Classification of Parkinson's disease patients using Bayesian deep learning based on fMRI data

Maryam Amirmazlaghani[a,*], Amin Amini[a]

[a]Department of Computer Engineering and Information Technology, Amirkabir University of Technology, Tehran, Iran

ARTICLE INFO

Keywords:
Functional magnetic resonance imaging (fMRI)
Resting-state
Disease diagnosis
Bayesian deep learning
Convolutional neural network (CNN)
Parkinson's disease

ABSTRACT

The goal of resting-state functional magnetic resonance imaging (fMRI) is to investigate the brain's functional connections by using the temporal similarity between blood oxygenation level-dependent (BOLD) signals in different regions of the brain "at rest" as an indicator of synchronous neural activity. However having different images of patients' brains despite the sequential time-based information, may contain enough data to investigate their brains and enables us to recognize some diseases. In this paper, we propose a novel method to automatically extract this non-time sequence-based information using a Bayesian deep learning algorithm based on a convolutional neural network (CNN). Instead of using some predefined points of interest (POIs) we use the whole data in the training phase so those points of the brain which do not contain related information about the disease will be ignored automatically by the trained model. Moreover, this method does not make any assumptions about the disease, patients, etc., makes it a possible universal disease diagnosis approach to differentiate diseases having an impact on brain functionality. This method is a supervised algorithm with a small number of calculations using three-dimensional CNN. Each fMRI scan (which contains t time slices of the brain) of patients will be divided into t different 3D images enabling us to make the dataset much bigger in number and calculations way simpler. Subsequently, all of these images are fed to a Bayesian network similar to LeNet-5 (but in three dimensions) to train our model. Then to determine if a person is suffering from Parkinson's or not, we test his/her t fMRI images and get t different results which leads to a fraction (probability) of how unhealthy his/her brain is and if that fraction is above 0.5 we can classify that sample as a Parkinson's patient.

1. Introduction

Intro

2. Materials and methods

2.1. Participants

We had 30 subjects with ACER over 85 and MMSE over 27, comprising 15 healthy controls `[HC, age 63.33 ± 5.25, 8 females ]` and 15 patients diagnosed with Parkinson's disease `[PD, age 70.73 ± 4.80, 8 females ]`. Significant differences in age and ACER among two groups were found `(age: p=0.0012 , ACER: p=0.015)`. However no significant difference in MMSE were found `(MMSE: p=0.315)`.

2.2. Data acquisition

fMRI scans were obtained from the OpenfMRI project which was acquired on a 3T Siemens Verio with repetition time (TR) = 2.5 s, echo time (TE) = 30ms, flip angle = 80°. Each sample contains 198 slices of $39 \times 64 \times 64$ voxels from the brains of patients. Subjects were instructed to close their eyes and to rest quietly during the scan session [6].

2.3. Data pre-processing

The fMRI data were pre-processed using the standard modules of FMRIB Software Library v5.0 [4]. The pre-processing steps for the anatomical data involved motion correction (MCFLRIT), skull stripping, and spatial smoothing (Gaussian kernel of 5-mm FWHM). Low level noise was removed using a high-pass temporal filtering. The functional images were then aligned to the individual's high-resolution T1-weighted scans, which were subsequently registered to the Montreal Neurological Institute standard space (MNI152) using affine linear registration. and resampled at 2mm cubic voxels. The product of the preprocessing step was used in the experiments.

2.4. Three dimensional convolutional neural networks

3D Convolutional Neural Networks which are inspired by human visual system are similar to classic neural networks. This architecture has been particularly designed based on the explicit assumption that raw data are three-dimensional that enables us to encode certain properties and also to reduce the amount of hyper parameters. The CNN topology utilizes spatial relationships to reduce the number of parameters which must be learned and thus improves upon general feed-forward back propagation training. Equation (1) shows how value of each voxel is calculated from previous layer based on a fixed kernel `(size = P×Q×R)` and a bias for `j`th feature in `i`th layers where v is value, f is the activation function and w is filter weight [2].

$$v_{ij}^{xyz} = f\left(b_{ij} + \sum_{m} \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} w_{ijm}^{pqr} \, v_{(i-1)m}^{(x+p)(y+q)(z+r)}\right) \quad (1)$$

Since fMRI images are 3D time series and our experiment tries to find out if images contain any useful informa-

*Corresponding author
mazlaghani@aut.ac.ir (M. Amirmazlaghani); amin-amini@aut.ac.ir (A. Amini)
ORCID(s):

