BIOINFOMATICS

From phenotype to genotype: an association study of longitudinal phenotypic markers to Alzheimer’s disease relevant SNPs

Hua Wang1, Feiping Nie1, Heng Huang1† Jingwen Yan2, Sungeun Kim2, Kwangskik Nho3, Shannon L. Risacher2, Andrew J. Saykin2, Li Shen2*, for the Alzheimer’s Disease Neuroimaging Initiative4

1Department of Computer Science and Engineering, University of Texas at Arlington, TX 76019, USA and 2Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN 46202, USA

ABSTRACT

Motivation: Imaging genetic studies typically focus on identifying single-nucleotide polymorphism (SNP) markers associated with imaging phenotypes. Few studies perform regression of SNP values on phenotypic measures for examining how the SNP values change when phenotypic measures are varied. This alternative approach may have a potential to help us discover important imaging genetic associations from a different perspective. In addition, the imaging markers are often measured over time, and this longitudinal profile may provide increased power for differentiating genotype groups. How to identify the longitudinal phenotypic markers associated to disease sensitive SNPs is an important and challenging research topic.

Results: Taking into account the temporal structure of the longitudinal imaging data and the interrelatedness among the SNPs, we propose a novel ‘task-correlated longitudinal sparse regression’ model to study the association between the phenotypic imaging markers and the genotypes encoded by SNPs. In our new association model, we extend the widely used $\ell_2^1$-norm for matrices to tensors to jointly select imaging markers that have common effects across all the regression tasks and time points, and meanwhile impose trace-norm regularization onto the unfolded coefficient tensor to achieve low rank such that the interrelationship among SNPs can be addressed.

The effectiveness of our method is demonstrated by both empirical evaluations and a compact set of selected imaging predictors relevant to disease sensitive SNPs.

Availability: Software is publicly available at:

http://ranger.uta.edu/~heng/Longitudinal

Contact: heng@uta.edu or shenli@inpui.edu

1 INTRODUCTION

Neuroimaging genetics is an emerging research field, where brain imaging is used as quantitative phenotypes to investigate the role of genetic variation in brain structure and function. It holds great promise for a systems biology of the brain to better understand complex neurobiological systems, from genetic determinants to cellular processes to the complex interplay of brain structure, function, behavior and cognition. Disorders of the nervous system are associated with complex neurobiological changes, which may lead to profound alterations at all levels of organization.

Genome-wide association studies (GWAS) have been increasingly performed to correlate high-throughput single-nucleotide polymorphism (SNP) data to large-scale imaging data. To facilitate such association analysis, many studies used a hypothesis-driven approach (Glahn et al. 2007) by making significant reduction in one or both data types. For example, some whole-brain studies focused on a small number of genetic variables (e.g. Brun et al. 2004, Filippini et al. 2004, Harris et al. 2004, Nichols and Inkster, 2009), and some whole-genome studies examined a limited number of imaging variables (e.g. Baranzini et al. 2006, Potkin et al. 2009, Seshadri et al. 2009). Many SNPs have been identified as risk factors for Alzheimer’s disease (AD), see those in the AlzGene database (www.alzgene.org).

So far most studies focus on selecting and associating SNPs to AD status or imaging phenotypes. Very few studies have been done to directly examine how the SNP values change when phenotypic measures are varied, i.e. via regression of SNP values on phenotypic measures. This alternative approach may have a potential to help us discover important imaging genetic associations from a different perspective. In this study, we perform such an initial analysis for finding phenotypic imaging markers that are related to SNPs from or proximal to AlzGene candidates.

Neuroimaging measures have been widely studied to predict disease status and/or cognitive performance (Batmanghelich et al. 2006, Shen et al. 2006). However, whether these measures coupled with their longitudinal profiles have sufficient power to infer relevant genotype groups is still an under-explored yet important topic in AD research. A simple strategy typically used in longitudinal studies (e.g. Risacher et al. 2010) is to analyze a single summarized value such as average change rate of change or slope. This approach may be inadequate to distinguish the complete dynamics of cognitive trajectories and thus become unable to identify the underlying genetic structure.

