

ОБНАРУЖЕНИЕ АНОМАЛИЙ НА ОСНОВЕ СУРРОГАТНЫХ МОДЕЛЕЙ

Бурнаев Е. В.¹

(Сколковский институт науки и технологий, Москва)

Методы предиктивного технического обслуживания используются для скорейшего обнаружения существенных изменений (разладок) в работе механизмов. Основной целью этого подхода к обслуживанию является непрерывный мониторинг и корректировка технического состояния или полная замена механизмов до того, как выявленные изменения станут критичными для работы отдельных компонент или системы в целом. Диагностические возможности методов предиктивного обслуживания значительно возросли в последние годы в связи с улучшением сенсорных технологий наблюдения и разработкой новых алгоритмов обработки информации. Использование предсказательного обслуживания имеет ряд преимуществ по сравнению с другими практикуемыми подходами к обслуживанию, а именно, возможность вести более точный и своевременный контроль работоспособности отдельных узлов и всей системы в целом; постоянный мониторинг и анализ внутреннего и внешнего состояния повышает безопасность работы и позволяет более оперативно и в некоторых случаях превентивно реагировать на возможные аварии и выходы из строя; значительное сокращение затрат на техническое обслуживание благодаря исключению плановых замен технически исправных и надежных узлов системы. В данной работе рассматривается задача построения предсказательных моделей (суррогатных моделей) для решения задачи предиктивного технического обслуживания. Проанализированы особенности задач обнаружения аномалий и прогнозирования поломок. Проведен обзор основных потребностей индустриальных приложений и описана структура соответствующих систем предиктивного технического обслуживания. Приведен пример применения методов обнаружения аномалий на основе суррогатных моделей для предиктивного технического обслуживания лопаточных машин.

Ключевые слова: предиктивная (предсказательная) аналитика, индустриальная инженерия, машинное обучение, предсказательное техническое обслуживание, предсказание поломок.

1. Введение

В сложных технических системах возникновение поломок и неисправностей зачастую приводит к серьезным убыткам.

¹ Евгений Владимирович Бурнаев, к.ф.-м.н., доцент (e.burnaev@skoltech.ru).

Например, при непрерывном производстве неисправность одного из элементов производственной цепочки может приводить к полной или частичной остановке конвейера и длительному простоя. В то же время современные промышленные станки и агрегаты комплектуются системами телеметрии, отражающими ключевые показатели работы устройств, а также их текущее техническое состояние.

Современное развитие математического аппарата (машинное обучение и когнитивные технологии) и возможности построения полного стека сбора, хранения и обработки собираемых данных позволяют внедрять новые подходы к проведению технического обслуживания, автоматизации и оптимизации производства. Так называемое предсказательное техническое обслуживание (ПТО), или обслуживание по состоянию, обладает рядом преимуществ по сравнению с другими подходами, позволяя

а) в режиме реального времени контролировать состояние большого числа подключенных устройств (информационный сервис);

б) проводить оптимизацию хранения и доставки запасных частей и составлять эффективное расписание ремонтов (логистический сервис);

в) снижать затраты на проведение технического обслуживания (экономический сервис);

г) повышать эффективность производства за счет сокращения числа незапланированных остановок и простоев (производственный сервис).

Внедрение предсказательного технического обслуживания на производстве предоставляет уникальные новые возможности¹, но в то же время сопряжено с рядом трудностей. Можно выделить две наиболее крупные задачи:

¹ Например, внедрение системы предсказательной аналитики для обслуживания паровых турбин позволяет сократить до 50% затраты на обслуживание и повысить время жизни паровых турбин до трех раз [35].

а) построение централизованной инфраструктуры сбора телеметрии, хранения и обработки данных, проведения вычислений, необходимых для определения текущего технического состояния и предсказания поломок, и т.д.;

б) разработка и построение прогнозных моделей на основе собираемой телеметрии для предсказания типичных поломок, определения оставшегося ресурса работы элементов системы и решения смежных задач.

Вторая задача является ключевой для построения системы предсказательного технического обслуживания, так как зачастую производители оборудования не могут предусмотреть всех внештатных ситуаций и предоставить точные физические модели поведения показателей телеметрии в предкритических и критических состояниях. Однако практика показывает, что такие модели можно строить на основе собираемых данных с датчиков на оборудовании. Для этого необходим современный набор инструментов машинного обучения и прогностической аналитики.

В общем случае входные данные для построения модели (в зависимости от конкретного оборудования будут требоваться различные подмножества указанных ниже данных) состоят из следующих компонентов:

а) телеметрия (значения различных характеристик, снимаемых с оборудования, с привязкой ко времени);

б) данные по отказам оборудования и другим инцидентам, классификатор типов поломок;

в) данные по техническому обслуживанию и ремонту, классификатор типов ремонта;

г) периоды времени, в пределах которых оборудование работало в штатном режиме, и т.п.

Обученная модель позволит в зависимости от постановки задачи и типов возникающих неисправностей:

1. Детектировать аномальное поведение системы.
2. Прогнозировать поломки.
3. Прогнозировать оставшийся ресурс работы узлов и агрегатов.

2. Особенности задач обнаружения аномалий и прогнозирования поломок

Главная идея, на основе которой проводится прогнозирование поломок, состоит в следующем очевидном наблюдении: оборудование (скажем, турбина) обычно входит в предаварийное состояние из-за появления каких-то мелких дефектов, таких как трещины или протечки. Эти дефекты со временем усугубляются и приводят к критическим сбоям, вплоть до полного отказа оборудования. Соответственно, неожиданно происходящие события – например, появление этих мелких дефектов – могут повлиять на поток информации с датчиков и привести к возникновению аномальных наблюдений в данных.

Аналогичные задачи возникают не только в промышленности, но и, например, в банковской сфере, когда необходимо детектировать аномальное поведение клиентов, которое может быть связано с мошенничеством.

Содержательная задача прогнозирования поломок формулируется как задача обнаружения «разладок» (аномалий, change points – «точек излома») в процессах, описывающих поведение исследуемой системы, и классификации выявленных «нестандартных» паттернов. Стандартным подходом к обнаружению аномалий в данных является использование контрольных карт и более общих методов, таких как методы скорейшего обнаружения разладки [15, 16, 21, 25].

Несмотря на то, что различные методы обнаружения разладок и их классификации предложены и исследованы достаточно давно, особенности задачи прогнозирования поломок потребовали решения задач прогнозирования, обнаружения разладки и классификации в принципиально новых математических постановках, связанных с особенностью наблюдаемых и исследуемых процессов.

