

Выбор признаков в задаче авторегрессионного прогнозирования биомедицинских сигналов*

В. Г. Новицкий, А. М. Катруца, В. В. Стрижов

В работе исследуется задача прогнозирования биомедицинских сигналов. Требуется построить устойчивую модель прогнозирования сигнала, которая на основе исторических данных предскажет значения сигнала для заданного горизонта прогнозирования. В работе используется авторегрессионная модель прогнозирования сигнала. Однако такая модель является неустойчивой из-за наличия мультиколлинеарности в исторических данных сигнала. Для повышения устойчивости модели предлагается использовать методы отбора признаков: метод частных наименьших квадратов (PLS) и метод, основанный на квадратичном программировании (QPFS). Результатом применения методов отбора признаков является снижение размерности задачи и повышение устойчивости модели без существенной потери точности прогноза. В работе проведено сравнение рассматриваемых методов отбора признаков. Также проведен вычислительный эксперимент на выборке полисомнографических сигналов Santa Fe Institute: электроэнцефалограмма, электрокардиограмма, электромиограмма, дыхательные движения, кровяное давление.

Ключевые слова: *прогнозирование временных рядов; отбор признаков; мультиколлинеарность; устойчивость модели; метод частных наименьших квадратов; PLS; квадратичное программирование; полисомнография*

1 Введение

В работе решается задача прогнозирования биомедицинских сигналов на основе исторических данных. Для решения этой задачи используется авторегрессионная модель, то есть предполагается, что значение сигнала в данный момент времени линейно зависит от предыдущих значений этого же сигнала. Авторегрессионная модель является неустойчивой из-за наличия мультиколлинеарности в исторических данных. Для решения этой проблемы необходимо использовать методы отбора признаков [1], в результате чего повышается устойчивость модели без существенного снижения качества прогноза.

Ранее были предложены следующие алгоритмы отбора признаков: полный перебор подмножеств признаков [1], генетический алгоритм [2], шаговая регрессия [3], последовательное добавление признаков с ортогонализацией FOS [4, 5], алгоритмы, основанные на регуляризации, такие как LASSO [6], LARS [7], Elastic Net [8]. Также был предложен метод главных компонент [9].

В работе исследуются два метода отбора признаков: метод частных наименьших квадратов (PLS) [10] и метод квадратичной оптимизации (QPFS) [11]. Метод частных наименьших квадратов основан на снижении размерности матрицы признаков и выделяет линейные комбинации признаков, которые больше всего влияют на вектор ответов. Выделение признаков происходит итеративно, в порядке уменьшения их влияния на вектор ответов [10]. Поэтому можно рассматривать только самые значимые комбинации, незначительно потеряв в точности прогноза.

Отбор признаков методом квадратичной оптимизации основан на решении задачи квадратичного программирования. В качестве целевой функции выступает квадратичная форма. Каждый элемент матрицы квадратичного члена – коэффициент корреляции

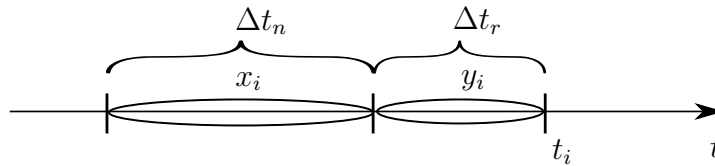
*Работа выполнена при частичной финансовой поддержке РФФИ, проекты № №00-00-00000 и 00-00-00001.

между соответствующими признаками или любая другая функция близости между признаками, гарантирующая положительную определённую матрицу квадратичного члена. Элементы линейного члена – коэффициенты корреляции между признаком и объектом или любые другие меры релевантности признаков целевому вектору. Аргументом задачи оптимизации является бинарный вектор, каждый элемент которого показывает входит ли соответствующий признак в подмножество отобранных признаков. Заменяя дискретное допустимое множество его выпуклой оболочкой, получаем задачу выпуклой оптимизации, которая может быть эффективно решена. Если компонента оптимального решения задачи выпуклой оптимизации, превышает некоторое пороговое значение, то соответствующий признак будет отобран.

В работе проведено сравнение двух методов отбора признаков в задаче авторегрессионного прогнозирования биомедицинских сигналов: метода частных наименьших квадратов и метода, основанного на квадратичном программировании. Сравнение двух методов проводится на выборке сигналов из полисомнографической базы данных Santa Fe Institute [12–14]. База содержит биомедицинские сигналы, такие как электроэнцефалограмма, электрокардиограмма, электромиограмма, дыхательные движения, кровяное давление. Результатом применения отбора признаков является снижение размерности задачи и повышение устойчивости модели без существенной потери точности прогноза.

