2017 International Symposium on Nonlinear Theory and Its Applications,
NOLTA2017, Cancun, Mexico, December 4-7, 2017

Deep Neural Generative Model for fMRI Image Based Diagnosis of Mental Disorder

Tashiro Tetsuo†, Matsubara Takashi† and Uehara Kuniaki†

†Graduate School of System Informatics, Kobe University, 1-1 Rokko-dai, Nada, Kobe, Hyogo 657-8501, Japan
Email: tashiro@ai.cs.kobe-u.ac.jp, matsubara@phoenix.kobe-u.ac.jp, and uehara@kobe-u.ac.jp

Abstract—Diagnosis of mental disorders based on fMRI brain image analysis often has two steps: unsupervised feature extraction and supervised classification. This is expected to prevent overfitting due to the typically small size of medical fMRI datasets. However, the unsupervised feature extraction process has a risk of extracting individual variability (such as brain shape) as features instead of disease-related brain activity. In this study, we propose a fMRI brain image analysis method based on conditional variational auto-encoder (CVAE), which is a deep learning model extracting features with given label information. The CVAE can classify fMRI images without another feature extraction process, suppresses overfitting, and achieves better diagnosis accuracy.

1. Introduction

Recently, the number of subjects who suffer from mental disorders is ever-increasing. Early diagnosis and early treatment of mental disorders are important, but mental disorders in their early phases induce no physical anomaly in a brain. Thus, doctors diagnose mental disorders by interview. To improve the accuracy of diagnosis, analytical diagnosis methods are considered to be promising. Recently, fMRI brain image analysis is used as an analytical diagnosis method. Mental disorders often have an abnormal activity or hypofunction of the brain. Thus, machine learning has been attempted as a method for finding anomaly activity of the brain. What matters here is the size of the dataset. The size of the medical dataset is often very small due to the difficulty in publication owing to personal information protection and the cost required for obtaining fMRI dataset. Therefore, a supervised classification such as deep learning, which requires big data, has been considered not to be suited for diagnosis. Instead, feature extraction is performed first by unsupervised learning that can be performed on even small-sized data, and then, diagnosis is performed with supervised learning manner [Suk 15]. However, this process has a risk of not obtaining features required for diagnosing because individual variabilities such as brain shape, which is more significant than disease-related brain activity. Therefore, to improve diagnosis accuracy of diseases by using brain image analysis, obtaining disease-related brain activity is important.

In this study, we propose a deep neural generative model for diagnosing of mental disorders based on fMRI brain images [Kingma 14]. In general, with a small dataset, a generative model achieves better performance than a discriminative model [Raina 03]. The proposed model uses a conditional variational auto-encoder, which is given a pair of a fMRI brain image and an assumed label (healthy or not) and reconstructs the original image. The reconstruction error can be considered as the posterior likelihood of the assumption. The proposed model estimates the condition of the given fMRI brain image based on the likelihoods. The experimental results demonstrate that the proposed model achieves more accurate diagnosis than baseline methods: support vector machine with feature extraction and multilayer perceptron.

2. Dataset and preprocessing

2.1. Dataset

In this study, we used a dataset of rs-fMRI images obtained from schizophrenia patients published on `https://openfmri.org/dataset/ds000030/`. Features of schizophrenia have been studied for decades. Although there is a change in the frontal lobe part, there is no clear difference that can be diagnosed by looking at the brain image and there is a difference as a function deterioration of the brain. The dataset consist of 52 patients with schizophrenia and 122 normal control subjects. time repetition = 3000 ms, acquisition matrix size = 64×64×34, 152 scans, and a voxel thickness = 3.0 mm.

2.2. Preprocessing

We performed the routine preprocessing procedure for rs-fMRI using the SPM12 software package [http://www.fil.ion.ucl.ac.uk/spm/software/spm12/]. First, we discarded the first 10 scans of each subject to ensure magnetization equilibrium. Second, we performed time slice adjustment with SPM12. We used time slice order option of {1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 2 4 6 8 10 1 2 14 16 18 20 22 24 26 28 30 32 34}. Next, we performed realignment in order to suppress the displacement of the position of the brain due to the movement of the subject. In realignment, rigid body return was performed so that the position is aligned with the first scan. Then, we normalized to suppress

