

О построении систем раннего оповещения для диагностического технического обслуживания в аэрокосмической промышленности

Е.В. Бурнаев

Сколковский институт науки и технологий, Москва, Россия
Институт проблем передачи информации, Российская академия наук, Москва, Россия
e-mail: e.burnaev@skoltech.ru

Поступила в редколлегию 01.07.2019

Аннотация—Рассматривается задача построения предсказательных моделей для систем раннего оповещения в целях диагностического технического обслуживания в аэрокосмической промышленности. Предложен новый подход к прогнозированию редких поломок, основанный на новой методологии, которая позволяет учесть свойства технических систем и специфические требования, накладываемые приложениями.

Ключевые слова: система мониторинга, обнаружение аномалий, предсказание поломок, машинное обучение, предсказательное техническое обслуживание, обработка сигналов

1. ВВЕДЕНИЕ

Авиационная промышленность является одной из наиболее регулируемых отраслей — качество, безопасность и эффективность работы авиационных систем оказывает непосредственное влияние на жизнь и здоровье людей.

В последние годы крупнейшие производители самолетов и авиакомпании пришли к выводу, что за счет использования методов предсказательной аналитики и инженерного знания данные, собираемые системами телеметрии самолета во время полета, а также другие данные о жизненном цикле самолета, могут быть использованы для обслуживания оборудования на основе автоматического контроля его состояния (т.н. предсказательное техническое обслуживание, ПТО). ПТО играет все большую роль в обеспечении надежности и эффективности — методы ПТО предоставляют возможности для предсказания и предотвращения возможных отказов, уменьшения стоимости обслуживания и увеличения коэффициента использования самолетного парка.

Сервисы ПТО могут оценивать текущее состояние оборудования, определять его операционный статус, своевременно обнаруживать аномальные условия, предотвращать возможный вынужденный простой [1–3, 20, 38, 43]. Сложности, возникающие при построении моделей для систем ПТО в авиации, состоят в том, что

- данные телеметрии самолетов имеют очень сложную структуру: временные ряды с большим числом показателей (часто больше нескольких сотен) и высокой частотой измерений (часто порядка десятков тысяч измерений за полет); большие объемы данных (типичный размер выборки данных для обучения измеряется в терабайтах); пропущенные измерения, нестационарный шум; сложная иерархическая структура типов поломок, которые необходимо прогнозировать;
- естественно, что поломки — редкие события, по этой причине задача прогнозирования сильно несбалансирована [13, 39], то есть типичные выборки данных содержат много примеров нормального функционирования системы и только небольшое число примеров системы, находящейся в аномальном состоянии.

В [33, 42] авторы рассмотрели модели типа авторегрессии для прогнозирования временных рядов телеметрии с последующей коррекцией прогноза на основе ядерной регрессии и применили полученную модель для обнаружения аномалий в показаниях датчиков. В [8, 10] рассматривается задача ПТО для прогнозирования времени замены выпускного клапана системы управления вентиляции в самолете. Для построения моделей используются как признаки, генерируемые по данным о тех или иных типах событий (например, поломки определенных типов) в определенные моменты времени, так и на основе авторегрессионных моделей показаний телеметрии. В [9] авторы рассмотрели подход к прогнозированию времени жизни критических компонент системы отбора воздуха самолета на основе линейной машины опорных векторов. В [44] решалась задача обнаружения поломок в некотором химическом процессе на основе стандартных методов моделирования временных рядов. Аналогичный подход на основе векторной линейной авторегрессии использовался и в [22]. В [21] разрабатывались подходы к онлайн анализу телеметрии с подшипников оси поезда на основе ядерной регрессии.

Фактически, в работах, перечисленных выше, для решения задачи ПТО конкретной технической системы используются следующие подходы:

- задача прогнозирования поломок фактически “заменяется” задачей обнаружения аномалий в потоке данных телеметрии; в свою очередь, задача обнаружения аномалий решается каким-либо стандартным методом, зачастую без учета зависимостей между наблюдаемыми сигналами; в том числе часто для обнаружения аномалий строят линейные модели прогноза будущих значений временных рядов, после чего сравнивая факт и прогноз принимают решение о наличии/отсутствии аномалии;
- в случае, если все же имеется “разметка”, то есть в исторической выборке присутствует информация о том, в какие моменты времени поломки какого типа происходили, авторы обычно используют стандартный в машинном обучении подход — генерируют входные признаки по данным телеметрии и далее применяют методы классификации типа случайных лесов.

Отметим, что оба указанных выше “общих” подхода в целом не применимы для построения систем раннего оповещения для диагностического технического обслу-

живания в аэрокосмической промышленности. Действительно, число примеров поломок обычно очень мало, то есть задача классификации для обучения с учителем будет крайне сильно несбалансирована. В тоже время переформулировка такой задачи как задачи обнаружения аномалий также не работает — соответствующая предсказательная модель должна иметь низкий уровень ложных тревог и при этом быть способной обнаружить достаточное для использования на практике количество поломок; стандартные же методы обнаружения аномалий, да еще и примененные к зашумленным высокоразмерным данным телеметрии, будут порождать большое количество ложных тревог.

Таким образом, разработка автоматизированной системы для предсказания поломок и раннего оповещения о возможных дорогостоящих неисправностях — очень сложная задача, и требуется разработка специализированной методологии, которая позволит учесть перечисленные выше особенности обрабатываемых данных, порождаемых техническими системами в авиации.

В данной работе рассматривается задача построения предсказательных моделей для систем раннего оповещения в целях диагностического технического обслуживания в аэрокосмической промышленности. Автором предложен новый подход к прогнозированию редких поломок, основанный на новой методологии, которая позволяет учесть свойства технических систем и специфические требования, накладываемые приложениями. Применение соответствующих алгоритмов проиллюстрировано на задаче прогнозирования определенных типов поломок двигателя пассажирского самолета.

2. ОБЩАЯ МЕТОДОЛОГИЯ

Имеет место следующее очевидное наблюдение: оборудование (скажем, двигатель) 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, появляющиеся дефекты изменяют статистические свойства данных телеметрии, которые можно выявлять методами обнаружения аномалий.

Таким образом, исходя из этого простого наблюдения, мы предлагаем следующий общий подход к построению моделей раннего оповещения для диагностического технического обслуживания, состоящий из последовательных этапов анализа данных телеметрии технической системы:

1. Идентификация подсистем: по накопленным данным телеметрии с датчиков все измеряемые параметры (признаки) разделяются на группы, соответствующие либо различным подсистемам, либо таким кластерам, для которых между параметрами имеются значительные зависимости, при этом между параметрами из разных кластеров зависимости не значимы;
2. Обнаружение аномалий: в каждой выделенной группе зависимых параметров или идентифицированной подсистеме используются методы обнаружения аномалий

- для выявления либо структурных изменений в моделях зависимости, либо простых экстремальных значений, либо любых других отклонений (см. рис. 1);
3. Формирование предупреждающих сигналов: с помощью правил булевой логики отдельные простые двоичные ряды аномалий, полученных на предыдущем шаге, объединяются в предупреждающие сигналы о возможной неисправности;
 4. Сопоставление событий: каждый предупреждающий сигнал тестируется на предсказательную способность с помощью статистических методов, разработанных для определения сигналов с наибольшей предсказательной силой, которые предшествуют поломке или пред-аварийным событиям и, как правило, не возникают больше нигде, кроме как за короткое время до таких событий (см. раздел 3.2);
 5. Формирование сигналов тревоги: предупреждающие сигналы с достаточно хорошей предсказательной способностью комбинируются в сигнал тревоги, которая срабатывает, когда активна хотя бы одна его составляющая;
 6. Валидация по отдельным объектам: вся процедура построения сигнала тревоги выполняется на наборе данных со всеми объектами, кроме одного, а затем проверяется на этом исключенном объекте для оценки производительности по ключевым метрикам раннего предупреждения (см. раздел 3.3).

