华南理工大学

South China University of Technology

# The Experiment Report of Machine Learning

SCHOOL: SCHOOL OF SOFTWARE ENGINEERING

SUBJECT: SOFTWARE ENGINEERING

Author:
潘楷乐 姚永铖 赵云潇

Supervisor:
Qingyao Wu

Student ID：201530612583
20153061
201530613757

Grade:
Undergraduate

December 9, 2017

# Recommender System Based on Matrix Decomposition

**Abstract—**

## I.INTRODUCTION

*Motivation*

1. Explore the construction of recommended system.
2. Understand the principle of matrix decomposition.
3. Be familiar to the use of gradient descent.
4. Construct a recommendation system under small-scale dataset, cultivate engineering ability.

*Reference*

Matrixactorization: A Simple Tutorial and Implementation in Python

1. Tan, W., Cao, L., & Fong, L. (2016, May). Faster and cheaper: Parallelizing large-scale matrix factorization on gpus. In Proceedings of the 25th ACM International Symposium on High-Performance Parallel and Distributed Computing (pp. 219-230). ACM.
2. Xie, X., Tan, W., Fong, L. L., & Liang, Y. (2017, June). CuMF_SGD: Parallelized Stochastic Gradient Descent for Matrix Factorization on GPUs. In Proceedings of the 26th International Symposium on High-Performance Parallel and Distributed Computing (pp. 79-92). ACM.

## II. METHODS AND THEORY

*Methods*

Using stochastic gradient descent method(SGD)

## III.EXPERIMENT

*Dataset*

1. Utilizing MovieLens-100k dataset.
2. u.data —Consisting 10,000 comments from 943 users out of 1682 movies. At least, each user comment 20 videos. Users and movies are numbered consecutively from number 1 respectively. The data is sorted randomly

| user id | item id | rating | timestamp |
|---|---|---|---|
| 196 | 242 | 3 | 881250949 |
| 186 | 302 | 3 | 891717742 |
| 22 | 377 | 1 | 878887116 |
| 244 | 51 | 2 | 880606923 |
| 166 | 346 | 1 | 886397596 |

3. u1.base / u1.test are train set and validation set respectively, seperated from dataset u.data with proportion of 80% and 20%. It also make sense to train set and validation set from u1.base / u1.test to u5.base / u5.test.
4. You can also construct train set and validation set according to your own evaluation method.

*Implementation*

alpha=0.02, beta=0.01,k=15

Experiment Step

The experiment code and drawing are both completed on jupyter.

*Using stochastic gradient descent method(SGD):*

1. Read the data set and divide it (or use u1.base / u1.test to u5.base / u5.test directly). Populate the original scoring matrix against the raw data, and fill 0 for null values.
2. Initialize the user factor matrix and the item (movie) factor matrix , where is the number of potential features.
3. Determine the loss function and hyperparameter learning rate and the penalty factor .
4. Use the stochastic gradient descent method to decompose the sparse user score matrix, get the user factor matrix and item (movie) factor matrix:

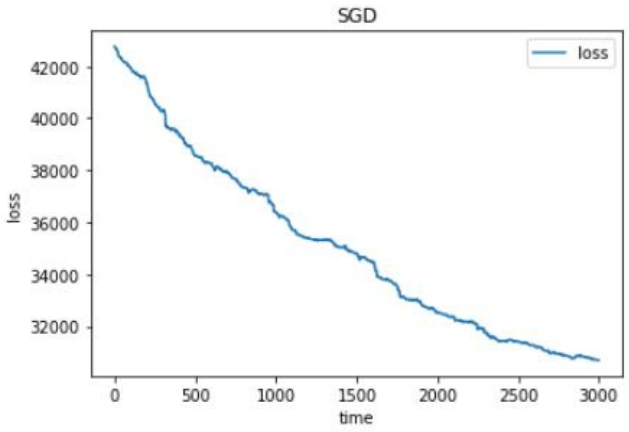
   4.1 Select a sample from scoring matrix randomly;

   4.2 Calculate this sample's loss gradient of specific row(column) of user factor matrix and item factor matrix;

   4.3 Use SGD to update the specific row(column) of and ;

   4.4 Calculate the on the validation set, comparing with the of the previous iteration to determine if it has converged.
5. Repeat step 4. several times, get a satisfactory user factor matrix and an item factor matrix , **Draw a curve with varying iterations**.
6. The final score prediction matrix is obtained by multiplying the user factor matrix and the transpose of the item factor matrix .

### Result




## IV. CONCLUSION

Through this experiment, we have a deeper understanding of machine learning, understand the principle of matrix decomposition in the process of exploring the recommendation system, and become more proficient in gradient descent algorithm, and cultivate our own teamwork ability. But there are also shortcomings, our team's practical ability is not strong enough, and the results are not very excellent. In the future of learning and life, we will continue to work hard in the field of machine learning to carry out more in-depth research.