

Прогнозирование намерений по сигналам ECoG*

Калиниченко О. И., Ремизова А. С.

¹ Московский физико-технический институт

Работа посвящена построению системы тестирования прогностических моделей, предлагающей различные критерии качества. Рассматривается случай коррелированных входных и выходных сигналов электрокортикограммы и фазовых траекторий движения конечностей высокой размерности. Ставится 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 анализирует ошибку. Вычислительный эксперимент проводится на данных ECoG проекта NeuroTycho.

Ключевые слова: декодирование временных рядов, отбор признаков, PLS, QPFS, генетический алгоритм, BCI, электрокортикограмма, траектории движения конечностей.

Введение

В работе исследуются методы 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 подключение 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: линейная регрессия, метод частных наименьших квадратов (PLS) [2], а также методы отбора признаков с помощью квадратичного программирования (QPFS) [4] и генетического алгоритма.

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

Задача предсказания

Научный руководитель: Стрижов В. В. Задачу поставил: Стрижов В. В. Консультант: Исаченко Р. В.

Задана выборка $\mathcal{D} = (\mathbf{X}, \mathbf{Y})$, где $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n] \in \mathbb{R}^{m \times n}$ — матрица объектов, $\mathbf{Y} = [\mathbf{y}_1 \dots \mathbf{y}_r] \in \mathbb{R}^{m \times r}$ — целевая матрица. Вектор $\mathbf{x}_j \in \mathbb{R}^m$ — значения j -го признака на элементах выборки; вектор $\mathbf{y}_i \in \mathbb{R}^m$ — i -й целевой столбец. Имеется линейная модель $\mathbf{f}(\mathbf{x}, \Theta)$ с набором параметров $\Theta \in \mathbb{R}^{n \times r}$, аппроксимирующая целевой вектор $\mathbf{y} \in \mathbb{R}^r$ по известным значениям признаков объекта $\mathbf{x} \in \mathbb{R}^n$:

$$\mathbf{f}(\mathbf{x}, \Theta) = \Theta^\top \mathbf{x}. \quad (1)$$

За $\mathbf{F}(\mathbf{X}, \Theta)$ обозначена матрица $[\mathbf{f}(\mathbf{x}_1, \Theta), \dots, \mathbf{f}(\mathbf{x}_m, \Theta)]^\top$.

На выборке \mathcal{D} и модели $\mathbf{f}(\mathbf{x}, \Theta)$ задана функция ошибки S . Задачей является поиск наилучших параметров Θ , доставляющих минимум функции ошибки S :

$$\Theta^* = \arg \min_{\Theta \in \mathbb{R}^{m \times r}} S(\Theta | \mathcal{D}, \mathbf{f}). \quad (2)$$

При наличии мультикоррелированных столбцов в матрице \mathbf{X} , что вероятно, если пространство признаков имеет высокую размерность, решение задачи (2) выбора матрицы параметров Θ нестабильно. Определение устойчивости (стабильности) модели приведено в разделе «Анализ ошибки».

Задача отбора признаков

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

$$\mathbf{f}(\mathbf{x}_{\mathcal{A}}, \Theta_{\mathcal{A}}) = \Theta_{\mathcal{A}}^\top \mathbf{x}_{\mathcal{A}}. \quad (3)$$

Задача предсказания (2) состоит в определении матрицы параметров $\Theta \in \mathbb{R}^{|\mathcal{A}| \times r}$ при заданном множестве \mathcal{A} такой, что:

$$\Theta^* = \arg \min_{\Theta \in \mathbb{R}^{|\mathcal{A}| \times r}} S(\Theta | \mathcal{D}, \mathbf{f}). \quad (4)$$

Для нахождения множества \mathcal{A} решается задача отбора признаков:

$$\mathcal{A}^* = \arg \min_{\mathcal{A} \in \mathcal{J}} Q(\mathcal{A} | \mathbf{X}, \mathbf{Y}), \quad (5)$$

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

Для решения этой задачи не обязательно требуется оценка оптимального вектора параметров Θ^* . Она использует зависимости между векторами $\mathbf{x}_j, j \in \mathcal{J}$ и целевыми векторами $\mathbf{y}_i, i \in \{1, \dots, r\}$.

