

Sparse Clustering with Resampling for Subject Classification in PET Amyloid Imaging Studies

Wenzhu Bi, George C. Tseng, Lisa A. Weissfeld, Julie C. Price

Abstract—Sparse k-means clustering (Sparse_kM) can exclude uninformative variables and yield reliable parsimonious clustering results, especially for $p \gg n$. In this work, Sparse_kM and data resampling were combined to identify variables of greatest interest and define confidence levels for the clustering. The method was evaluated by statistical simulation and applied to PiB PET amyloid imaging data to identify normal control (NC) subjects with (+) or without (-) evidence of amyloid, i.e., PiB(+/-). **Simulations.** A dataset of $n=60$ observations (3 groups of 20) and $p=500$ variables was generated for each simulation run; only 50 variables were truly different across groups. The dataset was resampled 20 times, Sparse_kM was applied to each sample and average variable weights were calculated. Probabilities of cluster membership, also called confidence levels, were computed ($n=60$). Simulations were performed 250 times. The 50 truly different variables were identified by variable weights that were 13-32 times greater than those for the 450 uninformative variables. **Human Data.** For the PiB PET dataset, images (ECAT HR+, 10-15 mCi, 90 min) were acquired for 64 cognitively normal subjects (74.1 ± 5.4 yrs). Parametric PiB distribution volume ratio images were generated (Logan method, cerebellum reference) and normalized to the MNI template (SPM8) to produce a dataset of $n=64$ subjects and $p=343,099$ voxels/image. The dataset was resampled 10 times and Sparse_kM was applied. An average voxel weight image was computed that indicated cortical areas of greatest interest that included precuneus and frontal cortex; these are key areas linked to early amyloid deposition. Seven of 64 subjects were identified as PiB(+) and 47 as PiB(-) with confidence $\geq 90\%$, where another subject was PiB(+) at lower confidence (80%) and the other 9 subjects were PiB(-) at confidence in the range of 50-70%. In conclusion, Sparse_kM with resampling can help to establish confidence levels for clustering when $p \gg n$ and may be a promising method for revealing informative voxels/spatial patterns that distinguish levels of amyloid load, including that at the transitional amyloid +/- boundary.

I. INTRODUCTION

A pathological hallmark of Alzheimer's disease (AD) is amyloid-beta ($A\beta$) plaque deposition in brain. PET amyloid imaging with Pittsburgh Compound-B (PiB) is now widely used to detect fibrillar $A\beta$ deposition in living humans [1]-[2]. It is now established that amyloid deposition occurs in some (~30%) cognitively normal control (NC) subjects, well before there is evidence of cognitive problems [3]. It is of

major research interest to classify NC subjects as PiB(+) or PiB(-) to improve the early detection of $A\beta$ pathology in vivo.

Statistical clustering methods have long been used to identify biologically distinct subgroups from an $n \times p$ dataset without knowledge of the true group membership, where n is the number of observations and p is the number of variables observed. Since some variables are irrelevant to clustering and may only introduce noise, especially in the case of $p \gg n$, sparse clustering has recently been developed to exclude these variables and to yield more reliable and parsimonious clustering results. Recently in 2010, Witten and Tibshirani [4] introduced a general framework for sparse clustering that utilizes a Lasso penalty. Earlier in 2005, Tseng and Wong proposed a method for generating tight clusters [5] for the analysis of genetic data. This method combined resampling with clustering to construct robust tight clusters without forcing all observations into clusters such that some observations were not included. The method was developed to identify robust biologically distinct gene expression patterns.

In the present work, we propose a method to combine Witten's sparse k-means clustering with resampling to identify variables of greatest interest and to generate confidence levels for the clustering results. The performance of the proposed method was examined by statistical simulation. The method was also applied to a PiB PET amyloid imaging dataset to categorize subjects as PiB(+) or PiB(-). Voxel-level data were evaluated with an overall goal of identifying a subset of voxels that are most important for robust subject classification.

II. METHODS

The main idea of the proposed method is to resample the original dataset $X_{n \times p}$ for B times and apply sparse k-means clustering (Sparse_kM) to each sample. Each time we randomly sample without replacement 70% of the original n observations to form the sample dataset, X_i [5]. Sparse_kM is then applied to the sample X_i , a set of sparse clustering criteria is obtained for this sample, and these parameters are used to predict the cluster membership for the remaining 30% of the original dataset. The number of clusters, K , should be pre-specified and remain the same for each resampling run. Variable weights are assigned in a way such that the more a given variable contributes to clustering, the larger the weight; some variables have zero weights. The variable weights for a given run are standardized by dividing each weight by the tuning parameter for that run. The tuning parameter is the upper bound in the Lasso penalty that dictates the level of sparsity among the variables [4].

