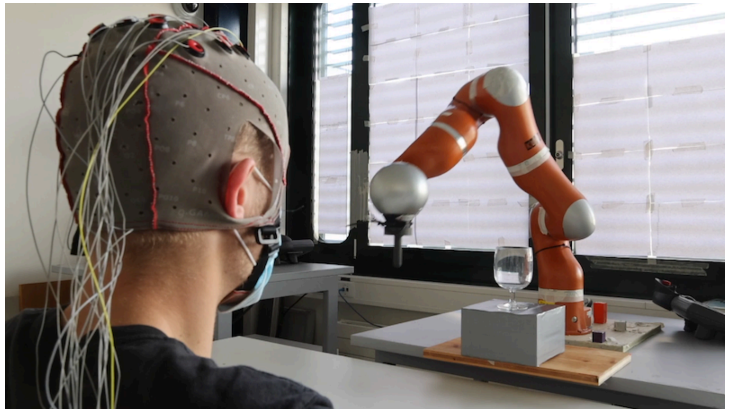
Hyper-Accelerated Learning for Brain-Computer Interfaces via Partial Target-Aware Optimal Transport

Ekansh Gupta, Cheng-Yeh Chen, Raghupathy Sivakumar
Georgia Institute of Technology
Atlanta GA, USA

Brain-Computer Interfaces (BCIs)

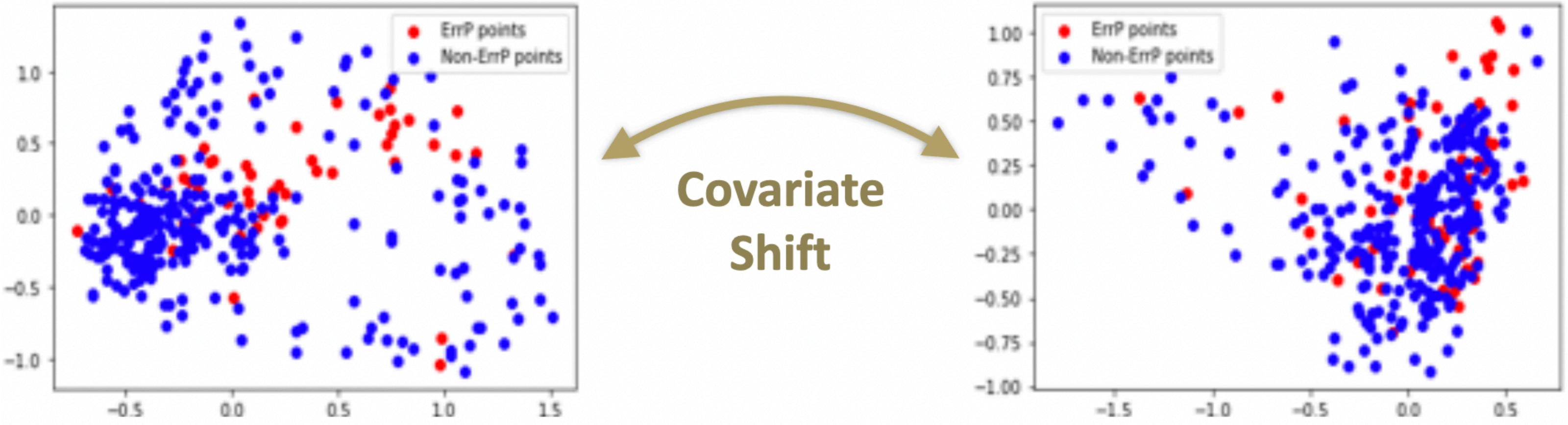
- A powerful modality for human-machine communication
- Useful for exciting and futuristic applications for wearables like robot control, gaming, virtual reality, etc.



