

Применение интеллектуальных методов анализа данных к задаче прогнозирования результатов промышленного испытания элементов конструкции по данным тензометрии

Е.Е.Истратова, А.Н. Кожевников, П.В. Ласточкин, Е.В. Глинин

Аннотация — В статье представлены результаты исследования трех интеллектуальных методов анализа данных для решения задачи прогнозирования результатов промышленного испытания элементов конструкции по данным тензометрии. В качестве объекта исследования была рассмотрена двутавровая балка, для которой была проведена серия экспериментов с нагружением и записью значений ее напряженно-деформированного состояния. Полученные экспериментальные значения были использованы для исследования и прогнозирования прочности в режиме реального времени. Научная новизна работы заключается в предложении как авторской методики прогнозирования напряженно-деформированного состояния балки в зависимости от параметров окружающей среды, так и инструмента, основанного на применении методов интеллектуального анализа данных. В результате работы была реализована информационная система, отличительной особенностью которой является возможность оперативной обработки больших объемов экспериментальных, то есть сырых данных в режиме реального времени, что достигается за счет применения методов интеллектуального анализа данных. В ходе работы были выявлены и проанализированы параметры окружающей среды как со стороны стенда, так и со стороны микроклимата, оказывающие влияние на точность измерений. Благодаря этому, экспериментальные данные могут быть скорректированы в любой момент времени при изменении тех или иных характеристик.

Ключевые слова — тензометрия, напряженно-деформированное состояние, двутавровая балка, интеллектуальный анализ данных, линейная регрессия, деревья решений, метод опорных векторов.

I. ВВЕДЕНИЕ

Промышленные предприятия получают большое количество данных, которые в большинстве случаев неструктурированные. Такие данные требуют обработки для извлечения полезной информации.

Истратова Евгения Евгеньевна, Новосибирский государственный технический университет, istratova@mail.ru

Кожевников Алексей Николаевич, Новосибирский государственный технический университет, ak-1993@bk.ru

Ласточкин Павел Валерьевич, Новосибирский государственный технический университет, l.pasha.v@gmail.com

Глинин Евгений Вадимович, Новосибирский государственный технический университет, eugeny.glinin@yandex.ru

Данные могут различаться по типу, в связи с этим возникает необходимость использовать интеллектуальный анализ данных. Интеллектуальный анализ данных (Data Mining) — это область, в которой используются информационные технологии для анализа структурированных и неструктурированных данных, которые могут быть собраны из различных источников, включая электронные таблицы, базы данных, тексты, изображения, видео, аудио, геоданные.

В рамках работы были рассмотрены несколько предметных областей, из которых преобладающей является процесс изучения и анализа промышленных испытаний двутавровой балки. При этом в качестве объекта исследования была рассмотрена двутавровая балка с длиной рабочей поверхности в 3,2 метра с профилем 16Б2, регламентируемым требованиями ГОСТ 26020-83.

Предметом исследования являлось прогнозирование прочности балки на основе получаемых экспериментальных данных.

Исходя из этого цель исследования заключалась в прогнозе того, при какой нагрузке и в какой точке в балке из определенного материала произойдет разрушение.

Для реализации указанной цели в рамках исследования были решены задачи по изучению предметной области; разработке алгоритма программного обеспечения для прогнозирования результатов промышленных испытаний двутавровой балки; разработке программного обеспечения на основе спроектированного алгоритма и исследованию готового программного продукта.

Результатом данной работы является программа для прогноза результатов промышленного испытания элементов конструкции по данным тензометрии.

Научная новизна исследования заключается в применении методов интеллектуального анализа данных для решения задачи прогнозирования результатов промышленных прочностных испытаний элементов конструкции.

Практическая значимость исследования напрямую связана с определением оптимальных параметров при проведении натурных испытаний балки для повышения ее прочностных характеристик, что в итоге способствует сокращению расходов на производство

конструкции и увеличению периода ее эксплуатации. Так как полученные в результате реальных промышленных испытаний данные позволяют отслеживать в режиме реального времени динамику изменения напряжений в элементах металлических конструкций, это позволяет предотвратить случаи нарушения целостности конструкции, а также оценить ее состояние до момента начала ее разрушения.

II. ХАРАКТЕРИСТИКА ОБЪЕКТА ИССЛЕДОВАНИЯ

В качестве объекта исследования в работе выступала двутавровая балка, имеющая 6 отверстий. Длина балки составляла 3,2 метра. Объект был изготовлен из различных материалов, среди которых была сталь марки Ст3сп5, а также образцы, в которых сталь была заменена на композитные материалы в процентном соотношении к исходному материалу.

Расположенные на поверхности балки, отверстия были разделены по функциональному принципу на две категории. К первой относилось отверстие, расположенное в левой части балки под срезанной полкой, и являющееся точкой приложения нагрузки к объекту. С помощью данного отверстия происходило крепление к балке гидропривода для нагружения, осуществлявшееся при помощи болтового соединения. Ко второй категории отверстий относились 5 отверстий, расположенных на рабочей поверхности балки. Их функциональное назначение заключалось в концентрации нагрузки, именно данная зона объекта является потенциально опасной с точки зрения разрушения. Геометрическая характеристика объекта исследования с указанием расположенных на нем отверстий приведена на рис. 1.

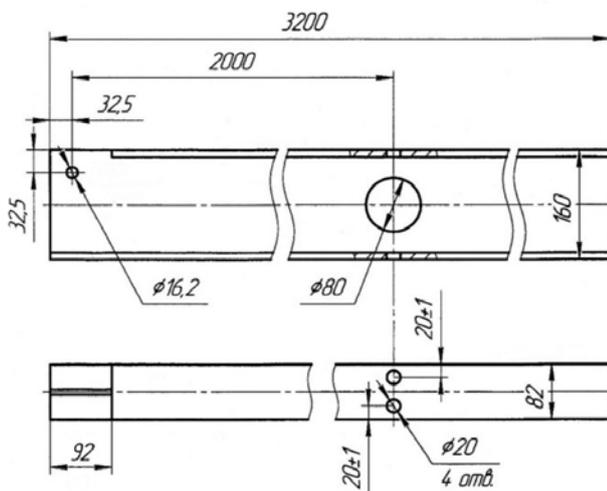


Рисунок 1 – Геометрическая характеристика балки

Для определения напряженно-деформированного состояния объекта исследования были использованы тензометрические датчики, установленные в зоне концентрации напряжений. Всего был установлен 21 тензометрический датчик. Из них по 3 датчика было установлено у каждого отверстия в полках, 6 датчиков были размещены вокруг отверстия в стенке балки, и по

1 датчику было расположено в каждой полке объекта исследования. С помощью тензодатчиков были зафиксированы и переданы для обработки значения напряженного-деформированного состояния балки под действием нагрузки.

III. СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

Интеллектуальный анализ данных используется для обнаружения закономерностей и взаимосвязей данных с учетом больших объемов данных, получаемых при помощи наблюдений [1]. Таким образом, объектом Data Mining является нахождение неизвестных данных характеристик, связей и зависимостей между ними.

Data Mining состоит из следующих действий по обработке данных [2]:

1. Извлечение, преобразование и загрузка транзакционных данных в хранилище системы.
2. Хранение и управление данными в многомерной базе данных системы с учетом распределения прав доступа пользователей к этим информационным ресурсам.
3. Анализ данных с помощью прикладного программного обеспечения посредством представления их в виде графиков или таблиц.

Исходя из этого, назначение методов интеллектуального анализа данных заключается в обработке больших объемов данных для выявления закономерностей и тенденций, которые могут быть полезны в принятии решений.

Помимо назначения также принято выделять цель и отличительную особенность методов интеллектуального анализа данных. Цель технологии Data Mining – нахождение скрытых закономерностей, позволяющих делать точные прогнозы, которые не могут быть найдены обычными методами. Основная особенность технологии Data Mining заключается в интеграции достаточно разностороннего ассортимента математических методов с последними достижениями в сфере информационных технологий [3]. Благодаря этому, данная технология включает в себе формализованные методы и методы неформального анализа, которые объединяют количественную и качественную характеристики объекта исследования.

Большинство аналитических методов, используемые в технологии Data Mining — это известные математические методы и алгоритмы. Новым в их применении является возможность их применения при решении конкретных проблем, обусловленная появившимися возможностями технических и программных средств [4]. Этим обусловлено тем, что большинство методов Data Mining были разработаны в рамках теории искусственного интеллекта.

К ключевым преимуществам методов интеллектуального анализа данных можно отнести следующие [5]:

1. Повышение эффективности за счет автоматизации обработки больших объемов данных. Методы интеллектуального анализа данных могут помочь организациям эффективнее использовать данные,

автоматизируя многие процессы, которые, в противном случае, пришлось бы выполнять вручную.

2. Повышение скорости принятия решений. Методы интеллектуального анализа данных могут помочь организациям принимать более обоснованные и оперативные решения, предоставляя им более полное представление данных.

3. Повышение конкурентоспособности за счет получения новых знаний закономерностей и формулирования идей для улучшения производимых продуктов и оказываемых услуг.

4. Повышение качества обслуживания клиентов. Методы интеллектуального анализа данных могут использоваться для выявления моделей и тенденций в поведении клиентов, что может помочь организациям улучшить обслуживание клиентов и повысить удовлетворенность последних.

Однако методы интеллектуального анализа данных также имеют некоторые недостатки, связанные в том числе со сложностью их внедрения и применения за счет необходимости обучения персонала организации; с высокими требованиями к вычислительным ресурсам предприятия и необходимостью переоснащения его материально-технической базы.

Таким образом, методы интеллектуального анализа данных обладают не только преимуществами, но и имеют недостатки, которые необходимо учитывать. Именно поэтому необходимо проанализировать и тщательно оценить соотношение рисков и выгод перед внедрением методов интеллектуального анализа данных.

Классификация методов интеллектуального анализа данных может быть основана на различных признаках, наиболее распространенными из которых являются следующие [6]: принцип работы с исходными данными; принцип обучения математических моделей; функциональный принцип.

Рассматривая принцип работы с входными информационными потоками, методы можно разделить на две группы, которые связаны как с непосредственным использованием и хранением данных, также и с выявлением формальных закономерностей в шаблонах данных. В первой группе исходные данные хранятся в явном детализированном виде и используются непосредственно на стадиях прогностического моделирования или анализа исключений. В данном случае у этой группы методов есть сложность, которая заключается в анализе сверхбольших объемов данных. К данной группе методов относятся: метод k ближайших соседей, кластерный анализ, рассуждение по аналогии. При технологии дистилляции шаблонов один образец (шаблон) информации извлекается из исходных данных и преобразуется в некие формальные конструкции, вид которых зависит от используемого метода Data Mining. Методы, которые относятся к данной группе: логические методы, методы, основанные на уравнениях, методы визуализации.

Относительно обучения математических моделей методы интеллектуального анализа данных делятся на

статистические, основанные на использовании усредненного накопленного опыта, который отражен в ретроспективных данных, и кибернетические, содержащие ряд разнородных математических подходов.

К статистическим методам технологии Data Mining относятся следующие [7]:

1. Дескриптивный анализ и описание исходных данных.

2. Анализ связей (корреляционный и регрессионный анализ, факторный анализ, дисперсионный анализ).

3. Многомерный статистический анализ (компонентный анализ, дискриминантный анализ, многомерный регрессионный анализ, канонические корреляции и др.).

4. Анализ временных рядов (динамические модели и прогнозирование).

К кибернетическим методам технологии Data Mining относятся следующие:

1. Искусственные нейронные сети (распознавание, кластеризация, прогноз).

2. Эволюционное программирование (алгоритмы метода группового учета аргументов).

3. Генетические алгоритмы (оптимизация).

4. Ассоциативная память (поиск аналогов, прототипов).

5. Нечеткая логика.

6. Деревья решений.

7. Системы обработки экспертных знаний.

