Volume 6 | Summer 2020

Intelligent Systems Program

Newsletter

Table of Contents

Director's Address

Dr. Vanathi Gopalakrishnan

Associate Professor,
Department of Biomedical Informatics & Intelligent Systems Program

Greetings! As I reflect upon my first year as Director of the ISP, I am in awe of our entire community as we continue to navigate successfully through this adventure of knowns and unknowns! There is an ancient verse in the Hitopadesha ("Hitam" = well-being, Upadesha = "advice") that goes as follows:

विद्या ददाति विनयं विनयाद्याति पात्रताम्
(vidyaa dadaathi vinayam, vinayaad yaati paathrataam)
पात्रत्वाद्धनमाप्नोति धनाद्धर्मं ततः सुखम्
(paathratvaad dhanam aapnothi, dhanaad dharmam thathah sukham)

This roughly translates into:

Knowledge bestows humility (discipline), from which arises capability (worthiness), from which one gets wealth (and enrichment), using which one does good deeds (is righteous), which leads to joy (contentment).

I feel truly privileged to be a part of this knowledgeable ISP community which has grown to approximately 35 faculty from across 11 departments within the University of Pittsburgh. This past year we had 5 new approved secondary faculty appointments, with the primary affiliations of these Pitt faculty being Computer Science, Biomedical Informatics, Political Science, or the Learning Research and Development Center (LRDC). We also welcomed our largest incoming batch of 9 PhD students during the 2019-2020 year.

The School of Computing and Information (SCI) was quick to adapt to online instruction when the novel coronavirus pandemic changed the way we work. This allowed the ISP to continue its operations as normally as possible, resulting in a total of 3 successful doctoral dissertation defenses and 1 Master's degree granted this past year. The ISP also participated in the German-American Chamber of Commerce sponsored East Coast Industry Forum, which solicited our expertise in AI applications for personalized education and medical imaging.

You will find herein interesting information about many of our new faculty, alumni, and what we have been up to as a group. I am fascinated by the vision that our founders had for the ISP, and how we have become central to applied AI research within the University of Pittsburgh. I predict that over the next decade, we will become even larger, stronger and more unified as a diverse, creative community of AI researchers and educators. This would align with SCI's vision of becoming a powerhouse of research and educational innovation in multidisciplinary AI.

The best of AI and its integrative applications are yet to take shape! Stay curious...

ISP Alumni Highlights

Dr. Vincent Aleven received the LRDC's 2019 Distinguished Alumni Award. This award is given every two years to alumni who have demonstrated excellence in research of significant impact and other indicators of national and international recognition.

Dr. Giuseppe Carenini became the Chief Scientific Officer of ConVISation Labs. In collaboration with the WeITel company, ConVISation Labs transfers their research on text analytics to the healthcare domain.

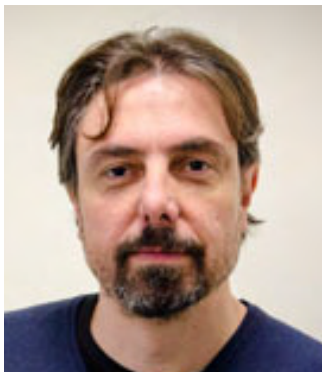
Dr. Violetta Cavalli-Sforza is promoted as Associate Professor of Computer Science in the School of Science and Engineering at Al Akhawayn University in Ifrane, Morocco. Starting in August 2019, she has also taken on the role of Coordinator of the Center for Advancement of Teaching, Learning and Scholarship (ATLAS) and is Assistant Vice President of Academic Affairs.

Dr. Bruce McLaren received the Fulbright Scholarship to further educational technology to teach a course on Educational Technology at the Universidad Té chnica Federico Santa Maria in Valparaiso, Chile. The goal of the Fulbright project is to help the development of educational technology in Chile, both as a research discipline and as a practical instructional tool for schools.

Dr. Gaurav Trivedi is volunteering to be on the jury of a program called WomenWhoTech. It is awarding $10,000 grants for women-led start-ups on a rolling basis. Contact Dr. Trivedi if you are interested!

Dr. Ioannis Tsamardinos founded a startup company called Gnosis Data Analysis and recently launched their first product: Just Add Data Bio (JADBIO). JADBIO fully automates the predictive modeling process, drastically boosting the productivity of the expert user. But, at the same time, it is so easy to use that it democratizes machine learning to non-experts. It requires no coding, math, statistics, or machine learning expertise. Our main design philosophy is that the user declares **what** they want to achieve from their analysis, **not how**. JADBIO has the intelligence to do the rest on its own. JADBIO is the only AutoML tool specifically designed for low-sample, high dimensional data, meaning it has no problem analyzing multi-omics and other molecular datasets with 500,000 features. It can handle classification, regression, but also time-to-event (survival analysis) outcomes, the latter being a unique functionality. JADBIO is also **correct**, in the sense that it does not overestimate predictive performance even with tiny sample sizes. Emphasis has also been placed on **feature selection** as quite often discovering the "relevant" features is the primary task in life sciences, while the actual predictive model may be secondary. JADBIO demanded the invention of plenty of new algorithms. In addition, they have published 13 scientific papers with novel results obtained with JADBIO and a few clicks, ranging from nanomaterial property prediction to bank failure prediction, and of course, several molecular biology results. Now, their focus is on adoption and wide use.

Give it a shot and let Dr. Tsamardinos know what you think by getting a free trial at jadbio.com . If you would like to contribute in any way, please contact Dr. Tsamardinos.



Dr. Tsamardinos started the idea of developing such tools when he was a PhD student at ISP!

Dr. Jaromir Savelka and Dr. Kevin Ashley won the Jurix 2019 Best Student Paper Award for their paper titled "Computer-Assisted Creation of Boolean Search Rules for Text Classification in the Legal Domain". Jurix 2019 was the 32nd International Conference on Legal Knowledge and Information Systems.

