

Н. Ю. Петухова

АНАЛИЗ ИНФОРМАТИВНОСТИ ДАННЫХ ПРИ ПОСТРОЕНИИ ПРОГНОЗНОЙ МОДЕЛИ ЦЕНЫ НА ЭЛЕКТРОЭНЕРГИЮ

Аннотация.

Актуальность и цели. Рассматривается задача построения эффективного метода прогноза показателей равновесной цены на электроэнергию на рынке на сутки вперед. Цель работы – предварительный анализ прогнозных данных для выявления наиболее важных прогнозных факторов.

Материалы и методы. Основой для построения прогноза являются модели с применением нейронных сетей и бустинга над композицией алгоритмов. Проведен математический анализ входных данных на основе оценки их информативности и метода выделения главных компонент в корреляционных зависимостях.

Результаты. Для выбора оптимального набора прогнозных факторов был проведен корреляционный и кластерный анализ. Уменьшение размерности данных проведено методом главных компонент. Информативность признаков определялась выборочной оценкой их взаимной энтропии.

Выводы. Рассмотренные методы анализа данных позволяют ранжировать признаки по их информативности, проводить кластерный анализ, своевременно добавлять новые факторы прогноза. Все это приводит к наиболее оптимальной прогнозной модели. Описано применение информационных критериев для создания дополнительных параметров.

Ключевые слова: рынок электроэнергии, нейронные сети, главные компоненты, информативность прогнозных данных, информационные критерии, константа Липшица.

N. Yu. Petukhova

AN ANALYSIS OF INFORMATIONAL CONTENT OF DATA WHEN CREATING A PREDICTIVE MODEL OF ELECTRICITY PRICE

Abstract.

Background. The problem of creation of an effective method of forecasting the indicators of the equilibrium price of the electric power in the market for the day ahead is considered. The work purpose is to give a preliminary analysis of predictive data to identify the most important predictive factors.

Materials and methods. The models with application of neural networks and gradient boosting over a composition of algorithms are the basis of forecast creation. A mathematical analysis of the input data on the basis of their informational content assessment and the method of allocation of main components in correlation dependences are carried out.

Results. To select the optimal set of predictive factors, the correlation and cluster analysis was carried out. The data dimensionality reduction was carried out by the method of principle components. The informational content of indicators was defined by selective assessment of their mutual entropy.

Conclusions. The considered methods of analysis of data allow to range the indicators by their informational content, to carry out the cluster analysis, in due time to add new factors to the forecast. All of these lead to the most optimal predictive model.

Keywords: energy market prediction, artificial neural networks, principle components method, informative forecasting, information criterion, Lipschitz constant.

Введение

Рассматривается задача прогноза равновесной цены на электроэнергию на основе применения современных методов анализа и прогнозирования. Для участников оптового рынка электроэнергии и мощности (ОРЭМ) наиболее актуальны прогнозы равновесной цены на электроэнергию на рынке на сутки вперед (РСВ). Точное прогнозирование будущих ценовых показателей позволяет потребителям рационально закупать необходимые объемы электроэнергии и снижать убытки при выходе на балансирующий рынок. Для решения задач прогнозирования равновесной цены РСВ обычно используются линейные регрессионные модели, основанные на анализе временных рядов. Основным недостатком линейных моделей является то, что они описывают именно линейную зависимость между прогнозными данными и результатом. Другим существенным недостатком линейных моделей является чувствительность к шумам в данных. Также успешность применения моделей линейной регрессии сильно зависит от выполнения предпосылок Гаусса – Маркова о нормальном законе распределения исследуемых данных, которые на практике обычно не выполняются. Для прогнозирования изменчивого рынка электроэнергии с множеством нелинейных закономерностей актуальны алгоритмы машинного обучения. Они дают хорошие показатели точности даже на сильно «зашумленных» данных, позволяют использовать большое количество информации, обнаруживать и анализировать сложные нелинейные закономерности в данных, выделять те, что оказывают наибольшее влияние на прогноз. В данной работе представлена математическая методика анализа качества прогнозных моделей и информативности прогнозных данных.

1. Прогнозная модель на базе машинного обучения

1.1. Математическая постановка задачи

Дана обучающая выборка $\{x_i, i = 1, \dots, n\}$ из множества X , на которой известны значения $y(x_i) = y_i, y_i \in Y$. Задача обучения: по выборке $\{x_i, y_i\}$ восстановить зависимость $y(x)$, т.е. построить решающую функцию $a: X \rightarrow Y$.

