

На правах рукописи



Спицын Сергей Владимирович

**МОДЕЛИ И АЛГОРИТМЫ ОБРАБОТКИ ТРАЕКТОРНОЙ ИНФОРМАЦИИ
НА ОСНОВЕ МЕТОДОВ КЛАССИФИКАЦИИ И
РЕГРЕССИОННОГО АНАЛИЗА**

Специальность 05.13.01 – Системный анализ, управление и обработка информации
(технические системы)

Автореферат

диссертации на соискание ученой степени

кандидата технических наук

Рязань – 2021 г.

Работа выполнена в Федеральном государственном бюджетном образовательном учреждении высшего образования «Рязанский государственный радиотехнический университет имени В.Ф. Уткина» (ФГБОУ ВО «РГРТУ») на кафедре «Систем автоматизированного проектирования вычислительных средств».

Научный руководитель: **Корячко Вячеслав Петрович**, заслуженный деятель науки и техники РФ, доктор технических наук, профессор, заведующий кафедрой систем автоматизированного проектирования вычислительных средств ФГБОУ ВО «РГРТУ»

Официальные оппоненты: **Прохоров Сергей Антонович**, заслуженный работник высшей школы РФ, доктор технических наук, профессор, заведующий кафедрой информационных систем и технологий ФГАОУ ВО «Самарский национальный исследовательский университет имени академика С.П. Королева» (Самарский университет), г. Самара

Козлов Евгений Владимирович, кандидат технических наук, старший преподаватель кафедры информационно-телеметрического и метрологического обеспечения ФГКВОУ ВО «Военная академия РВСН имени Петра Великого», г. Балашиха

Ведущая организация: Акционерное общество «Научно-инженерный центр Санкт-Петербургского электротехнического университета» (АО «НИЦ СПб ЭТУ»), г. Санкт-Петербург

Защита состоится **29 июня 2021 г.** в **14** часов на заседании диссертационного совета Д 212.211.01 в ФГБОУ ВО «РГРТУ» по адресу:

390005, г. Рязань, ул. Гагарина, д. 59/1.

С диссертацией можно ознакомиться в научной библиотеке ФГБОУ ВО «РГРТУ» и на сайте «www.rsreu.ru».

Автореферат разослан «_____» _____ 2021 года.

Ученый секретарь
диссертационного совета
Д 212.211.01
доктор технических наук, доцент



**Пруцков Александр
Викторович**

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы. Одной из приоритетных задач при проведении ракетных пусков (полигонных испытаний) является обработка, анализ и представление траекторных измерений, поступающих с борта изделия с общим потоком цифровой телеметрической информации (ТМИ) и описывающих баллистическую траекторию полёта. В связи с тем, что сложные изделия ракетно-космической техники (РКТ), как правило, состоят из нескольких ступеней, к таким задачам также относится и установление фактических географических координат точек падения отделяющихся частей. Исходя из этого, в том числе формируется и признак попадания или промаха – посредством сопоставления с зоной расчетных географических координат и учетом предельно допустимых отклонений. От эффективности обработки и безошибочности траекторных расчётов и основанной на них оперативной оценки состояния изделия во многом зависит качественное выполнение поставленных задач натурных работ.

Основные сложности, связанные с выделением из общего потока ТМИ цифрового кадра, содержащего траекторную информацию, его дальнейшей обработкой и представлением, заключаются:

- в жёстких временных ограничениях, характерных для лётного эксперимента в режиме реального масштаба времени (РМВ);
- во внешних факторах, влияющих на достоверность обрабатываемой информации, поступающей от подвижных и удалённых объектов;
- в ограниченных вычислительных мощностях программно-аппаратного обеспечения полигонного измерительного комплекса (ПИК), задействованного в лётном эксперименте (ракетном пуске);
- в разнородности телеметрической информации.

Для того чтобы минимизировать негативное влияние вышеперечисленных факторов и их комбинаций, требуется применение информационно-аналитических систем, основанных на современных методах, позволяющих сократить значительный объём непрерывных математических вычислений за счёт введения в методику (алгоритм) расчёта этапов прогнозирования численных величин с приемлемым уровнем точности, а также выделить в задачи бинарной классификации детектирование аварийных и нештатных ситуаций и отклонений определенных телеметрируемых параметров от их допустимых значений.

К наиболее эффективным способам решения данных задач следует отнести современные методы искусственного интеллекта: регрессионный анализ, машинное обучение, «data mining», нейронные сети, логистическая регрессия, а также другие методы интеллектуальной и предсказательной аналитики. Эффективное и правильное применение построенных на этих методах информационно-аналитических систем и их моделей не только увеличивает скорость обработки данных, снижая при этом влияние «человеческого фактора», но также в значительной степени минимизирует требования к аппаратному обеспечению и наличию квалифицированного персонала инженеров-испытателей (вплоть до их полного отсутствия), что в свою очередь выгодно отражается на экономической стороне вопроса.

Степень разработанности темы исследования. В настоящее время с ростом количества параметров ТМИ, подлежащих контролю во время натурных работ наблюдается и существенное сокращение временных отрезков, отведённых на оперативную обработку информации, принятие решений и выдачу заключений о результа-

тах пуска, что говорит о необходимости и важности внедрения в ракетно-космическую отрасль систем интеллектуального анализа данных не только на этапе лётного эксперимента (режима РМВ), но и на этапах подготовки и послеполётного анализа (отложенного сеанса). В области искусственного интеллекта, ранее относящегося к области кибернетики, фундаментальные исследования связывают с С.Ю. Масловым, Д.А. Поспеловым, Г.С. Поспеловым, Г.С. Осиповым, А. Ньюэллом (A.Newell) и Г. Саймоном (H.Simon), чьи труды получили дальнейшее развитие в работах современных отечественных исследователей В.П. Корячко, А.Н. Пылькина Л.А. Демидовой. Ключевые монографии Г.С. Осипова содержат изложение основных методов представления знаний, моделирования рассуждений, моделирования поведения, методов обучения и приобретения знаний интеллектуальными системами, а также методов автоматизации приобретения знаний. Основными работами, направленными на исследование проблем, связанных с информационно-аналитическим обеспечением полигонных испытаний сложных изделий ракетно-космической техники, а также с обработкой цифровой телеметрической информации являются работы В.И. Белицкого, В.А. Меньшикова, А.В. Назарова, А.И. Лоскутова, Г.И. Козырева.

Фундаментальные исследования основных методов классификации и регрессии, таких как «дерево решений», CART, «лес решений» связывают с именами Л. Бреймана (L. Breiman) и А. Катлер (A. Cutler). В методы обработки траекторных измерений значительный вклад внесли Р.Ф. Аппазов, О.Г. Сытин, Б.Ц. Бахшиян.

Объектом исследования являются процессы обработки и анализа цифровой телеметрической информации, содержащей кадры бортовой навигационной аппаратуры потребителя в части траекторной информации.

Предмет исследования – унифицированные модели и алгоритмы обработки и анализа траекторной информации, поступающей с борта изделия РКТ при проведении полигонных испытаний, включающие в себя этапы расчета географических координат точек падения отделяющихся частей изделия для формирования признака попадания и промаха, а также этапы детектирования возникновений нештатных ситуаций и прогнозирования численных значений координат посредством методов машинного обучения и регрессионного анализа.

Целью диссертационной работы является повышение производительности, объективности оценок и уровня автоматизации процессов обработки, анализа и представления траекторной информации объекта (сложного изделия ракетно-космической техники) в системах информационно-аналитического и телеметрического обеспечения полигонных измерительных комплексов. Для достижения поставленной цели необходимо решить следующие **основные задачи**:

- провести исследование и сравнительный анализ существующих современных и перспективных методов обработки траекторных измерений, моделей и алгоритмов расчёта баллистической траектории сложных изделий ракетно-космической техники;

- выработать технические предложения, обеспечивающие высокий уровень автоматизации и объективности оценок натурных работ, направленные на повышение скорости обработки, анализа и представления траекторной ТМИ и позволяющие минимизировать требования к производительности аппаратно-программных средств полигонного измерительного комплекса;

- провести анализ процессов обработки зарегистрированной в ходе лётного эксперимента ТМИ, и включающих в себя стадии выделения цифровой кадровой структуры, содержащей траекторную информацию объекта, её обработки и формиро-

вания единого унифицированного формата представления выделенных параметров для последующего анализа и передачи в смежные системы;

- разработать модели и алгоритм расчёта баллистической траектории отделяющихся частей сложного изделия РКТ для формирования признаков попадания и промаха, обеспечивающих автоматизацию процессов обработки и анализа траекторной ТМИ;

- в отдельные расчётные этапы разработанного алгоритма ввести этапы регрессионного анализа данных с прогнозированием соответствующими моделями численных значений траекторных параметров, с целью снижения объема математических вычислений и, как следствие, минимизации требований, предъявляемых к производительности аппаратно-программных средств полигонного измерительного комплекса;

- разработать эффективную методику оптимизации гиперпараметров модели классификатора, а также модель и алгоритм интеллектуального детектирования возникновения нештатных и аварийных ситуаций на борту изделия РКТ на основе неявных признаков траекторных отклонений, основанные на бинарной классификации состояний объекта (штатного и нештатного) с целью увеличения объективности и оперативности аналитической оценки хода и результатов пуска;

- провести анализ целесообразности и эффективности применения и внедрения предложенных моделей и алгоритмов.

Научная новизна диссертационной работы заключается в том, что в ней впервые предложены:

- модели и алгоритм расчёта баллистической траектории отделяющихся частей сложного многопараметрического объекта РКТ для формирования признаков попадания и промаха, отличающиеся от известных тем, что при их реализации предложено использовать элементы теории машинного обучения и прогнозирования численных значений траекторных параметров посредством методов регрессионного анализа, что позволяет обеспечить высокий уровень автоматизации процессов обработки и анализа траекторной ТМИ, минимизировать требования к производительности аппаратно-программных средств ПИК и снизить влияние «человеческого фактора»;

- эффективная методика оптимизации гиперпараметров модели классификатора бортовых состояний объекта, имеющая большую практическую значимость при описании штатного функционирования сложного многопараметрического объекта РКТ и исследовании причин возникновения нештатных состояний;

- модель и алгоритм детектирования возникновения нештатных и аварийных ситуаций бортовых систем объекта на основе неявных признаков траекторных отклонений, отличающиеся от известных тем, что при их реализации предложено использовать методы интеллектуального анализа данных с подбором оптимального бинарного классификатора состояний «норма/не норма» для перехода бортовых систем из штатного состояния функционирования в аварийное, что позволяет значительно повысить уровень объективности оценок как текущего состояния изделия, так и результатов пуска в целом, а также повысить оперативность выработки решений, направленных на предотвращение возможных нежелательных последствий.