- BCIs can be invasive or non-invasive
  - Invasive BCIs require specialized surgery but measure clear signal
  - Non-invasive BCIs can be deployed widely but measure noisy signal

Challenges

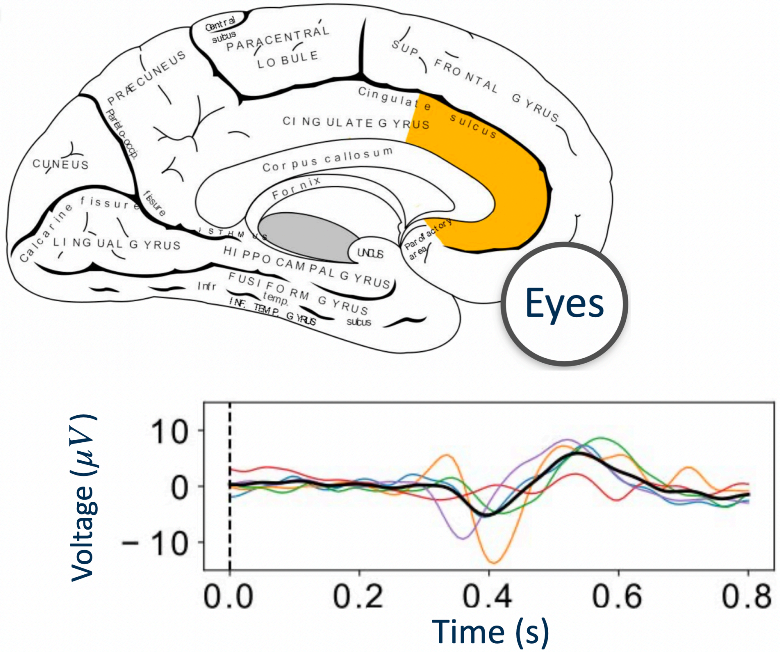
- Systems designed for non-invasive BCIs largely suffer from poor generalization
    - Between different subjects and environments, brain signals show considerable variance
    - This necessitates long retraining/calibration sessions
- The lack of generalization is typically attributed to the **covariate shift** of signals in the probability space, which manifests itself as disparate marginal and class conditional distributions across the source and target domains