Dr. Michael Young was appointed as the Director of the University of Utah's Entertainment Arts and Engineering (EAE) Program in January of 2020. He is also a professor in Utah's School of Computing and an adjunct professor of Philosophy. The EAE program is an academic unit focused on games research and teaching, and has been ranked by Princeton Review in the top five programs worldwide 4 times in the last five years.

ISP Faculty Highlights

Dr. Maria Chikina received the 2020 Chancellor's Distinguished Research Award. This award annually recognizes outstanding scholarly accomplishments of members of the University of Pittsburgh's faculty.

Dr. Rosta Farzan and her colleagues were awarded an NSF grant for a project titled "Quantifying Hyperlocal Digital Equity: A Path to Supporting Digital Participation". The project focuses on understanding skills and opportunity gaps in the marginalized communities of Pittsburgh, in face of the Covid-19 pandemic and the shift to online services in all areas of life, including health, education, and social services.

Dr. Farzan also received A fellowship by the Leverhulme Trust in the UK to visit the University College London for 3 months in summer 2021 to conduct research on a project entitled "Developing Research and Training Agenda in Promoting Data-Driven Civic Engagement among Youth in Marginalized Communities" in collaboration with Dr. Licia Capra at the Computer Science department at UCL.

Dr. Farzan spent 6 weeks at the Universidad Té cnica Federico Santa Maríain Valparaiso, Chile to collaborate on establishing research and education in the social computing area. The collaboration has also led to an ongoing research project with the graduate students in the social computing course on study of the 2019 Chilean Protests and an analysis of the framing of the week 1 of the protests in English and Spanish Social and Traditional News Media.

Dr. Vanathi Gopalakrishnan was chosen as a distinguished alumnus from the B.M.S. College of Engineering's Department of Computer Science and Engineering. Dr. Gopalakrishnan was among the first Computer Science and Engineering graduates of this popular institution.

Dr. Mike Lewis and collaborators from CMU and Northrop Grumman were awarded a DARPA/US DoD grant under the Artificial Social Intelligence for Successful Teams (ASIST) program for a project entitled "A Robust and Adaptive Agent That Supports High Performance Teams" in an attempt to develop agents capable of making theory-of-mind inferences using a belief-desire-intention (BDI) model.

Dr. Lewis and collaborators were also awarded Army Research Laboratory grants under the Strengthening Teamwork for Robust Operations in Novel Groups (STRONG) program for projects entitled "Individualized Adaptation in Human Agent Teams" and "Strengthening Robust Teamwork in Human Agent Teams".

Dr. Diane Litman received 2020 Provost's Faculty Award for Excellence in Doctoral Mentoring. This award annually recognizes outstanding mentoring of graduate students seeking a research doctorate degree.

Dr. Litman and Dr. Godly from LRDC were awarded an NSF grant from the division of Information and Intelligent Systems (IIS) for a project entitled "Discussion Tracker: Development of Human Language Technologies to Improve the Teaching of Collaborative Argumentation in High School".

Dr. Shandong Wu has been promoted to the rank of Associate Professor with Tenure in the Department of Radiology at the University of Pittsburgh.

The Pittsburgh Center for AI Innovation in Medical Imaging (CAIIMI), led by the director of Dr. Shandong Wu with two associate directors, was launched in January 2020 and to date it has engaged 98 members (researchers and clinicians) from 12 different departments at Pitt, UPMC, and CMU. CAIIMI's vision is to build a center for advanced innovative AI research, clinical translation, commercialization, and collaborations in medical imaging by synergizing computational expertise and clinical resources in Pittsburgh. The Center presents as a unique multi-disciplinary team and uses a convergence approach to integrate distinct expertise to develop trustworthy imaging AI. CAIIMI received a scaling grant ($400K) from the Pitt Momentum Funds provided by the Provost and Senior Vice Chancellor for Research, in addition to the startup funding form the Department of Radiology. The Center has been submitting proposals to apply for large center grants from NIH and NSF for research and training and is in the process of building strong academic-industry sponsorship with multiple industry partners.

Many PhD students at the ISP program are involved in the CAIIMI research efforts including Giacomo Nebbia, Saba Dadsetan, Degan Hao, Jun Luo. There are great research and career opportunities at CAIIMI, and ISP faculty and students are welcome to join CAIIMI's efforts. If interested, send an email to Dr. Wu.

Momentum Funds Awarding Ceremony

New ISP Faculty Members

We are delighted to welcome five new faculty members who are doing amazing research on such a diverse array of subjects.

Dr. Sofia Triantafillou

Assistant Professor,
Department of Biomedical Informatics

Dr. Triantafillou, it is a pleasure meeting you today; we are excited to have you in the ISP family.

How does being an ISP faculty member fit into your career goals? what things do you look forward to?

While I work in the school of medicine, my work is mostly on core machine learning algorithms, and my plan for my lab is to do work on both clinical/biological applications and algorithmic development. Being a member of ISP faculty is particularly important for the latter, and I am looking forward to working with ISP students on developing novel causal discovery and inference methods.

What are your current research interests? What would you like other ISP faculty and students to know about your research group?

My research is driven by the need to go beyond correlational relationships in data analysis. I devise methods to infer causal relationships from mixtures of datasets of different origins, for example, both observational and experimental data. The rationale for my work is that there is an unprecedented abundance of data measuring the same system under different conditions, and yet each of them is analyzed in isolation, and the synthesis of scientific knowledge is done manually by the researchers. I believe that causal modeling can help automate this process.

What is your vision for the future of your research field and where do you think it is headed in five years (or beyond)?

Causal discovery methods have been around for several decades, but are not yet established in standard data analysis. I believe one reason for this is that there is a gap between theoretical advances in causal discovery and applied causal inference.

