Localization of Heterogeneous Disease Features in Neuroimage: A Differential Inclusion Approach

Yuan Yao

HKUST
https://yao-lab.github.io

March 23, 2019

Acknowledgements

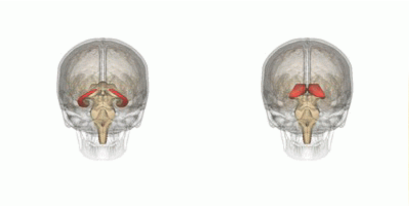
- Applications:
  - *Xinwei Sun, Yizhou Wang*, *Chendi Huang* (PKU)
  - *Lingjing Hu* (Capital University of Medical Sciences)
  - *Qianqian Xu* (CAS)
- Theory
  - *Chendi Huang, Xinwei Sun*, *Jiechao Xiong* (PKU & Tecent AI Lab)
- Discussions:
  - *Stan Osher*, *Wotao Yin* (UCLA), *Feng Ruan* (Stanford & PKU)
- Grants:
  - National Basic Research Program of China (973 Program), NSFC

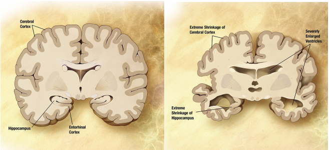
What is Alzheimer's Disease (AD)

- A complex chronically progressive neurodegenerative disease and the most common form of dementia in elderly people worldwide.
- In 2015, about 29.8 million AD people worldwide and dementia resulted in 1.9 million deaths.
- In Hong Kong alone, dementia was estimated to affect approximately 103,433 people over the age of 60 years in 2009 and is expected to increase 222% by 2039.
- The patient with AD is transitioned from Normal Control (NC), to Mild Cognitive Impairment (MCI) then to AD.
- There are no clear rules to define status of the diseases.
- However, the brain structure changes a lot during the process before functional changes detected.

Biomarkers from sMRI

- Biomarkers processed by structural Magnetic Resonance Imaging (sMRI) (PET etc.) as predictors used to classify NC, MCI, and AD, e.g. atrophy of Hippocampus and Thalamus.
- Symptoms can be alleviated or inhibited by drugs, it's hard to say whether it has structurally improvement.





Heterogeneity of Voxel-based Features

The preprocessed features on structural Magnetic Resonance Imaging (sMRI) images contain the following voxel-wise features:

- Lesion features that are contributed to the disease
- Procedural bias introduced during the preprocessing steps and shown to be helpful in classification
- Irrelevant or null features that are not due to disease status

Heterogeneity of Voxel-based Features

The preprocessed features on structural Magnetic Resonance Imaging (sMRI) images contain the following voxel-wise features:

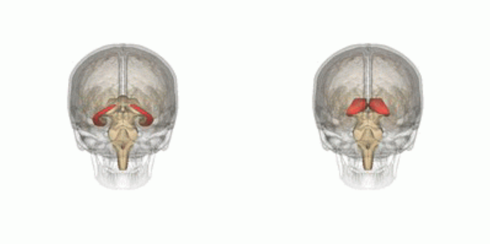
- Lesion features that are contributed to the disease
- Procedural bias introduced during the preprocessing steps and shown to be helpful in classification
- Irrelevant or null features that are not due to disease status

Our two goals in voxel-based neuroimage analysis for disease prediction:

- Accurate Classification: NC, MCI, AD
- Stable feature selection of lesion features and procedural bias, with high recall and low false discovery rate (FDR).

Lesion Features

- Have been the main focus in disease prediction.
- Only a few number of gray matter voxels are correlated with the disease.
- Geometrically clustered in atrophied regions in dementia disease such as Alzheimer's Disease (AD).



Procedural Bias

- Introduced during the preprocessing steps, are found to be helpful for disease prediction.
- Refer to the mistakenly enlarged Gray Matter (GM) voxels surrounding locations with cerebral spinal fluid (CSF) spaces enlarged, e.g. lateral ventricle in AD. [Sun et al. MICCAI 2017]

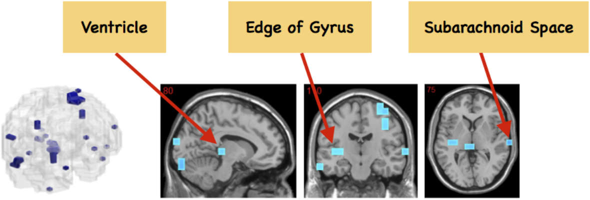


Fig. 1. The overlapped voxels among top 150 negative value voxels in each fold of β_{pre} at the time corresponding to the best average prediction result in the path of GSplit LBI using 10-fold cross-validation. For subjects with AD, they represent enlarged GM voxels surrounding lateral ventricle, subarachnoid space, edge of gyrus, etc.

