

Снижение размерности пространства зависимой переменной в задачах прогнозирования*

Мария Владимировна, Роман Исаченко

vladimirova.maria@phystech.edu, isa-ro@yandex.ru

Московский физико-технический институт

В работе решается задача обнаружения зависимостей в прогнозируемой переменной. Используется набор гомогенных моделей, восстанавливающих прогноз по общему для всех переменных описанию объектов. Рассматривается линейная модель метода частных наименьших квадратов и ее предложенная нелинейная модификация. Находятся оптимальные параметрические преобразования исходных пространств объектов и ответов. Проводится вычислительный эксперимент на реальных данных объемов потребления электроэнергии и данных сигналов кортикограмм.

Ключевые слова: прогнозирование временных рядов; мультиколлинеарность; метод частных наименьших квадратов; PLS; нелинейный PLS

1. Введение

В работе рассматривается задача прогнозирования временных рядов в случае наличия мультиколлинеарности в данных. Методы решения данной задачи сравниваются на двух наборах данных, имеющих избыточную информацию.

Первый набор данных представляет собой временные ряды объема потребления электроэнергии в Варшаве. Электрическая энергия является важной движущей силой экономического развития, а точность прогнозов спроса является важным фактором, который ведет к успешному эффективному планированию. По этой причине энергетическим анализам необходимо руководство для лучшего выбора наиболее подходящих методов прогнозирования, чтобы обеспечить точные прогнозы тенденций потребления электроэнергии. Предполагается, что значение сигнала в данный момент времени линейно зависит от предыдущих значений этого же сигнала, поэтому данные являются мультиколлинеарными.

Второй набор данных взят из проекта Project Tycho, в котором изучалась проблема проектирования нейро-компьютерного интерфейса (BCI) для обмена информацией между мозгом и электронным устройством. Решается задача выбора функций в моделях регрессии в приложении к декодированию движения на основе электрокардиограмм (ECoG). Проблема состоит в том, чтобы предсказать траектории руки из временных рядов напряжения кортикальной активности. Описание функции каждой точки находится в пространственно-временной частотной области включает в себя сами временные ряды напряжения и их спектральные характеристики. Выбор функции имеет решающее значение для адекватного решения проблемы регрессии, поскольку электрокортикальные данные являются высокомерными и измерения коррелируют как во временной, так и в пространственной областях.

Система BCI улучшает умственные и физические возможности пользователя, обеспечивая прямую связь между мозгом и компьютером. BCI направлены на восстановление поврежденных функциональных возможностей пациентов с механическими или когнитивными нарушениями. В данной статье предлагается новый метод выбора признаков

29 в прогнозировании движения и его реконструкции. Первый шаг к прогнозированию
30 предполагаемых движений — научиться реконструировать фактические перемещения из
31 кортикальной активности. Рассматривается проблема непрерывной реконструкции траек-
32 тории. Субдуральные сигналы ECoG измеряются через 32 или 64 канала, когда субъект
33 перемещает руку. Когда сигналы ECoG трансформируются в информационные функции,
34 проблема восстановления траектории является проблемой регрессии. Извлечение функ-
35 ции включает в себя применение некоторого спектрально-временного преобразования к
36 сигналам ECoG с каждого канала. Так как результирующее пространственно-временное
37 спектральное представление сильно избыточно, используются различные методы выбора
38 объектов и уменьшения размерности, чтобы извлечь только наиболее важные функции.

39 Для решения задачи прогнозирования используется авторегрессионная модель. Авто-
40 регрессионная модель является неустойчивой в случае наличия мультиколлинеарности в
41 исторических данных. Для решения этой проблемы необходимо используются методы от-
42 бора признаков [21], в результате чего повышается устойчивость модели без существенного
43 снижения качества прогноза.

44 В работе исследуются методы отбора признаков: метод частных наименьших квад-
45 ратов (PLS) [11] и предложенная его нелинейная модификация (cnlPLS). Метод частных
46 наименьших квадратов основан на снижении размерности матрицы признаков и выделя-
47 ет линейные комбинации признаков, которые оказывают наибольшее влияние на вектор
48 ответов. Выделение признаков происходит итеративно, в порядке уменьшения их влия-
49 ния на вектор ответов [11]. Рассматриваются только значимые комбинации признаков,
50 незначительно потеряв в точности прогноза.

51 Методы PLS регрессии подробно описаны в работах [8,9]. Разницу между методом PLS
52 и связанными с ним подходами, различные разновидности регрессии PLS можно найти
53 в [17].