This is because real problems are very complex, and interdisciplinary teams are necessary to successfully apply causal discovery methods. My vision is that in the future, we will bridge this gap and we will see more systematically successful, practical applications of causal discovery in real-world problems.

What are one or two of your proudest professional accomplishments?

I am very proud of my PhD. Being a grad student is very challenging on many levels, but I am happy that I followed this career trajectory, because I really enjoy research!

How many graduate students are you advising now?

At the moment, I am only co-advising a MSc student with my past advisor, so if any student is interested in causal discovery, applied or theoretical, come find me!

Are you accepting new students at this point? If yes, do you have any messages for the ISP students who wish to join your research group?

Yes, my lab is new and I have open positions! I am looking for students who have a strong mathematical background and are intrigued by causality and want to figure out if we can learn causal relationships from data.

Do you have any work experience or collaboration with industry?

I have some experience with tutoring causal discovery methods for data science teams in the industry. I personally prefer the academic environment, but I have many good friends (and a husband) who left academia for the industry and are extremely happy with their choice, so I always encourage my colleagues and my students to explore this option. Causal discovery is also a growing field in the industry, with many big companies interested in applying data-driven methods to guide their policies, so I welcome collaborations of any kind.

Do you have any new courses or publications that you would like ISP students to know about?

I just co-authored a paper with Greg Cooper on using published experimental data and large observational data to learn causal models. I am very excited about this research, because in my search for applications I have always found that the experimental data are not publicly available, particularly in the clinical domain. This work enables integrating large observational cohorts and published causal effects, and I believe is the first step towards my goal for automating the scientific process! I will also teach an introduction to causal models in neuromatch, an awesome conference organized by my PostDoc mentor, Konrad Kording.

What advice would you give to ISP students who are currently seeking a job?

I would advise them to be confident! I know it can be intimidating to be in the job market, but your skills are in high demand and you will be very successful in finding a job you enjoy, whatever this may be.

How do you spend your spare time?

I really enjoy baking, and I spend a lot of my spare time doing that!

Thank you Dr. Triantafillou for introducing yourself and we appreciate the time you took to speak with us.

Dr. Erin Walker

Associate Professor,
Learning Research and Development Center

Dr. Walker, it is a pleasure meeting you today; we are excited to have you in the ISP family.

How does being an ISP faculty member fit into your career goals? what things do you look forward to?

I am excited to be part of a program that focuses on multidisciplinary, use-inspired AI research, and to connect with other students and faculty who have those strong interests. I am looking forward to this program opening up new collaborations for me within Pitt.

What are your current research interests? What would you like other ISP faculty and students to know about your research group?

I do research in personalized learning environments, with a particular emphasis on how these technologies apply to social interaction between collaborating students or between a student and an agent. My lab designs, implement, and evaluate the effects of educational technologies in both formal and informal settings, spanning multiple learning domains and ages. My research group is highly collaborative and welcomes the exchange of ideas with other students and faculty.

What is your vision for the future of your research field and where do you think it is headed in five years (or beyond)?

I think we will see intelligent learning technologies that can support individual students, collaborating groups, and classroom interactions, and that are sensitive in a holistic way to students' past learning experiences and current learning context. I also think we'll see the rise of new kinds of immersive learning experiences, including robotic learning environments, VR/AR, and multimodal sensing and responsiveness to student learning states. We will see new blended models for learning that combine digital and physical learning experiences.

What are one or two of your proudest professional accomplishments?

I am generally proud of the way my lab builds intelligent systems that can function in real-world contexts and contribute meaningfully to our understanding of learning theory. I like enabling learning experiences that would not be possible without technology.

How many graduate students are you advising now?

Three PhD students. There are also two master's students who are quite active in my lab.

Are you accepting new students at this point? If yes, do you have any messages for the ISP students who wish to join your research group?

I have openings for new students in my lab who are interested in the intersection of AI, HCI, and learning technologies. Students who tend to be a good fit are self-motivated, collaborative, want to build new technologies, and open to working directly with users of the technologies they build.

Do you have any work experience or collaboration with industry?

No, very little.

Do you have any new courses or publications that you would like ISP students to know about?

My google scholar page is up to date if people want to see what my lab has been up to lately.

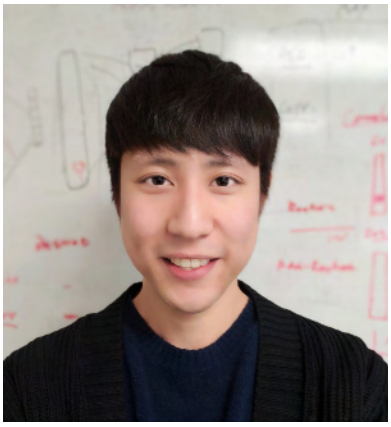
What advice would you give to ISP students who are currently seeking a job?

It depends on their goals. One piece of general advice is to think deeply about your strengths, both in terms of helping you focus in on the kind of job that would best meet your objectives, and in terms of helping you understand how to communicate those strengths to people making hiring decisions.

How do you spend your spare time?

I do a lot of running.

Thank you Dr. Walker for introducing yourself and we appreciate the time you took to speak with us.



Dr. Seong Jae Hwang

Assistant Professor,
Department of Computer Science

Dr. Hwang, it is a pleasure meeting you today; we are excited to have you in the ISP family.

How does being an ISP faculty member fit into your career goals? what things do you look forward to?

I believe any interdisciplinary researcher like myself would find ISP to be an amazing community to be a part of. For instance, meeting colleagues from various domains across the campus has been wonderful over the past months already. Importantly, this includes ISP students whom I find very exciting to talk to and interact with. With new ISP faculty bringing in unique expertise, I am excited to see how ISP grows and expands as more and more diverse research areas coalesce.

What are your current research interests? What would you like other ISP faculty and students to know about your research group?

