ACM Conference on Knowledge Discovery and Data Mining

Data-Efficient Brain Connectome Analysis via Multi-Task Meta-Learning

Presented by **Yanqiao Zhu**

yzhu@cs.ucla.edu

@ https://SXKDZ.github.io

The Scalable Analytics Institute (ScAi)
Department of Computer Science
University of California, Los Angeles (UCLA)

UCLA Samueli Computer Science

Joint work with Owen Yang, Hejie Cui, Xuan Kan, Lifang He, Ying Guo, Carl Yang

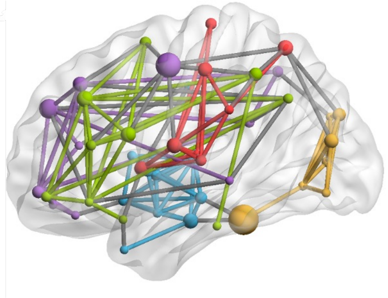
Outline

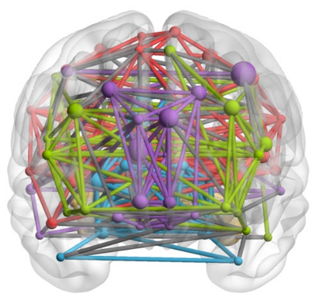
1. Background and Motivation
2. Preliminaries and Related Work
3. Methods and Analysis
4. Conclusions and Discussions

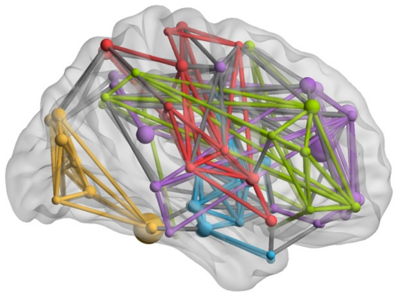
Outline

Background & Motivation

- **The Scarcity of Neuroimaging Data**
  - Brain network data is usually collected through various techniques such as functional magnetic resonance imaging (fMRI) or diffusion tensor imaging (DTI).
  - The data collection process is costly and expensive which leads to severe scarcity of available training resources and data instances.
  - Modern machine learning techniques on complex topological data requires sufficiently large samples to achieve effective domain knowledge extraction and discriminative power.







Outline

1. Background and Motivation
2. **Preliminaries and Related Work**
3. Methods and Analysis
4. Conclusions and Discussions

Related Work

- **Graph Neural Networks**
  - GNN is powerful in learning topological relational information among nodes and edges.
  - BrainGNN (Li et al. 21’) proposed ROI aware graph convolution layer and selective pooling layer.
  - With limited training samples, GNN suffers poor performance result and high variances.
- **Meta-Learning on Graphs**
  - Previous work surveyed on applicability and feasibility of disentangling mapped representation to perform multiple objectives and jointly train the parameters.
  - Alternatively, other approaches feature shared substructure learning which raises compatibility concerns to dense brain connectomes.

