DISCRIMINATION OF ADHD CHILDREN BASED ON DEEP BAYESIAN NETWORK

A. Junyu Hao, B. Lianghua He

Tongji University, Department of Computer Science and Technology

ABSTRACT

Attention deficit hyperactivity disorder (ADHD) is a threat for the public health all the time, so the effective discrimination of it is significant and meaningful. In current research, Functional Magnetic Resonance Imaging (fMRI) data has become a popular tool for the analysis of ADHD. In this paper, we introduce the Deep Bayesian Network, a combination of Deep Belief Network and Bayesian Network, to classify the ADHD children from the normal. In Deep Bayesian Network, The Deep Belief Network is applied to normalize and reduce dimension of the fMRI data in every brodmann area. And the Bayesian Network is used to extract the feature of relationships between several well-performed brain areas by structure learning. According to the information of structure and probability in Bayesian Network, we predicted the subjects as control,combined ,inattentive or hyperactive using SVM classifier. The final results perform better than using single Deep Belief Network and the best results in ADHD-200 competition.

Keywords: ADHD, Deep Learning, Bayesian Network, SVM, Deep Belief Network

1. INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is among the most common psychiatric disorders of childhood that persists into adulthood in the majority of cases[1]. According to American Psychiatric Associations Diagnostic and Statistical Manual, the prevalence of ADHD in the whole world is approximately 5%, especially in the United States the prevalence among 8 to 15-years-olds reaches to 8.7% during the past years. Therefore, the methods of diagnosing the ADHD are in urgent need.

As an excellent method of measuring brain activation, fMRI signals are studied to classify ADHD. Shengfu Liang et al.[2] utilizes the LDA classifier to discriminate ADHD by analyzing the rs-fMRI data, and the average accuracy of distinguishing normal and ADHD children reaches 80.08% through 50 times of 2-fold validation. Xunheng Wang[3] applies Kernel Principal Component Analysis (KPCA) method based on connectivity matrix of each functional meaningful brain region to find the abnormal pattern of ADHD. Then Support Vector Machine (SVM) as a classifier increases the accuracy rate to 81% using a leave-one-out cross validation. In the global ADHD-200 competition, Eloyan A et al.[4] has achieved relatively better scores by using rs-fMRI based on decomposition of CUR along with gradient boosting. In this paper, three datasets from ADHD-200 competition are applied to discriminate ADHD.

However, it will encounter a serious problem during analyzing the ADHD with fMRI data, which are the mass of data and data redundancy. Therefore, dimensionality reduction is essential. Although there are many dimensionality reduction methods proposed, but they only take mathematical requirements into consideration, rather than physical requirements. Luckily, Geoffrey E.Hinton et al.[5] derives a fast, greedy algorithm that can learn deep, directed belief networks one layer at a time, provided the top two layers form an undirected associative memory by using complementary prior. Because of the good performance in dimensionality reduction currently, Deep Learning has been applied for many areas including image processing[5], audio classification[6], natural language processing[7] and so on.

Due to the cause in ADHD cases is unknown, the relationships between different areas should be token into consideration rather than analysis the brain area separately, which meets the requirements of Bayesian network. It is a graphical model that can encode probabilistic relationships among variables of interest. On the one hand, the model can be used to learn causal relationships and gain understanding about a problem domain. For example, in bioinformatics Bayesian network has been used for the interpretation and discovery of gene regulatory pathways[8]. On the other hand, it is an ideal representation for combining prior knowledge and data. Based on these, Bayesian network has been used in information retrieval[9], natural language processing[10] and so on.

