

半监督偏好学习算法

赵敏¹ 刘惊雷¹

摘要 为了解决推荐系统的鲁棒性和重构精度问题,文中提出半监督偏好学习算法,通过偏好学习获得潜在偏好,实现推荐.使用 $l_{2,1}$ 范数作为优化目标函数的正则项,消除噪声和异常点.采用图的拉普拉斯调节整合用户-项目矩阵的单边信息,实现多图融合,提高推荐精度.在 Movielens 10M 数据集和 Netflix 数据集上的实验验证文中算法精度较高、速度较快、鲁棒性较高.

关键词 半监督, 偏好学习, 鲁棒性, 图拉普拉斯调节, $l_{2,1}$ 范数

引用格式 赵敏,刘惊雷.半监督偏好学习算法.模式识别与人工智能,2019,32(10):909-916.

DOI 10.16451/j.cnki.issn1003-6059.201910005

中图法分类号 TP 18

Semi-supervised Preference Learning Algorithm

ZHAO Min¹, LIU Jinglei¹

ABSTRACT To improve low robustness and reconstruction accuracy of recommendation system, a semi-supervised preference learning algorithm is proposed to obtain potential preferences via preference learning and implement recommendations. The $l_{2,1}$ norm is utilized as the regularization of the optimization objective function to eliminate the noises and outliers. The graph Laplacian regularization is employed to integrate the side information of UI matrix to realize multi-image fusion and improve recommendation precision. The experiments on Movielens 10M and Netflix datasets indicate that the proposed algorithm produces high precision, speed and robustness.

Key Words Semi-supervision, Preference Learning, Robustness, Laplace Regularization, $l_{2,1}$ Norm

Citation ZHAO M, LIU J L. Semi-supervised Preference Learning Algorithm. Pattern Recognition and Artificial Intelligence, 2019, 32(10): 909-916.

通过偏好学习,从具有噪声的数据中获取用户潜在偏好是现代机器学习中的重要研究挑战.随着推荐系统在电子商务网站中的应用,逐渐出现准确性、鲁棒性等问题,在整合用户信息时,会遇到一些噪声或异常,导致系统的鲁棒性和准确性降低.

在早期的推荐系统研究中,Mehta 等^[1]提出基于 M 估计的矩阵分解 (M -Estimator Based Matrix Factorization, MMF),通过 M 估计器构造鲁棒的矩阵分解算法. Cheng 等^[2]提出基于最小截平方和 (Least Trimmed Squares, LTS) 估计的矩阵分解 (Least Trimmed Squares Matrix Factorization, LTSMF),是鲁棒的矩阵分解模型,通过 LTS 估计器估计参数. Zhang 等^[3]提出基于最小中位数平方 (Least Median Squares, LMedS) 估计的矩阵分解方法 (LMedS Matrix Factorization, LMedSMF),为基于 MF 的鲁棒协同推荐算法. LMedSMF 引入 LMedS 估计器对具有最大残差的评级进行过滤,使用递归最小二乘 (Recursive Least Squares, RLS) 估计器实现用户特征矩阵和项目特征矩阵的鲁棒估计.

偏好推荐系统在采用损失函数刻画重构精度时,损失函数不能消除异常和噪声,导致推荐系统的鲁棒性和准确性较低.为了解决上述问题,本文提出

收稿日期:2019-05-17;录用日期:2019-09-02

Manuscript received May 17, 2019;

accepted September 2, 2019

国家自然科学基金项目 (No. 61572419, 61773331, 61703360) 资助

Supported by National Natural Science Foundation of China (No. 61572419, 61773331, 61703360)

本文责任编辑 陈松灿

Recommended by Associate Editor CHEN Songcan

1. 烟台大学 计算机与控制工程学院 烟台 264005

1. School of Computer and Control Engineering, Yantai University, Yantai 264005

具有鲁棒性的半监督偏好学习 (Semi-supervised Preference Learning, SSPL) 算法. 通过偏好学习获得潜在偏好, 实现推荐. 使用 $l_{2,1}$ 范数作为优化目标函数的正则项, 消除噪声和异常点. 采用图的拉普拉斯调节^[4-5] 整合用户-项目 (User-Item, UI) 矩阵的单边信息^[6], 实现多图融合, 提高推荐精度. 在 Movielens 10M 数据集和 Netflix 数据集上的实验验证文中算法精度较高、速度较快、鲁棒性较高.

1 半监督偏好学习的定义及概念

定义 1 偏好学习^[7] 给定矩阵 $X \in \mathbf{R}^{m \times n}$, 将该矩阵进行分解 ($X = WH^T$), 若能找到低秩矩阵 $W \in \mathbf{R}^{m \times k}$ 和 $H \in \mathbf{R}^{n \times k}$, 则能从矩阵中得到潜在偏好或特征. 这种得到潜在偏好特征的行为称为偏好学习.

