

Выбор оптимального набора признаков из мультикоррелирующего множества в задаче прогнозирования*

Р. Г. Нейчев¹, А. М. Камруца¹, В. В. Стрижов²
neychhev@phystech.edu, amkatrutsa@yandex.ru, strijov@gmail.com

В данной статье рассматривается проблема прогнозирования временных рядов. Для получения устойчивого прогноза предлагается рассматривать входные временные ряды как матрицу объект-признак и использовать отбор признаков. В условиях мультиколлинеарности признаков необходим критерий для ее обнаружения. Для этого предлагается применить подход, основанный на методе Белсли. Исключение коррелирующих признаков при отборе позволяет сократить размерность задачи и получить устойчивые оценки параметров модели. Для отбора признаков в работе предлагается метод добавления и удаления признаков. В качестве практической проверки данного метода в ходе вычислительного эксперимента решается задача прогнозирования почасовых значений цен на электроэнергию. Эксперименты были проведены на реальных данных о ценах на электроэнергию в Германии.

Ключевые слова: устойчивость модели, выбор признаков, метод Белсли, почасовое прогнозирование цен, прогнозирование временных рядов, метод добавления-удаления признаков, линейная регрессия.

Введение

В данной работе рассматривается задача прогнозирования временных рядов. В современном мире большинство наблюдаемых величин, особенно связанных с деятельностью человека, обладают периодичностью: цены на авиабилеты значительно зависят от месяца, загруженность дорог — от дня недели и времени суток, спрос на сезонные товары — от времени года. Поэтому в данной работе рассматриваются временные ряды, обладающие периодичностью. Для построения прогноза предлагается использовать метод авторегрессии [1], который сводит задачу прогнозирования к задаче линейной регрессии [2].

Кроме метода авторегрессии, для построения прогнозов используются другие подходы. Например, алгоритм *ARIMA*, являющийся обобщением алгоритма авторегрессионного скользящего среднего (*ARMA*), или алгоритм *Гусеницы* [3], который заключается в преобразовании одномерного временного ряда в многомерный ряд и применении к нему метода главных компонент [4]. Сложность данного алгоритма квадратична по отношению к длине ряда, поэтому для длинных временных рядов он менее предпочтителен, чем алгоритмы с линейной зависимостью. Под *сложностью* алгоритма понимается зависимость необходимых вычислительных ресурсов от длины подаваемого на вход временного ряда.

В задаче регрессии требуется восстановить значение целевого вектора на основе заданных признаков, наличие мультиколлинеарности между которыми приводит к получению в качестве решения неустойчивой модели. Модель будем называть *устойчивой*, если любые малые изменения вектора параметров приводят к слабым изменениям целевого вектора. *Мультиколлинеарность* — это наличие сильной зависимости между признаками, кото-

1 — Московский физико-технический институт (Moscow Institute of Physics and Technology),

2 — Вычислительный центр РАН им. Дородницына (Dorodnicyn Computing Centre of RAS),

* Работа выполнена при финансовой поддержке РФФИ, проект 14-07-31264.

рая приводит к значительному снижению устойчивости модели [5]. Для решения проблемы мультиколлинеарности предлагается применить отбор признаков.

Рассмотрим подходы, которые используются для отбора признаков. В методе *Lasso* [6] к среднеквадратичной ошибке добавляется регуляризационный член, равный l_1 -норме вектора параметров. Это приводит к занулению некоторых элементов вектора параметров и, как следствие, к отбору признаков. Метод *LAD-Lasso* [6, 7] имеет преимущество перед *Lasso*, которое заключается в большей устойчивости к ошибкам и более точном отборе признаков благодаря тому, что каждый признак имеет собственную величину штрафа. Метод наименьших углов *LARS* [8] является обобщенной версией *Lasso*. Он оценивает веса свободных переменных, изменяя их таким образом, чтобы доставить наибольшую корреляцию с вектором регрессионных остатков. Основное преимущество этого метода заключается в том, что он выполняется за число шагов, не превышающее число свободных переменных. Метод *группового учета аргументов* [9] основан на рекурсивном отборе моделей, на основе которых строятся более сложные модели, за счет чего точность моделирования увеличивается на каждом шаге. *Гребневая регрессия* [10] применяется в случае мультиколлинеарности признаков, в ней в качестве регуляризатора к величине среднеквадратичной ошибки добавляется l_2 -норма вектора параметров.

В данной работе в качестве способа отбора признаков предлагается модифицированная версия шаговой регрессии *Add-Del* [1]. Добавление признаков производится с помощью метода *FOS* [11]. Он последовательно добавляет признаки, максимально коррелирующие с вектором регрессионных остатков. Удаление признаков производится с помощью метода *Белсли* [12]. Данный метод позволяет обнаружить мультикоррелирующие признаки и получить более устойчивую модель, исключив их из рассмотрения. Предполагается, что использование данных методов при отборе признаков позволит получить устойчивую модель, для работы с которой не требуется больших вычислительных ресурсов. Сложность предлагаемого метода линейно зависит от числа признаков.

