Yan Zhu

Department of Computer Science and Engineering, University of California Riverside
Email: yzhu015@ucr.edu Mobile:(951)850-6308

Research Interests

Data Mining, Machine Learning

Education

Ph.D. Candidate, University of California, Riverside (UCR) 2013–2018(expected)
Major: Computer Science *Advisor*: Dr. Eamonn Keogh *GPA*: 3.9/4.0

M.S., Shanghai Jiao Tong University (SJTU) 2010–2013
Major: Integrated Circuit Engineering

B.S., Shanghai Jiao Tong University (SJTU) 2006–2010
Major: Microelectronics

Publications

Yan Zhu, Makoto Imamura, Daniel Nikovski and Eamonn Keogh. "Matrix Profile VII: Time Series Chains: A New Primitive for Time Series Data Mining," *IEEE ICDM 2017*, pp. 695-704 (**Best Student Paper Award**).

Yan Zhu, Zachary Zimmerman, Nader Shakibay Senobari, Chin-Chia Michael Yeh, Gareth Funning, Abdullah Mueen, Philip Brisk and Eamonn Keogh. "Exploiting a Novel Algorithm and GPUs to Break the Ten Quadrillion Pairwise Comparisons Barrier for Time Series Motifs and Joins," *Knowledge and Information Systems*, 54.1(2018): 203-236.

Michele Linardi, **Yan Zhu**, Themis Palpanas and Eamonn Keogh. "Matrix Prole X: VALMOD - Scalable Discovery of Variable-Length Motifs in Data Series," *SIGMOD 2018 (to appear).*

Yan Zhu, Chin-Chia Michael Yeh, Liudmila Ulanova, Nurjahan Begum, Yifei Ding, Hoang Anh Dau, Zachary Zimmerman, Diego Furtado Silva, Abdullah Mueen, Eamonn Keogh. "Time Series Joins, Motifs, Discords and Shapelets: a Unifying View that Exploits the Matrix Profile," *Data Mining and Knowledge Discovery* (2017): 1-41.

Yan Zhu, Zachary Zimmerman, Nader Shakibay Senobari, Chin-Chia Michael Yeh, Gareth Funning, Abdullah Mueen, Philip Brisk and Eamonn Keogh (2016). "Matrix Profile II: Exploiting a Novel Algorithm and GPUs to break the one Hundred Million Barrier for Time Series Motifs and Joins," *IEEE ICDM 2016*, pp. 739-748 (**Best Paper Award Candidate**).

Chin-Chia Michael Yeh, **Yan Zhu**, Liudmila Ulanova, Nurjahan Begum, Yifei Ding, Hoang Anh Dau, Diego Furtado Silva, Abdullah Mueen and Eamonn Keogh (2016). "Matrix Profile I: All Pairs Similarity Joins for Time Series: A Unifying View that Includes Motifs, Discords and Shapelets," *IEEE ICDM 2016*, pp. 1317-1322.

Yan Zhu and Eamonn Keogh, "Irrevocable-Choice Algorithms for Sampling from a Stream," *Data Mining and Knowledge Discovery*, 30.5 (2016): 998-1023.

Yan Zhu and Sheldon X.-D. Tan, "GPU-Accelerated Parallel Monte Carlo Analysis of Analog Circuits by Hierarchical Graph-Based Solver," *IEEE/ACM Asia and South Pacific Design Automation Conference (ASP-DAC)*, pp. 719-724, 2015.

Hanbin Hu, Guoyong Shi and **Yan Zhu**, "Incremental Symbolic Construction for Topological Modeling of Analog Circuits," in *Proc. IEEE 10th International Conference on ASIC(ASICON)*, Shenzhen, China, October 2013

Tian Huang, **Yan Zhu**, Qiannan Zhang, Yongxin Zhu, Dongyang Wang, Meikang Qiu and Lei Liu, "An LOF-based Adaptive Anomaly Detection Scheme for Cloud Computing," in *Proc. IEEE*

5th International Workshop on Security Aspects in Processes and Services Engineering(SAPSE), Kyoto, Japan, July 2013

Yan Zhu, Guoyong Shi, Frank Lee and Andy Tai, "Symbolic Time-Varying Root-Locus Analysis for Oscillator Design," in *Proc. IEEE 10th International New Circuits and Systems Conference (NEWCAS)*, Montreal, Canada, June 2012

Selected Research Projects

UCR Matrix Profile Research Jan. 2016–Present
Graduate Student Researcher, Data Mining Lab, UCR

- Proposed Time Series Chains, a new primitive for time series data mining that captures the evolving trend within the data.
- Proposed STOMP, the fastest known algorithm to evaluate time series joins. The GPU version of STOMP can evaluate the full join of a time series of length one hundred million within 5 days. This is the longest length ever tried in the literature for time series motif discovery or time series joins.
- Proposed STOMPI, the fastest known online time series join algorithm.

Irrevocable-Choice Algorithms for Streaming Data Jan. 2015–Dec. 2015
Graduate Student Researcher, Data Mining Lab, UCR

- Proposed an irrevocable-choice maximum diversity algorithm to capture samples from a data stream

Insect Detection and Classification Based on Wingbeat Sound Oct. 2014–Present
Graduate Student Researcher, Data Mining Lab, UCR

- Proposed various algorithms for insect detection, classification and activity analysis

Anomaly Detection on Cloud Platforms Jul. 2012–Mar. 2013
Team Member, Lab of Embedded Architecture, School of Microelectronics

- Implemented an adaptive multi-dimensional R-tree based K-nearest-neighbor search algorithm for fault detection

Industrial Experience

Google Inc., Kirkland, WA Jun. 2017–Sep. 2017
Summer Intern *Supervisor*: Dr. Kang Li

- Worked on machine learning algorithms for Youtube user profiling based on TensorFlow

Microsoft Research, Redmond, WA Jun. 2016–Sep. 2016
Research Intern *Supervisor*: Dr. Ethan Jackson

- Worked on Project Premonition, invented two light-weight insect classification algorithms

Mitsubishi Electric Research Laboratories, Cambridge, MA Jun. 2015–Sep. 2015
Research Intern *Supervisor*: Dr. Daniel Nikovski

- Worked in Data Analytics group, invented a novel time series prognostics algorithm, results have been summarized in a patent submission

Honors and Awards

- Best Student Paper Award ICDM 2017
- Student Travel Awards KDD 2017, ICDM 2016
- Department Fellowship Award, Department of Computer Science and Engineering, University of California, Riverside 2015
- Dean's Fellowship Award, Department of Electrical and Computer Engineering, University of California, Riverside 2013-2014

- Excellent Academic Scholarship (Third-Class) of SJTU 2011-2012
- Excellent Academic Scholarship (Second-Class) of SJTU 2010-2011
- Excellent Academic Scholarship (Second-Class) of SJTU 2006-2007

Skills

- Proficiency in C/C++, Perl, Tcl, Cadence SKILL language, Linux Shell, Verilog.
- Familiarity with Python, Hspice, Cadence Spectre, Matlab, Lapack, LaTeX.
- Familiarity with various machine learning algorithms
- Familiarity with TensorFlow