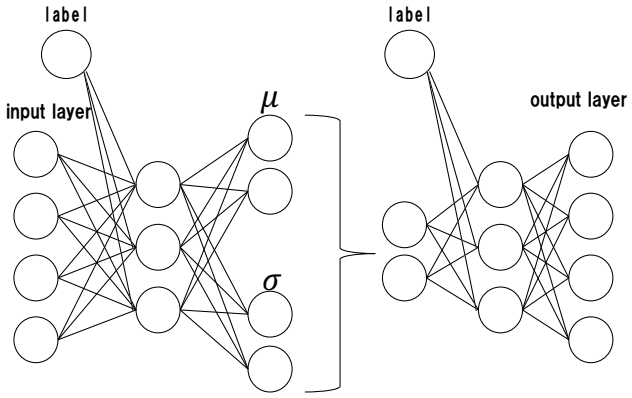

Fig. 1: Architecture of Conditional Variational Auto-encoder.

individual differences such as brain shape. Specifically, we normalized rs-fMRI images to the MNI space with a voxel size of $3 \times 3 \times 3$ mm. The normalized fMRI images were parcellated into 116 Regions-Of-Interest (ROIs) using Automated Anatomical Labeling (AAL) template [Tzourio-Mazoyer 02]. In addition, the averages of voxels were calculated for each ROI region. As a result, a fMRI image became a 116-dimensional vector. Since the 116-dimensional vector is the average of voxels in the brain, the change in the 116 dimensional-vector correlates to a functional change in the brain. Finally, numerical values were normalized in the spatial direction to eliminate noise during having fMRI. In addition, it has been shown to be reliable in the frequency range between 0.06 Hz and 0.025 Hz. Therefore, we bandpass-filtered the 116-dimensional data with a frequency band. This preprocessing follows the same procedure as Suk et al.'s work [Suk 15].

3. Proposed method

3.1. Conditional Variational Auto-encoder

Conditional variational auto-encoder (CVAE) is a model that extends VAE and learns using labeled data [Kingma 14]. CVAE can separate label information at the time of feature extraction. In generally, CVAE is used for image generation, semi supervised learning, etc. Fig. 1 illustrates the network architecture of the CVAE. Let y be label data (whether the subject is healthy or have a disorder) and $\boldsymbol{x}$ be the data vector (116-dimensional vector). CVAE learns to minimize simultaneous distribution $\log p(\boldsymbol{x}, y)$. Considering the log likelihood, like VAE, it can be transformed as follows. For detailed VAE formula deformation see the paper [Kingma 14].

$$
\begin{aligned}
&\log p_\theta(\boldsymbol{x}, y) \\
&\quad = \log \int p_\theta(\boldsymbol{x}|y, \boldsymbol{z}) p(\boldsymbol{z}) d\boldsymbol{z} + \log p(y) \\
&\quad = \log \int q_\phi(\boldsymbol{z}|\boldsymbol{x}, y) \frac{p_\theta(\boldsymbol{x}|y, \boldsymbol{z}) p(\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x}, y)} d\boldsymbol{z} + C \\
&\quad (\because C = \log p(y), const) \\
&\quad \geq \int q_\phi(\boldsymbol{z}|\boldsymbol{x}, y) \log \frac{p_\theta(\boldsymbol{x}|y, \boldsymbol{z}) p(\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x}, y)} d\boldsymbol{z} \\
&\quad = \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x}, y)}[\log p_\theta(\boldsymbol{x}|y, \boldsymbol{z})] - D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x}, y) \| p(\boldsymbol{z})) \\
&\quad = -\mathcal{L}(\boldsymbol{x}, y)
\end{aligned} \tag{1}
$$

Let $\mathcal{L}$ be the loss function of CVAE, to be minimized through learning. The feature representations are extracted as latent variables in the hidden layer. However, by adding the label y to the encoder and decoder, feature representations other than labels, that is, personal differences such as the shape and size of the brain are extracted in the hidden layer at the time of input reconstruction. As a result, label information is separated into y given as condition. By this process, the model can summarize the feature representations of the presence or absence of disease into y. In addition, $\mathcal{L}$ is the reconstruction error for the first term and Kullback-Leibler divergence for the second term. To balance these two errors, the coefficients l_b and l_z are multiplied expressed as follows.

$$
\begin{aligned}
\mathcal{L} = -l_b \cdot \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x}, y)}[\log p_\theta(\boldsymbol{x}|y, \boldsymbol{z})] \\
+ l_z \cdot D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x}, y) \| p_\theta(\boldsymbol{z}))
\end{aligned} \tag{2}
$$

The coefficients l_b and l_z are adjusted as a hyperparameters. In this paper, we consider using CVAE as a discriminator. Given the data $\boldsymbol{x}$, the probability that the label is y can be expressed as follows.

$$
\log p_\theta(y = \tilde{y}|\boldsymbol{x}) = \log \frac{p_\theta(\boldsymbol{x}, y = \tilde{y})}{\sum_k p_\theta(\boldsymbol{x}, y = k)} \tag{3}
$$

The denominator of Equation (3) does not change with respect to labels. Therefore, the numerator of Equation (3) is calculated for each label. Now $\log p_\theta(y = \tilde{y}|\boldsymbol{x}) \geq -\mathcal{L}(\boldsymbol{x}, y)$. Therefore, the probability of label $\tilde{y}$ can be approximated by the product $-\mathcal{L}$. The label with lowest loss is our result for classification.

