

# 论文 | 组推荐系统及其应用研究

原创 Thinkgamer 搜索与推荐Wiki 2020-04-13

收录于话题

#论文笔记

19个



这篇文章主要是说明一下**群组推荐系统**，众所周知，推荐系统已经应用十分广泛，群组推荐的应用不仅老用户上发挥了极大的作用，在新用户的冷启动上也发挥了很大的作用。

由于后续会有一篇文章介绍结合深度学习的群组推荐，所以这里先借用该篇论文，借花献佛，介绍一下群组推荐系统及其应用。

本文中所涉及的内容算是比较老了，但在不同公司或者推荐系统的不同阶段仍然有应用，其中群组推荐的思想更是经久不衰，所以本文比较适合以“温习知识”或“了解群组推荐”，至于群组推荐的更加先进和业界比较常用的做法可以持续关注公众号，后续会进行介绍。

## 传统推荐系统

推荐问题通常被形式化描述为一个定义在用户空间和项目空间上的效用函数，推荐系统的目的是把特定的项目推荐给用户，使效用函数最大化。

目前推荐系统的分类没有统一的标准，从不同的角度可以进行不同的划分。

- 从推荐模型的角度：
  - 协同过滤推荐
  - 基于内容的推荐
  - 基于知识的推荐
  - 组合推荐
  - 基于人口统计学的土建
  - 基于效用的推荐
  - 基于关联规则的推荐
  - 基于网络结构的推荐
- 从推荐系统的数据源和应用环境的角度：
  - 上下文感知推荐
  - 社会化推荐
  - 移动推荐
- 从推荐系统技术的角度：
  - 基于规则的推荐
  - 基于机器学习的推荐
  - 基于深度学习的推荐
  - 基于强化学习的推荐
  - 基于知识图谱的推荐

### 群组推荐系统的形式化定义

---

许多日常活动中是由多个用户以群组的形式参与的，因此推荐系统需要考虑群组中每个用户的偏好来进行推荐，这种推荐系统称为群组推荐系统（Group Recommend System），它将推荐对象由单个用户扩展为一个群组，这为推荐系统带来了一些挑战。

群组成员的兴趣偏好可能很相似，也可能存在较大差异，如何获取群组成员的共同偏好，来缓解群组成员之间的偏好冲突，使推荐的结果尽可能满足所有群组成员的需求，则是群组推荐研究需要解决的关键问题。

群组推荐目前还没有统一的形式化定义，有文献基于**共识分数**的概念给出形式化定义：

#### 1、群组预测评分

群组  $G$  对项目  $i$  的预测评分  $grouprate(G, i)$  由群组中每个用户预测评分  $rate(u, i)$  融合得到，使用的偏好融合策略不同，群组计算的预测方法也不同，其中最常用的则是均值策略。

$$grouprate(G, i) = \frac{1}{|G|} \sum_{u \in G} rate(u, i)$$

## 2、群组分歧度

群组对项目度分歧度  $dis(G, i)$  表示群组  $G$  中的用户对项目  $i$  预测评分的差异程度。群组分歧度也有多种计算方法，这里给出一种最常用的方法：**利用分歧方差 (disagreement variabce)** 计算群组分歧度。

$$dis(G, i) = \frac{1}{|G|} \sum_{u \in G} (rate(u, i) - mean(G, i))^2$$

其中  $mean(G, i)$  是群组  $G$  中的用户对项目  $i$  预测评分的均值。

## 3、共识函数

共识函数  $F(G, i)$  利用群组预测评分和组分歧计算得出群组  $G$  对项目  $i$  的共识分数。

$$F(G, i) = w_1 * grouprate(G, i) + w_2 * (1 - dis(G, i))$$

其中：

- $w_1$  表示群组预测分在共识函数中的权重
- $w_2$  表示群组分歧度在共识函数中的权重

组推荐的目标是对于给定的群组  $G$  和项目集合  $I$ ，找到特定的项目  $i$  使共识函数最大化。即群组预测评分最高，群组分歧度越低，就越能满足所有群组成员的偏好。