В следующих разделах для описанных выше этапов приведены детальные описания.

2.1. Идентификация подсистем

В общем случае пространство признаков может иметь высокую размерность, а также разработчику моделей раннего оповещения для предсказательного обслуживания может быть доступно не очень много информации о структуре технической системы. Чтобы преодолеть недостаток знаний в предметной области, можно применить процедуру кластеризации и разделить пространство признаков на значимо зависимые группы. Такой подход состоит из трех основных этапов:

1. Расчет взаимозависимости между парами признаков;
2. Отсечение незначимых зависимостей и построение неориентированного графа с признаками в виде вершин и ребрами между значимо зависимыми признаками;
3. Разбиение пространства признаков на основе полученного графа связности с использованием методов иерархической или спектральной кластеризации для выделения сообществ в графах [37], а также методов вложения графов и его вершин [24].

Следует отметить, что в редких случаях, когда имеется доступ к детальному описанию технической системы и/или есть возможность получить консультацию от соответствующего инженера, кластеризация, основанная на полученном описании, превосходит автоматическую, но, как показывают результаты реальных экспериментов, обычно не с очень большим отрывом. В данной работе при обработке показаний телеметрии с пассажирского самолета А380 кластеризация была получена на основе описания подробной структуры и измеряемых параметров, представленного в отчетах ACMS, см. раздел 4. В итоге удалось идеально разделить пространство при-

знаков на связанные с двигателем группы из четырех параметров каждая и группу параметров, связанных с самолетом в целом.

2.2. Обнаружение аномалий

Общий случай обнаружения аномального измерения, основанный на изучении многообразия зависимостей между признаками в данных, показан на рис. 1, см. [29–32]. Здесь d – это точка выборки, а \hat{d} – ее аппроксимация, полученная за счет проекции на построенное по исторической выборке многообразию данных \mathcal{M} . В общем случае точка выборки d может быть вытеснена из порождающего многообразия \mathcal{M} либо из-за шума, либо из-за структурного изменения в многообразии, либо из-за того и другого.

Другим базовым методом обнаружения аномалий является оценка простых выбросов или экстремальных значений наблюдаемых параметров.

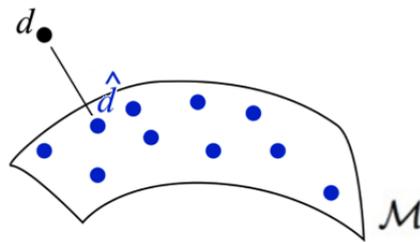


Рис. 1. Общая идея обнаружения аномалий основана на реконструкции низкоразмерного многообразия данных.

В целом, по причине того, что на практике приходится обрабатывать разнородные данные в самых разных ситуациях, приходится использовать различные методы обнаружения аномалий:

- некоторые данные телеметрии представляются в виде многомерных временных рядов, поэтому возможно обнаружение аномалий в потоках сенсорных данных на основе методов из [5–7, 18, 23, 36], при этом для прогнозирования редких событий и повышения надежности результата можно строить ансамбли моделей, см. [4, 26–28, 41], используя обнаруженные аномалии и их характеристики в качестве предвестников основных типов поломок;
- при построении моделей также возможно использование привилегированной информации о будущих событиях, которая доступна на этапе обучения моделей. Аналогичные подходы, использованные в [14, 16, 40] для обнаружения аномалий с одновременным выбором гиперпараметров модели, позволили существенно повысить точность прогноза;
- данные сенсоров, содержащиеся в обучающей выборке, зачастую имеют что-то типа пространственной размерности, поскольку различные компоненты временных рядов соответствуют различным узлам инженерной системы; таким образом,

можно построить граф зависимостей между потоками данным, регистрируемых различными сенсорами, и использовать современные методы выделения признаков из графовых данных [24, 25], а также методы выделения признаков из панельных временных рядов [34, 35], чтобы обогатить набор входных признаков, используемых для построения предсказательной модели;

- Другим типичным методом обнаружения аномалий является подход, основанный на построении суррогатной модели [11, 19, 45] для аппроксимации зависимостей между наблюдаемыми параметрами и последующей детекции аномалий на основе ошибки прогноза их значений с использованием непараметрических мер для оценки неопределенности [15, 17] в качестве индикатора уверенности в принятом решении.

2.3. Формирование предупреждающих сигналов

Предупреждающие сигналы (о возможных поломках в будущем) строятся путем объединения сигналов о появлении аномалий определенных типов. На практике, чтобы обеспечить интерпретируемость предупреждающего сигнала, обычно комбинируют сигналы об одновременном появлении аномалий не более чем 1–3 типов. Например, если x_{0t} и x_{1t} являются парой сигналов о различных типах аномалий в какой-то момент времени t , которые используются совместным образом, то результирующий предупреждающий сигнал будет

$$a_t = \begin{cases} 1, & \text{если } x_{0t} = 1 \text{ и } x_{1t} = 1, \\ 0, & \text{иначе.} \end{cases}$$

2.4. Сопоставление событий

Существует два возможных подхода к проблеме построения предупреждающего сигнала. Первый — применение вручную предметных и сопутствующих знаний для выбора и объединения аномалий в потенциально прогнозирующие сигналы. Второй — жадный алгоритм, который позволяет извлечь достаточно частые паттерны одновременного срабатывания сигнала и появления целевого события для автоматического построения наиболее эффективных предупреждающих сигналов из отдельных серий аномалий.

“Ручной” метод оправдан в тех (редких) случаях, когда число целевых событий в обучающей выборке недостаточно, чтобы можно было применять автоматический метод, основанный на оценке частоты одновременной встречаемости определенных последовательностей аномалий и целевых событий.

Более подробное описание алгоритма сопоставления событий есть в разделе 3.

2.5. Валидация

Проверка модели и оценка её точности – это один из наиболее важных шагов в анализе. Основной целью является проверка устойчивости прогнозирующих паттернов аномалий, извлеченных из данных. Проверка осуществляется за счет оценки

необходимых параметров для выбранных типов аномалий в объединенной выборке, содержащей данные телеметрии по всем объектам кроме одного, а затем тестового запуска полученной системы раннего оповещения на оставленном объекте.

3. СОПОСТАВЛЕНИЕ СОБЫТИЙ

Чтобы определить, какие аномалии в наборе данных являются прогнозирующими в отношении того или иного события отказа, можно использовать так называемый алгоритм сопоставления событий. По сути, это процедура выбора объектов, которая принимает на вход

- двоичные признаки, собранные в набор \mathcal{F} ; в данном случае, эти признаки кодируют появление аномалий того или иного типа,
- целевые переменные ($\{0, 1\}$ -сигналы),

и отфильтровывает те из признаков, которые не встречаются статистически чаще в коротком окне перед началом целевого события.

3.1. Разметка сигналов

Пусть $(X_{ft})_{t=1}^T \in \{0, 1\}$, $f \in \mathcal{F}$ являются предупреждающими сигналами или, в общем случае, бинарными признаками, а $(y_t)_{t=1}^T \in \{0, 1\}$ – индикаторами наступления целевого события. Для произвольного бинарного сигнала $(b_t)_{t=1}^T \in \{0, 1\}$ определим моменты срабатывания как

$$I_b = \{t = 1, \dots, T : b_t = 1\} \subseteq \{1, \dots, T\}.$$

Если имеется J событий, тогда положим $I_b = (t^j)_{j=1}^J$, где $t^j < t^{j+1}$.

