MMLS 2018 Booklet

Wednesday, June 6	
8 - 8:50	Breakfast.
8:50 - 9	Opening remarks.
9 - 9:50	**Plenary speaker:** Alice Zheng (Chair: Po-Ling Loh). *Metrics, Insights, and Lessons Learned from Operating ML Systems in Advertising.*
9:50 - 10:30	**Poster spotlights** (Chair: Suriya Gunasekar).
10:30 - 11:30	Coffee, posters.
11:30 - 12:10	**Invited talks** (chair: Matus Telgarsky). **11:30:** Wei Xu *Learning for Natural Language Understanding and Generation* **11:50:** Chinmay Hegde *Fast, provable algorithms for learning structured dictionaries and autoencoders*
12:10 - 1:30	Lunch.
1:30 - 2:50	Sponsored talks, Q&A.
2:50 - 3:30	**Invited talks** (chair: Suriya Gunasekar). **2:50:** Mesrob Ohannessian *One (categorical) distribution estimator for all dimensions* **3:10:** Alex Schwing *Stable Training of Generative Adversarial Nets via the Sliced Wasserstein Distance*
3:30 - 4:30	Coffee, posters.
4:30 - 5:20	**Plenary speaker:** Robert Schapire (Chair: Mike Franklin). *Sample-Efficient Reinforcement Learning with Rich Observations.*
5:30	Reception (sponsored by ServiceNow).
Thursday, June 7	
8 - 9	Breakfast.
9 - 9:50	**Plenary speaker:** Jon Kleinberg (Chair: Matus Telgarsky). *Inherent Trade-Offs in Algorithmic Fairness.*
9:50 - 10:30	**Invited talks** (chair: Po-Ling Loh). **9:50:** Evan Sparks *Large Scale Deep Learning in Practice* **10:10:** Yingyu Liang *Learning Mixtures of Linear Regressions with Nearly Optimal Complexity*
10:30 - 11:30	Coffee, posters.
11:30 - 12:30	**Invited talks** (chair: Suriya Gunasekar). **11:30:** Ermin Wei *A Fast Distributed Asynchronous Newton-Based Optimization Algorithm* **11:50:** Mladen Kolar *Estimation and Inference for Differential Networks* **12:10:** Jeff Naughton *Inverting OLAP; Text Classification and Text Creative Modeling.*
12:30 - 1:40	Lunch.
1:40 - 2:30	**Plenary speaker:** Patrick Wolfe (Chair: Mike Franklin). *Challenges and Opportunities for Data Science.*
2:30 - 3	Coffee.
3 - 4	Panel: MMLS advisory board.
4 - 4:10	Closing remarks.

- **General MMLS info:** visit `http://midwest-ml.org/2018/` .
- **Questions or concerns?** email `midwestml2018@gmail.com` .
- **Submit panel questions:** `https://app2.sli.do/event/rtaawtze/ask` .

We graciously thank our sponsors:

Platinum.

- 3M.
- American Family Insurance.
- College of Engineering, University of Illinois, Urbana-Champaign.
- Google Madison.
- Here.
- ServiceNow.
- Toyota Technological Institute at Chicago.

Gold.

- Department of Computer Science, University of Chicago.
- Grainger Institute, University of Wisconsin-Madison.
- Northwestern Mutual.

1 Plenary talks

- **Speaker:** Jon Kleinberg (Cornell University)

  Title: Inherent Trade-Offs in Algorithmic Fairness.

  Abstract: Recent discussion in the public sphere about classification by algorithms has involved tension between competing notions of what it means for such a classification to be fair to different groups. We consider several of the key fairness conditions that lie at the heart of these debates, and discuss recent research establishing inherent trade-offs between these conditions. We also consider a variety of methods for promoting fairness and related notions for classification and selection problems that involve sets rather than just individuals. This talk is based on joint work with Sendhil Mullainathan, Manish Raghavan, and Maithra Raghu.

  Bio: Jon Kleinberg is the Tisch University Professor in the Departments of Computer Science and Information Science at Cornell University. His research focuses on issues at the interface of algorithms, networks, and information, with an emphasis on the social and information networks that underpin the Web and other on-line media. He is a member of the National Academy of Sciences and the National Academy of Engineering, and the recipient of research fellowships from the MacArthur, Packard, Simons, and Sloan Foundations, as well as awards including the Harvey Prize, the Nevanlinna Prize, the SIGKDD Innovation Award, and the ACM Prize in Computing.

- **Speaker:** Robert Schapire (MSR NYC)

  Title: Sample-Efficient Reinforcement Learning with Rich Observations.

