

结合 L_1 和 L_2 正则化约束的隐语义预测模型研究

王德贤, 何先波, 贺春林, 周 坤, 陈敏治

西华师范大学 计算机学院, 四川 南充 637000

摘 要: 在大数据领域中预测高维稀疏矩阵中的缺失数据, 通常采用随机梯度下降算法构造隐语义模型来对缺失数据进行预测。在随机梯度下降算法来求解模型的过程中经常加入正则化项来提高模型的性能, 由于 L_1 正则化项不可导, 目前在隐语义模型中主要通过加入 L_2 正则化项来构建隐语义模型(SGD_LF)。但因为 L_1 正则化项能提高模型的稀疏性增强模型求解能力, 因此提出一种基于 L_1 和 L_2 正则化约束的隐语义(SPGD_LF)模型。在通过构建目标函数时, 同时引入 L_1 和 L_2 正则化项。由于目标函数满足利普希茨条件, 并通过二阶的泰勒展开对目标函数进行逼近, 构造出随机梯度下降的求解器, 在随机梯度下降求解隐语义模型的过程中通过软阈值来处理 L_1 正则化项所对应的边界优化问题。通过此优化方案, 可以更好地表达目标矩阵中的已知数据在隐语义空间中的特征和对应的所属社区关系, 提高了模型的泛化能力。通过在大型工业数据集上的实验表明, SPGD_LF模型的预测精度、稀疏性和收敛速度等性能都有显著提高。

关键词: 大数据应用; 高维稀疏矩阵; 隐语义

文献标志码: A **中图分类号:** TP391 **doi:** 10.3778/j.issn.1002-8331.1807-0140

王德贤, 何先波, 贺春林, 等. 结合 L_1 和 L_2 正则化约束的隐语义预测模型研究. 计算机工程与应用, 2019, 55(19): 121-127.

WANG Dexian, HE Xianbo, HE Chunlin, et al. Latent factor prediction model combining L_1 and L_2 regularization constraints. Computer Engineering and Applications, 2019, 55(19): 121-127.

Latent Factor Prediction Model Combining L_1 and L_2 Regularization Constraints

WANG Dexian, HE Xianbo, HE Chunlin, ZHOU Kun, CHEN Minzhi

School of Computer Science, China West Normal University, Nanchong, Sichuan 637000 China

Abstract: LF model is usually built by SGD method and it's used to predict the missing data of high-dimensional sparse matrix in big data field. LF model need to integrate regularization terms to improve its performance. Due to L_1 regularization term is non-differentiable, normally integrates L_2 regularization term into an LF model only. However, the L_1 regularization normal can improve the sparsity and solving ability of LF model. To solve the issue, this paper proposes a SPGD_LF model that simultaneously integrates both L_1 and L_2 regularization terms in to an LF model. Since the objective function satisfies the Lipschitz condition and approximates the objective function by second-order Taylor expansion, a solver for stochastic gradient descent is constructed. In the process of stochastic gradient descent, the soft threshold process deals with the boundary optimization problem corresponding to the L_1 regularization term and solves the implicit semantic model. Through this optimization scheme, the characteristics of the known data in the target matrix in the latent factor space and the corresponding community relationship can be better expressed, and the generalization ability of the model is improved. Empirical studies on two datasets from industrial applications and the results show that the prediction accuracy, sparsity and convergence rate of SPGD_LF model are improved significantly.

Key words: big data application; high-dimensional and sparse matrix; latent factor

基金项目: 国家大学生创新创业专项基金(No.201510638047); 四川省教育厅重点项目(No.15ZA048, No.13ZA0015); 南充市科技支撑项目(No.15A0068); 西华师范大学英才基金(No.17YC150)。

作者简介: 王德贤(1993—), 男, 硕士, 研究方向: 大数据智能计算; 何先波(1971—), 男, 博士, 教授, 研究方向: 大数据、嵌入式; 贺春林(1971—), 通讯作者, 男, 教授, 研究方向: 大数据、图像处理, E-mail: chunlin_he@163.com; 周坤(1993—), 男, 讲师, 研究方向: 大数据; 陈敏治(1994—), 女, 硕士, 研究方向: 人工智能。

收稿日期: 2018-07-18 **修回日期:** 2018-12-04 **文章编号:** 1002-8331(2019)19-0121-07

CNKI网络出版: 2019-01-17, <http://kns.cnki.net/kcms/detail/11.2127.tp.20190115.1403.006.html>

1 引言

在大数据领域中,高维稀疏(High Dimensional and Sparse, HiDS)矩阵在各种大数据智能应用相关的研究领域频繁可见,如生物信息、社交网络服务等^[1-3]。在大数据的应用场景中,一些基于学习系统的工业应用通常会涉及到处理大量的实体以及这些实体之间高维的关联关系,如推荐系统中的项目(items)、用户(users)和项目-用户之间的偏好关系。在HiDS矩阵中由于很难观察到数量庞大的实体之间所有关系,因此HiDS矩阵仅描述了实体间部分关系^[3](即高维稀疏矩阵中存在大量缺失值)。隐语义(Latent Factor, LF)分析模型是一种从HiDS矩阵的数据中提取有用信息的有效方法^[4-6],它起源于矩阵分解(Matrix-Factorization, MF)技术,并通过一个LF模型找到一个HiDS矩阵的低秩逼近,但传统的MF模型是处理全矩阵,而本文聚焦于目标矩阵是HiDS的矩阵^[4-6]。