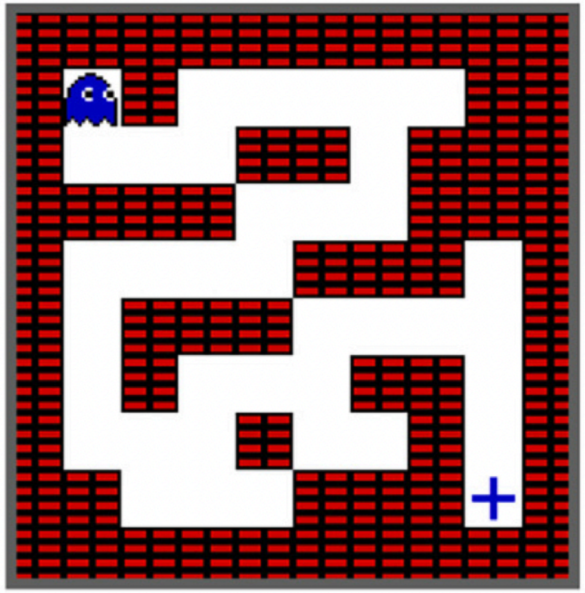


- In this paper, we propose adapting models to address covariate shift

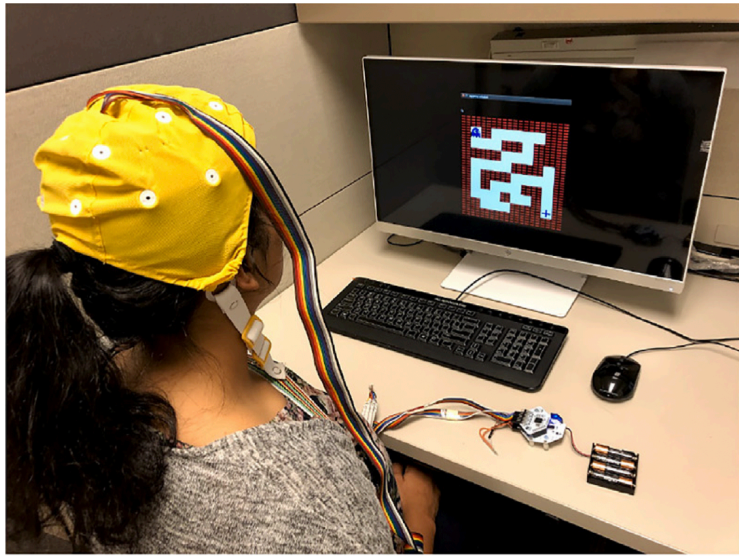
ErrP and Dataset

- Error Potential (ErrP) dataset collected in our lab from 10 subjects
- ErrP signal is elicited in the brain when a subject observes an erroneous activity
- Each subject observes an agent navigate a maze on their screen (the agent makes a wrong move with a probability of 0.2)



