

ОЦЕНКА ТЕХНИЧЕСКОГО СОСТОЯНИЯ ОБЪЕКТА: ВЫДЕЛЕНИЕ ИНФОРМАТИВНЫХ ПРИЗНАКОВ И ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

Голев А. В.¹, Московцев А. М.², Огородников О. В.³
*(ФГБУН Институт проблем управления
им. В.А. Трапезникова РАН, Москва)*

Предварительная обработка исходных данных повышает эффективность применения методов интеллектуального анализа данных (ИАД). Наряду с очисткой, нормализацией, преобразованием данных и т.д. она включает в себя этап выделения наиболее информативных признаков, т.е. позволяет снизить размерность пространства признаков. Многие исследователи считают этап выделения информативных признаков рассматриваемого объекта одним из самых важных и сложных этапов решения задач ИАД. В связи с этим становится актуальным вопрос выбора алгоритмов, обеспечивающих наибольшую эффективность методов ИАД. В докладе рассмотрены алгоритмы выделения информативных признаков в задаче ИАД на примере оценки и прогноза технического состояния (задача ранней диагностики) электродвигателя беспилотного летательного аппарата, где анализируемое множество значений контролируемых параметров относится к определенному классу или уровню опасности в зависимости от возможных последствий сложившейся ситуации (состояния технического объекта). Применение методов ИАД позволяет решать задачу контроля технического состояния технического объекта, когда отсутствуют достаточное количество данных для принятия статистических гипотез с необходимым уровнем доверия и подробные математические описания физических процессов развития деградаций в условиях воздействия всевозможных внутренних и внешних факторов.

Ключевые слова: интеллектуальный анализ данных, выделение информативных признаков, классификация, диагностика, электродвигатель, беспилотный летательный аппарат.

1. Введение

Актуальность разработки систем контроля технического состояния (СКТС) электродвигателя обусловлена развитием технологий, связанных с созданием перспективных беспилотных

¹ Артем Владимирович Голев, м.н.с. (oiw23@mail.ru).

² Александр Михайлович Московцев, инженер (salllok-98@yandex.ru).

³ Олег Викторович Огородников, н.с. (o.v.ogorodnikov@gmail.com).

электрических летательных аппаратов (ЛА) [2]. Внедрение высокоэффективных СКТС позволит обеспечить высокие показатели отказоустойчивости и снизить эксплуатационные затраты.

Задача разработки алгоритмов оценки состояния технических объектов на практике решается с использованием концептуальных подходов «Белого ящика», «Черного ящика» или их комбинирования [1].

При использовании подхода «Белого ящика» технический объект (ТО) рассматривают как динамический объект, для которого полностью известна его внутренняя структура, набор взаимодействующих элементов, характер связей между ними, внешние и внутренние условия и их влияние. В ходе анализа выполняется моделирование приведенного динамического объекта, изменение параметров которого описывается системой дифференциальных уравнений.

Многие динамические объекты имеют сложную структуру и состоят из значительного количества взаимодействующих элементов. При построении моделей таких объектов необходимо учитывать значительное количество внутренних и внешних факторов. Это часто обуславливает невысокую адекватность моделей, построенных на базе подхода «Белый ящик». К примеру, поведение даже простого динамического объекта, функционирующего в штатном или аварийном режиме, не будет соответствовать построенной модели, так как набор возможных неисправностей объекта широк и не может быть предусмотрен в модели заранее. В противном случае, для описания штатных ситуаций необходимо сильно усложнять модель.

На практике поведение большинства объектов плохо поддается формализации. Поэтому подход к построению алгоритмов оценки состояния технических объектов, основанный только на аналитическом моделировании, применяется для сравнительно простых объектов с очевидными свойствами. Аналитические модели динамического объекта не способны приспособиваться к изменившимся условиям функционирования, т.е. не способны обучаться и в конечном счете могут стать неадекватными.

При использовании концептуального подхода «Черный ящик» к построению алгоритмов оценки состояния ТО необходимость в точной информации о структуре ТО как динамического

объекта отсутствует. В рамках данного подхода считается, что о функциях и реакциях ТО можно судить по выходным параметрам, наблюдаемым как результат на внешнее воздействие.

Для упрощения задачи поиска и формализации закономерностей в данных о состоянии ТО в них выделяют подмножества прецедентов и интерпретируют их как классы, которые соответствуют подмножествам конкретных состояний ТО. Для решения задачи построения моделей классификации применяется дискриминантный анализ, например, линейный дискриминант Фишера, логистическая регрессия и т.д.