54 Нелинейное расширение метода PLS регрессии впервые введено в [6]. В литературе
55 были разработаны различные модификации PLS. Предложены нелинейные методы PLS,
56 основанные на различных моделях: искусственных нейронных сетей [4], функции актива-
57 ции радиальных оснований [5], логистическая функция активации и методы оптимизации
58 роевых частиц [7], используют прямые нейронные сети [3], искусственную нейронную сеть
59 Эльмана [10].

60 Предлагается провести модификацию алгоритма PLS: совершить криволинейное и
61 нелинейное преобразования пространства целевой переменной для учета зависимостей
62 между сигналами в разные моменты времени.

63 В работе проведено сравнение двух методов отбора признаков в задаче авторегрессион-
64 ного прогнозирования сигналов (PLSR и cnlPLSR). Цель регрессии PLS [18] — предсказать
65 \mathbf{Y} по \mathbf{X} и описать их общую структуру. Когда \mathbf{Y} — вектор, а \mathbf{X} — матрица полного ран-
66 га, эта цель может быть выполнена с использованием обычной линейной регрессии. Если
67 число предикторов велико по сравнению с числом наблюдений, то \mathbf{X} будет сингулярной и
68 регрессионный подход в этом случае невозможен из-за наличия мультиколлинеарности.

69 В качестве практической проверки данных методов в ходе вычислительного экспери-
70 мента решается задача прогнозирования на реальных данных. Результатом применения
71 отбора признаков является снижение размерности задачи и повышение устойчивости мо-
72 делей без существенной потери точности прогноза.

Постановка задачи прогнозирования

Рассматривается сигнал $\mathbf{x} = [x_t]$, $t = 1, \dots, n$, $x_t \in \mathbb{R}$. Необходимо спрогнозировать r значений сигнала или отклика: $\mathbf{y} = [y_t]$, $t = 1, \dots, r$, $y_t \in \mathbb{R}$. Прогнозом могут являться, например, r последующих значений временного ряда. Предполагается, что $r < n$.

Предполагается, что сигнал обладает следующими свойствами:

- значения сигнала получены через одинаковые промежутки времени,
- в сигнале нет пропущенных значений,
- сигнал имеет период $\tau > r$.

Для решения задачи прогнозирования сигнала предлагается использовать авторегрессионную модель. В авторегрессионной модели признаками являются предыдущие значения прогнозируемого сигнала. Предполагается, что значение сигнала в данный момент времени линейно зависит от предыдущих значений этого же сигнала.

Пусть $\mathcal{D} = (\mathbf{X}, \mathbf{Y})$ — выборка, $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^m \in \mathbb{R}^{m \times n}$ — матрица объектов, $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^m \in \mathbb{R}^{m \times r}$ — матрица ответов. Каждая строка \mathbf{x}_i матрицы \mathbf{X} — локальная история сигнала (n значений сигнала, начиная с момента i). Каждая строка \mathbf{y}_i матрицы \mathbf{Y} — локальный прогноз (например, r значений сигнала, начиная с момента $n+1$). Матрицы \mathbf{X} и \mathbf{Y} имеют следующую структуру:

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

В авторегрессионной модели матрица ответов представляется в виде

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

где $f(\mathbf{X}, \boldsymbol{\Theta}) = \mathbf{X}\boldsymbol{\Theta}$ — линейная модель, $\boldsymbol{\Theta} \in \mathbb{R}^{n \times r}$ — матрица параметров модели, а $\varepsilon(\mathbf{X}) \in \mathbb{R}^{n \times r}$ — матрица регрессионных остатков.

Введем функцию ошибки S на выборке \mathcal{D} :

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

Разобьем выборку \mathcal{D} на обучающую $\mathcal{D}_{\mathcal{L}}$ и контрольную $\mathcal{D}_{\mathcal{C}}$, где $\mathcal{D}_{\mathcal{L}} \cap \mathcal{D}_{\mathcal{C}} = \emptyset$, $\mathcal{D}_{\mathcal{L}} \cup \mathcal{D}_{\mathcal{C}} = \mathcal{D}$.

Прогноз отклика сигнала на r следующих значений ряда — это ответ модели при найденных оптимальных параметрах задачи авторегрессии: $\mathbf{y} = \mathbf{x}\boldsymbol{\Theta}^*$, где

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

Одним из основных препятствий эффективного применения регрессионного анализа является мультиколлинеарность. Она связана с линейной зависимостью между признаками объектов. В результате мультиколлинеарности матрица парных коэффициентов корреляции и матрица $\mathbf{X}^T \mathbf{X}$ становятся слабообусловленными, т.е. их определители близки к нулю. Это приводит к неустойчивости оценок коэффициентов регрессии, так как в их выражения входит обратная матрица $(\mathbf{X}^T \mathbf{X})^{-1}$, получение которой связано с делением на определитель матрицы $\mathbf{X}^T \mathbf{X}$. Значения сигнала в соседние моменты времени являются зависимыми. Таким образом, наблюдается мультиколлинеарность между признаками

авторегрессионной модели прогнозирования сигнала. Для решения проблемы мультиколлинеарности применяются методы отбора признаков. Оптимальные параметры задачи регрессии Θ настраиваются на обучающей выборке \mathcal{D}_L . Отбор признаков происходит на контрольной выборке \mathcal{D}_C .

