Distributed Data Mining Bibliography *

Kun Liu, Hillol Kargupta and Jessica Ryan
Computer Science and Electrical Engineering Department
University of Maryland Baltimore County
Baltimore, Maryland 21250
{kunliu1, hillol, jryan4}@cs.umbc.edu

March, 2004 (Release 1.3)

Abstract

Advances in computing and communication over wired and wireless networks have resulted in many pervasive distributed computing environments. Many of these environments deal with different distributed sources of voluminous data, multiple compute nodes, and distributed user community. Analyzing and monitoring these distributed data sources require a data mining technology designed for distributed applications. The field of distributed data mining (DDM) deals with this problem—mining distributed data by paying careful attention to the distributed resources. The goal of this paper is to maintain and distribute a bibliography of DDM-related publications. We hope that DDM researchers and practitioners find this service useful. We welcome every help from the community in maintaining the bibliography.

*The authors acknowledge supports from the NASA (NRA) NAS2-37143 and the United States National Science Foundation CAREER award IIS-0093353.

References

[1] H. Abe and T. Yamaguchi. Comparing the Parallel Automatic Composition of Inductive Applications with Stacking Methods. In *Parallel and Distributed computing for Machine Learning. In conjunction with the 14th European Conference on Machine Learning (ECML'03) and 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'03)*, Cavtat-Dubrovnik, Croatia, September 2003.

[2] T. Achalakul and S. Taylor. A distributed spectral-screening PCT algorithm. *Journal of Parallel and Distributed Computing*, 63(3):373–384, 2003.

[3] R. C. Agarwal, C. C. Aggarwal, and V. V. V. Prasad. A Tree Projection Algorithm for Generation of Frequent Item Sets. *Journal of Parallel and Distributed Computing*, 61(3):350–371, 2001.

[4] C. C. Aggarwal, J. L. Wolf, K.-L. Wu, and P. S. Yu. Horting Hatches an Egg: A New Graph-Theoretic Approach to Collaborative Filtering. In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 201–212, San Diego, CA, August 1999.

[5] G. Agrawal. High-level Interfaces for Data Mining: From Offline Algorithms on Clusters to Streams on Grids. In *Workshop on Data Mining and Exploration Middleware for Distributed and Grid Computing*, Minneapolis, MN, September 2003.

[6] R. Agrawal and J. C. Shafer. Parallel Mining of Association Rules. *IEEE Transactions On Knowledge And Data Engineering*, 8:962–969, 1996.

[7] N. Amado, J. Gama, and F. Silva. Exploiting Parallelism in Decision Tree Induction. In *Parallel and Distributed computing for Machine Learning. In conjunction with the 14th European Conference on Machine Learning (ECML'03) and 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'03)*, Cavtat-Dubrovnik, Croatia, September 2003.

[8] V. S. Ananthanarayana, D. K. Subramanian, and M. N. Murty. Scalable, Distributed and Dynamic Mining of Association Rules. In *Proceedings of HIPC'00*, pages 559–566, Bangalore, India, 2000.

[9] H. Andrade, T. Kurc, J. Saltz, and A. Sussman. Decision Tree Construction for Data Mining on Clusters of Shared Memory Multiprocessors. In *HPDM: High Performance, Pervasive, and Data Stream Mining 6th International Workshop on High Performance Data Mining: Pervasive and Data Stream Mining (HPDM:PDS'03). In conjunction with Third International SIAM Conference on Data Mining*, San Francisco, CA, May 2003.

[10] E. Ariwa, M. Senousy, and M. M. Gaber. Facilities Management and E-business Model Application for Distributed Data Mining using Mobile Agents. *The International Journal of Applied Marketing*, 2(1), 2003.

[11] Ezendu Ariwa and Medhat Medhat Gaber. Globalization and Informatization: Analysis of the Application of Distributed Data Mining to Facilities Management. In *32nd International Conference on Computers and Industrial Engineering Sustainability, Globalisation-The Engineering Challenge*, 2003.

[12] J. Aronis, V. Kolluri, F. Provost, and B. Buchanan. The WoRLD: Knowledge Discovery from Multiple Distributed Databases. Technical Report ISL-96-6, Intelligent Systems Laboratory, Department of Computer Science, University of Pittsburgh, Pittsburgh, PA, 1996.

[13] M. Z. Ashrafi, D. Taniar, and K. A. Smith. A Data Mining Architecture for Distributed Environments. *IICS 2002*, pages 27–38, 2002.

[14] M. J. Atallah and W. Du. Secure Multi-Party Computational Geometry. In *Proceedings of 7th International Workshop on Algorithms and Data Structures (WADS 2001)*, volume 2125 of *Lecture Notes in Computer Science*, pages 165–179, Providence, RI, August 2001. Springer Verlag.

[15] A. Atramentov, H. Leiva, and V. Honavar. A Multi-Relational Decision Tree Learning Algorithm - Implementation and Experiments. In *Proceedings of the Thirteenth International Conference on Inductive Logic Programming*, Berlin, 2003. Springer-Verlag.

[16] B. Babcock and C. Olston. Distributed Top-K Monitoring. In *Proceedings of the ACM SIGMOD 2003 International Conference on Management of Data*, pages 28–39, February 2003.

[17] S. Bailey, R. Grossman, H. Sivakumar, and A. Turinsky. Papyrus: A System for Data Mining Over Local and Wide Area Clusters and Super-clusters. In *Proceedings of the 1999 ACM/IEEE conference on Supercomputing*, page 63, Portland, OR, 1999. ACM Press.

[18] J. Bala, S. Baik, A. Hadjarian, B. K. Gogia, and C. Manthorne. Application of a Distributed Data Mining Approach to Network Intrusion Detection. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 1419–1420, Bologna, Italy, 2002. ACM Press.

[19] R. Bhargava, H. Kargupta, and M. Powers. Energy Consumption in Data Analysis for On-board and Distributed Applications. In *Proceedings of the 2003 International Conference on Machine Learning workshop on Machine Learning Technologies for Autonomous Space Applications*, 2003.

[20] P. B. Bhat, C. S. Raghavendra, and V. K. Prasanna. Efficient collective communication in distributed heterogeneous systems. *Journal of Parallel and Distributed Computing*, 63(3):251–263, 2003.

[21] R. Bhatnagar. Decision Tree Induction by Cooperating Agents. In *Workshop on Multi-Agent Learning*, Providence, RI, July 1997.

[22] J. P. Bradford and J. B. Fortes. Characterization and Parallelization of Decision-Tree Induction. *Journal of Parallel and Distributed Computing*, 61(3):322–349, 2001.

[23] M. Burl, C. Fowlkes, J. Roden, A. Stechert, and S. Mukhtar. Diamond Eye: A distributed architecture for image data mining. In *SPIE Thirteenth International Symposium. on Aerospace/Defence Sensing, Simulation, and Controls*, 1999.

[24] D. Cai, M. F. McTear, and S.I. McClean. Knowledge discovery in distributed databases using evidence theory. *International Journal of Intelligent Systems*, 15(8):745–761, 2000.

[25] J. Calmet, S. Jekutsch, P. Kullmann, and J. Schü. KOMET - A System for the Integration of Heterogeneous Information Sources. In *International Syposium on Methodologies for Intelligent Systems*, pages 318–327, 1997.

[26] M. Cannataro, A. Congiusta, D. Talia, and P. Trunfio. A Data Mining Toolset for Distributed High-Performance Platforms. In *Proceedings of Data Mining 2002*, Bologna, Italy, 2002. Wessex Institute Press.

[27] M. Cannataro and D. Talia. The Knowledge Grid. *Communications of the ACM*, 46(1):89–93, January 2003.

[28] M. Cannataro, D. Talia, and P. Trunfio. KNOWLEDGE GRID: High Performance Knowledge Discovery on the Grid. *GRID 2001*, pages 38–50, 2001.

[29] D. Caragea, J. Reinoso, A. Silvescu, and V. Honavar. Statistics Gathering for Learning from Distributed, Heterogeneous and Autonomous Data Sources. In *International Workshop on Information Integration on the Web, IJCAI 2003*. AAAI, Acapulco, Mexico, 2003.

[30] D. Caragea, A. Silvescu, and V. Honavar. Towards a Theoretical Framework for Analysis and Synthesis of Agents That Learn from Distributed Dynamic Data Sources. In *Emerging Neural Architectures Based on Neuroscience*, pages 547–559. Springer-Verlag, 2001.

