

Mid-Atlantic Student Colloquium on Speech, Language and Learning (MASC-SLL)

October 11, 2013

University of Maryland, Baltimore County

<http://bit.ly/masc13>

Organizing Committee:

- Snigdha Chaturvedi (UMCP)
- Ben Johnson (UMBC)
- Michael Paul (JHU)
- Jennifer Sleeman (UMBC)
- Bryan Wilkinson (UMBC)

Faculty Advisers

- Tim Finin (UMBC)
- Tim Oates (UMBC)

Program Committee:

- Dan Goldwasser (UMD)
- He He (UMD)
- Yuening Hu (UMD)
- Abhay Kashyap (UMBC)
- Elijah Mayfield (CMU)
- Priscilla Moraes (UDEL)
- Nanyun Peng (JHU)
- Alexander Rush (Columbia)
- Efsun Sarioglu (GWU)
- Dezhao Song (Lehigh)
- Shashank Srivastava (CMU)
- Adam Teichert (JHU)
- Svitlana Volkova (JHU)
- Travis Wolfe (JHU)

Sponsors:

Financial and logistical support for 2013 MASC-SLL is generously provided by UMBC's Office of the Provost, College of Engineering and Information Technology, UMBC Graduate School and the Computer Science and Electrical Engineering Department.

Guest wifi access: *UMBC Visitor*, Hashtag: *#MASC13*, Web Site: <http://bit.ly/masc13>

Program Schedule

09:00-09:45	Registration
09:45-10:00	Opening
10:00-11:20	<p>Oral presentations I</p> <ul style="list-style-type: none"> • <i>'Semantic Textual Similarity Systems'</i>, Lushan Han, Abhay Kashyap, Tim Finin, James Mayfield and Jonathan Weese (UMBC & JHU) • <i>'Toward Faster Audio Search Using Context-Dependent Hashing'</i>, Keith Levin, Aren Jansen and Ben Van Durme (JHU) • <i>'Learning to Interpret Natural Language Instructions'</i>, Shawn Squire, Monica Babes-Vroman, Marie Desjardins, Ruoyuan Gao, Michael Littman, James Mac-Glashan and Smaranda Muresan (UMBC & Brown) • <i>'Lexical and Hierarchical Topic Regression'</i>, Viet-An Nguyen, Jordan Boyd-Graber and Philip Resnik (UMCP)
11:20-12:10	Poster Session I
12:10-12:40	Lunch
12:40-1:40	<p>Panel</p> <p><i>'How to be a successful PhD student and transition to a great job'</i></p> <ul style="list-style-type: none"> • Marie desJardins (UMBC) • Mark Dredze (JHU) • Claudia Pearce (DoD) • Ian Soboroff (NIST) • Hanna Wallach (UMass)
1:50-3:10	<p>Oral presentations II</p> <ul style="list-style-type: none"> • <i>'Large-scale Semantic Parsing via Schema Matching and Lexicon Extension'</i>, Qingqing Cai and Alexander Yates (Temple) • <i>'Efficient First-Order Probabilistic Logic Programming for Natural Language Inference'</i>, William Yang Wang and William W. Cohen (CMU) • <i>'Semi-Markov Phrase-based Monolingual Alignment'</i>, Xuchen Yao, Benjamin Van Durme, Chris Callison-Burch, Peter Clark (JHU & UPenn & AI2) • <i>'Gathering and Generating Paraphrases from Twitter with Application to Normalization'</i>, Wei Xu, Alan Ritter and Ralph Grishman (NYU)
3:10-4:00	Poster session II
4:00-5:00	<p>Breakout Session I</p> <ul style="list-style-type: none"> • <i>'NLP in low resource settings'</i>, Ann Irvine (JHU)
4:00-5:00	<p>Breakout Session II</p> <ul style="list-style-type: none"> • <i>'Dynamic Programming: Theory and Practice'</i>, Alexander Rush (Columbia/MIT) and Nathaniel Wesley Filardo (JHU)
4:00-5:00	<p>Breakout Session III</p> <ul style="list-style-type: none"> • <i>'NELL: Never Ending Language Learning'</i>, Partha P Talukdar (CMU)
5:00	Closing
5:15 - 6:30	Wine down, Flat Tuesdays, UMBC Commons
6:00 - 8:00	Poetry Slam in the Performing Arts and Humanities Building Atrium

