

Data mining on very large databases

Bibliography

Main-memory approaches

[Goe04] B. Goethals. Memory issues in frequent itemset mining. In: ACM Symposium on Applied Computing, 2004.

[Vaa04] R. Vaarandi. A breadth-first algorithm for mining frequent patterns from event logs. In Intelligence in Communication Systems, 2004.

[Han00] J. Han, J. Pei, and Y. Yin. Mining frequent patterns without candidate generation. In ACM SIGMOD, 2000.

[Pie03] A. Pietracaprina and D. Zandolin. Mining frequent itemsets using patricia tries. In FIMI '03, November 2003.

[Agr94] R. Agrawal and R. Srikant. Fast algorithm for mining association rules. In VLDB '94, September. 1994.

[Zak00] M. Zaki. Scalable algorithms for association mining. IEEE Trans. In Knowledge Discovery and Data Engineering, vol. 12, no. 3, pp. 372--390, June 2000.

[Uno04] T. Uno, M. Kiyomi, and H. Arimura. LCM ver. 2: Efficient mining algorithms for frequent/closed/maximal itemsets. In FIMI '04, 2004.

[Gra03] G. Grahne and J. Zhu. Efficiently using prefix-trees in mining frequent itemsets. In FIMI '03, November 2003.

[El-04] El-Hajj, M. and Zaïane, O. R. 2004. COFI approach for mining frequent itemsets revisited. In ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (Paris, France). DMKD '04.

[Uno05] Uno, T., Kiyomi, M., and Arimura, H. 2005. LCM ver.3: collaboration of array, bitmap and prefix tree for frequent itemset mining. In ACM OSDM '05, Chicago, Illinois, August 21 - 21, 2005.

[Ino00]A. Inokuchi, T. Washio, and H. Motoda. An Apriori-based algorithm for mining frequent substructures from graph data. In PKDD 2000, pages 13-23, Lyon, France, Sept. 2000.

[Kur01] M. Kuramochi and G. Karypis. Frequent subgraph discovery. In ICDM 2001, pages 313-320, San Jose, CA, Nov. 2001.

[Yan02] Y. Yan and J. Han. gSpan: Graph-based substructure pattern mining. In ICDM 2002, Maebashi, Japan, December 2002.

[Zak02] M.J. Zaki. Efficiently mining frequent trees in a forest. In ACM SIGKDD Conference '02, Edmonton, Alberta, Canada, July 2002.

[Asa02] T. Asai, K. Abe, S. Kawasoe, H. Arimura, H. Satamoto, and S. Arikawa. Efficient substructure discovery from large semi-structured data. In SIAM 2002, April 2002.

[Bas06] S. Bashir and A. Rauf Baig. Performance Analysis of Frequent Itemset Mining Using Hybrid Database Representation Approach. In IEEE INMIC '06, pp.237-243, 23-24 Dec. 2006

[Lil03] G. Liu, H. Lu, J.X. Yu, W. Wei, and X. Xiao. AFOPT: A efficient implementation of pattern growth approach. In FIMI '03, November 2003.

[Par95] J. S. Park, M.-S. Chen, P. S. Yu. An Effective Hash Based Algorithm for Mining Association Rules. SIGMOD Conference 1995: 175-186

[Sav95] A. Savasere, E. Omiecinski, and S. Navathe. An efficient algorithm for mining association rules in large databases. In VLDB '95, 1995.

[Toi96] H. Toivonen: Sampling Large Databases for Association Rules. VLDB 1996: 134-145. 1996

[Bri97] S. Brin, R. Motwani, J.D. Ullman, and S. Tsur. Dynamic itemset counting and implication rules for market basket data. SIGMOD Record, Vol 6, No 2, ACM Press, pp 255-264. 1997

[Sil95] A. Silberschatz, A. Tuzhilin. On Subjective Measures of Interestingness in Knowledge Discovery. KDD 1995: 275-281. 1995

[Boul03] J.-F. Boulicaut, A. Bykowski, C. Rigitti, "Free-sets: A condensed representation of Boolean data for the approximation of frequent queries", Data Mining and Knowledge Discovery journal 2003.

[Man96] H. Mannila, H. Toivonen, "Multiple uses of frequent sets and condensed representations", in Int. ACM Conf. on Knowledge Discovery and Data Mining (KDD), 1996.

[Bay98] R. J. Bayardo Jr. Efficiently Mining Long Patterns from Databases. SIGMOD Conference 1998: 85-93. 1998

[Zak+02] M. J. Zaki, C.-J. Hsiao. CHARM: An Efficient Algorithm for Closed Itemset Mining. SDM 2002

[Pas99] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal. Discovering Frequent Closed Itemsets for Association Rules. In 7th international Conference on Database theory (January 10 - 12, 1999). Springer-Verlag, London, 398-416. 1999

Disk-Based approaches

[El03] M. El-Hajj and O. R. Zaiane. Inverted matrix: Efficient discovery of frequent items in large datasets in the context of interactive mining. In ACM SIGKDD Conference '03, 2003.