  Abstract: We study a version of reinforcement learning in which the agent must learn how to choose actions based on observations so as to maximize long-term reward. We focus especially on when the observations may be realistically rich, such as images, text documents, patient records, etc. We introduce a new algorithm for systematic exploration, in other words, for discovering through experimentation how best to choose actions. Along the way, we also propose a new measure called the Bellman rank which we argue captures the degree to which the learning problem exhibits underlying structure, and which can be favorably bounded in a number of previously studied cases. We show that the Bellman rank determines the statistical efficiency of our algorithm, which, although not computationally efficient, requires a number of samples that is polynomial in the Bellman rank as well as more standard parameters, but which is entirely independent of the size of the observation space.

  This work is joint with Nan Jiang, Akshay Krishnamurthy, Alekh Agarwal, and John Langford.

  Bio: Robert Schapire is a Principal Researcher at Microsoft Research in New York City. He received his PhD from MIT in 1991. After a short post-doc at Harvard, he joined the technical staff at AT&T Labs (formerly AT&T Bell Laboratories) in 1991. In 2002, he became a Professor of Computer Science at Princeton University. He joined Microsoft Research in 2014. His awards include the 1991 ACM Doctoral Dissertation Award, the 2003 Gdel Prize, and the 2004 Kanelakkis Theory and Practice Award (both of the last two with Yoav Freund). He is a fellow of the AAAI, and a member of both the National Academy of Engineering and the National Academy of Sciences. His main research interest is in theoretical and applied machine learning, with particular focus on boosting, online learning, game theory, and maximum entropy.

- **Speaker:** Patrick Wolfe (Purdue University)

  Title: Challenges and Opportunities for Data Science.

  Abstract: How do we draw sound and defensible conclusions from big data? This question lies at the heart of data science. In this talk I will first describe some of the challenges and opportunities inherent in this rapidly emerging field, and then discuss the current state of the art in one area of particular relevance to Sandia National Laboratories: big network data. Progress in this area includes the development of new large-sample theory that helps us to view and interpret networks as statistical data objects, along with the transformation of this theory into new statistical methods to model and draw inferences from network data in the real world. The insights that result from connecting theory to practice also feed back into pure mathematics and theoretical computer science, prompting new questions at the interface of combinatorics, analysis, probability, and algorithms.

  Bio: Patrick J. Wolfe (SM) received B.S.E.E. and B.Mus. degrees from the University of Illinois at Urbana-Champaign (1998) and his Ph.D. from the University of Cambridge (2003) as U.S. National Science Foundation Graduate Research Fellow. After teaching at Cambridge from 20012003, he joined the faculty of Harvard University (2004) and received the Presidential Early Career Award for Scientists and Engineers from the White House (2008). In 2012, he returned to the UK to take up an Established Career Fellowship in the Mathematical Sciences at University College London (UCL), where he also served as a Royal Society Research Fellow and as founding Executive Director of UCLs Big Data Institute. In 2017, he was appointed the Frederick L. Hovde Dean of Science at Purdue University.

  Dr. Wolfe is also a trustee and non-executive director of the Alan Turing Institute, the U.K.s National Institute for Data Science, and serves on the board of its commercial subsidiary. Previously the Institutes Deputy Director and recently named its first honorary fellow, he played a leading role in establishing the institute and shaping its priorities through an extensive program of engagement with a diverse range of experts and stakeholders. He has provided expert advice on applications of data science to policy, societal, and commercial challenges, including to the U.S. and U.K. governments

and to a range of public and private bodiesincluding most recently the U.K. Food Standards Agency as an inaugural member of its Science Council. Dr. Wolfe is currently Chair, IEEE SPS Big Data Special Interest Group and serves on the steering committee of the IEEE SPS Data Science Initiative, as well as Co-Chair, Data Science Section of the Institute for Mathematical Statistics.

Dr. Wolfe has received awards for his research from a number of international bodies, including the Royal Society, the Acoustical Society of America, and the IEEE. He is active in the global mathematics, statistics, and physical sciences communities, and most recently was an organizer and Simons Foundation Fellow at the Isaac Newton Institute for Mathematical Sciences 2016 semester research program on Theoretical Foundations for Statistical Network Analysis.

- **Speaker:** Alice Zheng (Amazon)

  Title: Metrics, Insights, and Lessons Learned from Operating ML Systems in Advertising.

  Abstract: The role of a machine learning researcher usually begins with inventing a novel algorithm to solve an important problem, and ends with a prototype and a paper. From an engineering perspective, however, this is just the beginning. The prototype will be turned into a robust system for production, and a team will monitor and operate the system throughout its life time. With probability 1, something will go wrong. What happens next are the three key steps of system operations: detection, mitigation, and root cause diagnosis. In this talk, we will discuss lessons learned from operating machine learning systems in programmatic advertising, including operational procedures and metrics that we've developed to guard against these near-certainty situations.

