УДК 519.237.5

С. И. Носков [1], Д. В. Пашков [1]

[1] Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация

ПРОЕКТИРОВАНИЕ ИНФОРМАЦИОННОЙ СИСТЕМЫ РЕАЛИЗАЦИИ НЕКОТОРЫХ ЭТАПОВ ПОСТРОЕНИЯ РЕГРЕССИОННЫХ МОДЕЛЕЙ

Аннотация. Качество регрессионной модели в значительной степени определяется опытом специалистов, выполняющих сбор и исследование данных, в частности: надежными результатами наблюдений, наиболее близко определенными формами связей между переменными и верно подобранными методами оценки регрессии. Однако даже при наличии этих составляющих нельзя гарантировать оптимальный характер итоговой математической зависимости, потому как объем необходимых расчетов и проверок для поиска решения задачи оптимизации выходного уравнения часто оказывается существенно больше, чем можно выполнить вручную. В статье рассмотрены некоторые этапы подхода к процессу математического моделирования объекта методами регрессионного анализа данных, выделены проблемные места. В качестве варианта решения предлагается реализация алгоритмического программного комплекса, специализированного под описанный процесс моделирования. Проведен краткий обзор аналогов. Выводы сопровождаются графическим представлением функциональных требований к целевой программной реализации. Проектирование выполнено на концептуальном уровне.

Ключевые слова: математическая модель, регрессионный анализ, IDEF0, функциональное моделирование, UML, функциональные требования, концептуальное проектирование.

S.I. Noskov [1], D.V. Pashkov [1]

[1] Irkutsk State Transport University, Irkutsk, Russian Federation

DESIGNING AN INFORMATION SYSTEM FOR THE IMPLEMENTATION OF SOME STAGES OF BUILDING REGRESSION MODELS

Abstract. The quality of the regression model is largely determined by the experience of specialists who collect and study data, in particular: reliable observational results, the most closely defined forms of relationships between variables and correctly selected regression assessment methods. However, even with these components, it is impossible to guarantee the optimal nature of the final mathematical dependence, because the amount of necessary calculations and checks to find a solution to the optimization problem of the output equation often turns out to be significantly more than can be performed manually. The article considers some stages of the fundamental approach to the process of mathematical modeling of an object by methods of regression analysis of data, identifies problem areas. As a solution, the implementation of an algorithmic software package specialized for the described modeling process is proposed. A brief review of analogues is carried out. The conclusions are accompanied by a graphical representation of the functional requirements for the target program implementation. The design is carried out at the conceptual level.

Keywords: mathematical model, regression analysis, IDEF0, functional modeling, UML, functional requirements, conceptual design.

Введение. С развитием цифровых технологий растет объем обрабатываемой информации, поэтому с каждым годом за анализом данных укрепляются позиции актуальной научной и прикладной сферы деятельности. В подтверждение ключевой важности этого направления растет спрос на более точные и совершенные инструменты аналитики. Существующие программные продукты применяются для решения разносторонних задач: в работе [1] продемонстрирован анализ данных о дефектах технологического процесса с применением пакета Stadia; в статье [2], используя модули пакета Statistica, рассматривается решение задачи интеллектуального анализа данных.

В статистических исследованиях особое внимание занимает раздел прогнозирования поведения исследуемых объектов, в связи с чем широко применяется регрессионный анализ. Для построения разных типов регрессий применяются как прикладные программы, так и языки программирования: в работе [3] на примере Stadia и MathCad сравниваются специализированное программное обеспечение (ПО) и программы широкого назначения; в статье [4] демонстрируются возможности офисного пакета Excel при построении разных типов регрессий; в

докладе [5] рассматриваются библиотеки языка программирования R, предназначенных для регрессионного анализа; в публикации [6] представлен подход к моделированию регрессий с помощью пакета GRETL.

Несмотря на широкий выбор программных средств, позволяющих выполнять разносторонний статистический анализ данных, применение этих инструментов сопровождается рядом проблем, чаще всего связанных с устаревшими подходами к анализу. Среди прочих: отсутствие выбора алгоритмов оценки связей между переменными, сильно ограниченный набор критериев адекватности, осложненная работа с данными. В лучшем случае недостатки одного инструмента можно устранить за счет другого, в ином случае конструируется новое программное обеспечение [7]. Редко встречаются реализации методов оптимизации модели, особенно это касается программ общего назначения, основные проблемы изложены в работе [8]. К одному из таких методов относится «конкурс» регрессий [9]. Смысл механизма «конкурса» заключается в построении нелинейных аналогов регрессионных уравнений, за счет которого допускается возможным подобрать вариант с лучшими аппроксимационными характеристиками [10]. В публикации [11] представлен пример реализации алгоритма в среде Excel с использованием VBA, для выбора лучшего уравнения использовались коэффициент детерминации и критерий Фишера. В статье [12] представлена реализация на Delphi, победитель конкурса определяется по заданным критериям адекватности, в сравнении с [11] увеличено максимальное число экзогенных переменных с 3 до 5. В работе [13] описана модернизация программного комплекса [12], реализованная на C++, что привело к общему увеличению скорости вычислений. Возможности комплекса расширили предельное число регрессоров до 6, обновлена функциональность программного продукта, в том числе добавлены инструменты прогнозирования.

Очевидно, что языки программирования обладают явным преимуществом для обеспечения оптимизированных вычислений в задачах анализа данных, к приоритетным вариантам выбора стоит отнести: Delphi, PHP, Java, Ruby, семейство C, Python, специализированный для статической обработки данных R, Scala и другие. Использование языков программирования описывается во многих работах: в исследовании [14] рассматривается применение Python для проектирования регрессионной модели автомобильного рынка; в публикациях [15 – 17] выполнен сравнительный анализ различных групп языков программирования, при этом в статьях [16, 17] сделан акцент на применении R. Однако, нельзя не отметить, что языки относятся к категории сложных инструментов. Их использование требует определенного уровня подготовки, знаний синтаксиса и парадигм программирования, что подчеркивается в работе [18]. Разработка программ с графическим интерфейсом позволяет существенно увеличить охват аудитории и сфер применения, а также смещает выбор в сторону языков общего назначения. В материале [19] устанавливается вычислительное превосходство C++ при организации вычислений обратной матрицы, поэтому C-подобные языки являются лучшим выбором для реализации алгоритмов конструирования регрессионных моделей.

Обобщенный подход к анализу данных с использованием некоторых методов регрессионного моделирования. В процессе построения регрессии выбор спецификации является основополагающим этапом исследования [20, 21]. Входящие в модель факторы должны быть обоснованы достаточным объемом наблюдений. Исследуемые данные не должны содержать пропусков, т. е. обладать целостной структурой. Набор независимых переменных выбирается таким образом, чтобы подходить по смыслу к зависимой. Особо специфичные задачи требуют участия узконаправленных специалистов для объективного формулирования спецификации [9].

Ввиду простоты вычислений начинать процесс моделирования принято с формирования линейной связи между факторами:

$$y_k = \sum_{i=1}^{m} \alpha_i x_{ki} + \varepsilon_k, k = \overline{1,n}, \quad (1)$$

где n – число наблюдений; m – число переменных; y_k и x_{ki} – значения зависимой и независимых переменных соответственно; α_i – параметры, подлежащие оцениванию; ε_k – ошибки аппроксимации.

Для оценки вектора параметров $\alpha = (\alpha_1, \dots, \alpha_m)$ линейной регрессии (1) применяется ряд методов, среди которых наиболее распространенным является метод наименьших квадратов [22]. Он входит в число базовых методов регрессионного анализа, прост в применении и часто универсален, ввиду своей умеренной восприимчивости к выбросам. Тем не менее, в зависимости от требуемого уровня чувствительности, используются методы наименьших модулей, робастного, антиробастного и смешанного оценивания [23].

Полученную модель следует проверить на адекватность, применив соответствующие критерии [9]. Регрессии оцениваются по множеству показателей, некоторые из них: средняя абсолютная и квадратичная ошибки аппроксимации, коэффициенты детерминации, корреляции, критерии Фишера, Дарбина-Уотсона, согласованности поведения [24, 25]. Чем шире используемый набор качественных характеристик, тем объективнее и надежнее итоговые выводы по результатам вычислений. Необходимо учитывать, что за исключением некоторых универсальных метрик, для каждого метода оценивания набор качественных показателей является уникальным.

В контексте регрессионного анализа построенная таким образом математическая модель, как правило, имеет промежуточный характер. В действительности, связи между переменными являются в значительной степени нелинейными, как следствие, полученное уравнение служит отправной точкой к дальнейшему анализу. Модель следует улучшить, воспользовавшись соответствующими методами [26, 27], например, алгоритмом проведения «конкурса» регрессий [9, 21, 24]. Число альтернативных вариантов уравнения регрессии может быть довольно велико, поэтому расчеты требуют обеспечения достаточными вычислительными мощностями. Неопределенность подхода к выбору наиболее привлекательной модели усложняет процесс идентификации победителя конкурса среди множества регрессионных уравнений. Преобладающий критерий или группа показателей определяются в зависимости от поставленной задачи. Исходя из выявленных проблем, задача конструирования специализированного ПО является крайне актуальной.

Концептуальное проектирование программной среды. Комплексный подход к проектированию информационных и программных систем гарантирует однозначный уровень понимания того, какие задачи должен решать программный продукт. На этапе концептуального проектирования обеспечивается определение и структурирование целевых требований к программе. Описанный способ к анализу данных с применением некоторых методов регрессионного анализа будем трактовать как модель “AS-IS” и использовать в качестве основы для архитектурного дизайна программы. Спецификацию состава требований начнем с формулирования желаемого результата, для этого воспользуемся методологией IDEF0 [28, 29] и визуализируем подход к исследованию регрессии в виде функциональной модели “TO-BE”, результат представлен на рис. 1 – 4, примечания к рисункам представлены в таблицах 1 – 4:

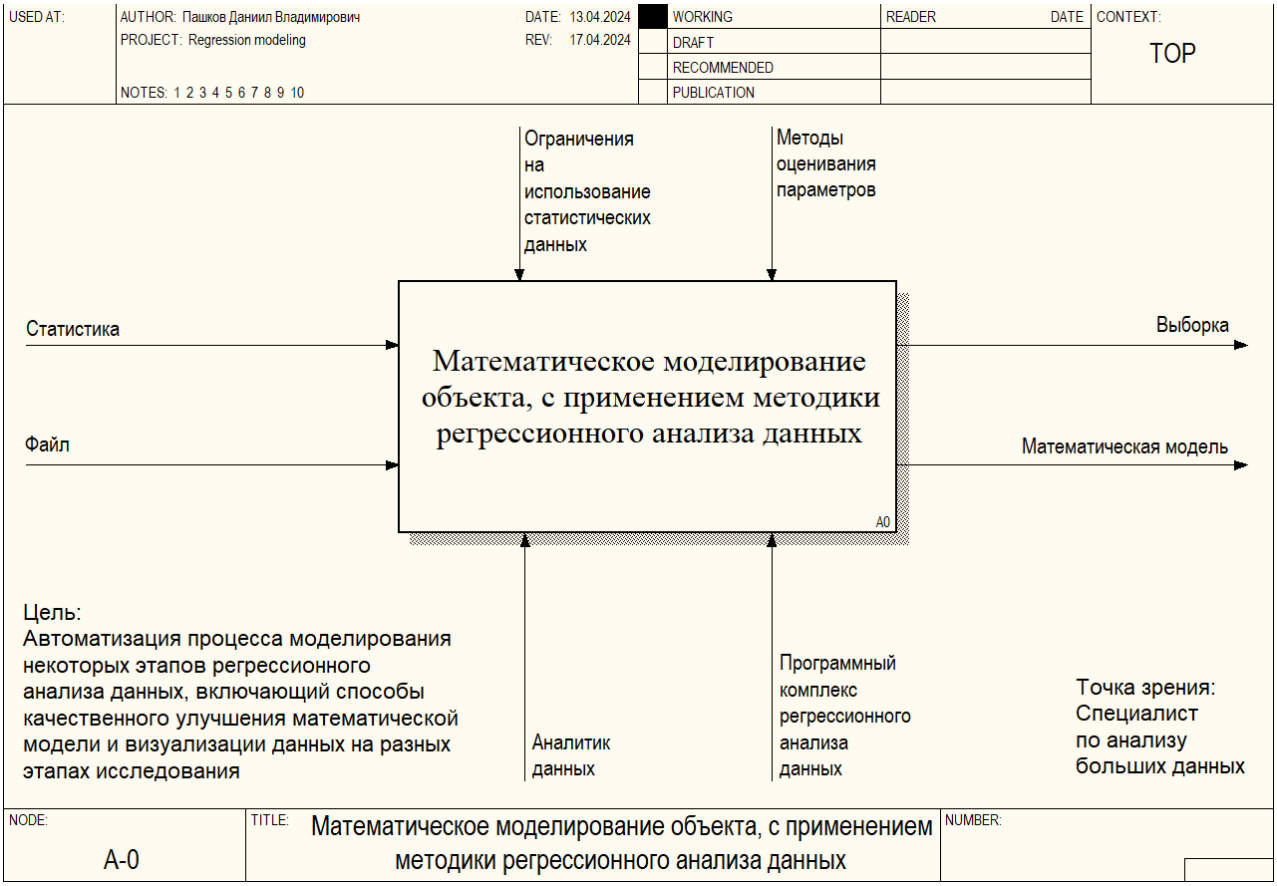

Рис. 1. Контекстная диаграмма процесса математического моделирования объекта исследования, с применением специализированного программного обеспечения

Таблица 1.

Описание дуг контекстной диаграммы модели “TO-BE”

Дуга	Пояснение
Статистика	Упорядоченный, необработанный массив статистических данных, вводимый вручную
Файл	Текстовый файл, содержащий массив статистических данных, перечисленных через разделители
Ограничения на использование статистических данных	Требования к обработке информации, содержащую коммерческую, государственную тайну или иное
Методы оценивания параметров	Арсенал методов для оценки параметров модели
Аналитик данных	Пользователь ПО, проводящий исследование
Программный комплекс регрессионного анализа данных	Целевой, разрабатываемый программный продукт
Выборка	Подготовленный к моделированию набор данных
Математическая модель	Специфицированное регрессионное уравнение

На данном этапе по описанному обобщенному подходу к анализу данных был выделен бизнес-процесс (рис. 1), подобраны входные данные, ограничения, участники процесса и выходные результаты, пояснения даны в таблице 1. Для дальнейшей детализации и декомпозируем бизнес-процесс на подпроцессы.

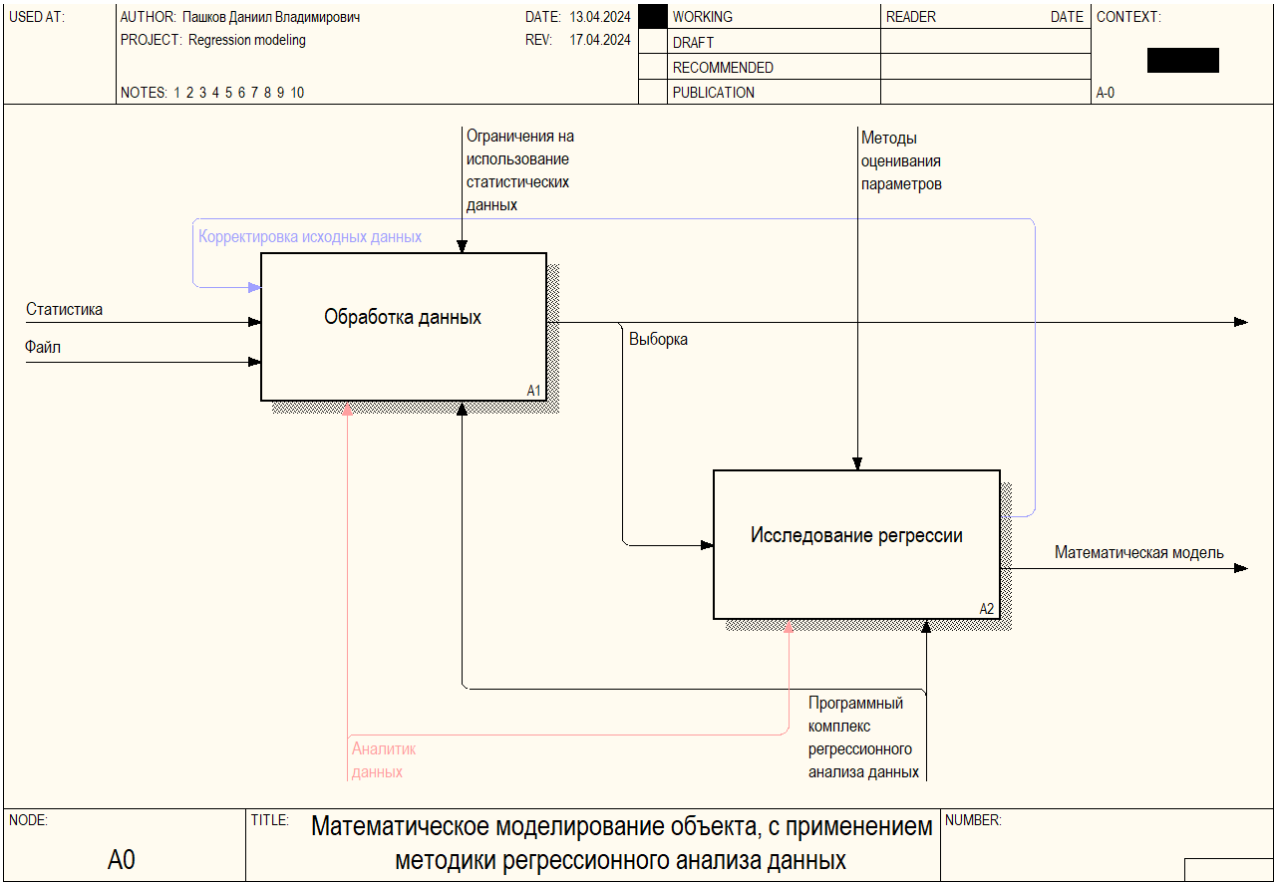

Рис. 2. 1 уровень декомпозиции модели “TO-BE” процесса математического моделирования объекта исследования, с применением специализированного программного обеспечения

Таблица 2.

Описание новых дуг 1-го уровня декомпозиции модели “TO-BE” процесса математического моделирования объекта исследования, с применением специализированного программного обеспечения

Дуга	Пояснение
Корректировка исходных данных	Редактирование исходного массива статистических данных в соответствии с новыми требованиями

На первом уровне декомпозиции (рис. 2) описана основная идея программного продукта: объединение механизмов подготовки данных, исследования и улучшения математической модели в единой программной среде, описание приведено в таблице 2. Углубимся ещё на один уровень декомпозиции для каждого из подпроцессов.

После идентификации исследуемого объекта (рис. 3), сбор и подготовка данных является главным звеном цепочки действий для построения математической модели. Выборка данных является уникальной, определяет минимальные требования к объемам наблюдений и напрямую влияет на результаты полученных вычислений. При наличии недостаточно полного набора данных следует заняться дополнительным анализом и уточнением описываемых факторов. Комментарии к дугам процесса обработки данных приведены в таблице 3.

Сам по себе процесс моделирования (рис. 4) является более тривиальным и рутинным этапом исследования по сравнению с задачей подготовки данных, но именно по результатам расчетов уточняются требования к спецификации модели и связям между переменными (рис. 3), а также на этой стадии сосредоточена основная потребность к вычислительным мощностям целевого программного продукта. Дополнения к дугам приведено в таблице 4.

