

СИСТЕМА ПРЕДИКТИВНОЙ АНАЛИТИКИ ДЛЯ ЭНЕРГЕТИЧЕСКИХ ОБЪЕКТОВ

Андрюшин А.В., Щербатов И.А., Цуриков Г.Н., Титов Ф.М.

Московский энергетический институт, Россия, г. Москва ул. Красноказарменная д.14

AndriushinAV@mpei.ru, ShcherbatovIA@mpei.ru, grishatsurikov9826@yandex.ru,

fedor.titov.335@gmail.com

Аннотация: Показана необходимость прогнозирования значений технологических параметров в рамках решения задачи прогнозирования технического состояния энергетического оборудования для целей организации ремонтов по состоянию. Рассмотрена архитектура и методы предобработки больших объемов данных и прогнозирования значений параметров. Показана эффективность искусственных нейронных сетей при решении данных задач.

Ключевые слова: предиктивная аналитика, техническое состояние, предобработка данных, прогноз, искусственная нейронная сеть, энергетика.

Введение

При разработке система предиктивной аналитики (СПА) для прогнозирования в основном используются характеристики, основанные на экспертных знаниях, сформулированных в процессе эксплуатации оборудования эвристиках, а также предположении о том, что динамика изменения ряда параметров свидетельствует об отказе оборудования с определенной уверенностью через некоторый интервал времени. Неоднородность больших данных (Big Data) [1] и методы их обработки неразрывно связаны при реализации СПА.

В систему предиктивной аналитики поступает неструктурированная телеметрическая информация в реальном масштабе времени от различных информационных и автоматизированных система, таких как АСКУЭ и SCADA, а также учетных систем (например, фиксирующих дефекты, результаты диагностики и пр.), MES, ERP и EAM. Использование накопленной информации в скором будущем должно обеспечить энергетике качественный скачок вперед [2]. Для большинства систем предиктивной аналитики характерна следующая структура [3]: импорт данных в систему; преобразование и подготовка данных; кластеризация для выявления тех данных, которые поступают затем в предиктивную модель; прогнозирование трендов и отказов на основе предиктивной модели.

Проведенный критический анализ литературных источников в области обработки больших данных и временных рядов для целей реализации оценки состояния оборудования и прогнозирования его надежности позволяет сделать вывод о целесообразности разработки оригинального подхода создания систем предиктивной аналитики для предприятий энергетики.

1 Структура системы предиктивной аналитики

Качество прогнозирования технического состояния оборудования энергетики и эффективность функционирования других систем предприятия, использующих данные прогнозов, зависит от архитектуры и функциональности системы предиктивной аналитики.

Одна из ключевых подзадач при создании СПА – предобработка данных, целью которой является получение генеральной совокупности данных как основы для синтеза или коррекции предиктивной модели, а также получения непосредственно прогноза. Поступающие данные должны быть по возможности отфильтрованы от помех, шумов и ошибок, а также привязаны к реальному масштабу времени. Совокупности данных должны быть равномерными, нормализованными, а в ряде случаев их требуется получить с использованием специальных методов, например, когда данных не достаточно для реализации предиктивной модели или ее корректировки. Для реализации СПА разработана структура, содержащая основные модули, реализующие один из ключевых подходов Индустрии 4.0 [4] кибер-физические объекты (рис. 1).

На структурной схеме представлены основные системы и подсистемы, а также информационные потоки и потоки данных, обеспечивающие прогнозирование трендов ключевых параметров, технического состояния оборудования энергетики и принятия решений.

Согласно концепции кибер-физических систем данные от датчиков, установленных на оборудовании, поступают в соответствующие информационные системы (ИС), к которым можно отнести SCADA, MES, PLM, EAM. Эти данные доступны для ЛПР, решающего задачу постановки глобальной цели системы энергетического оборудования. Данные информационные системы обеспечивают поддержку принятия решений при управлении жизненным циклом производственных активов предприятий энергетики, оптимизации генерации или транспорта энергии и т.д.

