SDML Final - HW2 Extension

B04902016曾奕青、B04902103蔡昀達、B04902105戴培倫

Abstract

- Data Problems
- Analysis
- Models

Data Problems

1. Not next day: 170 users
2. Incomplete training data: 2280 users

Training data

```
2015-01-27,85,247
2015-01-27,85,291
2015-01-27,85,217
2015-01-27,85,19
2015-01-27,85,42
2015-01-27,85,81
2015-01-28,85,97
2015-01-28,85,39
2015-01-28,85,97
2015-01-28,85,1315
2014-09-15,86,382
2014-09-15,86,34
```

Testing data

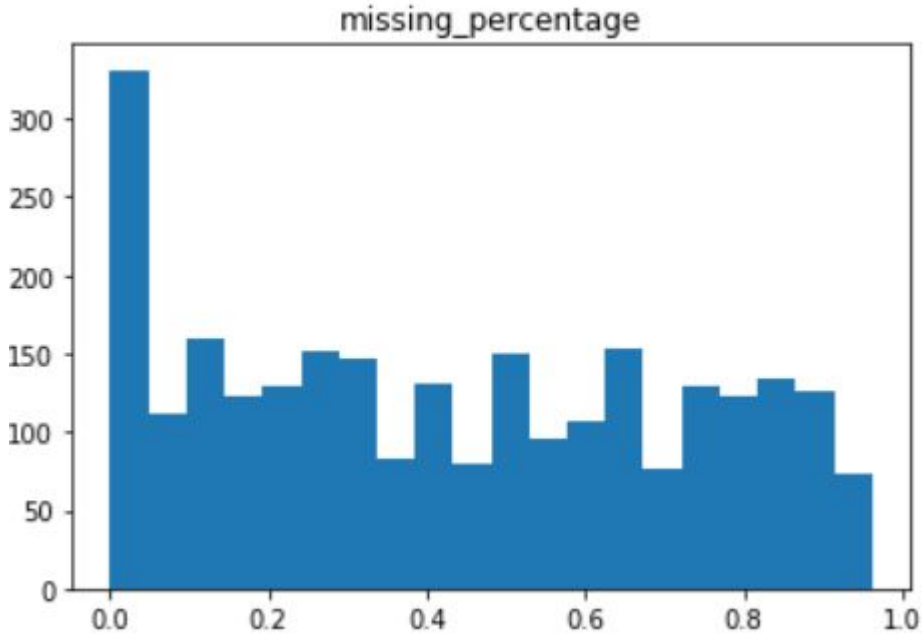
```
2015-03-12,82,81
2015-03-12,82,125
2015-01-28,85,39
2015-01-28,85,217
2015-01-28,85,1669
2015-02-01,85,19
2015-02-01,85,42
2015-02-01,85,42
2015-02-01,85,943
2015-02-01,85,1682
2015-02-01,85,228
2015-02-01,85,224
2015-02-01,85,1700
2015-02-01,85,39
2015-02-01,85,1451
2015-02-01,85,39
2015-02-01,85,39
2015-02-01,85,42
2015-02-02,85,228
2015-02-02,85,1682
```

Kaggle answer

85	{1682, 19, 1700, 39, 228, 224, 42, 1451, 943}

The answer is the food eaten on 2015/02/01, which is not the next day of 2015/01/28.

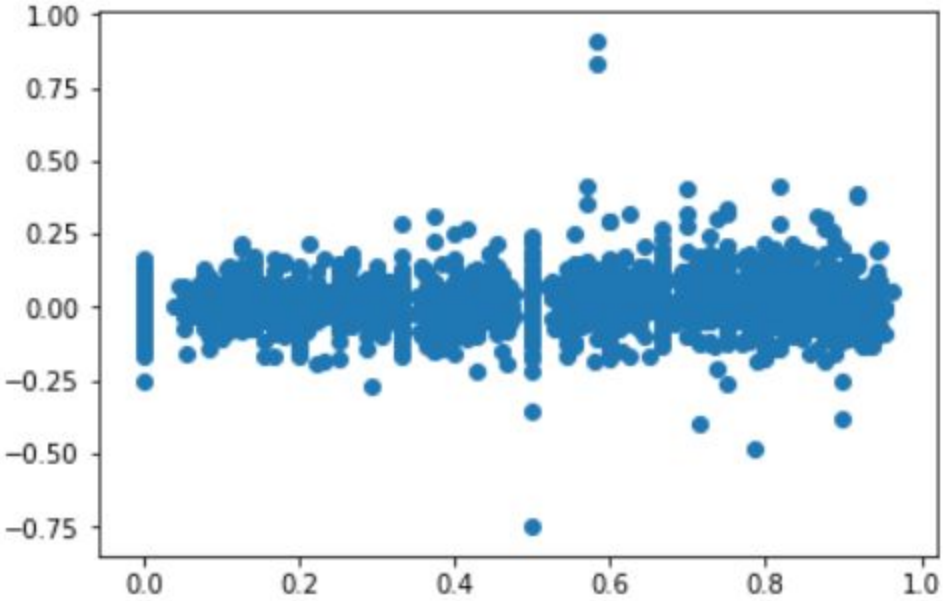
	All users	Next day	Not next day
Baseline	0.36164	0.36551	0.30603
RNN	0.29445	0.29661	0.26352


missing_percentage
300
250
200
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0

