

УДК 629.7

ИССЛЕДОВАНИЕ МЕТОДОВ ОТБОРА ИНФОРМАТИВНЫХ ПРИЗНАКОВ И МАШИННОГО ОБУЧЕНИЯ ДЛЯ ДИАГНОСТИКИ ТЕХНИЧЕСКОГО СОСТОЯНИЯ ЭЛЕКТРОМЕХАНИЧЕСКИХ ПРИВОДОВ ЛЕТАТЕЛЬНОГО АППАРАТА

А.В. Голев

Институт проблем управления им. В.А. Трапезникова РАН
Россия, 117997, Москва, Профсоюзная ул., 65
E-mail: oiw23@mail.ru

Д.А. Петров

Институт проблем управления им. В.А. Трапезникова РАН
Россия, 117997, Москва, Профсоюзная ул., 65
E-mail: daniel.petrov@ipu.ru

Ключевые слова: интеллектуальный анализ данных, классификация, диагностика, электромеханический привод, летательный аппарат, состояние технического объекта.

Аннотация: В докладе представлено сравнительное исследование методов классификации сигналов технического состояния электромеханического привода летательного аппарата. Также, предлагая подход к решению задачи классификации, рассматривается несколько типов алгоритмов отбора информативных признаков и интеллектуального анализа данных. Особое внимание уделяется анализу результатов классификации, где представлено сравнение эффективности различных методов классификации. В работе приводятся методы повышения надежности диагностических систем за счет оптимизации процесса отбора информативных признаков, что способствует своевременному выявлению потенциальных отказов и сокращению времени простоя аппаратуры.

1. Введение

Диагностика технического состояния критически важных узлов летательных аппаратов в контексте электромеханических систем является одной из востребованных задач современной авиационной промышленности. По мере развития концепции усовершенствования летательных аппаратов, например, таких как «электрический самолет», возрастает важность точной и своевременной диагностики. Это повышает безопасность полетов и способствует оптимизации эксплуатационных расходов за счет предотвращения нештатных ситуаций и нарушений работоспособности узлов.

Авиационная промышленность предъявляет строгие требования к точности и своевременной диагностике, что ставит на первый план задачи ранней диагностики и классификации технического состояния объекта. В данном докладе рассматривается сравнение методов классификации состояния электромеханических приводов (ЭМП)

летательных аппаратов, а также анализ эффективности различных методов отбора информативных признаков в контексте интеллектуального анализа данных.

Внедрение методов машинного обучения в системы ранней диагностики предотвращает количество аварийных ситуаций за счет своевременного выявления неисправностей. Данные методы способны обрабатывать и анализировать огромные массивы данных, выявляя сложные закономерности и взаимосвязи, что в совокупности позволяет формировать точные классификационные и прогностические модели технического состояния. Особое внимание в современных исследованиях уделяется методам отбора информативных признаков, которые решают проблему «проклятия размерности», улучшая качество и устойчивость классификационных моделей, снижая риск переобучения и упрощая вычислительные процессы.

В первом разделе доклада предлагается схема комплекса алгоритмов отбора информативных признаков для исследования влияния данных методов на способность алгоритмов решать задачи классификации. Во втором разделе приводятся результаты расчетных исследований решения задачи классификации по трем характеристикам (люфт, сухое и вязкое трение) технического состояния ЭМП.

2. Комплекс алгоритмов отбора информативных признаков

Пусть набор данных математической модели электромеханического привода представлен в виде множества $X = \{(M_1, \dots, M_D)^i\}$, где $i = 1, \dots, n$, составленного из подмножеств M_1, \dots, M_D временных последовательностей $M_d = \{m_d^1, \dots, m_d^k\}$, где $d = 1, \dots, D; k = 1, \dots, K$, характеризующих техническое состояние технического объекта, где D – это количество контролируемых параметров математической модели, K – число записей временной последовательности, n – количество итераций изменения внутренних моделируемых параметров системы: люфт (BL), сухое трение (M_{DF}) и вязкое трение (M_{VF}). Данные моделируемые параметры задают нам три характеристики технического состояния $y^i \in Y$, где $y^i = \langle y_{BL}, y_{M_{DF}}, y_{M_{VF}} \rangle^i$. В таблице 1 представлена матрица изменения внутренних параметров модели. Для модели ЭМП определены контролируемые параметры: положение ротора электродвигателя (ЭД), приведенное к выходному валу (град), положение выходного вала (град), напряжение ЭД (В), сила тока ЭД (А), потребляемая мощность ЭД (Вт), угловая скорость ротора ЭД (об/мин), момент ротора ЭД (мНм), мощность на валу ЭД (Вт), угловая скорость на выходном валу (град/с).

Таблица 1. Матрица значений внутренних моделируемых параметров ЭМП.