  Bio: Alice Zheng is the head of the Machine Learning Optimization team in Amazon advertising. She specializes in research and development of machine learning methods, tools, and applications. She recently co-authored an O'Reilly book on "Feature Engineering for Machine Learning." Previously, she worked at Turi/Dato/GraphLab, where she led the Toolkits team and helped with marketing and user education. She worked as a researcher at Microsoft Research, Redmond and as a postdoc at Carnegie Mellon University's Auton Lab and the Parallel Data Lab. Alice received B.A.s in Mathematics and Computer Science and a Ph.D. from U. C. Berkeley in Prof. Michael Jordan's lab.

2 Invited talks

- **Speaker:** Chinmay Hegde (Iowa State)

  Title: Fast, provable algorithms for learning structured dictionaries and autoencoders

  Abstract: Recent algorithmic advances have led to the emergence of provably accurate algorithms for learning a dictionary of atoms that represent a given dataset. However, these algorithms are beset with several challenges: high running time, large memory costs, and susceptibility to missing data. In this talk, we address these challenges by constructing a family of new algorithms where the dictionaries themselves obey conciseness assumptions (such as compositionality, sparsity and/or democracy). We also discuss implications of our algorithmic techniques for training (shallow) autoencoder architectures. Joint work with Thanh Nguyen, Raymond Wong, and Akshay Soni.

  Bio: Chinmay Hegde is an assistant professor, and Black and Veatch Faculty Fellow, in Electrical and Computer Engineering at Iowa State University. His research focuses on developing fast and robust algorithms for machine learning and statistical signal processing, with applications to imaging problems. Prior to this, he received his PhD in Electrical and Computer Engineering at Rice University, and was a Shell-MIT Postdoctoral Associate in CSAIL at the Massachusetts Institute of Technology. Chinmay is the recipient of multiple awards, including best paper awards at SPARS and ICML, the Budd Award for Best Engineering Thesis at Rice University in 2013, the Warren Boast Award for Undergraduate Teaching in 2016, the NSF CRII Award in 2016, and the NSF CAREER Award in 2018.

- **Speaker:** Mladen Kolar (UChicago)

  Title: Estimation and Inference for Differential Networks

  Abstract: We present a recent line of work on estimating differential networks and conducting statistical inference about parameters in a high-dimensional setting. First, we consider a Gaussian setting and show how to directly learn the difference between the graph structures. A debiasing procedure will be presented for construction of an asymptotically normal estimator of the difference. Next, building on the first part, we show how to learn the difference between two graphical models with latent variables. Linear convergence rate is established for an alternating gradient descent procedure with correct initialization. Simulation studies illustrate performance of the procedure. We also illustrate the procedure on an application in neuroscience. Finally, we will discuss how to do statistical inference on the differential networks when data are not Gaussian.

  Bio: Mladen Kolar is Assistant Professor of Econometrics and Statistics at the University of Chicago Booth School of Business. His research is focused on high-dimensional statistical methods, graphical models, varying-coefficient models and data mining, driven by the need to uncover interesting and scientifically meaningful structures from observational data. Particular applications arise in studies of dynamic regulatory networks and social media analysis. His research has appeared in several publications including the Journal of Machine Learning Research, Annals of Applied Statistics, and the

Electronic Journal of Statistics. He also regularly presents his research at the top machine learning conferences, including Advances in Neural Information Processing Systems and the International Conference of Machine Learning.

Kolar was awarded a prestigious Facebook Fellowship in 2010 for his work on machine learning and network models. He spent a summer with Facebooks ads optimization team working on a large scale system for click-through rate prediction. His other past research included work with INRIA Rocquencourt in Paris, France and Joint Research Center in Ispra, Italy.

Kolar earned his PhD in Machine Learning in 2013 from Carnegie Mellon University, as well as a diploma in Computer Engineering from the University of Zagreb. For his Ph.D. thesis work on Uncovering Structure in High-Dimensions: Networks and Multi-task Learning Problems, Kolar received from 2014 SIGKDD Dissertation Award honorable mention.

Outside of academia, Kolar enjoys chess, badminton, running, cycling and hiking.

- **Speaker:** Yingyu Liang (UW-Madison)

  Title: Learning Mixtures of Linear Regressions with Nearly Optimal Complexity

  Abstract: Mixtures of Linear Regressions (MLR) is an important mixture model with many applications. In this model, each observation is generated from one of the several unknown linear regression components, where the identity of the generated component is also unknown. Previous works either assume strong assumptions on the data distribution or have high complexity. This paper proposes a fixed parameter tractable algorithm for the problem under general conditions, which achieves global convergence and the sample complexity scales nearly linearly in the dimension. In particular, different from previous works that require the data to be from the standard Gaussian, the algorithm allows the data from Gaussians with different covariances. When the conditional number of the covariances and the number of components are fixed, the algorithm has nearly optimal sample complexity N = O(d) as well as nearly optimal computational complexity O(Nd), where d is the dimension of the data space. To the best of our knowledge, this approach provides the first such recovery guarantee for this general setting.

