Deep Hyperalignment

Muhammad Yousefnezhad, Daoqiang Zhang

31st Advances in Neural Information Processing Systems (NIPS-17)

Dec 2017
What is Hyperalignment?
Brain Function Analysis

Deep Hyperalignment

Image | fMRI scan | Voxel pattern | Output
-----|-----------|---------------|-------
Shoe  | ![Shoe fMRI](shoe_fMRI.png) | ![Shoe voxel pattern](shoe_voxel_pattern.png) | =SHOE
Cat   | ![Cat fMRI](cat_fMRI.png) | ![Cat voxel pattern](cat_voxel_pattern.png) | =CAT
Shoe? | ![Shoe? fMRI](shoe_fMRI.png) | ![Shoe? voxel pattern](shoe_voxel_pattern.png) | =SHOE?

A Generalized Approach

\[
\min_{R^{(i)}, R^{(j)}} \sum_{i=1}^{S} \sum_{j=i+1}^{S} \left\| f(X^{(i)}) R^{(i)} - f(X^{(j)}) R^{(j)} \right\|^2_F
\]

s.t. \( (R^{(\ell)})^T \left( \left( f(X^{(\ell)}) \right)^T + \epsilon I \right) f(X^{(\ell)}) R^{(\ell)} = I, \ \ell = 1:S \)

- If \( f(x) = x \) & \( \epsilon = 0 \), then we have the original HA
- If \( f(x) = x \) & \( \epsilon \neq 0 \), then we have the Regularized HA
- If \( f(x) \) is a nonlinear kernel, then we have the Kernel HA
Deep Hyperalignment
DHA: Objective Function

We want to optimize the following function:

\[
\min_{\theta^{(i)}, R^{(i)}} \sum_{i=1}^{S} \sum_{j=i+1}^{S} \left\| f_i(X^{(i)}; \theta^{(i)}) R^{(i)} - f_j(X^{(j)}; \theta^{(j)}) R^{(j)} \right\|_F^2
\]

subject to

\[
\left( R^{(\ell)} \right)^T \left( \left( f_\ell(X^{(\ell)}; \theta^{(\ell)}) \right)^T f_\ell(X^{(\ell)}; \theta^{(\ell)}) + \epsilon I \right) R^{(\ell)} = I, \quad \ell = 1:S
\]

where the deep network is defined as follows:

\[
f_\ell(X^{(\ell)}; \theta^{(\ell)}) = \text{mat}(h^{(\ell)}_C, T, V_{new})
\]

\[
h^{(\ell)}_m = g(W^{(\ell)}_m h^{(\ell)}_{m-1} + b^{(\ell)}_m), \quad \text{where} \quad h^{(\ell)}_1 = \text{vec}(X^{(\ell)}) \quad \text{and} \quad m = 2:C
\]
Generalized DHA

\[
\min_{G, R^{(i)}, \theta^{(i)}} \sum_{i=1}^{S} \left\| G - f_i(X^{(i)}; \theta^{(i)}) R^{(i)} \right\|_F^2
\]

s.t. \( G^T G = I \)

where

\[
G = \frac{1}{S} \sum_{j=1}^{S} f_j(X^{(j)}; \theta^{(j)}) R^{(j)}
\]
DHA: Optimization

- **rank-m SVD**
  \[
  f_\ell(X^{(\ell)};\theta^{(\ell)}) \overset{SVD}{=} \Omega^{(\ell)} \Sigma^{(\ell)} (\Psi^{(\ell)})^T, \quad \ell = 1:S
  \]

- **Projection Matrix**
  \[
  P^{(\ell)} = f_\ell(X^{(\ell)};\theta^{(\ell)}) \left( \left( f_\ell(X^{(\ell)};\theta^{(\ell)}) \right)^T f_\ell(X^{(\ell)};\theta^{(\ell)}) + \epsilon I \right)^{-1} f_\ell(X^{(\ell)};\theta^{(\ell)})^T
  \]
  \[
  = \Omega^{(\ell)} (\Sigma^{(\ell)})^T \left( \Sigma^{(\ell)} (\Sigma^{(\ell)})^T + \epsilon I \right)^{-1} \Sigma^{(\ell)} (\Omega^{(\ell)})^T = \Omega^{(\ell)} D^{(\ell)} (\Omega^{(\ell)} D^{(\ell)})^T
  \]

  where
  \[
  D^{(\ell)} (D^{(\ell)})^T = (\Sigma^{(\ell)})^T \left( \Sigma^{(\ell)} (\Sigma^{(\ell)})^T + \epsilon I \right)^{-1} \Sigma^{(\ell)}.
  \]

