

BisIE

这个实验包含了六个对比实验，和两个数据集的预处理和训练。

六个对比实验分别为：

- **BisIE**
- **atrank**
- **bpr**
- **rnn_att**
- **BisIE_mask**

对比实验的运行分为两个步骤：

1.先将原始数据集转换为pandas数据帧，并重新映射分类ID，并产生训练数据集和测试数据集

```
1 cd atrank; #进入到其中一个对比实验中
2 python 1_convert_pd.py ;
3 python 2_remap_id.py;
4 python build_dataset.py; #产生训练数据集和测试数据集
```

2.使用默认的参数开始训练和评估

```
1 python train.py #利用atrank中的auc评价指标
2 python train_auc.py #正常的auc进行评价
```

BisIE_mask为我们本实验的重点：

- model.py（这个模型中的embedding只考虑了用户的历史行为与商品类别和时间之间的关系）
- model_new.py（这个模型的embedding同时还考虑了历史行为的类别，例如每个行为可以分为转化或者非转化行为）
- train_auc.py（针对model的训练）
- train_newmodel.py（针对model_new的训练）

本次实验我们还分别用了两个不同的数据集进行了对比，在上述我们所说的数据集的预处理，都是针对亚马逊的商品购买的数据集。在对BisIE_mask的实验中，我们还利用了天池中的数据，下面就是对天池数据集的预处理：

1.先将原始数据集转化为pandas数据帧，并重新映射分类ID

```
1 cd tianchi;  
2 python 1_ConversionFile.py;  
3 python 2_Convert.py;  
4 python 3_Remap.py;
```

2.构建适合当前方法的数据集

```
1 python build_dataset.py #生成train_set和test_set
```