

УДК 621.6

**АНАЛИЗ ЭФФЕКТИВНОСТИ МЕТОДОВ ОБУЧЕНИЯ
ПРОГРАММНО-АППАРАТНОГО КОМПЛЕКСА
АНАЛИТИЧЕСКОЙ ДИАГНОСТИКИ УСТАНОВОК
ЭЛЕКТРОЦЕНТРОБЕЖНЫХ НАСОСОВ (ПАКАД УЭЦН)**

¹Большунов А.В., ²Волков П.В., ¹Мостакалов К.А.

¹ФГАОУ «Национальный исследовательский Томский политехнический университет»,
Томск, e-mail: avb5@tpu.ru, mostokalov93@mail.ru;

²АО «ТомскНИПИнефть», Томск, e-mail: VolkovPV@nipineft.tomsk.ru;

В данной работе представлен анализ производительности машины экстремального обучения (ELM) по сравнению с классификаторами опорных векторов (SVM) и K-Nearest Neighbor (K-NN) для автоматической диагностики состояния оборудования электроцентробежных насосов. Испытания проводились с использованием 1412 реальных примеров, извлеченных из электрических погружных насосов. Извлечение сигнала вибрации было выполнено в лаборатории, и образцы были маркированы экспертами. Использовались две модели выделения признаков: статистические характеристики из области времени и частоты и амплитудные пики гармоник и субгармоник частоты вращения вала. Последовательный выбор объектов был применен для повышения производительности классификатора и уменьшения размерности набора данных. Экспериментальные результаты показывают, что ELM может использоваться в качестве алгоритма классификации в системах автоматической диагностики. В определенных сценариях ELM может превзойти SVM в отношении качества результатов и скорости обучения. Перед окончательным развертыванием системы диагностики произведено сравнение количества времени на обучение, возможности ошибки, количества рассматриваемых признаков для решения поставленной задачи. Для эффективной и качественной оценки было выполнено более пяти тысяч исследований, прогнозов и сравнений.

Ключевые слова: классификатор, скважина, дисбаланс, эффективность алгоритма, вибрационный сигнал, установка электроцентробежных насосов, УЭЦН, диагностика

**ANALYSIS OF EFFICIENCY OF TRAINING METHODS OF SOFTWARE
AND HARDWARE COMPLEX OF ANALYTICAL DIAGNOSTICS
OF ELECTRIC CENTRIFUGAL SUBMERSIBLE PUMP (PACAD ESP)**

¹Bolshynov A.V., ²Volkov P.V., ¹Mostokalov K.A.

¹National Research Tomsk Polytechnic University, Tomsk, e-mail: avb5@tpu.ru, mostokalov93@mail.ru;

²TomskNIPIneft, Tomsk, e-mail: VolkovPV@nipineft.tomsk.ru

This work presents a performance analysis of the Extreme Learning Machine (ELM) compared to the Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN) classifiers for automatic diagnosis of machine conditions. Tests were performed using 1 412 real examples extracted from electrical submersible pumps. The vibration signal extraction was executed in laboratory and the samples were labeled by experts. Two feature extraction models were employed, statistical features from the time and frequency domains and amplitude peaks of harmonics and subharmonics of the shaft rotation frequency. Sequential feature selection was applied to improve classifier performance and to reduce dataset dimensionality. Experimental results suggest that the ELM may be used as a classification algorithm in automatic diagnosis systems. In certain scenarios, the ELM can outperform SVM regarding the quality of results and training speed. Comparison of the amount of training time, the possibility of errors, the number of detected signs to solve the problem. For an effective and high-quality assessment, more than five thousand studies, forecasts and comparisons were performed.

