

CSEE Research Review – Poster Abstracts¹

Friday, May 6, 2011

Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County (UMBC)

BS Students (2)

1. Ross A. Pokorny (Advisor: Tim Finin and Anupam Joshi), *ebiquity*
Tweet Collector: A Framework for Retrieving, Processing, and Storing Live Data from Twitter

Social media systems like Twitter and Facebook provide an important new source of information about emerging events, interests, opinions, and trends. While these posts are brief, they are rich in meta data and connected to complex social networks. Twitter is an especially interesting source due to its openness and high volume of over 100 million posts a day. Collecting, filtering, analyzing and storing information from a dynamic Twitter stream is an essential component for any system that derives information from it. I designed and implemented TweetCollector as a scalable system to automatically collect Twitter status updates matching a user-specified query. The received tweets are run through an extensible workflow, to which new components can be added as needed. After processing, the status updates, along with the data generated during the processing phase, are stored in a relational database for human inspection and further analysis. Scalability is achieved in a multicore environment through the use of multi-threading and resource pooling. TweetCollector ensures reliable collection of statuses with on-the-fly processing in order to allow social media researchers to rapidly discover and react to new information from a promising new data source. This work was funded, in part, by the Office of Naval Research through a subcontract from Lockheed Martin.

2. Morgan A. Madeira and Rachel Sweeton (Advisor: Anupam Joshi), *ebiquity*
Finding Communities through Social Media

Social media sites have rapidly increased in popularity over the past few years. We have investigated the potential value of collecting and analyzing data from these sites. We identified communities based on similarities in the data. One application of this procedure is determining areas affected by natural disasters based on data collected from networking sites. Our research focuses on identifying these communities automatically. To better understand how to approach the bigger problem, we started by collecting data from Twitter and analyzing it based on political sentiment. We examined Tweets to infer an individual's political party and used the sorted data to create our gold standard. The standard is used to train the computer to classify novel data sets. We have been working toward having the computer recognize communities and predict whether a person is associated with certain groups. This work was funded, in part, by Lockheed Martin and the Office of Naval Research.

¹ *Students who won research awards may also have posters, but they are ineligible for the three poster awards. For their abstracts, see the talk abstracts.*

3. Akshaya Iyengar (Advisor: Tim Finin), *ebiquity*
Estimating Temporal Boundaries for Events Using Social Media Data

Social media websites like Twitter, Flickr and YouTube generate a high volume of user generated content as a major event occurs. Our goal is to automatically determine as accurately as possible when the event starts and when it ends by analyzing the volume and content of social media data. We describe a technique that estimates the temporal boundaries of anticipated events like wildfires and hurricanes and helps to monitor changes as events unfold. Estimating these temporal boundaries segments the event-related data into three major phases: the buildup to the event, the event itself, and the post-event effects and repercussions. The technique can also detect the presence and scope of significant sub-events occurring during the course of an event. For events that transpire over time and space, such as the dispersal of an oil spill, a hurricane or a spreading wildfire, we can analyze how event progressed, traveled geographically and major sub-events that occurred within the event. When applied to natural disasters and man-made disturbances, the derived data can help organizations involved in mediation efforts to track and analyze evolving events.

4. Mohit Kewalramani (Advisors: Tim Finin and Anupam Joshi), *ebiquity*
Detecting Communities on Twitter

Twitter has recently evolved into a source of social, political and real time information in addition to being a means of communication and marketing. Monitoring and analyzing information on Twitter can lead to valuable insights which might otherwise be hard to get using conventional media. An important task in analyzing highly networked information sources like twitter is to identify communities that are formed. A community can be defined as a group of nodes that have more links within the set than outside it. We plan to present a technique for detecting communities in twitter using link structure, folksonomy, text and other available metadata of the tweets. The link structure on Twitter can be characterized using the twitter notion of followers, being followed and the @Mentions, @Reply and @RT tags in tweets. We intend to test the performance of clustering using the N-Cut algorithm and a bottom up agglomerative approach. The communities thus created can be characterized using the dominant tags, geo-location, sentiment associated with these tags and other meta-data used for clustering.