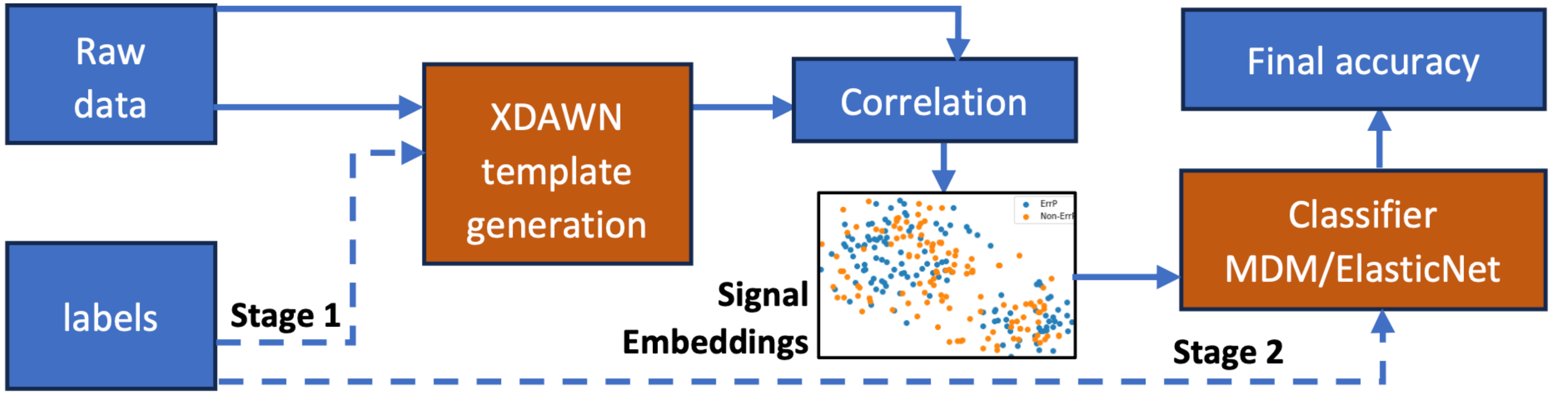


Maze

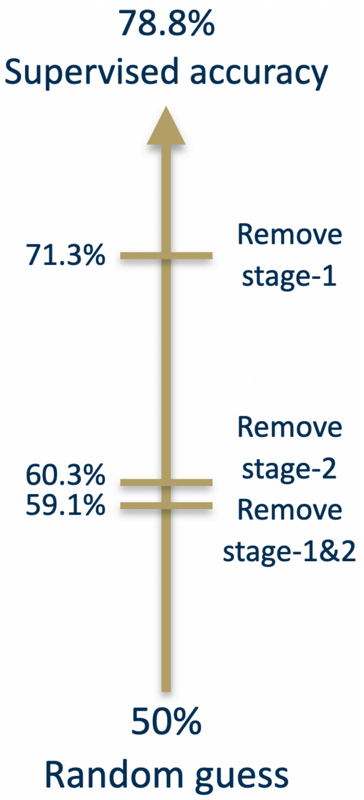


Supervised detection pipeline

- We start with the xDAWN + Riemannian Geometry (xRG) based supervised model that obtains state-of-the-art performance for ErrP generalization
- xRG contains two stages that use supervised learning
  - **Template generation stage (stage 1)**: generates signal embeddings from raw signal data using a template estimated from the ground-truth class labels
  - **Classification stage (stage 2)**: trains the classifier using the ground-truth class labels



Factors limiting detection generalization



MDM / ElasticNet	Label-assisted stage 1	Label-free stage 1
Label-assisted stage 2	76.0% / 78.8%	64.8% / 71.3%
Label-free stage 2	57.1% / 60.3%	55.8% / 59.1%
Silhouette score	0.0202	0.0116

- Original accuracy achieved by supervised learning:
  - 76.0% for MDM and 78.8% for ElasticNet
- Removing labels in stage 1:
  - Accuracy drops by 7.5% for ElasticNet due to **diminished class discrimination** in the label-free embeddings
- Removing labels in stage 2:
  - Accuracy drops by 18.5% for ElasticNet due to **covariate shift of signals**
- Removing labels in stage 1 & 2:
  - Accuracy drops by 19.7% for ElasticNet

Optimal transport: solution to covariate shift

- Optimal transport is the general problem of adapting one distribution to another as efficiently as possible
- Source distribution $\mathbf{a}$, target distribution $\mathbf{b}$, transport plan γ, and cost matrix $\mathbf{M}$
- Problem formulation:

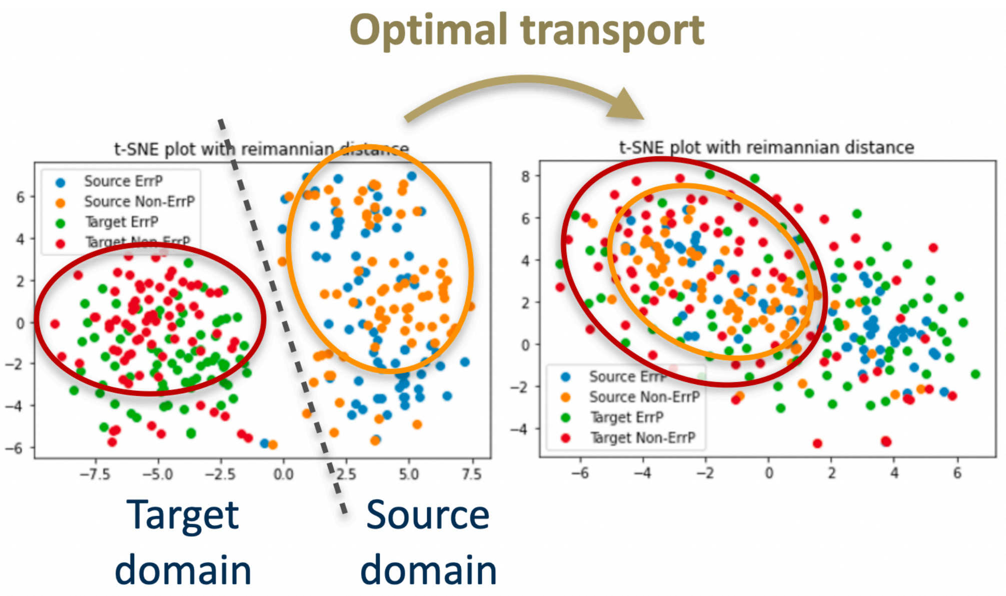
$$\gamma = \arg\min_{\gamma} \langle \gamma, \mathbf{M} \rangle_{\mathrm{F}} + \lambda \Omega_e(\gamma) + \eta \Omega_g(\gamma) \tag{2}$$

$$s.t.\ \gamma \mathbf{1} = \mathbf{a},\ \gamma^T \mathbf{1} = \mathbf{b},\ \gamma \geq 0 \tag{3}$$