Практическая значимость работы состоит в том, что предложенные модели и алгоритмы автоматизированной обработки и анализа цифровой телеметрической информации, содержащей траекторные данные сложного изделия РКТ, обеспечивают:

– снижение вычислительной нагрузки на аппаратно-программные средства полигонного измерительного комплекса за счёт минимизации объёма производимых математических вычислений, как в режиме реального масштаба времени, так и в режиме после сеанса;

– повышение эффективности и функциональных возможностей штатного персонала в процессе проведения лётного эксперимента и натурных работ со сложным изделием РКТ за счет повышения уровня автоматизации процессов обработки, анализа и представления ТМИ;

– повышение объективности анализа и оценки ТМИ, зарегистрированной во время лётных испытаний;

– интеллектуальное детектирование возникновения бортовых нештатных (аварийных) ситуаций по косвенному траекторному признаку, что даёт возможность оперативной выработки технических решений по предотвращению возможных негативных последствий.

Методы исследования. Для решения поставленных задач использовались методы машинного обучения, классификации, математической статистики, системного анализа, методы регрессионного анализа, а также библиотеки для статистической обработки данных высокоуровневых языков программирования R, C++ и Python.

Область исследования. Содержание диссертации соответствует следующим пунктам Паспорта специальности 05.13.01 «Системный анализ, управление и обработка информации (технические системы)»: п.5 «Разработка специального математического и алгоритмического обеспечения систем анализа, оптимизации, управления, принятия решений и обработки информации», п.11 «Методы и алгоритмы прогнозирования и оценки эффективности, качества и надежности сложных систем», п.12 «Визуализация, трансформация и анализ информации на основе компьютерных методов обработки информации», п.13 «Методы получения, анализа и обработки экспертной информации».

Апробация работы. Основные положения диссертационной работы докладывались и обсуждались на международной научно-технической конференции «MIST: Aerospace-III 2020: Передовые технологии в аэрокосмической отрасли, машиностроении и автоматизации» (Красноярск, 2020), индексируемой Scopus и Web of Science, а также на всероссийских и международных конференциях: «Современные технологии в науке и образовании. Вычислительная техника и автоматизированные системы» (Рязань, 2017, 2018, 2019, 2020), «Актуальные проблемы современной науки и производства» (Рязань, 2020), «V Всероссийской военно-научной конференции «Проблемы применения космических сил и средств в системе операций вооруженных сил Российской Федерации» (Санкт-Петербург, 2020). Положения диссертации апробированы при защите НИР, эскизных и технических проектов ОКР в АО «РКЦ «Прогресс» - ОКБ «Спектр» г. Рязань.

Внедрение. Результаты работы внедрены в филиале АО «РКЦ «Прогресс» - ОКБ «Спектр» г. Рязань, «4-ом Государственном Центральном Межвидовом Полигоне Министерства Обороны Российской Федерации», а также «1-ом Государственном испытательном космодроме Министерства обороны Российской Федерации» в аппаратно-программных комплексах сбора, обработки, анализа и представления телеметрической информации, что подтверждается соответствующими актами внедрения.

Достоверность результатов исследования подтверждается математическими и экспериментальными обоснованиями и доказательствами, а также успешным практическим внедрением и штатной эксплуатацией спроектированных на основе предложенных моделей и алгоритмов аппаратно-программных комплексов обработки, представления и анализа бортовой ТМИ.

Публикации. По теме диссертации опубликовано 11 печатных работ, из них 5 статьей опубликованы в изданиях, рекомендованных ВАК РФ для публикации результатов кандидатских диссертаций и 1 статья, индексируемая Scopus и Web of Science. Получено свидетельство о регистрации программы для ЭВМ.

Положения, выносимые на защиту:

1. Модели и алгоритм расчёта баллистической траектории отделяющихся частей сложного изделия РКТ с использованием линейной регрессии позволяют с допустимым уровнем точности прогнозировать районы падения отделившихся частей конструкции при повышении оперативности обработки траекторной ТМИ в среднем на 21,4% по сравнению с базовыми строгими моделями расчёта траектории на пассивных участках полёта.

2. Методика оптимизации гиперпараметров модели классификатора бортовых состояний сложного многопараметрического объекта РКТ решает задачу построения оптимальной модели бинарного классификатора состояний эффективнее (по показателям AUC) существующих базовых методик на 1,33% в условиях малых объемов обучающей выборки данных и на 6,25% в условиях равномерного распределения исходного набора траекторных измерений, частота ошибок первого и второго рода составляет 3,34% и 5,24% соответственно.

3. Модель и алгоритм интеллектуального детектирования возникновений нештатных и аварийных состояний на борту объекта РКТ на основе неявных признаков траекторных отклонений решают задачу повышения точности классификаций текущего состояния бортовых систем по косвенному признаку в зоне неопределённостей состояний эффективнее аналогичных на 7,74% в условиях малых объемов обучающей выборки и на 6,97% при равномерном распределении траекторной информации на обучающие и тестовые наборы.

Структура и объем работы. Диссертация состоит из введения, трёх глав с выводами, заключения, списка литературы и приложения. Основной текст содержит 135 страниц, 48 рисунков и 21 таблицу. Список литературы содержит 13 страниц и включает 131 наименование.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

В **первой главе** приводится краткий обзор и анализ современных методов обработки и представления ТМИ, содержащей траекторные данные объекта РКТ. Представлен общий порядок действий и технологические этапы при работе с траекторными данными при проведении лётных испытаний (ракетных пусков). Рассмотрен алгоритм извлечения цифровых параметров из форматов штатных систем (ЕГНАКУ, CCSDS) с последующим формированием выходных кадровых структур траекторных измерений. Проведён анализ первичных и вторичных величин, необходимых для решения траекторной задачи, соответствующих методов математического расчета, а также анализ существующих строгих моделей расчёта баллистической траектории на активном и пассивном (участок свободного полёта) участках. Представлена ER-диаграмма базы данных, характеризующая единую информационную модель предметной области анализа траекторной информации сложного многопараметрического объекта РКТ в информационно-аналитической среде полигонного измерительного комплекса (ПИК). Установлено, что для современных этапов создания и эксплуатации ПИК характерно непрерывное усложнение объектов испытаний, что в свою очередь приводит к усложнению средств и методов их испытаний, а также повышению требований к качеству обработки возрастающих объемов измерительной информа-

ции, получаемой в ходе натурных работ. Анализ процессов, связанных с обработкой и анализом траекторной телеметрии выявил следующие недостатки в организации соответствующих существующих систем:

- высокие требования к вычислительным мощностям аппаратно-программных средств полигонной инфраструктуры вследствие постоянно увеличивающейся информативности бортовой ТМИ и значительного объёма производимых сложных математических расчётов;

- недостаточное и малоэффективное внедрение механизмов статистического накопления базы знаний о штатных и аварийных состояниях бортовых систем, что приводит к снижению уровня аналитической оценки последующих натурных работ;

- необходимость непрерывного и досконального контроля множества бортовых телеметрических параметров (явных признаков), сопровождающаяся потребностью привлечения высококвалифицированных специалистов, обладающих знаниями и опытом идентификации нештатных и аварийных ситуаций, что в свою очередь является причиной снижения оперативности анализа;

- низкая степень автоматизированного интеллектуального анализа процессов аналитической оценки ТМИ и, как следствие, необъективность оценок ситуаций с высокой степенью зависимости от «человеческого фактора»;

- необходимость упрощения строгой модели расчёта пассивного участка траектории, в рамках которого предлагается рассмотреть возможный вариант эффективно-го дополнения строгой модели этапами прогнозирования траекторных параметров посредством регрессионного анализа выборки исходных траекторных данных, регистрируемых в ходе лётного эксперимента.

Выводы: По результатам выполненного анализа, задача диссертационного исследования сформулирована по двум направлениям, связанным с созданием информационно-аналитических и интеллектуальных экспертных систем:

1. *Первое направление исследований* – повышение эффективности полигонных программно-технических средств и систем анализа ТМИ в части обработки траекторных измерений на *пассивном участке траектории* за счет разработки моделей и алгоритма расчёта баллистической траектории отделяющихся частей объекта с использованием линейной регрессии. Данное направление является задачей регрессионного

анализа данных $\mathbf{X} \in \mathbb{R}^{l \times n}$, $\mathbf{y} \in \mathbb{R}$: $\hat{y} = f(\mathbf{x}) = \sum_{i=0}^n \mathbf{w}_i \mathbf{x}_i = \mathbf{x}\mathbf{w}$ (для l наблюдений n траекторных признаков), результатом решения которой являются прогнозы траекторных величин вещественного типа $\mathbf{y} = \mathbf{X}\mathbf{w}$: $\mathbf{w} = \mathbf{X}^{-1}\mathbf{y}$.

2. *Второе направление исследований* – повышение эффективности информационно-аналитического обеспечения полигонных испытаний в части обработки траекторной ТМИ на *активном участке траектории* за счет разработки эффективной методики оптимизации гиперпараметров классификатора, а также модели и алгоритма интеллектуального детектирования возникновений нештатных и аварийных ситуаций на борту объекта РКТ на основе неявных признаков траекторных отклонений. Данное направление является задачей классификации $\mathbf{X} \in \mathbb{R}^{l \times n}$, $\mathbf{y} \in \{0,1\}^l$, $\hat{y} = f(\mathbf{x}) = P(Y = 1 | \mathbf{x})$, результатом решения которой является максимизация функции правдоподобия целевой переменной $\prod_{i=1}^l P(Y = y_i | x_i) \longrightarrow \max$ с нахождением вероятности отнесения метки целе-

вой переменной к одному из классов, в бинарном случае, позволяющей по принципу «норма/не норма» классифицировать штатные и нештатные состояния объекта.