为了从HiDS矩阵中提取到有用信息,并高效地预测HiDS矩阵中的缺失数据,很多有效的机器学习预测方法被提出和研究^[1-11]。其中LF模型通过将HiDS矩阵中的已知实体对应的行和列映射到低维度的LF空间中,基于目标HiDS矩阵中的已知项和在LF空间中对应的隐语义来构造目标函数,最后通过最小化目标函数来得到LFs(Latent Factors)即为隐语义输出模型^[4-6],隐语义模型的建模原理如图1所示。代表性的LF模型包括:奇异值矩阵分解^[4-5]、概率矩阵分解模型^[6]、非参数化贝叶斯LF模型^[12]等。这些LFs可被解释为隐藏在HiDS矩阵中的实体特性,它们在各种HiDS数据分析任务中非常重要,例如缺少数据估计^[2]、移动用户追踪^[12]、社区发现^[13]等,但它们有一些自身的不足如奇异值矩阵分解模型,它是基于矩阵的运算,相对于本文中仅聚焦于目标矩阵的已知数据构建的LF模型来说增加了运算效率,且预测精度较低,概率矩阵分解模型它可以近似看成奇异值矩阵分解模型的拓展,非参数贝叶斯LF模型也会造成精度的损失,这些算法在模型的预测精度和模

型稀疏性表达上都有一定的局限性。

在对高维稀疏矩阵中缺失数据的预测中,因为 L_1 正则化项不可导,所以通常仅采用带有 L_2 正则化项的随机梯度下降(Stochastic Gradient Descent, SGD)算法,即为SGD_LF模型来进行未知数据的预测。但仅仅采用 L_2 范数来约束目标函数,有以下缺点:(1)不能将已知实体及其实体间的关系高效准确地映射到隐语义空间中;(2)由于增加了非相关特征信息的干扰,损失了模型的预测精度;(3)由于处理的目标矩阵数据庞大,非相关的信息的加入大大提高了模型的训练成本;(4)对应的隐语义不进行稀疏限制,不能很好地体现各实体所属的社区关系。为了解决以上问题我们在SGD_LF中加入 L_1 范数, L_1 会对映射到隐语义空间对应的实体特征进行限制,将实体映射到隐语义空间中并将不相关的属性值限制为0,排除非相关信息的干扰,提高模型的计算效率和预测精度,同时更好地体现实体所属关系。

本文立足于SGD算法,在SGD算法基础上进行高效拓展,并有以下创新性:(1)同时引入 L_1 和 L_2 正则化项,提出一种带 L_1 和 L_2 的随机近端梯度下降算法模型(Stochastic Proximal Gradient Descent Latent Factor, SPGD_LF);(2)本文对所提模型进行严格的数学推导;(3)在大规模真实工业数据集上对本文提出的算法模型进行了实验验证。

2 相关工作和问题定义

2.1 问题定义

定义1 令 M 和 N 表示两个实体集合, $R^{[M] \times [N]}$ 表示矩阵 R 。 R 矩阵中的第 m 行 n 列的值 $r_{m,n}$ 来量化两个实体 $m \in M$ 和 $n \in N$ 之间的关系, Δ 和 Γ 分别表示 R 矩阵中已知项集合和未知集合, $|\Delta| \ll |\Gamma|$ 。给定矩阵 R 。因此,一个LF模型被定义如下^[13-15]:

定义2 已知 R 和 Δ ,LF模型得到秩为 d 的一个近似 R 的矩阵 $\hat{R} = PQ^T$ 。 P 和 Q 分别为 $m \times d$ 和 $n \times d$

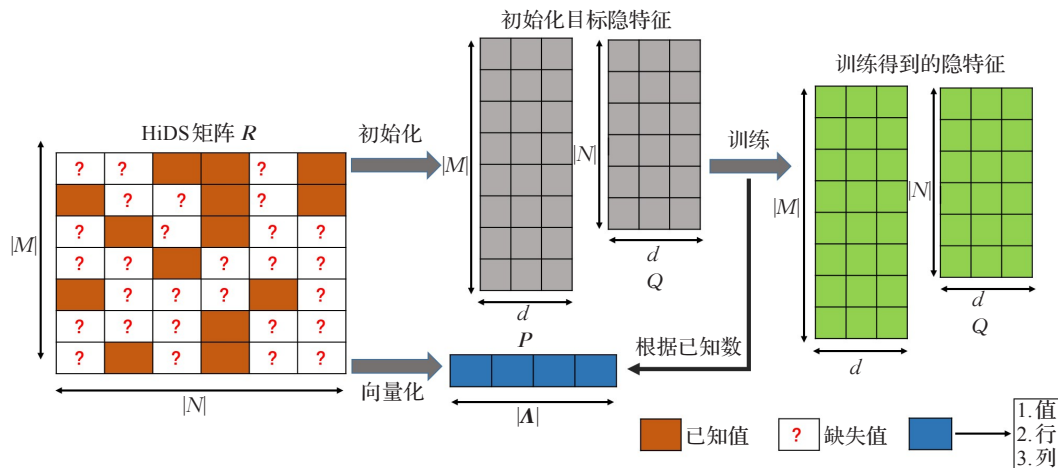


图1 隐语义模型建模过程

的LF矩阵即为隐语义空间。

其中 d 是隐语义空间对应的两个矩阵的维度, P 和 Q 组成LFs且反映了 M 和 N 的相应特征。为了获得 P 和 Q , 需要一个目标函数来衡量 Δ 和 \hat{R} 中的对应项之间的差异。通常采用欧几里得距离公式^[5-8]来构造目标函数,如下所示:

$$\arg \min_{P, Q} \epsilon(P, Q) = \frac{1}{2} \sum_{r_{m,n} \in \Delta} \left(r_{m,n} - \sum_{k=1}^d p_{m,k} q_{n,k} \right)^2 \quad (1)$$