Keywords: classifier, well, imbalance, algorithm efficiency, vibration signal, electric submersible pumps, ESP, diagnostics

Системы электрических погружных насосов (ЭЦН) часто используются в качестве метода искусственного подъема при морской добыче нефти. Системы ЭЦН работают внутри нефтяной скважины, и их установка и возможное удаление из-за технического обслуживания являются дорогостоящими операциями. Таким образом, чтобы избежать задних проблем на этапе эксплуатации, перед внедрением ЭЦН проводится тщательная оценка надежности [1]. Контролируемое обучение часто используется для разработки систем безмодельной диагностики [2]. Подход без модели имеет

то преимущество, что избегает явных экспертных знаний. Его основным недостатком является необходимость значительного количества полученных примеров. В [3] представлено исследование, в котором используются 1834 реальных примера, помеченных экспертами и двумя архитектурами классификаторов, K-ближайших соседей (К-БС) и поддержка опорных векторов (ПОВ), в качестве репрезентативных методов контролируемого обучения. Статистические и вейвлетные модели были использованы для описания состояния машины. В настоящей статье используются гармони-

ки и субгармоники амплитудных пиков вращения вала в качестве модельных признаков, помимо традиционных статистических характеристик, использованных в [4]. Большое количество образцов, необходимых для безмодельной системы диагностики, снабжено 1411 реальными примерами, также отмеченными экспертами.

Цель исследования: сравнить классификатор экстремальное машинное обучение (ЭМО) с К-БС и ООВ при применении системы диагностики. Изменять вид извлечения признаков для объективного сравнения производительности и обучаемости систем. Выяснить основные значимые критерии производительности.

*Диагностика погружных
масляных мотопомп*

На ранней стадии эксплуатации нефтяных скважин флюид естественным образом поступает на поверхность. Методы искусственного подъема используются в старых скважинах или для увеличения дебита из текущих скважин. Электропогружная насосная установка (ЭЦН), в которой используется погружной электродвигатель, приводящий в действие многоступенчатый центробежный насос, считается искусственным методом подъема. Система ЭЦН работает внутри скважины, а питание двигателя обеспечивается электрическим кабелем с поверхности [5]. На рис. 1 показана типичная схема системы ЭЦН. Тип ЭЦН, изучаемый здесь, имеет два двигателя, два протектора и два насоса.

Перед вводом системы УЭЦН в эксплуатацию ее необходимо тщательно проверить.

Акселерометры прикрепляются к системе в стратегических положениях до ее оценки в резервуар. Для УЭЦН, изученной здесь, 36 акселерометров соединены попарно с фазовым смещением на 90 градусов в осевом направлении, как показано на рис. 2, а. Акселерометры равномерно распределены вдоль двигателей, протекторов и насосов. Следовательно, одна пара акселерометров соединена внизу, в середине и сверху каждого компонента, как показано на рис. 2, б. Наконец, собранные данные анализируются и маркируются экспертом, используя в качестве основного метода визуальный контроль частотных спектров, полученных из преобразования Фурье необработанного сигнала вибрации.

Результаты, представленные здесь, были получены из 1411 образцов, помеченных как нормальные или несбалансированные, для каждой пары акселерометров. Априорное распределение классов составляет 92,35% для нормального режима работы и 7,45% для дисбаланса. Эти образцы были получены в лабораториях производителей ЭЦН, которые поставляют Petrobras – компания, спонсирующая этот проект. Функция эксперта заключается в предварительном просмотре срока службы системы ЭЦН. Настоящая работа является частью проекта, целью которого является предоставление эксперту вспомогательной информации о качестве системы для облегчения процесса принятия решений. Следовательно, желательно, чтобы выходные данные системы автоматической диагностики имели сильную корреляцию с мнением эксперта.

