Division of Statistics and Data Science, Department of Mathematical Sciences
University of Cincinnati

University of CINCINNATI

Multinomial Tensor Regression with Application to Whole-brain Structural MRI Analysis

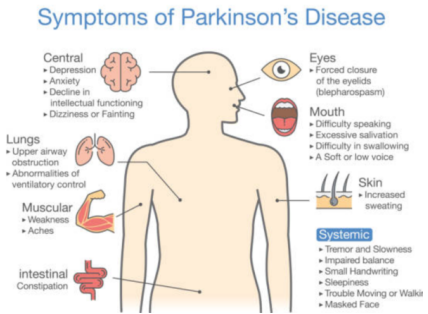
Xuan Cao, Fang Yang, Qingling Huang

June 03, 2021

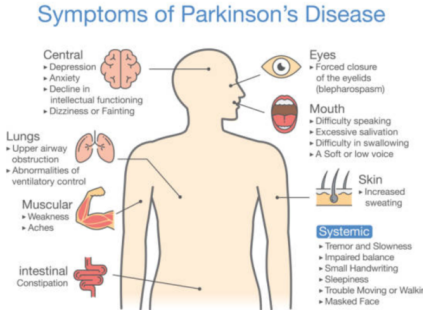
Overview

- Introduction
- Multinomial tensor regression
- Application to MRI data analysis
- Conclusions
- References

- Parkinson's disease (PD) is a major neurodegenerative disease influenced by both genetic and environmental factors. PD is characterized by the degeneration of dopamine-producing cells in the brain resulting in motor symptoms and nonmotor features.



- Parkinson's disease (PD) is a major neurodegenerative disease influenced by both genetic and environmental factors. PD is characterized by the degeneration of dopamine-producing cells in the brain resulting in motor symptoms and nonmotor features.

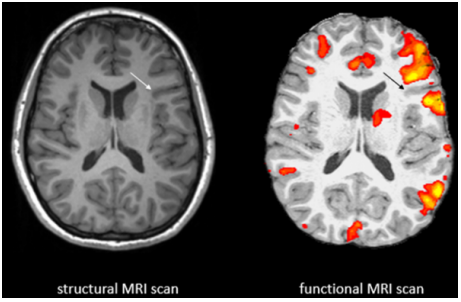


- Depression is a prominent nonmotor feature which is highly prevalent early in the disease process and has a significant impact on quality of life and disability.

Introduction

fMRI and sMRI

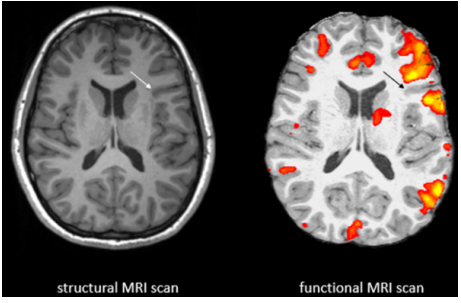
- A variety of neuroimaging technologies including functional magnetic resonance imaging (fMRI), structural MRI (sMRI) positron emission tomography and electroencephalography have been adopted, and provided new insights into PD.



Introduction

fMRI and sMRI

- A variety of neuroimaging technologies including functional magnetic resonance imaging (fMRI), structural MRI (sMRI) positron emission tomography and electroencephalography have been adopted, and provided new insights into PD.



- Structural MRI has recently received more research focus [*Jukka J. Remes in 2011 and Yael Jacob in 2019*] with better stability and repeatability compared to fMRI, where there exist concerns about accuracy due to noise.

Introduction

Whether the patient has Parkinson's disease (PD) is wondered to be derive from the sMRI. It is actually a classification problem. And there are several statistical methods to achieve it.

- Logistic regression / Multivariate logistic regression

Introduction

Whether the patient has Parkinson's disease (PD) is wondered to be derive from the sMRI. It is actually a classification problem. And there are several statistical methods to achieve it.

