# Table of Contents

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Director's Address</td>
<td>3</td>
</tr>
<tr>
<td>ISP 30th Anniversary</td>
<td>4</td>
</tr>
<tr>
<td>Alumni Highlights</td>
<td>5</td>
</tr>
<tr>
<td>Faculty Highlights</td>
<td>8</td>
</tr>
<tr>
<td>Faculty Grants</td>
<td>10</td>
</tr>
<tr>
<td>Student Highlights</td>
<td>12</td>
</tr>
<tr>
<td>ISP @ Conferences</td>
<td>14</td>
</tr>
<tr>
<td>New ISP Faculty Appointments</td>
<td>16</td>
</tr>
<tr>
<td>Provost Fellowships in Intelligent Systems</td>
<td>21</td>
</tr>
<tr>
<td>Thesis Proposals</td>
<td>24</td>
</tr>
<tr>
<td>Dissertation Defenses</td>
<td>29</td>
</tr>
<tr>
<td>New PhD Graduates</td>
<td>34</td>
</tr>
<tr>
<td>ISP Events</td>
<td>38</td>
</tr>
<tr>
<td>Contact Us</td>
<td>39</td>
</tr>
</tbody>
</table>
2017-2018 was a year to celebrate ISP’s 30 year past as well as to continue to chart its future as part of Pitt’s new School of Computing and Information (SCI).

On March 15-16, 2018, the ISP 30th Anniversary Celebration highlighted 30 years of excellence in Artificial Intelligence research and education at the University of Pittsburgh. Over 100 individuals participated in the ISP30 events, which included an alumni welcome reception, a full day academic symposium with two keynotes and four technical sessions, and a concluding program banquet with the Provost (followed by an unofficial excursion by many of us to one of Pittsburgh’s craft breweries). ISP30 was the first alumni event for both ISP and SCI, and a great time was had by all. The current faculty and students were delighted to have so many alumni participate, including the founding and almost all subsequent ISP Directors, some of the first students from the 1987 entering class, and many other alumni from later phases of the program. Since I myself didn’t join Pitt until 2001, I was personally excited to meet so many former students for the first time and to hear stories from the early days of ISP. The ISP30 webpage (http://www.isp.pitt.edu/ispanniversary/) continues to be maintained, with links to the final symposium program, lots of photos, and both short and long versions of an anniversary program video. Video excerpts (where we have speaker permissions) from the excellent and engaging symposium technical sessions are now linked as well.

More generally, thanks to the efforts of our webmaster Khushboo Thaker, there are now many ways to keep connected with ISP. Please check them out!

Dr. Diane Litman
Professor, Department of Computer Science, School of Computing and Information
Senior Scientist, Learning Research and Development Center

Browse recent publications authored by ISP faculty and students
http://www.isp.pitt.edu/resources/publications2017
http://isp.pitt.edu/resources/publications2018 (currently through mid year)

Friend ISP’s official Facebook account ISP@UPITT
https://www.facebook.com/groups/492930340863765/

Like ISP’s public Facebook page

Follow ISP’s Twitter account
https://twitter.com/ISPPitt

Grow your LinkedIn network

Browse ISP news archives
http://isp.pitt.edu/news

Finally, although my 2016-2018 term as ISP Director has come to an end, I am delighted to be moving to a Co-Directorship role with Vanathi Gopalakrishnan (Biomedical Informatics) for 2018-2020. I am also excited to welcome 6 new graduate students to the program, as well as 2 faculty who although not new to Pitt are new to ISP. New graduate student profiles are currently being featured on Facebook, while interviews with our newly appointed faculty are later in this newsletter. This newsletter also includes our student editor’s report on ISP30, dissertation abstracts, interviews with graduating students, alumni updates (thanks to all who responded!), and more. Happy reading!

Diane Litman
ISP Director, 9/1/16-8/31/18
ISP Co-Director, 9/1/18-8/31/20
On March 16, 2018, the Intelligent Systems Program (ISP)’s alumni, faculty and students joined together in celebration of the ISP 30th anniversary at the University Club, University of Pittsburgh. That day was special for all the people who have developed the ISP from scratch. They were enthusiastic about what they have brought up since the beginning of the program in 1986.

At this event, the two founders of ISP, Dr. Alan Lesgold (now Emeritus Faculty) and Dr. Rich Thomason (now at Michigan), shared the goal that made them create this program and the difficulties they have faced to keep this program successful. In the mid-80s, Artificial Intelligence (AI) began to find its way across various disciplines such as medicine and law and students weren’t satisfied to only work on a single problem. They wanted to expand their research while using the help of experts in their field as well as AI. So, the group of faculty from many different fields who were all interested in using Intelligent applications to solve their problems gathered together and started the program.

For the last 30 years, ISP has provided a variety of opportunities to expose faculty and students to a range of fields with the focus on AI and has remained committed to this vision. Our students and faculty conduct critical research in areas such as educational technology, machine learning, biomedical informatics, natural language processing and law. In addition, ISP’s unique approaches have inspired students to do novel research and find what they are passionate about as they begin their careers.

At the ISP 30th celebration, nearly 100 guests including students, alumni, past and present faculty were welcomed by Dr. Diane Litman, the Director of the Intelligent Systems Program along with Dr. Paul R. Cohen, the Dean of the School of Computing and Information.

At ISP@30 two keynote speakers, Dr. Martha Pollack and Mr. Bryan Salesky addressed the celebration’s guests. Dr. Pollack was a member of the ISP faculty from 1991 to 2000, and is currently the 14th president of Cornell University. She presented a fascinating talk titled “AI: The Good, The Bad and the Ugly” which went into how AI affects our life both positively and negatively. Also, there was an unparalleled opportunity to hear from one of the most successful CEOs in the United States, Mr. Salesky from Argo AI and a 2002 Pitt computer engineering graduate. He focused his talk on the impact of self-driving vehicles on the world from his professional perspective.

