

ОБРАБОТКА БОЛЬШИХ ДАННЫХ В ПОТОКОВОМ РЕЖИМЕ ДЛЯ ПРОГНОЗИРОВАНИЯ НЕИСПРАВНОСТЕЙ В МАШИНОСТРОИТЕЛЬНОМ ПРОИЗВОДСТВЕ

А.М. Павлов

andrey-mp@yandex.ru

SPIN-код: 6647-9400

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация

Аннотация

В современном промышленном сценарии обработка больших данных играет ведущую роль в повышении эффективности бизнеса — снижении издержек при обслуживании оборудования и повышении производительности. Все большее число станков оснащается интеллектуальными устройствами (такими как датчики и исполнительные механизмы), которые отвечают за мониторинг состояния станка в режиме реального времени и осуществляют корректирующие действия до того, как качество заготовки будет снижено или станок будет поврежден. Однако многие производственные компании все еще не используют преимущества больших данных, поступающих из производственных систем. В некоторых случаях аналитика больших данных является неизученной проблемой, поскольку считается, что она требует времени и ресурсов. Более того, реальные преимущества обработки промышленных данных в реальном времени обычно недооцениваются. Статья посвящена процессу производства компонентов, а также описанию основных разработок и уроков, извлеченных при настройке аналитической платформы больших данных для обработки и анализа данных с датчиков станков с числовым программным управлением.

Ключевые слова

Большие данные, прогнозирование неисправностей, предиктивное обслуживание оборудования, машиностроительное производство, повышение производительности, интеллектуальное управление

Поступила в редакцию 01.02.2021

© МГТУ им. Н.Э. Баумана, 2021

Введение. В настоящее время в рамках машиностроительного производства генерируется огромный объем разнообразных данных. Хотя объем данных, связанных с промышленными процессами и продуктами, растет в геометрической прогрессии, многие производства в настоящее время используют лишь небольшую их часть — во многих случаях обработка и анализ данных в реальном времени не проводятся. В таком промышленном контексте, основанном на данных, правильно примененная аналитика обещает значительно улучшить производительность по всей цепочке поставок. В частности, использование производственных данных в реальном времени означает своевременное профилактическое обслуживание, способное улучшить качество, эффективность и результа-

тивность производственных операций, а также повлиять на производительность и прибыльность всего производственного процесса.

В этом контексте появляются проекты, направленные на преодоление некоторых серьезных препятствий, которые до сих пор не позволяли многим компаниям воспользоваться всеми преимуществами аналитики больших данных (англ. *big data analytics* — BDA). Многие компании и организации осознали потенциальное конкурентное преимущество, которое они могут получить благодаря своевременной и точной аналитике больших данных, но им не хватает IT-опыта и бюджета для полноценного использования BDA. Чтобы преодолеть это препятствие, компании рассматривают уже готовые проекты, в которых применяется подход на основе модели BDA-as-a-Service (с англ. «база данных как сервис», MBDAaaS), предоставляя удобную и настраиваемую платформу, способную управлять всем процессом анализа больших данных и его артефактами.

Обзор предыдущих научных исследований. По оценкам последних исследований, затраты на незапланированные простои в обрабатывающей промышленности составляют 50 млрд долл. в год [1]. Использование профилактического обслуживания и обнаружения отказов может помочь сектору снизить затраты и улучшить конкурентные преимущества компаний.

Прогностическое обслуживание (англ. *predictive maintenance*, PdM) в настоящее время обеспечивается интеллектуальными технологиями, обычно встроенными в физические активы [2, 3]. В частности, используя огромный объем данных, производимых этими технологиями, и применяя правильную аналитику больших данных, можно прогнозировать сбои машин с очевидными преимуществами с точки зрения эффективности производства деталей и ненужной минимизации времени простоя. Кроме того, раннее обнаружение отказов широко применяется в отрасли, поскольку оно потенциально позволяет устранить катастрофические отказы машин [4].