1.1 $l_{2,1}$ 范数

损失函数的作用是衡量模型预测的好坏, 但是一般的损失函数并不能避免噪声和异常问题^[8], 所以本文使用 $l_{2,1}$ 范数对损失函数进行约束, 进而减少噪声和异常值的影响. 在 $l_{2,1}$ 范数中, 空间维度的距离通过 l_2 范数测量, 而不同数据点的总和和使用 l_1 范数. $l_{2,1}$ 范数对图像重建中的异常值或噪声数据具有鲁棒性^[9]. 矩阵 X 的 $l_{2,1}$ 范数定义为

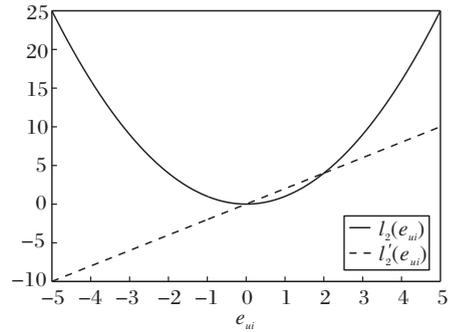
$$\|X\|_{2,1} = \sum_{i=1}^n \sqrt{\sum_{j=1}^m X_{i,j}^2} = \sum_{i=1}^n \|X_{i,:}\|_2.$$

对于任何旋转矩阵 R , 行是旋转不变的, 即

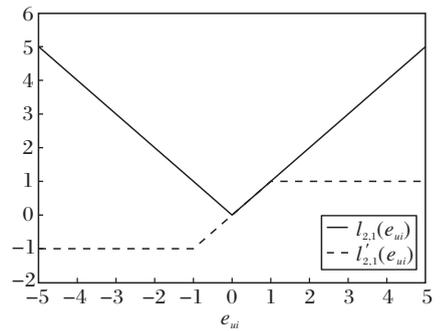
$$\|XR\|_{2,1} = \|X\|_{2,1}.$$

通过基于影响函数的方法可说明 $l_{2,1}$ 范数对于异常值或噪声数据比欧几里德距离 (即 l_2 范数) 更鲁棒. 在鲁棒统计中, 引入影响函数度量估计量的鲁棒性, 影响函数是估计量的一阶导数^[10]. 如果影响函数是有界的, 则异常值对估计量的影响也是有界的, 即估计量是鲁棒的. 这便是矩阵的 $l_{2,1}$ 范数的实际描述过程. 图 1 描绘 l_2 范数和 $l_{2,1}$ 范数损失函数的曲线及其具有不同残差的影响函数曲线.

从图 1 可看出, $l'_{2,1}(e_{ui})$ 是有界的. l_2 范数损失函数的影响函数是无界的 ($l'_2(e_{ui}) = 2e_{ui}$), 因此噪声或异常值对 l_2 范数的影响可能较大. 所以, 在存在异常值或噪声数据时, l_2 范数不鲁棒. $l_{2,1}$ 范数损失函数的影响函数是有界的, 这意味着可以控制噪声或异常值对 $l_{2,1}$ 范数的影响. 因此, $l_{2,1}$ 范数对于噪声或异常值比 l_2 范数更鲁棒. 本文选择 $l_{2,1}$ 范数作为损失函数的正则项. $l_{2,1}$ 范数在每个数据向量内保持特征旋转不变性, 并最小化数据向量之间的异常值影响, 因此本文的目标函数的鲁棒性优于其它函数^[11].



(a) l_2



(b) $l_{2,1}$

图 1 两种范数损失函数及其影响函数

Fig. 1 Two loss functions and their influence functions

1.2 图拉普拉斯调节

图拉普拉斯的结构如下: 顶点是样本数据点, 边权重是数据点间相似度, 由此描述数据点间的相似关系^[12].

定义 2 图的拉普拉斯调节 (L 调节) 通过加入图的先验信息约束矩阵分解的一种半监督调节方式^[13].

L 调节本质是半监督流形调节^[14], 即通过先验信息进行调节. 利用图的辅助信息将高维数据嵌入低维流形中, 当两个样例位于低维流形中的一个小局部邻域内时, 具有相似的类型标签. L 调节形式为 $\text{tr}(W^T L W)$, 通过 L 调节可缩减空间, 提高精度.

2 半监督偏好学习算法

2.1 算法思想

对于用户电影评分矩阵 X , 其中每项 X_{ij} 表示用户 i 对电影 j 的评分, 但实际上矩阵中有一些缺失值, 这些缺失值表示该用户未对这部电影进行评价.

推荐算法的目的就是精确预测这些缺失项的值,然后对这些缺失项进行预测,预测分数最高的项目推荐给目标用户。

本文将数据集中用户作为图的顶点,用户间的相似度作为边的权重,构造数据集的图形式 (V^w, E^w) ^[15]. 假设真实目标矩阵可分解为 $X^* = W^*(H^*)^T$,并且存在图 (V^w, E^w) 的邻接矩阵对 W^* 的 m 行与 H^* 的 n 行的图 (V^h, E^h) 之间的关系进行编码,由图中的边连接的两行(或列)在欧几里德距离中彼此接近,即

$$\frac{1}{2} \sum_{i,j} T_{i,j}^w (w_i - w_j)^2 = \text{tr}(W^T \text{Lap}(T^w) W), \quad (1)$$

$$\begin{aligned} \arg \min_{W, H, H^T H = I} \frac{1}{2} \|X - WH^T\|_{2,1} + \frac{\lambda_L}{2} \text{tr}(W^T \text{Lap}(E^w) W) + \text{tr}(H^T \text{Lap}(E^h) H) + \frac{\lambda_w}{2} \|W\|_F^2 + \lambda_h \|H\|_F^2 = \\ \arg \min_{W, H, H^T H = I} \frac{1}{2} \|X - WH^T\|_{2,1} + \frac{\lambda_L}{2} \text{tr}(W^T L_w W) + \text{tr}(H^T L_h H), \end{aligned} \quad (2)$$

