CIS 635 Knowledge Discovery & Data Mining

Crime Hotspots Forecasting

Lecturer: Dr. Yong Zhuang

Overview

This is National Institute of Justice's (NIJ) Real-Time Crime Forecasting Challenge, which hopes to provide researchers and the federal government with a better understanding of the potential for crime forecasting in Portland,Oregon.

Website: https://nij.ojp.gov/funding/real-time-crime-forecasting-challenge-posting

Calls-for-service (CFS) records

Temporal Information

Spatial information

CATEGORY	CALL GROUPS	final_case_type	CASE DESC	occ_date	x_coordinate	y_coordinate	census_tract
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	1/18/2013	7649793	662388	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	1/5/2013	7651202	661479	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	1/28/2013	7647818	663182	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	2/2/2013	7649298	661246	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	1/13/2013	7650935	661746	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	2/17/2013	7650248	660907	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	1/30/2013	7650289	662464	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	3/13/2013	7650182	664208	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	2/16/2013	7649859	665351	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	3/2/2013	7649894	664127	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	3/29/2013	7649298	661246	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	4/27/2013	7647366	665494	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	4/27/2013	7648668	662094	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	5/2/2013	7650785	661371	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	5/12/2013	7647366	665494	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	5/31/2013	7650022	663852	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	6/1/2013	7648386	663997	100
STREET CRIMES	DISORDER	DISTP	DISTURBANCE - PRIORITY	5/27/2013	7648851	663899	100

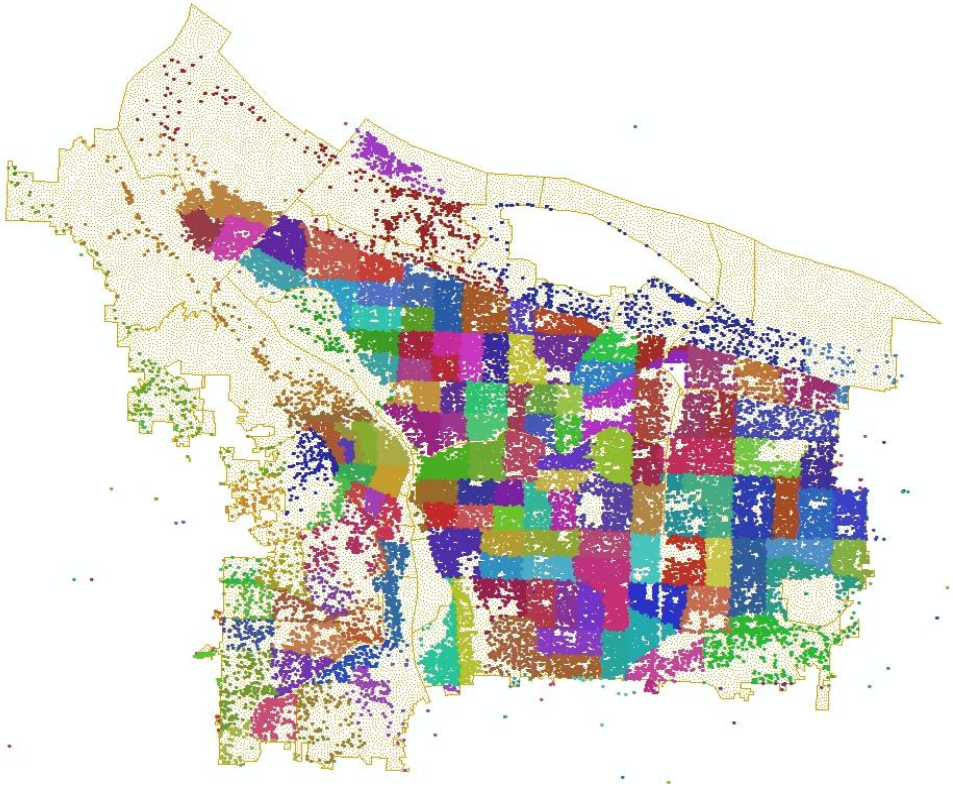
Categories of CFS

CFS Category	Code	Translation
Burglary	BURG	BURGLARY – COLD
	BURGP	BURGLARY – PRIORITY *H
	PROWLP	PROWLER
Theft of Auto	RSTLN	ROLLING STOLEN *H
	VEHREC	VEHICLE RECOVERED
	VEHST	VEHICLE STONE – COLD
	VEHSTP	VEHICLE STOLEN – PRIORITY
All CFS	This category includes all CFS including those in the above categories.	

