Outlier Detection for Temporal Data

— Proposal for a Tutorial at SDM'13 Conference —

Manish Gupta[1], Jing Gao[2], Charu Aggarwal[3], Jiawei Han[1]

[1]Univ. of Illinois at Urbana-Champaign, {gupta58,hanj}@illinois.edu

[2]State Univ of New York, Buffalo, jing@buffalo.edu

[3]IBM T. J. Watson Research Center, charu@us.ibm.com

Abstract

Outlier (or anomaly) detection is a very broad field which has been studied in the context of a large number of research areas like statistics, data mining, sensor networks, environmental science, distributed systems, spatio-temporal mining, etc. The first few articles in outlier detection focused on time series based outliers (in statistics). Since then, outlier detection has been studied on a large variety of data types including high-dimensional data, uncertain data, stream data, network data, time series data, spatial data, and spatio-temporal data. While there have been many tutorials and surveys for general outlier detection, we focus on outlier detection for temporal data in this tutorial.

A large number of applications generate temporal datasets. For example, in our everyday life, various kinds of records like credit, personnel, financial, judicial, medical, etc. are all temporal. This stresses the need for an organized and detailed study of outliers with respect to such temporal data. In the past decade, there has been a lot of research on various forms of temporal data including consecutive data snapshots, series of data snapshots and data streams. Besides the initial work on time series, researchers have focused on rich forms of data including multiple data streams, spatio-temporal data, network data, community distribution data, etc. Compared to general outlier detection, techniques for temporal outlier detection are very different, like AR models, Markov models, evolutionary clustering, etc.

In this tutorial, we will present an organized picture of recent research in temporal outlier detection. We begin by motivating the importance of temporal outlier detection and briefing the challenges beyond usual outlier detection. Then, we list down a taxonomy of proposed techniques for temporal outlier detection. Such techniques broadly include statistical techniques (like AR models, Markov models, histograms, neural networks), distance and density based approaches, grouping based approaches (clustering, community detection), network based approaches, and spatio-temporal outlier detection approaches. We summarize by presenting a collection of applications where temporal outlier detection techniques have been applied to discover interesting outliers.

1   Introduction

- **Why is a tutorial on this theme important?**

  With the rapid increase of stored data, the interest in the discovery of hidden information has exploded in the last decade. One important problem that arises during the discovery process is treating data with temporal dependencies. Given such huge amounts of temporal data, an important task is to find surprising instances efficiently. Recently, many effective and efficient temporal anomaly detection techniques have been proposed in a variety of research disciplines including data mining, sensor networks, environmental science, distributed systems, spatio-temporal mining, etc. Though there have been multiple surveys and tutorials on general outlier detection, there is no single tutorial or survey dedicated to a thorough study of the diverse techniques and extensive studies in temporal outlier detection. We believe that the organized and extensive coverage of work from multiple disciplines in this tutorial will greatly benefit researchers in these disciplines and motivate cross fertilization of ideas.

- **Intended audience**: Researchers and practitioners in data mining, distributed systems and sensor networks. While the audience with a good background on data mining would benefit most from this tutorial, we believe the material to be presented would give general audience and newcomers a complete picture of the current work, introduce important research topics in this field, and inspire them to learn more.

- **Background**: Preliminary knowledge about data mining and algorithms.

2 A preliminary outline of the tutorial

1. Introduction to Temporal Outlier Detection

   (a) Main challenges

   (b) Taxonomy of techniques for temporal outlier detection

   (c) Comparisons with general outlier detection

   (d) Motivations

2. Statistical Models

   (a) Regression models

      i. AutoRegressive (AR) model [12, 40, 50]

      ii. Polynomial curve fitting [1]

   (b) Markov models

      i. Dynamic Bayesian networks [41]

      ii. Markov chain model [84]

      iii. Markov models with suffix trees [48, 57]

   (c) Online sequential discounting algorithms [81, 82]

   (d) Histograms, neural networks, TSA-tree [64, 65, 73]

3. Distance and Density based Approaches

   (a) Incremental distance based outliers [9, 42, 83]

   (b) Incremental local outlier factor (LOF) [70]

4. Grouping based Approaches

   (a) Outlier detection using dynamic cluster maintenance [10, 27]

   (b) Community outlier detection

      i. Evolutionary community outlier detection [37]

      ii. Community trend outlier detection [36]

5. Network based Approaches

   (a) Centralized approaches

      i. Outlier detection by comparing graphs [67, 74]

      ii. Outlier detection using PCA on origin-destination flows in networks [51]

      iii. Graph outlier detection using time series of graph properties [69]

      iv. Link-based approches (LOADED and GOutlier) [3, 35]

   (b) Distributed approaches

      i. Distributed computation of distance based outliers [16, 76]

      ii. Distributed computation of link based outliers [66]

6. Spatio-Temporal Outlier Detection Approaches

   (a) Introduction to spatio-temporal outliers [15, 23, 77]

   (b) A spatio-temporal tree based approach [80]

   (c) PCA based approach [54]

7. Applications in Real Datasets

   (a) Environmental sensor data [31, 40, 41, 44, 50, 54, 77, 80]

   (b) Industrial sensor data [12, 25, 65]

   (c) Computer networks data [51, 52, 53]

   (d) Web graphs [68]

8. Summary

3   Length of the tutorial

Based on the outline above, we propose to give a regular (2 hours) tutorial.

