Learning to Detect Events with Markov-Modulated Poisson Processes

Alexander Ihler
Toyota Technological Institute at Chicago,
Jon Hutchins
Department of Computer Science
University of California, Irvine
and
Padhraic Smyth
Department of Computer Science
University of California, Irvine

Time-series of count data occur in many different contexts, including internet navigation logs, freeway traffic monitoring, and security logs associated with buildings. In this paper we describe a framework for detecting anomalous events in such data using an unsupervised learning approach. Normal periodic behavior is modeled via a time-varying Poisson process model, which in turn is modulated by a hidden Markov process that accounts for bursty events. We outline a Bayesian framework for learning the parameters of this model from count time series. Two large real world data sets of time series counts are used as test beds to validate the approach, consisting of freeway traffic data and logs of people entering and exiting a building. We show that the proposed model is significantly more accurate at detecting known events than a more traditional 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 to make inferences about different aspects of events such as number of vehicles or people involved. The results indicate that the Markov-modulated Poisson framework 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

Additional Key Words and Phrases: Event detection, Markov modulated, Poisson

1. INTRODUCTION

Advances in sensor and storage technologies allow us to record increasingly detailed pictures of human behavior. Examples include logs of user navigation and search

Portions of this work have appeared at the ACM Conference on Knowledge Discovery and Data Mining (SIGKDD), 2006.

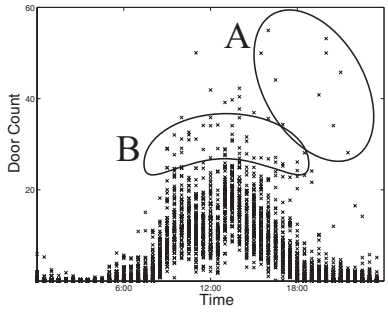


Fig. 1. Jittered scatterplot of the number of people entering on any weekday over a fifteen week period, shown as a function of the time of day (in half-hour intervals). Although certain points (e.g., set A) clearly represent unusual periods of increased activity, it is less clear which, if any, of the values in set B represent something similar.

on the internet, RFID traces, security video archives, and loop-sensor records of freeway traffic. These time series often reflect the underlying hourly, daily, and weekly rhythms of natural human activity. At the same time, the time series are often corrupted by events corresponding to bursty periods of unusual behavior. Examples include anomalous bursts of activity on a network, large occasional meetings in a building, traffic accidents, and so forth.

In this paper we address the problem of identifying such events, by learning the patterns of both normal behavior and events from historical data. While at first glance this problem might seem relatively straightforward, the problem becomes difficult in an unsupervised context when events are not labeled or tagged in the data. Learning a model of normal behavior requires the removal of the abnormal events from the historical record—but detecting the abnormal events can be accomplished reliably only by knowing the baseline of normal behavior. This leads to a "chicken and egg" problem that is the main focus of this paper.

We focus in particular on time series data where time is discrete and $N(t)$ is a measurement of the number of individuals or objects engaged in some activity over the time-interval $[t-1, t]$, e.g., counts of the number of people who enter a building every 15 minutes, or the number of vehicles that pass a certain location on the freeway every 5 minutes. As an example, Figure 1 shows counts of the estimated number of people entering a building over time from an optical sensor at the front door of a UC Irvine (UCI) campus building. The data are "jittered" slightly by Gaussian noise to give a better sense of the density of counts at each time. 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.

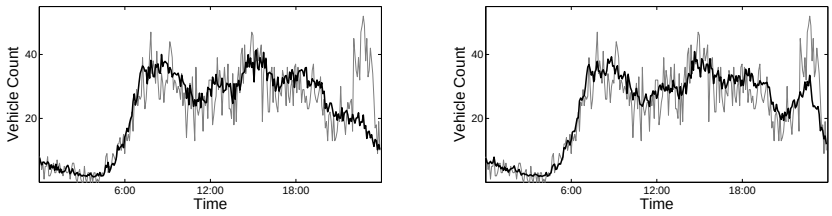


Fig. 2. Example of freeway traffic data for Fridays for a particular on-ramp. (a) Average time profile for normal, non game-day Fridays (dark curve) and data for a particular Friday (6/10/05) with a baseball game that night (light curve). (b) Average time profile over all Fridays (dark curve) superposed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

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 (LA) located near Dodger Stadium, where the LA 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 and occasional additional counts caused by aperiodic events.

2. RELATED WORK AND OUTLINE OF THE PAPER

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. [2002] described a technique that represents a real-valued time series by quantizing into it a finite set of symbols and then used a Markov model to detect surprising patterns in the symbol sequence. Guralnik and Srivastava [1999]

proposed an iterative likelihood-based method for segmenting a time series into piecewise homogeneous regions. Salmenkivi and Mannila [2005] 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 [2002] 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 framework we propose to address this problem is derived from the Markov–modulated Poisson processes used by Scott and Smyth [2003] for analysis of Web surfing behavior and Scott [2002] for telephone network fraud detection. We extend this work by employing a more flexible model of event-related counts as well as allowing for missing data. We adopt a Bayesian approach to learning and inference, allowing us to pose and answer a variety of queries within a probabilistic framework, 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 first describing, in Section 3, the two data sets used throughout the paper: freeway traffic data and entry/exit counts for a building. 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 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 describes how the model can be used to answer other, related inference questions, such as investigating different degrees of time-heterogeneity in the model and estimating event attendance. In Section 9 we conclude with a brief discussion of open research problems and summary comments.

