UVM

Computer Science

Research Day 2008

Program

Dec 12, 2008

8:30am-3:45pm

Jost Foundation Room
422 Davis Center

Organizer: Margaret J. Eppstein

SCHEDULE AND INDEX TO ABSTRACTS

8:30-9 Poster Setup, Coffee and bagels

9-10:20: WELCOME SESSION

9-9:05 Xindong Wu "Welcome to CS Research Day"

9:05-9:10 Maggie Eppstein: "Computer Science and Complex Systems"

(p.3) 9:10-10:10 Hava Siegelmann: KEYNOTE SPEECH: "What causes Jetlag?"

10:10-10:20 brief coffee break

10:25-12:30: COMPLEX SYSTEMS MODELING SESSION

10:20-1:25: Dean Domenico Grasso: "Complex Systems in CEMS and CS"

(p.4) 10:25-10:50: Josh Bongard "Using Genetic Programming to Explain Human Responses to an Aesthetic Preference Task"

(p.5) 10:50-11:15: K.N. Krishnanand (Krishna) "Rapid Synthesis of Approximate, Physically-plausible Models from Visual Data"

(p.6) 11:15-11:40: Josh Payne "Mixing patterns and Information Cascades"

(p.7) 11:40-12:05: Michael Pellon "Evolutionary Methods for Designing Multi-Species Heterogeneous Swarms with Self-organizing Oscillatory Behavior"

(p.8) 12:05-12:30: Jeff Sprenger "Interactive Simulation of Spiking Neural Networks"

12:30-1:45 LUNCH AND POSTER SESSION

(p.9) Mohammed Al-Kateb "Flash-based Reservoir Sampling over Data Streams"

(p.10) Song Wang "Providing location privacy and k-anonymity for mobile Users"

(p.11) Josh Auerbach "How Robot Morphology and Shaping Affect the Learning of Multiple Behaviors"

(p.12) Karim Chichakly "Optimizing Stormwater Management to Reduce Phosphorus Loads to Lake Champlain"

(p.13) Charley Robinson "A Distributed In Situ System for Snow Water Equivalence Measurement"

(p.14) Zhenyu Lu "Exploiting Multiple Model Types with Active Learning"

(p.15) Lei Chen "Self-synchronization in a Sensor Network"

1:45-3:45: AUTONOMOUS SENSORS AND COMPLEX SYSTEMS ANALYSIS SESSION

(p.16) 1:45-2:10: Sean Wang "Learning-based resource scheduling for autonomous sensor devices"

(p.17) 2:10-2:35: Peter Chapin "Programming Language Support for Wireless Sensor Network Reconfiguration"

(p.18) 2:35-3:00: Lingbo Yu "Classifying 3D Structures of Macro Molecules: A Case Study of Principal Component Analysis with Expectation Maximization"

(p.19) 3:00-3:25: Alison Pechenik "Crop differentiation from hyperspectral data: ANNs vs. SVM"

(p.20) 3:25-3:45: Gary Johnson "Generalized Source-Sink Models for Ecosystem Service Assessment: A probabilistic tool for environmental planning and research."

3:45: Maggie Eppstein – "Thanks for coming!"

9:10-10:10: KEYNOTE SPEECH: "What causes Jetlag?"

Hava T. Siegelmann
UMass Amherst (Computer Science) and Harvard University (Evolutionary Dynamics)

