Outlier Detection for Temporal Data

— Proposal for a Tutorial at CIKM'13 Conference —

Manish Gupta[1], **Jing Gao**[2], **Charu Aggarwal**[3], **Jiawei Han**[4]

[1]Microsoft, India, gmanish@microsoft.com

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

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

[4]Univ. of Illinois at Urbana-Champaign, hanj@illinois.edu

1 About the Instructor

- **Manish Gupta**, http://dais.cs.uiuc.edu/manish/ is an applied researcher at the Bing team in Microsoft India R&D Private Limited at Hyderabad, India. He received his Masters in Computer Science from IIT Bombay in 2007 and his Ph.D. from the University of Illinois at Urbana-Champaign in 2013. He worked for Yahoo! Bangalore for two years. His research interests are in the areas of web mining, data mining and information retrieval. 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.

2 Aims/Learning Objectives

Temporal data is omnipresent and growing rapidly. The aim of this tutorial is to present various recent techniques developed to perform outlier detection on various forms of such temporal data. 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. For each temporal data type, we will list down several interesting outlier definitions and present approaches suggested in the literature for efficient and effective detection of such outliers. We summarize by presenting a collection of applications where temporal outlier detection techniques have been applied to discover interesting outliers.

3 Description of Topics

Here is a brief outline of the tutorial with relevant references.

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. Outlier Detection for Time Series Data
   (a) Outliers in Time Series Database
      i. Direct Detection of Outlier Time Series
         A. Unsupervised Discriminative Approaches [67, 18, 36]
         B. Unsupervised Parametric Approaches [79, 77, 70]
         C. Unsupervised OLAP based Approach [54]
         D. Supervised Approaches [15, 55]
      ii. Window based Detection of Outlier Time Series
         A. Normal Pattern Database Approaches [15, 31, 51]
         B. Negative and Mixed Pattern Database Approaches [22, 23, 24]
      iii. Outlier Subsequences in a Test Time Series [46, 57, 7, 37]
   (b) Outliers Within a Given Time Series
      i. Points as Outliers: Prediction models, Profile Similarity based Approaches, Deviants [8, 59, 73, 42, 61]

ii. Subsequences as Outliers [45, 72, 13, 28, 47, 56, 78]

3. Outlier Detection for Stream Data
   (a) Evolving Prediction Models [75, 2]
   (b) Distance based Outliers for Sliding Windows
      i. Distance based Global Outliers [6, 76]
      ii. Distance based Local Outliers [11, 66]
      iii. Other Variants [16, 12]
   (c) Outliers in High-dimensional Data Streams [80]
4. Outlier Detection for Stream Data in Distributed Scenarios
   (a) Distributed Setting and Challenges [81]
   (b) Distributed Temporal Data [10, 63, 69]
   (c) Distributed Sensor Data Streams with Spatial Considerations [44, 26, 27]
5. Outlier Detection for Spatio-Temporal Data
   (a) Techniques for ST-Outlier Detection [9, 19]
   (b) Tracking of ST-Outliers [74, 58]
   (c) Trajectory Outliers [53, 30]
6. Outlier Detection for Temporal Information Networks
   (a) Graph Similarity based Outlier Detection Algorithms [64, 65, 68]
   (b) Online Graph Outlier Detection Algorithms [41, 5]
   (c) Community based Outlier Detection Algorithms [35, 33, 34]
7. Applications of Temporal Outlier Detection Techniques
   (a) Environmental sensor data [26, 38, 39, 43, 48, 52, 71, 74]
   (b) Industrial sensor data [8, 21, 62]
   (c) Computer networks data [49, 50, 51]
   (d) Web graphs [64]
8. Summary

4 Scope of the Tutorial

The tutorial covers outlier detection techniques for temporal data popular in data mining community. Many techniques have also been developed in statistics community and we would not cover them. Specifically, we would discuss techniques for time series data, data streams, distributed data streams, network data, and spatio-temporal data. We would not cover novelty detection techniques.