2 Постановка задачи прогнозирования

Для решения задачи прогнозирования сигнала предлагается использовать авторегрессионную модель. Рассмотрим сигнал $\mathbf{x} = [x_t], t = 1, \dots, n, x_t \in \mathbb{R}$. Необходимо спрогнозировать следующие r значений сигнала: $\mathbf{z} = [y_t], t = 1, \dots, r, y_t \in \mathbb{R}$. Предполагается, что $r < n$.



Предположим, что сигнал обладает следующими свойствами:

- его значения получены через одинаковые промежутки времени
- у него нет пропущенных значений
- сигнал имеет период $\tau \geq r$.

В авторегрессионной модели предполагается, что значение сигнала в данный момент времени линейно зависит от предыдущих значений этого же сигнала. В данной модели признаками являются предыдущие значения прогнозируемого сигнала.

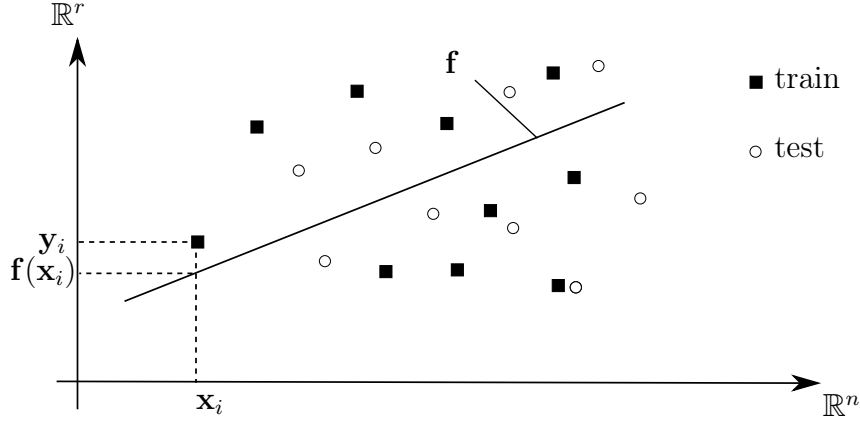
Пусть $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ – матрица плана, $\mathbf{x}_j \in \mathbb{R}^m$ – j -ый признак, $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_r]$ – матрица ответов, $\mathbf{y}_j \in \mathbb{R}^m$. Каждая строка \mathbf{x}_i матрицы \mathbf{X} – локальная история сигнала (n значений сигнала, начиная с момента i). Каждая строка \mathbf{z}_i матрицы \mathbf{Y} – локальный прогноз сигнала (r значений сигнала, начиная с момента $n + i$).

$$[\mathbf{X} | \mathbf{Y}] = \left[\begin{array}{cccc|cccc} x_1 & x_2 & \dots & x_n & y_1 & y_2 & \dots & y_r \\ x_2 & x_3 & \dots & x_{n+1} & y_2 & y_3 & \dots & y_{r+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_m & x_{m+1} & \dots & x_{m+n-1} & y_m & y_{m+1} & \dots & y_{r+m-1} \end{array} \right] = \left[\begin{array}{c|c} \mathbf{x}_1 & \mathbf{z}_1 \\ \mathbf{x}_2 & \mathbf{z}_2 \\ \vdots & \vdots \\ \mathbf{x}_m & \mathbf{z}_m \end{array} \right].$$

В авторегрессионной модели матрица ответов представляется в виде

$$\mathbf{Y} = \mathbf{f}(\mathbf{X}, \mathbf{W}) + \varepsilon(\mathbf{X}),$$

где $\mathbf{f}(\mathbf{X}, \mathbf{W}) = \mathbf{XW}$ – модель, $\mathbf{W} \in \mathbb{R}^{n \times r}$ – матрица параметров модели, а $\varepsilon(\mathbf{X})$ – вектор регрессионных остатков.



