

Jessica H. Lin

CONTACT INFORMATION	Computer Science George Mason University 4400 University Drive, MS 4A5 Fairfax, VA 22030	Office: (703) 993-4693 Fax: (703) 993-1710 Email: jessica@cs.gmu.edu Website: http://www.cs.gmu.edu/~jessica
---------------------	---	--

RESEARCH INTERESTS	<ul style="list-style-type: none">• Mining time series data, including indexing, efficient discovery of motifs, anomalies, rules, structures, hierarchies in massive time series databases, and their applications in medical, geoinformatics, and astronomy domains.• Visualizing large time series databases• Mining spatiotemporal data• Image/shape mining• Text mining• Stream data mining
--------------------	--

EDUCATION	Ph.D., Computer Science (Research: Database/Data Mining) University of California, Riverside Dissertation Title: Discovering Unusual and Non-trivial Patterns in Time Series Databases Advisor: Dr. Eamonn Keogh	June 2005
-----------	--	-----------

	M.S., Computer Science (Research: Database/Data Mining) University of California, Riverside Project: Dimensionality Reduction by Random Projection and Latent Semantic Indexing Advisor: Dr. Dimitrios Gunopulos	Mar 2002
--	--	----------

	B. S., Computer Science University of California, Riverside	July, 1999
--	---	------------

PROFESSIONAL EXPERIENCE	Assistant Professor Computer Science Department George Mason University	8/05 – Present
-------------------------	--	----------------

	Research Assistant Computer Science Department University of California, Riverside	03/02 – 06/05
--	---	---------------

	Research Intern AT&T Research, Information Visualization	6/04 – 9/04
--	--	-------------

	Research Intern Microsoft Research, Machine Learning and Applied Statistics.	6/03 – 9/03
--	--	-------------

	Summer Intern Environmental Services Research Institute (ESRI)	6/01 – 9/01
--	--	-------------

	Programmer George E. Brown Jr. Salinity Laboratory, U.S. Department of Agriculture (USDA)	6/98 – 6/00
--	--	-------------

TEACHING

Data Structures (Fall 2005, Spring 2005, Fall 2006)

Database Management Systems (Fall 2007, Spring 2008, Fall 2008, Fall 2009, Fall 2010, Spring 2011)

Data Mining for Multimedia Data (Fall 2008, Fall 2009, Fall 2010)

Special Topics in Data Mining (Fall 2006, Spring 2007, Fall 2007, Spring 2011)

PUBLICATIONS

Edited Book

1. Guido, C., Lin, J., and Waters, N. (Eds). (2012). *Data Mining for Geoinformatics: Methods and Applications*. Springer-Verlag. *Under contract*.

Book Chapters

2. Lin, J., Williamson, S., Borne, K. and DeBarr, D. (2011). **Pattern Recognition in Time Series**. *Advances in Machine Learning and Data Mining for Astronomy*. Eds. Kamal, A., Srivastava, A., Way, M., and Scargle, J. Chapman & Hall. *Accepted*.
3. Ratanamahatana, C. A., Lin, J., Gunopulos, D., Keogh, E., Vlachos, M., and Das, G. (2010): **Mining Time Series Data**. *Data Mining and Knowledge Discovery Handbook 2010*: 1049-1077
4. Lin, J., Vlachos, M., Keogh, E. & Gunopulos, D. (2006). **Multi-Resolution Time Series Clustering and Application to Images**. *Multimedia Data Mining and Knowledge Discovery*, Eds. Valery Al Petrushin and Latifur Khan. Springer.
5. Ratanamahatana, C. A., Lin, J., Gunopulos, D., Keogh, E. & Vlachos, M. (2005). **Mining Time Series Data**. *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*. Kluwer Academic Publishers.

Journal Papers

6. Lin, J., Keogh, E., Li, W. & Lonardi, S. (2007). **Experiencing SAX: A Novel Symbolic Representation of Time Series**. *Data Mining and Knowledge Discovery Journal*.
7. Lonardi, S., Lin, J., Keogh, E. & Chiu, B. (2006). **Efficient Discovery of Unusual Patterns in Time Series**. *Special Issue of New Generation Computing Journal*.
8. Keogh, E., Lin, J. & Fu, A. (2006). **Finding the Most Unusual Time Series Subsequence: Algorithms and Applications**. *Knowledge and Information Systems (KAIS)*. Springer-Verlag.
9. Keogh, E., Lin, J., Fu, A. & Van Herie, H. (2005). **Finding the Unusual Medical Time Series: Algorithms and Applications**. *IEEE Transactions on Information Technology in Biomedicine*.
10. Lin, J., Keogh, E. & Lonardi, S. (2005). **Visualizing and Finding Patterns in Time Series**. *Information Visualization Journal*.
11. Keogh, E., Lin, J. & Truppel, W. (2004). **Clustering of Time Series Subsequences is Meaningless: Implications for Past and Future Research**. *Knowledge and Information Systems (KAIS)*, Springer-Verlag.
12. Lin, J. & Keogh, E. (2004). **Finding or Not Finding Rules in Time Series**. *Special Issue of Advances in Econometrics*.

