

# Data Mining: Concepts and Techniques (2nd edition)

Jiawei Han and Micheline Kamber  
Morgan Kaufmann Publishers, 2006

## Bibliographic Notes for Chapter 8 Mining Stream, Time-Series, and Sequence Data

Stream data mining research has been active in recent years. Popular surveys on stream data systems and stream data processing include Babu and Widom [BW01], Babcock, Babu, Datar, et al. [BBD<sup>+</sup>02], Muthukrishnan [Mut03], and the tutorial by Garofalakis, Gehrke, and Rastogi [GGR02].

There have been extensive studies on stream data management and the processing of continuous queries in stream data. For a description of synopsis data structures for stream data, see Gibbons and Matias [GM98]. Vitter introduced the notion of reservoir sampling as a way to select an unbiased random sample of  $n$  elements without replacement from a larger ordered set of size  $N$ , where  $N$  is unknown [Vit85]. Stream query or aggregate processing methods have been proposed by Chandrasekaran and Franklin [CF02], Gehrke, Korn, and Srivastava [GKS01], Dobra, Garofalakis, Gehrke, and Rastogi [DGGR02], and Madden, Shah, Hellerstein, and Raman [MSHR02]. A one-pass summary method for processing approximate aggregate queries using wavelets was proposed by Gilbert, Kotidis, Muthukrishnan, and Strauss [GKMS01]. Statstream, a statistical method for the monitoring of thousands of data streams in real time, was developed by Zhu and Shasha [ZS02, SZ04].

There are also many stream data projects. Examples include Aurora by Zdonik, Cetintemel, Cherniack, et al. [ZCC<sup>+</sup>02], which is targeted towards stream monitoring applications; STREAM, developed at Stanford University by Babcock, Babu, Datar, et al. aims at developing a general purpose Data Stream Management System (DSMS) [BBD<sup>+</sup>02]; and an early system called Tapestry by Terry, Goldberg, Nichols, and Oki [TGNO92], which used continuous queries for content-based filtering over an append-only database of email and bulletin board messages. A restricted subset of SQL was used as the query language in order to provide guarantees about efficient evaluation and append-only query results.

A multidimensional stream cube model was proposed by Chen, Dong, Han, et al. [CDH<sup>+</sup>02] in their study of multidimensional regression analysis of time-series data streams. MAIDS (Mining Alarming Incidents from Data Streams), a stream data mining system built on top of such a stream data cube, was developed by Cai, Clutter, Pape, et al. [CCP<sup>+</sup>04].

For mining frequent items and itemsets on stream data, Manku and Motwani proposed sticky sampling and lossy counting algorithms for approximate frequency counts over data streams [MM02]. Karp, Papadimitriou and Shenker proposed a counting algorithm for finding frequent elements in data streams [KPS03]. Giannella, Han, Pei, et al. proposed a method for mining frequent patterns in data streams at multiple time granularities [GHP<sup>+</sup>04].

For stream data classification, Domingos and Hulten proposed the VFDT algorithm, based on their Hoeffding tree algorithm [DH00]. CVFDT, a later version of VFDT, was developed by Hulten, Spencer, and Domingos [HSD01] to handle concept drift in time-changing data streams. Wang, Fan, Yu and Han proposed an ensemble classifier to mine concept-drifting data streams [WFYH03]. Aggarwal, Han, Wang and Yu developed a  $k$ -nearest neighbor-based method for classify evolving data streams [AHWY04b].

Several methods have been proposed for clustering data streams. The  $k$ -median based STREAM algorithm was proposed by Guha, Mishra, Motwani, and O'Callaghan [GMMO00] and by O'Callaghan, Mishra, Meyerson, et al. [OMM<sup>+</sup>02]. Aggarwal, Han, Wang, and Yu proposed CluStream, a framework for clustering evolving data streams [AHWY03], and HPStream, a framework for projected clustering of high dimensional data streams [AHWY04a].