5. Adrian Rosebrock and Jesus Caban (Advisors: Charles Nicholas)
Automatic Pill Identification Using Polar Shape Matrices and Color Histograms

Medication errors are a leading cause of unexpected complications in hospitals and clinics. A crucial step often omitted by pharmacy stores and medical providers is an effective way to visually validate the prescription drugs given to patients. Automatic identification of prescription drugs opens new research opportunities in which statistical models can be created to help healthcare organizations at validating their drugs and patients at verifying their pills.

To describe shape and the imprint on the pill a Polar Shape Matrix (PSM) descriptor is used. The PSM is invariant to rotation, scale, and translation making it a robust feature. To compare the utility of the PSM, a simple method of finding the contour of the pill and calculating the triangle area between the centroid and consecutive contour points is used. As expected, the PSM outperforms the triangle method.

A color histogram is then used to describe color in both the normalized RGB and $L^*a^*b^*$ color space, where images in the $L^*a^*b^*$ color space perform slightly better. By using a comprehensive and high-resolution image inventory of the 658 most widely used solid dose medications we have analyzed the performance of PSM at describing the characteristics of a pill. Results show that 96.5% of the time the shape can be accurately identified and that supervised machine learning methods such as SVMs and k -NN perform similarly in this task. When combining shape, color, and imprint together, we are able to identify pills within images with 87.5% accuracy.

6. Aniket Bocharé, Darshana Dalvi and Yelena Yesha (Advisor: Yelena Yesha)
Genome-based Clinical Decision Support System

The complete sequencing of the Human Genome under Human Genome Project, a historic effort to decipher each of 3 billion letters has opened a new perspective in studying complex diseases. Over the few decades, epidemiological studies of large cohorts have been carried out which focus on individual's clinical behavior, signs, symptoms, medical and family history for diagnosis and treatment. Many studies do not take into account genetic variability of individuals within the population. There have been significant contributions made by researchers to identify potential risks behind complex chronic diseases such as Type II Diabetes; however the etiology has been long controversial. Our aim is to harness in individual strength of different approaches and build a model to provide an accurate, informative and appropriate diagnosis of Type II Diabetes. We intend to investigate the gathered data and tune it for optimum real time performance for using various data mining approaches in such a way that it will infer useful associations, patterns and clusters and help us understand factors in the pathology by statistically evaluating the results. Use of multi-core processors for data intensive computations and dissemination of the derived results and knowledge using Web Based Service Oriented system will provide real time support for integration, automation and scalability for various data sets from different medical institutions to the physicians at the point of care.

7. Nagapradeep Chinnam (Advisors: Tim Finin and Anupam Joshi), *ebiquity*
Group Recognition in Social Networks

In present scenario people are using social networks to get connected and to share information. On an average each individual has 300-400 connections. Managing these connections to easily share information without intruding into users' privacy poses a tricky challenge. An easy way to organize these connections is by grouping them and using the group semantics to develop an information sharing policy.

We propose to develop a machine learning system which can classify the user connections into groups given the training set which consists of groups and group instances. For practical demonstration of the system, we are using the data collected from Facebook. The major tasks are to identify the relevant features for the classification and to find a learning mechanism that best suits the problem. Another significant challenge is posed by hierarchical and overlapping groups. We show that our system classifies new connections into these groups with high accuracy even with 10-20% of labeled data.

8. Radhika Dharurkar (Tim Finin and Anupam Joshi), *ebiquity*
Context Modeling and Reasoning Approach in Context-Aware Middleware for Activity Recognition

With the advent of various wireless technologies, people want access to diverse kind of information anywhere and anytime with the personal devices they carry all the time. For such extremely mobile scenarios, instead of expecting help from the infrastructure, researchers focus on developing mobile applications which can work with existing infrastructure and explore the capability of smart devices and social networking.

We enhance the base framework for collaborative data gathering by capturing temporal, spatial information and also information from sources like user's Calendar for activity recognition. The system relies on machine learning and statistical analysis techniques which help the system better deal with the changing contexts in constantly changing environment. We introduce context-aware infrastructure for a location based smart phone system, the ontology based context modeling and reasoning approach. Furthermore, we show that our activity recognition system can be used in many applications to deliver high quality user experience to smart phone users.