Введем функцию ошибки S :

$$S(\mathbf{W}|\mathcal{D}) = \|\mathbf{f}(\mathbf{X}, \mathbf{W}) - \mathbf{Y}\|_2^2 = \|\mathbf{XW} - \mathbf{Y}\|_2^2.$$

Пусть $\mathcal{I} = \{1, \dots, m\}$ – множество индексов элементов выборки. Разобьем выборку $\mathcal{D} = (\mathbf{X}, \mathbf{Y})$ на обучающую $\mathcal{D}_{\mathcal{L}}$ и контрольную $\mathcal{D}_{\mathcal{C}}$ ($\mathcal{I} = \mathcal{L} \sqcup \mathcal{C}$).

Пусть $\mathbf{x} \in \mathbb{R}^n$ – история сигнала по настоящее время. Прогноз сигнала на горизонте r : $\mathbf{y} = \mathbf{xW}^*$, где \mathbf{W}^* – решение задачи авторегрессии:

$$\mathbf{W}^* = \arg \min_{\mathbf{W} \in \mathbb{R}^{n \times r}} S(\mathbf{W}|\mathcal{D}_{\mathcal{L}}). \quad (1)$$

Значения биомедицинского сигнала в соседние моменты времени не являются независимыми. Таким образом, наблюдается мультиколлинеарность между признаками в авторегрессионной модели прогнозирования сигнала. Для решения проблемы мультиколлинеарности применяются методы отбора признаков. Отбор признаков происходит на контрольной выборке $\mathcal{D}_{\mathcal{C}}$. В следующем разделе ставится задача отбора признаков.

3 Постановка задачи отбора признаков

Пусть $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ – матрица плана, $\mathbf{x}_j \in \mathbb{R}^m$ – j -ый признак. $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_r]$ – матрица ответов, $\mathbf{y}_j \in \mathbb{R}^m$, $\mathcal{D} = (\mathbf{X}, \mathbf{Y})$ – выборка.

Пусть $\mathcal{J} = \{1, \dots, n\}$ – множество индексов признаков.

Определение 1. Признаки из индексного подмножества $\mathcal{A} \subseteq \mathcal{J}$ – называются мультиколлинеарными, если существуют $j \in \mathcal{A}$, коэффициенты $\lambda_k \in \mathbb{R}$ и достаточно малое $\delta > 0$, такие что:

$$\left\| \mathbf{x}_j - \sum_{k \in \mathcal{A} \setminus j} \lambda_k \mathbf{x}_k \right\|_2^2 < \delta.$$

Определение 2. Признаки, которым соответствуют ненулевые строки матрицы \mathbf{W} , называются *активными*.

Введем обозначения: \mathcal{A} – множество всех индексов активных признаков, $\mathfrak{D}^{\mathcal{A}} = ([\chi_j, j \in \mathcal{A}], \mathbf{Y})$. Сформулируем задачу отбора признаков:

$$\mathcal{A}^* = \arg \min_{\mathcal{A} \in \mathcal{J}} \mathcal{Q}(\mathcal{A} | \mathfrak{D}_c), \quad (2)$$

где $\mathcal{Q} : \mathcal{A} \rightarrow \mathbb{R}$ – критерий качества подмножества признаков.

4 Метод частных наименьших квадратов

Идея метода частных наименьших квадратов состоит в том, чтобы перейти в пространство меньшей размерности с сохранением ковариации между признаками и ответами. Матрица плана $\mathbf{X} \in \mathbb{R}^{m \times n}$ и матрица ответов $\mathbf{Y} \in \mathbb{R}^{m \times r}$ проецируются на пространство меньшей размерности $m \times l$ так, что ковариация между новыми признаками и ответами максимальна:

$$\begin{aligned} \mathbf{X} &= \mathbf{T}\mathbf{P}^T + \mathbf{E} \\ \mathbf{Y} &= \mathbf{U}\mathbf{Q}^T + \mathbf{F} \end{aligned}$$

$\mathbf{T} \in \mathbb{R}^{m \times l}$, $\mathbf{U} \in \mathbb{R}^{m \times l}$ – значения признаков и ответов в пространстве, на которое они были спроектированы. $\mathbf{P} \in \mathbb{R}^{n \times l}$, $\mathbf{Q} \in \mathbb{R}^{r \times l}$ – матрицы перехода из нового пространства в старое. \mathbf{E} , \mathbf{F} – матрицы невязок. Так как ковариация между новыми признаками и ответами максимальна, то мы можем строить регрессионную модель в пространстве меньшей размерности с сохранением точности прогноза. Параметр метода частных наименьших квадратов $l \in \mathbb{N}$ определяет размерность этого пространства. Отбор признаков осуществляется в виде замены исходных признаков $[\chi_1, \dots, \chi_n]$ на l новых признаков – линейные комбинации исходных признаков.