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 fifteen weeks 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

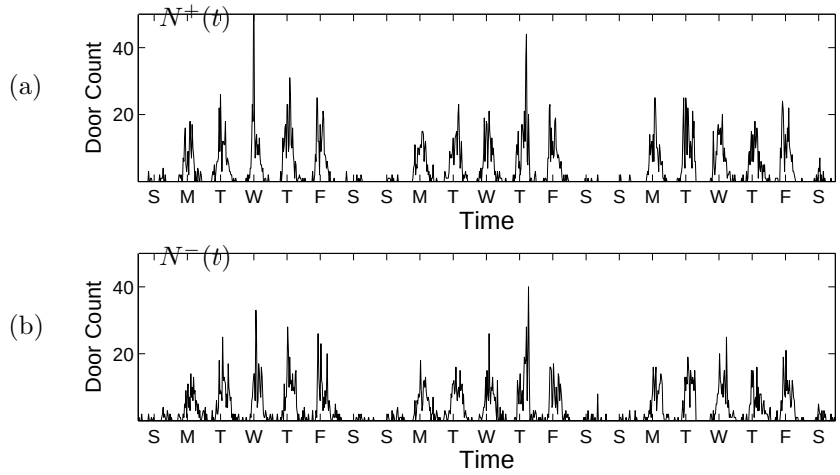


Fig. 3. (a) Entry data for the main entrance of the Calit2 building for three weeks, beginning 7/23/05 (Sunday) and ending 8/13/05 (Saturday). (b) Exit data for the same door over the same time period.

with internet access, where they are stored.

The observation sequences ("door counts") acquired at the front door form a noisy time series with obvious structure but many outliers (see Figure 3). The data are corrupted by the presence of events, namely, 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 additional 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.

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 [Chen et al. 2001]. 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 short periods of time with 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 (ground truth) events for this data (which are unknown to the model and only used for validation) corresponding to the dates and times of baseball games. Note that the baseball-related events at this on-ramp correspond to traffic leaving at the end of a 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.

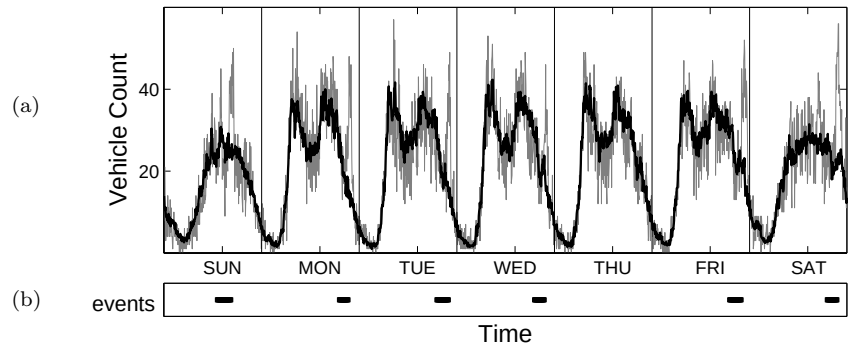


Fig. 4. (a) One week of traffic data (light curve) from Sunday to Saturday (June 5-11), with the estimated normal traffic profile (estimated by the proposed model described later in the paper) superposed as a dark curve. (b) Ground truth list of events (baseball games).

4. SIMPLE POISSON MODELS

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 door count processes.

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

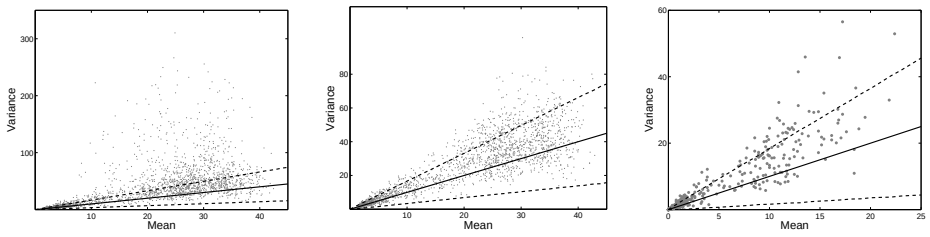
$$\mathrm{P}(N; \lambda) = e^{-\lambda} \lambda^N / N! \qquad N = 0, 1, \ldots \tag{1}$$

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. $\lambda(t)$, (1) becomes a nonhomogeneous Poisson distribution, in which the degree of heterogeneity depends on the function $\lambda(t)$.

