Supplementary Material
HYDRA: revealing Heterogeneity of imaging and genetic patterns through a multiple max-margin Discriminative Analysis framework

Erdem Varol[a,*], Aristeidis Sotiras[a], Christos Davatzikos[a],
for the Alzheimer's Disease Neuroimaging Initiative[1]

[a]*Section for Biomedical Image Analysis*
Center for Biomedical Image Computing and Analytics
University of Pennsylvania
Philadelphia, PA 19104, USA

Abstract

This serves as a supplement to "HYDRA: revealing Heterogeneity of imaging and genetic patterns through a multiple max-margin Discriminative Analysis framework". We provide here additional evidence for the presence of heterogeneity in the ADNI dataset by examining the classifier margin, as well as the reproducibility and hierarchy of the clustering result. We compare these values to the ones we find when applying HYDRA to a simulated non-heterogeneous dataset. In this case, contrarily to what we observe for the ADNI datasets, the margin does not increase for higher values of K and the clustering is not reproducible.

1. Additional measures of heterogeneity

To provide additional evidence for the ability of HYDRA to disentangle heterogeneous signal in neuroimaging and genetic datasets, we applied HYDRA to a simulated non-heterogeneous dataset and compared its performance to the performance in the ADNI datasets.

We used four evaluation measures to quantify the performance of the algorithm and assess the amount of heterogeneity (or lack thereof) in the simulated dataset and the ADNI datasets: 1) area under the curve (AUC), 2) adjusted Rand index (ARI), 3) cross-validated classification margin, and 4) clustering hierarchy. The AUC and the margin quantify the classification performance, while ARI and the clustering hierarchy characterize the clustering performance.

While AUC measures the classification performance, the margin is a related measurement that quantifies the degree of separation between the controls and patients. In cases where the datasets are very separable (*e.g.*, high dimensional settings), AUC may reach a plateau, while the margin may still increase. Likewise, while ARI measures the stability of clustering at a particular level, clustering hierarchy provides additional insight as to this stability by showing that clustering splits have structure.

The margin is defined as the separation between the control and patient populations. Namely, if $y_i \in \{-1, +1\}$ is the label of each sample with patients having label -1,

*Corresponding author at: Section for Biomedical Image Analysis, Center for Biomedical Image Computing and Analytics, University of Pennsylvania, 3600 Market Street Suite 380, Philadelphia, PA 19104, USA. Fax:+1 215 614 0266.

Email address: erdem.varol@uphs.upenn.edu (Erdem Varol)

[1]Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

and $\mathbf{x}_i$ is the features and $\{\mathbf{w}_j, b_j\}$ is the jth face of the polytope classifier of HYDRA, the margin is defined as:

$$m = \min_i y_i(\min_j \mathbf{w}_j^T \mathbf{x}_i), \quad (1)$$

which is the signed distance between the worst classified sample and the classifier boundary. Thus, a large positive margin indicates very well separated populations, while a negative margin indicates the presence of misclassified samples.

We evaluate clustering hierarchy both qualitatively and quantitatively. The qualitative way is to visualize in the form of a grid the clustering memberships of each patient evaluated by HYDRA at different clustering levels K. The rows of the grid correspond to the patients, while the columns correspond to clustering results obtained at different K. For each column, the clustering assignment of each patient is encoded by a different color, and the same color within each column indicates co-clustered patients. Thus, observing block repetition from column to column indicates that groups of co-clustered patients at one particular level, also co-cluster at a subsequent finer level of clustering. This is a qualitative indication of clustering hierarchy.

In order to quantitatively assess the hierarchy of the clustering, the ARI is calculated between clusters estimated at successive levels. If the clustering is hierarchical, high ARI is expected to occur between successive levels of clustering since the split should have structure. Otherwise, if the clustering is not hierarchical, the clusters at successive levels are expected to form independently from one another, yielding a low ARI.

