2019 Meet N Greet

1. Asudeh, Abholfaz -
2. Caragea, Cornelia – Information Retrieval group
3. Di Eugenio, Barbara – Natural Language Processing
4. Johnson, Andy – Electronic Visualization Laboratory (EVL)
5. Mansky, William – PL Theory + Verification
6. Michaelis, Joseph – Learning + Interest + Technology
7. Parde, Natalie – Semantics, Multimodal NLP, Robotics, Healthcare
8. Pina, Luis – Dynamic Software Updated, Multi-Version Execution, Java Fuzzing and Concolic Execution
10. Sidiropoulos, Tasos – Graph Algorithms, Computation Geometry
12. Solworth, Jon – Towards a new secure and private software
13. Stephens, Brent – In-network Computing, Programmable Networking, RDMA
14. Tan, Wei – Computer Vision, Pattern Recognition, Deep Learning
15. Wu, Xingbo – Memory and storage, Performance & Efficiency, Key-value Systems
17. Zheleva, Elena – Causal Data Science, Unbiased Machine Learning, Personalize Privacy Assistants
Abolfazl (Abol-faz) Asudeh

- asudeh@uic.edu, http://asudeh.github.io
- SEO 1131

- **Brainstorming** is the key to solving problems and that "good" research is teamwork, in opposed to individuals' effort

- To find efficient, effective, and scalable algorithmic solutions for data science problems
  - Data Management Community
Research

Top-k and Ranking Query Processing

Compact Representatives for Datasets

Responsible Data Science (Data Ethics)

Web Databases; P2P Marketplaces
Expedia, Airbnb, BlueNile, ...

Data Management for Machine Learning
Computer Networks
Research

- Problem Complexity
- Data Structure
- Combinatorial Geometry
- Approximation Algorithms
- Randomized Algorithms
- Sampling Techniques
- MithraLabel: Nutritional Labels for Datasets

- MithraRanking

### Data Overview
- **Attribute Name**: Recidivism_score, Violence_score
- **Histogram**: Max: 1.69, Min: -0.69, Mean: 0.33, Null Entries: 33
- **Functional Dependencies**: Violence_score → Recidivism_score
- **Maximal Uncovered Patterns**: See visual graph

### Functional Dependencies
- Violence_score → recidivism_score
- Violence_score → age
- Violence_score → event_first_name
- Violence_score → marriage_status

### Maximal Uncovered Patterns
- See visual graph

### Nutritional Labels for Datasets

### MithraRanking
- **Select Dataset**
- **Upload Your Dataset**
- **Ranking Attributes**
  - C_days_from_compas
  - Juvenile_c_count
  - Days_b_screening_arrest
- **Cosine Similarity**: 98%

### Fairness Criteria
- **Analyzing**: 30%
- **Fairness Constraint(s)**: at most 50% age <= 56
- **Direction**: Percentage
- **Select Attribute**: None
- **Select Condition**: None

### Ranking Attributes
- C_days_from_compas: 0.21
- Juvenile_c_count: 0.74
- Days_b_screening_arrest: 0.61

### Ranking
- **Suggestions**
  - Fair: 0.19, Most Stable: 0.24, Fair & More Stable: 0.20
- Days_b_screening_arrest: 0.65, Juvenile_c_count: 0.30

### Back to List
Research Interests

- Artificial Intelligence
- Information Retrieval and Extraction
- Natural Language Processing
- Opinion Mining, Sentiment and Subjectivity Analysis
- Machine Learning for Big Data

Acknowledgements
Information Extraction – Keyphrase Extraction

From scholarly documents: [WWW-19, ACL-17, AAAI-17, EMNLP-16, EMNLP-15]

Disaster Twitter data: [WWW-19]

<table>
<thead>
<tr>
<th>No.</th>
<th>Tweet text</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>we need help in Houston, our apartments are surrounded with water like an island we need rescue 10373 N Sam Houston Pkwy E need help</td>
</tr>
<tr>
<td>2.</td>
<td>@houstonpolice please help I’m stranded with my kids I need help fast my address is 8618 Banting st. houston tx 77078.</td>
</tr>
<tr>
<td>3.</td>
<td>Big tree fell on power lines and blocking Brown Ave near Washington St in Orlando’s Thornton Park neighborhood. #HurricaneIrma</td>
</tr>
<tr>
<td>4.</td>
<td>Very extensive damage sustained throughout #Wilmington, #ncwx... from #hurricane #Florence. Lots of trees split or uprooted, siding ripped from homes, powerlines down, flooding of downtown streets, etc.</td>
</tr>
<tr>
<td>5.</td>
<td>I am evacuated from my house but I’m safe. #fire #CampFire #WoolseyFire #wildfire #safe #Evacuation #evacuations #EVACUATED #scary #ThousandOaks #Camarillo</td>
</tr>
</tbody>
</table>
Is that Noah Smith?

