



单位代码 10635
学 号 112020334202216

西南大學

专业学位硕士学位论文

基于稀疏张量回归的高维数据预测

论文作者：向邱燕

指导教师：袁建军

专业学位类别：应用统计

专业领域：

提交论文日期： 2022 年 4 月 14 日

论文答辩日期： 2022 年 5 月 28 日

学位授予单位：西南大学

中 国 • 重 庆

2022 年 5 月

独创性声明

学位论文题目：基于稀疏张量回归的高维数据预测

本学位论文是作者在导师指导下独立完成的研究工作及取得的研究成果，恪守学术诚信，遵守学术准则。对本研究及学位论文撰写曾做出贡献的老师、朋友、同仁在文中作了明确说明并表示衷心感谢。

学位论文作者签名：阿邱燕

导师签名：袁建峰

签名日期：2022年6月13日

签名日期：2022年6月13日

学位论文版权使用授权书

本学位论文作者完全了解西南大学有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权西南大学研究生院可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

本论文公开时间：获学位当年；推迟1年；推迟2年。

学位论文作者签名：阿邱燕

导师签名：袁建峰

签名日期：2022年6月13日

签名日期：2022年6月13日

目 录

摘 要.....	I
Abstract.....	III
第 1 章 绪论.....	1
1.1 研究背景和意义.....	1
1.2 国内外研究现状.....	2
1.3 本文研究内容及结构安排.....	4
第 2 章 相关理论基础.....	7
2.1 张量定义.....	7
2.2 张量变换.....	8
2.3 张量运算.....	9
2.4 张量分解.....	9
2.5 张量岭回归.....	12
2.6 基于拉普拉斯的贝叶斯张量回归.....	16
第 3 章 稀疏张量回归算法.....	19
3.1 基于 CP 分解的张量 Lasso 回归.....	20
3.2 模拟实验.....	23
3.3 案例分析.....	28
3.4 本章小结.....	29
第 4 章 贝叶斯张量回归算法.....	31
4.1 基于拉普拉斯先验的贝叶斯张量回归.....	32
4.2 模拟实验.....	38
4.3 案例分析.....	41
4.4 本章小结.....	42
第 5 章 基于张量回归算法的用户行为预测.....	45
5.1 描述性统计分析.....	46
5.2 基于张量回归模型的用户行为预测分析.....	50
5.3 本章小结.....	62
第 6 章 总结与展望.....	65
6.1 总结.....	65
6.2 展望.....	65
参考文献.....	67
攻读硕士学位期间所做的工作.....	71

致谢..... 73

专用术语注释表

符号说明:

\circ	向量外积
\otimes	Kronecker 积
$\operatorname{argmin}()$	使括号内式子达到最小时, 变量的取值
prox	近端映射算子

缩略词说明:

英文缩写	英文全称	中文全称
CP 分解	CANDECAMP/PARAFAC Decomposition	CP 分解
TT 分解	Tensor Train Decomposition	张量链分解
TR 分解	Tensor Ring Decomposition	张量环分解
QR 分解	Orthogonal Triangular Decomposition	正交三角分解
ADMM	Alternating Direction Method of Multipliers	交替方向乘子法
MCMC	Markov Chain Monte Carlo	马尔科夫链蒙特卡洛法
SNR	Signal to Noise Ratio	信噪比水平

基于稀疏张量回归的高维数据预测

应用统计专业 硕士研究生 向邱燕

指导教师 袁建军 副教授

摘要

随着智能技术的发展，高维数据日益普遍，如图像、视频、社交网络关系和用户行为数据等，被广泛应用在特征预测、视频分类和推荐系统关系挖掘等各个领域。而传统的机器学习算法需要先对数据进行向量化或矩阵化处理，这通常会破坏高维数据携带的内部结构信息，并常常会带来过拟合的新问题。近年来，将张量形式的数据应用到经典的线性回归模型中引起了广泛关注，一方面，可以尽可能地保留数据的结构信息，达到更优的预测效果，另一方面，结合使用张量分解技术，可以有效简化模型并减少待估参数的数量。本文对于预测变量和响应变量均为任意阶张量数据的问题进行梳理研究，主要工作有以下三点：

第一，对于频率学派视角下的稀疏张量正则回归问题，梳理了近年来提出的两种张量岭回归算法，分别为基于 CP 分解的张量岭回归和基于 TT 分解的张量岭回归，并在此基础上，提出了基于 CP 分解的张量 Lasso 回归算法。首先，通过假设张量回归系数具有低 CP 秩结构，并施加 l_1 惩罚项，构建张量 Lasso 回归算法；其次，在理论上对回归系数的求解进行推演，采用了广义交替方向乘子法进行求解；最后，在数据集上对三个回归算法进行对比分析，结论表明本文提出的基于 CP 分解的张量 Lasso 回归算法对标已有的两种张量岭回归算法在多项评价指标上均有一定幅度的提升，如在小样本下相关系数提升了 14%。