Systems in the body demonstrate different levels of alertness with respect to the rotation of earth and as governed by the circadian system. Organs' desyncrhonicity as occurring after travel and in shift workers was assumed to result by different re-entrainment rates of system components when following the change of phase in the master clock, the Supra Chariasmatic Nuclei (SCN). While simple, this assumption did not explain data where relative desyncrhonicity reach complete relative opposition. Collecting available data, we created a multistage non-linear system level modeling, were organs are characterized based on their actual dynamical oscillation patterns and are connected by the coupling strengths and relative delays. Our model demonstrates that the chief source of desyncrhonicity among organs reaches its peak following eastward phase advance of 5-8 hours time zones, and it occurs since some (not all) organs advance their clocks, following the SCN, by repetitive delays. This phenomena is termed ``antidromic reentrainment." Antidromic reentrainmnet explains the relative anti-correlated phases among organs. Studying the effect of the light protocols on the reentrainment, light protocols were proposed that reduce both relative phases and total reentrainment time following harsh changes in time zones (T. Leise and H. T. Siegelmann, "Dynamics of a multistage circadian system," Journal of Biological Rhythms, 21(4):314-323. 2006).

SPEAKER BIO:
Dr. Siegelmann received her BA in Computer Science from Technion University Summa Cum Laude, was awarded an Artist Certificate in Piano from the Rubin Conservatory in 1982, and received the State of Israel Best Undergraduate Medal in 1984. She completed her MSc in Computer Science Cum Laude at Hebrew University in 1992 with a thesis entitled "Document Allocation in Multiprocessor Information Retrieval Systems: An Application of Genetic Algorithms". One year later in 1993 she received her PhD in Computer Science from Rutgers as a "Doctoral Fellow of Excellence" where she completed her dissertation work on "Foundation of Recurrent Neural Networks". She was awarded the Alon Fellowship of Excellence of the Israeli National Committee for Higher Education which funded her research from 1994-1997, and in 1998 she published her book "Neural Networks and Analog Computation: Beyond the Turing Limit". She is currently an associate professor of Computer Science at U. Mass, Amherts, where she is Director of the Biologically Inspired Neural and Dynamical Systems Lab. She has 48 refereed journal articles published or in press in several prestigious computer science and interdisciplinary journals (including Science), 19 book chapters, and 50 conference papers. Her work has also received wide recognition in the popular media such as NPR and numerous magazines and newspapers. She is an associate editor for Frontiers in Computational Neuroscience, and is on the editorial board of the American Inst. of Physics Journal Chaos: An interdisciplinary Journal of Nonlinear Science. In 2007 we were fortunate to have her join our UVM CS Board of Advisors.

10:25-10:50 TALK: Josh Bongard

"Using Genetic Programming to Explain Human Responses to an Aesthetic Preference Task"

Josh C Bongard, Peter S Dodds, Chris M Danforth

Genetic Programming is an attractive modeling methodology in that, when it is applied correctly, it can produce human-intelligible models of the system's hidden internal structure. In this work I discuss the application of GP to model individual human subject's responses to an aesthetic preference task. I will show how the resulting GP models can both predict the subjects' responses, and uncover similarities and differences among the subjects' response strategies. By gradually expanding the maximum depth of the GP trees, the models gradually reveal how the subjects' strategies can be encoded as nonlinear combinations of the underlying features used to generate the images. This provides insight into which aspects of the stimuli different subjects responded to. This approach has value in that determining which aspects of complex stimuli an animal or human responds to is a notoriously difficult problem. I will conclude with a discussion as to how this approach could be generalized to other psychology studies.

10:50-11:15 TALK: K.N. Krishnanand (Krishna)

"Rapid Synthesis of Approximate, Physically-plausible Models from Visual Data"

K.N. Krishnanand[1], Josh C. Bongard[1], and Andrew N. Meltzoff[2]

[1]Department of Computer Science, University of Vermont, kkrishna@uvm.edu
[2]Institute for Learning and Brain Sciences, University of Washington

Modeling of 3D objects from 2D images is a challenging and still unsolved computer vision problem. We present a novel modeling framework that generates near real-time, approximate, yet physically-plausible representations of objects observed with a camera rather than creating accurate models that are computationally expensive to generate. The vision literature abounds with techniques that are capable of localizing and tracking moving objects in a scene by analyzing the sequence of images from a video footage of the event. However, our approach is driven by the dual goals of extracting the 3D shape and relative motion of objects of interest. Our approach to the modeling of observed scenes is based on carefully selecting a small subset of the total pixels available for visual processing. To achieve this, we use the exploration estimation algorithm (EEA) to create the visual models: a population of three-dimensional models is optimized against a growing set of training pixels, and periodically a new pixel that causes disagreement among the models is selected from the observed scene and added to the training set. We show here that using only 0.2% of the available pixels, the algorithm can generate approximate models of a simple scene. Moreover, we show that this method results in selection of pixels around the edges of observed objects, therefore leading to automated edge detection.