Confidential Review Copy. DO NOT DISTRIBUTE.

Topics to Avoid:
Demoting Latent Confounds in Text Classification

Anonymous EMNLP-IJCNLP submission

Abstract

Despite impressive performance on many text classification tasks, deep neural networks tend to learn frequent superficial patterns that are specific to the training data and do not always generalize well. In this work, we observe this limitation with respect to the task of native language identification. We find that standard language (L1) of an individual based on their language production in a second language (L2, in our case English). In this scenario, a model trained to predict L1 is likely to predict that a person is, say, Swedish, if the texts authored by that person are about Sweden, because the training data exhibits such topical confounds. This problem is the focus of our work.

[EMNLP-19]
Other NLP Projects

- Emotion Detection: "My doctor’s office is very clean, who cares when he has prescribed me a wrong medication for six months!"
  [EMNLP-18, AAAI-18]

- Stance detection: @realDonaldTrump is the only honest voice of the @GOP and that should scare the shit out of everyone! #SemST. Target: Hillary Clinton; Stance: Against; Sentiment: Positive.
  [EMNLP-19]

- Pessimism/optimism detection: “Life’s about taking risks. Don’t be afraid to put yourself out there.”
  [EMNLP-18]
Thank you!

Hamed Khanpour
Sujatha Das
Florin Bulgarov
Corina Florescu
Lucas Sterckx
Yingjie Li
Andreea Godea
Alina Ciobanu
Ana Uban
Jishnu Ray Chowdhury
Ashwini Tonge
Krutarth Patel
Kishore Neppalli
Natural Language Processing @ UIC: NLP with a purpose

Barbara Di Eugenio -- nlp.cs.uic.edu

Thanks to: NSF, NIH, ONR, Motorola, Yahoo!, UIC CRB, Politecnico Torino, Qatar Research Foundation, Institute for Education Studies
Conversational Agents

NLP for educational technology

Articulate: NL dialogue for visualization

Multimodal interaction for assistive robots (NSF)

Health care: virtual health coaching via SMS (NSF)
Information Presentation and Summarization

✧ **PatientNarr**: integrates physician discharge notes, nursing records and patient perspective to generate patient-centered summaries of hospital stays (NIH)
Information Gathering & Summarization: Older projects

- **Tweets and Life Events** *(Yahoo)*: Who got engaged?

  Jenna Middleton @@AdamxTorres I am just think how dumb Eli is.. He and Jess are engaged.

- **SongRecommend** *(Motorola)*: Summarization of reviews for recommender systems: extract information about songs from album reviews, generate summaries
William Mansky, PL Theory + Verification

• Research interests: mathematical models of programming languages, proving programs correct, modeling concurrency

• Current projects: verified web server, reasoning about low-level concurrency, verifying database implementations, connecting verified systems
How can we write programs that we know are correct?
Writing a Correct Program

```java
i = 1;
while(i < n){
    r = r * i;
    i++;
}
```

- When $n$ is 1, should return 1
- When $n$ is 2, should return 2
Writing a Correct Program

```java
i = 1;
while(i <= n){
    r = r * i;
    i++;
}
```

- When n is 1, should return 1
- When n is 2, should return 2
- When n is 5, should return 120
Writing a Proved-Correct Program

∀n, f(n) = n!

Proof

i = 1;
while(i <= n){
    r = r * i;
    i++;
}

x = v;
rest

rest

Semantics

Undefined behavior?
Concurrency?

Memory? I/O?
Program Verification

• Result: real programs that are provably bug-free!
  — No out-of-bounds array accesses, null pointer dereferences, memory leaks, race conditions, ...
  — And they actually do the right thing!

• Model interesting features: I/O, concurrency

• Verify interesting programs: web servers, concurrent databases

• Looking for students!
Deep Understanding + Social Robotics + Interest Development + Social Robotics
Long-term Interactions
Research Process

Examine

Build HCI, Learning & Interest Theory

Design

Develop Interaction Models → Implement on Learning Technology → Integrate with real-world activities

Assess

Evaluate HCI, Learning & Interest
Research Process

Examine
Build HCI, Learning & Interest Theory

Develop Interaction Models

Design
Implement on Learning Technology
Integrate with real-world activities

Assess
Evaluate HCI, Learning & Interest
Research Process

Examine
Build HCI, Learning & Interest Theory

Develop Interaction Models

Design
Implement on Learning Technology

Integrate with real-world activities

Assess
Evaluate HCI, Learning & Interest
Research Process

Examine
- Build HCI, Learning & Interest Theory

Design
- Develop Interaction Models
- Implement on Learning Technology

Assess
- Integrate with real-world activities
- Evaluate HCI, Learning & Interest
Research Process

