Anomaly Detection in the WIPER System using A Markov Modulated Poisson Distribution

Ping Yan
Tim Schoenharl
Alec Pawling
Greg Madey
Outline

- Background
- WIPER
- MMPP framework
- Application
- Experimental Results
- Conclusions and Future work
Background

- **Time Series**
  - A sequence of observations measured on a continuous time period at time intervals
  - Example: economy (Stock, financial), weather, medical etc...

- **Characteristics**
  - Data are not independent
  - Displays underlying trends
WIPER System

- Wireless Phone-based Emergency Response System

Functions

- Detect possible emergencies
- Improve situational awareness

Cell phone call activities reflect human behavior
Data Characteristics

- **Two cities:**
  - **Small city A:**
    - Population – 20,000  Towers – 4
  - **Large city B:**
    - Population – 200,000  Towers – 31

- **Time Period**
  - Jan. 15 – Feb. 12, 2006
Tower Activity

Small City (4 Towers)
Tower Activity

Large City (31 Towers)
15-Day Time Period Data

Small City
15-Day Time Period Data

Large City
Observations

- Overall call activity of a city are more uniform than a single tower.
- Call activity for each day displays similar trend.
- Call activity for each day of the week shares similar behavior.
MMPP Modeling

\[ N(t) = N_0(t) + N_A(t) \]

- \( N(t) \): Observed Data
- \( N_0(t) \): Unobserved Data with normal behavior
- \( N_A(t) \): Unobserved Data with abnormal behavior

Both \( N_0(t) \) and \( N_A(t) \) can be modulated as a Poisson Process.
Modeling Normal Data

- Poisson distribution

\[
P(N; \lambda) = \frac{e^{-\lambda} \lambda^N}{N!} \quad N = 0, 1, \ldots
\]

- Rate Parameter: a function of time

\[
\lambda \sim \lambda(t)
\]
Adding Day/hour effects

\[ \lambda(t) = \lambda_0 \delta_{d(t)} \eta_{d(t),h(t)} \]

\( d(t) \in [1, 2, ..., 7] \)

Associated with Monday, Tuesday ... Sunday

\( h(t) \) : Time interval, such as minute, half hour, hour etc

\( \lambda_0 \) : Average rate of the Poisson process over one week
Requirements:

\[
\sum_{i=1}^{7} \delta_i = 7 \quad \sum_{j=1}^{D} \eta_{i,j} = D, \quad \forall i
\]

\(\delta_i\) : Day effect, indicates the changes over the day of the week

\(\eta_{i,j}\) : Time of day effect, indicates the changes over the time period \(j\) on a given day of \(i\)
Day Effect

![Graph showing call activities over time intervals with a peak at certain times and a linear trend indicating day effect.](image-url)
Time of Day Effect
Prior Distributions for Parameters

\[ \lambda_0 \sim \Gamma(\lambda; a^L, b^L) \quad \Gamma(.) \text{ is the Gamma distribution} \]

\[ \frac{1}{7}[\delta_1, \delta_2, \ldots, \delta_7] \sim \text{Dir}(\alpha^d_1, \alpha^d_2, \ldots, \alpha^d_7, \ldots) \]

\[ \frac{1}{D}[\eta_{i,1}, \eta_{i,2}, \ldots, \eta_{i,D}] \sim \text{Dir}(\alpha^h_1, \alpha^h_2, \ldots, \alpha^d_D) \]

\text{Dir}(.) \text{ is a Dirichlet distribution}
Modeling Anomalous Data

- $N_A(t)$ is also a Poisson process with rate $\lambda_A(t)$.

- Markov process $A(t)$ is used to determine the existence of anomalous events at time $t$.

$$A(t) = \begin{cases} 
1 & \text{an event is occurring at time } t \\
0 & \text{otherwise}
\end{cases}$$
Continued

- Transition probabilities matrix

\[ M_A = \begin{pmatrix} 1 - A_0 & A_1 \\ A_0 & 1 - A_1 \end{pmatrix} \]

\[ A_0 \sim \beta(A, a_0^A, b_0^A) \]

\[ A_1 \sim \beta(A, a_1^A, b_1^A) \]

\[ N_A(t) \sim \begin{cases} 0 & A(t) = 0 \\ P(N; \lambda_A(t)) & A(t) = 1 \end{cases} \]
MMPP ～ HMM

- **Typical HMM**
  (Hidden Markov Model)

- **MMPP**
  (Markov Modulated Poisson Process)
Apply MCMC

- **Forward Recursion**
  - Calculate conditional distribution of $P( A(t) \mid N(t) )$

- **Backward Recursion**
  - Draw sample of $N_A(t)$ and $N_0(t)$

- **Draw Transition Matrix from Complete Data**
Anomaly Detection

- Posterior probability of $A(t)$ at each time $t$ is an indicator of anomalies

- Apply MCMC algorithm:
  - 50 iterations
Results

Posterior Distribution Averages

- Observed
- Modeled
Continued
Conclusions

- Cell phone data reflects human activities on hourly, daily scale

- MMPP provides a method of modeling call activity, and detecting anomalous events
Future Work

- Apply on longer time period, and investigate monthly and seasonal behavior
- Implement MMPP model as part of real time system on streaming data
- Incorporate into WIPER system