

УДК 519.584

Л. Н. Сандуляну, студ., *Московский физико-технический институт*

В. В. Стрижов, к.ф.-м.н., н.с., *Вычислительный центр РАН*

Выбор признаков в авторегрессионных задачах прогнозирования¹

Исследуется проблема выбора модели оптимальной сложности при авторегрессионном прогнозировании. Задача состоит в отыскании наименее обусловленного набора признаков, доставляющего при этом заданное значение функции ошибки. Для выбора этого набора используется модифицированный алгоритм последовательного добавления и удаления признаков. В работе предложен метод поиска оптимальной модели прогнозирования временных рядов. В вычислительном эксперименте приведено сравнение прогнозов рядов почасовых цен на электроэнергию.

Ключевые слова: отбор признаков, мультиколлинеарность, шаговая регрессия, метод Белсли, прогнозирование временных рядов.

1 Введение

Решается задача восстановления линейной регрессии при наличии большого числа мультиколлинеарных признаков. Термин «мультиколлинеарность» введен Р. Фишером при рассмотрении линейных зависимостей между признаками [1]. Проблема состоит в том, что количество признаков значительно превосходит число зависимых переменных, то есть мы имеем дело с переопределенной матрицей. Для решения этой задачи необходимо исключить наиболее малоинформативные признаки. Для отбора признаков предлагается использовать модифицированный метод шаговой регрессии.

Ранее для решения подобных задач использовались следующие методы: метод наименьших углов LARS [2], Лассо [3], ступенчатая регрессия [4], последовательное добавление признаков с ортогонализацией FOS [5, 6], шаговая регрессия [4, 7, 8] и другие [9].

¹ Работа выполнена при поддержке РФФИ, проект 10-07-00422.

В работе предложен модифицированный метод шаговой регрессии. Шаговыми методами называются методы, заключающиеся в последовательном удалении или добавлении признаков согласно определенному критерию. Метод включает два основных этапа: последовательное добавление признаков и последовательное удаление признаков. Добавление признаков производится с помощью FOS [5, 6]. Данный метод последовательно добавляет признаки которые максимально коррелируют с вектором регрессионных остатков. Удаление признаков в нашей работе выполняется с помощью метода Белсли [10]. Он позволяет выявить мультиколлинеарность признаков используя сингулярное разложение матрицы признаков. Для нахождения алгоритма, который доставляет одновременно точную и устойчивую, в смысле минимизации числа признаков, модель, предложен новый метод останова этапов добавления и удаления признаков, а так же останова всего алгоритма.

Основной причиной использования метода шаговой регрессии в качестве алгоритма выбора модели послужило то, что метод Белсли невозможно использовать при вырожденной авторегрессионной матрице, а метод ортогонального добавления признаков, в свою очередь, выбирает мультикоррелирующий набор признаков. Пошаговый алгоритм позволяет избежать этих недостатков и получить хорошо обусловленный набор признаков, позволяющий сделать достаточно точный прогноз.

Предложенный метод выбора модели проиллюстрирован задачей прогнозирования почасовых цен на электроэнергию на сутки вперед. Ранее эта задача решалась с помощью гребневой регрессии [11], метода наименьших углов, построения локальных регрессионных моделей [12, 13].

Для получения прогноза в работе предлагается использовать авторегрессионный алгоритм. В основе этого алгоритма лежит построение авторегрессионной матрицы, в которую построчно укладывается временной ряд, причем длина строки (ширина матрицы авторегрессии) равна периодике — 24-м

часам. Таким образом, каждый столбец содержит цену в некоторый час по всем суткам, и рассматривается как признак в задаче регрессии, а каждая строка является элементом выборки. Строятся 24 регрессионные модели — для прогнозирования цен на каждый час следующих суток. При использовании линейных моделей задача может быть решена методом наименьших квадратов.

В вычислительном эксперименте проведено сравнение предлагаемого алгоритма с базовым методом SSA [14].

Работа состоит из трех основных частей. Первая часть посвящена прогнозированию с помощью авторегрессионной матрицы. Во второй части описан выбор признаков при прогнозировании, здесь же можно найти описание метода Белсли. В последнем разделе приведены результаты вычислительного эксперимента, проведенного на основе данных почасовых цен на электроэнергию.

