计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(11):373−379 [doi: 10.15888/j.cnki.csa.008766]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

聚类概率矩阵分解的变分推断及应用①

刘　杰[1,2], 叶子锋[1,2]

[1](中国科学技术大学 管理学院, 合肥 230026)
[2](中国科学技术大学 国际金融研究院, 合肥 230026)
通信作者: 叶子锋, E-mail: zfyeee@mail.ustc.edu.cn

摘　要: 概率矩阵分解模型根据用户历史交互信息个性化推荐商品, 是协同过滤中的经典方法之一. 传统矩阵分解假设下无法利用不同用户之间的相似性, 且在面对异常值时常预测失准. 根据用户聚类信息, 可构建共轭先验分布与类别相关的聚类概率矩阵分解模型, 同时改变相关共轭先验分布形式, 完成对参数作正则化处理. 通过变分推断, 理论推导变分参数的显式表达式, 从而建立相应评分预测算法. 模拟及真实数据集均表明该模型的预测性能优于基准模型, 并能对用户评分做出现实解释.

关键词: 推荐系统; 聚类; 矩阵分解; 变分近似推断; 坐标下降算法; 协同过滤; 预测模型

引用格式: 刘杰,叶子锋.聚类概率矩阵分解的变分推断及应用.计算机系统应用,2022,31(11):373–379. http://www.c-s-a.org.cn/1003-3254/8766.html

Variational Inference of Probabilistic Matrix Factorization Based on Clustering and Its Application

LIU Jie[1,2], YE Zi-Feng[1,2]

[1](School of Management, University of Science and Technology of China, Hefei 230026, China)
[2](International Institute of Finance, University of Science and Technology of China, Hefei 230026, China)

Abstract: Probabilistic matrix factorization model, making personalized item recommendations according to a user's historical interaction information, is one of the classic methods in collaborative filtering. Under the assumption of the traditional matrix factorization model, the similarities among different users cannot be used, and prediction is often inaccurate when outliers occur. A clustering-based probabilistic matrix factorization model with category-related conjugate prior distribution is built with user clustering information. Its parameters are regularized by changing the form of the conjugate prior distribution. Through variational inference, the explicit expressions of variational parameters are theoretically derived, and corresponding rating prediction algorithms are thereby established. Both simulation and real datasets show that the prediction performance of the proposed model is better than that of the benchmark model, and it can provide realistic explanations for users' rating behavior.

Key words: recommender system; clustering; matrix factorization; variational approximate inference; coordinate descent algorithm; collaborative filtering; prediction model

推荐系统作为一种有效的过滤手段, 可对体量庞大、类型繁多、结构复杂的数据进行挖掘和建模分析, 帮助用户快速、准确地筛选其感兴趣的信息, 大幅节省检索花费的时间和精力, 从而缓解信息过载问题. 其应用可涵盖诸多领域, 如电子商务领域中的"猜你喜欢"专栏、书籍影音领域中电影评分及推荐功能;社交网络领域中推荐可能认识的好友.

矩阵分解模型因其在 Netfilx 举办的推荐系统大赛上获得优胜而倍受研究者关注[1]. 该模型针对用户与项目间的评分进行建模, 假定用户的偏好、项目的特

① 基金项目: 国家自然科学基金 (71771201, 71874171, 71731010, 71631006, 71991464, 71871208, 72071193)
收稿时间: 2022-02-15; 修改时间: 2022-03-16; 采用时间: 2022-03-21; csa 在线出版时间: 2022-07-07

征由高维空间中的隐因子向量所决定, 隐因子各分量表示无法被直接观测到的属性特征, 进一步用其内积表示评分.

为更好地利用相似用户间的信息以提高预测性能, 本文采用了基于用户聚类的概率矩阵分解模型 (clustering Bayesian probabilistic matrix factorization, CBPMF), 在已知聚类结果的条件下, 通过指数族共轭先验形成概率生成模型, 利用矩阵分解思想进行评分预测.

1 相关工作

推荐系统可分为 3 大类方法: 基于内容、协同过滤和混合式方法[2,3]. 本文主要关注协同过滤方法, 此方法应用广泛, 预测表现优异, 其核心假设为相似的用户会与相似的商品产生相似的交互, 因此可仅依靠已存在的用户和项目之间的历史交互, 即观测评分, 完成目标用户推荐. 而矩阵分解则隶属协同过滤方法大类, 研究关注于在不同场景中拓展矩阵分解模型, 利用高效的算法估计模型参数, 同时保持预测精度.