Задача включает два этапа:

1) по выборке $\{x_i\}$ построить алгоритм $a \in A, A$ – множество базовых алгоритмов;

2) этап применения a : вводится неотрицательная функция потерь $L(a, x) \geq 0$, характеризующая ошибку алгоритма a , и функционал качества

$$Q(a, \{x_i\}) = \sum_{i=1}^n L(a, x_i). \quad (1)$$

Тогда классический метод должен минимизировать риск – найти $a \in A$, при котором $Q(a, \{x_i\}) \rightarrow \min_{a \in A}$. В качестве функции потерь обычно используют:

$$L(a, x) = (a(x) - y(x))^2. \quad (2)$$

1.2. Прогнозные данные и применяемые методы машинного обучения

Исторические данные представляют собой упорядоченный набор основных параметров торгов, собранных за последние 2–3 года, в почасовой привязке. Задача прогнозирования энергопотребления: кратко- и среднесрочные прогнозы равновесной цены на электроэнергию (целевой переменной) на рынке на сутки вперед. Требуется построить функцию, которая будет наиболее точно аппроксимировать значения целевой переменной. Для прогноза эффективно комбинированное использование нейронных сетей и различных вариаций решающих деревьев.

Градиентный бустинг. Основная идея бустинга: построение правильной композиции относительно простых алгоритмов для получения сильной модели. Каждая последующая модель строится таким образом, чтобы придавать больший вес и предпочтение ранее некорректно предсказанным наблюдениям [1, 2]. Задана модель прогноза: параметрическое множество функций $g(x, \alpha)$, α – набор параметров. $L(y, g(x, \alpha))$ – функция потерь. Для минимизации суммарной функции потерь

$$Q(\tilde{\alpha}) = \sum_{i=1}^N L(y_i, g(x_i, \alpha))$$

применяют метод градиентного спуска: значения $\tilde{\alpha}$ меняются вдоль линии антиградиента $L(g, \tilde{\alpha})$. Обновленное значение параметров: $\alpha_{i+1} = \alpha_i - \nabla L(\alpha)$. Для улучшения качества прогноза также используется функциональный градиентный спуск – модель Фридмана: бустинг над решающими деревьями *GBM*. Здесь при построении следующей модели используется дерево решений так, чтобы лучшим образом приближать общий градиент целевой функции. В качестве базовых алгоритмов используются: различные методы линейной регрессии (*ARMA*, *GARCH*, *SARIMA*, *LARS* и др.); нелинейные методы в области пиковых значений, различные критерии оценки значимости коэффициентов модели или вновь добавленных факторов, методы выявления автокорреляции или гетероскедастичности [3–5]. Именно решающие деревья позволяют получать интервальные оценки, выполнять проверки гипотез.

Нейронные сети. В модели использованы многослойные нейронные сети [1, 5, 6]. Эти сети позволяют использовать нелинейные прогнозные алгоритмы и имеют память (веса, присвоенные прогнозным данным). Нейронная сеть со скрытыми слоями принимает на входе вектор данных для прогноза; j -й скрытый нейрон дает на выходе величину $h_j = \Phi(\sum_i w_{ij} x_i)$, где Φ – функция активации. Скрытый слой связан с выходным слоем, в матричной

форме уравнения связи имеют следующий вид: $\mathbf{h} = \Phi(\mathbf{W}\mathbf{X})$, $\mathbf{y} = \mathbf{V}\mathbf{h}$, где \mathbf{x} – входной вектор; \mathbf{W} – матрица весов, соединяющая входной слой и скрытый слой; \mathbf{V} – матрица весов, соединяющая скрытый и выходной слой.

1.3. Прогнозные данные

Данные российского рынка электроэнергии и мощности взяты за период с 1 февраля 2015 г. по 1 ноября 2017 г. по первой ценовой зоне. Они имеют 24600 наблюдений. Собранные исторические данные с официальных сайтов «АТСЭнерго» и Системного Оператора ОРЭМ [7] позволяют выдавать прогнозы равновесной цены на РСВ, усредненные по первой ценовой зоне. Основные данные представлены с периодом 1 ч.

Основные данные для прогноза:

- дата (год-месяц-день час);
- количество потребленной энергии (фактическое, за указанный час);
- прогноз потребления энергии (предшествующий день, за указанный час);
- цена на газ;
- фактическая температура за указанный час и прогноз температуры;
- фактическая цена;
- планируемые и фактические объемы генерации и потребления, МВт · ч;
- цена в заявке со стороны генерации и потребления, руб./МВт · ч;
- генерирующее оборудование: генерация ГЭС (МВт) и активная мощность ТЭС (МВт);
- день недели, ч;
- равновесная цена на РСВ, руб./МВт · ч.

1.4. Предварительная обработка и анализ входных данных

Из рис. 1–3 видно, что выбросы (пиковые значения) могут превосходить средние значения в несколько раз. Для определения пиковых значений были использованы данные, превышающие двойное значение среднеквадратичного отклонения. Имеется большой разброс значений цены на электричество в зависимости от времени суток и от того, в какой день (рабочий или выходной) проводились измерения. Энергопотребление также имеет сезонный характер: в жаркие месяцы пик потребления приходится на середину дня, в зимние месяцы – на начало и конец рабочего дня.

2. Основные методы анализа качества модели и входных данных

Самые простые показатели качества прогнозной модели (прежде всего регрессионной) – это информационные критерии: мера качества модели, учитывающая ее простоту и точность прогноза. Лучшей по качеству будет та модель, у которой меньше значение информационного критерия. Были использованы следующие основные критерии [3, 5]:

1. Критерий Акаике:

$$AIC = \ln\left(\sum_{i=1}^N \varepsilon_i^2 / N\right) + 2k / N,$$

N – число наблюдений; k – число параметров модели; $\varepsilon_i = y(x_i) - y_i$.

