Adaptive Event Detection with Time–Varying Poisson Processes

Alexander Ihler
School of Information and Computer Science
University of California, Irvine
Irvine, CA 92697–3425
ihler@ics.uci.edu

Jon Hutchins
School of Information and Computer Science
University of California, Irvine
Irvine, CA 92697–3425
johutchi@uci.edu

Padhraic Smyth
School of Information and Computer Science
University of California, Irvine
Irvine, CA 92697–3425
smyth@ics.uci.edu

ABSTRACT

Time-series of count data are generated in many different contexts, such as web access logging, freeway traffic monitoring, and security logs associated with buildings. Since this data measures the aggregated behavior of individual human beings, it typically exhibits a periodicity in time on a number of scales (daily, weekly, etc.) that reflects the rhythms of the underlying human activity and makes the data appear non-homogeneous. At the same time, the data is often corrupted by a number of bursty periods of unusual behavior such as building events, traffic accidents, and so forth. The data mining problem of finding and extracting these anomalous events is made difficult by both of these elements. In this paper we describe a framework for unsupervised learning in this context, based on a time-varying Poisson process model that can also account for anomalous events. We show how the parameters of this model can be learned from count time series using statistical estimation techniques. We demonstrate the utility of this model on two data sets for which we have partial ground truth in the form of known events, one from freeway traffic data and another from building access data, and show that the model performs significantly better than a non-probabilistic, threshold-based technique. We also describe how the model can be used to investigate different degrees of periodicity in the data, including systematic day-of-week and time-of-day effects, and make inferences about the detected events (e.g., popularity or level of attendance). Our experimental results indicate that the proposed time-varying Poisson model provides a robust and accurate framework for adaptively and autonomously learning how to separate unusual bursty events from traces of normal human activity.

Categories and Subject Descriptors

I.5.1 [**Pattern Recognition**]: Models—*statistical*; G.3 [**Probability and Statistics**]: Probabilistic Algorithms

General Terms

Algorithms

Keywords

Poisson, Markov modulated, event detection

1. INTRODUCTION

Analyzing and understanding patterns of human behavior over time is an area of increasing interest in a number of different data mining applications. Examples include analysis and understanding of Web access logs, event detection and prediction with vehicular traffic and accident data, and classifying human activities from low–cost observation modalities used for ubiquitous sensing such as RFID, video, et cetera. In this paper we focus on time-series data where time is discrete and $N(t)$ is a measurement of the number of individuals or objects recorded over the time-interval $[t-1, t]$. For example, an optical sensor at a door might report an estimate of how many people have entered a building over a 30-minute period, or an inductive loop sensor on a freeway might report an estimate of how many vehicles have passed over the sensor in the previous 5 minutes. Since this type of data measures the aggregated behavior of many individuals, it typically exhibits a temporal periodicity on many scales (daily, weekly, etc.) reflecting the rhythms of underlying human activity. It is often also corrupted by sustained (bursty) periods of anomalous behavior, which we will refer to in this paper as *events*. Note that the term *event* is sometimes used in the time series literature to refer to individual measurements (e.g., the recording of a single person walking through a door at a particular time). Here, however, we will use *event* in a different manner, to refer to a large-scale activity that is unusual relative to normal patterns of behavior, such as a large meeting in a building, a malicious attack on a Web server, or a traffic accident on a freeway.

To fully understand such data, we often care about both the patterns of the typical and predictable behavior, and detecting and extracting information from the deviations from this behavior. However, this leads to an inherent "chicken and egg" deconvolution problem, since detecting anomalous periods of time requires some knowledge of what constitutes normal behavior, but our historical data consists of both normal and anomalous (event) data mixed together.

As an example, Figure 1 shows counts of the estimates number of people entering a building over time from an optical sensor at the front door of a UCI campus building. The

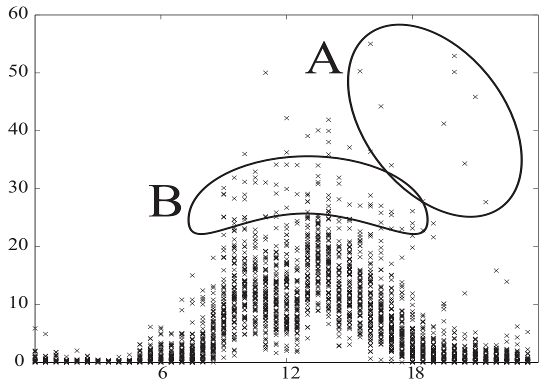

Figure 1: Jittered scatterplot of the number of people entering on any weekday over a four–month period, shown as a function of the time of day (in half-hour intervals). Although certain points (e.g., set A) are clearly "outliers" and represent unusual events with greater than normal attendance, it is less clear which, if any, of the values in set B represent something similar.

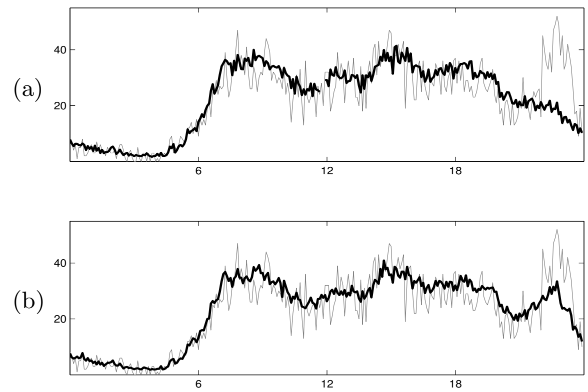

Figure 2: Example of freeway traffic data for Fridays for a particular on-ramp. (a) Average time profile for normal, non game-day Fridays (dark curve) and data for a particular Friday (6/10/05) with a baseball game that night (light curve). (b) Average time profile over all Fridays (dark curve) superposed on the same Friday data (light curve) as in the top panel.

data are "jittered" slightly by Gaussian noise to give a better sense of the density of counts at each time. Once again, there are parts of this signal which are clearly periodic, and other parts which are obvious outliers; but there are many samples which fall into a gray area. For example, set (A) in Figure 1 is clearly far from the typical behavior for their time period; but set (B) contains many points which are somewhat unusual but may or may not be due to the presence of an event. In order to separate the two, we need to define a model of uncertainty (*how* unusual is the measurement?), and additionally incorporate a notion of event *persistence*, i.e., the idea that a single, somewhat unusual measurement may not signify anything but several in a row could indicate the presence of an event.

