ОБНАРУЖЕНИЕ ФИШИНГОВЫХ ЭЛЕКТРОННЫХ ПИСЕМ С ПОМОЩЬЮ РЕКУРРЕНТНЫХ НЕЙРОННЫХ СЕТЕЙ

Болдырихин Н. В.[1], Ядрец Э. А.[2]

DOI: 10.21681/2311-3456-2025-4-134-141

Цель исследования: *рассмотреть особенности применения рекуррентных нейронных сетей при решении задачи обнаружения фишинговых писем.*

Метод(ы) исследования: *сравнение и сопоставление, математическое и программное моделирование, системный анализ.*

Результат(ы) исследования: *рассмотрены понятие и виды фишинговых атак. Проведен анализ современных публикаций по теме использования рекуррентных нейронных сетей в задачах обнаружения фишинга, который показал, что использование рекуррентных сетей даёт возможность с большой вероятностью обнаруживать фишинговые письма. Проанализированы датасеты, имеющиеся в открытом доступе: большинство датасетов ориентированы на обнаружение фишинговых URL – адресов. Немногочисленные датасеты, ориентированные на текст электронного письма, в подавляющем большинстве являются англоязычными, качественные русскоязычные датасеты в открытом доступе отсутствуют, поэтому был собран собственный датасет из русскоязычных электронных писем. Проведено также математическое и программное моделирование различных рекуррентных нейронных сетей для выявления фишинговых писем: RNN, LSTM, BiLSTM и проведен сравнительный анализ их характеристик. Выявлены зависимости характеристик потерь и точности от числа эпох. Сравнительный анализ рекуррентных сетей показал, что наиболее эффективной в решении задач обнаружения фишинга в рамках исследований оказалась сеть BiLSTM, которая обнаружила 91,43 % фишинговых писем. Худшие характеристики показала сеть RNN, которая обнаружила только 50,71 % фишинговых писем из тестовой выборки. Следует отметить, что данные результаты получены для сетей, обучаемых на датасетах малого объёма (300 писем).*

Научная новизна: *результаты исследований позволяют аргументировано заключить, что из рассмотренных рекуррентных нейронных сетей именно BiLSTM наилучшим образом справляются с задачами выявления фишинговых писем при небольших объемах обучающего датасета.*

Ключевые слова: *кибермошенничество, защита от фишинга, рекуррентные нейронные сети RNN, LSTM, BiLSTM.*

Введение

В настоящее время задача обеспечения информационной безопасности становится очень актуальной как для государственных учреждений, коммерческих организаций, так и для физических лиц. Повсеместное внедрение цифровых технологий стало причиной небывалого всплеска киберпреступной деятельности [1–16]. Большое количество таких преступлений совершается с использованием фишинговых технологий [4–9].

Фишинг – вид кибермошенничества, нацеленный на получение доступа к конфиденциальной информации посредством побуждения пользователя перейти по ссылке на интернет-ресурс, содержащий вредоносный код. Если пользователь переходит на данный ресурс, то ему на устройство загружается вредоносный код или пользователь сам вводит логин и пароль на поддельном сайте, имитирующем, например, ресурс банка [4–9].

Фишинговые ссылки могут распространяться при помощи электронных писем, СМС, через мессенджеры и т.д.

Главная проблема состоит в том, что мошенники постоянно подстраиваются под текущую информационную повестку, модифицируя и усложняя схемы фишинговых атак.

Самой распространённой разновидностью фишинга является «массовый» или «слепой» фишинг. На электронную почту жертв, вслепую, рассылаются шаблонные варианты сообщений, ведущие на заражённые или поддельные сайты. Письма маскируются под официальные, а темы могут быть разнообразными: сообщения от службы поддержки, почтового сервиса, уведомления о закрытии, открытии, блокировке банковских счетов, извещения из государственных органов или популярных услуг и сервисов. В таких письмах обычно пользователя просят обновить свои данные или войти в аккаунт для дальнейших манипуляций. Главная особенность такого метода в том, что атакующий заранее не знает кого именно он атакует.

Самой опасной формой является целевой фишинг, направленный на конкретного человека. Этот тип фишинга представляет собой одну из ключевых

1 Болдырихин Николай Вячеславович, кандидат технических наук, доцент кафедры «Кибербезопасность информационных систем» Донского государственного технического университета, г. Ростов-на-Дону, Россия, E-mail: boldyrikhin@mail

2 Ядрец Эдуард Александрович, магистрант кафедры «Кибербезопасность информационных систем» Донского государственного технического университета, г. Ростов-на-Дону, Россия, E-mail: xperia1058@gmail.com

тактик, применяемых в практике кибершпионажа. Его отличительная черта – целенаправленность и индивидуальный подход к реализации. Изучается сам человек и его окружение. Собирается различная персональная информация: фотографии, стилистика переписки, информация о местах отдыха, предпочтениях, увлечениях и т.д. Подобные атаки нацелены на известных медийных персон, бизнесменов и политических деятелей.

Процесс распознавания фишингового письма весьма затруднительный. Для этого применяются антивирусные программы, антифишинговые расширения браузеров, спам-фильтры, системы обнаружения вторжений, DLP-системы (Data loss prevention) и т.д., которые используют как сигнатурные методы, так и методы, основанные на применении искусственного интеллекта [4–14].

Искусственный интеллект на основе нейросетей в настоящее время широко применяется в решении задач информационной безопасности. Это связано с тем, что системы, основанные на машинном обучении, могут выявлять не только известные атаки, но и атаки «нулевого дня», при этом показывая хорошие результаты. Хотя конечно такие системы не являются панацеей, они, тем не менее, набирают большую популярность. Поэтому исследования в данном направлении представляют теоретический и практический интерес.

