

# 基于外出与日常两种签到行为的 POIs 推荐系统

2016 年 9 月 26 日

核心思想: 我们发现用户签到行为明显的分为两类, 一类是日常附近签到行为, 另一类是外出时签到行为, 这两类行为下产生的签到地点、类别、消费水平都是不一样的。因此我们根据距离的远近, 我们将用户 check-in 行为划分为两类: 就近 check-in 行为, 外出 check-in 行为。对两种 check-in 行为, 我们分别建立模型。

外出 check-in 行为:

考虑在城市中外出时, 用户常常会首先考虑去哪个商业圈, 其次再去考虑去这个商业圈中进行怎样的消费。基于这样的考虑, 我们将所有 check-in 聚类, 找出商业圈 (center), 依据用户对已有商业圈的评分, 预测他对所有商业圈的评分。

$$E_c = \frac{\lambda_c}{2} \sum_{i=1}^N \sum_{c=1}^G I_{i,c} (R_{i,c} - U_i^T L_c)^2 + \frac{\lambda_u}{2} \|U\|_F^2 + \frac{\lambda_l}{2} \|L\|_F^2 \quad (1)$$
$$P_{r_c} \propto l_{i,c} \times U_i^T L_c$$

上式中  $R_{i,c}$  是用户  $U_i$  对商业中心  $L_c$  的评分, 我们对商业圈评价进行矩阵分解, 预测人们对商业圈的打分  $P_{r_c}$ ,  $l_{i,c}$  是用户  $i$  到 center  $L_c$  的距离。

最终的外出行为推荐系统:

$$E = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M I_{i,j} (R_{i,j} - P_{i,c_j} U_i^T V_j)^2 + \frac{\lambda_u}{2} \|U\|_F^2 + \frac{\lambda_v}{2} \|V\|_F^2 \quad (2)$$
$$+ \frac{\lambda_k}{2} \sum_{i=1}^N \sum_{k=1}^H I_{i,k} Z_{i,k} \sum_{f \in N_{i,k}} S_{i,f} \|U_i - U_f\|_F^2$$

在 Baseline 中引入商圈因子  $P_{i,c_j}$ ,  $P_{i,c_j}$  表示用户对地点  $j$  所在的商业圈  $c_j$  的预测评分。

同时考虑了社交改进项, 很多论文中表示, 基于用户相似性的推荐在 POIs 推荐系统中表现并不好, 社交系统也没有起到很大的作用。这是因为 check-in 类别太分散, 签到太稀疏。事实上, 我们并不需要两个用户在所有方面都非常相似才可以做推荐, 只要两个用户在某方面相似度较高就可以了。因此上式中第三行是将用户按照兴趣重复性聚类, 同时考虑了用户对该类别东西的喜欢程度  $Z_{i,k}$  与用户的相似度。见图 figure1。

就近 check-in 行为:

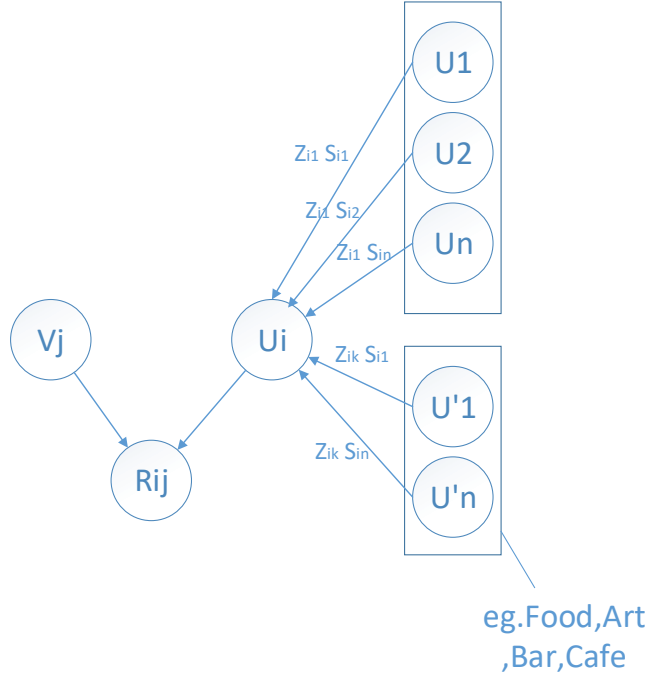


Figure 1: 基于用户兴趣聚类推荐图.