	Original	Complete
Baseline	0.35121	0.36164
RNN	0.29101	0.29445

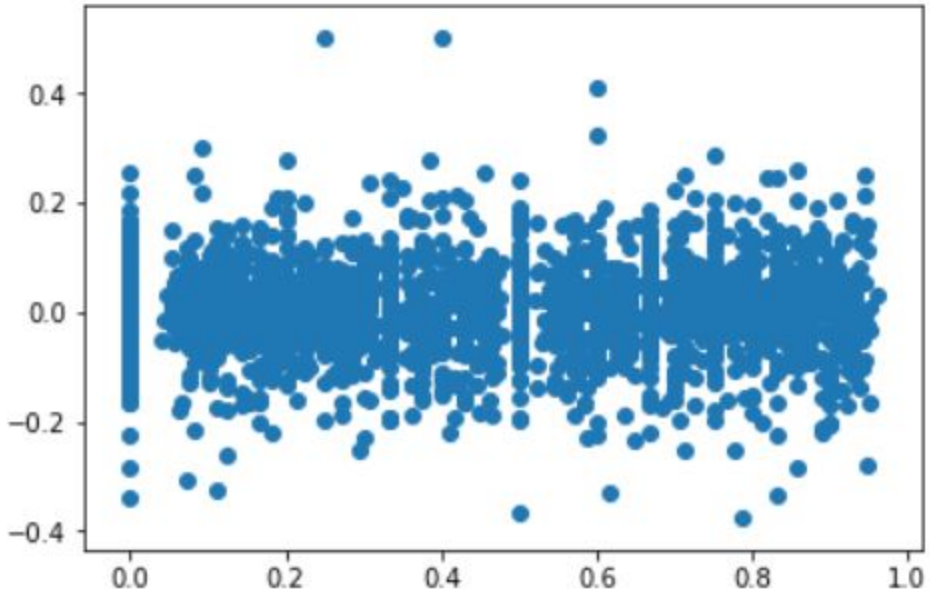
```
plt.scatter(missing_percentage, np.array(baseline_new_score) - np.array(baseline_old_score))
```

```
<matplotlib.collections.PathCollection at 0x7fa2798ae4e0>
```



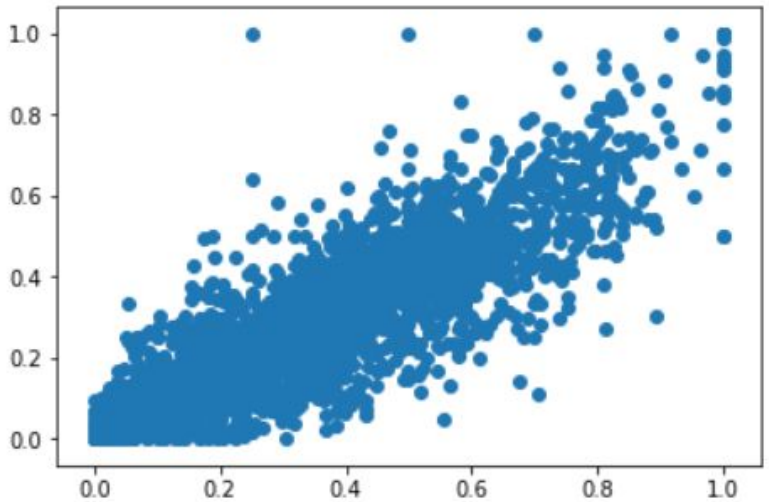
```
plt.scatter(missing_percentage, np.array(rnn_new_score) - np.array(rnn_old_score))
```

```
<matplotlib.collections.PathCollection at 0x7fa278557518>
```



