

# Выбор согласованных моделей для построения нейроинтерфейса

Кулаков Ярослав

February 2021

## 1 Аннотация

В работе решается задача построения нейро-компьютерного интерфейса. Требуется предсказать трехмерную траекторию движения кисти по сигналам с коры головного мозга. Сложность задачи состоит в том, что сигнал высокоразмерный и сильно скоррелированный. Предлагается применить методы снижения размерности исходного пространства с согласованием моделей. Для решения задачи используются линейные и нелинейные модели. Анализируются целевое и латентное пространства, получаемые парой моделей. Экспериментальные результаты подтверждают, что предлагаемый метод повышает качество предсказаний модели.

## 2 Введение

Нейрокомпьютерные интерфейсы (BCI) предоставляют возможность интерпретации активности мозга, использовать и обрабатывать ее моделями машинного обучения с целью предсказания действий или декодирования

мыслей человека [9] [2]. Из-за высокой сложности мозга и большого количества информации, содержащейся в нем в каждый момент времени, высокой размерности описания данных электроэнцефалограммы и электрокортикограммы. Сигналы представляют собой скоррелированные их временные ряды. Для получения нескоррелированных, но информативных признаков, решается задача снижения размерности исходного пространства.[8] [1] [11] В работе исследуются линейные и нелинейные модели декодирования сигналов. Оценивается качество, устойчивость и сложность рассматриваемых моделей.

В статье [4] проведены сравнения алгоритмов выбора признаков: QPFS [4] с LARS [7], Lasso [12], Ridge [6] и отбор признаков с генетическим алгоритмом [10]. Quadratic Programming Feature Selection [4] показывает наилучшие результаты и этот метод можно адаптировать для нашей задачи.

Так же проводится сравнение с методами PLS, PCA, других нелинейных моделей. При решении задачи выбора признаков, одновременно оптимизируются две задачи: минимизируется корреляция между признаками и максимизируется информативность признаков по отношению к таргету. Задача осложняется тем, что признаки и таргеты имеют разную природу.

В данной работе предложен устойчивый алгоритм модулирования сигнала в конечности от состояния активности мозга, состоящий из этапов:

- Построение латентного пространства меньшей размерности, с минимальной корреляцией признаков между собой и максимальной корреляцией признаков с предсказываемым сигналом.
- Построение прогностической модели в полученном пространстве.
- Восстановление обратной зависимости для предсказания активности мозга.

### 3 Постановка задачи

Рассматривается выборка  $(X, Y)$ .  $X \in \mathbb{R}^{T,K}$ , где  $T$  — количество временных отметок, а  $K$  — число электродов, используемых для снятия сигнала.  $Y \in \mathbb{R}^{C,T}$ , где  $C$  — номер координаты в трехмерном пространстве. Данные содержат записи о траектории движения руки в трехмерном пространстве и ECoG сигнала. ECoG сигнал снимался с 64х электродов, частотой 1кГц. Чтобы сформировать тензор признаков, каждая эпоха ECoG была сопоставлена с временно-частотно-пространственным пространством с помощью непрерывного волнового преобразования (CWT).

Снижение размерности. Задача состоит в поиске функций  $\phi : X^{n \times m} \rightarrow X^{n \times k}$  для объектов и функции  $\psi : Y^{n \times p} \rightarrow X^{n \times q}$  для кодирования целевых переменных. Причем  $k < m, q < p$ . Полученные матрицы являются матрицами представлений в латентном пространстве.

Рассматриваются линейные и нелинейные модели. Линейные модели менее подвержены проклятию размерности и слабее переобучаются, а нелинейные методы способны уловить сложные закономерности. Чтобы объединить преимущества обоих методов рассматриваются generalized linear models, additive models и их комбинация — GAM.

Линейная модель задается в виде  $\hat{y} = X^T \theta + \theta_0$ , где  $X$  — матрица признаков, а  $\theta$  — вектор параметров модели. Предполагается, что истинная зависимость так же является линейной, с шумом, распределенным нормально.

GLM — обобщение линейной регрессии, в котором мы можем применять разные функции к  $y$ , а так же предполагать разные его распределения.

AM — еще один способ по внедрению нелинейности в модель.  $\hat{y} = \theta_0 + \sum \theta_i f_i(x_i)$ .