Таким образом, *начальной системой координат*, как наиболее подходящей для реальных натуральных работ, принята *нецентральная стартовая система ССК* (x, y, z) , началом которой является жёстко связанная с Землёй точка старта, ось OY направлена от точки старта изделия на поверхности Земли вверх вдоль линии отвеса, OX – направлена по азимуту прицеливания ψ в горизонтальной плоскости с дополняющей систему до правой осью OZ . *Итоговой системой координат* (для определения географических координат районов падения) принята *геодезическая эллипсоидальная система B, L, H* , в которой отсчётной поверхностью является частная референц-версия общеземного эллипсоида вращения (эллипсоид Красовского), представляющая собой наиболее подходящий датум, адаптированный для описания европейской части России в геоцентрической системе координат.

Во второй главе в рамках первого направления исследований разработаны модели и алгоритм расчёта баллистической траектории отделяющихся частей сложного изделия РКТ с использованием линейной регрессии.

В качестве входных данных для расчёта принимаются: заранее известные массогабаритные данные объекта и данные о его текущей траектории (x, y, z) , получаемые из цифровой ТМИ в составе общего бортового телеметрического потока, предельно допустимые отклонения точек падения отделившихся частей и азимут прицеливания ψ . Выходными данными произведённых расчётов является файл событий, содержащий в себе координаты точек разделения, текущие координаты головной части (B, L, H) и координаты точек падения отделившихся частей $(B1, L1), (B2, L2)$, при $H1=H2=0$ – превышение над земным эллипсоидом), на основании чего, при наличии задачи, возможно формирование признаков попадания и промаха. Предлагается следующий порядок действий:

Шаг 1. Расчёт проекций вектора угловой скорости вращения Земли на оси ССК $w_{3x} = w_3 \cos(\varphi) \cos(\psi)$, $w_{3y} = w_3 \sin(\varphi)$, $w_{3z} = -w_3 \cos(\varphi) \sin(\psi)$, где φ – геоцентрическая широта точки старта; w_3 – угловая скорость вращения Земли.

Шаг 2. Расчёт постоянных коэффициентов $a_{11} = w_3^2 - w_{3x}^2$, $a_{22} = w_3^2 - w_{3y}^2$, $a_{33} = w_3^2 - w_{3z}^2$, $a_{23} = a_{32} = -w_{3y} w_{3z}$, $a_{12} = a_{21} = -w_{3x} w_{3y}$, $b_{23} = -b_{32} = 2w_{3x}$, $b_{31} = -b_{13} = 2w_{3y}$, $b_{12} = -b_{21} = 2w_{3z}$, и расстояния от объекта с текущими координатами $r = \sqrt{(x - x_c)^2 + (y - y_c)^2 + (z - z_c)^2}$, взятыми из телеметрического кадра, x_c, y_c, z_c – координаты центра Земли относительно стартовой СК, x, y, z – текущие координаты части конструкции.

Шаг 3. Расчёт скорости объекта в стартовой системе координат v в текущий момент времени $v = \sqrt{v_x^2 + v_y^2 + v_z^2}$, где v_x, v_y, v_z – скорости объекта, взятые из выделенного бортового телеметрического кадра, содержащего траекторную информацию.

Шаг 4. Расчёт радиальной $g_r = -\frac{\mu}{r^2} - \frac{3}{2} \mu J_2 \frac{R_e^2}{r^4} (1 - 3 \sin^2 \varphi)$ и меридиональной $g_w = -3 \mu J_2 \frac{R_e^2}{r^4} \sin \varphi \cos \varphi$ составляющих ускорения, где μ – произведение постоянной тяготения на массу Земли; J_2 – вторая зональная гармоника; R_e – средний экваториальный радиус Земли.

Шаг 5. Нахождение численного решения системы дифференциальных уравнений методом Рунге-Кутты 4 порядка:

$$\frac{dv_x}{dt} = \frac{-c_x p v S_{v_x}}{m} - \frac{g_r (x - x_c)}{r} - \frac{g_w w_{3_x}}{w_3} + a_{11}(x - x_c) + a_{12}(y - y_c) + a_{13}(z - z_c) + b_{12}v_y + b_{13}v_z, \quad (1)$$

$$\frac{dv_y}{dt} = \frac{-c_x p v S_{v_x}}{m} - \frac{g_r (y - y_c)}{r} - \frac{g_w w_{3_y}}{w_3} + a_{21}(x - x_c) + a_{22}(y - y_c) + a_{23}(z - z_c) + b_{21}v_x + b_{23}v_z, \quad (2)$$

$$\frac{dv_z}{dt} = \frac{-c_x p v S_{v_x}}{m} - \frac{g_r (z - z_c)}{r} - \frac{g_w w_{3_z}}{w_3} + a_{31}(x - x_c) + a_{32}(y - y_c) + a_{33}(z - z_c) + b_{31}v_x + b_{32}v_y, \quad (3)$$

$$\frac{dx}{dt} = v_x, \quad \frac{dy}{dt} = v_y, \quad \frac{dz}{dt} = v_z, \quad (4)$$

c_x – безразмерный аэродинамический коэффициент, применяемый в расчёте пассивного участка траектории; S – площадь миделя для расчета пассивного участка траектории; p – плотность воздуха в данной точке траектории; m – масса отделившейся части. Параметры c_x , S , m , выбираются исходя из массогабаритных данных объекта.

Шаг 6. Расчет высоты над поверхностью Земли $h = r - r_0$, где r_0 – радиус Земли в текущей точке полета.

Шаг 7. После достижения точки останова расчёта ($h = 0$, что соответствует точке падения отделившейся части на поверхность Земли) полученные x , y , z подлежат конвертации в B , L формата десятичных градусов $dd.ddddd$ ($x, y, z \rightarrow B, L, H$). Далее, при наличии задачи, возможно формирование признаков попадания/промаха. Формирование признаков производится посредством сопоставления фактических координат с заранее известной априорной (расчётной) зоной с учетом заданных предельно допустимых отклонений.

В таблице 1 представлен фрагмент результатов расчёта для объекта, состоящего из головной части и двух отделяющихся ступеней на отрезке бортового относительно времени от $t_0 = 51998,1$ до $t_4 = 51998,5$ сек («time»). Параметры «В», «L», «H» характеризуют широту, долготу и высоту головной части, «t1», «B1», «L1», «t2», «B2», «L2» – соответствующие параметры первой и второй ступеней.

Таблица 1 – Фрагмент результатов расчётов

time	B	L	H	t1	B1	L1	t2	B2	L2
51998,1	38,5248	36,8541	97451,6	744998	38,2421	43,3383	677998	38,2400	43,3617
51998,2	38,5248	36,8554	97655,8	746998	38,2413	43,3613	678998	38,2392	43,3850
51998,3	38,5248	36,8568	97860,1	747998	38,2409	43,3844	679998	38,2388	43,4084
51998,4	38,5248	36,8582	98064,9	747998	38,2402	43,4088	681998	38,2382	43,4306
51998,5	38,5248	36,8596	98270,0	747998	38,2390	43,4327	682998	38,2371	43,4538

Матрица корреляции полного набора результатов расчётов $n = 5348$ наблюдений показывает выраженную взаимосвязь линейного характера между параметрами (рисунок 1). На диагонали приведены наименования параметров и графики плотности их распределения. Сверху от диагонали, на пересечении строк и столбцов, показаны коэффициенты корреляции Пирсона.

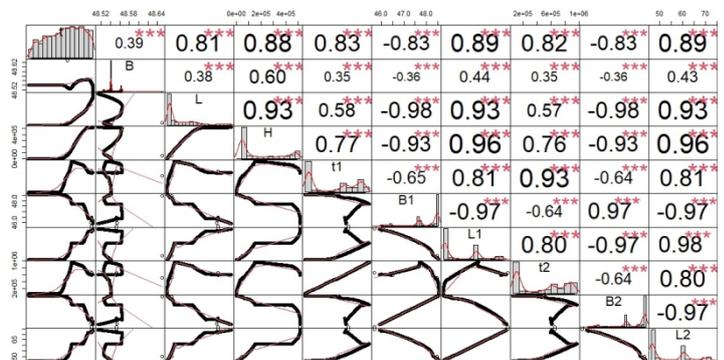


Рисунок 1 – Матрица корреляции

Для взаимозависимости широт это: $r_{B1B2} = \frac{\text{cov}_{B1B2}}{\sigma_{B1}\sigma_{B2}} = \frac{\sum (B1 - \overline{B1})(B2 - \overline{B2})}{\sqrt{\sum (B1 - \overline{B1})^2 \sum (B2 - \overline{B2})^2}}$, соответст-

вующих значений, где $\overline{B1} = \frac{1}{5348} \sum_{i=1}^n B1_i$ и $\overline{B2} = \frac{1}{5348} \sum_{i=1}^n B2_i$ – средние значения двух выборок широт; cov_{B1B2} – корреляционный момент (ковариация) $B1$ и $B2$, σ – среднеквадратическое отклонение. Для взаимозависимости долгот $L1, L2$ справедливы аналогичные выражения. Внизу, на декартовой плоскости – точечные диаграммы рассеяния (*scatter plot*) с выделенными линиями регрессии.