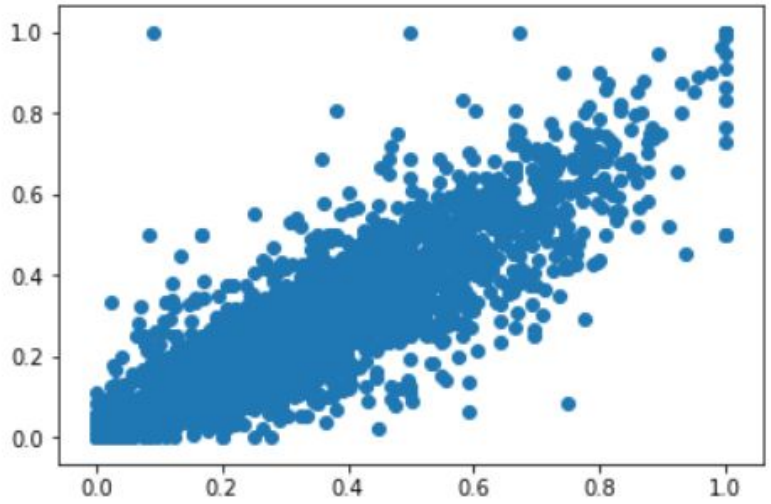
```
plt.scatter(baseline_new_score, rnn_new_score)
```

```
<matplotlib.collections.PathCollection at 0x7fa27
```

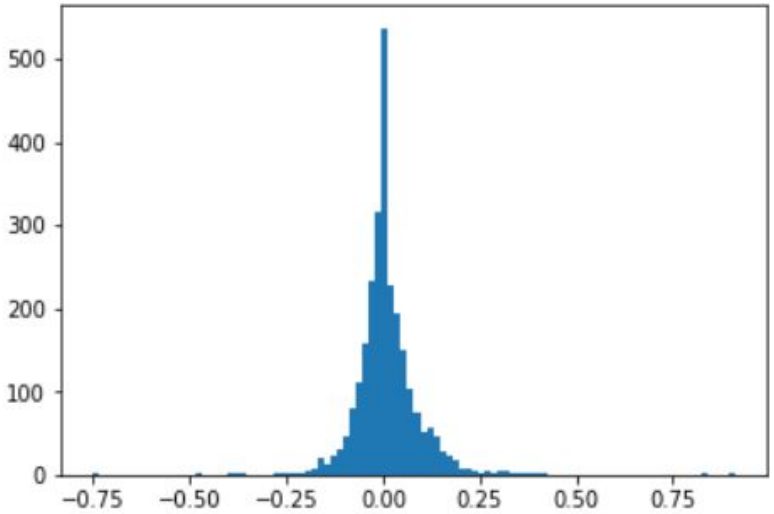



```
plt.scatter(baseline_old_score, rnn_old_score)
```

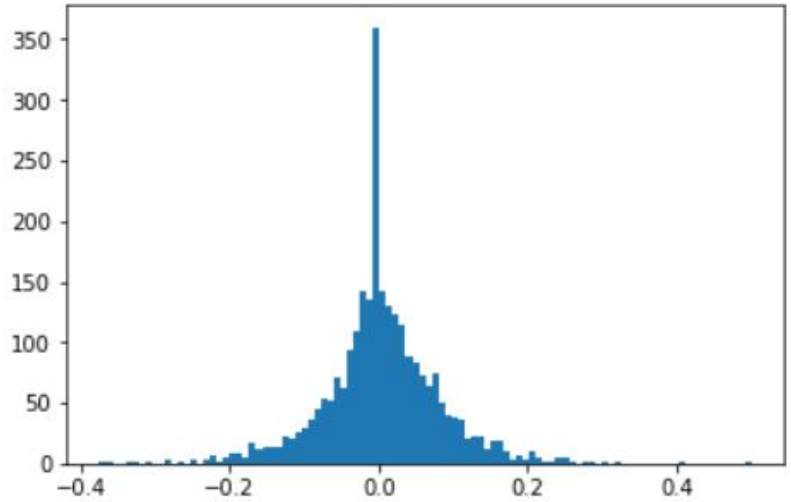
```
<matplotlib.collections.PathCollection at 0x7fa2786b79e8>
```



```
plt.hist(np.array(baseline_new_score) - np.array(
plt.show()
```




```
a = plt.hist(np.array(rnn_new_score) - np.array(rnn_old_score)
plt.show()
```



Metrics

	MAP@20	MRR@20	Recall@20	Find_all
Baseline	0.36164	0.79332	0.53795	0.0099
RNN	0.29445	0.76185	0.45176	0.0086

