Journal of Neuroscience Methods 221 (2014) 78–84

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

Computational Neuroscience

Application of fused lasso logistic regression to the study of corpus callosum thickness in early Alzheimer's disease

Sang H. Lee[a,*], Donghyeon Yu[b], Alvin H. Bachman[a], Johan Lim[b], Babak A. Ardekani[a]

[a] The Nathan S. Kline Institute for Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA
[b] Department of Statistics, Seoul National University, Seoul, South Korea

HIGHLIGHTS

- Development of an algorithm for fused lasso logistic regression (FLLR) model.
- We applied FLLR to classify Alzheimer disease (AD) subjects from normal controls.
- We built a classifier based on corpus callosum (CC) thickness profiles.
- Classifier accuracy was estimated to be 84% based on five-fold cross-validation.
- Regions in the genu and splenium differentiated between groups were identified.

ARTICLE INFO

Article history:
Received 26 July 2013
Received in revised form
19 September 2013
Accepted 25 September 2013

Keywords:
Alzheimer's disease
Brain
Corpus callosum
Fused lasso
Logistic regression
MRI

ABSTRACT

We propose a fused lasso logistic regression to analyze callosal thickness profiles. The fused lasso regression imposes penalties on both the l_1-norm of the model coefficients and their successive differences, and finds only a small number of non-zero coefficients which are locally constant. An iterative method of solving logistic regression with fused lasso regularization is proposed to make this a practical procedure. In this study we analyzed callosal thickness profiles sampled at 100 equal intervals between the rostrum and the splenium. The method was applied to corpora callosa of elderly normal controls (NCs) and patients with very mild or mild Alzheimer's disease (AD) from the Open Access Series of Imaging Studies (OASIS) database. We found specific locations in the genu and splenium of AD patients that are proportionally thinner than those of NCs. Callosal thickness in these regions combined with the Mini Mental State Examination scores differentiated AD from NC with 84% accuracy.

1. Introduction

Alzheimer's disease (AD) is an age-related, irreversible neurodegenerative disease that develops gradually and results in episodic memory loss and behavioral and personality changes. There is great interest in finding neuroimaging based markers to predict the progression of AD, which could be very helpful in managing care for individual patients, and in AD research to develop interventions for altering the course of disease. Gray matter, especially the medial temporal lobe (MTL) and hippocampus, is a major focus in these studies (see Risacher and Saykin (2013) for a comprehensive review), but it is hard to reliably measure the MTL and hippocampus automatically. On the other hand, a number of studies have reported structural changes in the corpus callosum (CC) in AD (Di Paola et al., 2010; Frederiksen et al., 2011; Teipel et al., 2002; Wang et al., 2006), and there are fully automated procedures to measure the CC with good accuracy (Ardekani et al., 2012). In this paper, we attempt to identify the regions of the CC whose thickness could differentiate patients with early AD from normal controls (NCs) and use them in a fused lasso logistic regression classifier for early AD.

Two major approaches have been used to localize the differences in the CC between two populations: (1) defining a small number of geometric regions and (2) measuring the thickness of the CC at a large number (50–100) of evenly spaced points along the midline and doing regression or factor analysis to determine locations of significant difference. In (1) there are a number of different geometric parcellations in use (Hampel et al., 2002; Weis et al., 1991; Witelson, 1989) and no general agreement on how many divisions are meaningful. Method (2) does not require *a priori* parcellations and can give much finer detail, but suffers from multiple comparisons and does not exploit the correlations expected between neighboring thickness. Factor analysis has been used to

* Corresponding author at: Center for Advanced Brain Imaging, Nathan Kline Institute, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA. Tel.: +1 845 398 6638; fax: +1 845 398 5472.
E-mail address: shlee@nki.rfmh.org (S.H. Lee).

0165-0270/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jneumeth.2013.09.017

Table 1
Sample characteristics of the data set as given by Marcus et al. (2007). Values are *mean* ± *SD* (minimum–maximum). The range of MMSE score is between 0 (worst) and 30 (best).

	CDR=0	CDR=0.5	CDR=1
Number	98	70	28
Sex (female/male)	72/26	39/31	19/9
Age (years)	75.9 ± 9.0 (60–94)	76.4 ± 7.0 (63–92)	82 ± 5.7 (78–86)
MMSE score	29.0 ± 1.2 (25–30)	25.6 ± 3.5 (14–30)	21.7 ± 3.8 (15–29)

divide the CC into anatomically "meaningful" subdivisions and a post-hoc ANOVA-type analysis has been applied to these subdivisions (Denenberg et al., 1991b; Hallam et al., 2008; Kilian et al., 2007). However, factor analysis can be subjective in determining the number of regions and it is not reliable in determining subdivisions.

We explore the second approach to corpus callosum analysis by measuring its thickness profile from rostrum to splenium. We introduce a new statistical approach to analyze the thickness profile to find regional differences in the CC between two populations. To this end, we extend a logistic regression model by incorporating fused lasso regularization to select contiguous regions rather than individual thicknesses. This approach does not require *a priori* divisions of the CC, but is able to identify regionally specific differences in CC thickness between two populations. We demonstrate the proposed method to determine regional specificity of the CC thickness differences between NCs and patients with very mild or mild AD. For this purpose, we utilize images from the Open Access Series of Imaging Studies (OASIS) MRI database (Marcus et al., 2007).

2. Method

2.1. MRI data

The OASIS cross-sectional data set (Marcus et al., 2007) contains 3D MPRAGE MRI brain scans from 416 right-handed subjects. Of these, we include in this study the 98 healthy normal subjects aged 60 or above without dementia (CDR=0) and all 98 subjects aged 60 or above with very mild/mild AD (CDR=0.5 or 1). See Marcus et al. (2007) for details of subject demographics, inclusion/exclusion criteria, MRI acquisition protocol, and preprocessing steps. Table 1 summarizes the demographics of the subjects included in this paper.

