

УДК 007.51:007.3



К.С. Серегичев, А.А. Сукончиков
Вологодский государственный университет

ПРОГНОЗИРОВАНИЕ ЭЛЕКТРОПОТРЕБЛЕНИЯ С ПОМОЩЬЮ ВЫДЕЛЕНИЯ ФУНКЦИЙ ИЗ ВРЕМЕННОГО РЯДА

В статье рассматриваются вопросы прогнозирования электропотребления промышленных предприятий. Исследование проводится с использованием библиотеки `tsfresh` на языке `python`. Особенность данной библиотеки – автоматическое извлечение функций и признаков из временного ряда для дальнейшего использования в модели машинного обучения.

Прогнозирование, машинное обучение, временные ряды.

Организация АО «НПП Энергопромсервис» занимается внедрением на предприятия систем учета и диспетчеризации энергоресурсов. В настоящий момент реализовано более 500 проектов по автоматизации в данном направлении. Взаимодействуя с заказчиками как по внедренным системам, так и касательно предприятий, где мы только планируем начать работу, мы регулярно сталкиваемся с запросом о необходимости краткосрочного прогнозирования потребления электроэнергии для повышения эффективности управления энергохозяйством и, как результат,

снижения стоимости потребляемой электроэнергии на ее розничном рынке для предприятия.

В предыдущей статье [1] было проведено исследование возможностей такого прогнозирования с помощью моделей, скользящего среднего, авторегрессии и нейронных сетей. Наилучший результат был достигнут с использованием алгоритма скользящего среднего SARIMAX и составил 14,6 % средней абсолютной ошибки при прогнозировании на массиве данных за несколько месяцев, что является недостаточным для поставленных целей.

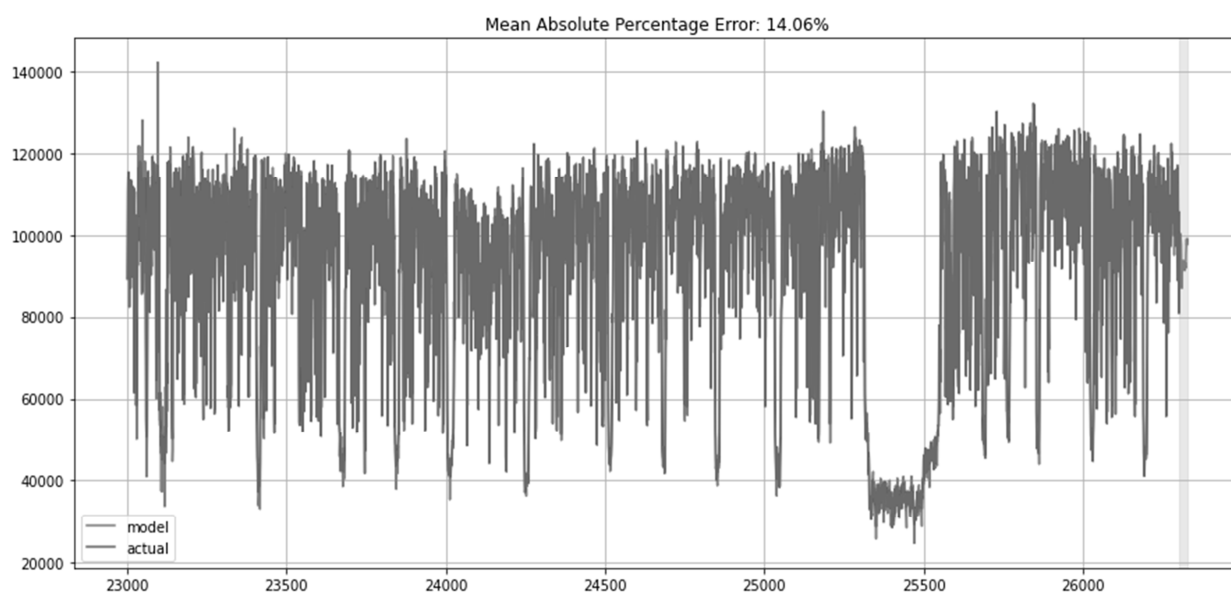


Рис. 1. График целевой переменной и предсказания, наложенные друг на друга, с выводом величины ошибки над графиком

Исходные данные для прогнозирования

	time	P, кВт
0	2021-07-20 09:00:00	26.210
1	2021-07-20 10:00:00	25.596
2	2021-07-20 11:00:00	23.652
3	2021-07-20 12:00:00	25.469
4	2021-07-20 13:00:00	24.318

Так как потребление электроэнергии на крупных предприятиях является коммерческой тайной, их регулярное получение и использование в исследованиях затруднено. Для преодоления данного момента было проведено сравнительное исследование формы и характера потребления на металлургическом предприятии и в нашей компании. Выяснилось, что в обоих случаях имеют место быть формы потребления в виде ступенек высокого и низкого потребления, суточная, недельная и месячная цикличность, достаточно резкие переходы между моментами низкого и высокого потребления электроэнергии.

