МЕТОДЫ И СИСТЕМЫ ЗАЩИТЫ ИНФОРМАЦИИ, ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

DOI 10.54398/20741707_2022_3_61
УДК 004.032.26

МЕТОДИКА ОБНАРУЖЕНИЯ АТАК СОЦИАЛЬНОЙ ИНЖЕНЕРИИ НА ОСНОВЕ АЛГОРИТМОВ АНАЛИЗА ЕСТЕСТВЕННОГО ЯЗЫКА

Статья поступила в редакцию 22.06.2022, в окончательном варианте – 31.08.2022.

Частикова Вера Аркадьевна, Кубанский государственный технологический университет, 350072, Российская Федерация, г. Краснодар, ул. Московская, 2,
кандидат технических наук, доцент, ORCID: 0000-0003-2372-8275, e-mail: chastikova_va@mail.ru
Гуляй Виктория Геннадьевна, Кубанский государственный технологический университет, 350072, Российская Федерация, г. Краснодар, ул. Московская, 2,
студентка, ORCID: 0000-0002-3131-5705, e-mail: ms.gulyay@bk.ru

В последнее время наравне с техническими атаками на пользователей различных электронных коммуникационных систем возросло количество атак социальной инженерии. Так, за последний год был зафиксирован прирост атак социальной инженерии более чем на 90 %. Существующие на данный момент готовые продукты различных производителей не способны в полной мере бороться с атаками подобного типа. В данной работе рассматривается новый подход к решению задач по обнаружению атак социальной инженерии – применение алгоритмов анализа естественного языка. С целью экспериментальной проверки возможности использования данных методов в рамках поставленной задачи были реализованы следующие алгоритмы: языковая модель bag-of-words, алгоритм вложения слов Word2Vec и метод BERT, базирующийся на архитектуре Трансформер. По результатам проведенных исследований выявлено, что лучшие результаты показала модель BERT, у которой точность обработки данных контрольной выборки составила 97,35 %. Также стоит отметить алгоритм bag-of-words, имеющий значительное преимущество относительно других моделей в скорости обработки данных – примерно 1–2 м/с на одну эпоху обработки данных. Алгоритм Word2Vec показал средние результаты относительно моделей bag-of-words и BERT.

Ключевые слова: атака социальной инженерии, алгоритм анализа естественного языка, bag-of-words, Word2Vec, BERT

METHODOLOGY OF SOCIAL ENGINEERING ATTACK DETECTION BASED ON NATURAL LANGUAGE ANALYSIS ALGORITHMS

The article was received by the editorial board on 22.06.2022, in the final version – 31.08.2022.

Chastikova Vera A., Kuban State Technological University, 2 Moskovskaya St., Krasnodar, 350072, Russian Federation,
Cand. Sci. (Engineering), Associate Professor, ORCID: 0000-0003-2372-8275, e-mail: chastikova_va@mail.ru
Gulyai Victoria G., Kuban State Technological University, 2 Moskovskaya St., Krasnodar, 350072, Russian Federation,
student, ORCID: 0000-0002-3131-5705, e-mail: ms.gulyay@bk.ru

Recently, along with technical attacks on users of various electronic communication systems, the number of social engineering attacks has increased. So, over the past year, an increase in social engineering attacks has been recorded by more than 90 %. Currently existing ready-made products from various manufacturers are not able to fully combat attacks of this type. In this paper, we consider a new approach to solving problems of detecting social engineering attacks - the use of natural language analysis algorithms. In order to experimentally test the possibility of using these methods within the framework of the task, the following algorithms were implemented: the bag-of-words language model, the Word2Vec word embedding algorithm and the BERT method based on the Transformer architecture. According to the results of the study, it was found that the best results were shown by the BERT model, in which the accuracy of processing data from the control sample was 97.35 %. It is also worth noting the bag-of-words algorithm, which has a significant advantage over other models in data processing speed - approximately 1-2 ms per data processing epoch. The Word2Vec algorithm showed average results relative to the bag-of-words and BERT models. This word embedding algorithm has a processing accuracy advantage over the bag-of-words language model, and a processing speed advantage over the BERT algorithm.

Keywords: social engineering, natural language processing algorithms, bag-of-words, Word2Vec, BERT

Graphical annotation (Графическая аннотация)

1 – входящий звонок; 2 – пользователь делает отметку в специальном приложении о возможной угрозе атаки социальной инженерии, содержащейся во входящем звонке (субъективная оценка); 3 – телефонная база данных, содержащая в себе номера телефонов и пометки, сделанные пользователями; 4 – оповещение пользователя об угрозе; 5 – атака социальной инженерии с номера телефона, не зарегистрированного в телефонной базе; 6 – программа считает звонок безопасным и не предупреждает пользователя о реальной угрозе; 7 – программный комплекс на основе алгоритмов обработки естественного языка анализирует разговор (объективная оценка); 8 - программа выдает пользователю отчет о проделанном анализе и предупреждает его в случае атаки социальной инженерии.

Введение. На фоне событий последних двух лет наравне с техническими атаками на пользователей сети участились атаки с использованием методов социальной инженерии. Так, по данным Центра мониторинга и реагирования на кибератаки Solar JSOC компании «Ростелеком» в 2021 году атаки с применением методов социальной инженерии составили около 30 % от всех осуществленных кибератак. Также согласно опросу, проведенному российской исследовательской компанией «Антифишинг» [7], доля пользователей, подвергшихся атакам мошенников, применяющих методы социальной инженерии, составила 86,4 % от всех пользователей сети, из которых порядка 18,7 % понесли серьезный ущерб. Таким образом, примерно каждый пятый человек, в адрес которого была осуществлена атака социальной инженерии, понес ощутимые материальные или финансовые потери, а также неизбежно подвергся утечке персональных данных (рис. 1).