Пусть $w \geq 1$ является целочисленной шириной “прогнозирующего окна”, $h \geq 0$ является “прогнозируемым горизонтом”, и $m \geq 0$ является целочисленной задержкой из-за “операций технического обслуживания”, проводимых в результате наступления события. Числа w и h определяют ширину и смещение соответствующего окна, в пределах которого считается, что сигнал корректно предупредил об этом предстоящем событии. В свою очередь, числа h и m определяют “период технического обслуживания”, в течение которого сигналы тревоги игнорируются, поскольку действие по обслуживанию все еще производится. Число h определяет за сколько моментов до наступления события тревога считается упреждающей: если $h = 0$, то любая тревога до наступления события, даже если она сработала непосредственно перед этим, является упреждающей.

Поскольку можно считать, что последовательные промежутки сигналов целевого события I относятся к одному и тому же основному отказу или его причине, события, которые находятся слишком близко друг к другу, объединяются и рассматриваются как одно “продолжительное” событие.

Введем следующие обозначения: в дальнейшем, если не указано иначе, под интервалами вида $[a, b)$ подразумевается закрытые слева и открытые справа целочисленные интервалы, т.е. для $a < b$

$$[a, b) := \{n \in \mathbb{Z} : a \leq n < b\},$$

и $[a, b) = \emptyset$ для $b \leq a$.

Параметры h , w и m определяют максимальное разделение по времени между последовательными событиями для их объединения: если $t^k - t^j < h + m + w$, тогда события j и k считаются за одно событие. Таким образом, “объединение” преобразует время срабатывания целевого ряда событий I_y в набор непересекающихся интервалов $[t_0^{jk}, t_1^{jk})$ с достаточным расстоянием (по времени) между ними: $t_0^{j_{k+1}} - t_1^{jk} \geq h + m + w$ для всех $k = 1, \dots, K$, где K является числом “различных” объединенных событий. Еще раз отметим, что интервалы — закрытые слева и открытые справа: t_0^{jk} — первый момент текущего начала события, t_1^{jk} — первый момент сразу после того, как событие закончилось.

“Прогнозирующая способность” сигнала $f \in \mathcal{F}$ относительно целевых событий y определяется путем подсчета того, как часто $(X_{ft})_{t=1}^T$ срабатывает слишком рано (ложная тревога) или вовремя (“правильная” тревога) относительно y . Аварийные сигналы во время действий по техническому обслуживанию игнорируются. Точная маркировка сигналов тревоги описана ниже.

Пусть s будет время срабатывания сигнала f , т.е. $X_{fs} = 1$. Если события $y_t = 1$ отсутствуют во время $t = 1, \dots, T$, $K = 0$, тогда тревога является ложной, если $s \in [0, T - w)$, и не релевантной, если $s \in [T - w, T]$. Если событие имеет место, то допустим $[t_{k0}, t_{k1})$ является интервалом, на котором происходит последнее K -тое событие. Этот случай аналогичен случаю “без события”, в котором тревога происходит слишком рано, если $s \in [t_{k1} + m, T - w)$, и не релевантна, если $s \in [T - w, T]$. В этих случаях предполагается, что $K + 1$ -ое событие произойдет в “неопределенном будущем”, т.е. на интервале $[t_{k+1,0}, t_{k+1,1}) = [T + h + 1, +\infty)$.

Наконец, рассмотрим случай $k < K$, и предположим, что k -тое событие происходит на интервале $[t_{k0}, t_{k1})$ (см. рис. 2). Если сигнал в момент s сработал до окончания события, $s < t_1$, тогда он относится к событию k и, в зависимости от того, где лежит s , тревога считается

1. Сработавшей вовремя, если $s \in [t_{k0} - h - w, t_{k0} - h)$;
2. Нерелевантной из-за технического обслуживания, когда $s \in [t_{k-1,1}, t_{k-1,1} + m)$;
3. Нерелевантной из-за наступления события, если s лежит либо в $[t_{k0} - h, t_{k0})$, либо в $[t_{k0}, t_{k1})$;
4. Ложной или сработавшей слишком рано, когда $s \in [t_{k-1,1} + m, t_{k0} - h - w)$.

Здесь предполагается, что при $k = 1$ предыдущее событие имело место в “бесконечном прошлом”, то есть во время $[t_{k-1,0}, t_{k-1,1}) = (-\infty, -m)$.

Пусть K^+ количество интервалов с корректными срабатываниями (совпадает с K), K^- — количество интервалов с ложными срабатываниями (либо K , либо $K + 1$).

Количество уникальных истинных срабатываний считается как $U_f^+ = \left| \{k = 1, \dots, K^+ : \exists t \in [t_{k0} - h - w, t_{k0} - h) \text{ при } x_{ft} = 1\} \right|$, и общее количество истинных срабатываний считается как $S_f^+ = \left| \{t = 1, \dots, T : x_{ft} = 1 \text{ и } \exists k = 1, \dots, K^+ \text{ при } t \in [t_{k0} - h - w, t_{k0} - h)\} \right|$.
 Количество уникальных ложных срабатываний считается как $U_f^- = \left| \{k = 1, \dots, K^- : \exists t \text{ при } x_{ft} = 1 \text{ и } t - \text{слишком рано}\} \right|$, и общее число ложных срабатываний считается как $S_f^- = \left| \{t = 1, \dots, T : x_{ft} = 1 \text{ и } t - \text{слишком рано}\} \right|$.

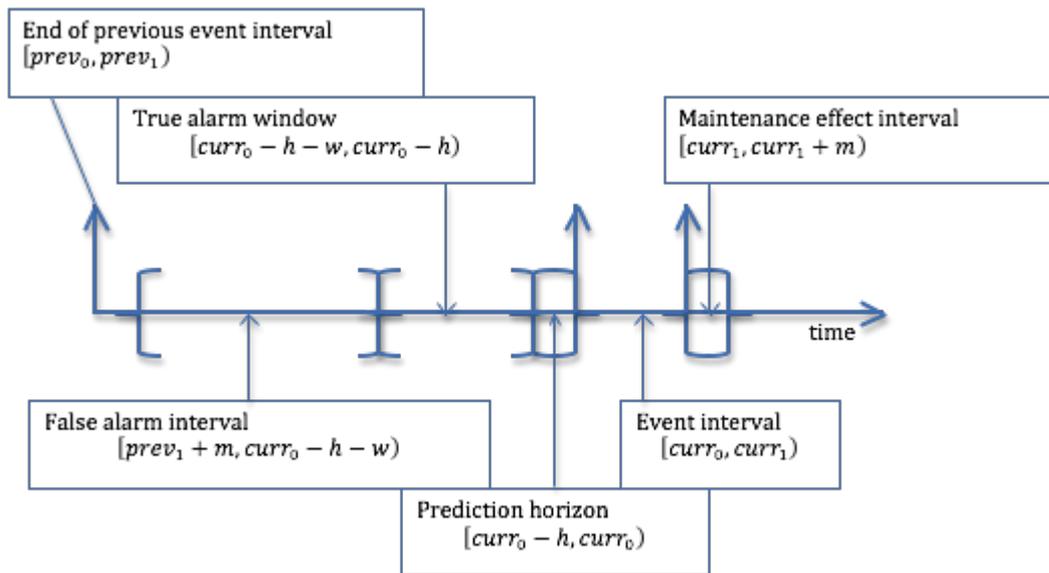


Рис. 2. Диаграмма порядка и геометрии ложных и истинных периодов срабатываний предупреждающих сигналов, прогнозируемого горизонта, интервалов события и интервалов с “эффектами” от технического обслуживания.