Основные отличия функциональности, которую необходимо разработать для решения задач предсказательной аналитики, от функциональности для анализа простоев оборудования на основе контрольных карт, состоят в следующем:

1. Стандартные методы обнаружения разладок/аномалий в большинстве своем используют достаточно ограничивающие параметрические предположения об источнике данных, не рассчитаны на обнаружение разладок в многомерных сигналах в условиях сверхбольших массивов данных, а также на автоматическое выявление каналов, по которым происходит разладка. Например, стандартные методы прогнозирования появления событий различных классов в потоке многомерных данных работают в предположениях малой размерности данных, а также сравнимого числа примеров для каждого класса событий.

2. Контрольные карты позволяют проводить анализ причин простоя оборудования только постфактум. В общем же случае необходима разработка методов в случае наличия исторической разметки данных (данные о времени и типе случившейся поломки), которые позволяют прогнозировать будущие (на заданном горизонте) поломки оборудования, обеспечивая низкий (контролируемый) уровень ложных тревог и значимое покрытие (долю поломок в будущем, «покрытых» прогнозом). При этом количества примеров рассматриваемых классов событий обычно не сбалансированы (поломки происходят редко) [19, 49].

3. После детектирования разладки или прогнозирования поломки необходимо автоматическое выделение небольшого подмножества параметров, наиболее значимых с точки зрения полученного результата применения модели. Такого рода автоматический анализ позволит локализовать возможные причины разладки/поломки. При этом стандартные методы выделения групп связанных параметров используют линейные меры корреляции и т.п.

4. Многие из вышеперечисленных методов содержат свободные гиперпараметры, которые приходится настраивать вручную при применении методов к новым данным. При этом разработка соответствующего алгоритмического ядра ПТО требует, чтобы была разработана и реализована единая методология анализа и обработки данных, которая комбинирует все необходимые методы (предобработки данных, выделения групп связанных параметров, снижения размерности и выделения наиболее значимых параметров, построения суррогатных моде-

лей, кластеризации и классификации на графах) и с минимальным привлечением человека и знаний из предметной области позволяет для нового источника данных строить предсказательные модели для прогнозирования состояния системы.

В свою очередь, для прогнозирования поломок необходимо одновременно наблюдать несколько процессов, отражающих различные характеристики исследуемой системы, так как «нестандартное» функционирование отдельных компонентов системы, предшествующих поломке, приводит к изменениям сразу в нескольких процессах, описывающих поведение системы. Для надежного прогнозирования поломок число наблюдаемых процессов должно быть достаточно велико, и феномен «проклятия размерности» присущ многим проблемам при исследовании поведения многомерных процессов. Поэтому для эффективного решения задачи прогнозирования поломок необходимо сводить задачу к исследованию «агрегированных» процессов, «сохраняющих» всю необходимую информацию для решения исходной задачи, но имеющих существенно меньшую размерность. Использование специальных процедур снижения размерности особенно важно при решении задачи несбалансированной классификации, в которой феномен проклятия размерности может приводить к полной «неработоспособности» большинства разработанных и используемых алгоритмов.

Второй принципиальной особенностью наблюдаемых многомерных процессов является тот факт, что имеются различные зависимости и взаимосвязи между компонентами этих процессов, описывающих поведение одной и той же системы, и различные «совместные» ограничения на их значения. Поэтому множество различных значений этих процессов, полученных при различных условиях функционирования системы, занимает лишь небольшую часть высокоразмерного пространства наблюдений, имеющую невысокую внутреннюю размерность. Эта особенность данных и позволяет решить задачу снижения размерности путем нахождения низкоразмерной параметризации носителя многомерных данных.

В силу указанных особенностей адекватной математической моделью наблюдаемых многомерных временных рядов

является так называемая модель многообразия, в соответствии с которой многомерные временные ряды, полученные при наблюдении реальных сложных систем, лежат на неизвестном многообразии невысокой размерности (низкоразмерной поверхности, обладающей некоторыми специфическими свойствами), вложенной в высокоразмерное пространство наблюдений. Эта модель, предложенная только в 2000 году и описывающая большинство процессов, встречающихся в реальной жизни, является в настоящее время общепринятой и используется в современных алгоритмах анализа многомерных данных. Методы анализа многомерных данных в условиях справедливости этой гипотезы определили новое направление в машинном обучении и анализе данных, получившее название «Моделирование многообразий» (Manifold learning) [14].

Таким образом, требуется проведение фундаментальных и поисковых исследований для разработки эффективных решений задач ПТО.

В рамках проведения фундаментальных исследований необходима разработка принципиально новых методов: восстановления характеристик объекта по неточной и/или неполной информации о других его характеристиках, включая методы восстановления зависимостей по данным; прогнозирования (предсказания) данных, генерируемых источником с неизвестными статистическими характеристиками; аппроксимации многообразий данных, которые будут использованы в задачах снижения размерности и выделения значимых признаков; обнаружения разладки и аномалий (выбросов) в данных; вычислительной геометрии и топологии с целью построения эффективных алгоритмов снижения размерности, прогнозирования и классификации данных.

В рамках проведения поисковых исследований необходимы: разработка новых и адаптация существующих алгоритмов и методов решения задачи несбалансированной классификации; разработка новых и адаптация существующих алгоритмов и методов отбора признаков в задачах несбалансированной классификации; разработка новых и адаптация существующих алгоритмов и методов выявления аномалий (выбросов) в данных;

развитие графовых моделей сложных систем, в том числе разработка и адаптация существующих алгоритмов построения графов взаимодействия параметров систем [29]; адаптация существующих алгоритмов прогнозирования временных рядов и оценки точности прогноза в случае данных высокой размерности; адаптация существующих алгоритмов кластеризации высокоразмерных данных к условию, когда размерность данных превышает число наблюдений.

В результате проведенных работ автором были предложены подходы к обработке временных рядов в режиме реального времени [12, 33, 34, 41–43], выбору модели в задаче обнаружения аномалий [20], обнаружению аномалий для данных, в которых присутствует привилегированная информация [23], обнаружению аномалий на фоне квазипериодических циклов [13], построению ансамблей детекторов для повышения робастности процедуры обнаружения аномалий [10], оценке параметров в процедурах несбалансированной классификации для прогнозирования редких событий [19, 47], непараметрическому обнаружению аномалий в потоках многомерных данных [24, 28, 45, 57]. Разработанные оригинальные методы позволили решить ряд важных прикладных задач, связанных с предсказательным техническим обслуживанием [1] и обнаружением аномалий: прогнозирование отказов вспомогательной силовой установки пассажирского самолета [7, 1, 17], интеллектуальный мониторинг и обнаружение аномалий в работе программно-нагруженных систем [11], обнаружение вредоносного ПО [23] и др. [22].