Figure 1 shows the framework of proposed method. It can be seen that there are totally four steps. The first step is that pre-process the voxels with max-pooling of frequencies(**A** in Fig.1). Then, Deep Belief Network is used to reduce dimensionality of data in every brodmann area(**B** in Fig.1), which reconstructs the different number of voxels in every areas into the same number of features, and the Bayesian network can extract the relationships with the normalized fMRI data(**C-E** in Fig.1). These two steps are the highlights of Deep Bayesian

Thanks to National Natural Sciences Foundation of China.(No.61272267, 61270220, 51075306,61273261)

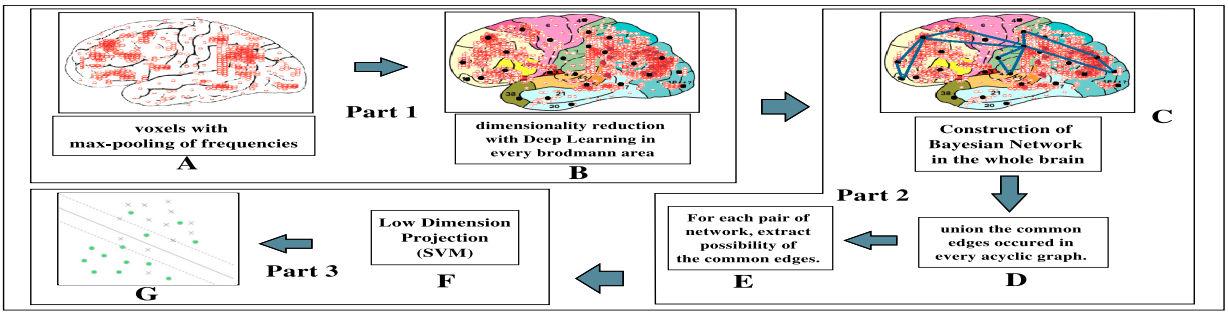


Fig. 1. The procedure of Deep Bayesian Network

Network. Finally, using SVM as a classifier to discriminate the ADHD children from normal(**F-G** in Fig.1).

2. METHOD

The proposed method, Deep Bayesian Network, can be divided into two main parts such as Deep Belief Network for normalization and Dimensionality reduction(Part 1 in Fig.1), Bayesian Network construction for feature extraction(Part 2 in Fig.1). The following sections describe each parts in details. Finally, the SVM is used as a classifier(Part 3 in Fig.1).

2.1. Deep Belief Network

Deep Belief Network is made of a stack of restricted Boltzmann machines(RBM)[11]. It is a two-layer, undirected, bipartite graphical model including visible units and hidden units. The weights W and biases a, b of the RBM determine the energy of a joint configuration of the hidden and visible units $E(v, h)$,

$$E(v, h; \theta) = -\sum_{i=1}^{V}\sum_{j=1}^{H} v_i h_j w_{ij} - \sum_{i=1}^{V} b_i v_i - \sum_{j=1}^{H} a_j h_j \quad (1)$$

with model parameters $\theta = \{W, b, a\}$ and $v_i, h_j \in \{0, 1\}$.

In general Boltzmann machines, the probability distributions are defined in terms of the energy function $p(v, h) = \frac{1}{z}\exp(-E(v, h))$, where z is defined as the sum of $\exp(-E(v, h))$. Based on $p(v, h)$, the conditional probability of v given h and of h given v.

By now, the pre-train of RBM network has mapped the input data into different feature space and kept much feature information. So back propagation is needed to spread the error message to each level of RBM from the top to down. Following the gradient of the log likelihood $\log P(v)$, we obtain the update rule for the weights as,

$$\Delta w_{ij} = \epsilon(< v_i, h_j >_{data} - < v_i, h_j >_{recon}) \quad (2)$$

Where ϵ is the learning rate and the angle brackets manifests the expectations relative to the distribution specified in the subscript.

2.2. Bayesian Network

A Bayesian network $B =< N, A, \Theta >$ is a directed acyclic graph (DAG) $< N, A >$ with a conditional probability table (CPT) for each node, collectively represented by Θ. Each node $n \in N$ represents a domain variable x_i, and each arc $a \in A$ between nodes represents a probabilistic dependency $p(x_i|x_j)$. The joint probability $p(X)$ can be calculated as follows:

$$p(X) = \prod_{i=1}^{m} p(x_i|\pi_i)) \quad (3)$$

where π_i denotes the parent $i = 1, \cdots, m$ nodes of x_i. There are two major tasks in learning a BN: structure learning and parameters learning. This paper mainly uses the structure to extract the information, so we apply Max-Min Hill Climbing (MMHC) algorithm to learn structure efficiently, which is most suitable to the brain areas.