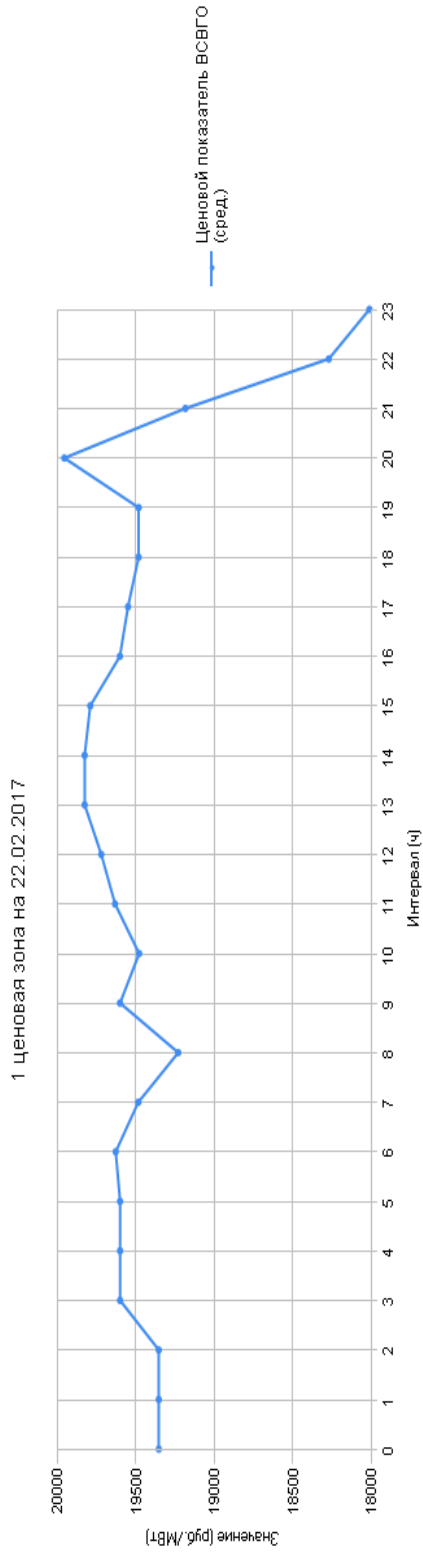


Рис. 1. Почасовое распределение цены на электроэнергию 2 февраля 2017 г.

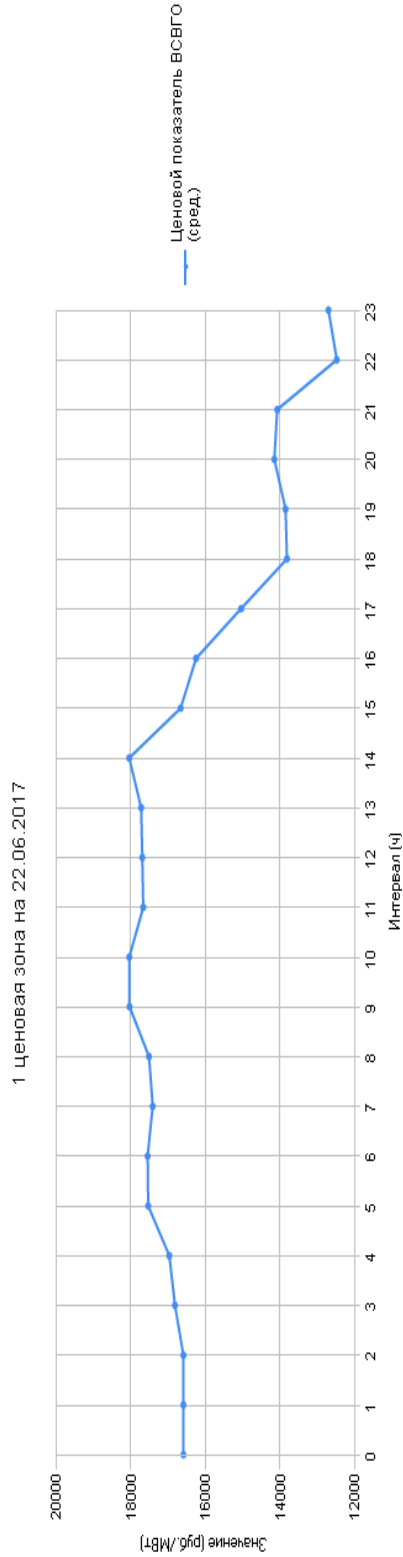


Рис. 2. Почасовое распределение цены на 30 декабря 2017 г.

