CSEE Research Review – Program
Friday, May 1, 2009
Technology Center, South Campus[1]

Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County (UMBC)
www.csee.umbc.edu/events/research review

A celebration of research accomplishments by UMBC CSEE faculty and students in 2008–2009.
Free and open to the public (space limited to 125). Ample free parking.

Reception (9:00am–9:20am), with coffee, orange juice, doughnuts

Opening Remarks, Charles Nicholas, Chair, Dept. of CSEE (9:20am–9:30am)

Session I (9:30am–10:40am) – Each talk is 20mins.

Charles Nicholas
"Who wrote this document?"

Adam Anthony (Advisor: Marie desJardins), MAPLE Lab – *Honorable mention research by PhD student*
Fast relational clustering using the block modularity clustering objective

Peter A. Hamilton (Advisor: Marie desJardins), MAPLE Lab – *Award for best Research by BS student*
Applying swarm rule abstraction to a wireless sensor network domain

Break (10:40am–11:00am)

Session II (11:00am–12:10noon)

Tim Finin, ebiquity Group
Creating and exploiting a web of (semantic) data

Aaron Curtis (Advisor: Marc Olano), VANGOGH Lab – *Honorable mention research by MS student*
Real-time soft shadows on the GPU via Monte Carlo Sampling

Deepak Chinavle (Advisor: Tim Oates), Coral Lab – *Award for best research by MS student*
Adversarial classification: An ensemble-based approach

Free Lunch, in Lobby (12:10noon–1:00pm),with sandwiches, hot entree, salad, drinks, and brownies

Student Poster Session (1:00pm–2:10pm) – *vote for three best posters using Range Voting*

Session III (2:10pm–3:30pm)

Marc Olano, VANGOGH Lab
Simulation in real-time computer games

Fusan Yaman, Research Assistant Professor, MAPLE & Coral Labs
A context driven approach for workflow mining

Muhammad A. Talukder (Advisor: Curtis Menyuk), Photonics Lab – *Award for best research by PhD student*
Analytical and computational study of self-induced transparency mode locking in quantum cascade lasers

Student Poster Awards and Reception (3:30-3:45pm)
Adjourn (3:45pm)

Organizer: Alan T. Sherman

---

[1] Directions: Take Gun Road off Rolling Road (Rt. 166). The UMBC "Satellite" Shuttle stops at South Campus.

CSEE Research Review – Talk Abstracts
Friday, May 1, 2009

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

Session I

Charles Nicholas
"Who wrote this document?"

Questions of authorship have fascinated historians, theologians, and other scholars for centuries. In recent years statisticians and now at last computer scientists are also addressing these issues. We will present an overview of the study of authorship attribution, including famous examples such as "The Federalist Papers," the various "Wizard of Oz" books, as well as the Hebrew Bible and the Christian New Testament. We will present results from our own work in applying *Latent Semantic Analysis (LSA)*, a well-known technique in information retrieval, to the authorship attribution problem.

Adam Anthony (Advisor: Marie desJardins), MAPLE Lab – *Honorable mention research by PhD student*
Fast relational clustering using the block modularity clustering objective

I discuss a new algorithm for clustering in relational data that emphasizes scalability to large relational data sets. The algorithm uses a unique clustering objective, referred to as *Block Modularity*, which measures the quality of any general relational pattern, including community, hierarchical and *k*-partite relational patterns. First, I motivate the area of relational clustering in general, and then compare and contrast two popular approaches to relational clustering. Then, I describe the block modularity objective and the resulting optimization algorithm, and present empirical evidence that the new algorithm finds relatively high quality clusterings for a very low cost in runtime. Finally, I briefly discuss areas of current and future work, including the remaining contributions in my dissertation: the probabilistic relational clustering framework, multi-relational clustering, and relation selection.

