

# Haixun Wang

---

## CONTACT INFORMATION

Haixun Wang, Ph.D.  
IBM T. J. Watson Research Center  
19 Skyline Dr.  
Hawthorne, New York 10533

*Voice:* (914)784-7131  
*Fax:* (914)784-7455  
*Email:* haixun@us.ibm.com  
*Homepage:* <http://wis.cs.ucla.edu/~hxwang>

## RESEARCH INTERESTS

Data Mining  
  
Information Retrieval  
  
Database System and Language

## RECENT PROFESSIONAL ACTIVITIES

SDM'09 Area PC Chair, ICDE'09 Demo PC Chair, ICDE'09 PC member, WWW'09 PC member, EDBT'09 PC member, APWeb'08 Industry PC Chair, ICDM'08 Exhibit and Demo Chair, SIGMOD'07 Sponsor Chair, SIGKDD'08 Senior PC member (Area Chair), VLDB'08 PC member, SDM'08 PC member, ICDM'08 PC member, etc.

## EDUCATION

**Ph.D. in Computer Science**      University of California, Los Angeles      Sept. 1996 - Jun. 2000  
Major: Database and Knowledge Base Systems  
Minor: Networking and Operating Systems  
Dissertation: "User-Defined Aggregates for Advanced Database Applications"  
Advisor: Prof. Carlo Zaniolo

**M.S. in Computer Science**      Shanghai Jiao Tong University      Sept. 1994 - Jun. 1996  
Major: Database Systems  
Minor: Operating System  
Thesis: "Full-Text Information Retrieval in Database Systems"

**B.S. in Computer Science**      Shanghai Jiao Tong University      Sept. 1990 - Jun. 1994  
Honor Program

## EXPERIENCE

**Technical Assistant**      IBM T. J. Watson Research Center      July 2007 - Present  
Technical Assistant to Mark N. Wegman, Head of Computer Science, IBM Research

**Technical Assistant**      IBM T. J. Watson Research Center      Aug. 2006 - Feb. 2007

Technical Assistant to Stuart Feldman, VP of Computer Science, IBM Research

As the technical assistant to IBM's VP of Computer Science who oversees research in all of IBM's 8 global labs, I have the opportunity to review hundreds of exploratory projects that aim at addressing the grand challenges of computer science in the next decade. Among everything else, I learned to judge the validity and worthiness of a project, both in terms of science and business.

**Research Staff Member**                      IBM T. J. Watson Research Center                      Aug. 2000 - July. 2006  
Research in the area of data mining, information retrieval, and data management.

My recent research interest lies in developing advanced database language and systems, indexing and searching advanced data structures, classifying and clustering data streams, providing secure database outsourcing, etc. I have also advised many students. Since 2000, I have published papers in SIGMOD (6), VLDB (7), ICDE (18), EDBT (4), SIGKDD (4), SDM (6), ICDM (14), etc.

## TEACHING EXPERIENCE

Students I have advised:

- I advised intern students who usually spent several summers at IBM. My recent student, Hao He, was from Duke University. I was on Hao's Ph.D. dissertation committee. Hao interned at IBM for 3 summers, and was awarded an IBM fellowship. Our work appeared in CIKM, ICDE, and SIGMOD. Hao graduated in 2007 and is now with Google. Similar to Hao, I have advised many other intern students in the previous years, including Andrea Chu (UCLA), Yiming Yang (UCI), etc.
- I also advised students remotely. Recent students include Shixi Chen of Fudan University, and our work appeared in ICDE'08; Peng Wang of Fudan University (now an Assistant Professor of Fudan University), and our work appeared in ICDM (two papers) and TKDE; Zhao Qi of Georgia Tech (now with the AT&T lab), and our work appeared in PODS'06; Yijian Bai of UCLA (now with Google), and our work appeared in SDM, ICDE, and CIKM; Min Xie of Renmin University of China, and our work appeared in VLDB and EDBT; etc.

Taught CS143 (Introduction to Database Systems) at UCLA. The class had over 100 enrolled students, and the teaching load was 4 hours per week.