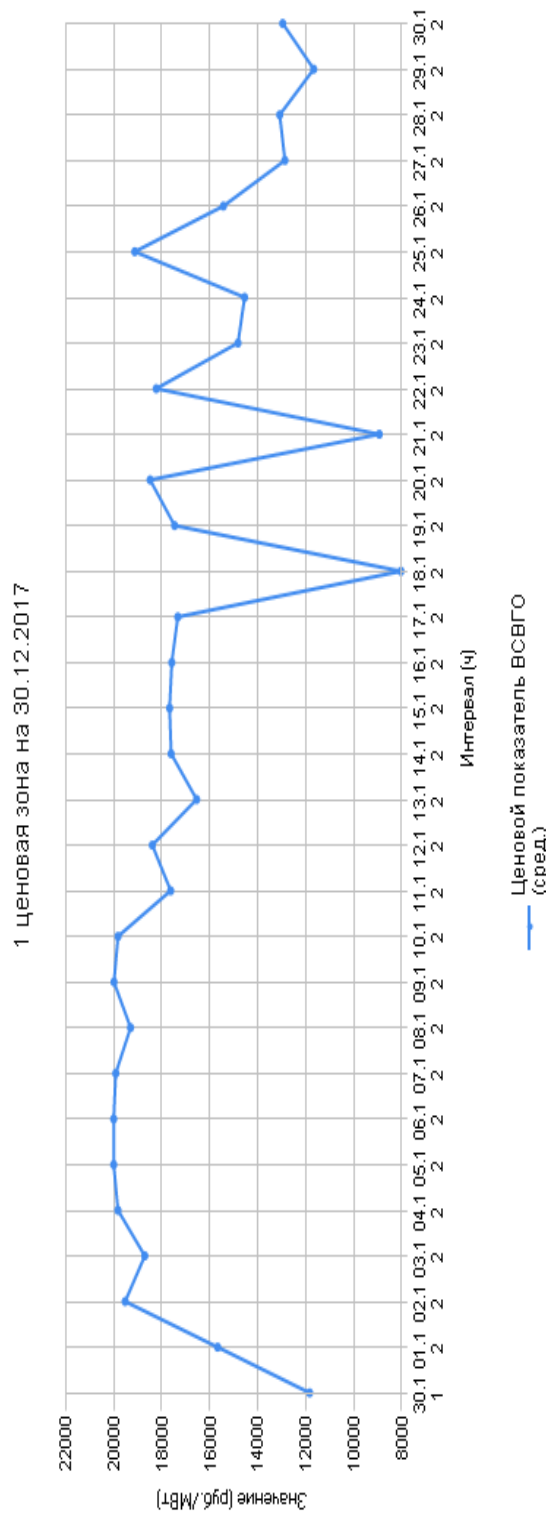


Рис. 3. Средняя цена за декабрь 2017 г.

2. Критерий Шварца:

$$SC = \ln\left(\sum_{i=1}^N \varepsilon_i^2 / N\right) + k \cdot \ln(N) / N. \quad (3)$$

3. Коэффициент детерминации R^2 : показывает долю дисперсии зависимой переменной, объясненную моделью

$$R^2 = Q_R / (Q_R + Q_\varepsilon), Q_R = \sum_{i=1}^N (\bar{y} - y_i)^2, Q_\varepsilon = \sum_{i=1}^N (y(x_i) - y_i)^2, \bar{y} - \quad (4)$$

среднее значение; $y(x_i)$ – прогнозируемое значение по модели; y_i – точное значение в обучающей выборке.

Выбор оптимальных информационных признаков. Существует несколько подходов к выбору оптимальных признаков. По распространенной классификации можно выделить три основных категории:

1. Статистические методы.
2. Методы на основе вероятностной оценки информации о наборе данных.
3. Алгоритмические методы оценки информативности, скорости работы программы и используемых прогнозных данных.

2.1. Статистические методы. Выделение главных компонент

Важное свойство нейронных сетей (НС) – их способность к обучению на тестовых данных. Одна из его целей – выявление во входных данных наиболее существенных признаков. Анализ главных компонент – это стандартный прием, используемый для уменьшения размерности данных. В процессе обучения на основе самоорганизации в каждом слое НС выполняется анализ признаков по следующим критериям (постулатам обучения Хебба) [1, 4]:

– сильные сигналы (верные входные данные) должны получать усиление своих весов, и это усиление должно компенсироваться ослаблением слабых (неподходящих входных данных);

– изменение весов сильного сигнала должно способствовать успешной совместной работе и других сигналов, расположенных в данном слое, т.е. усиливать также и их веса.

Метод главных компонент преобразует исходные данные в пространство признаков с сокращением размерности – выделением меньшего числа наиболее эффективных признаков, содержащих основную часть информации из входных данных. Множество данных – вектор \mathbf{X} размерности m должен быть передан с помощью l , $l < m$, чисел, при этом надо построить линейное преобразование A , при котором «обрезание» вектора $A\mathbf{X}$ будет давать минимальную среднеквадратичную ошибку – сумму дисперсий элементов, удаленных из вектора \mathbf{X} . Если \mathbf{X} – случайный вектор с $M(\mathbf{X}) = 0$, а \mathbf{q} – его проекция, то величину связи \mathbf{X} и \mathbf{q} характеризует их дисперсия $S = \mathbf{X}^T \mathbf{q}$, при этом $\sigma^2 = M(S^2) = \mathbf{q}^T \mathbf{R} \mathbf{q}$, где \mathbf{R} – матрица корреляции вектора \mathbf{X} [5, 8].

