Recommender System with Mapreduce - 2

赵敏 老师

扫描二维码关注微信/微博
获取最新IT面试情报及权威解答

微信: ninechapter
知乎专栏: http://zhuanlan.zhihu.com/jiuzhang
微博: http://www.weibo.com/ninechapter
官网: www.jiuzhang.com

Recap

- What is recommender System
- What is Item CF
- How to implement Item CF with MR
  - Build co-occurrence matrix

Divide Data by User ID

User_id	Movie_id: Rating	Movie_id: Rating
1	1:10	2: 8
2	1: 8	2: 5.5
3	2: 4.5	8: 9
4	2: 6.5	8: 2.5

Build Co-Occurrence Matrix

Co-Occurrence Matrix

MovieA: MovieB	Relation
1: 1	2
1: 2	2
2: 1	2
2: 2	4
2: 8	2
8: 2	2
8: 8	2

Build Co-occurrence Matrix

九章算法

Input

User_id	Movie_id: Rating	Movie_id: Rating
1	1:10	2: 8
2	1: 8	2: 5.5
3	2: 4.5	8: 9
4	2: 6.5	8: 2.5

Output

MovieA: MovieB	Relation
1: 1	2
1: 2	2
2: 1	2
2: 2	4
2: 8	2
8: 2	2
8: 8	2

Build Co-occurrence Matrix: Mapper

User_id	Movie_id: Rating	Movie_id: Rating
1	1:10	2: 8
2	1: 8	2: 5.5
3	2: 4.5	8: 9
4	2: 6.5	8: 2.5

MovieA: MovieB	Relation
1:1	1
1:2	1
2:1	1
2:2	1
1:1	1
1:2	1
2:1	1
2:2	1

....

Build Co-occurrence Matrix: Reducer

MovieA: MovieB	Relation
1:1	1
1:2	1
2:1	1
2:2	1
1:1	1
1:2	1
2:1	1
2:2	1

....

Merge

MovieA: MovieB	Relation
1: 1	2
1: 2	2
2: 1	2
2: 2	4
2: 8	2
8: 2	2
8: 8	2

Implement recommender system with MapReduce

We have co-occurrence matrix now, what do we miss?

Implement recommender system with MapReduce

- Normalize co-occurrence matrix
- Build rating matrix
- Multiply co-occurrence matrix and rating matrix
- Generate recommender list

Normalize Co-occurrence Matrix

	M1	M2	M3	M4	M5
M1	2	2	1	1	0
M2	2	4	2	2	1
M3	1	2	2	0	1
M4	1	2	0	2	0
M5	0	1	1	0	1

归一化处理

	M1	M2	M3	M4	M5
M1	2/6	2/6	1/6	1/6	0
M2	2/11	4/11	2/11	2/11	1/11
M3	1/6	2/6	2/6	0	1/6
M4	1/5	2/5	0	2/5	0
M5	0	1/3	1/3	0	1/3

Normalize Cooccurrence Matrix

MovieA: MovieB	Relation
1: 1	2
1: 2	2
2: 1	2
2: 2	4
2: 8	2
8: 2	2
8: 8	2

MovieB	MovieA=relation
1	1=2/4
2	1=2/4
1	2=2/8
2	2=4/8
8	2=2/8
2	8=2/4
8	8=2/4

	M1	M2	M3	M4	M5
M1	2	2	1	1	0
M2	2	4	2	2	1
M3	1	2	2	0	1
M4	1	2	0	2	0
M5	0	1	1	0	1

Rating matrix

User Rating Matrix

```
1,10001,5.0
1,10002,3.0
1,10003,2.5
2,10001,2.0
2,10002,2.5
2,10003,5.0
2,10004,2.0
3,10001,2.0
3,10004,4.0
3,10005,4.5
3,10007,5.0
4,10001,5.0
4,10003,3.0
4,10004,4.5
4,10006,4.0
5,10001,4.0
5,10002,3.0
5,10003,2.0
5,10004,4.0
5,10005,3.5
5,10006,4.0
```

Multiply co-occurrence matrix and rating matrix

Co-Occurrence Matrix

MovieB	MovieA=relation
1	1=2/4
2	1=2/4
1	2=2/8
2	2=4/8
8	2=2/8
2	8=2/4
8	8=2/4

User Rating Matrix

User_id	Movie_id: Rating
1	1:10
1	2: 8
2	1: 8
2	3: 4.5
4	2: 5.5
4	4: 6.5

Multiply co-occurrence matrix and rating matrix

Multiply two matrix?

Multiply co-occurrence matrix and rating matrix

No! Slow! OOM

Multiply co-occurrence matrix and rating matrix

九章算法

Co Matrix

	M1	M2	M3	M4	M5
M1	2/6	2/6	1/6	1/6	0
M2	2/11	4/11	2/11	2/11	1/11
M3	1/6	2/6	2/6	0	1/6
M4	1/5	2/5	0	2/5	0
M5	0	1/3	1/3	0	1/3

Rating Matrix

Movie	User B rating
M1	3
M2	7
M3	8
M4	0
M5	0

Result Matrix

Movie	User B rating
M1	2/6*3 + 2/6*7 +...
M2	2/11*3 + 4/11*7 + ...
M3	1/6*3 + 2/6*7 +...
M4	1/5*3 + 2/5*7 +...
M5	0*3 + 1/3*7 +...