2.2. Corpus callosum segmentation

A fully automated method (Ardekani et al., 1997) is applied to find the mid-sagittal plane (MSP) of the MRI volumes to bring the head yaw and roll angles as close as possible to zero. In addition, the anterior commissure (AC) and posterior commissure (PC) were located on the MSP using a fully automated model-based method (Ardekani and Bachman, 2009). Using this information, the original MRI volume was re-sliced to obtain a single image of matrix size: 512 × 512 and pixel size: 0.5 mm × 0.5 mm representing the true AC–PC aligned MSP. Using *a priori* information obtained from manually traced corpora callosa on this type of image, a rectangular CC search region was identified on the MSP. Finally, a multi-atlas model-based segmentation method (Aljabar et al., 2009; Ardekani et al., 2012, 2013; Cabezas et al., 2011) using the Automatic Registration Toolbox (ART) non-linear registration algorithm (Ardekani et al., 2005; Klein et al., 2009) was used to locate the CC within the search region as shown in Fig. 1(a). If necessary, small corrections to the detected CC were made manually using the ITK-SNAP software (Yushkevich et al., 2006) by an operator blinded to subject status. The final segmentation was represented as a binary image where pixels were assigned a value of 1 if they belonged to the CC and a value of 0 otherwise. The CC segmentation algorithm described above has been implemented in C++ and using message passing interface (MPI) parallel programming library. The software, *yuki*, is available publicly at www.nitrc.org/projects/art. The average processing time for a CC segmentation running *yuki* in parallel on 7 of the 8 existing Cores in a 2.4 GHz Dual Quad-Core Linux computer was less than 7 s. Running the program on a single Core in serial mode required approximately 28 s.

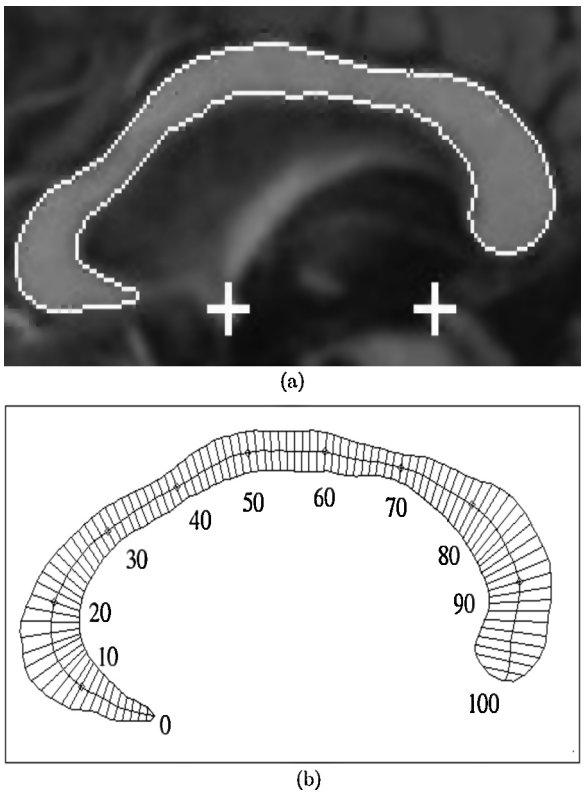


Fig. 1. (a) The outline shows the detected corpus callosum. The automatically detected anterior commissure (AC) and posterior commissure (PC) are shown by the plus signs. (b) An example CC thickness profile. Every tenth medial axis position $\mathbf{m}_i^*$ is indicated by a diamond. Numbering is 0 at the rostrum tip (**r**) and 100 at the inferior point of the splenium (**s**).

It must be emphasized that in the multi-atlas based approach that we have used, the non-linear registrations are applied to the CC atlases only, not to the test image being measured. The test image is treated as a fixed target to which floating atlases are matched. Therefore, the CC of the test image does not undergo any shape or size alternation due to the non-linear registration processes involved in the algorithm. The only alignment that is applied to the test image is a non-shape/size altering rigid-body transformation based on the detected MSP and AC and PC landmarks that brings the test image to a standardized orientation, as described above, before multi-atlas based segmentation is applied.

2.3. Corpus callosum thickness profile

The CC thickness profile is specified in terms of 99 non-zero thickness values at equally spaced intervals along the length of the CC. Our method for finding these values is similar to those of Clarke et al. (1989) and Denenberg et al. (1991a). More specifically, the thickness values are lengths of line segments that connect pairs of points on the *upper* and *lower* boundaries of the CC. The line segments are perpendicular to the medial axis of the CC and intersect the medial axis at equal intervals. The steps for finding the thickness values are as follows.

(a) The tip of the rostrum **r** (point 0 in Fig. 1(b)) and the inferior point of the splenium **s** (point 100 in Fig. 1(b)) are located automatically in the segmented binary CC image. For this purpose, we use information based on the known CC orientation with respect the AC and PC landmarks.
(b) Using the landmarks (rostrum tip and splenium bottom) found in step (a), the CC boundary is divided into two parts, an upper curve and a lower curve, 101 equally spaced points $\mathbf{u}_i$ are placed on the upper boundary and 101 equally spaced points $\mathbf{l}_i$ are located on the lower boundary, where $i = 0, 1, \ldots, 100$, $\mathbf{u}_0 = \mathbf{l}_0 = \mathbf{r}$, and $\mathbf{u}_{100} = \mathbf{l}_{100} = \mathbf{s}$.
(c) The mid-points $\mathbf{m}_i = (\mathbf{u}_i + \mathbf{l}_i)/2$ between the corresponding upper and lower CC boundary pairs $\mathbf{u}_i$ and $\mathbf{l}_i$ are found to define 101 points on a medial axis of the CC.
(d) A medial axis is reconstructed by linear interpolation between the $\mathbf{m}_i$ and smoothed. Then, a new set of 101 equally spaced points $\mathbf{m}_i^*$ are placed on the smoothed medial axis, such that $\mathbf{m}_0^* = \mathbf{r}$ and $\mathbf{m}_{100}^* = \mathbf{s}$.
(e) The normal lines to the medial axis at points $\mathbf{m}_i^*$ ($i = 1, 2, \ldots, 99$) are determined and their intersection points with the upper and lower boundaries of the CC are detected and denoted by $\mathbf{u}_i^*$ and $\mathbf{l}_i^*$, respectively.
(f) Finally, the CC thickness profile is defined as $\mathbf{x} = \{x_i\}$ ($i = 1, 2, \ldots, 99$) where x_i are the distances between points $\mathbf{u}_i^*$ and $\mathbf{l}_i^*$, that is, $x_i = \|\mathbf{u}_i^* - \mathbf{l}_i^*\|$.

