Exploiting the Power of Group Differences

Using Patterns to Solve Data Analysis Problems

Guozhu Dong
Wright State University

MORGAN & CLAYPOOL PUBLISHERS

Bibliography

[1] Neda Abdelhamid, Aladdin Ayesh, and Fadi Thabtah. Phishing detection based associative classification data mining. *Expert Systems with Applications*, 41(13):5948–5959, 2014. DOI: 10.1016/j.eswa.2014.03.019 36

[2] Pierre Accorsi, Arnaud Sallaberry, Nathalie Lalande, Sandra Bringay, Florence Le Ber, Pascal Poncelet, Mickaël Fabrègue, Flavie Cernesson, Agnès Braud, and Maguelonne Teisseire. Hydroqual: Visual analysis of river water quality. In *VAST: Visual Analytics Science and Technology*, pages 123–132, 2014. DOI: 10.1109/vast.2014.7042488 18

[3] Niusvel Acosta-Mendoza, Andrés Gago-Alonso, Jesús Ariel Carrasco-Ochoa, José Francisco Martínez-Trinidad, and José Eladio Medina-Pagola. Improving graph-based image classification by using emerging patterns as attributes. *Engineering Applications of Artificial Intelligence*, 50:215–225, 2016. DOI: 10.1016/j.engappai.2016.01.030 97

[4] Charu C. Aggarwal and Philip S. Yu. Outlier detection for high dimensional data. In *ACM SIGMOD Record*, vol. 30, pages 37–46, 2001. DOI: 10.1145/376284.375668 44

[5] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. In *ACM SIGMOD Conference*, pages 207–216, 1993. DOI: 10.1145/170036.170072 12, 22

[6] Javier Lopez Alberca. Natural emerging patterns: A reformulation for classification. Ph.D. thesis, Royal Institute of Technology, 2014. 25

[7] Marharyta Aleksandrova. Matrix factorization and contrast analysis techniques for recommendation. Ph.D. thesis, Université de Lorraine, 2017. 97

[8] Hamad Alhammady and Kotagiri Ramamohanarao. The application of emerging patterns for improving the quality of rare-class classification. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 207–211, Springer, 2004. DOI: 10.1007/978-3-540-24775-3_27 95

[9] Hamad Alhammady and Kotagiri Ramamohanarao. Using emerging patterns and decision trees in rare-class classification. In *Proc. of IEEE International Conference on Data Mining (ICDM)*, pages 315–318, 2004. DOI: 10.1109/icdm.2004.10058 95

[10] Hamad Alhammady and Kotagiri Ramamohanarao. Expanding the training data space using emerging patterns and genetic methods. In *Proc. of SIAM International Conference on Data Mining*, pages 481–485, 2005. DOI: 10.1137/1.9781611972757.45 95

[11] Hamad Alhammady and Kotagiri Ramamohanarao. Mining emerging patterns and classification in data streams. In *Proc. of IEEE/WIC/ACM International Conference on Web Intelligence*, pages 272–275, 2005. DOI: 10.1109/wi.2005.96 96

[12] Hamad Alhammady and Kotagiri Ramamohanarao. Using emerging patterns to construct weighted decision trees. *IEEE Transactions on Knowledge and Data Engineering*, 18(7):865–876, 2006. DOI: 10.1109/tkde.2006.116 95

[13] Zulfiqar Ali and Waseem Shahzad. EPACO: A novel ant colony optimization for emerging patterns based classification. *Cluster Computing*, pages 1–15, 2017. DOI: 10.1007/s10586-017-0894-4 96

[14] U. Alon, N. Barkai, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. of the National Academy of Science*, 96:6745–6750, 1999. DOI: 10.1073/pnas.96.12.6745 72

[15] William F. Anderson and Rayna Matsuno. Breast cancer heterogeneity: A mixture of at least two main types? *Journal of the National Cancer Institute*, 98:948–951, 2006. DOI: 10.1093/jnci/djj295 83

[16] Piotr Andruszkiewicz. Lazy approach to privacy preserving classification with emerging patterns. In *Proc. of the 19th International Symposium of Emerging Intelligent Technologies in Industry (ISMIS)*, pages 253–268, 2011. DOI: 10.1007/978-3-642-22732-5_21 95

[17] Yasmine Asses, Aleksey Buzmakov, Thomas Bourquard, Sergei O. Kuznetsov, and Amedeo Napoli. A hybrid classification approach based on FCA and emerging patterns-an application for the classification of biological inhibitors. In *CLA: The 9th International Conference on Concept Lattices and their Applications*, 2012. 8, 92, 99

[18] Jens Auer and Jurgen Bajorath. Emerging chemical patterns: A new methodology for molecular classification and compound selection. *Journal of Chemical Information and Modeling*, 46(6):2502–2514, 2006. DOI: 10.1002/chin.200708200 30, 33, 37, 38, 39, 87, 98

[19] Jens Auer and Jurgen Bajorath. Distinguishing between bioactive and modeled compound conformations through mining of emerging chemical patterns. *Journal of Chemical Information and Modeling*, 48(9):1747–1753, 2008. DOI: 10.1021/ci8001793 88, 98

[20] Jens Auer and Jurgen Bajorath. Simulation of sequential screening experiments using emerging chemical patterns. *Medicinal Chemistry*, 4(1):80–90, 2008. DOI: 10.2174/157340608783331452 88, 98

[21] James Bailey, Thomas Manoukian, and Kotagiri Ramamohanarao. Classification using constrained emerging patterns. In *International Conference on Web-Age Information Management*, pages 226–237, Springer, 2003. DOI: 10.1007/978-3-540-45160-0_22 95

[22] James Bailey, Thomas Manoukian, and Kotagiri Ramamohanarao. A fast algorithm for computing hypergraph transversals and its application in mining emerging patterns. In *Proc. of IEEE International Conference on Data Mining (ICDM)*, pages 485–488, 2003. DOI: 10.1109/icdm.2003.1250958 25

[23] Elena Baralis, Silvia Chiusano, and Paolo Garza. A lazy approach to associative classification. *IEEE Transactions on Knowledge and Data Engineering*, 20(2):156–171, 2008. DOI: 10.1109/tkde.2007.190677 36

[24] Yves Bastide, Rafik Taouil, Nicolas Pasquier, Gerd Stumme, and Lotfi Lakhal. Mining frequent patterns with counting inference. *ACM SIGKDD Explorations Newsletter*, 2(2):66–75, 2000. DOI: 10.1145/380995.381017 13

[25] Stephen D. Bay and Michael J. Pazzani. Detecting change in categorical data: Mining contrast sets. In *ACM International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 302–306, 1999. 2, 8, 17
DOI: 10.1145/312129.312263

[26] Stephen D. Bay and Michael J. Pazzani. Characterizing model errors and differences. In *Proc. of International Conference on Machine Learning (ICML)*, pages 49–56, 2000. 8

[27] Stephen D. Bay and Michael J. Pazzani. Detecting group differences: Mining contrast sets. *Data Mining Knowledge Discovery*, 5(3):213–246, 2001. 8, 17

[28] Kevin Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft. When is "nearest neighbor" meaningful? In *International Conference on Database Theory*, pages 217–235, Springer, 1999. DOI: 10.1007/3-540-49257-7_15 44, 58

[29] Philipp Blandfort, Jörn Hees, and Desmond U. Patton. An overview of computational approaches for analyzing interpretation. *ArXiv Preprint ArXiv:1811.04028*, 2018. 98

[30] Ilona Bluemke and Marcin Tarka. Detection of anomalies in a SOA system by learning algorithms. In *Complex Systems and Dependability*, pages 69–85, Springer, 2013. DOI: 10.1007/978-3-642-30662-4_5 50

[31] Anne-Laure Boulesteix and Gerhard Tutz. Identification of interaction patterns and classification with applications to microarray data. *Computational Statistics and Data Analysis*, 50(3):783–802, 2006. DOI: 10.1016/j.csda.2004.10.004 69