式中, $r_{m,n}$ 表示在 R 中的值, $p_{m,k}$ 和 $q_{n,k}$ 是 P 和 Q 中的对应值。LF模型在分析高维稀疏矩阵时通常需要加入 L_2 正则化项来防止过模型拟合。因此得到:

$$\begin{aligned} \arg \min_{P, Q} \epsilon(P, Q) &= \frac{1}{2} \sum_{r_{m,n} \in \Delta} ((r_{m,n} - \sum_{k=1}^d p_{m,k} q_{n,k})^2 + \\ &\lambda_2 \|p_{m,\cdot}\|_2^2 + \lambda_2 \|q_{n,\cdot}\|_2^2) = \frac{1}{2} \sum_{r_{m,n} \in \Delta} ((r_{m,n} - \sum_{k=1}^d p_{m,k} q_{n,k})^2 + \\ &\lambda_2 \sum_{k=1}^d p_{m,k}^2 + \lambda_2 \sum_{k=1}^d q_{n,k}^2) \end{aligned} \quad (2)$$

式中, $\|\cdot\|_2^2$ 计算对应向量的 L_2 正则化值, p_m 和 q_n , 分别表示在 P 和 Q 中第 m 行和第 n 行向量。 λ_2 为 P 和 Q 正则化系数。

2.2 基于随机梯度下降算法的隐语义模型

之前的研究^[16-17]表明,SGD算法实现简单,计算复杂度低,因此通常采用SGD算法来构建LF模型来分析HiDS矩阵。通过SGD算法来最小化带参数的目标函数,每个训练实例按照梯度的相反方向进行更新。因此可得每个单一训练实例 $r_{m,n} \in \Delta$ 的瞬时损失,如下:

$$\epsilon_{m,n}(p_{m,\cdot}, q_{n,\cdot}) = (r_{m,n} - \hat{r}_{m,n})^2 + \lambda_P \|p_{m,\cdot}\|_2^2 + \lambda_Q \|q_{n,\cdot}\|_2^2 \quad (3)$$

对目标函数求偏导,可以得到SGD算法的训练更新公式:

$$\begin{aligned} \arg \min_{P, Q} \epsilon(P, Q) &\stackrel{\text{SGD}}{\Rightarrow} \forall r_{m,n} \in \Delta, k \in \{1, 2, \dots, d\}: \\ &\begin{cases} p_{m,k} \leftarrow p_{m,k} - \eta \frac{\partial \epsilon_{m,n}}{\partial p_{m,k}} = p_{m,k} + \eta \cdot \text{err}_{m,n} \cdot q_{n,k} - \eta \cdot \lambda_P \cdot p_{m,k} \\ q_{n,k} \leftarrow q_{n,k} - \eta \frac{\partial \epsilon_{m,n}}{\partial q_{n,k}} = q_{n,k} + \eta \cdot \text{err}_{m,n} \cdot p_{m,k} - \eta \cdot \lambda_Q \cdot q_{n,k} \end{cases}, \\ \text{err}_{m,n} &= r_{m,n} - \sum_{k=1}^d p_{m,k} q_{n,k} \end{aligned} \quad (4)$$

式中, η 表示学习速率,取值范围为 $(0, 1]$ 。根据对SGD算法的推理,得到SGD_LF模型,具体描述总结如下所示。

Input: Δ, M, N, d

Operation

Initialize $P^{[M] \times d}, Q^{[N] \times d}$ at random in $(0, 0.5)$

Initialize $\eta, \lambda, i = 1, \text{Maximum_round_count}$

1. while not converge and $i \leq C$ do

2. for each $r_{m,n} \in \Delta$

```

3.    $\hat{r}_{m,n} = \sum_{k=1}^d p_{m,k} q_{n,k}$ 
4.    $\text{err}_{m,n} = r_{m,n} - \hat{r}_{m,n}$ 
5.   for  $k=1$  to  $d$ 
6.      $p_{m,k} = p_{m,k} + \eta \times \text{err}_{m,n} \times q_{n,k} - \eta \times \lambda \times p_{m,k}$ 
7.      $q_{n,k} = q_{n,k} + \eta \times \text{err}_{m,n} \times p_{m,k} - \eta \times \lambda \times q_{n,k}$ 
8.   end for
9.   end for
10.  $i = i + 1$ 
11.end while

```

Output: P, Q

以上描述了SGD_LF模型的执行过程,当在(3)和(4)中的 L_2 正则化项系数为零时,即为不带任何正则化项的标准SGD的算法构建的LF模型。因此,可得到标准的SGD算法更新公式如式下所示:

$$\begin{cases} p_{m,k} \leftarrow p_{m,k} + \eta \cdot \text{err}_{m,n} \cdot q_{n,k} \\ q_{n,k} \leftarrow q_{n,k} + \eta \cdot \text{err}_{m,n} \cdot p_{m,k} \end{cases}, \quad \text{err}_{m,n} = r_{m,n} - \sum_{k=1}^d p_{m,k} q_{n,k} \quad (5)$$

3 基于随机近端梯度下降的LF模型

SPGD^[18-19]是对SGD算法的高效拓展。在SGD_LF模型的目标函数中引入 L_1 正则化项来构造SPGD_LF模型,并根据分析SPGD_LF模型的目标函数的利普希茨条件^[20]和泰勒的二阶展开式得到相应的梯度更新。首先得到拓展的目标函数式如下:

$$\begin{aligned} \arg \min_{P, Q} \epsilon_1(P, Q) &= \\ &\frac{1}{2} \sum_{r_{m,n} \in \Delta} (\epsilon_{m,n}(p_{m,\cdot}, q_{n,\cdot}) + \lambda_1 \|p_{m,\cdot}\|_1 + \lambda_1 \|q_{n,\cdot}\|_1) \end{aligned} \quad (6)$$

