Robust AI Techniques for Large-Scale Surveys of Space and Planetary Science Data Sets

P. R. Gazis

The Problem

- Many problems in space and planetary science involve large-scale surveys, followed by analysis of the results.
- As data sets have grown larger, these surveys have grown difficult to perform.
  - Long multi-variate time series
  - Hyperspectral images
  - CMB data
- While it may be possible to use a priori knowledge to devise specialized detection algorithms for particular types of events (e.g., particle physics and cosmic ray analysis code) this is not always practical.

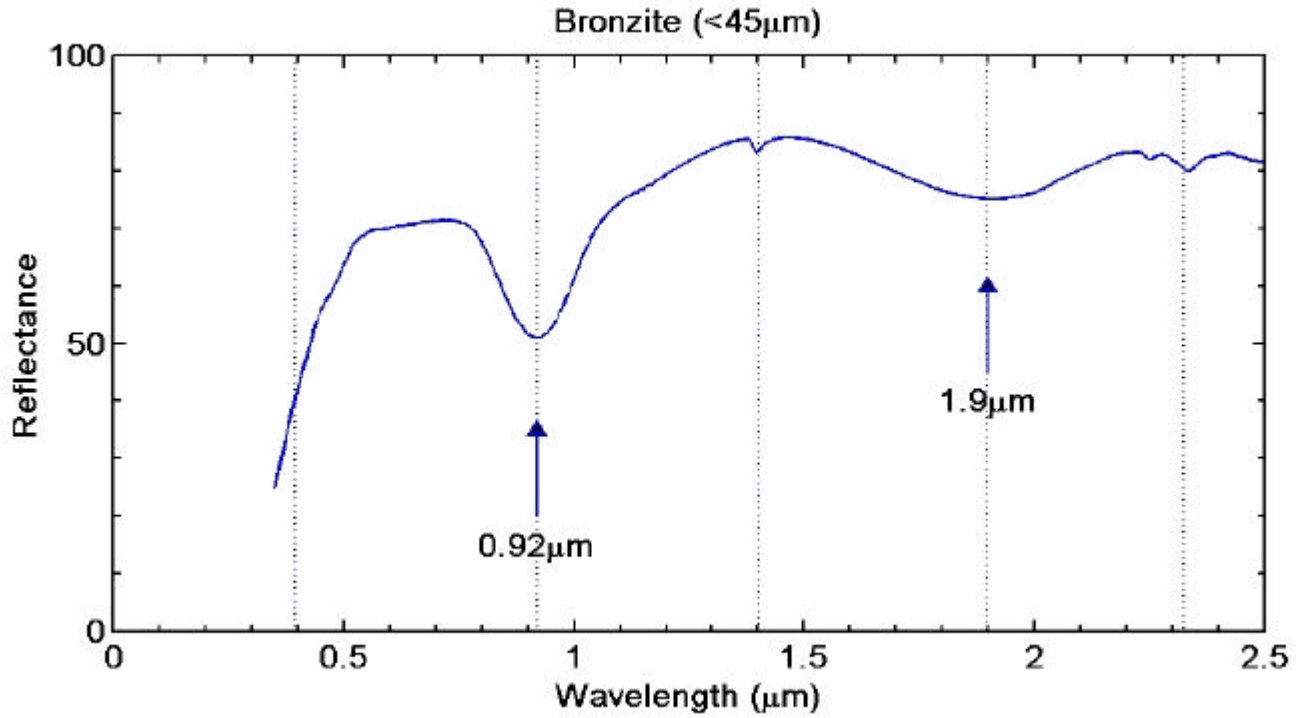
One Possible Solution

- Separate the survey and analysis parts of the problem.
- Use well-understood traditional AI methods such as expert systems or neural networks to perform the survey.
- Be willing to accept some loss of sensitivity or a modest number of false positives, depending on the application.
- Use the results as-is, if they are sufficiently well-understood, or…
- Perform the analysis step on the reduced data set, at a substantial saving in time and effort.

