Proceedings of the International Conference on Information and Automation, December 15-18, 2005, Colombo, Sri Lanka.

Efficient Analytics for Effective Monitoring of Biomedical Security

Maheshkumar Sabhnani*, Daniel Neill*, Andrew Moore*, Artur Dubrawski*, and Weng-Keen Wong+
*School of Computer Science
Carnegie Mellon University, Pittsburgh, PA 15213, USA
Email: {sabhnani+, neill, awm, awd}@cs.cmu.edu
+School of Electrical Engineering and Computer Science
Oregon State University, Corvallis, OR 97331-5501
Email: wong@eecs.oregonstate.edu

***Abstract*—This paper reviews three successful statistical data mining approaches developed recently at the Auton Lab of Carnegie Mellon University to support public health officials in their work towards protecting biomedical safety and security. The presented methods focus on monitoring health care data sources including hospital emergency department records, sales of over-the-counter medications, and consumer food complaints. Their purpose is to detect statistically significant signs of disease outbreaks, or food safety related concerns, as early as possible. These approaches have already been successfully deployed in the United States and other developed countries, but they also have a vast potential utility among developing societies. The Auton Lab is actively seeking additional deployments, and several pieces of the relevant software are available for download and use free of charge. This paper describes each of the presented methods, and provides results of their utilization so far.**

I. Introduction

BIOMEDICAL security has recently gained a lot of attention in the research community, both because of the threat of emerging global pandemics such as SARS and avian influenza, as well as the potential danger of bioterrorist attacks. In both cases, the early detection of emerging disease outbreaks can enable more rapid epidemiological response, possibly saving many lives. Many data sources may serve as useful indicators of an emerging outbreak of disease, including patient visits to emergency departments in hospitals, drug sales in medical stores, and consumer food complaints. Our focus is on the development of automatic bio-surveillance systems to analyze one or more of these data streams and generate alerts when abnormal patterns occur.

In this paper, we present overviews of three algorithms that have been successfully applied to monitor such data sources, enabling the timely and accurate detection of disease outbreaks and other patterns of adverse events. All these algorithms have been developed at the Auton Lab of Carnegie Mellon University; more detailed descriptions and software implementations are available on our website (http://www.autonlab.org).

II. What's Strange About Recent Events

A. Overview

Multidimensional data with a temporal component is available in numerous disciplines such as medicine, engineering, and astrophysics. This data is often used for monitoring purposes by a novelty detection system. These systems inspect the data for anomalies and raise an appropriate alert upon discovery of any deviations from the norm. The *What's Strange About Recent Events* (WSARE) is a statistical data mining technology for monitoring a stream of transactional database records to discover whether changes are occurring [1-3]. The changes might be in specific fields with means or variances that have changed significantly. More interestingly, WSARE also alerts when no individual records look strange, and when no individual fields of the sequence of records is changing, but when the interrelationship between fields is changing. When monitoring hospital admissions records, an example of such a pattern might be "*there is no significant increase in males, but among men living in zip code X there is double the rate of respiratory problems*".

WSARE requires records within a specified temporal period to be defined as recent records. Records preceding the recent data points in time are used to produce a baseline dataset that represents normal behavior. WSARE compares the recent data against the baseline data to find the most significant change in recent records. This change is described using a rule, which is made up of components of the form $X_i = V_i^j$, where X_i is the i^{th} feature and V_i^j is the j^{th} value of that feature. For example in case of emergency department cases, the one-component rule *Gender = Male* characterizes the subset of the data involving male visitors. Our most recent work [2-3] replaces our earlier baseline method [1] with a Bayesian network which produces the baseline distribution by taking the joint distribution of the data and conditioning on attributes that are responsible for the trends. We show that our algorithm, called WSARE 3.0,

is able to detect outbreaks in simulated data with almost the earliest possible detection time while keeping a low false positive count.

B. Algorithm

The WSARE algorithm consists of three steps. First of all, the baseline dataset is created. Secondly, the algorithm searches for the best scoring rule using both the recent and the baseline datasets. Finally, a p-value for the best scoring rule is calculated using a randomization test. We will briefly provide more details about the three steps above. We will be describing WSARE 3.0 [2-3], which uses a Bayesian network to create the baseline dataset.