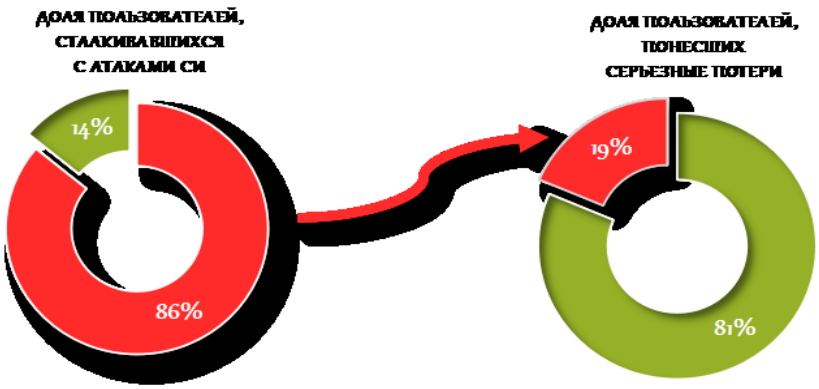

Рисунок 1 – Доля пользователей сети, подвергшихся атакам социальной инженерии и понесщих убытки соответственно

Также по данным, предоставленным российской исследовательской компанией «Антифишинг», было выявлено, что в последнее время мошенники все чаще стали использовать в своих целях вишинг *(англ. vishing, от Voice phishing)* – «выманивание» интересующей информации путем совершения телефонных звонков. Так, в рамках опроса [7] с вишингом сталкивались порядка 59 % опрошенных (рис. 2).

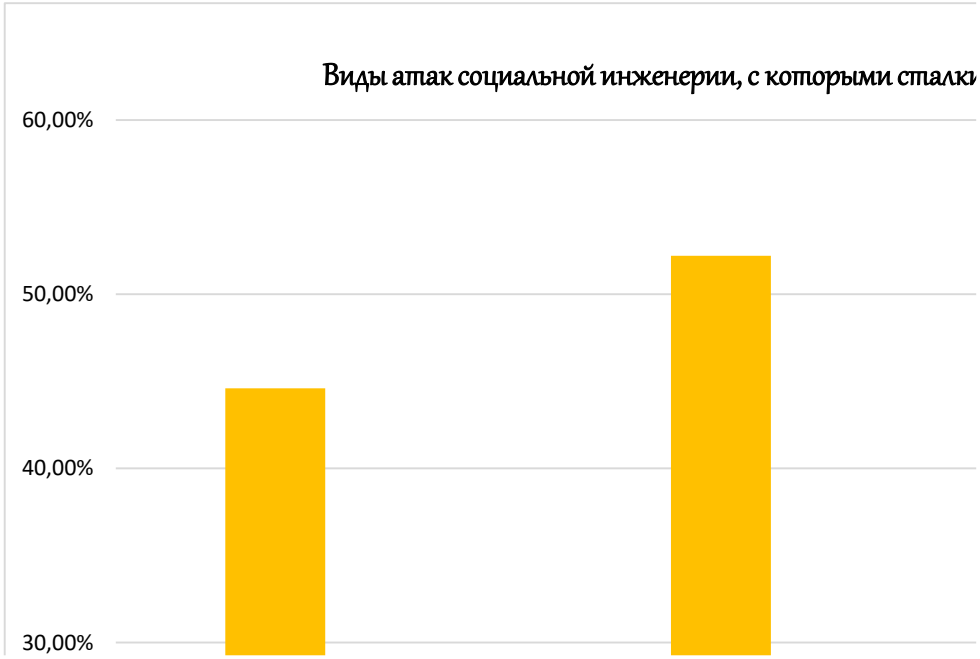

Рисунок 2 – Виды атак социальной инженерии, с которыми сталкивались пользователи за 2021 год

Обзор проблемной области. Неутихающий рост атак социальной инженерии объясняется тем, что такие атаки являются своего рода «взломом» человека, так как зачастую жертва самостоятельно выдает мошеннику всю необходимую ему конфиденциальную информацию [3]. В большинстве случаев такой способ требует от злоумышленника гораздо меньших затрат по сравнению с другими типами атак. Уже не приходится приобретать или создавать специализированное оборудование, позволяющее достигать мошеннику поставленных целей. Так, например, в случае вишинговых атак злоумышленнику достаточно лишь совершить телефонный звонок и представиться сотрудником банка/полиции/налоговой службы или придумать более изощренный способ выманивания персональных данных [3]. Главным остается одно: злоумышленнику не требуется иметь специальные навыки или особое оборудование, а соответственно, такой способ даже если не принесет выгоды, то и серьезных расходов не потребует. Благодаря этому мошенники для реализации своих целей все чаще обращаются именно к такому способу атаки.

Также стоит отметить, что некоторые способы атак социальной инженерии, такие как вишинг и фишинг, могут объединятся и с другими методами [4], представленными в таблице 1.

