

СРАВНЕНИЕ ЭФФЕКТИВНОСТИ МЕТОДОВ ОБРАБОТКИ ПРИЗНАКОВ В ЗАДАЧЕ РАННЕЙ ДИАГНОСТИКИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА¹

Вересников Г.С., Голев А.В., Скрябин А.В.

Институт проблем управления им. В.А. Трапезникова РАН, Москва, Россия
veresnikov@mail.ru, oiw23@mail.ru

Аннотация. В докладе рассматривается задача ранней диагностики и классификации технического состояния узла летательного аппарата. Основная цель доклада заключается в сравнении типов методов анализа временных сигналов – методов общего назначения и статистические методы для оценки их эффективности при решении задачи классификации технического состояния электромеханического привода.

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

Введение

Современные методы анализа данных играют ключевую роль в обеспечении безопасности и надежности сложных инженерных систем, таких как летательные аппараты. Диагностика и мониторинг критически важных компонентов, таких как электродвигатели, трансмиссии и редукторы, представляют собой важнейшие задачи, от которых зависит безопасность полетов и эффективность эксплуатации летательных аппаратов. Стратегия проведения ремонтных работ таких систем «после отказа» нецелесообразна из-за значительных затрат на ликвидацию возможных последствий отказа или остановку технологического процесса. Применение стратегии ремонта по техническому состоянию позволяет минимизировать объем затрат путём своевременного ремонта предупреждать аварийные ситуации. Поэтому точная оценка их технического состояния с использованием методов анализа данных становится все более актуальной.

Применение современных подходов к интеллектуальному анализу данных и машинному обучению открывает широкий спектр методов для эффективного анализа вибрационных характеристик ключевых узлов летательных аппаратов. Одним из наиболее значимых аспектов этого процесса является извлечение признаков, которое оказывает существенное влияние на точность и эффективность последующих этапов машинного обучения. Корректное извлечение признаков способствует сокращению размерности данных, улучшению интерпретируемости моделей и снижению шума в данных. Это, в свою очередь, повышает точность классификации, снижает риск переобучения и увеличивает вычислительную эффективность.

В данной работе рассматривается применение трех различных подходов к извлечению признаков: метод главных компонент, автокодировщик и группа методов статистического анализа [1] для проведения обработки данных. Данные типы методов позволяют решить задачу классификации технического состояния электромеханического объекта, но каждый из них имеет свои преимущества и недостатки.

Метод главных компонент представляет собой линейный метод снижения размерности, который преобразует исходные данные в новые координаты, упорядоченные по убыванию дисперсии. Это позволяет выделить важные особенности и уменьшить размерность данных без потери полезной информации. Автокодировщики представляют собой подвид нейронных сетей, которые обучаются восстанавливать входные данные на выходе. Таким образом возможно определить важные признаки и уменьшить размерность используемых данных без потери значимой информации. Статистические методы извлечения признаков позволяют вычислить различные статистические характеристики сигнала, такие как пиковое значение, среднеквадратичное значение, коэффициент пика, стандартное отклонение и другие. Данная группа методов позволяет получить характеристическое представление данных, которое возможно использовать в моделях машинного обучения для классификации технического состояния.

С учетом сложности и строгих требований к эксплуатации летательных аппаратов, возникает необходимость в интеграции современных методов анализа и классификации данных для обеспечения точной и своевременной диагностики технических объектов.¹

Для работы были сформированы задачи:

¹ Исследование выполнено частично за счет гранта Российского научного фонда (проект № 23-19-00464)

- Анализ и сравнение применения метода главных компонент, автокодировщика и группы методов статистического анализа на данных с использованием критериев оценки качества моделей классификации.
- Анализ и сравнение оценок точности решения задачи классификации технического состояния электромеханического привода на данных временных показателей для каждого из внутренних параметров.

Таким образом, это исследование направлено на изучение и сравнение различных методов извлечения признаков, что является важным шагом в направлении создания более надежных и безопасных систем диагностики в авиационной индустрии. В условиях быстрого развития технологий в области машинного обучения и анализа данных, это исследование способствует повышению безопасности и эффективности эксплуатации летательных аппаратов.

1. Обработка исходного набора данных

Исходный набор данных для сравнения методов извлечения признаков состоит из записей контролируемых параметров при отработке управляющего сигнала на математической модели электромеханического привода (ЭМП): 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). Данные представлены в формате временных рядов на Рис. 1 и в процессе моделирования изменялись внутренние параметры ЭМП: люфт, сухое и вязкое трение.