With specialties in machine learning, computer vision, and deep learning techniques, our group focuses on medical imaging analysis for Alzheimer's disease through close collaborations with experts in aging research. Specifically, we develop algorithms that address domain-specific challenges to effectively model disease progressions, derive sensitive biomarkers, and analyze longitudinal patterns using various neuroimaging modalities. Also, we are always interested in learning about problems from other domains and finding out potential collaborations.

What is your vision for the future of your research field and where do you think it is headed in five years (or beyond)?

Unfortunately, Alzheimer's disease is still incurable. As a computer scientist, I believe AI has great potential to further advance our understanding of the disease. Thankfully, incredibly valuable data are now becoming more accessible than ever before with greater quantity and quality. Therefore, within the next few years, I believe the role of AI will become increasingly more practical and valuable such that finding AI experts in research teams will be a common sight.

What are one or two of your proudest professional accomplishments?

My proudest accomplishments are starting my PhD program followed by joining SCI at Pitt.

How many graduate students are you advising now?

I am advising two PhD students and one masters students (and one undergraduate).

Are you accepting new students at this point? If yes, do you have any messages for the ISP students who wish to join your research group?

I cannot guarantee these things now, but I am always happy to chat with students interested in joining.

Do you have any work experience or collaboration with industry?

I don't have full-time work experience, but I have done 5 internships at 4 different companies including a start-up.

Do you have any new courses or publications that you would like ISP students to know about?

I had a seminar course on medical imaging analysis using AI in Spring 2020, and I will be teaching an undergraduate research course on similar topics in Fall 2020. I am working on several papers right now which will be available on my website soon.

What advice would you give to ISP students who are currently seeking a job?

I do not have specific advice to give because I think ISP students are already (or will be) very well equipped with critical skills and important research interests that many would find appealing nowadays.

How do you spend your spare time?

I typically watch movies or play video games at home during my spare time.

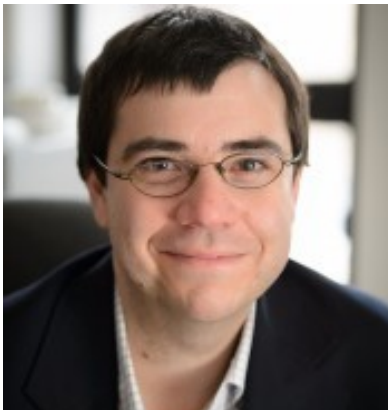
Thank you Dr. Hwang for introducing yourself and we appreciate the time you took to speak with us.

Dr. Richard Boyce

Associate Professor, Department of Biomedical Informatics

Research Interests:

The design and evaluation of intelligent clinical decision support interventions, distributed and collaborative knowledge bases, and advanced pharmacovigilance signal detection management.



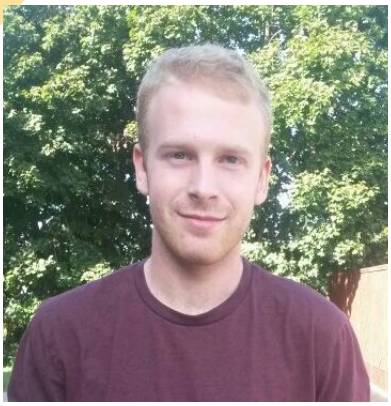
Dr. Michael Colaresi

Professor, Department of Political Science
Director, Institute of Cyber Law, Policy, and Security

Research Interests:

Computational social science, Bayesian inference, applied machine learning, natural language processing, national security secrecy and oversight, international and intrastate violence, and the measurement of human rights protections and violations.

ISP Dissertation Proposals



Bryan Andrews

Doctoral Student,
Intelligent Systems Program

Bryan successfully defended his dissertation proposal in December 2019:

Inducing Sets: A Novel Framework for Reasoning about Graphical Markov Models in the Presence of Latent Variables

Directed acyclic graphs (DAGs) have become widely studied and applied in causal modeling and discovery; their directed structure provides an interpretable representation for causality and their simplicity facilitates systematic learning procedures, however, DAGs are ill-equipped to handle systems with latent variables. Meanwhile, latent confounding and selection bias occur with some regularity in many domains. Maximal ancestral graphs (MAGs) naturally extend DAGs by implicitly taking the possibility of latent variables into account. In fact, MAGs represent the smallest superclass of DAGs closed under marginalization and conditioning. This makes MAGs an obvious choice for causal modeling and discovery in systems with latent confounding and selection bias.

In this work, we introduce inducing sets, a novel framework for reasoning about graphical Markov models in the presence of latent variables. In particular, inducing sets may be used to reason about MAGs. From inducing sets, we derive m-connecting sets as an alternative representation of MAGs and present two main theoretical results for MAGs: a characterization of Markov equivalence and a factorization criterion. We hypothesize that m-connecting sets may be used in conjunction with MAGs in order to create state-of-the-art causal structure learning algorithms. To this end, we conjecture that m-connecting sets allow for the efficient scoring of MAGs and may be used to characterize a set of local search operations for causal network discovery. Using m-connecting sets, we propose two algorithms: a global method for causal network discovery and a local method for query-based causal edge discovery.