Пусть $\mathbf{a} \in \mathbb{B}^n = \{0, 1\}^n$ — вектор принадлежности признаков подмножеству \mathcal{A} . Задача (5) представлена в виде:

$$\mathbf{a}^* = \arg \min_{\mathbf{a} \in \mathbb{B}^n} Q(\mathbf{a} | \mathbf{X}, \mathbf{Y}), \quad (6)$$

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

$$\mathcal{A}^* = \{j : \mathbf{a}_j^* = 1, j \in \mathcal{J}\}. \quad (7)$$

Модели

В данной работе рассматриваются модели линейной регрессии и PLS (Partial Least Squares Regression).

Линейная регрессия

Модель линейной регрессии является линейной моделью (1). Матрица параметров $\Theta \in \mathbb{R}^{n \times r}$ подбирается таким образом, чтобы минимизировать квадратичную функцию ошибки S (2):

$$\|\mathbf{F}(\mathbf{X}, \Theta) - \mathbf{Y}\|_2^2 \rightarrow \min_{\Theta \in \mathbb{R}^{n \times r}}. \quad (8)$$

Решением оптимизационной задачи является:

$$\Theta = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}. \quad (9)$$

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} .

$$\underset{m \times n}{\mathbf{X}} = \underset{m \times l}{\mathbf{T}}^\top \cdot \underset{l \times n}{\mathbf{P}}^\top + \underset{m \times n}{\mathbf{E}} \quad (10)$$

$$\underset{m \times r}{\mathbf{Y}} = \underset{m \times l}{\mathbf{U}}^\top \cdot \underset{l \times r}{\mathbf{Q}}^\top + \underset{m \times r}{\mathbf{F}} \quad (11)$$

Методы отбора признаков

В данной работе рассматриваются метод отбора признаков QPFS (Quadratic Programming Feature Selection) и генетический алгоритм.

QPFS

Идея метода отбора признаков состоит в минимизации числа схожих признаков и максимизации числа признаков, информативных для задачи предсказания целевого вектора $\boldsymbol{\nu} \in \mathbb{R}^m$. Формализация постановки задачи состоит в ее определении как задачи (6) со следующим критерием Q :

$$Q(\mathbf{a}|\mathbf{X}, \boldsymbol{\nu}) = \mathbf{a}^\top \mathbf{Q} \mathbf{a} - \mathbf{b}^\top \mathbf{a}, \quad (12)$$

где $\mathbf{Q} \in \mathbb{R}^{m \times m}$ — матрица попарного сходства признаков, $\mathbf{b} \in \mathbb{R}^m$ — вектор релевантностей признаков целевому вектору.

Для вычисления матрицы \mathbf{Q} и вектора \mathbf{b} определяются следующие функции:

$$\text{Sim} : \mathcal{J} \times \mathcal{J} \rightarrow [0, 1]; \quad (13)$$

$$\text{Rel} : \mathcal{J} \rightarrow [0, 1], \quad (14)$$

где функция Sim отражает схожесть между парой признаков, а функция Rel — релевантность признака целевому вектору.

Тогда матрица \mathbf{Q} вычисляется с использованием функции Sim:

$$\mathbf{Q} = [q_{ij}] = \text{Sim}(\chi_i, \chi_j), \quad (15)$$

а вектор \mathbf{b} — с использованием функции Rel:

$$\mathbf{b} = [b_i] = \text{Rel}(\chi_i). \quad (16)$$

Функции Sim, Rel выбираются в соответствии с решаемой задачей. Далее рассматриваются некоторые примеры таких функций.

Схожесть между парой признаков можно вычислить, используя коэффициент корреляции Пирсона для случайных величин χ_i, χ_j :

$$\rho_{ij} = \frac{\text{cov}(\chi_i, \chi_j)}{\sqrt{\text{var}(\chi_i) \cdot \text{var}(\chi_j)}}, \quad (17)$$

где cov — ковариация случайных величин, var — вариация случайной величины.