Рассматривая классификацию далее, выделяются две группы методов интеллектуального анализа данных на основе функционала, связанных с решением задач сегментации, также связанных с задачами классификации и кластеризации, и задачами прогнозирования.

Первая группа методов составляет инструменты для нахождения шаблонов или образцов, которые описывают данные, поддающиеся интерпретации с точки зрения аналитика. Методы, направленные на получение описательных результатов, включают итеративные методы кластерного анализа, такие как: алгоритм k -средних, k -медианы, методы иерархические кластерного анализа, самоорганизующиеся карты Кохонена, кросс-табличная визуализация, другие методы визуализации.

Методы прогнозирования используют значения показателей, которые были получены в результате предыдущего прогнозирования или измерения. Методы, которые направлены на получение прогнозирующих результатов, включают такие методы, как: нейронные сети, метод ближайшего соседа, линейная регрессия, деревья решений, метод опорных векторов.

Таким образом, для решения задачи прогнозирования результатов промышленных испытаний экспериментальных данных в качестве возможных вариантов были выбраны следующие методы интеллектуального анализа данных: линейная регрессия, деревья решений, метод опорных векторов. Для данных методов были определены критерии сравнения и выполнено сопоставление полученных результатов со

значениями, полученными в ходе натурных испытаний.

В рамках реализации проекта для решения задачи прогнозирования результатов промышленных испытаний экспериментальных данных в качестве исследуемых были выбраны следующие методы интеллектуального анализа данных: линейная регрессия, деревья решений, метод опорных векторов. Для данных методов были определены критерии сравнения и выполнено сопоставление полученных результатов со значениями, полученными в ходе натурных испытаний.

Линейная регрессия – это модель, которую можно использовать для прогнозирования значений переменных в зависимости от значений других переменных. В этом случае исследуемые переменные называются зависимыми переменными, а переменные, которые влияют на эти данные, называются независимыми переменными [8].

Деревья решений – это древовидная структура, подобная блок-схеме, где каждый узел представляет собой проверку значения атрибута, каждая ветвь обозначает результат теста, а листья дерева представляют классы или распределения классов. Точность данного метода высока только для небольшого набора данных, поскольку при работе с большими объемами данных в методе слабо реализовано упрощение некоторых типов логических задач [9].

Метод опорных векторов – это контролируемая стратегия обучения, используемая как для классификации, так и для прогнозирования. Когда выходные данные метода опорных векторов представляют собой непрерывное значение, утверждается, что методология обучения выполняет прогнозирование, и как только методология обучения предсказывает метку категории входного объекта, это называется классификацией [10].

IV. ПРОГНОЗИРОВАНИЕ РЕЗУЛЬТАТОВ ПРОМЫШЛЕННЫХ ИСПЫТАНИЙ НА ОСНОВЕ УРАВНЕНИЯ МНОЖЕСТВЕННОЙ ЛИНЕЙНОЙ РЕГРЕССИИ

В результате исследования параметров микроклимата на точность измерений напряженно-деформированного состояния изделия было получено уравнение регрессии (1), с помощью которого можно определить величину напряженно-деформированного состояния в зависимости от исследуемых параметров.

$$Y=1.2598+0.0639X_1-0.0031X_2+0.0155X_3+0.0163X_4 \quad (1)$$

Оценка значимости уравнения множественной регрессии была выполнена путем проверки гипотезы о равенстве нулю коэффициента детерминации, рассчитанного на основании данных генеральной совокупности [11]. Для ее проверки был использован F-критерий Фишера. При этом вычислили фактическое (наблюдаемое) значение F-критерия, через коэффициент детерминации R^2 , рассчитанный по данным конкретного наблюдения. Чем ближе этот коэффициент к единице, тем больше уравнение регрессии объясняет поведение Y [12]. Добавление в модель новых объясняющих

переменных осуществляется до тех пор, пока растет скорректированный коэффициент детерминации. Коэффициент детерминации равен $R^2 = 0.9966^2 = 0.9931$.

Табличное значение при степенях свободы $k_1=2$ и $k_2=n-m-1=10-4-1=5$, $F_{кр}(4;5)=5.19$. Поскольку фактическое значение $F > F_{кр}$, то коэффициент детерминации статистически значим и уравнение регрессии статистически надежно, а все его коэффициенты совместно значимы.

V. ПРОГНОЗИРОВАНИЕ РЕЗУЛЬТАТОВ ПРОМЫШЛЕННЫХ ИСПЫТАНИЙ НА ОСНОВЕ ПРИМЕНЕНИЯ МЕТОДА «ДЕРЕВЬЯ РЕШЕНИЙ»

Деревья классификации и регрессии – это инструменты поддержки принятия решений, которые применяются в машинном обучении, анализе данных и статистике. Структура подобного дерева определяется двумя понятиями: «листья» и «ветки». «Ветки» дерева решения представляют собой ребра логической конструкции, на которых записаны атрибуты. От «веток» зависит целевая функция. «Листья» фиксируют значения целевой функции, остальные узлы — атрибуты, по которым различаются данные случаи. Таким образом, для того чтобы определить новый вариант, необходимо спуститься по дереву решений до необходимого листа и выбрать соответствующее значение [13,14].

Для исследуемой совокупности данных объектом является значение напряженно-деформированного состояния, которое под действием атрибутов может оставаться в пределах нормы или отклоняться от нее. С физической точки зрения подобное отклонение означает наличие деформации балки.

К ключевым атрибутам дерева решений были отнесены следующие характеристики с соответствующими им значениями:

1. Температура воздуха в помещении (значения: низкая и высокая);
2. Влажность воздуха в помещении (значения: низкая и высокая);
3. Нагрузка на балку (значения: низкая и высокая);
4. Состав материала (значения: сталь и композит).

Исследуя атрибуты объекта пошагово, дерево решений должно дать ответ на вопрос о том, находится ли значение напряженно-деформированного состояния в пределах нормы или нет. Однако, при использовании деревьев решений результат должен быть максимально эффективным и за минимальное возможное количество шагов. Для решения данной задачи на каждом шаге для проверки определяется «наилучший» в текущей ситуации атрибут. Разделение множества на подмножества на основе значений атрибутов продолжается до момента, пока это необходимо, или пока не закончатся атрибуты.

