

# Прогнозирование намерений по сигналам ECoG\*

*Калиниченко О. И., Ремизова А. С.*

<sup>1</sup> Московский физико-технический институт

Работа посвящена построению системы тестирования  $L \times K$  прогностических моделей для различных критериев качества. Рассматривается случай коррелированных входных и выходных сигналов электрокортикограммы и фазовых траекторий движения конечностей высокой размерности. Рассматривается задача предсказания намерений по сигналам головного мозга. Входные данные – сигналы электрокортикограммы (ECoG). Для выявления и устранения скрытых зависимостей используются методы снижения размерности пространства и отбора признаков. Предложенная система тестирования оценивает качество прогноза моделей и ошибку. Вычислительный эксперимент проводится на данных ECoG проекта NeuroTycho.

**Ключевые слова:** декодирование временных рядов, PLS, QPFS, электрокортикограмма, траекторий движения конечностей.

## Введение

Работа посвящена исследованию методов моделирования нейросетевого интерфейса (BCI) [5]. Входные данные – сигналы мозга, полученные с помощью электрокортикографии (ECoG) и электроэнцефалографии (EEG). ECoG-сигналы имеют более высокое разрешение и большую амплитуду, однако для их получения требуется непосредственное подсоединение электродов к коре головного мозга. Одной из задач при построении систем BCI является предсказание намерений.

Предлагается декодировать исходные сигналы и спрогнозировать траекторию движения верхних конечностей. Исходное пространство имеет избыточно высокую размерность. Линейная зависимость между признаками приводит к мультиколлинеарности. Для устранения мультиколлинеарности предлагается применить методы понижения размерности и отбора признаков.

Признаковое описание многомерного временного ряда существует в пространствах независимых и зависимых переменных. Для учета существующих закономерностей в исходном и выходном пространстве используется скрытое пространство латентных переменных. В скрытом пространстве происходит согласование между образами исходных пространств.

В эксперименте рассматриваются следующие модели: метод частных наименьших квадратов (PLS) [2], отбор признаков с помощью квадратичного программирования (QPFS) [4], метод Белсли (Belsley) и вариации этих методов.

PLS является методом отбора признаков

Описание метода QPFS...

Предлагается система тестирования прогностических моделей с оценкой качества и анализом ошибки. Подобный инструмент может применяться не только в задаче анализа сигналов мозга, но и во многих других задачах, связанных с прогнозированием многомерных временных рядов.

## Постановка задачи

### Постановка задачи предсказания

Задана выборка  $\mathcal{D} = (\mathbf{X}, \mathbf{Y})$ , где  $\mathbf{X} \in \mathbb{R}^{m \times n}$  — матрица объектов,  $\mathbf{X} = [\chi_1 \dots \chi_n]$ , где  $\chi_j \in \mathbb{R}^m$  — вектор значений  $j$ -го признака на элементах выборки;  $\mathbf{Y} \in \mathbb{R}^{m \times r}$  — матрица ответов,  $\mathbf{Y} = [y_1 \dots y_r]$ . Имеется линейная модель  $\mathbf{f}$  с набором параметров  $\boldsymbol{\theta}$  из пространства  $\mathbb{R}^{n \times r}$ , предсказывающая  $\mathbf{y} \in \mathbb{R}^r$  по  $\mathbf{x} \in \mathbb{R}^n$ , следующего вида:

$$\mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) = \underset{1 \times n}{\mathbf{x}} \underset{n \times r}{\boldsymbol{\theta}}. \quad (1)$$

Задана функция ошибки  $S$  на выборке  $\mathcal{D}$  и модели  $\mathbf{f}$  с параметрами  $\boldsymbol{\theta}$ . Задачей является поиск наилучших параметров  $\boldsymbol{\theta}^*$ , то есть таких, при которых функция ошибки минимальна:

$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta} \in \Theta} S(\boldsymbol{\theta} | \mathcal{D}, \mathbf{f}). \quad (2)$$

В случае коррелированных данных  $\mathbf{X}$  решение задачи выбора оптимального вектора параметров  $\boldsymbol{\theta}$  нестабильно.

