

# 结合隐式反馈与相关性建模的 概率流式张量分解推荐模型

赵建立,姚 彬,陈建建,李 浩

(山东科技大学 计算机科学与工程学院,山东 青岛 266590)

**摘要:**在推荐系统中,流式张量分解模型常被用于分析和处理高维流式数据,并向用户推荐可能感兴趣的商品。然而现有模型常常忽略对象特征之间的相关性,缺乏有效的先验建模,并且模型大多仅使用显式评分数据,忽略对隐式反馈信息的利用,导致模型精度下降。针对以上问题,提出一种结合隐式反馈与相关性建模的概率流式张量分解推荐模型(POSTFR)。POSTFR 通过添加具有 Lewandowski-Kurowicka-Joe(LKJ)先验的协方差矩阵建模对象间的相关性,并在其中加入隐式反馈信息以提高频繁出现的对象的权重,提高了模型的准确度。在多个真实数据集上,POSTFR 比其他概率流式模型取得更好的效果。

**关键词:**张量分解;流式数据;隐式反馈;概率流式模型;推荐模型

中图分类号:TP181

文献标志码:A

## Probabilistic streaming tensor decomposition recommendation model with implicit feedback and correlation modeling

ZHAO Jianli, YAO Bin, CHEN Jianjian, LI Hao

(College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China)

**Abstract:** These applications generate streaming data based on users' interactive behavior. In recommender systems, streaming tensor decomposition models are often employed for analyzing and processing streaming data, and for recommending items that users may be interested in. However, existing models often neglect the correlation between object features and lack effective prior modeling. Besides, most models only use explicit rating data without the use of implicit feedback information, thus exacerbating the accuracy loss of models. In this paper, we proposed a probabilistic streaming tensor decomposition recommendation model with implicit feedback and relationship modeling (POSTFR). POSTFR modeled the relationship between objects by adding covariance matrices with Lewandowski-Kurowicka-Joe(LKJ) priors and incorporated implicit feedback information in the updating process to increase the weight of frequently occurring objects. In this way, the accuracy of the proposed model was improved. Compared to other probabilistic streaming models, POSTFR obtained better results in several real datasets.

**Key words:** tensor decomposition; streaming data; implicit feedback; probabilistic streaming model; recommendation model

随着电子商务的蓬勃发展,海量的商品以及交互数据造成了信息过载。推荐系统是解决信息过载的有

收稿日期:2022-10-15

基金项目:国家自然科学基金项目(62072288);山东省自然科学基金项目(ZR2021MF104,ZR2021MF113);青岛市科技计划重点研发专项(21-1-2-19-xx);青岛市西海岸新区科技计划项目(2020-1-6)

作者简介:赵建立(1977—),男,山东莱芜人,教授,博士生导师,主要从事个性化推荐算法、交通大数据分析、行为计算等方面研究.E-mail:jlzhao@sdust.edu.cn

姚 彬(1994—),男,山东青岛人,硕士研究生,主要从事个性化推荐算法研究。

效手段,根据用户的兴趣与行为习惯为用户推荐可能感兴趣的商品。以在线电影评分网站为例,用户在观看电影后进行评分,网站则依据评分和上下文信息对用户进行建模,通常使用张量分解对生成的数据进行分析,以便提供更加个性化的电影推荐。CANDECOMP/PARACAC(CP)分解<sup>[1]</sup>是一种经典的张量分解模型,使用交替最小二乘法(alternating least squares, ALS)<sup>[2]</sup>或梯度下降法<sup>[3]</sup>更新从分解中获得的隐因子矩阵。尽管 CP 分解模型在许多领域都取得成功,但只适用于静态数据,面对在线应用中生成的流式数据则无能为力<sup>[4]</sup>。此外,由于存储限制或隐私保护等原因,历史数据可能无法被重新读取<sup>[5]</sup>。因此人们提出流式 CP 分解模型。2016年,Zhou等<sup>[6]</sup>率先提出 OnlineCP 模型,利用互补矩阵对原始张量交替更新,取得较好效果。但 OnlineCP 被限制为仅在时间模式增长,存在一定局限性。2017年,Song等<sup>[7]</sup>结合低秩补全算法提出 MAST 模型,以适应多模式的增量场景。在此基础上,Najafi等<sup>[8]</sup>提出 OR-MSTC 模型,旨在消除由于数据缺失和异常值对流式张量分解带来的不利影响。但上述两种方法均假设流式张量是维度增长的,无法适用于数据增量发生在张量内部的情况,如一位原有用户对原有影片新增了评分。

在线应用场景中,应用产生的时序数据对张量分解模型提出了挑战:模型在某一时刻仅能利用有限的数  
据,限制了传统张量分解模型的性能;数据更新的速度可能非常快,对算法的效率提出一定要求。为解决  
这些问题,将贝叶斯框架与 CANDECOMP/PARAFAC(CP)分解相结合的概率流式张量分解模型应运而生。  
最具代表性的是 Du等<sup>[9]</sup>提出的 POST 模型,该模型使用多元高斯先验建模隐因子矩阵,并使用 KL 散度<sup>[10]</sup>  
作为目标函数进行后验更新,能够同时适应数据在不同位置上的增量更新(无论张量的维度是否增长)。  
2019年,Zheng等<sup>[11]</sup>在 POST 模型的基础上融合边信息提出 POST-si 模型,有效提高了模型的预测精度。  
但 POST 与 POST-si 都先验地对各个对象进行了条件独立的假设,意味着模型缺乏对对象之间的相关性建  
模,同时上述模型也均未对数据中的隐式反馈信息加以利用。同年,Chen等<sup>[12]</sup>提出一种结合威沙特分布的  
概率 CP 分解模型 BGCP,将其用于交通时空数据补全并获得置信度信息,但 BGCP 使用的威沙特先验对协  
方差矩阵的正定性有严格限制,会导致迭代过程中断或收敛困难。针对该问题,Lewandowski等<sup>[13]</sup>于  
2009年提出一种使用增量 Cholesky 分解方式的 LKJ 分布,能够有效地对正定矩阵进行更新和采样而不破  
坏其正定性。2020年,Bach等<sup>[14]</sup>将预训练和先验知识加入贝叶斯模型,旨在解决文本分类领域中面临的数  
据稀疏性问题。2021年,Thanh等<sup>[15]</sup>提出一种使用随机算法的自适应在线 CP 分解模型 ROLCP,结合矩阵  
近似对张量分解过程进行优化,但 ROLCP 对隐因子矩阵进行抽象表示,一定程度上降低了预测精度。