Таблица 1 – Методы социальной инженерии, совмещающиеся с фишинговыми и вишинговыми атаками

Название метода	Описание	Пример
Претекстинг	Использование некого предлога для выманивания у жертвы конфиденциальной информации; в большинстве случаев является неотъемлемой частью вишинга.	Мошенник совершает телефонный звонок, в котором представляется сотрудником банка и сообщает, что карта человека, которому он звонит заблокирована, для ее разблокировки необходимо сообщить номер карты и ее пин-код.
Троянский конь	Отправка пользователю сети сообщения, содержащего вредоносную ссылку, – один из самых распространенных вариантов фишинга.	На электронную почту человеку приходит письмо с описанием какого-либо очень выгодного предложения, например, скидки на нужный пользователю товар. Заинтересовавшейся жертве для более подробного ознакомления предлагается перейти по ссылке на сайт компании, продающей товар по сниженной цене. Однако ссылка оказывается ложной, и вместо перехода на сайт человек скачивает на свой компьютер или телефон вирус.

Продолжение таблицы 1

Кви про кво (услуга за услугу)	Способ, непосредственно связанный с такими видами атак, как фишинг и вишинг, так как зачастую является их основной составляющей.	Мошенник совершает звонок в крупную компанию, представившись техническим менеджером, узнает у сотрудников о наличии каких-либо неполадок в системе, после чего под предлогом «устранения» неполадок получает доступ к системе.
Обратная социальная инженерия	Способ, предполагающий ситуацию, когда жертва сама будет вынуждена обратиться к мошеннику за помощью. Такой способ часто является продолжением атаки вида «Троянский конь».	После предварительного запуска вируса на компьютере жертвы мошенник делает рассылку писем от лица фирмы, занимающейся устранением компьютерных вирусов, по электронным адресам жертв.
Лечевой серфинг	Самый простой способ получения персональных данных – подглядывание, подслушивание и т.д. Мошенник получает персональные данные жертвы путем плечевого серфинга, а затем использует их в ходе вишинговой атаки для достижения доверия со стороны жертвы.	Так, услышав, как человек в общественном месте диктует номер своего телефона и называет имя, мошенник получает возможность втереться в доверие при совершении звонка от имени сотрудника каких-либо служб. Также мошенник может пойти дальше и проследить за будущей жертвой с целью получить большее количество данных: подсмотреть номер карты на кассе в магазине, отследить, где человек работает и т.д.

Другие виды социальной инженерии, такие как «дорожное яблоко», – способ, аналогичный «троянскому коню», только имеющий не электронный, а физический вид, и реклама зачастую носит самостоятельный характер.

Возможные пути решения рассматриваемой проблемы. На данный момент существуют приложения, позволяющие вычислять мошенников, использующих в своих целях телефонную связь. Однако все они работают по одинаковому механизму, представленному на рисунке 3.

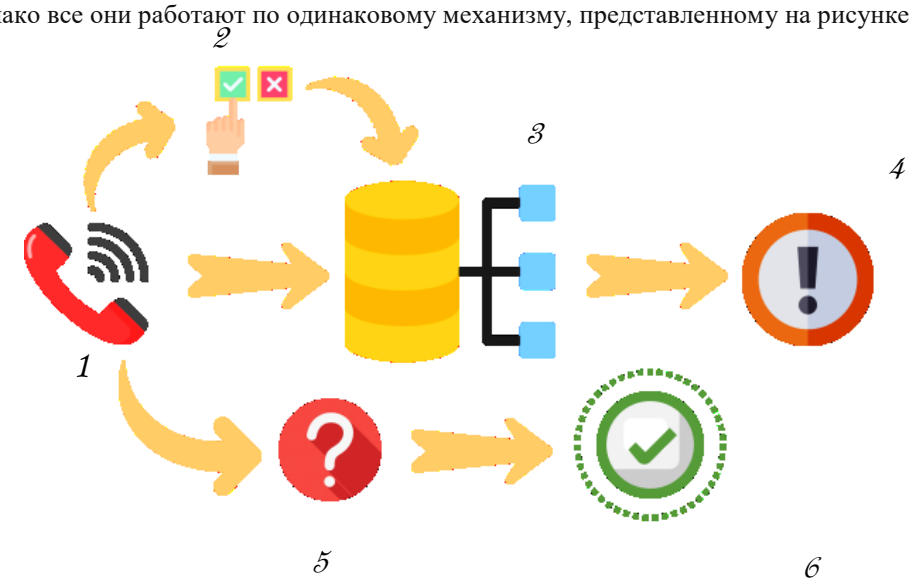

Рисунок 3 – Обобщенная схема функционирования существующих приложений для выявления телефонных мошенников: 1 – входящий звонок; 2 – пользователь делает отметку в специальном приложении о возможной угрозе атаки социальной инженерии, содержащейся во входящем звонке (субъективная оценка); 3 – телефонная база данных, содержащая в себе номера телефонов и пометки, сделанные пользователями; 4 – оповещение пользователя об угрозе; 5 – атака социальной инженерии с номера телефона, не зарегистрированного в телефонной базе; 6 – программа считает звонок безопасным и не предупреждает пользователя о реальной угрозе

Ядром каждого такого приложения является телефонная база данных, состоящая, главным образом, из номеров телефонов, содержащих различные пометки: «Спам», «Мошенник», «Робот» и т.д. Каждое конкретное приложение имеет свой набор меток. Данные приложения в зависимости от состава базы данных можно разделить на два типа:

- приложения, хранящие в базе данных только номера телефонов, имеющие метки;
- приложения, хранящие в базе данныхвсе контакты устройства, на котором данное приложение установлено.

Первый тип приложений является более безопасным, так как вероятность утечки персональных данных ниже по сравнению со вторым типом, где вопрос о сохранности личных сведений остается открытым. Это связано с тем, что подобные базы не раз из-за утечек различного рода попадали непосредственно в руки мошенников [6].