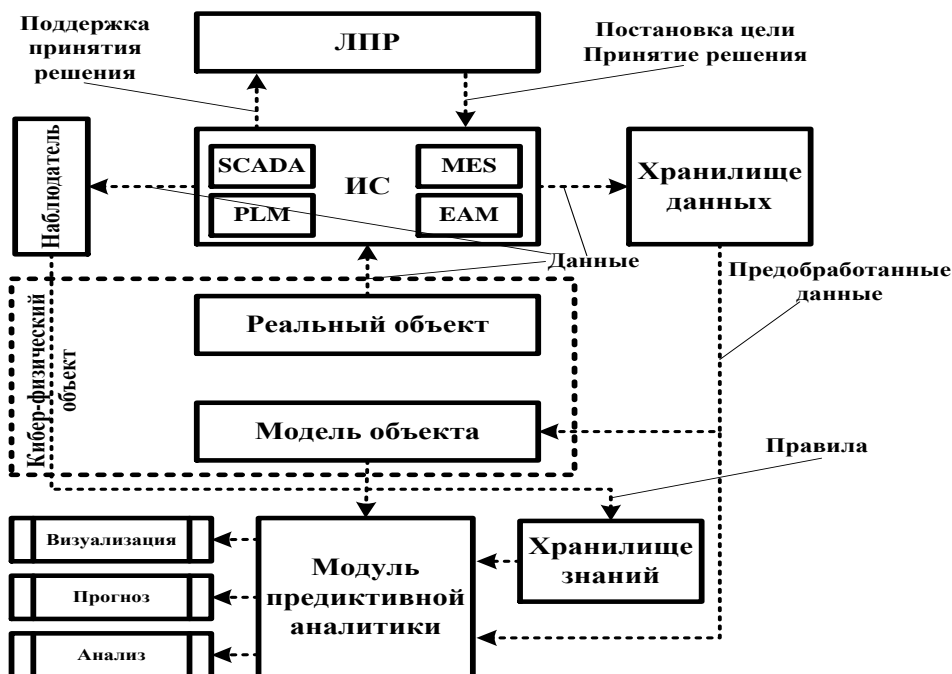


Рис. 1. Структура системы предиктивной аналитики

Таким образом, разрабатываемая СПА является частью общего подхода по цифровизации предприятий энергетики, которая в свою очередь направлена на повышение эффективности их функционирования. Система предиктивной аналитики на основе поступающих данных и имеющихся экспертных знаний обеспечивает прогнозирование технического состояния оборудования, характерных показателей и трендов, их интерпретацию и анализ, а также визуализацию происходящих в оборудовании процессов.

2 Сбор данных в системе предиктивной аналитики

Данные, поступающие в систему предиктивной аналитики, влияют на ряд аспектов, обуславливающих ключевые свойства ее применения: эффективность прогнозирования как мера минимизации затрат на поддержание оборудования энергетики в надлежащем техническом состоянии (обеспечении заданной надежности); качество прогнозирования как объективная характеристика возможности использования полученного прогноза для управления оборудованием или принятия решений на уровне предприятия.

Структурная схема сбора данных системы предиктивной аналитики представлена на рис. 2 (ЕО – единица оборудования энергетики, КО – компонента крупномасштабной энергетической системы [5]).

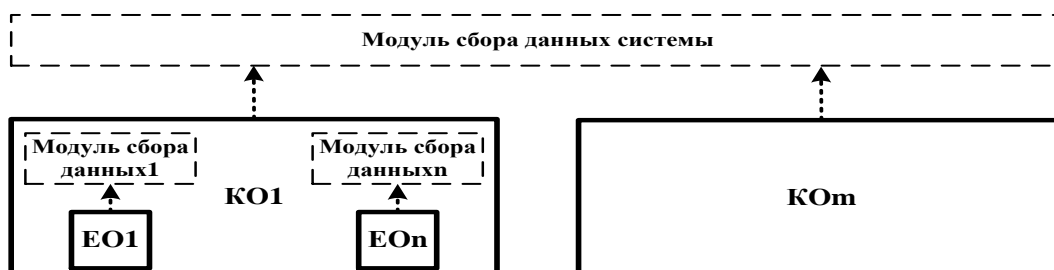


Рис. 2. Сбор данных системы предиктивной аналитики

Необходимо отметить, что согласно концепции промышленного интернета вещей целесообразно обеспечить максимальную унификацию аппаратных средств и протоколов передачи данных.

3 Предобработка данных

Рассмотрим наличие аномалий в измерительных данных систем предиктивной аналитики. Первый случай – наличие пропусков в получаемых данных. В литературных источниках описано большое число методов и алгоритмов восстановления данных при наличии пропусков [6-8]. Общая черта данных методов – заполнение пропущенного одного или нескольких значений.