В рамках данной статьи проведено исследование особенностей применения рекуррентных нейронных сетей при обнаружении фишинговых электронных писем.

Особенности применения рекуррентных нейронных сетей различных типов при решении задач обнаружения фишинга

В работах [4–9] подробно рассмотрены самые разнообразные методы решения задач обнаружения фишинга, в том числе применение нейронных сетей. На основе анализа данных работ выявлено, что наиболее часто для таких задач применяются рекуррентные (Recurrent neural network, RNN) и свёрточные нейронные сети (Convolutional neural network, CNN).

В работе [5] авторы Корнюхина С. П., Лапонина О. Р. подробно рассмотрели решение задачи обнаружения фишинга различными нейронными сетями, в том числе CNN и LSTM (Long short-term memory). Экспериментальные результаты показали, что LSTM в случаях обнаружения фишинговых URL-адресов, демонстрирует точность свыше 99 %, а при обнаружении фишинговых писем – 98 %. Сеть CNN показывает максимальные результаты при обнаружении фишинговых URL-адресов – свыше 98 %.

В научной статье [6] авторами Suleiman Y. Yerima и Mohammed K. Alzaylaee рассмотрен подход применения CNN для высокоточной классификации фишинговых сайтов. Основываясь на результатах обширных экспериментов, авторы пришли к выводу, что модели на основе CNN оказались очень эффективными в обнаружении неизвестных фишинговых сайтов. Более того, подход на основе CNN показал лучшие результаты, чем традиционные классификаторы машинного обучения, достигнув 98,2 % обнаружения фишинга.

В работе [7] авторами Weiping Wang, Feng Zhang, Xi Luo и Shigeng Zhang представлен быстрый подход к обнаружению фишинговых сайтов под названием Precise Phishing Detection with Recurrent Convolutional Neural Networks (PDRCNN), который опирается только на URL сайта. PDRCNN не нужно извлекать содержимое целевого сайта и использовать сторонние сервисы. Представленный метод кодирует информацию об URL-адресе в двумерный тензор и передаёт этот тензор в нейронную сеть глубокого обучения для классификации исходного URL-адреса. Авторы работы [7] используют двунаправленную LSTM-сеть (Bidirectional Long Short-Term Memory, BiLSTM) для извлечения глобальных признаков из построенного тензора и передают всю строковую информацию каждому символу в URL. После этого используют CNN для автоматического определения того, какие символы играют ключевую роль, захватывают ключевые компоненты URL и сжимают извлечённые характеристики в векторное пространство фиксированной длины. Комбинируя эти два типа сетей, PDRCNN достигает более высокой производительности, чем при использовании сетей по отдельности.

В работе [8] автором Cagatay Catal и др., приведён систематический обзор использования алгоритмов глубокого обучения для обнаружения фишинга.

Авторы работы [8] провели систематический обзор литературы, с целью обобщить результаты применения подходов глубокого обучения для обнаружения фишинга, представленные в отобранных научных публикациях. Были рассмотрены девять исследовательских вопросов и представлен обзор того, как применялись алгоритмы глубокого обучения для обнаружения фишинга. Во всех моделях применялись алгоритмы глубокого обучения с супервизией. В качестве источников использовались наборы данных, связанные с URL, информацией о третьих лицах на сайте, содержимое сайта и электронная почта.

Наиболее часто применялись различные алгоритмы глубокого обучения. Среди них выделяются нейронные сети, известные как глубокие нейронные сети (Deep Neural Network, DNN). DNN и гибридные алгоритмы глубокого обучения обеспечили наилучшую производительность среди других алгоритмов глубокого обучения. В 72 % исследований для построения модели прогнозирования не применялся какой-либо алгоритм отбора признаков. PhishTank был наиболее используемым набором данных среди

других наборов данных. Несмотря на то, что Keras и Tensorflow были наиболее предпочтительными фреймворками для глубокого обучения, в 46 % статей, исследуемых авторами, не упоминался ни один фреймворк [8].

Таким образом, можно утверждать, что рекуррентные сети очень популярный вид сетей при решении задач обнаружения фишинга.

В рамках данного исследования рассмотрены особенности применения различных типов рекуррентных нейронных сетей для выявления фишинговых писем: RNN, LSTM, BiLSTM.

Описание наборов данных

Задача подготовки наборов данных является в целом нетривиальной, поскольку за качественными наборами данных (датасетами) идёт настоящая охота. Компании-разработчики крайне редко раскрывают свои качественные датасеты, хотя алгоритмы публикуют довольно охотно. Это связано с тем, что сбор данных зачастую является сложным, трудоемким и долгим. Так же при подготовке датасетов следует учитывать необходимость очистки данных от выбросов и пропусков, необходимость нормализации и отбора признаков [5], что также сопряжено с существенными затратами.

При исследовании готовых датасетов, находящихся в открытом доступе на порталах VirusTotal, PhishTank OpenPhish, оказалось, они в основном содержат данные о URL-адресах.

Поскольку изначально ставилась задача определения фишинга не только по URL-адресам, содержащимся в письмах, но и по тексту самого письма, то указанные датасеты не подошли для проведения исследования.

На портале GitHub и Kaggle размещены датасеты фишинговых электронных писем, однако они англоязычные и по этой причине тоже не подошли, ввиду того, в рамках данного исследования интерес представляли именно русскоязычные фишинговые письма, поскольку анализировались не только ссылки, содержащиеся в письме, но и сам текст письма.