## 4、Top-k 组推荐

群组推荐将前  $k$  个共识分数最高的项目推荐给群组，即Top-k推荐，推荐列表中的项目  $i$  按照共识分数降序排序。

从上述的形式化定义看出，群组推荐系统首先通过传统的推荐算法为每个用户生成预测评分，然后根据用户预测评分得到群组预测评分。然而**这并不是所有群组推荐系统通用的方法**。此外，**该形式化定义与传统推荐系统的形式化定义并不统一不利于组推荐问题的描述与求解**。因此，群组推荐系统的形式化定义仍是一个开放问题。

虽然上述的群组推荐的形式化定义不通用，但是在具体的应用场景可以结合自己的**群组预测评分和群组分歧度**进行使用。

群组推荐与传统推荐的异同

下表展示了群组推荐系统与传统推荐系统的比较

-	传统推荐系统	群组推荐系统
推荐目标	单个用户	群组
用户偏好获取	不需要偏好共享	需要偏好共享
偏好融合	不需要	需要
推荐结果展示	独立方式	共享方式
支持用户交互	不需要	需要

当然群组推荐也和传统的推荐系统存在一些相同之处，比如：

- 常见的协同、基于内容、组合推荐等也会用在群组推荐中
- 群组推荐也可以作为一路召回为单个用户推荐
- 群组推荐可以帮助解决“用户冷启动”问题

群组推荐系统研究框架

群组推荐的生命周期分为4个阶段：

- 收集群组成员数据
- 获取群组成员偏好信息
- 生成群组推荐
- 推荐结果的评价和反馈

下面展示的是从面向过程角度提出的4层群组推荐系统研究框架：

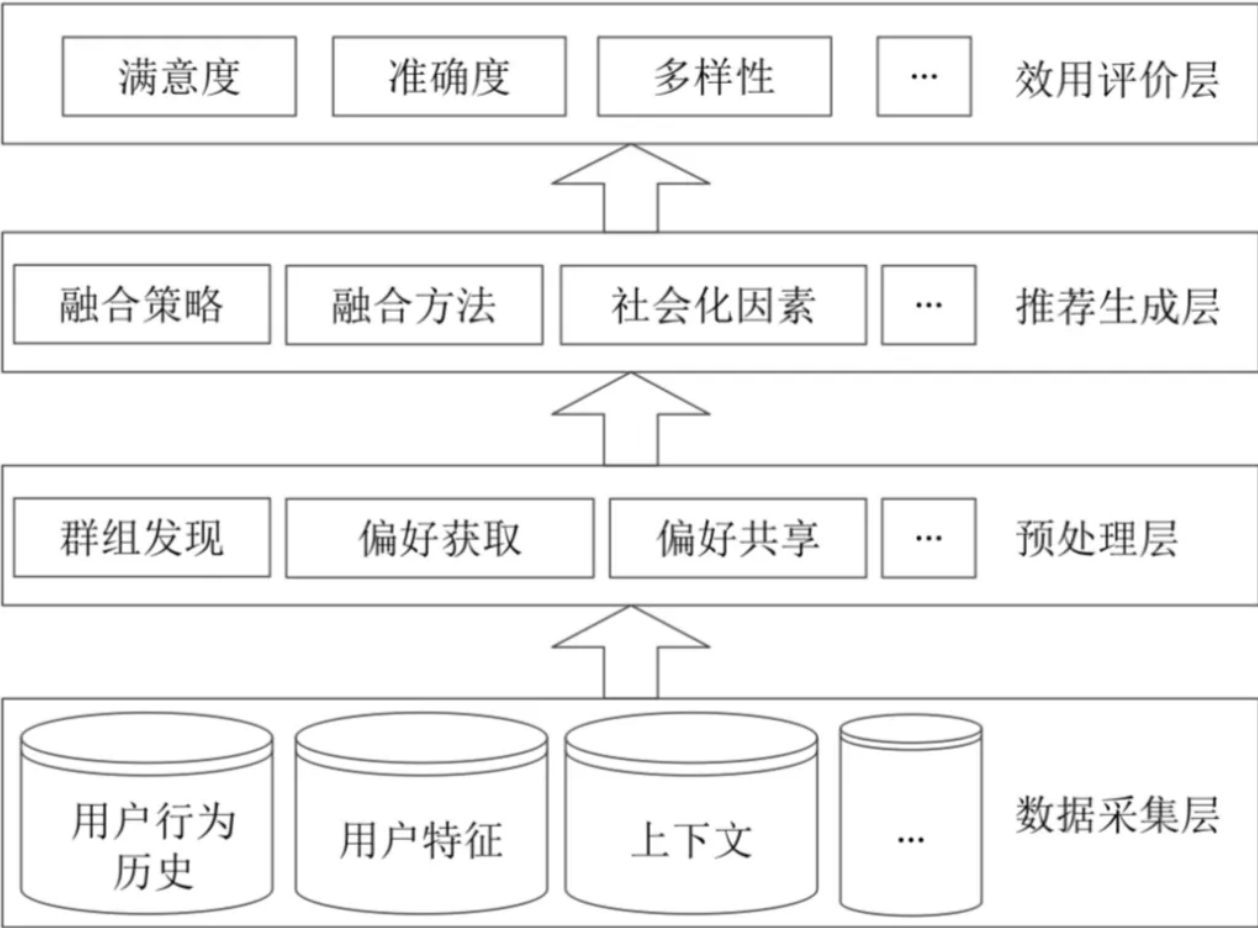


图 1 组推荐系统研究框架

搜索与推荐Wiki  
[https://blog.csdn.net/Gamer\\_gyt](https://blog.csdn.net/Gamer_gyt)