2 Задача прогнозирования с помощью авторегрессионной матрицы

Даны временной ряд $\mathbf{s}_1 = \{x_i\}_{i=1}^T$, будем называть его целевым рядом, и матрица признаков, столбцами которой являются временные ряды $\mathbf{s}_2, \mathbf{s}_3 \dots \mathbf{s}_p$. Необходимо спрогнозировать следующие τ значений ряда \mathbf{s}_1 . Предполагается, что

- отсчеты x_i сделаны через равные промежутки времени,
- ряд s имеет периодическую составляющую τ ,
- ряд s не имеет пропущенных значений,
- длина ряда s кратна периоду τ .

Для нахождения оптимальной модели, предлагается построить алгоритм прогноза, позволяющий решать задачи прогнозирования периодических рядов. С помощью этого алгоритма строится прогноз по выбранному набору признаков. В нашей работе предлагается использовать метод авторегрессии.

Сначала опишем как строится прогноз методом авторегрессии без учета вспомогательных рядов $\mathbf{s}_2, \mathbf{s}_3 \dots \mathbf{s}_p$, а затем обобщим метод на случай

многомерного ряда (метод многомерной авторегрессии).

Пусть длина временного ряда $T = m\tau$. Построим авторегрессионную матрицу \mathbf{X}^*

$$\mathbf{X}^* = \begin{pmatrix} x_1 & x_2 & \dots & x_{\tau-1} & x_{\tau} \\ \dots & \dots & \dots & \dots & \dots \\ x_{j\tau+1} & x_{j\tau+2} & \dots & x_{(j+1)\tau-1} & x_{(j+1)\tau} \\ \dots & \dots & \dots & \dots & \dots \\ x_{(m-2)\tau+1} & x_{(m-2)\tau+2} & \dots & x_{(m-1)\tau-1} & x_{(m-1)\tau} \\ x_{T-\tau+1} & x_{T-\tau+2} & \dots & x_{T-1} & x_T \end{pmatrix}.$$

Введем обозначения:

$$\mathbf{X}^* = \begin{pmatrix} \mathbf{X} & \mathbf{y} \\ \mathbf{x}_m & x_T \end{pmatrix}.$$

Необходимо построить линейную регрессию:

$$\mathbf{y} = \mathbf{X}\mathbf{w}, \quad (1)$$

где \mathbf{w} — вектор параметров. Тогда получим $x_T = \langle \mathbf{x}_m, \mathbf{w} \rangle$.

Требуется решить задачу минимизации евклидовой нормы вектора регрессионных остатков

$$\|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2 \rightarrow \min.$$

Вектор параметров \mathbf{w} отыскивается с помощью метода наименьших квадратов

$$\mathbf{w} = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{y}).$$

Однако зависимость $\mathbf{y} = \mathbf{f}(\mathbf{w}, \mathbf{X})$ может быть существенно нелинейной относительно свободных переменных, и для построения линейной модели удовлетворительного качества необходимо расширить множество признаков с помощью функциональных преобразований исходных признаков.

В случае многомерного ряда при построении авторегрессионной матрицы необходимо учитывать временные ряды $\mathbf{s}_2, \mathbf{s}_3 \dots \mathbf{s}_p$. Сначала строится авторегрессионная матрица \mathbf{X}_1 для ряда \mathbf{s}_1 . Для каждого следующего временного ряда \mathbf{s}_j , где $j = 1, \dots, p$, строится авторегрессионная матрица \mathbf{X}_j . Но для рядов $\mathbf{s}_2, \mathbf{s}_3 \dots \mathbf{s}_p$ не вводятся вектора \mathbf{y} в отличие от ряда \mathbf{s}_1 . То есть для

этих рядов авторегрессионная матрица будет содержать τ столбцов, а не $\tau - 1$, как для целевого ряда. Присоединив авторегрессионные матрицы всех временных рядов, получим матрицу

$$\mathbf{X} = [\mathbf{X}_1 | \mathbf{X}_2, \dots, \mathbf{X}_p],$$

откуда получим

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \dots & x_{\tau-1} & x_{\tau} & s_2^1 & \dots & s_p^1 \\ x_{\tau+1} & x_{\tau+2} & \dots & x_{2\tau-1} & x_{2\tau} & s_2^2 & \dots & s_p^2 \\ \dots & \dots \\ x_{T-\tau+1} & x_{T-\tau+2} & \dots & x_{T-1} & x_T & s_2^m & \dots & s_p^m \end{pmatrix},$$

где s_i^j — j -ое значение ряда s_i .