Statistical methods for time-series analysis have been proposed and studied extensively in statistics, such as in Chatfield [Cha03], Brockwell and Davis [BD02], and Shumway and Stoffer [SS05]. StatSoft's Electronic Textbook ([www.statsoft.com/textbook/stathome.html](http://www.statsoft.com/textbook/stathome.html)) is a useful online resource that includes a discussion on time-series data

analysis. The ARIMA forecasting method is described in Box, Jenkins, and Reinsel [BJR94]. Efficient similarity search in sequence databases was studied by Agrawal, Faloutsos, and Swami [AFS93]. A fast subsequence matching method in time-series databases was presented by Faloutsos, Ranganathan, and Manolopoulos [FRM94]. Agrawal, Lin, Sawhney, and Shim [ALSS95] developed a method for fast similarity search in the presence of noise, scaling, and translation in time-series databases. Language primitives for querying shapes of histories were proposed by Agrawal, Psaila, Wimmers, and Zait [APWZ95]. Other work on similarity-based search of time-series data includes Raffei and Mendelzon [RM97], and Yi, Jagadish, and Faloutsos [YJF98]. Yi, Sidiropoulos, Johnson, Jagadish, et al. [YSJ<sup>+</sup>00] introduced a method for on-line mining for coevolving time sequences. Chen, Dong, Han, et al. [CDH<sup>+</sup>02] proposed a multidimensional regression method for analysis of multidimensional time-series data. Shasha and Zhu present a state-of-the-art overview of the methods for high performance discovery in time series [SZ04].

The problem of mining sequential patterns was first proposed by Agrawal and Srikant [AS95]. In the Apriori-based GSP algorithm, Srikant and Agrawal [SA96] generalized their earlier notion to include time constraints, a sliding time window, and user-defined taxonomies. Zaki [Zak01] developed a vertical format-based sequential pattern mining method called SPADE, which is an extension of vertical-format-based frequent itemset mining methods, like Eclat and Charm [Zak98, ZH02]. PrefixSpan, a pattern growth approach to sequential pattern mining, and its predecessor, FreeSpan, were developed by Pei, Han, Mortazavi-Asl, et al. [HPMA<sup>+</sup>00, PHMA<sup>+</sup>01, PHMA<sup>+</sup>04]. The CloSpan algorithm for mining closed sequential patterns was proposed by Yan, Han, and Afshar [YHA03]. BIDE, a bidirectional search for mining frequent closed sequences was developed by Wang and Han [WH04].

The studies of sequential pattern mining have been extended in several different ways. Mannila, Toivonen, and Verkamo [MTV97] consider frequent episodes in sequences, where episodes are essentially acyclic graphs of events whose edges specify the temporal before-and-after relationship but without timing-interval restrictions. Sequence pattern mining for plan failures was proposed in Zaki, Lesh, and Ogihara [ZLO98]. Garofalakis, Rastogi, and Shim [GRS99] proposed the use of regular expressions as a flexible constraint specification tool that enables user-controlled focus to be incorporated into the sequential pattern mining process. The embedding of multidimensional, multilevel information into a transformed sequence database for sequential pattern mining was proposed by Pinto, Han, Pei, et al. [PHP<sup>+</sup>01]. Pei, Han, and Wang studied issues regarding constraint-based sequential pattern mining [PHW02]. CLUSEQ is a sequence clustering algorithm, developed by Yang and Wang [YW03]. An incremental sequential pattern mining algorithm, IncSpan, was proposed by Cheng, Yan, and Han [CYH04]. SeqIndex, efficient sequence indexing by frequent and discriminative analysis of sequential patterns, was studied by Cheng, Yan, and Han [CYH05]. A method for parallel mining of closed sequential patterns was proposed by Cong, Han, and Padua [CHP05].

Data mining for periodicity analysis has been an interesting theme in data mining. Özden, Ramaswamy, and Silberschatz [ORS98] studied methods for mining periodic or cyclic association rules. Lu, Han, and Feng [LHF98] proposed intertransaction association rules, which are implication rules whose two sides are totally ordered episodes with timing-interval restrictions (on the events in the episodes and on the two sides). Bettini, Wang, and Jajodia [BWJ98] consider a generalization of intertransaction association rules. The notion of mining partial periodicity was first proposed by Han, Dong, and Yin, together with a max-subpattern hit set method [HDY99]. Ma and Hellerstein [MH01] proposed a method for mining partially periodic event patterns with unknown periods. Yang, Wang, and Yu studied mining asynchronous periodic patterns in time-series data [YWY03].

Methods for the analysis of biological sequences have been introduced in many textbooks, such as Waterman [Wat95], Setubal and Meidanis [SM97], Durbin, Eddy, Krogh and Mitchison [DEKM98], Baldi and Brunak [BB01], Krane and Raymer [KR03], Jones and Pevzner [JP04], and Baxevanis and Ouellette [BO04]. BLAST was developed by Altschul, Gish, Miller, et al. [AGM<sup>+</sup>90]. Information about BLAST can be found at the NCBI Web site [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/). For a systematic introduction of the BLAST algorithms and usages, see the book “BLAST” by Korf, Yandell, and Bedell [KYB03].