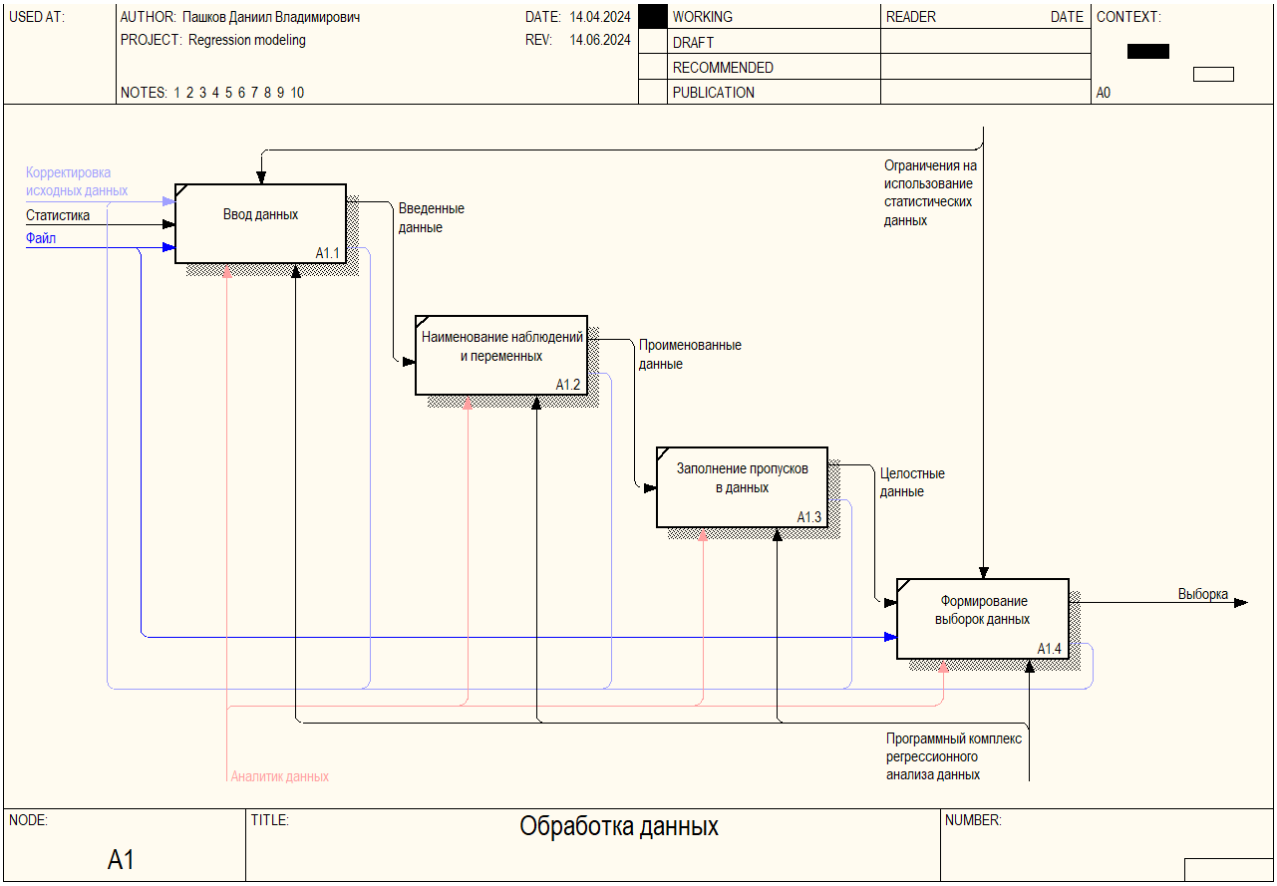

Рис. 3. Декомпозиция процесса обработки данных

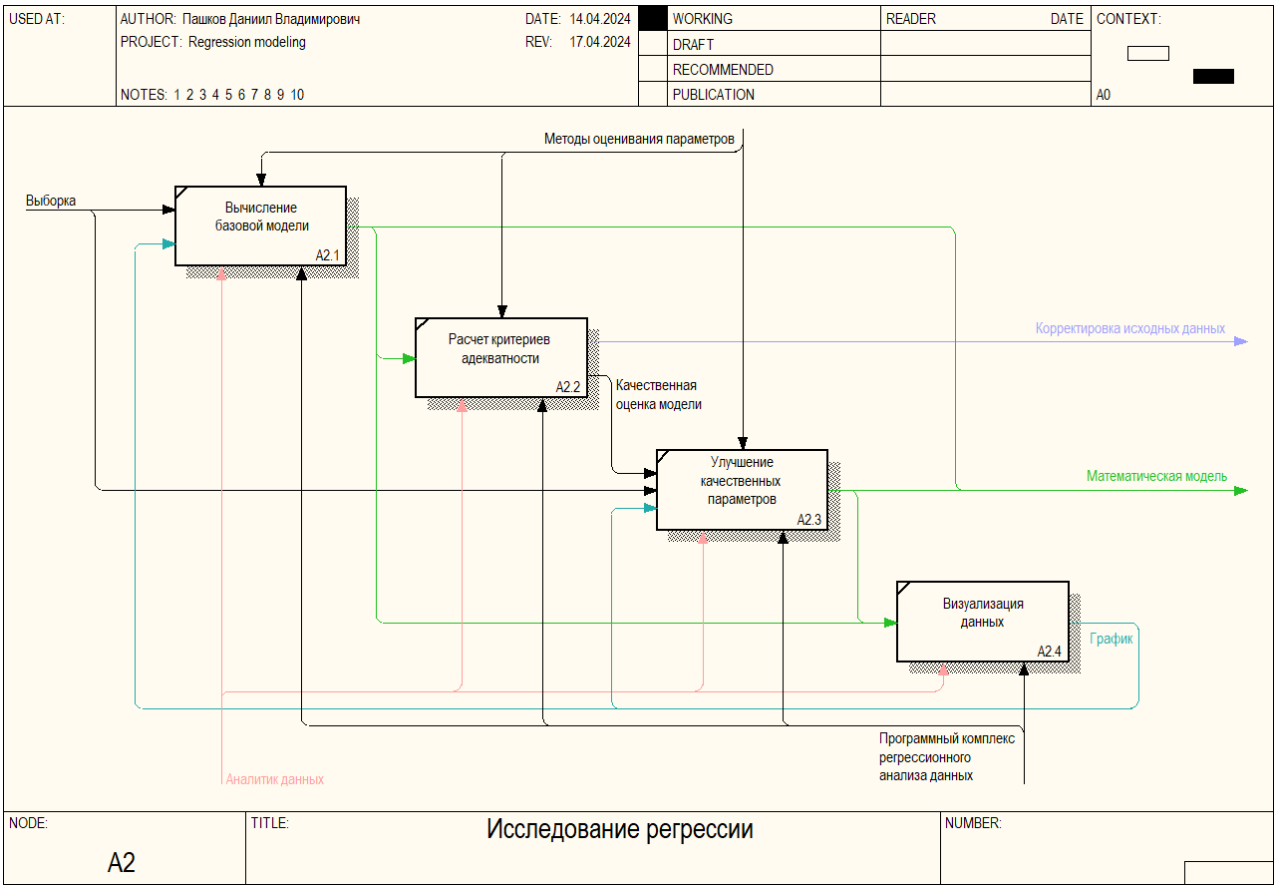

Рис. 4. Декомпозиция процесса исследования регрессии

Таблица 3.

Описание новых дуг декомпозиции процесса обработки данных

Дуга	Пояснение
Введенные данные	Массив данных, введенный в программу
Проименованные данные	Массив данных, с указанным периодом наблюдения и наименованным перечнем переменных
Целостные данные	Массив данных, не содержащий пропусков по указанному периоду наблюдений

Таблица 4.

Описание новых дуг декомпозиции процесса исследования регрессии

Дуга	Пояснение
Качественная оценка модели	Группа критериев адекватности регрессионной модели
График	Визуальное отображения наблюдаемых переменных

Декомпозицию можно продолжать дальше, но текущего описания достаточно для концептуального понимания работы системы. Таким образом, был описан бизнес-процесс моделирования объекта с применением специализированного программного обеспечения с декомпозицией на 2 уровня в глубину.

Дальнейшим этапом является формирование функциональных требований к программе. Более того, при моделировании процессов с помощью методологии IDEF0 намеренно не уделяется внимание к временной последовательности проведенных работ в пользу логического отношения между работами, поэтому отдельно следует выделить потоки исполнительных процессов. Сформулируем требования по описанной “TO-BE” модели и изобразим их графически, воспользовавшись нотацией UML [30, 31]. Представим функциональность системы на диаграмме вариантов использования (Use Case), рис. 5:

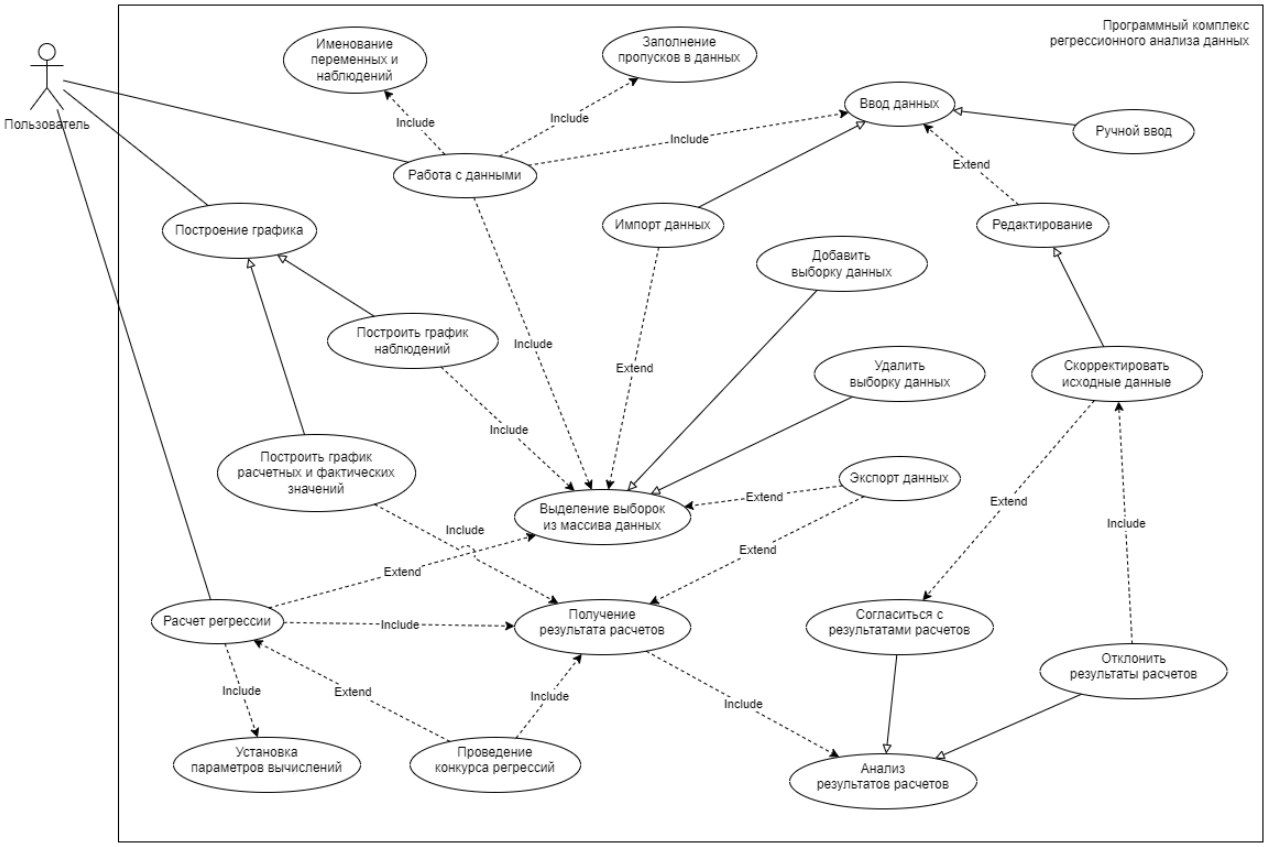

Рис. 5. Диаграмма вариантов использования процесса математического моделирования объекта с применением программного комплекса регрессионного анализа данных

На диаграмме прецедентов (рис. 5) изложены основные возможности системы и предполагаемое взаимодействие с пользователем. Переложим описанные требования на потоки действий, также воспользовавшись UML- моделированием. Типовой сценарий поведения пользователя при работе с программой отобразим на диаграмме активностей (Activity).

На диаграмме активностей, изображенной на рис. 6, 7, отражены последовательности действий участвующих в описываемом процессе акторов, к промежуточным результатам даны пояснения в блоках комментариев.

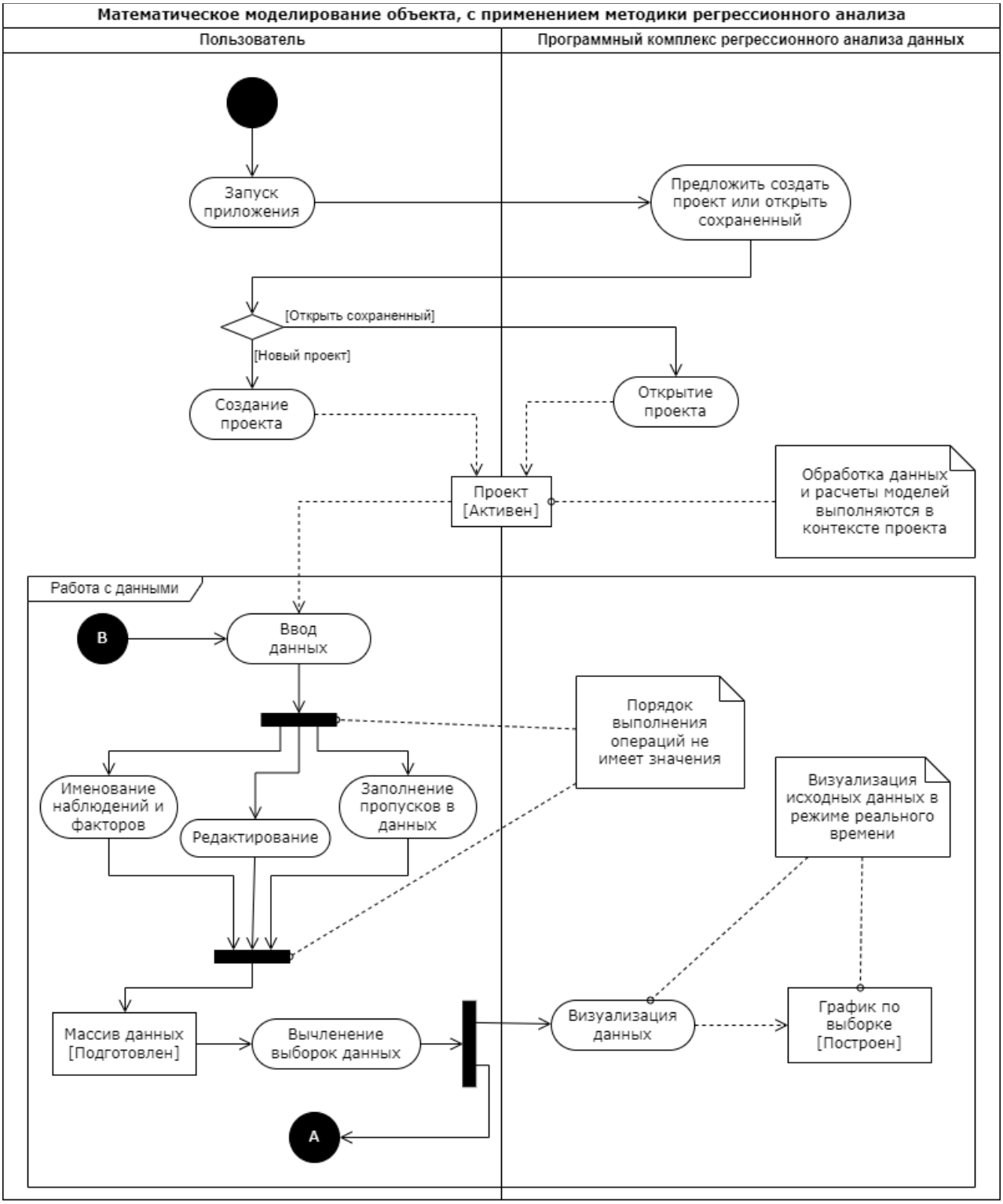

Рис. 6. Диаграмма активностей процесса математического моделирования объекта с применением программного комплекса регрессионного анализа данных

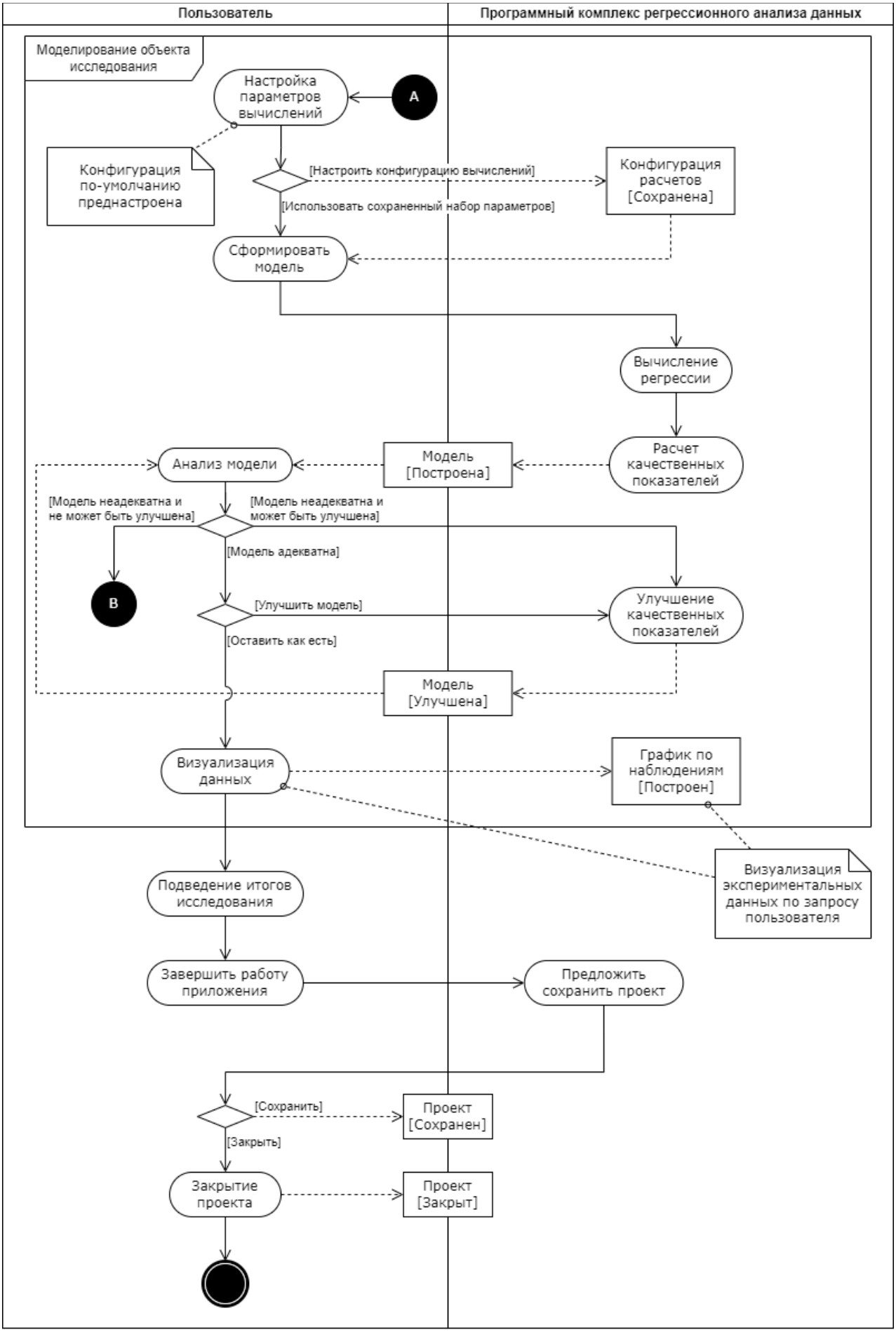

Рис. 7. Продолжение диаграммы активностей

Таким образом, был подготовлен минимальный набор графической документации, описывающий требования к целевому программному обеспечению на концептуальном уровне. Результатов текущего этапа достаточно, чтобы перейти к проектированию архитектуры программного продукта.

