Choosing agreed models for building a neural interface

Kulakov Yaroslav

February 2021

1 Annotation

The work solves the problem of building a neurocomputer interface. It is required to predict the three-dimensional trajectory of limb movement based on signals from the cerebral cortex. The complexity of the problem lies in the fact that the description of the original signals is redundant and highly correlated. It is proposed to apply methods of reducing the dimensionality of the original space with matching models in the hidden space. To solve the problem, linear and nonlinear agreement models are used. The target and latent spaces obtained by a pair of models are analyzed. Experimental results confirm that the proposed method improves the quality of model predictions.

2 Introduction

Neurocomputer interfaces (BCI) decode brain activity, make it possible to process it with machine learning models in order to predict actions by extracting useful information from the data obtained and presenting it in a human-interpretable form [12] [3]. The descriptions of the EEG and EEC data are high-dimensional

due to the high complexity of the brain and the large amount of information at each point in time. The signals are highly correlated time series. To obtain uncorrelated and informative features, the problem of reducing the dimension of the original space is solved.[11] [2] [15]

The article [7] compares feature selection algorithms. The Quadratic Programming Feature Selection [7] algorithm is considered in comparison withc LARS [10], Lasso [17], Ridge [9] and feature selection with the genetic algorithm [13].

A comparison is made with the methods PLS, PCA[1] [16], other nonlinear models. When solving the problem of feature selection, two functionalities are simultaneously optimized: the correlation between features is minimized and the information content of features in relation to the target is maximized. The task is complicated by the fact that features and targets are of different nature. Therefore, different models are built for different spaces, and the final model is obtained by matching the base ones.

In this paper, linear and nonlinear signal decoding models are investigated. The quality, stability and complexity of the models under consideration are assessed.

In this paper, we propose a robust algorithm for decoding the brain activity signal. The proposed algorithm consists of the following stages:

- construction of a latent space of a lower dimension, with a minimum correlation of features with each other and a maximum correlation of features with a predicted signal;

- building a predictive model in the resulting space;

2

3   Problem statement

The sample is considered $(\mathbf{X}, \mathbf{Y})$. $\mathbf{X} \in \mathbb{R}^{m \times n}, \mathbf{Y} \in \mathbb{R}^{m \times r}$, where $\mathbf{X}$ — time series of electrocorticogram, $\mathbf{Y}$ — time series of hand position in 3D space. Here m — number of timestamps, n — the number of electrodes used to pick up the signal, $r = 3$ — the number of coordinates in three-dimensional space. The data contains records of the trajectory of the hand movement in 3D space and the ECoG signal. It is required to build a pair of agreed models predicting the path of the brush Y_{t+1}, according to the series $X_0 \ldots X_{t+1}$ and $Y_0 \ldots Y_t$, where $\mathbf{X_t}, \mathbf{Y_t}$ — multidimensional feature descriptions of brain activity and coordinates at a moment in time t.

3.1   Regression in the original data space

Consider a family of models $f : \mathbf{X} \to \mathbf{Y}$, where $\mathbf{X} \in \mathbb{R}^{m \times np}$ — the obtained time series of features after the wavelet transformation, and $\mathbf{Y} \in \mathbb{R}^{m \times r}$ — a series of brush paths. Here p — the number of features obtained for each time point and each electrode by the wavelet transform. The problem is to find the model f^*, that minimizes the given error functional $\mathcal{L}$.

$$f^* = \arg\min_{f} \mathcal{L}(f, \mathbf{X}, \mathbf{Y}). \tag{1}$$

We will consider a parametric family of models $f(x, \theta)$, where θ — model parameters. Then the problem of finding a model f^* is equivalent to the problem of finding parameters θ^*:

$$\theta^* = \arg\min_{\theta} \mathcal{L}(\theta, X, Y). \tag{2}$$

The linear regression model is considered as a basic model: $f(\mathbf{X}, \theta) = \theta^T \mathbf{X}$.

3.2 Regression in the target signal space

Regression models of time series are considered $f : \mathbf{Y} \to \mathbf{Y}$. Where $\mathbf{Y} \in \mathbb{R}^{m \times r}$ For each coordinate, consider the time series $\{y_t\}$. The problem is posed of finding a model f^*, that predicts from the last $y_{t'-p} \ldots y_{t'}$ the value $y_{t'+1}$ for all $t' \in \{0, \ldots t\}$, minimizing some error functional $\mathcal{L}$.

$$f^* = \arg\min_f \mathcal{L}(f, \mathbf{Y}). \tag{3}$$

Similarly, the problem of finding a model f^* is equivalent to the problem of finding parameters θ^*.

$$\theta^* = \arg\min_\theta \mathcal{L}(\theta, \mathbf{Y}). \tag{4}$$

Autoregressive model is used as basic predictive models AR, as well as modify it ($ARIMAX$ [14] and etc).