Along with these spectacular talks we had four sessions of discussions titled “The Past, Present and Future”, “Uses of AI to Enhance Education”, “Intelligent Systems in Medicine: From Early Dreams to Routine Systems that Advance Human Health and Transform Healthcare” and “Machine Learning: From Research to Practice”. Each session included a panel of ISP faculty as well as ISP alumni who have become experts in their field. The panelists contributed by sharing the story of their experiences while in ISP, or described the novel projects they have done or their perspective about the future of their fields.

The event ended with the poster presentation from the current ISP students which allowed them to talk about their research and to hear feedback from their colleagues in attendance.

ISP is indebted to all the people who contributed to making this event interesting. This event reunited many old friends and colleagues in a friendly and professional atmosphere. For 30 years ISP has been educating students to become leaders in their field and we are looking forward to another 30 years of success!

Saba Dadsetan
ISP PhD Student, 2017-present
Alumni Highlights
Alumni Highlights

Dr. Giuseppe Carennini (2000) has been promoted to Full Professor at the University of British Columbia.

Dr. Rosta Farzan (2009) has been promoted to the rank of Associate Professor with Tenure at the School of Computing and Information, University of Pittsburgh.

Dr. Bruce McLaren (1999) is the president of the International Society of Artificial Intelligence in Education for the years 2017-2019. Also, he has been co-author of 3 best paper awards over the past two years, one with alumna Rosta Farzan (also a Pitt professor) and the other two with alumnus Vincent Aleven (also a CMU professor).

Dr. Noboru Matsuda (2004) has joined the Department of Computer Science at North Carolina State University as an Associate Professor over the summer of 2018. He is also an affiliate at the Center for Educational Informatics lead by Dr. James Lester. Dr. Matsuda will continue leading research on the application of a teachable agent (SimStudent) and evidence-based learning engineering methods (PASTEL). He recently received a new IES grant (Education Technology, Sep 1, 2018 - Aug 30, 2021) for developing an intelligent synthetic tutee that asks questions for students while they are learning by teaching to understand how scaffolding their reflective thinking facilitates the effect of tutor learning.

Dr. Shaghayegh Sahebi (2017) has been awarded funds by the National Science Foundation for her CRII proposal "Modeling Student Knowledge and Improving Performance when Learning from Multiple Types of Materials." Dr. Sahebi and her husband had their wedding ceremony in Alanya, Turkey on August 5th, 2018.
Alumni Highlights

Dr. Min Chi (2009) along with Dr. Diane Litman (ISP Director) elected as Executive Committee Members of the International Society of Artificial Intelligence in Education

Dr. Violetta Cavalli Sforza (1998) will be the Local Arrangements Co-Chair and Dr. Bruce McLaren will be the General Co-Chair for AIED 2020 (International Conference on Artificial Intelligence in Education) in Ifrane, Morocco.

The summer 2018 issue of Pitt Magazine featured an article regarding Dr. Steve Casner’s (1990) recently published book titled "Careful: A User’s Guide to Our Injury-Prone Minds". This story can be found here.
Faculty Highlights
Faculty Highlights

- Dr. Yu-Ru Lin talked about the Las Vegas tragedy and her research in the Las Vegas Sun's article titled "Americans' 'unique connection' to Las Vegas could impact grieving.
  - Dr. Lin received a Service Award at the SBP-BRiMS in 2017
  - Dr. Lin invited to be a keynote speaker at the International Workshop on the Social Web for Environmental and Ecological Monitoring (SWEEM 2018), Stanford University

- Dr. Lin served as a program committee co-chair for two conferences: Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS 2017) and The International AAAI Conference on Web and Social Media (ICWSM 2019)
  - Dr. Lin will also be serving as an academic editor for PLOS ONE journal

- Dr. Diane Litman became an Association of Computational Linguistics Fellow.
  - Dr. Litman also served as the Program Co-Chair for the Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL)

- Dr. Michael Lewis's paper "Transparency and Explanation in Deep Reinforcement Learning Neural Networks" won the Best Paper Award at the Conference on Artificial Intelligence, Ethics and Society 2018.

- Dr. Adriana Kovashka received a Google Faculty Research Award

- Dr. Rebecca Hwa served as the Program Co-Chair for the Conference on Empirical Methods in Natural Language Processing (EMNLP)
Faculty Grants
**Faculty Grants**

- **Dr. Yu-Ru Lin**, along with a multidisciplinary team of Pitt investigators, has received a three-year NSF grant to build and evaluate a marketplace and a mobile application for multi-modal transportation.

- **Dr. Peter Brusilovsky** has been awarded a NSF grant under the program "Cyberlearn And Future Learn Tech" for a project titled **Collaborative Research: CSEdPad: Investigating and Scaffolding Students' Mental Models during Computer Programming Tasks to Improve Learning, Engagement, and Retention**

- **Dr. Kayhan Batmanghelich** was awarded a R01 grant to develop an approach to integrate Radiomic data with Genetic for characterization of Chronic Obstructive Pulmonary Disease (COPD)

- **Dr. Diane Litman** was awarded an Institute of Education Sciences (IES) grant for "Enhancing Undergraduate STEM Education by Integrating Mobile Learning Technologies with Natural Language Processing"

- **Dr. Litman** and Dr. Amanda Godley from the School of Education have been awarded a NSF grant under the program "Cyberlearn And Future Learn Tech" for a project titled **EAGER: Discussion Tracker: Development of Human Language Technologies Collaborative Argumentation in High School English Classrooms**

- Drs. **Yu-Ru Lin** and **Rebecca Hwa** were awarded a DARPA/US DoD grant under the program "Understanding Group Biases" for the project "TRIBAL: A Tripartite Model for Group Bias Analytics". DARPA Understanding Group Biases (UGB) Disruptioneering Program, 2018–2019
Student Highlights
This year the ISP family welcomed two new tiny members - both named Daniel. Congratulations to both proud fathers, Javad Rahimikollu and Jaromir Savelka.

ISP student Gaurav Trivedi won the 3-minute thesis competition hosted as a part of SCI week.
http://sci.pitt.edu/sci-week/3mt/

ISP student Mohammadamin Tajgardoon won the Best Student Paper Award for the paper titled "Patient-Specific Explanations from Risk Prediction Models" in Clinical Research Informatics at AMIA Joint Summits.