In a high dimensional space, such as the one encountered in neuroimaging and genetics, it is very likely that the control and patient populations will be linearly separable. Therefore, introducing non-linearity, as done in HYDRA, may not greatly increase the classification performance. However, since HYDRA's objective is to find clusters that increase the margin, the clustering results are inextricably linked to the margin. Therefore, observing an increase in margin coupled with reproducible and hierarchical clustering is a sign that there is heterogeneous signal in the dataset that is revealed by the algorithm. Likewise, based on the high dimensionality argument, stable AUC is not a sufficient indication to confirm, or contradict, the existence of heterogeneity in a dataset since the non-linearity may not improve separability if the data are already linearly separable. In cases where AUC does not increase, the margin, the clustering reproducibility and the clustering hierarchy may shed light on whether there is a heterogeneous signal that is picked up by the method.

2. Results

The AUC, ARI, margin and clustering hierarchy of the simulated non-heterogeneous dataset and the ADNI datasets were compared. The non-heterogeneous dataset was simulated in the following fashion. 50 control subjects were drawn from a 1000 dimensional Gaussian distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{I}$, and another 50 patient subjects were drawn from a 1000 dimensional Gaussian distribution with mean $0.1 \times \mathbf{1}$ and covariance matrix $\mathbf{I}$. By construction, this simulated dataset does not possess any heterogeneity in the patient group since the samples all differ from the controls along one direction.

Table 1 reports the values for the different measures in the case of the simulated non-heterogeneous dataset, as well as the ADNI structural MRI and genetic datasets. While we observe a stable level of AUC for increasing values of K in all three cases (see Table 1, first row), the distributions of the ARI, the margin and the clustering hierarchy differ significantly between datasets.

While in the case of the ADNI datasets we observe important clustering reproducibility indicated by ARI, in the case of the simulated non-heterogeneous dataset the ARI hovers close to zero for all values of K (see Table 1, second row). This indicates that the algorithm randomly splits the data for all values of K.

Furthermore, while the margin is steadily increasing with K for the ADNI datasets, the margin remains stable as K increases for the simulated dataset (see Table 1, third row). In other words, in the ADNI datasets, adding additional clusters, or hyperplanes, allows HYDRA to harness the heterogeneous structure of the data and increase the separation between the groups. On the contrary, as expected, adding more clusters does not improve the separation between the groups for the simulated dataset.

Lastly, and most importantly, we observe that the clustering assignments in both the structural MRI dataset and the genetic ADNI dataset have block repetitions from column to column. This indicates that patients that co-cluster

at one level tend to consistently co-cluster at successive finer levels (see Table 1, fourth row). However, such a hierarchy is not readily observable in the simulated non-heterogeneous dataset. This observation is further quantified by the ARI measured between clusterings at successive levels (see Table 1, fifth row). In the ADNI datasets, there is a high agreement between the clusters obtained at successive levels. This indicates that clusters obtained at higher levels are more likely to be subsets of the clusters at lower levels. Such a structure is not observed for the simulated non-heterogeneous dataset. This means that clustering obtained at higher levels does not bear any resemblance to those obtained at lower levels due to lack of structure in the data.

3. Conclusion

Comparing the ARI, the margin and the clustering hierarchy results for the simulated non-heterogeneous dataset and the ADNI datasets provides evidence that there is heterogeneous structure in the ADNI datasets which HYDRA is able to disentangle. On the other hand, HYDRA does not give a false indication of heterogeneity in the case that there is none as for example in the simulated dataset. Additionally, the comparison of the AUC for the three datasets shows that the AUC is not an indicator for presence/absence of heterogeneity, but rather a test to control that the non-linearity that is introduced by HYDRA does not lead to over-fitting.

