Feature Selection for Miltivariate Correlated ECoG-based data

Roman Isachenko

Skoltech advisor: Maxim Fedorov

MIPT advisor: Vadim Strijov

May 16, 2018.

Signal decoding problem

Goal

- Investigate dependencies in input and target spaces for signal decoding problem.
- Build a stable model for time series decoding in the case of multicorrelated object description.

Challenge

- Measurements of spatial-temporal data are interacted since sensors are close to each other.
- Target variable is a vector whose elements are dependent.

Solution

Propose feature selection algorithms which take into account dependencies in both input and target spaces.

Related works

- Katrutsa A., Strijov V. Comprehensive study of feature selection methods to solve multicollinearity problem according to evaluation criteria // *Expert Systems with Applications* 76, 2017.

- Li J. et al. Feature selection: A data perspective //*ACM Computing Surveys (CSUR)* 50(6), 2017.

- Eliseyev A. et al. Iterative N-way partial least squares for a binary self-paced brain–computer interface in freely moving animals //*Journal of neural engineering* 4(8), 2011.

- Rodriguez-Lujan I. et al. Quadratic programming feature selection // *Journal of Machine Learning Research* 11(Apr), 2010.

- Motrenko A., Strijov V. Multi-way Feature Selection for ECoG-based Brain-Computer Interface // *Expert Systems with Applications* Submitted to the journal.

Application: Brain Computer Interface (BCI)

Aim

Develop systems to help people with a severe motor control disability recover mobility.

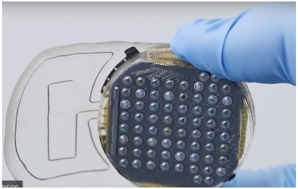
Hypothesis

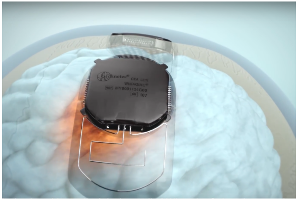
"When we imagine making a movement, we trigger the same electrical activity in the motor cortex of the brain as when we actually perform that activity."*

Solution

Record electrical signals – electrocorticograms (ECoG), decode them to drive complex objects, for example, to move the limbs of an exoskeleton.

* `http://clinatec.fr`





Eliseyev A. et al. CLINATEC BCI platform based on the ECoG-recording implant WIMAGINE and the innovative signal-processing: preclinical results, 2014.

Multivariate regression

Given

Dataset $(\mathbf{X}, \mathbf{Y})$, design matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$, target matrix $\mathbf{Y} \in \mathbb{R}^{m\times r}$,

$$\mathbf{X} = [\chi_1, \dots, \chi_n]; \quad \mathbf{Y} = [\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_r].$$

Model

Forecast a dependent variable $\mathbf{y} \in \mathbb{R}^r$ from an independent input object $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{y} = \boldsymbol{\Theta}\mathbf{x} + \varepsilon, \quad \boldsymbol{\Theta} \in \mathbb{R}^{r\times n}.$$

Loss function

$$\mathcal{L}(\boldsymbol{\Theta}|\mathbf{X}, \mathbf{Y}) = \left\| \underset{m\times r}{\mathbf{Y}} - \underset{m\times n}{\mathbf{X}} \cdot \underset{r\times n}{\boldsymbol{\Theta}}^{\mathsf{T}} \right\|_2^2 \to \min_{\boldsymbol{\Theta}}.$$

$$\boldsymbol{\Theta}^{\mathsf{T}} = (\mathbf{X}^{\mathsf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathsf{T}}\mathbf{Y}.$$

The linear dependent columns of the matrix $\mathbf{X}$ leads to an instable solution.
To avoid the strong linear dependence, feature selection techniques are used.

Feature selection problem

Goal

Find a boolean vector $\mathbf{a} = \{0, 1\}^n$ of indicators for selected features.

Feature selection error function

$$\mathbf{a} = \arg\min_{\mathbf{a}' \in \{0,1\}^n} S(\mathbf{a}' | \mathbf{X}, \mathbf{Y}).$$

Relaxed problem

From discrete domain $\{0, 1\}^n$ to continuous relaxation $[0, 1]^n$:

$$\mathbf{z} = \arg\min_{\mathbf{z}' \in [0,1]^n} S(\mathbf{z}' | \mathbf{X}, \mathbf{Y}), \quad a_j = [z_j > \tau].$$

Once the solution $\mathbf{a}$ is known:

$$\mathcal{L}(\mathbf{\Theta}_\mathbf{a} | \mathbf{X}_\mathbf{a}, \mathbf{Y}) = \left\| \mathbf{Y} - \mathbf{X}_\mathbf{a} \mathbf{\Theta}_\mathbf{a}^\mathsf{T} \right\|_2^2 \to \min_{\mathbf{\Theta}_\mathbf{a}},$$

where the subscript $\mathbf{a}$ indicates the submatrix with the columns for which $a_j = 1$.

