
**ПРОГРАММНАЯ ИНЖЕНЕРИЯ, ТЕСТИРОВАНИЕ
И ВЕРИФИКАЦИЯ ПРОГРАММ**

УДК 004.42

**РАЗРАБОТКА КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ
ЭНЕРГОСБЕРЕЖЕНИЯ УМНЫХ ЗДАНИЙ
С ПРИМЕНЕНИЕМ КОМПЬЮТЕРНОЙ АЛГЕБРЫ**

© 2020 г. Е. Ю. Щетинин*

*Финансовый университет при правительстве РФ
125993 Москва, Ленинградский пр-т, д. 49, Россия*

**E-mail: riviera-molto@mail.ru*

Поступила в редакцию 10.07.2019 г.

После доработки 15.09.2019 г.

Принята к публикации 03.10.2019 г.

Интеллектуальные технологии энергосбережения и энергоэффективности являются современным масштабным мировым трендом в развитии энергетических систем. Точные оценки энергосбережения важны для продвижения строительных проектов в области энергоэффективности и демонстрации их экономической привлекательности. Растущее количество цифровой измерительной инфраструктуры, используемой в коммерческих зданиях, привело к повышению доступности данных высокой частоты, которые можно использовать для обнаружения неисправностей и диагностики оборудования, отопления, вентиляции, и оптимизации кондиционирования воздуха. Это обусловило применение современных и эффективных методов машинного обучения, предоставляющих перспективные возможности для получения более точных прогнозов энергопотребления здания, и, таким образом, повышения показателей энергосбережения. В настоящей работе на основе модели градиентного бустинга предложен метод моделирования и прогнозирования энергопотребления комплекса зданий, а также разработаны компьютерные алгоритмы, их программно реализующие в системе компьютерной алгебры SymPy. Для оценки энергоэффективности были использованы данные о энергопотреблении 300 коммерческих зданий. Результаты компьютерного моделирования показали, что использование разработанных алгоритмов и программ повысило точность прогнозирования более чем в 80 процентах случаев по сравнению с другими алгоритмами машинного обучения.

DOI: 10.31857/S0132347420050088

1. 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. Например, на государственном и федеральном уровнях в России были установлены долгосрочные целевые показатели энергосбережения, и эти целевые показатели должны быть достигнуты с помощью программ энергоэффективности. Анализ энергоэффективности имеет решающее значение для определения тарифов для владельцев зданий, плательщиков коммунальных тарифов и поставщиков услуг [1–3].

Развитие интеллектуальных сетей в производстве, финансах и услугах создает новые возможности для разработки и применения эффективных методов машинного обучения и анализа данных, проектирования новых модулей управления киберфизическими энергетическими системами. 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 способствуют планированию и эксплуатации распределительной сети, а также управлению спросом. В связи с этим точные интеллектуального учета позволяют более точно прогнозировать спрос, повысить эффективность эксплуатации распределительных сетей, сократить время восстановления поставок, а также снизить эксплуатационные издержки сетей.

Интеллектуальные технологии сбора, регистрации и мониторинга данных о потреблении энергии создают огромное количество данных различного характера для использования поставщиками энергии и сетевыми операторами. Объем данных варьируется в зависимости от количества установленных интеллектуальных счетчиков, количества полученных сообщений интеллектуальных счетчиков, размера сообщения и частоты записи измерений, например, каждые 15 или 30 минут. Эти данные могут быть использованы для оптимального управления сетью, повышения точности прогнозирования нагрузки, выявления аномальных эффектов электроснабжения (условия пиковой нагрузки), формирования гибких ценовых тарифов для различных групп потребителей.