- Examples:
  - Self-organizing maps for the detection of interplanetary shocks
  - Expert systems for the classification of hyperspectral data
  - Self-organizing maps for clustering

Expert Systems for the Classification of Hyperspectral Data

- Apply a straightforward algebraic scheme to identify local minima and inflection points that might be associated with spectral bands in near-IR reflectance spectra



Classification of Hyperspectral Data

- Characterize the bands as lists of attributes

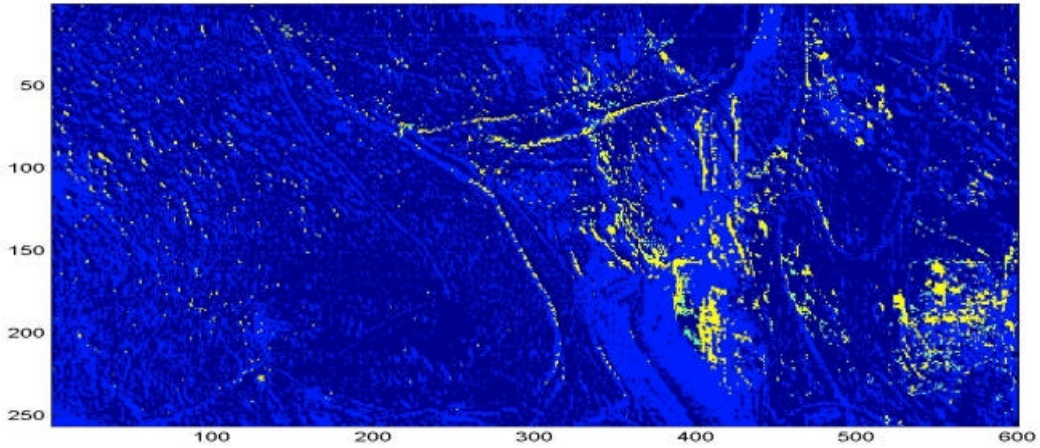
```
((ANIR000   1 0.395  2.548 0.350  0.000 0.450  0.000 0.099  1.243 0.100 T)
 (ANIR001   2 0.920 24.094 0.590  0.000 1.343  0.000 5.779  0.754 0.753 T)
 (ANIR002   3 1.401  1.641 1.365  0.000 1.459  0.000 0.061  0.892 0.094 T)
 (ANIR003   4 1.897  9.178 1.474  0.000 2.207  0.348 3.609  1.254 0.733 T)
 (ANIR004   5 2.241  1.130 2.174  0.766 2.417  0.000 0.209  1.547 0.243 T))
```

- Use prior knowledge and empirical tests on libraries of sample data to develop a rule set that can identify spectra that might be associated with particular mineral classes of interest

```
(RULE001 (IF   (ANIR* * (0.92 +/- 0.05) ( > 0.0) * * * * * * (> 0.25) T)
               (ANIR* * (2.22 +/- 0.07) ( > 0.0) * * * * * * * T)
               (ANIR* * (2.55 +/- 0.07) ( > 0.0) * * * * * * * T))
         (THEN (SAMPLE IS JAROSITE)))
(RULE002 (IF   (ANIR* * (1.00 +/- 0.10) ( > 3.0) * * * * * * ( > 0.3) T))
         (THEN (SAMPLE IS IRON_BEARING)))
(RULE003 (IF   (ANIR* * (0.92 +/- 0.05) ( > 0.0) * * * * * * * T)
               (ANIR* * (1.90 +/- 0.05) ( > 0.0) * * * * * * (> 0.6) T))
         (THEN (SAMPLE IS PYROXENE_HIGH_FE)))
(RULE004 (IF   (ANIR* * (1.00 +/- 0.05) ( > 0.0) * * * * * * * T)
               (ANIR* * (2.20 +/- 0.05) ( > 0.0) * * * * * * (> 0.6) T))
         (THEN (SAMPLE IS PYROXENE_LOW_FE)))
(RULE005 (IF   (SAMPLE IS PYROXENE_HIGH_FE))
```

