ПРОГНОЗИРОВАНИЕ ЭЛЕКТРОПОТРЕБЛЕНИЯ В АГРЕГИРОВАННЫХ ГРУППАХ, АНСАМБЛЕВЫЕ МОДЕЛИ И МОДЕЛИ СЕЛЕКТИВНОГО ВЫБОРА

АВТОР:

А.В. КЫЧКИН
К.Т.Н., НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ВЫСШАЯ ШКОЛА ЭКОНОМИКИ, ФИЛИАЛ В Г. ПЕРМЬ (НИУ ВШЭ – ПЕРМЬ)

Прогнозные модели активно используются в электроэнергетике при построении нагрузки энергопотребителей. Предопределенный прогнозируемый профиль энергопотребления агрегированной группы позволяет центру управления спросом на электроэнергию принять корректное решение о целесообразности участия центра в происходящих событиях в системе.

Ключевые слова: машинное обучение; управление спросом на электроэнергию; электропотребление; краткосрочное прогнозирование; нейросетевые модели; ансамблевые модели; модели селективного выбора.

Прогнозирование уровней энергопотребления позволяет формировать точный алгоритм управления энергосистемой

ВВЕДЕНИЕ

Прогнозные модели активно используются в электроэнергетике для построения профилей нагрузки одиночных энергопотребителей на сутки вперед, где показывают достаточно высокую точность, например 90 % и более [2]. Аналогичные модели применяются и для групп энергопотребителей, подключенных к одной подстанции или находящихся в одном энергорайоне. Предопределенный на сутки вперед профиль энергопотребления агрегированной группы позволяет принять решение о целесообразности участия в событиях ценозависимого управления спросом на электроэнергию [11, 16]. Возможность точного предсказания энергопотребления позволяет также формировать сценарии для управления накопителями энергии, что в совокупности с моделями ценообразования может быть использовано для балансировки генерации и потребления, в том числе с применением возобновляемых источников энергии [5].

В случае работы с профилем групповой нагрузки обеспечить требуемое качество прогноза становится сложнее по сравнению с прогнозом для одиночного потребителя. Кроме того, одни и те же прогнозные модели могут обеспечивать разную точность на протяжении года. Во многом это связано с тем, что характер отдельных нагрузок меняется со временем, сильное влияние могут оказывать изменяющиеся технологические режимы не согласованных друг с другом энергопотребителей, а также пользователи (персонал) и климатические факторы [3]. Практика построения прогнозов в группах показывает необходимость постоянного обновления параметров моделей, изменения их конфигураций и даже выбора новых (табл. 1).

В некоторых случаях сочетание различных моделей прогнозирования может уменьшить суммарное отклонение прогнозных значений относительно исходного ряда данных, тем самым достигается более высокая точность по сравнению с любой из моделей по отдельности [7]. Данный подход подразумевает комбинирование моделей (объединение нескольких прогнозов) с некоторыми весовыми коэффициентами. Значение коэффициентов необходимо пересчитывать каждый раз по мере получения новых фактических данных, что становится доступным в условиях развития технологий интернета вещей и умного учета [1, 4, 5]. Также известны методы, реализующие переключение между лучшими прогнозными моделями. В этом случае для каждого периода прогноза выбирается только одна модель, показывающая лучший результат в предыдущих периодах или в аналогичных условиях.

Использование ансамблей или адаптивного переключения между моделями можно рассматривать как вариации экспертного обучения [6, 7, 12]. В статье приведена оценка распределения относительных весов, назначенных отдельным моделям в результате применения различных стратегий. Показано,

ОБОЗНАЧЕНИЯ И АББРЕВИАТУРЫ

Обозначение/аббревиатура	Значение
d	Индекс дня
t	Индекс времени
N	Количество измерений/прогнозов электрической нагрузки в пределах суток
T	Дискретность измерений/прогнозов электрической нагрузки в минутах (15 мин)
$Y_d(t)$	Нагрузка в период t в день d
$\hat{Y}_d^x(t)$	Прогноз нагрузки в период t в день d моделью x, где x будет заменено на название конкретной модели
AMSG1	Алгоритм выбора модели по точности прогноза за предыдущие сутки
AMSGN	Алгоритм отбора по прогнозу за N предыдущих дней
AMSMLP	Алгоритм отбора, основанный на многослойном персептроне
AR	Авторегрессионная модель
CLD	Устойчивая модель, учитывающая N дней одного и того же дня недели
EMLR	Ансамблевая регрессионная модель
EMMLP	Ансамблевая модель многослойного персептрона
EMWAF	Ансамблевая модель взвешенного среднего предсказания
HW	Модель Хольта–Винтерса
N-day	Устойчивая модель для N предыдущих дней
N-same-day	Устойчивая модель для N предыдущих одних и тех же дней
PAR	Устойчивая авторегрессионная модель
SPNN	Сезонная устойчивая нейросетевая модель
SPR	Сезонная устойчивая регрессионная модель
STLF	Краткосрочное прогнозирование нагрузки
RLS	Рекурсивный метод наименьших квадратов
RMSE	Среднеквадратическая ошибка

Таблица 1

что данные модели могут демонстрировать немного лучшие результаты при большом временном горизонте, так как они всегда приспосабливаются к лучшей на данный момент модели прогнозирования.