Опишем в чем состоит задача выбора оптимальной модели. Задана выборка $D = (\{\mathbf{x}_i, y_i\}), i \in \mathcal{J}$, где множество свободных переменных — вектор $\mathbf{x} = [x_1, \dots, x_j, \dots, x_n]$, проиндексированно $j \in \mathcal{J} = \{1, \dots, n\}$. Задано разбиение множества индексов элементов выборки $\mathcal{J} = \mathcal{L} \sqcup \mathcal{C}$. Так же задан класс регрессионных моделей $f(\mathbf{w}, \mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle$ — параметрических функций, линейных относительно параметров. Функция ошибки задана следующим образом

$$S = \sum_{i \in \mathcal{X}} (y_i - f(\mathbf{w}, \mathbf{x}_i))^2, \quad (2)$$

где $\mathcal{X} \subseteq \mathcal{J}$ — некоторое множество индексов. Требуется найти такое подмножество индексов $\mathcal{A} \subseteq \mathcal{J}$, которое бы доставляло минимум функции

$$\mathcal{A}^* = \operatorname{argmin}_{\mathcal{A} \subseteq \mathcal{J}} S(f_{\mathcal{A}} | \mathbf{w}^*, \mathcal{D}_{\mathcal{C}}) \quad (3)$$

на множестве индексов \mathcal{C} . При этом параметры \mathbf{w}^* модели должны доставлять минимум функции

$$\mathbf{w}^* = \operatorname{argmin}_{\mathbf{w} \in \mathcal{W}} S(\mathbf{w} | \mathcal{D}_{\mathcal{L}}, f_{\mathcal{A}}) \quad (4)$$

на множестве индексов \mathcal{L} . Здесь $f_{\mathcal{A}}$ обозначает модель f , включающую только столбцы матрицы X с индексами из множества \mathcal{A} , а обозначение вида $S(\mathbf{w} | \mathcal{D})$ означает, что переменная \mathcal{D} фиксирована, а переменная \mathbf{w} изменяется.

3 Выбор признаков при прогнозировании

Предположим, что мы имеем данные о цене на электроэнергию за год. Тогда матрица \mathbf{X} имеет размерность $(\tau - 1) \times (n - 1)$, где $\tau = 24$, а $n = 365$. То есть \mathbf{X} является матрицей 23×364 . Как видно, авторегрессионная матрица является переопределенной, поэтому необходим отбор признаков. Для отбора признаков в предлагается модифицировать метод шаговой регрессии.

3.1 Процедура выбора оптимального набора признаков

Опишем два этапа алгоритма Add и Del. На первом этапе последовательно добавляются признаки согласно (4), доставляющие минимум S на обучающей выборке, заданной множеством индексов \mathcal{L} . На втором этапе происходит последовательное удаление признаков согласно методу Белсли. Пусть на k -ом шаге алгоритма имеется активный набор признаков $\mathcal{A}_k \in \mathcal{J}$. На нулевом шаге \mathcal{A}_0 пуст. Опишем этапы Add и Del.

Этап Add. Находим признак доставляющий минимум S на обучающей выборке

$$j^* = \arg \min_{j \in \mathcal{J} \setminus \mathcal{A}_{k-1}} S(\mathbf{w} | \mathcal{D}_{\mathcal{L}}, f_{\mathcal{A}_{k-1} \cup \{j\}}).$$

Затем добавляем новый признак j^* к текущему активному набору

$$\mathcal{A}_k = \mathcal{A}_{k-1} \cup \{j^*\}$$

и повторяем эту процедуру до тех пор пока $S(f_{\mathcal{A}_k} | \mathbf{w}^*, \mathcal{D})$ превосходит свое минимальное значение на данном этапе не более чем на некоторое заданное значение ΔS_1 .

Этап Del. Находим индексы обусловленности и долевы коэффициенты для текущего набора признаков \mathcal{A}_{k-1} согласно методу Белсли, описание которого приведено ниже. Далее находим максимальный индекс обусловленности

$$i^* = \arg \max_{i \in \mathcal{A}_{k-1}} \eta_i. \quad (5)$$

Затем ищем максимальный долевы коэффициент соответствующий

найденному индекс обусловленности η_{i^*}

$$j^* = \arg \max_{j \in \mathcal{A}_{k-1}} q_{i^*j}. \quad (6)$$

Удаляем j^* -ый признак из текущего набора

$$\mathcal{A}_k = \mathcal{A}_{k-1} \setminus j^*$$

и повторяем эту процедуру до тех пор пока $S(f_{\mathcal{A}_k} | \mathbf{w}^*, \mathcal{D})$ превосходит свое минимальное значение на данном этапе не более чем на некоторое заданное значение ΔS_2 .