虽然现有模型能够较好地适应在线推荐场景,但大多忽略了对对象之间的相关性,且未考虑数据中的隐式  
反馈信息,不能对隐因子矩阵中对象间的相关性进行良好建模,无法准确捕获数据的潜在特征。本研究提出  
一种结合隐式反馈与相关性建模的概率流式张量分解推荐模型(probabilistic streaming tensor decomposi-  
tion recommendation model with implicit feedback and relationship modeling,POSTFR)以解决上述问题并  
获得了良好的推荐性能。

## 1 模型的符号表示及 CP 分解介绍

### 1.1 符号表示

本研究中所使用的符号如表 1 所示。表 1 中, $d_k$  表示张量  $\mathbf{Y}$  中第  $k$  阶的维度,对应数据中第  $k$  个模式中的  
 $d_k$  个对象。张量中某个元素  $y_i$  在  $\mathbf{Y}_{(k)}$  中的坐标为  $(i_k, j)$ ,  $j$  可由式(1)计算得到:

$$j = 1 + \sum_{s=1, s \neq k}^K (i_s - 1) J_s, J_s = \prod_{t=1, t \neq k}^{s-1} d_t. \quad (1)$$

### 1.2 CP 分解简介

张量分解将原始张量表示为多个简单结构的组合,目标是利用数据的低秩特性补全张量中缺失的部分。  
在推荐系统中,张量分解常被用于预测未知的用户评分并提供个性化推荐。CP 分解是一种经典的张量分解  
方式,将张量表示为多个隐因子矩阵的乘积,记为  $\mathbf{y} = [[\mathbf{U}^1, \mathbf{U}^2, \dots, \mathbf{U}^K]]$ ,  $K$  表示张量的阶数。在 CP 分解  
中,  $y_i$  可用式(2)计算:

$$\mathbf{y}_i = \sum_{r=1}^R \prod_{k=1}^K \mathbf{u}_{i_k, r}^k = \mathbf{I}^T (\mathbf{u}_{i_1}^1 \otimes \cdots \otimes \mathbf{u}_{i_K}^K), \quad (2)$$

式中:  $\mathbf{u}_{i_k}^k$  是长度为  $R$  的行向量,  $\otimes$  表示哈达玛乘积。对于某个已知的隐因子矩阵集合  $\mathbf{V} = [\mathbf{U}^1, \mathbf{U}^2, \dots, \mathbf{U}^K]$ , CP 分解的目标函数可表示为:

$$L(\mathbf{V}) = \frac{1}{2} \|\mathbf{y}_{(k)} - \mathbf{U}^k (\mathbf{U}^1 \otimes \cdots \otimes \mathbf{U}^{k-1} \otimes \mathbf{U}^{k+1} \otimes \cdots \otimes \mathbf{U}^K)^T\|^2. \quad (3)$$

式(3)通过最小化原始张量与预测值的误差  $L(\mathbf{V})$  对集合  $\mathbf{V}$  进行求解, 通常使用交替最小二乘法<sup>[2]</sup>对  $\mathbf{V}$  中所有的隐因子矩阵交替更新, 直至收敛。对于其中某个隐因子矩阵  $\mathbf{U}^j$ , 更新过程可表示为:

$$\mathbf{U}^j \leftarrow \mathbf{y}_{(j)} (\bigotimes_{k \neq j}^K \mathbf{U}^k) (\bigotimes_{k \neq j}^K \mathbf{U}^{kT} \mathbf{U}^k)^\dagger. \quad (4)$$

式中,  $\dagger$  表示 Moore-Penrose 伪逆。

相比其他的张量分解方式, CP 分解的结构更加简单且计算复杂度更低。特别是在流式张量分解框架下, CP 分解能够更高效地对数据更新作出响应。而 Tensor-Train(TT)分解和 Tensor-Ring(TR)分解倾向于保留原始张量的结构信息, 更适合应用于图像处理领域。现有的流式张量分解推荐模型多是在 CP 分解基础上的改进模型, 为更好地进行对比, 本研究选用 CP 分解作为基础模型。

## 2 POSTFR 模型

在推荐系统领域, 传统的概率模型通常假设对象之间的相关性(如用户的社交关系)是已知且不变的。然而许多研究都已证实这种相关性会随时间逐渐发生变化。因此可以使用先验分布对对象间的相关性进行初步建模, 并在数据的更新过程中对其进行修正。为有效利用已知信息, 可将对象出现的频率作为隐式反馈信息加入模型更新过程。例如在某一时段内用户的评分次数可用于表征该用户的独立性程度: 如果所有用户都均衡地进行打分, 可认为这些用户都具备较弱的独立性, 更容易受到当前时段中其他人评分的影响。但如果某个用户在当前时段中存在更多的打分记录, 可认为该用户与其他用户相比更加独立, 因此在后验更新时可以相应地增加该用户的权重。

基于以上出发点, 本研究提出一种概率流式张量分解模型 POSTFR, 该模型首先使用多元高斯先验对 CP 分解的隐因子矩阵建模, 并利用 LKJ 先验建模协方差矩阵以表示对象间的相关性。进一步地, 利用数据中的隐式反馈信息对协方差进行后验更新。下面将分三部分详细介绍模型 POSTFR。

### 2.1 贝叶斯框架

贝叶斯框架是利用已知先验分布与数据似然推理参数后验的过程, 其一般形式可表示为:

$$p(\theta | D_{\text{old}} \cup D_{\text{new}}) \propto p(\theta) \cdot p(D_{\text{new}} | \theta). \quad (5)$$

