

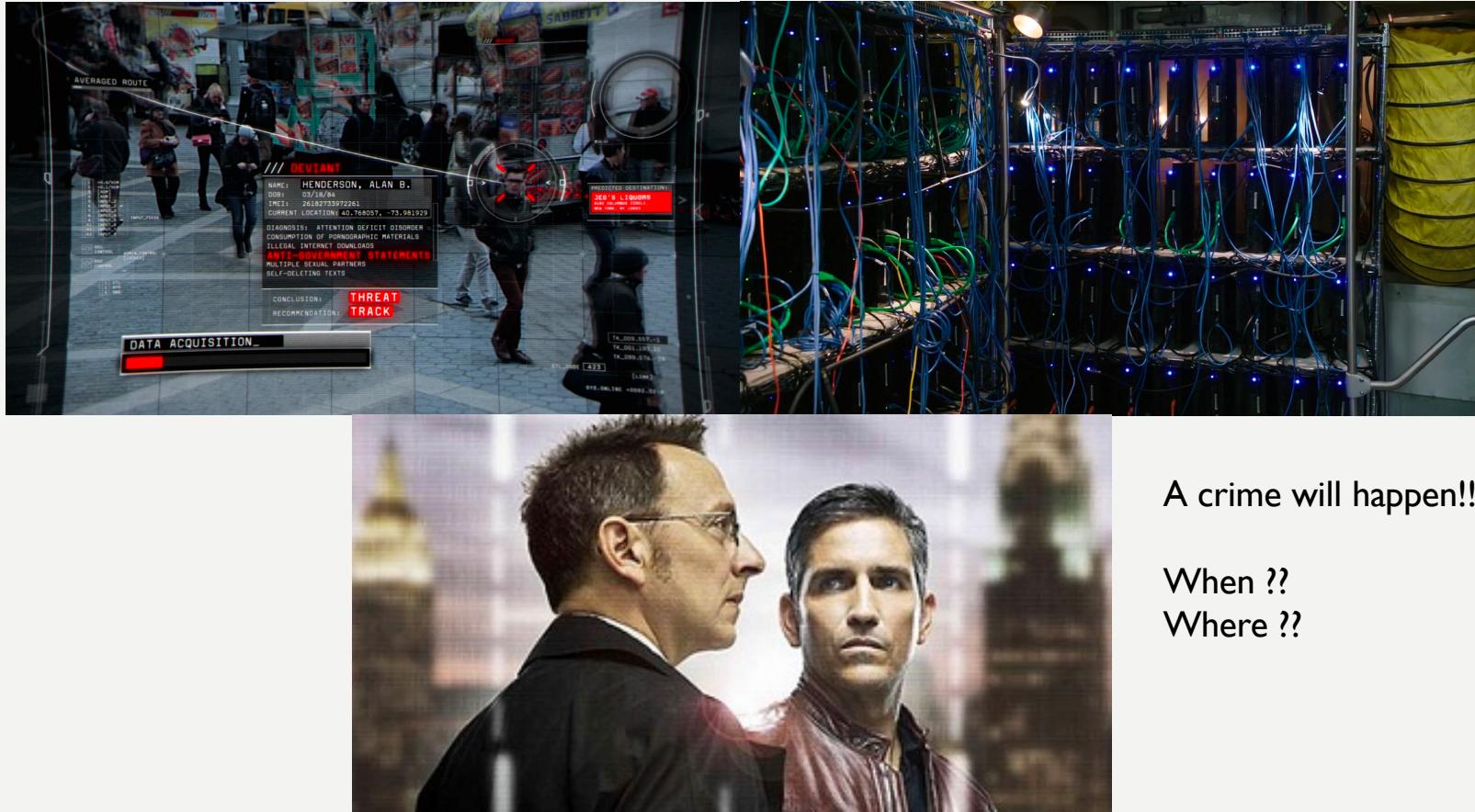
PREDICT CRIME

USING SPATIAL AND TEMPORAL APPROACH

A COMPARISON STUDY BETWEEN CHICAGO
AND SEATTLE

Yangyang Dai

RESEARCH BACKGROUND AND IDEAS



A crime will happen!!!