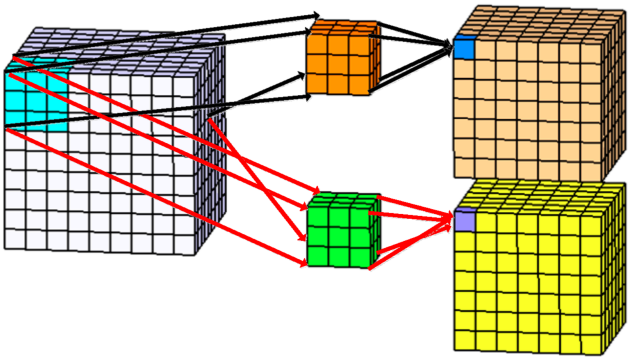

Figure 1: 3D CNN demonstration (kernel size = 3×3×3).

tion despite their sequences, three dimensional CNN can be used as our computational core. A simple 3D CNN can be seen in Figure 1

2.5. Bayesian deep learning

Deep learning has achieved significant success in many perception tasks including seeing (visual object recognition), reading (text understanding), and hearing (speech recognition). These are undoubtedly fundamental tasks for a functioning comprehensive artificial intelligence (AI) or data engineering (DE) system. However, in order to build a real AI/DE system, simply being able to see, read, and hear is far from enough. It should master the ability of thinking.

Take medical diagnosis as an example. Besides seeing visible symptoms (or medical images from MRI) and hearing descriptions from patients, a doctor has to look for relations among all the symptoms and preferably infer the corresponding etiology. Only after that can the doctor provide medical advice for the patients. In this example, although the abilities of seeing and hearing allow the doctor to acquire information from the patients, it is the thinking part that defines a doctor. Specifically, the ability of thinking here could involve causal inference, logic deduction, and dealing with uncertainty, which is apparently beyond the capability of conventional deep learning methods [5].

A neural network with arbitrary depth and non-linearities, with dropout applied before every weight layer, is mathematically equivalent to an approximation to the probabilistic deep Gaussian process [1].

Let $\hat{y}$ be the output of a NN model with L layers and a loss function $E(\cdot, \cdot)$ such as the softmax loss or the Euclidean loss (square loss).We denote by W_i the NN's weight matrices of dimensions $K_i \times K_{i-1}$, and by b_i the bias vectors of dimensions K_i for each layer $i = 1, ..., L$. We denote by y_i the observed output corresponding to input x_i for $1 \leq i \leq N$ data points, and the input and output sets as X, Y. During NN optimization a regularization term is often added. We often use $L2$ regularization weighted by some weight decay λ, resulting in a minimization objective (often referred to as cost) and it can written as Equation (2) [1].

$$\mathcal{L}_{dropout} := \frac{1}{N}\sum_{i=1}^{N} E(y_i, \hat{y}_i) + \lambda \sum_{i=1}^{L}(||W_i||_2^2 + ||b_i||_2^2) \quad (2)$$

Now of if we add dropout to the calculations each unit may remain the same with probability p_i or dropped out (its value is set to zero). We use the same values in the backward pass propagating the derivatives to the parameters. Assume we are given a covariance function of the form of Equation (3).

$$K(x, y) = \int p(w)p(b)\sigma(w^T x + b)\sigma(w^T y + b)dwdb \quad (3)$$

with some element-wise non-linearity $\sigma(\cdot)$ and distributions $p(w), p(b)$. It had been shown that a deep Gaussian process with L layers and covariance function $K(x, y)$ can be approximated by placing a variational distribution over each component of a spectral decomposition of the GPs' covariance functions.

Let W_i be a random matrix of dimensions $K \times K_{i-1}$ for each layer i, and write $\omega = \{W_i\}_{i=1}^{L}$. A priori,we let each row of W_i distribute according to the $p(W)$ above. In addition, assume vectors m_i of dimensions K_i for each GP layer. The predictive probability of the deep GP model given some precision parameter $\tau > 0$ can be parameterized as Equation (4).

$$\begin{aligned} p(y|x, X, Y) &= \int p(y|x, \omega)p(\omega|X, Y)d\omega \\ p(y|x, \omega) &= \mathcal{N}(y; \hat{y}(x, \omega), \tau^{-1} I_D) \\ \hat{y}(x, \omega = \{W_1, ..., W_L\}) &= \\ \sqrt{\frac{1}{K_L}} W_L \sigma(...\sqrt{\frac{1}{K_1}} W_2 \sigma(W_1 x + m_1)...) & \end{aligned} \quad (4)$$

The posterior distribution $p(\omega|X, Y)$ in Equation (4) is intractable. It had been shown that if we use a distribution over matrices whose columns are randomly set to zero, we can recover Equation (2) [1]. Note that previously mentioned distribution $q(\omega)$ can be defined as:

$$\begin{aligned} W_i &= M_i \cdot diag([z_{i,j}]_{j=1}^{K_i}) \\ z_{i,j} &= Bernoulli(p_i)\ \forall\ 1 \leq i \leq L\ ,\ 1 \leq j \leq K_{i-1} \end{aligned}$$

given some probabilities p_i and matrices M_i as variational parameters. The variable $z_{i,j} = 0$ corresponds then to unit j in layer $i-1$ being dropped out as an input to layer i. Subsequently if we use a dropout in 3D convolutional neural network, it would be a practice of Bayesian 3D CNN.