- **Sum of Projection Matrices**
  \[
  A = \sum_{i=1}^S P^{(i)} = \widetilde{A} \widetilde{A}^T, \quad \text{where} \quad \widetilde{A} \in \mathbb{R}^{T \times m_S} = [\Omega^{(1)} D^{(1)} … \Omega^{(S)} D^{(S)}].
  \]

Deep Hyperalignment

Cholesky Decomposition
DHA: Optimization

★ Objective Function can be reformulated as follows:

$$\min_{G, R^{(i)}, \theta^{(i)}} \sum_{i=1}^{S} \left\| G - f_i(X^{(i)}; \theta^{(i)}) R^{(i)} \right\| \equiv \max_{G} \left( \text{tr}(G^\top A G) \right).$$

★ So, we have:

$$A G = G \Lambda, \text{ where } \Lambda = \{\lambda_1 \ldots \lambda_T\}$$

$$\tilde{A} = G \tilde{\Sigma} \tilde{\Psi}^\top \text{ Incremental PCA}$$

★ DHA mappings can be calculated as follows:

$$R^{(\ell)} = \left( \left( f_\ell(X^{(\ell)}; \theta^{(\ell)}) \right)^\top f_\ell(X^{(\ell)}; \theta^{(\ell)}) + \epsilon I \right)^{-1} \left( f_\ell(X^{(\ell)}; \theta^{(\ell)}) \right)^\top G.$$
In order to use back-propagation algorithm for seeking an optimized parameters for the deep network, we also have:

\[
\frac{\partial Z}{\partial f_\ell(X^{(\ell)}; \theta^{(\ell)})} = 2R^{(\ell)}G^T - 2R^{(\ell)}(R^{(\ell)})^T \left(f_\ell(X^{(\ell)}; \theta^{(\ell)}) \right)^T.
\]

where

\[
Z = \sum_{\ell=1}^{T} \lambda_\ell
\]
Empirical Studies
## Datasets

Table S2: The datasets.

<table>
<thead>
<tr>
<th>Title</th>
<th>ID</th>
<th>S</th>
<th>K</th>
<th>T</th>
<th>V</th>
<th>X</th>
<th>Y</th>
<th>Z</th>
<th>Scanner</th>
<th>TR</th>
<th>TE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixed-gambles task</td>
<td>DS005</td>
<td>48</td>
<td>2</td>
<td>240</td>
<td>450</td>
<td>53</td>
<td>63</td>
<td>52</td>
<td>S 3T</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>Visual Object Recognition</td>
<td>DS105</td>
<td>71</td>
<td>8</td>
<td>121</td>
<td>1963</td>
<td>79</td>
<td>95</td>
<td>79</td>
<td>G 3T</td>
<td>2.5</td>
<td>30</td>
</tr>
<tr>
<td>Word and Object Processing</td>
<td>DS107</td>
<td>98</td>
<td>4</td>
<td>164</td>
<td>932</td>
<td>53</td>
<td>63</td>
<td>52</td>
<td>S 3T</td>
<td>2</td>
<td>28</td>
</tr>
<tr>
<td>Auditory and Visual Oddball</td>
<td>DS116</td>
<td>102</td>
<td>2</td>
<td>170</td>
<td>2532</td>
<td>53</td>
<td>63</td>
<td>40</td>
<td>P 3T</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>Multi-subject, multi-modal</td>
<td>DS117</td>
<td>171</td>
<td>2</td>
<td>210</td>
<td>524</td>
<td>64</td>
<td>61</td>
<td>33</td>
<td>S 3T</td>
<td>2</td>
<td>30</td>
</tr>
<tr>
<td>Forrest Gump</td>
<td>DS113</td>
<td>20</td>
<td>10</td>
<td>451</td>
<td>2400</td>
<td>160</td>
<td>160</td>
<td>36</td>
<td>S 7T</td>
<td>2.3</td>
<td>22</td>
</tr>
<tr>
<td>Raiders of the Lost Ark</td>
<td>N/A</td>
<td>10</td>
<td>7</td>
<td>924</td>
<td>980</td>
<td>78</td>
<td>78</td>
<td>54</td>
<td>S 3T</td>
<td>3</td>
<td>30</td>
</tr>
</tbody>
</table>