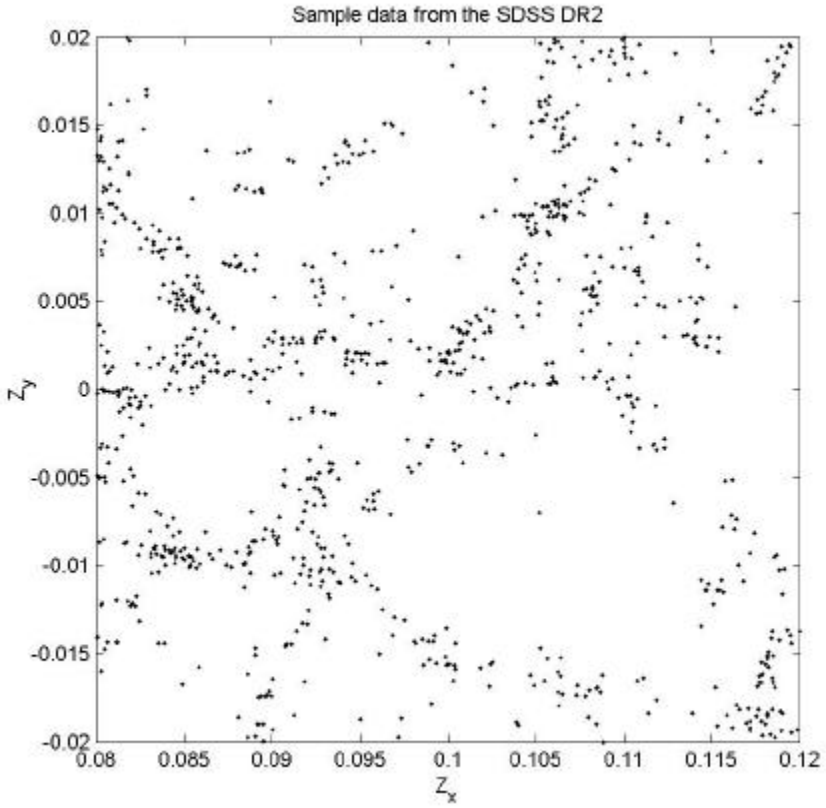
Classification of Hyperspectral Data - Example

HyperSpecTIR image of Virginia City: Possible Iron-bearing minerals and jarosite



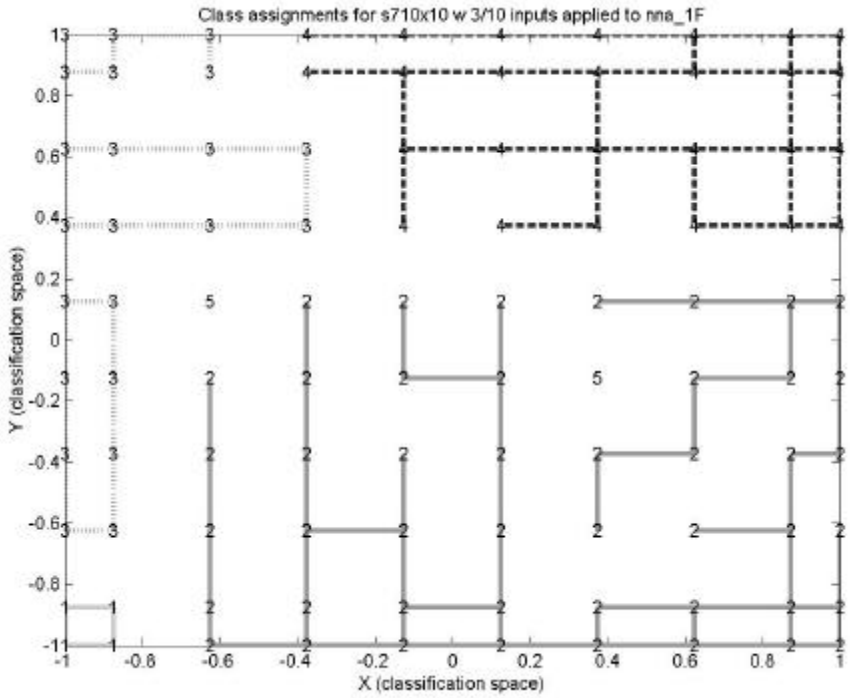
But what about ground truth?

