

# Curriculum Vitae

## WEI FAN

IBM T.J.Watson Research  
Hawthorne, NY 10532  
weifan@us.ibm.com, wei.fan@gmail.com (preferred)  
Office +1-914-784-7027  
<http://www.weifan.info>  
<http://www.cs.columbia.edu/~wfan>

## EMPLOYMENT

### IBM Research

Research Staff Member, IBM T.J.Watson Research, Hawthorne, NY 10532, Dec 2000 - Present.

### AT&T Research

Summer Intern, AT&T Research, Florham Park, NJ, Summer 1998/1999

### IBM Research

Summer Intern, IBM T.J.Watson Research, Yorktown Heights, NY, Summer 1995/1996

## EDUCATION

- Ph.D.** May 2001, Computer Science, Columbia University, New York, New York. Thesis: “Cost-sensitive, Scalable and Adaptive Learning using Ensemble-based Methods”. Advisor: Salvatore J. Stolfo. US Patent 7,424,619.
- M.Phil.** October 1997, Computer Science, Columbia University, New York, New York, GPA 4.0/4.0.
- M.Sc.** July 1995, Computer Science, Tsinghua University, Beijing, China. Dissertation: Design and Prototype of Distributed and Parallel C++, Advisor: Weidu Jiang, GPA 3.8/4.0.
- B.Eng.** July 1993, Computer Science & Mechanical Engineering, Tsinghua University, Beijing, China, GPA 3.7/4.0.

## MAIN RESEARCH AREAS AND INTERESTS

- Simple Non-parametric Methods - Random Decision Trees
- Stream Mining
- Sample Selection Bias - Data Shift
- Multiple Model Approach - Ensembles
- Transfer Learning
- Cost-sensitive Learning - Skewed Distribution
- Feature Construction

- Intrusion Detection
- Bioinformatics
- Applications (weather forecasting, manufacturing, traffic, fraud detection, catalogue mailing, late and default payment prediction, etc.)

## **BEST PAPER AWARDS and CONTEST AWARDS**

- 2008 ICDM'08 Data Mining Contest Crown Award (highest awards). Our submission uses my previously proposed random decision tree or RDT (Fan et al., ICDM'03) based approach. Wei Fan, Erheng Zhong, Sihong Xie, Yuzhao Huang, Kun Zhang, Jing Peng and Jiangtao Ren.
- 2006 Best Paper Award (Application Category), "Forecasting Skewed Stochastic Biased Ozone Days: Analyses and Solutions" by Kun Zhang, Wei Fan, Xiaojing Yuan, Ian Davidson, and Xiangshang Li, IEEE International Conference on Data Mining (ICDM'06), December 2006, Hong Kong, China.
- 1997 Runner Up Best Paper Award, "JAM: Java Agents for Meta-learning" by Salvatore Stolfo, Andreas Promdromidis, Shelley Tselepis, Wenke Lee, Wei Fan and Philip Chan, Third International Conference on Knowledge Discovery and Data Mining (KDD-97), Newport Beach, California, August 1997.

## **CITATION H-INDEX**

According to Google scholar, the H-index is 23.

Among these 23 papers

- 2 papers are single-authored
- 5 papers (excluding the 2 single authored) are 1st-authored.
- More than 200 citations: 2 papers
- 100 to 200 citations: 4 papers
- 50 to 100 citations: 10 papers
- 13 papers among the 23 were published after joining IBM in Dec 2000.

## **MAIN APPLICATION AREAS AND INTERESTS**

- Financial Service: various types of fraud detection, trading anomaly detection, money laundry.
- Business Intelligence: marketing catalogue mailing campaign decision making, charity donation solicitation.
- Web User Behavior Mining/Intelligence: personalized campaign based on user behavior, recommendation system.
- Manufacturing Control: chip testing process optimization.
- Intrusion Detection: cost-sensitive intrusion detection, adhoc wireless network anomaly detection.
- Weather Forecasting: ozone pollution detection.
- Bioinformatics: DNA ALU sequence and Maze Helicon sequence detection.

## IMPACT of PRIOR ACADEMIC WORK

PhD thesis work that builds cost-sensitive data mining based intrusion detection, ensemble-based real-time high throughput detection system, anomaly detection system are widely cited, known, patented, licensed by a startup (SystemDetection) and used to build the first data mining intrusion detection prototype. US Patent 7,424,619.

Various works on data stream mining including ensemble-based methods, StreamMiner, systematic data selection, active stream mining, etc, has been widely cited, used in graduate level courses, tutorials and encyclopedia.

Random decision tree (RDT) based work has been demonstrated independently to be more accurate than complex methods such as AdaBoost, SVM for many applications, successfully used to save 50% of testing time for P6 chip CLY test, and won the best application paper awards in ICDM'06. Most recently, our submission based on RDT has won the ICDM08 Data Mining Contest Crown Award (championship),

Delivered various solutions for both financial and business intelligent related services and manufacturing optimization.

## SELECTED PUBLICATIONS

### Conferences - Refereed

- 2009 Erheng Zhong, Wei Fan, Jing Peng, Olivier Verscheure, and Jiangtao Ren, "Universal Learning over Related Distributions and Adaptive Graph Transduction, accepted at The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2009) September 7-11, 2009, Bled, Slovenia
- 2009 Xiaoxiao Shi, Wei Fan, Qiang Yang and Jiangtao Ren, "Relaxed Transfer of Different Classes via Spectral Partition , accepted at The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD 2009) September 7-11, 2009, Bled, Slovenia
- 2009 Erheng Zhong, Wei Fan, Jing Peng, Kun Zhang, Jiangtao Ren, and Olivier Verscheure, "Cross Domain Distribution Adaptation via Kernel Mapping,15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09)
- 2009 Jing Gao, Wei Fan, Yizhou Sun, and Jiawei Han, "Heterogeneous Source Consensus Learning via Decision Propagation and Negotiation, 15th SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'09)
- 2009 Sihong Xie, Wei Fan, Jing Peng, Olivier Verscheure, and Jiangtao Ren, "Latent Space Domain Transfer between High Dimensional Overlapping Distributions??. accepted at 2009 18th International World Wide Web Conference (WWW'09).
- 2008 Erheng Zhong, Sihong Xie, Wei Fan, Jiangtao Ren, Jing Peng, and Kun Zhang "Graph-based Iterative Hybrid Feature Selection", 2008 IEEE International Conference on Data Mining (ICDM'08), Pisa, Italy.
- 2008 Xiaoxiao Shi, Wei Fan, and Jiangtao Ren "Actively Transfer Domain Knowledge", 2008 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD'08), Antwerp, Belgium.

