

Multi-task Learning for Commercial Brain Compute Interfaces

George Panagopoulos 22-October-2017 Computational Physiology Lab University of Houston



- > Brain Computer Interfaces
- Subject-subject generalization
- Multi-task learning
- Experiment
- Future work



- Medical imaging devices, typically EEG-based
- Monitor the activity of certain areas of the brain
- Map certain recording patterns to specific activities (Active)
- Classify some aspect of the human's cognitive state (Passive)



www.bio-signal.com



Commercial BCI

- BCI technology getting out of the lab
- Wireless headsets with EEG sensors and cool design

Advantages

- Economic
- Easy to use
- Application friendly
- Broader audience



Disadvantages

Limited data Noisy signals



Train one specific model for each subject

Problems:

- Need subject-specific training
- The algorithms adapt to noise
- Prior knowledge not utilized
- Knowledge extraction through different models, to derive general conclusions



Train one model with data from all subjects (pooled)

Problem:Neural signals exhibit variability across brains





Learning on training tasks is performed simultaneously to capture intrinsic relatedness and share knowledge.

Advantages:

- Theoretically proved to increase accuracy in new tasks
- Extracts common patterns from all tasks
- Already used in clinical neuroimaging studies to overcome inter-subject variability issues



Regression with group sparsity constraint on the coefficients K of all tasks T

$$min\frac{1}{N_t}\sum_{t=1}^T J(X_t, W_t, Y_t) + \lambda \sum_{k=1}^K |W_k|_2$$
$$X_T \in R^{N_s \times K}, Y_s \in R^{N_s} \quad W = \begin{bmatrix} w_1^1 & \cdots & w_1^K \\ \vdots & \ddots & \vdots \\ w_T^1 & \cdots & w_T^K \end{bmatrix}$$

Obozinksi, Taskar and Jordan, 2006



Difference in Coefficients

Each approach yields a different W:





Bayesian estimation of the coefficients prior distribution, which is shared for all subjects

$$p(W; X, Y) \sim \prod_{t=1}^{T} p(y_t, X_t; w_t) p(w_t) \quad , \qquad p(w_t) \sim N(\mu, \Sigma) , \forall t \in T$$

$$\min_{w_t,\mu,\Sigma} \frac{1}{\sigma^2} \sum_{t=1}^{T} \left| |X_t w_t - Y_t| \right|^2 + \frac{1}{2} \sum_{t=1}^{T} (w_t - \mu)^T \Sigma^{-1} (w_t - \mu) + \frac{T}{2} \log det(\Sigma)$$

A coefficient vector w_n for a new subject is sampled from $N(\mu, \Sigma)$.

Alamgir, Grosse-Wentrup and Altun, 2010



Datasets

Carnegie Mellon Experiment

- 9 subjects
- Ten 2-minute sessions with MOOC videos
- Self-classified levels of confusion for each session
- Classes: Confused or not

Berkeley Experiment

- 30 subjects
- One 5-minute session
- Two types of stimuli during session
 - Math, memorizing colors, think of items
 - Listen to music, watch video ads, relax
- Classes: Mental activity or relaxation



Accuracy

10-fold subject cross validation



Also used: MLP, RBF NN, LVQ, KNN, XGBOOST and Ensemble of NN



Consistent pattern extraction agreeing with the field's literature *Pooled Multi-task*





Multi-task algorithms are more robust than conventional pooled approaches to the subject generalization problem

 Steady or increase accuracy as the number of subjects increases

Consistent feature selection

This may apply to all studies that include humans



Address session-session generalization:

- Compare fully adaptive and multi-task learning approaches to same subject recordings
- Use the models derived for each subject in MTL to new recordings of that subject

Sequential multi-task learning

 Exploit the sequential nature of the data to achieve more accuracy results



Pipeline (R and MATLAB) with run instructions in: https://github.com/GiorgosPanagopoulos/Multi-task-Learning-for-Commercial-Brain-Computer-Interfaces



Thank you

gpanagopoulos@uh.edu

https://giorgospanagopoulos.github.io