№	Наименование параметра	Минимальное значение	Максимальное значение	Шаг
1	Люфт	0,2	2,72	0,04
2	Сухое трение	0,36	2,13	0,03
3	Вязкое трение	0,001	0,06	0,001

Для каждой итерации моделируемые параметры $y \in Y$ имеют класс состояния, ставящийся в соответствие элементу $x \in X$, где $x = \langle M_1, \dots, M_D \rangle$. Для решения задачи классификации требуется построить алгоритм оценки технического состояния ЭМП $\Phi: X \rightarrow Y$, позволяющий для произвольного элемента $x_k \in X$ получить элемент $y_k \in Y$.

План эксперимента обработки данных для построения алгоритма классификации технического состояния представлен на рис. 1.



Рис.1. Схема эксперимента работы с набором данных для решения задачи классификации состояния ЭМП.

В данной схеме набор данных математической модели будет представлен в виде векторов вида $\bar{x} = \{m_1^1, \dots, m_1^k, m_2^1, \dots, m_D^k\}$, в котором объединены временные последовательности контролируемых параметров с суммарной длиной более 3000 записей.

Использование алгоритмов для выделения информативных признаков позволяет уменьшить количество подаваемых признаков классификаторам. Выделение осуществляется на основе методов двух типов [1-4]:

- методы фильтрации: метод mutual information, критерий хи-квадрат, корреляция Пирсона, метод абсолютного отклонения, метод Relief, ти-тест;
- встроенные методы: метод дерева решений, метод случайного леса, метод градиентного бустинга, метод extra tree, регрессия Лассо, метод Ridge.

Каждым методом выделения информативных признаков была отобрана сотня признаков с высоким уровнем значимости, при этом обучение и тренировки методов выделения производились на 10% данных. Дополнительно осуществлялся выбор из отобранных признаков оберточными методами: метод последовательного отбора, метод рекурсивного удаления. Также используется дополнительный метод отбора на основе взаимной корреляции между информативными признаками, полученными фильтрационными и оберточными методами. Таким образом, были сформированы наборы обучающих выборок, в которых количество признаков уменьшилось до 10 штук. Примеры графиков уровней значимости представлены на рис. 2-3.

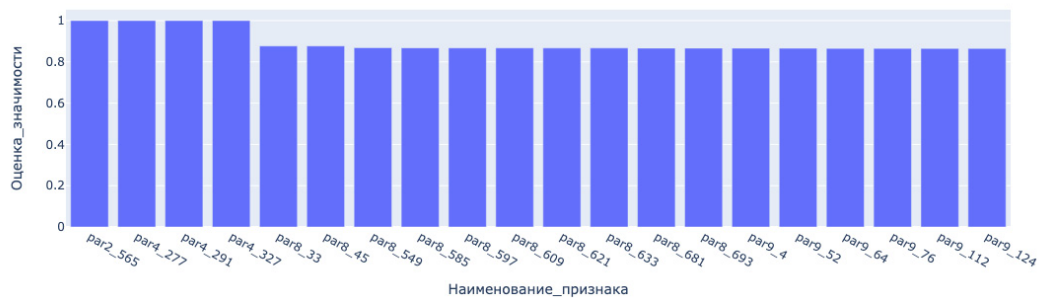


Рис.2. График уровня значимости информативных признаков, отобранных методом ти-тест.



Рис. 3. График уровня значимости информативных признаков, отобранных методом случайного леса.

3. Результаты расчетных исследований

Для решения задачи оценки технического состояния ЭМП использовались модели классификаций [5-7]: метод опорных векторов, метод k-ближайших соседей, логистическая регрессия, деревья решений, нейронная сеть. На основе отобранных множеств признаков производилось обучение с учителем на 10% данных, а оставшиеся 90% использовались на тесте. В таблице 2 представлены примеры оценок классификации на тестовой выборке.

Таблица 2. Примеры результатов классификации по типу характеристики «Люфт».

№	Вид данных	Метод отбора	Класс	Название модели	F1 мера
333	полный вектор	Дерево решений	Люфт	Нейронная сеть	0.998269
332	полный вектор	Дерево решений	Люфт	Решающее дерево	0.995438
331	полный вектор	Дерево решений	Люфт	Метод опорных векторов	0.994155
343	полный вектор	Метод extra_tree	Люфт	Нейронная сеть	0.994141

Были получены статистические результаты по точности алгоритмов классификации. Они представлены на рис. 4. Результаты показывают, что наибольшее количество вхождений в порог выше 0.9 по метрике F1 мера [8] имеет классификатор на основе нейронных сетей.