Determining the baseline is difficult due to the presence of various trends in data, such as trends caused by the day of week and by seasonal variations in temperature and weather. Creating the baseline distribution without taking these trends into account can lead to unacceptably high false positive counts and slow detection times. WSARE 3.0 accounts for these trends by using a Bayesian network to model the joint probability distribution of the features of the data. These features can be divided into *environmental attributes*, which are features such as the season and the day of week that cause trends in the data, and *response attributes*, which are the remaining features such as syndrome and gender. WSARE 3.0 learns a Bayesian network from all data from before the recent period. During this Bayesian network structure learning phase, environmental attributes are prevented from having parents because we are not interested in predicting their distributions, but rather, we want to use them to predict the distributions of the response attributes. Once the Bayesian network structure is learned, we can then produce a conditional probability distribution that represents the baseline behavior given the environmental attributes for the current day. As an example, suppose we are monitoring Emergency Department data and that the environmental attributes Season, Day of Week, and Weather cause fluctuations in this data. Also, let the response attributes be $X_1, \ldots, X_n$. Assuming that today is a snowy winter Saturday, we can use the joint probability distribution captured by the Bayesian network to produce the conditional probability distribution $P(X_1, \ldots, X_n \mid Season = Winter, Day\ of\ Week = Saturday, Weather = Snow)$, which intuitively represents the baseline distribution given the conditions for the current day. The baseline dataset can consequently be produced by sampling a large number of records from this conditional probability distribution.

Once the baseline dataset is generated, we need to search for the best scoring rule, which characterizes the group with the most unusual shift in proportions between the baseline and recent datasets. In order to prevent over-fitting, additional components are only added to the best rule so far if the addition of those components is statistically significant.

Let *BR* be the best scoring rule found and let Score(*BR*) be the score of *BR*. We cannot interpret Score(*BR*) as its actual *p-value* because the process for finding the best scoring rule involves multiple hypothesis tests. The final step of WSARE accounts for the multiple hypothesis testing problem by calculating a compensated p-value for the best scoring rule through a randomization test in which the date and the remaining features are assumed to be independent. The randomization test consists of several iterations, typically around 1000. On iteration *j*, we shuffle the dates between records in the recent and the baseline datasets to produce a randomized dataset called DB^j_{rand}. We then find the best scoring rule BR^j on DB^j_{rand}. At the end, we determine where Score(*BR*) would be ranked among the values of Score(BR^j) from all the iterations. The compensated p-value CPV is calculated as:

$$CPV = \frac{\#\,Score(BR^j) < Score(BR)}{\#\,randomization\quad test\quad iterations}$$

Finally, an alarm sounds if the compensated p-value CPV is lower than a threshold, for example 0.01.

C. Results

WSARE has been downloaded by more than 100 public health departments from the USA and the rest of the world since we made it publicly available in January 2003. Recently, the Israel Center for Disease Control evaluated WSARE 3.0 retrospectively using an unusual outbreak of influenza type B that occurred in an elementary school in central Israel. WSARE 3.0 was applied to patient visits to community clinics between the dates of May 24, 2004 to June 11, 2004. The attributes in this dataset included the visit date, area code, ICD-9 code, age category, and day of week. The day of week was used as the only environmental attribute. WSARE 3.0 reported two rules with p-values at 0.002 and five other rules with p-values below 0.0001. Two of the five anomalous patterns with p-values below 0.0001 corresponded to the influenza outbreak in the data. The rules that characterized the two anomalous patterns consisted of the same three attributes of primary complaint code, area code and age category, indicating that an anomalous pattern was found involving children aged 6-14 having viral symptoms within a specific geographic area. WSARE 3.0 successfully detected the outbreak on the second day from its onset. Similarly, in a recent retrospective analysis of the Walkerton (Canada) fatal outbreak of *E. coli* infection due to tap water contamination, it was determined that WSARE 3.0 would have detected the outbreak one day before a boil-water advisory was released if its alarm threshold was set to a level that permitted two false positives per year.

III. Spatial Scan Statistics

A. Overview

In the field of bio-surveillance, epidemiologists are interested in detecting significant clusters of disease cases; these clusters may be indicative of an emerging disease epidemic. We have developed a system for automatic detection of disease clusters that can detect potential outbreaks, pinpoint their spatial location, and distinguish between significant clusters and those simply due to chance. More generally, our methods can be applied to any spatial data mining problem where the goal is detection of *overdensities*: spatial regions with higher than expected values of some quantity (the "count") with respect to some underlying "baseline" information. This is useful not only for bio-surveillance (including both detection of naturally occurring outbreaks and terrorist bio-attacks) but also in many other application domains, such as medical imaging, astrophysics, and military surveillance.