其中

$$L_w = \text{Lap}(E^w) + \lambda_w I_m,$$

L_w 和 L_h 的定义相似. 将正则化参数放入 L_w 、 L_h 的定义中,

$$\|W\|_F^2 = \text{tr}(W^T I_m W).$$

$X - WH^T$ 就是把 X 分解为 $W \in \mathbf{R}^{m \times k}$ 和 $H \in \mathbf{R}^{n \times k}$ ^[17].

使用 W 和 H 的线性组合预测特定用户对特定项目的偏好. 通过式(2)得到特征矩阵 W 、 H 满足约束条件,并且是与实际值偏差最小的结果.

本文加入附加约束 $H^T H = I$ 的作用是保证解的唯一性和降低优化算法的计算成本.

$$L(E, W, G, H, F, C_1, C_2, C_3) =$$

$$\begin{aligned} \|E\|_{2,1} + \text{tr}(H^T L_h H) + \frac{\lambda_L}{2} \text{tr}(W^T L_w W) + \langle C_1, X - WH^T - E \rangle + \langle C_2, W - G \rangle + \langle C_3, H - F \rangle = \\ \|E\|_{2,1} + \text{tr}(H^T L_h H) + \frac{\lambda_L}{2} \text{tr}(W^T L_w W) + \frac{\mu}{2} (\|W - G\|_F^2 + \|X - WH^T - E\|_F^2 + \|H - F\|_F^2), \end{aligned} \quad (3)$$

其中, $\langle G, F \rangle = \text{tr}(G^T F)$, C_1 、 C_2 、 C_3 为拉格朗日乘子, $\mu > 0$ 为惩罚参数. 在每次迭代时沿一个坐标方向最小化增广拉格朗日量,即相对于一个变量的损失最小化,其它变量固定. 共有5个变量,迭代更新方法如下.

初始化. 每次迭代都会更新 μ ,通常从较小的值开始,初始值通常在 $(10^{-5}, 10^{-3})$ 内,具体取决于数据集. C_1 、 C_2 、 C_3 会以目标变量和辅助变量之间的近

其中, $\text{Lap}(T^w) := D^w - T^w$ 为 (V^w, E^w) 的图拉普拉斯矩阵, D^w 为对角矩阵. 将式(1)添加到最小化问题

$$\begin{aligned} \widehat{W}, \widehat{H} = \arg \min_{W, H} \frac{1}{2} \|X - WH^T\|_{2,1}^2 + \frac{\lambda_w}{2} \|W\|_F^2 + \\ \frac{\lambda_h}{2} \|H\|_F^2 \end{aligned}$$

中解出,当 $E_{i,j}^w$ 较大时, $w_i \approx w_j$ 的解. H^* 和相关图拉普拉斯 $\text{Lap}(T^h)$ 与 W^* 和 W^* 相关图拉普拉斯 $\text{Lap}(T^w)$ 类似.

本文不仅希望得到的目标矩阵是低秩的^[16],而且还希望变量 W 、 H 适用于基础图结构,得到如下目标函数:

2.2 算法求解

与其它矩阵分解方法问题相同,式(2)中 W 、 H 是非凸的,很难找到全局最优解^[18].

首先,引入3个辅助变量:

$$E = X - WH^T, \quad G = W, \quad F = H,$$

目标函数(2)改写为

$$\begin{aligned} \arg \min_{E, W, G, H, F} \|E\|_{2,1} + \frac{\lambda_L}{2} \text{tr}(W^T L_w W) + \text{tr}(H^T L_h H), \\ \text{s. t. } E = X - WH^T, G = W, F = H, H^T H = I. \end{aligned}$$

使用不精确的增广拉格朗日乘子法得到如下函数:

似间隙更新,因此为了简单起见,设置为 $\mathbf{0}$ 矩阵.

更新辅助矩阵 E . 固定其它变量更新 E ,改写式(3)为

$$\begin{aligned} L = \arg \min_E \|E\|_{2,1} + \langle C_1^t, X - WH^{tT} - E \rangle = \\ \arg \min_E \|E\|_{2,1} + \frac{\mu^t}{2} \left\| X - WH^{tT} - E + \frac{C_1^t}{\mu^t} \right\|_F^2, \end{aligned} \quad (4)$$

为了解决问题(4),需要如下定理^[19].

