

基于局部代表性矩阵分解的冷启动推荐方法

石磊, 赵鑫, 沈一栋
计算机科学国家重点实验室
shilei@ios.ac.cn

Lei Shi, Wayne Xin Zhao and Yi-Dong Shen. Local Representative based Matrix Factorization for Cold Start Recommendation. ACM Transactions on Information Systems (TOIS) 36(2) 22:1-28, 2017

背景介绍

- 冷启动问题是推荐领域中的一类最具挑战性问题
- 缺乏用户偏好标注, 无法估计用户兴趣
- 基于问询的方法选择物品推送给用户并获得用户的偏好标注, 从而估计用户的兴趣

代表性矩阵分解

- 假设:** 用户对所有物品的偏好可以通过用户对少量代表性物品的偏好进行重构
- 步骤 1:** 选择代表性物品并获取用户的反馈
- 步骤 2:** 根据用户对代表性物品的反馈预测用户对所有物品的偏好
- 现有方法仅选择**全局代表性物品集合**对用户进行刻画, 然而全局代表性物品需要对整个物品集合有很高的覆盖度, 导致其**无法刻画细粒度用户偏好信息**

局部代表性矩阵分解

基本思想: 我们采用两轮问询机制: 第一轮问询进行**动态用户分组**; 第二轮问询采用局部数据重构选择**局部代表性物品**进行用户兴趣刻画。

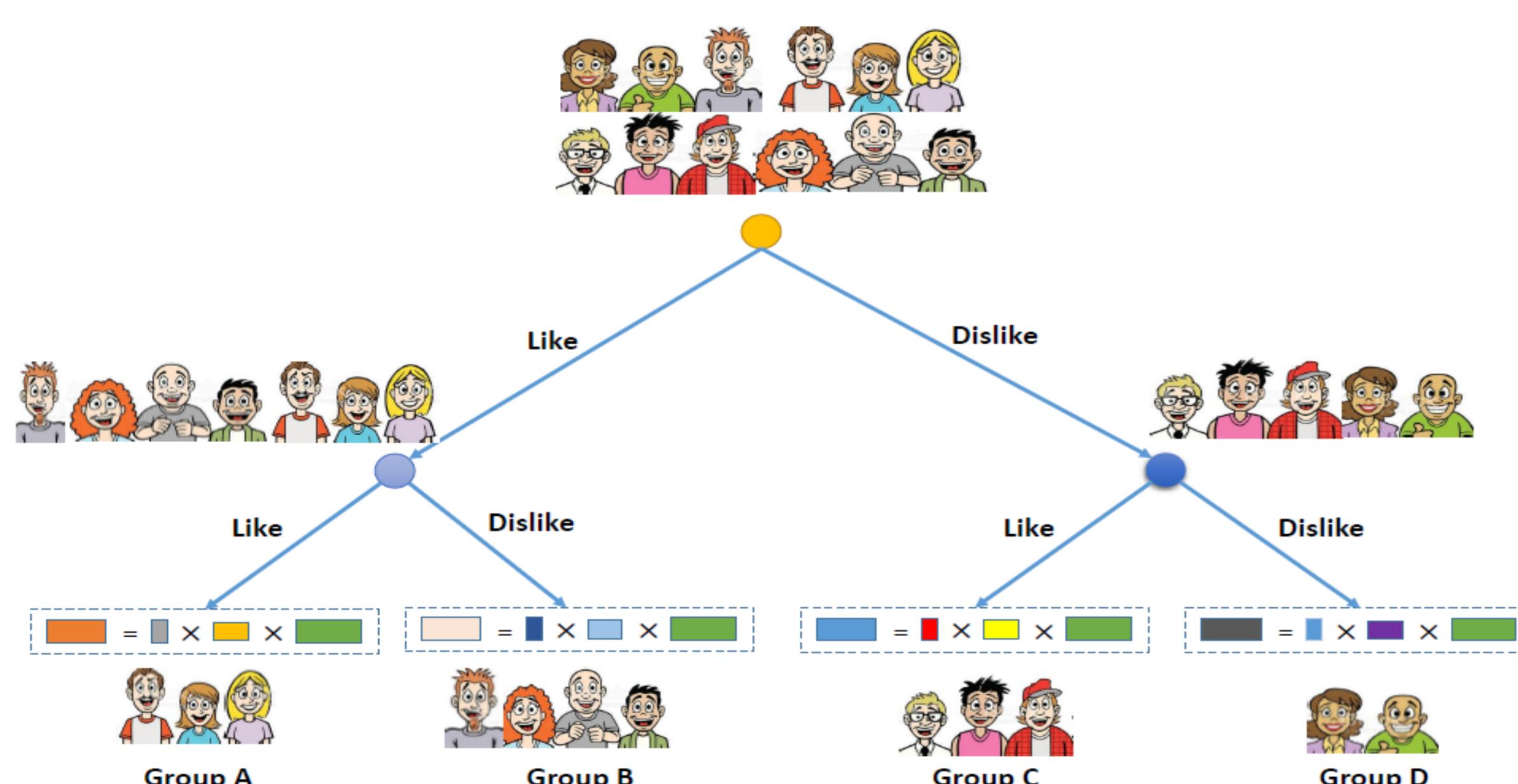


图1. 用户分组示意图

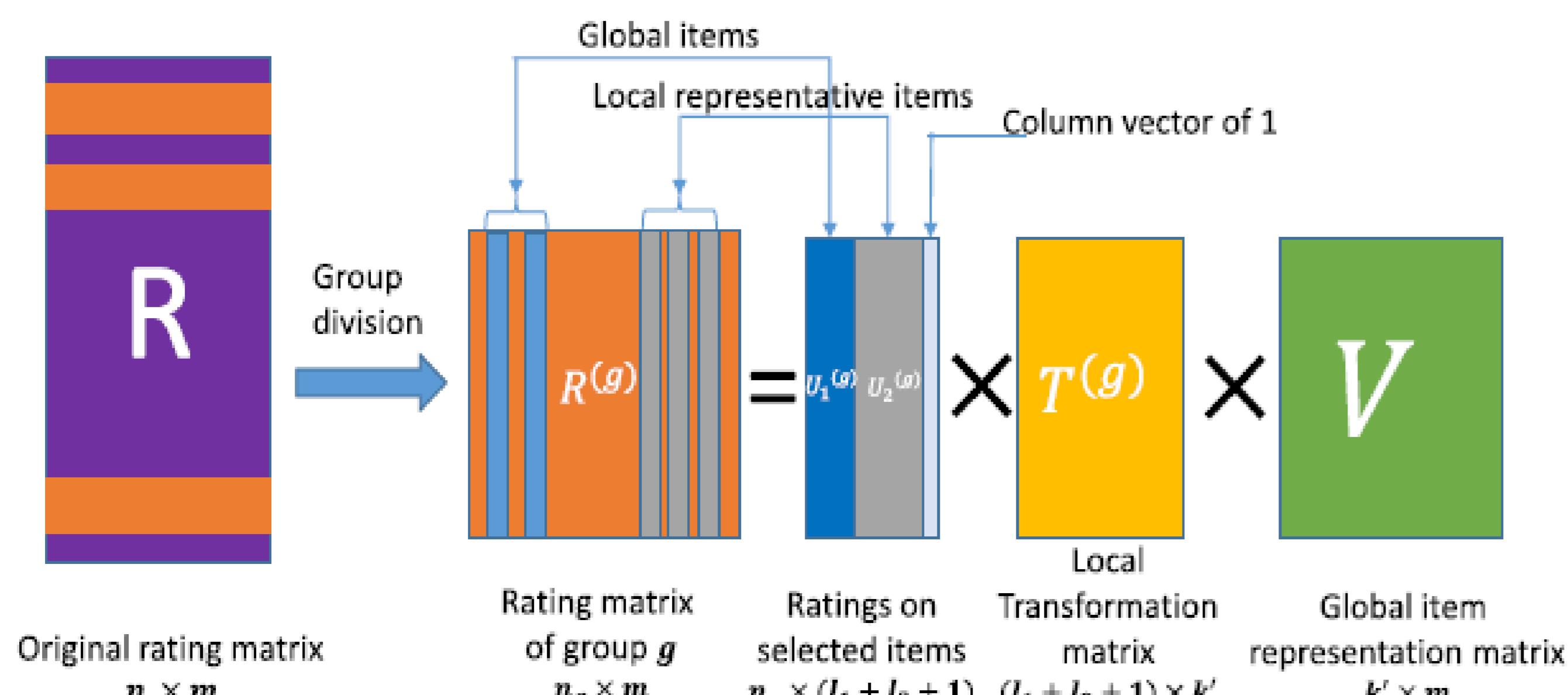


图2. 局部数据重构示意图

优化目标

$$\begin{aligned} \min_{\mathcal{G}, \mathbf{T}^{(g)}, \mathbf{U}_1^{(g)}, \mathbf{U}_2^{(g)}, \mathbf{V}} & \sum_{g \in \mathcal{G}} \{ \|\mathbf{R}^{(g)} - [\mathbf{U}_1^{(g)}; \mathbf{U}_2^{(g)}; \mathbf{e}] \mathbf{T}^{(g)} \mathbf{V} \|_F^2 + \alpha \|\mathbf{T}^{(g)}\|_F^2 \} + \beta \|\mathbf{V}\|_F^2 \\ \text{s.t.