В авторских работах, посвященных современным методам моделирования многообразий и их использованию при решении различных прикладных задач, описаны новые методы нахождения низкоразмерной структуры высокоразмерных данных, лежащих в основе новых эффективных алгоритмов моделирования многообразий [46, 18, 36–38]. В частности, автором были разработаны и уже успешно применены различные методы прогнозирования случайных процессов, сосредоточенных на многообразии данных, обнаружения разладки в их поведении и классификации обнаруженных отклонений. Эти новые методы

являются существенной частью принципиально нового подхода к решению задачи прогнозирования поломок.

3. Применение методов обнаружений аномалий на основе суррогатных моделей для ПТО лопаточных машин

В соответствии с указом Президента РФ одним из главных направлений обеспечения национальной безопасности до 2020 года в экономической сфере на долгосрочную перспективу является энергетическая безопасность, в частности, обеспечение стабильности функционирования систем энерго- и теплоснабжения. Наиболее технологически сложной, дорогостоящей и критически важной частью систем энерго- и теплоснабжения фактически является силовая установка (ГТУ или ПТУ). Области применения лопаточных машин практически не ограничены: нефтегазодобывающая промышленность, промышленные предприятия, муниципальные образования и т.д.

Традиционные методы технического обслуживания лопаточных машин, включая плановое обслуживание и обслуживание непосредственно при поломках и авариях, оказываются недостаточно эффективными, поскольку эти методы не позволяют вести точный и своевременный контроль работоспособности отдельных узлов и всей установки в целом, а также оперативно и в некоторых случаях превентивно реагировать на возможные аварии и выходы из строя.

Таким образом, критически важным для системного развития единой энергетической системы России является необходимость разработки системы управления, включающей мониторинг и диагностику для предсказательного технического обслуживания силовых установок, реализованной на базе технологических решений российских производителей, а также разработка центра управления телеметрическими данными с различных установок, позволяющего унифицированным образом накапливать, систематизировать и анализировать большие объемы телеметрических данных.

Автоматизированная система управления, мониторинга и диагностики (СУМиД) для предсказательного технического обслуживания лопаточных машин позволит обеспечить отслеживание состояния лопаточных машин в процессе эксплуатации с целью предсказания моментов времени, в которые необходимо провести их техническое обслуживание, в том числе расположенных в удаленных и труднодоступных районах.

Автоматизированный комплекс управления, мониторинга и диагностики энергетического оборудования (СУМиД) в составе конкретной силовой установки состоит из следующих частей, см. рис. 1:

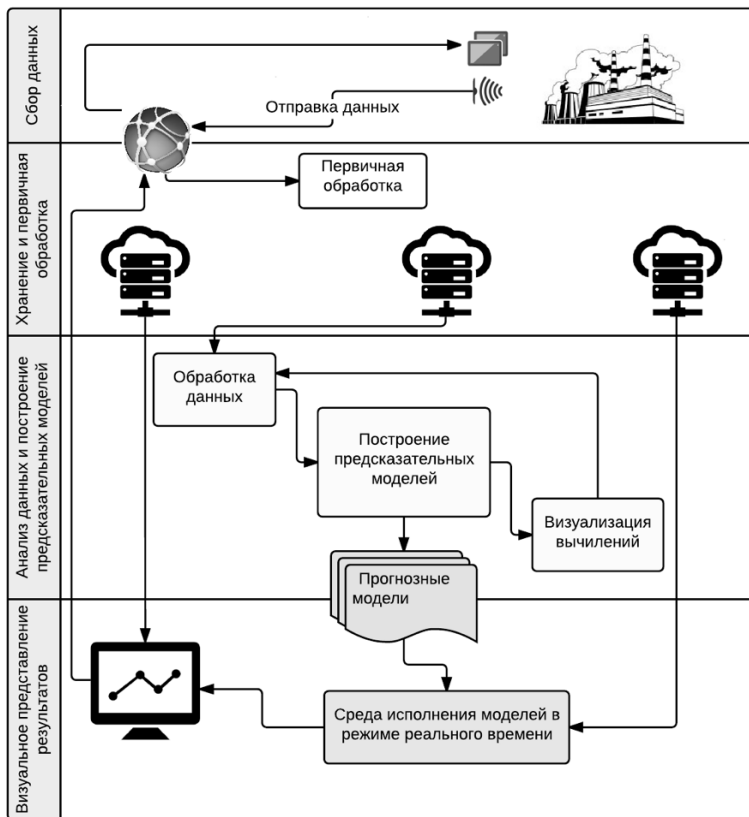


Рис. 1. Автоматизированный комплекс управления, мониторинга и диагностики

1. Автоматизированная система сбора, обработки и хранения больших объемов телеметрических данных (big data) с лопаточных машин, в том числе

- подсистема сенсоров для синхронного снятия телеметрических данных;
- подсистема приема, передачи и обработки телеметрических данных;
- база данных для хранения поступающих, в том числе и в режиме реального времени, данных, удаленного доступа к данным и манипуляций с ними;
- подсистема управления потоковыми данными, обеспечивающая первичную обработку данных и потоковую передачу данных через потоковую базу данных, сохранение в реляционную базу данных.

2. Система для анализа данных, построения предсказательных моделей и прогнозирования аномалий и разладок, в том числе

- подсистема анализа собираемых телеметрических данных, в том числе и в режиме реального времени (в зависимости от критичности информации и периодичности поступления данных);
- подсистема моделирования поведения наблюдаемых характеристик лопаточных машин на основе гибридных моделей, сочетающих в себе высокоточные вычислительно тяжелые модели, основанные на детальном описании физики процессов, и суррогатные модели на основе экспериментальных и расчетных данных. Такого рода подходы необходимы для прогнозирования моментов времени, в которые надо проводить техническое обслуживание лопаточной машины;
- база знаний из предметной области (конкретная постановка задач, минимизация числа «ложных тревог», учет дополнительных, возможно, ненаблюдаемых факторов и т.п.) и набор индикаторов работы различных частей лопаточных машин, в том числе в режиме онлайн;

- подсистема для симуляции на исторических данных и визуализации.

3. Система исполнения предсказательных моделей на потоковых данных, визуализации показателей работы лопаточной машины и результатов исполнения предсказательных моделей, отображение предупреждений и тревог, в том числе

- подсистема для взаимодействия (в том числе и обмена данными) ситуационного центра с производителем лопаточных машин и инженерным отделом генерирующей компании для проведения консультаций в режиме реального времени;
- подсистема для обеспечения инженерного отдела генерирующей компании необходимой информацией о техническом состоянии лопаточной машины, а также результатами анализа данных сенсоров, проведенного ситуационным центром;
- подсистема для наблюдения за состоянием силовой установки операторами (в режиме урезанной функциональности через веб-интерфейс).

