

Contrast pattern mining and its application for building robust classifiers

Kotagiri Ramamohanarao

Department of Computer Science and Software Engineering

The University of Melbourne

Kotagiri@unimelb.edu.au

Abstract:

The ability to distinguish, differentiate and contrast between different data sets is a key objective in data mining. Such ability can assist domain experts to understand their data and can help in building classification models. This presentation will introduce the techniques for contrasting data sets. It will also focus on some important real world applications that illustrate how contrast patterns can be applied effectively for building robust classifiers.

Useful references:

- 1) Arunasalam, Bavani and Chawla, Sanjay and Sun, Pei. Striking Two Birds with One Stone: Simultaneous Mining of Positive and Negative Spatial Patterns. In Proceedings of the Fifth SIAM International Conference on Data Mining, April 21-23, pp, Newport Beach, CA, USA, SIAM 2005
- 2) Eric Bae, James Bailey, Guozhu Dong: Clustering Similarity Comparison Using Density Profiles. Australian Conference on Artificial Intelligence 2006: 342-351
- 3) James Bailey, Thomas Manoukian, Kotagiri Ramamohanarao: Fast Algorithms for Mining Emerging Patterns. PKDD 2002: 39-50.
- 4) J. Bailey and T. Manoukian and K. Ramamohanarao: A Fast Algorithm for Computing Hypergraph Transversals and its Application in Mining Emerging Patterns. Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM). Pages 485-488. Florida, USA, November 2003.
- 5) Stephen D. Bay, Michael J. Pazzani: Detecting Change in Categorical Data: Mining Contrast Sets. KDD 1999: 302-306.
- 6) Guozhu Dong, Jinyan Li: Efficient Mining of Emerging Patterns: Discovering Trends and Differences. KDD 1999: 43-52.
- 7) Guozhu Dong, Jinyan Li: Mining border descriptions of emerging patterns from dataset pairs. Knowl. Inf. Syst. 8(2): 178-202 (2005).

- 8) Xiaonan Ji, James Bailey, Guozhu Dong: Mining Minimal Distinguishing Subsequence Patterns with Gap Constraints. ICDM 2005: 194-201.
- 9) Xiaonan Ji, James Bailey, Guozhu Dong: Mining Minimal Distinguishing Subsequence Patterns with Gap Constraints. Knowl. Inf. Syst. 11(3): 259--286 (2007).
- 10) Haiquan Li, Jinyan Li, Limsoon Wong, Mengling Feng, Yap-Peng Tan: Relative risk and odds ratio: a data mining perspective. PODS 2005: 368-377
- 11) Jinyan Li, Guimei Liu and Limsoon Wong. Mining Statistically Important Equivalence Classes and delta-Discriminative Emerging Patterns. KDD 2007.
- 12) Jinyan Li, Thomas Manoukian, Guozhu Dong, Kotagiri Ramamohanarao: Incremental Maintenance on the Border of the Space of Emerging Patterns. Data Min. Knowl. Discov. 9(1): 89-116 (2004).

- 13) Jinyan Li and Qiang Yang. Strong Compound-Risk Factors: Efficient Discovery through Emerging Patterns and Contrast Sets. IEEE Transactions on Information Technology in Biomedicine. To appear.
- 14) Elsa Loekito, James Bailey: Fast Mining of High Dimensional Expressive Contrast Patterns Using Zero-suppressed Binary Decision Diagrams. KDD 2006: 307-316.
- 15) Hamad Alhammady, Kotagiri Ramamohanarao: The Application of Emerging Patterns for Improving the Quality of Rare-Class Classification. PAKDD 2004: 207-211
- 16) Hamad Alhammady, Kotagiri Ramamohanarao: Using Emerging Patterns and Decision Trees in Rare-Class Classification. ICDM 2004: 315-318
- 17) Hamad Alhammady, Kotagiri Ramamohanarao: Expanding the Training Data Space Using Emerging Patterns and Genetic Methods. SDM 2005
- 18) Hamad Alhammady, Kotagiri Ramamohanarao: Using Emerging Patterns to Construct Weighted Decision Trees. IEEE Trans. Knowl. Data Eng. 19) 18(7): 865-876 (2006).
- 20) Hamad Alhammady, Kotagiri Ramamohanarao: Mining Emerging Patterns and Classification in Data Streams. Web Intelligence 2005: 272-275
- 21) James Bailey, Thomas Manoukian, Kotagiri Ramamohanarao: Classification Using Constrained Emerging Patterns. WAIM 2003: 226-237
- 22) Guozhu Dong, Xiuzhen Zhang, Limsoon Wong, Jinyan Li: CAEP: Classification by Aggregating Emerging Patterns. Discovery Science 1999: 30-42.
- 23) Hongjian Fan, Kotagiri Ramamohanarao: An Efficient Single-Scan Algorithm for Mining Essential Jumping Emerging Patterns for Classification. PAKDD 2002: 456-462
- 24) Hongjian Fan, Kotagiri Ramamohanarao: Efficiently Mining Interesting Emerging Patterns. WAIM 2003: 189-201