第二，对于贝叶斯学派下的张量回归问题，阐述了张量回归系数基于高斯分布的贝叶斯张量回归算法，并提出了回归系数基于拉普拉斯先验的贝叶斯张量回归算法。首先，为以回归系数为主的参数选定合适的先验分布；其次，推导各个参数的后验分布形式，使其后验分布具备可采样的现实性，使用 MCMC 方法中的吉布斯采样算法进行高效采样；最后，在数据集上考察了模型的表现，结论表明基于拉普拉斯分布的贝叶斯张量回归算法比现有的基于高斯分布的贝叶斯张量回归表现得更好，95%置信区间覆盖率均有 1%-2%的提升。

第三，将本文提出的张量回归算法应用在阿里天池所共享的飞猪平台用户行为数据集中，分别从用户特征、商品特征和时间维度出发，以三个视角为例对用户行为数据进行了建模预测分析，取得了较为理想的预测效果。

关键词：高维数据，张量回归，贝叶斯回归，吉布斯采样，用户行为预测

Sparse Tensor Regression for High-Dimensional Data Prediction

Qiuyan Xiang

Directed by Assoc. Prof. Jianjun Yuan

Abstract

With the development of intelligent technology, high-dimensional data, such as images, videos, social network relationships and user behaviors, are becoming increasingly common. They are widely used in various fields, such as feature prediction, video classification, and recommendation system relationship mining. Traditional machine learning algorithms need to vectorize or matrix the data first, which usually destroys the internal structural information carried by high-dimensional data, and often leads to overfitting. In recent years, the application of using tensors in the classical linear regression model has attracted significant attention. On the one hand, the structural information of data can be preserved as much as possible to achieve better prediction accuracy. On the other hand, tensor decomposition techniques can effectively simplify the regression model and reduce the number of parameters to be estimated. This research sorts out and researches the sparse regression algorithms that predict a tensor from a tensor. The main research results are as follows:

First, for the sparse tensor regression model, this research introduces two ridge tensor regression algorithms proposed in recent years, namely ridge tensor regression based on CP decomposition and ridge tensor regression based on tensor train decomposition. On this basis, this research proposes a Lasso tensor regression algorithm based on CP decomposition. Firstly, the Lasso tensor regression is constructed by assuming that the tensor coefficient has a low CP rank structure and imposing the l_1 penalty term. Secondly, the solution of the coefficient of the regression model is theoretically deduced, and an augmented alternating direction multiplier method is used to solve this problem. Finally, three regression algorithms are compared

on the datasets. The conclusion shows that the Lasso tensor regression algorithm based on CP decomposition proposed in this research has a certain degree of improvement in several evaluation indicators compared with the two existing tensor ridge regression algorithms. For example, the correlation coefficient increased by 14% in the case of small samples.

Second, for the tensor regression model under the Bayesian school, the Bayesian tensor regression algorithm based on Gaussian distribution is introduced. On this basis, this research proposes a Bayesian tensor regression algorithm based on Laplace prior distribution. Firstly, we choose appropriate prior distributions for parameters dominated by coefficients. Secondly, derive the posterior distribution form of each parameter to sample the posterior distribution realistically. Using Gibbs sampling, which is one of the MCMC methods, to sample efficiently. Finally, the performance of the two algorithms is examined on the datasets, and the conclusion shows that the Laplace Bayesian tensor regression algorithm performs better than the existing Gaussian Bayesian tensor regression. The coverage of the 95% confidence interval has been improved by 1%-2%.

Third, the tensor regression algorithms proposed in this research are applied to predict the user behaviors of the Fliggy platform, which is shared by Ali Tianchi. From the user characteristics, product characteristics and time dimensions, this research takes three perspectives as an example to predict and analyze user behaviors, and achieves a relatively ideal prediction accuracy.