Для создания обучающего набора данных был разработан скрипт взаимодействия с сервером электронной почты по протоколу IMAP. Скрипт способен работать с различными кодировками текста, такими как UTF-8 и Windows-1252. Скрипт подключается к почтовому серверу перебирает электронные письма на сервере, извлекая из них данные и сохраняет их. Перед сохранением происходит декодирование заголовков и тела письма электронной почты, уделяя особое внимание теме письма и текстовому содержимому тела. Это важная деталь для выявления потенциальных фишинговых писем, поскольку эти поля часто содержат вводящее в заблуждение или вредоносное содержимое. В результате работы данного скрипта, из личного электронного почтового ящика был выполнен сбор данных, ставший основой для формирования обучающего набора данных, включающего в себя 300 писем, 120 из которых не относились фишингу, и 180 писем являлись фишинговыми. Особенностью собранного набора данных является его структурирование, где каждое письмо сохранено в виде отдельного текстового файла, обеспечивая таким образом удобство для последующего анализа и обработки информации.

Дополнительно была создана проверочная выборка из 140 фишинговых писем, которая не участвовала в обучении и валидации модели, но имеет схожие стилистические паттерны текста и ключевые фразы для проверки работоспособности выбранных нейросетевых моделей.

Эффективность LSTM, как и других нейронных сетей в задачах обработки текста в значительной степени зависит от качества входных данных, оптимальная подготовка которых, существенно повышает способность сети к обучению и обобщению. В рамках исследования для подготовки данных использовалась SpaCy. Библиотека SpaCy — является мощным инструментом, позволяющим решать такие задачи подготовки данных, как лемматизация, токенизация, анализ зависимостей и множество других задач.

В качестве предобработки применялась токенизация [15], а затем лемматизация [16] слов, на основе средней русскоязычной модели SpaCy «ru_core_news_md» для выделения лемм алфавитных токенов.

Модель «ru_core_news_md» в SpaCy обеспечивает высокую точность токенизации и лемматизации. Эта модель обучена на большом объёме текстовых данных, позволяя ей эффективно справляться с разнообразными текстовыми задачами и даёт преимущество в точности и надёжности предобработки данных. На рисунке 1 приведены свойства языковой модели, представленные на официальном сайте https://spacy.io/.

Рис. 1. Свойства языковой модели

```
nlp = spacy.load("ru_core_news_md")

def tokenize_text(text):
    doc = nlp(text)
    return [token.lemma_ for token in doc if token.is_alpha]

def vectorize_text(tokens):
    vectors = [nlp.vocab[token].vector for token in tokens if nlp.vocab[token].has_vector]
    if vectors:
        vectors = np.array(vectors)
        return torch.tensor(vectors, dtype=torch.float32)
    else:
        return torch.empty((0, nlp.vocab.vectors_length), dtype=torch.float32)
```

Рис. 2. Код для предобработки текстовых данных

Токенизация – это процесс разделения текста на составляющие его элементы, называемые токенами. В контексте текста токены обычно представляют собой слова, но также могут включать пунктуацию и другие символы. Цель токенизации – преобразовать непрерывный текст в управляемый набор данных, которые далее могут быть обработаны и проанализированы.

Лемматизация – это процесс приведения слова к его базовой форме (лемме). Лемматизация учитывает морфологический анализ слов, чтобы вернуть правильную базовую форму слова с учётом его использования в тексте.

Реализация функций предварительной обработки текста с помощью библиотеки SpaCy представлена на рисунке 2.

Лемматизированные токены затем векторизуются с использованием предварительно обученных векторов слов, доступных в модели SpaCy. Учитываются только те токены, для которых в словаре SpaCy имеются векторы, каждый токен преобразуется в вектор с использованием соответствующего вложения. Затем векторы преобразуются в тензоры, обеспечивая совместимость с моделью LSTM или любыми другими.

Подготовленные векторы и соответствующие метки сохраняются, затем данные в рамках своего класса случайным образом перемешиваются и разделяются на обучающую и тестовую выборки в соответствии с заданным соотношением: 80 % на обучение и 20 % для валидации модели. В результате предобработки получились хорошо структурированные и нормализованные входные данные для обучения нейросети.

Характеристики обучения выбранных топологий нейросетей

Следует отметить, что задача обнаружения фишинга относится к задаче бинарной классификации [6]. В этом случае матрица ошибок имеет следующий вид (см. рис. 3).

В качестве показателей качества работы нейронной сети использовались бинарная кросс-энтропийная функция потерь (Loss)

Реальное событие / Результат работы нейросети	**Фишинг присутствует (Positive)**	**Фишинг отсутствует (Negative)**
Фишинг присутствует (Positive)	True Positive (TP) — правильное определение фишинга	FP (False Positive) — ошибочное определение фишинга, хотя фактически его не было
Фишинг отсутствует (Negative)	FN (False Negative) — когда письмо определено как безопасное, хотя фактически оно фишинговое	True Negative (TN) — правильное определение отсутствия фишинга

Рис. 3. Матрица ошибок

$$\text{Loss} = P_p * \ln(P_{np}) + (1 - P_p) * \ln(1 - P_{np}), \quad (1)$$

и точность (Accuracy),

$$\text{Accuracy} = \frac{TP + TR}{TN + FP + FN + TP}, \quad (2)$$

где P_p – вероятность наличия фишинга, P_{np} – вероятность отсутствия фишинга; TP, TR, TN, FP, FN – переменные, обозначенные на рисунке 3.