2.3. The Construction of Deep Bayesian Network

2.3.1. Deep Feature Extraction

In this paper, the original fMRI data D is a time series, so it can not be used as the input for deep belief network in every brodmann area. The frequency information can be used as the characteristics of each voxel by the study in our research group. So the input of Deep Belief Network X is pre-processed by realign, slice time, co-register, normalize,smooth which are described in this paper[12] in detail after fast Fourier transform algorithm (FFT).

$$X_f = fft(D) \quad (4)$$

To construct the Deep Belief Network, the forward calculation and backward propagation are needed in every layer.

$$[\theta, X] = RBN(X) \quad (5)$$

$$\theta = backprop(\theta, X) \quad (6)$$

The pseudo-code for application of Deep Belief Network on ADHD data can be seen as following Algorithm 1 .

Algorithm 1 Deep Feature Extraction

Input: original fMRI data D

1: Get frequencies in every voxel X_f

$$X_f \leftarrow fft(D)$$

2: Pre-process frequency information: $X \leftarrow prepro(X_f)$

3: **for** $k \leftarrow 1, 3$ **do**

4: train the weights $\theta = [w, a, b]$ of single rbm

$$[\theta, X] \leftarrow RBN(X)$$

5: **end for**

6: Adjust weights with Backpropagation

$$\theta \leftarrow backprop(\theta, X)$$

7: Reduce dimensionality: $Y \leftarrow DBN(\theta, X)$

Output: dimensionality-reduced data Y

Algorithm 2 Structure Learning of Deep Bayesian Network

Input: dimensionality-reduced data Y, labels of data L

1: Filter the data: $X_f \leftarrow filter(Y)$

2: Test Conditional independence

$$I \leftarrow indep(X_f)$$

3: union the parent's limitation: $U \leftarrow union(I)$

4: Learning structure of Bayesian Network

$$[G, P] \leftarrow BN(U, Y_f)$$

5: Extract feature: $F \leftarrow extract(G, P)$

6: Classify the features: $C \leftarrow svm(F, L)$

7: Computer accuracy: $T \leftarrow com(L, C)$

Output: accuracy of discrimination T

2.3.2. Structure Learning of Deep Bayesian Network

Duo to the reason of ADHD is unknown, it is meaningful to analysis the dimensionality-reduced data Y and sample label L in the whole brain with Bayesian Network. First, filter the data Y to wipe out the noise. Second, the MMHC algorithm can speed up the structure of Bayesian network, but it needs the limited parent nodes of each node. So the dependent nodes I are computed by conditional independence testing in filtered data Y_f.

$$I = indep(Y_f) \tag{7}$$

Third, we union the limited parent nodes U and use it to learn the structure of all data including training data and testing data. By this, we can get a DAG G and a table P including the information of probability of edge shown in equation below.

$$[G, P] = BN(U, Y_f) \tag{8}$$

Forth, the probability of each edge in Bayesian network will be extracted out and viewed as feature F of ADHD children and normal children. Finally, SVM will be acted as classifier to train the training samples and classify the testing samples, which will get the classified label of testing sample C. The accuracy T can be computed by predicted labels L and true labels Y. The pseudo-code for application of BN on ADHD data can be seen as following Algorithm 2.

3. EXPERIMENTS AND RESULTS

3.1. Data

The data used in this paper can be downloaded from the ADHD-200 Global Competition website. DBN and BN model is built upon the ADHD dataset for NYU, Peking-1 and KKI respectively.

For NYU, the training subjects are 216, and testing subjects are 41; for Peking-1 dataset, the training subjects are 85, and testing subjects are 50; and for KKI the training subjects and testing subjects are 83 and 11 respectively. The detail information for the subjecets is shown in Table 1.

