Dr. Gregory J. Hamerly

Last update: 2020/01/14

Baylor University
Department of Computer Science
School of Engineering and Computer Science
One Bear Place #97141
Waco, TX 76798-7141, USA

Phone: +1-254-710-6846
Fax: +1-254-710-3889
Email: hamerly@cs.baylor.edu
http://cs.baylor.edu/~hamerly

Education

- Doctor of Philosophy in Computer Science   June, 2003
  University of California, San Diego
  Advisor: Charles Elkan
  Dissertation: "Learning structure and concepts in data through data clustering"
- Master of Science in Computer Science   June, 2001
  University of California, San Diego
- Bachelor of Science in Computer Science (magna cum laude)   June, 1999
  California Polytechnic State University, San Luis Obispo

Academic employment history

- Associate Professor   September, 2010 - present
  Department of Computer Science, Baylor University
  Graduate program director (2011 - 2017)
- Assistant Professor   September, 2004 - 2010
  Department of Computer Science, Baylor University
- Postdoctoral research appointment   September, 2003 – July, 2004
  Department of Computer Science, Katholieke Universiteit Leuven (Belgium)
- Lecturer   Summer 2002
  Department of Computer Science, California Polytechnic University, San Luis Obispo

Professional employment history

- Developer – Scrool AB   2013 - present
- Machine learning researcher and software developer – Mohomine, Inc.   2000 - 2001
- Software developer intern – Netscape Communications Corporation   Summer 1997
- Software developer intern – DigitalStyle Corporation   Summer 1996
- Website programmer – ElectriCiti Corporation   Summer 1995

Scholarly activity

Refereed publications (Journal and Proceedings)

1. Micheal C. Munson, Devon L. Plewman, Katelyn M. Baumer, Ryan Henning, Collin T. Zahler, Alexander T. Kietzman, Alexandra A. Beard, Shizuo Mukai, Lisa Diller, Greg Hamerly, Bryan F. Shaw. **Autonomous early detection of eye disease in childhood photographs**. In Science Advances, 2019.

2. Petr Ryšavý and **Greg Hamerly**. **Geometric methods to accelerate k-means algorithms**. SIAM Conference on Data Mining (SDM), 2016.

3. Petr Ryšavý and **Greg Hamerly**. **Geometric methods to accelerate k-means algorithms**. SIAM Conference on Data Mining (SDM), 2016.

4. Pablo Rivas-Perea, Erich Baker, **Greg Hamerly**, and Bryan F. Shaw. **Detection of leukocoria using a soft fusion of expert classifiers under non-clinical settings**. In BMC Opthamology, 2014.

5. Jonathan Drake and **Greg Hamerly**. **Accelerated k-means with adaptive distance bounds**. In OPT 2013: the 5th NIPS workshop on optimization for machine learning, 2012.

6. **Greg Hamerly** and Greg Speegle. **Efficient model selection for large-scale nearest-neighbor data mining**. In the *British National Conference on Databases (BNCOD)*, 2010.

7. **Greg Hamerly**. **Making k-means even faster**. In *Proceedings of the SIAM International Conference on Data Mining*, 2010.
   Acceptance rate: 23.4%.

8. Bing Yin and **Greg Hamerly**. **Hierarchical stability-based model selection for clustering algorithms**. In *Proceedings of the 2009 International Conference on Machine Learning and Applications*, 2009.
   Acceptance rate: 46%.

9. Joshua Johnston and **Greg Hamerly**. **Improving SimPoint accuracy for small simulation budgets with EDCM clustering**. In *Proceedings of 2nd workshop on Statistical and Machine learning approaches to ARchitecture and compilaTion (SMART'08)*, January 2008.
   Acceptance rate: 70%.

10. Erez Perelman, Jeremy Lau, Harish Patil, Aamer Jaleel, **Greg Hamerly**, and Brad Calder. **Cross binary simulation points**. In *2007 IEEE International Symposium on Performance Analysis of Systems and Software, ISPASS 2007*, pages 179–189, March 2007.
    Acceptance rate: 30.3%.

11. Yu Feng and **Greg Hamerly**. **PG-means: learning the number of clusters in data**. In Bernhard Schölkopf, John C. Platt, and Thomas Hoffman, editors, *Advances in Neural Information Processing Systems 19, Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems*, pages 393–400, December 2006.
    Acceptance rate: 24%.

