Distributed Data Mining Bibliography *

Kanishka Bhaduri, Kamalika Das, Kun Liu, Hillol Kargupta and Jessica Ryan
Computer Science and Electrical Engineering Department
University of Maryland Baltimore County
Baltimore, Maryland 21250
{kanishk1, kdas1, kunliu1, hillol, jryan4}@cs.umbc.edu

July, 2008 (Release 1.9)

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.

Distributed Association Rule Mining Bibliography

References

[1] 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.

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

[3] 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.

[4] 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.

[5] H. Cheng, P. Tan, S. Jon, and W. Punch. Recommendation via Query Centered Random Walk on K-partite Graph. In *Proceedings of the IEEE International Conference on Data Mining (ICDM '07)*, pages 457–462, Omaha, NE, 2007.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] A. Javed and A. Khokhar. Frequent Pattern Mining on Message Passing Multiprocessor Systems. *Distributed and Parallel Databases* , 16(3):321–334, November 2004.

[11] Asif Javed and Ashfaq Khokhar. Frequent pattern mining on message passing multiprocessor systems. *Distributed and Parallel Databases*, 16(3):321 – 334, November 2004.

[12] 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.

[13] S. Li, T. Wu, and W. M. Pottenger. Distributed Higher Order Association Rule Mining Using Information Extracted from Textual Data. *SIGKDD Exploration*, 7(1):26–35, 2005.

[14] A. Manjhi, V. Shkapenyuk, K. Dhamdhere, and C. Olston. Finding (Recently) Frequent Items in Distributed Data Streams. In *Proceedings of the 21st International Conference on Data Engineering (ICDE'05)*, Tokyo, Japan, April 2005.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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, February 2001.

[20] I. Pramudiono and M. Kitsuregawa. Parallel FP-Growth on PC Cluster. In *Proceedings of the Seventh Pacific-Asia Conference of Knowledge Discovery and Data Mining (PAKDD03)*, pages 467–473, Seoul, Korea, April – May 2003.

[21] Iko Pramudiono and Masaru Kitsuregawa. Parallel FP-Growth on PC cluster. In *Advances in Knowledge Discovery and Data Mining: 7th Pacific-Asia Conference (PAKDD)*, Seoul, Korea, April-May 2003.

[22] A. Schuster and R. Wolff. Communication-Efficient Distributed Mining of Association Rules. *Data Mining and Knowledge Discovery*, 8(2), March 2004.

[23] Assaf Schuster and Ran Wolff. Communication Efficient Distributed Mining of Association Rules. In *Proceedings of the 2001 ACM SIGMOD International Conference on Management of Data*, volume 30, pages 473–484, California, USA, June 2001.

[24] 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.

[25] Assaf Schuster, Ran Wolff, and Dan Trock. A High-Performance Distributed Algorithm for Mining Association Rules . In *Third IEEE International Conference on Data Mining*, Florida , USA, November 2003.

[26] D. B. Skillicorn. Parallel Frequent Set Counting. *Parallel Computing*, 28(5):815–825, May 2002.

[27] D. B. Skillicorn. Parallel frequent set counting. *Distributed and Parallel Databases*, 28(5):815 – 825, May 2002.

[28] 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.

[29] 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)*, November 2003.

[30] Ran Wolff and Assaf Schuster. Association Rule Mining in Peer-to-Peer Systems . In *Third IEEE International Conference on Data Mining*, Melbourne, FL, November 2003.

[31] 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.

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

[33] 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.

Distributed Classification Bibliography

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] 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.

[3] 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.

[4] A. Bar-Or, A. Schuster, R. Wolff, and D. Keren. Hierarchical Decision Tree Induction in Distributed Genomic Databases. *Accepted for IEEE Transactions on Knowledge and Data Engineering – Special Issue on Mining Biological Data*, 2005.

[5] A. Bar-Or, R. Wolff, A. Schuster, and D. Keren. Decision Tree Induction in High Dimensional, Hierarchically Distributed Databases. In *Proceedings of 2005 SIAM International Conference on Data Mining (SDM'05)*, Newport Beach, CA, April 2005.

[6] J. Basak and R. Kothari. A Classification Paradigm for Distributed Vertically Partitioned Data. *Neural Computation*, 16(7):1525–1544, July 2004.

[7] K. Bhaduri, R. Wolff, C. Giannella, and H. Kargupta. Distributed decision-tree induction in peer-to-peer systems. *Stat. Anal. Data Min.*, 1(2):85–103, 2008.

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

[9] 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.

[10] D. Caragea. *Learning Classifiers from Distributed, Semantically Heterogeneous, Autonomous Data Sources*. PhD thesis, Iowa State University, 2004.

[11] D. Caragea, J. Pathak, and V. Honavar. Learning Classifiers from Semantically Heterogeneous Data. In *Proceedings of third International Conference on Ontologies, DataBases and Applications of Semantics for Large Scale Information Systems (ODBASE)*, Agia Napa, Cyprus, October 2004.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

[19] 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.

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

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

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

[23] 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.

[24] 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.

[25] Nitesh V. Chawla, Lawrence O. Hall, Kevin W. Bowyer, and W. Philip Kegelmeyer. Learning Ensembles from Bites: A Scalable and Accurate Approach. *Journal of Machine Learning Research*, 5:421–451, April 2004.

[26] 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.

[27] 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.

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

[29] R. Chen, K. Sivakumar, and H. Kargupta. Collective Mining of Bayesian Networks from Distributed Heterogeneous Data. *Knowledge and Information Systems*, 6(2):164–187, March 2004.

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

[31] A. D'Costa, V. Ramachandran, and A. Sayeed. Distributed Classification of Gaussian Space-Time Sources in Wireless Sensor Networks. *IEEE Journal of Selected Areas in Communications*, 22(6), August 2004.

[32] M. Duarte and Y.-H. Hu. Distance Based Decision Fusion in a Distributed Wireless Sensor Network. *Telecommunication Systems*, 26(2–4):339–350, 2004.

[33] 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.

[34] 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.

[35] 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.

[36] Chris Giannella, Kun Liu, Todd Olsen, and Hillol Kargupta. Communication Efficient Construction of Decision Trees Over Heterogeneously Distributed Data. In *Proceedings of The Fourth IEEE International Conference on Data Mining (ICDM'04)*, Brighton, UK, November 2004.

[37] Vladimir Gorodetsky, Oleg Karsaeyv, and Vladimir Samoilov. Multi-agent technology for distributed data mining and classification. In *IEEE/WIC International Conference on Intelligent Agent Technology (IAT 2003)*, October 2003.

[38] 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.

[39] 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.

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

[41] V. Guralnik and G. Karypis. Parallel Tree-projection-based Sequence Mining Algorithms. *Parallel Computing*, 30:443–472, April 2004.

[42] 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.

[43] Lawrence O. Hall, Nitesh Chawla, Kevin W. Bowyer, and W. Philip Kegelmeyer. Learning Rules from Distributed Data. *Large-Scale Parallel Data Mining*, 1729:211–220, July 2003.

[44] 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.

[45] 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.

