Day 2

Please run the following code to update the recoflow package

In [89]:

!pip install recoflow --upgrade
import recoflow

```
Requirement already up-to-date: recoflow in /usr/local/lib/python3.6/dist-packages (0.0.7)
Requirement already satisfied, skipping upgrade: altair<4.0,>=3.2 in /usr/local/lib/python3.6/dist-packages
(from recoflow) (3.2.0)
Requirement already satisfied, skipping upgrade: pandas<0.26.0,>=0.25.1 in /usr/local/lib/python3.6/dist-pac
kages (from recoflow) (0.25.1)
Requirement already satisfied, skipping upgrade: scikit-learn<0.22.0,>=0.21.3 in /usr/local/lib/python3.6/di
st-packages (from recoflow) (0.21.3)
Requirement already satisfied, skipping upgrade: numpy<2.0,>=1.17 in /usr/local/lib/python3.6/dist-packages
(from recoflow) (1.17.2)
Requirement already satisfied, skipping upgrade: matplotlib<4.0,>=3.1 in /usr/local/lib/python3.6/dist-packa
ges (from recoflow) (3.1.1)
Requirement already satisfied, skipping upgrade: requests<3.0,>=2.22 in /usr/local/lib/python3.6/dist-packag
es (from recoflow) (2.22.0)
Requirement already satisfied, skipping upgrade: keras<3.0,>=2.3 in /usr/local/lib/python3.6/dist-packages
(from recoflow) (2.3.0)
Requirement already satisfied, skipping upgrade: entrypoints in /usr/local/lib/python3.6/dist-packages (from
altair<4.0,>=3.2->recoflow) (0.3)
Requirement already satisfied, skipping upgrade: six in /usr/lib/python3/dist-packages (from altair<4.0,>=3.
2->recoflow) (1.11.0)
Requirement already satisfied, skipping upgrade: jsonschema in /usr/local/lib/python3.6/dist-packages (from
altair<4.0,>=3.2->recoflow) (3.0.2)
Requirement already satisfied, skipping upgrade: toolz in /usr/local/lib/python3.6/dist-packages (from altai
r<4.0,>=3.2->recoflow) (0.10.0)
Requirement already satisfied, skipping upgrade: jinja2 in /usr/local/lib/python3.6/dist-packages (from alta
ir<4.0,>=3.2->recoflow) (2.10.1)
Requirement already satisfied, skipping upgrade: pytz>=2017.2 in /usr/local/lib/python3.6/dist-packages (fro
m pandas<0.26.0,>=0.25.1->recoflow) (2019.2)
Requirement already satisfied, skipping upgrade: python-dateutil>=2.6.1 in /usr/local/lib/python3.6/dist-pac
kages (from pandas<0.26.0,>=0.25.1->recoflow) (2.8.0)
Requirement already satisfied, skipping upgrade: scipy>=0.17.0 in /usr/local/lib/python3.6/dist-packages (fr
om scikit-learn<0.22.0,>=0.21.3->recoflow) (1.3.1)
Requirement already satisfied, skipping upgrade: joblib>=0.11 in /usr/local/lib/python3.6/dist-packages (fro
m scikit-learn<0.22.0,>=0.21.3->recoflow) (0.13.2)
Requirement already satisfied, skipping upgrade: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local/lib/
python3.6/dist-packages (from matplotlib<4.0,>=3.1->recoflow) (2.4.2)
Requirement already satisfied, skipping upgrade: kiwisolver>=1.0.1 in /usr/local/lib/python3.6/dist-packages
(from matplotlib<4.0,>=3.1->recoflow) (1.1.0)
Requirement already satisfied, skipping upgrade: cycler>=0.10 in /usr/local/lib/python3.6/dist-packages (fro
m matplotlib<4.0,>=3.1->recoflow) (0.10.0)
Requirement already satisfied, skipping upgrade: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/lib/p
ython3.6/dist-packages (from requests<3.0,>=2.22->recoflow) (1.24.2)
Requirement already satisfied, skipping upgrade: idna<2.9,>=2.5 in /usr/local/lib/python3.6/dist-packages (f
rom requests<3.0,>=2.22->recoflow) (2.8)
Requirement already satisfied, skipping upgrade: chardet<3.1.0,>=3.0.2 in /usr/local/lib/python3.6/dist-pack
ages (from requests<3.0,>=2.22->recoflow) (3.0.4)
Requirement already satisfied, skipping upgrade: certifi>=2017.4.17 in /usr/local/lib/python3.6/dist-package
s (from requests<3.0,>=2.22->recoflow) (2019.6.16)
```

```
s<3.0,>=2.3->recoflow) (5.1.2)
Requirement already satisfied, skipping upgrade: keras-applications>=1.0.6 in /usr/local/lib/python3.6/dist-
packages (from keras<3.0,>=2.3->recoflow) (1.0.8)
Requirement already satisfied, skipping upgrade: h5py in /usr/local/lib/python3.6/dist-packages (from keras<
3.0, >= 2.3 - \text{recoflow} (2.9.0)
Requirement already satisfied, skipping upgrade: keras-preprocessing>=1.0.5 in /usr/local/lib/python3.6/dist
-packages (from keras<3.0,>=2.3->recoflow) (1.1.0)
Requirement already satisfied, skipping upgrade: setuptools in /usr/local/lib/python3.6/dist-packages (from
jsonschema->altair<4.0,>=3.2->recoflow) (41.0.1)
Requirement already satisfied, skipping upgrade: pyrsistent>=0.14.0 in /usr/local/lib/python3.6/dist-package
s (from jsonschema->altair<4.0,>=3.2->recoflow) (0.15.4)
Requirement already satisfied, skipping upgrade: attrs>=17.4.0 in /usr/local/lib/python3.6/dist-packages (fr
om jsonschema->altair<4.0,>=3.2->recoflow) (19.1.0)
Requirement already satisfied, skipping upgrade: MarkupSafe>=0.23 in /usr/local/lib/python3.6/dist-packages
(from jinja2->altair<4.0,>=3.2->recoflow) (1.1.1)
WARNING: You are using pip version 19.1.1, however version 19.2.3 is available.
You should consider upgrading via the 'pip install --upgrade pip' command.
```