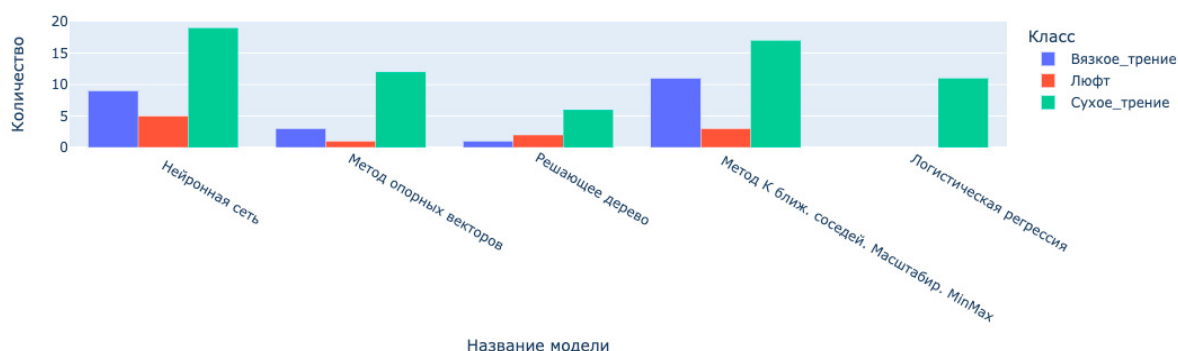


Рис. 4. Сводная диаграмма отношения количества высокоточных оценок по F1 мере для каждого из классификаторов.

На рис. 5 представлены статистические результаты по методам отбора информативных признаков. Данная диаграмма показывает, что информативные признаки, отобранные встроенными методами (методом дерева решений, методом extra tree и методом случайного леса) справились с классификацией по всем характеристикам ЭМП.

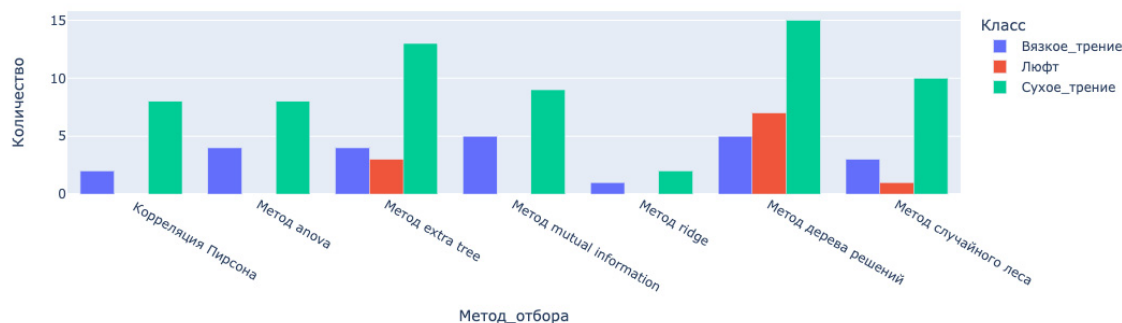


Рис. 5. Сводная диаграмма отношения количества высоких оценок точности по F1 мере к методам отбора информативных признаков.

4. Заключение

Разработан комплекс алгоритмов, позволяющие выделить информативные признаки для классификации технического состояния ЭМП. Эти алгоритмы можно использовать в различных связках для улучшения показателей классификации. Результаты расчетных исследований на примере классификации технического состояния ЭМП, связанных с изменением люфта, сухого и вязкого трений, показывают необходимость дальнейшего развития разработанных алгоритмов. Дальнейшие исследования предлагается направить на повышение эффективности разработанных алгоритмов посредством поиска критериев качества отобранных информативных признаков, позволяющих повысить точность классификационных моделей.

Исследование выполнено частично за счет гранта Российского научного фонда № 23-19-00464.

Список литературы

1. Мыльник В.В., Титаренко Б.П., Волочиенко В.А. Исследование систем управления / 2-е изд., перераб. и доп. М.: Академический Проект, 2003. 352 с.
2. Уоссермен Ф. Нейрокомпьютерная техника: теория и практика / Пер. с англ. М.: Мир, 1992. 157 с.
3. 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.
4. Yu L., Liu H. Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution // Proceedings of the 12'th International Conference on Machine Learning (ICML-2003). Washington DC, 2003. P. 856-863.
5. He Q.P., Wang J. Fault Detection Using the k-Nearest Neighbor Rule for Semiconductor Manufacturing Processes // IEEE Transactions On Semiconductor Manufacturing. 2007. Vol. SM-20, No. 4.
6. Jayadeva, Khemchandani R., Chandra S. Twin Support Vector Machines: Models, Extensions and Applications. Springer, 2016. 211 p.
7. Rokach L. Data Mining With Decision Trees: Theory and Applications / 2nd Edition. World Scientific, 2014. 328 p.
8. Panthong R., Srivihok A. Wrapper Feature Subset Selection for Dimension Reduction Based on Ensemble Learning Algorithm // The Third Information Systems International Conference // Procedia Computer Science. 2015. Vol. 72. P. 162-169.