4   About the Instructor

- **Manish Gupta**, http://web.engr.illinois.edu/~gupta58/ received his Masters in Computer Science from IIT Bombay, in 2007. He worked for Yahoo! Bangalore for two years. Since 2009, he has been working towards his Ph.D. with Dr. Jiawei Han at the Department of Computer Science, University of Illinois at Urbana-Champaign. He has interned over summers at Microsoft Research, IBM Research and NEC Labs America. His research interests are in the areas of data mining and information retrieval. Specifically his recent interests are in the area of mining of information networks. He has published more than 20 research papers in referred journals and conferences, including KDD, PKDD, SDM, WWW conferences.

- **Jing Gao**, http://www.cse.buffalo.edu/~jing/, received the BEng and MEng degrees, both in Computer Science from Harbin Institute of Technology, China, in 2002 and 2004, respectively. She received her Ph.D. degree in Computer Science from University of Illinois at Urbana Champaign, in 2011. She is currently an assistant professor in the Computer Science and Engineering Department of the State University of New York at Buffalo. She is broadly interested in data and information analysis with a focus on data mining and machine learning. In particular, her research interests include ensemble methods, transfer learning, mining data streams and anomaly detection. She has published more than 40 papers in refereed journals and conferences, including KDD, NIPS, ICDCS, ICDM and SDM conferences.

- **Charu Aggarwal**, http://charuaggarwal.net/, is a Research Scientist at the IBM T. J. Watson Research Center in Yorktown Heights, New York. He completed his B.S. from IIT Kanpur in 1993 and his Ph.D. from Massachusetts Institute of Technology in 1996. His research interest during his Ph.D. years was in combinatorial optimization (network flow algorithms), and his thesis advisor was Professor James B. Orlin . He has since worked in the field of performance analysis, databases, and data mining. He has published over 200 papers in refereed conferences and journals, and has applied for or been granted over 80 patents. Because of the commercial value of the above-mentioned patents, he has received several invention achievement awards and has thrice been designated a Master Inventor at IBM. He is a recipient of an IBM Corporate Award (2003) for his work on bio-terrorist threat detection in data streams, a recipient of the IBM Outstanding Innovation Award (2008) for his scientific contributions to privacy technology, and a recipient of an IBM Research Division Award (2008) for his scientific contributions to data stream research. He has served on the program committees of most major database/data mining conferences, and served as program vice-chairs of the SIAM Conference on Data Mining , 2007, the IEEE ICDM Conference, 2007, the WWW Conference 2009, and the IEEE

ICDM Conference, 2009. He served as an associate editor of the IEEE Transactions on Knowledge and Data Engineering Journal from 2004 to 2008. He is an associate editor of the ACM TKDD Journal , an action editor of the Data Mining and Knowledge Discovery Journal , an associate editor of the ACM SIGKDD Explorations, and an associate editor of the Knowledge and Information Systems Journal. He is a fellow of the IEEE for "contributions to knowledge discovery and data mining techniques", and a life-member of the ACM.

- **Jiawei Han**, http://www.cs.uiuc.edu/~hanj/ (Ph.D., Univ. of Wisconsin at Madison), is Abel Bliss Professor in Engineering, in the Department of Computer Science at the University of Illinois. He has been researching into data mining, information network analysis, and database systems, with over 600 publications. He served as the founding Editor-in-Chief of ACM Transactions on Knowledge Discovery from Data (TKDD) and on the editorial boards of several other journals. Jiawei has received IBM Faculty Awards, HP Innovation Awards, ACM SIGKDD Innovation Award (2004), IEEE Computer Society Technical Achievement Award (2005), IEEE Computer Society W. Wallace McDowell Award (2009), and Daniel C. Drucker Eminent Faculty Award at UIUC (2011). He is a Fellow of ACM and a Fellow of IEEE. He is currently the Director of Information Network Academic Research Center (INARC) supported by the Network Science-Collaborative Technology Alliance (NS-CTA) program of U.S. Army Research Lab. His book "Data Mining: Concepts and Techniques" (Morgan Kaufmann) has been used worldwide as a textbook.

5 Related Tutorials Given by the Instructors

There is no related tutorial presented by the instructors.

6 Related Events

Many related events dedicated to outlier detection were held by other people in data mining, sensor networks, communication networks, and distributed systems communities.

1. **Tutorial**: "*Anomaly Detection in Military Communication Networks*", (Ritu Chadha and Akshay Vashist), The 2012 Military Communications Conf. (MILCOM'12), Orlando, Florida, Oct 2012.

2. **Tutorial**: "*Network Anomaly Detection: From Basics to Advanced Methods*", (Christian Callegari), The 8th Intl. Wireless Communications and Mobile Computing Conf. (IWCMC'12), Limassol, Cyprus, Aug 2012.