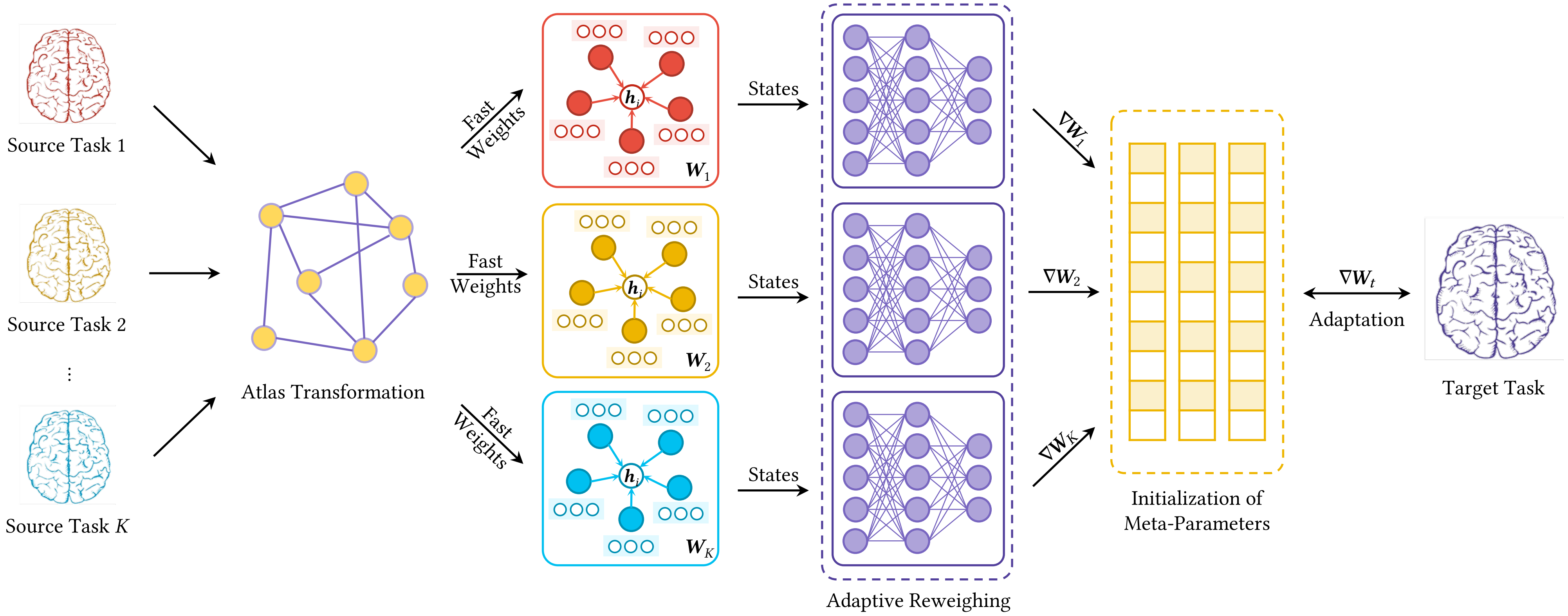
Problem Formulation

- We consider the brain network data with carefully parcellated ROIs and correlations as edge weighted graph $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i, A_i)$ where $\mathcal{V}$ is the node set and $\mathcal{E} = \mathcal{V} \times \mathcal{V}$ is the edge set.
- Our objective: train an encoder model $f_\theta(\cdot)$ such that θ efficiently converges to optimal θ^* on target dataset $\mathcal{D}^t$ given that θ_0 is initialized via proper pre-training or meta-training on source datasets $\mathcal{D}^s = \{S_1, S_2, \cdots, S_k\}$, where $|\mathcal{D}^s| > |\mathcal{D}^t|$.
- Considering the multiview and multimodality nature of brain network dataset, we regard each view as an independent training objective. That is, given a dataset $\mathcal{D}$ with k modalities, our learning pipeline will extract k tasks into the task distribution τ.

Outline

1. Background and Motivation
2. Preliminaries and Related Work
3. **Methods and Analysis**
4. Conclusions and Discussions

Overall Pipeline



Data Efficient Training (1)

- **Single Task Transfer Learning**
  - The first stage involves pre-training the encoder $f_\theta(\cdot)$ on a single source task and its corresponding objective
  - We use the binary cross entropy loss as our objective function for graph classification where y_i refers to class label and $\sigma(\cdot)$ is the sigmoid non-linear activation
  - We then fine-tune the encoder on target task and evaluate its performance under k-fold cross validation
  - Sensitive and vulnerable to knowledge gaps between source and target domain

Data Efficient Training (2)

- **Transitioning to multi-task transfer learning (MTT)**
  - We extend into multi-task setting where $f_\theta(\cdot)$ is pre-trained according to an aggregated (e.g., sum) loss objective
  - MTT has limited generalizability power due to the joint training objective which leads to suboptimal convergence and ineffective capturing of shared domain knowledge
  - The state-of-the-art meta-learning (Finn et al. 17') architecture demonstrates robustness in generalizing knowledge across domains by leveraging task-specific adaptation, feature reuse, and second-order gradient update

Datasets

- Bipolar Disorder (BP)
  - Consists of 52 bipolar I individuals and 45 healthy controls with fMRI and DTI modalities
- Human Immunodeficiency Virus (HIV)
  - Consists of 35 early HIV patients and 35 seronegative controls constructed with fMRI and DTI modalities
- Parkinson's Progression Markers Initiative (PPMI)
  - Consists of 569 Parkinson's infected patients and 149 healthy controls with Probabilistic Index of Connectivity (PICo), Hough voting (Hough), and FSL modalities. For experimental purposes, we balance the training and testing set by randomly sampling 149 infected instances

Experimental Configurations

- Backbone encoders:
  - BrainNetCNN (Kawahara et al. 17')
  - GCN (Kipf et al. 17')
  - GAT (Veličković et al. 18')
- Training configurations:
  - We consider PPMI to be sources dataset due to larger sample abundance whereas BP and HIV are regarded as target dataset for fine-tuning and evaluation.

Results

Table 1: Performance comparison of our proposed methodologies and baselines in terms of area under the ROC curve (AUC) and accuracy (ACC). The best performing model is highlighted in boldface.

