Social Recommendation with Optimal Limited Attention 阅读报告

背景

作为一个热门的研究课题，推荐系统无疑受到了学术界和业界的广泛关注。但传统的推荐系统存在 *数据稀疏性* 和 *冷启动* 的问题，导致模型的性能下降。于是，**社会化推荐(Social Recommendation)**就被提出，并通过用户之间的社会关系(如好友)来缓解上述两个问题。

然而，大多数现有的方法都不考虑注意力因素造成的限制，即由于人脑的注意力有限**(Limited Attention)**，人们只能接受有限数量的信息，这已经被社会科学视为人类内在的生理特性。事实上，在社交网络中人们很容易就能建立联系，以至于一些在线下比较陌生的人，在线上也能成为好友。这意味着用户之间的社会关系中存在许多噪声和无用的信息。

因此，作者认为对于单个用户的社会化推荐中，不应考虑其所有好友对其的影响。而应该为每个用户选择最优好友子集，使这些好友的偏好能够最好地影响目标用户。

主要工作

- 作者将机器学习技术与社会科学概念相结合，并形式化了社会化推荐中的最优有限关注问题。
- 作者提出了一种新的算法，能够有效地为每个目标用户选择一组社会联系（好友），使得这些被选中好友的偏好最能影响目标用户。然后学习目标用户对这些被选中好友的最优关注(optimal attentions)。

相关研究

- Kang 等人首先提出在推荐系统中考虑 limited attention，但他们只是在所有的社会关系上简单地增加非零权重，无法模拟现实世界的场景，即注意力有限的人只考虑少数朋友的信息。
- 此外，现有的工作中大多使用皮尔森相关系数来计算用户之间的相似性。作者举例说明了这种方法的缺陷。

全局符号

Notations	Description
$\mathbf{R}$	Rating matrix
$\mathbf{U}$	User latent features matrix
$\mathbf{V}$	Item latent features matrix
M	The number of users
N	The number of items
$F(i)$	The set of user i's friends
d	$\mathbf{U} \in \mathbb{R}^{d\times M}$, $\mathbf{V} \in \mathbb{R}^{d\times N}$

问题定义：Optimal Limited Attention（OLA)

作者将OLA问题定义为：

- 输入：
    1. 一组用户
    2. 用户之间的社交连接信息
    3. 一组项目
    4. 用户-项目评级子集
- 输出（对于每个目标用户）：
    1. 一个最优好友子集，使得这个子集中的好友的偏好可以最好地影响目标用户
    2. 针对最优好友子集中的每一个用户，学习目标用户对其的最佳的注意力/注意程度(optimal attention)

算法：OLA-Rec(Social Recommendation with Optimal Limited Attention)

首先，为每一个目标用户 i 引入一个向量 $\phi_i \in \mathbb{R}^{d\times 1}$，视为用户 i 的好友对其偏好产生的影响的聚合：

$$\phi_i = \sum_{u\in F(i)} \alpha_{iu}\mathbf{U}_u \tag{1}$$

其中，α_{iu} 为 i 对 u 的注意力，且满足$\alpha_{iu} \geq 0, \sum_{u=1}^{|F(i)|} \alpha_{iu} = 1$. 这里 (1) 应该是指最优的情况，因为作者接着提出了最小化绝对值：

$$\begin{aligned} &\min_{\alpha_i} \left| \sum_{u\in F(i)} \alpha_{iu}\mathbf{U}_u - \phi_i \right| \\ &s.t.\ \ \alpha_{iu} \geq 0, \\ &\sum_{u=1}^{|F(i)|} \alpha_{iu} = 1 \end{aligned} \tag{2}$$

然后作者通过变换找到了 (2) 的上界：

$$\begin{aligned} &\left| \sum_{u\in F(i)} \alpha_{iu}\mathbf{U}_u - \phi_i \right| \leq C \parallel \alpha_\mathrm{i} \parallel_2 + L \sum_{u\in F(i)} \alpha_{iu} d(U_u, U_i) \\ \Leftrightarrow &\left| \sum_{u\in F(i)} \alpha_{iu}\mathbf{U}_u - \phi_i \right| \leq C \left(\parallel \alpha_\mathrm{i} \parallel_2 + \alpha_i^T \beta_i \right) \end{aligned} \tag{3}$$

其中 C 和 L 为常数，$\beta_i \in \mathbb{R}^{|F(i)|}$，且：

$$\beta_{iu} = L \cdot d(\mathbf{U}_u, \mathbf{U}_i)/C$$

$d(\cdot\,,\,\cdot)$为欧氏距离，$L_c = \frac{L}{C}$为超参。**同时规定** $u \in F(i)$ **按照** $d(\mathbf{U}_u, \mathbf{U}_i)$ **升序排列**，即与 i 离得近的 u 排在前面。

于是最小化 (2) 变为最小化它的上界：

$$\begin{aligned} &\min_{\alpha_\mathrm{i}} C \left(\parallel \alpha_\mathrm{i} \parallel_2 + \alpha_i^T \beta_i \right) \\ s.t.\ \ &\alpha_{iu} \geq 0 \\ &\sum_{u=1}^{|F(i)|} \alpha_{iu} = 1 \end{aligned} \tag{4}$$

应用拉格朗日乘数法：

$$L(\alpha_i, \lambda, \theta) = \parallel \alpha_i \parallel_2 + \alpha_i^T \beta_\mathrm{i} + \lambda(1 - \sum_{u=1}^{|F(i)|} \alpha_{iu}) - \sum_{u=1}^{|F(i)|} \theta_{iu}\alpha_{iu} \tag{5}$$