Another example of this "chicken and egg" problem is illustrated in Figure 2. The top panel shows vehicle counts every five minutes for an on-ramp on the 101 freeway in Los Angeles located near Dodger Stadium, where the Los Angeles Dodgers baseball team plays their home games. The darker line shows the average count for the set of "normal" Fridays when there were no baseball games (averaged over every non game-day Friday for each specific 5-minute time slice). The daily rhythm of normal Friday vehicle flow is clear from the data: little traffic in the early hours of the morning, followed by a sharp consistent increase during the morning rush hour, relatively high volume and variability of traffic during the day, another increase for the evening rush hour, and a slow decay into the night back to light traffic.

The light line in the top panel shows the counts for a particular Friday when there was a baseball game: the "event" can be seen in the form of significantly increased traffic around 22:00 hours, corresponding to a surge of vehicles leaving from the baseball stadium. It is clear that relative to the average profile (the darker line) that the baseball traffic is anomalous and should be relatively easy to detect.

Now consider what would happen if we did not know when the baseball games were being held. The lower panel shows the time series for the same Friday as the top panel (the lighter line) but now with the average over *all* Fridays superposed, i.e., the average time-profile including both game-day and non game-day Fridays. This average profile has now been pulled upwards around 22:00 hours and sits roughly halfway between normal traffic for that time of night (the darker line in the top panel) and the profile that corresponds to a baseball event (the light curve). Ideally we would like to learn both the patterns of normal behavior and to detect events that indicate departures from the norm. For example, given the time series shown in Figure 2, we would like to learn a model that reflects the bimodal nature of such data, namely a combination of the normal traffic patterns to which is occasionally added additional counts caused by aperiodic events.

In this paper we investigate the use of Markov-Poisson models for this purpose, and illustrate how to learn such models from data to both characterize normal behavior and detect anomalous events. The model consists of a time-varying Poisson process that includes both systematic diurnal (time of day) and calendar (day of week) variation in Poisson rates over time, as well as a hidden Markov event process. We adopt a Bayesian approach to learning and inference, allowing us to pose and answer a variety of queries within a probabilistic framework, queries such as "did any events occur in this time-period?", "how many additional counts were caused by a particular event?", "what is the estimated duration of an event?", and so forth. Different high-level questions about the data can also be addressed, such as "are Monday and Tuesday normal patterns the same?" or "are the patterns of normal behavior consistent over time or changing?" using Bayesian model selection techniques.

The remainder of the paper proceeds by discussing related work in Section 2 and then, in Section 3, describing in more detail the two data sets, freeway traffic data and "people counter" building data, that we use throughout the paper. Section 4 illustrates the limitations of a simple baseline approach to event detection based on thresholding. In Section 5 we describe our proposed probabilistic model and

Section 6 describes how this model can be learned from data using a Bayesian estimation framework. Section 7 discusses how we can use the learned model for event detection and validates the model's predictions of anomalous events by using known ground-truth schedules of events. We show that our proposed approach is significantly more accurate in practice than a baseline threshold-based method. Section 8 uses Bayesian model selection techniques to investigate different levels of time-heterogeneity in the model, and Section 9 illustrates how the model can be used for estimating event attendance. In Section 10 we conclude with a brief discussion of open research problems and summary comments.

2. RELATED WORK

There has been a significant amount of prior work in both data mining and statistics on finding surprising patterns, outliers, and change–points in time series. For example, Keogh et al. [1] described a technique that represents a real-valued time-series by quantizing into it a finite set of symbols and then uses a Markov model to detect surprising patterns in the symbol sequence. Guralnik and Srivastava [2] proposed an iterative likelihood-based method for segmenting a time-series into piecewise homogeneous regions. Salmenkivi and Mannila [3] investigated the problem of segmenting sets of low-level time-stamped events into time-periods of relatively constant intensity, using a combination of Poisson models and Bayesian estimation methods. Kleinberg [4] demonstrated how a method based on an infinite automaton could be used to detect bursty events in text streams.

All of these approaches share a common goal with that of this paper, namely detection of novel and unusual data points or segments in time-series. However, none of this earlier work focuses on the specific problem we address here, namely detection of bursty events embedded in time series of counts that reflect the normal diurnal and calendar patterns of human activity.

The model proposed here is derived from the Markov–modulated Poisson processes used by Scott and Smyth [5] for analysis of Web surfing behavior and Scott [6] for telephone network fraud detection. We extend the latter model by employing a more flexible model of event–related counts and including missing data, and show that not only is it accurate at detecting the presence of events in two new data sets by using ground truth for validation, but also how it can be used to perform additional tasks such as model selection and inference over other quantities of interest about the event.

3. DATA SET CHARACTERISTICS

We use two different data sets throughout the paper to illustrate our approach. In this section we describe these data sets in more detail.

The first data set will be referred to as the *building* data, consisting of 3 months of count data automatically recorded every 30 minutes at the front door of the Calit2 institute building on the UC Irvine campus. The data are generated by a pair of battery–powered optical detectors that measure the presence and direction of objects as they pass through the building's main set of doors. The number of "counts" in each direction are then communicated via a wireless link to a base station with internet access, at which they are stored.

The observation sequences ("people counts") acquired at the front door form a noisy time series with obvious structure but many outliers (see Figure 3). The data is corrupted by the presence of "events"—non-periodic activities which take place in the building and (typically) cause an increase in foot traffic entering the building before the event, and leaving the building after the event, possibly with some "churn" (people going in and out) during the event. Some of these events can be seen easily in the time–series, for example the two large spikes in both entry and exit data on days four and twelve in Figure 3. However, many of these events may be less obvious and only become visible when compared to the behavior over a long period of time.

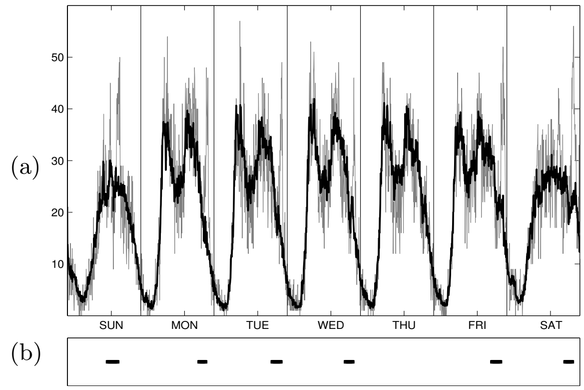

Figure 4: (a) One week of traffic data (light curve) from Sunday to Saturday (June 5-11), with the estimated normal traffic profile (estimated by the proposed model described later in the paper) superposed as a dark curve. (b) Ground truth list of events (baseball games).

