

MODELING SUBJECT VARIABILITY IN LARGE SCALE FUNCTIONAL MAGNETIC RESONANCE IMAGING DATA

Christoffer Hjort¹, Alessandro Montemurro¹, Edoardo Nemni¹

¹Technical University of Denmark, DTU Compute

ABSTRACT

The aim of the paper is predicting behavioral traits using Functional Magnetic Resonance Imaging brain scan connectome. In particular, a multi-way array decomposition is involved and the obtained matrices are used for classification and regression tasks. The specific decomposition method used in this study is a constrained version of PARAFAC2 model. The classification task focuses on gender and age, the regression uses the subject-specific scores obtained by PARAFAC2 in order to predict personality. The scores obtained by the variant of the PARAFAC2 model can be actually used to predict behavioral traits from the HCP data. Using subject-specific scores from PARAFAC2 model cannot guarantee a better result compared to an easier approach, e.g., making prediction using the correlogram.

Index Terms— fMRI, behavioral traits, Multi-way unsupervised decomposition, PARAFAC2.

1. INTRODUCTION

The main objective of this study is related functional connectomes to behavioral traits using the subject-specific scores obtained from a variant of the constrained PARAFAC2 model. The data used are taken from the **Human Connectome Project** (HCP), which is a project to construct a map of the complete structural and functional neural connections in vivo within and across individuals. The HCP represents the first large-scale attempt to collect and share data of a scope and detail sufficient to begin the process of addressing deeply fundamental questions about human connective anatomy and variation. The goal after fitting the variant of the PARAFAC2 model is studying if the extracted matrices can predict behavioral traits from the HCP in order to define how brain activity relates to behavioral data. In particular, whether any specific patterns of brain connectivity are associated with personality defined using the **Five Factor Model (NEO-FFI)** by Costa & McCrae [1]. The Five Factor Model examines a person's Big Five personality traits such as: **openness to experience**, **conscientiousness**, **extroversion**, **agreeableness**, and **neuroticism** capturing the major facets of human personality across cultures. There are different versions of personality

inventories and the one used in this paper, NEO-FFI, uses only 60 items (question), 12 per domain.

Prior to the analysis, the HCP data were pre-processed with Independent Component Analysis (ICA) in order to get a parcellation of the brain scans into brain regions, each represented by one independent component.

2. METHODS

Consider an fMRI scan $\mathbf{X}^{(s)} \in \mathbb{R}^{V \times T}$, where V is the number of brain voxels and T is the number of time-samples for subject s . In particular, V is the number of independent component used in the data pre-processing to obtain a brain parcellation; each independent component represents a brain voxel. The $V \times V$ voxel-wise covariance matrix can be used to model the functional connectivity in the brain for subject s ; if the voxel means are subtracted, then the covariance matrix is equivalent to the correlogram $\mathbf{X}^{(s)}\mathbf{X}^{(s)\top}$ [2].

Alternative representations of this connectivity involve the use of low-rank decomposition. These models are capable of extracting relevant features, lower the dimension of the data and improve the interpretability of this smaller subset of features. The model used in this paper is PARAFAC2 model [3]. PARAFAC2 aims to extend the idea of Principal Component Analysis (PCA), where the factor scores matrices are not unique. The problem of choosing a particular orientations in PCA can be avoided using the principle of PARALLEL FACTORS: multiple two-ways arrays, also referred as to *slices* of a three-ways array are simultaneously fitted in terms of a common set of factors with different weights for each slice. Under mild conditions, the model is fully identified, i.e. the uniqueness of the factors is ensured [4].

The general form of PARAFAC2 model with K components is given by

$$\mathbf{X}^{(s)} = \mathbf{A}\mathbf{C}^{(s)}\mathbf{R}\mathbf{P}^{(s)} + \mathbf{E}^{(s)}, \quad s = 1, \dots, S \quad (1)$$

$$\text{subject to } \mathbf{P}^{(s)}\mathbf{P}^{(s)\top} = \mathbf{I} \quad s = 1, \dots, S$$

where $\mathbf{C}^{(s)} \in \mathbb{R}^{K \times K}$ is a diagonal matrix containing subject specific scores for each voxel, $\mathbf{P}^{(s)} \in \mathbb{R}^{K \times T}$ is the ma-

trix containing the T time sampling of the PARAFAC2 components and $\mathbf{E}^{(s)}$ is an error term, assumed to be zero-mean Gaussian and isotropic covariance matrix $\mathbf{Q}^{(s)}\mathbf{Q}^{(s)\top}$, where $\mathbf{Q}^{(s)} = \sigma_s \mathbf{I}$.

