
Blog: Introduction to functional Magnetic Resonance Imaging

Subject: Scientific Research; fMRI Data Analysis

Author: Yunfei Luo

Date: Jun 26th, 2020

Abstract

With the advanced technologies of Magnetic Resonance Imaging (MRI) and Statistical Analysis, scientists are able to study human's brain in an appropriate way. In the fields of Cognitive Science, Neural Science, and Psychology, studying the relationships between the brain activity and the behaviors or psychological states reveals great potentials in clinical applications. With the inclusion of techniques in Computer Science, the scientific studies becomes more effective, which can be revealed in the interdisciplinary field: functional Magnetic Resonance Imaging (fMRI), that dealing with series of MRI images across time. This blog provides a clear and comprehensive representation of the bird-view of fMRI field based on my personal learning note of the course *Principles of fMRI 1 & 2* offered by Johns Hopkins University and University of Colorado Boulder.

1. Architecture Overview

The final goals of fMRI can be summarized as:

- i) **Localization:** Determine which parts of the brain are activated during specific task(s).
- ii) **Connectivity:** How regions are connected with one another.
- iii) **Prediction:** Use brain activity to predict perceptions, behavior, or health status.

Before we analyze toward the interests above, we first need to design an appropriate experiment to approach the tasks that we are interested in. Block Design and Event-related Design are the two main types of experimental design which will be discussed in details in the following sections.

Then we need to collect data, and analyze the data.

The proper analyzing modes are Tradition Brain Mapping, which is a classical approach in science, and Multivoxel (or Multivariate) Pattern Analysis (MVPA) which is often related to techniques of Machine Learning. These modes will be discussed in details in the later sections.

Moreover, considering that there are many variability in experimental design, model selection, and the varies source of noises such as head movements, heart rate, and respiration, etc., the quality control is significant across the entire fMRI project as shown in Figure 1. Again, more details will be discussed in the later sections.

2. Experimental Design

During the experiments, we want the subjects to do the tasks we want, and detect the effects based on the observations on the fMRI data of the brain. For example, we may want to see how brain responses if we ask the participants to look at images that could be either positive (make people feel comfortable, could be happy or enjoyment) or negative (make people feel bad, such as sad and pain). The question is now revealed: how should we arrange the order of tasks that exposes to or done by the subjects in order to effectively detect the contrasts or responses from the brain activity? The two main strategies are Block-Design and Event-Related-Design, as shown in Figure 2.

2.1. Block Design

In Block design, the similar events are grouped into a single block. The length of the block is fixed once set during the experiment.

Each blocks have the same length, often about 16 to 20 seconds. If the block length is too short, there might not have enough time space for the brain to show differences among different tasks, which often caused by vascular inelasticity. If use long blocks, more than 40 seconds for example, it is risky that there will be more unwanted activation caused by the out-of-scope association and imagination done by the subjects.