12. **Greg Hamerly**, Erez Perelman, Jeremy Lau, Tim Sherwood, and Brad Calder. **Using machine learning to guide architecture simulation**. *Journal of Machine Learning Research (JMLR)*, 7:343–378, February 2006.
    Impact factor: 3.116 (ranked 11/92 in Artificial intelligence).
    Acceptance rate: 27%.

13. **Greg Hamerly**, Erez Perelman, and Brad Calder. **Comparing multinomial and k-means clustering for simpoint**. In *2006 IEEE International Symposium on Performance Analysis of Systems and Software, ISPASS 2006*, pages 131–142. IEEE Computer Society, 2006.
Acceptance rate: 29.6%.

14. **Greg Hamerly**, Erez Perelman, Jeremy Lau, and Brad Calder. **SimPoint 3.0: Faster and more flexible program phase analysis**. *Journal of Instruction-Level Parallelism (JILP)*, September 2005.
Acceptance rate: ∼22%.

15. **Greg Hamerly**, Erez Perelman, Jeremy Lau, and Brad Calder. **SimPoint 3.0: Faster and more flexible program analysis**. In *In proceedings of the Workshop on Modeling, Benchmarking and Simulation (MoBS)*, June 2005.
Acceptance rate: 50%.

16. Jeremy Lau, Jack Sampson, Erez Perelman, **Greg Hamerly**, and Brad Calder. **The strong correlation between code signatures and performance**. In *Proceedings of the IEEE international symposium on performance analysis of systems and software (ISPASS)*, pages 236–247, 2005.
Acceptance rate: 29%.

17. Jeremy Lau, Erez Perelman, **Greg Hamerly**, Tim Sherwood, and Brad Calder. **Motivation for variable length intervals to find hierarchical phase behavior**. In *IEEE International Symposium on Performance Analysis of Systems and Software, ISPASS 2005*, pages 135–146, 2005.
Acceptance rate: 29%.

18. John Seng and **Greg Hamerly**. **Exploring perceptron-based register value prediction**. In *Proceedings of the second value-prediction and value-based optimization workshop*, pages 10–16. International conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS-XI), October 2004.
Acceptance rate: 67%.

19. **Greg Hamerly** and Charles Elkan. **Learning the k in k-means**. In Sebastian Thrun, Lawrence K. Saul, and Bernhard Schölkopf, editors, *Advances in Neural Information Processing Systems 16*, pages 281–288. MIT Press, December 2003.
Acceptance rate: 27.6%.

20. Erez Perelman, **Greg Hamerly**, and Brad Calder. **Picking statistically valid and early simulation points**. In *12th International Conference on Parallel Architectures and Compilation Techniques (PACT 2003)*, pages 244–255. IEEE Computer Society, October 2003.
Acceptance rate: 16.7%.

21. Erez Perelman, **Greg Hamerly**, Michael Van Biesbrouck, Tim Sherwood, and Brad Calder. **Using SimPoint for accurate and efficient simulation**. In *Proceedings of the international conference on measurement and modeling of computer systems (SIGMETRICS)*, pages 318–319, June 2003.

22. **Greg Hamerly** and Charles Elkan. **Alternatives to the k-means algorithm that find better clusterings**. In *Proceedings of the 2002 ACM CIKM International Conference on Information and Knowledge Management*, pages 600–607. ACM, November 2002.
Acceptance rate: 25%.

23. Tim Sherwood, Erez Perelman, **Greg Hamerly**, and Brad Calder. **Automatically characterizing large scale program behavior**. In *Proceedings of the tenth international conference on architectural support for programming languages and operating systems (ASPLOS-X)*, pages 45–57, October 2002.
Acceptance rate: 13.7%.

24. **Greg Hamerly** and Charles Elkan. **Bayesian approaches to failure prediction for disk drives**. In *Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001)*, pages 202–209. Morgan Kaufmann, June 2001.
Acceptance rate: 32.1%.

Invited publications

1. **Greg Hamerly**, Erez Perelman, and Brad Calder. **How to use SimPoint to pick simulation points**. *SIGMETRICS Performance Evaluation Review*, 31(4):25–30, March 2004.

