Homework 2: Crime Event Detection Considering Spatial-temporal Factors

Hsun-Ping Hsieh

EE, CCE

Schedule

- 12/5: Homework 2 release
- next:
  - 12/26: Homework 2 deadline
  - 12/12, 12/19, 12/26 paper presentation
  - 1/9: Project Presentation
  - 1/2: no course

Crime Event Detection Considering Spatial-temporal Factors

- Goal: given a set of candidate locations in the city and timeslots, our aim is to identify the ones with high crime rate.

- Motivation: patrol path planning, crime alert

Task

- Given:
    - Historical crime events
        - Positions (lat, lng)
        - time information(date and time)
        - crime type
    - Some candidate locations
        - positions(lat, lng)
        - time information (with date and time slot)

- You need to
    - Detect/predict the ones will occur at least one crime event

Historical crime data: Crimes2016.csv (comma division)

- 1. date and time
- 2. crime type
- 3. latitude
- 4. longitude

POI information: locCategory.csv (comma division)

- 1. latitude
- 2. longitude
- 3. POI type

Weather data: Weather.csv (comma division)

- 1. datetime
- 2. humidity
  - max:100
- 3. pressure
- 4. Temperature
- 5. weather_description
  - category-based
- 6. wind_direction
  - <=360
- 7. wind_speed

Testing instances: questionNode_2017.csv (comma division)

- 1. latitude
- 2. longitude
- 3. date
- 4. time slot
  - midnight(0-5:59)
  - morning(6-11:59)
  - afternoon(12-17:59)
  - night(18-23:59)

```
1  Latitude,Longitude,Date,Time slot
·  41.82516963,-87.6192093,2017-06-06,night
·  41.79038534,-87.61911513,2017-07-16,afternoon
·  41.90557165,-87.740311,2017-01-10,midnight
-  41.8596613,-87.70547774,2017-03-10,afternoon
·  41.72824964,-87.53694824,2017-09-12,midnight
·  41.93505987,-87.67472305,2017-05-09,night
·  41.90700301,-87.66665767,2017-11-20,afternoon
·  41.76039682,-87.72717942,2017-04-09,morning
10 41.69540558,-87.58393983,2017-01-06,midnight
·  41.97739898,-87.67210979,2017-08-28,afternoon
·  41.72854561,-87.6231489,2017-12-09,afternoon
·  41.98646096,-87.6552839,2017-11-19,afternoon
·  41.66246257,-87.6048654,2017-10-16,morning
-  41.66592272,-87.57834943,2017-04-23,morning
16 41.8388o011,-87.71982484,2017-10-24,night
·  41.95470512,-87.7166153,2017-01-01,night
·  41.76559301,-87.68816325,2017-07-10,night
·  41.930906,-87.64401704,2017-01-28,afternoon
20 41.96293165,-87.71496963,2017-07-23,afternoon
·  41.66924495,-87.61156171,2017-10-11,afternoon
·  41.78429824,-87.66797436,2017-09-02,night
·  41.65052621,-87.59457404,2017-11-22,midnight
·  41.81954712,-87.6748287,2017-01-19,morning
-  41.76109969,-87.73650481,2017-06-19,midnight
·  41.78439557,-87.64490296,2017-07-23,midnight
·  41.73663961,-87.56241854,2017-06-05,midnight
·  41.85237397,-87.62549165,2017-05-16,midnight
·  41.69455035,-87.56520604,2017-04-12,morning
30 41.70083314,-87.57046628,2017-10-28,midnight
·  41.74271022,-87.63408818,2017-07-27,afternoon
·  41.78840613,-87.63283645,2017-04-13,morning
·  41.76938082,-87.60184892,2017-08-08,afternoon
·  41.94234256,-87.65904126,2017-11-07,morning
-  41.81158555,-87.67959973,2017-10-19,night
   41.77954061,-87.65966427,2017-06-04,midnight
   41.82213036,-87.60327658,2017-05-14,afternoon
·  41.70580401,-87.54563464,2017-05-05,night
·  41.85522939,-87.65435931,2017-06-23,night
·  41.93066074,-87.70807475,2017-09-02,night
   41.89329878,-87.62636218,2017-01-18,afternoon
```

Effectiveness

- F1 score = 2 * P * R / (P + R)

- Precision= Σ True positive/Σ Test outcome positive
  - The number of guessed positive

- Recall= Σ True positive/Σ Condition positive
  - Total numbers of positive in data

Report format(1)

- No formal format, you just can find a way to introduce your (but not limited to)
    - methodologies & thinking
    - or experimental process/framework
    - or experimental results