[46] R. Jin and H. Liu. SWITCH: A Novel Approach to Ensemble Learning for Heterogeneous Data. In *The 15th European Conference on Machine Learning (ECML) and the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)* , Pisa, Italy, September 2004.

[47] Murat Kantarcioglu and Chris Clifton. Privately Computing a Distributed k-nn Classifier. In *15th European Conference on Machine Learning (ECML) and the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*, Pisa, Italy, September 2004.

[48] 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.

[49] 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.

[50] Hillol Kargupta and Haimonti Dutta. Orthogonal Decision Trees. In *Proceedings of The Fourth IEEE International Conference on Data Mining (ICDM'04)*, Brighton, UK, November 2004.

[51] 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.

[52] Chak-Man Lam, Xiao-Feng Zhang, and William K. Cheung. Mining local data sources for learning global cluster models. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI04)*, 2004.

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

[54] 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.

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

[56] Elio Lozano and Edgar Acuna. Parallel algorithms for distance-based and density-based outliers. In *Proceedings of the Fifth IEEE International Conference on Data Mining*, Houston, Texas, August 2005.

[57] P. Luo, H. Xiong, K. Lu, and Z. Shi. Distributed Classification in Peer-to-Peer Networks. In *Proceedings of the 13th International Conference on Knowledge Discovery and Data Mining (KDD '07)*, pages 968–976, New York NY, 2007.

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

[59] 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.

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

[61] 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.

[62] 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.

[63] 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.

[64] 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.

[65] 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.

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

[67] 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.

[68] 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.

[69] Martin Scholz. On the complexity of rule discovery from distributed data. In *Proceedings of the Fifth IEEE International Conference on Data Mining*, Houston, Texas, August 2005.

[70] 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.

[71] 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.

[72] 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.

[73] 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.

[74] 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.

[75] Tsoumakas, Katakis, and Vlahavas. Effective Voting of Heterogeneous Classifiers. In *15th European Conference on Machine Learning (ECML) and the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*, Pisa, Italy, September 2004.

[76] G. Tsoumakas, I. Katakis, and I. Vlahavas. Effective Voting of Heterogeneous Classifiers. In *Proceedings of the 15th European Conference on Machine Learning, ECML'04*, volume LNAI 3201, pages 465–476, Pisa, Italy, Spetember 2004. Springer-Verlag.

[77] G. Tsoumakas, L. Katakis, and I. Vlahavas. Effective Voting of Heterogeneous Classifiers. In *The 15th European Conference on Machine Learning (ECML) and the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)* , Pisa, Italy, September 2004.

[78] 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.

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

[80] Grigorios Tsoumakas, Lefteris Angelis, and Ioannis Vlahavas. Clustering Classifiers for Knowledge Discovery from Physically Distributed Databases. *Data and Knowledge Engineering*, 49(3):223–242, June 2004.

[81] Grigorios Tsoumakas, Lefteris Angelis, and Ioannis Vlahavas. Clustering classifiers for knowledge discovery from physically distributed databases. *Data and Knowledge Engineering*, 49(3), June 2004.

[82] Grigorios Tsoumakas, Nick Bassiliades, and Ioannis Vlahavas. *A Knowledge-based Web Information System for the Fusion of Distributed Classifiers*, chapter 8, pages 271–308. IDEA Group, 2004.

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

[84] 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.

[85] 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.

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

[87] S. Wu, K. Chuang, C. Chen, and M. Chen. DIKNN: An Itinerary-based KNN Query Processing Algorithm for Mobile Sensor Networks. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE '07)*, pages 456–465, Istanbul, Turkey, 2007.

[88] W. Wu, W. Guo, and K. Tan. Distributed Processing of Moving K-Nearest-Neighbor Query on Moving Objects. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE '07)*, pages 1116–1125, Istanbul, Turkey, 2007.

[89] 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.

[90] 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.

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

[92] 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.

[93] 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.

Distributed Clustering Bibliography

References

[1] 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.

[2] S. Bandyopadhyay, C. Gianella, U. Maulik, H. Kargupta, K. Liu, and S. Datta. Clustering Distributed Data Streams in Peer-to-Peer Environments. *Information Science Journal (In Press)*, 2004.

[3] S. Datta, C. Giannella, and H. Kargupta. K-Means Clustering over a Large, Dynamic Network. In *Proceedings of 2006 SIAM Conference on Data Mining*, Bethesda, MD, April 2006.

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

[5] 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.

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

[7] 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.

[8] 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.

[9] A. Gionis, H. Mannila, and P. Tsaparas. Clustering Aggregation. In *Proceedings of the 21st International Conference on Data Engineering (ICDE'05)*, Tokyo, Japan, April 2005.

[10] 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.

[11] 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.

[12] K. Hammouda and M. Kamel. HP2PC: Scalable Hierarchically-Distributed Peer-to-Peer Clustering. In *Proceedings of the 2007 SIAM International Conference on Data Mining (SDM '07)*, Philadelphia, PA, 2007.

[13] E. Hung and D. Cheung. Parallel Mining of Outliers in Large Database. *Distributed and Parallel Databases*, 12:5–26, July 2002.

[14] E. Januzaj, H. P. Kriegel, and M. Pfeifle. Scalable Density Based Distributed Clustering. In *Proceedings of EDBT*, volume 2992 of *Lecture Notes in Computer Science*, pages 88–105, March 2004.

[15] E. Januzaj, H.-P. Kriegel, and M. Pfeifle. Scalable Density-Based Distributed Clustering. In *The 15th European Conference on Machine Learning (ECML) and the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)* , Pisa, Italy, September 2004.

[16] 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.

[17] 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.

[18] 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.

[19] M. Klusch, S. Lodi, and G. L. Moro. Distributed Clustering Based on Sampling Local Density Estimates. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI 2003)*, pages 485–490, Mexico, August 2003.

[20] H.-P. Kriegel, P. Kröger, A. Pryakhin, and M. Schubert. Effective and Efficient Distributed Model-based Clustering. In *The Fifth IEEE International Conference on Data Mining (ICDM'05)*, Houston, TX, November 2005.

[21] P. Kunath, H.-P. Kriegel, M. Pfeifle, and M. Renz. Approximated Clustering of Distributed High-Dimensional Data. In *Proceedings of the Ninth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'05)*, Hanoi, Vietnam, May 2005.

[22] 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.

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

[24] E. Lozano and E. Acuna. Parallel Algorithms for Distance-based and Density-based Outliers. In *The Fifth IEEE International Conference on Data Mining (ICDM'05)*, Houston, TX, November 2005.

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

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

[27] Samer Nassar, Jörg Sander, and Corrine Cheng. Incremental and Effective Data Summarization for Dynamic Hierarchical Clustering. In *Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data*, pages 467–478, Paris, France, June 2004.

[28] 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.

[29] 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.

[30] 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.

[31] 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.

[32] 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.

[33] G. Tsoumakas, L. Angelis, and I. Vlahavas. Clustering Classifiers for Knowledge Discovery from Physically Distributed Databases. *Data and Knowledge Engineering*, 49(3):223–242, June 2004.