Примем за матрицы координат целевой переменной матрицы значений $B2, L2$ (широты и долготы точек падения второй ступени):

$$\begin{aligned} \mathbf{B2} = \overline{\mathbf{b}_{B2_{\text{pred}}}} \mathbf{B2}_{\text{pred}} + \overline{\boldsymbol{\varepsilon}_{B2}}, \quad \mathbf{L2} = \overline{\mathbf{b}_{L2_{\text{pred}}}} \mathbf{L2}_{\text{pred}} + \overline{\boldsymbol{\varepsilon}_{L2}}, \\ \mathbf{B2} = \begin{pmatrix} B2_1 \\ B2_2 \\ \dots \\ B2_{5348} \end{pmatrix} = \begin{pmatrix} 38,5411 \\ 38,5411 \\ \dots \\ 38,2371 \end{pmatrix}, \quad \mathbf{L2} = \begin{pmatrix} L2_1 \\ L2_2 \\ \dots \\ L2_{5348} \end{pmatrix} = \begin{pmatrix} 36,3011 \\ 36,3011 \\ \dots \\ 43,4538 \end{pmatrix}, \end{aligned} \quad (5)$$

$$\begin{aligned} B2 = b_0 + b_H H + b_{L1} L1 + b_{\text{time}} \text{time} + \varepsilon_{B2}, \\ L2 = b_0 + b_H H + b_{B1} B1 + b_{\text{time}} \text{time} + \varepsilon_{L2}, \end{aligned}$$

Первичными рассматриваемыми моделями примем многомерные регрессионные модели по типу (5). Подбор коэффициентов регрессии, минимизирующих сумму квадратов регрессионных остатков на гиперплоскость произведем методом наименьших квадратов (МНК): $\sum \varepsilon_i^2 = \sum ((B2_i, L2_i) - f_i(B2_{\text{pred}}, L2_{\text{pred}}))^2 \rightarrow \min$.

$$\begin{aligned} \mathbf{B2}_{\text{pred}} = \begin{pmatrix} 1 & H_1 & L1_1 & \text{time}_1 \\ 1 & H_2 & L1_2 & \text{time}_2 \\ \dots & \dots & \dots & \dots \\ 1 & H_{5348} & L1_{5348} & \text{time}_{5348} \end{pmatrix}, \quad \mathbf{L2}_{\text{pred}} = \begin{pmatrix} 1 & H_1 & B1_1 & \text{time}_1 \\ 1 & H_2 & B1_2 & \text{time}_2 \\ \dots & \dots & \dots & \dots \\ 1 & H_{5348} & B1_{5348} & \text{time}_{5348} \end{pmatrix}, \\ \overline{\mathbf{b}_{B2_{\text{pred}}}} = \begin{pmatrix} b_0 \\ b_H \\ b_{L1} \\ b_{\text{time}} \end{pmatrix}, \quad \overline{\mathbf{b}_{L2_{\text{pred}}}} = \begin{pmatrix} b_0 \\ b_H \\ b_{B1} \\ b_{\text{time}} \end{pmatrix}, \quad \overline{\boldsymbol{\varepsilon}_{B2}} = \begin{pmatrix} \varepsilon_{B2_1} \\ \varepsilon_{B2_2} \\ \dots \\ \varepsilon_{B2_{5348}} \end{pmatrix}, \quad \overline{\boldsymbol{\varepsilon}_{L2}} = \begin{pmatrix} \varepsilon_{L2_1} \\ \varepsilon_{L2_2} \\ \dots \\ \varepsilon_{L2_{5348}} \end{pmatrix}, \\ \overline{\mathbf{b}_{B2_{\text{pred}}}} = (\mathbf{B2}_{\text{pred}}^T \mathbf{B2}_{\text{pred}})^{-1} \mathbf{B2}_{\text{pred}}^T \mathbf{B2} = \begin{pmatrix} b_0 \\ b_H \\ b_{L1} \\ b_{\text{time}} \end{pmatrix} = \begin{pmatrix} -0,85 \\ 1,56 \\ 1,22 \\ 0,17 \end{pmatrix}, \quad \overline{\mathbf{b}_{L2_{\text{pred}}}} = (\mathbf{L2}_{\text{pred}}^T \mathbf{L2}_{\text{pred}})^{-1} \mathbf{L2}_{\text{pred}}^T \mathbf{L2} = \begin{pmatrix} b_0 \\ b_H \\ b_{B1} \\ b_{\text{time}} \end{pmatrix} = \begin{pmatrix} 0,45 \\ 3,45 \\ 2,4 \\ -0,88 \end{pmatrix}, \end{aligned} \quad (6)$$

Полученные коэффициенты (6) дают соответствующие регрессионные модели для прогнозирования широты и долготы точек падения второй ступени:

$$B2 = -0,85 + 1,56H + 1,22L1 + 0,17\text{time} + \varepsilon_{B2}, \quad L2 = 0,45 + 3,45H + 2,4B1 - 0,88\text{time} + \varepsilon_{L2}. \quad (7)$$

Моделям (7) соответствует система коэффициентов детерминации:

$$\left\{ \begin{aligned} R_{B2}^2 &= 1 - \frac{\frac{1}{5348} \sum_{i=1}^n (B2_i - (1,56H_i + 1,18L1_i + 0,17\text{time}_i - 0,85))^2}{\frac{1}{5348} \sum_{i=1}^n (B2_i - \overline{B2_n})^2} = 0,81616, \\ R_{L2}^2 &= 1 - \frac{\frac{1}{5348} \sum_{i=1}^n (L2_i - (0,45 + 3,45H_i + 2,4B1_i - 0,88\text{time}_i))^2}{\frac{1}{5348} \sum_{i=1}^n (L2_i - \overline{L2_n})^2} = 0,78158, \end{aligned} \right.$$

Однако, аналогичные расчёты (8) для системы одномерных парных регрессионных моделей (9), исключающих наличие коллинеарности между предикторами, показывают более высокие значения R^2 (10), а также скорректированного коэффициента детерминации R^2_{adj}

(11), учитывающего добавление в модель новых предикторов. Фрагмент резуль-

$$\begin{aligned} \mathbf{B2}_{\text{pred}} = \begin{pmatrix} 1 & B1_1 \\ 1 & B1_2 \\ \dots & \dots \\ 1 & B1_{5348} \end{pmatrix} = \begin{pmatrix} 1 & 38,5411 \\ 1 & 38,5411 \\ \dots & \dots \\ 1 & 38,2390 \end{pmatrix}, \quad \mathbf{B2} = \begin{pmatrix} B2_1 \\ B2_2 \\ \dots \\ B2_{5348} \end{pmatrix} = \begin{pmatrix} 38,5411 \\ 38,5411 \\ \dots \\ 38,2371 \end{pmatrix}, \quad \overline{\mathbf{b}_{B2_{\text{pred}}}} = \begin{pmatrix} b_0 \\ b_{B1} \end{pmatrix} = \begin{pmatrix} -2,34 \\ 1,18 \end{pmatrix}, \\ \mathbf{L2}_{\text{pred}} = \begin{pmatrix} 1 & L1_1 \\ 1 & L1_2 \\ \dots & \dots \\ 1 & L1_{5348} \end{pmatrix} = \begin{pmatrix} 1 & 36,3011 \\ 1 & 36,3011 \\ \dots & \dots \\ 1 & 43,4327 \end{pmatrix}, \quad \mathbf{L2} = \begin{pmatrix} L2_1 \\ L2_2 \\ \dots \\ L2_{5348} \end{pmatrix} = \begin{pmatrix} 36,3011 \\ 36,3011 \\ \dots \\ 43,4538 \end{pmatrix}, \quad \overline{\mathbf{b}_{L2_{\text{pred}}}} = \begin{pmatrix} b_0 \\ b_{L1} \end{pmatrix} = \begin{pmatrix} 1,87 \\ 3,77 \end{pmatrix}, \end{aligned} \quad (8)$$

$$B2 = -2,34 + 1,18B1 + \varepsilon_{B2}, \quad L2 = 1,87 + 3,77B2 + \varepsilon_{L2}, \quad (9)$$

$$R_{B2}^2 = 1 - \frac{\frac{1}{5348} \sum_{i=1}^n (B2_i - (1,18B1_i - 2,34))^2}{\frac{1}{5348} \sum_{i=1}^n (B2_i - \overline{B2_n})^2} = 0,9299863, \quad R_{L2}^2 = 1 - \frac{\frac{1}{5348} \sum_{i=1}^n (L2_i - (1,87 + 3,77L1_i))^2}{\frac{1}{5348} \sum_{i=1}^n (L2_i - \overline{L2_n})^2} = 0,9199757, \quad (10)$$

татов оценок качества полученных регрессионных моделей, проведённых на основе относительных критериев R^2 и R^2_{adj} , определяющих степень взаимозависимостей дисперсий соответствующих траекторных параметров, представлены в таблице 2 (тип моделей указан в терминологии языка программирования «R»). $R^2_{adj} = 1 - (1 - R^2) \frac{(n-1)}{(n-k)} \leq R^2$, (11)

Таблица 2 – R^2 и R^2_{adj} моделей L2

Тип модели L2 (отклик ~ предикторы)	R^2	R^2_{adj}
$L2 \sim H$	0,7295953	0,7295821
$L2 \sim H+B1$	0,7574387	0,7574265
$L2 \sim H+B1+time$	0,7815899	0,7815739
$L2 \sim H+B1+time+L$	0,8179307	0,8179217
$L2 \sim H+B1+time+L+t1$	0,8572334	0,8572305
$L2 \sim H+B1+time+L+t1+B$	0,8772338	0,8772307
$L2 \sim L1$	0,9199757	0,9199756

для n наблюдений k параметров. Совместные графики прогнозов (*fitted*) некоторых моделей и расчётных значений долготы второй степени ($L2$) показаны на рисунке 2. Построение и оценка качества моделей для параметра широты $B2$ показывают аналогичные зави-

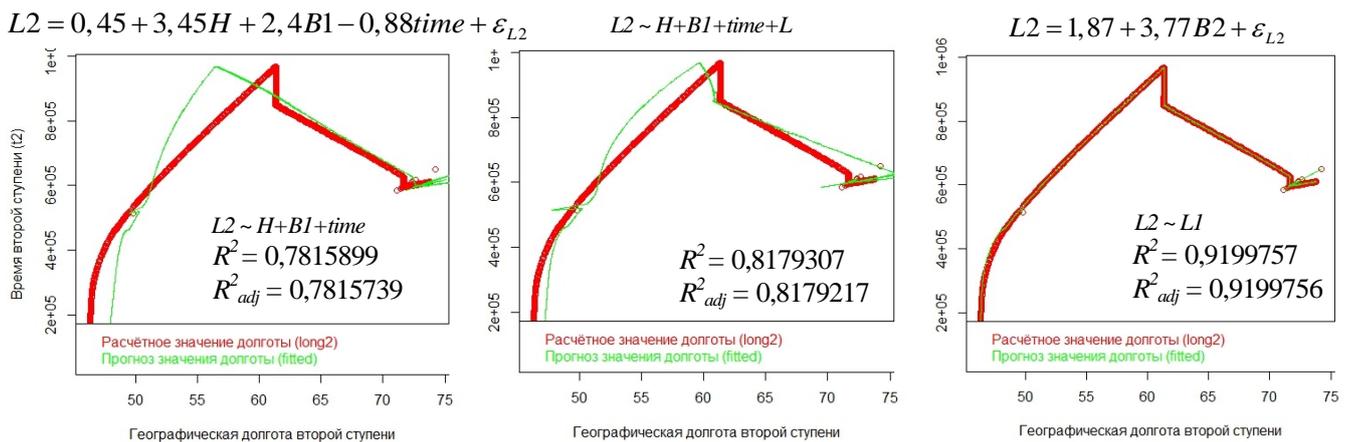


Рисунок 2 – Расчётные значения параметра долготы $L2$ и их прогнозы

симости. Выбросы на совместных графиках (точки в правом углу) обусловлены зашумлённостью или некачественной бортовой ТМИ, зарегистрированной в ходе лётного эксперимента.