B. Algorithm

We consider spatial datasets that have been aggregated to a uniform, two-dimensional grid. Let G be an $N \times N$ grid of cells, where each cell $s_i \in G$ is associated with a *count* c_i and an underlying *baseline* b_i. For example, a cell's count may be the number of disease cases in that geographical location in a given time period, while its baseline may be the total population "at-risk" for the disease. Our goal is to search over all rectangular regions $S \subseteq G$, and find the region S^* with the highest score according to some *score function* F: $S^* = \text{argmax}_S F(S)$. This score function can be derived from our models of how the data is generated under the null hypothesis H_0 (of no clusters) and the set of alternative hypotheses $H_1(S)$, each representing a cluster in some region S. We typically use the likelihood ratio statistic, which is the likelihood of the data under the alternative hypothesis divided by the likelihood of the data under the null hypothesis. For many of the models used in the bio-surveillance domain, the score function $F(S)$ can be expressed as a function of two sufficient statistics: the total count of the region, $C(S) = \sum_S c_i$, and the total baseline of the region, $B(S) = \sum_S b_i$. This enables us to efficiently compute the score function for any given region S.

One such efficiently computable score function, which is also of great interest to epidemiologists, is Kulldorff's *spatial scan statistic* [4]. This statistic assumes that counts c_i are generated by an inhomogeneous Poisson process with mean qb_i, where b_i is the at-risk population and q is the underlying "disease rate." Kulldorff proved that the following likelihood ratio statistic is most powerful for finding a single region of elevated disease rate:

$$F(S) = C \log \frac{C}{B} + (C_{tot} - C) \log \frac{C_{tot} - C}{B_{tot} - B} - C_{tot} \log \frac{C_{tot}}{B_{tot}}$$

In the above score function, C represents the total count of region S, B represents its total baseline, and C_{tot} and B_{tot} represent the total count and total baseline of the grid respectively. Our algorithms can use either Kulldorff's score function or many other possible statistics; in certain cases, other score functions will have better detection power than Kulldorff's.

Though much previous work focuses on the case of detecting compact (circular or square) clusters, we presented an efficient algorithm [5] that can detect both compact and elongated (rectangular) clusters. This extension is important in epidemiological applications because disease clusters are often elongated: airborne pathogens may be blown by the wind, creating an ellipsoid "plume," and waterborne pathogens may be carried along the path of a river. In each of these cases, the resulting clusters have high aspect ratios, and tests for squares and circles will have low power for detecting these clusters. Naively, the algorithmic complexity of searching an $N \times N$ grid is $O(N^3)$ for squares and $O(N^4)$ for rectangles. In [5], we used an "overlap-kd tree" data structure to reduce the complexity of searching rectangular regions to $O((N \log N)^2)$. This algorithm makes spatial scanning practical and computationally feasible even for massive real world datasets, which require high grid resolutions for accurate searching. Although [5] only considered *axis-aligned rectangles,* the work can be easily extended to search for non-axis aligned rectangles. One simple method of doing this is to examine multiple "rotations" of the data, mapping each to a separate grid and computing the most significant region for each grid.

Once we have found the most significant region of grid G according to our score function, we can compute its statistical significance (p-value) using randomization. To do so, we generate a large number R (typically 1000) of replica grids, where a replica has the same underlying baselines b_i as G, but has counts randomly generated under the assumption of a uniform disease rate. For each replica grid, we find the maximum region score. The p-value can then be calculated as $(R_{beat} + 1) / (R + 1)$, where R_{beat} is the number of replica grids with maximum scores higher than the original grid. If the p-value is less than 0.05, we can conclude that the discovered region is unlikely to have occurred by chance, and is thus a "significant spatial cluster." Otherwise, no significant clusters exist.

