Alzheimer’s Disease Detection Using A Projection Based Learning Meta-cognitive RBF Network

G. Sateesh Babu\textsuperscript{1}, S. Suresh\textsuperscript{1} and B. S. Mahanand\textsuperscript{2}

\textsuperscript{1} School of Computer Engineering  
Nanyang Technological University, Singapore

\textsuperscript{2} Department of Information Science and Engineering  
Sri Jayachamarajendra College of Engineering, Mysore, India

Presented at International Joint Conference on Neural Networks, (IJCNN 2012), Brisbane
Outline

1 Introduction
   - Motivation
   - Objective

2 Extraction
   - VBM analysis

3 Reduction
   - ICA

4 PBL-McRBFN

5 Experiments
   - MRI Data
   - Setup
   - VBM Results
   - Classification

6 Conclusion
Motivation

• Alzheimer’s disease (AD) is main threat to public health
• Approximately 30 million AD patients worldwide
  • 5.3 million Americans
  • 3.7 million Indians
• AD is a progressive, degenerative disease that leads to memory loss, poor judgement and problems in learning
• Only way of diagnosing AD definitely is by post-mortem examinations of the brain
• Early detection of AD using magnetic resonance imaging (MRI) and machine learning methods is a promising area of research
Objectives of the study

- Detection of Alzheimer’s disease from magnetic resonance images using voxel-based morphometry and projection based learning meta-cognitive radial basis function network.
Feature extraction using VBM

- Voxel-based morphometry (VBM) is a fast and fully automated method for voxel-wise comparison of gray matter tissue concentration between normal persons and patients.
- Steps involved in the VBM analysis:
  - Unified Segmentation
  - Smoothing
  - Statistical Analysis
- VBM detected voxel locations of the significant areas are used to extract features from MRI.
Voxel-based morphometry

J. Ashburner et. al,

Voxel-based
morphometry,

Neuroimage, vol. 11,
no. 6, pp. 805-821,
2000.
Feature reduction using ICA

- Independent Component Analysis (ICA), as one of important techniques of blind signal separation, has been shown to provide a powerful method for neuroimaging data.

- One of multivariate data-driven techniques that enable an exploratory analysis of MRI data sets to provide useful information about the relationship between voxels in local substructures of the brain.

- ICA finds a transformation in which the components of transformed data are statistically as independent from each other as possible.

- The main idea is to apply ICA on VBM detected morphometric features to reduce the number of features which are later used as input to classifier.
Meta-cognitive Radial Basis Function Network

**Definition**

A RBF network which employs human meta-cognitive principles

---

**Key learning principles in McRBFN**

- **Sample Deletion**: Sample deleted without being used for learning
- **Sample Learning**: Sample used to grow network/update parameters
- **Sample Reserve**: Sample pushed to the back of the sample pool
Meta-cognitive Radial Basis Function Network: Cognitive Component

Input layer: \( m \) nodes
Hidden layer: \( K \) nodes
Output layer: \( n \) nodes

Output layer: \( W: K \times n \)

\[
\begin{align*}
\mathbf{x}^t &= [x_1^t, \ldots, x_m^t]^T \in \mathbb{R}^m \quad \text{input of the } t^{th} \text{ sample} \\
h_k^t &= \exp \left( - \frac{||x^t - \mu_k^l||^2}{(\sigma_k^l)^2} \right) \\
\hat{y}_j^t &= \sum_{k=1}^{K} w_{kj} h_k^t
\end{align*}
\]
Meta-cognitive Radial Basis Function Network: Cognitive Component

