

## МОДЕЛИРОВАНИЕ ЭНЕРГОПОТРЕБЛЕНИЯ УМНЫМИ ЗДАНИЯМИ С ИСПОЛЬЗОВАНИЕМ СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

ЕВГЕНИЙ ЮРЬЕВИЧ ЩЕТИНИН (ORCID 0000-0003-3651-7629)

ФГОБУ ВО «Финансовый университет при Правительстве Российской Федерации»

### *Аннотация.*

Интеллектуальные технологии энергосбережения и энергоэффективности являются современным масштабным мировым трендом в деловом бизнесе. Спрос на «умные» здания растет не только в мире, но и в России, прежде всего на рынке строительства и эксплуатации крупных торгово-развлекательных и бизнес-центров и др. В настоящей работе для моделирования энергопотребления применен алгоритм градиентного бустинга. На его основе предложен метод моделирования дневного профиля энергопотребления и разработан реализующий его численный алгоритм. Для оценки его эффективности были использованы данные о энергопотреблении 380 коммерческих зданий. Результаты показали, что использование модели градиентного бустинга улучшило точность прогнозирования более чем в 80 процентах случаев по сравнению с алгоритмами случайного леса и регрессии.

**Ключевые слова:** умные здания, искусственный интеллект, машинное обучение, градиентный бустинг.

### *Abstract.*

Intelligent technologies of energy saving and energy efficiency are a modern large-scale global trend in the development business. The demand for smart buildings is growing not only in the world, but also in Russia, especially in the market of construction and operation of large shopping, entertainment and business centers. In this paper the gradient boosting algorithm is used to simulate energy consumption. On its base the method of modeling the daily energy consumption profile is proposed and a numerical algorithm is developed. Data on the energy consumption of 380 commercial buildings were used to assess its efficiency. The computer experiments showed that the use of the gradient boosting model improved the prediction accuracy in more than 80 percent of cases compared to the random forest algorithms.

**Keywords:** smart buildings, artificial intelligence systems, gradient boosting, neural nets.

### ВВЕДЕНИЕ

Важнейшим из направлений развития экономики является повышение энергоэффективности производственных и потребительских секторов экономики. Одним из важнейших направлений развития экономики России утверждена государственная программа Российской Федерации "Энергосбережение и повышение энергетической эффективности на период до 2030 года". В целях снижения воздействия на окружающую среду и затрат, связанных с сектором коммерческих зданий, было реализовано несколько программ по повышению энергоэффективности промышленных предприятий. Сегодня растущая доступность данных от интеллектуальных счетчиков и устройств

в сочетании с интеллектуальным анализом данных позволяет оптимизировать процесс за счет повышения уровня автоматизации при сохранении или повышении точности результата. Развитие интеллектуальных сетей в производстве, финансах и услугах создает новые возможности для разработки и применения эффективных методов машинного обучения и анализа данных. Внедрение интеллектуальных счетчиков обеспечивает преимущества конечным потребителям, поставщикам энергии и сетевым операторам, предоставляя потребителям информацию о режиме потребления, близком к реальному времени, что поможет им управлять реальным потреблением энергии, экономить деньги и сокращать выбросы парниковых газов. В то же время

интеллектуальные счетчики будут способствовать планированию и эксплуатации распределительной сети, а также управлению спросом. В связи с этим данные интеллектуального учета позволят более точно прогнозировать спрос, повысить эффективность использования активов в распределительных сетях, локализовать простои и сократить время восстановления поставок, а также снижение эксплуатационных расходов сетей.