Многие методы обнаружения аномалий уже были разработаны в области машинного обучения и статистики с применением контролируемых, частично контролируемых и неконтролируемых методов [5]. Хотя большинство из них были предназначены для общих областей приложений, недавние исследования также были сосредоточены на методах, разработанных для конкретных из них. В частности, мало внимания уделяется проблеме обнаружения производственных аномалий [6]. Сегодня для выявления аномального поведения подсистемы широко используются несколько алгоритмов, таких как метод главных компонент (англ. *Principal component analysis*, PCA), К-средние, С-средние, иерархическая кластеризация, нейронные сети, гауссовская смешанная модель (англ. *Generalized method of moments*, GMM) и модифицированная кластеризация порядка ранжирования (англ. *Modified ranking order clustering*, MROC) [5]. Пожалуй, основной проблемой остается выбор алгоритма, наиболее полно соответствующего промышленному контексту и бизнес-целям, а также подходящего подхода к его реализации с учетом уже существующей технологической инфраструктуры.

Исследования и промышленный контекст. Для проектов на основе модели BDA-as-a-Service рассматривают четыре пилотных направления, по которым разрабатывают и тестируют все методологические и технологические решения проектов. Для данного исследования были определены три различных варианта использования:

1) осуществляется прогнозирующая аналитика по размерам деталей в процессе производства, которая касается анализа рабочих циклов в различных производственных ячейках для прогнозирования потенциальных аномалий продукта;

2) анализируются первопричины уведомлений о качестве на этапе проверки, касающиеся расследования на этапе проверки первопричины, которая породила аномалию в продукте;

3) прогнозируются неисправности, возникающие при анализе данных, собранных как в ходе рабочего цикла в производственных ячейках, так и в ходе инспекционных мероприятий, с целью прогнозирования неисправностей оборудования до того, как качество продукта окажется под угрозой.

Основное внимание в этой статье уделено варианту использования прогнозирования сбоев, который потребовал более индивидуальной и глубокой разработки.

Цель разработки сценария использования прогнозирования неисправностей — обнаружение и прогнозирование возможных неисправностей станков, которые могут возникнуть во время рабочего цикла. В частности, цель состоит в том, чтобы определить пороговые значения, связанные с различными сбоями, которые возникают на станках. Для достижения этой цели необходимо отслеживать состояние работоспособности станка, собирая рабочие данные станков, участвующих в рабочем цикле и определяя диапазон значений, на основе которых может быть выдано предупреждение. В этом случае анализ данных осуществляется почти в реальном времени, и набор данных постоянно обновляется — в него добавляются данные, полученные в течение рабочего цикла.

Первый шаг к развертыванию аналитики обнаружения неисправностей на платформе MBDAaaS состоит в определении периметра тематического исследования. Исходя из потребностей пилотного бизнеса было выбрано производство конических зубчатых колес. Ячейка с конической зубчатой передачей представляет собой две подъячейки обработки в последовательном порядке (одна до и другая после термообработки деталей) и подъячейку для окончательной проверки произведенных компонентов. В пилотном проекте рассматривали три станка ячейки. Станки подключены к сети через датчики, которые собирают рабочие данные станка, такие как температуру, угловую скорость, скорость шпинделя и т. д.

Современные станки с числовым программным управлением (ЧПУ), такие как устройства с конической зубчатой передачей, способны распознавать зара-

нее определенный набор состояний отказа и снабжены сервомеханизмами, которые заставляют станок выдавать диагностический код на экране контроллера, корректировать его поведение или даже полностью останавливать производственный процесс всякий раз, когда обнаруживается состояние отказа.