Each year, ISP students intern at great companies. Here are some of this year’s internships.

Omid Kashefi at Expedia
Mohammadamin Tajgardoon at Expedia
Fattaneh Jabbari at Honda Research
Zahra Rahimi at Adobe
ISP
@Conferences
ISP @ Conferences

Each year ISP faculty, alumni and students attend many research conferences to present their publications and reunite with friends and colleagues.

Educational Data Mining Conference (EDM), 2018.
From left to right:
Ilya Goldin - Alum
Shaghayegh Sahebi - Alum
Peter Brusilovsky - Professor
Collin Lynch - Alum
Mengdi Wang - Student
Khushboo Thaker - Student

Association for the Advancement of Artificial Intelligence (AAAI), 2018.
From left to right:
Mahdi Pakdaman - Alum
Huma Hashemi - Alum
Diane Litman - Professor
(and Jesse Thomason, Huy Nguyen, CS Alum)
New ISP Faculty Appointments

ISP is excited to welcome two new faculty members who bring with them exciting new areas of research.
It is a pleasure meeting you today; we are excited to have you in the ISP family.

Are you excited about joining ISP? What are the things you look forward to with the ISP community?

I am very excited to join ISP as I think machine learning is becoming more and more critical in every discipline. It is good to be part of a multidisciplinary program where we are able to exchange ideas across fields. Often you have a problem in your field which is completely analogous to a solved problem in another field and I think this kind of idea exchange will be very important in the future.

Before joining the faculty you gave a presentation in one of the ISP seminars. How do you think it went?

I had a good time and the students were very engaged. I got many questions and a few of the students stayed around afterwards. Naturally, the students who were already working on biological problems got a little bit more interested, but I spoke to some students from different backgrounds and there were some nice connections to methods in their area.

You've brought a new field of research to ISP. What would you like others to know about your research group?

We see ourselves as data facilitators because a lot of biomedical research now happens through genome-scale data collection but translating these data into scientific knowledge is quite difficult. So, in general, there is a lot of computational problems that arise, and I see us as having the role of translating the raw data into representations that biologists and clinicians can interpret.

Tell us about your vision of the field: where do you think the field is headed in the next five years?

My work spans several different fields but the general field of genomics is very much driven by technological innovation. I personally feel the biological questions have always been the same; we want to figure out how some complex disease works and how we can improve health or understanding the interaction between genetics and environment. New technologies allow us to look at different aspects of biology and also pose computational problems. The exciting things right now are 3D genome organization and various single-cell assays. Both of these generate massive amounts of data but also provide some novel insights that weren’t possible before.

Another way in which technology drives the field forward is simply by making data collection cheaper. As a result we are getting bigger and more complex datasets that require increasing amounts of computational expertise. Currently I am part of a big project that will try to understand how exercise works on the molecular level. There is about one thousand people participated and we are going to get a variety of genome scale measurements while they are going through exercises training. In this case, even if we measure absolutely everything that may be important we just have piles of data but what we ultimately want is an actual model of how exercise contributes to health on the molecular level. For such complex projects we need new ways of integrating across data-types as well as incorporating prior knowledge.

What are one or two of your proudest professional accomplishments?

I am proud of the constrained matrix decomposition work that I talked about in the ISP seminar because despite the fact that the method is not published yet many people are already using it. I have gotten a very positive response from the community because our method is really easy to use and works out of the box on many different datasets. I have a colleague that has already applied our method to the analysis of huge data compendia with very nice results. My other proudest accomplishment would have to be my graduate school work where I have made some predictions and I went into the lab, and I verified all of them. They had all worked and in one case the predictions we made had contradicted a previous study but it still turned out to be right.

How many graduate students are you advising now?
I have one full-time student now and will be a co-advisor for a couple of students. **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 am looking to take up to three students depending on who else we hire. I do not want to say that you have to have a particular skill set because I feel that people come to graduate school to learn. You can learn everything you need to do genomics and you do not need to be a math genius. Many of the most important tools are technically relatively simple but are built on some conceptual insight.

For me the most important thing is to be genuinely interested in figuring out how things work. It is not enough to say I have one benchmark and I optimized it so I am satisfied with the results. You need to have a good understanding of why the method works. It is also good for one’s career longterm because when you understand why methods work at an intuitive level you can explain it to anyone. Overall, the thing I look for in students is being interested in data problems and ideally also biology.

**Do you have any experience working in or collaborating with industry?**

I do some consulting for the pharmaceutical industry. I do a lot of work with biomarker studies mostly working with animal models which surprisingly can be as heterogeneous as people. For example if you have a group of rats, and subject them to a protocol to model a human disease so that one can test new drugs, some rats do not respond to the protocol and some will not respond to the drug. It would be really helpful to design a test that can predict which animals are which ahead of time to reduce cost. So even if you can do a little bit better than chance, companies are very interested in that. Recently we have been doing more of this kind of biomarker predictive studies for human clinical trials.

**What are your hobbies?**

Well, I have three kids, so I do not have much time. We like to go camping with our camper-van. I work on things other than biology sometimes in my spare time so to me that is a hobby though other people may consider it work. Recently I have worked on voting data and specifically gerrymandering which is an issue that has received a lot of news coverage. Pennsylvania has a new map for the coming election and I am happy to have been involved in that effort.

**Thank you Dr. Chikina for introducing yourself and we appreciate the time you took to speak with us.**
New ISP Faculty Appointments

Dr. Jeffrey Cohn

*Professor, Department of Psychology & Psychiatry,*

*Adjunct Professor, Robotics Institute of CMU*

It is a pleasure meeting you today, we are excited to have you in the ISP family.

**What encouraged you the most to join ISP?**

**What are the things you look forward to with the ISP community?**

During my years in affective computing, most of my Pittsburgh collaborators have been at the Robotics Institute at CMU. When I learned of the new School of Computing and Information (SCI)'s interests in affective computing, I sought out Dean Cohen. From that, I wanted to support new initiatives in affective computing/computational behavioral science. And, of course, I had known Diane and her work for some time and sought a way that we might collaborate. We recently wrote an NSF proposal together with CMU colleagues. ISP seemed a perfect fit.