Conference and Workshop Papers

13. Ngan, C. K., Brodsky, A. & Lin, J. (2011) **A Service Framework for Learning, Querying, and Monitoring Multivariate Time Series**. In *Proceedings of the 13th International Conference on Enterprise Information Systems*. June 8-11, 2011. Beijing, China. To Appear.
14. Cervone G., Lin J., Franzese P., **Addressing Wind Direction Uncertainty in Source Estimation Through Dynamic Time Warping**, In *Proceedings of the 91st American Meteorological Society Annual Meeting*, Session 2: Computational intelligence methods and their applications to environmental science, Seattle, WA, January 2011
15. Lin, J., Cervone, G. & Waters, N. (Eds.) (2010). **Proceedings of the International Workshop on Data Mining for Geoinformatics. DMG/SIGSPATIAL GIS 2010**. San Jose, CA. Nov 2, 2010.
16. Lin, J., Cervone, G. & Franzese, P. (2010). **Assessment in Error in Air Quality Models Using Dynamic Time Warping**. In *proceedings of the International Workshop on Data Mining for Geoinformatics, in conjunction with SIGSPATIAL GIS 2010*. San Jose, CA. Nov 2, 2010.
17. Lin, J. & Li, Y. (2010) **Finding Approximate Frequent Patterns in Streaming Medical Data**. In *proceedings of the 23rd IEEE International Symposium on Computer-Based Medical Systems*. Perth, Australia. Oct 12-15, 2010.
18. Li, Y & Lin, J. (2010). **Approximate Variable-Length Time Series Motif Discovery Using Grammar Inference**. In *proceedings of the 10th International Workshop on Multimedia Data Mining, in conjunction with SIGKDD 2010*.
19. Ngan, C. K., Brodsky, A. & Lin, J. (2010). **Decisions on Multivariate Time Series: Combining Domain Knowledge with Utility Maximization**. In *proceedings of the 15th International Conference on Decision Support Systems*.
20. Lin, J. & Li, Y. (2009). **Finding Structurally Different Medical Data**. In *proceedings of the 22nd IEEE International Symposium on Computer-Based Medical Systems*. August 3-4. Albuquerque, New Mexico.
21. Lin, J. & Li, Y. (2009). **Finding Structural Similarity in Time Series Using Bag-of-Patterns Representation**. In *proceedings of the 21st International Conference on Scientific and Statistical Database Management (SSDBM)*. New Orleans. June 2-4, 2009.
22. Al-Shammari, E. & Lin, J. (2008). **Automated Corpora Creation Using a Novel Arabic Stemming Algorithm**. In *proceedings of the 2008 International Symposium on Using Corpora in Contrastive and Translation Studies (UCCTS)*. Hanzhou, China. Sept 25-27.
23. Al-Shammari, E. & Lin, J. (2008). **A New Arabic Stemming Algorithm**. In *proceedings of the 2008 ISCA Workshop on Experimental Linguistics*. Athens, Greece. Aug 25-27.
24. Al-Shammari, E. & Lin, J. (2008). **A Novel Arabic Lemmatization Algorithm**. In *proceedings of the 2nd SIGIR Workshop on Analytics for Noisy Unstructured Text Data*. Singapore, July 24-27, 2008. p. 113-118.
25. Lin, J. & Etter, D. (2008). **Exact and Approximate Reverse Nearest Neighbor Search in Multimedia Data**. *SIAM International Conference on Data Mining*.
26. DeBarr, D. & Lin, J. (2007). **Time Series Classification Challenge Experiments**. In *proceedings of the Workshop and Challenge on Time Series Classification, at the 13th ACM SIGKDD International Conference on*

Knowledge Discovery and Data Mining. San Jose, CA. Aug 12-15.