Также важно учесть, что маркируют номера телефонов сами пользователи, опираясь на свои личные ощущения, которые могут меняться и становиться обманчивыми под влиянием различных факторов, как внутренних, так и внешних. Человек может поддаться умелым убеждением мошенника и не заподозрить обмана или, наоборот, решив перестраховаться, поставить метку на номер телефона, звонок с которого не нес в себе никакой угрозы. Также в спешке человек может просто забыть пометить подозрительный номер телефона. Таким образом, такие данные являются весьма субъективными.

Ключевым моментом в механизме работы данного рода приложений является поиск в базе данных номера телефона, с которого осуществляется входящий звонок на телефон пользователя приложения. При наличии какой-либо метки у данного номера пользователю выводится оповещение с информацией о наличии возможной угрозы.

Однако такие приложения имеют еще один серьезный недостаток: если мошенник производит звонок с номера телефона, незарегистрированного в базе данных, то система не сможет заподозрить опасность, даже если этот разговор будет носит реальную угрозу для сохранности персональных данных или финансовых средств пользователя. Это связано с тем, что все приложения такого рода сфокусированы на фиксации и дальнейшем поиске номера телефона звонившего в базе данных, а не на анализе семантической значимости телефонного разговора.

Предлагаемый вариант решения. В рамках данной работы было проведено исследование, в ходе которого было выявлено, что в борьбе с мошенничеством путем применения методов социальной инженерии наиболее эффективными будут алгоритмы анализа естественного языка, так как обработка естественного языка на данный момент является одной из наиболее современных и стремительно развивающихся областей искусственного интеллекта, а также с ее помощью становится возможным анализировать семантическую нагрузку текстов, воспроизводимых на естественном языке [4, 5].

Так, одним из вариантов решения задач, связанных с обнаружением и нейтрализацией атак социальной инженерии, является создание программного комплекса, базирующегося на алгоритмах обработки естественного языка. Обобщенный принцип работы такого программного комплекса представлен на рисунке 4.

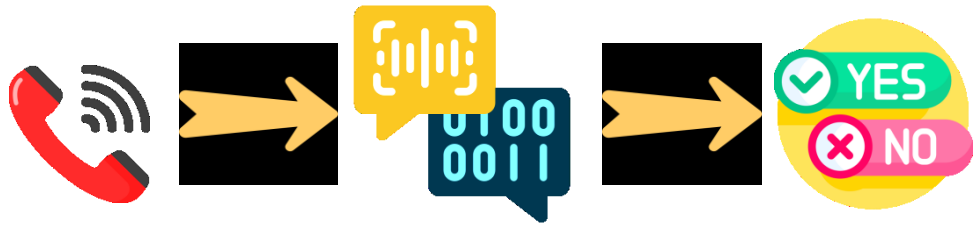

Рисунок 4 – Обобщенная схема работы предлагаемой методики: 1 – входящий звонок; 2 – программный комплекс на основе алгоритмов обработки естественного языка анализирует разговор (объективная оценка); 3 – программа выдает пользователю отчет о проделанном анализе и предупреждает его в случае атаки социальной инженерии

Для функционирования данного программного комплекса не требуется наличие телефонной базы данных, так как механизм его работы главным образом сосредоточен на анализе семантической составляющей входящего звонка. То есть программный комплекс в режиме реального времени обрабатывает телефонный разговор и при необходимости оповещает пользователя о возможной опасности.

Таким образом, приложение, базирующееся на данном программном комплексе, будет иметь ряд преимуществ относительно уже существующих продуктов:

- обработка непосредственно смысловой нагрузки разговора, позволяющая обнаруживать атаки независимо от того, производились звонки с этого номера телефона ранее или нет;
- результат анализа в режиме реального времени, то есть приложение позволяет моментально пресекать возможную атаку со стороны мошенников, не дожидаясь того, когда данный номер телефона будет помечен одним из пользователей;
- объективная оценка, не зависящая от личностных качеств и предубеждений пользователей приложения;

- безопасность приложения, обусловленная тем, что для функционирования данного программного комплекса не требуется предоставлять доступ к персональной информации, например, к номерам контактных телефонов, записанных на устройстве.

Выбор алгоритмов обработки естественного языка на основе проведенного анализа. С целью дальнейшей реализации программного комплекса был проведен сравнительный анализ существующих методов обработки естественного языка [1], в результате которого были выбраны следующие алгоритмы:

- языковая модель bag-of-words;
- алгоритм вложения слов Word2Vec;
- метод BERT, базирующийся на архитектуре Трансформ.

Языковая модель bag-of-words. Данная модель была выбрана, несмотря на ряд ее недостатков в связи с тем, что она обладает наиболее высокой скоростью обработки данных. Скорость обработки данных является важным аспектом при выборе алгоритма обработки, так как приложение, базирующееся на рассматриваемом программном комплексе, должно обрабатывать данные в режиме реального времени. Соответственно, следует отдать предпочтение модели, требующей меньшее время на обработку поступающей информации относительно других методов.

Обработка информации с помощью модели bag-of-words происходит следующим образом [1]: данные из датасета разбиваются на предложения, непосредственно составляющие датасет, и слова, входящие в него, из которых создается словарь данной модели. На основе полученных данных строится матрица, позволяющая определить наличие того или иного слова в каждом предложении [9]. Столбцами в полученной матрице являются слова из составленного словаря, а строками – исходные предложения. Общий вид механизма обработки данных с помощью bag-of-words представлен на рисунке 5.

Значительным недостатком модели bag-of-words является отсутствие связей контекстных связей между словами в предложении и, соответственно, в тексте [1]. Это связано с тем, что модель хранит лишь данные о наличии или отсутствии слова в том или ином предложении. Такой метод обработки информации может быть эффективен, но только при работе с однотипными данными.

