National Aeronautics and Space Administration

Schedule and Abstracts

October 5-6, 2010

Computer History Museum
Mountain View, CA

Sponsored by

NASA Aviation Safety Program
NASA Applied Information Systems
Research Program

www.nasa.gov

October 5-6, 2010

Computer History Museum
Mountain View, CA

CIDU 2010 brings together top researchers and practitioners in the field of data mining, focusing on research and development activities in the Earth Sciences, Space Sciences, and Systems Health application areas. The proceedings of CIDU 2010 will be published by NASA, archived in the NASA Center for Aerospace Information, and indexed by DBLP. Selected papers will be published in the journal *Statistical Analysis and Data Mining*.

Organizing Committee

General Chair: Ashok Srivastava, NASA Ames Research Center

Program Co-Chairs: Nitesh Chawla, University of Notre Dame
Philip Yu, University of Illinois at Chicago

Area Chairs

Earth Science Applications:
Sara Graves, University of Alabama in Huntsville
Steve Sain, NCAR

Space Science Applications
Kiri Wagstaff, NASA Jet Propulsion Laboratory
Kirk Borne, George Mason University

Aerospace and Engineering Systems Applications:
Sylvain Letourneau, NRC, Canada
Dimitry Gorinevsky, Stanford University

Posters Chair: Kanishka Bhaduri, MCT Inc./NASA Ames Research Center

Proceedings Chair: Paul Melby, MITRE

Publicity Chair: Hui Xiong, Rutgers University

Local Arrangements Chair:
Elizabeth Foughty, MCT Inc./NASA Ames Research Center

Communications Chair:
Kamalika Das, SGT Inc./NASA Ames Research Center

PROGRAM COMMITTEE

Aditi Chattopadhyay, Arizona State University
Aleksandar Lazarevic, United Technology Research Center
David Thompson, NASA Jet Propulsion Laboratory
Dragos Margineantu, Boeing
Jianping Zhang, MITRE
Mark Last, Ben-Gurion University of the Negev
Naresh Iyer, GE Global Research
Philip Kegelmeyer, Sandia National Laboratory
Alessandro Sperduti, University of Padua, Via Trieste
Alfredo Cuzzocrea, ICAR-CNR and University of Calabria, Italy
Arindam Banerjee, University of Minnesota
Jiawei Han, University of Illinois at Urbana-Champaign
Joydeep Ghosh, University of Texas at Austin
Olfa Nasraoui, University of Louisville
Sugato Basu, Google Inc.
Tim Oates, University of Maryland Baltimore County
Xindong Wu, University of Vermont
Zhi-Hua Zhou, Nanjing Universit, China
Auroop Ganguly, Oak Ridge National Laboratory
Claire Monteleoni, Columbia University
Karsten Steinhaeuser, University of Notre Dame
Latifur Khan, University of Texas at Dallas
Olufemi Omitaomu, Oak Ridge National Laboratory
Rahul Ramachandran, University of Alabama in Huntsville
Ranga Raju Vatsavai, Oak Ridge National Laboratory
Shashi Shekhar, University of Minnesota
Vipin Kumar, University of Minnesota
Zoran Obradovic, Temple University
Gary Weiss, Fordham University
Amy McGovern, University of Oklahoma
Douglas Burke, Harvard University
George Djorgovski, California Institute of Technology
Jeff Scargle, NASA Ames Research Center
Massimo Brescia, University of Naples
Michael Burl, NASA Jet Propulsion Laboratory
Mike Way, NASA Ames and NASA Goddard
Rick White, Space Telescope Science Institute
Robert Brunner, UIUC/NCSA
Terran Lane, University of New Mexico
Trey Smith, NASA Ames Research Center

Conference Agenda

Tuesday, October 5

8:00 AM	Registration
8:30 AM	*Opening Remarks* - Elizabeth Foughty, Ashok Srivastava, Nitesh Chawla
8:45 AM	**Keynote Address** *Distributed Convex Optimization for Large Scale Statistical Modeling and Data Analysis* – Stephen Boyd, Stanford University
9:45 AM	* ***break*** *
	Session 1 – Sara Graves, Chair
10:10 AM	**Invited Talk** *Mining in the Tropics: Risks and Rewards* – Ramakrishna Nemani, NASA Ames Research Center
10:40 AM	*Understanding Severe Weather Processes Through Spatiotemporal Relational Random Forests* – Amy McGovern, University of Oklahoma
11:00 AM	*Complex Networks in Climate Science: Progress, Opportunities, and Challenges* – Karsten Steinhaeuser, University of Notre Dame
11:20 AM	*Spatially Adaptive Semi-supervised Learning with Gaussian Processes for Hyperspectral Data Analysis* – Joydeep Ghosh, University of Texas at Austin
11:45 PM	**Lunch and Keynote Address** *Soft Computing in the Design of Anomaly Detection Models* – Piero Bonissone, GE Global Research
	Session 2 – Kirk Borne, Chair
1:30 PM	**Invited Talk** *Exploration of the Time Domain in Astronomy: Towards the Real-Time Mining of Petascale Data Streams* – George Djorgovski, California Institute of Technology
2:00 PM	*Improving Cause Detection Systems with Active Learning* – Vincent Ng, University of Texas at Dallas
2:20 PM	*Classification of Mars Terrain Using Multiple Data Sources* – Alan Kraut, Carnegie Mellon University
2:40 PM	*Lunar Terrain and Albedo Reconstruction from Apollo Imagery* – Ara Nefian, SGT Inc./NASA Ames
3:00 PM	*Data Mining the Galaxy Zoo Mergers* – Steven Baehr, George Mason University
3:20 PM	* ***break*** *
	Session 3 – Claire Monteleoni, Chair
3:40 PM	*Optimal Partitions of Data in Higher Dimensions* – Jeff Scargle, NASA Ames Research Center

4:00 PM	*Probability Calibration by the Minimum and Maximum Probability Scores in One-Class Bayes Learning for Anomaly Detection* – Guichong Li, University of Ottawa
4:20 PM	*Scalable Time Series Change Detection for Biomass Monitoring Using Gaussian Process* – Ranga Raju Vatsavai, Oak Ridge National Laboratory
4:40 PM	*A Comparative Study of Algorithms for Land Cover Change* – Ashish Garg, University of Minnesota
5:00 PM	* ***break*** *
5:30 – 7:30 PM	**Poster Session and Reception**

