Group Analysis by Visualized Distributional Representation for Resting-state Functional Brain Connectivity

Jiating Zhu[1] and Jiannong Cao[2]

Computing Department, The Hong Kong Polytechnic University
Hong Kong, China
[1]sophie._z@163.com
[2]csjcao@polyu.edu.hk

***Abstract*— The functional brain connectivity based on resting-sate functional magnetic resonance imaging (fMRI) data has been widely studied. However, most of the analyzes, based on functional brain network construction, either limit themselves to descriptors from graph-theoretical measures and subgraph patterns, or fail to gain the whole picture of the large-scale functional brain network. Due to the curse of the dimensionality, the millions-dimension data of the correlation information from fMRIs can not be directly processed by the current machine learning method, even with the state-of-art deep learning architecture. Hence, a good representation that can be both informative and compressed is strongly desired.**

In this paper, we introduce a distributional representation that can provide information across high resolution regions to low resolution regions in a brain. It is believed that a good representation of functional brain connectivity can easily separate patients with brain disorders from normal controls. In our case, the difference between groups is visualized clearly with the distributional representation. The visualized group difference not only gives a visible explanation for existing techniques in functional brain network construction, but also provides extra clues for the informativeness of weak connections, which are neglected in traditional analysis.

Moreover, we propose an approach with distributional representation mappings to validate the effectiveness of the features visualized in the distributional representation. We find that the mapped distributional representation improves the performance dramatically in distinguishing subjects with a given mental disorder from the control ones. The results suggests that the distributional representation is promising for further design of algorithms, such as clustering and classification, for brain disorders.

I. Introduction

The connectivity information from resting state functional magnetic resonance imaging (R-fMRI) data of a single subject is usually stored as a correlation matrix with ten thousand by ten thousand dimensions. To investigate the underlying functional brain connectivity in such a high dimensional space is challenging itself [1].

Conventionally, the functional brain connectivity is represented by features derived from the functional brain network constructed from the correlation matrix [2]. This approach ensures that computational analysis can be realized with traditional machine learning methods. The popular features that have been extensively studied are graph-theoretical measures [3][4] and subgraph patterns [5] in the functional brain networks. Such features, however, are limited to their formalized definition and can not summarize the whole picture of the data.

The deep-learning method, on the other hand, is a powerful approach to represent high dimension data. However, it mostly takes no more than thousands dimension data as input once at a time [6]. Hence, simply putting the high spatial resolution fMRI data into a deep network architecture is not feasible. The deep neural network has to store input data, weight parameters and activations to propagate their information through the network while training [6]. Current memory capacity is not enough for millions-dimension data such as the voxel-based brain network constructed directly from fMRIs. Furthermore, low dimension representations learned from deep neural networks are task-related, which are entangled with the predefined architecture and are not interpretable.

Existing works, such as tensor-based network embedding [7] and deep learning for functional network mining [8][9], have tried state-of-art machine learning methods on a region-based network, where the network is smaller than 100×100 dimensions. Nevertheless, the region-based network neglected the inter-region information. It has been studied that the voxel-based network shows more prominent characteristics than region-based network, such as small-world property [10].

In this paper, we study a novel representation, namely *distributional* [11], which can capture the whole picture of the functional connectivity information in a brain. We propose a framework for analyzing distributional representations in a visible way. Moreover, we suggest an approach to find better distributional representations for individuals and centroids of groups. By computational evaluation and investigation on the distributional representation, we demonstrated that our approach obtains the following findings:

- Visual distinction of the brain disorder group from the control group. Such distinction can be enhanced dramatically by either suppressing the variation across subjects in a group or augmenting the difference between centroids of groups, which are suggested by our designed

mappings.

- The potential to capture some valuable information which is neglected from the traditional ones. This is promising for further designs of better algorithms such as clustering and classification. The reported state-of-art results on large data set achieved 71.2% accuracy in identifying a brain disorder [8], which still has a lot of room to be improved.
- The (visible) reason why the threshold goes down from 0.4 in a voxel-based network to 0.2 in a region-based network. This can guide the threshold selection for traditional brain network construction.
- The (visible) reason whether a voxel-based or a region-based analysis is suitable for analyzing a given brain disorder. This can give a (visible) guide to future study for identifying a mental disease in ether voxel-level or region-level.

II. Definitions and terminologies

A. Functional brain data

Resting state functional magnetic resonance imaging (R-fMRI) measures fluctuations in blood-oxygen-level dependent (BOLD) signals in subjects at rest [4]. The correlations in neural activity between distant brain regions can be measured from the correlation values between the time series of BOLD signals in pairwise regions. The correlation measures can be Granger causality, independent components, mutual information, and most commonly used Pearsons correlation coefficient values. A functional brain network can be constructed by regarding distant brain regions as nodes and the correlations between regions as edges.

The smallest region unit in R-fMRI is voxel which is a three dimensional pixel. Usually the size of the voxel-based networks are approximately 16,000×16,000, while region-based networks are 90×90 [10]. Region-based networks extracted brain regions by structurally defined anatomical masks. The number of high-resolution regions is close to the scale of the voxel-based network, whereas the number of low-resolution regions is close to the scale of the region-based network.

B. Distributional representation for a single subject

1) Correlation distribution: A functional brain network can be represented by a correlation matrix R, where each entry r_{ij} represents the Pearsons correlation coefficient between region i and region j. The correlation distribution counts the number of occurrences of each different correlation value in the correlation matrix R. As the correlation matrix R is symmetric, we only count the value of the entry r_{ij} that satisfies $i > j$. The correlation distribution can be approximated by calculating the histogram for discretized correlation values.