3.3 The problem of high dimensionality and correlation

The high dimension of the space X and the linear dependence of the columns leads to data redundancy and instability of models. Therefore, the problem is posed of finding the functions $\phi : X^n \to T^l$ и $Y^r \to U^s$, representing the original spaces X, Y in hidden spaces of lesser dimension T, U ($l < n, s < r$), maximizing the covariance between the independent and target variables in these spaces. The resulting matrices are matrices of hidden representations in the latent space.

Definition. Let's call the space $\mathbb{T} \subset \mathbb{R}^l$ a hidden space for space $\mathbb{X} \in \mathbb{R}^n (l \leqslant n)$, if there is a function $\varphi_e : \mathbb{X} \to \mathbb{T}$ and function $\varphi_d : \mathbb{T} \to \mathbb{X}$ such that

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

Function $\varphi_e(\mathbf{x})$ called the object encoding function $\mathbf{x}$, function $\varphi_d(\mathbf{t})$ called the

decoding function.

Similarly, we introduce the definition of the hidden space $\mathbb{U} \subset \mathbb{R}^s$ for the target space $\mathbb{Y}$, encoding functions $\psi_e : \mathbb{Y} \to \mathbb{U}$ and decoding $\psi_d : \mathbb{U} \to \mathbb{Y}$

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

The general scheme of the decoding problem takes the form of the following commutative diagrams:

$$
\begin{array}{ccc}
\mathbb{X} \subset \mathbb{R}^n & \xrightarrow{\ f\ } & \mathbb{Y} \subset \mathbb{R}^r \\
\varphi_e \big\Updownarrow \varphi_d & & \psi_d \big\Updownarrow \psi_e \\
\mathbb{T} \subset \mathbb{R}^l & \xrightarrow{\ h\ } & \mathbb{U} \subset \mathbb{R}^s
\end{array}
\tag{5}
$$

Model PLS is used to construct the latent space.

3.4 Model agreement

Definition. Model matching is a method of combining models that predict a target variable across different spaces. Weighted summation is used as a way to match the predictions of the two models. There are two predictions of the coordinate at a time $t+1 - \hat{y}_{PLS,t+1}[2], \hat{y}_{AR,t+1}[4]$. The final prediction will be the weighted sum of these predictions.

$$\hat{y}_{t+1} = \alpha \times \hat{y}_{PLS,t+1} + (1 - \alpha) \times \hat{y}_{AR,t+1}.$$

Где α — the hyperparameter of the supermodel is fitted over the grid, minimizing the error functional.

3.5 Metrics

Let y be the test segment of a multidimensional time series of coordinates, and $\hat{y}$ be the predicted one. $y, \hat{y} \in \mathbb{R}^{m' \times r}$, where m' — segment length. The following external quality criteria were selected in the work:

- $MSE(\|y - \hat{y}\|_2^2)$ — Root mean square error.

- $MAE(\|y - \hat{y}\|_1)$ — Mean absolute error.

4 Theory

4.1 PLS

The pseudocode of the PLS regression method is given in Algorithm 1. The algorithm iteratively calculates one column t_k, u_k, p_k, q_k at each of the l steps from matrices $\mathbf{T}$, $\mathbf{U}$, $\mathbf{P}$, $\mathbf{Q}$ respectively. After calculating the next set of vectors, the next peer-to-peer approximations are subtracted from the matrices X, Y. It is assumed that the original matrices X and Y are normalized (have zero mean and unit mean deviation).

4.2 AR, SARIMAX

The series $\{y_t\}$ is given. Let's fix the parameter p — the number of the last values of the series by which the next prediction will be built. For each value y_t' from the training part of the series, select the p preceding it. According to the resulting matrix $X^{m,p}$ train linear regression $X\theta = Y$, minimizing the functionality MSE.

SARIMAX [14] extends the basic autoregressive model, takes into account the linear dependence on the past values of the series and on the errors of past predictions, and takes into account the seasonality.