Only work on All CFS

CFS Category	Code	Translation
Street Crime	ASSLT	ASSAULT – COLD
	ASSLTP	ASSAULT –PRIORITY *Note: This code initially was listed erroneously as "ASSLTT"*
	ASSLTW	ASSAULT WITH WEAPON *H
	DIST	DISTURBANCE – COLD
	DISTP	DISTURBANCE – PRIORITY
	DISTW	DISTURBANCE – WITH WEAPON *H
	GANG	GANG RELATED
	ROB	ROBBERY – COLD
	ROBP	ROBBERY – PRIORITY *H
	ROBW	ROBBERY – WITH WEAPON *H
	SHOOT	SHOOTING – COLD
	SHOOTW	SHOOTING – WITH WEAPON *H
	SHOTS	SHOTS FIRED
	STAB	STABBING COLD
	STABW	STABBING WITH WEAPON *H
	VICE	VICE-DRUGS, LIQUOR, PROSTITUTION, GAMBLING

Calls-for-service (CFS) records (2013)



Hot spot map

The smallest cell size is **250*250 sq.ft**

The Biggest cell size is **600*600 sq.ft**

Suppose the cell size is **250*250 sq.ft,** the range of number of hotspots is:
0.25 mi² / 62500 ft² = 112
0.75 mi² / 62500 ft² = 335

Table 2: Requirements for Entries

Requirement	Description of Requirement
Required files	.dbf .prj .sbn .sbx .shp .shx
Projection of files	NAD_1983_HARN_StatePlane_Oregon_North_FIPS_3601_Feet_Intl
Required variables	Unique ID for each cell A binary variable (1 – hot spot, 0 – not) Area for each cell measured in square feet to 4 decimal places
Cell shape	Any shape
Individual cell area*	62,500 ft^2 – 360,000 ft^2
Total forecasted area	0.25 mi^2 – .75 mi^2

*Cells forming the outer boundary of the study area (Portland Police Districts, see file in Section X.) must be trimmed so that the total area of all cells equals 147.71 square miles (+/-0.02 square miles). The area of each interior cell must be equal to one another.

Cell size

- Cell size **250*250 sq.ft**

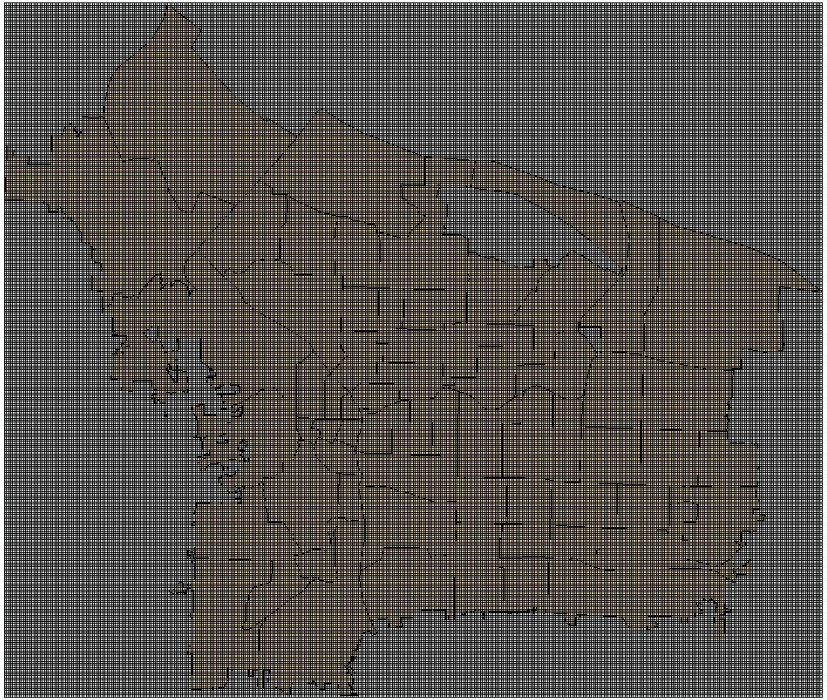
- Cell size **600*600 sq.ft**

Grid overlays Portland (fishnet)