式中:  $\theta$  表示需要进行推断的模型参数,  $D_{\text{old}}$  是已经获得的旧数据集合,  $D_{\text{new}}$  是当前的新数据集合,  $p(\theta)$  表示参数先验,  $p(D_{\text{new}} | \theta)$  是新数据似然,  $p(\theta | D_{\text{old}} \cup D_{\text{new}})$  表示更新后的参数后验。贝叶斯推断假设模型能够利用新数据不断对参数进行迭代更新, 且新数据与旧数据都服从于整体数据的分布。

LKJ 分布是贝叶斯分析中建模协方差矩阵的常用先验分布, 能够通过参数空间的正交化和标准化提高模型的稳定性和泛化能力。与 Gamma 分布类似, LKJ 分布中使用参数  $\eta$  控制先验分布的形态, 当  $\eta = 1$  时, LKJ 分布退化为均匀分布; 当  $\eta > 1$  时, 分布具有一个中心峰和长尾特征。

### 2.2 POSTFR 的先验设置

首先为隐因子矩阵  $\mathbf{U}^k$  中的每个行向量定义一个高斯先验  $\mathbf{u}_{i_k}^k \in \mathbf{R}^r$ ,

$$\mathbf{u}_{i_k}^k \sim N(\boldsymbol{\mu}_{i_k}^k, (\boldsymbol{\Lambda}^k)^{-1}). \quad (6)$$

隐因子矩阵的联合分布可用多元高斯分布进行定义:

表 1 符号对照表

Table 1 Symbol Legend

中文名称	符号
$K$ 阶张量 $\mathbf{y}$	$\mathbf{y} \in \mathbf{R}^{d_1 \times d_2 \times \cdots \times d_K}$
张量 $\mathbf{y}$ 中某个元素	$\mathbf{y}_i$ , 其中 $i = (i_1, i_2, \dots, i_K)$
张量 $\mathbf{y}$ 的模 $k$ 展开矩阵	$\mathbf{Y}_{(k)} \in \mathbf{R}^{d_k \times (d_1 \cdots d_{k-1} d_{k+1} \cdots d_K)}$
CP 分解的第 $k$ 个隐因子矩阵	$\mathbf{U}^k$
CP 秩	$R$
$\mathbf{U}^k$ 中第 $i$ 行向量	$\mathbf{u}_i^k \in \mathbf{R}^R$
CP 张量积	$[[ \cdot ]]$

$$p(\mathbf{V}) = \prod_{k=1}^K \prod_{i_k=1}^{d_k} N(\mu_{i_k}^k, (\mathbf{\Lambda}^k)^{-1}). \quad (7)$$

式中:  $\mu_{i_k}^k$  表示隐因子矩阵  $\mathbf{U}^k$  第  $i_k$  个行向量对象的均值;  $(\mathbf{\Lambda}^k)^{-1}$  是隐因子矩阵  $\mathbf{U}^k$  的精确度矩阵,  $\mathbf{\Lambda}^{(k)} \sim \text{LKJ}(\mathbf{W}_0, v_0)$  [13], 用于建模行向量对象之间的相关性, 其中  $\mathbf{W}_0$  表示 Cholesky 分解后的观测协方差矩阵,  $v_0$  表示协方差的离散度系数,  $v_0$  越大代表先验信息越少, 分布越扁平。

为对数据条目值  $y_i$  进行建模, 定义一个服从 Gamma 分布的方差参数  $\tau$ ,

$$p(\tau | a_0, b_0) = \text{Gam}(\tau | a_0, b_0) = \frac{b_0^{a_0}}{\Gamma(a_0)} \tau^{a_0-1} e^{-b_0\tau}. \quad (8)$$

Gamma 分布会在  $\frac{a_0}{b_0}$  附近给予  $\tau$  一个强烈的先验。为避免为模型引入太多人工信息, 设置初始值  $a_0 = b_0 =$

1。对于观测的目标值  $y_i$ , 使用高斯似然建模为:

$$p(y_i | \mathbf{V}) = N(y_i | (\mathbf{u}_{i_k}^k)^T (\mathbf{u}_{i_1}^1 \otimes \cdots \otimes \mathbf{u}_{i_{k-1}}^{k-1} \otimes \mathbf{u}_{i_{k+1}}^{k+1} \otimes \cdots \otimes \mathbf{u}_{i_K}^K), \tau). \quad (9)$$

式中,  $\otimes$  是向量的哈达玛乘积。

### 2.3 结合隐式反馈的后验更新

对于 LKJ 分布定义的协方差矩阵  $\mathbf{\Lambda}_i^k$ , 其分布可表示为:

$$p(\mathbf{\Lambda}_i^k) \propto |\mathbf{\Lambda}_i^k|^{-\eta} \prod_{i=1}^k \frac{1}{|2\pi \mathbf{\Lambda}_i^k|^{n/2}} \exp\left(-\frac{1}{2} \mathbf{x}_i^T \mathbf{\Lambda}_i^{k-1} \mathbf{x}_i\right). \quad (10)$$

式中:  $\eta$  是 LKJ 分布的形状参数;  $n$  为常数, 表示矩阵中包含的对象数量。式(10)整理之后得:

$$p(\mathbf{\Lambda}_i^k | \mathbf{U}^k) \propto |\mathbf{\Lambda}_i^k|^{-\eta} \prod_{i=1}^k \exp\left(-\frac{1}{2} \mathbf{x}_i^T \mathbf{\Lambda}_i^{k-1} \mathbf{x}_i\right). \quad (11)$$

在对隐因子矩阵的后验更新过程中, 每当获取到新的观测数据集  $S_i$ , 使用式(12)对协方差后验  $\mathbf{\Lambda}_i^{k*}$  进行更新:

$$\mathbf{\Lambda}_i^{k*} = (\mathbf{\Lambda}_i^{k-1} + \tau \cdot \sum_i \langle \mathbf{t}_{i, \rightarrow k}, \mathbf{t}_{i, \rightarrow k}^T \rangle)^{-1}, i \in S_i. \quad (12)$$

