Intelligent Data Understanding for Earth and Space Science

The Ames Intelligent Data Understanding Group uses its expertise in Data Mining & Data Analysis to support the fields of Earth & Space Science.

Background

The Intelligent Data Understanding (IDU) Group is collaborating with domain scientists to answer pressing scientific questions in their fields and aid in knowledge discovery. The group builds on the significant successes and reputation that previous Machine Learning researchers have had at NASA Ames. Our main goal is to create tools and methods to aid in the assimilation and understanding of scientific data to best advance NASA's missions. We also advance scientific data understanding through research, and develop methodology for assisting in the efficient and cost effective collection of scientific data.

Research Overview

Virtual Sensors- Using Data Mining Techniques to Efficiently Estimate Remote Sensing Spectra
Various instruments are used to create images of the earth and other objects in the universe in a diverse set of wavelength bands. Many applied science questions that are relevant to the Earth Science remote sensing community require analysis of enormous amounts of data that were generated by instruments with disparate measurement capabilities. The IDU Group addresses this problem using Virtual Sensors. Virtual Sensors is a method that uses models trained on spectrally rich data to fill in unmeasured spectral channels in spectrally poor data. We demonstrate this method by using models trained on the high resolution Terra Moderate Resolution Imaging Spectroradiometer (MODIS) instrument to estimate what the equivalent of the MODIS 1.6- m channel would be for the National Oceanic and Atmospheric Administration Advanced Very High Resolution Radiometer (AVHRR/2) instrument. Simulation of the 1.6 micron channel would improve the ability of the AVHRR/2 sensor to detect clouds over snow and ice.

Figure 1

Greenland, from MODIS year 2000 day 140 time 1830 true channel 6.

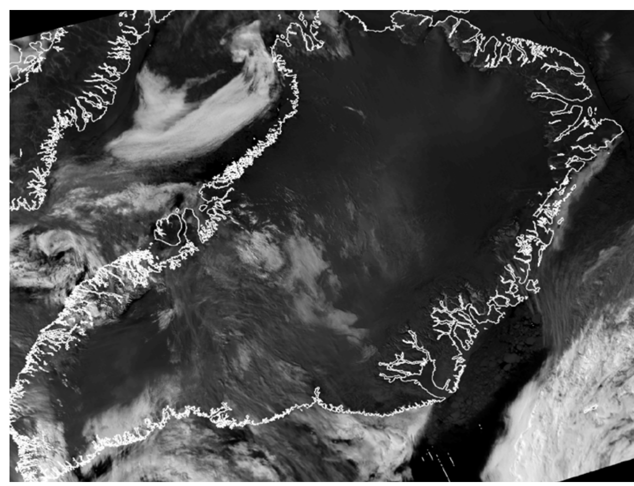
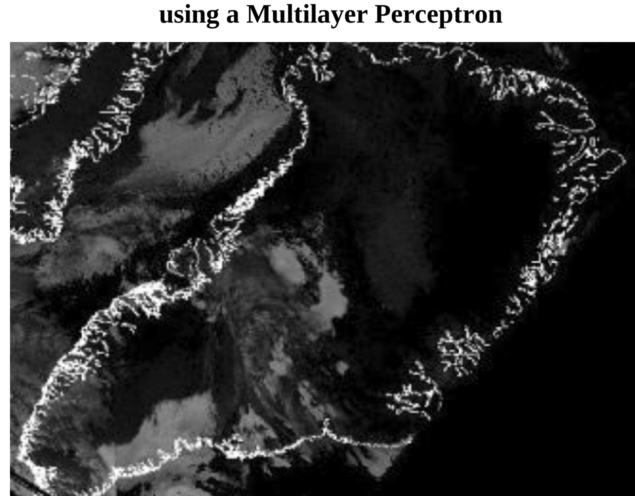


Figure 2
AVHRR prediction from year 2000, day 140, time 1839 using a Multilayer Perceptron