定理 1 给定一个矩阵 $A = [a_1, a_2, \dots, a_n] \in \mathbf{R}^{m \times n}$ 和一个正标量 λ , 定义 Q^* 为

$$\min \frac{1}{2} \|Q - A\|_F^2 + \lambda \|Q\|_{2,1}$$

的最优解. Q^* 的第 i 列

$$Q^*(:, i) = \begin{cases} \frac{\|a_i\| - \lambda}{\|a_i\|} a_i, & \lambda < \|a_i\| \\ \mathbf{0}, & \text{其它} \end{cases}$$

式(4) 改写成

$$L = \arg \min_H \text{tr}(H^T L_h F) + \langle C_1^t, X - W^t H^T - E \rangle + \langle C_3, H - F^t \rangle =$$

$$\arg \min_H \text{tr}(H^T L_h F) + \frac{\mu^t}{2} \left(\left\| X - W^t H^T - E^t + \frac{C_1^t}{\mu^t} \right\|_F^2 + \left\| H - F^t + \frac{C_3}{\mu^t} \right\|_F^2 \right),$$

令

$$N = F - \frac{C_3}{\mu} + \frac{1}{\mu} F L_h + \left(X - E + \frac{C_1}{\mu} \right)^T W,$$

可以使用紧凑的形式改写优化问题如下:

$$\arg \min_H \|H - N\|_F^2.$$

根据约束条件, 上式可等价成

$$\arg \min_{H^T H = I} \text{tr}(H^T N). \quad (5)$$

使用定理 2^[19] 解决此问题, 该定理与经典正交

$$L = \arg \min_W \frac{\lambda_L}{2} \text{tr}(W^T L_w G^t) + \langle C_1^t, X - W H^T - E^t \rangle + \langle C_2^t, W - G^t \rangle =$$

$$\arg \min_W \frac{\lambda_L}{2} \text{tr}(W^T L_w G^t) + \frac{\mu^t}{2} \left(\left\| W - G^t + \frac{C_2^t}{\mu^t} \right\|_F^2 + \left\| X - W H^T - E + \frac{C_1}{\mu^t} \right\|_F^2 \right),$$

更新方法与更新 H 类似. 令

$$M = G - \frac{C_2}{\mu} + \frac{\lambda_L}{2\mu} G L_w + \left(X - E + \frac{C_1}{\mu} \right) H^T,$$

得到最优解 W :

$$W^{t+1} = Z^t D^{tT},$$

其中 Z, D 为 M 的 SVD 分解的左右奇异值.

更新辅助矩阵 G . 通过固定其它变量更新 G :

$$L = \arg \min_G \frac{\lambda_L}{2} \text{tr}(W^T L_w G^t) + \langle C_2^t, W^t - G \rangle =$$

$$\arg \min_G \frac{\lambda_L}{2} \text{tr}(W^T L_w G^t) + \frac{\mu^t}{2} \left\| W^t - G + \frac{C_2^t}{\mu^t} \right\|_F^2,$$

扩展上式中的目标函数并删除与 G 无关的项之后, 得

$$\arg \min_E \frac{1}{2} \|E - Y\|_F^2 + \frac{1}{\mu^t} \|E\|_{2,1},$$

$$Y^t = X - W^t H^{tT} - E + \frac{C_1^t}{\mu^t}.$$

根据定理 1, 式(4) 的解如下:

$$E^{t+1}(:, i) = \begin{cases} \frac{\|y_i\| - \lambda}{\|y_i\|} y_i^t, & \frac{1}{\mu} < \|y_i^t\| \\ 0, & \text{其它} \end{cases}$$

其中 y_i^t 为 Y 的第 i 列.

更新电影特征矩阵 H . 固定其它变量更新 H :

Procrustes 问题有关.

定理 2 给定一个方程与式(5) 中的目标函数类似, H 定义为

$$H = UV^T,$$

所以更新 H 的解为

$$H^{t+1} = U^t V^{tT},$$

其中 U, V 为 N 的奇异值分解 (Singular Value Decomposition, SVD) 的左右奇异值.

更新用户特征矩阵 W . 固定其它变量更新 W :

$$\min_G \text{tr} \left(G^T G - 2GW^T - \frac{2}{\mu} GC_2^T + \frac{\lambda_L}{\mu} W^T L_w G \right),$$

通过代数, 得

$$\min_G \left\| G - W + \frac{C_2}{\mu} - \frac{\lambda_L}{2\mu} W^T L_w \right\|_F^2.$$

令

$$P = W + \frac{C_2}{\mu} - \frac{\lambda_L}{2\mu} W^T L_w,$$

得出 G 的解:

$$G_{i,j}^{t+1} = \max(P_{i,j}^t, 0), \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, k.$$

更新辅助矩阵 F . 固定其它变量更新 F :

$$L = \arg \min_F \text{tr}(H^{tT} L_h F) + \langle C_3^t, H^t - F \rangle =$$

$$\arg \min_F \text{tr}(H^{tT} L_h F) + \frac{\mu^t}{2} \left\| H^t - F + \frac{C_3}{\mu^t} \right\|_F^2,$$

更新方式与 G 类似. 令

$$J = H + \frac{C_3}{\mu} - \frac{1}{\mu} H^T L_h,$$

得出 F 的解:

$$F_{i,j}^{t+1} = \max(J_{i,j}^t, 0), \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, k.$$

更新参数 C_1, C_2, C_3, μ :

$$C_1^{t+1} = C_1^t + \mu^t (X - W^t H^{tT} - E^t),$$

$$C_2^{t+1} = C_2^t + \mu^t (W^{tT} - G^t),$$

$$C_3^{t+1} = C_3^t + \mu^t (H^t - F^t),$$

$$\mu^{t+1} = \rho \mu^t,$$

其中, $\rho > 1$ 为控制收敛速度的参数. ρ 越大, 获得收敛所需的迭代次数越少, 但可能会失去最终目标函数值的精度.