The second data set will be referred to as the *freeway traffic data* and consists of estimated vehicle counts every 5 minutes over 6 months from an inductive loop–sensor located on the Glendale on-ramp to the 101-North freeway in Los Angeles [7]. Figure 4 shows the temporal pattern for a particular week starting with Sunday morning and ending Saturday night. The daily rhythms of traffic flow are clearly visible as is the distinction between weekdays and weekends. Also visible are "local excursions" corresponding to significantly different counts compared to relatively smooth normal pattern, such as the baseball games on Sunday afternoon and every evening except Thursday. The lower panel of Figure 4 shows a set of known events ("ground truth") for this data (which will be unknown to the model and only used for validation) corresponding to the dates and times of baseball games. Note that the "on-ramp events" correspond to traffic leaving at the end of a baseball game when large numbers of individuals leave the stadium and get on the freeway—thus, the event has a signature in the data that will tend to lag in time that of the baseball game itself.

Both data sets included a small number of holidays (1-3) which were removed before modeling, since these days were known apriori to involve relatively different (atypical) behavior. (Treating these days as normal tends to slightly decrease their respective days' profiles and may increase probabilities of false alarms, etc.) Alternatively, the model described later could be augmented to estimate the profile of holiday behavior separately, if desired.

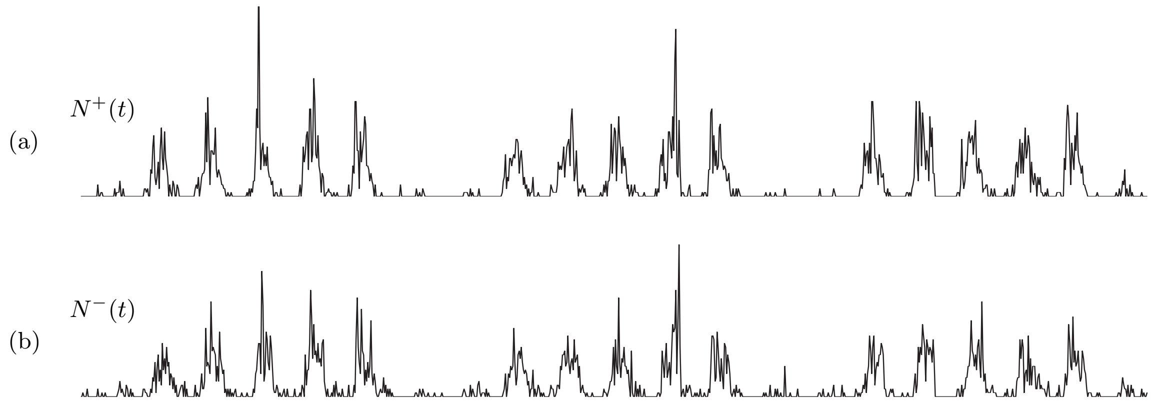

Figure 3: (a) Entry data for the main entrance of the Calit2 building for three weeks, beginning 7/23/05 (Sunday) and ending 8/13/05 (Saturday). (b) Exit data for the same door over the same time period.

4. A BASELINE MODEL AND ITS LIMITATIONS

One relatively straightforward baseline for detecting unusual events in count data is to perform a simple threshold test based on a Poisson model for each time period. Specifically, let us estimate the Poisson rate λ of a particular time and day by averaging the observed counts on similar days (e.g., Mondays) at the same time, i.e., the maximum likelihood estimate. Then, we detect an event of increased activity when the observed count N is sufficiently greater than the average, as measured by the Poisson distribution:

$$\mathrm{P}(N;\lambda) = e^{-\lambda}\lambda^N/N! < \epsilon$$

and $\lambda < N$.

For some data sets, this approach can be quite adequate—in particular, if the events interspersed in the data are sufficiently few compared to the amount of non–event observations, and if they are sufficiently noticeable in the sense that they cause a large increase in activity. However, these assumptions do not always hold, and we can observe several modes of failure in such a simple model.

One way this model can fail is because of the "chicken and egg" problem referred to in the introduction and illustrated in Figure 2. As discussed earlier, the presence of large events distorts the estimated rate of "normal" behavior, increasing it slightly, which causes the threshold test to miss the presence of other events around that same time.

A second type of failure occurs when there is a slight increase in traffic level which is not of sufficient magnitude to be noticed; however, the increase is *sustained* over a period of several observations signaling the presence of a persistent event. In Figure 5, the event indicated for the first day can be easily found by the threshold model by setting the threshold sufficiently high enough to detect the event but low enough so that there are no false alarms. In order for the threshold model to detect the event on the second day, however, the threshold must be increased, which also causes the detection of two false alarms over the two-day period. Anomalies detected by the threshold model are shown in the second panel of the figure while known events (baseball games) are displayed in the third panel.

A third weakness of the threshold model is its difficulty in capturing the duration of an event. In order to detect

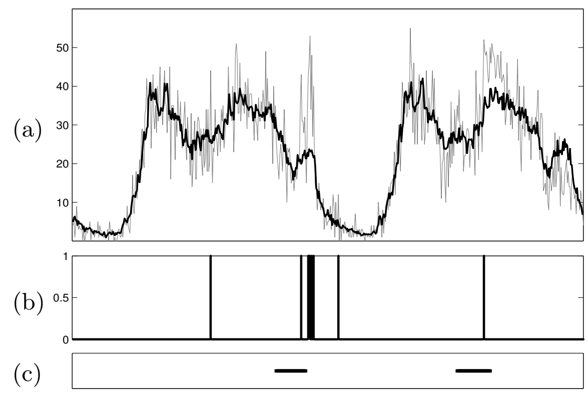

Figure 5: Illustration of the baseline threshold model set to detect the event on the second day, with (a) original freeway traffic time series in the top panel (light curve) for May 17-18, and mean profile as used by the threshold model (dark curve), (b) events detected by the threshold method in the center panel, and (c) ground truth (known events) in the bottom panel. Note the false alarms.

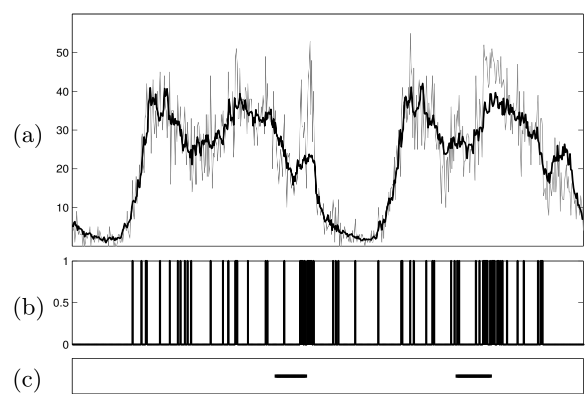

Figure 6: Same as Figure 5 but with an even lower threshold to detect the full duration of the large event on the second day, causing multiple false alarms.

not only the presence of the event on the second day but also its duration, the threshold must be raised to the point that the number of false alarms becomes quite prohibitive, as illustrated in Figure 6. (Note that the traffic event, corresponding to people departing the game, begins at or near the end of the actual game time.)