[34] 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.

[35] 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.

[36] X. Xu, N. Yuruk, Z. Feng, and T. Schweiger. SCAN: A Structural Clustering Algorithm for Networks. In *Proceedings of the 13th International Conference on Knowledge Discovery and Data Mining (KDD '07)*, pages 824–833, New York NY, 2007.

[37] Q. Zhang, J. Liu, and W. Wang. Approximate clustering on distributed data streams. In *ICDE*, pages 1131–1139, 2008.

Distributed Stream Mining Bibliography

References

[1] 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.

[2] 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.

[3] Magdalena Balazinska, Hari Balakrishnan, Sam Madden, and Michael Stonebraker. Fault-tolerance in the borealis distributed stream processing system. In *ACM SIGMOD/PODS 2005 Conference*, Baltimore, Maryland, June 2005.

[4] Gong Chen, Xindong Wu, and Xingquan Zhu. Sequential pattern mining in multiple streams. In *Proceedings of the Fifth IEEE International Conference on Data Mining*, Houston, Texas, August 2005.

[5] J. Chi, M. Koyuturk, and A. Grama. Conquest: A Distributed Tool for Constructing Summaries of High-Dimensional Discrete Attribute Data Sets. In *Proceedings of 2004 SIAM International Conference on Data Mining (SDM'04)*, Lake Buena Vista, FL, April 2004.

[6] G. Cormode, S. Muthukrishnan, and W. Zhuang. Conquering the Divide: Continuous Clustering of Distributed Data Streams. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE '07)*, pages 1036–1045, Istanbul, Turkey, 2007.

[7] Graham Cormode, Minos Garofalakis, S. Muthukrishnan, and Rajeev Rastogi. Holistic aggregates in a networked world: Distributed tracking of approximate quantiles. In *ACM SIGMOD/PODS 2005 Conference*, Baltimore, Maryland, June 2005.

[8] Graham Cormode and S. Muthukrishnan. Space efficient mining of multigraph streams. In *ACM SIGMOD/PODS 2005 Conference*, Baltimore, Maryland, June 2005.

[9] Graham Cormode, S. Muthukrishnan, and Wei Zhuang. What's different: Distributed, continuous monitoring of duplicate-resilient aggregates on data streams. In *Proceedings of the twenty-second International Conference on Data Engineering (ICDE 2006)*, Atlanta, Georgia, April 2006.

[10] 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.

[11] Sumit Ganguly, Minos Garofalakis, Amit Kumar, and Rajeev Rastogi. Join-distinct aggregate estimation over update streams. In *ACM SIGMOD/PODS 2005 Conference*, Baltimore, Maryland, June 2005.

[12] P. Gibbons and S. Tirthapura. Estimating simple functions on the union of data streams. In *ACM Symposium on Parallel Algorithms and Architectures*, pages 281–291, 2001.

[13] Jeong-Hyon Hwang, Magdalena Balazinska, Alexander Rasin, Ugur Cetintemel, Mike Stonebraker, and Stan Zdonik. High-availability algorithms for distributed stream processing. In *The 21st International Conference on Data Engineering (ICDE 2005)*, Tokyo, Japan, April 2005.

[14] H. Kargupta, R. Bhargava, K. Liu, M. Powers, P. Blair, S. Bushra, J. Dull, K. Sarkar, M. Klein, M. Vasa, and D. Handy. VEDAS: A Mobile and Distributed Data Stream Mining System for Real-time Vehicle Monitoring. In *Proceedings of 2004 SIAM International Conference on Data Mining (SDM'04)*, Lake Buena Vista, FL, April 2004.

[15] 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.

[16] 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.

[17] K. K. Loo, I. Tong, B. Kao, and D. Cheung. Online Algorithms for Mining Inter-Stream Associations From Large Sensor Networks. In *Proceedings of the Ninth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'05)*, Hanoi, Vietnam, May 2005.

[18] Amit Manjhi, Suman Nath, and Phillip B. Gibbons. Tributaries and deltas: Efficient and robust aggregation in sensor network streams. In *ACM SIGMOD/PODS 2005 Conference*, Baltimore, Maryland, June 2005.

[19] 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.

[20] C. Olston, J. Jiang, and J. Widom. Adaptive filters for continuous queries over distributed data streams. In *Proceedings of the 2003 ACM SIGMOD international conference on Management of data*, pages 563–574, San Diego, California, 2003.

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

[22] 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.

[23] 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.

[24] I. Sharfman, Assaf Schuster, and Daniel Keren. A geometric appraoch to monitoring threshold functions over distributed data streams. In *Proceedings of the SIGMOD 2006*, Chicago, Illinois, June 2006.

[25] Utkarsh Srivastava, Kamesh Munagala, and Jennifer Widom. Operator placement for in-network stream query processing. In *ACM SIGMOD/PODS 2005 Conference*, Baltimore, Maryland, June 2005.

[26] 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.

[27] A. Zhou, F. Cao, Y. Yan, C. Sha, and X. He. Distributed Data Stream Clustering: A Fast EM-based Approach. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE '07)*, pages 736–745, Istanbul, Turkey, 2007.

[28] Y. Zhou, B. Chin Ooi, and K.-L. Tan. Dynamic Load Management for Distributed Continuous Query Systems. In *Proceedings of the 21st International Conference on Data Engineering (ICDE'05)*, Tokyo, Japan, April 2005.

Distributed Systems (P2P, Sensor net, Grids and more) Bibliography

References

[1] A. Agostini and G. Moro. Identification of communities of peers by trust and reputation. In *Proceedings of AIMSA'04*, volume 3192 of *Lecture Notes in Computer Science*, pages 85–95, September 2004.

[2] Ali Shaikh Ali and Ian J. Taylor. Web services composition for distributed data mining. In *Proceedings of the 2005 International Conference on Parallel Processing Workshops (ICPPW'05)*, Washington DC, 2005.

[3] Kamal Ali and Wijnand Van Stam. TiVo: Making Show Recommendations Using a Distributed Collaborative Filtering Architecture. In *Proceedings of The Tenth ACM SIGKDD Conference (KDD'04)*, Seattle, WA, August 2004.

[4] B. Arai, S. Lin, and D. Gunopulos. Efficient Data Sampling in Heterogeneous Peer-to-Peer Networks. In *Proceedings of the IEEE International Conference on Data Mining (ICDM '07)*, pages 23–32, Omaha, NE, 2007.

[5] Benjamin Arai, Gautam Das, Dimitrios Gunopulos, and Vana Kalogeraki. Approximating aggregations in peer-to-peer databases. In *Proceedings of the twenty-second International Conference on Data Engineering (ICDE 2006)*, Atlanta, Georgia, April 2006.

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

[7] Asad Awan, Ronaldo A. Ferreira, Suresh Jagannathan, and Ananth Grama. Distributed uniform sampling in unstructured peer-to-peer networks. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences*, volume 9, page 223.3, January 2006.

[8] W.-T. Balke, W. Nejdl, W. Siberski, and U. Thaden. DL meets P2P - Distributed Document Retrieval based on Classification and Content. In *European Conference on Digital Libraries (ECDL)*, Vienna, Austria, 2005.