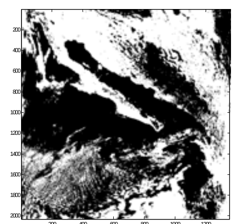


Intelligent Data Understanding for Earth and Space Science

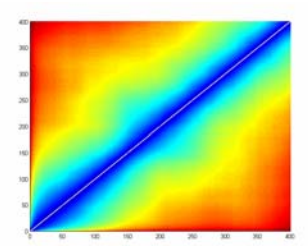
Novel Methods for Predicting Photometric Redshifts from Broad Band Photometry using Virtual Sensors: We calculate photometric redshifts from the Sloan Digital Sky Survey Main Galaxy Sample, The Galaxy Evolution Explorer All Sky Survey, and The Two Micron All Sky Survey using two machine learning methods. We utilize the broad-band photometry from the three surveys alongside Sloan Digital Sky Survey measures of photometric quality and galaxy morphology. Our first method draws from the theory of ensemble learning while the second employs Gaussian process regression both of which allow for the estimation of redshift along with a measure of uncertainty in the estimation. These two methods are compared to a well known Artificial Neural Network training-set method and to simple linear and quadratic regression. We also demonstrate the need to provide confidence bands on the error estimation made by both classes of models. A key contribution of our group is to quantify the variability in the quality of results as a function of model and training sample. We show how simply choosing the "best" model given a data set and model class can produce misleading results. We show that, using additional quality and morphology indicators rather than only the Sloan Digital Sky Survey broad-band imaging data, redshift accuracy can be improved by 10s of percent. Our work can be expanded to other photometric surveys where sufficient redshift calibration objects exist.

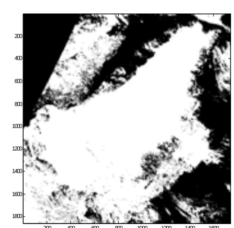
Self-Dissimilarity – A New Data Analysis Tool
We are developing a family of novel methods to probe aspects of data sets usually left unexamined by machine learning algorithms. By mimicking the process that goes on under a microscope as the magnification is changed, we probe a data set at different scales to build models of the structures present in the data at different scales. By comparing the resultant models at different scales information flow between scales can be explored.
The method has been applied to one- and two-dimensional artificial and real images. Work is under way analyzing the data for the cosmic microwave background recently obtained from the Wilkinson Microwave Anisotropy Probe (WMAP). The method promises a new tool through which to probe datasets. When applied to a data archive the method could be used to index the data. In analysis of remote sensing data, Self-Dissimilarity could reflect hidden levels of bio-diversity; indicate progressive die-backs before they become apparent or reflect changes to spatial distribution of climates.

Figures 3 (a & b)
Self dissimilarity results for satellite images. The original images are shown on the left (pixel values have been digitized to either 0 or 1) and self dissimilarity signatures are shown on the right. Even with the crude pixel thresholding the two images are seen to have different structure at different scales. In particular image 3(b) has structures at very large scales that differ significantly from the small scale structure. This may be due to the relative lack of clouds in the second image.

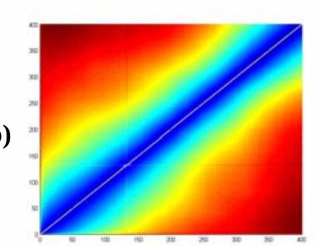


3(a)





3(b)



Relevance to Science Mission Directorate

Data mining & data analysis techniques developed by the Intelligent Data Understanding Group will enable space & earth scientists to better answer important questions in their fields. Our methods help scientists obtain results sooner, extracting new information from old records. We minimize their costs by using more effective methods to get the information they need.

Points of Contact:

Dr. Ashok Srivastava (Principal Scientist and Group Leader)
650-604-2409; Ashok.N.Srivastava@nasa.gov
http://ti.arc.nasa.gov/people/ashok

Dr. David Wolpert
650-604-3362; David.H.Wolpert@nasa.gov
http://ti.arc.nasa.gov/people/dhw/

Group Web Page:
http://ti.arc.nasa.gov/datamining

Ames Research Center