- 25) Hongjian Fan, Kotagiri Ramamohanarao: Noise Tolerant Classification by Chi Emerging Patterns. PAKDD 2004: 201-206
- 26) Hongjian Fan, Ming Fan, Kotagiri Ramamohanarao, Mengxu Liu: Further Improving Emerging Pattern Based Classifiers Via Bagging. PAKDD 2006: 91-96
- 27) Hongjian Fan, Kotagiri Ramamohanarao: A weighting scheme based on emerging patterns for weighted support vector machines. GrC 2005: 435-440
- 28) Hongjian Fan, Kotagiri Ramamohanarao: Fast Discovery and the Generalization of Strong Jumping Emerging Patterns for Building Compact and Accurate Classifiers. IEEE Trans. Knowl. Data Eng. 18(6): 721-737 (2006)
- 29) Jinyan Li, Guozhu Dong, Kotagiri Ramamohanarao: Instance-Based Classification by Emerging Patterns. PKDD 2000: 191-200
- 30) Jinyan Li, Guozhu Dong, Kotagiri Ramamohanarao: Making Use of the Most Expressive Jumping Emerging Patterns for Classification. PAKDD 2000: 220-232
- 31) Jinyan Li, Guozhu Dong, Kotagiri Ramamohanarao: Making Use of the Most Expressive Jumping Emerging Patterns for Classification. Knowl. Inf. Syst. 3 (2): 131-145 (2001)
- 32) Jinyan Li, Kotagiri Ramamohanarao, Guozhu Dong: Emerging Patterns and Classification. ASIAN 2000: 15-32
- 33) Jinyan Li, Guozhu Dong, Kotagiri Ramamohanarao, Limsoon Wong: DeEPs: A New Instance-Based Lazy Discovery and Classification System. Machine Learning 54(2): 99-124 (2004).
- 34) Wenmin Li, Jiawei Han, Jian Pei: CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules. ICDM 2001: 369-376
- 35) Jinyan Li, Kotagiri Ramamohanarao, Guozhu Dong: Combining the Strength of Pattern Frequency and Distance for Classification. PAKDD 2001: 455-466
- 36) Bing Liu, Wynne Hsu, Yiming Ma: Integrating Classification and Association Rule Mining. KDD 1998: 80-86
- 37) Kotagiri Ramamohanarao, James Bailey: Discovery of Emerging Patterns and Their Use in Classification. Australian Conference on Artificial Intelligence 2003: 1-12
- 38) Ramamohanarao, K. and Bailey, J. and Fan, H. Efficient Mining of Contrast Patterns and Their Applications to Classification, Third International Conference on Intelligent Sensing and Information Processing, 2005 (39--47).
- 39) Ramamohanarao, K. and Fan, H. Patterns Based Classifiers, World Wide Web 2007: 10(71--83).

- 40) Qun Sun, Xiuzhen Zhang, Kotagiri Ramamohanarao: Noise Tolerance of EP-Based Classifiers. Australian Conference on Artificial Intelligence 2003: 796-806
- 41) Xiuzhen Zhang, Guozhu Dong, Kotagiri Ramamohanarao: Information-Based Classification by Aggregating Emerging Patterns. IDEAL 2000: 48-53
- 42) Xiuzhen Zhang, Guozhu Dong, Kotagiri Ramamohanarao: Building Behaviour Knowledge Space to Make Classification Decision. PAKDD 2001: 488-494
- 43) Zhou Wang, Hongjian Fan, Kotagiri Ramamohanarao: Exploiting Maximal Emerging Patterns for Classification. Australian Conference on Artificial Intelligence 2004: 1062-1068