[31] D. Caragea, A. Silvescu, and V. Honavar. A Framework for Learning from Distributed Data Using Sufficient Statistics and its Application to Learning Decision Trees. *International Journal of Hybrid Intelligent Systems.*, 2003.

[32] D. Caragea, A. Silvescu, and V. Honavar. Decision Tree Induction from Distributed, Heterogeneous, Autonomous Data Sources. In *Proceedings of the Conference on Intelligent Systems Design and Applications (ISDA 03)*, Tulsa, Oklahoma, 2003.

[33] Doina Caragea, Adrian Silvescu, and Vasant Honavar. Decision Tree Induction from Distributed Data Sources. In *Proceedings of the Conference on Intelligent Systems Design and Applications*, Tulsa, Oklahoma, 2003.

[34] R. J. Cavanaugh. Sphinx: A Scheduling Middleware for Data Intensive Applications on a Grid. In *Workshop on Data Mining and Exploration Middleware for Distributed and Grid Computing*, Minneapolis, MN, September 2003.

[35] P. Chan, W. Fan, A. Prodromidis, and S. Stolfo. Distributed Data Mining in Credit Card Fraud Detection. *IEEE Intelligent Systems*, pages 67–74, Nov/Dec 1999.

[36] P. Chan and S. J. Stolfo. Toward Parallel and Distributed Learning by Meta-learning. In *Working Notes AAAI Work. Knowledge Discovery in Databases*, pages 227–240. AAAI, 1993.

[37] P. Chan and S. J. Stolfo. A Comparative Evaluation of Voting and Meta-learning on Partitioned Data. In *Proceedings of Twelfth International Conference on Machine Learning*, pages 90–98, 1995.

[38] P. Chan and S. J. Stolfo. On the Accuracy of Meta-learning for Scalable Data Mining. *Intelligent Information System*, 8:5–28, 1996.

[39] P. Chan and S. J. Stolfo. Sharing Learned Models among Remote Database Partitions by Local Meta-Learning. In E. Simoudis, J. Han, and U. Fayyad, editors, *The Second International Conference on Knowledge Discovery and Data Mining*, pages 2–7. AAAI Press, 1996.

[40] P. Chan and S. J. Stolfo. Toward Scalable Learning with Non-uniform Class and Cost Distribution: A Case Study in Credit Card Fraud Detection. In *Proceeding of the Fourth International Conference on Knowledge Discovery and Data Mining*. AAAI Press, September 1998.

[41] J. Chattratichat, J. Darlington, Y. Guo, S. Hedvall, M. Koler, and J. Syed. An Architecture for Distributed Enterprise Data Mining. In *High Performance Computing Networking*, pages 573–582, Amsterdam, Netherlands, 1999.

[42] N. Chawla, S. Eschrich, and L. O. Hall. Creating Ensembles of Classifiers. *IEEE International Conference on Data Mining*, pages 580–581, 2001.

[43] N. V. Chawla. *RiDE: Rule-learning in a Distributed Environment*, 1999.

[44] N. V. Chawla. *Learning on extremes — size and imbalance — of data*. PhD thesis, University of South Florida, 2002.

[45] N. V. Chawla, L. O. Hall, K. W. Bowyer, T. E. Moore, and W. P. Kegelmeyer. Distributed Pasting of Small Votes. In *Multiple Classifier Systems*, 2002.

[46] N. V. Chawla, T. E. Moore, L. O. Hall, K. W. Bowyer, W. P. Kegelmeyer, and C. Springer. Distributed Learning With Bagging-like Performance. *Pattern Recognition Letters*, 24:455–471, 2003.

[47] R. Chen and S. Krishnamoorthy. A New Algorithm for Learning Parameters of a Bayesian Network from Distributed Data. In *Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM 2002)*, pages 585–588, Maebashi City, Japan, December 2002. IEEE Computer Society.

[48] R. Chen, S. Krishnamoorthy, and H. Kargupta. Distributed Web Mining using Bayesian Networks from Multiple Data Streams. In *Proceedings of the IEEE International Conference on Data Mining*, pages 281–288. IEEE Press, November 2001.

[49] R. Chen, K. Sivakumar, and H. Kargupta. Distributed Bayesian Mining from Heterogeneous Data. *Knowledge and Information Systems Journal*, 2003. Accepted for publication. In Press.

[50] Z. Chen, X. Meng, B. Zhu, and R. H. Fowler. WebSail: From On-line Learning to Web Search. *Knowledge and Information Systems*, 4(2):219–227, 2002.

[51] A. Chervenak, I. Foster, C. Kesselman, C. Salisbury, and S. Tuecke. The Data Grid: Towards an Architecture For the Distributed Management and Analysis of Large Scientific Datasets, 1999.

[52] D. Cheung and Y. Xiao. Effect of Data Skewness in Parallel Mining of Association Rules. In *12th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 48–60, Melbourne, Australia, April 1998.

[53] D. W. Cheung, J. Han, V. T. Ng, A. W. Fu, and Y. Fu. A Fast Distributed Algorithm for Mining Association Rules. In *Proceedings of 1996 International Conference on Parallel and Distributed Information Systems (PDIS'96)*, pages 31–44, Miami, FL, 1996.

[54] D. W. Cheung, V. T. Ng, A. W. Fu, and Y. Fu. Efficient Mining of Association Rules in Distributed Databases. *IEEE Transactions On Knowledge And Data Engineering*, 8:911–922, 1996.

[55] V. Cho and B. Wüthrich. Toward Real Time Discovery from Distributed Information Sources. In *12th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 376–377, Melbourne, Australia, April 1998.

[56] V. Cho and B. Wüthrich. Distributed Mining of Classification Rules. *Knowledge and Information Systems*, 4(1):1–30, 2002.

[57] A. Choudhary, M. Kandemir, J. No, G. Memik, X. Shen, W. Liao, H. Nagesh, S. More, V. Taylor, R. Thakur, and R. Stevens. Data Management for Large-Scale Scientific Computations in High Performance Distributed Systems. *Cluster Computing: the Journal of Networks, Software Tools and Applications*, 3(1):45–60, 2000.

[58] A. Choudhury, P. B. Nair, and A. J. Keane. A Data Parallel Approach for Large-Scale Gaussian Process Modeling. In *Proceedings of the Second SIAM International Conference on Data Mining*, Arlington, VA, April 2002.

[59] C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, , and M. Zhu. Tools for Privacy Preserving Distributed Data Mining. *ACM SIGKDD Explorations*, 4(2), December 2002.

[60] F. Coenen, P. Leng, and A. Shakil. T-trees, Vertical Partitioning and Distributed Association Rule Mining. In *The Third IEEE International Conference on Data Mining (ICDM'03)*, Melbourne, FL, November 2003.

[61] D. J. Cook, L. B. Holder, G. Galal, and R. Maglothin. Approaches to Parallel Graph-Based Knowledge Discovery. *Journal of Parallel and Distributed Computing*, 61(3):427–446, 2001.

[62] V. Crestana and N. Sopkar. Mining Decentralized Data Repositories. Technical Report CSE-TR-385-99, University of Michigan, Ann Arbor, 1999.

[63] H. Dail, F. Berman, and H. Casanova. A Decoupled Scheduling Approach for Grid Application Development Environments. *Journal of Parallel and Distributed Computing*, 63(5):505–524, 2003.

[64] S. Dao and B. Perry. Applying a Data Miner To Heterogeneous Schema Integration. In *Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95)*, pages 63–68, Montreal, Canada, August 1995.

[65] W. Davies and P. Edwards. Agent-Based Knowledge Discovery. In *AAAI Spring Symposium on Information Gathering*, 1995.

[66] W. Davies and P. Edwards. Distributed Learning: An Agent-Based Approach to Data-Mining. In Diana Gordon, editor, *Proceedings of Machine Learning-95 Workshop on Agents That Learn From Each Other*, Tahoe City, CA, 1995. AAAI Press.

[67] A. Demers, J. E. Gehrke, and M. Riedewald. Research Issues in Distributed Mining and Monitoring. In *Proceedings of the National Science Foundation Workshop on Next Generation Data Mining (NGDM 2002)*, Baltimore, MD, November 2002.

[68] A. Demiriz. webSPADE: A Parallel Sequence Mining Algorithm to Analyze Web Log Data. In *Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM 2002)*, pages 755–758, Maebashi City, Japan, December 2002. IEEE Computer Society.

[69] I. Dhillon and D. Modha. A Data-clustering Algorithm on Distributed Memory Multiprocessors. In *Proceedings of the KDD'99 Workshop on High Performance Knowledge Discovery*, pages 245–260, 1999.