Projection Based Learning Algorithm

- The considered energy function:
  \[ J(W) = \frac{1}{2} \sum_{i=1}^{t} \sum_{j=1}^{n} \left( y^i_j - \sum_{k=1}^{K} w^i_k h^i_k \right)^2 \]
- Optimum solution:
  \[ W^* = \arg \min_{W \in \mathbb{R}^{K \times n}} J(W) \]
  \[ \sum_{k=1}^{K} a_{kp} w^i_k = b^i_p, \ p = 1, \ldots, K; \ j = 1, \ldots, n \]
- Matrix form: \( AW = B \)
- Projection matrix (\( A \in \mathbb{R}^{K \times K} \)):
  \[ a_{kp} = \sum_{i=1}^{t} h^i_k h^i_p, \ k = 1, \ldots, K; \ p = 1, \ldots, K \]
- Output matrix (\( B \in \mathbb{R}^{K \times n} \)):
  \[ b_{pj} = \sum_{i=1}^{t} h^i_p y^i_j, \ p = 1, \ldots, K; \ j = 1, \ldots, n \]
- Output weight Matrix: \( W^* = A^{-1}B \)
**Meta-cognitive Radial Basis Function Network:**

**Meta-cognitive Component**

**Monitors** novelty and error of cognitive component

**Control** *what-to-learn, when-to-learn and how-to-learn*

**Monitory Signals**

- **Novelty:** Class-wise Significance
  - Average distance between current sample and existing neurons of same class: 
    \[ \psi_c = \frac{1}{K^c} \sum_{k=1}^{K^c} h(x^t, \mu_k^c) \]

- **Error:** Maximum Hinge Error \((E^t) = \max_{j \in 1,2,\ldots,n} |e^t_j|\)
  
  \[ e^t_j = \begin{cases} 
  0 & \text{if } y_j^t \hat{y}_j^t > 1 \\
  y_j^t - \hat{y}_j^t & \text{otherwise} 
  \end{cases} \quad j = 1, 2, \ldots, n \]
Meta-cognitive Radial Basis Function Network: Meta-cognitive Component

Control Signals

- **What-to-learn**: Sample deletion strategy
  - IF $c^t == \hat{c}^t$ AND $E^t \leq \beta_d$ THEN
  - Delete the sample from data stream

- **How-to-learn**: Sample learning strategy
  - Add new neurons
  - Update output weights of neurons
  - Update self-regulatory thresholds

- **When-to-learn**: Sample reserve strategy
  - If what-to-learn and how-to-learn fails
Meta-cognitive Radial Basis Function Network: Meta-cognitive Component how-to-learn

Neuron Growth: IF $\left( \hat{c}^t \not= c^t \text{ OR } E^t \geq \beta_a \right)$ AND $\psi_c(x^t) \leq \beta_c$

- **No-overlapping with any class:** $(d_S \gg \sigma_{nrs}^c \text{ AND } d_I \gg \sigma_{nri}^l)$
  
  $\mu_{K+1}^c = x^t; \quad \sigma_{K+1}^c = \kappa \sqrt{x^T x}$

- **No-overlapping with the inter-class:** $(d_S \not< \frac{d_S}{d_I} < 1)$
  
  $\mu_{K+1}^c = x^t; \quad \sigma_{K+1}^c = \kappa \|x^t - \mu_{nrs}^c\|$  

- **Minimum Overlapping with the inter-class:** $(1 < \frac{d_S}{d_I} < 1.5)$
  
  $\mu_{K+1}^c = x^t + \zeta (\mu_{nrs}^c - \mu_{nri}^l)$
  
  $\sigma_{K+1}^c = \kappa \|\mu_{K+1}^c - \mu_{nrs}^c\|$
Neuron Growth: Output weight allocation and adaptation

- The size of the matrix $A$ is increased from $K \times K$ to $(K + 1) \times (K + 1)$

$$A_{(K+1) \times (K+1)} = \begin{bmatrix} A_{K \times K} + (h^T)^T h^T & a_{K+1}^T \\ a_{K+1} & a_{K+1, K+1} \end{bmatrix}$$

where $h^T = [h_1^T, h_2^T, \ldots, h_K^T]$, $a_{K+1} \in \mathbb{R}^{1 \times K} = \sum_{i=1}^{t} h_{i}^{K+1} h_{p}^{i}$, $p = 1, \ldots, K$ and $a_{K+1, K+1} \in \mathbb{R}^{+} = \sum_{i=1}^{t} h_{i}^{K+1} h_{K+1}^{i}$