Большой интерес представляют различные методы ИАД, обеспечивающие высокую эффективность в задачах поиска трудно формализуемых закономерностей в данных. В настоящий момент имеется значительное число современных методов ИАД, пригодных для решения задач оценки состояния ТО. Самые популярные из них – это нейронные сети [3], деревья решений [7], метод рассуждения на основе аналогичных случаев (метод k -ближайших соседей) [4], метод опорных векторов [5]. Представленные методы и их различные модификации имеют одно общее свойство – используют информацию о ранее возникающих состояниях ТО (прецедентах) для поиска и формализации закономерностей в его поведении, однако различаются между собой по принципу формирования конечного результата.

Для повышения эффективности применения методов ИАД нередко производится предварительная обработка исходных данных, позволяющая выделить наиболее информативные признаки (ИП) для решения задачи классификации (снизить размерность пространства признаков).

Многие исследователи считают этап выделения информативных признаков одним из самых важных и сложных этапов решения задач классификации. В связи с этим становится актуальным вопрос выбора методов и алгоритмов выделения ИП, обеспечивающих наибольшую эффективность методов ИАД.

Задача выделения информативных признаков (снижения размерности пространства признаков) заключается в уменьшении количества исходных признаков (входных параметров), поступающих на вход алгоритмов или моделей классификации, в которых формализованы закономерности, отражающие взаимосвязи

между параметрами и техническим состоянием ТО. Особенно значительное внимание уделяется этой проблеме теорией машинного обучения в прикладных задачах, решение которых основано на методах ИАД, так как часто поиск и формализация закономерностей могут быть затруднены из-за избыточности информации (проблема «проклятия размерности») и шума, что может приводить к переобучению (overfitting). Во многих случаях снижение размерности пространства признаков позволяет увеличить точность и упростить модели/алгоритмы классификации, создаваемые с применением методов ИАД.

Обычно при уменьшении размерности пространства признаков исходное множество признаков подвергается процедурам отбора или проекции (преобразования) признаков.

Под отбором признаков понимается формирование подмножества значимых признаков для модели/алгоритма классификации. Применение «оберточных» методов (wrapper method) [6] является наиболее универсальным подходом к отбору признаков. Данные методы обеспечивают оптимизацию подмножества параметров моделей/алгоритмов классификации на основе информации, полученной в результате итерационного выполнения полноценного поиска и формализации закономерностей с использованием методов ИАД. Главной особенностью «оберточных» методов является то, что на каждой итерации оптимизационного алгоритма создаются модели/алгоритмы, которые в дальнейшем могут непосредственно применяться для решения прикладной задачи. Таким образом для каждого подмножества признаков, сгенерированного «оберточным» методом, выполняется обучение нейронной сети, строится дерево решений, растущая пирамидальная сеть и т.д.

Оптимизация подмножества информативных признаков может сочетаться с оптимизацией параметров применяемого метода ИАД (выбор архитектуры нейронной сети – количество слоев, нейронов, вид активационной функции; глубины дерева решения; параметров метода k -ближайших соседей – метрика сходства объектов, количество соседей).

Анализ мировой научной литературы показал актуальность применения методов ИАД в интеграции с методами выделения информативных признаков. Использование оберточных методов,

методов фильтрации и проекции для выделения информативных признаков позволит повысить эффективность создания и применения алгоритмов оценки и прогноза технического состояния электродвигателей летательных аппаратов. В качестве методов фильтрации выбраны поиск корреляционных зависимостей, критерий Information gain, статистический критерий хи-квадрат-тест. В качестве методов проекции предлагается использовать спектральный анализ, позволяющий учитывать особенности функционирования электродвигателя, метод главных компонент и ядерный метод главных компонент. В качестве «оберточных» методов выбраны метод полного перебора для малого количества исходных признаков (< 6) и генетические алгоритмы – при большей размерности.

В результате решения задачи ранней диагностики с использованием методов ИАД выполняется оценка и прогноз технического состояния электродвигателя ЛА, при которых анализируемое множество значений контролируемых параметров относится к определенному классу или уровню опасности в зависимости от возможных последствий сложившейся ситуации. При этом наиболее важным является прогноз технического состояния электродвигателя. Возникновение в процессе эксплуатации нештатной или аварийной ситуации может свидетельствовать о случившемся выходе из строя электродвигателя, поэтому получение выводов о текущем состоянии контролируемого объекта является недостаточным. Основной задачей анализа информации при диагностике электродвигателя является определение негативных тенденций изменения технического состояния электродвигателя заранее.

Возможность предотвращения нештатных и аварийных ситуаций, возникающих в процессе эксплуатации электродвигателя, в основном определяется способностью алгоритмов, в которых реализованы методы анализа информации о контролируемых параметрах, точно идентифицировать и прогнозировать техническое состояние электродвигателя. Точность оценки и прогноза технического состояния электродвигателя во многом определяется возможностями алгоритмов выделения информативных признаков, которые применяются при построении алгоритмов

оценки технического состояния электродвигателя с использованием методов ИАД.