Quadratic Programming Feature Selection

$$\|\boldsymbol{\nu} - \mathbf{X}\boldsymbol{\theta}\|_2^2 \to \min_{\boldsymbol{\theta} \in \mathbb{R}^n}.$$

Quadratic programming problem

$$S(\mathbf{z}'|\mathbf{X}, \boldsymbol{\nu}) = (1-\alpha) \cdot \underbrace{\mathbf{z}^{\mathsf{T}}\mathbf{Q}\mathbf{z}}_{\mathrm{Sim}(\mathbf{X})} - \alpha \cdot \underbrace{\mathbf{b}^{\mathsf{T}}\mathbf{z}}_{\mathrm{Rel}(\mathbf{X},\boldsymbol{\nu})} \to \min_{\substack{\mathbf{z} \geq \mathbf{0}_n \\ \mathbf{1}_n^{\mathsf{T}}\mathbf{z}=1}}.$$

- $\mathbf{z} \in [0,1]^n$ – feature importances;
- $\mathbf{Q} \in \mathbb{R}^{n \times n}$ – pairwise feature similarities;
- $\mathbf{b} \in \mathbb{R}^n$ – feature relevances to the target vector.

$$\mathbf{Q} = \left[\left|\mathrm{corr}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)\right|\right]_{i,j=1}^n, \quad \mathbf{b} = \left[\left|\mathrm{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu})\right|\right]_{i=1}^n.$$

Statement

In the case of semidefinite matrix $\mathbf{Q}$ the QPFS problem is convex. Shift spectrum for semidefinite relaxation:

$$\mathbf{Q} \to \mathbf{Q} - \lambda_{\min}\mathbf{I}.$$

Multivariate QPFS

Feature selection error function

$$\mathbf{z} = \mathop{\arg\min}_{\mathbf{z}' \in [0,1]^n} S(\mathbf{z}' | \mathbf{X}, \mathbf{Y}) = \mathop{\arg\min}_{\mathbf{z} \geq \mathbf{0}_n,\, \mathbf{1}_n^{\mathsf{T}} \mathbf{z} = 1} \left[(1 - \alpha) \cdot \mathbf{z}^{\mathsf{T}} \mathbf{Q} \mathbf{z} - \alpha \cdot \mathbf{b}^{\mathsf{T}} \mathbf{z} \right].$$

Relevance Aggregation (RelAgg)

$$\mathbf{b} = [|\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu})|]_{i=1}^n \to \mathbf{b} = \left[\sum_{k=1}^r |\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu}_k)| \right]_{i=1}^n.$$

This approach does not use the dependencies in the columns of the matrix $\mathbf{Y}$.

Symmetric Importance (SymImp)

Penalize correlated targets by Sim($\mathbf{Y}$)

$$\alpha_1 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{Q}_x \mathbf{z}_x}_{\text{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{z}_x^{\mathsf{T}} \mathbf{B} \mathbf{z}_y}_{\text{Rel}(\mathbf{X},\mathbf{Y})} + \alpha_3 \cdot \underbrace{\mathbf{z}_y^{\mathsf{T}} \mathbf{Q}_y \mathbf{z}_y}_{\text{Sim}(\mathbf{Y})} \to \min_{\substack{\mathbf{z}_x \geq \mathbf{0}_n,\, \mathbf{1}_n^{\mathsf{T}} \mathbf{z}_x = 1 \\ \mathbf{z}_y \geq \mathbf{0}_r,\, \mathbf{1}_r^{\mathsf{T}} \mathbf{z}_y = 1}} .$$

$$\mathbf{Q}_x = \left[\left| \text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j) \right| \right]_{i,j=1}^{n}, \ \mathbf{Q}_y = \left[\left| \text{corr}(\boldsymbol{\nu}_i, \boldsymbol{\nu}_j) \right| \right]_{i,j=1}^{r}, \ \mathbf{B} = \left[\left| \text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu}_j) \right| \right]_{\substack{i=1,\dots,n \\ j=1,\dots,r}} .$$

$$\alpha_1 + \alpha_2 + \alpha_3 = 1 \quad \alpha_i \geq 0, \ i = 1, 2, 3.$$

Statement

Balance between Sim($\mathbf{X}$), Rel($\mathbf{X}, \mathbf{Y}$), and Sim($\mathbf{Y}$) is achieved by:

$$\alpha_1 \propto \overline{\mathbf{Q}}_y \overline{\mathbf{B}}; \quad \alpha_2 \propto \overline{\mathbf{Q}}_x \overline{\mathbf{Q}}_y; \quad \alpha_3 \propto \overline{\mathbf{Q}}_x \overline{\mathbf{B}},$$

where $\overline{\mathbf{Q}}_x$, $\overline{\mathbf{B}}$, $\overline{\mathbf{Q}}_y$ are mean values of $\mathbf{Q}_x$, $\mathbf{B}$, and $\mathbf{Q}_y$, respectively.