3.2. Алгоритм сопоставления событий

Если $(X_{ft})_{t=1}^T \in \{0, 1\}$ являются бинарными сигналами тревоги, $f \in \mathcal{F}$ и $(y_t)_{t=1}^T \in \{0, 1\}$ – индикатор наступления целевого события, тогда алгоритм сопоставления событий с аварийными сигналами выглядит следующим образом:

1. Создаются различные интервалы событий $(y_t)_{t=1}^T$ вида $[t_{k0}, t_{k1})$, для $k = 1, \dots, K$ (K может быть нулем);
2. Пусть K^+ является количеством интервалов с истинными срабатываниями (совпадает с K), K^- – количество интервалов с ложными срабатываниями (либо K , либо $K + 1$);
3. Для каждого $f \in \mathcal{F}$ вычисляются базовые метрики:

- (a) Подсчитывается количество общих и уникальных истинных срабатываний S_f^+ и U_f^+ , соответственно;
- (b) Считается S_f^- и U_f^- , количество общих и уникальных ложных срабатываний, соответственно;
- (c) Вычисляется следующая производная метрика:
- оценки вероятности истинных и ложных срабатываний $p_f^\varepsilon = \frac{S_f^\varepsilon}{K^\varepsilon}$ для $\varepsilon \in \{+, -\}$;
 - доля ложных срабатываний: $fa/cf_f = \frac{S_f^-}{U_f^+}$;
 - р-значение PV_f статистического теста на равенство между p_f^+ и p_f^- , когда $p_f^+ > p_f^-$;
4. Применяется одна из следующих стратегий, чтобы получить $\mathcal{F}_{\text{predictive}} \subseteq \mathcal{F}$:
- правило на основе t-теста : оставить $f \in \mathcal{F}_{\text{predictive}}$, если $U_f^+ > 1$ и $PV_f \geq \alpha$, где α — некоторый фиксированный уровень значимости $[0, 1]$; обычно $\alpha = 0.05$;
 - “жесткое” правило : оставить $f \in \mathcal{F}_{\text{predictive}}$, когда $U_f^+ \geq \theta$ и $U_f^- = 0$; здесь θ — в роли жесткого порога, который контролирует поддержку гипотезы, что f прогнозирует возникновение события y ;
 - “мягкое” правило : оставить $f \in \mathcal{F}_{\text{predictive}}$, если $fa/cf_f \leq \theta$, где θ — некоторый “мягкий” порог, контролирующий отношение количества ложных тревог к количеству истинных тревог.

Возможно, что ни один сигнал в \mathcal{F} не является достаточно прогнозирующим в соответствии с указанными стратегиями фильтрации и их параметрами. Конечно, в этом случае $\mathcal{F}_{\text{predictive}} = \emptyset$. Более того, вполне вероятно, что разные целевые события y^1 и y^2 имеют разные наборы прогнозирующих признаков $\mathcal{F}_{\text{predictive}}(y^1) \neq \mathcal{F}_{\text{predictive}}(y^2)$.

3.3. Метрики прогнозирования

Ключевыми показателями, которые характеризуют эффективность любой системы раннего предупреждения с фиксированным горизонтом прогнозирования и шириной окна, являются:

- количество ложных срабатываний или частота ложных срабатываний;
- отношение охваченных событий (истинных срабатываний) к общему количеству событий.

Точность системы раннего предупреждения обратно пропорциональна частоте ложных тревог (или их числу). Чем точнее система, тем больше уверенности ее пользователей в том, что тревоги действительно предвосхищают события, а не просто являются ложным сигналом.

Чувствительность системы измеряет долю событий, которые система может отловить, и она прямо пропорциональна количеству охватываемых событий (коэффициенту охвата).

Эти показатели лежат в основе алгоритма сопоставления событий и используются непосредственно для выбора признаков. Для конкретного целевого события и сигналов тревоги $(y_t)_{t=1}^T \in \{0, 1\}$ и $(x_t)_{t=1}^T \in \{0, 1\}$, соответственно, эти метрики вычисляются следующим образом:

1. Вычисление интервалов событий и ложных и истинных срабатываний для y_t с параметрами $h \geq 0$, $w \geq 1$ и $m \geq 0$, как в разделе 3.1;
2. Для сигнала x_t подсчитываются K^+ , U^+ , и S^- по заданной структуре интервалов;
3. Подсчет метрик:
 - степень ложных срабатываний $fa_x = \frac{S_f^-}{K^+} \in [0, +\infty)$;
 - доля покрытия $cf_x = \frac{U_f^+}{K^+} \in [0, 1]$.

Обратите внимание, что ложные тревоги учитываются как общее количество совпадений с ложными интервалами срабатываний, тогда как истинные тревоги вычисляются только как уникальные совпадения.

4. ОПИСАНИЕ ИСХОДНЫХ ДАННЫХ

Данные телеметрии пассажирского самолета представлены в Central Management System (CMS) отчетах и в Aircraft Condition Monitoring System (ACMS) отчетах. Отчеты CMS содержат бинарные (есть/нет) данные о тех или иных типах событий (например, поломки определенных типов) в определенные моменты времени. Отчеты ACMS содержат информацию об одном или нескольких моментальных снимках значений набора заданных параметров, а также сообщения с описаниями тех или иных поломок, собранными из разных подсистем самолета.

Данные о появлении событий в CMS и необработанные данные отчетов ACMS хранятся в простых текстовых файлах с табличной структурой. В данной работе представлены результаты, полученные по историческим данным, которые охватывали период с апреля 2011 г. по ноябрь 2013 г. Данные были доступны для 31-го самолета.

Каждая запись в файлах CMS содержит как минимум время появления события, номер самолета и фазу полета. Дополнительные поля не используются при подготовке набора данных и последующем анализе.

Каждый отчет ACMS имеет тип (число в диапазоне 1 до 30), содержит время сбора данных, фазу полета, значения измерений параметров и дополнительные данные, такие как коды пунктов отправления и прибытия, тип самолета и т.д. ACMS отчеты соответствуют спецификации Engine Alliance; доступны следующие типы отчетов: 1 до 15, 17, 21, 23, 25, 26, и 30.

Наборы измеряемых параметров, содержащихся в отчете, зависят от его типа. Диапазон времени, охватываемый доступными отчетами, также зависит от их типа, например, все доступные отчеты типа 1 могут охватывать диапазон времени, отличный от диапазона времени, для которого представлены отчеты типа 2. В таблице 1 представлена сводка доступных отчетов ACMS, в т.ч. и информация о том, для сколько самолетов из 31-го имеются отчеты ACMS соответствующего типа.

Отчет	Столбцы	Записи	Самолеты (из 31)	Фазы
R1	322	58917	31	8, 9
R2	406	62919	31	8, 9
R3	318	33831	31	7.1
R4	339	33467	31	7.1
R5	1452	22	12	1, 2, 8
R6	differ	80	17	1, 2, 8, 11
R7	315	122	4	8
R8	differ	5701	25	1.1, 2
R9	differ	30823	31	3
R10	differ	355	28	1, 1.1, 8
R12	differ	27432	31	11
R13	differ	2574	31	9
R14	83	34490	31	11
R15	differ	359	30	9
R17	225	16964	31	1, 2
R21	differ	24016	31	8
R23	differ	6426	31	12
R25	differ	33533	31	11
R26	differ	264	19	2, 8, 8.1
R30	831	33557	31	8.1

Таблица 1. Информация об ACMS-отчетах

Для анализа поломок некоторых типов в A380 использовались отчеты ACMS, доступные для всех самолетов и имеющие одинаковый набор столбцов (параметров). Согласно таблице 1, эти отчеты имеют номера 1, 2, 3, и 4.