Peter A. Hamilton (Advisor: Marie desJardins), MAPLE Lab – *Best Research by BS student*
Applying swarm rule abstraction to a wireless sensor network domain

Rule abstraction is a powerful tool for modeling abstract behaviors in swarm systems. The research presented in this paper examines the application of rule abstraction to the wireless sensor network domain. I analyze the potential of rule abstraction to accurately model and control the connectivity, coverage, and density of a simple sensor network. I also present a simulation tool developed to facilitate the discovery and creation of new abstract rules and discuss preliminary experimental results that will lead to the development of new abstract rules.

Tim Finin, ebiquity Group
Creating and exploiting a web of (semantic) data

Twenty years ago Tim Berners-Lee proposed a distributed hypertext system based on standard Internet protocols. The Web that resulted fundamentally changed the ways we share information and services, both on the public Internet and within organizations. That original proposal contained the seeds of another effort that has not yet blossomed: a Semantic Web designed to enable computer programs to share and understand structured and semi-structured information easily. We will review the evolution of the idea and technologies to realize a Web of Data and describe how we are exploiting them to enhance information retrieval and information extraction.

Aaron Curtis (Advisor: Marc Olano), VANGOGH Lab – *Honorable mention research by MS student*
Real-time soft shadows on the GPU via Monte Carlo Sampling

Realistic shadows present a difficult problem in real-time rendering. While techniques for rendering hard edged shadows from point light sources are well established, attempts to incorporate soft shadows typically suffer from inaccuracies or poor performance.

Our algorithm makes use of recent advances in GPU randomization to perform Monte Carlo sampling of points on an area light source. Rays are then traced to the sampled points, using the shadow map as a discretized representation of occluders in the scene. The accuracy of this method can be improved through the use of multiple shadow maps, which together are able to better approximate the scene geometry.

As with conventional shadow mapping, our method is performed entirely on the GPU, does not require any precomputation, and can handle fully dynamic scenes with arbitrary geometric complexity. The quality of the generated shadows is comparable to that of off-line rendering algorithms such as ray tracing, while performance remains real-time, on par with existing techniques.

Deepak Chinavle (Advisor: Tim Oates), Coral Lab – *Best research by MS student*
Adversarial classification: An ensemble-based approach

Spam has been studied and dealt with extensively in the email, web and, recently, the blog domain. Recent work has addressed the problem of non-stationarity of data using ensemble based approaches. Adversarial classification has been handled by retraining base classifiers using labeled samples obtained from the ensemble. However, frequent retraining is expensive. The need is to determine dynamically when the classifiers should be retrained and to retrain only those classifiers that are performing poorly. We show how mutual agreement between classifiers can be use to reduce retraining time, measure runtime performance, and keep track of the weakest performing classifier. We back our research with experimental results using real life data from blogs as a special case of spam

Marc Olano, VANGOGH Lab
Simulation in real-time computer games

Modern computer games perform physical simulation in a variety of contexts. It is perhaps most visible in kinematic simulation of articulated bodies, the so-called "rag doll physics" for character animation. However, simulation can also be found for cloth, fluid, optical effects with participating media, and models of surface reflectance. These methods must run effectively on a range of consumer hardware, and must be fast, finishing within a fraction of the 10–30 ms time available per frame. This talk will present some of the mainstream and research methods, demonstrate their results, and discuss the state of consumer-level hardware to accelerate physical simulation.

Fusan Yaman, Research Assistant Professor, MAPLE & Coral Labs
A context driven approach for workflow mining

Workflows play an important role in automation. The key challenge in building workflows is the need for a domain expert. However, in many realistic domains, experts are either not readily accessible or cannot express their knowledge in a declarative way. A solution to this bottleneck is workflow mining, which aims to discover a workflow given sample executions. In this talk, I will present a novel workflow mining algorithm, *WIT,* which applies grammar inferencing techniques to discover the target workflow with the help of the dataflow. WIT is designed to cope with the challenge of learning when training examples are scarce. It has been deployed as a major component in the POIROT system, which is built to take DARPA's integrated learning challenge, where several learners interact to solve different parts of a complex learning problem given only one example solution.