式中:  $\langle \mathbf{t}_{i, \rightarrow k}, \mathbf{t}_{i, \rightarrow k}^T \rangle = \langle \mathbf{u}_{j_1}^1 \mathbf{u}_{j_1}^{1T} \rangle \otimes \cdots \otimes \langle \mathbf{u}_{j_{k-1}}^{k-1} \mathbf{u}_{j_{k-1}}^{k-1T} \rangle \otimes \langle \mathbf{u}_{j_{k+1}}^{k+1} \mathbf{u}_{j_{k+1}}^{k+1T} \rangle \otimes \cdots \otimes \langle \mathbf{u}_{j_K}^K \mathbf{u}_{j_K}^{KT} \rangle$ ,  $\langle \cdot \rangle$  表示向量积。

LKJ 分布能够捕捉对象之间的相关性, 尤其适用于协方差矩阵的建模, 其允许在没有足够数据的情况下为模型提供先验信息。然而, LKJ 先验仍然存在一些局限性。首先在初始状态下, 假设对象间的相关性是均匀分布的, 这可能不符合真实情况; 其次, 可能无法充分学习数据中的复杂结构。因此本研究在模型中加入隐式反馈分布  $p(\mathbf{h}_i)$ 。根据贝叶斯公式, 将协方差后验  $\mathbf{\Lambda}_i^{k*}$  构造为:

$$p(\mathbf{\Lambda}_i^{k*} | \mathbf{U}^k, \mathbf{h}_i) \propto |\mathbf{\Lambda}_i^k|^{-\eta} \prod_{i=1}^k \exp\left(-\frac{1}{2} (\mathbf{x}_i^T \mathbf{\Lambda}_i^{k-1} \mathbf{x}_i + \mathbf{h}_i^T \mathbf{h}_i)\right). \quad (13)$$

式中,  $\mathbf{h}_i = [|\omega_1|_i, |\omega_2|_i, \dots, |\omega_{d_k}|_i]$  表示当前数据批次中的归一化观测数量向量,  $|\omega_{d_k}|_i$  表示第  $d_k$  个对象在此批次中的观测数量。对式(13)取对数并求导后可得  $\mathbf{\Lambda}_i^{k*}$  的更新公式:

$$\mathbf{\Lambda}_i^{k*} = (\mathbf{\Lambda}_i^{k-1} + \langle \mathbf{h}_i, \mathbf{h}_i^T \rangle + \tau \cdot \sum_i \langle \mathbf{t}_{i, \rightarrow k}, \mathbf{t}_{i, \rightarrow k}^T \rangle)^{-1}, i \in S_i. \quad (14)$$

这种方法的思想来源是, 在当前数据中对象出现的越频繁, 模型对该对象的分布就越有信心。收敛性方面, 由于高斯先验的协方差矩阵  $\mathbf{\Lambda}_i^k$  符合正定且目标函数是关于  $\mathbf{\Lambda}_i^k$  的二次函数, 可以证明式(13)中后验的对数似然是关于  $\mathbf{\Lambda}_i^k$  的凸函数, 因此满足收敛条件。

在获得  $\mathbf{\Lambda}_i^{k*}$  后, 均值后验  $\boldsymbol{\mu}_i^{k*}$  可表示为:

$$\boldsymbol{\mu}_i^{k*} = \mathbf{\Lambda}_i^{k*} (\mathbf{\Lambda}_i^{k-1} \boldsymbol{\mu}_i^k + \tau \cdot \sum_i y_i \langle \mathbf{t}_{i, \rightarrow k} \rangle), i \in S_i. \quad (15)$$

式中,  $\langle \mathbf{t}_{i, \rightarrow k} \rangle = \boldsymbol{\mu}_{j_1}^1 \otimes \cdots \otimes \boldsymbol{\mu}_{j_{k-1}}^{k-1} \otimes \boldsymbol{\mu}_{j_{k+1}}^{k+1} \otimes \cdots \otimes \boldsymbol{\mu}_{j_K}^K$ 。

此外, 还需对观测值方差  $\tau$  的控制系数  $a_0$  和  $b_0$  进行后验更新, 其后验分别记为  $a^*$  和  $b^*$ 。更新过程分别如式(16)和式(17)所示。

$$a^* = a + \frac{1}{2} |S_t|, \tag{16}$$

$$b^* = b + \frac{1}{2} \sum_i (y_i^2 - 2y_i \langle \mathbf{I}^T \mathbf{t}_i \rangle + \langle \mathbf{I}^T \mathbf{t}_i \rangle^2), i \in S_t. \tag{17}$$

式中,  $|S_t|$  为当前批次观测到的数据条目数量,  $\mathbf{t}_i = \boldsymbol{\mu}_{j_1}^1 \otimes \dots \otimes \boldsymbol{\mu}_{j_K}^K$ 。整个更新过程的伪代码如算法 1 所示。

POSTFR 模型的整体结构如图 1 所示。其中, 虚线左半部分建模  $\mathbf{y}_i$  的均值: A 部分表示对隐因子矩阵先验的建模过程, B 部分表示对融合隐式反馈信息的特征向量后验建模; 虚线右半部分建模  $\mathbf{y}_i$  的方差: C 和 D 部分分别表示对方差  $\tau$  与噪声  $\kappa$  的建模; E 部分用于计算  $\mathbf{y}_i$ 。箭头表示各部分之间的参数传递过程。相比于原有模型, POSTFR 模型的主要改进在图中用红色框出: A 部分添加了对协方差  $\sigma$  的先验建模, B 部分添加了隐式反馈信息, 用于增强特征向量的相关性建模。

在时间、空间复杂度方面, POSTFR 模型在计算过程中进行了 Cholesky 分解与 Khatri-Rao (KR) 积运算, 时间复杂度是多项式级的, 为  $O(KR^3 + KD^2)$ , 其中  $K$  表示张量模式数,  $R$  表示 CP 秩,  $D$  表示张量最大维度。而 POST 模型、BGCP 模型与 ROLCP 模型的时间复杂度均为  $O(KR^3)$ , 时间复杂度保持在同一数量级。POSTFR 模型的空间复杂度为  $O(KR^2 + KD^2 + KR D)$ , 略高于其余模型的  $O(KR^2 + KR D)$ 。

