Timothy N. Rubin, PhD

Team Lead / Senior Data Scientist
(215) 990-4012
Tim.Rubin@gmail.com
Personal Website | Github Profile

WORK AND RESEARCH EXPERIENCE

DATA SCIENCE TEAM LEAD / SENIOR DATA SCIENTIST: CHANGE HEALTHCARE: 2018-PRESENT

- Lead a cross-functional team of data scientists and engineers on multiple healthcare-related ML products
- Contributed to a framework for building, evaluating, and deploying ML pipelines
- Developed NLP models for automatically interpreting doctor and radiologist reports
- Assisted on development of in-house active-learning solution for efficiently collecting and labeling text data

SENIOR 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.

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 (a) automatically identifying functional brain regions using GC-LDA (b) empirically evaluating semantic models, and (c) improving prediction methods for Latent Dirichlet Allocation models
- Secured a $65,000 grant for studying linguistic features related to schizophrenia

DATA SCIENCE 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)
- Teaching assistant for 8 semesters. Ran discussion and laboratory sections for undergraduate classes

RELEVANT SKILLS

Programming languages:
- Python; SQL; HIVE, Spark; Java (some); C++ (some)

Statistical analysis software:
- MATLAB; R; SPSS; Excel; BUGS

Analytical Skills:
- Machine learning; Natural language processing; Experimental research and design, Probability theory and statistics

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

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 / Cognitive Science, *Cum Laude* — May 2004

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., 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)