

Sparse Methods for Dimensionality Reduction in Medical Imaging

Ben Kandel

Specific Aims

In this proposal, we aim to advance linear sparse data decompositions by 1) explicitly enforcing smoothness and contiguity constraints; 2) incorporating prior data. In addition, we aim to 3) extend nonlinear dimensionality reduction techniques to incorporate sparsity, which as far as I can tell has not been attempted in the way we're doing it before. Finally, 4) we aim to apply the methods developed to a dataset of Alzheimer's Disease patients (not ADNI). Specifically:

Specific Aim IA: Contiguous and Grouped Linear Sparse Medical Image Decomposition (to be combined with next section)

Sparsity in matrix decompositions, unless otherwise regularized, can often result in a small number of voxels scattered throughout a medical image. As a result, sparse decomposition techniques that are applied to data in which spatial information is important, such as genomic data (if the location on a chromosome is important) or image data, usually have some sort of constraint that encourages the data to be distributed in only a few relatively large areas. We propose to use an explicit smoothing constraint to further ensure that our decompositions are spatially coherent. Using this smoothing constraint enables us to relax other requirements of methods that do not have an explicit spatial coherence constraint.

Specific Aim IB: Group Sparse Matrix Decomposition

One issue that sparse data decompositions have is that they often do not offer a convenient way to incorporate prior knowledge or labels into the decomposition. We propose to use a group data decomposition algorithm. This group algorithm will enforce sparsity only on labels (i.e., only a few pre-defined regions of the brain will be allowed to be included in the decomposition), but will not enforce sparsity *within* a given region, so that within a region, all voxels of the region may vary freely. This will encourage sparsity in an anatomically-informed way, so that anatomically meaningful regions as a whole will be allowed to stay in the model.

Specific Aim II: Non-Linear Sparse Dimensionality Reduction

We propose to extend nonlinear dimensionality reduction techniques to the sparse setting. Non-linear dimensionality reduction techniques are useful for analysis of data that does not lie on a linear subspace of the original, high-dimensional data. Such nonlinear dimensionality reduction techniques have proven very useful in analysis of image data. In this work, we propose a sparse extension of one of the most popular non-linear dimensionality reduction techniques, Locally Linear Embedding (LLE). We also demonstrate how the ideas presented here can be easily applied to other related non-linear dimensionality reduction techniques.

Specific Aim III: Application to Neurodegenerative Diseases

Although the dimensionality reduction techniques outlined in the previous section are of general interest and can be applied to many varieties of image data, we are specifically interested in medical applications of the data. Traditionally, medical image analysis is performed on a voxel-wise, mass univariate basis. This technique, however, is not ideally suited for prediction of disease or clinical data because of the very large dimensionality of the analyzed data. We propose to use the dimensionality reduction procedures outlined above to predict the rate of cognitive decline in patients suffering from neurodegenerative diseases, including frontotemporal dementia and Alzheimer's Disease.

Background, Significance, and Innovation

Medical imaging forms an increasingly indispensable role in medical diagnosis and disease characterization because it offers the ability to gain detailed information about the structure and function of patients' bodies. At the same time, though, the richness and subtleties in medical images present significant methodological challenges for applying traditional statistical methodologies to medical imaging data. Because of the extremely high dimensionality and spatial information contained within medical images, traditional regression and classification techniques are not ideally suited for using medical images to make predictions as to a patient's disease state or clinical data. Voxel-based morphometry (VBM) [1], one of the most widely used methods for analyzing neuroimaging data, looks for voxel-wise differences between images. Because the significance maps that are the output of VBM are still very high-dimensional, the output from VBM is not ideally suited to predictions of disease state or clinical data, and some other dimensionality reduction technique is still required. In addition, because VBM operates in a mass-univariate manner, it may lose some statistical power that a multivariate statistical method may recover.

Although dimensionality reduction is helpful for

Tra

medical imaging data into traditional statistical methods. Medical images offer information about patients that is impossible to obtain using other non-invasive means. Medical imaging offers information about a patient's anatomy, physiology, and other characteristics that is impossible to otherwise obtain. However, quantitative methods for analyzing medical images are hampered by the extremely high dimensionality of medical images. Because medical images can easily contain more than a million voxels, the extremely high dimensionality presents significant issues for incorporating medical images into statistical and machine learning methodologies. When faced with data of such high dimensionality, dimensionality reduction techniques are often used to convert the high-dimensional data into a form that is more amenable to statistical analysis.

In high-dimensional data analysis, sparsity, or parts-based representation, is often used to encourage dimensionality reduction or other analysis of high-dimensional data to have a relatively small number of non-zero components in the final solution. In terms of the data, this means that only a small portion of the dimensions of the original data have an impact on the final solution. In terms of medical imaging, this means that only a relatively small proportion of the image (and therefore anatomical region captured by the image) is retained in the final solution. Enforcing sparsity on solutions of dimensionality reduction means that only a subset of the data is retained for the final reduction. This enables solutions to be clinically interpretable—because only a small portion of the image is retained, it is possible to link that part of the image with a known anatomical function and thereby gain insight into the relation between a region and some clinical data.

Literature Cited

- [1] J Ashburner and K J Friston. Voxel-based morphometry—the methods. *NeuroImage*, 11(6 Pt 1):805–821, June 2000. PMID: 10860804.