I would like to have an opportunity to work with students at Pitt who have interests in computer vision and signal processing with applications in emotion, and computational behavioral science. My work is increasingly multimodal with applications in smart and intelligent health, psychopathology, and neuroscience. Two of my current NIH grants enjoin neuroscience, surgery, psychiatry, and psychology. A major goal is to develop a closed-loop system for adaptive DBS (Deep Brain Stimulation). Micro-electrodes are implanted in sub cortex for treatment of intractable obsessive compulsive disorder and depression. With bioengineers from Brown University, we will train classifiers to titrate stimulation from real-time inputs of the microelectrode arrays, EEG from surface of the brain, autonomic physiology, and affective behavior and social interaction. We will learn much about brain circuits and behavior and contribute to patient welfare. In other work we are developing an automated system of annotating verbal and nonverbal affect between parents and their adolescent children. Many of them have current or previous depression disorder. Other projects are biomedical and developmental. All of these projects are multimodal and seek breakthroughs in computational behavioral science that will have high impact. ISP, as a multidisciplinary program in the new school of information and computer science, is just the place to be.

**Before joining the faculty you gave a presentation in one of the ISP seminars. How do you think it went?**

ISP students and faculty showed high level of interest, asked great questions, and were welcoming.

**You've brought a new field of research to ISP. What would you like others to know about your research group?**

They are multidisciplinary and inter-institutional. We have expertise in computer vision, machine learning, and voice quality and timing. We are motivated by basic and applied problems in affective computing, computational behavioral science, psychotherapy, and neuroscience-based treatments for psychological disorders. Because of its importance to emotion and its disorders, we have added expertise in natural language processing and understanding. Our collaborators include behavioral, computer, and biomedical members from CMU, Baylor College of Medicine, Children's Hospital of Seattle, Brown University, and University of Miami. Students in my lab gain experience and opportunity to contribute to novel, high impact research in multi-institution interdisciplinary teams. They are well prepared for both academic research and industry.

**You've had a lot of impressive achievements. What are one or two of your proudest professional accomplishments?**

I guess the role I have been able to play in developing automated facial expression analysis. I started work in this before much existed. One of my first databases may be the most highly cited database in the field. Over time we have released over six more high-impact databases. I've organized major conferences in the field and trained talented, creative students. These and the contributions we have made to computer vision, machine learning, and affective computing have been highly rewarding. I look forward to next frontiers.
**What courses have you created or proposed from the start of your career? Do you have any new courses or publications that you would like us to know about?**

I would like to teach or co-teach courses in affective computing.

**How many graduate students are you advising now?**

A Pitt student in clinical psychology just graduated and has begun a postdoc at Language Technology Institute at CMU, from which he continues to collaborate. I have a continuing CMU student and another begins this fall.

**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 would like to take one or two new students.

**Do you have any experience working in or collaborating with industry?**

Several projects are relevant. I am Co-Investigator on an SBIR grant to develop scalable facial measurement for respirators. The idea is to use IPhone camera and motion detector to measure distances between facial landmarks in millimeters. This project comes out of our experience in 3D registration from 2D video. In other projects, a recent startup in face processing (Intraface) was recently acquired by Facebook. And I have consulted with a number of companies.

**We have a number of students who are in the job market. Do you have any advice for them?**

Some students come to graduate school because they want employable skills. Others seek future in academia or are undecided between academia and industry. A fair number of students are conflicted, especially when they look around and see faculty who have left for industry. I encourage students to stay in school. For one thing, technologies are moving fast and people who have a masters degree and jump to industry risk losing out in the next downturn. They also miss out on the wonderful intellectual freedom and independence we enjoy in academia. Think carefully.

**What are your hobbies?**

I enjoy road-biking. I have a hybrid bike for daily back and forth between home and university, and a road bike for 20- to 50-mile loops when the chance avails. Other interests include the Pittsburgh Arts and Lectures series, theater at the Public, and CMU drama series.

Thank you Dr. Cohn for taking the time to speak to us.
The Provost Fellowships in Intelligent Systems are awarded to two students of exceptional ability and promise. With the move to SCI, these Fellowships were created to replace the Dietrich School's Mellon Fellowship program.
Towards Interactive Natural Language Processing in Clinical Care

Free text makes it convenient for clinicians to conveniently capture rich information about patients. For the foreseeable future care-providers are likely to continue using narratives and first-person stories in Electronic Medical Records (EMRs), even though they lack common structure and standardized vocabulary. This makes it hard to extract relevant information from the EMR for computation and analysis. Despite advances in Natural Language Processing (NLP) techniques, building models is often expensive and time-consuming. Current approaches require a long collaboration between clinicians and data-scientists. Clinicians provide annotations and training data, while data-scientists build the models. With the current approaches, the domain experts—clinicians and clinical researchers, do not have provisions to inspect these models and give feedback. This forms a barrier to NLP adoption in the clinical domain, resulting in fewer real world applications.

Interactive methods are particularly attractive for clinical text due to the diversity of tasks that need customized training data. I demonstrated this approach in our prior work on NLPReViz (nlpreviz.github.io), an interactive tool for clinicians to train and build binary NLP models on their own for retrospective review of colonoscopy procedure notes. During the period of my fellowship, I will extend this effort to extend this effort from making predictions at document-level to identify relevant parts (or sentences) within a note. I propose an intelligent tool that can be used by clinicians to identify text for inclusion in summaries of patient notes, as ‘signouts’, from full-text notes. Lessons learned from the development and evaluation of this tool will provide insight into the generalized design of interactive NLP systems for wider clinical applications.

