УДК 681.5:004.8

АВТОМАТИЗИРОВАННЫЙ ВЫБОР ПРОГНОЗНОЙ МОДЕЛИ ДЛЯ МНОГОЗОННЫХ IoT-СИСТЕМ ТЕПЛОВОГО УПРАВЛЕНИЯ

Майлыханова Б.А.[1,2], Кашаганова Г.Б.[3]

[1]Казахский национальный университет имени аль-Фараби, Алматы, Казахстан, bulgyn.mail@gmail.com

[2]Казахский национальный аграрный исследовательский университет, Алматы, Казахстан,

[3]Казахский национальный исследовательский технический университет имени К. И. Сатпаева, Алматы, Казахстан

Аннотация: предложена процедура автоматизированного выбора модели краткосрочного прогнозирования температуры для децентрализованного управления многозонным зданием. Процедура совместно учитывает MAE, RMSE и R^2, нормированные внутри каждой комбинации режима HVAC, зоны и горизонта. Проверка выполнена по 96 результатам для четырёх зон, двух режимов и горизонтов 15, 30 и 60 мин. Random Forest обеспечил минимальный MAE в 14 из 24 задач и минимальный интегральный критерий в 13 задачах; XGBoost выбран в 9 задачах. Полученные правила позволяют автоматически назначать модель локальному контроллеру с учётом режима и горизонта.

Ключевые слова: автоматизированный выбор модели, IoT, многозонное здание, прогнозирование, Random Forest, XGBoost.

I. ВВЕДЕНИЕ

Прогноз температуры на горизонте от нескольких минут до одного часа необходим для опережающего управления HVAC-оборудованием. В многозонной киберфизической системе единая модель не всегда обеспечивает одинаковую точность: динамика зависит от этажа, тепловой инерции зоны, наружных условий и режима нагрева или охлаждения. Поэтому задача проектирования прогнозного слоя состоит не только в обучении моделей, но и в автоматизированном назначении подходящего алгоритма каждой локальной задаче управления.

Открытый набор данных Oak Ridge National Laboratory содержит минутные измерения двухэтажного экспериментального офисного здания при различных сценариях работы HVAC [1]. Методы Random Forest [2] и XGBoost [3] широко применяются для нелинейного прогнозирования, тогда как линейная регрессия и модель постоянства позволяют оценить практический выигрыш усложнения алгоритма. Обзоры методов прогнозирования энергопотребления зданий подчёркивают зависимость качества модели от данных и режима эксплуатации [4, 5]. Однако выбор модели только по усреднённой ошибке скрывает локальные отказы и не учитывает согласованность нескольких метрик.

Цель работы – разработать воспроизводимую процедуру выбора прогнозной модели для многозонной IoT-системы на основе локально нормированного интегрального критерия. В отличие от агрегированного ранжирования, решение принимается отдельно для каждой комбинации режима HVAC, зоны и горизонта прогнозирования.

II. МАТЕРИАЛЫ И МЕТОДЫ

Исходные экспериментальные результаты получены на минутных данных четырёх репрезентативных зон 102, 105, 202 и 205 для режимов охлаждения и нагрева. Для каждой зоны сформированы прогнозы на 15, 30 и 60 мин. Выборка разделена хронологически: первые 80 % наблюдений использованы для обучения, последние 20 % – для тестирования. Такое разделение исключает перенос информации из будущего в обучающую часть.

В признаки включены текущая температура и относительная влажность зоны, температура и расход приточного воздуха VAV, наружная температура и влажность, прямая, рассеянная и глобальная солнечная радиация, циклические признаки времени, а также лаги температуры 1, 5 и 15 мин. Сравнивались Persistence, Linear Regression, Random Forest и XGBoost. Для Random Forest использованы 200 деревьев и min_samples_leaf=2; XGBoost включал 200 деревьев глубиной 4 при learning_rate=0,05. Во всех стохастических процедурах random_state=42 [6].