Requirement already satisfied, skipping upgrade: pyyaml in /usr/local/lib/python3.6/dist-packages (from kera

Data Preprocessing

```
import numpy as np
import pandas as pd

In [4]:

ratings = pd.read_csv("/tf/data/ratings.csv")
users = pd.read_csv("/tf/data/users.csv")
items = pd.read_csv("/tf/data/items.csv")
```

In [6]:

from recoflow.preprocessing import EncodeUserItem, StratifiedSplit

```
In [8]:
# Encoding the data
interaction, n users, n items, user encoder, item encoder = EncodeUserItem(ratings,
                                                                           "user id",
                                                                           "movie id",
                                                                           "rating",
                                                                           "unix timestamp")
Number of users: 943
Number of items: 1682
In [18]:
train, test = StratifiedSplit(interaction, [0.8, 0.2])
train.shape, test.shape
Out[18]:
((80000, 7), (20000, 7))
In [46]:
min rating = interaction.RATING.min()
max rating = interaction.RATING.max()
min rating, max rating
Out[46]:
(1, 5)
```

Model - Deep Facorization

```
In [47]:
```

```
from keras.models import Model
from keras.layers import Embedding, Input, Dot, Add, Activation, Lambda, Concatenate, Dense, Flatten, Dropout
from keras.regularizers import 12
from keras.optimizers import Adam
```

In [48]:

Embedding?

```
Init signature:
Embedding(
    input dim,
    output dim,
    embeddings initializer='uniform',
    embeddings regularizer=None,
    activity regularizer=None,
    embeddings constraint=None,
    mask zero=False,
    input length=None,
    **kwargs,
Docstring:
Turns positive integers (indexes) into dense vectors of fixed size.
eg. [[4], [20]] \rightarrow [[0.25, 0.1], [0.6, -0.2]]
This layer can only be used as the first layer in a model.
# Example
```python
 model = Sequential()
 model.add(Embedding(1000, 64, input length=10))
 # the model will take as input an integer matrix of size (batch, input length).
 # the largest integer (i.e. word index) in the input should be
 # no larger than 999 (vocabulary size).
 # now model.output shape == (None, 10, 64), where None is the batch dimension.
 input array = np.random.randint(1000, size=(32, 10))
 model.compile('rmsprop', 'mse')
 output array = model.predict(input array)
 assert output array.shape == (32, 10, 64)
Arguments
 input dim: int > 0. Size of the vocabulary,
 i.e. maximum integer index + 1.
 output dim: int >= 0. Dimension of the dense embedding.
 embeddings initializer: Initializer for the `embeddings` matrix
 (see [initializers](../initializers.md)).
 embeddings regularizer: Regularizer function applied to
 the `embeddings` matrix
 (see [regularizer](../regularizers.md)).
 activity regularizer: Regularizer function applied to
 the output of the layer (its "activation").
 (see [regularizer](../regularizers.md)).
```

```
embeddings constraint: Constraint function applied to
 the `embeddings` matrix
 (see [constraints](../constraints.md)).
 mask zero: Whether or not the input value 0 is a special "padding"
 value that should be masked out.
 This is useful when using [recurrent layers](recurrent.md)
 which may take variable length input.
 If this is `True` then all subsequent layers
 in the model need to support masking or an exception will be raised.
 If mask zero is set to True, as a consequence, index 0 cannot be
 used in the vocabulary (input dim should equal size of
 vocabulary + 1).
 input length: Length of input sequences, when it is constant.
 This argument is required if you are going to connect
 `Flatten` then `Dense` layers upstream
 (without it, the shape of the dense outputs cannot be computed).
Input shape
 2D tensor with shape: `(batch size, sequence length)`.
Output shape
 3D tensor with shape: `(batch size, sequence length, output dim)`.
References
 - [A Theoretically Grounded Application of Dropout in
 Recurrent Neural Networks (http://arxiv.org/abs/1512.05287)
 /usr/local/lib/python3.6/dist-packages/keras/layers/embeddings.py
File:
Type:
 type
```

Subclasses:

```
In [49]:
```

```
def DeepMF(n users, n items, n factors, min rating, max rating):
 # Item Layer
 item input = Input(shape=[1], name='Item')
 item embedding = Embedding(n items, n factors, embeddings regularizer=12(1e-6),
 embeddings initializer="glorot normal",
 name='ItemEmbedding')(item input)
 item vec = Flatten(name='FlattenItemE')(item embedding)
 # Item Bias
 item bias = Embedding(n items, 1, embeddings regularizer=12(1e-6),
 embeddings initializer="glorot normal",
 name='ItemBias')(item input)
 item bias vec = Flatten(name='FlattenItemBiasE')(item bias)
 # User Layer
 user input = Input(shape=[1], name='User')
 user embedding = Embedding(n users, n factors, embeddings regularizer=12(1e-6),
 embeddings initializer="glorot normal",
 name='UserEmbedding')(user input)
 user vec = Flatten(name='FlattenUserE')(user embedding)
 # User Bias
 user bias = Embedding(n users, 1, embeddings regularizer=12(1e-6),
 embeddings initializer="glorot normal",
 name='UserBias')(user input)
 user bias vec = Flatten(name='FlattenUserBiasE')(user bias)
 # Concatenation of Item and User & then Add Bias
 Concat = Concatenate(name="Concat")([item vec, user vec])
 ConcatDrop = Dropout(0.5)(Concat)
 # Use the Dense layer for non-linear interaction learning
 Dense 1 = Dense(32, name="Dense1", activation="relu")(ConcatDrop)
 Dense 1 Drop = Dropout(0.5)(Dense 1)
 Dense 2 = Dense(1, name="Dense2")(Dense 1 Drop)
 # Add the Bias
 AddBias = Add(name="AddBias")([Dense 2, item bias vec, user bias vec])
 # Scaling for each user
 y = Activation('sigmoid')(AddBias)
 rating output = Lambda(lambda x: x * (max rating - min rating) + min rating)(y)
 # Model Creation
```