传统矩阵分解方法假设观测到的数据来自低维的线性子空间, 重构便可得到矩阵的低秩结构, 可利用特征值分解、非负矩阵分解等方法求解参数[4,5]. 而 Salakhutdinov 等人对矩阵分解施加了严格的概率分布, 从而建立概率矩阵分解模型的基本框架, 用于处理体量大、数据稀疏、结构不平衡的电影评分数据集[6]. 进一步地, 为了解决超参数选择困难的问题, Salakhutdinov 等人引入共轭先验, 提出贝叶斯概率矩阵分解模型, 对应地使用 Gibbs 采样算法估计参数[7].

矩阵分解还融合除观测评分外的其他信息, 提出了概率矩阵分解模型的变体, 以缓解数据稀疏问题, 从而提高预测性能. 为了利用用户间关系, Ma 等人将用户信任矩阵与评分矩阵结合, 提出了基于社交网络的概率矩阵分解模型, 缓解了只考虑评分信息时的数据稀疏问题[8]. Liu 等人融合了社交关系和项目内容信息, 假设用户和项目隐因子各自服从不同的先验分布, 以此提高预测精度[9]. Peng 等人考虑现实中具有信任关系的用户偏好其实并不相似的情况, 将用户偏好相似性约束在特定领域, 提出了基于信任和偏好切分的矩阵分解模型[10]. Feng 等人同时考虑用户评分和局部关系相似性, 以解决推荐系统场景中数据稀疏的问题[11].

在参数估计方面, 目前求解隐因子的算法主要有 MCMC[12]、变分近似推断[13,14] 以及梯度下降. Lim 等人提出了基于变分推断的 SVD 矩阵分解算法[15]. Luo 等人利用伽马共轭先验缓解正态假设不稳健及数据缺失的问题, 采用了 MCMC 方法进行参数估计, 提供了统计意义上的解释[16]. Zhao 等人构造了 L1 范数惩罚, 利用变分近似推断求解低秩矩阵分解模型中的参数, 在计算机视觉领域取得了较好结果[17]. 王娟等人使用随机梯度下降算法进行矩阵分解模型的参数求解[18]. Blei 等人指出相比于 MCMC 方法, 变分推断以损失部分精度为代价获得更快的收敛速度, 因此更适合于体量较大的数据集[19].

目前研究大多考虑将用户集合视作整体, 或认为各用户之间互不相同, 两种建模思路均未采用聚类方法, 无法利用相似用户之间的信息提升预测表现, 且后者还会因待估参数过多而造成过拟合. 本文采用的聚类概率矩阵分解模型可有效解决用户先验分布设置过于粗糙或过于精细而导致评分预测不准确的问题. 具体来说, 此模型有两处创新, 其一, 通过贝叶斯分层使同一水平之间的随机变量共享信息, 对用户聚类并使同类用户服从同一先验, 进而提升用户偏好预测的准确性; 其二, 更换先验形式, 构建贝叶斯框架, 完成对模型参数的正则化, 并且考虑多个模型的平均而非某一特定模型, 防止过拟合现象. 此外, 本文将采用变分推断方法, 对应 CBPMF 模型, 推导得到各随机变量对应变分分布的显式表达式, 从而建立变分坐标下降算法以估计参数.

2 聚类概率矩阵分解模型

本节主要对 CBPMF 模型的评分生成过程作详细说明. 概率矩阵分解模型主要针对推荐系统中的评分数据进行建模. 假定用户 i 对项目 j 的评分R_{ij}可分解为 D 维空间中用户隐因子U_i和项目隐因子V_j的内积, 即$R_{ij} \approx U_i^{\mathrm{T}} V_j$, 对应矩阵形式为$R \approx U^{\mathrm{T}} V$, 其中$U_{D\times N} = [U_1, \cdots, U_N]$, $V_{D\times M} = [V_1, \cdots, V_M]$. 各个可观测评分$R_{ij}$之间相互独立, 且服从均值为$U_i^{\mathrm{T}} V_j$, 方差为$\sigma_0^2$的正态分布, 如式 (1):

$$p(R_{ij}|U_i^{\mathrm{T}} V_j, \sigma_0^2) = N(R_{ij}|U_i^{\mathrm{T}} V_j, \sigma_0^2) \tag{1}$$

给定用户聚类结果, CBPMF 模型假定同一类别中用户的先验分布相同, 在每一类下利用各自先验参数分别进行矩阵分解, 最后根据隐因子估计预测评分. 具体来说, 假定用户的类别为常数 K, 类别向量为$Z =$

