

УДК 004.6: 338.27:519.222

ПОВЫШЕНИЕ КАЧЕСТВА КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ ЭЛЕКТРОПОТРЕБЛЕНИЯ НА ОСНОВЕ КЛАСТЕРИЗАЦИИ ПАРАМЕТРОВ

Дулесов А.С., Шилов А.В.

ФГБОУ ВПО «Хакасский государственный университет имени Н.Ф. Катанова»,
Абакан, e-mail: dulesov@khsu.ru, antonshilov2015@yandex.ru

Проанализированы результаты прогнозирования почасового электропотребления (с 0-го по 23-й час суток) в республике Хакасия с использованием методов линейной регрессии, скользящих средних, нейронных сетей. Выявлена необходимость корректировки полученных прогнозных решений с помощью метода кластеризации параметров. Проведена процедура кластеризации согласно фактору «статус дня» (будний или праздничный/выходной). Представлено графическое изображение фактических и прогнозных значений электропотребления на сутки вперёд, полученных как без учёта, так и с использованием кластерного подхода. Сопоставительный анализ прогнозных ошибок позволил сделать вывод о повышении эффективности прогноза при выполнении предварительной кластеризации. Детальный анализ значений ошибок на участках перелома тренда, а именно: 8–9; 12–13; 17–18 часов, позволил выявить необходимость и дальнейшей корректировки прогнозных значений.

Ключевые слова: краткосрочный прогноз, кластеризация, ошибка прогноза, параметр, линейная регрессия, скользящие средние, нейросетевые методы

IMPROVING OF THE POWER CONSUMPTION SHORT-RANGE FORECASTING ACCURACY BASED ON CLUSTERING OF PARAMETERS

Dulesov A.S., Shilov A.V.

The Khakas State University by N.F. Katanov, Abakan,
e-mail: dulesov@khsu.ru, antonshilov2015@yandex.ru

The results of the hourly power consumption (from the zero to the twenty-third hour) forecasting in the Republic of Khakassia using linear regression, moving averages and neural networks are analyzed. The necessity of the correction of obtained predictive solutions using the clustering of parameters method is identified. The clustering procedure based on the «Day status» factor (workday or day off) is made. The graphical interpretation of actual and forecast power consumption values for the twenty-four hours both without accounting and with using clustering approach is given. The comparative analysis has led to the conclusion about the accuracy improving of the forecast values with applying of preliminary clustering based on the «Day status». The detailed analysis of errors in trend turning points: 8–9; 12–13; 17–18 hours revealed that there is the need to correct the forecast values further.

Keywords: short-range forecast, clustering, forecast error, parameter, linear regression, moving averages, neural networks

При краткосрочном прогнозировании электропотребления возникает проблема, заключающаяся в учете действия определённых факторов на процесс электропотребления, к которым можно отнести статус дня (будний или праздничный/выходной) и фактор сезонности (в зависимости от включения/отключения системы отопления).

В силу различной степени действия перечисленных выше факторов на объект исследования значения параметров могут варьироваться достаточно широко, что влечёт за собой наличие выбросов в выборке, и, как следствие, увеличение величины ошибки прогноза. В связи с этим эффективным методом повышения качества получаемых прогнозных решений является кластеризация [8–9]. Данная процедура позволяет выделить кластеры, в которых значения необходимых показателей будут достаточно близки между собой. Последующие дей-

ствия связаны с непосредственным прогнозированием на основе данных, находящихся в пределах обозначенных кластеров.

Далее в работе изложен анализ почасового прогнозирования электропотребления как без учёта кластерного подхода, так и с его применением.

Анализ прогнозных значений

В качестве методов прогнозирования широкое распространение получили статистические и интеллектуальные. Среди них были выбраны, согласно их преимуществам, приведённым в работах [5–6, 10], линейная регрессия, скользящие средние и нейросетевые методы. При прогнозировании почасового потребления электроэнергии Хакасии, МВт·ч, на временном лаге 3 месяца (6552 значения), получены значения средней абсолютной ошибки аппроксимации (MAPE), %, которые представлены

в табл. 1. В качестве факторов, влияющих на динамику величины электропотребления, согласно [3–4, 7], использована информация о буднем (рабочем) или праздничном/выходном дне, сезонности в зависимости от подключения/отключения системы отопления зданий. Инструментом прогнозирования являлась программа статистического анализа данных *IBM SPSS Statistics v. 20* и стандартный офисный программный продукт *MS Excel 2007*.