For an introduction to Markov chains and hidden Markov models from a biological sequence perspective, see Durbin, Eddy, Krogh, and Mitchison [DEKM98], and Jones and Pevzner [JP04]. A general introduction can be found in Rabiner [Rab89]. Eddy and Krogh have each respectively headed the development of software packages for hidden Markov models for protein sequence analysis, namely HMMER (pronounced “hammer”, available at <http://hmmerr.wustl.edu/>) and SAM ([www.cse.ucsc.edu/research/compbio/sam.html](http://www.cse.ucsc.edu/research/compbio/sam.html)).

# Bibliography

- [AFS93] R. Agrawal, C. Faloutsos, and A. Swami. Efficient similarity search in sequence databases. In *Proc. 4th Int. Conf. Foundations of Data Organization and Algorithms*, pages 69–84, Chicago, IL, Oct. 1993.
- [AHWY03] C. C. Aggarwal, J. Han, J. Wang, and P. S. Yu. A framework for clustering evolving data streams. In *Proc. 2003 Int. Conf. Very Large Data Bases (VLDB'03)*, pages 81–92, Berlin, Germany, Sept. 2003.
- [AHWY04a] C. Aggarwal, J. Han, J. Wang, and P. S. Yu. A framework for projected clustering of high dimensional data streams. In *Proc. 2004 Int. Conf. Very Large Data Bases (VLDB'04)*, pages 852–863, Toronto, Canada, Aug. 2004.
- [AHWY04b] C. Aggarwal, J. Han, J. Wang, and P. S. Yu. On demand classification of data streams. In *Proc. 2004 ACM SIGKDD Int. Conf. Knowledge Discovery in Databases (KDD'04)*, pages 503–508, Seattle, WA, Aug. 2004.
- [ALSS95] R. Agrawal, K.-I. Lin, H. S. Sawhney, and K. Shim. Fast similarity search in the presence of noise, scaling, and translation in time-series databases. In *Proc. 1995 Int. Conf. Very Large Data Bases (VLDB'95)*, pages 490–501, Zurich, Switzerland, Sept. 1995.
- [APWZ95] R. Agrawal, G. Psaila, E. L. Wimmers, and M. Zait. Querying shapes of histories. In *Proc. 1995 Int. Conf. Very Large Data Bases (VLDB'95)*, pages 502–514, Zurich, Switzerland, Sept. 1995.
- [AS95] R. Agrawal and R. Srikant. Mining sequential patterns. In *Proc. 1995 Int. Conf. Data Engineering (ICDE'95)*, pages 3–14, Taipei, Taiwan, Mar. 1995.
- [BB01] P. Baldi and S. Brunak. *Bioinformatics: The Machine Learning Approach* (2nd ed.). MIT Press, 2001.
- [BBD<sup>+</sup>02] B. Babcock, S. Babu, M. Datar, R. Motwani, and J. Widom. Models and issues in data stream systems. In *Proc. 2002 ACM Symp. Principles of Database Systems (PODS'02)*, pages 1–16, Madison, WI, June 2002.
- [BD02] P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting* (2nd ed.). Springer, 2002.
- [BJR94] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time Series Analysis: Forecasting and Control* (3rd ed.). Prentice-Hall, 1994.
- [BO04] A. Baxevasis and B. F. F. Ouellette. *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins* (3rd ed.). John Wiley & Sons, 2004.
- [BW01] S. Babu and J. Widom. Continuous queries over data streams. *SIGMOD Record*, 30:109–120, 2001.
- [BWJ98] C. Bettini, X. Sean Wang, and S. Jajodia. Mining temporal relationships with multiple granularities in time sequences. *Bull. Technical Committee on Data Engineering*, 21:32–38, 1998.