Committee Members:
Gregory F. Cooper, Professor, Department of Biomedical Informatics and Intelligent Systems, University of Pittsburgh
Peter Spirtes, Professor, Department of Philosophy, Carnegie Mellon University
Thomas Richardson, Professor, Department of Statistics, University of Washington
Takis Benos, Professor, Department of Computational Biology and Intelligent Systems, University of Pittsburgh
Sofia Triantafillou, Assistant Professor, Department of Biomedical Informatics and Intelligent Systems, University of Pittsburgh

ISP Dissertation Defenses

Jaromir successfully defended his doctoral dissertation in April 2020:

Discovering Sentences for Argumentation about Meaning of Statutory Terms

Jaromir Savelka, *PhD*

Data Scientist,
Reed Smith LLP

In this work I studied, designed, and evaluated computational methods to support interpretation of statutory terms. Understanding statutes is difficult because the abstract rules they express must account for diverse situations, even those not yet encountered. The interpretation involves an investigation of how a particular term has been referred to, explained, interpreted, or applied in the past. Going through the list of results manually is labor intensive. A response to a search query may consist of hundreds or thousands of documents. I investigated the feasibility of developing a system that would respond to a query with a list of sentences that mention the term in a way that is useful for understanding and elaborating its meaning. I treat the discovery of sentences for argumentation about the meaning of statutory terms as a special case of ad hoc document retrieval. The specifics include retrieval of short texts (sentences), specialized document types (legal case texts), and, above all, the unique definition of document relevance.

This work makes a number of contributions to the areas of legal information retrieval and legal text analytics. First, a novel task of discovering sentences for argumentation about the meaning of statutory terms is proposed. This is a task lawyers routinely perform using a combination of manual and computational approaches. Second, a data set comprising 42 queries (26,959 sentences) was assembled to support the experiments presented here. Third, by systematically assessing the performance of a number of traditional information retrieval techniques, I position this novel task in the context of a large body of

work on ad hoc document retrieval. Fourth, I assembled a unique list of 129 descriptive features that model the retrieved sentences, their relationships to the terms of interest, as well as the statutory provisions they come from. I demonstrate how the proposed feature set could be utilized in learning-to-rank settings by showing how a number of machine learning algorithms learn to rank the sentences with very reasonable effectiveness. Fifth, I analyze the effectiveness of fine-tuning pre-trained language models in the context of this special task and demonstrate a very promising direction for future work.

Committee members:
Kevin D. Ashley, Professor, School of law, Learning Research & Development Center, and Intelligent Systems, University of Pittsburgh
Milos Hauskrecht,Professor, Department of Computer Science and Intelligent Systems, University of Pittsburgh
Daqing He, Professor, School of Information Sciences and Intelligent Systems, University of Pittsburgh
Diane Litman, Professor, Department of Computer Science and Intelligent Systems, University of Pittsburgh

Jeya successfully defended his doctoral dissertation in August 2019:

Knowledge Discovery with Bayesian Rule Learning Methods for Actionable Biomedicine

Jeya Balaji Balasubramanian, *PhD*

Postdoctoral Researcher,
National Cancer Institute (NCI)

Discovery of precise biomarkers are crucial for improved clinical diagnostic, prognostic, and therapeutic decision-making. They help improve our understanding of the underlying physiological (and pathophysiological processes) within an individual. To discover precise biomarkers, we must take a personalized medical approach that accounts for an individual's unique clinical, genetic, omic, and environmental information. The molecular-level omic information provides an opportunity to understand complex physiological processes at an unprecedented resolution. The reducing costs and improvements in high-throughput technologies, which collect omic data from an individual, has now made it feasible to include a person's omic information as a standard component to their medical record. This information can only be clinically actionable if it is understandable to a clinician and applicable in the correct medical context. Biomarker discovery from omic data is challenging because they are— 1) high-dimensional, which increases the chance of false positive discoveries from traditional data mining methods; 2) most diseases are multifactorial, where many factors influence the disease outcome, making it challenging to be modeled by most data mining algorithms while keeping it interpretable to a clinician; and 3) traditional data mining methods discover only statistically significant biomarkers but do not account for clinical relevance, therefore they do not translate well in clinical practice.

In this dissertation, I formulate the problem of learning both statistically significant and clinically relevant biomarkers as a knowledge discovery problem. In computer science, knowledge discovery in databases is "a non-trivial process of the extraction of valid, novel, potentially useful, and ultimately understandable patterns in data". Clinical practice guidelines in decision support systems are often presented as explicit propositional logic rules because they are easy for a clinician to understand and are often actionable instructions themselves. Bayesian rule learning (BRL) is a rule-learning classifier that learns patterns as a set of probabilistic classification rules. I develop BRL to efficiently learn from high-dimensional data and obtain a robust set of rules by identifying context-specific independencies in the data. To help model multifactorial diseases, I study various ensemble methods with BRL, collectively called Ensemble Bayesian Rule Learning (EBRL). I also develop a novel ensemble model visualization method called Bayesian Rule Ensemble Visualization tool (BREVity) to make EBRL more human-readable for a researcher or a clinician. I develop BRL with informative priors (BRLp) to enable BRL to incorporate prior domain knowledge into the model learning process, thereby further reducing the chance of discovering false positives. Finally, I develop BRL for knowledge discovery (BRL-KD) that can incorporate a clinical utility function to learn models that are clinically more relevant. Collectively, I use these BRL methods, developed for the task of biomarker discovery, as the knowledge engine of an intelligent clinical decision support system called Bayesian Rules for Actionable Informed Decisions or BRAID, a concept framework that can be deployed in clinical practice.

Committee members:
Vanathi Gopalakrishnan, Associate Professor, Department of Biomedical Informatics and Intelligent Systems, University of Pittsburgh
Gregory F. Cooper, Professor, Department of Biomedical Informatics and Intelligent Systems, University of Pittsburgh
Shyam Visweswaran, Associate Professor, Department of Biomedical Informatics and Intelligent Systems, University of Pittsburgh
Steven E. Reis, Professor, Associate Vice Chancellor Clinical Research, University of Pittsburgh

Jonathan successfully defended his doctoral dissertation in February 2020:

Deep Learning for Causal Structure Learning Applied to Cancer Pathway Discovery

Jonathan Young, *MD, PhD*

Intelligent Systems Program

In general, the cellular mechanisms leading to cancer in an individual are heterogeneous, nuanced, and not well understood. It is well appreciated that cancer is a disease of aberrant signaling, and the state of a cancer cell can be described in terms of abnormally functioning cellular signaling pathways. Identifying all of the abnormal cellular signaling pathways causing a patient's cancer would enable more patient specific and effective treatments – including targeting multiple abnormal pathways during a treatment regime. Here we interpret the cellular signaling system as a causal graphical model and apply a modified deep neural network (DNN) to learn latent causal structure that represents the cancer cellular signaling system.