4.1 Testing the Poisson Assumption

The assumption of a Poisson distribution may not always hold in practice, and it is reasonable to ask whether it fits the data at hand. For a simple Poisson distribution there are a number of classical goodness-of-fit tests which can be applied, such as the chi-squared test [Papoulis 1991]. However, in this case we typically have a large number of potentially different rates, each of which has only a few observations associated with it. For example in the freeway traffic data, there are 2016 five-minute time slices in a week, each of which may be described by a different rate, and for each time-slice there are between 22 and 25 non-missing observations over the 25 weeks of our study.

Although classical extensions for hypothesis testing exist in such situations [Svensson 1981], here we use a more qualitative approach since our goal is simply to assess whether the Poisson assumption is reasonable for the data. Specifically, we know that, if the data are Poisson, the true mean and variance of the observations at each time slice should be equal. Given finite data, the empirical mean and variance at each time provide noisy estimates of the true mean and variance, and we can visually assess whether these estimates are close to equal by comparing them to the



(a) (b) (c)

Fig. 5. (a) Scatter-plot of empirical (mean,variance) pairs observed in the data, compared with the theoretical distribution expected for Poisson-distributed random variables (dashed lines show ±2-sigma interval). The presence of events makes these distributions quite dissimilar. (b) After removing about 5% of data thought to be influenced by events, the data show much closer correspondence to the Poisson assumption. (c) Building data after removing about 5% of observations thought to have events present, along with corresponding confidence intervals.

distribution expected for Poisson-distributed data.

This comparison, performed using all observations, is shown in Figure 5(a)—empirical estimates of observed (mean,variance) pairs at each time slice for the traffic data are shown as a scatter-plot, while the two-sigma confidence region for the null hypothesis (that the data are Poisson) is demarcated by dashed lines. Visually, the two distributions seem quite different, indicating that the raw data is itself not Poisson. However, this should not be surprising, since the raw data is corrupted by the presence of occasional bursts of events.

A better evaluation of the model's appropriateness is given by first running the Markov-modulated Poisson model described in the sequel, then using the results to remove the fraction of the data thought to be anomalous. After removing only about 5% of the data, the scatter-plot looks considerably more consistent with the Poisson hypothesis, as shown in Figure 5(b). Visually, one can see that the data may be slightly over-dispersed (higher variance than mean), a common issue in count data, but that the approximation looks fairly reasonable. Figure 5(c) shows the same plot for the 15 weeks of building data (including points for both entry and exit data). Here too, the distribution of (mean, variance) pairs appears to be reasonable in the context of our primary goal of event detection, although again there is evidence that the data are slightly over-dispersed compared to that of a Poisson distribution.

4.2 A Baseline Model and its Limitations

Given that the non-event observations are close to Poisson, 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 different than the average, as indicated by having low probability under the estimated Poisson distribution, $\mathrm{P}(N;\lambda) < \epsilon$.

For some data sets, this approach can be quite adequate—in particular, if the

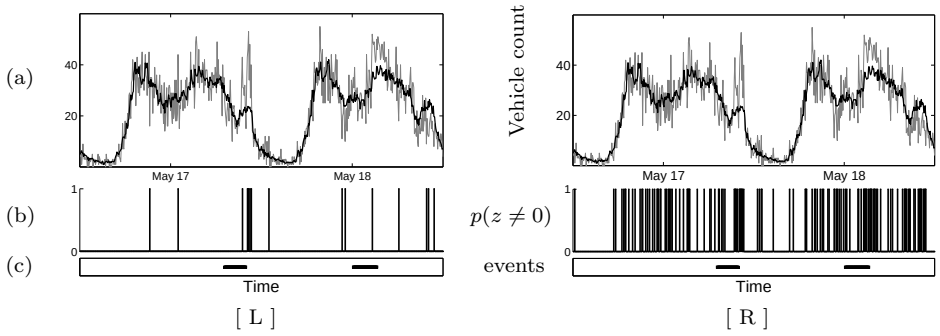


Fig. 6. [L]: Illustration of the baseline threshold model set to detect the event on the second day, with (a) original freeway traffic time series (light curve) for May 17-18, and mean profile as used by the threshold model (dark curve), (b) events detected by the threshold method, and (c) ground truth (known events) in the bottom panel. Note the false alarms. [R]: Using a lower threshold to detect the full duration of the large event on the second day, causing many more false alarms.

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 dramatic change 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 due to the "chicken and egg" problem mentioned in the introduction and illustrated in Figure 2. As discussed earlier, the presence of large events distorts the estimated rate of "normal"behavior, 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 change in traffic level which is not of sufficient magnitude to be noticed; however, the change is *sustained* over a period of several observations signaling the presence of a persistent event. In Figure 6[L], 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 decreased, which also causes the detection of a few false alarms over the two-day period. Anomalies detected by the threshold model are shown in Figure 6[L](b), while the known events (baseball games) are displayed in panel (c).

