Research Portfolio - PhD Application

Thanh Tu, Do

February 1, 2024

Outline

1. Self-Introduction

2. Research Experience
   - Publications
   - Signal Processing
   - Missing Data Imputation
   - Computer vision
   - Simulation-based Inference

3. Research Interest

Education and Professional Affiliation

Education

Vietnam National University, Ho Chi Minh University of Science
Master Student - Falcuty of Mathematics and Computer Science (*Jan 2022 - now*)

Foreign Trade University of Vietnam, Hanoi Campus
Bachelor of International Business and Economics (*July 2011 - May 2015*)

Working experience

Vigo Retail Ho Chi Minh City, Vietnam
Data Scientist (*Jan 2023 - now*)

- Designed and implemented recommendation algorithm to personalize use experience using Torch Lightning framework.

Outline

1 Self-Introduction

2 Research Experience
- Publications
- Signal Processing
- Missing Data Imputation
- Computer vision
- Simulation-based Inference

3 Research Interest

Accepted

- (**BME 2020**) **Tu, Do Thanh**, Thuong Nguyen, Anh Tho Le, Sinh Nguyen, Huong Ha. *"Automated EOG removal from EEG signal using Independent Component Analysis and Machine Learning Algorithms"* at The 8th International Conference in Vietnam on the Development of Biomedical Engineering.
- (**ICHST 2023**) **Tu, Do Thanh**, Luan Van Tran, Tho Anh Le, Thao Mai Thi Le, Lan-Anh Hoang Duong, Thuong Hoai Nguyen, Anh Minh Hoang An, Duy The Phan, Khiet Thu Thi Dang, Quyen Hoang Quoc Vo, Nam Phuong Nguyen, Huong Thanh Thi Ha. *"Stress prediction using machine-learning technique on physiological signal"*

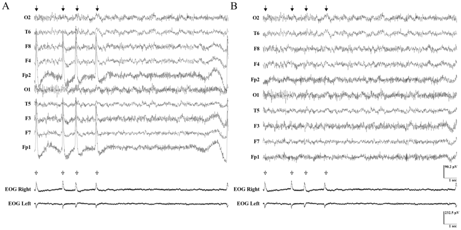
Submitted

- Mai Anh Vu*, Thu Nguyen*, **Tu T. Do***, Nhan Phan, Nitesh V. Chawla, Pål Halvorsen, Michael A. Riegler and Binh T. Nguyen. *"Conditional expectation with regularization for missing data imputation"*
- **Tu T. Do**, Mai Anh Vu, Hoang Thien Ly, Thu Nguyen, Steven A. Hicks, Michael A. Riegler, Pål Halvorsen Halvorsen and Binh T. Nguyen. *"Blockwise Principal Component Analysis for monotone missing data imputation and dimensionality reduction"*

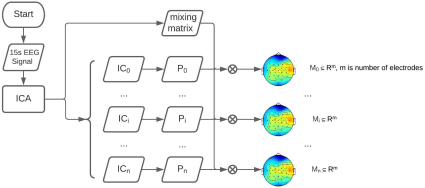
Publications II

- **Tu T. Do**, Mai Anh Vu, Hoang Thien Ly, Thu Nguyen, Steven A. Hicks, Michael A. Riegler, Pål Halvorsen Halvorsen and Binh T. Nguyen. *"Estimating lower-dimensional space representation in Principal Component Analysis under missing data condition"*

Automated EOG removal from EEG signal using Independent Component Analysis and Machine Learning Algorithms.
Supervisor: Dr. Huong Ha, Brain Health Lab.



- Worked on data analysis and visualization: Visualizing the topographical map of EEG signal's power on the scalp.
- Main idea: train a model to classify the topological map of each IC to identify whether the the IC represent ocular activity.



Missing Data Imputation I

Low-dimension Representation Estimation in Principal Component Analysis under Missing Data

Supervisor: Dr. Thu Nguyen, SimulaMet

Utilizing the estimation of the covariance matrix, we can compute the projection matrix $\mathbf{V}$, and estimated the missing entries $\mathbf{x}_m$ using conditional Gaussian Expectation given observed value $\mathbf{x}_o$

$$\hat{\mathbf{x}}_m = \mathbb{E}[\mathbf{x}_o] + \mathbf{\Sigma}_{om}\mathbf{\Sigma}_o^{-1}(\mathbf{x}_o - \mathbb{E}[\mathbf{x}_m]) \tag{11}$$

Recall that $\mathbf{x}$ is centered, so that $\mathbb{E}[\mathbf{x}_o]$ and $\mathbb{E}[\mathbf{x}_m]$ are $\mathbf{0}$, hence,

$$\hat{\mathbf{x}}_m = \mathbf{\Sigma}_{om}\mathbf{\Sigma}_o^{-1}\mathbf{x}_o. \tag{12}$$