Цель состоит в том, чтобы распознать модель поведения станка, которая приводит к каждому отказу, и своевременно сообщить об этом пользователям, чтобы они приняли решения, которые предотвратят снижение качества. Основные проблемы, которые необходимо решить, включают: изучение модели классификации данных, правильное определение неисправностей и обнаружение подходящей модели для прогнозирования. В этом случае компания готова выйти за рамки предварительно установленных производителем состояний отказа ЧПУ. В частности, он идентифицирует неизвестные диапазоны значений параметров станка, что позволяет обнаруживать особые состояния отказа, которые не были заранее определены производителем, но могут повредить станок и также повлиять на качество заготовки. Распознавание модели поведения (с точки зрения значений параметров), которая предшествует каждому отдельному неисправному состоянию, закрывает вариант использования, позволяющий прогнозировать.

Таким образом, вариант использования состоит из следующих шагов:

- обнаружение и прогнозирование новых неисправностей;
- описание и прогнозирование непредопределенных неисправностей.

Примечательно, что обнаружение настроенных сбоев может быть основано на порогах, определенных техническими специалистами, как следствие того факта, что алгоритм аналитики для обнаружения сбоев часто решает описательную задачу на основе набора данных, предоставленного инженером предметной области, который направлен на построение диагностической модели для обнаружения неисправности.

Реализованное приложение. Следуя модельно-ориентированному подходу к разработке программного обеспечения (англ. *Model driven architecture*, MDA) для разработки инфраструктуры MBDAaaS, предоставляются три модели (CIM, PIM и PSD), представляющие вариант использования прогнозирования сбоев, как также описано в [7].

На первом шаге генерируется json-файл, содержащий спецификации модели CIM (декларативной) с точки зрения целей с индикаторами и задачами (рис. 1). Например, аналитика для обнаружения поведенческой модели станка (неконтролируемая кластеризация машинного обучения) была определена тремя целями: G1 (задача, кластеризация), G2 (модели, описательные), G3 (подход к обучению, неконтролируемый).

Второй шаг касается определения PIM (процедурной модели или платформенно-независимой модели), описывающей в рабочем процессе OWL-S, как аналитика должна выполняться и распараллеливаться в соответствии с требованиями и ограничениями, указанными в декларативной модели.

```

"@type": "tdm:Area",
"tdm:label": "Data Analytics", "tdm:incorporates": [
  [...]
  {
    "@type": "tdm:Indicator", [...]
    {
      "@type": "tdm:Objective",
      "tdm:constraint": "{}",
      "tdm:label": "Clustering"
    }
  }
  [...]
  {
    "@type": "tdm:Indicator",
    "tdm:label": "Learning Approach",
    "tdm:visualisationType": "Option",
    "tdm:incorporates": [{
      "@type": "tdm:Objective",
      "tdm:constraint": "{}",
      "tdm:label": "Unsupervised"
    }]
  }
  [...]

```

Рис. 1. Фрагмент декларативной модели в формате JSON для процесса прогнозирования неисправностей

На третьем шаге используется PSM (платформенно-зависимая модель), т. е. модель развертывания, которая сопоставляет каждую операцию и инструмент с конкретными услугами. Модель развертывания была разработана на основе локальной инфраструктуры компании для обработки данных, оркестрации, аналитики и визуализации. Преобразование от процедурной модели к модели развертывания основано на компиляторе, который принимает в качестве входных данных файл рабочего процесса OWL-S и информацию о целевой платформе и создает на выходе технологически зависимый рабочий процесс.

Выбор языка OWL был обусловлен тем, что он поддерживает автоматическое обнаружение, вызов, композицию и взаимодействие веб-служб. Онтология OWL-S, напротив, структурирована в трех взаимосвязанных субонтологиях: профиль, модель процесса и основание. Онтология профиля используется для выражения того, что «делает» сервис; модель процесса описывает, «как это работает»; конечная схема отображает конструкции модели процесса на подробные спецификации форматов сообщений, протоколов и т. д. Кроме того, чтобы указать службу с помощью выбранной методологии, OWL-S был расширен для определения сопоставления с целями декларативной модели, используемой на этапе выбора службы. Таким образом, предлагаемая модель предполагает построение сложных сервисов как композицию внутри модели процесса.

