David M. Andrzejewski

Recent professional experience

Industry

2018 to present — **Senior Engineering Manager, Data Insights**,
Sumo Logic, Redwood City, CA.

2015 - 2018 — **Engineering Manager, Unified Logs and Metrics**,
Sumo Logic, Redwood City, CA.
Team responsible for new time-series data analysis platform

2014 - 2015 — **Tech Lead / Manager, Data Sciences Engineering**,
Sumo Logic, Redwood City, CA.
Lead the development of advanced analytics for machine-generated log data

2013 - 2014 — **Lead Data Sciences Engineer**,
Sumo Logic, Redwood City, CA.

2011 - 2013 — **Data Sciences Engineer**,
Sumo Logic, Mountain View, CA.

2010 - 2011 — **Postdoctoral Research Staff Member**,
Lawrence Livermore National Laboratory, Livermore, CA.
Applied statistical modeling to knowledge discovery in text corpora

2008 — **Research Intern**, *Microsoft Research*, Redmond, WA.
Developed analysis tool for investigating system performance anomalies

Academic

2008 - 2010 — **Research Assistant (Professors Mark Craven and Xiaojin Zhu)**,
UW–Madison, Madison, WI.
Knowledge-augmented topic models
Developed new latent topic models to allow prior knowledge and user feedback
Proposed, implemented, and conducted experiments on new models and techniques

Education

2007–2010 — **PhD**, *University of Wisconsin–Madison*, Madison, WI.
Computer Sciences
Research focus: Machine Learning
Advisors: Mark Craven and Xiaojin Zhu
Thesis: *Incorporating Domain Knowledge in Latent Topic Models*

2005–2007 — **MS**, *University of Wisconsin–Madison*, Madison, WI.
Computer Sciences

2000–2005 — **BS**, *University of Wisconsin–Madison*, Madison, WI.
Computer Engineering, Mathematics, Computer Sciences

Selected publications

David Andrzejewski and David Buttler. Latent topic feedback for information retrieval. In *KDD '11: Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, 2011. (8% of submissions accepted for oral presentation).

David Andrzejewski, Xiaojin Zhu, Mark Craven, and Benjamin Recht. A framework for incorporating general domain knowledge into latent Dirichlet allocation using first-order logic. In *IJCAI '11: Proceedings of the 22nd International Joint Conference on Artificial Intelligence*. AAAI Press, 2011. (17% of submissions accepted).

David Andrzejewski, Xiaojin Zhu, and Mark Craven. Incorporating domain knowledge into topic modeling via Dirichlet forest priors. In *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*, pages 25–32. Association for Computing Machinery, 2009. (25% of submissions accepted).

David Andrzejewski, Anne Mulhern, Ben Liblit, and Xiaojin Zhu. Statistical debugging using latent topic models. In *ECML '07: Proceedings of the 18th European conference on Machine Learning*, pages 6–17. Springer-Verlag, 2007. (9% of submissions accepted).

Selected technical talks

- Understanding Software System Behavior With ML and Time Series Data. QCon.ai, San Francisco (April 2018)
- Functional Programming for Machine Learning (panel). Scale By the Bay, San Francisco (November 2017)
- Economical machine learning via functional programming. Big Data Scala by the Bay, Oakland (August 2015)
- Graph mining for log data. Strata + Hadoop World, San Jose (February 2015)
- Machine learning for machine data. Strata Conference, Santa Clara (February 2014)
- Latent Topic Feedback for Information Retrieval. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (August 2011)
- A Framework for Incorporating General Domain Knowledge into Latent Dirichlet Allocation using First-Order Logic. International Joint Conference on Artificial Intelligence (July 2011)
- Incorporating domain knowledge into topic modeling via Dirichlet forest priors. International Conference on Machine Learning (June 2009)
- Statistical debugging using latent topic models. European Conference on Machine Learning (September 2007)

Patents and applications

- **System and method of drug identification through radio frequency identification (RFID)**
  United States Patent Application (11/465993)
  Ronald Makin, Kyle Jansson, Silas Zirn, David Andrzejewski,Timothy Flink
- **Visualization tool for system tracing infrastructure events**
  United States Patent Application (US8464221)
  Alice X. Zheng, Trishul A. Chilimbi, Shuo-Hsien Hsiao, Danyel A. Fisher, David M. Andrzejewski