The advantage of such distributional representation over the correlation matrix is that it will not be affected by the errors from the anatomical alignment across brains.

2) Joint distribution: However, the spatial information between two regions can not be captured by the correlation distribution. We introduce a distance distribution which counts the number of occurrences of each different distance value in a distance matrix D, where the entry d_{ij} is the Euclidean distance between the centers of regions i and j in a brain. Again, we only count the value of the entry d_{ij} that satisfies $i > j$, as the distance matrix D is symmetric.

The joint distribution, denoted by $h(r, d)$, for a subject is a correlation-distance joint distribution. It represents the probabilities of co-occurrence of different possible correlation values and different possible distance values between regions within a brain. It can be approximated by calculating a two-dimensional histogram, called *join histogram*, for pairwise discretized correlation values and discretized distance values.

3) Conditional distribution: Because the distance distribution is a non-uniform distribution, different distance values will give different contribution of spatial information to the joint probability distribution. To reduce this side effect, we consider a conditional distribution of correlation given distance.

The conditional distribution, denoted by $h(r|d)$, for a subject is a correlation-distance conditional distribution. It represents the correlation distribution when the distance between the regions is given to be a particular value. It can also be approximated by calculating a two-dimensional histogram, called *conditional histogram*, for pairwise discretized correlation values and discretized distance values.

The conditional distribution can be regarded as a representation mapped from the joint distribution. Such a mapping can be denoted as a distribution function Γ, e.g., $h(r|d) = \Gamma(h(r, d)) = h(r, d)/D(d)$, where $D(d)$ is the distance distribution.

4) Distributional representation definition: The distributional representation $h_b(r, d)$ for a subject b can be the joint distribution $h(r, d)$, the conditional distribution $h(r|d)$, or any other distributions to which the joint distribution $h(r, d)$ is mapped.

C. Group of subjects

Investigating the difference of patients with brain disorders from normal controls can facilitate the study of disease mechanisms and for informing therapeutic interventions.

All the subjects that we concern are numbered $1, 2, ..., N$. The group that consists of all the subjects is viewed as the set $B = \{1, 2, ..., N\}$. The group of subjects with brain disorders is regarded as a subset A of the set B, whereas the control group is denoted by another subset T of the set B, where $T \cap A = \emptyset$ and $T \cup A = B$.

The centroid of subjects in a group can be statistically characterized as the mean of the distributional representations for the subjects in the group.

The mean, denoted by $\mu_Y(r, d)$, of the distributional representations for subjects in a group Y is a binary function as follow:

$$\mu_Y(r, d) = \frac{1}{|Y|} \sum_{b \in Y} h_b(r, d), \tag{1}$$

where $|Y|$ is the number of subjects in the group Y and Y is a subset of B, i.e., $Y \subset B$. We call the mean $\mu_Y(r,d)$ the centroid of the group Y.

D. Variation across subjects

The centroid is not enough to characterize a group, due to the variation across subjects within the group.

One way to measure the variation across subjects is the entropy of the group. The entropy, denoted by $S_Y(r,d)$, of a group Y is a binary function as follow:

$$S_Y(r,d) = -\sum_{b \in Y} h_b(r,d) \cdot \log h_b(r,d). \tag{2}$$

The higher the entropy at a given pair (r,d), the lower the variation at the pair (r,d) is.

An alternative approach is the standard deviation of distributional representations. The standard deviation, denoted by $\sigma_Y(r,d)$, of subjects in a group Y is a binary function as follow:

$$\sigma_Y(r,d) = \sqrt{\frac{1}{|Y|}\sum_{b \in Y}[h_b(r,d) - \mu_Y(r,d)]^2}. \tag{3}$$

However, the standard deviations for different groups are not comparable. Hence, we consider the coefficient of variation (CV), which is a standardized measure of dispersion of a probability distribution. The coefficient of variation, denoted by $V_Y(r,d)$, of a group Y is also a binary function, which is

$$V_Y(r,d) = \frac{\sigma_Y(r,d)}{\mu_Y(r,d)}. \tag{4}$$

III. Investigation of Distributional Representations

In this paper, we demonstrate that the distributional representation is valuable in designing algorithms for differentiating groups of subjects, not just useful for finding outliers (alienate shape in visualized distribution [11]).

Figure 1 shows the proposed framework for analyzing distributional representations. We first highlight the body information in the joint distributional representation by blocking the visual noise. Then we try to find group characteristics by visualizing the variation across subjects in the group . The group difference can be visualized by the contrast between the variation of the patient group and the variation of the control group. After that, we can analyze the group difference and extract group features to design distribution mapping functions for better representations. Finally, the quality of the mapped distributional representation will be evaluated with further analysis.

A. Data description

The functional brain data we analyzed here is from the ADHD-200 sample. This public data set collected resting-state fMRI scans from subjects with ADHD and typically developing controls in 8 international imaging sites. We only analyzed the preprocessed data from NYU with the largest number of subjects. The preprocessed data used NIAK pipeline

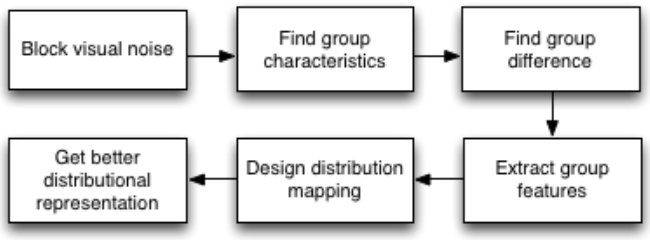

Fig. 1. Framework for analyzing distributional representations.

and has extracted time courses from around 3000 regions of interests [12]. One or two resting-state fMRI scans were acquired for each subject in the NYU data. We picked the first scans (216 subjects) and eliminated the questionable subjects (marked along with the data) and the outliers (see [11]), which leads to 173 subjects (for the set B) for analysis in this study. 84 subjects (for the set T) are typically developing children (TDC) and 89 subjects (for the set A) have ADHD.