The final weight for a variable is then computed as the average of the weights across B samples. A confidence level

Manuscript received November 15, 2011. This work was supported by NIH (R01AG033042, P50AG005133, R01MH070729, R37AG025516, P01AG025204, K02AG027998), Dana Foundation and Alzheimer's Association.

Wenzhu Bi is with the Department of Biostatistics, University of Pittsburgh, Pittsburgh PA 15261 USA (telephone: 412-605-1552, e-mail: web10@pitt.edu). George C. Tseng and Lisa A. Weissfeld are with the Department of Biostatistics and Julie C. Price is with the Department of Radiology, University of Pittsburgh, Pittsburgh, PA, USA.

for cluster membership, defined as the observed proportion of each observation belonging to a given cluster, is then computed. Tight clusters can be constructed to only include observations associated with high confidence levels, such as $\geq 90\%$ or 95% . Some observations, therefore, may not fall in any of the tight clusters (remainder).

III. SIMULATIONS

Data were simulated as described by Witten and Tibshirani [4]. Briefly, for each of 250 simulation runs, an $X_{n \times p}$ dataset was created with $n=60$ observations (3 groups, 20 observations per group) and $p=500$ variables. Each value in the dataset was randomly generated from normal distributions with standard deviation equal to 1. The group differences corresponded to differences between the means of the simulated normal distributions, but only for the first $q=50$ variables. For the remaining $p-q=450$ variables, values were generated using the same standard normal distribution (mean 0, standard deviation 1) for all three groups. For the first $q=50$ variables, the mean values were -0.7 (group 1), 0 (group 2), and 0.7 (group 3). It is important to note that the choice of mean values ensured some group overlap that was needed to test the capability of the combined clustering and resampling method to define tight clusters and exclude uninformative variables. A total of 250 simulated datasets were generated.

Each simulated dataset was resampled $B=20$ times and Sparse_kM with $K=3$ clusters was applied to each sample. The final weight for each variable was computed as the average weight across $B=20$ samples. Probabilities of cluster membership were computed for each of the 60 observations to obtain confidence levels for the cluster membership.

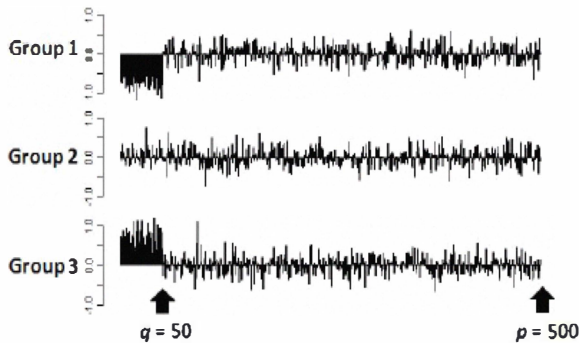


Fig. 1. Group averages for the 500 variables in an example simulated dataset $X_{60 \times 500}$. There are 3 groups with 20 samples per group. Groups differ in first $q=50$ of $p=500$ total variables.

Simulation results: Fig. 1. shows the group averages for the 500 variables in an example simulated dataset and reflects differences in the mean values for the first $q=50$ variables of each group (i.e., -0.7 , 0 and 0.7). Across 250 simulations, on average, the weights successfully identified the 50 variables (weight range: 0.012 - 0.016 , standard error: 0.0006 - 0.0018) that were truly different as indicated by weights that were about 13-32 times greater than those associated with the remaining uninformative 450 variables (weight range: 0.0005 - 0.0009 , standard error: 0.00003 - 0.0002).

TABLE I. NUMBER OF OBSERVATIONS IN TIGHT CLUSTERS FOR ONE SIMULATED DATASET

Confidence Level	Cluster 1	Cluster 2	Cluster 3	Remainder
[90% - 100%]	14 (1)	9 (0)	8 (0)	29
[70% - 100%]	21 (1)	17 (1)	19 (0)	3
(# misclassified)				

Table I shows the number of the observations in each tight cluster for one simulated dataset. With a confidence level $\geq 90\%$, i.e. at least 18 out of 20 times of resampling, there were 14 observations in cluster 1, 9 in cluster 2 and 8 in cluster 3 with one observation in cluster 1 being misclassified. The remaining 29 observations were left outside of these three tight clusters. When the confidence level threshold for the tight clusters was relaxed to be 70%, there were 21 observations in cluster 1, 17 in cluster 2, and 19 in cluster 3 with 2 observations (one in cluster 1 and the other in cluster 2) being misclassified. The 3 remaining observations were left outside of these clusters. These results reflect the overlap between groups in the original simulated dataset.

