TIME AND LOCATION RECOMMENDATION FOR CRIME PREVENTION

Yihong Zhang
Panote Siriaraya
Yukiko Kawai
Adam Jatowt

Overview

- Background
- Modeling crime inference as a recommendation problem
- Solutions to crime inference as a recommendation problem
- Experimental Evaluation
- Conclusion

Why Crime Prediction

- Police departments would like to send targeted patrol to places where crime is likely to happen
- Residents and tourists would like to avoid dangerous places
- Policy makes would want to investigate the cause of high crime rates in certain area

Availability of Crime Data

- In recent years, more and more cities put its crime data online
  - New York City
  - Los Angeles
  - Chicago
  - Philadelphia
- Crime data contains rich information
  - Exact GPS coordinates of crime
  - Time of crime specified to minutes
  - Type of crime

Crime Prediction: Current State-of-Art

- Crime rate prediction: the number of crimes that will occur in the area
  - Regression
  - Large geographical unit: zip-code area, grids of 2km x 2km cell size
- Next moment prediction: the likelihood a crime will occur in the next time unit in an area
  - Binary classification
  - Smaller geographical unit
  - Techniques usually need to down sample negative instances
- Granularity vs Sparsity

Recommendation Problem

		3	5	
		4	3	2
	1	5		
		4	4	3

Recommendation Problem

	?	3	5	?
	?	4	3	2
	1	5	?	?
	?	4	4	3

Similarity Between User-Item Data and Spatio-Temporal Crime Data

- Spatial consistency of crime
  - Certain crime is said to be location-dependent (see near repeat theory [1]), and the crime number depends on the potential criminal living in the area
- Temporal consistency of crime
  - To some degree crime follows daily routine patterns of residents around the area
  - For example, theft happens more in the night during the dinner hour than in the day, because people (potential victims) are more likely to be outside

Time and Location Recommendation for Crime Prevention

		3	5	
		4	3	2
	1	5		
		4	4	3

Why Location as User and Time as Item?

- Reason 1: the number of location is much larger than the number of time units
  - in a typical recommendation problem, the number of users is far more than the number of items
- Reason 2: crimes are mostly caused by the criminals living in the neighborhood
  - it is more appropriate to represent the human factor as the location

Sparsity

- Typical sparsity in a product review dataset: 90%
- Our spatio-temporal unit for crime data
  - 200m x 200m blocks
  - 24 x 7 hours in a week
- Sparsity (San Francisco crimes):

Table 1. Crime number and sparsity for crimes in SF.

	0	1	2	3	4	≥ 5	sparsity
theft	346,970	18,565	13,264	5,990	3,971	7,888	0.87
assault	30,6278	76,56	5,446	2,055	1,421	1,720	0.94

Solving Recommendation Problem

- Collaborative filtering
- Context-based methods
  - Tensor Decomposition Analysis
  - Latent Topic Analysis

Collaborative Filtering

- Item-based / user-based
- Similarity matrix

$$sim(i,j) = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{||\vec{i}||_2 * ||\vec{j}||_2}$$

- Predicting scores based on similarity

$$P_{u,i} = \frac{\sum_{j=1}^{N} sim(i,j) * R_{u,j}}{\sum_{j=1}^{N} sim(i,j)}$$

Adding Context Information

- In product recommendation, it is common to consider user review comments
- Features extracted from review text can be added as another dimension of information in recommendation

Generating Context

- Row data: tweets with coordinate tags
- Align tweets to crimes in spatio-temporal units
- Feature extraction from tweets
  - Bag-of-words
  - Convert tweets to vectors using pre-trained word embeddings (GloVe)

Context-based Methods

- Tensor decomposition
  - Using decomposed low rank matrix to approximate data
  - Learn by gradient descent
  - Missing values can be computed by multiplication of low rank matrix
- Latent Topic Analysis
  - Models contextual data as latent topics
  - Parameters controls the influence of user, item, and context
  - Can make prediction with a small number of rating records

Experimental Evaluation - Non-recommendation Baselines

- Historical sum
  - consider the same amount of crimes are likely to happen in the same location and the same hour in the future as in the past
- Auto-Regressive Integrated Moving Average (ARIMA)
  - a common method used in time series forecasting
- Vector Auto-Regression (VAR)
  - a popular forecast method that combines multiple signals together in an AR model
- Kernel Density Estimation (KDE)
  - a popular interpolation method to estimate crime in areas where there is lack of previous crime records
  - A separate KDE model for each hour