[70] W. Du and G. Agrawal. Using General Grid Tools and Compiler Technology for Distributed Data Mining: Preliminary Report. In *5th International Workshop on High Performance Data Mining: Resource and Location Aware Mining (HPDM:RLM'02). In conjunction with Second International SIAM Conference on Data Mining*, Arlington, VA, April 2002.

[71] W. Du and M. J. Atallah. Privacy-Preserving Cooperative Scientific Computations. In *14th IEEE Computer Security Foundations Workshop*, pages 273–282, Nova Scotia, Canada, June 2001.

[72] W. Du and M. J. Atallah. Secure Multi-Party Computation Problems and Their Applications: A Review and Open Problems. In *New Security Paradigms Workshop*, pages 11–20, Cloudcroft, NM, September 2001.

[73] W. Du and Z. Zhan. Building Decision Tree Classifier on Private Data. In *Workshop on Privacy, Security, and Data Mining at The 2002 IEEE International Conference on Data Mining (ICDM'02)*, Maebashi City, Japan, December 2002.

[74] E. Durfee, V. R. Lesser, and D. D. Corkill. Cooperative Distributed Problem Solving. In A. Barr, , P. R. Cohen, and E. I. Feigenbaum, editors, *Handbook of Artificial Intelligence*, volume 4, 1989.

[75] F. Barillari E. Nardelli and M. Pepe. Distributed Searching of Multi-dimensional Data: A Performance Evaluation Study. *Journal of Parallel and Distributed Computing*, 49(1):111–134, 1998.

[76] M. Faerman, A. Su, R. Wolski, and F. Berman. Adaptive Performance Prediction for Distributed Data-intensive Applications. In *Proceedings of the 1999 ACM/IEEE conference on Supercomputing (CDROM)*, page 36, Portland, OR, 1999. ACM Press.

[77] W. Fan, S. J. Stolfo, and J. Zhang. The Application of AdaBoost for Distributed, Scalable and On-Line Learning. In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 362–366, San Diego, CA, August 1999.

[78] X. Fern and C. Brodley. Random Projection for High Dimensional Data Clustering: A Cluster Ensemble Approach. In *The Twentieth International Conference on Machine Learning (ICML2003)*, Washington, DC, August 2003.

[79] N. Fonseca, R. Camacho, and F. Silva. A parallel ILP algorithm that incorporates incremental batch learning. In *Parallel and Distributed computing for Machine Learning. In conjunction with the 14th European Conference on Machine Learning (ECML'03) and 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'03)*, Cavtat-Dubrovnik, Croatia, September 2003.

[80] G. Forman and B. Zhang. Distributed Data Clustering Can Be Efficient and Exact. *SIGKDD Explorations*, 2(2):34–38, 2000.

[81] D. Foti, D. Lipari, C. Pizzuti, and D. Talia. Scalable Parallel Clustering for Data Mining on Multicomputers. In *3rd Workshop on High Performance Data Mining. In conjunction with International Parallel and Distributed Processing Symposium 2000 (IPDPS'00)*, Cancun, Mexico, May 2000.

[82] Mohamed Medhat Gaber. *A Framework for a Scalable Distributed Data Mining Model*, 2002.

[83] Mohamed Medhat Gaber. A Model of Distributed Data Mining as a Knowledge Acquisition Tool in Knowledge Management Systems. In *10th Scientific Conference on Information Systems and Computer Technology: Knowledge Management in the Era of Globalization*, 2003.

[84] J. Ghosh, A. Strehl, and S. Merugu. A Consensus Framework for Integrating Distributed Clusterings Under Limited Knowledge Sharing. In *Proceedings of NSF Workshop on Next Generation Data Mining*, pages 99–108, Baltimore, MD, November 2002.

[85] J. Ghosh and K. Tumer. Robust Order Statistics Based Ensembles for Distributed Data Mining. In Hillol Kargupta and Philip Chan, editors, *Advances in Distributed and Parallel Knowledge Discovery*, pages 185–210. MIT/AAAI Press, 2000.

[86] Nikolaos Giannadakis, Anthony Rowe, Moustafa Ghanem, and Yi ke Guo. InfoGrid: providing information integration for knowledge discovery. *Information Sciences. Special Issue: Knowledge Discovery from Distributed Information Sources*, 155(3–4):199–226, October 2003.

[87] M. Gillmann, J. Weißenfels, G. Shegalov, W. Wonner, and G. Weikum. A Goal-driven Auto-Configuration Tool for the Distributed Workflow Management System Mentorlite. *SIGMOD Record*, 29(2):595, 2000.

[88] S. Goil and A. Choudhary. High Performance OLAP and Data Mining on Parallel Computers. *Journal of Data Mining and Knowledge Discovery (Special Issue on Scalable High-Performance Computing for KDD)*, 1(4):391–417, 1997.

[89] S. Goil and A. Choudhary. PARSIMONY: An Infrastructure for Parallel Multidimensional Analysis and Data Mining. *Journal of Parallel and Distributed Computing*, 61(3):285–321, 2001.

[90] D. L. Grecu and L. A. Becker. Coactive Learning for Distributed Data Mining. In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98)*, pages 209–213, New York, NY, August 1998.

[91] R. L. Grossman, S. Bailey, A. Ramu, B. Malhi, and A. Turinsky. The Preliminary Design of Papyrus: A System for High Performance, Distributed Data Mining over Clusters. In Hillol Kargupta and Philip Chan, editors, *Advances in Distributed and Parallel Knowledge Discovery*, pages 259–275. MIT/AAAI Press, Menlo Park, CA, 2000.

[92] R. L. Grossman and A. Turinsky. A Framework for Finding Distributed Data Mining Strategies That are Intermediate Between Centralized Strategies and In-Place Strategies. In *Workshop on Distributed and Parallel Knowledge Discovery*, Boston, MA, 2000.

[93] P. Gu and A. B. Maddox. A Framework for Distributed Reinforcement Learning. In Gerhard Weiß and Sundip Sen, editors, *Adaption and Learning in Multi-Agent Systems*, number 1042 in Lecture Notes in Computer Science : Lecture Notes in Artificial Intelligence, pages 97–112, New York, NY, 1995. Springer-Verlag. Proceedings IJCI'95 Workshop, Montreal, Canada, 1995.

[94] Y. Guo. Discovery Net: A UK E-science Pilot Project for Grid Based Knowledge Discovery Service. In *Workshop on Data Mining and Exploration Middleware for Distributed and Grid Computing*, Minneapolis, Minnesota, September 2003.

[95] Y. Guo and J. Sutiwaraphun. Distributed learning with Knowledge Probing: A New Framework for Distributed Data Mining. In Hillol Kargupa and Phillip Chan, editors, *Advances in Distributed and Parallel Knowledge Discovery*, pages 113–131. MIT/AAAI Press, 2000.

[96] N. Gupta and S. Sen. Faster Output-sensitive Parallel Algorithms for 3D Convex Hulls and Vector Maxima. *Journal of Parallel and Distributed Computing*, 63(4):488–500, 2003.

[97] I. J. Haimowitz, Özden Gür-Ali, and H. Schwarz. Integrating and Mining Distributed Customer Databases. In *The Third ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Newportbeach, CA, August 1997.

[98] L. Hall and K. Bowyer. Comparing Pure Parallel Ensemble Creation Techniques against Bagging. In *The Third IEEE International Conference on Data Mining (ICDM'03)*, Melbourne, FL, November 2003.

[99] D. E. Hershberger and H. Kargupta. Distributed Multivariate Regression Using Wavelet-Based Collective Data Mining. *Journal of Parallel and Distributed Computing*, 61(3):372–400, 2001.

[100] J. Himber, J. Tikänmaki, Hannu T.T. Toivonen, K. Korpiaho, and H. Mannila. Time Series Segmentation for Context Recognition in Mobile Devices. In *2001 IEEE International Conference on Data Mining (ICDM'01)* , San Jose, California, November 2001.

[101] B. Hollebeek. NDMA: Collecting and Organizing a Large Scale Collection of Medical Image Records. In *Workshop on Data Mining and Exploration Middleware for Distributed and Grid Computing*, Minneapolis, MN, September 2003.

[102] V. Honavar, L. Miller, and J. Wong. Distributed Knowledge Networks. In *IEEE Information Technology Conference*, Syracuse, NY, 1998.

