

## 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
- 基于询问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:** 我们采用两轮询问机制: 第一轮询问进行**动态用户分组**; 第二轮询问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。

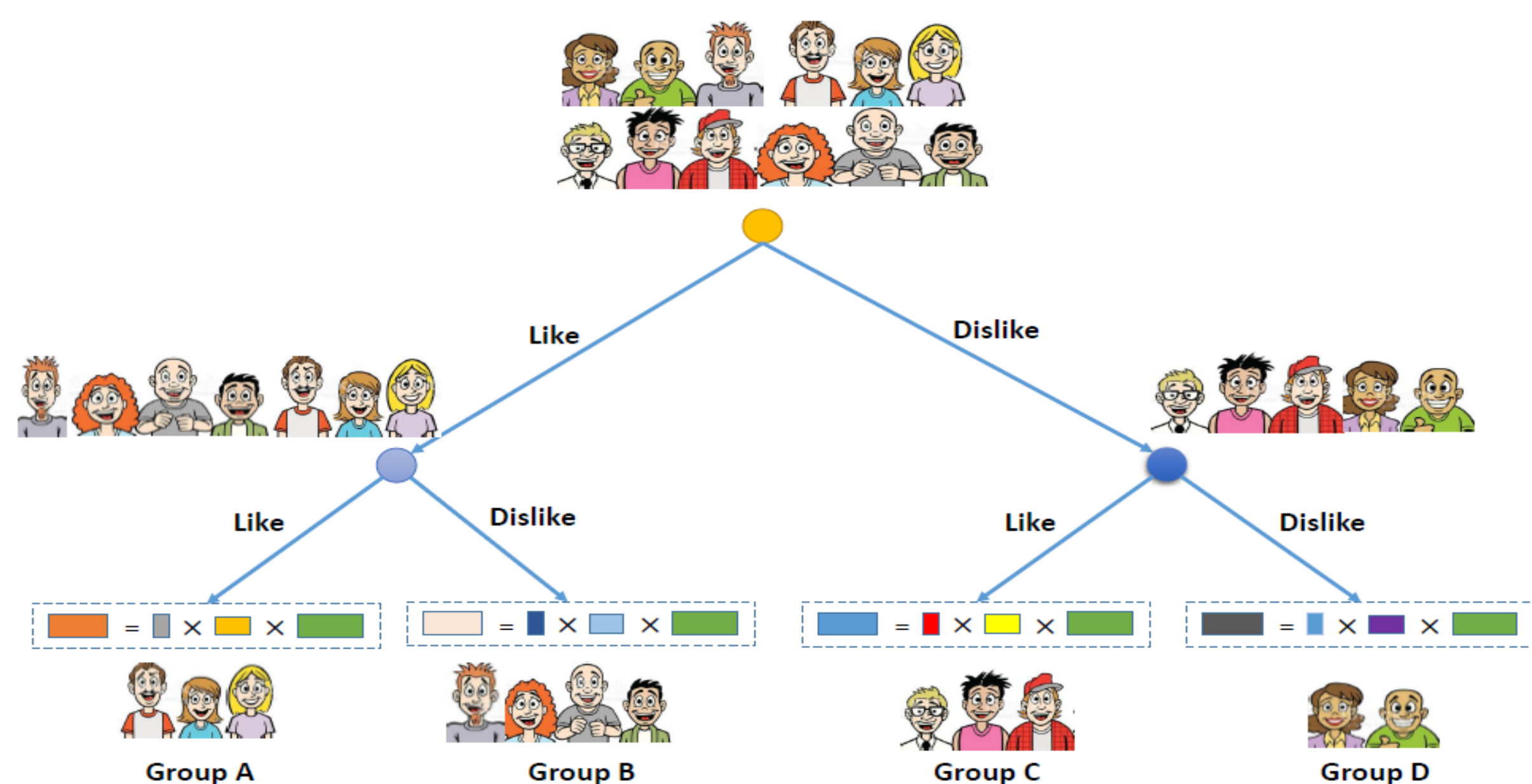