Осуществлена проверка используемых данных об электропотреблении на нормальность закона распределения с применением критерия Колмогорова – Смирнова.

Полученные значения, представленные в табл. 2, указывают на несоответствие данных нормальному закону распределения, за исключением 5-го часа, где значение асимптотической значимости равно 0,078.

Таким образом, перечисленные выше методы прогнозирования не позволяют получить достаточно точные прогнозные решения, в результате чего необходимо использование механизма корректировки. Одним из эффективных методов повышения качества прогноза является предварительная кластеризация данных.

Кластеризация параметров заключается, согласно [1–2], в разбиении объектов на относительно схожие, однородные группы согласно некоторым признакам.

В качестве данных, подлежащих кластеризации, использованы и проанализированы ретроспективные статистические значения потребляемой электроэнергии на выбранном временном лаге (3 месяца). Поскольку динамика электропотребления по будним и праздничным/выходным дням отличается, то целесообразно разбить значения по выделенным ранее кластерам.

Таблица 1

Величины средних абсолютных ошибок прогнозных решений, полученных с помощью различных методов прогнозирования

№ п/п	Метод прогнозирования	Вид метода прогнозирования	Средняя абсолютная ошибка (MAPE, %)
1.	Регрессия	Линейная	2,692
2.	Скользящая средняя	Простая ($n = 14$)	2,07
		Взвешенная ($n = 16$)	2,344
		Экспоненциальная ($n = 6, K = 0,143$)	2,599
3.	Нейронные сети	Многослойный перцептрон	2,072
		Радиальная базисная функция	2,799

Примечание. Параметр n – порядок скользящей средней; K – коэффициент сглаживания.

Таблица 2

Результаты одновыборочного критерия Колмогорова – Смирнова. Часы с ненормальным распределением данных

Параметры	Час 0	Час 1	Час 2	Час 3	Час 4	Час 6	Час 7	Час 8	Час 9	Час 10	Час 11	Час 12
Статистика Z Колмогорова – Смирнова	1,518	1,596	1,512	1,447	1,403	1,379	1,364	1,476	1,435	1,681	1,508	1,646
Асимптотическая значимость (двухсторонняя)	0,02	0,012	0,021	0,03	0,039	0,045	0,048	0,026	0,033	0,007	0,021	0,009

Окончание табл. 2

Параметры	Час 13	Час 14	Час 15	Час 16	Час 17	Час 18	Час 19	Час 20	Час 21	Час 22	Час 23
Статистика Z Колмогорова – Смирнова	1,663	1,646	1,663	1,595	1,496	1,386	1,463	1,597	1,674	1,672	1,629
Асимптотическая значимость (двухсторонняя)	0,008	0,009	0,008	0,012	0,023	0,043	0,028	0,012	0,007	0,007	0,01

Массив статистических данных электропотребления $W(T, H, D)$ включает в себя факторы: T – время; H – сезон в зависимости от отопления; D – день (будний или праздничный/выходной). Кластеры, выделенные на основании фактора (признака) «Будний или праздничный/выходной день», могут быть записаны соответственно в виде $W_{fd}(T, H)$ и $W_{do}(T, H)$, где $W_{fd}(T, H)$ – кластер, содержащий данные об электропотреблении в будние дни; $W_{do}(T, H)$ – кластер, содержащий данные об электропотреблении в праздничные/выходные дни.

После проведения процедуры кластеризации выполнено прогнозирование параметра электропотребления, используя данные вышеобозначенных кластеров. Сопоставительные величины полученных средних абсолютных ошибок аппроксимации без использования кластеризации и при работе с кластерами $W_{fd}(T, H)$ и $W_{do}(T, H)$ приведены в табл. 3.

Графические представления фактических и прогнозных значений на сутки вперёд с учётом и без использования кластеризации представлены на рис. 1–3.

Согласно табл. 3 и рис. 1–3, после проведения процедуры кластеризации по будним

и праздничным/выходным дням средняя абсолютная ошибка прогноза уменьшена, при этом наибольшая эффективность наблюдается у метода простой скользящей средней.