$[z_1,\cdots,z_N]$, 分量z_i表示用户 i 所属的特定类别, 且$z_i \in \{1,\cdots,K\}$, 若$z_i = k$, 则说明用户 i 属于类别 k. 在类别向量Z已知的情况下, 用户、项目隐因子分布分别如式 (2) 所示:

$$\begin{cases} p(U_i|\mu_{z_i}^U,\tau_{z_i}^U) = N_D(U_i|\mu_{z_i}^U,(\tau_{z_i}^U)^{-1}I) \\ p(V_j|\mu^V,\tau^V) = N_D(V_j|\mu^V,(\tau^V)^{-1}I) \end{cases} \tag{2}$$

其中, $N_D(x|\mu,\Sigma)$表示均值向量为μ, 协方差矩阵为Σ的多元正态密度函数, 参数τ为精度系数, 从中不难发现, 由于用户聚类结果存在, 用户隐因子与其所属的特定类别向量有关, 而项目隐因子则无关类别. 记用户和项目隐因子对应先验参数集合$\Theta_U = \{\mu_1^U,\cdots,\mu_K^U,\tau_1^U,\cdots,\tau_K^U\}$, $\Theta_V = \{\mu^V,\tau^V\}$. 假定用户均值向量服从均值为$\mu_0$, 协方差阵为$\sigma_0^U I$的多元正态分布, 精度系数$\tau$服从参数为$a_0,b_0$的伽马分布, 如式 (3) 所示:

$$\begin{cases} p(\mu_k^U|\mu_0,\sigma_U^2) = N_D(\mu_k^U|\mu_0,\sigma_U^2 I) \\ p(\tau_k^U|a_0,b_0) = \Gamma(\tau_k^U|a_0,b_0) \end{cases} \tag{3}$$

其中, $\Gamma(x|a,b)$表示参数为a,b的伽马密度函数, μ_0,σ_0^U,a_0,b_0为给定的超参数. 类似地, 对于项目隐因子参数类似有如式 (4) 所示的先验分布, 区别在于分布参数是否与聚类结果有关.

$$\begin{cases} p(\mu^V|\mu_0,\sigma_V^2) = \mathrm{N_D}(\mu^V|\mu_0,\sigma_V^2\mathrm{I}) \\ p(\tau^V|a_0,b_0) = \Gamma(\tau^V|a_0,b_0) \end{cases} \tag{4}$$

此处模型所采用的先验为正态-正态、正态-伽马双重共轭分布, 对应的概率生成过程如图 1 所示, 其中实心圆R_{ij}表示可观测的评分, 空心圆圈中的变量为隐因子及其先验分布参数, 其余均为给定的超参数, 记为$\Theta_0 = \{a_0,b_0,\mu_0,\sigma_U,\sigma_V,\sigma_0\}$.

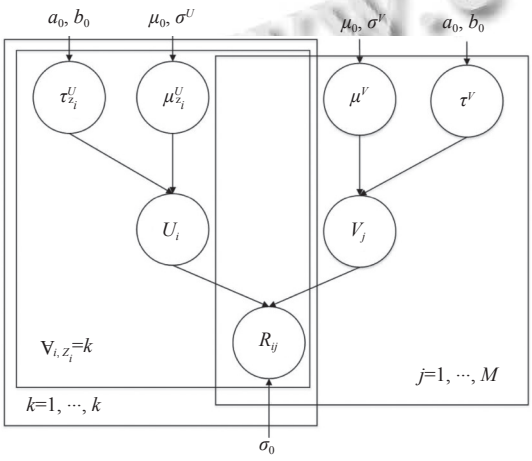

图 1 CBPMF 模型概率图

3 变分近似算法

模型过于复杂而导致参数无法求解是概率生成模型常见的困难之一. 本文将采用变分推断近似估计模型参数, 从而完成评分的预测工作.