Сформулирован основной вклад исследования в рамках проекта Международной исследовательской группы «Разработка цифровой модели прогнозирования и ценозависимого управления спросом на электроэнергию, потребляемую подземными горнодобывающими предприятиями» (2020) в решение поставленной задачи представлена вариация сезонной устойчивой регрессионной модели (SPR), детально рассмотренная автором в статье, опубликованной в предыдущем номере журнала [2]. Модель позволяет определить относительную значимость низко- и высокочастотных устойчивых и временных характеристик с использованием нелинейной аппроксимации на нейронной сети. На основе реальных данных модели сравниваются со стандартными методами прогнозирования.

ВКЛАД ИССЛЕДОВАНИЯ В РАМКАХ ПРОЕКТА МЕЖДУНАРОДНОЙ ИССЛЕДОВАТЕЛЬСКОЙ ГРУППЫ В РЕШЕНИЕ ЗАДАЧИ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ ЭНЕРГОПОТРЕБЛЕНИЯ АГРЕГИРОВАННЫХ НАГРУЗОК НА ОСНОВЕ АНСАМБЛЕВЫХ МОДЕЛЕЙ И МОДЕЛЕЙ СЕЛЕКТИВНОГО ВЫБОРА

Одна из самых больших трудностей при краткосрочном прогнозировании агрегированной нагрузки заключается в неопределенности, возникающей при одиночных нагрузках даже на относительно коротком горизонте прогнозирования. Неопределенность вызвана случайными факторами, связанными с персоналом и режимами работы электрооборудования. Тем не менее, каким бы случайным ни было поведение объекта исследования, существуют определенные признаки (features), с помощью которых можно охарактеризовать нагрузку (подробнее об извлечении признаков см. [8, 9]). На основе данного подхода создана вариация модели SPR — модель SPNN, оперирующая ансамблями признаков, характерными для электрической нагрузки в зданиях и сооружениях, энергопотребление которых связано с деятельностью человека и погодными условиями.

Использование регрессионных моделей (PAR и SPR) также позволяет определить значимость (вес) рассматриваемых признаков, повышая тем самым интерпретируемость производных моделей. Однако с учетом того, насколько сильным может быть предположение моделей PAR и SPR о линейных зависимостях, в исследовании также используется многоуровневая архитектура персептрона (SPNN) для изучения возможных нелинейных зависимостей между прогнозом и признаками, рассмотренными в модели SPR. В работе показано отличие точности полученных моделей от точности исходной линейной регрессии, а также то, что линейные зависимости, обнаруженные моделью SPR, обеспечивают более точный прогноз по сравнению с нелинейными зависимостями, найденными с помощью модели SPNN.

Для более эффективного использования различных признаков, рассмотренных в упомянутых ранее моделях, в данной работе предлагаются ансамблевые модели и адаптивные стратегии переключения между моделями. Поскольку различные модели могут работать по-разному при меняющихся погодных условиях или различном поведении пользователей/оборудования, сочетание моделей или переключение между моделями может еще больше повысить точность прогнозирования. С этой целью проведен сравнительный анализ между альтернативными ансамблевыми стратегиями и стратегиями переключения между моделями, показавший потенциальные преимущества таких стратегий по сравнению с использованием отдельных моделей.

НЕЙРОСЕТЕВОЕ ПРЕДСТАВЛЕНИЕ МОДЕЛИ SPNN, ГДЕ $F_1 \ldots F_{15}$ — ПРИЗНАКИ ПРОФИЛЯ НАГРУЗКИ [2]

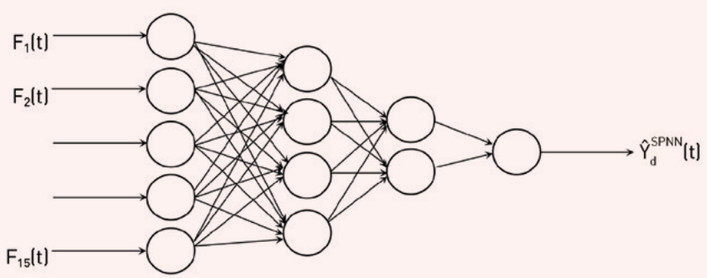

Рис. 1

Вклад исследования в рамках проекта Международной исследовательской группы заключается в следующем.