Experimental Evaluation - Dataset

- Theft and assault data from San Francisco
  - Studied period: 100 weeks starting from 2016
  - theft: 151k
  - assault: 42k
- Contextual data
  - Tweets collected in 2016 and 2017
  - Total number: 371k tweets
  - To avoid bots, we remove tweets from the top 1% most frequently posting users

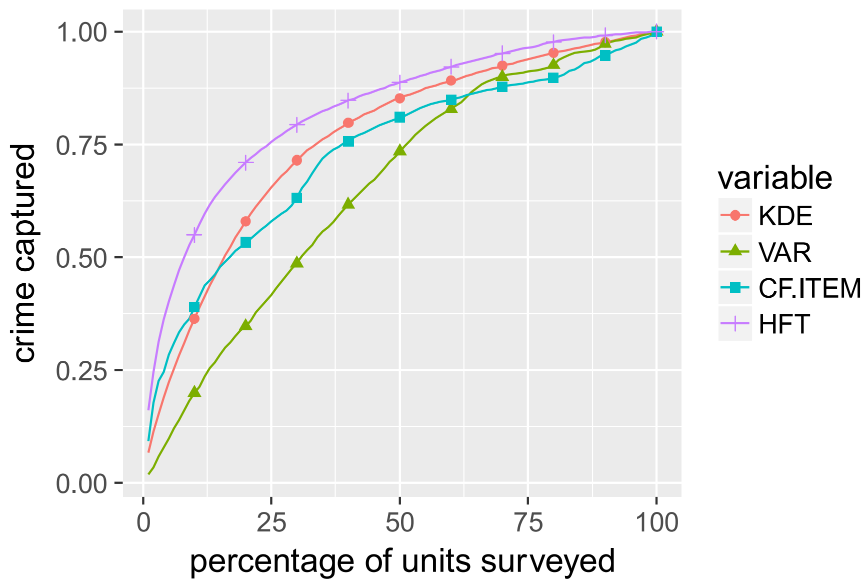
Rating Approximation Evaluation

- The accuracy of the recommendation model in the training data, measure by MAE
- Lower MAE means better approximation (e.g., by low rank matrix multiplication)

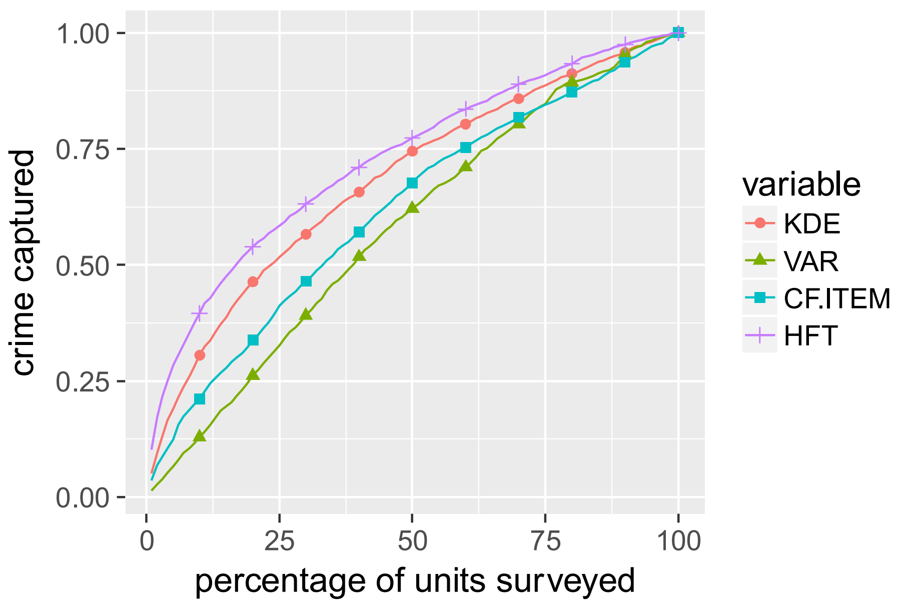
Table 2. MAE of recommendation and non-recommendation methods

	ARIMA	VAR	TD	HFT
theft	2.306	1.946	2.299	**1.408**
assault	2.089	1.978	2.085	**1.645**