Muhammad A. Talukder (Advisor: Curtis Menyuk), Photonics Lab – *Best research by PhD student*
Analytical and computational study of self-induced transparency mode locking in quantum cascade lasers

The possibility of using the self-induced transparency effect to modelock lasers has been discussed since the late 1960s, but has never been observed. It is proposed that quantum cascade lasers are the ideal tool to realize self-induced transparency modelocking due to their rapid gain recovery times and relative long coherence times, and because it is possible to interleave gain and absorbing layers. Designs of quantum cascade lasers are presented here that satisfy the requirements for self-induced transparency modelocking. Analytical modelocked solutions of the coupled Maxwell-Bloch equations that define the dynamics in quantum cascade lasers that have both gain and absorbing layers have been found under the conditions that there is no frequency detuning, the absorbing layers have a dipole moment twice that of the gain layers, the input pulse is a π pulse in the gain medium, and the gain recovery times in the gain and the absorbing layers are much longer than the coherence time T_2 and are short compared to the round-trip time. It is shown that the modelocked pulse durations are on the order of T_2, which is typically about 100 fs. The Maxwell-Bloch equations have been solved computationally to determine the robustness of the modelocked solutions when frequency detuning is present, the dipole moment of the absorbing layers differs from twice that of the gain layers, the gain relaxation time is on the order of 1–10 ps, as typically obtained in quantum cascade lasers, and the initial pulse is not a π pulse in the gain medium. We find that modelocked solutions exist over a broad parameter range.

CSEE Research Review – Poster Abstracts
Friday, May 1, 2009

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

Note: The students who won research awards also have posters, but they are ineligible for the three poster awards. For their abstracts, see the talk abstracts.

MS Students

1. Wesley Griffin, VANGOGH Lab
  Creating Lines on the Geometry Shader

  Recent advances in graphics hardware include capabilities that have yet to be fully explored. One specific capability is the processing of mesh adjacency information in the programmable geometry shader. Mesh adjacency information is the set of triangles that are adjacent to the primary triangle currently being processed. Having access to adjacent triangles allows a shader program to compute information normally not available in the graphics pipeline, such as edge orientation or surface curvature data.
     A popular domain in non-photorealistic rendering is line drawing. There have been many techniques developed to draw lines in an artistic manner. These techniques can be divided into two groups: image-based and object-based. Image-based techniques work by rendering objects to a frame buffer and then processing the image to create lines. Object-based techniques, on the other hand, work with the polygonal mesh representation of objects, and analyze various aspects of the geometry to create different types of lines.
     Recent object-based techniques specifically analyze the surface curvature of polygonal meshes to extract lines. Typically these algorithms would be run on the host computer and the lines would be textured and rendered on the graphics hardware. These algorithms are potentially uniquely suited to the geometry shader hardware capabilities. We will develop a system that creates and renders stylized lines for generic models and runs completely in graphics hardware.

2. Albert Kir, Tejaswini Kavallappa (Advisor: Joel Morris)
 Develop statistical and deterministic signal processing algorithms for enhanced mid-IR sensor performance

  A mid-IR laser-based gas sensing system based on *Laser-based Spectroscopy (LAS)* is an important application in numerous fields. We used detection theory to develop a generic statistical analysis model for mid-IR gas sensing systems to compute the different performance probabilities. A new signal strength measure, (delta) SNR, is defined to address the detection problem for the gas sensing system and used to perform system improvement analysis for the detection performance of the sensing system.

3. Lushan Han (Advisors: Tim Finin, Yelena Yesha and Anupam Joshi), ebiquity Group
   Finding the Most Appropriate Semantic Web Terms from Words

   The Semantic Web language RDF was designed to unambiguously define and use ontologies to encode data and knowledge on the Web. Many people find it difficult, however, to write complex RDF statements and queries because doing so requires familiarity with the appropriate ontologies and the terms they define. We develop a system that maps a set of ordinary English words to a set of most appropriate (considering both term consistency and ontology popularity) RDF terms among an interconnected ontology network. We use the *Swoogle Semantic Web* search engine to provide RDF term and ontology co-occurrence statistics, the WorldNet lexical ontology to resolve synonyms, and a practical three step approach to find the best suitable ontology context as well as the most appropriate terms for the input words.