9. Amit Mahale (Advisors: Tim Finin and Anupam Joshi), *ebiquity*
Group Centric Information Sharing Using Hierarchical Models

Traditional Information security frameworks are based on the principle of “Need to know,” However post 9/11 the focus is now on “Need to Share”. The role of Information sharing is growing in mission critical as well as social media applications. However Information sharing has to deal with the risks and benefits involved in the transaction. In this thesis, we focus on the Group centric information sharing model and develop an application framework to demonstrate its usefulness in current applications. The framework uses semantic web technology to represent and reason about the hierarchy in a group centric environment and uses java environment to make access decisions implementing group operation semantics.

10. Pavan Kumar Kamaraju and Zary Segall (Advisor: Zary Segall), *Convergent Design Lab*
Framework to investigate context-aware solutions for opportunistic content delivery

The last few years has seen an enormous growth in the number of smart phones in use and the trend is expected to continue in the near future as well with the availability of more processing power and larger form factors. This explosive growth has contributed to an unprecedented adoption of mobile data services which results in severe network congestion during peak hours due to limited availability of cellular capacity and the prohibitive cost of installing new infrastructure. Even the recent adoption of LTE by carriers has not kept pace with increasing data needs and the networks continue to be under stress. However, it is of grave importance for the provider to maintain superior Quality of Experience (QoE) and Quality of Service (QoS) even under constrained network capacity. One possible approach for achieving high QoE is through the use of context information from smart phones that allow the network to take pro-active decisions for content delivery.

Key players within the telco industry seem to be unanimous in foreseeing, for the next few years that the M2M traffic trends will grow rapidly. The Apps on mobiles, serving content, can be considered the present manifestation of the growing trend. Understanding the potential impact of M2M traffic on current network dimensioning and architectures is of paramount importance for the success and sustainability of these novel services in future wireless mobile networks.

This poster presents a tool “Melange” which can be used to measure and demonstrate the benefits of using context awareness for content delivery. Using the tool, the user can remotely control a number of mobile handsets by deploying different types of synthetic and natural workloads, emulating the traffic, to evaluate the impact on battery consumption and user perception of context-aware content pre-fetching.

11. Yan Kang, Jaewook Kim and Yun Peng (Advisor: Yung Peng), *ebiquity*
Extensible Dynamic Form for Supplier Discovery

Discovery of suppliers (supplier discovery) is essential for building a flexible network of suppliers in a supply chain. The first step for supplier discovery is to collect manufacturing capabilities of suppliers and requirements of customers. In traditional e-marketplaces, online form interfaces are typically used to collect the requirements and capabilities. However, those forms are mostly lack of flexibility to capture a variety of requirements and capabilities in a structured way. We propose a new innovative form architecture called eXtensible Dynamic Form (XDF) to facilitate data collection process of supplier discovery. This architecture provides several key innovations including: 1) architecture for users (suppliers or customers) to create new structure of form for their own contents; 2) an synonym-based intelligent search engine facilitating users to reuse the existing form components 3) hierarchical representation of the requirements and capabilities as XML instances. Experimental results demonstrate that the proposed architecture is valuable for facilitating the supplier discovery process.

12. Zhong Ren and Mohamed Younis (Advisor: Mohammed Yonis), *Wireless Sensor Networks Lab*
Exploiting Architectural Techniques for Boosting Base Station Anonymity in Wireless Sensor Networks

Wireless Sensor Networks (WSNs) are deployed to serve mission-critical applications in hostile environments such as battlefield and territorial borders. In these setups, WSNs may be subject to attacks in order to disrupt the network operation. The most effective way for an adversary to do so is by targeting the Base-Station (BS), where the sensor data are collected in the field. By identifying and locating the BS, the adversary can launch attacks to damage or disrupt the operation of the BS. Therefore, maintaining the BS anonymity is of utmost importance in WSNs.

Three approaches are proposed to boost the anonymity of the BS nodes to protect them from potential threats. We first explore the deployment of more BS nodes. We compare the BS anonymity of one versus multiple stationary BS under different network topologies. Results show that having more base-stations can boost both the average and max anonymity of BS nodes. We further provide guidelines on a cost versus anonymity trade-off to determine the most suitable BS count for a network. Secondly we exploit the mobility of base-stations and explore the effect of relocating some of the existing BS nodes to the lowest anonymity regions. Results show that having one mobile BS can dramatically boost the anonymity of the network and moving multiple BS does not provide much value. Finally, we propose to pursue dynamic sensor to cluster re-association to confuse the adversary. This can be employed when base-stations cannot safely move.