- Logistic regression / Multivariate logistic regression
- Classification tree

Introduction

Whether the patient has Parkinson's disease (PD) is wondered to be derive from the sMRI. It is actually a classification problem. And there are several statistical methods to achieve it.

- Logistic regression / Multivariate logistic regression
- Classification tree
- Support Vector Machine (SVM)

Introduction

Whether the patient has Parkinson's disease (PD) is wondered to be derive from the sMRI. It is actually a classification problem. And there are several statistical methods to achieve it.

- Logistic regression / Multivariate logistic regression
- Classification tree
- Support Vector Machine (SVM)
- Convolutional Neural Networks (CNN)

Introduction

Whether the patient has Parkinson's disease (PD) is wondered to be derive from the sMRI. It is actually a classification problem. And there are several statistical methods to achieve it.

- Logistic regression / Multivariate logistic regression
- Classification tree
- Support Vector Machine (SVM)
- Convolutional Neural Networks (CNN)

...

Introduction

- Those classical methods treat covariates as a vector and estimate a corresponding vector of regression coefficients.

Introduction

- Those classical methods treat covariates as a vector and estimate a corresponding vector of regression coefficients.

- Modern applications in medical imaging generate covariates of more complex form such as multidimensional arrays (tensors).

Introduction

- Those classical methods treat covariates as a vector and estimate a corresponding vector of regression coefficients.

- Modern applications in medical imaging generate covariates of more complex form such as multidimensional arrays (tensors).

- In particular, tensor regression model [*Hua Zhou, Lexin Li, and Hongtu Zhu in 2013*] is a regression framework that treats binary clinical outcome as response, and images as covariates in the form of multi-dimensional array. These tensor regression methods could not only resolve the computational and modeling challenges of large-scale imaging data, but also achieve perfect accuracy even in smaller sample sizes.

Introduction

- However, most of the existing methods focused on either the individual diagnosis of PD or the progression of depression comorbidity without simultaneously inferring the onset as well as the stage of PD.

Introduction

- However, most of the existing methods focused on either the individual diagnosis of PD or the progression of depression comorbidity without simultaneously inferring the onset as well as the stage of PD.

- In this study, our goal was to build and validate a multinomial tensor regression based framework that leveraged three-dimensional (3D) sMRI data to differentiate between nondepressed PD, depressed PD and healthy subjects.

Introduction

- This study was approved by the Medical Research Ethical Committee of Nanjing Brain Hospital (Nanjing,China) in accordance with the Declaration of Helsinki, and written informed consent was obtained from all subjects.

Introduction

- This study was approved by the Medical Research Ethical Committee of Nanjing Brain Hospital (Nanjing,China) in accordance with the Declaration of Helsinki, and written informed consent was obtained from all subjects.

- 84 depressed PD (DPD) patients, 196 non-depressed PD (NDPD) patients and 200 matched healthy controls (HC) were recruited. 3D sMRI images from NDPD, DPD patients and HCs were normalized using Statistical Parametric Mapping on the Matlab platform.

Introduction

- This study was approved by the Medical Research Ethical Committee of Nanjing Brain Hospital (Nanjing,China) in accordance with the Declaration of Helsinki, and written informed consent was obtained from all subjects.

- 84 depressed PD (DPD) patients, 196 non-depressed PD (NDPD) patients and 200 matched healthy controls (HC) were recruited. 3D sMRI images from NDPD, DPD patients and HCs were normalized using Statistical Parametric Mapping on the Matlab platform.

- As a result, the images of original size (512, 512, 128) were converted into images of size (79, 95, 79) such that the complexity of the following analysis was dramatically reduced without losing relevant information.

Multinomial tensor regression

Background

- The normalized 3D sMRI scans in our cases have $79 \times 95 \times 79 = 592,895$ voxels, i.e., 592,895 parameters to be estimated in a regression setup, if each voxel is treated as a covariate.