Для практической проверки предлагаемого метода в рамках вычислительного эксперимента решается задача построения почасового прогноза цены на электроэнергию [13, 14]. Эксперимент ставится на реальных данных и состоит из двух частей. В первой части на основе отобранных предлагаемым методом признаков, строится прогноз методом авто-регрессии. Результаты сравниваются с реальным поведением цен после рассматриваемого периода и с результатами работы других алгоритмов. Во второй части рассматривается непосредственно отбор признаков и производится сравнение результатов работы предлагаемого и других методов.

Постановка задачи прогнозирования

Рассмотрим набор временных рядов $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_p$ вида $\mathbf{s}_j = \{x_{ij}\}$, $i = 1, \dots, T - 1$, $j = 1, \dots, p$. Ряд \mathbf{s}_1 будем называть *целевым*. Необходимо спрогнозировать следующие τ значений целевого ряда \mathbf{s}_1 , опираясь на ряды $\mathbf{s}_1, \dots, \mathbf{s}_p$. Набор рядов обладает следующими свойствами: отсчеты времени i сделаны через равные промежутки, ряд \mathbf{s}_j имеет периодическую составляющую τ и не имеет пропущенных значений, длина ряда $T - 1$ кратна периоду τ .

Построим авторегрессионную матрицу для целевого ряда \mathbf{s}_1 :

$$\mathbf{X}_1 = \left(\begin{array}{cccc|c} x_1 & x_2 & \dots & x_{\tau-1} & x_{\tau} \\ \dots & \dots & \dots & \dots & \dots \\ x_{j\tau+1} & x_{j\tau+2} & \dots & x_{(j+1)\tau-1} & x_{(j+1)\tau} \\ \dots & \dots & \dots & \dots & \dots \\ \hline x_{(m-2)\tau+1} & x_{(m-2)\tau+2} & \dots & x_{(m-1)\tau-1} & x_{(m-1)\tau} \\ x_{T-\tau+1} & x_{T-\tau+2} & \dots & x_{T-1} & x_T \end{array} \right).$$

Представим матрицу \mathbf{X}_1 в виде

$$\mathbf{X}_1 = (\mathbf{X}|\mathbf{y}).$$

Для временных рядов $\mathbf{s}_j, j = 2, \dots, p$, построим авторегрессионные матрицы \mathbf{X}_j . Полученные матрицы припишем справа к матрице \mathbf{X} для целевого ряда. В результате получим матрицу

$$\mathbf{X}^* = (\mathbf{X}|\mathbf{X}_2, \dots, \mathbf{X}_p|\mathbf{y}).$$

Разобьем полученную авторегрессионную матрицу \mathbf{X}^* следующим образом:

$$\mathbf{X}^* = \left(\begin{array}{c|c} \mathbf{X} & \mathbf{y} \\ \mathbf{x}_m & x_T \end{array} \right),$$

выделив последний столбец и строку. Заметим, что вектор \mathbf{y} и матрица \mathbf{x} были переобозначены. Для нахождения значения x_T поставим задачу линейной регрессии:

$$\|\mathbf{X}\mathbf{w} - \mathbf{y}\|^2 \longrightarrow \min_{\mathbf{w}}, \quad (1)$$

где \mathbf{w} — вектор параметров. Тогда x_T найдем как как

$$x_T = \mathbf{x}_m \mathbf{w}. \quad (2)$$

Чтобы не перестраивать авторегрессионную матрицу для каждого следующего значения временного ряда, в качестве вектора \mathbf{x}_m выбираются $\tau - 1$ значений ряда предыдущего периода. Необходимо найти такой вектор параметров \mathbf{w}^* , на котором достигается минимум функции ошибки S . Оптимальный вектор параметров для задачи (1) имеет вид:

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

При работе с временными рядами большой длины авторегрессионная матрица является недоопределенной. Например, рассмотрим временной ряд с почасовыми показаниями за два месяца. Матрица \mathbf{X} будет иметь размеры $(24 - 1) \times (31 \cdot 2 - 1) = 23 \times 61$. Чтобы понизить ее размерность, поставим задачу отбора признаков. Признаками будем считать столбцы матрицы \mathbf{X} .