Key words: High-dimensional data, Tensor regression, Bayesian regression, Gibbs sampling, User behavior prediction

第1章 绪论

1.1 研究背景和意义

随着大数据时代的发展，许多应用程序产生了大量的高维数据，如神经成像、视频音频和社交网络关系等。以电影推荐系统数据为例，可以很自然地将其表示为三阶张量形式：用户-电影-偏好程度，其中偏好程度可以用观看、评论、打分等指标加以衡量。如果使用传统的机器学习算法进行建模，需要先对数据进行向量化或矩阵化处理，而简单展开的方式通常会破坏数据的内部结构，导致数据的有效信息丢失，并且常常伴随着过拟合现象的出现。张量则为这样的高维数据提供了自然的表现形式，使用张量对数据进行表示和存储能够保留数据的结构信息，因此近年来直接对张量数据进行建模分析在各个领域得到了广泛的应用，如磁共振成像分析^[1-3]、图像分类^[4-6]、人脸识别^[7,8]和社交行为挖掘^[9-11]等。

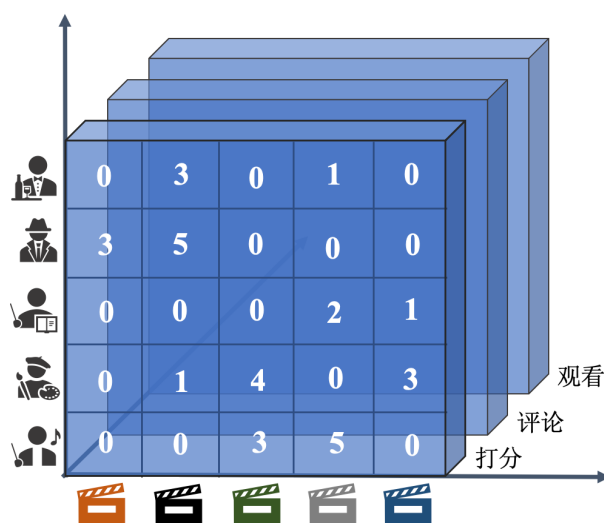


图 1-1 电影推荐系统

Figure 1-1 Movie recommendation system

现实中的大部分张量数据具有稀疏性，如图 1-1 电影推荐系统数据所示，用户对于不同的电影具有不同的偏好，当用户和电影的数量很大时，绝大多数用户只看过少部分电影，并对其进行打分、评论等行为，而绝大部分电影也仅仅只收到部分用户的观影反馈。数据稀疏性为张量问题的求解带来了便利，可以通过张量分解技术对稀疏张量进行低秩估计，使用尽可能简单的结构来近似逼近原始的复杂张量，这不仅使得模型更加符合现实解释意义，而且很大程度上解决了数据运算量大的问题。目前最主要的张量分解技术是 CP 分解、Tucker 分

解和高阶奇异值分解,都已经得到了广泛的应用^[12-14]。近年来,张量网络技术在降维领域的应用也催生出了张量链分解和张量环分解,并且也都取得了相对可观的重构效果^[15]。综上所述,使用张量形式的数据进行建模分析不仅在实际意义上具有合理性,同时在张量分解技术的背景下也具有现实可行性。

虽然使用张量分解技术进行低秩重构已经得到了较为充分的研究和发展,但是把张量数据和张量分解技术应用到经典的回归模型中,使用张量数据进行预测分析的算法研究相对而言还有较大的空间。近年来,学者们尝试直接对张量回归系数本身进行稀疏约束,并结合张量分解技术进行求解,取得了令人惊喜的效果^[16]。综上所述,基于张量数据的回归预测分析不仅应用前景广阔,并且已经取得了一定程度的发展。本文将以内国外学者已有的研究成果为指导,对基于张量数据的高维预测问题进行梳理和进一步研究。

1.2 国内外研究现状

1.2.1 张量分解

张量分解被认为是学习张量结构信息最有用的技术之一,起源于 1927 年,目前已经取得了较为充分的发展。1927 年, Hitchcock 提出张量分解的概念,其思想是将高阶张量分解为一组简单的成分来进行表示,比如向量、矩阵或是更低阶的张量^[17]。但 Hitchcock 仅仅提出了简单的基于多元形式的张量分解,具有一定局限性,直到 1963 年,经典的 Tucker 分解正式登场,将一个复杂的高阶张量分解为一个核心张量乘以每个维度下的因子矩阵的形式^[18]。1970 年, Carroll 和 Harshman 在其论文中介绍了 CANDECOMP 和 PARAFAC 分解方法,简称 CP 分解^[19,20],其主要思想是将一个高阶张量分解为多个秩一张量的总和, Kiers 和 Mocks 也在其论文中使用了 CP 分解来分析大脑成像数据^[21,22]。近年来,张量网络也被用于张量数据的降维上,它利用矩阵乘积态 (MPS) 的方法将一个高阶张量分解为多个低阶张量的归并。例如, Oseledets 在 2011 年提出了张量链分解^[23],通过一系列低秩张量链式乘积的组合来表示高阶张量^[24],随后,张量环分解也引起了学者们的关注^[25,26],通过一系列简单张量以圆形多重线性乘积的形式来表示张量,结论表明张量网络也能够较好地学习到数据中的结构信息。