4. Ситуационный центр управления телеметрическими данными с различных лопаточных машин, позволяющий унифицированным образом накапливать, систематизировать и анализировать большие объемы телеметрических данных, представлять агрегированные результаты анализа на едином информационном табло, проводить совместные совещания.

При управлении, мониторинге и диагностике энергетического оборудования важным является использование исторических данных для получения информации о протекании технологических процессов, а также для построения функциональных связей и генерации статистических моделей зависимости ключевых показателей эффективности (KPI) процессов от параметров технологического режима работы для того или иного объекта. Математически данная задача формулируется как задача моделирования многообразий и построения суррогатных моделей зависимостей различных групп параметров.

На основе построенных моделей возможно проводить симуляции различных режимов работы объектов, не экспериментируя на реальном оборудовании. Помимо этого становится возможным решать широкий класс задач по усовершенствованию технологических процессов, среди которых:

- диагностика и поиск причин отклонений параметров процесса от желаемых значений;
- мониторинг состояния оборудования;
- построение виртуальных датчиков;
- задачи стабилизации и оптимизации целевых показателей эффективности.

Опишем основную математическую идею метода обнаружения аномалий на основе суррогатных моделей для целей ПТО:

- Обозначим через $\{x_i\}_{t \geq 1}$ d -мерный наблюдаемый сигнал телеметрии;
- Для каждого $i = 1, \dots, d$ строится суррогатная модель зависимости $x_{i,t}$ от значений остальных компонент сигнала $\{x_{1,s}, \dots, x_{i-1,s}, x_{i+1,s}, \dots, x_{d,s}\}$ для $s = t - L_i, \dots, t$ как в текущий, так и в прошлые моменты времени, на основе исторических данных, соответствующих нормальному режиму функционирования системы;
- Определим порог для детектирования аномалий h как такой порог, что в нормальном состоянии вероятность превышения порога значением ошибки прогноза мала. Для моделирования указанной вероятности используется модель на основе гауссовских процессов;
- Сигнал тревоги о появлении аномалии объявляется, если для текущих наблюдений ошибка прогноза превышает пороговое значение h .

Рассмотрим применение указанной методологии на примере решения задачи обнаружения аномалий в работе парогазовой установки (ПГУ). Исторические данные были представлены в виде временных рядов 26 измеряемых параметров за период с 1 января 2012 года по 31 декабря 2016 года с поминутным разрешением. Задача состояла в том, чтобы построить индикатор

тор деградации системы и выявить моменты начала периодов, когда начиналась деградация системы.

На рис. 2 приведен пример зависимости значений измеряемых параметров от времени. Видно, что поведение временных рядов «неоднородным» образом зависит от времени, данные сильно зашумлены. Более того, фактически нет размеченных событий поломок, т.е. невозможно использование методов обучения с учителем.

При построении индикатора деградации системы мы использовали следующие предположения, см. рис. 3:

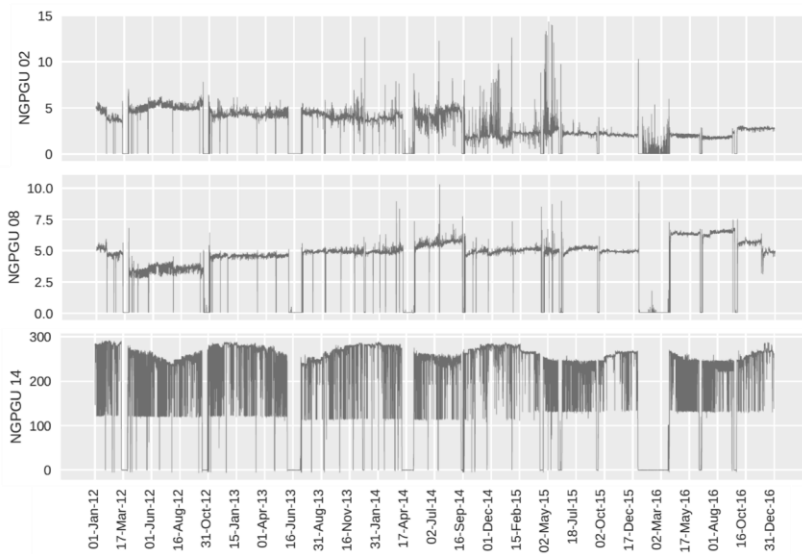


Рис. 2. Пример зависимости значений измеряемых параметров от времени

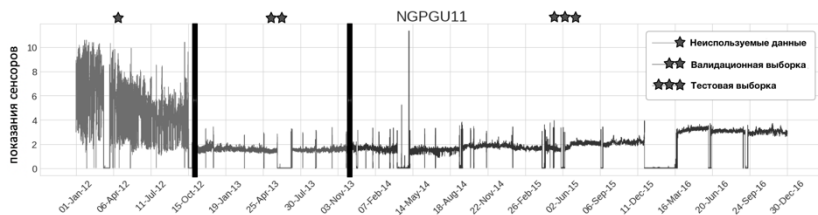


Рис. 3. Различные периоды в работе ПГУ

- Данные, соответствующие периоду Nov-12 – Nov-13, рассматривались как нормальные, так как в это время не происходило каких-либо значимых событий технического обслуживания ПГУ или поломок (на рис. 3 интервалы с этими наблюдениями отмечены одной и двумя звездочками соответственно);

- С начала регистрации наблюдений и до момента Oct-2012 временные ряды имели экстремально высокие вариации, что потенциально соответствует либо ненормальному режиму работы ПГУ, либо шуму в данных. По этой причине указанные наблюдения не использовались для обучения модели.

Для построения индикатора деградации системы были реализованы следующие шаги:

1. Построить суррогатные модели зависимости различных наблюдаемых параметров друг от друга, как это было описано выше. Таким образом, по сути, мы учитываем, что наблюдаемые значения многомерного временного ряда описывают на некотором многообразии нелинейную траекторию, см. рис. 4 и раздел 2.

