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:
ZiJia Zhang and ZhiQuan Wen

Supervisor:
Qingyao Wu

Student ID:
201530613726 and 201530612996

Grade:
Undergraduate

December 9, 2017

Matrix-based Recommendation System

Abstract—The short abstract is intended to give the reader an overview of the experiment. It should be brief and to the point.

I. Introduction

THE motivation of this experiment:

1. Understand Matrix-based Recommendation System further
2. Get familiar with the basic method of recommendation system
3. Learn to use Matrix-based to solve the recommendation system question and combine the theory with the actual project
4. Experience the complete process

II. Methods and Theory

Algorithm 2 ALS Algorithm

1: **Require** rating matrix $\mathbf{R}$, feature matrices $\mathbf{P}$, $\mathbf{Q}$ and regularization parameter λ.
2: **Optimize** $\mathbf{P}$ while fixing $\mathbf{Q}$:
$$\mathbf{p}_u = (\mathbf{q}_i\mathbf{q}_i^\top + \lambda n_{\mathbf{p}_u} I)^{-1} \cdot \mathbf{Q}^\top \cdot \mathbf{R}_{u*}^\top.$$
3: **Optimize** $\mathbf{Q}$ while fixing $\mathbf{P}$:
$$\mathbf{q}_i = (\mathbf{p}_u\mathbf{p}_u^\top + \lambda n_{\mathbf{q}_i} I)^{-1} \cdot \mathbf{P}^\top \cdot \mathbf{R}_{*i}.$$
4: **Repeat** the above processes until **convergence**.

III. Experiments

A. Dataset

We use the dataset named “u1.base” as the train dataset and the dataset named “u1.test” as the test dataset.

1. MovieLens-100k data set

2. u.data - Consists of 10,000 users rated 16,000 movies by 943 users. Each user scored at least 20 movies. Users and movies numbered consecutively from number one. 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. Data sets u1.base / u1.test to u5.base / u5.test are u.data dataset in accordance with the proportion of 80% / 20% split training set and test set

B. The step of the experiments

1. Read the data and divide the data set (or use u1.base / u1.test to u5.base / u5.test directly). Fill the original scoring matrix with the raw data, zeroes for null values.

2. Initialize user factor matrices and item (movie) factor matrices, where K is the number of potential features.

3. Determine the loss function and determine the super-parametric learning rate and penalty coefficient.

Use stochastic gradient descent method to decompose the sparse user score matrix, get the user factor matrix and objects (movie)

Factor matrix:

4.1 randomly select a sample in the user rating matrix;

4.2 Calculation of the loss of the sample function value of the user factor matrix of a row (column) and an item factor matrix

Gradient of rows (columns);

4.3 Gradient descent Update one row (column) and one row (column);

4.4 Calculated on the validation set, compared with the previous iteration to determine whether the convergence.

5. Repeat step 4. Several times, get the satisfactory user factor matrix and item factor matrix, draw the variable with the number of iterations

The curve.

6. Multiply the user factor matrix by the transpose of the item factor matrix to get the final score prediction matrix.

IV. Conclusion