Top-N Recommender System via Matrix Completion（AAAI 2016）

Zhao Kang, Chong Peng, Qiang Cheng

摘要

Top-N 推荐系统已被广泛研究无论是在工业界还是学术界。然而，推荐质量远远不能令人满意。在本文中，我们建议一种简单但有前途的算法。我们基于低秩假设填充用户项矩阵，同时保持原始信息。为此，采用了非凸秩松弛而非核规范来提供设计了更好的秩逼近和有效的优化策略。该方法提高了 Top-N 的推荐提升到一个新的水平

关键词：推荐; 协同过滤

1 引言

随着网上商城的爆发式增长，现在一个用户想要找到自己感兴趣并且有用的商品已经变成了非常难的事情。Top-N 推荐系统目前已经普遍应用在各大电子商务网站平台上，用于给用户推荐他们可能喜欢的或者觉得有用的商品。一个准确率高的推荐系统能给电商平台带来巨大的订单成交量。推荐基于用户自己生成的数据，用户的数据分为显示反馈以及隐式反馈和二者同时存在的混合式反馈。在过去的几十年，提出了很多推荐系统的算法。主要分为基于邻居的协同过滤、基于模型的协同过滤以及基于排名的方法。本次论文复现拟通过解决一个非凸优化问题增加 Top-N 推荐系统的准确率。

2 相关工作

2.1 输入数据分类

基于用户行为主要有一下四种

1. 显示反馈：例如用户对物品的 1 到 5 分的数值评分（numerical rating）行为和喜欢/不喜欢的二值评分（binary rating）行为，以及点赞（like）等单值行为
2. 隐式反馈（又称为单类反馈），例如用户对物品的点击（click）、浏览（browse/view/examination）、收藏（collect/favorite）、加入购物车（add□to□cart）和购买（purchase）等行为。
3. 异构反馈，例如同时包含两种或两种以上的显式反馈和/或隐式反馈。
4. 序列反馈，例如在隐式反馈中包含时序信息。

2.2 传统矩阵分解

[1]最早提出了矩阵分解的概念。即将一个矩阵分解成两个矩阵的乘积，使用矩阵中已有的数据学习，预测未知数据。预测公式如下：

$$X = UV^T + b_u + b_i$$

其中 U 表示用户特征矩阵，V 表示物品特征矩阵，b_u, b_i 分别表示用户和物品的 bias。

2.3 评分与排序

Top-N 推荐系统的主要输出形式有两种

1. 评分：根据用户对其他物品的评分，预测用户对未知物品的评分

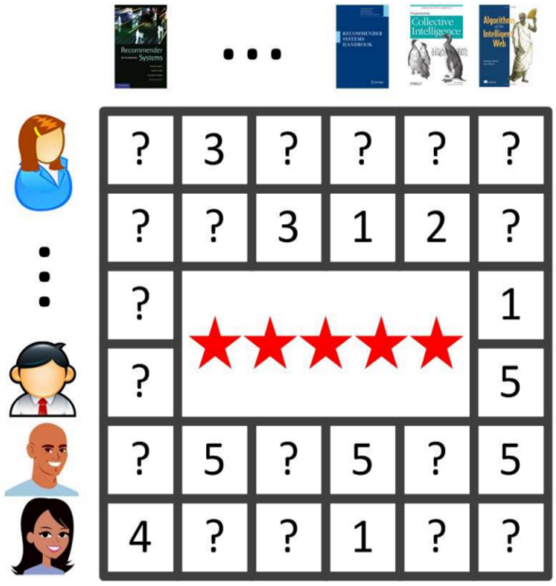

图 1: 输入输出形式

2. 排序：根据用户的数据，对所有未交互物品，选出 N 个该用户最可能喜欢的物品

3 本文方法

3.1 本文方法概述

- Present a matrix completion based method for the Top-N recommendation problem.
- The proposed method recovers the user-item matrix by solving a rank minimization problem.
- To better approximate the rank, a nonconvex function is applied.
- Conduct a comprehensive set of experiments on multiple datasets and compare its performance against that of other state-of-the-art Top-N recommendation algorithms.