When ??
Where ??

AVAILABLE STUDIES AND RESOURCES

- Example studies:
 - Predict London crime rates, predict crime types in Denver, Los Angeles and etc.
 - Eg. A. Bogomolov, B. Lepri, J. Staiano, N. Oliver, F. Pianesi and A. Pentland, 'Once Upon a Crime: Towards Crime Prediction from Demographics and Mobile Data', CoRR, vol. 14092983, 2014.
- Methods overview:
 - SVM: using this method is still slow and computationally expensive
 - Fuzzy time series: works well on binary data, but not so much on more levels
 - Artificial neural network (ANN): relatively accurate, but takes a long time in training phases
 - Unsupervised method by using multivariate time series based on parametric Minkowski model and dynamic time wrapping (DTW), however, this method is difficult to handle a missing value in order to get more accurate results
 - Bayesian Network: it may lead to some deviation during conducted the experiment, so the more factors unrelated to the geography should be considered to improve the accuracy of the model.
 - decision tree: eg, the result of crime cases has been classified into two classes such as neutral and danger. However, this method does not work well on all type of datasets.
 - logistic regression: the limitation of this approach is difficult to identify the probability of burglary activity and specific locations.

AVAILABLE STUDIES AND RESOURCES

- **Contribution and focuses:**
- mid-west vs. west coast
- the newly developed vs. Early developed
- Crime city vs. civilized city

CRIME DATA

- Mid-West
 - With relatively long history, early-developed
 - infamous for the crime, gangs
- Chicago crime – 2001 to present
- extracted from the Chicago Police Department's **CLEAR (Citizen Law Enforcement Analysis and Reporting) system**
- 22 attributes (id, date, block, x, y coordinate etc.), 6.57M instances of crimes
- <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>

ID	Case Num	Date ↓	Block	IUCR	Primary Type	Description	Location
11267718	JB201047	03/26/2018 11:55:00 PM	066XX S WESTERN AVE	0860	THEFT	RETAIL THEFT	GAS ST.
11267724	JB201048	03/26/2018 11:45:00 PM	007XX S WELLS ST	1330	CRIMINAL TRESPASS	TO LAND	CONST
11267725	JB201049	03/26/2018 11:45:00 PM	117XX S THROOP ST	0910	MOTOR VEHICLE THEFT	AUTOMOBILE	STREET
11267755	JB201053	03/26/2018 11:40:00 PM	064XX W IRVING PARK RD	1330	CRIMINAL TRESPASS	TO LAND	GROCE
11267705	JB201038	03/26/2018 11:35:00 PM	068XX S ASHLAND AVE	502P	OTHER OFFENSE	FALSE/STOLEN/ALTERED TRP	STREET
11267737	JB201040	03/26/2018 11:30:00 PM	031XX S KEELER AVE	0326	ROBBERY	AGGRAVATED VEHICULAR HIJACKING	STREET
11267784	JB201062	03/26/2018 11:30:00 PM	040XX W MAYPOLE AVE	031A	ROBBERY	ARMED: HANDGUN	STREET
11267753	JB201044	03/26/2018 11:19:00 PM	039XX W 63RD ST	1812	NARCOTICS	POSS: CANNABIS MORE THAN 30GMS	STREET
11267813	JB201046	03/26/2018 11:15:00 PM	003XX W 109TH PL	0486	BATTERY	DOMESTIC BATTERY SIMPLE	RESIDE
11268448	JB202026	03/26/2018 11:15:00 PM	007XX W 47TH ST	1320	CRIMINAL DAMAGE	TO VEHICLE	STREET
11267714	JB201023	03/26/2018 11:10:00 PM	071XX S ASHLAND AVE	0320	ROBBERY	STRONGARM - NO WEAPON	STREET
11267792	JB201032	03/26/2018 11:08:00 PM	077XX S CORNELL AVE	1310	CRIMINAL DAMAGE	TO PROPERTY	RESIDE
11267701	JB201028	03/26/2018 11:05:00 PM	033XX W WARNER AVE	1320	CRIMINAL DAMAGE	TO VEHICLE	STREET
11268384	JB201839	03/26/2018 11:05:00 PM	056XX S WABASH AVE	0620	BURGLARY	UNLAWFUL ENTRY	APARTN