  Bio: Yingyu Liang is an assistant professor in the Department of Computer Sciences at the University of Wisconsin-Madison. His research interests are providing theoretical analysis for machine learning models and designing efficient algorithms with provable guarantees for applications. He received a B.S. in 2008 and an M.S. in 2010 in Computer Science from Tsinghua University, and a Ph.D. degree in Computer Science from Georgia Institute of Technology in 2014. He was a postdoctoral researcher in 2014-2017 in the Computer Science Department at Princeton University.

- **Speaker:** Jeff Naughton (Google)

  Title: Inverting OLAP; Text Classification and Text Creative Modeling.

  Abstract: Inverting OLAP is a popular paradigm for business data analysis. In this paradigm analysts specify measures and multi-dimensional points or sub-regions in a data set, and the system returns the value of the specified metrics in the specified region. In this talk we consider inverting this process - that is, rather than the analyst telling the system where to look, the system tells the analyst where to look. We will describe some problems that arise in this approach.

  Text creatives are the snippets of text that appear in search ads. We will briefly and at a high level discuss how text classification models can be used to understand user queries and text creatives and specifically how TF-HUB can be useful in text classification tasks with small amount of data to improve the relevance and quality of ads served.

  Bio: Jeff Naughton is a Principal Scientist and Site Lead at Google-Madison. Prior to February 2016 he was a professor with a research focus on database management systems at the University of Wisconsin-Madison, where he held the John P. Morgridge Chair of Computer Sciences. He received his PhD in Computer Science from Stanford University.Xi Wu is a software engineer at Google. He received his PhD in Computer Science from the University of Wisconsin-Madison. He received a Google PhD Fellowship in 2016 in the area of privacy and security.Kazoo Sone is a software engineer for Google. Since he joined in 2011, he has led several machine learning & natural language processing projects for AdWords. His current research interests are weakly supervised learning & transfer learning for text modeling. He earned his Ph.D in Aeronautics with minors in Applied Math & Computer Science from Caltech.

- **Speaker:** Mesrob Ohannessian (TTIC)

  Title: One (categorical) distribution estimator for all dimensions

  Abstract: Categorical models are a natural fit for many problems. When learning the distribution of categories from samples, high-dimensionality may dilute the data. Minimax optimality is too pessimistic to remedy this issue. A serendipitously discovered estimator, absolute discounting, corrects empirical frequencies by subtracting a constant from observed categories, which it then redistributes among the unobserved. It outperforms classical estimators empirically and has been used extensively in natural language modeling. We'll rigorously explain the prowess of this estimator using less pessimistic notions. We show that (1) absolute discounting recovers classical minimax KL-risk rates and that (2) it is *adaptive* to an effective dimension rather than the true dimension, thus the same estimator can be used in all dimensions. We explore the practical implications with an application to the Global Terrorism Database.

  Bio: Mesrob I. Ohannessian is a Research Assistant Professor at the Toyota Technological Institute at Chicago. He was previously a postdoc at UCSD, MSR-Inria, and Universit Paris-Sud. He received his PhD in EECS from MIT. He has a

passion for teaching at all levels. His research interests are in machine learning, statistics, information theory, and their applications, particularly to problems marked by data scarcity.

- **Speaker:** Alex Schwing (UIUC)

  Title: Stable Training of Generative Adversarial Nets via the Sliced Wasserstein Distance

  Abstract: Modeling of ambiguity in data is important for applications where a single correct result does not exist, such as image captioning, image inpainting and even machine translation. Impressive advances have been reported in recent years, demonstrating that probability distributions or sampling over complex domains can be learned by transforming classical distributions via deep nets, which are trained via a saddle-point formulation. Optimization of this generative adversarial net formulation is however challenging and many tricks have proven indispensable, some theoretically justified and others empirically validated. In this talk we discuss promising findings based on an adversarial net formulation which uses the sliced Wasserstein distance. It results in stable and fast training. We will show results of generators which sample images of size 1024 x 1024 and are directly trained from scratch, i.e., no sequential or hierarchical training process is necessary.

  Bio: Alex Schwing is an Assistant Professor at the University of Illinois at Urbana-Champaign working with talented students on computer vision and machine learning topics. He received his B.S. and diploma in Electrical Engineering and Information Technology from Technical University of Munich in 2006 and 2008 respectively, and obtained a PhD in Computer Science from ETH Zurich in 2014. His PhD thesis was awarded an ETH medal. Afterwards he joined University of Toronto as a postdoctoral fellow before joining UIUC in 2016. His research interests are in the area of computer vision and machine learning, where he has co-authored numerous papers on topics in scene understanding, inference and learning algorithms, deep learning, image and language processing and generative modeling.