MinMax / MaxMin

SymImp penalizes targets that are correlated and are not sufficiently explained by the features.

$$\alpha_1 \cdot \underbrace{\mathbf{z}_x^\mathsf{T} \mathbf{Q}_x \mathbf{z}_x}_{\mathrm{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{z}_x^\mathsf{T} \mathbf{B} \mathbf{z}_y}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} \to \min_{\substack{\mathbf{z}_x \geq \mathbf{0}_n, \\ \mathbf{1}_n^\mathsf{T} \mathbf{z}_x = 1}}; \quad \alpha_3 \cdot \underbrace{\mathbf{z}_y^\mathsf{T} \mathbf{Q}_y \mathbf{z}_y}_{\mathrm{Sim}(\mathbf{Y})} + \alpha_2 \cdot \underbrace{\mathbf{z}_x^\mathsf{T} \mathbf{B} \mathbf{z}_y}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} \to \min_{\substack{\mathbf{z}_y \geq \mathbf{0}_r, \\ \mathbf{1}_r^\mathsf{T} \mathbf{z}_y = 1}}.$$

Minimax problem

$$\min_{\substack{\mathbf{z}_x \geq \mathbf{0}_n \\ \mathbf{1}_n^\mathsf{T} \mathbf{z}_x = 1}} \max_{\substack{\mathbf{z}_y \geq \mathbf{0}_r \\ \mathbf{1}_r^\mathsf{T} \mathbf{z}_y = 1}} \left(\text{or } \max_{\substack{\mathbf{z}_y \geq \mathbf{0}_r \\ \mathbf{1}_r^\mathsf{T} \mathbf{z}_y = 1}} \min_{\substack{\mathbf{z}_x \geq \mathbf{0}_n \\ \mathbf{1}_n^\mathsf{T} \mathbf{z}_x = 1}} \right) \left[\alpha_1 \cdot \underbrace{\mathbf{z}_x^\mathsf{T} \mathbf{Q}_x \mathbf{z}_x}_{\mathrm{Sim}(\mathbf{X})} - \alpha_2 \cdot \underbrace{\mathbf{z}_x^\mathsf{T} \mathbf{B} \mathbf{z}_y}_{\mathrm{Rel}(\mathbf{X},\mathbf{Y})} - \alpha_3 \cdot \underbrace{\mathbf{z}_y^\mathsf{T} \mathbf{Q}_y \mathbf{z}_y}_{\mathrm{Sim}(\mathbf{Y})} \right].$$

Theorem

For positive definite matrices $\mathbf{Q}_x$ and $\mathbf{Q}_y$ the max-min and min-max problems have the same optimal value.

Theorem

Minimax problem is equivalent to the quadratic problem with $n + r + 1$ variables.

Shift spectrum to obtain the convex problem.

Maximum Relevance (MaxRel)

Drop the term $\text{Sim}(\mathbf{Y})$.

Minimax problem

$$\min_{\substack{\mathbf{z}_x \geq \mathbf{0}_n \\ \mathbf{1}_n^\mathsf{T}\mathbf{z}_x = 1}} \max_{\substack{\mathbf{z}_y \geq \mathbf{0}_r \\ \mathbf{1}_r^\mathsf{T}\mathbf{z}_y = 1}} \left[(1-\alpha) \cdot \mathbf{z}_x^\mathsf{T} \mathbf{Q}_x \mathbf{z}_x - \alpha \cdot \mathbf{z}_x^\mathsf{T} \mathbf{B} \mathbf{z}_y\right].$$

Theorem

For positive definite matrices $\mathbf{Q}_x$ the max-min and min-max problems have the same optimal value and the final quadratic problem is convex.