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

Section 4, and in particular the failures of Figure 6, motivate the use of a probabilistic model capable of reasoning simultaneously about the rate of normal behavior (intuitively corresponding to the periodic portion of the data) and the presence and duration of events (relatively rare deviations from the norm). Let us assume that the two processes are additive, so that

$$N(t) = N_0(t) + N_E(t), \qquad N(t) \geq 0 \tag{2}$$

where $N_0(t)$ is the number of occurrences attributed to the normal building occupancy, and $N_E(t)$ represents the change in the number of occurrences which is attributed to an event at time t (positive or negative); the non-negativity condition indicates that we cannot observe fewer than zero counts. We discuss modeling each of the variables N_0, N_E in turn. Note that, although the models described here are defined for discrete time periods, it may also be possible to extend them to continuous time measurements [Scott 1998; 2002].

5.1 Modeling Periodic Count Data

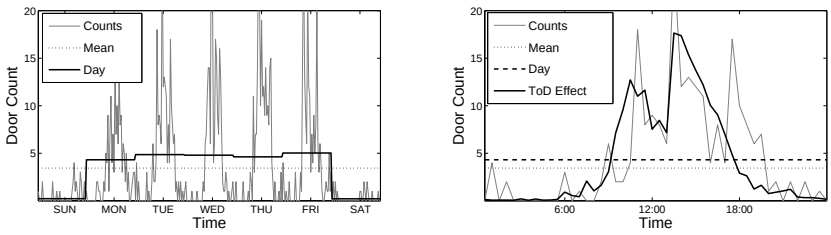
To model the periodic, predictable portion of behavior corresponding to normal activity, we use a nonhomogeneous Poisson process (see Section 4) with a particular parameterization of the rate $\lambda(t)$; our model is derived from that of Scott [1998], and has been used to detect and segment fraud patterns in telephone network usage [Scott 2002]. 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$\sum_{j=1}^{7} \delta_j = 7$ and $\sum_{i=1}^{D} \eta_{j,i} = D \quad \forall j$, 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 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_j is the *day effect*, or the relative change for day j (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).

Figure 7 illustrates these two effects for the building data. Figure 7(a) 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 7(b) then shows how $\eta_{d(t),h(t)}$ then modulates $\lambda(t)$ over a single day to achieve a sensible time–dependent rate value.

Graphical models provide a useful and general formalism for characterizing the dependence structure among a set of random variables [Jordan 1998]. In a directed graphical model or Bayesian network, directed edges indicate a factorization of the joint distribution into a product of conditionals, in which each node depends only on the values of its parents in the graph. Figure 8(a) shows a graphical model in the form of a plate diagram for the periodic data $N_0(t)$ and associated parameters. The plate notation (shown as a rectangle in Figure 8(a)) is used in graphical models to indicate sets of replicated variables that are conditionally independent given



(a) (b)

Fig. 7. (a) The effect of $\delta_{d(t)}$, as seen over a week of building exit data. The relative rates over the weekend (Sunday, Saturday) are much lower than those on weekdays. (b) The effect of $\eta_{d(t),h(t)}$ in modulating the Poisson rate of building exit data over a single day. There is a noticeable peak around lunchtime, and a heavy bias towards the end of the day.

common parent nodes that are outside the plate [Buntine 1994]. The plates indicate that there are multiple variables $\lambda(t)$ and $N_0(t)$, one for each value of $t \in \{1 \dots T\}$, and that the $\lambda(t)$ variables are conditionally independent of each other given λ_0, σ, and η. 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:

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

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 process changes the observed behavior, increasing or decreasing the number of observed counts. In cases of increased activity ("positive" events), these deviations may arise from the presence of some cause (e.g., people arriving for an event in the building), while decreased activity patterns ("negative" events) can be thought of as a suppression or removal of counts which would normally have been observed.

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

$$z(t) = \begin{cases} 0 & \text{if there is no event at time } t \\ +1 & \text{if there is a positive event} \\ -1 & \text{if there is a negative event} \end{cases}$$

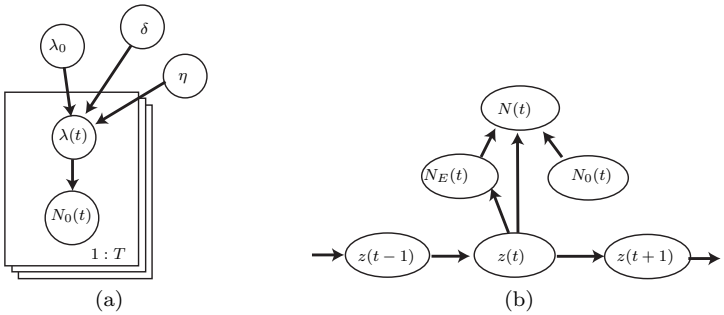


Fig. 8. (a) Graphical model for $\lambda(t)$ and $N_0(t)$. The parameters λ_0, δ, and η (the periodic components of $\lambda(t)$) couple the distributions over time. (b) Graphical model for $z(t)$ and $N(t)$. The Markov structure of $z(t)$ couples the variables over time [in addition to the coupling of $N_0(t)$ from (a)].