4. Experiment

4.1. Comparative approach

For comparison, we evaluated the following two methods. First, we use multi-layer perceptron (MLP), a feedforward neural network. Compared to this model, we showed that the proposed method could learn with suppressing overfitting. In the second method, we used auto-encoder for feature extraction, and support vector machine (SVM)

Table 1: Accuracy of identification of schizophrenia.

Method	Deep learning paramaters	SVM paramaters	Test accuracy	SEN	SPEC
Auto-encoder + SVM	units:[200,100,20]	cost = 0.1, gamma = 0.1	60.74	62.7%	65.2%
MLP	units:[100,50,1]	—	70.10%	41.9%	98.3%
CVAE	units:[100,100,20], $l_b = 0.1$, $l_z = 0.1$	—	78.60%	67.0%	90.7%

for classification. Compared to this method, we showed that better diagnosis accuracy could be obtained by learning in one step by CVAE than performing diagnosis after first extracting features.

4.2. Parameter settings

Adjustment of hyperparameters is essential for deep learning. The hyperparameters we adjusted were as follows: Number of units, l_b and l_z. We selected the best parameters by a grid search. For verification, we performed leave-2-out-cross validation. For other network architectures, following previous study [Kingma 14], we set the number of hidden layers to 2 for all the networks. because it is to reduce the search scope and to suppress overfitting. we used ReLU as the activation function and Adam [Kingma 14] as the optimization for the learning algorithm. In addition, we performed Layer Normalization on the output of each layer to improve learning accuracy. For the comparison methods, hyperparameters adjustment were performed in the same way.

4.3. Classification Results

Table 1 shows the diagnosis accuracy. As mentioned above, the diagnosis accuracy in this study is the result obtained by leave-2-out-cross validation. As indicated by chapter 2.1, this dataset is unbalanced data. Therefore, we adjusted the unbalance of classes by oversampling. The best hyperparameters found by grid search for each method is shown Table 1. we obtained the best test diagnosis accuracy. In this study, we fixed the number of layers to narrow down the comparison and search range, but in fact, the result of auto-encoder + SVM achieved the worst result among the three methods, which is only about 61%. The accuracy of MLP is about 70%. This result shows that Deep Learning can obtain a certain effect on fMRI data as well. The method we proposed obtained the accuracy of 78%. This result is more accurate than MLP and auto-encoder + SVM which learned in the traditional two step learning. Compared to MLP, we demonstrated it is possible to obtain high diagnosis accuracy with a generative model instead of a discriminatory model. Compared to auto-encoder + SVM, we showed that accuracy improves by making two step learning one step. Consider the parameters of CVAE here. The following Table 2 shows the top four diagnosis accuracy in CVAE. From Table 2, we could see that we obtained high diagnosis accuracy when the ratio of l_b and l_z (parameter of the loss function) to be $1:1$. The difference between auto-encoder and VAE is the term of Kullback-Leibler divergence amount of loss function. When l_z is smaller than l_b, learning is performed as auto-encoder. Since the best diagnosis accuracy was obtained when the ratio of l_b and l_z was 1: 1, we could see that using VAE extracted better features than an auto-encoder.

Table 2: Accuracy of CVAE.

units	l_b	l_z	train	test
100,100,20	0.1	0.1	97.0%	78.6%
100,50,20	0.1	0.1	95.7%	77.2%
100,100,20	0.1	0.01	98.4%	75.2%
100,100,20	0.1	0.001	97.0%	74.7%

4.4. Identification of disease-related regions

Since the proposed method is a generative model learning in one step, the method can identify the parts related to the mental disorders by using the following indicator:

$$\{\mathcal{B}_{\tilde{y}=0}(y=1) - \mathcal{B}_{\tilde{y}=0}(y=1)\} + \{\mathcal{B}_{\tilde{y}=1}(y=0) - \mathcal{B}_{\tilde{y}=1}(y=0)\} \quad (4)$$

Where $\tilde{y}$ represents the true label of the data, and y is the estimated label. $\mathcal{B}$ is the reconstruction error based on $\mathcal{L}$. In particular, $\mathcal{B}$ is given as the following indicator:

$$\mathcal{B}_{\tilde{y}} = \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x},\tilde{y})}[\log p_\theta(x|\tilde{y}, \boldsymbol{z})] \quad (5)$$