Wednesday, October 6

	Session 4 – Paul Melby, Chair
8:30 AM	**Invited Talk** *Vehicle Level Reasoning and Data Mining* – Dinkar Mylaraswamy, Honeywell
9:00 AM	*Keyword Search in Text Cube: Finding Top-k Aggregated Cell Documents* – Cindy Xide Lin
9:20 AM	*Dynamic Strain Mapping and Real-Time Damage State Estimation Under Biaxial Random Fatigue Loading* – Clyde Coelho, Arizona State University
9:40 AM	*Analyzing Aviation Safety Reports: From Topic Modeling to Scalable Multi-Label Classification* – Amrudin Agovic, University of Minnesota
10:00 AM	* ***break*** *
	Session 5 – Kamalika Das, Chair
10:20 AM	*Usage of Dissimilarity Measures and Multidimensional Scaling for Large Scale Solar Data Analysis* – Juan Banda, Montana State University
10:40 AM	*A Knowledge Discovery Strategy for Relating Sea Surface Temperatures to Frequencies of Tropical Storms and Generating Predictions of Hurricanes Under 21st-century Global Warming Scenarios* – Vipin Kumar, University of Minnesota
11:00 AM	*Tracking Climate Models* – Claire Monteleoni, Columbia University
11:20 AM	*Adaptive Model Refinement for the Ionosphere and Thermosphere* – Anthony D'Amato, University of Michigan
11:45 PM	**Lunch and Keynote Address** *Monitoring of the Changes in the Global Forest Cover Using Data Mining* – Vipin Kumar, University of Minnesota
	Session 6 – Karsten Steinhaeuser, Chair
1:30 PM	**Invited Talk** *Applying Avatar Machine Learning to NIF Optics Inspection Analysis* – Laura Kegelmeyer, Lawrence Livermore National Laboratory

2:00 PM	*PADMINI: A Peer-to-Peer Distributed Astronomy Data Mining System and a Case Study* – Hillol Kargupta, University of Maryland Baltimore County
2:20 PM	*Multi-temporal Remote Sensing Image Classification: A Multi-view Approach* – Ranga Raju Vatsavai, Oak Ridge National Laboratory
2:40 PM	*Distributed Anomaly Detection Using Satellite Data from Multiple Modalities* – Kamalika Das, SGT Inc./NASA Ames
3:00 PM	*Multi-label ASRS Dataset Classification Using Semi-supervised Subspace Clustering* – Mohammad Salim Ahmed, University of Texas at Dallas
3:20 PM	* ***break*** *
3:40 – 5:00	**Panel Discussion – Nitesh Chawla, Moderator** *Data to Understanding to Knowledge Discovery in Earth, Climate, and Aero Sciences* Piero Bonissone, GE Global Research George Djorgovski, California Institute of Technology Vipin Kumar, University of Minnesota Pat Moran, NASA Ames Research Center Ramakrishna Nemani, NASA Ames Research Center Ramasamy Uthurusamy, General Motors

Presentations

Tuesday

Keynote Address

Session 1

Invited Talk

Lunch Keynote

Session 2

Invited Talk

Presentations

Keynote Address

Distributed Convex Optimization for Large Scale Statistical Modeling and Data Analysis

Stephen Boyd
Stanford University

Many statistical modeling and data analysis methods are based on convex optimization, a special type of optimization problem which can be solved effectively. These include many regression models, support vector machines, logistic and probit classifiers, and many new techniques, such as compressed sensing, Lasso, and basis pursuit for finding sparse or otherwise simple models from data. General-purpose convex optimization techniques can handle medium-size problems; special algorithms for specific models have been developed to handle large-scale data sets. In this task we will present another approach to handling truly large-scale data sets, using distributed convex optimization. The idea (which goes back to the 1970s, used in other contexts) is to solve a huge problem on a set of machines, with distributed data. Each machine carries out a local data fit, with high-level coordination to bring the problems into consensus, at the globally optimal fit. We will argue that these methods are well suited to modeling modern massive data sets in a cloud or other distributed computing framework.

This research was joint work with Neal Parikh, Eric Chu, Borja Peleato, and Dimitri Gorinevsky.

Stephen P. Boyd is the Samsung Professor of Engineering, and Professor of Electrical Engineering in the Information Systems Laboratory at Stanford University. His current research focus is on convex optimization applications in control, signal processing, and circuit design.

Professor Boyd received an AB degree in Mathematics, summa cum laude, from Harvard University in 1980, and a PhD in EECS from U.C. Berkeley in 1985. In 1985 he joined the faculty of Stanford's Electrical Engineering Department. He has held visiting Professor positions at Katholieke University (Leuven), McGill University (Montreal), Ecole Polytechnique Federale (Lausanne), Qinghua University (Beijing), Universite Paul Sabatier (Toulouse), Royal Institute of Technology (Stockholm), Kyoto University, and Harbin Institute of Technology. He holds an honorary doctorate from Royal Institute of Technology (KTH), Stockholm.

Professor Boyd is the author of many research articles and three books: *Linear Controller Design: Limits of Performance* (with Craig Barratt, 1991), *Linear Matrix Inequalities in System and Control Theory* (with L. El Ghaoui, E. Feron, and V. Balakrishnan, 1994), and *Convex Optimization* (with Lieven Vandenberghe, 2004).

Professor Boyd has received many awards and honors for his research in control systems engineering and optimization, including an ONR Young Investigator Award, a Presidential Young Investigator Award, and an IBM faculty development award. In 1992 he received the AACC Donald P. Eckman Award, which is given annually for the greatest contribution to the field of control engineering by someone under the age of 35. In 1993 he was elected Distinguished Lecturer of the IEEE Control Systems Society, and in 1999 he was elected Fellow of the IEEE, with citation: "For contributions to the design and analysis of control systems using convex optimization based CAD tools." He has been invited to deliver more than 30 plenary and keynote lectures at major conferences in both control and optimization.

In addition to teaching large graduate courses on Linear Dynamical Systems, Nonlinear Feedback Systems, and Convex Optimization, Professor Boyd has regularly taught introductory undergraduate Electrical Engineering courses on Circuits, Signals and Systems, Digital Signal Processing, and Automatic Control. In 1994 he received the Perrin Award for Outstanding Undergraduate Teaching in the School of Engineering, and in 1991, an ASSU Graduate Teaching Award. In 2003, he received the AACC Ragazzini Education award, for contributions to control education, with citation: "For excellence in classroom teaching, textbook and monograph preparation, and undergraduate and graduate mentoring of students in the area of systems, control, and optimization."

Invited Talk

Mining in the Tropics: Risks and Rewards

Ramakrishna Nemani
NASA Ames Research Center

The tropical regions of the Earth are the energy engines for the global climate system. On interannual timescales much of the global climate variability originates in the tropics through changes in El Nino Southern Oscillation (ENSO), the Asian monsoon, and the location of the Inter-Tropical Convergence Zone (ITCZ). Unraveling the complex interplay of these large-scale atmospheric phenomena and their impact on regional carbon cycling is difficult due to the lack of a sufficient network of ground observations. While orbiting satellites provide a unique perspective of the tropical regions, the observations are often difficult to interpret due to contamination by perennial clouds during the rainy season and aerosols from fires during the dry season. Knowledge discovery under these challenging conditions is difficult but highly rewarding because hypotheses generated through machine learning and data mining techniques could form the basis for field campaigns or satellite missions.