Taught a summer course (Advances in Database Research) at Renmin University of China in 2003. The class had over 50 enrolled students, and the teaching load was 20 hours per week for 3 weeks.

## PUBLICATIONS

### 2009

- [ 1 ] Tao Li, Wei Peng, Chang-shing Perng, Sheng Ma, and **Haixun Wang**, *An Integrated Data-Driven Framework for Computing System Management*, IEEE Transactions on Systems, Man, and Cybernetics - part A., 2009, .
- [ 2 ] Shixi Chen, **Haixun Wang**, and Shuigeng Zhou, *Concept Clustering for Evolving Data*, 24th International Conference on Data Engineering (ICDE), 2009, Shanghai, China.
- [ 3 ] Risi Thonang, Hao He, AnHai Doan, **Haixun Wang**, and Jun Yang, *Weighted Proximity Best-Joins for Information Retrieval*, 24th International Conference on Data Engineering (ICDE), 2009, Shanghai, China.

[ 4 ] Xiaohui Gu and **Haixun Wang**, *Online Anomaly Prediction for Robust Cluster Systems*, 24th International Conference on Data Engineering (ICDE), 2009, Shanghai, China.

## 2008

[ 5 ] Lipyeow Lim, **Haixun Wang**, and Min Wang, *Modeling and Querying E-Commerce Data in Hybrid Relational-XML DBMSs*, 27th International Conference on Conceptual Modeling (ER), 2008, Barcelona, Spain. (**Best Paper Award**)

[ 6 ] Yijian Bai, Hetal Thakkar, **Haixun Wang**, and Carlo Zaniolo, *Time-Stamp Management and Query Execution in Data Stream Management Systems*, Journal of IEEE Internet Computing, 2008, Vol. 12, No. 6, pages 13-21.

[ 7 ] **Haixun Wang**, Jian Yin, Chang-shing Perng, and Philip S. Yu, *Dual Encryption for Query Integrity Assurance*, ACM 17th Conference on Information and Knowledge Management (CIKM), 2008, Napa Valley, California.

[ 8 ] **Haixun Wang** and Jian Pei, *Clustering by Pattern Similarity*, Journal of Computer Science and Technology (JCST), 2008, Vol. 23, No. 4, pages 481-496.

[ 9 ] Ruoming Jin, Yang Xiang, Ning Ruan, and **Haixun Wang**, *Efficiently Answering Reachability Query on Very Large Directed Graphs*, ACM International Conference on Management of Data (SIGMOD), 2008, Vancouver, Canada.

[10] Jiangming Yang, **Haixun Wang**, Ning Gu, Yiming Liu, Chunsong Wang, and Qiwei Zhang, *Lock-Free Consistency Control for Web 2.0 Applications*, 17th International World Wide Web Conference (WWW), 2008, Beijing, China.

[11] Wei-Shinn Ku, Roger Zimmermann, and **Haixun Wang**, *Location-based Spatial Query Processing in Wireless Broadcast Environment*, IEEE Transactions on Mobile Computing (TMC), 2008, Vol. 7, No. 1.

[12] Jiefeng Cheng, Jeffrey Xu Yu, Xuemin Lin, **Haixun Wang**, and Philip S. Yu, *Fast Computing Reachability Labelings for Large Graphs with High Compression Rate*, 11th International Conference on Extending Database Technology (EDBT), 2008, Nantes, France.

[13] Min Xie, **Haixun Wang**, Jian Yin, and Xiaofeng Meng, *Providing Freshness Guarantees for Outsourced Databases*, 11th International Conference on Extending Database Technology (EDBT), 2008, Nantes, France.

[14] Shixi Chen, **Haixun Wang**, Shuigeng Zhou, and Philip S. Yu, *Stop Chasing Trends: Discovering High Order Models in Evolving Data*, 23rd International Conference on Data Engineering (ICDE), 2008, Cancun, Mexico.

[15] Jiefeng Cheng, Jeffrey Xu Yu, Bolin Ding, Philip S. Yu, and **Haixun Wang**, *Fast Graph Pattern Matching*, 23rd International Conference on Data Engineering (ICDE), 2008, Cancun, Mexico.