从上述张量分解的文献梳理中可以发现,在张量分解领域,相关研究起步早且发展较为成熟。不仅有经典且广为应用的 CP 分解、Tucker 分解等,还有近年来引起关注的张量网络分解,如张量链分解和张量环分解等较为新颖的技术。

1.2.2 频率学派下的张量回归

由于张量分解技术的日渐成熟,许多领域已经将张量分解技术运用在回归模型中,并且得到了切实有效的结论。例如,稀疏正则低秩张量回归算法假设回归系数具有低 Tucker 秩,利用回归模型来分析基因组数据,在数据的预测上取得了较好的效果^[16]。Zhou 使用张量形式的神经影像数据来预测医学上的临床症状,实现了高维数据下的高效回归^[27],Li 提出了基于 Tucker 分解的广义线性模型,在磁共振图像上进行分析探究^[28]。Zhang 使用了基于 CP 分解的逻辑回归解决视频分类任务^[29]。Kossaifi 将张量回归应用到神经网络中,进行图像特征预测,不仅保留了数据本身的多线性结构,还有效地减少了全连接层的待估参数^[30]。通过文献调研,可以发现张量回归模型应用场景广阔,在多个领域都取得了较好的表现。

尽管张量回归的应用领域广泛,但以上的大多数方法只能处理预测变量为张量形式或是响应变量为张量形式的数据,而对于预测变量和响应变量均为任意维度张量的回归问题研究较少。但如果按照广泛使用的方法,即将原始的高阶张量元素堆叠在一个维度较低的矩阵或向量形式中再进行模型建立,会导致维数灾难、结构信息丢失和过度拟合等新的问题。为此,对于预测变量和响应变量均是高维张量数据的回归问题,Lock 首先提出了具有低 CP 秩结构的张量对张量 (Tensor On Tensor) 回归模型^[31],它可以从一个任意维度的张量自变量去预测任意维度的张量响应变量。同时,为了防止数据过度拟合,对系数张量施加了 l_2 惩罚项,建立了基于 CP 分解的张量岭回归模型。类似地,其他张量分解方法也被用于张量对张量的回归问题上,如 Gahrooei 假设张量系数具有低 Tucker 秩结构,并且对系数施加 l_2 惩罚项,建立了多元张量对张量回归模型^[32]。另外,张量网络,如张量链分解和张量环分解,也被使用在张量对张量的回归模型上,Liu 假设回归系数具有张量链 (Tensor Train) 或是张量环 (Tensor Ring) 结构,并对其施加了 l_2 惩罚项,分别建立了基于 TT 分解的张量岭回归模型^[33]和基于 TR 分解的张量岭回归模型^[34]。以上基于各种分解形式而建立的张量岭回归模型都在模拟数据和真实数据集上取得了较为可观的效果。

在非张量形式的数据上,除了上述的岭回归,被广泛使用的还有 Lasso 回归,即对回归系数施加 l_1 惩罚项。Lasso 回归可以取得比岭回归更为稀疏的解,通常也更为稳健,具有更强的泛化能力。但在预测变量和响应变量均为任意阶张量形式的数据下,Lasso 回归的研究相对较少,本文将以此为方向进行张量 Lasso 回归算法的构建。

1.2.3 贝叶斯学派下的张量回归

在前述的频率学派中，待估参数被看作是一个已知的固定值，通过极大似然的思想进行估计。在贝叶斯学派中，待估参数被看作是一个随机变量，有自己的分布形式和统计特征。尽管两个学派在对参数的定义上具有质的区别，但其实是一脉相承、相互关联的。如在低维数据中，为回归系数假设服从均值为0高斯先验分布时，即希望系数能够在0附近随机波动，与频率学派下的正则化技术具有异曲同工之处。

在低维数据中，贝叶斯回归模型已经发展得相对比较完善，基于高斯分布的贝叶斯回归和基于拉普拉斯分布的贝叶斯回归在很早就被提出^[35,36]，时至今日，仍然被广泛应用于各个领域的问题研究当中。如 Yang 将基于高斯先验的贝叶斯岭回归与集成经验模型分解相结合，进行了短期风速预测，取得了较高的预测准确性^[37]。在对于临床数据的缺失问题，Yamaguchi 根据可能存在的影响因素对临床表现进行缺失值预测^[38]。除此之外，Ge 在疾病预测问题中应用了具有连续收缩先验的贝叶斯回归模型，使得其表现具有更强的鲁棒性，在六种复杂疾病的预测中表现出较高的预测准确性^[39]。