Рисунок 5 – Механизм работы языковой модели bag-of-words

Алгоритм вложения слов Word2Vec. Данный алгоритм компенсирует недостаток модели bag-of-words за счет процесса «встраивания» слов, т.е. создания вещественного вектора, соответствующего каждому слову в тексте, в многомерном пространстве таким образом, что слова, близкие по семантическому значению, находятся на более близком расстоянии [8]. Работа непосредственно с семантической нагрузкой слов позволяет данной модели устанавливать контекстные связи с ближайшими словами, что заметно улучшает эффективность обработки данных по сравнению с моделью bag-of-words. При этом благодаря применению в базе модели сразу двух алгоритмов обработки данных (CBOW и Skip-gram) становится возможным как определение контекста заданного слова, так и обратная операция – нахождение слова по известному контексту. Благодаря этому модель Word2Vec работает эффективнее аналогичных алгоритмов вложения слов fastText и GloVe. Однако метод Word2Vec не всегда справляется с обработкой неизвестных или редких слов [1].

Обобщенная схема обработки данных с помощью алгоритма Word2Vec представлена на рисунке 6. Можно заметить, что в данном случае модель работает непосредственно со словами и их контекстом, а не с предложениями, в которых они находятся. Такой метод обработки данных как раз и способствует определению семантической нагрузки каждого слова с целью дальнейшего построения многомерного векторного пространства.

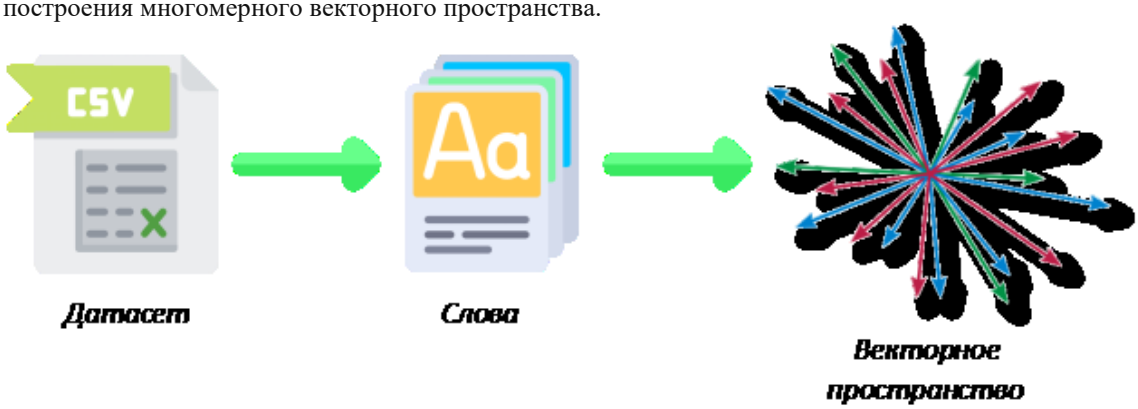

Рисунок 6 – Механизм работы алгоритма вложения слов Word2Vec

Метод BERT. Наиболее современным из рассматриваемых методов является метод BERT, базирующийся на современнейшей архитектуре Трансформер. Такая архитектура позволяет производить обучение модели даже на небольшой выборке данных за счет обработки данных по принципу «каждый с каждым» (рис. 7).

Основным отличием новой архитектуры является применение в качестве энкодера и декода алгоритма, включающего в себя механизм Multi-head Attention (рис. 8), вместо рекуррентных или сверточных нейронных сетей [2]. Данный механизм позволяет каждому входному вектору взаимодействовать с другими векторами благодаря наличию нескольких взаимосвязанных алгоритмов Self-Attention, выполняющих параллельную обработку данных. В свою очередь, каждый механизм Self-Attention имеет свой набор весовых коэффициентов, что позволяет воссоздавать полный контекстный разбор ситуации. Результаты работы всех таких механизмов объединяются в общий тензор, который в ходе последующей обработки становится итоговым результатом. Такой подход значительно увеличивает эффективность работы модели по сравнению с рекуррентными и сверточными нейронными сетями, где используется hidden state (скрытое состояние модели, необходимое для прогнозирования последовательности на выходе). В связи с этим языковые модели, базирующиеся на архитектуре Трансформер, имеют ряд преимуществ:

1. Алгоритмы обработки естественного языка, основывающиеся на архитектуре-трансформере, могут обрабатывать бесконечно длинные последовательности данных, не теряя семантическую зависимость между словами при обработке, в отличие от рекуррентных нейронных сетей, где даже применение механизмов LSTM и GPU с высокой степенью рекурсии не позволяет решить данную проблему [1].
2. Модели, базирующиеся на архитектуре Трансформер, не имеют присущей рекуррентным нейронным сетям проблемы взрывающихся и затухающих градиентов [2]. Это обусловлено тем, что благодаря архитектуре-трансформеру модель обрабатывает все имеющиеся данные за один проход, в отличие от рекуррентных и сверточных нейронных сетей, где обработка данных происходит последовательно.
3. Из второго пункта также следует, что модель, базирующаяся на данной архитектуре, требует меньшее количество шагов обучения относительно рекуррентных и сверточных нейронных сетей.