Remove some data

	Original			
RNN	0.294			

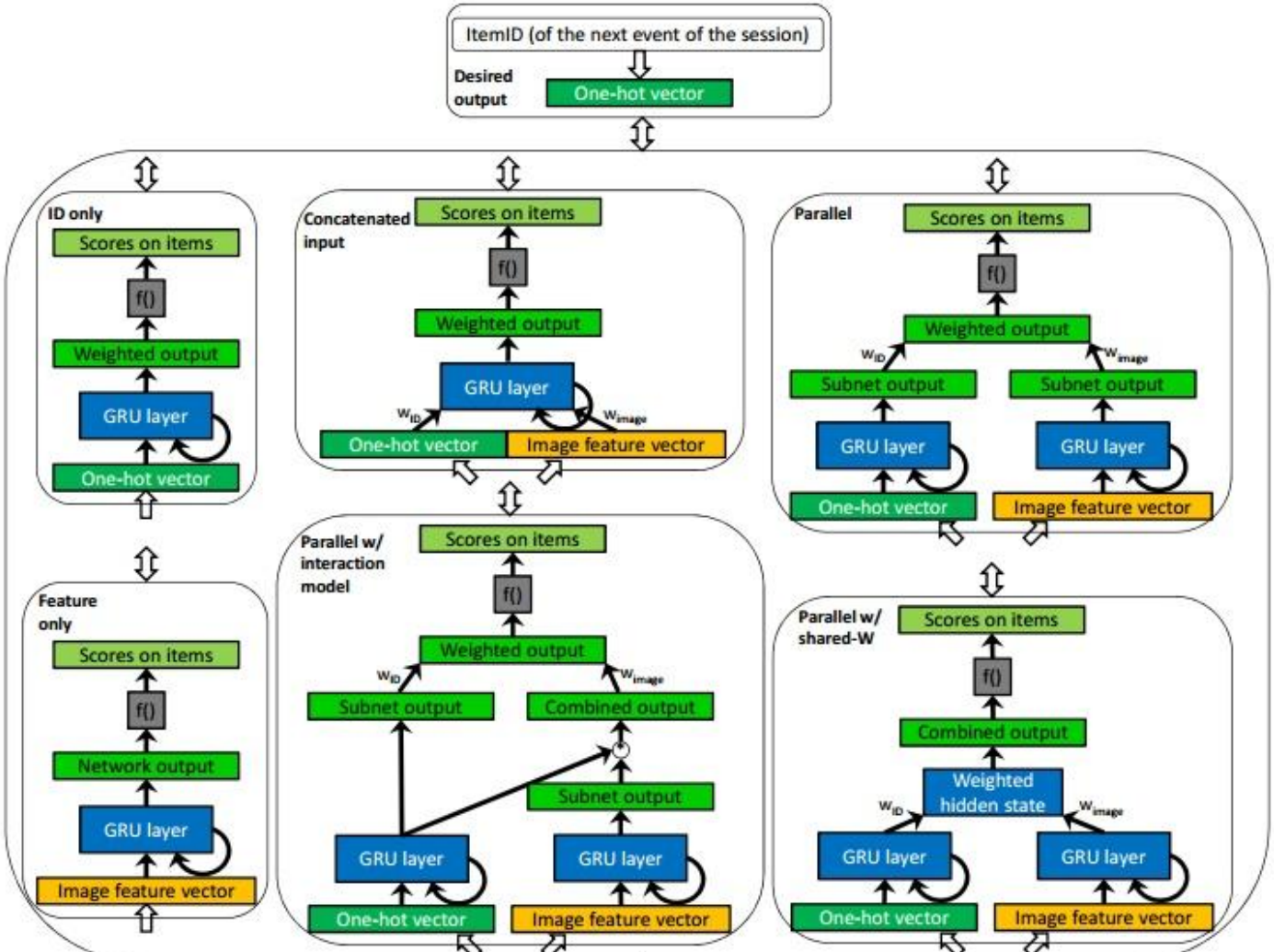
Remove some data

	Original	Remove 3	Remove 5	
RNN	0.294	0.291	0.292	

Remove some data

	Original	Remove 3	Remove 5	Remove 5, Keep recent
RNN	0.294	0.291	0.292	0.287

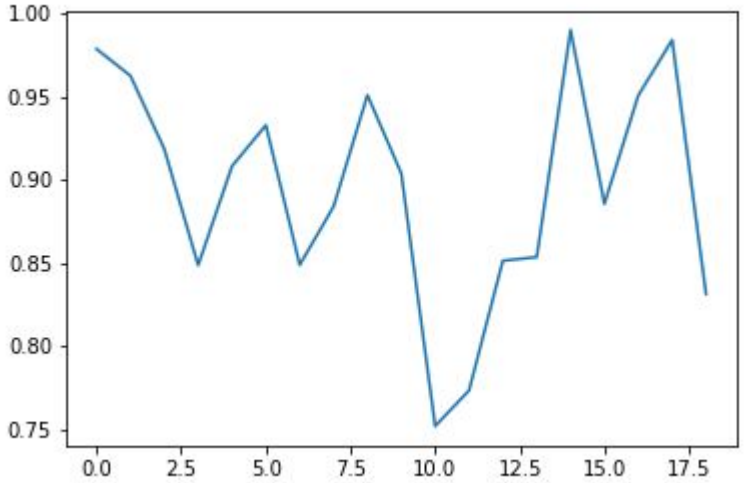
Parallel Recurrent Neural Network Architectures for Feature-rich Session-based Recommendations