2. Tim Sherwood, Erez Perelman, **Greg Hamerly**, Suleyman Sair, and Brad Calder. **Discovering and exploiting program phases**. *IEEE Micro: Micro's top picks from computer architecture conferences*, 23(6):84–93, November 2003.
Impact factor: 2.565 (ranked 9/45 in Hardware and Architecture).

Invited book chapters

1. **Greg Hamerly**, Erez Perelman, Timothy Sherwood, and Brad Calder. **Representative sampling using SimPoint**. In *Processor and System-On-Chip Simulation*. Springer, 2010.

2. Brad Calder, Timothy Sherwood, **Greg Hamerly**, and Erez Perelman. **SimPoint: Picking representative samples to guide simulation**. In Lizy Kurian John and Lieven Eeckhout, editors, *Performance Evaluation and Benchmarking*, chapter 7. CRC Press, 2005.

Doctoral dissertation

1. **Greg Hamerly**. *Learning structure and concepts in data through data clustering.* PhD thesis, University of California San Diego, June 2003.

Invited presentations

- "Competitive Learning at Baylor University," New Mexico Tech; CSE Speaker Series; February 1, 2013.

- "Hypothesis Testing and Model Selection for Clustering," Texas A&M University; Parasol Seminar; December 5, 2008.

- "Improving SimPoint Accuracy for Small Simulation Budgets with EDCM Clustering," University of Texas, Austin; Computer Science colloquium; April 30, 2008.

- "PG-means: learning the number of clusters in data," University of Texas, Austin; Computer Science colloquium; May 4, 2007.

Patents

- Method and apparatus for identifying similar regions of a program's execution
United States Patent 7,802,236

Software packages developed and released

- Fast k-means (written in C++)

  This software replicates Lloyd's classical k-means algorithm, but is much faster (up to 40 times faster). It uses carefully-kept distance bounds (between points and centers) to reduce the number of distance calculations that are necessary. This algorithm enjoys a very small increase in required memory and is easily parallelized.

- SimPoint, versions 1.0 - 3.2 (written in C++)

  This software is designed to efficiently identify the number of phase behaviors in the trace of a software program (or any vector-based event frequency trace), partition the trace into those phases, and select the most representative example(s) from each phase. Designed to deal with high-dimensional, sparse inputs. I developed this software for the 2002 ASPLOS publication, and have maintained it since then.

- G-means (written in Matlab)

  An implementation of the G-means clustering algorithm that finds the number of clusters when using the k-means algorithm. Accompanies my 2003 NIPS publication.

- PG-means (written in Matlab and C++)

  An implementation of the PG-means clustering algorithm that finds the number of clusters when using the Expectation-Maximization algorithm to train a mixture of Gaussians. Accompanies my 2006 NIPS publication.

Students advised

Postdoctoral students:

- Pablo Rivas-Perea, Ph.D. 2012-2015.

Master's degrees supervised:

- Jan Sladek, 2017.
- Vaclav Cibur, 2016.
- James Boer, 2016.
- Petr Ryšavý, 2015.
- Ryan Yan, 2015.
- Ryan Henning, 2014.
- Paniz Karbasi, 2014.
- Li Guo, 2014.
- Jonathan Drake, 2013. Thesis topic: accelerated methods for the k-means algorithm.
- Tak-Chien Chiam, 2012. Thesis title: Age Classification from Facial Images for Detecting Retinoblastoma.
- Hao Guo, 2012. Project: Leukocoric Eye Detection in Digital Images
- George Montañez, Spring 2011 (Committee chairman; principally advised by Bob Marks). Thesis title: Information Storage Capacity of Genetic Algorithm Fitness Maps.
- Winston Ewert, Spring 2011 (Committee chairman; principally advised by Bob Marks). Thesis title: Studies of active information in search.

- Lei Meng, Summer 2010 (Co-advised with Greg Speegle). Topic: efficient cross-validation for k-nearest neighbor models using kd-trees.
- Bing Yin, Summer 2009. Thesis title: Hierarchical Stability-Based Model Selection For Clustering Algorithms
- Joshua Johnston, Spring 2007. Thesis title: Clustering in high dimension and choosing cluster representatives for SimPoint
- Yu Feng, Fall 2006. Thesis title: PG-means: Learning the Number of Clusters in Data

Outside committee member for the following students:

- Stephen Dark, M.S. in Electrical & Computer Engineering, November 2010.
- Chris Homan, M.S. in Computer Science, May 2009.
- Adam Sealey, M.S. in Computer Science, May 2009.
- Jason V. Davis, Ph.D. in Computer Science at University of Texas, Austin, April 2008.
- Ashish Arte, M.S. in Computer Science, November 2005.
- Lin Zhang, oral examination in Statistics, October 2005.
- Michael Coons, M.S. in Mathematics, May 2005.
- Jerry Knight, M.S. in Computer Science, April 2005.