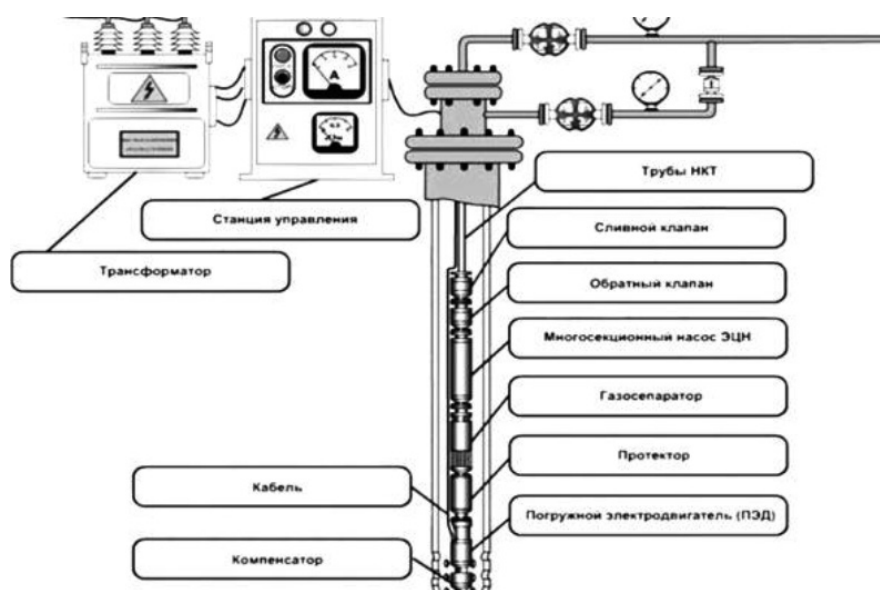


Рис. 1. Схема УЭЦН

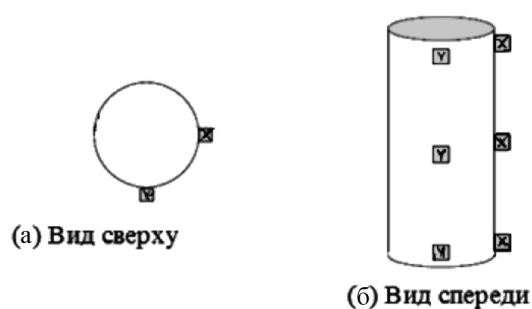


Рис. 2. Схема подключения акселерометра

Особенности экстракционных моделей

Эта работа использует методологию обучения под надзором [6–8], чтобы создать систему диагностики неисправностей без модели. Особенности должны быть извлечены из сигналов, полученных в ходе лабораторных испытаний. Вибрационные сигналы, собираемые акселерометрами, широко используются при автоматической диагностике отказов вращающихся механизмов [9–11]. Как для эксперта, так и для автоматической системы диагностики, сигналы, собранные от оборудования, не могут быть непосредственно использованы для принятия решения, поэтому необходимо извлечь описательные признаки.

В этой работе для экспериментов использовались две основные модели методов выделения признаков. Во-первых, статистическая модель применяется во временной и частотной области. Во-вторых, пики частоты вращения вала, его гармоник и субгармоник используются в качестве признаков, извлеченных из частотной области. Эксперты используют пару акселерометров X и Y со смещением фазы на 90 градусов для маркировки образцов, чтобы избежать пропуска компонентов ортогональной вибрации. Таким образом, стремясь предоставить те же метки, что и эксперты, каждый образец также состоит из элементов, извлеченных из двух акселерометров.

В предварительных экспериментах по отбору признаков модель признаков пика гармоник показала наивысшую дискриминационную способность разделять классы при нормальной работе и дисбаланс по сравнению со статистическими признаками. Вычисление для нахождения пика определенной частоты вычислительно недорого, но необходимо знать частоту вращения вала анализируемого сигнала. Частота вращения вала контролируется оборудованием, называемым приводом с регулируемой скоростью (ПРС) [5], которое управляет источником питания систе-

мы ЭЦН. Следовательно, частота вращения вала является параметром, влияющим на частоты, на которых можно ожидать пиков. Из-за естественного трения компонентов системы ЭЦН реальная частота вращения вала часто ниже, чем у соответствующего источника питания, примерно на 5%. Способ эмпирически решить эту проблему – найти пик в окне, который на 7% ниже, чем источник питания. Недостаток этого решения заключается в том, что в редких случаях сигнал может быть загрязнен источником питания, что приводит к пику. Если результирующий пик выше, чем пик, соответствующий скорости вращения, извлечение будет осуществляться на основе источника питания, а не на основе вращения вала. Таким образом, гармоники и субгармоники не будут извлекать полезную информацию для выявления сбоя. Чтобы избежать этой проблемы, пики были взяты из ожидаемой полосы частот, пониженной на 0,5 Гц. Например, если источник питания 40 Гц, 1.