近年来，由于张量数据的普遍存在，许多学者将低维的贝叶斯回归拓展到高维张量数据中。如 Guhaniyogi 提出贝叶斯张量回归模型，其假设回归系数服从于多维收缩先验，在不同稀疏程度和信号强度下进行了模拟分析，在神经影像领域进行了预测分析^[40]。Hoff 对社交网络的张量数据进行了探究，结论表明在多变量关系的数据中，线性贝叶斯张量回归模型具有较强的解释能力^[41]。

对于预测变量和响应变量均为任意阶张量的贝叶斯回归模型，其研究相对匮乏，Lock 在提出了基于 CP 分解的岭回归时，也完善了其思路下的贝叶斯推理框架，为回归系数假设了高维形式下的球形高斯先验，提出了基于高斯分布的贝叶斯张量回归模型^[31]。基于此，本文也延承其思路，给出基于 CP 分解的 Lasso 回归下的贝叶斯推理框架，对参数进行不确定性量化。在低维数据中，假设回归系数服从于拉普拉斯先验，其最大后验估计与 Lasso 回归极大似然估计一致，本文将扩展至高维张量数据中，建立基于拉普拉斯分布的贝叶斯张量回归模型。

1.3 本文研究内容及结构安排

1.3.1 研究内容

从上述文献综述中可以发现，国内外学者对于预测变量和响应变量均为任意阶张量的回归模型的研究还有进一步研究的空间，受此启发，本文将从频率

学派和贝叶斯学派的角度研究预测变量和响应变量均为任意阶张量的回归模型。首先，在频率学派的体系中，通过对回归系数施加 l_1 惩罚项，构建张量 Lasso 回归模型，可以有效解决维数灾难、结构信息丢失和过度拟合等问题。其次，在贝叶斯推理框架下，建立回归系数具有拉普拉斯先验的贝叶斯张量回归模型，通过对以回归系数为主的各个参数进行合理的先验分布假设实现对参数的自动选择，有效地降低人为调参的困难。最后，将本文所提出的算法模型应用在阿里天池所共享的用户行为数据集上进行预测分析，利用商品和用户的特征信息对用户行为进行预测，为平台和商家的精准营销提供思路。

本文研究框架如图 1-2 所示。

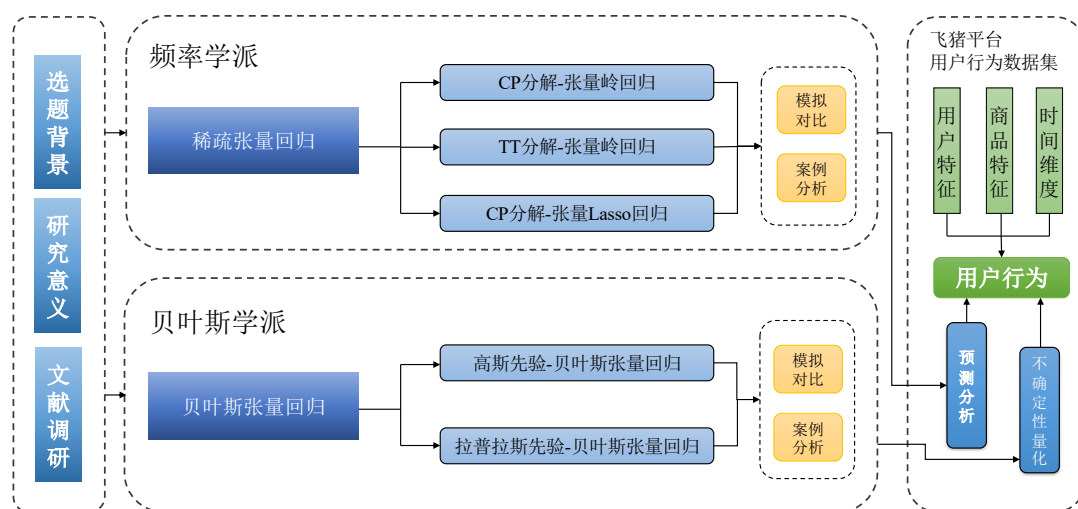


图 1-2 研究框架

Figure 1-2 Research framework

1.3.2 结构安排

第一章为绪论，简要介绍了选题背景、研究意义和研究现状，梳理了张量分解的发展历程，成熟的张量分解技术为后续回归问题的求解奠定了坚实的技术基础。另外，分别梳理了频率学派和贝叶斯学派下张量回归的发展及应用领域，确定可供深入研究的方向。

第二章为相关理论基础，由于张量数据相对比较新颖，其变换形式复杂多样，有必要对其定义、变换、运算和分解方法加以说明。

第三章详细阐述了本文提出的基于 CP 分解的张量 Lasso 回归算法，并在数据集上与近年来两个高效回归算法进行对比分析，评价预测效果，分别为基于 CP 分解的张量岭回归和基于 TT 分解的张量岭回归。

