

# Crime Hot-spots Forecasting – Phase I

## 1 Educational Goal

Apply data mining techniques on real-world crime call service data to estimate the hot-spots in the future.

## 2 Phase I

- **Goal:** Get baseline results of the crime hot-spots forecasting problem.
- **Due Date:** Thursday, October 20th, 2016.
- **Programming language:** python.
- **Dataset:** Calls-for-service (CFS) records provided by the Portland Police Bureau (PPB) from March 1, 2012 to September 30, 2016. This data contains four crime categories: all calls-for-service; burglary; street crime; and motor vehicle theft, by date and GPS coordinates.
- **Evaluation:** Use PAI and PEI\* to evaluate your results, the definitions of PAI and PEI\* are as follows:

**Prediction Accuracy Index (PAI):** The PAI will measure the effectiveness of the forecasts with the following equation:

$$PAI = \frac{\frac{n}{N}}{\frac{a}{A}} \quad (1)$$

where  $n$  equals the number of crimes that occur in the forecasted area,  $N$  equals the total number of crimes,  $a$  equals the forecasted area, and  $A$  equals the area of the entire study area.

**Prediction Efficiency Index\* (PEI\*):** The PEI\* will measure the efficiency of the forecasts with the following equation:

$$PEI^* = \frac{PAI}{PAI^*} \quad (2)$$

where PEI\* equals the maximum obtainable PAI value for the amount of area forecasted,  $a$ . As such:

$$PEI^* = \frac{n}{n^*} \quad (3)$$

where  $n^*$  equals the maximum obtainable  $n^*$  for the amount of area forecasted,  $a$ .

## 3 Approach

### 3.1 Step 1. Training and test data generation:

Using different parameters (different cell size, different cell shape, for example), create a grid of cells which covers all of the city of Portland. Transfer the CFS data (select a crime category) to the grid, such

that each cell contains the number of CFS records that occurred in that cell over a given time period. For example, you can use the burglary records in January 2016 to build a grid with  $600ft * 600ft$  square cells, where each cell has the total number of burglary calls that occurred in the cell in January. Code examples will be available for converting GPS coordinates to grid cells.

**Requirement:** Create monthly cell grids for one year's worth of a single crime category's data. 1. Individual cell area:  $62,500ft^2 - 360,000ft^2$ . 2. Minimum cell height:  $125ft$  3. Cell shape: square or rectangle.

### **3.2 Step 2. Calculate PAI and PEI of predicted hot-spot cells based on a single month of data:**

Choose the number of hot spots,  $n$ , and find the  $n$  cells with the highest concentrations of crimes and use these cells as a prediction of the next month's hot spots, using PAI and PEI as an evaluation method. For example, if you choose  $n = 20$ , and January 2016 data for training, you should pick the top-20 cells in your January grid as hotspots, and compare these 20 hotspots to the hotspots in the grid data for February 2016, and then determine the effectiveness(PAI) and efficiency(PEI) index values of these predictions.

**Requirement:** 1. Total hotspots area:  $0.25mi^2 - 0.75mi^2$ . 2. Draw a plot (or multiple plots) to show all the results with different parameters (cell size, cell shape, category, number of hotspots).

### **3.3 Step 3. Calculate PAI and PEI of multiple-month-based prediction:**

Choose a number of months  $m$ , then using the best parameters you tried in step 2, perform a prediction based on an  $m$ -month time period, then evaluate that prediction against the month following the last month of the training window using PAI and PEI.

**Requirement:** 1. Draw a plot to show the results with different values of  $m$ .

## **4 Submission Requirements**

1. Write an experiment report to discuss your experimental results, including detailed parameter settings, plots in step 2 and step 3, and the best parameter settings. Submit the paper copy of the report with the cover page in class. Paper copy should be bound firmly together as one pack (for example, staple, but not limited to, at the left corner). 5 points will be deducted for unbounded homework.
2. Submit the soft copy of the report through your UMassOnline account.
3. No hard copies or soft copies results in 0 points.