3. **Tutorial**: "*Discovering Roles and Anomalies in Graphs: Theory and Applications*", (Tina Eliassi-Rad and Christos Faloutsos), 2012 SIAM Intl. Conf. on Data Mining (SDM'12), Anaheim, CA, Apr 2012.

4. **Tutorial**: "*Anomaly Detection: A Tutorial*", (Sanjay Chawla and Varun Chandola), The 2011 IEEE Intl. Conf. on Data Mining (ICDM'11), Vancouver, Canada, Dec 2011.

5. **Tutorial**: "*Wavelets and Network Anomaly Detection*", (Michele Pagano and Christian Callegari), The 3rd Intl. Conf. on Security of Information and Networks (SIN'10), Taganrog, Russia, Sep 2010.

6. **Tutorial**: "*Advanced Statistical Approaches for Network Anomaly Detection*", (Christian Callegari), The 2010 Intl. Symposium on Performance Evaluation of Computer and Telecommunication Systems (SPECTS'10), Ottawa, Canada, Jul 2010.

7. **Tutorial**: "*Outlier Detection Techniques*", (Hans-Peter Kriegel, Peer Kröger, Arthur Zimek), The 16th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (KDD'10), Washington, D.C., Jul 2010.

8. **Tutorial**: "*Outlier Detection Techniques*", (Hans-Peter Kriegel, Peer Kröger, Arthur Zimek), The 2010 SIAM Intl. Conf. on Data Mining (SDM'10), Columbus, Ohio, Apr 2010.

9. **Tutorial**: "*Statistical Approaches for Network Anomaly Detection*", (Christian Callegari), The 2nd Intl. Conf. on Security of Information and Networks (SIN'09), Gazimagusa, North Cyprus, Oct 2009.

10. **Article**: "*A Tutorial Overview of Anomaly Detection in Hyperspectral Images*", (Stefania Matteoli, Marco Diani, Giovanni Corsini), Aerospace and Electronic Systems Magazine, IEEE, 25:7 pages 5–28, May 2009.

11. **Tutorial**: "*Outlier Detection Techniques*", (Hans-Peter Kriegel, Peer Kröger, Arthur Zimek), The 13th Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD'09), Bangkok, Thailand, Apr 2009.

12. **Tutorial**: "*Data Mining for Anomaly Detection*", (Aleksander Lazarevic, Jaideep Srivastava, Vipin Kumar, Arindam Banerjee and Varun Chandola), The 2008 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD'08), Antwerp, Belgium, Sep 2008.

13. **Tutorial**: "*Anomaly Detection*", (Arindam Banerjee, Varun Chandola, Vipin Kumar and Jaideep Srivastava), The 2008 SIAM Intl. Conf. on Data Mining (SDM'08), Atlanta, Georgia, Apr 2008.

14. **Tutorial**: "*Outlier Detection: Principles, Techniques and Applications*", (Sanjay Chawla), The 2006 Pacific-Asia Conf. on Knowledge Discovery and Data Mining(PAKDD'06), Singapore, Apr 2006.

15. **Unpublished Report**: "*Summary of Biosurveillance-relevant Statistical and Data Mining Technologies*", (Andrew Moore, Gregory Cooper, Rich Tsui and Michael Wagner), Auton Lab, Carnegie Mellon University, Feb 2002.

Outlier detection is so popular and useful for industry that many tools have been built for efficient outlier detection. E.g. Package 'outliers' in R[1], RapidMiner[2], Oracle[3], etc. Besides these, the following workshops also focused on the general area of outlier detection.

1. **Workshop**: "*Intl. Workshop on Maritime Anomaly Detection*", (Tilburg center for Cognition and Communication), Tilburg, The Netherlands, Jun 2011.

2. **Workshop**: "*Intimate 2006 – Network Anomaly Diagnosis Workshop*", (Benoit Claise, Christophe Diot, Serge Fdida, Laurent Bernaille and Chantal Perrichon), Paris, France, Jul 2006.

3. **Workshop**: "*Data Mining Methods for Anomaly Detection*", (Dragos Margineantu, Stephen Bay, Philip Chan and Terran Lane), The 2005 ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining (KDD'05), Chicago, IL, Aug 2005.

4. **Workshop**: "*The 1st Workshop on Intrusion Detection and Network Monitoring*", (Marcus J. Ranum), Santa Clara, CA, Apr 1999.

5. **Workshop**: "*The 1st Intl. Workshop on the Recent Advances in Intrusion Detection (RAID'98)*", (Marc Wilikens), Louvain-la-Neuve, Belgium, Sep 1998.

All of the above events focused on the general area of outlier detection. However, none of these have focused specifically on temporal outlier detection. In our tutorial, we will provide an overview of techniques for temporal outlier detection in a very organized way. Moreover, we will also point out the new challenges involved in handling temporal data in various forms. Aspects derived from research in other disciplines will be interesting to the data mining audience.

7 Other Tutorials Given by the Instructors

The instructors have rich experience at teaching various kinds of courses on data mining. The following tutorials were delivered in the international conferences in the last several years.