1. В работе представлена модель SPNN как вариация модели SPR, специально разработанная для прогнозирования электрической нагрузки с учетом выявления нелинейных зависимостей между признаками с помощью нейронной сети.

2. Исследуется потенциал применения ансамблевых моделей и стратегии адаптивного переключения между моделями для сокращения ошибок прогнозирования. Так, комбинирование прогнозов различных моделей или переключение между прогнозами моделей используется для уменьшения непостоянства точности прогноза отдельных моделей.

3. В результате проведенного эксперимента с использованием реальных данных об энергопотреблении группы объектов оценена точность рассмотренных моделей прогнозирования.

Данная статья расширяет информацию, представленную в [2], поскольку содержит более подробное сравнение со стандартными методами прогнозирования, в том числе описание модели SPNN. Кроме того, выполнено исследование нескольких вариаций ансамблевых моделей и адаптивных методов переключения между моделями.

УСТОЙЧИВЫЕ РЕГРЕССИОННЫЕ МОДЕЛИ НЕЙРОННОЙ СЕТИ

Сезонные устойчивые регрессионные модели (SPR) нацелены на комбинирование большого количества признаков, что позволяет лучше учитывать сезонную составляющую [2]. Обучение модели SPR сводится к задаче оптимизации линейной регрессии, которая итеративно решается с использованием рекурсивного метода наименьших квадратов.

Зависимость электрической нагрузки от признаков, рассмотренных в модели SPR, может быть нелинейной. Такие зависимости неизвестны заранее, но могут быть аппроксимированы с помощью нейронной сети. Для этого реализована вариация модели SPR — сезонная устойчивая модель в виде нейронной сети (SPNN) — многослойного персептрона (MLP). Данная вариация предназначена для выявления возможных нелинейных зависимостей между нагрузкой и признаками, используемыми в SPR-модели. Обучение модели SPNN — это обучение многослойного персептрона, т. е. нейронной сети с прямой связью, который активно применяется при решении задач прогнозирования энергопотребления [10, 15]. В качестве входных данных многослойного персептрона используются признаки профиля нагрузки [2]. Архитектура сети для SPNN схематично показана на рис. 1.

Поскольку речь идет о прогнозировании временных рядов, обучение и валидация модели должны проводиться итеративно с разделением данных на небольшие информационные пакеты. Например, тренировочный набор данных был разделен на пакеты с фиксированным размером N = 96 точек, что соответствует данным, полученным за одни сутки.

Такой тренировочный набор за день d может быть представлен в виде матрицы $F_d \triangleq [f_1(t), f_2(t) \ldots f_{15}(t)]_{t = 1 \ldots N}$, дополняющей уже имеющиеся данные за все дни до дня d. После обновления весов нейронной сети с помощью дополнительных данных матрицы F_d обновленная модель прогнозирует нагрузку на следующий день (d + 1), а затем проводится тестирование точности модели — процедура кросс-валидации.

Вычисления продолжаются итеративно, одни и те же действия повторяются в конце каждого дня d: а) дополнение обучающей выборки данными, собранными за прошедший день d; б) тестирование модели путем сравнения прогноза, полученного в конце дня (d — 1) с фактическими измерениями нагрузки за день d. Схема этого итеративного процесса приведена на рис. 2.

Для настройки многослойного перцептрона использовались следующие рекомендации: входной слой состоит из количества нейронов, совпадающего с количеством признаков [2]; скрытые слои содержат по несколько нейронов на каждом слое; выходной слой состоит из одного нейрона, так как решается задача регрессии. Этот нейрон на выходе дает прогноз нагрузки для определенного временного интервала. В качестве функции активации нейронов скрытых слоев применялась усеченная линейная функция активации (rectified linear activation function), а для нейронов выходного слоя — линейная функция активации (linear activation function). В качестве функции потерь использовалась среднеквадратическая ошибка.

Для вычисления количества нейронов на скрытом слое применялась следующая формула [15]:

$$N_{nodes} = \frac{4N_i^2 + 3}{N_i^2 - 8}, \quad (1)$$

где N_i — число нейронов входного слоя.

КРОСС-ВАЛИДАЦИЯ МОДЕЛИ SPNN

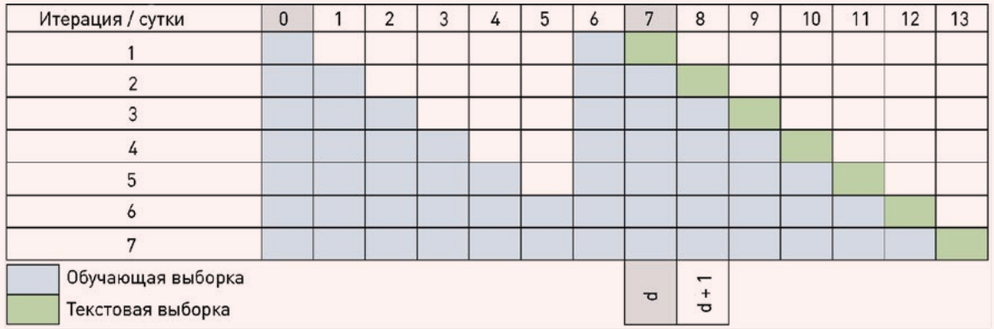

Рис. 2

В результате была получена нейронная сеть с $N_{nodes} = 4$.