[103] M. Indrawan, S. Krishnaswamy, and T. Ranjan. Using Mobile Agents to Support Unreliable Database Operations. In *The International Conference on Advanced Information Networking and Applications (AINA2003)*, Xian, China, April 2003. IEEE Press.

[104] M. Indrawan, S. Krishnaswamy, and D. Sethi. Supporting Database Retrieval in Mobile Computing Environments Using Mobile Agents. In *International Conference on Advances in Mobile Multimedia (MoMM 2003)*, Jakarta, Indonesia, September 2003.

[105] Bai Z. J. A Parallel Algorithm for Computing the Generalized Singular Value Decomposition. *Journal of Parallel and Distributed Computing*, 20(3):280–288, 1994.

[106] V. C. Jensen and N. Soparkar. Frequent Itemset Counting Across Multiple Tables. In *4th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 49–61, 2000.

[107] R. Jin and G. Agrawal. Shared Memory Parallelization of Data Mining Algorithms: Techniques, Programming Interface, and Performance. In *Proceedings of the Second SIAM International Conference on Data Mining*, Arlington, VA, April 2002.

[108] R. Jin and G. Agrawal. Combining Distributed Memory and Shared Memory Parallelization for Data Mining Algorithms. In *HPDM: High Performance, Pervasive, and Data Stream Mining 6th International Workshop on High Performance Data Mining: Pervasive and Data Stream Mining (HPDM:PDS'03). In conjunction with Third International SIAM Conference on Data Mining*, San Francisco, CA, May 2003.

[109] R. Jin and G. Agrawal. Communication and Memory Efficient Parallel Decision Tree Construction. In *Proceedings of the Third SIAM International Conference on Data Mining*, San Francisco, CA, May 2003.

[110] R. Jin and G. Agrawal. A Middleware for Developing Parallel Data Mining Applications. In 2001, editor, *Proceedings of the First SIAM International Conference on Data Mining*, April Chicago, IL.

[111] X. Jin, Y. Lu, and C. Shi. Distribution Discovery: Local Analysis of Temporal Rules. In *The Sixth Pacific- Asia Conference on Knowledge Discovery and Data Mining (PAKDD2002)*, Taiwan, China, May 2002.

[112] E. Johnson and H. Kargupta. Collective, Hierarchical Clustering From Distributed, Heterogeneous Data. In M. Zaki and C. Ho, editors, *Large-Scale Parallel KDD Systems. Lecture Notes in Computer Science*, volume 1759, pages 221–244. Springer-Verlag, 1999.

[113] C. Jones, J. Hall, and J. Jale. Secure Distributed Database Mining: Principles of Design. In Hillol Kargupa and Phillip Chan, editors, *Advances in Distributed Data Mining*, pages 273–291. MIT/AAAI Press, 2000.

[114] S.-H. Jou and S.-J. Kao. Agent-Based Infrastructure and an Application to Internet Information Gathering. *Knowledge and Information Systems*, 4(1):80–95, 2002.

[115] Pierre-Emmanuel Jouve and Nicolas Nicoloyannis Laboratoire Eric. A New Method for Combining Partitions, Applications for Cluster Ensembles in KDD. In *Parallel and Distributed computing for Machine Learning. In conjunction with the 14th European Conference on Machine Learning (ECML'03) and 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'03)*, Cavtat-Dubrovnik, Croatia, September 2003.

[116] R. Kanagasabai and A.-H.Tan. Mining Semantic Networks for Knowledge Discovery. In *The Third IEEE International Conference on Data Mining (ICDM'03)*, Melbourne, FL, November 2003.

[117] M. Kantarcioglu and C. Clifton. Privacy-preserving Distributed Mining of Association Rules on Horizontally Partitioned Data. In *ACM SIGMOD Workshop on Research Issues on DMKD'02*, June 2002.

[118] M. Kantarcioglu and J. Vaidya. A New Architecture for Privacy Preserving Data Mining. In *Privacy, Security and Data Mining of the ACS Series Conferences in Research and Practice in Information Technology*, volume 14, 2002.

[119] M. Kantarcioglu and J. Vaidya. An Architecture for Privacy-preserving Mining of Client Information. In Chris Clifton and Vladimir Estivill-Castro, editors, *IEEE International Conference on Data Mining Workshop on Privacy, Security, and Data Mining*, volume 14, pages 37–42, Maebashi City, Japan, December 2002.

[120] H. Kargupta and P. Chan. Distributed Data Mining. *AI Magazine*, 20(1):126, 1999.

[121] H. Kargupta and P. Chan. *Advances in Distributed and Parallel Knowledge Discovery*. AAAI/MIT Press, 2000.

[122] H. Kargupta, J. Ghosh, V. Kumar, and Z. Obradovic. Report from the Workshop on Distributed and Parallel Knowledge Discovery. *SIGKDD Explorations*, 2(2):108–109, 2000.

[123] H. Kargupta, I. Hamzaoglu, and B. Stafford. Scalable, Distributed Data Mining Using An Agent Based Architecture. In David Heckerman, Heikki Mannila, Daryl Pregibon, and Ramasamy Uthurusamy, editors, *Proceedings of Knowledge Discovery And Data Mining*, pages 211–214, Menlo Park, CA, 1997. AAAI Press.

[124] H. Kargupta, W. Huang, S. Krishnamrthy, B. Park, and S. Wang. Collective PCA from Distributed, Heterogeneous Data. In *Proceedings of The Fourth European Conference on Principles and Practice of Knowledge Discovery in Databases*, pages 452–457. Springer Verlag, September 2000.

[125] H. Kargupta, W. Huang, K. Sivakumar, and E. Johnson. Distributed Clustering Using Collective Principal Component Analysis. *Knowledge and Information Systems*, 3(4):422–448, 2001.

[126] H. Kargupta, C. Kamath, and P. Chan. Distributed and Parallel Data Mining: Emergence, Growth, and Future Directions. In *Advances in Distributed and Parallel Knowledge Discovery*, pages 409–416. AAAI/MIT Press, 2000.

[127] H. Kargupta, K. Liu, and J. Ryan. Random Projection and Privacy Preserving Correlation Computation from Distributed Data. In *HPDM: High Performance, Pervasive, and Data Stream Mining 6th International Workshop on High Performance Data Mining: Pervasive and Data Stream Mining (HPDM:PDS'03). In conjunction with Third International SIAM Conference on Data Mining*, San Francisco, CA, May 2003.

[128] H. Kargupta and B. Park. Mining Decision Trees from Data Streams in a Mobile Environment. In *Proceedings of the IEEE International Conference on Data Mining*, pages 75–82. IEEE Press, November 2001.

[129] H. Kargupta and B. Park. Mining Time-critical Data Stream Using the Fourier Spectrum of Decision Trees. In *Proceedings of the IEEE International Conference on Data Mining*, pages 281–288. IEEE Press, 2001.

[130] H. Kargupta and B. Park. Mining Time-Critical Data Streams from Mobile Devices using Decision Trees and Their Fourier Spectrum. *IEEE Transaction on Knowledge and Data Engineering*, 2003.

[131] H. Kargupta, B. Park, D. Hershberger, and E. Johnson. Collective Data Mining: A New Perspective Towards Distributed Data Mining. In Hillol Kargupta and Philip Chan, editors, *Advances in Distributed and Parallel Knowledge Discovery*, pages 133–184. MIT/AAAI Press, 2000.

[132] H. Kargupta, B. Park, E. Johnson, E. Sanseverino, L. Silvestre, and D. Hershberger. Collective Data Mining From Distributed Vertically Partitioned Feature Space. In *Workshop on distributed data mining. International ConferenceonKnowledge Discovery and Data Mining.*, 1998.

[133] H. Kargupta, B. Park, S. Pittie, L. Liu, D. Kushraj, and K. Sarkar. MobiMine: Monitoring the Stock Market from a PDA. *ACM SIGKDD Explorations*, 3(2):37–46, January 2002.

[134] H. Kargupta, K. Sivakumar, and S. Ghosh. A Random Matrix-based Approach for Dependency Detection. In *Proceedings of the 2002 Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD'2002)*, pages 24–29, Madison, WI, June 2002. ACM SIGMOD.

[135] H. Kargupta, K. Sivakumar, and S. Ghosh. Dependency Detection in MobiMine and Random Matrices. In *Proceedings of the 6th European Conference on Principles and Practice of Knowledge Discovery in Databases*, pages 250–262, Helsinki, Finland, 2002. Springer Verlag.