ItemID (of the next event of the session)
Desired output
One-hot vector
ID only
Scores on items
f()
Weighted output
GRU layer
One-hot vector
Concatenated input
Scores on items
f()
Weighted output
GRU layer
W_{ID}
W_{image}
One-hot vector
Image feature vector
Parallel
Scores on items
f()
Weighted output
W_{ID}
W_{image}
Subnet output
Subnet output
GRU layer
GRU layer
One-hot vector
Image feature vector
Feature only
Scores on items
f()
Network output
GRU layer
Image feature vector
Parallel w/ interaction model
Scores on items
f()
Weighted output
W_{ID}
W_{image}
Subnet output
Combined output
Subnet output
GRU layer
GRU layer
One-hot vector
Image feature vector
Parallel w/ shared-W
Scores on items
f()
Combined output
Weighted hidden state
W_{ID}
W_{image}
GRU layer
GRU layer
One-hot vector
Image feature vector

Food Features

58 columns

整理後 → 19 class + 10 ingredients



Performance

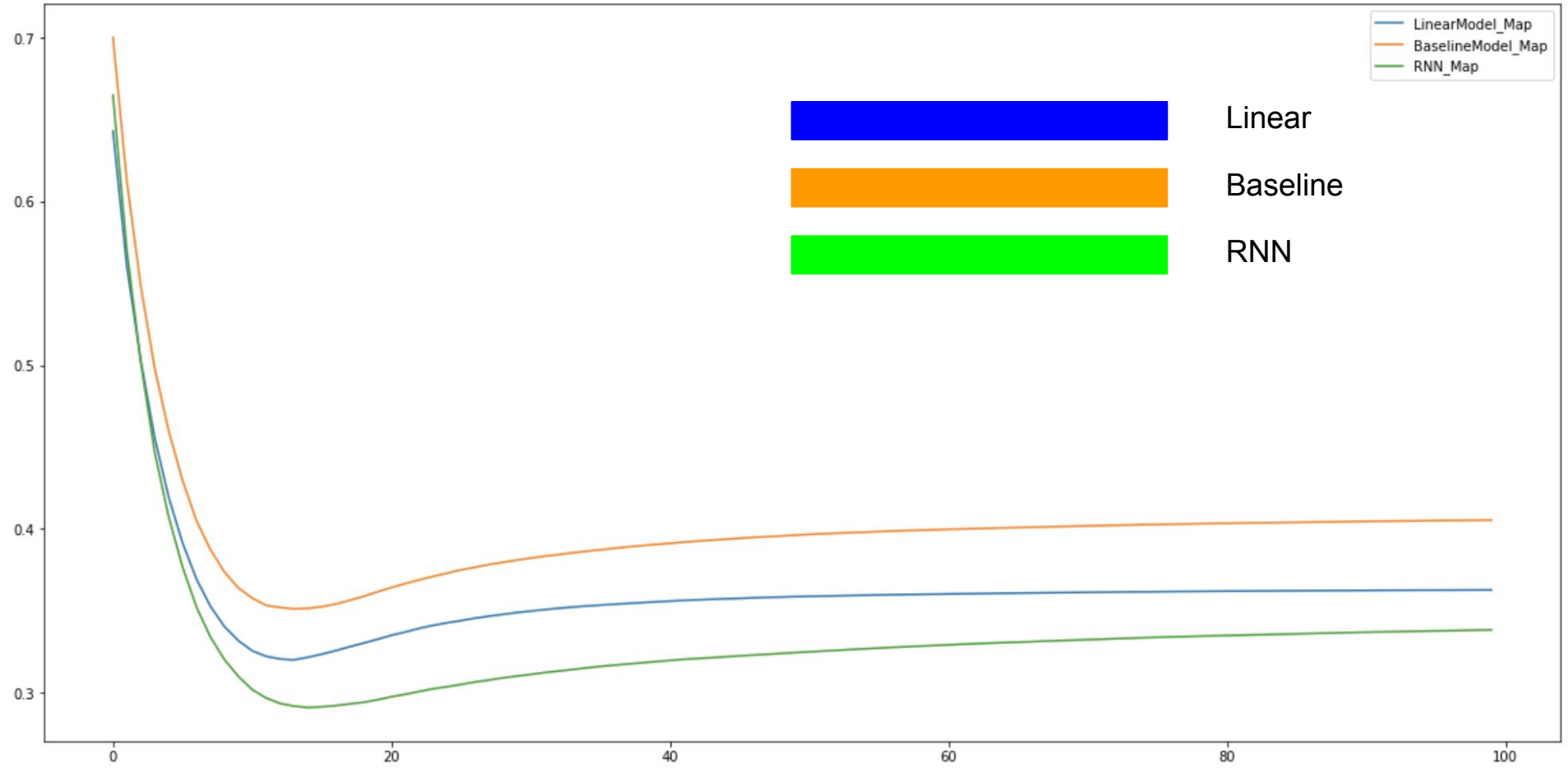
My performance

→ 加入feature後穩定下降0.01

Method	Recall@20	MRR@20
Item-kNN	0.6263	0.3740
ID only	0.6831 (+9.07%)	0.3847 (+2.85%)
ID only (200)	0.6963 (+11.17%)	0.3881 (+3.77%)
Feature only	0.5367 (−14.30%)	0.3065 (−18.05%)
Concatenated	0.6766 (+8.03%)	0.3850 (+2.94%)
Parallel (sim)	0.6765 (+8.01%)	0.4014 (+7.34%)
Parallel (alt)	0.6874 (+9.76%)	0.4331 (+15.81%)
Parallel (res)	0.7028 (+12.21%)	**0.4440** (+18.72%)
Parallel (int)	**0.7040** (+12.41%)	0.4361 (+16.60%)
Shared-W (sim)	0.6681 (+6.66%)	0.4007 (+7.13%)
Shared-W (alt)	0.6804 (+8.63%)	0.4035 (+7.88%)
Shared-W (res)	0.6425 (+2.58%)	0.3541 (−5.31%)
Shared-W (int)	0.6658 (+6.31%)	0.3715 (−0.66%)
Int. model (sim)	0.6751 (+7.78%)	0.3998 (+6.90%)
Int. model (alt)	0.6847 (+9.32%)	0.4104 (+9.74%)
Int. model (res)	0.6749 (+7.76%)	0.4098 (+9.56%)
Int. model (int)	0.6843 (+9.25%)	0.4040 (+8.02%)

