

# Ryan R. Curtin

439 Calhoun St. NW  
Atlanta, GA 30318  
443.534.0378  
ryan@ratml.org

## OBJECTIVE

---

Understand algorithms. Make them faster.

## EDUCATION

---

- Ph.D in Electrical and Computer Engineering** August 2015  
Georgia Institute of Technology, Atlanta, GA  
Thesis: **“Improving Dual-Tree Algorithms”**
- Master of Science in Electrical and Computer Engineering** May 2009  
Georgia Institute of Technology, Atlanta, GA
- Bachelor of Science with Highest Honors in Electrical Engineering** May 2008  
Georgia Institute of Technology, Atlanta, GA

## RESEARCH OVERVIEW

---

Numerous machine learning problems are bottlenecked by computation; the training process for neural networks, SVMs, HMMs, or nearest neighbor models are just a few instances. One prominent example is the training of neural networks: this can take days, even on modern hardware, and even with massive computational resources. For some classes of problems—those that can be expressed as pairwise problems (including nearest neighbor search and some related problems)—trees can provide a fast approximate solution. My current interests lie in exploring how hierarchical data structures can be used to accelerate current machine learning problems, such as the training of deep neural networks.

## PUBLICATIONS

---

- [0] **“Automatic batch size selection for fast RNN training and evaluation”**. R.R. Curtin, K. Kenemer. *In preparation*.
- [1] **“Exploiting the structure of furthest neighbor search for fast approximate results”**. R.R. Curtin, J. Echaüz, A.B. Gardner. *Information Systems*, 2018.
- [2] **“gmm\_diag and gmm\_full: C++ classes for multi-threaded Gaussian mixture models and Expectation-Maximisation”**. C. Sanderson, R.R. Curtin. *Journal of Open Source Software*, vol. 2, 2017.
- [3] **“An Open Source C++ Implementation of Multi-Threaded Gaussian Mixture Models, k-Means and Expectation Maximisation”**. C. Sanderson, R.R. Curtin. In *Proceedings of the 11th International Conference on Signal Processing and Communication Systems (ICSPCS '17)*, p. 1–8, 2017.
- [4] **“Designing and building the mlpack open-source machine learning library”**. R.R. Curtin, M. Edel. *Submitted to The Fourth International Conference of PUST (ICOPUST 2017)*.
- [5] **“pfsuper: simulated-based prognostics to monitor and predict sparse time series”**. J. Echaüz, A.B. Gardner, R.R. Curtin, N. Vasiloglou, G.J. Vachtsevanos. In *Proceedings of the Annual Conference of*

*the Prognostics and Health Management Society 2017 (PHM '17)*, p. 1–9, 2017.

- [6] “**Detecting adversarial samples from artifacts**”. R. Feinman, R.R. Curtin, S. Shintre, A.B. Gardner. *arXiv preprint arXiv:1703.00410*, 2017.
- [7] “**A dual-tree algorithm for fast  $k$ -means clustering with large  $k$** ”. R.R. Curtin. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, p. 300-308, 2017.
- [8] “**Fast approximate furthest neighbors with data-dependent candidate selection**”. R.R. Curtin, A.B. Gardner. In *Similarity Search and Applications 2016 (SISAP 2016)*, p. 221–235, 2016. *Nominated for Best Paper award, invited for journal submission.*
- [9] “**Armadillo: a template-based C++ library for linear algebra**”. C. Sanderson, R.R. Curtin. *Journal of Open Source Software*, vol. 1:26, p. 1–2, 2016.
- [10] “**Improving dual-tree algorithms**”. R.R. Curtin. Ph.D. thesis, Georgia Tech, 2015.
- [11] “**Faster dual-tree traversal for nearest neighbor search**”. R.R. Curtin. In *Similarity Search and Applications 2015 (SISAP 2015)*, p. 77–89, 2015.
- [12] “**Single-tree GMM training**”. R.R. Curtin. *Technical report GT-CSE-2015-01*, Georgia Institute of Technology, School of Computational Science and Engineering, 2015.
- [13] “**Collaborative filtering via matrix decomposition in mlpack**”. S. Agrawal, R.R. Curtin, S. Ghaisas, M.R. Gupta. In *Workshop on Machine Learning Open Source Software 2015*, Lille, France, 2015.
- [14] “**Plug-and-play runtime analysis for dual-tree algorithms**”. R.R. Curtin, D. Lee, W.B. March, P. Ram. *Journal of Machine Learning Research*, vol. 16, p. 3269–3297, 2015.
- [15] “**An automatic benchmarking system**”. M. Edel, A. Soni, R.R. Curtin. In *NIPS 2014 Workshop on Software Engineering for Machine Learning*, Montreal, Canada, 2014.
- [16] “**Classifying broiler chicken condition using audio data**”. R.R. Curtin, W. Daley, D.V. Anderson. *GlobalSIP 2014 Symposium on Signal Processing Applications Related to Animal Environments*, 2014.
- [17] “**Dual-tree fast exact max-kernel search**”. R.R. Curtin, P. Ram. *Statistical Analysis and Data Mining*, vol. 7, issue 4, p. 229–253, 2014.
- [18] “**Tree-independent dual-tree algorithms**”. R.R. Curtin, W.B. March, P. Ram, D.V. Anderson, A.G. Gray, C.L. Isbell, Jr. In *Proceedings of the 30th International Conference on Machine Learning (ICML '13)*, Atlanta, Georgia, 2013.
- [19] “**mlpack: a scalable C++ machine learning library**”. R.R. Curtin, J.R. Cline, N.P. Slagle, W.B. March, P. Ram, N.A. Mehta, A.G. Gray. *Journal of Machine Learning Research*, vol. 14, p. 801–805, 2013.
- [20] “**Fast exact max-kernel search**”. R.R. Curtin, P. Ram, A.G. Gray. In *SIAM International Conference on Data Mining (SDM '13)*, p. 1–9, Austin, Texas, 2013. *Nominated for Best Paper.*
- [21] “**mlpack: a scalable C++ machine learning library**”. R.R. Curtin, J.R. Cline, N.P. Slagle, M.L. Amidon, A.G. Gray. In *NIPS 2011 Workshop on Big Learning*, Granada, Spain, 2011.
- [22] “**Learning distances to improve phoneme classification**”. R.R. Curtin, N. Vasiloglou, D.V. Anderson. In *Proceedings of the 2011 IEEE International Workshop on Machine Learning in Signal Processing (MLSP 2011)*, Beijing, China, 2011.