Заключение. В статье описаны основные процессы при математическом моделировании объекта с применением инструментов регрессионного анализа. Даны понятия и способы для улучшения качества характеристик моделей, сформулированы сопутствующие проблемы. Рассмотрены альтернативные реализации механизма. Комплексно смоделирован ожидаемый подход к проведению анализа данных с использованием специализированного программного обеспечения.

Описанные и смоделированные графически процессы ложатся в основу требований к целевому программному продукту, в дальнейшем авторы намерены продолжать развивать материал и предоставить решение описанных проблем. Целью ставится реализация программного комплекса регрессионного анализа данных.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Букша Д. Р. Применение статистического пакета STADIA для анализа данных / Научно-технические достижения студентов, аспирантов, молодых ученых строительно-архитектурной отрасли : Сборник научных трудов VIII Республиканской конференции молодых ученых, аспирантов, студентов. В 3-х томах, Макеевка, 22 апреля 2022 года. Том 1. – Макеевка: Донбасская национальная академия строительства и архитектуры, 2022. – С. 23-28.

2. Гаврилко К. А. Реализация методов интеллектуального анализа данных на основе модулей интегрированной статистической системы statistic // Научно-технические достижения студентов, аспирантов, молодых ученых строительно-архитектурной отрасли : Сборник научных трудов VI Республиканской конференции молодых учёных, аспирантов, студентов. В 3-х томах, Макеевка, 17 апреля 2020 года. Том 1. – Макеевка: Донбасская национальная академия строительства и архитектуры, 2020. – С. 16-20.

3. Болотина Н. В., Швалева А. В. Построение регрессионных моделей зависимости механических свойств от химического состава с помощью программного обеспечения Mathcad, Stadia // Наука и производство Урала. – 2015. – № 11. – С. 221-224.

4. Яковлев В. Б. Линейное и нелинейное оценивание параметров регрессии в Microsoft Excel // Вестник МГПУ. Серия: Информатика и информатизация образования. – 2019. – № 2(48). – С. 58-71.

5. Лебедева О. А., Зарядов И. С. Обзор инструментов для регрессионного анализа в R: от фундаментальных методов до нейронных сетей // Information and Telecommunication Technologies and Mathematical Modeling of High-Tech Systems : Материалы XIV международной научной конференции, Москва, 8–12 апреля 2024 года. – Москва: Российский университет дружбы народов, 2024.

6. Тусков А. А. Применение Gretl для построения многофакторной модели // Модели, системы, сети в экономике, технике, природе и обществе. – 2011. – № 1(1). – С. 154-159.

7. Кузьмин Е. С., Зайцев О. В., Кузьмина Е. В. Комплекс программ для проведения статистических расчетов Statistic 2.1 // Математическая морфология: электронный математический и медико-биологический журнал. – 1997. – Т. 2, № 1. – С. 96-98.

8. Базилевский М. П., Носков С. И. Анализ специализированного программного обеспечения для автоматизации "конкурса" регрессионных моделей // Информационные технологии и проблемы математического моделирования сложных систем. – 2010. – № 8. – С. 50-56.

9. Носков С.И. Технология моделирования объектов с нестабильным функционированием и неопределенностью в данных. – Иркутск : Облинформпечать, 1996. – 320 с.

10. Пашков Д. В., Носков С. И. Реализация конкурса регрессионных моделей эффективности интеллектуальной деятельности // Электронный сетевой политематический журнал "Научные труды КубГТУ". – 2022. – № 6. – С. 40-51.

11. Хубаев Г. Н., Денисенко В. А., Коротин Д. В. Метод всех возможных регрессий: программная реализация в системе MS Excel // Информационные системы, экономика, управление трудом и производством : Ученые записки. Том Выпуск 16. – Ростов-на-Дону : Ростовский государственный экономический университет "РИНХ", 2014. – С. 126-129.

12. Носков С. И., Базилевский М. П. Программный комплекс автоматизации процесса построения регрессионных моделей // Международный журнал прикладных и фундаментальных исследований. – 2010. – № 1. – С. 93-94.

13. Базилевский М. П. Программно-математическое обеспечение автоматизации многокритериального выбора регрессионных моделей : специальность 05.13.18 "Математическое моделирование, численные методы и комплексы программ" : диссертация на соискание ученой степени кандидата технических наук. – Иркутск, 2012. – 153 с.

14. Утакаева И. Х. Применение пакета статистического анализа Python для анализа данных автомобильного рынка // Вестник Алтайской академии экономики и права. – 2019. – № 2-2. – С. 346-351.

15. Черман А. Н., Чубенко Д. О., Филонова Е. С. Анализ языков программирования PYTHON, RUBY, SCALA для обработки больших данных на примерах алгоритмов сортировки // Наука и современность : Материалы Всероссийской научно-практической конференции студентов и молодых ученых, Таганрог, 10 ноября 2023 года. – Таганрог: ДиректСайнс, 2023. – С. 131-134.

16. Захаренков А. О., Евдокимова Г. С. Сравнительный анализ языков Ruby, R и Python в вопросах анализа данных на примере задачи кластеризации // Системы компьютерной математики и их приложения. – 2020. – № 21. – С. 38-44.

17. Багдади М. А., Мархабаев Б. А., Мысева Е. Р. Обзор возможностей и сравнение языков программирования Python и R в области анализа данных // Инновационные механизмы управления цифровой и региональной экономикой : Материалы V Международной студенческой научной конференции, Москва, 15–16 июня 2023 года. – Москва: Национальный исследовательский ядерный университет "МИФИ", 2023. – С. 282-294.

18. Ивин В. В. Применение языка R и среды RStudio для статистического анализа данных // Педагогический опыт: от теории к практике : Сборник материалов VI Международной научно-практической конференции, Чебоксары, 06 августа 2018 года / Редколлегия: О.Н. Широков [и др.]. – Чебоксары: Общество с ограниченной ответственностью "Центр научного сотрудничества "Интерактив плюс", 2018. – С. 47-53.

19. Базилевский М. П., Носков С. И. Анализ систем программирования для решения вычислительной задачи проведения "конкурса" регрессионных моделей // Информационные технологии и проблемы математического моделирования сложных систем. – 2011. – № 9. – С. 47-51.

20. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: Исследование зависимостей. – М.: Финансы и статистика, 1985. – 488 с.

21. Базилевский М. П., Носков С. И. Статистический анализ критериальных матриц при организации "конкурса" регрессионных моделей // Информационные технологии и математическое моделирование в управлении сложными системами. – 2019. – № 1(2). – С. 13-26.

22. Голованчиков А. Б., Доан М. К., Петрухин А. В., Меренцов Н. А. Сравнение точности аппроксимации экспериментальных данных методом наименьших относительных квадратов с методом наименьших квадратов // Моделирование, оптимизация и информационные технологии. – 2020. – Т. 8, № 1(28).

23. Носков С. И., Перфильева К. С., Хоняков А. А., Торопов В. Д. Возможная альтернативность подходов к регрессионному моделированию объектов // Вестник транспорта Поволжья. – 2021. – № 6(90). – С. 68-70.

24. Носков С. И., Пашков Д. В., Улыбин Т. Т., Улыбина А. Ю. Организация конкурса регрессионных моделей выгрузки вагонов на железнодорожном транспорте // Инженерный вестник Дона. – 2023. – № 6(102). – С. 301-309.

25. Носков С. И. Критерий "согласованность поведения" в регрессионном анализе // Современные технологии. Системный анализ. Моделирование. – 2013. – № 1(37). – С. 107-110.

26. Рычка О. В. Разработка алгоритма реализации методов повышения качества регрессионных моделей, используемых при проектировании технических систем // Информатика и кибернетика. – 2020. – № 3(21). – С. 13-19.

27. Михайлова Т. М., Михайлов Б. А. Методы повышения качества регрессионной модели при ее использовании для прогнозирования // Ученые записки Санкт-Петербургского имени В.Б. Бобкова филиала Российской таможенной академии. – 2003. – № 1(20). – С. 168-193.

28. Приходько Н. А., Кулаченок А. К. Моделирование в нотации IDEF0 // Моя профессиональная карьера. – 2022. – Т. 1, № 36. – С. 137-141.

29. Карпычев В. Ю. Функциональное моделирование (IDEF0) как метод исследования блокчейн-технологии // Труды НГТУ им. Р.Е. Алексеева. – 2018. – № 4(123). – С. 22-32.

30. Suriya Dr. S., S. N. Design of UML Diagrams for WEBMED - Healthcare Service System Services // EAI Endorsed Transactions on e-Learning. – 2023. – Vol. 8, No. 1. – P. e5.

31. Барклаевская Н. В. Использование унифицированного языка моделирования UML в проектном подходе при обучении студентов по специальности "бизнес-информатика" // Материалы научно-методической конференции СЗИУ РАНХиГС. – 2015. – № 1. – С. 21-29.

REFERENCES

1. Buksha D. R. Application of the STADIA statistical package for data analysis / Scientific and technical achievements of students, postgraduates, young scientists of the construction and architectural industry : A collection of scientific papers of the VIII Republican Conference of young Scientists, postgraduates, students. In 3 volumes, Makeyevka, April 22, 2022. Volume 1. – Makeyevka: Donbass National Academy of Construction and Architecture, 2022. – pp. 23-28.

2. Gavrilko K. A. Implementation of data mining methods based on modules of the integrated statistical system statistical // Scientific and technical achievements of students, postgraduates, young scientists of the construction and architectural industry : Collection of scientific papers of the VI Republican Conference of Young scientists, postgraduates, students. In 3 volumes, Makeyevka, April 17, 2020. Volume 1. – Makeyevka: Donbass National Academy of Construction and Architecture, 2020. – pp. 16-20.

3. Bolotina N. V., Shvaleva A.V. Construction of regression models of the dependence of mechanical properties on chemical composition using Mathcad, Stadia software // Science and production of the Urals. - 2015. – No. 11. – pp. 221-224.

4. Yakovlev V. B. Linear and nonlinear estimation of regression parameters in Microsoft Excel // Bulletin of the Moscow State Pedagogical University. Series: Informatics and Informatization of education. – 2019. – № 2(48). – Pp. 58-71.

5. Lebedeva O. A., Zaryadov I. S. Overview of tools for regression analysis in R: from fundamental methods to neural networks // Information and Telecommunication Technologies and Mathematical Modeling of High-Tech Systems : Proceedings of the XIV International Scientific Conference, Moscow, April 8-12, 2024. – Moscow: Russian University of Peoples' Friendship, 2024.

6. Tuskov A. A. Application of Gretl for building a multifactorial model // Models, systems, networks in economics, technology, nature and society. – 2011. – № 1(1). – Pp. 154-159.

7. Kuzmin E. S., Zaitsev O. V., Kuzmina E. V. A set of programs for statistical calculations Statistical 2.1 // Mathematical morphology: electronic mathematical and biomedical journal. – 1997. – Vol. 2, No. 1. – pp. 96-98.

8. Bazilevsky M. P., Noskov S. I. Analysis specialized software for automation of the "competition" of regression models // Information technologies and problems of mathematical modeling of complex systems. - 2010. – No. 8. – pp. 50-56.

9. Noskov S.I. Technology of modeling objects with unstable functioning and uncertainty in data. Irkutsk : Oblinformpechat, 1996. 320 p.

10. Pashkov D. V., Noskov S. I. Implementation of the competition of regression models of intellectual activity effectiveness // Electronic network polythematic journal "Scientific works of KubSTU". - 2022. – No. 6. – pp. 40-51.

11. Khubaev G. N., Denisenko V. A., Korotin D. V. The method of all possible regressions: software implementation in the MS Excel system // Information systems, economics, labor and production management : Scientific notes. Volume Issue 16. – Rostov-on-Don : Rostov State University of Economics "RINH", 2014. – pp. 126-129.

12. Noskov S. I., Bazilevsky M. P. Software package for automating the process of constructing regression models // International Journal of Applied and Fundamental Research. - 2010. – No. 1. – pp. 93-94.

13. Bazilevsky M. P. Software and mathematical support for automation of multi-criteria selection of regression models : specialty 05.13.18 "Mathematical modeling, numerical methods and software packages" : dissertation for the degree of Candidate of technical Sciences. – Irkutsk, 2012. – 153 p.

14. Utakaeva I. H. Application of the Python statistical analysis package for analyzing automotive market data // Bulletin of the Altai Academy of Economics and Law. – 2019. – No. 2-2. – pp. 346-351.

15. Cherman A. N., Chubenko D. O., Filonova E. S. Analysis of programming languages PYTHON, RUBY, SCALA for big data processing using examples of sorting algorithms // Science and modernity : Materials of the All-Russian Scientific and Practical Conference of Students and Young Scientists, Taganrog, November 10, 2023. – Taganrog: Direct Science, 2023. – pp. 131-134.

16. Zakharenkov A. O., Evdokimova G. S. Comparative analysis of Ruby, R and Python languages in data analysis using the clustering problem as an example // Computer mathematics systems and their applications. - 2020. – No. 21. – pp. 38-44.

17. Baghdadi M. A., Marhabaev B. A., Myseva E. R. Overview of the possibilities and comparison of Python and R programming languages in the field of data analysis // Innovative mechanisms of digital and regional economy management : Proceedings of the V International Student Scientific Conference, Moscow, June 15-16, 2023. – Moscow: National Research Nuclear University "MEPhI", 2023. – pp. 282-294.

18. Ivin V. V. Application of the R language and the RStudio environment for statistical analysis of data // Pedagogical experience: from theory to practice : Collection of materials of the VI International Scientific and Practical Conference, Cheboksary, August 06, 2018 / Editorial Board: O.N. Shirokov [et al.]. – Cheboksary: Society Limited Liability Company "Center for Scientific Cooperation "Interactive Plus", 2018. – pp. 47-53.

19. Bazilevsky M. P., Noskov S. I. Analysis of programming systems for solving the computational problem of conducting a "competition" of regression models // Information technologies and problems of mathematical modeling of complex systems. - 2011. – No. 9. – pp. 47-51.

20. Ayvazyan S.A., Enyukov I.S., Meshalkin L.D. Applied statistics: A study of dependencies. – M.: Finance and Statistics, 1985. – 488 p.

21. Bazilevsky M. P., Noskov S. I. Statistical analysis of criterion matrices in the organization of the "competition" of regression models // Information technologies and mathematical modeling in the management of complex systems. – 2019. – № 1(2). – Pp. 13-26.

22. Golovanchikov A. B., Doan M. K., Petrukhin A.V., Merentsov N. A. Comparison of the accuracy of approximation of experimental data by the method of least relative squares with the method of least squares // Modeling, optimization and information technologies. - 2020. – Vol. 8, No. 1(28).

23. Noskov S. I., Perfilieva K. S., Honyakov A. A., Toropov V. D. Possible alternative approaches to regression modeling of objects // Bulletin of transport of the Volga region. – 2021. – № 6(90). – Pp. 68-70.

24. Noskov S. I., Pashkov D. V., Ulybin T. T., Ulybina A. Yu. Organization of a competition for regression models of unloading wagons on railway transport // Engineering Bulletin of the Don. – 2023. – № 6(102). – Pp. 301-309.

25. Noskov S. I. Criterion "consistency of behavior" in regression analysis // Modern technologies. System analysis. Modeling. – 2013. – № 1(37). – Pp. 107-110.

26. Rychka O. V. Development of an algorithm for the implementation of methods to improve the quality of regression models used in the design of technical systems // Informatics and Cybernetics. – 2020. – № 3(21). – Pp. 13-19.

27. Mikhailova T. M., Mikhailov B. A. Methods of improving the quality of regression model when using it for forecasting // Scientific notes of the St. Petersburg branch of the Russian Customs Academy named after V.B. Bobkov. – 2003. – № 1(20). – pp. 168-193.

28. Prikhodko N. A., Kulachenok A. K. Modeling in IDEF0 notation // My professional career. - 2022. – Vol. 1, No. 36. – pp. 137-141.

29. Karpychev V. Yu. Functional modeling (IDEF0) as a research method for blockchain technology // Proceedings of the R.E. Alekseev NSTU. – 2018. – № 4(123). – Pp. 22-32.

30. Suriya Dr. S., S. N. Design of UML Diagrams for WEBMED - Healthcare Service System Services // EAI Endorsed Transactions on e-Learning. – 2023. – Vol. 8, No. 1. – P. e5.

31. Barclayevskaya N. V. The use of the unified modeling language UML in the project approach when teaching students in the specialty "business informatics" // Materials of the scientific and methodological conference of the SZIU RANHiGS. – 2015. – No. 1. – pp. 21-29.

Информация об авторах

Сергей Иванович Носков – д. т. н., профессор, профессор кафедры «Информационные системы и защита информации», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: sergey.noskov.57@mail.ru

Даниил Владимирович Пашков – аспирант кафедры «Информационные системы и защита информации», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: dvp-2000@mail.ru

Authors

Sergey Ivanovich Noskov, Doctor of Technical Science, Professor, the Subdepartment Information systems and information security, Irkutsk State Transport University, Irkutsk, e-mail: sergey.noskov.57@mail.ru

Daniel Vladimirovich Pashkov – postgraduate student of the Department Information Systems and Information Security, Irkutsk State Transport University, Irkutsk, e-mail: dvp-2000@mail.ru

Для цитирования

Носков С.И., Пашков Д.В. Проектирование информационной системы реализации некоторых этапов построения регрессионных моделей // «Информационные технологии и математическое моделирование в управлении сложными системами»: электрон. науч. журн. – 2024. – №2. – С. 1-14 – Режим доступа: http://ismm-irgups.ru/toma/2222024, свободный. – Загл. с экрана. – Яз. рус., англ. (дата обращения: 19.06.224)

For citations

Noskov S.I., Pashkov D. V. Designing an information system for the implementation of some stages of building regression models // Informacionnye tehnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami: ehlektronnyj nauchnyj zhurnal [Information technology and mathematical modeling in the management of complex systems: electronic scientific journal], 2024. No. 2. P. 1-14. [Accessed 19/06/24]

УДК 519.87

Г. Д. Гефан [1], В. С. Попова [1], Н. С. Попова [1]

[1] Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация

МЕТОД РАЗЛИЧЕНИЯ СХОДНЫХ ПОЧЕРКОВ С ПОМОЩЬЮ ЛИНЕЙНОГО КЛАССИФИКАТОРА

Аннотация. Метод простого линейного классификатора (ПЛК) предложено использовать для решения задачи различения сходных почерков. Почерк – уникальное свойство каждого человека, по которому возможна идентификация личности. Однако различить сходные почерки непросто, особенно при большом объеме данных. Результаты экспериментов, приведённые в работе, показали, что метод ПЛК позволяет достичь достаточно высокого уровня точности и надёжности при классификации графических образцов и может быть эффективным инструментом для различения сходных почерков.

Ключевые слова: задачи классификации, метод опорных векторов, линейный классификатор, различение сходных почерков, моделирование в Python.

G.D.Gefan [1], V.S. Popova [1], N.S.Popova [1]

[1] Irkutsk State Transport University, Irkutsk, Russian Federation

METHOD OF RECOGNIZING SIMILAR HANDWRITING USING A LINEAR CLASSIFIER

Abstract. The method of a simple linear classifier (SLC) is proposed to be used to solve the problem of recognizing similar handwritings. Handwriting is unique to each person, it can be used to identify a person. However, to recognize similar handwritings is not easy, especially with a large amount of data. The experimental results presented in this paper show that the SLC method can achieve a sufficiently high level of accuracy and reliability in the classification of graphical samples and can be an effective tool for recognizing similar handwritings.