Так как на вход нейронной сети данные подаются пакетами фиксированной длины с большими выбросами, данные должны быть нормализованы, т. е. приведены к единому диапазону от 0 до 1. Нормализация осуществляется с помощью выражения

$$y'_t = \left[\frac{x_t - x_{min}}{x_{max} - x_{min}}\right] [Y_{max} - Y_{min}] + Y_{min}, \qquad (2)$$

где x_t, x_{min}, x_{max} — фактическое, минимальное и максимальное значение нагрузки;

Y_{min}, Y_{max} — минимальное и максимальное целевые значения нагрузки.

АНСАМБЛЕВЫЕ МОДЕЛИ

Необходимо учитывать, что каждая из рассмотренных моделей может работать по-разному при различных условиях эксплуатации и/или поведении персонала (пользователей). Кроме того, моделям может потребоваться разное время на обучение, например, для модели SPNN необходимо больше времени, чем для модели SPR. Это может привести к различным результатам на этапах моделирования. Например, модель SPR может работать лучше на начальном этапе эксперимента, в то время как SPNN может дать хорошие результаты позже. В данном разделе описано несколько вариантов создания ансамблей на основе рассмотренных моделей, что позволяет более эффективно использовать их возможности, а также применять модели, показывающие в настоящий момент времени лучшие результаты. Иными словами, ансамблевые модели — это комбинации базовых моделей. Под базовыми моделями подразумеваются все рассмотренные нами ранее модели, включая стандартные и устойчивые [2].

АНСАМБЛЕВАЯ РЕГРЕССИОННАЯ МОДЕЛЬ (EMLR)

Эта ансамблевая модель представляет собой решение задачи линейной регрессии, в которой регрессорами выступают выходы базовых прогнозных моделей, описанные в [2]. Для вычисления весов каждой из базовых моделей используется рекурсивный метод наименьших квадратов (RLS). Предполагается, что наибольшие веса следует присваивать моделям, дающим наиболее точные прогнозы.

АНСАМБЛЕВАЯ МОДЕЛЬ МНОГОСЛОЙНОГО ПЕРСЕПТРОНА (EMMLP)

Эта модель основана на том же принципе, что и модель EMLR. Тем не менее веса вычисляются не методом RLS, а с помощью нелинейной регрессии (MLP — многослойный персептрон).

АНСАМБЛЕВАЯ МОДЕЛЬ ВЗВЕШЕННОГО СРЕДНЕГО ПРЕДСКАЗАНИЯ (EMWAF)

Выбор модели, показавшей лучшие результаты по сравнению с другими моделями в течение определенного количества дней (например, в течение N предыдущих дней, как упоминалось выше), не обязательно гарантирует, что последовательность выбранных моделей будет лучшей из возможных.

Такой критерий отбора может привести к необоснованному переключению между моделями (например, из-за небольших вариаций в RMSE предыдущего дня), что может ухудшить совокупную эффективность последовательности выбранных моделей в долгосрочной перспективе.

Для обеспечения высокой точности при адаптивном выборе модели в долгосрочной перспективе используется алгоритм экспоненциально взвешенного среднего предсказания [7, 13], относящегося к экспертным моделям. В таком контексте базовые модели рассматриваются как «эксперты» i = 1 ... m, каждый из которых представляет свой прогноз временного ряда на следующие сутки.

Пусть $Y_d \triangleq \{y_{d,t}\}_{t=1\ldots96}$ — это последовательность измерений нагрузки за день d,

$F_{i,d} \triangleq \{f_{i,d,t}\}_{t=1\ldots96}$ — последовательность предсказаний модели-эксперта i на следующий день,

$P_d \triangleq \{p_{d,t}\}_{t=1\ldots96}$ — последовательность предсказаний алгоритма, который в общем виде может быть взвешенной комбинацией доступных моделей-экспертов (базовых моделей).