[32] Evan A. Boyle, Yang I. Li, and Jonathan K. Pritchard. An expanded view of complex traits: From polygenic to omnigenic. *Cell*, 169(7):1177–1186, 2017. DOI: 10.1016/j.cell.2017.05.038 68

[33] Björn Bringmann, Siegfried Nijssen, and Albrecht Zimmermann. Pattern-based classification: A unifying perspective. *ArXiv Preprint ArXiv:1111.6191*, 2011. 36

[34] Anna L. Buczak and Erhan Guven. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys and Tutorials*, 18(2):1153–1176, 2016. DOI: 10.1109/comst.2015.2494502 44

[35] Cristóbal J. Carmona, M. J. del Jesus, and Francisco Herrera. A unifying analysis for the supervised descriptive rule discovery via the weighted relative accuracy. *Knowledge-Based Systems*, 139:89–100, 2018. DOI: 10.1016/j.knosys.2017.10.015 8

[36] Olivier Cavadenti, Victor Codocedo, Jean-François Boulicaut, and Mehdi Kaytoue. What did I do wrong in my MOBA game? Mining patterns discriminating deviant behaviours. In *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 662–671, 2016. DOI: 10.1109/dsaa.2016.75 18, 97

[37] Michelangelo Ceci, Annalisa Appice, Costantina Caruso, and Donato Malerba. Discovering emerging patterns for anomaly detection in network connection data. In *International Symposium on Methodologies for Intelligent Systems*, pages 179–188, Springer, 2008. DOI: 10.1007/978-3-540-68123-6_20 50

[38] Michelangelo Ceci, Annalisa Appice, Lucrezia Macchia, and Donato Malerba. Relational classification based on emerging patterns. In *Proc. of the 16th Italian Symposium on Advanced Database Systems (SEBD)*, pages 45–56, 2008. 33

[39] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011. DOI: 10.1145/1961189.1961199 48

[40] Elaheh Alipour Chavary, Sarah M. Erfani, and Christopher Leckie. Summarizing significant changes in network traffic using contrast pattern mining. In *Proc. of ACM Conference on Information and Knowledge Management*, pages 2015–2018, 2017. DOI: 10.1145/3132847.3133111 97

[41] Lijun Chen and Guozhu Dong. Masquerader detection using OCLEP: One class classification using length statistics of emerging patterns. In *International Workshop on Information Processing over Evolving Networks (WINPEN)*, 2006. DOI: 10.1109/waimw.2006.19 45, 49

[42] Lijun Chen and Guozhu Dong. Using emerging patterns in outlier and rare-class prediction. In *Contrast Data Mining: Concepts, Algorithms, and Applications*, Guozhu Dong and James Bailey, Eds., pages 171–186, Chapman & Hall/CRC, 2012. DOI: 10.1201/b12986-17 49

[43] Xiangtao Chen, Yajing Gao, and Siqi Ren. A new contrast pattern-based classification for imbalanced data. In *Proc. of the 2nd International Symposium on Computer Science and Intelligent Control*, page 45, ACM, 2018. DOI: 10.1145/3284557.3284708 96

[44] Hong Cheng, Xifeng Yan, Jiawei Han, and Philip S. Yu. Direct discriminative pattern mining for effective classification. In *Proc. of IEEE International Conference on Data Engineering*, pages 169–178, 2008. DOI: 10.1109/icde.2008.4497425 8, 25

[45] Michael W. K. Cheng, Byron Choi, and William Kwok-Wai Cheung. Hiding emerging patterns with local recoding generalization. In *Proc. of Advances in Knowledge Discovery and Data Mining (PAKDD)*, pages 158–170, 2010. DOI: 10.1007/978-3-642-13657-3_19 98

[46] Darrell Conklin. Antipattern discovery in folk tunes. *Journal of New Music Research*, 42(2):161–169, 2013. DOI: 10.1080/09298215.2013.809125 90

[47] Nicholas Cook. Computational and comparative musicology. *Empirical Musicology: Aims, Methods, Prospects*, pages 103–126, 2004. DOI: 10.1093/acprof:oso/9780195167498.003.0006 90

[48] Laurence Coquin, Steven J. Canipa, William C. Drewe, Lilia Fisk, Valerie J. Gillet, Mukesh Patel, Jeffrey Plante, Richard J. Sherhod, and Jonathan D. Vessey. New structural alerts for ames mutagenicity discovered using emerging pattern mining techniques. *Toxicology Research*, 4(1):46–56, 2015. DOI: 10.1039/c4tx00071d 99

[49] Val Curtis, Robert Aunger, and Tamer Rabie. Evidence that disgust evolved to protect from risk of disease. *Proc. of the Royal Society of London B: Biological Sciences*, 271(Suppl 4):S131–S133, 2004. DOI: 10.1098/rsbl.2003.0144 2

[50] Kang Deng and Osmar R. Zaïane. Contrasting sequence groups by emerging sequences. In *International Conference on Discovery Science*, pages 377–384, Springer, 2009. DOI: 10.1007/978-3-642-04747-3_29 96

[51] Guozhu Dong and James Bailey. *Contrast Data Mining: Concepts, Algorithms, and Applications*. CRC Press, 2012. DOI: 10.1201/b12986 25

[52] Guozhu Dong and Sanjeev Bhatta. Subpopulation-wise conditional correlation modeling and analysis. In *IEEE International Conference on Big Knowledge (ICBK)*, pages 17–24, 2017. DOI: 10.1109/icbk.2017.19 85

[53] Guozhu Dong and Kaustubh Deshpande. Efficient mining of niches and set routines. In *Proc. of Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, pages 234–246, 2001. DOI: 10.1007/3-540-45357-1_27 97

[54] Guozhu Dong and Neil Fore. Discovering dynamic logical blog communities based on their distinct interest profiles. In *Proc. of International Conference on Social Eco-Informatics (SOTICS)*, 2011. 63

[55] Guozhu Dong, Chunyu Jiang, Jian Pei, Jinyan Li, and Limsoon Wong. Mining succinct systems of minimal generators of formal concepts. In *Database Systems for Advanced Applications (DASFAA)*, pages 175–187, 2005. DOI: 10.1007/11408079_17 13

[56] Guozhu Dong and Jinyan Li. Efficient mining of emerging patterns: Discovering trends and differences. In *Proc. of ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pages 43–52, 1999. DOI: 10.1145/312129.312191 2, 13, 16, 17, 18, 46, 71

[57] Guozhu Dong and Jinyan Li. Mining border descriptions of emerging patterns from dataset pairs. *Knowledge and Information Systems*, 8(2):178–202, 2005. DOI: 10.1007/s10115-004-0178-1 16, 18, 25

[58] Guozhu Dong, Jinyan Li, Guimei Liu, and Limsoon Wong. *Mining Conditional Contrast Patterns. Chapter in Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction*. Yanchang Zhao, Chengqi Zhang, and Longbing Cao, Eds., IGI Global, 2009. 25

[59] Guozhu Dong, Jinyan Li, and Limsoon Wong. *The Use of Emerging Patterns in the Analysis of Gene Expression Profiles for the Diagnosis and Understanding of Diseases*. Chapter in *New Generation of Data Mining Applications*, Mehmed Kantardzic and Jozef Zurada, Eds., IEEE Press, 2005. 89

[60] Guozhu Dong and Sai Kiran Pentukar. OCLEP+: One-class anomaly and intrusion detection using minimal length of emerging patterns. *CSE Tech Report*, Wright State University, available at arXiv:1811.09842, 2018. 49

[61] Guozhu Dong and Vahid Taslimitehrani. Pattern aided regression modeling and prediction model analysis. *IEEE Transactions on Knowledge Data Engineering (TKDE)*, 27:2452–2465, 2015. DOI: 10.1109/icde.2016.7498398 78, 80, 82, 83

[62] Guozhu Dong and Vahid Taslimitehrani. Pattern aided classification. In *Proc. of SIAM International Conference on Data Mining*, pages 225–233, 2016. DOI: 10.1137/1.9781611974348.26 80, 82, 84