Используемые метрики и критерии качества:  $MSE(\|y - \hat{y}\|_2^2)$ ,  $MAE(\|y -$

$$\hat{y}||_1), MAPE(\frac{1}{n} \sum \frac{|Y_i - \hat{Y}_i|}{Y_i}).$$

Соответственно задача ставится как минимизация этих Loss-function:

$$L(X, Y, \Theta) \rightarrow \min.$$

## 4 Теоритическое обоснование

Определение. Назовём пространство  $\mathbb{T} \subset \mathbb{R}^l$  скрытым пространством для пространства  $\mathbb{X} \in \mathbb{R}^n (l \leq n)$ , если существуют функция  $\varphi_e : \mathbb{X} \rightarrow \mathbb{T}$  и функция  $\varphi_d : \mathbb{T} \rightarrow \mathbb{X}$  такие что

$$\mathbf{x} \in \mathbb{X} \quad \exists \mathbf{t} \in \mathbb{T} : \varphi_d(\varphi_e(\mathbf{x})) = \varphi_d(\mathbf{t}) = \mathbf{x}$$

Функция  $\varphi_e(\mathbf{x})$  называется функцией кодирования объекта  $\mathbf{x}$ , функция  $\varphi_d(\mathbf{t})$  называется функцией декодирования.

Аналогично введём определение скрытого пространства  $\mathbb{U} \subset \mathbb{R}^s$  для целевого пространства  $\mathbb{Y}$ , функции кодирования  $\psi_e : \mathbb{Y} \rightarrow \mathbb{U}$  и декодирования  $\psi_d : \mathbb{U} \rightarrow \mathbb{Y}$

$$\mathbf{y} \in \mathbb{Y} \quad \exists \mathbf{u} \in \mathbb{U} : \psi_d(\psi_e(\mathbf{y})) = \psi_d(\mathbf{u}) = \mathbf{y}$$

Общая схема задачи декодирования принимает вид следующей коммутативной диаграммы:

$$\mathbb{X} \subset \mathbb{R}^n \xrightarrow{f} \mathbb{Y} \subset \mathbb{R}^r$$

$$\varphi_e \downarrow \varphi_d \quad \psi_d \downarrow \psi_e$$

$$\mathbb{T} \subset \mathbb{R}^l \xrightarrow{h} \mathbb{U} \subset \mathbb{R}^s$$

---

**Algorithm 1:** PLS Algorithm

---

**Result:** T,P,Q

**while** *While condition* **do**

$$\mathbf{w}_k := \mathbf{X}_k^\top \mathbf{u}_{k-1} / (\mathbf{u}_{k-1}^\top \mathbf{u}_{k-1}); \quad \mathbf{w}_k := \frac{\mathbf{w}_k}{\|\mathbf{w}_k\|}$$

$$\mathbf{t}_k := \mathbf{X}_k \mathbf{w}_k$$

$$\mathbf{c}_k := \mathbf{Y}_k^\top \mathbf{t}_k / (\mathbf{t}_k^\top \mathbf{t}_k); \quad \mathbf{c}_k := \frac{\mathbf{c}_k}{\|\mathbf{c}_k\|}$$

$$\mathbf{u}_k := \mathbf{Y}_k \mathbf{c}_k$$

**end**

$$\mathbf{p}_k := \mathbf{X}_k^\top \mathbf{t}_k / (\mathbf{t}_k^\top \mathbf{t}_k), \mathbf{q}_k := \mathbf{Y}_k^\top \mathbf{t}_k / (\mathbf{t}_k^\top \mathbf{t}_k)$$

$$\mathbf{X}_{k+1} := \mathbf{X}_k - \mathbf{t}_k \mathbf{p}_k^\top$$

$$\mathbf{Y}_{k+1} := \mathbf{Y}_k - \mathbf{t}_k \mathbf{q}_k^\top$$

---

## 5 Вычислительный эксперимент

### 5.1 Цель

Проверить применимость модели PLS на преобразованных данных и сравнить с моделью линейной регрессии. Визуализировать полученные результаты.

## 5.2 Описание датасета

Датасет состоит из 20-ти записей двух обезьян, которые пытались достать кусочек еды правой рукой. Преобразованные с помощью непрерывного волнового преобразования данные представляют собой тензор  $X \in \mathbb{R}^{T,K,W+1}$ , где  $W$  — размерность для волновых коэффициентов преобразования. Кроме того, к данным добавлен исходная матрица временных рядов для каждого датчика.  $Y$  остается неизменной. Данные подготовлены Анастасией Мотренко и уже подеелны на обучающую и тестовую выборки. Обучающая выборка имеет следующую размерность:  $X : (12801, 32, 27)$ , где 12801 — количество временных отметок, 32 — количество электродов, 27 — количество частот для построения коэффициентов преобразования и еще одно значение, отвечающее напряжению на датчике при фиксированном моменте времени и номере датчика.  $Y : (12801, 3)$ , соответственно для каждой отметки времени имеется три координаты позиции кисти.