Для каждой задачи g и модели m рассчитаны MAE, RMSE и R^2. Ошибочные метрики нормировались по диапазону среди четырёх моделей внутри одной задачи:

$$\tilde{e}_k(g,m) = [e_k(g,m) - \min_m e_k(g,m)] / [\max_m e_k(g,m) - \min_m e_k(g,m)], \quad k \in \{MAE, RMSE\}.$$

Для R^2 направление шкалы инвертировалось:

$$\tilde{r}(g,m)=[\max R^2(g,m)-R^2(g,m)]/[\max R^2(g,m)-\min R^2(g,m)].$$

Интегральная оценка $S(g,m)=(\tilde{e}MAE+\tilde{e}RMSE+\tilde{r})/3$ принимает значения от 0 до 1; меньшая величина соответствует лучшему компромиссу. Равные веса выбраны для прозрачности правила. Алгоритм автоматически назначает $\arg\min S(g,m)$, а при равенстве выбирает модель с меньшим MAE.

III. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Всего проанализировано 96 комбинаций «режим–зона–горизонт–модель», образующих 24 локальные задачи выбора. Таблица 1 показывает устойчивость моделей на полном наборе задач. Random Forest получил наименьший средний MAE 0,0930 °C и не дал отрицательных R^2. XGBoost имел близкий средний MAE 0,0941 °C, меньший средний RMSE 0,1411 °C и максимальный MAE 0,2227 °C. У Persistence отрицательный R^2 наблюдался в семи случаях, поэтому его использование как рабочего предиктора нецелесообразно.

Таблица 1. Агрегированные показатели точности и устойчивости

Модель	MAE, °C	SD MAE	Макс. MAE	RMSE, °C	Средний R^2	$R^2<0$
Persistence	0,1303	0,0807	0,3235	0,1924	0,302	7
Linear Regression	0,1110	0,0661	0,2788	0,1565	0,581	4
Random Forest	0,0930	0,0561	0,2185	0,1460	0,671	0
XGBoost	0,0941	0,0550	0,2227	0,1411	0,672	0

Минимальный MAE был получен Random Forest в 14 из 24 задач, XGBoost – в 8, а Linear Regression и Persistence – по одному разу. После совместного учёта трёх метрик Random Forest выбран для 13 задач, XGBoost – для 9, Linear Regression – для 2; Persistence не выбран ни разу (рис. 1). Изменение победителя показывает, что минимизация только MAE и многокритериальный выбор неэквивалентны. Интегральный критерий исключает решения, у которых небольшое преимущество по одной ошибке сопровождается ухудшением RMSE или R^2.

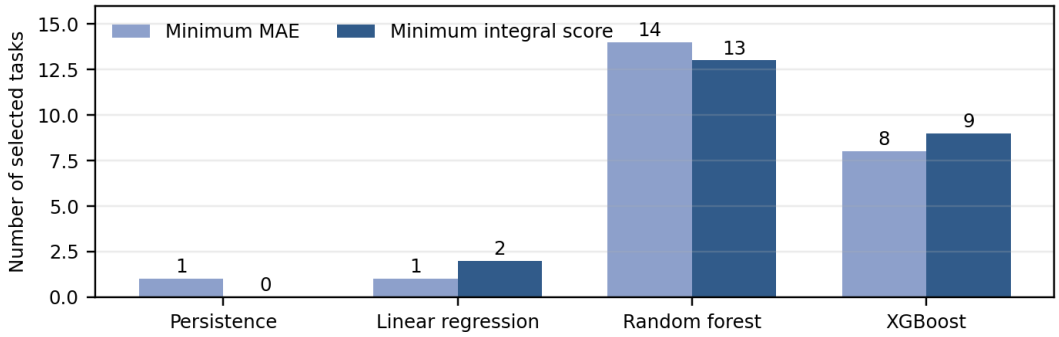


Рисунок 1. Частота автоматического выбора моделей в 24 локальных задачах

При попарном сравнении Random Forest имел меньший MAE, чем XGBoost, в 16 задачах, тогда как XGBoost – в 8. Однако преимущество зависит от горизонта. На 15 мин средние MAE практически совпали: 0,0654 °C для Random Forest и 0,0651 °C для XGBoost. На 30 мин XGBoost был лучше (0,0952 против 0,1015 °C), а на 60 мин Random Forest показал более низкую ошибку (0,1122 против 0,1222 °C). Следовательно, фиксированное назначение одной модели всем контроллерам упрощает внедрение, но не является оптимальным для каждого горизонта.