I will discuss the challenges of using the Earth observations from NASA satellites over Amazon ecosystems, followed by a discussion of a framework for detecting, classifying, and understanding anomalous behavior in satellite data. I will conclude by presenting a collaborative framework for large-scale data mining in Earth sciences consisting of a massive data center attached to the NASA supercomputing system with a community portal for exchanging algorithms and models.

Ramakrishna Nemani is a senior research scientist with the Biospheric Sciences Branch in the Earth Science Division at NASA's Ames Research Center, Moffett Field, Calif. His work focuses on integrating satellite data with simulation models to understand and predict vegetation responses to changes and variations in climate and using this information to support natural resources management. He is a member of NASA missions such as the Earth Observing System and the Landsat Data Continuity Mission. He authored and co-authored more than 140 papers and has received several NASA performance awards, including an Exceptional Scientific Achievement Medal in 2008.

Understanding Severe Weather Processes Through Spatiotemporal Relational Random Forests

Amy McGovern, Timothy Supinie, David Gagne II,
Nathaniel Troutman, Matthew Collier
University of Oklahoma
Roger Brown
NOAA/National Severe Storms Laboratory
Jeffrey Basara
Oklahoma Climatological Survey
John Williams
National Center for Atmospheric Research

Major severe weather events can cause a significant loss of life and property. We seek to revolutionize our understanding of and ability to predict such events through the mining of severe weather data. Because weather is inherently a spatiotemporal phenomenon, mining such data requires a model capable of representing and reasoning about complex spatiotemporal dynamics, including temporally and spatially varying attributes and relationships. We introduce an augmented version of the Spatiotemporal Relational Random Forest, which is a Random Forest that learns with spatiotemporally varying relational data. Our algorithm maintains the strength and performance of Random Forests but extends their applicability, including the estimation of variable importance, to complex spatiotemporal relational domains. We apply the augmented Spatiotemporal Relational Random Forest to three severe weather data sets: predicting atmospheric turbulence across the continental United States, examining the formation of tornadoes near strong frontal boundaries, and understanding the translation of drought across the southern plains of the United States. The results on such a wide variety of real-world domains demonstrate the extensive applicability of the Spatiotemporal Relational Random Forest. Our long-term goal is to significantly improve the ability to predict and warn about severe weather events.

Complex Networks in Climate Science: Progress, Opportunities, And Challenges

Karsten Steinhaeuser, Nitesh Chawla
University of Notre Dame
Auroop Ganguly
Oak Ridge National Laboratory

Networks have been used to describe and model a wide range of complex systems, both natural and man-made. One particularly interesting application in the earth sciences is the use of complex networks to represent and study the global climate system. In this paper, we motivate this general approach, explain the basic methodology, report on the state of the art (including our contributions), and outline open questions and opportunities for future research.

Spatially Adaptive Semi-supervised Learning with Gaussian Processes for Hyperspectral Data Analysis

Goo Jun, Joydeep Ghosh
University of Texas at Austin

A semi-supervised learning algorithm for the classification of hyperspectral data, Gaussian process expectation maximization (GP-EM), is proposed. Model parameters for each land cover class are first estimated by a supervised algorithm using Gaussian process regressions to find spatially adaptive parameters, and the estimated parameters are then used to initialize a spatially adaptive mixture-of-Gaussians model. The mixture model is updated by expectation-maximization iterations using the unlabeled data, and the spatially adaptive parameters for unlabeled instances are obtained by Gaussian process regressions with soft assignments. Two sets of hyperspectral data taken from the Botswana area by the NASA EO-1 satellite are used for experiments. Empirical evaluations show that the proposed framework performs significantly better than baseline algorithms that do not use spatial information, and the results are also better than any previously reported results by other algorithms on the same data.

Keynote Address

Soft Computing in the Design of Anomaly Detection Models

Piero Bonissone
GE Global Research

Soft Computing (SC) is a term that originally defined in 1994 by L.A. Zadeh as "an association of computing methodologies that includes as its principal members fuzzy logic (FL), neuro-computing (NC), evolutionary computing (EC), and probabilistic computing (PC)." Since then, this concept has evolved into a methodology and a set of techniques to cover the aspects of data-driven model design, model generation, and model tuning.

In our modeling process, we propose a strong separation between offline meta-heuristics (MH), used for design and tuning the models, and online MH, used for models selection or aggregation. This view suggests a broader use of SC components, since it enables us to use hybrid SC techniques at each of the MH levels as well as at the object level. We can manage model complexity by finding the best model architecture to support problem decomposition, generate local models with high performance in focused applicability regions, provide smooth interpolations among local models, and increase robustness to imperfect data by aggregating diverse models. Furthermore, this separation facilitates the model lifecycle management, which is required to maintain the models' vitality and prevent their obsolescence over time.

To illustrate this concept, we will use a case study. We will address the problem of anomaly detection for a fleet of physical assets, such as aircraft engines or gas turbines. Anomaly detection typically uses unsupervised learning techniques to extract the underlying structural information from the data, define normal structures and regions, and identify departures from such regions. We analyze potential causes of anomalies, as they vary from incipient system failures to malfunctioning sensors, operating the asset in unusual regions, using inappropriate anomaly detection models, etc. We focus on one of the most neglected causes for anomalies: the inadequate accuracy of the anomaly detection models, which may create false alarms. To address this issue, we propose a hybrid approach based on a fuzzy supervisory system and an ensemble of locally trained auto associative neural networks (AANNs.) The design and tuning of this hierarchical model is performed using evolutionary algorithms. In our approach we interpolate among the outputs of the local models (AANNs) to assure smoothness in operating regime transition and provide continuous condition monitoring to the system. Experiments on

simulated data from a high-bypass, turbofan aircraft engine model demonstrated promising results.

A Chief Scientist at GE Global Research, Piero Bonissone has been a pioneer in the field of fuzzy logic, AI, soft computing, and approximate reasoning systems applications since 1979. He is a Fellow of IEEE, AAAI, IFSA, and a Coolidge Fellow at GE Global Research. He served as Editor in Chief of the *International Journal of Approximate Reasoning* for 13 years. He has co-edited six books and has over 150 publications. He received 51 patents issued from the USPTO (plus 51 pending). He has (co-)chaired 12 scientific conferences and symposia focused on Multi-Criteria Decision-Making, Fuzzy sets, Diagnostics, Prognostics, and Uncertainty Management in AI. In 2002, he was President of the IEEE Neural Networks Society (now Computational Intelligence Society). He has been an Executive Committee member of NNC/NNS/CIS society since the past 16 years and an IEEE CIS Distinguished Lecturer since 2004.