Далее приведены характеристики обучения для различных топологий нейросетей.

Стандартная рекуррентная нейронная сеть (RNN) обладает простой архитектурой. На рисунке 4 представлены основные параметры модели, а также проиллюстрирован процесс её обучения.

```
Модель:  RNNClassifier(
  (rnn): RNN(300, 256, num_layers=2, batch_first=True)
  (fc): Linear(in_features=256, out_features=2, bias=True)
)
Начало обучения
Epoch 1, Train Loss: 0.6933, Train Acc: 60.00%, Val Loss: 0.6325, Val Acc: 63.33%
Epoch 2, Train Loss: 0.6753, Train Acc: 60.37%, Val Loss: 0.6279, Val Acc: 56.67%
Epoch 3, Train Loss: 0.6593, Train Acc: 62.22%, Val Loss: 0.7825, Val Acc: 56.67%
Epoch 4, Train Loss: 0.6483, Train Acc: 61.85%, Val Loss: 0.6949, Val Acc: 56.67%
Epoch 5, Train Loss: 0.6498, Train Acc: 61.48%, Val Loss: 0.8026, Val Acc: 56.67%
Epoch 6, Train Loss: 0.6861, Train Acc: 61.11%, Val Loss: 0.7081, Val Acc: 56.67%
Epoch 7, Train Loss: 0.6617, Train Acc: 61.11%, Val Loss: 0.7368, Val Acc: 56.67%
Epoch 8, Train Loss: 0.6508, Train Acc: 61.85%, Val Loss: 0.6735, Val Acc: 56.67%
Epoch 9, Train Loss: 0.6833, Train Acc: 61.11%, Val Loss: 0.6809, Val Acc: 56.67%
Epoch 10, Train Loss: 0.6613, Train Acc: 61.85%, Val Loss: 0.6067, Val Acc: 63.33%
```

Рис. 4. Демонстрация процесса обучения стандартной RNN

На рисунке 5 приведена динамика изменения потерь на обучающей и валидационной выборках в ходе обучения модели. График наглядно демонстрирует, как изменяются значения потерь по мере увеличения числа эпох.

Рис. 5. Изменение потерь в процессе обучения RNN

Рисунок 6 отображает изменение точности модели с каждой новой эпохой, иллюстрируя эти изменения как для обучающих, так и для валидационных данных.

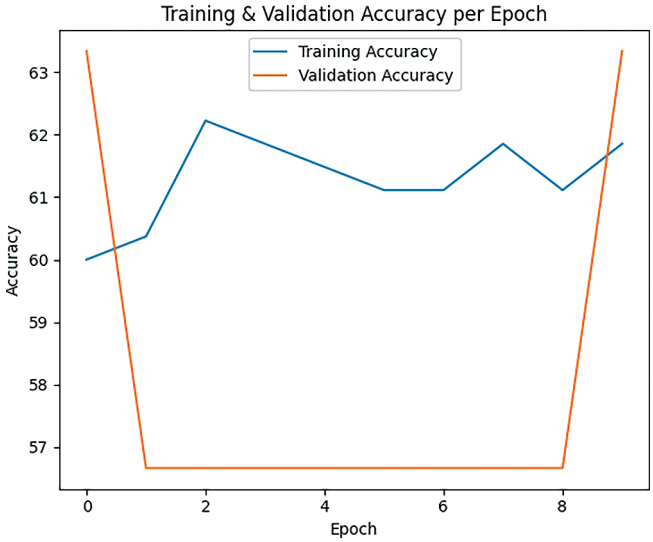

Рис. 6. Изменение точности RNN в процессе обучения

Валидационная точность остаётся на уровне 56,67 % на протяжении большинства эпох, она ниже, чем точность на обучающей выборке, такое поведение модели вероятно является следствием проблемы с обобщающей способностью.

Долгосрочная кратковременная память (LSTM) отличается более сложной архитектурой по сравнению с стандартной рекуррентной сетью (RNN). На рисунке 7 демонстрируются ключевые параметры модели и показан процесс её обучения.

На рисунке 8 показана динамика изменения потерь на обучающей и валидационной выборках

```
Модель:  LSTMClassifier(
  (lstm): LSTM(300, 256, num_layers=2, batch_first=True)
  (fc): Linear(in_features=256, out_features=2, bias=True)
)
Начало обучения
Epoch 1, Train Loss: 0.6991, Train Acc: 47.08%, Val Loss: 0.6799, Val Acc: 63.33%
Epoch 2, Train Loss: 0.6737, Train Acc: 61.25%, Val Loss: 0.6679, Val Acc: 63.33%
Epoch 3, Train Loss: 0.6639, Train Acc: 61.67%, Val Loss: 0.6604, Val Acc: 63.33%
Epoch 4, Train Loss: 0.6646, Train Acc: 61.67%, Val Loss: 0.6751, Val Acc: 61.67%
Epoch 5, Train Loss: 0.6600, Train Acc: 62.08%, Val Loss: 0.6686, Val Acc: 61.67%
Epoch 6, Train Loss: 0.6547, Train Acc: 62.50%, Val Loss: 0.6972, Val Acc: 61.67%
Epoch 7, Train Loss: 0.6760, Train Acc: 59.58%, Val Loss: 0.6741, Val Acc: 61.67%
Epoch 8, Train Loss: 0.6468, Train Acc: 63.33%, Val Loss: 0.6718, Val Acc: 61.67%
Epoch 9, Train Loss: 0.6521, Train Acc: 61.67%, Val Loss: 0.6685, Val Acc: 61.67%
Epoch 10, Train Loss: 0.6508, Train Acc: 62.08%, Val Loss: 0.6569, Val Acc: 61.67%
```

Рис. 7. Демонстрация хода обучения LSTM

в ходе обучения модели LSTM. Важно отметить, что кривая потерь на обучающей выборке показывает тенденцию к снижению, поскольку модель постепенно адаптируется к данным, минимизируя ошибку предсказания.