Постановка задачи отбора признаков

Представим целевой вектор \mathbf{y} в виде

$$\mathbf{y} = \mathbf{X}\mathbf{w} + \boldsymbol{\varepsilon}(\mathbf{X}),$$

где $\boldsymbol{\varepsilon}$ — вектор регрессионных остатков. Пару (\mathbf{X}, \mathbf{y}) назовем выборкой и обозначим как \mathcal{D} . Признаки, которым соответствуют ненулевые члены вектора \mathbf{w} , будем называть *активными*, а остальные признаки будем считать *исключенными*. Множество индексов элементов выборки обозначим \mathcal{I} и разобьем на непересекающиеся подмножества: $\mathcal{I} = \mathcal{L} \sqcup \mathcal{C}$.

Функцию ошибки $S(\mathbf{f}(\mathbf{w}, \mathbf{X}), \mathbf{y})$ зададим как квадрат нормы вектора регрессионных остатков:

$$S = \|\boldsymbol{\varepsilon}(\mathbf{X})\|^2. \quad (3)$$

Подмножество индексов активных признаков обозначим $\mathcal{A} \subset \mathcal{J}$, где \mathcal{J} есть множество индексов всех признаков. Назовем *моделью* пару $(\mathbf{f}, \mathcal{A})$ и обозначим ее как $\mathbf{f}_{\mathcal{A}}$. При постановке задачи линейной регрессии функция \mathbf{f} фиксирована, поэтому для выбора модели необходимо найти множество индексов \mathcal{A} , минимизирующее функцию ошибки S на элементах выборки $\mathcal{D}_{\mathcal{C}}$, состоящей из элементов выборки \mathcal{D} с индексами из множества \mathcal{C} :

$$\mathcal{A}^* = \arg \min_{\mathcal{A} \subset \mathcal{J}} S(\mathcal{A} | \mathbf{w}^*, \mathcal{D}_{\mathcal{C}}). \quad (4)$$

Запись вида $S(\mathcal{A} | \mathcal{D})$ означает, что выборка \mathcal{D} фиксирована, а выборка \mathcal{A} меняется. Для решения задачи (4) требуется найти вектор параметров \mathbf{w}^* , доставляющий минимум функции ошибки S на элементах выборки $\mathcal{D}_{\mathcal{C}}$, определяемой аналогично \mathcal{D} :

$$\mathbf{w}^* = \arg \min_{\mathbf{w} \in \mathbb{W}} S(\mathbf{w} | \mathcal{D}_{\mathcal{C}}, \mathcal{A}), \quad (5)$$

где \mathbb{W} — множество возможных значений вектора \mathbf{w} .

Метод Белсли

Между признаками возможно существование мультиколлинеарной зависимости. Например, если временной ряд обладает посуточной периодичностью, его значение в один и тот же час сильно коррелирует со значениями в соседние часы. В случае мультиколлинеарности признаков оценка вектора параметров (5) является неустойчивой. Для устранения данной проблемы необходимо найти мультиколлинеарные признаки. Применим для этого метод Белсли. Рассмотрим сингулярное разложение матрицы \mathbf{X} :

$$\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^{\top}. \quad (6)$$

\mathbf{U} и \mathbf{V} — ортогональные матрицы, \mathbf{D} — диагональная матрица, состоящая из сингулярных чисел λ_j таких, что

$$\lambda_1 < \lambda_2 < \dots < \lambda_r,$$

где r — ранг матрицы \mathbf{D} . В нашей задаче мы предполагаем, что матрица невырождена, поэтому $r = n$. Ковариационная матрица вектора параметров \mathbf{w} оценивается как $\mathbf{A}^{-1} = \mathbf{X}^{\top}\mathbf{X}$. Сингулярные числа λ_i являются собственными значениями, а столбцы матрицы \mathbf{V} — собственными векторами ковариационной матрицы $\hat{\mathbf{A}}^{-1}$. Используя сингулярное разложение (6), запишем:

$$\hat{\mathbf{A}}^{-1} = \mathbf{X}^{\top}\mathbf{X} = \mathbf{V}\mathbf{D}^{\top}\mathbf{U}^{\top}\mathbf{U}\mathbf{D}\mathbf{V}^{\top} = \mathbf{V}\mathbf{D}^2\mathbf{V}^{\top},$$

Определим как i -ый *индекс обусловленности* отношение $\eta_i = \lambda_{\max}/\lambda_i$, где λ_{\max} — максимальное сингулярное число. Большое значение η_i указывает на близкую к линейной зависимость между признаками. Максимальный индекс обусловленности матрицы \mathbf{X} характеризует, насколько велико будет изменение компонент вектора параметров \mathbf{w} при изменении матрицы признаков \mathbf{X} . Назовем его *числом обусловленности* θ . Заметим, что число обусловленности матрицы $\hat{\mathbf{A}}$ есть квадрат числа обусловленности матрицы \mathbf{X} .

В рамках нашей задачи для нахождения максимально коррелирующих признаков необходимо найти индекс i^* вида

$$i^* = \arg \max_{i \in \mathcal{A}} \eta_i, \quad (7)$$

где \mathcal{A} — текущее множество активных признаков.