算法 1 POSTFR 算法

```

输入: 均值先验  $\mathbf{u}_{i_k}^k$ 、协方差先验  $\boldsymbol{\Lambda}_k^k$ 、观测数据集  $S_t$ 
输出: 均值后验  $\boldsymbol{\mu}_i^{k*}$ 、协方差后验  $\boldsymbol{\Lambda}_i^{k*}$ 
1: 若不存在先验, 则首先进行初始化:
2:   设置 Gamma 分布参数  $a_0 = b_0 = 1$ 
3:   使用式(8)初始化方差系数  $\tau$ 
4:   使用式(7)随机初始化隐因子矩阵均值先验
5:   使用式(8)随机初始化协方差先验
6: for 每个新的观测数据切片  $S_t$ :
7:   for 每个模式  $k$ :
8:     使用式(14)计算协方差后验  $\boldsymbol{\Lambda}_i^{k*}$ 
9:     使用式(15)计算协方差后验  $\boldsymbol{\mu}_i^{k*}$ 
10:  end for
11: 使用式(9)计算当前预测值  $\mathbf{y}_i$ 
12:  if 未达到收敛阈值: continue
13: end for
14: 使用式(16)、式(17)计算  $a^*$  和  $b^*$ 
15: 使用式(8)计算更新后的  $\tau$ 
16: 结束

```

### 3 实验

本节在 3 个不同数据集上将本研究提出的 POSTFR 模型与 POST<sup>[9]</sup>、BGCP<sup>[12]</sup> 和 ROLCP<sup>[15]</sup> 进行对比实验, 并通过消融实验对 POSTFR 模型的两方面改进效果加以验证。

#### 3.1 数据集及实验设置

实验使用的 3 个数据集分别是 Movielens 100 k 数据集、Movielens 1 M 数据集和豆瓣电影评分数据集。数据集基本信息如表 2 所示。由于豆瓣数据集中包含大量的不活跃用户, 为避免对实验结果产生影响, 首先对豆瓣数据集进行数据清洗: 将评分数量少于 10 条的用户和被评分数量少于 10 条的电影数据剔除。

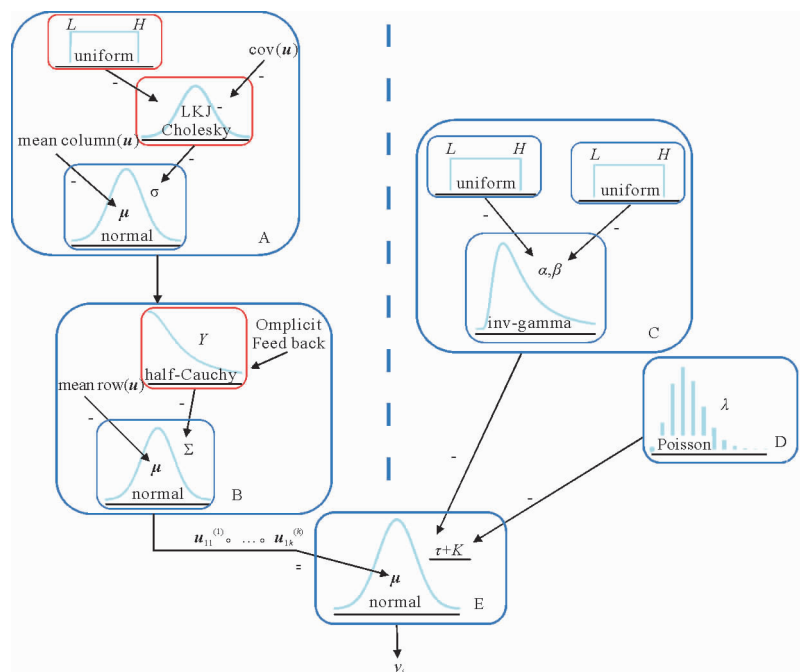


图 1 POSTFR 模型结构

Fig. 1 Structure of POSTFR

表 2 数据集基本信息  
Table 2 Basic information of datasets

数据集	用户数	电影数	评分数	稀疏度/%	统计起始时间
Movielens 100 k	943	1 682	80 000	5.04	1997-09-19
Movielens 1 M	6 040	3 883	1 000 209	4.26	2000-01-01
豆瓣电影评分数据集(清洗前)	94 532	1 762	404 972	0.24	2005-06-12
豆瓣电影评分数据集(清洗后)	8 687	1 604	223 689	1.61	2005-06-12

然后,按照时间戳先后顺序对数据重新排序,为控制张量的规模,将原本的时间戳属性转换为星期一到星期日,最终获得 3 阶(用户-电影-星期)的数据张量。按照 8 : 2 的比例划分训练集与测试集。为减少不确定性因素,将 CP 秩固定为 5,收敛阈值设置为 0.1。需要注意的是,本节中的模型均对数据切片的大小(或称数据量规模)不敏感。每个数据集的训练集都被平均分割为 20 个数据切片以模拟流式数据,在每个时间步都仅使用一个数据切片对模型进行训练。

评价指标使用推荐系统中最常用的均方根误差(root mean squared error, RMSE)和平均绝对误差(mean absolute error, MAE)。其中, RMSE 衡量模型预测值与观测值之间的距离,数值越小证明模型对观测值的预测越准确; MAE 衡量模型预测值与观测值之间绝对值的差异,同样数值越小模型效果越好。RMSE( $R_{MSE}$ )和 MAE( $M_{AE}$ )的计算方法分别如式(18)和式(19)所示:

$$R_{MSE} = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}, \tag{18}$$

$$M_{AE} = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i|. \tag{19}$$

式中,  $m$  为观测值的数量。

所有实验均在 AMD Ryzen5 2600@3.40 GHz CPU、16 GB RAM 上进行,编译平台使用 PyCharm Community Edition 2021。

