

Dimension Reduction of Diffusion MRI measures of White Matter Tractometry

Leqi Tang

Junior Paper, Psychology Honor Program

Department of Psychology

University of Washington

September 23, 2021

Abstract

Diffusion magnetic resonance imaging (dMRI) is one of the most popular and widely used MRI techniques to assess white matter tracts and connections between different brain areas. Derived from each imaging voxel and each vertex, the high-dimensional data of various diffusion MRI measures makes it difficult to analyze and visualize certain tissue microstructural properties. In this proposal, we investigate three dimensionality reduction algorithms: principal components analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and non-negative matrix factorization (NMF) to perform data-reduction analysis. A sample of 652 adults aged 18-88 years collected by The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) was scanned using MRI. The most commonly used two dMRI measures: the fractional anisotropy (FA) and mean diffusivity (MD) across 18 brain pathways were profiled in the data. We compare the three techniques in terms of the loss of quality produced by them and assess the effects of different numbers of reduced dimensions. Together, we select the most appropriate dimensional reduction method to further unfold the characteristics of white matter regions in terms of minimum loss of quality.

Introduction

White matter in the tissue in the human brain is composed of millions of nerve fibers, which connect various grey matter cortex. The axon conducts nerve impulses between neurons, which are special cells designed to transmit information between different areas of the brain, muscles, and other nerve cells. With rising observations of white matters, the properties of the brain white matter tissues have been introduced to correlate with cognitive abilities, decision making, emotional states, and developmental changes (Fields et al. 2014; Leong et al. 2016; Wandell & Yeatman 2013; Yeatman et al. 2011, 2012a). Thus, it is crucial to examine the properties of white matters in further depth.

Multiple techniques, including diffusion MRI, tractography, and quantitative MRI (qMRI), have been developed to measure the tissue property of human brain development. As a non-invasive technique, diffusion MRI has become the preferred tool to measure molecular diffusion in brain tissues. The MRI instrument excites the spins, hydrogen in free water in white matter, to produce a signal, and the spin attenuates the signal over time (Wandell, 2016). Moreover, the size of signal attenuation can be measured in each direction by applying appropriate magnetic field spatial gradients. Signal attenuation in a particular direction, θ , is measured by the apparent diffusion coefficient $A(\theta)$ (Le Bihan & Johansen-Berg 2012, Stejskal & Tanner 1965). The signal is attenuated by an amount that depends on the direction of the gradient and also the amount of diffusion along that direction (Wandell, 2016). With high values of the diffusion at a specific direction of a white matter voxel, the signal attenuates to a greater extent. And the size of the signal measured becomes weaker. In order to acquire the diffusion coefficient, a b-value which is crucial to note presents a set of physical

constant and experimental parameters. Since a b-value is integrated with a set of measurement conditions, the same b-value could represent different combinations of gradient strength and duration in order to measure different tissue properties.

Diffusion MRI measures, derived from the diffusion signals, can be obtained in each imaging voxel from all orientations with diffusion tensor imaging (DTI) (Basser et al., 1994). A diffusion tensor model has also been developed to model the dMRI signal within a voxel. The diffusion tensor model treats diffusion as an anisotropic Gaussian process (Basser & Jones 2002, Filler 2009). The tensor model uses six parameters to describe the Gaussian diffusion within each voxel. Besides, the DTI provides several scalar statistics: fractional anisotropy (FA), which reflects the degree of diffusion anisotropy (Pierpaoli and Basser, 1996) and mean diffusivity (MD), an indicator of the overall magnitude of diffusion. The combination of streamlines would visually reconstruct the geometry of the brain white matter tracts (Mori and Van Zijl, 2002). In gray matter, cell membranes hinder the movement of water molecules. The neuroimaging researches of dMRI measures gives plenty of insights into the correlation between white matter and human cognitive, developmental, and behavioral activities. It is remarkable to note that the measures of diffusion anisotropy and mean diffusivity acquired by diffusion tensor magnetic resonance imaging (DT-MRI) demonstrate the age effect in schizophrenia patients. The measures are capable of presenting the age-related differences in brain maturation (Jones et al., 2006). In this proposal, we will apply k-means clustering, a method of vector quantization, to examine the age effect on our high-dimensional data.