Keywords: classification problems, Support Vector Machines, linear classifier, recognition of similar handwritings, modeling in Python.

Введение. На протяжении многих лет машинное обучение использовалось для решения множества задач в таких областях, как медицинская диагностика, техническая диагностика (компьютерное зрение, распознавание речи), экономика (кредитный скоринг, обнаружение мошенничества, биржевой анализ), офисная автоматизация (распознавание текста или рукописного ввода, обнаружение спама, категоризация документов) [1]. В общих чертах выделяют задачи классификации, кластеризации и регрессии.

Задачи классификации и их практическое значение. **Методы классификации**. Задачи классификации возникают в случае, когда необходимо присвоить объекту метку принадлежности к определённому классу на основе различных характеристик, называемых признаками [2, 3]. Классификация объектов находит широкое применение в системах безопасности, в управлении и контроле доступа, в системах по распознаванию знаков, человеческих лиц.

Существует немало методов классификации. К наиболее распространенным методам относятся: деревья решений, метод k-ближайших соседей, наивный байесовский классификатор, логистическая регрессия и метод опорных векторов. Последний метод рассмотрим чуть подробнее.

Основная идея метода опорных векторов (Support Vector Machine, SVM) состоит в построении линии (или гиперплоскости), оптимально разделяющей объекты выборки на два класса [4]. Необходимо разделить множества некоторой полосой так, чтобы, во-первых, эта полоса была как можно шире (для лучшего разделения двух классов), и, во-вторых, чтобы были минимизированы ошибки разделения [5-7]. Перечисленные оптимизационные требования (максимизации ширины полосы и минимизация ошибок) противоречат друг другу, и критерий оптимизации должен быть сконструирован так, чтобы можно было регулировать их

относительную важность. Векторы, оказывающиеся на границах разделительной полосы, называются опорными.

Описание простого линейного классификатора (ПЛК), его преимущества. Несмотря на свою популярность, SVM обладает рядом недостатков. Во-первых, он неустойчив по отношению к «шуму» (ошибочным точкам) в исходных данных. Если обучающая выборка содержит шумовые выбросы («объекты-нарушители»), то они будут существенным образом учтены при построении разделяющей гиперплоскости [8]. Во-вторых, есть регулирующий параметр алгоритма C, который надо подбирать [9, 10]. Более того, число переменных при решении задачи оптимизации равно числу обучающих векторов, что приводит к замедлению процесса обучения.

Алгоритм ПЛК следующий. Пусть имеется 2 класса тренинговых (обучающих) векторов $\mathbf{x_i}$. Присваиваем им метки z_i (одному классу +1, другому -1). Задаем условие оптимизации (1):

$$\sum_{i=1}^{n}(\mathbf{x}_i \cdot \mathbf{w} - b)z_i \to \max. \quad (1)$$

Здесь величина b – неизвестное расстояние от начала координат до границы (2):

$$b = \tfrac{1}{2}\left[\overline{\mathbf{x}\cdot\mathbf{w}}\Big|_{(1)} + \overline{\mathbf{x}\cdot\mathbf{w}}\Big|_{(2)}\right], \quad (2)$$

где слагаемые в скобках соответствуют усреднённым скалярным произведениям векторов одного и другого классов на неизвестный нормальный вектор разделительной гиперплоскости $\mathbf{w}$. Задаем ограничение вида $|\mathbf{w}| = 1$. Находим оптимальный нормальный вектор $\mathbf{w}^*$ и величину b^* как решение сформулированной задачи оптимизации.

Для классификации нового вектора $\mathbf{x}$ находим $\mathbf{x}\cdot\mathbf{w}^*$. Если величина $\mathbf{x}\cdot\mathbf{w}^* - b^* < 0$, то относим этот вектор к классу $z_i = -1$, иначе – к классу $z_i = 1$ [11].

Следовательно, преимущество данного метода состоит в том, что задача не зависит от какого-либо дополнительного параметра в отличие от метода опорных векторов, в котором необходимо задавать управляющий параметр C. Число переменных при решении задачи оптимизации равно размерности векторного пространства и не зависит от числа тренинговых векторов, что кардинально сокращает требуемые ресурсы.

Тестирование ПЛК на случайных векторах (двумерное нормальное распределение). Для тестирования ПЛК используем набор случайных двумерных векторов, координаты которых имеют нормальное распределение. Через a и σ обозначим математическое ожидание и среднеквадратическое отклонение (СКО) нормального распределения соответственно (таблица 1).

Таблица 1.

Данные для тестирования ПЛК

	1-ый класс		2-ой класс	
	X_1	X_2	X_1	X_2
a	3	5	5	7
σ	1	1	1	1

Решая задачу оптимизации, найдем параметры границы. Также посмотрим, сколько «ошибок» возникает, когда вектор одного класса опознаётся как вектор другого класса. Для этого возьмем 25 тренинговых векторов каждого класса. На графике (рис. 1) показаны данные двух классов, обозначенных красными и синими кружками, вместе с границей решения (линия черного цвета), найденной методом ПЛК. Из рисунка видно, что в результате работы классификатора количество «ошибок» равно 4 (2 + 2).

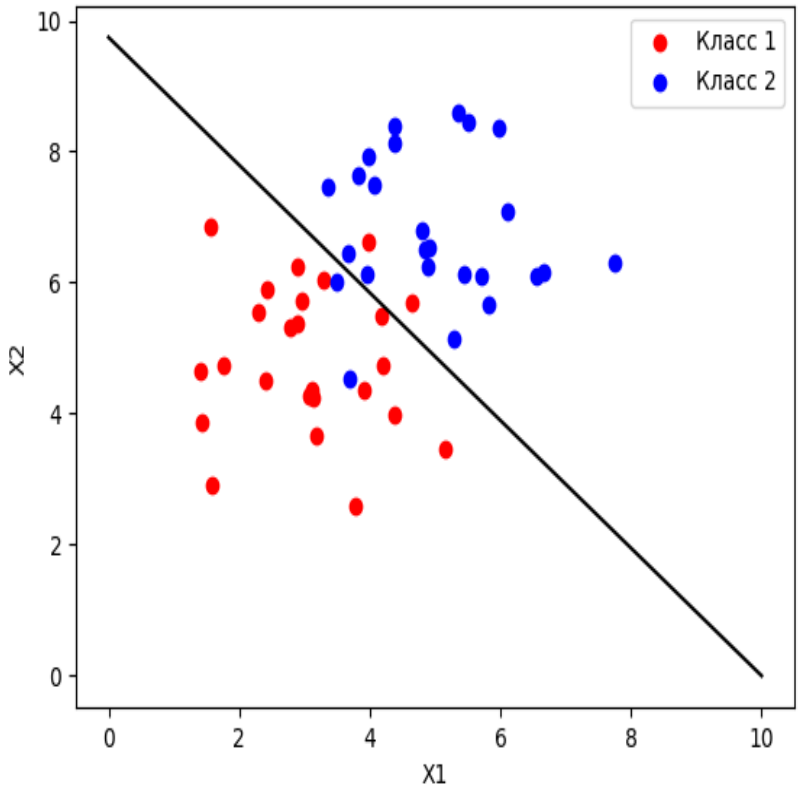

Рис. 1. Результат эксперимента со случайными векторами (двумерное нормальное распределение)

В таблицу 2 сведены результаты двадцати таких экспериментов, средние значения параметров границы по всей серии, а также теоретические параметры границы, которые в этом случае из соображений симметрии легко получить: $x_{(1)} + x_{(2)} = 10$ или $\mathbf{w}^* = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$, $b^* = \frac{10}{\sqrt{2}}$.

Таблица 2.

Результаты экспериментов

№	b^*	$\mathbf{w}^*$	
1	7,026	0,712	0,702
2	7,302	0,653	0,757
3	6,877	0,780	0,626
4	6,974	0,698	0,716
5	6,864	0,828	0,560
6	7,261	0,595	0,804
7	6,763	0,732	0,681
8	7,158	0,746	0,666
9	7,262	0,621	0,783
10	6,516	0,845	0,535
11	7,225	0,676	0,737
12	6,835	0,788	0,615
13	7,041	0,730	0,683
14	7,110	0,645	0,764
15	7,089	0,683	0,730

16	7,098	0,730	0,683
17	6,785	0,753	0,658
18	7,150	0,709	0,705
19	6,937	0,667	0,745
20	7,338	0,628	0,778
Среднее	7,031	0,711	0,696
Теоретическое	7,071	0,707	0,707

Таким образом, убеждаемся, что полученные оценки нормального вектора и расстояния до границы от начала координат не смещены, то есть колеблются вокруг теоретических значений.

Поэкспериментируем с другими классами векторов. Посмотрим, что будет происходить, если изменять только значение σ (например, в таблице 1 уменьшим её значение до 0,1). Результат приведен на рис. 2.

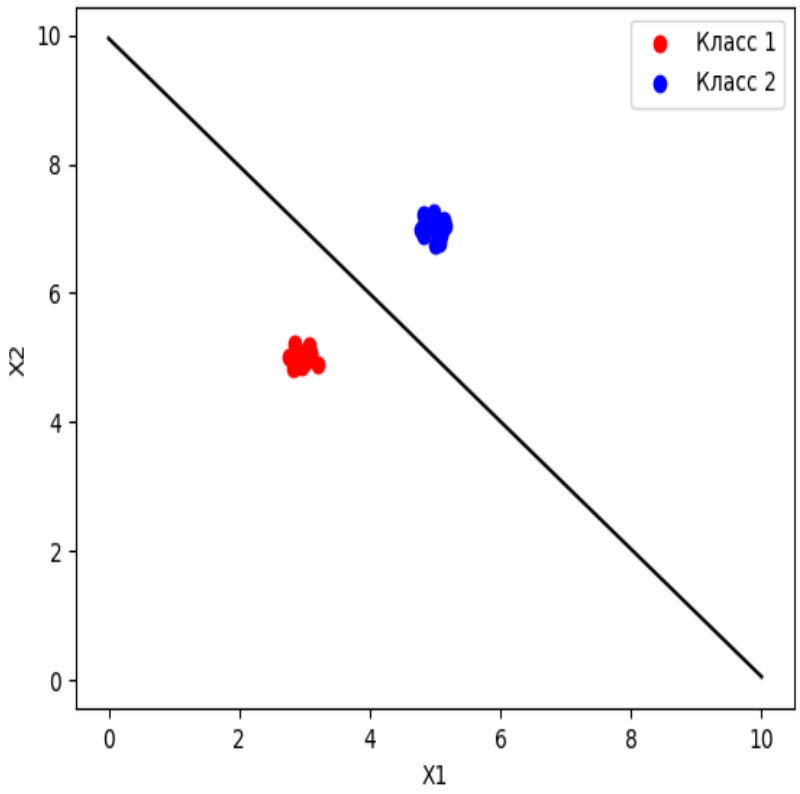

Рис. 2. Эксперимент с СКО равным 0,1

Очевидно, при одних и тех же значениях a с увеличением значения σ количество «ошибок» увеличивается, а при уменьшении значения σ – уменьшается. С изменением a центры классов могут располагаться достаточно близко (далеко) относительно друг друга, что может увеличить (уменьшить) количество «ошибок». Помимо этого возможно изменение теоретической границы.

В нашем случае модель реализовывалась на языке программирования Python, а генерация случайных векторов – с помощью его дополнительного модуля numpy.

Постановка задачи различения сходных почерков. Построение модели. Итак, решим одну из широко распространенных задач в системах по распознаванию рукописных текстов – задачу различения сходных почерков – методом классификации данных, основанным на

решении задачи линейного программирования. На данном этапе задача имеет следующую постановку: по имеющимся k росписям двух лиц построить модель, которая с наибольшей надёжностью определяла бы, какому из двух лиц принадлежит каждая роспись [12].

Для эксперимента были отобраны два лица («А» и «Б» – близнецы) со сходными почерками. Очевидно, что для одного и того же человека характерен некоторый разброс характеристик почерка от одного акта к другому. Так, было сфотографировано по 60 букв «з», написанных каждым лицом (то есть в нашем случае k равно 120). На рис. 3 приведены образцы почерков.

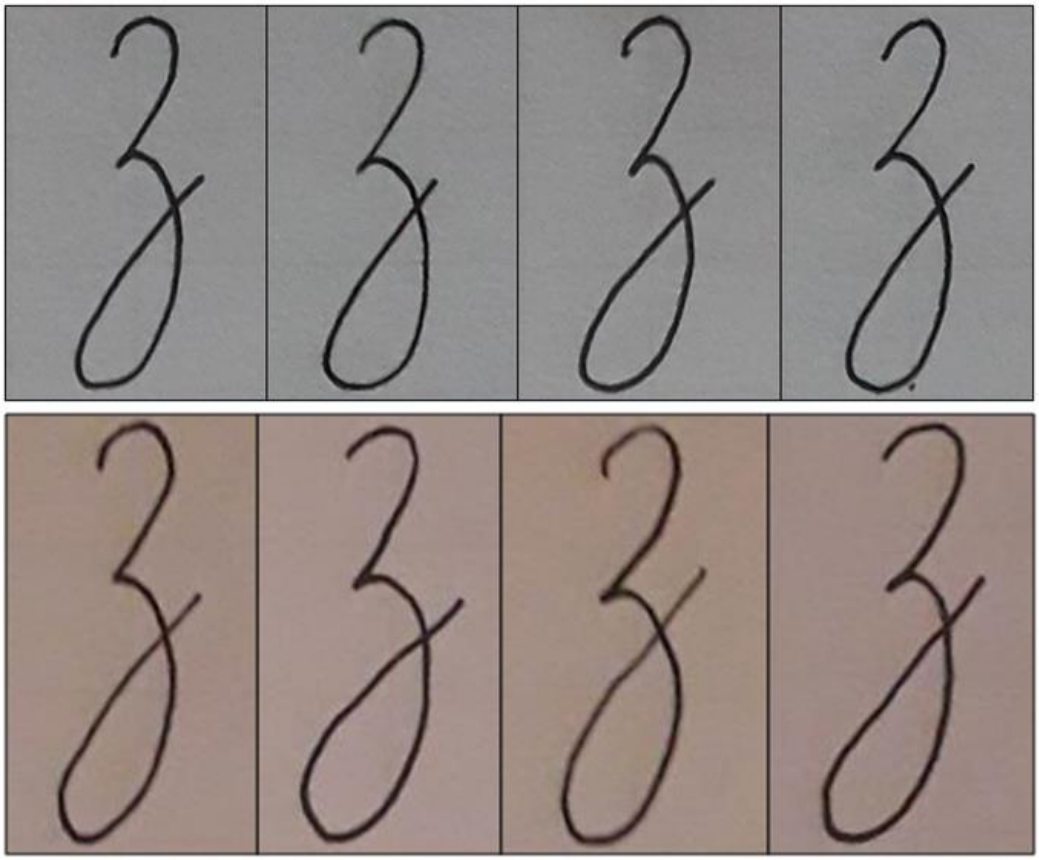

Рис. 3. Примеры почерков лиц «А» и «Б» соответственно

Все фотоснимки букв масштабировались и центрировались одинаково (рис. 4). Обработку получившихся изображений проводили с помощью языка программирования Python.

Сначала для получения векторов необходимо определить контур почерка, то есть кривую, соединяющую все непрерывные точки вдоль границы объекта. Затем – контрольные точки, на которых будет основываться способ параметризации векторов. В данной работе для обнаружения краев была применена специально предназначенная для этого функция cv.Canny().

Чтобы определить векторы, поставим задачу выделить, например, четыре «контрольные» точки (рис. 4), которые имеются на каждой росписи. На всех изображениях эти точки выбираются единообразно.

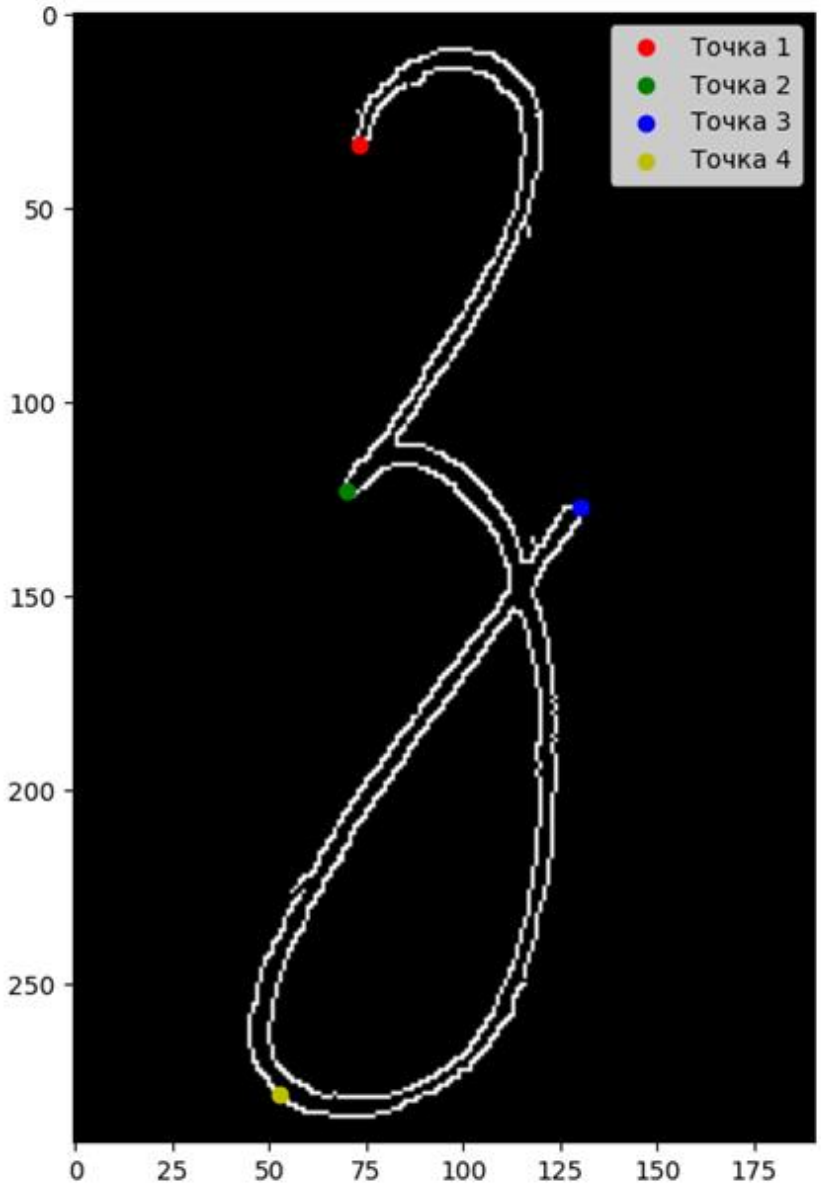

Рис. 4. Отображение «контрольных» точек

Далее рассчитываются расстояния между выбранными точками: 1-2, 2-3 и 3-4. Тем самым для каждой росписи получаем трёхмерный тренинговый вектор.

Эксперимент со 120 росписями дал следующие результаты для нормального вектора и расстояния до границы от начала координат (рис. 5): $\mathbf{w}^* = (0{,}379 \quad -0{,}246 \quad -0{,}892)$; $b^* = -131{,}764$.