4层群组推荐系统研究框架

底层的数据源包括：

- 群组成员行为历史
- 群组成员浏览记录
- 群组用户-项目评分
- 群组成员人口统计学特征
- 项目特征
- 上下文信息
- ...

数据预处理层则是通过分析用户和项目的源数据信息构建用户偏好模型。与传统推荐系统相比，群组推荐系统的用户偏好过程更加关注用户的负向偏好（因为，负向偏好数据更容易获取，且在构建群组推荐候选池的时候可以过滤掉负向物品）和群组成员偏好的共享。

推荐生成层则是将群组成员偏好映射到群组偏好，通过一定的融合策略和推荐融合方法生成群组推荐过程。

群组推荐效果评价层则是采用一些评价方法对群组推荐进行评价，常见的：

- 用户满意度
- 准确度
- 覆盖率
- 多样性
- 精细度
- ...

### 群组推荐的关键技术

---

#### 用户偏好获取

用户行为包括：

- 显式反馈：能明确表示用户对内容喜好等级的数据，比如评分，打星等
- 隐式反馈：用户与物品之间产生的交互行为，没有明确的表示喜好等级

#### 群组推荐的群组发现

目前群组发现方法主要是根据用户的偏好和人口统计学信息等特征对用户进行聚类（**当然这已经不是目前的主流方法了，能对用户聚类的方式有很多，比如用户的偏好、人口统计学、使用设备、各种Embedding向量等**）。

进行聚类的算法主要有：

- 层次化聚类
- 划分式聚类
  - KMeans聚类
  - 二分KMeans聚类
- 基于滑窗的聚类
  - Mean-Shift 聚类
- 基于密度的聚类
  - DBSCAN
- 基于网格的聚类
- 新发展的一些方法
  - 基于约束（eg: EM聚类）