[63] Guozhu Dong, Xiuzhen Zhang, Limsoon Wong, and Jinyan Li. CAEP: Classification by aggregating emerging patterns. In *Proc. of Discovery Science*, pages 30–42, 1999. DOI: 10.1007/3-540-46846-3_4 33, 36

[64] Fabrice Dosseville, François Rioult, and Sylvain Laborde. Why do sports officials dropout? In *Machine Learning and Data Mining for Sports Analytics*, 2013. 97

[65] James Dougherty, Ron Kohavi, and Mehran Sahami. Supervised and unsupervised discretization of continuous features. In *Proc. of International Conference on Machine Learning (ICML)*, pages 194–202, 1995. DOI: 10.1016/b978-1-55860-377-6.50032-3 10, 66

[66] Wouter Duivesteijn and Julia Thaele. Understanding where your classifier does (not) work. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 250–253, Springer, 2015. DOI: 10.1109/icdm.2014.10 8

[67] William DuMouchel, Wen-Hua Ju, Alan F. Karr, Matthias Schonlau, Martin Theusan, and Yehuda Vardi. Computer intrusion: Detecting masquerades. *Statistical Science*, 16(1):1–17, 2001. DOI: 10.1214/ss/998929476 48

[68] Hongjian Fan, Ming Fan, Kotagiri Ramamohanarao, and Mengxu Liu. Further improving emerging pattern based classifiers via bagging. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 91–96, Springer, 2006. DOI: 10.1007/11731139_13 36, 95

[69] Hongjian Fan and Kotagiri Ramamohanarao. An efficient single-scan algorithm for mining essential jumping emerging patterns for classification. In *Proc. of Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, pages 456–462, 2002. DOI: 10.1007/3-540-47887-6_45 96

[70] Hongjian Fan and Kotagiri Ramamohanarao. A Bayesian approach to use emerging patterns for classification. In *Proc. of Australasian Database Conference*, pages 39–48, 2003. 25, 36, 95

[71] Hongjian Fan and Kotagiri Ramamohanarao. Noise tolerant classification by chi emerging patterns. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 201–206, Springer, 2004. DOI: 10.1007/978-3-540-24775-3_26 36, 95

[72] Hongjian Fan and Kotagiri Ramamohanarao. A weighting scheme based on emerging patterns for weighted support vector machines. In *Proc. of IEEE International Conference on Granular Computing*, pages 435–440, 2005. DOI: 10.1109/grc.2005.1547329 95

[73] Gang Fang, Majda Haznadar, Wen Wang, Haoyu Yu, Michael Steinbach, Timothy R. Church, William S. Oetting, Brian Van Ness, and Vipin Kumar. High-order SNP combinations associated with complex diseases: Efficient discovery, statistical power and functional interactions. *PloS One*, 7(4):e33531, 2012. DOI: 10.1371/journal.pone.0033531 69

[74] Gang Fang, Gaurav Pandey, Wen Wang, Manish Gupta, Michael Steinbach, and Vipin Kumar. Mining low-support discriminative patterns from dense and high-dimensional data. *Knowledge and Data Engineering, IEEE Transactions on*, 24(2):279–294, 2012. DOI: 10.1109/tkde.2010.241 8

[75] Gang Fang, Wen Wang, Benjamin Oatley, Brian Van Ness, Michael Steinbach, and Vipin Kumar. Characterizing discriminative patterns. *Computing Research Repository*, abs/1102.4, 2011. 69

[76] Fabio Fassetti, Simona E. Rombo, and Cristina Serrao. Exceptional pattern discovery. In *Discriminative Pattern Discovery on Biological Networks*, pages 23–30, Springer, 2017. DOI: 10.1007/978-3-319-63477-7_3 8

[77] Usama M. Fayyad and Keki B. Irani. Multi-interval discretization of continuous-valued attributes for classification learning. In *Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1022–1029, 1993. DOI: 10.18297/etd/18 11

[78] Mengling Feng, Guozhu Dong, Jinyan Li, Yap-Peng Tan, and Limsoon Wong. Pattern space maintenance for data updates and interactive mining. *Computational Intelligence*, 26(3):282–317, 2010. DOI: 10.1111/j.1467-8640.2010.00360.x 62

[79] Juan L. Fernández-Martínez, Stephen T. Sonis, et al. Sensitivity analysis of gene ranking methods in phenotype prediction. *Journal of Biomedical Informatics*, 64:255–264, 2016. DOI: 10.1016/j.jbi.2016.10.012 66

[80] Neil Fore and Guozhu Dong. *CPC: A Contrast Pattern Based Clustering Algorithm*. Chapter in *Contrast Data Mining: Concepts, Algorithms and Applications*, Guozhu Dong and James Bailey, Eds., Chapman & Hall/CRC, 2013. 61, 62

[81] Min Gan and Honghua Dai. Detecting and monitoring abrupt emergences and submergences of episodes over data streams. *Information Systems*, 39:277–289, 2014. DOI: 10.1016/j.is.2012.05.009 96

[82] Milton García-Borroto, José Fco Martínez-Trinidad, and Jesús Ariel Carrasco-Ochoa. Fuzzy emerging patterns for classifying hard domains. *Knowledge and Information Systems*, 28(2):473–489, 2011. DOI: 10.1007/s10115-010-0324-x 36, 95

[83] Milton García-Borroto, José Fco Martínez-Trinidad, and Jesús Ariel Carrasco-Ochoa. A survey of emerging patterns for supervised classification. *Artificial Intelligence Review*, 42(4):705–721, 2014. DOI: 10.1007/s10462-012-9355-x 36, 95

[84] Milton García-Borroto, José Fco Martínez-Trinidad, Jesús Ariel Carrasco-Ochoa, Miguel Angel Medina-Pérez, and José Ruiz-Shulcloper. LCMine: An efficient algorithm for mining discriminative regularities and its application in supervised classifica-

tion. *Pattern Recognition*, 43(9):3025–3034, 2010. DOI: 10.1016/j.patcog.2010.04.008 25, 36

[85] A. M. García-Vico, C. J. Carmona, D. Martín, M. García-Borroto, and M. J. del Jesus. An overview of emerging pattern mining in supervised descriptive rule discovery: Taxonomy, empirical study, trends, and prospects. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2017. DOI: 10.1002/widm.1231 8

[86] Behzad Ghanbarian, Vahid Taslimitehrani, Guozhu Dong, and Yakov A. Pachepsky. Sample dimensions effect on prediction of soil water retention curve and saturated hydraulic conductivity. *Journal of Hydrology*, 528:127–137, 2015. DOI: 10.1016/j.jhydrol.2015.06.024 84

[87] Shameek Ghosh, Mengling Feng, Hung Nguyen, and Jinyan Li. Risk prediction for acute hypotensive patients by using gap constrained sequential contrast patterns. In *AMIA Annual Symposium Proceedings*, page 1748, American Medical Informatics Association, 2014. 96, 99

[88] Shameek Ghosh, Jinyan Li, Longbing Cao, and Kotagiri Ramamohanarao. Septic shock prediction for ICU patients via coupled HMM walking on sequential contrast patterns. *Journal of Biomedical Informatics*, 66:19–31, 2017. DOI: 10.1016/j.jbi.2016.12.010 91, 99

[89] Tao Gu, Liang Wang, Hanhua Chen, Guimei Liu, Xianping Tao, and Jian Lu. Mining emerging sequential patterns for activity recognition in body sensor networks. In *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*, pages 102–113, Springer, 2010. DOI: 10.1007/978-3-642-29154-8_9 94, 99

[90] Tao Gu, Zhanqing Wu, XianPing Tao, Hung Keng Pung, and Jian Lu. epSICAR: An emerging patterns based approach to sequential, interleaved and concurrent activity recognition. In *Proc. of IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 1–9, 2009. DOI: 10.1109/percom.2009.4912776 94, 99

[91] Jiawei Han, Jian Pei, and Micheline Kamber. *Data Mining: Concepts and Techniques*. Elsevier, 2011. 44