[136] H. Kargupta, K. Sivakumar, W. Huang, R. Ayyagari, R. Chen, B. Park, and E. Johnson. Toward Ubiquitous Mining of Distributed Data. In Robert Grossman, Chandrika Kamath, Philip Kegelmeyer, Vipin Kumar, and Raju Namburu, editors, *Data Mining for Scientific and Engineering Applications*, pages 281–306. Kluwer Academic Publishers, 2001.

[137] Stasinos Th. Konstantopoulos. A Data-Parallel Version of Aleph. In *Parallel and Distributed computing for Machine Learning. In conjunction with the 14th European Conference on Machine Learning (ECML'03) and 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'03)*, Cavtat-Dubrovnik, Croatia, September 2003.

[138] H. Kosch, D. Skillicorn, and D. Talia. Parallel and Distributed Databases. In *Data Mining and Knowledge Discovery. Euro-Par 2002*, pages 319–320, 2002.

[139] S. Krishnaswamy, S. Loke, and A. Zaslavsky. Cost Models For Distributed Data Mining. In *Proceedings of the Twelfth International Conference on Software Engineering and Knowledge Engineering (SEKE'2000)*, pages 31–38, Chicago, IL, 2000.

[140] S. Krishnaswamy and W. S. Loke. On Modelling Agent Mobility in Multiagent Methodologies. In *Workshop on Agent Oriented Information Systems (AOIS 2003). Held in conjunction with the Second International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2003)*, Melbourne, Australia, July 2003.

[141] S. Krishnaswamy, W. S. Loke, and A. Zaslasvky. Supporting the Optimisation of Distributed Data Mining by Predicting Application Run Times. In M. Piatting, J. Filipe, and J. Braz, editors, *Enterprise Information Systems IV*. Kluwer Academic, 2002.

[142] S. Krishnaswamy, W. S. Loke, and A. Zaslavsky. Supporting the Optimisation of Distributed Data Mining by Predicting Application Run Times. In *Proceedings of the Fourth International Conference on Enterprise Information Systems (ICEIS 2002)*, pages 374–381, Ciudad Real, Spain, April 2002.

[143] S. Krishnaswamy, W. S. Loke, and A. Zaslavsky. Towards Anytime Anywhere Data Mining E-Services. In S. J. Simoff, G. J. Williams, and M. Hegland, editors, *Proceedings of the Australian Data Mining Workshop (ADM'02) at the 15th Australian Joint Conference on Artificial Intelligence*, pages 47–56, Canberra, Australia, December 2002. University of Technology Sydney.

[144] S. Krishnaswamy, A. Zaslasvky, and W. S. Loke. Internet Delivery of Distributed Data Mining Services: Architectures, Issues and Prospects. In V. K. Murthy and N. Shi, editors, *Architectural Issues of Web-enabled Electronic Business*, pages 113–127. Idea Group, 2003.

[145] S. Krishnaswamy and A. Zaslavsky. Activating a Passive Database Using Knowledge Discovery Techniques. *Journal of Computing and Information (JCI)*, 3(1), 1998.

[146] S. Krishnaswamy, A. Zaslavsky, and S. W. Loke. An Architecture to Support Distributed Data Mining Services in E-Commerce Environments. In *Second International Workshop on Advance Issues of E-Commerce and Web-Based Information Systems (WECWIS 2000)*, pages 239–246, Milpitas, CA, June 2000.

[147] S. Krishnaswamy, A. Zaslavsky, and W. S. Loke. Towards Data Mining Services on the Internet with a Multiple Service Provider Model: An XML Based Approach. *Electronic Commerce Research (Special issue on Electronic Commerce and Service Operations)*, 2(3), August 2001.

[148] S. Krishnaswamy, A. Zaslavsky, and W. S. Loke. Techniques for Estimating the Computation and Communication Costs of Distributed Data Mining. In *Proceedings of International Conference on Computational Science (ICCS2002) - Part I*, volume 2331 of *Lecture Notes in Computer Science (LNCS)*, pages 603–612. Springer Verlag, 2002.

[149] T. Krüger, J. Wickel, and K.-F. Kraiss. Parallel and Distributed Computing for an Adaptive Visual Object Retrieval System. In *Proceedings of the 17th International Parallel and Distributed Processing Symposium IPDPS 2003*, France, April 2003.

[150] A. D. Kshemkalyani and M. Singhal. Communication Patterns in Distributed Computations. *Journal of Parallel and Distributed Computing*, 62(6):1104–1119, 2002.

[151] C. Kuengkrai and C. Jaruskulchai. A Parallel Learning Algorithm for Text Classification. In *The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Edmonton, Canada, July 2002.

[152] V. Kumar. Network Intrusion Detection Using Distributed Data Mining. In *Workshop on Data Mining and Exploration Middleware for Distributed and Grid Computing*, Minneapolis, MN, September 2003.

[153] T. Kurc. On-demand Exploration of Very Large, Distributed Datasets in Large-scale Simulation Studies. In *Workshop on Data Mining and Exploration Middleware for Distributed and Grid Computing*, Minneapolis, MN, September 2003.

[154] W. Lam and A. M. Segre. Distributed Data Mining of Probabilistic Knowledge. In *Proceedings of the 17th International Conference on Distributed Computing Systems*, pages 178–185, Washington, DC, 1997. IEEE Computer Society Press.

[155] S. Lander and V. R. Lesser. Customizing Distributed Search Among Agents with Heterogeneous Knowledge. In *Proceedings of the First International Conference on Information and Knowledge Management*, 1992.

[156] S. Lander and V. R. Lesser. Understanding the Role of Negotiation in Distributed Search Among Heterogeneous Agents. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 1993.

[157] A. Lazarevic and Z. Obradovic. The Distributed Boosting Algorithm. In *Knowledge Discovery and Data Mining*, pages 311–316, 2001.

[158] A. Lazarevic and Z. Obradovic. Boosting Algorithms for Parallel and Distributed Learning. *Distributed and Parallel Databases: An International Journal, Special Issue on Parallel and Distributed Data Mining*, 2:203–229, 2002.

[159] A. Lazarevic, D. Pokrajac, and Z. Obradovic. An E-commerce System for Mining Distributed Spatial Databases. In *Proceedings of International Conference on Advances in Infrastructure for Electronic Business, Science, and Education on the Internet*, pages 129–134, LÁquila, Italy, August 2000.

[160] A. Lazarevic, D. Pokrajac, and Z. Obradovic. Distributed Clustering and Local Regression for Knowledge Discovery in Multiple Spatial Databases. In *Proceedings of 8th European Symposium on Artificial Neural Networks*, pages 129–134, Bruges, Belgium, April 2000.

[161] C. Leckie and R. Kotagiri. Learning to Share Distributed Probabilistic Beliefs. In *The Nineteenth International Conference on Machine Learning (ICML2002)*, Sydney, Australia, July 2002.

[162] B. Lesyng, P. Baa, and D. Erwin. EUROGRIDCCEuropean computational grid testbed. *Journal of Parallel and Distributed Computing*, 63(5):590–596, 2003.

[163] K. Li. Scalable Parallel Matrix Multiplication on Distributed Memory Parallel Computers. *Journal of Parallel and Distributed Computing*, 61(12):1709–1731, 2001.

[164] T. Li, S. Zhu, and M. Ogihara. A New Distributed Data Mining Model Based on Similarity. *ACM SAC Data Mining Track*, March 2003.

[165] T. Li, S. Zhu, and M. Ogihara. Algorithms for Clustering High Dimensional and Distributed Data. *Intelligent Data Analysis Journal*, 7(4), 2003.

[166] S.-J. Lim and Y.-K. Ng. A Hybrid Fragmentation Approach for Distributed Deductive Database Systems. *Knowledge and Information Systems*, 3(2):198–224, 2001.

[167] C.-R. Lin, C.-H. Lee, M.-S.Chen, and P. S. Yu. Distributed data Mining in a Chain Store Database of Short Transactions. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 576–581, Edmonton, Canada, 2002. ACM Press.

[168] X. Lin, C. Clifton, and M. Zhu. Privacy Preserving Clustering with Distributed EM Mixture Modeling. *Knowledge and Information Systems*, to appear.

[169] Yehuda Lindell and Benny Pinkas. Privacy Preserving Data Mining. *Journal of Cryptology*, 15(3):177–206, 2002. An extended abstract appeared at the CRYPTO 2000 conference.

[170] J. Liu, H. Li, F. Chan, and F. Lam. A Novel Approach to Fast Discrete Fourier Transform. *Journal of Parallel and Distributed Computing*, 54(1):48–58, 2001.