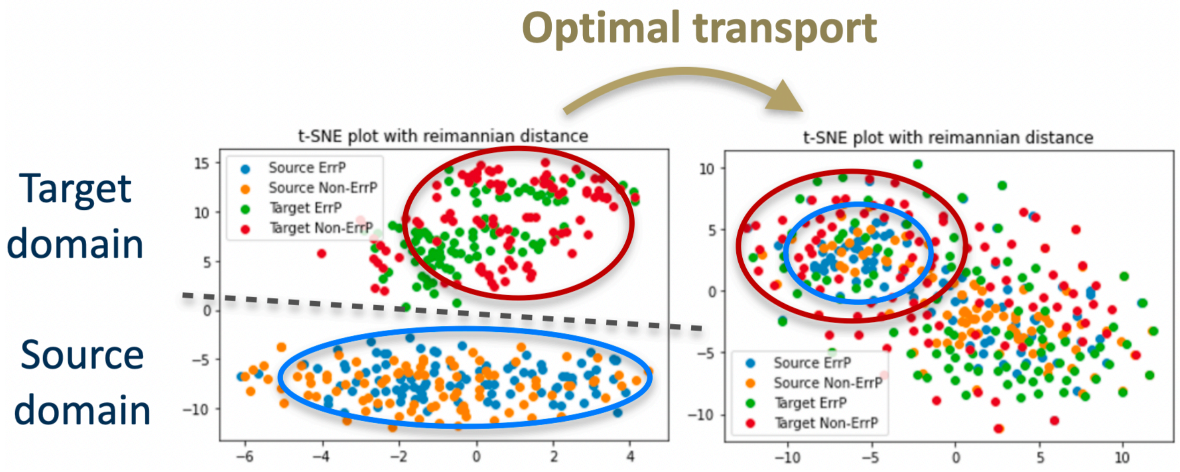
- Entropic regularization term Ω_e and group lasso regularization term Ω_g

Types of transport maps

- Positive transport
  - The source domain maintains its class discrimination after transport, and its ErrP points are adjacent to the target ErrP points and vice versa



- Negative transport
  - The source ErrP points are adjacent to the non-ErrP points in the target dataset and vice versa



Partial target-aware optimal transport (PTA-OT)

- To mitigate negative transport, we propose “partial target-aware optimal transport” by modifying the cost matrix M to establish the desired relationship between the source and target points
- Outline of PTA-OT
  - Calculate Riemannian mean of the centroids of the target data class by only a few labeled samples from the target dataset
  - Bias the transport map to avoid transporting source labels to an area that is close to the centroids of another class
  - Solve the biased optimal transport problem

Partial target-aware optimal transport (PTA-OT)

Algorithm 1 Partial target-aware optimal transport

Input: Source set $\{\mathbf{S}_i | i = 1, \cdots, n_s\}$ with its density $\mathbf{a} \in \mathbb{R}^{n_s}$ and target set $\{\mathbf{T}_i | i = 1, \cdots, n_t\}$ with its density $\mathbf{b} \in \mathbb{R}^{n_t}$. Few-shot class labeled target sets $\{\mathbf{L_0}^i | i = 1, \cdots, m_0 \| m_0 \ll n_t\}$ and $\{\mathbf{L_1}^i | i = 1, \cdots, m_1 | m_1 \ll n_t\}$

Initialization: $\mathbf{M}_{i,j} = \| \log(\mathbf{T}_j^{-1/2} \mathbf{S}_i \mathbf{T}_j^{-1/2}) \|_2^2$.