Основные модели, используемые в оценивании профилей энергопотребления, являются эмпирическими моделями, которые связывают объемы потребления электроэнергии в зданиях с такими параметрами, как, температура воздуха внешней среды, влажность, характеристики самого здания и т.д. 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 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 сложных моделей с более высокими прогностическими свойствами. В настоящей работе предложен новый компьютерный алгоритм построения модели потребления электроэнергии группой коммерческих зданий, использующих умные датчики регистрации показаний. Для этого нами были использованы алгоритмы машинного обучения на ансамблях, такие как случайный лес и градиентный бустинг GBM [4, 5, 13]. Также разработаны эффективные численные алгоритмы оценки и тонкой настройки параметров модели GBM и оценивания точности прогноза энергопотребления, которые были программно реализованы в системе компьютерной алгебры SymPy и языке программирования Python.

2. МЕТОДЫ МОДЕЛИРОВАНИЯ ЭНЕРГОПОТРЕБЛЕНИЯ С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ

На сегодняшний день растущая доступность данных со smart-счетчиков и в сочетании с ин-

теллектуальным анализом данных позволяет оптимизировать процесс энергопотребления за счет повышения уровня автоматизации при сохранении и повышении точности прогнозирования [1–3]. Основными моделями, используемыми при оценке профилей энергопотребления, являются регрессионные модели, связывающие потребление энергии в зданиях с такими параметрами, как, температура внешней среды, влажность, индивидуальные характеристики здания и т.д. [3, 7]. Для построения таких моделей традиционно использовались данные ежемесячных счетов за коммунальные услуги, однако рост доступных данных со smart-счетчиков с часовым и 30-минутным интервалом позволяют создать новые методы для более точного прогнозирования. Однако, с внедрением интеллектуальных технологий сбора, анализа и контроля данных энергопотребления достигнут значительный прогресс в разработке новых методов с применением алгоритмов машинного обучения, среди которых наиболее перспективными с точки зрения точности прогнозирования являются семейства алгоритмов ансамблей. Хорошо известны алгоритмы, применяемые в этих целях, такие как регрессионные деревья, случайный лес [4], бэггинг [5], а также бустинг [13]. Методы ансамблей создают модель, обучающую несколько относительно простых моделей, а затем объединяющую их для создания более сложной, но с лучшими прогнозирующими свойствами. Хотя эти алгоритмы машинного обучения с большим успехом используются во многих областях, они только начинают применяться к задачам моделирования энергосбережения. Например, в работах [8, 12, 15] авторы использовали случайный лес для прогнозирования почасового потребления энергии.

В данной работе предложен новый алгоритм выбора и тонкой настройки параметров модели энергопотребления зданиями, входящими в бизнес-комплекс, использующий алгоритм градиентного бустинга GBM, а также разработаны алгоритмы оценивания точности прогнозирования энергопотребления. Применение алгоритмов ансамбля и деревьев решений в качестве метода регрессии имеет несколько преимуществ, одним из которых является то, что правила разделения представляют собой интуитивно понятный и простой графический способ визуализации результатов. Кроме того, по своей структуре они могут одновременно обрабатывать числовые и категориальные входные параметры. Они устойчивы к выбросам и могут эффективно работать с пропусками данных в пространстве входных параметров. Иерархическая структура дерева решений автоматически моделирует взаимодействие между входными параметрами и выполняет эффективный отбор переменных, например, если входной параметр вообще не используется, то

прогноз не зависит от этого входного параметра. Наконец, алгоритмы деревьев решений просты в реализации и вычислительно эффективны с большими объемами данных [9, 10]. Алгоритм GBM был впервые предложен для задач классификации в работе [13]. Его основной принцип заключается в том, что несколько простых моделей, называемых “слабыми” моделями обучения, должны быть объединены в одну итерационную схему для выбора параметров с целью получения так называемой “сильной” модели обучения, то есть модели с лучшей точностью прогнозирования. Таким образом, алгоритм GBM добавляет новое дерево решений на каждой итерации так, что наилучшим образом уменьшает функцию потерь. Алгоритм продолжает работать, пока не будет достигнуто максимальное количество итераций или пока не будет достигнута указанная точность. Это означает, что на каждом новом шаге дерева решений, добавленные в модель на предыдущих шагах, фиксируются. Таким образом, модель может быть улучшена в тех ее частях, где она еще оценивает остатки недостаточно хорошо.