Выводы:

- 1) Подтверждена возможность нахождения численных значений географических координат посредством прогнозирования с помощью линейных регрессионных моделей (9) с приемлемым уровнем точности;
- 2) Оптимальными регрессионными моделями, объясняющими максимальное количество дисперсии для данного случая (рисунок 1) являются одномерные парные регрессионные модели с наименьшим количеством регрессоров (« $B2 \sim B1$ », « $L2 \sim L1$ »);
- 3) С целью минимизации требований, предъявляемых к производительности аппаратно-программных средств ПИК и повышения уровня и оперативности аналитической оценки хода и результатов проведения пуска, расчётным путём (шаг 1 - 7) с нахождением численного решения системы дифференциальных уравнений методом Рунге-Кутты 4 порядка (1)-(4) целесообразно находить только численные значения

координат первой ступени объекта. Значения координат второй ступени и объекта в аварийной ситуации допускается получать с помощью прогнозирования (*вывод 1*).

В третьей главе в рамках второго направления исследований разработана эффективная методика оптимизации гиперпараметров модели классификатора, а также модель и алгоритм интеллектуального детектирования нештатных и аварийных состояний объекта на основе неявных признаков траекторных отклонений.

Состояние объекта РКТ в ходе проведения лётного эксперимента контролируется посредством сравнения обширного набора телеметрируемых параметров (контактных, функциональных и др.) с их эталонными значениями (допусковый контроль) в ручном или автоматическом режиме. В интересах оптимизации массива передаваемой измерительной информации исследована возможность контроля состояний объекта на основе неявных признаков, в качестве которых выступают траекторные отклонения объекта от его априорной траектории в зоне неопределённости состояний. С целью имитации возникновений нештатных ситуаций в исходный набор траекторных параметров (5348 наблюдений) внесены синтетические изменения, не характерные для естественного хода лётного эксперимента по принципу, показанному в координатной форме в таблице 3 (курсив) и таблице 4 (бинарное описание первых 8 наблюдений). Колонка «alarm» описывает возможные состояния объекта по принципу «норма (0)/возникла не норма (1)», что соответствует переменной состояния A таблицы 4. $A(x)$, $A(y)$, $A(z)$ описывают зависимость значения переменной состояния A от значений соответствующих координат. Построение модели первичного классификатора произведём по алгоритму решающего дерева, основанном на снижении критерия

Таблица 3 – Координатное описание состояний энтропии Шеннона:

time	x	y	z	alarm
51946,5	2922944,516	3058962,123	4757023,510	0
51946,6	2922944,550	3058962,230	4757023,920	1
51946,7	2922944,623	3058962,563	4757023,168	1
51946,8	2922944,672	3058963,805	4757024,102	1
...
51954,1	2922944,516	3058964,951	4757023,365	0

$$E = -p_{(+)} \log_2 p_{(+)} - p_{(-)} \log_2 p_{(-)}, \quad (12)$$

Таким образом, алгоритм снижения неопределённости зависимой переменной состояния A согласно (12), т.е. повышения точности классификации A , сводится к последовательному и эффективному учёту влияния

Таблица 4 – Бинарное описание состояний

x	y	z	$A(x)$	$A(y)$	$A(z)$
+	-	-	<i>+(alarm)</i>	<i>+(alarm)</i>	<i>+(alarm)</i>
+	-	+	<i>+(alarm)</i>	<i>+(alarm)</i>	<i>+(alarm)</i>
+	-	-	<i>+(alarm)</i>	<i>+(alarm)</i>	<i>+(alarm)</i>
-	+	+	<i>+(alarm)</i>	<i>+(alarm)</i>	<i>+(alarm)</i>
-	-	-	<i>-(normal)</i>	<i>-(normal)</i>	<i>-(normal)</i>
-	-	-	<i>-(normal)</i>	<i>-(normal)</i>	<i>-(normal)</i>
-	-	-	<i>-(normal)</i>	<i>-(normal)</i>	<i>-(normal)</i>
+	+	+	<i>-(normal)</i>	<i>-(normal)</i>	<i>-(normal)</i>

координат-предикторов (регрессоров x , y , z), а точнее – учёту частных энтропий по каждой из координат (13): *Шаг 1*. Найдем положительные и отрицательные частотные вероятности нештатных (аварийных) и штатных состояний $p_{(a)i}$, $p_{(n)i}$, при $i \in (x, y, z)$:

$$\left\{ \begin{array}{l} A(x_{(+)}) = (+++-) \\ p_{(a)x_{(+)}} = 3/4 = 0,75, \\ p_{(n)x_{(+)}} = 1/4 = 0,25, \\ A(x_{(+)}) \subset A(x), \end{array} \right\} \left\{ \begin{array}{l} A(x_{(-)}) = (+---) \\ p_{(a)x_{(-)}} = 1/4 = 0,25, \\ p_{(n)x_{(-)}} = 3/4 = 0,75, \\ A(x_{(-)}) \subset A(x), \end{array} \right\} \left\{ \begin{array}{l} A(y_{(+)}) = (+-) \\ p_{(a)y_{(+)}} = 1/2 = 0,5, \\ p_{(n)y_{(+)}} = 1/2 = 0,5, \\ A(y_{(+)}) \subset A(y), \end{array} \right\} \left\{ \begin{array}{l} A(y_{(-)}) = (+++---) \\ p_{(a)y_{(-)}} = 3/6 = 0,5, \\ p_{(n)y_{(-)}} = 3/6 = 0,5, \\ A(y_{(-)}) \subset A(y), \end{array} \right\} \left\{ \begin{array}{l} A(z_{(+)}) = (++-) \\ p_{(a)z_{(+)}} = 2/3 = 0,66, \\ p_{(n)z_{(+)}} = 1/3 = 0,33, \\ A(z_{(+)}) \subset A(z), \end{array} \right\} \left\{ \begin{array}{l} A(z_{(-)}) = (++---) \\ p_{(a)z_{(-)}} = 2/5 = 0,4, \\ p_{(n)z_{(-)}} = 3/5 = 0,6, \\ A(z_{(-)}) \subset A(z), \end{array} \right\}$$

Шаг 2. Найдем частные энтропии для каждого из возможных исходов согласно (12):

$$E_{1x} = -p_{(a)x(+)} \log_2 p_{(a)x(+)} - p_{(n)x(+)} \log_2 p_{(n)x(+)} = 0,81125, E_{2x} = -p_{(a)x(-)} \log_2 p_{(a)x(-)} - p_{(n)x(-)} \log_2 p_{(n)x(-)} = 0,81125,$$

$$E_{1y} = -p_{(a)y(+)} \log_2 p_{(a)y(+)} - p_{(n)y(+)} \log_2 p_{(n)y(+)} = 1, E_{2y} = -p_{(a)y(-)} \log_2 p_{(a)y(-)} - p_{(n)y(-)} \log_2 p_{(n)y(-)} = 1, \quad (13)$$

$$E_{1z} = -p_{(a)z(+)} \log_2 p_{(a)z(+)} - p_{(n)z(+)} \log_2 p_{(n)z(+)} = 0,923, E_{2z} = -p_{(a)z(-)} \log_2 p_{(a)z(-)} - p_{(n)z(-)} \log_2 p_{(n)z(-)} = 0,971,$$

Шаг 3. Введём дополнительный критерий прироста информативности ig_i по каждой координате: $ig_x = E(A) - E(A|x)$, $ig_y = E(A) - E(A|y)$, $ig_z = E(A) - E(A|z)$, $E(A) = 1$ – полная энтропия переменной состояния A , $E(A|x)$, $E(A|y)$, $E(A|z)$ – условные энтропии переменной состояния A , т.е. критерий неопределённости после учёта влияния координат-регрессоров, рассчитанный для n групп пропорционально общему числу наблюдений в группе $N = n_1 + n_2$.