Пусть $\bar{\chi}_i = [\bar{\chi}_i \dots \bar{\chi}_i] \in \mathbb{R}^m$ — вектор, значения элементов которого равны среднему $\bar{\chi}_i$ вектора χ_i . Выборочный коэффициент корреляции Пирсона для векторов χ_i, χ_j определяется как:

$$\hat{\rho}_{ij} = \frac{(\chi_i - \bar{\chi}_i)^\top (\chi_j - \bar{\chi}_j)}{\|\chi_i - \bar{\chi}_i\|_2 \|\chi_j - \bar{\chi}_j\|_2} \quad (18)$$

Полагая $\chi_y = \nu$, можно аналогичным образом вычислить выборочный коэффициент корреляции Пирсона $\hat{\rho}_{iy}$ для вектора χ_i значений признака и целевого вектора ν , чтобы измерить релевантность i -го признака целевому вектору.

Таким образом, элементы матрицы $\mathbf{Q} = [q_{ij}]$ и вектора $\mathbf{b} = [b_i]$ вычисляются как:

$$q_{ij} = \text{Sim}(\chi_i, \chi_j) = |\hat{\rho}_{ij}|; \quad (19)$$

$$b_i = \text{Rel}(\chi_i) = |\hat{\rho}_{iy}|. \quad (20)$$

Пусть $p(\chi_i)$ — плотность случайной величины χ_i , $p(\chi_i, \chi_j)$ — совместная плотность случайных величин χ_i и χ_j . Для измерения нелинейной зависимости между случайными величинами χ_i и χ_j применяется взаимная информация:

$$I(\chi_i, \chi_j) = \int \int p(\chi_i, \chi_j) \log \frac{p(\chi_i, \chi_j)}{p(\chi_i)p(\chi_j)} d\chi_i d\chi_j. \quad (21)$$

Выборочная взаимная информация $\hat{I}(\chi_i, \chi_j)$ вычисляется на основе оценок соответствующих плотностей распределений.

Тогда элементы матрицы $\mathbf{Q} = [q_{ij}]$ и вектора $\mathbf{b} = [b_i]$ вычисляются как:

$$q_{ij} = \text{Sim}(\chi_i, \chi_j) = \hat{I}(\chi_i, \chi_j); \quad (22)$$

$$b_i = \text{Rel}(\chi_i) = \hat{I}(\chi_i, \nu). \quad (23)$$

Генетический алгоритм

Изначально имеется "популяция" прогностических моделей одной и той же природы, но построенных с участием различных множеств признаков. Каждая модель характеризуется множеством признаков, которые будут использоваться при ее построении. Назовем это множество признаковым множеством модели. Начальный набор моделей выбирается случайным образом.

Алгоритм состоит из итеративно повторяемых шагов. На каждом шаге происходит отбор, скрещивание и мутация. Обозначим P_k - популяция на шаге k , f — число лучших моделей в популяции, g — число моделей для скрещивания, p — вероятность выбора модели для мутации, $R(\mathbf{f}, D)$ - метрика качества. На каждом шаге алгоритм выполняет следующие действия:

1. Отбор. Из P_k выбирается f моделей с наименьшим значением метрики K .

$$F = f_{min_values}(R(\mathbf{f}|\mathcal{D})). \quad (24)$$

Они добавляются в P_{k+1} .

2. Из множества F случайным образом выбираются g моделей для скрещивания.
3. Скрещивание. Выбранные модели случайным образом разбиваются на пары. Для каждой пары моделей в одной паре происходит процедура "скрещивания": формируется новое множество признаков, в котором каждый признак присутствует, если он есть в признаковых множествах обеих моделей-родителей, не присутствует, если его нет в обоих признаковых множествах моделей-родителей, и присутствует с вероятностью $\frac{1}{2}$, если он есть ровно в одном из признаковых множеств. Модели, полученные с помощью скрещивания (их $\frac{g}{2}$), добавляются в P_{k+1} .
4. Мутация. Каждая модель из P_{k+1} подвергается мутации: каждый элемент ее признакового множества с вероятностью p удаляется из множества, каждый признак не из ее признакового множества с вероятностью p добавляется в множество. Таким образом получается новое признаковое множество. Вероятность p обычно выбирают небольшой.