Алгоритм GBM более эффективен, если на каждой итерации вклад добавленного дерева решений учитывается с использованием некоторого гиперпараметра, описывающий одну из важных характеристик алгоритма, именно, скорость обучения модели v . Одна из проблем выбора гиперпараметра v состоит в том, что для достижения требуемой точности ϵ , в зависимости от значения v , необходимо соответствующее число итераций m . А именно, чем меньше значение v , тем большее число итераций необходимо выполнить. Таким образом, необходимо разработать процедуру оптимального выбора гиперпараметров v , m модели GBM при заданной точности ϵ . Еще одной проблемой является наличие автокорреляции, которая, как известно, вносит дополнительные искажения в оценки параметров модели [16]. Решением изложенных выше проблем на наш взгляд является применение рандомизации в процессе построения алгоритма градиентного бустинга GBM. На каждой итерации вместо всей выборки для оценки дерева решений применяется ее подвыборка, извлекаемая случайным образом. На практике, однако, редко можно выделить достаточное количество точек данных для точной оценки прогностической эффективности моделей без влияния на качество оценки. Когда количество наблюдений недостаточно, уменьшение размера обучающего набора может привести к существенному снижению точности [6, 9, 17]. Поэтому, для оценки влияния размера подвыборки на качество подгонки модели необходимо использовать подвыборки различных размеров. В этой ситуации для решения поставленных задач нами был

разработан численный алгоритм выбора оптимальных значений гиперпараметров модели GBM с применением процедуры k -кратной перекрестной проверки и рандомизации (k -fold cross validation). Метод k -кратной перекрестной проверки включает деление набора данных на k подвыборок примерно одинакового размера. Сначала модель оценивается с использованием $(k - 1)$ блоков в качестве обучающей выборки, а k -й блок (тестовая выборка) используется для определения точности прогноза. Далее процедура повторяется k раз, и всякий раз в качестве тестовой выборки используется новый блок. Для оценки точности прогноза на k -м блоке мы используем среднеквадратичную ошибку $RMSE_k = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2}$, где \hat{y} – прогнозируемое значение, y_i – значение набора реальных данных. Метод перекрестной проверки k -кратности использует среднеквадратичную ошибку в виде

$$RMSE(CV) = \frac{1}{k} \sum_{l=1}^k RMSE_l, \quad (2.1)$$

k – число блоков. Предложенная и исследуемая в работе модель GBM имеет четыре гиперпараметра:

1. d – глубина деревьев решений;
2. m – количество итераций;
3. v – скорость обучения, которая обычно составляет значение между 0 и 1, уменьшение которого приводит к замедлению процедуры оценки, что требует увеличения количества итераций m ;
4. k – количество разбиений выборки на подвыборки, используемые на каждом шаге итерации в алгоритме k -кратной перекрестной проверки (2.1).

Как и для любого алгоритма машинного обучения, проблема переобучения актуальна и для алгоритма GBM. Переобучение обычно является недостатком чрезмерно сложной модели. В случае модели GBM это может произойти, если выбрано слишком много итераций m или слишком большая глубина d деревьев решений. Таким образом, необходимо выбрать оптимальную комбинацию гиперпараметров, чтобы избежать переобучения и в то же время обеспечить наилучшую точность прогноза. Эффективным методом для решения этой проблемы является поиск по сетке (Grid Search) [14]. Этот подход состоит из определения множества различных значений гиперпараметров, построения модели для каждой их комбинации и выбора оптимальной комбинации с использованием показателей, которые количественно определяют модель с точки зрения точности прогнозирования (2.1).

Псевдокод представленного алгоритма имеет следующий вид:

1. Задать глубину деревьев решений d , количество итераций m , скорость обучения v , достижимую точность оценивания параметров модели ϵ .