Poster Session I 11:20 - 12:10

#	Institution	Authors	Title
1	IITSL	Rivindu Perera and Udayangi Perera	Anaphora resolution through semantic distance weighting
2	UMD	Mohit Iyyer and Jordan Boyd-Graber	Applications of Deep Learning to Framing and Question Answering
3	UMD	Ning Gao and Douglas Oard	Resolving Person Names for First-Person Entity Linking
4	UMBC	Varish Mulwad	Semantic Message Passing for Generating Linked Data from Tables
5	UMBC	Lushan Han	Schema Free Querying of Semantic Data
6	JHU	Jonathan Wintrode and Sanjeev Khudanpur	Can you repeat that? Using word repetition to improve Spoken Term Detection
7	UMBC	Abhay Kashyap and Tim Finin	Top Down Hierarchical Clustering of Streaming Document Entities
8	UMD	Ke Zhai, Jordan Boyd-Graber and Shay Cohen	Online Adaptor Grammar with Hybrid Inference
9	UMD	John Morgan, Jordan Boyd-Graber and Alvin Grissom	Thinking On Your Feet: A Model Of Simultaneous Translation
10	JHU	Chris Paxton, Amy Babay and Danning He	Developing a Markov Decision Process to Model Medical Interventions in the ICU
11	UMD	Alvin Grissom and Jordan Boyd-Graber	Computational Verb Prediction in Japanese
12	UPenn	Hezekiah Akiva Bacovcin and Kobey Shwayder	A Theoretically Informed Morphological Parser
13	UMD	Dan Goldwasser and Hal Daume III	Modeling Latent Pragmatic Effects in Courtroom Dialogues
14	UMBC	Oleg Aulov, Adam Price and Milton Halem	Human Sensor Networks for Disaster Management
15	JHU	Matthew R. Gormley, Mark Dredze and Jason Eisner	Markov Random Field Parsing
16	GWU	Cheng Tang	Testing number of clusters for k-means clustering based on the Classification Likelihood Model
17	JHU	Anna M. Kruspe	Automatic Language Identification for Singing

Poster Session II 3:10 - 4:00

#	Institution	Authors	Title
18	IITSL	Rivindu Perera and Udayangi Perera	Recent advances in conceptual graph based knowledge representation
19	NYU	Wei Xu, Raphael Hoffmann, Le Zhao and Ralph Grishman	Filling Knowledge Base Gaps for Distant Supervision of Relation Extraction
20	UMD	Jyothi Vinjumur, Jiaul Piaik and Douglas Oard	Evaluating the Evaluation: Assessing the Utility of a Test Collection for Privilege in E-discovery
21	UMD	John Morgan, Ann Irvine, Hal Daume, Marine Carpuat and Dragos Munteanu	Measuring Machine Translation Errors in New Domains
22	UMD	Yuening Hu, Ke Zhai, Vladimir Eidelman and Jordan Boyd-Graber	Tree-based Topic Models for Translation Domain Adaptation
23	PSU	Siddhartha Banerjee and Prasenjit Mitra	Automatic Generation of Wikipedia Play Articles using Playscript Classification
24	UMD	Thang Nguyen, Peratham Wiriyathamabhum, Yuening Hu and Jordan Boyd-Graber	Spectral Algorithms for Topic Models: An Evaluation
25	UMD	Hal Daume III, Samir Khuller, Manish Purohit and Gregory Sanders	Kuhn meets Rosenblatt: Combinatorial Algorithms for Online Structured Prediction
26	UMBC	Jennifer Sleeman and Tim Finin	Type Prediction for Efficient Coreference Resolution in Heterogeneous Semantic Graphs
27	JHU	Nanyun Peng, Yiming Wang and Mark Dredze	Learning Polylingual Topic Models from Code-Switched Social Media Conversations
28	UMBC	Ben Johnson, Jesse English and Benjamin Bengfort	OpenWIMs: An Open Source Semantic Text Analyzer
29	UMBC	Morgan Madeira and Anupam Joshi	Analyzing Close Friend Interactions in Social Media
30	UMD	Snigdha Chaturvedi, Vittorio Castelli, Ramesh Nallapati, Hema Raghavan and Radu Florian	Joint Probabilistic Models for Relevance Prediction in Non-Factoid Question Answering
31	GHES	Angel Garrido, Sergio Ilarri and Eduardo Mena	Hypatia - Towards a Support Expert System for Documentation Departments

Panel 12:40-1:40

How to be a successful PhD student and transition to a great job

Marie desJardins is a full professor of computer science at the University of Maryland, Baltimore County, with research interests in artificial intelligence and computer science education. She has graduated 10 Ph.D. students and 25 M.S. students and has supervised 50 undergraduate research projects, and is frequently asked to participate in mentoring and outreach events.

Mark Dredze is an assistant research professor at Johns Hopkins University and a researcher in the Human Language Technology Center of Excellence. He already has a Ph.D. from the University of Pennsylvania, but thinks a lot about how he would do it all again.