[92] Jiawei Han, Jian Pei, Yiwen Yin, and Runying Mao. Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, 8(1):53–87, 2004. DOI: 10.1023/b:dami.0000005258.31418.83 62

[93] Franciso Herrera, Cristóbal José Carmona, Pedro González, and María José Del Jesus. An overview on subgroup discovery: Foundations and applications. *Knowledge and Information Systems*, 29(3):495–525, 2011. DOI: 10.1007/s10115-010-0356-2 8

[94] Victoria Hodge and Jim Austin. A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2):85–126, 2004. DOI: 10.1023/b:aire.0000045502.10941.a9 44

[95] Robert C. Holte, Liane Acker, Bruce W. Porter, et al. Concept learning and the problem of small disjuncts. In *International Joint Conferences on Artificial Intelligence (IJCAI)*, vol. 89, pages 813–818, 1989. 28, 33

[96] Oleg Iegorov, Vincent Leroy, Alexandre Termier, Jean-François Méhaut, and Miguel Santana. Data mining approach to temporal debugging of embedded streaming applications. In *Proc. of the 12th International Conference on Embedded Software*, pages 167–176, IEEE Press, 2015. DOI: 10.1109/emsoft.2015.7318272 98

[97] James Jaccard, Robert Turrisi, and Jim Jaccard. *Interaction Effects in Multiple Regression*, no. 72, Sage, 2003. DOI: 10.4135/9781412984522 68

[98] James Jaccard, Choi K. Wan, and Robert Turrisi. The detection and interpretation of interaction effects between continuous variables in multiple regression. *Multivariate Behavioral Research*, 25(4):467–478, 1990. DOI: 10.1207/s15327906mbr2504_4 68

[99] Ioannis Kavakiotis, George Tzanis, and Ioannis Vlahavas. Polyadenylation site prediction using polyA-iEP method. In *Polyadenylation*, pages 131–140, Springer, 2014. DOI: 10.1007/978-1-62703-971-0_11 97, 99

[100] Hyunjoong Kim, Wei-Yin Loh, Yu-Shan Shih, and Probal Chaudhuri. Visualizable and interpretable regression models with good prediction power. *IIE Transactions*, 39(6):565–579, 2007. DOI: 10.1080/07408170600897502 83

[101] Young Bun Kim, Jung Hun Oh, and Jean Gao. Emerging pattern based subspace clustering of microarray gene expression data using mixture models. In *Advances in Bioinformatics and its Applications*, pages 13–24, World Scientific, 2005. DOI: 10.1142/9789812702098_0002 98

[102] Sofia Kleisarchaki. Difference analysis in big data: Exploration, explanation, evolution. Ph.D. thesis, Université Grenoble Alpes, 2016. 98

[103] Willi Klösgen. Explora: A multipattern and multistrategy discovery assistant. In *Advances in Knowledge Discovery and Data Mining*, pages 249–271, American Association for Artificial Intelligence, 1996. 8

[104] Ludmila I. Kuncheva and Christopher J. Whitaker. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, 51(2):181–207, 2003. DOI: 10.1023/A:1022859003006 83

[105] Paul Labute. Binary QSAR: A new method for the determination of quantitative structure activity relationships. In *Biocomputing*, pages 444–455, World Scientific, 1999. DOI: 10.1142/9789814447300_0044 38

[106] Himabindu Lakkaraju, Stephen H. Bach, and Jure Leskovec. Interpretable decision sets: A joint framework for description and prediction. In *Proc. of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1675–1684, ACM, 2016. DOI: 10.1145/2939672.2939874 99

[107] Heon Gyu Lee, Kiyong Noh, Bum Ju Lee, Ho-Sun Shon, and Keun Ho Ryu. Cardiovascular disease diagnosis method by emerging patterns. In *Proc. of the 2nd International Conference on Advanced Data Mining and Applications (ADMA)*, pages 819–826, 2006. DOI: 10.1007/11811305_89 98, 99

[108] Hochang B. Lee and Constantine G. Lyketsos. Depression in Alzheimer's disease: Heterogeneity and related issues. *Biological Psychiatry*, 54(3):353–362, 2003. DOI: 10.1016/s0006-3223(03)00543-2 83

[109] Jong Bum Lee, Minghao Piao, and Keun Ho Ryu. Incremental emerging patterns mining for identifying safe and non-safe power load lines. In *IEEE International Conference on Computer and Information Technology (CIT)*, pages 1424–1429, IEEE, 2010. DOI: 10.1109/cit.2010.255 97

[110] Alban Lepailleur, Guillaume Poezevara, and Ronan Bureau. Automated detection of structural alerts (chemical fragments) in (eco) toxicology. *Computational and Structural Biotechnology Journal*, 5(6), 2013. DOI: 10.5936/csbj.201302013 98

[111] Gang Li, Rob Law, Huy Quan Vu, Jia Rong, and Xinyuan Roy Zhao. Identifying emerging hotel preferences using emerging pattern mining technique. *Tourism Management*, 46:311–321, 2015. DOI: 10.1016/j.tourman.2014.06.015 97

[112] Jinyan Li, Guozhu Dong, and Kotagiri Ramamohanarao. Making use of the most expressive jumping emerging patterns for classification. In *Proc. of Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, pages 220–232, 2000. DOI: 10.1007/3-540-45571-x_29 17

[113] Jinyan Li, Guozhu Dong, and Kotagiri Ramamohanarao. Making use of the most expressive jumping emerging patterns for classification. *Knowledge and Information Systems*, 3(2):131–145, 2001. DOI: 10.1007/3-540-45571-x_29 35

[114] Jinyan Li, Guozhu Dong, Kotagiri Ramamohanarao, and Limsoon Wong. DeEPs: A new instance-based lazy discovery and classification system. *Machine Learning*, 54(2):99–124, 2004. DOI: 10.1023/b:mach.0000011804.08528.7d 34, 36

[115] Jinyan Li, Haiquan Li, Limsoon Wong, Jian Pei, and Guozhu Dong. Minimum description length principle: Generators are preferable to closed patterns. In *AAAI Conference on Artificial Intelligence*, pages 409–414, 2006. 13

[116] Jinyan Li, Guimei Liu, and Limsoon Wong. Mining statistically important equivalence classes and delta-discriminative emerging patterns. In *Proc. of ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pages 430–439, 2007. DOI: 10.1145/1281192.1281240 25, 55

[117] Jinyan Li, Huiqing Liu, James R. Downing, Allen Eng-Juh Yeoh, and Limsoon Wong. Simple rules underlying gene expression profiles of more than six subtypes of acute lymphoblastic leukemia (ALL) patients. *Bioinformatics*, 19(1):71–78, 2003. DOI: 10.1093/bioinformatics/19.1.71 30, 89, 99

[118] Jinyan Li, Kotagiri Ramamohanarao, and Guozhu Dong. Combining the strength of pattern frequency and distance for classification. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 455–466, Springer, 2001. DOI: 10.1007/3-540-45357-1_48 35, 95

[119] Jinyan Li and Limsoon Wong. Identifying good diagnostic gene groups from gene expression profiles using the concept of emerging patterns. *Bioinformatics*, 18(10):1406–1407, 2002. DOI: 10.1093/bioinformatics/18.10.1406 72

[120] Wenmin Li, Jiawei Han, and Jian Pei. CMAR: Accurate and efficient classification based on multiple class-association rules. In *Proc. of IEEE International Conference on Data Mining (ICDM)*, pages 369–376, 2001. DOI: 10.1109/icdm.2001.989541 35

[121] Kah Leong Lim, Prasanna R. Kolatkar, Kwok Peng Ng, Chee Hoe Ng, and Catherine J. Pallen. Interconversion of the kinetic identities of the tandem catalytic domains of receptor-like protein-tyrosine phosphatase PTPα by two point mutations is synergistic and substrate-dependent. *Journal of Biological Chemistry*, 273(44):28986–28993, 1998. DOI: 10.1074/jbc.273.44.28986 25

