

# PMF模型介绍

## 模型基本假设

评分矩阵R可由用户特征矩阵U.T和电影特征矩阵V相乘得到;  
矩阵R, U, V都服从高斯分布

$$\hat{R} = U^T V,$$

# PMF模型介绍

在简单无约束PMF的基础上引入“用户为电影打过分的影响”矩阵W，W表征了“用户对某电影打过分”这件事本身对用户偏好的影响，它将作为用户特征矩阵的约束被引入目标函数

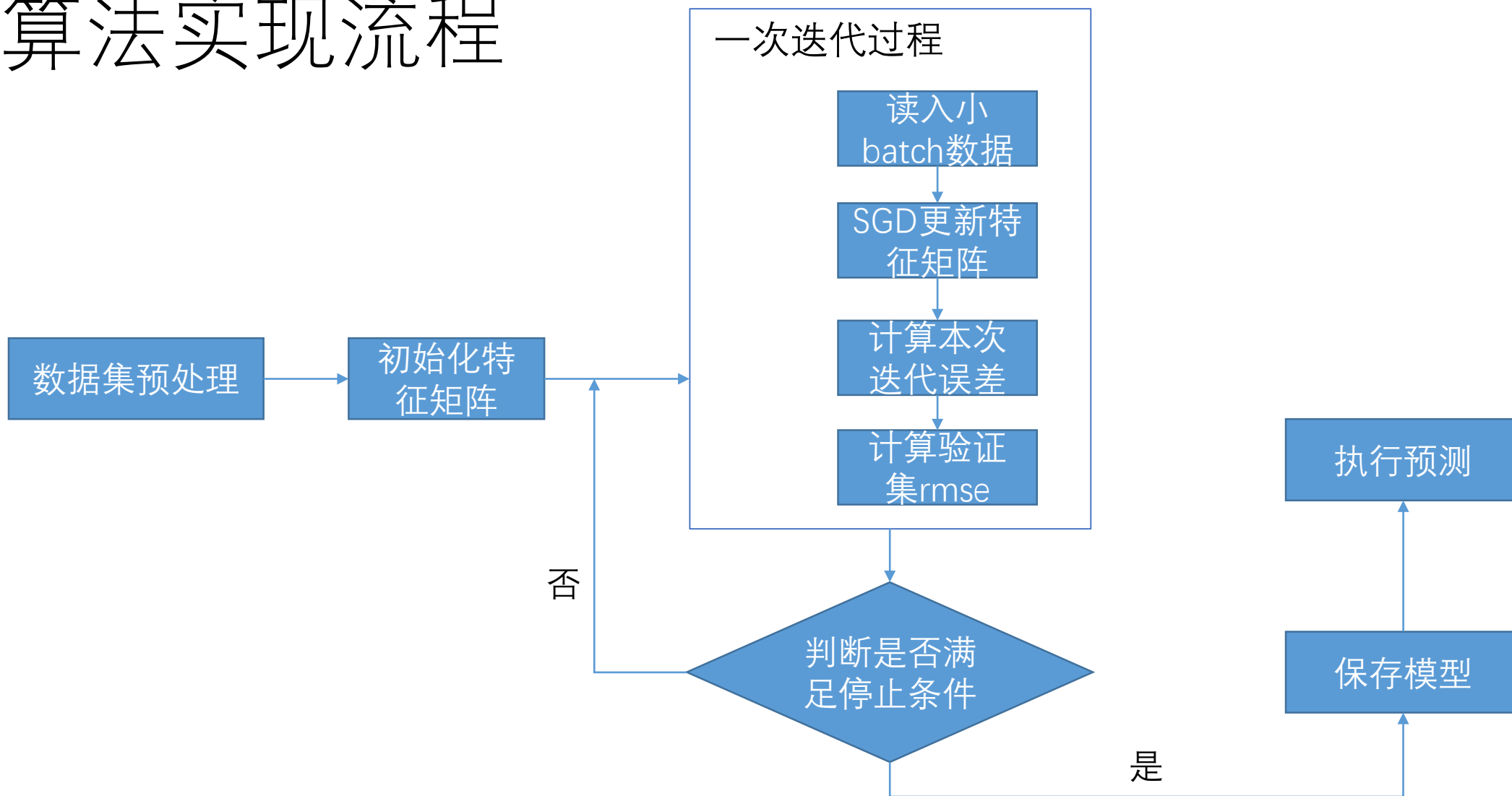
这样将简单PMF的优化目标转变为：

$$\begin{aligned} E = & \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M I_{ij} \left( R_{ij} - g \left( \left[ Y_i + \frac{\sum_{k=1}^M I_{ik} W_k}{\sum_{k=1}^M I_{ik}} \right]^T V_j \right) \right)^2 \\ & + \frac{\lambda_Y}{2} \sum_{i=1}^N \| Y_i \|^2_{Fro} + \frac{\lambda_V}{2} \sum_{j=1}^M \| V_j \|^2_{Fro} + \frac{\lambda_W}{2} \sum_{k=1}^M \| W_k \|^2_{Fro}, \end{aligned}$$

# 算法训练策略

- 采用随机梯度下降算法对目标函数做迭代优化
- 优化参数：
  - $U$ ——用户特征矩阵
  - $V$ ——电影特征矩阵
  - $W$ ——“用户对电影评过分”约束矩阵

# 算法实现流程



# 代码主要函数说明

```
def train(self, num_user, num_mv, user_list, mv_list, data_train, data_test, max_iter=100):| ...
```

# 训练函数，使用SGD算法优化目标函数

```
def save_model(self, path, U, V): ... # 模型保存函数，将训练得到的参数保存到.csv文件
```

```
def load_model(self, path): ... # 模型加载函数，读取.csv文件中保存的模型，得到预测需要的矩阵
```

```
def predict(self, path, U, V): ... # 预测函数，预测R矩阵缺失值
```

```
def rmse(self, U, V, data_test): ... # RMSE计算函数，训练过程中计算每轮迭代后的rmse值
```

```
def logistic(self, prediction): ... # logistic函数，将预测值映射至 (0,1) 区间，防止预测值超出评分区间
```

```
def rating_map(self, rating): ... # 评分映射函数，将观测值映射至 (0,1) 区间
```

```
def data_read(path, ratio): ... # 数据加载函数，读取原始数据集，生成训练所需要的格式
```