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

$$M_z = \begin{pmatrix} z_{00} & z_{0+} & z_{0-} \\ z_{+0} & z_{++} & z_{+-} \\ z_{-0} & z_{-+} & z_{--} \end{pmatrix}$$

with each row summing to one, e.g., $z_{00} + z_{0+} + z_{0-} = 1$. These variables can be interpreted in terms of intuitive characteristics of the system; for example, the length of each time period between events is geometric with expected value $1/(1 - z_{00})$, the length of each positive event is geometric with expected value $1/(1-z_{++})$, and so forth. We give the transition probability variables priors specified as

$$[z_{00},\, z_{0+},\, z_{0-}] \sim \mathrm{Dir}(z\,;\, [a^Z_{00},\, a^Z_{0+},\, a^Z_{0-}])$$

and similarly for the other matrix rows, where $\mathrm{Dir}(\cdot)$ is again the Dirichlet distribution.

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

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

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

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

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 8(b). Here, $z(t)$ provides the time–dependent structure of the process; from Figure 8(a)–(b), one can see that $N(t)$ has temporal structure both from $\lambda(t)$ and $z(t)$.

This type of gated Poisson contribution, called a Markov–modulated Poisson model, is a common component of many network traffic models [Heffes and Lucantoni 1984; Scott 2002]. In our application we are specifically interested in detecting the periods of time in which the event process $z(t)$ is active, and we can use 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 [Scott 2002]. In fact it is not 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 M_z, since all variables λ_0, δ, η, and M_z are conditionally independent (see Figure 8 or Section 6.2).

We can thus infer posterior distributions over each of the variables of interest using Markov chain Monte Carlo (MCMC) methods [Geman and Geman 1984; Gelfand and Smith 1990]. 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. Experimentally we have found that the sampler converges quite rapidly on the data sets used in this paper, where convergence is informally assessed by monitoring the parameter values and values of the marginal likelihood. For both data sets used in this paper, 10 burn-in iterations followed by 50 more sampling iterations appeared quite sufficient. In practice, on a 3GHz Xeon desktop machine in Matlab, the building data (15 weeks of 30-minute intervals) took about 3 minutes while the traffic data (25 weeks of 5-minute intervals) took about one hour. The samples obtained from MCMC 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. If this degree of uncertainty is not of interest (for example, if the data are sufficiently many that the uncertainty is very small) we could use an alternative method such as expectation–maximization (EM) to learn the parameter values [Buntine 1994].

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 [Baum et al. 1970]. We provide below the necessary equations for completeness. Specifically, in the forward pass we compute, for each $t \in \{1, \ldots, T\}$ the conditional distribution $p(z(t)|\{N(t'), t' \leq t\})$ using the

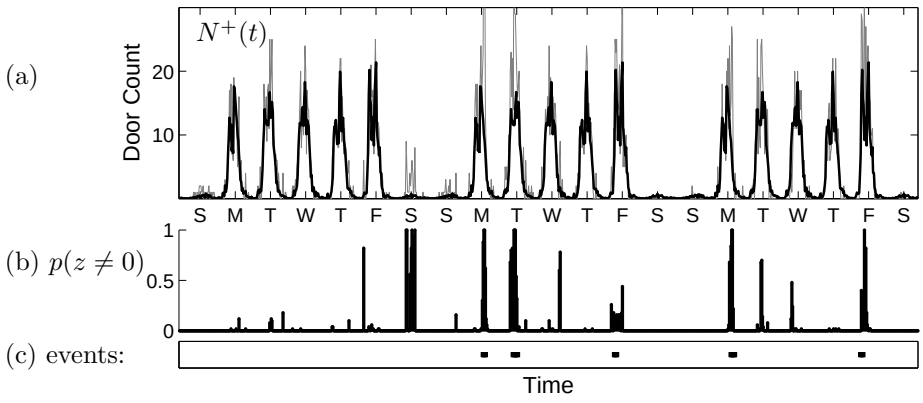


Fig. 9. (a) Entry data, along with $\lambda(t)$, over a period of three weeks (Sept. 25–Oct. 15). Also shown are (b) the posterior probability of an event being present, $p(z(t) \neq 0)$, and (c) the periods of time in which an event was scheduled for the building. All but one of the scheduled events are detected, along with a few other time periods (such as a period of greatly heightened activity on the first Saturday).

likelihood functions

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

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

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

Given $z(t) = Z(t)$, we can then determine $N_0(t)$ and $N_E(t)$ by sampling. If $z(t) = 0$, we simply take $N_0(t) = N(t)$; if $z(t) = +1$ we draw $N_0(t)$ from the discrete distribution

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

and if $z(t) = -1$ from the distribution

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

then setting $N_E(t) = N(t) - N_0(t)$. Note that, if $z(t) = +1$, N_0 takes values in $\{0 \ldots N\}$; if $z(t) = -1$, however, N_0 has no fixed upper limit. In practice, for computational efficiency we truncate the distribution (imposing an upper limit) at the point given by $\mathrm{P}(N(t)+i;\lambda(t)) < 10^{-4}$.