- 基于模糊
- 基于粒度
- 量子聚类
- 核聚类
- 谱聚类

群组推荐的偏好融合算法

偏好融合策略

基本融合策略

群组推荐中常用的融合策略如下：

表 2 组推荐系统常用的融合策略			
名称	公式	注释	说明
公平策略	$p^G = \{ loop(d^u) : \exists d^u \in \bigcup_{u \in G} p^u \}$	$p^u$ 为用户偏好模型 $p^G$ 为群组偏好模型, $d^u$ 为用户对项目或项目特征的评分, $loop$ 为循环取值	每次推荐轮流选择群组成员的评分作为群组评分
均值策略	$p^G = \{ avg(d^u) : \exists d^u \in \bigcup_{u \in G} p^u \}$	$avg$ 为求平均值	所有群组成员评分的平均值作为群组评分
痛苦避免均值策略	$p^G = \{ avg(d^u) : \exists d^u \in \bigcup_{u \in G} p^u \wedge d^u > \alpha \}$	$\alpha$ 为痛苦阈值	设定一个痛苦阈值, 所有高于痛苦阈值的群组成员评分的平均值作为群组评分
最受尊敬者策略	$p^G = \{ d^u : \exists d^u \in p^{u_{resp}} \wedge u_{resp} \in G \}$	$u_{resp}$ 为最受尊敬用户	从群组成员中选出一个用户, 将这个用户的评分作为群组评分
最小痛苦策略	$p^G = \{ \min(d^u) : \exists d^u \in \bigcup_{u \in G} p^u \}$	$\min$ 为求最小值	选择所有群组成员评分中的最低评分作为群组评分
最开心策略	$p^G = \{ \max(d^u) : \exists d^u \in \bigcup_{u \in G} p^u \}$	$\max$ 为求最大值	选择所有群组成员评分中的最高评分作为群组评分

群组推荐中常用的融合策略

加权模型在进行偏好融合时，应当考虑到不同的群组成员对群组偏好的不同影响。加权模型根据群组成员的特征、角色、影响力等因素，为每个群组成员分配不同的权重，这类加权模型可分为：

- 静态模型：根据群组成员的年龄、性别等人口统计学信息对用户进行分类 然后对不同类别赋予不同的权重，比如：一个针对00后的社区对青年和老年人分配较小的权重
- 交互模型：根据用户与内容之间的交互行为来分配权重。比如：用户的行为越多，说明用户越活跃，则权重越大

研究表明：交互模型的推荐效果通常比静态模型要好。

一种常见的交互模型加权方法为：根据用户 $u_x$ 的角色 $role(u_x)$ 和活跃度 $act(u_x)$ 来计算用户在群组 $f_a$ 中所占的权重可以用以下两种方式计算：

- The role-based model

$$w(u_x, f_a) = \frac{\sum_{y \in U} act(u_y) | role(u_y) = role(u_x)}{\sum_{y \in U} act(u_y)}$$

$role(u_y) = role(u_x)$ 表示和用户 $u_x$ 角色相同的用户

- The family-log model

$$w(u_x, f_a) = \frac{act(u_x)}{\sum_{y \in U} act(u_y)}$$

上述公式来自论文：Group-Based Recipe Recommendations: Analysis of Data Aggregation Strategies (<https://dl.acm.org/doi/10.1145/1864708.1864732>)

## 多种策略融合

在传统的推荐系统中，组合推荐方法通过组合多种推荐算法能够提高推荐效果，同样群组推荐系统也可以通过多种融合策略来应对单一融合策略的不足。同样是上文中提到的一篇论文，介绍的一种组合策略，公式如下：

$$w'(u_x, f_a) = \begin{cases} 1 & , rat(u_x, item_i) \notin (r_1, r_2) \\ (u_x, f_a) & , other \end{cases}$$

其中：

- $r_1$ 表示最小痛苦策略的最小阈值
- $r_2$ 表示最开心策略的最大阈值
- $rat(u_x, item_i) \notin (r_1, r_2)$ 表示  $u_x$ 对 $item_i$ 的评分不在 $(r_1, r_2)$ 时权重为1，其他为 $w(u_x, f_a)$