Алгоритм:

```

Вход:  $\mathbf{X}, \mathbf{Y}, l$ ;
Выход:  $\mathbf{T}, \mathbf{U}, \mathbf{P}, \mathbf{Q}$ ;
1: для  $i = 1, \dots, l$ 
2:    $\mathbf{u} := \mathbf{y}_1$ 
3:   повторять
4:      $\mathbf{p} := \frac{\mathbf{X}^T \mathbf{u}}{\|\mathbf{X}^T \mathbf{u}\|_2}$ 
5:      $\mathbf{t} := \mathbf{X} \mathbf{p}$ 
6:      $\mathbf{q} := \frac{\mathbf{Y}^T \mathbf{t}}{\|\mathbf{Y}^T \mathbf{t}\|_2}$ 
7:      $\mathbf{u} := \mathbf{Y} \mathbf{q}$ 
8:   пока  $\mathbf{t}$  не перестанет меняться
9:   сохранить  $\mathbf{t}, \mathbf{u}, \mathbf{p}, \mathbf{q}$ 
10:   $\mathbf{X} := \mathbf{X} - \mathbf{t} \mathbf{p}^T$ 
11:   $\mathbf{Y} := \mathbf{Y} - \mathbf{u} \mathbf{q}^T$ 

```

Примечание к алгоритму: i -ый столбец матриц \mathbf{T} , \mathbf{U} , \mathbf{P} , \mathbf{Q} – вектор \mathbf{t} , \mathbf{u} , \mathbf{p} , \mathbf{q} , полученный на i -ом шаге алгоритма ($i = 1, \dots, l$).

В [10] показано, что предельное значение \mathbf{p} во внутреннем цикле максимизирует ковариацию между новыми признаками и матрицей ответов:

$$\mathbf{p} = \arg \max_{\|\mathbf{p}\|=1} \text{Var}(\mathbf{X}\mathbf{p}) (\text{Corr}(\mathbf{Y}, \mathbf{X}\mathbf{p}))^T (\text{Corr}(\mathbf{Y}, \mathbf{X}\mathbf{p})) \quad (3)$$

5 Метод квадратичной оптимизации

Идея метода квадратичной оптимизации – минимизировать число похожих признаков и максимизировать число значимых признаков. Пусть $\mathbf{a} \in \mathbb{B}^n$ – вектор, такой что $a_j = 1$, если и только если $j \in \mathcal{A}$. Переформулируем задачу (2):

$$\mathbf{a}^* = \arg \min_{\mathbf{a} \in \mathbb{B}^n} \mathcal{Q}(\mathbf{a} | \mathcal{D}_c). \quad (4)$$

Вектор \mathbf{a}^* и индексное множество \mathcal{A}^* эквивалентны в следующем смысле:

$$a_j^* = 1 \Leftrightarrow j \in \mathcal{A}^*.$$

В качестве минимизируемого функционала качества $\mathcal{Q}(\mathbf{a})$ предлагается использовать квадратичную форму:

$$\mathcal{Q}(\mathbf{a}) = \mathbf{a}^\top \Sigma \mathbf{a} - \mathbf{b}^\top \mathbf{a},$$

где $\Sigma = [q_{ij}]$, q_{ij} – коэффициент корреляции Пирсона между χ_i и χ_j : $q_{ij} = \frac{\text{cov}(\chi_i, \chi_j)}{\sqrt{\text{Var}(\chi_i) \text{Var}(\chi_j)}}$, b_i – коэффициент корреляции Пирсона между χ_i и \mathbf{y} : $b_i = \frac{\text{cov}(\chi_i, \mathbf{y})}{\sqrt{\text{Var}(\chi_i) \text{Var}(\mathbf{y})}}$

Множество отобранных признаков определяется решением следующей оптимизационной задачи:

$$\mathbf{a}^* = \arg \min_{\mathbf{a} \in \mathbb{B}^n} \mathbf{a}^\top \Sigma \mathbf{a} - \mathbf{b}^\top \mathbf{a}. \quad (5)$$

Эта задача является NP-трудной, так как допустимое множество \mathbb{B}^n дискретное. Аппроксимируем задачу (5) следующей задачей:

$$\mathbf{a}^* = \arg \min_{\mathbf{a} \in [0,1]^n} \mathbf{a}^\top \Sigma \mathbf{a} - \mathbf{b}^\top \mathbf{a}. \quad (6)$$

Дискретное допустимое множество \mathbb{B}^n было заменено его выпуклой оболочкой $[0,1]^n$. Задача (6) – задача квадратичного программирования, которая может быть эффективно решена. Чтобы вернуться к дискретному вектору и осуществить отбор признаков, будем использовать *порог значимости* τ .