Задача условной оптимизации:
$$\begin{cases} \sigma^2(\mathbf{q}) \rightarrow \text{extr} \\ \|\mathbf{q}\| = 1 \end{cases}$$
 имеет решение, если \mathbf{q}

является собственным вектором матрицы \mathbf{R} : $\mathbf{R} \mathbf{q} = \lambda \mathbf{q}$, λ – собственные числа

корреляционной матрицы. Известно, что корреляционная матрица имеет вещественные неотрицательные собственные числа $\lambda_1, \lambda_2, \dots, \lambda_m$, $i = 1, \dots, m$. Собственные векторы \mathbf{q} определяют основные направления, вдоль которых $\sigma^2(\mathbf{q})$ принимает наибольшее значение $\sigma^2(\mathbf{q}_i) = \lambda_i$. Количество признаков, необходимых для эффективного представления данных, можно сократить, устранив в них линейно зависимые комбинации, имеющие малые дисперсии. Пусть $\lambda_1, \lambda_2, \dots, \lambda_l$ – наибольшие l собственных значений матрицы \mathbf{R} . Тогда вектор \mathbf{X} можно аппроксимировать первыми l главными компонентами

$$\bar{\mathbf{X}} = \sum_{i=1}^l a_i \mathbf{q}_i, l < m.$$

Вектор ошибки такой аппроксимации равен

$$|\mathbf{e}| = |\mathbf{X} - \bar{\mathbf{X}}| = \sum_{i=l+1}^m \sigma_i^2 = \sum_{i=l+1}^m \lambda_i^2.$$

Чем меньше будут собственные числа $\lambda_{l+1}, \dots, \lambda_m$, тем более эффективным будет сокращение размерности в представлении исходных данных. Выделение главных компонент может быть полезно и при проведении кластеризации входных данных. Новые данные, поступающие в нейронную сеть на этапе ее обучения, можно проверить на степень корреляции с главными компонентами. При сильной зависимости они не несут в себе никакой принципиально новой информации; если корреляция мала и новые данные можно считать независимыми от основной выборки, их надо оставить; веса, которые они получат, покажут их влияние на качество прогнозной модели.

2.2. Применение элементов теории информации Шеннона

Важным элементом теории информации в работе нейронных сетей является принцип максимума взаимной информации Линскера [4, 5]: синаптические связи многослойной нейронной сети организуются таким образом, чтобы максимизировать объем информации, которая сохраняется при преобразовании сигнала на каждой из стадий его обработки. Этот принцип позволяет не только делать прогноз, но и преобразовать исходные данные для уменьшения их избыточности. Поступление на вход нейронной сети сигнала x_k или наступление случайного события $X = x_k$ должно приводить к увеличению объема информации, и, наоборот, поступление на вход сигнала x_k , не влияющего на прогноз, должно восприниматься как достоверное событие, происходящее с вероятностью $p_k = 1$ и не содержащее никакой новой информации. Объем информации, связанный с наступлением случайного события $X = x_k$ с вероятностью p_k , определяется следующим образом [1, 8]:

$$I(x_k) = \log(1/p_k) = -\log(p_k);$$

$I(p_k)$ является неотрицательной возрастающей функцией, и, чем меньше вероятность события, тем больше информации несет в себе его наступление.

Если рассматривать x_k как дискретную случайную величину, распределенную на отрезке $[-K; K]$, то энтропия X – это среднее значение $I(x_k)$:

$$H(X) = M(I(X)) = \sum_{k=-K}^K p_k I(x_k) = - \sum_{k=-K}^K p_k \log(p_k), \quad (5)$$

она является мерой среднего объема информации, которую несет в себе входной сигнал X . Ответ на вопрос выбора модели прогноза дает применение принципа максимальной энтропии [1, 2]: если выводы основаны на неполной информации, то модель должна строиться на тех данных, у которых допустимые распределения вероятностей максимизируют энтропию. Таким образом, задача оценки входных данных является задачей условной оптимизации:

$$\begin{cases} H(X) \rightarrow \max, \\ X \in F, \end{cases}$$

F – множество допустимых в модели функций распределения входных данных. Решить эту задачу можно методом множителей Лагранжа при известной ковариационной матрице \mathbf{R} .