© The Author(s) 2012. Published by Oxford University Press.
The SNP data used in this study (Saykin, 2006) are coupled (Argyriou et al., 2006), so that the identified imaging phenotypes have common influence on all the SNPs. Based on the emerging structured sparse learning techniques, we propose to use the trace/nuclear norm regularization (Candès and Recht, 2009; Nie, 2012) to approximately minimize the rank of regression coefficient matrix, which is linearly dependent. Finally, through enforcing the $\ell_2,1$-norm regularization, the imaging feature selection across most SNPs are linearly dependent. Through enforcing the $\ell_2,1$-norm regularization, the imaging feature selection across most SNPs are coupled (Argyriou et al., 2006; Dobnikar et al., 2004), so that the identified imaging phenotypes have common influence on all the SNPs.

We apply the proposed method to the Alzheimer’s Disease Neuroimaging Initiative (ADNI) cohort (Mueller et al., 2005) to identify the longitudinal phenotypic markers related to candidate AD genes. With these observations, in this work, we propose a new task-correlated longitudinal sparse regression framework to effectively identify the longitudinal phenotypic markers related to candidate AD SNPs. Based on the emerging structured sparse learning techniques, which has been effectively applied in imaging genetics studies (Wang et al., 2012a, 2012b), the new combined structured sparse regularizations are introduced to tackle the longitudinal phenotypic patterns and biological genotypic correlations. The proposed new computational biology model consists of three major components. First, due to the serial measures of the imaging phenotypes over time, we propose a novel longitudinal regression analysis method. As a result, the regression coefficients assess the relationships between longitudinal phenotypes and their genetic makeup. Second, certain SNPs are naturally correlated via different ways, e.g. multiple SNPs from one single gene often jointly carry out similar genetic functionalities, SNPs in high linkage disequilibrium (LD) are linked together in meiosis. To incorporate such SNP correlations in our association studies, we propose to use the trace/nuclear norm regularization (Candes and Recht, 2006; Nie et al., 2013) to approximately minimize the rank of regression coefficient matrix, such that the coefficients of phenotypes associated to correlated SNPs are linearly dependent. Finally, through enforcing the $\ell_2,1$-norm regularization, the imaging feature selection across most SNPs are coupled (Argyriou et al., 2006; Dobnikar et al., 2004), so that the identified imaging phenotypes have common influence on all the SNPs.

The SNP data used in this study (Saykin, 2006) were genotyped using the Human 610-Quad BeadChip (Illumina, Inc., San Diego, CA, USA). Among all SNPs, only SNPs belonging to the top 40 AD candidate genes (Bertram et al., 2011a, b, 2012a, b) were selected after the standard quality control (QC) and imputation steps. The QC criteria for the SNP data include (i) call rate check per subject and per SNP marker, (ii) gender check, (iii) sibling pair identification, (iv) the Hardy–Weinberg equilibrium test, (v) marker removal by the minor allele frequency and (vi) population stratification. As the second pre-processing step, the quality-controlled SNPs were imputed using the MaCH software (Sawhill et al., 2011) to estimate the missing genotypes. After that, the Illumina annotation information based on the Genome build 36.2 was used to select a subset of SNPs, belonging to the top 40 AD candidate genes (Bertram et al., 2007). The above procedure yielded 1224 SNPs from 37 genes. For the remaining three genes, no SNPs were available on the genotyping chip.

### 2.2 MRI analysis and extraction of imaging phenotypes

Two widely used automated MRI analysis techniques were used to process and extract imaging phenotypes across the brain from all the MRI scans of ADNI participants as previously described (Shen et al., 2008). First, voxel-based morphometry (VBM) (Ashburner and Friston, 2005) was performed to define modulated gray matter (GM) maps and extract local GM values for target regions. Second, automated parcellation via FreeSurfer V4 (Fischl et al., 2004) was conducted to define volumetric and cortical thickness values for regions of interest (ROIs) and to extract total intracranial volume (ICV). Further information is available in (Shen et al., 2010). The time points examined in this study for imaging markers included baseline (BL), Month 6 (M6), Month 12 (M12) and Month 24 (M24). All the participants with no missing BL/M6/M12/M24 MRI measurements were included in this study. Figure 2 shows the names of these ROIs in the brain space. All these measurements were adjusted for baseline ICV using the regression weights derived from the healthy control (HC) participants.