---

[1] http://cran.r-project.org/web/packages/outliers/outliers.pdf
[2] http://www.youtube.com/watch?v=C1KNb1Kw-As
[3] http://docs.oracle.com/cd/B28359_01/datamine.111/b28129/anomalies.htm

1. **Conference tutorial**: "*On the Power of Heterogeneous Information Networks*", (Yizhou Sun, Jiawei Han, Xifeng Yan, and Philip S. Yu), 2012 Intl. Conf. on Advances in Social Network Analysis and Mining (ASONAM'12), Istanbul, Turkey, Aug 2012.

2. **Conference tutorial**: "*Mining Knowledge from Interconnected Data: A Heterogeneous Information Network Analysis Approach*", (Yizhou Sun, Jiawei Han, Xifeng Yan, and Philip S. Yu), 2012 Intl. Conf. on Very Large Data Bases (VLDB'12/PVLDB), Istanbul, Turkey, Aug 2012.

3. **Invited tutorial**: "*Data Mining with Social and Trajectory Data: Urban Computing in the Big Data Age*", (Jiawei Han), China Computer Federation - Advanced Disciplines Lectures (CCF-ADL), Beijing, China, Aug 2012.

4. **Conference tutorial**: "*Mining Knowledge from Data: An Information Network Analysis Approach*", (Jiawei Han, Yizhou Sun, Xifeng Yan, Philip S. Yu), 2012 IEEE Intl. Conf. on Data Engineering (ICDE'12), Arlington, VA, Apr 2012.

5. **Conference tutorial**: "*Data Stream Mining: Challenges and Techniques*", (Latifur Khan, Wei Fan, Jiawei Han, Jing Gao, Mohammad Mehedy Masud), The 15th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'11), Shenzhen, China, May 2011.

6. **Conference tutorial**: "*Mining Heterogeneous Information Networks*", (Jiawei Han, Yizhou Sun, Xifeng Yan, and Philip S. Yu), 2010 ACM SIGKDD Conf. on Knowledge Discovery and Data Mining (KDD'10), Washington, D.C., Jul 2010.

7. **Conference tutorial**: "*Mining Knowledge from Databases: An Information Network Analysis Approach*", (Jiawei Han, Yizhou Sun, Xifeng Yan, and Philip S. Yu), 2010 ACM SIGMOD Intl. Conf. on Management of Data (SIGMOD'10), Indianapolis, Indiana, Jun 2010.

8. **Conference tutorial**: "*On the Power of Ensemble: Supervised and Unsupervised Methods Reconciled*", (Jing Gao, Wei Fan, and Jiawei Han), 2010 SIAM Data Mining Conf. (SDM'10), Columbus, Ohio, May 2010.

9. **Conference tutorial**: "*Mining Moving Object and Traffic Data*", (Jiawei Han, Zhenhui Li, and Lu An Tang), 2010 Intl. Conf. on Database Systems for Advanced Applications (DASFAA'10), Tsukuba, Japan, Apr 2010.

10. **Conference tutorial**: "*A Geometric Perspective on Dimensionality Reduction*", (Deng Cai, Jiawei Han, and Xiaofei He), 2009 SIAM Data Mining Conf. (SDM'09), Sparks, NV, Apr 2009.

11. **Conference tutorial**: "*Scalable OLAP and Mining of Information Networks*", (Jiawei Han, Xifeng Yan, and Philip S. Yu), 2009 Intl. Conf. on Extending Database Technology (EDBT'09), Saint-Petersburg, Russia, Mar 2009.

12. **Conference tutorial**: "*Integration of Classification and Pattern Mining: A Discriminative and Frequent Pattern-based Approach*", (Hong Cheng, Jiawei Han, Xifeng Yan, and Philip S. Yu), 2008 Intl. Conf. on Data Mining (ICDM'08), Pisa, Italy, Dec 2008.

13. **Conference tutorial**: "*Mining Massive RFID, Trajectory, and Traffic Data Sets*", (Jiawei Han, Jae-Gil Lee, Hector Gonzalez, Xiaolei Li), 2008 ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining (KDD'08), Las Vegas, NE, Aug 2008.

14. **Conference tutorial**: "*Exploring the Power of Links in Scalable Data Analysis*", (Jiawei Han, Xiaoxin Yin, and Philip S. Yu), 2008 Intl. Conf. on Data Engineering (ICDE'08), Cancun, Mexico, Apr 2008.

15. **Conference tutorial**: "*Mining for Software Reliability,*" (Chao Liu, Tao Xie, and Jiawei Han), 2007 Intl. Conf. on Data Mining (ICDM'07), Omaha, NE, Oct 2007.

16. **Conference tutorial**: "*Exploring the Power of Links in Data Mining,*" (Jiawei Han, Xiaoxin Yin, and Philip S. Yu), 2007 Intl. Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD'07), Warsaw, Poland, Sep 2007.

17. **Conference tutorial**: *"Mining and Searching Graphs and Structures,"* (Jiawei Han, Xifeng Yan, and Philip S. Yu), 2006 ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining (KDD'06), Philadelphia, PA, Aug 2006.