Without loss of generality, \mathbf{A} is assumed to be the identity matrix, meaning that number of PARAFAC2 components equals the number of brain voxels.

The constraint in Eq. (1) specifies that the Gramian matrix of $\mathbf{P}^{(s)}$ is constant over all the subjects. In this paper, a modified version of PARAFAC2 model is proposed. To understand the reason why the matrix $\mathbf{P}^{(s)}$ is constrained to be orthogonal, consider the correlogram $\mathbf{X}^{(s)}\mathbf{X}^{(s)\top}$ using the decomposition described in Eq. (1).

$$\begin{aligned}\mathbf{X}^{(s)}\mathbf{X}^{(s)\top} &= \mathbf{C}^{(s)}\mathbf{R}\mathbf{P}^{(s)}(\mathbf{C}^{(s)}\mathbf{R}\mathbf{P}^{(s)})^\top \\ &= \mathbf{C}^{(s)}\mathbf{R}\mathbf{P}^{(s)}\mathbf{P}^{(s)\top}\mathbf{R}^\top\mathbf{C}^{(s)\top} \\ &= \mathbf{C}^{(s)}\mathbf{R}(\mathbf{I})\mathbf{R}^\top\mathbf{C}^{(s)\top} \\ &= \mathbf{C}^{(s)}\mathbf{R}\mathbf{R}^\top\mathbf{C}^{(s)\top}\end{aligned}$$

where we used the property that $\mathbf{P}^{(s)}$ is an orthogonal matrix. In the last equality we see that the correlogram can be written as a product of matrices where we have a *core* matrix, $\mathbf{R}\mathbf{R}^\top$, which is not subject-dependent. This ensures a consistent representation of the correlation between voxels across all the subjects.

We propose an extended version on the original PARAFAC2 model, such that each subject has its own voxel specific noise $\mathbf{Q}^{(s)} \in \mathbb{R}^{V \times V}$. Hence, for each subject, the covariance matrix of the noise is diagonal and the i^{th} diagonal element represents the standard deviation of the noise of the i^{th} brain voxel. We will assume that the noise covariance matrix is $\mathbf{Q}^{(s)}\mathbf{Q}^{(s)\top}$, where

$$\mathbf{Q}^{(s)} = \text{diag} \left[\frac{1}{\sigma_{1,s}}, \frac{1}{\sigma_{2,s}}, \dots, \frac{1}{\sigma_{V,s}} \right].$$

The i^{th} diagonal element $\sigma_{i,s}^{-1}$ represents the standard deviation of the error of region i for subject s .

Once the neuro-image $\mathbf{X}^{(s)}$ is decomposed and the matrices of PARAFAC2 are estimated, regularized regression and regularized classification is carried out, using those matrices as predictors. The results will be compared with the ones obtained when the correlogram is used as predictor.

2.1. Training PARAFAC2 model

In order to obtain the estimates of the matrices $\mathbf{C}^{(s)}$, \mathbf{R} , $\mathbf{P}^{(s)}$ and $\mathbf{Q}^{(s)}$ we minimized a specific loss function using an Alternating Least Squares (ALS) approach. ALS It is a multi-step iterative optimization process. In every iteration

it solves for one matrix keeping the other fixed. Alternating between the steps guarantees reduction of the cost function, until convergence.

No-noise estimation. A first approach consists of ignoring the error term. The cost function used is the Frobenius norm of the difference between the image and its reconstruction:

$$\begin{aligned}\mathcal{L} &= \sum_{s=1}^S \|\mathbf{X}^{(s)} - \mathbf{C}^{(s)}\mathbf{R}\mathbf{P}^{(s)}\|_F^2 \\ &= \sum_{s=1}^S \text{tr} \left[(\mathbf{X}^{(s)} - \mathbf{C}^{(s)}\mathbf{R}\mathbf{P}^{(s)})(\mathbf{X}^{(s)} - \mathbf{C}^{(s)}\mathbf{R}\mathbf{P}^{(s)})^\top \right].\end{aligned}$$