Определение 3. Порог значимости τ – неотрицательное число, такое что $j \in \mathcal{A}^* \Leftrightarrow a_j^* > \tau$.

Число отобранных признаков – количество индексов j , таких что $a_j^* > \tau$. Увеличивая τ , мы уменьшаем число отобранных признаков. Порог значимости τ – параметр метода.

6 Вычислительный эксперимент

В рамках вычислительного эксперимента проводится построение прогноза сигнала энергопотребления. В ходе эксперимента сравниваются методы PLS и QPFS. Сравнение проведение на реальных данных объемов потребления электроэнергии в Польше за 12 лет [15]. Для иллюстрации суточной и недельной периодичности на рис. 1 приведен график зависимости энергопотребления за 1 месяц. Для иллюстрации годовой периодичности на рис. 2 приведен график зависимости энергопотребления за 3 года.

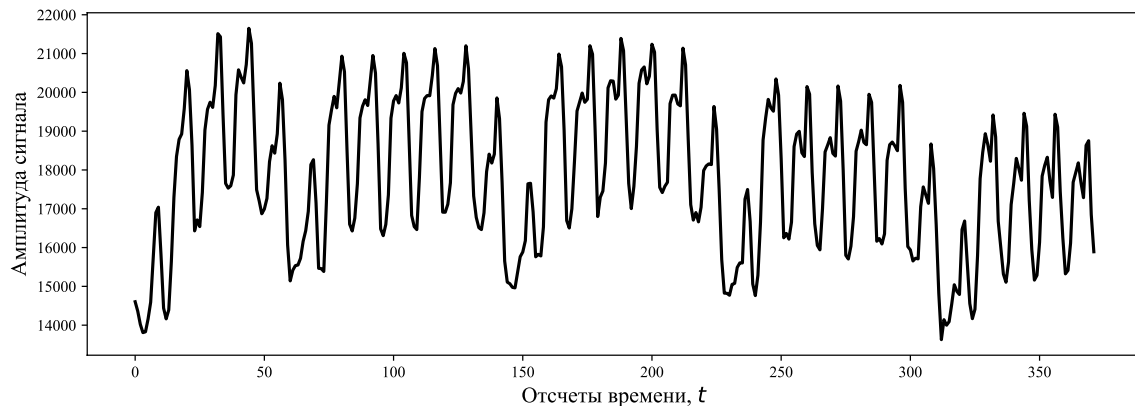


Рис. 1 График зависимости зависимости энергопотребления за 1 месяц

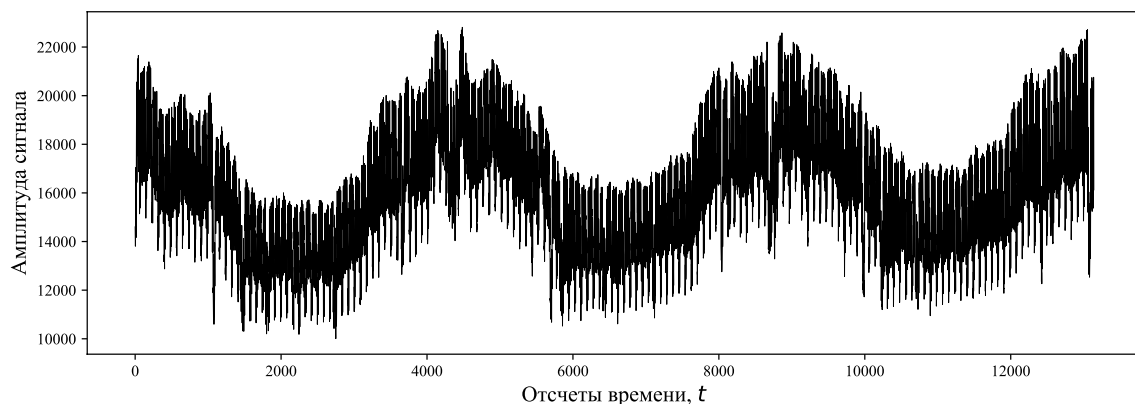


Рис. 2 График зависимости зависимости энергопотребления за 3 года

Сигнал представляет собой объем энергопотребления за 2 часа, в сутках 12 значений сигнала. Решается задача прогнозирования на 1 день. В качестве признаков были выбраны значения сигнала за предшествующую неделю, а также значения сигнала в день за неделю до текущего и за год до текущего (всего $(7 + 2) \times 12$ значений). Обучающая выборка содержит данные за 1000 дней, тестовая выборка – за 2000 дней. Отбор признаков осуществлялся на обучающей выборке, контроль качества отбора признаков производился на тестовой выборке.

Пусть \mathbf{X}_A – матрица отобранных признаков (в ней из матрицы \mathbf{X} из контрольной выборки оставлены только столбцы с индексами из множества A). Матрица \mathbf{Y}_A – в случае PLS совпадает с матрицей \mathbf{U} – проекциями значений матрицы \mathbf{Y} из контрольной выборки на пространство меньшей размерности, в случае QPFS является столбцом матрицы \mathbf{Y} . Определим $r = S(\mathbf{W}^* | (\mathbf{X}, \mathbf{Y}))$, $r_A = S(\mathbf{W}^* | (\mathbf{X}_A, \mathbf{Y}_A))$, где \mathbf{W}^* – решение задачи авторегрессии (1).