式中, $\|\cdot\|_1$ 计算对应向量的 L_1 正则化值,类似SGD算法,可以得到在每个单一训练实例 $r_{m,n} \in \Delta$ 的瞬时损失:

$$\begin{aligned} \arg \min_{p_{m,\cdot}, q_{n,\cdot}} \epsilon_{m,n}(p_{m,\cdot}, q_{n,\cdot}) &= \\ &\frac{1}{2} (\epsilon_{m,n}(p_{m,\cdot}, q_{n,\cdot}) + \lambda_1 \|p_{m,\cdot}\|_1 + \lambda_1 \|q_{n,\cdot}\|_1) \end{aligned} \quad (7)$$

$\epsilon_{m,n}(p_{m,\cdot}, q_{n,\cdot})$ 函数是一个关于 p_m 和 q_n 的一个双凸函数,因此,固定 p_m (或 q_n) 对应 P (或 Q) 中的LFs, 则 $\epsilon_{m,n}(p_{m,\cdot}, q_{n,\cdot})$ 就关于 $q_{n,k}$ (或 $p_{m,k}$), $\forall k \in \{1, 2, \dots, d\}$ 的凸函数。例如,在第 t 次迭代的时候,固定在 P 中除了 $p_{m,k}$ 的LFs,表达式如下:

$$\begin{cases} \forall j \in \{1, 2, \dots, d\} - \{k\}: p_{m,j} = p_{m,j}^{(t)} \\ \forall j \in \{1, 2, \dots, d\}: q_{n,j} = q_{n,j}^{(t)} \end{cases} \quad (8)$$

这里的 $p_{m,j}^{(t)}$ 和 $q_{n,j}^{(t)}$ 表示 $p_{m,j}$ 和 $q_{n,j}$ 第 t 次迭代后的状态值。对于式(7)可写为如下形式:

$$\arg \min_{p_{m,k}} \epsilon_{m,n}(p_{m,k}) = \frac{1}{2} (\epsilon_{m,n}(p_{m,k}) + \lambda_1 \|p_{m,k}\|_1) \quad (9)$$

$$\varepsilon_{m,n}(p_{m,k}) = \left(r_{m,n} - \sum_{k=1}^d p_{m,k} q_{n,k} \right)^2 + \lambda_2 \|p_{m,k}\|_2^2 \quad (10)$$

由于 L_1 正则化项是不可微的, 因此根据式(10)的利普希茨条件, 可得:

$$\|\nabla \varepsilon_{m,n}(p'_{m,k}) - \nabla \varepsilon_{m,n}(p_{m,k})\|_2^2 \leq L \|p'_{m,k} - p_{m,k}\|_2^2, \forall p'_{m,k}, p_{m,k} \quad (11)$$

式中, L 代表为利普希茨系数, $p'_{m,k}$ 表示 $p_{m,k}$ 的一个驻点。令 $p'_{m,k} = p_{m,k}^{(t)}$, 即 $p'_{m,k}$ 表示 $p_{m,k}$ 第 t 次迭代后的状态值。式(11)在 $p_{m,k}^{(t)}$ 处进行二阶的泰勒展开, 可得:

$$\begin{aligned} \hat{\varepsilon}_{m,n}(p_{m,k}) &\cong \varepsilon_{m,n}(p_{m,k}^{(t)}) + \\ &\frac{\partial \varepsilon_{m,n}(p_{m,k}^{(t)})}{\partial p_{m,k}} (p_{m,k} - p_{m,k}^{(t)}) + \frac{L}{2} (p_{m,k} - p_{m,k}^{(t)})^2 = \\ &\frac{L}{2} \left(p_{m,k} - p_{m,k}^{(t)} + \frac{1}{L} \frac{\partial \varepsilon_{m,n}(p_{m,k}^{(t)})}{\partial p_{m,k}} \right)^2 + \varphi(p_{m,k}^{(t)}) \quad (12) \end{aligned}$$

$$\varphi(p_{m,k}^{(t)}) = \varepsilon_{m,n}(p_{m,k}^{(t)}) - \frac{1}{2L} \left(\frac{\partial \varepsilon_{m,n}(p_{m,k}^{(t)})}{\partial p_{m,k}} \right)^2 \quad (13)$$

式(12)和式(13)中的决策参数 $p_{m,k}$ 是无关, 因此在 $\hat{\varepsilon}_{m,n}(p_{m,k})$ 的最小化过程中 $\varphi(p_{m,k}^{(t)})$ 可视为常数, 可得到如下推论:

$$\hat{\varepsilon}_{m,n}(p_{m,k}) \cong \frac{L}{2} \left(p_{m,k} - p_{m,k}^{(t)} + \frac{1}{L} \frac{\partial \varepsilon_{m,n}(p_{m,k}^{(t)})}{\partial p_{m,k}} \right)^2 \quad (14)$$

$$\hat{\varepsilon}_{m,n}(p_{m,k}) = 0 \Rightarrow p_{m,k} = p_{m,k}^{(t)} - \frac{1}{L} \frac{\partial \varepsilon_{m,n}(p_{m,k}^{(t)})}{\partial p_{m,k}} \quad (15)$$