- 2008 Wei Fan, Kun Zhang, Hong Cheng, Jing Gao, Xifeng Yan, Jiawei Han, Philip Yu, and Olivier Verscheure “Direct Mining of Discriminative and Essential Frequent Patterns via Model-based Search Tree”, 2008 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’08), Las Vegas, NV.
- 2008 Jing Gao, Wei Fan, Jing Jiang, and Jiawei Han “Knowledge Transfer via Multiple Model Local Structure Mapping”, 2008 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’08), Las Vegas, NV.
- 2008 Jiangtao Ren, Xiaoxiao Shi, Wei Fan, and Philip S. Yu “Type Independent Correction of Sample Selection Bias via Structural Discovery and Re-balancing”, 2008 SIAM International Conference on Data Mining (SDM’08), Atlanta, GA, Apr 2008.
- 2008 Jiangtao Ren, Zhengyuan Qiu, Wei Fan, and Philip S. Yu “Forward Semi-supervised Feature Selection”, 2008 Pacific-Asian Conference on Knowledge Discovery and Data Mining (PAKDD’08), Osaka, Japan, April 2008.
- 2007 Jing Gao, Wei Fan, and Jiawei Han “On Appropriate Assumptions to Mine Data Streams: Analysis and Practice”, 2007 IEEE International Conference on Data Mining (ICDM’07), Omaha, NE, Oct 2007.
- 2007 Kun-Lung Wu, Philip S. Yu, Bugra Gedik, Kirsten Hildrum, Charu Aggarwal, Eric Bouillet, Wei Fan, David George, Xiaohui Gu, Gang Luo, Haixun Wang “Challenges and Experiences in Prototyping a Multi-Modal Stream Analytic and Monitoring Application on System S”, 2007 Very Large DataBase Conference (VLDB’07), Vienna, Austria, Sep 2007.
- 2007 Wei Fan, and Ian Davidson “On Sample Selection Bias and Its Efficient Correction via Model Averaging and Unlabeled Examples”, 2007 SIAM International Conference on Data Mining (SDM’07), Minneapolis, MN, April 2007.
- 2007 Jing Gao, Wei Fan, Jiawei Han, and Philip S. Yu, “A General Framework for Mining Concept-Drifting Streams with Skewed Distribution”, 2007 SIAM International Conference on Data Mining (SDM’07), Minneapolis, MN, April 2007.
- 2006 Kun Zhang, Wei Fan, Xiaojing Yuan, Ian Davidson, and Xiangshang Li, “Forecasting Skewed Stochastic Biased Ozone Days: Analyses and Solutions”, 2006 IEEE International Conference on Data Mining (ICDM’2006), December 2006. *Best Paper Award: Application Category*
- 2006 Kun Zhang, Wei Fan, Bill Buckles, Xiaojing Yuan, and Zujia Xu: “Discovering Unrevealed Properties of Probability Estimation Trees: On Algorithm Selection and Performance Explanation”, 2006 IEEE International Conference on Data Mining (ICDM’2006), December 2006.
- 2006 Ian Davidson, and Wei Fan “When Efficient Model Averaging Out-performs Boosting and Bagging”, PKDD 2006, Berlin, September 2006.
- 2006 Wei Fan, Joe McClosky, and Philip S. Yu, “A General Framework for Accuracy and Fast Regression by Data Summarization in Random Decision Trees”, KDD2006, Philadelphia, August 2006.
- 2006 Wei Fan, and Ian Davidson “ReverseTesting: An Efficient Framework to Select Amongst Classifiers under Sample Selection Bias”, KDD2006, Philadelphia, August 2006.
- 2005 Wei Fan, Ed Greengrass, Joe McCloskey, and Philip S. Yu “Effective Estimation of Posterior Probabilities: Explaining the Accuracy of Randomized Decision Tree Approaches”. The Fifth IEEE International Conference on Data Mining (ICDM’05), Houston, Texas, November 2005.
- 2005 Wei Fan, Ian Davidson, Bianca Zadrozny, and Philip S. Yu “An Improved Categorization of Classifier’s Sensitivity on Sample Selection Bias”. The Fifth IEEE International Conference on Data Mining (ICDM’05), Houston, Texas, November 2005.

- 2005 Wei Fan, Janek Mathuria, and Chang-tien Lu “Making Data Mining Models Useful to Model Non-paying Customers of Exchange Carriers”. 2005 SIAM International Conference on Data Mining (SDM’05), Newport Beach, CA, April 2005.
- 2005 Fei Tony Liu, Kai Ming Ting, and Wei Fan “Maximizing Tree Diversity by Building Complete-Random Decision Trees” (short paper), 2005 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD’05), Hanoi, Vietnam, May 2005.
- 2004 Wei Fan “On the Optimality of Probability Estimation by Random Decision Trees”, The Nineteenth National Conference on Artificial Intelligence (AAAI’04), San Jose, CA, July, 2004.
- 2004 Wei Fan “StreamMiner: A Classifier-Ensemble based Engine to Mine Concept-Drifting Data Streams” (demo), The 2004 International Conference on Very Large Database (VLDB’04), Toronto, Canada, August 2004.
- 2004 Wei Fan “Systematic Data Selection to Mine Concept-drifting Data Streams”, The Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’05), Seattle, Washington, August 2004.
- 2004 Wei Fan, Yi-an Huang, Haixun Wang, and Philip S. Yu “Active Mining of Data Streams” (short paper), 2004 SIAM International Conference on Data Mining (SDM’04), Orlando, Florida, April 2004.
- 2004 Wei Fan, Yi-an Huang, and Philip S. Yu “Decision Tree Evolution with Limited Number of Examples from Drifting Data Streams” (short paper), The Fourth IEEE International Conference on Data Mining (ICDE’04), Brighton, UK, November 2004.
- 2004 Wei Fan, Philip S. Yu, and Haixun Wang “Mining Extremely Skewed Trading Anomalies” (industrial track), The Ninth International Conference on Extending Database Technology (EDBT’04), Heraklion, Greece, March 2004.
- 2004 Haixun Wang, Fang Chu, Wei Fan, and Philip S. Yu “A Fast Algorithm for Subspace Clustering by Pattern Similarity”, 16th International Conference on Scientific and Statistical Database Management (SSDBM 2004), 21-23 June 2004, Santorini Island, Greece.
- 2003 Yi-an Huang, Wei Fan, Wenke Lee, Philip S. Yu “Cross-Feature Analysis for Detecting Ad-Hoc Routing Anomalies”, 23rd International Conference on Distributed Computing Systems (ICDCS 2003), 19-22 May 2004, Providence, RI.
- 2003 Haixun Wang, Chang-Shing Perng, Wei Fan, Sanghyun Park, Philip S. Yu “Indexing Weighted-Sequences in Large Databases”, 19th IEEE International Conference on Data Engineering (ICDE 2003), March 5-8, 2003, Bangalore, India: 63-74.
- 2003 Wei Fan, Haixun Wang, Philip S. Yu, Sheng Ma “Is random model better? On its accuracy and efficiency”, 3rd IEEE International Conference on Data Mining (ICDM 2003), November 19-22, Melbourne, FL: 51-58.
- 2003 Wei Fan, Haixun Wang, Philip S. Yu, Shaw-hwa Lo “Inductive Learning in Less Than One Sequential Data Scan”, 18th International Joint Conference on Artificial Intelligence (IJCAI 2003), August 15-19, 2003, Acapulco, Mexico: 595-600.
- 2003 Haixun Wang, Wei Fan, Philip S. Yu, Jiawei Han “Mining concept-drifting data streams using ensemble classifiers”, 9th ACM International Conference on Knowledge Discovery and Data Mining (KDD 2003), August 24-27, 2003, Washington DC: 226-235.
- 2003 Haixun Wang, Sanghyun Park, Wei Fan, Philip S. Yu “ViST: A Dynamic Index Method for Querying XML Data by Tree Structures” 2003 ACM SIGMOD International Conference on Management of Data (SIGMOD 2003): 110-121.
- 2002 Wei Fan, Fang Chu, Haixun Wang, Philip S. Yu “Pruning and Dynamic Scheduling of Cost-Sensitive Ensembles”, Eighteenth National Conference on Artificial Intelligence (AAAI 2002): 146-151.