Архитектура. Экземпляр платформы был настроен в облаке Amazon Web Services (AWS) путем установки диспетчера кластеров Kubernetes — обоснованность применения облачных вычислительных ресурсов была показана в работах [8, 9]. Платформа основана на развертывании единиц микросервисов в кластерной инфраструктуре. Kubernetes — это оркестратор кластера, который позволяет службам на основе Docker совместно использовать ресурсы кластера и работать в распределенной среде. Инфраструктура кластера включает три рабочие виртуальные машины и одну виртуальную машину с ролью мастера со следующими компонентами:

- экземпляры сервисов (Pod в терминологии Kubernetes) на одной машине, таких как Redis, MySQL, Spring Dataflow server, Service Catalog;
- экземпляры масштабируемых реплик (Deployment в терминологии Kubernetes) без сохранения состояния, таких как Spring Cloud Dataflow (SCDF), Spark, HDFS, Kafka, Nexus Server и Viz;
- экземпляры сервисов с отслеживанием состояния (Stateful Set в терминологии Kubernetes), где требуется уникальный сетевой идентификатор и стабильное хранилище, например узлы Cassandra или Spark/HDFS.

Spring Cloud Dataflow (SCDF) — это библиотека, которая позволяет пользователям определять и запускать поток данных приложения на шине сообщений (Kafka). Платформа предоставляет API для создания потоков данных между службами пакетной и потоковой аналитики. Каталог сервисов, развернутый как Kubernetes Pod, состоит из Catalog-API (интерфейс REST для каталога сервисов) и базы данных MySQL для хранения определений сервисов. Сервер Nexus использовался в качестве менеджера реестра для размещения образов Docker (опциональный компонент). Также была развернута платформа визуализации (Viz) для представления графических результатов аналитики.

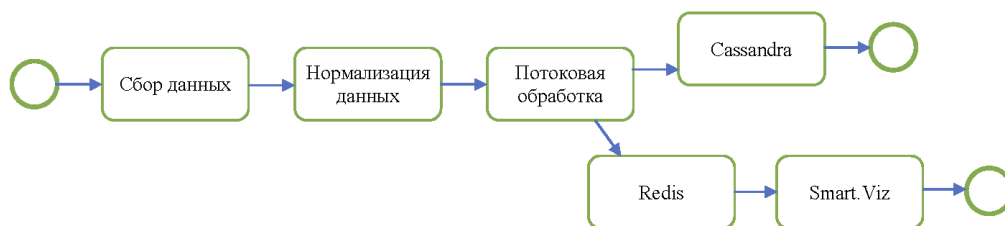


Рис. 2. Конвейер обработки данных прогноза неисправностей

Конвейер обработки данных. В соответствии с целями методологии [10], конвейер обработки данных прогнозирования неисправностей представляет собой композицию из ряда простых «атомарных процессов». Атомарный процесс — это описание службы, которая ожидает одно сообщение и возвращает одно сообщение в ответ. Составной процесс — это процесс, который поддерживает некоторое состояние и может иметь несколько различных целей: получение или преобразование данных или знаний; изменение состояния объектов или сервисов в сервисной среде; и, наконец, применение изменений для реаль-

ных объектов. Сервис может генерировать и возвращать некоторую новую информацию на основе ввода и состояния мира. Другими словами, производство информации описывается входами и выходами процесса. Компилятор извлекает те службы, которые совместимы с целями и ограничениями, указанными в декларативной модели. В частности, компилятор получает свою композицию OWL-S, сериализованную в формат RDF/XML, автоматически анализирует эту спецификацию и таким образом генерирует новый файл, определяющий рабочий процесс на основе языка, принятого SCDF.

Конвейер обработки данных прогнозирования неисправностей на основе шести сервисов показан на рис. 2.