Оценками дисперсии параметров будут диагональные элементы матрицы $\hat{\mathbf{A}}^{-1}$. *Дисперсионные доли* q_{ij} определим как вклад j -го признака в дисперсию i -го элемента вектора параметров \mathbf{w} :

$$q_{ij} = \frac{v_{ij}^2 / \lambda_j^2}{\sum_{j=1}^n v_{ij}^2 / \lambda_j^2}. \quad (8)$$

Из определения дисперсионных долей следует, что их большие значения указывают на наличие зависимости между признаками. Определим индекс j^* , вносящий максимальный вклад в дисперсию i -го элемента вектора \mathbf{w} :

$$j^* = \arg \max_{j \in \mathcal{A}} q_{i^*j}, \quad (9)$$

где максимальный индекс обусловленности i^* определяется (7).

Будем называть модель $\mathbf{f}_{\mathcal{A}}$ *неустойчивой*, если число обусловленности матрицы признаков \mathbf{X} велико: $\theta \gg 1$.

Исключение из модели $\mathbf{f}_{\mathcal{A}}$ параметров, максимально влияющих на минимальное сингулярное значение матрицы \mathbf{X} , доставляет максимальную устойчивость модели.

Действительно, пусть N объектов описываются моделью $\mathbf{f}_{\mathcal{A}}$, состоящей из n признаков χ_i . Некоторые признаки мультиколлинеарны. Обозначим максимальный индекс обусловленности η_{\max} , из ковариационной матрицы $\hat{\mathbf{A}}$ найдем максимальную соответствующую ему дисперсионную долю. Пусть она соответствует признаку χ_{j^*} . Это указывает на принадлежность признака χ_{j^*} к множеству мультикоррелирующих признаков. Мультиколлинеарность влечет близость к нулю одного или нескольких сингулярных значений λ_k .

Следовательно, исключение j^* -го признака из модели $\mathbf{f}_{\mathcal{A}}$ приведет к увеличению минимального сингулярного числа и уменьшению числа обусловленности θ' новой матрицы \mathbf{X}' . Меньшее число обусловленности характеризует лучшую устойчивость модели, а значит, исключение j^* -го признака ее повышает.

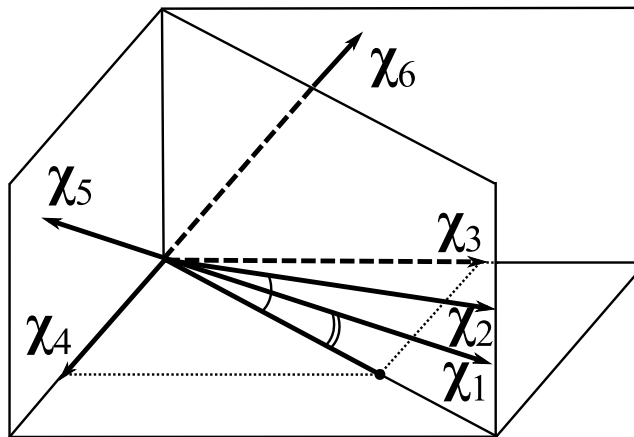


Рис. 1. Мультикоррелирующие, ортогональные векторы значений свободных переменных.

Для лучшего понимания принципа работы метода Белсли, рассмотрим его на конкретном примере. На Рис. 1 приведены 3 объекта, каждый из которых описывается шестью признаками χ_1, \dots, χ_6 . Расстоянием от объекта до признака будем считать величину угла между ними. Табл. 1 содержит разложение вектора дисперсий $\text{var}(\chi_j)$ по индексам обусловленности η_i . На Рис. 2 приведена визуализация оценки ковариационной матрицы параметров $\hat{\mathbf{A}}$.

Таблица 1. Разложение $\text{var}(\chi_j)$

	Индекс обусловленности	Дисперсия параметров					
		$\text{var}(\chi_1)$	$\text{var}(\chi_2)$	$\text{var}(\chi_3)$	$\text{var}(\chi_4)$	$\text{var}(\chi_5)$	$\text{var}(\chi_6)$
η_1	0	0	0	0	0	0	0
η_2	0	0	0	0	0	0	0
η_3	0	0	0	0	0	0	0
η_4	1	0.02	0.05	0	0.2	0.01	0.53
η_5	2.29	0.08	0.1	0.15	0.39	0.48	0.23
η_6	4.87	0.9	0.85	0.84	0.4	0.5	0.24

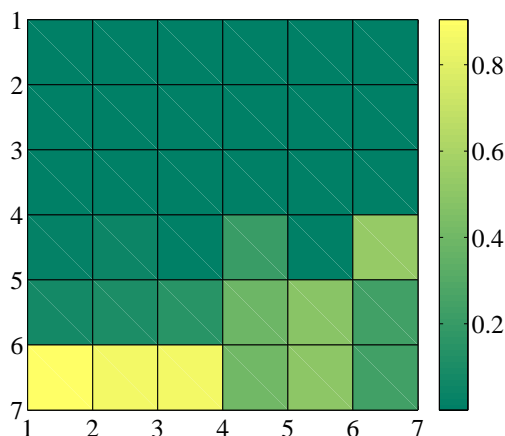


Рис. 2. Визуализация оценки ковариационной матрицы параметров $\hat{\mathbf{A}}$.