При построении модели дерева решений в исследовании был использован алгоритм CART (Classification and Regression Trees), который предназначается для построения бинарного дерева решений. У бинарного дерева при разбиении у каждого узла существует только два потомка, которые

называются дочерними ветвями. Дальнейшее разделение ветвей зависит от того, на сколько много исходных данных описывает данная ветвь. На каждом этапе построения дерева, правило, которое формируется в узле, разделяет заданное множество примеров на два. На первой итерации строятся все возможные гиперплоскости таким образом, чтобы разбивали исходное пространство на две части. Для каждого такого разбиения пространства подсчитывается количество наблюдений в каждом из подпространств различных классов.

В результате выбирается разбиение, которое максимально выделило в одном из подпространств наблюдения одного из классов. Это разбиение становится корнем дерева решений с двумя разбиениями. Атрибуты, на которых основывается разбиение, выбираются в соответствии с какой-либо мерой, например, индексом Джини, мерой энтропии.

В данном исследовании для дерева решений, которое строится на наборе данных, для оценки оптимальности разбиения использовалась метрика «критерий Джини».

В последующих итерациях разбиение проводится для каждой ветви, пока либо не будут получены узлы, содержащие объекты только одного класса (эти узлы будут объявлены листьями, то есть элементами дерева, которые не имеют разбиений), либо не будет достигнуто заданное значение ошибки в узле дерева (по отношению количества неправильно определенных примеров к общему числу примеров), либо количество наблюдений в узле не окажется меньше заданного минимального количества, либо не будет достигнуто заданное значение глубины дерева, то есть количество последовательных разбиений.

Однако в каждом из случаев может произойти переобучение дерева. Чтобы уменьшить вероятность такого исхода может применяться метод перекрестной проверки, который является оригинальной частью алгоритма CART.

Перекрестная проверка представляет метод выбора окончательного дерева, когда набор данных представляет собой небольшое количество или же записи в наборе данных настолько специфичны, что набор не может быть разделен на обучающую и тестовую выборки.

Дерево, которое построено на основе пула экспериментальных данных (рис. 2), сначала предлагает выполнить разбиение по значению атрибута «Нагрузка», а потом в зависимости от значения объекта по атрибутам «Влажность» или «Температура».

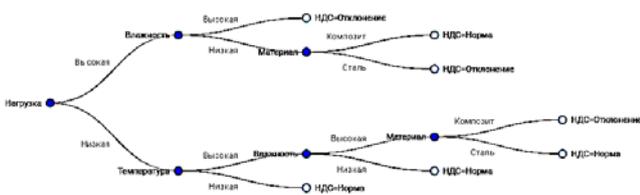


Рисунок 2 – Схема дерева решений

Для проведения компьютерного моделирования,

направленного на выявление наилучшей модели для прогнозирования значений напряженно-деформируемого состояния балки, количество наблюдений составило 1126.

В процессе реализации алгоритма обучения исходное множество было разделено на обучающее подмножество, которое включает 75 % наблюдений, и тестовое подмножество, которое составило 25 % наблюдений.

При прогнозировании с помощью метода «дерева решений» использовалась факторная модель на основе алгоритма CART, где в качестве факторов для прогноза значений напряженно-деформированного состояния балки были использованы параметры микроклимата помещения, состава материала и нагрузки.

На тестовом множестве в качестве оценки результатов моделирования использовалась мера среднеквадратической ошибки, которая характеризует степень близости значений напряженно-деформированного состояния и его прогноза.

Предварительно была проведена предварительная обработка всех значений факторов, так что все значения ряда стали стандартизованными со средним значением, которое равно 0, и дисперсией, равной 1. Это было сделано из-за чувствительности алгоритмов машинного обучения к масштабу, также к размерности исходных значений признаков.

Алгоритм CART, для выявления оптимального разбиения в узле, использует два наиболее значимых параметра: глубина дерева (depth) и число входных переменных (features). Ключевой особенностью разработанной модели является автоматический выбор наилучшего значения одного параметра при заданном, фиксированном значении другого параметра. Так, в описываемом компьютерном эксперименте в качестве фиксируемого параметра выступает параметр features, в зависимости от значения которого выбирается наилучшее значение параметра глубины дерева. Возможными значениями параметра features является множество с дискретными значениями от 1 до 4, где максимальное значение, равное 4, соответствует количеству входных переменных модели. Параметр depth принимает значения от 2 до 7. В соответствии с результатами исследования, минимум среднеквадратической ошибки (3.29 %) был получен при значениях параметра features, равного 3, и параметра depth, равного 6.

VI. ПРОГНОЗИРОВАНИЕ ЗНАЧЕНИЙ НАПРЯЖЕННО-ДЕФОРМИРОВАННОГО СОСТОЯНИЯ БАЛКИ НА ОСНОВЕ МЕТОДА ОПОРНЫХ ВЕКТОРОВ

Помимо линейной регрессии и деревьев решений, задача прогнозирования результатов натуральных испытаний балки была осуществлена с использованием метода опорных векторов. Таким образом, в рамках реализации проекта были исследованы основные возможности метода опорных векторов для решения задачи прогнозирования экспериментальных данных промышленного испытания балки.

Метод опорных векторов (Support Vector Machine

(SVM)) представляет собой аппроксимирующую функцию, применяемую в качестве инструмента минимизации структурного риска [15]. Данный метод основывается на том, что уровень ошибок обучаемой машины на тестовом множестве является суммой ошибок обучения и константы, которая определяется значением измерения Вапника–Червоненкиса.

Исходная модель данных, которая поступает на вход для каждого значения напряженно-деформируемого состояния, подлежит преобразованию следующим образом: для формирования последующего значения ряда исходный набор разбивается на обучающий, который состоит из первые 844 наблюдения (75%), и тестовый набор, состоящий из оставшиеся 282 наблюдения (25%).