[122] Bing Liu, Wynne Hsu, and Yiming Ma. Integrating classification and association rule mining. In *Proc. of ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pages 80–86, 1998. 35

[123] Guimei Liu, Haojun Zhang, Mengling Feng, Limsoon Wong, and See-Kiong Ng. Supporting exploratory hypothesis testing and analysis. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 9(4):31, 2015. DOI: 10.1145/2701430 24, 97

[124] Lu Liu, Lei Duan, Hao Yang, Jyrki Nummenmaa, Guozhu Dong, and Pan Qin. Mining distinguishing customer focus sets for online shopping decision support. In *International*

Conference on Advanced Data Mining and Applications, pages 50–64, Springer, 2016. DOI: 10.1007/978-3-319-49586-6_4 97

[125] Qingbao Liu and Guozhu Dong. A contrast pattern based clustering quality index for categorical data. In *Proc. of IEEE International Conference on Data Mining (ICDM)*, pages 860–865, 2009. DOI: 10.1109/ICDM.2009.105 51, 52, 53, 54, 55, 56

[126] Qingbao Liu and Guozhu Dong. CPCQ: Contrast pattern based clustering quality index for categorical data. *Pattern Recognition*, 45(4):1739–1748, 2012. DOI: 10.1016/j.patcog.2011.10.007 51, 52, 53, 54, 55, 56

[127] Quanzhong Liu, Jiangning Song, and Jinyan Li. Using contrast patterns between true complexes and random subgraphs in PPI networks to predict unknown protein complexes. *Scientific Reports*, 6, 2016. DOI: 10.1038/srep21223 92, 99

[128] Xiaoqing Liu, Jun Wu, Feiyang Gu, Jie Wang, and Zengyou He. Discriminative pattern mining and its applications in bioinformatics. *Briefings in Bioinformatics*, 16(5):884–900, 2014. DOI: 10.1093/bib/bbu042 8

[129] Yong Liu, Jianzhong Li, and Jinghua Zhu. A novel graph classification approach based on frequent closed emerging patterns. *Journal of Computer Research and Development*, 44(7):1169–1176, 2007. DOI: 10.1360/crad20070711 95

[130] Elsa Loekito and James Bailey. Fast mining of high dimensional expressive contrast patterns using zero-suppressed binary decision diagrams. In *Proc. of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 307–316, 2006. DOI: 10.1145/1150402.1150438 8

[131] Octavio Loyola-González, José Fco Martínez-Trinidad, Jesús Ariel Carrasco-Ochoa, and Milton García-Borroto. Study of the impact of resampling methods for contrast pattern based classifiers in imbalanced databases. *Neurocomputing*, 175:935–947, 2016. DOI: 10.1016/j.neucom.2015.04.120 33, 36

[132] Octavio Loyola-González, José Fco Martínez-Trinidad, Jesús Ariel Carrasco-Ochoa, and Milton García-Borroto. A novel contrast pattern selection method for class imbalance problems. In *Mexican Conference on Pattern Recognition*, pages 42–52, Springer, 2017. DOI: 10.1007/978-3-319-59226-8_5 96

[133] Octavio Loyola-González, Miguel Angel Medina-Pérez, Dayton Hernández-Tamayo, Raúl Monroy, Jesús Ariel Carrasco-Ochoa, and Milton García-Borroto. A pattern-based approach for detecting pneumatic failures on temporary immersion bioreactors. *Sensors*, 19(2):414, 2019. DOI: 10.3390/s19020414 96

[134] Octavio Loyola-González, Miguel Angel Medina-Pérez, José Fco Martínez-Trinidad, Jesús Ariel Carrasco-Ochoa, Raúl Monroy, and Milton García-Borroto. PBC4cip: A new contrast pattern-based classifier for class imbalance problems. *Knowledge-Based Systems*, 115:100–109, 2017. DOI: 10.1016/j.knosys.2016.10.018 36, 96

[135] Sylvain Lozano, Guillaume Poezevara, Marie-Pierre Halm-Lemeille, Elodie Lescot-Fontaine, Alban Lepailleur, Ryan Bissell-Siders, Bruno Crémilleux, Sylvain Rault, Bertrand Cuissart, and Ronan Bureau. Introduction of jumping fragments in combination with QSARs for the assessment of classification in ecotoxicology. *Journal of Chemical Information and Modeling*, 50(8):1330–1339, 2010. DOI: 10.1021/ci100092x 98

[136] Hadi Tabatabaee Malazi and Mohammad Davari. Combining emerging patterns with random forest for complex activity recognition in smart homes. *Applied Intelligence*, 48(2):315–330, 2018. DOI: 10.1007/s10489-017-0976-2 94, 99

[137] Shihong Mao and Guozhu Dong. Discovery of highly differentiative gene groups from microarray gene expression data using the gene club approach. *Journal of Bioinformatics and Computational Biology*, 3(6):1263–1280, 2005. DOI: 10.1142/s0219720005001545 70, 75

[138] Shihong Mao and Guozhu Dong. *Towards Mining Optimal Emerging Patterns Amidst 1000s of Genes*. Chapter in *Contrast Data Mining: Concepts, Algorithms and Applications*, Guozhu Dong and James Bailey, Eds., Chapman & Hall/CRC, Data Mining and Knowledge Discovery Series, 2012. 75

[139] Shihong Mao, Charles Wang, and Guozhu Dong. Evaluation of inter-laboratory and cross-platform concordance of DNA microarrays through discriminating genes and classifier transferability. *Journal of Bioinformatics and Computational Biology*, 7(01):157–173, 2009. DOI: 10.1142/s0219720009004011 98

[140] Jean-Philippe Métivier, Bertrand Cuissart, Ronan Bureau, and Alban Lepailleur. The pharmacophore network: A computational method for exploring structure—activity relationships from a large chemical data set. *Journal of Medicinal Chemistry*, 61(8):3551–3564, 2018. DOI: 10.1021/acs.jmedchem.7b01890 88, 98

[141] Jean-Philippe Métivier, Alban Lepailleur, Aleksey Buzmakov, Guillaume Poezevara, Bruno Crémilleux, Sergei O. Kuznetsov, J'erémie Le Goff, Amedeo Napoli, Ronan Bureau, and Bertrand Cuissart. Discovering structural alerts for mutagenicity using stable emerging molecular patterns. *Journal of Chemical Information and Modeling*, 55(5):925–940, 2015. DOI: 10.1021/ci500611v 25, 89, 99

[142] Dean Mobbs, Cindy C. Hagan, Tim Dalgleish, Brian Silston, and Charlotte Prévost. The ecology of human fear: Survival optimization and the nervous system. *Frontiers in Neuroscience*, 9:55, 2015. DOI: 10.3389/fnins.2015.00055 2

[143] MOE (Molecular Operating EnVironment). Chemical Computing Group Inc., Montreal, Quebec, Canada H3B 3X3. 39

[144] David S. Moore, George P. McCabe, and Bruce A. Craig. *Introduction to the Practice of Statistics*. W H Freeman New York, 2009. DOI: 10.2307/1269120 24

[145] Hiroyuki Morita and Yukinobu Hamuro. A classification model using emerging patterns incorporating item taxonomy. In *Recent Progress in Data Engineering and Internet Technology*, pages 187–192, Springer, 2013. DOI: 10.1007/978-3-642-28807-4_26 36

[146] Takanobu Nakahara. Use of personal color and purchasing patterns for distinguishing fashion sensitivity. In *International Conference on Social Computing and Social Media*, pages 258–267, Springer, 2018. DOI: 10.1007/978-3-319-91485-5_20 98

[147] Mike A. Nalls, Nathan Pankratz, Christina M. Lill, Chuong B. Do, Dena G. Hernandez, Mohamad Saad, Anita L. DeStefano, Eleanna Kara, Jose Bras, Manu Sharma, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nature Genetics*, 46(9):989, 2014. DOI: 10.1038/ng.3043 67