18. **Conference tutorial**: *"Mining, Indexing, and Similarity Search in Graphs and Complex Structures,"* (Jiawei Han, Xifeng Yan, and Philip S. Yu), 2006 Intl. Conf. on Data Engineering (ICDE'06), Atlanta, GA, Apr 2006.

19. **Conference invited tutorials**: *"Mining and Searching of Graph-Structured Databases,"* (Jiawei Han, Xifeng Yan and Philip S. Yu), 2005 Intl. Conf. on Data Mining (ICDM'05), Houston, TX, Nov 2005.

20. **Conference industry portal tutorials**: *"Traditional Data Mining,"* and *"Data Mining Today,"* (Jiawei Han), 2004 Asian-Pacific Conf. on Knowledge Discovery and Data Mining (PAKDD'04), Sydney, Australia, May 2004.

21. **Conference tutorial**: *"Data Mining for Machine Learners,"* (Johannes Gehrke and Jiawei Han), 2003 Intl. Conf. on Knowledge Discovery and Data Mining (KDD'03), Washington, D.C., Aug 2003.

22. **Conference tutorial**: *"Frequent Pattern Mining Methods and Applications: An Overview,"* (Jiawei Han, Laks V.S. Lakshmanan, and Jian Pei), 2001 Intl. Conf. on Knowledge Discovery and Data Mining (KDD'01), San Jose, California, Aug 2001.

23. **Conference tutorial**: *"Sequential Pattern Mining: From Shopping History Analysis to Weblog Mining and DNA Mining,"* (Jiawei Han), 2001 Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD'01), Hong Kong, April 2001.

8 References for Tutorial Preparation

The following is a list of references which will be used in the preparation of the tutorial material. Many other papers will be referenced as well.

References

[1] C. C. Aggarwal. On Abnormality Detection in Spuriously Populated Data Streams. In *Proc. of the 2005 SIAM Intl. Conf. on Data Mining (SDM)*, pages 80–91, 2005.

[2] C. C. Aggarwal, editor. *Data Streams - Models and Algorithms*, volume 31 of *Advances in Database Systems*. Springer, 2007.

[3] C. C. Aggarwal, Y. Zhao, and P. S. Yu. Outlier Detection in Graph Streams. In *Proc. of the 27th Intl. Conf. on Data Engineering (ICDE)*, pages 399–409. IEEE Computer Society, 2011.

[4] J. Allan. Introduction to Topic Detection and Tracking. In J. Allan, editor, *Topic Detection and Tracking*, pages 1–16, 2002.

[5] J. Allan, editor. *Topic Detection and Tracking: Event-based Information Organization*. Kluwer Academic Publishers, 2002.

[6] J. Allan, J. Carbonell, G. Doddington, J. Yamron, and Y. Yang. Topic Detection and Tracking Pilot Study: Final Report. In *Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop*, pages 194–218, Feb 1998.

[7] J. Allan, V. Lavrenko, and R. Swan. Explorations Within Topic Tracking and Detection. In J. Allan, editor, *Topic Detection and Tracking*, pages 197–224, 2002.

[8] J. Allan, R. Papka, and V. Lavrenko. On-line New Event Detection and Tracking. In *Proc. of the 21st Annual Intl. ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR)*, pages 37–45, 1998.

[9] F. Angiulli and F. Fassetti. Detecting Distance-based Outliers in Streams of Data. In *Proc. of the 16th ACM Conf. on Information and Knowledge Management (CIKM)*, pages 811–820, 2007.

[10] I. Assent, P. Kranen, C. Baldauf, and T. Seidl. AnyOut: Anytime Outlier Detection on Streaming Data. In *Database Systems for Advanced Applications*, volume 7238 of *Lecture Notes in Computer Science*, pages 228–242. Springer Berlin / Heidelberg, 2012.

[11] V. Barnett and T. Lewis. *Outliers in Statistical Data.* John Wiley & Sons Ltd., 1978.

[12] S. Basu and M. Meckesheimer. Automatic Outlier Detection for Time Series: An Application to Sensor Data. *Knowledge and Information Systems - Special Issue on Mining Low-Quality Data*, 11(2):137–154, Feb 2007.

[13] M. Berndtsson and J. Hansson. Issues in Active Real-Time Databases. In *Active Real-Time Databases (ARTDB)*, pages 142–157, 1995.

[14] A. M. Bianco, M. García Ben, E. J. Martnez, and V. J. Yohai. Outlier Detection in Regression Models with ARIMA Errors Using Robust Estimates. *Journal of Forecasting*, 20(8):565–579, Dec 2001.

[15] D. Birant and A. Kut. Spatio-Temporal Outlier Detection in Large Databases. *Journal of Computing and Information Technology (CIT)*, 14(4):291–297, 2006.

[16] J. Branch, B. Szymanski, C. Giannella, R. Wolff, and H. Kargupta. In-Network Outlier Detection in Wireless Sensor Networks. In *Proc. of the 26th IEEE Intl. Conf. on Distributed Computing Systems (ICDCS)*, pages 51–81, 2006.

[17] T. Brants, F. Chen, and A. Farahat. A System for New Event Detection. In *Proc. of the 26th Annual Intl. ACM SIGIR Conf. on Research and Development in Informaion Retrieval (SIGIR)*, pages 330–337, 2003.