Mentor for the following undergraduates:

- University Scholars thesis: Jonathan Drake, 2011.
- Honors thesis: Tak-Chien Chiam, 2009 - 2010.
- Honors directed reading: Tak-Chien Chiam, Spring 2009.
- Honors contract projects:
    - Chris Vaszauskas, CSI 3334, Fall 2009.
    - Mitchell Mebane, CSI 3334, Fall 2006.

Grant activity

	RFP	Agency	Title	Date submitted	Amount requested	Years	Status
1	IIS	NSF	HDR DSC: Collaborative Research: Modernizing Water and Wastewater Treatment through Data Science Education & Research (MoWaTER)	October 2019	$716,126	3	Funded
2	URC	Baylor	Learning to identify critical soliloquies in the Browning letters	Feb. 2013	$6,000	1	Funded
3	Sabbatical	Baylor	Efficient algorithms for large-scale clustering	Oct. 2009	$11,000	Summer	Funded
Continued on next page							

	RFP	Agency	Title	Date submitted	Amount requested	Years	Status
4	CNS	NSF	An interdisciplinary, problem-based, collaborative course on computational thinking	April 2009	$156,277	2	Funded
5		ICPC	ICPC Competitive Learning Institute	Feb. 2009	$2,000	1	Funded
6		JetPay	Baylor programming contest support	Aug. 2007	$500	1	Funded
7		Intel	Accurate phase analysis for efficient simulation across multiple architectures	Feb. 2007	$50,000	2	Funded
8	YIDP-06	Baylor	Improving SimPoint Functionality, Accuracy and Performance	June 2006	$25,000	1	Funded
9	-	BAEF	Postdoctoral research fellowship	Oct. 2003	$17,000	1	Funded
10	-	JL Moore	Ph.D. fellowship	Sep. 1999	$40,000	4	Funded
11	-	Powell	Ph.D. fellowship	Sep. 1999	$15,000	1	Funded

Honors and awards

- Baylor award for Outstanding Teaching (among tenured faculty from schools of Engineering and Computer Science, Business, and Law), 2012
- Two-time world finalist in the ACM International Programming Competition, 2000 and 2001

Academic service

I have been a program committee member for the following conferences:

- IEEE International Symposium on Performance Analysis for Systems and Software (ISPASS), 2009
- ACM Knowledge Discovery in Databases (KDD), 2008
- ACM Conference on Information and Knowledge Management (CIKM), 2006-2008
- International Conference on Machine Learning, 2004

I have been a reviewer for the following conferences, journals, and organizations:

- National Science Foundation (NSF)
- Journal of Machine Learning Research (JMLR)
- International Conference on Machine Learning (ICML)
- Neural Information Processing Systems (NIPS)
- AAAI Conference
- International Symposium on High Performance Computing Architectures (HPCA)
- IEEE Transactions on Knowledge and Data Engineering (TKDE)

- IEEE Transactions on Parallel and Distributed Systems (TPDS)
- IEEE Transactions on Computers (TC)
- IEEE Transactions on Neural Networks (TNN)
- Journal of Intelligent and Fuzzy Systems (JIFS)
- International Conference on Very Large Data Bases (VLDB)
- International Conference on Discovery Science

I have been on the journal review board for:

- The International Journal of Artificial Intelligence, Neural Networks and Complex-Problem Solving Technologies

Teaching experience

Teaching activity

Courses taught at Baylor University:

- CSI 3334 Data Structures
- CSI 4144 Competitive Learning
- CSI 4330 Foundations of Computing
- CSI 4336 Theory of Computer Science
- CSI 5010 Graduate Seminar
- CSI 5325 Machine Learning

Courses taught at California Polytechnic State University, San Luis Obispo:

- CSc 101 Fundamentals of Computer Science I

Teaching assistant at University of California, San Diego:

- CSE 134A Web programming and databases
- CSE 130 Programming languages
- CSE 127 Computer security
- CSE 240 Graduate computer architecture