11:15-11:40 TALK: Josh Payne

"Mixing patterns and Information Cascades"

Joshua L. Payne, Margaret J. Eppstein, & Peter S. Dodds

How does an arcane song rise from obscurity to become a mainstream hit? How do obscure, low-budget videos become viral? In this talk, we will discuss a class of models designed to address these questions, and investigate how the structure of inter-individual interactions affects the propagation of such information. In particular, we will consider the influence of mixing patterns (*i.e.*, the propensity with which individuals of similar popularity interact with one another) on the size and duration of information cascades and on the ability of a single individual to trigger a cascade, as a function of their popularity.

11:40-12:05 TALK: Michael Pellon

"Evolutionary Methods for Designing Multi-Species Heterogeneous Swarms
with Self-organizing Oscillatory Behavior"

Michael Pellon, University of Vermont, Department of Computer Science
Margaret J. Eppstein, University of Vermont, Department of Computer Science

In recent years, artificial swarms have gained a tremendous amount respect for their ability to produce and maintain complex spatio-temporal patterns while lacking any form of centralized control. While inspired by naturally occurring systems such as ant colonies and beehives, many of these artificial systems have been embraced by the robotics, telecommunications and the entertainment industries as novel solutions to a number of the longstanding technical challenges. Despite a rich appreciation for their theoretical beauty, to date there are no formal methods for designing a swarm system were multiple species of agents each with their own unique set of rules to interact by work together to produce a desired spatio-temporal pattern. This work investigates a purely algorithmic approach for designing multi-species swarm systems that evolves both the size of and rule set employed by each species as a part of a much larger population of complete candidate solutions. Solutions with various numbers of species are then tested for their ability to produce and maintain simple oscillatory patterns even in the presence of minor disturbances in their genome. Approaches such as this that require little to no human intervention hold the promise of fostering developments in the emerging fields of micro-machining, nano-robotics and drug delivery.

12:05-12:30 TALK: Jeff Sprenger

Interactive Simulation of Spiking Neural Networks

Jeff Sprenger

Spiking neural networks (SNNs) offer several advantages over their sigmoidal counterparts. Some human visual recognition tasks execute in the order of 100ms, or ten synaptic stages. SNNs offer a more biologically plausible model that can process faster than sigmoidal neural networks and more suitable for recognizing temporal patterns. The author demonstrates a graphical, interactive SNN simulator to build basic circuits for temporal pattern recognition. Once derived, these basic circuits can be assembled into more complex recognizers using evolutionary algorithms.