Первый сервис рабочего процесса — это сервис по сбору данных, который принято называть Machine Connector. Он считывает поток данных датчиков, которые измеряют различные параметры производственного процесса, такие как скорость шпинделя, нагрузка и мощность двигателя. Набор данных также включает метаданные о станке и данные измерений. Метаданные содержат информацию о типе станка и количестве датчиков на нем. Каждое измерение имеет метку времени и может быть разного типа: числом с плавающей точкой, целым числом, строкой, временным типом. Выдержки из списка значений, передаваемых конкретным датчиком, приведены в таблице.

Выдержка из списка значений датчика

| Дата | Значение | Дата | Значение |
|---------------------|----------|---------------------|----------|
| 2018-05-29 12:41:18 | 56,263 | 2018-05-29 12:42:21 | 0,037 |
| 2018-05-29 12:41:25 | 56,263 | 2018-05-29 12:42:28 | 0,037 |
| 2018-05-29 12:41:33 | 56,263 | 2018-05-29 12:42:36 | -0,008 |
| 2018-05-29 12:41:41 | 56,263 | 2018-05-29 12:42:44 | -0,008 |
| 2018-05-29 12:41:49 | 56,263 | 2018-05-29 12:42:52 | -0,008 |
| 2018-05-29 12:41:57 | 0,037 | 2018-05-29 12:43:00 | -0,038 |
| 2018-05-29 12:42:05 | 0,037 | 2018-05-29 12:43:08 | -0,038 |
| 2018-05-29 12:42:13 | 0,037 | 2018-05-29 12:43:16 | -0,038 |

Обработка данных для получения выгоды для бизнеса в режиме реального времени имеет решающее значение. Для этого могут потребоваться модели машинного обучения или расчет простых операций, в любом случае это должен быть быстрый и масштабируемый процесс для обеспечения доступности системы. На первом этапе элементы группируются с помощью службы Streaming K-Means, это потоковое выполнение алгоритма K-средних. Эта служба должна поддерживать K центров кластеров, где число K определяется в метаданных для каждого программного статуса каждого отдельного станка. Все элементы данных от датчиков используются для вычисления центроидов кластера. Если последовательность из последних N точек данных имеет объединенную вероятность перехода ниже порогового значения, возникает предупреждение о станке. При обнаружении аномалии первая отметка времени среди последних N точек

данных выводится как предупреждение об обнаружении аномалии в этот момент времени.

Среди всех доступных сервисов хранения данных в реальном времени был выбран сервис Cassandra, который позволяет принимать временные данные в реальном времени и поддерживать эти записи с длительным периодом хранения; в то время как сервис Redis был выбран для достижения чистого приема записей «ключ-значение» в память и для поддержки аналитического отчета, выполняемого Smart.viz. В частности, процесс визуализации, делегированный Smart.viz, приводит к созданию диаграмм в пять этапов: выбор источника данных и базы данных; запрос к базе данных, спецификация модели данных; отображение; настройка и визуализация графиков.

Сервис Redis использовали в качестве источника данных в памяти, поскольку он предлагает время отклика в миллисекундах, что позволяет создавать приложения, которые обрабатывают миллионы запросов в секунду в реальном времени. Экземпляр Redis выбран для хранения всех объемов данных, связанных с результатами анализа в процессе прогнозирования сбоев.

Заключение. Целью исследования была реализация в сложном промышленном контексте аналитики больших данных, способной обрабатывать потоковые данные от станков с ЧПУ, с целью визуализации в реальном времени рабочего состояния станка и прогнозирования возможных неисправности станка. В частности, реализация была реализована по трехэтапной методике, предложенной технологией MBDAaaS. Ниже приводится краткий список основных трудностей, с которыми пришлось столкнуться:

- настройка аналитики в соответствии с модельным подходом BDA-as-a-service (MBDAaaS);
- настройка и установка платформы в сети компании в связи с конфиденциальностью управляемых данных;
- разработка и интеграция сервисов аналитики больших данных на платформе.