Self-Organizing Maps (SOMs) for Clustering SDSS data – The Problem: How to Identify Clusters of Galaxies?



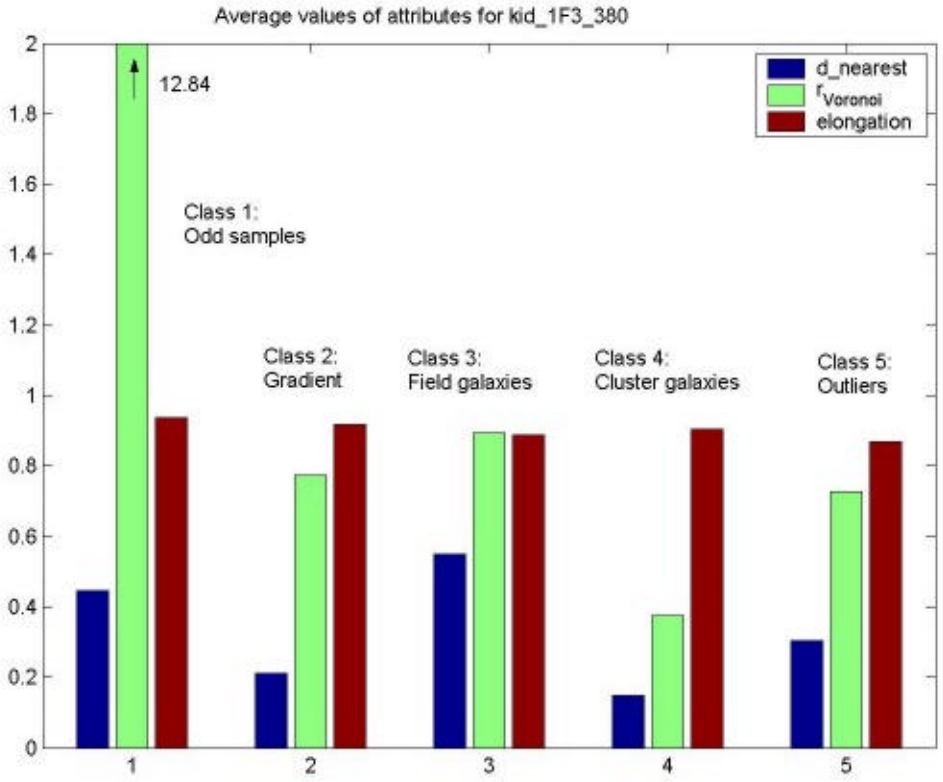
SOMs for Clustering SDSS data – The Method

- Generate a Voronoi tessellation and use the dimensions of the Voronoi cells to associate a set of attributes with each sample – one possibility might be $d_{\text{Nearest Neighbor}}/R_{\text{Voronoi}}$, $R_{\text{Voronoi}}/d_{\text{Uniform}}$, and Elongation.
- Use a SOM to map these 3-dimensional vectors onto a 2-D 'classification space' and search for boundaries in this 'classification space' to identify different classes of samples