第四章介绍了贝叶斯学派下的张量回归思想，阐述了本文所提出的基于拉普拉斯先验的贝叶斯张量回归算法，并将其在数据集上进行比较分析。

第五章为实际的案例分析，数据集选自阿里天池所共享的电商行业用户行为数据。首先，对数据进行了描述性和探索性分析，了解其本身的分布情况；其次，以三个视角为例，建立了张量回归模型，分别根据用户特征和商品特征对用户行为次数进行预测分析；最后，从贝叶斯角度给出用户行为次数预测值的后验分布，对预测值的不确定性进行量化。

第六章为总结与展望，系统总结了本文进行的主要工作和相应的结论，并在此基础上展望了进一步研究的方向。

1.3.3 创新点

本文创新点有以下三点，前两点为模型上的创新，第三点为应用上的创新：

第一，针对频率学派下稀疏张量回归问题，提出了基于 CP 分解的张量 Lasso 回归模型。首先，该模型假设张量回归系数具有低 CP 秩结构，并对其施加 l_1 惩罚项，构建张量 Lasso 回归模型；其次，在理论上对回归系数的求解进行推演，采用了广义交替方向乘子法进行求解；最后，在数据集上进行对比分析，结论表明本文提出的基于 CP 分解的张量 Lasso 回归算法对标已有的两种张量岭回归算法在多项评价指标上均有一定幅度的提升。

第二，针对贝叶斯学派下稀疏张量回归问题，提出了张量回归系数基于拉普拉斯先验的回归模型。首先，通过假设回归系数服从于拉普拉斯先验分布来构建贝叶斯推理框架；其次，推导各个参数的后验分布形式，使其后验分布具备可采样的现实性，使用马尔科夫链蒙特卡洛方法中的吉布斯采样算法对参数后验分布进行高效采样；最后，在数据集上考察了模型的表现，结论表明基于拉普拉斯分布的贝叶斯张量回归比基于高斯分布的贝叶斯张量回归表现得更好。

第三，将张量回归算法应用在阿里天池所共享的飞猪平台的用户行为数据集中，对于用户行为数据进行了统计分析并进行了模型拟合，预测的准确率最高达到了 98.11%。

第 2 章 相关理论基础

本章将对文中出现的数学符号进行说明，并介绍必要的张量变换、运算及分解方法。在本文中，花体字母 (\mathcal{A}) 表示张量，大写加粗字母 (\mathbf{A}) 表示矩阵，小写加粗字母 (\mathbf{a}) 表示向量，小写字母 (a) 表示标量。

2.1 张量定义

张量是指多维数组，如 K 阶张量可以表示为 $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_K}$ 。以三阶张量为例，如图 2-1 所示，该张量有三个维度的索引，分别为 I_1 、 I_2 和 I_3 ，通过这三个索引 $[i_1, i_2, i_3]$ 的具体值，可以唯一表示张量中的一个元素 $a_{i_1 i_2 i_3}$ 。可以看出一阶张量即向量，二阶张量即矩阵，三阶或更高阶的张量被统称为高阶张量。

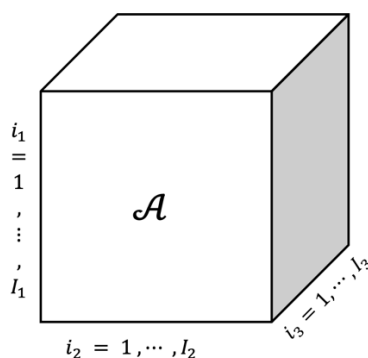
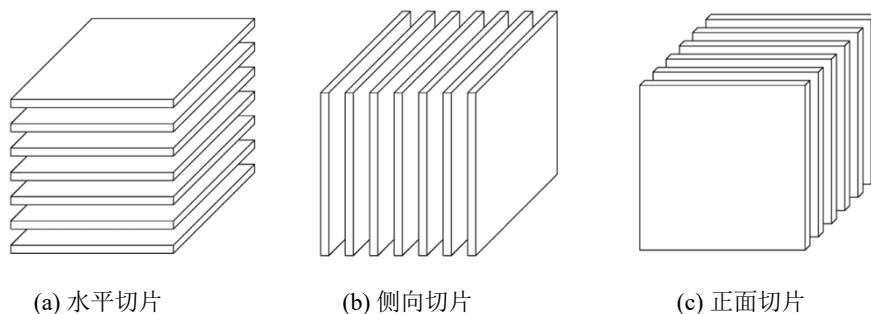