I am excited to be supported by the Provost Fellowships in Intelligent Systems this year. It provides me with great flexibility and opportunity to shape my dissertation research. I am looking forward to working on my new tool!
Instance-specific Causal Discovery from Observational Data

Discovering causal knowledge from data is fundamental in many areas of science. In the past 25 years, there has been considerable progress in developing general computational methods for modeling and discovering causal knowledge from observational data. A primary use of such methods is to analyze observational scientific data to generate causal Bayesian networks (CBNs) that model the causal relationships among variables. Almost all of these methods aim to learn the causal relationships that are common to instances (e.g., patient) in the population. They do not attempt to learn the causal relationships for each instance, which may differ in important ways from the common set of causal relationships in the population.

While obtaining accurate population-wide CBNs is useful, developing causal models that are specific to each instance is often important. For example, to understand and treat a patient, it is critical to understand the causal mechanisms that are operating in that patient. Our goal is to develop an instance-specific CBN learning method that searches the space of Bayesian networks to build a model that is specific to an instance X by guiding the search based on X's attributes (e.g., patient symptoms, signs, lab results, and genotype). We hypothesize that such an instance-specific algorithm will often model the causal relationships for X better than a population-wide one.

The significance of this project is in identifying causal relationships among variables of interest for each given instance. Such discoveries can affect many scientific fields, especially biomedicine and health care. For example, in cancer it is important to identify the typically small number genetic mutations that are causing (driving) a patient's cells to be cancerous, in contrast to the usually large number of “passenger” mutations that exist. That is, we wish to know the patient-specific genetic causes of the patient's cancer. In doing so, we better understand the mechanisms driving the cancer and can plan treatments that address each of them in order to arrest or eradicate the cancer. In summary, the development of instance-specific causal discovery methods is an important and largely unstudied area of research.

The Provost Fellowship supports me to develop and evaluate such instance-specific CBN learning methods in order to complete my doctoral dissertation work.
Thesis Proposals
Data and Knowledge-driven Iterative Bayesian Classification Rule Learning for Actionable Biomedicine

Abstract:
Knowledge Discovery in Databases (KDD) is a non-trivial process of the extraction of valid, novel, potentially useful, and ultimately understandable patterns in data. The KDD process is an interactive, iterative, and a multi-step process that includes data preparation, pattern search, knowledge evaluation, and refinement. Data mining is an important computational step in the KDD process that enables automated (or semi-automated) pattern search from the data. The most important patterns in the data are not necessarily the ones that have the largest support in the dataset. Instead, patterns that are actionable are subjectively more interesting to a user. This can only be accomplished with an integrated user-knowledge-driven and data-driven data mining. Typically in data-driven sciences, it can be tedious to specify all the relevant knowledge for the knowledge discovery task. Instead, it is more convenient to iteratively offer feedback to the model and suggest ways to make it subjectively more interesting. Biomedicine is such a data-driven science, where data mining methods are routinely used to discover novel patterns in datasets. This is largely due to the developments in high-throughput ‘omic’ technologies. Idiosyncrasies in such datasets, such as high-dimensionality and skewness, present challenges to data mining algorithms and motivate the need for developing new methods to address them. Bayesian Rule Learning (BRL) is a data mining algorithm with demonstrated success in biomedical datasets. BRL learns patterns as probabilistic classification rule models represented in form of explicit propositional logic. In this thesis project, we propose a knowledge and data-driven, iterative Bayesian Rule Learning framework that can extract subjectively interesting knowledge, in terms of its actionability, when compared to a purely-data driven approach. We propose to evaluate iBRL by searching for subjectively interesting knowledge on a dataset studying risk factors associated with cardiovascular diseases, guided by a domain expert.

Committee: Vanathi Gopalakrishnan (chair) (ISP, DBMI, PITT), Gregory Cooper (ISP, DBMI, PITT), Shyam Visweswaran (ISP, DBMI, PITT), Steven Reis (VC Clinical Research, PITT)
Entrainment Measures for Multi-Party Spoken Dialogues

Zahra Rahimi
Doctoral Student
since 2012

Abstract:

Entrainment or alignment of interlocutors during a conversation, becoming more similar to each other, has shown to be an important phenomenon that is correlated with several dialogue and task qualities in both human-human and human-computer interactions. It has been shown effective on increasing the human perspective of naturalness and friendliness of dialogue in spoken dialogue systems. Although dyadic interactions and multi-party text-based communications, such as twitter and online forums, get a lot of attention, multi-party spoken dialogues have been less studied. But, it is becoming more and more important. There are several applications for multi-party spoken dialogues such as human-human meetings, computer assisted collaborative learning, and multi-party chat bots.

There are several gaps and challenges in analysis of entrainment at multi-party spoken dialogue. At the input level of the algorithms, the challenge is how to approximate addressers and addressees for construction of the entrainment source and target pairs. At the model level, the challenge is how one can effectively extend the existing pair measures to multi-party measures. At the output level, the challenge is how we can use the entrainment information to predict multi-party interaction qualities. Moreover, there is a gap in studying the relation of linguistic modalities, such as acoustic-prosodic and lexical.

My dissertation aims at enhancing measurement of entrainment at multi-party spoken dialogues at two linguistic modalities of lexical and acoustic-prosodic; from two major perspectives of directional and unidirectional; and in multiple granularity of pair, speaker and group levels. To address source and target pair construction challenge, I propose to use a window-based algorithm as opposed to using only the immediate preceding turn or all preceding turns.

To enhance the extension of the model from dyad to multi-party, I propose two new models based on the entrainment behavior of speakers in groups for both frequency-based and generative model-based approaches. These new model extensions will replace the simple averaging method that is used in existing multi-party measurements. Finally, I propose to perform a comprehensive investigation on the relation of group outcomes and entrainment with respect to different granularities, measurement approaches, and linguistic modalities that was explored or introduced in this study. I expect that my dissertation work will provide an effective approach for measuring entrainment at multi-party spoken dialogue, contribute to evaluating human-human and human computer interactions, providing effective measures required to implement entrainment in multi-party spoken dialogue systems, and gives better insight to researchers studying human interactions.