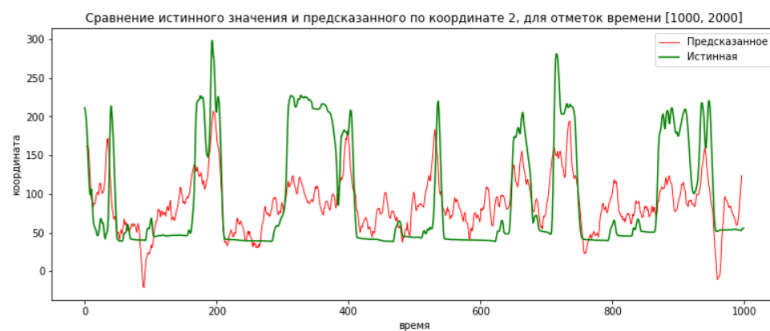
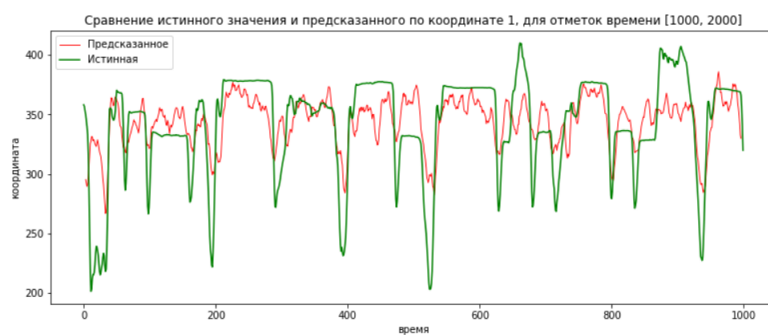
## 5.3 План

Прежде чем применять алгоритмы сгенерируем дополнительно признаки — экспоненцируем существующие. Такая синтезация будет полезной, так как значения признаков слабо отличаются друг от друга и алгоритму сложнее будет искать закономерности. А при экспоненцировании разницы между значениями станут существенными. После этого обучим базовую модель PLS. Опытным путем подберем оптимальную размерность латентного пространства для этого алгоритма. Посмотрим на графики предсказаний и истинных значений по всем трем координатам. Так же построим истинную и предсказанную траекторию в трехмерном пространстве. В конце сравним с результатом работы линейной регрессии и убедимся, что применение PLS было оправданным, размерность и правда очень большая для обычной

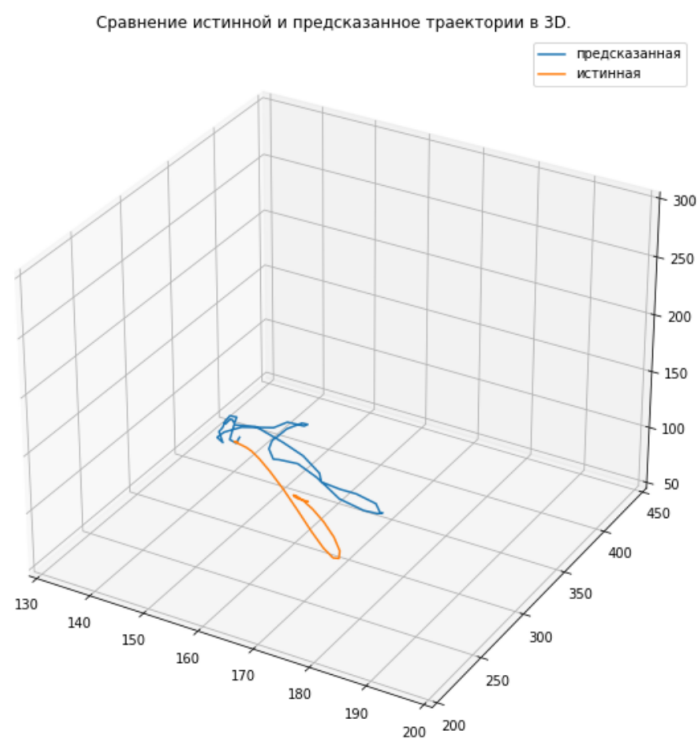
линейной регрессии.

## 5.4 Выполнение

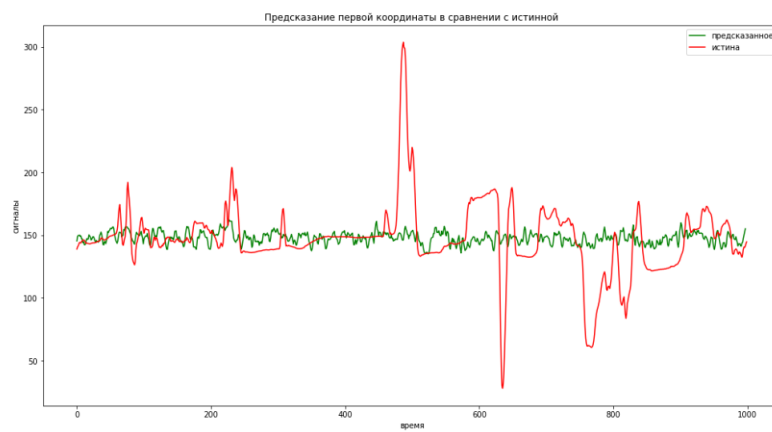
После генерации признаков и применения алгоритма PLS получили на тестовых данных следующие предсказания.