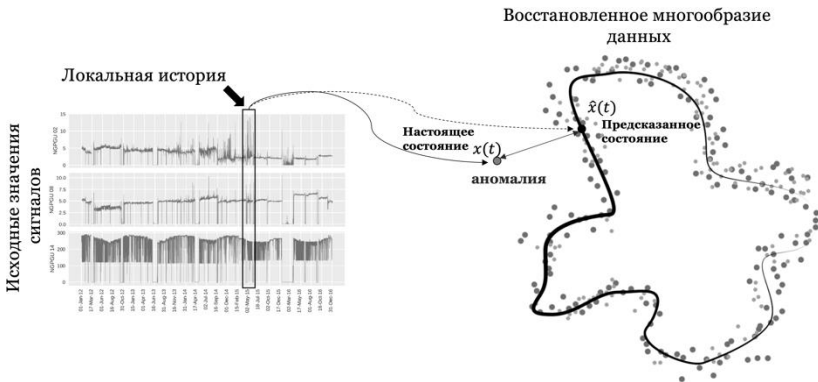


Рис. 4. Схематическая иллюстрация подхода к обнаружению аномалий на основе моделирования многообразия данных

2. Спрогнозировать значения временных рядов на тестовом множестве, для которого надо выявить наличие аномальных режимов системы.

3. Подсчитать ошибки прогнозирования.

4. Подсчитать ошибки прогнозирования на обучающей выборке, используя подход на основе кросс-проверки.

5. Сравнить ошибки на обучающей и тестовой выборках на основе индикаторов аномальности. В качестве таковых мы использовали

- а) индикатор превышения текущим значением ошибки границы, задаваемой максимальной наблюдаемой ранее (в нормальном режиме) ошибкой;
- б) p -значение гипотезы о превышении текущим значением ошибки границы, задаваемой максимальной наблюдаемой ранее (в нормальном режиме) ошибкой.

На рис. 5 и 6 представлены зависимости построенных индикаторов деградации системы от времени.

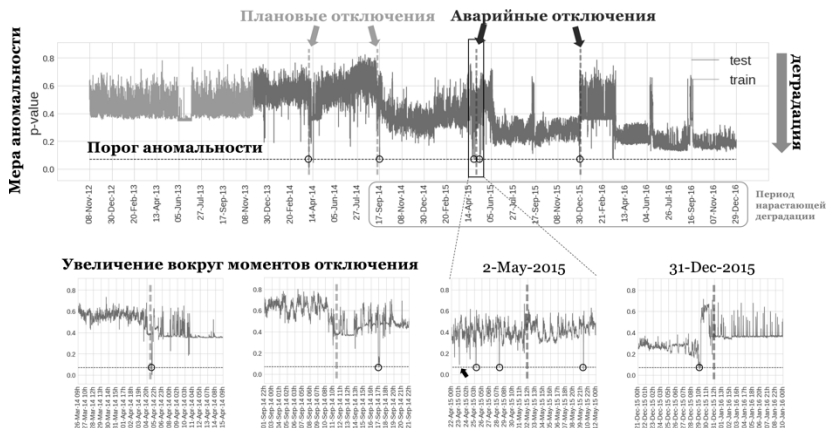


Рис. 5. Индикатор деградации, задаваемый p -значением гипотезы о превышении текущим значением ошибки границы, определяемой максимальной наблюдаемой ранее (в нормальном режиме) ошибкой

Таким образом, исходя из полученных результатов можно сделать следующие выводы:

1. Наши показатели свидетельствуют о постепенной деградации системы, начиная с середины 2014 года.

2. В ходе «слепого теста» (на данных, неизвестных до момента проведения тестирования полученной модели) было обнаружено несколько аномалий (исключительно на основе анализа исходных данных, т.е. отсутствовала какая-либо исходная информация о фактических датах остановок/критических событий в ходе работы ПГУ):

- найденные аномалии соответствуют плановым и аварийным отключениям;
- аномалии, соответствующие аварийным остановкам ПГУ, были детектированы существенно раньше моментов времени, когда возникли аварийные ситуации:
 - а) первая аномалия, обнаруженная 29 декабря 2015 года, соответствуют аварийной остановке из-за критического события 31 декабря 2015 года;
 - б) вторая аномалия, обнаруженная в период с 23 по 28 апреля 2015 года, соответствует аварийной остановке 2 мая 2015 года.

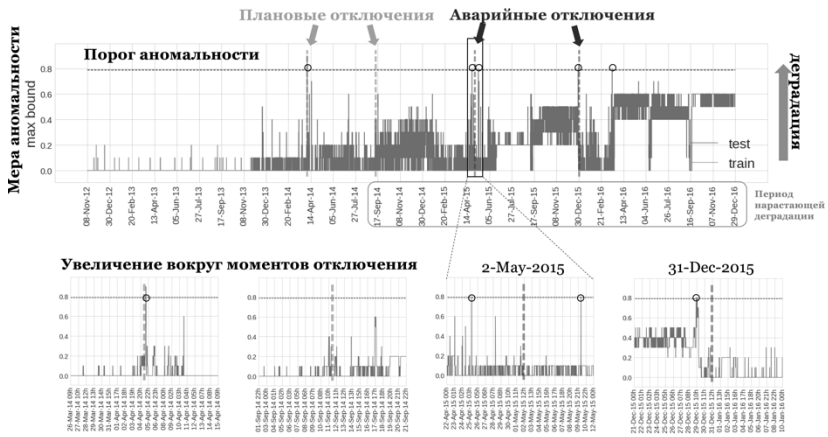


Рис. 6. Индикатор деградации, основанный на сравнении текущего (нормированного) значения ошибки с границей, определяемой максимальной наблюдаемой ранее (в нормальном режиме) ошибкой

4. Заключение

В данной работе была рассмотрена задача построения предсказательных моделей (суррогатных моделей) для решения задач индустриальной предиктивной аналитики. Автор, проанализировав потребности индустриальных приложений, сформулировал ряд новых математических и алгоритмических задач и разработал соответствующие методы моделирования по данным. Применение этих методов позволило эффективно решить практически важные задачи.

В рамках проведенных научных исследований автором был получен ряд новых результатов. В частности, был предложен целый ряд ключевых концепций в разработанных методах предсказательной аналитики: подходы к иерархической байесовской регуляризации в регрессионных моделях, в том числе на основе гауссовских процессов, и методы адаптивного планирования экспериментов; непараметрические методы обнаружения аномалий и прогнозирования редких событий; вычислительно эффективные алгоритмы построения регрессии на основе гауссовских процессов для данных со специальной структурой; непараметрические методы оценки точности прогноза суррогатной модели и др.