СТРАТЕГИИ АДАПТИВНОГО ПЕРЕКЛЮЧЕНИЯ МЕЖДУ ПРОГНОЗАМИ МОДЕЛЕЙ

АЛГОРИТМ ВЫБОРА МОДЕЛИ ПО ТОЧНОСТИ ПРОГНОЗА ЗА ПРЕДЫДУЩИЕ СУТКИ (AMSG1)

Метод основан на подходе, подразумевающем, что только одна модель прогноза, которая была наиболее точной вчера, должна быть использована сегодня. Метод работает следующим образом. В конце каждого дня d вычисляется RMSE на текущий день для каждой из моделей, перечисленных в [2]. Полученные значения RMSE на день d сравниваются друг с другом, затем определяются наименьшее значение ошибки и модель прогноза, показавшая эту наименьшую ошибку. Для дня прогноза профиля нагрузки на день (d + 1) будет выбрана модель, предоставившая наименьшее значение RMSE для дня (d + 1).

Этот метод, несмотря на свою простоту, реализует ежедневный отбор лучшей модели прогноза из предыдущего дня. Очевидно, что если только одна конкретная модель ежедневно показывает низкую ошибку, то она будет выбрана для последующих прогнозов. Если другая модель показывает лучший результат в день d, то метод будет применен к дню (d + 1). К сожалению, он имеет низкую эффективность, если каждый день лучший результат показывают разные модели. В таком случае метод будет постоянно приводить к ошибочному выбору и его точность будет низкой. Для преодоления этого недостатка необходимо выбрать модель по результатам оценки за несколько дней.

RMSE МОДЕЛЕЙ, ОБУЧЕННЫХ НА ОСНОВЕ ДАННЫХ ПО ЭНЕРГОПОТРЕБЛЕНИЮ ГРУППЫ А

Модель	Один день 03.02.2017	Одна неделя 01.02.2017 – 07.02.2017	Один месяц 17.11.2016 – 16.12.2016	Полгода 16.11.2016 – 15.05.2017
Устойчивые модели				
N-same-day	405	343	304	306
N-day	369	305	265	340
HW	372	343	282	316
SARIMA	389	318	264	306
PAR	375	301	268	312
SPR	348	255	223	296
SPRMLP	364	304	276	324
Ансамблевые модели				
EMLR	368	297	257	303
EMMLP	384	303	269	317
EMWAF	340	251	222	296
Модели селективного выбора				
AMSG1	348	270	236	295
AMSGN	376	259	224	285
AMSMLP	348	255	223	296

Таблица 2

RMSE МОДЕЛЕЙ, ОБУЧЕННЫХ НА ОСНОВЕ ДАННЫХ ПО ЭНЕРГОПОТРЕБЛЕНИЮ ГРУППЫ Б

Модель	Один день 03.02.2017	Одна неделя 01.02.2017 – 07.02.2017	Один месяц 17.11.2016 – 16.12.2016	Полгода 16.11.2016 – 15.05.2017
Устойчивые модели				
N-same-day	105	107	93	96
N-day	69	90	84	88
HW	82	95	91	94
SARIMA	71	91	85	88
PAR	75	94	81	87
SPR	66	93	84	87
SPRMLP	68	92	84	86
Ансамблевые модели				
EMLR	72	95	86	89
EMMLP	65	89	**80**	**84**
EMWAF	**62**	**88**	81	86
Модели селективного выбора				
AMSG1	71	94	86	87
AMSGN	68	93	83	86
AMSMLP	66	95	85	87

Таблица 3

АЛГОРИТМ ОТБОРА ПО ПРОГНОЗУ ЗА N ПРЕДЫДУЩИХ ДНЕЙ (AMSGN)

Метод выбирает лучшую модель по сумме ее ошибок RMSE за предыдущие N дней. Модель с наименьшей суммой считается лучшей и используется для прогнозирования нагрузки на следующий день. Важным наблюдением является то, что иногда день ото дня лучшими оказываются разные модели прогноза. Другими словами, если в день d самую низкую ошибку показала одна модель, то в день (d + 1) лучшей может стать другая модель. В этом случае следует обратить внимание на то, как часто одна модель была лучшей. Для более корректной оценки следует рассмотреть накопленную сумму ошибок.

Проведенный анализ поведения моделей в разные интервалы времени показал, что выбор модели нельзя делать на основе большого интервала, так как это может привести к потере качества прогноза на локальных участках энергетического профиля, например в рамках одной или двух недель. В связи с этим метод выбора модели прогноза должен учитывать наименьшую накопленную сумму ошибки RMSE и работать внутри скользящего окна, размер которого сопоставим с недельным интервалом. Предложенный метод выбирает лучшую модель в соответствии с меньшей суммой RMSE каждой из моделей, приведенных в [2] за предыдущие N дней; рекомендуется рассматривать N = 10 дней.

