Report for Delta IoT Tutorials

H.P.

NCKU EE

UCLAB

Urban Science & Computing Lab

Report for Delta IoT Tutorials

- Deadline: 12/8 17:00
  - Penalty: each day late -5
  - Submit to Moodle
- Write down your comments/thinking/reviews for the presentation of each week
  - Total 4 tutorials
  - 250 words for each tutorial
  - Combine into a pdf or a word file

Homework 2: Route/POI Promotion for First Time Visitors

H.P.

NCKU EE

UCLAB

Urban Science & Computing Lab

Route/POI Promotion for First Time Visitors

- **Goal:** Given the large volume of heterogeneous data generated by location-based services, one target is to estimate the visiting preference of users who haven't visited target Point of Interest (POI) / route yet, and return the target user list based on their visiting probabilities
- **Motivation**
  - The location promotion problem is different from existing studies on POI recommendation
  - Help local companies to promote their business more effectively
  - Route package for finding target users
- **Point of Interest (POI)** :A certain location/attraction/place
- **Route**: The visiting location sequence of a certain day.

Challenges

- Difficult cases in real data
  - **Problem 1**: Unsupervised nature:
    - Unobserved" POI-user
    - data sparsity
    - new POIs with no check-in records
  - **Problem 2**: Complex user preference
    - geographical coordinates
    - category information (semantic)
    - footprint orders
    - Time-sensitive behaviors
  - **Problem 3**: Different role in a city
    - Victors or non-visitors

Task

- Given:
  - Users who visited New York and Boston City before
  - Users' check-in records (route-based) in these two cities
    - with holiday information
  - The attribute of user id in NY is known, but in testing set (in Boston) the user id is unknown
    - But I will let you know the corresponding user ids of 10% routes in Boston.
  - The category information of all POIs
- You need to
  - For each unknown route, guess a ranked user list
    - The list is ranked by their visiting probabilities

NewYork_Data.txt
(row: one-day trajectory)

- **Uid, holiday_indicator, location_id1,hour1, location_id2, hour2, location_id3, hour3, …**

- 44339:0,4afc15ebf964a520cf1f22e3,20,447bf8f1f964a520ec331fe3,21
- 44339:0,4ad10f08f964a5205bdc20e3,21
- 44339:1,4f1d8d0c77166bf6d1425c4a,21,49ef2aabf964a52083681fe3,23
- 44339:1,4d86ae0a99b78cfa7b2bea1f,0,3fd66200f964a52006e91ee3,18
- 44339:0,4e8cb88577c86de9487eda3c,8
- 44339:1,45226caef964a520d73a1fe3,18,4f1d8d0c77166bf6d1425c4a,21
- 15772:0,4ace6c89f964a52078d020e3,9
- 15772:0,3fd66200f964a5205ae91ee3,10,4a6a38cef964a520decc1fe3,11,4a02d6f9f964a5206e711fe3,13,4afc8966f964a520a52322e3,14,4b072abdf964a52099f822e3,14,4c452ca28c1f20a112833d99,15
- 15772:1,438d6b12f964a520322b1fe3,13,4b072abdf964a52099f822e3,14,3fd66200f964a5209ee81ee3,15

  …

BSID_info.txt and NYID_info.txt

- **Location_id latitude, longitude, location_type, country**
- 3fd66200f964a52000ec1ee3 42.345907 -71.087001 Indian Restaurant US
- 3fd66200f964a52003ec1ee3 42.346127 -71.080363 French Restaurant US
- 3fd66200f964a52008ec1ee3 42.348212 -71.085207 Mexican Restaurant US
- 3fd66200f964a5200aec1ee3 42.251665 -71.037348 French Restaurant US
- 3fd66200f964a5200bec1ee3 42.348751 -71.083938 Middle Eastern Restaurant US
- 3fd66200f964a5200dec1ee3 42.347190 -71.079912 Pizza Place US
- 3fd66200f964a5200eec1ee3 42.347729 -71.085331 Steakhouse US
- 3fd66200f964a52010ec1ee3 42.349284 -71.084513 Mexican Restaurant US
- 3fd66200f964a52011ec1ee3 42.347366 -71.079911 Sports Bar US
- 3fd66200f964a52014ec1ee3 42.346080 -71.075363 Bar US

Boston: validation_set1.csv (row: one-day trajectory)

- **Uid, holiday_indicator, location_id1, hour1, location_id2, hour2, location_id3, hour3,….**
- 37491,1,46e55504f964a520b94a1fe3,13
- 159161,0,4acd2941f964a52025cb20e3,18
- 80711,0,40b28c80f964a5200ff81ee3,18
- 80711,1,4a1b1a86f964a520d07a1fe3,11,3fd66200f964a520b9eb1ee3,19,4d21532b04f55941ec8ef485,21
- 80711,0,4e7e2d68722e1a6d66af650f,20
- 80711,0,4b05a60cf964a5202de022e3,9,40b28c80f964a520e3f71ee3,18

Boston: testfile_set1.csv (row: one-day trajectory)

- **Route_id, holiday_indicator, location_id1, hour1, location_id2, hour2, location_id3, hour3,….**
- 1,0,4a649249f964a52048c61fe3,15,4bab6d05f964a52089a83ae3,21
- .
  .
- .
- 79,0,4a02465ef964a52045711fe3,21
- 80,1,4bb7beb27421a5939efcc040,12,40b28c80f964a520a4f71ee3,13,3fd66200f964a52062ec1ee3,18,3fd66200f964a520f6eb1ee3,21
- 81,1,4de19091d22d2a4ecdb18204,19
- 82,1,4aca8793f964a52042c220e3,21,4bde4f0cffdec9281174e8a1,22
- 83,1,40b28c80f964a520f6f71ee3,0
- 84,0,4b05a60cf964a5202de022e3,11
- 85,0,4b05a60cf964a5202de022e3,17,40b28c80f964a520fdf71ee3,18
  .
  .
  .
- 24884,1,3fd66200f964a520e4eb1ee3,12
- 24885,1,4afdc0dff964a520902a22e3,15,40b28c80f964a52038f81ee3,16,4a440986f964a52011a71fe3,18
- 24886,1,43be0c22f964a5200c2d1fe3,13

Effectiveness

- Average Ranking Ratio (ARR)
  - For each unknown route, we calculate ranking percentage of recommending list
    - the hit position/ total number of users
  - Average all users

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 initial results
  - or how you exploit the labeled data from Boston city

Report format (2) Questions

- What are the physical meanings of your proposed methods? Why do you want to do that?
- If same approach results in much different performance among cities, what are the possible reasons you think?
- Do you have other idea for solving this problem? (but no time to try)

Testing

- We will provide an opportunity to let your try your answer before 12/11 23:59pm
  - Please submit your files to our server (the link will be provided later)

Policy

- 1-2 members for each team
- Deadline: 12/20 23:59pm
  - Penalty: each day late -5
  - Submit your file 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
  - One route one file(so total 24886 files), you should name it using route id
    - each row has only one user id
      - E.g.
        userID1
        userID2
        userID3
        .
        .
        .
    - Total number of lines for each file: 2563
    - Ranked by visiting probabilities of users (from high to low)
- Compress all files into a rar or zip file
- 30% ~ 40% of your final grade
- Grade: ARR: 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

Hints

- Capture personal behaviors from some perspectives (some are mutually influenced)
  - Route length
  - Temporal preference
    - hour or date
  - Category preference (category hierarchy)
    - https://developer.foursquare.com/docs/resources/categories
- Divide cases for holidays
- Try some methods on validation data and see the effectiveness