The preprocessed data is available at http://preprocessed-connectomes-project.org/adhd200/.

B. Visual variation

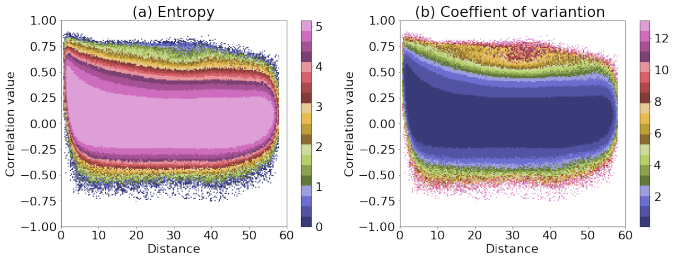

Fig. 2. Measures of the variation across joint histograms, each with 200×200 bins, of all subjects.

We visualize the variation of distributional representations across all the subjects by two measurements, the entropy $S_B(r,d)$ and the coefficient of variation $V_B(r,d)$.

Figure 2(a) shows the entropy $S_B(r,d)$, whereas Figure 2(b) shows the coefficient of variation $V_B(r,d)$. We can see that the entropy $S_B(r,d)$ has a pattern similar to the one of the coefficient of variation $V_B(r,d)$. The pink area of the former occupies the same region as the blue one of the latter does in the distance-correlation plane. This implies that both of them are able to measure the variation across of subjects.

However, the area with a higher value of the entropy $S_B(r,d)$ corresponds to the one with a lower value of the coefficient of variation $V_B(r,d)$. For instance, the pink area, which reaches at the highest value, of the entropy $S_B(r,d)$ corresponds to the blue area, which reaches at the lowest value, of the coefficient of variation $V_B(r,d)$. This means that the higher the entropy of an area, the lower value the variation of the area has. Moreover, the coefficient of variation $V_B(r,d)$ shows a wider range of values than the entropy $S_B(r,d)$ does, see the color bars in Figure 2.

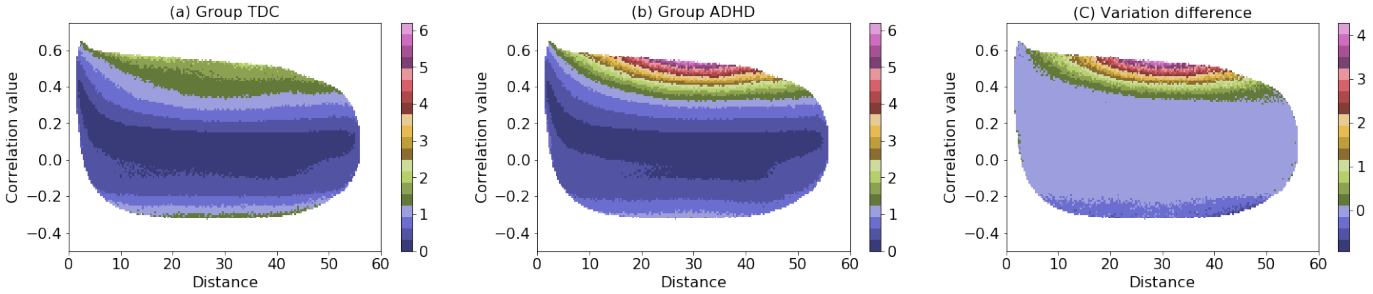

Fig. 3. The coefficient of variation across the joint distributional representations of subjects in a group.

One advantage of visualized variation is that it helps to remove visual noise of distributional representations.

There are many dots around the boundary of the colorful part in both (a) and (b) of Figure 2. These high values of variation affect the appearance of the representation, which makes the important area overwhelmed by the visual noise. We regard as visual noise the count value smaller than a given value th in the joint histogram of the centroid $\mu_B(r, d)$. Here, the given value th is set to 5, which is 5% of the mean of the count values in the histogram. The visual noise can be removed to highlight the main body for visualization. Figure 3 shows the variations after the visual noise removal. The number of bins after noise removal becomes 14092 instead of 200×200.

Notice that the visual noise should not be removed in computational analysis. Our experiments show that any arbitrary removal will influence the performance of the representation. We think that the visual noise is indispensable to the integrity of the distributional representation.

C. Distinction of brain disorders against controls

The coefficient of variation across the joint distributional representations of subjects in either of the groups TDC and ADHD is shown in Figure 3(a) and (b), respectively. The variation of the group ADHD is apparently greater than the variation of the group TDC in the upper region of the distance-correlation plane. Subtracting the variation of the group TDC from the variation of the group ADHD results in the variation difference in Figure 3(c). This implies that the connectivity in the group ADHD has larger variation than the control group in the region where the correlation value is greater than 0.4. On the other hand, small variation means stable connectivity when the correlation value is below 0.4.