UserB: Movie1 = Co[M1][M1]* Rating[M1] + Co[M1][M2]* Rating[M2] + Co[M1][M3]* Rating[M3] + Co[M1][M4]* Rating[M4] + Co[M1][M5]* Rating[M5]

Multiply co-occurrence matrix and rating matrix

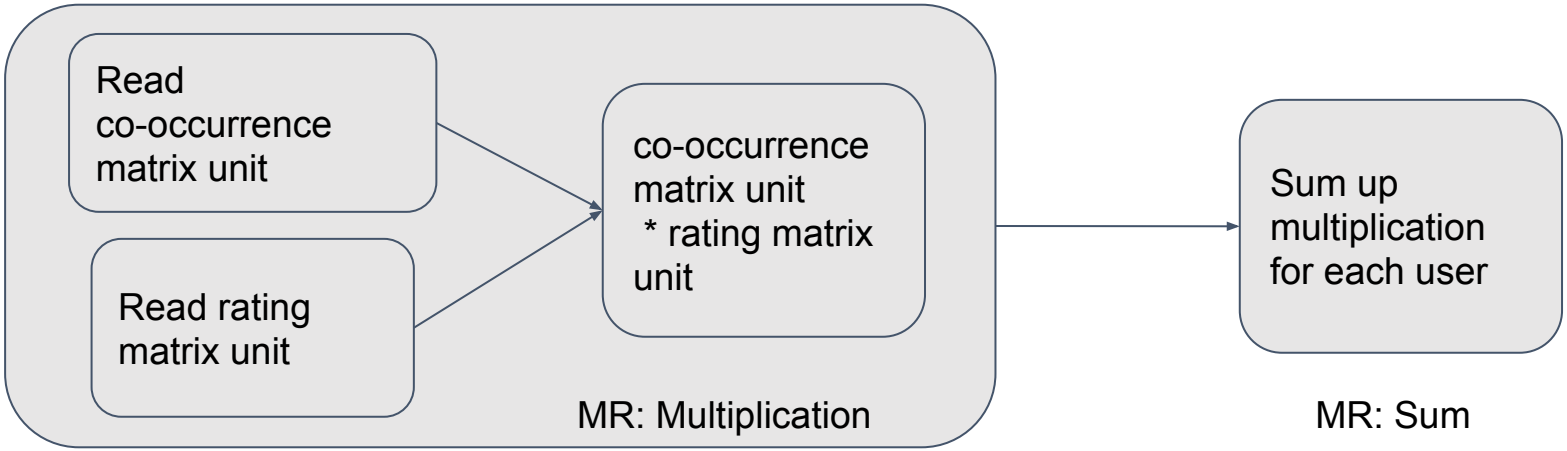
Co-Occurrence Matrix

MovieB	MovieA=relation
1	1=2/4
2	1=2/4
1	2=2/8
2	2=4/8
8	2=2/8
2	8=2/4
8	8=2/4

User Rating Matrix

User_id	Movie_id: Rating
1	1:10
1	2: 8
2	1: 8
2	3: 4.5
4	2: 5.5
4	4: 6.5

Mapper



Mapper1

MovieB	MovieA=relation
1	1=2/4
2	1=2/4
1	2=2/8
2	2=4/8
8	2=2/8
2	8=2/4
8	8=2/4

MovieB	MovieA=relation
1	1=2/4
2	1=2/4
1	2=2/8
2	2=4/8
8	2=2/8
2	8=2/4
8	8=2/4

Mapper2

User_id	Movie_id: Rating
1	1:10
1	2: 8
2	1: 8
2	3: 4.5
4	2: 5.5
4	4: 6.5

Movie_id	User_id:Rating
1	1:10
2	1:8
1	2:8
3	2:4.5
2	4:5.5
4	4:6.5

Reducer

Key = movieB_id
Value = <movieA=relation, user:rating>

MovieB	MovieA=relation
1	1=2/4
2	1=2/4
1	2=2/8
2	2=4/8
8	2=2/8
2	8=2/4
8	8=2/4

Movie_id	User_id:Rating
1	1:10
2	1:8
1	2:8
3	2:4.5
2	4:5.5
4	4:6.5

user:movie	rating*relation
1: 1	10*2/4
1: 1	10*2/8
2: 1	8*2/4
2: 1	8*2/8

Sum

user:movie	rating*relation
1: 1	10*2/4
1: 1	10*2/8
2: 1	8*2/4
2: 1	8*2/8

user:movie	rating*relation
1: 1	10*2/4 + 10*2/8
2: 1	8*2/4 + 8*2/8

Thanks~~