[148] Vigneshwaran Namasivayam, Disha Gupta-Ostermann, Jenny Balfer, Kathrin Heikamp, and Jurgen Bajorath. Prediction of compounds in different local structure-activity relationship environments using emerging chemical patterns. *Journal of Chemical Information and Modeling*, 54(5):1301–1310, 2014. DOI: 10.1021/ci500147b 88, 98

[149] Vigneshwaran Namasivayam, Ye Hu, Jenny Balfer, and Jurgen Bajorath. Classification of compounds with distinct or overlapping multi-target activities and diverse molecular mechanisms using emerging chemical patterns. *Journal of Chemical Information and Modeling*, 53(6):1272–1281, 2013. DOI: 10.1021/ci400186n 88, 98

[150] Vigneshwaran Namasivayam, Preeti Iyer, and Jurgen Bajorath. Prediction of individual compounds forming activity cliffs using emerging chemical patterns. *Journal of Chemical Information and Modeling*, 53(12):3131–3139, 2013. DOI: 10.1021/ci400597d 88, 98

[151] Kerstin Neubarth and Darrell Conklin. Contrast pattern mining in folk music analysis. In *Computational Music Analysis*, pages 393–424, Springer, 2016. DOI: 10.1007/978-3-319-25931-4_15 90

[152] Mao Nishiguchi and Hiroyuki Morita. CAECP and CRPD: Classification by aggregating essential contrast patterns, and contrast ranked path diagrams. *Journal of Information and Knowledge Management*, 15(04), 2016. DOI: 10.1142/s0219649216500453 36

[153] Petra Kralj Novak, Nada Lavrac, and Geoffrey I. Webb. Supervised descriptive rule discovery: A unifying survey of contrast set, emerging pattern and subgroup mining. *Journal of Machine Learning Research*, 10:377–403, 2009. 8

[154] Megan Oaten, Richard J. Stevenson, and Trevor I. Case. Disgust as a disease-avoidance mechanism. *Psychological Bulletin*, 135(2):303, 2009. DOI: 10.1037/a0014823 2

[155] Katsuhiko Okada, Takahiro Azuma, Masakazu Nakamoto, and Yukinobu Hamuro. Stock performance after securities analyst's rating downgrades. *Transactions of the Japanese Society for Artificial Intelligence*, 27(6):355–364, 2012. DOI: 10.1527/tjsai.27.355 97

[156] Dijana Oreski, Irena Kedmenec, and Bozidar Klicek. Exploring capabilities of contrast mining application in SWOT analysis. *Proc. of the 8th MAC*, page 210, 2016. 97

[157] Stephan Pabinger, Andreas Dander, Maria Fischer, Rene Snajder, Michael Sperk, Mirjana Efremova, Birgit Krabichler, Michael R. Speicher, Johannes Zschocke, and Zlatko Trajanoski. A survey of tools for variant analysis of next-generation genome sequencing data. *Briefings in Bioinformatics*, 15(2):256–278, 2014. DOI: 10.1093/bib/bbs086 66, 68

[158] Xianchao Pan, Hu Mei, Sujun Qu, Shuheng Huang, Jiaying Sun, Li Yang, and Hua Chen. Prediction and characterization of P-glycoprotein substrates potentially bound to different sites by emerging chemical pattern and hierarchical cluster analysis. *International Journal of Pharmaceutics*, 502(1–2):61–69, 2016. DOI: 10.1016/j.ijpharm.2016.02.022 33

[159] Jin Hyoung Park, Heon Gyu Lee, Gyoyong Sohn, Jin-Ho Shin, and Keun Ho Ryu. Emerging pattern based classification for automated non-safe power line detection. In *Proc. of International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, pages 169–173, 2009. DOI: 10.1109/fskd.2009.769 97

[160] Nicolas Pasquier, Yves Bastide, Rafik Taouil, and Lotfi Lakhal. Discovering frequent closed itemsets for association rules. In *Proc. of International Conference on Database Theory*, pages 398–416, 1999. DOI: 10.1007/3-540-49257-7_25 13

[161] Jun Pei and Min Zhang. A new classification algorithm based on emerging patterns. In *Proc. of the 6th International Asia Conference on Industrial Engineering and Management Innovation*, pages 57–65, Springer, 2016. DOI: 10.2991/978-94-6239-145-1_6 95

[162] Minghao Piao, Heon Gyu Lee, Gyoyong Sohn, Gouchol Pok, and Keun Ho Ryu. Emerging patterns based methodology for prediction of patients with myocardial ischemia. In *Proc. of International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, pages 174–178, 2009. DOI: 10.1109/fskd.2009.638 33, 98, 99

[163] Guillaume Poezevara, Bertrand Cuissart, and Bruno Crémilleux. Discovering emerging graph patterns from chemicals. In *Proc. of International Symposium on Foundations of Intelligent Systems (ISMIS)*, pages 45–55, 2009. DOI: 10.1007/978-3-642-04125-9_8 98

[164] Guillaume Poezevara, Sylvain Lozano, Bertrand Cuissart, Ronan Bureau, Pierre Bureau, Vincent Croixmarie, Philippe Vayer, and Alban Lepailleur. A computational selection of metabolite biomarkers using emerging pattern mining: A case study in human hepatocellular carcinoma. *Journal of Proteome Research*, 16(6):2240–2249, 2017. DOI: 10.1021/acs.jproteome.7b00054 88, 99

[165] Adriana Prado, Marc Plantevit, Céline Robardet, and Jean-Francois Boulicaut. Mining graph topological patterns: Finding covariations among vertex descriptors. *IEEE Transactions on Knowledge and Data Engineering*, 25(9):2090–2104, 2013. DOI: 10.1109/tkde.2012.154 97

[166] Hemant Purohit, Guozhu Dong, Valerie Shalin, Krishnaprasad Thirunarayan, and Amit Sheth. Intent classification of short-text on social media. In *IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, pages 222–228, 2015. DOI: 10.1109/smartcity.2015.75 97

[167] J. Ross Quinlan. Learning efficient classification procedures and their application to chess end games. In *Machine Learning*, pages 463–482, Springer, 1983. DOI: 10.1016/b978-0-08-051054-5.50019-4 2

[168] J. Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993. 35

[169] Murugesan Raju, Danlu Liu, Frederick W. Fraunfelder, and Chi-Ren Shyu. Discovering multifactorial associations with the development of age-related cataract using contrast mining. In *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2297–2299, 2017. DOI: 10.1109/bibm.2017.8218033 98

[170] Kotagiri Ramamohanarao and Hongjian Fan. Patterns based classifiers. *World Wide Web*, 10(1):71–83, 2007. DOI: 10.1007/s11280-006-0012-7 36

[171] Anna Rampini, Gloria Bordogna, Paola Carrara, Monica Pepe, Massimo Antoninetti, Alessandro Mondini, and Paola Reichenbach. Modelling landslides' susceptibility by fuzzy emerging patterns. In *Landslide Science and Practice*, pages 363–370, Springer, 2013. DOI: 10.1007/978-3-642-31325-7_48 97

[172] Jenna M. Reps, Uwe Aickelin, and Richard B. Hubbard. Refining adverse drug reaction signals by incorporating interaction variables identified using emergent pattern mining. *Computers in Biology and Medicine*, 2015. DOI: 10.2139/ssrn.2822199 68, 90, 99

[173] Marylyn D. Ritchie, Lance W. Hahn, and Jason H. Moore. Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, 24(2):150–157, 2003. DOI: 10.1002/gepi.10218 69

[174] Marylyn D. Ritchie, Lance W. Hahn, Nady Roodi, L. Renee Bailey, William D. Dupont, Fritz F. Parl, and Jason H. Moore. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *The American Journal of Human Genetics*, 69(1):138–147, 2001. DOI: 10.1086/321276 68

[175] Guillaume A. Rousselet, Cyril R. Pernet, and Rand R. Wilcox. Beyond differences in means: Robust graphical methods to compare two groups in neuroscience. *European Journal of Neuroscience*, 46(2):1738–1748, 2017. DOI: 10.1111/ejn.13610 2, 24