$\mathbf{C}_0 = mean(\mathbf{L}_0), \mathbf{C}_1 = mean(\mathbf{L}_1)$, the initial approximation of target class centroids.

for $i = 1, \cdots, n_s$ **do**

  for $j = 1, \cdots, n_t$ **do**

    $\mathbf{D}_{j0} = dist(\mathbf{T}_j, \mathbf{C}_0), \mathbf{D}_{j1} = dist(\mathbf{T}_j, \mathbf{C}_1)$

    if $class(\mathbf{S}_i) == 0$ **then**

      $\Delta = \mathbf{D}_{j0} / \mathbf{D}_{j1}$

    else

      $\Delta = \mathbf{D}_{j1} / \mathbf{D}_{j0}$

    end if

    $\mathbf{M}'_{i,j} = \mathbf{M}_{i,j} * \Delta$

  end for

end for

$\gamma = \arg\min_\gamma \langle \gamma, \mathbf{M}' \rangle_F + \lambda \Omega_e(\gamma) + \eta \Omega_g(\gamma)\ s.t.\ (3).$

Return: γ.

Initializing cost matrix $\mathbf{M}$ by Riemannian distance between source and target sample

Approximating target class centroids $\mathbf{C}_0$ and $\mathbf{C}_1$ by few-shot labels

Decreasing cost toward centroid of the same class
Increasing cost toward centroid of the different class

Solving OT with respect to target-aware cost matrix $\mathbf{M}'$

Performance evaluation

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	Mean
xRG MDM	59.2%	53.8%	54.8%	60.4%	54.2%	56.7%	54.3%	55.2%	53.5%	55.5%	55.8%
PTA-OT MDM	61.2%	63.0%	61.4%	65.8%	59.9%	60.6%	59.0%	64.4%	60.9%	63.8%	**62.0%**
xRG ElasticNet	62.6%	56.2%	61.4%	61.4%	58.6%	59.4%	57.9%	53.8%	58.0%	62.0%	59.1%
PTA-OT ElasticNet	66.6%	67.6%	63.2%	63.9%	68.5%	64.1%	60.1%	74.0%	62.2%	71.54%	**66.2%**

Table 2: Subject-wise cross-user transfer learning accuracy for label-free xRG vs our algorithm.

- Highlight 1:
  - The improvement is universal for all the subjects

Comparison with label-assisted/free stages

MDM / ElasticNet	Label-assisted stage 1	Label-free stage 1
Label-assisted stage 2	76.0% / 78.8%	64.8% / 71.3%
Label-free stage 2	57.1% / 60.3%	55.8% / 59.1%

Closely recover

62.0% / 66.2%

Substantially outperform

- Highlight 2:
  - The average accuracy is improved by 11.1% / 12.0% for MDM / ElasticNet
- Highlight 3:
  - Given the label-free embeddings (stage 1), we are able to reach within 95.6% and 92.8% of the accuracy for MDM and ElasticNet
- Highlight 4:
  - Given the label-free classification (stage 2), we are able to outperform by 8.6% and 9.8% of the accuracy for MDM and ElasticNet
- Highlight 5:
  - We use only a small fraction (5%) of the target labels, thereby accelerating model generalization by an order of magnitude

Conclusion and future work

- Our algorithm is a general-purpose algorithm that works with data distributions which suffer from covariate shift and minimizes the disparity between marginal source and target distributions while also preserving the class conditional probabilities
- Our preliminary results show significant potential in using PTA-OT
  - Using 5% of labels to achieved 95% of supervised performance
- Future work
  - incorporate both temporal and spatial information into optimal transport
  - increase the granularity of domain adaptation
  - reduce the required number of labels
  - increase both within-subject and cross-subject accuracy

Contact Information

- Ekansh Gupta
  - egupta8@gatech.edu
  - Currently on the job market for both internship and full positions
- Cheng-Yeh Chen
  - cchen847@gatech.edu
- Raghupathy Sivakumar (Siva)
  - siva@gatech.edu

Thank you!

PAPER

Georgia Tech