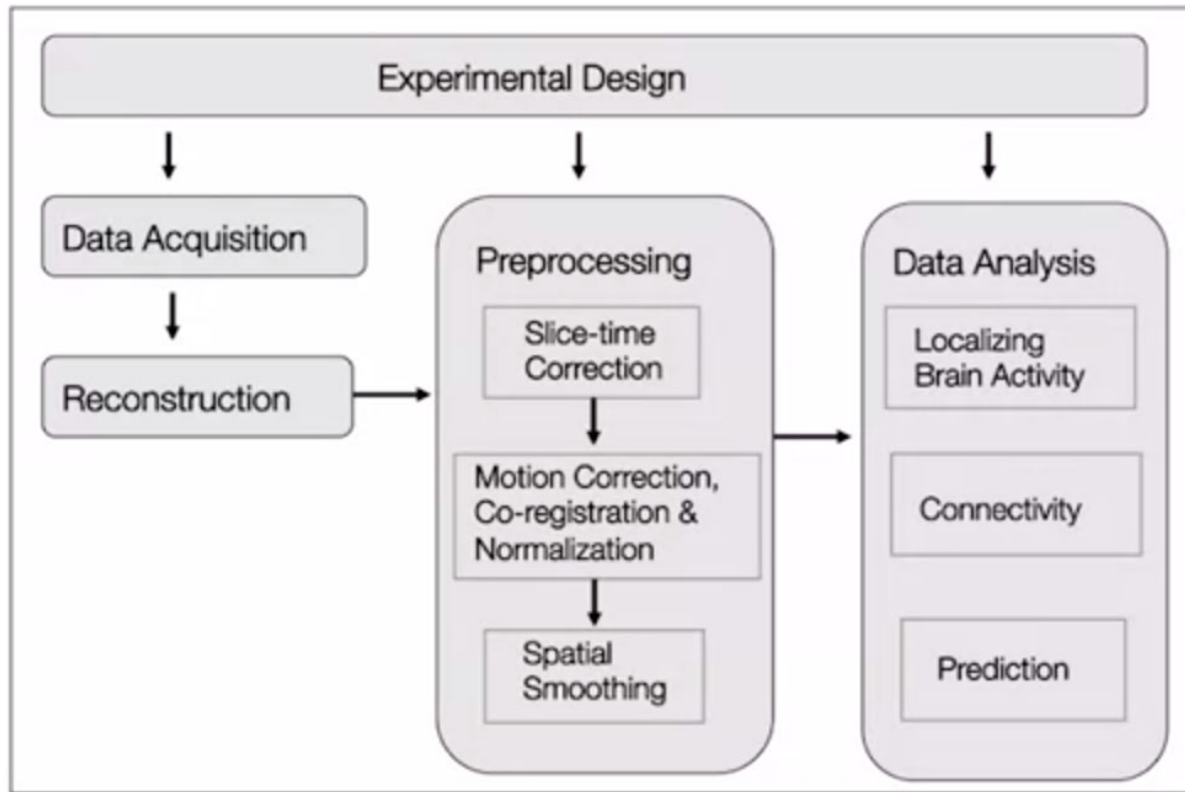


Figure 1. Bird-view of Goals of fMRI (from course slide of Principles of fMRI 1 & 2)

2.2. Event-Related Design

In this design mode, the tasks are mixed in orders, with unequal length.

For this kind of design, there are no strict requirements on the boundaries of the events' length, but if there are low-frequency components, such as a long gap between two events, it is still risky with the same reasons as mentioned in Block-Design.

2.3. Quality Control

The Rule-of Thumb is to have 30 to 40 minutes of functional scan time for each subjects considering the psychological effects such as fatigue and habituation. And scan as many participants as possible for better generalization and robustness of the study.

The tradeoffs among different factors always exist. The block design is robust in detect the contrast of responses of brain among different tasks. However, the blocks always comes in same length and same order. Considering the effects of Neural Habituation, i.e. the subjects will know what will be the next block, and will be

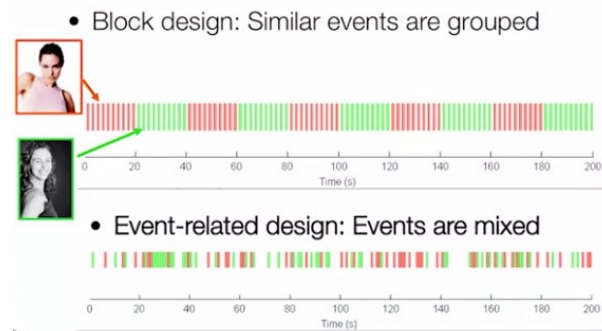


Figure 2. Sketch figure of Block Design and Event-Related Design (from course slide of Principles of fMRI 1 & 2)

subconsciously prepared for that, the contrast results is not sufficient for making specific psychological or behavioral inferences. On the other hand, although event-related design might not provide a clear contrast in responses among the tasks, it has better estimation to the shape of hemodynamic responses which is significant for making specific inferences.