Table 1. Demographic Information of three Datasets

type	NYU		Peking-1		KKI	
	train	test	train	test	train	test
	216	41	85	50	83	11
control	98	12	61	27	61	8
combined	73	22	7	9	16	3
inattentive	2	0	0	1	5	0
hyperactive	43	7	17	13	1	0

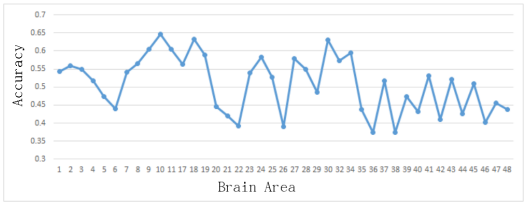
3.2. The Experiment of Parameter Design

To construct a Deep Bayesian Network, we must set up the number of selected areas. If the value of number is small, Bayesian network will not extract the enough information of relationships between brain areas. If the value of number is too large, the probability that the unrelated brain areas will be chose as a node will get higher, which has a side effect on the final classification. In addition, with the increase of the variables, the search space of network structure will present exponentially. So the experiment of parameter designing is significant.

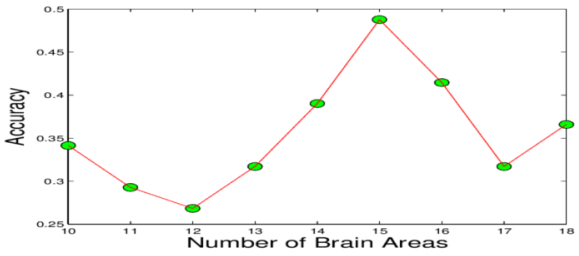
3.2.1. Parameter Design in NYU dataset

As NYU dataset in the ADHD-200 competition achieved the lowest discrimination results, Deep Belief Network is particularly tested on the NYU dataset of 48 areas and softmax

as a classifier generates the accuracies of different brain areas,which is convenient for the choice of node in the Bayesian network. The results of 41 regions are shown in Figure 2(a).



(a) Accuracy of NYU in 48 Areas



(b) Accuracy in Different Number of Brain Areas

Fig. 2. Parameter Design in NYU dataset

From Figure 2(a), we can see that different brain areas have different performance of discrimination. The areas of 10, 18, 30, 9, 11, 19, 34, 32, 8, 17, 2, 28, 3, 1, 7, 23, 25, 41 perform well. According to brodmann definition, it is clear that prefrontal cortex(9,10,25), visual cortex(8,17,18,19), somatosensory cortex(1,2,3,7) and cingulate cortex(23,30,32) is related to the ADHD closely. Therefore, they are selected as the input of Bayesian Network. The Figure 2(b) shows that the accuracy of discriminate fluctuate a little from 10 to 18, but it reaches the peak when the number of brain areas is equal 15. So we will choose 15 brain areas in the front to construct the Bayesian network. Besides, using the relationships between brain areas to discriminate ADHD is better than using information in single brain area and the best result in ADHD-200 competition.

3.3. Performances on NYU, Peking-1 and KKI dataset

The experiments are also excuted on the NYU, Peking-1 and KKI dataset. The results released by ADHD-200 competition are 35.19% for NYU, 51.05% for Peking-1 and 61.90% for KKI respectively. The prediction accuracies of Attributed graph distance[13] (AGD) are none, 58.82%,54.55%. The Deep Bayesian Network gains a higher prediction accuracies than Deep Belief Network single, which are 48.78% for NYU, 54.00% for Peking-1 and 72.72% for KKI. The details show in Figure 3.

From this chart, we can see that Deep Bayesian Network improves the prediction accuracies in these three datasets

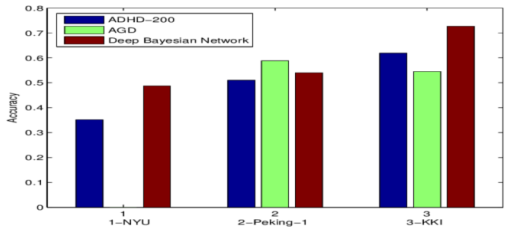


Fig. 3. Performance in Different Datasets

compared with the results of ADHD-200 competition. Besides, the increase of accuracy is the highest in the NYU dataset than other two datasets. The different number of training samples have a big effect on the accuracy. This effect is shown in Figure 4.

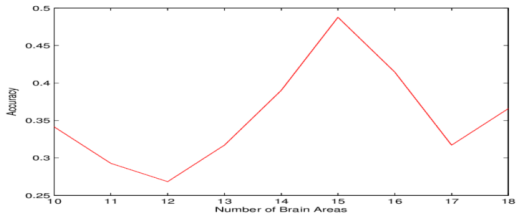


Fig. 4. Performance in Different Number of training samples

Considering that discriminating the ADHD is important and meaningful, here we take the prediction accuracy of the method along with the specificity and sensitivity values into consideration. The detail information is shown in Tab 2.