733,940

Cell Size: 250*250 sq.ft
Grid Size: 331 * 390

7,603,950

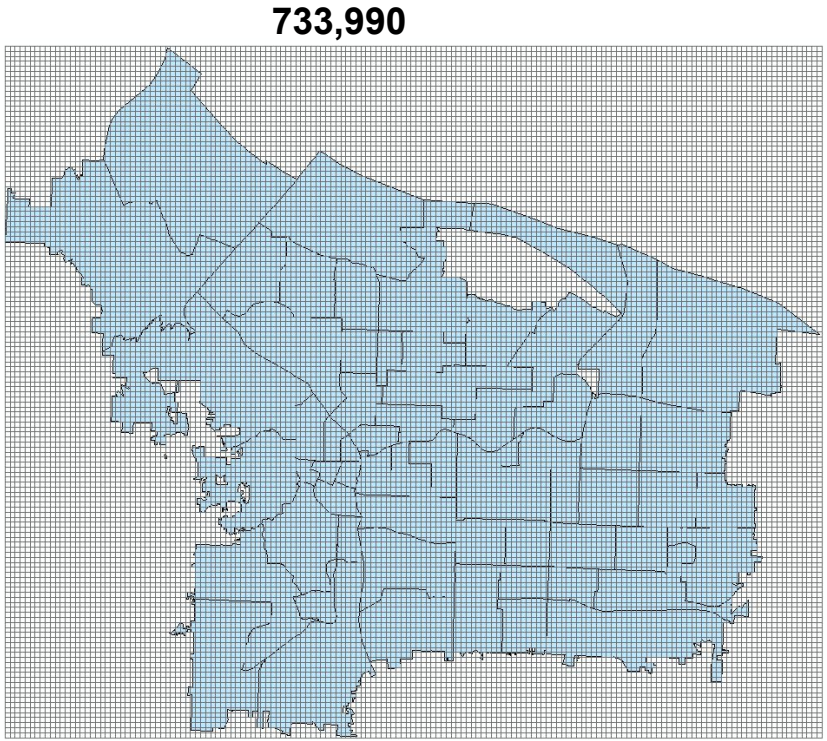


7,701,450

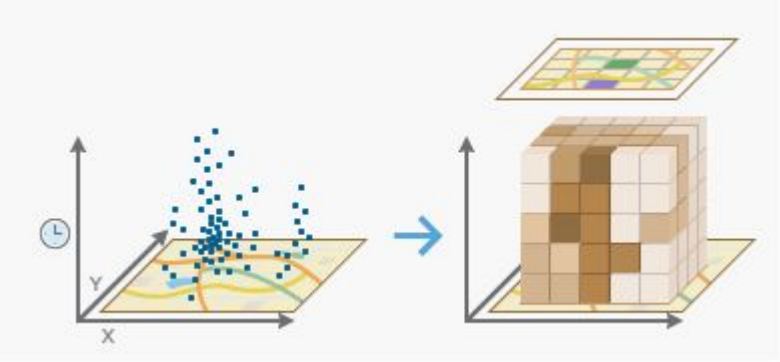
651,190

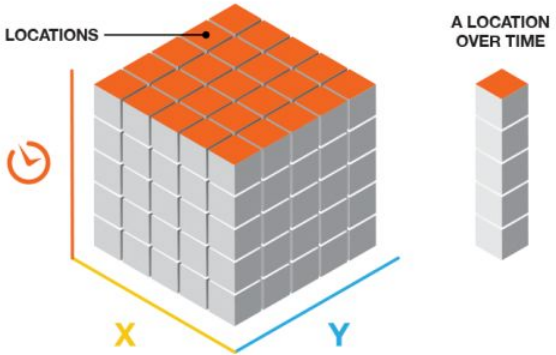
Grid overlays Portland (fishnet)