В процессе обучения нейронной сети потребуется также использовать элементы теории информации, чтобы определить информативность входного сигнала X на основе наблюдаемого значения Y . Для этого потребуется условная энтропия множества наблюдений X [5]:

$$H(X|Y) = H(X, Y) - H(Y),$$

где

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log(p(x, y)),$$

$p(x, y)$ – плотность совместного распределения случайных величин X и Y . Условная энтропия представляет собой неопределенность системы после получения выходных данных, поэтому величина $I(X, Y) = H(X) - H(X|Y)$, называемая взаимной информацией, показывает ту часть неопределенности, которая была разрешена после наблюдения этих данных:

$$\begin{aligned} I(X, Y) &= H(X) - H(X|Y) = \\ &= \sum_{x \in X} \sum_{y \in Y} p(x, y) \log(p(x, y) / (p(x)p(y))). \end{aligned} \quad (6)$$

Функция $I(X, Y)$ является неотрицательной и симметричной: $I(X, Y) = I(Y, X)$, при этом величина $I(Y, X)$ – мера неопределенности выхода Y , разрешенной при получении входных данных X , может рассматриваться как критерий важности признака X в прогнозной модели (известно, что $I(Y, X) = 0$, если вход и выход модели являются независимыми случайными величинами). При построении прогнозной модели и входные данные, и свободные параметры (веса) системы должны быть настроены так, чтобы оптимизировать взаимную информацию (принцип Informax Линскера [5, 6]).

2.3. Аналитические методы

Здесь входные данные анализируются непосредственно в процессе обучения нейронной сети. Для этого используют обучающую выборку $A = \{(x_i, y_i), i = 1, \dots, N\}$, N – число примеров. При завершении процесса обучения, когда веса решающей функции $F(x, \omega)$ стабилизируются, ее константа Липшица L должна быть мала. Поэтому при анализе тестового набора используют ее выборочную оценку:

$$\bar{L} = \max_{i \neq j, xi \neq xj} \frac{\|y_i - y_j\|}{\|x_i - x_j\|}. \quad (7)$$

Уменьшение \bar{L} ускоряет обучение нейронной сети. При этом оценка константы Липшица потребует порядка N^2 действий [3, 6].

Другой подход к исследованию процесса обучения – это проверка процесса минимизации функционала качества модели $Q(x, \omega)$ (1) методом градиентного спуска. Стабильность его работы связана с гессианом целевой функции по параметрам: $G = \{\partial^2 Q / \partial \omega_i \partial \omega_j\}$. Собственные числа матрицы

Гессе λ и ее число обусловленности $\mu = \frac{\max |\lambda_i|}{\min |\lambda_i|}$ оказывают влияние и на

скорость обучения, и на реализацию численного метода поиска экстремума. Известно, что скорость сходимости этих численных алгоритмов прямо пропорциональна значению μ [9], тогда уменьшение числа обусловленности также ведет к оптимизации модели. Вместе с определением константы Липшица полезно рассмотреть и те множества примеров $\{x_i, y_i\}$ из обучающей выборки, которые дают ее максимальное значение. Значения $\{x_i\}$ из этих примеров стоит добавить к главным компонентам множества X . Также, возможно, следует произвести увеличение масштаба в этих данных: это приведет к уменьшению константы Липшица и к тому, что строящаяся прогнозная модель $F(x, \omega)$ не будет иметь вдоль этих переменных резких скачков.

Стандартные методы определения собственных векторов корреляционной матрицы при большой размерности входных данных имеют большую вычислительную сложность, поэтому на практике также более эффективно использовать аналитические адаптивные методы. Эти методы непосредственно преобразуют вектор входных данных, выделяя в нем главные направления по их корреляции. Так, для определения первой главной компоненты \bar{y}_1 достаточно простой нейронной сети из одного нейрона, где

$$\bar{y}_1 = \bar{w}_1^T \bar{x} = \sum_{j=1}^N w_{1j} x_j,$$

$$\bar{w}_1(k+1) = \bar{w}_1(k) + \eta(k) \bar{y}_1^T(k) (\bar{x}(k) - \bar{w}_1(k) \bar{y}_1(k)), \quad (8)$$

$\eta(k) < 1$ – темп обучения [8].

В процессе обучения одни и те же обучающие выборки предъявляются сети многократно, пока процесс не стабилизируется. Определение следующих l главных компонент потребует нейронной сети с l выходными нейронами. Построение i -го собственного вектора идет по аналогичной формуле

$$\bar{y}_i(k) = \sum_{j=1}^N w_{ij}(k)x_j(k),$$

где изменение весов происходит с учетом входного вектора X и весов, полученных для предыдущих главных компонент.

3. Практическое применение методов анализа входных данных

В ходе построения прогнозной модели был проведен корреляционный анализ входных данных. Наивысшая линейная корреляция выявлена у целевой переменной с ценой в реальном времени, спросом на рынке и ценой на газ; также имеется сильная линейная зависимость со значением целевой переменной за предыдущий час и за этот час предшествующего дня. Вычисление корреляции Спирмена показало наличие нелинейной зависимости целевой переменной от прогнозных данных. Это подтверждает необходимость использования в прогнозе алгоритмов, реализующих нелинейные зависимости. Этого же требует большой разброс значений цены на электричество в зависимости от времени суток. Была проведена проверка гипотез на независимость входных данных и на характер их распределения. Критерии независимости типа χ^2 и др. [5, 6] подтверждают данные корреляционного анализа и необходимость использования регрессионных моделей с гетероскедастичностью типа GARCH. Гипотеза о нормальном законе распределения основных входных данных опровергается. Это подтверждает необходимость предобработки данных, их кластеризации, применения метода выделения главных компонент.