Finally, the projection can be estimated by transforming sample $\mathbf{x}$ by a linear transformation $\mathbf{V}^\top$

$$\begin{aligned}\mathbf{V}^\top\mathbf{x} &= (\mathbf{V}_o^\top, \mathbf{V}_m^\top)\begin{pmatrix}\mathbf{x}_o \\ \mathbf{x}_m\end{pmatrix} \\ &= \mathbf{V}_o^\top\mathbf{x}_o + \mathbf{V}_m^\top\mathbf{x}_m \\ &= \mathbf{V}_o^\top\mathbf{x}_o + \mathbf{V}_m^\top\mathbf{\Sigma}_{om}\mathbf{\Sigma}_o^{-1}\mathbf{x}_o \\ &= (\mathbf{V}_o^\top + \mathbf{V}_m^\top\mathbf{\Sigma}_{om}\mathbf{\Sigma}_o^{-1})\mathbf{x}_o.\end{aligned} \tag{13}$$

Classifier guided MCMC for imbalance learning problem

- Utilized MCMC to oversample minority classes in imbalance learning problem.
- Implemented and compare proposed method against other baseline methods.

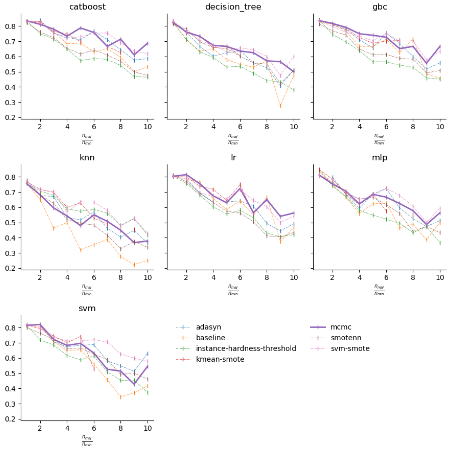
MCMC

To sample from from class c_i with probability density function $p(X|y=c_i)$, we can utilize Markov Chain Monte Carlo, we need quantity:

$$H = \frac{p(x|y=c_i)}{p(x_t|y=c_i)} = \frac{p(x,y)/p(y)}{p(x_t,y)/p(y)} = \frac{p(y|x)}{p(y|x_t)} \times \frac{p(x)}{p(x_t)}$$

The first term $p(y|x)$, we can train a classifier to approximate this quantity. The choice of $d_\theta(.)$ is very flexible, can be `scikit-learn` implementation of `LogisticRegression()` for binary classification problem or a simple fully connected network.

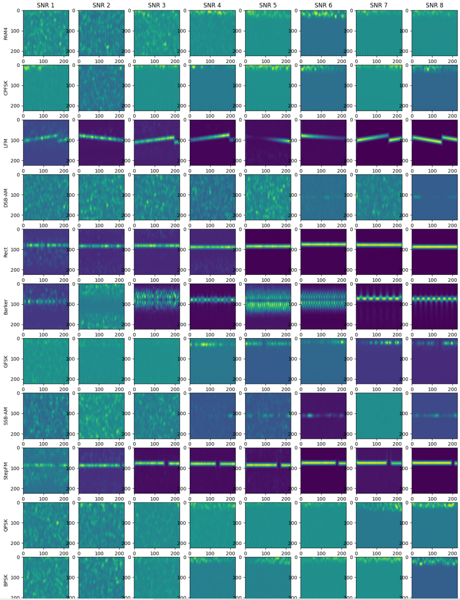
$$p(y|x) = d_\theta(x)$$



Computer vision

Wave form classfication using Deep Convolutional Neural Network

- Implemented baseline method Racomnet (previous work)
- Implemented and train various common architecture on given dataset, namely Vision Transformer, EfficientNet, MobileNet, etc.
- Proposed pretraining appoach using self-supervised approach, namely combine supervised loss and self-supervised loss (Barlow-Twin loss, Constrastive Loss)



Simulation-based Inference

- Studied the problem of Simulation Based Inference (SBI)
- Surveyed current methods, namely Approximate Bayesian Computation,Likelihood-free MCMC with Amortized Likelihood Ratio Estimator (AMCMC).
- Proposed to use Iterative AutoEncoder Dynamics to sample from posterior distribution.

Adapting MCMC for SBI task

We want to sample from $p(\theta|\mathbf{x})$ using MCMC, we need this quantity

$$\frac{p(\theta|\mathbf{x})}{p(\theta_t|\mathbf{x})} = \frac{p(\theta)p(\mathbf{x}|\theta)/p(\mathbf{x})}{p(\theta_t)p(\mathbf{x}|\theta_t)/p(\mathbf{x})} = \frac{p(\theta)}{p(\theta_t)} \times \frac{p(\mathbf{x}|\theta)}{p(\mathbf{x}|\theta_t)} = \frac{p(\theta)}{p(\theta_t)} \times r(\mathbf{x}|\theta, \theta_t) \quad (1)$$