[176] Khalid E. K. Saeed, Heon Gyu Lee, Wun-Jae Kim, Eun Jong Cha, and Keun Ho Ryu. Using emerging subsequence in classifying protein structural class. In *Proc. of International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, pages 349–353, 2009. DOI: 10.1109/fskd.2009.752 98, 99

[177] David Savage. Detection of illicit behaviours and mining for contrast patterns. Ph.D. thesis, RMIT, Australia, 2017. 98

[178] Chinar C. Shah, Xingquan Zhu, Taghi M. Khoshgoftaar, and Justin Beyer. Contrast pattern mining with gap constraints for peptide folding prediction. In *FLAIRS Conference*, pages 95–100, 2008. 97, 99

[179] Jingbo Shang, Meng Jiang, Wenzhu Tong, Jinfeng Xiao, Jian Peng, and Jiawei Han. DPPred: An effective prediction framework with concise discriminative patterns. *IEEE Transactions on Knowledge and Data Engineering*, 2017. DOI: 10.1109/tkde.2017.2757476 8, 25

[180] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:623–656, 1948. 11

[181] Richard Sherhod, Valerie J. Gillet, Philip N. Judson, and Jonathan D. Vessey. Automating knowledge discovery for toxicity prediction using jumping emerging pattern mining. *Journal of Chemical Information and Modeling*, 52(11):3074–3087, 2012. DOI: 10.1021/ci300254w 89, 98

[182] Richard Sherhod, Philip N. Judson, Thierry Hanser, Jonathan D. Vessey, Samuel J. Webb, and Valerie J. Gillet. Emerging pattern mining to aid toxicological knowledge discovery. *Journal of Chemical Information and Modeling*, 54(7):1864–1879, 2014. DOI: 10.1021/ci5001828 98

[183] Nicholas Skapura and Guozhu Dong. Class distribution curve based discretization with application to wearable sensors and medical monitoring. *International Journal of Monitoring and Surveillance Technologies Research (IJMSTR)*, 5(4):23–37, 2017. DOI: 10.4018/ijmstr.2017100102 11

[184] Hee Seok Song, Jae kyeong Kim, and Soung Hie Kim. Mining the change of customer behavior in an internet shopping mall. *Expert Systems with Applications*, 21(3):157–168, 2001. DOI: 10.1016/s0957-4174(01)00037-9 97

[185] Arnaud Soulet, Bruno Crémilleux, and François Rioult. Condensed representation of emerging patterns. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 127–132, Springer, 2004. DOI: 10.1007/978-3-540-24775-3_16 25

[186] Arnaud Soulet and C. Hébert. Using emerging patterns from clusters to characterize social subgroups of patients affected by atherosclerosis. In *Proc. of Discovery Challenge Workshop co-located with ECML/PKDD*, 2004. 2

[187] Matt Spencer, Nicole Takahashi, Sounak Chakraborty, Judith Miles, and Chi-Ren Shyu. Heritable genotype contrast mining reveals novel gene associations specific to autism subgroups. *Journal of Biomedical Informatics*, 77:50–61, 2018. DOI: 10.1016/j.jbi.2017.11.016 8, 96

[188] Michael Steinbach, Haoyu Yu, Gang Fang, and Vipin Kumar. Using constraints to generate and explore higher order discriminative patterns. In *Proc. of Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, pages 338–350, 2011. DOI: 10.1007/978-3-642-20841-6_28 8

[189] Qun Sun, Xiuzhen Zhang, and Kotagiri Ramamohanarao. Noise tolerance of ep-based classifiers. In *Australasian Joint Conference on Artificial Intelligence*, pages 796–806, Springer, 2003. DOI: 10.1007/978-3-540-24581-0_68 33

[190] Yanmin Sun, Andrew K. C. Wong, and Mohamed S. Kamel. Classification of imbalanced data: A review. *International Journal of Pattern Recognition and Artificial Intelligence*, 23(04):687–719, 2009. DOI: 10.1142/s0218001409007326 28

[191] Atsushi Takizawa. Classification and feature extraction of criminal occurrence points using CAEP with transductive clustering. *Procedia—Social and Behavioral Sciences*, 21:83–92, 2011. DOI: 10.1016/j.sbspro.2011.07.036 90

[192] Guanting Tang, Jian Pei, James Bailey, and Guozhu Dong. Mining multidimensional contextual outliers from categorical relational data. *Intelligent Data Analysis*, 19(5):1171–1192, 2015. DOI: 10.3233/ida-150764 50

[193] Vahid Taslimitehrani and Guozhu Dong. A new CPXR based logistic regression method and clinical prognostic modeling results using the method on traumatic brain injury. In *IEEE International Conference on BioInformatics and BioEngineering (BIBE)*, pages 283–290, 2014. DOI: 10.1109/bibe.2014.16 84, 99

[194] Vahid Taslimitehrani, Guozhu Dong, Naveen L. Pereira, Maryam Panahiazar, and Jyotishman Pathak. Developing EHR-driven heart failure risk prediction models using CPXR(Log) with the probabilistic loss function. *Journal of Biomedical Informatics*, 60:260–269, 2016. DOI: 10.1016/j.jbi.2016.01.009 84, 99

[195] Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, and Ali A. Ghorbani. A detailed analysis of the KDD CUP 99 data set. In *IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)*, pages 1–6, 2009. DOI: 10.1109/cisda.2009.5356528 48

[196] Paweł Terlecki. On the relation between jumping emerging patterns and rough set theory with application to data classification. In *Transactions on Rough Sets XII*, pages 236–338, Springer-Verlag, 2010. DOI: 10.1007/978-3-642-14467-7_13 8

[197] Pawel Terlecki and Krzysztof Walczak. Adaptive classification with jumping emerging patterns. In *Proc. of 3rd International Conference on Rough Sets and Knowledge Technology*, pages 39–46, 2008. DOI: 10.1007/978-3-540-79721-0_11 96

[198] Takahisa Toda. Hypergraph transversal computation with binary decision diagrams. In *International Symposium on Experimental Algorithms*, pages 91–102, Springer, 2013. DOI: 10.1007/978-3-642-38527-8_10 25

[199] George Tzanis, Ioannis Kavakiotis, and Ioannis Vlahavas. PolyA-iEP: A data mining method for the effective prediction of polyadenylation sites. *Expert Systems with Applications*, 38(10):12398–12408, 2011. DOI: 10.1016/j.eswa.2011.04.019 97, 99

[200] Joseph L. Usset, Rama Raghavan, Jonathan P. Tyrer, Valerie McGuire, Weiva Sieh, Penelope Webb, Jenny Chang-Claude, Anja Rudolph, Hoda Anton-Culver, Andrew Berchuck, et al. Assessment of multifactor gene—environment interactions and ovarian cancer risk: Candidate genes, obesity, and hormone-related risk factors. *Cancer Epidemiology and Prevention Biomarkers*, 25(5):780–790, 2016. DOI: 10.1158/1538-7445.am2015-4684 68

[201] Adriano Veloso, Wagner Meira Jr., and Mohammed J. Zaki. Lazy associative classification. In *6th International Conference on Data Mining (ICDM)*, pages 645–654, IEEE, 2006. DOI: 10.1109/icdm.2006.96 36

[202] Sebastián Ventura and José María Luna. *Supervised Descriptive Pattern Mining*. Springer, 2018. DOI: 10.1007/978-3-319-98140-6 8

[203] Winn Voravuthikunchai, Bruno Crémilleux, and Frédéric Jurie. Histograms of pattern sets for image classification and object recognition. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 224–231, 2014. DOI: 10.1109/cvpr.2014.36 97

[204] Jilles Vreeken, Matthijs Van Leeuwen, and Arno Siebes. KRIMP: Mining itemsets that compress. *Data Mining and Knowledge Discovery*, 23(1):169–214, 2011. DOI: 10.1007/s10618-010-0202-x 24