- **Speaker:** Evan Sparks (Determined AI)

  Title: Large Scale Deep Learning in Practice

  Abstract: Deep learning has yielded a step function improvement at an array of important problems ranging from computer vision to natural language processing, and there is enormous excitement about its potential. However, building practical applications powered by deep learning remains an enormous challenge: the necessary expertise is scarce, the hardware requirements can be prohibitive, and current software tools are immature and limited in scope. In this talk, we will first describe how deep learning workflows are supported by existing software tooling. We will then describe several promising opportunities to drastically improve these workflows via novel algorithmic and software solutions, including automated hyperparameter optimization, efficient utilization of distributed resources via performance models, and reproducible workflow management. This talk draws on academic work done at CMU, Berkeley, and UCLA, as well as our experiences at Determined AI, a startup that provides software to make deep learning engineers dramatically more productive.

  Bio: Evan R. Sparks is co-founder and CEO of Determined AI. While earning his PhD in Computer Science in Berkeley's AMPLab, he contributed to the design and implementation of much of the large-scale machine learning ecosystem around Apache Spark, including MLlib and KeystoneML. Prior to Berkeley, Evan worked in quantitative finance and web intelligence. He also holds an AB in Computer Science from Dartmouth College.

- **Speaker:** Ermin Wei (Northwestern)

  Title: A Fast Distributed Asynchronous Newton-Based Optimization Algorithm

  Abstract: One of the important problems in the field of distributed optimization is the problem of minimizing a sum of local convex objective functions over a networked system. Most of existing work in this area focuses on developing distributed algorithms in a synchronous setting under the presence of a central clock, where the agents need to wait for the slowest one to finish the update, before proceeding to the next iterate. Asynchronous distributed algorithms remove the need for a central coordinator, reduce the synchronization wait, and allow some agents to compute faster and execute more iterations. In the asynchronous setting, the only known algorithms for solving this problem could achieve either linear or sublinear rate of convergence. In this work, we built upon the existing literature to develop and analyze an asynchronous Newton-based method to solve a penalized version of the problem. We show that this algorithm guarantees almost sure convergence with global linear and local quadratic rate in expectation. Numerical studies confirm superior performance of our algorithm against other asynchronous methods.

  This is a joint work with Fatemeh Mansoori.

  Bio: Ermin Wei is currently an Assistant Professor at the EECS Dept of Northwestern University. She completed her PhD studies in Electrical Engineering and Computer Science at MIT in 2014, advised by Professor Asu Ozdaglar, where she also obtained her M.S.. She received her undergraduate triple degree in Computer Engineering, Finance and Mathematics with a minor in German, from University of Maryland, College Park. Wei has received many awards, including the Graduate Women of Excellence Award, second place prize in Ernst A. Guillemen Thesis Award and Alpha Lambda Delta National Academic Honor Society Betty Jo Budson Fellowship. Wei's research interests include distributed optimization methods, convex optimization and analysis, smart grid, communication systems and energy networks and market economic analysis.

- **Speaker:** Wei Xu (OSU)

Title: Learning for Natural Language Understanding and Generation

Abstract: Human language is notoriously complex due to the multitude of ways people can express the same meaning (i.e. paraphrases). I will present our work on robust machine learning methods for large-scale paraphrasing, including 1) automatic paraphrase acquisition that exploited multi-instance learning and deep neural networks for semantics; and 2) utilizing paraphrases for various natural language generation tasks with machine translation techniques. I will also show how similar multi-instance learning models can learn large knowledge bases and resolve time expressions with limited labeled data.

Bio: Wei Xu is an assistant professor of Computer Science and Engineering at the Ohio State University. Her research lies at the intersections of machine learning, natural language processing, and social media. She received her PhD in Computer Science from New York University where she was a MacCracken fellow. Between 2014 and 2016, she was a postdoctoral researcher at the University of Pennsylvania. She recently received the NSF CRII Award, CrowdFlower AI for Everyone Award in addition to funding from DARPA.

3 Posters

3.1 Spotlights

1. **Author(s):** Cesar Uribe (UIUC), Soomin Lee, Alexander Gasnikov, Angelia Nedic

   Title: A dual approach for optimal algorithms in distributed optimization over networks

2. **Author(s):** Sathya Ravi (UW-Madison), Tuan Dinh, Vishnu Lokhande, Vikas Singh

   Title: Constrained deep learning using conditional gradient and applications in computer vision

3. **Author(s):** Daniel Vial (University of Michigan), Vijay Subramanian

   Title: Personalized PageRank dimensionality and algorithmic implications

4. **Author(s):** Mehrnaz Amjadi (University of Illinois at Chicago), Theja Tulabandhula

   Title: Block structure based models for graph sequences

5. **Author(s):** Yucheng Chen (UIUC), Bolton Bailey, Daniel Hsu, Jian Peng, Matus Telgarsky

   Title: Neural network solutions to optimal transport

6. **Author(s):** Niladri S. Chatterji (UC Berkeley), Xiang Cheng, Peter Bartlett, Michael Jordan

   Title: Underdamped Langevin MCMC

7. **Author(s):** Girish Joshi (UIUC), Girish Chowdhary

   Title: Cross-domain transfer in deep reinforcement learning using policy adaptation