Multinomial tensor regression

Background

- The normalized 3D sMRI scans in our cases have $79 \times 95 \times 79 = 592,895$ voxels, i.e., 592,895 parameters to be estimated in a regression setup, if each voxel is treated as a covariate.

- *Hua Zhou, Lexin Li, and Hongtu Zhu [2013]* proposed the family of tensor regression models that incorporate the special structure of tensor covariates encoded in these images for binary classification.

Multinomial tensor regression

Background

- The normalized 3D sMRI scans in our cases have $79 \times 95 \times 79 = 592,895$ voxels, i.e., 592,895 parameters to be estimated in a regression setup, if each voxel is treated as a covariate.

- *Hua Zhou, Lexin Li, and Hongtu Zhu [2013]* proposed the family of tensor regression models that incorporate the special structure of tensor covariates encoded in these images for binary classification.

- Adapted to our case, we propose the multinomial tensor regression model to discriminate between NDPD, DPD and HC.

Multinomial tensor regression

Model

Let n denote the generic sample size representing the number of 3D images exploited for fitting the model. Let Y_{ij} be the binary variable that indicates whether the ith subject has NDPD, DPD, or this person is healthy. That is:

Multinomial tensor regression

Model

Let n denote the generic sample size representing the number of 3D images exploited for fitting the model. Let Y_{ij} be the binary variable that indicates whether the ith subject has NDPD, DPD, or this person is healthy. That is:

- $Y_{i1} = 1$ if the ith patient is healthy and 0, otherwise;

Multinomial tensor regression

Model

Let n denote the generic sample size representing the number of 3D images exploited for fitting the model. Let Y_{ij} be the binary variable that indicates whether the ith subject has NDPD, DPD, or this person is healthy. That is:

- $Y_{i1} = 1$ if the ith patient is healthy and 0, otherwise;
- $Y_{i2} = 1$ if the ith patient has DPD and 0, otherwise;

Multinomial tensor regression

Model

Let n denote the generic sample size representing the number of 3D images exploited for fitting the model. Let Y_{ij} be the binary variable that indicates whether the ith subject has NDPD, DPD, or this person is healthy. That is:

- $Y_{i1} = 1$ if the ith patient is healthy and 0, otherwise;
- $Y_{i2} = 1$ if the ith patient has DPD and 0, otherwise;
- $Y_{i3} = 1$ if the ith patient has NDPD and 0, otherwise.

Multinomial tensor regression

Model

Let n denote the generic sample size representing the number of 3D images exploited for fitting the model. Let Y_{ij} be the binary variable that indicates whether the ith subject has NDPD, DPD, or this person is healthy. That is:

- $Y_{i1} = 1$ if the ith patient is healthy and 0, otherwise;
- $Y_{i2} = 1$ if the ith patient has DPD and 0, otherwise;
- $Y_{i3} = 1$ if the ith patient has NDPD and 0, otherwise.

Denote $X_i \in \mathbb{R}^{79 \times 95 \times 79}$ as the three-dimensional array encoded in the sMRI scan for the ith subject.

Multinomial tensor regression

Model

The multinomial tensor classification model can be expressed as,

$$Y_{ij} \sim \text{Multinomial}\left(\mu_{i1}, \mu_{i2}, \mu_{i3}\right), \tag{1}$$

$$\log\left(\frac{\mu_{i1}}{\mu_{i3}}\right) = \langle B_1, X_i \rangle, \tag{2}$$

$$\log\left(\frac{\mu_{i2}}{\mu_{i3}}\right) = \langle B_2, X_i \rangle, \tag{3}$$

where $\langle B_k, X_i \rangle$ represents the inner product of tensor B_k and X_i for $k = 1, 2$. $B_k \in \mathbb{R}^{79 \times 95 \times 79}$ is a weight tensor in the form of

$$B_k = \sum_{r=1}^{R} \beta_{k,r}^1 \circ \beta_{k,r}^2 \circ \beta_{k,r}^3,$$

i.e., a rank-R CP decomposition of B_k, where $\beta_{k,r}^1 \in \mathbb{R}^{79}, \beta_{k,r}^2 \in \mathbb{R}^{95}$ and $\beta_{k,r}^3 \in \mathbb{R}^{79}$ are three vector components and $\beta_{k,r}^1 \circ \beta_{k,r}^2 \circ \beta_{k,r}^3$ denotes their outer product.