[9] W.-T. Balke, W. Nejdl, W. Siberski, and U. Thaden. Progressive Distributed Top k Retrieval in Peer-to-Peer Networks. In *Proceedings of the 21st International Conference on Data Engineering (ICDE'05)*, Tokyo, Japan, April 2005.

[10] Wolf-Tilo Balke, Wolfgang Nejdl, Wolf Siberski, and Uwe Thaden. Progressive distributed top-k retrieval in peer-to-peer networks. In *The 21st International Conference on Data Engineering (ICDE 2005)*, Tokyo, Japan, April 2005.

[11] S. Banyopadhyay, C. Giannella, U. Maulik, H. Kargupta, S. Datta, and K. Liu. Clustering distributed data streams in peer-to-peer environments. *Information Science*, 176(14):1952–1985, 2006.

[12] Mayank Bawa, Aristides Gionis, Hector Garcia-Molina, and Rajeev Motwani. The price of validity in dynamic networks. In *Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, Paris, France, 2004.

[13] 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.

[14] A. Bhattacharya, A. Meka, and A. Singh. MIST: Distributed Indexing and Querying in Sensor Networks using Statistical Models. In *Proceedings of the 33rd Very Large Data Bases Conference (VLDB '07)*, pages 854–865, New York, NY, 2007.

[15] Yitzhak Birk, Idit Keidar, Liran Liss, Assaf Schuster, and Ran Wolff. Veracity radius - capturing the locality of distributed computations. In *Proceedings of the 25th ACM Symposium on Principles of Distributed Computing (PODC '06)*, pages 102–111, July 2006.

[16] M. Bonifacio, P. Bouquet, G. Mameli, and M. Nori. Peer-mediated distributed knowledge management. In *Agent-Mediated Knowledge Management: International Symposium (AMKM'03)*, Stanford, CA, March 2003.

[17] J. Branch, B. Szymanski, R. Wolff, C. Gianella, and H. Kargupta. In-network outlier detection in wireless sensor networks. In *Proceedings of the 26th International Conference on Distributed Computing Systems (ICDCS)*, pages 102–111, July 2006.

[18] 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.

[19] 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.

[20] 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.

[21] M. Cannataro, A. Congiusta, C. Mastroianni, A. Pugliese, D. Talia, and P. Trunfio. *Grid-Based Data Mining and Knowledge Discovery*, chapter 2, pages 19–45. Intelligent Technologies for Information Analysis. Springer-Verlag, 2004.

[22] M. Cannataro, A. Congiusta, A. Pugliese, D. Talia, and P. Trunfio. Distributed data mining on grids: Services, tools, and applications. *IEEE Transactions on Systems, Man, Cybernetics, Part B*, 34(6), 2004.

[23] 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.

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

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

[26] Mario Cannataro, Domenico Talia, and Paolo Trunfio. Distributed data mining on the grid. *Grid computing: Towards a new computing infrastructure*, 18(8):1101–1112, 2002.

[27] 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.

[28] 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.

[29] 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.

[30] S. Castano and S. Montanelli. Semantic self-formation of communities of peers. In *Proceedings of the ESWC Workshop on Ontologies in Peer-to-Peer Communities*, Heraklion, Greece, May 2005.

[31] 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.

[32] 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.

[33] 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.

[34] V. Cholvi, P. Felber, and E. Biersack. Efficient Search in Unstructured Peer-to-Peer Networks. In *Proceedings of the Sixteenth Annual ACM Symposium on Parallelism in Algorithms and Architectures*, pages 271–272, Barcelona, Spain, 2004.

[35] Laurence Cholvy and Christophe Garion. Answering Queries Addressed to Several Databases According to a Majority Merging Approach. *Journal of Intelligent Information Systems*, 22:175–201, March 2004.

[36] 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.

[37] A. Congiusta, D. Talia, and P. Trunfio. *Parallel and Grid-Based Data Mining*, chapter 1, pages 1–25. n Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers. Spinger Verlag, 2005.

[38] C. Cooper, M. Dyer, and C. Greenhill. Sampling Regular Graphs and a Peer-to-Peer Network. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'05)*, pages 980–988, Vancouver, British Columbia, 2005. Society for Industrial and Applied Mathematics.

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

[40] 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.

[41] 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.

[42] K. Das, K. Bhaduri, K. Liu, and H. Kargupta. Distributed identification of top-l inner product elements and its application in a peer-to-peer network. *IEEE Trans. on Knowl. and Data Eng.*, 20(4):475–488, 2008.

[43] S. Datta, K. Bhaduri, C. Giannella, R. Wolff, and H. Kargupta. Distributed data mining in peer-to-peer networks. *IEEE Internet Computing*, 10(4):18–26, 2006.

[44] S. Datta, C. Giannella, and H. Kargupta. K-means clustering over peer-to-peer networks. In *Proceedings of the 8th International Workshop on High Performance and Distributed Mining (HPDM'05). In conjunction with the SIAM International Conference on Data Mining*, Newport Beach, California, April 2005.

[45] S. Datta and H. Kargupta. Uniform Data Sampling from a Peer-to-peer Network. In *Proceedings of the 27th International Conference on Distributed Computing Systems (ICDCS '07)*, page 50, Toronto, Canada, 2007.

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

[47] 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.

[48] J. Davits, J. Yu, S. Basu, D. Gutelis, and A. Harris. iLink: Search and Routing in Social Networks. In *Proceedings of the 13th International Conference on Knowledge Discovery and Data Mining (KDD '07)*, pages 931–940, New York NY, 2007.

[49] Paul Donachy, Terrence J harmer, Ron H Perrott, Jens Rasch, Sarah Bearder, and Martin Beckett. Grid Enabled Distributed Data Mining and Conversion of Unstructured Data. In *Proceedings of UK e-Science All Hands Meeting*, Nottingham, UK, September 2003.

[50] W. Du and G. Agrawal. Developing Distributed Data Mining Implementations for a Grid Environment. In *Workshop on Peer-to-Peer and Global Computing, with Conference on Cluster Computing and Grid (CCGRID)*, May 2002.

[51] 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.

[52] M. F. Duarte and Y. H. Hu. Vehicle Classification in Distributed Sensor Networks. *Journal of Parallel and Distributed Computing*, 64(7):826–838, 2004.

[53] M. Eisenhardt, W. Muller, and A. Henrich. Classifying Documents by Distributed P2P Clustering. In *Proceedings of Informatik 2003*, GI Lecture Notes in Informatics, Frankfurt, Germany, September 2003.

[54] 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.

[55] G. W. Flake, S. Lawrence, C. L. Giles, and F. M. Coetzee. Self organization and identification of web communities. *IEEE Computer*, 35(3):66–71, March 2002.

[56] Ariel Fuxman, Phokion G. Kolaitis, Rene J. Miller, and Wang-Chiew Tan. Peer data exchange. In *ACM SIGMOD/PODS 2005 Conference*, Baltimore, Maryland, June 2005.

