

David M. Andrzejewski

Recent professional experience

Industry

- 2019 to present **Director of Engineering, AI Experiences,**
Sumo Logic, Redwood City, CA.
- 2011 - 2019 **Various engineering roles,**
Sumo Logic, Mountain View and Redwood City, CA.
- 2010 - 2011 **Postdoctoral Research Staff Member,**
Lawrence Livermore National Laboratory, Livermore, CA.
Applied statistical modeling to knowledge discovery in text corpora

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

- Reliable machine learning. Scale By the Bay, Oakland (2019)
- Privacy-aware data science in Scala with monads and type level programming. Scale By the Bay, SF (2018)
- Understanding Software System Behavior With ML and Time Series Data. QCon.ai, SF (2018)
- Functional Programming for ML (panel). Scale By the Bay, SF (2017)
- Economical ML via functional programming. Big Data Scala by the Bay, Oakland (2015)
- Graph mining for log data. Strata + Hadoop World, San Jose (2015)
- Machine learning for machine data. Strata Conference, Santa Clara (2014)
- Latent Topic Feedback for Information Retrieval. ACM SIGKDD Conference on Knowledge Discovery and Data Mining (2011)
- A Framework for Incorporating General Domain Knowledge into Latent Dirichlet Allocation using First-Order Logic. IJCAI, Barcelona (2011)
- Incorporating domain knowledge into topic modeling via Dirichlet forest priors. ICML, Montreal (2009)
- Statistical debugging using latent topic models. ECML, Warsaw (2007)

Patents and applications

- **Clustering of structured log data by key schema**
United States Patent (11321158)
Udit Saxena, Reetika Roy, Ryley Higa, David M. Andrzejewski, Bashyam TCA
- **Clustering of structured log data by key-values**
United States Patent (11663066)
Udit Saxena, Reetika Roy, Ryley Higa, David M. Andrzejewski, Bashyam TCA
- **Cardinality of time series**
United States Patent (11182434)
Christian Friedrich Beedgen, David M. Andrzejewski, Weijia Che
- **Anomaly detection**
United States Patent (10445311B1)
Kumar Saurabh, David M. Andrzejewski, Yuchen Zhao, Christian Friedrich Beedgen, Bruno Kurtic
- **Data enrichment and augmentation**
United States Patent (11397726)
Christian Friedrich Beedgen, David M Andrzejewski, Benjamin Everette Newton, Kumar Avijit, Stefan Christoph Zier
- **Logs to metrics synthesis**
United States Patent (11042534)
Christian Friedrich Beedgen, David M Andrzejewski, Benjamin Everette Newton, Kumar Avijit, Stefan Christoph Zier

- **Key name synthesis**
United States Patent (11481383)
Christian Friedrich Beedgen, David M. Andrzejewski
- **Visualization tool for system tracing infrastructure events**
United States Patent (8464221)
Alice X. Zheng, Trishul A. Chilimbi, Shuo-Hsien Hsiao, Danyel A. Fisher, David M. Andrzejewski
- **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