Moreover, the high resolution of dMRI data also provides novel findings between white matter tracts and social impairment in autism spectrum disorders (Ameis & Catani 2015), white matter anomalies detected in anterior corpus callosum associated with a speech disorder (Civier et al. 2015), and other human behavior. However, although the quantitative measurements may specify various white matter properties, diffusion tensor imaging (DTI) has limitations that could not be overlooked. For example, FA and MD lack specificity to the various physical properties of white matter, such as crossing fibers (Jeurissen et al., 2013). Besides, the average profile of the dMRI measures may vary along a given white matter pathway depending on the underlying fiber architecture (Yeatman et al., 2012).

With the influx and rapid development of data science and modeling, several new frameworks have been proposed to have a better understanding of white matter imaging data. High angular resolution diffusion imaging (HARDI) is capable of discriminating multiple fiber populations crossing within the same voxel. The higher angular resolution allows to have a more precise representation of the 3D pattern of water diffusion within a voxel (Berman, Lanza, Blaskey, Edgar & Roberts, 2013). Besides, multi-shell acquisitions have also been shown to improve the angular resolution of orientation distribution functions (ODFs) (Descoteaux et al., 2011; Jeurissen et al., 2014; Chamberland et al., 2018). The dMRI data combined with tractography algorithms would be able to estimate relatively streamlines, which are the underlying trajectories at each voxel between cortical regions. Multiple tractography algorithms have been developed to estimate relatively long fascicles (streamlines) (Wandell, B., 2016). Tractography allows us to disentangle the white matter that carry biological information between cortical

regions. It is expedient for researchers to have a more comprehensive assessment of white matter microstructures and their connections.

Diffusion Imaging in Python (Dipy) is a free and open source software project focused on 3D and 4D+ neuroimaging. It is developed using the Python language. Dipy contains many different methods for computational anatomy including diffusion signal pre-processing, reconstruction of diffusion distribution in individual voxels; fiber tractography and fiber track post-processing, diffusion and structural visualization (Garyfallidis et al., 2014). Python is a general purpose, object-oriented programming language which was designed with an emphasis on code readability. The algorithm investigated in this proposal will also be implemented in Python. The new frameworks proposed, including along-tract profile and various tractography algorithms are worthwhile to note that they are capable of providing higher sensitivity to microstructural features of fiber pathways by mapping different MR-derived measures over white matter bundles (Chamberland et al., 2019).

In the field of functional MRI, independent component analysis (ICA) has been analyzed to be a successful example of unsupervised dimensionality reduction that allows the extraction of temporally segregated resting-state networks (Beckmann et al., 2009).

However, the high-dimensional data derived from dMRI imaging of each voxel and bundle and the large number of participants give much difficulties on processing and analyzing neuroimaging scan data. The dMRI measures often share redundant predictor variable and experimental noise and ultimately decrease statistical power if strictly correcting for Type I errors. With machine learning techniques are increasing being used in making inferences on

individual subjects neuroimaging scan data and relevant statistical analysis, dimension reduction methods, which have been used in the neuroimaging voxel data, become an essential step before training a machine learning mode to avoid overfitting and therefore improving model prediction accuracy and generalization ability (Mwangi, Tian & Soares, 2013).

