The Impact of Community Safety on House Ranking

Zijun Yao, Yanjie Fu, Bin Liu, Hui Xiong

Department of Management Science and Information System, Rutgers University

{zijun.yao, yanjie.fu, binben.liu, hxiong}@rutgers.edu

Community Safety of Houses

- Community safety describes the degree that people live without fear of crime, such as burglary, robbery, or assault.
- Community safety is a fundamental buying factor of houses.
- Serious safety issues impact house values by (1) endangering residents, (2) degrading neighborhood, and (3) obstructing area development.

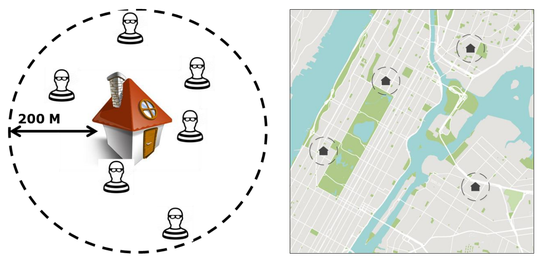
House Ranking

- **Designing ranking systems for residential houses (single family type), based on house values, for house investment decision support.**

How to Know Local Safety Condition?

```
OFFENSE_ID:        2010200143220200
OFFENSE_TYPE_ID:   burglary-residence-by-force
OCCURRENCE_DATE:   2010-04-28 08:29:59
INCIDENT_ADDRESS:  2909 N IVANHOE ST, DENVER
LATITUDE:          39.7586462
LONGITUDE:         -104.9214312
```

A crime record.



Crime collection for houses.

- **Historical crime records:** spatio-temporal crime incident data.
- **Crime influences:** the collected crime information for a house should happened nearby.

How to Analyze Crimes for House Investment?

- **Residential forcible burglary (break-in):** a widely spread crime type which directly impacts housing experiences.
- **Community crime evidence:** the extracted crime information which infers the underlying community safety condition.
- **Safety-aware house ranking:** a ranker which combines comprehensive house safety features to predict house values.

Problem Definition

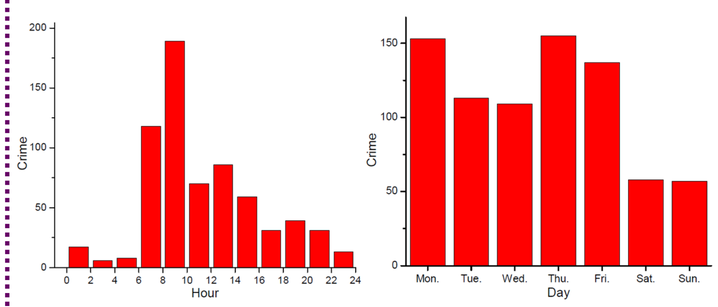
- **Given**
  Houses with corresponding location and square footage values.
  Historical burglary crimes with location, address and time.
  House profiles (characteristics of neighborhood, surrounding, build).
- **Objective**
  Ranking houses according to house value by the degree of safety concerns.
- **Three challenges**
  1. What crime evidences to mine for understanding comprehensive house safety situation?
  2. How does the severity of community safety impact house values?
  3. How to consider impact from factors other than safety in house ranking?

Community Crime Evidence Mining (1)

- **Crime severity category:** inferring the severity of a crime by the losses which are not disclosed in crime records.
- **Occurrence address evidence (E1):** use the appraisal of houses tracked by victimized address to assess the possible losses.

$$E_1(c) = Appraisal(add_c)$$

- **Occurrence Time Evidence (E2):** use the time slot during which the crime happened to assess the burglary losses.



Crime statistics by hours, and days.

Strong correlation between time of burglary and schedule of people.

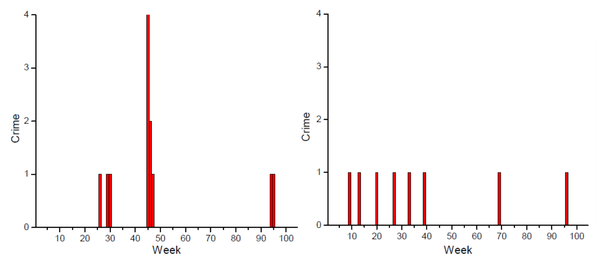
$$Entropy(k) = -\sum_{i,P_{k,i}\neq 0} P_{k,i} \cdot logP_{k,i}$$

$$P_{k,i} = |C_{k,i}|/|C_k|$$

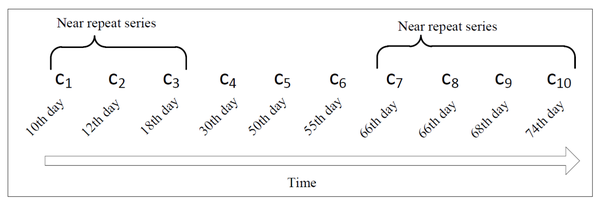
$$E_2(c) = Entropy(ts_c)$$

Community Crime Evidence Mining (2)

- **Temporal correlation category:** aiming to consider the temporal proximity among adjacent burglaries for revealing safety condition.



(a) Low value house (b) Normal value house



An example of near repeat series mining.

- **Near repeat series:** consists of *n* adjacent crimes around a house and meets two conditions, (1) a minimum size threshold (e.g., 3 crime), and (2) a maximum interval threshold (e.g., 7 days).
- **Series Size Evidence (E3):** $E_3(s) = |s|$ where $s = \{c_1, \dots, c_n\}$ is a series.
- **Series Length Evidence (E4):** $E_4(s) = t_{c_n} - t_{c_1} + 1$ where *t* is timestamp.
- **Series Intensity Evidence (E5):** $E_5(s) = \max_{2\leq n\leq N}\{\tau - (t_{c_n} - t_{c_{n-1}}) + 1\}$.