13. Charles Lohr (Advisor: Zary Segall), *Center for Convergent Design*
Semantic Light: Building Blocks

The concept of Semantic Light is simply that lighting systems can be aware of what they are lighting. This offers a number of potential advantages over conventional lighting in quality and efficiency. Semantic Light requires fine grained control of the output of many lights and requires sensors to take in information about what is being lit. It uses this information to control the output lighting in great detail. By running various algorithms, semantic light can provide information to the user and has a number of applications including augmented reality.

14. Kavin Rathinam Kasinathan and Mohamed Younis (Advisor: M. Younis), *Embedded Systems & Networks*. **Distributed Approach for Mitigating Coverage Loss in Heterogeneous Wireless Sensor Networks**

In a heterogeneous wireless sensor network, nodes with different sensing capabilities are dispersed throughout an area of interest. Nodes with similar capabilities are not necessarily collocated. When a node fails, the area in the vicinity of such a node is left uncovered and the application may be negatively impacted. In this paper, we present a distributed algorithm for mitigating the coverage loss caused by the failure of a sensor node. The proposed algorithm looks for one or multiple nodes that can be repositioned in order to fill the coverage gap. A search is conducted within the 2-hop neighborhood to identify a node or a combination of more than one node that can collectively possess the capabilities of the failed node. If no such combination of replaceable nodes exists, among the 2-hop neighbors, a further node that has similar capabilities is tasked to periodically substitute the failed node to provide the necessary coverage. In order to maximize the performance of the network, the replaceable nodes are chosen such that there is a minimum coverage and connectivity impact on the network after replacement. In the addition, the distance a node will travel to replace the failed node is also taken into consideration to limit the recovery overhead.

15. Sanket Gupte and Mohamed Younis (Advisor: Mohammad Younis), *Embedded Systems & Networks*
Vehicular Networking for Intelligent and Autonomous Traffic Management

Despite the push for using mass transit, the number of vehicles on the road is growing at a steady rate and traffic congestion has become a daily problem that most people suffer. This not only impacts the productivity of the population but also poses a safety risk. Most of the technologies for intelligent highways focus on safety measures and increased driver awareness, and expect a centralized management for the traffic flow. The aim of our research is to enable autonomous and adaptive traffic management through vehicular networks. By allowing data exchange between vehicles about route choices, congestions and traffic alerts, a vehicle makes a decision on the best course of action. Unlike centralized schemes that provide recommendations, a driver factors in the destination and routes of nearby vehicles in deciding on whether rerouting is advisable. The collective effect of all vehicles will reshape the traffic pattern based on their destination and road conditions. To validate our approach we have developed a graphical tool that not only enables the collection of performance statistics but also allows visualizing the effect on traffic. The implementation also supports smart traffic lights and configurable roads.

16. Jerome Lourdu Vikram Stanislaus and Tinoosh Mohsenin (Advisor: Tinoosh Mohsenin), *Energy Efficient High Performance Computing Lab*
Implementation of Full Parallel Low Density Parity Check Decoder for 802.11ad WLAN

Low Density Parity Check (LDPC) decoding has received significant attention due to its superior error correction performance, and has been adopted by recent communication standards such as IEEE 802.11ad, the next generation of WLAN standard. While there has been much research on LDPC decoders, high-speed systems that require many processing nodes typically suffer from large wire-dominated circuits operating at low clock rates due to large critical path delays caused by the codes' inherently irregular and global communication patterns.

This work presents an FPGA implementation of a full parallel single pipelined LDPC decoder for the 802.11ad WLAN standard. The decoder uses a (672, 588) rate 7/8 LDPC code, with BPSK modulation over AWGN channel. The 5-bit fixed-point decoder implementation has an approximate 0.4 dB SNR loss when compared to the floating point implementation which shows a tradeoff between the error correction performance and routing complexity. The synthesis results for a Virtex-6 XC6VLX760 FPGA indicate that the decoder utilizes 83964 slice counts, which is 70% of total FPGA slice resources. It operates at 45 MHz and with 11 decoding iterations it delivers 2.7 Gbps throughput.