2.3 算法步骤

半监督偏好学习方法假设存在少量可以解释用户评级行为的潜在因素(特征). 用户和项目表示为潜在空间中的特征向量, 其中用户项对之间的相似性表示用户对该项目进行评级的趋势. 将用户项目评级矩阵 X 分解为 2 个低秩因子矩阵 W 和 H , 使用它们的线性组合预测特定用户对特定项目的偏好.

算法 1 半监督偏好学习算法

输入 数据集 $X \in \mathbf{R}^{m \times n}$, 最大迭代次数 T

输出 矩阵 X 分解后得到的特征矩阵 W, H

1. 初始化 $t = 0$, 矩阵 W_0, H_0 使用随机值初始化, 初始化 L_w, L_h , 初始化 $\mu = 0.001, \rho = 1.05, \lambda_L = 0.001$, 将 C_1, C_2, C_3 初始化为 $\mathbf{0}$ 矩阵.

2. Repeat

3. 构造 W, H 的图拉普拉斯矩阵.

4. 更新辅助矩阵 E :

$$E^{t+1}(:, i) = \begin{cases} \frac{\|y_i\| - \lambda}{\|y_i\|} y_i^t, & \frac{1}{\mu} < \|y_i^t\| \\ \mathbf{0}, & \text{其它} \end{cases}$$

其中

$$Y = X - W H^T + \frac{C_1}{\mu},$$

y_i 为 Y 的第 i 列.

5. 更新电影特征矩阵 H :

$$H^{t+1} = U^t V^{tT},$$

其中 U, V 为 N 的 SVD 分解的左右奇异值,

$$N = F - \frac{C_3}{\mu} + \frac{1}{\mu} F L_h + \left(X - E + \frac{C_1}{\mu} \right)^T W.$$

6. 更新用户特征矩阵 W :

$$W^{t+1} = Z^t D^{tT},$$

其中 Z, D 为 M 的 SVD 分解的左右奇异值,

$$M = G - \frac{C_2}{\mu} + \frac{\lambda_L}{2\mu} G L_w + \left(X - E + \frac{C_1}{\mu} \right) H^T.$$

7. 更新辅助矩阵 G :

$$G_{i,j}^{t+1} = \max(P_{i,j}^t, 0), \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, k,$$

$$P = W + \frac{C_2}{\mu} - \frac{\lambda_L}{2\mu} W^T L_w.$$

8. 更新辅助矩阵 F :

$$F_{i,j}^{t+1} = \max(J_{i,j}^t, 0), \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, k,$$

$$J = H + \frac{C_3}{\mu} - \frac{1}{\mu} H^T L_h.$$

9. 更新参数 C_1, C_2, C_3, μ :

$$C_1^{t+1} = C_1^t + \mu^t (X - W^t H^{tT} - E^t),$$

$$C_2^{t+1} = C_2^t + \mu^t (W^{tT} - G^t),$$

$$C_3^{t+1} = C_3^t + \mu^t (H^t - F^t),$$

$$\mu^{t+1} = \rho \mu^t.$$

10. $t = t + 1$

11. 直到达到最大收敛次数 T .

Return W, H .

用户 u 对项目 i 的预测评分 \hat{r}_{ui} , 可表示为

$$\hat{r}_{ui} = w_u h_i,$$

其中, w_u 为用户 u 的 k 维特征向量, h_i 为项目 i 的 k 维特征向量^[20].

本文将 SSPL 应用到推荐系统, 得到基于 SSPL 算法的推荐算法, 具体步骤如下.

算法 2 基于 SSPL 的推荐算法

输入 电影评分数据集 $X \in \mathbf{R}^{m \times n}$

输出 用户 u 对电影 i 的前 k 个预测评分 \hat{r}_{ui}

step 1 调用算法 1 得到 W, H .

step 2 根据 W, H 计算用户对电影的预测评分 $\hat{r}_{ui} = w_u h_i$, 降序排列预测评分.

step 3 返回前 k 个预测评分.

2.4 时间复杂度分析

E 的计算涉及 Y 的计算和更新, 时间复杂度分别为 $O(mn + k^3)$ 和 $O(mn)$. 由于 H, L 的稀疏性, F 的时间复杂度为 $O(k^2)$. H 主要计算包括 N 及其 SVD 分解的计算, 时间复杂度分别为 $O(m^3)$ 和 $O(nk^2)$. W 的时间复杂度与 H 相同, 为 $O(m^3 + nk^2)$. G 的时间复杂度为 $O(k^2)$. 所以, 每次迭代的总复杂度为 $O(m^3 + mn)$.