When we subtract the centroid of the group TDC from the centroid of the group ADHD, we get the group difference as shown in Figure 4(d). The middle part of the body has higher contrast than the upper part. Considering the fact that the middle part has almost none variation as shown in Figure 3(c), we can deduce that the middle part of the distributional

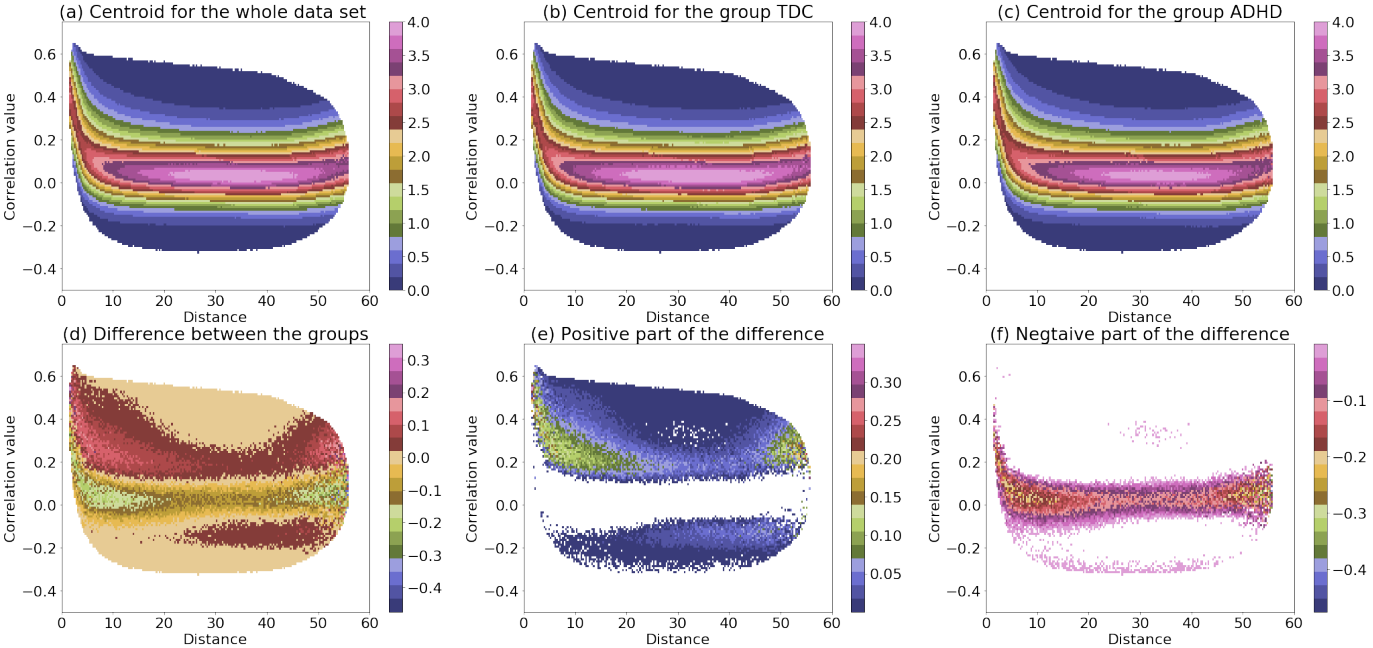

Fig. 4. The centroid of the conditional histograms in a group and the centroid difference between two groups.

representation has stable characteristics to differentiate the group ADHD from the group TDC. More interesting finding is that there are positive and negative parts from the subtraction (Figure 4(d) and (e)). This means that the two distributional representations for the centroids of the two groups cross over with each other in a three dimensional space. This difference between the two groups can not been seen directly from visualized distributional representations in Figure 4(a)-(c).

IV. Better distributional representations

Inspired from the above investigation, we would like to 1) suppress areas with large variation across subjects in a group, and 2) enhance the group difference to provide strong indicators of group characteristics in the representation. We will discuss the effectiveness of this two principles to improve the distributional representation in the following subsections.

It is believed that a good representation can easily separate groups of subjects. Hence, we design a (rough but quick) test to the performance of the distributional representation in categorizing subjects.

A. Group differentiation method

Here, we use a simple approach to distinguish which group a subject belongs to.

The difference for two distributional representations $h_1(r,d)$ and $h_2(r,d)$ can be measured by the Kullback-Leibler (KL) divergence

$$D_{\mathrm{KL}}(h_1||h_2) = \sum_{i,j} h_1(r_i,d_j)\log\frac{h_1(r_i,d_j)}{h_2(r_i,d_j)} \tag{5}$$

where r_i and d_j are discretized values of correlation and distance, respectively.

We regard that a subject $b \in B$ falls into the group A if its distributional representation $h_b(r,d)$ is closer to the centroid of the group A than the centroid of the group T. Hence, by simply comparing KL values $KL_T(h_b)$ and $KL_A(h_b)$ of an individual b, where

$$KL_T(h_b) = D_{\mathrm{KL}}(h_b||\mu_T), \tag{6}$$

$$KL_A(h_b) = D_{\mathrm{KL}}(h_b||\mu_A), \tag{7}$$

we can guess that the individual b falls into the group that has the smaller KL divergence to it.

We use sensitivity (true positive rate) to measure the performance of a representation in identifying subjects. The sensitivity for the group TDC is the percentage of healthy subjects that are correctly identified as TDC, whereas the sensitivity for the group ADHD is the percentage of individuals with ADHD that are correctly identified as ADHD. The total performance is measured as the percentage of individuals that are correctly identified in the whole data set B.