式(15)为梯度下降的更新形式, 可对其进行梯度下降, 可求得函数最小值时。其中 $1/L$ 等同于式(4)中的学习率 η 。令 $z^{[M] \times d}$ 为对应 P 矩阵的一个辅助矩阵且

$$z_{m,k} = p_{m,k}^{(t)} - \frac{1}{L} \frac{\partial \varepsilon_{m,n}(p_{m,k}^{(t)})}{\partial p_{m,k}}, \text{ 结合式(9)和(15)可得:}$$

$$\arg \min_{p_{m,k}} \varepsilon_{m,n}(p_{m,k}) = \frac{1}{2} \left(\frac{L}{2} (p_{m,k} - z_{m,k})^2 + \lambda_1 \|p_{m,k}\|_1 \right) \quad (16)$$

对式(16)进行软阈值处理, 可以得到关于 $p_{m,k}$ 的更新公式如下:

$$\begin{aligned} p_{m,k} &\leftarrow \begin{cases} z_{m,k} - \lambda_1/L, & \lambda_1/L < z_{m,k} \\ 0, & |z_{m,k}| \leq \lambda_1/L \\ z_{m,k} + \lambda_1/L, & z_{m,k} < -\lambda_1/L \end{cases}, \\ z_{m,k} &= p_{m,k}^{(t)} + \frac{1}{L} \text{err}_{m,n}^{(t)} \cdot q_{n,k} - \frac{\lambda_2}{L} p_{m,k}^{(t)}. \quad (17) \end{aligned}$$

同理, 令 $y^{[N] \times d}$ 为对应 Q 矩阵的一个辅助矩阵且

$$y_{n,k} = q_{n,k}^{(t)} + \frac{1}{L} \frac{\partial \varepsilon_{m,n}(p_{m,k}^{(t)})}{\partial q_{n,k}}, \text{ 得到关于 } q_{n,k} \text{ 的相关更新公}$$

式如下:

$$\begin{aligned} q_{n,k} &\leftarrow \begin{cases} y_{n,k} - \lambda_1/L, & \lambda_1/L < y_{n,k} \\ 0, & |y_{n,k}| \leq \lambda_1/L \\ y_{n,k} + \lambda_1/L, & y_{n,k} < -\lambda_1/L \end{cases}, \\ y_{n,k} &= q_{n,k}^{(t)} + \frac{1}{L} \text{err}_{m,n}^{(t)} \cdot p_{m,k}^{(t)} - \frac{\lambda_2}{L} q_{n,k}^{(t)}. \quad (18) \end{aligned}$$

对比 SGP_LF 模型和 SPGD_LF 模型的更新公式(5)和式(17)、(18), 其中很明显的可以得到 SPGD_LF 模型增加了边界优化策略, 如式(17)当更新值为 $|z_{m,k}| \leq \lambda_1/L$ 边界条件时, 通过软阈值处理时数学推导得出在该条件范围时, 更新值实际应为零, 但在 SGP_LF 中的更新公式并未对其进行限制, 引入了很多非相关的特征数据, 对模型精度和模型的训练都带来了损失; 在式(17)中模型对不同的实体对应的参数值具有不同的更新策略, 对不同实体对应的不同参数进行不同策略的更新, 能更好地适应不同的数据情况, 提高模型的应用能力。因此, 可以得到 SPGD_LF 模型能更好地体现出模型的稀疏性并能很好地提高模型的泛化能力。

根据 SPGD 算法推理, 可得 SPGD_LF 模型, 如下所示。

Input: Δ, M, N, d

Operation

Initialize $P^{[M] \times d}, Q^{[N] \times d}$ at random in $(0, 0.05)$

Initialize $L, \lambda, \lambda_1, i=1, C=\text{Maximum_round_count}$

1. while not converge and $i \leq C$ do

2. for each $r_{m,n} \in \Delta$

3. $\hat{r}_{m,n} = \sum_{k=1}^d p_{m,k} q_{n,k}$

4. $\text{err}_{m,n} = r_{m,n} - \hat{r}_{m,n}$

5. for $k=1$ to d

6. $p_{m,k} = p_{m,k} + 1/L \times \text{err}_{m,n} \times q_{n,k} - 1/L \times \lambda \times p_{m,k}$

7. if $\lambda_1/L < p_{m,k}$ then $p_{m,k} = p_{m,k} - \lambda_1/L$

8. else if $-\lambda_1/L < p_{m,k} < \lambda_1/L$ then $p_{m,k} = 0$

9. else if $p_{m,k} < -\lambda_1/L$ then $p_{m,k} = p_{m,k} + \lambda_1/L$

10. end if

11. $q_{n,k} = q_{n,k} + 1/L \times \text{err}_{m,n} \times p_{m,k} - 1/L \times \lambda \times q_{n,k}$

12. if $\lambda_1/L < q_{n,k}$ then $q_{n,k} = q_{n,k} - \lambda_1/L$

13. else if $-\lambda_1/L < q_{n,k} < \lambda_1/L$ then $q_{n,k} = 0$

14. else if $q_{n,k} < -\lambda_1/L$ then $q_{n,k} = q_{n,k} + \lambda_1/L$

15. end if

16. end for

17. end for

18. $i = i + 1$

19. end while

Output: P, Q

4 实验结果及分析

实验数据集:本文中实验在两个数据集进行,它们都是来自工业应用中的真实数据。

数据集1 MovieLens 20M数据集(简称D1),由明尼苏达大学通过MovieLens网站收集,包含了138 493个用户对26 744个项目的评分信息,但是其中20 000 263已知评分项,该数据集的稀疏性为0.54%,其稀疏性计算如下:

稀疏性=(矩阵中已知数量/矩阵总数) $\times 100\%$

数据集2 Douban 数据集(简称D2),是用于测试推荐预测算法模型性能的重要数据集,共有129 489用户对58 540个项目评分,其中已知评分数据项为1 700万条,该数据集的稀疏性为0.22%。