2. Начальный шаг равен $x_0 = \bar{y}$, где \bar{y} – среднее значение y , $f_0 = 0$.

3. Для $j = 1, \dots, m$ выполнить:

- Произвольно извлечь подвыборку (x_i, y_i) , $i = 1, \dots, N$ из обучающей выборки, где N – количество измерений, соответствующих размеру подвыборки;

- Используя значения (x_i, y_i) , построить дерево решений \tilde{f}^j глубины d с учетом остатков z^j ;

- Обновить значение остатков $z^{j+1} = z^j + v \tilde{f}^j(x)$.

4. Если $RMSE(CV) < \epsilon$ или $j = m$, то переход к п. 5. Если нет, то возврат к п. 3. (здесь ϵ – заданная точность оценки параметров модели).

5. Конец.

При реализации рассмотренного алгоритма перед нами возникла задача выбора системы компьютерной программы. Ряд содержательных рекомендаций по этому вопросу дан в работе [23]. Наиболее интересной на наш взгляд представляется система SymPy [24, 25]. Эта система появилась как библиотека символьных вычислений для языка Python [26]. Быстрое развитие этого языка в последние годы наряду с библиотеками Matplotlib, SciPy, NumPy и др. обеспечило устойчивый интерес исследователей к применению в своей работе систем символьных вычислений. Удобный интерфейс SymPy позволяет передавать выходные результаты в библиотеки Python для их визуализации или дальнейших вычислений. В компьютерной среде языка Python широко представлены численные алгоритмы, реализующие метод *GridSearch* [26]. Для этого нужно импортировать класс *GridSearchCV* из библиотеки *sklearn.model_selection* и задать начальные значения основных параметров. Соответствующие операторы, программно реализующие выше изложенный алгоритм в Python, выглядят следующим образом

```
import numpy as np
from sympy import *
from sklearn.model_selection import
(cross_val_score, train_test_split,
GridSearchCV, RandomizedSearchCV)
from sklearn.metrics import r2_score
from lightgbm.sklearn import LGBMRegressor
from sklearn.ensemble import
RandomForestRegressor
X_train, X_test, y_train, y_test =
train_test_split(X, y, test_size=0.2,
random_state=2018)
```

Далее необходимо определить начальные значения гиперпараметров с использованием следующего оператора

```
hyper_space =
{'m_iter': [100, 200, 300, 400, ...1000],
'max_depth': [4, 5, 8],
'num_leaves': [15, 31, 63, 127],
'subsample': [0.6, 0.7, 0.8, 1.0]}
```

провести оценивание методом градиентного бустинга

```
est = LGBMRegressor(boosting='gbdt',
n_jobs=-1, random_state=2018)
```

и найти оптимальные значения гиперпараметров методом поиска по сетке *GridSearch*

```
gs = GridSearchCV(est, hyper_space,
scoring='r2', cv=4, verbose=1)
gs_results = gs.fit(X_train, y_train)
print("BEST PARAMETERS: "
+ str(gs_results.best_params_))
print("BEST CV SCORE: "
+ str(gs_results.best_score_)).
```

Далее построим прогноз энергосбережения и оценим его ошибку

```
# Predict (after fitting GridSearch-
CV is an estimator with best parameters)
y_pred = gs.predict(X_test)
# Score
score = r2_score(y_test, y_pred)
print("R2 SCORE: {}".format(score)).
```

При необходимости возможен вывод результатов работы программы в формат *LATEX*. Для этого следует импортировать модуль *LATEX*

```
from IPython.display import Latex
```

и вызвать функцию

```
sympy.init_printing(use_unicode=True).
```

Кроме этого, возможно экспортировать результаты в формат ряда языков программирования (*C*, *C++*, *Fortran*). Данные функции доступны с использованием класса *sympy.printing*.