5 Tutorial History

"*Outlier Detection for Temporal Data*", (Manish Gupta, Jing Gao, Charu Aggarwal and Jiawei Han), 2013 SIAM Intl. Conf. on Data Mining (SDM'13), Austin, Texas, May 2013.
"*Outlier Detection for Graph Data*", (Manish Gupta, Jing Gao, Charu Aggarwal and Jiawei Han), 2013 IEEE/ACM Intl. Conf. on Social Networks Analysis and Mining (ASONAM'13), Niagara Falls, Canada, Aug 2013.

6 Format of Tutorial

1/2 day

7 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.

8 Keywords

Temporal outlier detection
Time series data
Data streams
Distributed data streams
Temporal networks
Spatiotemporal outliers

9 Full Description

Outlier detection is a very broad field and has been studied in the context of a large number of application domains. Chandola et al. [17], Hodge et al. [40] and Zhang et al. [81] provide an extensive overview of outlier detection techniques. Outlier detection is also referred to as anomaly detection, event detection, novelty detection, deviant discovery, change point detection, fault detection, intrusion detection or misuse detection. Three main types of outliers studied in the literature are point outliers, contextual outliers and collective outliers. A variety of supervised, semi-supervised and unsupervised techniques have been used for outlier detection. These include classification based, clustering based, nearest neighbor based, density based, statistical, information theory based, spectral decomposition based, visualization based, depth based, and signal processing based techniques. In the past few decades, outlier detection has been studied for high-dimensional data [3], uncertain data [4], streaming data [1, 2, 5], network data [5, 29, 32, 34, 35] and time series data [14, 25]. Outlier detection has been used extensively for intrusion detection, fraud detection, fault detection, system health monitoring, event detection in sensor networks, and detecting eco-system disturbances. Outlier detection is very popular in industrial applications, and many software tools have been built for efficient outlier detection. Some examples include R (packages 'outliers'[1] and 'outlierD' [20]), SAS[2], RapidMiner[3], Oracle[4], etc.

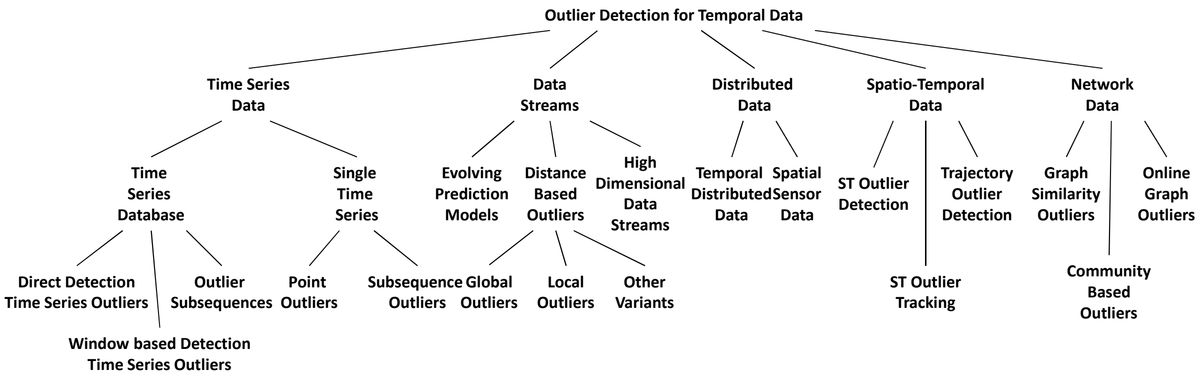

Figure 1: Organization of the Tutorial

Different kinds of data in our everyday lives, such as credit, personnel, financial, judicial, medical, web usage data are temporal. Social network data streams, astronomy data, sensor data, computer network traffic, commercial transactions are all examples of massive amounts of temporal data. As a result, over time, besides time series, a large variety of temporal datasets have become quite popular. These include temporal networks, temporal databases, data streams, distributed data streams, and spatio-temporal data. The quest to mine information from these new forms of temporal data have led to the rise of the field of temporal data mining [60] which includes tasks such as temporal data similarity computation, representation, and summarization, temporal data classification and clustering, prediction, temporal pattern discovery, spatio-temporal data mining, and outlier detection for temporal data. This tutorial will focus on the aspect of outlier detection.

Specific Challenges for Outlier Detection for Temporal Data: While temporal outlier detection aims to find rare and interesting instances, as in the case of traditional outlier detection, new challenges arise due to the nature of temporal data. We list them below.