Examine

Build HCI, Learning & Interest Theory

Develop Interaction Models

Design

Implement on Learning Technology

Integrate with real-world activities

Assess

Evaluate HCI, Learning & Interest
Project 1: In-home learning companion

Design companion robot to support science interest

Partner with teachers, parents, classrooms and community org

Goodman Community Center
Project 2: Mobile classroom assistant

Design learning assistant robot to support STEM interest in classrooms

Partner with teachers, parents, classrooms and local schools
Natalie Parde
Assistant Professor

Office location: 1132 SEO

Courses taught recently: 594, 421

Lab: Natural Language Processing Laboratory (https://nlp.lab.uic.edu)
Core Research Areas

- **Semantics**
  - Metaphor, sarcasm

- **Multimodal NLP**
  - Grounded language learning, visual storytelling

- **Robotics**
  - Human-robot dialogue, interactive language learning, social robotics

- **Healthcare**
  - Cognitive health promotion, dementia detection
Luís Pina (Loo-eesh Pee-na)
Research - Dynamic Software Updates

- DSU for Java
  - How to write state transformation
  - How to transform the state efficiently
  - How to test that updates are correct

- Initial prototypes with Software Transactional Memory
Research - Multi-Version Execution

Followers:

- Different configurations
  - E.g., logging enabled
  - Client works if feature crashes

- Heavyweight incompatible analyses
  - E.g., Valgrind with asan
  - Client retains native latency

- Updating new version
  - Reliable DSU in the background
Research - Java Fuzzing and Concolic Execution

```java
void method(int input) {
    if (input == 9082374) {
        crash();
    }
}
```

1. Fuzzing
   - Generate random inputs
   - Fitness function to maximize coverage
   - Use generators for structured input
     - E.g., XML, HTML, Javascript

1. Concolic Execution
   - Gather constraints in execution
     - E.g., (input ≠ 9082374)
   - Negate them and solve
   - Use the solution as input

1 is fast but gets stuck, 2 can reach deep inputs but is very slow

Combining the speed of 1 and the insights of 2 is the holy grail I’m working on
DEEP LEARNING + X

Sathya Ravi
08/29/2019
DEEP LEARNING + CONSTRAINTS

Fairness

Privacy

Interpretability
DEEP LEARNING + DYNAMICS

Feedback

GAN
DIRECTIONS

- Theory: Understanding existing algorithms
- Application: Identify relationships and exploit them to design better training algorithms
Tasos Sidiropoulos

**Graph algorithms**

social networks  
brain networks

“Graph theory is the new calculus”  
-- Daniel Spielman, Yale

**Computational geometry**

data  
geometry  
algorithms

discovering structure  
exploiting structure

"Graph theory is the new calculus" -- Daniel Spielman, Yale
Bob Sloan: Security and Privacy Policy

- Fundamental questions:
  - What surveillance schemes and limitations are technically feasible?
  - How do we make good tradeoffs between privacy versus free good stuff (Gmail? FB?) and crime and terrorism fighting?
Example Topics, themes

• Analysis of 50 State’s definition of Personally Identifiable Information (PII) and connection to data mining, technical feasibility, etc. (former student)
• Many: Use of Simple Game Theory
• Current questions: (1) Do companies have enough info on expected losses due to breaches? (2) Algorithmic transparency in AI/Machine Learning era.

IEEE Security & Privacy Magazine, May/June 2018:

When Is an Algorithm Transparent? Predictive Analytics, Privacy, and Public Policy

Robert H. Sloan | University of Illinois at Chicago
Richard Warner | Chicago-Kent College of Law

The problem of algorithmic transparency is pressing. Predictive systems are transparent for consumers if they can ascertain the risks and benefits associated with the predictive systems to which they are subject. We examine three ways to meet this condition: disclosing source code, transparency without disclosing source code, and informational norms.
Towards new systems software for security and privacy

Jon A. Solworth

Dept. of Computer Science and
solworth@ethos-os.org

www.ethos-os.org
4224 SEL
CS 587 MW 6:00-7:15
The state of software today:

- Every widely used system today has been broken by attackers
- Software lacks appropriate security services
- It's too complex and fails easily
- So far, the attacker *always* wins

Every effort to fix it has failed, resulting in massive compromise

- US National Security Agency
- US Office of Personnel Management
- Mass surveillance by governments
A better approach: Design-for-security systems semantics