## 偏好融合方法

偏好融合的方法分为：

- 推荐结果融合
- 评分融合
- 群组偏好建模

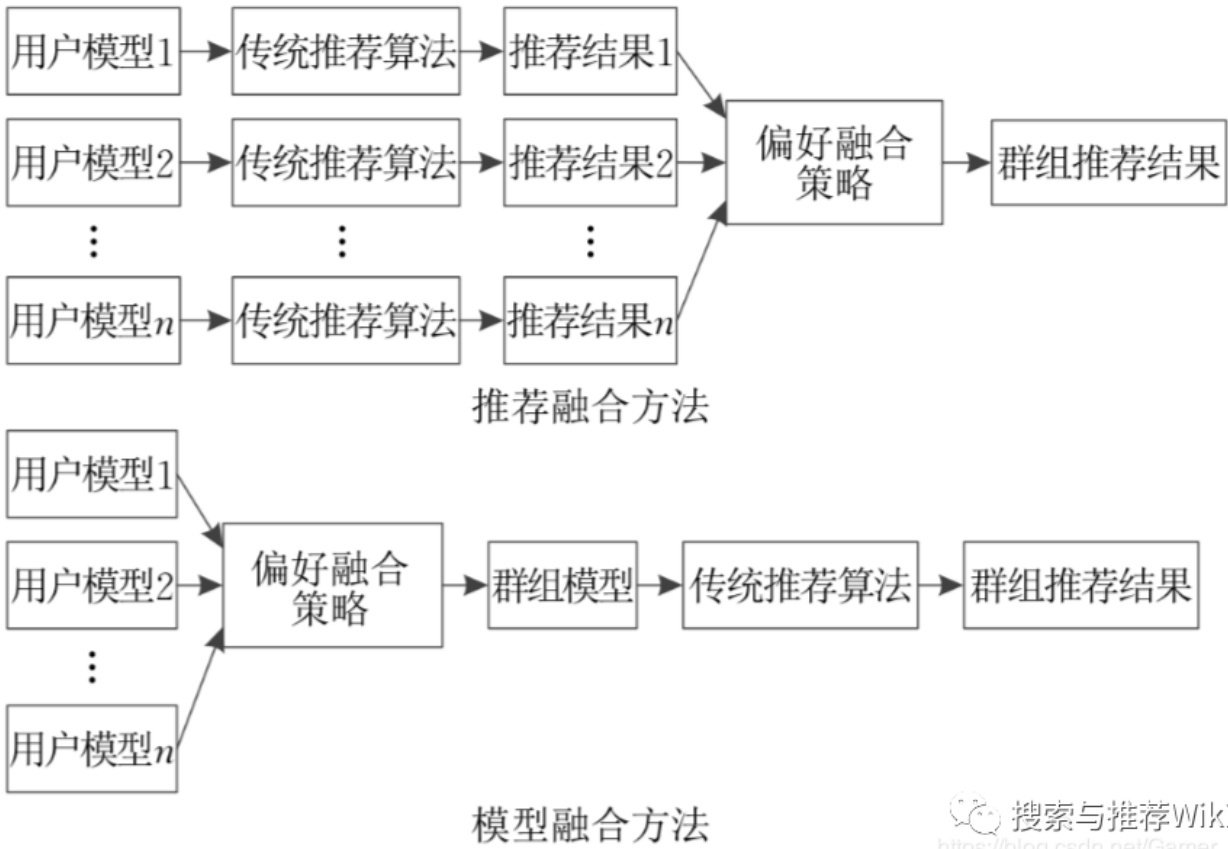
其中推荐结果融合和评分融合都是先为每个群组成员生成推荐结果，再汇聚到群组上，群组偏好建模则是先把群组成员的偏好模型融合生成群组的偏好模型，再根据群组偏好模型生成群组推荐。

因此根据偏好融合的时机，可以将偏好融合方法分为：



- 模型融合
- 推荐融合

其区别如下：



搜索与推荐Wiki  
[https://blog.csdn.net/Garner\\_gyt](https://blog.csdn.net/Garner_gyt)

模型融合

模型融合的形式化表示如下：

$$rat(g_a, item_i) = \frac{\sum_{u_x \in g_a} w(u_x, g_a) rat(u_x, item_i)}{\sum_{u_x \in g_a} w(u_x, g_a)}$$
$$pred(g_a, item_n) = \frac{\sum_{g_b \in G} sim(g_a, g_b) rat(g_b, item_n)}{\sum_{g_b \in G} sim(g_a, g_b)}$$