Решение проводилось для нескольких методов (Cubic / Liner SVM и искусственные нейронные сети с алгоритмом обучения Левенберга-Марквардта). Результаты представлены в табл. 1.

Таблица 1. Результаты экспериментов по применению алгоритмов восстановления пропусков

Пропуски	Регрессионные методы		Нейронная сеть	
	Алгоритм машинного обучения	СКО (RMSE)	Алгоритм машинного обучения	СКО (RMSE)
1 столбец (10%)	Cubic SVM	0.27517	Levenberg-Marquardt	0.82636
1 строка (10%)	Liner SVM	0.37958	Levenberg-Marquardt	1.64428
3 столбца (30%)	Liner SVM	0.31582	Levenberg-Marquardt	1.1475
3 строки (30%)	Liner SVM	0.19736	Levenberg-Marquardt	4.8081
5 столбцов (50%)	Liner SVM	0.35404	Levenberg-Marquardt	2.3039
5 строк (50%)	Linear Regression	0.38985	Levenberg-Marquardt	2.6446

Результаты свидетельствуют о целесообразности применения регрессионных моделей для решения задачи восстановления пропусков в данных.

Следующей не менее важной задачей по предобработке больших массивов данных для СПА является необходимость выявления и устранения выбросов (аномально больших или малых значений технологических параметров, присутствующих в генеральной совокупности). Выброс – элемент, существенно отличающийся от других элементов выборки.

Рассмотрим примеры решения задачи определения выбросов в данных, поступающих от средств измерений объектов энергетики в систему предиктивной аналитики. В качестве базового математического аппарата выберем искусственные нейронные сети. Обучающая выборка содержит информацию о температуре окружающего воздуха (X_1), давлении в конденсаторе паровой турбины (X_2) и значениях мощности парогазовой установки при данных параметрах. Для выборки, после представления данных в трехмерном пространстве, отобрано 100 элементов, среди которых визуально легко определить значения, заметно отличающиеся от остальных. Для анализа точности определения выбросов с использованием разработанного способа первоначально выбросы были определены на основании методики из ГОСТ Р ИСО 16269-4-2017 (значения параметра мощность, определенные как выброс кодировались единицей, в противном случае нулем).

Выбросы значений для параметра мощность (Y) идентифицируются путем нахождения студентизированной ошибки и ее последующим сравнением с показателем распределения Стьюдента (при доверительной вероятности 0.995, числе степеней свободы $k = n - p - 2$, где n - количество элементов в выборке, p - количество переменных, влияющих на значения Y).

Для определения потенциальных выбросов по переменным X используется параметр h , характеризующий расстояние между значением, принимаемым случайной величиной X в i -й точке и средним арифметическим всех n значений, принимаемых X . Элемент выборки считается выбросом, если значение данного параметра меньше, чем результат выражения $h \leq 2(p+1)/n$.

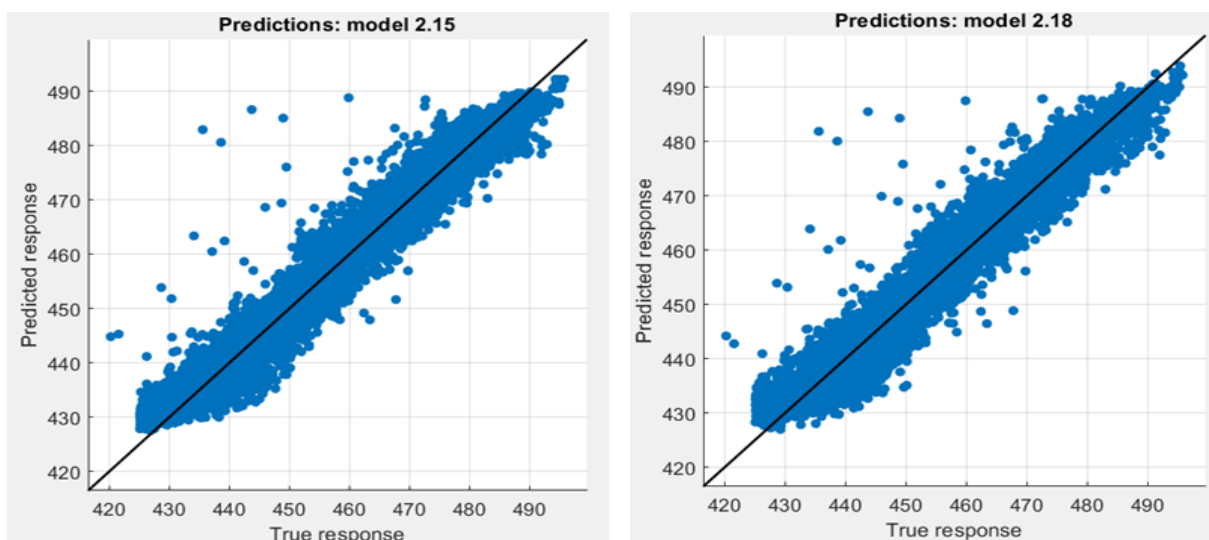
Потенциальный выброс считается реальным если расстояние Кука и показатель DFFITS принимают определенные значения. Например, расстояние Кука сравнивается с показателем F-распределения (перцентилем уровня, равным 50%, значением $p+1$, значением $n-p-1$), а показатель DFFITS с результатом выражения $2\sqrt{(p+1)/n}$.