Прогнозирование на основе машин опорных векторов было основано на концепции факторного прогнозирования, когда процесс развития динамики целевой переменной описывается набором определенных переменных, взятых в определенных единицах измерения и форматах.

В исследовании, показатель среднеквадратической ошибки, который рассчитывается на основе тестового множества, используется как мера сравнения качества получаемых моделей. Данный показатель максимально точно характеризует степень близости значений изучаемого процесса и его прогноза.

Использование метода опорных векторов требует предварительной обработки исходных данных, которая заключается в их стандартизации [16]. Для проекта были взяты параметры, при которых среднее значение равно 0, а дисперсия принимает значение, равное 1.

При реализации в качестве свободно настраиваемых параметров были приняты: параметр C , который влияет на качество построения поверхности решений за счет ее упрощения; параметр σ , который регулирует возможность учета ошибки; полиномиальная, сигмоидальная и радиально-базисная функции ядра.

Компьютерное моделирование проводилось с использованием пакета MATLAB и программы SVR. В результате эксперимента для разных значений параметров были получены значения среднеквадратической ошибки для модели напряженно-деформируемого состояния балки при разных уровнях параметров C и σ , а также различных функциях ядра. Наилучшие результаты по значению среднеквадратической ошибки для исследуемого показателя демонстрируют модели машин опорных векторов, которые используют в качестве параметра ядра полиномиальную функцию. Среди моделей, которые построены с использованием этой функции, наилучшие результаты при любом значении параметра σ демонстрируют реализации, которые имеют минимально значение C , равное 1. Однако лучшие результаты были получены при значении параметра σ , равного 0.1.

VII. ВЫБОР И ОБОСНОВАНИЕ МЕТОДА ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ДЛЯ ПРОГНОЗИРОВАНИЯ НАТУРНЫХ ИСПЫТАНИЙ ОБЪЕКТА

Каждый метод технологии Data Mining характеризуется различными свойствами, которые могут быть определяющими при выборе конкретного метода анализа данных. Рассматривая основные свойства и характеристики методов Data Mining были выбраны следующие: точность, масштабируемость, трудоемкость, быстрота.

Точность метода интеллектуального анализа данных определяет близость полученных результатов с экспериментальными значениями.

Масштабируемость — это свойство вычислительной системы, которое обеспечивает предсказуемый рост системных характеристик, например, быстроты реакции, общей производительности, при добавлении к ней вычислительных ресурсов.

Трудоемкость метода интеллектуального анализа данных представляет собой временные и материальные затраты на внедрение и развертывание данного метода в условиях промышленности.

Быстрота метода интеллектуального анализа данных напрямую связана с необходимостью обработки данных в режиме реального времени.

В табл. 1 приведена сравнительная характеристика рассмотренных методов. Оценка каждой из характеристик была проведена в соответствии со следующими категориями, расположенными в порядке возрастания: низкая, нейтральная, высокая.

Таблица 1. Результаты сравнительного анализа методов интеллектуального анализа данных

Критерий сравнения	Методы интеллектуального анализа данных		
	Линейная регрессия	Деревья решений	Метод опорных векторов
Точность	нейтральная	низкая	низкая
Масштабируемость	высокая	высокая	низкая
Трудоемкость	нейтральная	высокая	нейтральная
Быстрота	высокая	нейтральная	высокая

Несмотря на то, что ни один метод интеллектуального анализа данных не может обеспечить решение всего спектра задач технологии Data Mining, согласно данным в таблице, наиболее оптимальным вариантом для рассматриваемой задачи является линейная регрессия, что обусловлено относительно высокой точностью и оперативностью обработки данных по сравнению с другими рассматриваемыми методами.

В качестве ключевого критерия сравнения выступила точность, характеризующая близость полученных результатов с экспериментальными значениями. Полученные при помощи различных методов интеллектуального анализа данных, значения напряженно-деформированного состояния балки были сопоставлены с экспериментальными данными, результаты приведены в табл. 2. Сравнение было выполнено при следующих значениях параметров: температура воздуха в помещении — 21°C; влажность

воздуха в помещении — 34 %; нагрузка — 300 кгс; материал балки — сталь Ст3сп5.

Таблица 2. Результаты сравнительного анализа данных, полученных различными способами

Методы	Значения НДС, кгс/мм ²	Точность, %
Линейная регрессия	7.4011	97.30
Деревья решений	7.3562	96.71
Метод опорных векторов	7.3547	96.69
Эксперимент	7.6065	-

Как видно из таблицы, результаты аналитического расчета, полученные при помощи метода опорных векторов и деревьев решений практически совпадают. Однако наиболее точно экспериментальные данные описывает линейная регрессия, что вероятно связано с тем, что при разработке математической модели было учтено действие факторов окружающей среды, то есть микроклимата помещения, на напряженно-деформированное состояние объекта исследования.

Таким образом, в работе была реализована конкурентная оценка математических моделей, полученных при помощи разных методов интеллектуального анализа данных. Данная оценка была основана на применении различных моделей к одному и тому же набору данных с последующим сравнением их характеристик для выбора наилучшего варианта.

VIII. РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ ДЛЯ ПРОГНОЗИРОВАНИЯ РЕЗУЛЬТАТОВ ПРОМЫШЛЕННЫХ ИСПЫТАНИЙ ЭЛЕМЕНТОВ КОНСТРУКЦИИ

На основе анализа собранных экспериментальных данных было установлено, что наиболее оптимальным решением для реализации информационной системы является вариант, основанный на микросервисной архитектуре.

Преимущество подобного подхода заключается в возможности масштабирования и тестирования каждого из элементов готового программного продукта вне зависимости от других.

Структура программного решения включает базу данных и три модуля для сбора, обработки и визуализации экспериментальных данных.

Структура информационной система представлена на рис. 3.



Рисунок 3 – Структура информационной системы

Для создания информационной системы первым этапом является сбор данных, которые непосредственно будут переданы в базу данных. Необработанные

данные, которые хранятся в базе данных, передаются в модуль визуализации данных и в модуль обработки данных. После выполнения необходимых действий в модуле для прогноза данных результат также перенаправляется в модуль визуализации для отображения и возможного сравнения с необработанными данными.