[16] Junyi Xie, Jun Yang, Yuguo Chen, **Haixun Wang**, and Philip S. Yu, *A Sampling-Based Approach to Information Recovery*, 23rd International Conference on Data Engineering (ICDE), 2008, Cancun, Mexico.

## 2007

[17] Lipyeow Lim, **Haixun Wang**, and Min Wang, *Unifying Data and Domain Knowledge Using*

- Virtual Views*, 33rd International Conference on Very Large Data Bases (VLDB), 2007, Vienna, Austria.
- [18] Min Xie, **Haixun Wang**, Jian Yin, and Xiaofeng Meng, *Integrity Auditing of Outsourced Data*, 33rd International Conference on Very Large Data Bases (VLDB), 2007, Vienna, Austria.
- [19] Kun-Lung Wu, Philip S. Yu, Bugra Gedik, Kristen Hildrum, Charu Aggarwal, Eric Bouillet, Wei Fan, David George, Xiaohui Gu, Gang Luo, and **Haixun Wang**, *Challenges and Experience in Prototyping a Multi-Modal Stream Analytic and Monitoring Application on System S (industry paper)*, 33rd International Conference on Very Large Data Bases (VLDB), 2007, Vienna, Austria.
- [20] Wei Peng, Charles Perng, Tao Li, and **Haixun Wang**, *Event Summarization for System Management*, ACM Int'l Conf. on Knowledge Discovery and Data Mining (SIGKDD), 2007, San Jose, USA.
- [21] Hao He, **Haixun Wang**, Jun Yang, and Philip Yu, *BLINKS: Ranked Keyword Searches on Graphs*, ACM International Conference on Management of Data (SIGMOD), 2007, Beijing, China.
- [22] Chengkai Li, Min Wang, Lipyeow Lim, **Haixun Wang**, and Kevin Chang, *Supporting Ranking and Clustering as Generalized Order-By and Group-By*, ACM International Conference on Management of Data (SIGMOD), 2007, Beijing, China.
- [23] Yijian Bai, **Haixun Wang**, and Carlo Zaniolo, *Load Shedding in Classifying Multi-Source Streaming Data: A Bayes Risk Approach*, 6th SIAM International Conference on Data Mining (SDM), 2007, Minnesota, USA.
- [24] Xiaohui Gu, Philip S. Yu, and **Haixun Wang**, *Adaptive Load Diffusion for Multiway Windowed Stream Joins*, 22nd International Conference on Data Engineering (ICDE), 2007, Istanbul, Turkey.
- [25] Jian Pei, Ada Wai-Chee Fu, Xuemin Lin, and **Haixun Wang**, *Computing Compressed Multi-dimensional Skyline Cubes Efficiently*, 22nd International Conference on Data Engineering (ICDE), 2007, Istanbul, Turkey.
- [26] Haoliang Jiang, **Haixun Wang**, Shuigeng Zhou, and Philip S. Yu, *GString: A Novel Approach for Efficient Search in Graph Databases*, 22nd International Conference on Data Engineering (ICDE), 2007, Istanbul, Turkey.
- [27] Wei-Shinn Ku, Roger Zimmermann, and **Haixun Wang**, *Location-based Spatial Queries with Data Sharing in Wireless Broadcast Environments (short paper)*, 22nd International Conference on Data Engineering (ICDE), 2007, Istanbul, Turkey.
- [28] Yijian Bai, Hetal Thakkar, **Haixun Wang**, and Carlo Zaniolo, *Optimizing Timestamp Management in Data Stream Management Systems (short paper)*, 22nd International Conference on Data Engineering (ICDE), 2007, Istanbul, Turkey.
- [29] Lipyeow Lim, **Haixun Wang**, and Min Wang, *Semantic Data Management: Towards Querying Data with their Meaning (short paper)*, 22nd International Conference on Data Engineering (ICDE), 2007, Istanbul, Turkey.
- [30] Yijian Bai, Hetal Thakkar, **Haixun Wang**, and Carlo Zaniolo, *A Flexible Query Graph Based Model for the Efficient Execution of Continuous Queries*, First International Workshop on Scalable Stream Processing Systems (SSPS), 2007, Istanbul, Turkey.
- [31] Peng Wang, **Haixun Wang**, Xiaochen Wu, Wei Wang, and Baile Shi, *A Low Granularity*

*Classifier for Data Streams with Concept Drifts and Biased Class Distribution*, IEEE Transactions on Knowledge and Data Engineering (TKDE), 2007, Vol. 19, No. 9.