12:30-1:45 POSTER PRESENTATION: Mohammed Al-Kateb

Flash-based Reservoir Sampling over Data Streams

Mohammed Al-Kateb and Byung Suk Lee

Department of Computer Science, The University of Vermont

Reservoir sampling is arguably the most common technique for obtaining a uniform random sample from a stream of tuples. The existing research on reservoir sampling assumes that a reservoir is maintained in either RAM or hard disk. With the flash memory gaining attention as a new storage medium, especially for embedded devices, the reservoir sampling in the flash memory naturally becomes an issue. This flash-based reservoir sampling, however, is quite challenging because of the erase-block-and-write-page characteristics of updating data in flash memory, which makes the conventional reservoir sampling inefficient. The conventional reservoir sampling relies on the basic operation of deleting a randomly selected tuple from the reservoir and then replacing the deleted tuple with a new sampled tuple. This random deletion over flash memory incurs the significant overhead of erase-block-and-write-page and, therefore, should be avoided as much as possible. In this paper we present *Flash-Reservoir*, a novel flash-based reservoir sampling algorithm that achieves this goal. The key approach is to order the tuples in the flash in such a way that the number of random deletions is minimized. This ordering can be done by looking ahead and probabilistically predicting which future tuples will replace which tuples in the reservoir and making sure the tuples currently in the in-flash reservoir are arranged in the same order as they will be replaced in the future. For this, we use a look-ahead list which keeps track of the probabilistic prediction of tuple replacements and a list of the position indices of tuples as they should be in an imaginary in-RAM reservoir, in addition to the in-flash reservoir storing the actual tuples.

12:30-1:45 POSTER PRESENTATION: Song Wang

"Providing location privacy and k-anonymity for mobile Users"

Song Wang and X. Sean Wang

Location **B**ased **S**ervices (LBS) is described as those information services delivering differentiated information based on the location from where a mobile user issues the request. A privacy concern arises as the request might contain the location information of the user and the attacker might connect the location information in a request with a user. Spatial cloaking is a classical method providing k-anonymity for mobile users in location based services, but it does not provide location privacy. After using spatial cloaking, the attacker could still be able to narrow down the request issuer to a specific enough area. SpaceTwist is a new framework that secures location privacy but not k-anonymity. To bridge this gap, we propose two new algorithms to achieve both location privacy and k-anonymity for mobile users in nearest neighbor query.

12:30-1:45 POSTER PRESENTATION: Josh Auerbach

"How Robot Morphology and Training Order Affect the Learning of Multiple Behaviors"

Joshua Auerbach and Josh C. Bongard

Automatically synthesizing behaviors for robots with articulated bodies poses a number of challenges beyond those encountered when generating behaviors for simpler agents. One such challenge is how to optimize a controller that can orchestrate dynamic motion of different parts of the body at different times. This work presents an incremental shaping method that addresses this challenge: it trains a controller to both coordinate a robot's leg motions to achieve directed locomotion toward an object, and then coordinate gripper motion to achieve lifting once the object is reached. It is shown that success is dependent on the order in which these behaviors are learned, and that despite the fact that one robot can master these behaviors better than another with a different morphology, this learning order is invariant across the two robot morphologies investigated here. This suggests that aspects of the task environment, learning algorithm or the controller dictate learning order more than the choice of morphology.

12:30-1:45 POSTER PRESENTATION: Karim Chichakly

"Optimizing Stormwater Management to Reduce Phosphorus Loads to Lake Champlain"

Karim J. Chichakly[1], Margaret J. Eppstein[1], William B. Bowden[2]

[1]Department of Computer Science, Complex Systems Center, UVM
[2]RubensteinSchool of Environment and Natural Resources, UVM

High-levels of phosphorus loading to Lake Champlain have been linked to algal blooms within the lake and to eutrophication of sections of the lake. Impervious land from development in the surrounding watersheds has increased the strength of rainfall runoff, and hence sediment and phosphorous loads, leaving many watersheds impaired. In this project, a process-based model will be incorporated into a multiscale, multiobjective genetic algorithm for determining optimal watershed management practices, trading off between effectiveness and cost. Automatic innovization techniques will be developed to extract fundamental design principles for "best solutions" from the optimal management practices selected by the genetic algorithm. This will be accomplished by examining characteristics of solutions along and orthogonal to the Pareto-optimal front.