#### МОДЕЛИРОВАНИЕ ПРОФИЛЯ ПОТРЕБЛЕНИЯ ЭЛЕКТРОЭНЕРГИИ

Основные модели, используемые в оценивании профилей энергопотребления, являются эмпирическими моделями, которые связывают объемы потребления электроэнергии в зданиях с такими параметрами, как, например, температура воздуха внешней среды, влажность, характеристики самого здания и т.д. [2,4,6]. Традиционно для построения таких моделей использовались данные ежемесячных счетов за коммунальные услуги, однако увеличение доступности данных счетчиков с часовым и 15-минутным интервалами позволило создать новые модели для более точных прогнозов. За последние два десятилетия были достигнуты значительные успехи в разработке новых методов машинного обучения, среди которых наиболее перспективными с точки зрения точности прогнозирования являются подходы семейства алгоритмов ансамблевого обучения. Ансамблевые методы строят модель путем обучения нескольких относительно простых моделей, а затем объединяют их для создания более сложной модели с более высокими прогностическими свойствами. Наиболее известные алгоритмы обучения ансамбля используют бэггинг [1]; случайные леса RF [2]; и различные варианты бустинга [3]. Алгоритм бустинга GBM является наиболее мощным методом, его основной принцип заключается в том, чтобы несколько простых моделей, называемых "слабо обучающимися моделями", объединить в одну итерацион-

ную схему подбора параметров с целью получения так называемого «сильного учащегося», т.е. модели с улучшенной точностью прогнозирования. Таким образом, алгоритм GBM итеративно добавляет на каждом шаге новое дерево решений, которое наилучшим образом уменьшает функцию потерь. А именно, в регрессионной модели алгоритм начинается с ее инициализации, которая обычно представляет собой дерево решений, минимизирующее функцию потерь (RMSE), а затем на каждом шаге новое дерево решений подгоняется к текущему остатку и добавляется к предыдущей модели для обновления остатков. Алгоритм продолжает работать до тех пор, пока не будет достигнуто максимальное число итераций или не достигнута заданная точность. Этот процесс является пошаговым, т.е. на каждом новом этапе дерева решений, добавленные в модель на предыдущих этапах, не изменяются. Алгоритм GBM будет более эффективным, если на каждой итерации вклад добавленного дерева решений учитывается с помощью некоторого гиперпараметра, который можно охарактеризовать как скорость обучения. Идея процедуры подбора гиперпараметра заключается в том, что большее число малых шагов обеспечивает более высокую точность, чем меньшее число больших шагов. Параметр может принимать значение от 0 до 1, и чем он меньше, тем точнее будет модель. Однако выбор более сильного сжатия подразумевает большее число итераций для достижения сходимости, так как значение обратно пропорционально числу итераций.

Другим способом повышения точности прогнозирования алгоритма GBM является добавление рандомизации в процесс оценивания. На каждой итерации, вместо использования полного набора данных, для оценивания дерева решений используется подвыборка. Однако, чтобы оценить влияние уменьшения числа точек данных на качество подгонки модели,

необходимо проверить несколько подвыборок различной размерности [5].

В настоящей работе осуществлена попытка создать новый численный алгоритм выбора и тонкой настройки параметров модели базового профиля энергопотребления коммерческих зданий. Для решения поставленной задачи нами использовался алгоритм градиентного бустинга GBM, а также разработан ряд алгоритмов тестирования точности прогнозирования энергопотребления [5,6,7].

#### ОЦЕНКИ И ПРОГНОЗ

##### ЭНЕРГОЭФФЕКТИВНОСТИ УМНЫХ ЗДАНИЙ

Для каждого здания данные временных рядов были разбиты на периоды обучения и прогнозирования. Период прогнозирования был определен как последние 12 месяцев имеющихся данных. Модели обучаются с использованием двух различных сроков обучения, которые составляют 6 месяцев и 12 месяцев. Для всех зданий из базы данных в распоряжении имеется 36 месяцев потребления электроэнергии и данные температуры наружного воздуха.

Гиперпараметры GBM были настроены автоматически, используя метод сетки поиска с помощью методов перекрестной проверки. Таким образом, глубина деревьев решений  $d$  была выбрана из множества  $\{3, \dots, 10\}$ , скорость обучения была выбрана между значениями 0,5 и 0,01, количество итераций  $K$  было выбрано в пределах от 10 итераций до 500 с шагом 10 итераций. Показатели точности  $R^2$ ,  $CV(RMSE)$  были рассчитаны по всему набору данных зданий и продемонстрировали ухудшение точности по мере сокращения периода обучения с 12 месяцев до 6 месяцев.