When $N(t)$ is unobserved (missing), $N_0(t)$ and $N_E(t)$ are coupled only through $z(t)$ and the positivity condition on $N(t)$. Thus, when $z(t) \neq -1$ (positive or no event), N_0 and N_E can be drawn independently, and when $z(t) = -1$ (negative event) they can be drawn fairly easily through rejection sampling, i.e., repeatedly drawing the variables independently until they satisfy the positivity condition. Overall, missing data are relatively rare, with essentially no observations missing

in the building data and about 7% of observations missing in the traffic data (due to loop sensor errors or down-time).

6.2 Sampling the Parameters Given the Complete Data

Because T is an integral number of weeks, $T = 7 * D * W$, the complete data likelihood is

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

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

$$e^{-T\lambda_0} \lambda_0^{\sum N_0(t)} \prod_j \delta_j^{\sum_{d(t)=j} N_0(t)} \prod_{j,i} \eta_{j,i}^{(\ldots)}$$

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

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

the posterior distributions are

$$\begin{aligned} \lambda_0 &\sim \Gamma(\lambda; a^L + S, b^L + T) \\ \frac{1}{7}\,[\delta_1, \ldots, \delta_7] &\sim \mathrm{Dir}(\alpha_1^d + S_1, \ldots, \alpha_7^d + S_7) \\ \frac{1}{D}\,[\eta_{j,1}, \ldots, \eta_{j,D}] &\sim \mathrm{Dir}(\alpha_1^h + S_{j,1}, \ldots, \alpha_D^h + S_{j,D}). \end{aligned}$$

Sampling the transition matrix parameters $\{z_{00},\ z_{0+},\ \ldots z_{--}\}$ is similarly straightforward—we compute

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

to obtain the posterior distribution

$$[z_{00},\ z_{0+},\ z_{0-}] \sim \mathrm{Dir}(z; [a_{00}^Z + Z_{00},\ a_{0+}^Z + Z_{0+},\ a_{0-}^Z + Z_{0-}])$$

and similar forms for the other z_{ij}. As noted by Scott [2002], Markov–modulated Poisson processes appear to be relatively sensitive to the selection of prior distributions over the z_{ij} 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 [2002]. However, for an event detection application such as those under consideration, we have fairly strong ideas of what constitutes a "rare" event, e.g., approximately how often we expect to see events occur (say, 1–2 per day) and how long we expect them to last (perhaps an hour or two). We can leverage this information to form relatively strong priors on the transition parameters of $z(t)$ which force the marginal probability of $z(t)$. This avoids over-explanation of the data, such as using the event process to compensate for the fact

that the "normal" data exhibits slightly larger than expected variance for Poisson data (see Section 4.1). By adjusting these priors one can also increase or decrease the model's sensitivity to deviations and thus 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 $p(z(t) \neq 0 | \{N(t)\})$ 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 78 home games in the LA Dodgers 2005 regular season were used as the validation set. Three 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.

The results obtained by performing MCMC for the building data are shown in Figure 9. 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 9 shows incoming (entry) data for the building. Displayed below the time series is the posterior probability of $z(t) \neq 0$ 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 10 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 11–12. Figure 11 shows a Friday game that is more sparsely attended than the Friday game plotted in Figure 2 and provides an example in which 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 12 displays the same two–day period as Figure 6, where the threshold model was shown to detect false alarms when the threshold level was set low enough

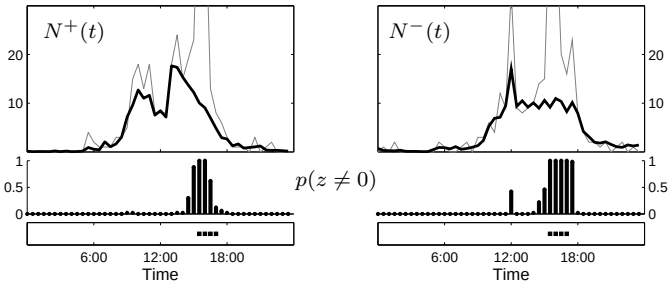


Fig. 10. Data for Oct. 3, 2005, along with rate $\lambda(t)$ and probability of event $p(z \neq 0)$. At 3:30 P.M. an event was held in the building atrium, causing anomalies in both the incoming and outgoing data over most of the time period.

