

BIBLIOGRAPHY

- [1] A. Abbott and A. Tsay. Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect. In *Sociological Methods and Research* Vol 29 pp3-33, 2000.
- [2] R. Agrawal and R. Srikant. Mining sequential patterns. In *ICDE 95*, pages 3–14, Taipei, Taiwan, Mar. 1995.
- [3] M. S. Aldenderfer and R. K. Blashfield. Cluster analysis Beverly Hills : Sage Publications, 1984.
- [4] J. Ayres, J. Flannick, J. Gehrke, and T. Yiu. Sequential pattern mining using a bitmap representation. In *SIGKDD*, pages 429–435, July 2002.
- [5] R. Agrawal, C. Faloutsos, and A. Swami. Efficient similarity search in sequence databases. In *The fourth international conference on foundations of data organization and algorithms* 1993.
- [6] P. Arabie and L. J. Hubert. An overview of combinatorial data analysis. In *Clustering and classification* Arabie, P. Hubert, L. J., & De Soete, G. (Eds.). River Edge, NJ: World scientific Publ. 1996.
- [7] D. Barbara, Y. Li, and J. Couto COOLCAT: an entropy-based algorithm for categorical clustering. In *CIKM*, pages 582–589, 2002.
- [8] R. J. Bayardo. Efficiently mining long patterns from databases. In *SIGMOD*, pages 85–93, June 1998.
- [9] Z. Bi, C. Faloutsos, F. Korn. The “DGX” distribution for mining massive, skewed data. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM Press, 2001.
- [10] S. Brin, R. Motwani, and C. Silverstein. Beyond market baskets: generalizing association rules to correlations. In *Proc. of ACM International Conference On Management of Data (SIGMOD)*, pp265-276, 1997.
- [11] P. Berkhin. Survey Of Clustering Data Mining Techniques. *Technical Report - Accrue Software*, San Jose, CA, 2002.
- [12] D. Chudova and P. Smyth. Pattern discovery in sequences under a markov assumption. In *Proc. 2002 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD'02)*, Edmonton, Alberta, Canada, July 2002.
- [13] J. Coggins. Dissimilarity measures for clustering strings. In *Time warps, string edits, and macromolecules: the theory and practice of sequence comparison*. D. Snakoff, & J. Kruskal, (Eds.), pp 253-310. Addison-Wesley Pub. Co. MA. 1983.

- [14] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. Dynamic Programming. In *Introduction to Algorithms*. pp 301-328. Cambridge, Mass. : MIT Press. 1990.
- [15] R. O. Duda, P. E. Hart, and D. G. Stork. Unsupervised Learning and Clustering. In *Pattern Classification.*, New York: Wiley, pages 517–599, 2001.
- [16] R. C. Dubes and A. K. Jain. Clustering methodologies in exploratory data analysis. In *Advances in Computers*. Vol 19, pages113–228. 1980.
- [17] L. Ertoz, M. Steinbach, and V. Kumar. Finding Clusters of Dierent Sizes, Shapes, and Densities in Noisy, High Dimensional Data In *Third SIAM International Conference on Data Mining*, San Fransico. CA, 2003.
- [18] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From Data Mining to Knowledge Discovery: An Overview. In *Advances in Knowledge Discovery and Data Mining*, Fayyad et al (Eds), pp 1-34. Cambridge, MA: AAAI/MIT Press, 1996.
- [19] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. The KDD process for extracting useful knowledge from volumes of data. In *Communications of ACM* Vol 36 Issue 11, pp 27-34. New York, NY: ACM Press, Nov. 1996.
- [20] U. Fayyad and R. Uthurusamy. Data mining and knowledge discovery in databases. In *Communications of ACM* Vol 36 Issue 11, pp 24-26. New York, NY: ACM Press, Nov. 1996.
- [21] J. Foley, A. VanDam, S. Feiner, and J. Hughes. Computer graphics : principles and practice Reading, Mass. : Addison-Wesley, 1996.
- [22] K. K. Fukunaga and P.M. Narendra. A branch and bound algorithm for computing k-nearest neighbours. In *IEEE Transactions on Computers*, 24:750-753, 1975.
- [23] V. Ganti, J. Gehrke, and R. Ramakrishnan CACTUS - Clustering Categorical Data Using Summaries. In *KDD*, pages 73–83, 1999.
- [24] D. Gibson, J. M. Kleinberg, and P. Raghavan Clustering Categorical Data: An Approach Based on Dynamical Systems. In *VLDB*, pages 311–322, 1998.
- [25] Osamu Gotoh. Multiple sequence alignment: Algorithms and applications. In *Adv. Biophys.*, Vol. 36, pp159-206. 1999.
- [26] R. Goerge, D. Duncan, L. Usher, B. J. Lee, M. Reidy, B. Needell, and A. Brookhar. Dynamics of Children’s Movement among the AFDC, Medicaid, and Foster Care Programs. *Technical report to U.S. Dept. of Health and Human Services (HHS)*. 2000.
- [27] S. Guha, R. Rastogi, and K. Shim ROCK: A Robust Clustering Algorithm for Categorical Attributes. In *ICDE*, pages 512-521, 1999.
- [28] M. Garofalakis, R. Rastogi, and K. Shim. SPIRIT: Sequential pattern mining with regular expression constraints. In *Proc. 1999 Int. Conf. Very Large Data Bases (VLDB’99)*, pages 223–234, Edinburgh, UK, Sept. 1999.