Шаг 4. Рассчитаем условные энтропии переменной состояния. Для первых 8 наблюдений исходной траекторной выборки ($N_i=8$) будут справедливы выражения:

$$E(A|x) = E_{1x}n_{1x} / N_x + E_{2x}n_{2x} / N_x = 0,5E_{1x} + 0,5E_{2x}, \text{ при } n_{1x} = n_{2x} = 4,$$

$$E(A|y) = E_{1y}n_{1y} / N_y + E_{2y}n_{2y} / N_y = 0,75E_{1y} + 0,25E_{2y}, \text{ при } n_{1y} = 6, n_{2y} = 2,$$

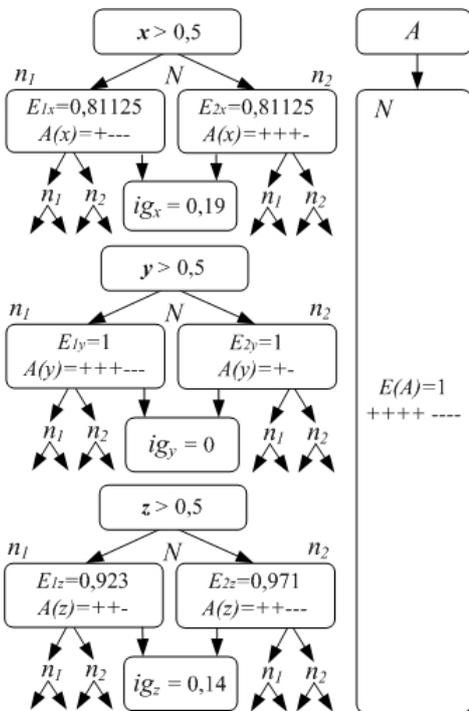
$$E(A|z) = E_{1z}n_{1z} / N_z + E_{2z}n_{2z} / N_z = 0,25E_{1z} + 0,75E_{2z}, \text{ при } n_{1z} = 3, n_{2z} = 5, \quad (14)$$

Шаг 5. Посредством (15) выявим наиболее информативную координату для первичного классификационного разбиения (*split-процесса*): $ig_x > ig_y > ig_z$.

$$ig_x = 1 - E(A|x) = 0,18875 \approx 0,19, \quad ig_y = 1 - E(A|y) = 0, \quad ig_z = 1 - E(A|z) = 0,143125 \approx 0,14, \quad (15)$$

Из (15) следует, что в отличие от координат y, z , учёт регрессора x больше помогает снизить значение полной энтропии за счёт $E(A|x)$. Помимо снижения критерия неопределённости отклика, учёт x также даёт большее значение критерия информативности $ig_x = 0,19$ (рисунок 3). Процесс учёта влияния координат-регрессоров с расчётом

Рисунок 3 – Split-процесс



критериев ig_i продолжается до достижения минимального значения $E(A)$, что позволяет безошибочно разделить бинарные наблюдения переменной состояния A на два класса. Таким образом, алгоритм построения бинарного классификатора сводится к следующим действиям: **Шаг 1.** В обучающем наборе траекторных данных $X(x,y,z)$ (обучающей выборке) на основании заранее заданного критерия $Q(X(x,y,z), j, s)$ находится наилучшее ее разбиение на две части: $R_1(j, s) = \{(x, y, z) | (x, y, z)_j \leq s\}$, $R_2(j, s) = \{(x, y, z) | (x, y, z)_j > s\}$.

Шаг 2. После того как найдены наилучшие значения j и s , в соответствии с условием $[(x, y, z)_j \leq s]$ создаётся корневая вершина дерева классификации. Родительские объекты разбиваются на две части: часть из них направляется в правое поддерево, оставшаяся часть – в левое. **Шаг 3.** Для каждой из полученных групп многократно повторяется процедура построения дочерних вершин. **Шаг 4.** В случае, когда после очередного разбиения вы-

борки в одной из частей оказываются объекты только одного класса, создаётся листовая вершина, описывающая класс бортовых состояний попавших в нее объектов.

Предлагаемая методика оптимизации гиперпараметров классификатора минимизирует переобучение первичного алгоритма классификации бортовых состояний $a = \mu X_{(x,y,z)}^m$ относительно контрольной выборки координат $X_{(x,y,z)}^k$, т.е. разность

$\delta(a, X_{(x,y,z)}^m, X_{(x,y,z)}^k) = Q(a, X_{(x,y,z)}^k) - Q(a, X_{(x,y,z)}^m)$, $X_{(x,y,z)}^m = \{x_1, \dots, x_m; y_1, \dots, y_m; z_1, \dots, z_m\}$ – выборка координат из множества $X_{(x,y,z)}$, за счёт:

1) Разбиения исходного набора координат $X_{(x,y,z)}^L = X_{(x,y,z)}^m \amalg X_{(x,y,z)}^k$ на обучающую $X_{(x,y,z)}^m$ и временно недоступную алгоритму тестовую $X_{(x,y,z)}^k$ выборки;

2) Скользящей перекрёстной валидации (разновидности «*k-fold*»-обработки исходного набора данных, устраняющей проблему обучения модели на одной и той же обучающей выборке путём разбиения исходного набора данных на *k* частей с последующим обучением модели на *k-1* частях. Оставшаяся часть данных при этом отводится тестовой выборке. В результате *k* повторений, каждая из *k* частей проходит через тестовую выборку, обеспечив наиболее равномерное использование имеющегося набора данных для оценки эффективности модели классификатора). Вариант, применительный к координатным временным рядам показан на рисунке 4;

3) Ограничения максимальной глубины ветвления дерева классификатора согласно алгоритму (рисунок 5):

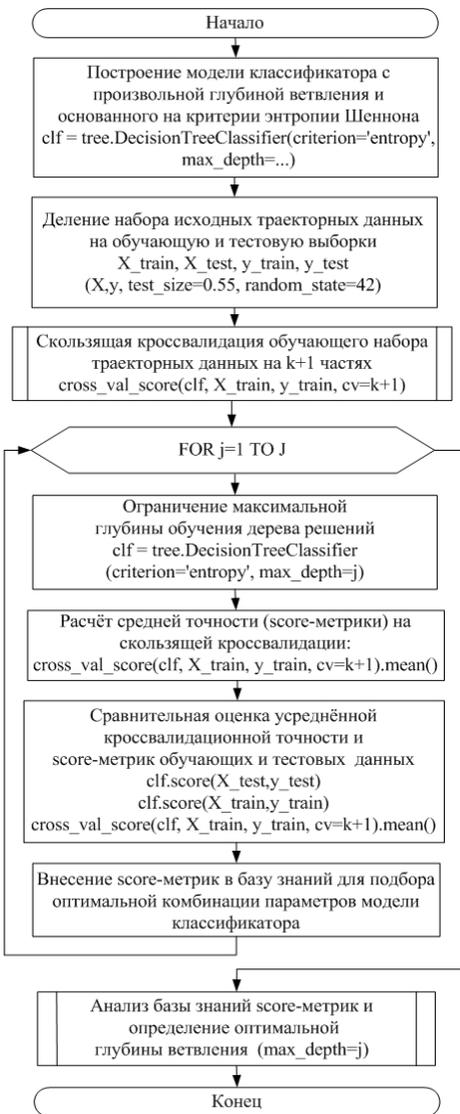


Рисунок 4 – Скользящая валидация координатного временного ряда на *k+1* частях

Шаг 1. Построение первичной модели классификатора с произвольной глубиной ветвления дерева решений и разделение данных на обучающую и тестовую выборки.

Шаг 2. Сравнительная оценка количества верных классификаций первичной модели по отношению к сумме всех возможных классификаций, включая ложные на обучающих и тестовых координатных выборках (*score*).

Шаг 3. Скользящая кроссвалидация обучающей выборки координат $X_{(x,y,z)}^m$ (применительно к временным рядам «*time series data*») на *k+1* частях с поэтапной оценкой *score*-метрик и средней точности (рисунок 4).

Шаг 4. Цикл подбора оптимальной глубины обучения (*max_depth*) с расчётом соответствующих *score*-метрик и внесением их в базу знаний (таблица 5).

Рисунок 5 – Подбор оптимальной глубины ветвления

Шаг 5. Анализ базы знаний и выбор наиболее эффективной глубины обучения первичной модели классификатора, соответствующей оптимальному соотношению метрик обучающей, тестовой и кроссвалидационной выборок (*max_depth*=7 при *k-fold_score*=0,805434 в таблице 5 и рисунке 6).

Таблица 5 – Фрагмент базы знаний

max_depth	score-метрика		
	train	test	k-fold
1	0.781879	0.796610	0.781961
...
6	0.867450	0.769492	0.799031
7	0.882550	0.776271	0.805434
...
20	0.978188	0.772881	0.771877

Для улучшения вычленения классификатором генерализованной закономерности детектирования нештатных состояний на борту объекта, требуется переформулировка задачи в контекст метрик эффективности, т.е. оптимизации процесса классификации посредством ввода дополнительных критериев качества модели помимо её *score*-точности. Такая необходимость также продиктована и тем, что возникновение нештатных и аварийных ситуаций при проведении лётных испытаний сложных изделий РКТ является достаточно редким, но чреватым серьезными последствиями явлением, что в свою очередь, приводит к выраженному дисбалансу классов при формировании обучающего и тестового поднаборов данных.

В качестве меры выравнивания соотношения классов предлагается искусственная модификация исходных данных (сегментация и корректировка весов): $\beta_{mnc} = 5\beta_{mjc}$, где β_{mnc} – вес минорного класса, а β_{mjc} – вес доминирующего класса.

Качественными метриками (метриками эффективности) выступают:

- 1) точность классификации – доля истинно положительных состояний объекта, определенных моделью классификатора как положительные (*p*-метрика – способность различать классы);
- 2) полнота классификации – доли истинных положительных классификаций из всех положительных классификаций, сделанных моделью (*r*-метрика – способность модели к распознаванию определённого класса).