Encoder	Dataset	Modality	DSL		STT		MTT		MML	
			AUC	ACC	AUC	ACC	AUC	ACC	AUC	ACC
BrainNetCNN	BP	fMRI	0.50±0.13	0.51±0.15	0.55±0.07	0.56±0.08	0.56±0.09	0.56±0.11	0.57±0.10	0.57±0.07
		DTI	0.47±0.16	0.49±0.14	0.53±0.11	0.54±0.12	0.54±0.07	0.54±0.09	0.55±0.13	0.56±0.08
	HIV	fMRI	0.60±0.15	0.59±0.13	0.66±0.14	0.65±0.10	0.66±0.13	0.66±0.11	0.67±0.12	0.67±0.09
		DTI	0.54±0.16	0.53±0.15	0.60±0.09	0.60±0.09	0.60±0.10	0.60±0.12	0.57±0.11	0.61±0.14
GAT	BP	fMRI	0.51±0.13	0.52±0.16	0.57±0.07	0.58±0.05	0.59±0.10	0.59±0.07	0.61±0.07	0.60±0.09
		DTI	0.50±0.09	0.50±0.13	0.53±0.08	0.54±0.10	0.51±0.06	0.55±0.08	0.55±0.08	0.57±0.05
	HIV	fMRI	0.61±0.15	0.61±0.14	0.65±0.07	0.66±0.11	0.66±0.09	0.68±0.06	0.68±0.10	0.69±0.08
		DTI	0.56±0.17	0.55±0.15	0.61±0.07	0.60±0.08	0.62±0.09	0.61±0.10	0.64±0.09	0.62±0.12
GCN	BP	fMRI	0.55±0.11	0.54±0.14	0.59±0.12	0.58±0.13	0.61±0.10	0.60±0.11	**0.62±0.08**	**0.62±0.10**
		DTI	0.51±0.12	0.52±0.11	0.52±0.10	0.54±0.12	0.55±0.09	0.56±0.14	**0.59±0.07**	**0.58±0.11**
	HIV	fMRI	0.63±0.18	0.64±0.12	0.65±0.14	0.68±0.15	0.67±0.12	0.68±0.11	**0.69±0.10**	**0.70±0.09**
		DTI	0.60±0.12	0.58±0.13	0.61±0.11	0.60±0.12	0.63±0.13	0.63±0.15	**0.65±0.12**	**0.64±0.13**

Atlas Transformation

- Another challenge of cross-dataset brain analysis: incompatible and non-convertible ROI definitions across datasets, which may lead to negative knowledge transfer.
- Two solutions:
  - Simple Auto-encoding (AE): obtains semantic preserving and fixed representation of original feature space. The encoding process is parameterized by a projection matrix $\boldsymbol{W} \in \mathbb{R}^{n \times k}$ and the objective is $\underset{\boldsymbol{W}}{\operatorname{argmin}} \parallel \boldsymbol{X} - \boldsymbol{X}\boldsymbol{W}\boldsymbol{W}^{\boldsymbol{T}} \parallel^{\boldsymbol{2}}$.
  - Learnable Linear Projections (LP): performs dimension reduction or projection of the original input feature space by attaching a parameterized projection head $\boldsymbol{W} \in \mathbb{R}^{n \times k}$ in front encoder f_θ. The optimization can be potentially hindered by the inability of the encoder to learn from fixed input signals.

Atlas Transformation (cont.)

- Another challenge of cross-dataset brain analysis: incompatible and non-convertible ROI definitions across datasets, which may lead to negative knowledge transfer.

Table 2: Performance with three different atlas transformation techniques.

Dataset	Modality	Zero Pad		LP		AE	
		AUC	ACC	AUC	ACC	AUC	ACC
BP	fMRI	$0.62_{\pm 0.08}$	$0.62_{\pm 0.10}$	$0.62_{\pm 0.13}$	0.63 ± 0.12	$\mathbf{0.63}_{\pm 0.09}$	$\mathbf{0.64}_{\pm 0.09}$
	DTI	$0.59_{\pm 0.07}$	$0.58_{\pm 0.11}$	$0.59_{\pm 0.09}$	$0.60_{\pm 0.14}$	$\mathbf{0.60}_{\pm 0.04}$	$\mathbf{0.61}_{\pm 0.10}$
HIV	fMRI	$0.69_{\pm 0.10}$	$0.70_{\pm 0.09}$	$0.71_{\pm 0.13}$	$0.70_{\pm 0.11}$	$\mathbf{0.73}_{\pm 0.10}$	$\mathbf{0.72}_{\pm 0.08}$
	DTI	$0.65_{\pm 0.12}$	$0.64_{\pm 0.13}$	$0.68_{\pm 0.14}$	$0.66_{\pm 0.13}$	$\mathbf{0.69}_{\pm 0.06}$	$\mathbf{0.69}_{\pm 0.08}$