3.2 命名

表 1: Notations

Parameters	
m	number of user
n	number of item
$i \in \{1, 2, ..., m\}$	user_ID
$j \in \{1, 2, ..., n\}$	item_ID
......	......
$\Omega = \{(i, j)\}$	(user, item) pairs in data
$M_{ij} \in \{0, 1\}$	indicator variable, $M_{ij} = 1$ if (i,j) $\in \Omega$
$\hat{X} \in \mathbb{R}^{m \times n}$	reference predictor
$Y \in \mathbb{R}^{m \times n}$	auxiliary variable
$Z \in \mathbb{R}^{m \times n}$	Z is a Lagrange multiplier
μ	penalty parameter

3.3 形式化方法

Consider the following matrix completion problem.

$$\min_{X} logdet((X^TX)^{\frac{1}{2}}+I) \quad s.t. \quad X_{ij}=M_{ij},(i,j)\in\Omega \tag{1}$$

This is a nonconvex optimization problem, which is not easy to solve in general. Based on augmented Lagrangian multiplier (ALM) method, introducing an auxiliary variable Y

$$\min_{XY} logdet((X^TX)^{\frac{1}{2}}+I)+l_{R_+}(Y) \quad s.t. \quad X_{ij}=M_{ij}, X=Y \tag{2}$$

where l_{R_+} is the indicator function, defined element-wisely as

$$l_{R_+}=\begin{cases}0, if \quad x>=0;\\ +\infty, otherwise\end{cases}$$

which has an augmented Lagrangian function of the form.

$$\begin{aligned} L(X,Y,Z) =& logdet((X^TX)^{\frac{1}{2}}+I)+l_{R_+}(Y)\\ &+\frac{\mu}{2}\left\|X-Y+\frac{Z}{\mu}\right\|_F^2 \quad s.t. X_{ij}=M_{ij}\end{aligned} \tag{3}$$

where Z is a Lagrange multiplier and μ >0 is a penalty parameter.

Given the current pointX^t,Y^t,Z^t, we update X^{t+1} by

$$\begin{aligned} X^{t+1}=&\arg\min_{X} \quad logdet((X^TX)^{\frac{1}{2}}+I)+l_{R_+}(Y)\\ &+\frac{\mu^t}{2}\left\|X-Y^t+\frac{Z^t}{\mu^t}\right\|_F^2\end{aligned} \tag{4}$$

[2] post a theorem which initially in[3] Let A = $U\sum_A V^T$ be the SVD of A and $\sum_A=diag(\sigma_A)$ Lef F(Z) = f $\circ$ σ(Z) be a unitarilyinvariant function and $\mu>0$,Then an optimal solution to the following problem

$$\min_{Z} F(Z)+\frac{\mu}{2}\left\|Z-A\right\|_F^2 \tag{5}$$

is $Z^*=U\sum_{Z^*}V^T$,where $\sum_{Z^*}=diag(\sigma^*)$ and $\sigma^*=prox_{f,u}(\sigma_A)$ Here $prox_{f,u}(\sigma_A)$ is the Moreau-Yosida operator, defined as

$$prox_{f,u}(\sigma_A):=\arg\min_{\sigma>=0} f(\sigma)+\frac{\mu}{2}\left\|\sigma-\sigma_A\right\|_2^2 \tag{6}$$

according to the first-order optimality condition, the gradient of the objective function of (6) with respect to each singular value should vanish. For logdet function, we have

$$\frac{1}{1+\sigma_i}+\beta(\sigma_i-\sigma_{i,A})=0 \quad s.t. \quad \sigma_i>=0 \tag{7}$$