27. Lin, J. & Keogh, E. (2006). **Group SAX: Extending the Notion of Contrast Sets to Time Series and Multimedia Data**. In *proceedings of the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases*. Berlin, Germany. Sept 18-22.
28. Fu, A., Leung, O., Keogh, E. & Lin, J. (2006). **Finding Time Series Discords Based on Haar Transform**. In *proceedings of the 2nd International Conference on Advanced Data Mining and Applications*. Xi'an, China. Aug 14-18.
29. Keogh, E., Lin, J. & Fu, A. (2005). **HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence**. In *proceedings of the 5th IEEE International Conference on Data Mining (ICDM)*. Nov 27-30. Houston, TX.
30. Lin, J., Keogh, E., Fu, A., & van Herie, H. (2005). **Approximations to Magic: Finding Unusual Medical Time Series**. In *proceedings of the 18th International Symposium on Computer-Based Medical Systems*. June 23-24. Dublin, Ireland.
31. Lin, J., Vlachos, M., Keogh, E., Gunopulos, D., Liu, J. W., Yu, S. J., Le, J. J. (2005): **A MPAA-Based Iterative Clustering Algorithm Augmented by Nearest Neighbors Search for Time-Series Data Streams**. *PAKDD 2005*: 333-342
32. Lin, J., Keogh, E., Lonardi, S., Lankford, J. & Nystrom, D. (2004). **Visually Mining and Monitoring Massive Time Series**. In *proceedings of the International Conference on Knowledge Discovery and Data Mining (SIGKDD)*. Aug 22-25. Seattle, WA.
33. Lin, J., Keogh, E., Lonardi, S., Lankford, J. & Nystrom, D. (2004). **VizTree: a Tool for Visually Mining and Monitoring Massive Time Series**. In *proceedings of the 30th International Conference on Very Large Databases (VLDB)*. Aug 29-Sep 3. Toronto, CA.
34. Lin, J., Vlachos, M., Keogh, E., & Gunopulos, D. (2004). **Iterative Incremental Clustering of Time Series**. *IX Conference on Extending Database Technology (EDBT)*. March 14-18, 2004.
35. Keogh, E., Lin, J. & Truppel, W. (2003). **Clustering of Time Series Subsequences is Meaningless: Implications for Past and Future Research**. In *proceedings of the 3rd IEEE International Conference on Data Mining (ICDM)*. Melbourne, FL. Nov 19-22, 2003.
36. Lin, J., Vlachos, M, Keogh, E., & Gunopulos, D. (2003). **Multi-Resolution K-Means Clustering of Time Series and Application to Images**. In *proceedings of the 4th SIGKDD Workshop on Multimedia Data Mining*. Washington D.C., Aug 24-27, 2003.
37. Lin, J., Keogh, E., Lonardi, S. & Chiu, B. (2003). **A Symbolic Representation of Time Series, with Implication for Streaming Algorithms**. In *proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*. San Diego, CA. June 13, 2003. <http://www.cs.gmu.edu/~jessica/sax.htm>
38. Lin, J. Keogh, E. & Truppel, W. (2003). **Clustering of Streaming Time Series is Meaningless**. In *proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*. San Diego, CA. June 13, 2003. (Acceptance rate: 28%).
39. Lin, J., Vlachos, M., Keogh, E. & Gunopulos, D. (2003). **A Wavelet-Based Anytime Algorithm for K-Means Clustering of Time Series**. In *proceedings of the Workshop on Clustering High Dimensional Data and Its Applications, at the 3rd SIAM International Conference on Data Mining*. San Francisco, CA.

May 3, 2003.

40. Lin, J. & Gunopulos, D. (2003). **Dimensionality Reduction by Random Projection and Latent Semantic Indexing**. In *proceedings of the Text Mining Workshop, at the 3rd SIAM International Conference on Data Mining*. San Francisco, CA. May 3, 2003.
41. Lin, J. Keogh, E. & Truppel, W. (2003). **(Not) Finding Rules in Time Series: A Surprising Result with Implications for Previous and Future Research**. In *proceedings of the 2003 International Conference on Artificial Intelligence, Special Session on Applications of AI in Finance and Economics*. Las Vegas, NV. June 23-26, 2003.
42. Patel, P., Keogh, E., Lin, J., & Lonardi, S. (2002). **Mining Motifs in Massive Time Series Databases**. In *proceedings of the 2002 IEEE International Conference on Data Mining (ICDM)*. Maebashi City, Japan. Dec 9-12, 2002.
43. Lin, J., Keogh, E., Patel, P. & Lonardi, S. (2002). **Finding Motifs in Time Series**. In the *2nd Workshop on Temporal Data Mining, at the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Edmonton, Alberta, Canada. July 23-26, 2002.

PROFESSIONAL ACTIVITIES

Workshop Chair

- SIGSPATIAL International Workshop on Data Mining for Geoinformatics, in cooperation with SIGSPATIAL GIS 2010.

Workshop Program Chair

- Data Mining for Geoinformatics in Cooperation with the 17th International Conference on Geoinformatics, 2009.
- International Workshop on Spatial and Spatiotemporal Data Mining (SSTDM 09) in Cooperation with IEEE ICDM 2009
- International Workshop on Spatial and Spatiotemporal Data Mining (SSTDM 08) in Cooperation with IEEE ICDM 2008

Program Committee Member

- International Conference on Data Mining (ICDM), 2007, 2008, 2009, 2010, 2011
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2006, 2007, 2011
- International Conference on Machine Learning and Applications (ICMLA), 2010
- European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD) 2007, 2008, 2009, 2010, 2011
- ECML/PKDD Workshop on Knowledge Discovery in Health Care and Medicine, 2011

Journal Reviewer

- Transaction on Knowledge and Data Engineering (TKDE) 2006, 2007, 2008, 2009, 2010, 2011
- Knowledge and Information Systems (KAIS) 2006, 2007, 2008, 2009
- Data Mining and Knowledge Discovery Journal (DAMI) 2006, 2008, 2010
- ACM Computing Survey 2009, 2011
- The Visual Computer Journal, 2006
- *Special Issue of Advances in Econometrics - Applications of AI in Finance & Econ*, 2004

External Reviewer

- ICDE 2006, CBMS 2005, TDM 2005, VLDB 2004, DMWI 2003, ICML 2003, CIKM 2002.