3.1 贝叶斯变分推断理论

根据贝叶斯推断, 参数估计需根据后验分布进行, 但其因为积分运算过于复杂而无显式表达, 故无法直接通过最大化后验分布估计参数U,V. 为了解决此问题, 本文引入可处理的变分分布$q(\cdot)$近似真实后验分布, 假定其随机变量之间相互独立且由各变分参数控制, 如式 (5) 所示. 通过最小化变分分布和后验分布之间 KL (Kullback-Leibler) 距离寻找最优$q^*(\cdot)$, 当且仅当变分分布$q(\cdot)$等于真实后验分布$p(U,V|R,\Theta_0)$时, KL 距离达到最小.

$$\begin{aligned} q(U,V,\Theta_U,\Theta_V) = &\sum_{i=1}^{N} q(U_i)\sum_{j=1}^{M} q(V_j) \\ &\cdot\sum_{k=1}^{K}\left\{q(\mu_k^U)q(\tau_k^U)\right\}q(\mu^V)q(\tau^V) \end{aligned} \tag{5}$$

CBPMF 模型似然函数$\log p(R|\Theta_0)$可以分解为变分下界 (evidence low bound) 与 KL 距离之和. 给定评分数据的情况下, 最小化 KL 距离等价于最大化变分下界, 故可将感兴趣参数U,V的估计问题转化为对应的优化问题求解. 根据平均场变分推断 (mean-field variational inference), 固定隐因子U_i,V_j, 对其余无关随机变量求期望即可得到对应的变分分布, 如式 (6) 所示:

$$\begin{cases} q(U_i) \propto \exp\left\{E_{-U_i}[\log p(U,V,\Theta_U,\Theta_V,R|\Theta_0)]\right\} \\ q(V_j) \propto \exp\left\{E_{-V_j}[\log p(U,V,\Theta_U,\Theta_V,R|\Theta_0)]\right\} \end{cases} \tag{6}$$

其中, E_{-X_i}表示对除X_i外所有随机变量做期望. 给定评分矩阵R及超参数集合Θ_0的条件下, 先验参数μ_k^U, $k=1,\cdots,K$和μ^V服从多元正态分布, 即:

$$\begin{cases} \mu_k^U \sim N_D(\mu_k^{U^*},\Sigma_k^{U^*}),\ k=1,\cdots,K \\ \mu^V \sim N_D(\mu^{V^*},\Sigma^{V^*}) \end{cases} \tag{7}$$

其中, 均值向量为:

$$\begin{cases} \mu_k^{U^*} = \Sigma_k^{U^*-1}\cdot(\sigma_U^{-2}\mu_0 + E[\tau_k^U]\sum\limits_{i\in U(k)} E[U_i]) \\ \mu^{V^*} = \Sigma^{V^*-1}\cdot(\sigma_V^{-2}\mu_0 + E[\tau^V]\sum\limits_{j=1} ME\left[V_j\right]) \end{cases} \tag{8}$$

协方差矩阵为:

$$\begin{cases} \Sigma_k^{U^*} = (\sigma_U^{-2} + E\left[\tau_k^U\right]N_k^U)^{-1}I \\ \Sigma^{V^*} = (\sigma_V^{-2} + E\left[\tau^V\right]M)^{-1}I \end{cases} \tag{9}$$

其中, $U(k)$表示属于 k 类的用户.

同样条件下, 先验参数τ_k^U, $k=1,\cdots,K$和τ^V服从伽马分布, 即:

$$\begin{cases}\tau_k^U \sim Gamma(a_k^{U^*}, b_k^{U^*}),\ k=1,\cdots,K\\ \tau^V \sim Gamma(a^{V^*}, b^{V^*})\end{cases} \tag{10}$$

其中, 形状参数为:

$$a_k^{U^*}=\frac{DN_k^U}{2}+a_0,\ k=1,\cdots,K;\ a^{V^*}=\frac{DM}{2}+a_0 \tag{11}$$

且N_k^U表示类别 k 的用户数量, 逆尺度参数为:

$$\begin{cases}b_k^{U^*}=b_0+\dfrac{1}{2}\displaystyle\sum_{i\in U(k)}E\left[\left(U_i-\mu_k^U\right)^{\mathrm{T}}\cdot\left(U_i-\mu_k^U\right)\right]\\ b^{V^*}=b_0+\dfrac{1}{2}\displaystyle\sum_{j=1}^{M}E\left[\left(V_j-\mu_k^V\right)^{\mathrm{T}}\cdot\left(V_j-\mu_k^V\right)\right]\end{cases} \tag{12}$$

在此基础上, 可推导得预测评分所必须的用户、项目隐因子参数服从多元正态分布, 即:

$$\begin{cases}U_i \sim N_D(\mu_{U_i}^*, \Sigma_{U_i}^*),\ i=1,\cdots,N\\ V_j \sim N_D(\mu_{V_j}^*, \Sigma_{V_j}^*),\ j=1,\cdots,M\end{cases} \tag{13}$$