[205] Chunyu Wang, Yizhou Wang, and Alan L. Yuille. An approach to pose-based action recognition. In *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 915–922, 2013. DOI: 10.1109/cvpr.2013.123 91, 99

[206] Dawei Wang, Wei Ding, Henry Lo, Melissa Morabito, Ping Chen, Josue Salazar, and Tomasz Stepinski. Understanding the spatial distribution of crime based on its related variables using geospatial discriminative patterns. *Computers, Environment and Urban Systems*, 39:93–106, 2013. DOI: 10.1016/j.compenvurbsys.2013.01.008 8, 18

[207] Liang Wang, Yizhou Wang, Tingting Jiang, and Wen Gao. Instantly telling what happens in a video sequence using simple features. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3257–3264, 2011. DOI: 10.1109/cvpr.2011.5995377 93, 99

[208] Lusheng Wang, Hao Zhao, Guozhu Dong, and Jianping Li. On the complexity of finding emerging patterns. *Theoretical Computer Science*, 335(1):15–27, 2005. DOI: 10.1109/cmpsac.2004.1342691 25

[209] Xiaoting Wang. Trajectory mining in the context of the Internet of Things. Ph.D. thesis, University of Melbourne, 2017. 97

[210] Xiaoting Wang, Christopher Leckie, Hairuo Xie, and Tharshan Vaithianathan. Discovering the impact of urban traffic interventions using contrast mining on vehicle trajectory data. In *Pacific–Asia Conference on Knowledge Discovery and Data Mining*, pages 486–497, Springer, 2015. DOI: 10.1007/978-3-319-18038-0_38 97

[211] Zhou Wang, Hongjian Fan, and Kotagiri Ramamohanarao. Exploiting maximal emerging patterns for classification. In *Australasian Joint Conference on Artificial Intelligence*, pages 1062–1068, Springer, 2004. DOI: 10.1007/978-3-540-30549-1_102 95

[212] Geoffrey I. Webb, Shane M. Butler, and Douglas A. Newlands. On detecting differences between groups. In *Proc. of ACM Conference on Knowledge Discovery and Data Mining (KDD)*, pages 256–265, 2003. DOI: 10.1145/956755.956781 8

[213] Wei Wei, Jinjiu Li, Longbing Cao, Yuming Ou, and Jiahang Chen. Effective detection of sophisticated online banking fraud on extremely imbalanced data. *World Wide Web*, 16(4):449–475, 2013. DOI: 10.1007/s11280-012-0178-0 94

[214] Gary M. Weiss. The impact of small disjuncts on classifier learning. In *Data Mining*, pages 193–226, Springer, 2010. DOI: 10.1007/978-1-4419-1280-0_9 28

[215] Cheng-Hsiung Weng and Cheng-Kui Huang Tony. Observation of sales trends by mining emerging patterns in dynamic markets. *Applied Intelligence*, pages 1–15, 2018. DOI: 10.1007/s10489-018-1231-1 97

[216] Daniela M. Witten and Robert Tibshirani. A comparison of fold-change and the t-statistic for microarray data analysis. *Analysis*, 2007. 66

[217] Stefan Wrobel. An algorithm for multi-relational discovery of subgroups. In *Proc. of European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*, pages 78–87, 1997. DOI: 10.1007/3-540-63223-9_108 2

[218] Stefan Wrobel. Inductive logic programming for knowledge discovery in databases. In *Relational Data Mining*, pages 74–101, Springer, 2001. DOI: 10.1007/978-3-662-04599-2_4 8

[219] Baohua Wu, Lei Duan, Zhonghua Yu, Changjie Tang, and Jun Zhu. Birth defects detection algorithm based on emerging patterns. *Journal of Computer Applications*, 31(4):885–889, 2011. DOI: 10.3724/sp.j.1087.2011.00885 93, 99

[220] Dongkuan Xu and Yingjie Tian. A comprehensive survey of clustering algorithms. *Annals of Data Science*, 2(2):165–193, 2015. DOI: 10.1007/s40745-015-0040-1 58

[221] Rui Xu and Donald Wunsch. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678, 2005. DOI: 10.1109/tnn.2005.845141 51, 58

[222] Jingfeng Xue, Changzhen Hu, Kunsheng Wang, Rui Ma, and Jiaxin Zou. Metamorphic Malware detection technology based on aggregating emerging patterns. In *Proc. of International Conference Interaction Sciences*, pages 1293–1296, 2009. DOI: 10.1145/1655925.1656162 33

[223] S. Ben Yahia, Tarek Hamrouni, and E. Mephu Nguifo. Frequent closed itemset based algorithms: A thorough structural and analytical survey. *ACM SIGKDD Explorations Newsletter*, 8(1):93–104, 2006. DOI: 10.1145/1147234.1147248 13

[224] Yulan Yan, Chikara Hashimoto, Kentaro Torisawa, Takao Kawai, Jun'ichi Kazama, and Stijn De Saeger. Minimally supervised method for multilingual paraphrase extraction from definition sentences on the Web. In *Proc. of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 63–73, 2013. 97

[225] Eng-Juh Yeoh, Mary E. Ross, Sheila A. Shurtleff, W. Kent Williams, Divyen Patel, Rami Mahfouz, Fred G. Behm, Susana C. Raimondi, Mary V. Relling, Anami Patel, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell*, 1(2):133–143, 2002. DOI: 10.1016/s1535-6108(02)00032-6 89, 99

[226] Xiaoxin Yin and Jiawei Han.   CPAR: Classification based on predictive association rules. In *Proc. of SIAM International Conference on Data Mining (SDM)*, 2003. DOI: 10.1137/1.9781611972733.40 35

[227] Kui Yu, Wei Ding, Dan A. Simovici, Hao Wang, Jian Pei, and Xindong Wu. Classification with streaming features: An emerging-pattern mining approach. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 9(4):30, 2015. DOI: 10.1145/2700409 96

[228] Kui Yu, Xindong Wu, Wei Ding, Hao Wang, and Hongliang Yao. Causal associative classification. In *11th IEEE International Conference on Data Mining (ICDM)*, pages 914–923, 2011. DOI: 10.1109/icdm.2011.30 33, 35

[229] Reza Zafarani and Huan Liu. Social computing data repository at ASU, `http://soci alcomputing.asu.edu`, 2009. 63

[230] Shaoyi Zhang, Kotagiri Ramamohanarao, and James C. Bezdek.   EP-based robust weighting scheme for fuzzy SVMs. In *Proc. of the 21st Australasian Conference on Database Technologies*. 95

[231] Xiuzhen Zhang, Guozhu Dong, and Kotagiri Ramamohanarao. Information-based classification by aggregating emerging patterns. In *Proc. of Intelligent Data Engineering and Automated Learning (IDEAL)*, pages 48–53, 2000. DOI: 10.1007/3-540-44491-2_8 31, 35, 96

[232] Xiuzhen Zhang, Guozhu Dong, and Kotagiri Ramamohanarao.   Building behaviour knowledge space to make classification decision. In *Pacific–Asia Conference on Knowledge Discovery and Data Mining*, pages 488–494, Springer, 2001. DOI: 10.1007/3-540-45357-1_51 95

[233] Zhigang Zheng, Wei Wei, Chunming Liu, Wei Cao, Longbing Cao, and Maninder Bhatia. An effective contrast sequential pattern mining approach to taxpayer behavior analysis. *World Wide Web*, 19(4):633–651, 2016. DOI: 10.1007/s11280-015-0350-4 96

[234] Yun Zhou, James Bailey, Ioanna Ioannou, Sudanthi Wijewickrema, Stephen O'Leary, and Gregor Kennedy. Pattern-based real-time feedback for a temporal bone simulator. In *Proc. of the 19th ACM Symposium on Virtual Reality Software and Technology*, pages 7–16, 2013. DOI: 10.1145/2503713.2503728 33, 93, 99

[235] Xiaojin Zhu. Semi-supervised learning literature survey. *Computer Science*, 2(3):4, University of Wisconsin-Madison, 2006. 40