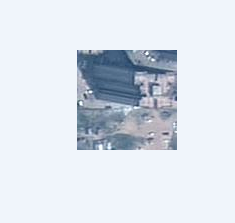
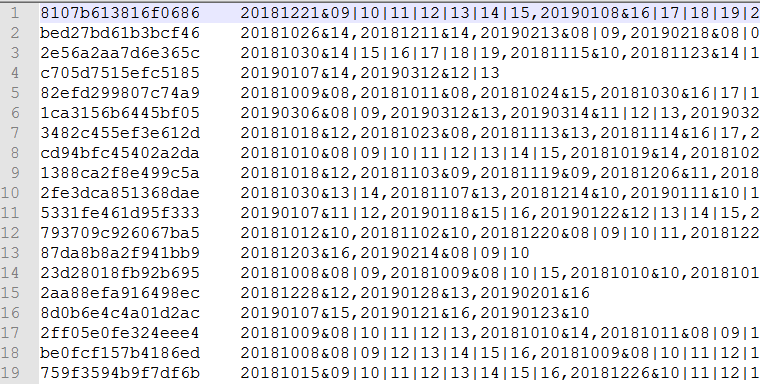
# Urban Region Function Classification

Website: <https://dianshi.baidu.com/competition/30/question>

## The goal:

Class the image (100\*100 pixels)/User visit time data to 9 classes.

|  |  |
| --- | --- |
| CategoryID | Functions of Areas |
| 001 | Residential area |
| 002 | School |
| 003 | Industrial park |
| 004 | Railway station |
| 005 | Airport |
| 006 | Park |
| 007 | Shopping area |
| 008 | Administrative district |
| 009 | Hospital |

### Initial analyses:

1. Neural network must be used (huge data, no hard rule to classify them).
2. Images can only offer limited information (textures are very similar, such as residential area, school, shopping area and others). The User visits information is very important.

### Things need to do:

1. How to extract the user visit information?
2. How to combine the image and visit information together?
3. Data augmentation?

## 5/4/19

Data preparing.

1. Extract the visit information.
2. Total visitor (the number of records) (170000 -1, how to normalize it?)
3. 24/12 hours, the number of visits.
4. ~~12 months, the number of visits. (September,2018 ----February, 2019, no enough data)~~
5. The return times of each visitor.
6. The average stay time of every day.
7. The size of data of the record. (9MB-1k)

## 5/7/19

Normalization: assume all data are following the normal distributing, then estimate the parameters, then normalize them to 0-1.

Mean-std:

1. 1401.671325,3315.93847939731🡪 /1400-2
2. 3.5713300572068145,1.469428439705176🡪/3.57-2
3. 0.3036313519477844,0.4598593726143714🡪/0.30-2

## 5/8/19

1. Do the training sample balance.
2. Prepare the CNN.
3. Train the CNN.

5/10/19

1. Retrain the CNN, added the image augmentation.

5/13/19

The first try results.



Be careful about the space in the txt.

5/14/19

1. Include the raw visit data for the training.
2. Train the network without visit data and generate the confusion matrix.
3. Generate the CNN image features.

The test class spy:

1: 0.2354

2: 0.1833

3: 0.0923

1. Generate the confusion matrix.
2. Check the combination of the image and visit features.

The features from image: most of them are 0 and the range is around(0,8);

1. Try re-normalize the visit features into 0-8.// no obvious improvement.
2. Try add one fcl before merge.
3. Reduce the image feature dimension. // no obvious improvement.
4. Try different combination length.
5. Try the other baseline.