Framework for Understanding Non-linear Data with Missing Elements

Kevin Wheeler
Dogan Timucin
Maura Rabbette
Peter Pilewskie

Goals & Objectives

- Provide scientists with high-throughput interactive capability to analyze data and to graphically manipulate machine learning algorithms.
- Reduce time in programming for specific domains and in understanding algorithms.
- Contribute to atmospheric understanding through physical modeling and statistical analysis of hyper-spectral data.
- Contribute to atmospheric knowledge discovery through domain-specific Bayesian learning.

Technical Problem Statement

Graphical Programming: At current and expected data collection rates scientist have less time to program and learn new machine learning techniques. To minimize this time our framework concentrates on making interactive algorithms which are graphically accessible.

Atmospheric Understanding: The effect of the atmospheric aerosols on the solar heating of the Earth is regarded as the biggest missing piece in the global climate puzzle. We are developing a Bayesian framework to infer, from hyper-spectral data, the impact of aerosols on the solar radiation incident on the Earth's atmosphere.

Technical Approach

Graphical Framework

Implemented a graphical framework capable of user interaction and analysis of on-line data. This framework has been applied to 3 different problem domains:

- analysis of EEG for brain based interfaces,
- analysis of EMG for gesture based interfaces,
- analysis of hyper-spectral data.

Bayesian Framework

Develop a Bayesian learning algorithm that will be integrated into the graphical framework to estimate relevant aerosol parameters from data:

- derive a first-principles model of atmospheric optics suitable for use in the Bayesian algorithm,
- integrate calibration and navigation corrections for pre-processing of hyper-spectral data.

Data & NASA Relevance

- Hyper-spectral (380-1000 nm) down-welling and up-welling data used to determine radiative energy transport variations (helps to determine what atmospheric constituents are important).
- Electromyogram (EMG) data for gesture recognition for use with robotics and wearable computers.
- Electroencephalogram (EEG) data for brain machine interfaces.

Accomplishments

- Graphical framework has been used for closed-loop EEG experiments where participants have controlled computer movement with imagined motion.
- Graphical framework has been used to perform principal component analysis (PCA) and Independent Component Analysis (ICA) on hyper-spectral data.
- Independent Component Analysis has been shown to be effective for detecting instrument anomalies (such as saturation).
- A “custom” physical-optics model of the atmospheric propagation and scattering of solar radiation has been developed from the classical Maxwell–Lorentz theory. This model will constitute the heart of the Bayesian learning algorithm, obviating the use of general-purpose packages such as MODTRAN.

Significance of Progress

- Framework has made closed-loop EEG experiments possible by providing an easy means to reconfigure algorithms on the fly, very fast graphics for appropriate subject and experimenter feedback.
- The framework has shown that it is feasible to detect instrument anomalies during data collection prior to data archiving. This will provide the capability to scientists to change experiments quickly when experiments are not performing as expected.
- Derivation of a physical model that is specifically tailored to the class of data under study provides a powerful and efficient foundation for the successful subsequent application of pertinent statistical analysis and learning algorithms.

Facilities & Personnel

Facilities:

Framework is developed and used in NASA ARC Neuro-Engineering Laboratory. It runs on laptops, Linux workstations, Macs (OS X), and SGI clusters. Display capabilities currently include 3 high-definition plasma displays and a life-size immersive display system.

Personnel:

Personnel include 2 contractors, 1.5 civil servants, and 2 students.

References

Rabbette, M. and P. Pilewskie, Multivariate Analysis of Solar Spectral Irradiance Measurements,
J. Geophys. Res., 106, 9685-9696, 2001.

Rabbette, M. and P. Pilewskie, Principal Component Analysis of Arctic Solar Irradiance Spectra,
in press, J. Geophys. Res. (oceans), (2002).