Проведенные исследования в значительной степени были мотивированы потребностями задач индустриального проектирования, а также задачами обработки биомедицинских и экономических данных. Это позволило использовать разработанные математические методы, реализованные компанией ООО «ДАТАДВАНС» (см. [23]) в виде алгоритмического ядра программной системы pSeven Core (первое название программной системы – MACROS library) суррогатного моделирования и инженерной оптимизации для решения целого ряда прикладных инженерных задач. Разработанное алгоритмическое ядро получило сертификацию на финальном уровне Technology Readiness Level (TRL, см. [54]). Согласно официальному пресс-релизу компании Airbus, разработанные методы позволили сэкономить до 10% в некоторых важных этапах полного цикла проектирования новых моделей самолетов (см. [4–6]).

Использование разработанных алгоритмов и их дальнейшее развитие позволило успешно реализовать ряд инженерных приложений в авиакосмической промышленности [3], а также применить их для решения задач предсказательной аналитики в других прикладных областях, например, таких как

- моделирование прочностных характеристик элементов обшивки самолета А350-900 (компания Airbus), см. [27, 50];
- моделирование аэродинамических характеристик суборбитального космического летательного аппарата (компания Astrium, см. [9]), вертолета (компания Airbus Helicopters, см. [51]);
- Снижение массы кокпита болида одной из команд «Формула-1»;
- прогнозирование концентрации воздуха в разных точках трубопровода в зависимости от внешних условий и геометрии (компания Airbus);
- Прогнозирование отказов вспомогательной силовой установки пассажирского самолета (компания Airbus), см. [7, 8, 1, 17];
- метамоделирование в задаче нестационарной трехфазной фильтрации в пласте со скважиной сложной геометрии (компания Газпром-НТЦ), см. [55, 56]; также был решен ряд других задач в нефтегазовой отрасли, см. [30–32, 39, 44, 52, 53];
- интеллектуальный мониторинг и обнаружение аномалий в работе программнонагруженных систем (компания ООО «Яндекс»), см. [11, 13], в том числе и для обнаружения вредоносного ПО, см. [23];
- обработка информации с автоматических дорожных метеорологических станций и прогнозирование состояния дорожного покрытий автомагистралей (компания АО «Минимакс-94»), см. [48, 49];
- Анализ данных сортировки и секвенирования нового поколения для исследования влияния 5'-нетранслируемого участка на эффективность трансляции у *Escherichia coli* (биомедицина), см. [26]; моделирование геофизических процессов Земли, см. [40], и др.

Е. Бурнаев выражает благодарность компании ООО «Датадванс» за предоставление данных и постановки задачи в рамках договоров с ведущими индустриальными компаниями.

Литература

1. БУРНАЕВ Е. *Предсказательное обслуживание* // В кн. «Перспективные рынки и технологии интернета вещей: публичный аналитический доклад» / Под ред. И.Г. Дежиной]. – М.: ООО «Лайм», 2019. – 272 с.
2. БУРНАЕВ Е., ГУБАРЕВ Ф., МОРОЗОВ С., ПРОХОРОВ А., ХОМИНИЧ Д. *Автоматизация инженерных расчетов, анализ данных и оптимизация с помощью программного комплекса PSE/MACROS* // Межотраслевая информационная служба. – 2013. – №4. – С. 41–50.
3. БУРНАЕВ Е., БЕРНШТЕЙН А. *Методы анализа данных, предсказательного моделирования и обслуживания* // В кн. «Перспективные технологии для авиационной промышленности: Аналитический обзор». – М.: Наука, 2017. – 463 с.
4. *Airbus achieves multi-objective optimization of its aircraft families with datadance's "macros" software*. [Электронный ресурс]. – URL: <https://www.airbus.com/newsroom/press-releases/en/2013/08/airbus-achieves-multi-objective-optimization-of-its-aircraft-families-with-datadance-s-macros-software.html> (дата обращения: 07.02.2020).
5. *Airbus to reduce lead times in numerical analysis activities for aircraft design*. [Электронный ресурс]. – URL: <https://www.airbus.com/newsroom/press-releases/en/2013/06/airbus-to-reduce-lead-times-in-numerical-analysis-activities-for-aircraft-design.html> (дата обращения: 07.02.2020).
6. *Airbus uses datadance's macros software for structural analysis of a350 xwb aircraft*. [Электронный ресурс]. – URL: <https://www.aerospace-technology.com/news/newsairbus-uses-datadance-s-macros-software-structural-analysis-a350-xwb-aircraft-4362144/> (дата обращения: 07.02.2020).

7. ALESTRA S., BORDRY C., BRAND C., BURNAEV E., EROFEEV P., PAPANOV A., SILVEIRA-FREIXO C. *Application of rare event anticipation techniques to aircraft health management* // Advanced Materials Research. – 2014. – Vol. 1016. – P. 413–417.
8. ALESTRA S., BORDRY C., BRAND C., BURNAEV E., EROFEEV P., PAPANOV A., SILVEIRA-FREIXO C. *Rare event anticipation and degradation trending for aircraft predictive maintenance* // Proc. of the joint WCCM-ECCM-ECFD 2014 Congress, 20-25 July, Barcelona, Spain. – 2014.
9. ALESTRA S., KAPUSHEV E., BELYAEV M., BURNAEV E., DORMIEUX M., CAVAILLES A., CHAILLOT D., FERREIRA E. *Surrogate models for spacecraft aerodynamic problems* // Proc. of the joint WCCM-ECCM-ECFD 2014 Congress, 20-25 July, Barcelona, Spain. – 2014.
10. ARTEMOV A., BURNAEV E. *Ensembles of detectors for online detection of transient changes*// Proc. SPIE. – 2015. – Vol. 9875. – P. 9875–9875–5.
11. ARTEMOV A., BURNAEV E. *Detecting performance degradation of software-intensive systems in the presence of trends and long-range dependence* // IEEE 16th Int. Conference on Data Mining Workshops (ICDMW-2016). – 2016. – P. 29–36.
12. ARTEMOV A., BURNAEV E. *Optimal estimation of a signal perturbed by a fractional brownian noise* // Theory of Probability and Its Applications. – 2016. – Vol. 60, No. 1. – P. 126–134.
13. ARTEMOV A., BURNAEV E., LOKOT A. *Nonparametric decomposition of quasi-periodic time series for change-point detection* // Proc. SPIE. – 2015. – Vol. 9875. – P. 9875–9875–5.
14. BERNSTEIN A., KULESHOV A. *Manifold learning: Generalization ability and tangent proximity* // Int. Journal of Software and Informatics. – 2013. – Vol. 7, No. 3.
15. BURNAEV E. *Disorder problem for poisson process in generalized bayesian setting* // Theory of Probability and Its Applications. – 2009. – Vol. 53, No. 3. – P. 500–518.
16. BURNAEV E. *Nonlinear minimax problem of the quickest detection of disorder for brownian motion* // Proc. of the Russian-Japanese Symposium on “A Stochastic Analysis of Advanced