[57] Prasanna Ganesan, Qixiang Sun, and Hector Garcia-Molina. Adlib: A self-tuning index for dynamic peer-to-peer systems. In *The 21st International Conference on Data Engineering (ICDE 2005)*, Tokyo, Japan, April 2005.

[58] H. Ghosh and S. Chaudhury. Distributed and Reactive Query Planning in R-MAGIC: An Agent-Based Multimedia Retrieval System. *IEEE Transactions on Knowledge and Data Engineering*, 16(9):1082–1095, September 2004.

[59] 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.

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

[61] C. Gkantsidis, M. Mihail, and A. Saberi. Random Walks in Peer-to-Peer Networks. In *Proceedings of IEEE INFOCOM*, 2004.

[62] 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.

[63] 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.

[64] M. Greenwald and S. Khanna. Power-Conserving Computation of Order-Statistics over Sensor Networks. In *Proceedings of the Twenty-third ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pages 275–285, Paris, France, June 2004.

[65] 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.

[66] Carlos Guestrin, Peter Bodk, Romain Thibaux, Mark A. Paskin, and Samuel Madden. Distributed regression: an efficient framework for modeling sensor network data. In *Proceedings of the Third International Symposium on Information Processing in Sensor Networks (IPSN 2004)*, Berkeley, California, April 2004.

[67] Ramakrishna Gummadi, Xin Li, Ramesh Govindan, Cyrus Shahabi, and Wei Hong. Energy-efficient data organization and query processing in sensor networks. In *The 21st International Conference on Data Engineering (ICDE 2005)*, Tokyo, Japan, April 2005.

[68] 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.

[69] P. Han, R. Shen, and F. Yang. A Novel Distributed Collaborative Filtering Algorithm and Its Implementation on P2P Overlay Network. In *Proceedings of the Eighth Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Sydney, Australia, May 2004.

[70] 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.

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

[72] K. Horowitz and D. Malkhi. Estimating network size from local information. *The Information Processing Letters Journal*, 88(5):237–243, December 2003.

[73] H.-C. Hsiao and C.-T. King. Similarity discovery in structured p2p overlays. In *2003 International Conference on Parallel Processing (ICPP'03)*, page 636, 2003.

[74] 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.

[75] 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.

[76] Mark Jelasity, Wojtek Kowalczyk, and Maarten van Stteen. An approach to massively distributed aggregate computing on peer-to-peer networks. In *Proceedings of the 12th Euromicro Conference on Parallel, Distributed and Network-Based Processing (PDP'04)*, A Coruna, Spain, February 2004.

[77] Wu-Shan Jiang and Ji-Hui Yu. Distributed data mining on the grid. In *Proceedings of 2005 International Conference on Machine Learning and Cybernetics*, Guangzhou, August 2005.

[78] 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.

[79] 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.

[80] 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.

[81] 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.

[82] S. Kamvar, M. Schlosser, and H. Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th international conference on World Wide Web (WWW)*, pages 640–651, Budapest, Hungary, 2003.

[83] 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.

[84] 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.

[85] 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.

[86] 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.

[87] 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.

[88] David Kempe, Alin Dobra, and Johannes Gehrke. Gossip-based computation of aggregate information. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 482–491, October 2003.

[89] M. Khambatti, K. D. Ryu, and P. Dasgupta. Efficient discovery of implicitly formed peer-to-peer communities. *International Journal of Parallel and Distributed Systems and Networks*, 5(4):155–164, 2002.

[90] M.S. Khambatti, K.D. Ryu, and P. Dasgupta. Push-pull gossiping for information sharing in peer-to-peer communities. In *Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA)*, pages pp. 1393–1399., Las Vegas, Nevada, June 2003.

[91] M. Klusch, S. Lodi, and G. L. Moro. Agent-Based Distributed Data Mining: The KDEC Scheme. In *Intelligent Information Agents: The AgentLink Perspective*, LNAI 2586, pages 104–122. Springer, July 2003.

[92] J. H. Kotecha, V. Ramachandran, and A. M. Sayeed. Distributed Multitarget Classification in Wireless Sensor Networks. *IEEE Journal of Selected Areas in Communications*, 23(4):703–713, April 2005.

[93] 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.

[94] 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.

[95] 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.

[96] 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.

[97] 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.

[98] 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.

[99] 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.

[100] Denis Krivitski, Assaf Schuster, and Ran Wolff. A local facility location algorithm for sensor networks. In *Proceedings of International Conference on Distributed Computing in Sensor Systems (DCOSS'05)*, Marina del Rey, CA, June-July 2005.

[101] A. Kumar, M. Kantardzic, P. Ramaswamy, and P. Sadeghian. An extensible service oriented distributed data mining framework. In *Proceedings of International Conference on Machine Learning and Applications, 2004*, December 2004.

[102] 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.

[103] 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.

[104] J. Kurhinen and J. Vuori. Information diffusion in a single-hop mobile peer-to-peer. In *Proceedings of the 10th IEEE Symposium on Computers and Communication(ISCC 2005)*, Cartagena, Spain, 2005.

[105] 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.

[106] Wang-Chien Lee and Baihua Zheng. Dsi: A fully distributed spatial index for wireless data broadcast. In *The 21st International Conference on Data Engineering (ICDE 2005)*, Tokyo, Japan, April 2005.

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

[108] 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.

[109] D. Liu. Resilient Cluster Formation for Sensor Networks. In *Proceedings of the 27th International Conference on Distributed Computing Systems (ICDCS '07)*, page 40, Toronto, Canada, 2007.

[110] K. Liu, K Bhaduri, K. Das, P. Nguyen, and H. Kargupta. Client-side web mining for community formation in peer-to-peer environments. In *SIGKDD workshop on web usage and analysis (WebKDD)*, 2006.

[111] Mihai Lupu and Bei Yu. Hiwarpp - hierarchical wavelet-based retrieval on peer-to-peer network. In *Proceedings of the twenty-second International Conference on Data Engineering (ICDE 2006)*, Atlanta, Georgia, April 2006.

[112] Amit Manjhi, Suman Nath, and Phillip B. Gibbons. Tributaries and deltas: Efficient and robust aggregation in sensor network streams. In *ACM SIGMOD/PODS 2005 Conference*, Baltimore, Maryland, June 2005.

[113] 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.

[114] 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.

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

[116] 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.

[117] M. Mehyar, D. Spanos, J. Pongsajapan, S. Low, and R. Murray. Distributed averaging on a peer-to-peer network. In *Proceedings of IEEE Conference on Decision and Control*, 2005.

[118] A. Metwally, D. Agrawal, and A. E. Abbadi. Efficient computation of frequent and top-k elements in data streams. In *Proceedings of the tenth International Conference on Database Theory (ICDT'05)*, Edinburgh, Scotland, January 2005.

[119] W. Nejdl, W. Siberski, U. Thaden, and W.-T. Balke. *Top-k Query Evaluation for Schema-Based Peer-to-Peer Networks*, 2004.

[120] 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.

[121] 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.

[122] Michael Rabbat and Robert Nowak. Distributed optimization in sensor networks. In *Proceedings of the third international symposium on Information processing in sensor networks (IPSN '04)*, pages 20–27, Berkeley, California, 2004.