**2006**

[32] Peng Wang, **Haixun Wang**, Wei Wang, Baile Shi, and Philip S. Yu, *LOCI: Load Shedding through Class-Preserving Data Acquisition*, 6th IEEE International Conference on Data Mining (ICDM), 2006, Hong Kong, China.

[33] Chang-shing Perng, **Haixun Wang**, and Sheng Ma, *Fast Relevance Discovery in Time Series (short paper)*, 6th IEEE International Conference on Data Mining (ICDM), 2006, Hong Kong, China.

[34] Gabriel Pui Cheong Fung, Jeffrey Yu, **Haixun Wang**, Huan Liu, and David W Cheung, *A Balanced Ensemble Approach to Weighting Classifiers for Text Classification (short paper)*, 6th IEEE International Conference on Data Mining (ICDM), 2006, Hong Kong, China.

[35] **Haixun Wang** and Ying Yang, *Predictive Learning on Data Streams (tutorial)*, 6th IEEE International Conference on Data Mining (ICDM), 2006, Hong Kong, China.

[36] Yijian Bai, Hetal Thakkar, Richard Luo, **Haixun Wang**, and Carlo Zaniolo, *A Data Stream Language and System Designed for Power and Extensibility*, ACM Conference on Information and Knowledge Management (CIKM), 2006, Arlington, USA.

[37] Jian Pei, **Haixun Wang**, Jian Liu, Ke Wang, Jianyong Wang, and Philip S. Yu, *Discovering Frequent Closed Partial Orders from Strings*, IEEE Transactions on Knowledge and Data Engineering (TKDE), 2006, Vol. 18, Iss. 11, Page(s): 1467 - 1481.

[38] **Haixun Wang**, Jian Yin, Jian Pei, Philip S. Yu, and Jeffrey X. Yu, *Suppressing Model Overfitting in Mining Concept-Drifting Data Streams*, 12th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), 2006, Philadelphia, USA.

[39] Qi Zhao, Mitsunori Ogihara, **Haixun Wang**, and Jun Xu, *Finding Global Icebergs over Distributed Data Sets*, 25th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS), 2006, Chicago.

[40] **Haixun Wang**, Hao He, Jun Yang, Philip Yu, and Jeffrey Xu Yu, *Dual Labeling: Answering Graph Reachability Queries in Constant Time*, 21st International Conference on Data Engineering (ICDE), 2006, Atlanta, USA.

[41] Jeffrey Xu Yu, Jiefeng Cheng, Xuemin Lin, **Haixun Wang**, and Philip Yu, *Fast Computation of Reachability Labeling for Large Graphs*, 10th International Conference on Extending Database Technology (EDBT), 2006, Munich, Germany.

[42] Yun Chi, **Haixun Wang**, Philip S. Yu, and Richard R. Muntz, *Catch the moment: maintaining closed frequent itemsets over a data stream sliding window*, Knowledge Information Systems, 2006, 10(3): 265-294.

[43] Peng Wang, **Haixun Wang**, Xiaochen Wu, Wei Wang, and Baile Shi, *On Reducing Classifier Granularity in Mining Concept-Drifting Data Streams*, 5th IEEE International Conference on Data Mining (ICDM), 2006, New Orleans, Louisiana, USA.

[44] Jian Pei, Jian Liu, **Haixun Wang**, Ke Wang, Philip S. Yu, and Jianyong Wang, *Efficiently Mining Frequent Closed Partial Orders*, 5th IEEE International Conference on Data Mining (ICDM), 2006, New Orleans, Louisiana, USA.

