Identification of Autism Spectrum Disorders on Brain Structural MRI with Variational Autoencoders

Mélanie Garcia
Institut Hypercube
51, esplanade du Général de Gaulle, Paris La Défense, France
melanie.garcia@institut-hypercube.org

Annabelle Blangero
Institut Hypercube
51, esplanade du Général de Gaulle, Paris La Défense, France
ablangero@octo.com

Jean-Marc Orgogozo
Institut Hypercube
51, esplanade du Général de Gaulle, Paris La Défense, France
jean-marc.orgogozo@institut-hypercube.org

Abstract

Autism Spectrum Disorders (ASD) are diverse, which makes diagnostic and therapeutic research complicated in the absence of a more detailed categorization. Nowadays, the evaluation of a patient is based on behavior observation, making the diagnosis long and complex, with the necessity of different experts' involvement. The goal of this study is to contribute to the research on objective metrics, quickly quantifiable, for the evaluation of autism. Despite the acknowledgment that ASD is a neurological disorder, there is still no consensus on the underlying affected neural substrate. Thus, we are working on brain imaging variables from the Autism Brain Imaging Data Exchange 1, focusing on structural MRI. Neuroimaging tools enable to compute many relevant characteristics like the volume of segmented brain areas. However, it takes a lot of time to compute all the parameters separately, and then to analyze it with a multivariate approach (a few hours per image).

Our proposal is to build and implement an autoencoder architecture which generates latent features of interest, overcoming the limitations of the « classical » approach: reducing the cost of time during the processing (a few seconds per image) while improving classification performance. Among unsupervised neural networks, we have been interested in Variational Autoencoders (VAE), which has recently shown great abilities in data compression by extracting relevant information from images or text. Inspired from Bayesian inference, VAE interests in the input data distribution $p_\theta(x)$. Guessing it is Gaussian, VAE estimates latent vector z mean and variance with the encoder ($q_\psi(z|x)$) being an estimator of the true posteriori distribution ($p_\theta(z|x)$). This approach makes the model more robust, giving a mathematical sense to the latent vector. Thus, we have studied if VAE can create pertinent features in the ASD context.

On the one hand, we will introduce our benchmark model: 1/ selecting an optimal set of « classical » neuroimaging variables for the classification autist/non autist with Mutual Information and Recursive Feature Elimination in cross-validation; 2/ applying a logistic regression on the selected features to get a prediction score, with metrics Receiver Operating Characteristic area under curve (ROC) whose

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

significance was evaluated with permutation tests. Best features corresponded to the ASD brain morphological particularities found in publications. We obtained a ROC AUC score of 0.67 (p-value = 0.01).
On the other hand, we will introduce our model based on VAE: 1/ encoding MRI data with VAE; 2/ feature selection and binary classification similar to the benchmark model. We obtained a ROC AUC score of 0.65 (p-value = 0.01).
Finally, we compared both sets of features to interpret VAE encoded features in terms of neurological markers.
Our study shows the pertinence of looking for a different kind of features from brain imaging data in order to reach faster and better performances in Autism diagnostic.