8. **Author(s):** Pedro Savarese (TTIC), Michael Maire

   Title: Soft parameter sharing for deep neural networks

9. **Author(s):** Bowen Shi (TTIC), Aurora Martinez del Rio, Jon Michaux, Greg Shakhnarovich, Daine Brentari, Karen Livescu

   Title: Fingerspelling recognition in the wild

10. **Author(s):** Chaoyue Liu (Ohio State University), Mikhail Belkin

    Title: Parametrized accelerated methods free of condition number

3.2 Remaining posters

11. **Author(s):** Antigoni Georgiadou (Florida State University)

    Title: Simplifying the sky: Mitigating baryonic effects in cosmological weak lensing measurements

12. **Author(s):** Aniket Anand Deshmukh (University of Michigan), Srinagesh Sharma, Clayton Scott, James W. Cutler, Mark Moldwin

    Title: Contextual bandits for spacecraft magnetometer interference cancellation

13. **Author(s):** Shengchao Liu (UW-Madison), Thevaa Chandereng, Yingyu Liang

    Title: An order-invariant structure learning method for molecule classification

14. **Author(s):** Sen Na (University of Chicago), Mladen Kolar, Zhuoran Yang, Zhaoran Wang
    Title: Stein's estimator for single-index varying coefficient model

15. **Author(s):** Rahul Ghosal (North Carolina State University), Arnab Maity, Timothy Clark, Stefano Longo
    Title: Variable selection in functional linear concurrent regression

16. **Author(s):** Mireille Boutin (Purdue), Tarun Yellamraju
    Title: "Real" high-dimensional data clustering

17. **Author(s):** Sumeet Katariya (UW-Madison), Robert Nowak
    Title: Coarse ranking with smart clustering

18. **Author(s):** Debjyoti Saharoy (University of Illinois at Chicago), Theja Tulabandhula
    Title: A realistic online algorithm to learn buyer behavior

19. **Author(s):** Ashley Hou (UW-Madison), Po-Ling Loh
    Title: Robust budget allocation

20. **Author(s):** Yunyang Xiong (UW-Madison), Hyunwoo J. Kim, Vijayaraghavan Thiruvengadam, Kathikeyan Sankaralingam, Vikas Singh
    Title: Gaze estimation under severe power and energy constraints

21. **Author(s):** Cem Subakan (UIUC), Oluwasanmi Koyejo, Paris Smaragdis
    Title: Learning the base distribution in implicit generative models

22. **Author(s):** Xiaowu Dai (UW-Madison), Yuhua Zhu
    Title: On large-batch training for deep learning: A theoretical understanding of the generalization gap

23. **Author(s):** Lifu Tu (TTIC), Kevin Gimpel
    Title: Learning approximate inference networks for structured prediction

24. **Author(s):** Y. Samuel Wang (University of Chicago), Mathias Drton
    Title: Causal discovery with unobserved confounding and non-Gaussian data

25. **Author(s):** Byol Kim (University of Chicago), Song Liu, Mladen Kolar
    Title: Statistical inference for high-dimensional differential networks

26. **Author(s):** Joshua Whitman (UIUC), Girish Chowdhary
    Title: Learning dynamics across similar spatiotemporally evolving systems

27. **Author(s):** Hongsheng Liu (UNC Chapel Hill), Tianxiao Sun
    Title: An inexact subsampled proximal Newton-type method under generalized self-concordance

28. **Author(s):** Taylor V. Williams (Purdue), Tarun Yellamraju, Kerrie A. Douglas, Mireille Boutin
    Title: Application of the n-TARP binary clustering method to survey data

29. **Author(s):** David Waller (Purdue), Hillary Merzdorf, Nathan Hicks, Dan Goldwasser, Kerrie Douglas
    Title: Comprehensive validation of naive Bayes algorithm through qualitative methods

30. **Author(s):** Shripad Gade (UIUC), Nitin Vaidya
    Title: POLAR: Private Optimization and Learning Algorithm

31. **Author(s):** Hyebin Song (UW-Madison), Garvesh Raskutti
    Title: PULasso: High-dimensional variable selection with presence-only data

32. **Author(s):** Xiaomin Zhang (UW-Madison), Xuezhou Zhang, Xiaojin Zhu, Po-Ling Loh
    Title: Theoretical support of machine learning debugging via weighted M-estimation

33. **Author(s):** Duzhe Wang (UW-Madison), Po-Ling Loh
    Title: Robust estimation in high-dimensional heteroscedastic regression with graph trend filtering

34. **Author(s):** Belinda Tzen (UIUC), Tengyuan Liang, Maxim Raginsky
    Title: Local optimality and generalization guarantees for the Langevin algorithm via empirical metastability

35. **Author(s):** Runyu Mao (Purdue), Shaobo Fang, Sri Kalyan Yarlagadda, Fengqing Zhu
    Title: Automated noisy image removal for online food image collection