Algorithm 1: Algorithm PLS

Require: $\mathbf{X}, \mathbf{Y}, l$;

Ensure: $\mathbf{T}, \mathbf{P}, \mathbf{Q}$;

  1: normalize matrices X and Y by columns

  2: initialize u_0 (first column of matrix $\mathbf{Y}$)

  3: $\mathbf{X}_1 = \mathbf{X}; \mathbf{Y}_1 = \mathbf{Y}$

  4: **for** $k = 1, \ldots, l$ **do**

  5:    **repeat**

  6:       $w_k := \mathbf{X}_k^T u_{k-1} / (u_{k-1}^T u_{k-1}); \quad w_k := \frac{w_k}{\|w_k\|}$

  7:       $t_k := \mathbf{X}_k w_k$

  8:       $c_k := \mathbf{Y}_k^T t_k / (t_k^T t_k); \quad c_k := \frac{c_k}{\|c_k\|}$

  9:       $u_k := \mathbf{Y}_k c_k$

 10:   **until** t_k does not stabilize

 11:   $p_k := \mathbf{X}_k^T t_k / (t_k^T t_k), \; q_k := \mathbf{Y}_k^T t_k / (t_k^T t_k)$

 12:   $\mathbf{X}_{k+1} := X_k - t_k p_k^T$

 13:   $\mathbf{Y}_{k+1} := Y_k - t_k q_k^T$

 14: **end for**
---

5 Experiment

ECoG signal was taken from 64x electrodes, 1 kHz frequency. To form the feature tensor, each ECoG epoch was mapped to the time-frequency-spatial space using a wavelet transform [5], [4].The purpose of the experiment is to analyze the proposed matching procedure.

5.1 Dataset description

The dataset [compiled by A. Motrenko] consists of 20 records of two monkeys trying to reach a piece of food with their right hand. The data transformed using the wavelet transform is a tensor $X \in \mathbb{R}^{T \times K \times W + 1}$, where W — dimension for wave conversion coefficients. In addition, the original time series matrix for each sensor is added to the data. Y remains unchanged. The data were prepared by Anastasia Motrenko and have already been divided into training and test samples. The training sample has the following dimensions: $X : (12801, 32, 27)$,

where $T = 12801$ — number of timestamps, $K = 32$ — number of electrodes, $W = 26$ — the number of frequencies for constructing the conversion factors and one more value corresponding to the voltage on the sensor at a fixed point in time and the number of the sensor. $Y : (12801, 3)$, accordingly, there are three hand position coordinates for each timestamp.
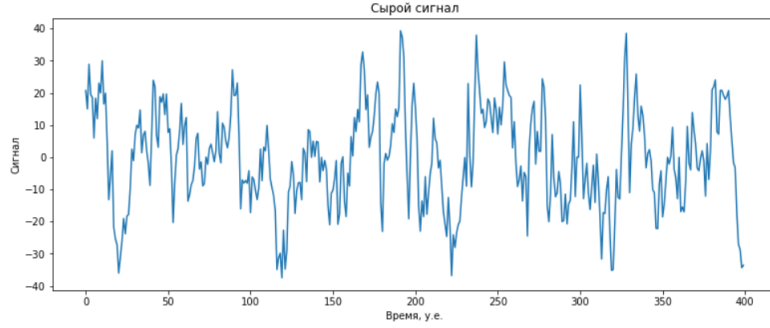

Fig. 0. Visualization of the ECoG signal through one of the channels.

5.2 Model analysis

Using the grid, select the optimal dimension of the latent space for the PLS algorithm. Train PLS, AR, SARIMAX, LR models. Loop through the alpha parameter for agreement. Compare results.