12:30-1:45 POSTER PRESENTATION: Charley Robinson

"A Distributed In Situ System for Snow Water Equivalence Measurement"

Christian Skalka, Jeff Frolik, and Beverley Wemple

We describe a ground-based system that provides quasi real-time measurement and collection of snow-water equivalent (SWE) data in remote settings. The system is significantly cheaper and easier to deploy than current methods and is more robust to terrain and snow bridging effects. The system also enjoys several possible remote data recovery solutions. Compared to current infrastructure using existing technology, our system features will combine to allow more sites to be installed for the same cost and effort, in a greater variety of terrain, enabling data collection at improved spatial resolutions. The system integrates a new computational architecture with new sensor technologies. Our computational architecture is based on wireless sensor networks, comprised of programmable, low-cost, low-powered nodes capable of sophisticated sensor control and remote data communication. Our sensor technology works by measuring attenuation of electromagnetic radiation, an approach that is immune to snow bridging and significantly reduces sensor footprints.

12:30-1:45 POSTER PRESENTATION: Zhenyu Lu

"Exploiting Multiple Model Types with Active Learning"

Zhenyu Lu and Josh C. Bongard

Computer Science Dept., University of Vermont

Active learning is a technique that interactively chooses data points and trains classifiers on them. Many approaches to active learning involve iteratively training one classifier and choosing data points with the lowest confidence, but designing a confidence measure without bias is not trivial. An alternative approach is to use an ensemble of classifiers and choose data points that cause maximal disagreement among them. Many classifiers with different underlying structures could fit this framework, but some classifiers are more suitable for different data sets than others. The question then arises as to how to find the most suitable classifier for a certain data set. In this work, an evolutionary algorithm is proposed to address this problem. The algorithm starts with a combination of different types of classifiers and proceeds through a number of iterations. During each iteration, a data point is chosen based on the classifiers' uncertainty, the relative numbers of classifiers of each type are adapted according to their performances, and the set of classifiers are re-trained using an evolutionary algorithm. Currently, Artificial Neural Networks (ANN) and Decision Trees (DT) serve as the model types, but other types could be included as well. Our experiments with three synthetic data sets show that the algorithm performs more consistently across the sets than if only one model type was applied to each data set.

12:30-1:45 POSTER PRESENTATION: Lei Chen

"Self-synchronization in a Sensor Network"

Lei Chen , Dr. Jeff Frolik & Dr. Sean Wang

Self-synchronization mechanism has been shown capable of achieving global distributed decisions in a wireless sensor network with no need for a fusion center, which improves resilience against node failure and reduces the congestion probability. In this poster, a brief history of self-synchronization design is first summarized. Then a sensor network of 11 fully-connected nodes is used as an example to illustrate the design philosophy behind the self-synchronization. The dynamic states of the nodes are modeled as a coupled differential equation system and convergence of the nodes is studied numerically with these equations. Lastly, some ideas for future research work are presented, aiming at improving the current design. Possible applications of self-synchronization are also discussed.

1:45-2:10 TALK: Sean Wang

"Learning-based resource scheduling for autonomous sensor devices"

Sean Wang and Jeff Frolik

The goal of resource scheduling in a network of sensor devices is to provide sufficient sensing and actuation capabilities while avoiding excessive redundancy. Although redundancy provides needed resilience and robustness, too much of it can degenerates network performance due to unnecessary depletion of the limited energy reserve and transmission collisions within the limited bandwidth. In our work, we study a learning-based approach in resource scheduling to reduce redundancy but at the same time, to provide sufficient coverage to satisfy application requirements. An added requirement is that the network must be sufficiently resilient to survive random failures of individual devices or elimination of groups of devices. Our basic idea is that in addition to learning-based adaptation, we instill the devices with a probabilistic behavior that, combined with adaptation, gives a necessary jitter to respond to random events and to move towards an optimal resource scheduling (as seen in evolutionary computing). The talk will outline our approach and show some initial results.