При решении этих задач исследование промышленной области с точки зрения технологических активов и физических процессов оказалось необходимым для создания аналитической системы, способной обрабатывать промышленные данные в режиме реального времени. В качестве доказательства разработанная система требовала знаний в следующих областях:

- датчики, установленные на рабочих станках, выдающие потоковые данные;
- промышленные протоколы связи (такие как OPC-UA), отвечающие за отправку таких данных от датчиков на серверы;
- приложения с поддержкой OPC, настроенные для использования данных датчиков с помощью аналитики больших данных.

Более того, еще одним важным моментом для успешного внедрения аналитики больших данных в исследуемом контексте стало активное сотрудничество с партнерами проекта. Все взаимодействия имели решающее значение для определения и спецификации архитектуры платформы с целью выявления технических и функциональных требований.

Архитектура платформы была определена для предоставления аналитики больших данных как услуги путем интеграции и использования наборов инструментов и сервисов для управления и анализа ценности больших данных в нескольких областях. На основе архитектурных моделей сервисы и наборы инструментов были интегрированы в эту платформу, чтобы дать пользователям возможность разрабатывать процессы обработки данных и аналитики, а также настраивать спецификации и сервисы. Целенаправленный подход, при котором менеджеры могут фиксировать этапы бизнес-процесса с помощью целей и подцелей, наблюдать за выполнением процессов, выявлять аномалии и предпринимать соответствующие действия, оказался полезным. В частности, платформа позволила:

- организовать сбор и интеграцию больших данных, связанных с производственными процессами;
- применять методы прогнозной аналитики для оптимизации производительности оборудования;
- сообщать сотрудникам компании о результатах аналитики для оптимизации процесса принятия решений.

Технические требования и требования к производительности были собраны для развертывания системы применительно к производственному процессу, способному производить данные в (почти) реальном времени. В сотрудничестве с другими экспертами проекта были выбраны правильные алгоритмы, методы анализа данных и отчеты по данным для управления сложным производственным поведением. В частности, было обнаружено, что необходимо определить соответствующие технологии для восстановления данных с датчиков в течение рабочего цикла оборудования.

Еще один урок, полученный на этапе разработки, касается постоянно растущего объема данных датчиков, которыми необходимо управлять с помощью технологий, отличных от традиционных. В частности, традиционное моделирование данных было отброшено, поскольку регистратор неприменим в сценарии машиностроительной промышленности.

Выводы. В статье описаны разработки и извлеченные уроки, основанные на применении методологии MBDAaaS при создании ведущих решений для больших данных в предметной области. Предлагаемая методология базируется на модельном инженерном подходе, который основан на специализированной архитектуре, управляемой данными, с семантическими свойствами для обработки большого количества этих данных. Это позволяет упростить цепочку компетенций, необходимых для управления всем конвейером больших данных и поддержки задачи аналитики.

Среди трех возможных вариантов использования (например, прогнозирующая аналитика размеров заготовок в процессе производства, анализ первопричин уведомлений о качестве на этапе проверки и прогнозирование неис-

правностей) статья фокусируется на прогнозировании неисправностей для анализа данных, собранных в ячейке конического зубчатого колеса, где имеется ряд современных станков с ЧПУ. Они способны распознавать заранее определенный набор состояний отказа и корректировать свое поведение. Разработанное решение нацелено на то, чтобы выйти за рамки предварительно установленных производителем ЧПУ состояний отказа.

Технические вопросы и извлеченные уроки позволят вести дальнейшие разработки и применять аналогичные приложения в других сложных промышленных секторах.

Внедрение этого пилотного проекта и взаимодействие с мощным инструментом MBDAaaS принесло несколько улучшений, таких как реализация сбора данных и потоковой передачи в реальном времени с набора станков с ЧПУ путем применения различных агентов для стандартизации потока в соответствии с протоколом связи OPC-UA и независимое описание отдельных атомарных услуг в реализации варианта использования, что позволяет легко использовать их в новых различных сценариях.

Дальнейшие исследования будут посвящены применению и тестированию на пилотном сайте, а также другим аналитическим услугам и услугам для управления большими данными, поступающими со станков с ЧПУ.