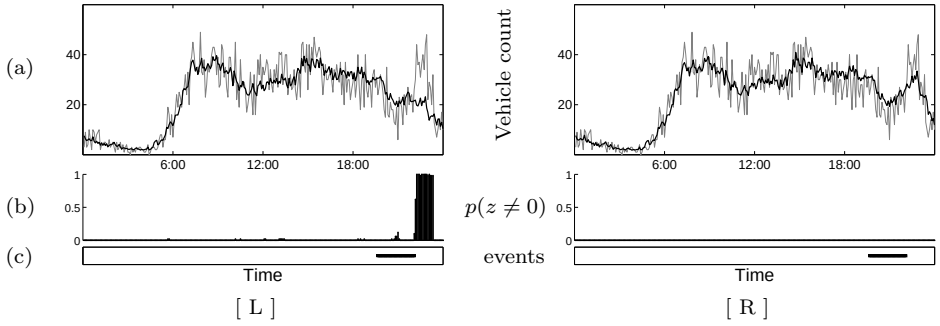


Fig. 11. [L]: A Friday evening game, Apr. 29, 2005. Shown are (a) the prediction of normal activity, $\lambda(t)$; (b) the estimated probability of an event, $p(z \neq 0)$; and (c) the actual game time. [R]: The threshold model's prediction for the same day.

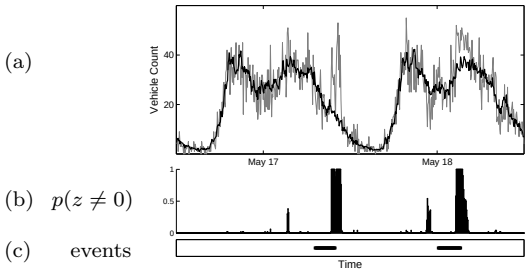


Fig. 12. (a) Freeway data for May 17-18,2005, along with rate $\lambda(t)$; (b) probability of event $p(z \neq 0)$; (c) actual event times.

to detect the event on day two. Our model detects both events with no false alarms, and nicely shows the duration of the predicted events.

Table I compares the accuracies of the Markov-modulated Poisson process (MMPP) model described in Section 5 and the baseline threshold model of Section 4.2 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

Building Data:			**Freeway Data:**		
Total # of Predicted Events	MMPP Model	Threshold Model	Total # of Predicted Events	MMPP Model	Threshold Model
149	100.0%	89.7%	355	100.0%	85.9%
98	86.2%	82.8%	264	100.0%	82.1%
68	79.3%	65.5%	154	100.0%	66.7%
			129	97.4%	55.1%

Table I. Accuracies of predictions for the two data sets, in terms of the percentages of known events found by each model, for different total numbers of events predicted. There were 29 known events in the building data, and 78 in the freeway data.

were adjusted so that a specific number of events were detected, by adjusting 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 4.2 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 may plausibly correspond to real events rather than false alarms, such as unscheduled activities for the building data and accidents and non-sporting events for the freeway traffic data.

Negative events (corresponding to lower than expected activity) tend to be more rare than positive events (higher than expected activity), but can play an important role in the model. For example, in the building data the presence of holidays (during which very little activity is observed) can corrupt the estimates of normal behavior. If known, these days can be explicitly removed (marked as missing) [Ihler et al. 2006], but by treating such periods as negative events the model can be made robust to such periods. Although negative events are quite rare in the building data (comprising about 5% of the events detected), they are more common in the traffic data set (about 40% of events). Figure 13 shows a typical example of a negative traffic event. Here, only three cars were observed on the ramp during a 15 minute period with normally high traffic followed by a 30 minute period with much higher than normal activity. We might speculate that the initial negative event could be due to an accident or construction, which shut down the ramp for a short period, followed by a positive event during which the resulting build-up of cars were finally allowed onto the highway.

8. OTHER INFERENCES

Given that our model is capable of detecting and separating the influence of unusual periods of activity, we may also wish to use the model to estimate other quantities of interest. For example, we can separate out the normal patterns of behavior and use a goodness-of-fit test to answer questions about the degree of heterogeneity in the data and thus the underlying human behavior. Alternatively, we might wish to use our estimates of the *amounts* of abnormal behavior to infer other, indirectly

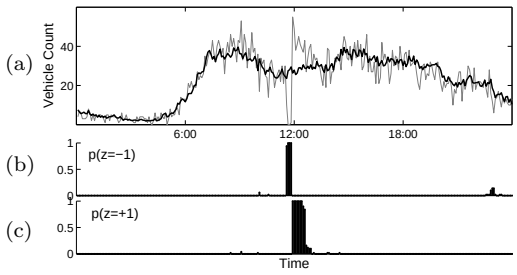


Fig. 13. Negative and positive events: (a) Freeway data and estimated profile for May 6; (b) when the number of observed cars drops sharply, the probability of a negative event is high; (c) the decrease is followed by a short but large increase in traffic, detected as a positive event.

related aspects of the event, such as its popularity or importance. We discuss each of these cases next.

8.1 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 data we have acquired thus far supports a particular degree of heterogeneity.

We can phrase many of these questions as tests over sub-models 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:

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

We can compare these various models by estimating each of their marginal likelihoods [Gelfand and Dey 1990]. 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 [1995].