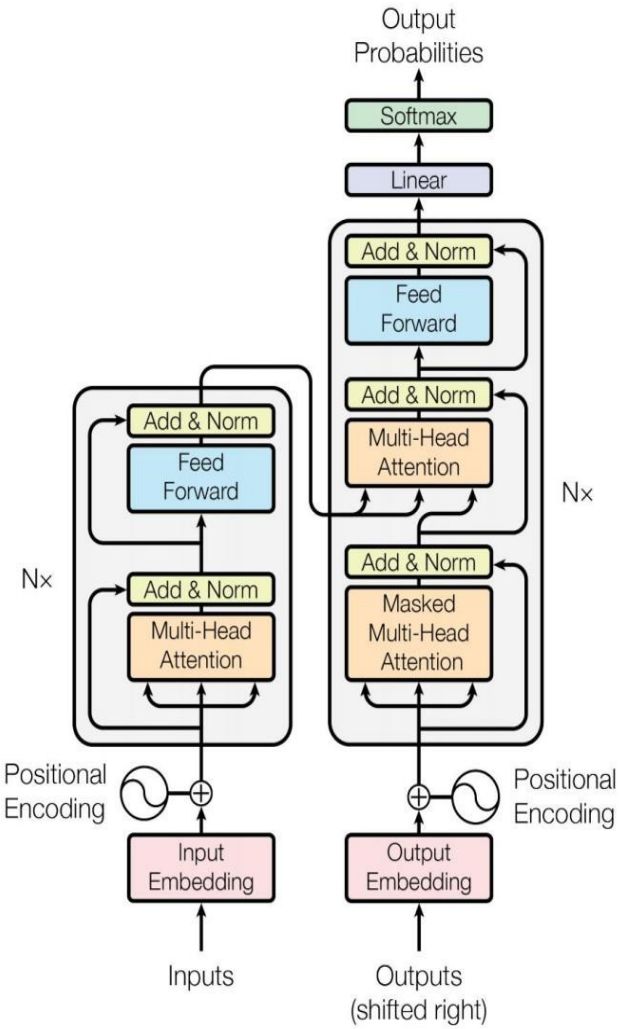

Рисунок 7 – Структура алгоритма BERT, функционирующего на базе архитектуры Трансформер [2]

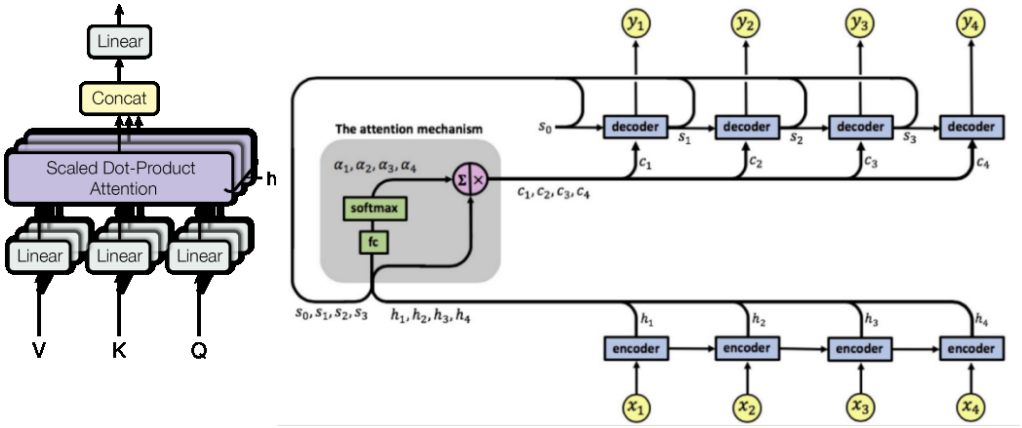

Рисунок 8 – Схема работы механизма Multi-head Attention, где Q (query) – тензор Запросов, K (key) – тензор Ключей, V (value) – тензор Значений [2]

Таким образом, метод BERT (рис. 9) аналогично алгоритму вложения слов работает непосредственно со словами, их значением и контекстом. Однако рассмотренный выше механизм Multi-head attention позволяет строить контекстные связи не только между рядом стоящими словами, но и между словами во всем предложении или тексте. Обработка данных таким образом позволяет определять семантическую нагрузку всего предложения или логически связного текста. Важно учесть, что метод затрачивает гораздо большее время на обработку информации относительно рассматриваемых ранее методов.

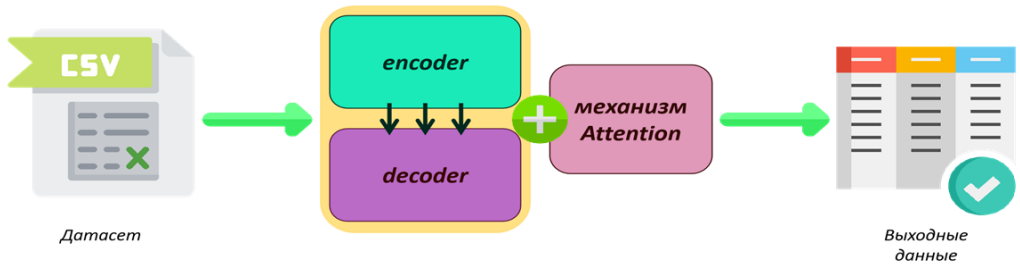

Рисунок 9 – Механизм работы метода BERT, базирующегося на архитектуре Трансформер

Разработка датасета. Для обучения любой модели NLP (natural language processing) необходим набор соответствующих данных – датасет. От корректности подобранного материала зависит точность обучения и эффективность работы алгоритма. Социальная инженерия является динамическим типом кибератак, так как регулярно появляются новые методы атак, меняются их формы и содержание. Соответственно, набор данных, разработанный для обучения программного комплекса (ПК) с целью борьбы с атаками социальной инженерии, должен содержать наиболее актуальные примеры атак на момент реализации ПК.