在迭代过程中, 引用一些变量, 并且使用 SVD 分解进行变量的更新, 所以, 算法时间复杂度并未减少, 仍在多项式时间内. 本文算法时间复杂度与一般算法相似, 相比其它算法, 本文算法准确度与鲁棒性更优.

3 实验及结果分析

实验使用计算机系统环境如下:Windows 7 64 bit,主频为 3.2 GHz 的 i5-6500 英特尔处理器,配有 4 GB DDR 4 内存. 程序语言使用 Matlab. 运行环境是 Matlab R2014a.

选择如下两个数据集作为实验数据.

1) MovieLens 10M 数据集. 数据集包含 MovieLens 的 71 567 位用户,共有对 10 681 部电影 的 10 000 054 个评分和 95 580 个标签,评分值范围为 0.5 ~ 5,其中 0.5 为最低(不喜欢),5 为最高(最喜欢).

2) Netflix 数据集. 这是一个真实的数据集,包含 278 858 位用户,共有 17 770 部电影中的 100 480 507 个评分. 评分都是整数,其中 1 和 5 分别表示不喜欢和最喜欢,也是最低分和最高分.

对于每个数据集,本文对 2 000 位用户和 3 000 个项目的子集进行抽样. 选择的 2 000 位用户是社交网络中拥有最多朋友的用户,选择的 3 000 个项目总体评价最高. 表 1 给出数据集的统计数据.

表 1 数据集的统计信息

Table 1 Statistics of datasets

	MovieLens 10M	Netflix
用户	2000	2000
项目	3000	3000
评级	173172	60485
关系	32548	74575
评级密度 / %	2.98	1.00

将每个数据集随机分成两组,80% 数据用作训练集,其余 20% 数据用作测试集. 为了评估半监督偏好学习算法的鲁棒性,将噪声或异常注入训练集.

使用如下评价指标.

1) 推荐准确度. 通过均方根误差 (Root Mean Squared Error, RMSE) 计算,即

$$RMSE = \sqrt{\frac{1}{|U|} \sum_{u \in U, i \in I} |r_{ui} - \hat{r}_{ui}|^2},$$

其中, r_{ui} 为用户 u 对电影 i 的实际评分, \hat{r}_{ui} 为使用推荐算法得到的预测评分.

2) 预测偏移 (Prediction Shift, PS). 测量处理噪声或异常点前后项目 i 的预测偏差. 预测偏移越小,算法的鲁棒性越好. PS 定义如下:

$$PS = \frac{1}{|U|} \sum_{u \in U} |\hat{r}'_{ui} - \hat{r}_{ui}|,$$

其中, \hat{r}_{ui} 、 \hat{r}'_{ui} 分别为处理噪声或异常点前和处理噪声

或异常点后用户 u 的预测评分, U 为推荐系统中的用户集.

为了说明 SSPL 的优势,使用如下对比算法: MMF^[1], LTSMF^[2], LMedSMF^[3], 基于非负矩阵分解 (Non-negative Matrix Factorization, NMF) 的鲁棒协同过滤 (Robust Collaborative Filtering NMF, RCF-NMF)^[20].

为了评估 SSPL 性能,在 MovieLens 10M、Netflix 数据集上进行实验,将 SSPL 与 4 个鲁棒推荐算法进行对比.

图 2 为 5 种算法在 MovieLens 10M 数据集上的 RMSE 对比. 由图可看出,在 MovieLens 10M 数据集上,MMF 的 RMSE 约为 0.87,LTSMF 和 LMedSMF 的 RMSE 约为 0.84,RCF-NMF 的 RMSE 约为 0.83,SSPL 的 RMSE 约为 0.825. 所以,SSPL 的 RMSE 最小,原因是采用图的拉普拉斯调节,通过加入先验信息进行推荐,保证推荐精度. 从图还可看出,随着噪声的变化,SSPL 的 RMSE 值无明显变化且最小,说明 SSPL 是鲁棒的. RCF-NMF 的 RMSE 值也无明显变化,因为其使用 R_1 范数,准确度低于 SSPL. 因此,在 MovieLens 10M 数据集上,SSPL 的推荐精度优于 MMF、LTSMF、LMedSMF、RCF-NMF. 由图可看出,每种算法的 RMSE 值随着噪声的改变都有轻微的浮动,是因为每种算法鲁棒性不同,抑制的噪声也不同,如 MMF 对中度噪声或异常更具鲁棒性, RMSE 值会随噪声的改变有轻微上浮或下降.