- A successful attack causes a software failure
- Too hard to make all software immune to failure
- Our goal is design system semantics to make applications more resilient to attack, e.g.
  - A C program might have a buffer overflow
  - A Java program cannot
- This is an effective mechanism because the vast majority of security holes are an artifact of system software semantics
- We analyze attacks, find underlying issue, design systems components to withstand broad classes of attacks
Projects

Security is Semantics

**Ethos** A clean-slate OS with
- Strong security services
- Simple, composable semantics
- Ethos applications are inherently more robust

**CRISP** A much more secure and private web experience
- Replace HTTP+Javascript
- Provides WebApps with Privacy and Security

**FASOR** Fast and Strong Onion Routing
- Onion Routing protocol for private Internet
- Simpler and more flexible than Tor

**Disruptor Stores** Using distributed anonymous storage to provide privacy on the Internet
Brent Stephens
Assistant Professor

Office location: 1330 SEO

Courses taught recently:
594: Data Center Networking
494: Data Center Systems

Research interests: In-network Computing, Programmable Networking, RDMA
I work with programmable networking technologies

Examples:

Mellanox Innova-2 Flex Open Programmable SmartNIC (2x100Gbps w/ RDMA and FPGA)

and

Barefoot Tofino programmable switch (33x100Gbps)
I create new network protocols and services
I work with real cloud applications and systems
I like to ride bicycles
Wei Tang

• Assistant Professor
• Ph.D., Northwestern University, 2019
• B.E. (2012) and M.E. (2015), Beihang University, Beijing, China

• Computer vision, pattern recognition, deep learning
• CS 415 Computer Vision I

Website: https://tangw.people.uic.edu
Email: tangw@uic.edu
Research interests

• Computer vision, pattern recognition, deep learning

• Current focus: human-centered visual computing
  • Human motion and action analysis
  • Applications in human-computer interaction and virtual reality

Website: https://tangw.people.uic.edu

Email: tangw@uic.edu
Xingbo Wu
Joined UIC CS in Fall’18

Research interests
● Memory and storage
● Performance & efficiency
● Key-value systems

Teaching
● CS 461 Operating Systems
● CS 594 High-perf. NoSQL DB
We Make System Software More Efficient

Applications
- Hadoop
- Spark

Databases
- MySQL
- ORACLE
- DB2
- Redis
- levelDB

OS
- Linux
- Docker
- Xen Project

Hardware
Our Efforts on Efficient Data Management

- [SoCC’17] 15x faster index search with a small cache
- [ICS’19] Better CPU cache utilization with Software-Defined Cache
- [EuroSys’19] An asymptotically and practically faster kv index
- [APSys’16] Remove expensive flushes for NVM KV caches
- [EuroSys’16] Remove 50% of misses in KV cache without adding DRAM
- [APSys’15] Enable comprehensive copy-on-write for overlay file systems
- [ATC’15] Reduce write-amplification of KV stores by 10x-20x
- [Systor’15] Eliminate small metadata writes in virtual disks

Back to List
Machine Learning for Intelligent Design of Power Converters

Xinhua Zhang
Department of Computer Science
University of Illinois at Chicago
https://www.cs.uic.edu/~zhangx/
Motivation

• Use machine learning to automatically design, optimize, and synthesize electrical devices
  • Topology
  • Parameter

• Huge impact

• Great opportunity

“Circuit-GNN: Graph Neural Networks for Distributed Circuit Design”
He, Zhang, Katabi, ICML 2019.

(a) 6-resonator filter
(b) 4-resonator filter
Approach

- Approximate the behaviour of a circuit by neural networks

High-fidelity simulator

compare

High-fidelity simulator
Reinforcement learning for both parameter & topology

topology synthesis + parameter optimization
Advisor and skills

- Xinhua Zhang (Computer Science)
  - Machine Learning
- Sudip Mazumder (Electrical and Computer Engineering)
  - Power electronics
  - IEEE Fellow
- Deep learning and reinforcement learning
- Coding in Python (PyTorch or TensorFlow)
- Ph.D. students only (basic analog circuit background)
- Reference: Electrical Power Converter part of solicitation: https://arpa-e-foa.energy.gov/FileContent.aspx?FileID=e14e478b-6e50-47a9-bbd3-be0a8ee0a880
Imagine you have collected or been given a network dataset.

**Research goal:** Identify and resolve barriers to causal inference from relational data for real-world applications.

- Does social media make us more “hateful” and why?
- What interventions can reduce bullying in schools?
- Did Juul ads lead to increase in youth vaping?
- What makes people feel empathy for others?

**Research goal:** Improve machine learning models by addressing inherent biases in (found) data.

**Research goal:** Empower people in their privacy choices through personalized privacy assistants.

---

Back to List