Средний MAE Random Forest составил 0,0958 °C в охлаждении и 0,0902 °C в нагреве; для XGBoost – 0,0963 и 0,0920 °C соответственно. Средний R^2 в охлаждении был около 0,49 для обеих ансамблевых моделей, а в нагреве достигал 0,848–0,851. Это различие подтверждает необходимость режимно-зависимой проверки: одинаковая абсолютная ошибка не означает одинаковую долю объяснённой изменчивости.

Предлагаемая процедура пригодна для этапа автоматизированного проектирования прогнозного слоя. После переобучения моделей система формирует таблицу метрик, нормирует их внутри локальной задачи и обновляет назначение модели. В рабочем контуре прогноз выполняет только выбранная модель, поэтому многомодельное сравнение не увеличивает постоянную вычислительную нагрузку. Практическое правило по результатам эксперимента состоит в использовании XGBoost как предпочтительного кандидата для горизонта 30 мин и Random Forest как более устойчивого кандидата для 60 мин с обязательной локальной проверкой по S.

Ограничение исследования состоит в том, что вычислительное время и объём памяти не включены в критерий, поскольку представленные результаты содержат только показатели качества прогноза. Для развёртывания на маломощных edge-устройствах оценку S следует расширить нормированными задержкой вывода и размером модели. Кроме того, пороги повторного выбора должны учитывать сезонный дрейф и накопление новых данных.

IV. ПРАКТИЧЕСКАЯ ИНТЕГРАЦИЯ ПРОЦЕДУРЫ ВЫБОРА

Автоматизированный выбор модели реализуется как отдельный модуль конфигурирования прогнозного слоя. После накопления очередного временного окна данные проверяются на пропуски и упорядочиваются по времени. Затем для каждой комбинации режима HVAC, зоны и горизонта формируются обучающая и контрольная части, рассчитываются три метрики и определяется модель с минимальной оценкой S. Результат сохраняется в таблице конфигурации, которую локальные контроллеры считывают перед выполнением прогноза.

Сравнение с моделью постоянства дополнительно показывает практический эффект применения Random Forest. Во всех шести сочетаниях режима и горизонта MAE уменьшилась. Наиболее заметное снижение получено для нагрева на горизонте 60 мин – 43,14 %, наименьшее – для нагрева на 30 мин, 4,33 %. Следовательно, преимущество нелинейной модели нельзя переносить между режимами без локальной проверки.

Таблица 2. Снижение MAE модели Random Forest относительно Persistence

Режим	Горизонт, мин	MAE Persistence, °C	MAE Random Forest, °C	Снижение MAE, %
Охлаждение	15	0,1002	0,0680	32,12
Охлаждение	30	0,1304	0,0948	27,29
Охлаждение	60	0,1807	0,1246	31,09
Нагрев	15	0,0817	0,0627	23,21
Нагрев	30	0,1132	0,1083	4,33
Нагрев	60	0,1755	0,0998	43,14

В эксплуатационном контуре повторный выбор следует выполнять по расписанию либо при обнаружении дрейфа: росте контрольной MAE, изменении режима HVAC или существенном изменении распределения входных признаков. Чтобы исключить частое переключение моделей из-за случайных колебаний метрик, новая конфигурация должна приниматься только при устойчивом улучшении на нескольких последовательных контрольных окнах. Предыдущее назначение необходимо сохранять для быстрого отката.

Для edge-реализации процедура разделяется на два контура. Обучение и сравнительная оценка выполняются на шлюзе или сервере, где доступны память и вычислительные ресурсы. Локальный контроллер получает выбранную модель, её версию, перечень признаков и допустимый диапазон входных данных. Перед активацией проверяются целостность файла модели, совместимость схемы признаков и контрольный прогноз. Такой порядок снижает нагрузку на устройство и обеспечивает воспроизводимость обновления.

Расширение критерия для промышленного применения должно учитывать задержку вывода, объём памяти и энергозатраты. Эти показатели нормируются внутри той же локальной задачи и включаются в S с весами, определяемыми требованиями контроллера. При ограничении по времени сначала исключаются модели, не удовлетворяющие предельной задержке, после чего среди допустимых вариантов выбирается решение с минимальной интегральной оценкой качества.