其中, 均值向量为:

$$\begin{cases}\mu_{U_i}^*=\Sigma_{U_i}^*\cdot\left(E\left[\tau_{z_i}^U\cdot\mu_{z_i}^U\right]+\sigma_0^{-2}\displaystyle\sum_{j=1}^{M}I_{ij}R_{ij}E\left[V_j\right]\right)\\ \mu_{V_j}^*=\Sigma_{V_j}^*\cdot\left(E\left[\tau^V\cdot\mu^V\right]+\sigma_0^{-2}\displaystyle\sum_{i=1}^{N}I_{ij}R_{ij}E\left[U_i\right]\right)\end{cases} \tag{14}$$

及协方差矩阵为:

$$\begin{cases}\Sigma_{U_i}^*=\left(E\left[\tau_{z_i}^U\right]I+\sigma_0^{-2}\displaystyle\sum_{j=1}^{M}I_{ij}E\left[V_jV_j^{\mathrm{T}}\right]\right)^{-1}\\ \Sigma_{V_j}^*=\left(E\left[\tau^V\right]I+\sigma_0^{-2}\displaystyle\sum_{i=1}^{\mathrm{N}}\mathrm{I}_{ij}E\left[U_iU_i^{\mathrm{T}}\right]\right)^{-1}\end{cases} \tag{15}$$

由于 CBPMF 模型采用指数族共轭先验, 故导出变分分布形式上与原分布保持一致. 上述期望计算均可根据基本矩阵运算得到, 在此不过多赘述.

3.2 评分预测算法

根据上述理论推导, 本节给出 CBPMF 模型评分预测算法. 具体来说, 该算法可视作坐标下降过程, 首先初始化变分参数集$\Theta_q^{(0)}$, 在第 t 次更新中, 根据上一时刻的变分参数$\Theta_q^{(t)}$, 利用可迭代得到下一时刻的参数集$\Theta_q^{(t+1)}$, 在给定阈值ε情况下, 若变分下界的变化量小于阈值, 即$L\left(\Theta_q^{(t+1)}\right)-L\left(\Theta_q^{(t)}\right)<\varepsilon$, 则认为参数更新至收敛.

算法 1. CBPMF 模型评分预测算法

(1) 清洗数据得到评分矩阵R;
(2) 初始化变分参数集$\Theta_q^{(0)}$, 超参数集Θ_0, 阈值ε;
(3) **while** 变分下界未达到收敛 **do:**
(4)
1) 根据式 (7)、式 (8), 更新每一类别中用户先验均值向量μ_k^U以及精度参数τ_k^U;
2) 根据式 (7)、式 (8), 更新单一项目先验均值向量μ^V以及精度参数τ^V;
3) 根据式 (14)、式 (15), 更新每一用户的隐因子向量的多元正态分布参数;
4) 根据式 (14)、式 (15), 更新每一项目的隐因子向量的多元正态分布参数;
5) 根据当前变分参数$\Theta_q^{(t)}$计算变分下界$L\left(\Theta_q^{(t)}\right)$;
(5) **end while**
(6) 根据收敛的用户、项目隐因子, 预测未观测的评分, 即$\hat{R}_{ij}=U_i^{\mathrm{T}}V_j\approx\mu_{U_i}^{*\mathrm{T}}\mu_{V_j}^*$;
(7) 将目标用户的预测评分进行排序, 推荐排名最靠前的 k 个项目;

得到变分参数集合Θ_q后, 用变分分布近似代替后验分布. 由于用户、项目隐因子参数的变分参数服从多元正态分布, 其 MAP 估计为对应的均值向量, 因此可得到U_i^*, V_j^*的近似估计为$\mu_{U_i}^*, \mu_{V_j}^*$, 进一步可估计缺失评分. 最后根据缺失评分的预测结果, 给出目标用户最感兴趣的 top-k 项目推荐.

本算法基于变分近似推断的统计原理构建, 运用变分参数近似 CBPMF 模型中的真实后验分布, 从而根据最大化后验分布估计参数, 其理论复杂度为$O((N+M)D+K)$, 故效率较高, 可快速进行更新预测. 相比于 MCMC 和单纯的坐标梯度下降算法, 此算法收敛速度较快且保持一定的预测准确性, 适合推荐系统场景.

4 实验与分析

4.1 模拟分析

为了契合推荐系统应用场景, 本文根据协同过滤思想来产生模拟数据, 同时结合聚类方法, 考虑不同类别中用户表现存在的差异[20], 具体来说, 模拟数据假定存在$N=300$位用户、$M=300$个项目, 隐空间的维度$D=10$; 异常数据占比为 10%; 且设置数据缺失率为 0.9 以对应评分的稀疏性. 在类别数量K分别为 3、5 两种情形下, 针对用户平均评分倾向进行聚类, 最后使用 5 折交叉验证预测缺失评分, 在 100 次重复试验下得到指标 MAE 和 MSE 的均值和标准差用以评价模型.