Качество отдельных составляющих прогнозной модели и используемых в них параметров проверялось стандартными информационными критериями (3) и оценкой константы Липшица (7). Важность каждого из прогнозных признаков оценивалась по энтропии этих признаков, вычисленной на основе выборочных функций распределения (4). Ниже приводятся признаки, отсортированные в порядке значимости согласно критерию ожидаемого уменьшения энтропии:

- lag1: 13805.4;
- текущая цена на РСВ: 1872.7;
- произведено тепловыми станциями: 148.2;
- час суток: 103.6;
- количество потребленной энергии за час: 36.7;
- цена на газ: 20.1;
- фактическая температура: 7.9;
- lag12: 6.3.

График автокорреляции на рис. 4 показывает, что целевая переменная имеет ярко выраженную зависимость на саму себя, в особенности при лаге равном 1; лаги близкие к 24 тоже существенны. Зависимость от предыдущих значений в остатках модели слишком велика. Поэтому в прогноз взяты дополнительные признаки: лаги целевой переменной с шагом 1 и 24.

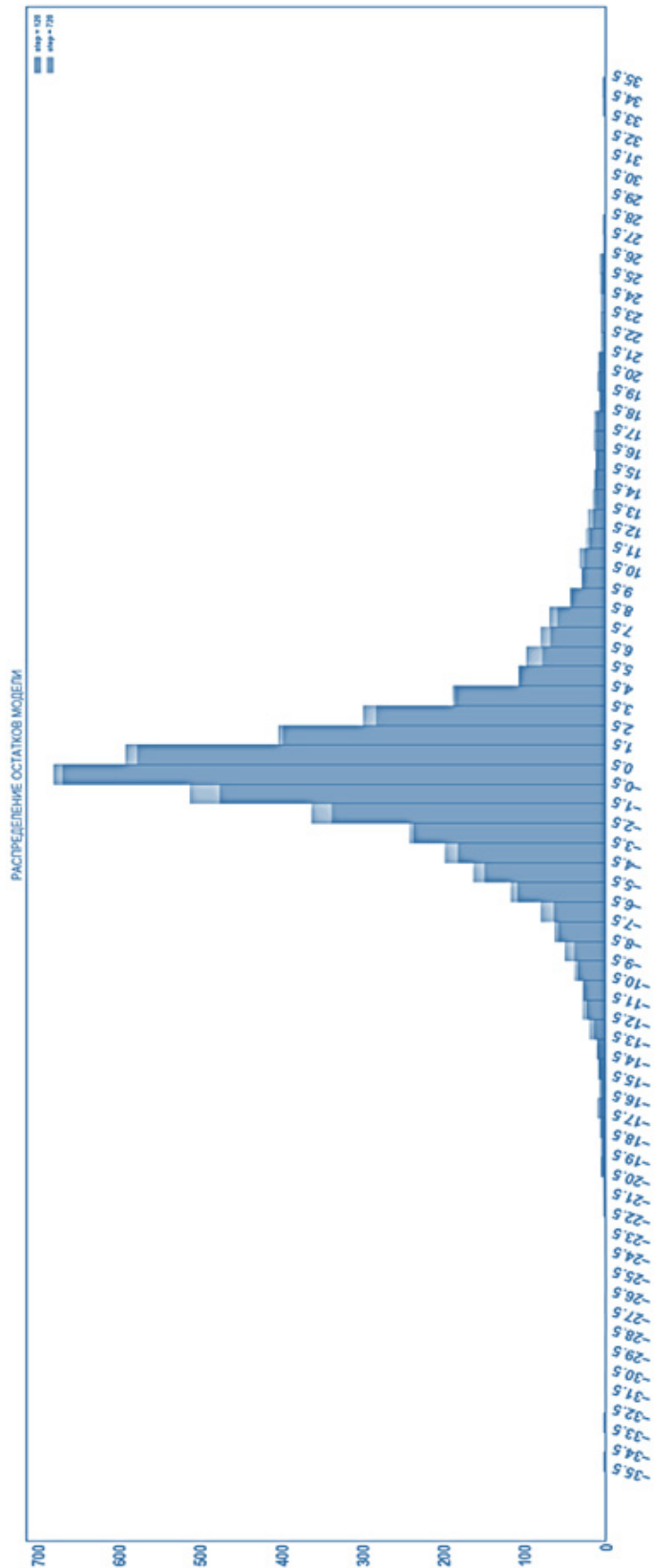


Рис. 4. График автокорреляции остатков модели

В качестве дополнительных признаков были сгенерированы:

- 1) фактическая температура, фактическая генерация на тепловых станциях, объем потребления, фактические значения целевой переменной;
- 2) разница между прогнозными значениями и фактическими;
- 3) лаги целевой переменной с различным шагом (1, 2, 12, 24);
- 4) лаги различных признаков (фактическое, прогнозное потребление);
- 5) вероятность принадлежности к кластеру;
- 6) минимальное и максимальное значения целевой переменной, взятые за предыдущий день и их разница;
- 7) переменная, отмечающая наблюдения, попадающие на выходные и праздничные дни и период отпусков.

