

Scala: 2.11 Spark: 2.2.1

## Task 1

### Jaccard-based LSH

#### *Command line:*

```
$ bin/spark-submit --class JaccardLSH Yi_Wei_hw3.jar video_small_num.csv Yi_Wei_SimilarProducts_Jaccard.txt
```

```
bin/spark-submit --class JaccardLSH Yi_Wei_hw3.jar video_small_num.csv  
Yi_Wei_SimilarProducts_Jaccard.txt
```

The first parameter is the input path of ratings file

The second parameter is the output path of the result

**Precision:** 1.0

**Recall:** 0.9334389857369255

**Time:** 63.776161138 sec

### Cosine-based LSH

**Precision:** 0.9434225438061198

**Recall:** 0.7099652299416126

## Task 2

### Model-based CF:

#### *Command line:*

```
$ bin/spark-submit --class ModelBasedCF Yi_Wei_hw3.jar video_small_num.csv video_small_testing_num.csv Yi_Wei_ModelBasedCF.txt
```

```
bin/spark-submit --class ModelBasedCF Yi_Wei_hw3.jar video_small_num.csv  
video_small_testing_num.csv Yi_Wei_ModelBasedCF.txt
```

The first parameter is the input path of ratings file

The second parameter is the input path of testing file

The third parameter is the output path of the result

**Result table:**

>=0 and <1	4628
>=1 and <2	2175
>=2 and <3	579
>=3 and <4	317
>=4	1
RMSE	1.285640853360315
Time (sec)	12.30224146

## User-based CF

**Command line:**

```
$ bin/spark-submit --class UserBasedCF Yi_Wei_hw3.jar video_small_num.csv video_small_testing_num.csv Yi_Wei_UserBasedCF.txt
```

```
bin/spark-submit --class UserBasedCF Yi_Wei_hw3.jar video_small_num.csv  
video_small_testing_num.csv Yi_Wei_UserBasedCF.txt
```

The first parameter is the input path of ratings file

The second parameter is the input path of testing file

The third parameter is the output path of the result

**Result table:**

$\geq 0$ and $< 1$	4038
$\geq 1$ and $< 2$	2306
$\geq 2$ and $< 3$	865
$\geq 3$ and $< 4$	423
$\geq 4$	68
RMSE	1.4476578316927355
Time (sec)	11.065969438