In the remaining sections of the paper we will discuss a more sophisticated probabilistic model that accounts for these different aspects of the problem, and show (in Section 7) that it can be used to obtain significantly more accurate detection performance than the simple thresholding method.

5. PROBABILISTIC MODELING

Let $N(t)$, for $t \in \{1, \ldots, T\}$, generically refer to the observed count at time t for any of the time-dependent counting processes, such as the freeway traffic 5-minute aggregate count process or either of the two (entering or exiting) building 30-minute aggregate people count processes. In order to model $N(t)$, we require both a model of the "normal", typical behavior (intuitively corresponding to the periodic portion of the data), and a model of the event process (intuitively corresponding to rare increases in the number of observed counts). Let us assume that the two processes are additive, so that

$$N(t) = N_0(t) + N_E(t) \tag{1}$$

where $N_0(t)$ is the number of occurrences attributed to the normal building occupancy, and $N_E(t)$ represents the number of occurrences attributed to an event at time t. We discuss modeling each of these in turn. Note that, although the models described here are defined for discrete time periods, they can also be extended to continuous time measurements [6, 8].

5.1 Modeling Periodic Count Data

Perhaps the most common probabilistic model for count data is the Poisson distribution, whose probability mass function is given by

$$\mathrm{P}(N; \lambda) = e^{-\lambda} \lambda^N / N! \qquad N = 0, 1, \ldots \tag{2}$$

where the parameter λ represents the rate, or average number of occurrences in a fixed time interval. When λ is a function of time, i.e. $\lambda(t)$, (2) becomes a nonhomogeneous Poisson distribution, in which the degree of heterogeneity depends on the function $\lambda(t)$.

We employ a model derived from that of Scott [8], which has been used to detect and segment fraud patterns in telephone network usage [6]. Specifically, we decompose $\lambda(t)$ as

$$\lambda(t) = \lambda_0 \; \delta_{d(t)} \; \eta_{d(t),h(t)} \tag{3}$$

where $d(t)$ takes on values $\{1, \ldots, 7\}$ and indicates the day on which time t falls (so that Sunday = 1, Monday = 2, and so forth), and $h(t)$ indicates the interval (e.g., half-hour periods for the building data) in which time t falls. By further requiring that

$$\sum_{i=1}^{7} \delta_i = 7 \qquad \text{and} \qquad \sum_{i=1}^{D} \eta_{j,i} = D \quad \forall j,$$

where D is the number of time intervals in a day (48 for the building data and 288 for the freeway traffic data), we can

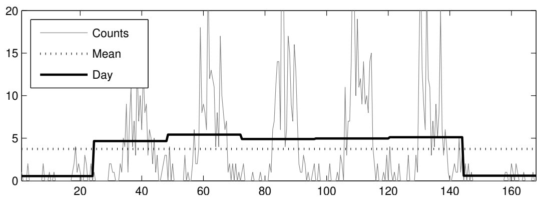

Figure 7: The effect of $\delta_{d(t)}$, as seen over a week of building exit data. Clearly, the relative rates over the weekend (Sunday, Saturday) are much lower than those on weekdays.

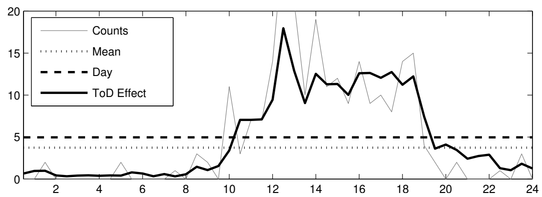

Figure 8: The effect of $\eta_{d(t),h(t)}$ in modulating the Poisson rate of building exit data over a single day. There is a clear peak around lunchtime, and a heavy bias towards the end of the day.

ensure that the values λ_0, δ, and η are easily interpretable: λ_0 is the average rate of the Poisson process over a full week, δ_i is the *day effect*, or the relative change for day i (so that, for example, Sundays have a lower rate than Mondays), and $\eta_{j,i}$ is the relative change in time period i given day j (the *time of day* effect).

Figures 7–8 illustrate these two effects for the building data. Figure 7 shows one week's worth of data alongside the estimated rate with day effect only, i.e., $\lambda_0 \, \delta_{d(t)}$; this is the full Poisson rate $\lambda(t)$ averaged over the time of day. Figure 8 then shows how $\eta_{d(t),h(t)}$ then modulates $\lambda(t)$ over a single day to achieve a sensible time–dependent rate value.

Figure 9 shows a graphical model in the form of a plate diagram for the periodic data $N_0(t)$ and its parameters. A key point is that, given $N_0(t)$, the parameters λ_0, δ, and η are all independent of $N(t)$.

By choosing conjugate prior distributions for these variables we can ensure that the inference computations in Section 6 have a simple closed form:

$$\lambda_0 \sim \Gamma(\lambda; a^L, b^L)$$
$$\frac{1}{7}\,[\delta_1, \ldots, \delta_7] \sim \mathrm{Dir}(\alpha_1^d, \ldots, \alpha_7^d)$$
$$\frac{1}{D}\,[\eta_{j,1}, \ldots, \eta_{j,D}] \sim \mathrm{Dir}(\alpha_1^h, \ldots, \alpha_D^h)$$

where Γ is the Gamma distribution,

$$\Gamma(\lambda; a, b) \propto \lambda^{a-1} e^{-b\lambda}$$

and $\mathrm{Dir}(\cdot)$ is a Dirichlet distribution with the specified parameter vector.

5.2 Modeling Rare, Persistent Events

In the data examined in this paper, the anomalous measurements can be intuitively thought of as being due to relatively short, rare periods in which an additional random

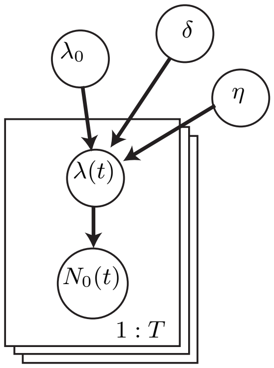

Figure 9: Graphical model for $\lambda(t)$ and $N_0(t)$. The parameters λ_0, δ, and η (the periodic components of $\lambda(t)$) couple the distributions over time.

process also contributes to the observations (e.g., people arriving for an event), increasing the number of observed counts. The model can be easily modified to capture alternative situations, i.e., the presence of an event suppressing or otherwise altering the number of "normal" counts by changing (1) into a more general relationship; however, in practice our simple additivity assumption seems sufficient.