对 α_i 求偏导数，并置为0：

$$\frac{\partial L}{\partial \alpha_{iu}} = \alpha_{iu} - \parallel \alpha \parallel_2 \times (\lambda - \beta_{iu} + \theta_{iu}) = 0$$

$$\frac{\alpha_{iu}}{\|\alpha_i\|_2} = \lambda - \beta_{iu} + \theta_{iu}$$

根据KKT条件： $\forall \alpha_{iu} > 0, \theta_{iu} = 0($即 $\beta_{iu} < \lambda)$ 以及 $\forall \alpha_{iu} > 0, \theta_{iu} \geq 0($即 $\beta_{iu} \geq \lambda)$。

因此对任意的最优解 $\alpha^*_{iu} > 0$:

$$\frac{\alpha^*_{iu}}{\|\alpha^*_{\mathbf{i}}\|_2} = \lambda - \beta_{iu} \tag{6}$$

考虑到 $\sum_{u=1}^{|F(i)|} \alpha_{iu} = 1$，任意 $\alpha^*_{iu} > 0$可以如下计算:

$$\alpha_{iu} = \frac{\lambda - \beta_{iu}}{\sum_{\alpha_{iu}>0}(\lambda - \beta_{iu})} \tag{7}$$

由于 $\beta_{\mathbf{i}} = (\beta_{i1}, \ldots, \beta_{iu}, \ldots, \beta_{i|F(i)|})^T$ 是按照 $d(U_u, U_i)$**升序**排列的，故α_{iu}是**降序**排列的，因此:
$\exists\, k^*_i$，$1 \leq k^*_i \leq |F(i)|$ 使得 $\forall u > k^*_i, \alpha^*_{iu} = 0$ 且 $\forall u \leq k^*_i, \alpha^*_{iu} > 0$. 这里的$k^*_i$就是用户 i 最优好友子集的基数。

- 计算 α_i

  (6)的两边平方和:

$$\sum_{\alpha^*_{iu}>0} \frac{(\alpha^*_{iu})^2}{\|\alpha^*_{\mathbf{i}}\|_2^2} = \sum_{\alpha^*_{iu}>0} (\lambda - \beta_{iu})^2 = 1 \tag{8}$$

  将(8)展开得:

$$k^*_i \lambda^2 - 2\lambda \sum_{u=1}^{k^*_i} \beta_{iu} + (\sum_{u=1}^{k^*_i} \beta^2_{iu} - 1) = 0 \tag{9}$$

$$\lambda = \frac{1}{k^*_i}\left(\sum_{u=1}^{k_i*} \beta_{iu} + \sqrt{k^*_i + \left(\sum_{u=1}^{k_i*} \beta_{iu}\right)^2 - k^*_i \sum_{u=1}^{k^*_i} \beta^*_{iu}}\right) \tag{10}$$

  因此，给定k^*_i，可以将 (10)中的λ代入(7)得到α^*_i .

- 计算 k^*_i

  作者提出了一种简单的算法计算k^*_i，即从 1 开始升序遍历k_i，直到直到满足条件的最大的k_i为止:

Algorithm 1 Optimal Limited Attention

Require: target user i, i's social connections (friends) $u \in F(i)$, $\beta_{iu} \in \mathbb{R}$ sorted in ascending order.
Initialization: $\lambda_0 = \beta_{i,1} + 1$, $k = 0$
while $\lambda_k > \beta_{i,k+1}$ and $k \leq |F(i)|$ **do**
  $k \leftarrow k + 1$

$$\lambda_k = \frac{1}{k}\left(\sum_{u=1}^{k} \beta_{iu} + \sqrt{\left(k + \left(\sum_{u=1}^{k} \beta_{iu}\right)^2 - k \sum_{u=1}^{k} \beta^2_{iu}\right)}\right).$$

end while
return k and $\boldsymbol{\alpha}_i$ (whose u-th element $\alpha_{iu} = \frac{\lambda_k - \beta_{iu}}{\sum_{\alpha_{iu}>0}(\lambda_k - \beta_{iu})}$);

（不过这里有一个问题就是，作者并没有证明 λ_k是随k递增的，那么最后一次循环的结果难道不会出现 $\lambda_k < \lambda_{k-1}$的情况吗？）

- Objective Function 和 EM style optimization strategy

  将所有的优化目标组合在一起得到损失函数:

$$
\begin{aligned}
\mathcal{L} \quad = \quad & \min \Bigg[\frac{1}{2} \sum\nolimits_{i=1}^{M} \sum\nolimits_{j=1}^{N} I_{ij}^{R} \left(R_{ij} - \phi_i^T V_j\right)^2 \\
+ \quad & \frac{\delta_\phi}{2} \sum_{i=1}^{M} \Big(\phi_i - \sum_{u \in F(i)_{k^*}} \alpha_{iu}^{*} U_u\Big)^T \Big(\phi_i - \sum_{u \in F(i)_{k^*}} \alpha_{iu}^{*} U_u\Big) \\
+ \quad & \frac{\delta_\phi}{2} \sum\nolimits_{i=1}^{M} (U_i - \phi_i)^T (U_i - \phi_i) \\
+ \quad & \frac{\delta_U}{2} \sum\nolimits_{i=1}^{M} U_i^T U_i + \frac{\delta_V}{2} \sum\nolimits_{j=1}^{N} V_j^T V_j \Bigg], \qquad (
\end{aligned}
$$

  最后作者提出使用 EM style的优化策略，即先通过算法1计算$\mathbf{k}^*$和α^*，再通过SGD对目标函数进行优化。