Age-related Classification and Prediction Based on MRI: A Sparse Representation Method

Longfei Su, Lubin Wang, Hui Shen, Dewen Hu*

College of Mechatronics and Automation, National University of Defense Technology, Changsha, Hunan 410073, PR China

Abstract

Through analysis of structural magnetic resonance imaging (MRI) images, classification and prediction of the age of adults (19-79) were implemented to make analysis of the age-related changes of the grey matter (GM) concentration. Due to the distributed nature of the aging spatial pattern in human brain, a multivariate voxel selection method based on sparse representation was introduced to identify the most discriminative brain regions. It can effectively pick out the isolating voxels as well as the clustered voxels which contribute enormously to the classification and age prediction. By using this multivariate voxel selection method, the binary classification can get a higher accuracy compared to univariate voxel selection method. Age prediction of all the subjects via sparse representation (SR) was carried out in our study. Four different models were used to fit the predicted age of all subjects in maturity index (MI) space. Difference trend of the brain development between the senior and the junior was observed. That the development or decline of GM of the senior over 60 accelerates was found in our study.

© 2011 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of the Asia-Pacific Chemical, Biological & Environmental Engineering Society (APCBEES) Open access under CC BY-NC-ND license.

Keywords: MRI; classification; adults; prediction; age; GM concentration; sparse representation;

1. Introduction

It is well known that aging is followed by neuroanatomical changes in the cerebral cortex. These changes include changes of the structural connectivity based on white matter (WM) and changes of the GM concentration. That Reorganization of anatomical connectivity based on the change of white matter (WM) within cerebral cortex is contributable to aging has already been confirmed [1]. Machine learning
based approach according to approximate structural connectivity patterns which were extracted from diffusion magnetic resonance images has also been documented [2]. Predicting individual brain maturity by using only 5 minutes of resting state functional connectivity magnetic resonance (fcMRI) has made significant progress recently [3]. Substantial correspondence between resting state functional connectivity and structural connectivity measured in the same subjects also has been observed [4]. However, age effects on the GM concentration have many different even controversial conclusions. In this study, we have introduced a voxel selection approach to improve binary classification and age prediction about individuals based on the GM concentration map.

Activity levels of individual voxels are known to correlate highly with each other, and strong evidence for the distributed nature of the pattern representation in the brain has been documented [5]. The distributed nature of the aging spatial patterns in the brain makes the use of multivariate pattern analysis (MVPA) necessary. Voxel selection is critical for classification and prediction based on MRI. Reducing the dimension of the GM concentration data can significantly improve the performance of the classifier [6][7]. If correlation among voxels carries important information, the individual voxels chosen by univariate approach could be suboptimal [8]. A multivariate voxels selection algorithm based on sparse representation was introduced to select the voxels which changed due to aging. Due to distributed nature of aging spatial pattern in the brain, this algorithm is especially fit for voxel selection from GM concentration map. Furthermore, this voxels selection algorithm was evaluated by the performance of classification and prediction.

2. Material and Methods

2.1. Participants

Eighty four healthy volunteers participated in this study (43 females: 43.3±19.3 years old, range 19-79 and 41 males: 44.7±15.4 years old, range 19-73). Two subjects were ambidextrous, four were left-handed and the rest were right-handed. All subjects were recruited for the International Consortium of Brain Mapping (ICBM) dataset at Montreal Neurological Institute (MNI) and have no history of neurological and psychiatric disorders. All scans were performed on the same Siemens Sonata 1.5 T MRI scanner.

2.2. Imaging Protocol

T1-weighted structural MRI was acquired with the following parameters: Common parameters were TR=22 ms, TE=9.2 ms, slice thickness=1 mm, flip angle=30°, FOV=256×256 mm², and in-plane resolution=256×256. All data is part of the 1000 Functional Connectomes project (http://icon_1000.projects.nitrc.org/fcpClassic/FcpTable.html). There are 86 subjects’ data on the website and every data consist of three resting-state functional magnetic resonance imaging (fMRI) and one T1-weighted structural MRI images, only T1-weighted structural MRI images were used for classification and age prediction. For one subject has one fMRI not integrated, and one subject is too old for modulated, the number of subjects in this study is reduced to 84.

2.3. Data preprocessing

In order to obtain the GM concentration map, structural MRI images were preprocessed using SPM8 toolbox (http://www.fil.ion.ucl.ac.uk/spm/) and the VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm8/). First, structural MRI data were spatially normalized to the standard space of Talairach and Tournoux; Second, the every T1-weighted structural MRI were segmented into three partitions-GM, WM and
cerebro-spinal fluid (CSF); Third, the normalized GM images with modulation were then smoothed with an isotropic Gaussian filter of 8 mm full-width half-maximum kernel to obtain GM density map. We used voxels with GM density value greater than 0.05 as original inputs for classification to ensure that voxels selected in the GM partition is informative.