Значения электропотребления по часам были проверены на соблюдение нормального закона распределения случайной величины. Электропотребление по будним дням, согласно критерию Колмогорова – Смирнова, с 0-го по 6-й часы, а также с 18-го по 19-й, подчиняется нормальному закону. Касательно асимптотической значимости в пределах другого выделенного кластера $W_{do}(T, H)$, она стала выше по сравнению с данными без использования кластеризации, указывая на соответствие величин нормальному закону распределения.

Детальные значения абсолютных прогнозных ошибок, отмеченные на проблемных часовых промежутках, а именно: с 8-го по 9-й, с 12-го по 13-й, с 17-го по 18-й часы, представлены в табл. 4. Данные участки охарактеризованы резкой сменой динамики изменения электропотребления, связанной с временем обеда, прихода и ухода с рабочих мест сотрудников в организациях и на предприятиях.

Таблица 3

Величины средних абсолютных ошибок прогнозных решений, кластеры «Будние дни» и «Праздничные/Выходные дни»

№ п/п	Метод прогнозирования	Вид метода прогнозирования	Средняя абсолютная ошибка (MAPE, %)		
			Все дни	Будние дни	Праздничные/выходные дни
1.	Регрессия	Линейная	2,69	1,6	2,44
2.	Скользящая средняя	Простая	2,07	1,61	1,45
		Взвешенная	2,34	1,63	1,47
		Экспоненциальная	2,6	2,32	2,3
3.	Нейронные сети	Многослойный перцептрон	2,07	1,93	1,73
		Радиальная базисная функция	2,8	2,63	1,91

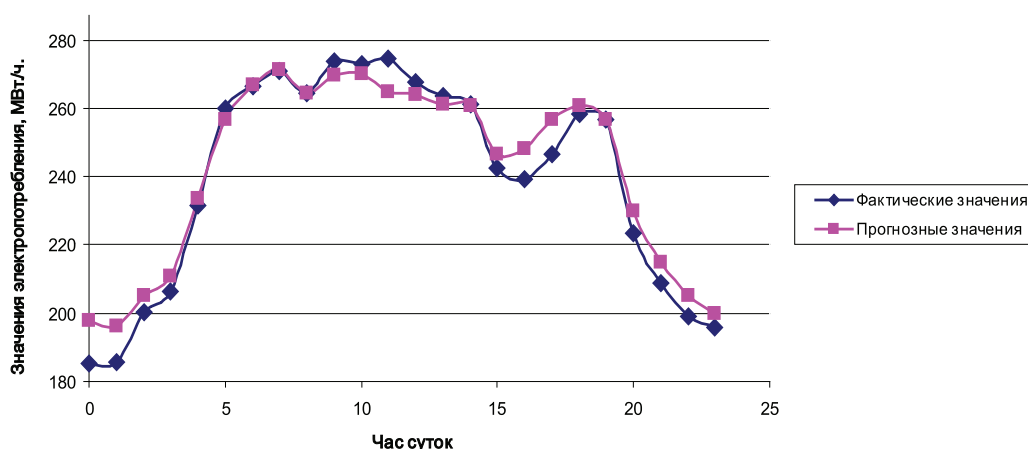


Рис. 1. Графики фактических и прогнозных значений электропотребления в Хакасии на сутки вперёд, метод простой скользящей средней, без кластеризации