**2005**

- [45] Jiong Yang, **Haixun Wang**, Wei Wang, and Philip S. Yu, *An Improved Biclustering Method for Analyzing Gene Expression Profiles*, International Journal on Artificial Intelligence Tools, 2005, Vol. 14, No. 5.
- [46] Hao He, **Haixun Wang**, Jun Yang, and Philip S. Yu, *Compact Reachability Labeling for Graph-Structured Data*, 14th ACM Conference on Information and Knowledge Management (CIKM), 2005, Bremen, Germany.
- [47] **Haixun Wang** and Jian Pei, *A Random Method for Quantifying Changing Distributions in Data Streams*, 9th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), 2005, Porto, Portugal.
- [48] **Haixun Wang**, Philip S. Yu, and Jian Pei, *Pattern-based Similarity Search for Microarray Data*, 11th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), 2005, Chicago.
- [49] Moonjung Cho, Jian Pei, **Haixun Wang**, and Wei Wang, *Preference-based Frequent Pattern Mining*, International Journal of Data Warehousing and Mining, 2005, Vol. 1, No. 4.
- [50] **Haixun Wang**, Chang-Shing Perng, Sheng Ma, and Philip S. Yu, *Demand-driven Frequent Itemset Mining Using Pattern Structures*, Knowledge and Information Systems, 2005, Vol. 8, No. 1, pages 85-102.
- [51] Yun Chi, **Haixun Wang**, and Philip S. Yu, *Loadstar: Load Shedding in Data Stream Mining (demo)*, 31th International Conference on Very Large Data Bases (VLDB), 2005, Trondheim, Norway.
- [52] Chang Luo, **Haixun Wang**, and Carlo Zaniolo, *A Native Extension of SQL for Mining Data Streams (demo)*, ACM International Conference on Management of Data (SIGMOD), 2005, Baltimore, Maryland.
- [53] **Haixun Wang** and Xiaofeng Meng, *On the Sequencing of Tree Structures for XML Indexing*, 21st International Conference on Data Engineering (ICDE), 2005, Tokyo, Japan.
- [54] **Haixun Wang**, Jian Pei, and Philip S. Yu, *Online Mining of Data Streams: Problems, Applications and Progress (tutorial)*, 21st International Conference on Data Engineering (ICDE), 2005, Tokyo, Japan.
- [55] **Haixun Wang**, Chang-Shing Perng, and Philip S. Yu, *Near-Neighbor Search in Pattern Distance Spaces*, 4th SIAM International Conference on Data Mining (SDM), 2005, Newport Beach, USA.
- [56] Yun Chi, Philip S. Yu, **Haixun Wang**, and Richard Muntz, *Loadstar: A Load Shedding Scheme for Classifying Data Streams*, 4th SIAM International Conference on Data Mining (SDM), 2005, Newport Beach, USA.
- [57] Naoki Abe, Edwin Pednault, Bianca Zadrozny, **Haixun Wang**, Wei Fan, Chid Apte, and (Book Chapter), *Sequential Risk Management in E-Business by Reinforcement Learning*, Handbook of Integrated Risk Management for E-Business (A. Labbi, Ed.), 2005, ISBN 1-932159-07-X, J. Ross Publishing, Inc.
- [58] **Haixun Wang** and Wei Wang, *Stay Current and Relevant in Data Mining Research (panel)*,

10th International Conference on Database Systems for Advanced Applications (DASFAA), 2005, Beijing, China.

[59] **Haixun Wang**, Philip S. Yu, and Jiawei Han, *Mining Data Streams*, The Data Mining and Knowledge Discovery Handbook, 2005, p777-792.

## 2004

[60] Yun Chi, **Haixun Wang**, Philip Yu, and Richard Muntz, *Moment: Maintaining Closed Frequent Itemsets over a Stream Sliding Window*, 4th IEEE International Conference on Data Mining (ICDM), 2004, Brighton, UK.

[61] Yu Wang, **Haixun Wang**, Xiaofeng Meng, and Shan Wang, *Estimating the Selectivity of XML Path Expression with Predicates by Histograms*, 5th International Conference on Web-Age Information Management (WAIM), 2004, Dalian, China.