The above algorithm makes two assumptions that may not be necessarily true. The first is that, as shown in Fig. 1(b), the perpendicular line segments to the medial axis do not cross each other within the CC. The second assumption is that a perpendicular to the medial axis intersects each of the upper and lower boundaries exactly once. In occasional cases, these assumptions are violated especially in the genu-rostral area. To deal with this situation, the following two steps are added to the above algorithm.

(g) If a line segment perpendicular to the medial axis at a point $\mathbf{m}_i^*$ crosses another line segment within the CC or if a line segment does not cross the upper and lower boundaries exactly once, delete the corresponding points $\mathbf{u}_i^*$ and $\mathbf{l}_i^*$.
(h) Replace the missing $\mathbf{u}_i^*$ and $\mathbf{l}_i^*$ points by placing equi-distance points between the closest existing neighboring points. For example, if $\mathbf{u}_i^*$ were deleted in step (g), replace it with a point that bisects the arc on the upper boundary that joints $\mathbf{u}_{i-1}^*$ and $\mathbf{u}_{i+1}^*$. Extension to cases where more than two adjacent boundary points are missing is trivial. Once the boundary points are corrected, compute the thicknesses $x_i = \|\mathbf{u}_i^* - \mathbf{l}_i^*\|$ as before.

2.4. Logistics regression with fused lasso regularization

The logistic regression models the probability of an event of interest, denoted by p_i, using a univariate score of many predictive variables as $p_i = \exp(s_i)/\{1 + \exp(s_i)\}$ where s_i is a weighted linear sum of predictive variables. The relationship between s_i and p_i is the well-known s-shaped curve. Often s_i is used as a classifier in the sense that we classify the ith subject into the positive class (the class of an event of interest) if $s_i \geq 0$, which is equivalent to $p_i \geq 1/2$. Otherwise, the subject is classified into the normal group. The ℓ_1 regularization (lasso: least absolute shrinkage and selection operator) was introduced by Tibshirani (1996) to simultaneously select and estimate the model. The lasso method finds only a few nonzero coefficients (i.e., weights of the predictive variables that form s_i) and becomes a useful tool when the number of predictors is large but only a few of them are actually effective.

The classical lasso regression by Tibshirani (1996), which penalizes the ℓ_1 norm of the coefficients vector (the sum of absolute values of coefficients) of ordinary least squares, has many variants to incorporate additional structural information on covariates. For example, the elastic net by Zou and Hastie (2005) is designed to simultaneously select strongly correlated variables in the model. The group lasso by Yuan and Lin (2007) again jointly selects or deselects a group of variables to explain a single categorical variable. The fused lasso regression by Tibshirani et al. (2005) assumes that the predictors are observed in an order and penalizes a convex combination of the ℓ_1 norm of differences of adjacent coefficients and the ℓ_1 norm of the coefficient vector itself. It is also known that, like the elastic net, the fused lasso regression selects correlated variables together but it further makes the coefficients to be locally equal to each other.

The callosal thicknesses are spatially correlated in the sense that the thickness at one point is correlated to thicknesses of its neighboring point, and we are interested in finding regions rather than individual thickness points that would differentiate between groups. For this reason, the fused lasso logistic regression (FLLR) is a suitable choice for the analysis of callosal thickness profiles, and results in locally homogeneous and spatially contiguous regions that would indicate the existence of the event of interest (e.g., a subject belonging to the AD group).

Suppose that we have n samples $\{(y_i, \mathbf{d}_i, \mathbf{x}_i), i = 1, 2, \ldots, n\}$ ($n = 196$ in the current study), where $\mathbf{d}_i = (d_{i1}, d_{i2}, \ldots, d_{is})$ is an s-dimensional covariate vector such as demographic information, $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{it})$ is a t-dimensional set of callosal thicknesses ($t = 99$ in the current study), and y_i is a binary variable indicating status of the ith subject. In this paper, $y_i = 1$ if the ith subject is in the AD group, otherwise 0. In order to find regional differences of the CC thickness between two populations controlling for covariates, we consider the logistic regression model:

$$g(p_i) = \log \frac{p_i}{1 - p_i} = \mathbf{d}_i\boldsymbol{\gamma} + \mathbf{x}_i\boldsymbol{\beta},$$

where $p_i = \Pr(y_i = 1)$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_s)^{\mathrm{T}}$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_t)^{\mathrm{T}}$. The likelihood function of (γ, β) is given by

$$\begin{aligned} L(\gamma, \beta; \mathbf{y}, \mathbf{d}, \mathbf{x}) &= \prod_{i=1}^{n} p_i^{y_i}(1 - p_i)^{1-y_i} \\ &= \prod_{i=1}^{n} \exp\left\{ y_i \log \frac{p_i}{1 - p_i} + \log(1 - p_i) \right\} \\ &= \prod_{i=1}^{n} \{ y_i(\mathbf{d}_i\gamma + \mathbf{x}_i\boldsymbol{\beta}) - \log(1 + e^{\mathbf{d}_i\gamma + \mathbf{x}_i\boldsymbol{\beta}}) \}, \end{aligned}$$

where $\mathbf{y} = (y_1, \ldots, y_n)^{\mathrm{T}}$. Subsequently, the log-likelihood function of (γ, β) becomes

$$\begin{aligned} \ell(\beta, \gamma; \mathbf{y}, \mathbf{d}, \mathbf{x}) &= \log L(\gamma, \beta; \mathbf{y}, \mathbf{d}, \mathbf{x}) \\ &= \sum_{i=1}^{n} \{ y_i(\mathbf{d}_i\gamma + \mathbf{x}_i\boldsymbol{\beta}) - \log(1 + e^{\mathbf{d}_i\gamma + \mathbf{x}_i\boldsymbol{\beta}}) \}. \end{aligned} \quad (1)$$