实验评判度量:由于本文是针对HiDS矩阵缺失数据的预测,因此用预测数据作为模型性能的验证。采取均方根误差(Root Mean Square Error, RMSE)来衡量HiDS矩阵中缺失数据的预测精度。令 Ω 为测试数据集,它和训练数据集中的数据不重合:

$$RMSE = \sqrt{\left(\sum_{r_{u,v}} (r_{u,v} - \hat{r}_{u,v})^2 \right) / |\Omega|} \quad (19)$$

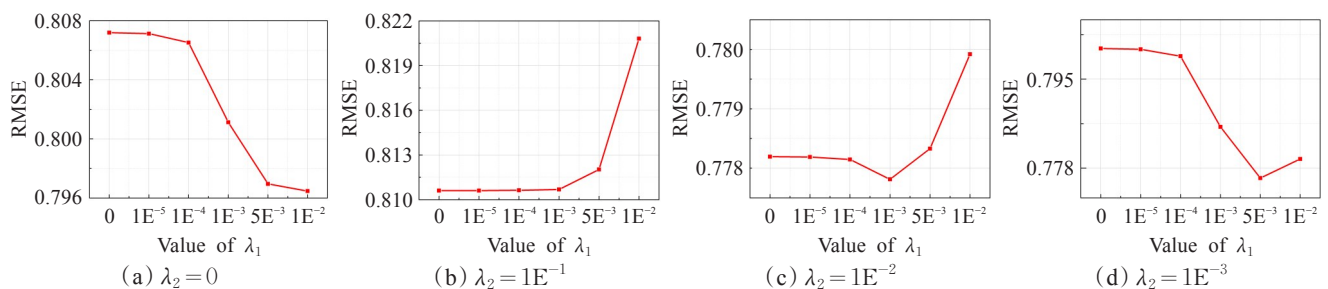


图2 D1数据集上的RMSE

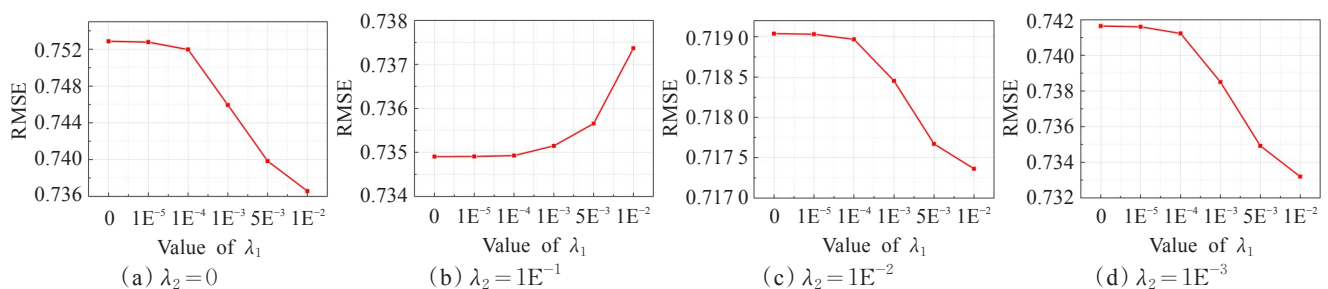


图3 D2数据集上的RMSE

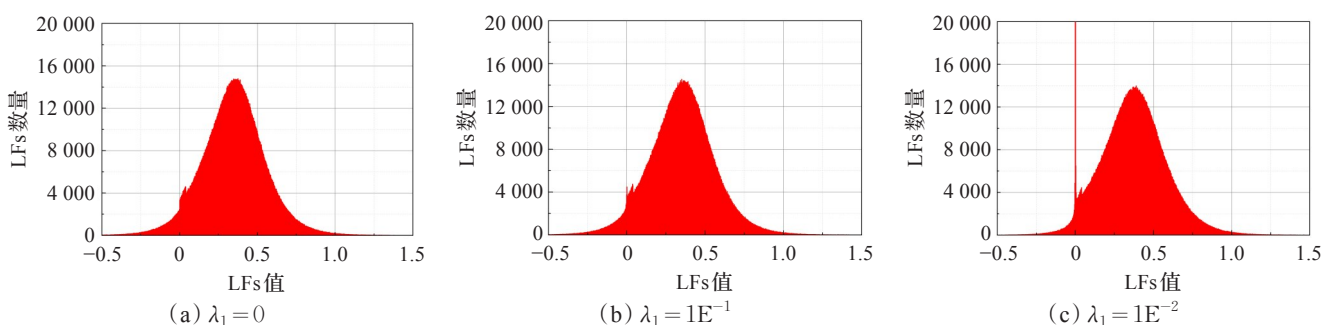


图4 D1上的LFs值分布

$\hat{r}_{u,v}$ 表示测试实例 $r_{u,v} \in \Omega$ 的预测。RMSE值越高,表示预测精度越低,模型性能越差;相反,值越低表示预测精度越高,模型性能越好。

本文实验中隐语义空间维度为 $d=20$,且隐语义空间对应的矩阵都采用相同的初始化状态值。实验采用了20%~80%训练验证设置,用5折交叉验证来评估每个对比模型的性能。将数据集分为5部分,随机选取其中的1部分作为测试数据集,剩余的4部分作为训练数据集。训练过程的停止条件设置为:(1)训练的迭代次数达到1 000轮;(2)相邻两次迭代得到的精度差小于 1.0×10^{-5} 。

设置SGD_LF算法模型和本文提出的SPGD_LF算法模型进行性能比较,实验结果如下所示:

图2表示D1数据集上的实验结果,图3表示D2数据集上的实验结果,图4和图5分别为D1和D2上的LFs的分布图,图6为在不同参数条件下SGD_LF和SPGD_LF模型的收敛迭代比较。

首先,在图2和图3分别给出了SPGD_LF模型在D1和D2数据集上正则化系数 λ_1 和 λ_2 网格搜索训练实验效果图(η 均设置为0.001),得到如下的结论:

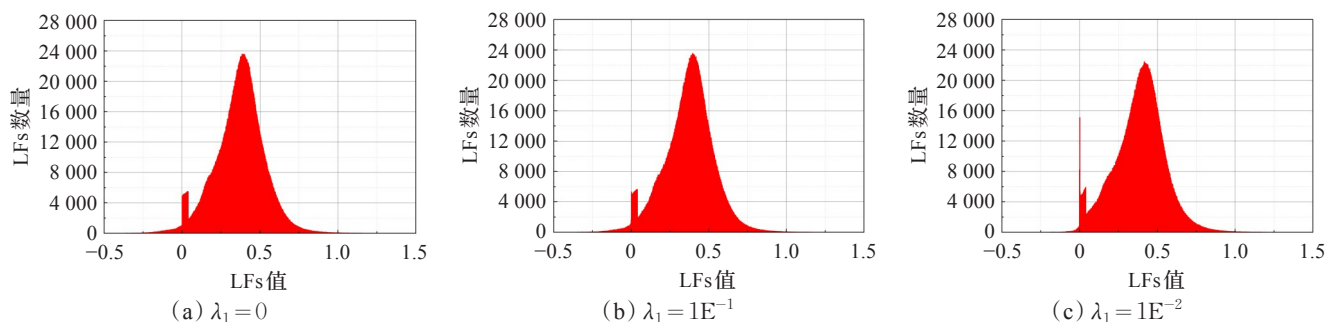


图5 D2上的LFs值分布

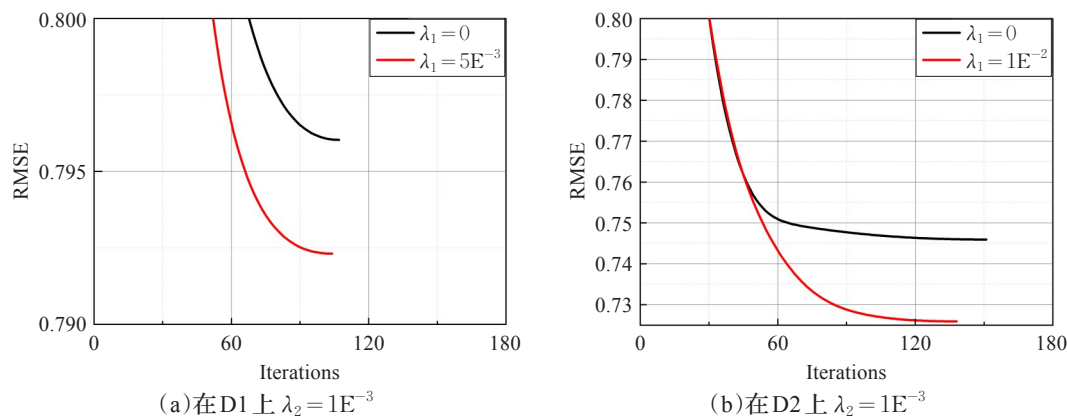


图6 SGD-LF和SPGD-LF模型的精度比较

(1)在模型中同时加入 L_1 和 L_2 正则化项,在SGD-LF基础上很大程度的提高模型的预测精度。在图2(d)中可得当仅有 L_2 正则化项和同时有 L_1 和 L_2 时,预测精度分别为0.796和0.791,其精度差为0.63%,同样的情况在图3(d)其精度分别为0.741和0.733其精度差为1.07%。同样的实验结果在图2和图3的(b)和(c)中也能得到。

(2)在模型中同时加入 L_1 和 L_2 正则化项,能使模型精度达到最优。根据本文中SPGD算法推导可得,当 $\lambda_1=0$ 且 $\lambda_2=0$ 时,SPGD算法构造的LF模型,实为不带任何正则化项的SGD算法构造的LF模型;当 $\lambda_1=0$ 且 $\lambda_2 \neq 0$ 时,其为SGD-LF模型;当 $\lambda_1 \neq 0$ 且 $\lambda_2 \neq 0$ 时,即为SPGD-LF模型。如图2所示,这三种情况在D1上的最好精度分别为:0.807、0.778和0.777,其精度差分别为3.86%、0.13%和0;同理在D2上的最好精度分别为:0.752、0.719和0.717,其精度差分别为4.88%、0.28%和0。

(3)SPGD-LF模型很大程度地提高了模型的稀疏性。在所有数据集上设置 $\lambda_2=0.01$ 。在D1上,如图4可得,当 $\lambda_1=0$ 时,为零的LFs比重最小;随着 λ_1 值增加时,为零的LFs比重也随之增加,当 $\lambda_1=1E^{-2}$ 时,为零的LFs比重增加的最为明显。同样的情况在D2上,图5中也可以看到。根据实验效果,相比于SGD-LF模型,SPGD-LF模型能提高模型的稀疏性,更好地表达了HiDS矩阵中行列间的关系,去除不相关的干扰信息。

(4)SPGD-LF模型在精度和收敛速度都优于SGD-LF模型,减少了模型的计算成本。为了进一步比较SPGD-LF

模型和SPGD-LF模型性能,在图2(d)和图3(d)中各找出一组对比实验如图6所示。图6为分别给出在D1和D2上的一组迭代收敛过程。在图6(a)上可得,SPGD-LF和SGD-LF的精度分别0.972和0.976,其精度提高了0.41%;其收敛轮数分别为104和112,收敛速度提高了7.69%。同样的情况在图6(b)上可见,SPGD-LF和SGD-LF的精度分别0.725和0.745,其精度提高了2.76%,其收敛轮数分别为138和151,收敛速度提高了9.42%。