Согласно Табл. 1, максимальный индекс обусловленности $\eta_6 = 4.87$. В соответствующей ему шестой строке дисперсионные доли в первом и втором столбцах (выделены красным) максимальны. Они соответствуют признакам χ_1 и χ_2 , между которыми есть ярко выраженная зависимость, что наглядно видно на Рис. 1.

Метод отбора признаков Add-Del

Применим метод Белсли для отбора признаков. Предлагаемый метод состоит из двух этапов **Add** и **Del**. Метод предполагает, что регрессионные остатки имеют нормальное распределение с нулевым матожиданием и произвольной неотрицательно определенной ковариационной матрицей. В начальный момент множество активных признаков $\mathcal{A}_0 = \emptyset$. Рассмотрим работу алгоритма.

Этап **Add**. Выбираем признак j^* из множества исключенных признаков согласно

$$j^* = \arg \min_{j \in \mathcal{J} \setminus \mathcal{A}} S(\mathbf{w} | \mathcal{A} \cup \{j\}, \mathcal{D}) \quad (10)$$

и добавляем его к модели:

$$\mathcal{A} = \mathcal{A} \cup \{j^*\}. \quad (11)$$

Признаки добавляются к активному набору \mathcal{A} , пока функция ошибки не удовлетворит критерию останова:

$$S(\mathbf{w}_{\mathcal{A}} | \mathcal{D}) \geq S_{min} + \delta S_1. \quad (12)$$

Этап **Del**. В соответствии с критерием (9) находим признак j^* , максимально коррелирующий с другими, и удаляем его из множества :

$$\mathcal{A} = \mathcal{A} \setminus \{j^*\}. \quad (13)$$

Повторяем данную операцию, пока полученная ошибка $S(f_{\mathcal{A}_k} | \mathbf{w}^*, \mathcal{D})$ превосходит свое минимальное значение на данном этапе не более, чем на заданную величину δS_2 . Критерий останова этапа *Del* на данном шаге

$$S(\mathbf{w}_{\mathcal{A}} | \mathcal{D}_{\mathcal{L}}) \geq S_{min} + \delta S_2. \quad (14)$$

Этапы **Add** и **Del** повторяются до тех пор, пока значение ошибки не стабилизируется, то есть ее значение будет слабо изменяться от итерации к итерации. Величина колебания значения от итерации к итерации должна составлять величину порядка $\frac{1}{100}S$, более конкретную величину можно подобрать опытным путем.

Остановимся подробнее на критериях останова (12),(14). Для определения величин $\delta S_{1,2}$ рассмотрим свойства суммы квадратов регрессионных остатков

$$\mathbf{S} = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon}. \quad (15)$$

Вектор регрессионных остатков $\boldsymbol{\varepsilon}$, определяется выражением

$$\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\mathbf{w} = \left(\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \mathbf{y}, \quad (16)$$

где матрица $(\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T) = \mathbf{P}$ симметрична и идемпотентна: $\mathbf{P}^T = \mathbf{P}$ и $\mathbf{P}^2 = \mathbf{P}$.

Подставляя выражение (16) в (15) и учитывая, что

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

получаем

$$\mathbf{S} = \mathbf{y}^T \left(\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \mathbf{y} = \mathbf{y}^T \mathbf{y} - \mathbf{w}^T \mathbf{X}^T \mathbf{y}. \quad (17)$$

Здесь \mathbf{S} записана как квадратичная форма вектора \mathbf{y} . В предположении, что

$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\mathbf{w}, \sigma^2 \mathbf{I}), \quad (18)$$

математическое ожидание \mathbf{S} имеет вид [15].