На рис. 3 показан пример спектрального графика, используемого экспертами. Данные о вибрации, используемые для генерации этого спектра, были собраны из системы ЭЦН с частотой вращения 45 Гц. Восемь гармоник представлены на этом рисунке и отмечены вертикальными пунктирными линиями. Высокий пик, близкий к 45 Гц, указывает на проблему дисбаланса [4]. Можно заметить, что реальная частота этого пика немного ниже, чем у источника питания 45 Гц. Также можно отметить появление высоких амплитудных значений вблизи гармоник в 90, 135, 270 и 315 Гц. Эти значения предполагают, что 7-я и 8-я гармоники могут иметь потенциально полезную информацию. Это первоначальная мотивация использовать гармоники за пределами основной частоты. Высокие субгармоники между 45 и 90 Гц и 90 и 135 Гц также могут наблюдаться. Это мотивирует дополнительное извлечение субгармоник.

На рис. 4 представлен пример спектрального графика для нормального состояния, извлеченного из системы ЭЦН с частотой вращения 50 Гц. Сравнивая рис. 3 с рис. 4, можно видеть, что второй не имеет высоких пиков. Однако в обоих случаях некоторые пики могут быть найдены около 1,0х, его гармоники и субгармоники, как, например, можно наблюдать в 7,0х и 1,5х.

Архитектура и алгоритм обучения системы опорных векторов (ООВ) [6] создают гиперплоскость разделения максимального предела между двумя классами.

Чтобы улучшить линейное разделение в исходном евклидовом пространстве, ООВ отображает входные векторы в многомерное пространство признаков посредством некоторого нелинейного отображения [9], используя функцию ядра, которая реализует внутреннее произведение индивидуально отображенных шаблонов, обычно не доступны по своему усмотрению. Чтобы классифицировать более двух классов, можно использовать повторный подход. Классификатор ООВ до сих пор считается одним из самых мощных методов двоичной классификации [4]. В этой работе используется архитектура классификации С-ООВ радикальная базисная функция (РБФ), реализованная с помощью библиотеки LIBSVM. Параметры ООВ были установлены следующим образом, основываясь на предыдущих экспериментах: γ в функции ядра был установлен на 0,013423; стоимость C С-ООВ была установлена на 1,0; и допуск критерия прекращения был установлен на 0,001.

Оценка эффективности

Для сравнения эффективности алгоритмов классификации была использована наиболее распространенная мера, а именно оценочная точность. Разбивая имеющиеся данные на обучающие и тестовые наборы данных, точность – это доля выборок, правильно классифицированных в количестве выборок тестового набора данных. k -кратная оценка эффективности перекрестной проверки разделяет данные D на k приблизительно равных частей D_1, D_2, \dots, D_k и изучает их с уменьшенным набором данных, с одной пропущенной частью. Неиспользованная часть D_i используется

в качестве тестового набора [8]. Частным случаем перекрестной проверки является исключение одного, где число сгибов k равно значению количества выборок. Поскольку оценочная точность используется в качестве критерия эффективности в этой статье, общая точность перекрестной проверки оценивается как среднее значение k кратных как:

$$ACC_{global} = \frac{1}{k} \sum_{i=1}^k ACC_i. \quad (1)$$

Выбранный алгоритм является сильным алгоритмом выбора неоптимального признака, а именно последовательного прямого выбора (ППВ), который является хорошим компромиссом между достаточным исследованием пространства поиска и вычислительными затратами. Чтобы выбрать d из общего числа Q объектов, ППВ инициализируется с пустым набором объектов $Y \leftarrow \emptyset$. Объекты итеративно добавляются в Y , согласно некоторому критерию выбора. Алгоритм останавливается и возвращает Y , когда $|Y| = d$. Эксперименты, выполненные в этой статье, использовали в качестве критерия оценочную среднюю вероятность ошибки, полученную при 10-кратной перекрестной проверке или исключении одного. ППВ с оцененной средней вероятностью ошибки выполняет полную оценку производительности для каждого доступного признака $f_i \in X, f_i \notin Y$ с набором кандидатов $(f_i \notin Y) \cup Y, i \in \{1, \dots, Q\}$, где X – полный набор всех функций. Та особенность, которая увеличивает критерий производительности больше всего, присоединяется к Y . Связи решаются произвольно.