Claudia Pearce received a PhD in CS in 1994 from UMBC. Her area of research is information retrieval and text processing systems. Over the years she has been a practicing researcher and a research manager in this area. In her current position, she develops technical health programs for computer scientists.

Ian Soboroff is the manager of the Retrieval Group at NIST, where he conducts research on measuring the effectiveness of search, NLP, and video systems. He received his Ph.D. from UMBC in January 2001.

Hanna Wallach is an assistant professor in the School of Computer Science at UMass Amherst and a core faculty member in UMass's newly-formed computational social science research initiative. In her not-so-spare time, Hanna (or Logistic Aggression, as she is better known) likes to put on roller skates and hit people really, really hard.

Breakout Sessions 4:00-5:00

NLP in low resource settings

Many state of the art approaches to NLP tasks assume that a large amount of data of a particular type is available to supervise the learning of statistical models. Examples include sentiment annotations for sentiment analysis, parallel sentences for machine translation, and transcriptions of spoken language for speech recognition. However, for many languages and domains, such data isn't available in large quantities, and we must adapt standard approaches accordingly. In this breakout discussion, we'll share ideas about how to approach NLP in low resource settings. Possible points for discussion include: (1) unsupervised and semi-supervised learning, (2) alternative types of data, (3) crowdsourcing, and (4) active learning.

Ann Irvine is a final year PhD student in the Computer Science Department and the Center for Language and Speech Processing at Johns Hopkins. Her advisor is Chris Callison-Burch, and her thesis topic is machine translation for low resource languages and domains.

Dynamic Programming: Theory and Practice

Dynamic programming is a crucial practical tool for implementing NLP systems, but unfortunately it can be very hard to get correct. It is all too common for students to start a project with vague notions of chart variables and pseudo-code for CKY, and end up with unsalvageable code. In this break-out session, we give an opinionated overview of alternative representations of dynamic programming as a logical system, as a hypergraph, and as linear optimization. We then discuss frameworks for implementing DP in practice and what possibilities may exist in the near-future. We end with a collaborative discussion of techniques researchers use in practice and what pragmatic problems should be addressed going forward.

Alexander Rush is a Ph.D. candidate in Computer Science at the Massachusetts Institute of Technology studying with Prof. Michael Collins and is currently a visiting scholar at Columbia University in New York. He received his A.B. in Computer Science from Harvard University in 2007. Before starting graduate study, he worked as lead engineer on the Platform/API team at Facebook. His research interest is in formally sound, but empirically fast inference methods for natural language processing, with a focus on models for syntactic parsing, translation, and speech. Last spring, he was the lead TA for NLP on Coursera, Columbia's first MOOC, with over 30,000 registered students. He has received best paper awards from EMNLP 2010 and NAACL 2012, the latter for work completed as an intern at Google research.

Nathaniel Wesley Filardo is a fourth-year Ph.D. student at the Johns Hopkins Center for Language and Speech Processing. He did his undergraduate work at Carnegie Mellon University, getting a degree in Physics and another in Computer Science. he is a graduate fellow of the Human Language Technology Center Of Excellence (HLTCOE).

NELL: Never Ending Language Learning

One of the great technical challenges in big data is to construct computer systems that learn continuously over years, from a continuing stream of diverse data, improving their competence at a variety of tasks, and becoming better learners over time. This discussion will describe Carnegie Mellon University's research to build a Never-Ending Language Learner (NELL) that runs 24 hours per day, forever, learning to read the web. Each day NELL extracts (reads) more facts from the web, and integrates these into its growing knowledge base of beliefs. Each day NELL also learns to read better than yesterday, enabling it to go back to the text it read yesterday, and extract more facts, more accurately, today. NELL has been running 24 hours/day for over three years now. The result so far is a collection of 50 million interconnected beliefs (e.g., `servedWith(coffee, applePie)`, `isA(applePie, bakedGood)`), that NELL is considering at different levels of confidence, along with hundreds of thousands of learned phrasings, morphological features, and web page structures that NELL has learned to use to extract beliefs from the web. Track NELL's progress at <http://rtw.ml.cmu.edu>

Partha Pratim Talukdar is a Postdoctoral Fellow in the Machine Learning Department at Carnegie Mellon University, working with Tom Mitchell on the Never Ending Language Learning (NELL) project. Partha received his PhD (2010) in CIS from the University of Pennsylvania, working under the supervisions of Fernando Pereira, Zack Ives, and Mark Liberman. Partha is broadly interested in Machine Learning, Natural Language Processing, Data Integration, and Cognitive Science. His dissertation introduced novel graph-based weakly-supervised methods for Information Extraction and Integration. His past industrial research affiliations include HP Labs, Google Research, and Microsoft Research. He is currently co-authoring a book on graph-based semi-supervised learning.

Notes

Notes

Notes