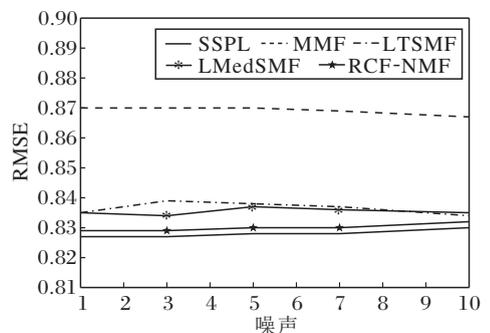


图 2 五种算法在 MovieLens 10M 数据集上的 RMSE 对比
Fig. 2 RMSE comparison of 5 algorithms on MovieLens 10M dataset

图 3 为 5 种算法在 Netflix 数据集上的 RMSE 对比. 由图可看出,MMF 的 RMSE 在 0.905 附近,而 LTSMF 和 LMedSMF 的 RMSE 在 0.89 和 0.895 之间,RCF-NMF 的 RMSE 在 0.88 附近,SSPL 的 RMSE 在 0.877 左右波动. 显然 5 种算法在 Netflix 数据集上的

RMSE 与 MovieLens 10M 数据集相比有一定的增加,说明 Netflix 数据集的稀疏性对 5 种算法的推荐精度有所影响.然而,Netflix 数据集上 SSPL 的 RMSE 仍然最小.在 Netflix 数据集上,SSPL 的推荐精度优于 MMF、LTSMF、LMedSMF、RCF-NMF. SSPL 的 RMSE 值无明显波动,所以仍可证明 SSPL 具有鲁棒性.

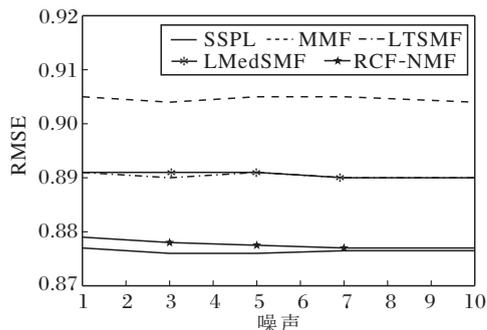


图 3 五种算法在 Netflix 数据集上的 RMSE 对比

Fig. 3 RMSE comparison of 5 algorithms on Netflix dataset

图 4 为 5 种算法在 MovieLens 10M 数据集上的 PS 值对比,由图可知,在 MovieLens 10M 数据集有噪声的情况下,MMF 的 PS 值在 (0.60,1.26) 内,LTSMF 的 PS 值最高,随着噪声的增加在 (0.57,1.32) 内,LMedSMF 的 PS 值在 (0.49,1.17) 内,RCF-NMF 的 PS 值在 (0.43,1.07) 内,SSPL 的 PS 值在 (0.43,1.05) 内.由此可知,SSPL 与 RCF-NMF 的鲁棒性相差不多,但 SSPL 鲁棒性稍优.

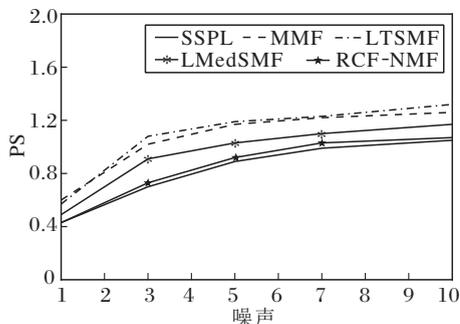


图 4 五种算法在 MovieLens 10M 数据集上的 PS 对比

Fig. 4 PS comparison of 5 algorithms on MovieLens 10M dataset

图 5 为 5 种算法在 Netflix 数据集上的 PS 对比.由图可知,在 Netflix 数据集上,相同的噪声情况下的 SSPL 的 PS 值小于 MMF 和 LTSMF 的 PS 值.这表明 SSPL 在相同噪声环境下的鲁棒性优于 MMF 和 LTSMF.由图中还可看出,SSPL 在某些情况下的 PS

值接近 LMedSMF 和 RCF-NMF,而大多数情况下 PS 值小于 LMedSMF 和 RCF-NMF.这意味着 SSPL 和 LMedSMF、RCF-NMF 在鲁棒性方面无显著差异.因此,相比 MMF 和 LTSMF,SSPL 具有更好的鲁棒性,但在 Netflix 数据集上 SSPL 具有与 LMedSMF、RCF-NMF 相似的鲁棒性.

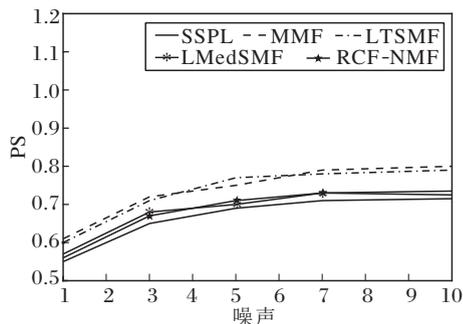


图 5 五种算法在 Netflix 数据集上的 PS 对比

Fig. 5 PS comparison of 5 algorithms on Netflix dataset

4 结束语

随着推荐系统在电子商务网站中的广泛应用,如何确保推荐的可信度和鲁棒性已成为亟待解决的问题.本文提出半监督偏好学习算法,引入 $l_{2,1}$ 范数函数,用于减轻由噪声点和异常值引起的不可靠正则化的影响.使用图的拉普拉斯半监督调节方式,提高偏好学习推荐算法的准确性,通过矩阵分解获取用户的潜在偏好或特征,实现偏好学习.针对方法的优化问题,提出迭代更新算法.实验表明本文算法的有效性鲁棒性.