[18] J. P. Burman and M. C. Otto. Census Bureau Research Project: Outliers in Time Series, 1988.

[19] V. Chandola, A. Banerjee, and V. Kumar. Anomaly Detection: A Survey. *ACM Computing Surveys*, 41(3):15:1–15:58, 2009.

[20] I. Chang, G. C. Tiao, and C. Chen. Estimation of Time Series Parameters in the Presence of Outliers. *Technometrics*, 30(2):193–204, 1988.

[21] C. Chen and L.-M. Liu. Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of the American Statistical Association*, 88(421):284–297, 1993.

[22] X.-y. Chen and Y.-y. Zhan. Multi-scale Anomaly Detection Algorithm based on Infrequent Pattern of Time Series. *Journal of Computational and Applied Mathematics*, 214(1):227–237, Apr 2008.

[23] T. Cheng and Z. Li. A Multiscale Approach for Spatio-Temporal Outlier Detection. *Transactions in Geographic Information Systems (T. GIS)*, 10(2):253–263, 2006.

[24] A. M. da Silva, C. C. Young-Molling, and S. Levitus. Atlas of Surface Marine Data 1994 Vol. 4: Anomalies of Fresh Water Fluxes. NOAA Atlas NESDIS 9. Technical report, U.S. Gov. Printing Office, Wash., D.C., 1994.

[25] D. Dasgupta and S. Forrest. Novelty Detection in Time Series Data using Ideas from Immunology. In *In Proc. of the 5th Intl. Conf. on Intelligent Systems*, 1996.

[26] P. Dickinson, H. Bunke, A. Dadej, and M. Kraetzl. Median Graphs and Anomalous Change Detection in Communication Networks. In *Proc. of the Intl. Conf. on Information, Decision and Control*, pages 59–64, Feb 2002.

[27] M. Elahi, K. Li, W. Nisar, X. Lv, and H. Wang. Efficient Clustering-Based Outlier Detection Algorithm for Dynamic Data Stream. In *Proc. of the 5th Intl. Conf. on Fuzzy Systems and Knowledge Discovery (FSKD)*, volume 5, pages 298–304, Oct 2008.

[28] T. Fawcett and F. Provost. Activity Monitoring: Noticing Interesting Changes in Behavior. In *Proc. of the 5th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 53–62, 1999.

[29] J. G. Fiscus, G. R. Doddington, J. S. Garofolo, and A. F. Martin. NIST's 1998 Topic Detection and Tracking Evaluation (TDT2). In *Proc. of the 6th European Conference on Speech Communication and Technology (EUROSPEECH)*, 1999.

[30] A. J. Fox. Outliers in Time Series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(3):350–363, 1972.

[31] C. Franke and M. Gertz. Detection and Exploration of Outlier Regions in Sensor Data Streams. In *Proc. of the 2008 IEEE Intl. Conf. on Data Mining Workshops (ICDMW)*, pages 375–384, 2008.

[32] C. Franke and M. Gertz. ORDEN: Outlier Region Detection and Exploration in Sensor Networks. In *Proc. of the 2009 ACM SIGMOD Intl. Conf. on Management of data (SIGMOD)*, pages 1075–1078, 2009.

[33] P. H. Franses and A. Lucas. Outlier Detection in Cointegration Analysis. *Journal of Business and Economic Statistics*, 16(4):459–468, 1998.

[34] P. Galeano, D. Pea, and R. S. Tsay. Outlier Detection in Multivariate Time Series by Projection Pursuit. *Journal of the American Statistical Association*, 101(474):654–669, 2006.

8

[35] A. Ghoting, M. E. Otey, and S. Parthasarathy. LOADED: Link-Based Outlier and Anomaly Detection in Evolving Data Sets. In *Proc. of the 4th IEEE Intl. Conf. on Data Mining (ICDM)*, pages 387–390, 2004.

[36] M. Gupta, J. Gao, Y. Sun, and J. Han. Community Trend Outlier Detection using Soft Temporal Pattern Mining. In *Proc. of the European Conf. on Machine Learning and Knowledge Discovery in Databases (ECML PKDD)*, page To appear, 2012.

[37] M. Gupta, J. Gao, Y. Sun, and J. Han. Integrating Community Matching and Outlier Detection for Mining Evolutionary Community Outliers. In *Proc. of the 18th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 859–867, 2012.

[38] J. Han, M. Kamber, and J. Pei. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers Inc., 2011.

[39] D. M. Hawkins. *Identification of Outliers*. Chapman and Hall, 1980.

[40] D. J. Hill and B. S. Minsker. Anomaly Detection in Streaming Environmental Sensor Data: A Data-driven Modeling Approach. *Environmental Modelling and Software*, 25(9):1014–1022, Sep 2010.

[41] D. J. Hill, B. S. Minsker, and E. Amir. Real-time Bayesian Anomaly Detection for Environmental Sensor Data. In *Proc. of the 32nd Conference of International Association of Hydraulic Engineering and Research (IAHR)*, 2007.