- The size of matrix $B$ is increased from $K \times n$ to $(K + 1) \times n$

$$B_{(K+1) \times n} = \begin{bmatrix} B_{K \times n} + (h^T)^T (y^T)^T \\ b_{K+1} \end{bmatrix}$$

where $b_{K+1} \in \mathbb{R}^{1 \times n} = \sum_{i=1}^{t} h_{i}^{K+1} y_{j}^{i}$, $j = 1, \ldots, n$

- Output weights are estimated as

$$\begin{bmatrix} \mathbf{W}_K \\ \mathbf{w}_{K+1} \end{bmatrix} = \left( A_{(K+1) \times (K+1)} \right)^{-1} B_{(K+1) \times n}$$

- Update $\beta_a := \delta \beta_a + (1 - \delta) E^t$
Meta-cognitive Radial Basis Function Network: Meta-cognitive Component how-to-learn

**Neuron Growth: Output weight allocation and adaptation**

- After calculating inverse of the matrix $A_{(K+1)\times(K+1)}$ recursively using matrix identities, the resultant equations are

$$\mathbf{W}_K = \left[\mathbf{I}_{K\times K} + \frac{(A_{K\times K})^{-1} a_{K+1} a_{K+1}}{\Delta} \right] \left[ \mathbf{W}_K + (A_{K\times K})^{-1} (h^t)^T (y^T)^T \right] - \frac{(A_{K\times K})^{-1} a_{K+1} b_{K+1}}{\Delta}$$

$$\mathbf{w}_{K+1} = -\frac{a_{K+1}}{\Delta} \left( \mathbf{W}_K + (A_{K\times K})^{-1} (h^t)^T (y^T)^T \right) + \frac{b_{K+1}}{\Delta} \Delta$$

where $\Delta = a_{K+1} = a_{K+1} a_{K+1} - a_{K+1} (A_{K\times K} + (h^T) (h^T)^{-1} a_{K+1})$

**Parameter Update:** IF $c^t = \hat{c}^t$ AND $E^t \geq \beta_u$

- Update the weights of existing neurons using projection based learning

$$\mathbf{A} = \mathbf{A} + (h^T)^T h^t; \quad \mathbf{B} = \mathbf{B} + (h^T)^T (y^T)^T$$

$$\mathbf{W}_K = \mathbf{W}_K + A^{-1} (h^T)^T (e^T)^T$$

- Update $\beta_u := \delta \beta_u + (1 - \delta) E^t$
1. Predict the output
2. Update the McRBFN knowledge base measures
3. Select the learning strategy
   - Sample Deletion
   - Sample Learning
   - Sample Reserve
4. Go to the next sample

McRBFN Schematic diagram
Open Access Series of Imaging Studies (OASIS) database has been used in this study (http://www.oasisbrains.org).

Whole brain T1-weighted 3D MPRAGE (Magnetization-prepared Rapid Acquisition Gradient Echo) volumes were acquired on a Siemens 1.5T scanner. The acquired volumes had 128 sagittal 1.25 mm slices without gaps and pixel resolution of $256 \times 256 (1 \times 1 \text{mm})$.

The demographic and dementia details are:

<table>
<thead>
<tr>
<th>Group</th>
<th>Healthy Persons</th>
<th>AD Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Persons</td>
<td>98</td>
<td>100</td>
</tr>
<tr>
<td>Percentage of Male</td>
<td>26.5%</td>
<td>41.0%</td>
</tr>
<tr>
<td>Age(mean±std)</td>
<td>75.92±8.99</td>
<td>74.76±7.12</td>
</tr>
<tr>
<td>MMSE(mean±std)</td>
<td>28.96±1.21</td>
<td>24.32±4.17</td>
</tr>
<tr>
<td>No. of persons</td>
<td></td>
<td></td>
</tr>
<tr>
<td>with CDR 0/0.5/1/2</td>
<td>98/0/0/0</td>
<td>0/70/28/2</td>
</tr>
</tbody>
</table>
Experimental setup