- [CCP<sup>+</sup>04] Y. D. Cai, D. Clutter, G. Pape, J. Han, M. Welge, and L. Auvil. MAIDS: Mining alarming incidents from data streams. In *Proc. 2004 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'04)*, pages 919–920, Paris, France, June 2004.
- [CDH<sup>+</sup>02] Y. Chen, G. Dong, J. Han, B. W. Wah, and J. Wang. Multi-dimensional regression analysis of time-series data streams. In *Proc. 2002 Int. Conf. Very Large Data Bases (VLDB'02)*, pages 323–334, Hong Kong, China, Aug. 2002.
- [CF02] S. Chandrasekaran and M. Franklin. Streaming queries over streaming data. In *Proc. 2002 Int. Conf. Very Large Data Bases (VLDB'02)*, pages 203–214, Hong Kong, China, Aug. 2002.
- [Cha03] C. Chatfield. *The Analysis of Time Series: An Introduction* (6th ed.). Chapman and Hall, 2003.
- [CHP05] S. Cong, J. Han, and D. Padua. Parallel mining of closed sequential patterns. In *Proc. 2005 ACM SIGKDD Int. Conf. Knowledge Discovery in Databases (KDD'05)*, pages 562–567, Chicago, IL, Aug. 2005.
- [CYH04] H. Cheng, X. Yan, and J. Han. IncSpan: Incremental mining of sequential patterns in large database. In *Proc. 2004 ACM SIGKDD Int. Conf. Knowledge Discovery in Databases (KDD'04)*, pages 527–532, Seattle, WA, Aug. 2004.
- [CYH05] H. Cheng, X. Yan, and J. Han. Seqindex: Indexing sequences by sequential pattern analysis. In *Proc. 2005 SIAM Int. Conf. Data Mining (SDM'05)*, pages 601–605, Newport Beach, CA, April 2005.
- [DEKM98] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison. *Biological Sequence Analysis: Probability Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998.
- [DGGR02] A. Dobra, M. Garofalakis, J. Gehrke, and R. Rastogi. Processing complex aggregate queries over data streams. In *Proc. 2002 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'02)*, pages 61–72, Madison, WI, June 2002.
- [DH00] P. Domingos and G. Hulten. Mining high-speed data streams. In *Proc. 2000 ACM SIGKDD Int. Conf. Knowledge Discovery in Databases (KDD'00)*, pages 71–80, Boston, MA, Aug. 2000.
- [FRM94] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos. Fast subsequence matching in time-series databases. In *Proc. 1994 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'94)*, pages 419–429, Minneapolis, MN, May 1994.
- [GGR02] M. Garofalakis, J. Gehrke, and R. Rastogi. Querying and mining data streams: You only get one look (a tutorial). In *Proc. 2002 ACM-SIGMOD Int. Conf. on Management of Data (SIGMOD'02)*, page 635, Madison, WI, June 2002.
- [GHP<sup>+</sup>04] C. Giannella, J. Han, J. Pei, X. Yan, and P. S. Yu. Mining frequent patterns in data streams at multiple time granularities. In H. Kargupta, A. Joshi, K. Sivakumar, and Y. Yesha, editors, *Data Mining: Next Generation Challenges and Future Directions*. AAAI/MIT Press, 2004.
- [GKMS01] A. C. Gilbert, Y. Kotidis, S. Muthukrishnan, and M. Strauss. Surfing wavelets on streams: One-pass summaries for approximate aggregate queries. In *Proc. 2001 Int. Conf. on Very Large Data Bases (VLDB'01)*, pages 79–88, Rome, Italy, Sept. 2001.
- [GKS01] J. Gehrke, F. Korn, and D. Srivastava. On computing correlated aggregates over continuous data streams. In *Proc. 2001 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'01)*, pages 13–24, Santa Barbara, CA, May 2001.
- [GM98] P. B. Gibbons and Y. Matias. New sampling-based summary statistics for improving approximate query answers. In *Proc. 1998 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'98)*, pages 331–342, Seattle, WA, June 1998.
- [GMMO00] S. Guha, N. Mishra, R. Motwani, and L. O'Callaghan. Clustering data streams. In *Proc. 2000 Symp. Foundations of Computer Science (FOCS'00)*, pages 359–366, Redondo Beach, CA, 2000.