Литература

- [1] Coleman C., Chandramouli M., Damodaran S., et al. Making maintenance smarter. *deloitte.com: веб-сайт*. URL: <https://www2.deloitte.com/us/en/insights/focus/industry-4-0/using-predictive-technologies-for-asset-maintenance.html> (дата обращения: 15.12.2020).
- [2] Власов А.И., Григорьев П.В., Кривошеин А.И. Модель предиктивного обслуживания оборудования с применением беспроводных сенсорных сетей. *Надежность и качество сложных систем*, 2018, № 2, с. 26–35. DOI: <https://doi.org/10.21685/2307-4205-2018-2-4>
- [3] Amruthnath N., Gupta T. A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. *5th ICIEA*, 2018, pp. 355–361. DOI: <https://doi.org/10.1109/IEA.2018.8387124>
- [4] Ермаков В.А., Корзун Д.Ж. Цифровые сервисы предиктивной аналитики при обслуживании производственного оборудования. *Цифровые технологии в образовании, науке, обществе*, 2019, с. 73–75.
- [5] Chandola V., Banerjee A., Kumar V. Anomaly detection: a survey. *ACM Comput. Surv.*, 2009, vol. 41, no. 3, art. 15. DOI: <https://doi.org/10.1145/1541880.1541882>
- [6] Harding J.A., Shahbaz M., Kusiak A. Data mining in manufacturing: a review. *J. Manuf. Sci. Eng.*, 2006, vol. 128, no. 4, pp. 969–976. DOI: <https://doi.org/10.1115/1.2194554>
- [7] Crespino A.M., Di Biccari C., Lazoi M., et al. Fault prediction in aerospace product manufacturing: a model-based big data analytics service. In: *Enterprise Interoperability*. Wiley, 2018, pp. 193–200. DOI: <https://doi.org/10.1002/9781119564034>
- [8] Wan J., Tang S., Li D., et al. A manufacturing big data solution for active preventive maintenance. *IEEE Trans. Ind. Informat.*, 2017, vol. 13, no. 4, pp. 2039–2047. DOI: <https://doi.org/10.1109/TII.2017.2670505>

- [9] Yu W., Dillon T., Mostafa F., et al. A global manufacturing big data ecosystem for fault detection in predictive maintenance. *IEEE Trans. Ind. Informat.*, 2020, vol. 16, no. 1, pp. 183–192. DOI: <https://doi.org/10.1109/TII.2019.2915846>
- [10] Ardagna C.A., Bellandi V., Ceravolo P., et al. A model-driven methodology for big data analytics-as-a-service. *IEEE BigData Congress*, 2017, pp. 105–112. DOI: <https://doi.org/10.1109/BigDataCongress.2017.23>

Павлов Андрей Михайлович — аспирант кафедры «Промышленная логистика», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Научный руководитель — Александров Александр Анатольевич, кандидат технических наук, доцент кафедры «Промышленная логистика», МГТУ им. Н.Э. Баумана, Москва, Российская Федерация.

Ссылку на эту статью просим оформлять следующим образом:

Павлов А.М. Обработка больших данных в потоковом режиме для прогнозирования неисправностей в машиностроительном производстве. *Политехнический молодежный журнал*, 2021, № 03(56). <http://dx.doi.org/10.18698/2541-8009-2021-03-679>

STREAMING BIG DATA FOR FAILURE PREDICTION IN MECHANICAL ENGINEERING

A.M. Pavlov

andrey-mp@yandex.ru
SPIN-code: 6647-9400

Bauman Moscow State Technical University, Moscow, Russian Federation

Abstract

In today's industrial scenario, big data processing plays a leading role in improving business efficiency - reducing equipment maintenance costs and increasing productivity. An increasing number of machine tools are equipped with intelligent devices (such as sensors and actuators) that monitor the condition of the machine in real time and take corrective action before the quality of the workpiece is reduced or the machine is damaged. However, many manufacturing companies are still not taking advantage of big data coming from manufacturing systems. In some cases, big data analytics is an unexplored problem because it is believed to take time and resources. Moreover, the real benefits of real-time industrial data processing are usually underestimated. The article focuses on the component manufacturing process, as well as a description of the main developments and lessons learned when setting up a big data analytical platform for processing and analyzing data from sensors of numerically controlled machines.