Подготовленные таким образом данные использовались для обучения и тестирования искусственной нейронной сети. Результат идентификации выбросов составил 100%, т.е. все выбросы были определены верно. Среди 100 элементов выборки 9 выбросов идентифицированы верно, также как и 91 элемент, который к выбросам не относится.

И, наконец, заключительным этапом является нормализация данных.

4 Прогнозирование значений параметров

В качестве примера прогнозирования значений параметров покажем прогнозирование мощности Y в зависимости от значений X_1 и X_2 . Для выбора наиболее подходящего метода прогнозирования были исследованы искусственные нейронные сети, SVM и алгоритм бэггинга деревьев принятия решений. На рис. 5 представлены результаты прогнозирования.



а) метод бэггинга деревьев решений

б) SVM-метод

Рис. 5. Результаты прогноза с помощью различных методов

При использовании алгоритма бэггинга деревьев решений и SVM-алгоритма, среднеквадратичное отклонение прогнозируемой величины составило 3.4487 и 3.5648 МВт соответственно, а применение искусственных нейронных сетей показало результат 3.4271 МВт. При этом именно нейронные сети показали наименьшее время получения результата.

Заключение

Разработана структура системы предиктивной аналитики для крупномасштабных энергетических систем. Прогностические возможности системы могут использоваться для прогнозирования значений различных технологических параметров, которые являются характеристиками технического состояния оборудования энергетики. Решена задача подготовки данных для использования в системах предиктивной аналитики. В качестве базового математического аппарата применяется высокоэффективные методы искусственных нейронных сетей, которые обеспечивают высокую точность прогнозирования при требуемой скорости получения прогноза.

Литература

1. Wang J., Zhang W., Shi Y., Duan S., Liu J. Industrial Big Data Analytics: Challenges, Methodologies, and Applications. 2018 : [Электронный ресурс] Режим доступа : <https://arxiv.org/ftp/arxiv/papers/1807/1807.01016.pdf>.
2. Chidambaram V., Evans H., Etheredge K. Big Data: Is the Energy Industry Starting to See Real Applications? // Supply Chain Management Review. 2015. No. 12. P. 62-64.
3. Silipo R., Winters P. Big Data, Smart Energy, and Predictive Analytics Time Series Prediction of Smart Energy Data. 2013. 2009. [Электронный ресурс]. - Режим доступа: https://files.knime.com/sites/default/files/inline-images/knime_bigdata_energy_timeseries_whitepaper.pdf.
4. Цуриков Г.Н., Щербатов И.А. Применение промышленного интернета вещей на объектах энергетики // Мехатроника, автоматика и робототехника. 2018. № 2. С. 97-100.
5. Щербатов И.А. Управление сложными слабоформализуемыми многокомпонентными системами. - Ростов н/Д: ЮНЦ РАН, 2015. - 288. с.
6. Литтл Р.Дж.А., Рубин Д.Б. Статистический анализ данных с пропусками. - М.: Финансы и статистика, 1990. - 336 с.

7. Злоба Е., Яцкие И. Статистические методы восстановления пропущенных данных // Computer Modelling & New Technologies. -2002. - Vol. 6.- № 1. - Pp. 51-61.
8. Абраменкова И.В., Круглов В.В. Методы восстановления пропусков в массивах данных // Программные продукты и системы. 2005. № 2. С. 18-22.