуга, Россия

Аннотация: В работе рассматривается задача автоматической классификации киберугроз на основе текстовых данных с использованием методов глубокого обучения. Предлагается подход, основанный на рекуррентной нейронной сети типа Long Short-Term Memory (LSTM), предназначенной для анализа последовательностей естественного языка. В качестве экспериментальной базы использован открытый датасет Cyber Threat Intelligence, содержащий текстовые описания киберинцидентов и соответствующие им классы угроз. Описаны этапы предварительной обработки данных, архитектура модели, процесс обучения и методы оценки качества. Экспериментальные результаты показывают, что предложенная модель достигает точности классификации 85,33 % на тестовой выборке, что подтверждает целесообразность применения LSTM для анализа текстов в задачах кибербезопасности.

Ключевые слова: кибербезопасность, анализ текстов, глубокое обучение, LSTM, классификация угроз, машинное обучение.

APPLICATION OF RECURRENT LSTM NEURAL NETWORKS FOR CLASSIFICATION OF TEXT CYBER THREATS

Larin S.E.
student,
Kaluga State University named after K. E. Tsiolkovsky
Kaluga, Russia
Belash V.Yu.
Ph.D., Associate Professor,
Kaluga State University named after K. E. Tsiolkovsky
Kaluga, Russia

Abstract: This paper examines the problem of automatically classifying cyberthreats based on text data using deep learning methods. We propose an approach based on a Long Short-Term Memory (LSTM) recurrent neural network designed for analyzing natural language sequences. The open Cyber Threat Intelligence dataset, containing text descriptions of cyber incidents and their corresponding threat classes, is used as an experimental base. The data preprocessing stages, model architecture, training process, and quality assessment methods are described. Experimental results demonstrate that the proposed model achieves 85.33% classification accuracy on the test set, confirming the feasibility of using LSTM for text analysis in cybersecurity problems.

Keywords: cybersecurity, text analysis, deep learning, LSTM, threat classification, machine learning.

Рост цифровизации и развитие сетевых технологий приводят к значительному увеличению числа и сложности кибератак. Современные инциденты информационной безопасности сопровождаются большими объёмами текстовой информации, включая журналы событий, аналитические отчёты специалистов и уведомления центров реагирования на компьютерные инциденты (Computer Emergency Response Team, CERT). Ручной анализ таких

данных является трудоёмким и плохо масштабируемым, что обуславливает необходимость автоматизации процессов выявления и классификации киберугроз.

Методы машинного обучения и, в частности, глубокого обучения демонстрируют высокую эффективность при решении задач обработки естественного языка. В отличие от традиционных сигнатурных подходов, обучаемые модели способны извлекать скрытые закономерности из данных и адаптироваться к новым типам атак. Особенно перспективными для анализа текстовых последовательностей являются рекуррентные нейронные сети, такие как LSTM, обеспечивающие учёт контекста и порядка слов.

Целью данной работы является исследование возможности применения LSTM-модели для многоклассовой классификации текстовых описаний киберугроз и оценка её эффективности на реальном открытом наборе данных.

В последние годы активно ведутся исследования, посвящённые применению нейронных сетей в задачах информационной безопасности. Работы показывают эффективность искусственных нейронных сетей при анализе журналов событий, выявлении аномалий и снижении числа ложноположительных срабатываний. Отдельное направление исследований связано с использованием рекуррентных архитектур и трансформеров для обработки текстовых данных кибераналитики. Несмотря на это, задача классификации текстовых описаний угроз остаётся актуальной и требует эмпирической оценки на разнообразных датасетах.

В качестве источника данных использован открытый датасет Cyber Threat Intelligence, содержащий текстовые описания инцидентов и метки классов угроз. Исходные данные включают неструктурированные текстовые поля, характеризующие тип атаки, используемую инфраструктуру и контекст инцидента.

Предварительная обработка данных включала следующие этапы:

1) приведение текста к нижнему регистру;

2) удаление URL, специальных символов и служебных токенов;

3) удаление стоп-слов английского языка;

4) токенизацию текстов;

5) приведение последовательностей к фиксированной длине с использованием дополнения и усечения.

Классы угроз были закодированы в числовом виде и преобразованы с использованием one-hot encoding для обучения модели многоклассовой классификации. Данные были разделены на обучающую и тестовую выборки в пропорции 80/20.

Для решения задачи классификации была выбрана рекуррентная нейронная сеть LSTM. Данный выбор обусловлен способностью архитектуры эффективно моделировать долгосрочные зависимости в текстовых последовательностях.

Архитектура модели включает:

1) слой Embedding для преобразования токенов в плотные векторные представления размерности 100;

2) LSTM-слой с 150 нейронами для извлечения контекстных признаков;

3) слой Dropout для снижения риска переобучения;

4) полносвязный выходной слой с функцией активации softmax, обеспечивающий многоклассовую классификацию.