4. Palanivel Kodeswaran (Advisor: Anupam Joshi), ebiquity Group
   Towards a Declarative Framework for Managing Application and Network Adaptations

   Cross layer optimizations are increasingly being used for a variety of applications for pushing application intelligence into the network layer with the overall goal of improving application specific metrics. However most of these implementations are ad hoc and performed on a per application basis. In this paper we propose a declarative framework for managing application and network adaptations. The declarative framework provides a much needed clean line of separation between the high level goals and the low level implementations. Our framework exposes the tunable features of both the application and the network across layers which can then be jointly optimized through operator specified policies. This allows operators to control the adaptation process and retain control over their networks while the application and the network adapt in response to changing conditions. We pursue an ontology based approach and use semantic web languages such as OWL, RDF etc. in our framework for the policy and declarative specifications, thereby leveraging the inherent reasoning and conflict resolution features of these languages. We then describe our simulator developed on top of the NS2 simulator to demonstrate the utility of our approach in the easy implementation of cross layer optimizations through sample application scenarios.

5. John Krautheim (Advisors: Dhananjay Phatak and Alan Sherman), Cyber Defense Lab
   Virtual Private Infrastructure

   Cloud computing places an organization's sensitive data in the control of a third party, introducing a significant level of risk on the privacy and security of the data. We propose a new management and security model for cloud computing called the *Private Virtual Infrastructure (PVI)* that shares the responsibility of security in cloud computing between the service provider and client, decreasing the risk exposure to both. The PVI datacenter is under control of the information owner while the cloud fabric is under control of the service provider. A cloud Locator Bot pre-measures the cloud for security properties, securely provisions the datacenter in the cloud, and provides situational awareness through continuous monitoring of the cloud security. PVI and Locator Bot provide the tools organizations require to maintain control of their information in the cloud and realize the benefits cloud computing.

6. Wenjia Li (Advisors: Anupam Joshi and Tim Finin), ebiquity Group
  Policy-based Malicious Peer Dectection in Mobile Ad Hoc Networks

  Mobile Ad Hoc Networks (MANETs) are susceptible to various node misbehaviors due to their unique features, such as a highly dynamic network topology, rigorous power constraints and error-prone transmission media. While significant research efforts have been made to address the problem of misbehavior detection, little research work has been done on distinguishing truly malicious from simply faulty behaviors. We are developing a policy-based malicious peer detection framework that collects and use context information to determine the likely intent of a peer that is misbehaving. The context information includes features such as the communication channel status, buffer status, and transmission power levels. Our simulation results show that framework can distinguish malicious from faulty peers with high confidence. Moreover, the mechanism converges to a consistent view of malicious nodes amongst all the nodes with a limited communication overhead.

7. Justin Martineau (Advisor: Tim Finin), ebiquity Group
  Delta TFIDF: An Improved Feature Space for Sentiment Analysis

  Mining opinions and sentiment from social networking sites is a popular application for social media systems. Common approaches use a machine learning system with a bag of words feature set. We present *Delta TFIDF*, an intuitive general purpose technique to weight word scores efficiently before classification. Delta TFIDF is easy to compute, implement, and understand. We use Support Vector Machines to show that Delta TFIDF significantly improves accuracy for sentiment analysis problems using three well known data sets.