Суть работы информационной системы заключается в следующем. Сначала собирают экспериментальные данные. На их основе получают теоретические зависимости значений напряженно-деформированного состояния от ряда параметров. На основе этих зависимостей проводят расчеты и прогнозы. Периодически в параллель с расчетами ведут эксперименты для контроля расчетных значений. Новые опытные данные добавляются в базу данных. При накоплении данных или появлении новых параметров возникает необходимость в корректировке теоретических зависимостей, для этого производят их перерасчет, который осуществляется на основе всего объема данных, хранящихся в базе данных.

Основные этапы работы информационной системы и задействованные при этом модули приведены в табл. 3.

Таблица 3. Взаимосвязь этапов работы программы и ее модулей

Этапы работы информационной системы	Модули информационной системы
Сбор экспериментальных данных	Модуль сбора данных База данных
Создание математической модели	База данных Модуль обработки данных
Проведение новых опытов	Модуль сбора данных База данных
Проведение расчетов	База данных Модуль обработки данных
Сравнение расчетных и экспериментальных данных	База данных Модуль обработки данных Модуль визуализации данных

Таким образом, на основе анализа состава и функций модулей информационной системы был составлен алгоритм ее работы (рис. 4).

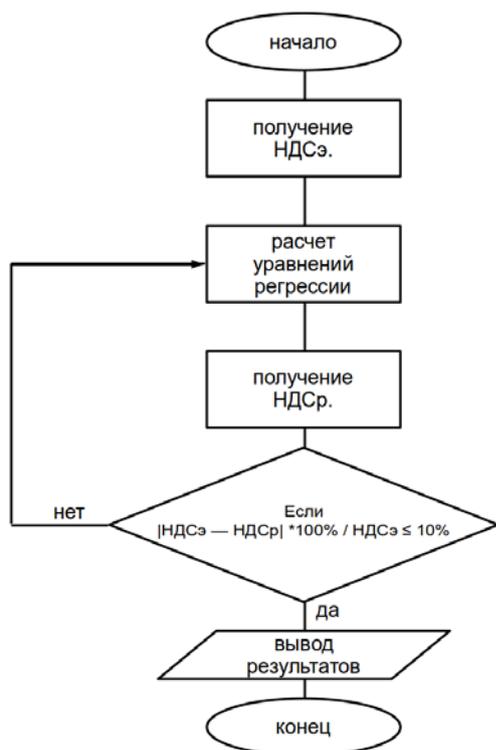


Рисунок 4 – Алгоритм работы программного обеспечения

Следующим этапом работы по созданию программного обеспечения для анализа данных и прогнозирования результатов промышленного испытания элементов конструкции по данным тензометрии стало проектирование базы данных.

Готовая база данных содержит следующие основные таблицы. Таблица employee включает информацию о работниках компании, осуществляющих как непосредственное выполнение, так и контроль за ходом натурных испытаний. Помимо фамилии, имени и отчества сотрудника, в таблице содержится информация о его должности и табельном номере.

Таблица experiment содержит информацию о дате проведения натурных испытаний, характеристике микроклимата помещения (температура и влажность воздуха), величине нагрузки, создаваемой гидроприводом, а также об изменении напряженно-деформированного состояния изделия на основании данных, передаваемых с тензометрических датчиков.

В таблице detector хранится информация о коде датчика, представляющая собой номер, который позволяет его идентифицировать; информация о номере датчика, описывающая его местоположение на изделии (причем, в зависимости от конкретного опыта, номер датчика может меняться за счет изменения его локации, в то время, как код датчика меняться не будет); а также информация о характеристике датчика, его технических особенностях.

В таблице balk содержится информация об инвентарном номере балки, материале, из которого она изготовлена, а также о ее характеристиках.

Также для реализации в физической модели данных связи «многие-ко-многим» была дополнительно создана

таблица connection, содержащая коды датчиков и коды проводимых экспериментов.

Таким образом, в результате проектирования была создана база данных, позволяющая хранить и обрабатывать информацию, необходимую для работы программного обеспечения в достаточном объеме.

Проектирование модуля для обработки и прогноза данных было реализовано на основе множественной линейной регрессии. В качестве основного инструмента для интеллектуального анализа данных был применен программный продукт Rstudio. Принцип работы модуля включает в себя три ключевых этапа, показанных на схеме (рис. 5).

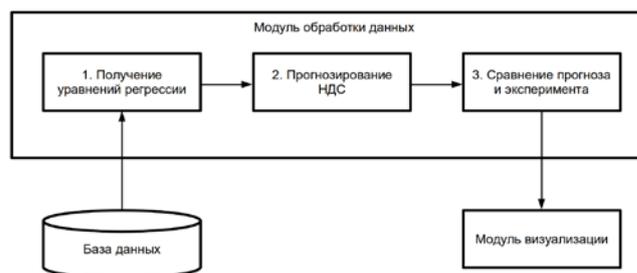


Рисунок 5 – Схема работы модуля обработки данных

На первом этапе выполняется построение математических регрессионных моделей для каждой исследуемой точки на балке. Исходными данными для указанного этапа являются все экспериментальные данные, хранящиеся в базе данных. В результате получается двадцать уравнений множественной регрессии, отличающихся по интенсивности влияния независимых переменных на величину напряженно-деформированного состояния объекта исследования.

На втором этапе на основе полученных уравнений множественной регрессии и вводимых пользователем значений независимых переменных осуществляется прогнозирование величины напряженно-деформированного состояния балки. Проведение натурных испытаний двутавра при этом не завершается, а осуществляется параллельно.

Третий этап необходим для контроля работы модуля обработки данных, он заключается в периодическом осуществлении сравнения прогнозных, то есть расчетных, и экспериментальных данных. В качестве периода используется интервал времени, составляющий две недели, что соответствует проведению примерно 64-76 экспериментальных исследований. За этот период идет накопление новых результатов натурных испытаний, которые могут повлиять на изменение регрессионных зависимостей. Рассчитанные значения напряженно-деформированного состояния балки из модуля обработки данных передаются в модуль визуализации для отображения в виде отчетов и графиков.