- Statistical Models”. Theory Probab. Appl. / Eds.: A. Shiryaev, V. Ulyanov. – 2010. – Vol. 55. – P. 536–557.
17. BURNAEV E. *Rare failure prediction via event matching for aerospace applications* // Proc. of the 3rd Int. Conference on Circuits, System and Simulation (ICCSS-2019). – 2019. – P. 214–220.
 18. BURNAEV E., CHERNOVA S. *On an iterative algorithm for calculating weighted principal components* // Journal of Communications Technology and Electronics. – 2015. – Vol. 60, No. 6. – P. 619–624.
 19. BURNAEV E., EROFEEV P., PAPANOV A. *Influence of resampling on accuracy of imbalanced classification* // Proc. SPIE. – 2015. – Vol. 9875. – P. 9875–9875–5.
 20. BURNAEV E., EROFEEV P., SMOLYAKOV D. *Model selection for anomaly detection* // Proc. SPIE. – 2015. – Vol. 9875. – P. 9875–9875–6.
 21. BURNAEV E., FEINBERG E., SHIRYAEV A. *On asymptotic optimality of the second order in the minimax quickest detection problem of drift change for brownian motion* // Theory of Probability and Its Applications. – 2009. – Vol. 53, No. 3. – P. 519–536.
 22. BURNAEV E., KOPTILOV I., NOVIKOV G., KHANIPOV T. *Automatic construction of a recurrent neural network-based classifier for vehicle passage detection* // Proc. SPIE. – 2017. – Vol. 10341. – P. 10341–10341–6.
 23. BURNAEV E., SMOLYAKOV D. *One-class svm with privileged information and its application to malware detection* // IEEE 16th Int. Conference on Data Mining Workshops (ICDMW–2016). – 2016. – P. 273–280.
 24. BURNAEV E., GOLUBEV G. *On one problem in multichannel signal detection* // Problems of Information Transmission. – 2017. – Vol. 53, No. 4. – P. 368–380.
 25. BURNAEV E. *Disorder problem for a poisson process in the generalized bayesian setting* // Russian Mathematical Surveys. – 2007. – Vol. 62, No. 4. – P. 790.
 26. EVFRATOV S., OSTERMAN I., KOMAROVA E., POGORELSKAYA A., RUBTSOVA M., ZATSEPIN T., SEMASHKO T., KOSTRYUKOVA E.S., MIRONOV A., BUR-

- NAEV E., KRYMOVA E., GELFAND M., GOVORUN V., BOGDANOV A., SERGIEV P., DONTSOVA O. *Application of sorting and next generation sequencing to study 5'-utr influence on translation efficiency in escherichia coli* // Nucleic Acids Research. – 2017. – Vol. 45, No. 6. – P. 3487–3502.
27. GRIHON S., BURNAEV E., BELYAEV M., PRIKHODKO P. *Surrogate Modeling of Stability Constraints for Optimization of Composite Structures*. – Springer New York, New York, 2013. – P. 359–391.
28. ISHIMTSEV V., BERNSTEIN A., BURNAEV E., NAZAROV I. *Conformal k-nn anomaly detector for univariate data streams* // Proc. of the Sixth Workshop on Conformal and Probabilistic Prediction and Applications. Proceedings of Machine Learning Research. – 2017. – Vol. 60. – P. 213–227.
29. IVANOV S., BURNAEV E. *Anonymous walk embeddings* // Proc. of the 35th Int. Conference on Machine Learning. Proceedings of Machine Learning Research. – 2018. – Vol. 80. – P. 2186–2195.
30. KANIN E., VAINSHTEIN A., OSIPSOV A., BURNAEV E. *The method of calculation the pressure gradient in multiphase flow in the pipe segment based on the machine learning algorithms* // IOP Conference Series: Earth and Environmental Science. – 2018. – Vol. 193.
31. KANIN E., OSIPTSOV A., VAINSHTEIN A., BURNAEV E. *A predictive model for steady-state multiphase pipe flow: Machine learning on lab data* // Journal of Petroleum Science and Engineering. – 2019. – Vol. 180. – P. 727–746.
32. KLYUCHNIKOV N., ZAYTSEV A., GRUZDEV A., OVCHINNIKOV G., ANTIPOVA K., ISMAILOVA L., MURAVLEVA E., BURNAEV E., SEMENIKHIN A., CHEREPANOV A., KORYABKIN V., SIMON I., TSURGAN A., KRASNOV F., KOROTEEV D. *Data-driven model for the identification of the rock type at a drilling bit* // Journal of Petroleum Science and Engineering. – 2019. – Vol. 178. – P. 506 – 516.
33. KOROTIN A., V'YUGIN V., BURNAEV E. *Aggregating strategies for long-term forecasting* // Proc. of the Seventh Workshop on Conformal and Probabilistic Prediction and Applica-

- tions. Proceedings of Machine Learning Research. – 2018. – Vol. 91. – P. 63–82.
34. KOROTIN A., VYUGIN V., BURNAEV E. *Adaptive hedging under delayed feedback*. – 2019. Neurocomputing.
35. KOUL A., SHANKAR S., WALLACE W. *Predictive Maintenance – A new approach to lifecycle management of turbines*. – Life Prediction Technologies Inc, 2014.
36. KULESHOV A., BERNSTEIN A., BURNAEV E. *Conformal prediction in manifold learning* // Proc. of the Seventh Workshop on Conformal and Probabilistic Prediction and Applications. Proceedings of Machine Learning Research. – 2018. – Vol. 91. – P. 234–253.
37. KULESHOV A., BERNSTEIN A., BURNAEV E. *Kernel regression on manifold valued data* // Proc. of IEEE 5th Int. Conference on Data Science and Advanced Analytics (DSAA-2018). – 2018. – P. 120–129.
38. KULESHOV A., BERNSTEIN A., BURNAEV E. *Manifold learning regression with non-stationary kernels* // Proc. Of Artificial Neural Networks in Pattern Recognition. – Springer International Publishing, 2018. – P. 152–164.
39. MAKHOTIN I., KOROTEEV D., BURNAEV E. *Gradient boosting to boost the efficiency of hydraulic fracturing* // Journal of Petroleum Exploration and Production Technology. – 2019.
40. PROSKURA P., ZAYTSEV A., BRASLAVSKY I., EGOROV E., BURNAEV E. *Usage of multiple rtl features for earthquakes prediction* // Proc. of Computational Science and Its Applications – ICCSA. – Springer International Publishing, 2019. – P. 556–565.
41. RIVERA R., NAZAROV I., BURNAEV E. *Towards forecast techniques for business analysts of large commercial data sets using matrix factorization methods* // Journal of Physics: Conference Series. – 2018. – Vol. 1117, No. 1.
42. RIVERA R., PLETNEV A., PILYUGINA P., NAZAROV I., DIAZ G., ZHU W., BURNAEV E. *Topology based clusterwise regression for user segmentation and demand forecasting* //