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 II. From these values, it appears that D_0 (all days the same)

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 II. 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 III. Average log marginal likelihood under various time-of-day dependency models for the building data: T_0, all days have the same time profile; T_1, weekend days and weekdays share time profiles; T_2, each day has its own individual time profile. The model appears to slightly prefer T_1, indicating strong profile similarities among weekdays and among weekends. Parameters δ_j were unconstrained.

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

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

The results, shown in Table III, 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.

8.2 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 (or fewer) people seem to be entering or leaving the building or the number of extra (or missing) 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

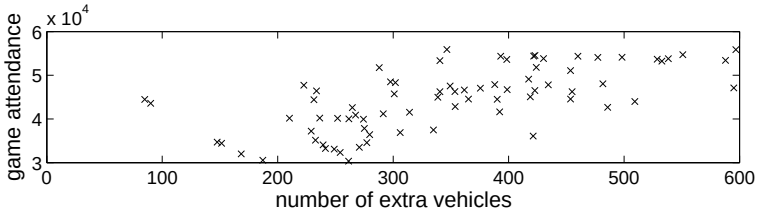


Fig. 14. The attendance of each baseball game (y-axis) shows correlation with the number of additional (event–related) vehicles detected by the model (x-axis).

and sum the posterior mean of extra (event-related) vehicles observed, $N_E(t)$, during the duration of the event detection. Figure 14 shows that our estimate of the number of additional cars is positively correlated with the actual overall attendance recorded for the games (correlation coefficient 0.67). Similar attendance scores can be computed for the building data, or other quantities such as duration estimated, though for these examples no ground truth exists for comparison.

9. 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 or decreases in the counts. This detection process naturally accumulates information over time, and by virtue of having a model of uncertainty provides a natural way to compare potentially anomalous events occurring on different days or times.

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.

Although the current model is very effective in accurately detecting events in the data sets we looked at, it is also a relatively simple model and there are a number of potential extensions that could improve its performance for specific applications. For example, the Poisson parameters for nearby time-periods could be coupled in the model to share information during learning and encourage smoothness in the inferred mean profiles over time. Similarly, it could be valuable to incorporate additional exogenous variables into the proposed model, e.g., allowing both normal and event-based intensities to be dependent on factors such as weather. The event process could be generalized from a Markov to a semi-Markov process to handle events with specific (non-geometric) duration signatures. In principle all of these extensions could be handled via appropriate extensions of the graphical models

and Bayesian estimation techniques described earlier in the paper, with attendant increases in modeling and computational complexity.

A further interesting direction for future work is to simultaneously model multiple correlated time series, such as those arising from door counts from multiple doors (and perhaps from more than one type of sensor) 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 material is based upon work supported in part by the National Science Foundation under Award Numbers ITR-0331707, IIS-0431085, and IIS-0083489.

REFERENCES

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. 1970. A maximization technique occurring in statistical analysis of probabilistic functions of Markov chains. *Annals of Mathematical Statistics 41,* 1 (February), 164–171.

Buntine, W. 1994. Operations for learning with graphical models. *Journal of Artificial Intelligence Research 2*, 159–225.

Chen, C., Petty, K., Skabardonis, A., Varaiya, P., and Jia, Z. 2001. Freeway performance measurement system: mining loop detector data. 80th Annual Meeting of the Transportation Research Board, Washington, D.C. `http://pems.eecs.berkeley.edu/`.

Chib, S. 1995. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association 90,* 432 (Dec.), 1313–1321.

Gelfand, A. E. and Dey, D. K. 1990. Bayesian model choice: asymptotics and exact calculations. *Journal of Royal Statistical Society, Series C 56,* 3, 501–514.

Gelfand, A. E. and Smith, A. F. M. 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association 85*, 398–409.

Geman, S. and Geman, D. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 6,* 6 (Nov.), 721–741.

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

Heffes, H. and Lucantoni, D. M. 1984. A Markov-modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE Journal on Selected Areas in Communications 4,* 6, 856–868.

Ihler, A., Hutchins, J., and Smyth, P. 2006. Adaptive event detection with time–varying Poisson processes. In *KDD '06: Proceedings of the twelfth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM Press, New York, NY, USA, 207–216.

Jordan, M. I., Ed. 1998. *Learning in Graphical Models*. MIT Press, Cambridge, MA.

Keogh, E., Lonardi, S., and chi' Chiu, B. Y. 2002. 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*. ACM Press, New York, NY, USA, 550–556.

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

Papoulis, A. 1991. *Probability, Random Variables, and Stochastic Processes*, 3rd ed. McGraw-Hill Inc., New York, NY.

Salmenkivi, M. and Mannila, H. 2005. Using Markov chain Monte Carlo and dynamic programming for event sequence data. *Knowledge and Information Systems 7,* 3, 267–288.

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

Scott, S. 2002. Detecting network intrusion using a Markov modulated nonhomogeneous Poisson process. `http://www-rcf.usc.edu/~sls/mmnhpp.ps.gz`.

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

Svensson, A. 1981. On a goodness of fit test for multiplicative Poisson models. *The Annals of Statistics 9,* 4, 697–704.