S is the number of subject; K denotes the number of stimulus categories; T is the number of scans in unites of TRs (Time of Repetition); V denotes the number of voxels in ROI; X, Y, Z are the size of 3D images; Scanners include S=Siemens, G = General Electric, and P = Philips in 3 Tesla or 7 Tesla; TR is Time of Repetition in millisecond; TE denotes Echo Time in second; Please see [openfmri.org](http://openfmri.org) for more information.
Simple Tasks Analysis

Table 1: Accuracy of HA methods in post-alignment classification by using simple task datasets

<table>
<thead>
<tr>
<th>Algorithms, Datasets</th>
<th>DS005</th>
<th>DS105</th>
<th>DS107</th>
<th>DS116</th>
<th>DS117</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu$-SVM [17]</td>
<td>71.65±0.97</td>
<td>22.89±1.02</td>
<td>38.84±0.82</td>
<td>67.26±1.99</td>
<td>73.32±1.67</td>
</tr>
<tr>
<td>HA [1]</td>
<td>81.27±0.59</td>
<td>30.03±0.87</td>
<td>43.01±0.56</td>
<td>74.23±1.40</td>
<td>77.93±0.29</td>
</tr>
<tr>
<td>RHA [2]</td>
<td>83.06±0.36</td>
<td>32.62±0.52</td>
<td>46.82±0.37</td>
<td>78.71±0.76</td>
<td>84.22±0.44</td>
</tr>
<tr>
<td>KHA [3]</td>
<td>85.29±0.49</td>
<td>37.14±0.91</td>
<td>52.69±0.69</td>
<td>78.03±0.89</td>
<td>83.32±0.41</td>
</tr>
<tr>
<td>SVD-HA [4]</td>
<td>90.82±1.23</td>
<td>40.21±0.83</td>
<td>59.54±0.99</td>
<td>81.56±0.54</td>
<td>95.62±0.83</td>
</tr>
<tr>
<td>SRM [5]</td>
<td>91.26±0.34</td>
<td>48.77±0.94</td>
<td>64.11±0.37</td>
<td>83.31±0.73</td>
<td>95.01±0.64</td>
</tr>
<tr>
<td>SL [9]</td>
<td>90.21±0.61</td>
<td>49.86±0.4</td>
<td>64.07±0.98</td>
<td>82.32±0.28</td>
<td>94.96±0.24</td>
</tr>
<tr>
<td>CAE [6]</td>
<td>94.25±0.76</td>
<td>54.52±0.80</td>
<td>72.16±0.43</td>
<td>91.49±0.67</td>
<td>95.92±0.67</td>
</tr>
<tr>
<td>DHA</td>
<td>97.92±0.82</td>
<td>60.39±0.68</td>
<td>73.05±0.63</td>
<td>90.28±0.71</td>
<td>97.99±0.94</td>
</tr>
</tbody>
</table>

Table 2: Area under the ROC curve (AUC) of different HA methods in post-alignment classification by using simple task datasets