- 2002 Haixun Wang, Chang-Shing Perng, Wei Fan, Philip S. Yu “An Index Structure for Pattern Similarity Searching in DNA Microarray Data”, 1st IEEE Bioinformatics Conference CSB 2002: 256-267.
- 2002 Wei Fan, Haixun Wang, Philip S. Yu, Salvatore J. Stolfo “A Fully Distributed Framework for Cost-Sensitive Data Mining” 22nd International Conference on Distributed Computing Systems (ICDCS 2002), July 2-5, 2002, Vienna, Austria: 445-446.
- 2002 Wei Fan, Haixun Wang, Philip S. Yu, Shaw-hwa Lo, Salvatore J. Stolfo “Progressive Modelling”. 2nd IEEE International Conference on Data Mining (ICDM 2002), December 9-12, 2003, Maebashi, Japan: 163-170.
- 2002 Naoki Abe, Edwin P. D. Pednault, Haixun Wang, Bianca Zadrozny, Wei Fan, Chidanand Apte “Empirical Comparison of Various Reinforcement Learning Strategies for Sequential Targeted Marketing”, 2nd IEEE International Conference on Data Mining (ICDM 2002), December 9-12, 2002, Maebashi, Japan: 3-10.
- 2002 Edwin P. D. Pednault, Naoki Abe, Bianca Zadrozny, Haixun Wang, Wei Fan, and Chidnant Apte “Sequential Cost-sensitive Decision Making with Reinforcement Learning”, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’02), Edmonton, Canada.
- 2002 Wei Fan, Haixun Wang, Philip S. Yu, Salvatore J. Stolfo “A Framework for Scalable Cost-sensitive Learning Based on Combining Probabilities and Benefits”, 2nd SIAM International Conference on Data Mining (SDM 2002), April 11-13, 2002, Arlington, VA.
- 2002 Wei Fan, Salvatore J. Stolfo “Ensemble-based Adaptive Intrusion Detection”, 2nd SIAM International Conference on Data Mining (SDM 2002), April 11-13, 2002, Arlington, VA.
- 2001 Wei Fan, Matthew Miller, Salvatore J. Stolfo, Wenke Lee, Philip K. Chan “Using Artificial Anomalies to Detect Unknown and Known Network Intrusions”, 1st IEEE International Conference on Data Mining (ICDM 2001), 29 Nov- 1 Dec, 2001, San Jose, CA: 123-130.
- 2001 Wenke Lee, Salvatore J. Stolfo, Philip K. Chan, Eleazar Eskin, Wei Fan, Matthew Miller, Shlomo Hershkop, and Junxin Zhang “Real-time Data Mining-based Intrusion Detection”, Second DARPA Information Survivability and Exposition (DISCEX’01), June 2001, pp, 89-100.
- 2000 Wei Fan, Wenke Lee, Salvatore Stolfo, Matthew Miller “A Multiple Model Cost-Sensitive Approach for Intrusion Detection” The Eleventh European Conference on Machine Learning (ECML00).
- 2000 William Cohen, Wei Fan “Web-Collaborative Filtering: Recommending Music by Spidering The Web” The Ninth International Conference on World Wide Web (accepted and to appear).
- 2000 Salvatore J. Stolfo, Wenke Lee, Wei Fan, Andreas Prodromidis, and Philip K. Chan “Cost-based Modelling for Fraud and Intrusion Detection: Results from the JAM Project” DARPA Information Survivability Conference 2000 (accepted and to appear).
- 1999 Wei Fan, Salvatore J. Stolfo, Junxin Zhang and Philip Chan “AdaCost: Misclassification Cost-sensitive Learning” Proceedings of the Sixteenth International Conference on Machine Learning (ICML’99), pp.97-105, Bled, Slovenia, June 1999.
- 1999 Wei Fan, Salvatore J. Stolfo and Junxin Zhang “The Application of AdaBoost for Distributed, Scalable and On-line Learning” Proceedings of ACM Sixth International Conference on Knowledge Discovery and Data Mining (KDD’99). pp.362-366, San Diego, California, August 1999.
- 1997 Salvatore J. Stolfo, Andreas Prodromidis, Shelley Tselepis, Wenke Lee, Wei Fan, and Philip Chan “JAM - Java Agents of Meta-learning over Distributed Networks” Proceeding of Third International Conference on Knowledge Discovery and Data Mining (KDD-97), pp.74 - 81, Newport Beach, California, August 1997. *Runner-up best application paper*. The JAM Project is a featured article in Trusted Information Systems Magazine, March, 1998

- 1996 Jose Moreira, Vijay Naik and Wei Fan “Design and Implementation of Computational Steering for Parallel Scientific Applications”, Proceedings of the Eighth SIAM Conference on Parallel Processing for Scientific Computing, Minneapolis, Minnesota, March 14–17, 1997.

### **Journals and Magazines - Refereed**

- 2008 Jing Gao, Boling Ding, Wei Fan, and Jiawei Han: “Classifying Data Streams with Skewed Class Distribution and Concept Drifts”, IEEE Internet Computing. accepted to appear.
- 2008 Kun Zhang, and Wei Fan: “Forecasting Skewed Biased Stochastic Ozone Days: analyses, solutions and beyond”, Knowl. Inf. Syst. 14(3):507-527 (2008).
- 2004 Wei Fan, Matthew Miller, Salvatore J. Stolfo, Wenke Lee, Philip K. Chan: “Using artificial anomalies to detect unknown and known network intrusions”. Knowl. Inf. Syst. 6(5): 507-527 (2004)
- 2002 Wenke Lee, Wei Fan, Matthew Miller, Salvatore J. Stolfo, Erez Zadok “Toward Cost-Sensitive Modelling for Intrusion Detection and Response”. Journal of Computer Security 10(1/2): 5-22 (2002)
- 2001 Wenke Lee, and Wei Fan “Mining System Audit Data: Opportunities and Challenges”, SIGMOD Record 30(4):34-44 (2001)
- 2001 Salvatore J. Stolfo, Wenke Lee, Philip K. Chan, Wei Fan, and Eleazar Eskin “Data Mining-based Intrusion Detection: An Overview of the Columbia IDS project”, SIGMOD Record 30(4):5-14 (2001).
- 2000 William W. Cohen, Wei Fan: “Web-collaborative filtering: recommending music by crawling the Web”. Computer Networks 33(1-6): 685-698 (2000)
- 1999 Philip Chan, Andreas Prodromidis, Wei Fan, and Salvatore J. Stolfo “Distributed Data Mining in Credit Card Fraud Detection” IEEE Intelligent Systems Journal November/December 1999, pp.67-74.
- 1999 William Cohen and Wei Fan “Learning Page-independent Heuristics for Extracting Data from Web-pages”, International Journal of Computer and Telecommunication Networking, Vol.31, pp.1641-1652.

### **Workshops - Refereed**

- 1999 Wei Fan, Philip K. Chan and Salvatore J. Stolfo “Using Conflicts among Base Classifiers to Measure the Performance of Stacking”, Proceedings of Recent Advances in Meta-learning and Future Work, pp.10-17. (ICML-99), Bled, Slovenia, June 1999.
- 1997 Salvatore J. Stolfo, Wei Fan, Wenke Lee and Philip Chan “Credit Card Fraud Detection using Meta-learning: Issues and Initial Results”, AI Approaches to Fraud Detection and Risk Management Workshop, pp.83-90, AAAI-97, Providence, Rhode Island, July 1997.
- 1996 Wei Fan, Philip Chan and Salvatore J. Stolfo “A Comparative Evaluation of Combiner and Stacked Generalizations”, Proceedings of Integrating Multiple Learning Models for Improving and Scaling Machine Learning Algorithms Workshop (AAAI-96), pp.40-47, Portland, Oregon, 1996.

