

Publications of the From Data to Knowledge Research Unit

Publications are in alphabetical order by the first author.

2002

1. Bingham, E., Mannila, H. & Seppänen, J. K. Topics in 0-1 Data. *Proc. of 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Edmonton, Alberta, Canada, July 23-26, 2002. pp. 450-455.
2. Bykowski, A., Seppänen, J. & Hollmén, J. Model-independent Bounding of the Supports of Boolean Formulae in Binary Data. *Proc. of First International Workshop Knowledge Discovery in Inductive Databases (KDID'02)*, Helsinki, Finland, August 2002. pp. 20-31.
3. de Raedt, L., Jaeger, M., Lee, S. D. & Mannila, H. A Theory of Inductive Query Answering. *Proc. of 2002 IEEE International Conference on Data Mining (ICDE 2002)*, Maebashi City, Japan, December 9-12, 2002. pp. 123-130.
4. Elomaa, T., Mannila, H. & Toivonen, H. (eds.) *Proc. of ECML 2002 - 12th European Conference on Machine Learning, LNCS 2430*. Springer, 2002.
5. Elomaa, T., Mannila, H. & Toivonen, H. (eds.) *Proc. of Principles of Data Mining and Knowledge Discovery - 6th European Conference, PKDD 2002, LNCS 2431*. Springer, 2002.
6. Grossman, R., Han, J., Kumar, V., Mannila, H. & Motwani, R. (eds.) *Proceedings of the Second SIAM International Conference on Data Mining, SIAM 2002*; ISBN 0-89871-517-2. 2002.
7. Han, J., Altman, R. B., Kumar, V., Mannila, H. & Pregibon, D. Emerging Scientific Applications in Data Mining. *Communications of the ACM*, 2002. Vol. 45, No. 8, pp. 54-58.
8. Koivisto, M., Perola, M., Varilo, T., Hennah, W., Ekelund, J., Lukk, M., Peltonen, L., Ukkonen, E. & Mannila, H. An MDL Method for Finding Haplotype Blocks and for Estimating the Strength of Haplotype Block Boundaries. *Proc. of Pacific Symposium on Biocomputing 2003*, Kauai, Hawaii, January 3-7, 2003. 2002, World Scientific, pp. 502-513.
9. Leung, C.K., Ng, R. & Mannila, H. OSSM: A Segmentation Approach to Optimize Frequency Counting. *Proc. of 18th International Conference on Data Engineering (ICDE 2002)*, San Jose, CA, USA, February 26 - March 1, 2002. pp. 583-593.

10. Mannila, H. Global and Local Methods in Data Mining: Basic Techniques and Open Problems. *Proc. of 29th International Colloquium on Automata, Languages, and Programming, ICALP 2002*, Malaga, Spain, July 2002. 2002, Springer-Verlag, pp. 57-68.
11. Mannila, H., Patrikainen, A., Seppänen, J. K. & Kere, J. Long-Range Control of Expression in Yeast. *Bioinformatics*, 2002. Vol. 18, pp. 482-483.
12. Niini, T., Vettenranta, K., Hollmén, J., Larramendy, M. L., Aalto, Y., Wikman, H., Nagy, B., Seppänen, J. K., Salvador, A. F., Mannila, H., Saarinen-Pihkala, U. M. & Knuutila, S. Expression of Myeloid-Specific Genes in Childhood Acute Lymphoblastic Leukemia – a cDNA Array Study. *Leukemia*, 2002. Vol. 16, No. 11, pp. 2213-2221.
13. Onkamo, P., Ollikainen, V., Sevon, P., Toivonen, H.T.T., Mannila, H. & Kere, J. Association Analysis for Quantitative Traits by Data Mining: QHPM. *The Annals of Human Genetics*, 2002. Vol. 66, pp. 419-429.
14. Ruosaari, S. & Hollmén, J. Image Analysis for Detecting Faulty Spots from Microarray Images. *Proc. of 5th International Conference on Discovery Science (DS'2002)*, Lübeck, Germany, November 24-26, 2002. Springer, pp. 259-266.
15. Salmenkivi, M., Kere, J. & Mannila, H. Genome Segmentation using Piecewise Constant Intensity Models and Reversible Jump MCMC. *Bioinformatics (European Computational Biology Conference 2002)*, 2002. Vol. 18, Supplement 2, pp. S211-S218.
16. Seppänen, J. K., Hollmén, J., Bingham, E. & Mannila, H. Nonnegative Matrix Factorization on Gene Expression Data. *Bioinformatics 2002*, Bergen, Norway, April 4-7, 2002. poster 49. <http://www.ii.uib.no/bio2002/abstractsbio2002.pdf>
17. Wikman, H., Kettunen, E., Seppänen, J. K., Karjalainen, A., Hollmén, J., Anttila, S. & Knuutila, S. Identification of Differentially Expressed Genes in Pulmonary Adenocarcinoma by Using a cDNA Array. *Oncogene*, 2002. Vol. 21, No. 37, pp. 5804-5813.
18. Zhu, Y., Hollmén, J., Rätty, R., Aalto, Y., Nagy, B., Elonen, E., Kere, J., Mannila, H., Franssila, K. & Knuutila, S. Investigatory and Analytical Approaches to Differential Gene Expression Profiling in Mantle Cell Lymphoma. *British Journal of Haematology*, 2002. Vol. 119, No. 4, pp. 905-915.