В качестве функции потерь использовалась categorical cross-entropy, оптимизация выполнялась методом Adam. На рисунке 1 представлены графики точности и функции потерь по эпохам, демонстрирующие устойчивую сходимость модели. Видно, что на 7–8 эпохах достигается стабилизация метрик, что подтверждает эффективность применения EarlyStopping.

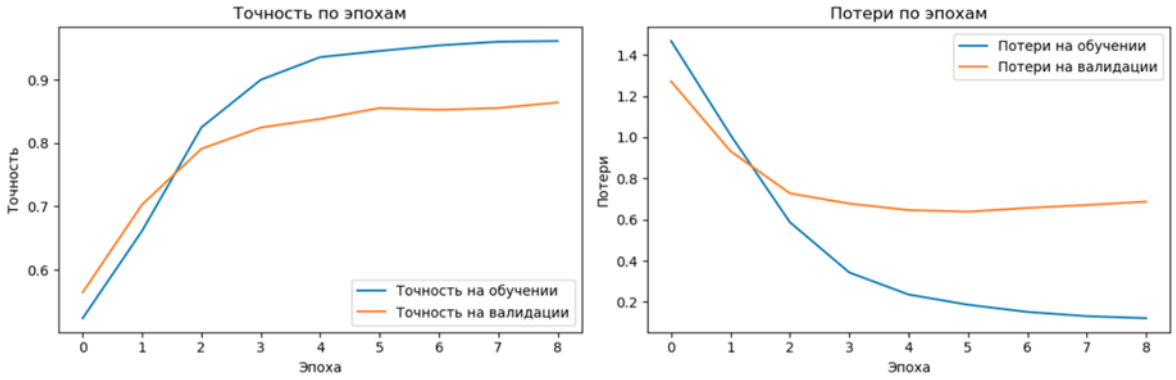

Рисунок 1 – График точности (accuracy) и функции потерь (loss) по эпохам обучения[1]

Обучение проводилось на протяжении 10 эпох с размером батча 64. Также применялась техника EarlyStopping, которая завершала обучение при отсутствии улучшений в валидационной метрике в течение 3 эпох.

По результатам тестирования была достигнута точность классификации 85,33 % при значении функции потерь 0,6541. Анализ матрицы ошибок показал, что модель демонстрирует устойчивую способность различать основные классы угроз, при этом основные ошибки приходятся на семантически близкие категории.