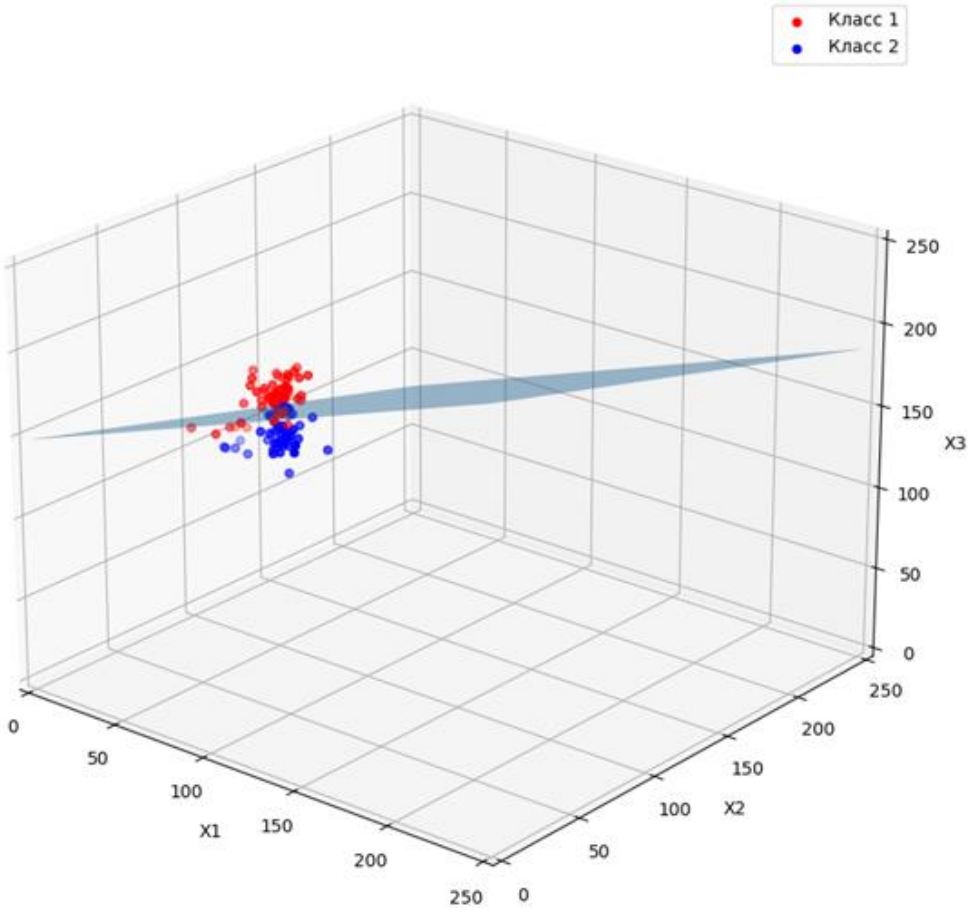

Рис. 5. Результат построения модели по 120 росписям двух лиц

Количество «ошибок» равно 20 (для 1-го класса их 9, для 2-го – 11).

Таким образом, доля правильных опознаний почерка составляет примерно 83,3%. Это указывает на то, что алгоритм классификатора прошел проверку успешно и может быть применен для последующих экспериментов.

Тестирование модели. Для более «честной» оценки пропустим через созданную модель уже новый, тестовый набор из 70 векторов от лица «А» и 30 векторов от лица «Б».

Результат работы ПЛК следующий: для 1-го класса количество ошибок равно 8, для 2-го – 8. Тем самым надёжность распознавания составила 84%.

Выводы. **Перспективы**. **Оценка эффективности метода для реальной ситуации**. В данной работе был создан и протестирован алгоритм ПЛК при решении задачи различения сходных почерков двух лиц (пары близнецов). Качество распознавания составляет около 84%. Модель реализована на языке программирования высокого уровня, Python.

Наша модель имеет узкую область применения, поскольку построена на данных образцах почерка пары близнецов (может применяться только в отношении них везде, где требуется аутентификация автора росписи: кем именно из близнецов была выполнена роспись).

В дальнейшем этот алгоритм может быть использован в качестве инструмента для проверки подлинности пользователя (речь идет о создании новой модели). Ведь различение сходных почерков является важной задачей в области криминалистики. Так, например, мошенники часто подделывают подписи в документах (договорах купли-продажи, дарения, завещаниях, накладных и других). Задача заключается в определении различий и сходств между почерками

разных людей с целью идентификации личности или выявления подделки. Допустим, первый класс остается неизменным (эталонным), а второй класс изменяемым (поддельным). В качестве образцов второго класса будут добавляться данные образцов почерка других лиц, которые пытаются подражать эталонной подписи.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

32. Журавлев Ю.И., Рязанов В.В., Сенько О.В. Распознавание. Математические методы. Программная система. Практические применения. – М.: Фазис, 2005. – 159 с.

33. Кугаевских А.В., Муромцев Д.И., Кирсанова О.В. Классические методы машинного обучения. – СПб: Университет ИТМО, 2022. – 53с.

34. Marsland S. Machine Learning: An Algorithmic Perspective. CRC Press. – 2009. – 406 p.

35. Vapnik V.N. The Nature of Statistical Learning Theory. – Berlin : Springer – Verlag, 1995. – 334 p.

36. Brink H., Richards J., Fetherolf M. Real-World Machine Learning. Manning. – 2016. – 264 p.

37. Вьюгин В.В. Элементы математической теории машинного обучения: учебное пособие. – М.: МФТИ: ИППИ РАН, 2010. – 252 с.

38. Shalev-Shwartz S., Ben-David S. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press. – 2014. – 410 p.

39. Harrington P. Machine Learning in Action. Manning. – 2012. – 384 p.

40. Nefedov A. Support Vector Machines: A Simple Tutorial, 2016. – 35 p.

41. Bishop M. Christopher Pattern Recognition and Machine Learning. Springer. – 2006. – 738 p.

42. Гефан Г.Д., Иванов В.Б. Метод опорных векторов и альтернативный ему простой линейный классификатор // Информационные технологии и проблемы математического моделирования сложных систем. – Иркутск : ИрГУПС, 2012. – Вып. 10. – С. 84-94.

43. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов (статистические проблемы обучения). – М.: Наука, 1974. – 416 с.

REFERENCES

32. Zhuravlev Yu.I., Ryazanov V.V., Sen'ko O.V. *Raspoznavanie. Matematicheskie metody. Programmnaya sistema. Prakticheskie primeneniya* [Recognition. Mathematical methods. Program system. Practical applications]. Moscow, «Fazis» Publ., 2005, 159 p.

33. Kugaevskikh A.V., Muromtsev D.I., Kirsanova O.V. *Klassicheskie metody mashinnogo obucheniya* [Classical machine learning methods]. Saint Petersburg, 2022, 53 p.

34. Marsland S. Machine Learning: An Algorithmic Perspective. CRC Press. – 2009. – 406 p.

35. Vapnik V.N. The Nature of Statistical Learning Theory. – Berlin : Springer – Verlag, 1995. – 334 p.

36. Brink H., Richards J., Fetherolf M. Real-World Machine Learning. Manning. – 2016. – 264 p.

37. V'yugin V.V. *Elementy matematicheskoy teorii mashinnogo obucheniya: uchebnoe posobie* [Elements of the mathematical theory of machine learning: student textbook]. Moscow: MFTI: IPPI RAN, 2010, 252 p.

38. Shalev-Shwartz S., Ben-David S. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press. – 2014. – 410 p.

39. Harrington P. Machine Learning in Action. Manning. – 2012. – 384 p.

40. Nefedov A. Support Vector Machines: A Simple Tutorial, 2016. – 35 p.

41. Bishop M. Christopher Pattern Recognition and Machine Learning. Springer. – 2006. – 738 p.

42. Gefan G.D., Ivanov V.B. *Metod opornykh vektorov i al'ternativnyy yemu prostoy lineynyy klassifikator* [Support vector machine and its alternative simple linear classifier]. *Informacionnye tekhnologii i problemy matematicheskogo modelirovaniya slozhnyh system* [Information technologies

and problems of mathematical modeling of complex systems]. Irkutsk, IrGUPS, 2012, no. 10, pp. 84-94.

43. Vapnik V.N., Chervonenkis A.Ya. *Teoriya raspoznavaniya obrazov (statisticheskie problemy obucheniya)* [Theory of pattern recognition (statistical learning problems)]. Moscow, «Nauka» Publ., 1974, 416 p.

Информация об авторах

Григорий Давыдович Гефан – к. ф.-м. н., доцент кафедры «Математика», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: grigef@rambler.ru

Виктория Сергеевна Попова – студентка гр. БАС.5-22-1, Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: popovavika2017@yandex.ru

Надежда Сергеевна Попова – студентка гр. БАС.5-22-1, Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: Nm2nadia@yandex.ru

Authors

Grigory Davydovich Gefan – candidate of physical and mathematical sciences, associate professor of department of mathematics, Irkutsk State Transport University, Irkutsk, e-mail: grigef@rambler.ru

Victoria Sergeevna Popova – student, Irkutsk State Transport University, Irkutsk, e-mail: popovavika2017@yandex.ru

Nadezhda Sergeevna Popova – student, Irkutsk State Transport University, Irkutsk, e-mail: Nm2nadia@yandex.ru

Для цитирования

Гефан Г.Д., Попова В.С., Попова Н.С. Метод различения сходных почерков с помощью линейного классификатора // «Информационные технологии и математическое моделирование в управлении сложными системами»: электрон. науч. журн. – 2024. – №2. – С. 15-23 – Режим доступа: http://ismm-irgups.ru/toma/2222024, свободный. – Загл. с экрана. – Яз. рус., англ. (дата обращения: 19.06.224)

For citations

Gefan G.D., Popova V.S., Popova N.S. Method of recognizing similar handwriting using a linear classifier // Informacionnye tehnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami: ehlektronnyj nauchnyj zhurnal [Information technology and mathematical modeling in the management of complex systems: electronic scientific journal], 2024. No. 2. P. 15-23. [Accessed 19/06/24]

УДК *004.514*

Кириллова Т.К.[1], Знайдюк А.Н.[1], Павлов П.С.[1],
[1] *Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация*

ПРОЕКТИРОВАНИЕ ВЕБ-ПРИЛОЖЕНИЯ МОНИТОРИНГА ОПАСНОСТИ ВОЗМОЖНОГО РАЗМЫВА УЧАСТКОВ ЖЕЛЕЗНОЙ ДОРОГИ НА ОСНОВЕ ГЕОИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Аннотация. *В статье рассматривается результаты проектирования программного обеспечения в форме веб-приложения для мониторинга опасности возможного размыва участков земляного полотна и искусственных сооружений Улан-Баторской железной дороги, с учётом возможностей современных геоинформационных технологий, а также возможность мониторинга состояния искусственных сооружений (ИССО). Веб-приложение дает возможность централизованно и оперативно информировать всех участников об изменении погодных условий и характеристиках пути. Для обеспечения безопасного и бесперебойного движения поездов*

железная дорога должна быть защищена от размывов. Размывы в летние месяцы являются серьезным препятствием для ритмичной работы. В настоящее время, когда планируется серьезное повышение размеров перевозок как местных, так и транзитных в направлении международного коридора Россия – Монголия – Китай, обеспечение сохранности грузов и пассажиров, безусловное выполнение графика движения поездов являются залогом повышения доходности работы дороги.

Ключевые слова: *веб-приложение, геоинформационные технологии, размыв пути, выпавшие осадки, железнодорожный путь.*

T.K. Kirillova[1], A.N. Znaidyuk, P.S. Pavlov
[1] *Irkutsk State Transport University, Irkutsk, the Russian Federation*

DESIGNING A DIGITAL ECOSYSTEM OF TOURISM ACTIVITIES

Abstract. The article considers the results of the design and development of software in the form of a web application for monitoring the danger of possible erosion of sections of the roadbed and artificial structures of the Ulaanbaatar railway, taking into account the capabilities of modern geoinformation technologies, as well as the possibility of monitoring the state of the ISSO. The web application makes it possible to centrally and promptly inform all participants about changes in weather conditions and path characteristics. To ensure the safe and uninterrupted movement of trains, the railway must be protected from washouts. Washouts in the summer months are a serious obstacle to rhythmic work. At present, when it is planned to seriously increase the size of both local and transit traffic in the direction of the international corridor Russia – Mongolia – China, ensuring the safety of goods and passengers, unconditional fulfillment of the train schedule are the key to increasing profitability.

Keywords: web application, geoinformation technologies, track erosion, precipitation, railway track.

Введение. Актуальной является задача автоматизации системы мониторинга опасности возможного размыва участков земляного полотна и искусственных сооружений, обеспечение безопасности во внештатных ситуациях дороги в соответствии с комплексом требований, таких, как минимизация времени реагирования, сокращение времени оповещения об инциденте и т.д. Вопросы проектирования веб-ориентированных архитектур программных решений на основе геоинформационных систем освещаются в трудах отечественных ученых и специалистов, среди которых Воробьева Г.Р. Автор указывает на преимущества визуализации объектов благодаря использованию геоинформационных технологий, которые предоставляют больше информации, чем другие известные системы или технологии. Важность этой функциональности подчеркивается динамическими свойствами и многоуровневым масштабированием геопространственного изображения. В научных работах сформулированы и теоретически обоснованы научные подходы к проблемам создания сервис-ориентированных программных решений [1,2].

Постановка задачи. Цель исследования – рассмотреть проектирование программного обеспечения в форме веб-приложения для мониторинга опасности возможного размыва участков земляного полотна и искусственных сооружений Улан-Баторской железной дороги. Задачами является осуществление проектирования программного обеспечения с учётом возможностей современных геоинформационных технологий, и обеспечением возможности мониторинга состояния искусственных сооружений (ИССО). Разработка модели прогнозирования рисков размыва железнодорожного пути выполнена с использованием методики факторного анализа рисков размывов пути, предложенной специалистами службы пути АО «Улан-Баторская железная дорога» и доработанная специалистами университета путей сообщения в 2024 году [3].

Для обеспечения безопасного и бесперебойного движения поездов железная дорога должна быть защищена от размывов [5-6]. Размывы в летние месяцы являются серьезным препятствием для ритмичной работы УБЖД. В настоящее время, когда планируется серьезное повышение размеров перевозок по УБЖД как местных, так и транзитных в направлении международного коридора Россия – Монголия – Китай, обеспечение сохранности грузов и пассажиров, безусловное выполнение графика движения поездов являются залогом повышения доходности работы дороги [7].

Согласованная функциональность веб-приложения:

– при вводе данных об ожидаемом количестве осадков (ручной и автоматический ввод) программа прогнозирует места (участки дороги), где есть риск размыва пути с указанием степени опасности (жёлтая зона - средний уровень риска, оранжевая зона – опасная, с вероятностью размыва, красная зона – очень высокий уровень риска);

– выдавать рекомендации о необходимости принятия управленческих решений при оперативном планировании мероприятий;

– формирование отчёта в двух форматах в PDF, EXCEL;

– организация автоматических уведомлений о рисках размыва пути, выводимые на рабочий стол пользователя;

– ввод исходных данных в базу данных приложения осуществляется полностью специалистами УБЖД;

– рассылка публичных и личных сообщений пользователям;

– выдача отчётов и построение аналитических графиков за разные промежутки времени, возможность делать выборку по рисковым километрам дороги.

Проектирование. Этапы проектирование веб-приложения:

- проведен анализ предметной области и выбран стек технологий;
- осуществлено проектирование веб-приложения;
- разработан прототип интерфейса;

Для реализации веб-приложения выбран язык программирования Python, фреймворк Django, HTML, CSS. Системой управления базами данных (СУБД) был выбран PostgreSQL, так как это СУБД с открытым исходным кодом. С помощью PostgreSQL можно создавать, хранить базы данных и работать с данными с помощью запросов на языке SQL [9-12].

Алгоритм прогнозирования рисков размыва. Процесс начинается, с того, что гидрометцентр Монголии присылает прогноз погоды по участкам. Для корректной работы алгоритма данные необходимо обработать. После обработки полученные данные заносятся в базу данных и далее считаются, как динамические т.к. он изменяются каждый день. После обработки данных гидрометцентра, происходит подсчет рисков размыва железнодорожного пути. Для этого понадобиться получить раннее внесённые данные и параметры объектов ИССО, после чего происходит подсчет риска размыва и занесения его в базу данных. В самом конце сотрудникам выдаются управленческие рекомендации в зависимости от балла. Блок схема представлена на рисунке 1.

Проект веб-приложения программы «Система контроля опасности размыва дороги» («СКОРД»), предполагается спроектировать с использованием геоинформационных систем (ГИС, географическая информационная система) — это компьютерные технологии, которые применяют для создания карт и оценки фактически существующих объектов, а также происшествий. Информация на карту наносится слоями, что позволяет любой слой данных добавить, или удалить, что делает обновление удобнее. Пользователь может указать место или объект на цифровой карте, чтобы найти информацию о нем.

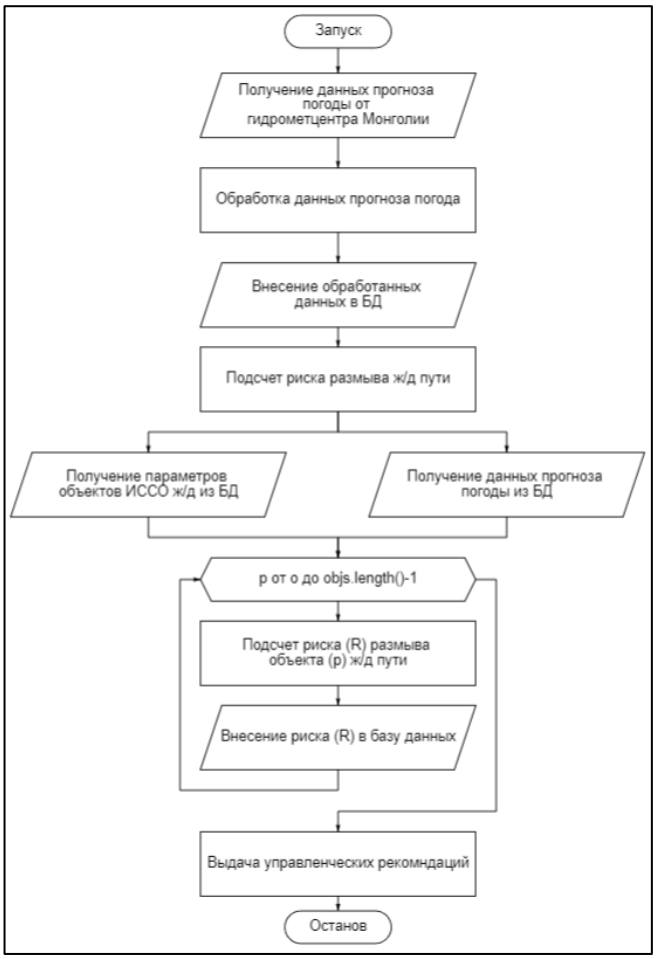

Рисунок 1. Блок схема алгоритма

Например, пользователь может щелкнуть на значок станции, чтобы узнать дополнительные количественные данные по ней (состояние земляного полотна, характеристика пути, наличие ИССО и т.п.) Характеристики железнодорожных станций, включенные в описательную часть программы, освещаются в трудах ученых [13-14]. Такие системы собирают, хранят и анализируют информацию, а также обеспечивают ее графическую интерпретацию, подобная логика реализована в базе данных мобильного приложения мониторинга технического состояния локомотива и ремонтных работ [15]. При запуске веб-приложения пользователя встречает форма входа в систему. Окно авторизации предоставляет пользователю возможность входить в систему имеет поля: «Логин», «Пароль». Также кнопка «Войти» и «Забыли пароль?», как показано на рисунке 2.