5. ОБРАБОТКА ДАННЫХ

“Характеристика” или “измерение” – это замер параметра, содержащийся в отчете ACMS, созданном на определенной фазе полета. Параметры из разных отчетов, но с одинаковыми именами, рассматриваются как одно измерение одного и того же явления. Например:

- Параметр MN_D10 (Mach Number) из отчета R1, замеряемый в фазе 8, является одним признаком;
- MN_D10 из R2, замеряемый в фазе 8, является тем же самым признаком, хотя он измеряется для отчета другого типа;
- MN_D11 из R1, замеряемый в фазе 8, является другим признаком, потому что, как указывает его имя, он использует другой источник данных для определения своего значения;
- MN_D10 из R1, замеряемый в фазе 9, тоже является другим признаком, так как полетная фаза отличается (см. список полетных фаз на рис. 3).

Для каждого самолета были сформированы матрицы X и Y со следующими свойствами:

- X является матрицей полетов-признаков, т.е., X_{ij} – это значение j -го признака, замеренного во время i -го полета;

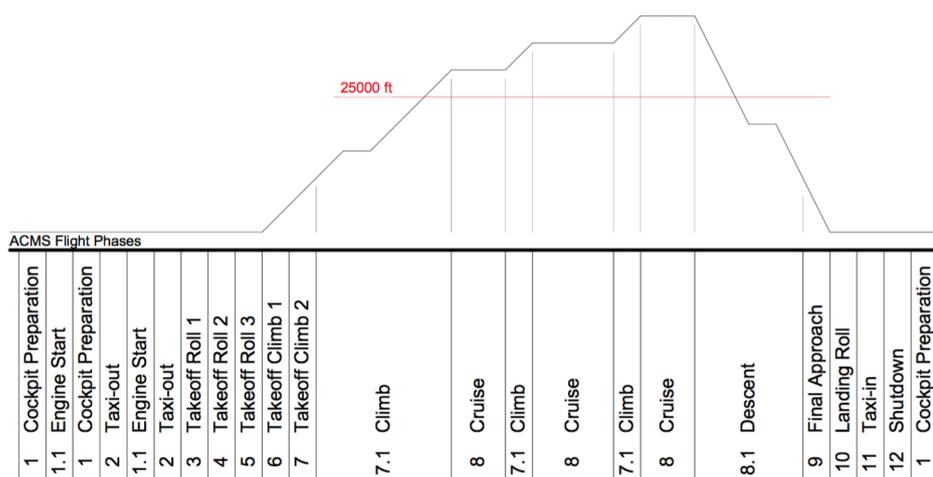


Рис. 3. Фазы полета.

- X может содержать пропущенные значения по одной из следующих причин:
 - отсутствующее значение в исходных данных ACMS, значение, помеченное как ошибочное, как определено префиксом в необработанном файле отчета;
 - не каждый отчет генерируется во время полета;
- Y является бинарной матрицей полетов-отказов, т.е. $Y_{ij} = 1$ если сообщение CMS с j -ом кодом ошибки возникает во время i -го полета, иначе $Y_{ij} = 0$;

Каждая строка X и Y представляет один полет, тогда как столбец X – отдельный признак ACMS, а столбец Y – индикатор появления кода ошибки CMS. Строки обеих матриц синхронизируются: X_i и Y_i соответствуют одному и тому же полету, и упорядочены хронологически.

Необработанные данные ACMS и CMS напрямую не идентифицируют рейс, в течение которого были созданы отчеты, и не предоставляют информацию о времени и продолжительности полета. Эта информация имеет решающее значение для правильной синхронизации набора данных, так как в противном случае при формировании обучающей выборки могут возникнуть следующие артефакты:

- группировка экземпляров отчета, которые соответствуют различным рейсам;
- отчеты с разных рейсов могут объединяться в один и тот же рейс;
- рейс, во время которого отчеты не создавались вообще, что означает, что некоторые последовательные строки в наборе данных могут представлять непоследовательные рейсы;
- неправильное назначение записей о сбоях CMS может привести к неверным результатам.

5.1. Детали обработки данных

Указанные выше проблемы были решены путем объединения данных строк из всех отчетов и упорядочения их по временной метке получения отчета и его фазе. Это решение основывается на следующих предположениях:

- Во время полета воздушное судно проходит этапы полета в их естественном порядке: последовательно от 1 до 7, затем 7.1, 8, 8.1 в произвольном порядке, и, наконец, от 9 до 12 последовательно;
- Во время каждого полета, по крайней мере, один отчет ACMS генерируется на этапах 1–7, как минимум один на фазах {7.1, 8, 8.1,}, и еще как хотя бы один на фазах 9–12;

Алгоритм построения матриц данных X и Y для каждого самолета:

1. Распознавание полетов и сопоставление им отчетов
 - (a) Сортировка всех отчетов ACMS, доступных для воздушного судна, по отметке времени их получения;
 - (b) Нахождение пересечения временных диапазонов, охватываемых отчетами ACMS R1 – R4, и отбрасывание всех отчетов, полученных за пределами этого пересечения;
 - (c) Обозначение оставшихся отчетов как r_1, \dots, r_n (в хронологическом порядке). Обозначение значения параметра `param` в отчете r_i как $r_i[\text{param}]$.
 - (d) Нахождение всех $i > 1$, таких, что r_i имеет более раннюю полетную фазу, чем r_{i-1} , исключая случаи, когда фазы и отчета r_i , и отчета r_{i-1} с точки зрения знаний о предметной области не должны содержать информации о тех типах поломок, прогноз которых предполагается делать.
 - (e) Обозначим i из предыдущего шага как s_2, \dots, s_m , также обозначим $s_1 = 1, s_{m+1} = n + 1$.
 - (f) В таком случае, отчеты, полученные в полете j , — это $r_{s_j}, r_{s_{j+1}}, \dots, r_{s_{j+1}-1}$, и m — полное число полетов в данных. Далее, используя указанное правило, происходит сопоставление отчетов полетам.
 - (g) Оценивание времени начала и времени окончания каждого рейса согласно формуле $t_j^{\text{start}} = r_{s_j}[\text{TIMESTAMP INDEX}]$, $t_j^{\text{end}} = r_{s_{j+1}-1}[\text{TIMESTAMP INDEX}]$. Таким образом, полет j начинается не позже, чем t_j^{start} , и заканчивается раньше, чем t_j^{end} .

6. ПОСТРОЕНИЕ ПРОГНОЗОВ НЕКОТОРЫХ ТИПОВ ПОЛОМОК ДЛЯ A380

Данные, используемые в рассматриваемом примере, взяты из отчетов ACMS с 1 по 4; в этих данных в том числе содержится информация, необходимая для идентификации ошибок индикации, планирования, мониторинга газового тракта и контроля работы двигателя.

Эти отчеты содержат данные, которые являются основным элементом в планировании демонтажа двигателя и работ по его обслуживанию, с возможностью увеличения времени работы двигателя на крыле.

Ниже приведены основные описания отчетов 1–4:

Отчет <001 > :

“Отчет о работе двигателя” предоставляет данные о таких неисправностях индикации, как ошибки датчика (смещение, дрейф, шум) и ошибки обработки датчика (схема формирования сигнала, проводка). Примерами сбоев планирования являются неправильная оснастка, разорванные или липкие соединения для отводов и лопастей, утечки отводов, неисправности ТСС (клапаны) и неисправности воздушного масляного радиатора (клапаны). Обнаружению ошибок планирования часто помогает информация из отчетов, соответствующих фазе “Взлет” и/или “Набор высоты”. Мониторинг газового тракта предоставляет данные для отслеживания снижения производительности модуля.