- Proc. of IEEE 5th Int. Conference on Data Science and Advanced Analytics (DSAA-2019). – 2019.
43. RIVERA-CASTRO R., NAZAROV I., XIANG YU., PLETNEEV A., MAKSIMOV I., BURNAEV E. *Demand forecasting techniques for build-to-order lean manufacturing supply chains* // Proc. of Advances in Neural Networks – ISNN. – Springer International Publishing, 2019. – P. 213–222.
 44. ROMANENKOVA E., ZAYTSEV A., KLYUCHNIKOV N., GRUZDEV A., ANTIPOVA K., ISMAILOVA L., BURNAEV E., SEMENIKHIN A., KORYABKIN V., SIMON I., KOROTEEV D. *Real-time data-driven detection of the rock type alteration during a directional drilling* // arXiv e-prints arXiv:1903.11436. – 2019.
 45. SAFIN A., BURNAEV E. *Conformal kernel expected similarity for anomaly detection in time-series data* // Advances in Systems Science and Applications. – 2017. – Vol. 17, No. 3. – P. 22–33.
 46. SHAHRUKH A., BURNAEV E., LEMPISKIY V. *Latent convolutional models* // Proc. Int. Conference on Learning Representations. – 2019.
 47. SMOLYAKOV D., KOROTIN A., EROFEEV P., PAPANOV A., BURNAEV E. *Meta-learning for resampling recommendation systems* // Proc. SPIE 11041, Eleventh Int. Conference on Machine Vision (ICMV-2018). – 2019. – Vol. 110411S.
 48. SMOLYAKOV D., SVIRIDENKO N., BURIKOV E., BURNAEV E. *Anomaly pattern recognition with privileged information for sensor fault detection* // Proc. of Artificial Neural Networks in Pattern Recognition. – Springer International Publishing, 2018. – P. 320–332.
 49. SMOLYAKOV D., SVIRIDENKO N., ISHIMTSEV V., BURIKOV E., BURNAEV E. *Learning ensembles of anomaly detectors on synthetic data* // Proc. of Advances in Neural Networks – Springer International Publishing, 2019. – P. 292–306.
 50. STERLING G., PRIKHODKO P., BURNAEV E., BELYAEV M., GRIHON S. *On approximation of reserve factors dependency on loads for composite stiffened panels* // Advanced Materials Research. – 2014. – Vol. 1016. – P. 85–89.

51. STRUZIK A., BURNAEV E., PRIKHODKO P. *Surrogate models for helicopter loads problems* // Proc. of the 5th European Conference for Aeronautics and Space Sciences. Germany, Munich. – 2013.
52. SUDAKOV O., BURNAEV E., KOROTEEV D. *Driving digital rock towards machine learning: Predicting permeability with gradient boosting and deep neural networks* // Computers and Geosciences. – 2019. – Vol. 127. – P. 91–98.
53. SUDAKOV O., KOROTEEV D., BELOZEROV B., BURNAEV E. *Artificial neural network surrogate modeling of oil reservoir: A case study* // Proc. of Advances in Neural Networks. – Springer International Publishing, 2019. – P. 232–241.
54. *Technology readiness level*. Электронный ресурс]. – URL: https://en.wikipedia.org/wiki/Technology_readiness_level (дата обращения: 07.02.2020).
55. TEMIRCHEV P., KOSTOEV R., KOROTEEV D., SIMONOV M., BURNAEV E., OSELEDETS I., AKHMETOV A. *Application of machine learning technologies for rapid 3d modelling of inflow to the well in the development system* // Proc. of SPE Russian Petroleum Technology Conference. – 2018.
56. TEMIRCHEV P., SIMONOV M., KOSTOEV R., BURNAEV E., OSELEDETS I., AKHMETOV A., MARGARIT A., SITNIKOV A., KOROTEEV D. *Deep Neural Networks Predicting Oil Movement in a Development Unit* // arXiv e-prints arXiv:1901.02549. – 2019.
57. VOLKHONSKIY D., BURNAEV E., NOURETDINOV I., GAMMERMAN A., VOVK V. *Inductive conformal martingales for change-point detection* // Proc. of the Sixth Workshop on Conformal and Probabilistic Prediction and Applications. Proceedings of Machine Learning Research. – 2017. – Vol. 60. – P. 132–153.

ANOMALY DETECTION BASED ON SURROGATE MODELS

Evgeny Burnaev, Skolkovo Institute of Science and Technology, Moscow, Cand.Sc., assistant professor (e.burnaev@skoltech.ru).

Abstract: Predictive maintenance methods are used to detect as soon as possible significant changes (disorders) in the operation of mechanisms. The main purpose of this approach to maintenance is to continuously monitor and correct the technical condition or completely replace the mechanisms before the detected changes become critical for the operation of individual components or the system as a whole. The diagnostic capabilities of predictive maintenance methods have significantly increased in recent years due to the improvement of sensory observation technologies and the development of new information processing algorithms. The use of predictive maintenance has a number of advantages in comparison with other practiced approaches to maintenance, namely, the ability to conduct more accurate and timely monitoring of the health of individual parts and the entire system as a whole; continuous monitoring and analysis of internal and external conditions improves the safety of operation and allows for a more rapid and in some cases preventive response to possible accidents and failures; significant reduction in maintenance costs, due to the exclusion of planned replacements of technically sound and reliable system components. In this paper, we consider the problem of constructing predictive models (surrogate models) to solve the predictive maintenance problem. The special features of the problems of detecting anomalies and predicting failures are analyzed. An overview of the main needs of industrial applications and a description of the structure of the corresponding predictive maintenance systems is provided. An example of using methods for detecting anomalies based on surrogate models for predictive maintenance of shovel machines is given.

Keywords: predictive analytics, industrial engineering, machine learning, predictive maintenance, failure prediction.

УДК 519.2 + 519.8

ББК 22.18

DOI: 10.25728/ubs.2020.86.1

Статья представлена к публикации членом редакционной коллегии А.И. Михальским.

Поступила в редакцию 28.05.2020.

Опубликована 31.07.2020.