TABLE II. AVERAGE NUMBER OF OBSERVATIONS IN TIGHT CLUSTERS IN 250 SIMULATION RUNS

Confidence Level	Cluster 1	Cluster 2	Cluster 3	Remainder
[90% - 100%]	12 (0)	4 (0)	10 (0)	34
standard error	0.5	0.3	0.5	
[70% - 100%]	18 (0)	12 (0)	16 (0)	14
standard error	0.8	0.6	0.8	
(# misclassified)				

The clustering results for the 250 simulations were summarized based on the mean number of observations that appeared in the correct group, along with corresponding measures of standard error and confidence level (Table II). Also shown in Table II is the average number of misclassified observations (in parentheses) that was zero for each cluster. Across simulations, on average, there were a total of 26 of the 60 observations that were correctly classified in the tight clusters with confidence levels $\geq 90\%$. At a confidence level $\geq 70\%$, 46 of the 60 observations were correctly classified in the tight clusters. The standard error measurements indicated that the method performed consistently across the 250 simulations.

Overall, the simulation study illustrated how the proposed method works. It is demonstrated that the method was able to identify the subset of variables that were truly different. The method left some of the intermediate overlapping observations out of the tight clusters, which may serve as a more desirable approach than a clustering method that forces all observations into the clusters.

IV. HUMAN DATA

The proposed method was applied to $[^{11}\text{C}]\text{PiB}$ PET (or PiB) retention images in order to classify NC subjects as either PiB(+) or PiB(-) (based on the level of PiB binding to fibrillar amyloid). PiB PET data (ECAT HR+ 10-15 mCi, 90 min) were acquired in 64 NC subjects (Age 74.1 ± 5.4 years; 43 female, 21 male). The mini-mental state exam (MMSE) scores

for these subjects were 28 ± 2 , where 30 is a perfect score. MR images were acquired for co-registration, atrophy-related CSF dilution correction, and spatial normalization. The PiB PET data were analyzed on regional and voxel basis. Regions-of-interest (ROIs) were defined on the MR image and applied to extract regional PiB PET time-activity data for subsequent analysis. The primary ROIs included anterior cingulate (ACG), precuneus (PRC), frontal (FRC), parietal (PAR), lateral temporal (LTC) cortex and a cortical mean of these five regions (CTX5).

The reference Logan graphical method was applied to generate regional PiB distribution volume ratio (DVR, cerebellum reference) values and parametric DVR images (no CSF correction). Although this graphical method is prone to biases, particularly that arising from noisy voxel-level data, the method has demonstrated good test-retest reproducibility with regional and parametric image results that are consistent with results obtained using model-based approaches [6]. The parametric PiB DVR images were spatially normalized to the MNI template (Fig. 3., left) using SPM8. A dataset of $n=64$ subjects and $p=343,099$ voxels/image was then analyzed by the proposed Sparse_kM-resampling clustering method.

Human Data results: Fig. 2. shows the PiB DVR values (with and without correction for CSF dilution) for the 5 cortical regions and the cortical mean (CTX5). The PiB DVR cutoff values that were used to distinguish PiB(+) and PiB(-) subjects were generated using an iterative boxplot outlier approach as described in [3]. Without CSF correction, 25 subjects were PiB(+) as a result of one or more of the individual cortical DVR values exceeding its regional threshold value, while 15 were PiB(+) based on the single CTX5 global cortical mean cutoff.

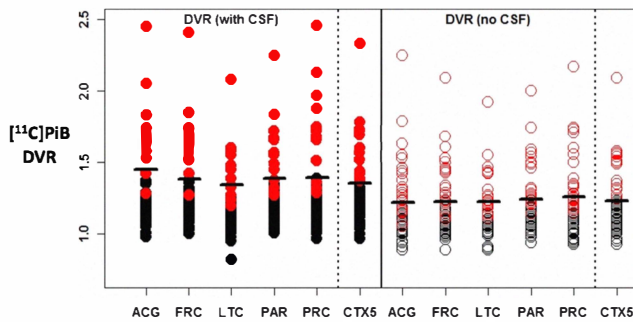


Fig. 2. Example of regional $[^{11}\text{C}]\text{PiB}$ DVR values for the 64 controls that shows range of retention from negligible (black) to high (red). The black line is PiB(+) cut-off based on single ROIs or CTX5 DVR values shown with and without correction for CSF dilution.

On a voxel basis, the clustering method was applied to the PiB DVR images of the $n=64$ subjects with $p=343,099$ voxels per image for each subject. The dataset was resampled 10 times (70% without replacement) and sparse k-means clustering with $K=2$ clusters was applied to each sample. The weights for each voxel were averaged (across the 10 resamplings) to obtain an average weight for that voxel that reflects its level of importance for the subject classification (Fig. 3., right). The magnitude and distribution of the voxels in the average weight image indicate cortical regions of that were

most informative in clustering subjects into two levels of PiB retention. These areas include precuneus, parietal and frontal cortex; these are areas that show high PiB retention in Alzheimer's disease and early deposition in NC subjects [7].