- [29] V. Guralnik and G. Karypis. A scalable algorithm for clustering sequential data. In *Proc. 2001 Int. Conf. Data Mining (ICDM'01)*, pages 179–186, San Jose, CA, Nov. 2001.
- [30] D. Gusfield. *Algorithms on strings, trees, & sequences: Computer Science and Computational Biology*. Cambridge Univ. Press, Cambridge, England. 1997.
- [31] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. New York, NY: Morgan Kaufmann Publishers, 2001.
- [32] A. Hinneburg and D. Keim. Clustering Techniques for Large Data Sets: From the past to the future. In *Tutorial notes fro ACM SIGKDD 1999 International Conference on Knowledge discovery and data mining*, Han et al, (Eds), ACM. 1999.
- [33] D. Hand, H. Mannila, and P. Smyth. *Principles of Data Mining*. Cambridge, MA: MIT Press, 2001.
- [34] J. Han, J. Pei, et al. FreeSpan: Frequent pattern-projected sequential pattern mining. In *SIGKDD*, pages 355–359, Aug. 2000.
- [35] A. K. Jain, M. N. Murty, and P. J. Flynn. Data Clustering: A Review. In *ACM Computing Surveys*, Vol. 31, No. 3, pages 264–323, Sep 1999.
- [36] A. K. Jain and R. C. Dubes. *Algorithms for clustering data*. Englewood Cliffs, NJ: Prentice-Hall. 1988.
- [37] L. Kaufman and P. J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: Wiley. 1990.
- [38] W. L. G. Koontz, Patrenahalli M. Narendra, and K. Fukunaga. A Branch and Bound Clustering Algorithm. In *IEEE Transactions on Computers*. 24(9): 908-915 (1975).
- [39] H.C. Kum, J. Pei, W. Wang, and D. Duncan. ApproxMAP : Approximate Mining of Consensus Sequential Patterns. *Technical Report TR02-031*, UNC-CH, 2002.
- [40] J. Kruskal. An overview of sequence comparison. In *Time warps, string edits, and macromolecules: the theory and practice of sequence comparison*. D. Snakoff, & J. Kruskal, (Eds.), pp 253-310. Addison-Wesley Pub. Co. MA. 1983.
- [41] A. Large, L. Tedd, and R. Hartley. *Information seeking in the online age: principles and practice*. Bowker. 1999.
- [42] B. Liu, M. Hu, and W. Hsu. Multi-level organization and summarization of the discovered rules. In *Proc. 6th ACM Intl. Conf. On Knowledge Discovery and Data Mining (SIGKDD)*, pp208-217, 2000.
- [43] B. Liu, W. Hsu, and Y. Ma. Mining association rules with multiple minimum supports. In *Proc. 5th ACM Intl. Conf. On Knowledge Discovery and Data Mining (SIGKDD)*, pp337-341, 1999.
- [44] G. R. McPherson and S. DeStefano. *Applied Ecology and Natural Resource Management*. Cambridge University Press, Cambridge, England. 2002.