Committee: Diane Litman (chair) (ISP, SCI, PITT), Rebecca Hwa (ISP, SCI, PITT), Kevin Ashley (ISP, LAW, PITT), Louis-Philippe Morency (CS, CMU)
Abstract:

In this dissertation, we attempt to exploit the unlabeled data to facilitate the classification of documents in terms of substantive legal concepts. Particularly, we focus on legal documents in a common law setting where each document contains case facts and legal reasoning by a fact finder to determine the final decision. For certain legal claims or issues, the mapping between case facts and the final decision often follows a stereotypical fact pattern, namely Legal Factors, that is fairly consistent under the rule of law. Since one might consider factors as concepts or topics that are discussed across decisions, we should be able to learn a document representation from text that reflects these concepts similar to topic modeling. This dissertation should make progress toward this goal. In particular, we try to answer the following research questions: How can a computer system utilize vector space models to automatically represent documents in a semantic vector space based on an underlying domain model rather than the corpus vocabulary space? Moreover, we are interested in how a computer system can identify portions of text that are relevant to the underlying domain model.

In order to answer these research questions, we have formed a number of hypotheses including the following: Dimensionality reduction techniques on the term-document matrix of a larger corpus will enable learning VSMs that were more expressive of the underlying legal factors. In particular, the target representation of judicial decisions will outperform TF-IDF baselines as well as a previously developed model of the document classification task with respect to macro F1 scores as the measure.

In order to frame how to answer our research question we conducted a preliminary study. We started with a term-document matrix of a larger corpus. We used a variety of unsupervised methods and undertake a series of studies to learn useful mappings between texts and the domain model. We assessed the performance of our models by extrinsic evaluation. To evaluate our document representations, we used a corpus of 179 judicial decisions that are labeled by expert annotators with respect to the underlying legal factors and tried to predict the factors for each document in a multi-label classification framework. Our preliminary results show that one can detect some signal (i.e., features) for classifying factors in case texts. Our VSMs performed better than a previously published attempt at learning to identify factors in cases. We plan to improve upon these results and develop techniques with which a target representation of judicial decisions will filter out case specific information and identify portions of text that discuss the legal factors.

Committee: Kevin Ashley (chair) (ISP, LAW, PIT), Diane Litman (ISP, SCI, PIT), Daqing He (ISP, SCI, PIT), Kayhan Batmanghelich (ISP, DBMI, PIT)
Abstract:

Free text makes it convenient for clinicians to conveniently capture rich information about patients. For the foreseeable future care-providers are likely to continue using narratives and first-person stories in Electronic Medical Records (EMRs), even though they lack common structure and standardized vocabulary. This makes it hard to extract relevant information from the EMR for computation and analysis. Despite advances in Natural Language Processing (NLP) techniques, building models is often expensive and time-consuming. Current approaches require a long collaboration between clinicians and data-scientists. Clinicians provide annotations and training data, while data-scientists build the models. With the current approaches, the domain experts-- clinicians and clinical researchers, do not have provisions to inspect these models and give feedback. This forms a barrier to NLP adoption in the clinical domain, resulting in fewer real world applications.

Interactive methods are particularly attractive for clinical text due to the diversity of tasks that need customized training data. I demonstrated this approach in our prior work on NLPReViz (nlpreviz.github.io), an interactive tool for clinicians to train and build binary NLP models on their own for retrospective review of colonoscopy procedure notes. During the period of my fellowship, I will extend this effort to extend this effort from making predictions at document-level to identify relevant parts (or sentences) within a note. I propose an intelligent tool that can be used by clinicians to identify text for inclusion in summaries of patient notes, as ‘signouts’, from full-text notes. Lessons learned from the development and evaluation of this tool will provide insight into the generalized design of interactive NLP systems for wider clinical applications.

Committee: Harry Hochheiser (chair)(ISP, DBMI, Pitt), Shyam Visweswaran (ISP, DBMI, Pitt), Rebecca Hwa (ISP, SCI, Pitt), Wendy Chapman (DBMI, Utah)
Dissertation Defenses
Robust Parsing for Ungrammatical Sentences

Natural Language Processing (NLP) is a research area that specializes in studying computational approaches to human language. However, not all of the natural language sentences are grammatically correct. Sentences that are ungrammatical, awkward, or too casual/colloquial tend to appear in a variety of NLP applications, from product reviews and social media analysis to intelligent language tutors or multilingual processing. In this thesis, we focus on syntactic parsing, an essential component of many NLP applications. We investigate the impact of ungrammatical sentences on statistical parsers. We also hypothesize that breaking up parse trees from problematic parts prevents NLP applications from degrading due to incorrect syntactic analysis.

A parser is robust if it can overlook problems such as grammar mistakes and produce a parse tree that closely resembles the correct analysis for the intended sentence. We develop a robustness evaluation metric and conduct a series of experiments to compare the performances of state-of-the-art parsers on the ungrammatical sentences. The evaluation results show that ungrammatical sentences present challenges for statistical parsers, because the well-formed syntactic trees they produce may not be appropriate for ungrammatical sentences. We also define a new framework for reviewing the parses of ungrammatical sentences and extracting the coherent parts whose syntactic analyses make sense. We call this task parse tree fragmentation. The experimental results suggest that the proposed overall fragmentation framework is a promising way to handle syntactically unusual sentences; they also validate the utility of parse tree fragmentation methods in two external tasks of sentential grammaticality judgment and semantic role labeling.

Committee: Rebecca Hwa (Chair) (ISP & SCI, PITT), Diane Litman (ISP & SCI, PITT), Christian Schunn (ISP & Psychology, PITT), Na-Rae Han (Linguistics, PITT)

Defense Date: October 17, 2017
Human-Data Interaction in Large and High-Dimensional Data

Human-Data Interaction (HDI) is an emerging field which studies how humans make sense of large and complex data. Visual analytics tools are a central component of this sense-making process. However, the growth of big data has affected their performance, resulting in latency in interactivity or long query-response times, both of which degrade one's ability to do knowledge discovery. To address these challenges, a new paradigm of data exploration has appeared in which a rapid but inaccurate result is followed by a succession of gradually more accurate answers. As the primary objective of this thesis, we investigated how this incremental latency affects the quantity and quality of knowledge discovery in an HDI system. We have developed a big data visualization tool and studied 40 participants in a think-aloud experiment, using this tool to explore a large and high-dimensional data. Our findings indicate that although incremental latency reduces the rate of discovery generation, it does not affect one's chance of making a discovery per each generated visualization, and it does not affect the correctness of those discoveries. However, in the presence of latency, utilizing contextual layers such as a map result in fewer mistakes while exploring higher-dimensional visualizations lead to more incorrect discoveries.