В соответствии с представленной формализацией в алгоритмах диагностики технического состояния электродвигателя реализуются:

- методы поиска и формализации закономерностей в эмпирических данных, представленных в виде набора прецедентов;
- формальное представление найденных закономерностей и методы для их использования.

Для повышения эффективности формализации закономерностей, отражающих зависимости между контролируруемыми параметрами и оценкой технического состояния электродвигателя, предлагается схема формирования и анализа данных для выделения информативных признаков, которая рассматривается в следующем разделе.

2. Анализ данных о техническом состоянии электродвигателя

Для решения задачи ранней диагностики электродвигателя с использованием методов ИАД предлагается следующая схема формирования и анализа данных.



Рис. 1. Схема анализа данных в задаче диагностики электродвигателя

Решение задачи состоит из нескольких этапов: получение эмпирических данных и их обработка, создание первичного набора векторов из эмпирических данных, оптимизация множества признаков и обучение алгоритмов ИАД на данных.

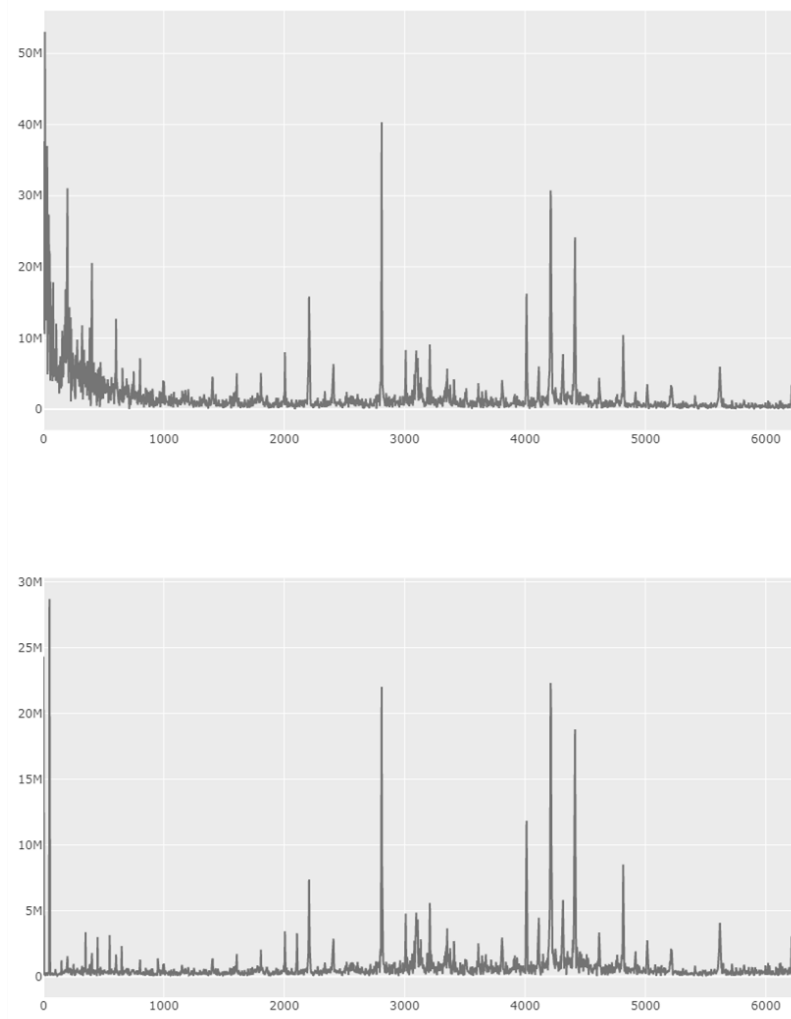


Рис. 2. Пример преобразования Фурье вибрационного (сверху) и звукового (снизу) сигналов

Для проведения ресурсных испытаний был создан стенд с жестко закрепленным двигателем U5 KV400. Нагрузкой служит пропеллер, создающий поток воздуха. На стенде установлены:

- датчик виброускорений для измерения вибраций статора двигателя;
- измерительный микрофон для акустических шумов двигателя;
- фотодатчик, фиксирующий пересечение светового потока лопастью пропеллера;
- делители напряжения для записи фазных токов и напряжений.

На этапе обработки экспериментальных данных были рассмотрены данные за последние 150 дней. По каждой записи для каждого канала были построены с помощью преобразования Фурье спектры сигналов.

Для того чтобы убрать избыточность в данных и снизить размерность пространства признаков, были использованы метод отбора признаков с помощью теста χ^2 и с использованием случайного леса.

Используя метод χ^2 удалось отобрать 24 признака с уровнем значимости $P > 0,8$, используя метод отбора признаков с помощью случайного леса удалось отобрать 16 признаков из набора данных.

