Vaidas Simkus

vsimkus.github.io
github.com/vsimkus
simkus.vaidotas@gmail.com
+44 7784 132166

Profile

I enjoy tackling complex data challenges with generative machine learning methods. During my PhD, I addressed the key issue of missing data in real-world domains, improving the reliability and applicability of machine learning. My experience spans diverse roles at Disney, Avature, and J.P. Morgan, where I consistently delivered effective solutions and demonstrated strong adaptability and problem-solving skills. I am now eager to apply my probabilistic and generative ML skills to tackle key challenges in applied sciences and healthcare.

Education

PhD in Machine Learning, University of Edinburgh, UK Mar 2020 - Aug 2024
Thesis: *"Deep unsupervised machine learning in the presence of missing data"*.
Advisors: Dr. Michael Gutmann and Prof. Chris Williams.

MSc in Artificial Intelligence, Distinction, University of Edinburgh, UK Sep 2018 - Aug 2019
Received the Howe Masters Prize for the *best overall student performance* in the cohort.

BEng in Software Engineering, First-class, University of Southampton Sep 2013 - Jul 2016

Work experience

Machine Learning Consultant, Freelance, Remote Jul 2024 - *Present*
Consulting pharma and research organisations on design of efficient experiments for healthcare applications.

- Increased yield of chemical synthesis for drug production with transfer learning for Bayesian optimisation.
- Set up a shared compute environment for flexible experimentation using AWS SageMaker Script Mode.

Doctoral Researcher at the School of Informatics, University of Edinburgh Mar 2020 - *Present*
Identified limitations of popular machine learning methods in the presence of missing data. These insights led to new methods that increase the applicability of machine learning to domains affected by missing data.

- Proposed the *first general-purpose estimation method* for deep statistical models from incomplete data.
- Discovered and explained *why fitting VAEs from incomplete data is difficult*, and proposed variational-mixture approaches that mitigate these issues.
- Identified limitations of existing approaches for *conditional sampling of VAEs*, and developed two new approaches, based on adaptive MCMC and importance sampling, that mitigate the discovered limitations.
- In-progress paper on fast state-of-the-art missing data imputation for time-series and tabular data.
- Worked as a teaching assistant, tutor, and exam marker for postgraduate courses in machine learning.

Research Scientist Intern at Disney Research, Switzerland Sep 2022 - Dec 2022
Accelerated expensive rendering processes via machine learning and neural network inference optimisation.

- Accelerated Monte Carlo-rendered (ray-traced) video denoising by >20% via neural network pruning.
- Developed a pruning library for TensorFlow/Keras models, to optimise ML models for the Disney studios.

Natural Language Processing Engineer at Avature, Spain Dec 2019 - Feb 2020
Built ML solutions for resume parsing, interpretable semantic search, and job–candidate recommendation.

- Developed a method based on bag-of-words for learning skill embeddings from parsed job descriptions.
- Developed a character-level BiLSTM model for predicting skill-based embeddings directly from job titles.
- Implemented an efficient semantic search and a bi-directional recommender for candidates and jobs.

Software Engineer at J.P. Morgan Chase, United Kingdom Aug 2016 - Jul 2018
Developed greenfield projects and existing business-critical software systems for banking and enterprise.

- Developed a robust and low-latency middleware system for real-time payment processing and routing.
- Built an internal service-desk automation with a chat-bot interface, reducing new support tickets by 16%.
- Implemented a tool for orchestrating employee desktop environment migrations of >200K employees.

Skills summary

Probabilistic Models	Variational Autoencoders, Normalizing Flows, Diffusion Models, Flow Matching, Hidden Markov Models, Graphical Models
Probabilistic Inference	Variational Inference, Markov Chain Monte Carlo, Importance Sampling
Model Training	Maximum Likelihood Estimation, Bayesian Inference, Stochastic Optimisation
Other Core ML skills	Bayesian Optimisation, Bayesian Experimental Design, Missing Data Imputation, Deep Learning
Programming Languages	Python, Java, Bash, SQL, C/C++
ML Toolset	PyTorch, PyTorch Lightning, TensorFlow/Keras, NumPy, Pandas, Scikit, Slurm
Other	git, Test-Driven Development, CI/CD, AWS, SageMaker

Publications

Improving Variational Autoencoder Estimation from Incomplete Data with Mixture Variational Families 2024
Vaidotas Simkus and Michael Gutmann. *Transactions on Machine Learning Research (**TMLR**).*
Also presented at Data-centric Machine Learning Research (DMLR) Workshop at **ICLR** 2024.
Paper: openreview.net/forum?id=lLVmIvZfry. Code: github.com/vsimkus/demiss-vae.

Conditional Sampling of Variational Autoencoders via Iterated Approximate Ancestral Sampling 2023
Vaidotas Simkus and Michael Gutmann. *Transactions on Machine Learning Research (**TMLR**).*
Paper: openreview.net/forum?id=I5sJ6PU6JN. Code: github.com/vsimkus/vae-conditional-sampling.

Variational Gibbs Inference for Statistical Model Estimation from Incomplete Data 2023
Vaidotas Simkus, Benjamin Rhodes, and Michael Gutmann. *Journal of Machine Learning Research (**JMLR**).*
Also presented at **NeurIPS** 2023: neurips.cc/virtual/2023/poster/73921.
Paper: jmlr.org/papers/v24/21-1373.html. Code and demo: github.com/vsimkus/variational-gibbs-inference.

Learning Job Titles Similarity from Noisy Skill Labels 2022
Rabih Zbib, Lucas Lacasa Alvarez, Federico Retyk, Rus Poves, Juan Aizpuru, Hermenegildo Fabregat, *Vaidotas Simkus*, and Emilia García-Casademont. *International workshop on Fair, Effective And Sustainable Talent management using data science (FEAST) at **ECML-PKDD**.*
Paper: arxiv.org/abs/2207.00494. Original dataset: github.com/rabihzbib/jobtitlesimilarity_dataset.

Reviewing experience

Transactions on Machine Learning Research (TMLR) 2023 - *Present*

Teaching experience (postgraduate-level courses)

Head Teaching Assistant and Marker in Probabilistic Modelling and Reasoning, U. of Ed. 2022 - 2024
Lab Demonstrator in Data Mining and Exploration, University of Edinburgh 2021
Tutor and Marker in Machine Learning and Pattern Recognition, University of Edinburgh 2020

Open-source contribution

I believe in the power of open-source software to empower the disadvantaged and reduce technological inequality. Therefore, to give back my share I try to make my code open-source and contribute to OSS projects.

- Personal projects are open sourced at github.com/vsimkus.
- Contribution to large OSS projects:
  - *PyTorch*: An optimised library for deep learning.
  - *TensorFlow-Probability*: An optimised library for probabilistic machine learning and inference.
  - *Joplin*: A cross-platform note-taking app based on markdown.