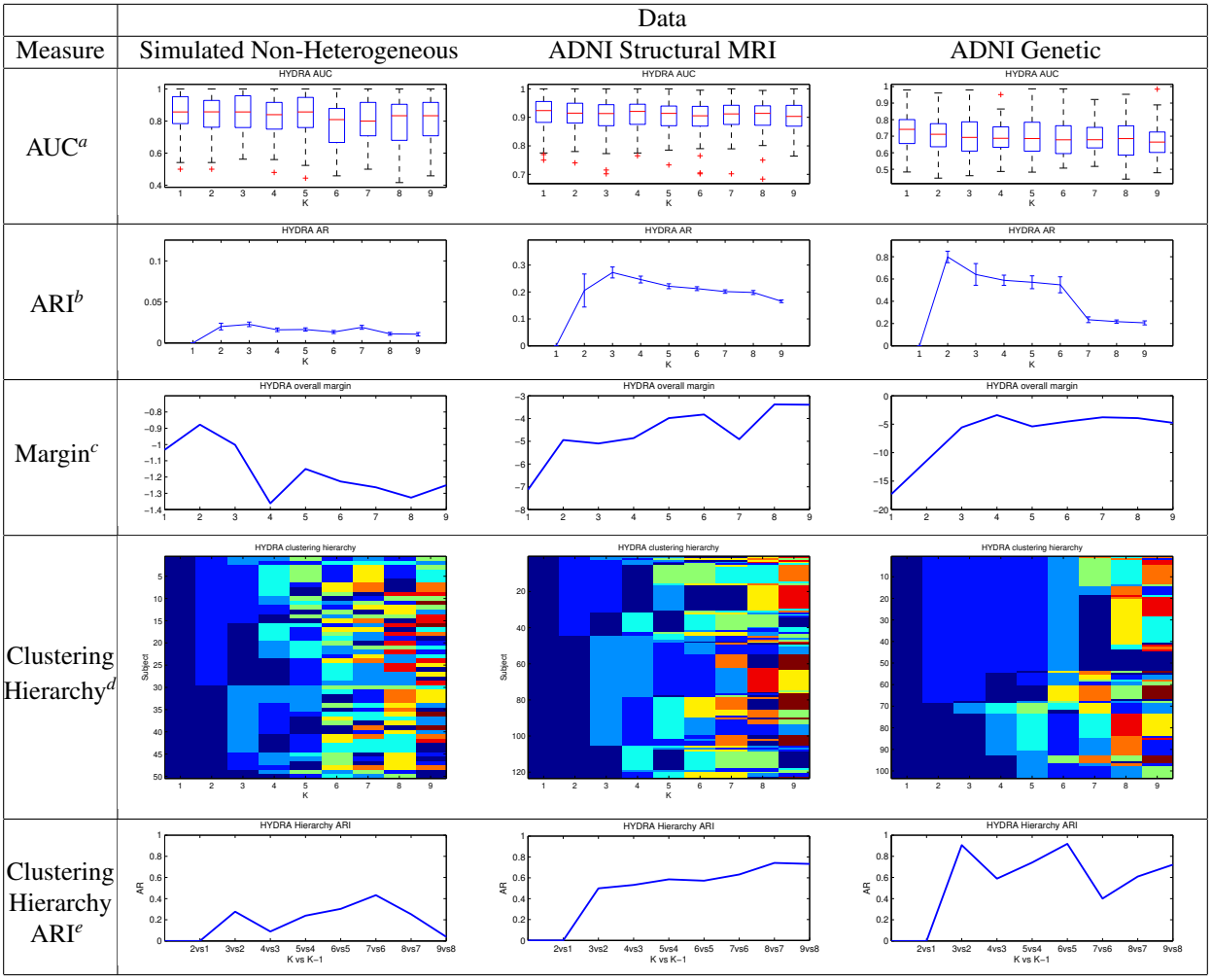

Table 1: Comparison of classification AUC, clustering reproducibility, classifier margin and clustering hierarchy for the three different datasets at varying levels of K for HYDRA.[a]AUC—Area under ROC curve, [b]ARI—Adjusted Rand Index, higher values of ARI indicate more reproducible clustering, values close to zero indicate near random clustering. [c]Margin—Cross validated classification margin in Eq. (1). This shows the separability between patients and controls for increasing values of K. [d]Clustering hierarchy — Cluster membership of patients (rows) at different levels of clustering (columns) indicated by colors corresponding to different clusters. Same colors within columns indicate co-clustered patients. Block repetition between columns indicates that patients that co-cluster at a particular level of clustering, also co-cluster at another level of clustering, resulting in a hierarchical structure. [e] Clustering hierarchy ARI — The ARI between the clustering results at successive levels of clustering. High overall values indicate a highly hierarchical clustering, while low values indicate unstructured splitting. These results demonstrate that, due to lack of structure, HYDRA does not find reproducible and hierarchical clustering for the simulated non-heterogeneous data. Moreover, in the non-heterogeneous case, the margin does not increase for increasing K. On the other hand, there is underlying structure in the ADNI datasets. HYDRA is able to take advantage of this structure and produce a hierarchical and reproducible clustering while improving the separation between the classes.