Метод частных наименьших квадратов (PLS)

Метод частных наименьших квадратов проецирует матрицу объектов \mathbf{X} и матрицу ответов \mathbf{Y} в латентное пространство меньшей размерности. В отличие от метода главных компонент, метод частных наименьших квадратов учитывает взаимосвязь между матрицами \mathbf{X} и \mathbf{Y} и снижает размерности обоих пространств.

Матрица плана \mathbf{X} и матрица ответов \mathbf{Y} проецируются на пространство меньшей размерности l следующим образом:

$$\begin{array}{ccccc} \boxed{\mathbf{X}} & = & \boxed{\mathbf{T}} & \times & \boxed{\mathbf{P}^\top} & + & \boxed{\mathbf{E}} \\ m \times n & & m \times l & & l \times n & & m \times n \\ \\ \boxed{\mathbf{Y}} & = & \boxed{\mathbf{U}} & \times & \boxed{\mathbf{Q}^\top} & + & \boxed{\mathbf{F}} \\ m \times r & & m \times l & & l \times r & & m \times r \end{array}$$

Рис. 1. Размерности векторов в алгоритме PLS

$$\begin{array}{ccccc} \mathbf{X} & = & \mathbf{T} & \mathbf{P}^\top & + & \mathbf{E} \\ m \times n & & m \times l & l \times n & & m \times n \end{array}, \quad (1)$$

$$\begin{array}{ccccc} \mathbf{Y} & = & \mathbf{U} & \mathbf{Q}^\top & + & \mathbf{F} \\ m \times r & & m \times l & l \times r & & m \times r \end{array}, \quad (2)$$

где \mathbf{T} , \mathbf{U} — матрицы объектов и ответов в спроектированном пространстве, причём $\mathbf{T}^\top \mathbf{T} = \mathbf{I}_l$; \mathbf{P} , \mathbf{Q} — матрицы перехода из нового пространства в старое; \mathbf{E} , \mathbf{F} — матрицы невязок.

Псевдокод метода регрессии PLS приведен в алгоритме 1. Алгоритм итеративно на каждом из l шагов вычисляет по одному столбцу \mathbf{t} , \mathbf{u} , \mathbf{p} , \mathbf{q} , из которых путем присоединения формируются матрицы \mathbf{T} , \mathbf{U} , \mathbf{P} , \mathbf{Q} соответственно. Предполагается, что вектора новых признаков \mathbf{t} и \mathbf{u} являются линейными комбинациями столбцов матриц \mathbf{X} и \mathbf{Y} соответственно.

Во внутреннем цикле алгоритма вычисляются вектора \mathbf{w} и \mathbf{c} . Правила обновления векторов \mathbf{w} , \mathbf{c} совпадают с итерацией Power method [20]. Вектора \mathbf{w} , \mathbf{c} являются собственными векторами матриц $\mathbf{X}^\top \mathbf{Y} \mathbf{Y}^\top \mathbf{X}$ и $\mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y}$ соответственно

$$\mathbf{w} \propto \mathbf{X}^\top \mathbf{u} = \mathbf{X}^\top \mathbf{Y} \mathbf{c} = \mathbf{X}^\top \mathbf{Y} \mathbf{Y}^\top \mathbf{t} = \mathbf{X}^\top \mathbf{Y} \mathbf{Y}^\top \mathbf{X} \mathbf{w},$$

$$\mathbf{c} \propto \mathbf{Y}^\top \mathbf{t} = \mathbf{Y}^\top \mathbf{X} \mathbf{w} = \mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{u} = \mathbf{Y}^\top \mathbf{X} \mathbf{X}^\top \mathbf{Y} \mathbf{c}.$$