На основании этого было принято решение о возможности предварительных исследований на данных о потреблении электроэнергии в помещениях, занимаемых группой компаний АО «НПП Энергопром-сервис», ООО «Мависмарт», АО «ПЛЦ АС». Занимаемый ими этаж имеет собственное ВРУ с 2-мя вводами, которое охватывается собственной системой автоматического учета и диспетчеризации энергоресурсов. Здесь находятся рабочие места более 100 сотрудников, а также в качестве потребителей электроэнергии имеются стенды для проверки, наладки, проверки различного измерительного оборудования и средств автоматизации; климатическая система; центральная серверная, обеспечивающая облачные решения для взаимодействия внутри компаний, а также демо-комплекс внедряемых систем. В качестве основных исходных данных будут использованы усредненные за 1 час значения полной мощности по обоим вводам ВРУ. Прогнозирование будет выполнено только на основе предыдущих значений этого же временного ряда.

Прогнозирование временного ряда только на основе его предыдущих значений – сложная задача. Для ее решения обычно создают вспомогательные временные ряды на основе функций, выделенных из основного. Извлеченные таким образом функции используются для описания временного ряда, т.е. часто эти функции дают новое представление о временном ряде и его динамике. И далее их можно использовать для обучения моделей машинного обучения.

Современные достижения в области машинного обучения позволяют автоматизировать процесс извлечения функций и признаков из временного ряда для последующего использования в моделях машинного обучения. В нашем исследовании для автоматизации решения указанной задачи воспользуемся библиотекой `tsfresh`. Исследования с использованием данной библиотеки регулярно проводятся в мировом научном сообществе [2–4]. С ее помощью для наших исходных данных мы сможем дополнительно получить более 300 значимых функций из порядка 1300 доступных. В таблице 2 представлены некоторые из них.

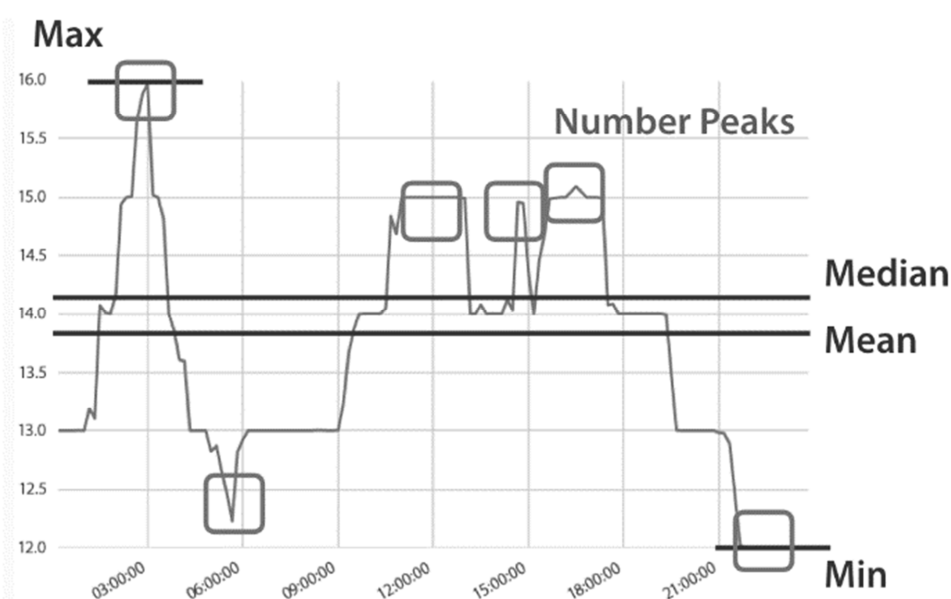


Рис. 2. Визуализация принципа выделения функций из временного ряда

Список некоторых функций, поддерживаемых библиотекой tsfresh

№	Наименование	Описание функции
1	<code>abs_energy</code> (x)	Возвращает абсолютную энергию временного ряда, которая представляет собой сумму квадратов значений.
2	<code>absolute_maximum</code> (x)	Вычисляет наибольшее абсолютное значение временного ряда x.
3	<code>absolute_sum_of_changes</code> (x)	Возвращает сумму по абсолютному значению последовательных изменений в серии x.
4	<code>agg_autocorrelation</code> (x, параметр)	Описательная статистика автокорреляции временных рядов.
5	<code>agg_linear_trend</code> (x, параметр)	Вычисляет линейную регрессию методом наименьших квадратов для значений временных рядов, которые были агрегированы по фрагментам, в сравнении с последовательностью от 0 до количества фрагментов минус один.
6	<code>approximate_entropy</code> (x, m, r)	Реализует векторизованный алгоритм приближенной энтропии.
7	<code>ar_coefficient</code> (x, параметр)	Этот калькулятор функций соответствует безусловной максимальной вероятности процесса AR (k) с авторегрессией.