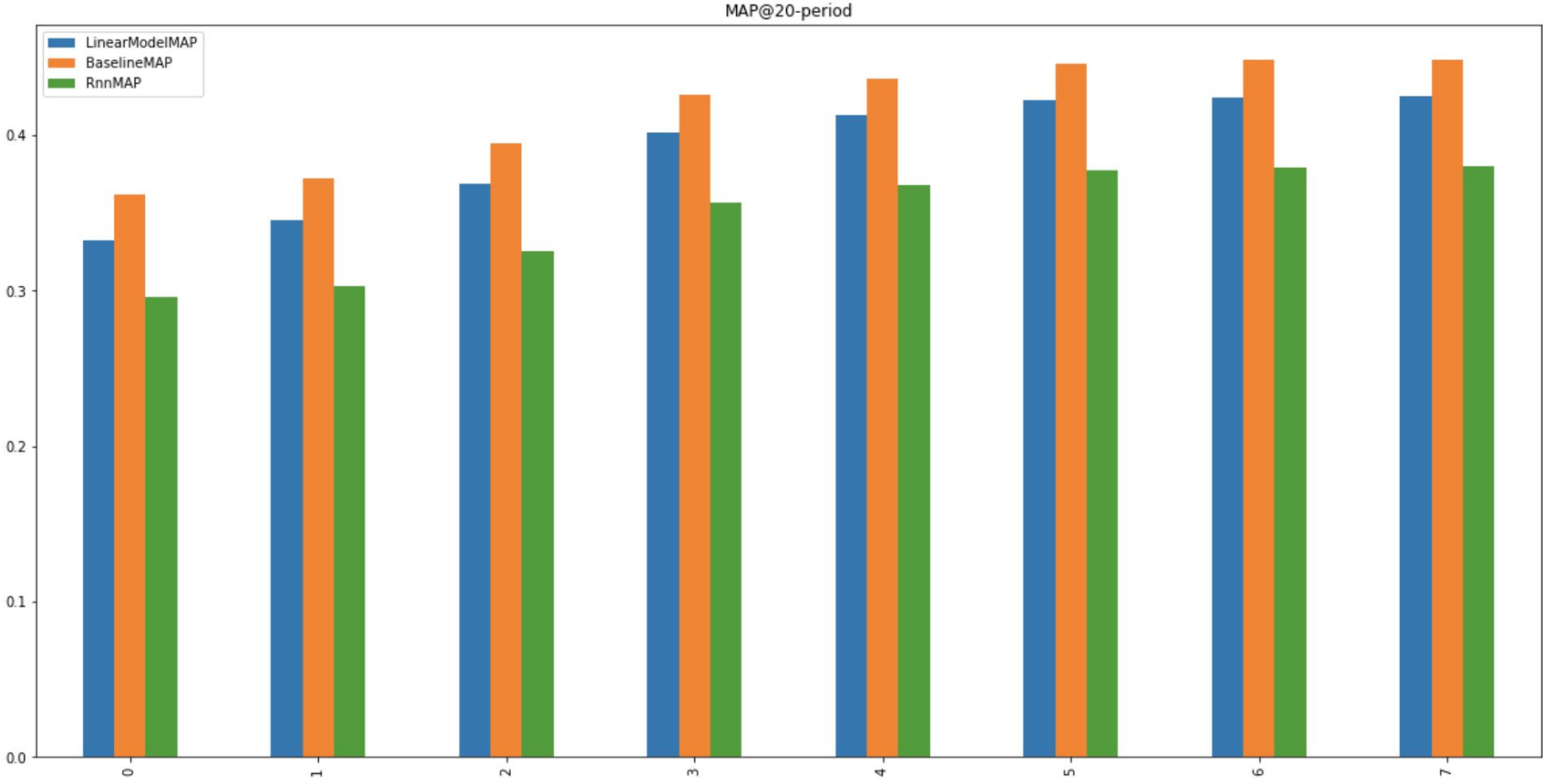
Model Comparison

- MAP@k, k in [1, 100]


LinearModel_Map
BaselineModel_Map
RNN_Map
0.7
0.6
0.5
0.4
0.3
0
20
40
60
80
100
Linear
Baseline
RNN

Model Comparison

- MAP@k, k in [1, 100]
- Ans = foods eaten in next k days


MAP@20-period
LinearModelMAP
BaselineMAP
RnnMAP
0.0
0.1
0.2
0.3
0.4
0
1
2
3
4
5
6
7

Model Comparison

- MAP@k, k in [1, 100]
- Ans = foods eaten in next k days
- Predict on different day