Обновляя вектора по данным правилам, мы максимизируем ковариацию между векторами \mathbf{t} и \mathbf{u}

$$\begin{aligned} \max_{\mathbf{t}, \mathbf{u}} \text{cov}(\mathbf{t}, \mathbf{u})^2 &= \max_{\substack{\|\mathbf{w}\|=1 \\ \|\mathbf{c}\|=1}} \text{cov}(\mathbf{X}\mathbf{w}, \mathbf{Y}\mathbf{c})^2 = \max_{\substack{\|\mathbf{w}\|=1 \\ \|\mathbf{c}\|=1}} \text{cov}(\mathbf{c}^\top \mathbf{Y}^\top \mathbf{X} \mathbf{w})^2 = \\ &= \max_{\|\mathbf{w}\|=1} \text{cov} \left\| \mathbf{Y}^\top \mathbf{X} \mathbf{w} \right\|^2 = \max_{\|\mathbf{w}\|=1} \mathbf{w}^\top \mathbf{X}^\top \mathbf{Y} \mathbf{Y}^\top \mathbf{X} \mathbf{w} = \\ &= \lambda_{\max}(\mathbf{X}^\top \mathbf{Y} \mathbf{Y}^\top \mathbf{X}). \end{aligned}$$

После завершения внутреннего цикла вычисляются вектора \mathbf{p} , \mathbf{q} . Для перехода на следующий шаг необходимо вычесть из матриц \mathbf{X} и \mathbf{Y} одноранговые аппроксимации.

Если ковариация между новыми признаками и ответами максимальна, то можно строить регрессионную модель в пространстве меньшей размерности с сохранением уровня точности прогноза. Параметр метода частных наименьших квадратов $l \in \mathbb{N}$ определяет размерность нового пространства. Отбор признаков осуществляется в виде замены исходных признаков на l новых признаков — линейные комбинации исходных признаков.

Чтобы получить модель регрессии, связывающую \mathbf{Y} и \mathbf{X} , на каждом шаге алгоритма 1 находятся коэффициенты $\beta \in \mathbb{R}$ такие, что $\mathbf{u} \approx \mathbf{t}\beta$. Найденные коэффициенты образуют вектор параметров $\beta \in \mathbb{R}^l$. Подставляя вектор параметров в модель (1), получаем

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^\top + \mathbf{F} \approx \mathbf{T} \text{diag}(\beta) \mathbf{Q}^\top = \mathbf{X}\mathbf{W} \text{diag}(\beta) \mathbf{Q}^\top = \mathbf{X}\Theta.$$

Алгоритм 1 Алгоритм PLSR

Вход: $\mathbf{X}, \mathbf{Y}, l$;

Выход: $\mathbf{T}, \mathbf{U}, \mathbf{P}, \mathbf{Q}$;

- 1: инициализировать \mathbf{u} (первый столбец матрицы \mathbf{Y})
 - 2: $\mathbf{T} := \mathbf{U} := \mathbf{P} := \mathbf{Q} := \emptyset$
 - 3: для $i = 1, \dots, l$
 - 4: **повторять**
 - 5: $\mathbf{w} := \mathbf{X}^\top \mathbf{u} / (\mathbf{u}^\top \mathbf{u})$; $\mathbf{w} := \frac{\mathbf{w}}{\|\mathbf{w}\|}$
 - 6: $\mathbf{t} := \mathbf{X}\mathbf{w}$
 - 7: $\mathbf{c} := \mathbf{Y}^\top \mathbf{t} / (\mathbf{t}^\top \mathbf{t})$; $\mathbf{c} := \frac{\mathbf{c}}{\|\mathbf{c}\|}$
 - 8: $\mathbf{u} := \mathbf{Y}\mathbf{c}$
 - 9: **пока** \mathbf{t} не стабилизируется
 - 10: $\mathbf{T} := \text{concat}[\mathbf{T}; \mathbf{t}]$; $\mathbf{U} := \text{concat}[\mathbf{U}; \mathbf{u}]$
 - 11: $\mathbf{p} := \mathbf{X}^\top \mathbf{t} / (\mathbf{t}^\top \mathbf{t})$, $\mathbf{q} := \mathbf{Y}^\top \mathbf{u} / (\mathbf{u}^\top \mathbf{u})$
 - 12: $\mathbf{P} := \text{concat}[\mathbf{P}; \mathbf{p}]$; $\mathbf{Q} := \text{concat}[\mathbf{Q}; \mathbf{q}]$
 - 13: регрессия \mathbf{u} на \mathbf{t} : $\beta := \mathbf{u}^\top \mathbf{t} / (\mathbf{t}^\top \mathbf{t})$
 - 14: $\mathbf{X} := \mathbf{X} - \mathbf{t}\mathbf{p}^\top$
 - 15: $\mathbf{Y} := \mathbf{Y} - \beta \mathbf{t}\mathbf{q}^\top$
-

Модификация метода частных наименьших квадратов (cnlPLS)