[171] A. M. Manning and J. A. Keane. Data Allocation Algorithm for Parallel Association Rule Discovery. In *The Fifth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2001)*, Hong Kong, China, April 2001.

[172] C. Mastroianni, D. Talia, and P. Trunfio. Managing Heterogeneous Resources in Data Mining Applications on Grids Using XML-Based Metadata. In *IPDPS*, Nice, France, 2003.

[173] M. Mazzucco, A. Ananthanarayan, R. Grossman, J. Levera, and G. Rao. Merging Multiple Data Streams On Common Keys Over High Performance Networks. In *Proceedings of the 2002 ACM/IEEE Conference on Supercomputing*, pages 1–12, Baltimore, MD, 2002. IEEE Computer Society Press.

[174] S. McClean, R. Páircéir, B. Scotney, and K. Greer. A Negotiation Agent for Distributed Heterogeneous Statistical Databases. *SSDBM 2002*, pages 207–216, 2002.

[175] S. McClean, B. Scotney, and K. Greer. Conceptual Clustering Heterogeneous Distributed Databases. In *Workshop on Distributed and Parallel Knowledge Discovery*, Boston, MA, 2000.

[176] S. McClean, B. Scotney, and K. Greer. A Scalable Approach to Integrating Heterogeneous Aggregate Views of Distributed Databases. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, pages 232–235, 2003.

[177] S. McClean, B. Scotney, and F. Palmer. Conceptual Clustering of Heterogeneous Sequences via Schema Mapping. *ISMIS 2002*, pages 85–93, 2002.

[178] S. McClean, B. Scotney, and F. Palmer. Temporal Probabilistic Concepts from Heterogeneous Data Sequences. *Soft-Ware 2002*, pages 191–205, 2002.

[179] S. I. McClean and B. W. Scotney. Using Evidence Theory for the Integration of Distributed Databases. *International Journal of Intelligent Systems*, 12(10):763–776, 1997.

[180] Mohamed Medhat. Distributed Classification Using OIKI DDM Model.

[181] S. Merugu and J. Ghosh. Privacy-preserving Distributed Clustering using Generative Models. In *The Third IEEE International Conference on Data Mining (ICDM'03)*, Melbourne, FL, November 2003.

[182] S. Morinaga, K. Yamanishi, and Jun ichi Takeuchi. Distributed Cooperative Mining for Information Consortium. In *The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, DC, August 2003.

[183] D. Neiman et al. Exploiting Meta-Level Information in a Distributed Scheduling System. In *Proceeding of the 12th National Conference on Artificial Intelligence*, 1994.

[184] S. Nestorov. Mining Qualified Association Rules in Distributed Databases. In *Workshop on Data Mining and Exploration Middleware for Distributed and Grid Computing*, Minneapolis, MN, September 2003.

[185] C. Nowak. Multiple Databases, Partial Reasoning, and Knowledge Discovery. In X. Wu, R. Kotagiri, and K. B. Korb, editors, *Research and Development in Knowledge Discovery and Data Mining*, volume 1394 of *Lecture Notes in Computer Science : Lecture Notes in Artificial Intelligence*, pages 403–404, New York, NY, 1998. Springer-Verlag.

[186] T. Oates, M. Schmill, and P. R. Cohen. Parallel and Distributed Search for Structure in Multivariate Time Series. In *Machine Learning: ECML-97*, volume 1224 of *Lecture Notes in Computer Science : Lecture Notes in Artificial Intelligence*, pages 191–198, New York, NY, 1997. Springer-Verlag. 9th European Conference on Machine Learning.

[187] M. Oguchi and M. Kitsuregawa. Parallel Data Mining on ATM-connected PC cluster and Optimization of its Execution Environment. In *3rd Workshop on High Performance Data Mining. In conjunction with International Parallel and Distributed Processing Symposium 2000 (IPDPS'00)*, Cancun, Mexico, May 2000.

[188] M. Otey, A. Veloso, C. Wang, S. Parthasarathy, and Wagner Meira Jr. Incremental Techniques for Mining Dynamic and Distributed Databases. In *The Third IEEE International Conference on Data Mining (ICDM'03)*, Melbourne, FL, November 2003.

[189] A. Paccanaro and G. Hinton. Learning Distributed Representations by Mapping Concepts and Relations into a Linear Space. In *The Seventeenth International Conference on Machine Learning (ICML2000)*, Stanford University, CA, June 2000.

[190] R. Pairceir, S. McClean, and B. Scotney. Automated Discovery of Rules and Exceptions from Distributed Databases Using Aggregates. In *Proceedings of 3rd European Conference on Principles and Practice of Knowledge Discovery in Databases*, Prague, Czech Republic, September 1999.

[191] Rónán Páircéir, Sally I. McClean, and Bryan W. Scotney. Discovery of Multi-level Rules and Exceptions From a Distributed Database. In *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 523–532, Boston, MA, 2000. ACM Press.

[192] B. Park. *Knowledge Discovery from Heterogeneous Data Streams Using Fourier Spectrum of Decision Trees*. PhD thesis, Washington State University, 2001. PhD. Dissertation.

[193] B. Park, R. Ayyagari, and H. Kargupta. A Fourier Analysis-Based Approach to Learn Classifier from Distributed Heterogeneous Data. In *Proceedings of the First SIAM International Conference on Data Mining*, Chicago, IL, April 2001.

[194] B. Park and H. Kargupta. The Fourier Spectrum of Decision Trees: Theoretical Issues and Application in Ensemble-based Learning from Data Streams. , 2001. In communication.

[195] B. Park and H. Kargupta. Constructing Simpler Decision Trees from Ensemble Models Using Fourier Analysis. In *Proceedings of the 7th Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD'2002)*, pages 18–23, Madison, WI, June 2002. ACM SIGMOD.

[196] B. Park and H. Kargupta. Distributed Data Mining: Algorithms, Systems, and Applications. In Nong Ye, editor, *Data Mining Handbook*, pages 341–358. IEA, 2002.

[197] B. Park, H. Kargupta, E. Johnson, E. Sanseverino, D. Hershberger, and L. Silvestre. Distributed, Collaborative Data Analysis from Heterogeneous Sites Using a Scalable Evolutionary Technique. *Journal of Applied Intelligence*, 16(1):19–42, 2002.

[198] J. S. Park, M.-S.Chen, and P. S. Yu. Efficient Parallel Data Mining for Association Rules. In *Proceedings of ACM International Conference on Information and Knowledge Management*, pages 31–36, Baltimore, MD, November 1995.

[199] S. Parthasarathy and S. Dwarkadas. Shared State for Distributed Interactive Data Mining Applications. In *International Journal on Distributed and Parallel Databases*, March 2002.

[200] S. Parthasarathy and M. Ogihara. Clustering Distributed Homogeneous Datasets. In *Proceedings of the Fourth European Conference on Principles of Data Mining and Knowledge Discovery*, volume 1910 of *Springer-Verlag Lecture Notes in Computer Science*, pages 566–574, 2000.

[201] S. Parthasarathy and M. Ogihara. Exploiting Dataset Similarity for Distributed Mining. In *3rd Workshop on High Performance Data Mining. In conjunction with International Parallel and Distributed Processing Symposium 2000 (IPDPS'00)*, Cancun, Mexico, May 2000.

[202] S. Parthasarathy, M. Zaki, and W. Li. Memory Placement Techniques for Parallel Association Mining. In *The Fourth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, August 1998.

[203] S. Parthasarathy, M. J. Zaki, M. Ogihara, and W. Li. Parallel Data Mining for Association Rules on Shared-Memory Systems. *Knowledge and Information Systems*, 3(1):1–29, 2001.

[204] José M. Pena and E. Menasalvas. Towards Flexibility in a Distributed Data Mining Framework. In *Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD 2001)*, 2001.

[205] D. W. Pfitzner and J. K. Salmon. Parallel Halo Finding in N-body Cosmology Simulations. In *The Second ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Portland, OR, August 1996.

[206] José C. Pinheiro and D. X. Sun. Methods for Linking and Mining Massive Heterogeneous Databases. In *The Fourth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, August 1998.

[207] Sweta Pittie, Hillol Kargupta, and Byung-Hoon Park. Dependency detection in MobiMine: a systems perspective. *Information Sciences. Special Issue: Knowledge Discovery from Distributed Information Sources*, 155(3–4):227–243, October 2003.