Sparse regression using ℓ_1-regularization is one of the most popular tools for high-dimensional data. The FLLR applied in this paper proposes minimizing

$$\sum_{i=1}^{n} \left\{ -y_i(\mathbf{d}_i\gamma + \mathbf{x}_i\beta) + \log(1 + \exp(\mathbf{d}_i\gamma + \mathbf{x}_i\beta)) \right\} + \Omega(\beta; \lambda_1, \lambda_2), \quad (2)$$

where

$$\Omega(\beta;\lambda_1,\lambda_2)=\lambda_1\sum_{j=1}^{t}|\beta_j|+\lambda_2\sum_{j=2}^{t}|\beta_j-\beta_{j-1}|.$$

Note that penalties are only on $\boldsymbol{\beta}$.

Lee et al. (2006) show that the problem of maximizing the log-likelihood function (1) can be reformulated as a weighted ordinary least squares problem by taking the 2nd order Taylor expansion. Following the steps described in Lee et al. (2006), the objective function (2) can be quadratically approximated. If the current estimate of γ and β are denoted $\widehat{\gamma}^{(r)}$ and $\widehat{\beta}^{(r)}$, respectively, the objective function is given by:

$$Q(\gamma,\beta|\widehat{\gamma}^{(r)},\widehat{\beta}^{(r)})=\sum_{i=1}^{n}w_i(z_i-\mathbf{d}_i\gamma-\mathbf{x}_i\beta)^2+\Omega(\beta;\lambda_1,\lambda_2), \tag{3}$$

by letting

$$w_i=\hat{p}_i(1-\hat{p}_i),$$

$$z_i=\mathbf{d}_i\hat{\gamma}^{(r)}+\mathbf{x}_i\hat{\beta}^{(r)}+\frac{y_i-\hat{p}_i}{\hat{p}_i(1-\hat{p}_i)}\quad :\textit{adjusted}\quad\textit{response}$$

$$\hat{p}_i=g^{-1}(\mathbf{d}_i\hat{\gamma}^{(r)}+\mathbf{x}_i\hat{\beta}^{(r)}).$$

Algorithms to solve the original fused lasso problem using a squared loss have been widely discussed in the literature. However, algorithms for logistic loss have rarely been discussed. In this paper, we consider a modified coordinate descent (CD) algorithm for the FLLR. The proposed algorithm can be considered as a logistic modification of the pathwise coordinate descent algorithm of Friedman et al. (2007). However, our procedure does not utilize the monotonicity (in fusion penalty) of the solution, which implies that, given λ_1, if coefficients of two variables are fused at $\lambda_2=c$, then they are also fused for all $\lambda_2>c$. As they have proved, monotonicity is true only for a few special cases of the covariate matrix.

Our CD algorithm also has "descent" and "fusion" steps. The algorithm iterates these two steps. The descent step is applied for β_k with $\widehat{\beta}_j, j\neq k$, for each k. When the descent step fails to improve the objective function, the fusion step is applied, that is, we set $\beta_k=\beta_{k-1}=\eta$ (or $\beta_k=\beta_{k+1}=\eta$) and do the descent step with respect to η. In other words, we move two parameters β_k and β_{k-1} (or β_{k+1}) together. The detailed steps are as follows:

1. Let $\widehat{\gamma}_j=0$ for $j=1,2,\ldots,s$, $\widehat{\beta}_k=0$, for $k=1,2,\ldots,t$, and, for $i=1,2,\ldots,n$, $\widehat{p}_i=1/\{1+\exp(-\mathbf{d}_i\widehat{\gamma}-\mathbf{x}_i\widehat{\beta})\}$, $w_i=\widehat{p}_i(1-\widehat{p}_i)$, $z_i=\mathbf{d}_i\widehat{\gamma}+\mathbf{x}_i\widehat{\beta}+(y_i-\widehat{p}_i)/\{\widehat{p}_i(1-\widehat{p}_i)\}$, $z_i^*=\sqrt{w_i}z_i$, $\mathbf{d}_i^*=\sqrt{w_i}\mathbf{d}_i$, and $\mathbf{x}_i^*=\sqrt{w_i}\mathbf{x}_i$.
2. (update of γ) We consider the derivative of (3) for γ_j, for $j=1,2,\ldots,s$, given $\beta=\widehat{\beta}$, which is

$$\frac{\partial Q(\gamma,\widehat{\beta}|\widehat{\gamma},\widehat{\beta})}{\partial\gamma_j}=\sum_{i=1}^{n}\left\{2(d_{ij}^*)^2\gamma_j-d_{ij}^*\left\{z_i^*-\sum_{l\neq j}d_{il}^*\gamma_l-\sum_{k=1}^{t}x_{ik}^*\widehat{\beta}_k\right\}\right\},$$

$$\textit{for } j=1,2,\ldots,s. \tag{4}$$

We update $\widehat{\gamma}$ by solving the above set of normal equations by setting (4) equal to 0.