A House Safety-Aware (HSA) Ranking Model

- **Optimizing both accuracy of house value prediction and consistency of house ranking prediction**

$$Pr(\mathcal{O}|W,\Phi) = Pr(Y|W,\Phi)Pr(R|W,\Phi)$$

where $\mathcal{O} = \{Y, R\}$ is the observed values and rankings, $\Phi = \{\sigma^2, b^2\}$ is the hyperparameters.

- **Prediction Accuracy** $Pr(Y|W,\Phi) = \prod_{i=1}^{I} \mathcal{N}(y_i|f_i,\sigma^2)$ **where** $f_i = W^T X_i$
- **Ranking Consistency (pairwise: house is in order with value)**

$$Pr(R|W,\Phi) = \prod_{i=1}^{I-1}\prod_{j=i+1}^{I} Pr(h_i \to h_j|W,\Phi)$$

Incomparable houses: treating house pairs with different importances in pair-wise ranking optimization by using Euclidean based similarity [0,1] on house profile

$$D(h_i,h_j) = -\sqrt{\sum_{l=1}^{L}|p_l^i - p_l^j|^2}$$

$$Pr(R|W,\Phi) = \prod_{i=1}^{I-1}\prod_{j=i+1}^{I}\left(\frac{1}{1+exp(-(f_j - f_i))}\right)^{D_{ij}}$$

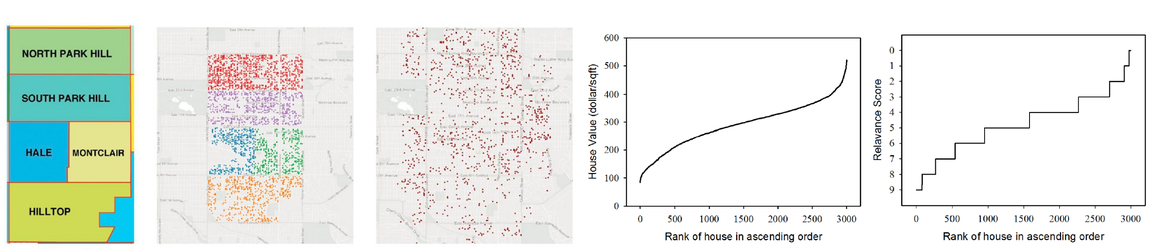
- **Log Posterior**

$$\mathcal{L}(W|Y,R,\sigma^2,b^2) = -\frac{1}{2\sigma^2}\sum_{i=1}^{I}(y_i - f_i)^2 + \sum_{i=1}^{I-1}\sum_{j=i+1}^{I} D_{ij} ln\frac{1}{1+exp(-(f_j - f_i))} - \frac{1}{2b^2}\sum_{m=1}^{M} w_m^2$$

Characteristics	Description
Neighborhood Characteristics	
Household Income	Average annual household income
High Educated Ratio	Ratio of residents with least bachelors degree to residents who are at least 25 years old
Population Growth	Percentage growth of population from 2000 to 2010
Surrounding Characteristics	
School Rating	Average rating of the nearest public high, middle and primary schools (A school has rating 1 to 5)
Point-of-Interest (POI) Diversity	Number of diverse categorical tags extracted from all the POIs which locate within *d* meters of the house
Check-in Density	Average number of social network check-in within *d* meters in every workday after hours 6 PM to 6 AM
Build Characteristics	
Land Area (in sqft), Bedroom Number and Build Year	

Experimental Results

- **The Experimental Data:** We use 3000 houses (single family type) with appraised values in 2015. 1131 forcible burglary crimes 2009 – 2014.

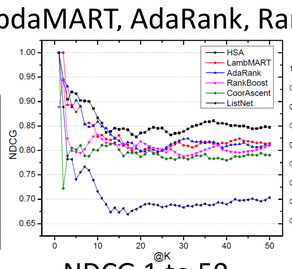


Neighborhood, houses, and crimes. House value grading.

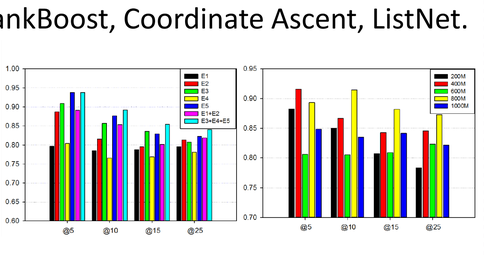
- **Evaluation Metrics:** NDCG@N, P@*N*, R@*N*, Kendall's Tau Coefficient.
- **Baseline Algorithms:** LambdaMART, AdaRank, RankBoost, Coordinate Ascent, ListNet.
- **Performances:**

Metrics	Lamb MART	Ada Rank	Rank Boost	Coor Ascnt	List Net	HSA
NDCG@5	0.8788	0.9023	0.7967	0.8015	0.7408	**0.9160**
NDCG@7	0.8537	0.8532	0.8026	0.7890	0.7664	**0.9013**
NDCG@10	0.8280	0.8590	0.8320	0.8003	0.7159	**0.8666**
NDCG@15	0.8309	0.8197	0.8113	0.8015	0.6733	**0.8429**
NDCG@25	0.8007	0.7993	0.8025	0.7859	0.6839	**0.8457**
Tau	0.2137	0.2733	0.1611	0.2326	0.2471	**0.3146**

NDCG and Tau coefficient.



NDCG 1 to 50.



Feature and radius study by NDCG

Conclusions

- Extracting comprehensive community crime evidences (learned from historical crime data) to reveal underlying house safety conditions.
- Incorporating community safety to enhance house ranking.

RUTGERS
THE STATE UNIVERSITY OF NEW JERSEY