Так как наиболее правдоподобная оценка параметров \mathbf{w} при предположении (18) имеет вид $\mathbf{w} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$, и функция ошибки в этом случае задана как $S(\mathbf{w}) = \mathbf{S}$, то математическое ожидание суммы квадратов регрессионных остатков имеет вид

$$E(S(\mathbf{w})) = \text{tr} \left(\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \mathbf{I} \sigma^2 + (\mathbf{X}\mathbf{w})^T \left(\mathbf{I} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \right) \mathbf{X}\mathbf{w}. \quad (19)$$

Так как след идемпотентной матрицы (в данном случае это матрица Мура-Пенроуза $\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$), то

$$\begin{aligned} E(S(\mathbf{w})) &= \text{rank} \left((\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) \right) \mathbf{I} \sigma^2 = \left(m - \text{rank}(\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) \right) \sigma^2 \\ &= (m - \text{rank}(\mathbf{X})) \sigma^2 = (m - n) \sigma^2, \end{aligned}$$

здесь m — число элементов выборки и строк матрицы \mathbf{X} . Если матрица плана \mathbf{X} не содержит коллинеарных столбцов и её ранг $\text{rank}(\mathbf{X}) = n$, то несмещенной оценкой σ^2 является оценка

$$\hat{\sigma}^2 = \frac{S(\mathbf{w})}{m - n}.$$

Поэтому величины $\delta S_{1,2}$ можно определить как дисперсию вектора регрессионных остатков S .

Вычислительный эксперимент

В рамках вычислительного эксперимента производится построение прогноза цен на электроэнергию. В ходе эксперимента сравниваются результаты прогнозирования с помощью предлагаемого метода (*Add-Del* и авторегрессия) и алгоритма *Гусеницы*. В качестве входных данных используются временные ряды с реальными ценами на электроэнергию в Германии [16].

Для иллюстрации недельной и суточной периодичности на Рис. 3 приведен график зависимости цены на электроэнергию от времени за четыре недели. На Рис. 4 показана авторегрессионная матрица для того же периода. Из Рис. 4 видно, что цена имеет недельную и суточную периодичности, а значит, может иметь место мультиколлинеарность признаков, в роли которых в данной задаче выступают ежедневные значения цен.

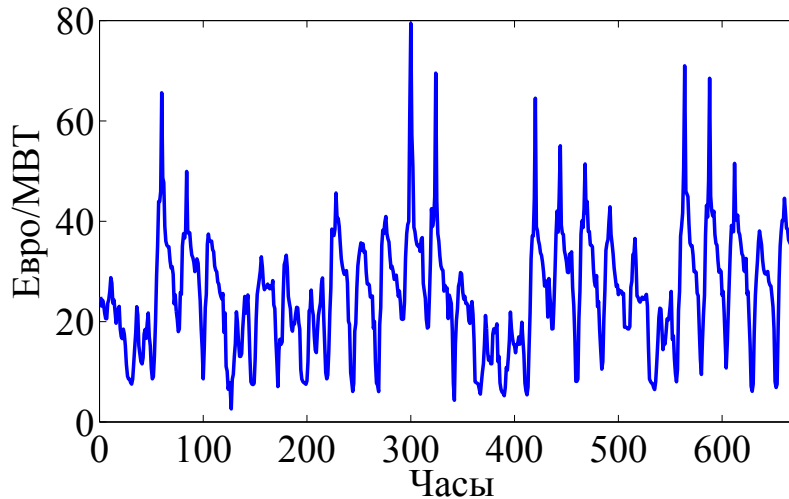


Рис. 3. Изменение цены в течение четырех недель.

На Рис. 5 приводится сравнение прогноза цены на пять суток с помощью авторегрессии с отбором признаков а) *Add-Del* и б) *Гусеницы*. Прогноз, полученный с помощью предложенного метода предоставляет большую точность при сопоставимом времени работы. Обучающая выборка содержит информацию за предшествующие 90 дней. Как видно из

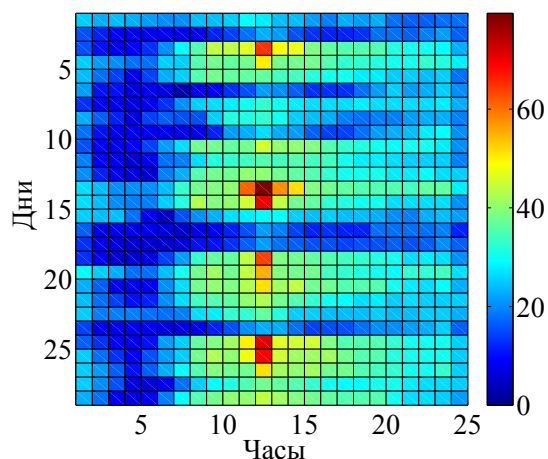


Рис. 4. Авторегрессионная матрица за четыре недели.

графика, алгоритмы получают сходные по качеству результаты. Время работы алгоритмов практически одинаково.