Результаты в виде графика точности модели изображены на рисунке 9. Модель LSTM показывает

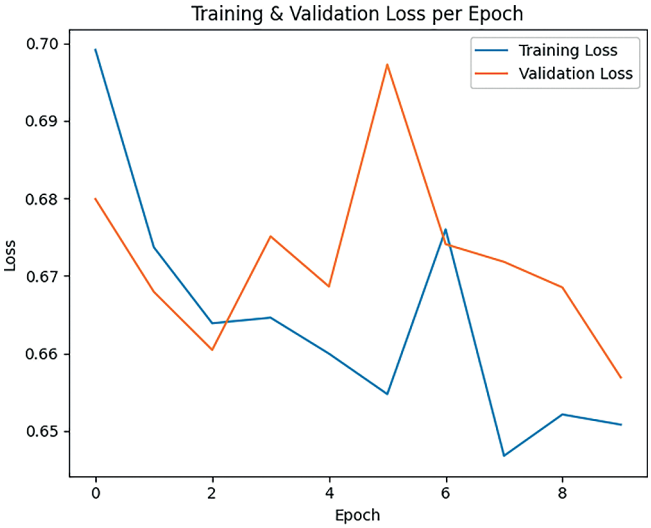

Рис. 8. Потери LSTM при обучении

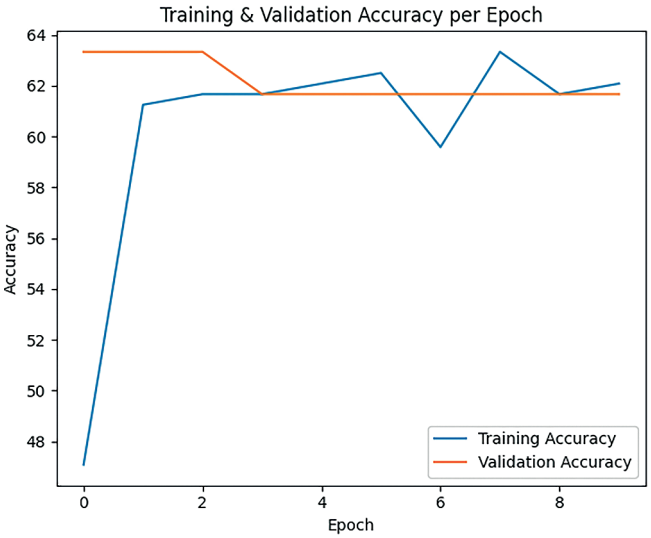

Рис. 9. Точность LSTM по пройденным эпохам

```
Модель:  BiLSTMClassifier(
  (lstm): LSTM(300, 256, num_layers=2, batch_first=True, bidirectional=True)
  (fc): Linear(in_features=512, out_features=2, bias=True)
)
Начало обучения
Epoch 1, Train Loss: 0.6611, Train Acc: 62.22%, Val Loss: 0.6444, Val Acc: 60.00%
Epoch 2, Train Loss: 0.5659, Train Acc: 71.85%, Val Loss: 0.5580, Val Acc: 66.67%
Epoch 3, Train Loss: 0.4113, Train Acc: 80.37%, Val Loss: 0.4133, Val Acc: 73.33%
Epoch 4, Train Loss: 0.2391, Train Acc: 90.74%, Val Loss: 0.3616, Val Acc: 83.33%
Epoch 5, Train Loss: 0.1738, Train Acc: 94.44%, Val Loss: 0.2044, Val Acc: 90.00%
Epoch 6, Train Loss: 0.1346, Train Acc: 94.81%, Val Loss: 0.1653, Val Acc: 96.67%
Epoch 7, Train Loss: 0.0636, Train Acc: 98.89%, Val Loss: 0.3152, Val Acc: 90.00%
Epoch 8, Train Loss: 0.0728, Train Acc: 97.04%, Val Loss: 0.1137, Val Acc: 96.67%
Epoch 9, Train Loss: 0.0463, Train Acc: 98.52%, Val Loss: 0.1826, Val Acc: 90.00%
Epoch 10, Train Loss: 0.0326, Train Acc: 99.12%, Val Loss: 0.1126, Val Acc: 96.67%
```