```
model = Model([user_input, item_input], rating_output, name="DeepFM")

Compile Model
model.compile(loss='mean_squared_error', optimizer=Adam(lr=0.001))

return model
```

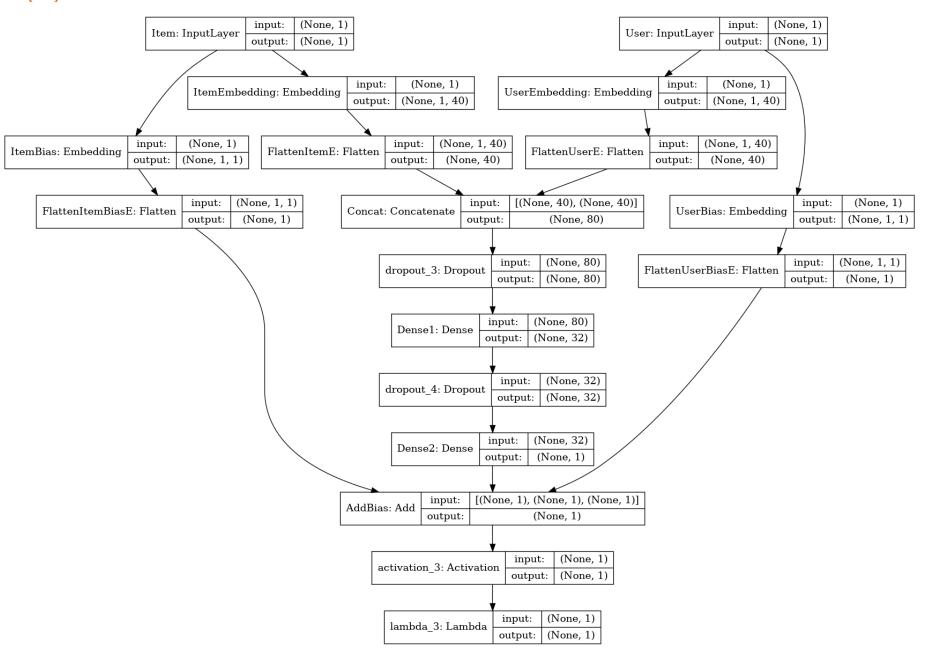
#### In [50]:

```
n_factors = 40
model = DeepMF(n_users, n_items, n_factors, min_rating, max_rating)
```

#### In [51]:

```
from keras.utils import plot_model
plot_model(model, show_layer_names=True, show_shapes=True)
```

Out[51]:



#### In [52]:

model.summary()

Model: "DeepFM"

Layer (type)	Output	Shape	Param #	Connected to
Item (InputLayer)	(None,	1)	0	
User (InputLayer)	(None,	1)	0	
ItemEmbedding (Embedding)	(None,	1, 40)	67280	Item[0][0]
UserEmbedding (Embedding)	(None,	1, 40)	37720	User[0][0]
FlattenItemE (Flatten)	(None,	40)	0	<pre>ItemEmbedding[0][0]</pre>
FlattenUserE (Flatten)	(None,	40)	0	UserEmbedding[0][0]
Concat (Concatenate)	(None,	80)	0	FlattenItemE[0][0] FlattenUserE[0][0]
dropout_3 (Dropout)	(None,	80)	0	Concat[0][0]
Densel (Dense)	(None,	32)	2592	dropout_3[0][0]
dropout_4 (Dropout)	(None,	32)	0	Dense1[0][0]
ItemBias (Embedding)	(None,	1, 1)	1682	Item[0][0]
UserBias (Embedding)	(None,	1, 1)	943	User[0][0]
Dense2 (Dense)	(None,	1)	33	dropout_4[0][0]
FlattenItemBiasE (Flatten)	(None,	1)	0	ItemBias[0][0]
FlattenUserBiasE (Flatten)	(None,	1)	0	UserBias[0][0]
AddBias (Add)	(None,	1)	0	Dense2[0][0] FlattenItemBiasE[0][0] FlattenUserBiasE[0][0]
activation_3 (Activation)	(None,	1)	0	AddBias[0][0]
lambda_3 (Lambda)	(None,	1)	0	activation_3[0][0]

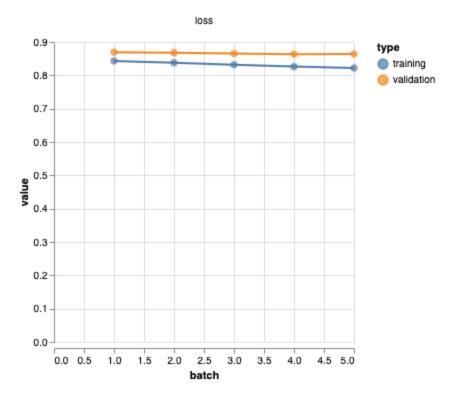
Total params: 110,250 Trainable params: 110,250 Non-trainable params: 0

#### In [61]:

## In [62]:

from recoflow.vis import MetricsVis
MetricsVis(output.history)





```
In [63]:
```

```
from recoflow.recommend import GetPredictions

predictions = GetPredictions(model, interaction)
predictions.head()
```

#### Out[63]:

	USER	ITEM	RATING
0	195	241	4.077020
1	195	301	4.219788
2	195	376	2.237257
3	195	50	3.569890
4	195	345	3.739755

## In [64]:

GetPredictions?

```
Signature: GetPredictions(model, data)
Docstring:
Get predictions for all user-item combinations

Params:
 data (pandas.DataFrame): DataFrame of entire rating data
 model (Keras.model): Trained keras model

Returns:
 pd.DataFrame: DataFrame of rating predictions for each user and each item

File: /usr/local/lib/python3.6/dist-packages/recoflow/recommend.py
Type: function

In [65]:
```

from recoflow.metrics import RatingMetrics

```
In [68]:
RatingMetrics(test, predictions)
Out[68]:

MSE RMSE MAE
```

## **Recommendations**

0 0.8619 0.9284 0.7355

```
In [75]:
```

```
from recoflow.recommend import ItemEmbedding, UserEmbedding
item_embedding = ItemEmbedding(model, "ItemEmbedding")
user_embedding = UserEmbedding(model, "UserEmbedding")
```

#### **Getting Recommendation**

- Given an Item, what are the similar Items
- Given a User, what items should I recommend

```
In [76]:
```

```
In [80]:
```

```
from recoflow.recommend import GetSimilar, ShowSimilarItems
similar_items = GetSimilar(item_embedding, k=5, metric="cosine")
similar_items
Out[80]:
```

```
In [85]:
```

ShowSimilarItems(0, similar\_items, item\_encoder, items, image\_path='/tf/data/posters/')





New Jersey Drive (1995) Bitter Sugar (Azucar Amargo) (1996) Wings of Courage (1995)









In [91]:

recoflow.\_\_version\_\_

Out[91]:

'0.0.6'

In [93]:

from recoflow.metrics import RatingMetrics

## For a user I want to find relevant

In [102]:

## For a user I want to find relevant
from recoflow.recommend import GetRankingTopK
GetRankingTopK
ranking\_topk = GetRankingTopK

#### Ranking Top K

- Get all the predictions for all users
- Remove all the items the user has already seen (train)
- Sort the remaining data by predicted ratings
- Cut off at K

```
In [107]:
```

```
ranking_topk = GetRankingTopK(model, interaction, train, k=5)
```

#### In [108]:

```
ranking_topk.head()
```

#### Out[108]:

	USER	ITEM	RATING	rank
0	0	1448	4.681684	1
1	0	407	4.654933	2
2	0	482	4.587464	3
3	0	602	4.564211	4
4	0	317	4.539346	5

```
In [109]:
```

```
n_users * 5
```

## Out[109]:

4715

#### In [111]:

```
ranking_topk.shape
```

```
Out[111]:
```

(4715, 4)

# **Model - Neural Collaborative Filtering**

In [116]:

from recoflow.models import NeuralCollaborativeFiltering

## In [118]:

NeuralCollaborativeFiltering??

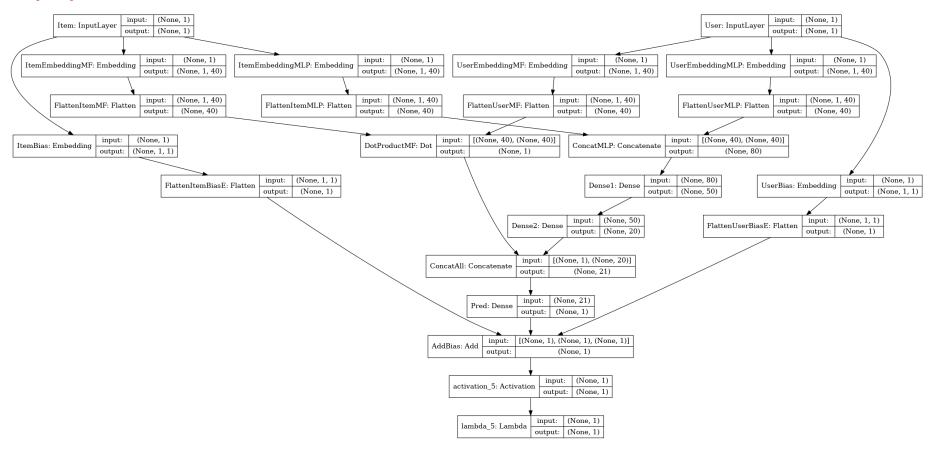
```
Signature:
NeuralCollaborativeFiltering(
 n users,
 n items,
 n factors,
 min rating,
 max rating,
Docstring: <no docstring>
Source:
def NeuralCollaborativeFiltering(n users, n items, n factors, min rating, max rating):
 # Item Layer
 item input = Input(shape=[1], name='Item')
 # Item Embedding MF
 item embedding mf = Embedding(n items, n factors, embeddings regularizer=12(1e-6),
 embeddings initializer='he normal',
 name='ItemEmbeddingMF')(item input)
 item vec mf = Flatten(name='FlattenItemMF')(item embedding mf)
 # Item embedding MLP
 item embedding mlp = Embedding(n items, n factors, embeddings regularizer=12(1e-6),
 embeddings initializer='he normal',
 name='ItemEmbeddingMLP')(item input)
 item vec mlp = Flatten(name='FlattenItemMLP')(item embedding mlp)
 # User Layer
 user input = Input(shape=[1], name='User')
 # User Embedding MF
 user embedding mf = Embedding(n users, n factors, embeddings regularizer=12(1e-6),
 embeddings initializer='he normal',
 name='UserEmbeddingMF')(user input)
 user vec mf = Flatten(name='FlattenUserMF')(user embedding mf)
 # User Embedding MF
 user embedding mlp = Embedding(n users, n factors, embeddings regularizer=12(1e-6),
 embeddings initializer='he normal',
 name='UserEmbeddingMLP')(user input)
 user vec mlp = Flatten(name='FlattenUserMLP')(user embedding mlp)
 # Multiply MF paths
 DotProductMF = Dot(axes=1, name='DotProductMF')([item_vec_mf, user_vec_mf])
```