In this proposal, we mainly focus on analyzing the capability of successful feature reduction of three commonly used dimension reduction algorithms: principal components analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and non-negative matrix factorization (NMF). Specifically, we will perform dimension reduction on the data with different parameters and hyperparameters, including the number of reduced components, to investigate the effect of various parameters on the reduction's accuracy. Besides, we will explore appropriate statistical inference analysis to compare the three-dimension reduction methods, in terms of the loss of quality comparing with the original data information. The low dimension data reduced and extracted will be used in further analysis of white matter region closeness and characteristics (to be discussed). The algorithms that is the most appropriate for neuroimaging data analysis will be capable to be implemented in relevant Python module.

Methods

Participants

This study uses the data collected from the Stage 1 and Stage 2 from the Cam-CAN project. Nearly three-thousands adults aged 18-90 completed a home interview, and a subset of six-hundred and fifty-two participants aged from 18-88 years (mean = 54 ± 18) (the "CC700") were scanned using structural Magnetic Resonance Imaging (MRI) and completed multiple cognitive experiments. The sample has three-hundred and thirty females and three-hundreds and twenty-two males.

Tractography and tractometry

Diffusion weighted imaging (DWI) was performed to acquire structure data with using 2D twice-refocused spin-echo echo-planar imaging (DW-SE-EPI) (TR=9100ms and TE=104ms, TI=900ms; FOV=192x192mm; 66 axial slices with 2mm isotropic voxels) implemented. The imaging was gotten with the use of 30 encoding directions, two b-values (1000, 2000 s/mm^2), and B0=0. The echo-planer imaging (EPI) had 0.0684 readout time with echo spacing=0.72ms and EPI factor=96 (Minett, 2011).

Eighteen bundles of interest were then interactively dissected in the native space of each subject using a combination of include and exclude regions of interests (ROIs).

The virtual dissection plan includes:

Corticospinal Tract (CST) (bilateral): the CST ascends from the brainstem, paralleling the ventricles to the cortex. This is the major neuronal pathway providing voluntary motor function.

Uncinate Fasciculus (UNC) (bilateral): the UNC connects parts of the limbic system, traveling in a posterior-medial direction, to the orbitofrontal cortex.

Inferior Fronto-Occipital Fasciculus (IFOF) (bilateral): the IFOF located from the occipital lobe through the external capsule to the inferior frontal cortex.

Inferior Longitudinal Fasciculus (ILF) (bilateral): the ILF is a long-range pathway that connects the occipital and temporal-occipital areas of the brain to the anterior temporal areas.

Superior Longitudinal Fasciculus (SLF) (bilateral): the SLF runs through the pathway between dorsal parietal lobes to frontal lobes and provides partial communication with the temporal lobe.

Anterior Thalamic Radiation (ATR) (bilateral): the ATR is the bundle that connects the prefrontal cortex and the thalamus through the anterior limb of the internal capsule.

Corticospinal Tract (CST) (bilateral): the corticospinal tract is a white matter tract that starts with cerebral cortex and terminates at lower motor neurons and interneurons in the spinal cord. CST is responsible for controlling movements of the limbs and trunk.

Cingulum (CGC) (bilateral): a white matter pathway that connects between the frontal lobe, parietal lobe and temporal lobe.

Using the Pandas Python library, we will access to the diffusion measures data and take average along each segment of the pathway for each participant. This is a commonly used method to merge the quantitative nature of diffusion measures with the qualitative nature of tractography is to collapse voxel-based measures into a single scalar value per bundle (Jones

et al. (2006); Kanaan et al. (2006); Jones et al. (2005a)). Then the “SimpleImputer” Python object will be applied to impute missing values on the data with a calculated statistic.

If necessary, bundles will be pruned to remove outliers as in (Cousineau et al., 2017; Garyfallidis et al., 2018). The original data with outliers will also be computed by three-dimension reduction algorithms to examine and compare the influence of outliers to the algorithms.

Dimension reduction

The dMRI measures data comprised 18 white matter regions with 641 participants’ imaging data whether collected or imputed. In this proposal, we investigate three dimension reduction methods: Principal component analysis (PCA), Non-negative matrix factorization (NMF), and T-distributed Stochastic Neighbor Embedding (t-SNE) to explore the possible redundancy of each white matter region and compare the accuracy of each technique with regards to the loss of the original data’s quality.