### 3. TASK-CORRELATED LONGITUDINAL SPARSE REGRESSION

For the association study of longitudinal imaging phenotypes to the genotypes, the input imaging features are a set of matrices $X = \{X_1, X_2, \ldots, X_T\} \in \mathbb{R}^{d \times n \times T}$ corresponding to the measurements at $T$ consecutive time points, where $X_t$ is the imaging measurements for a certain type of imaging markers, such as VBM or FreeSurfer markers used in this study, at time $t\ (1 \leq t \leq T)$. Obviously, $X$ is a tensor data with $d$ imaging features, $n$ subject samples and $T$ time points. The output genetic variations described by $c$ SNPs for the $n$ subject samples forms a matrix $Y = [y_{1,1}, \ldots, y_{1,n}] \in \mathbb{R}^{n \times c}$, where the $y_{t,i} \in \mathbb{R}^c$ is the SNP values of the $t$th subject sample. Our goal is to learn from $\{X, Y\}$ a model that can reveal the associations between the longitudinal imaging phenotypes $X$ and the genotypes $Y$.

A straightforward method for relating imaging phenotypes and SNPs is to perform regression at each time point separately, which, though, does not take into account the valuable information conveyed by the longitudinal patterns of the phenotypic inputs. To overcome this limitation, different from previous studies that learned the regression coefficient matrix for each time point individually, we aim to learn a unified longitudinal regression model to find the genetic features that are associated to the longitudinal imaging patterns over all the measurement time points. To this end, we expect to learn a coefficient tensor ($\hat{B}$) to reveal the temporal changes of the coefficient matrices. In this article, we propose to use the low-rank structured sparse regularizations to explore the temporal patterns and the interrelatedness between SNPs in a new task-correlated longitudinal sparse regression model.
3.1 Task-correlated longitudinal sparse regression using low-rank structured sparse regularizations

The simplest model to associate the phenotypic markers to the genotypes is the multivariate regression model, which solves the following optimization problem:

$$\min_{B} J_0 = \mathcal{L}(B) + \gamma \sum_{k=1}^{d} \left\|b_k^T \right\|_2^2,$$

where $b_k^T$ denotes the $k$th row of coefficient matrix $B_t$ at time $t$, and $\mathcal{L}(B)$ is the proposed longitudinal loss and defined as

$$\mathcal{L}(B) = \|B \otimes 1 - X^T - Y^T\|_F^2 = \sum_{t=1}^{T} \|X_t^T B_t - Y_t^T\|_F^2.$$  
(2)

Because the objective $J_0$ in Equation (3) can be decoupled for each individual time point and does not consider the longitudinal correlations between the imaging features and the SNPs, it is not suitable for longitudinal data analysis and feature selection. Because the selected imaging markers with temporal changes are desired to connect all the SNPs, the $T$ groups of regression tasks at different time points should not be decoupled and have to be performed simultaneously. Thus, we introduce the structured sparse regularization [Avery et al. 2012, Nie et al. 2013, Obozinski et al. 2009] into the longitudinal data regression and feature selection model as following:

$$\min_{B} J_1 = \mathcal{L}(B) + \gamma \sum_{k=1}^{d} \left\|b_k^T \right\|_2^2.$$  
(3)

It has been shown that Candès and Recht 2011, Candès and Recht 2007 the trace-norm is the best convex approximation of the rank-norm. Therefore, the third term of $J_2$ in Equation (4) indeed minimizes the rank of the unfolded learning model $B$, such that the correlations among the SNPs are captured. Due to its both capabilities for imaging marker selection and task correlation integration, we call $J_2$ defined in Equation (4) as the proposed 'task-correlated longitudinal sparse regression model'.

3.2 A new optimization algorithm and its global convergence

Because our new objective $J_2$ is non-smooth, the problem in Equation (4) is difficult to solve in general. Some existing methods, such as LARS [Efron et al. 2004], linear gradient search [Lin et al. 2009], proximal [Beck and Teboulle 2009] methods, can solve it, but not efficiently. Thus, in this subsection we derive a new efficient algorithm to solve $J_2$ with rigorous proof of its global convergence.