To model the behavior of anomalous periods of time, we use a binary process $z(t)$ to indicate the presence of an event, i.e.,

$$z(t) = \begin{cases} 1 & \text{if there is an event at time } t \\ 0 & \text{otherwise;} \end{cases}$$

and define the probability distribution over $z(t)$ to be Markov in time, with transition probability matrix

$$M_z = \begin{pmatrix} 1 - z_0 & z_1 \\ z_0 & 1 - z_1 \end{pmatrix}$$

so that the length of each time period between events is geometric with expected value $1/z_0$, and the length of each event is geometric with expected value $1/z_1$. We give z_0, z_1 priors

$$z_0 \sim \beta(z; a_0^Z, b_0^Z) \qquad\qquad z_1 \sim \beta(z; a_1^Z, b_1^Z)$$

where $\beta(\cdot)$ is the Beta distribution.

Given $z(t)$, we can model the increase in observation counts due to the event, $N_E(t)$, as Poisson with rate $\gamma(t)$

$$N_E(t) \sim \begin{cases} 0 & z(t) = 0 \\ \mathrm{P}(N; \gamma(t)) & z(t) = 1 \end{cases}$$

and $\gamma(t)$ as independent at each time t

$$\gamma(t) \sim \Gamma(\gamma; a^E, b^E).$$

In fact, $\gamma(t)$ may be marginalized over analytically, since

$$\int \mathrm{P}(N; \gamma)\Gamma(\gamma; a^E, b^E) = \mathrm{NBin}(N; a^E, b^E/(1 + b^E)) \qquad (4)$$

where NBin is the negative binomial distribution. A graphical model representing the distribution over $z(t)$, $N_E(t)$, and $N(t)$ is shown in Figure 10. Here, $z(t)$ provides the time–dependent structure of the process; from Figures 9–10, one can see that $N(t)$ has temporal structure both from $\lambda(t)$ and $z(t)$.

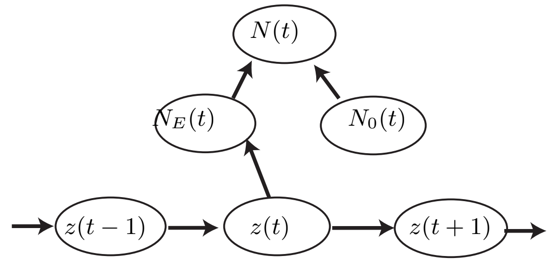

Figure 10: Graphical model for $z(t)$ and $N(t)$. The Markov structure of $z(t)$ couples the variables over time (in addition to the coupling of $N_0(t)$ from Figure 9).

This type of gated Poisson contribution, called a Markov–modulated Poisson model, is a common component of many network traffic models [6,9]. In our application we are specifically interested in detecting the periods of time in which Our event process $z(t)$ is active, and perhaps using the rate $\gamma(t)$ or the associated count $N_E(t)$ to provide information about its "popularity". While it is also possible to couple the rates $\gamma(t)$ in order to capture the idea that, for example, two detections at times t and $t+1$ are likely to be related and thus have correlated count increases, we do not address this additional complexity here.

6. LEARNING AND INFERENCE

Let us initially assume that our total length of observation comprises some integral number of weeks, so that $T = 7 * D * W$ for some integer W. Although not strictly necessary, this assumption greatly simplifies the inference procedure for estimating the parameters of the model [6]; nor is it at all restrictive in our setting, since we can always extend a region of interest to cover an integer number of weeks by taking the additional data to be unobserved.

Given the complete data $\{N_0(t), N_E(t), z(t)\}$, it is straightforward to compute maximum a posteriori (MAP) estimates or draw posterior samples of the parameters $\lambda(t)$ and $\{z_0, z_1\}$, since all variables λ_0, δ, η, z_0, and z_1 are conditionally independent (see Figures 9–10, or Section 6.2).

We can thus infer posterior distributions over each of the variables of interest using Markov chain Monte Carlo (MCMC) methods [10,11]. Specifically, we iterate between drawing samples of the hidden variables $\{z(t), N_0(t), N_E(t)\}$ (described in Section 6.1) and the parameters given the complete data (described in Section 6.2). The complexity of each iteration of MCMC is $\mathcal{O}(T)$, linear in the length of the time series, and experimentally converges quite rapidly. These samples can be used to not only to provide a point estimate of the value of each parameter (for example, its posterior mean) but also to gauge the amount of uncertainty about that value.

6.1 Sampling the hidden variables given parameters

Given the periodic Poisson mean $\lambda(t)$ and the transition probability matrix M, it is relatively straightforward to draw a sample sequence $z(t)$ using a variant of the forward–backward algorithm [12]; we provide the necessary equations for completeness. Specifically, in the forward pass we compute, for each $t \in \{1, \ldots, T\}$ the conditional distribution

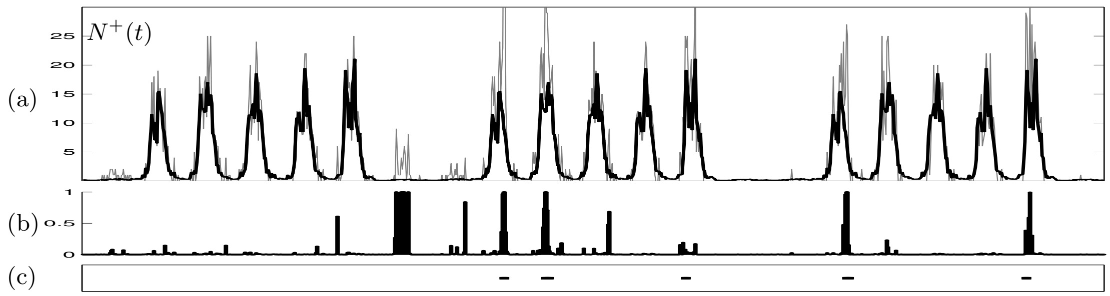

Figure 11: (a) Entry data, along with $\lambda(t)$, over a period of three weeks (Sept. 25–Oct. 15). Also shown are (b) the posterior probability of an event, $p(z)$, and (c) the periods of time in which an event was scheduled for the building. Most of the scheduled events are detected, along with a few other time periods (such as a period of greatly heightened activity on the first Saturday).