Principal component analysis (PCA)

The PCA is a technique for reducing data’s dimensionality to increase interpretability but at the same time minimize information loss. The PCA breaks down a multidimensional data into a set of orthogonal ‘modes’ and corresponding scalar values.

For a high-dimensional matrix X , we have the covariance matrix which is a symmetric matrix as: $C = \frac{1}{n-1} AA^T$. The matrix C can be diagonalized and give:

$$C = VSV^T$$

where we have V as a matrix of eigenvectors and S as a diagonal matrix with eigenvalues σ_n . The values σ_n are the singular values that are nonnegative and ordered from largest to smallest. The eigenvectors are called principal directions. The projections of the data on principal axes are principal components, which are orthogonal to each other. In this way, a significant proportion of the variance in the white matter data can be explained by a reduced number of orthogonal components, compared to the total number of raw input variables (Chamberland et al., 2019).

The PCA technique has benefits that tries to preserve the global structure of the data and process independent features of the data. However, it might get highly affected by some outliers and overfit.

Non-negative matrix factorization (NMF)

The NMF decompresses the non-negative data matrix into the product of two matrices. For a matrix X , we have:

$$X = W * H$$

where X is the data set matrix and we assume it is an $n*m$ matrix. The matrix W is the basis matrix that has $n*p$ values; while the matrix H is the weight matrix that consists $p*m$ values. Both lower-dimensional vectors are non-negative. Every column of H is the projection of the data X onto W .

As a feature extraction algorithm, NMF obtains a parts-based representation and is compatible with the intuitive notion of picking out additive combinations and combining parts to form a whole.

T-distributed Stochastic Neighbor Embedding (t-SNE)

T-SNE is one method of nonlinear dimensionality reduction methods, which are capable of separating data that cannot be separated by any straight line. The t-SNE has been applied in several research areas including bioinformatics, biomedical signal processing, and cancer research. The t-SNE algorithm first calculates the Euclidian distances of each data from all of the other data. Then the conditional probability $p_{j|i}$ of each data x_j to be next to the other data x_i is represented by a Gaussian distribution centered at x_i .

The conditional probability is stated as:

$$p_{j|i} = \frac{g(|x_i - x_j|)}{\sum_{k \neq i} g(|x_i - x_k|)}$$

where $g(|x_i - x_j|)$ represents the value of $\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)$, which is the perplexity of a data.

The condition probability will be divided by the sum of all the other points placed at the Gaussian center at x_i for the purpose of normalization. We calculate the joint probability from the conditional distributions with following equation:

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$$

Then we calculate the joint probability for another time but with the t-distribution for the original data, because of the heavy-tails property of the t-distributions. The probability here with t-distribution is marked as q .

$$q_{ij} = \frac{(1 + \|x_i - x_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|x_k - x_l\|^2)^{-1}}$$

We use the Kullback-Leiber divergence (KL divergence) to make the joint probability distribution of the data in the low dimension. The t-SNE is capable of preserving the local structure of the data. This is due to the use of the KL divergence as a measure of the difference between the two distributions with respect to the distance of data (van der Maaten and Hinton, 2008). With the focus on local structure, the model will not be highly affected by outliers.

Each dimension reduction method will be applied to the dMRI dataset, and we will explore their effects and the effects of different number of low dimensions. For t-SNE, we will manually tune the multiple hyperparameters such as the cost function parameter perplexity, and the optimization parameters the learning rate, and the number of iterations for better resolution of the dimension reduction. The low dimension datasets acquired by three methods will be displayed as a scatterplot. The X and Y axis in all the scatterplots are the values of any two dimensions from the low dimension data.

We will try to reconstruct the original data from each of the reduced datasets by certain calculation. To compare the loss of quality, we will compute the square root of the average of the square of the difference between the reduced data and the original data for each participant.

