Michael Boratko

11 Rumford Street – West Hartford, CT 06107 – USA

+1 (860) 294 9393 • boratko@gmail.com • www.mboratko.com

michaelboratko • mboratko

I am a machine learning researcher, currently a postdoc in the Information Extraction and Synthesis Laboratory at the University of Massachusetts Amherst, under the direction of Andrew McCallum. My recent work is at the intersection of **representation learning** and **probabilistic modeling** with applications to **natural language processing** and **structured prediction**. I care about improving foundational elements of machine learning to develop new solutions to real problems of practical importance with a large impact.

Education

2018-Present **Postdoctoral Research**, *University of Massachusetts Amherst*.
Measure-Theoretic Set Representation Learning with Box Embeddings
Advisor: Andrew McCallum
Research Areas: Representation Learning, Probabilistic Modeling, Natural Language Processing
- Project coordinator for IBM AI Horizons Network Grant, DARPA Machine Common Sense Grant
- Co-authored NSF CISE RI Medium Grant (https://www.nsf.gov/awardsearch/showAward?AWD_ID=2106391)
- Served in an advisory role for 5 PhD students.
- Supervised research projects for 25+ masters students with IBM, Amazon, and Meta.
- Ran independent study courses on reasoning over knowledge representations, language modeling, and graph representation.

2012–2018 **Ph.D. Mathematics**, *University of Massachusetts Amherst*.
On the Growth of Sobolev Norms for the Nonlinear Schrödinger Equation on Tori and Boundary Unique Continuation for Elliptic PDE
Advisors: Andrea Nahmod and Nestor Guillen
Research Areas: Variational Analysis, Harmonic Analysis, Geometric Partial Differential Equations
Selected Coursework: Differential Equations, Real and Complex Analysis, Measure Theoretic Probability, Stochastic Processes, Topology and Differential Geometry, Cryptography

2015–2018 **MS Computer Science**, *University of Massachusetts Amherst*.
Selected Coursework: Machine Learning, Deep Learning, Artificial Intelligence, Information Theory, Data Science, Combinatorics and Graph Theory, Programming Languages, Theory of Computation

2008–2010 **BA Mathematics, Economics**, *Central Connecticut State University*, (GPA: 4.0/4.0).
Selected Coursework: Mathematical Statistics, Econometrics, Game Theory, Philosophy of Logic

Vocational Experience

2012–2018 **Graduate Teaching Assistant**, *University of Massachusetts Amherst*.
As part of my teaching fellowship, I taught, graded, and held office hours for undergraduate classes in Mathematics and Computer Science.
- Teaching: Calculus I, II, III and Differential Equations
- Teaching Assistant: Artificial Intelligence (undergraduate and graduate level)

2002–Present **Owner**, *Starstreak (www.starstreak.com)*, West Hartford, CT.

Established company specializing in IT Consulting and Website Development.

- Designed servers and networks from the ground up, as well as leveraging cloud-based solutions.
  - Server Technologies: Linux (Ubuntu, Debian, and Arch), TrueNAS, and Docker.
  - Network Technologies: VPNs and VLANs on Ubiquiti, pfSense and custom iptables.
- Full-stack web development:
  - Clients include Thomson Reuters, Sensitech (Carrier), and Nexamp (Mitsubishi).
  - Leveraged open-source frameworks including FastAPI, Flask, Django, Drupal, and WordPress.
  - Developed extensions and fully custom web-based applications for specific niche industries.
  - Setup managed Linux hosting on bare-metal, VPS, and cloud (AWS, Azure) infrastructure
- Hired and managed subcontractors.
  - Created internal wiki for systems and procedures to expedite onboarding.
  - Developed custom invoice and ticket system to integrate work orders for multiple contractors.

Military

2003–Present **Musician**, *Connecticut Army National Guard*, Rockville, CT.

Staff Sergeant (E6) in the 102nd Army National Guard Band.
Play several instruments, arrange music, and provide supervision and mentorship as a squad leader.
Awarded Army Commendation Medal (ARCOM) and 4 Army Achievement Medals (AAM).