Most causal discovery algorithms have been developed to find causal structure and parameterizations of causal structure relative to the observed variables of a dataset. A smaller number of casual discovery algorithms also find latent causal structure, but these methods are often highly constrained or less applicable to the problem explored here, suggesting that new methods are needed. In this dissertation, we address a problem for which it is known that a set of variables X causes another set of variables Y (e.g., mutations in DNA cause changes in gene expression), and these causal relationships are encoded by a causal network among a set of an unknown number of latent variables. We develop a modified deep learning model, referred to as redundant input neural network (RINN), with an L1 regularized objective function to find causal relationships between input (X), hidden, and output (Y) variables. More specifically, our model allows input variables to directly interact with all latent variables in a neural network to influence what information latent variables encode in order to generate the output variables accurately. In a series of simulation experiments, we show that the RINN model successfully recovers latent causal structure from various simulated datasets with different levels of noise better than other models.

We hypothesize that training a RINN on multiple omics data will enable us to map the functional impacts of genomic alterations to latent variables in a deep learning model, allowing us to discover the hierarchical causal relationships between variables perturbed by different genomic alterations. We apply the RINN to cancer genomic data, where it is known that genomic alterations cause changes in gene expression. We show that differentially expressed genes can be predicted from somatic genome alterations with reasonable AUROCs by a RINN (or DNN). We also show that a RINN is able to discover many real cancer signaling pathway relationships, especially relationships between genes in the PI3K, Nrf2, and TGFb pathways, including some causal relationships. In this setting, the connections between input and latent variables make the latent variables partially interpretable, as they can be easily mapped to input space. However, despite relatively large levels of regularization, the returned causal graphs were still somewhat too dense to be easily and directly interpretable as causal graphs. Future versions of the RINN, with differential regularization, autoencoder pre−trained representations, and optimization with parallelized and constrained evolutionary algorithms, will have a high probability of capturing more easily interpretable cancer pathways.

Committee members:
Xinghua Lu, Professor, Department of Biomedical Informatics and Intelligent Systems, University of Pittsburgh
Gregory F. Cooper, Professor, Department of Biomedical Informatics and Intelligent Systems, University of Pittsburgh
Vanathi Gopalakrishnan, Associate Professor, Department of Biomedical Informatics and Intelligent Systems, University of Pittsburgh
Harry Hochheiser, Associate Professor, Department of Biomedical Informatics and Intelligent Systems, University of Pittsburgh
Songjian Lu , Associate Professor, Department of Biomedical Informatics and Intelligent Systems, University of Pittsburgh

Sanya successfully defended her Master's dissertation in April 2020:

Bayesian Networks for Diagnosing Childhood Malaria in Malawi

Sanya Taneja,

Doctoral Student,
Intelligent Systems Program

Infectious diseases such as malaria are responsible for the majority of under-five deaths in low- and middle-income countries. Accurate diagnosis and management of illnesses can help in reducing the global burden of childhood morbidity and mortality. While trained healthcare workers deliver treatment for common childhood illnesses in healthcare facilities in Malawi, there is a significant lack of diagnostic support in rural health centers. With recent trends in artificial intelligence in global health, we hypothesize that a data-driven approach to diagnosis of childhood illnesses may address the challenges faced in health centers in low-resource countries such as Malawi. In this study, we aim to utilize Bayesian networks to diagnose cases of childhood malaria in Malawi. We develop two Bayesian diagnostic models for classification of malaria using clinical signs and symptoms. The first model is created manually, while the other combines an Augmented Naïve Bayes approach with expert knowledge. The models are learned using a national survey dataset which contains sick child observations including patient information, diagnosis, and symptoms. The target malaria diagnosis is taken as the result of the malaria rapid diagnostic test (mRDT). The performance of the Bayesian models is further compared to traditional machine learning classifiers on the basis of accuracy, AUC, precision, F1 score, sensitivity, and specificity. We also present an experimental framework that can be used to model the malaria diagnostic support in the rural health centers. The manually created Bayesian model achieves an accuracy of 63.6% with an AUC of 0.58. The augmented naïve Bayes model considers associations between the variables and achieves an accuracy of 62.7%. The Bayesian models outperform the logistic regression and random forest models in the classification of the disease. Bayesian models provide a powerful, efficient and data-driven tool for diagnosis of childhood illness that can lead to a more evidence-based clinical practice in Malawi. The simplicity and interpretability of Bayesian models offer a unique approach to diagnostic support in low-resource countries. As Bayesian models are representative of the population from which the data has been derived, this approach can be generalized to other childhood illnesses in different regions of the world.

Committee members:
Shyam Visweswaran, Associate Professor, Department of Biomedical Informatics and Intelligent Systems, University of Pittsburgh
Gregory F. Cooper, Professor, Department of Biomedical Informatics and Intelligent Systems, University of Pittsburgh
Gerald P. Douglas, Assistant Professor, Department of Biomedical Informatics, University of Pittsburgh

New ISP PhD Graduates

Jeya Balaji Balasubramanian, *PhD*

Postdoctoral Researcher,
National Cancer Institute (NCI)

Jeya Balaji Balasubramanian joined ISP in 2013 and graduated in 2019. Jeya's advisor during his PhD was Dr. Vanathi Gopalakrishnan from the Department of Biomedical Informatics.

Would you briefly describe your dissertation topic, its importance, and your experience with your defense?