[1] Составлено авторами

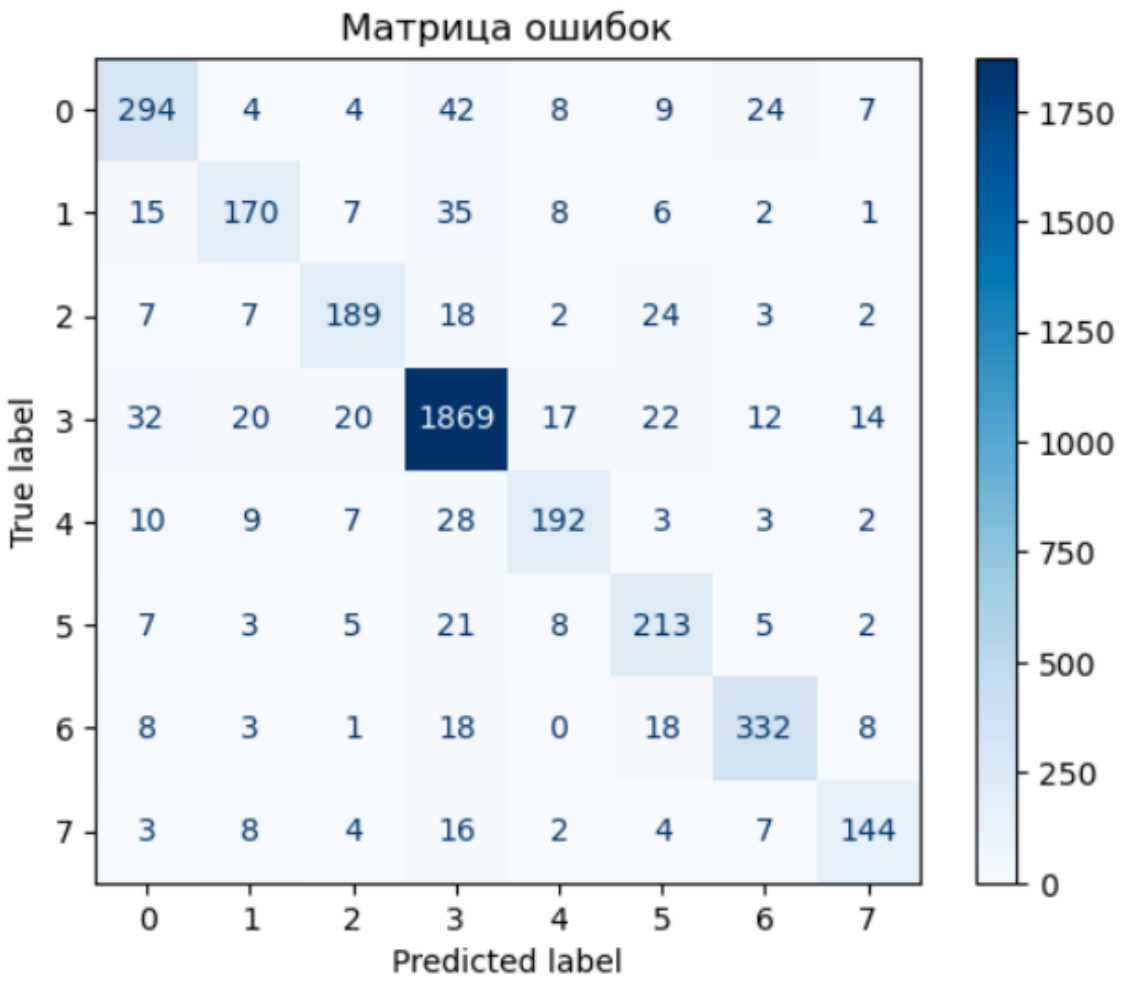

Рисунок 2 – Матрица ошибок модели на тестовой выборке[2]

Матрица ошибок, представленная на рисунке 2, позволяет выявить, какие именно классы были наиболее подвержены ошибочной классификации. Например, наблюдается частая путаница между классами X и Y.

Для качественной оценки извлекаемых признаков была выполнена визуализация скрытых представлений LSTM с использованием метода главных компонент (PCA), что подтвердило формирование различимых кластеров для различных классов угроз.

На рисунке 3 представлен Classification Report — количественная оценка точности, полноты и F1-меры по каждому классу.

[2] Составлено авторами

```
125/125 ━━━━━━━━━━━━━━━━━━━━ 6s 52ms/step
              precision    recall  f1-score   support

           0       0.78      0.75      0.77       392
           1       0.76      0.70      0.73       244
           2       0.80      0.75      0.77       252
           3       0.91      0.93      0.92      2006
           4       0.81      0.76      0.78       254
           5       0.71      0.81      0.76       264
           6       0.86      0.86      0.86       388
           7       0.80      0.77      0.78       188

    accuracy                           0.85      3988
   macro avg       0.80      0.79      0.80      3988
weighted avg       0.85      0.85      0.85      3988
```

Рисунок 3 – Отчет классификации по классам (precision, recall, F1-score)[3]

На рисунке 4 представлена PCA-визуализация признаков. Она подтверждает, что обученная модель формирует хорошо различимые кластеры для различных типов угроз.