Таблица 1 показывает, что обе модели GBM превосходят модели RF и регрессию по показателям точности  $R^2$  и  $RMSE$ . При сокращении периода обучения до 6 месяцев значительное снижение было отмечено по стандартной модели GBM (со стандартной  $K$ -кратным  $CV$ ) и

незначительное снижение в  $R^2$  для GBM-1d, GBM7d и для RF модели, в то время как точность улучшилась в регрессионной модели, что означает, что для данного набора данных регрессионная модель не повышает точность при увеличении числа наблюдений.

#### Выводы

В работе предложен метод построения модели базового профиля потребления электроэнергии крупными коммерческими центрами и бизнес-зданиями. Он основан на алгоритме градиентного бустинга с адаптивной настройкой гиперпараметров с использованием процедуры  $K$ -fold-blocks [5]. Работоспособность и эффективность его применения к решению задачи повышения энергоэффективности была протестирована как на модельных, так и реальных данных. Модель GBM показала более высокую точность прогнозирования, чем модели регрессии и случайных лесов на всех протестированных периодах обучения. Результаты проведенных компьютерных экспериментов показали, что использование модели GBM позволяет повысить точность оценки энергосбережения как отдельного здания так всего комплекса. Также оказалось, что использование 6-месячного периода обучения для построения моделей GBM привело к незначительному снижению точности прогноза энергопотребления, по сравнению с теми, которые были получены на 12-месячном периоде обучения, который обычно используется для всего здания. Это позволяет в целом сократить общее время, необходимое для проведения оценки энергосбережения всего комплекса зданий.

Сравнение различных алгоритмов настройки гиперпараметров показало, что важно принять во внимание автокорреляцию временных рядов. Действительно, результаты показывают, что использование стандартной кросс-валидации  $K$ -fold  $CV$  снижает точность алгоритма GBM. Это связано с тем, что при ис-

пользовании стандартного подхода K-fold CV наблюдения в тестовых и обучающих наборах данных не являются независимыми (из-за автокорреляции измерений, получаемых от умных датчиков), что приводит к переобучению на модельных

данным. Также было показано, что разница в использовании качестве блока прогноза неделя или день не оказывает существенного влияния на результаты.

Таблица 1.

Сравнительный анализ различных алгоритмов оценивания точности прогноза энергопотребления.

Model/Criteria	R2 6m	RMSE(CV) 6m	R2 12m	RMSE(CV) 12m
GBM	33	47	61	76
GBM-1day	57	63	67	81
GBM-7day	77	70	81	86
R-forest (случайный лес)	28	40	35	48
Regression	17	30	27	38

### Список источников

1. Breiman L., Bagging predictors, Machine learning, 1996, vol. 24(2), p. 123–140.
2. Щетинин Е.Ю., Бережков М.В., Повышение эффективности сетей энергопотребления с применением систем искусственного интеллекта, Экономика и управление: проблемы, решения. 2018, т.7, 5, с. 164-168.
3. Breiman L., Random forests, Machine learning, 2001, vol. 45(1), p.5–32.
4. Щетинин Е.Ю., Эффективные компьютерные алгоритмы моделирования спотовых цен на электроэнергию, Научное обозрение, 2016, 22, с.237-242.
5. Щетинин Е.Ю., Бережков М.В., Любин П.Г., Разработка эффективного алгоритма краткосрочного прогнозирования электропотребления с использованием метода ансамбля, В книге: Информационные технологии и математическое моделирование высокотехнологических систем. Материалы Всероссийской конференции с международным участием, 2018, с. 413-415.
6. Shchetinin E.Yu., Lyubin P.G., Fast two-dimensional smoothing with discrete cosine transform, Springer Communications in Computer and Information Science (CCIS), Springer: Berlin, 2016, 678, p. 646-656.
7. Shchetinin E. Yu., Cluster-based energy consumption forecasting in smart grids, Springer Communications in Computer and Information Science (CCIS), Springer, Bern, 2018, 919, p. 454-466.