National Security Education Center

A CONSORTIUM OF LANL INSTITUTES

UNCLASSIFIED

Information Science and Technology Seminar Series

Kiri Wagstaff
Jet Propulsion Laboratory

"Eyes Wide Open: Iterative Discovery in Large Data Sets without Premature Specialization"

Wednesday, March 21, 2012
3:00 - 4:00 PM
TA-3, Bldg. 1690, Room 102 (CNLS Conference Room)

Abstract: What is the best way to dive in and explore a new data set? We pose a new machine learning problem, iterative discovery, that seeks to enable users to interactively explore a large data set and quickly identify items of interest. At each iteration, the system selects an item for the user to review, and the user provides feedback as to whether it is interesting or not. The system must retain a strong exploratory bias to avoid premature specialization (seeking items similar to those known to be interesting, and missing out on different but equally interesting items). Unlike active learning, the goal is to select items of most use to the user, not to the system. I will describe a solution called Discovery through Eigenbasis Modeling of Uninteresting Data (DEMUD) that avoids premature specialization on the positive (interesting) class by modeling the negative (uninteresting) feedback only. DEMUD is especially effective when the class of interest is rare and/or heterogeneous. I will share results of experiments with image, planetary science, and astronomy data, in which we find that DEMUD discovers items of interest faster than methods that model the positive class, including active learning.

Biography: Kiri Wagstaff is a senior researcher in artificial intelligence and machine learning at the Jet Propulsion Laboratory. Her focus is on developing new machine learning and data analysis methods, particularly those that can be used for in situ analysis onboard spacecraft such as orbiters, landers, rovers, and so on. She holds a Ph.D. in Computer Science from Cornell University and an M.S. in Geological Sciences from the University of Southern California. She received a 2008 Lew Allen Award for Excellence in Research, with which she investigated the ability of machine learning methods to operate correctly in high-radiation space environments. Her current projects include automated change detection in Mars orbital images, texture-based characterization of rover images, and fast detection of transient signals in radio array data.

Los Alamos NATIONAL LABORATORY EST.1943

UNCLASSIFIED

For more information contact James Theiler, theiler@lanl.gov, 665-5682.
Hosted by the Information Science and Technology Center (ISTC)