### **Book Chapter - Peer Refereed**

- 2004 One Chapter in “Next Generation of Data-Mining Applications” edited by Mehmet Kantardzic and Jozef Zurada.
- Chapter II.6 “A Fully Distributed Framework for Cost-sensitive Data Mining”

- 1995 Three Chapters in the book “Object-Oriented Programming Language C++”, Tsinghua University Press, Beijing, China, August 1995. ISBN 7-302-01734-4/TP.759
- Chapter 1 Object-Oriented Methodology and Introduction to C++
  - Chapter 4 Class
  - Chapter 7 Class Inheritance

## SELECTED PUBLICATIONS - listed by main topic

### Data Stream Mining

- 2008 Jing Gao, Boling Ding, Wei Fan, and Jiawei Han: “Classifying Data Streams with Skewed Class Distribution and Concept Drifts”, IEEE Internet Computing. accepted to appear.
- 2008 Kun Zhang, and Wei Fan: “Forecasting Skewed Biased Stochastic Ozone Days: analyses, solutions and beyond”, Knowl. Inf. Syst. 14(3):507-527 (2008).
- 2007 Kun-Lung Wu, Philip S. Yu, Bugra Gedik, Kirsten Hildrum, Charu Aggarwal, Eric Bouillet, Wei Fan, David George, Xiaohui Gu, Gang Luo, Haixun Wang “Challenges and Experiences in Prototyping a Multi-Modal Stream Analytic and Monitoring Application on System S”, 2007 Very Large DataBase Conference (VLDB’07), Vienna, Austria, Sep 2007.
- 2007 Jing Gao, Wei Fan, and Jiawei Han “On Appropriate Assumptions to Mine Data Streams: Analysis and Practice”, 2007 IEEE International Conference on Data Mining (ICDM’07), Omaha, NE, Oct 2007.
- 2007 Jing Gao, Wei Fan, Jiawei Han, and Philip S. Yu, “A General Framework for Mining Concept-Drifting Streams with Skewed Distribution”, 2007 SIAM International Conference on Data Mining (SDM’07), Minneapolis, MN, April 2007.
- 2006 Kun Zhang, Wei Fan, Xiaojing Yuan, Ian Davidson, and Xiangshang Li, “Forecasting Skewed Stochastic Biased Ozone Days: Analyses and Solutions”, 2006 IEEE International Conference on Data Mining (ICDM’2006), December 2006. *Best Paper Award: Application Category*
- 2004 Wei Fan “StreamMiner: A Classifier-Ensemble based Engine to Mine Concept-Drifting Data Streams” (demo), The 2004 International Conference on Very Large Database (VLDB’04), Toronto, Canada, August 2004.
- 2004 Wei Fan “StreamMiner: A Classifier-Ensemble based Engine to Mine Concept-Drifting Data Streams” (demo), The 2004 International Conference on Very Large Database (VLDB’04), Toronto, Canada, August 2004.
- 2004 Wei Fan, Yi-an Huang, Haixun Wang, and Philip S. Yu “Active Mining of Data Streams” (short paper), 2004 SIAM International Conference on Data Mining (SDM’04), Orlando, Florida, April 2004.
- 2004 Wei Fan, Yi-an Huang, and Philip S. Yu “Decision Tree Evolution with Limited Number of Examples from Drifting Data Streams” (short paper), The Fourth IEEE International Conference on Data Mining (ICDE’04), Brighton, UK, November 2004.
- 2003 Haixun Wang, Wei Fan, Philip S. Yu, Jiawei Han “Mining concept-drifting data streams using ensemble classifiers”, 9th ACM International Conference on Knowledge Discovery and Data Mining (KDD 2003), August 24-27, 2003, Washington DC: 226-235.

### Sample Selection Bias

- 2008 Wei Fan and Masashi Sugiyama, “Sample Selection Bias - Covariate Shift: Problems, Solutions and Applications (Tutorial)”, 2008 IEEE International Conference on Data Mining (ICDM’08), Pisa, Italy.
- 2008 Jiangtao Ren, Xiaoxiao Shi, Wei Fan, and Philip S. Yu “Type Independent Correction of Sample Selection Bias via Structural Discovery and Re-balancing”, 2008 SIAM International Conference on Data Mining (SDM’08), Atlanta, GA, Apr 2008.
- 2008 Kun Zhang, and Wei Fan: “Forecasting Skewed Biased Stochastic Ozone Days: analyses, solutions and beyond”, *Knowl. Inf. Syst.* 14(3):507-527 (2008).
- 2006 Ian Davidson, and Wei Fan “When Efficient Model Averaging Out-performs Boosting and Bagging”, PKDD 2006, Berlin, September 2006.
- 2007 Wei Fan, and Ian Davidson “On Sample Selection Bias and Its Efficient Correction via Model Averaging and Unlabeled Examples”, 2007 SIAM International Conference on Data Mining (SDM’07), Minneapolis, MN, April 2007.
- 2006 Kun Zhang, Wei Fan, Xiaojing Yuan, Ian Davidson, and Xiangshang Li, “Forecasting Skewed Stochastic Biased Ozone Days: Analyses and Solutions”, 2006 IEEE International Conference on Data Mining (ICDM’2006), December 2006. *Best Paper Award: Application Category*
- 2006 Wei Fan, and Ian Davidson “ReverseTesting: An Efficient Framework to Select Amongst Classifiers under Sample Selection Bias”, KDD2006, Philadelphia, August 2006.
- 2005 Wei Fan, Ian Davidson, Bianca Zadrozny, and Philip S. Yu “An Improved Categorization of Classifier’s Sensitivity on Sample Selection Bias”. The Fifth IEEE International Conference on Data Mining (ICDM’05), Houston,

## **Transfer Learning**

- 2009 Sihong Xie, Wei Fan, Jing Peng, Olivier Verscheure, and Jiangtao Ren, “Latent Space Domain Transfer between High Dimensional Overlapping Distributions??”, accepted at 2009 18th International World Wide Web Conference (WWW’09).
- 2008 Xiaoxiao Shi, Wei Fan, and Jiangtao Ren “Actively Transfer Domain Knowledge”, 2008 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD’08), Antwerp, Belgium.
- 2008 Jing Gao, Wei Fan, Jing Jiang, and Jiawei Han “Knowledge Transfer via Multiple Model Local Structure Mapping”, 2008 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’08), Las Vegas, NV.

## **Feature Construction**

- 2008 Wei Fan, Kun Zhang, Hong Cheng, Jing Gao, Xifeng Yan, Jiawei Han, Philip Yu, and Olivier Verscheure “Direct Mining of Discriminative and Essential Frequent Patterns via Model-based Search Tree”, 2008 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’08), Las Vegas, NV.

## **Ensemble Methods**

- 2008 Jing Gao, Wei Fan, Jing Jiang, and Jiawei Han “Knowledge Transfer via Multiple Model Local Structure Mapping”, 2008 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’08), Las Vegas, NV.