PhD Students (11)

17. Josiah Dykstra (Advisor: Alan Sherman), *Cyber Defense Lab*
Understanding Issues in Cloud Forensics

The inevitable vulnerabilities and criminal targeting of cloud environments demand an understanding of how digital forensic investigations of the cloud can be accomplished. We present two hypothetical case studies of cloud crimes; child pornography being hosted in the cloud, and a compromised cloud-based website. Our cases highlight shortcomings of current forensic practices and laws. We describe significant challenges with cloud forensics, including forensic acquisition, evidence preservation and chain of custody, and open problems for continued research.

18. Lushan Han and Tim Finin (Advisor: Tim Finin), *ebiquity*
PowerRelations: A Question Answering System for Semantic Web Data

Large amounts of structured and semi-structured semantic data are available on the Web. A well-known example is DBpedia, which extracts data from Wikipedia, encodes it in the Semantic Web language RDF, and stores it in a triplestore. Although a formal query language, SPARQL, is available for accessing such data, it remains challenging for users to query the knowledge unless they are familiar with SPARQL and the particular ontologies used. We have developed an intuitive system for users to express queries by describing entities and relations using natural language terms in a simple graphical interface. Our system automatically translates the user's intuitive description into a corresponding SPARQL query that produces an answer. Our key contribution is the robust techniques mapping user terms in a variety of expressions to the most appropriate concepts and relations used in DBpedia, even though its ontologies are diverse and relatively informal due to the nature of Wikipedia and the noisy information extraction process. Our approach combines a statistical analysis of the DBpedia knowledge base that includes correlations of concepts and relations and lexical semantic similarity metrics learned from WordNet and a large text corpus. We disambiguate user input terms by exploring all possible interpretations and selecting the best one based on correlation and similarity. To improve recall, we further harvest properties similar to best choices taking into account the context just disambiguated. Initial experiments show that the system works very well on a collection of test questions from the 2011 Workshop on Question-Answering over Linked Data.

19. Varish Mulwad (Advisor: Tim Finin), *ebiquity*
Generating Linked Data by Inferring the Semantics of Tables

A vast amount of information is encoded in tables on the web, spreadsheets and databases. Considerable work has been focused on exploiting unstructured free text; however techniques that are effective for documents and free text do not work well with tables. Early work in table interpretation in the field of document analysis and later on the Web, focused mainly on understanding and extracting tables from scanned documents and html web pages. Relatively little work has addressed the understanding and interpretation of the semantics and meaning associated with tables. In this work, we present a framework for understanding and interpreting the “semantics” of tables. The meaning of the table is often encoded in the column headers, the relations implicit between the various columns, the table’s caption, as well the text surrounding the table. Using this evidence, augmented with a background knowledge base such as the Linked Open Data cloud, our framework will map every column header to a class from an appropriate ontology, link the data values to existing entities in the linked data cloud (or map them as values of a property wherever appropriate) and discover and identify relations between various columns. The interpreted semantics will be represented as linked RDF assertions which can be used for further reasoning.

20. Wenjia Li, Anupam Joshi and Tim Finin (Advisor: Anupam Joshi), *ebiquity*
CARE-CPS: Context-Aware tRust Evaluation for Cyber-Physical System Using Policies

A Cyber-Physical System (CPS) involves a tight coupling between the physical and computational elements to form a situation-aware system that responds intelligently to dynamic changes of the real-world scenarios. Security is a key challenge for the deployment of CPS: Due to environmental influences, the CPS data are inherently noisy that may contain many false reports; Moreover, the reporting devices (e.g., sensors, cameras) in CPS may even be hacked by adversaries to deliberately generate fake readings. Therefore, it is highly desirable to extract correct information from a large volume of noisy data and properly evaluate the reputation of reporting devices in CPS. In this paper, we propose a Context-Aware tRust Evaluation scheme for wireless networks in CPS (CARE-CPS), in which the trustworthiness of each reporting device is assessed based on both the reporting history of the device and the environment in which the abnormal readings occur. We declare a set of policy rules to accurately describe how we determine the reputation of each reporting device based on these factors. To validate the CARE-CPS scheme, we have conducted experiments in terms of both simulation and real deployment on smart phones. Experimental results show that the CARE-CPS scheme can properly evaluate the trustworthiness of the report devices in CPS.