For the model specified in (1) to (3), we estimate the parameters by maximizing the likelihood.Given the observed imaging data X_i and the binary indicators of the three classes Y_{i1}, Y_{i2}, Y_{i3} for $i = 1, \ldots, n$, the log-likelihood function can be expressed as

$$l\left(\beta_1^1, \beta_1^2, \beta_1^3, \beta_2^1, \beta_2^2, \beta_2^3\right) = \sum_{i=1}^{n} \sum_{k=1}^{3} y_{ik} \log(\mu_{ik}), \tag{4}$$

where

$$\mu_{i1} = \frac{\exp(\langle \beta_1^1 \circ \beta_1^2 \circ \beta_1^3, X_i \rangle)}{1 + \exp(\langle \beta_1^1 \circ \beta_1^2 \circ \beta_1^3, X_i \rangle) + \exp(\langle \beta_2^1 \circ \beta_2^2 \circ \beta_2^3, X_i \rangle)},$$

$$\mu_{i2} = \frac{\exp(\langle \beta_2^1 \circ \beta_2^2 \circ \beta_2^3, X_i \rangle)}{1 + \exp(\langle \beta_1^1 \circ \beta_1^2 \circ \beta_1^3, X_i \rangle) + \exp(\langle \beta_2^1 \circ \beta_2^2 \circ \beta_2^3, X_i \rangle)},$$

$$\mu_{i3} = \frac{1}{1 + \exp(\langle \beta_1^1 \circ \beta_1^2 \circ \beta_1^3, X_i \rangle) + \exp(\langle \beta_2^1 \circ \beta_2^2 \circ \beta_2^3, X_i \rangle)}.$$

Multinomial tensor regression

Estimation

The parameters $\Theta = \left(\beta_1^1, \beta_1^2, \beta_1^3, \beta_2^1, \beta_2^2, \beta_2^3\right)$ are solved by block relaxation [*Jan de Leeuw, 1994*] as in Algorithm 1. The parameters are updated in a blockwise manner until convergence.

Algorithm 1 Block relaxation algorithm for maximizing (4).

Initialize $\left\{\beta_k^{j(0)}\right\}_{j=1,2,3;k=1,2}$ with random values

repeat $(t \geq 1)$

  for $j = 1, 2, 3$ and $k = 1, 2$ **do**

    $\beta_k^{j(t+1)} = \text{argmax}_{\beta_k^j} l\left(\beta_1^{1(t+1)}, \beta_1^{2(t+1)}, \ldots, \beta_k^j, \ldots, \beta_3^{2(t)}\right)$

  end for

until $|l\left(\Theta^{(t+1)}\right) - l\left(\Theta^{(t)}\right)| < \epsilon$

Multinomial tensor regression

Properties

The multinomial tensor regression model enjoys similar properties as the logistic tensor regression model [*Hua Zhou, Lexin Li, and Hongtu Zhu in 2013*].

Multinomial tensor regression

Properties

The multinomial tensor regression model enjoys similar properties as the logistic tensor regression model [*Hua Zhou, Lexin Li, and Hongtu Zhu in 2013*].

- Asymptotic and Consistent properties

Multinomial tensor regression

Properties

The multinomial tensor regression model enjoys similar properties as the logistic tensor regression model [*Hua Zhou, Lexin Li, and Hongtu Zhu in 2013*].