- 2007 Jing Gao, Wei Fan, and Jiawei Han “On Appropriate Assumptions to Mine Data Streams: Analysis and Practice”, 2007 IEEE International Conference on Data Mining (ICDM’07), Omaha, NE, Oct 2007.
- 2007 Wei Fan, and Ian Davidson “On Sample Selection Bias and Its Efficient Correction via Model Averaging and Unlabeled Examples”, 2007 SIAM International Conference on Data Mining (SDM’07), Minneapolis, MN, April 2007.
- 2007 Jing Gao, Wei Fan, Jiawei Han, and Philip S. Yu, “A General Framework for Mining Concept-Drifting Streams with Skewed Distribution”, 2007 SIAM International Conference on Data Mining (SDM’07), Minneapolis, MN, April 2007.
- 2006 Ian Davidson, and Wei Fan “When Efficient Model Averaging Out-performs Boosting and Bagging”, PKDD 2006, Berlin, September 2006.
- 2006 Wei Fan, Joe McClosky, and Philip S. Yu, “A General Framework for Accuracy and Fast Regression by Data Summarization in Random Decision Trees”, KDD2006, Philadelphia, August 2006.
- 2005 Fei Tony Liu, Kai Ming Ting, and Wei Fan “Maximizing Tree Diversity by Building Complete-Random Decision Trees” (short paper), 2005 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD’05), Hanoi, Vietnam, May 2005.
- 2004 Wei Fan “On the Optimality of Probability Estimation by Random Decision Trees”, The Nineteenth National Conference on Artificial Intelligence (AAAI’04), San Jose, CA, July, 2004.
- 2004 Wei Fan “StreamMiner: A Classifier-Ensemble based Engine to Mine Concept-Drifting Data Streams” (demo), The 2004 International Conference on Very Large Database (VLDB’04), Toronto, Canada, August 2004.
- 2004 Wei Fan “Systematic Data Selection to Mine Concept-drifting Data Streams”, The Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’05), Seattle, Washington, August 2004.
- 2004 Wei Fan, Philip S. Yu, and Haixun Wang “Mining Extremely Skewed Trading Anomalies” (industrial track), The Ninth International Conference on Extending Database Technology (EDBT’04), Heraklion, Greece, March 2004.
- 2003 Yi-an Huang, Wei Fan, Wenke Lee, Philip S. Yu “Cross-Feature Analysis for Detecting Ad-Hoc Routing Anomalies”, 23rd International Conference on Distributed Computing Systems (ICDCS 2003), 19-22 May 2004, Providence, RI.
- 2003 Wei Fan, Haixun Wang, Philip S. Yu, Sheng Ma “Is random model better? On its accuracy and efficiency”, 3rd IEEE International Conference on Data Mining (ICDM 2003), November 19-22, Melbourne, FL: 51-58.
- 2003 Wei Fan, Haixun Wang, Philip S. Yu, Shaw-hwa Lo “Inductive Learning in Less Than One Sequential Data Scan”, 18th International Joint Conference on Artificial Intelligence (IJCAI 2003), August 15-19, 2003, Acapulco, Mexico: 595-600.
- 2003 Haixun Wang, Wei Fan, Philip S. Yu, Jiawei Han “Mining concept-drifting data streams using ensemble classifiers”, 9th ACM International Conference on Knowledge Discovery and Data Mining (KDD 2003), August 24-27, 2003, Washington DC: 226-235.
- 2002 Wei Fan, Fang Chu, Haixun Wang, Philip S. Yu “Pruning and Dynamic Scheduling of Cost-Sensitive Ensembles”, Eighteenth National Conference on Artificial Intelligence (AAAI 2002): 146-151.
- 2002 Wei Fan, Haixun Wang, Philip S. Yu, Salvatore J. Stolfo “A Fully Distributed Framework for Cost-Sensitive Data Mining” 22nd International Conference on Distributed Computing Systems (ICDCS 2002), July 2-5, 2002, Vienna, Austria: 445-446.

- 2002 Wei Fan, Haixun Wang, Philip S. Yu, Shaw-hwa Lo, Salvatore J. Stolfo “Progressive Modelling”. 2nd IEEE International Conference on Data Mining (ICDM 2002), December 9-12, 2003, Maebashi, Japan: 163-170.
- 2002 Wei Fan, Haixun Wang, Philip S. Yu, Salvatore J. Stolfo “A Framework for Scalable Cost-sensitive Learning Based on Combining Probabilities and Benefits”, 2nd SIAM International Conference on Data Mining (SDM 2002), April 11-13, 2002, Arlington, VA.
- 2002 Wei Fan, Salvatore J. Stolfo “Ensemble-based Adaptive Intrusion Detection”, 2nd SIAM International Conference on Data Mining (SDM 2002), April 11-13, 2002, Arlington, VA.
- 2000 Wei Fan, Wenke Lee, Salvatore Stolfo, Matthew Miller “A Multiple Model Cost-Sensitive Approach for Intrusion Detection” The Eleventh European Conference on Machine Learning (ECML00).
- 1999 Wei Fan, Salvatore J. Stolfo, Junxin Zhang and Philip Chan “AdaCost: Misclassification Cost-sensitive Learning” Proceedings of the Sixteenth International Conference on Machine Learning (ICML'99), pp.97-105, Bled, Slovenia, June 1999.
- 1999 Wei Fan, Salvatore J. Stolfo and Junxin Zhang “The Application of AdaBoost for Distributed, Scalable and On-line Learning” Proceedings of ACM Sixth International Conference on Knowledge Discovery and Data Mining (KDD'99). pp.362-366, San Diego, California, August 1999.
- 1997 Salvatore J. Stolfo, Andreas Prodromidis, Shelley Tselepis, Wenke Lee, Wei Fan, and Philip Chan “JAM - Java Agents of Meta-learning over Distributed Networks” Proceeding of Third International Conference on Knowledge Discovery and Data Mining (KDD-97), pp.74 - 81, Newport Beach, California,

### **Cost-sensitive Learning Skewed Distribution**

- 2008 Jing Gao, Boling Ding, Wei Fan, and Jiawei Han: “Classifying Data Streams with Skewed Class Distribution and Concept Drifts”, IEEE Internet Computing. accepted to appear.
- 2007 Jing Gao, Wei Fan, Jiawei Han, and Philip S. Yu, “A General Framework for Mining Concept-Drifting Streams with Skewed Distribution”, 2007 SIAM International Conference on Data Mining (SDM'07), Minneapolis, MN, April 2007.
- 2006 Kun Zhang, Wei Fan, Xiaojing Yuan, Ian Davidson, and Xiangshang Li, “Forecasting Skewed Stochastic Biased Ozone Days: Analyses and Solutions”, 2006 IEEE International Conference on Data Mining (ICDM'2006), December 2006. *Best Paper Award: Application Category*
- 2005 Wei Fan, Janek Mathuria, and Chang-tien Lu “Making Data Mining Models Useful to Model Non-paying Customers of Exchange Carriers”. 2005 SIAM International Conference on Data Mining (SDM'05), Newport Beach, CA, April 2005.
- 2004 Wei Fan, Yi-an Huang, Haixun Wang, and Philip S. Yu “Active Mining of Data Streams” (short paper), 2004 SIAM International Conference on Data Mining (SDM'04), Orlando, Florida, April 2004.
- 2004 Wei Fan, Yi-an Huang, and Philip S. Yu “Decision Tree Evolution with Limited Number of Examples from Drifting Data Streams” (short paper), The Fourth IEEE International Conference on Data Mining (ICDE'04), Brighton, UK, November 2004.
- 2004 Wei Fan, Philip S. Yu, and Haixun Wang “Mining Extremely Skewed Trading Anomalies” (industrial track), The Ninth International Conference on Extending Database Technology (EDBT'04), Heraklion, Greece, March 2004.
- 2003 Wei Fan, Haixun Wang, Philip S. Yu, Shaw-hwa Lo “Inductive Learning in Less Than One Sequential Data Scan”, 18th International Joint Conference on Artificial Intelligence (IJCAI 2003), August 15-19, 2003, Acapulco, Mexico: 595-600.