Поэтому в рамках данной исследовательской работы был создан набор, состоящий из 1000 примеров атак, наиболее актуальных и часто встречаемых на начало 2022 года. Для создания датасета были использованы материалы исследований компаний «Антифишинг» и «Positive Technologies». Часть полученного обучающего набора данных приведена на рисунке 10.

```
import zipfile
import os
import time
from google.colab import drive
from PIL import Image
drive.mount('/content/drive')
df = pd.read_csv('Dataset_SI.csv')
df['dialog']
```

```
1. Добрый день! Я менеджер банка, вы знакомы с инвестициями? Если нет, скажите номер вашего счета, чтобы я мог оценить ваши...
2. Здравствуйте! Это ваша управляющая компания, Произошел перерасчет оплаты за коммунальные услуги, , с вас были сняты изли...
3. Добрый день! Я сотрудник банка. На данный момент в банке снижена ставка по кредиту, не желаете ознакомиться с условиями?
4. Я Ваш новый коллега из вашего отдела, я забыл пароль от входа в систему, не подскажите его?
5. Здравствуйте! В период с 14 по 18 марта вам были предоставлены услуги ржд. Как вы оцениваете работу проводников, обслужи...
6. Я работник пенсионного фонда России, вас зовут? Вам назначена дополнительная выплата к пенсии. Укажите номер карты, на к...
7. Здравствуйте! Вы подключились к услуге «Погода онлайн». Если вы желаете отписаться позвоните по номеру *7543*004#
8. Не желаете вы принять участие в переписи населения по телефону? Для этого продиктуйте ваше полное имя и имена членов ваш...
9. Добрый день! Это работник отдела коммуникации банка, оцените работу нашего отдела от 1 до 10 баллов
10. Здравствуйте! Вас беспокоит управляющая компания, вам необходимо принять мастера для проверки счетчиков воды, иначе ваш...
...
996. С вашей карты списано рублей 5000. Продиктуйте номер вашей карты и ее CVC-код, чтобы предотвратить возможную атаку мош...
997. Вас беспокоит менеджер федеральной компании мегафон, поступил запрос от вашего имени на смену номера телефона. Вы дейс...
998. Здравствуйте! Это администрация социальной сети Вконтакте ваш аккаунт пытались взломать, сообщите ваши логин и пароль ...
999. Я сотрудник службы безопасности банка, поступила информация, что кто-то пытался поменять номер вашего мобильного банка...
1000. Я являюсь работником следственного комитета, Киров Ярослав работает у вас? Он проходит свидетелем по расследуемому дел...
```

Рисунок 10 – Вывод 15 предложений полученного датасета на экран

Полученные результаты. На основе полученного датасета было произведено обучение языковой модели, алгоритма вложения слов и метода таким образом, что 80 % предложений из датасета подавались на вход алгоритмов обработки естественного языка в качестве тренировочной выборки, а оставшиеся 20 % – в качестве проверочной. В результате обучения моделей NLP были получены результаты, приведенные в таблице 2.

Таблица 2 – Результаты исследования

	Модель bag-of-words	**Модель Word2Vec**	**Модель BERT**
Работа с предложениями целиком, без учета значения и связи отдельных слов	✓	✗	✗
Обработка текста с учетом синонимичности слов	✗	✓	✓

Продолжение таблицы 2

Возможность обучаться на датасете малого объема			
Примерное время обучения одной эпохи модели	1–2 мс	25–27 мс	122–126 мс
Результаты обработки обучающей выборки	95,37 %	98,42 %	99,16 %
Результаты обработки контрольной выборки	91,84 %	96,60 %	97,35 %

Выводы. Таким образом, исходя из результатов проведенного исследования, можно сделать вывод, что наилучшие результаты показала модель BERT, у которой точность обработки данных контрольной выборки составила 97,35 %. При этом стоит отметить, что алгоритм bag-of-words, несмотря на то, что точность обработки данных с его помощью была ниже относительно других методов, имеет значительное преимущество в скорости обработки данных. Так, время обучения одной эпохи модели bag-of-words в среднем составляло 1–2 мс. Алгоритм Word2Vec показал средние результаты относительно моделей bag-of-words и BERT.

Однако стоит заметить, что разница в точности обработки данных у рассматриваемых моделей в рамках проведенного исследования не столь значительна – 5,51 % между максимальным и минимальным показателями. Это обусловлено тем, что для обучения и дальнейшей проверки работы моделей использовался датасет большого объема – 1000 предложений различного рода. Но так как социальная инженерия является динамическим типом атак, то для эффективной работы алгоритмов обучающий датасет должен регулярно пополняться актуальными примерами атак. То есть дальнейшее обучение будет производиться на малой выборке, что, соответственно, приведет к значительному снижению точности обработки данных такими моделями, как bag-of-words и Word2Vec. Это связано с тем, что данные алгоритмы гарантируют хорошие результаты анализа данных только при наличии большой обучающей выборки. При этом модель BERT способна показывать высокую точность обработки данных и при обучении на небольшом датасете, это обусловлено архитектурой-трансформером, на которой базируется данный метод.

Библиографический список

1. Частикова, В. А. Методы обработки естественного языка в решении задач обнаружения атак социальной инженерии / В. А. Частикова, К. В. Козачек, В. Г. Гуляй // Вестник Адыгейского государственного университета. Серия 4: Естественно-математические и технические науки. – 2021. – № 4 (291). – С. 95–108. – DOI 10.53598/2410-3225-2021-4-291-95-108.