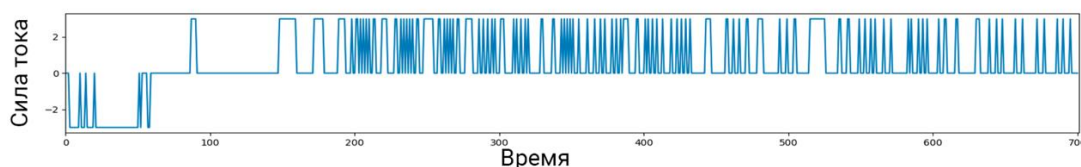


Рис. 1. Пример временного ряда показателя силы тока ЭД

В таблице 1 представлена матрица изменения внутренних параметров модели.

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 имеют свой тип класса состояния, ставящийся в соответствие элементу каждого внутреннего параметра. Для решения задачи классификации требуется построить алгоритмы оценки технического состояния ЭМП по всем типам классов технического состояния.

План эксперимента обработки данных для построения алгоритма классификации технического состояния представлен на Рис. 2

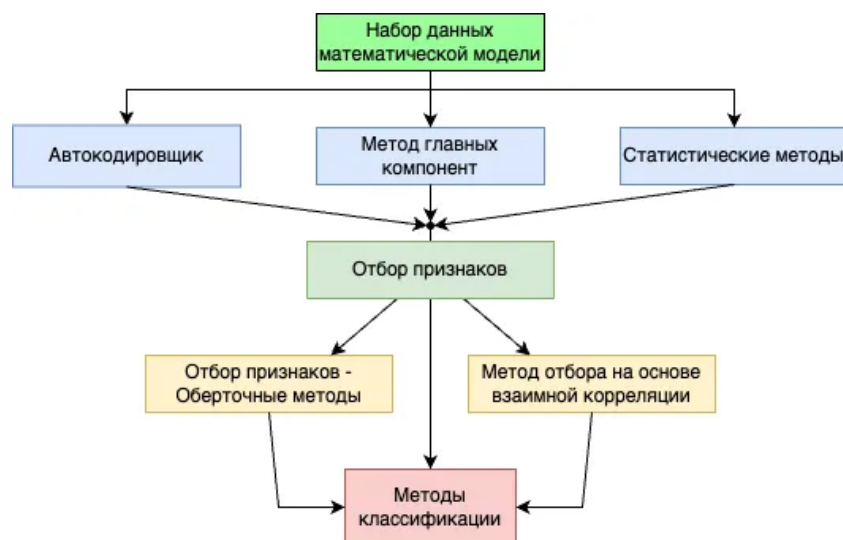


Рис. 2. Схема эксперимента работы с набором данных для решения задачи классификации состояния ЭМП

Решение задачи состоит из нескольких этапов: получение данных, создание набора векторов методами извлечения признаков, интеграция методов выделения информативных признаков и применение методов классификации.

В процессе обработки экспериментальных данных анализируемые сигналы были представлены в виде временных рядов. С использованием дополнительных преобразований были получены три набора векторов.

Первый набор векторов представляет собой набор признаков, полученный с использованием автокодировщика. Это достигается благодаря двум основным компонентам автокодировщика: энкодер и декодер. Энкодер преобразует входные данные в сжатое представление меньшей размерности. Декодер восстанавливает данные из сжатого представления обратно в исходное пространство. При обучении автокодировщика происходит минимизация функции потерь, которая представляет собой среднеквадратичную ошибку между входными и выходными данными. Цель состоит в обучении сети кодировать и декодировать данные таким образом, чтобы потеря информации была минимальной. Тогда на основе обученного автокодировщика возможно использовать сжатое представление, которое представляет собой уменьшенное количество нелинейных признаков изначального набора данных

Второй набор векторов получен с использованием метода главных компонент, который является одним из наиболее используемых методов снижения размерности данных. Он позволяет выделить основные направления вариации в данных, что упрощает анализ и визуализацию многомерности в данных. Данный метод является линейным методом, который преобразует исходные данные в новое пространство координат, упорядоченное по убыванию дисперсии. Для использования метода главных компонент требуется вычислить среднее значение каждого признака, что определения вариаций данных вокруг их среднего значения. Далее проводится вычисление ковариационной матрицы для оценки корреляции между различными признаками в данных и вычисление собственных вектор и собственных значений ковариационной матрицы для определения направления максимальной вариации и величины. Собственные вектора указывают направления главных компонент, а собственные значения показывают, какая доля общей дисперсии объясняется каждой главной компонентой. Последний шаг работы метода главных компонент связан с созданием вектора проекции данных для нового признакового пространства. Данный вектор проекций и будет использоваться для решения задачи классификации.