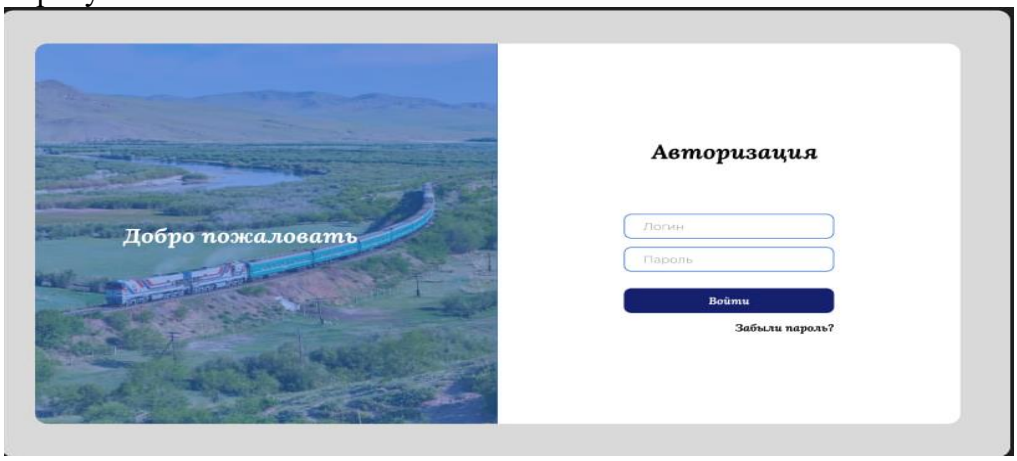

Рисунок 2. Страница входа в систему

Для визуализации объектов ИССО предусмотрена возможность прикреплять несколько фото и один видеоматериал на карту, рисунок 3.

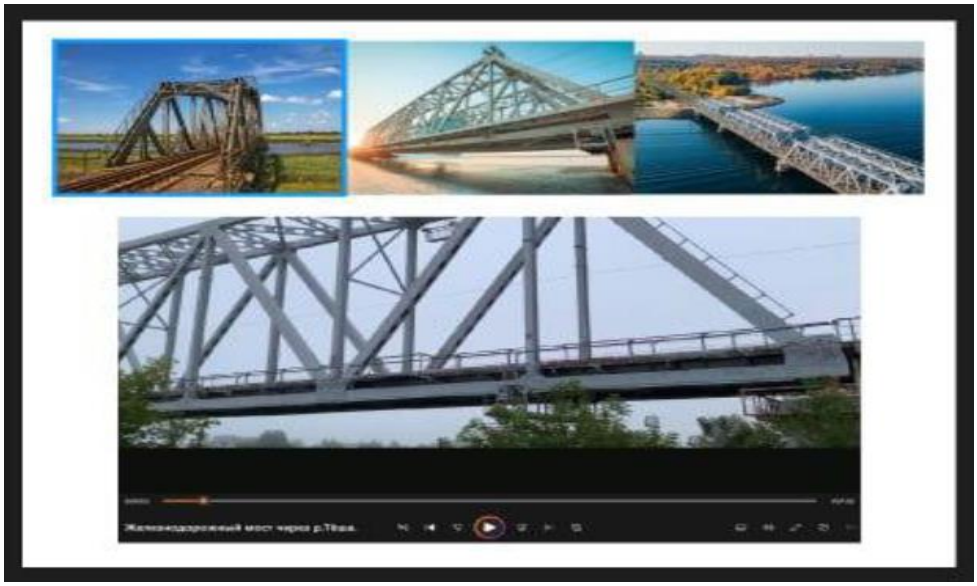

Рисунок 3. Пример функции визуализация ИССО

Окно с интерактивной картой и отметками участков, а также списком станций справа. При нажатии на маркер, отметка выделяется, а объект в списке станций разворачиваться, как показано на рисунке 4.

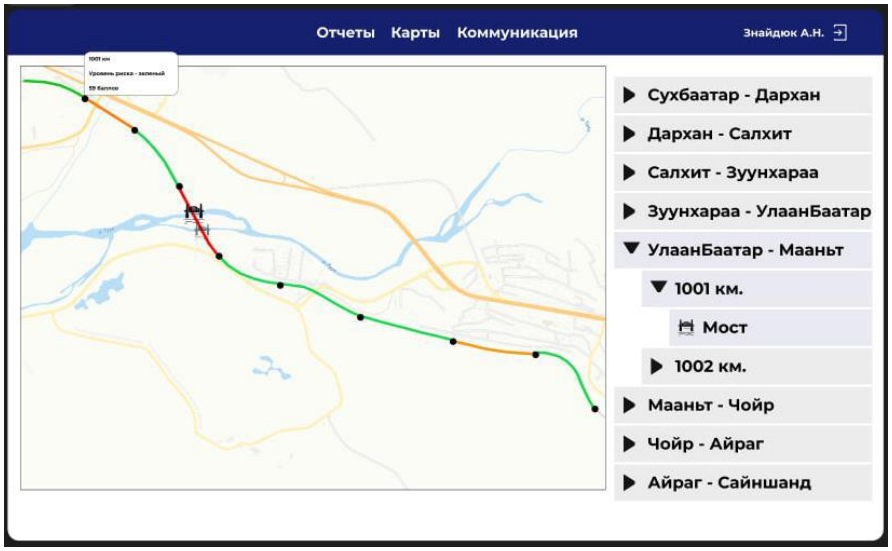

Рисунок 4. Функция «Нажатие на маркер»

Заключение. Веб-приложение дает возможность централизованно и оперативно информировать всех участников об изменении погодных условий и характеристиках пути. Внедрение веб-приложения даст следующие результаты и функции:

– формирование единой базы данных о техническом состоянии пути, сбор аналитики и последующая ее обработка;

– ремонтный персонал, сможет в режиме «онлайн» дать пояснения по устранению неисправности в пути следования;

– внедрение мессенджера позволит централизованно и оперативно отправлять информацию всем пользователям или выборочно по сформированным группам;

– уменьшение времени реагирования при наступлении опасных условий размыва пути;

– повышение эффективности мониторинга технического состояния пути;

– для руководителей есть возможность получения отчета в табличной и графической форме за выбранный календарный период.

Внедрение веб-приложения «Система контроля опасности размыва дороги», в долгосрочной перспективе повысит удобство работы благодаря упрощению процесса взаимодействия между работниками, единой базе данных, появится общее представление о всех инженерных сооружениях и других характеристиках пути.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Воробьев А.В., Воробьева Г.Р. Геоинформационная система динамической пространственной кластеризации распределенных источников данных // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2023. № 64. С. 61-73.

2. Пудовик Б.Р., Пушкаревский Ю.С. Особенности проектирования изделий с использованием систем автоматизированного проектирования// Системы управления и обработки информации. 2018. № 4 (43). С. 86-90.

3. Перфильева Е.В., Мелехова А.Д., Благоразумова О.В., Подвербный В.А. Программа «Элегия» для принятия решения в условиях риска// Фундаментальные и прикладные исследования в условиях геоплитической нестабильности. Материалы XXIII Всероссийской научно-практической конференции. Ростов-на-Дону, 2023. С. 11-14.

4. Кириллова Т.К. Управление рационализаторской деятельностью на Восточном полигоне как объект автоматизации // Экономика и предпринимательство. 2022. № 5 (142). С. 997-1000.

5. По приглашению Национального агентства по метеорологии и мониторингу окружающей среды Монголии делегация Забайкальского УГМС приняла участие в научно-практической конференции «Влияние изменений климата на режим и ресурсы трансграничных вод», которая состоялась в период 12–13 августа 2010 года в Монголии // URL: https://www.meteorf.gov.ru/press/news/3890/ (дата обращения: 24.06.2023).

6. Балжир Мунхдэлгэр Организация и развитие грузовых перевозок на сети Монгольской железной дороги : диссертация кандидата технических наук. – М.: МГУПС (МИИТ), 2015. – 129 с

7. Программа технической модернизации и развития АО «УБЖД» на период 2014 − 2020 годы / ОАО «ИЭРТ», ИПИИ «Иркутскжелдорпроект» − филиал ОАО «Росжелдорпроект», 2013. – 5 этап. – Том 2. – ПЗ. – Часть 1. – 263 с.

8. Абасова Н.И., Доржиева Э.Л., Кириллова Т.К., Нитежук М.С. Разработка информационной системы для управления программными проектами предприятия //Вестник Бурятского государственного университета. Экономика и менеджмент. 2022. № 4. С. 3-9.

9. Lyashenko I.I. About the use of case-technologies in the process of designing information systems//Вестник Инновационного Евразийского университета. 2022. № 2 (86). С. 126-133.

10. Кряжева Е.В., Дереглазов К.Ю. Проектирование интерфейса и выбор технологий реализации для веб-приложения «Путеводитель по городу» //Заметки ученого. 2021. № 13. С. 62-68.

11. Аршинский В.Л., Аршинский Л.В., Доржсурэн Х. Методика агрегированной оценки состояния производственно-экономической системы на примере станции Улан-Баторской железной дороги // Вестник Иркутского государственного технического университета. 2018. Т. 22. № 2 (133). С. 34-44.

12. Аршинский В.Л., Аршинский Л.В., Доржсурэн Х. Проблемы формирования и использования баз знаний при логико-аксиологической оценке систем на примере железнодорожной станции // Транспортная инфраструктура Сибирского региона. 2018. Т. 1. С. 390-396.

13. Аршинский Л.В., Хишигсурен Д. Разработка онтологии для агрегированного оценивания качества функционирования станции Улан-Баторской железной дороги //Транспортная инфраструктура Сибирского региона. 2017. Т. 1. С. 396-401.

14. Бурэн-Итгэл Г. Повышение эффективности использования автономных локомотивов для грузоперевозок на железных дорогах Монголии [Текст] дис. … канд. техн. наук: 2.4.2 /Гантумур Бурэн-Итгэл Московский энергет. институт М. – 2022. – 131 с.

15. Мунгунхуяг Г., Кириллова Т.К. Проектирование мобильного приложения мониторинга технического состояния локомотива и ремонтных работ ТО-2 на Улан-Баторской железной дороге// Молодая наука Сибири. 2023. № 3 (21). С. 123-129.

REFERENCES

1. Vorob'ev A.V., Vorob'eva G.R. Geoinformacionnaya sistema dinamicheskoj prostranstvennoj klasterizacii raspredelennyh istochnikov dannyh // Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika. 2023. № 64. P. 61-73.

2. Pudovik B.R., Pushkarevskij Yu.S. *Osobennosti proektirovaniya izdelij s ispol'zovaniem sistem avtomatizirovannogo proektirovaniya* [Sistemy upravleniya i obrabotki informacii]. 2018. № 4 (43). P. 86-90.

3. Perfil'eva E.V., Melekhova A.D., Blagorazumova O.V., Podverbnyj V.A. Programma «Elegiya» dlya prinyatiya resheniya v usloviyah riska// Fundamental'nye i prikladnye issledovaniya v usloviyah geopliticheskoj nestabil'nosti. Materialy XXIII Vserossijskoj nauchno-prakticheskoj konferencii. Rostov-na-Donu, 2023. S. 11-14.

4. Kirillova T.K. *Upravlenie racionalizatorskoj deyatel'nost'yu na Vostochnom poligone kak ob"ekt avtomatizacii* [Ekonomika i predprinimatel'stvo]. 2022. № 5 (142). S. 997-1000.

5. Po priglasheniyu Nacional'nogo agentstva po meteorologii i monitoringu okruzhayushchej sredy Mongolii delegaciya Zabajkal'skogo UGMS prinyala uchastie v nauchno-prakticheskoj konferencii «Vliyanie izmenenij klimata na rezhim i resursy transgranichnyh vod», kotoraya sostoyalas' v period 12–13 avgusta 2010 goda v Mongolii // URL: https://www.meteorf.gov.ru/press/news/3890/ (data obrashcheniya: 24.06.2023).

6. Balzhir Munhdelger Organizaciya i razvitie gruzovyh perevozok na seti Mongol'skoj zheleznoj dorogi : dissertaciya kandidata tekhnicheskih nauk. – M.: MGUPS (MIIT), 2015. – p. 129

7. Programma tekhnicheskoj modernizacii i razvitiya AO «UBZhD» na period 2014 − 2020 gody / OAO «IERT», IPII «Irkutskzheldorproekt» − filial OAO «Roszheldorproekt», 2013. – 5 etap. – Tom 2. – PZ. – Chast' 1. – 263

8. Abasova N.I., Dorzhieva E.L., Kirillova T.K., Nitezhuk M.S. *Razrabotka informacionnoj sistemy dlya upravleniya programmnymi proektami predpriyatiya* [Vestnik Buryatskogo gosudarstvennogo universiteta. Ekonomika i menedzhment]. 2022. № 4. P. 3-9.

9. Lyashenko I.I. *About the use of case-technologies in the process of designing information systems*. [Vestnik Innovacionnogo Evrazijskogo universiteta]. 2022. № 2 (86). P. 126-133.

10. Kryazheva E.V., Dereglazov K.Yu. *Proektirovanie interfejsa i vybor tekhnologij realizacii dlya veb-prilozheniya «Putevoditel' po gorodu»* [Zametki uchenogo]. 2021. № 13. P. 62-68.

11. Arshinskij V.L., Arshinskij L.V., Dorzhsuren H. *Metodika agregirovannoj ocenki sostoyaniya proizvodstvenno-ekonomicheskoj sistemy na primere stancii Ulan-Batorskoj zheleznoj dorogi.* [Vestnik Irkutskogo gosudarstvennogo tekhnicheskogo universiteta]. 2018. T. 22. № 2 (133). P. 34-44.

12. Arshinskij V.L., Arshinskij L.V., Dorzhsuren H. *Problemy formirovaniya i ispol'zovaniya baz znanij pri logiko-aksiologicheskoj ocenke sistem na primere zheleznodorozhnoj stancii* [Transportnaya infrastruktura Sibirskogo regiona]. 2018. T. 1. S. 390-396.

13. Arshinskij L.V., Hishigsuren D. *Razrabotka ontologii dlya agregirovannogo ocenivaniya kachestva funkcionirovaniya stancii Ulan-Batorskoj zheleznoj dorogi.* [Transportnaya infrastruktura Sibirskogo regiona]. 2017. T. 1. S. 396-401.

14. Buren-Itgel G. *Povyshenie effektivnosti ispol'zovaniya avtonomnyh lokomotivov dlya gruzoperevozok na zheleznyh dorogah Mongolii*. dis. … kand. tekhn. nauk: 2.4.2 /Gantumur Buren-Itgel Moskovskij energet. institut M. – 2022. – 131 P.

15. Mungunhuyag G., Kirillova T.K. *Proektirovanie mobil'nogo prilozheniya monitoringa tekhnicheskogo sostoyaniya lokomotiva i remontnyh rabot TO-2 na Ulan-Batorskoj zheleznoj doroge.* [Molodaya nauka Sibiri]. 2023. № 3 (21). P. 123-129.

Информация об авторах

Кириллова Татьяна Климентьевна - заведующий кафедрой «Информационные системы и защита информации», доцент, Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: kirillova_tk@irgups.ru

Знайдюк Алексей Николаевич - студент 4 курса направления подготовки «Программная инженерия», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: Znaidyuk00@gmail.com

Павлов Павел Сергеевич - студент 4 курса направления подготовки «Информационные системы и технологии», Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: Kakadyfff@mail.ru

Authors

Kirillova Tatiana Klimentevna, Head of the ISiZI Department, Associate Professor, Irkutsk State University of Railway Transport, Irkutsk, e-mail: kirillova_tk@irgups.ru

Znaidyuk Alexey Nikolaevich, 4th year student of the direction of training "Software Engineering", Irkutsk State Transport University, Irkutsk, e-mail: Znaidyuk00@gmail.com

Pavlov Pavel Sergeevich, 4th year student of the direction of training " Information systems and technologies", Irkutsk State Transport University, Irkutsk, e-mail: Kakadyfff@mail.ru.

Для цитирования

Кириллова Т.К., Знайдюк А.Н., Павлов П.С. Проектирование веб-приложения мониторинга опасности возможного размыва участков железной дороги на основе геоинформационных технологий // «Информационные технологии и математическое моделирование в управлении сложными системами»: электрон. науч. журн. – 2024. – №2. – С.23-30. – Режим доступа: http://ismm-irgups.ru/toma/2222024.

For citations

Kirillova T.K., Znajdyuk A.N., Pavlov P.S. Proektirovanie veb-prilozheniya monitoringa opasnosti vozmozhnogo razmyva uchastkov zheleznoj dorogi na osnove geoinformacionnyh tekhnologij // «Informacionnye tekhnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami»: elektron. nauch. zhurn. – 2024. – №2. – S.23-30. – Rezhim dostupa: http://ismm-irgups.ru/toma/2222024

УДК 004.056.52

П.В. Бабак [1], ***Т.К. Кириллова***[1], ***И.Е. Пинин*** [1]

[1] *Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация*

БЕЗОПАСНОСТЬ КРИПТОВАЛЮТ

Аннотация. В данном исследовании авторы фокусируются на описании угроз безопасности криптовалютных платформ. Выполнен обзор уязвимостей, которые могут быть обнаружены на криптовалютных

биржах, кошельках и смарт-контрактах. Задача исследования заключается в понимании механизмов, на которых эти платформы основаны, а также определения возможных проблем, связанных с ними. В фокусе исследования находятся уязвимости к атакам типа «51%», фишингу и социальной инженерии. Кроме того, исследование включает рекомендаций по улучшению безопасности системы, на основе результатов проведенных исследований. Результаты исследования будут полезны разработчикам, пользователям и регуляторам криптовалютных платформ, что особенно важно в свете участившихся случаев взломов и мошенничества в криптоиндустрии.

Ключевые слова: *криптовалюта, уязвимость, криптовалютная биржа, криптокошелек, смарт-контракт, «атака51%», фишинг, социальная инженерия, блокчейн.*

P.V. Babak [1], T.K. Kirillova [1], I.E. Pinin[1]

[1] Irkutsk State Transport University, Irkutsk, Russian Federation

CRYPTOCURRENCY SECURITY

Abstract. In this study, the authors focus on describing the security threats of cryptocurrency platforms. An overview of vulnerabilities that can be found on cryptocurrency exchanges, wallets and smart contracts has been performed. The aim of the study is to understand the mechanisms on which these platforms are based, as well as to identify possible problems associated with them. The study focuses on vulnerabilities to attacks such as "51%", phishing and social engineering. In addition, the study will include the proposal of recommendations for improving the security of the system, based on the results of the conducted research. The results of the study will be useful to developers, users and regulators of cryptocurrency platforms, which is especially important in light of the increasing cases of hacking and fraud in the crypto industry.

Keywords: *cryptocurrency, vulnerability, cryptocurrency exchange, crypto wallet, smart contract, «51% attack», phishing, social engineering, blockchain.*

Введение. Криптовалюты, основанные на технологии блокчейн [1], стремительно ворвались в мир финансов, предлагая новые возможности для транзакций, инвестиций и создания децентрализованных систем. Вместе с тем, стремительный рост популярности криптовалют сопровождается увеличением числа киберугроз, направленных на эксплуатацию уязвимостей в этой новой и сложной экосистеме.

Данное исследование посвящено анализу безопасности криптовалютных платформ [3-4], сфокусированном на выявлении уязвимостей, присущих криптовалютным биржам, кошелькам и смарт-контрактам. Актуальность исследования обусловлена участившимися случаями взломов и мошенничества в криптоиндустрии [2], что приводит к значительным финансовым потерям и подрывает доверие к криптовалютам как к надежному инструменту.

Целью исследования является глубокое понимание механизмов, лежащих в основе функционирования криптовалютных платформ, а также идентификация потенциальных уязвимостей, связанных с ними. Особое внимание будет уделено анализу следующих угроз:

- атаки 51%: Исследование рассмотрит возможность реализации атак 51% на различные криптовалюты, оценивая вероятность таких атак и их потенциальные последствия;
- фишинг: будут проанализированы распространенные методы фишинга, используемые злоумышленниками для кражи криптовалют, и предложены рекомендации по защите от них;
- социальная инженерия: Исследование рассмотрит тактики социальной инженерии, применяемые для получения доступа к криптовалютным активам, и предложит меры противодействия.