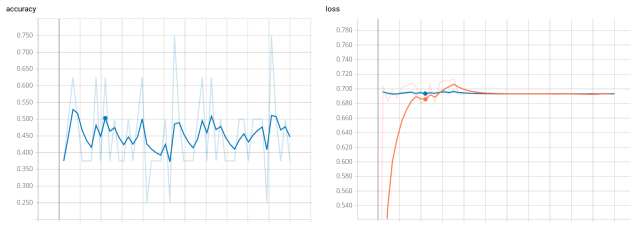

Figure 2: Parkinson's disease diagnosis accuracy and loss using Sarraf and Tofighi's algorithm.

2.6. Proposed method

As mentioned before, our experiment aims to find out if fMRI scans carry any hidden information without taking time-based sequences in action. Accordingly, we examined some previous researches in this field and an eye-catching example of this glance is Sarraf and Tofighi's adopted LeNet-5 architecture [3]. Their algorithm concatenates fMRI images based on z and t axis and does not make any assumptions about the disease or the data (except their unknown patient selection criteria), making it a potential applicable solution to different disease diagnosis tasks. In brief, each patient's scan is presented as a huge two-dimensional matrix which is then fed to a CNN network. Although this method has shown huge success in the case of Alzheimer's (based on large ADNI dataset), our experiments clarify that it is not suitable for Parkinson's disease diagnosis with a small dataset (such as OpenfMRI ds000245). In addition to that, Sarraf and Tofighi's method consumes a huge amount of memory and demands a powerful processor due to its large dimensions of data.

Without a doubt, fMRI scans are nothing but spatial and temporal connectivities. As we aim to focus on spatial relations and put temporal ones aside, our algorithm must try to keep all of that information in calculations. Besides, Sarraf and Tofighi's method loses a lot of these connectivities by flattening data toward the z axis. In addition to that, the mentioned data conversion will increase our model complexity and variance, especially in a case like Parkinson's disease dataset. In fact, by resizing 198 slices of $39 \times 64 \times 64$ voxels to a 2D 2496×12672 pixels, we are adding a lot of unnecessarily and incorrect spatial connectivities based on the t axis and losing some important ones in the z axis by misplacing them. As a result, it is predictable that flattening algorithm may lead to overfitting and compute-intensive calculations which can be seen in Figure 2.

A. My Appendix

Appendix sections are coded under `\appendix`.

`\printcredits` command is used after appendix sections to list author credit taxonomy contribution roles tagged using `\credit` in frontmatter.

CRediT authorship contribution statement

Amin Amini: Conceptualization of this study, Methodology, Software.

References

[1] Gal, Y., Ghahramani, Z., 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning, in: international conference on machine learning, pp. 1050–1059.

[2] Ji, S., Xu, W., Yang, M., Yu, K., 2013. 3d convolutional neural networks for human action recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 35, 221–231. doi:10.1109/tpami.2012.59.

[3] Sarraf, S., Tofighi, G., 2016. Classification of alzheimer's disease using fmri data and deep learning convolutional neural networks arXiv:arXiv:1603.08631.

[4] Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., et al., 2004. Advances in functional and structural mr image analysis and implementation as fsl. Neuroimage 23, S208–S219.

[5] Wang, H., Yeung, D.Y., 2016. Towards bayesian deep learning: A framework and some existing methods. IEEE Transactions on Knowledge and Data Engineering 28, 3395–3408. doi:10.1109/tkde.2016.2606428.

[6] Yoneyama, N., Watanabe, H., Kawabata, K., Bagarinao, E., Hara, K., Tsuboi, T., Tanaka, Y., Ohdake, R., Imai, K., Ogura, R., Kato, Y., Masuda, M., Hattori, T., Ito, M., Maesawa, S., Mori, D., Katsuno, M., Sobue, G., 2018. Olfactory dysfunction and functional connectivity changes in cognitively normal parkinson's disease. https://openfmri.org/dataset/ds000245. This data was obtained from the OpenfMRI database. Its accession number is ds000245 [Online; accessed 25-September-2019].