Limitation of Existing Models

- The lesion features have been the only focus of existing models for their medical interpretability.
- In VBM analysis, procedure biases are introduced on some features during the commonly used preprocessing procedure of T_1 weighted image (e.g. DARTEL [Ashburner 2007, Neuroimage])
- Procedure bias can be helpful for classification, however they are ignored in the literature.

ADNI Dataset

- Consider AD/NC classification (namely ADNC) and MCI (Mild Cognitive Impairment)/NC (namely MCINC) classification
- The data are obtained from ADNI[1] database, which is split into 1.5T and 3.0T (namely 15 and 30) MRI scan magnetic field strength datasets.
- DARTEL VBM pipeline [Ashburner (2007) Neuroimage] is implemented to preprocess the data.
- The input features consist of 2,527 $8\times8\times8$ mm^3 size voxels with average values in GM population template greater than 0.1.
- Experiments are designed on 15ADNC, 30ADNC and 15MCINC tasks.

[1] http://adni.loni.ucla.edu

The efficacy of Procedural bias in prediction

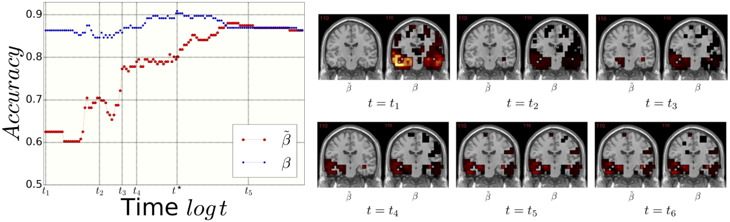
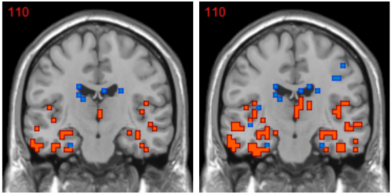


Figure: Exploitation of both lesion features and procedure bias (blue) improves prediction accuracy by dominating the curve merely by lesion features (red). $\tilde{\beta}$ corresponds to the lesion features interpretable for AD, while β additionally leverages the procedure bias to improve the prediction.

Lesion Features vs. Procedure Bias



Local FDR T-test($p < 0.001$)

Figure: Leision features (red) as degenerate GM voxels, while Procedure Bias (blue) contains ‘enlarging’ GM voxels in preprocessing due to the enlarging cerebral spinal fluid (CSF), surrounding lateral ventricle and subarachnoid space etc.

Results of Prediction

Exploiting both the procedure bias and lesion features, our method (GSplitLBI [Sun-Hu-Y.-Wang 2017]) achieves the state-of-the-art performance in the 10-fold cross-validation error:

Table 1. Comparison of GSplit LBI with other models

	MLDA	SVM	Lasso	Graphnet	Elastic Net	TV + l_1	n^2GFL	GSplit LBI (β_{pre})
15ADNC	85.06%	83.12%	87.01%	86.36%	88.31%	83.77%	86.36%	**88.96%**
30ADNC	86.93%	87.50%	87.50%	88.64%	89.20%	87.50%	87.50%	**90.91%**
15MCINC	61.41%	70.13%	69.80%	72.15%	70.13%	73.83%	69.80%	**75.17%**

Stability of Feature Selection

Our method (GSplitLBI [Sun-Hu-Y.-Wang 2017]) also champions in stability of heterogeneous feature selection:

Table 2. mDC comparison between GSplit LBI and other models

	Lasso	Elastic Net	Graphnet	TV + l_1	n^2 GFL	GSplit LBI (β_{les})
Accuracy	87.50%	89.20%	88.64%	87.50%	87.50%	88.64%
mDC	0.1992	0.5631	0.6005	0.5824	0.5362	**0.7805**
$\sum_{k=1}^{10} \lvert S(k)\rvert/10$	50.2	777.8	832.6	712.6	443.9	129.4

Where

$$mDC := \frac{10|\cap_{k=1}^{10} S(k)|}{\sum_{k=1}^{10} |S(k)|}$$

with $S(k)$ denoting the support set of β_{les} in k-th fold.

How do we reach this?

Below we are going to introduce a new methodology:

- Boosting with Structural Sparsity
- mathematically, a differential inclusion method as restricted gradient flows, whose discretization meets the sparse mirror descent algorithm
- variable splitting enables us to exploit both lesion features and procedure bias effectively
- it can be generalized to FDR heterogeneous smoothing for new algorithms in Empirical Bayes