3. (descent step) In this step, we consider the derivative of (3) for β_k, given $\beta_l=\widehat{\beta}_l$, $l\neq k$, which is

$$\frac{\partial Q(\widehat{\gamma},\beta|\widehat{\gamma},\widehat{\beta})}{\partial\beta_k}=\sum_{i=1}^{n}\left\{2(x_{ik}^*)^2\beta_k-x_{ik}^*(z_i^*-\sum_{j=1}^{s}d_{ij}^*\widehat{\gamma}_j-\sum_{l\neq k}x_{il}^*\widehat{\beta}_l)\right\}$$

$$+\lambda_1\mathrm{sgn}(\beta_k)-\lambda_2\mathrm{sgn}(\widehat{\beta}_{k+1}-\beta_k)+\lambda_2\mathrm{sgn}(\beta_k-\widehat{\beta}_{k-1}), \tag{5}$$

where $\mathrm{sgn}(x)$ is the subgradient of $|x|$. The function (5) is piecewise linear with breaks at $\{0,\widehat{\beta}_{k-1},\widehat{\beta}_{k+1}\}$ and let $\tilde{\beta}_k$ be the solution to $\partial Q(\beta_k)/\partial\beta_k=0$. If $Q(\tilde{\beta}_k)\leq Q(\widehat{\beta}_k)$, then let $\widehat{\beta}_k=\tilde{\beta}_k$ and do the descent step to update β_{k+1}. If not, we do the fusion step below to update β_k and β_{k-1}, simultaneously.
4. (fusion step) In this step, we move β_k and β_{k-1} together. We set $\eta=\beta_k=\beta_{k-1}$. The derivative of (3) with respect to η for given $\beta_j=\widehat{\beta}_j, j\neq k,(k-1)$ is

$$\sum_{i=1}^{n}\left\{2\left(x_{i(k-1)}^*+x_{ik}^*\right)^2\eta-\left(x_{i(k-1)}^*+x_{ik}^*\right)\left(z_i^*-\sum_{j=1}^{s}d_{ij}^*\widehat{\gamma}_j-\sum_{l\neq k,(k-1)}x_{il}^*\widehat{\beta}_l\right)\right\}$$

$$+2\lambda_1\mathrm{sgn}(\eta)-\lambda_2\mathrm{sgn}(\widehat{\beta}_{k+1}-\eta)+\lambda_2\mathrm{sgn}(\eta-\widehat{\beta}_{k-2}). \tag{6}$$

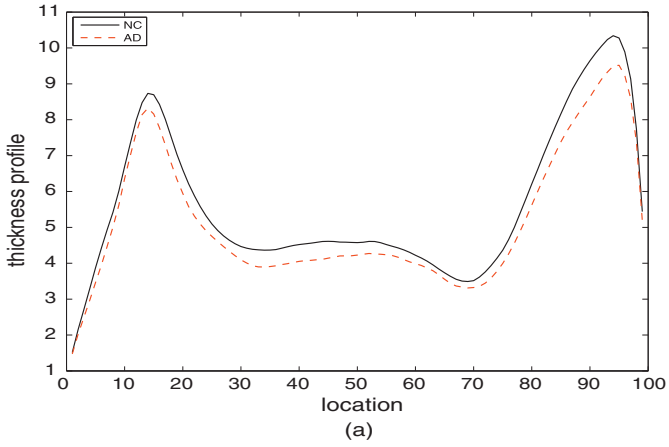
The same procedure in the descent step finds the solution, say $\tilde{\eta}$, to (6)=0. Update β_{k-1} and β_k with $\tilde{\eta}$, and go to the descent step for β_{k+1}.
5. We repeat the descent and fusion steps until convergence.
6. For $i=1,\ldots,n$, update $\widehat{p}_i$, w_i, z_i^*, $\mathbf{d}_i^*$, and $\mathbf{x}_i^*$.
7. Repeat steps 2–6 until convergence.

Generally, difficulties in the fused lasso regularized model arise in the computation speed and finding the optimal tuning parameters. When the number of features and samples are very large, say, over 100,000 features with 200 samples, the algorithm needs to be improved to solve the problem efficiently. The choice of tuning parameters is critical since tuning parameters determine the number of non-zero coefficients. For example, if tuning parameters are small enough, then many coefficients will be zero. Several criteria such as the AIC and the BIC are suggested for finding the optimal tuning parameters. In this study, since we are interested in building a predictive model of logistic regression, tuning parameters (λ_1,λ_2) are chosen to minimize the predictive (Pearson) residual sum of squares, that is, for a new observation $(y,\mathbf{d},\mathbf{x})$,

$$\mathrm{PR}(\lambda)=\mathrm{E}\left\{\frac{(y-\hat{p}_\lambda))^2}{\hat{p}_\lambda(1-\hat{p}_\lambda)}\right\}, \tag{7}$$

where $\hat{p}_\lambda=1/\{1+\exp(-\mathbf{d}\widehat{\gamma}_\lambda-\mathbf{x}\widehat{\beta}_\lambda)\}$, and $\widehat{\gamma}_\lambda$ and $\widehat{\beta}_\lambda$ are the estimates of γ and β with tuning parameters $\lambda=(\lambda_1,\lambda_2)$.