For a more detailed biography, see http://www.rpi.edu/~bonisp/bonissone-bio.htm

Invited Talk

Exploration of the Time Domain in Astronomy: Towards the Real-Time Mining of Petascale Data Streams

George Djorgovski
California Institute of Technology

A new generation of synoptic digital sky surveys, which cover large areas of the sky repeatedly, looking for transient and variable phenomena, is opening a vibrant new area of research in astrophysics, ranging from the Solar system, through stellar evolution, Galactic structure, to cosmology and extreme relativistic phenomena. Today, we have surveys with Terascale data rates, but we are heading into the Petascale regime within a decade, with facilities like the LSST or SKA. In addition to the usual challenges of knowledge discovery in massive and complex data sets, the time dimension brings additional challenges along with the scientific opportunities: transient events have to be discovered, characterized, and classified in (near) real time, with the decisions made for an optimal follow-up of the most interesting ones, given the sparse and valuable available assets (e.g., spectroscopy with large telescopes, target-of-opportunity interrupts, etc.). I will describe the work done with PQ and CRTS surveys, which, in addition to the scientific returns of their own, provide scientific and technological testbeds for the more ambitious surveys in the future.

George Djorgovski is a Professor of Astronomy and a Co-Director of the Center for Advanced Computing Research at Caltech, and the Director of the Meta Institute for Computational Astrophysics, the first professional scientific organization based entirely in virtual worlds. After receiving his PhD from UC Berkeley, he was a Harvard Junior Fellow, before joining the Caltech faculty in 1987. He was a Presidential Young Investigator and an Alfred P. Sloan Foundation Fellow, among other honors and distinctions, and he is an author or coauthor of several hundred professional publications. He was one of the founders of the Virtual Observatory concept, and was the Chairman of the US Nat'l Virtual Observatory Science Definition Team, and the PI of DPOSS, PQ, and CRTS digital sky surveys. His e-Scientific interests include definition and development of the universal methodology, tools and frameworks for data-intensive and computationally enabled science, various aspects of data mining, virtual scientific organizations, etc.

Improving Cause Detection Systems with Active Learning

Isaac Persing, Vincent Ng
University of Texas at Austin

Active learning has been successfully applied to many natural language processing tasks for obtaining annotated data in a cost-effective manner. We propose several extensions to an active learner that adopts the margin-based uncertainty sampling framework. Experimental results on a cause detection problem involving the classification of aviation safety reports demonstrate the effectiveness of our extensions.

Classification of Mars Terrain Using Multiple Data Sources

Alan Kraut, David Wettergreen
Carnegie Mellon University

Images of Mars are being collected faster than they can be analyzed by planetary scientists. Automatic analysis of images would enable more rapid and more consistent image interpretation and could draft geologic maps where none yet exist. In this work we develop a method for incorporating images from multiple instruments to classify Martian terrain into multiple types. Each image is segmented into contiguous groups of similar pixels, called superpixels, with an associated vector of discriminative features. We have developed and tested several classification algorithms to associate a best class to each superpixel. These classifiers are trained using three different manual classifications with between two and six classes. Automatic classification accuracies of 50 to 80% are achieved in leave-one-out cross-validation across 20 scenes using a generalized boosting classifier.

Lunar Terrain and Albedo Reconstruction from Apollo Imagery

Ara Nefian, Zachary Morotto
SGT Inc./NASA Ames Research Center
Taemin Kim
NASA Marshall Space Flight Center
Michael Broxton
NASA Ames Research Center

Generating accurate three-dimensional planetary models and albedo maps is becoming increasingly important as NASA plans more robotics missions to the Moon in the coming years. This paper describes a novel approach for separation of topography and albedo maps from orbital Lunar images. Our method uses an optimal Bayesian correlator to refine the stereo disparity map and generate a set of accurate digital elevation models (DEMs). The albedo maps are obtained using a multi-image formation model that relies on the derived DEMs and the Lunar-Lambert reflectance model. The method is demonstrated on a set of high-resolution scanned images from the Apollo-era missions.

Data Mining the Galaxy Zoo Mergers

Steven Baehr, Arun Vedachalam, Kirk Borne, Daniel Sponseller
George Mason University

Collisions between pairs of galaxies usually end in the coalescence (merger) of the two galaxies. Collisions and mergers are rare phenomena, yet they may signal the ultimate fate of most galaxies, including our own Milky Way. With the onset of massive collection of astronomical data, a computerized and automated method will be necessary for identifying those colliding galaxies worthy of more detailed study. This project researches methods to accomplish that goal. Astronomical data from the Sloan Digital Sky Survey (SDSS) and human-provided classifications on merger status from the Galaxy Zoo project are combined and processed with machine learning algorithms. The goal is to determine indicators of merger status based solely on discovering those automated pipeline-generated attributes in the astronomical database that correlate most strongly with the patterns identified through visual inspection by the Galaxy Zoo volunteers. In the end, we aim to provide a new and improved automated procedure for classification of collisions and mergers in future petascale astronomical sky surveys. Both information gain analysis (via the C4.5 decision tree algorithm) and cluster analysis (via the Davies-Bouldin Index) are explored as techniques for finding the strongest correlations between human-identified patterns and existing database attributes. Galaxy attributes measured in the SDSS green waveband images are found to represent the most influential of the attributes for correct classification of collisions and mergers. Only a nominal information gain is noted in this research; however, there is a clear indication of which attributes contribute so that a direction for further study is apparent.

Optimal Partitions of Data in Higher Dimensions

Jeff Scargle
NASA Ames Research Center
Bradley Jackson
San Jose State University

Consider piece-wise constant approximations to a function of several parameters, and the problem of finding the best such approximation from measurements at a set of points in the parameter space. We find good approximate solutions to this problem in two steps: (1) partition the parameter space into cells, one for each of the N data points, and (2) collect these cells into blocks, such that within each block the function is constant to within measurement uncertainty. We describe a branch-and-bound algorithm for finding the optimal partition into connected blocks, as well as an $O(N^{2})$ dynamic programming algorithm that finds the exact global optimum over this exponentially large search space, in a data space of any dimension. This second solution relaxes the connectivity constraint, and requires additivity and convexity conditions on the block fitness function, but in practice none of these items cause problems. From the wide variety of intelligent data understanding applications (including cluster analysis, classification, and anomaly detection) we demonstrate two: partitioning of the State of California (2D) and the Universe (3D).