### 3.2 实验结果及分析

图 2~4 分别展示 4 种算法在 Movielens 100 k 数据集、Movielens 1 M 数据集和豆瓣电影评分数据集上的 RMSE 与 MAE 对比。由图 2 可见,相比于仅对隐因子矩阵整体建模的 BGCP 模型,POST 和本研究提出的 POSTFR 由于在更细粒度上对各个对象建模,因此在较为稠密的 Movielens 100 k 数据集中获得了更好的预测结果。相比于 BGCP、POST 与 ROLCP,POSTFR 模型在 RMSE 上分别下降 5.6%、3.1%和 4.7%,在 MAE 上分别下降 5.4%、3.0%和 4.6%。在 RMSE 指标下,POSTFR 模型优势更加明显。这是由于 MAE 指标更加关注整体的误差情况,而 RMSE 指标则反映个体的偏差程度,POSTFR 模型中由于加入隐式反馈信息,因此能够更好地建模个体特征。

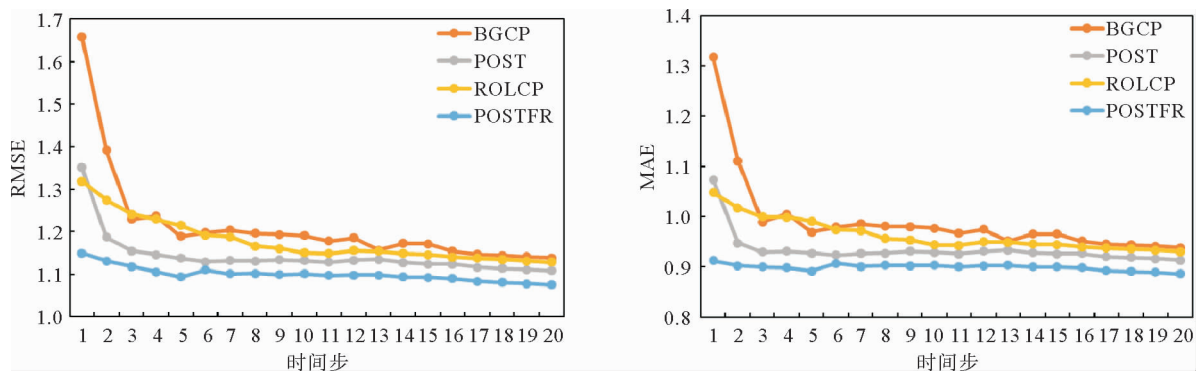


图 2 4 种算法在 Movielens 100 k 数据集上的 RMSE 与 MAE 对比  
Fig. 2 RMSE and MAE of 4 algorithms on Movielens 100 k dataset

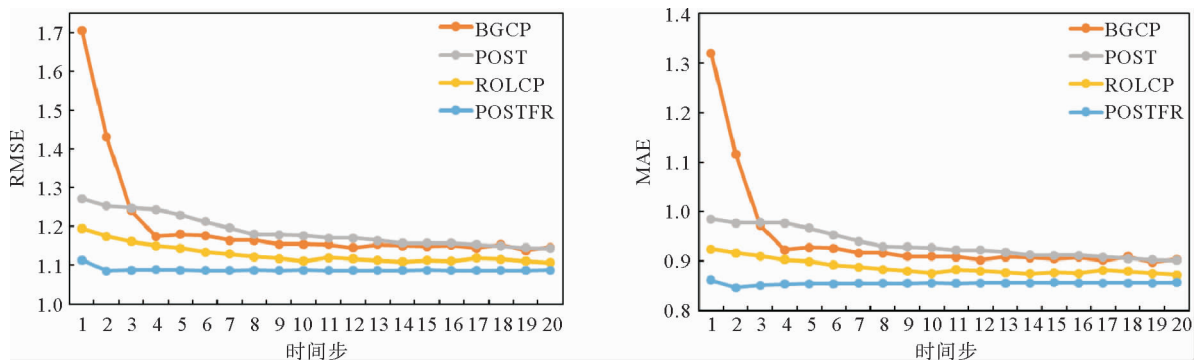


图 3 4 种算法在 Movielens 1 M 数据集上的 RMSE 与 MAE 对比

Fig. 3 RMSE and MAE of 4 algorithms on Movielens 1 M dataset

由图 3 可见,相比于 POST 模型,在 Movielens 1 M 数据集中 BGCP 在迭代初期获得了更好的效果。这是因为 Movielens 1 M 数据集比 Movielens 100 k 更加稀疏,因此关注个体模式的 POST 模型在最初的时间步中难以从数据切片中学习足够的对象信息。而由于加入协方差矩阵建模观测对象的相关性,POSTFR 模型在迭代初期就获得了较好的收敛。在更大规模的数据集中,ROLCP 由于能够学习到更多信息也取得较好的效果。相比于 BGCP、POST 和 ROLCP,POSTFR 模型的 RMSE 分别下降 6.3%、6.4% 和 2.9%, MAE 分别下降 6.3%、6.3% 和 2.9%。