# 代码执行结果

用户特征矩阵U

	1	2	3	4	5	6	7
0	0.367059	-1.14706	-0.04231	0.495125	-1.36548	-1.06093	0.859416
1	0.323483	-0.21858	-1.601	0.778174	0.063412	-1.02019	0.042137
2	1.375823	0.595502	-2.07645	1.182702	1.069751	1.922235	-0.54458
3	1.089877	0.516555	0.406064	-2.66736	0.835411	-0.35408	-0.72643
4	-0.36778	-0.57233	1.393175	-0.86143	-0.61557	-0.12409	-0.5406
5	1.259483	-0.15259	-1.05633	1.388795	-1.61965	0.055637	-0.76412
6	0.326548	0.275649	-0.7364	-0.88604	-1.04662	-1.50819	-0.17574
7	1.213308	-1.51687	0.073459	0.560507	0.362622	-1.37842	0.858498
8	-1.49274	-0.08369	0.017515	0.044467	-0.79287	0.722088	0.729846
9	-2.19299	-0.51586	0.00039	0.641675	-0.01846	-1.61702	1.175727
10	1.071761	0.197097	-1.29619	-0.38117	1.072602	-1.10014	-0.74672
11	0.607822	0.472779	1.952482	-1.01989	0.408561	1.015955	-1.70688
12	-0.2569	-2.66291	-0.61797	-0.08058	-0.3776	2.150545	1.185635
13	0.758198	0.005116	1.457935	0.409694	1.786899	0.302759	1.221719
14	-0.69713	0.237042	-1.29493	0.509756	1.380089	0.307097	0.150223
15	2.078159	1.241961	2.048496	0.182477	-1.77247	1.17496	-0.58622
16	-0.46726	0.794891	-0.54904	-0.99918	0.961927	1.980071	-2.18359
17	-0.09487	1.611853	1.490536	-0.2383	-0.10951	-0.78861	-0.80538
18	-0.36092	-2.3502	0.715602	0.766887	-1.2237	-0.14791	1.819262
19	-2.1166	-0.82678	2.135585	-0.41524	0.108164	-1.94673	-0.23803
20	-1.53327	-0.19171	0.617136	-0.30213	-0.14379	-0.46492	-1.84158
21	1.349152	-1.08796	-0.58242	-0.67863	1.654509	-0.08931	0.965264
22	0.454178	-0.04331	-0.67063	-0.93666	0.173075	0.739286	0.273747
23	0.450112	1.463629	-0.03508	-0.44569	0.947282	0.691892	-0.45956
24	-0.29507	0.019608	-1.00927	0.203073	2.979127	0.393658	-0.04424
25	-0.52701	-0.03383	-0.3602	0.325895	0.774252	-0.62769	-0.24554
26	-1.67436	-0.13402	0.081498	1.582749	-0.13264	0.672639	0.008742
27	-0.14908	-0.83189	0.783813	-0.7541	-0.64633	0.689285	0.995311
28	-1.22811	-1.13925	-0.93271	1.800344	-0.39585	-1.70162	-0.51103

电影特征矩阵U

	1	2	3	4	5	6
0	-0.11346	0.018833	0.042207	-0.06236	-0.03717	-0.0966
1	-0.15486	0.044705	0.099896	0.065307	-0.08665	-0.20652
2	-0.11932	-0.04221	0.164972	0.097374	-0.07403	0.04354
3	0.086005	0.003736	-0.09903	0.090564	-0.10215	0.289623
4	-0.12934	-0.01044	0.187111	-0.0697	-0.1551	-0.03304
5	0.201601	0.042741	-0.11534	-0.03888	-0.05533	-0.05833
6	-0.02167	-0.01721	-0.33513	0.026821	0.000951	0.218727
7	0.163642	-0.1074	-0.12744	0.089323	0.130905	0.118607
8	-0.14073	-0.1245	-0.0169	0.087259	-0.0081	0.086231
9	0.016457	0.060574	-0.07382	0.005965	0.093978	0.070596
10	0.01519	-0.07423	-0.0626	-0.15559	0.14258	-0.03071
11	0.070751	0.02588	0.045753	0.017045	0.026079	0.176309
12	-0.06592	-0.08628	0.135268	0.022306	0.020203	-0.08972
13	-0.02263	0.004397	-0.09616	-0.03811	0.03187	0.176948
14	0.039863	-0.03503	-0.12458	0.158979	0.026899	0.029173
15	0.145417	0.110359	0.151143	0.217122	-0.05917	-0.06729
16	-0.09779	0.079472	0.176915	-0.089	0.160883	0.128463
17	0.103849	0.0694	-0.03139	0.132369	-0.06824	0.120425
18	-0.08582	-0.00946	0.221442	0.030419	-0.03951	-0.13408
19	0.216479	0.081284	-0.27437	0.036626	0.133586	0.298441
20	0.045118	-0.11168	-0.11424	-0.08003	-0.14707	-0.01489
21	0.491165	-0.03199	-0.16841	0.115451	0.070175	-0.04252
22	-0.01075	0.085921	-0.1255	-0.06408	0.062158	0.085851
23	0.08847	0.048225	-0.14938	-0.00487	0.038097	-0.25085
24	0.064303	0.039075	-0.1179	0.053293	0.132858	0.115944
25	-0.14597	-0.00852	0.06022	-0.00381	-0.10077	-0.05925
26	-0.30647	0.090109	-0.0738	0.172691	0.056735	0.198137
27	0.062115	0.089331	-0.11206	0.087055	-0.06546	0.223663
28	0.187043	-0.00168	-0.06834	0.005831	0.012694	0.010647