АЛГОРИТМ ОТБОРА, ОСНОВАННЫЙ НА МНОГОСЛОЙНОМ ПЕРСЕПТРОНЕ (AMSMLP)

Аналогично методу AMSGN модель AMSMLP выбирает лучшую модель в соответствии с ошибкой RMSE за предыдущий день, но выбор основан на работе нейронной сети. В этом случае нейронная сеть используется для решения проблемы классификации. Число входов и число выходов в сети равно числу моделей, приведенных в [2].

Обучение сети происходит с использованием метода обратного распространения ошибки следующим образом. На каждый из входов подается значение RMSE соответствующей по номеру модели прогноза за день (d — 1). На выход, соответствующий номеру той модели, которая показала самую низкую ошибку в день d, подается 1. На остальные выходы подается 0. Далее производится подстройка весов нейронной сети. Так повторяется для всех дней, доступных в обучающей выборке данных. Как только модель обучилась, ее можно использовать для выбора метода прогноза на сутки вперед. Для этого следует подать на каждый вход значения RMSE соответствующих моделей прогноза, доступных за один предыдущий день. Нейронная сеть, используя полученные в ходе обучения веса, проведет классификацию и активирует один из своих выходов, соответствующих номеру модели прогноза. Этот метод хорошо улавливает нелинейные зависимости в поведении величины RMSE за предыдущие дни для различных моделей и пытается выбрать наилучшую модель для прогноза на следующий день.

ОЦЕНКА ТОЧНОСТИ ПРОГНОЗНЫХ МОДЕЛЕЙ

В качестве метрики для оценки точности предсказания моделей была выбрана среднеквадратическая ошибка (RMSE). В табл. 2 и 3 представлены оценки RMSE рассмотренных ранее моделей, полученные в результате обучения моделей на основе реальных данных нагрузок при энергопотреблении двух агрегированных групп зданий.

Из данных таблиц видно, что стабильно худшей моделью стала модель N-same-day, использующая данные за три предыдущих таких же дня в неделе, например, три предыдущих понедельника. Удовлетворительные результаты показала модель N-days, использующая среднее значение нагрузки за предыдущие последовательные N = 10 дней. Также видно, что ошибки устойчивых моделей обычно выше ошибок ансамблевых моделей и моделей селективного выбора.

На рис. 3 представлено распределение абсолютных значений ошибок RMSE для устойчивых моделей за каждые сутки в течение интервала времени июнь 2016 — февраль 2017 г. Абсолютное значение ошибки в течение года сильно плавает, что обусловлено изменениями величины энергопотребления. Таким образом, в разные интервалы времени лучшей оказываются различные модели, однако в среднем самую низкую ошибку из устойчивых показала модель SPR.

АБСОЛЮТНЫЕ ЗНАЧЕНИЯ ОШИБОК RMSE ДЛЯ УСТОЙЧИВЫХ МОДЕЛЕЙ ЗА КАЖДЫЕ СУТКИ

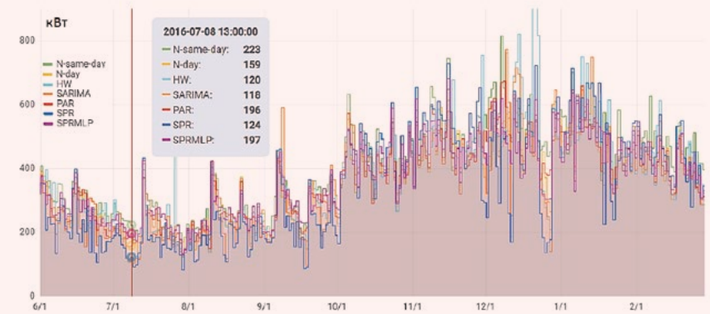

Рис.3

СРЕДНЕЕ НАКОПЛЕННОЕ ЗНАЧЕНИЕ ОШИБКИ RMSE ДЛЯ АНСАМБЛЕВЫХ МОДЕЛЕЙ И МОДЕЛЕЙ СЕЛЕКТИВНОГО ВЫБОРА В ИНТЕРВАЛЕ ВРЕМЕНИ ОДИН ГОД