2.4. Dimension reduction and feature selection

A total of 84 subjects were enrolled in the studies, we give subjects who is younger than 35 a label (1) as junior group, give subjects older than 55 another label (-1) as senior group. The junior group consists of 32 subjects; the senior group consists of 27 subjects.

For some reasons, such as noise, inter-individual anatomical differences, only a small number of the voxels are highly discriminative [9]. Voxel selection is thus necessary to transform this high-dimensional dataset to a low-dimensional space, which decreased the computational complexity and removed the redundancy.

Here, a Multivariate method based on sparse representation was used for voxel selection [10]. The model of SR method can be described as bellow:

\[ y = Aw \]

Here, \( y \in \mathbb{R}^n \) is a given class label, when it is used for binary classification, every element of \( y \) is either 1 or -1. Also, when it is used for regression, \( y \) can be other measurement, for example age. \( A \in \mathbb{R}^{n \times m} \) is a basis matrix of which each column is the same voxels of different subjects, each row is all the voxels of the same subject. \( w \in \mathbb{R}^m \) is unknown, the object of this sparse representation is to find the weight vector \( w \) which is satisfy with function (1), and as sparse as possible.

Consider of the following optimization problem:

\[
\begin{align*}
\min & \|w\|_0, \\
\text{subject to} & \quad Aw = y,
\end{align*}
\]

(1)

The number of nonzero of a vector is defined as 0-norm, which is the sparsest solution of (1). But, it is very hard to get seek the solution of (2). Instead, we turn to an alternative optimization problem:

\[
\begin{align*}
\min & \|w\|_1, \\
\text{subject to} & \quad Aw = y,
\end{align*}
\]

(3)

Sum of the absolute value of a vector is defined as 1-norm. The problem of (3) is a linear programming problem which can be solved by MATLAB easily. Although the solution of (3) is often not the sparsest and not the same as the solution of (2), under some conditions the two solutions can be viewed as equivalent. Sufficient condition for the two sparse solutions are equal is too hard to satisfy, so a new method based on probability was adopted. The two solution can be seen the same under a high probability (e.g. 0.95), then we can use the solution of (3) instead of the solution of (2).

In order to facilitate the solving of the optimization problem of (3), we can transform it into another format as bellow: define new nonnegative variable \( u \) and \( v \), \( u - v = w \), and \( u, v \in \mathbb{R}^m \),

\[
\begin{align*}
\min & \sum_{i=1}^{m} (u_i + v_i), \\
\text{subject to} & \quad [A, -A][u^T, v^T] = y \quad \text{and} \quad u \geq 0, v \geq 0,
\end{align*}
\]

(4)
Now, the problem is converted to a typical nonnegative linear programming problem, under the MATLAB we can easily get the u, v, then w. The value of the element of the weight vector w represents the weight of the corresponding column of A for classification or regression.

2.5. Pattern classification

Classification is performed on feature vectors after voxel selection. Classification of subjects into two classes (junior vs. senior) is a task of learning a model from the training data, and then tested on a new set of samples to predict the samples’ group. Support Vector Machine (SVM) belongs to a learning system based on recent advances in statistical learning theory, seeking the largest margin hyperplane and minimizing the structural risk. It works well when the number of training samples is small, so it has a superior generate ability [11].

2.6. Classifier performance and discriminative pattern

Because of lack of samples, leave one out cross validation (LOOCV) was used in this study to estimate the performance of classifier. Here each subject was viewed as test sample by turns and all the rest samples were viewed as training samples So, we had to do cross validation as many times as the number of the samples. In every LOOCV round, the discrimination function computed from the training sample was applied to the test sample to make a classification or prediction. After all LOOCV rounds were completed, we could get the average accuracy of the classification or prediction.

In our study, a univariate voxel selection method was also referred as the contrast. Ranking by the score of Pearson correlation coefficient (PCC) was commonly used in dimension reduction and feature selection [3]. Due to the reasons of huge feature dimensions of the original input and computation complexity of the SR, before we made use of SR for voxel selection, ranking by Pearson correlation coefficient was first used to reduce the feature dimensions to 4000, then the SR algorithm was applied to the 4000 dimensional data, finally the dimensions were reduced to 1000 which were used for classification. As contrast, the first 1000 dimensions ranked by the score of Pearson correlation coefficient were also used for classification.