Keywords

Big data, fault prediction, predictive equipment maintenance, mechanical engineering, productivity improvement, intelligent management

Received 01.02.2021

© Bauman Moscow State Technical
University, 2021**References**

- [1] Coleman C., Chandramouli M., Damodaran S., et al. Making maintenance smarter. *deloitte.com: website*. URL: <https://www2.deloitte.com/us/en/insights/focus/industry-4-0/using-predictive-technologies-for-asset-maintenance.html> (accessed: 15.12.2020).
- [2] Vlasov A.I., Grigor'yev P.V., Krivoshein A.I. Model of predictive equipment maintenance with application of wireless touch networks. *Nadezhnost' i kachestvo slozhnykh system* [Reliability & Quality of Complex Systems], 2018, no. 2, pp. 26–35. DOI: <https://doi.org/10.21685/2307-4205-2018-2-4> (in Russ.).
- [3] Amruthnath N., Gupta T. A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance. *5th ICIEA*, 2018, pp. 355–361. DOI: <https://doi.org/10.1109/IEA.2018.8387124>
- [4] Ermakov V.A., Korzun D.Zh. [Digital services of predictive analytics at maintenance of production equipment]. *Tsifrovyye tekhnologii v obrazovanii, nauke, obshchestve* [Digital Technologies in Education, Science, Society], 2019, pp. 73–75 (in Russ.).
- [5] Chandola V., Banerjee A., Kumar V. Anomaly detection: a survey. *ACM Comput. Surv.*, 2009, vol. 41, no. 3, art. 15. DOI: <https://doi.org/10.1145/1541880.1541882>
- [6] Harding J.A., Shahbaz M., Kusiak A. Data mining in manufacturing: a review. *J. Manuf. Sci. Eng.*, 2006, vol. 128, no. 4, pp. 969–976. DOI: <https://doi.org/10.1115/1.2194554>

- [7] Crespino A.M., Di Biccari C., Lazoi M., et al. Fault prediction in aerospace product manufacturing: a model-based big data analytics service. In: Enterprise Interoperability. Wiley, 2018, pp. 193–200. DOI: <https://doi.org/10.1002/9781119564034>
- [8] Wan J., Tang S., Li D., et al. A manufacturing big data solution for active preventive maintenance. *IEEE Trans. Ind. Informat.*, 2017, vol. 13, no. 4, pp. 2039–2047. DOI: <https://doi.org/10.1109/TII.2017.2670505>
- [9] Yu W., Dillon T., Mostafa F., et al. A global manufacturing big data ecosystem for fault detection in predictive maintenance. *IEEE Trans. Ind. Informat.*, 2020, vol. 16, no. 1, pp. 183–192. DOI: <https://doi.org/10.1109/TII.2019.2915846>
- [10] Ardagna C.A., Bellandi V., Ceravolo P., et al. A model-driven methodology for big data analytics-as-a-service. *IEEE BigData Congress*, 2017, pp. 105–112. DOI: <https://doi.org/10.1109/BigDataCongress.2017.23>

Pavlov A.M. — PhD Student, Department of Industrial Logistics, Bauman Moscow State Technical University, Moscow, Russian Federation.

Scientific advisor — Aleksandrov A.A., Cand. Sc. (Eng.), Assoc. Professor, Department of Industrial Logistics, Bauman Moscow State Technical University, Moscow, Russian Federation.

Please cite this article in English as:

Pavlov A.M. Streaming big data for failure prediction in mechanical engineering. *Politekhnikeskiiy molodezhnyy zhurnal* [Politechnical student journal], 2021, no. 03(56). <http://dx.doi.org/10.18698/2541-8009-2021-03-679.html> (in Russ.).