Equation (4) is the error of the wrong label minus the error of the correct label. Since the reconstruction error has the same number of dimensions as the input, 116-dimensions, which is the number of regions divided by AAL. Therefore, we could identify the region with the largest value among the results of Equation (4). We could consider that the region to be most effective at discriminating. In this case, we identified the region by using CVAE (units: (100, 100, 20), l_b: 0.1, l_z: 0.1) of the parameter with the highest identification accuracy. The results are as shown in Table 3. The fMRI image of the regions in the brain is shown in Fig. 1 and Fig. 2

According to Fig. 1-3, we found that the regions related to schizophrenia mainly exists around the cerebellum and the frontal lobe.

Table 3: Significant regions for diagnosis of schizophrenia

rank	name	plot colar
1	Vermis 10	red
2	Temporal Pole Mid L	magenta
3	Amygdala R	blue
4	Amygdala L	blue
5	Frontal Mid R	cyan
6	Paracentral Lobule L	green
7	Vermis 9	red
8	Cerebelum 3 L	yellow
9	Vermis 6	red
10	Frontal Inf Orb L	cyan

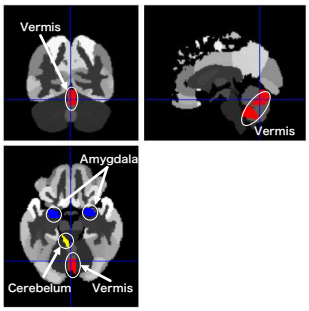

Fig. 1: Vermis , Amygdala and celeblum L.

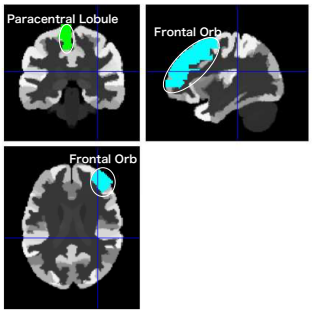

Fig. 2: Frontal Mid R and Paracentral L Lobule.

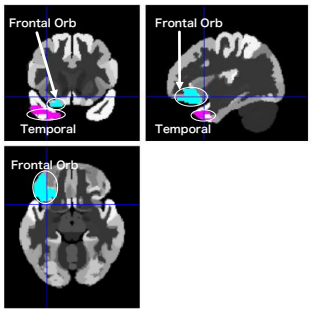

Fig. 3: Frontal inf Orb L and Temporal Pole Mid L.

The cerebellum is said to control the integration of perceptual information, emotion, and movement control [Mitchell 08]. Frontal lobes are said to control emotion [Michael 12]. This is consistent with the regions that had been suggested as possibly related to schizophrenia. Especially, anomalous activity in the frontal lobes is greatly related to schizophrenia. This result proved that the model obtained by this method extracts feature related to schizophrenia. In addition, identification of sites important to diagnosing the disease is possible at the same time by this method.

4.5. Conclusion

In this study, we proposed a method using deep generative neural model. we used CVAE for diagnosis of patients with schizophrenia and obtained higher diagnosis accuracy than MLP, auto-encoder+SVM. By using the generative model, we showed that overfitting could be prevented even in one-step of learning method. According to this result, we verified that the proposed method showed a certain effect in diagnosis using fMRI images. Therefore, we believe that small data such as a medical data could be learned by generative model. In addition, this model helps to identify disease-related regions of mental disorders, a previously difficult task.

Acknowledgments

This study was partially supported by the JSPS KAKENHI (16K12487), Kayamori Foundation of Information Science Advancement, and SEI Group CSR Foundation.

References

[Raina 03] Raina et al. (2003), Advances in Neural Information Processing Systems 16, pp.545-552

[Kingma 13] Kingma et al. (2013), In Proc. of International Conference on Learning Representations

[Suk 15] Suk et al. (2015), Brain Structure and Function, vol. 220, pp. 841-859

[Tzourio-Mazoyer 02] Tzourio-Mazoyer et al. (2002), NeuroImage, vol.15, pp.273-289

[Malinen 10] Malinen et al. (2010), Proceedings of the National Academy of Sciences of the United States of America, vol.107, pp.6493-6497

[Suk 16] , Suk et al. (2016), NeuroImage, vol.129, pp. 292-307

[Kingma 14] Kingma et al. (2014), Advances in Neural Information Processing Systems 27, pp.3581-3589

[Mitchell 08] Mitchell et al. (2008), The Cerebellum, vol.7, pp. 589-594

[Michael 12] Michael et al. (2012), Rev Neurosci, vol. 23, pp. 253-262