36. **Author(s):** Jonathan Eskreis-Winkler (University of Chicago), Risi Kondor
    Title: Multiresolution representations of graphs using random forests

37. **Author(s):** Glenn Fung (American Family Insurance), Devin Conathan, Sukrat Gupta, Luisa Polania
    Title: RoCKET: Robust Classification and Knowledge Extraction from Text

38. **Author(s):** Ming Yu (University of Chicago), Zhaoran Wang, Varun Gupta, Mladen Kolar
    Title: Recovery of simultaneous low rank and two-way sparse coefficient matrices, a nonconvex approach

39. **Author(s):** Ozge Surer (Northwestern University), Daniel Apley, Edward C. Malthouse
    Title: Coefficient tree regression for discovering hidden structure

40. **Author(s):** Nathan Hicks (Purdue), Hillary Merzdorf, David Waller, Gaurav Nanda, Dan Goldwasser, Kerrie Douglas
    Title: Comprehensive qualitative validation of topic modeling results for open-ended survey questions

41. **Author(s):** Sathya Ravi (UW-Madison), Ronak Mehta, Vikas Singh
    Title: Robust blind deconvolution via mirror descent

42. **Author(s):** Ankit Pensia (UW-Madison), Varun Jog, Po-Ling Loh
    Title: Generalization error bounds for noisy, iterative algorithms

43. **Author(s):** Siyuan Ma (Ohio State University), Raef Bassily, Mikhail Belkin
    Title: The power of interpolation: Understanding the effectiveness of SGD in modern over-parametrized learning

44. **Author(s):** Ashok Makkuva (UIUC), Sreeram Kannan, Pramod Viswanath
    Title: Globally consistent algorithms for mixture of experts

45. **Author(s):** Yuheng Bu (UIUC), Jiaxun Lu, Venugopal V. Veeravalli
    Title: Active and adaptive sequential learning

46. **Author(s):** Hao-Jun Shi (Northwestern University), Raghu Bollapragada, Dheevatsa Mudigere, Jorge Nocedal, Peter Ping Tak Tang
    Title: A progressive batching L-BFGS method for machine learning

47. **Author(s):** Zhihan Guo (UW-Madison), Theodoros Rekatsinas
    Title: Data profiling methods for interactive data cleaning

48. **Author(s):** Maya Lozinski (University of Chicago)
    Title: Properties of best subset selection and implications for practitioners

49. **Author(s):** Seong Jae Hwang (UW-Madison), Ronak Mehta, Vikas Singh
    Title: Sampling-free uncertainty estimation in gated recurrent units with exponential families

50. **Author(s):** Nasim Sonboli (DePaul University), Robin Burke
    Title: Fairness-aware recommendation systems

51. **Author(s):** Julian Katz-Samuels (University of Michigan), Clayton Scott
    Title: Feasible arm identification

52. **Author(s):** Pedro Savarese (TTIC), Nathan Srebro
    Title: Continuation methods for deep neural networks

53. **Author(s):** Harshal Maske (UIUC), Girish Chowdhary
    Title: Learning instructional policy from demonstration

54. **Author(s):** Hsu Kao (University of Michigan), Vijay Subramanian

    Title: Localization, approximations, and nonlinear consensus for distributed non-convex optimization

55. **Author(s):** Muni Sreenivas Pydi (UW-Madison), Varun Jog, Po-Ling Loh

    Title: Graph-based ascent algorithms for function maximization

56. **Author(s):** Yunyang Xiong (UW-Madison), Hao Zhou, Vishnu Lokhande, Vikas Singh

    Title: Domain adaptation with conditional distribution matching with applications

4 Sponsored talks

- **Speaker:** Jamal Afridi (3M)

  Title: AI in a Materials World

  Abstract: This talk will discuss the opportunities for Artificial Intelligence (AI) in a materials company, 3M. 3M has a diverse portfolio of products (50,000+), in markets ranging from medical informatics to worker safety to air quality and more. This diversity provides unique opportunities for Data Scientists to positively impact peoples lives in new ways through the integration of materials with AI. This talk will highlight some examples where 3M is leveraging AI to create smart products for the future.

  Bio: Jamal Afridi is a Data Scientist at 3M. Jamal's research interests lie at the intersection of computer vision, machine learning and information theory. Jamal has been invited to present his research at various national and international venues. In 2014, at the International Conference on Pattern Recognition (ICPR), his research paper was selected as one of the four nominees for the prestigious 'Best Industry Related Paper Award (BIRPA)'. Before joining 3M, Jamal completed his PhD in Computer Science at Michigan State University. His PhD thesis focused on Deep Learning algorithms and how they can be used for analyzing 3D MRI data. His research efforts helped secure NIH R01 grant. Jamal continues to review new research in machine learning and computer vision for several well-known conferences and for the journal Pattern Recognition.