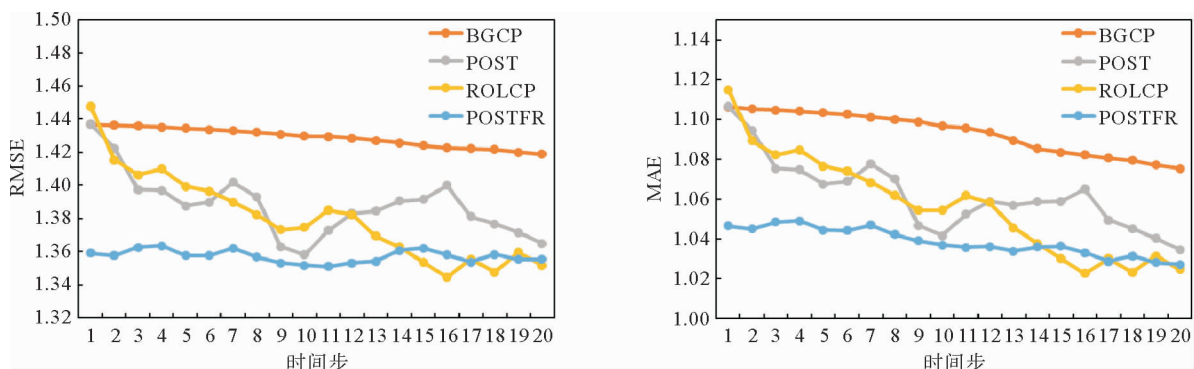


图 4 4 种算法在豆瓣电影评分数据集上的 RMSE 与 MAE 对比

Fig. 4 RMSE and MAE of 4 algorithms on Douban movie rating dataset

由图 4 可见,相比于 BGCP 和 POST,POSTFR 模型的 RMSE 和 MAE 均取得更好的预测精度。RMSE 分别下降 6.5% 和 1.3%,MAE 分别下降 6.3% 和 1.2%。而 ROLCP 在豆瓣数据集上获得了更好的效果,这是由于豆瓣评分中用户的评分受到更多因素影响,使用高斯先验难以捕获这种复杂的结构关系。但相比于其他模型,POSTFR 模型由于建模对象之间的相关性,在迭代过程中明显更加稳定。综上,在不同数据集的实验中,本研究提出的 POSTFR 模型均取得最佳或接近最佳的预测精度,并且具有更好的稳定性。

### 3.3 消融实验

为验证本研究进行的改进对模型预测性能提升的有效性,分别在 POSTFR 模型的协方差矩阵中去除 LKJ 先验或隐式反馈信息进行消融实验。在 3 个数据集上的实验结果分别如图 5~7 所示。

由图 5 可见,去除 LKJ 先验或者去除隐式反馈信息均会对结果造成不利影响。具体而言,去除 LKJ 先验的协方差项会导致在前几次迭代中难以收敛到合理的精度,并且令模型最终的精度更差。而去除隐式反馈信息则使模型在运行中期效果变差。与 POSTFR 相比,去除 LKJ 先验或去除隐式反馈信息分别使模型的 RMSE 升高 4.8% 和 3.7%,MAE 升高 4.7% 和 3.8%。由图 6 可以得出与图 5 类似的结论。另外,如同在 3.2 节中所提到的,相比于 MAE,在 RMSE 上两种改进方法的效果差异更大,这说明添加协方差矩阵更

有利于消除结果中的异常值。去除 LKJ 先验和去除隐式反馈信息在 Movielens 1 M 数据集上分别使模型的 RMSE 升高 5.5% 和 7.8%, 使 MAE 升高 5.3% 和 8.0%。