Третий набор данных представляет собой разнообразные статистические методы, направленные на вычисление характеристик сигналов. Эти методы просты в реализации и интерпретации, что делает их популярными в анализе временных рядов. Они позволяют получить информативные признаки, которые могут использоваться в моделях машинного обучения для классификации технического состояния. К данной группе методов относятся: пиковое значение, среднеквадратичное значение, коэффициент пика, стандартное отклонение ускорения, среднеквадратичное значение скорости, среднеквадратичное значение смещения, частота пиков, пик псевдоскорости, среднее изменение сигнала, дисперсия изменения сигнала, площадь под графиком сигнала, максимальное отклонение от

медианы. Применение статистических методов позволяет достаточно быстро выявить важные особенности сигналов и интерпретировать результаты. Пример такого вектора представлен на Рис. 3.



Рис. 3. Пример вектора набора данных, полученного с использованием группы статистических методов

По итогу работы методов извлечения признаков были сформированы наборы:

- Вектора, полученные методом автокодировщика, где каждый вектор имеет размерность 500 признаков.
- Вектора, полученные методом главных компонент, где каждый вектор имеет размерность 1200 признаков.
- Вектора, полученные методами статистического анализа, где каждый вектор имеет размерность 90 признаков

2. Методы отбора информативных признаков

Для ранней диагностики и классификации технического состояния электромеханического привода были использованы алгоритмы выделения информативных признаков, базирующиеся на трех ключевых типах [2-4]:

1. Методы фильтрации: χ^2 , корреляция Пирсона, абсолютное отклонение, взаимная информация, ти-тест, апова
2. Методы встроенные: дерево решений, случайный лес, extra tree, ridge

Применение этих методов к трем наборам данных привело к формированию множеств из 10 информативных признаков по каждому из методов. При этом варьировалось количество отбираемых признаков 1% и 10%. Ниже приведены примеры полученных диаграмм результатов работы моделей Рис. 5-6.



Рис. 4. Пример диаграммы уровня значимости признаков с использованием метода χ^2 по набору данных, сформированному с использованием метода главных компонент



Рис. 5. Пример диаграммы уровня значимости признаков с использованием метода дерева решений из набора данных, сформированном с использованием автокодировщика

Дополнительно к отобранным признакам фиксировалось время работы моделей для их сравнения (Таблица 2). Данное время работы алгоритмов замерялось по 1% и 10% данных векторов.

Таблица 2. Время работы методов отбора информативных признаков

Название модели	Время работы по 1% данных (с.)	Время работы по 10% данных (с.)
Хи ²	1.022675	11.051955
Корреляция Пирсона	10.065775	16.705694
Взаимная информация	4.395174	8.185104
Абсолютное отклонение	0.037700	0.025953
Ти-тест	0.207748	5.479003
Аноva	4.574227	10.301854
Дерево решений	0.064216	0.259153
Extra tree	0.029623	0.031091
Ridge	8.664022	10.262581
Случайный лес	0.023529	0.042969

3. Результаты моделей классификации

Для решения задачи классификации технического состояния электромеханического привода использовались модели классификаций [5-7]: логистическая регрессия, метод k-ближайших соседей, метод опорных векторов, дерево решений, нейронная сеть. На основе отобранных множеств признаков производилось обучение с учителем, где каждый вектор признаков был отнесен к одному из трёх классов состояния «исправное», «предаварийное» и «аварийное» для каждого типа неисправности.

Были получены следующие оценки точности наивысшим по F1 мере [8] на тестовых данных. В таблице представлена нижняя граница 50 наилучших результатов (Таблица 3). Полученные результаты представлены на Рис.6-8.

Таблица 3. Оценки точности классификации

Вид неисправности	Минимальная оценка 50 «наилучших» на 1%	Минимальная оценка 50 «наилучших» на 10%
Люфт	0.835828	0.870326
Сухое трение	0.927557	0.974017
Вязкое трение	0.844766	0.881618



Рис. 6. Количественное распределение методов извлечения признаков в группе из 50 наилучших результатов моделей классификации

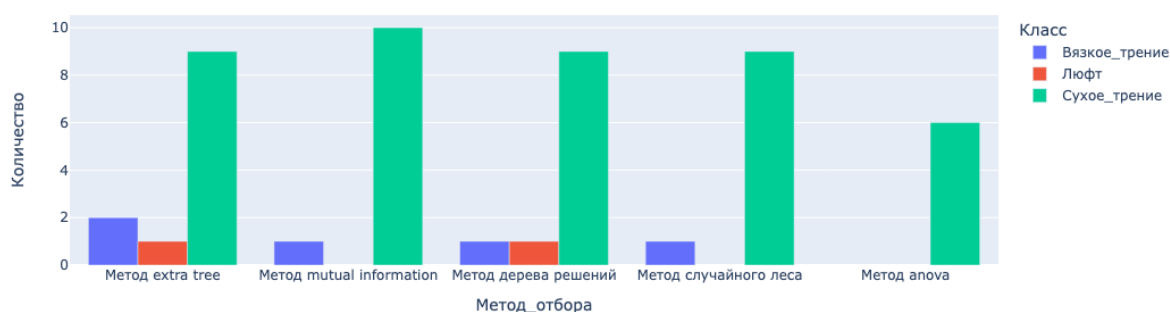


Рис. 7. Количественное распределение методов отбора признаков в группе из 50 наилучших результатов моделей классификации

