



## The Experiment Report of Machine Learning

---

**SCHOOL: SCHOOL OF SOFTWARE ENGINEERING**

**SUBJECT: SOFTWARE ENGINEERING**

Author:  
Ke Ding/Qiuling Chen/Rong Li

Supervisor:  
Qingyao Wu

Student ID: 201530611449  
201530612026  
201530611210

Grade:  
Undergraduate

December 28, 2017

# 基于矩阵分解的推荐系统

## Abstract—

使用 SGD 实现的基于小数据集上的矩阵分解的推荐系统。

## I. INTRODUCTION

- 1.探索推荐系统的搭建过程
- 2.理解矩阵分解的原理
- 3.熟练对梯度下降的运用
- 4.在简单小规模数据集上实现推荐系统，培养工程能力

## II. METHODS AND THEORY

实验环境:

python3, 至少包含下列 python 包: sklearn, numpy, jupyter, matplotlib。

建议直接安装 anaconda3, 其已经内置了以上 python 包。

实验步骤:

本次实验代码及画图均在 jupyter 上完成。

优化方法:

利用随机梯度下降方法优化

读取数据, 并划分数据集 (或直接用 u1.base/u1.test 到 u5.base/u5.test)。根据原始数据填充原始评分矩阵, 对于空值可以填充为零。

初始化用户因子矩阵和物品 (电影) 因子矩阵, 其中为潜在特征数。

确定损失函数和确定超参数学习率和惩罚系数。

利用随机梯度下降法分解稀疏的用户评分矩阵, 得到用户因子矩阵和物品 (电影) 因子矩阵:

4.1 随机选择用户评分矩阵中的一个 sample ;

4.2 计算该 sample 的损失函数值对用户因子矩阵某一行 (列) 和物品因子矩阵某一行 (列) 的梯度;

4.3 梯度下降更新一行 (列) 与一行 (列);

4.4 计算在验证集上的, 可与上一次迭代的比较判断是否收敛。

重复步骤 4. 若干次, 得到满意的用户因子矩阵和物品因子矩阵, 画出随迭代次数变化的曲线图。

将用户因子矩阵与物品因子矩阵的转置相乘即可得到最终的评分预测矩阵。

## III. EXPERIMENT

数据集描述

采用 MovieLens-100k 数据集

u.data -- 由 943 个用户对 1682 个电影的 10000 条评分组成。每个用户至少评分 20 部电影。用户和电影从 1 号开始连续编号。数据是随机排序的。

例:

```
user id item id rating timestamp
196 242 3 881250949
186 302 3 891717742
22377 1 878887116
244 512 880606923
166 346 1 886397596
```

3. 数据集 u1.base / u1.test 到 u5.base / u5.test 都是将 u.data 数据集按照 80% / 20% 的比例分割的训练集和测试集。

4. 也可按照自己的评估方法划分训练集和验证集。

执行:

1. 初始化

(1) 矩阵全部采用随机初始化, 空值填充为 0

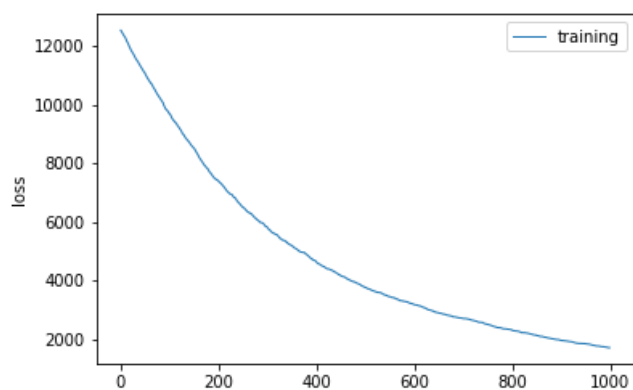
(2) SGD:

alpha = 0.05, k = 20, iteration = 1000, p = 0.02, q = 0.02

2. 过程

导入数据, 确认用户因子矩阵和物品矩阵, 确认损失函数为 RMSE 和惩罚系数, 利用随机梯度下降法分解稀疏的用户评分矩阵, 得到用户因子矩阵和物品 (电影) 因子矩阵, 若干次, 得到满意的用户因子矩阵和物品因子矩阵, 画出随迭代次数变化的曲线图。将用户因子矩阵与物品因子矩阵的转置相乘即可得到最终的评分预测矩阵。

3. 结果



#### IV. CONCLUSION

经过这次的实验，我们明白了推荐系统是如何搭建的，理解了矩阵分解的原理，并且对梯度下降的运用更加熟练了，算法并没有想象中的那么困难与复杂，初步了解了机器学习对实际工程解决实际问题的重要作用，同时，小组通过团队合作，完成实验的效率比一个人的时候效率要高很多，在现实中解决实际问题时，团队合作是必不可少的。