**任务描述**

商品分享是用户在电商平台中主要的社交互动形式，商品分享行为包含了用户基于商品的社交互动偏好信息，用户的商品偏好和社交关系随时间变化，其基于商品的社交互动属性也随时间改变。本任务关注社交图谱动态链接预测问题，用四元组 (u, i, v, t) 表示用户在t (time) 时刻的商品分享互动行为，其中i (item) 标识特定商品，u (user) 表示发起商品链接分享的邀请用户，v (voter) 表示接收并点击该商品链接的回流用户。因此在本任务中链接预测指的是，在已知邀请用户u，商品i和时间t的情况下，预测对应回流用户v。

**任务目标**

针对社交图谱动态链接预测，参赛队伍需要根据已有动态社交图谱四元组数据，对于给定u，i，t，对v进行预测。

**数据集介绍**

**初赛训练集&测试集**

在初赛阶段，我们会发布用户动态商品分享数据作为训练集。训练集由三部分构成。

* item\_share\_train\_info.json：用户动态商品分享数据，每行是一个json串，具体字段信息如下：

| **字段** | **字段说明** |
| --- | --- |
| inviter\_id | 邀请用户id |
| item\_id | 分享商品id |
| voter\_id | voter的用户id |
| timestamp | 分享行为发生时间 |

* user\_info.json：用户信息数据，每行是一个json串，具体字段信息如下：

| **字段** | **字段说明** |
| --- | --- |
| user\_id | 用户的id，包含数据集中所有的inviter\_id和voter\_id |
| user\_gender | 用户性别，0表示女性用户，1表示男性用户，未知为-1 |
| user\_age | 用户所在年龄段，数值范围为1～8，数值越大表示年龄越大，未知为-1 |
| user\_level | 用户的平台积分水平，数值范围为1~10，数值越大表示用户的平台积分水平越高 |

* item\_info.json：商品信息数据，每行是一个json串，具体字段信息如下：

| **字段** | **字段说明** |
| --- | --- |
| item\_id | 商品id |
| cate\_id | 商品叶子类目id |
| cate\_level1\_id | 商品一级类目id |
| brand\_id | 商品所属的品牌id |
| shop\_id | 商品所属的店铺id |

在初赛阶段，我们会发布测试集item\_share\_ preliminary\_test\_info.json，每行是一个json串，具体字段信息如下：

| **字段** | **字段说明** |
| --- | --- |
| inviter\_id | 邀请用户id |
| item\_id | 分享商品id |
| timestamp | 分享行为发生时间 |

**复赛测试集**

在复赛阶段我们发布复赛训练集item\_share\_final\_train\_info.json和复赛测试集item\_share\_final\_test\_info.json，复赛结果提交方式和初赛保持一致，具体参见任务提交说明章节。

**任务提交说明**

参赛队伍可在初赛和复赛阶段下载对应的测试集数据。每一行包括 triple\_id, inviter\_id, item\_id, timestamp。

测试集示例如下：

[

{"triple\_id":"0", "inviter\_id":"0", "item\_id":"12423", "timestamp":"2023-01-05 10:11:12"}

]

本次比赛的目标对给定的邀请者inviter\_id，物品item\_id和时间信息timestamp，预测此次交互行为的回流者voter\_id。参赛队伍需要从用户信息数据 (见user\_info.json) 中选择用户，给出一个长度为5的候选回流者 (voter\_id) 列表，其中预测概率大的voter排在前面，把对应的user\_id填入candidate\_voter\_list，最后以json文件的形式提交。初赛和复赛的提交文件形式相同。

提交文件示例如下：

[

{"triple\_id":"0",

"candidate\_voter\_list": ["1", "2", "3", "4", "5"] }

]

**评价指标**

本任务采用MRR（Mean Reciprocal Rank，平均倒数排名）指标来评估模型的预测效果。

对于一个query：(inviter, item, ?, timestamp )，我们需要根据概率找出可能的候选voter列表，为candidate\_voter\_list，概率高的排在前面。若真实的回流者排在第n位，则该次query的分数为1/n。由于候选voter列表长度为5，所以此处n最大为5，如果列表中不包含真实的回流者，则此次query的得分为0。最终对多个query的分数求平均，计算公式如下：