In terms of efficiency, having 2 conditions (2 different tasks) is optimal in either design mode. And the experimental efficiency has order:
Block-Design > Dense-Event-Related-Design > Sparse-Event-Related-Design

In the cases where the number of conditions ≥ 2 , Genetic Algorithm (GA) is often used for searching in the random space of the order of tasks exposes to the subjects. Details of GA will not be discussed here.

3. Data Preparation

3.1. Data Structure

There are two main types of data, Structural (T1) images, which is a single MRI images; and Functional (T2*) images, which is a series of MRI images across time, which we refer to as fMRI. At each time step, the images consist of 3 images from different direction: Coronal, Sagittal, and Axial, as shown in Figure 3.

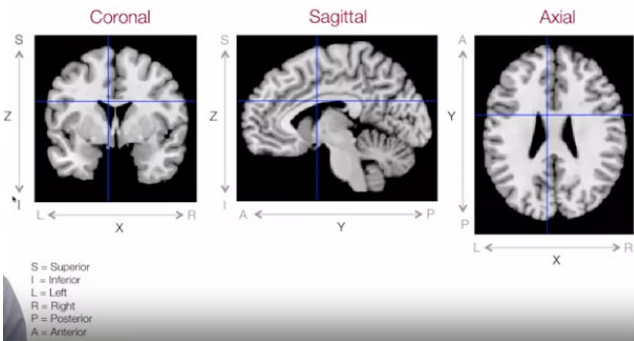


Figure 3. data structure of fMRI (from course slide of Principles of fMRI 1 & 2)

Basically an MRI images consist of number of voxels with their value of intensity, which can also be interpret as the pixels in 3 dimensions. So in fMRI, we have a time series data for each voxel as shown in Figure 4.

3.2. Data Acquisition

Physically, the data is collected by controlling the motion of nuclei of hydrogen atoms of the subjects. Each nuclei is spinning at some directions, and has its own magnetic field. When the subjects is placed in the Magnetic Resonance (MR) scanner, the nuclei will align with the magnetic field of the equipment and form an **Longitudinal Magnetization**. The nuclei now have the same directions of spinning, but could have different phases.

Then during the scanning, a Radio Frequency (RF)

- Each voxel has a corresponding time course.

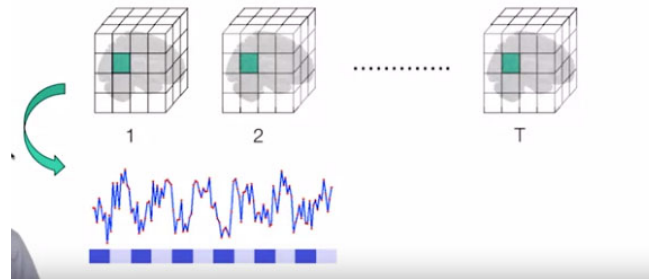


Figure 4. Indication of voxels (from course slide of Principles of fMRI 1 & 2)

pulse is performed. The phases of the nuclei will be aligned, and will be tipped over and form a **Transversal Magnetization**. The degree for such tip-over is often 90 degrees.

Then we remove the RF pulse, and the transversal Magnetization will decrease exponentially, and longitudinal magnetization will increase exponentially back to its original size. The signal created during this process will then be caught by a receiver coil. The signal is then map to a image as shown in Figure 3 by Fourier Transfer.

3.3. Blood Oxygenation Level Dependent (BOLD)

Studying oxygenation changes in the brain across time is the most common approach in fMRI. It measures the ratio of oxygenated to deoxygenated hemoglobin in the blood. Because oxyhemoglobin is diamagnetic and deoxyhemoglobin is paramagnetic, their difference in magnetism allow us to catch the signal of how these two types of hemoglobin distributed and changes in the brain.