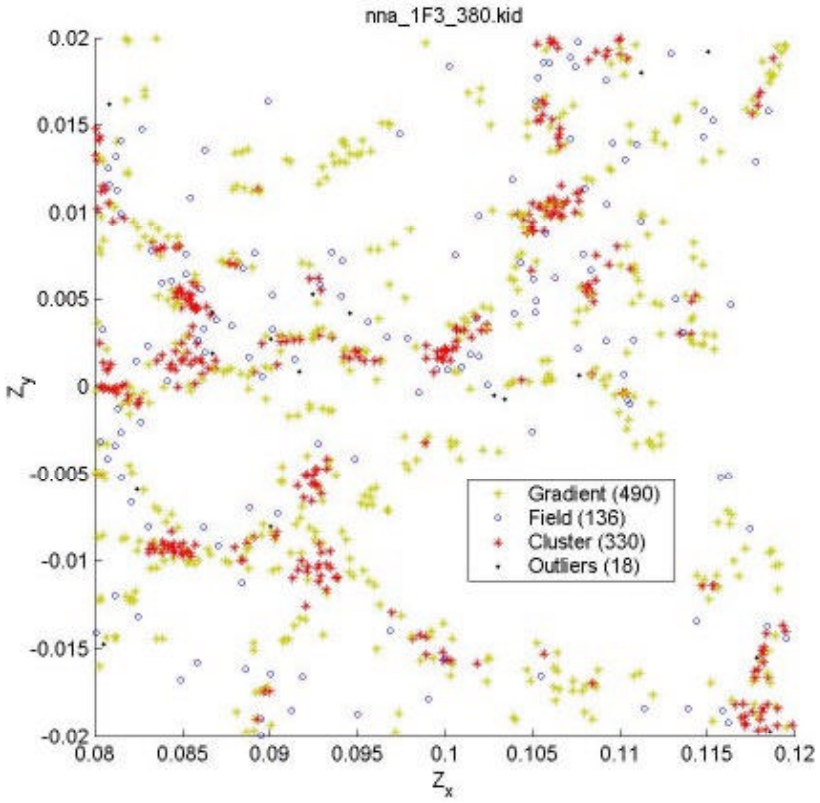


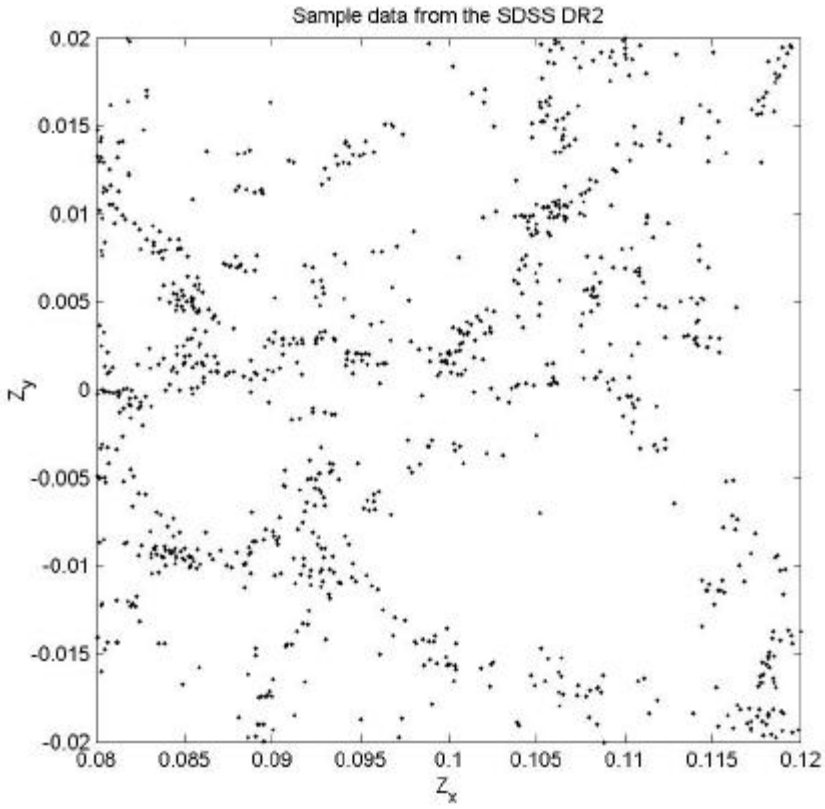
SOMs for Clustering SDSS data – The Classes

This method identifies three classes were identified in addition to outliers. The identification is reasonably stable and comparatively insensitive to details regarding the determination of attributes.



SOMs for Clustering SDSS data - Results





Summary

- Traditional well-understood AI methods such as neural networks, SOMs, and expert systems can be used to perform surveys of large space science and planetary science data sets that might defy analysis by more sophisticated methods.
- The results can be useful by themselves, or serve as a starting point for later examination by problem-specific analysis or statistical schemes.
- How can we take advantage of this?
- Surveys alone are not enough. To support a viable proposalk, they must be part of a larger coherent plan.