## PROFESSIONAL EXPERIENCE

---

**Symantec Corporation**, Atlanta, GA

Fall 2015 – present

Center for Advanced Machine Learning *Principal Research Scientist*

My responsibilities at Symantec fall into roughly three categories:

- Pursue a research programme loosely focused on Symantec-relevant applications such as malware classification and related tasks
- Continue work as lead developer of mlpack (<http://www.mlpack.org>), a C++ machine learning library
- Apply machine learning approaches to internal Symantec problems, or help other internal Symantec groups improve their machine learning approaches

**Georgia Institute of Technology**, Atlanta GA

Spring 2010 – Fall 2015

*Graduate Research Assistant*

At various times I worked for these four labs:

FASTLAB (<http://www.fast-lab.org>)

Cooperative Analog and Digital Signal Processing Group

The pfunk research group

HPC Garage

advisor Dr. Alexander G. Gray

advisor Dr. David V. Anderson

advisor Dr. Charles L. Isbell, Jr.

advisor Dr. Rich W. Vuduc

I was/am also the primary developer and maintainer for mlpack (<http://www.mlpack.org>), an open-source scalable C++ machine learning library that is in use by scientists worldwide, currently with over 40000 downloads and 60 contributors.

I have also been involved as a TA or guest lecturer for multiple courses and groups:

- **ECE4001** (Professional Practice), Fall 2008: primary TA
- **CSE6740** (Machine Learning I), Fall 2012: guest lecture on the mlpack machine learning library
- **CSE6730** (Modeling and Simulation: Fundamentals and Implementation), Spring 2014: primary TA
- **ECE4811** (Computational humor), Spring 2014: guest lecture on collaborative filtering for recommendation systems
- **CS2804** (C++ Programming), Spring 2014: guest lectures on template metaprogramming
- **Linux Users Group at Georgia Tech**, Fall 2006–Fall 2015: frequent lecturer on various topics related to Linux, C++, or open-source software

**Compuglobalhypermegamet, L.L.C.**, Atlanta, GA

Spring 2013 – present

*CEO/Founder*

Machine learning consulting.

**Google, Inc.**, Mountain View, CA

Summer 2010

*Software Engineering Intern*

Worked with the Similar Pages team to provide improved search results.

**Georgia Tech Research Institute**, Atlanta, GA

Fall 2009 – Spring 2010

Food Processing Technology Division

*Graduate Research Assistant*

Applied machine learning techniques for stress detection in broiler chickens.

**Georgia Tech Research Institute**, Atlanta, GA  
ELSYS Lab

Spring 2009 – Fall 2009

*Graduate Research Assistant*

Investigated techniques for the A-to-D frontend of a radar warning receiver.

**Nexidia, Inc.**, Buckhead, GA

Summer 2007

*Research Intern*

Created voice synthesizers that can generate missing samples and still be comprehensible.

## SKILLS

---

- Extensive knowledge of Linux and related UNIX-like systems (as well as Windows)
- Good understanding of and experience with 1930s automotive technology
- Extremely comfortable with C and C++ as well as a plethora of other languages and design paradigms
- Experience with distributed, multicore, and GPU technologies such as MPI, OpenMP, OpenCL, CUDA, and others
- Basic machining knowledge: lathes, mills, drill presses, routers, saws, etc.
- Knowledgeable with state-of-the-art machine learning techniques for classification, regression, density estimation, and other similar tasks
- Experienced with hand-optimizing programs for substantial runtime improvement
- Amateur metallurgist
- Conversant in circuit design and physical implementation
- Skilled woodworker
- Nationally-known indoor kart racer (multiple national-level wins)

## PROFESSIONAL SERVICE

---

- *President*, Linux Users Group at Georgia Tech (2006-2011)
- *Treasurer*, Eta Kappa Nu, Beta Mu chapter (2007-2009)
- Reviewer for *The Journal of Machine Learning Research*, *WACV 2017*, *MLOSS 2015*

## ADVISING AND MENTORING

---

Through both Google Summer of Code and the labs I have worked for, I have advised and mentored a number of students.

- **5** Masters students from 2010 to 2014
- **6** undergraduate students from 2010 to 2015
- **14** Summer of Code students from 2013 to 2018

**References available upon request.**