Рис. 10. Параметры BiLSTM модели и процесс обучения

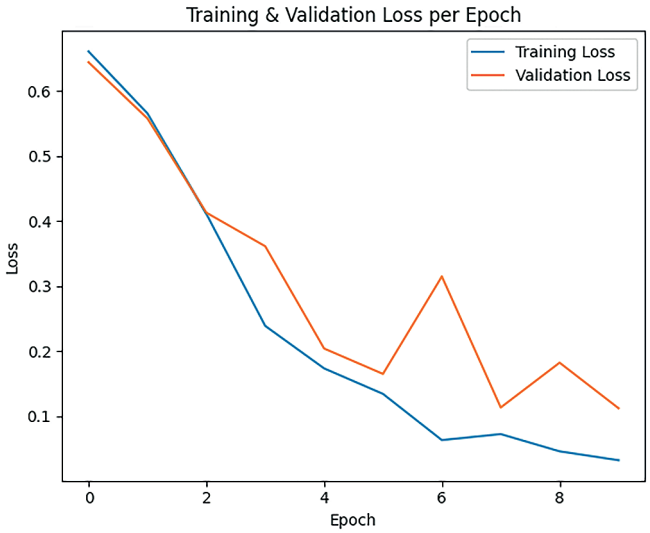

Рис. 11. Потери BiLSTM при обучении

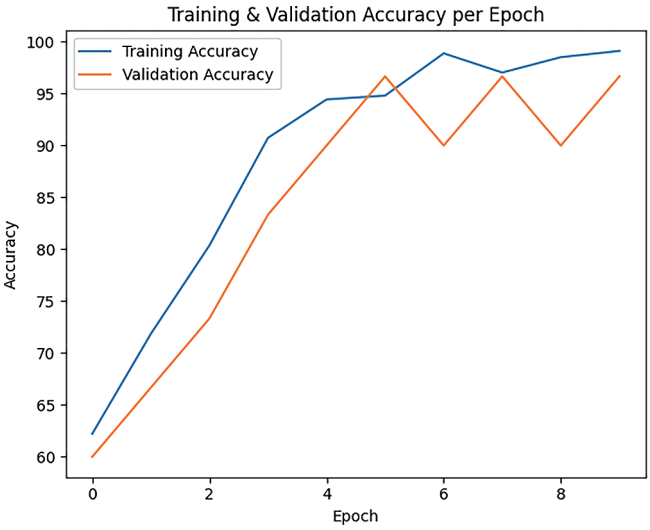

Рис. 12. Точность BiLSTM модели при обучении

средние результаты точности на тренировочных данных и на валидационном наборе она остаётся практически неизменной (около 62 %). Модель явно сталкивается с трудностями в обобщении информации и эффективной обработке длинных текстов, так как важные зависимости между словами могут находиться далеко друг от друга.

Двунаправленная модель LSTM или Bidirectional LSTM (BiLSTM) отличается от стандартной LSTM своей архитектурой, которая позволяет анализировать последовательность данных в двух направлениях вперёд и назад. На рисунке 10 представлены ключевые параметры модели BiLSTM, а также отображён процесс её обучения.

График потерь модели при обучении представлен на рисунке 11.

График точности BiLSTM модели изображён на рисунке 12.

Изображения графиков, доказывают высокую эффективность модели BiLSTM в обработке и понимании обучающего набора данных, поскольку обработка данных проводится в двух направлениях, как было сказано ранее. Достоинство модели состоит в понимании контекста и более точной обработке входных данных, в отличие от простой LSTM.

Тестирование нейросетей на основе дополнительной выборки

На рисунке 13 приведён результат обработки 140 фишинговых писем RNN моделью. Из них модель неверно определила 69 писем как легитимные, что соответствует 49 % от общего количества писем.

```
Файл: ph_9_8.txt - Результат: Легитимное - Вероятности: 0.70, 0.30
Файл: ph_9_9.txt - Результат: Легитимное - Вероятности: 0.71, 0.29
RNNClassifier(
  (rnn): RNN(300, 256, num_layers=2, batch_first=True)
  (fc): Linear(in_features=256, out_features=2, bias=True)
)
Общее количество фишинговых писем: 140
Определены как Легитимные письма: 69 49.29%
Определены как Фишинговые письма: 71 50.71%
```

Рис. 13. Результат обработки RNN

Результат обработки LSTM моделью представлен на рисунке 14.

```
Файл: ph_9_8.txt - Результат: Легитимное - Вероятности: 0.64, 0.36
Файл: ph_9_9.txt - Результат: Легитимное - Вероятности: 0.62, 0.38
LSTMClassifier(
  (lstm): LSTM(300, 256, num_layers=2, batch_first=True)
  (fc): Linear(in_features=256, out_features=2, bias=True)
)
Общее количество фишинговых писем: 140
Определены как Легитимные письма: 36 25.71%
Определены как Фишинговые письма: 104 74.29%
```

Рис. 14. Результаты обработки LSTM

Исходя из результата обработки, можно заметить, что 36 писем были неверно классифицированы, а это 25% от общего их количества. Данный результат намного лучше, чем у RNN сети, но все же далёк от идеала.

Результаты обработки для модели BiLSTM приведены на рисунке 15.

```
Файл: ph_9_8.txt - Результат: Фишинг - Вероятности: 0.00, 1.00
Файл: ph_9_9.txt - Результат: Фишинг - Вероятности: 0.10, 0.90
BiLSTMClassifier(
  (lstm): LSTM(300, 256, num_layers=2, batch_first=True, bidirectional=True)
  (fc): Linear(in_features=512, out_features=2, bias=True)
)
Общее количество фишинговых писем: 140
Определены как Легитимные письма: 12 8.57%
Определены как Фишинговые письма: 128 91.43%
```

Рис.15. Результаты обработки для BiLSTM

Здесь результаты выглядят заметно лучше, по сравнению с ранее рассматриваемыми сетями: всего 12 писем из 140 были определены как легитимные, что составило приблизительно 8 % от их общего количества.

Выводы

В результате проведённого исследования была достигнута основная цель данной статьи: рассмотрены особенности применения рекуррентных нейронных сетей при решении задачи обнаружения фишинговых писем.

В процессе работы проведен анализ предметной области, который показал, что применение рекуррентных нейронных сетей значительно повышает точность обнаружения фишинга.

В рамках данного исследования авторами произведено формирование собственного датасета на основе собранных фишинговых электронных писем на русском языке. Это связано с отсутствием в открытом доступе качественных русскоязычных датасетов, содержащих тексты фишинговых писем.

