

# Publications of the From Data to Knowledge Research Unit

Publications are in alphabetical order by the first author.

## 2004

1. Afrati, F.; Gionis, A.; Mannila, H. Approximating a Collection of Frequent Sets. *Proc. of 10th International Conference on Knowledge Discovery and Data Mining (KDD)*, Seattle, Washington, USA, Aug. 22-25, 2004.
2. Bykowski, A.; Seppänen, J.; Hollmén, J. Model-Independent Bounding of the Supports of Boolean Formulae in Binary Data. In: Meo, R., Lanzi, P. & Klemettinen, M. (eds.), *Database Support for Data Mining Applications: Discovering Knowledge with Inductive Queries*. Lecture Notes in Computer Science Vol. 2682. Heidelberg 2004, Springer-Verlag, pp. 234-249.
3. Geerts, F.; Mannila, H.; Terzi, E. Relational Link-Based Ranking. *Proc. of 30th International Conference on Very Large Data Bases (VLDB'04)*, Toronto, Canada, August 29 - September 3, 2004.
4. Gionis, A.; Mannila, H.; Seppänen, J. K. Geometric and Combinatorial Tiles in 0-1 Data. *Proc. of 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2004)*, Pisa, Italy, September 20-24, 2004. pp. 173-184.
5. Gionis, A.; Mannila, H.; Terzi, E. Clustered Segmentations. *Proc. of 3rd Workshop on Mining Temporal and Sequential Data (TDM)*, Seattle, Washington, USA, Aug. 22-25, 2004.
6. Hiisilä, H.; Bingham, E. Dependencies Between Transcription Factor Binding Sites: Comparison Between ICA, NMF, PLSA and Frequent Sets. *Proc. of 4th IEEE International Conference on Data Mining*, Brighton, UK, November 1-4, 2004. pp. 114-121.
7. Himberg, J.; Hyvärinen, A.; Esposito, F. Validating the Independent Components of Neuroimaging Time-Series via Clustering and Visualization. *Neuroimage*, 2004. Vol. 22, No. 3, pp. 1214-1222.
8. Kaban, A.; Bingham, E.; Hirsimäki, T. Learning to Read Between the Lines: The Aspect Bernoulli Model. *Proc. of 4th SIAM International Conference on Data Mining*, Lake Buena Vista, Florida, USA, April 22-24, 2004. pp. 462-466.

9. 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). Expression Profiling of Cancer-Related Genes in Squamous Cell Lung Cancer. *Cancer Genetics and Cytogenetics*, 2004. Vol. 149, No. 2, pp. 98-106.
10. Luysaert, S.; Sulkava, M.; Raitio, H.; Hollmén, J. Evaluation of Forest Nutrition Based on Large-Scale Foliar Surveys: Are Nutrition Profiles the Way of the Future? *Journal of Environmental Monitoring*, 2004. Vol. 6, No. 2, pp. 160-167.
11. Mäntyjärvi, J.; Himberg, J.; Kangas, P.; Tuomela, U.; Huuskonen, P. Sensor Signal Data Set for Exploring Context Recognition of Mobile Devices. *Proc. of Workshop "Benchmarks and a database for context recognition"*, in conjunction with the 2nd Int. Conf. on Pervasive Computing (PERVASIVE 2004), Linz/Vienna, Austria, April 18-23, 2004. (Electronic publication, Swiss Federal Institute of Technology Zurich, Electronics laboratory)
12. Mäntyjärvi, J.; Nybergh, K.; Himberg, J.; Hjelt, K. Touch Detection System for Mobile Terminals. *Proc. of Mobile Human-Computer Interaction - MobileHCI 2004: 6th International Symposium*, Glasgow, UK, September 13-16, 2004. Heidelberg 2004, Springer-Verlag, pp. 331-336.
13. Patrikainen, A.; Mannila, H. Subspace Clustering of Binary Data - A Probabilistic Approach. *Proc. of SIAM Data Mining 2004, Workshop on Clustering High-Dimensional Data*, Lake Buena Vista, Florida, USA, April 22-24, 2004.
14. Puolamäki, K.; Savia, E.; Sinkkonen, J.; Kaski, S. Two-Way Latent Grouping Model for User Preference Prediction. Espoo, Finland: Helsinki University of Technology, 2004. (Publications in Computer and Information Science, Report A80).
15. Salojärvi, J.; Puolamäki, K.; Kaski, S. Relevance Feedback from Eye Movements for Proactive Information Retrieval. *Proc. of Workshop on Processing Sensory Information for Proactive Systems (PSIPS 2004)*, Oulu, Finland, June 14-15, 2004. pp. 37-42.
16. Seppänen, J.; Mannila, H. Dense Itemsets. *Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD-2004)*, Seattle, WA, USA, August 22-25, 2004. pp. 683-688.
17. Sulkava, M.; Tikka, J.; Hollmén, J. Sparse Regression for Analyzing the Development of Foliar Nutrient Concentrations in Coniferous Trees. *Proc. of Fourth International Workshop on Environmental Applications of Machine Learning (EAML 2004)*, Bled, Slovenia, September 2004. pp. 57-58.
18. Tikka, J.; Hollmén, J. Learning Linear Dependency Trees from Multivariate Time-series Data. *Proc. of Workshop on Temporal Data Mining: Algorithms, Theory and Applications*, in conjunction with The Fourth IEEE International Conference on Data Mining, Brighton, UK, November 2004.
19. Vesanto, J. & Hollmén, J. An Automated Report Generation Tool for the Data Understanding Phase. In: Abraham, A. , Jain, L. & van der Zwaag, B. J. (eds.), *Innovations in Intelligent Systems: Design, Management and Applications, Studies in Fuzziness and Soft Computing Vol. 140*. Heidelberg 2004, Springer (Physica) Verlag, chapter 5.