8. Don Miner (Advisor: Marie desJardins), MAPLE Lab
  Learning Non-Explicit Control Parameters of Swarm Systems

  Swarm-level behavior of a swarm system is easily measurable. For example, the density of a boid flock can be measured by observing the area covered by the flock, divided by the number of agents. However, adjusting the *Explicit Control Parameters (ECP)* of the system's program to generate specific non-explicit behavior is non-trivial. Determining the agent-level parameter values of the boid flock that results in the swarm exhibiting a specific density is difficult. Our approach is to use common and novel machine learning approaches to learn correlations between the ECP and user-defined *Non-Explicit Control Parameters (NECP),* which represent more abstract concepts in the system. Users adjust the value for an NECP, which is then translated to values for the ECPs that the program can handle directly. NECPs provide more intuitive and more efficient user control of these swarm systems since they represent more abstract or swarm-level concepts of the system. In addition, NECPs can be used as predictors to determine what how the swarm will behave without running an experiment. Our main contribution is a general framework for defining non-explicit control parameters. Our work focuses on a few popular domains: Reynolds boid flocking, particle swarm optimization, wireless sensor network layout, and traffic simulations. Approaches we have investigated for this purpose include linear regression, gradient descent, perceptrons and *k*-nearest neighbor.

9. Rory Mulvaney (Advisor: Dhananjay Phatak)
  Regularization and Diversification against Overfitting and Over-specialization

  In machine learning, regularization against over fitting attempts to solve the problem caused when a learning algorithm assumes that all training patterns are always present together. Cross validation provides a more realistic distribution of training data, since it trains with various subsets, weighted by appropriate probabilities. *Basis Function Regularization (BFR)* is a new augmentation to an objective function that attempts to efficiently emulate cross validation without requiring cross validation subsets, and furthermore provides a cross validation routine with a single regularization parameter to optimize, reducing risk of over fitting. Experiments confirmed BFR's usefulness for regularizing least squares regression against over fitting.

  It is felt that over-specialization is an analogous phenomenon, in human researchers, to over fitting. Just as financial advisors recommend broad diversification, it is imagined that, in order to diversify, researchers should publish their work in a homogenized "Priority-setting Market," where the "price" of an issue is proportional to its priority.

  In the spirit of diversification and systems engineering, several other promising projects have taken shape, including: a cache-motivated clustering objective, update lists for relocatable data objects, in-cache *B*-trees for extension words, a universal programming language framework with loadable syntax modules, and process management note-taking techniques.

10. Zareen Syed (Advisor: Tim Finin), ebiquity Group
  Wikitology: A Wikipedia Derived Hybrid Knowledgebase

  We are developing *"Wikitology"* as a Wikipedia derived novel hybrid knowledge base using Wikipedia and other related knowledge resources to expose the knowledge hidden in different forms such as RDF triples, links, graphs, tables and free text to applications thereby enabling effective access and utilization of world knowledge. We have successfully developed and evaluated Wikitology 1.0 system, a blend of the statistical and ontological approach for predicting concepts in documents. An enhanced version of Wikitology, "Wikitology 2.0" was constructed as a knowledge base of known individuals and organizations as well as general concepts for use in the ACE cross document co-reference task by incorporating structured data in RDF from DBpedia and Freebase and encoded in an RDFa-like format. The evaluation results showed high precision (0.966) and reasonably high recall (0.72). We are currently working on our Wikitology 3.0 system by focusing on enhancements targeted towards TAC 2009 Knowledge base population task for persons, organizations and locations. It involves extraction of information about entities with reference to an external knowledge source. The main tasks include entity linkage and ontology slot filling. We have incorporated and integrated data from Freebase, DBpedia and Wikipedia to construct entity link graphs for persons, organizations and geo-locations. We plan to implement graph based algorithms on the entity link graphs to aid both the entity linking and slot filling tasks. We are also employing Wikitology for named entity disambiguation which would support the entity linking task.

Abstracts from Students Unable to Attend

Patricia Ordonez (Advisor: Marie desJardins), MAPLE Lab
A Multivariate Time-Series Visualization of Clinical and Physiological Data

We present an approach that creates a multivariate time-series representation, a Multivariate Time Series Amalgam (MTSA), of physiological and clinical data that medical providers can visually interpret. It enables medical providers to receive more personalized and well rounded information about a patient to make more informed and efficient decisions about an individual's care. The representation serves as a visual model of a patient's state over time and is organized in a manner so that a dependency between the data and the state of four vital organs- the heart, lung, liver and kidney- is emphasized. The objective of the visualization is to provide an integrated, visual patient history in a time critical situation which emphasizes the change in parameter values over time.