Graph-Based Shape Analysis for MRI Classification

Seth Long, Washington State University, USA

Lawrence B. Holder, Washington State University, USA

ABSTRACT

Searching for correlations between brain structure and attributes of a person's intellectual state is a process which may be better done by automation than by human labor. Such an automated system would be capable of performing classification based on the discovered correlation, which would be means of testing how accurate the discovered correlation is. The authors have developed a system which generates a graph-based representation of the shape of the third and lateral ventricles based on a structural MRI, and classifies images represented in this manner. The system is evaluated on accuracy at classifying individuals showing cognitive impairment to Alzheimer's Disease. Classification accuracy is 74.2% when individuals with CDR 0.5 are included as impaired in a balanced dataset of 166 images, and 79.3% accuracy when differentiating individuals with CDR at least 1.0 and healthy individuals in a balanced dataset of 54 images. Finally, the system is used to classify MR images according to level of education, with 77.2% accuracy differentiating highly-educated individuals from those for whom no higher education is listed, in a balanced dataset of 178 images.

Keywords: *Alzheimer's, Automatic Classification, Brain, Education, Frequent Subgraph Discovery, Graph, Graph Mining, Machine Learning, Ventricle*

The physical structure of the human brain is highly variable from one individual to another. Some of these variations correlate with known psychological characteristics and life experiences, such as a high level of education, or a childhood traumatic brain injury. In other cases, the effect of structural characteristics is unknown. Likewise, some psychological phenomena are known to correlate with differences in brain structure, some with neurotransmitter levels or other chemical changes, and others do not correlate with any known physical phenomena. Discovering what these correlations are is an active problem in neuroscience.

DOI: 10.4018/jkdb.2011040102

An automatic system to discover these correlations could improve the size of dataset which may be feasibly considered and reduce the cost of examining the data in terms of human time. Such a system should be capable of examining the shape of a particular structure, finding correlations between shape and neurological conditions, and classifying images based on the discovered correlations. A successful automatic system should have general utility for varying classification criteria.

Analyzing the shape of structures in the brain can be a particular challenge due to lack of clear, defined boundaries around many of them, as noted in Nestor et al. (2008). However, the ventricular system provides a sharp contrast to the rest of the brain, appearing dark in color in MR images due to the cerebrospinal fluid inside the ventricles. This provides a natural target for automatic analysis. Given this, our system classifies images based on the shape of the ventricular system. The shape of the third and lateral ventricles is represented as a graph, which is then classified using graph classification techniques.

The volume of the ventricular system has been shown to increase with the progression of Alzheimer's Disease (Nestor et al., 2008). As such, a system which can use the shape of the ventricular system to classify images should be successful on recognition of Alzheimer's Disease. The hippocampus is known to shrink with the progression of Alzheimer's Disease (Du et al., 2001), which may indicate that the shape of the ventricular system changes early in the progression of the disease, expanding in the direction of the hippocampus. This leaves the possibility open that shape will be a more accurate predictor of Alzheimer's Disease than overall volume. Regardless, success on the task of automatic Alzheimer's Disease classification is an indication that a classification system may be viable for discovery of other correlations. Figure 1 gives examples of healthy and impaired MRI scans.

This paper describes a method of creating a tree representing the 3D shape of the third and lateral ventricles, by both the tree structure and node and edge labels. Trees created by this method are then classified by finding frequent subtrees which are present only in trees derived from cognitively-impaired or healthy brains. Feature vectors are constructed for each image by presence or absence of the frequent subtrees, and the image is classified by a support vector machine operating on these feature vectors. The hypothesis that such a system should be capable of distinguishing with better than random accuracy between cognitively-impaired and healthy individuals is tested and found to be true, with an accuracy of 79.3% when mildly impaired individuals (CDR 0.5) are excluded from the test, and 74.2% when they are included. All images used are part of the OASIS dataset (Marcus et al., 2007).

A classifier which can accurately classify structural MRI scans based on ventricular shape can be expected to perform well on classification of Alzheimer's Disease vs. healthy individuals. However, such a system may have other applications of general utility. For example, taxi driving is known to affect hippocampus size (Maguire et al., 2000). This is a fairly specific case, but learning in general has also been found to affect brain shape and structure (Blakemore & Frith, 2005). Given this, it may be possible to determine something about an individual's level of education based on the shape of the ventricular system. After testing our system on Alzheimer's Disease recognition, we apply it to classification based on level of education, with the result that classification based on this measure is possible. Our system demonstrates 77.2% accuracy classifying individuals based on a high level of education (4-5 years) and no listed education. This validates the hypothesis that the system can perform classification using a variety of class labels.

Previous Work

Some previous work has been focused on the particular problem of automatic recognition of Alzheimer's Disease from MRI data. For example, Klöppel et al. (2008) consider each voxel to be a feature in a feature vector, and then use a support vector machine to classify the resulting feature vectors. Cocosco et al. (2003) consider features of the image which are points in the gray matter, white matter, or CSF, and prune the resulting set of features to develop a minimum spanning tree which can be used for classification based on a clustering approach. Cuingnet et al. (2010) discuss and compare 10 different methods using a large dataset from 509 participants. As such, the study of automatic detection of Alzheimer's Disease is well-studied. In contrast, our study of classification of Alzheimer's Disease is used to validate a general-purpose classifier, which is

Related Content