就近 check-in 行为不需要考虑商业中心，额外考虑就近行为签到时的消费水平比较稳定，且在社交部分我们更多的考虑生活轨迹相同的用户作为朋友进行推荐。

$$\begin{aligned}
 E &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M I_{i,j} (R_{i,j} - U_i^T V_j)^2 + \frac{\lambda_u}{2} \|U\|_F^2 + \frac{\lambda_v}{2} \|V\|_F^2 \\
 &+ \frac{\lambda_k}{2} \sum_{i=1}^N \sum_{k=1}^H I_{i,k} Z_{i,k} \sum_{f \in N_{i,k}} G_{i,f} \|U_i - U_f\|_F^2 \\
 &+ \frac{\lambda_d}{2} \sum_{k \in M} P_{j,k} \|V_i - V_k\|_F^2 \\
 E &= E \times l_{i,j}
 \end{aligned} \tag{3}$$

其中  $G_{i,f}$  代表用户  $i$  与用户  $f$  的轨迹相似度，我们认为就近签到时应该更多地考虑轨迹相同的用户做推荐，外出时则考虑更多的品味相同的用户做推荐；因此上式中第二行，我们同时考虑了用户对分类的兴趣与用户之间的轨迹

$P_{j,k}$  代表物品  $j$  与物品  $k$  的价位相似度，我们认为就近签到时消费价格相对比较稳定，foursquare 中有价位打分，分为 1 4 分，可以利用这个计算消费相似度。

在上述两种推荐系统中引入对时间的考虑，考虑不同时间段用户对 center 的选择不同，还有不同时间段用户对 catalog 的选择不同。

t 如何分段根据具体数据分析得到。

外出 check-in 行为:

$$E_c^t = \frac{\lambda_c}{2} \sum_{i=1}^N \sum_{c=1}^G I_{i,c}^t (R_{i,c}^t - U_i^{t'} L_c^t)^2 + \frac{\lambda_u}{2} \|U^{t'}\|_F^2 + \frac{\lambda_l}{2} \|L\|_F^2 \quad (4)$$

$$P_{r,c,t} \propto l_{i,c} \times U_i^{t'} L_c^T \quad (5)$$

$$\begin{aligned} E^t &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M I_{i,j} (R_{i,j} - P_{i,c_j,t} U_i^T V_j)^2 + \frac{\lambda_u}{2} \|U\|_F^2 + \frac{\lambda_v}{2} \|V\|_F^2 \\ &+ \frac{\lambda_k}{2} \sum_{i=1}^N \sum_{k=1}^H I_{i,k} Z_{i,k,t} \sum_{f \in N_{i,k}} S_{i,f} \|U_i - U_f\|_F^2 \end{aligned} \quad (6)$$

就近 check-in 行为:

$$\begin{aligned} E &= \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^M I_{i,j} (R_{i,j} - U_i^T V_j)^2 + \frac{\lambda_u}{2} \|U\|_F^2 + \frac{\lambda_v}{2} \|V\|_F^2 \\ &+ \frac{\lambda_k}{2} \sum_{i=1}^N \sum_{k=1}^H I_{i,k} Z_{i,k,t} \sum_{f \in N_{i,k}} G_{i,f} \|U_i - U_f\|_F^2 \\ &+ \frac{\lambda_d}{2} \sum_{k \in M} P_{j,k} \|V_i - V_k\|_F^2 \\ E &= E \times l_{i,j} \end{aligned} \quad (7)$$