评分预测结果

	0	1	2	3	4	5	6	7	8
0	4.213708	2.964481	3.597474	3.191574	2.695635	2.141177	2.624841	2.878213	4.397493
1	2.80342	3.597615	2.475136	3.195842	2.591489	3.002325	2.971369	3.79274	4.1464
2	3.7001	3.581104	3.060168	3.372549	2.564788	4.219492	3.507958	2.183894	1.89626
3	2.266408	2.951779	3.426436	3.410476	2.977591	1.7797	2.734935	2.471904	2.797488
4	3.742095	2.830054	2.068986	3.032057	4.101086	3.899245	2.492302	3.52855	3.046446
5	1.895272	3.125467	4.606501	3.366452	2.945871	2.669239	3.309783	3.122394	3.457067
6	2.822703	2.75088	2.801812	3.471471	3.219791	2.44749	2.996222	2.368354	2.40541
7	2.352597	3.291624	3.335552	2.360858	3.107348	3.153497	3.862385	2.848957	1.904503
8	3.179377	3.109882	3.598432	2.945445	3.615803	2.117946	2.401052	3.817184	3.551718
9	4.374659	1.803987	1.91631	3.422815	2.530634	3.829743	1.588476	1.727932	2.952345
10	4.391168	2.679246	2.184473	3.354135	3.102283	2.68111	2.290019	3.016818	2.979379
11	3.243232	3.081911	2.864463	2.681383	2.700484	2.28477	2.102088	3.610615	2.652177
12	2.584755	3.066282	3.808817	3.569657	2.710365	3.126255	2.428396	3.571592	3.908292
13	2.29753	3.224114	2.737709	3.496353	3.420372	3.064934	3.141888	3.897432	3.025372
14	3.561872	2.817215	2.266445	2.216118	2.610981	1.79417	2.476854	3.205117	3.517527
15	2.269204	2.524777	3.221586	2.237285	2.929786	3.209192	2.071188	3.546332	3.590606
16	2.656821	2.972662	2.370038	3.811465	2.757455	4.626459	3.542395	1.48328	3.323545
17	3.934925	4.098897	1.593719	2.601235	3.988438	4.282278	3.007218	2.314744	2.984791
18	4.25255	3.202835	3.543001	1.804593	3.259956	1.993015	2.742153	2.262613	3.778517
19	3.523772	3.506007	2.661227	3.854123	2.893569	4.138042	3.271007	1.697498	2.764871
20	3.276734	2.668887	3.984663	3.815999	3.096158	2.289878	2.659443	3.094066	3.177387
21	3.187132	3.684173	1.623937	3.052133	3.431757	3.810217	3.4654	3.31795	3.758252
22	4.370109	2.998348	2.834991	1.851447	3.094369	1.431495	2.211066	3.398463	4.276795
23	4.21911	2.587607	1.512281	3.239335	3.313543	2.082982	2.253904	4.10639	3.756744
24	2.932677	2.907078	3.160548	2.714023	3.585083	4.033558	3.449083	3.56366	2.883965
25	1.447169	2.626621	2.651222	1.355601	2.00826	1.712393	1.48252	4.037718	2.556713
26	4.228397	2.983129	2.634397	3.125099	3.262261	4.134758	2.78389	3.4925	3.116273
27	3.497686	3.289455	1.772116	4.070138	3.162642	3.991454	3.303311	3.26023	2.805182
28	3.171808	3.327452	1.955609	2.876966	2.982388	3.41563	3.021801	2.217325	3.463924
29	2.585332	3.150545	3.49821	2.748589	2.580219	2.489384	3.683611	2.169945	2.699369