Probability Calibration By The Minimum And Maximum Probability Scores In One-Class Bayes Learning For Anomaly Detection

Guichong Li, Nathalie Japkowicz
University of Ottawa
Ian Hoffman, R. Kurt Ungar
Radiation Protection Bureau, Health Canada

One-class Bayes learning such as one-class Naïve Bayes and one-class Bayesian Network employs Bayes learning to build a classifier on the positive class only for discriminating the positive class and the negative class. It has been applied to anomaly detection for identifying abnormal behaviors that deviate from normal behaviors. Because one-class Bayes classifiers can produce probability score, which can be used for defining anomaly score for anomaly detection, they are preferable in many practical applications as compared with other one-class learning techniques. However, previously proposed one-class Bayes classifiers might suffer from poor probability estimation when the negative training examples are unavailable. In this paper, we propose a new method to improve the probability estimation. The improved one-class Bayes classifiers can exhibit high performance as compared with previously proposed one-class Bayes classifiers according to our empirical results.

Scalable Time Series Change Detection for Biomass Monitoring Using Gaussian Process

Varun Chandola, Ranga Raju Vatsavai
Oak Ridge National Laboratory

Biomass monitoring, specifically, detecting changes in the biomass or vegetation of a geographical region, is vital for studying the carbon cycle of the system and has significant implications in the context of understanding climate change and its impacts. Recently, several time series change detection methods have been proposed to identify land cover changes in temporal profiles (time series) of vegetation collected using remote sensing instruments. In this paper, we adapt Gaussian process regression to detect changes in such time series in an online fashion. While Gaussian process (GP) has been widely used as a kernel-based learning method for regression and classification, its applicability to massive spatio-temporal data sets, such as remote sensing data, has been limited owing to the high computational costs involved. In our previous work we proposed an efficient Toeplitz matrix-based solution for scalable GP parameter estimation. In this paper we apply this solution to a GP-based change detection algorithm. The proposed change detection algorithm requires a memory footprint which is linear in the length of the input time series and runs in time which is quadratic to the length of the input time series. Experimental results show that both serial and parallel implementations of our proposed method achieve significant speedups over the serial implementation. Finally, we demonstrate the effectiveness of the proposed change detection method in identifying changes in Normalized Difference Vegetation Index (NDVI) data.

A Comparative Study Of Algorithms For Land Cover Change

Shyam Boriah, Varun Mithal, Ashish Garg,
Vipin Kumar, Michael Steinbach
University of Minnesota
Chris Potter
NASA Ames Research Center
Steve Klooster
CSU Monterey Bay

Ecosystem-related observations from remote sensors on satellites offer huge potential for understanding the location and extent of global land cover change. This paper presents a comparative study of three time series-based algorithms for detecting changes in land cover. The techniques are evaluated quantitatively using forest fire ground truth from the state of California for 2000–2009. On relatively high quality data sets, all three schemes perform reasonably well, but their ability to handle noise and natural variability in the vegetation data differs dramatically. In particular, one of the algorithms significantly outperforms the other two since it accounts for variability in the time series.

Invited Talk

Vehicle Level Reasoning and Data Mining

Dinkar Mylaraswamy
Honeywell

In this talk, I will describe the significance of vehicle level reasoner for safety, then describe an the underlying data model to support the vehicle level reasoning. I will talk about how data mining can help to discover unknown structure and tune parameters of this data model using data collected on a typical aircraft, and iIllustrate this using 1–2 examples.

Dinkar Mylaraswamy is a Research Fellow in Honeywell's Aerospace Advanced Technology group. His research interests include fault diagnosis, condition-based control, and modeling. His thesis on blackboard-based cooperative fault diagnosis problem solving was a foundation for Honeywell's asset management assetMax offering. For this work, he was recognized with Honeywell's highest technical award (H. W. Sweatt Award). He has served as the principal investigator on several health management programs both in the aerospace and industrial domains. Recent ones include Honeywell's PTMD and several Army engine analytics programs.

As part of the Technology Strategy team, Dr. Mylaraswamy is responsible for identifying strategic direction for IVHM technology areas within Honeywell which cut across multiple products and services. Dr. Mylaraswamy has authored over two dozen papers and holds ten patents in the area of fault diagnosis and its applications.

Keyword Search in Text Cube: Finding Top-k Aggregated Cell Documents

Bolin Ding, Yintao Yu, Bo Zhao, Cindy Xide Lin,
Jiawei Han, Chengxiang Zhai
University of Illinois at Urbana-Champaign

We study the problem of keyword search in a data cube with text-rich dimensions (the so-called text cube). The text cube is built on a multidimensional text database, where each row is associated with some text data (e.g., a document) and other structural dimensions (attributes). A cell in the text cube aggregates a set of documents with matching attribute values in a subset of dimensions. A cell document is the concatenation of all documents in a cell. Given a keyword query, our goal is to find the top-k most relevant cells (ranked according to the relevance scores of cell documents with regard to the given query) in the text cube. We define a keyword-based query language and apply IR-style relevance model for scoring and ranking cell documents in the text cube. We propose two efficient approaches to find the top-k answers. The proposed approaches support a general class of IR-style relevance scoring formulas that satisfy certain basic and common properties. One of them uses more time for pre-processing and less time for answering online queries; and the other one is more efficient in pre-processing and consumes more time for online queries. Experimental studies on the ASRS dataset are conducted to verify the efficiency and effectiveness of the proposed approaches.

Dynamic Strain Mapping and Real-Time Damage State Estimation Under Biaxial Random Fatigue Loading

Subhasish Mohanty, Aditi Chattopadhyay,
John N. Rajadas, Clyde Coelho
Arizona State University

Fatigue damage and its prediction is one of the foremost concerns of structural integrity research community. The current research in structural health monitoring (SHM) is to provide continuous (or on-demand) information about the state of a structure. The SHM system can be based on either active or passive sensor measurements. Though the current research on ultrasonic wave propagation based active sensing approach has the potential to estimate very small damage, it has severe drawbacks in terms of low sensing radius and external power requirements. To alleviate these disadvantages, passive sensing based SHM techniques can be used. Currently, few efforts have been made towards time-series fatigue damage state estimation over the entire fatigue life (stages I, II, and III). A majority of the available literature on passive sensing SHM techniques demonstrates the clear trend in damage growth during the final failure regime (stage-III regime) or during when the damage is comparatively large enough. The present paper proposes a passive sensing technique that demonstrates a clear trend in damage growth almost over the entire stage II and III damage growth regime. A strain gauge measurement based passive SHM frameworks that can estimate the time-series fatigue damage state under random loading is proposed. For this purpose, a Bayesian Gaussian process nonlinear dynamic model is developed to map the reference condition dynamic strain at a given instant of time. The predicted strains are compared with the actual sensor measurements to estimate the corresponding error signals. The error signals estimated at two different locations are correlated to estimate the corresponding fatigue damage state. The approach is demonstrated for an Al-2434 complex cruciform structure applied with biaxial random loading.