2. Ashish, Vaswani. Attention Is All You Need / Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin // ArXiv:1706.03762. – 2017.

3. Сахно, В. В. Социальная инженерия, ее техники и способы защиты / В. В. Сахно, А. С. Пищаева // Modern Science. – 2020. – № 2–2. – С. 349–351.

4. Свищев, А. В. Искусственный интеллект как средство защиты от атак методами социальной инженерии / А. В. Свищев, Я. А. Акатьев // Colloquium-journal. – 2020. – № 7–1 (59). – С. 52–55. – DOI 10.24411/2520-6990-2020-11490.

5. Частиков, А. П., Алешин, А. В., Частикова, В. А. Выявление аномалий в базах знаний интеллектуальных систем // Пятьдесят лет развития кибернетики : труды Международной научно-технической конференции. – 1999. – С. 123–124.

6. Российская компания, специализирующаяся на разработке решений в сфере информационной безопасности Positive Technologies. Актуальные киберугрозы: итоги 2021 года // Cybersecurity Threatscape – 2022. – Режим доступа: https://www.ptsecurity.com/ru-ru/research/analytics/cybersecurity-threatscape-2021/, свободный. – Заглавие с экрана. – Яз. рус. (дата обращения: 17.05.2022).

7. Российская исследовательская компания и разработчик программного обеспечения ООО «Антифишинг». 86,4 % россиян стали жертвами цифровых мошенников // Антифишинг. – 2021. – Режим доступа: https://www.itweek.ru/security/news-company/detail.php?ID=215424&, свободный. – Заглавие с экрана. – Яз. рус. (дата обращения: 17.05.2022).

8. Yilmaz, S. A deep learning analysis on question classification task using Word2vec representations / S. Yilmaz, S. Toklu // Neural Computing & Applications. – 2020. – Vol. 32, № 7. – P. 2909–2928. – DOI 10.1007/s00521-020-04725-w.

9. Катенко, Ю. В. Применение методов машинного обучения для анализа текстовой информации / Ю. В. Катенко // Охрана, безопасность, связь. – 2019. – Т. 3, № 4 (4). – С. 90–94.

References

1. Chastikova, V. A., Kozachek, K. V., Gulyai, V. G. Metody obrabotki estestvennogo yazyka v reshenii zadach obnaruzheniya atak socialnoy inzhenerii [Methods of natural language processing in solving problems of detecting social engineering attacks]. *Vestnik Adygeyskogo gosudarstvennogo universiteta. Seriya 4: Estestvenno-matematicheskie i tekhnicheskie nauki* [Bulletin of the Adyghe State University. Series 4: Natural-mathematical and technical sciences], 2021, no. 4 (291), pp. 95–108.

2. Ashish, Vaswani, Noam, Shazeer, Niki, Parmar, Jakob, Uszkoreit, Llion, Jones, Aidan, N. Gomez, Lukasz, Kaiser, Illia, Polosukhin. Attention Is All You Need. *ArXiv:1706.03762*, 2017.

3. Sakhno, V. V., Pishchaeva, A. S. Sotsialnaya inzheneriya, ee tekhniki i sposoby zashchity [Social engineering, its techniques and methods of protection]. *Modern Science*, 2020, no. 2–2, pp. 349–351.

4. Svishchev A. V., Akatiev, Ya. A. Iskusstvennyy intellekt kak sredstvo zashchity ot atak metodami sotsialnoy inzhenerii [Artificial intelligence as a means of protection against attacks by social engineering]. *Colloquium-journal*, 2020, no. 7–1 (59), pp. 52–55. – DOI 10.24411/2520-6990-2020-11490.

5. Chastikov, A. P., Aleshin, A. V., Chastikova, V. A. Vyyavleniye anomaliy v bazakh znaniy intellektualnykh sistem [Identification of anomalies in the knowledge bases of intelligent systems]. *Pyatdesyat let razvitiya kibernetiki : trudy Mezhdunarodnoy nauchno-tekhnicheskoy konferentsii* [Fifty Years of the Development of Cybernetics : Proceedings of the International Scientific and Technical Conference], 1999, pp. 123–124.

6. Rossiyskaya kompaniya, specializiruyushchayasya na razrabotke resheniy v sfere informatsionnoy bezopasnosti Positive Technologies. Aktualnye kiberugrozy: itogi 2021 goda [Russian company specializing in the development of information security solutions Positive Technologies. Actual cyber threats: results of 2021]. *Cybersecurity Threatscape – 2022*. Available at: https://www.ptsecurity.com/ru-ru/research/analytics/cybersecurity-threatscape-2021/ (accessed 05.17.2022).

7. Rossiyskaya issledovatelskaya kompaniya i razrabotchik programmnogo obespecheniya OOO «Antifishing». 86,4 % rossiyan stali zhertvami tsifrovykh moshennikov [Russian research company and software developer Antiphishing LLC 86.4 % of Russians became victims of digital scammers]. *Antifishing* [Antiphishing], 2021. Available at: https://www.itweek.ru/security/news-company/detail. php?ID=215424& (accessed 05.17.2022).

8. Yilmaz, S., Toklu, S. A deep learning analysis on question classification task using Word2vec representations. *Neural Computing & Applications*, 2020, vol. 32, no 7, pp. 2909–2928. – DOI 10.1007/s00521-020-04725-w.

9. Katenko, Yu. V. Primenenie metodov mashinnogo obucheniya dlya analiza tekstovoy informatsii [Application of machine learning methods for the analysis of textual information]. *Okhrana, bezopasnost, svyaz* [Security, safety, communication, 2019, vol. 3, no. 4 (4), pp. 90–94.