The performance of a classifier can be measured by Generalization Rate (GR), Sensitivity (SS) and Specificity (SC). Here Sensitivity SS is defined as the proportion of junior subjects correctly predicted, while Specificity SC represents the proportion of senior subjects correctly predicted. The proportion of all subjects correctly predicted is evaluated by Generalization Rate GR.

3. Results

3.1. Accuracy of Classification

For SVM has many parameters that can be adjusted, now values of the main parameters are given bellow: C=1, p1=5, and kernel function was setting to linear kernel function. As two voxel selection methods were adopted to the original input data before classification, thus two classification results were obtained for each voxel selection method, corresponding to the PCC+SR+SVM and PCC+SVM, respectively.

The number of selected features for classification was started from 100 features, each time another 100 features were added to the prior features for classification until the number of all features for classification achieves 1000. Before taking linear kernel SVM for classification, the PCC+SR and PCC
were applied for dimension reduction. Relationship of the GR and the number of features was shown in Fig. 1.

![Fig.2 Relationship between classification accuracy and number of features](image)

From Fig. 1, we can see that the SR method can take as few voxels as possible to get a higher accuracy of classification. PCC+SR+SVM can reach a high GR when only 200 voxels were used. But PCC+SVM had to use at least 600 voxels to get a high GR but lack of stability.

In consistent with previous research about age classification, we rearranged the subjects. The subjects aged from 19 to 30 constitute the junior group (24 subjects); the subjects aged from 60 to 79 constitute the senior group (17 subjects) [2]. Under this arrangement, with only 100 voxels retained for classification, the SVC having linear kernel function, we can get better classification accuracies. The classification procedure is the same as previous. Only different is the constitution of the two groups. For the three methods of voxel selection, the accuracies were shown in table. 1.

<table>
<thead>
<tr>
<th>Method</th>
<th>SC</th>
<th>SS</th>
<th>GR</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCC+SR+SVC</td>
<td>95.8%</td>
<td>100%</td>
<td>97.6%</td>
</tr>
<tr>
<td>PCC+SVC</td>
<td>91.7%</td>
<td>76.5%</td>
<td>85.4%</td>
</tr>
</tbody>
</table>

From table. 1 we can see more clearly that the GR of PCC+SR+SVC is apparently higher than that of PCC+SVC. It indicates that the SR method is suitable for voxels selection. On the other hand, this is also evidence that the voxels which can reflect the age of subjects is not only partially clustered, but also distribute in the whole brain.

The GR of PCC+SVC which do not consider the correlation between the voxels was comparably low. Individual voxels are known to correlate highly with each other, so when the number of selected voxels is small, the selected features with the highest univariate score may not be able to include all of the relevant voxels. But the PCC+SR+SVC method can efficiently find the typical representation of aging voxels.

If the linear kernel function of the SVC was replaced by a nonlinear kernel function, the GR can reach an even higher level. For example, when a polynomial kernel function is used in the PCC+SR+SVC, the GR can reach 100%. So there is biomarker which can label the subjects’ group in our cerebral cortex. The GM concentration can reflect the age of one subject. There is age information indeed in our brain.

### 3.2. Age Prediction
There were 84 subjects available. All subjects were used for age prediction. Also, LOOCV was used to estimate the age prediction accuracy. Every subject was selected as the test group for once, the rest were served as the training data. So the training process repeated 84 times. The chronological age was used as training measure in every LOOCV fold. Then the PCC+SR+SVR was applied to predict the age of every subject. After PCC, 4000 voxels were retained for SR, during SR, 3400 voxels were eliminated, and at last 600 voxels were left for age prediction. The parameters of the SVR were set the same as previous. At last a predicted age was obtained for every subject.

Thus, this predicted age represents the maturity of the cerebral cortex of the corresponding subject. The mean predicted age of the subjects aged from 35 to 55 was defined as 1, and then the predicted age of all the subjects can be converted to a new maturity index (MI). The MI here represents 600 voxels that best reflect the brain maturity. The relationship of the MI and chronological age is shown in fig.2. (Red circles).