The following update rules for the matrices are obtained:

$$(1a) \quad \mathbf{P}^{(s)_{new}} = \mathbf{V}^{(s)}\mathbf{U}^{(s)\top} = (\mathbf{U}^{(s)}\mathbf{V}^{(s)\top})^\top$$

$$(2a) \quad [\mathbf{C}^{(s)}]_{ii}^{new} = \frac{[\mathbf{R}\mathbf{P}^{(s)}\mathbf{X}^{(s)\top}]_{ii}}{[\mathbf{R}\mathbf{R}^\top]_{ii}}$$

$$(3a) \quad \mathbf{R}^{new} = \left[\sum_s \mathbf{C}^{(s)\top}\mathbf{X}^{(s)}\mathbf{P}^{(s)\top} \right] \times \left[\sum_s \mathbf{C}^{(s)\top}\mathbf{C}^{(s)} \right]^{-1}$$

In update (1a), the matrices $\mathbf{U}^{(s)}$ and $\mathbf{V}^{(s)}$ come from the singular value decomposition

$$\mathbf{U}^{(s)}\Sigma^{(s)}\mathbf{V}^{(s)\top} = \text{svd} \left(\mathbf{X}^{(s)\top}\mathbf{C}^{(s)}\mathbf{R} \right).$$

The orthogonality of $\mathbf{P}^{(s)}$ is ensured by the fact that $\mathbf{U}^{(s)}$ and $\mathbf{V}^{(s)}$ are orthogonal (by definition) and a product of orthogonal matrices is still an orthogonal matrix.

Notice how update (2a) refers only to the diagonal element of the matrix $\mathbf{C}^{(s)}$. Moreover, the summation over s in the update (3a) shows how \mathbf{R} can take into account all the subjects.

Noise estimation. A likelihood approach is used when the noise is to be estimated together with the other matrices. Due to the monotonic behavior of the logarithmic function, the log-likelihood function is used:

$$\begin{aligned}\log \mathcal{L} &= \sum_{s=1}^S \sum_{v=1}^V \sum_{t=1}^T \left[-\frac{1}{2} \log(2\pi\sigma_{v,s}^2) + \right. \\ &\quad \left. - \frac{1}{2\sigma_{v,s}^2} \left(\mathbf{X}^{(s)}(v, t) - (\mathbf{C}^{(s)}\mathbf{R}\mathbf{P}^{(s)})(v, t) \right)^2 \right].\end{aligned}$$

The following updating rules are obtained using ALS:

$$(1b) \quad \mathbf{Q}^{(s)new} = \left[\frac{1}{T} \sum_{t=0}^T \left(\mathbf{X}^{(s)}(v) - (\mathbf{C}^{(s)} \mathbf{R} \mathbf{P}^{(s)})(v) \right)^2 \right]^{-}$$

$$(2b) \quad \mathbf{P}^{(s)new} = \mathbf{V}^{(s)} \mathbf{U}^{(s)\top} = (\mathbf{U}^{(s)} \mathbf{V}^{(s)\top})^\top$$

$$(3b) \quad [\mathbf{C}^{(s)}]_{ii}^{new} = \frac{[\mathbf{R} \mathbf{P}^{(s)} \mathbf{X}^{(s)\top} \mathbf{Q}^{(s)} \mathbf{Q}^{(s)\top}]_{ii}}{[\mathbf{Q}^{(s)} \mathbf{Q}^{(s)\top} \mathbf{R} \mathbf{R}^\top]_{ii}}$$

$$(4b) \quad \mathbf{R}^{new} = \left[\sum_s \mathbf{C}^{(s)\top} \mathbf{Q}^{(s)} \mathbf{Q}^{(s)\top} \mathbf{X}^{(s)} \mathbf{P}^{(s)\top} \right] \times \left[\sum_s \mathbf{C}^{(s)\top} \mathbf{Q}^{(s)} \mathbf{Q}^{(s)\top} \mathbf{C}^{(s)} \right]^{-1}.$$

In update (2b), the matrices $\mathbf{U}^{(s)}$ and $\mathbf{V}^{(s)}$ come from the singular value decomposition

$$\mathbf{U}^{(s)} \Sigma^{(s)} \mathbf{V}^{(s)\top} = \text{svd} \left(\mathbf{X}^{(s)\top} \mathbf{Q}^{(s)} \mathbf{Q}^{(s)\top} \mathbf{C}^{(s)} \mathbf{R} \right).$$

Notice how the update for the matrix $\mathbf{Q}^{(s)}$ is given by a time average of the error (image minus its reconstruction). Moreover, all the other updates are a weighed version of the "without noise" estimates, weighed by the covariance matrix of the error, $\mathbf{Q}^{(s)} \mathbf{Q}^{(s)\top}$.