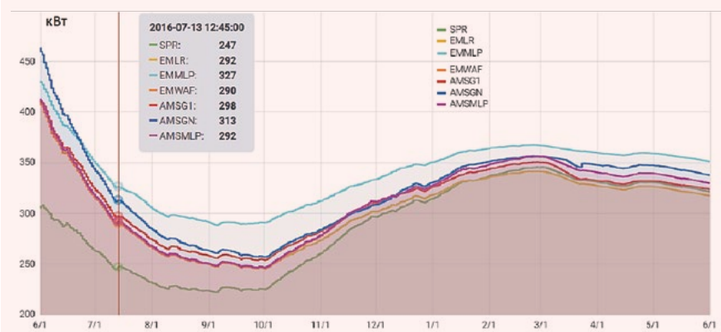

Рис. 4

РАСПРЕДЕЛЕНИЕ ВЕСОВ ДЛЯ МОДЕЛЕЙ ПРОГНОЗА В СОСТАВЕ АНСАМБЛЕВОЙ РЕГРЕССИОННОЙ МОДЕЛИ

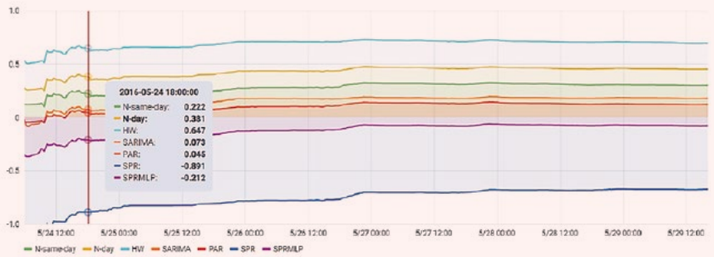

Рис. 5

РАСПРЕДЕЛЕНИЕ ВЕСОВ ДЛЯ МОДЕЛЕЙ ПРОГНОЗА В СОСТАВЕ ВЗВЕШЕННОГО СРЕДНЕГО ПРЕДСКАЗАНИЯ

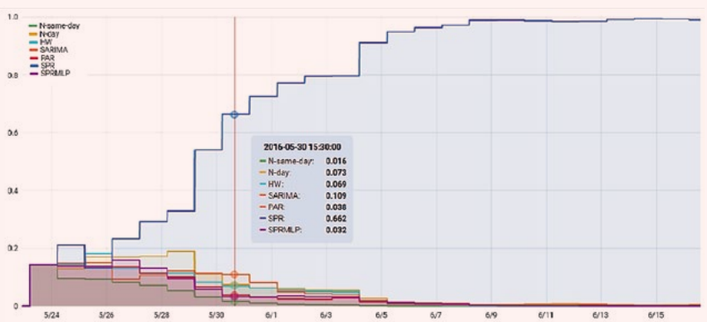

Рис. 6

На рис. 4 показано среднее накопленное значение ошибки RMSE для ансамблевых моделей и моделей селективного выбора в течение интервала времени июнь 2016 — май 2017 г. по сравнению с ошибкой SPR. Из рисунка видно, что модель SPR в первые полгода лидирует по точности. Однако с увеличением периода времени ее обгоняет по точности модель EMWAF. Также аналогичные результаты начинает показывать модель AMSG1, к ней приближается AMSMLP. Экспериментально установлено, что в интервалы времени свыше одного года адаптивные модели будут всегда превосходить модель SPR.

На рис. 5 показано, каким образом ансамблевая модель EMLR проводит адаптацию коэффициентов. Видно, что модель смогла достаточно быстро рассчитать веса устойчивых моделей. Полученные примерно на третий день веса меняются в течение года незначительно, что говорит об успешном решении задачи линейной регрессии рекурсивным способом. К сожалению, EMLR не смогла обеспечить высокую точность по сравнению с другими моделями, что говорит о наличии нелинейной зависимости между значениями прогнозов.

Модель EMMLP успешно определяет нелинейную зависимость в данных для группы А (см. табл. 3). Для обучения нейронной сети необходима выборка данных об энергопотреблении примерно за месяц. Однако при аналогичном периоде обучения эта же модель показала низкую точность в случае с данными группы Б в табл. 2, что говорит о низкой робастности метода.

Модель EMWAF оказалась более робастной по отношению к прочим, она не чувствительна к сезонности, ее точность была всегда лучшей (или равной лучшей модели) на обучающих выборках продолжительностью до одного месяца. Данная ансамблевая модель быстро успевает выбирать веса устойчивых прогнозных моделей; на рис. 6 представлен процесс адаптации при распределении весов.

Результаты селективного выбора прогнозных моделей с использованием AMSG1, AMSGN, AMSMLP приведены на рис. 7, на котором показано, что в течение года разные стратегии ведут себя по-разному, например, AMSMLP применяла в 99 % случаев только модель SPR. Точность AMSMLP была близка к точности AMSG1, которая показала большое число переключений между моделями. Следует отметить, что лучшей из предложенных стала модель AMSGN. Данная модель использовала в 61 % случаев SPR, в остальное время были задействованы модели N-day, PAR и SARIMA.

ЗАКЛЮЧЕНИЕ И ПЛАНИРУЕМЫЕ ИССЛЕДОВАНИЯ

Предложенные ансамблевые модели и модели селективного выбора значительно повышают качество прогнозирования энергопотребления для агрегированных нагрузок. Они обладают большей робастностью по отношению к устойчивым моделям, что делает их более устойчивыми к изменениям характера одиночных энергопотребителей, формирующих групповой профиль нагрузки, поведения пользователей и сезонным климатическим изменениям.