[123] 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.

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

[125] 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.

[126] 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.

[127] I. Sharfman, Assaf Schuster, and Daniel Keren. A geometric appraoch to monitoring threshold functions over distributed data streams. In *Proceedings of the SIGMOD 2006*, Chicago, Illinois, June 2006.

[128] 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.

[129] Heng Tao Shen, Yanfeng Shu, and Bei Yu. Efficient Semantic-Based Content Search in P2P Network. *IEEE Transactions on Knowledge and Data Engineering*, 16:813–826, July 2004.

[130] S. Shi, J. Yu, G. Yang, and D. Wang. Distributed page ranking in structured p2p networks. In *Proceedings of 2003 International Conference on Parallel Processing (ICPP'03)*, page 179, 2003.

[131] 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.

[132] A. Silberstein, A. Gelfand, K. Munagala, G. Puggioni, and J. Yang. Making Sense of Suppressions and Failures in Sensor Data: A Bayesian Approach. In *Proceedings of the 33rd Very Large Data Bases Conference (VLDB '07)*, pages 842–853, New York, NY, 2007.

[133] A. Silberstein and J. Yang. Many-to-Many Aggregation for Sensor Networks. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE '07)*, pages 986–995, Istanbul, Turkey, 2007.

[134] Utkarsh Srivastava, Kamesh Munagala, and Jennifer Widom. Operator placement for in-network stream query processing. In *ACM SIGMOD/PODS 2005 Conference*, Baltimore, Maryland, June 2005.

[135] 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.

[136] 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.

[137] 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.

[138] 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.

[139] E. Tanin, A. Harwood, and H. Samet. A Distributed Quadtree Index for Peer-to-Peer Settings. In *Proceedings of the 21st International Conference on Data Engineering (ICDE'05)*, Tokyo, Japan, April 2005.

[140] C. Tempich, A. Löser, and J. Heizmann. *Community Based Ranking in Peer-to-Peer Networks*, volume 3761, chapter Lecture Notes in Computer Science, pages 1261–1278. Springer-Verlag GmbH, October 2005.

[141] 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.

[142] Y. Wang and J. Vassileva. Trust-based community formation in peer-to-peer file sharing networks. In *Proceedings IEEE International Conference on Web Intelligence (WI'04)*, pages 341–338, Beijing, China, October 2004.

[143] 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.

[144] 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.

[145] 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.

[146] R. Wolff, K. Bhaduri, and H. Kargupta. Local L2 Thresholding Based Data Mining in Peer-to-Peer Systems. In *Proceedings of 2006 SIAM Conference on Data Mining*, Bethesda, MD, April 2006.

[147] Qishi Wu, Nageswara S. V. Rao, Jacob Barhen, S. Sitharama Iyengar, Vijay K. Vaishnavi, Hairong Qi, and Krishnendu Chakrabarty. On Computing Mobile Agent Routes for Data Fusion in Distributed Sensor Networks. *IEEE Transactions on Knowledge and Data Engineering*, 16:740–753, June 2004.

[148] Lin Xiao, Stephen Boyd, and Sanjay Lall. A space-time diffusion scheme for peer-to-peer least-squares estimation. In *Proceedings of the Fifth International Symposium on Information Processing in Sensor Networks (IPSN 2006)*, Nashville, TN, April 2006.

[149] L. Xiong, S. Chitti, and L. Liu. Top-k Queries across Multiple Private Databases. In *Proceedings of the 25th IEEE International Conference on Distributed Computing Systems (ICDCS'05)*, pages 145–154. IEEE Computer Society, 2005.

[150] 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.

[151] X. Yang, H. Lim, T. Ozsu, and K. Tan. In-Network Execution of Monitoring Queries in Sensor Networks. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data and Principles of Database Systems*, pages 521–532, New York, NY, 2007.

[152] Liangzhong Yin and Guohong Cao. Dup: Dynamic-tree based update propagation in peer-to-peer networks. In *The 21st International Conference on Data Engineering (ICDE 2005)*, Tokyo, Japan, April 2005.

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

[154] D. Zeinalipour-Yazti, Z. Vagena, D. Gunopulos, V. Kalogeraki, V. Tsotras, M. Vlachos, N. Koudas, and D. Srivastava. The Threshold Join Algorithm for Top-k Queries in Distributed Sensor Networks. In *Proceedings of the 2nd International Workshop on Data Management for Sensor Networks*, pages 61–66, Trondheim, Norway, 2005. ACM Press.

[155] 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.

[156] J. Zhao, R. Govindan, and D. Estrin. Computing aggregates for monitoring wireless sensor networks. In *Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications*, pages 139–148, May 2003.

[157] M. Zhong, K. Shen, and J. Seiferas. Non-uniform random membership management in peer-to-peer networks. In *In Proceedings of the IEEE INFOCOM*, Miami, FL, March 2005.

[158] A. Zhou, W. Qian, S. Zhou, B. Ling, L. Xu, N. W. Siong, O. B. Chin, and T. Kian-Lee. Data management in peer-to-peer environment: a perspective of BestPeer. *Journal of Computer Science and Technology*, 18(4):452–461, 2003.

[159] D. Zhou, S. Orshanskiy, H. Zha, and C. Giles. Co-Ranking Authors and Documents in a Heterogeneous Network. In *Proceedings of the IEEE International Conference on Data Mining (ICDM '07)*, pages 739–744, Omaha, NE, 2007.

[160] Zhuang, Y. Chen, L. Wang, X. S. Lian, and Jie. A Weighted Moving Average-based Approach for Cleaning Sensor Data. In *Proceedings of the 27th International Conference on Distributed Computing Systems (ICDCS '07)*, page 38, Toronto, canada, 2007.

Privacy Preserving Distributed Data Mining Bibliography

References

[1] J. Al-Muhtadi, R. Campbell, A. Kapadia, M. D. Mickunas, and S. Yi. Routing through the mist: Privacy preserving communication in ubiquitous computing environments. In *Proceedings of ICDCS'02*, pages 74–83, Vienna, Austria, June July.

[2] 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.

[3] Roberto J. Bayardo and Rakesh Agrawal. Data privacy through optimal k-anonymization. In *The 21st International Conference on Data Engineering (ICDE 2005)*, Tokyo, Japan, April 2005.

[4] Elisa Bertino, Beng Chin Ooi, Yanjiang Yang, and Robert H. Deng. Privacy and ownership preserving of outsourced medical data. In *The 21st International Conference on Data Engineering (ICDE 2005)*, Tokyo, Japan, April 2005.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] Cynthia Dwork and Kobbi Nissim. Privacy-preserving Data mining on Vertically Partitioned Databases. In *Proceedings of The 24rd Annual International Cryptology Conference (CRYPTO 2004)*, Santa Barbara, CA, August 2004.

[10] Benjamin C. M. Fung, Ke Wang, and Philip S. Yu. Top-down specialization for information and privacy preservation. In *The 21st International Conference on Data Engineering (ICDE 2005)*, Tokyo, Japan, April 2005.

