STA 571 Project Proposal: Spatio - Temporal Modelling of Crimes Rates Using Gaussian Processes.

Devin Ti Tze Hong[1]

[1]Duke University

1. Introduction

Modeling of crime rates across space and time is a challenge that is of use to policy makers and police departments. For this, statistical models are appropriate due to the lack of precise physical or economical models for predicting crime rates. Gaussian Processes provide a flexible statistical approach to modelling crime rates across space and time. Spatio-temporal modeling is motivated by the fact that crime rates are likely to be correlated across:

- Space (Eg. High crime areas are likely to be near high crime areas)
- Time (Eg. High crime at one point of time is likely to be correlated with high crime in a nearby time period, with influence degrading overtime)
- Crime Type

2. Related Work

A similar work is done in [1] .A gaussian process with a poisson process likelihood is used to predict weekly aggregated counts of thefts in Chicago in prespecified fixed neighborhood areas. Here, the author essentially treats the data as K different time series of theft counts, where K is the number of neighborhoods aggregated over, then a kernel which accounts for temporal and spatial covariance is used to model the data.

The data however is richer than that, with individual incidents being tagged with both a location and time, I thus seek to model these incidences of various crimes as a point process of locations and times. This leads to richer models such as the log gaussian cox process in [2]. Beside this another extension is multi-task learning of various different crime types such as assault, theft, gun crimes, which are likely to show strong positive correlation.

3. Model

The model proposed is a Log Gaussian Cox Process. Suppose we have incident data in the form $D = \{(l_1, t_1), (l_2, t_2)..., (l_N, t_N)\}$ where l_i denotes the location (in Lat Lon) of the incident and t_i denotes the time in which it occurs. Then the counts Y_A of incidences in a given space - time area A is given by:

$$Y_A|\lambda(x) \sim \text{Poisson}\left(\int_{x \in A} \lambda(x)dx\right)$$

where:

$$\lambda(x) = exp(f(x))$$
$$f(x) \sim \text{GP}(m(x), K(x, x^{'}))$$

Where x refers to the space - time coordinates (l_i, t_i). I plan to fit this model using either PyMC or GPflow.

4. Data

The dataset will be the open source Chicago crime dataset[1]. This contains raw incidences of various crimes both with time and location.

5. Evaluation

The model will be evaluated for its predictive accuracy on the incident of thefts in aggregated spatial areas and times. With a probabilistic model, it will also be interesting to evaluate the forecast uncertainties produced by the models.

References

[1] S. R. Flaxman, "A general approach to prediction and forecasting crime rates with gaussian processes," *Heinz College Second Paper. Pittsburg: Carnegie Mellon University*, 2014.

[2] P. J. Diggle, P. Moraga, B. Rowlingson, and B. M. Taylor, "Spatial and spatio-temporal log-gaussian cox processes: extending the geostatistical paradigm," *Statistical Science*, vol. 28, no. 4, pp. 542–563, 2013.

[1]Dataset publicly available at: https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2