- Since new data arrives at every time instant, the scale of the data is very large. This often leads to processing and resource challenges.
- In the case of temporal data, the data model needs to be updated when there is significant model drift. These often correspond to change points in the data.
- In order to detect change points, temporal data outlier detection needs to be very efficient. In the streaming content, a single scan is allowed. Traditional outlier detection is much easier, since it is typically an offline task.
- Outlier detection for temporal data in distributed scenarios poses significant challenges of minimizing communication overhead and computational load in resource-constrained environments.
- How to combine the properties of the data itself, the network, and the space dimension with the time dimension to identify practically useful outlier definitions is very challenging.

Over the past decade, several useful techniques have been proposed in the literature in data mining with applications in sensor networks, environmental science and distributed systems communities. In this work, we aim to provide a comprehensive and structured overview of such techniques. We will discuss temporal

[1] http://cran.r-project.org/web/packages/outliers/outliers.pdf
[2] http://www.nesug.org/Proceedings/nesug10/ad/ad07.pdf
[3] http://www.youtube.com/watch?v=C1KNb1Kw-As
[4] http://docs.oracle.com/cd/B28359_01/datamine.111/b28129/anomalies.htm

outlier detection techniques for time series data, stream data, distributed data streams, spatio-temporal data and temporal networks. We will also present a few applications where such temporal outlier detection techniques have been successfully employed. Figure 1 shows the organization of the material that will be covered in the tutorial.

10 Target Audience

Researchers and practitioners in knowledge management, 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.

11 Relation to CIKM Areas

The theme of the proposed tutorial is most related to knowledge management track of CIKM 2013. Specifically in that track, the tutorial falls under the "Data mining theory, methods, and applications" and "Big data analytics" categories.

12 Pre-requisite Knowledge of Audience

Preliminary knowledge about data mining and algorithms.

13 Relevant References

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. Aggarwal and K. Subbian. Event Detection in Social Streams. In *Proc. of the* 12^{th} *SIAM Intl. Conf. on Data Mining (SDM)*, pages 624–635, 2012.

[2] 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.

[3] C. C. Aggarwal and P. S. Yu. Outlier Detection for High Dimensional Data. *SIGMOD Records*, 30:37–46, May 2001.

[4] C. C. Aggarwal and P. S. Yu. Outlier Detection with Uncertain Data. In *Proc. of the 2008 SIAM Intl. Conf. on Data Mining (SDM)*, pages 483–493, 2008.

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

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

[7] M. Atallah, R. Gwadera, and W. Szpankowski. Detection of Significant Sets of Episodes in Event Sequences. In *Proc. of the* 4^{th} *IEEE Intl. Conf. on Data Mining (ICDM)*, pages 3–10, 2004.

[8] 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.

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

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

[11] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander. LOF: Identifying Density-based Local Outliers. In *Proc. of the 2000 ACM SIGMOD Intl. Conf. on Management of Data (SIGMOD)*, pages 93–104, 2000.

[12] Y. Bu, L. Chen, A. W.-C. Fu, and D. Liu. Efficient Anomaly Monitoring over Moving Object Trajectory Streams. In *Proc. of the* 15^{th} *ACM Intl. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 159–168, 2009.

[13] Y. Bu, O. T.-W. Leung, A. W.-C. Fu, E. J. Keogh, J. Pei, and S. Meshkin. WAT: Finding Top-K Discords in Time Series Database. In *Proc. of the* 7^{th} *SIAM Intl. Conf. on Data Mining (SDM)*, pages 449–454, 2007.

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

[15] J. a. B. D. Cabrera, L. Lewis, and R. K. Mehra. Detection and Classification of Intrusions and Faults using Sequences of System Calls. *SIGMOD Records*, 30(4):25–34, Dec 2001.

[16] H. Cao, Y. Zhou, L. Shou, and G. Chen. Attribute Outlier Detection over Data Streams. In *Proc. of the* 15^{th} *Intl. Conf. on Database Systems for Advanced Applications - Volume Part II (DASFAA)*, pages 216–230, 2010.