Taking the derivative of $J_2$ w.r.t. $B_t$ and set it to zeros, we have:

$$2X_t^T B_t - 2X_t Y_t + 2\gamma D_t B_t + 2\gamma_2 D_t^T B_t = 0,$$  
(5)

where $D$ is a diagonal matrix with $D(k, k) = \frac{1}{n_{\text{tr}}(k)}$ and $D = (BB^T)^{-1/2}$. Thus, we can derive

$$B_t = (X_t X_t^T + \gamma_1 D + \gamma_2 D_t^{-1} Y_t Y_t^T)B_t.$$  
(6)

When the time $t$ changes from 1 to $T$, we can compute $B_t \{1 \leq t \leq T\}$ by Equation (6). Because $D$ and $D$ depend on $B$ and can be seen as latent variables, we propose an iterative algorithm to obtain the global optimum solutions of $B$ in Algorithm 1.

**Algorithm 1:** A new algorithm to minimize $J_2$ in Equation (4)

1. Initialize $B_0 \in \mathbb{R}^{d \times m \times T}$ using the regression results at each individual time point;
2. Calculate the diagonal matrix $D$, where the $k$-th diagonal element is computed as $D(k, k) = \frac{1}{n_{\text{tr}}(k)}$;
3. Update $B_t$ by $B_t = (X_t X_t^T + \gamma_1 D + \gamma_2 D_t^{-1} Y_t Y_t^T)B_t$;
4. Until Converges;
5. Result $\hat{B} = \{B_t; B_2; \ldots; B_T\} \in \mathbb{R}^{d \times m \times T}$.

We summarize the convergence of Algorithm 1 as following.

**Theorem 1:** Algorithm 1 monotonically decreases $J_2$ in Equation (4) in each iteration, and converges to the globally optimal solution.

**Proof:** In Algorithm 1 in each iteration we denote the updated $B_t$ as $\tilde{B}_t$ and the updated $\hat{L}$ as $\tilde{L}$. From step 4 we know that:

$$\tilde{L} + \gamma_1 \sum_{t=1}^{T} \text{Tr}(\tilde{B}_t^T D_t \tilde{B}_t) + \gamma_2 \sum_{t=1}^{T} \text{Tr}(\tilde{B}_t^T D_t \tilde{B}_t) \leq$$

$$\tilde{L} + \gamma_1 \sum_{t=1}^{T} \text{Tr}(\tilde{B}_t^T D_t \tilde{B}_t) + \gamma_2 \sum_{t=1}^{T} \text{Tr}(\tilde{B}_t^T D_t \tilde{B}_t).$$  
(7)

By this theorem, the updated $\tilde{L}$ and $\tilde{B}_t$ will converge in each iteration, and the global optimum solution can be obtained.
In each iteration, denote the updated $B$ as $\bar{B}$ and the updated $b_t^f$ as $\bar{b}_t$, according to the definitions of $D$ and $D_t$, we can write

\[
\mathcal{L} + \frac{\gamma_1}{2} \sum_{k=1}^{d} \frac{\|\sum_{i=1}^{T} d_i |b_{k,i}^f|^2\|_2^2}{\gamma_2} + \frac{\gamma_2}{2} \text{tr}(B B^T (B B^T)^{-1}) \leq \mathcal{L} + \frac{\gamma_1}{2} \sum_{k=1}^{d} \frac{\|\sum_{i=1}^{T} d_i |b_{k,i}^f|^2\|_2^2}{\gamma_2} + \frac{\gamma_2}{2} \text{tr}(\bar{B} B^T (\bar{B} B^T)^{-1}).
\]  

Following Nie et al. (2010) it can be verified that

\[
\sum_{i=1}^{T} d_i |b_{k,i}^f|^2 \leq 2 \sum_{i=1}^{T} \sum_{j=1}^{T} |b_{k,i}^f b_{k,j}^f|^2 + 2 \sum_{i=1}^{T} \sum_{j=1}^{T} |b_{k,i}^f b_{k,j}^f|^2.
\]  

\[
\text{tr}(B B^T) - \text{tr}(\bar{B} B^T) \leq 2 \sum_{i=1}^{T} \sum_{j=1}^{T} |b_{k,i}^f b_{k,j}^f|^2.
\]  