[62] Jian Pei, **Haixun Wang**, and Philip S. Yu, *Online Mining Data Streams: Problems, Applications and Progress (tutorial)*, 10th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), 2004, Seattle, Washington.

[63] Yan-Nei Law, **Haixun Wang**, and Carlo Zaniolo, *Query Languages and Data Models for Database Sequences and Data Streams*, 30th International Conference on Very Large Data Bases (VLDB), 2004, Toronto, Canada.

[64] **Haixun Wang**, Fang Chu, Wei Fan, Philip S Yu, and Jian Pei, *A Fast Algorithm for Subspace Clustering by Pattern Similarity*, 16th International Conference on Scientific and Statistical Database Management (SSDBM), 2004, Santorini Island, Greece.

[65] Xiaofeng Meng, Yu Jiang, Yan Chen, and **Haixun Wang**, *XSeq: An Index Infrastructure for Tree Pattern Queries (demo)*, ACM International Conference on Management of Data (SIGMOD), 2004, Paris, France.

[66] Wei Fan, Yi-an Huang, Haixun Wang, and Philip S Yu, *Active Mining of Data Streams*, 3rd SIAM International Conference on Data Mining (SDM), 2004, Florida, USA.

[67] Wei Fan, Philip S Yu, and **Haixun Wang**, *Mining Extremely Skewed Security Trading Anomalies*, 9th International Conference on Extending Database Technology (EDBT), 2004, Crete, Greece.

[68] Cindy Chen, **Haixun Wang**, and Carlo Zaniolo, *Toward Extensible Spatio-Temporal Databases: an approach based on User-Defined Aggregates*, "Flexible querying and reasoning in spatio-temporal databases: theory and applications, 2004, Springer Geosciences/Geoinformation series.

## 2003

[69] **Haixun Wang**, Sanghyun Park, Wei Fan, and Philip S Yu, *ViST: A Dynamic Index Method for Querying XML Data by Tree Structures*, ACM International Conference on Management of Data (SIGMOD), 2003, San Diego, California, USA.

[70] **Haixun Wang**, Wei Fan, Philip S Yu, and Jiawei Han, *Mining Concept-Drifting Data Streams using Ensemble Classifiers*, 9th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), 2003, Washington DC, USA.

[71] **Haixun Wang**, Chang-Shing Perng, Wei Fan, Sanghyun Park, and Philip S Yu, *Indexing Weighted Sequences in Large Databases*, IEEE International Conference on Data Engineering

(ICDE), 2003, Bangalore, India.

[72] **Haixun Wang**, Carlo Zaniolo, and Chang R. Luo, *ATLaS: a Small but Complete SQL Extension for Data Mining and Data Streams (demo)*, 29th International Conference on Very Large Data Bases (VLDB), 2003, Berlin, Germany.

[73] **Haixun Wang** and Carlo Zaniolo, *ATLaS: A Native Extension of SQL for Data Mining*, Second SIAM International Conference on Data Mining (SDM), 2003, San Francisco.

[74] Wei Fan, **Haixun Wang**, Philip S Yu, and Shaw-hwa Lo, *Inductive Learning in Less Than One Sequential Scan*, 18th International Joint Conference on Artificial Intelligence (IJCAI), 2003, Acapulco, Mexico.

[75] Wei Fan, **Haixun Wang**, Philip S Yu, and Sheng Ma, *Is random model better? Its accuracy and efficiency*, 3rd IEEE International Conference on Data Mining (ICDM), 2003, Florida, USA.

[76] Jian Pei, Xiaoling Zhang, Moonjung Cho, **Haixun Wang**, and Philip S. Yu, *MaPle: A Fast Algorithm for Maximal Pattern-based Clustering*, 3rd IEEE International Conference on Data Mining (ICDM), 2003, Florida, USA.

[77] Carlo Zaniolo, Chang R. Luo, Yan N. Law, and Haixun Wang, *Incompleteness of Database Languages for Data Streams and Data Mining*, Invited talk for 11th Italian Symposium on Advanced Database Systems (SEBD), 2003, Cetraro.