- 2002 Wei Fan, Fang Chu, Haixun Wang, Philip S. Yu “Pruning and Dynamic Scheduling of Cost-Sensitive Ensembles”, Eighteenth National Conference on Artificial Intelligence (AAAI 2002): 146-151.
- 2002 Wei Fan, Haixun Wang, Philip S. Yu, Salvatore J. Stolfo “A Fully Distributed Framework for Cost-Sensitive Data Mining” 22nd International Conference on Distributed Computing Systems (ICDCS 2002), July 2-5, 2002, Vienna, Austria: 445-446.
- 2002 Wei Fan, Haixun Wang, Philip S. Yu, Shaw-hwa Lo, Salvatore J. Stolfo “Progressive Modelling”. 2nd IEEE International Conference on Data Mining (ICDM 2002), December 9-12, 2003, Maebashi, Japan: 163-170.
- 2002 Naoki Abe, Edwin P. D. Pednault, Haixun Wang, Bianca Zadrozny, Wei Fan, Chidanand Apte “Empirical Comparison of Various Reinforcement Learning Strategies for Sequential Targeted Marketing”, 2nd IEEE International Conference on Data Mining (ICDM 2002), December 9-12, 2002, Maebashi, Japan: 3-10.
- 2002 Edwin P. D. Pednault, Naoki Abe, Bianca Zadrozny, Haixun Wang, Wei Fan, and Chidnant Apte “Sequential Cost-sensitive Decision Making with Reinforcement Learning”, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’02), Edmonton, Canada.
- 2002 Wei Fan, Haixun Wang, Philip S. Yu, Salvatore J. Stolfo “A Framework for Scalable Cost-sensitive Learning Based on Combining Probabilities and Benefits”, 2nd SIAM International Conference on Data Mining (SDM 2002), April 11-13, 2002, Arlington, VA.
- 1999 Wei Fan, Salvatore J. Stolfo, Junxin Zhang and Philip Chan “AdaCost: Misclassification Cost-sensitive Learning” Proceedings of the Sixteenth International Conference on Machine Learning (ICML’99), pp.97-105, Bled, Slovenia, June 1999.

## **Feature Selection**

- 2008 Erheng Zhong, Sihong Xie, Wei Fan, Jiangtao Ren, Jing Peng, and Kun Zhang “Graph-based Iterative Hybrid Feature Selection”, 2008 IEEE International Conference on Data Mining (ICDM’08), Pisa, Italy.
- 2008 Jiangtao Ren, Zhengyuan Qiu, Wei Fan, and Philip S. Yu “Forward Semi-supervised Feature Selection”, 2008 Pacific-Asian Conference on Knowledge Discovery and Data Mining (PAKDD’08), Osaka, Japan, April 2008.

## **Fraud Detection**

- 2007 Wei Fan, and Ian Davidson “On Sample Selection Bias and Its Efficient Correction via Model Averaging and Unlabeled Examples”, 2007 SIAM International Conference on Data Mining (SDM’07), Minneapolis, MN, April 2007.
- 2007 Kun-Lung Wu, Philip S. Yu, Bugra Gedik, Kirsten Hildrum, Charu Aggarwal, Eric Bouillet, Wei Fan, David George, Xiaohui Gu, Gang Luo, Haixun Wang “Challenges and Experiences in Prototyping a Multi-Modal Stream Analytic and Monitoring Application on System S”, 2007 Very Large DataBase Conference (VLDB’07), Vienna, Austria, Sep 2007.
- 2006 Wei Fan, and Ian Davidson “ReverseTesting: An Efficient Framework to Select Amongst Classifiers under Sample Selection Bias”, KDD2006, Philadelphia, August 2006.
- 2005 Wei Fan, Ed Greengrass, Joe McCloskey, and Philip S. Yu “Effective Estimation of Posterior Probabilities: Explaining the Accuracy of Randomized Decision Tree Approaches”. The Fifth IEEE International Conference on Data Mining (ICDM’05), Houston, Texas, November 2005.

- 2005 Wei Fan, Janek Mathuria, and Chang-tien Lu “Making Data Mining Models Useful to Model Non-paying Customers of Exchange Carriers”. 2005 SIAM International Conference on Data Mining (SDM’05), Newport Beach, CA, April 2005.
- 2004 Wei Fan “On the Optimality of Probability Estimation by Random Decision Trees”, The Nineteenth National Conference on Artificial Intelligence (AAAI’04), San Jose, CA, July, 2004.
- 2004 Wei Fan “StreamMiner: A Classifier-Ensemble based Engine to Mine Concept-Drifting Data Streams” (demo), The 2004 International Conference on Very Large Database (VLDB’04), Toronto, Canada, August 2004.
- 2004 Wei Fan “Systematic Data Selection to Mine Concept-drifting Data Streams”, The Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’05), Seattle, Washington, August 2004.
- 2004 Wei Fan, Yi-an Huang, Haixun Wang, and Philip S. Yu “Active Mining of Data Streams” (short paper), 2004 SIAM International Conference on Data Mining (SDM’04), Orlando, Florida, April 2004.
- 2004 Wei Fan, Yi-an Huang, and Philip S. Yu “Decision Tree Evolution with Limited Number of Examples from Drifting Data Streams” (short paper), The Fourth IEEE International Conference on Data Mining (ICDE’04), Brighton, UK, November 2004.
- 2004 Wei Fan, Philip S. Yu, and Haixun Wang “Mining Extremely Skewed Trading Anomalies” (industrial track), The Ninth International Conference on Extending Database Technology (EDBT’04), Heraklion, Greece, March 2004.
- 2004 Wei Fan “On the Optimality of Probability Estimation by Random Decision Trees”, The Nineteenth National Conference on Artificial Intelligence (AAAI’04), San Jose, CA, July, 2004.
- 2003 Wei Fan, Haixun Wang, Philip S. Yu, Sheng Ma “Is random model better? On its accuracy and efficiency”, 3rd IEEE International Conference on Data Mining (ICDM 2003), November 19-22, Melbourne, FL: 51-58.
- 2003 Wei Fan, Haixun Wang, Philip S. Yu, Shaw-hwa Lo “Inductive Learning in Less Than One Sequential Data Scan”, 18th International Joint Conference on Artificial Intelligence (IJCAI 2003), August 15-19, 2003, Acapulco, Mexico: 595-600.
- 2003 Haixun Wang, Wei Fan, Philip S. Yu, Jiawei Han “Mining concept-drifting data streams using ensemble classifiers”, 9th ACM International Conference on Knowledge Discovery and Data Mining (KDD 2003), August 24-27, 2003, Washington DC: 226-235.
- 2002 Wei Fan, Fang Chu, Haixun Wang, Philip S. Yu “Pruning and Dynamic Scheduling of Cost-Sensitive Ensembles”, Eighteenth National Conference on Artificial Intelligence (AAAI 2002): 146-151.
- 2002 Wei Fan, Haixun Wang, Philip S. Yu, Salvatore J. Stolfo “A Fully Distributed Framework for Cost-Sensitive Data Mining” 22nd International Conference on Distributed Computing Systems (ICDCS 2002), July 2-5, 2002, Vienna, Austria: 445-446.
- 2002 Wei Fan, Haixun Wang, Philip S. Yu, Shaw-hwa Lo, Salvatore J. Stolfo “Progressive Modelling”. 2nd IEEE International Conference on Data Mining (ICDM 2002), December 9-12, 2003, Maebashi, Japan: 163-170.
- 2002 Wei Fan, Haixun Wang, Philip S. Yu, Salvatore J. Stolfo “A Framework for Scalable Cost-sensitive Learning Based on Combing Probabilities and Benefits”, 2nd SIAM International Conference on Data Mining (SDM 2002), April 11-13, 2002, Arlington, VA.
- 2000 Salvatore J. Stolfo, Wenke Lee, Wei Fan, Andreas Prodromidis, and Philip K. Chan “Cost-based Modelling for Fraud and Intrusion Detection: Results from the JAM Project” DARPA Information Survivability Conference 2000