基准模型则采用贝叶斯概率矩阵分解模型 BPMF 以及社交关系矩阵分解模型 SRBPMF, 其中 BPMF 模

型对应未考虑用户的相似性, 不进行聚类, 假定所有用户隐因子服从同一先验分布的情况, 因此对模型参数的惩罚较弱; 而 SRBPMF 模型则对应所有用户各自服从参数不同的先验分布的情况, 即每个用户各为一类, 此方法假设过于强, 存在相当大的过拟合风险. 本文采用的 CBPMF 模型则可视作通过聚类对上述两种方法作了折中处理.

从表 1 MAE、MSE 可明显看出本文采用的 CBPMF 模型在预测上的表现优于 BPMF 模型和 SRBPMF 模型, 且三者中 SRBPMF 模型表现最差; MAE、MSE 的标准差于括号中显示, 从中可以看出模拟数据中 3 类模型预测波动不大. 因 SRBPMF 模型需要待估参数与用户数量成线性增长, 当用户数量较多时, 过拟合的风险非常高; 而 BPMF 模型则无法利用同一类别用户之间的相似关系, 两者均无法较好处理模拟数据中的异常值, 因此预测表现均不如 CBPMF 模型. 另一方面, CBPMF 模型具有分层贝叶斯结构, 在对用户进行聚类后, 可共享同一水平的用户信息, 当某用户观测到异常值而导致估计偏差时, 该类别下的参数先验分布通过赋予一个较低的先验概率值而达到正则化的作用, 降低异常值影响, 表现更为鲁棒.

表 1 评分预测情况汇总表

类别数量	模型	评价指标	
		MAE	MSE
K=3	CBPMF	0.573 5 (0.026)	0.601 3 (0.062)
	BPMF	0.577 0 (0.025)	0.603 1 (0.060)
	SRBPMF	0.578 6 (0.027)	0.610 0 (0.063)
K=5	CBPMF	0.567 2 (0.016)	0.571 9 (0.034)
	BPMF	0.572 9 (0.015)	0.581 6 (0.032)
	SRBPMF	0.575 7 (0.014)	0.585 7 (0.031)

注: 括号中为标准差

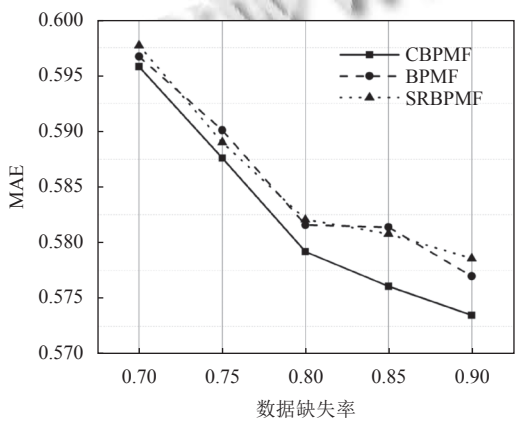

图 2 各模型 MAE 随数据缺失率变化折线图

模拟考虑不同数据缺失率对模型预测准确性的影响, 结果如图 2 所示, 对于任一数据缺失率而言, CBPMF 模型预测表现均优于基准模型. 由于异常值占比恒定为 10%, 随着可观测数据的增加, 异常值的数量也随之同步增加, 故在本模拟中数据缺失率和模型预测准确性之间成反比关系. 另外, 随着异常值总数降低, 相比于基准模型, CBPMF 模型可以获得更大的提升, 意味着其能够更好地应对异常值对模型的干扰.

4.2 实证分析

本小节运用 MovieLens 电影评分公共数据集进行实证分析. 该数据集包含 943 名用户和 1 682 部电影, 交互得到评分数据共计 10 000 条, 具体还包括用户的性别与年龄、电影的名称及类型等描述性信息. 为保证评分有效性, 该数据集中每位用户至少对 20 部以上电影给出过评分. 评分取值于离散有限集合, 即$R_{ij} \in \{1,2,3,4,5\}$.