[11] B. Gilburd, A. Schuster, and R. Wolff. Privacy-Preserving Data Mining on Data Grids in the Presence of Malicious Participants. In *Proceedings of the 13th International Symposium on High-Performance Distributed Computing*, pages 225–234, Honolulu, Hawaii, June 2004.

[12] Bobi Gilburd, Assaf Schuster, and Ran Wolff. k-TTP: a New Privacy Model for Large-scale Distributed Environments. In *10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 563–568, Seattle, WA, August 2004.

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

[14] P. Kamat, W. Xu, W. Trappe, and Y. Zhang. Temporal Privacy in Wireless Sensor Networks. In *Proceedings of the 27th International Conference on Distributed Computing Systems (ICDCS '07)*, page 23, Toronto, Canada, 2007.

[15] 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.

[16] 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.

[17] 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.

[18] Murat Kantarcioglu and Chris Clifton. Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data. *IEEE Transactions on Knowledge and Data Engineering*, 16:1026–1037, September 2004.

[19] H. Kargupta, K. Das, and K. Liu. Multi-Party, Privacy-Preserving Distributed Data Mining using a Game Theoretic Framework. In *Proceedings of the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases(PKDD)*, pages 523–531, Warsaw, Poland, 2007.

[20] 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.

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

[22] 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.

[23] Kun Liu, Hillol Kargupta, and Jessica Ryan. Random Projection-Based Multiplicative Data Perturbation for Privacy Preserving Distributed Data Mining. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 18(1):92–106, January 2006.

[24] Da Meng, Krishnamoorthy Sivakumar, and Hillol Kargupta. Privacy Sensitive Bayesian Network Parameter Learning. In *Proceedings of The Fourth IEEE International Conference on Data Mining (ICDM'04)*, Brighton, UK, November 2004.

[25] 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.

[26] 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.

[27] Ashish Sanil, Alan Karr, Xiaodong Lin, and Jerome Reiter. Privacy Preserving Regression Modelling via Distributed Computation. In *10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, WA, August 2004.

[28] Assaf Schuster, Ran Wolff, and Bobi Gilburd. Privacy-Preserving Association Rule Mining in Large-Scale Distributed Systems . In *4th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGrid04)*, Illinois , USA, April 2004.

[29] 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.

[30] 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.

[31] 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.

[32] Rebecca Wright and Zhiqiang Yang. Privacy-Preserving Bayesian Network Structure Computation on Distributed Heterogeneous Data. In *Proceedings of The Tenth ACM SIGKDD Conference (KDD'04)*, Seattle, WA, August 2004.

[33] Sheng Zhong, Zhiqiang Yang, and Rebecca N. Wright. Privacy-enhancing k-anonymization of customer data. In *ACM SIGMOD/PODS 2005 Conference*, Baltimore, Maryland, June 2005.

Distributed Preprocessing Bibliography

References

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

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

[3] 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.

[4] Panagis Magdalinos, Christos Doulkeridis, and Michalis Vazirgiannis. K-landmarks: Distributed dimensionality reduction for clustering quality maintenance. In *Proceedings of PKDD*, 2006.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] Alan Wagner, James Chilson, Raymond Ng, and Ruben Zamar. Parallel Computation of High Dimensional Robust Correlation and Covariance Matrices. In *10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, WA, August 2004.

[10] 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.

Surveys

References

[1] 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.

[2] A.A. Freitas and S.H. Lavington. *Mining Very Large Databases with Parallel Processing.* Kluwer Academic Publishers, 1998.

[3] Yike Guo and Robert Grossman (editors). *High Performance Data Mining: Scaling Algorithms, Applications and Systems.* Kluwer Academic Publishers, 1999.

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

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

[6] 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.

[7] 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.

[8] 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.

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

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

[11] F. Perich, A. Joshi, T. Finin, and Y. Yesha. On Data Management in Pervasive Computing Environments. *IEEE Transactions on Knowledge and Data Engineering*, 16(5):621–634, May 2004.

[12] 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.

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

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

[15] Mohammed J. Zaki and Ching-Tien Ho (editors). *Large-scale Parallel Data Mining*, volume 1759 of *Lecture Notes in Artificial Intelligence, State-of-the-Art-Survey*. Springer-Verlag, 2000.

Miscellaneous Topics

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] O. Benjelloun, H. Garcia-Molina, H. Gong, H. Kawai, T. Larson, D. Menestrina, and S. Thavisomboon. D-Swoosh: A Family of Algorithms for Generic, Distributed Entity Resolution. In *Proceedings of the 27th International Conference on Distributed Computing Systems (ICDCS '07)*, page 37, Toronto, Canada, 2007.

[6] K. Bhaduri and H. Kargupta. An efficient local algorithm for distributed multivariate regression in peer-to-peer networks. In *SDM*, pages 153–164, 2008.

[7] 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.

[8] P. Cao and Z. Wang. Efficient top-K Query calculation in Distributed Networks. In *Proceedings of the Twenty-third Annual ACM Symposium on Principles of Distributed Computing*, pages 206–215, New York, NY, 2004. ACM Press.

[9] 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.

[10] 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.

[11] 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.

[12] G. Cong, W. Fan, and A. Kementsietsidis. Distributed Query Evaluation with Performance Guarantees. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, pages 509–520, New York, NY, 2007.

[13] Graham Cormode, Minos Garofalakis, S. Muthukrishnan, and Rajeev Rastogi. Holistic aggregates in a networked world: Distributed tracking of approximate quantiles. In *ACM SIGMOD/PODS 2005 Conference*, Baltimore, Maryland, June 2005.

[14] Alin Deutsch, Yannis Katsis, and Yannis Papakonstantinou. Determining source contribution in integration systems. In *ACM SIGMOD/PODS 2005 Conference*, Baltimore, Maryland, June 2005.

[15] Christos Doulkeridis, Kjetil Nrvg, and Michalis Vazirgiannis. Scalable semantic overlay generation for p2p-based digital libraries. In *Proceedings of ECDL*, 2006.

[16] Christos Doulkeridis, Kjetil Nrvg, and Michalis Vazirgiannis. Desent: decentralized and distributed semantic overlay generation in p2p networks. *IEEE Journal on Selected Areas in Communications 25(1): 25-34 (2007)*, 2007.

[17] C. du Mouza, W. Litwin, and P. Rigaux. SD-Rtree: A Scalable Distributed Rtree. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE '07)*, pages 296–305, Istanbul, Turkey, 2007.

[18] 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.

[19] H. Dutta, C. Giannella, K. Borne, and H. Kargupta. Distributed Top-k Outlier Detection from the Astronomy Catalogs using the DEMAC System. In *Proceedings of the 2007 SIAM International Conference on Data Mining (SDM '07)*, Philadelphia, PA, 2007.

[20] 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.

[21] Ying-Wu Fang, Xiu-Bing Zhao, Guang-Peng Zhang, Yi Wang, Yi Sun, and Yong-Fang Zhang. Study on algorithms of parallel and distributed data mining calculating process. In *Proceedings of 2005 International Conference on Machine Learning and Cybernetics*, Guangzhou, August 2005.