За  $\mathbf{F}(\mathbf{X}, \boldsymbol{\theta})$  обозначена матрица  $[\mathbf{f}(\mathbf{x}_1, \boldsymbol{\theta}), \mathbf{f}(\mathbf{x}_2, \boldsymbol{\theta}), \dots, \mathbf{f}(\mathbf{x}_m, \boldsymbol{\theta})]^\top$ . Рассматривается квадратичная функция ошибки:

$$\|\mathbf{F}(\mathbf{X}, \boldsymbol{\theta}) - \mathbf{Y}\|_2^2 \rightarrow \min_{\boldsymbol{\theta} \in \mathbb{R}^{n \times r}} \quad (3)$$

Если пространство признаков имеет высокую размерность, вероятно, что матрица  $\mathbf{X}$  близка к вырожденной, а потому решение проблемы оптимизации (3) будет нестабильным. Для решения указанной задачи применяются методы отбора признаков.

### Постановка задачи отбора признаков

Пусть  $\mathcal{J} = \{1, \dots, n\}$  — множество индексов признаков, а  $\mathcal{A} \subseteq \mathcal{J}$  — его некоторое подмножество. Тогда линейную модель  $\mathbf{f}$  на подмножестве признаков  $\mathcal{A}$  можно определить как:

$$\mathbf{f}(\mathbf{x}, \mathcal{A}, \boldsymbol{\theta}) = \underset{1 \times |\mathcal{A}|}{\mathbf{x}_{\mathcal{A}}} \underset{|\mathcal{A}| \times r}{\boldsymbol{\theta}} \quad (4)$$

Функция ошибки  $S$  теперь вычисляется на выборке  $\mathcal{D}$ , модели  $\mathbf{f}$  с параметрами  $\boldsymbol{\theta}$  и на  $\mathcal{A}$ .

Тогда задача предсказания определяется как поиск набора признаков  $\mathcal{A}$  и матрицы параметров  $\boldsymbol{\theta}^* \in \mathbb{R}^{|\mathcal{A}| \times r}$  такого, что:

$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^{|\mathcal{A}| \times r}} S(\boldsymbol{\theta}, \mathcal{A} | \mathcal{D}, \mathbf{f}) \quad (5)$$

Задача отбора признаков поставлена как:

$$\mathcal{A}^* = \arg \min_{\mathcal{A} \subseteq \mathcal{J}} Q(\mathcal{A} | \mathbf{X}, \mathbf{y}), \quad (6)$$

где  $Q : \mathcal{A} \rightarrow \mathbb{R}$  — это некоторый критерий качества, который определяет качество выбранного подмножества признаков  $\mathcal{A} \subseteq \mathcal{J}$ .

Для решения этой задачи не обязательно требуется оценка оптимального вектора параметров  $\boldsymbol{\theta}^*$ . Она использует зависимости между векторами  $\chi_j, j \in \mathcal{J}$  и целевыми векторами  $y_i, i \in \{1, \dots, r\}$ .

Пусть  $\mathbf{a} \in \mathbb{B}^n = \{0, 1\}^n$  – вектор индикаторов принадлежности признаков подмножеству  $\mathcal{A}$  – то есть  $\mathbf{a}_j = 1$  тогда и только тогда, когда  $j \in \mathcal{A}$ . Тогда задача 6 может быть записана как:

$$\mathbf{a}^* = \arg \min_{\mathbf{a} \in \mathbb{B}^n} Q(\mathcal{A} | \mathbf{X}, \mathbf{y}), \quad (7)$$

где  $Q : \mathbb{B}^n \rightarrow \mathbb{R}$  – критерий  $Q$ , но определенный на  $\mathbb{B}^n$ ; связь между вектором  $\mathbf{a}^*$  и множеством  $\mathcal{A}^*$  определяется через:

$$\mathbf{a}_j^* = 1 \Leftrightarrow j \in \mathcal{A}^*, j \in \mathcal{J}. \quad (8)$$

## Методы отбора признаков

Существуют различные методы отбора признаков. В данной работе используются:

1. PCA
2. QPFS
3. Генетический алгоритм
4. Elastic

Приведем краткие описания используемых алгоритмов.