- [45] H. Mannila, H. Toivonen, and A. I. Verkamo. Discovery of frequent episodes in event sequences. In *Data Mining and Knowledge Discovery: An International Journal*, Vol 1(3):pp256-289, 1997.
- [46] C. E. Metz. Basic principles of ROC analysis. In *Seminars in Nuclear Medicine*, vol. 8, pp283-298, 1978.
- [47] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal. Discovering frequent closed itemsets for association rules. In *ICDT*, pages 398–416, Jan. 1999.
- [48] J. Pei, J. Han, et al. PrefixSpan: Mining sequential patterns efficiently by prefix-projected pattern growth. In *ICDE*, pages 215–224, April 2001.
- [49] J. Pei, J. Han, and W. Wang. Constraint-based sequential pattern mining in large databases. In *submitted to publication*, March 2002.
- [50] J. Pei, A. K. H. Tung, and J. Han. Fault-tolerant frequent pattern mining: Problems and challenges. In *Proc. 2001 ACM-SIGMOD Int. Workshop Data Mining and Knowledge Discovery (DMKD'01)*, pages 7–12, Santa Barbara, CA, May 2001.
- [51] J. Sander, M. Ester, H. P. Kriegel, and X. Xu. Density based clustering in spatial databases: The algorithm gdbscan and its applications. In *Data Mining and Knowledge Discovery*, 2(2):169.194, 1998.
- [52] D. Sankoff, R. Cedergren. Simultaneous comparison of three or more sequences related by a tree. In *Time warps, string edits, and macromolecules: the theory and practice of sequence comparison*. D. Snakoff, & J. Kruskal, (Eds.), pp 253-310. Addison-Wesley Pub. Co. MA. 1983.
- [53] Sas Institute. Proc Modeclust. In SAS/STAT User Guide. Sas online Document. 2000
- [54] Myra Spiliopoulou. Managing interesting rules in sequence mining. In *Proc. European Conf. on Principles and Practice of Knowledge Discovery in Databases*, pp554-560, 1999.
- [55] R. Srikant and R. Agrawal. Mining sequential patterns: Generalizations and performance improvements. In *EDBT*, pages 3–17, Mar. 1996.
- [56] P. Stolorz and R. Musick (Eds). Guest Editorial. In *Data mining and knowledge discovery* Vol 1 No. 4. Special issue: Scalable High Performance Computing for Knowledge Discovery and Data Mining. Boston, MA: Kluwer Academic Publishers, 1998.
- [57] J. Thompson, F. Plewniak, and O. Poch. A comprehensive comparison of multiple sequence alignment programs. In *Nucleic Acids Research*. Vol. 27. No. 13: pp 2682-2690. Oxford University Press. 1999.
- [58] C. L. Usher, K. A. Randolph, and H. C. Gogan. Placement patterns in foster care. In *Social Service Review*, vol 73, pp22-36. 1999.
- [59] M. Vingron, M. Waterman. Sequence alignment and penalty choice. In *Journal of molecular biology*, Vol 235: pp1-12.

- [60] K. Wang, Y. He, and J. Han. Mining frequent itemsets using support constraints. In Proc. 26th Intl. Conf. on Very Large Data Bases (VLDB), 2000.
- [61] M. A. Wong and T. Lane. A kth Nearest Neighbor Clustering Procedure. In *Journal of the Royal Statistical Society, Series B*, 45, pp362-368, 1983.
- [62] C. Yang, U. Fayyad, and P.S. Bradley. Efficient discovery of error-tolerant frequent itemsets in high dimensions. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 194–203. ACM Press, 2001.
- [63] X. Yan, J. Han, and R. Afshar. CloSpan: Mining Closed Sequential Patterns in Large Datasets. In *Third SIAM International Conference on Data Mining*, San Francisco, CA, 2003.
- [64] J. Yang, P. S. Yu, W. Wang, and J. Han. Mining long sequential patterns in a noisy environment. In *Proc. 2002 ACM-SIGMOD Int. Conf. on Management of Data (SIGMOD'02)*, Madison, WI, June 2002.
- [65] M.J. Zaki. Sequence mining in categorical domains: incorporating constraints. In *Proceedings of the ninth international conference on Information and knowledge management*, pages 422–429. ACM Press, 2000.
- [66] M. J. Zaki. Efficient enumeration of frequent sequences. In *7th International Conference Information and Knowledge Management*, Nov 1998.
- [67] M. J. Zaki, S. Parthasarathy, et al. Parallel algorithm for discovery of association rules. *Data Mining and Knowledge Discovery*, 1:343–374, 1997.
- [68] Y. Zhang, A. W. Fu, C. H. Cai, and P. Heng. Clustering Categorical Data. In *ICDE*, page 305, 2000.