21. Robert Weiblen, Andrew Docherty, Jonathan Hu and Curtis Menyuk (Advisor: Curtis Menyuk)
Calculation of the Expected Bandwidth for a Mid-Infrared Supercontinuum Source Based on AS₂S₃ Chalcogenide Photonic Crystal Fibers

We computationally investigate supercontinuum generation in an As₂S₃ solid core photonic crystal fiber (PCF) with a hexagonal cladding of air holes using a pump wavelength of 2.8 μm . We find that significant variation of the measured bandwidth with small changes in the system parameters implies that one cannot accurately calculate the experimentally-expected bandwidth from a single numerical simulation, since the output spectrum of a supercontinuum generation source is extremely sensitive to the input pulse parameters, including pulse duration, peak power, and noise. Additionally, supercontinuum output spectra display a complex structure that makes an unambiguous bandwidth definition nontrivial. We propose a novel method for bandwidth calculation that is robust to variations in the output spectrum. We also propose the use of a smoothed and ensemble-averaged spectrum that is expected to be a better predictor of the bandwidth that would be seen in experimental systems. We show that bandwidth fluctuations are considerably reduced, allowing us to calculate the bandwidth more accurately. Using this method, we maximize the bandwidth in our source to obtain a spectrum that extends from 2.5 μm to 6.2 μm with an uncertainty of $\pm 0.5 \mu\text{m}$. The optimized bandwidth is consistent with prior work, but with significantly increased accuracy.

22. Lloyd Emokpae and Mohamed Younis (Advisor: Mohammed Younis)
Surface-Reflection Based Communication and Localization in Underwater Sensor Networks

Most communication and localization algorithms in underwater environments have been constrained by dependencies on the line of sight (LOS), which is hard to guarantee due to the inherent node mobility. This constraint hinders node discovery and ad-hoc formation in underwater networks and limits the performance of routing protocols. This paper introduces a novel surface-based reflection (SBR) model that uses a homomorphic deconvolution technique to establish water-surface-reflected communication links. We then propose a surface-based reflection anchor-free localization (SBR-AL) algorithm that can be employed by the individual nodes to establish a relative coordinate system. Our approach also employs a switch-beamed directional antenna model that allows each node to use the LOS estimated from SBR-AL to enable directional communication which is beneficial for higher signal-to-noise ratios (SNR). The relative locations can facilitate the various network operation functions such as geo-routing and collision-free medium access. The simulation results confirm the effectiveness of the proposed approach.

23. Yichuan Gui, James Lo and Yun Peng (Advisors: Yun Peng and James Lo), *ebiquity*
Training Neural Networks by Using Risk-Averting Error Criteria with a Dynamically Selected Risk- Sensitivity Index

In training a neural network, the convexity region of a risk-averting error (RAE) with a risk-sensitivity index (RSI) expands as the RSI increases. Moreover, the RAE converges to the mean squared error (MSE) as the RSI decreases to zero, suggesting a method to solve the local-minimum problem. The method consists of a convexification phase to create tunnels or wormholes leading to a global minimum or global near minimum and a deconvexification phase to return to the MSE, for which the training is intended.

This paper proposes a procedure of dynamically selecting the RSI in the convexification phase that is insensitive to the architecture of the neural network under training or the selection of the initial guess of its synaptic weights. This procedure thus eliminates the need for multiple training sessions with different initial weight vectors, and consistently produces a satisfactorily trained neural network.

Many numerical examples are provided, demonstrating the effectiveness of the procedure and showing the superior training results as compared with those obtained from the MSE. Equally important, the procedure is illustrated to work on training data with a high level of noise.

24. Kostas Kalpakis, Shiming Yang and Yaacov Yesha
Near Real-time Data Assimilation for the HYSPLIT Aerosol Dispersion Model

HYSPLIT is a model developed by NOAA's Air Resources Laboratory for forecasting aerosol trajectories, dispersion, and concentration from emission sources. It is used extensively by NOAA to routinely provide a number of data products.