虽然现在有很多关于推荐系统的研究,但对于推荐系统的鲁棒性研究较少,特别是随着大数据时代的到来,推荐系统面临着大幅增加的数据量,要求推荐算法具有良好的可扩展性以响应评级量表的变化.在本文中只采用电影评分数据进行实验,具有一定的局限性,应考虑推荐系统的扩展性.如何开发高效、稳健的推荐算法将是未来需要进行的工作.

参 考 文 献

[1] MEHTA B, HOFMANN T, NEJDL W. Robust Collaborative Filtering // Proc of the ACM Conference on Recommender Systems. New York, USA: ACM, 2007: 49-56.

[2] CHENG Z P, HURLEY N. Robust Collaborative Recommendation by Least Trimmed Squares Matrix Factorization // Proc of the 22nd IEEE International Conference on Tools with Artificial Intelligence.

- Washington, USA: IEEE, 2010; 105–112.
- [3] ZHANG F Z, SUN S X. A Robust Collaborative Recommendation Algorithm Based on Least Median Squares Estimator. *Journal of Computers*, 2014, 9(2): 308–314.
- [4] DING C H Q, LI T, JORDAN M I. Convex and Semi-nonnegative Matrix Factorizations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(1): 45–55.
- [5] LIU H F, WU Z H, CAI D, *et al.* Constrained Nonnegative Matrix Factorization for Image Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(7): 1299–1311.
- [6] CHIANG K Y, DHILLON I S, HSIEH C J. Using Side Information to Reliably Learn Low-Rank Matrices from Missing and Corrupted Observations. *Journal of Machine Learning Research*, 2018, 19: 1–35.
- [7] 雷恒鑫, 刘惊雷. 基于行列联合选择矩阵分解的偏好特征提取. *模式识别与人工智能*, 2017, 30(3): 279–288.
(LEI H X, LIU J L. Preference Feature Extraction Based on Column Union Row Matrix Decomposition. *Pattern Recognition and Artificial Intelligence*, 2017, 30(3): 279–288.)
- [8] CANDÈS E J, PLAN Y. Matrix Completion with Noise. *Proceedings of the IEEE*, 2010, 98(6): 925–936.
- [9] LI Z C, LIU J, TANG J H. Robust Structured Subspace Learning for Data Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(10): 2085–2098.
- [10] DEBRUYNE M, HUBERT M, SUYKENS J A K. Model Selection in Kernel Based Regression Using the Influence Function. *Journal of Machine Learning Research*, 2008, 9: 2377–2400.
- [11] HUBERT M, ROUSSEEUW P J. ROBPCA: A New Approach to Robust Principal Component Analysis. *Technometrics*, 2005, 47(1): 64–79.
- [12] ZHAO Z, ZHANG L J, HE X F. Expert Finding for Question Answering via Graph Regularized Matrix Completion. *IEEE Transactions on Knowledge and Data Engineering*, 2015, 27(4): 993–1004.
- [13] SHANG F H, JIAO L C, WANG F. Graph Dual Regularization Non-negative Matrix Factorization for Co-clustering. *Pattern Recognition*, 2012, 45(6): 2237–2250.
- [14] HUANG S D, WANG H J, LI T, *et al.* Robust Graph Regularized Nonnegative Matrix Factorization for Clustering. *Data Mining and Knowledge Discovery*, 2018, 32(2): 483–503.
- [15] CHEN Y, JALALI A, SANGHAVI S, *et al.* Clustering Partially Observed Graphs via Convex Optimization. *Journal of Machine Learning Research*, 2014, 15: 2213–2238.
- [16] NIE F P, XU D, TSANG I W. Flexible Manifold Embedding: A Framework for Semi-supervised and Unsupervised Dimension Reduction. *IEEE Transactions on Image Processing*, 2010, 19(7): 1921–1932.
- [17] MACKEY L, TALWALKAR A, JORDAN M I. Distributed Matrix Completion and Robust Factorization. *Journal of Machine Learning Research*, 2015, 16: 913–960.
- [18] CANDÈS E J, RECHT B. Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics*, 2009, 9(6): 717–772.
- [19] HUANG J, NIE F P, HUANG H. Robust Manifold Nonnegative Matrix Factorization. *ACM Transactions on Knowledge Discovery from Data*, 2014, 8(3). DOI: 10.1145/2601434.
- [20] ZHANG F Z, LU Y L, CHEN J M, *et al.* Robust Collaborative Filtering Based on Non-negative Matrix Factorization and R_1 -norm. *Knowledge-Based Systems*, 2017, 118: 177–190.

作者简介



赵敏, 硕士研究生, 主要研究方向为半监督偏好学习的鲁棒性. E-mail: ytdxzhaomin@163.com.

(ZHAO Min, master student. Her research interests include robustness of semi-supervised preference learning.)



刘惊雷(通讯作者), 博士, 教授, 主要研究方向为人工智能、理论计算机科学. E-mail: jinglei_liu@sina.com.

(LIU Jinglei (Corresponding author), Ph. D., professor. His research interests include artificial intelligent and theoretical computer science.)