[78] Guozhu Dong, Jiawei Han, Laks V. S. Lakshmanan, Jian Pei, **Haixun Wang**, and Philip S. Yu, *Online mining of changes from data streams: Research problems and preliminary results*, ACM SIGMOD Workshop on Management and Processing of Data Streams (MPDS), 2003, San Diego, California, USA.

[79] Jiong Yang, **Haixun Wang**, Wei Wang, and Philip S. Yu, *Enhanced Biclustering on Gene Expression data*, 3rd IEEE Symposium on Bioinformatics and Bioengineering (BIBE), 2003, Washington DC.

[80] Natraj Arni, KayLiang Ong, Shalom Tsur, **Haixun Wang**, and Carlo Zaniolo, *The Deductive Database System LDL++*, Theory and Practice of Logic Programming, 2003, 3(1):61-94.

## 2002

[81] **Haixun Wang**, Wei Wang, Jiong Yang, and Philip S. Yu, *Clustering by Pattern Similarity in Large Data Sets*, ACM International Conference on Management of Data (SIGMOD), 2002, Madison, Wisconsin, USA.

[82] Wei Fan, **Haixun Wang**, Philip S. Yu, and Salvatore Stolfo, *A Framework for Scalable Cost-sensitive Learning Based on Combining Probabilities and Benefits*, Second SIAM International Conference on Data Mining (SDM), 2002, Arlington, USA.

[83] Haixun Wang and Carlo Zaniolo, *ATLaS: a Powerful Database Language and System Based on Simple Extensions of SQL (short paper)*, 18th International Conference on Data Engineering (ICDE), 2002, San Jose, USA.

[84] Jiong Yang, Wei Wang, Haixun Wang, and Philip S. Yu, *Delta-Cluster: Capturing Subspace Correlation in a Large Data Set*, 18th International Conference on Data Engineering (ICDE), 2002, San Jose, USA.



- [85] **Haixun Wang**, Chang-Shing Perng, Sheng Ma, and Philip Yu, *Mining Associations by Pattern Structure in Large Relational Tables*, 2nd IEEE International Conference on Data Mining (ICDM), 2002, Maebashi, Japan.
- [86] Naoki Abe, Edwin Pednault, **Haixun Wang**, Bianca Zadrozny, Wei Fan, and Chidanand Apte, *Empirical Comparison of Various Reinforcement Learning Strategies for Sequential Targeted Marketing*, 2nd IEEE International Conference on Data Mining (ICDM), 2002, Maebashi, Japan.
- [87] Wei Fan, **Haixun Wang**, Philip S. Yu, Shaw-hwa Lo, and Salvatore J. Stolfo, *Progressive Modeling*, 2nd IEEE International Conference on Data Mining (ICDM), 2002, Maebashi, Japan.
- [88] Chang-Shing Perng, **Haixun Wang**, Sheng Ma, and Joseph Hellerstein, *User-directed Exploration of Mining Space with Multiple Attributes*, 2nd IEEE International Conference on Data Mining (ICDM), 2002, Maebashi, Japan.
- [89] Edwin Pednault, Wei Fan, **Haixun Wang**, Naoki Abe, Bianca Zadrozny, and Chidanand Apte, *Sequential Cost-Sensitive Decision Making with Reinforcement Learning*, 8th ACM SIGKDD International Conference on Data Mining, 2002, Edmonton, Canada.
- [90] **Haixun Wang**, Charles Perng, Wei Fan, and Philip S. Yu, *An Indexing Structure for Similarity Searching in Microarray Data*, Proceedings of the First IEEE Computer Society Bioinformatics Conference (CSB 2002), 2002, Palo Alto, California, USA.
- [91] Haixun Wang and Carlo Zaniolo, *Extending SQL for Decision Support Applications*, Design and Management of Data Warehouses (DMDW 2002), 2002, .
- [92] Charles Perng, **Haixun Wang**, Sheng Ma, and Joe Hellerstein, *User-directed Discovery of Patterns in Multi-attribute Data*, KDD Explorations, 2002, Vol. 4, Iss. 1.
- [93] Jiong Yang, Wei Wang, **Haixun Wang**, and Philip S. Yu, *Improving performance of bicluster discovery in a large data set*, Proceedings of the 6th ACM International Conference on Research in Computational Molecular Biology (RECOMB), 2002, .
- [94] Wei Fan, Fang Chu, **Haixun Wang**, and Philip S. Yu, *Pruning and dynamic scheduling of cost-sensitive ensembles*, 18th National Conference on Artificial Intelligence (AAAI-02), 2002, Edmonton, Canada.
- [95] Wei Fan, **Haixun Wang**, Philip S. Yu, and Salvatore Stolfo, *A Fully Distributed Framework for Cost-sensitive Data Mining*, the 22nd International Conference on Distributed Computing Systems (ICDCS), 2002, Vienna, Austria.
- 2001**
- [96] **Haixun Wang** and Philip S. Yu, *SSDT: A Scalable Subspace-Splitting Classifier for Biased Data*, First IEEE International Conference on Data Mining (ICDM), 2001, San Jose, California.
- [97] Charles Perng, **Haixun Wang**, Sheng Ma, and Joe Hellerstein, *FARM: A Framework for Exploring Mining Spaces with Multiple Attributes*, First IEEE International Conference on Data Mining (ICDM), 2001, San Jose, California.
- [98] **Haixun Wang** and Charles Perng, *The S2-Tree: An Index Structure for Subsequence Matching of Spatial Objects*, 5th Pacific-Asic Conference on Knowledge Discovery and Data Mining (PAKDD), 2001, Hong Kong.