В работе рассмотрены сервисы децентрализованных финансов и атаки на эти сервисы. На основе сопоставления достоинств, недостатков и ограничений к применению для существующих решений, ряд авторов делают выводы о перспективах развития систем распределенного реестра, обеспечивающие конфиденциальность транзакций [15]. На основании полученных оценок предложены рекомендации по снижению рисков от угроз воздействия крипто вымогателей [17]. Полученные результаты могут быть применены для противодействия компьютерным преступлениям на корпоративные информационные системы.

Защита криптовалют, проблемы и решения. В целях исследования данной темы рассмотрим несколько статей от экспертов данной области. А далее представим свои исследования в данной области.

«Единственное, что стоит между злоумышленниками и вашей криптовалютой – установленные вами степени безопасности. Некоторые люди думают, что хакеру нет смысла атаковать конкретно их, но это не значит, что нужно легкомысленно относиться к защите. Зачастую злоумышленники атакуют не кого-то определенного, а сразу массу людей. А если случится потеря денег, то никто не сможет их вам вернуть. Эксперты Profinvestment изучили тему и подготовили ряд советов по безопасному хранению монет в некастодиальных кошельках, основываясь на собственном опыте и известных правилах [5]».

Предлагаемые решения:

- владелец криптовалютного кошелька должен быть единственным, кто имеет доступ к своим цифровым активам. Это подразумевает безопасное хранение приватного ключа и резервной фразы;
- аппаратные (холодные) кошельки считаются одними из самых безопасных для хранения криптовалют, поскольку они не подключены к интернету и, следовательно, менее подвержены хакерским атакам;
- горячие кошельки (мобильные, десктопные, веб), хотя и обеспечивают удобство доступа к цифровым активам в любое время, более подвержены риску взлома по сравнению с оффлайн кошельками;
- бумажные кошельки – другой вид оффлайн-хранения, где ключи для доступа к цифровым активам хранятся на бумаге. Они могут быть утеряны или уничтожены, поэтому важно хранить их в безопасном месте;
- некоторые пользователи предпочитают хранить свои цифровые активы на биржах, но это большой риск, биржа может стать жертвой взлома или заблокировать аккаунт по каким-то свои соображениям;
- владелец кошелька должен регулярно обновлять программное обеспечение и устанавливать надежные пароли/PIN-коды. Некоторые приложения позволяют ставить биометрическую защиту;
- наконец, следует знать о фишинговых атаках и других видах скама в криптоиндустрии [4].

«Лучший способ защиты биткоинов и других цифровых активов от кражи - хранить приватные ключи в холодном кошельке. Холодные кошельки не подключены к Интернету или даже другому устройству. Ни одно устройство хранения данных не является на 100% безопасным, но есть несколько методов, которым можно воспользоваться для защиты ключей от криптовалюты [6]».

Предлагаемые решения:

- холодные кошельки, также называемые холодным хранилищем, являются лучшим способом обезопасить ваши приватные ключи от биткоинов;
- некоторые биржи предоставляют безопасное холодное хранение ключей пользователей на институциональном уровне, но некоторые критики советуют отказаться от этого метода;
- некоторые биржи имеют страховку от кражи криптовалюты при определенных обстоятельствах, но охватываемые инциденты очень ограничены;
- совмещение использования холодного и горячего кошельков, чтобы в подключенном

кошельке была только криптовалюта, которая необходима в данный момент [6].

«Защита криптовалюты — необходимость любого пользователя. В настоящее время криптовалюта является привлекательным предметом воровства для хакеров. Для защиты своих средств необходимо быть предельно внимательными и следовать простым правилам».

Предлагаемые решения:

- для повседневных нужд рекомендуется использовать небольшую сумму. Распределение средств между несколькими депозитариями, сводит к нулю возможность кражи всех средств за один раз;
- создавать резервную копию кошелька. Хранение резервной копии криптокошелька позволит восстановить его в случае выхода компьютера из строя, а также если ваш мобильный телефон был украден;
- использование надежного пароля, а также офлайн-кошельки. Последние считаются наиболее надежными, так как позволяют хранить криптовалюту без выхода в интернет [7]. Таким образом, кошелек не может быть взломан злоумышленниками.

Анализ наиболее популярных уязвимостей. В данном разделе проведем анализ трех ключевых уязвимостей, характерных для криптовалютной сферы в виде таблицы:

Таблица 1.

Классификация уязвимостей

№	Уязвимость	Определение уязвимости	Основа уязвимости	Примеры атак	Методы защиты
1	Атака 51%	Ситуация, в которой злоумышленник или группа злоумышленников получают контроль над 51% вычислительной мощности сети блокчейна [11].	Криптовалюты с низким хешрейтом, такие как Bitcoin Gold, Verge, Ethereum Classic, наиболее уязвимы к атаке 51%.	В марте 2022 года хакеры взяли под контроль пять из девяти проверяющих узлов сайдчейна, связанного с эфириумом. Хакеры подделали вывод из сети 56.25 миллиардов рублей [8].	Proof-of-Work (PoW) - наиболее распространенный механизм защиты от атак 51% [12]. Proof-of-Stake (PoS) и Delegated Proof-of-Stake (DPoS) - альтернативные механизмы консенсуса, где влияние пользователей на сеть определяется количеством монет [13].
	Фишинг	Вид кибератаки, целью которой является получения конфиденциальных данных пользователя.	Злоумышленники используют различные уловки для введения пользователей в заблуждение.	2022 год - «год утечек». За год общий объем выставленных на продажу или размещенных в открытом доступе данных россиян превысил 2,8 терабайта [9].	-Распознавание фишинговых сайтов и писем (проверка доменных имен, ссылок, грамматики и т.д.). Использование антивирусного ПО.
	Социальная инженерия	Совокупность приемов манипуляций	Злоумышленники используют различные	В 2023 году мошенники выманили у GlowToken	Повышение осведомленности пользователей о тактиках социальной

		человеком или группой людей с целью получения выгоды [14].	психологические тактики.	LLC сумму около 22,5 млн. руб [10].	инженерии. Обучение сотрудников компаний.

Анализируя таблицу 1, приведем основные методы защиты от уязвимостей в криптосфере, в виде рисунка 1.

Рисунок 1. Методы защиты криптовалюты от различных уязвимостей

Заключение. В настоящем исследовании были рассмотрены ключевые уязвимости, присущие криптовалютной сфере, такие как атака 51%, фишинг и социальная инженерия. Анализ выявил, что, несмотря на растущую популярность и внедрение передовых технологий в индустрии криптовалют, существуют серьезные риски, связанные с безопасностью криптовалютных платформ.

Результаты исследования показали, что атака 51% представляет значительную угрозу для многих криптовалют, особенно для тех, которые имеют низкий хешрейт и децентрализованы в меньшей степени. Фишинг и социальная инженерия активно используются злоумышленниками для кражи криптовалютных активов, эксплуатируя человеческие слабости, такие как доверчивость, жадность и невнимательность.

В рамках данной работы были предложены эффективные методы защиты от выявленных уязвимостей. Обзор практических работ показал, что для предотвращения атак 51% рекомендуется использовать механизмы консенсуса, такие как Proof-of-Stake и Delegated Proof-of-Stake, а также стремиться к более равномерному распределению вычислительной мощности в сети. Для борьбы с фишингом и социальной инженерией необходимо повышать осведомленность пользователей, использовать надежные технические средства защиты и строго следовать протоколам безопасности.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Н. Федосеев, А. Патрушева. «Связанные одной цепочкой: как блокчейн защищает данные» // URL: https://practicum.yandex.ru/blog/chto-takoe-blokchain-i-kak-eto-rabotaet (дата обращения: 24.05.2024).
2. А. Брисколини, «Все о криптоиндустрии за 30 минут. История Blockchain, стратегии заработка, криптоценность Bitbon» // URL: https://www.youtube.com/watch?v=_cHAyBueGiw (дата обращения: 24.05.2024).
3. А.Пасютина, «Криптовалюта» // URL: https://secrets.tinkoff.ru/glossarij/kriptovalyuta/ (дата обращения: 24.05.2024).
4. С.Воробей, «Лучшие криптовалютные платформы и площадки: ТОП-25 криптоплатформ в 2024». // URL: https://profinvestment.com/cryptocurrency-trading-platforms/ (дата обращения: 24.05.2024).
5. С.Воробей, «Защита криптовалютного кошелька от взлома: 19 рекомендаций и способов для безопасного использования криптовалют» // URL: https://profinvestment.com/cryptocurrency-wallet-protection/ (дата обращения: 24.05.2024).
6. Н.Райфф, «Защитите свои биткоины от кражи и взлома». // URL: https://www.investopedia.com/tech/ways-protect-your-bitcoin-investment-against-theft-and-hacks/ (дата обращения: 24.05.2024).
7. А.Макаров, «Как защитить свои криптовалютные активы». // URL: https://www.anti-malware.ru/practice/solutions/how-protect-your-cryptocurrency-assets (дата обращения: 26.05.2024).
8. Атака 51%: что нужно знать об этом? // URL: https://dzen.ru/a/Y-VSVwAcr0_3oLaz (дата обращения: 26.05.2024).
9. На крючке: как изменился фишинг в 2022 году и на что мошенники ловили своих жертв, // URL: https://habr.com/ru/companies/solarsecurity/articles/708694/ (дата обращения: 26.05.2024).
10. В. Кодолова, «Пострадавший от мошенников основатель Glow Token потребовал компенсации от Crypto.com». // URL: https://bits.media/postradavshiy-ot-moshennikov-osnovatel-glow-token-potreboval-kompensatsii-ot-crypto-com/ (дата обращения: 26.05.2024).
11. «Атака 51%», // URL: https://academy.binance.com/ru/articles/what-is-a-51-percent-attack

12. «Что такое Proof of Work (PoW)?» //URL: https://academy.binance.com/ru/articles/proof-of-work-explained (дата обращения: 26.05.2024).
13. «Что такое социальная инженерия?». // URL: https://www.kaspersky.ru/resource-center/definitions/what-is-social-engineering (дата обращения: 26.05.2024).
14. С.Погудин, «Хешрейт простыми словами». // URL: https://www.finam.ru/publications/item/kheshreyt-prostymi-slovami-20230922-1909/ (дата обращения: 26.05.2024).
15. Помогалова А.В., Донсков Е.А., Котенко И.В. Децентрализованные финансовые сервисы: общий алгоритм атаки. // URL: https://elibrary.ru/item.asp?id=49265247 (дата обращения: 26.05.2024).
16. Запечников С.В. Системы распределенного реестра, обеспечивающие конфиденциальность транзакций. // URL: https://elibrary.ru/item.asp?id=44365097 (дата обращения: 26.05.2024).
17. Сердечный А.Л., Скогорева Д.А., Длинный Е.П., Ле Т.Ч., Чьеу Д.В. Сердечный А.Л. Картографическое исследование blockchain-транзакций и смарт-контрактов киберпреступников, атакующих автоматизированные информационные системы, и оценка ущербов от реализации их атак // Информационная безопасность. 2021. Т. 24. ВЫП. 4. С. 471-500. URL: https://elibrary.ru/item.asp?id=48158181 (дата обращения: 26.05.2024).

REFERENCES

1. N. Fedoseev, A. Patrusheva. «Svyazannye odnoj cepochkoj: kak blokchejn zashchishchaet dannye» // URL: https://practicum.yandex.ru/blog/chto-takoe-blokchain-i-kak-eto-rabotaet (data obrashcheniya: 24.05.2024).
2. Briskolini, «Vse o kriptoindustrii za 30 minut. Istoriya Blockchain, strategii zarabotka, kriptocennost' Bitbon» // URL: https://www.youtube.com/watch?v=_cHAyBueGiw (data obrashcheniya: 24.05.2024).
3. A.Pasyutina, «Kriptovalyuta» // URL: https://secrets.tinkoff.ru/glossarij/kriptovalyuta/ (data obrashcheniya: 24.05.2024).
4. S.Vorobej, «Luchshie kriptovalyutnye platformy i ploshchadki: TOP-25 kriptoplatform v 2024». // URL: https://profinvestment.com/cryptocurrency-trading-platforms/ (data obrashcheniya: 24.05.2024).
5. S.Vorobej, «Zashchita kriptovalyutnogo koshel'ka ot vzloma: 19 rekomendacij i sposobov dlya bezopasnogo ispol'zovaniya kriptovalyut» // URL: https://profinvestment.com/cryptocurrency-wallet-protection/ (data obrashcheniya: 24.05.2024).
6. N.Rajff, «Zashchitite svoi bitkoiny ot krazhi i vzloma». // URL: https://www.investopedia.com/tech/ways-protect-your-bitcoin-investment-against-theft-and-hacks/ (data obrashcheniya: 24.05.2024).
7. A.Makarov, «Kak zashchitit' svoi kriptovalyutnye aktivy». // URL: https://www.anti-malware.ru/practice/solutions/how-protect-your-cryptocurrency-assets (data obrashcheniya: 26.05.2024).51% Attack: What do you need to know about it? https://changelly.com/blog/51-percent-attack/
8. Ataka 51%: chto nuzhno znat' ob etom? // URL: https://dzen.ru/a/Y-VSVwAcr0_3oLaz (data obrashcheniya: 26.05.2024).
9. Na kryuchke: kak izmenilsya fishing v 2022 godu i na chto moshenniki lovili svoih zhertv, // URL: https://habr.com/ru/companies/solarsecurity/articles/708694/ (data obrashcheniya: 26.05.2024).
10. V. Kodolova, «Postradavshij ot moshennikov osnovatel' Glow Token potreboval kompensacii ot Crypto.com». // URL: https://bits.media/postradavshiy-ot-moshennikov-osnovatel-glow-token-potreboval-kompensatsii-ot-crypto-com/ (data obrashcheniya: 26.05.2024).
11. «Ataka 51%», // URL: https://academy.binance.com/ru/articles/what-is-a-51-percent-

attack
12. «Chto takoe Proof of Work (PoW)?» //URL: https://academy.binance.com/ru/articles/proof-of-work-explained (data obrashcheniya: 26.05.2024).
13. «Chto takoe social'naya inzheneriya?». // URL: https://www.kaspersky.ru/resource-center/definitions/what-is-social-engineering (data obrashcheniya: 26.05.2024).
14. S.Pogudin, «Heshrejt prostymi slovami». // URL: https://www.finam.ru/publications/item/kheshreyt-prostymi-slovami-20230922-1909/ (data obrashcheniya: 26.05.2024).
15. Pomogalova A.V., Donskov E.A., Kotenko I.V. Decentralizovannye finansovye servisy: obshchij algoritm ataki. // URL: https://elibrary.ru/item.asp?id=49265247 (data obrashcheniya: 26.05.2024).
16. Zapechnikov S.V. Sistemy raspredelennogo reestra, obespechivayushchie konfidencial'nost' tranzakcij. // URL: https://elibrary.ru/item.asp?id=44365097 (data obrashcheniya: 26.05.2024).
17. Serdechnyj A.L., Skogoreva D.A., Dlinnyj E.P., Le T.Ch., Ch'eu D.V. Serdechnyj A.L. Kartograficheskoe issledovanie blockchain-tranzakcij i smart-kontraktov kiberprestupnikov, atakuyushchih avtomatizirovannye informacionnye sistemy, i ocenka ushcherbov ot realizacii ih atak // Informacionnaya bezopasnost'. 2021. T. 24. VYP. 4. S. 471-500. URL: https://elibrary.ru/item.asp?id=48158181 (data obrashcheniya: 26.05.2024).

Информация об авторах

Бабак Павел Вячеславович – студент 1 курса кафедры «Информационные системы и защита информации», направления подготовки «Информационная безопасность», Иркутский государственный университет путей сообщения, pasha.babak.3000@gmail.com

Кириллова Татьяна Климентьевна – заведующий кафедрой «Информационные системы и защита информации», доцент, Иркутский государственный университет путей сообщения, г. Иркутск, e-mail: kirillova_tk@irgups.ru

Пинин Илья Евгеньевич - студент 1 курса кафедры «Информационные системы и защита информации», направления подготовки «Информационная безопасность», Иркутский государственный университет путей сообщения, ilapinin@gmail.com

Information about the authors

Babak Pavel Vyacheslavovich – 1st year student of the Department "Information Systems and Information Protection", training area "Information Security", Irkutsk State University of Railway Engineering, pasha.babak.3000@gmail.com

Kirillova Tatiana Klimentevna – Head of the ISiZI Department, Associate Professor, Irkutsk State University of Railway Transport, Irkutsk, e-mail: kirillova_tk@irgups.ru

Pinin Ilya Evgenievich - 1st year student of the Department "Information Systems and Information Protection", training area "Information Security", Irkutsk State University of Railway Engineering, ilapinin@gmail.com

Для цитирования

Бабак П.В., Кириллова Т.К, Пинин И.Е. Безопасность крипотовалют// «Информационные технологии и математическое моделирование в управлении сложными системами»: электрон. науч. журн. – 2024. – №2. – С.30-38. – Режим доступа: http://ismm-irgups.ru/toma/2222024.

For citations

Babak P.V., Kirillova T.K., Pinin I.E. Security of cryptocurrencies // «Informacionnye tekhnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami»: elektron. nauch. zhurn – 2024. – No.2. – P. 30-38. – Rezhim dostupa: http://ismm-irgups.ru/toma/2222024

УДК 004.056

Бутин А. А.[1], Соколова А. И.[1]

[1]Иркутский государственный университет путей сообщения, г. Иркутск, Российская Федерация

ОСОБЕННОСТИ ИНТЕГРАЦИИ SIEM-СИСТЕМЫ С ДРУГИМИ СРЕДСТВАМИ ЗАЩИТЫ ИНФОРМАЦИИ

Аннотация. В данной статье освещены особенности интеграции SIEM-системы с другими средствами защиты информации, включающие в себя проблемы, которые необходимо учитывать специалисту при разработке системы защиты информации. Проанализирована статистика увеличения компьютерных атак за последние годы. Описано назначение и приведена архитектура SIEM-системы, интегрированной с другими средствами защиты информации. Проанализировано, какие средства наиболее часто выступают в роли источников событий информационной безопасности. Отражены принципы функционирования программного обеспечения, предназначенного для сбора, нормализации и корреляции событий. Описан процесс сбора событий посредством сканирования сетевых узлов в разных режимах. Рассмотрена необходимость настройки сетевых протоколов и специализированного программного обеспечения для сбора событий, а также используемые для этого методы. Приведен принцип срабатывания правил корреляции событий. Для актуализации существующих правил предложено использование матрицы MITRE ATT&CK, в которой описаны техники, используемые злоумышленниками для реализации атак. Отражен параметр для подсчета количества событий, поступающих в SIEM-систему от средств защиты информации. Также рассмотрен современный способ масштабирования системы защиты информации с помощью улучшения технической оснащенности системы хранения.

Ключевые слова: информационная безопасность, событие информационной безопасности, инцидент информационной безопасности, SIEM-система.

Butin A. A.[1], Sokolova A. I.[1]

[1]Irkutsk State Transport University, Irkutsk, Russian Federation

FEATURES OF INTEGRATION OF THE SIEM SYSTEM WITH OTHER INFORMATION SECURITY MEANS

Annotation. This article highlights the features of the integration of a SIEM system with other information security means, including problems that must be taken into account by a specialist when developing an information security system. The statistics of the increase in computer attacks in recent years are analyzed. The purpose and architecture of a SIEM system integrated with other information security means are described. It is analyzed which tools most often act as sources of information security events. The principles of the functioning of software designed to collect, normalize and correlate events are reflected. The process of collecting events by scanning network nodes in different modes is described. The necessity of configuring network protocols and specialized software for event collection, as well as the methods used for this purpose, is considered. The principle of operation of the correlation rules is given. To update the existing rules, it is proposed to use the MITRE ATT&CK matrix, which describes the techniques used by attackers to implement attacks. The parameter for counting the number of events received by the SIEM system from information security means is reflected. A modern way of scaling the information security system by improving the technical equipment of the storage system is also considered.