Повторение этапов Add и Del осуществляется до тех пор пока значение $S(f_{\mathcal{A}_k} | \mathbf{w}^*, \mathcal{D})$ не стабилизируется.

3.2 Метод Белсли для удаления признаков

Рассмотрим матрицу признаков \mathbf{X} . Она имеет размерность $m \times n$. Выполним ее сингулярное разложение:

$$\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T,$$

где \mathbf{U} , \mathbf{V} — ортогональные матрицы размерностью соответственно $m \times m$ и $n \times n$ и $\mathbf{\Lambda}$ — диагональная матрица с элементами (сингулярными числами) на диагонали, такими что

$$\lambda_1 > \lambda_2 > \dots > \lambda_r,$$

где r — ранг матрицы \mathbf{X} . Заметим, что в нашем случае $r = n$. Это связано с тем, что в алгоритме шагового выбора на каждом шаге мы имеем мультиколлиниарный, но невырожденный набор признаков.

Столбцы матрицы \mathbf{V} являются собственными векторами, а квадраты сингулярных чисел — собственными значениями корреляционной матрицы $\mathbf{X}^T\mathbf{X}$.

$$\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{\Lambda}^T\mathbf{U}^T\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T = \mathbf{V}\mathbf{\Lambda}^2\mathbf{V}^T,$$

$$\mathbf{X}^T\mathbf{X}\mathbf{V} = \mathbf{V}\mathbf{\Lambda}^2.$$

Отношение максимального сингулярного числа к j -му сингулярному числу назовем индексом обусловленности с номером j :

$$\eta_j = \frac{\lambda_{\max}}{\lambda_j}.$$

Если матрица X неполноранговая, то значительная часть индексов обусловленности неопределено. Однако, в нашем случае, как упоминалось выше, матрица признаков X является матрицей полного ранга.

Используя сингулярное разложение, дисперсия параметров, найденных методом наименьших квадратов $\mathbf{w} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$, где \mathbf{w} — вектор параметров модели, может быть записана как

$$\mathbf{var}(\mathbf{w}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 (\mathbf{V}^T)^{-1} \mathbf{\Lambda}^{-2} \mathbf{V}^{-1} = \sigma^2 \mathbf{V} \mathbf{D}^{-2} \mathbf{V}^T.$$

где σ^2 — это дисперсия регрессионных остатков. Таким образом дисперсия j -го регрессионного коэффициента — это j -й диагональный элемент матрицы $\mathbf{Var}(\mathbf{w})$.

Для обнаружения мультиколлинеарности признаков построим таблицу, в которой каждому индексу обусловленности η_j соответствуют значения q_{ij} — долевые коэффициенты. Сумма долевых коэффициентов по индексу j равна единице,

$$\sigma^{-2} \mathbf{var}(w_i) = \sum_{j=1}^n \frac{v_{ij}^2}{\lambda_j^2} = (q_{i1} + q_{i2} + \dots + q_{in}) \sum_{j=1}^n \frac{v_{ij}^2}{\lambda_j^2},$$

где q_{ij} — отношение соответствующего слагаемого в разложении вектора $\sigma^{-2} \mathbf{var}(w_i)$ ко всей сумме, а $\mathbf{V} = (v_{ij})$.

Мультиколлинеарность определяется с помощью таблицы 1: большие величины η_j означают, что возможно есть зависимость между признаками. Большие значения q_{ij} в соответствующих строках относятся к признакам, между которыми эта зависимость существует. Малые значения η_j также исследуются: между признаками, соответствующими большим значениям q_{ij} , зависимости не существует. Для нахождения мультиколлинеарных признаков решаются задачи (5) и (6).

4 Описание базового алгоритма прогноза

В вычислительном эксперименте проведено сравнение предлагаемого алгоритма с базовым методом SSA. Приведем его краткое описание.