图1. 用户分组示意图

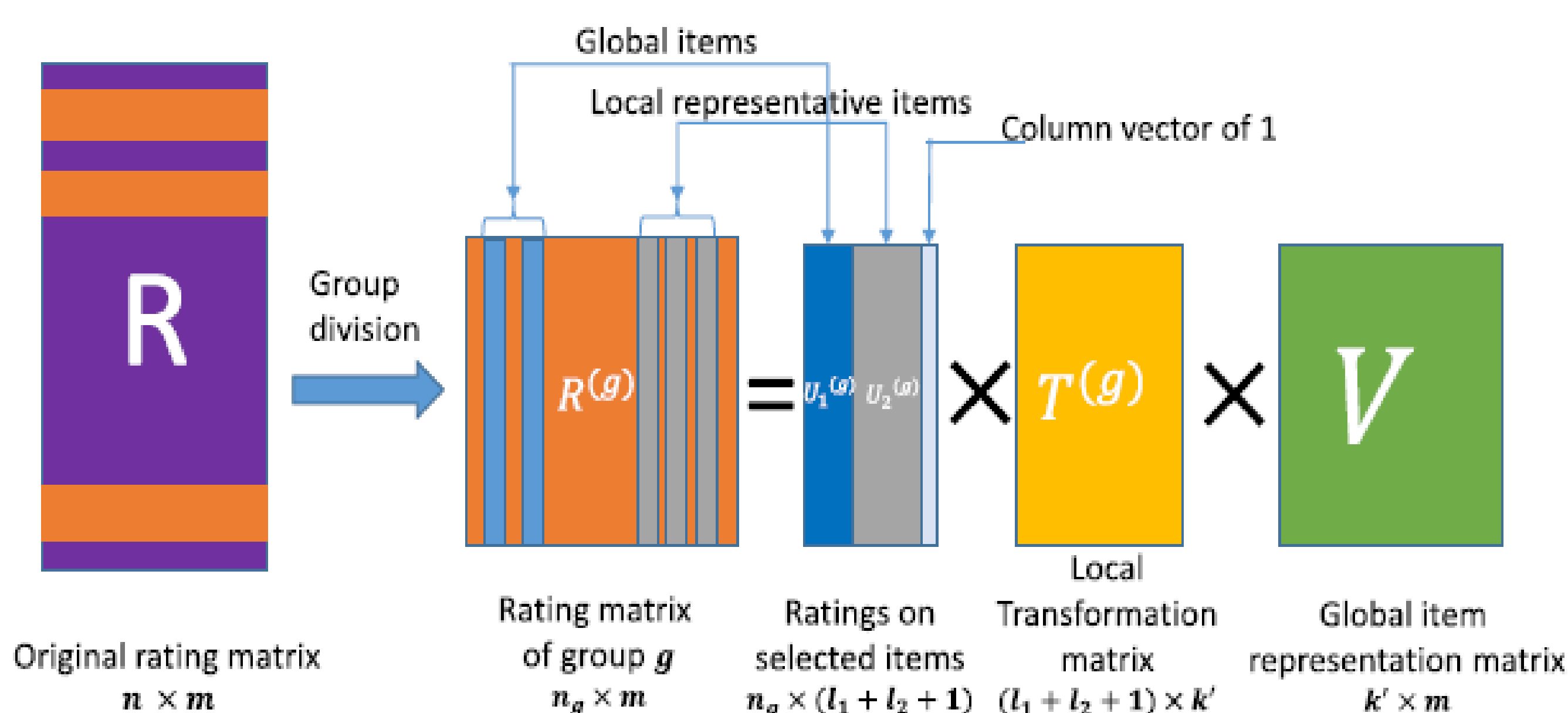


图2. 局部数据重构示意图

## 优化目标

$$\min_{\mathcal{G}, T^{(g)}, U_1^{(g)}, U_2^{(g)}, V} \sum_{g \in \mathcal{G}} \{ \|R^{(g)} - [U_1^{(g)}; U_2^{(g)}; e]T^{(g)}V\|_F^2 + \alpha \|T^{(g)}\|_F^2 \} + \beta \|V\|_F^2$$