2:10-2:35 TALK: Peter Chapin

"Programming Language Support for Wireless Sensor Network Reconfiguration"

Peter Chapin and Chris Skalka

The nesC programming language is commonly used for the development of wireless sensor network applications. nesC allows different software components to be statically wired together to form node-level applications. We are developing an extension to nesC that allows components on different nodes to be wired together in a manner that is similar to that provided by the existing nesC language. Our extension hides all node to node communication behind a remote procedure call abstraction. Because of the volatile nature of wireless sensor networks, the inter-node wiring must be adaptable to changes in network topology or function. Thus unlike the current nesC language, our extension provides a mechanism for supporting dynamic wiring configurations so that the structure of the overall program can change at run time.

2:35-3:00 TALK: Lingbo Yu

"Classifying 3D Structures of Macro Molecules: A Case Study of Principal Component Analysis with Expectation Maximization"

Lingbo Yu, Robert Snapp, Michael Radermacher

Principal Component Analysis (PCA) is one of the common methods for feature extraction. It has been successfully applied to classify 2D Electron Microscopy images based on the distance in the feature subspace. However, 3D reconstructions of heterogeneous samples in electron microscopy are incomplete due to experimental limitations, thus PCA cannot be applied directly. The goal of this research is to develop a method that is capable of correctly classifying 3D structures independent of missing data. An iterative probabilistic principal components analysis algorithm (Tipping and Bishop, 1999. and Roweis, 1998) was extended. Expectation Maximization (EM) algorithm is used to estimate principal features, as well as the missing data. The performance of this algorithm is studied with a simulated 3D data set from which different amount of data is removed (i.e., "missing"). Finally we proposed two different methods to recover the missing data in 3D reconstructions. Either class averaging can be applied to acquire a complete reconstruction followed a correct classification, or the missing data can be filled with the estimated data.

3:00-3:25 TALK: Alison Pechenik

"Crop differentiation from hyperspectral data: ANNs vs. SVM"

Ralf Wieland[1], Bernd Zbell[1] and Wilfred Mirschel[1], Alison Pechenick[2]

[1]Leibniz-Center for Agricultural Landscape Research (ZALF), Institute of Landscape Systems Analysis, Muencheberg, Germany
[2]College of Engineering & Mathematical Sciences, University of Vermont

Hyperspectral data can be used to indicate crop health. However, the amount of data collected across many bandwidths over time imposes a heavy computational overhead, especially when involving physics-based models or matrix manipulations (e.g., principal component and discriminant analyses). In this work, two alternative approaches are investigated: support vector machines and neural networks. To evaluate their accuracy and computational efficiency, these methods were applied to the task of classifying eleven different crops grown at the ZALF agricultural research institute east of Berlin.

3:25-3:45 TALK: Gary Johnson

"Generalized Source-Sink Models for Ecosystem Service Assessment: A probabilistic tool for environmental planning and research."

Gary Johnson, Ferdinando Villa, Robert Snapp, Sergey Krivov

A Generalized Source-Sink Model (GSSM) provides a spatial conceptualization of the probabilistic dispersion of some quantity, be it matter, energy, or information (which we shall call the asset) through an abstract space whose location elements are arranged according to some monotonic distance function. This space can be 2-dimensional like a plane or geospatial map, 3-dimensional like a volumetrically-specified world, or N-dimensional for abstract spaces such as communication networks. The model space is made asset-specific by assigning each location three key parameters, which may be calculated as functions of an asset-driven set of spatially dependent variables: 1) expected units of asset production, 2) likelihood of destructively absorbing a unit of the asset, 3) likelihood of non-destructively caching a unit of the asset. Additionally, an asset-specific network flow function will be used to compute the probabilistic flow topology over the location graph. By calculating point-to-point path probabilities (or remote effect bandwidths) between every source and target location and then measuring the change in these bandwidths due to varying the three location-specific parameters listed above, we can determine a global dependency relationship in asset conveyance between each target and every other location in the model space. This algorithm has been designed to be used in the context of probabilistically evaluating the flow of ecosystem services over landscapes from natural systems to human beneficiaries of the services, enabling the first ever quantitative mapping of provisionsheds and benefitsheds in the arena of ecosystem service assessment and valuation (ESAV).