Report format(2)

	Physical meaning
Variable 1	
Variable 2	
Variable 3	
Variable 4	
Variable 5	

Submitted file formation(result.csv)

- 1. latitude
- 2. longitude
- 3. date
- 4. time slot
  - midnight(0-5:59)
  - morning(6-11:59)
  - afternoon(12-17:59)
  - night(18-23:59)
- 5. your prediction
  - 1: crime occur
  - 0: no crime event
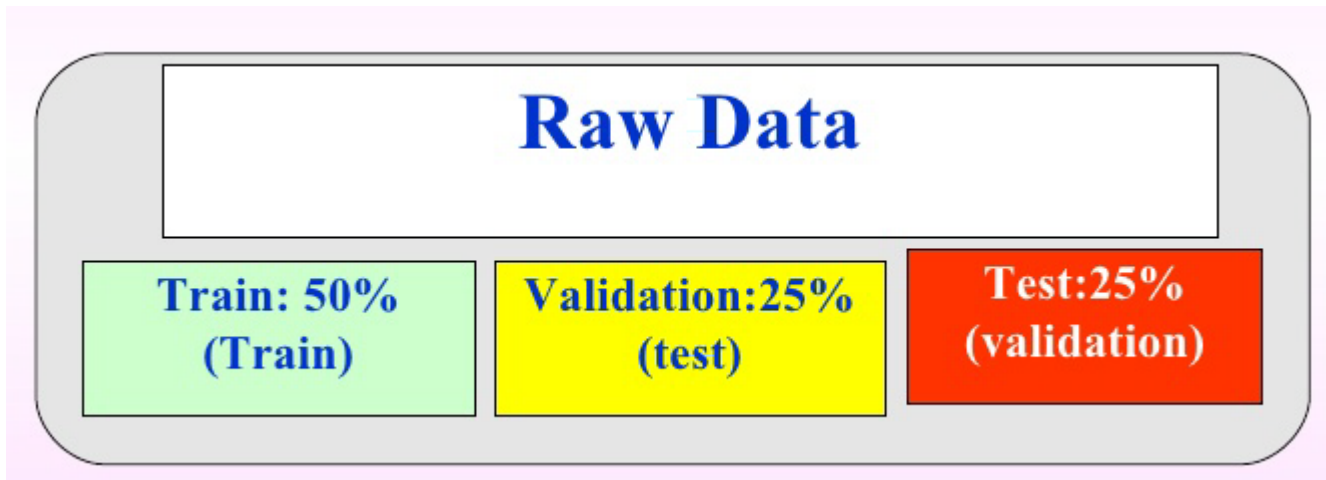
Hint1: feature extraction

- For each candidate location, predicting the yes or no by extracting their neighborhood dynamic & static variables
    - For example, number of night spots->high crime rate?
    - For example, number of all locations->low crime rate?
    - For example, the location of police stations?-> low crime rate?
    - For example, rainy->low crime rate?
    - For example, morning->low crime rate?
    - For example, midnight+rainny->low crime rate?

- Dynamic data: weather & time
- Static data: POI-based distribution(e.g. density, entropy)
- Neighborhood definition: radius = r meters(you need the program of calculating distance using the pair of lat and lng)

Hint2: generate training instance using 2016 crime data

- Select the locations(with lat and lng) in testing data and extract the availability of crime on different time slots

- Finding balanced positive and negative instances

Hint3: validation

- If you want to tune parameters and test the performance…(e.g. radius or model's parameters)
  - You need to divide validation set from the training set

Hint4: Classification

- Machine learning-based
    - SVM(笨蛋也可以用的libsvm)
    - Neural network
    - naive bayes
    - Random forest
- KNN-based
    - by feature distance(similarity)
- Score-based
    - For each feature, extract the probability of yes or no
    - Rule-based

Testing

- We will provide an opportunity to let your try your answer before 12/19 23:59pm

- Please submit your files to moodle
    - 6 testing results:12/24:3,12/19:3

Policy

- 1-6 members for each team
- Testing deadline: 12/19 23:59pm
- Final Deadline: 12/26 23:59pm(submit one result)
    - Penalty: each day late -5
    - Submit your file(with your student ids) to our server (the link will be provided later)
- Submit your source code of your methods (source_code.rar)
- Submit your report (report.docx)
- Submit your answer(result.csv)
- Compress all files into a rar or zip file
- 15~20% of your final grade
- Grade: F-score: 50%, report: 50%
    - Normalized score
- Encourage propose your own idea
    - Reporting some failing methods is ok (if it spend much time)
- Discussion but no plagiarism