MAP-i - PhD Work Plan:

Event Detection

Student: Hadi Fanaee Tork (hadi.fanaee@fe.up.pt)
Supervisor: Joao Gama (jgama@fep.up.pt)
Institute: LIAAD-INESC TEC

Motivation: Event detection is an emerging research area since recent years. It has a wide range of application in many industries from bio-surveillance to the health, safety, economics and monitoring systems, etc. Event detection methods can be classified into four types as temporal, spatial, spatio-temporal and logic-based. In spatio-temporal and spatial event detection we have to deal with high dimensional data continuously flowing as stream. For this purpose, we believe that spectral techniques which are able to handle such data can be very helpful. In this work plan, we first state the problem, its applications and its definition and later explain related works and then finally state our ideas.

1- What is event?

An event is a significant occurrence or pattern that is unusual comparing to the normal patterns of the behavior of a system [1]. This can be natural phenomena or manual system interaction. Some examples of events can be an attack on the network, bioterrorist activities, epidemic disease, damage in an aircraft, pipe-breaks, forest fires, etc. In a real system, most of the times everything is going normal, however in rare occasions something happens which can be very harmful and cause many losses.

2- Event Detection

Event detection process, in its simplest scenario, can be a direct observation of a system state. In more advanced and complex scenarios, the system state is not observable directly and must be inferred from its behavior [1]. In this PhD proposal, we focus on event detection from time-evolving and high speed data streams. Moreover, we assume that the behavior of the system is changing over the time.

Event detection tries to discover emerging behaviors, different from the normal behavior of a system. We need to detect these abnormal behaviors as soon as possible, to prevent severe damages to the system. In fact, Event detection Can be compared as looking for a needle in a haystack in order to 1) find the desired needle 2) check if it’s a needle or something else and 3) specify what the properties of that needle are[2].

3- Applications

Identification emerging event and its position in the data is very important task in many fields and is a growing research topic in many fields such as bio-surveillance, safety, health and economics, etc. For instance in field of disease outbreaks, DARPA [3] has estimated that a two-day gain in detection time could reduce fatalities by a factor of six. Also Neill and Wong [3,4] demonstrated that hotspot violent crime can be predicated 1-3 weeks in advance by detecting the clusters of leading indication crimes such as disorderly conduct, trespass and simple assault. Several other applications regarding the temporal event detection are presented in [1]. Also a tutorial on spatial and spatio-temporal event detection is presented in [3].

4- Event Detection and relevant problems

Event detection is similar to other problems such as change detection, anomaly detection, clustering, one-class classification and novelty detection; however it differs from them in some aspects.

- **Change detection**

By comparing the distribution of most recent observations with previous observations in a data stream, changes can be detected, however that changes are not necessarily start point of events since many of them can be happened by the chance[6], Thus event detection is a different concept.

- **Anomaly and outlier Detection**

Event Detection can be considered as a sub-field of anomaly detection. Anomaly is as a pattern in the data that does not conform to the expected behavior [9]. What make anomaly detection different from event detection is that in anomaly detection we search for points of interest which are different from the rest of the data, while in event detection we search for group of points or patterns which their member can be normal but when are aggregated over time or space or both, they can be appeared surprisingly.

- **Clustering**

Event detection is different from clustering. In clustering we try to simply partition the entire dataset into groups while in event detection we search for the part of dataset in which some quantity is significantly higher (or lower) than expected [6]. Some clustering methods such as density-based clustering, partition the data based on the density of points in space. These partitions may correspond to the anomalous spatial regions. For three reasons they are different from events clusters [6]:

1) They just look for the dense regions regardless of checking if they are really significant clusters or just have occurred by the chance.
2) In Event detection we look for small (but significant) changes among all the data while in density-based clustering this kind of changes cannot be detected.

3) The density-based clustering methods are not proper for dealing with spatially and temporally varying baselines because their nature is based on density of points per area. In event detection we should be able to allow dealing with counts and baselines in a principled probabilistic framework.