RNN vs Baseline Model

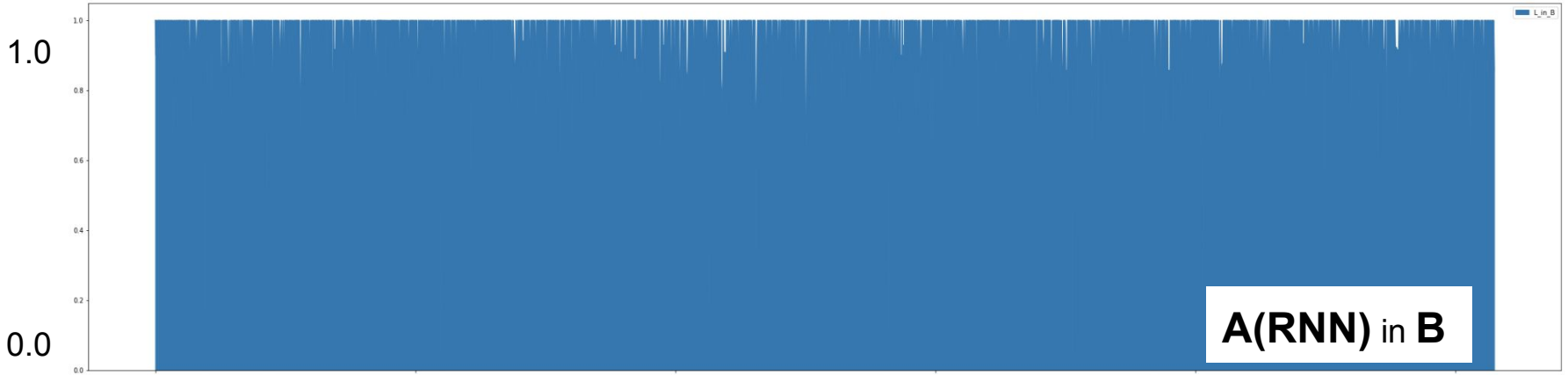
A: correct foods in RNN

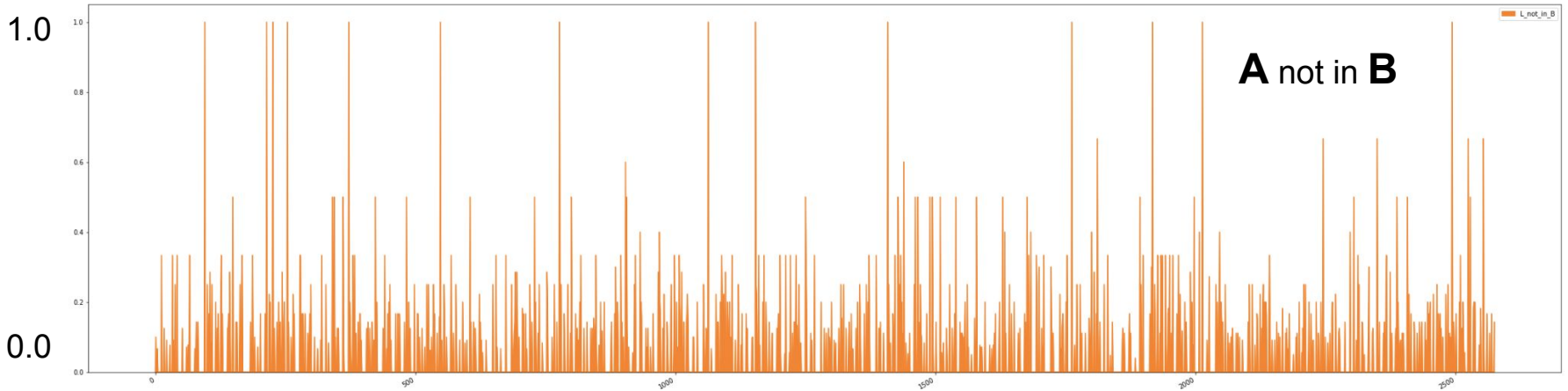
B: correct foods in Baseline Model

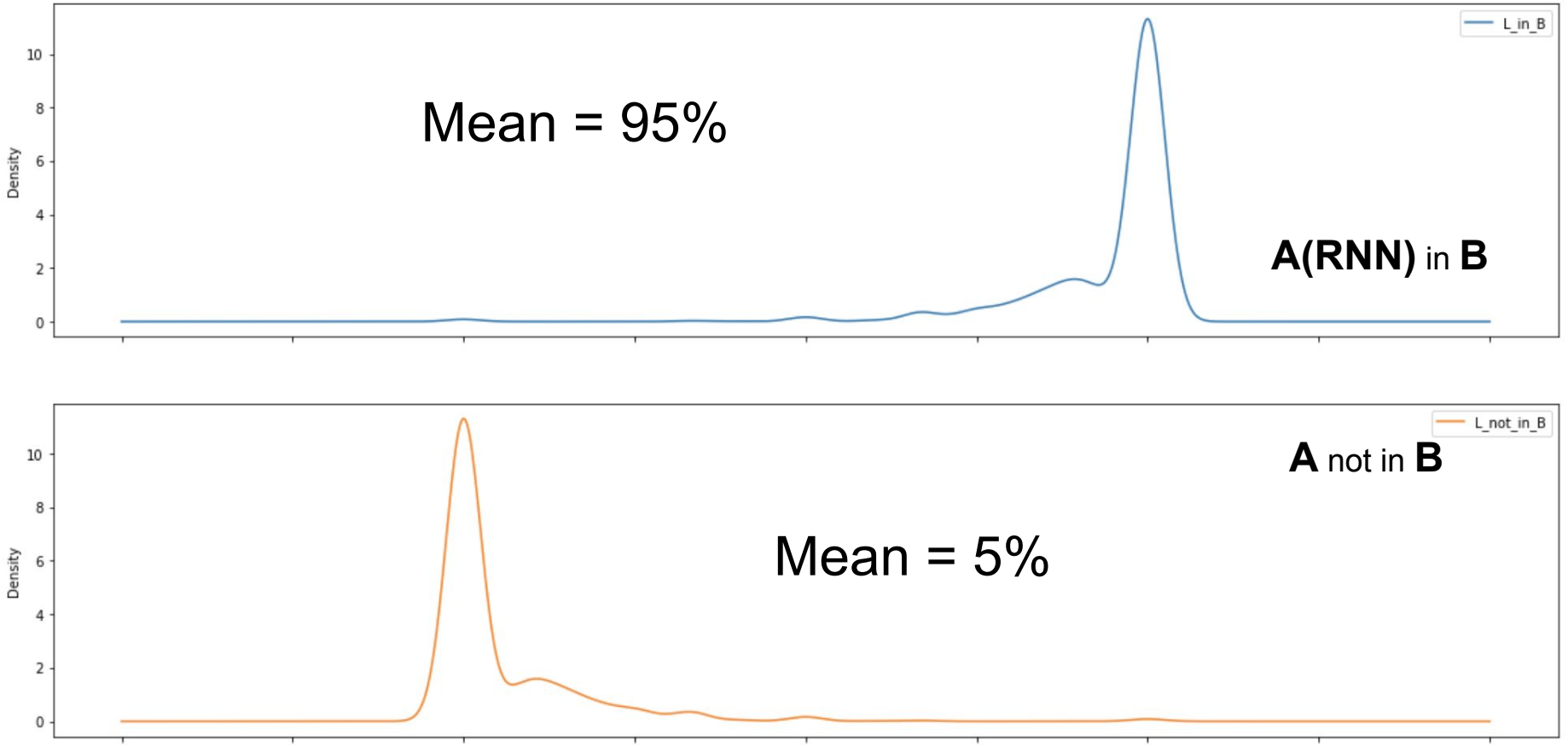
RNN vs Baseline Model

A: correct foods in RNN

B: correct foods in Baseline Model


1.0
0.0
L_in_B
A(RNN) in B


1.0
0.0
L_not_in_B
A not in B
0
500
1000
1500
2000
2500


L_in_B
Density
10
8
6
4
2
0
Mean = 95%
A(RNN) in B
L_not_in_B
A not in B
Mean = 5%

MF+RNN+GAN model

MF+RNN+GAN model