- Asymptotic and Consistent properties

  Assume the true coefficients Θ_0 are (globally) identifiable up to permutation and the array covariates X_i are iid from a bounded distribution. The maximum likelihood estimator is consistent and converges to Θ_0 in probability.

Real data analysis

- We test the performance of our multivariate tensor regression model in the real brain image MRI data, which included healthy controls, DPD and NDPD patients.

Real data analysis

- We test the performance of our multivariate tensor regression model in the real brain image MRI data, which included healthy controls, DPD and NDPD patients.
- The sMRI scans for all 480 subjects were first randomly divided into training set (80%) and testing set (20%) while retaining the DPD:NDPD:HC ratio in both sets.

Real data analysis

- We test the performance of our multivariate tensor regression model in the real brain image MRI data, which included healthy controls, DPD and NDPD patients.
- The sMRI scans for all 480 subjects were first randomly divided into training set (80%) and testing set (20%) while retaining the DPD:NDPD:HC ratio in both sets.
- We first apply our multivariate tensor regression model on training dataset to get the parameters estimates. Next, we evaluated the fitting performance and prediction performance on the training and testing sets respectively.

Real data analysis

We use the following measurement metrics to evaluate the model performance.

Real data analysis

We use the following measurement metrics to evaluate the model performance.

- Rand Index (RI)

Real data analysis

We use the following measurement metrics to evaluate the model performance.

- Rand Index (RI)
- Prediction Accuracy (PA)

Real data analysis

We use the following measurement metrics to evaluate the model performance.

- Rand Index (RI)
- Prediction Accuracy (PA)
- Multi-class extension of the AUC approach (MAUC)

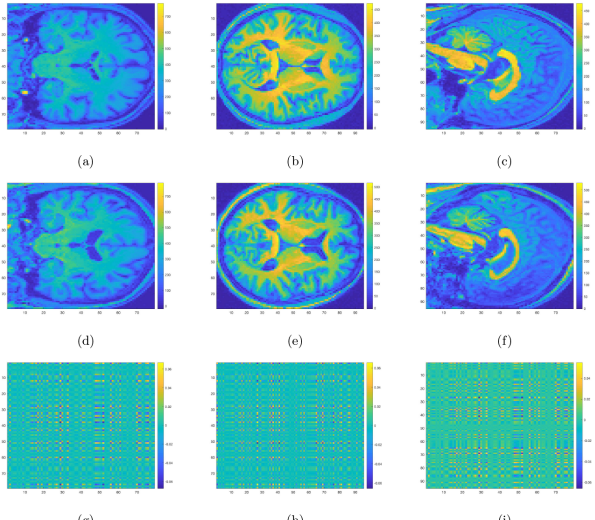
Real data analysis

The results were provided in Table 1. We found that the proposed method achieved a perfect fitting accuracy in the training set and a high prediction accuracy in the testing set. Furthermore, based on the MAUC values, our method was shown to be quite robust with respect to the varying thresholds.

Table: Model performance table in the training and testing sets.

	RI	PA	MAUC
Training	1	1	1
Testing	0.9460	0.9592	0.9457

Real data analysis



(a) (b) (c)

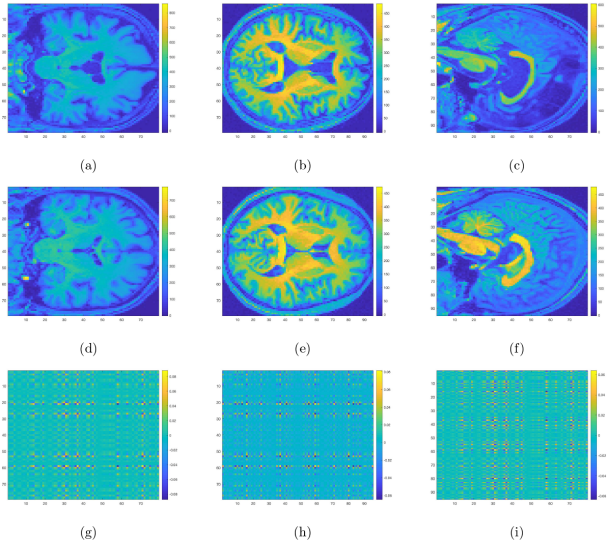
(d) (e) (f)