- **One-class classification**

In one-class classification we assume that we have the label data with normal and event label. Then a machine learning algorithm classifies the new data into one of the normal and event data. For instance, in anomaly detection there are several works related to intrusion detection systems [5] which use label data to discover the pre-defined attacks. Sometimes these labels can just be available for normal data (semi-supervised). In this case we just have label data for normal data and label of events are unknown [3]. Since most events data are not labeled data this approaches is not so applicable.

- **Novelty Detection**

Novelty Detection is a process of discovery of unknown concepts that has not been learned during training in machine learning. This happens when sometimes the test data contains information about objects that were not known at the time of training the model. The limitation of novelty detection is need for label data which in most of event detection problem is not available.

5- Sensor-based Event detection

When a system is not observable directly, it can be monitored by some sensors. Due to two problems sensor-based event detection is a challenging task [1]: 1) it needs a high amount of computation 2) It needs a large space for marinating huge data. For instance, in order to detect a tsunami event a large number of sensors in the sea are needed to be processed and be kept.

Usually two approaches can be used for detection of events using sensors [1]. In the first approach for each sensor one threshold is determined and if the measure beat the threshold, alarm will be raised. One direct obstacle of this approach is sensor failures. In the second approach, multiple sensors are hired. The problem of this approach is its increasing complexity: we need to monitor correlations between multiple sensors.

From now on whenever we use the term “event detection” we refer “sensor-based event detection”.

6- Event detection challenges

Common event detection challenges are [1, 3]:

- **Domain-Dependence:** one approach which is helpful for one domain can be failed in another one.
- **Time constraint**: Events should be detected in a short time (e.g. seconds or minutes depending on the domain) for prevention and proper reaction.
- **High True Alarm:** Sometimes such as what happens in medical imaging, if a true event was ignored, we miss important information such as cancer events. So the significance of high true alarm rates in some domains is undeniable.
- **Low False Alarm:** Event detection performs well when it produces high true alarms and low false alarms. Sometimes false alarms impose extreme costs and they should be minimized.

7- Event Detection Data Nature

Event detection methods in literature can be categorized in to four groups based on the nature of the data as follow:

- **Temporal event detection**: Event is defined as a detecting different behavior in a single or multiple signals.
- **Spatial event detection**: First law of geography says "Everything is related to everything else, but near things are more related than distant things". Spatial relationship can relate objects as well as their temporal sequences. In this case Event can be defined as detecting different region in a spatial space in terms of its corresponding values. Many contributions in bio-surveillance and medical imaging are based on Kulldorff spatial scan statistic [7]. However there are some recent works like Neil trying to apply Bayesian techniques to spatial scan statistic [8] or similarly, Bayesian network [9] for categorical data. Also Neill has proposed some solutions for fast detection of events [6].
- **Spatio-temporal event detection**: Event is defined as detection of spatio-temporal region in data which its corresponding values are very different from the rest of the data. Most works of Kulldorff and Neill methods are extendable to spatio-temporal.
- **Logic-based temporal methods**: in this type of event detection such as works like [10], first data changes to a semantically higher level of data and then try to discover the rare occurrence among the higher leveled data.

8- Data Types

In most event detection problems we deal with three different types of data as follow:

- **Count data**: is common in sensor-based event detection. Single or multiple sensors are collecting count or measurement data from the sensors.
- **Categorical data**: is more complicated form of data which instead of count of data a records of categorical data such as sex, age range, etc. is available.
- **Higher level of data**: if a lower level of data such as GPS trajectories turn into higher level of data such as "U-turn", "High-speed”. Strange occurrence will be discovered among the obtained higher level data.