Для оценки качества модели, полученной с помощью *Add-Del* приводится ее сравнение с моделями, полученным *LARS* и *Lasso*. В данных методах также происходит отбор признаков путем обнуления коэффициентов вектора признаков \mathbf{w} . Для сравнения вычислялись информационные критерии Акаике (AIC) и Байеса (BIC):

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

$$\text{BIC} = \tau \left(\ln \frac{RSS}{\tau} \right) + |\mathcal{A}| \ln \tau,$$

где RSS — величина среднеквадратичной ошибки соответствующего алгоритма, вычисленная по набору активных признаков \mathcal{A} . Так же учитывается число вошедших в модель признаков. Результаты сравнения приведены в Табл. 2. На Рис. 6 для визуального сравнения приведены результаты прогнозирования методом авторегрессии на пять недель с использованием полученных моделей.

Таблица 2. Результаты построения модели

	RSS	AIC	BIC	Число признаков, n
Add-Del	57.97	237.97	343.99	11
LARS	0.82	178.82	283.67	23
Lasso	104.01	282.01	386.86	18

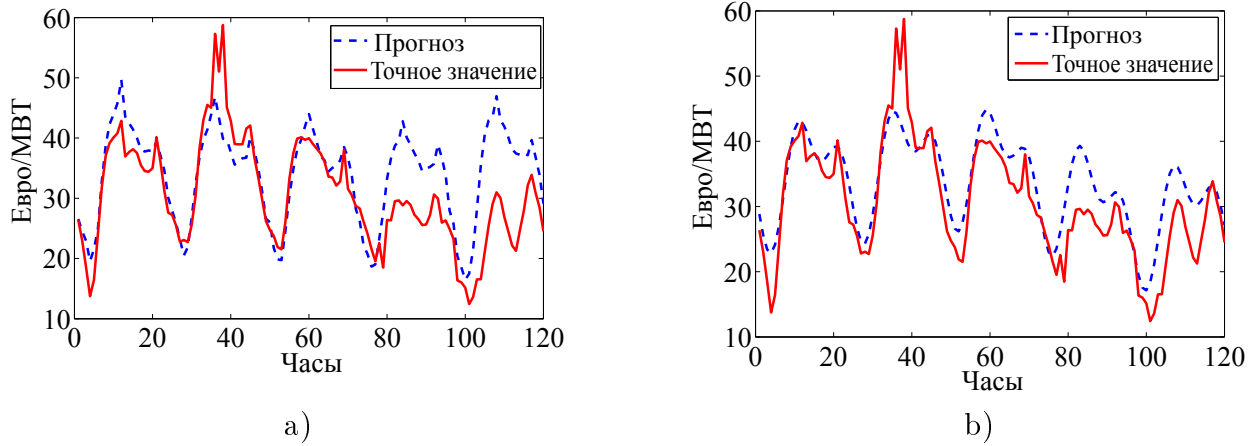


Рис. 5. Сравнение прогнозируемого и реального поведения цены с использованием: а) алгоритма Add-Del, б) алгоритма Гусеницы.

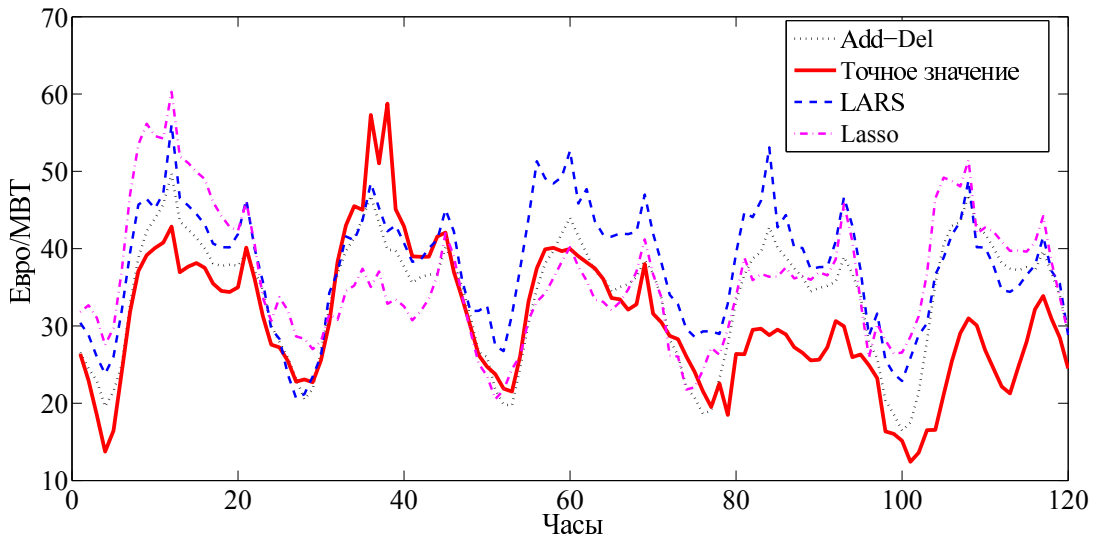


Рис. 6. Сравнение результатов работы алгоритмов, включающих отбор признаков.