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

[18] V. Chandola, V. Mithal, and V. Kumar. A Comparative Evaluation of Anomaly Detection Techniques for Sequence Data. In *Proc. of the 2008* 8^{th} *IEEE Intl. Conf. on Data Mining (ICDM)*, pages 743–748, 2008.

[19] T. Cheng and Z. Li. A Hybrid Approach to Detect Spatial-Temporal Outliers. In *Proc. of the* 12^{th} *Intl. Conf. on Geoinformatics Geospatial Information Research*, pages 173–178, 2004.

[20] H. Cho, Y. jin Kim, H. J. Jung, S.-W. Lee, and J. W. Lee. Outlierd: An r package for outlier detection using quantile regression on mass spectrometry data. *Bioinformatics*, 24(6):882–884, 2008.

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

[22] D. Dasgupta and N. Majumdar. Anomaly Detection in Multidimensional Data using Negative Selection Algorithm. In *Proc. of the 2002 Congress on Evolutionary Computation (CEC)*, volume 2, pages 1039–1044, 2002.

[23] P. D'haeseleer, S. Forrest, and P. Helman. An Immunological Approach to Change Detection: Algorithms, Analysis and Implications. In *Proc. of the 1996 IEEE Symposium on Security and Privacy*, pages 110–119, May 1996.

[24] S. Forrest, A. S. Perelson, L. Allen, and R. Cherukuri. Self-Nonself Discrimination in a Computer. In *Proc. of the 1994 IEEE Symposium on Security and Privacy*, pages 202–212, 1994.

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

[26] 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.

[27] 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.

[28] A. W.-c. Fu, O. T.-W. Leung, E. Keogh, and J. Lin. Finding Time Series Discords based on Haar Transform. In *Proc. of the* 2^{nd} *Intl. Conf. on Advanced Data Mining and Applications (ADMA)*, pages 31–41, 2006.

[29] J. Gao, F. Liang, W. Fan, C. Wang, Y. Sun, and J. Han. On Community Outliers and their Efficient Detection in Information Networks. In *Proc. of the* 16^{th} *ACM Intl. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 813–822, 2010.

[30] Y. Ge, H. Xiong, Z.-h. Zhou, H. Ozdemir, J. Yu, and K. C. Lee. Top-Eye: Top-K Evolving Trajectory Outlier Detection. In *Proc. of the* 19^{th} *ACM Intl. Conf. on Information and Knowledge Management (CIKM)*, pages 1733–1736, 2010.

[31] A. Ghosh, A. Schwartzbard, M. Schatz, et al. Using Program Behavior Profiles for Intrusion Detection. In *Proc. of the SANS Intrusion Detection Workshop*, 1999.

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

[33] M. Gupta, C. C. Aggarwal, J. Han, and Y. Sun. Evolutionary Clustering and Analysis of Bibliographic Networks. In *Proc. of the 2011 Intl. Conf. on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 63–70, 2011.

[34] M. Gupta, J. Gao, Y. Sun, and J. Han. Community Trend Outlier Detection using Soft Temporal Pattern Mining. In *Proc. of the 2012 European Conf. on Machine Learning and Knowledge Discovery in Databases (ECML PKDD)*, pages 692–708, 2012.

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

[36] M. Gupta, A. B. Sharma, H. Chen, and G. Jiang. Context-Aware Time Series Anomaly Detection for Complex Systems. In *Proc. of the SDM Workshop on Data Mining for Service and Maintenance*, 2013.

[37] R. Gwadera, M. J. Atallah, and W. Szpankowski. Reliable Detection of Episodes in Event Sequences. *Knowledge and Information Systems (KAIS)*, 7(4):415–437, May 2005.

[38] 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.

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

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

[41] T. IDÉ and H. KASHIMA. Eigenspace-based Anomaly Detection in Computer Systems. In *Proc. of the* 10^{th} *ACM Intl. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 440–449, 2004.

[42] H. V. Jagadish, N. Koudas, and S. Muthukrishnan. Mining Deviants in a Time Series Database. In *Proc. of the* 25^{th} *Intl. Conf. on Very Large Data Bases (VLDB)*, pages 102–113, 1999.

[43] 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.

[44] 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.