As the secondary objective, we investigated what strategies improved a subject's performance. Our observations suggest that successful participants explore the data methodically, by first examining simple and familiar concepts and then gradually adding complexity to the visualizations, until they build a correct mental model of the inner workings of the tool. With this model, they generate several discovery patterns, each acting as a blueprint for forming new insights. Ultimately, some participants combined their discovery patterns to create multifaceted data-driven stories. Based on these observations, we propose design guidelines for developing HDI platforms for large and high-dimensional data.

Committee: Micheal Lewis (Chair) (ISP & SCI, PITT), Christian Schunn (ISP & Psychology, PITT), Yu-Ru Lin (ISP & SCI, PITT), Randy Sargent (IR, CMU)

Defense Date: November 27, 2017
Mastery development requires not only acquiring component skills, but also practicing their integration into more complex skills. When learning programming, an example is to first learn \( += \) and loops, then learn how to combine them into a loop that sums a sequence of numbers. The existence of integration skills has been supported by cognitive science research, yet it has rarely been considered in learner modeling, the key component for adaptive assistance in an intelligent tutoring system (ITS). Without this, early assertions of mastery in ITSs after only basic component skill practice or practice in limited contexts may be merely indicating shallow learning.

My dissertation introduces integration skills, widely acknowledged by cognitive science research, into learner modeling. To demonstrate this, I chose program comprehension with a complex integrative nature. To provide grounds for skill modeling, I applied a difficulty factor assessment (DFA) approach (from cognitive science) to classroom studies, and identified integration skills along with generalizable integration difficulty factors in common basic programming patterns. I used the DFA data to inform the construction of the learner model, CKM-HI, which incorporates integration skills in a hierarchical structure in a Bayesian network (BN). Compared with other machine learning approaches, BN naturally utilizes domain knowledge and maintains interpretable knowledge states for adaptation decisions. To address the limitation of prediction metrics to evaluate such multi-skill learner models, I proposed and applied a multifaceted evaluation framework. Data-driven evaluations on a real-world dataset show that CKM-HI is superior to two popular multi-skill learner models, CKM and WKT, regarding predictive performance, parameter plausibility, and expected instructional effectiveness. To evaluate its real-world impact, I built a program comprehension ITS driven by learner models and a classroom study deploying this system suggests that CKM-HI could lead to better learning than the CKM model.

My dissertation work is the first to systematically demonstrate the value of integration skill modeling, and offers novel integration-level learner modeling and multifaceted evaluation approaches applicable to a broader context. Further, my work contributes recent ITS infrastructure and techniques to programming education, and also contributes an example of taking an interdisciplinary approach to ITS research.

Committee: Peter Brusilovsky (Chair) (ISP & SCI, PITT), Christian D. Schunn (ISP & Psychology, PITT), Marek Druzdzel (ISP & SCI, PITT), Kenneth Koedinger (HCII, CMU)

Defense Date: July 5, 2018
Program Construction Examples in Computer Science Education: From Static Text to Adaptive and Engaging Learning Technology

My dissertation is situated in the field of computer science education research, specifically, the learning and teaching of programming. This is a critical area to be studied, since, primarily, learning to program is difficult, but also, the need for programming knowledge and skills is growing, now more than ever. This research is particularly focused on how to support a student's acquisition of program construction skills through worked examples, one of the best practices for acquiring cognitive skills in STEM areas.

While learning from examples is superior to problem-solving for novices, it is not recommended for intermediate learners with sufficient knowledge, who require more attention to problem-solving. Thus, it is critical for example-based learning environments to adapt the amount and type of assistance given to the student's needs. This important matter has only recently received attention in a few select STEM areas and is still unexplored in the programming domain. The learning technologies used in programming courses mostly focus on supporting student problem-solving activities and, with few exceptions, examples are mostly absent or presented in a static, non-engaging form.

To fill existing gaps in the area of learning from programming examples, my dissertation explores a new genre of worked examples that are both adaptive and engaging, to support students in the acquisition of program construction skills. My research examines how to personalize the generation of examples and how to determine the best sequence of examples and problems, based on the student's evolving level of knowledge. It also includes a series of studies created to assess the effectiveness of the proposed technologies and, more broadly, to investigate the role of worked examples in the process of acquiring programming skills.

Results of our studies show the positive impact that examples have on student engagement, problem-solving, and learning. Adaptive technologies were also found to be beneficial: The adaptive generation of examples had a positive impact on learning and problem-solving performance. The adaptive sequencing of examples and problems engaged students more persistently in activities, resulting in some positive effects on learning.

Committee: Peter Brusilovsky (Chair) (ISP & SCI, PITT), Christian D. Schunn (ISP & Psychology, PITT), Diane Litman (ISP & SCI, PITT), Vincent Aleven (HCII, CMU)

Defense Date: July 24, 2018
New PhD Graduates
An Interview with Dr. Huma Baradaran Hashemi:

Huma Baradaran Hashemi joined ISP in 2011 and graduated in 2017. Her advisor during her PhD was Dr. Rebecca Hwa from the Computer Science department.

Can you briefly describe your dissertation work and your experience with your thesis defense?

My dissertation was about robust parsing of ungrammatical sentences. With the help of my advisor, Rebecca, we proposed a new framework to revise parse trees of ungrammatical sentences. We introduced several methods to extract coherent parts of the trees using grammatical mistakes in the sentences.

As I look back at my PhD journey, I think the thesis defense is only the last gate. The miles that one has traveled to reach that gate are as important. In my opinion, a clear dissertation proposal is crucial to have a direct and shorter path to the defense gate.

How was your experience in ISP?