Разделим наш табличный набор данных для обучения модели дерева регрессии с повышением градиента (GBRT) и оценим ее, используя в качестве функции ошибки «квантиль 95 %». Использование данной функции ошибки позволит научить модель действовать более «смело» при прогнозировании переходных процессов, что важно при прогнозировании электропотребления. Если наша модель нацелена на прогнозирование, т.е. прогнозирование будущих данных на основе прошлых данных, нам не следует использовать обучающие данные, которые предшествуют данным тестирования. При машинном обучении временным рядам предположение о независимом и идентично распределенном неверно, поскольку точки данных не являются независимыми и имеют временную взаимосвязь. Производительность модели составила 0,168 % средней абсолютной погрешности на тестовой выборке данных.

Ошибка обобщения, измеренная с помощью разделения обученных тестов в случайном порядке, может быть слишком оптимистична. Обобщение с помощью разделения по времени, вероятно, будет более репрезентативным для истинной производительности регрессионной модели. Оценим эту изменчивость нашей оценки ошибок с помощью надлежащей перекрестной проверки.

Ниже средние баллы CV и их стандартное отклонение. Вариабельность между срезами не велика, что говорит о правильности подготовки данных для обучения и тестирования.

$$CV \text{ MAPE: } 0.206 \pm 0.050 \%$$

В данном исследовании было выполнено прогнозирование потребления электроэнергии с использованием библиотеки tsfresh для автоматического извлечения функций из временного ряда. Полученный результат в 0,206 % средней абсолютной ошибки прогнозирования значительно превосходит лучший результат из предыдущего исследования в 14,06 %, который был получен с использованием модели скользящего среднего. Результаты исследования показали правильность выбора модели для прогнозирования потребления электроэнергии и возможность

проведения дальнейших экспериментальных исследований в данном направлении.

Литература

1. Серегичев, К. С. Кратковременное прогнозирование электропотребления современных энергоемких предприятий / К. С. Серегичев, А. А. Суконщиков // Вестник Вологодского государственного университета. – 2024. – № 2. – С. 36–39.

2. Крист, М. Извлечение объектов распределенных и параллельных временных рядов для промышленных приложений с большими данными / Кемп-Лир А. В., Файндт М. // Азиатская конференция по машинному обучению (ACML), семинар по обучению работе с большими данными (WLBD). – 2016. – URL: <https://arxiv.org/abs/1610.07717v1> (дата обращения: 09.12.2024). – Текст : электронный.

3. Крист, М. Извлечение признаков временных рядов на основе масштабируемых тестов гипотез (tsfresh – пакет Python) / Браун Н., Нойффер Дж., Кемпа-Лир А. В. – doi: 10.1016/j.neucom.2018.03.067 // Нейрокомпьютинг 307. – 2018. – С. 72–77.

4. Моделирование прогнозируемого разделения событий микролинизирования с использованием систематического анализа характеристик временных рядов / Кеннеди, А., Джемма Н., Раттенбери Н., Кемпа-Лир А. В. – doi: 10.1016/j.ascom.2021.100460 // Астрономия и вычислительная техника. – 2021. – С. 1–14.

5. Прогнозирование временных рядов TensorFlow. – URL: https://www.tensorflow.org/tutorials/structured_data/time_series?hl=ru (дата обращения: 09.12.2024). – Текст : электронный.

6. Lagged features for time series forecasting Scikit-learn. – URL: https://scikit-learn.org/stable/auto_examples/applications/plot_time_series_lagged_features.html#lagged-features-for-time-series-forecasting (дата обращения: 09.12.2024). – Текст : электронный.

7. Feature extraction settings tsfresh. – URL: https://tsfresh.readthedocs.io/en/latest/text/feature_extraction_settings.html (дата обращения: 09.12.2024). – Текст : электронный.

K.S. Seregichev, A.A. Sukontschikov
Vologda State University

**FORECASTING ELECTRICITY CONSUMPTION USING TIME SERIES
FEATURE EXTRACTION**

The article examines the issues of forecasting electricity consumption in the commercial real estate. The study is conducted using the tsfresh library in Python. The peculiarity of this library is the automatic extraction of functions and features from a time series for further use in a machine learning model.

Forecasting, machine learning, time series.