Для последующего разложение ряда $\mathbf{s}_1 = \{x_i\}_{i=1}^T$ по главным компонентам преобразуем ряд в траекторную матрицу (матрицу Ганкеля) \mathbf{Y} , которую строим следующим образом:

$$\mathbf{Y} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ x_2 & x_3 & \dots & x_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_m & x_{m+1} & \dots & x_T \end{pmatrix}, \quad (7)$$

где $n = T - m + 1$ — время жизни гусеницы. Матрицу (7) будем называть нецентрированной траекторной матрицей, порожденной гусеницей длины n . Проводимый в дальнейшем анализ главных компонент может проводиться как по центрированной, так и по нецентрированной выборкам. Для упрощения выкладок рассмотрим простейший нецентрированный вариант.

Построим ковариационную матрицу следующим образом:

$$\mathbf{C} = \frac{1}{n} \mathbf{Y}^T \mathbf{Y}.$$

Так как матрица Ганкеля \mathbf{Y} невырождена, то матрица \mathbf{C} является полноранговой, то есть ее ранг равен n . Выполним её сингулярное разложение:

$$\mathbf{C} = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^T,$$

где $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$ — диагональная матрица собственных чисел, $\mathbf{V} = [v^1, \dots, v^n]$ — ортогональная матрица собственных векторов-столбцов. При этом будем предполагать, что собственные векторы упорядочены по убыванию соответствующих собственных чисел, т. е. $\lambda_1 > \lambda_2 > \dots > \lambda_n$.

Перейдем к прогнозированию временных рядов методом гусеницы. Рассмотрим систему уравнений:

$$\begin{cases} \sum_{j=1}^n h_j v_1^j & = & x_{m+1}, \\ & \dots & \\ \sum_{j=1}^n h_j v_{n-1}^j & = & x_T. \end{cases} \quad (8)$$

Пусть $\mathbf{h}^* = [h_1^*, \dots, h_n^*]$ — решение системы (8), тогда для продолжения ряда получим

$$x_{T+1} = \sum_{j=1}^n h_j^* v_n^j.$$

Выбираем r главных компонент из матрицы V

$$\mathbf{V}^* = \begin{pmatrix} v_1^{i_1} & v_1^{i_2} & \dots & v_1^{i_r} \\ v_2^{i_1} & v_2^{i_2} & \dots & v_2^{i_r} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n-1}^{i_1} & v_{n-1}^{i_2} & \dots & v_{n-1}^{i_r} \end{pmatrix},$$

где i_1, \dots, i_r — номера выбранных главных компонент. Введем следующие обозначения $\mathbf{v} = (v_n^{i_1}, v_n^{i_2}, \dots, v_n^{i_r})$, $\mathbf{q} = (x_{m+1}, \dots, x_n)^T$ и $\tilde{\mathbf{h}} = (h_{i_1}, \dots, h_{i_r})^T$. В этих обозначениях система (8) запишется как

$$\mathbf{V}^* \tilde{\mathbf{h}} = \mathbf{q}. \quad (9)$$

Учитывая (9), можно записать для прогнозируемого значения x_{T+1} следующую формулу:

$$x_{T+1} = \mathbf{v}((\mathbf{V}^*)^T \mathbf{V}^*)^{-1} (\mathbf{V}^*)^T \mathbf{q}.$$

Таким образом мы построили прогноз — следующее по времени значение x_{T+1} временного ряда $\mathbf{s}_1 = \{x_i\}$.

5 Вычислительный эксперимент

С целью сравнить предложенный в настоящей работе подход с базовым алгоритмом SSA была проведена серия экспериментов по краткосрочному прогнозированию временных рядов. Данные — почасовые цены на электроэнергию в Германии за период с 1 января 2003г. по 10 июля 2009г., опубликованы в [15]. Мы прогнозируем почасовые цены на ближайшие сутки по предыдущей истории. На рис. 1 приведен график отражающий колебание цен на электроэнергию в течение недели. На нем виден спад цен в ночное время и в выходные дни. Диаграмма на рис. 2 показывает изменение цен за четыре недели в зависимости от времени суток. На рис. 3 приведены два графика. Первый отражает зависимость реальных данных о ценах от времени, а второй —

зависимость спрогнозированных, с помощью предложенного в работе алгоритма, на сутки цен от времени. На рис. 4 показано, как изменяется квадратичная ошибка S , определенная в (2), на тестовой выборке \mathcal{C} , от итерации к итерации.