B. Improvement by variation suppression

Large variation across distributional representations of subjects in a group will lead to wrong judgment in distinguishing group characteristics. Hence, we transform the larger variations into smaller ones, to reduce their contributions to the representation. For each subject b in a group Y, we have the mapping function

$$\Psi_Y(h_b(r,d)) = \exp(-\alpha V_Y)\cdot h_b(r,d), \tag{8}$$

where α is a coefficient parameter that can be changed. Sometimes, the mapping function Ψ_Y is written as $\Psi_Y(h_b(r,d))_\alpha$ to emphasis the parameter α.

Thus, the mapped centroid of the group Y will be

$$\begin{aligned}\mu_Y\Psi_Y(r,d) &= \frac{1}{|Y|}\sum_{b\in Y}\Psi_Y(h_b(r,d)) \\ &= \frac{1}{|Y|}\sum_{b\in Y}\exp(-\alpha V_Y)\cdot h_b(r,d).\end{aligned} \tag{9}$$

To evaluate the mapping function Ψ, we calculate the mapped centroids $\mu_T\Psi_T(r,d)$ and $\mu_A\Psi_A(r,d)$ for groups TDC and ADHD, respectively. For each subject b in the group B (the whole data set), we calculate two mapped representations $\Psi_T(h_b(r,d))$ and $\Psi_A(h_b(r,d))$, and then calculate the KL divergence from either of groups TDC and ADHD respectively by

$$KL_T(h_b) = D_{\mathrm{KL}}(\Psi_T(h_b)||\mu_T\Psi_T), \tag{10}$$

$$KL_A(h_b) = D_{\mathrm{KL}}(\Psi_A(h_b)||\mu_A\Psi_A). \tag{11}$$

Figure 5 illustrates the computing procedure of the KL divergence with the mapping function Ψ.

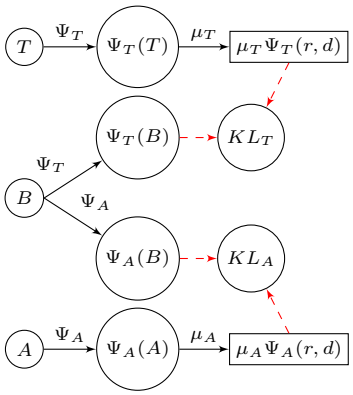

Fig. 5. Illustration for the mapping function Ψ and the KL divergences from the centroids to each subjects. The concrete arrows represent unary functions, whereas the dashed arrows represent binary operations.

Figure 6 shows the performance of Ψ with different values of α, ranged from 0 to 100 with an interval length of 0.5. We can see that the total performance climbs steadily as α increases and reaches to 90% eventually. The performance for identifying ADHD keeps raising until about 97%, whereas the performance for identifying TDC remains around 85% with some fluctuations.

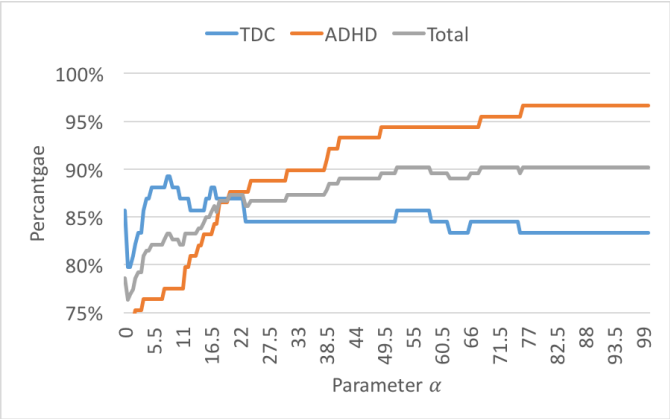

Fig. 6. Performance of $\Psi(h(r|d))$ with different values of α.

The number of individuals that are correctly identified in either of groups TDC and ADHD with the mapping function Ψ is shown in Table I, which transform the conditional representation $h(r|d)$ into the mapped representation $\Psi(h(r|d))$. We can see that the parameter α that gives the best performance is 52.5 or 76.0. It can distinguish 90.17% subjects correctly, i.e., 156 out of 173 subjects in the set B are identified correctly (see the last column). The mapped representation improves 11.56% than the one without mapping. The performance without mapping is 78.61% (see the row at $\alpha = 0.0$).

TABLE I

RESULTS OF $\Psi(h(r|d))$ WITH DIFFERENT VALUES OF α.

α	TDC	ADHD	Total
0.0	72(85.71%)	64(71.91%)	136(78.61%)
0.5	67(79.76%)	65(73.03%)	132(76.30%)
8.0	75(89.29%)	69(77.53%)	144(83.24%)
11.0	73(86.90%)	69(77.53%)	142(82.08%)
16.5	74(88.10%)	74(83.15%)	148(85.55%)
19.5	73(86.90%)	77(86.52%)	150(86.71%)
23.5	71(84.52%)	78(87.64%)	149(86.13%)
52.5	72(85.71%)	84(94.38%)	**156(90.17%)**
62.5	70(83.33%)	84(94.38%)	154(89.02%)
76.0	70(83.33%)	86(96.63%)	**156(90.17%)**

The season why the mapping Ψ can largely improve the performance for the group ADHD is that there are some ares with large variation across distributional representations of subjects in the group ADHD as shown in Figure 3(b), which can be suppressed effectively by the mapping Ψ.