9- Event Detection Methods

Temporal vent detection methods mostly are borrowed from change detection, time series analysis and signal processing and can be divided in to four types [1] :
1)Statistical which benefits from methods such as static threshold, quality control charts family such as Shewhart control charts, Moving Average, EWMA and CUSUM, regression, time series analysis methods such as ARMA and ARIMA, Kalman filter[11], Wavelets, Hidden Markov Models and Principal component analysis[12], etc. 2)Probabilistic which instead of statistic test, probability of event occurrence are considered 3)Machine Learning-based which benefit from machine learning methods such as Neural networks, Genetic algorithms, etc. 4)Composite which combine techniques from different methods together. A full review of temporal methods are presented in [1], [3], [10]

10- Computation Improvement

Computation time is the critical issue in event detection problem specially spatial and spatio-temporal event detection, Neill employed a kind of KD-tree namely "overlap-KD tree"[6] to speed up the event detection performance. In another work he employed Bayesian [14] methods to reach more speed. Other works such as [15, 16, and 17] which are suitable for categorical data benefited from association rule and Bayesian network for mining the events in categorical data and also speeding up of performance.

11- Research Proposal

We are interested in research the application of spectral methods on temporal, spatial and spatio-temporal event detection and performing some contributions over spectral-based event detection. The reasons why we think that spectral techniques might be useful in event detection problem are:

- Our approaches will be data driven and we must address the challenge about the streaming behavior of the data and the time-evolving characteristic of the system.
- In event detection problem we are usually dealing with high dimensional data including spatio-temporal dimensions. Spectral techniques have proved their ability to reduce the dimensionality of multi-dimensional data [5]. They also can be employed in the pre-processing step to reduce the dimensionality. Therefore, these approaches can be helpful and need more attention.
- In most event detection real-world problems, label data is not available [6]. Thus unsupervised methods such as spectral approaches which are suitable for both unsupervised settings and forecasting can be a good choice.
- Spectral methods are able to detect anomalies from the multivariate systems.
- Spectral methods doesn’t make any assumption about the nature and distribution of the data.

- Event data are huge and changes over time and also need to be processed quickly. Spectral methods can reduce the required data amount need to be processed and meanwhile can work in incremental way such as [19].

As far as we know, there are not much works around this topic. There are two problems that make the application of spectral-based methods to event detection a challenge [5]. The first one is related to the high computational cost of this method. The second one, is that event detection must be possible to detect after the projecting the raw data in lower dimensions.

Most of the spatial and spatio-temporal event detection methods are around spatial scan statistic and its variations [3, 6, 7, 8, and 9]. Also since our target data is at least three-dimension data (space, time and measurements), we are trying to focus on three-way analysis [18] which had been under less attention by the field researchers. Our goal is to provide a spectral-based event detection framework with the following issues:

- Modeling the spatial and the spatio-temporal data in any kind including count, categorical and higher level data types as a multi dimension tensor.
- Extension of incremental spectral approaches such as [19] to sliding window computational model [22] in event detection.
- Application of multi-way analysis and tensor decomposition methods such as Tucker3 [20] to the spatio-temporal tensor to enable monitoring the system to discover events.
- Extension of Tucker3 to the sliding window computational model.
- Computation and performance improvement.
- Applying techniques in Neighborhood based anomaly detection [5].
- Incorporating the prior knowledge to the system
- Using the model for forecasting of events.

During the research we would like to investigate:

- Comparison of our framework with the state of the art methods in terms of detection accuracy, delay in detection and computational time on real and benchmark data sets.
- Improving the performance and accuracy of our framework by combining some solutions already provided in the literature such as Bayesian techniques.
- Incorporating the spatial index techniques like r-tree and KD-tree in our spectral-based framework to increase the performance.

We expect our final event detection system could

- Provide fast and accurate solutions for detection of events from spatial and spatio-temporal data in comparison of the state of the art methods.
- Retrieving the corresponding part of data that caused the event and its properties.
- Handle three different kinds of event data: count, categorical and higher level.