3. КОМПЬЮТЕРНЫЕ ЭКСПЕРИМЕНТЫ И АНАЛИЗ РЕЗУЛЬТАТОВ

Для тестирования разработанных в статье алгоритмов были использованы данные показателей электропотребления городского конгломерата зданий в г. Дублин, Ирландия [11]. Данные представляют собой 15-минутные измерения потребления электроэнергии в кВт за период 29.03.2011–20.02.2013. Типовой график исследуемых временных рядов энергопотребления одного из зданий приведен на рис. 1. Для каждого здания

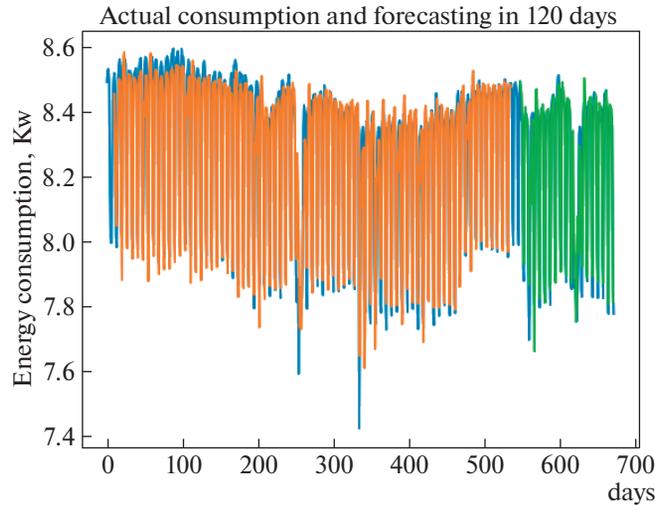


Рис. 1. График энергопотребления одного из зданий в логарифмической шкале.

данные разбиты на периоды обучения и прогнозирования. Периоды прогнозирования были определены в пределах последних 12 месяцев исследуемых данных. Модели регрессии, случайного леса и градиентного бустинга обучались с использованием двух разных периодов, которые составили 6 и 12 месяцев. Глубина деревьев решений d была выбрана из набора (3, ..., 10), скорость обучения ν была выбрана между 0.05 и 1, число итераций m принимало значения от 1 до 1000 с шагом в 10 итераций. Гиперпараметры GBM настраивались с использованием алгоритма поиска по сетке совместно с методами перекрестной проверки. Для этого были использованы три варианта стандартной процедуры перекрестной проверки (CV): 5-кратная CV, 5-кратная CV и сутки в качестве блока (CV-1day), 5-кратная CV с одной неделей в качестве блока для CV (CV-7day). Таким образом, с учетом выше предложенных алгоритмов перекрестной проверки с тонкой настройкой параметров, имеем три различные реализации модели градиентного бустинга: 1) модель GBM с выбором параметров стандартным методом 5-кратной перекрестной проверки CV; 2) модель GBM с CV-1day (GBM-1d); 3) модель GBM с CV-7day (GBM-7d).

Результаты компьютерных экспериментов показали, что значение скорости обучения $\nu = 0.1$ слишком мало, так как алгоритм оказался слишком чувствителен как к количеству итераций, так и к глубине деревьев решений. Также выяснилось, что при скорости обучения 0.2 и глубине дерева решений $d = 5$ алгоритм не достиг оптимального числа итераций при $m = 500$. При увеличении скорости обучения до $\nu = 1$ из-за необходимости увеличения количества итераций и в силу возросшей

сложности модели алгоритм начинает переобучаться. Таким образом, оптимальная скорость обучения была принята равной 0.5. Характер поведения ошибки RMSE (2.1) модели GBM в зависимости от различных значений ν представлен на рис. 2.