Awards

2022 ICML Outstanding Reviewer (Top 10%)

2021 NeurIPS Outstanding Reviewer (Top 8%)

2018 Distinguished Teaching Award from Department of Mathematics and Statistics, University of Massachusetts Amherst

2010 Departmental Award for Economics, Central Connecticut State University

2004 Fifth Place in TopCoder competition at Fairfield University / General Electric

Computer Skills

Languages Python, Julia, Go, Rust, OCaml, C
Web HTML, Javascript, CSS, PHP, SQL
Libraries PyTorch, TensorFlow
Frameworks Django, Flask, FastAPI, Vue, React
Tools Git, Travis CI, Docker, Weights & Biases, Poetry, NPM, Webpack, Slurm

Software

Dec, 2021 **Geometric Graph Embeddings**: a Python library for training representations which can model directed graphs, including Euclidean, Complex, and Hyperbolic Vectors, Euclidean, Hyperbolic, and Probabilistic Cones, as well as Gumbel and t-Box models.

July, 2021 **Box Embeddings**: a Python library for geometric representation learning compatible with PyTorch and TensorFlow. This provides box embeddings as a module, following current recommended practices for training and numerical stability, with comprehensive documentation.

Jan, 2021 **ProtoQA Evaluator**: extensible evaluation framework implimenting all three similarity measures (exact match, WordNet similarity, and RoBERTa) for the **ProtoQA Dataset**, a question answering dataset for prototypical commonsense reasoning.

Publications

Conferences

[1] Dongxu Zhang, **Michael Boratko**, Cameron Musco, and Andrew McCallum. "Modeling Transitivity and Cyclicity in Directed Graphs via Binary Code Box Embeddings". *NeurIPS* (2022).

[2] Shib Sankar Dasgupta*, **Michael Boratko***, Siddhartha Mishra, Shriya Atmakuri, Dhruvesh Patel, Xiang Lorraine Li, and Andrew McCallum. "Word2Box: Capturing Set-Theoretic Semantics of Words Using Box Embeddings". *ACL* (2022).

[3] Dhruvesh Patel, Pavitra Dangati, Jay-Yoon Lee, **Michael Boratko**, and Andrew McCallum. "Modeling Label Space Interactions in Multi-label Classification Using Box Embeddings". *ICLR* (2022).

[4] Siddhartha Mishra, Nicholas Monath, **Michael Boratko**, Ari Kobren, and Andrew McCallum. "An Evaluative Measure of Clustering Methods Incorporating Hyperparameter Sensitivity". *AAAI* (2022).

[5] **Michael Boratko***, Dongxu Zhang*, Nicholas Monath, Luke Vilnis, Kenneth L. Clarkson, and Andrew McCallum. "Capacity and Bias of Learned Geometric Embeddings for Directed Graphs". *NeurIPS* (2021).

[6] Tejas Chheda, Purujit Goyal, Trang Tran, Dhruvesh Patel, **Michael Boratko**, Shib Sankar Dasgupta, and Andrew McCallum. "Box Embeddings: An Open-Source Library for Representation Learning Using Geometric Structures". *EMNLP Demo* (2021).

[7] **Michael Boratko**, Javier Burroni, Shib Sankar Dasgupta, and Andrew McCallum. "Min/Max Stability and Box Distributions". *UAI* (2021). **(Long Presentation, $\frac{48}{777} \approx 6\%$)**.

[8] Yasumasa Onoe, **Michael Boratko**, Andrew McCallum, and Greg Durrett. "Modeling Fine-Grained Entity Types with Box Embeddings". *ACL* (2021).

[9] Xuelu Chen*, **Michael Boratko***, Muhao Chen, Shib Sankar Dasgupta, Xiang Lorraine Li, and Andrew McCallum. "Probabilistic Box Embeddings for Uncertain Knowledge Graph Reasoning". *NAACL* (2021).

[10] Shib Sankar Dasgupta*, **Michael Boratko***, Dongxu Zhang, Luke Vilnis, Xiang Li, and Andrew McCallum. "Improving Local Identifiability in Probabilistic Box Embeddings". *NeurIPS* (2020).

[11] **Michael Boratko***, Xiang Lorraine Li*, Tim O'Gorman*, Rajarshi Das*, Dan Le, and Andrew McCallum. "ProtoQA: A Question Answering Dataset for Prototypical Common-Sense Reasoning". *EMNLP* (2020).

[12] Dhruvesh Patel, Shib Sankar Dasgupta, **Michael Boratko**, Xiang Li, Luke Vilnis, and Andrew McCallum. "Representing Joint Hierarchies with Box Embeddings". *AKBC* (2020).