- **Speaker:** Debu Chatterjee (ServiceNow)

  Title: Enterprise AI and the importance of Text

  Abstract: ServiceNow was founded on a very simple idea: that work should be easier. That getting simple stuff done shouldnt be so hard and complex stuff should be manageable. It started with ITcreating a System of Action to streamline and automate unstructured work, eliminating the back and forth emails, phone calls, and manual processes that waste time, money, and sap productivity. Today, an entire enterpriseHR, customer service, security, and beyondcan tap into the power of the Now Platform to create a better experience for employees, users, and customers, and transform the way work is done.

  There is history. Each enterprise has their own unique history. Much of the history is textual information emails, notes, FAQs, knowledge bases, communities, and now chats and conversations. Each enterprise has its own understanding of textual information being exchanged. Better we get to know the enterprise language, easier it becomes to make work easy for our customers. To transform the way work is done, resolve issues faster and increase agent efficiency, ServiceNow uses machine learning to automatically categorize, route, and prioritize issues primarily from text. It learns from patterns in your historical data, becoming increasingly accurate in its predictive recommendations. Operationalizing this learning and applying it scalably out of the box for each of our customer one at a time is what we define as Enterprise AI.

  Bio: Debu Chatterjee heads the Machine Learning, ArtificiaI Intelligence, and Analytics Product Engineering Organisation of ServiceNow. Debu joined ServiceNow in Jan 2017 when his company DxContinuum was acquired by ServiceNow to seed ServiceNow intelligent automation roadmap. The fully integrated DxContinuum Fathom engine in the NOW platform now powers the Agent Intelligence product of ServiceNow. He led the acquisition of Parlo.io a Natural Language Understanding engine to deliver conversational AI agents and enterprise language modeling.

  Earlier, in 2012, he founded a business decisions company DxContinuum specializing in operationalizing predictive business outcomes. As the CEO of the company, he brought the transformative power of predictive analytics to B2B sales and marketing operations for a roster of customers like Adobe, Cisco and VMWare. The patented, machine- learning software platform accelerated the model generation and delivery for a variety of supervised machine learning use cases.

  Prior, Debu has demonstrated successful technology leadership in diverse settings and scale. He was a key member of Oracle's flagship product, the database from version 6 to version 11 in various capacities. Every SQL access in Oracle goes through the code he has written. He led Informatica's PowerCenter and Connectors group delivering the flagship V9 release, and subsequently the Health Care Fraud Analytics team at FICO. His enterprise software development experience spans 30 years in the areas of machine learning, predictive analytics, metadata services, distributed systems, databases, big data, systems management and fraud detection.

He has a BS in CS from IIT Kharagpur, a MS in Computer Science from UNC Chapel Hill, and an MBA from Wharton. He holds 23 patents, and has filed numerous others. He is a technical advisory member at Benhamou Global Ventures, and a member of BOV and GEAB at UNC Chapel Hill. He also contributes his time to various Silicon Valley professional and mentorship organizations.

- **Speaker:** Ian Endres (HERE)

  Title: Wrangling location data at HERE Technologies

  Abstract: Map making on a global scale is a monumental undertaking. It requires extensive data collection, feature detection, aggregation of tens of thousands of heterogeneous sources, validation, and continuous monitoring to adapt to changes in real-time. Machine learning research at HERE Technologies is not only focused on solving these core map making problems, but also how to leverage this location data to provide services across many different industries.

  Bio: Ian Endres is a Lead Research Engineer at HERE Technologies. His research focuses on applying computer vision and machine learning to feature detection from a range of sensors, and automatically deriving detailed geospatial maps from those features. He received his BS and PhD in Computer Science from University of Illinois at Urbana-Champaign.

- **Speaker:** Derrick Higgins (American Family Insurance)

  Title: Data Science and Machine Learning at American Family Insurance

  Abstract: American Family increasingly uses machine learning and artificial intelligence to drive its insurance business, but is also cultivating out new data sources and methods to develop entirely new business models related to financial services, property and risk. This talk will outline our portfolio of work and showcase projects of particular interest to machine learning practitioners.

  Bio: Derrick Higgins is an R&D strategist, manager, data scientist and computational linguist; since 2016, he has led American Familys DSAL data science team in Chicago. Prior to joining American Family, Dr. Higgins was lead data scientist at Civis Analytics, and used deep learning to uncover latent factors in political discussions on social media. Before that, he was the director of NLP and speech research at the Educational Testing Service, where he and his team developed tools for analyzing student responses that are now used in leading testing programs around the world, including the GRE and TOEFL. Dr. Higgins earned a Ph.D. in Linguistics from the University of Chicago in 2002. He has contributed to research in many fields of natural language processing and educational measurement, including semantic representation, discourse structure analysis, item generation, off-topic essay identification, and the automated scoring of spoken responses. His research has been published in leading conferences and journals in the fields of computational linguistics, speech processing, and language testing, and has resulted in ten patents.