- VBM analysis using the Statistical Parametric Map (SPM) software package
- The complete data set consists of 198 samples with 19879 features
- ICA analysis performed using FastICA algorithm
- 19879 VBM detected morphometric features reduced to different combinations of 5, 10, 20, 30, 40, 50, 100, 150, 200, 250, 294 features using FastICA algorithm
- 50% samples are randomly chosen for training and remaining for testing at each trial
- The performance of PBL-McRBFN classifier is studied by generating 10 random trails on the training and testing sets
VBM analysis results

• Significant areas with increased gray matter density in the normal persons relative to AD patients overlaid on the standard template in (a) Sagittal view (b) Coronal view (c) Axial view.
VBM analysis results

- Maximum intensity projections of the significant areas with increased gray matter density in the healthy persons relative to the AD patients in (a) Sagittal view (b) Coronal view (c) Axial view
### PBL-McRBFN Classifier Results

<table>
<thead>
<tr>
<th>No. of Features</th>
<th>Training Efficiency</th>
<th>Testing Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>STD</td>
</tr>
<tr>
<td>5</td>
<td>92.85</td>
<td>2.60</td>
</tr>
<tr>
<td>10</td>
<td>95.30</td>
<td>2.34</td>
</tr>
<tr>
<td>20</td>
<td>95.22</td>
<td>1.79</td>
</tr>
<tr>
<td>30</td>
<td>95.71</td>
<td>1.11</td>
</tr>
<tr>
<td>40</td>
<td>92.44</td>
<td>1.85</td>
</tr>
<tr>
<td>50</td>
<td>95.51</td>
<td>1.98</td>
</tr>
<tr>
<td>100</td>
<td>94.69</td>
<td>3.49</td>
</tr>
<tr>
<td>150</td>
<td>93.67</td>
<td>3.33</td>
</tr>
<tr>
<td>200</td>
<td>94.28</td>
<td>2.11</td>
</tr>
<tr>
<td>250</td>
<td>95.91</td>
<td>2.97</td>
</tr>
<tr>
<td>294</td>
<td>92.04</td>
<td>2.43</td>
</tr>
<tr>
<td>19879</td>
<td>93.26</td>
<td>2.66</td>
</tr>
</tbody>
</table>
Performance comparison

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>No. of Features</th>
<th>Classification Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>VBM + ICA + PBL-McRBFN</td>
<td>50</td>
<td>71.11</td>
</tr>
<tr>
<td>VBM + ICA + PBL-McRBFN</td>
<td>100</td>
<td>70.14</td>
</tr>
<tr>
<td>VBM + PBL-McRBFN</td>
<td>19879</td>
<td>77.56</td>
</tr>
<tr>
<td>VBM + PCA + SVM</td>
<td>-</td>
<td>66.9</td>
</tr>
<tr>
<td>VBM + IPCA + SVM</td>
<td>-</td>
<td>69.7</td>
</tr>
<tr>
<td>VBM + ICA + SVM</td>
<td>200</td>
<td>62.8</td>
</tr>
</tbody>
</table>
Conclusion

- Proposed a new approach for AD detection using PBL-McRBFN classifier based on VBM detected and ICA reduced features from MRI.
- The study shows that accurate classification of AD subjects can be performed with ICA reduced features or complete features using VBM-PBL-McRBFN classifier.
- PBL-McRBFN accuracy on complete VBM detected features set is 6 – 7% more than ICA reduced features sets, this is due to data consists of 100 ‘very mild to moderate AD’ patients and 98 healthy elder persons. Hence, PBL-McRBFN classifier requires all morphometric features to classify AD patients group consists of wide range of CDR from 0.5 – 2 and healthy elder persons.
- The performance comparison with the well-known SVM based classifiers in the literature clearly indicated the superior performance of the proposed approach.
Thank you