Since when the neurons are activated, they will absorb oxyhemoglobin from the neighbor blood vessels, we doesn't measure the neural activity directly, but metabolic demands (oxygen consumption) of the active neurons. As the neurons absorb oxyhemoglobin, there will be a quick decrease in oxyhemoglobin and increase in deoxyhemoglobin, which will be shown as an decreases of magnetic resonance signal. But suddenly, as the body will sending more oxyhemoglobin to reply the oxygen consumption, there will be a significant increase in oxyhemoglobin and decrease in deoxyhemoglobin, which will result in more magnetic resonance signal. After about 4 to 6 seconds, the intensity of the signal will reach the peak, and start reduces below the baseline, then go back to the original situations.

The entire process demonstrate an delayed responses and is quite slow that it would take about 20 to 30 seconds. The signal vs. time wave is known as Hemodynamic Response Function (HRF), which we will discuss more about it in the section of analysis.

3.4. Noise & Registration

There are always noise during the scanning, could be artifacts or errors that are inevitable. Head motion, heart beat, and respiration are often great concerns in fMRI. In order to dealing with such noise, an Intrasubject registration is needed, where we need to ensure that each voxels across time always represent the same area. However, the motion often influence the the magnetic field which will leave a "spin-history" of the nuclei, which cannot be entirely removed.

Moreover, for the purpose of robustness and generalization of the study, a spatial normalization is needed, which often refer to as Intersubject registration. One subject will often be selected as the "standard", and the data from other subjects will be manipulated to align with the standard. Such normalization allows researchers to come up with the results making sense across the population.

3.5. Quality Control

As we mentioned in the previous section, there are always tradeoffs that need to deal with. Echo-Planer Imaging (EPI), a standard and widely used techniques for obtaining functional images, collects images slice-by-slice from bottom to top of the brain. Figure 5 briefly shows the tradeoff relationships of EPI among brain coverage, Spatial and Temporal resolution, and susceptibility artifacts.

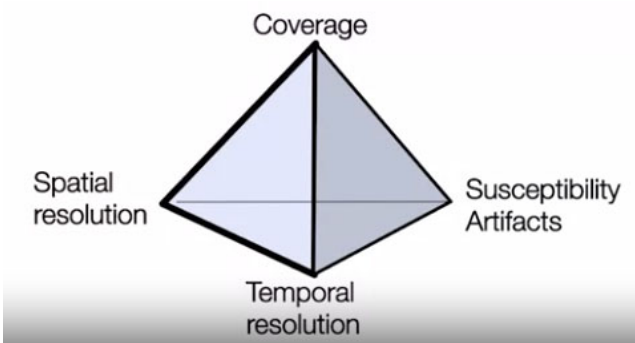


Figure 5. Tradeoffs in different aspects of data (from course slide of Principles of fMRI 1 & 2)

Firstly, more coverage of the brain will help a lot in registration and normalization, but there will be less temporal resolution and more susceptibility artifacts. Secondly, although

higher resolution often reduce the susceptibility artifacts, spatial and temporal resolution always cannot be compatible. Higher spatial resolution lead to better decoding and prediction, especially in small regions activity. But there are cost in temporal resolution. Higher temporal resolution has better ability to separate and remove artifacts but has cost in spatial resolution, time, and coverage. Finally, higher susceptibility artifacts will provide better fidelity, especially localization in problematic brain areas.

4. Data Analysis

4.1. Traditional Brain Mapping

Usually, the analyzing tasks use mass univariate approach, which is also known as traditional brain mapping, as shown in Figure 6. In this approach we modeling for each voxel separatly and independently.