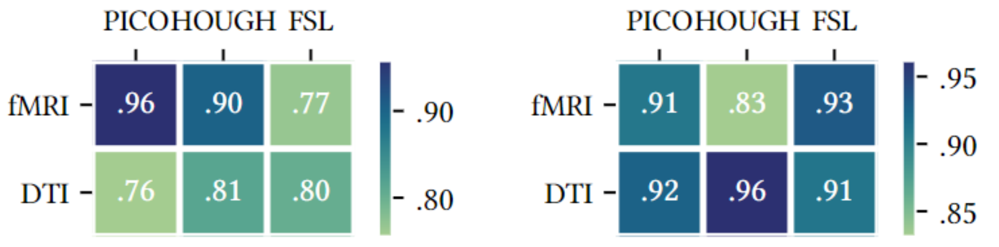
Task Adaptive Reweighing

- Motivations:
  - Base meta-learning framework fails to consider learning difficulty of different individual task. That is, during meta-training, each task would reflect varying rate of optimization convergence, which leads to skewed and biased overall generalization.
- Analysis:
  - We first verify the assumption of underlying clinical and semantical relatedness of source (PPMI) and target (BP, HIV) data by visualizing a correlation matrix derived from computed task embeddings (Achille et al. 19').

Task Adaptive Reweighing (cont.)

- Observation:
  - High and varying level of correlation can be observed which corroborates with clinical studies on inter-connections among the investigated diseases. This suggests that the meta-training on each source task does not contribute equally to the adaptation on target task.
- Solution:
  - We leverage a dynamic and optimizable scheme for inner-loop hyperparameter (i.e. learning rate, weight decay) selection inspired by (Baik et al. 20’). The task-specific optimization is governed by a non-linear mapping function that the determines rate of convergence. In other words, the task-wise adaptation in meta-training is now weighted according to learning “difficulty” and generalizability contribution.

Task Adaptive Reweighing (cont.)



(a) PPMI and BP Task Correlation (b) PPMI and HIV Task Correlation

Figure 2: Task correlations with different source and target datasets. We compute the Fisher information estimation derived from the Hessian matrix by training each task using the same architecture as in Section 4.2. The task embedding is then composed of layer-wise concatenation of the flattened Fisher information matrix.

Table 3: Performance with task reweighing techniques.

Dataset	Modality	DSL		MML		MMAR	
		AUC	ACC	AUC	ACC	AUC	ACC
BP	fMRI	$0.55_{\pm0.11}$	$0.54_{\pm0.14}$	$0.62_{\pm0.08}$	$0.62_{\pm0.10}$	$\mathbf{0.68_{\pm0.10}}$	$\mathbf{0.66_{\pm0.08}}$
	DTI	$0.51_{\pm0.12}$	$0.52_{\pm0.11}$	$0.59_{\pm0.07}$	$0.58_{\pm0.11}$	$\mathbf{0.64_{\pm0.06}}$	$\mathbf{0.64_{\pm0.10}}$
HIV	fMRI	$0.63_{\pm0.18}$	$0.64_{\pm0.12}$	$0.69_{\pm0.10}$	$0.70_{\pm0.09}$	$\mathbf{0.74_{\pm0.10}}$	$\mathbf{0.76_{\pm0.08}}$
	DTI	$0.60_{\pm0.12}$	$0.58_{\pm0.13}$	$0.65_{\pm0.12}$	$0.64_{\pm0.13}$	$\mathbf{0.72_{\pm0.08}}$	$\mathbf{0.72_{\pm0.07}}$

Outline

1. Background and Motivation
2. Preliminaries and Related Work
3. The Proposed Method
4. Experiments
5. **Conclusions**

Concluding Remarks

- We proposed a data-efficient learning framework on small-sized brain network dataset using meta-learning incorporated with brain-network-oriented design consideration.
- The framework is naturally generic and can be applied to a wide spectrum of backbone encoders, task distributions, objective functions, and datasets.

Discussions

- Current limitations:
  - Brain networks are generally multimodal and a comprehensive feature extraction requires a parametric model to capture informative shared knowledge across modalities.
  - Sampling source data for model meta-training could be costly which motivates further investigation on achieving comparable optimization with less data instances.
- Future directions:
  - We extend our investigation by applying experiments under unsupervised setting for model meta-training and explore on sampling-efficient strategies for brain network learning


THANKS