(g) (h) (i)

Real data analysis

In Figure 1, we drew the heatmaps for the coefficient matrices $\beta_1^1 \circ \beta_1^2, \beta_1^1 \circ \beta_1^3$ and $\beta_1^2 \circ \beta_1^3$ in model (2) corresponding to three different surfaces, and aligned the locations with significant values to the sMRI images for both NDPD patients and healthy subjects. Specifically, the disease-related alterations were found mainly in bilateral frontotemporal, occipital lobes, basal ganglia, thalamus, corpus callosum, midbrain and cerebellum.

Real data analysis



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Real data analysis

- Figure 2 allowed us to structurally visualize the differences in DPD and NDPD by plotting the coefficient matrices $\beta_2^1 \circ \beta_2^2$, $\beta_2^1 \circ \beta_2^3$ and $\beta_2^2 \circ \beta_2^3$ in model (3). In particular, compared with NDPD, DPD group displayed distinguishable differences in corpus callosum, cerebellum and the right superior temporal gyrus. At the same time, bilateral fronto-occipital lobe, left temporal lobe, bilateral basal ganglia and thalamus also showed significant differences.

Real data analysis

- Figure 2 allowed us to structurally visualize the differences in DPD and NDPD by plotting the coefficient matrices $\beta_2^1 \circ \beta_2^2$, $\beta_2^1 \circ \beta_2^3$ and $\beta_2^2 \circ \beta_2^3$ in model (3). In particular, compared with NDPD, DPD group displayed distinguishable differences in corpus callosum, cerebellum and the right superior temporal gyrus. At the same time, bilateral fronto-occipital lobe, left temporal lobe, bilateral basal ganglia and thalamus also showed significant differences.

- One major advantage of our method was that one could simultaneously reveal the most discriminative structural changes in NDPD and DPD patients.

Conclusions

- We built and validated a multinomial tensor regression based framework that leveraged 3D sMRI scans to differentiate among non-depressed PD, depressed PD and healthy subjects.

Conclusions

- We built and validated a multinomial tensor regression based framework that leveraged 3D sMRI scans to differentiate among non-depressed PD, depressed PD and healthy subjects.
- Our findings were in good agreement with the alterative features in previous studies.

Conclusions

- We built and validated a multinomial tensor regression based framework that leveraged 3D sMRI scans to differentiate among non-depressed PD, depressed PD and healthy subjects.
- Our findings were in good agreement with the alterative features in previous studies.
- Structural differences in corpus callosum, cerebellum and the right superior temporal gyrus, as well as the bilateral fronto-occipital lobe, left temporal lobe, bilateral basal ganglia and thalamus were detected between DPD and NDPD. These findings suggested disease-related alterations of structure as the basis for faulty information processing in this disorder.

References

- Hua Zhou, Lexin Li, and Hongtu Zhu. 2013. Tensor Regression with Applications in Neuroimaging Data Analysis. *J. Amer. Statist. Assoc. 108, 502 (2013), 540–552. https://doi.org/10.1080/01621459.2013.776499*

- Yajing Zhu, Xiaopeng Song, Mingze Xu, Xiao Hu, Erfeng Li, Jiajia Liu, Yonggui Yuan, Jiahong Gao, and Weiguo Liu. 2016. Impaired interhemispheric synchrony in Parkinson's disease with depression. *Scientific Reports 6 (2016). https://doi.org/10.1038/srep27477*

- David J. Hand and Robert J. Till. 2001. A simple gen-eralisation of the area under the roc curve for multiple class classification problems.*Machine Learning,45:171–186*

Thank You !