I developed major extensions to a machine learning algorithm that automatically learns probabilistic rules from data. The rule induction algorithm is called Bayesian Rule Learning (BRL). Rule models tend to be easy to read and understand because they take the form of simple propositional logic sentences as follows— "if the condition is true, then the consequent must also be true within a certain degree of confidence". BRL searches over a space of Bayesian network structures learned from data and infers rules from them. Using the Bayesian framework has certain advantages including— 1) having a mechanism to account for uncertainty (from noise or lack of data), and 2) being able to incorporate prior knowledge about the domain. Using a method to handle uncertainty in data called Bayesian Model Combination (BMC), I was able to build predictive models with BRL that beat state-of-the-art methods traditionally used in machine learning. Using BMC makes it hard to read rules. For this reason, I developed a visualization method to be able to read rule structures from BMC models. Using the Bayesian framework, I developed the first rule learning method, to my knowledge, that could incorporate prior domain knowledge (causal relationships) into the modeling process. These functionalities allowed BRL to be able to find rules subjectively interesting to a specific user.

The dissertation defense went smoothly. I think by this milestone, ISP students would have the confidence to be able to defend their work. All of ISP's milestones are research focused, which prepared me well for the defense.

What were the obstacles you encountered in conducting your research? How did you overcome those problems?

Earlier in my PhD, I struggled with my scientific writing. Dr. Greg Cooper recommended a book called "Four Steps to Funding" by Morgan Giddings. It was very informative. While the book

was about writing grants, it benefited me in both reading and writing technical papers. The book breaks down each decision in a research project (and sub-projects) into simply answering a sequence of 4 questions— 1) why (problem and motivation) 2) who (related work and background) 3) what (proposed work), and 4) how (how does the proposed work solve the problem?). It helped me pinpoint the gaps in my own understanding of any research question.

What was your most proud moment?

Each time I had a paper published; it was a proud moment. It marked an end of a research cycle upon which I could build upon.

How was your experience as an ISP student? What did you learn from this experience?

I had a wonderful time at ISP. I especially miss my friends there. The experience taught me a lot about how to become an independent researcher. Being able to find and prioritize interesting problems. Systematically approaching the problem. Also, a great deal about showing integrity in research.

Do you have any suggestions or comments to improve the ISP program?

I think the engineering aspect of machine learning is just as important as the scientific aspect. Engineering enables science. The scientific aspect covers the necessary mathematics, statistical theory, algorithms, and domain-specific knowledge surrounding artificial intelligence. The engineering aspect covers the ability to actually build and deploy the machine learning model. This requires you to be aware of version control, various cloud resources, GPU programming, and web technologies. Being able to efficiently build and deploy machine learning methods greatly improves your understanding of the method and also makes your work more accessible to the scientific community. Only then can you receive feedback from the scientific community to improve your work and elevate it to cutting-edge research. My suggestion would be to include more engineering courses in ISP.

What have you been doing since graduation?

I am working as a postdoctoral researcher at the National Institutes of Health. I am working on developing artificial intelligence theory, its application to oncology, and more recently its application to the COVID-19 pandemic.

Do you have any word of advice for new ISP students?

I benefitted immensely from speaking with the more senior students of ISP. ISP already sets up a student mentor for the first-year students. I recommend making good use of this resource and speaking to more students advanced into their PhD milestones. They offer great guidance on the various coursework and milestones. They also push you to finish your milestones.

Also, beyond the coursework, take interest in the engineering aspect of machine learning. It is not enough to be good in theory. You must also learn how to efficiently implement and deploy these methods for others to use it conveniently.

Thank you Dr. Balasubramanian for the time you took to speak with us.

Jaromir Savelka, *PhD*

Data Scientist,
Reed Smith LLP

Jaromir Savelka joined ISP in 2013 and graduated in 2020. Jaromir's advisor during his PhD was Dr. Kevin Ashley from the Learning Research and Development Center.

Would you briefly describe your dissertation topic, its importance, and your experience with your defense?

My work focuses on interpretation of statutory law. The goal is to retrieve short pieces of text, such as sentences, that are useful for argumentation about the meaning of concepts mentioned in statutes. The work is important because understanding statutes is difficult and the investigation of how a particular concept has been used in past is labor intensive. The work presents methods to fully automate the task which lawyers routinely perform using a combination of manual and computational approaches. I had a great experience with my defense. Over 30 people attended and there were a lot of interesting questions. Everything went well and I enjoyed it a lot.

What were the obstacles you encountered in conducting your research? How did you overcome those problems?

From my perspective, the most difficult part was the scale of the whole effort starting from taking the courses and finishing with the dissertation defense. It is a lot of work to just get to the PhD candidacy and it is probably even more work to conduct the dissertation work. Looking back, I do not really think there was one single step which clearly stood above the others in terms of its difficulty.

What was your most proud moment?

It was definitely the dissertation defense. It was the moment when everything came together and I realized that all the effort of the last 6 years finally came to a successful conclusion. It felt great.

How was your experience as an ISP student? What did you learn from this experience?

I really enjoyed the experience. I learned a ton taking the courses as well as passing the milestones. I had a great relationship with my advisor (Kevin Ashley) and with his help, I was able to publish multiple papers at international conferences. I also feel that the whole group of people around ISP is very friendly and supportive. I have met many new friends with whom I am sure I will stay in touch even after I leave the program.

Do you have any suggestions or comments to improve the ISP program?

I cannot think of anything. The program is really well designed.

What have you been doing since graduation?

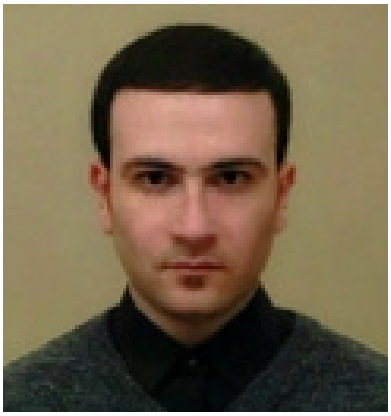
I work as a data scientist for Reed Smith LLP here in Pittsburgh.