We estimate the above predictive residual using a q-fold cross-validation (CV) method in order to choose the tuning parameters in an adaptive way. To be specific, let $\{S_k:k=1,\ldots,q\}$ be a partition of all samples $\{(y_1,\mathbf{d}_1,\mathbf{x}_1),(y_2,\mathbf{d}_2,\mathbf{x}_2),\ldots,(y_n,\mathbf{d}_n,\mathbf{x}_n)\}$, whose size of S_k is approximately equal to $[n/q]$. For each $k=1,\ldots,q$, let $\widehat{\gamma}_\lambda^{(-k)}$ and $\widehat{\beta}_\lambda^{(-k)}$ be estimates from a training



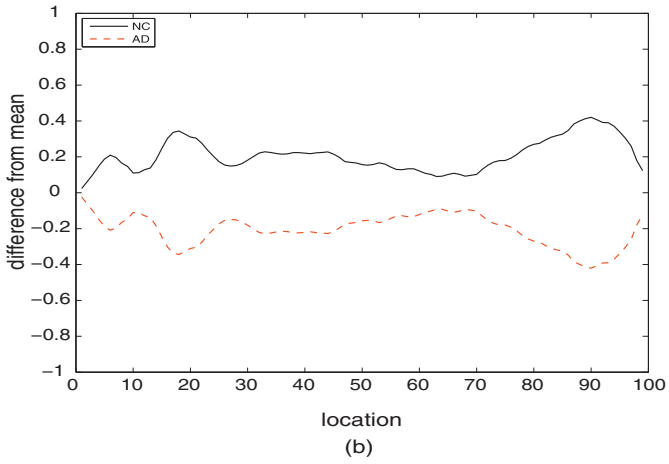
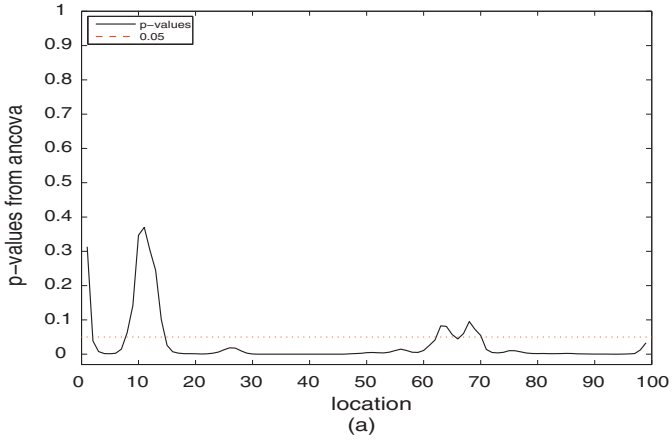


Fig. 2. (a) Mean of raw callosal thickness profiles in NC (solid line) and AD (dotted line). (b) Mean adjusted callosal thickness profiles for each group after removing the effects of age, ICV and sex.



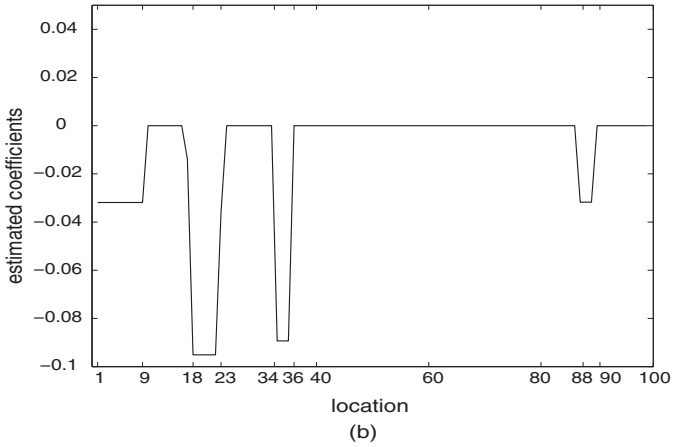


Fig. 3. (a) The p-values obtained from ANCOVA at each location. The p-values are below 0.05 at most locations on CC. (b) Estimated coefficients from the proposed FLLR method. Non-zero coefficients are found at the locations of 1–9, 18–23, 34–36, 88–90.

set $\{S_j : j \neq k\}$ with $\lambda = (\lambda_1, \lambda_2)$. Then, the predictive residual is estimated by

$$\widehat{\mathrm{PR}}(\lambda) = \sum_{k=1}^{q} \sum_{i \in S_k} \frac{(y_i - \widehat{p}_{\lambda,i}^{(-k)})^2}{\widehat{p}_{\lambda,i}^{(-k)}(1 - \widehat{p}_{\lambda,i}^{(-k)})},$$

where $\widehat{p}_{\lambda,i}^{(-k)} = 1/\{1 + \exp(-\mathbf{d}_i \widehat{\gamma}_{\lambda}^{(-k)} - \mathbf{x}_i \widehat{\beta}_{\lambda}^{(-k)})\}$. We perform a grid search to find the minimum point of $\widehat{\mathrm{PR}}(\lambda)$.

3. Results

3.1. Preliminary analysis

It is known that both intra-cranial volume (ICV) and CC cross-sectional areas decrease with age (Holloway et al., 1993). Larger ICV also correlates with larger CC (Bruner et al., 2012). Also, there is a potential of sexual dimorphism in the CC (Ardekani et al., 2013). Therefore, it is necessary to remove the effects of age, ICV and sex, which are considered as potential confounding effects, in examining the callosal differences for AD. In OASIS, the ICV is estimated by the eTIV (estimated total intra-cranial volume) variable and distributed along with the imaging data. Fig. 2 shows the effects of age, ICV (represented by $\mathrm{eTIV}^{1/3}$) and sex on the CC thickness. Fig. 2(a) shows the mean curves of raw CC thickness for each group while Fig. 2(b) shows the callosal differences of NC and AD from the mean after removing the effects of those variables. To obtain the thickness profiles of Fig. 2(b), we take the residuals from the regression model of callosal thickness against age, ICV and sex at each location, then take the average of the residuals for each group at each location. Fig. 2(b) shows a general thinning of the CC in AD as well as greater than average thinning in the areas of the genu and splenium.

It is natural to consider the ANCOVA model with covariates of age, sex and ICV to test the callosal difference between NC and AD at each location, i.e., for each $j = 1, \ldots, 99$,

$$x_{ij} = \alpha_0 + \alpha_1 \mathrm{AGE}_i + \alpha_2 \mathrm{SEX}_i + \alpha_3 \mathrm{eTIV}_i^{1/3} + \alpha_4 \mathrm{AD}_i + \epsilon_i$$

where SEX and AD are the indicator functions for female and AD, respectively, and ϵ_i follows the standard normal distribution. Fig. 3(a) presents the p-values from the ANCOVA indicating that most locations show significance at the level of 0.05 without any correction for multiple testing, which is not very informative in showing regional specific differences for predictive differentiation between AD and NC.