Предлагается провести модификацию алгоритма PLS: совершить криволинейное и нелинейное преобразования пространства целевой переменной и независимой переменной для учета мультиколлинеарности между значениями сигнала в различные моменты времени. Схема модифицированного алгоритма представлена на рис. ?? . После применения к исходным матрицам \mathbf{X} и \mathbf{Y} нелинейных преобразований F_x и F_y соответственно используется линейный алгоритм метода частных наименьших квадратов.

$$\begin{aligned} \mathbf{X} &\xrightarrow{F_x(\mathbf{X}, \mathbf{v}_x)} \tilde{\mathbf{X}} = \mathbf{T} \times \mathbf{P}^\top + \mathbf{E} \\ \mathbf{Y} &\xrightarrow{F_y(\mathbf{Y}, \mathbf{v}_y)} \tilde{\mathbf{Y}} = \mathbf{U} \times \mathbf{Q}^\top + \mathbf{F} \end{aligned}$$

Рис. 2. Размерности векторов в алгоритме cnlPLS

Нелинейные преобразования

Рассматриваются нелинейное параметрическое преобразование пространства зависимой переменной \mathbf{Y} и независимой переменной \mathbf{X} (примеры преобразований представлены в табл. 1). Преобразование и вектор параметров, относящиеся к зависимой переменной и независимой переменной, обозначим соответственно $F_y(\mathbf{Y}, \mathbf{v}_y)$ и $F_x(\mathbf{X}, \mathbf{v}_x)$ и введем переменные для преобразованных пространств

$$\tilde{\mathbf{Y}} = F_y(\mathbf{Y}, \mathbf{v}_y), \quad \tilde{\mathbf{X}} = F_x(\mathbf{X}, \mathbf{v}_x). \quad (3)$$

Функции для криволинейных преобразований удовлетворяют следующим условиям:

- $F : \mathbb{R} \rightarrow \mathbb{R}$,
- $F(0) = 0$,
- F дифференцируется по параметрам \mathbf{v}_y ,
- существует F^{-1} .

№	Функция	Параметры
1	$F(x) = \text{sign}(x) \exp(a)(\exp(b x) - 1)$	$a, b > 0$
2	$F(x) = \text{sign}(x) \exp(a)(\exp(b \ln(1 + x)) - 1)$	$a, b > 0$
3	$F(x) = \text{sign}(x) \exp(a)(\exp(b x ^{1/2}) - 1)$	$a, b > 0$
4	$F(x) = \text{sign}(x) \exp(a)(\exp(b x ^{1/3}) - 1)$	$a, b > 0$
5	$F(x) = \text{sign}(x) \exp(a)(\exp(b x ^{1/4}) - 1)$	$a, b > 0$
6	$F(x) = \text{sign}(x) \exp(a)(\exp(b x ^2) - 1)$	$a, b > 0$

Таблица 1. Нелинейные преобразования

Для обучения параметров \mathbf{v}_y используется градиентный метод. Предлагается подход для обновления весов \mathbf{v}_y , основанный на линеаризации функции преобразования. Разложим (3) в ряд Тейлора до второго порядка:

$$\mathbf{u} \approx \mathbf{u}_0 + \frac{\partial \mathbf{u}}{\partial \mathbf{v}_y} \Delta \mathbf{v}_y.$$

Для вычисления $\Delta \mathbf{v}_y$ предложены следующие шаги. Рассматривается разница $\mathbf{u} - \mathbf{u}_0 = \frac{\partial \mathbf{u}}{\partial \mathbf{v}_y} \Delta \mathbf{v}_y$. Определется рассогласование

$$\mathbf{u} - \mathbf{u}_0 \approx \frac{\partial \mathbf{u}}{\partial \mathbf{v}_y} \Delta \mathbf{v}_y = \mathbf{J}_u \Delta \mathbf{v}_y,$$

141 где матрица \mathbf{J}_u состоит из частных производных $\left\{ \frac{\partial \mathbf{u}}{\partial \mathbf{v}_y} \right\}$, вычисленных при известном зна-
142 чении переменной \mathbf{u} :

$$\mathbf{J}_u = \frac{\partial \mathbf{u}}{\partial \mathbf{v}_y} = \frac{\partial}{\partial \mathbf{v}_y} (\tilde{\mathbf{Y}} \mathbf{c}) = \frac{1}{(\mathbf{t}^\top \mathbf{t})} \frac{\partial}{\partial \mathbf{v}_y} (\tilde{\mathbf{Y}} \tilde{\mathbf{Y}}^\top \mathbf{t}) = \frac{1}{(\mathbf{t}^\top \mathbf{t})} \left(\frac{\partial \tilde{\mathbf{Y}}}{\partial \mathbf{v}_y} \cdot \tilde{\mathbf{Y}}^\top + \tilde{\mathbf{Y}} \cdot \frac{\partial \tilde{\mathbf{Y}}^\top}{\partial \mathbf{v}_y} \right) \mathbf{t}.$$