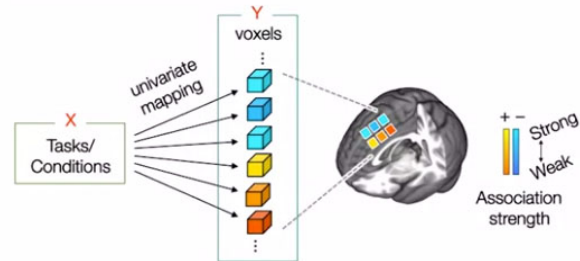


Figure 6. Traditional brain mapping (from course slide of Principles of fMRI 1 & 2)

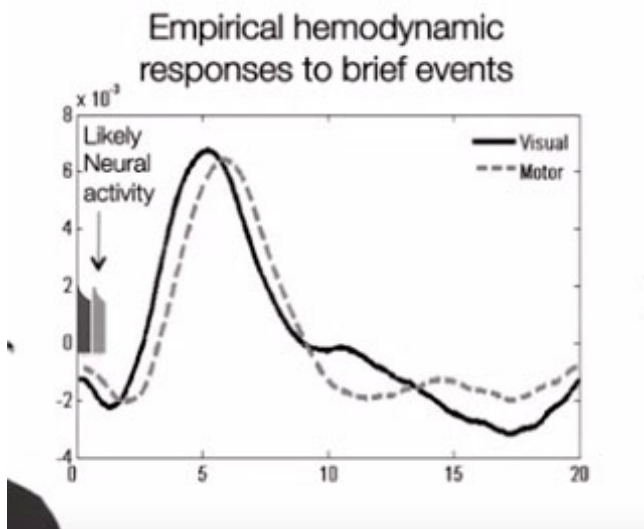


Figure 7. Fitting with assumed HRF (from course slide of Principles of fMRI 1 & 2)

4.1.1. GENERAL LINEAR MODEL (GLM)

First of all, we will fit a General Linear Model (GLM) for each voxel as shown in Figure 7 with:

$$Y = X\beta + \epsilon$$

Where Y is the observed time series data of the single voxel. X is the design matrix, which is often the assumed canonical HRF, or matrix that contain canonical HRF, its derivative, and its dispersion derivative. Finite Impulse Response (FIR) is also a proper choices for the design matrix. β is the regressors, and ϵ is the error term. By perform the fitting task, we are solving the optimization problem:

$$\min_{\beta} ||Y - X\beta||_2^2$$

We can easily solve this problem by taking the derivative with respect to β and set it to zero. We finally arrive the Ordinary Least Squares (OLS):

$$\beta^* = (X^T X)^{-1} X^T Y$$

Where the diagnol of the matrix $(X^T X)^{-1}$ is defined as the design efficiency, which is a significant metric to evaluate the tradeoffs. However, OLS is only optimal if the error or the noise term ϵ is i.i.d (Independent and Identical Distributed). Otherwise, we will introduce a weighted matrix W that corresponds to the actual variance:

$$\beta^* = (X^T W X)^{-1} X^T W Y$$

which is the Weighted Least Squares (WLS) that can be solved by iterative algorithms.

4.1.2. STATISTICAL TEST

After we fit the GLM for each voxel, we can perform a statistical test with the regressors we have. Usually, we construct a contrast vector c that represent the null hypothesis, denoted as H_0 . Then we conduct a t-test:

$$t = \frac{c^T \beta^*}{\sqrt{\text{Var}(c^T \beta^*)}}$$

Where

$$\text{Var}(c^T \beta^*) = \sigma^2 c^T (X^T X)^{-1} c$$

where σ corresponds to the residual noise, and $(X^T X)^{-1}$ is the efficiency matrix as mentioned before.

The t score represent the relative likelihood of the null hypothesis to be rejected. It could also be F-test if we have contrast matrix, or we could calculate the p-values, z-values, or Fisher scoring (The choices of the statistical test depends on varies aspects of the performed experiment). We then reconstruct the brain images with the scores of each voxel, and setting threshold to visualize which area voxels are activated.

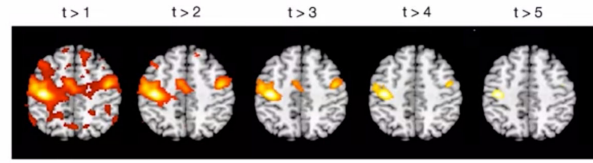


Figure 8. Comparisons of selecting different threshold on t-scores (from course slide of Principles of fMRI 1 & 2)

4.1.3. QUALITY CONTROL

Quality Control exists every where in fMRI project. Selecting appropriate threshold should be treat seriously. As shown in Figure 8, different setting in threshold will directly influence the conclusion of which areas of the brain are activated. Two types of errors are now revealed:

- i) False Positive, where H_0 is true, but we reject it, i.e. the voxel is not activated, but we determine it as activated.
- ii) False Negative, where H_0 is false, but we failed to reject it, i.e. the voxel is activated, but we determine it as not activated.