- 1999 Wei Fan, Salvatore J. Stolfo, Junxin Zhang and Philip Chan “AdaCost: Misclassification Cost-sensitive Learning” Proceedings of the Sixteenth International Conference on Machine Learning (ICML’99), pp.97-105, Bled, Slovenia, June 1999.
- 1999 Philip Chan, Andreas Prodromidis, Wei Fan, and Salvatore J. Stolfo “Distributed Data Mining in Credit Card Fraud Detection” IEEE Intelligent Systems Journal November/December 1999, pp.67-74.
- 1999 Wei Fan, Salvatore J. Stolfo and Junxin Zhang “The Application of AdaBoost for Distributed, Scalable and On-line Learning” Proceedings of ACM Sixth International Conference on Knowledge Discovery and Data Mining (KDD’99). pp.362-366, San Diego, California, August 1999.
- 1997 Salvatore J. Stolfo, Andreas Prodromidis, Shelley Tselepis, Wenke Lee, Wei Fan, and Philip Chan “JAM - Java Agents of Meta-learning over Distributed Networks” Proceeding of Third International Conference on Knowledge Discovery and Data Mining (KDD-97), pp.74 - 81, Newport Beach, California, August 1997. *Runner-up best application paper*. The JAM Project is a featured article in Trusted Information Systems Magazine, March, 1998

### **Weather Forecasting Environmental Protection**

- 2008 Kun Zhang, and Wei Fan: “Forecasting Skewed Biased Stochastic Ozone Days: analyses, solutions and beyond”, Knowl. Inf. Syst. 14(3):507-527 (2008).
- 2006 Kun Zhang, Wei Fan, Xiaojing Yuan, Ian Davidson, and Xiangshang Li, “Forecasting Skewed Stochastic Biased Ozone Days: Analyses and Solutions”, 2006 IEEE International Conference on Data Mining (ICDM’2006), December 2006. *Best Paper Award: Application Category*

### **Stock Trading Anomaly Detection**

- 2004 Wei Fan, Philip S. Yu, and Haixun Wang “Mining Extremely Skewed Trading Anomalies” (industrial track), The Ninth International Conference on Extending Database Technology (EDBT’04), Heraklion, Greece, March 2004.

### **Catalogue Mailing Campaign**

- 2004 Wei Fan “Systematic Data Selection to Mine Concept-drifting Data Streams”, The Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’05), Seattle, Washington, August 2004.
- 2002 Wei Fan, Haixun Wang, Philip S. Yu, Salvatore J. Stolfo “A Fully Distributed Framework for Cost-Sensitive Data Mining” 22nd International Conference on Distributed Computing Systems (ICDCS 2002), July 2-5, 2002, Vienna, Austria: 445-446.
- 2002 Naoki Abe, Edwin P. D. Pednault, Haixun Wang, Bianca Zadrozny, Wei Fan, Chidanand Apte “Empirical Comparison of Various Reinforcement Learning Strategies for Sequential Targeted Marketing”, 2nd IEEE International Conference on Data Mining (ICDM 2002), December 9-12, 2002, Maebashi, Japan: 3-10.
- 2002 Edwin P. D. Pednault, Naoki Abe, Bianca Zadrozny, Haixun Wang, Wei Fan, and Chidnant Apte “Sequential Cost-sensitive Decision Making with Reinforcement Learning”, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD’02), Edmonton, Canada.

2002 Wei Fan, Haixun Wang, Philip S. Yu, Salvatore J. Stolfo “A Framework for Scalable Cost-sensitive Learning Based on Combing Probabilities and Benefits”, 2nd SIAM International Conference on Data Mining (SDM 2002), April 11-13, 2002, Arlington, VA.

### **Late and Default Payment Prediction**

2005 Wei Fan, Janek Mathuria, and Chang-tien Lu “Making Data Mining Models Useful to Model Non-paying Customers of Exchange Carriers”. 2005 SIAM International Conference on Data Mining (SDM’05), Newport Beach, CA, April 2005.

### **Intrusion Detection**

2004 Wei Fan, Matthew Miller, Salvatore J. Stolfo, Wenke Lee, Philip K. Chan: “Using artificial anomalies to detect unknown and known network intrusions”. *Knowl. Inf. Syst.* 6(5): 507-527 (2004)

2003 Yi-an Huang, Wei Fan, Wenke Lee, Philip S. Yu “Cross-Feature Analysis for Detecting Ad-Hoc Routing Anomalies”, 23rd International Conference on Distributed Computing Systems (ICDCS 2003), 19-22 May 2004, Providence, RI.

2002 Wenke Lee, Wei Fan, Matthew Miller, Salvatore J. Stolfo, Erez Zadok “Toward Cost-Sensitive Modelling for Intrusion Detection and Response”. *Journal of Computer Security* 10(1/2): 5-22 (2002)

2001 Wenke Lee, and Wei Fan “Mining System Audit Data: Opportunities and Challenges”, *SIGMOD Record* 30(4):34-44 (2001)

2001 Salvatore J. Stolfo, Wenke Lee, Philip K. Chan, Wei Fan, and Eleazar Eskin “Data Mining-based Intrusion Detection: An Overview of the Columbia IDS project”, *SIGMOD Record* 30(4):5-14 (2001).

2000 Wei Fan, Wenke Lee, Salvatore Stolfo, Matthew Miller “A Multiple Model Cost-Sensitive Approach for Intrusion Detection” The Eleventh European Conference on Machine Learning (ECML00).

### **TUTORIALS**

- Wei Fan and Masashi Sugiyama, “Sample Selection Bias - Covariate Shift: Problems, Solutions and Applications”, 2008 IEEE International Conference on Data Mining (ICDM’08), Pisa, Italy.

### **FUNDED PROPOSALS**

- Jiangtao Ren and Wei Fan, “Semi-supervised Feature Selection” funded by National Science Foundation of China NSFC No. 60703110 (2007 - 2009).
- Obtained \$20,000 sponsorship to SDM’09 from Google, NEC Labs America, and NSF.

### **MAJOR HONOURS and AWARDS**

2008 ICDM’08 Data Mining Contest Crown Award (highest awards). Our submission uses my previously proposed random decision tree (RDT, Fan et al., ICDM’03) based approach. Wei Fan, Erheng Zhong, Sihong Xie, Yuzhao Huang, Kun Zhang, Jing Peng and Jiangtao Ren.