Посмотрим на предсказания в 3D.



И наконец, применим модель линейной регрессии.



Из графиков можно сделать вывод, что уже сейчас, используя только



баловый алгоритм PLS удалось определять основные закономерности и пики в движении. Стоит отметить, что особенно хорошо определялись именно резкие изменения по координатам, а когда кисть не двигалась, предсказания были не точными и шумными. Это не удивительно, ведь когда мозг посылает сигналы руке, он думает не о координате руки, а о силе, которую надо приложить мышце, то есть по второму закону Ньютона, об ускорении руки. Иными словами, при одинаковых сигналах мозга траектория движения руки будет одинаковая, не зависимо от того, находимся ли мы в точке  $(0, 0, 0)$  или в точке  $(100, 100, 100)$ , так что предсказывать напрямую координату по сигналам мозга не совсем правильно. Кроме того, видно, что обычная линейная регрессия не добилаь хоть каких-то результатов. Таким образом имеет смысл использовать алгоритмы понижения размерности. В будущем предполагается оценить предсказания алгоритма по функциям риска, указанных выше.

## 6 Сравнение моделей, анализ ошибок

Рассматривается прогнозирование временного ряда координаты кисти, по прошлым точкам траектории. Предлагается сравнить три модели — *PLS* в чистом виде, *SARIMAX* в чистом виде и их микс. Сравнение происходит по метрикам  $MSE$ ,  $MAE$ .

	MSE	MAE
<b>Clear PLS</b>	840.791	19.0604
<b>Clear SARIMAX</b>	290.792	8.65382
<b>Mix</b>	174.495	7.60459

Как видно, микс моделей показал наилучший результат.

## 7 Заключение

### Список литературы

- [1] Adham Atiyabi, Martin H. Luerksen, and David M.W. Powers. Pso-based dimension reduction of eeg recordings: Implications for subject transfer in bci. *Neurocomputing*, 119:319–331, 2013. Intelligent Processing Techniques for Semantic-based Image and Video Retrieval.
- [2] Benjamin Blankertz, Claudia Sannelli, Sebastian Halder, Eva M. Hammer, Andrea Kübler, Klaus-Robert Müller, Gabriel Curio, and Thorsten Dickhaus. Neurophysiological predictor of smr-based bci performance. *NeuroImage*, 51(4):1303–1309, 2010.
- [3] Aksenova T. Elisseyev A. Stable and artifact-resistant decoding of 3d hand trajectories from ecog signals using the generalized additive model. *Journal of neural engineering.*, 2014.
- [4] Strijov V.V. Isachenko R.V. Quadratic programming optimization with feature selection for non-linear models. *Lobachevskii Journal of Mathematics*, pages 39(9) : 1179–1187., 2018.
- [5] Aleksandr Katrutsa and Vadim Strijov. Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria. *Expert Systems with Applications*, 76, 01 2017.
- [6] Donald W. Marquardt and Ronald D. Snee. Ridge regression in practice. *The American Statistician*, 29(1):3–20, 1975.
- [7] Yoan Miche, Mark van Heeswijk, Patrick Bas, Olli Simula, and Amaury Lendasse. Trop-elm: A double-regularized elm using lars and tikhonov regularization. *Neurocomputing*, 74(16):2413–2421, 2011. Advances in

Extreme Learning Machine: Theory and Applications Biological Inspired Systems. Computational and Ambient Intelligence.

- [8] Strijov V.V. Motrenko A.P. Multi-way feature selection for ecog-based brain-computer interface. *Expert Systems with Applications*, 114(30), pages 402–413., 2018.
- [9] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw. Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on Biomedical Engineering*, 51(6):1034–1043, 2004.
- [10] Feng Tan, Xuezheng Fu, Yanqing Zhang, and Anu G Bourgeois. A genetic algorithm-based method for feature subset selection. *Soft Computing*, 12(2):111–120, 2008.
- [11] L. Vega-Escobar, A. E. Castro-Ospina, and L. Duque-Muñoz. Feature extraction schemes for bci systems. pages 1–6, 2015.
- [12] Peng Zhao and Bin Yu. Stagewise lasso. *The Journal of Machine Learning Research*, 8:2701–2726, 2007.
- [13] Исаченко Р.В. Выбор модели декодирования сигналов в пространствах высокой размерности. *Рукопись*, 2021.
- [14] Стрижов В.В. Яушев Ф.Ю., Исаченко Р.В. Модели согласования скрытого пространства в задаче прогнозирования. *Системы и средства информатики*, page 31(1), 2021.