[42] N. Hoang Vu, V. Gopalkrishnan, and P. Namburi. Online Outlier Detection Based on Relative Neighbourhood Dissimilarity. In *Proc. of the 9th Intl. Conf. on Web Information Systems Engineering (WISE)*, pages 50–61, 2008.

[43] V. J. Hodge and J. Austin. A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review*, 22(2):85–126, 2004.

[44] V. Jakkula and D. J. Cook. Anomaly Detection using Temporal Data Mining in a Smart Home Environment. *Methods of information in medicine*, 47(1):70–75, 2008.

[45] M. Jesús Sánchez and D. Peña. The Identification of Multiple Outliers in ARIMA Models. *Communications in Statistics - Theory and Methods*, 32(6):1265–1287, 2003.

[46] M. Jun, H. Jeong, and C. Kuo. Distributed Spatio-temporal Outlier Detection in Sensor Networks. In *Society of Photo-Optical Instrumentation Engineers (SPIE)*, pages 760–763, 2005.

[47] A. Justel, D. Pena, and R. S. Tsay. Detection of Outlier Patches in Autoregressive Time Series. *Statistica Sinica*, 11(3):651–674, 2001.

[48] E. Keogh, S. Lonardi, and B. Y.-c. Chiu. Finding Surprising Patterns in a Time Series Database in Linear Time and Space. In *Proc. of the 8th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 550–556, 2002.

[49] A. Kontostathis, L. Galitsky, W. M. Pottenger, S. Roy, and D. J. Phelps. A Survey of Emerging Trend Detection in Textual Data Mining. In *Survey of Text Mining I: Clustering, Classification, and Retrieval, Volume 1*, pages 185–228, 2003.

[50] W. F. Krajewski and K. L. Krajewski. Real-time Quality Control of Streamflow Data: A Simulation Study. *Journal of the American Water Resources Association (JAWRA)*, 25(2):391–399, 1989.

[51] A. Lakhina, M. Crovella, and C. Diot. Characterization of Network-wide Anomalies in Traffic Flows. In *Proc. of the 4th ACM SIGCOMM Conf. on Internet Measurement (IMC)*, pages 201–206, 2004.

[52] T. Lane and C. E. Brodley. An Application of Machine Learning to Anomaly Detection. In *Proc. of the 20th National Information Systems Security Conf. (NISSC)*, pages 366–380, 1997.

[53] T. Lane and C. E. Brodley. Temporal Sequence Learning and Data Reduction for Anomaly Detection. In *Proc. of the 5th ACM Conf. on Computer and Communications Security (CCS)*, pages 150–158, 1998.

[54] R. Lasaponara. On the Use of Principal Component Analysis (PCA) for Evaluating Interannual Vegetation Anomalies from SPOT/VEGETATION NDVI Temporal Series. *Ecological Modelling*, 194(4):429–434, 2006.

[55] N. D. Le, R. D. Martin, and A. E. Raftery. Modeling Flat Stretches, Bursts, and Outliers in Time Series Using Mixture Transition Distribution Models. *Journal of the American Statistical Association*, 91(436):1504–1515, 1996.

[56] J. Lin, E. Keogh, A. Fu, and H. Van Herle. Approximations to Magic: Finding Unusual Medical Time Series. In *Proc. of the 18th IEEE Symposium on Computer-Based Medical Systems (CBMS)*, pages 329–334, 2005.

[57] J. Lin, E. Keogh, S. Lonardi, and B. Chiu. A Symbolic Representation of Time Series, with Implications for Streaming Algorithms. In *Proc. of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, pages 2–11, 2003.

[58] A. Luceno. Detecting Possibly Non-Consecutive Outliers in Industrial Time Series. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 60(2):295–310, 1998.

[59] M. Markou and S. Singh. Novelty Detection: A Review - Part 1: Statistical Approaches. *Signal Processing*, 83(12):2481–2497, 2003.

[60] M. Markou and S. Singh. Novelty Detection: A Review - Part 2: Neural Network Based Approaches. *Signal Processing*, 83(12):2499–2521, 2003.

[61] M. M. Masud, T. Al-Khateeb, L. Khan, C. C. Aggarwal, J. Gao, J. Han, and B. M. Thuraisingham. Detecting Recurring and Novel Classes in Concept-Drifting Data Streams. In *Proc. of the 11th IEEE International Conference on Data Mining (ICDM)*, pages 1176–1181, 2011.

[62] M. M. Masud, Q. Chen, L. Khan, C. C. Aggarwal, J. Gao, J. Han, and B. M. Thuraisingham. Addressing Concept-Evolution in Concept-Drifting Data Streams. In *Proc. of the 10th IEEE International Conference on Data Mining (ICDM)*, pages 929–934, 2010.

[63] R. E. McCulloch and R. S. Tsay. Bayesian Inference and Prediction for Mean and Variance Shifts in Autoregressive Time Series. *Journal of the American Statistical Association*, 88(423):968–978, 1993.

[64] S. Muthukrishnan, R. Shah, and J. Vitter. Mining Deviants in Time Series Data Streams. In *Proc. of the 16th Intl. Conf. on Scientific and Statistical Database Management*, pages 41–50, Jun 2004.