Table 2. The Detail information in Different Datasets

Dataset	Accuracy	Specificity	Sensitivity
NYU	64.7	68.8	43.9
Peking-1	66.3	87.7	22.9
KKI	59.0	83.0	55.6

4. CONCLUSION

In this paper, a novel method called Deep Bayesian Network, is proposed to classify fMRI ADHD image data. Because of the combination Deep Belief network and Bayeisan network, Deep Bayesian Network can compute relationships among brodmann brain areas more effectively. A series of experimental results also prove that Deep Bayesian network improves the classification performatnce of ADHD greatly comparing with the ADHD-200 competition resutls.

5. REFERENCES

[1] Sandra JJ Kooij, Susanne Bejerot, Andrew Blackwell, Herve Caci, Miquel Casas-Brugué, Pieter J Carpentier, Dan Edvinsson, John Fayyad, Karin Foeken, Michael Fitzgerald, et al., "European consensus statement on diagnosis and treatment of adult adhd: The european network adult adhd," *BMC psychiatry*, vol. 10, no. 1, pp. 67, 2010.

[2] Sheng-Fu Liang, Tsung-Hao Hsieh, Pin-Tzu Chen, Ming-Long Wu, Chun-Chia Kung, Chun-Yu Lin, and Fu-Zen Shaw, "Differentiation between resting-state fmri data from adhd and normal subjects: Based on functional connectivity and machine learning," in *Fuzzy Theory and it's Applications (iFUZZY), 2012 International Conference on*. IEEE, 2012, pp. 294–298.

[3] Xunheng Wang, Yun Jiao, and Zuhong Lu, "Discriminative analysis of resting-state brain functional connectivity patterns of attention-deficit hyperactivity disorder using kernel principal component analysis," in *Fuzzy Systems and Knowledge Discovery (FSKD), 2011 Eighth International Conference on*, vol. 3.

[4] Ani Eloyan, John Muschelli, Mary Beth Nebel, Han Liu, Fang Han, Tuo Zhao, Anita D Barber, Suresh Joel, James J Pekar, Stewart H Mostofsky, et al., "Automated diagnoses of attention deficit hyperactive disorder using magnetic resonance imaging," *Frontiers in systems neuroscience*, vol. 6, 2012.

[5] Geoffrey Hinton, Simon Osindero, and Yee-Whye Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[6] Honglak Lee, Peter Pham, Yan Largman, and Andrew Y Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," in *Advances in neural information processing systems*, 2009, pp. 1096–1104.

[7] Ruhi Sarikaya, Geoffrey E Hinton, and Anoop Deoras, "Application of deep belief networks for natural language understanding.," *IEEE/ACM Transactions on Audio, Speech & Language Processing*, vol. 22, no. 4, pp. 778–784, 2014.

[8] Nir Friedman, Michal Linial, Iftach Nachman, and Dana Pe'er, "Using bayesian networks to analyze expression data," *Journal of computational biology*, vol. 7, no. 3-4, pp. 601–620, 2000.

[9] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, et al., *Modern information retrieval*, vol. 463, ACM press New York, 1999.

[10] Wendy Webber Chapman, Marcelo Fizman, Brian E Chapman, and Peter J Haug, "A comparison of classification algorithms to automatically identify chest x-ray reports that support pneumonia," *Journal of biomedical informatics*, vol. 34, no. 1, pp. 4–14, 2001.

[11] Geoffrey E Hinton, "Training products of experts by minimizing contrastive divergence," *Neural computation*, vol. 14, no. 8, pp. 1771–1800, 2002.

[12] Scott A Huettel, Allen W Song, and Gregory McCarthy, *Functional magnetic resonance imaging*, vol. 1, Sinauer Associates Sunderland, MA, 2004.

[13] Shah M Dey S, Rao A R, *Attributed graph distance measure for automatic detection of attention deficit hyperactive disordered subjects*, vol. 8: 64., Frontiers in Neural Circuits, 2014.