C. Discovery of the reason for threshold selection

If we subtract the centroid $\mu_T \Psi_T(r,d)$ of the group TDC from the centroid $\mu_A \Psi_A(r,d)$ of the group ADHD, we have the contrast difference shown in Figure 7. In comparison to Figure 4, the transformed distributional representation leads to a more balanced difference between positive and negative values.

The axis of the distance in the distance-correlation plane reflects the region resolution. The bigger the distance, the lower the region resolution, and vice versa. This gives visible reason to determine the region resolution for analysis. For example, it is suitable for high region resolution analysis if a brain disorder shows significant difference around small distance values; otherwise, it is suitable for low region resolution analysis.

In Figure 7(b), we can tell that the prominent feature for the group ADHD is the green part. The changes along distance values indicate the connectivity changes along with resolutions of the regions. At a distance smaller than 3, the correlation ranging from 0.4 to 0.7 has higher probabilities. This implies that the range (from 0.4 to 0.7) of the correlation is suitable for analyzing the brain disorder ADHD in a high resolution region level. This is consistent with the common agreement for the correlation threshold selection in constructing functional brain networks. In constructing voxel-based networks, usually the lower bound of the correlation threshold is 0.4 to ensure the sparsity of the adjacency matrix [2], while the upper bound is 0.7 to avoid fragmentation and graininess [10].

Furthermore, the middle parts in Figure 7(b) and (c) show that the prominent group difference is in a correlation range from 0 to 0.25, which implies that it is suitable for a lower resolution region (e.g., the distance around 30). Again, such a correlation range is consistent with the threshold reported in [13], whose authors studied 400 regions in the ADHD-200 sample.

Moreover, the correlation value below 0.4 also contains valuable information as it is shown in Figure 7(b). We believe that the green part covering the area of the correlation range from 0.2 to 0.4 with distance values smaller than 25, which is often neglected for the network analysis, will be useful with further studies.

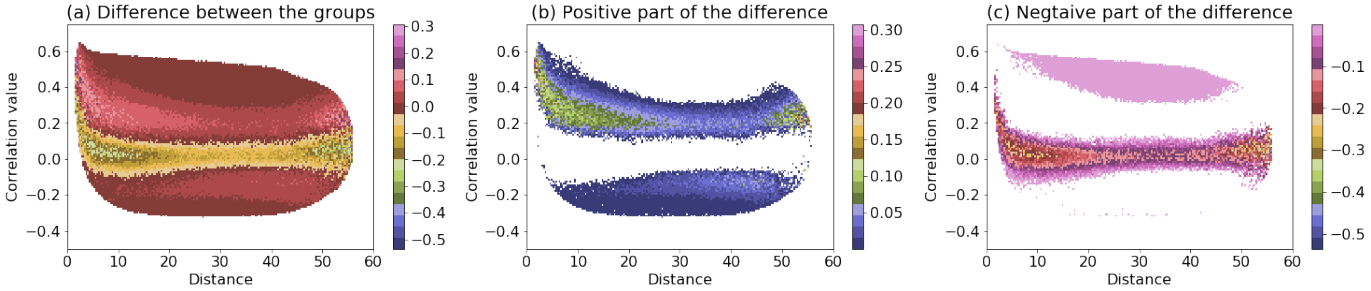

Fig. 7. Group difference with the mapping function Ψ of conditional histograms given $\alpha = 0.5$.

The discussion above not only suggests that our visualized distributional representation supports existing works, but also implies that our work can guide for functional network analysis for other datasets if the group connectivity difference is unclear.

D. Feature enhancement

To further improve the representation, the difference between groups TDC and ADHD can be enhanced by the transformation

$$\Phi(h_b(r,d)) = [\mu_T\Psi_T(r,d)_\beta - \mu_A\Psi_A(r,d)_\beta]^2 \cdot h_b(r,d), \quad (12)$$

where $b \in B$ and β is a coefficient parameter that can be changed.

We compare the mapped representation $\Phi(h_b(r,d))$ for each subject $b \in B$ with fixed centroids $\mu_A\Psi_A(r,d)$ and $\mu_T\Psi_T(r,d)$ respectively by

$$KL_T(h_b) = D_{\mathrm{KL}}(\Phi(h_b)||\mu_T\Psi_T), \quad (13)$$

$$KL_A(h_b) = D_{\mathrm{KL}}(\Phi(h_b)||\mu_A\Psi_A). \quad (14)$$

Figure 8 illustrates the procedure of calculating the KL divergence for the mapping Ψ.

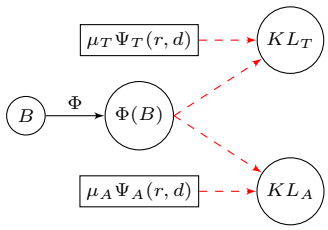

Fig. 8. Illustration for the mapping function Φ and the KL divergences from the centroids to each subjects. The concrete arrows represent unary functions, whereas the dashed arrows represent binary operations.

From Equation (9), we have $\mu_A\Psi_A(r,d) = \mu_A(r,d)$ and $\mu_T\Psi_T(r,d) = \mu_T(r,d)$ when $\alpha = 0.0$. For this reason, we regard that the performance of the mapping Φ has nothing to do with the mapped centroids in the case of $\alpha = 0.0$. Table II shows the number of individuals that are correctly found in each group when $\alpha = 0.0$, where the conditional representation $h(r|d)$ is transformed into the mapped representation $\Phi(h(r|d))$. Figure 9 shows the performance with different values of β, ranged from 0.1 to 9.9 with an interval length of 0.1. The best performance is 94.22%, which improves 15.61% in comparison to 78.61%, the performance without mapping (see the row at $\alpha = 0.0$ in Table I). This reveals that the mapping Φ is also effective.