So basically, we are dealing with the tradeoffs between sensitivity (true positive rate) and specificity (true negative rate). There are two main approaches to the tradeoffs:

- i) voxel based, i.e. seek in finding an upper-bound of threshold normalized by the number of voxel.
- ii) Cluster-level Inference, an intuitive way to filter out the peaks.

There are numbers of algorithms in each approach. Here I only provide the most common used strategies.

A voxel based approach called False Discovery Rate (FDR) becomes popular recently that focus on the proportion of false positive rate among rejected tests. Benjamini-Hochberg is the most popular FDR algorithm:

- i) Rank our scores, with $p_1 \leq p_2 \leq \dots \leq p_m$ where m is the total number of voxels.
- ii) find the largest i such that $p_i \leq \frac{i}{m} * q$ where q is our selected desired limit (usually 0.05).
- iii) Finally, we reject all hypothesis corresponding to p_1, p_2, \dots, p_r .

Cluster-level Inference is more widely used than voxel based. Threshold Free Cluster Enhancement (TFCE) is the popular one, where we control the threshold by two hyper-parameters: u_c and the area, as shown in Figure 9.

Though many packages implemented these algorithms with some default values, it is still worth to manually check and manipulate for the proper threshold.

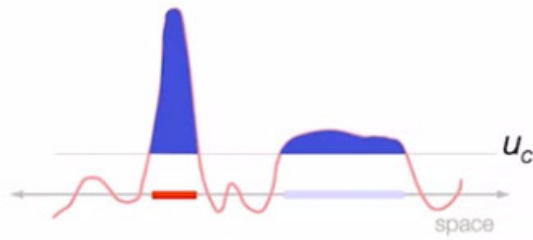


Figure 9. Combine cluster size with intensity information (from course slide of Principles of fMRI 1 & 2)

4.2. Connectivity

4.2.1. FUNCTIONAL CONNECTIVITY

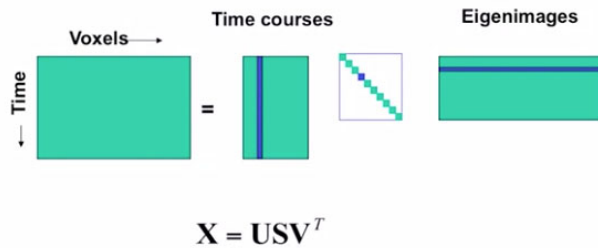


Figure 10. SVD Decomposition (from course slide of Principles of fMRI 1 & 2)

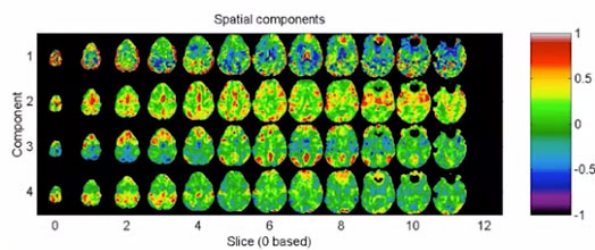


Figure 11. Visualize Principal Components (from course slide of Principles of fMRI 1 & 2)

This type of connectivity seek to make inferences on the structure of the relationships among brain regions. In terms of graph theory, the relationships are often undirected graph. Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are widely used algorithm for studying function connectivity. Let's say we have matrix X where each row represent the reshaped voxels at each time step. There are some underlying "components" in X

that represent the extent of association of different areas of the brain. PCA decompose the matrix X in three matrix as shown in Figure 10, where each $S_{ii}U_i \otimes V_i^T$ is a principal component, and S_{ii} represent the extent of importance of this component. As we can see, each component has the same shape with X , so they are also the time series images, and can be visualized as shown in Figure 11.