C. Multidimensional Spatial Scans

In [6], we extended our efficient spatial scan statistics (*SSS*) methods to multidimensional data. This extension makes spatial scanning computationally feasible for a variety of datasets with more than two spatial dimensions: for example, we successfully applied the spatial scan to detect clusters of brain activity in three-dimensional fMRI (functional magnetic resonance imaging) data. We can also

scan efficiently over space-time clusters, using time as an additional dimension, or include various other attributes of the data. For example, for emergency department visits, in addition to hospital location and time, we can include patient age and gender, allowing us to more accurately detect epidemics with different effects on different groups. For over-the counter drug sales, we can include store location, time, and information about promotional sales. In both of these bio-surveillance domains, multi-dimensional scans result in more meaningful results, as they not only indicate the spatial location of an outbreak but also present characteristics of that outbreak, based on the location and extent of the detected region in the multidimensional space.

D. Space Time Clusters

By extending spatial scan statistics to multi-day analysis in a space-time framework, we are able to detect emerging clusters of disease, whether these clusters emerge rapidly or gradually. In [7], we presented a class of spatio-temporal cluster detection methods that combine univariate time series analysis with spatial scan statistics for rapid detection of emerging space time clusters. These "expectation-based" approaches have two steps: first inferring how many cases we expect to see in each spatial location, then detecting spatial regions where the recent counts are significantly higher than expected. We have developed space-time scan statistics to detect both persistent clusters (regions that have constant relative risk over time) and emerging clusters (regions with risk that is monotonically increasing over time). To estimate baseline counts, a wide variety of univariate time series algorithms were used: these include exponentially weighted linear regression (EWLR), and exponentially weighted moving average (EWMA), with various methods of adjusting for day of week and seasonality. Through extensive testing, many of these methods were shown to be highly successful on the task of prospective detection of disease outbreaks.

E. Bayesian Scan Statistics

In [8], we proposed a Bayesian method for cluster detection, the "Bayesian spatial scan statistic." This method allows us to incorporate prior information about the likelihood, size, and impact of an outbreak; by combining these priors with the observed data, we can compute the posterior probability of an outbreak in each spatial region. We demonstrated that the Bayesian method has several advantages over the standard (frequentist) approach, including higher power to detect clusters and much faster run time. Other advantages include easier calibration, easier interpretation and visualization of results, and easier extension to multivariate data; see [8] for more details.

F. Results

We have demonstrated that our spatial scan techniques are useful for fast and accurate cluster detection in a variety of applications, including bio-surveillance and medical imaging. In the bio-surveillance domain, we have shown that our techniques can rapidly detect disease outbreaks, with both a high probability of detection and a low false positive rate.

We have also designed ultra-fast versions of spatial scan algorithms using the overlap-kd tree data structure, making these algorithms practical even for massive real-world datasets. Our "fast spatial scan" techniques achieve 100-1000x speedups as compared to naïve methods, with no loss of accuracy. Use of the Bayesian approach can further speed up investigation, since randomization testing is unnecessary in the Bayesian framework.

We have tested these algorithms on real-world datasets including emergency department visits (from four different states) and over the-counter sales data (throughout the United States). Results show that our algorithms can scale to such datasets containing millions of records. Every day, our SSS system [9] receives data from over 20,000 stores and hospitals nationwide, uses our automatic cluster detection methods to find potential outbreaks, and makes these results available to state and local public health officials through a web-based graphical interface. These public health officials also provide us with valuable feedback on the system, enabling us to continually improve our models and methods.

We have demonstrated that our space-time methods can rapidly and accurately detect emerging disease epidemics. We have also retrospectively studied various real world outbreaks and found impressive results. As an example, our retrospective analysis of the Walkerton E. Coli outbreak (discussed above) determined that the space-time scan would have detected the outbreak two days before the boil-water advisory, that is, one day before WSARE. The tradeoff, of course, is that WSARE is a more general detector than the space-time scan, and can detect a greater range of outbreak types. Tests on semi-synthetic outbreaks (simulated outbreaks injected into real baseline Emergency Department and over-the-counter sales data) likewise demonstrated that our method is fast, accurate, and has high power to detect outbreaks [7].

IV. TIP MONITOR

A. Overview

TIP MONitor's (TIPMON) job is to identify small subsets (pairs and triplets) of significantly correlated records in an incoming stream of multidimensional event-based data such as anonymous individual patient health events involving chief complaint strings, prescription orders, public safety hotlines or customer complaints. Very often, such data is sparse, noisy and it may contain very little and spotty

evidence of potentially crucial coincidences. The last feature makes it very hard to detect important events with more traditional approaches used by bio-surveillance analysts (such as spatial scan statistics, WSARE, or multivariate time series analysis), since they are designed to benefit from ample evidence.