143 Правило обновления для вектора $\Delta \mathbf{v}$ является решением задачи регрессии рассогла-
144 сования

$$\Delta \mathbf{v}_y = (\mathbf{J}_u^\top \mathbf{J}_u)^{-1} \mathbf{J}_u^\top (\mathbf{u} - \mathbf{u}_0). \quad (4)$$

145 Аналогично преобразованию зависимой переменной сводим задачу обновления вектора
146 параметров \mathbf{v}_x к задаче линейной регрессии:

$$\begin{aligned} \mathbf{t} - \mathbf{t}_0 &\approx \frac{\partial \mathbf{t}}{\partial \mathbf{v}_x} \Delta \mathbf{v}_x = \mathbf{J}_t \Delta \mathbf{v}_x \\ \Delta \mathbf{v}_x &= (\mathbf{J}_t^\top \mathbf{J}_t)^{-1} \mathbf{J}_t^\top (\mathbf{t} - \mathbf{t}_0). \end{aligned}$$

147 Преобразование независимой переменной

148 Алгоритм cnPLSR

149 В данном разделе представлен модифицированный метод PLSR, содержащий шаги
150 преобразования целевой переменной. Аналогично методу PLSR (алгоритм 1), алгоритм
151 ?? начинается с инициализации вектора \mathbf{u} , а обновления весов преобразования считается
152 с помощью рассогласования \mathbf{e} для вектора \mathbf{u} , вычисленного в цикле и на предыдущей
153 итерации.

Алгоритм 2 Алгоритм snlPLSR с преобразованием пространства объектов 2

Вход: $\mathbf{X}, \mathbf{Y}, l$;**Выход:** $\mathbf{T}, \mathbf{U}, \mathbf{P}, \mathbf{Q}, \mathbf{v}_x, \mathbf{v}_y$;

```

1: инициализировать  $\mathbf{v}_x$  и  $\mathbf{v}_y$ 
2:  $\mathbf{T} := \mathbf{U} := \mathbf{P} := \mathbf{Q} := \emptyset$ 
3: для  $i = 1, \dots, l$ 
4:   инициализировать  $\mathbf{t}$  (первый столбец матрицы  $\mathbf{X}$ )
5:   инициализировать  $\mathbf{u}$  (первый столбец матрицы  $\mathbf{Y}$ )
6:   повторять
7:      $\mathbf{t}_0 := \mathbf{t}, \mathbf{u}_0 = \mathbf{u}$ 
8:      $\tilde{\mathbf{X}} := F_x(\mathbf{X}, \mathbf{v}_x); \quad \tilde{\mathbf{Y}} = F_y(\mathbf{Y}, \mathbf{v}_y)$ 
9:      $\mathbf{w} := \tilde{\mathbf{X}}^\top \mathbf{u} / (\mathbf{u}^\top \mathbf{u}); \quad \mathbf{w} := \frac{\mathbf{w}}{\|\mathbf{w}\|}$ 
10:     $\mathbf{t} := \tilde{\mathbf{X}} \mathbf{w}$ 
11:     $\Delta \mathbf{v}_x = (\mathbf{J}_t^\top \mathbf{J}_t)^{-1} \mathbf{J}_t^\top (\mathbf{t} - \mathbf{t}_0)$ , где  $\mathbf{J}_t := \frac{\partial \mathbf{t}}{\partial \mathbf{v}_x}$ 
12:     $\mathbf{v}_x := \mathbf{v}_x + \Delta \mathbf{v}_x$ 
13:     $\mathbf{c} := \tilde{\mathbf{Y}}^\top \mathbf{t} / (\mathbf{t}^\top \mathbf{t}); \quad \mathbf{c} := \frac{\mathbf{c}}{\|\mathbf{c}\|}$ 
14:     $\mathbf{u} := \tilde{\mathbf{Y}} \mathbf{c}$ 
15:     $\Delta \mathbf{v}_y = (\mathbf{J}_u^\top \mathbf{J}_u)^{-1} \mathbf{J}_u^\top (\mathbf{u} - \mathbf{u}_0)$ , где  $\mathbf{J}_u := \frac{\partial \mathbf{u}}{\partial \mathbf{v}_y}$ 
16:     $\mathbf{v}_y := \mathbf{v}_y + \Delta \mathbf{v}_y$ 
17:   пока  $\mathbf{t}$  не стабилизируется
18:    $\mathbf{T} := \text{concat}[\mathbf{T}; \mathbf{t}]; \mathbf{U} := \text{concat}[\mathbf{U}; \mathbf{u}]$ 
19:    $\mathbf{p} := \tilde{\mathbf{X}}^\top \mathbf{t} / (\mathbf{t}^\top \mathbf{t}), \mathbf{q} := \tilde{\mathbf{Y}}^\top \mathbf{u} / (\mathbf{u}^\top \mathbf{u})$ 
20:    $\mathbf{P} := \text{concat}[\mathbf{P}; \mathbf{p}]; \mathbf{Q} := \text{concat}[\mathbf{Q}; \mathbf{q}]$ 
21:   регрессия  $\mathbf{u}$  на  $\mathbf{t}$ :  $\beta := \mathbf{u}^\top \mathbf{t} / (\mathbf{t}^\top \mathbf{t})$ 
22:    $\tilde{\mathbf{X}} := \tilde{\mathbf{X}} - \mathbf{t} \mathbf{p}^\top$ 
23:    $\tilde{\mathbf{Y}} := \tilde{\mathbf{Y}} - \beta \mathbf{t} \mathbf{q}^\top$ 
24:    $\mathbf{X} = F_x^{-1}(\tilde{\mathbf{X}}, \mathbf{v}_x); \mathbf{Y} = F_y^{-1}(\tilde{\mathbf{Y}}, \mathbf{v}_y)$ 