- 2006 Best Paper Award (Application Category), “Forecasting Skewed Stochastic Biased Ozone Days: Analyses and Solutions” by Kun Zhang, Wei Fan, Xiaojing Yuan, Ian Davidson, and Xiangshang Li, IEEE International Conference on Data Mining (ICDM’06), December 2006, Hong Kong, China.
- 1997 Runner Up Best Paper Award, “JAM: Java Agents for Meta-learning” by Salvatore Stolfo, Andreas Promdromidis, Shelley Tselepis, Wenke Lee, Wei Fan and Philip Chan, Third International Conference on Knowledge Discovery and Data Mining (KDD-97), Newport Beach, California, August 1997.
- 1994 Guanhua Fellowship Award, Tsinghua University, November 1994.
- 1991 Nantian Fellowship Award, Tsinghua University, November 1991.
- 1990 Best Student Fellowship Award, Tsinghua University, November 1990.

## INVITED TALKS

- ICMLA’06, Orlando, Florida “Let the Data Speak for Themselves - Simplistic Approaches to Difficult Data Mining Problems”

## UNIVERSITY/LAB SEMINARS

- Various talks on “Random Decision Tree” have been given in
  - Tsinghua University
  - Southwestern Jiaotong University
  - Peking University
  - Microsoft Asia May 2008
  - Hefei University of Technology May 2008
  - Sun-Yat Sun University May 2008
  - Hong Kong Polytechnic University May 2008
  - NEC Labs American Cupertino June 2006
  - SRC International June 2006
  - National University of Singapore September 2005
  - Tsinghua University September 2005
  - Rutgers University February 2005
- “From Feature Construction, to Simple but Effective Modelling, to Domain Transfer”
  - Tsinghua University
  - Southwestern Jiaotong University
  - Peking University
  - Montclair State University Oct 2008
  - NASA Ames Research Center Sep 2008
  - Facebook Inc. Aug 2008

## PROFESSIONAL ACTIVITIES

- Senior Position

- Award Committee member of ICDM'09
- Sponsorship co-chair of SDM'09
- Area Chair of ICDM'08
- Area Chair of SDM'08
- Senior PC member of KDD'06
- Program Committee member
  - PC member KDD'09
  - PC member PAKDD'09
  - PC member SDM'09
  - PC member PRICAI'08
  - PC member WWW'09 Data Mining Track
  - PC member CIKM'08 Data Mining Track
  - PC member SIGKDD'08
  - PC member WWW08
  - PC member PAKDD08
  - PC member ICDM'07
  - PC member SIGKDD'07
  - PC member CIKM'07
  - PC member SDM'07
  - PC member WWW'07
  - PC member PAKDD'07
  - PC member ECML/PKDD'06
  - PC member of ICDM'06
  - PC member of SDM'06
  - PC member of ICDM'05
  - PC member of DASFFA'05
  - PC member of SIGKDD'05
  - PC member of SDM'05
  - PC member of ICDE'04
  - PC member of EDBT'04
  - PC member of ICDM'03
- Workshop Organizer
  - DSMM'07 Workshop co-located with ICDM'07
- NSF Panelist
  - IIS April 2008
  - IIS April 2007
- Journal Reviewer
  - TKDE
  - KAIS

- Journal of Machine Learning
- Data Mining and Knowledge Discovery
- Pattern Recognition
- Statistical Analysis and Data Mining
- Advisor Board
  - Member of KD2U.

## **PATENT and PATENT APPLICATION**

- 2008 Haixun Wang, Wei Fan, and Philip Yu “System and Method for Tree Structure Indexing that Provides at least One Constraint Sequence to Preserve Query-equivalence between XML document structure match and subsequence match”, No. 7,475,070.
- 2008 Hong Cheng, Xifeng Yan, Wei Fan, and Philip S. Yu “System and Method for Graph Classification with Skewed Class Distribution”. Filed.
- 2007 Joe Wolf, Kirstin Hildrum, Deepak Ragan, Wei Fan, Lisa Amini, Jim Giles “System and Method for Dynamic Resource Allocation in Streaming Systems”. Filed.
- 2006 Julia Stanoivich, Jun Rao, Wei Fan, Guy Lohmen “System and Method for Query Optimization with Inductive Learning Model”. Filed.
- 2005 Wei Fan, and Philip S. Yu “System and Method for Continuously Diagnosing and Modelling Changes of Real-valued Streaming Variables”. No. 7,464,068.
- 2005 Haixun Wang, Wei Fan, and Philip S. Yu “Index Structure for Supporting structural XML queries”, No. 7,287,023.
- 2004 Wei Fan, Haixun Wang, and Philip S. Yu “System and Method for Sequential Modelling in Less Than One Sequential Scan”. No. 7,337,161.
- 2004 Wei Fan, Haixun Wang, and Philip S. Yu “System and Method for Continuous Diagnosis of Data Streams”. Filed.
- 2004 Wei Fan, Haixun Wang, and Philip S. Yu “System and Method for Sequence-based Subspace Pattern Clustering”. Filed.
- 2004 Haixun Wang, Wei Fan, and Philip S. Yu “System and Method for Mining Time-changing Data Streams”. Filed.
- 2004 Haixun Wang, Wei Fan, and Philip S. Yu “System and Method for Indexing Weighted-Sequences in Large Databases”. No. 7,418,455.
- 2003 Wei Fan, and Philip S. Yu “Cross-Feature Analysis”. Filed.
- 2003 Haixun Wang, Wei Fan, and Philip S. Yu “An index Structure for Supporting XML Queries”. No. 7,287,023.
- 2003 Wei Fan, Haixun Wang, and Philip S. Yu “System and Method for Adaptive Pruning”. Filed.
- 2003 Wei Fan, Haixun Wang, and Philip S. Yu “System and Method for Scalable Cost-sensitive Learning”. Filed.
- 2001 Wei Fan, and Salvatore J. Stolfo “System and Method for Anomaly Detection and Adaptive Learning” (**Thesis work**). No. 7,424,619.

1999 Salvatore Stolfo, Wei Fan, and Wenke Lee “Method and System for Using Intelligent Agents for Financial Transactions, Services, Accounting and Advice”, United States Patent No. 5,920,848.

## **STUDENT ADVICE EXPERIENCE**

- Kun Zhang (now faculty at Computer Science in Xavier University of Louisiana). Our work has been accepted in ICDM’06 (two papers), KDD’08 and KAIS journal. One of our papers in ICDM’06 has won the Best Application Paper Award. Under my guidance, Kun Zhang’s proposal on ”ALU insertion site” has been awarded a 3 years of funding and ranked 3 among over 100 submissions.
- Xiaoxiao Shi (now master student from Sun-Yat Sun University in Guangzhou China, and visiting student in Hong Kong University of Science and Technology, invited by Prof. Qiang Yang). Our work on sample selection bias and transfer learning has been accepted in SDM’08 and ECML/PKDD’08.
- Erheng Zhong and Sihong Xie (now master student from Sun-Yat Sun University in Guangzhou China). Our work on hybrid feature selection has been accepted in ICDM’08.
- Jing Gao (PhD student from UIUC and student of Prof. Jiawei Han). Our joint work has been accepted in SDM’07, ICDM’07 and KDD’08 (two papers).
- Zhengyuan Qiu (former master student from Sun-Yat Sun University, now a research scientist at Tencent Inc or QQ). Our work has been accepted in PAKDD’08.
- Yi-an Huang (former PhD student from Georgia Tech, now in Google). Our work has been accepted in ICDCS’03, ICDM’03 and SDM’04. Yi-an was nominated an “IBM Outstanding Summer Student” and worked in IBM for two summers. Our ICDCS’03 has 68 citations according to Google scholar.
- Fang Chu (former PhD student from UCLA now at Google China, co-supervised with Dr. Haixun Wang). Our joint work was accepted at AAAI’02 and SSDBM’04).

## **REFERENCES**

- Available upon request.