Отчет <002> :

В “Отчете о характеристиках крейсерского режима самолета” (Aircraft Cruise Performance Report) собраны данные, необходимые для мониторинга характеристик самолета на этапе “крейсерского режима”. Данный отчет дополняет “Отчет о состоянии двигателя в течение крейсерского режима” (Engine Cruise Report), за счет записи значений большого количества характеристик самолета, двигателя и других его внутренних компонентов.

Отчеты <003> и <004> :

“Отчет о состоянии двигателя во время взлета” (Engine Climb Report) и “Отчет о состоянии двигателя во время набора высоты” (Engine Takeoff Report), соответственно, предоставляют данные о неисправностях индикации и планирования, которые подробно описаны в “Отчете о состоянии двигателя в течение крейсерского режима” (Engine Stable Cruise Report); измерения мониторинга производительности двигателя включают в себя “максимальный дополнительный запас температуры выходящих газов” (Maximum Continuous EGT Margin), ухудшение снижения тяги при подъеме во время фазы подъема (Climb Reduced Thrust Derate during Climb phase), а также запас температуры выходящих газов (Exhaust Gas Temperature Margin, EGT margin) при наборе высоты, и ухудшение снижения тяги при взлете во время фазы взлета (Take Off Reduced Thrust Derate during the Take off phase). Эти отчеты дополняют отчет <001>, поскольку скрытые ошибки планирования требуют данных в двух или более эксплуатационных режимах.

Цель работы состояла в том, чтобы найти сбои, связанные с работой двигателей (коды АТА 60-80; АТА codes — Air Transport Association numbering system), которые можно было предвидеть с помощью сигналов раннего предупреждения. Физически нерелевантные позиционные, угловые и контрольные параметры были исключены. Результирующие сигналы оповещения основаны в основном на данных, связанных с двигателем, но иногда включают измерения, относящиеся к самолету в целом.

Набор данных был подготовлен по методике, описанной в разделе 5

6.1. Разбиение признаков

Исходный формат отчетов 1-4, представленный в спецификациях для движков Engine Alliance, имеет следующий вид: блок с заголовками, за которым следует блок данных. Каждый блок представляет собой набор групп, состоящий из строки заголовка и до четырех строк данных в столбце. Каждая строка пронумерована по следующей схеме: имя блока || номер строки внутри блока, где “||” обозначает конкатенацию. Имя блока имеет формат ‘ [A – Z][0 – 9] ’, т.е. буква и десятичная цифра. Номер строки представляет собой однозначное число. Заголовок блока определяет измерение, сообщаемое в его столбце, тогда как количество строк варьируется в зависимости от источника данных.

В этих отчетах в столбце параметров, каждый из которых содержит 4 записанные строки измерений, каждое измерение имеет 1 из 4 каналов ACMS в качестве источника. Каждый канал может быть идентифицирован с определенной точкой крепления двигателя на летательном аппарате, которая, как предполагается, проложена через фюзеляж самолета и никогда не ремонтируется в течение всего срока службы самолета, пока он полностью не выведен из эксплуатации.

В исходных данных этих отчетов столбцы параметров свернуты в одну строку с идентификаторами параметров (измерений), имеющими следующий формат: имя параметра || “_” || имя блока || номер строки внутри блока.

Важно отметить, что имена параметров представляют собой буквенно-цифровые последовательности, отделенные от идентификаторов столбцов подчеркиванием. Для этих отчетов имеется очень простое правило, которое определяет, относится ли данный параметр к летательному аппарату или является измерением некоторой характеристики, связанной с двигателем:

- относится к двигателю, если имеется четыре строки в блоке;
- относится к летательному аппарату в противном случае.

Например “EGTRM_E10, EGTRM_E11, EGTRM_E12, и EGTRM_E13” считаются относящимися к двигателю, тогда как “CAS_A10, CAS_A11, и CAS_A12”, или “GWT_V10” являются различными измерениями, относящимися к летательному аппарату. Таким образом, связанные с двигателем функции разделены на группы по 4 измерения — по одному на каждый двигатель (канал ACMS).

Группы измерений, используемые в анализе, формируются из имени параметра и блока, в то время как элементы группы являются значениями в строках соответствующего столбца. Важно отметить, что данные группируются только внутри каждого отчета, а не между отчетами. Это порождает следующие группы: “R” || номер отчета || “:” имя параметра || “_” || имя блока || “x”.

Например, упомянутые ранее измерения получают имя: “R1::EGTRM_E1x”. Данные по самолетам вообще не группируются, и каждое измерение используется как отдельный признак.

6.2. Выделение аномалий

Аномалии в каждой группе параметров, связанных с двигателем, извлекаются детектором на основе метода главных компонент [12]:

1. Тренировочные данные используются для конструирования одномерного линейного многообразия, которое объясняет большую часть совместной дисперсии элементов в группе; см. рис. 1 с \mathcal{M} , представляющим собой линейное подпространство, построенное из главных направлений оцененной ковариационной матрицы измерений в группе;
2. Тестовые данные d проецируются на многообразие в точку \hat{d} , по отношению к которой вычисляется ошибка линейной реконструкции $Q = \|d - \hat{d}\|^2$;
3. Величина аномальности вычисляется путем взятия логарифма от Q ;
4. Порог, который классифицирует величины на “нормальные” и “аномальные”, устанавливается равным $(1 - \alpha)$ -квантили эмпирического распределения таких величин на тренировочном наборе данных, где $\alpha \in \{95\%, 99\%\}$.

Естественно, вместо этого конкретного метода обнаружения аномалий на основе РСА, можно использовать любой другой метод. В рассматриваемом случае именно детектор на основе РСА позволил получить наиболее точные результаты.

Процесс извлечения аномалий для измерений, относящихся с самолету, устроен гораздо проще: в этом случае мы используем стандартное пороговое правило; порог отклонения установлен равным $(1 - \alpha)$ -квантили эмпирического распределения измерений в тренировочном наборе данных, где $\alpha \in \{95\%, 99\%\}$.

6.3. Генерация предупреждающих сигналов

Для каждого целевого сигнала события $(y_t)_{t=1}^T$ (появление кода ошибки в отчетах CMS) строится подмножество пар отдельных аномалий $\mathcal{P} \subseteq \{\{a_1, a_2\} : a_1, a_2 \in \mathcal{F}\}$ такое, что $A \in \mathcal{P}$ тогда и только тогда, когда y_t считается достаточно предсказуемым по сигналу $A_t = \bigwedge_{a \in A} a_t$. Интуитивно, если аномалии происходят точно в течение одного и того же полета, см. раздел 2.3, и значительно чаще перед событием сбоя, то они находятся в \mathcal{P} .

Сигналы тревоги для события y синтезируются с использованием: $S_t = \bigvee_{A_t \in \mathcal{P}} A_t$, то есть сигнал раннего предупреждения срабатывает, если была обнаружена хотя бы одна совместная аномалия. Преимущество этого “хотя бы одно срабатывание” правила заключается в том, что объединение прогнозирующих аномалий увеличивает шансы на успешное предвидение события, в то время как недостатком является то, что такой ансамбль имеет повышенное количество ложных тревог, поскольку не всегда аномалии идеально предсказывают событие сбоя (то есть никогда не возникают, кроме как незадолго до сбоя).