$p(z(t)|\{N(t'), t' \le t\})$ using the likelihood functions

$$p(N(t)|z(t)) = \begin{cases} \mathrm{P}(N(t); \lambda(t)) & z(t) = 0 \\ \sum_i \mathrm{P}(N(t) - i; \lambda(t))\, \mathrm{NBin}(i) & z(t) = 1 \end{cases}$$

(where the parameters of NBin$(\cdot)$ are as in (4)). Then, for $t \in \{T, \dots, 1\}$, we draw samples

$$Z(t) \sim p(z(t)|z(t+1) = Z(t+1), \{N(t'), t' \le t\}).$$

Given $z(t) = Z(t)$, we can then determine $N_0(t)$ and $N_E(t)$, by taking $N_0(t) = N(t)$ if $z(t) = 0$ and drawing $N_0(t)$ from the discrete distribution

$$N_0(t) \sim f(i) \propto \mathrm{P}(N(t) - i; \lambda(t))\, \mathrm{NBin}(i; a^E, b^E/(1 + b^E))$$

if $z(t) = 1$, and setting $N_E(t) = N(t) - N_0(t)$. If $N(t)$ is unobserved (missing), $N_0(t)$ and $N_E(t)$ are decoupled given $z(t)$ and we may draw them independently.

6.2 Sampling the parameters given the complete data

Because T is an integral number of weeks, $T = 7 * D * W$, we have that the complete data likelihood is given by

$$\prod_t e^{-\lambda(t)} \lambda(t)^{N_0(t)} \prod_t p(Z(t)|Z(t-1)) \prod_{Z(t)=1} \mathrm{NBin}(N_E(t))$$

Considering the first term, which only involves λ_0, δ, and η, we have

$$e^{-T\lambda_0} \lambda_0^{\sum N_0(t)} \prod_i \delta_i^{\sum_{d(t)=i} N_0(t)} \prod_{j,i} \eta_{j,i}^{(\dots)}$$

By virtue of choosing conjugate prior distributions, we have that the posteriors are given by distributions of the same form, but with parameters given by the sufficient statistics of the data. Defining

$$S_{i,j} = \sum_{t: \substack{d(t)=i, \\ h(t)=j}} N_0(t) \qquad S_i = \sum_j S_{i,j} \qquad S = \sum_i S_i$$

we have the posterior distributions

$$\lambda_0 \sim \Gamma(\lambda; a^L + S, b^L + T)$$

$$\frac{1}{7}\,[\delta_1, \dots, \delta_7] \sim \mathrm{Dir}(\alpha_1^d + S_1, \dots, \alpha_7^d + S_7)$$

$$\frac{1}{D}\,[\eta_{j,1}, \dots, \eta_{j,D}] \sim \mathrm{Dir}(\alpha_1^h + S_{j,1}, \dots, \alpha_D^h + S_{j,D}).$$

Sampling z_0, z_1 is similarly straightforward—we merely compute

$$Z_{ij} = \sum_{t: z(t)=i, z(t+1)=j} 1 \qquad \text{for } i = 0, 1,\ j = 0, 1$$

to obtain posterior distributions

$$z_0 \sim \beta(z; a_0^Z + Z_{01}, b_0^Z + Z_{00}) \quad z_1 \sim \beta(z; a_1^Z + Z_{10}, b_1^Z + Z_{11})$$

As noted by Scott [6], Markov–modulated Poisson processes appear to be relatively sensitive to the selection of prior distributions over z_0, z_1 and $\gamma(t)$, perhaps because there are no direct observations of the processes they describe. This appears to be particularly true for our model, which has considerably more freedom in the anomaly process (i.e., in $\gamma(t)$) than the telephony application of Scott [6]. We avoid over-explanation of the data by applying relatively strong priors to the transition parameters of $z(t)$ which force the marginal probability of $z(t)$ to 1–2 incidents per day, on average. By adjusting these priors one can increase or decrease the number of events detected; see Section 7.

7. ADAPTIVE EVENT DETECTION

One of the primary goals in our application is to automatically detect the presence of unusual events in the observation sequence. The presence or absence of these events is captured by the process $z(t)$, and thus we may use the posterior probability $p(z(t)|\{N(t)\})$ as an indicator of when such events occur.

Given a sequence of data, we can use the samples drawn in the MCMC procedure (Section 6) to estimate the posterior marginal distribution over events. For comparison to a ground truth of the events in the building data set, we obtained a list of the events which had been scheduled over the entire time period from the building's event coordinator. For the freeway traffic data set, the game times for 76 home games in the LA Dodgers 2005 regular season were used as the validation set. Five additional regular season games were not included in this set because they occurred during extended periods of missing loop sensor count information. Note that both sets of "ground truth" may represent an underestimate of the true number of events that occurred (e.g., due to unscheduled meetings and gatherings, concerts held at the baseball stadium, etc.). Nonetheless this ground truth is very useful in terms of measuring how well a model can detect a known set of events.

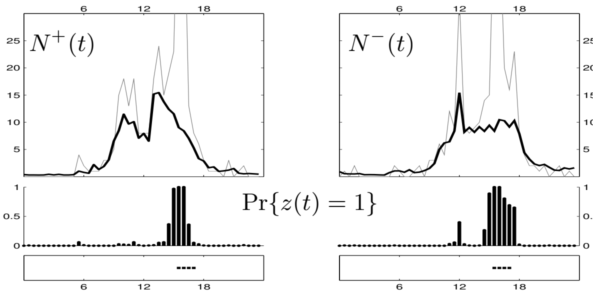

Figure 12: Data for Oct. 3, 2005, along with rate $\lambda(t)$ and probability of event $p(z)$. At 3:30 P.M. an event was held in the building atrium, causing anomalies in both the incoming and outgoing data over most of the time period.

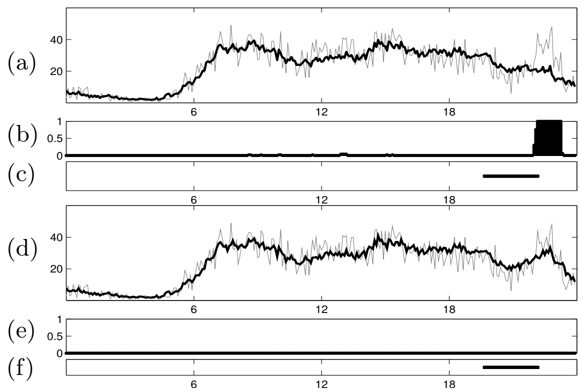

Figure 13: A Friday evening game, Apr. 29, 2005. Shown are (a) the prediction of normal activity, $\lambda(t)$; (b) the estimated probability of an event, $p(z)$; and (c) the actual game time. Panels (d)-(f) show the threshold model's prediction for the same day.

