

INFO411/INFO911

Data Mining and Knowledge Discovery

Project 5

Instructions:

You are required to prepare a set of presentation slides which must include (1) the full name and student number of each student in the group, the contribution (in percent) of each group member, (2) a description of the task, (3) your proposed data mining approach and methodologies; (4) your results and an analysis of your results; (6) the results a brief discussion and a conclusion.

Below is the recommended structure of your slides:

- Introduction (define the problem and the goal)
- Methods (propose approaches, and discuss their strengths and weaknesses)
- Results (include illustrative Figures and Tables and explanations)
- Discussion (discovered knowledge?)

Task: Data Reduction in Brain Image Classification Problems

Background:

The data available from:

<http://www-2.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/>

contains a time series of images of brain activation (fMRI scans). Images are taken at 500msec intervals during which a human subject performs the task of reading a sentence, observing a picture, and determining whether the sentence correctly described the picture. Each human subject is presented with 40 sentence & picture pairs. Each of these 40 exercises last approximately 30 seconds. The dataset contains the scans of 12 human participants. Each of the images is being described by about 5,000 3D pixels (voxels) measuring the brain activity across a portion of the brain.

The dataset has been used to detect cognitive states of the brain – a classification task.

Definition of the task:

Given the extremely high dimension of the input (5000 voxels times 8 images) to the classifier, it is sensible to explore methods for reducing the dimensionality. For example, consider PCA, hidden layers of neural nets, Self-Organizing Maps, and other relevant dimensionality reducing methods. Perform a thorough investigation. Which dimensionality reduction method works better for brain image classification systems?

Thus, the task is to deploy and compare dimensionality reduction methods for use in models that detect the cognitive states in brains. You will therefore need to show how the various dimensionality reduction methods affect the results of a classifier.

Relevant reading on the classifiers is available at the end of the following web-page:

<http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/>

with a particular emphasize on the paper “Learning to Decode Cognitive States from Brain Images” by T.M. Mitchell.

Requirements:

1. Present a general description of the corpus and present its general properties.
2. Describe the method(s) that you used for this task. Discuss the strengths and limitation of these methods.
3. Present and discuss your results.
4. Summarize: What new and interesting things did you discover while working on this project?