为了更加直观地观测用户对电影评分的倾向, 本小节挑选了平均评分位于正态分布前 25% 的严格用户 (用深蓝色节点表示), 以及位于后 25% 的宽容用户 (用红色节点表示) 两类用户群体, 根据评分情况绘制网络结构图, 如图 3 所示. 若电影节点的度越大, 则表明该电影被观看次数越多, 反之亦然; 若用户节点发出的边为红色, 则说明该用户给出评分为 4 或 5 的积极评价, 反之则用蓝色边表示消极评价.

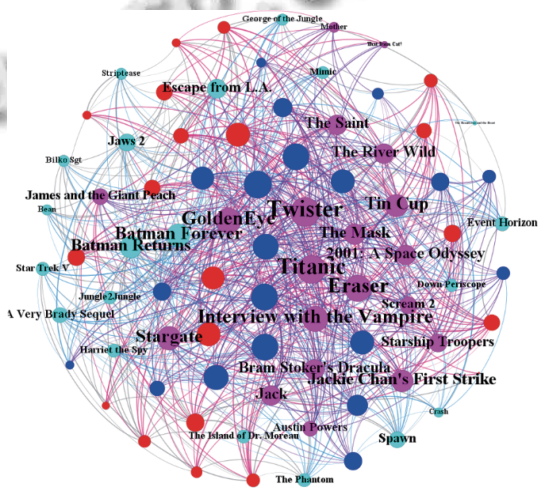

图 3 用户电影评分球状网络结构图

对球形网络图进行分析可得以下结论: (1) 电影的评分与其自身质量紧密相关. 高质量电影 (粉色节点) 集中在球形网络中间且边为红色, 标签对应为现实中

《007》《泰坦尼克号》等电影; 反之低质量电影 (浅蓝色) 游离于网络边缘, 边为蓝色, 对应的电影知名度较低; (2) 电影所受关注度与其自身质量紧密相关. 半径较大的节点所多为红色, 表示高质量电影受到关注度高; 反之则半径小的节点颜色多为蓝色, 表示低质量电影较少被人观看; (3) 用户评分表现出明显的倾向性. 对于同一电影而言, 宽容用户更倾向于给出积极评价, 表现为红色节点的边多为红色; 反之严格用户倾向于给出消极评价, 表现为浅蓝色节点的边多为蓝色.

根据用户评分均值, 事先得到类别数量为 K=3 的聚类结果. 根据不同随机数种子进行 5 折交叉验证, 进一步得到 MovieLens 数据集中的评分预测的评价指标 MAE 和 MSE 的均值如表 2 所示. 与模拟表现类似, 在隐空间维度和 10 和 20 的情况下, CBPMF 模型预测表现均明显优于基准模型 BPMF 和 SRBPMF. 标准差于括号中显示, 从标准差表现上看, CBPMF 模型表现更为稳定, 随 5 折交叉验证抽取数据的变化较小. 另外, 不难发现, 随着隐空间维数的增加, 模型预测性能增强.

表 2 MovieLens 数据集上模型预测情况汇总表

隐空间维度	模型	评价指标	
		MAE	MSE
D=10	CBPMF	0.8825 (0.0006)	1.1380 (0.0012)
	BPMF	0.8848 (0.0014)	1.1431 (0.0031)
	SRBPMF	0.8849 (0.0013)	1.1437 (0.0029)
D=20	CBPMF	0.8781 (0.0011)	1.1278 (0.0018)
	BPMF	0.8822 (0.0023)	1.1399 (0.0028)
	SRBPMF	0.8838 (0.0021)	1.1418 (0.0027)

注: 括号中为标准差

5 结论与展望

本文考虑现实中不同用户之间存在相似性的特点, 通过正态-正态、正态-伽马两种共轭指数先验构建贝叶斯分层结构, 将矩阵分解模型拓展为用户聚类情形下的概率矩阵分解模型. 一方面, 聚类概率矩阵分解模型 (CBPMF) 将具有相似评分倾向的用户聚为一类, 使其共享同一先验参数, 以此提高模型的预测准确性. 另一方面, 贝叶斯框架可通过对先验参数积分来考虑多个模型的平均, 从而达到防止过拟合的作用, 事实上在模型角度, 先验可视为对参数的一种正则化. 在参数估计上, 本文运用平均场变分推断思想, 使用变分分布近似真实后验分布, 严格推导得到变分参数的显式表达式, 从而对应拓展得到 CBPMF 模型评分预测算法, 此算法与推荐系统内的用户和项目数量呈线性关系, 可快速得到变分参数的近似估计.

虽然实证表明用户聚类可显著提高预测表现, 但本文未考虑不同的聚类方法所带的影响. 如 K-means、层次聚类法等, 从而进一步挖掘评分数据中用户的潜在倾向, 讨论预测工作. 此外, 类别数量K选取对应着模型选择, 也是值得研究的方向.