Curve fitting was carried out to show the process of the brain maturity clearly. Polynomials of degree n (n=2, 3, 4) were all used to fit the predicted age. When n was set equal to 2, the curve was a quadratic model; when n=3, the curve is cubic model, when n=4, the curve is quartic model. As a contrast, Von Bertalanffy growth curve was also introduced to fit the data. So, in Fig.2, four models run at the predicted age of all the subjects. The parameters of the four models were all shown in table.2

Table.2 Parameters of the four fitting models

<table>
<thead>
<tr>
<th>Model</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ax^2+bx+c</td>
<td>-6.01 x 10^-7</td>
<td>0.0211</td>
<td>0.18</td>
<td>—</td>
<td>—</td>
<td>0.6702</td>
</tr>
<tr>
<td>ax^3+bx^2+cx+d</td>
<td>7.57 x 10^-6</td>
<td>-0.0011</td>
<td>0.068</td>
<td>-0.42</td>
<td>—</td>
<td>0.6565</td>
</tr>
<tr>
<td>ax^4+bx^3+cx^2+dx+e</td>
<td>-4.26 x 10^-7</td>
<td>8.97 x 10^-5</td>
<td>-0.007</td>
<td>0.23</td>
<td>-1.93</td>
<td>0.6929</td>
</tr>
<tr>
<td>a(1-e^-bx)</td>
<td>1.5</td>
<td>0.025</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.6859</td>
</tr>
</tbody>
</table>

Akaike Information Criterion (AIC) is a method for model selection which can evaluate whether the curve fit the data. It can be derived as bellow:

\[ AIC = \ln(\sigma^2) + \frac{2m}{T} \]  \hspace{1cm} (5)
where \( m \) is the number of parameters in the model, \( T \) is the number of the samples, here \( T=84 \), \( s^2 \) is the estimated residual variance: 
\[
\hat{s}^2 = \frac{\text{sum of squared residuals for the model}}{T}.
\]
That is, the average squared residual for model.

The criterion may be minimized over choices of \( m \) to form a tradeoff between the fit of the model (which minimizes the sum of squared residuals) and the model’s complexity, which is measured by \( m \). The curve best fit the data can get the minimum value of AIC. So, AIC values for different curves in Fig.2 are also listed in table.2

From table.2, we can see that the AIC of the cubic model is the lowest of the four models. Because given a set of candidate models for the data, the preferred model is the one with the minimum AIC value. So the cubic model is best fit the predicted age data. Thus, two dash lines parallel to the polynomial of degree 3 were plotted in Fig.2 between which there are 95% of all the predicted age points. The Von Bertalanffy fits the predicted age of the subjects aged from 7 to 30 best based on fMRI [3], but here it is not the best.

4. Discussion

4.1. Distributed nature of the aging patterns

Distributed aging spatial patterns based on MRI data have already been proved by researchers, which can discriminate subjects of one group from another effectively [8]. According to Fig.1, the distinctiveness in the GM of the junior and the senior is not simply due to voxels that most correlate with the class label. Compared with PCC, voxels selected by PCC+SR is fewer, but the accuracy of the classification is higher. If the voxels in the same brain region which help to improve the classifier performance, only a small part of representative voxels are selected by SR method. But these voxels constitute a pattern which best represent the class information in the GM density map. So, the same as activity patterns based on fMRI data, the structural representative GM patterns also has a distributed nature.

For a better vision of the distribution, we increase the number of retained voxels to 600 which also have a good result according to Fig.1. The distribution was shown in Fig.3

![Fig.3 Typical voxels selected by the SR for age classification](image)

From Fig.3, we can see that the selected voxels were mostly in cingulate cortex, medial frontal cortex and part of cerebella. This is correspondence with previous research. Changes of Cortical thickness of the cingulate cortex and medial frontal cortex were referred previously [12]. Discovery of our research is evidence that the GM concentration of cingulate cortex and medial frontal cortex has a big difference between the junior and the senior. But, the SR method only selected a small part of voxels in the correlated region for classification. To pick out all the voxels is a difficulty for SR method.
4.2. Trend of brain maturation

Age prediction is the other task of our research. Through converting the predicted age to the MI space, we estimate the development trend of human brain. Four models were run at the predicted age in MI space and AIC was used to evaluate the fitness of the four models. The cubic model has got the lowest value of AIC, which prove it is the best model for the predicted age. According to Fig.2, from 19 to 30, differential coefficient of the curve decreases. It is corresponding with the maturity level of the human brain predicted based on fcMRI [3]. Between 30 and 60, differential coefficient of the curve has not significant change, just like a linear model. But, when the age is over 60, differential coefficient of the curve increases. A speed development or decline of the brain cortex can be deduced. No surprisingly, growth curve like Von Bertalanffy curve does not fit the data better. This fact indicates that decline of GM in brain accelerates when one is over 60.

5. Acknowledgments

This work was supported by the National 973 program of China (2011CB707802) and the National Natural Science Foundation of China (61003202).

References