<table>
<thead>
<tr>
<th>Algorithms, Datasets</th>
<th>DS005</th>
<th>DS105</th>
<th>DS107</th>
<th>DS116</th>
<th>DS117</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\nu$-SVM [17]</td>
<td>68.37±1.01</td>
<td>21.76±0.91</td>
<td>36.84±1.45</td>
<td>62.49±1.34</td>
<td>70.17±0.59</td>
</tr>
<tr>
<td>HA [1]</td>
<td>70.32±0.92</td>
<td>28.91±1.03</td>
<td>40.21±0.33</td>
<td>70.67±0.97</td>
<td>76.14±0.49</td>
</tr>
<tr>
<td>RHA [2]</td>
<td>82.22±0.42</td>
<td>30.35±0.39</td>
<td>43.63±0.61</td>
<td>76.34±0.45</td>
<td>81.54±0.92</td>
</tr>
<tr>
<td>KHA [3]</td>
<td>80.91±0.21</td>
<td>36.23±0.57</td>
<td>50.41±0.92</td>
<td>75.28±0.94</td>
<td>80.92±0.28</td>
</tr>
<tr>
<td>SVD-HA [4]</td>
<td>88.54±0.71</td>
<td>37.61±0.62</td>
<td>57.54±0.31</td>
<td>78.66±0.82</td>
<td>92.14±0.42</td>
</tr>
<tr>
<td>SRM [5]</td>
<td>90.23±0.74</td>
<td>44.48±0.75</td>
<td>62.41±0.72</td>
<td>79.20±0.98</td>
<td>93.65±0.93</td>
</tr>
<tr>
<td>SL [9]</td>
<td>89.79±0.25</td>
<td>47.32±0.92</td>
<td>61.84±0.32</td>
<td>80.63±0.81</td>
<td>93.26±0.72</td>
</tr>
<tr>
<td>CAE [6]</td>
<td>91.24±0.61</td>
<td>52.16±0.63</td>
<td>72.33±0.79</td>
<td>87.53±0.72</td>
<td>91.49±0.33</td>
</tr>
<tr>
<td>DHA</td>
<td>96.91±0.82</td>
<td>59.57±0.32</td>
<td>70.23±0.92</td>
<td>89.93±0.24</td>
<td>96.13±0.32</td>
</tr>
</tbody>
</table>
Complex Tasks Analysis

Figure 1: Comparison of different HA algorithms on complex task datasets by using ranked voxels.
Classification analysis by using feature selection

(A) DS105

(B) DS107
which can significantly reduce the time complexity of the optimization procedure [11, 13].

Performance of other techniques in fMRI analysis, e.g. Representational Similarity Analysis (RSA), to other state-of-the-art HA algorithms. In the future, we will plan to employ DHA for improving the studies on multi-subject fMRI analysis confirm that the DHA method achieves superior performance nonlinear function for each subject to transfer the brain activities to a common space. Experimental because the kernel in DHA is a multi-layer neural network, which can separately implement any alignment for a new subject. Further, DHA is not limited by a restricted fixed representational space fairly scales with data size and the training data is not referenced when DHA computes the functional rank-

Deep Hyperalignment (DHA) can handle fMRI functional alignment in multi-subject fMRI analysis. Deep Hyperalignment (DHA) can handle fMRI because it concurrently employs modified versions of SRM and SL for functional alignment. Further, methods, where runtime of other methods are scaled based on the DHA (runtime of the proposed generated by a PC with certain specifications. Figure 3 illustrates the runtime of the mentioned methods because it provides better feature representation in comparison with other techniques.

Figure 2: Classification by using feature selection.

Figure 2: Classification by using feature selection.

Clausen et al. [70 TRs as previous studies [1, 4, 7, 9]. Then, the experiments are repeated by using the different number of voxels in ROI are ranked by the method proposed in [1] based on their neurological priorities same by using

Deep Hyperalignment

Run Time Analysis

Run Time Analysis

(A) DS105

(B) DS107

Deep Hyperalignment
Future
Works
Future Works

- This paper extended a deep approach for hyperalignment methods in order to provide accurate functional alignment in multi-subject fMRI analysis.

- Deep Hyperalignment (DHA) can handle fMRI datasets with nonlinearity, high-dimensionality (broad ROI), and a large number of subjects. Further, its time complexity fairly scales with data size and the training data is not referenced when DHA computes the functional alignment for a new subject.

- In the future, we will plan to employ DHA for improving the performance of other techniques in fMRI analysis, e.g. Representational Similarity Analysis (RSA).
Thank You!

Q & A

For more details, contact:

myousefnezhad@nuaa.edu.cn
myousefnezhad@outlook.com
https://myousefnezhad.github.io/