The statistical inference for verifying the redundancy reduced by utilizing three algorithms will be further discussed.

Results

Predicted Results

We predict that the t-SNE model will outperform the other two models on reducing dMRI measures data while preserving the maximum amount of information contained in the original data. This prediction arises from the fact that t-SNE has been shown to have great potential in enhancing quality control for large neuroimaging data sets and identifying novel patterns in the data sets (Panta et al., 2016). However, after data has been pruned to remove outliers, we expect the PCA and t-SNE methods to have similar performance on dimension reduction.

For the t-SNE methods, we will also investigate the effects of different hyperparameters to enhance the accuracy of the neurological information that might be reduced by the dimension reduction method. As we know that the tunable parameter, ‘perplexity,’ balances attention between local and global aspects of the data, we will attempt to analyze and visualize the reduced data with different perplexities. We predict that the most suitable perplexity chosen for the number of influential neighbors for each data is between 5 to 20. The appropriate hyperparameters will make the model robust. In this way, we will be able to identify meaningful information from big data sets rapidly.

Using Seaborn, a Python data visualization library based on the Matplotlib Python library, we will also visualize a hierarchically clustered heatmap that plots the closeness of eighteen white matter bundles with the cluster map function. We expect that white matter pathways that are bilateral in the two hemispheres will have the closest relation.

Open Questions

While the current data set allows researchers to extract significant components with the dimension reductions methods, the interpretation of the resulting components can become more challenging as datasets grow larger. In addition, the confound variables, such as age and gender, has not been fully considered and excluded in the performance of dimension reduction algorithms.

Discussion

While the dimension reduction approach is general and can be implemented in various fields, the proposed three methods enable easy access to pre-process the large neuroimaging data sets for either prospective data currently being captured or retrospective data uploaded after the fact for comparison. We will take advantage of the dimension reduction methods to advance the application of diffusion models in health and disease. Resulting from this investigation, the framework of reduced dimension models could be implemented in relative data processing and visualization library in Python to facilitate the feature extraction process. The proposed framework of the appropriate dimension reduction method for dMRI data may be used further to examine brain microstructure and other related research questions. It is evident that with the ever-growing acquisition of a large number of subjects, feature extraction techniques become fundamental tools for processing multi-dimensional brain imaging datasets (e.g., the Human Connectome Project with >1000 young adults scans; Van Essen et al. (2013)). The approach of dimension reduction for brain data, combined with multiple current ongoing pieces of research about building software modules, suggests a promising future for the visualization of patterns of the large-scale dMRI data. Thus, taken together, the findings from the proposed research have broad and essential theoretical and real-world implications.

Reference

Basser, P.J., Mattiello, J., LeBihan, D., 1994. Mr diffusion tensor spectroscopy and imaging. *Biophys. J.* 66, 259–267.

Beckmann, C.F., Mackay, C.E., Filippini, N., Smith, S.M., 2009. Group comparison of resting-state fmri data using multi-subject ica and dual regression. *Neuroimage* 47, S148.

Berman, J., Lanza, M., Blaskey, L., Edgar, J., & Roberts, T. (2013). High Angular Resolution Diffusion Imaging Probabilistic Tractography of the Auditory Radiation. *American Journal Of Neuroradiology*, 34(8), 1573-1578. doi: 10.3174/ajnr.a3471

Chamberland, M., Raven, E., Genc, S., Duffy, K., Descoteaux, M., & Parker, G. et al. (2019). Dimensionality reduction of diffusion MRI measures for improved tractometry of the human brain. *Neuroimage*, 200, 89-100. doi: 10.1016/j.neuroimage.2019.06.020

Cousineau, M., Jodoin, P.M., Garyfallidis, E., Cote, M.A., Morency, F.C., Rozanski, V., GrandMaison, M., Bedell, B.J., Descoteaux, M., 2017. A test-retest study on Parkinson's ppmi dataset yields statistically significant white matter fascicles. *Neuroimage: Clin.* 16, 222–233.