Keywords: information security, information security threat, information security event, information security incident, SIEM system.

Введение. Тема расследования инцидентов информационной безопасности в настоящее время не теряет своей актуальности, растет количество кибератак, банк угроз пополняется новыми угрозами, находят новые уязвимости.

Для сравнения рассмотрим статистику увеличения числа атак на веб-ресурсы компаний с 2022 по 2023 год, приведённую на рис. 1. [1]

Рис. 1. – Число атак на веб-ресурсы компаний

Согласно статистике, приведённой Positive Technologies, количество проектов по восстановлению нарушенных бизнес-процессов возросло на 76% за первые девять месяцев 2023 года, в сравнении с показателями за весь 2022 год. Данная статистика отражена на рис.2. [2]

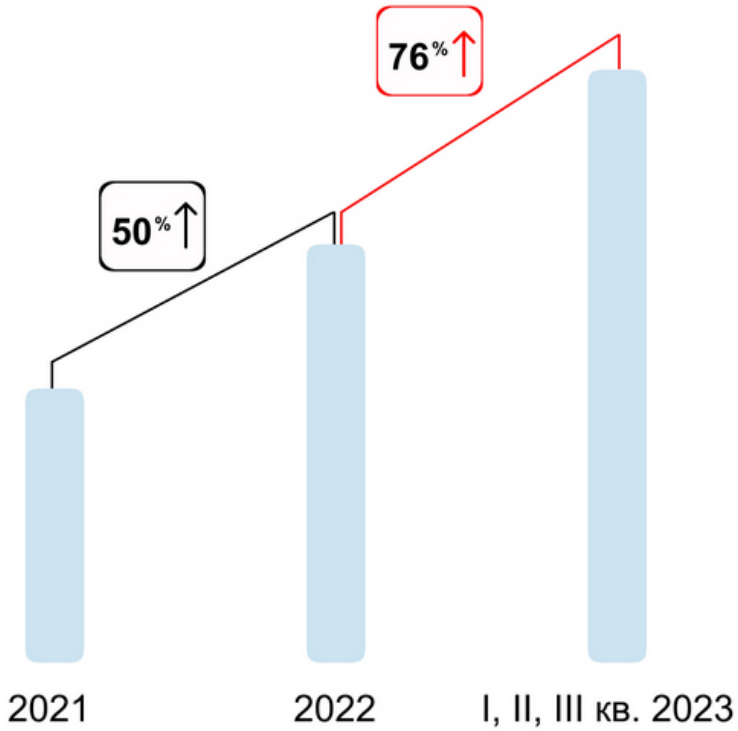

Рис. 2 – Количество проектов по расследованию инцидентов в 2021-м, 2022-м и за I—III квартал 2023 года

В связи с возросшей необходимостью отражать атаки всё больше компаний задумывается над построением системы защиты, способной обработать огромное количество информационных поток предприятия.

С целью сбора и анализа информации с различных средств защиты информации таких, как DLP, IDS, IPS, антивирусы, средства VPN, разработаны SIEM-системы. Но интеграция таких систем является сложным процессом, включающим в себя ряд особенностей, которые будут рассмотрены в данной статье.

Назначение SIEM-системы. SIEM (Security information and event management) — объединение двух терминов, обозначающих область применения ПО: SIM (Security information management) — управление информацией о безопасности, и SEM (Security event management) — управление событиями безопасности. [3]

К функциям SIM-системы относятся сбор, хранение и анализ записей журналов, а также формирование необходимой отчетности. К функциям SEM-системы относится мониторинг событий безопасности в реальном времени, а также выявление уязвимости и реагирование на инциденты безопасности. [4]

Пример архитектуры данной системы представлен на рис. 3.

Рис. 3. – Архитектура SIEM-системы

Источники событий. Интеграция SIEM-системы с средствами защиты имеет ряд особенностей, которые необходимо учитывать при разработке архитектуры системы защиты информации.

Важным шагом в разворачивании SIEM-системы является выбор источников событий.

Источники событий информационной безопасности — специализированное программное и аппаратное обеспечение для информационной безопасности, порождающее события информационной безопасности.

Источники событий сообщают о тех или иных явлениях в автоматизированной системе без оценки уровня их защищенности.

Компонент сбора событий сканирует IT-инфраструктуру предприятия, собирает сведения о сетевых узлах и события с источников. Собранные данные передаются компоненту управления. [5]

Примеры источников событий информационной безопасности: IDS/IPS (для сбора данных о сетевых атаках), средства антивирусной защиты (обнаружение вредоносных программ). Статистика наиболее часто подключаемых источников событий информационной безопасности представлена на рис. 4. [6]

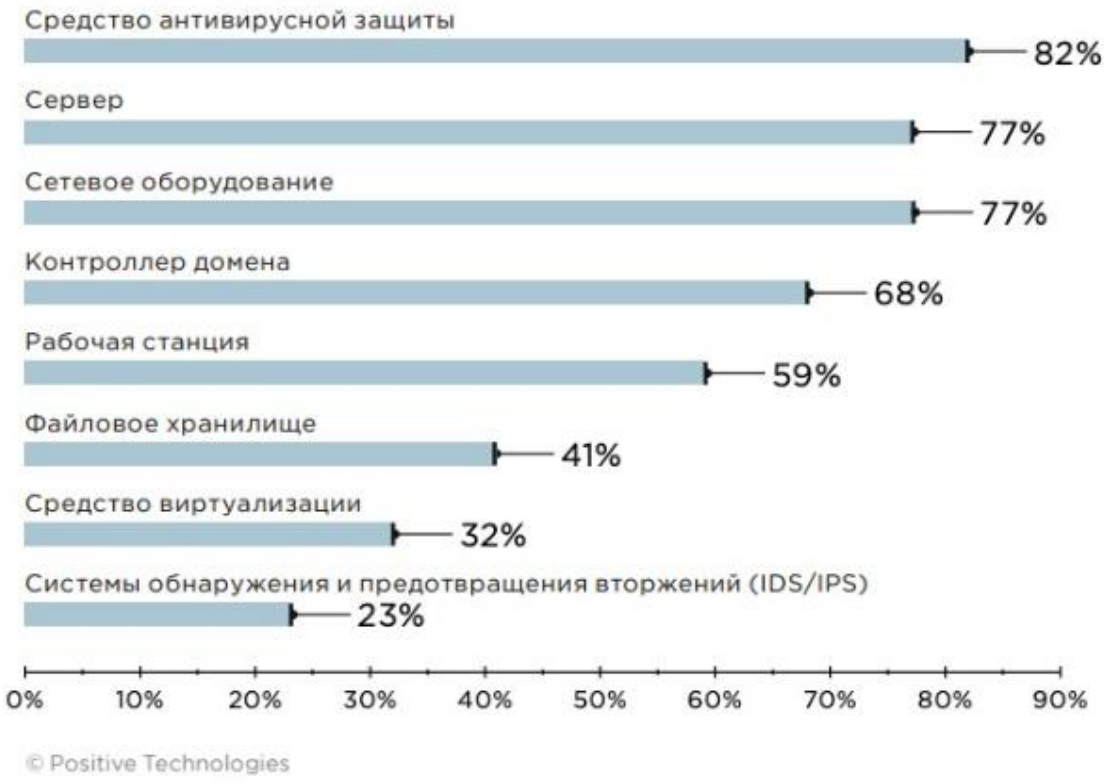

Рис. 4. – Наиболее часто подключаемые источники событий

Оператор SIEM-системы может обнаруживать новые активы источников событий с помощью встроенных модулей. Это происходит посредством сканирования сетевых узлов в режимах белого ящика (модулем audit) и черного ящика (модулем pentest).

Настройка отправки событий. Также одной из основных задач для специалиста, разворачивающего систему в инфраструктуре предприятия, является настройка отправки событий.

Зачастую это реализуется посредством настройки протокола syslog.

Syslog-сервер – это внешний сервер для сбора событий. Он хранит и анализирует полученные события, а также выполняет другие действия по управлению журналами. [7]

Стоит учитывать, что для сбора событий с некоторых источников событий требуются специализированные коннекторы.

Коннектор – специализированное ПО, предназначенное для сбора событий информационной безопасности, хранящихся в базе данных, и последующей первичной обработки полученных событий к единому внутреннему стандарту SIEM (преобразование к нормализованному виду).

Сбор событий может осуществляться в активном режиме – сборщик событий сам подключается к источнику по различным протоколам (RPC, SMB и т.д.) и собирает у него события. Или же источник сам присылает события посредством протокола Syslog или SNMP. [8]

Нормализация событий. Не менее важным при интеграции SIEM-системы со сторонними средствами защиты информации являются правила нормализации событий. События, получаемые от источников в «сыром» виде, необходимо привести к единой структуре, чтобы SIEM-система могла распознать их для последующей корреляции. При нормализации определяются, как минимум, основные сущности события: субъект, объект, источник, канал взаимодействия. [9]

Правила корреляции. Кроме того, для конкретных источников событий пишут отдельные правила корреляции. В них описываются критерии возникновения угрозы и реакция на них.

В SIEM-систему поступает огромный поток событий. Для того, чтобы связать их в одно событие информационной безопасности в правилах корреляции прописываются необходимые для этого условия (симптомы). Например, попытка неуспешного входа. Сервер корреляции отвечает за понимание инцидентов и отсеивание простых событий от общего потока. [10]

Счетчик подсчитывает количество совпадений по одному правилу. При определённом количестве событий SIEM-система может создать инцидент информационной безопасности. [11]

Принцип срабатывания правила корреляции приведен на рис. 5.

Рис. 5 – Принцип срабатывания правила корреляции

Правила корреляции должны постоянно актуализироваться экспертами. Поскольку меняются не только угрозы, но и инфраструктура предприятия. Для актуализации правил, например, можно использовать матрицу MITRE ATT&CK, отслеживая техники, которые используют злоумышленники в инцидентах информационной безопасности. [12]

Производительность при обработке большого количества событий. При внедрении SIEM-системы также важно рассчитать число обрабатываемых событий в секунду – EPS (events per second), получаемых от источников событий. [13]

Чтобы увеличить производительность хранилища событий и сократить расходы на аппаратное обеспечение, разработали гибридную схему хранения данных. В этом случае последние суточные индексы будут записываться на высокоскоростные твердотельные накопители (SSD) и со временем будут постепенно перезаписываться на более доступные накопители на жестких магнитных дисках. Это позволяет увеличить скорость обработки событий при одновременном выполнении поисковых запросов. [14]

Немаловажным является возможность масштабирования и отказоустойчивости системы, в особенности, если планируется увеличение количества средств защиты информации в инфраструктуре предприятия. [15]

Заключение. Интеграция SIEM-системы с другими средствами защиты информации является многоэтапным и сложным процессом. После проведения анализа того, какие средства защиты информации станут источниками событий для системы мониторинга, специалисту необходимо настроить процесс отправки событий, проработать правила нормализации и

корреляции, а также рассчитать количество обрабатываемых событий. Необходимо постоянно учитывать особенности архитектуры информационной системы и меняющиеся угрозы информационной безопасности. Только при непрерывной работе над улучшением системы защиты информации SIEM-система станет эффективным инструментом для контроля и устранения уязвимостей информационной системы предприятия.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Кибербезопасность в 2023-2024 гг.: тренды и прогнозы. Часть третья (ptsecurity.com) [Электронный ресурс]. – URL: https://www.ptsecurity.com/ru-ru/research/analytics/kiberbezopasnost-v-2023-2024-gg-trendy-i-prognozy-chast-tretya/ (Дата обращения: 10.04.2024).
2. Итоги расследований инцидентов ИБ в 2021–2023 годах (ptsecurity.com) [Электронный ресурс]. – URL: https://www.ptsecurity.com/ru-ru/research/analytics/outcomes-of-IS-incident-investigations-in-2021-2023-years/ (Дата обращения: 10.04.2024).
3. SIEM (ru.wikipedia.org) [Электронный ресурс]. – URL: https://ru.wikipedia.org/wiki/SIEM (Дата обращения: 10.04.2024).
4. Абденов А. Ж., Трушин В. А., Сулайман К. Анализ, описание и оценка функциональных узлов SIEM-системы: учебное пособие. – Новосибирск: НГТУ, 2018. – 122 с. (Дата обращения: 10.04.2024).
5. Алгоритм работы MaxPatrol SIEM и схема взаимодействия компонентов (help.ptsecurity.com) [Электронный ресурс]. – URL: https://help.ptsecurity.com/ru-RU/projects/siem/8.0/help/2189382283 (Дата обращения: 10.04.2024).
6. Выявление инцидентов ИБ с помощью SIEM: типичные и нестандартные задачи, 2020 (ptsecurity.com) [Электронный ресурс]. – URL: https://www.ptsecurity.com/ru-ru/research/analytics/incidents-siem-2020/ (Дата обращения: 11.04.2024).
7. Настройка параметров интеграции с SIEM (support.kaspersky.com) [Электронный ресурс]. – URL: https://support.kaspersky.com/KSWS/11/ru-RU/146650.htm (Дата обращения: 10.04.2024).
8. И снова про SIEM (habr.com) [Электронный ресурс]. – URL: https://habr.com/ru/companies/otus/articles/773430/ (Дата обращения: 11.04.2024).
9. Глубины SIEM: корреляции «из коробки». Часть 3.2. Методология нормализации событий (securitylab.ru) [Электронный ресурс]. – URL: https://www.securitylab.ru/blog/company/pt/345379.php (Дата обращения: 12.04.2024).
10. SIEM – Security Information and Event Management (www.securityvision.ru) [Электронный ресурс]. – URL: https://www.securityvision.ru/blog/siem-security-information-and-event-management/ (Дата обращения: 12.04.2024).
11. Корреляция SIEM – это просто. Сигнатурные методы (securitylab.ru) [Электронный ресурс]. – URL: https://www.securitylab.ru/analytics/431459.php?ysclid=luwe2xur4h87421031 (Дата обращения: 12.04.2024).
12. Сколько правил нужно SIEM-системе (kaspersky.ru) [Электронный ресурс]. – URL: https://www.kaspersky.ru/blog/siem-rules/35597/ (Дата обращения: 12.04.2024)
13. Внедрение SIEM – что нужно знать про него (habr.com) [Электронный ресурс]. – URL: https://habr.com/ru/sandbox/97147/ (Дата обращения: 12.04.2024)
14. MaxPatrol SIEM теперь обрабатывает до 60 000 событий в секунду (ptsecurity.com) [Электронный ресурс]. – URL: https://www.ptsecurity.com/ru-ru/about/news/maxpatrol-siem-teper-obrabatyvaet-do-60-000-sobytij-v-sekundu/ (Дата обращения: 12.04.2024).
15. Как правильно выбрать и внедрить SIEM-систему (anti-malware.ru) [Электронный ресурс]. – URL: https://www.anti-malware.ru/practice/methods/How-to-choose-and-implement-SIEM-correctly (Дата обращения: 12.04.2024).

REFERENCES

1. Cybersecurity in 2023-2024 гг.: trends and forecasts. Part three (ptsecurity.com) [Electronic resource]. – URL: https://www.ptsecurity.com/ru-ru/research/analytics/kiberbezopasnost-v-2023-2024-gg-trendy-i-prognozy-chast-tretya/ (Date of the operation: 10.04.2024).

2. Results of investigations of information security incidents in 2021-2023 (ptsecurity.com) [Электронный ресурс]. – URL: https://www.ptsecurity.com/ru-ru/research/analytics/outcomes-of-IS-incident-investigations-in-2021-2023-years/ (Date of the operation: 10.04.2024).

3. SIEM (ru.wikipedia.org) [Electronic resource]. – URL: https://ru.wikipedia.org/wiki/SIEM (Date of the operation: 10.04.2024).

4. Abdenov A. J., Trushin V. A., Sulaiman K. Analysis, description and evaluation of functional nodes of the SIEM system: a training manual. – Novosibirsk: NSTU, 2018. – 122 p. (Date of the operation: 10.04.2024).

5. MaxPatrol SIEM algorithm and component interaction scheme (help.ptsecurity.com) [Electronic resource]. – URL: https://help.ptsecurity.com/ru-RU/projects/siem/8.0/help/2189382283 (Date of the operation: 10.04.2024).

6. Identification of information security incidents using SIEM: typical and non-standard tasks, 2020 (ptsecurity.com) [Electronic resource]. – URL: https://www.ptsecurity.com/ru-ru/research/analytics/incidents-siem-2020/ (Date of the operation: 11.04.2024).

7. Configuring the parameters of integration with SIEM (support.kaspersky.com) [Electronic resource]. – URL: https://support.kaspersky.com/KSWS/11/ru-RU/146650.htm (Date of the operation: 10.04.2024).

8. And again about SIEM (habr.com) [Electronic resource]. – URL: https://habr.com/ru/companies/otus/articles/773430/ (Date of the operation: 11.04.2024).

9. SIEM depths: out-of-the-box correlations. Part 3.2. Event normalization methodology (securitylab.ru) [Electronic resource]. – URL: https://www.securitylab.ru/blog/company/pt/345379.php (Date of the operation: 12.04.2024).

10. SIEM – Security Information and Event Management (www.securityvision.ru) [Electronic resource]. – URL: https://www.securityvision.ru/blog/siem-security-information-and-event-management/ (Date of the operation: 12.04.2024).

11. SIEM correlation is simple. Signature methods (securitylab.ru) [Electronic resource]. – URL: https://www.securitylab.ru/analytics/431459.php?ysclid=luwe2xur4h87421031 (Date of the operation: 12.04.2024).

12. How many rules does the SIEM system need (kaspersky.ru) [Electronic resource]. – URL: https://www.kaspersky.ru/blog/siem-rules/35597/ (Date of the operation: 12.04.2024)

13. SIEM implementation – what you need to know about it (habr.com) [Electronic resource]. – URL: https://habr.com/ru/sandbox/97147/ (Date of the operation: 12.04.2024)

14. MaxPatrol SIEM now processes up to 60,000 events per second (ptsecurity.com) [Electronic resource]. – URL: https://www.ptsecurity.com/ru-ru/about/news/maxpatrol-siem-teper-obrabatyvaet-do-60-000-sobytij-v-sekundu/ (Date of the operation: 12.04.2024).

15. How to choose and implement a SIEM system correctly (anti-malware.ru) [Electronic resource]. – URL: https://www.anti-malware.ru/practice/methods/How-to-choose-and-implement-SIEM-correctly (Date of the operation: 12.04.2024).

Информация об авторах

Александр Алексеевич Бутин – к. ф.-м. н., доцент, доцент кафедры «кафедры «Информационные системы и защита информации», Иркутский государственный университет путей сообщения, г. Иркутск

Алена Игоревна Соколова – студент, Иркутский государственный университет путей сообщения, г. Иркутск

Authors

Aleksander Alekseevich Butin, Candidate of Physico-Mathematical Sciences, Doctor, Associate Professor, the Subdepartment of Information systems and information security, Irkutsk State Transport University, Irkutsk

Alena Igorevna Sokolova, student, Irkutsk State Transport University, Irkutsk

Для цитирования

Соколова А.И., Бутин А.А. Особенности интеграции SIEM-системы с другими средствами защиты информации // «Информационные технологии и математическое моделирование в управлении сложными системами»: электрон. науч. журн. – 2024. – №2. – С. 38-46. – Режим доступа: https://ismm.irgups.ru/toma/222-2024, свободный. – Загл. с экрана. – Яз. рус., англ.

For citations

Sokolova A.I., Butin A.A. Features of integration of the SIEM system with other information security means // Informacionnye tehnologii i matematicheskoe modelirovanie v upravlenii slozhnymi sistemami: ehlektronnyj nauchnyj zhurnal [Information technology and mathematical modeling in the management of complex systems: electronic scientific journal], 2024. No. 2. P. 38-46.