5.3 Performance

After generating features and applying the PLS algorithm on the training data, the following predictions are obtained:
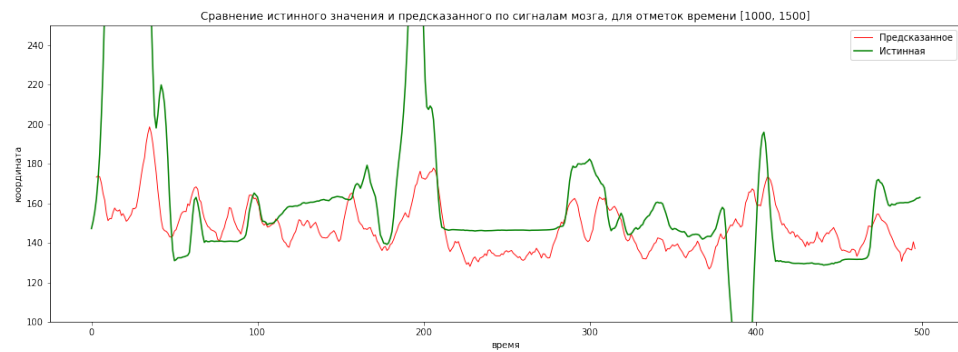
Fig. 1. Coordinate prediction by PLS model from time series of brain activity
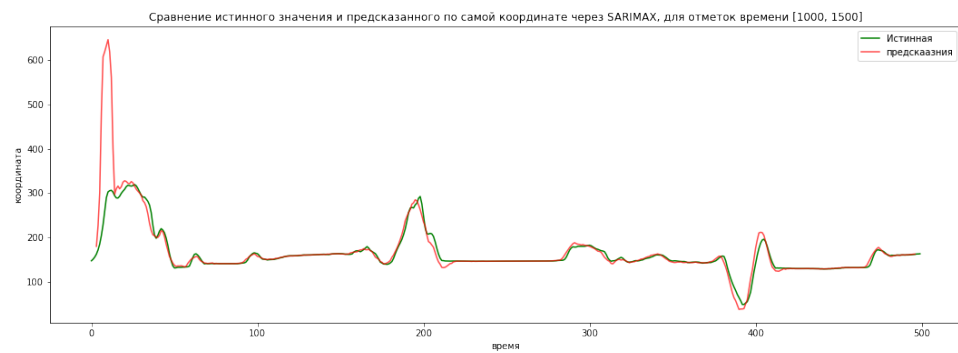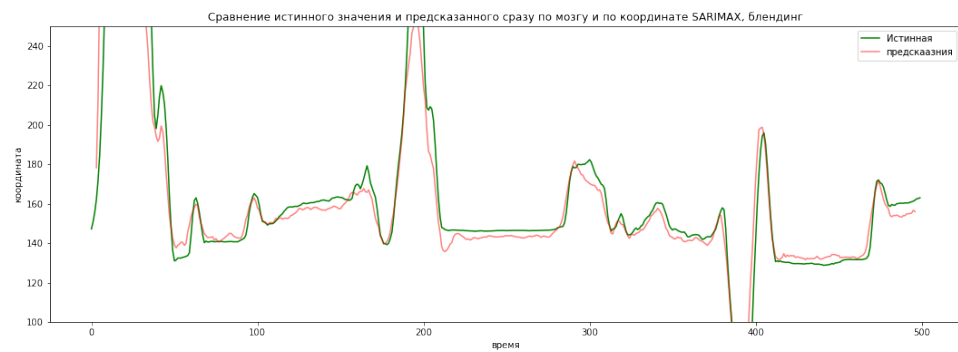
X

Fig. 2. Coordinate prediction by SARIMAX model.

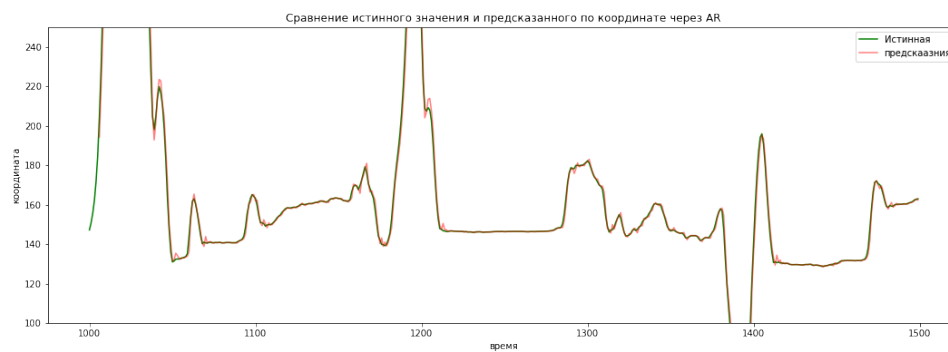Fig. 3. Result of model agreement, $\alpha = 0.86$.
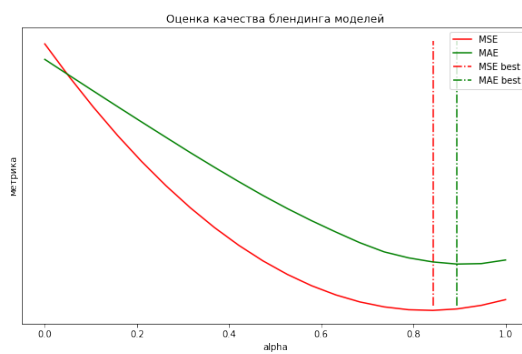

Fig. 4. Coordinate prediction by AR model.