The results obtained by performing MCMC for the building data are shown in Figure 11. We plot the observations $N(t)$ together with the posterior mean of the rate parameters $\lambda(t)$ over a three week period (Sept. 25–Oct. 15); Figure 11 shows incoming (entry) data for the building. Displayed below the time series is the posterior probability of $z(t)$ at each time t, drawn as a sequence of bars, below which dashes indicate the times at which scheduled events in the building took place. In this sequence, all of the known events are successfully detected, along with a few additional detections that were not listed in the building schedule. Such unscheduled activities often occur over weekends where the baseline level of activity is particularly low.

Figure 12 shows a detailed view of one particular day, during which there was an event scheduled in the building atrium. Plots of the probability of an unusual event for both the entering and exiting data show a high probability over the entire period allocated to the event, while slight increases earlier in the day were deemed much less significant due to their relatively short duration.

The results obtained by performing MCMC for the freeway traffic data for three game-days are shown in Figures 13–14. Figure 13 shows a Friday game that is more sparsely attended than the Friday game plotted in Figure 2 and is

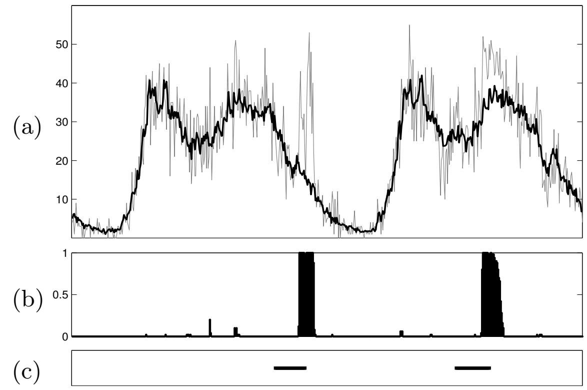

Figure 14: (a) Data for May 17-18,2005, along with rate $\lambda(t)$; (b) probability of event $p(z)$; (c) actual event times.

Total Number of Predicted Events	MMPP Model	Threshold Model
104	100.0%	86.2%
70	96.6%	75.9%
48	79.3%	65.5%

Table 1: Accuracies of predictions for the building data: the percentage of the 29 known events correctly predicted by each model, for different numbers of total events predicted.

an example of where our model successfully separates the normal Friday evening activity from game-day evening activity. The threshold model was able to detect the Friday games with heavy attendance, but more sparsely attended games such as this one were missed.

Figure 14 displays the same two–day period where the threshold model was shown to detect false alarms when the threshold level was set appropriately to detect the event on day two (Figure 5–6). Our model detects the two events with no false alarms, and nicely shows the duration of the predicted events.

Tables 1 and 2 compare the accuracies of the Markov-modulated Poisson process (MMPP) model described in Section 5 and the baseline threshold model of Section 4 on validation data not used in training the models for both the building and freeway traffic data respectively. For each row in the table, the MMPP model parameters were adjusted so that a specific number of events were detected, by adjusting

Total Number of Predicted Events	MMPP Model	Threshold Model
203	100.0%	86.8%
186	100.0%	81.6%
134	100.0%	72.4%
98	98.7%	60.5%

Table 2: Accuracies of predictions for the freeway traffic data: the percentage of the 76 known events correctly predicted by each model, for different numbers of total events predicted.

the priors on the transition probability matrix. The threshold model was then modified to find the same number of events as the MMPP model by adjusting its threshold ϵ.

In both data sets, for a fixed number of predicted events (each row), the number of true events detected by the MMPP model is significantly higher than that of the baseline model. This validates the intuitive discussion of Section 4 in which we outlined some of the possible limitations of the baseline approach, namely its inability to solve the "chicken and egg" problem and the fact that it does not explicitly represent event persistence. As mentioned earlier, the events detected by the MMPP model that are not in the ground truth list in many cases plausibly correspond to real events rather than false alarms, such as unscheduled building activities for the building data and accidents and non-sporting events for the freeway traffic data.

8. TESTING HETEROGENEITY

One question we may wish to ask about the data is, how time–varying is the process itself? For example, how different is Friday afternoon from that of any other weekday? By increasing the number of degrees of freedom in our model, we improve its potential for accuracy but may increase the amount of data required to learn the model well. This also has important consequences in terms of data representation (for example, compression), which may need to be a time–dependent function as well. Thus, we may wish to consider *testing* whether the amount of data we have thus far acquired supports a particular degree of heterogeneity.

We can phrase many of these questions as tests over submodels which require equality among certain subsets of the variables. For example, we may wish to test for the presence of the day effect, and determine whether a separate effect for each day is warranted. Specifically, we might test between three possibilities:

$$D_0 : \ \delta_1 = \ldots = \delta_7 \qquad \text{(all days the same)}$$
$$D_1 : \ \delta_1 = \delta_7, \delta_2 = \ldots = \delta_6 \quad \text{(weekends, weekdays the same)}$$
$$D_2 : \ \delta_1 \neq \ldots \neq \delta_7 \qquad \text{(all day effects separate)}$$

We can compare these various models by estimating each of their marginal likelihoods [13]. The marginal likelihood is the likelihood of the data under the model, having integrated out the uncertainty over the parameter values, e.g.,

$$p(N|D_2) = \int p(N|\lambda_0, \delta, \eta) p(\lambda_0, \delta, \eta) \ \partial\lambda_0 \, \partial\delta \, \partial\eta$$

Since uncertainty over the parameter values is explicitly accounted for, there is no need to penalize for an increasing number of parameters. Moreover, we can use the same posterior samples drawn during the MCMC process (Section 6) to find the marginal likelihood, using the estimate of Chib [14].

Computing the marginal likelihoods for each of the models $D_1, \ldots, D_3$ for the building data, and normalizing by the number of observations T, we obtain the values shown in Table 3. From these values, it appears that D_0 (all days the same) is a considerably worse model, and that D_1 and D_2 are essentially equal, indicating that either model will do an equally good job of predicting behavior.

We can derive similar tests for other symmetries that might exist. For example, we might wonder whether every day has the same time profile. (Note that this is possible, since Sunday might be a severely squashed version of Monday, i.e., fewer people come to work, but they follow a similar hourly pattern.) Alternatively, is each day of the week unique, or (again) might all weekdays be the same, and similarly weekend days? Our tests become

$$T_0 : \ \forall i, \eta_{1,i} = \ldots = \eta_{7,i} \qquad \text{(same time every day)}$$
$$T_1 : \ \forall i, \eta_{1,i} = \eta_{7,i}, \eta_{2,i} = \ldots = \eta_{6,i} \quad \text{(weekends, weekdays)}$$
$$T_2 : \ \forall i, \eta_{1,i} \neq \ldots \neq \eta_{7,i} \qquad \text{(all time effects separate)}$$

The results, shown in Table 4, show a small but distinct preference for T_1, indicating that although weekends and weekdays have differing profiles, one can better predict behavior by combining data across weekdays and weekends. Other tests, such as whether Fridays differ from other days, can be accomplished using similar estimates.