Descoteaux, M., Deriche, R., Le Bihan, D., Mangin, J.F., Poupon, C., 2011. Multiple q-shell diffusion propagator imaging. *Med. Image Anal.* 15, 603–621.

Garyfallidis, E., Brett, M., Amirbekian, B., Rokem, A., van der Walt, S., Descoteaux, M., & Nimmo-Smith, I. (2014). Dipy, a library for the analysis of diffusion MRI data. *Frontiers In Neuroinformatics*, 8. doi: 10.3389/fninf.2014.00008

Garyfallidis, E., Co[^]te, M.A., Rheault, F., Sidhu, J., Hau, J., Petit, L., Fortin, D., Cunanne, S., Descoteaux, M., 2018. Recognition of white matter bundles using local and global

streamline-based registration and clustering. *Neuroimage* 170, 283–295.

Jeurissen, B., Leemans, A., Tournier, J.D., Jones, D.K., Sijbers, J., 2013. Investigating the prevalence of complex fiber configurations in white matter tissue with diffusion magnetic resonance imaging. *Hum. Brain Mapp.* 34, 2747–2766.

Jones, D.K., Catani, M., Pierpaoli, C., Reeves, S.J., Shergill, S.S., O’sullivan, M., Golesworthy, P., McGuire, P., Horsfield, M.A., Simmons, A., et al., 2006. Age effects on diffusion tensor magnetic resonance imaging tractography measures of frontal cortex connections in schizophrenia. *Hum. Brain Mapp.* 27, 230–238.

Jones, D.K., Catani, M., Pierpaoli, C., Reeves, S.J., Shergill, S.S., O’sullivan, M., Maguire, P., Horsfield, M.A., Simmons, A., Williams, S.C., et al., 2005a. A diffusion tensor magnetic resonance imaging study of frontal cortex connections in very-late- onset schizophrenia-like psychosis. *Am. J. Geriatr. Psychiatry* 13, 1092–1099.

Minett, T. (2011). P1-245 Cambridge centre for ageing and neuroscience (CamCan) study protocol. *Journal Of Epidemiology & Community Health*, 65(Suppl 1), A134-A134. doi: 10.1136/jech.2011.142976e.37

Mori, S., Van Zijl, P.C., 2002. Fiber tracking: principles and strategies—a technical review. *NMR Biomed.* 15, 468–480.

Mwangi, B., Tian, T., & Soares, J. (2013). A Review of Feature Reduction Techniques in Neuroimaging. *Neuroinformatics*, 12(2), 229-244. doi: 10.1007/s12021-013-9204-3

Panta, S., Wang, R., Fries, J., Kalyanam, R., Speer, N., & Banich, M. et al. (2016). A Tool for Interactive Data Visualization: Application to Over 10,000 Brain Imaging and Phantom MRI Data Sets. *Frontiers In Neuroinformatics*, 10. doi: 10.3389/fninf.2016.00009

Pierpaoli, C., Basser, P.J., 1996. Toward a quantitative assessment of diffusion anisotropy. *Magn. Reson. Med.* 36, 893–906.

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.M.H., et al., 2013. The Wu-minn human connectome project: an overview. *Neuroimage* 80, 62–79.

Wandell, B. (2016). Clarifying Human White Matter. *Annual Review Of Neuroscience*, 39(1), 103-128. doi: 10.1146/annurev-neuro-070815-013815

Yeatman, J., Dougherty, R., Rykhlevskaia, E., Sherbondy, A., Deutsch, G., Wandell, B., & Ben-Shachar, M. (2011). Anatomical Properties of the Arcuate Fasciculus Predict Phonological and Reading Skills in Children. *Journal Of Cognitive Neuroscience*, 23(11), 3304-3317. doi: 10.1162/jocn_a_00061