[13] Xiang Li, Luke Vilnis, Dongxu Zhang, **Michael Boratko**, and Andrew McCallum. "Smoothing the Geometry of Probabilistic Box Embeddings". *ICLR* (2019). **(Oral, $\frac{24}{1591} \approx 1.5\%$)**.

[14] **Michael Boratko**, Harshit Padigela, Divyendra Mikkilineni, Pritish Yuvraj, Rajarshi Das, Andrew McCallum, Maria Chang, Achille Fokoue, Pavan Kapanipathi, Nicholas Mattei, Ryan Musa, Kartik Talamadupula, and Michael J. Witbrock. "An Interface for Annotating Science Questions". *EMNLP Demo* (2018).

Workshops

[15] Shib Sankar Dasgupta, Xiang Lorraine Li, **Michael Boratko**, Dongxu Zhang, and Andrew McCallum. "Box-To-Box Transformations for Modeling Joint Hierarchies". *RepL4NLP Workshop at ACL* (2021).

[16] **Michael Boratko**, Harshit Padigela, Divyendra Mikkilineni, Pritish Yuvraj, Rajarshi Das, Andrew McCallum, Maria Chang, Achille Fokoue, Pavan Kapanipathi, Nicholas Mattei, Ryan Musa, Kartik Talamadupula, and Michael J. Witbrock. "A Systematic Classification of Knowledge, Reasoning, and Context within the ARC Dataset". *MRQA Workshop at EMNLP* (2018). **(Best Paper Award, Oral Presentation)**.

Reviewing

Journals	TMLR (2022)
Conferences	ICML 2021, 2022 (Outstanding Reviewer Award - Top 10%)
	NeurIPS 2021 (Outstanding Reviewer Award - Top 8%)
Workshops	AAAI 2019 Workshop on Reasoning for Complext Question Answering
	NeurIPS 2019 Workshop on Sets and Partitions

Mentorship

Ph.D. Mentorship

- Xiang Lorraine Li
- Dongxu Zhang
- Shib Sankar Dasgupta
- Dhruvesh Patel
- Wenlong Zhao

Industry Research Projects

2019 **Complex Question Answering**.
Masters Students: Hang Liu, Keshav Set, Wei Xie, Yi-Pei Chen
IBM: Achille Fokoue, Maria Chang, Katik Talamadupula, Pavan Kapanipathi, Ryan Musa

2020 **Stock Fundamentals Prediction Using 10-K Filings**.
Masters Students: Deepanshu Pariyani, Kajal Tiwari, Nikita Agarwal
Voya Financial: Rajessh Ramachandran; Lexalytics: Paul Barba

2021 **Knowledge Distillation for Autoregressive Generational Models**.
Masters Students: Harold Rubio, Hieu Phan
Amazon Alexa AI: Haidar Khan, Varun Kumar

2022 **Learning Visual Representations Through Communication**.
Masters Students: Rushikesh Dudhat, Haowen Yu, Fenil Doshi, Vengal Rao Guttha
FAIR at Meta: Marco Baroni, Roberto Dessi

Research Projects

2019 **Probabilistic Box Embeddings for Entailment**.
Dhruvesh Patel (now Ph.D. student at UMass Amherst)

2020 **Knowledge Representation and Reasoning**.
EunJeong Hwang (now Ph.D. student at the University of British Columbia)
Ankita Naik (now Ph.D. student at UMass Amherst)

2021 **Box Embeddings and Large Language Models for Hypernym Representation**.
Wenlong Zhao (now Ph.D. student at UMass Amherst)

2021 **PyTorch and Tensorflow Library for Box Representation Learning**.
Purujit Goyal (now Software Developer for OpenReview)
Tejas Chheda (completing MS at UMass Amherst)
Trang Tran (Data Scientist at MassMutual)

2021 **Evaluation of Commonsense Knowledge**.
Pranay Kumar Yelugam (completing MS at UMass Amherst)
Melnita Dabre (now Software Engineer at Meta)
Nalini Singh (now Software Developer at Meta)

2021 **Word Representations with Box Embeddings**.
Shriya Atmakuri (completing MS at UMass Amherst)

2021 **Evaluating Hyperparameter Sensitivity of Clustering Methods**.
Siddhartha Mishra (completing MS at UMass Amherst)

Personal

US Citizen, Eagle Scout