Показатели точности коэффициент детерминации R^2 , среднеквадратическая ошибка RMSE(CV) были рассчитаны для всего набора данных и продемонстрировали снижение точности при сокращении периода обучения с 12 до 6 месяцев. При этом компьютерные эксперименты показали, что R^2 для моделей GBM превосходит соответствующие значения R^2 для моделей RF и регрессии. Это осо-

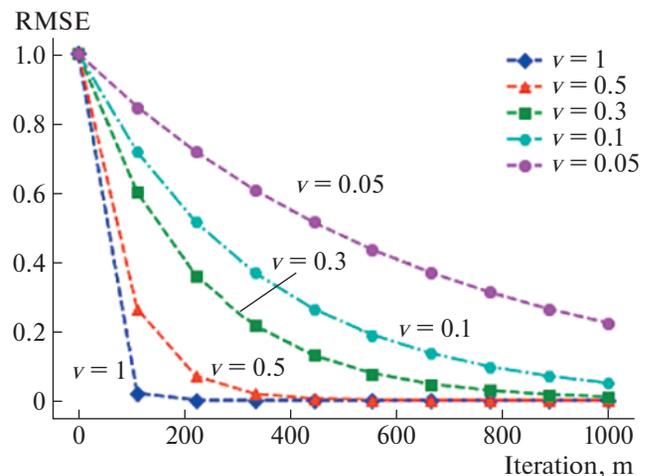


Рис. 2. RMSE для различных значений гиперпараметра ν при обучении модели GBM.

Таблица 1. Оценки точности моделей энергопотребления на тестовой выборке для 6 и 12 месяцев

Модель	R^2 (6)	RMSE(CV) (6)	R^2 (12)	RMSE(CV) (12)
GBM	33	47	61	76
GBM - 1д	57	63	67	81
GBM - 7д	77	70	81	86
RF	28	40	35	48
Регрессия	17	30	27	38

бенно заметно для моделей GBM-1д и GBM-7д. Как и ожидалось, модель GBM, использующая стандартную процедуру CV, имеет меньшую точность, чем две другие версии, в которых используется k -кратная проверка. С уменьшением периода обучения до 6 месяцев в стандартной модели GBM наблюдалось некоторое снижение R^2 и его небольшое снижение для моделей GBM-1д, GBM-7д и RF. В то же время точность модели регрессии не улучшилась, что означает, что для этого набора данных регрессионная модель не повышает точность при увеличении количества наблюдений.

Что касается результатов оценки показателя точности RMSE(CV), то модели GBM-1 и GBM-7д превосходят модели RF и GBM в обоих периодах обучения. Точность прогноза моделей GBM значительно повысилась, когда их период обучения был увеличен с 6 до 12 месяцев, в то время как показатель точности регрессионной модели R^2 несколько снизился. Напомним, что для показателя R^2 желательны более высокие значения, тогда как для значений RMSE(CV) желательно иметь их близкими к нулю. В Таблице 1 показано количество зданий в процентах, для которых модели GBM оказались более точными, чем регрессионная и модель случайного леса RF. Для RMSE(CV) столбцы представляют процент зданий с меньшим значением RMSE(CV), чем у регрессионной модели и RF. Эти результаты подтверждают, что алгоритмы GBM-1д и GBM-7д имеют более высокую точность, чем регрессионные и RF модели. На основе построенных моделей также были проведены моделирование и прогноз энергопотребления на примере показателей одного из зданий. На рис. 1. представлен график прогноза на 120 дней с использованием алгоритма GBM-7д.

4. РЕЗУЛЬТАТЫ РАБОТЫ И ВЫВОДЫ

В настоящей работе предложен метод моделирования и оценивания потребления электроэнергии

крупными коммерческими центрами и зданиями. Для оценивания параметров модели был использован алгоритм градиентного бустинга с адаптивной настройкой гиперпараметров с использованием k -кратной процедуры перекрестной проверки. Для вычисления оптимальных значений гиперпараметров был разработан компьютерный алгоритм с использованием поиска по сетке. Работоспособность и эффективность применения алгоритма GBM к решению задачи повышения точности прогноза потребления электроэнергии была протестирована на реальных данных электропотребления. Модель GBM показала более высокую точность прогнозирования, чем модели регрессии и случайного леса на всех протестированных периодах обучения. Результаты проведенных компьютерных экспериментов показали, что использование модели GBM позволяет повысить точность оценки энергосбережения как отдельного здания, так и комплекса зданий в целом. Также установлено, 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. Подобное утверждение оказалось несправедливым для линейных моделей регрессионного анализа, а также ряда алгоритмов машинного обучения, что подтвердило результаты работ [7, 15]. Таким образом, применение алгоритмов GBM позволяет не только повысить точность оценки энергосбережения в целом, но и сократить общее время, необходимое для проведения оценки энергосбережения всего комплекса зданий.