CRIME DATA

West coast

- Newly developed, tech-oriented
- Seattle crime – 2010 to 2017
- all the Police responses to 9-1-1 calls within the city
- 19 attributes, 1.47M instances
- https://data.seattle.gov/Public-Safety/Seattle-Police-Department-911-Incident-Response/3k2p-39jp?category=Public-Safety&view_name=Seattle-Police-Department-911-Incident-Response

Hundred Block Location	:	District/Sect	:	Zone/Blo	:	Census Tr	:	Longitude	:	Latitude	:	Incident Location
3 AV S / S WASHINGTON ST		K		K3		9200.2014		-122.330271593		47.600875809		(47.600875809°, -122.330271593°)
20XX BLOCK OF 15 AV W		Q		Q1		5802.2003		-122.37613941		47.636336049		(47.636336049°, -122.37613941°)
6 AV / YESLER WY		K		K3		9200.1002		-122.326350868		47.601708802		(47.601708802°, -122.326350868°)
86XX BLOCK OF 24 AV SW		F		F2		11401.2005		-122.363172642		47.525585666		(47.525585666°, -122.363172642°)
135XX BLOCK OF 23 AV NE		L		L1		200.6017		-122.304248161		47.727498035		(47.727498035°, -122.304248161°)
63XX BLOCK OF 29 AV SW		F		F1		10700.4001		-122.369833395		47.546493546		(47.546493546°, -122.369833395°)
MARTIN LUTHER KING JR WY S / S GENESEE ST		R		R2		10001.3006		-122.295370641		47.563805602		(47.563805602°, -122.295370641°)
CALIFORNIA AV SW / SW ALASKA ST		W		W2		10500.4003		-122.386778535		47.561104368		(47.561104368°, -122.386778535°)
24XX BLOCK OF AURORA AV N		Q		Q2		6000.2045		-122.346642846		47.641419741		(47.641419741°, -122.346642846°)
43XX BLOCK OF S FERDINAND ST		R		R3		10300.1002		-122.278843236		47.558197565		(47.558197565°, -122.278843236°)
68XX BLOCK OF 30 AV NE		U		U3		3800.1004		-122.29516869		47.678505415		(47.678505415°, -122.29516869°)
1XX BLOCK OF NW 81 ST		J		J2		2900.1014		-122.359311741		47.687664142		(47.687664142°, -122.359311741°)
26XX BLOCK OF S DEARBORN ST		G		G3		8900.4011		-122.298174668		47.595534943		(47.595534943°, -122.298174668°)
30XX BLOCK OF E UNION ST		C		C3		8800.1000		-122.293157653		47.612937729		(47.612937729°, -122.293157653°)

METHODS – FRAMEWORK

- Ideas:
- Build models using Chicago and Seattle crime data, train and test the two models, and predict the crime in each two cities
- Compare the temporal and spatial crime distributions
- Understand the differences and similarities between two cities
 - Crime distribution over the 12 months, Weekly distribution, daily distribution
 - Location or neighborhood differences (apart, street, sidewalk, school...)
 - Crime type differences

METHODS – DATA PREPARATION

- Data cleaning: missing values
- Data reduction: eg. possible dimension reduction
- Data integration: eg. unify attribute naming, adopt military time system
- Data Transformation and Discretization:
 - create crime_type, crime_time variables, group into smaller subsets if too many, into hour intervals

METHODS – MODELS AND ANALYSIS

- **Apriori Algorithm**
 - to find all possible crime frequent patterns regardless of the committed crime type
 - come up with a list of all crime hotspots along with its related frequent time
- **Naïve Bayesian Classifier**
 - assumes the independent effect between attribute value
 - sklearn
 - Multinomial Naïve Bayes, crime type prediction
- **Decision Tree Classifier**
 - sklearn
 - entropy function for the information gain to measure the quality of the split

Evaluation:

- prediction accuracy
- Running time

POTENTIAL RESULTS

- Spatial and temporal hot spots in Chicago and Seattle
- Comparison of crime patterns
- Predicted Crime type in a given location within a give period of time
- Questions?