Analyzing Aviation Safety Reports: From Topic Modeling to Scalable Multi-Label Classification

Amrudin Agovic, Hanhuai Shan, Arindam Banerjee
University of Minnesota

The Aviation Safety Reporting System (ASRS) is used to collect voluntarily submitted aviation safety reports from pilots, controllers, and others. As such it is particularly useful in researching aviation safety deficiencies. In this paper we address two challenges related to the analysis of ASRS data: (1) the unsupervised extraction of meaningful and interpretable topics from ASRS reports and (2) multi-label classification of ASRS data based on a set of manually predefined categories. For topic modeling we investigate the practical usefulness of Latent Dirichlet Allocation (LDA) when it comes to modeling ASRS reports in terms of interpretable topics. We also utilize LDA to generate a more compact representation of ASRS reports to be used in multi-label classification. For multi-label classification we propose a novel and highly scalable multi-label classification based on multi-variate regression. Empirical results indicate that our approach is superior to several baseline and state-of-the-art approaches.

Usage of Dissimilarity Measures and Multidimensional Scaling for Large Scale Solar Data Analysis

Juan Banda, Rafal Angryk
Montana State University

This work describes the application of several dissimilarity measures combined with multidimensional scaling for large scale solar data analysis. Using the first solar domain-specific benchmark data set that contains multiple types of phenomena, we investigated combinations of different image parameters with different dissimilarity measures in order to determine which combinations will allow us to differentiate our solar data within each class and versus the rest of the classes. In this work we also address the issue of reducing dimensionality by applying multidimensional scaling to our dissimilarity matrices produced by the previously mentioned combinations. By applying multidimensional scaling we can investigate how many resulting components are needed in order to maintain a good representation of our data (in a artificial dimensional space) and how many can be discarded in order to economize our storage costs. We present a comparative analysis between different classifiers in order to determine the amount of dimensionality reduction that can be achieve with said combination of image parameters, similarity measure, and multidimensional scaling.

A Knowledge Discovery Strategy for Relating Sea Surface Temperatures to Frequencies of Tropical Storms and Generating Predictions of Hurricanes Under 21st-century Global Warming Scenarios

Caitlin Race, Michael Steinbach, Vipin Kumar
University of Minnesota
Auroop Ganguly,
Oak Ridge National Laboratory
Fred Semazzi
North Carolina State University

The connections among greenhouse-gas emissions scenarios, global warming, and frequencies of hurricanes or tropical cyclones are among the least understood in climate science but among the most fiercely debated in the context of adaptation decisions or mitigation policies. Here we show that a knowledge discovery strategy, which leverages observations and climate model simulations, offers the promise of developing credible projections of tropical cyclones based on sea surface temperatures in a warming environment. While this study motivates the development of new methodologies in statistics and data mining, the ability to solve challenging climate science problems with innovative combinations of traditional and state-of-the-art methods is demonstrated. Here we develop new insights, albeit in a proof-of-concept sense, on the relationship between sea surface temperatures and hurricane frequencies, and generate most likely projections with uncertainty bounds for storm counts in the 21st-century warming environment based in turn on the Intergovernmental Panel on Climate Change Special Report on Emissions Scenarios. Our preliminary insights point to the benefits that can be achieved for climate science and impacts analysis, as well as adaptation and mitigation policies, by a solution strategy that remains tailored to the climate domain and complements physics-based climate model simulations with a combination of existing and new computational and data science approaches.

Tracking Climate Models

Claire Monteleoni, Shailesh Saroha
Columbia University
Gavin Schmidt
Columbia University and NASA Goddard Institute for Space Studies

Climate models are complex mathematical models designed by meteorologists, geophysicists, and climate scientists to simulate and predict climate. Given temperature predictions from the top 20 climate models worldwide, and over 100 years of historical temperature data, we track the changing sequence of which model currently predicts best. We use an algorithm due to Monteleoni and Jaakkola that models the sequence of observations using a hierarchical learner, based on a set of generalized Hidden Markov Models (HMM), where the identity of the current best climate model is the hidden variable. The transition probabilities between climate models are learned online, simultaneous to tracking the temperature predictions. On historical data, our online learning algorithm's average prediction loss nearly matches that of the best-performing climate model in hindsight. Moreover, its performance surpasses that of the average model prediction, which was the current state of the art in climate science, the median prediction, and least squares linear regression. We also experimented on climate model predictions through the year 2098. Simulating labels with the predictions of any one climate model, we found significantly improved performance using our online learning algorithm with respect to the other climate models, and techniques.

Adaptive Model Refinement for the Ionosphere and Thermosphere

Anthony D'Amato, Aaron Ridley, Dennis Bernstein
University of Michigan

Mathematical models of physical phenomena are of critical importance in virtually all applications of science and technology. This paper addresses the problem of how to use data to improve the fidelity of a given model. We approach this problem using retrospective cost optimization, a novel technique that uses data to recursively update an unknown subsystem interconnected to a known system. Applications of this research are relevant to a wide range of applications that depend on large-scale models based on first-principles physics, such as the Global Ionosphere-Thermosphere Model (GITM). Using GITM as the truth model, we demonstrate that measurements can be used to identify unknown physics. Specifically, we estimate static thermal conductivity parameters, and we identify a dynamic cooling process.

Keynote Address

Monitoring of the Changes in the Global Forest Cover Using Data Mining

Vipin Kumar
University of Minnesota

Assessing change in forest cover is of critical importance in studying natural and anthropogenic impacts on natural ecosystems. In particular, forest degradation accounts for almost 20% of anthropogenic greenhouse gas emissions (GHG) and thus is a significant driver of climate change, which, in turn can impact the health of the global ecosystem. Hence there has been a significant increase in international efforts such as the United Nations Program on Reducing Emissions from Deforestation and Forest Degradation (UN-REDD). In addition to politically negotiated treaties, a market-based approach has been proposed in which corporations or countries that are significant emitters of atmospheric carbon offer monetary payments for forest preservation in exchange for carbon credits to be used in a carbon trading system.

A key ingredient for effective forest management, whether for carbon trading or other purposes, is quantifiable knowledge about changes in forest cover. Rich amounts of data from remotely sensed images are now becoming available for detecting changes in forests or, more generally, land cover. However, in spite of the importance of this problem and the considerable advances made over the last few years in high-resolution satellite data acquisition, data mining, and online mapping tools and services, end users still lack practical tools to help them manage and transform this data into actionable knowledge of changes in forest ecosystems that can be used for decision making and policy planning purposes. Providing this actionable knowledge requires innovations in a number of technical areas: (1) identification of changes in global forest cover, (2) characterization of those changes, and (3) discovery of relationships between the number, magnitude, and type of these changes with natural and anthropogenic variables. To realize progress in the above areas, a number of computational challenges in spatio-temporal data mining need to be addressed. Specifically, analysis and discovery approaches need to be cognizant of climate and ecosystem data characteristics such as seasonality, inter-region variability, multi-scale nature, spatio-temporal autocorrelation, high dimensionality, and massive data size. This talk describes our initial efforts, achievements, and challenges in addressing some of the above areas.