Для оценки качества отбора признаков, осуществленного PLS и QPFS, используются следующие критерии. Критерий Маллоуза C_p определяется как баланс между ошибкой

$r_{\mathcal{A}}$ и количеством отобранных признаков $|\mathcal{A}|$:

$$C_p = \frac{r_{\mathcal{A}}}{r} - n + 2|\mathcal{A}|.$$

Меньшие значения критерия C_p соответствуют лучшему набору признаков.

Критерий устойчивости \varkappa определяется как логарифм числа обусловленности матрицы $\mathbf{X}_{\mathcal{A}}^T \mathbf{X}_{\mathcal{A}}$:

$$\varkappa = \log \frac{|\lambda|_{\max}}{|\lambda|_{\min}},$$

где λ – собственные числа матрицы $\mathbf{X}_{\mathcal{A}}^T \mathbf{X}_{\mathcal{A}}$. Чем ближе \varkappa к нулю, тем больше устойчивость модели.

Определим фактор инфляции дисперсии j -го признака

$$\text{VIF}_j = \frac{1}{1 - R_j^2},$$

где $R_j^2 = 1 - \frac{\text{RSS}_j}{\text{TSS}_j}$ – коэффициент детерминации j -го признака относительно остальных,

$$\text{RSS}_j = \|\boldsymbol{\chi}_j - \mathbf{X}_{\mathcal{A} \setminus j} \mathbf{w}^*\|_2^2, \text{ где } \mathbf{w}^* = \arg \min_{\mathbf{w} \in \mathbb{R}^{|\mathcal{A}|-1}} S(\mathbf{w} | (\mathbf{X}_{\mathcal{A} \setminus j}, \boldsymbol{\chi}_j)),$$

$$\text{TSS}_j = \sum_{i=1}^m (\chi_{ji} - \bar{\chi}_j), \text{ где } \bar{\chi}_j = \frac{1}{m} \sum_{i=1}^m \chi_{ji}.$$

Фактор VIF_j показывает наличие линейной зависимости j -го признака от других признаков из множества \mathcal{A} . Если значение VIF_j велико, то R_j^2 близко к 1, что показывает почти линейную зависимость j -го признака от остальных. Критерий VIF определяется по формуле $\text{VIF} = \max_{j \in \mathcal{A}} \text{VIF}_j$ и определяет наличие мультиколлинеарности.

Критерий SMAPE определяется как

$$\text{SMAPE} = \frac{2}{p} \sum_{i,j} \frac{|f_{ij}(\mathbf{X}_{\mathcal{A}}, \mathbf{W}^*) - (\mathbf{Y}_{\mathcal{A}})_{ij}|}{|f_{ij}(\mathbf{X}_{\mathcal{A}}, \mathbf{W}^*)| + |(\mathbf{Y}_{\mathcal{A}})_{ij}|},$$

где суммирование ведется по всем элементам матрицы $\mathbf{Y}_{\mathcal{A}}$, а p – число элементов матрицы $\mathbf{Y}_{\mathcal{A}}$. Критерий показывает среднюю ошибку модели на контрольной выборке.

