

References

Book References:

- [1] Patrick Winston. [Artificial Intelligence](#). Addison Wesley, 1992.
- [2] Michael Anderberg, [Clustering for Applications](#). Academic Press, 1973.
- [3] Kaufman and Rousseeuw, [Finding Groups in Data: An Introduction to Cluster Analysis](#), Wiley, 1990.
- [4] T. Kohonen, [Self-Organizing Maps.](#), Second Extended Edition, Springer Series in Information Sciences, Vol. 30, Springer, Berlin, Heidelberg, New York, 1997.
- [5] Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy Uthurasamy (eds.), [Advances in Knowledge Discovery and Data Mining](#), AAAI Press/ The MIT Press, 1996.
- [6] Michael Berry and Gordon Linoff, [Data Mining Techniques \(For Marketing, Sales, and Customer Support\)](#), John Wiley & Sons, 1997.
- [7] Sholom M. Weiss and Nitin Indurkha, [Predictive Data Mining \(a practical guide\)](#), Morgan Kaufmann Publishers, 1998.
- [8] Alex Freitas and Simon Lavington, [Mining Very Large Databases with Parallel Processing](#), Kluwer Academic Publishers, 1998.
- [9] A. K. Jain and R. C. Dubes, [Algorithms for Clustering Data](#), Prentice Hall, 1988.
- [10] Vipin Kumar, Ananth Grama, Anshul Gupta, and George Karypis, [Introduction to Parallel Computing: Algorithm Design and Analysis](#), Benjamin Cummings/Addison Wesley, Redwood City, 1994.

References

Book References:

[11] J. Ross Quinlan, [*C4.5: Programs for Machine Learning*](#), Morgan Kaufmann, 1993.

[12] Robert L. Grossman (Editor), Chandrika Kamath (Editor), Philip Kegelmeyer (Editor), Vipin Kumar (Editor), [*Data Mining for Scientific and Engineering Applications*](#), Kluwer Academic Publishers / November 2001.

References...

Research Paper References:

- [1] M. Mehta, R. Agarwal, and J. Rissanen, [*SLIQ: A Fast Scalable Classifier for Data Mining*](#), Proc. Of the fifth Int. Conf. On Extending Database Technology (EDBT), Avignon, France, 1996.
- [2] J. Shafer, R. Agrawal, and M. Mehta, [*SPRINT: A Scalable Parallel Classifier for Data Mining*](#), Proc. 22nd Int. Conf. On Very Large Databases, Mumbai, India, 1996.
- [3] A. Srivastava, E.H. Han, V. Kumar, and V.Singh, [*Parallel Formulations of Decision-Tree Classification Algorithms*](#), Proc. 12th International Parallel Processing Symposium (IPPS), Orlando, 1998.
- [4] M.Joshi, G.Karypis, and V. Kumar, [*ScalParC: A New Scalable and Efficient Parallel Classification Algorithms for Mining Large Datasets*](#), Proc. 12th International Parallel Processing Symposium (IPPS), Orlando, 1998.
- [5] N. Friedman, D. Geiger, and M. Goldszmidt, [*Bayesian Network Classifiers*](#), Machine Learning 29:131--163, 1997.
- [6] R. Agrawal, T.Imielinski, and A.Swami, [*Mining Association Rules Between Sets of Items in Large Databases*](#), Proc. 1993 ACM-SIGMOD Int.Conf. On Management of Data, Washington, D.C., 1993.
- [7] R. Agrawal and R. Srikant, [*Fast Algorithms for Mining Association Rules*](#), Proc. Of 20th VLDB Conference, 1994.

References...