Заключение

В данной работе рассмотрена проблема построения прогноза временных рядов. Подход, предлагаемый для ее решения подразумевает исключение из рассмотрения некоторых входных данных для улучшения качества прогноза, то есть отбор признаков. На устойчивость прогноза сильно влияет наличие мультиколлинеарных признаков, для обнаружения которых применяется метод Белсли. В ходе эксперимента были оценены результаты работы предлагаемого метода. Так же произведено сравнение качества моделей, полученных с помощью предлагаемого метода Add-Del, и алгоритмов LARS и Lasso. Предлагаемый метод получает модель, включающую меньшее количество признаков, но предоставляющую при этом сравнимые по качеству результаты. Из теоретического описания метода следует, что область его применения не ограничивается задачами прогнозирования, и он может быть использован в любых задачах, требующих отбора признаков при построении модели по признаковому описанию.

Литература

- [1] Леонтьева, Л. Н. Выбор моделей прогнозирования цен на электроэнергию / Л. Н. Леонтьева // *JMLDA*. — 2011. — Vol. 1, no. 2. — Pp. 127–137.
- [2] Крымова Е. А., С. В. В. Алгоритм выбора признаков линейных регрессионных моделей из конечного и счетного множеств / С. В. В. Крымова Е. А. // *Заводская лаборатория. Диагностика материалов*. — 2011. — Vol. 77, no. 5. — Pp. 63–68.
- [3] J.B. Elsner, A. T. Singular Spectrum Analysis. A New Tool in Time Series Analysis / A. T. J.B. Elsner. — Springer US, 1996. — 164 p.
- [4] A. N. Gorban N. R. Sumner, A. Y. Z. Topological grammars for data approximation / A. Y. Z. A. N. Gorban, N. R. Sumner // *Applied Mathematics Letters*. — 2007. — 4. — Vol. 20, no. 4. — P. 382–386.
- [5] Il-Gyo Chong, C.-H. J. Performance of some variable selection methods when multicollinearity is present. / C.-H. J. Il-Gyo Chong // *Chemometrics and Intelligent Laboratory Systems*. — 2005. — Vol. 78, no. 1-2. — Pp. 103–112.
- [6] Hansheng Wang Guodong Li, G. J. Robust regression shrinkage and consistent variable selection through the lad-lasso / G. J. Hansheng Wang, Guodong Li // *Journal of Business Economics Statistics*. — 2008. — Vol. 25. — Pp. 347–355.
- [7] Marcus Hildmann, F. H. Robust calculation and parameter estimation of the hourly price forward curve / F. H. Marcus Hildmann / 17th Power Systems Computation Conference. — Stockholm Sweden: 2011. — August. — Pp. 1–7.
- [8] Tibshirani, B. E. T. H. I. J. R. Least angle regression / B. E. T. H. I. J. R. Tibshirani // *The Annals of Statistics*. — 2004. — Vol. 32, no. 3. — Pp. 407–499.
- [9] А. Г. Ивахненко, В. С. С. Помехоустойчивость моделирования. / В. С. С. А. Г. Ивахненко. — Киев: Наукова думка, 1985. — 216 с.
- [10] N. R. Draper, H. S. Applied regression analysis / H. S. N. R. Draper. — 3 edition. — John Wiley and Sons, 1998. — May. — 736 p.
- [11] S. Chen C. F. N Cowan, P. M. G. Orthogonal least squares learning algorithm for radial basis function network / P. M. G. S. Chen, C. F. N Cowan // *Neural Networks*. — 1991. — Vol. 2, no. 2. — Pp. 302–309.
- [12] Belsley, D. A. Conditioning Diagnostics: Collinearity and Weak Data in Regression / D. A. Belsley. — New York: John Wiley and Sons, 1991. — 396 p.
- [13] Abdolkhalig, A. Optimized calculation of hourly price forward curve (hpfc) / A. Abdolkhalig // *International Journal of Electrical, Computer, Electronics and Communication Engineering*. — 2008. — Vol. 2, no. 9. — Pp. 840–850.
- [14] Marcus Hildmann, G. C. What makes a good hourly price forward curve? / G. C. Marcus Hildmann // *European Energy Market / 10th International Conference*. — Stockholm: IEEE, 2013. — May. — Pp. 1–7.
- [15] I. Kachapov, F. K. Orthogonal projection in teaching regression and financial mathematics / F. K. I. Kachapov // *Journal of Statistics Education*. — 2010. — Vol. 18, no. 1. — Pp. 1–18.
- [16] Временной ряд цен на электроэнергию: <https://svn.code.sf.net/p/dmba/code/data/germanspotprice.csv>.