图 2-1 三阶张量

Figure 2-1 A third-order tensor

对于三阶张量部分元素更为简洁的表示有切片和纤维两种，常见的切片方式有水平切片、侧向切片和正面切片三种。水平切片如图 2-2 (a) 所示，当固定 I_1 这个维度，一个三阶张量就沿着该方向被切割为 I_1 个矩阵，每个矩阵表示为 $\mathcal{A}_{i_1::}$ ，其中符号 $:$ 用于表示该维度上的所有元素。固定不同的维度可以实现各个方向上的切片操作。



(a) 水平切片

(b) 侧向切片

(c) 正面切片

图 2-2 三阶张量切片

Figure 2-2 Slices of a third-order tensor

而当同时固定两个维度时，如固定前两个维度，即管纤维，可以使用符号 $\mathcal{A}_{i_1 i_2}$ 表示。此时，每一个 $\mathcal{A}_{i_1 i_2}$ 代表一个向量(如同一根纤维)。有三种形式不同的纤维，如图 2-3 所示，分别为列纤维、行纤维和管纤维。

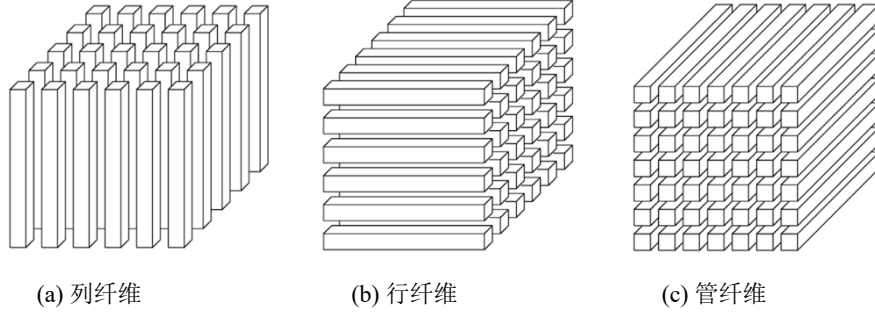


图 2-3 三阶张量纤维

Figure 2-3 Fibers of a third-order tensor

2.2 张量变换

对于张量数据，常做的变换有向量化和矩阵化，由于切片和切管的方式有多种，本节给出本文使用的具体公式，也是最为广泛使用的公式。

2.2.1 向量化

对于一个 K 阶张量，将其向量化是指将所有元素堆叠在一个维度，通过向量化算子 $vec(\cdot)$ 将张量展开为向量，其定义如式 (2.1) 所示。

$$vec(\mathcal{A})[i_1 + \sum_{k=2}^K (\prod_{l=1}^{k-1} I_l)(i_k - 1)] = \mathcal{A}[i_1, \dots, i_K] \quad (2.1)$$

2.2.2 模- j 展开

对于一个 K 阶张量，将其矩阵化是指将所有元素平铺成一个矩阵，以 K 阶中的某一维度作为矩阵的第一个维度，剩余的所有维度平铺成第二个维度。例如，对于张量 $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_K}$ ，进行模- j 展开，可以得到矩阵 $\mathbf{A}_{[j]} \in \mathbb{R}^{I_j \times I_{-j}}$ ，如式 (2.2) 所示。其中， $I_{-j} = I_1 \times \dots \times I_{j-1} \times I_{j+1} \times \dots \times I_K$ 。

$$\mathbf{A}_{[j]} \left[i_j, 1 + \sum_{k=1, k \neq j}^K \left((i_k - 1) \prod_{m=1, m \neq j}^{k-1} I_m \right) \right] = \mathcal{A}[i_1, \dots, i_K] \quad (2.2)$$

2.2.3 j -展开

张量的 j -展开取自文献^[33]，对于一个 K 阶张量 $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_K}$ ，对其进行 j -展开可以得到矩阵 $\mathbf{A}_{(j)} \in \mathbb{R}^{(I_1 \dots I_j) \times (I_{j+1} \dots I_K)}$ ，如式 (2.3) 所示。

$$\mathbf{A}_{(j)} \left[1 + \sum_{k=1}^j \left((i_k - 1) \prod_{m=1}^{k-1} I_m \right), 1 + \sum_{k=j+1}^K \left((i_k - 1) \prod_{m=j+1}^{k-1} I_m \right) \right] = \mathcal{A}[i_1, \dots, i_K] \quad (2.3)$$

2.3 张量运算

2.3.1 收缩张量积

对于两个任意阶的张量 $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_K \times P_1 \times \dots \times P_L}$ 和 $\mathcal{B} \in \mathbb{R}^{P_1 \times \dots \times P_L \times Q_1 \times \dots \times Q_M}$, 收缩张量积的定义如式 (2.4) 所示, 其中 L 是指收缩的阶数。

$$\langle \mathcal{A}, \mathcal{B} \rangle_L [i_1, \dots, i_K, q_1, \dots, q_M] = \sum_{p_1=1}^{P_1} \dots \sum_{p_L=1}^{P_L} \mathcal{A} [i_1, \dots, i_K, p_1, \dots, p_L] \times \mathcal{B} [p_1, \dots, p_L, q_1, \dots, q_M] \quad (2.4)$$