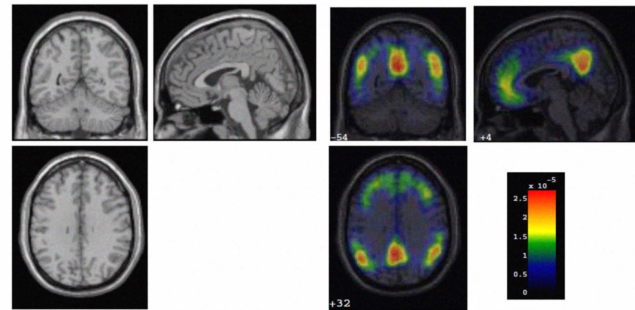


Fig. 3. Images of the MNI MRI template (Left) superimposed with the average weights from cluster analysis of $[^{11}\text{C}]\text{PiB}$ DVR images for 64 control subjects (Right). Bright voxels contribute most to classification of PiB(+) and PiB(-) subjects. Weight range: $0 - 2.71 \times 10^{-5}$, which is also 0 to 9.31 fold of $1/p$ ($1/p = 2.91 \times 10^{-6}$).

Among all 64 subjects, 7 subjects were identified as PiB(+) at a confidence level of 100% and another subject was positive at a confidence level of 80%. Forty-two subjects were PiB(-) with a confidence level of 100%, 5 subjects were negative with a confidence level of 90%, 7 subjects were negative with a confidence level of 70%, and 1 negative was with a confidence level of 60%. The last subject was positive or negative with a confidence level of 50%. With a confidence level threshold of 90%, we can construct tight clusters, where there were 7 subjects in the PiB(+) tight cluster and 47 subjects were in the PiB(-) tight cluster. The remaining 10 subjects were left outside of these two tight clusters. The lack of clear separation of the PiB(+) and PiB(-) groups could indicate that there is an intermediate group, which is of great current interest [7].

Fewer subjects were identified as PiB(+) by the voxel-level analysis, relative to the ROI approach. Factors that contribute to this difference include the fact that the ROI approach was based only on specific subsets of cortical brain voxels that were averaged across hemispheres, in contrast to the voxel approach that considered voxels throughout brain. Perhaps more importantly, there are substantial differences in the classification approaches. The Sparse_kM-resampling method provides stable tight cluster groupings defined at high confidence levels that results in more conservative criteria for amyloid-positivity than those obtained by the iterative boxplot outlier method that was used for the ROI approach.

V. CONCLUSIONS

The simulation study demonstrated that the proposed method of combining sparse clustering and resampling can help to stabilize variable selection and to establish confidence levels for the clustering results, which is especially important in the case of $p \gg n$. This approach appears to be promising as an objective method for identifying informative voxels that distinguish PiB(+) and PiB(-) subjects. This method may

provide useful insight on spatial patterns for subjects with different levels of amyloid load, including those at the transitional amyloid +/- boundary. For future work, more comprehensive statistical simulations will be performed to investigate method performance for different distribution parameters and to compare the proposed method to other types of clustering methods.

ACKNOWLEDGMENT

We thank Jeffrey James and Charles Laymon for their assistance.

REFERENCES

- [1] C. A. Mathis, Y. Wang, D. P. Holt, G. F. Huang, M. L. Debnath, and W. E. Klunk. "Synthesis and evaluation of ^{11}C -labeled 6-substituted 2-arylbenzothiazoles as amyloid imaging agents," *J Med Chem*, vol. 46, pp. 2740-54, Jun 2003.
- [2] W. E. Klunk, *et al.*, "Imaging the pathology of Alzheimer's disease: amyloid-imaging with positron emission tomography," *Neuroimaging Clin N Am*, vol. 13, pp. 781-9, ix, Nov 2003.
- [3] H. J. Aizenstein, *et al.*, "Frequent amyloid deposition without significant cognitive impairment among the elderly," *Arch Neurol*, vol. 65, pp. 1509-17, Nov 2008.
- [4] D. M. Witten and R. Tibshirani, "A framework for feature selection in clustering," *J Am Stat Assoc*, vol. 105, pp. 713-726, Jun 2010.
- [5] G. C. Tseng and W. H. Wong, "Tight clustering: a resampling-based approach for identifying stable and tight patterns in data," *Biometrics*, vol. 61, pp. 10-6, Mar 2005.
- [6] M. Yaquib, *et al.*, "Simplified parametric methods for ^{11}C PIB studies," *Neuroimage*, vol. 42, pp. 76-86, Aug 2008.
- [7] E. C. Mormino, *et al.*, "Not quite PIB-positive, not quite PIB-negative: Slight PIB elevations in elderly normal control subjects are biologically relevant," *Neuroimage*, Aug 2011.