Оценка модели, полученной посредством предложенной методики оптимизации гиперпараметров, производится с помощью графика ROC-кривой, показывающий оптимальное соотношения показателей полноты и точности классификаций (рисунок 7). Для выборки $X_{(x,y,z)}$ из распределения P^X задан статистический критерий $f := \mathbb{R}^n \longrightarrow \{H_0, H_1\}$ для сопоставления каждой реализации $X_{(x,y,z)}$ нулевой (H_0 – штатное состояние) либо альтернативной (H_1 – нештатная или аварийная ситуация) гипотезе. Возможны ошибки 1 рода при $f(x,y,z) = H_1$ при соответствии P^X гипотезе H_0 и ошибки 2 рода при $f(x,y,z) = H_0$ при соответствии P^X гипотезе H_1 . Для построения ROC-кривой бинарная классификация определялась путём варьирования специфического порога классификации, в зависимости от которого определялись значения статистических ошибок 1 и 2 рода: для каждого значения порога в интервале $[0;1]$ с шагом $d_x = 0,01$ рассчитывались значения метрик точности и полноты с построением графика их взаимозависимости. Оценка ROC-кривой, произведённая с помощью определения *AUC*-площади, описанного ей многогранника и рассчитанная с помощью численного метода трапеций (10) выявила показатель $AUC = 0,85$ (высокий).

$$AUC = \int f(x)dx = \sum_i \left[\frac{ДЛА_{i+1} + ДЛА_i}{2} \right] (r_{i+1} - r_i) = 0,85, \quad (10)$$

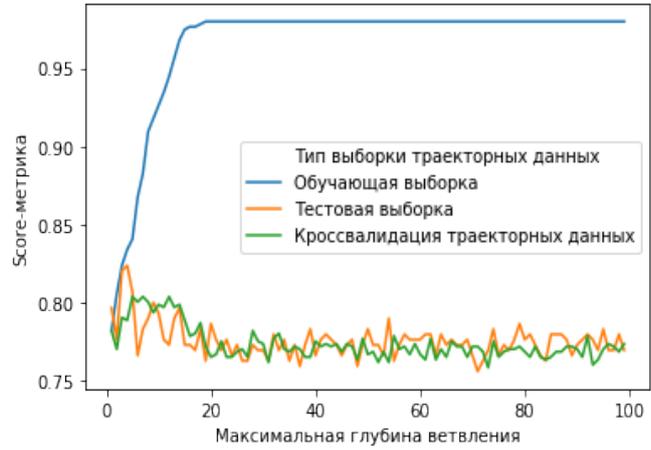


Рисунок 6 – Результаты анализа базы знаний

где $ДЛА_i$ и r_i – это численные значения метрик специфичности (доли ложных аварий) и полноты (r -метрики) на осях абсцисс и ординат соответственно. Матрица вероятных ошибок полученной первичной модели классификатора на тестовом наборе траекторных данных объекта представлена на рисунке 8.

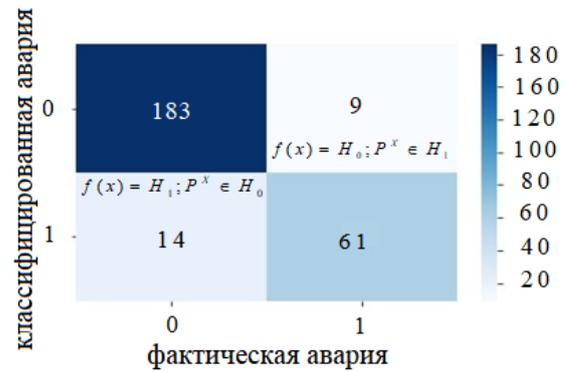
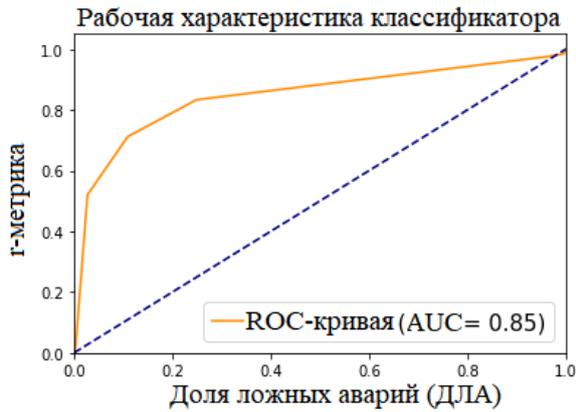


Рисунок 7 – ROC-кривая модели Рисунок 8 – Матрица вероятных ошибок

Помимо максимальной глубины ветвления, подбор оптимальных соотношений метрик качества также реализован с помощью коррекции допустимого размера выборки, после которого допускается последующее разбиение (`min_samples_split`) и размера выборки, допустимого для отнесения в отдельный исход (`min_samples_leaf`).

С учётом всех вышеперечисленных мер и применения алгоритма «поиска по сетке» оптимальных параметров модели классификатора получена оптимальная первичная модель классификатора бортовых состояний объекта РКТ на основе неявных признаков траекторных отклонений, основанная на энтропийном критерии с максимальной глубиной ветвления 7 при показателе метрики $AUC = 0,85$ (таблица 6).

Таблица 6 – Характеристики первичной модели

Параметр классификатора	Значение
критерий	энтропийный
максимальная глубина ветвления	7
<code>min_samples_split</code>	2
<code>min_samples_leaf</code>	1
$score$ -метрика	0,80543
r -метрика	0,87235
p -метрика	0,81032

Анализ значимости предикторов (регрессоров) первичной модели выявил следующие показатели: « x » = 0,427552 (43%), « y » = 0,302553 (30%), « z » = 0,269896 (27%). Полученная первичная модель классификатора показана на рисунке 9. Штатным состояниям объекта, соответствующим нормальному

функционированию объекта в вершинах модели классификатора присвоена метка «`class = Normal state`», возникновению на борту аварийных или нештатных состояний соответствует метка «`class = Alarm`». Сравнительный ROC-график, подтверждающий эффективность предложенной методики оптимизации гиперпараметров относительно существующих методов при равномерном распределении траекторных данных на обучающий и тестовый наборы показан в левом верхнем углу рисунка 9.

Заключительным этапом построения модели оптимального классификатора бортовых состояний объекта на основе неявных признаков траекторных отклонений является интеграция полученной первичной модели классификатора в алгоритм «случайного леса» – единого комитета классификаторов решающих деревьев. Результат классификации каждого из субклассификаторов на соответствующей бутстреп выборке (уникальной подвыборки исходного набора $X(x, y, z)$ малого объёма) усредняется, исходя из чего делается итоговое заключение о целевой метке класса. Характеристики итоговой модели классификатора представлены в таблице 7. Сравнительный ROC-график, подтверждающий эффективность итоговой модели относительно

существующих методов при равномерном распределении данных показан в левом нижнем углу рисунка 9.

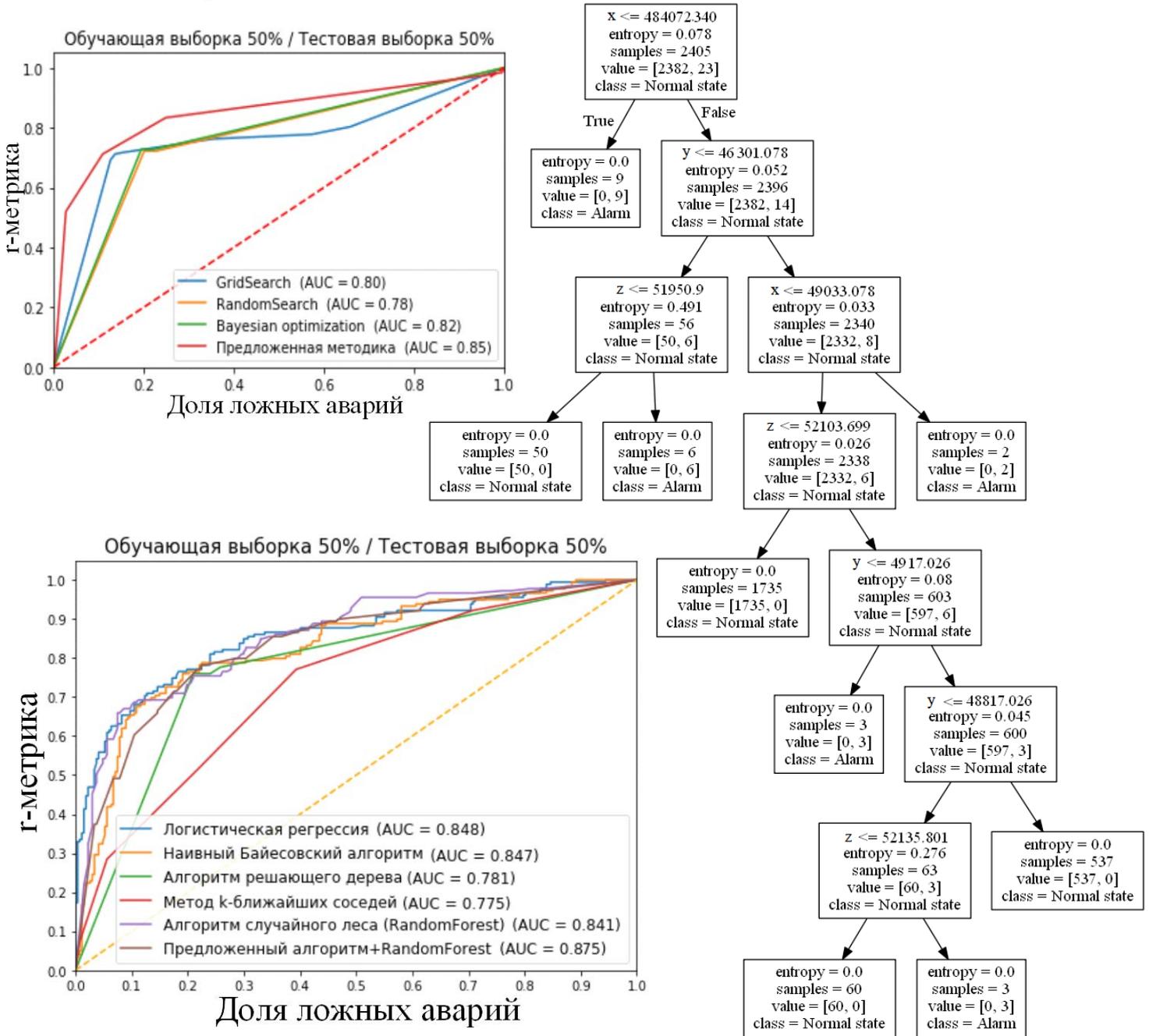


Рисунок 9 – Первичная модель классификатора бортовых состояний

Таблица 7 – Характеристики итоговой модели

Параметр классификатора	Значение
количество поддеревьев	30
критерий ветвления	энтропийный
порог глубины ветвления	7
score	0,87213
recall	0,89532
precision	0,85235

Таким образом, общая последовательность действий принимает следующий вид:

Шаг 1. Построение первичной модели классификатора состояний объекта на основе алгоритма решающего дерева.

Шаг 2. Зарегистрированная бортовая телеметрическая информация, содержащая траекторные измерения принимается в качестве исходного набора данных $X_{(x,y,z)}$ для первичной модели классификатора состояний.

Определяется критерий качества $Q(X_{(x,y,z)}, j, s)$. Находится наилучшее разбиение $X_{(x,y,z)}$ на обучающую и тестовую выборки: $R_1(j, s) = \{(x, y, z) | (x, y, z)_j \leq s\}$, $R_2(j, s) = \{(x, y, z) | (x, y, z)_j > s\}$.