- [GRS99] M. Garofalakis, R. Rastogi, and K. Shim. SPIRIT: Sequential pattern mining with regular expression constraints. In *Proc. 1999 Int. Conf. Very Large Data Bases (VLDB'99)*, pages 223–234, Edinburgh, UK, Sept. 1999.
- [HDY99] J. Han, G. Dong, and Y. Yin. Efficient mining of partial periodic patterns in time series database. In *Proc. 1999 Int. Conf. Data Engineering (ICDE'99)*, pages 106–115, Sydney, Australia, April 1999.
- [HPMA<sup>+</sup>00] J. Han, J. Pei, B. Mortazavi-Asl, Q. Chen, U. Dayal, and M.-C. Hsu. FreeSpan: Frequent pattern-projected sequential pattern mining. In *Proc. 2000 ACM SIGKDD Int. Conf. Knowledge Discovery in Databases (KDD'00)*, pages 355–359, Boston, MA, Aug. 2000.
- [HSD01] G. Hulten, L. Spencer, and P. Domingos. Mining time-changing data streams. In *Proc. 2001 ACM SIGKDD Int. Conf. Knowledge Discovery in Databases (KDD'01)*, San Fransisco, CA, Aug. 2001.
- [JP04] N. C. Jones and P. A. Pevzner. *An Introduction to Bioinformatics Algorithms*. MIT Press, 2004.
- [KPS03] R. M. Karp, C. H. Papadimitriou, and S. Shenker. A simple algorithm for finding frequent elements in streams and bags. *ACM Trans. Database Systems*, 28, 2003.
- [KR03] D. Krane and R. Raymer. *Fundamental Concepts of Bioinformatics*. Benjamin Cummings, 2003.
- [KYB03] I. Korf, M. Yandell, and J. Bedell. *BLAST*. O'Reilly, 2003.
- [LHF98] H. Lu, J. Han, and L. Feng. Stock movement and n-dimensional inter-transaction association rules. In *Proc. 1998 SIGMOD Workshop Research Issues on Data Mining and Knowledge Discovery (DMKD'98)*, pages 12:1–12:7, Seattle, WA, June 1998.
- [MH01] S. Ma and J. L. Hellerstein. Mining partially periodic event patterns with unknown periods. In *Proc. 2001 Int. Conf. Data Engineering (ICDE'01)*, pages 205–214, Heidelberg, Germany, April 2001.
- [MM02] G. Manku and R. Motwani. Approximate frequency counts over data streams. In *Proc. 2002 Int. Conf. Very Large Data Bases (VLDB'02)*, pages 346–357, Hong Kong, China, Aug. 2002.
- [MSHR02] S. Madden, M. Shah, J. M. Hellerstein, and V. Raman. Continuously adaptive continuous queries over streams. In *Proc. 2002 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'02)*, Madison, WI, June 2002.
- [MTV97] H. Mannila, H. Toivonen, and A. I. Verkamo. Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery*, 1:259–289, 1997.
- [Mut03] S. Muthukrishnan. Data streams: algorithms and applications. In *Proc. 2003 Annual ACM-SIAM Symp. Discrete Algorithms (SODA'03)*, pages 413–413, Baltimore, MD, Jan. 2003.
- [OMM<sup>+</sup>02] L. O'Callaghan, A. Meyerson, R. Motwani, N. Mishra, and S. Guha. Streaming-data algorithms for high-quality clustering. In *Proc. 2002 Int. Conf. Data Engineering (ICDE'02)*, pages 685–696, San Fransisco, CA, Apr. 2002.
- [ORS98] B. Özden, S. Ramaswamy, and A. Silberschatz. Cyclic association rules. In *Proc. 1998 Int. Conf. Data Engineering (ICDE'98)*, pages 412–421, Orlando, FL, Feb. 1998.
- [PHMA<sup>+</sup>01] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu. PrefixSpan: Mining sequential patterns efficiently by prefix-projected pattern growth. In *Proc. 2001 Int. Conf. Data Engineering (ICDE'01)*, pages 215–224, Heidelberg, Germany, April 2001.
- [PHMA<sup>+</sup>04] J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, and M.-C. Hsu. Mining sequential patterns by pattern-growth: The prefixspan approach. *IEEE Trans. Knowledge and Data Engineering*, 16:1424–1440, 2004.
- [PHP<sup>+</sup>01] H. Pinto, J. Han, J. Pei, K. Wang, Q. Chen, and U. Dayal. Multi-dimensional sequential pattern mining. In *Proc. 2001 Int. Conf. Information and Knowledge Management (CIKM'01)*, pages 81–88, Atlanta, GA, Nov. 2001.