Результаты сравнения методов отбора признаков приведены в таблице 1.

QP			
$ \mathcal{A} $	R	VIF	SMAPE
2	5.62	0.09	0.038
3	7.12	1.43	0.041
4	8.12	4.83	0.045
5	10.07	1.75	0.047
10	inf	4.94	0.050
PLS			
$ \mathcal{A} $	R	VIF	SMAPE
2	1.52	1.005	0.061
3	1.96	1.005	0.055
4	2.01	1.006	0.051
5	2.80	1.009	0.050
10	4.09	1.05	0.051
13	5.10	1.09	0.039
30	8.013	1.70	0.035

Таблица 1 Сравнение методов отбора признаков

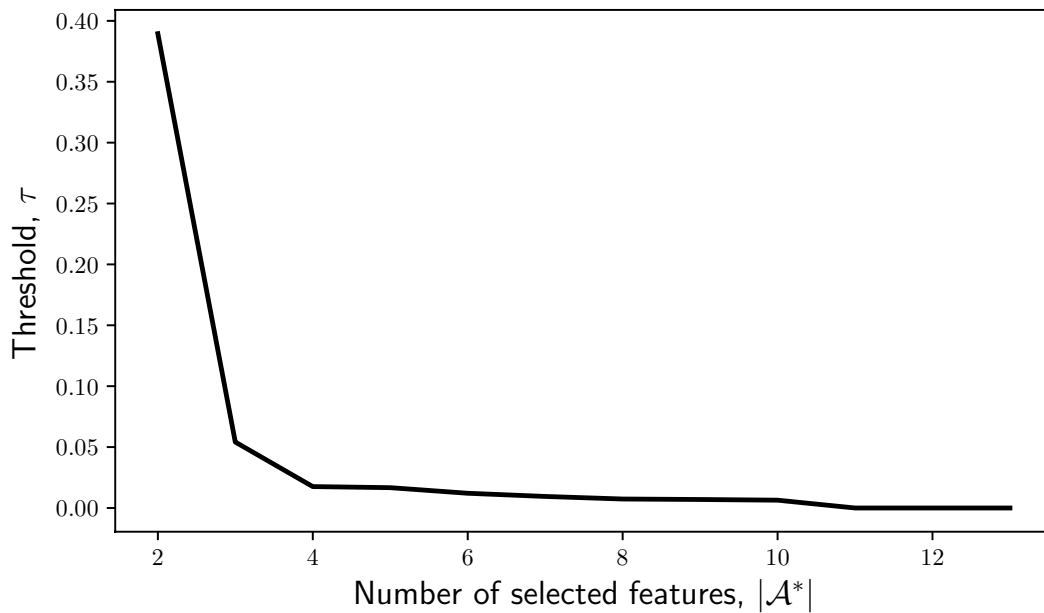


Рис. 3 Зависимость порога значимости τ от числа отобранных признаков в QPFS

-4377	0.095
-4373	0.010
-81	0.200
-69	0.080
-57	0.076
-45	0.083
-33	0.101
-21	0.058
-19	0.015
-9	0.227
-5	0.017
-1	0.038