20. Wikman, H.; Seppänen, J. K.; Sarhadi, V. K.; Kettunen, E.; Salmenkivi, K.; Kuosma, E.; Vainio-Siukola, K.; Nagy, B.; Karjalainen, A.; Sioris, T.; Salo, J.; Hollmén, J.; Knuutila, S.; Anttila, S. Caveolins as Tumor Markers in Lung Cancer Detected by Combined Use of cDNA and Tissue Microarrays. *The American Journal of Pathology*, 2004. Vol. 203, pp. 584-593.

## 2005

1. Afrati, F.; Das, G.; Gionis, A.; Mannila, H.; Mielikäinen, T.; Tsaparas, P. Mining Chains of Relations. *Proc. of 5th International Conference on Data Mining, Houston (ICDM 2005)*, Texas, USA, November 27-30, 2005. pp. 553-556.
2. Boulicaut, J.-F.; de Raedt, L.; Mannila, H. (eds.) *Constraint-Based Mining and Inductive Databases*, Springer-Verlag LNCS Volume 3848. Berlin 2005, Springer-Verlag.
3. Bounsaythip, C.; Hollmén, J.; Kaski, S.; Oresic, M. (eds.) *Proceedings of KR BIO05, Symposium on Knowledge Representation in Bioinformatics*, Espoo, Finland, June 15-17, 2005. Espoo, Finland 2005, Helsinki University of Technology. 50 p.
4. Esposito, F.; Scarabino, T.; Hyvärinen, A.; Himberg, J.; Formisano, E.; Comani, S.; Tedeschi, G.; Goebel, R.; Seifritz, E.; Di Salle, F. Independent Component Analysis of fMRI Group Studies by Self-Organizing Clustering. *Neuroimage*, 2005. Vol. 25, No. 1, pp. 193-205.
5. Gionis, A.; Mannila, H.; Tsaparas, P. Clustering Aggregation. *Proc. of 21st International Conference on Data Engineering (ICDE 2005)*, Tokyo, Japan, April 5-8, 2005. pp. 341-352.
6. Hyvönen, S.; Junninen, H.; Laakso, L.; Dal Maso, M.; Grönholm, T.; Bonn, B.; Keronen, P.; Aalto, P.; Hiltunen, V.; Pohja, T.; Launiainen, S.; Hari, P.; Mannila, H.; Kulmala, M. A Look at Aerosol Formation using Data Mining Techniques. *Atmos. Chem. Phys.*, 2005. Vol. 5, pp. 3345-3356.
7. Kettunen, E.; Nicholson, A.G.; Nagy, B.; Wikman, H.; Seppänen, J.K.; Stjernvall, T.; Ollikainen, T.; Kinnula, V.; Nordling, S.; Hollmén, J.; Anttila, S.; Knuutila, S. L1CAM, INP10, P-cadherin, tPA and ITGB4 Over-Expression in Malignant Pleural Mesotheliomas Revealed by Combined Use of cDNA and Tissue Microarray. *Carcinogenesis*, 2005. Vol. 26, No. 1, pp. 17-25.
8. Luyssaert, S.; Sulkava, M.; Raitio, H.; Hollmén, J. Are N and S Deposition Altering the Chemical Composition of Norway Spruce and Scots Pine Needles in Finland? *Environmental Pollution*, 2005. Vol. 138, No. 1, pp. 5-17.
9. Mannila, H.; Salmenkivi, M. Piecewise Constant Modeling of Sequential Data Using Reversible Jump Markov Chain Monte Carlo. In: Wang, J.; Zaki, M.; Toivonen, H.; Shasha, D. (eds.), *Data Mining in Bioinformatics*. Berlin 2005, Springer, pp. 85-103.
10. Papadimitriou, S.; Gionis, A.; Tsaparas, P.; Vaisanen, R.A.; Mannila, H.; Faloutsos, C. Parameter-Free Spatial Data Mining Using MDL. *Proc. of 5th International Conference on Data Mining (ICDM 2005)*, Houston, Texas, USA, November 27-30, 2005. pp. 346-353.