We develop a data assimilation system for assimilating observational data into the forecasting model in order to improve its forecasting accuracy. Our system is based on the Local Ensemble Transform Kalman Filter (LETKF) algorithm and it is computationally efficient. We evaluate our data assimilation system with real in-situ observational data, and find that our system improves upon HYSPLIT's forecast by reducing the normalized mean squared error and the bias.

We are also experimenting with assimilating MODIS data with HYSPLIT model forecasts. To this end, we extrapolate ground concentrations from MODIS Aerosol Optical Depth (AOD) data. Our extrapolation approach relies on spatially localized linear regressions of aerosol concentrations from ground stations in the Air Quality System (AQS) network and MODIS AOD data. We expect that assimilating the extrapolated concentrations leads into further improvements of HYSPLIT forecasts.

Our near real-time data assimilation system helps in bridging the gap between predictions and real-time observations, for more accurate and timely aerosol dispersion forecasts.

25. James Macglashan and Marie Desjardins (Advisor: Marie desJardin), *MAPLE Lab*
General Online Policy Transfer between Problems with Different State Representations

A typical challenge in reinforcement learning paradigms is generalizing experience so that learning is accelerated. Ways to generalize experience include transferring the policy from one problem to a similar new problem (where a problem is defined by the state space and reward function) and generalizing the experience from one state to additional similar states within the same problem. To share policies between different problems, a direct mapping between states of different problems may be assigned so that the policy from a source problem may be directly queried. For generalizing experience between states of the same problem, function approximation may be used. In function approximation, states are typically represented by a feature vector. This feature vector is then used as input to a learned function that estimates Q-values (the utility of taking an action in a given state), thereby sharing experience between states with similar feature vectors.

Ideally, an agent should be able to generalize experience in both of these ways. However, policy transfer between problems with different state space representations becomes problematic when function approximation is used because there is not a one-to-one state mapping between problems and because the learned function for a source problem takes a different kind of feature vector as input.

We define Object-oriented Sparse Distributed Memories, a means of function approximation that allows policy transfer to occur between problems with different state representations. Additionally, we describe how to perform transfer online, before source problems are finished learning.

26. James Cahill, Olukayode Okusaga, Weimin Zhou and Gary Carter (Advisor: Gary Carter)
Stimulated Rayleigh Scattering (STRS) in Optical Fibers

The effects of Stimulated Rayleigh Scattering (STRS) in optical fibers have largely been disregarded in favor of more dramatic effects such as stimulated Brillouin scattering (SBS). However, unlike the SBS spectrum, the STRS spectrum does not experience a significant frequency shift relative to the incident light beam. Therefore, while the magnitude of STRS is less than that of SBS in optical fiber, the noise induced by STRS is well within 1 MHz of the incident light. This spectral proximity to the incident light beam leads to greater degradation of a low-phase-noise radio frequency (RF) modulation signal than is induced by SBS. Therefore, STRS is a significant limiting factor for low noise fiber systems such as optoelectronic oscillators (OEOs).

In this poster, I present an experimental study illustrating the nonlinear gain of STRS by measuring the magnitude of counter-propagating light induced within 1 MHz of the incident light beam in 6 km single mode fiber (SMF) and dispersion shifted fiber (DSF) spools. My data shows a threshold power near -1 dBm and a saturation power near 8 dBm. Furthermore, I report that the threshold power can be increased by frequency modulation of the incident light. Finally, I report a decrease in OEO phase noise of about 10 dBc/Hz at 10 Hz when the input laser is frequency modulated to raise the STRS threshold above the oscillating power.

27. Karuna P. Joshi (Advisor: Yelena Yesha and Tim Finin), *ebiquity and Multicore Computational Center*.
Lifecycle of Virtualized Services on the Cloud

The delivery of IT services is increasingly based on the composition of multiple services and assets that are supplied by one or more service providers distributed across the network -- in the cloud. Managing these virtualized services efficiently over the cloud is an open challenge. There is no holistic view of what would constitute a lifecycle of virtualized services delivered on a cloud environment. We are developing a new integrated framework for automating the configuration, negotiation and procurement of IT services in a cloud computing environment using semantic web technologies. As part of this framework we have proposed a novel lifecycle for virtualized services on the cloud and divided it into five phases of requirements, discovery, negotiation, composition, and consumption. We have developed high level ontologies for each phase and are prototyping the system using technologies like OWL, RDF and SPAQRL.