3.2. FLLR analysis

Next we applied our proposed model to callosal thickness profiles after removing the potential confounding effects of age, sex and ICV. OASIS also provides a measure for cognitive impairment, the Mini-Mental State Examination (MMSE) score. MMSE is a simple questionnaire test that is widely used by clinicians to screen for dementia. The range of MMSE score is from 0 (worst) to 30 (best). To utilize all the information available, we include MMSE in our model, which might increase the prediction power of the model.

Table 2
Classification error rate by 5-fold cross-validation: accuracy=0.84(=165/196) with 78% sensitivity and 91% specificity. Out of 22 false negatives, 20 subjects with CDR=0.5 and 2 subjects with CDR=1 were misclassified as normal.

Est.	True	
	NC	AD
NC	89	22
AD	9	76
Total	98	98

If other biomarkers such as CSF beta-amyloid and tau protein levels are available, they can also be incorporated in the model. Thus, we generate a classifier by combining our callosal thicknesses with subject's MMSE scores as follows:

$$\log \frac{p_i}{1-p_i} = \gamma_0 + \gamma_1 \mathrm{MMSE} + \sum_{j=1}^{99} \beta_j \tilde{x}_{ij}, \tag{8}$$

with the fused lasso penalty of $\lambda_1 \sum_{j=1}^{99} |\beta_j| + \lambda_2 \sum_{j=2}^{99} |\beta_j - \beta_{j-1}|$ where $\tilde{x}_{ij}$s are residuals from the regression model, i.e., $\tilde{x}_{ij} = x_{ij} - (\hat{\alpha}_0 + \hat{\alpha}_1 \mathrm{AGE}_i + \hat{\alpha}_2 \mathrm{eTIV}_i^{1/3} + \hat{\alpha}_3 \mathrm{SEX}_i)$, and $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$ are estimates by the ordinary least squares method. The tuning parameters (λ_1, λ_2) are chosen using 5-fold CV as described in Section 2.4 in the range of $(\lambda_1, \lambda_2) \in [1.0, 1.0] \times [5.0, 5.0]$ by interval 0.1. The computation time for the grid search was approximately 2 h on a 2.9 GHz Quad-core Linux computer. In practice, one can start with arbitrarily large λ_1 to preserve most coefficients, followed by finding the optimal λ_2. Then, decrease λ_1 slightly and repeat the search for λ_2 in reasonable range around the previous optimal value of λ_2, and repeat the procedure until the minimum is reached.

Fig. 3(b) shows the estimated coefficients $\hat{\beta}_i$s of the proposed model. We found sparse but contiguous regions that indicate possible callosal differences where $\hat{\beta}_i \neq 0$. Regions [1–9], [18–23], [36–40] and [88–90] have negative coefficients suggesting that these regions are narrower in AD than in NC, and differentiate patients with early AD from NC. Table 2 reports the classification error rate based on our proposed method using 5-fold CV method. The accuracy of our classifier was 0.84 (165/196) with 78% sensitivity and 91% specificity.

4. Discussion and conclusion

In this paper, we have proposed using FLLR (fused lasso logistic regression) to find spatially correlated regions of the CC where thickness may be predictive of AD. Analysis of callosal thickness profiles provide a more detailed view of the differences between AD and NC groups than can be provided by using CC subdivisions. The proposed approach has several advantages in comparison with subdivision procedures. Subdivision procedures are usually done in two steps: subdivisions are determined either geometrically (e.g., splenium as the posterior 1/5 of the CC) or by factor analysis; and equality of the subdivision areas between the two populations is tested. In this approach, CC subdivisions are chosen subjectively or by noisy factor loadings which can be unreliable. The proposed FLLR method reformulates the problem as a feature selection problem in the framework of a prediction model, and finds differential subregions adaptively. For this reason, the FLLR method finds differential subregions between two populations more objectively and reliably. Our method is more specific than analysis based on subdivisions of the CC since we consider entire callosal thickness profiles and take neighborhoods into account in our model. In addition, it simultaneously provides a predictive model which may be valuable for diagnosis.

Finally, we have applied the proposed FLLR to find subregions of the CC where thicknesses differentiate early AD from NC. Our proposed method found that sections of the genu and splenium of AD are proportionally thinner than those of NC with 84% prediction accuracy. Nine normal subjects were misclassified as patients. In clinical practice, such subjects could be candidates for further evaluation. Twenty-two patients with AD were misclassified as normal. Of the 22 false negatives, 20 subjects with CDR=0.5 and 2 subjects with CDR=1 were misclassified as normal. A possibility of misdiagnosis should be considered further in these cases since thinning of the CC appears in the normal range incongruent with there clinical symptoms. A careful study with more extensive follow up is required to determine if these misclassified subjects of the CDR=0.5 group have clinical significance.

The regions identified by this method include parts of the genu that are connected to the frontal lobes, and a region in the splenium that seems to contain fibers that run to the medial temporal lobes (Hofer and Frahm, 2006). Both are regions that are implicated in AD. The region in the splenium seems to have a lesser weight, possibly due to its being a narrow tract whose position suffers from variability.

The classifier we have described is based on two commonly used diagnostic procedures, the MMSE and a structural MRI. Our automatic analysis programs require little supervision or intervention. The MRI protocol involved does not require long or complex acquisitions. The CC can be segmented in a fast, automatic, and reliable way. The FLLR model can be easily extended to include other brain or chemical biomarkers of AD.

Acknowledgments

J. Lim's research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2011-0029104). We appreciate the data made available by the OASIS project. OASIS was supported under grants P50 AG05681, P01 AG03991, R01 AG021910, P20 MH071616, and U24 RR021382.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jneumeth.2013.09.017.