Такие шаги происходят фиксированное число раз или до выполнения какого-либо заданного критерия останова. Выбор критерия, а также другие детали, зависят от реализации.

Ответом алгоритма является признаковое множество модели с наименьшим значением метрики, полученной на каком-либо шаге алгоритма.

Общее описание задачи анализа ошибки

В данной работе особое внимание уделяется исследованию не только точности, но и устойчивости алгоритма отбора признаков и построенной на отобранном множестве признаков модели.

Устойчивостью модели будем называть отсутствие статистически значимых изменений предсказаний при статистически незначимых изменениях выборки или множества признаков.

Для изучения устойчивости модели предлагается исследовать поведение функции ошибки при изменении выборки.

Введем некоторые обозначения.

Напомним, $\mathcal{A} \subseteq \mathcal{J}$ — некоторое подмножество множества признаков.

$\Theta^* = \arg \min_{\Theta \in \mathbb{R}^{|\mathcal{A}| \times r}} S(\theta, \mathcal{A}|\mathcal{D}, \mathbf{f})$ - предсказанные линейной моделью коэффициенты.

Пусть $\hat{y} = X\Theta \in \mathbb{R}^{m \times r}$ - предсказания линейной модели.

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

N - множество отобранных множеств признаков (см. Методы отбора признаков).

L - множество исследуемых типов моделей (см. Модели)

M - множество метрик качества модели (см. ниже в секции Метрики качества)

Для каждой пары алгоритм-тип модели можно построить модель данного типа, использующую отобранный данным алгоритмом набор признаков.

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 $r < n$. ??? мы этого не делаем вроде, пусть пока лежит.

Для каждого элемента из $N \times L \times K$ считается ошибка линейной модели при применении данного алгоритма отбора признаков. Можно считать, что полученная выборка сгенерирована из некоторого распределения функции ошибки. Задача анализа ошибки состоит в исследовании свойств данного распределения. Более подробно это будет описано в следующей секции (пока не придумала название).

В данной работе особое внимание уделяется исследованию зависимости устойчивости алгоритма от количества отобранных признаков. Для этого каждый алгоритм отбора признаков реализован таким образом, чтобы имелась возможность задавать желаемое количество отобранных признаков как гипер-параметр. Таким образом, в исследовании зависимости устойчивости модели от количества признаков, $N = Alg \times \mathcal{R}_m$, где Alg - множество исследуемых алгоритмов отбора признаков, $\mathcal{R}_m = \{1, 2, \dots, n\}$

Анализ распределения ошибок

Метрики

Характеристикой является сложность

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):

$$\text{MSE}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{1}{m} \sum_{i=1}^m \|\mathbf{y}_i - \hat{\mathbf{y}}_i\|_2^2 \quad (25)$$

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

$$\text{RMSE}(\mathbf{Y}, \hat{\mathbf{Y}}) = \sqrt{\text{MSE}(\mathbf{Y}, \hat{\mathbf{Y}})} \quad (26)$$

Нормированная среднеквадратичная ошибка (scaled mean squared error):

$$\text{sMSE}(\mathbf{Y}, \hat{\mathbf{Y}}) = \frac{\sum_{i=1}^m \|\mathbf{y}_i - \hat{\mathbf{y}}_i\|_2^2}{\sum_{i=1}^m \|\mathbf{y}_i - \bar{\mathbf{y}}\|_2^2}, \quad \bar{\mathbf{y}} = \frac{1}{m} \sum_{i=1}^m \mathbf{y}_i \quad (27)$$

Мультиколлинеарность Предположим, что приз

Число обусловленности матрицы $\mathbf{X}^T \mathbf{X}$

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}} \quad (28)$$

Рассматриваются числа обусловленности для $\mathbf{X}^T \mathbf{X}$, $\hat{\mathbf{Y}}^T \hat{\mathbf{Y}}$. В случае применения модели PLS, будут рассматриваться также числа обусловленности $\mathbf{T}^T \mathbf{T}$

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 // ???.