ICA on the other hand, decompose matrix X into two matrix, as shown in Figure 12, and we can visualize the components as we do with PCA, as shown in Figure 13.

PCA assumes an orthogonality among the components, and ICA assumes statistical independence among the components. Usually, independence is a stronger requirement than orthogonality (which we can also see that ICA provide better visualization than PCA), but ICA didn't rank the components which PCA does.

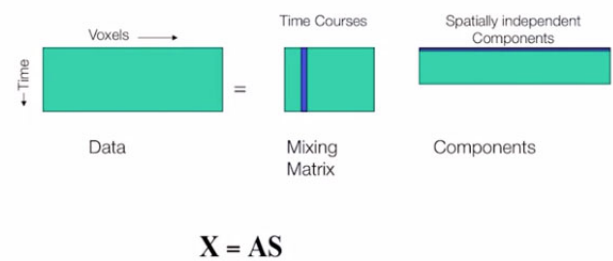


Figure 12. ICA decomposition (from course slide of Principles of fMRI 1 & 2)

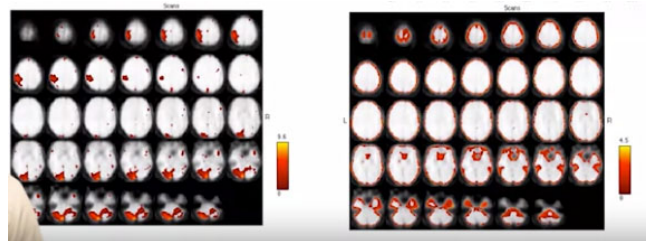


Figure 13. Visualize Independent Components (from course slide of Principles of fMRI 1 & 2)

4.2.2. EFFECTIVE CONNECTIVITY

Effective connectivity makes stronger conclusions than functional connectivity. In terms of graph theory, the relationships often represented as directed graph. To study the effective connectivity, there two main approaches: mediation and modulation.

As shown in Figure 14, mediation is the study that seek in

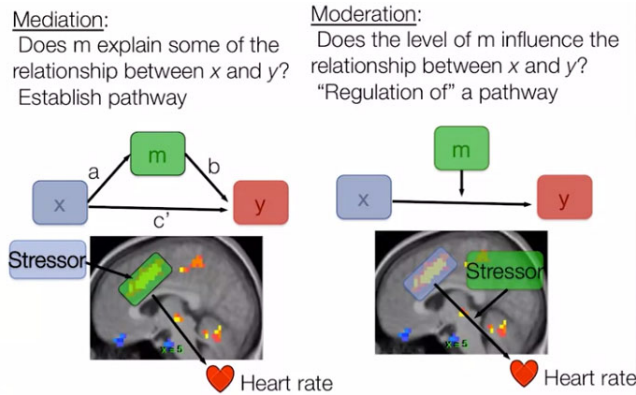


Figure 14. Mediation and Moderation (from course slide of Principles of fMRI 1 & 2)

establish the relationships among factors, and moderation seek in finding the conditional relationships among factors. For example, we want to study the relationship between the activation of brain area x and the heart rate. With mediation, we will explore what factor will cause the activation in area x , pain or enjoyment for instance, and come up with conclusion based on the experimental results. With moderation, if we assume that pain will activate area x , then we will change the extent of pain, and see if the correlation between the activation of area x and heart rate still exists.