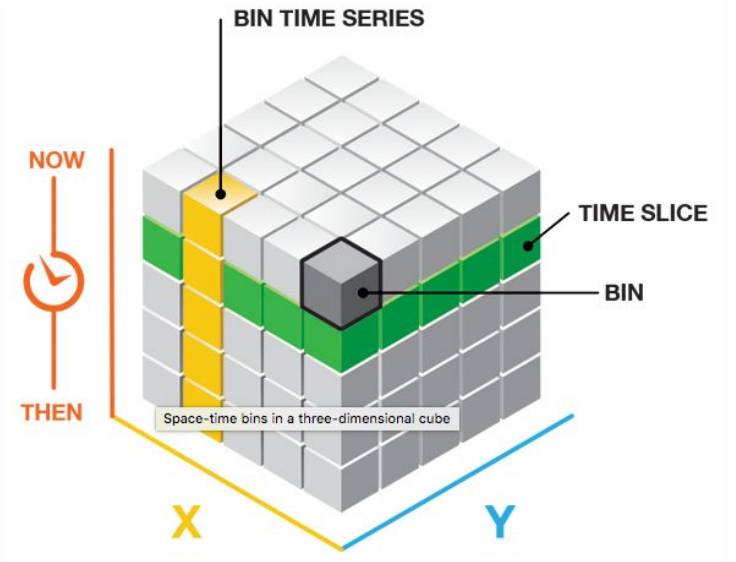
Cell Size: 600*600 sq.ft
Grid Size: 138 * 163



Spatio-Temporal Data







Judging Criteria - Prediction Accuracy Index (PAI)

$$\frac{\left(\frac{n}{N}\right)*100}{\left(\frac{a}{A}\right)*100} = \frac{HitRate}{AreaPercentage} = \text{Prediction Accuracy Index} \qquad (1)$$

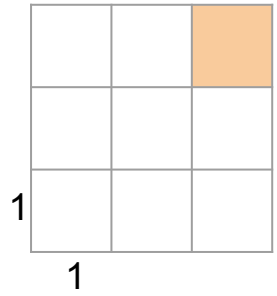
where n is the number of crimes in areas where crimes are predicted to occur (e.g. hotspots), N the number of crimes in study area, a the area (e.g. km^2) of areas where crimes are predicted to occur (e.g. area of hotspots), and A the area (e.g. km^2) of the study area.

Reference:

- **Chainey, S., Thompson, L., & Uhligh, S. (2008). The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. Security(21), 4-28.**

Evaluation (1 hot spot)

Hot spot map



1 week (3/1-3/7) evaluation

2	3	6
1	0	7
2	5	1

PAI = (6/27) / (1/9) =2

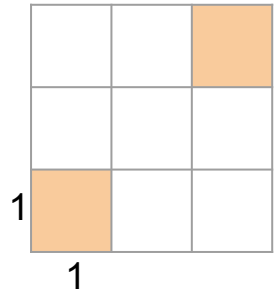
PEI* = 6 / 7

$$PAI = \frac{\left(\frac{n}{N}\right)*100}{\left(\frac{a}{A}\right)*100} = \frac{HitRate}{AreaPercentage} = \text{Prediction Accuracy Index}$$

$PEI^* = \frac{n}{n^*}$ where *n** equals the maximum obtainable *n* for the amount of area forecasted, *a*.

Evaluation (1 hot spot)

Hot spot map



1 week (3/1-3/7) evaluation

2	3	6
1	0	7
2	5	1

PAI = (8/27) / (2/9) =4/3

PEI* = 8 / 13

$$PAI = \frac{\left(\frac{n}{N}\right)*100}{\left(\frac{a}{A}\right)*100} = \frac{HitRate}{AreaPercentage} = \text{Prediction Accuracy Index}$$

$PEI^* = \frac{n}{n^*}$ where *n** equals the maximum obtainable *n* for the amount of area forecasted, *a*.

Which features are strongly relate to crime? burglary 1st week 2013.

Challenges in Crime Forecasting

- **Predictive Features**
  - What characteristics are most predictive of crime to optimize metrics like PAI or PEI*?
- **Memory Issues**: High-resolution grids demand significant memory during training.
  - How can this constraint be addressed effectively?
- **Data Challenges**
  - **Imbalance Issue**: Hotspot cells are scarce, leading to severe class imbalance.
  - **Missing Values**: How can we handle incomplete data to improve forecasting accuracy?
- **Additional Considerations**
  - Incorporating external datasets to enhance predictions.
  - Balancing efficiency and effectiveness in model performance.