The mapped centroids given different values of α, ranged from 0.1 to 0.5 with an interval length of 0.1, have been tried. The best performance is found at $\alpha = 0.1$ and $\beta = 0.7$. The number of individuals that are correctly found in each group given $\alpha = 0.1$ is shown in Table III. Figure 10 shows the performance with different values of β from 0.1 to 9.9 with an interval length of 0.1.

The best performance under this condition can distinguish 95.38% subjects correctly, i.e., 165 out of 173 subjects in the

TABLE II

RESULTS OF $\Phi(h(r|d))$ WITH DIFFERENT β GIVEN $\alpha = 0.0$.

β	TDC	ADHD	Total
0.1	75(89.29%)	80(89.89%)	155(89.60%)
0.5	79(94.05%)	84(94.38%)	**163(94.22%)**
0.8	78(92.86%)	82(92.13%)	160(92.49%)
1.0	79(94.05%)	82(92.13%)	161(93.06%)
1.8	80(95.24%)	73(82.02%)	153(88.44%)
2.7	84(100.00%)	70(78.65%)	154(89.02%)
8.5	84(100.00%)	72(80.90%)	156(90.17%)

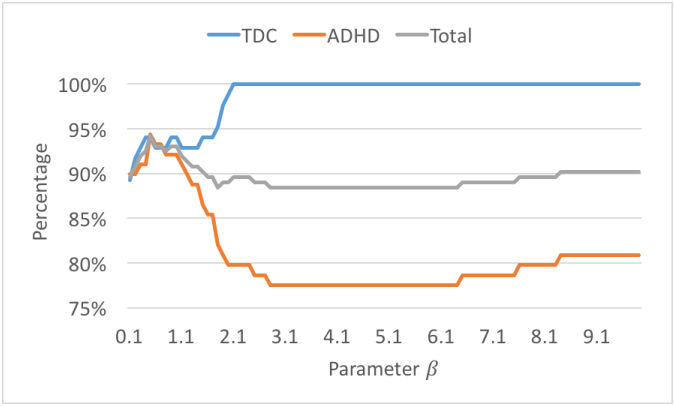

Fig. 9. Performance of $\Phi(h(r|d))$ with different β given $\alpha = 0.0$.

TABLE III

RESULTS OF $\Phi(h(r|d))$ WITH DIFFERENT β GIVEN $\alpha = 0.1$.

β	TDC	ADHD	Total
0.1	77(91.67%)	80(89.89%)	157(90.75%)
0.7	82(97.62%)	83(93.26%)	**165(95.38%)**
1.0	82(97.62%)	81(91.01%)	163(94.22%)
1.1	82(97.62%)	82(92.13%)	164(94.80%)
3.0	84(100.00%)	69(77.53%)	153(88.44%)
8.4	84(100.00%)	72(80.90%)	156(90.17%)

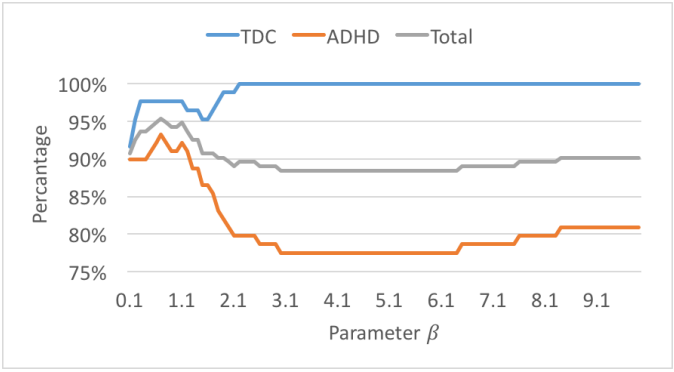

Fig. 10. Performance of $\Phi(h(r|d))$ with different β given $\alpha = 0.1$.

set B are identified correctly (see the last column), where 97.62% and 93.26% subjects in groups TDC and ADHD are correctly identified respectively. In comparison to the results of the conditional distribution representation, it improves 16.77%. This reveals that the mapping Φ is effective when the centroids are fixed.

In Figure 9 and Figure 10, the peaks of the total performance both appear in a range of β from 0.1 to 1.8. The performance for either of groups TDC and ADHD diverges apart each other when β is bigger than 1.8 and the total performance goes down to the start point around 90%.

TABLE IV
PERFORMANCE FOR DISTRIBUTIONAL REPRESENTATIONS WITH DIFFERENT MAPPINGS.

Representation	TDC	ADHD	Total
$h(r,d)$	52(61.90%)	45(50.56%)	97(56.07%)
$\Gamma(h(r,d))$	72(85.71%)	64(71.91%)	136(78.61%)
$\Psi(\alpha = 52.5)$	72(85.71%)	84(94.38%)	156(90.17%)
$\Psi(\alpha = 76.0)$	70(83.33%)	86(96.63%)	156(90.17%)
$\Phi(\alpha = 0.0, \beta = 0.5)$	79(94.05%)	84(94.38%)	163(94.22%)
$\Phi(\alpha = 0.1, \beta = 0.7)$	**82(97.62%)**	**83(93.26%)**	**165(95.38%)**
$\Phi(\alpha = 0.3, \beta = 1.1)$	84(100.00%)	81(91.01%)	165(95.38%)

From our experiments, we conclude that all mapped distributional representations improve the performance dramatically than the joint distributional representation $h(r,d)$, see Table IV. The representation produced corporately by three mapping functions Ψ, Φ and Γ raises the performance to 95.38% (see the second row from the bottom), which improves 39.31% in contrast to 56.07%, the performance of the joint distributional representation $h(r,d)$.