```
Concat MLP paths
 ConcatMLP = Concatenate(name='ConcatMLP')([item vec mlp, user vec mlp])
 # Use Dense to learn non-linear dense representation
 Dense 1 = Dense(50, name="Dense1")(ConcatMLP)
 Dense 2 = Dense(20, name="Dense2")(Dense 1)
 # Concatenate MF and MLP paths
 Concat = Concatenate(name="ConcatAll")([DotProductMF, Dense 2])
 # Use Dense to learn non-linear dense representation
 Pred = Dense(1, name="Pred")(Concat)
 # Item Bias
 item bias = Embedding(n items, 1, embeddings regularizer=12(1e-5), name='ItemBias')(item input)
 item bias vec = Flatten(name='FlattenItemBiasE')(item bias)
 # User Bias
 user bias = Embedding(n users, 1, embeddings regularizer=12(1e-5), name='UserBias')(user input)
 user bias vec = Flatten(name='FlattenUserBiasE')(user bias)
 # Pred with bias added
 PredAddBias = Add(name="AddBias")([Pred, item bias vec, user bias vec])
 # Scaling for each user
 y = Activation('sigmoid')(PredAddBias)
 rating output = Lambda(lambda x: x * (max rating - min rating) + min rating)(y)
 # Model Creation
 model = Model([user input, item input], rating output, name="NeuralCF")
 # Compile Model
 model.compile(loss='mean squared error', optimizer="adam")
 return model
File:
 /usr/local/lib/python3.6/dist-packages/recoflow/models.py
Type:
 function
```

#### In [120]:

from recoflow.models import NeuralCollaborativeFiltering
ncf = NeuralCollaborativeFiltering(n\_users, n\_items, n\_factors, min\_rating, max\_rating)
plot\_model(ncf, show\_layer\_names=True, show\_shapes=True)

#### Out[120]:



## In [121]:

ncf.summary()

Model: "NeuralCF"

Layer (type) 	Output Shape	Param # 	Connected to
Item (InputLayer)	(None, 1)	0	
User (InputLayer)	(None, 1)	0	
ItemEmbeddingMLP (Embedding)	(None, 1, 40)	67280	Item[0][0]
JserEmbeddingMLP (Embedding)	(None, 1, 40)	37720	User[0][0]
FlattenItemMLP (Flatten)	(None, 40)	0	<pre>ItemEmbeddingMLP[0][0]</pre>
FlattenUserMLP (Flatten)	(None, 40)	0	UserEmbeddingMLP[0][0]
ItemEmbeddingMF (Embedding)	(None, 1, 40)	67280	Item[0][0]
UserEmbeddingMF (Embedding)	(None, 1, 40)	37720	User[0][0]
ConcatMLP (Concatenate)	(None, 80)	0	FlattenItemMLP[0][0] FlattenUserMLP[0][0]
FlattenItemMF (Flatten)	(None, 40)	0	<pre>ItemEmbeddingMF[0][0]</pre>
FlattenUserMF (Flatten)	(None, 40)	0	UserEmbeddingMF[0][0]
Densel (Dense)	(None, 50)	4050	ConcatMLP[0][0]
DotProductMF (Dot)	(None, 1)	0	<pre>FlattenItemMF[0][0] FlattenUserMF[0][0]</pre>
Dense2 (Dense)	(None, 20)	1020	Dense1[0][0]
ConcatAll (Concatenate)	(None, 21)	0	DotProductMF[0][0] Dense2[0][0]
ItemBias (Embedding)	(None, 1, 1)	1682	Item[0][0]
JserBias (Embedding)	(None, 1, 1)	943	User[0][0]
Pred (Dense)	(None, 1)	22	ConcatAll[0][0]
FlattenItemBiasE (Flatten)	(None, 1)	0	ItemBias[0][0]
FlattenUserBiasE (Flatten)	(None, 1)	0	UserBias[0][0]

			<pre>FlattenItemBiasE[0][0] FlattenUserBiasE[0][0]</pre>	
activation_5 (Activation)	(None, 1)	0	AddBias[0][0]	_
lambda_5 (Lambda)	(None, 1)	0	activation_5[0][0]	_

0

Pred[0][0]

(None, 1)

Total params: 217,717
Trainable params: 217,717
Non-trainable params: 0

AddBias (Add)