В рамках разработки нейросетей были рассмотрены этапы предобработки входных данных, а также выполнен полный цикл обучения и валидации.

В результате сравнения характеристик обучения выбранных рекуррентных нейронных сетей, можно аргументировано заключить, что наиболее эффективной для решения задачи выявления фишинговых писем оказалась BiLSTM. Созданная сеть BiLSTM продемонстрировала достаточно высокую точность в обнаружении фишинговых сообщений: 91,43 % несмотря на то, что объем используемого датасета небольшой. Остальные рекуррентные сети в этих условиях показали худший результат.

Следует отметить, что результаты, полученные в рамках данного исследования, являются в основном практическими и справедливы только в конкретных условиях данного программного моделирования. При больших объемах датасета различие в точности обнаружения фишинга между BiLSTM и другими рекуррентными сетями скорее всего будет не столь значительным.

Литература

1. Кострикина А. О., Лазунин К. А. Информационная безопасность в критической информационной безопасности // Проблемы научной мысли. 2024. Т. 4. № 1. С. 82–85.
2. Чапис М. А. Информационная безопасность государства как правовой порядок обеспечения национальной безопасности в информационной сфере // Наукосфера. 2024. № 6(1). С. 551–557. DOI: 10.5281/zenodo.11638587.
3. Добродеев А. Ю. Показатели информационной безопасности как характеристика (мера) соответствия сетей и организаций связи требованиям информационной безопасности // Труды ЦНИИС. Санкт-Петербургский филиал. 2020. Т. 2. № 10. С. 50–78.
4. Лукманова К. А., Картак В. М. Разработка системы защиты от фишинговых атак с использованием программно-аппаратной реализации методов машинного обучения // Моделирование, оптимизация и информационные технологии. 2024. 12(4). DOI: 10.26102/2310-6018/2024.47.4.033.
5. Корнюхина С. П., Лапонина О. Р. Исследование возможностей алгоритмов глубокого обучения для защиты от фишинговых атак // International Journal of Open Information Technologies. 2023. Т. 11. № 6. С. 163–174.
6. Yerima S. Y., Alzaylaee M. K. High accuracy phishing detection based on convolutional neural networks // 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS). IEEE, 2020. С. 1–6. DOI:10.1109/ICCAIS48893.2020.9096869.
7. Wang W. et al. PDRCNN: Precise phishing detection with recurrent convolutional neural networks //Security and Communication Networks. 2019. Т. 2019. С. 1–15. DOI:10.1155/2019/2595794.
8. Catal C. et al. Applications of deep learning for phishing detection: a systematic literature review // Knowledge and Information Systems. 2022. Т. 64. №. 6. С. 1457–1500. DOI:10.1007/s10115-022-01672-x.
9. Dhanavanthini P., Chakkravarthy S. S. Phish-armour: phishing detection using deep recurrent neural networks. Soft Comput. 2023. DOI: 10.1007/s00500-023-07962-y.
10. Филимонов А. В., Осипов А. В., Плешакова Е. С., Гатауллин С. Т. Нейросетевые методы распознавания эмоций речи для противодействия мошенничеству в телекоммуникационных системах // Вопросы кибербезопасности. 2022. № 6(52). С. 83–92. DOI:10.21681/2311-3456-2022-6-83-92.
11. Технологии искусственного интеллекта и кибербезопасность: монография / А. Б. Менисов. – М: Ай Пи Ар Медиа, 2022. 133 с.
12. Применение искусственного интеллекта для решения задачи обеспечения безопасности информации, передаваемой в сетях / В. И. Юхнов, А. И. Сосновский, Н. В. Болдырихин, И. А. Сосновский // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2023. № 2. С. 26–28.
13. Бимолдина Ж. А. Как искусственный интеллект меняет правила игры в кибербезопасности // Форум. Серия: Роль науки и образования в современном информационном обществе. 2024. № S2(32). С. 235–240.
14. Букин А. В., Самонов А. В., Тихонов Э. И. Обнаружение инцидентов информационной безопасности на основе технологии нейронных сетей // Вопросы кибербезопасности. 2022. № 5(51). С. 61–73. DOI: 10.21681/2311-3456-2022-5-61-73.
15. Карпенко М. П. Токенизация как метод количественного измерения информации и знаний в учебных текстах профессионального образования // Инновации в образовании. 2025. № 3. С. 40–50.
16. Обработка естественного языка в действии / Л. Хобсон, Х. Ханнес, Х. Коул / перевод с английского И. Пальти, С. Черников. – СПб: Питер, 2020. 575 с.

DETECTION OF PHISHING EMAILS USING RECURRENT NEURAL NETWORKS

Boldyrikhin N. V.[3], Yadrets E. A.[4]

Keywords: *cyberbullying, phishing protection, recurrent neural networks RNN, LSTM, BiLSTM.*

Purpose of the study: *to consider the features of the use of recurrent neural networks in solving the problem of detecting phishing emails.*

Methods of research: *comparison, mathematical and software modeling, system analysis.*

Result(s): *The concept and types of phishing attacks are considered. The analysis of modern publications on the use of recurrent neural networks in phishing detection tasks has been carried out, which has shown that the use of recurrent networks makes it possible to detect phishing emails with high probability. Publicly available datasets have been analyzed: most datasets are focused on detecting phishing URLs. The few datasets focused on the text of an email are overwhelmingly in English, and high-quality Russian-language datasets are not publicly available, so our own dataset of Russian-language emails was compiled. Mathematical and software modeling of various recurrent neural networks for detecting phishing emails has also been carried out: RNN, LSTM, BiLSTM and a comparative analysis of their characteristics has been carried out. The dependences of loss characteristics and accuracy on the number of epochs are revealed. A comparative analysis of recurrent networks has shown that the BiLSTM network, which detected 91.43 % of phishing emails, was the most effective in solving phishing detection problems in the framework of research. The RNN network showed the worst characteristics, which detected only 50.71 % of phishing emails from the test sample. It should be noted that these results were obtained for networks trained on small-volume datasets (300 emails).*