- [PHW02] J. Pei, J. Han, and W. Wang. Constraint-based sequential pattern mining in large databases. In *Proc. 2002 Int. Conf. Information and Knowledge Management (CIKM'02)*, pages 18–25, McLean, VA, Nov. 2002.
- [Rab89] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proc. IEEE*, 77:257–286, 1989.
- [RM97] D. Rafiei and A. Mendelzon. Similarity-based queries for time series data. In *Proc. 1997 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'97)*, pages 13–25, Tucson, AZ, May 1997.
- [SA96] R. Srikant and R. Agrawal. Mining sequential patterns: Generalizations and performance improvements. In *Proc. 5th Int. Conf. Extending Database Technology (EDBT'96)*, pages 3–17, Avignon, France, Mar. 1996.
- [SM97] J. C. Setubal and J. Meidanis. *Introduction to Computational Molecular Biology*. PWS Pub Co., 1997.
- [SS05] R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications*. Springer, 2005.
- [SZ04] D. Shasha and Y. Zhu. *High Performance Discovery In Time Series: Techniques and Case Studies*. Springer, 2004.
- [TGNO92] D. Terry, D. Goldberg, D. Nichols, and B. Oki. Continuous queries over append-only databases. In *Proc. 1992 ACM-SIGMOD Int. Conf. Management of Data (SIGMOD'92)*, pages 321–330, 1992.
- [Vit85] J. S. Vitter. Random sampling with a reservoir. *ACM Trans. Math. Softw.*, 11:37–57, 1985.
- [Wat95] M. S. Waterman. *Introduction to Computational Biology: Maps, Sequences, and Genomes (Interdisciplinary Statistics)*. CRC Press, 1995.
- [WFYH03] H. Wang, W. Fan, P. S. Yu, and J. Han. Mining concept-drifting data streams using ensemble classifiers. In *Proc. 2003 ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD'03)*, pages 226–235, Washington, DC, Aug. 2003.
- [WH04] J. Wang and J. Han. BIDE: Efficient mining of frequent closed sequences. In *Proc. 2004 Int. Conf. Data Engineering (ICDE'04)*, pages 79–90, Boston, MA, Mar. 2004.
- [YHA03] X. Yan, J. Han, and R. Afshar. CloSpan: Mining closed sequential patterns in large datasets. In *Proc. 2003 SIAM Int. Conf. Data Mining (SDM'03)*, pages 166–177, San Fransisco, CA, May 2003.
- [YJF98] B.-K. Yi, H. V. Jagadish, and C. Faloutsos. Efficient retrieval of similar time sequences under time warping. In *Proc. 1998 Int. Conf. Data Engineering (ICDE'98)*, pages 201–208, Orlando, FL, Feb. 1998.
- [YSJ<sup>+</sup>00] B.-K. Yi, N. Sidiropoulos, T. Johnson, H. V. Jagadish, C. Faloutsos, and A. Biliris. Online data mining for co-evolving time sequences. In *Proc. 2000 Int. Conf. Data Engineering (ICDE'00)*, pages 13–22, San Diego, CA, Feb. 2000.
- [YW03] J. Yang and W. Wang. CLUSEQ: Efficient and effective sequence clustering. In *Proc. 2003 Int. Conf. Data Engineering (ICDE'03)*, pages 101–112, Bangalore, India, March 2003.
- [YWY03] J. Yang, W. Wang, and P. S. Yu. Mining asynchronous periodic patterns in time series data. *IEEE Trans. Knowl. Data Eng.*, 15:613–628, 2003.
- [Zak98] M. J. Zaki. Efficient enumeration of frequent sequences. In *Proc. 7th Int. Conf. Information and Knowledge Management (CIKM'98)*, pages 68–75, Washington, DC, Nov. 1998.
- [Zak01] M. Zaki. SPADE: An efficient algorithm for mining frequent sequences. *Machine Learning*, 40:31–60, 2001.

- [ZCC<sup>+</sup>02] S. Zdonik, U. Cetintemel, M. Cherniack, C. Convey, S. Lee, G. Seidman, M. Stonebraker, N. Tatbul, and D. Carney. Monitoring streams—a new class of data management applications. In *Proc. 2002 Int. Conf. Very Large Data Bases (VLDB'02)*, pages 215–226, Hong Kong, China, Aug. 2002.
- [ZH02] M. J. Zaki and C. J. Hsiao. CHARM: An efficient algorithm for closed itemset mining. In *Proc. 2002 SIAM Int. Conf. Data Mining (SDM'02)*, pages 457–473, Arlington, VA, April 2002.
- [ZLO98] M. J. Zaki, N. Lesh, and M. Ogihara. PLANMINE: Sequence mining for plan failures. In *Proc. 1998 Int. Conf. Knowledge Discovery and Data Mining (KDD'98)*, pages 369–373, New York, NY, Aug. 1998.
- [ZS02] Y. Zhu and D. Shasha. Statstream: Statistical monitoring of thousands of data streams in real time. In *Proc. 2002 Int. Conf. Very Large Data Bases (VLDB'02)*, pages 358–369, Hong Kong, China, Aug. 2002.