Повысить точность модели SPR за счет добавления нелинейных зависимостей не удалось. Нейронная сеть не всегда была самой точной и при работе в ансамблевой модели. Наиболее точной и адаптивной оказалась модель взвешенного среднего предсказания при размере обучающей выборки до одного месяца, что позволяет сделать вывод о целесообразности использования этой модели на ранних стадиях прогнозирования энергопотребления. По мере накопления обучающей выборки рекомендуется переключаться на модели AMSGN и EMMLP. Данная стратегия прогнозирования позволяет гарантированно повысить точность устойчивых моделей не менее чем на 5 %, что может стать критически важным при планировании событий управления спросом на электроэнергию.

РЕЗУЛЬТАТЫ РАБОТЫ СТРАТЕГИЙ СЕЛЕКТИВНОГО ВЫБОРА ПРОГНОЗНЫХ МОДЕЛЕЙ

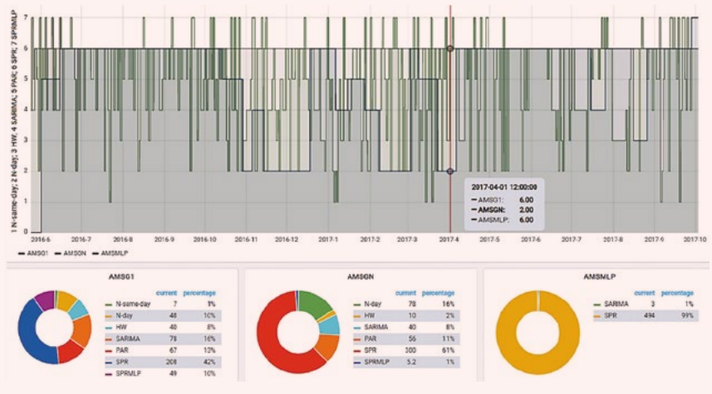

Рис. 7

ЛИТЕРАТУРА

1. Кычкин А.В., Артемов С.А., Белоногов А.В. Распределенная система энергомониторинга реального времени на основе технологии IoT//Датчики и системы. 2017. № 8–9. С. 49–55.
2. Кычкин А.В. Краткосрочное прогнозирование электропотребления на сутки вперед с использованием устойчивых моделей//Энергия единой сети. 2020. 12. С. 17–26.
3. Кычкин А.В., Микрюков Г.П. Метод обработки результатов мониторинга группы энергопотребителей//Энергобезопасность и энергосбережение. 2016. № 6. С. 9–14.
4. Кычкин А.В. Синтез системы удаленного энергетического мониторинга производства//Металлург. 2015. № 9. С. 20–27.
5. Файзрахманов Р.А., Франк Т., Кычкин А.В., Федоров А.Б. Рациональное управление потреблением энергии с помощью информационной системы энергоменеджмента My JEVis//Электротехника. 2011. № 11. С. 35–40.
6. Cancelo J.R., Espasa A., Grafe R. Forecasting the electricity load from one day to one week ahead for the Spanish system operator//International Journal of Forecasting. 2008. N 24 (4). P. 588–602.
7. Cesa-Bianchi N., Lugosi G. Prediction, Learning, and Games. New York, Cambridge University Press, 2006.
8. Christ M., Braun N., Neuffer J., Kempa-Liehr A.W. Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (TSFRESH — A Python package)//Neurocomputing. 2018. N 307. P. 72–77.
9. Christ M., Kempa-Liehr A.W., Feindt M. Distributed and parallel time series feature extraction for industrial big data applications. ArXiv:1610.07717 [cs] (May 2017).
10. Hippert H., Bunn D., Souza R. Large neural networks for electricity load forecasting: Are they overfitted?//International Journal of Forecasting. 2005. N 21 (3). P. 425–434.
11. Hong T., Fan S. Probabilistic electric load forecasting: A tutorial review, International//Journal of Forecasting. 2016. N 32 (3). P. 914–938.
12. Hong T., Gui M., Baran M.E., Willis H.L. Modeling and forecasting hourly electric load by multiple linear regression with interactions/In: IEEE PES General Meeting, 2010, P. 1–8.
13. Ljung L. System Identification: Theory for the User. 2nd ed. NJ: Prentice Hall Ptr, 1999.
14. Sheela K., Deepa S.N. Review on methods to fix number of hidden neurons in neural networks//Mathematical Problems in Engineering. 2013.06. doi:10.1155/2013/425740
15. Soares L.J., Medeiros M.C. Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data//International Journal of Forecasting. 2008. N 24 (4). P. 630–644.
16. Xu Y., Li N., Low S.H. Demand Response with Capacity Constrained Supply Function Bidding//IEEE Transactions on Power Systems. 2016. N 31 (2). P.1377–1394.