Fig. 5. Error analysis depending on the coefficient $alpha$ of the SARIMAX and PLS models.
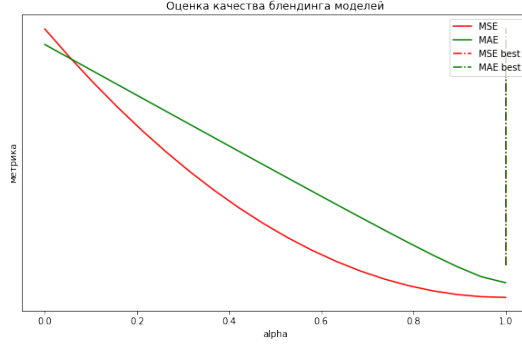
Fig. 6. Error analysis depending on the coefficient $alpha$ of AR and PLS models.

6   Error analysis

Prediction of the time series of the brush coordinates, based on the past points of the trajectory, is considered. It is proposed to compare five models $-$ PLS in its pure form, $SARIMAX$ in its pure form and their mix. And also AR in its pure form and its coordination with PLS. Comparison is based on metrics MSE, MAE.

	MSE	MAE
Clear PLS	791.429	18.7611
Clear SARIMAX(5)	115.063	5.49812
Mix_pls_sarimax	87.8215	5.28448
Clear AR	5.49391	1.21899
Mix_pls_ar	5.49391	1.21899

As you can see, agreement the SARIMAX and PLS models gives a better result than each of them separately. The AR model itself is so good that even

matching with PLS does not improve prediction, it only spoils it. It can be concluded that simple AR is the best model.

7 Conclusion

In the course of the work, various models were proposed and experimentally tested, as well as the agreement of the models. Based on prediction errors, it is argued that prediction by conventional autoregression gives much less error than other models and their combinations. Also it is shown that agreement models gives better results than each of them separately. Thus, a variation of the problem should be considered when, when testing models, the test series of coordinates of past values is not available, but only the series is available during training.

Список литературы

[1] Hervé Abdi. Partial Least Squares (PLS) Regression. *Encyclopedia for research methods for the social sciences*, pages 792–795, 2003.

[2] Adham Atyabi, Martin H. Luerssen, 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.

[3] 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.

[4] Zenas C Chao, Yasuo Nagasaka, and Naotaka Fujii. Long-term

asynchronous decoding of arm motion using electrocorticographic signals in monkey. *Frontiers in neuroengineering*, 3:3, 2010.

[5] Andrey Eliseyev and Tetiana Aksenova. Penalized multi-way partial least squares for smooth trajectory decoding from electrocorticographic (ecog) recording. *PloS one*, 11(5):e0154878, 2016.

[6] Aksenova T. Eliseyev A. Stable and artifact-resistant decoding of 3d hand trajectories from ecog signals using the generalized additive model. *Journal of neural engineering.*, 2014.

[7] 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.

[8] 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.

[9] Donald W. Marquardt and Ronald D. Snee. Ridge regression in practice. *The American Statistician*, 29(1):3–20, 1975.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] Stylianos I. Vagropoulos, G. I. Chouliaras, E. G. Kardakos, C. K. Simoglou, and A. G. Bakirtzis. Comparison of sarimax, sarima, modified sarima and ann-based models for short-term pv generation forecasting. pages 1–6, 2016.

[15] L. Vega-Escobar, A. E. Castro-Ospina, and L. Duque-Muñoz. Feature extraction schemes for bci systems. pages 1–6, 2015.

[16] Jacob A Wegelin et al. A survey of partial least squares (pls) methods, with emphasis on the two-block case. *University of Washington, Department of Statistics, Tech. Rep*, 2000.

[17] Peng Zhao and Bin Yu. Stagewise lasso. *The Journal of Machine Learning Research*, 8:2701–2726, 2007.

[18] Исаченко Р.В. Выбор модели декодирования сигналов в пространствах высокой размерности. *Рукопись*, 2021.

[19] Стрижов В.В. Яушев Ф.Ю., Исаченко Р.В. Модели согласования скрытого пространства в задаче прогнозирования. *Системы и средства информатики*, page 31(1), 2021.