[45] E. Keogh, J. Lin, and A. Fu. HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence. In *Proc. of the* 5^{th} *IEEE Intl. Conf. on Data Mining (ICDM)*, pages 226–233, 2005.

[46] 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* 8^{th} *ACM Intl. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 550–556, 2002.

[47] E. Keogh, S. Lonardi, and C. A. Ratanamahatana. Towards Parameter-free Data Mining. In *Proc. of the* 10^{th} *ACM Intl. Conf. on Knowledge Discovery and Data Mining (KDD)*, pages 206–215, 2004.

[48] 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.

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

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

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

[52] 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.

[53] J.-G. Lee, J. Han, and X. Li. Trajectory Outlier Detection: A Partition-and-Detect Framework. In *Proc. of the 2008 IEEE* 24^{th} *Intl. Conf. on Data Engineering (ICDE)*, pages 140–149, 2008.

[54] X. Li and J. Han. Mining Approximate Top-K Subspace Anomalies in Multi-dimensional Time-series Data. In *Proc. of the* 33^{rd} *Intl. Conf. on Very Large Data Bases (VLDB)*, pages 447–458, 2007.

[55] X. Li, J. Han, S. Kim, and H. Gonzalez. ROAM: Rule- and Motif-Based Anomaly Detection in Massive Moving Object Data Sets. In *Proc. of the* 7^{th} *SIAM Intl. Conf. on Data Mining (SDM)*, pages 273–284, 2007.

[56] J. Lin, E. Keogh, A. Fu, and H. Van Herle. Approximations to Magic: Finding Unusual Medical Time Series. In *Proc. of the* 18^{th} *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* 8^{th} *ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*, pages 2–11, 2003.

[58] C.-T. Lu and L. R. Liang. Wavelet Fuzzy Classification for Detecting and Tracking Region Outliers in Meteorological Data. In *Proc. of the* 12^{th} *Annual ACM Intl. Workshop on Geographic Information Systems (GIS)*, pages 258–265, 2004.

[59] J. Ma and S. Perkins. Online Novelty Detection on Temporal Sequences. In *Proc. of the* 9^{th} *ACM Intl. Conf on Knowledge Discovery and Data Mining (KDD)*, pages 613–618, 2003.

[60] T. Mitsa. *Temporal Data Mining*. Chapman and Hall/CRC, 2010.

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

[62] 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.

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

[64] 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.

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

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

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

[68] P. Shoubridge, 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.

[69] 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* 32^{nd} *Intl. Conf. on Very Large Data Bases (VLDB)*, pages 187–198, 2006.

[70] P. Sun, S. Chawla, and B. Arunasalam. Mining for Outliers in Sequential Databases. In *Proc. of the* 6^{th} *SIAM Intl. Conf. on Data Mining (SDM)*, pages 94–105, 2006.

[71] 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.

[72] L. Wei, E. Keogh, and X. Xi. SAXually Explicit Images: Finding Unusual Shapes. In *Proc. of the* 6^{th} *Intl. Conf. on Data Mining (ICDM)*, pages 711–720, 2006.

[73] A. W. Williams, S. M. Pertet, and P. Narasimhan. Tiresias: Black-box Failure Prediction in Distributed Systems. In *Proc. of the* 21^{st} *Intl. Parallel and Distributed Processing Symposium (IPDPS)*, pages 1–8, 2007.

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

[75] 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.

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

[77] J. Yang and W. Wang. CLUSEQ: Efficient and Effective Sequence Clustering. In *Proc. of the* 19^{th} *Intl. Conf. on Data Engineering (ICDE)*, pages 101–112, 2003.

[78] D. Yankov, E. Keogh, and U. Rebbapragada. Disk Aware Discord Discovery: Finding Unusual Time Series in Terabyte Sized Datasets. *Knowledge and Information Systems (KAIS)*, 17(2):241–262, Nov 2008.

[79] N. Ye. A Markov Chain Model of Temporal Behavior for Anomaly Detection. In *Proc. of the 2000 IEEE Systems, Man, and Cybernetics - Information Assurance and Security Workshop*, volume 166, pages 171–174, 2000.

[80] 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* 24^{th} *Intl. Conf. on Data Engineering (ICDE)*, pages 1628–1631, 2008.

[81] 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.