References

[1] Kerman, M. C. et al. (2009), "Event Detection Challenges, Methods, and Applications in Natural and Artificial Systems". In Proceedings of 14th International Command and Control Research and Technology Symposium: "C2 and Agility".
[2] Ludovic Zimmerlin, Vera S. Donnenberg and Albert D. Donnenberg , "Rare Event Detection and Analysis in Flow Cytometry: Bone Marrow Mesenchymal Stem Cells, Breast Cancer Stem/Progenitor Cells in Malignant Effusions, and Pericytes in Disaggregated Adipose Tissue" , Humana Press, 2010
[3] Neill, D. B., Wong, W-K (2009). "Tutorial on Event Detection", KDD 2009.
[4] D.B. Neill and W.L. Gorr. "Detecting and preventing emerging epidemics of crime. Advances in Disease Surveillance" 4: 13, 2007.
[5] V. Chandola, A. Banerjee, V. Kumar, "Anomaly detection: a survey", ACM Computing Surveys 41 (3) (2009) 1–58. ISSN: 0360-0300, doi: <http://doi.acm.org/10.1145/1541880.1541882>.
[6] D.B. Neill. "Detection of spatial and spatio-temporal clusters". Ph.D. thesis, Carnegie Mellon University , Department of Computer Science, 2006.
[7] M. Kulldorff. "A spatial scan statistic". Communications in Statistics: Theory and Methods, 26(6): 1481–1496, 1997.
[8] D.B. Neill, A.W. Moore, and G.F. Cooper. A Bayesian spatial scan statistic. In Y. Weiss, et al., eds. Advances in Neural Information Processing Systems 18, 1003-1010, 2006.
[9] D.B. Neill, G.F. Cooper, K. Das, X. Jiang, and J. Schneider. Bayesian network scan statistics for multivariate pattern detection. In J. Glaz, et al., eds., Scan Statistics: Methods and Applications, 2009.
[10] Artikis A., Skarlatidis A., Portet F. and Paliouras G. |Logic-Based Event Recognition", to appear in Knowledge Engineering Review, 2012.
[11] Welch G and Bishop G. 2006. "An Introduction to the Kalman Filter." University of North Carolina at Chapel Hill, Chapel Hill, NC. TR 95-041. Retrieved October 4, 2008, from
[12] Jolliffe, I. T. 2002. "Principal Component Analysis", 2nd ed. Springer.
[13] J. L. Bentley. Multidimensional binary search trees used for associative searching. Communications of the ACM, 18:509–517, 1975.
[14] Neill DB, Moore AW, Cooper GF, "A Bayesian spatial scan statistic". Neural Information Processing Systems 18, 2006.
[15] W.-K. Wong, A. W. Moore, G. F. Cooper, and M. M. Wagner. Rule-based anomaly pattern detection for detecting disease outbreaks. In Proc. 18th National Conference on Artificial Intelligence, 2002.
[16] W.-K. Wong, A. W. Moore, G. F. Cooper, and M. M. Wagner. Bayesian network anomaly pattern detection for disease outbreaks. In Proc. 20th International Conference on Machine Learning, 2003.
[17] W.-K.Wong, A.W.Moore, G. F. Cooper, andM.M.Wagner. WSARE:What's strange about recent events? *Journal of Urban Health*, 80(2 Suppl. 1):i66–i75, 2003.
[18] Kroonenberg, P. M. (2008). APPLIED MULTIWAY DATA ANALYSIS. WILLEY ,2008.
[19] J. Sun, D. Tao, S. Papadimitriou, P. S. Yu, and C. Faloutsos, "Incremental tensor analysis: Theory and applications," ACM Trans. Know. Data Discovery, vol. 2, no. 3, pp. 1–37, Oct. 2008.
[20] Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. Psychometrika, 31,279-3 1 1.

[21] M. Markou, S. Singh, Novelty detection: A review, part 1: Statistical approaches, Signal Processing 83, 2481–2497, 2003
[22] J. Gama, M. Gaber (Eds) (2007), "Learning from Data Streams" -- Processing Techniques in Sensor Networks , Springer.