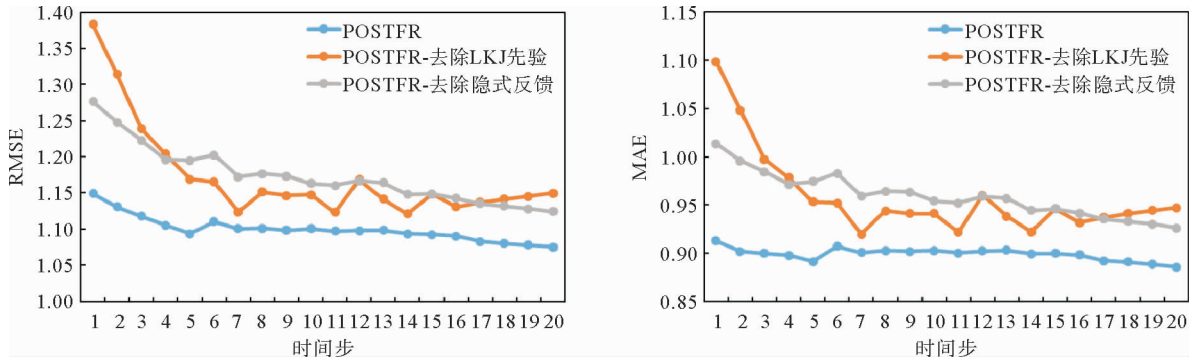


图 5 POSTFR 在 Movielens 100 k 数据集上的消融实验结果

Fig. 5 Ablation experiment results of POSTFR on Movielens 100 k dataset

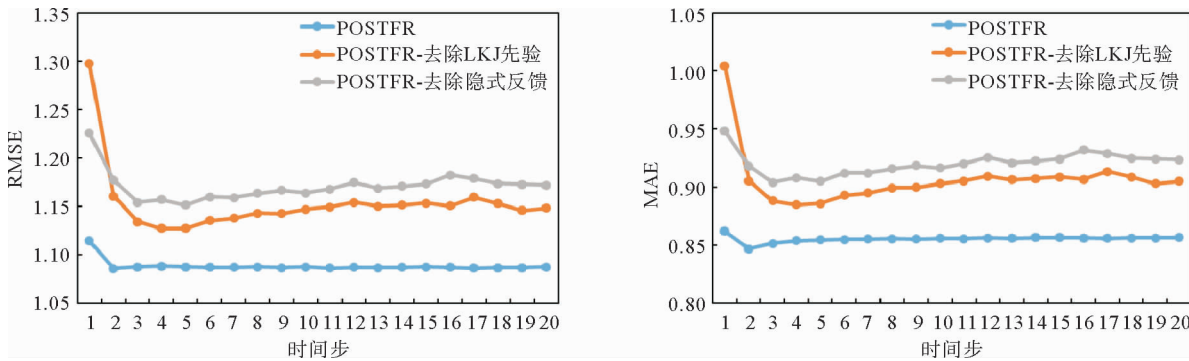


图 6 POSTFR 在 Movielens 1 M 数据集上的消融实验结果

Fig. 6 Ablation experiment results of POSTFR on Movielens 1 M dataset

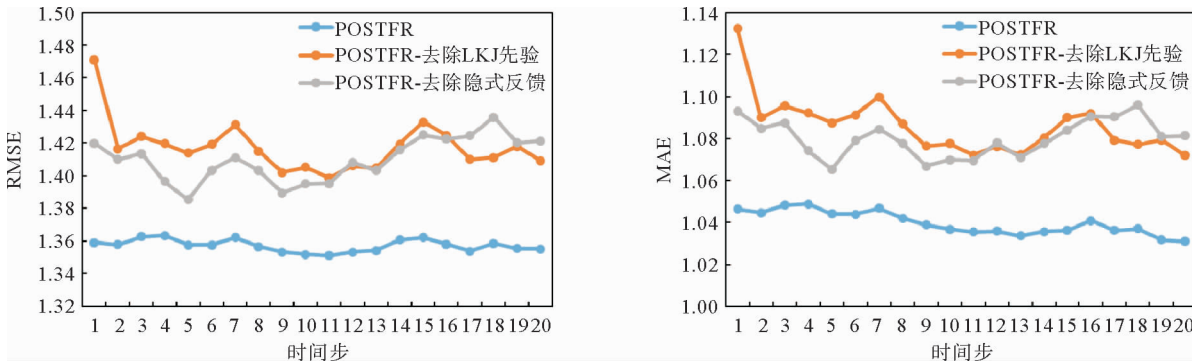


图 7 POSTFR 在豆瓣电影评分数据集上的消融实验结果

Fig. 7 Ablation experiment results of POSTFR on Douban movie rating dataset

由图 7 可见,综合具有 LKJ 先验的协方差项和在更新过程中结合隐式反馈信息,可以提高模型的预测准确性,同时使模型在迭代初期较好的收敛,减少由于当前观测值的偏差带来的预测精度波动。与 POSTFR 相比,去除 LKJ 先验或隐式反馈信息在豆瓣电影评分数据集上分别使模型的 RMSE 升高 3.9% 和 4.7%, 使 MAE 升高 3.6% 和 5.3%。

综上,在概率流式分解模型中加入隐式反馈信息和使用 LKJ 先验建模对象间相关性的方法是有效的。



#### 4 总结与展望

针对现有概率流式分解模型缺乏对对象间的相关性建模以及对隐式反馈信息利用的问题,提出一种使用协方差矩阵建模对象间相关性并且利用隐式反馈信息对后验进行更新的概率流式张量分解模型 POST-FR,模型对隐因子矩阵中的每个对象使用高斯先验,并使用具有 LKJ 先验的协方差矩阵对对象间的相关性建模。而在协方差矩阵的更新过程中,结合当前批次的隐式反馈信息以提高模型收敛效果。基于对 CP 分解的并行化是十分自然的认识,下一步工作将在两个方面展开:一是将模型的更新任务分配到多个节点进行并行计算以提升效率,二是将概率分解与 TT 分解、TR 分解等张量分解方式相结合。

#### 参考文献:

- [1] HITCHCOCK F L. The expression of a tensor or a polyadic as a sum of products[J]. *Journal of Mathematical Physics*, 1927, 6(1): 164-189.
- [2] CARROLL J D, CHANG J J. Analysis of individual differences in multidimensional scaling via an n-way generalization of “Eckart-Young” decomposition[J]. *Psychometrika*, 1970, 35(3): 283-319.
- [3] ACAR E, KOLDA T G, DUNLAVY D M, et al. Scalable tensor factorizations for incomplete data[J]. *Chemometrics and Intelligent Laboratory Systems*, 2011, 106(1): 41-56.
- [4] SMITH S, HUANG K, SIDIROPOULOS N D, et al. Streaming tensor factorization for infinite data sources[C]// 2018 SIAM International Conference on Data Mining. San Diego: SIAM Press, 2018: 81-89.
- [5] MARDANI M, MATEOS G, GIANNAKIS G B. Subspace learning and imputation for streaming big data matrices and tensors[J]. *IEEE Transactions on Signal Processing*, 2014, 63(10): 2663-2677.
- [6] ZHOU S, VINH N X, BAILEY J, et al. Accelerating online CP decompositions for higher order tensors[C]// 22nd ACM SIGKDD International Conference. San Francisco: Association for Computing Machinery, 2016: 1375-1384.
- [7] SONG Q, HUANG X, GE H C, et al. Multi-aspect streaming tensor completion[C]// 23rd ACM SIGKDD International Conference. Halifax: Association for Computing Machinery, 2017: 13-17.
- [8] NAJAFI M, HE L, YU P S. Outlier-robust multi-aspect streaming tensor completion and factorization[C]// 28th International Joint Conference on Artificial Intelligence. Macao: AAAI Press, 2019: 3187-3194.
- [9] DU Y, ZHENG Y, LEE K C, et al. Probabilistic streaming tensor decomposition[C]// 2018 IEEE International Conference on Data Mining. Singapore: IEEE Press, 2018: 99-108.
- [10] BIGOT J, DELEDALLE C. Low-rank matrix denoising for count data using unbiased Kullback-Leibler risk estimation[J/OL]. *Computational Statistics & Data Analysis*, 2022, 169. DOI: 10.1016/j.csda.2022.107423.
- [11] ZHENG Y. Probabilistic streaming tensor decomposition with side information[D]. Salt Lake City: The University of Utah, 2019: 1-26.
- [12] CHEN X Y, HE Z C, SUN L J. A Bayesian tensor decomposition approach for spatiotemporal traffic data imputation[J]. *Transportation Research Part C: Emerging Technologies*, 2019, 98(1): 73-84.
- [13] LEWANDOWSKI D, KUROWICKA D, JOE H. Generating random correlation matrices based on vines and extended onion method[J]. *Journal of Multivariate Analysis*, 2009, 100(9): 1989-2001.
- [14] BACH T X, ANH N D, LINH N V, et al. Dynamic transformation of prior knowledge into Bayesian models for data streams[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 35(4): 3742-3750.
- [15] THANH L T, ABED-MERAIM K, TRUNG N L, et al. A fast randomized adaptive CP decomposition for streaming tensors[C]// 2021 IEEE International Conference on Acoustics, Speech and Signal Processing. Toronto: IEEE Press, 2021: 2910-2914.

(责任编辑:傅游)