其中：

- $rat(g_a, item_i)$ 表示群组 $g_a$ 对项目 $item_i$ 的评分
- $rat(u_x, item_i)$ 表示用户 $u_x$ 对项目 $item_i$ 的评分
- $w(u_x, g_a)$ 表示用户 $u_x$ 在群组 $g_a$ 中的权重
- $pred(g_a, item_n)$ 表示群组 $g_a$ 对未评分项目 $item_n$ 的预测评分
- $sim(g_a, g_b)$ 表示群组 $g_a$ 和 $g_b$ 的偏好相似度

当然这里只是举例说明模型融合的策略，也有很多其他方式的形式化表示。

推荐融合

推荐融合的形式化表示为：

$$pred(g_a, item_i) = \frac{\sum_{u_x \in g_a} w(u_x, g_a) pred(u_x, item_i)}{\sum_{u_x \in g_a} w(u_x, g_a)}$$

模型融合和推荐融合的形式化表示公式来自论文：Group-Based Recipe Recommendations: Analysis of Data Aggregation Strategies (<https://dl.acm.org/doi/10.1145/1864708.1864732>)

偏好融合策略比较

下表给出了推荐准确率较高的几种偏好融合策略、偏好融合方法、推荐算法组合

偏好融合策略	偏好融合方法	推荐算法
最开心策略	模型融合	基于内容的推荐
最小痛苦策略	推荐融合	基于内容的推荐
痛苦避免均值策略	模型融合	基于用户的协同过滤
均值策略	推荐融合	组合推荐

群组推荐的评价

- 准确度：RMSE、MAP、F1、E(k)、Recall、Precision、AUC、NDCG
- 用户满意度
- 多样性
- 覆盖率
- 惊喜度

群组推荐研究的难点和发展方向

群组推荐的偏好融合

包括偏好融合策略和偏好融合方法。目前的融合策略主要关注的是公平性，但是这些预定义的策略和数据本身无关。

基于模型的群组推荐方法

现有的群组推荐方法也可以分为：

- 基于内存
- 基于模型

基于偏好融合的群组推荐算法则属于基于内存的方式，主要缺陷是使用启发式的融合策略，忽略了群组成员之间的相互影响。

基于模型的方法通过构建一个生成模型，对群组选择项目的生成过程进行建模，从而能够通过考虑群组成员之间的交互来进行更精准的推荐。

---

### 群组推荐中的不确定性

大部分推荐系统假设获取的用户偏好是准确的，但用户偏好本身存在不确定性，特别是通过隐式反馈获取的数据本身就具有很大的不确定性。

---

### 群组推荐的解释和可视化

对群组推荐的结果进行合理的解释并以合适的方式呈现给群组能够提高透明度，使用户能够更好的理解推荐机制的其他成员的偏好，从而更加容易接受推荐系统。

---

### 群组推荐的效用评价

由于群组类型、规模以及群组推荐生成方式的多样性使得群组推荐的评价比传统推荐系统的评价更加困难。

此外，目前推荐系统的评价集中在推荐准确度方面，而对推荐的覆盖率、多样性、惊喜度等方面关注较少。

---

### 上下文感知的群组推荐

除了用户和内容之间的交互信息，如何将用户的地理位置、时间、环境、情绪等因素应用到群组推荐中也是一大难点。

---

### 移动群组推荐

移动群组推荐关注的是在移动设备上，用户的需求变得更加不可控，如何实时有效的捕获用户的兴趣也是非常值得关注的地方。

---

### 群组推荐的用户隐私和安全

偏好共享和提供用户交互有利于群组推荐系统性能的提升，但同时也带来了用户隐私问题。

虽然在传统的推荐系统中，用户隐私和安全也是值得关注的，但是和群组推荐相比，群组推荐则更加复杂。

群组推荐中存在传统推荐问题

比如冷启动等。

OK，普及完毕！

真正的努力，都不喧嚣！



搜索与推荐Wiki

All In CTR、DL、ML、RL、NLP

❤️原创不易，点个“**在看**”鼓励鼓励吧

收录于话题 #论文笔记·19个

上一篇

论文 | 看腾讯如何玩转实时推荐-TencentRec

下一篇

【论文】序列推荐模型 TransRec

阅读原文