Vipin Kumar is currently William Norris Professor and Head of Computer Science and Engineering at the University of Minnesota. His research interests include High Performance computing and data mining. He has authored over 200 research articles, and co-edited or coauthored 10 books including the widely used textbooks *Introduction to Parallel Computing* and *Introduction to Data Mining*, both published by Addison-Wesley. Kumar has served as chair/co-chair for over a dozen conferences/workshops in the area of data mining and parallel computing. In 2001, Kumar co-founded SIAM International Conference on Data Mining and served as its steering committee chair until 2007. Kumar is a founding co-editor-in-chief of *Journal of Statistical Analysis and Data Mining*, editor-in-chief of *IEEE Intelligent Informatics Bulletin*, and series editor of the Data Mining and Knowledge Discovery Book Series published by CRC Press/Chapman Hall. Kumar is a Fellow of the AAAS, ACM, and IEEE. He received the 2009 Distinguished Alumnus Award from the Computer Science Department, University of Maryland College Park, and 2005 IEEE Computer Society's Technical Achievement Award for contributions to the design and analysis of parallel algorithms, graph-partitioning, and data mining.

Invited Talk

Applying Avatar Machine Learning to NIF Optics Inspection Analysis

Laura Kegelmeyer
Lawrence Livermore National Laboratory

The National Ignition Facility (NIF) at Lawrence Livermore National Laboratory (LLNL) is a complex optical laser system designed and built for the purpose of achieving controlled nuclear fusion in a laboratory setting. This enormously interdisciplinary project creates tremendous amounts of data that must be mined and processed.

Machine learning is emerging as the way to handle huge quantities of data that are too cumbersome to handle with traditional mathematical or human analysis methods. The NIF Optics Inspection Analysis group is using a self-configuring pattern recognition tool called Avatar (WP Kegelmeyer, Sandia National Laboratories) to apply machine learning techniques to various aspects of the optics inspection project. We train Avatar to categorize flaws on optics and then use the resulting ensemble of decision trees during shot operations to quickly report on the status of a beamline. The use of Avatar has resulted in the status report going from unusable (more false alarms reported than real flaws) to fully functional with very high accuracy.

The advantage of Avatar over other available machine learning tools is that it has innovative advances built in to automatically determine necessary parameters so that the user can operate in a “hands-off” fashion. One of these innovations is an automatic, data-driven stopping condition for out-of-bag validation which determines how many trees are needed in the ensemble, or forest, to maximize accuracy of the output. Another innovation, Hellinger trees, improves accuracy when data is skewed (where one class, usually the most interesting, is far less represented than the other classes).

Through use of data understanding and decision-making technologies presented here and elsewhere at this conference, it will be possible to manage the onslaught of incoming data and use the computer to focus our attention on the most important areas of interest.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Laura Mascio Kegelmeyer has specialized in Image Processing & Analysis for the past 20+ years at Lawrence Livermore National Laboratory. For the first 12 years she focused on biomedical imaging, where she developed automated applications to evaluate digitized mammograms, genome sequence (probe mapping) and genetic expression (blots and microarrays), chromosome abnormalities (fluorescence microscopy), and fetal cells in maternal blood and bone structures.

Since then, she has concentrated on automating analysis of the optics on NIF, the world's most powerful laser. In all of these applications, she has used pattern recognition and machine learning as an applied tool, to evaluate features and determine how to best make conclusions or decisions in real time.

Kegelmeyer received an BS in Biomedical Engineering in 1986 and an MS in Electrical Engineering 1988, both from Boston University. For a list of her publications and patents, see https://www-eng.llnl.gov/bios/bios_kegelmeyer.html

PADMINI: A Peer-to-Peer Distributed Astronomy Data Mining System and a Case Study

Tushar Mahule, Sandipan Dey, Sugandha Arora, Hillol Kargupta
University of Maryland Baltimore County
Kirk Borne
George Mason University

Peer-to-Peer (P2P) networks are appealing for astronomy data mining from virtual observatories because of the large volume of the data, compute-intensive tasks, potentially large number of users, and distributed nature of the data analysis process. This paper offers a brief overview of PADMINI—a Peer-to-Peer Astronomy Data MINIng system. It also presents a case study on PADMINI for distributed outlier detection using astronomy data. PADMINI is a web-based system powered by Google Sky and distributed data mining algorithms that run on a collection of computing nodes. This paper offers a case study of the PADMINI evaluating the architecture and the performance of the overall system. Detailed experimental results are presented in order to document the utility and scalability of the system.

Multi-temporal Remote Sensing Image Classification: A Multi-view Approach

Varun Chandola, Ranga Raju
Oak Ridge National Laboratory

Multispectral remote sensing images have been widely used for automated land use and land cover classification tasks. Often thematic classification is done using single date image, however in many instances a single date image is not informative enough to distinguish between different land cover types. In this paper we show how one can use multiple images, collected at different times of year (for example, during crop growing season), to learn a better classifier. We propose two approaches, an ensemble of classifiers approach and a co-training based approach, and show how both of these methods outperform a straightforward stacked vector approach often used in multi-temporal image classification. Additionally, the co-training based method addresses the challenge of limited labeled training data in supervised classification, as this classification scheme utilizes a large number of unlabeled samples (which comes for free) in conjunction with a small set of labeled training data.

Distributed Anomaly Detection Using Satellite Data From Multiple Modalities

Kanishka Bhaduri
MCT Inc./NASA Ames Research Center
Kamalika Das
SGT Inc./NASA Ames Research Center
Petr Votava
CSU Monterey Bay/NASA Ames Research Center

There has been a tremendous increase in the volume of Earth Science data over the last decade from modern satellites, in-situ sensors, and different climate models. All these datasets need to be co-analyzed for finding interesting patterns or for searching for extremes or outliers. Information extraction from such rich data sources using advanced data mining methodologies is a challenging task, not only due to the massive volume of data, but also because these datasets are physically stored at different geographical locations. Moving these petabytes of data over the network to a single location may waste a lot of bandwidth, and can take days to finish. To solve this problem, in this paper, we present a novel algorithm which can identify outliers in the global data without moving all the data to one location. The algorithm is highly accurate (close to 99%) and requires centralizing less than 5% of the entire dataset. We demonstrate the performance of the algorithm using data obtained from the NASA MODerate-resolution Imaging Spectroradiometer (MODIS) satellite images.