4.3. Multivoxel (or Multivariate) Pattern Analysis (MVPA)

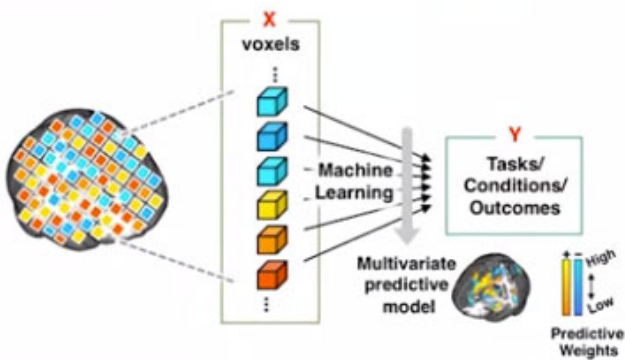


Figure 15. Multivoxel (or Multivariate) Pattern Analysis (from course slide of Principles of fMRI 1 & 2)

Because the traditional brain mapping uses mass-univariate analysis approach, every voxel is modeled separately and independently. There are no interaction information included, which is a core limitation of such approach. With more data becomes available, techniques in machine learning becomes applicable on fMRI data. MVPA as shown in Figure 15,

seek in assessing whether a pattern across voxels predicts a behavior or outcome. This approach takes whole brain MRI or fMRI image(s) X as input, and supervised by the output Y that could be the conditions at next time steps (Regression), or the labels of either the behavioral or psychological states (Classification).

4.3.1. ALGORITHM SELECTION

With regression tasks, here are some example choices:

- i) Linear Regression (e.g. Ridge, Lasso, etc.)
- ii) K-Nearest-Neighbor (KNN)
- iii) Decision Tree

With classification tasks, the example choices are:

- i) Logistic Regression (Linear Classification with logistic loss function)
- ii) Support Vector Machine (SVM)
- iii) Neural Networks

Especially for the choices of Neural Networks, since MRI data is essentially image data, Convolution Neural Networks could be an appropriate choices; and since fMRI data is essentially a time series data, Recurrent Neural Networks such as Long-Term-Short-Memory (LSTM) and Gated Recurrent Units (GRU) are worth to try.

For the space concerning, details about these machine learning models will not be discussed here. If you are interested in these techniques, there are abundant online resources available. You could also read my other original articles about the applications in other fields such as:

- i) *Report: Short Term and Long Term Cases Forecasting of COVID-19 with Varies Neural Networks*, available at: https://yunfeiluo.com/share.html?article=articles/reports/covid_19_cases_forecasting.pdf
- ii) *Paper review: A Neural Algorithm of Artistic Style*, by Gatys, Ecker, and Bethge, available at: https://yunfeiluo.com/share.html?article=articles/blogs/paper_review_artistic_style.pdf

4.3.2. MODEL EVALUATION

As always, the quality control in this analysis mode is done by evaluating the performance of the model. For regression tasks, the performance of the model on test set are often Mean Square Error (MSE), or Mean Absolute Error (MAE). For classification tasks, the evaluating matrices are often F1-Score, or AUC-ROC.

For splitting the training and testing set, k-fold cross validation is a widely used method. The procedure is:

- i) split data evenly into k folds (Could be stratified by the

classes, in order to ensure evenly exposure of each class for training)
ii) for each fold $i \in \{1, 2, \dots, k\}$, train model on the rest folds, and testing on fold i
iii) average the score (depends on the evaluating metric) across all the folds.

Epilogue

As a interdisciplinary subject that converged with Biology, Cognitive Science, Computer Science, Physics, Psychology, and Neural Science, there are a plenty of studying directions in the field of fMRI. This blog only provide an bird-view of conducting fMRI experiments and analyzing brain data. Personally, I'm interested in applying the machine learning techniques into the analysis. I believe that the developing with reasonable algorithms and models are extreme beneficial to the clinical applications. I will keep learning and exploring in this direction.

Bibliography

Martin Lindquist, Tor Wager, *Principles of fMRI 1*, Coursera, Access at:
<https://www.coursera.org/learn/functional-mri>
Last Accessed: Jun 25 th 2020

Martin Lindquist, Tor Wager, *Principles of fMRI 2*, Coursera, Access at:
<https://www.coursera.org/learn/functional-mri-2>
Last Accessed: Jun 25 th 2020