The above equation is quadratic and gives two roots. If$\frac{\sigma_{i,A}-1}{2}<=0$, the minimizer $\sigma^*=0$; otherwise there exists a unique minimizer $\sigma^*=\frac{\sigma_{i,A}-1}{2}$ by

$$\sigma^2+(1-\sigma_{i,A})\sigma-\sigma_{i,A}+\frac{1}{\mu}=0 \tag{8}$$

Finally, we obtain the update of X variable with

$$X^{t+1}=Udiag(\sigma^*)V^T \tag{9}$$

Then we fix the values at the observed entries and obtain

$$X^{t+1} = where(m > 0, M, X^{t+1}) \tag{10}$$

Then

$$Y^{t+1} = max(X^{t+1} + \frac{Z^t}{\mu^t}, 0) \tag{11}$$

Update the Lagrangian multipliers Z by

$$Z^{t+1} = Z^t + \mu^t (X^{t+1} - Y^{t+1}) \tag{12}$$

Update the parameter μ by

$$\mu^{t+1} = \gamma \mu^t \tag{13}$$

4 复现细节

4.1 与已有开源代码对比

本次复现所有代码全部由笔者编写，没有参考任何相关源代码，复现的论文没有开源的代码库。

4.2 实验环境

本次实验使用的操作系统为 windows10，python3，部分功能调用了百度飞桨平台的 API。编辑工具使用了 PyCharm2021 版本。为了便于操作和调试，代码采用了 jupyternotebook 格式。能够在一个单元框中运行例子。

4.3 算法伪代码

Algorithm 1 Solve (2)

Input: Original imcomplete data matrix$M_\Omega \in \mathbb{R}^{m \times n}$,

parameters $\mu^0 > 0, \gamma > 1$

Initialize: $Y = M, Z = 0$

REPEAT:

1:Obtain X through (10).

2:Update Y as (11).

3:Update the Lagrangian multipliers Z by (12).

4:Update the parameter μ^{t+1} by (13)

UNTIL stopping criterion is met.

4.4 总体框架

代码总体框架包括输入数据的读入模块、模型处理模块以及输出及评测模块。其中模型处理包括了初始化参数、训练以及预测等小模块。模型的输出为一个预测矩阵，测评模块主要有两个评估参数，HR 和 ARHR。

5 实验结果分析

5.1 评估标准

- HR:

$$HR = \frac{\#hits}{\#users} \tag{14}$$

- ARHR:

$$ARHR = \frac{1}{\#users} \sum_{i=1}^{\#hits} \frac{1}{p_i} \tag{15}$$

$\#hits$ 在这里是指成功猜到的用户喜好的数量，而在笔者的观念中，应该理解为正确推荐到的物品数量。但是最后的评测结果还是以论文中的公式进行计算。

5.2 结果展示

	HR	ARHR
第一折	0.32873807	0.12505428
第二折	0.471898197	0.19768554
第三折	0.553552492	0.24501759
第四折	0.587486744	0.27122869
第五折	0.567338282	0.27165665
平均	0.501802757	0.22212855
标准	0.428	0.215

图 2: 评测结果

模型采用了五折交叉验证。最终取平均与论文标准结果进行比较如图 3 所示。

6 总结与展望

6.1 不足

- 论文中很多涉及到数值分析以及线性代数的知识，本人还是不太理解，只能照着流程编程。
- 实验使用了六个数据集，本人只是用了一个。一部分原因是时间关系，还因为根据论文中的数据集官网找到的数据规模与论文中给出的不同，无法作比较。
- 时间紧迫，没有考虑优化相关事情。

6.2 展望

- 应用在多个数据集上，观察算法的鲁棒性。
- 优化算法的时间复杂度，提高预测以及学习效率。
- 考虑是否能够优化模型，提高准确率。
- 将模型在深度学习框架中实现缩短训练时间。

参考文献

[1] SALAKHUTDINOV R, MNIH A. Probabilistic Matrix Factorization[C]//NIPS. 2007.

[2] KANG Z, PENG C, CHENG Q. Robust PCA Via Nonconvex Rank Approximation[C/OL]//2015 IEEE International Conference on Data Mining. IEEE, 2015. https://doi.org/10.1109%2Ficdm.2015.15. DOI: 10.1109/icdm.2015.15.

[3] LEWIS H S, Adrian S.and Sendov. Nonsmooth Analysis of Singular Values. Part I: Theory[J/OL]. Set-Valued Analysis, 2005, 13(3): 213-241. https://doi.org/10.1007/s11228-004-7197-7. DOI: 10.1007/s11228-004-7197-7.