[208] H. Polat and W. Du. Privacy-Preserving Collaborative Filtering using Randomized Perturbation Techniques. In *The Third IEEE International Conference on Data Mining (ICDM'03)*, Melbourne, FL, November 2003.

[209] F. Poulet. Multi-way Distributed SVM algorithms. In *Parallel and Distributed computing for Machine Learning. In conjunction with the 14th European Conference on Machine Learning (ECML'03) and 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'03)*, Cavtat-Dubrovnik, Croatia, September 2003.

[210] I. Pramudiono and M. Kitsuregawa. Parallel WAP-Mine on PC Cluster. In *HPDM: High Performance, Pervasive, and Data Stream Mining 6th International Workshop on High Performance Data Mining: Pervasive and Data Stream Mining (HPDM:PDS'03). In conjunction with Third International SIAM Conference on Data Mining*, San Francisco, CA, May 2003.

[211] A. Prodromidis and P. Chan. Meta-learning in Distributed Data Mining Systems: Issues and Approaches. In Hillol Kargupta and Philip Chan, editors, *Advances of Distributed Data Mining*. MIT/AAAI Press, 2000.

[212] A. Prodromidis and S. J. Stolfo. Mining Databases with Different Schemas: Integrating Incompatible Classifiers. In *Knowledge Discovery and Data Mining*, pages 314–318, 1998.

[213] A. L. Prodromidis, S. J. Stolfo, and P. K. Chan. Pruning Classifiers in a Distributed Meta-Learning System. In *Proceedings of the First National Conference on New Information Technologies*, pages 151–160, 1998.

[214] F. Provost. Distributed Data Mining: Scaling Up and Beyond. In Hillol Kargupta and Philip Chan, editors, *Advances in Distributed Data Mining*. MIT/AAAI Press, 2000.

[215] V. Ramos and J. J. Merelo. *Self-Organized Stigmergic Document Maps: Environment as a Mechanism for Context Learning*, volume 1 of *AEB2002 1st Spanish Conference on Evolutionary and Bio-Inspired Algorithms*, chapter 7, pages 284–293. Centro University de Mérida, Feburary 2002.

[216] V. Ramos and F. Muge. Less is More - Genetic Optimisation of Nearest Neighbour Classifiers. In F. Muge, C. Pinto, and M. Piedade, editors, *10th Portuguese Conference on Pattern Recognition*, pages 293–301, Technical University of Lisbon, March 1998. RecPad.

[217] Vitorino Ramos and Ajith Abraham. Evolving a Stigmergic Self-Organized Data-Mining. In IADIS, editor, *IADIS-04, International Conference on Web Based Communities*, March 2004.

[218] O. Rana, D. Walker, M. Li, S. Lynden, and M. Ward. PaDDMAS: Parallel and Distributed Data Mining Application Suite. In *Fourteenth International Parallel and Distributed Processing Symposium*, pages 387–392, Cancun, Mexico, May 2000.

[219] J. Reinoso-Castillo. Ontolgy-Driven Information Extraction and Integration from Autonomous, Heterogeneous, Distributed Data Sources – A Federated Query-Centric Approach (Master Thesis), 2002.

[220] J. Reinoso-Castillo, A. Silvescu, D. Caragea, J. Pathak, and V. Honavar. Information Extraction and Integration from Heterogeneous, Distributed, Autonomous Information Sources: A Federated, Query-Centric Approach. In *IEEE International Conference on Information Integration and Reuse*, 2003.

[221] N. F. Samatova, G. Ostrouchov, A. Geist, and A. Melechko. RACHET: An Efficient Cover-Based Merging of Clustering Hierarchies from Distributed Datasets. *Distributed and Parallel Databases*, 11(2):157–180, 2002.

[222] S. Sarawagi and S. H. Nagaralu. Data Mining Models as Services on the Internet. *SIGKDD Explorations*, 2(1):24–28, 2000.

[223] M. Sayal and P. Scheuermann. Distributed Web Log Mining Using Maximal Large Itemsets. *Knowledge and Information Systems*, 3(4):389–404, 2001.

[224] J. Schneider, W.-K. Wong, A. Moore, and M. Riedmiller. Distributed Value Functions. In *The Sixteenth International Conference on Machine Learning (ICML99)*, Bled, Slovenia, June 1999.

[225] A. Schuster and R. Wolff. Communication Efficient Distributed Mining of Association Rules. In *ACM SIGMOD*, Santa Barbara, CA, April 2001.

[226] Assaf Schuster, Ran Wolff, and Bobi Gilburd. Privacy-Preserving Association Rule Mining in Large-Scale Distributed Systems. In *Proceedings of Cluster Computing and the Grid (CCGrid)*, 2004.

[227] B. W. Scotney, S. I. McClean, and M. C. Rodgers. Optimal and Efficient Integration of Heterogeneous Summary Tables in a Distributed Database. *Data and Knowledge Engineering*, 29:337–350, 1999.

[228] Bryan Scotney and Sally McClean. Database aggregation of imprecise and uncertain evidence. *Information Sciences. Special Issue: Knowledge Discovery from Distributed Information Sources*, 155(3–4):245–263, October 2003.

[229] C. Shahabi, L. Khan, and D. McLeod. A Probe-Based Technique to Optimize Join Queries in Distributed Internet Databases. *Knowledge and Information Systems*, 2(3):373–385, 2001.

[230] E. C. Shek, R. R. Muntz, E. Mesrobian, and K. W. Ng. Scalable Exploratory Data Mining of Distributed Geoscientific Data. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 32–37, Portland, OR, 1996.

[231] Mei-Ling Shyu, Choochart Haruechaiyasak, and Shu-Ching Chen. Category cluster discovery from distributed WWW directories. *Information Sciences. Special Issue: Knowledge Discovery from Distributed Information Sources*, 155(3–4):181197, October 2003.

[232] K. Sivakumar, R. Chen, and H. Kargupta. Learning Bayesian Network Structure from Distributed Data. In *Proceedings of the 3rd SIAM International Data Mining Conference*, pages 284–288, San Franciso, CA, May 2003.

[233] D. B. Skillicorn and Y. Wang. Parallel and Sequential Algorithms for Data Mining Using Inductive Logic. *Knowledge and Information Systems*, 3(4):405–421, 2001.

[234] R. Sterritt, K. Adamson, C. M. Shapcott, and E. P. Curran. Parallel Data Mining of Bayesian Networks From Telecommunications Network Data. In *3rd Workshop on High Performance Data Mining. In conjunction with International Parallel and Distributed Processing Symposium 2000 (IPDPS'00)*, Cancun, Mexico, May 2000.

[235] S. Stolfo, H. Dewan, D. Ohsie, and M. Hernandez. A Parallel and Distributed Environment for Database Rule Processing, Open Problems and Future Directions. In *Emerging Trends in Database and Knowledge-based Machines IEEE Press*, 1995.

[236] S. Stolfo et al. JAM: Java Agents for Meta-Learning over Distributed Databases. In *Proceedings of Third International Conference on Knowledge Discovery and Data Mining*, pages 74–81, Menlo Park, CA, 1997. AAAI Press.

[237] S. Stolfo, W. Fan, W. Lee, A. Prodromidis, and P. Chan. Cost-based Modeling for Fraud and Intrusion Detection: Results from the JAM Project. In *Proceedings of the 2000 DARPA Information Survivability Conference and Exposition (DISCEX '00)*, 2000.

[238] R. Subramonian and S. Parthasarathy. An Architecture for Distributed Data Mining. In *Fourth International Conference of Knowledge Discovery and Data Mining*, pages 44–59, New York, NY, 1998.

[239] V. Sunderam. Towards Service-Based Approaches to Data Mining in Grids. In *Workshop on Data Mining and Exploration Middleware for Distributed and Grid Computing*, Minneapolis, MN, September 2003.

[240] W. Sutandiyo, B. M. Chhetri, S. Krishnaswamy, and W. S. Loke. From m-GAIA to Grasshopper: Engineering Mobile Agent Applications. In *The Fifth International Conference on Information Integration and Web-based Applications and Services (IIWAS2003)*, Jakarta, Indonesia, September 2003.

[241] D. Talia. Grid-based Data Mining and the Knowledge Grid Framework. In *Workshop on Data Mining and Exploration Middleware for Distributed and Grid Computing*, Minneapolis, MN, September 2003.

[242] A. Topchy, A. Jain, and W. Punch. Combining Multiple Weak Clusterings. In *The Third IEEE International Conference on Data Mining (ICDM'03)*, Melbourne, FL, November 2003.