参考文献

1 Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. Computer, 2009, 42(8): 30–37. [doi: 10.1109/MC.2009.263]

2 Bobadilla J, Ortega F, Hernando A, *et al*. Recommender systems survey. Knowledge-based Systems, 2013, 46: 109–132. [doi: 10.1016/j.knosys.2013.03.012]

3 Najafabadi MK, Mohamed AH, Mahrin MN. A survey on data mining techniques in recommender systems. Soft Computing, 2019, 23(2): 627–654. [doi: 10.1007/s00500-017-2918-7]

4 Shi JR, Zheng XY, Yang W. Survey on probabilistic models of low-rank matrix factorizations. Entropy, 2017, 19(8): 424. [doi: 10.3390/e19080424]

5 Qiao HL. New SVD based initialization strategy for non-negative matrix factorization. Pattern Recognition Letters, 2015, 63: 71–77. [doi: 10.1016/j.patrec.2015.05.019]

6 Salakhutdinov R, Mnih A. Probabilistic matrix factorization. Proceedings of the 20th International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2007. 1257–1264.

7 Salakhutdinov R, Mnih A. Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. Proceedings of the 25th International Conference on Machine Learning. Helsinki: ACM, 2008. 880–887.

8 Ma H, Yang HX, Lyu MR, *et al*. SoRec: Social recommendation using probabilistic matrix factorization. Proceedings of the 17th ACM Conference on Information and Knowledge Management. Napa Valley: ACM, 2008. 931–940. [doi: 10.1145/1458082.1458205]

9 Liu JT, Wu CH, Liu WY. Bayesian probabilistic matrix factorization with social relations and item contents for recommendation. Decision Support Systems, 2013, 55(3): 838–850. [doi: 10.1016/j.dss.2013.04.002]

10 Peng W, Xin BG. A social trust and preference segmentation-based matrix factorization recommendation algorithm. EURASIP Journal on Wireless Communications

and Networking, 2019, 2019(1): 272. [doi: 10.1186/s13638-019-1600-4]

11 Feng CJ, Liang JY, Song P, *et al*. A fusion collaborative filtering method for sparse data in recommender systems. Information Sciences, 2020, 521: 365–379. [doi: 10.1016/j.ins.2020.02.052]

12 Brooks S. Markov chain Monte Carlo method and its application. Journal of the Royal Statistical Society: Series D, 1998, 47(1): 69–100. [doi: 10.1111/1467-9884.00117]

13 Beal MJ, Ghahramani Z. The variational Bayesian EM algorithm for incomplete data: With application to scoring graphical model structures. Proceedings of the Bayesian Statistics 7: The 7th Valencia International Meeting. Tenerife, 2002. 453–464.

14 Wainwright MJ, Jordan MI. Graphical models, exponential families, and variational inference. Foundations and Trends® in Machine Learning, 2008, 1(1–2): 1–305. [doi: 10.1561/2200000001]

15 Lim Y, Teh Y. Variational Bayesian approach to movie rating prediction. Proceedings of Knowledge Discovery and Data Mining. 2006.

16 Luo C, Zhang B, Xiang Y, *et al*. Gaussian-Gamma collaborative filtering: A hierarchical Bayesian model for recommender systems. Journal of Computer and System Sciences, 2017, 102: 42–56. [doi: 10.1016/j.jcss.2017.03.007]

17 Zhao Q, Meng DY, Xu ZB, *et al*. L_1-norm low-rank matrix factorization by variational Bayesian method. IEEE Transactions on Neural Networks and Learning Systems, 2015, 26(4): 825–839. [doi: 10.1109/TNNLS.2014.2387376]

18 王娟, 熊巍. 基于矩阵分解的最近邻推荐系统及其应用. 统计与决策, 2019, 35(6): 17–20. [doi: 10.13546/j.cnki.tjyjc.2019.06.004]

19 Blei DM, Kucukelbir A, McAuliffe JD. Variational inference: A review for statisticians. Journal of the American Statistical Association, 2017, 112(518): 859–877. [doi: 10.1080/01621459.2017.1285773]

20 Godoy-Lorite A, Guimerà R, Moore C, *et al*. Accurate and scalable social recommendation using mixed-membership stochastic block models. Proceedings of National Academy of Sciences of the United States of America, 2016, 113(50): 14207–14212. [doi: 10.1073/pnas.1606316113]

(校对责编: 牛欣悦)