```

Вычислительный эксперимент

В рамках вычислительного эксперимента строится прогноз временных рядов. В ходе эксперимента сравниваются методы PLSR, нелинейных автоэнкодеров и snlPLS. Сравнение проводится на реальных данных объемов потребления электроэнергии в Польше.

Вычислительный эксперимент, продемонстрированный в этом разделе, основан на данных электроэнергии. Данные состоят из временного ряда польских электрических нагрузок и временных рядов погоды в Варшаве (долгота: 21,25, широта: 52,30, высота над уровнем моря: 94). Временные ряды энергии состоят из почасовых записей (всего 52512 наблюдений), а погодные измерения проводились раз в день и содержат 2188 наблюдений. Многомасштабные временные ряды соответствуют периоду 1999-2004 годов. Результаты, полученные с этим набором данных, являются иллюстрацией предлагаемых методов, поскольку данные содержат многомасштабные временные ряды, имеющие различный характер.

Алгоритм	N=3	N=5	N=10	N=20
PLS	0,00404	0,00337	0,00151	0,00135
cnlPLS $g(x) = \text{sign}(x) \exp(a)(\exp(b x) - 1)$	0.00529	0.00514	0.00536	0.00506
cnlPLS $g(x) = \text{sign}(x) \exp(a)(\exp(b \ln(1 + x) - 1)$	0.00362	0.00386	0.00326	0.00317
cnlPLS $g(x) = \text{sign}(x) \exp(a)(\exp(b x ^{1/2}) - 1)$	0.00272	0.00236	0.00287	0.00128
cnlPLS $g(x) = \text{sign}(x) \exp(a)(\exp(b x ^{1/3}) - 1)$	0.00241	0.00233	0.00221	0.00173
cnlPLS $g(x) = \text{sign}(x) \exp(a)(\exp(b x ^{1/4}) - 1)$	0.00796	0.00768	0.00737	0.00803
cnlPLS $g(x) = \text{sign}(x) \exp(a)(\exp(b x ^2) - 1)$	0.00816	0.00798	0.00796	0.00775

Таблица 2. Значения ошибки MSE для разных чисел компонент и разных функций

Примеры работы алгоритма приведены на рис. ???. Метод успешно делает краткосрочный прогноз (до 10 дней). С увеличением горизонта прогнозирования предсказание смещается.

Результаты вычислительного эксперимента для предложенного модифицированного алгоритма cnlPLS представлены на рис. ???. На графиках изображены сглаженные зависимости ошибки MSE от числа компонент в алгоритме для разных функций. Из графиков видно, что для функций (a) – (e) ошибка при увеличении числа компонент падает, затем колеблется, слабо меняясь. Ошибка алгоритма с функцией (f) увеличивается при увеличении числа компонент. Это означает, что преобразование, выполненное в пространстве целевой переменной с помощью функции (f), плохо описывает зависимость. Меньшую ошибку имеют функции, растущие медленнее, а именно (d) и (e).

В табл. 2 продемонстрировано увеличение точности прогнозирования при использовании криволинейного преобразования в пространстве зависимой переменной, но увеличение точности в пределах погрешности алгоритма (0.0005-0.0010). Функции с быстрым ростом не позволяют описать зависимость.

Заключение

В данной работе предложен новый подход к обнаружению зависимостей в пространстве зависимой переменной задачи прогнозирования временных рядов. Сравнивались результаты прогнозирования временных рядов, полученных с помощью метода частных наименьших квадратов и предложенной модификации. Проведен вычислительный эксперимент на реальных данных потребления электроэнергии в Варшаве. Построенная прогностическая модель показала высокое качество предсказания электрической нагрузки.