Перед вводом выбранной модели в эксплуатацию рекомендуется проводить теневую проверку без воздействия на управляющий сигнал. В течение контрольного периода новая и действующая модели параллельно формируют прогнозы, а система регистрирует MAE, RMSE, R^2 и долю пропущенных вычислений. Замена допускается, если новая модель сохраняет преимущество для требуемого горизонта и не нарушает ограничения по времени отклика. Такой протокол отделяет статистический выигрыш на исторической выборке от пригодности модели в реальном контуре управления.

Отдельной обработки требуют нештатные ситуации. При пропуске входного признака контроллер может использовать последнее корректное значение только в пределах заранее установленного интервала; после его превышения следует перейти к безопасному базовому алгоритму. Выход температуры или расхода воздуха за диапазон обучающей выборки должен фиксироваться как признак дрейфа. Если контрольная ошибка превышает допустимый порог, назначение модели блокируется до повторной оценки. Журнал выбора должен содержать режим, зону, горизонт, версию данных, значения метрик и причину переключения, что обеспечивает проверяемость решения.

V. ЗАКЛЮЧЕНИЕ

Разработана процедура автоматизированного выбора краткосрочной прогнозной модели для многозонной IoT-системы теплового управления. На 24 локальных задачах Random Forest обеспечил минимальный MAE в 14 случаях и минимальную интегральную оценку в 13 случаях; XGBoost выбран по интегральному критерию в 9 случаях, Linear Regression – в 2. Ансамблевые модели не имели отрицательных R^2, тогда как базовые методы демонстрировали локальные отказы. Результаты показывают, что назначение модели должно учитывать режим HVAC, зону и горизонт, а не основываться только на общей средней ошибке. Предложенный критерий может использоваться как модуль конфигурирования прогнозного слоя децентрализованной системы управления.

ЛИТЕРАТУРА

[1] Yoon Y., Jung S., Im P., Gehl A. C. Datasets of a multizone office building under different HVAC system operation scenarios // Scientific Data. 2022. Vol. 9. Art. 775. DOI: 10.1038/s41597-022-01858-6.

[2] Breiman L. Random forests // Machine Learning. 2001. Vol. 45. P. 5–32. DOI: 10.1023/A:1010933404324.

[3] Chen T., Guestrin C. XGBoost: A scalable tree boosting system // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. P. 785–794. DOI: 10.1145/2939672.2939785.

[4] Amasyali K., El-Gohary N. M. A review of data-driven building energy consumption prediction studies // Renewable and Sustainable Energy Reviews. 2018. Vol. 81. P. 1192–1205. DOI: 10.1016/j.rser.2017.04.095.

[5] Zhao H.-X., Magoulès F. A review on the prediction of building energy consumption // Renewable and Sustainable Energy Reviews. 2012. Vol. 16, No. 6. P. 3586–3592. DOI: 10.1016/j.rser.2012.02.049.

[6] Bergstra J., Bengio Y. Random search for hyper-parameter optimization // Journal of Machine Learning Research. 2012. Vol. 13. P. 281–305.

AUTOMATED PREDICTIVE MODEL SELECTION FOR MULTI-ZONE IoT THERMAL CONTROL SYSTEMS

B. Mailykhanova[1,2], G. Kashaganova[3]

[1]Al-Farabi Kazakh National University, Almaty, Kazakhstan, bulgyn.mail@gmail.com

[2]Kazakh National Agrarian Research University, Almaty, Kazakhstan

[3]K. I. Satbayev Kazakh National Research Technical University, Almaty, Kazakhstan

Abstract: an automated procedure is proposed for selecting a short-term temperature prediction model for decentralized control of a multi-zone building. MAE, RMSE, and R^2 are normalized within each HVAC mode, zone, and horizon combination and combined into an integral score. The procedure was evaluated using 96 results for four zones, cooling and heating modes, and 15-, 30-, and 60-min horizons. Random Forest achieved the lowest MAE in 14 of 24 tasks and the lowest integral score in 13 tasks; XGBoost was selected in 9 tasks. The resulting rules support mode- and horizon-aware model assignment to local controllers.

Keywords: automated model selection, IoT, multi-zone building, forecasting, Random Forest, XGBoost.