2003

1. Bykowski, A., Seppänen, J. & Hollmén, J. Model-Independent Bounding of the Supports of Boolean Formulae in Binary Data. In: Lanzi, P. & Meo, R. (eds.), *Database Support for Data Mining Applications*. Heidelberg, Springer-Verlag, in press.
2. Gionis, A., Kujala, T. & Mannila, H. Fragments of Order. *Proc. of ACM SIGKDD 2003*, Washington, D.C., USA, August 24-27, 2003. pp. 129-136.
3. Gionis, A. & Mannila, H. Finding Recurrent Sources in Sequences. *Proc. of ACM ReCOMB 2003*, Berlin, Germany, April 10-13, 2003. pp. 123-130.

4. Gunopulos, D., Khardon, R., Mannila, H., Saluja, S., Toivonen, H. & Sarma, R.S. Discovering All Most Specific Sentences. *ACM Transactions on Database Systems*, 2003. Vol. 28, No. 2, pp. 140-174.
5. Hollmén, J., Seppänen, J.K. & Mannila, H. Mixture Models and Frequent Sets: Combining Global and Local Methods for 0-1 Data. *Proc. of Third SIAM International Conference on Data Mining*, San Francisco, CA, USA, May 1-3, 2003. Society of Industrial and Applied Mathematics, pp. 289-293.
6. Katajamaa, M. & Hollmén, J. Simulation Model for Gene Expression Data. *Currents in Computational Molecular Biology 2003, Proc. of 7th Annual International Conference on Research in Computational Molecular Biology, RECOMB*, April 2003. pp. 249-250.
7. Kettunen, E., Anttila, S., Seppänen, J.K., Karjalainen, A., Edgren, H., Lindström, I., Salovaara, R., Nissén, A.-M., Salo, J., Mattson, K., Hollmén, J., Knuutila, S. & Wikman, H. Differentially Expressed Genes in Non-Small Cell Lung Cancer (NSCLC). *Cancer Genetics and Cytogenetics*, 2003. In press.
8. Leino, A., Mannila, H., Pitkänen, R.-L. Rule Discovery and Probabilistic Modeling for Onomastic Data. *Knowledge Discovery in Databases: PKDD 2003, Proc. of 7th European Conference on Principles and Practice of Knowledge Discovery in Databases*, Cavtat-Dubrovnik, Croatia, September 22-26, 2003. pp. 291-302.
9. Mielikäinen, T. & Mannila, H. The Pattern Ordering Problem. *Knowledge Discovery in Databases: PKDD 2003, Proc. of 7th European Conference on Principles and Practice of Knowledge Discovery in Databases*, Cavtat-Dubrovnik, Croatia, September 22-26, 2003. pp. 327-338.
10. Pavlov, D., Mannila, H. & Smyth, P. Beyond Independence: Probabilistic Methods for Query Approximation on Binary Transaction Data. *IEEE Transactions on Data and Knowledge Engineering*, 2003. Vol. 15, No. 6, pp. 1409-1421.
11. Ruosaari, S. & Hollmén, J. Identifying Differentially Expressed Genes. *Proc. of TICSP Workshop on Computational Systems Biology*, Tampere, Finland, June 16-17, 2003.
12. Ruosaari, S. & Hollmén, J. Identifying Differentially Expressed Genes with Bootstrap-Based Testing. *Proc. of Bioinformatics 2003*, Helsinki, Finland, May 22-24, 2003.
13. Seppänen, J. K., Bingham, E. & Mannila, H. A Simple Algorithm for Topic Identification in 0-1 Data. *Knowledge Discovery in Databases: PKDD 2003, Proc. of 7th European Conference on Principles and Practice of Knowledge Discovery in Databases*, Cavtat-Dubrovnik, Croatia, September 22-26, 2003. pp. 423-434.