ISP was our home. I cannot imagine spending tough years of PhD somewhere other than ISP. Mahdi, my husband, and I were lucky to be part of ISP. Although ISP was small, it was a diverse program which presented us with so many perspectives on artificial intelligence. The faculties taught us the range of things one can possibly learn in graduate school, from doing great research to writing papers. We were also lucky to have nice friends at ISP that made our six years at Pittsburgh such a memorable experience. Our wonderful administrators had also a great impact on our survival in grad school and for us in the new country.

Do you have any suggestions to improve the program?

My suggestion is that current ISP students be more involved in ISP related decisions. It would be nice to convey this feeling to the students that ISP is for them and they are part of ISP.

What have you been doing since graduation?

I am currently an applied scientist at Microsoft knowledge graph team which is part of Microsoft AI and Research group. My role is to understand natural language questions so that we can extract answers from the knowledge graph. After my graduation, I found a job and exactly four days after I signed my job offer, we went back home to see our families after 6 years, 6 months and 24 days.

"Although ISP was small, it was a diverse program which presented us with so many perspectives on artificial intelligence."

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

The first year at ISP is a great opportunity to talk to all the members, either students or professors, to find out different aspects of AI that you can work on. Also, try to build your AI foundations by taking challenging courses and truly spend time to learn them. You will use this knowledge at some point in the future for sure.
An Interview with Dr. Saman Amirpour Amraii:

Saman Amirpour Amraii joined ISP in 2009 and graduated in 2017. His advisor during his PhD was Dr. Michael Lewis from the School of Computing and Information.

Can you briefly describe your dissertation work and your experience with your thesis defense?

I started my research with Prof. Lewis on the topic of Human-Swarm Interaction. Here, the goal is to investigate how an operator can control hundreds of robots, where one should constantly observe the state of a complex system and then guide that system accordingly to achieve a desired task. During my work, I realized this problem can be generalized to the recently proposed domain of Human-Data Interaction where the goal is to study how humans make sense of large and complex data. A central problem in HDI is the issue of latency: as we start dealing with big data, our analytical tools become slow, e.g. a database query can take hours. Such a latency have been shown to have a detrimental effect on the productivity of operators. One of the suggested solutions to reduce latency is to build systems that start with a rapid but inaccurate response and then gradually improve their results over time. For example, progressive visualization techniques start with a small sample of the data and then gradually increase the details of the visualization. Although progressive techniques scale well with big data, there are still many open questions on the efficacy of utilizing such system in a knowledge discovery task. In my thesis, I built a progressive visualization system which enabled participants to explore a large dataset on demographics. I then investigated how such a system differs from an ideal case of no latency and in particular, how does it affect the quality and quantity of discoveries achieved by the users. Surprisingly, the user studies indicated that in most tasks, the depth and breadth of discoveries achieved in a progressive system are as good as an instantaneous one, even though the user is often looking at an incomplete and dynamic representation of the data.

Writing the thesis is a demanding process. It was interesting to me that the hardest part was to define the research question in the first place! For me, it was an iterative process where I started with a rough question, built a prototype tool, did some user studies, and then used the results to further improve the question. I also greatly benefited from the suggestions given to me by my advisers. I sent weekly reports to my committee and asked for their comments on various roadblocks I faced. This approach was immensely helpful as it provided me with a much needed guidance and also allowed me to be cognizant of the expectations my committee had for the defense day. I strongly suggest to ISP students to keep their committee members in the loop and benefit from their advice as they work on their research.

How was your experience in ISP?

ISP is excellent in giving a lot of freedom to students to choose from a wide range of topics in AI, and work with a world-class faculty from multiple departments. This allows ISP to be a fertile ground for multidisciplinary research, and also allows ISP researchers to provide direct and tangible benefit to other disciplines as well.

Do you have any suggestions to improve the program?

I would advocate for even more interdisciplinary and inter-institute research. I believe that there are many more departments in the university that can greatly benefit from the AI research performed at ISP. Such collaborations will in turn provide numerous real world use cases for developing and testing AI algorithms, and can lead to a diverse set of learning and cross pollination opportunities.
What have you been doing since graduation?

During my PhD research, I saw a great interest from other researchers and even from industry in the novel technologies being developed for exploring and understanding large and complex data. I realized the progressive visualization paradigm can play a vital role in addressing the challenges faced in analyzing big data, and therefore I decided to continue my work by turning that into a startup. Our company, xSeer.io, is now on the leading edge of designing and developing human-oriented big data analytics system.

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

Student life is tough; no money, away from family and friends, and a lot of stress for delivering on the research and coursework. However, I still find it miraculous that one can be paid to sit behind a desk, read scientific papers, and contemplate on how she can contribute to the science! This is an extraordinary opportunity which you will miss when you leave the academia. The time you have at ISP and in Pittsburgh are some of the best years of your life. Cherish every moment of it.

Saman Amirpour Amraii and his advisor Dr. Michael Lewis in the graduation ceremony
ISP Events

ISP is hosting the annual picnic to welcome new students and faculty. The picnic event is going to be held on Friday, September 14th at the Bartlett Shelter in Schenley Park. We would like to invite all local ISP alumni to attend as well and hope to see all of you and your family there. We will have lots of food, beverages and fun.

Here are the details of this event:

Where:

   Bartlett Street Shelter at Schenley Park, Overlook Drive
   Pittsburgh, PA, 15207

When:

   Friday, September 14th
   3:00 pm to 6:00 pm

Please let us know if you will be able to join us by emailing Michele Thomas: paum4b@pitt.edu
Contact Us

Getting in touch with ISP is easy. ISP's pages on Twitter and Facebook are active and we are always reachable by Gmail. We encourage all the students, faculty, alumni and whoever is interested to find the latest news about ISP to follow us:

![Twitter](image) ![Facebook](image) ![Gmail](image)

This is the fourth edition of the ISP Newsletter and every effort has been taken to make it as thorough as possible. If you have any further comments, news or updates please send an email to the ISP administrator, Michele Thomas (paum4b@pitt.edu)

The digital copy of the newsletter is available at:
http://www.isp.pitt.edu/newsletter

**Director:** Dr. Diane Litman
**Editor:** Saba Dadsetan