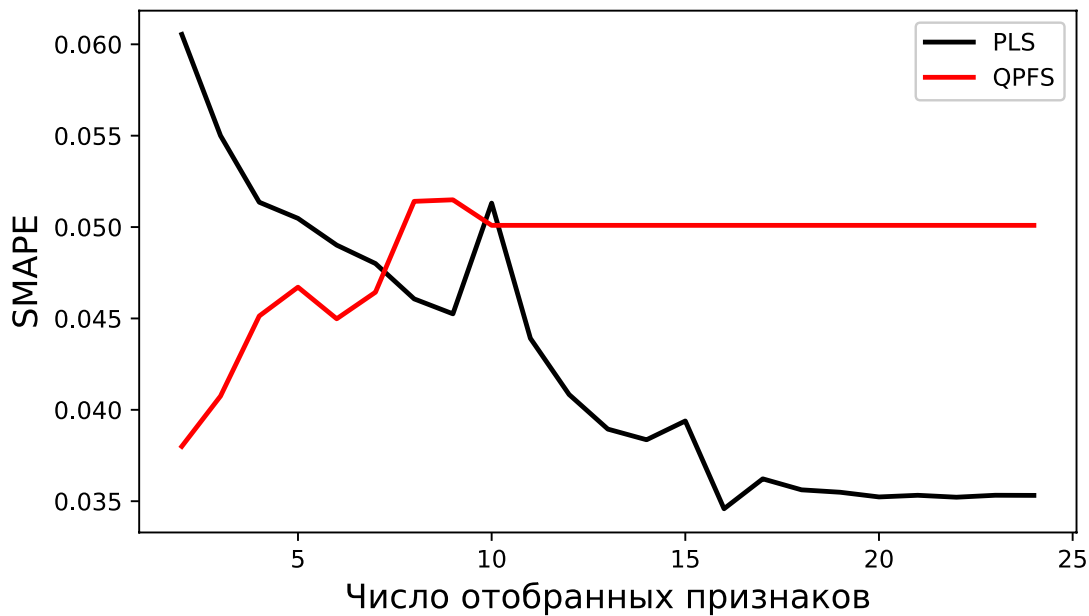


Рис. 4 График зависимости SMAPE от числа отобранных признаков

7 Заключение

В работе рассмотрена проблема мультиколлинеарности в авторегрессионной модели прогнозирования сигналов. Предлагается осуществить отбор признаков данной модели, в результате чего снижается размерность задачи и повышается устойчивость модели. В работе проведено сравнение двух методов отбора признаков: метод частных наименьших квадратов (PLS) и метод, основанный на квадратичном программировании (QPFS).

Литература

- [1] Chong Il-Gyo, Jun Chi-Hyuck. Performance of some variable selection methods when multicollinearity is present // Chemometrics and Intelligent Laboratory Systems, 2005. Vol. 78. No. 1. P. 103–112.

- [2] *Holland JH, Goldberg D.* Genetic algorithms in search, optimization and machine learning // ed: Addison-Wesley, Reading, MA, 1989.
- [3] *Draper Norman R, Smith Harry.* Fitting a straight line by least squares // Applied Regression Analysis, Third Edition, 1998. P. 15–46.
- [4] *Chen Sheng, Billings Stephen A, Luo Wan.* Orthogonal least squares methods and their application to non-linear system identification // International Journal of control, 1989. Vol. 50. No. 5. P. 1873–1896.
- [5] *Chen Sheng, Cowan Colin FN, Grant Peter M.* Orthogonal least squares learning algorithm for radial basis function networks // IEEE Transactions on neural networks, 1991. Vol. 2. No. 2. P. 302–309.
- [6] *Tibshirani Robert@articletibshirani1996regression, title=Regression shrinkage and selection via the lasso, author=Tibshirani, Robert, journal=Journal of the Royal Statistical Society. Series B (Methodological), pages=267–288, year=1996, publisher=JSTOR .* Regression shrinkage and selection via the lasso // Journal of the Royal Statistical Society. Series B (Methodological), 1996. P. 267–288.
- [7] *Efron Bradley, Hastie Trevor, Johnstone Iain, Tibshirani Robert et al.* Least angle regression // The Annals of statistics, 2004. Vol. 32. No. 2. P. 407–499.
- [8] *Zou Hui, Hastie Trevor.* Regularization and variable selection via the elastic net // Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2005. Vol. 67. No. 2. P. 301–320.
- [9] *Jolliffe Ian.* Principal component analysis. — 2002.
- [10] *Ng Kee Siong.* A simple explanation of partial least squares. — 2013.
- [11] *Katrutsa Alexandr, Strijov Vadim.* Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria // Expert Systems with Applications, 2017. Vol. 76. P. 1–11.
- [12] *Goldberger Ary L, Amaral Luis AN, Glass Leon, Hausdorff Jeffrey M, Ivanov Plamen Ch et al.* Physiobank, physiotoolkit, and physionet // Circulation, 2000. Vol. 101. No. 23. P. e215–e220.
- [13] *Ichimaru Y, Moody GB.* Development of the polysomnographic database on cd-rom // Psychiatry and Clinical Neurosciences, 1999. Vol. 53. No. 2. P. 175–177.
- [14] *Rigney David R, Goldberger Ary L, Ocasio W, Ichimaru Yuhei, Moody George B, Mark R.* Multi-channel physiological data: Description and analysis // Time Series Prediction: Forecasting The Future And Understanding The Past, 1993.
- [15] URL: <http://gdudek.el.pcz.pl/varia/stlf-data>.

Received