} & \cup_{g \in \mathcal{G}} \mathbf{R}^{(g)} = \mathbf{R} \text{ and } \mathbf{R}^{(g)} \cap \mathbf{R}^{(g')} = \emptyset, \\ & \mathbf{R}^{(g)} \in \mathbb{R}^{n_g \times m} \text{ and } \mathbf{R}^{(g)} \subset_{n_g, m} \mathbf{R}, \\ & \mathbf{U}_1^{(g)} \in \mathbb{R}^{n_g \times l_1} \text{ and } \mathbf{U}_1^{(g)} \subset_{n_g, l_1} \mathbf{R}^{(g)}, \\ & \mathbf{U}_2^{(g)} \in \mathbb{R}^{n_g \times l_2} \text{ and } \mathbf{U}_2^{(g)} \subset_{n_g, l_2} \mathbf{R}^{(g)}, \\ & \mathbf{V} \in \mathbb{R}^{k' \times m}, \mathbf{T}^{(g)} \in \mathbb{R}^{k' \times k'} \end{aligned}$$

我们提出上面的优化目标将**动态用户分组**、**局部数据重构**进行联合建模。具体的优化算法见我们的论文。

实验结果

Table 3. Performance Comparisons of Different Methods on the EachMovie Dataset (the Number of Questions Varies from 1 to 5)

Metrics	Methods	1	2	3	4	5
P@1	Rand	0.013	0.013	0.013	0.013	0.013
	Pop	0.160	0.160	0.160	0.160	0.160
	RBMF	0.402	0.370	0.405	0.442	0.467
	DMF	0.354	0.366	0.416	0.426	0.479
P@5	LRMF	0.394	0.425	0.471	0.492	0.500
	Rand	0.013	0.013	0.013	0.013	0.013
	Pop	0.282	0.282	0.282	0.282	0.282
	RBMF	0.269	0.322	0.352	0.361	0.373
P@10	DMF	0.340	0.354	0.361	0.379	0.394
	LRMF	0.348	0.377	0.394	0.411	0.416
	Rand	0.013	0.013	0.013	0.013	0.013
	Pop	0.231	0.231	0.231	0.231	0.231