Литература

- [1] *Thrun, Sebastian and Pratt, Lorien* Learning to learn // Springer Science & Business Media, 2012.
- [2] *Chong, Il Gyo and 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.
- [3] *Xuefeng, Yan* Hybrid artificial neural network based on BP-PLSR and its application in development of soft sensors // *Chemometrics and Intelligent Laboratory Systems*, 2010. Vol. 103. No. 2. P. 152–159.
- [4] *Mcavovt, J. and Process, Chemical Title* // *Journal name*, 2005. Vol. 16. No. 4. P. 379–391.
- [5] *Yan, Xuefeng F. and Chen, Dezhaoh Z. and Hu, Shangru X.* Chaos-genetic algorithms for optimizing the operating conditions based on RBF-PLS model // *Computers and Chemical Engineering*, 2003. Vol. 27. No. 10. P. 1393–1404.
- [6] *Frank, Ildiko E.* A nonlinear PLS model // *Chemometrics and Intelligent Laboratory Systems*, 1990. Vol. 8. No. 2. P. 109–119.
- [7] *Zhou, Yan Ping and Jiang, Jian Hui and Lin, Wei Qi and Xu, Lu and Wu, Hai Long and Shen, Guo Li and Yu, Ru Qin* Artificial neural network-based transformation for nonlinear partial least-square regression with application to QSAR studies // *Talanta*, 2007. Vol. 71. No. 2. P. 848–853.
- [8] *Chong, Il Gyo and Jun, Chi Hyuck* Notes on the history and nature of partial least squares (PLS) modelling // *Journal of Chemometrics*, 1988. Vol. 2. No. January. P. 231–246.
- [9] *Höskuldsson, Agnar* PLS regression // *Chemometrics and Intelligent Laboratory Systems*, 1987. Vol. 2. No. August. P. 581–591.
- [10] *Bulut, Elif and Egrioglu, Erol* A New Partial Least Square Method Based on Elman Neural Network // *Chemometrics and Intelligent Laboratory Systems*, 2005. Vol. 4. No. 4. P. 154–158.
- [11] *Ng, Kee Siong* A Simple Explanation of Partial Least Squares // *Journal title*, 2013. Vol. volume. No. number. P. 1–10.
- [12] *Rosipal, Roman* Nonlinear partial least squares: An overview // *Chemoinformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques*, 2011. Vol. number. No. number. P. 169–189.
- [13] *Wold, Svante and Kettaneh-Wold, Nouna and Skagerberg, Bert* Nonlinear PLS modeling // *Chemometrics and Intelligent Laboratory Systems*, 1989. Vol. 7. No. 1-2. P. 53–65.
- [14] *Rosipal, Roman and Kramer, Nicole* Overview and Recent Advances in Partial Least Squares // *????? C. Saunders et al. (Eds.): SLSFS 2005, LNCS 3940*, 2006. Vol. ?. No. ?. P. 34–51.
- [15] *Lu, Wen-Cong and Chen, Nian-Yi and Li, Guo-Zheng and Yang, Jie* Multitask Learning Using Partial Least Squares Method // *Proceedings of the Seventh International Conference on Information Fusion; International Society of Information Fusion*, 2004. Vol. 1. P. 79–84.
- [16] *Varnek, Alexandre and Baskin, Igor* Machine learning methods for property prediction in chemoinformatics: Quo Vadis? // *Journal of Chemical Information and Modeling*, 2012. Vol. 52. No. 6. P. 1413–1437.
- [17] *Lehky, Sidney R. and Kiani, Roozbeh and Esteky, Hossein and Tanaka, Keiji* Dimensionality of object representations in monkey inferotemporal cortex // *Neural computation*, 2014. Vol. 1872. No. 10. P. 1840–1872.
- [18] *Abdi, Hervé* Partial Least Squares (PLS) Regression // *Encyclopedia for research methods for the social sciences*, 2003. P. 792–795.
- [19] *Caruana, Rich and de Sa, Virginia R.* Benefitting from the Variables that Variable Selection Discards // *Journal of Machine Learning Research*, 2003. Vol. 3. No. 7-8. P. 1245–1264.

- 234 [20] *Mises R. V., Pollaczek-Geiringer H.* Praktische Verfahren der Gleichungsauflösung // ZAMM-
235 Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und
236 Mechanik, 1929. Vol. 9. No. 1. P. 58–77.
- 237 [21] *Li J. et al.* Feature selection: A data perspective // arXiv preprint arXiv:1601.07996, 2016.