Model	$E[\log_2 p(N^-(t))]$	$E[\log_2 p(N^+(t))]$
D_0	-2.86	-2.58
D_1	-2.55	-2.29
D_2	-2.55	-2.29

Table 3: Average log marginal likelihood of the data (exit and entry) under various day–dependency models: D_0, all days the same; D_1, weekends and weekdays separate; and D_2, each day separate. There does not appear to be a significant change in behavior among weekend days or among weekdays. Parameters $\eta_{i,j}$ were unconstrained.

Model	$E[\log_2 p(N^-(t))]$	$E[\log_2 p(N^+(t))]$
T_0	-2.58	-2.30
T_1	-2.52	-2.27
T_2	-2.55	-2.29

Table 4: Average log marginal likelihood under various time-of-day dependency models for the building data: T_0, all days have the same time profile; D_1, weekend days and weekdays share time profiles; D_2, each day has its own individual time profile. There appears to be a only slight improvement at each stage. Parameters δ_i were unconstrained.

9. ESTIMATING EVENT ATTENDANCE

Along with estimating the probability that an unusual event is taking place, as part of the inference procedure our model also estimates the number of counts which appear to be associated with that event. Marginalizing over the other variables, we obtain a distribution over how many additional people seem to be entering or leaving the building or the number of extra vehicles entering the freeway during a particular time period. One intriguing use for this information is to provide a score, or some measure of popularity, of each event.

As an example, taking our collection of LA Dodgers baseball games, we compute and sum the posterior mean of extra (event-related) vehicles observed, $N_E(t)$, during the duration of the event detection. Figure 15 shows that our estimate of the number of additional cars is positively correlated with the actual overall attendance recorded for the games (correlation coefficient 0.66). Similar attendance scores can

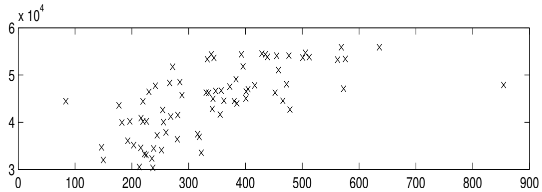

Figure 15: The attendance of each baseball game (y-axis) shows correlation with the number of additional (event–related) vehicles detected by the model (x-axis).

be computed for the building data, or other quantities such as duration estimated, though for these examples no ground truth exists for comparison.

10. CONCLUSION

We have described a framework for building a probabilistic model of time–varying counting processes, in which we observe a superposition of both time–varying but regular (periodic) and aperiodic processes. We then applied this model to two different time series of counts of the number of people entering and exiting through the main doors of a campus building and the number of vehicles entering a freeway, both over several months. We described how the parameters of the model may be estimated using MCMC sampling methods, while simultaneously detecting the presence of anomalous increases in the counts. This detection process naturally accumulates information over time, and by virtue of having a model of uncertainty gives a natural way to compare potentially anomalous events occurring on different days or times. We also showed that the detection can be performed in real–time by fixing the parameter distributions obtained during MCMC and performing a simplified form of forward inference.

Using a probabilistic model also allows us to pose alternative models and test among them in a principled way. Doing so, we can answer questions about how the observed behavior varies over time, and how predictable that behavior is. Finally, we described how the information obtained in the inference process can be used to provide an interesting source of feedback, for example estimating event popularity and attendance.

An interesting direction for future work is to simultaneously model multiple correlated time-series, such as those arising from people counts from multiple doors (and perhaps from multiple different types of sensors) as well as multiple time-series from different loop sensors along a freeway. More sensors provide richer information about occupancy and behavioral patterns, but it is an open question how these co-varying data streams should be combined, and to what degree their parameters can be shared.

Acknowledgments

The authors would like to thank Chris Davison and Anton Popov for their assistance with logistics and data collection, and Shellie Nazarenus for providing a list of scheduled events for the Calit2 building. This work was supported in part by the National Science Foundation under grants No. ITR-0331707 and IIS-0431085.

11. REFERENCES

[1] E. Keogh, S. Lonardi, and B. Y. chi' Chiu, "Finding surprising patterns in a time series database in linear time and space," in *KDD '02: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining.* New York, NY, USA: ACM Press, 2002, pp. 550–556.

[2] V. Guralnik and J. Srivastava, "Event detection from time series data," in *KDD '99: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining.* New York, NY, USA: ACM Press, 1999, pp. 33–42.

[3] M. Salmenkivi and H. Mannila, "Using markov chain monte carlo and dynamic programming for event sequence data," *Knowledge and Information Systems*, vol. 7, no. 3, pp. 267–288, 2005.

[4] J. Kleinberg, "Bursty and hierarchical structure in streams," in *KDD '02: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining.* New York, NY, USA: ACM Press, 2002, pp. 91–101.

[5] S. L. Scott and P. Smyth, "The Markov modulated Poisson process and Markov Poisson cascade with applications to web traffic data," *Bayesian Statistics*, vol. 7, pp. 671–680, 2003.

[6] S. Scott, "Detecting network intrusion using a Markov modulated nonhomogeneous Poisson process," http://www-rcf.usc.edu/~sls/mmnhpp.ps.gz.

[7] Freeway Performance Measurement System (PeMS), "http://pems.eecs.berkeley.edu/."

[8] S. Scott, "Bayesian methods and extensions for the two state Markov modulated Poisson process," Ph.D. dissertation, Harvard University, Dept. of Statistics, 1998.

[9] H. Heffes and D. M. Lucantoni, "A Markov-modulated characterization of packetized voice and data traffic and related statistical multiplexer performance," *IEEE J. Sel. Areas Comm.*, vol. 4, no. 6, pp. 856–868, 1984.

[10] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. PAMI*, vol. 6, no. 6, pp. 721–741, Nov. 1984.

[11] A. E. Gelfand and A. F. M. Smith, "Sampling-based approaches to calculating marginal densities," *J. Amer. Stat. Assoc.*, vol. 85, pp. 398–409, 1990.

[12] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Stat.*, vol. 41, no. 1, pp. 164–171, February 1970.

[13] A. E. Gelfand and D. K. Dey, "Bayesian model choice: asymptotics and exact calculations," *J. R. Stat. Soc. B*, vol. 56, no. 3, pp. 501–514, 1990.

[14] S. Chib, "Marginal likelihood from the Gibbs output," *J. Amer. Stat. Assoc.*, vol. 90, no. 432, pp. 1313–1321, Dec. 1995.