**2000**

- [99] **Haixun Wang** and Carlo Zaniolo, *Using SQL to Build New Aggregates and Extenders for Object-Relational Systems*, Proc. 26th Intl. Conf. on Very Large Databases (VLDB), 2000, Cairo, Egypt.
- [100] **Haixun Wang** and Carlo Zaniolo, *Database System Extensions for Decision Support: the AXL Approach*, ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD), 2000, Dallas, TX.
- [101] **Haixun Wang** and Carlo Zaniolo, *CMP: A Fast Decision Tree Classifier Using Multivariate Predictions*, 16th International Conference on Data Engineering (ICDE), 2000, San Diego, USA.
- [102] **Haixun Wang** and Carlo Zaniolo, *User Defined Aggregates in Object-Relational Systems*, 16th International Conference on Data Engineering (ICDE), 2000, San Diego, USA.
- [103] Chang-Shing Perng, **Haixun Wang**, Sylvia R. Zhang, and D. Stott Parker, *Landmarks: a New Model for Similarity-based Pattern Querying in Time Series Databases*, 16th International Conference on Data Engineering (ICDE), 2000, San Diego, USA.
- [104] **Haixun Wang** and Carlo Zaniolo, *Nonmonotonic Reasoning in LDL++: A Second-Generation Deductive Database System*, (Book Chapter) Logic-Based Artificial Intelligence, (J. Minker, Ed.), 2000, .

**1999**

- [105] **Haixun Wang** and Carlo Zaniolo, *User Defined Aggregates in Database Languages*, Seventh International Workshop on Database Programming Languages, 1999, Scotland.
- [106] **Haixun Wang** and Carlo Zaniolo, *User-Defined Aggregates for Datamining*, ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD), 1999, Philadelphia, USA.

**1998**

- [107] **Haixun Wang** and Carlo Zaniolo, *User-Defined Aggregates for Logical Data Languages*, The Sixth International Workshop on Deductive Databases and Logic Programming, 1998, Manchester, UK.
- [108] Carlo Zaniolo and **Haixun Wang**, *Logic-Based User-Defined Aggregates for the Next Generation of Database Systems*, (Book Chapter) The Logic Programming Paradigm: Current Trends and Future Directions, K.R. Apt, V. Marek, M. Truszczynski, D.S.Warren (eds.), 1998, Springer Verlag.

**1996**

- [109] **Haixun Wang**, Jinyuan You, Zhou Wang, and Kan Wang, *An Information Retrieval Algorithm for Database Applications (in Chinese)*, Computer Engineering, 1996, Vol. 22, No.3.