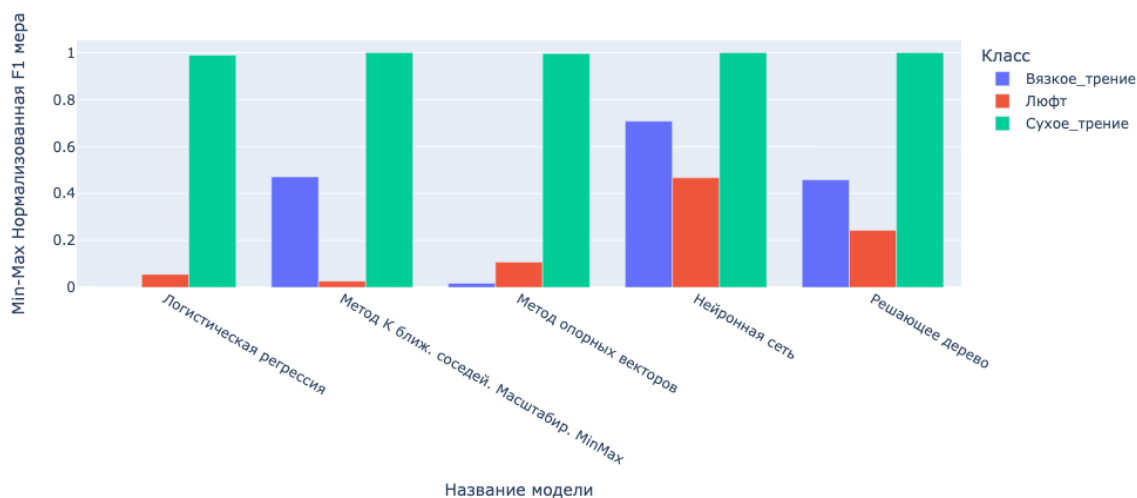


Рис. 8. График оценки распределения моделей классификации в группе из 50 наилучших результатов

Результаты показывают, что для обработки исходных данных предпочтительнее использовать метод главных компонент и методы из группы статистических функций. Для выделения информативных признаков предпочтительнее использовать встроенные методы, основанные на деревьях решений. Лучшим классификатором себя показали нейронные сети по каждому из видов неисправности.

4. Заключение

Выполнен анализ и сравнение эффективности методов извлечения признаков по данным различных показателей электромеханического привода летательного аппарата. Выполнено сравнение алгоритмов

выделения информативных признаков на сформированных наборах данных. Результаты проведенной работы показали, что применение методов извлечения признаков с выделением информативных признаков в задаче классификации технического состояния с использованием методов машинного обучения может обеспечить снижение размерности исходных данных и улучшить точность классификации. В частности, встроенные методы, показывают отличную оценку точности классификации.

Данная работа показывает возможность применения методов извлечения признаков для ранней диагностики технического состояния критически важных узлов летательного аппарата. Полученные результаты могут быть применены для создания более эффективных систем мониторинга и диагностики, что способствует повышению безопасности и надежности эксплуатации летательных аппаратов.

Литература

1. *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.
2. *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, pp. 856-863.*
3. *Мыльник В.В., Тутаренко Б.П., Волочиенко В.А.* Исследование систем управления. 2-е изд., перераб. и доп. // М.: Академический Проект, 2003. – 352 с.
4. *Уоссермен Ф.* Нейрокомпьютерная техника: теория и практика: Пер. с англ. // М.: Мир, 1992. – 157 с.
5. *Rokach L.* Data Mining With Decision Trees: Theory and Applications (2nd Edition) // *World Scientific Publishing Company/ 2014, P. 328*
6. *He Q. P., Wang J.* Fault Detection Using the k-Nearest Neighbor Rule for Semiconductor Manufacturing Processes // *IEEE Transactions On Semiconductor Manufacturing.* Vol. 20. 2007, № 4.
7. *Jayadeva, Khemchandani R., Chandra S.* Twin Support Vector Machines: Models, Extensions and Applications, // *Springer.* 2016, P. 211.
8. *Grzesica D., Wiecek P.* Advanced Forecasting Methods Based on Spectral Analysis // *Procedia Engineering.* – 2016. – Vol. 161. – P. 253–258.