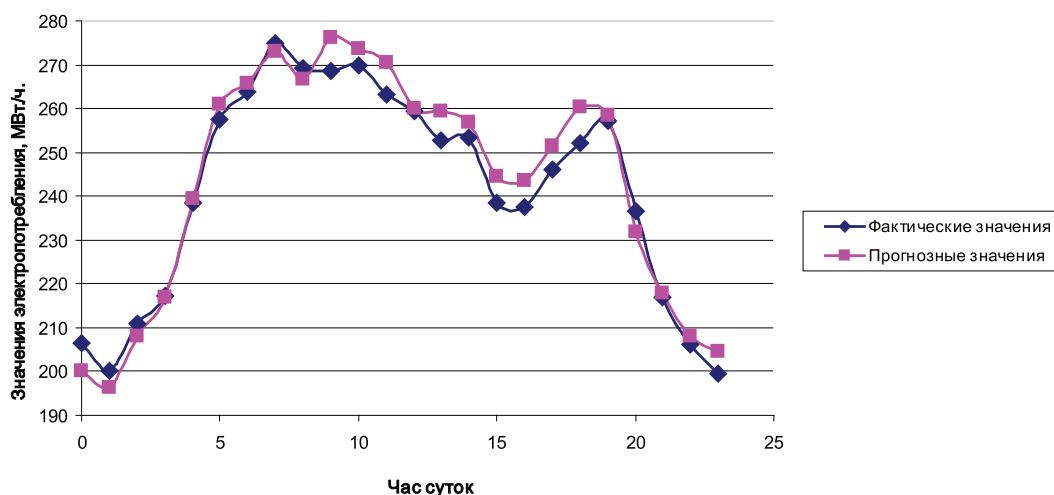


Рис. 2. Графики фактических и прогнозных значений электропотребления в Хакасии на сутки вперёд, метод простой скользящей средней, кластер «Будние дни»

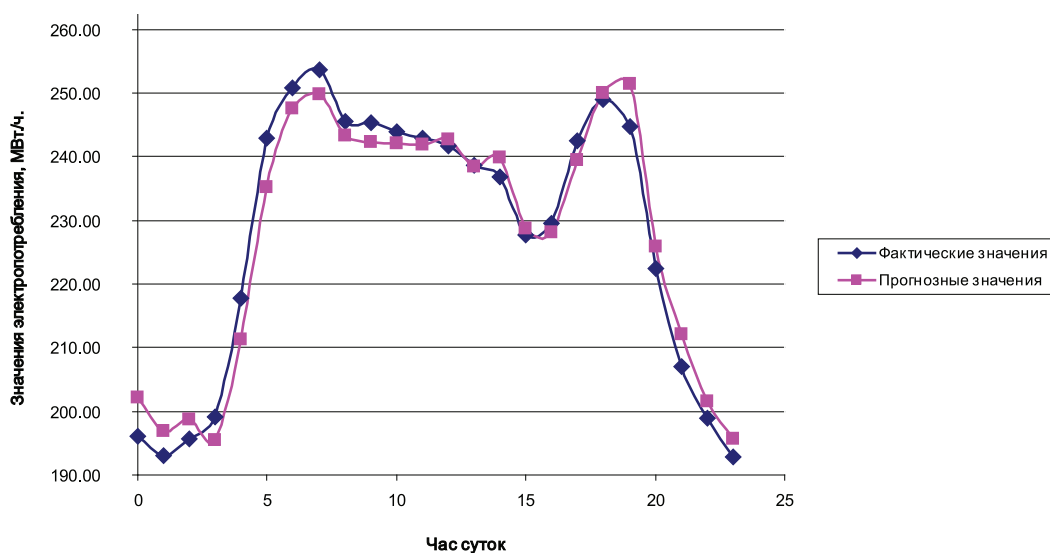


Рис. 3. Графики фактических и прогнозных значений электропотребления в Хакасии на сутки вперёд, метод простой скользящей средней, кластер «Праздничные/Выходные дни»

Таблица 4

Сравнение величин абсолютной ошибки прогноза, метод взвешенной скользящей средней, до и после проведения кластеризации

Час	Значение абсолютной ошибки прогноза, %, без кластеризации	Значение абсолютной ошибки прогноза, %, кластер «Будние дни»	Значение абсолютной ошибки прогноза, %, кластер «Праздничные/выходные дни»
8	0,09	0,96	0,94
9	1,55	2,89	1,25
12	1,36	0,25	0,34
13	0,89	2,65	0,1
17	4,1	2,17	1,26
18	0,95	3,23	0,44

Анализ табл. 4 позволил выявить частичное снижение величины прогнозной ошибки на указанных выше проблемных промежутках. Следует отметить, что до учёта кластерного подхода средняя абсолютная ошибка прогноза равнялась 1,5%. При выполнении кластеризации значений электропотребления согласно признаку «Будний день» ошибка составила 2%. При этом кластеризация с учётом признака «Праздничный/выходной день» позволила уменьшить значение прогнозной ошибки до 0,7%. Однако проблема неудовлетворительного качества прогнозных решений по-прежнему остаётся на часах 9-й, 13-й и промежутке с 17-го по 18-й часы.