Сравним результаты работы предложенного алгоритма с базовым алгоритмом SSA. Результаты экспериментов показаны в таблице 2. Для каждого алгоритма вычислялись ошибки MSE (средний квадратичное отклонение) и MAPE (средняя абсолютная процентная ошибка):

$$\text{MSE} = \frac{1}{\tau} \sum_{i=T+1}^{T+\tau} (\tilde{x}_i - x_i)^2, \quad \text{MAPE} = \frac{1}{\tau} \sum_{i=T+1}^{T+\tau} 100 \frac{|\tilde{x}_i - x_i|}{|x_i|},$$

где \tilde{x}_i — спрогнозированное значение целевого ряда в точке i , x_i — фактическое значение этого ряда в точке i . Так же сравнивались значения информационных критериев Акаике (AIC) и Байеса (BIC)

$$\text{AIC} = \tau \left(\ln \frac{S}{\tau} \right) + 2|\mathcal{A}|, \quad \text{BIC} = \tau \left(\ln \frac{S}{\tau} \right) + |\mathcal{A}| \ln \tau,$$

где S — среднеквадратичная ошибка, вычисленная по набору активных признаков \mathcal{A} . В таблице так же указано число n признаков, входящих в модель.

6 Заключение

Предложенный подход значительно эффективнее базового алгоритма, особенно случае, когда выборка содержит большое число мультиколлинеарных признаков. Вычислительный эксперимент показал, что увеличение числа признаков, из которых производится выбор, позволяет добиться улучшения качества модели. Однако при этом требуется введение дополнительных условий, позволяющих избежать появление мультиколлинеарных признаков. Предлагаемый алгоритм включает процедуру анализа мультиколлинеарности и позволяет выбирать хорошо обусловленные наборы признаков.

Литература

[1] Frisch R. *Statistical confluence analysis by means of complete regression systems*,

- Universitetets Okonomiske Institute, 1934.
- [2] Efron B., Hastie T., Johnstone I., Tibshirani R. *Least angle regression*, The Annals of Statistics, 2004, Vol. 32, no. 3. Pp. 407-499.
 - [3] Tibshirani R. *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society, 1996, Vol. 32, no. 1. Pp. 267-288.
 - [4] Draper N. R., Smith H. *Applied regression analysis*, John Wiley and Sons, 1998.
 - [5] Chen Y. W., Billings C. A., Luo W. *Orthogonal least squares methods and their application to non-linear system identification*, International Journal of Control, 1989, Vol. 2, no. 50. Pp. 873-896.
 - [6] Chen S., Cowan C. F. N., Grant P. M. *Orthogonal least squares learning algorithm for radial basis function network*, Transaction on neural network, 1991, Vol. 2, no. 2. Pp. 302-309.
 - [7] Efron B., Tibshirani R., Hastie T., Johnstone I., Wager S., Taylor T., Van der Vaart A. *Multiple regression analysis*, New York: Wiley, 1996.
 - [8] Rawlings J. O., Pantula S. G., Dickey D. A. *Applied regression analysis: a research tool*, New York: Springer-Verlag, 1998.
 - [9] Стрижов В. В., Крымова Е. А. *Выбор моделей в линейном регрессионном анализе*, Информационные технологии, № 10, 2011. С. 21—26.
 - [10] Belsley D. A. *Conditioning Diagnostics: Collinearity and weak data in regression*, New York: John Wiley and Sons, 1991.
 - [11] Tarantola A. *Inverse problem theory*, SIAM, 2005.
 - [12] Johnstone I., Tibshirani R., Efron B., Hastie T. *Least angle regression*, Springer, 2004.
 - [13] McNamee J. *Innovations in local modeling for time series prediction*, 1999.
 - [14] В. Н. Солнцев, Д. Л. Данилов, А. А. Жигляевский. *Главные компоненты временных рядов: Метод «Гусеница»*, С.-Пб. ГУ, 1997.
 - [15] V.Strijov, Lectures on Data Mining and Business Analytics, Time Series Set, <https://dmba.svn.sourceforge.net/svnroot/dmba/Data/GermanSpotPrice.csv>, 2011.