Adding the both sides of Equations (8) and (9) together, we obtain

\[
\mathcal{L} + \gamma_1 \sum_{k=1}^{d} \frac{\sum_{i=1}^{T} |b_{k,i}^f|^2}{\gamma_2} + \gamma_2 \text{tr}(B B^T) \leq \mathcal{L} + \gamma_1 \sum_{k=1}^{d} \frac{\sum_{i=1}^{T} |b_{k,i}^f|^2}{\gamma_2} + \gamma_2 \text{tr}(\bar{B} B^T).
\]  

Thus, our algorithm decreases the objective value of Equation (8) in each iteration. When the objective value keeps unchanged, Equation (9) is satisfied, i.e. the KKT condition of the objective is satisfied. Our algorithm reaches one of the optimal solutions. Because our objective in Equation (8) is a convex problem, our Algorithm 1 will converge to one of the globally optimal solution.

**Computational analysis.** In the iteration loop of Algorithm 1 Step 2 is computationally trivial. Step 3 solves a singular value decomposition (SVD) problem, and Step 4 solves a system of linear equations, both of which, thereby the whole algorithm, are well studied in literature and can be solved very efficiently by existing numerical packages.

## 4 EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we evaluate the proposed method by applying it to the ADNI cohort, where a wide range of imaging markers measured over a period of 2 years are examined and associated to SNPs that are relevant to AD. The goal is to discover a compact set of phenotypic imaging markers that are closely related to AD-sensitive genotypes encoded by SNPs.

### 4.1 Improved prediction of SNPs from longitudinal phenotypic imaging markers

We first evaluate the proposed method by applying it to the ADNI cohort to predict the SNPs of the participants from each of their two types of imaging phenotypes, i.e. VBM markers and FreeSurfer markers, tracked over four different time points, including BL and M06/M12/M24. Because some subjects of the ADNI cohort do not have complete imaging marker measurements over all the four time points, in our experiments we use the subject samples that have both SNPs data and complete imaging measurements. As a result, two subsets of ADNI subjects are included in our experiments, one for each type of imaging phenotypes, as detailed in Table 1.

We compare the proposed method against its three close counterparts including multivariate linear regression (LR) method, ridge regression (RR) method and absolute shrinkage and selection operator (Lasso) method. LR method is the most broadly used association model in both statistical learning and imaging genetics. RR method is the regularized version of LR method to avoid over-fitting. Lasso method replaces the squared $\ell_2$-norm regularization in RR method by the $\ell_1$-norm regularization, from which sparse results can be achieved. Different to these compared methods, our new association model imposes structured sparsity via the tensor $\ell_{2,1}$-norm regularization for phenotypic marker selection and the trace-norm regularization for capturing the interrelationships among different SNPs. We implement two versions of the proposed method as follows. First, we implement our method by only imposing the trace-norm regularization, denoted as ‘Ours’, which only makes use of the SNPs’ correlations, but does not select longitudinal imaging markers. Second, we implement the full version of the proposed method, denoted as ‘Ours’, which solves the problem in Equation (8). For measuring the regression performance of the five compared association models, we use a 5-fold cross-validation strategy by computing the Pearson’s correlation coefficient (CORR) and the root mean square error (RMSE) between the predicted and the actual SNP values, which are reported in Figure 1.

As can be seen from Figure 1, if we only use the baseline data, the proposed method is reduced into a conventional multi-task regression model, which appears as a matrix but not a tensor and achieves only the slightly better performance than the RR and Lasso methods. On the other hand, by using the longitudinal data, the performance of the proposed method is significantly improved, e.g. for predicting SNPs using the longitudinal data over all the four time points, the proposed (BL+M24) method achieves the CORR of 0.793 and 0.812 and the RMSE of 0.314 and 0.301, respectively, which are much better than the case of using only the baseline data. In addition, Figure 1 also shows that the usage of longitudinal data can improve the performances of all the LR, RR and Lasso methods, although the improvements are much less than the proposed method. These results demonstrate the effectiveness of using longitudinal data for improved regression from imaging phenotypes to genotypes.

