Tyler KepTrine

11 Walley Street, Boston MA 02128 | (315)-350-1812 | trine.tyler@gmail.com | ttrine | tyler-trine

Summary

Experienced data scientist, software engineer, and technical team lead. 7+ years experience applying advanced statistical modeling to complex problems. Proven excellence in clearly communicating complex technical concepts, including key contributions to a grant proposal selected for a $5.8 million award.

Education

University of Rochester Rochester NY
B.S. Data Science *September 2012 - May 2016*

- Rigorous education in both computer science and statistical learning theory
- Selected coursework
  - *Data Science Practicum.* Open source replication of AlphaGo, a Go AI built on CNNs
  - *Database Systems.* SQL, data models, DDLs, ACID, normalization, NoSQL databases
  - *Computer Science Courses.* Computer Programming, Data Structures, Algorithms
  - *Math Courses.* Probability, Mathematical Statistics, Linear Algebra, Discrete Math

Technical Skills

Machine learning Supervised training, testing, and validation; unsupervised density estimation, probabilistic graphical models, recommender systems, neural attention, diffusion
Deep learning PyTorch, transformers, CNNs, distributed GPU deployment, AWS cloud training

---

Software engineering Git workflows, Docker, CI/CD, test-driven development, object-oriented programming
Data engineering PostgreSQL, NoSQL databases, ETL pipelines, data warehousing, REST APIs
Programming languages Python, TypeScript, R, Java, Bash

Experience

Kateri Carbon
Data Scientist and Technical Team Lead *October 2023 - Present*

Founding member of technical staff and manager of the entire technical division.

- Managed and mentored a small technical team
- Implemented cross-team agile development practices, including a task board, CI/CD, test-driven development, and code reviews
- Independently designed and implemented all cloud infrastructure and data pipelines for the organization
- Created end-to-end automated mapping software, speeding the manual mapping process by over 50x and greatly reducing the burden on the environmental science team
- Perfomed several advanced geospatial analyses leveraging both raster and vector data

Independent Contractor
Data Scientist *August 2023 - October 2023*

Biogeochemical model validation and geospatial data ETL.

- Worked with environmental scientist to parameterize and validate a biogeochemical model
- Prototyped end-to-end geospatial ETL pipeline

Nifty Island Austin TX
Lead Data Scientist *May 2022 - June 2023*

Data lifecycle, modeling, analysis, and visualization on blockchain and in-game player data.

Key Achievements

- Lead cross-team effort to curate high-quality user data aligned across core services
- Clearly communicated complex technical concepts in simple terms, including their business impact
- Designed, developed, and implemented a novel recommender system with powerful contextual modeling capabilities using a neural attention mechanism

Systems and Technology Research Woburn MA
Senior Researcher *July 2017 - May 2022*

Data science R&D on challenging national defense problems.

Responsibilities

- Prototype, tune, productionize, and deploy a wide variety of statistical models to automate complex tasks
- Design, implement, test, dockerize, and deploy efficient, massive-scale data processing algorithms
- Technical communication, including presenting results to stakeholders and writing grant proposals

Key Achievements

- Proposed Bayesian approach to automated software assurance. Proposal selected for $5.8 million award
- Designed and implemented fast probabilistic solution to NP-hard subgraph isomorphism problem
- Designed and implemented matrix factorization model of individual behaviors in online groups

1010data New York NY
Data Scientist *July 2016 - July 2017*

Software development and data analysis over a distributed NoSQL database for billion-record datasets.

Responsibilities

- Optimize database queries for distributed execution
- Write custom database access management software
- Develop linear models of sales as a function of ad exposure

Personal projects

The Nature Conservancy Fisheries Monitoring *December 2016 - April 2017*

- Implemented novel CNN to classify fish species. Regularized via batch normalization, data augmentation
- Trained models on cloud GPUs, evaluated and analyzed results in R

AlphaGo replication *January 2016 - May 2016*

- Analyzed AlphaGo paper and construct good-faith implementation with a group of graduate students
- Implemented convolutional policy and value networks in Tensorflow

Selected Projects

Group Dynamics Modeling, Forecasting, and Validation Systems and Technology Research
Dr. Kirill Trapeznikov *September 2017 - August 2020*

- *Research goal.* Investigate behavior of indivduals in group contexts on social media websites. Develop predictive models of group membership.
- *Role.* Primary technical contributor. Independently read papers, propose research directions, preprocess data, implement, train, and tune models, evaluate and present results.
- *Approach.* Inspired by recommender systems, formulate problem as matrix factorization. Learn latent factor matrices pointwise. To regularize, constrain model to learn distribution of associated unstructured data. Engineer pipeline to yield reproducible experiments despite nondeterminic preprocessing. Bayesian optimization efficiently tunes highly sensitive hyperparameters.
- *Results.* Ongoing. Models reliably outperform baselines. Regularization with unstructured data boosts performance overall, but unevenly across users.

Synchronized Plans and Analytics for aiR-Cyber-Space (SPARCS) Systems and Technology Research
Mr. Nick Pioch *September 2017 - September 2018*

- *Research goal.* Database alignment. Build model to automatically align schemas from disparate databases.
- *Role.* Lead analyst. Implemented, trained models both collaboratively and independently.
- *Approach.* Draw ideas from knowledge base completion literature. Train neural network to infer facts about entities at the ontological (schema) level.
- *Results.* Schemas of sufficient size required to obtain reasonable results. Model correctly infers unobserved ontological facts when trained against large schemas with some explicit overlap.

AlphaGo Replication DSC 531 Data Science Practicuum
Dr. Henry Kautz *January 2016 - May 2016*

- *Research goal.* Build open-source implementation of AlphaGo.
- *Role.* Sole undergraduate student. Independently built initial prototype of the four deep nets in AlphaGo system. Implemented Markov Chain Monte Carlo (MCMC) rollouts with a partner. Helped lead group discussions to resolve apparent implementation inconsistencies.
- *Approach.* Study paper closely, understand implementation in sufficient detail to reconstruct it.
- *Results.* Many individual components finished (models, rollouts, game simulator, etc). Overall implementation incomplete due to insufficient time and compute resources. Broadly useful effort nevertheless; Github repository forked several thousand times, still under active development.

Online Interpretations of Scalar Adjectives BCS 206 Undergraduate Research in Cognitive Science
Dr. Chigusa Kurumada *September 2015 - May 2016*

- *Research question.* In discourse, people integrate context with speaker utterances to resolve their meaning. Two opposing theories about the mechanism behind this process: One holds that people rely on fixed rules to accomplish contextual integration, the other that the mechanism flexibly adapts to specific situations.
- *Role.* Team member, programmer. Independently analyzed experimental data, from raw eyetracking time series to hypothesis testing. Presented results with team.
- *Approach.* Conceptual replication of The Effect of Speaker-Specific Information on Pragmatic Inferences (Grodner and Sedivy, 2011). Eyetracking study with 40 subjects. Control group hears instructions from normal speaker, experimental group from speaker who overcontextualizes. Hypothesis: Control group integrates context, while experimental group learns to disregard it.
- *Results.* Hypothesis confirmed, supporting theory of adaptable contextual integration. Possible flaw discovered in reference experiment; they did not report interaction significance for a within-subject factor.

Grant Proposals

Bayesian ML for Assurance Case Evaluation in Complex Systems Systems and Technology Research
Dr. Steven Jilcott *July 2018 - May 2022*

- *Description.* Bayesian machine learning framework to automate safety assurance processes for complex, modular systems.
- *Role.* Contributed many of the main ideas of the proposal.
- *Outcome.* Won 5.8 million dollar contract.