Suppose that among the chief complaint strings of two unrelated patients in the same city on the same date there was a mention of bloody stools in pediatric cases. The multiple mentions of "bloody stools" or "pediatric" might not be surprising, but the tying together of these two factors, given matching geographic locations and timings of reporting, is sufficiently rare that seeing only two such cases is of interest. This was precisely the evidence that was the first noticeable signal of the tragic Walkerton, Canada, waterborne bacterial gastroenteritis outbreak caused by contamination of tap water in May 2000. That weak signal was spotted by an astute physician, not by a surveillance system. Reliable automated detection of such signals in multivariate data requires new analytic approach such as that in TIPMON.

B. Algorithm

TIPMON analyzes a stream of data. Given a new case X_n and an old case X_o, TIPMON computes the probability of these two cases being *noisy copies* of each other, under hypothesis that they have been generated by the same underlying common cause, C_k. Figure 1 shows this scenario pictorially. For instance, two food consumer complaint reports stemming from the same root cause related to contamination at some food processing plant are likely to have uniquely similar characteristics, as compared to other reports not related to that particular cause.

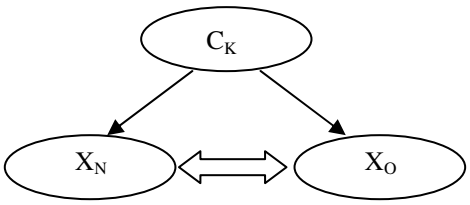

Fig. 1. "Noisy Copy" model in TIPMON**.**

For every new case, TIPMON simultaneously analyzes multiple causal models to find out if any cases in the past would have matched this new case based on each of the predefined causal scenarios. These causal models can be defined as conditional probability of new case and old case being related given the particular cause, mathematically $P(X_n, X_o \mid C_k)$. These models could either be learned, given a sufficient amount of labeled data on hand, or alternatively they could be defined by the domain experts. Each individual scenario of interest defines a new causal model that would be monitored by TIPMON. New causes and corresponding scenarios could also be incrementally added through active learning.

Consider an application that has a total of *K* defined causal models and *N-1* past cases. Let Q_{nok} be the probability of case X_n being similar to a past case X_o given a cause C_k. For this new case X_n, TIPMON analyzes the past historical data to answer the following three critical questions. First, it finds the probability of X_n being similar to any past case, $\left(\sum_{o=1}^{N-1}\sum_{k=1}^{K} Q_{nok}\right)$. This indicates how critical this case could be for further investigation, and the complement of that value directly characterizes anomalousness of the new case given the collection of past cases. Secondly, it finds the chances of the new case being generated according to each specific scenario *k*, $\left(\sum_{o=1}^{N-1} Q_{nok}\right)$. Thirdly, we could also compute chances of each old case being similar to the new case by computing the corresponding marginal probability $\left(\sum_{k=1}^{K} Q_{nok}\right)$. Looking at pairs of cases corresponding to the highest scores defined above, human analysts can more efficiently allocate their time to investigate most probable emerging patterns of recently occurring adverse events.

C. Results

A variant of TIPMON, named EPFC (*Emerging Patterns in Food Complaints*) is the analytical core of the *Consumer Complaint Monitoring System* (CCMS II) deployed at the United States Department of Agriculture (USDA). It is monitoring food complaints related to meat, eggs and poultry. EPFC is designed to screen sparse and noisy data for potential linkages between individual reports of adverse effects of food on its consumers.

These consumer complaint reports, collected in a passive surveillance mode, contain multi-dimensional and heterogeneous snippets of specific information about the consumers' demographics, the kinds, brands and sources of the food they ate, symptoms of sickness they may be experiencing, characteristics of foreign objects which could have been found in food, involved locations and times of occurrences, and so on. The EPFC estimates how likely it is for a newly reported complaint case to be a close copy of some other case in the past data, if both have been generated by the same specific underlying cause, such as for instance a

malicious contamination of raw food at plant. The top matches are reported to human analysts for investigation. The EPFC resolves the problem of manually checking for all possible associations between all possible pairs of complaints over recent weeks, and it helps to efficiently allocate limited analytical and investigative resources. Its unique feature is the ability to remain sensitive to signals supported by very little data – significant alerts can be raised on the basis of a very few complaints from consumers, provided that the few complaints contain significantly similar and explicable root causes.