### Table 1. Numbers of participants in the experiments using two different types of imaging markers

| Imaging phenotypes | # Total | # AD | # MCI | # HC |
|--------------------|---------|------|-------|------|
| VBM                | 424     | 86   | 194   | 144  |
| FreeSurfer         | 474     | 100  | 216   | 158  |

Downloaded from https://academic.oup.com/bioinformatics/article-abstract/28/18/i619/250340 by guest on 30 July 2018
Association study of longitudinal phenotypic markers to AD relevant SNPs

4.2 Identification of longitudinal imaging markers

One primary goal of this study is to identify a subset of imaging phenotypes that are highly correlated to certain SNPs to capture important imaging genomic associations in AD research. Thus, we examine the phenotypic imaging markers identified by the proposed methods, which are relevant to the genotypes encoded by SNPs.

4.2.1 Identified imaging markers with high AD risks

Shown in Figure 1 are the overall regression coefficients for all the VBM and FreeSurfer measures with respect to the 1224 SNPs used in this study. Because these SNPs are AlzGene candidates or proximal to the candidates, the results in Figure 1 can help identify SNP-relevant imaging phenotypes and have a potential to gain biological insights from gene to brain to symptoms. Besides, the top 10 selected VBM imaging features, as well as their association coefficients, are visualized in Figure 2 by mapping them onto the human brain.

A first glance at the association weight maps shows that the selected imaging markers have clear patterns that span across all the four studied time points, which demonstrates that these phenotypic markers are longitudinally stable thus can serve as screening targets over the course of AD progression. We also observe that hippocampal measures (LHippocampus, RHippocampus, LHippVol and RHippVol) are identified, which is in accordance with the fact that in the pathological pathway of AD, medial temporal lobe including hippocampus is firstly affected, followed by progressive neocortical damage (Braak and Braak, 1991; Delacourte et al., 1999). The thickness measures of isthmus cingulate (LInsthCing and RInsthCing), frontal pole (LFrontalPole and RFrontalPole) and posterior cingulate gyrus (LPPostCingulate and RPPostCingulate) are also selected, which, again, is in accordance with the fact that the GM atrophy of these regions is high in AD (Ehrmann et al., 2010; McEvoy et al., 2009).

In summary, the identified longitudinally stable markers strongly agree with the existing findings, which warrants the correctness of the discovered phenotype-genotype associations, and reveals the complex relationships among MRI measures, genetic variations and diagnosis status. This is of clear importance for theoretical research and clinical practices for a better understanding of AD mechanism.

4.2.2 Case studies: markers identified for rs423958-APOE and rs11136600-CLU

We provide two case studies to show the top 10 FreeSurfer markers associated with two major AD risk SNPs: rs423958-APOE and rs11136600-CLU. We explore the associations between the FreeSurfer markers and the two SNPs in four different subject groups induced from the ADNI data, i.e. the groups of All, AD, MCI and HC participants, respectively. The number of the subjects in each group is available in Table 1. We select the imaging markers by their average regression coefficients over all the four time points. The top 10 FreeSurfer markers relevant to rs423958-APOE and their regression coefficients are shown in Figure 3 and those relevant to rs11136600-CLU are shown in Figure 4. From Figure 3 we can see that most of the top 10 FreeSurfer markers for rs423958-APOE in the four different testing groups are well-known AD-sensitive phenotypes, such as hippocampal volume in All, AD, MCI and HC groups; amygdala volume in All, AD, MCI and HC groups; and entorhinal cortex thickness in AD and HC groups. Similar patterns are also observed for rs11136600-CLU, as shown in Figure 4. Although data are not shown due to space limit, our VBM analyses have also yielded similar results. The complete imaging marker identification results by our method for both VBM and FreeSurfer markers on the top 10 identified SNPs are available at the author’s website at http://ranger.uta.edu/~heng/img.snp/. These results have again demonstrated the promise of the proposed method in terms of its capability to identify imaging markers relevant to AD-sensitive SNPs.