$$s.t. \quad U_{g \in \mathcal{G}} R^{(g)} = R \text{ and } R^{(g)} \cap R^{(g')} = \emptyset,$$

$$R^{(g)} \in \mathbb{R}^{n_g \times m} \text{ and } R^{(g)} \subset_{n_g, m} R,$$

$$U_1^{(g)} \in \mathbb{R}^{n_g \times l_1} \text{ and } U_1^{(g)} \subset_{n_g, l_1} R^{(g)},$$

$$U_2^{(g)} \in \mathbb{R}^{n_g \times l_2} \text{ and } U_2^{(g)} \subset_{n_g, l_2} R^{(g)},$$

$$V \in \mathbb{R}^{k' \times m}, T^{(g)} \in \mathbb{R}^{k \times k'},$$

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	<b>0.402</b>	0.370	0.405	0.442	0.467
	fMF	0.354	0.366	0.416	0.426	0.479
	LRMF	0.394	<b>0.425</b>	<b>0.471</b>	<b>0.492</b>	<b>0.500</b>
P@5	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
	fMF	0.340	0.354	0.361	0.379	0.394
	LRMF	<b>0.348</b>	<b>0.377</b>	<b>0.394</b>	<b>0.411</b>	<b>0.416</b>
P@10	Rand	0.013	0.013	0.013	0.013	0.013
	Pop	0.231	0.231	0.231	0.231	0.231
	RBMF	0.248	0.297	0.308	0.321	0.326
	fMF	0.287	0.301	0.322	0.333	0.344
	LRMF	<b>0.309</b>	<b>0.325</b>	<b>0.342</b>	<b>0.354</b>	<b>0.360</b>
MAP	Rand	0.017	0.017	0.017	0.017	0.017
	Pop	0.251	0.251	0.251	0.251	0.251
	RBMF	0.252	0.302	0.314	0.326	0.334
	fMF	0.294	0.291	0.330	0.330	0.355
	LRMF	<b>0.319</b>	<b>0.334</b>	<b>0.355</b>	<b>0.367</b>	<b>0.374</b>

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
	fMF	0.427	<b>0.436</b>	0.448	0.451	0.480
	LRMF	<b>0.433</b>	<b>0.444</b>	<b>0.467</b>	<b>0.479</b>	<b>0.493</b>
P@5	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
	fMF	0.381	0.394	0.407	0.416	0.429
	LRMF	<b>0.388</b>	<b>0.403</b>	<b>0.420</b>	<b>0.426</b>	<b>0.439</b>
P@10	Rand	0.007	0.007	0.007	0.007	0.007
	Pop	0.322	0.322	0.322	0.322	0.322
	RBMF	0.257	0.296	0.316	0.352	0.365
	fMF	0.347	0.360	0.372	0.379	0.388
	LRMF	<b>0.352</b>	<b>0.367</b>	<b>0.378</b>	<b>0.387</b>	<b>0.395</b>
MAP	Rand	0.008	0.008	0.008	0.008	0.008
	Pop	0.178	0.178	0.178	0.178	0.178
	RBMF	0.150	0.180	0.189	0.205	0.213
	fMF	0.198	0.207	0.215	0.222	0.232
	LRMF	<b>0.202</b>	<b>0.212</b>	<b>0.226</b>	<b>0.232</b>	<b>0.241</b>