Some Reference

- Osher, Ruan, Xiong, Yao, and Yin, "Sparse Recovery via Differential Equations", *Applied and Computational Harmonic Analysis*, 2016
- Xiong, Ruan, and Yao, "A Tutorial on Libra: R package for Linearized Bregman Algorithms in High Dimensional Statistics", *Handbook of Big Data Analytics*, Eds. by Wolfgang Karl Härdle, Henry Horng-Shing Lu, and Xiaotong Shen, Springer, 2017
- Xu, Xiong, Cao, and Yao, "False Discovery Rate Control and Statistical Quality Assessment of Annotators in Crowdsourced Ranking", ICML 2016, arXiv:1604.05910
- Huang, Sun, Xiong, and Yao, "Split LBI: an iterative regularization path with structural sparsity", NIPS 2016
- Sun, Hu, Yao, and Wang, "GSplit LBI: taming the procedure bias in neuroimaging for disease prediction", MICCAI 2017
- Huang, Yao, "A Unified Dynamic Approach to Sparse Model Selection", AISTATS 2018
- Sun, Hu, Zhang, Yao, and Wang, "FDR-HS: An Empirical Bayesian Identification of Heterogenous Features in Neuroimage Analysis", MICCAI 2018
- Huang, Sun, Xiong, and Yao, "Boosting with Structural Sparsity: A Differential Inclusion Approach", *Applied and Computational Harmonic Analysis*, 2018

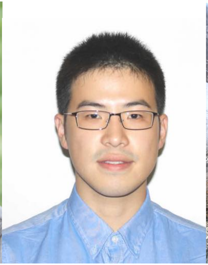
- R package:
  - `http://cran.r-project.org/web/packages/Libra/index.html`
- Matlab package:
  - `https://github.com/yuany-pku/split-lbi`
- Python package to appear

Summary: A Renaissance of Boosting?

- Boosting as gradient descent is arguably the **best off-the-shelf machine learning algorithm** (Leo Breiman): **AdaBoost** (Freund-Schapire), **L2Boost** (Buhlman-Yu), **LogitBoost** (Friedman)
- The (sparsity) restricted gradient descent dynamics: **differential inclusions**
  - has a simple discretized algorithm to follow the regularization path (LBI), amiable for parallel realizations in **big data analytics**
  - improves model selection or prediction, **better than generalized LASSO**
  - is widely adapted to **various sparsity** constraints
- This is a gift of Applied Mathematics to High Dimensional Statistics

Acknowledgements

Chao Gao (Chicago)

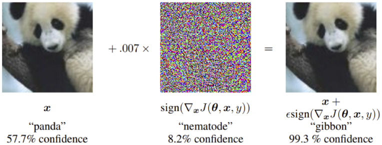


Jiyu Liu (Yale)

Weizhi Zhu (HKUST)

Robustness?

- How can one achieve robustness against adversarial?
- Robust statistics: for Huber's ϵ-contamination model, Generative Adversarial Networks (GANs) provably achieve statistically optimal robust estimates
- Reference
  - Gao, Liu, Yao, and Zhu, "Robust Estimation and Generative Adversarial Networks", ICLR 2019, `arXiv:1810.02030`.
  - Gao, Yao, and Zhu, "Generative Adversarial Nets for Robust Scatter Estimation: A Proper Scoring Rule Perspective.", `arXiv:1903.01944`.



[Goodfellow et al., 2014]

Decentralized Training of Neural Networks

- A Proximal Block Coordinate Descent Algorithm for Deep Neural Network Training. Tim Tsz-Kit Lau, Jinshan Zeng, Baoyuan Wu, Yuan Yao. The 6th International Conference on Learning Representations (ICLR 2018), Workshop Track
- Global Convergence of Block Coordinate Descent in Deep Learning Jinshan Zeng, Tim Tsz-Kit Lau, Shaobo Lin, Yuan Yao. ICML.
- A Convergence Analysis of Nonlinearly Constrained ADMM in Deep Learning. Jinshan Zeng, Shao-Bo Lin, Yuan Yao.

The END


рахмат
danke
謝謝
ngiyabonga
teşekkür ederim
Баярлалаа
спасибо
faafetai lava
mersi
barka
welalin
tack
kiitos
dankie
vinaka
спасибі
blagodaram
dank je
misaotra
matondo
paldies
grazzi
mahalo
tapadh leat
thank you
gracias
хвала
asante
manana
obrigada
tenki
nanni
nandri
dhanyavad
bedankt
hvala
mauruuru
köszönöm
dankon
akun
djiere dieuf
mochchakkeram
дякую
mamnun
murakoze
enkosi
bayarlalaa
gracie
dziękuję
chnorakaloutioun
gratias ago
gràcies
sulpáy
taiku
go raibh maith agat
chokrane
sobodi
dėkui
obrigado
mèsi
sagolun
sukriya
kop khun krap
arigatô
takk
dakujem
trugarez
ありがとう
tanemirt
grazie
najis tuke
didi madloba
kam sah hamnida
rahmat
terima kasih
rahmet
diolch
dhanyavadagalu
shukriya
mercê
мерси
তোমাকে ধন্যবাদ
감사합니다
xièxie
ευχαριστώ
merci