[3] Составлено авторами

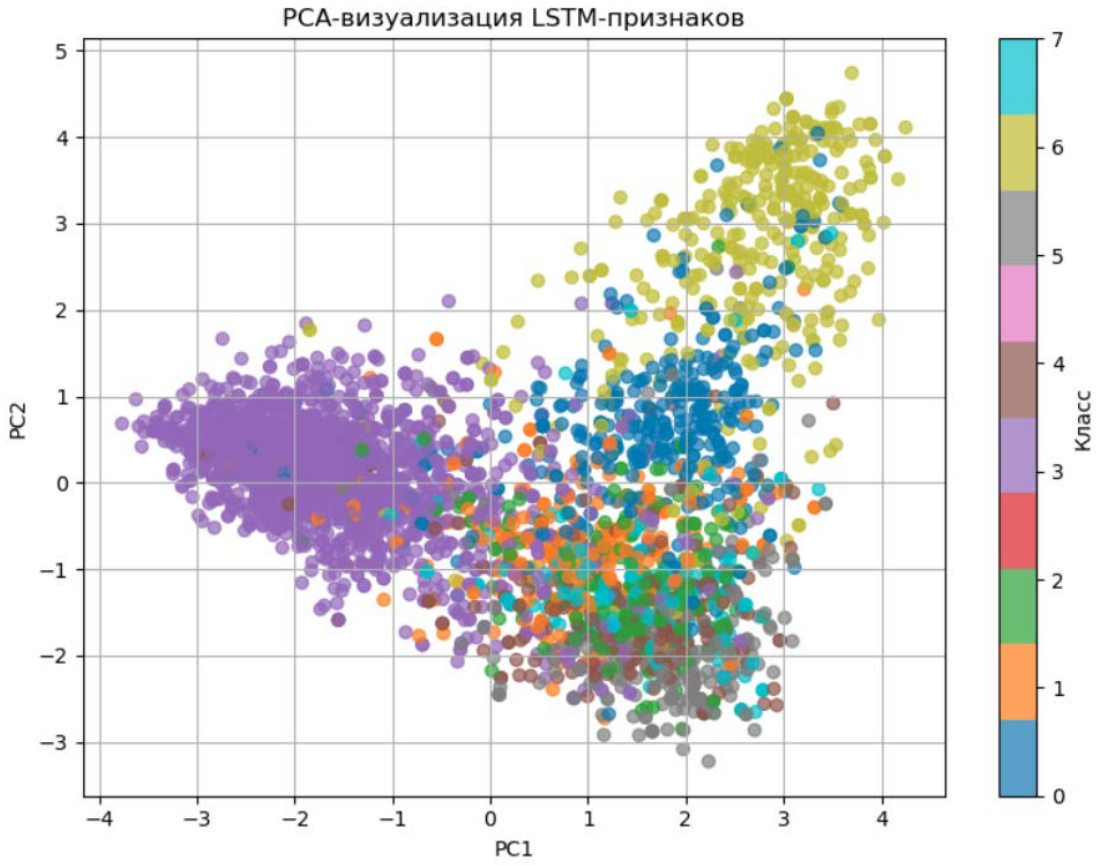

Рисунок 4 – Визуализация признаков методом PCA после обучения LSTM[4]

Полученные результаты подтверждают эффективность применения LSTM для анализа текстовой информации в задачах кибербезопасности. Несмотря на относительно простую архитектуру, модель демонстрирует конкурентоспособные показатели качества. Ограничением подхода является зависимость от объёма и качества размеченных данных, а также возможная потеря семантики при агрессивной очистке текста.

В статье представлен подход к автоматической классификации текстовых киберугроз с использованием рекуррентной нейронной сети LSTM. Проведённые эксперименты показали, что предложенная модель обладает высокой точностью и может использоваться в системах мониторинга информационной безопасности и поддержки принятия решений. В дальнейшем

[4] Составлено авторами

целесообразно исследовать применение более современных архитектур, таких как трансформеры, а также интеграцию текстовых признаков с сетевыми и временными характеристиками инцидентов.

Библиографический список

1. LSTM – сети долгой краткосрочной памяти // Хабр URL: https://habr.com/ru/companies/wunderfund/articles/331310/ (дата обращения: 21.09.2025).

2. Исхаков, АА. Выявление ложноположительных инцидентов кибербезопасности на основе искусственных нейронных сетей / А.А. Исхаков, А.З. Махмутова, И.В. Аникин // ИВД. – 2024. – №8 (116). – URL: https://cyberleninka.ru/article/n/vyyavlenie-lozhnopolozhitelnyh-intsidentov-kiberbezopasnosti-na-osnove-iskusstvennyh-neyronnyh-setey (дата обращения: 21.09.2025).

3. Аль-аммори, А. Методы и средства защиты информации / А. Аль-аммори, П.В. Дяченко, А.Е. Клочан, Е.В. Бакун, И.К. Козелецкая // The Scientific Heritage. – 2020. – №51-1. – URL: https://cyberleninka.ru/article/n/metody-i-sredstva-zaschity-informatsii (дата обращения: 21.09.2025).

4. Антонио, Д. Библиотека Keras – инструмент глубокого обучения. Реализация нейронных сетей с помощью библиотек Theano и TensorFlow / Д. Антонио, П. Суджит ; перевод с английского А. А. Слинкин. – Москва : ДМК Пресс, 2018. – 294 с.

5. Гольдберг, Й. Нейросетевые методы в обработке естественного языка : руководство / Й. Гольдберг ; перевод с английского А. А. Слинкина. – Москва : ДМК Пресс, 2019. – 282 с.

6. Джулли, А. Библиотека Keras – инструмент глубокого обучения. Реализация нейронных сетей с помощью библиотек Theano и TensorFlow / А. Джулли, С. Пал ; перевод А. А. Слинкин. – Москва : ДМК Пресс, 2018. – 294 с.

7. Искусственный интеллект в киберзащите // Хабр – URL: https://habr.com/ru/companies/pt/articles/904936/ (дата обращения: 21.09.2025).

8. Модели LSTM и GRU // Машинное и глубокое обучение – URL: https://deepmachinelearning.ru/docs/Neural-networks/Recurrent-neural-nets/LSTM-GRU (дата обращения: 21.09.2025).

9. Орлов, А. И. Искусственный интеллект: статистические методы анализа данных : учебник / А. И. Орлов. – Москва : Ай Пи Ар Медиа, 2022. – 843 с.

10. Цуканова, Н. И. Программирование глубоких нейронных сетей на языке Python : учебное пособие / Н. И. Цуканова. – Москва : КУРС, 2024. – 224 с.