Timothy N. Rubin, PhD

Data Scientist and Cognitive Scientist
(215) 990-4012
Tim.Rubin@gmail.com
Personal Website | Github Profile

WORK AND RESEARCH EXPERIENCE

SENIOR DATA SCIENTIST: CHANGE HEALTHCARE: 2018-PRESENT

- Leveraged machine learning and NLP models for analyzing and classifying medical record data.
- Develop named entity recognition models for automatically extracting lab measurements from OCR'd medical records.

DATA SCIENTIST: SURVEYMONKEY 2016-2018

- Developed all algorithms underlying SurveyMonkey's *SurveyMonkey Genius* platform—a user-facing ML product that provides customers with predictions about their survey (e.g., estimated completion times) as well as personalized recommendations for how to improve their survey. SurveyMonkey Genius has received industry press as part of SurveyMonkey's 2017 company rebrand, and had significant financial and brand impact.
- Developed a new SurveyMonkey use-case ontology and automated classification model. Applications for this product include sales assistance and driving a variety of personalization opportunities for users.

POSTDOCTORAL RESEARCH SCIENTIST: INDIANA UNIVERSITY 2013 – 2016

- Led and collaborated on research projects leading to numerous publications in top-tier journals and conferences
- Developed and implemented novel algorithms for identifying functional brain regions (Python and MATLAB)
- Secured a $65,000 grant for studying linguistic features related to schizophrenia
- Developed a computational modeling framework for empirically evaluating semantic models
- Developed an improved prediction method for Latent Dirichlet Allocation models

DATA CONSULTANT: UNIVERSITY OF WASHINGTON 2013

- Performed statistical analyses and hypothesis tests on previously collected mental health data
- Applied unsupervised learning algorithms for interpreting and summarizing a corpus of open-ended questionnaire responses collected in clinical settings

GRADUATE STUDENT RESEARCHER: UNIVERSITY OF CALIFORNIA, IRVINE 2006 - 2012

- Developed and implemented novel probabilistic topic models that achieved state-of-the art performance on multi-labeled document classification (MATLAB and C)
- Developed and implemented a novel algorithm for movie recommendations using Netflix data (MATLAB)
- Developed a novel variant of Latent Dirichlet Allocation applicable to graph hierarchies (MATLAB and Python)
- Teaching assistant for 8 semesters. Ran discussion and laboratory sections for undergraduate classes

RELEVANT SKILLS

Analytical Skills:

- Machine learning and pattern recognition; Natural language processing; Experimental design; Data mining; Implementation and development of novel modeling tools; Probability theory and statistics

Programming languages:

- Python; SQL; Spark; Java (some); C++ (some)

Statistical analysis software:

- MATLAB; R; SPSS; Excel; BUGS

Communication Skills:

- Public speaking; Teaching; Writing and presentation of research, for both technical and non-technical audiences

RESEARCH GRANTS AWARDED

2014-2015: "Building Statistical Language Processing Algorithms for the Automated Coding of Semi-Structured Interview Data in Clinical Schizophrenia." Indiana University Collaborative Research Grants fund (IUCRG). Funding acceptance rate: <20%. Role: Co-PI. (PI: Michael N. Jones). $65,434

EDUCATION

University of California, Irvine — Irvine, CA
Ph.D., Department of Cognitive Sciences — 2012
M.A., Department of Cognitive Sciences — 2009

Tufts University — Medford, MA
B.S. Psychology, *Cum Laude* — May 2004
Minor in Cognitive Science

SELECTED PUBLICATIONS

Papanikolaou, Y., **Rubin, T.N.**, Tsoumakas, G. (2017) Dense Distributions from Sparse Samples: Improved Gibbs Sampling Parameter Estimators for LDA. *Journal of Machine Learning Research (JMLR)*.

Rubin, T.N., Koyejo, O., Jones, M.N., Yarkoni, Y., (2016). Generalized Correspondence-LDA Models (GC-LDA) for Identifying Functional Regions in the Brain. *30th Annual Conference on Neural Information Processing Systems (NIPS)*.

Rubin, T.N., Kojeyo, O., Gorgolewski, K.J., Jones, M.N., Poldrack, R.A., Yarkoni, T. (2016) Decoding brain activity using a large-scale probabilistic functional-anatomical atlas of human cognition. *PLOS Computational Biology*.

Gruenenfelder, T.M., Recchia, G., **Rubin, T.N.**, Jones, M.N. (2015). Graph-Theoretic Properties of Networks Based on Word Association Norms: Implications for Models of Lexical Semantic Memory, *Cognitive Science*.

Rubin, T.N., Kievit-Kylar, B., Willits, J.A., Jones, M.N., (2014). Organizing the Space and Behavior of Semantic Models, *36th Annual Conference of the Cognitive Science Society*.

Rubin, T.N., Chambers, A., Smyth, P., Steyvers, M., (2012). Statistical Topic Models for Multi-Label Document Classification, *Machine Learning: special issue on Learning from Multi-Label Data*.

Rubin, T.N., Steyvers, M., (2009). A Topic Model For Movie Choices and Ratings, *9th International Conference on Cognitive Modeling (ICCM)*, (Supplementary Material)