Predicting Future Crime Number



theft



assault

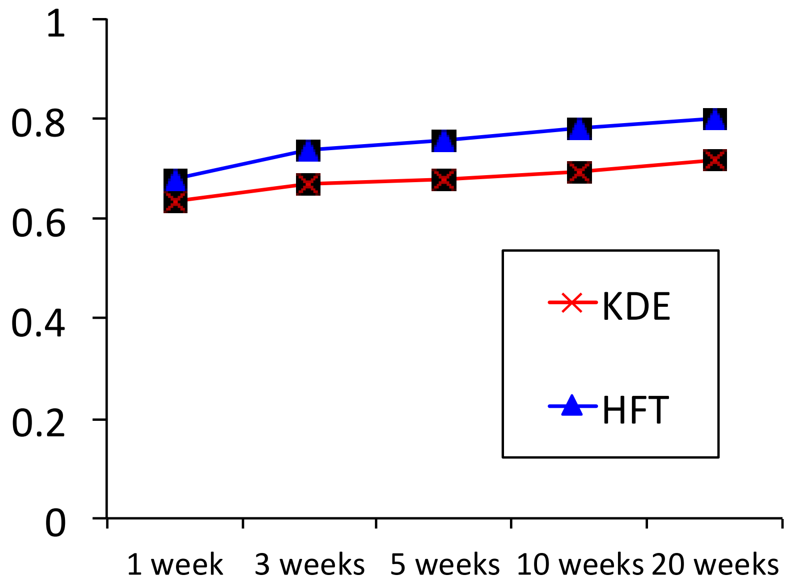
Predicting Future Crime Number

- 80 weeks training, 20 weeks testing

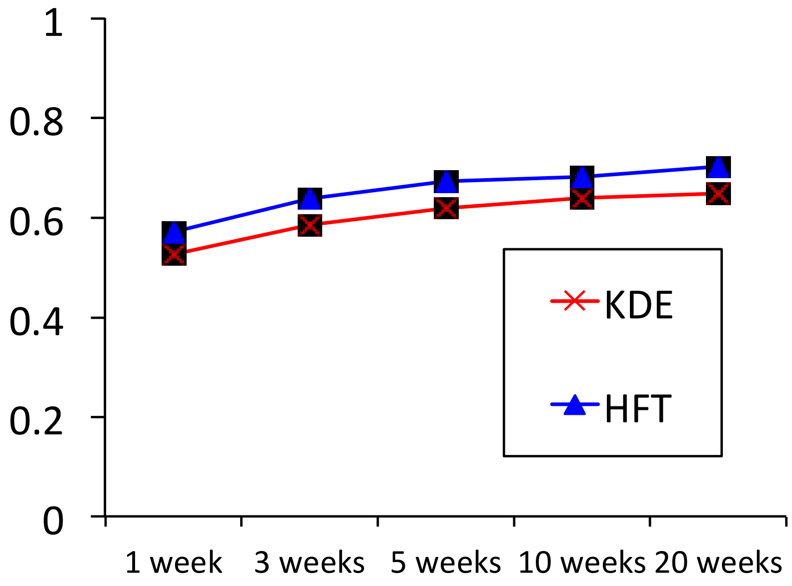
Table 3. Future crime prediction accuracy with recommendation and non-recommendation methods

	theft			assault		
	AUC	surv 20 %	surv 50%	AUC	surv 20%	surv 50%
historical sum	0.705	0.202	0.735	0.588	0.352	0.598
KDE	0.766	0.579	0.856	0.681	0.463	0.744
ARIMA	0.633	0.263	0.734	0.563	0.241	0.621
VAR	0.653	0.347	0.734	0.578	0.261	0.621
CF Item	0.728	0.533	0.810	0.617	0.338	0.677
CF User	0.657	0.463	0.682	0.536	0.239	0.528
TD	0.812	0.698	0.885	0.703	0.498	0.767
HFT	**0.824**	**0.709**	**0.886**	**0.725**	**0.539**	**0.773**

Effect of Training Data Size



theft



assault

Conclusion

- We model crime prediction as a recommendation problem
- Solutions in recommendation systems can solve the sparsity issue when assigning crimes into fine-grain spatio-temporal units
- We studied theft and assault in San Francisco. In the future, we plan to test more crime types and more cities

Thanks!

- And Questions?

Yihong Zhang
yhzhang7@gmail.com