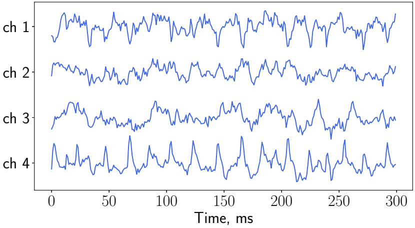
Statement

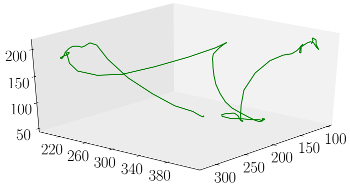
For the case $r = 1$ the proposed strategies SymImp, MinMax, MaxMin, MaxRel coincide with the original QPFS algorithm.

Summary

Algorithm	Strategy	Error function $S(\mathbf{z}\|\mathbf{X},\mathbf{Y})$
RelAgg	$\min\left[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y})\right]$	$\min_{\mathbf{z}_x}\left[(1-\alpha)\cdot \mathbf{z}_x^\top \mathbf{Q}_x \mathbf{z}_x - \alpha \cdot \mathbf{z}_x^\top \mathbf{B}\mathbf{1}_r\right]$
SymImp	$\min\left[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y}) + \mathrm{Sim}(\mathbf{Y})\right]$	$\min_{\mathbf{z}_x,\,\mathbf{z}_y}\left[\alpha_1 \cdot \mathbf{z}_x^\top \mathbf{Q}_x \mathbf{z}_x - \alpha_2 \cdot \mathbf{z}_x^\top \mathbf{B}\mathbf{z}_y + \alpha_3 \cdot \mathbf{z}_y^\top \mathbf{Q}_y \mathbf{z}_y\right]$
MinMax	$\min\left[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y})\right]$ $\max\left[\mathrm{Rel}(\mathbf{X},\mathbf{Y}) + \mathrm{Sim}(\mathbf{Y})\right]$	$\min_{\mathbf{z}_x}\max_{\mathbf{z}_y}\left[\alpha_1 \cdot \mathbf{z}_x^\top \mathbf{Q}_x \mathbf{z}_x - \alpha_2 \cdot \mathbf{z}_x^\top \mathbf{B}\mathbf{z}_y - \alpha_3 \cdot \mathbf{z}_y^\top \mathbf{Q}_y \mathbf{z}_y\right]$
MaxMin	$\max\left[\mathrm{Rel}(\mathbf{X},\mathbf{Y}) + \mathrm{Sim}(\mathbf{Y})\right]$ $\min\left[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y})\right]$	$\max_{\mathbf{z}_y}\min_{\mathbf{z}_x}\left[\alpha_1 \cdot \mathbf{z}_x^\top \mathbf{Q}_x \mathbf{z}_x - \alpha_2 \cdot \mathbf{z}_x^\top \mathbf{B}\mathbf{z}_y - \alpha_3 \cdot \mathbf{z}_y^\top \mathbf{Q}_y \mathbf{z}_y\right]$
MaxRel	$\min\left[\mathrm{Sim}(\mathbf{X}) - \mathrm{Rel}(\mathbf{X},\mathbf{Y})\right]$ $\max\left[\mathrm{Rel}(\mathbf{X},\mathbf{Y})\right]$	$\min_{\mathbf{z}_x}\max_{\mathbf{z}_y}\left[(1-\alpha)\cdot \mathbf{z}_x^\top \mathbf{Q}_x \mathbf{z}_x - \alpha \cdot \mathbf{z}_x^\top \mathbf{B}\mathbf{z}_y\right]$

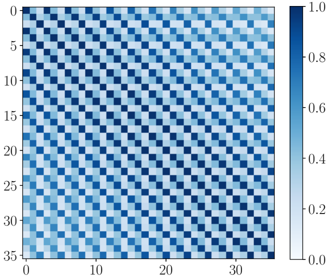
Computational experiment





$$\mathbf{X} \in \mathbb{R}^{m\times(32\cdot 27)};\quad \mathbf{Y} \in \mathbb{R}^{m\times 3k}.$$

$$\mathbf{Y} = \begin{pmatrix} x_1\ y_1\ z_1 & \dots & x_k & y_k & z_k \\ x_2\ y_2\ z_2 & \dots & x_{k+1} & y_{k+1} & z_{k+1} \\ \dots & \dots & & \dots & \\ x_m\ y_m\ z_m & \dots & x_{m+k} & y_{m+k} & z_{m+k} \end{pmatrix}$$



Y correlation matrix

---

http://neurotycho.org

Metrics

RMSE

The prediction quality:

$$\text{RMSE}(\mathbf{Y}, \widehat{\mathbf{Y}}_\mathbf{a}) = \sqrt{\text{MSE}(\mathbf{Y}, \widehat{\mathbf{Y}}_\mathbf{a})} = \|\mathbf{Y} - \widehat{\mathbf{Y}}_\mathbf{a}\|_2, \quad \text{where} \quad \widehat{\mathbf{Y}}_\mathbf{a} = \mathbf{X}_\mathbf{a}\mathbf{\Theta}_\mathbf{a}^\mathsf{T}.$$