[Gra04] G. Grahne and J. Zhu. Mining frequent itemsets from secondary memory. In ICDM '04, pp. 91–98.

[Ram02] G. Ramesh, W. Maniatty, and M. Zaki. Indexing and data access methods for database mining. In DMKD '02, 2002.

[Bue06] G. Buehrer, S. Parthasarathy, and A. Ghoting. Out-of-core frequent pattern mining on a commodity PC. In ACM SIGKDD Conference '06. August 20 - 23, 2006.

[Bar05] E. Baralis, T. Cerquitelli, S. Chiusano. Index Support for Frequent Itemset Mining in a Relational DBMS. In IEEE ICDE '05: 754-765, 2005.

[Bar09] E. Baralis, T. Cerquitelli, S. Chiusano, "IMine: Index Support for Item Set Mining," IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 4, pp. 493-506, 2009.

[Adn09] M. Adnan and R. Alhajj. DRFP-tree: disk-resident frequent pattern tree. In Journal on Applied Intelligence Springer. Volume 30, Number 2, Pages 84-97. April, 2009

[Wan04] Wang, C., Wang, W., Pei, J., Zhu, Y., and Shi, B. 2004. Scalable mining of large disk-based graph databases. In ACM SIGKDD Conference '04, Seattle, August 22 - 25, 2004.

[Wan05] Wang, W., Wang, C., Zhu, Y., Shi, B., Pei, J., Yan, X., and Han, J. 2005. GraphMiner: a structural pattern-mining system for large disk-based graph databases and its applications. In ACM SIGMOD '05, 879-881, 2005

[Pho07] Phoophakdee, B. and Zaki, M. J. 2007. Genome-scale disk-based suffix tree indexing. In ACM SIGMOD '07, Beijing, China, June 11 - 14, 2007.

[Tat04] S. Tata, R. Hankins, and J. Patel. Practical suffix tree construction. In VLDB '04, 2004.

[Tia05] Y. Tian, S. Tata, R. Hankins, and J. Patel. Practical methods for constructing suffix trees. VLDB Journal, 14(3):281–299, 2005.

[Liu03] G. Liu, H. Lu, W. Lou, and J. X. Yu. On computing, storing and querying frequent patterns. In ACM SIGKDD Conference '03, Washington, August 24 - 27, 2003.

[Liu07] G. Liu, H. Lu, and J. Xu Yu. CFP-tree: A compact disk-based structure for storing and querying frequent itemsets. Inf. Syst. 32, 2295-319, Apr. 2007.

[Mor98] T. Morzy and M. Zakrzewicz. Group bitmap index: a structure for association rules retrieval. In ACM SIGKDD Conference '98, pp. 284–288, 1998.

[Tuz02] A. Tuzhilin and B. Liu. Querying multiple sets of discovered rules. In ACM SIGKDD Conference '02, pp. 52–60, 2002.

Incremental mining of frequent patterns

[Adn06] M. Adnan, R. Alhajj, K. Barker. Alternative method for incrementally constructing the FP-tree. In IEEE International Conference on Intelligent Systems, UK, September 2006

[Che03] W. Cheung, O. R. Zaiane. Incremental Mining of Frequent Patterns without Candidate Generation or Support Constraint. In IDEAS'03, 2003.

[Leu05] C.K.-S. Leung, Q.I. Khan, T. Hoque. CanTree: a tree structure for efficient incremental mining of frequent patterns. In IEEE Data Mining Conference '05, 27-30 Nov. 2005

[HeZ07] H.-T. He, S.-L. Zhang. A New Method for Incremental Updating Frequent Patterns Mining. In ICICIC '07, pp.561-561, 5-7 Sept. 2007

[Bar08] E. Baralis, T. Cerquitelli, S. Chiusano. Indexing Evolving Databases for Itemset Mining. In "Intelligent Techniques and Tools for Novel System Architectures" book in the series of "Studies in Computational Intelligence", Chountas Panagiotis, Petrounias Ilias, Kacprzyk Janusz, Springer, pp. 305-323, 2008, Vol. 109/2008.

Mining languages

[Han96] J. Han, Y. Fu, W. Wang, K. Koperski, and O. Zaiane. DMQL: A data mining query language for relational databases. In DMKD, 1996.

[Meo96] R. Meo, G. Psaila, and S. Ceri. A new SQL-like operator for mining association rules. In VLDB, 1996.

[Tuz02] A. Tuzhilin and B. Liu, Querying multiple sets of discovered rules. In ACM SIGKDD Conference '02, pp. 52–60, 2002.