Сравнительный анализ разработанных в статье алгоритмов настройки гиперпараметров показал, что важно принять во внимание автокорреляцию временных рядов. Действительно, результаты компьютерных экспериментов продемонстрировали, что использование стандартной процедуры перекрестной проверки снижает точность алгоритма GBM. Это связано с тем, что при использовании стандартного подхода к выбору гиперпараметров наблюдения в тестовых и обучающих выборках не являются независимыми (из-за автокорреляции измерений), что приводит к переобучению. С целью преодоления влияния автокорреляции в алгоритм оценки точности прогноза энергопотребления была включена процедура рандомизации. Также было показано, что различие в использовании в качестве периода прогноза одной недели или одного дня не оказывает существенного влияния на результаты оценивания и прогноза энергопотребления. Поэтому можно сделать вывод, что для большинства случаев использование одного дня в качестве стандартного блока при оце-

нивании точности прогноза энергопотребления является хорошим выбором.

Известно, что одним из главных достоинств модели ансамблей является их гибкость и надежность при использовании с большим количеством входных параметров [18]. Кроме того, по сравнению с регрессионными моделями, нет необходимости модифицировать алгоритм для обработки дополнительных входных параметров. Вместо этого достаточно включить эти переменные в качестве входных данных алгоритма без необходимости определять конкретную форму модели для каждого из параметров. Кроме того, дополнительные возможности выбора переменных в модели GBM позволяют исключать параметры, не влияющие на модель, без снижения свойств модели. Модель GBM обладает рядом очевидных преимуществ по сравнению с регрессионными моделями, состоящих в ее способности сохранять точность на более коротких периодах обучения, повышение общей точности показателей энергоэффективности и простота включения дополнительных объясняющих переменных. Такими параметрами могут служить, например, заполняемость здания, его освещенность и влажность. Дальнейшее развитие настоящей работы будет связано с применением модели GBM к решению проблем, связанных с энергоэффективностью [16, 17, 21], таких как прогнозирование потребления энергии при долгом срочном планировании нагрузки, непрерывное обнаружение аномалий и количественная оценка изменения сетевой нагрузки, реагирующей на спрос [19, 20, 22].

Рассмотренные алгоритмы были программно реализованы в системе компьютерной алгебры SymPy и языка программирования Python. В работе приведены отдельные элементы программного комплекса и продемонстрированы результаты его работы для задач энергосбережения. Его применение значительно повысило эффективность исследований в указанной области. Можно сделать вывод, что использование систем компьютерной алгебры стало неотъемлемой частью в разработке компьютерных технологий в области энергоэффективности.