Multicorrelation

Mean value of miltiple correlation coefficient:

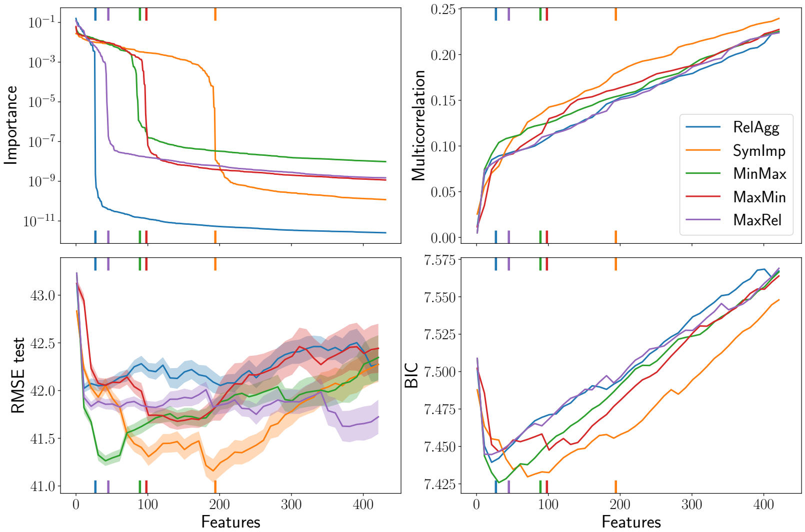
$$R^2 = \frac{1}{r}\text{tr}\left(\mathbf{C}^\mathsf{T}\mathbf{R}^{-1}\mathbf{C}\right); \quad \mathbf{C} = [\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\nu}_j)]_{\substack{i=1,\dots,n,\\ j=1,\dots,r}} \ \mathbf{R} = [\text{corr}(\boldsymbol{\chi}_i, \boldsymbol{\chi}_j)]_{i,j=1}^n.$$

BIC

Bayesian Information Criteria is a trade-off between prediction quality and the number of selected features $\|\mathbf{a}\|_0$:

$$\text{BIC} = m \ln\left(\text{MSE}(\mathbf{Y}, \widehat{\mathbf{Y}}_\mathcal{A})\right) + \|\mathbf{a}\|_0 \cdot \log m.$$

Metrics evaluation



Stability of selected feature subsets

Experiment design

- generate bootstrap data

$$(\mathbf{X}, \mathbf{Y}) \to \{(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_s, \mathbf{Y}_s)\};$$

- solve feature selection problem

$$\{(\mathbf{X}_1, \mathbf{Y}_1), \dots, (\mathbf{X}_s, \mathbf{Y}_s)\} \to \{\mathbf{z}_1, \dots, \mathbf{z}_s\};$$

- calculate statistics

$$\{\mathbf{z}_1, \dots, \mathbf{z}_s\} \to \{\text{RMSE}, \|\mathbf{a}\|_0, \text{Spearman } \rho, \ell_2 \text{ dist}\}.$$

	RMSE	$\|\mathbf{a}\|_0$	Spearman ρ	ℓ_2 dist
RelAgg	41.9 ± 0.1	27.0 ± 2.4	0.941 ± 0.005	0.14 ± 0.02
SymImp	41.1 ± 0.1	198.6 ± 3.8	0.942 ± 0.008	0.03 ± 0.00
MinMax	41.4 ± 0.2	92.4 ± 8.2	0.933 ± 0.007	0.10 ± 0.01
MaxMin	41.7 ± 0.1	97.0 ± 4.4	0.950 ± 0.004	0.07 ± 0.01
MaxRel	41.7 ± 0.0	37.6 ± 1.6	0.893 ± 0.012	0.17 ± 0.02

Conclusion

- BCI signal decoding problem is investigated.
- Feature selection algorithms for multivariate spatio-temporal data are proposed.
- Suggested algorithms are explored and compared.

Publications

- Isachenko R. et al. Feature Generation for Physical Activity Classification. *Artificial Intellegence and Decision Making*, submitted.
- Isachenko R., Strijov V. Quadratic programming optimization for Newton method. *Lobachevskii Journal of Mathematics*, accepted.
- Isachenko R., Vladimirova M., Strijov V. Dimensionality reduction for time series decoding and forecasting problems. Ready for submission.