Заключительным этапом является проверка соответствия значений напряженно-деформированного состояния, полученных из уравнений и из экспериментов. Если в результате проверки расхождение между расчетными и экспериментальными

значениями составляет менее 10%, то полученные уравнения множественной регрессии продолжают использовать для предсказания величины напряженно-деформированного состояния балки в процессе ее эксплуатации. В случае, если расхождение будет более 10%, выполняется перерасчет уравнений.

Таким образом, в результате работы модуля обработки данных осуществляется прогнозирование значений промышленного испытания элементов конструкции промышленных прочностных испытаний объекта исследования по данным тензометрии без проведения натурных испытаний. Это позволяет отслеживать в режиме реального времени динамику изменения напряжений в элементах металлических конструкций, что, в свою очередь, дает возможность предотвратить случаи нарушения целостности конструкции, а также оценить ее состояние до момента начала ее разрушения. Все это способствует сокращению расходов на производство конструкции и увеличению периода ее эксплуатации.

IX. ЗАКЛЮЧЕНИЕ

В ходе реализации работы был проведен анализ предметной области; была изучена характеристика экспериментального стенда; была выполнена классификация методов интеллектуального анализа данных; был сделан выбор инструментальных средств для реализации информационной системы; была спроектирована и протестирована информационная система для прогнозирования результатов промышленного испытания элементов металлической конструкции. Для решения основной задачи были выбраны следующие интеллектуальные методы анализа данных: линейная регрессия, деревья решений, метод опорных векторов. Для выполнения сравнительного анализа методов были определены следующие критерии: точность, масштабируемость, трудоемкость, быстрота. Полученные результаты позволили выбрать наиболее оптимальный метод реализации задачи прогнозирования. В качестве данного метода была определена линейная регрессия. Таким образом, в работе была реализована конкурентная оценка математических моделей, полученных при помощи разных методов интеллектуального анализа данных. Данная оценка была основана на применении различных моделей к одному и тому же набору данных с последующим сравнением их характеристик для выбора наилучшего варианта.

В результате выполнения работы была создана программа для прогноза результатов промышленного испытания элементов конструкции по данным тензометрии на основе применения интеллектуальных методов анализа данных, то есть прогнозирования того, при какой нагрузке и в какой точке в балке из определенного материала произойдет разрушение. При этом применение микросервисной архитектуры позволило решить такие вопросы, как: масштабируемость системы, независимость ее элементов друг от друга, их взаимозаменяемость. В

дальнейшем готовое программное обеспечение может быть использовано для прогнозирования напряженно-деформированного состояния других объектов при проведении их прочностных испытаний, а также в качестве метода неразрушающего контроля при исследовании напряженно-деформированного состояния элементов металлических конструкций по данным тензометрии.

БИБЛИОГРАФИЯ

- [1] Дядичев В.В. Задачи и методы интеллектуального анализа данных / В.В. Дядичев, Е.В. Ромашка, Т.В. Голуб // Геополитика и эко геодинамика регионов. – 2021. – Т. 1. – № 3. – С. 23-29.
- [2] Спас К.В. Анализ данных и интеллектуальные системы / К.В. Спас, А.Д. Азерников, К.А. Чичеров // Аллея науки. – 2018. – Т. 1. – № 5. – С. 234-239.
- [3] Сафин С.О. Проблемы анализа и обработки Big Data на основе использования метода Data Mining / С.О. Сафин // Материалы XVI Всероссийской молодежной научной конференции. – Уфа, 2022. – С. 701-712.
- [4] Проходский Д.В. Методы и стадии Data Mining / Д.В. Проходский, А.С. Гриб // Студенческий вестник. – 2020. – № 16-9. – С. 8-9.
- [5] Смолина Е.М. Преимущества применения методов интеллектуального анализа данных в образовании / Е.М. Смолина, Л.В. Черенькая // Системный анализ в проектировании и управлении. – 2021. – Т. 25. – № 1. – С. 537-542.
- [6] Губченко Н.О. Методы интеллектуального анализа данных: виды и их визуализация (Data Mining) / Н.О. Губченко, В.А. Сорокина // Инновационные подходы в современной науке. – 2021. – С. 68-74.
- [7] Семенов В.А. Анализ методов и задач интеллектуального анализа данных / В.А. Семенов, М.М. Семенова, Л.В. Мызникова // Известия Великолукской государственной сельскохозяйственной академии. – 2021. – С. 209-212.
- [8] Антонов Г.В. Линейная регрессия как один из методов статистического исследования / Г.В. Антонов, С.И. Иванов // Известия Великолукской государственной сельскохозяйственной академии. – 2021. – № 2. – С. 64-75.
- [9] Сулейманова А.Н. Обзор развития алгоритмов деревьев решений / А.Н. Сулейманова // Социология: методология, методы, математическое моделирование. – 2020. – № 50-51. – С. 64-97.
- [10] Гасымов К.Н. Метод опорных векторов / К.Н. Гасымов // Развитие науки в современном Мире. – 2017. – № 3. – С. 53-55.
- [11] Колентеев Н.Я. Коэффициенты корреляции и детерминации / Н.Я. Колентеев, О.А. Гончарова // Специальная техника и технологии транспорта. – 2022. – № 1. – С. 206-212.
- [12] Пичугин О.Н. Деревья решений как эффективный метод анализа и прогнозирования / О.Н. Пичугин, Ю.З. Прокофьева, Д.М. Александров // Нефтепромышленное дело. – 2021. – № 11. – С. 69-75.
- [13] Захарова О.И. Деревья решений и алгоритмы их построения / О.И. Захарова, Е.С. Артюшкина, С.В. Холопов // Евразийское научное объединение. – 2020. – № 4-2. – С. 97-99.
- [14] Кувайскова Ю.Е. Статистические методы прогнозирования: учебное пособие // Ю.Е. Кувайскова, В.Н. Клячкин. – Ульяновск: УлГТУ, 2019. – 197 с.
- [15] Вьюгин В.В. «Математические основы теории машинного обучения и прогнозирования» М.: 2022 - 387 с.
- [16] Омарова Ш.Е. Сравнительный анализ инструментов Data Mining / Ш.Е. Омарова, А.М. Медеубаева // Заметки ученого. – 2020. – № 11. – С. 185-193.