СПИСОКИ ЛИТЕРАТУРЫ

1. *Gellings C.W.* The smart grid: enabling energy efficiency and demand response. The Fairmont Press, Inc., 2009.
2. *Arghira N., Hawarah L., Ploix S., Jacomino M.* Prediction of appliances energy use in smart homes. *Energy*. 2012. V. 48. № 1. P. 128–134.
3. *Hawarah L., Jacomino M.* Smart Home – From User’s Behavior to Prediction of Energy Consumption. ICINCO 2010, June 15–18, Proceedings of the 7th International Conference on Informatics in Control, Automation and Robotics, Volume 1, Funchal, Madeira, Portugal.
4. *Breiman L.* Random forests, *Machine learning*. 2001. V. 45. № 1. P. 5–32.
5. *Breiman L.* Bagging predictors, *Machine learning*. 1996. V. 24. № 2. P. 123–140.
6. *Bergmeir C., Benitez, J.M.* On the use of cross-validation for time series predictor evaluation. *Information Sciences*. 2012. V. 191. P. 192–213.
7. *Ediger V., Akar S.* ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy*. 2007. V. 35. P. 1701–1708.
8. *Cincotti S., Gallo G., Ponta L.* Modeling and forecasting of electricity spot-prices: Computational intelligence vs. classical econometrics. *AI Commun*. 2014. V. 27. P. 301–314.
9. *Ardakani F.J., Ardehali M.M.* Novel effects of demand side management data on accuracy of electrical energy consumption modeling and long-term forecasting. *Energy Convers. Manag.* 2014. V. 78. P. 745–752.
10. *Shchetinin E.Yu.* Cluster-based energy consumption forecasting in smart grids, *Springer Communications in Computer and Information Science (CCIS)*. Springer, Berlin. 2018. V. 919. P. 446–456.
11. <https://data.gov.ie/dataset/energy-consumption-gas-and-electricity-civic-offices-2009-2012/resource/>
12. *Liaw A., Wiener M.* Classification and Regression by Random Forest. *R News*, 2002. V. 2. № 3. P. 18–22.
13. *Friedman J.* Greedy function approximation: a gradient boosting machine. *Ann. Stat.* 2001. V. 29. № 5. P. 1189–1232.
14. *Burnham K.P., Anderson D.R.* Model Selection and Multi-Model Inference: A Practical, Information-theoretic Approach, Springer Verlag, 2002.
15. *Kane M.J., Price N., Scotch M.* Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks, *BMC Bioinformatics*. 2014. V. 15. P. 276–284; <https://doi.org/10.1186/1471-2105-15-276>
16. *Granderson J., Touzani S., Custodio C.* Accuracy of automated measurement and verification techniques for energy savings in commercial buildings, *Applied Energy*. 2016. V. 173. P. 296–308.
17. *Jain R.K., Smith K.M., Culligan P.J., Taylor J.E.* Forecasting energy consumption of multi-family residential buildings using support vector regression: investigating the impact of temporal and spatial monitoring granularity on performance accuracy, *Appl. Energy*. 2014. V. 123. P. 168–178.
18. *Zhou Z.H.* Ensemble learning, *Enycl. Biom.* 2015. P. 411–416.
19. *Zorita A.L., Fernandez-Temprano M.A., Garcia-Escudero L.-A., Duque-Perez O.* A statistical modeling approach to detect anomalies in energetic efficiency of buildings, *Energy Build.* 2016. V. 110. P. 377–386.
20. *Amozegar M., Khorasani K.* An ensemble of dynamic neural network identifiers for fault detection and isolation of gas turbine engines, *Neural Network*. 2016. V. 76. P. 106–121.

21. *Shchetinin E.Yu., Melezhik V.S., Sevastyanov L.A.* Improving the energy efficiency of the smart buildings with the boosting algorithms, Proceedings of the Selected Papers of the 12th International Workshop on Applied Problems in Theory of Probabilities and Mathematical Statistics in the framework of the Conference on Information and Telecommunication Technologies and Mathematical Modeling of High-Tech Systems (АТРР+MS'2018), CEUR Workshop Proceedings. 2018. V. 2332. P. 69–78.
22. *Lyubin P., Shchetinin E.Yu.* Fast two-dimensional smoothing with discrete cosine transform, Communications in Computer and Information Science (CCIS). Springer, Berlin. 2016. V. 678. P. 646–656.
23. *Геворкян М.Н., Демидова А.В., Велиева Т.Р., Королькова А.В., Кулябов Д.С., Севастьянов Л.А.* Реализация метода стохастизации одношаговых процессов в системе компьютерной алгебры. Программирование. 2018. № 2. С. 18–27.
24. <https://github.com/sympy>.
25. <https://www.sympy.org/ru/index.html>.
26. <https://www.python.org>.