3. EXPERIMENTS

3.1. Real fMRI data

The large scale fMRI data is available for parcellations of 15, 25, 50, or 100 brain regions per subject. In this project, only the data of 15 brain regions were used due to memory limitations and computational limitations. Each brain region is sampled over 4800 time points, which gives $\mathbf{X} \in \mathbb{R}^{15 \times 4800 \times 1003}$. Clearly the huge dimensions of the real data is the motivation for applying tensor decomposition models, in an attempt to extract only relevant features for predicting behavioural traits in subjects. The data will be modeled both as standardized and non-standardized, where the standardized data has been scaled over the brain region dimension. This results in a factor of two for the number of experiments.

3.2. Baseline models

The features of the covariance baseline are given by the upper triangle and diagonal of the covariance matrix between the 15 brain regions of a subject. I.e. for each subject their covariance matrix is computed and the upper triangle and diagonal is extracted. This results in 120 features for each subject. Covariance of the 15 brain regions for each subject, i.e. the upper

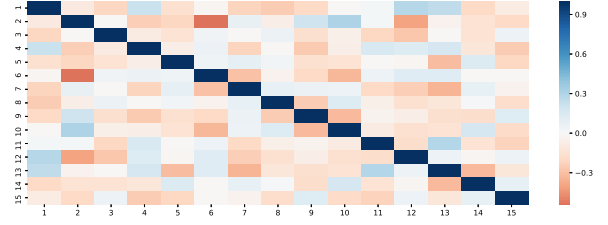


Fig. 1. Correlogram of the 15 brain regions for a random subject.

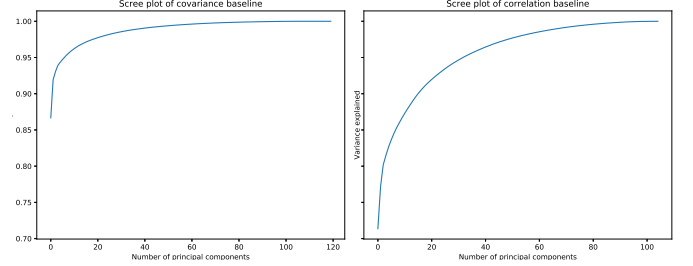


Fig. 2. Variance explained by principal components of covariance features (left), and variance explained by principal components of correlation features (right).

triangle and diagonal of the covariance matrix. This gives 120 features for each subject.

The correlation baseline is given as the correlation between the 15 brain regions of a subject. For each subject their correlation matrix is computed and only the upper triangle is extracted, which results in 105 features for each subject. Figure 1 shows the correlation matrix of a single randomly chosen subject. There is a great amount of brain regions with significant correlation, which illustrates what regions interacts with each other during the resting state of the subject. The rationale for choosing these baselines is whether this interactivity between brain regions is a better predictor than the scores obtained by the PARAFAC2 model.

Furthermore, the principal components of the covariance and correlation matrices are extracted. As shown in Figure 2, the first principal component of the covariance features explains 87% of the variance and the first principal component of the correlation features explains 71% of the variance, however, it takes 39 and 67 components to explain 99% of the variance for the covariance and correlation features respectively, so the incentive to take only a subset of the components is not major.

3.3. PARAFAC2 models

Compared to the baselines, PARAFAC2 should be able to extract more specific and directly relevant features. Both of the

PARAFAC2 models were trained using the ALS approach, and convergence was reached after approximately 6 epochs, however, the models were trained for 15 epochs each for a tiny improvement.

PARAFAC2 with no noise estimation decomposes each $\mathbf{X}^{(s)}$ into the matrices $\mathbf{C}^{(s)}\mathbf{R}\mathbf{P}^{(s)}$. The diagonal of $\mathbf{C}^{(s)}$ contains scores for each of the 15 brain regions of subject s , and these scores are used as features, which gives 15 features for each subject.

PARAFAC2 with noise estimation decomposes each $\mathbf{X}^{(s)}$ into the matrices $\mathbf{C}^{(s)}\mathbf{R}\mathbf{P}^{(s)} + \mathbf{Q}^{(s)}$. Again, the diagonal of $\mathbf{C}^{(s)}$ contains scores for each brain region of subject s , while $\mathbf{Q}^{(s)}$ contains estimates of region and subject specific noise. Both $\mathbf{C}^{(s)}$ and $\mathbf{Q}^{(s)}$ are used as features, and each gives 15 features for each subject.

