ADNI Progress report

Devendra Goyal

January 19, 2015

Part I

Using HMMs to predict disease progression

1 Open Problems/Possible venues

Transfer Learning/Manifolds Due to difference in protocols between MRI images from ADNI1 and ADNIGO/ADNI2 (1.5T vs 3T) we are unable to combine information from all three cohorts to learn a unified classifier/model.

A recent work explores this theme by learning a shared manifold using data from all 3 protocols. Not only does this utilize all data, but it simultaneously projects the data onto a more manageable lower dimension.

Missing data This problem manifests in two ways; particular patients missing certain visits, and the more severe problem of every patient not having data from every modality. This severely restricts any approaches that attempt to use multi-modal approaches for classification.

A recent work attempts to solve the problem by learning separate classifiers on groups of data (e.g. MRI+PET, MRI+CSF, etc.) and combining information from separate classifiers by using ensemble approaches.

Another simple solution to the problem worth exploring is the use of whatever data we have from each visit. In particular, rather than taking the intersection of visits from MRI, PET and CSF, we could simply use every visit that has any of the three modalities available to us. This would likely involve some kind of joint optimization over parameters or some minor mathematical tweaking of a traditional problem formulation (however this seems to be inexistent in the literature for some reason...).

Unsupervised Feature Learning We currently rely on features that are highly pre-processed. For instance, the only PET data available to us is the mean tracer retention in 5 ROIs in the brain. These ROIs are selected mainly on the basis of positive correlation with clinical disease label. However, such a highly processed dataset maybe obfuscating useful information that can be mined effectively.

Noisy Labels/Lack of Gold Standard Often, the clinical labels assigned to patients (NL/MCI/AD) are noisy/incorrect. Furthermore, labeling patients as MCI-converters/MCI-non-converters is problematic due to the incomplete information available to us.

The only method of obtaining gold standard diagnosis for patients (currently) is through post-mortem analysis of brain tissue, and this information is currently only available for a very small number of patients.

Amyloid-retention in brain Amyloid retention in PET-scans is a recently popularized modality that provides us information about the level of amyloid deposits in the brain. However, there is very little consensus about the effects of amyloid on the onset of AD/conversion likelihood from MCI. That is, how are amyloid-positive patients diferent from amyloid-negative patients in terms of risk of conversion/chance of getting Alzheimer's? There has been work showing a weak correlation between onset of AD and amyloid-deposits, however there is likely a need to consider more variables in more non-linear models.

Evaluating HMMs Common approaches to evaluating similar unsupervised learning approaches in the literature (besides those we already explored) are to threshold the unsupervised models at some probability/state and evaluate it as a discriminative classifier. Another common approach is to

use the Pearson correlation coeffecient by examining correlation with clinical scores such as the MMSE.

Non-gaussian nature of clinical data Clinical scores (MMSE/CDR/ADAS-Cog) are scored in a binary manner (correct/incorrect), with the scores then being aggregated over several questions. A Gaussian emission model fails in such situations, and a Bernoulli distribution might be better suited. This is currently ongoing.

2 Results

Using δPET as features for the HMM.

