Lab: DNN

Prof. Chia-Yu Lin

Behavior Classification of Exposition Visitors

Data Source

- [AIdea-Behavior Classification of Exposition Visitors](#)

Dataset Description

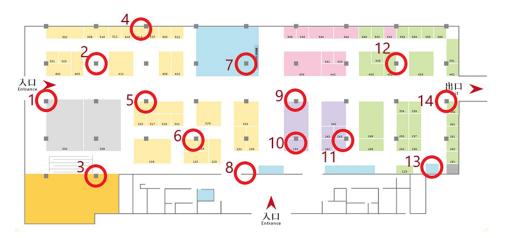1) **train.csv & test.csv**
   - mac_hash：User ID
   - sniffer_loc：User location (14 loc)
   - created_time：User data collection time

2) **training-label.csv**
   - mac_hash： User ID
   - label： User class (5 class, 0~4)

Data Preprocessing

1) Drop **'created time'** column.

2) Count whether each mac_hash (user) has passed through these 14 locations, and generate a training dataframe, the format as follows.

	mac_hash	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	0003ae8541d0e925fcee242287e2ad27	0	0	0	0	0	0	0	0	0	0	1	0	0	0
1	00078611037990f7f36b722f22595fe7	0	1	1	1	1	1	0	1	0	1	1	0	0	0
2	000ce31739b333d931813de403156844	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	0027240b2aeb1f7f3c053c073cafb452	0	1	0	1	1	1	1	1	1	1	1	1	1	1
4	0028a89964416dc9a8a66cc7fe25ef4e	0	1	1	1	0	1	0	0	1	1	1	1	1	1

3) Combine "training-label.csv" to add 'label' column.

	mac_hash	1	2	3	4	5	6	7	8	9	10	11	12	13	14	label
0	0003ae8541d0e925fcee242287e2ad27	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
1	00078611037990f7f36b722f22595fe7	0	1	1	1	1	1	0	1	0	1	1	0	0	0	1
2	000ce31739b333d931813de403156844	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3
3	0027240b2aeb1f7f3c053c073cafb452	0	1	0	1	1	1	1	1	1	1	1	1	1	1	4
4	0028a89964416dc9a8a66cc7fe25ef4e	0	1	1	1	0	1	0	0	1	1	1	1	1	1	4

Load Data

1) Load **train.csv** and **training-label.csv** file as dataframe

	mac_hash	sniffer_loc	created_time
0	10f6ad2e3fa3b3da8ceb3fbd95658681	2	2018-12-08 11:47:48
1	10f6ad2e3fa3b3da8ceb3fbd95658681	10	2018-12-08 12:37:34
2	10f6ad2e3fa3b3da8ceb3fbd95658681	9	2018-12-08 12:37:35
3	10f6ad2e3fa3b3da8ceb3fbd95658681	6	2018-12-08 12:44:48
4	10f6ad2e3fa3b3da8ceb3fbd95658681	4	2018-12-08 12:44:54

	mac_hash	label
0	10f6ad2e3fa3b3da8ceb3fbd95658681	4
1	5272841b630d74290ec129c64dd6378b	0
2	8ab74d094708a27c9660e897b93a50f9	3
3	3d00887dd6e574dd63974bc6c69b23b0	1
4	6461fce7f6f184401c961fda75069262	1

Training

1) Use sklearn **train_test_split()** to split your dataset
   – training dataset : validation dataset = 8 : 2
   – random_state = 0
2) Build **DNN model** and start training
3) validate your model using validation dataset
   – Use sklearn **classification_report()** to show your validation result

	precision	recall	f1-score	support
0	0.98	1.00	0.99	436
1	0.98	0.95	0.97	190
2	0.99	0.99	0.99	136
3	0.92	1.00	0.96	93
4	1.00	0.95	0.98	259
accuracy			0.98	1114
macro avg	0.97	0.98	0.98	1114
weighted avg	0.98	0.98	0.98	1114

Submit

1) Use your model to predict the class of each mac_hash in test.csv, and generate submit.csv. The format of submit.csv is as follows.

	mac_hash	label
0	004c76bfe9f7b20f3651e0bcbd8eee9e	1
1	0084303d531091816387956ea26b86eb	1
2	00872a1ee635c2619454350b96b94103	4
3	00a5f69fc7719ba29c7b202f292dbafb	4
4	00ad41313f1f637e079c1a9d9a857d12	4
...	...	...
1386	ff6883d04de66c6d4859928ec9d3f642	0
1387	ff7b41d3b33fb7d0b6704f1093173c40	4
1388	ffa5692a2dac937f456280c433961ea6	0
1389	ffcaf0f29c6bede7b5cd9bc807c31715	1
1390	ffd0c4403c3c44dcbe4772821b032252	4

1391 rows × 2 columns

2) Finally, compress your code and submit.csv together as .zip file and upload it (file name is s+student id)

Deadline: 2022/12/30 23:30

Fighting !