5 CONCLUSIONS

Elucidating the associations between longitudinal phenotypic imaging markers and AD sensitive SNPs is of important value for both scientific research and clinical practice. In this article, we presented a new task-correlated longitudinal sparse regression method to identify longitudinal imaging markers to AD-relevant SNPs. In our newly proposed regression model, we imposed a tensor $\ell_{2,1}$-norm regularization extended from the standard matrix $\ell_{2,1}$-norm to capture the temporal patterns in the longitudinal data over all the tasks of interest, and meanwhile imposed the trace-norm regularization onto the unfolded coefficient tensor such that the interrelatedness among the SNPs during the progression of AD conversion is addressed. Due to the additional time dimension of the input data and the non-smoothness of the tensor $\ell_{2,1}$-norm and trace-norm, solving the formulated objective of our new method was very challenging. Therefore, we presented an efficient iterative algorithm and rigorously proved its convergence to the global optimum. We applied the proposed method to the ADNI cohort and evaluated it in both SNPs prediction and longitudinal imaging marker identification. The clearly improved regression performance in the prediction and highly suggestive imaging markers selected by our new method have validated its effectiveness.

Fig. 1. Regression performance with respect to the use of different number of longitudinal time points by three different methods especially by the proposed method, which has the capability to make use of the input data through longitudinal feature selection; and the integration of the interrelatedness among the SNPs.
This research was supported by the National Science Foundation Grant (NSF) CCF-0830780, CCF-0917274, DMS-0915228 and IIS-1117965 at UTA; and by NSF IIS-1117335, NIH UL1 RR025761, U01 AG024904, NIA RC2 AG036535, NIA R01 AG19771 and NIA P30 AG10133-18S1 at IU.

Data collection and sharing for this project was funded by the ADNI (National Institutes of Health Grant U01 AG024904). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: Abbott; Alzheimer’s Association; Amarin; Baxalta; Biogen; Elanco; Eli Lilly and Company;骏力; Genentech; Anne and Robert H. Horsley Foundation; Janssen; Merck; Novartis; Pfizer; Servier; Synarc; and Takeda. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (FNIH), which is managed jointly by banging byNIH, Department of Veterans Affairs, and Office of the Director, NIH. View the磁盘在这里.
Association study of longitudinal phenotypic markers to AD relevant SNPs

Association; Alzheimer’s Drug Discovery Foundation; Amorfix Life Sciences Ltd.; AstraZeneca; Bayer HealthCare; Bioclinica, Inc.; Biogen Idec Inc.; Bristol-Myers Squibb Company; Eisai Inc.; Elan Pharmaceuticals Inc.; Eli Lilly and Company; F Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; GE Healthcare; Innogenetics, N.V.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Medpace, Inc.; Merck & Co., Inc.; Mesro Scale Diagnostics, LLC.; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Servier; Synarc Inc. and Takeda Pharmaceutical Company. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector support are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education and the study is coordinated by the Alzheimer’s Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of California, Los Angeles. This research was also supported by NIH grants P30 AG010129, K01 AG030514 and the Dana Foundation.