**PCA** PCA, или метод главных компонент, позволяет обнаружить линейные закономерности в данных, благодаря чему многомерные данные могут быть хорошо приближены линейными комбинациями векторов из пространства более низкой размерности размера без значительной потери информации. Базисные вектора этого пространства и называются главными компонентами. Они выбираются таким образом, чтобы дисперсия исходных данных по их направлениям была максимальна. В работах (доб. ссылки) показано, что  $k$ -мерное пространство размерности  $k$  имеет базисом вектора, отвечающие  $k$  максимальным собственным значениям ковариационной матрицы исходных данных.

### PLS.

Метод частных наименьших квадратов PLS рассматривает в качестве признаков линейные комбинации исходных. Предполагается, что существует скрытое пространство латентных переменных малой размерности  $l$  ( $l < n, r$ ). Происходит поиск матрицы  $\mathbf{T} \in \mathbb{R}^{m \times l}$ , которая наилучшим образом описывает матрицы  $\mathbf{X}$  и  $\mathbf{Y}$ .

$$\mathbf{X}_{m \times n} = \mathbf{T}_{m \times l}^T \cdot \mathbf{P}_{l \times n}^T + \mathbf{E}_{m \times n} \quad (9)$$

$$\mathbf{Y}_{m \times r} = \mathbf{U}_{m \times l}^T \cdot \mathbf{Q}_{l \times r}^T + \mathbf{F}_{m \times r} \quad (10)$$

$\mathbf{T}$  —

### QPFS

Чтобы вычислить матрицу  $\mathbf{Q}$  и вектор  $\mathbf{b}$

$$\text{Sim} \quad (11)$$

$$\text{Rel} \quad (12)$$

Коэффициент корреляции Пирсона определяется как:

$$\rho_{ij} = \frac{\text{cov}(\mathbf{x}_i, \mathbf{x}_j)}{\sqrt{\text{var}(\mathbf{x}_i) \cdot \text{var}(\mathbf{x}_j)}}$$

**Генетический алгоритм** Этот алгоритм основан на идеях, заимствованных у природы. Изначально имеется "популяция" прогностических моделей. Каждая модель (в исследовании, которому посвящена данная работа, это линейная модель) характеризуется множеством признаков, которые будут использоваться при ее построении. Изначально множества выбираются случайно. Алгоритм итеративный. На каждом шаге сначала вычисляется метрика качества модели. Выбирается некоторое количество моделей с наилучшей метрикой. Далее из этого множества случайно формируются пары моделей, которые "скрещиваются": формируется новое множество признаков, в котором каждый признак присутствует, если он есть в обеих моделях-родителях, не присутствует, если его нет в обеих моделях-родителях, и присутствует с вероятностью  $\frac{1}{2}$ , если он есть ровно в одной модели-родителе. Также на каждом шаге могут происходить мутации: случайное изменение множества признаков некоторых полученных моделей. Алгоритм продолжает работу, пока не выполняется критерий останова. Выбор критерия, а также другие детали, зависят от реализации.

## Анализ ошибки

В данной работе особое внимание уделяется исследованию не только точности, но и устойчивости алгоритма. Для этого применяются различные критерии качества.

Неустойчивостью алгоритма будем называть статистически значимое изменение ошибки при незначимых изменениях выборки или множества признаков.

Для того, чтобы исследовать неустойчивость изучаемых алгоритмов, предлагается некоторым образом немного изменить выборку и сравнить результат работы алгоритма. Чем больше неустойчивость алгоритма, тем больше будут отличаться результаты.