После того как были отобраны признаки, из полного набора данных были сформированы 4 набора:

- 1) полный вектор признаков;
- 2) масштабированный вектор признаков;
- 3) масштабированные признаки, отобранные методом χ^2 ;
- 4) масштабированные признаки, отобранные методом случайного леса.

Все вышеперечисленные наборы данных использовались в решении задачи классификации, где имелось два класса – А и Б. Классы были сбалансированы – 55% векторов класса А к 45% векторов класса Б. Ниже приведены различные методы и их оценки на исследуемых наборах данных.

Таблица 1. Результаты применения различных методов ИАД

Модели	Метод k -ближайших соседей	Метод k -ближайших соседей
Данные	χ^2	Случайный лес
F -мера	0,983	0,991
Точность	0,967	0,983
Полнота	1	0,998
Валидация	0,982	0,990
Обучение	0,987	0,991
Тест	0,989	0,988

При использовании полного и масштабированного наборов данных происходило переобучение на тренировочных наборах, чего нельзя сказать про результаты моделей с отобранными признаками. Оценки на тестовом наборе и $F1$ -мера на методе k -ближайших соседей у данных с отобранными признаками наиболее высокие.

3. Заключение

На основе экспериментальных наборов данных о тестовой эксплуатации электродвигателя с применением преобразования Фурье получен набор прецедентов, в котором были выделены информативные признаки. С использованием отобранных информативных признаков решена задача классификации и диагностики технического состояния.

Направление дальнейшей работы предполагает дополнительную обработку данных и удаление выбросов и шумов, внесённых температурой и влажностью, и оценку результатов работы моделей ИАД с информативными признаками, полученными с помощью «обёрточных методов», а также генетических алгоритмов и сверточных нейронных сетей.

Литература

1. МЫЛЬНИК В.В., ТИТАРЕНКО Б.П., ВОЛОЧИЕНКО В.А. *Исследование систем управления*. – М.: Академический Проект, 2003. – 352 с.

2. СКРЯБИН А.В. *Системы контроля технического состояния и прогнозирования неисправностей электромеханических рулевых приводов летательного аппарата. Современный уровень развития.* // Общероссийский научно-технический журнал «Полет». – 2018. – №2. – С. 50–64.
3. УОССЕРМЕН Ф. *Нейрокомпьютерная техника: теория и практика.* – М.: Мир, 1992. – 157 с.
4. HE Q.P., WANG J. *Fault Detection Using the k-Nearest Neighbor Rule for Semiconductor Manufacturing Processes* // IEEE Trans. On Semiconductor Manufacturing. – 2007. – Vol. 20, No. 4. – P. 345-354
5. JAYADEVA KHEMCHANDANI R., CHANDRA S. *Twin Support Vector Machines: Models, Extensions and Applications.* – Springer, 2016. – P. 211.
6. PANTHONG R., SRIVIHOK A. *Wrapper Feature Subset Selection for Dimension Reduction Based on Ensemble Learning Algorithm* // The Third Information Systems Int. Conference // Procedia Computer Science. – 2015. – Vol. 72. – P. 162–169.
7. ROKACH L. *Data Mining With Decision Trees: Theory and Applications.* – World Scientific Publishing Company, 2014. – P. 328

OBJECT TECHNICAL STATUS ESTIMATION: SELECTION OF INFORMATIVE FEATURES AND DATA ANALYSIS

Artem Golev, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, (oiw23@mail.ru).

Aleksandr Moskovcev, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, (salllok-98@yandex.ru).

Oleg Ogorodnikov, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, (o.v.ogorodnikov@gmail.com).

Abstract: Pre-processing of the source data increases the efficiency of using data mining methods. Along with cleaning, normalization, data transformation, etc., it includes the stage of selecting the most informative features, i.e. it allows you to reduce the dimension of the feature space. Many researchers consider the stage of identifying informative features of the object under consideration to be one of the most important and complex stages of solving data mining problems. In this regard, the question of choosing algorithms that ensure the greatest efficiency of data mining methods becomes relevant. The report considers algorithms for selecting informative features in the data mining problem on the example of assessing and predicting the technical

status (the problem of early diagnosis) of an unmanned aerial vehicle electric motor, where the analyzed set of values of controlled parameters belongs to a certain class or level of danger, depending on the possible consequences of the current situation (the state of the technical object). The use of data mining methods allows us to solve the problem of monitoring the technical status of a technical object when there is not enough data to accept statistical hypotheses with the necessary level of confidence and detailed mathematical descriptions of the physical processes of degradation development under the influence of various internal and external factors.

Keywords: intelligent data analysis, identification of informative features, classification, diagnostics, electric motor, unmanned aerial vehicle.

УДК 621.33 + 629.735.7

ББК 16.23 + 31.291