5 总结

本文实现了一个同时带有 L_1 和 L_2 正则化项的SPGD-LF模型。其既可通过 L_1 正则化项来提高模型的稀疏性并增强模型的泛化能力,也能通过 L_2 正则化项来避免模型过拟合。该模型在收敛速度、预测精度和稀疏性等性能等方面都比SPD-LF模型有明显提高,大大提高了模型的性能。

参考文献:

- [1] Chua H N, Wong L. Increasing the reliability of protein interactomes[J]. Drug Discovery Today, 2008, 13: 652.
- [2] Luo X, Zhou M C, Wang Z, et al. An effective scheme for QoS estimation via alternating direction method-based matrix factorization[J]. IEEE Transactions on Services Computing, 2019, 12(4): 503-518.
- [3] You Z H, Zhou M C, Xin L, et al. Highly efficient framework for predicting interactions between proteins[J].

- IEEE Transactions on Cybernetics,2017,47(3):731-743.
- [4] Pan J J, Pan S J, Jie Y, et al. Tracking mobile users in wireless networks via semi-supervised co-localization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2012,34(3):587-600.
- [5] Zhang S, Wang W, Ford J, et al. Learning from incomplete ratings using non-negative matrix factorization[C]// Proc of the SIAM Int Conf on Data Mining, Bethesda, USA,2006:549-553.
- [6] Fidel C, Victor C, Diego F, et al. Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems[J]. ACM Transactions on the Web,2011,5(1):1-33.
- [7] 罗俊,李劲华. LSHBMRPK-means 算法及其应用[J]. 计算机工程与应用,2017,53(21):62-67.
- [8] 王建飞,亢良伊,刘杰,等. 分布式随机方差消减梯度下降算法 topkSVRG[J]. 计算机科学与探索,2018,12(7):1047-1054.
- [9] 张以文,艾晓飞,崔光明,等. 联合用户兴趣矩阵及全局偏好的推荐算法[J]. 计算机科学与探索,2018,12(2):197-207.
- [10] 陈婷,朱青,周梦溪,等. 社交网络环境下基于信任的推荐算法[J]. 软件学报,2017,28(3):721-731.
- [11] Luo X, Zhou M C, Xia Y N, et al. An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems[J]. IEEE Transactions on Industrial Informatics,2014,10(2):1273-1284.
- [12] Luo X, Shang M, Li S. Efficient extraction of non-negative latent factors from high-dimensional and sparse matrices in industrial applications[C]//Proc of the 16th IEEE Int Conf on Data Mining, Barcelona, Spain,2016:311-319.
- [13] Brunet J P, Tamayo P, Golub T R, et al. Metagenes and molecular pattern discovery using matrix factorization[J]. Proceedings of the National Academy of Sciences of the United States of America,2003,101(12):4164-4169.
- [14] Pan J J, Pan S J, Jie Y, et al. Tracking mobile users in wireless networks via semi-supervised co-localization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2012,34(3):587-600.
- [15] Xu Y, Yin W, Wen Z, et al. An alternating direction algorithm for matrix completion with nonnegative factors[J]. Frontiers of Mathematics in China,2012,7(2):365-384.
- [16] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems[J]. IEEE Computer,2009,42(8):30-37.
- [17] Salakhutdinov R, Mnih A. Probabilistic matrix factorization[J]. Advances in Neural Information Processing Systems,2008,20:1257-1264.
- [18] Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems[J]. SIAM Journal on Imaging Sciences,2009,2(1):183-202.
- [19] Combettes P L, Wajs V R. Signal recovery by proximal forward-backward splitting[J]. SIAM Journal on Multi-scale Modeling & Simulation,2006,4(4):1168-1200.
- [20] Polyak B T. Introduction to optimization, translations series in mathematics and engineering[M]. New York: Publications Division,1987.

(上接第120页)

- [9] Han K, Li Q, Deng Z. Security and efficiency data sharing scheme for cloud storage[J]. Chaos Solitons & Fractals the Interdisciplinary Journal of Nonlinear Science & Non-equilibrium & Complex Phenomena,2016,86:107-116.
- [10] Zhao Q, Zhang Y, Zhang G, et al. Ciphertext-policy attribute based encryption supporting any monotone access structures without escrow[J]. Chinese Journal of Electronics,2017,26(3):640-646.
- [11] Goldreich O. Foundations of cryptography: Volume 2, basic applications[M]. Cambridge: Cambridge University Press,2004.615-626.
- [12] 巩林明,李顺东,窦家维,等. 同态加密方案及安全两点直线计算协议[J]. 软件学报,2017(12):3274-3292.
- [13] 王会勇,冯勇,赵岭忠,等. 基于多密钥同态技术的安全多方计算协议[J]. 华南理工大学学报(自然科学版),2017(7):69-76.
- [14] 葛永. 若干安全多方数值计算基础协议的设计[D]. 合肥:安徽大学,2017.
- [15] Ardehali M. Quantum oblivious transfer and bit commitment protocols based on two non-orthogonal states coding[J]. Computer Science,2017,41(12):2445-2454.
- [16] Zhang G, Liu L, Liu Y. An attribute-based encryption scheme secure against malicious KGC[C]//Proceedings of IEEE, International Conference on Trust, Security and Privacy in Computing and Communications,2012:1376-1380.