Введем некоторые обозначения.

Пусть  $m$  - количество наблюдений,

$n$  - количество признаков,

$q$  - размерность ответа,

$X \in R^{m \times n}$  - матрица признаков,

$y \in R^{m \times q}$  - истинные ответы,

$A \in \{0, 1\}^n$  - булева матрица отобранных признаков ( $A_i = 1 \Leftrightarrow i$  - признак выбран),

$\theta \in R^{n \times q}$  - предсказанные линейной моделью коэффициенты,

$\hat{y} = X\theta \in R^{m \times q}$  - предсказания линейной модели (построенной только с учетом отобранных признаков)

$S(y, \hat{y})$  - функция ошибки (метрика качества). Мы считаем, что чем меньше ее значение, тем больше точность алгоритма.

$\varepsilon = y - \hat{y}$  - ошибка предсказания

$l \in \{0, 1, 2, 3, 4\}$  - номер исследуемого алгоритма

$L$  - множество алгоритмов

Для того, чтобы вызвать незначительные изменения выборки, используется процедура бутстрепа. Мы будем создавать новые выборки, выбирая с возвращением  $m$  элементов исходной выборки. Фактически, таким образом будут генерироваться выборки из эмпирического распределения исходных данных.

Пусть  $k$  - количество бутстрепных выборок.  $K$  - множество бутстрепных выборок.

Для изменения множества параметров используется метод семплирования пространства параметров. Выбираются случайные подвыборки размера  $r < n$ . ???

Для каждой пары из  $L \times K$  считается ошибка линейной модели при применении данного алгоритма отбора признаков. Чем больше ошибки разнятся от выборки к выборке, тем менее устойчив метод отбора.

Далее, наверное, надо конкретно уже описать.

### Метрики

Пусть имеются истинный прогноз  $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m)$  и предсказание  $\hat{\mathbf{Y}} = (\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \dots, \hat{\mathbf{y}}_m)$ ;  $\mathbf{y}_i$  и  $\hat{\mathbf{y}}_i$ ,  $i = 1, 2, \dots, m$  – вектора размерности  $r$ .

Среднеквадратичная ошибка (mean squared error):

$$\text{MSE}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{1}{m} \sum_{i=1}^m \|\mathbf{y}_i - \hat{\mathbf{y}}_i\|_2^2 \quad (13)$$

Корень среднеквадратичной ошибки (root-mean-squared error):

$$\text{RMSE}(\mathbf{Y}, \hat{\mathbf{Y}}) = \sqrt{\text{MSE}(\mathbf{Y}, \hat{\mathbf{Y}})} \quad (14)$$

Нормированная среднеквадратичная ошибка (scaled mean squared error):

$$\text{sMSE}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{\sum_{i=1}^m \|\mathbf{y}_i - \hat{\mathbf{y}}_i\|_2^2}{\sum_{i=1}^m \|\mathbf{y}_i - \bar{\mathbf{y}}\|_2^2}, \quad \bar{\mathbf{y}} = \frac{1}{m} \sum_{i=1}^m \mathbf{y}_i \quad (15)$$

MAE, MADE, Коэффициент корреляции. Различные параметры модели из [4].

Сложность модели

### Литература

- [1] J. del R. Millán, F. Renken, J. Mourino and W. Gerstner Brain-actuated interaction // *Artif. Intell.*, 159(2004) 241–259.
- [2] Isachenko R., Vladimirova M., Strijov V. Dimensionality reduction for time series decoding and forecasting problems // *Machine Learning and Data Analysis*.
- [3] Katrutsa A., Strijov V. Stress test procedure for feature selection algorithms, 2015 // *Expert System with Applications*, 142, 172–183
- [4] Katrutsa A., Strijov V. Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria, 2017 // *Expert System with Applications*, 76, 1–11.
- [5] Motrenko A., Strijov V. Multi-way Feature Selection for ECoG-based Brain-Computer Interface // ???.