[243] G. Tsoumakas and I. Vlahavas. Distributed Data Mining of Large Classifier Ensembles. In *Proceedings of Companion Volume of the Second Hellenic Conference on Artificial Intelligence*, pages 249–256, Thessaloniki, Greece, April 2002.

[244] Grigorios Tsoumakas, Lefteris Angelis, and Ioannis Vlahavas. Similarity Based Distributed Classification.

[245] Grigorios Tsoumakas and Ioannis Vlahavas. Effective Stacking of Distributed Classifiers. In *Proceedings of the 15th European Conference on Artificial Intelligence*, pages 340–344, 2002.

[246] K. Tumer and J. Ghosh. Robust Order Statistics based Ensemble for Distributed Data Mining. In Philip Chan Hillol Kargupta, editor, *Advances in Distributed and Parallel Knowledge Discovery*. MIT/AAAI Press, 2000.

[247] A. Tveit and H. Engum. Parallelization of the Incremental Proximal Support Vector Machine Classifier using a Heap-based Tree Topology. In *Parallel and Distributed computing for Machine Learning. In conjunction with the 14th European Conference on*

Machine Learning (ECML'03) and 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'03), Cavtat-Dubrovnik, Croatia, September 2003.

[248] V. Ćurčin, M. Ghanem, Y. Guo, M. Köhler, A. Rowe, J. Syed, and P. Wendel. Discovery Net: Towards a Grid of Knowledge Discovery. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 658–663, Edmonton, Canada, 2002. ACM Press.

[249] J. Vaidya and C. Clifton. Privacy Preserving Association Rule Mining in Vertically Partitioned Data. In *The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Edmonton, Canada, July 2002.

[250] J. Vaidya and C. Clifton. Leveraging the "Multi" in Secure Multi-Party Computation. In *Workshop on Privacy in the Electronic Society held in association with the 10th ACM Conference on Computer and Communications Security*, Washington, DC, October 2003.

[251] J. Vaidya and C. Clifton. Privacy-Preserving K-Means Clustering over Vertically Partitioned Data. In *The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, DC, August 2003.

[252] S. Vucetic and Z. Obradovic. Discovering Homogeneous Regions in Spatial Data through Competition. In *The Seventeenth International Conference on Machine Learning (ICML2000)*, Stanford University, CA, June 2000.

[253] S. Vucetic and Z. Obradovic. Performance Controlled Data Reduction for Knowledge Discovery in Distributed Databases. In *The Fourth Pacific- Asia Conference on Knowledge Discovery and Data Mining (PAKDD2000)*, Kyoto, Japan, April 2000.

[254] Changzhou Wang and Xiaoyang Sean Wang. High-Dimensional Nearest Neighbor Search with Remote Data Centers. *Knowledge and Information Systems*, 4(4):440–465, 2002.

[255] F.-H. Wang, J.-M. Chang, Y.-L. Wang, and S.-J Huang. Distributed Algorithms for Finding the Unique Minimum Distance Dominating Set in Directed Split-stars. *Journal of Parallel and Distributed Computing*, 63(4):481–487, 2003.

[256] G. Weiß. A Multiagent Perspective of Parallel and Distributed Machine Learning. In K. P. Sycara and M. Wooldridge, editors, *Proceedings of the 2nd International Conference on Autonomous Agents (Agents'98)*, pages 226–230, New York, NY, 1998. ACM Press.

[257] P. Wendel and Y.-K. Guo. The Design of a Platform for Distributed KDD Components. In *5th International Workshop on High Performance Data Mining: Resource and Location Aware Mining (HPDM:RLM'02). In conjunction with Second International SIAM Conference on Data Mining*, Arlington, VA, April 2002.

[258] R. Wirth, M. Borth, and J. Hipp. When Distribution is Part of the Semantics: A New Problem Class for Distributed Knowledge Discovery. In *Proceedings of PKDD-2001 Workshop on Ubiquitous Data Mining for Mobile and Distributed Environments*, pages 56–64, Freiburg, Germany, September 2001.

[259] R. Wolff and A. Schuster. Mining Association Rules in Peer-to-Peer Systems. In *The Third IEEE International Conference on Data Mining (ICDM'03)*, Melbourne, FL, November 2003.

[260] R. Wolff, A. Schuster, and D. Trock. A High-Performance Distributed Algorithm for Mining Association Rules. In *The Third IEEE International Conference on Data Mining (ICDM'03)*, Melbourne, FL, November 2003.

[261] J. Wu. A Distributed Formation of Smallest Faulty Orthogonal Convex Polygons in 2-D Meshes. *Journal of Parallel and Distributed Computing*, 62(7):1168–1185, 2002.

[262] Y. Xing, M. G. Madden, J. Duggan, and G. J. Lyons. Context-based Distributed Regression in Virtual Organizations . In *Parallel and Distributed computing for Machine Learning. In conjunction with the 14th European Conference on Machine Learning (ECML'03) and 7th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'03)*, Cavtat-Dubrovnik, Croatia, September 2003.

[263] Y. Xing, M. G. Madden, J. Duggan, and G. J. Lyons. Distributed Regression for Heterogeneous Data Sets. In M. R. Berthold, H.-J. Lenz, E. Bradley, R. Kruse, and C. Borgelt, editors, *Proceedings of 5th International Symposium on Intelligent Data Analysis (IDA2003)*, LNCS 2810, pages 544–553, Berlin, German, August 2003. Springer.

[264] K. Yamanishi. Distributed Cooperative Bayesian Learning Strategies. In *Proceedings of COLT 97*, pages 250–262, New York, NY, 1997. ACM.

[265] J. Yang, V. Honavar, L. Miller, and J. Wong. Intelligent Mobile Agents for Information Retrieval and Knowledge Discovery from Distributed Data and Knowledge Sources. In *IEEE Information Technology Conference*, Syracuse, NY, 1998.

[266] B. Yi, N. Sidiropoulos, T. Johnson, H. V. Jagadish, C. Faloutsos, and A. Biliris. Online Data Mining for Co-Evolving Time Sequences. In *Proceedings of the 2000 International Conference on Data Engineering*, pages 13–22, 2000.

[267] H. Yu and Ee-Chien Chang. Distributed Multivariate Regression Based on Influential Observations. In *The Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, DC, August 2003.

[268] H. Yu, K. Chang, and J. Han. Heterogeneous Learner for Web Page Classification. In *Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM 2002)*, pages 538–545, Maebashi City, Japan, December 2002. IEEE Computer Society.

[269] O. Zaiane, M. El-Hajj, and P. Lu. Fast Parallel Association Rules Mining without Candidacy Generation. In *IEEE 2001 International Conference on Data Mining (ICDM'2001)*, pages 665–668, 2001.

[270] M. Zaki. Parallel and Distributed Association Mining: A Survey. *IEEE Concurrency*, 1999.

[271] M. Zaki. Parallel Sequence Mining on Shared-Memory Machines. *Journal of Parallel and Distributed Computing*, 61(3):401–426, 2001.

[272] M. Zaki, M. Ogihara, S. Parthasarathy, and W. Li. Parallel Data Mining for Association Rules on Shared-Memory Multiprocessors. In *Proceedings of Supercomputing'96*, pages 17–22, Pittsburg, PA, November 1996.

[273] M. Zaki and Y. Pin. Introduction: Recent Development in Parallel and Distributed Data Mining. *Distributed and Parallel Databases*, 11(2), 2002.

[274] B. Zeigler, H. Cho, J. Kim, H. Sarjoughian, and J. Lee. Quantization-based filtering in distributed discrete event simulation. *Journal of Parallel and Distributed Computing*, 62(11):1629–1647, 2002.

[275] Y. Zhang, G. Owen, S. Prasad, R. Sunderraman, and G. Vachtsevano. Intelligent Internet2 Agents for Distributed Data Mining. In *The Internet2 Network Research Workshop*, June 2000.

[276] Y.-Q. Zhang. On Distributed Fuzzy Relational Databases. *Microcomputers*, 7(3):42–49, 1987.

[277] Y.-Q. Zhang. Research in the Distributed Fuzzy Relational Database. In *The 3rd Sino-Japanese Shenyang-Sapparo International Conference on Computer Applications*, 1988.

[278] K. Zhao, B. Liu, T. Tirpak, and A. Schaller. Detecting Patterns of Change Using Enhanced Parallel Coordinate Visualization. In *The Third IEEE International Conference on Data Mining (ICDM'03)*, Melbourne, FL, November 2003.