Results for A380 engines	Warnings Codes	Covered Warnings	False Alarms	Total Number of Warnings
ENG 3 STALL	7100W430	4	0	21
ENG 1 IGN A FAULT	7400W090	2	0	5
ENG 4 IGN A FAULT	7400W120	3	0	14
ENG 3 REVERSER INHIBITED	7830W130	2	0	23
ENG 2 REVERSER UNLOCKED	7830W220	7	1	22
ENG 3 REVERSER UNLOCKED	7830W230	2	0	10
ENG 2 REVERSER CTL FAULT	7830W320	6	0	23
ENG 3 FADEC SYS FAULT	7325W230	2	1	4

Рис. 4. Результаты прогнозирования поломок различного типа, связанных с двигателем (см. описание в тексте раздела 6.4).

6.4. Оценка точности прогнозов

Параметры качества системы раннего предупреждения (см. раздел 3.3) были установлены для горизонта $h = 0$, ширины окна полетов $w = 30$, и продолжительности эффектов от технического обслуживания $m = 0$.

На рис. 4 представлены результаты прогнозирования поломок различного типа, связанных с двигателем:

- в первом столбце таблицы перечислены названия различных типов поломок, связанных с двигателем;
- во втором столбце таблицы перечислены АТА-коды этих типов поломок;
- последующие три столбца таблицы для каждого типа поломок содержат информацию о количестве поломок, которое удалось правильно спрогнозировать, количестве ложных тревог, и, наконец, об общем количестве поломок этого типа, соответственно.

Таким образом, из результатов, представленных на рис. 4, следует, что за счет использования предложенного в данной работе алгоритма удалось получить точность прогноза, которая обеспечивает выявление $\sim 28\%$ важных поломок, и при этом количество ложных тревог составляет менее $\sim 10\%$ (то есть в среднем на 9 правильно предсказанных поломок приходится 1 ложная тревога). Данные показатели точности соответствуют требованиям, предъявляемым на практике к подобного рода моделям, и поэтому модели были интегрированы в систему заказчика для раннего оповещения о поломках.

Отметим, что построенные модели раннего оповещения о поломках каждая основаны на комбинации нескольких “простых” моделей. Каждая из этих “простых” моделей либо детектирует аномалии в поведении параметров из группы параметров, связанных с двигателем, либо по пороговому правилу обнаруживает аномалии в значениях параметров, относящихся к самолету (см. раздел 6.2). Таким образом, в каждом конкретном случае раннего оповещения о поломке возможно выяснить в результате изменения значений каких именно параметров произошло объявление сигнала тревоги. Это в свою очередь делает такого рода модели интерпретируемы-

ми и помогает инженерам, обслуживающим техническую систему, быстрее находить причины поломки, если она происходит.

Если применять стандартные методы машинного обучения для построений предиктивных моделей на основе случайных лесов, градиентного бустинга над решающими деревьями, нейронных сетей и др., то не удастся достичь сравнимых показателей точности, которые представлены на рис. 6.4. Этот вывод имеет место как в случае, когда используются те же самые входные признаки, которые были использованы при построении предложенных моделей раннего оповещения о поломках, так и в случае, когда используются любые другие доступные входные сигналы телеметрии и их комбинации. Возможно, причина в том, что для достижения эффективного соотношения между “ложными тревогами” и “пропуском цели” необходимо использовать именно те модели, которые эффективно учитывают структуру задачи. Речь идет о том, что на практике, прежде чем поломка происходит, обычно начинаются небольшие изменения в поведении системы, вызванные появляющимися дефектами, которые можно обнаружить по изменению свойств потока информации с датчиков. Эти дефекты со временем усугубляются и приводят к критическим сбоям, вплоть до полного отказа оборудования. Соответственно, описанный в разделе 3 подход позволяет строить модели, имитирующие указанное наблюдение, и при этом эти модели имеют достаточно “простую” структуру — а потому они будут робастные. В свою очередь, предиктивные модели, построенные с помощью стандартных методов машинного обучения (конструирование признаков на основе имеющихся данных телеметрии и последующее применение методов типа градиентного бустинга над решающими деревьями), избыточно сложны и имеют завышенный уровень ложных тревог.

7. ЗАКЛЮЧЕНИЕ

В данной работе рассмотрена задача построения модели для прогнозирования редких событий в ситуации, когда наблюдаются сотни различных показателей о состоянии технической системы и требуется спрогнозировать поломку, примеров которой в имеющейся выборке исторических данных достаточно мало (выборка несбалансирована). Благодаря разработанному подходу удалось спрогнозировать некоторые возможные отказы авиационной техники. За счет специальной структуры модели мы обеспечили низкий уровень ложных тревог при значимой доле обнаруженных поломок. Эксперименты с применением стандартных методов машинного обучения для построения предиктивных моделей не позволили получить прогнозы со сравнимой точностью.

Е. Бурнаев выражает благодарность И. Назаров и П. Ерофееву за помощь в обработке данных, Н. Ключникову — за техническую помощь в подготовке текста статьи, и компании ООО Датадванс — за предоставление данных и постановку задачи в рамках договора с компанией, производящей пассажирские самолеты.

Исследование было частично поддержано грантом РФФИ 16-29-09649 офи_м.

СПИСОК ЛИТЕРАТУРЫ

1. Maintenance optimization. Airplane health management, 2015.
2. S. Alestra, C. Bordry, C. Brand, E. Burnaev, P. Erofeev, A. Papanov, and C. Silveira-Freixo. Application of rare event anticipation techniques to aircraft health management. *Advanced Materials Research*, 1016:413–417, 2014.
3. S. Alestra, E. Burnaev, and et al. Rare event anticipation and degradation trending for aircraft predictive maintenance. In *Proc. of the joint WCCM – ECCM – ECFD 2014 Congress*, pages 1–12, 2014.
4. A. Artemov and E. Burnaev. Ensembles of detectors for online detection of transient changes. In *Proc. SPIE*, volume 9875, pages 9875 – 9875 – 5, 2015.
5. A. Artemov and E. Burnaev. Detecting performance degradation of software-intensive systems in the presence of trends and long-range dependence. In *2016 IEEE 16th ICDMW*, pages 29–36, 2016.
6. A. Artemov and E. Burnaev. Optimal estimation of a signal perturbed by a fractional brownian noise. *Theory of Probability and Its Applications*, 60(1):126–134, 2016.
7. A. Artemov, E. Burnaev, and A. Lokot. Nonparametric decomposition of quasi-periodic time series for change-point detection. In *Proc. SPIE*, volume 9875, pages 9875–9875–5, 2015.
8. Marcia Baptista, Ivo P. de Medeiros, Joao P. Malere, Cairo Nascimento, Helmut Prendinger, and Elsa M.P. Henriques. Comparative case study of life usage and data-driven prognostics techniques using aircraft fault messages. *Computers in Industry*, 86:1 – 14, 2017.
9. Marcia Baptista, Ivo P. de Medeiros, Joao P. Malere, Helmut Prendinger, Cairo L. Jr. Nascimento, and Elsa Henriques. Improved time-based maintenance in aeronautics with regressive support vector machines. In *Annual Conference of the Prognostics and Health Management Society 2016*, pages 1–10, 2016.
10. Marcia Baptista, Shankar Sankararaman, Ivo. P. de Medeiros, Cairo Nascimento, Helmut Prendinger, and Elsa M.P. Henriques. Forecasting fault events for predictive maintenance using data-driven techniques and arma modeling. *Computers and Industrial Engineering*, 115:41 – 53, 2018.
11. M. Belyaev, E. Burnaev, E. Kapushev, M. Panov, P. Prikhodko, D. Vetrov, and D. Yarotsky. Gtapprox: Surrogate modeling for industrial design. *Advances in Engineering Software*, 102:29–39, 2016.
12. E. Burnaev and S. Chernova. On an iterative algorithm for calculating weighted principal components. *Journal of Communications Technology and Electronics*, 60(6):619–624, 2015.
13. E. Burnaev, P. Erofeev, and A. Papanov. Influence of resampling on accuracy of imbalanced classification. In *Proc. SPIE*, volume 9875, pages 9875–9875–5, 2015.
14. E. Burnaev, P. Erofeev, and D. Smolyakov. Model selection for anomaly detection. In *Proc. SPIE*, volume 9875, pages 9875 – 9875 – 6, 2015.
15. E. Burnaev and I. Nazarov. Conformalized kernel ridge regression. In *15th IEEE ICMLA*, pages 45–52, 2016.
16. E. Burnaev and D. Smolyakov. One-class svm with privileged information and its application to malware detection. In *IEEE 16th ICDMW*, pages 273–280, 2016.