更多的实验结果见我们的论文。

## 案例分析

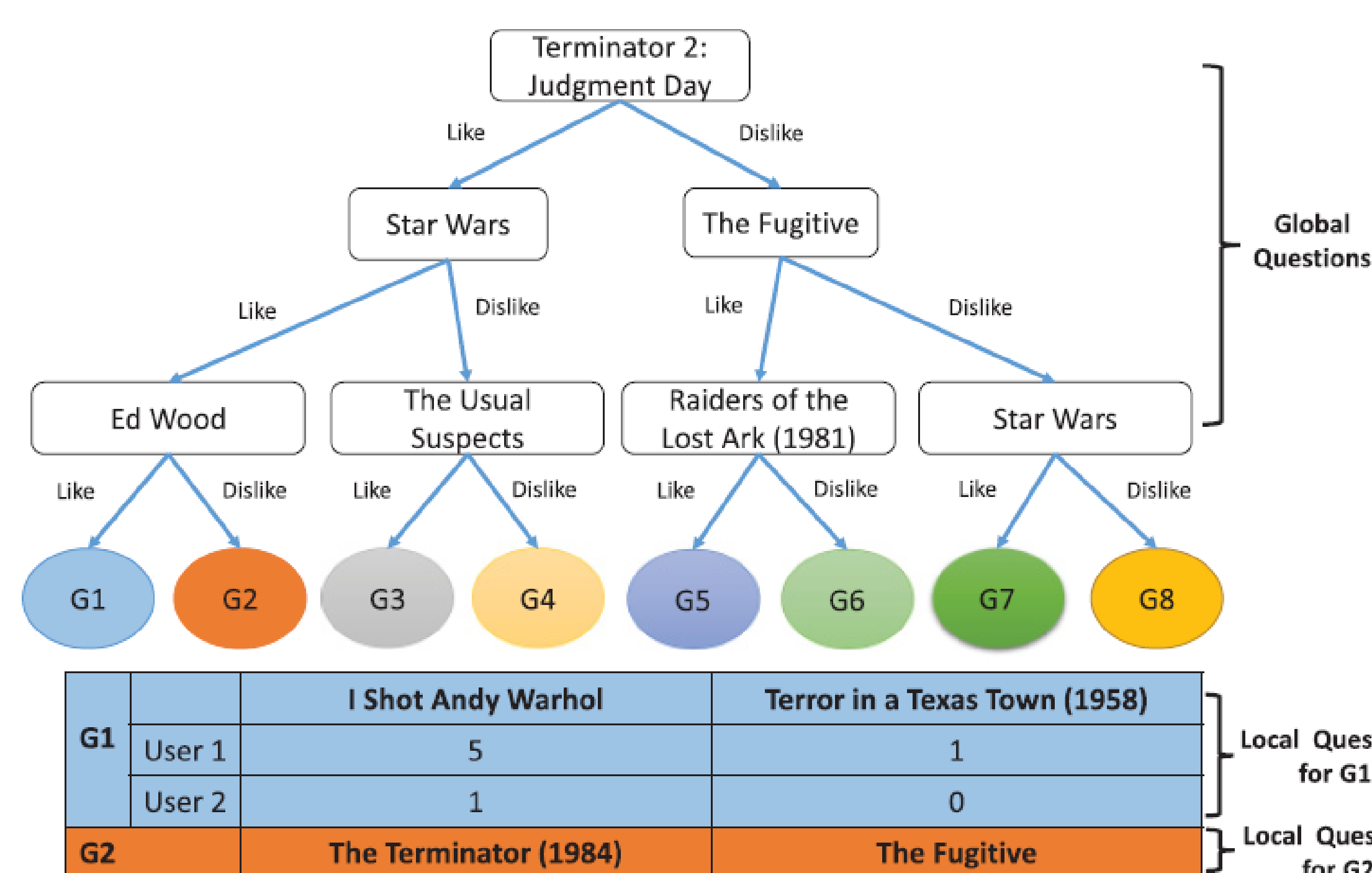


图3. 在EachMovie数据集上学习到的模型

## 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