Do you have any word of advice for new ISP students?

My advice would be to keep your pace and pass the milestones as fast as possible. The biggest mistake I have done was slowing down after defending my proposal. It was very difficult to get back to work on the dissertation and I feel like it cost me at least a year (perhaps more).

Thank you Dr. Savelka for speaking with us today.

ISP Provost Fellowships



A Generative Approach Towards Unsupervised Word-Level Classification

Omid Kashefi

Doctoral Student,
Intelligent Systems Program

Suppose you ask Amazon's Alexa for a quick summary of critic reviews for a movie you plan to watch; Alexa will go through hundreds of comments from different critics, pinpoint the word(s) that are strong indicators of their opinion, and then aggregate the overall consensus of the critics' opinion. In this example, a core problem is identifying certain words and understand their roles (e.g., positive or negative opinions). This is a difficult problem because the roles of the words depend on the context in which they are used. A straightforward approach, such as a lexicon lookup, may not correctly identify the intended usage. While a machine learning method may better take the context into account, it relies on the availability of the training corpus containing class labels for each word.

The challenge is not limited to the above example; it is a common problem for many NLP applications, especially for newer applications that do not have established resources. Some examples include: to fix a grammatical error, a grammatical error corrector needs to first pinpoint the word(s) that causes that grammatical error in a poorly written sentence and then suggest a replacement; to ensure the conciseness of an essay, a writing aid system needs to locate the semantically redundant word(s) and delete them; in order to automatically detect and hide comments with abusive language, Facebook needs to check for inappropriate words in that comment; or to determine if the best audience of an article are experts or general people, a recommendation system needs to detect the words that are expressing a specific or general level of information.

To address this problem, I propose an alternative training framework that is informed by class predictions of greater contexts, which may be easier to obtain. That is, suppose we know that the sentence (or some other larger text span) containing the word belongs to some Class A; if we determined that this word is a significant contributor to that prediction, we might also infer a label for the word itself. For example, suppose we knew a social media comment was deemed inappropriate (perhaps because someone reported it), if we find out which word(s) in the comment are contributed the most to the user's decision to report the comment, we already found the inappropriate word(s) and confirms whether the sentence is correctly reported or not.

Rather than directly estimating the relative contributions of each word, we first cast the problem as sentence rewriting exercise: rewrite the original sentence so that it becomes more likely to belong to Class B (e.g., appropriate) than Class A (e.g., inappropriate); then, those words that were changed in the rewriting are likely to be the heavy contributor to Class A so are inappropriate themselves. Modeled after style transfer and text generation methods, we propose a transfer function for class transference. To train an appropriate transfer function, however, we need a corpus of training examples for the sentence-level classification. I also explore some textual data augmentation options and investigates how they may facilitate the training of the transfer function on low-resource problem domains.

Concept Drift and Model Adaptation in Clinical Predictive Models

Amin Tajgardoon

Doctoral Student,
Intelligent Systems Program

Concept drift is a critical issue in the clinical domain, as it may compromise the performance of deployed predictive models and influence clinical decision-making. Clinical data is collected at an unprecedented pace in electronic health record (EHR) systems, particularly in intensive care units where large amounts of data streams flow into the EHR system every day. Concept drift in the EHR data can occur for several reasons such as shifts in the patients' population, evolving clinical practice, or changes in the coding systems, measurement methods, EHR interface, or data entry workflows. A drift in the EHR data may deteriorate the performance of predictive models, affect the clinicians' decision-making process, and impose threats on the patients' safety. Therefore, it is crucial to detect concept drift in the EHR data and update the models accordingly.

Several methods for detecting concept drift has been proposed in the machine learning literature. Most of these methods are based on the assumption that a significant drop in the performance of the model signals an underlying drift in the data distribution. However, the usability of concept drift detection methods has not been thoroughly investigated in the context of clinical predictive models. In addition, when concept drift occurs, deployed predictive models should be adapted to the changes in the data distribution; this process is referred to as model adaptation. State-of-the-art model families such as deep neural networks are yet to be comprehensively studied in the context of model adaptation in dynamic clinical environments.

The objective of this project is in identifying the underlying mechanisms of concept drift in dynamic clinical environments, and developing and evaluating new methods to automatically detect concept drift and accordingly, update deployed predictive models. Adapting clinical predictive models to concept drift can prevent model deterioration and ultimately will improve the robustness of clinical decision-support systems.

ISP @ Conferences

East Coast Industry Forum (ECIF)

Venue: Carnegie Mellon University , Date: October 18, 2019

From left to right:
Mahbaneh Eshaghzadeh - Student
Kayhan Batmanghelich - Professor
Peter Brusilovsky - Professor
Diane Litman - Professor
Vanathi Gopalakrishnan - Professor
Sanya Taneja - Student

ISP Forums

From left to right: Dr. Kevin Ashley, Dr. Panos Chrysanthis, Dr. Ganesh Mani, and Dr. Vanathi Gopalakrishnan

We sincerely thank all the faculty and students who participated in the ISP AI Forum seminar series this past year. We particularly thank Dr. Ganesh Mani for his insightful lecture on
"Intelligent Systems: Past, Present and Future"
This was much appreciated especially by first-year students who got a birds-eye view into the field. Our past events can be accessed here:
http://isp.pitt.edu/events

Contact Us

We encourage all the students, faculty, alumni, and whoever is interested to find the latest news about ISP to follow us through our social media.

The newsletter is available at: http://www.isp.pitt.edu/newsletter

ISP Director: Dr. Vanathi Gopalakrishnan

Editor: Fattaneh Jabbari