Шаг 3. После нахождения наилучших значений j и s , в соответствии с условием $[(x, y, z)_j \leq s]$ создаётся корневая вершина дерева классификации модели классификатора состояний объекта.

Шаг 4. Запускается процесс оптимизации гиперпараметров первичной модели классификатора состояний методом согласно предложенной методике, включающий в себя анализ базы знаний метрик точностей и полноты классификации на обучающем, тестовом и кроссвалидационном наборах данных.

Шаг 5. По результатам анализа ROC-кривых базы знаний метрик рассматриваемых классификаторов, определяется оптимальная первичная модель классификатора состояний объекта с наиболее эффективными параметрами максимальной глубины ветвления дерева классификации, минимального размера выборки для отнесения в отдельный исход и минимального размера выборки после которого допускается последующее разбиение.

Шаг 6. Производится построение итоговой модели классификатора состояний объекта по принципу алгоритма «случайного леса», в качестве поддеревьев которой выступают экземпляры полученной ранее первичной модели.

Шаг 7. Определение оптимальных параметров итоговой модели.

Выводы:

1) В результате применения предложенной модели и алгоритма, достигается положительный эффект, заключающийся в возможности оперативной классификации состояний бортовых систем сложного многопараметрического объекта РКТ исходя из неявных признаков отклонений объекта от его априорной траектории.

2) Эффективность предложенного алгоритма подтверждается уровнем точности и полноты классификаций реализованной на его основе модели классификатора бортовых состояний 87% и 89% соответственно с качественным показателем $AUC = 0,875$.

3) Подтверждена возможность эффективного использования предложенного алгоритма в качестве дополнения к основным методам контроля бортовых состояний (контроля массива параметров цифровой ТМИ), а также самостоятельного использования в качестве интеллектуального инструментария парирования нештатных ситуаций, таких как запуск аварийных циклограмм и минимизация рисков возможных нежелательных последствий аварий.

Анализ эффективности выносимых на защиту положений. Выполненные экспериментальные исследования в режиме послесекансной обработки с привлечением натурной траекторной информации 7 типов сложных многопараметрических изделий РКТ выявили следующее:

По первому положению. Предложенный алгоритм расчёта баллистической траектории с использованием линейной регрессии предъявляет меньшие требования к вычислительным средствам при повышении оперативности обработки измерительной информации в среднем на 21,4% по сравнению с существующими алгоритмами.

По второму положению. Первичная модель классификатора, полученная посредством предложенной методики оптимизации гиперпараметров по показателям AUC уступает 1,25% Байесовской оптимизации в условиях больших объёмов обучающей выборки (train 90% / test 10%), однако эффективнее существующих на 1,33% в условиях малых объёмов обучающей выборки данных и на 6,25% в условиях равномерного распределения с частотой ошибок первого и второго рода 3,34% и 5,24% соответственно.

По третьему положению. Разработанный алгоритм интеллектуального детектирования аварийных состояний обеспечивает повышение точности классификаций текуще-

го состояния бортовых систем по косвенному признаку в зоне неопределённости состояний эффективнее аналогичных, построенных на основе логистической регрессии, наивного Байесовского подхода, дерева решений, простого случайного леса и метода k -ближайших соседей в условиях любого распределения исходных траекторных данных на обучающий и тестовый наборы, в том числе на 6,97% при равномерном распределении исходных траекторных данных (train 50% / test 50%).

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

1. Рассмотрены и проанализированы современные подходы к обработке и анализу траекторной информации сложного изделия РКТ в ходе проведения лётных экспериментов. По результатам анализа существующих подходов, выявлена необходимость в проведении научного исследования возможности эффективного применения методов предиктивной аналитики и интеллектуального анализа данных по двум направлениям: регрессионного анализа данных и задач их классификации.

2. В рамках первого направления исследований разработаны модели и алгоритм расчёта баллистической траектории отделяющихся частей объекта РКТ с применением линейной регрессии, позволяющий с допустимым уровнем точности (значения скорректированных коэффициентов детерминации регрессионных моделей $R^2_{adj} = 0,9299862$ и $R^2_{adj} = 0,9199756$) формировать признаки попадания и промаха на основе прогнозов значений географических координат районов падения. Разработанные модели и алгоритм отличаются от известных тем, что при их реализации применяются элементы теории машинного обучения и прогнозирования численных значений траекторных параметров посредством методов регрессионного анализа, что позволяет оперативно определять координаты отделяющихся частей объекта и обеспечивает высокий уровень автоматизации процессов обработки и анализа траекторной ТМИ, снижая при этом требования к производительности аппаратно-программных средств полигонного измерительного комплекса и влияние «человеческого фактора»;

3. В рамках второго направления исследований разработана эффективная методика оптимизации гиперпараметров модели классификатора, а также модель классификатора и алгоритм детектирования возникновения нештатных и аварийных состояний объекта по неявным признакам траекторных отклонений. Разработанная модель и алгоритм выгодно отличаются от известных тем, что за счёт современных методов классификации, применяемых при их реализации, позволяют существенно повысить оперативность выработки решений, направленных на парирование аварийных ситуаций в ходе лётных испытаний с целью предотвращения нежелательных последствий. Также предложенная модель и алгоритм способны выступать и в качестве самостоятельных средств интеллектуального анализа траекторных измерений, что подтверждается показателями метрик эффективности итоговой модели классификатора бортовых состояний с уровнем точности классификации 87% и полноты классификации 89%.

Использование результатов диссертации позволяет:

- снизить вычислительную нагрузку на аппаратно-программные средства полигонного измерительного комплекса за счёт минимизации объёма производимых математических вычислений в пользу интеллектуального прогнозирования;

- повысить эффективность и расширить функциональные возможности штатного персонала испытателей в ходе натурных работ со сложным многопараметрическим изделием РКТ за счёт повышения уровня автоматизации процессов обработки, анализа и представления ТМИ;

- повысить объективности анализа и оценки телеметрической информации, зарегистрированной в ходе лётных испытаний;
- оперативно детектировать возникновения нештатных и аварийных ситуаций на борту объекта и сократить время принятия решений по их парированию;
- автоматизировать процесс накопления базы знаний состояний конкретных типов изделий и её эффективного использования;
- разрабатывать перспективные интеллектуальные системы обработки и анализа ТМИ на основе современных методов искусственного интеллекта.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в изданиях, рекомендованных ВАК РФ

1. Балашов С.Ю., Спицын С.В. Особенности обработки, анализа и представления траекторных измерений при проведении пусков изделий РКТ // Вестник Рязанского государственного радиотехнического университета. 2019. №69. С. 175-184. DOI:10.21667/1995-4565-2019-69-175-184
2. Спицын С.В., Товпеко А.В., Тихомиров С.А. Алгоритм обработки траекторных измерений при проведении ракетных пусков // Вестник Рязанского государственного радиотехнического университета. 2019. № 70. С. 118-126. DOI: 10.21667/1995-4565-2019-70-118-126
3. Спицын С.В., Товпеко А.В. Алгоритм расчета целеуказаний измерительным средствам // Радиотехника №11(18), 2019., С. 28-34. DOI: 10.18127/j00338486-201911(18)-04
4. Богданов А.А., Спицын С.В., Тихомиров С.А. Методы регрессионного анализа в обработке траекторной информации изделий ракетно-космической техники // Вестник Рязанского государственного радиотехнического университета. 2020. №74. С. 80-88. DOI:10.21667/1995-4565-2020-74-80-88
5. Спицын С.В., Тихомиров С.А. Модель классификатора бортовых состояний изделия ракетно-космической техники на основе неявных признаков траекторных отклонений // Вестник Рязанского государственного радиотехнического университета. 2021. №75. С. 133-141. DOI:10.21667/1995-4565-2021-75-133-141

Публикации в изданиях, рекомендованных ВАК РФ и входящих в международные реферативные базы данных (Scopus b Web of Science)

6. Spitsyn S.V., Potapov V.I., Koryachko V.P. The algorithm of processing and presenting the digital trajectory data obtained when launching rocket vehicles // IOP Conference Series: Materials Science and Engineering (MSE). 2021. Vol. 1047, 012007(1-7). DOI:10.1088/1757-899X/1047/1/012007

Публикации в других изданиях

7. Спицын С. В., Королев А. В. Циклический способ формирования электронных графиков работ и эксплуатации сложных технических объектов // Международный научно-технический форум СТНО-2017. Сборник трудов. Том 4. Рязань. 2017. С. 268-271.
8. Спицын С. В., Товпеко А. В. Основные показатели качества систем сбора и обработки измерений при проведении пусков изделий РКТ // Международный научно-технический форум СТНО-2018. Сборник трудов. Том 5. Рязань. 2018. С. 211-217.
9. Спицын С. В., Товпеко А. В. Актуальные проблемы расчета целеуказаний измерительным средствам при обеспечении ракетных пусков // Международный научно-технический форум СТНО-2019. Сборник трудов. Том 4. Рязань. 2019. С. 188-193.
10. Спицын С. В. Обработка и анализ траекторных измерений при проведении полигонных испытаний // Международный научно-технический форум СТНО-2020. Сборник трудов. Том 6. Рязань. 2020. С. 173-176.
11. Спицын С. В. Применение линейной регрессии в прогнозировании баллистической траектории // Всероссийская научно-техническая конференция «Актуальные проблемы современной науки и производства». Сборник трудов. Рязань. 2020.

Спицын Сергей Владимирович

**МОДЕЛИ И АЛГОРИТМЫ ОБРАБОТКИ ТРАЕКТОРНОЙ ИНФОРМАЦИИ
НА ОСНОВЕ МЕТОДОВ КЛАССИФИКАЦИИ И
РЕГРЕССИОННОГО АНАЛИЗА**

А в т о р е ф е р а т

диссертации на соискание ученой степени
кандидата технических наук

Бумага офисная. Печать цифровая.
Тираж 100 экз.

ФГБОУ ВО Рязанский государственный радиотехнический университет
имени В.Ф. Уткина
390005, Рязань, ул. Гагарина, д. 59/1