Scientific novelty: *the research results allow us to reasonably conclude that of the considered recurrent neural networks, BiLSTM is the one that best copes with the tasks of detecting phishing emails with small amounts of training dataset.*

References

1. Kostrikina A. O., Lazunin K.A. Informacionnaya bezopasnost' v kriticheskoj informacionnoj bezopasnosti // Problemy nauchnoj mysli. 2024. Vol. 4. № 1. pp. 82–85.
2. Chapis M. A. Informacionnaya bezopasnost' gosudarstva kak pravovoj poryadok obespecheniya nacional'noj bezopasnosti v informacionnoj sfere // Naukosfera. 2024. № 6(1). pp. 551–557. DOI: 10.5281/zenodo.11638587.
3. Dobrodeev A. Yu. Pokazateli informacionnoj bezopasnosti kak xarakteristika (mera) sootvetstviya setej i organizacij svyazi trebovaniyam informacionnoj bezopasnosti // Trudy CNIIS. Sankt-Peterburgskij filial. 2020. Vol. 2. № 10. pp. 50–78.
4. Lukmanova K. A., Kartak V. M. Razrabotka sistemy zashhity ot fishingovyx atak s ispol'zovaniem programmno-apparatnoj realizacii metodov mashinnogo obucheniya // Modelirovanie, optimizaciya i informacionnye texnologii. 2024. 12(4). DOI: 10.26102/2310-6018/2024.47.4.033.
5. Kornyuxina S. P., Laponina O. R. Issledovanie vozmozhnostej algoritmov glubokogo obucheniya dlya zashhity ot fishingovyx atak // International Journal of Open Information Technologies. 2023. Vol. 11. № 6. pp. 163–174.
6. Yerima S. Y., Alzaylaee M. K. High accuracy phishing detection based on convolutional neural networks // 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS). IEEE, 2020. pp. 1–6. DOI:10.1109/ICCAIS48893.2020.9096869.
7. Wang W. et al. PDRCNN: Precise phishing detection with recurrent convolutional neural networks // Security and Communication Networks. 2019. Vol. 2019. pp. 1–15. DOI:10.1155/2019/2595794.
8. Catal C. et al. Applications of deep learning for phishing detection: a systematic literature review // Knowledge and Information Systems. 2022. Vol. 64. № 6. pp. 1457–1500. DOI:10.1007/s10115-022-01672-x.
9. Dhanavanthini P., Chakkravarthy S. S. Phish-armour: phishing detection using deep recurrent neural networks. Soft Comput (2023). DOI: 10.1007/s00500-023-07962-y.
10. Filimonov A. V., Osipov A. V., Pleshakova E. S., Gataullin S. T. Nejrosetevye metody raspoznavaniya emocij rechi dlya protivodejstviya moshennichestvu v telekommunikacionnyx sistemax // Voprosy kiberbezopasnosti [Cybersecurity issues]. 2022. № 6(52). pp. 83–92. DOI:10.21681/2311-3456-2022-6-83-92.
11. Texnologii iskusstvennogo intellekta i kiberbezopasnost': monografiya / A. B. Menisov. – M: Aj Pi Ar Media, 2022. 133 p.
12. Primenenie iskusstvennogo intellekta dlya resheniya zadachi obespecheniya bezopasnosti informacii, peredavaemoj v setyax / V. I. Yuxnov, A. I. Sosnovskij, N. V. Boldyrixin, I. A. Sosnovskij // Trudy Severo-Kavkazskogo filiala Moskovskogo texnicheskogo universiteta svyazi i informatiki. 2023. № 2. pp. 26–28.
13. Bimoldina Zh. A. Kak iskusstvennyj intellekt menyaet pravila igry v kiberbezopasnosti // Forum. Seriya: Rol' nauki i obrazovaniya v sovremennom informacionnom obshhestve. 2024. № S2(32). pp. 235–240.
14. Bukin A. V., Samonov A. V., Tixonov E. I. Obnaruzhenie incidentov informacionnoj bezopasnosti na osnove texnologii nejronnyx setej // Voprosy kiberbezopasnosti [Cybersecurity issues]. 2022. № 5(51). pp. 61–73. DOI: 10.21681/2311-3456-2022-5-61-73.
15. Karpenko M. P. Tokenizaciya kak metod kolichestvennogo izmereniya informacii i znanij v uchebnyx tekstax professional'nogo obrazovaniya // Innovacii v obrazovanii. 2025. № 3. pp. 40–50.
16. Obrabotka estestvennogo yazyka v dejstvii / L. Xobson, X. Xannes, X. Koul. SPb: Piter, 2020. 575 p.

3 Nikolay N. Boldyrikhin, Ph.D. (in Engineering sciences), Associate Professor of the Department of Cyber security of Information Systems at the Don State Technical University, Rostov-on-Don, Russia. E-mail: boldyrikhin@mail.ru

4 Eduard A. Yadrets, master's student of the Department of Cybersecurity of Information Systems at the Don State Technical University. Rostov-on-Don, Russia, E-mail: xperia1058@gmail.com