REFERENCES

Argyriou,A. et al. (2007) Multi-task feature learning. NIPS, 41–48.
Ashburner,J. and Friston,K. (2000) Voxel-based morphometry—the methods. Neuroimage, 11, 805–821.
Braissant,S. et al. (2008) Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis. Human Mol. Genet., 18, 767–778.
Bramsichel,E. et al. (2009) A general and unifying framework for feature construction, in image-based pattern classification. Inf. Proc. Med. Imag., 21, 423–434.
Beck,A. and Touboul,M. (2009) A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci., 2, 183–202.
Brett,M. et al. (2007) Systematic meta-analyses of Alzheimer disease genetic association studies the alzheimer database. Nat. Genet., 39, 17–23.
Braak,H. and Braak,E. (1991) Neuropathological staging of Alzheimer-related changes. Acta Neuropathologica, 82, 239–259.
Braak,H. et al. (2009) Mapping the regional influence of genetics on brain structure variability—a tensor-based morphometry study. Neuroimage, 48, 37–49.
Budai,E. and Re bic,B. (2009) Exact matrix completion via convex optimization. FoCM, 9, 713–772.
Candès,E. and Tao,T. (2010) The power of convex relaxation: near-optimal matrix completion. Inform. Theory IEEE Trans., 56, 2035–2050.
Delacourte,A. et al. (1999) The biochemical pathway of neurofibrillary degeneration in aging and Alzheimer’s disease. Neurology, 52, 1156–1158.
Elston,R. et al. (2004) Least angle regression. Ann. Stat., 32, 407–499.
Filippin,N. et al. (2009) Anatomically-distinct genetic associations of APOE ε4 allele load with regional cortical atrophy in Alzheimer’s disease. Neuroimage, 44, 724–728.
Fischl,B. et al. (1999) Cortical surface-based analysis: ii: Inflation, flattening, and a surface-based coordinate system. Neuroimage, 9, 195–207.
Fischl,B. et al. (2002) Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron, 33, 341–355.
Glahn,D. et al. (2007) Neuroimaging endophenotypes: strategies for finding genes influencing brain structure and function. Human Brain Mapping, 28, 488–501.
Hani,R. et al. (2006) Imaging genetics: perspectives from studies of genetically driven variation in serotonin function and corticolimbic affective processing. Biol. Psychiatry, 58, 888–897.
Lehmann,M. et al. (2010) Reduced cortical thickness in the posterior cingulate gyrus is characteristic of both typical and atypical Alzheimer’s disease. J. Alzheimers Dis., 20, 567–598.
Li,Y. et al. (2010) Mach using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet. Epidemiol., 34, 813–834.
Liu,J. et al. (2009) Large-scale sparse logistic regression. In: SIGKDD09, Paris, France, pp. 547–556.
McEvoy,L. (2009) Alzheimer disease: quantitative structural neuroimaging for detection and prediction of clinical and structural changes in mild cognitive impairment. Radiology, 251, 195–205.
Mueller,S. et al. (2005) Ways toward an early diagnosis in alzheimer’s disease: the alzheimer’s disease neuroimaging initiative (adni). Alzheimers Dement., 1, 55–66.
Nichols,T. and Ashburner,J. (2009) Comparison of Whole Brain Multiloci Association Methods. Neuroimage, 47, 8161.
Nie,F. et al. (2010) Efficient and robust feature selection via joint l1-norm minimization. Adv. Neural Inform. Processing Syst., 23, 1813–1821.
Nie,F. et al. (2012) Schatten-p norm minimization for low-rank matrix recovery. In: Proceedings of Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI 2012).
Obst et al. (2006) Multi-task feature selection. Technical report, Department of Statistics, University of California, Berkeley.
Polkin,S. et al. (2009) Genome-wide strategies for discovering genetic influences on cognition and cognitive disorders: methodological considerations. Cogn. Neuropsychiatry, 14, 391–418.
Risacher,S. et al. (2010) Longitudinal MRI atrophy biomarkers: relationship to conversion in the ADNI cohort. Neurotol. Aging, 31, 1401–1418.
Saykin,A. et al. (2010) Alzheimer’s disease neuroimaging initiative biomarkers as quantitative phenotypes: genetics core aims, progress, and plans Alzheimers Dement., 6, 265–273.
Seshadri,S. et al. (2007) Genetic correlates of brain aging on MRI and cognitive test measures: a genome-wide association and linkage analysis in the Framingham Study. BMC Med. Genet., 8(Suppl 1), s15.
Shen,L. et al. (2011a) Sparse bayesian learning for identifying imaging biomarkers in AD prediction. Med Image Comput Comput Assist. Interv. (LNP 3), 611–618.
Shen,L. et al. (2011b) Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: a study of the ADNI cohort. Neuroimage, 53, 1051–1063.
Tibshirani,R. (1996) Regression shrinkage and selection via the LASSO. J. Royal. Statist. Soc B., 58, 267–288.
Wang,H. et al. (2011a) Identifying ad-sensitive and cognitive-relevant imaging biomarkers via joint classification and regression. In: Proceedings of the 14th Medical Image Computing and Computer-Aided Intervention (MICCAI 2011), Springer, Toronto, Ontario, Canada, pp. 115–123.
Wang,H. et al. (2011b) Sparse multi-task regression and feature selection to identify brain imaging predictors for memory performance. In: Proceedings of the 11th 2011 IEEE International Conference on Computer Vision (ICCV 2011), Barcelona, Spain, pp. 557–562, IEEE.
Wang,H.et al. (2012a) Identifying disease sensitive and quantitative trait relevant biomarkers from multi-dimensional heterogeneous imaging genetics data via sparse multi-modal multi-task learning Bioinformatics, 28, 1327–1336.
Wang,H.et al. (2012b) Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the adhd cohort. Bioinformatics, 28, 229–237.