[22] Nuno Fonseca, Fernando Silva, and Rui Camacho. Strategies to parallelize ILP systems. In *Proceedings of ILC 2005*, Bonn, Germany, August 2005.

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

[24] 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.

[25] Vladimir Gorodetsky, Oleg Karsaeyv, and Vladimir Samoilov. Software tool for agent-based distributed data mining. In *International Conference on Integration of Knowledge Intensive Multi-Agent Systems (KIMAS)*, Boston, MA, October 2003.

[26] D. Gu. Distributed em algorithm for gaussian mixtures in sensor networks. *IEEE Transactions on Neural Network*, 19(7):1154–1166, July 2008.

[27] 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.

[28] 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.

[29] Y. Hu, H. Chen, J. Lou, and J. Li. Distributed Density Estimation Using Nonparametric Statistics. In *Proceedings of the 27th International Conference on Distributed Computing Systems (ICDCS '07)*, page 28, Toronto, Canada, 2007.

[30] R. Huebsch, M. Garofalakis, J. Hellerstein, and I. Stoica. Sharing Aggregate Computation for Distributed Queries. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, pages 509–520, New York, NY, 2007.

[31] 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.

[32] K. Kim and S. Choi. Neighbor Search with Global Geometry: A Minimax Message Passing Algorithm. In *Proceedings of the 24th International Conference on Machine Learning (ICML '07)*, pages 401–408, Madison, WI, 2007.

[33] 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.

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

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

[36] Shonali Krishnaswamy, Seng W. Loke, and Arkady Zaslasvky. A hybrid model for improving response time in distributed data mining. *IEEE Transactions on Systems, Man and CyberneticsPart B: Cybernetics*, 34(6):2466–2479, December 2006.

[37] 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.

[38] S. Kutten and D. Peleg. Fault-local distributed mending. In *Proc. of the ACM Symposium on Principle of Distributed Computing (PODC)*, pages 20–27, Ottawa, Canada, August 1995.

[39] 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.

[40] 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.

[41] 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.

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

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

[44] 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.

[45] N. Linial. Locality in distributed graph algorithms. *SIAM Journal of Computing*, 21:193–201, 1992.

[46] B. Liu and E. Rundensteiner. Cost-Driven General Join View Maintenance over Distributed Data Sources. In *Proceedings of the 21st International Conference on Data Engineering (ICDE'05)*, Tokyo, Japan, April 2005.

[47] 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.

[48] Y. Lu, V. Roychowdhury, and L. Vandenberghe. Distributed parallel support vector machines in strongly connected networks. *IEEE Transactions on Neural Network*, 19(7):1167–1178, July 2008.

[49] Michael I. Jordan M. Wainwright. *A Variational Principle for Graphical Models*. In New Directions in Statistical Signal Processing (chapter 11). MIT Press, 2005.

[50] 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.

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

[52] S. Michel, P. Triantafillou, and G. Weikum. KLEE: a Framework for Distributed Top-k Query Algorithms. In *Proceedings of the 31st International Conference on Very Large Databases (VLDB'05)*, pages 637–648, Trondheim, Norway, 2005.

[53] 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.

[54] Moni Naor and Larry Stockmeyer. What can be computed locally? *SIAM Journal on Computing*, 24(6):1259–1277, December 1995.

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

[56] X. Nguyen, M. J. Wainwright, and Michael I. Jordan. Nonparametric Decentralized Detection Using Kernel Methods. *IEEE Transactions on Signal Processing (To Appear)*, 2005.

[57] 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.

[58] 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.

[59] J. Ocenasek, J. Schwarz, and M. Pelikan. Design of multithreaded estimation of distribution algorithms. In *Proceedings of Genetic and Evolutionary Computation Conference - GECCO 2003*, Berlin, Germany, August 2003.

[60] 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.

[61] B. Paechter, T. Back, M. Schoenauer, M. Sebag, A. E. Eiben, J. J. Merelo, and T. C. Fogarty. A distributed resource evolutionary algorithm machine (dream). In *Proceedings of the 2000 Congress on Evolutionary Computation*, July 2000.

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

[63] 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.

[64] 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.

[65] 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.

[66] 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.

[67] Theoni Pitoura and Peter Triantafillou. Self-join size estimation in large-scale distributed data systems. In *ICDE*, pages 764–773, 2008.

[68] 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.

[69] M. Saerens, F. Fouss, L. Yen, and P. Dupont. The principal component analysis of a graph and its relationships to spectral clustering. In *Proceedings of the 15th European Conference on Machine Learning (ECML)*, Pisa, Italy, 2004.

[70] 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.

[71] Martin Scholz. On the complexity of rule discovery from distributed data. In *Proceedings of the Fifth IEEE International Conference on Data Mining*, Houston, Texas, August 2005.

[72] H. Schweitzer. A Distributed Algorithm for Content Based Indexing of Images by Projections on Ritz Primary Images . *Data Mining and Knowledge Discovery*, 1(4):375–390, December 1997.

[73] Jimeng Sun, Huiming Qu, Deepayan Chakrabari, and Christos Faloutsos. Neighborhood formation and anomaly detection in bipartite graphs. In *Proceedings of the Fifth IEEE International Conference on Data Mining*, Houston, Texas, August 2005.

[74] Raz Tamir. A random walk through human associations. In *Proceedings of the Fifth IEEE International Conference on Data Mining*, Houston, Texas, August 2005.

[75] Pang-Ning Tan and Rong Jin. Ordering Patterns by Combining Opinions from Multiple Sources. In *10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, WA, August 2004.

[76] F. Tanudjaja and L. Mui. Persona: A contextualized and personalized web search. In *In Proceedings of the 35 Annual Hawaii International Conference on System Sciences (HICSS'02)*, volume 3, page 53, Hawaii, 2002.

[77] Michalis Vazirgiannis, Kjetil Nrvg, and Christos Doulkeridis. Peer-to-peer clustering for semantic overlay network generation. In *Proceedings of PRIS*, 2006.

[78] Y. Wang and D. DeWitt. Computing PageRank in a Distributed Internet Search Engine System. In *Proceedings of the 30th International Conference on Very Large Data Bases (VLDB 2004)*, Toronto, Canada, August 2004.

[79] 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.

[80] Syed Zahid Hassan Zaidi, Syed Sibte Raza Abidi, and Selvakumar Manickam. Distributed data mining from heterogeneous healthcare data repositories: towards an intelligent agent-based framework. In *Proceedings of the 15th IEEE Symposium on Computer-Based Medical Systems (CBMS 2002)*, 2002.

[81] Xiaofeng Zhang and William K. Cheung. Visualizing global manifold based on distributed local data abstractions. In *Proceedings of the Fifth IEEE International Conference on Data Mining*, Houston, Texas, August 2005.

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

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

[84] 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.

[85] W. Zhu, P. Bridges, and A. Maccabe. Embedded Gossip: Lightweight Online Measurement for Large-Scale Applications. In *Proceedings of the 27th International Conference on Distributed Computing Systems (ICDCS '07)*, page 58, Toronto, Canada, 2007.