11. Puolamäki, K.; Salojärvi, J.; Savia, E.; Simola, J.; Kaski, S. Combining Eye Movements and Collaborative Filtering for Proactive Information Retrieval. *SIGIR 2005: Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Salvador, Brasil, August 15-19, 2005. New York, USA 2005, ACM Press, pp. 146-153.
12. Rastas, P.; Koivisto, M.; Mannila, H.; Ukkonen, E. A Hidden Markov Technique for Haplotype Reconstruction. *Proc. of Algorithms in Bioinformatics: 5th International Workshop (WABI 2005)*, Lecture Notes in Computer Science, 3692. Berlin 2005, Springer, pp. 140-151.
13. Salmenkivi, M.; Mannila, H. Using Markov Chain Monte Carlo and Dynamic Programming for Event Sequence Data. *Knowl. Inf. Syst.*, 2005. Vol. 7, No. 3, pp. 267-288.
14. Salojärvi, J.; Puolamäki, K.; Kaski, S. Expectation Maximization Algorithms for Conditional Likelihoods. *Proc. of 22nd International Conference on Machine Learning (ICML 2005)*, Bonn, Germany, August 7-11, 2005. New York, USA 2005, ACM Press, pp. 753-760.
15. Salojärvi, J.; Puolamäki, K.; Kaski, S. Expectation Maximization Algorithms for Conditional Likelihoods. Espoo, Finland: Helsinki University of Technology, 2005. (Publications in Computer and Information Science Report A83).
16. Salojärvi, J.; Puolamäki, K.; Kaski, S. Implicit Relevance Feedback From Eye Movements. *Artificial Neural Networks: Biological Inspirations - ICANN 2005: 15th International Conference*, Warsaw, Poland, 11-15 September, 2005. Lecture Notes in Computer Science 3696. Berlin, Germany 2005, Springer-Verlag, pp. 513-518.
17. Salojärvi, J.; Puolamäki, K.; Kaski, S. On Discriminative Joint Density Modeling. *Machine Learning: ECML 2005, European Conference on Machine Learning*, Porto, Portugal, October 3-7, 2005. Lecture Notes in Artificial Intelligence 3270. Berlin, Germany 2005, Springer-Verlag, pp. 341-352.
18. Salojärvi, J.; Puolamäki, K.; Simola, J.; Kovanen, L.; Kojo, I.; Kaski, S. Inferring Relevance from Eye Movements: Feature Extraction. Espoo, Finland: Helsinki University of Technology, 2005. 23 p. (Publications in Computer and Information Science Report A82).
19. Savia, E.; Puolamäki, K.; Sinkkonen, J.; Kaski, S. Two-Way Latent Grouping Model for User Preference Prediction. *Proc. of 21st Conference on Uncertainty in Artificial Intelligence (UAI 2005)*, Edinburgh, Scotland, July 26-29, 2005. pp. 518-525.
20. Seppänen, J. Upper Bound for the Approximation Ratio of a Class of Hypercube Segmentation Algorithms. *Information Processing Letters*, 2005. Vol. 93, No. 3, pp. 139-141.
21. Seppänen, J.K.; Mannila, H. Boolean Formulas and Frequent Sets. In: Boulicaut, J.-C.; de Raedt, L.; Mannila, H. (eds.), *Constraint-Based Mining and Inductive Databases*, LNCS Volume 3848. Berlin 2005, Springer-Verlag, pp. 348-361.
22. Sulkava, M.; Rautio, P.; Hollmén, J. Combining Measurement Quality into Monitoring Trends in Foliar Nutrient Concentrations. *Artificial Neural Networks: Formal Models and Their Applications, International Conference on Artificial Neural*

- Networks (ICANN'05)*, Warsaw, Poland, September 11-15, 2005. Lecture Notes in Computer Science 3697. Berlin 2005, Springer, pp. 761-767.
23. Tikka, J.; Hollmén, J.; Lendasse, A. Input Selection for Long-Term Prediction of Time Series. *Proc. of 8th International Work-Conference on Artificial Neural Networks (IWANN 2005)*, Barcelona, Spain, June 8-10, 2005. Berlin 2005, Springer-Verlag, pp. 1002-1009.
  24. Ukkonen, A.; Fortelius, M.; Mannila, H. Finding Partial Orders from Unordered 0-1 Data. *Proc. of Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Chicago, Illinois, USA, August 21-24, 2005. pp. 285-293.