17. E. Burnaev and V. Vovk. Efficiency of conformalized ridge regression. In Proceedings of The 27th Conference on Learning Theory, volume 35, pages 605–622. PMLR, 2014.
18. E. V. Burnaev and G. K. Golubev. On one problem in multichannel signal detection. Problems of Information Transmission, 53(4):368–380, 2017.
19. E. V. Burnaev and P. V. Prikhod'ko. On a method for constructing ensembles of regression models. Automation and Remote Control, 74(10):1630–1644, 2013.
20. J. Dai and H. Wang. Evolution of aircraft maintenance and logistics based on prognostic and health management technology. In Lecture Notes in Electrical Engineering. Proceedings of the First Symposium on Aviation Maintenance and Management-Volume II, volume 297, pages 665–672. Springer, 2014.
21. Emanuele Fumeo, Luca Oneto, and Davide Anguita. Condition based maintenance in railway transportation systems based on big data streaming analysis. Procedia Computer Science, 53:437 – 446, 2015. INNS Conference on Big Data 2015 Program San Francisco, CA, USA 8-10 August 2015.
22. Fausto P. García, Diego J. Pedregal, and Clive Roberts. Time series methods applied to failure prediction and detection. Reliability Engineering and System Safety, 95(6):698 – 703, 2010.
23. V. Ishimtsev, A. Bernstein, E. Burnaev, and I. Nazarov. Conformal k-nn anomaly detector for univariate data streams. In Proceedings of the Sixth COPA Workshop, volume 60, pages 213–227. PMLR, 2017.
24. S. Ivanov and E. Burnaev. Anonymous walk embeddings. In Proc. of the 35th ICML, volume 80, pages 2186–2195. PMLR, 2018.
25. S. Ivanov, N. Durasov, and E. Burnaev. Learning node embeddings for influence set completion. In Proc. of IEEE ICDMW, pages 1034–1037, 2018.
26. A. Korotin, V. V'yugin, and E. Burnaev. Long-Term Online Smoothing Prediction Using Expert Advice. arXiv e-prints, abs/1711.03194, 2017.
27. A. Korotin, V. V'yugin, and E. Burnaev. Aggregating strategies for long-term forecasting. In Proceedings of the Seventh Workshop on Conformal and Probabilistic Prediction and Applications, volume 91, pages 63–82. PMLR, 2018.
28. A. Korotin, V. V'yugin, and E. Burnaev. Adaptive Hedging under Delayed Feedback. arXiv e-prints, abs/1902.10433, 2019.
29. A. Kuleshov, A. Bernstein, and E. Burnaev. Conformal prediction in manifold learning. In Proceedings of the Seventh COPA Workshop, volume 91, pages 234–253. PMLR, 2018.
30. A. Kuleshov, A. Bernstein, and E. Burnaev. Kernel regression on manifold valued data. In Proceedings of IEEE 5th DSAA, pages 120–129, 2018.
31. A. Kuleshov, A. Bernstein, and E. Burnaev. Manifold learning regression with non-stationary kernels. In Artificial Neural Networks in Pattern Recognition, pages 152–164. Springer, 2018.
32. A. Kuleshov, A. Bernstein, E. Burnaev, and Y. Yanovich. Machine learning in appearance-based robot self-localization. In 2017 16th IEEE ICMLA, pages 106–112, 2017.
33. Liu Datong, Peng Yu, and Peng Xiyuan. Fault prediction based on time series with online combined kernel svr methods. In IEEE Instrumentation and Measurement Technology Conference, pages 1163–1166, May 2009.

34. R. Rivera, I. Nazarov, and E. Burnaev. Towards forecast techniques for business analysts of large commercial data sets using matrix factorization methods. *Journal of Physics: Conference Series*, 1117(1):012010, 2018.
35. R. Rivera-Castro, I. Nazarov, Yu. Xiang, A. Pletneev, I. Maksimov, and E. Burnaev. Demand forecasting techniques for build-to-order lean manufacturing supply chains. *arXiv e-prints*, abs/1905.07902, 2019.
36. A. Safin and E. Burnaev. Conformal kernel expected similarity for anomaly detection in time-series data. *Advances in Systems Science and Applications*, 17(3):22–33, 2017.
37. B. Saha, A. Mandal, S.B. Tripathy, and D. Mukherjee. Complex networks, communities and clustering: A survey. *CoRR*, abs/1503.06277, 2015.
38. T. Shen, F. Wan, W. Cui, and B. Son. Application of prognostic and health management technology on aircraft fuel system. In *IEEE Proceedings of 2010 Prognostics and System Health Management Conference*, pages 1–7, Macao, 12–14 Jan 2010.
39. D. Smolyakov, A. Korotin, P. Erofeev, A. Papanov, and E. Burnaev. Meta-learning for resampling recommendation systems. In *Proc. SPIE 11041, 11th ICMV*, 110411S, 2019.
40. D. Smolyakov, N. Sviridenko, E. Burikov, and E. Burnaev. Anomaly pattern recognition with privileged information for sensor fault detection. In *ANNPR*, pages 320–332. Springer, 2018.
41. D. Smolyakov, N. Sviridenko, V. Ishimtsev, E. Burikov, and E. Burnaev. Learning Ensembles of Anomaly Detectors on Synthetic Data. *arXiv e-prints*, abs/1905.07892, 2019.
42. S. Su, W. Zhang, and S. Zhao. Online fault prediction for nonlinear system based on sliding arma combined with online ls-svr. In *Proceedings of the 33rd Chinese Control Conference*, pages 3287–3291, July 2014.
43. L. Tegtmeier. Math and maintenance. *Aviation Week and Space Technology*, 174(39), 2012.
44. Wende TIAN, Minggang HU, and Chuankun LI. Fault prediction based on dynamic model and grey time series model in chemical processes. *Chinese Journal of Chemical Engineering*, 22(6):643 – 650, 2014.
45. A. Zaytsev and E. Burnaev. Large scale variable fidelity surrogate modeling. *Annals of Mathematics and Artificial Intelligence*, 81(1):167–186, 2017.

ON CONSTRUCTION OF EARLY WARNING SYSTEMS FOR PREDICTIVE MAINTENANCE IN AEROSPACE INDUSTRY

Burnaev E.V.

Skolkovo Institute of Science and Technology, Moscow, Russia
Institute for Information Transmission Problems, Moscow, Russia

The problem of constructing predictive models for early warning systems for diagnostic maintenance in the aerospace industry is considered. A new approach to predicting rare failures based on a new methodology that takes into account the properties of technical systems and specific requirements imposed by applications is proposed.

KEYWORDS: Monitoring system, anomaly detection, failure prediction, machine learning, predictive maintenance, signal processing