Multi-label ASRS Dataset Classification Using Semi-supervised Subspace Clustering

Mohammad Salim Ahmed, Latifur Khan
University of Texas at Dallas
Nikunj Oza
NASA Ames Research Center
Mandava Rajeswari
Universiti Sains Malaysia

There has been a lot of research targeting text classification. Many of them focus on a particular characteristic of text data, multi-labelity. This arises due to the fact that a document may be associated with multiple classes at the same time. The consequence of such a characteristic is the low performance of traditional binary or multi-class classification techniques on multi-label text data. In this paper, we propose a text classification technique that considers this characteristic and provides very good performance. Our multi-label text classification approach is an extension of our previously formulated multi-class text classification approach called SISC (Semi-supervised Impurity based Subspace Clustering). We call this new classification model as SISC-ML (SISC Multi-Label). Empirical evaluation on real-world multi-label NASA ASRS (Aviation Safety Reporting System) data set reveals that our approach outperforms state-of-the-art text classification as well as subspace clustering algorithms.

Poster Session

A Comparative Study of Algorithms for Land Cover Change

Shyam Boriah/Varun Mithal/Ashish Garg/Vipin Kumar/Michael Steinbach, University of Minnesota; Chris Potter, NASA Ames Research Center; Steve Kloostyr, CSU Monterey Bay

Adapting to the Temporal Variations in ASRS Reports

Muhammad Abedi/Mohammad Masud/Raquibur Rahma/Latifur Khan, University of Texas at Dallas

Analyzing Aviation Safety Reports: From Topic Modeling to Scalable Multi-Label Classification

Amrudin Agovic/Hanhuai Sha/Arindam Banerjee, University of Minnesota

Automatic Data Representation, Regularization, and Visualization: Applications of Shannon Sampling on Graphs and Manifolds

Meyer Pesenson/Bruce McCollum, Caltech; Isaac Pesenson, Temple University; Michael Byalsky, Hebrew University of Jerusalem and Ariel University Center

Clustering Techniques for One-Class Support Vector Machine

Lian Yang/Rongliang Li/Qingwen Miao/Gang Liu/Guichong Li, University of Ottawa

Composite Risk Measures for Adaptive Dimensionality Reduction in Systems Health Monitoring

Nisheeth Srivastava/Jaideep Srivastava, University of Minnesota

Content-based Planetary Data Mining System

Mustafa Acer/Snehal Chennuru/Xinyao Hu/Russel Reed/Ed Katz, Carnegie Mellon Silicon Valley; Ara Nefian, NASA Ames Research Center

Correlated Topics in a Scalable Multidimensional Text Cube: Algorithms and Aviation Safety Case Study

Bo Zhao/Xide Lin/Jiawei Han, University of Illinois at Urbana-Champaign UIUC; Ashok Srivastava/Nikunj Oza, NASA Ames Research Center

Data Mining the Galaxy Zoo Mergers

Steven Baehr/Arun Vedachalam/Kirk Borne/Daniel Sponseller, George Mason University

Detection of Reconnection Exhausts in the Solar Wind Using Multi-Variate Time Series Classification

Tamara Sipes/Homa Karimabadi, SciberQuest, Inc; Jack Gosling, University of Colorado

Distributed Anomaly Detection Using Satellite Data From Multiple Modalities

Kanishka Bhaduri, MCT Inc./NASA Ames Research Center; Kamalika Das, SGT Inc./NASA Ames Research Center; Petr Votava, CSU Monterey Bay

Effective Outlier Detection In Science Data Streams

Kirk Borne/Arun Vedachalam, George Mason University

Efficient K-Median Clustering With Bit Sliced Column Data

Amal Perera, University of Moratuwa; Wiliam Perrizo, North Dakota State University

Four-Dimensional Compression for the Lossless Transmission of Remotely Sensed Data
Md. Al Mamun/Xiuping Jia/Michael Ryan, University of New South Wales

Gaussian Process Regression for Risk Forecasting in a PI Thermal Control System
Luis Gallo, NASA Goddard Space Flight Center

Intelligent Understanding and Processing of Airborne Sense and Avoid Radar Data with Antenna Diversities
Zhengzheng Li/Rockee Zhang, University of Oklahoma

Keyword Search in Text Cube: Finding Top-k Aggregated Cell Documents
Bolin Ding/Yintao Yu/Bo Zhao/Cindy Xide Lin/Jiawei Han/ Chengxiang Zhai, University of Illinois at Urbana-Champaign

Optimal Prediction of Adverse Events in Aviation Data
Rodney Martin, NASA Ames Research Center; Santanu Das, UARC/NASA Ames

Probability Calibration by the Minimum and Maximum Probability Scores in One-Class Bayes Learning for Anomaly Detection
Guichong Li, University of Ottawa

Recipes for the Estimation of Information-Theoretic Quantities to Analyze the Information Flow Between Different Variables
Deniz Gencaga/William B. Rossow, NOAA-CREST, The City College of New York; Kevin H. Knuth, University at Albany

Spacecraft Telemetry Monitoring by Dimensionality Reduction and Reconstruction
Takehisa Yairi/Minoru Inui, University of Tokyo; Yoshinobu Kawahara, Osaka University; Noboru Takata, Japan Aerospace Exploration Agency

Spatially Adaptive Semi-supervised Learning with Gaussian Processes for Hyperspectral Data Analysis
Goo Jun/Joydeep Ghosh, University of Texas Austin

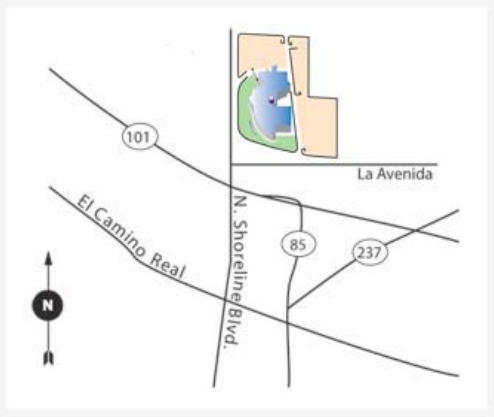
Tracking Climate Models
Claire Monteleoni/ Shailesh Saroha, Columbia University; Gavin Schmidt, Columbia University and NASA GISS

Usage of Dissimilarity Measures and Multidimensional Scaling for Large Scale Solar Data Analysis
Juan Band/Rafal Angryk, Montana State University

Information of Interest

To find the Conference Proceedings, please visit: https://dashlink.arc.nasa.gov

More information about the Computer Science Museum can be found at: http://www.computerhistory.org/



WiFi Access: Google Wifi is available and requires no credentials to access.