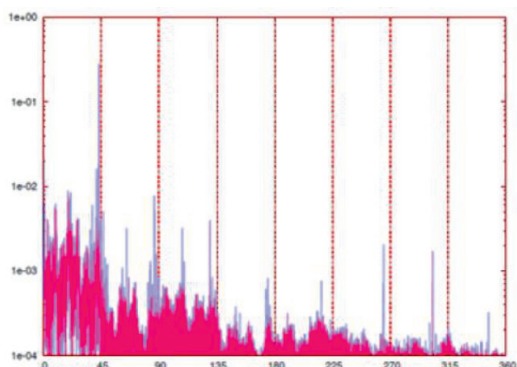


Рис. 3. Пример спектра для ошибки дисбаланса. Горизонтальная ось представляет частоту, а вертикальная ось представляет амплитуду в логарифмическом масштабе

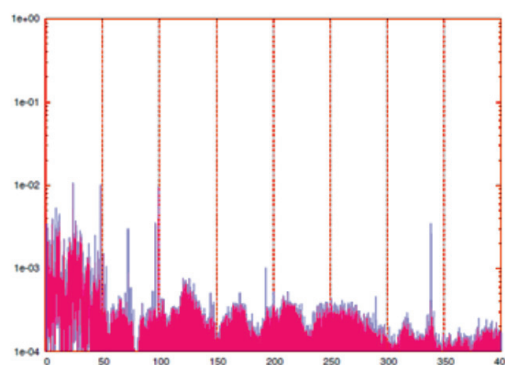


Рис. 4. Пример спектра для нормального состояния. Горизонтальная ось представляет частоту, а вертикальная ось представляет амплитуду в логарифмическом масштабе

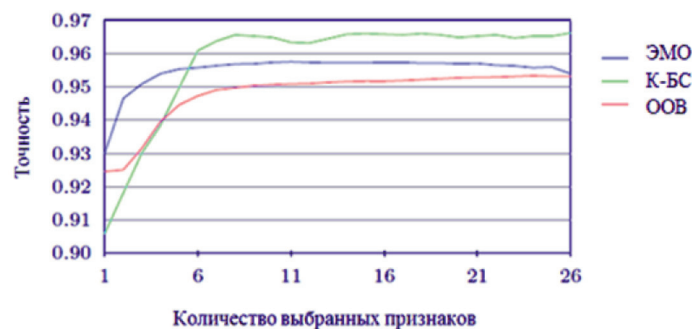


Рис. 5. Производительность классификаторов для каждого числа признаков, выбранных алгоритмом ППВ со статистическими признаками

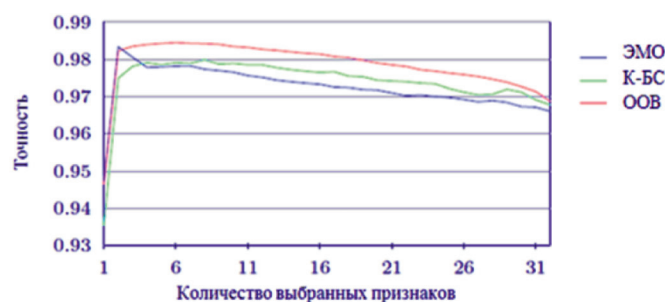


Рис. 6. Производительность классификаторов для каждого числа признаков, выбранных алгоритмом ППВ с гармоническими пиками

Экспериментальные результаты

Эксперименты были проведены с использованием 1413 реальных примеров, где признаки были извлечены двумя методами экстракции. В целях уменьшения дисперсии эксперименты, выполненные с этими классификаторами, были повторены 10 раз. Каждое значение, представленное на рис. 5 и 6, является средним из этих 10 повторений. К-БС работал очень быстро с использованными данными о сбоях, поэтому выбор его функций выполнялся с помощью перекрестной проверки без остановки. На рис. 5 и 6 показана средняя точность каждого классификатора с использованием статистических и гармонических пиковых характеристик соответственно, в соответствии с количеством выбранных признаков.