Таблица 1: Разложение $\text{var}(w_i)$

Индекс обусловленности	$\text{var}(w_1)$	$\text{var}(w_2)$...	$\text{var}(w_n)$
η_1	q_{11}	q_{21}	...	q_{n1}
η_2	q_{12}	q_{22}	...	q_{n2}
\vdots	\vdots	\vdots	\ddots	\vdots
η_n	q_{1n}	q_{2n}	...	q_{nn}

Таблица 2: Результаты работы алгоритмов

Алгоритмы выбора признаков	MSE	MAPE в рабочие дни	MAPE в выходные дни	AIC	BIC	Число признаков
Предложенный алгоритм	8.18	6.17	10.33	96.44	123.53	23
Метод SSA	13.25	16.16	29.01	130.74	149.59	16

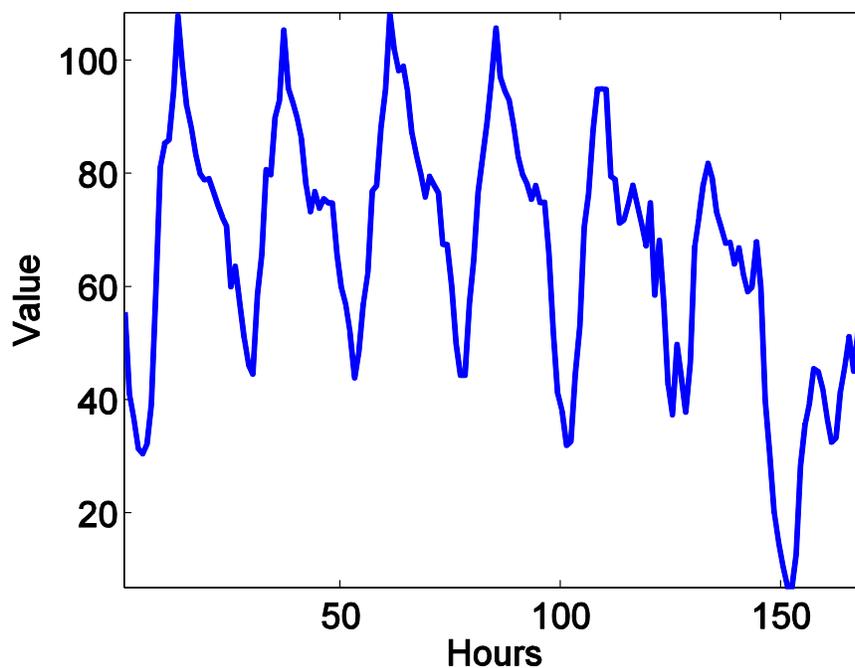


Рис. 1: Диаграмма зависимости цен от времени суток и дня недели

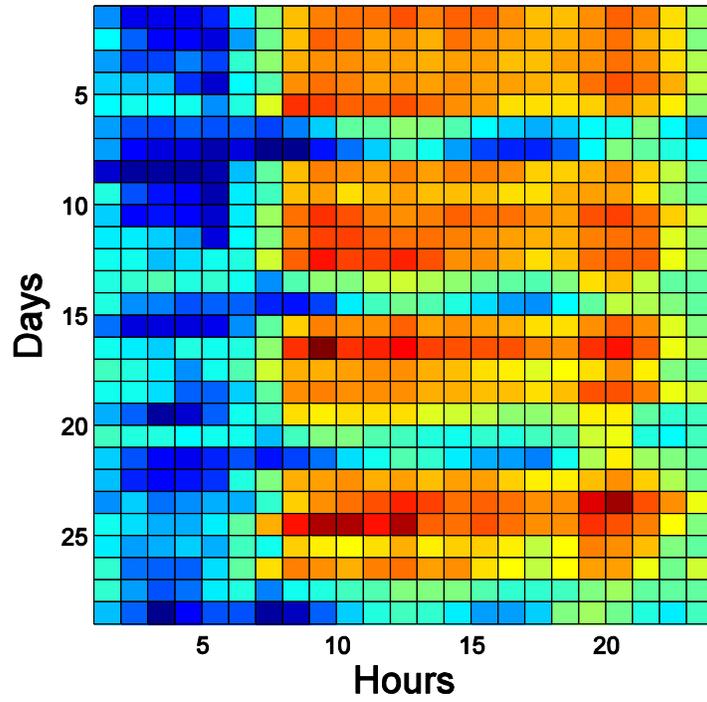


Рис. 2: Визуализация автокорреляционной матрицы

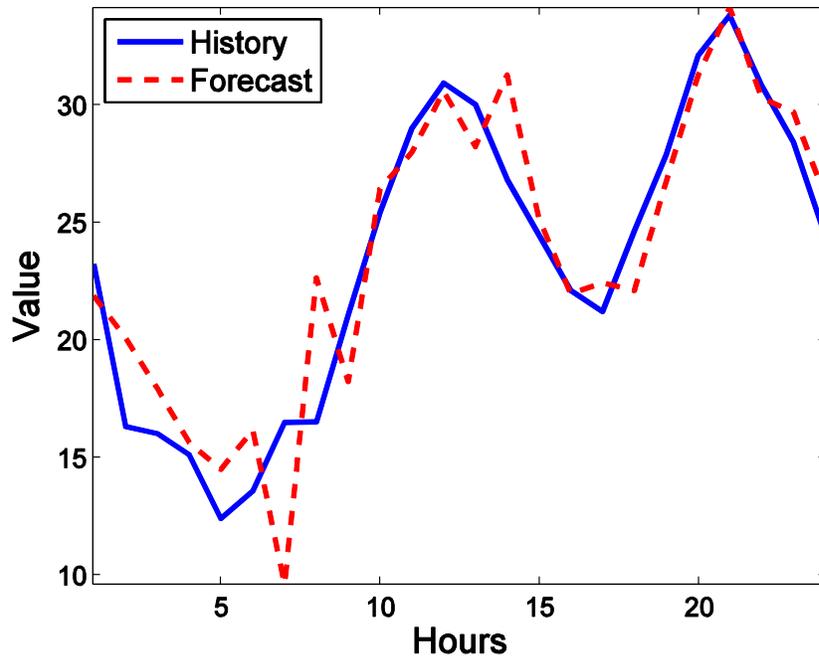


Рис. 3: Прогнозирование почасовых значений ряда на сутки вперед

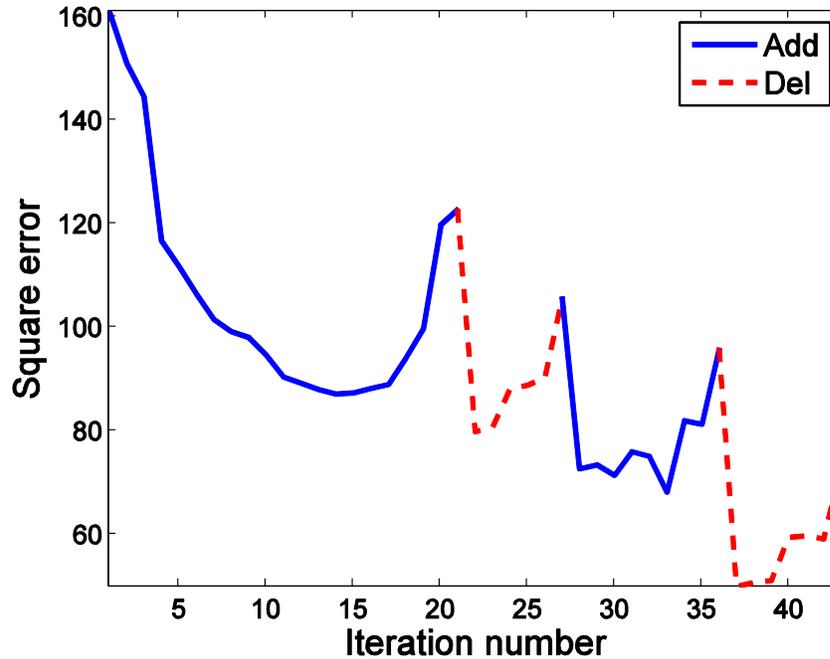


Рис. 4: Зависимость функции ошибки S от номера итерации

L. N. Sanduleanu, Moscow Institute of Physics and Technology

V. V. Strijov, Computing Center of the Russian Academy of Sciences

Feature selection in autoregression forecasting

The authors investigate the optimal model selection problem with application to the auto-regression forecasting. To solve the problem one has to select a maximum well-defined feature subset, subject to some given value of the error function. To select the feature set the modified add-del feature selection algorithm is used. This paper suggests a method of time series forecasting model selection. The computational experiment compares the electricity hourly prices forecasts.

Keywords: feature selection, multicollinearity, stepwise regression, Belsley method, time series forecasting.