The EPFC is receiving very positive feedback from its users. When tested on historical CCMS data, it managed to instantaneously flag an outbreak of a food borne illness (E. coli), which took experienced analysts over two weeks to identify. This has been possible due to the ability of TIPMON to remain sensitive to small signals in multivariate data; even when data is spotty, noisy and even if it comes in a short supply (the CCMS system currently logs only about 100 complaints per week).

V. Conclusion

We presented three algorithms for prospective bio-medical surveillance: WSARE, spatial scan statistics (SSS), and TIPMON. Table 1 shows a brief comparison of the utility of these methods in the field of biomedical security.

TABLE I
Functionality Comparison Chart

Algorithm	Prior Knowledge Required	Amount of Evidence Needed	Ability to Rapidly Detect Complex Patterns
Traditional Approaches	What and Where to look for?	Large	Low to Moderate
WSARE	---	Large	High
SSS	What to look for?	Large	Moderate to High
TIPMON	Causal Scenarios*	Small	High

* Causal scenarios can be learned from data if sufficient supply of it is available.

Traditional approaches in bio-medical security (e.g. univariate time series analysis) are highly dependent on the prior knowledge about the type and location of disease outbreaks under consideration. SSS has been developed to scan for anomalous spatial regions, and hence only needs to know the type of outbreak. WSARE searches for a variety of anomalous rules or patterns, and thus does not need either a prespecified type or location. The strength of TIPMON relies on the ability to early detect meaningful patterns based on small amount of evidence.

These three methods have already demonstrated utility in detecting outbreaks of infectious diseases and in discovering problems in safety of food supply in their early stages. The WSARE and SSS algorithms as well as related software are available for free download at http://www.autonlab.org. We highly encourage readers to download and apply them for bio-surveillance and adverse events monitoring tasks. We are hoping that these algorithms can be added to already existing systems at no cost or at a limited cost, dramatically improving system performance, as well as serving as stand-alone outbreak detection tools. We are always interested in resolving any practical issues related to applying these algorithms to real datasets.

VI. Acknowledgments

We would like to thank the software development team at Auton Lab for making these algorithms publicly available. We would also like to thank our sponsors for the numerous grants which helped develop these algorithms. TIPMON has been supported by the United States Department of Agriculture (CCMS II project). WSARE was funded by the DARPA BIOALIRT program. SSS research has been funded by grants from the National Science Foundation and the Pennsylvania Department of Health.

References

[1] W. –K. Wong, A. Moore, G. Cooper, M. Wagner, "Rule-based anomaly pattern detection for detecting disease outbreaks", *Proceedings of the 18th National Conference on Artificial Intelligence,* 2002.

[2] W. –K. Wong, A. Moore, G. Cooper, M. Wagner, "Bayesian network anomaly pattern detection for disease outbreaks", *Proceedings of the International Conference on Machine Learning,* 808-815, 2003.

[3] W. –K. Wong, A. Moore, G. Cooper, M. Wagner, "What's strange about recent events", *Journal of Urban Health*, 80: 66-75, 2003.

[4] M. Kulldorff, "A spatial scan statistic", *Communications in Statistics: Theory and methods,* 26(6): 1481-1496, 1997.

[5] D. Neill, A. Moore, "Rapid detection of significant spatial clusters", *Proceedings of the 10th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 256-265, 2004.

[6] D. Neill, A. Moore, F. Pereira, and T. Mitchell, "Detecting significant multidimensional spatial clusters", in *Advances in Neural Information Processing Systems 17,* MIT Press, 969-976, 2005.

[7] D. Neill, A. Moore, M. Sabhnani, K. Daniel, "Detection of emerging space-time clusters", *Proceedings of the 11th ACM SIGKDD Conference on Knowledge Discovery and Data Mining ,* 218-227, 2005.

[8] D. Neill, A. Moore, G. Cooper, "A Bayesian spatial scan statistic", in *Advances in Neural Information Prcessing Systems 18*, MIT Press, 2006, in press.

[9] M. Sabhnani, D. Neill, A. Moore, F.-C. Tsui, M. Wagner, J. Espino, "Detecting anomalous patterns for pharmacy retail data", *Proceedings of the KDD 2005 Workshop on Data Mining Methods for Anomaly Detection*, 2005.