References

Aljabar P, Heckemann RA, Hammers A, Hajnal JV, Rueckert D. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. Neuroimage 2009;46:726–38.
Ardekani BA, Kershaw J, Braun M, Kanno I. Automatic detection of the mid-sagittal plane in 3-D brain images. IEEE Trans Med Imaging 1997;16:947–52.
Ardekani BA, Guckemus S, Bachman AH, Hoptman MJ, Wojtaszek M, Nierenberg J. Quantitative comparison of algorithms for inter-subject registration of 3D volumetric brain MRI scans. J Neurosci Methods 2005;142:67–76.
Ardekani BA, Bachman AH. Model-based automatic detection of the anterior and posterior commissures on MRI scans. Neuroimage 2009;46(3):677–82.
Ardekani BA, Toshikazu I, Bachman AH, Szeszko PR. Multi-atlas corpus callosum segmentation with adaptive atlas selection. In: Proceedings of the ISMRM; 2012.
Ardekani BA, Figarsky K, Sidtis JJ. Sexual dimorphism in the human corpus callosum: an MRI study using the OASIS brain database. Cereb Cortex 2013;23(10):2514–20.
Bruner E, de la Cuetara JM, Colom R, Martin-Loeches M. Gender-based differences in the shape of the human corpus callosum are associated with allometric variations. J Anat 2012;220:417–21.
Cabezas M, Oliver A, LladA3 X, Freixenet J, Cuadra MB. A review of atlas-based segmentation for magnetic resonance brain images. Comput Methods Prog Biomed 2011;104:e158–177.
Clarke M, Kraftsik R, van Derk Loos H, Innocenti GM. Forms and measures of adult and developing human corpus callosum: is there sexual dimorphism? J Comp Neurol 1989;280:213–30.
Denenberg VH, Cowell PE, Fitch RH, Kertesz A, Kenner GH. Corpus callosum: multiple parameter measurements in rodents and humans. Phys Behav 1991a;49:433–7.

Denenberg VH, Kertesz A, Cowell PE. A factor analysis of the human's corpus callosum. Brain Res 1991b;548:126–32.
Di Paola M, Luders E, Di Iulio F, Cherubini A, Passafiume D, Thompson PM, et al. Callosal atrophy in mild cognitive impairment and Alzheimer's disease: different effects in different stages. Neuroimage 2010;49(1):141–9.
Frederiksen KS, Garde E, Skimminge A, Ryberg C, Rostrup E, BaarA(c) WFC, et al. Corpus callosum atrophy in patients with mild Alzheimer's disease. Neurodegener Dis 2011;8:476–82.
Friedman J, Hastie T, Hofling H, Tibshirani R. Pathwise coordinate optimization. Ann Appl Stat 2007;1(2):302–32.
Hallam BJ, Brown WS, Ross C, Buckwalter JG, Bigler ED, Tschanz JT, et al. Regional atrophy of the corpus callosum in dementia. J Int Neuropsychol Soc 2008;14:414–23.
Hampel H, Teipel SJ, Alexander GE, Pogarell O, Rapoport SI, Möller H-I. In vivo imaging of regions and cell type specific neocortical neurodegeneration in Alzheimer's disease. J Neural Transm 2002;109:837–55.
Hofer S, Frahm J. Topography of the human corpus callosum revisited - comprehensive fiber tractography using diffusion tensor magnetic resonance imaging. Neuroimage 2006;32:989–94.
Holloway RL, Anderson PJ, Defendini R, Harper C. Sexual dimorphism of the human corpus callosum from three independent samples: relative size of the corpus callosum. Am J Phys Anthropol 1993;92(4):481–98.
Kilian S, Brown WS, Hallam BJ, McMahon W, Lu J, Johnson M, et al. Regional callosal morphology in autism and macrocephaly. Dev Neuropsychol 2007;33(1):74–99.
Klein A, Andersson J, Ardekani BA, Ashburner J, Avants B, Chiang M-C, et al. Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 2009;46(3):786–802.
Lee S, Lee H, Abbeel P, Ng A. Efficient L_1 regularized logistic regression. In: Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06); 2006.
Marcus DS, Wang TH, Parker J, Csernansky JG, Morris JC, Buckner RL. Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. J Cogn Neurosci 2007;19:1498–507.
Risacher SL, Saykin AJ. Neuroimaging and other biomarkers for Alzheimer's disease: the changing landscape of early detection. Annu Rev Clin Psychol 2013;9:621–48.
Teipel SJ, Bayer W, Alexander GE, Zebuhr Y, Teichberg D, Kulic L, et al. Progression of corpus callosum atrophy in Alzheimer disease. Arch Neurol 2002;59(2):243–8.
Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B Stat Methodol 1996;58(1):267–88.
Tibshirani R, Saunders M, Rosset S, Zhu J, Knight K. Sparsity and smoothness via the fused lasso. J R Stat Soc Ser B Stat Methodol 2005;67(1):267–88.
Wang PJ, Saykin AJ, Flashman LA, Wishart HA, Rabin LA, Santulli RB, et al. Regionally specific atrophy of the corpus callosum in AD, MCI and cognitive complaints. Neurobiol Aging 2006;27(11):1613–7.
Weis S, Jellinger K, Wenger E. Morphometry of the corpus callosum in normal aging and Alzheimer's disease. J Neural Transm Suppl 1991;33:35–8.
Witelson SF. Hand and sex differences in the isthmus and genu of the human corpus callosum. A post mortem morphological study. Brain 1989;112(3):799–835.
Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. J R Stat Soc Ser B Stat Methodol 2007;68(1):49–67.
Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 2006;31(3):1116–28.
Zou H, Hastie T. Regularization and variable selection via the elastic net. J R Stat Soc Ser B Stat Methodol 2005;67(2):301–20.