Истратова Евгения Евгеньевна. Новосибирский государственный технический университет, г. Новосибирск, Россия. Кандидат технических наук, доцент кафедры автоматизированных систем управления. Количество печатных работ: 121. Область научных интересов: информационные технологии, информационные сети, системы компьютерного зрения. e-mail: istratova@mail.ru (ответственная за переписку).

Кожевников Алексей Николаевич. Новосибирский государственный технический университет, г. Новосибирск, Россия. Кандидат технических наук, доцент кафедры прочности. Количество печатных

работ: 42. Область научных интересов: информационные системы, тензометрия, прочностные расчеты.

Ласточкин Павел Валерьевич. Новосибирский государственный технический университет, г. Новосибирск, Россия. Аспирант факультета летательных аппаратов. Количество печатных работ: 22. Область научных интересов: информационные системы и технологии, информационные сети.

Глинин Евгений Вадимович. Новосибирский государственный технический университет, г. Новосибирск, Россия. Магистрант факультета автоматики и вычислительной техники. Количество печатных работ: 6. Область научных интересов: информационные технологии, информационные сети.

Application of intelligent data analysis methods to the problem of predicting the results of industrial testing of structural elements based on tensometry data

E.E. Istratova, A.N. Kozhevnikov, P.V. Lastochkin, E.V. Glinin

Abstract — The article presents the results of a study of three intelligent data analysis methods for solving the problem of predicting the results of industrial testing of structural elements according to strain gauge data. As an object of study, an I-beam was considered, for which a series of experiments was carried out with loading and recording the values of its stress-strain state. The obtained experimental values were used to study and predict the strength in real time. The scientific novelty of the work lies in the proposal of both the author's method for predicting the stress-strain state of the beam depending on the environmental parameters, and a tool based on the use of data mining methods. As a result of the work, an information system was implemented, a distinctive feature of which is the ability to quickly process large amounts of experimental, that is, raw data in real time, which is achieved through the use of data mining methods. In the course of the work, environmental parameters were identified and analyzed both from the side of the stand and from the side of the microclimate, which affect the accuracy of measurements. Due to this, experimental data can be corrected at any time when certain characteristics change.

Keywords — strain gauge, stress-strain state, I-beam, data mining, linear regression, decision trees, support vector machine.

REFERENCES

- [1] Dyadichev V.V. Tasks and methods of data mining / V.V. Dyadichev, E.V. Chamomile, T.V. Golub // Geopolitics and eco-geodynamics of regions. - 2021. - T. 1. - № 3. - S. 23-29.
- [2] Spas K.V. Data analysis and intelligent systems / K.V. Spas, A.D. Azernikov, K.A. Chicherov // Science Alley. - 2018. - T. 1. - № 5. - S. 234-239.
- [3] Safin S.O. Problems of analysis and processing of Big Data based on the use of the Data Mining method / S.O. Safin // Proceedings of the XVI All-Russian Youth Scientific Conference. - Ufa, 2022. - S. 701-712.
- [4] Prokhodsky D.V. Methods and stages of Data Mining / D.V. Prokhodsky, A.S. Mushroom // Student Bulletin. - 2020. - № 16-9. - S. 8-9.
- [5] Smolina E.M. Benefits of using data mining methods in education / E.M. Smolina, L.V. Chernenkaya // System analysis in design and management. - 2021. - T. 25. - № 1. - S. 537-542.
- [6] Gubchenko N.O. Methods of data mining: types and their visualization (Data Mining) / N.O. Gubchenko, V.A. Sorokina // Innovative approaches in modern science. - 2021. - S. 68-74.
- [7] Semenov V.A. Analysis of methods and tasks of data mining / V.A. Semenov, M.M. Semenova, L.V. Myznikova // Information technologies, system analysis and management (ITSAU-2021). - 2021. - S. 209-212.
- [8] Antonov G.V. Linear regression as one of the methods of statistical research / G.V. Antonov, S.I. Ivanov // Proceedings of the Velikie Luki State Agricultural Academy. - 2021. - № 2. - S. 64-75.
- [9] Suleymanova A.N. Overview of the development of decision tree algorithms / A.N. Suleimanova // Sociology: methodology, methods, mathematical modeling. - 2020. - № 50-51. - S. 64-97.
- [10] Gasimov K.N. Support vector method / K.N. Gasimov // Development of science in the modern world. - 2017. - № 3. - S. 53-55.
- [11] Kolenteev N.Ya. Coefficients of correlation and determination / N.Ya. Kolenteev, O.A. Goncharova // Special equipment and transport technologies. - 2022. - № 1. - S. 206-212.
- [12] Pichugin O.N. Decision trees as an effective method of analysis and forecasting / O.N. Pichugin, Yu.Z. Prokofiev, D.M. Aleksandrov // Oilfield business. - 2021. - № 11. - S. 69-75.
- [13] Zakharova O.I. Decision trees and algorithms for their construction / O.I. Zakharova, E.S. Artyushkina, S.V. Kholopov // Eurasian Scientific Association. - 2020. - № 4-2. - S. 97-99.
- [14] Kuvaiskova Yu.E. Statistical forecasting methods: textbook // Yu.E. Kuvaiskova, V.N. Klyachkin. - Ulyanovsk: UIGTU, 2019. - 197 p.
- [15] Vyugin V.V. "Mathematical foundations of the theory of machine learning and forecasting" M.: 2022 - 387 p.
- [16] Omarova Sh.E. Comparative analysis of Data Mining tools / Sh.E. Omarova, A.M. Medeubaeva // Notes of a scientist. - 2020. - № 11. - S. 185-193.