[65] A. Nairac, N. Townsend, R. Carr, S. King, P. Cowley, and L. Tarassenko. A System for the Analysis of Jet Engine Vibration Data. *Integrated Computer-Aided Engineering*, 6(1):53–66, Jan 1999.

[66] M. E. Otey, A. Ghoting, and S. Parthasarathy. Fast Distributed Outlier Detection in Mixed-Attribute Data Sets. *Data Mining and Knowledge Discovery*, 12(2-3):203–228, May 2006.

[67] P. Papadimitriou, A. Dasdan, and H. Garcia-Molina. Web Graph Similarity for Anomaly Detection. In *Proc. of the 17th Intl. Conf. on World Wide Web (WWW)*, pages 1167–1168, 2008.

[68] P. Papadimitriou, A. Dasdan, and H. Garcia-Molina. Web Graph Similarity for Anomaly Detection. *Journal of Internet Services and Applications*, 1(1):19–30, 2010.

[69] B. Pincombe. Anomaly Detection in Time Series of Graphs using ARMA Processes. *ASOR Bulletin*, 24(4):2–10, 2005.

[70] D. Pokrajac, A. Lazarevic, and L. J. Latecki. Incremental Local Outlier Detection for Data Streams. In *IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, pages 504–515. IEEE, Apr 2007.

[71] P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*. John Wiley & Sons, Inc., 1987.

[72] K. Sequeira and M. Zaki. ADMIT: Anomaly-based Data Mining for Intrusions. In *Proc. of the 8th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 386–395, 2002.

[73] C. Shahabi, X. Tian, and W. Zhao. TSA-Tree: A Wavelet-Based Approach to Improve the Efficiency of Multi-Level Surprise and Trend Queries on Time-Series Data. In *Proc. of the 12th Intl. Conf. on Scientific and Statistical Database Management (SSDBM)*, pages 55–68, 2000.

[74] P. Showbridge, M. Kraetzl, and D. Ray. Detection of Abnormal Change in Dynamic Networks. In *Proc. of the Intl. Conf. on Information, Decision and Control*, pages 557–562, 1999.

[75] G. Silvestri, F. Verona, M. Innocenti, and M. Napolitano. Fault Detection using Neural Networks. In *Proc. of the 1994 IEEE Intl. Conf. on Neural Networks*, volume 6, pages 3796–3799, Jun 1994.

[76] S. Subramaniam, T. Palpanas, D. Papadopoulos, V. Kalogeraki, and D. Gunopulos. Online Outlier Detection in Sensor Data using Non-parametric Models. In *Proc. of the 32nd Intl. Conf. on Very Large Data Bases (VLDB)*, pages 187–198, 2006.

[77] Y. Sun, K. Xie, X. Ma, X. Jin, W. Pu, and X. Gao. Detecting Spatio-Temporal Outliers in Climate Dataset: A Method Study. In *Proc of the 2005 IEEE Intl. Geoscience and Remote Sensing Symposium (IGARSS)*, pages 760–763, 2005.

[78] R. S. Tsay. Time Series Model Specification in the Presence of Outliers. *Journal of the American Statistical Association*, 81(393):132–141, 1986.

[79] R. S. Tsay, D. Pena, and A. E. Pankratz. Outliers in Multivariate Time Series. *Biometrika*, 87(4):789–804, 2000.

[80] E. Wu, W. Liu, and S. Chawla. Spatio-Temporal Outlier Detection in Precipitation Data. In *Proc. of the 2nd Intl. Conf. on Knowledge Discovery from Sensor Data (Sensor-KDD)*, pages 115–133, 2010.

[81] K. Yamanishi and J.-i. Takeuchi. A Unifying Framework for Detecting Outliers and Change Points from Non-stationary Time Series Data. In *Proc. of the 8th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 676–681, 2002.

[82] K. Yamanishi, J.-I. Takeuchi, G. Williams, and P. Milne. On-Line Unsupervised Outlier Detection Using Finite Mixtures with Discounting Learning Algorithms. *Data Mining and Knowledge Discovery*, 8(3):275–300, May 2004.

[83] D. Yang, E. A. Rundensteiner, and M. O. Ward. Neighbor-based Pattern Detection for Windows over Streaming Data. In *Proc. of the 12th Intl. Conf. on Extending Database Technology: Advances in Database Technology (EDBT)*, pages 529–540, 2009.

[84] N. Ye. A Markov Chain Model of Temporal Behavior for Anomaly Detection. In *Proc. of the Workshop on Information Assurance and Security (WAIS)*, Jun 2000.

[85] J. Zhang, Q. Gao, and H. Wang. SPOT: A System for Detecting Projected Outliers From High-dimensional Data Streams. In *Proc. of the 2008 IEEE 24th Intl. Conf. on Data Engineering (ICDE)*, pages 1628–1631, 2008.

[86] Y. Zhang, N. Meratnia, and P. J. M. Havinga. Outlier Detection Techniques For Wireless Sensor Networks: A Survey. Technical Report TR-CTIT-08-59, Centre for Telematics and Information Technology University of Twente, Oct 2008.