P@50	RBMF	0.248	0.297	0.308	0.321	0.326
	DMF	0.287	0.301	0.322	0.333	0.344
	LRMF	0.309	0.325	0.342	0.354	0.360
	Rand	0.017	0.017	0.017	0.017	0.017
MAP	Pop	0.251	0.251	0.251	0.251	0.251
	RBMF	0.252	0.302	0.314	0.326	0.334
	DMF	0.294	0.291	0.330	0.330	0.355
	LRMF	0.319	0.334	0.355	0.367	0.374

Table 4. Performance Comparisons of Different Methods on the MovieLens Dataset (the Number of Questions Varies from 1 to 5)

Metrics	Methods	1	2	3	4	5
P@1	Rand	0.007	0.007	0.007	0.007	0.007
	Pop	0.384	0.384	0.384	0.384	0.384
	RBMF	0.292	0.325	0.354	0.436	0.459
	DMF	0.427	0.436	0.448	0.451	0.480
P@5	LRMF	0.433	0.444	0.467	0.479	0.493
	Rand	0.007	0.007	0.007	0.007	0.007
	Pop	0.353	0.353	0.353	0.353	0.353
	RBMF	0.274	0.314	0.338	0.390	0.403
P@10	DMF	0.381	0.394	0.407	0.416	0.429
	LRMF	0.388	0.403	0.420	0.426	0.439
	Rand	0.007	0.007	0.007	0.007	0.007
	Pop	0.322	0.322	0.322	0.322	0.322
P@50	RBMF	0.257	0.296	0.316	0.352	0.365
	DMF	0.347	0.360	0.372	0.379	0.388
	LRMF	0.352	0.367	0.378	0.387	0.395
	Rand	0.008	0.008	0.008	0.008	0.008
MAP	Pop	0.178	0.178	0.178	0.178	0.178
	RBMF	0.150	0.180	0.189	0.205	0.213
	DMF	0.198	0.207	0.215	0.222	0.232
	LRMF	0.202	0.212	0.226	0.232	0.241