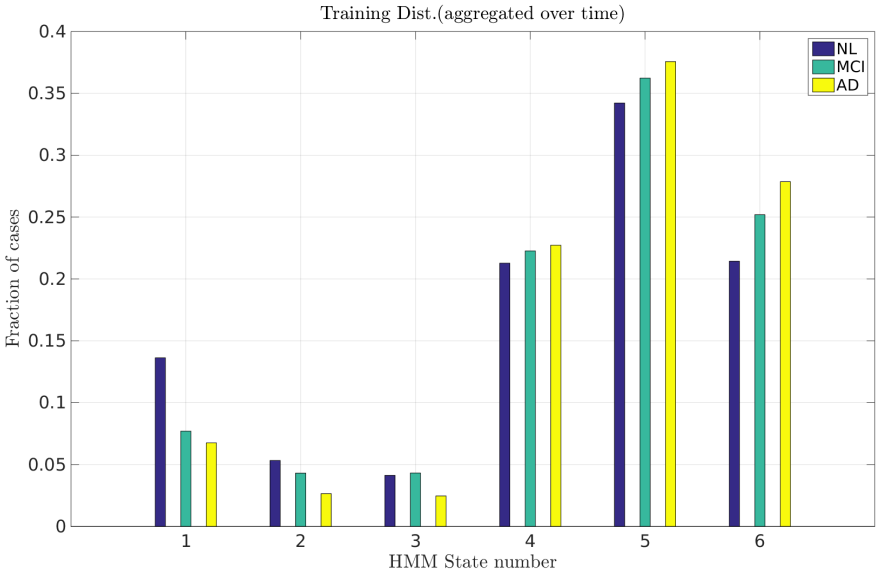

Figure 1: Distribution of labels given HMM state (Training set)

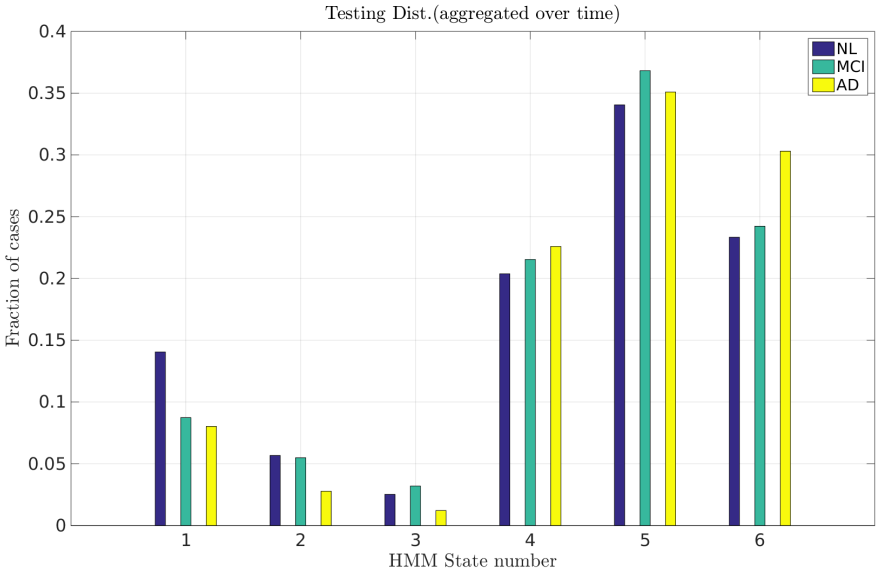

Figure 2: Distribution of labels given HMM state (Test set)

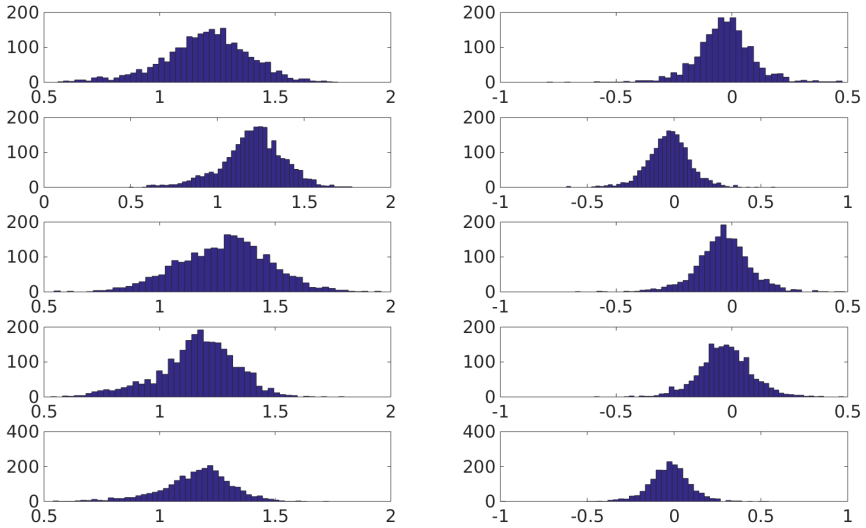

Figure 3: Distribution of values in the PET scans (left column), and the distribution of values in δPET scans (right column). Each row represents a particular ROI in the brain