These results indicate that our findings through the data visualization for distributional representations are promising. More importantly, the two proposed principles, namely *variation suppression* and *feature enhancement*, for designing mappings of distributional representations are effective for distinguishing subjects between groups, though our test is not a classification task.

V. CONCLUSIONS

Our analysis on the group difference with visualized distributional representations can explain the threshold selection range for the main-stream functional network construction in both voxel-based and region-based. Moreover, the area with its correlation below 0.4 also shows a prominent group difference in our distributional representations. This phenomena supports the idea from researchers who think that weak connections in the brain also have potential valuable information for functional network construction [14][15]. Hence, we believe that visualized distributional representations can guide the brain connectivity study for other unclear brain disorders.

Two simple mapping functions have been designed to enhance the group features that are found in the visualized distributional representation. The improvement on distinguishing subjects from different groups indicates not only the effectiveness of the visualization approach but also the informativeness of the distributional representation. The performance, raised by the two mapping functions, in differentiating groups is already highly above chance without further exploration in more distribution mappings and parameters' optimization. This is promising for applying distributional representations to further studies of algorithms, such as clustering and classification for brain disorders.

In conclusion, the distributional representation provides 1) a whole picture of the functional connectivity in the brain in contrast to traditional functional network analysis, 2) interpretable features for functional brain networks other than task-related representation with deep-learning approaches, 3) informative visualization analysis, and 4) effective guidance for further analysis in group difference.

REFERENCES

[1] K. A. Norman, S. M. Polyn, G. J. Detre, and J. V. Haxby, "Beyond mind-reading: multi-voxel pattern analysis of fmri data," *Trends in cognitive sciences*, vol. 10, no. 9, pp. 424–430, 2006.

[2] M. P. van den Heuvel, C. J. Stam, M. Boersma, and H. E. H. Pol, "Small-world and scale-free organization of voxel-based resting-state functional connectivity in the human brain." *NeuroImage*, 2008.

[3] X. Wang, Y. Ren, and W. Zhang, "Depression Disorder Classification of fMRI Data Using Sparse Low-Rank Functional Brain Network and Graph-Based Features," *Computational and Mathematical Methods in Medicine*, vol. 2017, no. 6684, pp. 1–11, 2017.

[4] J. Cabral, M. L. Kringelbach, and G. Deco, "Exploring the network dynamics underlying brain activity during rest," *Progress in Neurobiology*, vol. 114, pp. 102–131, Mar. 2014.

[5] B. Cao, X. Kong, J. Zhang, P. S. Yu, and A. B. Ragin, "Mining Brain Networks Using Multiple Side Views for Neurological Disorder Identification," in *2015 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2015, pp. 709–714.

[6] G. E. Hinton, "Reducing the Dimensionality of Data with Neural Networks," *Science*, vol. 313, no. 5786, pp. 504–507, July 2006.

[7] B. Cao, L. He, X. Wei, M. Xing, P. S. Yu, H. Klumpp, and A. D. Leow, "t-bne: Tensor-based brain network embedding," in *Proceedings of the 2017 SIAM International Conference on Data Mining*. SIAM, 2017, pp. 189–197.

[8] S. Wang, L. He, B. Cao, C.-T. Lu, P. S. Yu, and A. B. Ragin, "Structural deep brain network mining," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2017, pp. 475–484.

[9] H.-I. Suk, C.-Y. Wee, S.-W. Lee, and D. Shen, "State-space model with deep learning for functional dynamics estimation in resting-state fMRI," *NeuroImage*, vol. 129, no. C, pp. 292–307, Apr. 2016.

[10] S. Hayasaka and P. J. Laurienti, "Comparison of characteristics between region-and voxel-based network analyses in resting-state fmri data," *Neuroimage*, vol. 50, no. 2, pp. 499–508, 2010.

[11] J. Zhu and J. Cao, "Distributional Representation for Resting-state Functional Brain Connectivity Analysis," in *Submission to the 11th International Conference on Brain Informatics*, 2018.

[12] P. Bellec, C. Chu, F. Chouinard-Decorte, Y. Benhajali, D. S. Margulies, and R. C. Craddock, "The neuro bureau adhd-200 preprocessed repository," *Neuroimage*, vol. 144, pp. 275–286, 2017.

[13] A. dos Santos Siqueira, C. E. Biazoli Junior, W. E. Comfort, L. A. Rohde, and J. R. Sato, "Abnormal Functional Resting-State Networks in ADHD: Graph Theory and Pattern Recognition Analysis of fMRI Data," *BioMed Research International*, vol. 2014, no. 4, pp. 1–10, 2014.

[14] A. Goulas, A. Schaefer, and D. S. Margulies, "The strength of weak connections in the macaque cortico-cortical network," *Brain Structure and Function*, vol. 220, no. 5, pp. 2939–2951, 2014.

[15] E. Santarnecchi, G. Galli, N. R. Polizzotto, A. Rossi, and S. Rossi, "Efficiency of weak brain connections support general cognitive functioning," *Human Brain Mapping*, vol. 35, no. 9, pp. 4566–4582, Mar. 2014.