Как показано на рис. 5, при использовании статистических функций, когда выбрано от одной до пяти функций, ЭМО работает лучше, чем ООВ и К-БС. На самом деле, ЭМО работает лучше, чем ООВ, в течение всего процесса выбора функций. С другой стороны, простой алгоритм 1-БС превосходит оба сложных алгоритма с более чем пятью функциями. Поведение, показанное на рис. 5, с использованием модели при-

знаков пика гармоника не повторялось на рис. 6. Здесь алгоритм К-БС работает лучше, чем модель статистической функции, но не дает лучших результатов, чем ООВ с любым количеством функций. ЭМО работает лучше, чем ООВ, для одной и двух функций, достигая максимальной точности с двумя функциями. Во всех подходах этого эксперимента были выбраны две функции: гармоника 1,0х, собранные каждым акселерометром. Таким образом, наилучший результат с учетом максимальной расчетной точности был достигнут с помощью ООВ, а конкретный лучший результат с минимальным количеством функций был достигнут с помощью ЭМО. Помимо наблюдаемых, различия производительности, очевидно, статистически незначимы.

Выводы

Классификатор ЭМО не всегда работает лучше, чем К-БС и ООВ, но его можно рассматривать как вариант перед окончательным развертыванием системы диагностики из-за его простоты внедрения и небольшого количества времени на обучение. Варьируя вид извлечения признаков, его производительность может быть выше, чем у К-БС или ООВ, особенно когда применя-

ются методы выбора признаков. Эти эксперименты показывают, что методика выбора характеристик улучшает результаты ЭМО. Из-за непропорционального количества выборок каждой метки в будущей работе будут использоваться другие критерии производительности, которые могут иметь дело с этим набором неисправностей, где нормальный класс является доминирующим в предоставленных выборках, например РПЦ-АУК. Также будут исследованы другие методы извлечения элементов, области применения и методы выбора признаков.

Список литературы

1. Большунов А.В. Современные методы диагностики фактического состояния установок электроцентробежных насосов // Успехи современного естествознания. 2019. № 12. С. 249–253.
2. Трупова В.А. Система дифференциальных уравнений в пространстве одномерных касательных элементов второго порядка // Вестник ИрГТУ. 2010. № 6. С. 295–298.
3. Большунов А.В., Волков П.В., Зятиков П.Н. Экспериментальное исследование аэромеханического метода для предотвращения гидратообразования при подготовке и транспортировке природного газа и конденсата // Успехи современного естествознания. 2018. № 9. С. 52–56.
4. Dmitriev V.V., Gandzha T.V., Dolganov I.M., Aksenova N.V. An algorithm to improve the speed and accuracy of analysis of chemical process systems operation. *Petroleum and Coal*. 2017. vol. 59. I. 3. P. 429–441.
5. Такач Г.Н. Руководство по электрическим погружным насосам: проектирование, эксплуатация и техническое обслуживание // Технический журнал. 2012. № 1. С. 33–36.
6. Maystrenko A.V., Svetlakov A.A., Gandzha T.V., Dmitriev V.V., Aksenova N.V. Application of numerical signal differentiation methods to determine stationarity of A process. *Petroleum and Coal*. 2017. vol. 59. I. 3. P. 311–318.
7. Крюков О.В., Серебряков А.В. Превентивный мониторинг мощных электрических машин энергетических объектов для технического обслуживания по реальному состоянию // Приборы и системы. Управление, контроль, диагностика. 2017. № 8. С. 50–55.
8. Большунов А.В., Мостакалов К.А. Методика диагностики фактического состояния и неисправностей установок электроцентробежных насосов (УЭЦН) в реальном времени с использованием промышленных, электротехнических и геологических параметров // Успехи современного естествознания. 2016. № 10. С. 15–17.
9. Мендель Э.Н., Константинов В.И. Диагностика множественных неисправностей в насосах моторов нефтяной вышки с использованием ансамблей классификаторов машин опорных векторов // Интегрированное автоматизированное проектирование. 2017. № 18. С. 61–74.
10. Елисеев А.В. Введение в переменную и выбор функций // Журнал исследований машинного обучения. 2010. № 3. С. 157–182.