3.4. Predicting behavioural traits

The specific choice of behavioural traits is governed by the motivation of predicting the personality, gender, and age of subjects given the models' extracted features of resting state fMRI brain scans. The personality traits to be predicted are given by *Agreeableness*, *Openness*, *Conscientiousness*, *Neuroticism*, and *Extroversion*. The five personality traits are quantified by a continuous range of numbers, making this a regression problem. *Gender* is a classification problem, and so is *age*, as the age of subjects have been divided into four intervals of 22-25, 26-30, 31-35, 36+.

For classification a simple Logistic Ridge Regression model is used, and for regression a Linear Ridge Regression model. The regularization parameter for the L_2 -norm and the generalization performance is estimated with two-layer 10-fold cross-validation.

4. RESULTS

First, it's important to note that the majority class of *gender* is 53%, and for *age* it is 34%. For *gender* the Logistic Regression with L_2 penalty, achieved 82% accuracy for the standardized data on the PCA projection of features extracted by the covariance baseline, where as it only achieved 70% accuracy on the scores extracted by the PARAFAC2 models with and without noise estimation.

For *age*, an accuracy of 48% was achieved on non-standardized data by using the PCA projection of the features extracted by the covariance baseline. Using the scores extracted by the PARAFAC2 model without noise achieved 45% accuracy.

On the standardized data, predicting only the mean value on all five regression problems for *Agreeableness*, *Openness*,

Conscientiousness, *Neuroticism*, and *Extroversion* respectively, resulted in a mean squared error of 0.99 or 1. As the best result for any of the personality traits, the Ridge Regression achieved a MSE of 0.94 on the correlation features for predicting *openness*, where the scores by the PARAFAC2 models resulted in a MSE of 1. On all other personality traits, neither the baselines or the PARAFAC2 models were able to extract features that gave the Ridge Regression model a performance better than 1 MSE.

5. DISCUSSION

From the performance on the personality traits, it is clear that neither the baselines or the PARAFAC2 models were better than simply predicting the mean value. For gender and age, all models were better than predicting the majority class, but the PARAFAC2 models did not outperform the baselines.

5.1. Limitations of the model

The model proposed in the paper relies on some assumptions and it's subject to limitations. All the experiments have been carried out using a resolution of 15, i.e. the brain is partitioned into 15 regions. In order to get a deeper understanding of the connectivity between brain region and get a more realistic model, a *finer* partition of the brain could be used. For a general number V of voxels, obtained considering V independent components, we would get a $V \times V$ diagonal matrix $\mathbf{C}^{(s)}$ containing scores for all the regions. Hence, by increasing the number of regions, we could be able to capture more interactions between brain voxels, and perhaps the PARAFAC2 model would extract more relevant features as the data would contain larger amounts of information about the brain.

Throughout the study, we focused on the unsupervised decomposition of the brain scan $\mathbf{X}^{(s)}$ to get specific scores for the regions. We used those scores for regression for behavioral traits. We used *linear* regression, ignoring that the behavioral traits may be described better by *non-linear* interactions of the different voxels. The use of non-linear models might allow us to get a better fit of the data and a lower error.

6. CONCLUSION

The scores obtained by the variant of the PARAFAC2 model can be used to predict behavioral traits from the HCP data. The classification based on the PARAFAC2 scores performs better than a random prediction and than the majority class classification, but worse than the baseline. Regression using the PARAFAC2 scores is not better than predicting using the mean. Using the variant of the PARAFAC2 model cannot guarantee a better result compared to an easy approach as the covariance of brain regions.

References

- [1] Robert R McCrae and Paul T Costa Jr. “A contemplated revision of the NEO Five-Factor Inventory”. In: *Personality and individual differences* 36.3 (2004), pp. 587–596.
- [2] Kristoffer H Madsen, Nathan W Churchill, and Morten Mørup. “Quantifying functional connectivity in multi-subject fMRI data using component models”. In: *Human brain mapping* 38.2 (2017), pp. 882–899.
- [3] Henk AL Kiers, Jos MF Ten Berge, and Rasmus Bro. “PARAFAC2-Part I. A direct fitting algorithm for the PARAFAC2 model”. In: *Journal of Chemometrics* 13.3-4 (1999), pp. 275–294.
- [4] Richard A Harshman and Margaret E Lundy. “PARAFAC: Parallel factor analysis”. In: *Computational Statistics & Data Analysis* 18.1 (1994), pp. 39–72.