Заключение

Вычислительный эксперимент показал, что рост потенциальной прибыли при увеличении точности прогноза на 1 % колеблется от 1,7 до 4,7 % прибыли для различных ТЭЦ и зависит от сезона года и технико-экономических особенностей работы станций [7].

Варианты дальнейшего углубленного анализа прогнозных данных. В ходе исследования будет полезно:

- провести дальнейший анализ корреляционных зависимостей во входных данных, исходя из их нелинейного характера: выделить главные компоненты, определить содержащуюся в них информативность;
- выделить базовые алгоритмы прогнозной модели и главные прогнозные составляющие, оценить влияние новых данных на эти величины; получить вероятностные оценки информативности новых признаков по формулам взаимной информации;
- исследовать влияние гессиана целевой функции на скорость сходимости, провести кластеризацию прогнозных данных на основе максимизации взаимной информации.

Библиографический список

1. **Хайкин, С.** Нейронные сети / С. Хайкин. – Москва : Вильямс, 2006. – Ч. 1, 2. – 654 с.
2. **Friedman, H.** Greedy function approximation: a gradient boosting machine / H. Friedman // *Annals of Statistics & Data Analysis*. – 2002. – Vol. 38, № 4. – P. 1189–1232.
3. **Бидюк, П. И.** Структурный анализ методик построения регрессионных моделей по временным рядам наблюдений / П. И. Бидюк, Т. Ф. Зворыгина // *Управляющие системы и машины*. – 2003. – № 2. – С. 93–99.
4. **Загоруйко, Н. Г.** Прикладные методы анализа данных и знаний / Н. Г. Загоруйко. – Новосибирск : ИМ СО РАН, 1999. – 271 с.
5. **Пупков, К. А.** Нестационарные системы автоматического управления: анализ, синтез и оптимизация / К. А. Пупков, Н. Д. Егупов. – Москва : Изд-во МГТУ им. Баумана, 2007. – 340 с.
6. **Комарцева, Л. Г.** Нейрокомпьютеры / Л. Г. Комарцева, А. В. Максимов. – Москва : Изд-во МГТУ им. Баумана, 2004. – 410 с.
7. URL: <http://www.atsenergo.ru/>
8. **Осовский, С.** Нейронные сети для обработки информации / С. Осовский. – Москва : Финансы и статистика, 2002. – 485 с.

9. **Бахвалов, Н. С.** Численные методы / Н. С. Бахвалов, Н. П. Жидков, Г. М. Кобельков. – 7-е изд. – Москва : БИНОМ. Лаборатория знаний, 2011. – 630 с.

References

1. Khaykin S. *Neyronnye seti* [Neural networks]. Moscow: Vil'yams, 2006, part 1, 2, 654 p. [In Russian]
2. Friedman H. *Annals of Statistics & Data Analysis*. 2002, vol. 38, no. 4, pp. 1189–1232.
3. Bidyuk P. I., Zvorygina T. F. *Upravlyayushchie sistemy i mashiny* [Control systems and machines]. 2003, no. 2, pp. 93–99. [In Russian]
4. Zagoruyko N. G. *Prikladnye metody analiza dannykh i znaniy* [Applied methods of data and knowledge analysis]. Novosibirsk: IM SO RAN, 1999, 271 p. [In Russian]
5. Pupkov K. A., Egupov N. D. *Nestatsionarnye sistemy avtomaticheskogo upravleniya: analiz, sintez i optimizatsiya* [Nonstationary systems of automatic control: analysis, synthesis and optimization]. Moscow: Izd-vo MGTU im. Baumana, 2007, 340 p. [In Russian]
6. Komartseva L. G., Maksimov A. V. *Neyrokomp'yutery* [Neurocomputer]. Moscow: Izd-vo MGTU im. Baumana, 2004, 410 p. [In Russian]
7. Available at: <http://www.atsenergo.ru/>
8. Osovskiy S. *Neyronnye seti dlya obrabotki informatsii* [Neural networks for information processing]. Moscow: Finansy i statistika, 2002, 485 p. [In Russian]
9. Bakhvalov N. S., Zhidkov N. P., Kobel'kov G. M. *Chislennyye metody* [Numerical methods]. 7th ed. Moscow: BINOM. Laboratoriya znaniy, 2011, 630 p. [In Russian]

Петухова Наталья Юрьевна

кандидат физико-математических наук,
доцент, старший научный сотрудник,
лаборатория численного моделирования,
ООО «Базовые Технологии»
(Россия, г. Москва, 3-я улица
Ямского Поля, 2)

E-mail: natpet967@gmail.com

Petukhova Natal'ya Yur'evna

Candidate of physical and mathematical
sciences, associate professor, senior staff
scientist, the laboratory of numerical
modeling, «Bazovye tehnologii» LLC
(2 3-ya Yamskogo Polya street,
Moscow, Russia)

Образец цитирования:

Петухова, Н. Ю. Анализ информативности данных при построении прогнозной модели цены на электроэнергию / Н. Ю. Петухова // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2019. – № 3 (51). – С. 17–30. – DOI 10.21685/2072-3059-2019-3-2.