Заключение

Выполнен прогноз значений электропотребления на сутки вперёд с использованием методов прогнозирования, таких как линейная регрессия, скользящие средние и нейросети. Выявлено, что ретроспективные данные об электропотреблении не подчиняются нормальному закону распределения. По этой причине качество полученных прогнозных решений является неудовлетворительным.

С целью уменьшения ошибки прогноза выполнена процедура кластеризации с учетом фактора, влияющего на процесс потребления электроэнергии, а именно «Будний или праздничный/выходной день». Значения величин ошибок, полученные при прогнозировании с применением кластеризации, указывают на эффективность проведённой корректировки. Наименьшая величина средней абсолютной ошибки прогноза достигнута с помощью метода простой скользящей средней, которая с применением кластерного подхода равняется 1,6% («Будние дни») и 1,5% («Праздничные/выходные дни»). При этом до использования кластеризации средняя ошибка прогноза принимала значение, равное 2,1%.

Проверка с помощью критерия Колмогорова – Смирнова позволила сделать вывод о том, что проведённая процедура кластеризации по будним и праздничным/выходным

дням позволяет, в частности, избавиться от выбросов параметров в данных и, как следствие, повысить качество получаемых прогнозных решений.

Детальный анализ значений ошибки прогноза на часах, являющихся участками перелома, позволил зафиксировать частичное снижение средней ошибки прогнозных решений, но проблема низкого качества при этом осталась на 9-м, 13-м часах и на интервале с 17-го по 18-й часы, что говорит о необходимости использования других методов и подходов в прогнозировании.

Список литературы

1. Кластеризация [Электронный ресурс] // BaseGroupLabs: технологии анализа данных. – URL: <https://basegroup.ru/deductor/function/algorithm/clustering> (дата обращения: 11.03.2016).
2. Кластерный анализ. Научный подход при изучении сложных явлений [Электронный ресурс] // ФБ. – URL: <http://arcticaoy.ru/fb.ru/article/22911/klasternyy-analiz-nauchniy-podhod-pri-izuchenii-slojnyih-yavleniy.html> (дата обращения: 13.02.2016).
3. Макоклоев Б.И. Анализ и планирование электропотребления. – М.: Энергоатомиздат, 2008. – 296 с.
4. Методы принятия решений в электроэнергетике [Электронный ресурс] // ИД «Библио-Глобус». – URL: <http://bgscience.ru/lib/7261/> (дата обращения: 21.03.2016).
5. Сглаживание скользящих средних – достоинства и недостатки [Электронный ресурс] // ВебМастерМаксим.ру. – URL: <http://webmastermaksim.ru/foreks/sglazhivanie-skolzyashhix-srednix-dostoinstva-nedostatki.html> (дата обращения: 21.11.2015).
6. Хайкин С. Нейронные сети. – М.: Вильямс, 2008. – 1104 с.
7. Чувилкин А.В., Гордеев А.С. Влияние метеорологических и производственных факторов на потребление электроэнергии предприятий АПК [Электронный ресурс] // Профессиональный образовательный портал по Тамбовской области. – URL: <http://www.edu68.ru/nauka/290-vliyanie-meteorologicheskix-i-proizvodstvennyx-faktorov-na-potreblenie-.html> (дата обращения: 17.03.2016).
8. Чучуева И. Два способа повысить точность прогноза [Электронный ресурс] // Математическое бюро. – URL: <http://www.mbureau.ru/blog/dva-sposoba-povysit-tochnost-prognoza> (дата обращения: 19.03.2016).
9. Чучуева И.А. Модель прогнозирования временных рядов по выборке максимального правдоподобия // Математическое бюро. – URL: <http://www.mbureau.ru/articles/dissertaciya-model-prognozirovaniya-vremennyh-ryadov-glava-1> (дата обращения: 11.03.2016).
10. Шилов А.В. Анализ методов в системе выработки и принятия прогнозных решений // С.С. Чернов. – Новосибирск: Изд-во ЦРНС, 2014. – С. 27–34.