- [8]** R. Agrawal and J.C. Shafer, [*Parallel Mining of Association Rules*](#), IEEE Trans. On Knowledge and Data Eng., 8(6):962-969, December 1996.
- [9]** E.H.Han, G.Karypis, and V.Kumar, [*Scalable Parallel Data Mining for Association Rules*](#), Proc. 1997 ACM-SIGMOD Int. Conf. On Management of Data, Tucson, Arizona, 1997.
- [10]** R. Srikant and R. Agrawal, [*Mining Sequential Patterns: Generalizations and Performance Improvements*](#), Proc. Of 5th Int. Conf. On Extending Database Technology (EDBT), Avignon, France, 1996.
- [11]** M. Joshi, G. Karypis, and V. Kumar, [*Parallel Algorithms for Sequential Associations: Issues and Challenges*](#), Mini-symposium Talk at Ninth SIAM International Conference on Parallel Processing (PP'99), San Antonio, 1999.
- [12]** George Karypis, Eui-Hong (Sam) Han, and Vipin Kumar, [*CHAMELEON: A Hierarchical Clustering Algorithm Using Dynamic Modeling \(1999\)*](#), IEEE Computer, Vol. 32, No. 8, August, 1999. pp. 68-75.
- [13]** George Karypis, Eui-Hong (Sam) Han, and Vipin Kumar, [*Multilevel Refinement for Hierarchical Clustering \(1999\)*](#), Technical Report # 99-020.
- [14]** Daniel Boley, Maria Gini, Robert Gross, Eui-Hong (Sam) Han, Kyle Hastings, George Karypis, Vipin Kumar, Bamshad Mobasher, and Jerome Moore, [*Partitioning-Based Clustering for Web Document Categorization \(1999\)*](#). To appear in Decision Support Systems Journal.

References...

- [15] Eui-Hong (Sam) Han, George Karypis, Vipin Kumar and B. Mobasher, [Hypergraph Based Clustering in High-Dimensional Data Sets: A Summary of Results \(1998\)](#). Bulletin of the Technical Committee on Data Engineering, Vol. 21, No. 1, 1998.
- [16] K. C. Gowda and G. Krishna, [Agglomerative Clustering Using the Concept of Mutual Nearest Neighborhood](#), Pattern Recognition, Vol. 10, pp. 105-112, 1978.
- [17] Rakesh Agrawal, Johannes Gehrke, Dimitrios Gunopulos, Prabhakar Raghavan, [Automatic Subspace Clustering of High Dimensional Data for Data Mining Applications](#), *Proc. of the ACM SIGMOD Int'l Conference on Management of Data*, Seattle, Washington, June 1998.
- [18] Peter Cheeseman and John Stutz, ["Bayesian Classification \(AutoClass\): Theory and Results"](#), in U. M. Fayyad, G. Piatetsky-Shapiro, P. Smith, and R. Uthurusamy (eds.), "Advances in Knowledge Discovery and Data Mining", pp. 153-180, AAAI/MIT Press, 1996.
- [19] Tian Zhang, Raghu Ramakrishnan, Miron Livny, ["BIRCH: An Efficient Data Clustering Method for Very Large Databases"](#), Proc. of ACM SIGMOD Int'l Conf. on Data Management, Canada, June 1996.

References...

- [20] Venkatesh Ganti, Raghu Ramakrishnan, and Johannes Gehrke, "[Clustering Large Datasets in Arbitrary Metric Spaces](#)", Proceedings of IEEE Conference on Data Engineering, Australia, 1999.
- [21] Jarvis and E. A. Patrick, "[Clustering Using a Similarity Measure Based on Shared Nearest Neighbors](#)", IEEE Transactions on Computers, Vol. C-22, No. 11, November, 1973.
- [22] Sudipto Guha, Rajeev Rastogi, Kyuseok Shim, "[CURE: An Efficient Clustering Algorithm for Large Databases](#)", ACM SIGMOD Conference, 1998, pp. 73-84.
- [23] Sander J., Ester M., Kriegel H.-P., Xu X., "[Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and its Applications](#)", Data Mining and Knowledge Discovery, An International Journal, Kluwer Academic Publishers, Vol. 2, No. 2, 1998, pp. 169-194.
- [24] R. Ng and J. Han., "[Efficient and effective clustering method for spatial data mining](#)", In *Proc. 1994 Int. Conf. Very Large Data Bases*, pp. 144--155, Santiago, Chile, September, 1994.
- [25] Chris Fraley and Adrian E. Raftery, "[How many clusters? Which clustering method? - Answers via Model-Based Cluster Analysis](#)", *Computer Journal*, 41(1998):578-588.

References...

[26] Sudipto Guha, Rajeev Rastogi, Kyuseok Shim, "[ROCK: A Robust Clustering Algorithm for Categorical Attributes](#)", Proceedings of IEEE Conference on Data Engineering, Australia, 1999.

[27] Paul S. Bradley, Usama M. Fayyad and Cory A. Reina, "[Scaling Clustering Algorithms to Large Databases](#)", Proceedings of the 4th International Conference on Knowledge Discovery & Data Mining (KDD98).

***Over 100 More Data Mining References are available at
<http://www.cs.umn.edu/~mjoshi/dmrefs.html>***

Our group's papers are available via <http://www.cs.umn.edu/~kumar>