更多的实验结果见我们的论文。

案例分析

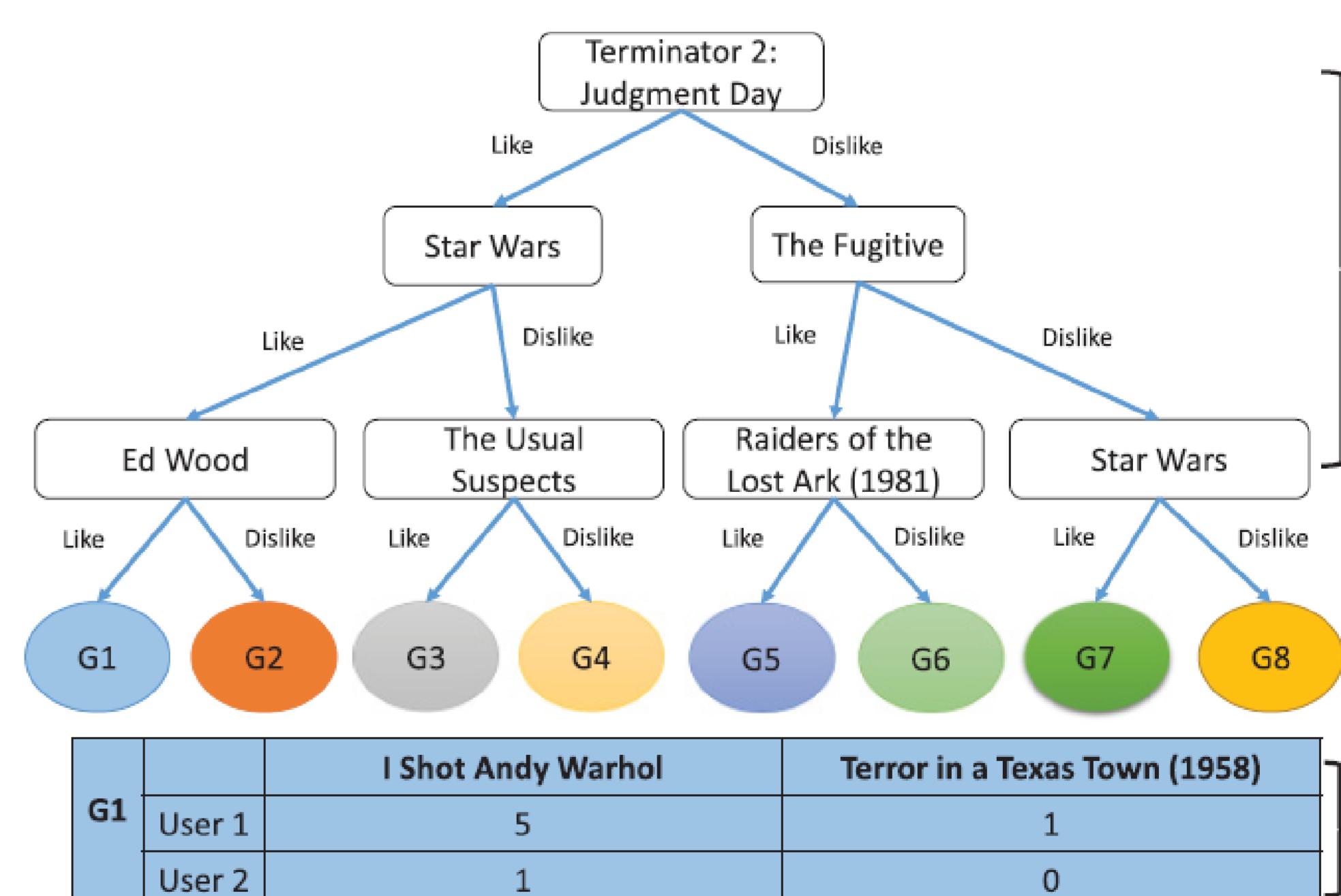


图3. 在EachMovie数据集上学习到的模型

结论

- 提出了基于局部代表性矩阵分解的冷启动推荐方法, 并验证了其有效性
- 局部代表性物品可以更有效地刻画用户细粒度偏好, 从而取得更好的推荐效果