2.3.2 Kronecker 积

Kronecker 积是两个矩阵之间的一种特殊乘积, 若有矩阵 $\mathbf{A} \in \mathbb{R}^{m \times n}$ 和矩阵 $\mathbf{B} \in \mathbb{R}^{p \times q}$, 其 Kronecker 积 $\mathbf{C} \in \mathbb{R}^{mp \times nq}$ 的定义如式 (2.5) 所示。

$$\mathbf{C} = \mathbf{A} \otimes \mathbf{B} = [\mathbf{a}_1 \mathbf{B}, \dots, \mathbf{a}_n \mathbf{B}] = \begin{pmatrix} a_{11} \mathbf{B} & a_{12} \mathbf{B} & \dots & a_{1n} \mathbf{B} \\ a_{21} \mathbf{B} & a_{22} \mathbf{B} & \dots & a_{2n} \mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} \mathbf{B} & a_{m2} \mathbf{B} & \dots & a_{mn} \mathbf{B} \end{pmatrix} \quad (2.5)$$

2.3.3 向量外积

K 个向量的外积如式 (2.6) 所示, 其结果为一个 K 阶秩一张量 $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_K}$ 。

$$\mathcal{A} = \mathbf{a}_1 \circ \dots \circ \mathbf{a}_K \quad (2.6)$$

2.3.4 张量范数

与矩阵 l_2 范数类似, 张量的 F 范数为所有元素的平方和开根号。张量 l_1 范数为所有元素的绝对值之和, 其数学表达如式 (2.7) 所示。

$$\begin{aligned} \|\mathcal{A}\|_F &= \sqrt{\sum_{i_1=1}^{I_1} \dots \sum_{i_K=1}^{I_K} a_{i_1 \dots i_K}^2} \\ \|\mathcal{A}\|_1 &= \sum_{i_1=1}^{I_1} \dots \sum_{i_K=1}^{I_K} |a_{i_1 \dots i_K}| \end{aligned} \quad (2.7)$$

2.4 张量分解

张量分解的思路是寻求更简单或是更有意义的方式来表示张量, 常用的张量分解方法有 CP 分解、Tucker 分解、TT 分解和 TR 分解。由于本文的模型仅使用到 CP 分解和 TT 分解, 故只阐述此两种张量分解方法。

2.4.1 CP 分解

CP 分解也称平行因子分解，其思想是将一个高阶的张量分解成多个秩一张量的和，分解而得的成分矩阵 $\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(K)}$ 具有相同的列数 R 。如式 (2.8) 所示，CP 分解可以被简写为 $[[\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(K)}]]$ ，其中 $\mathbf{U}_r^{(K)}$ 是指 $\mathbf{U}^{(K)}$ 的第 r 列。

$$\begin{aligned} \mathcal{A}[i_1, \dots, i_k, \dots, i_K] &\approx [[\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(k)}, \dots, \mathbf{U}^{(K)}]] \\ &= \sum_{r=1}^R \mathbf{U}_r^{(1)} \circ \dots \circ \mathbf{U}_r^{(k)} \dots \circ \mathbf{U}_r^{(K)} \end{aligned} \quad (2.8)$$

图 2-4 使用图形的形式展示了三阶张量 CP 分解的思想，使用 R 个秩一张量近似地表示原始的复杂张量。

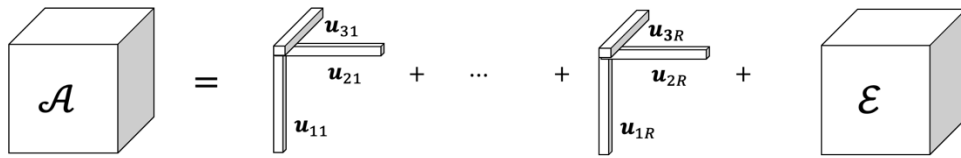


图 2-4 三阶张量的 CP 分解

Figure 2-4 CP decomposition of a third-order tensor

2.4.2 TT 分解

当张量的基本运算涉及的指标较少、缩并关系简单时，常用抽象指标记号去作解析运算，但是，当张量数目多、指标缩并关系复杂时，张量网络可以以图形的形式表示它们的缩并关系，TT (Tensor Train) 分解，也称张量链分解，正是张量网络分解中的一种，其图结构如图 2-5 所示。

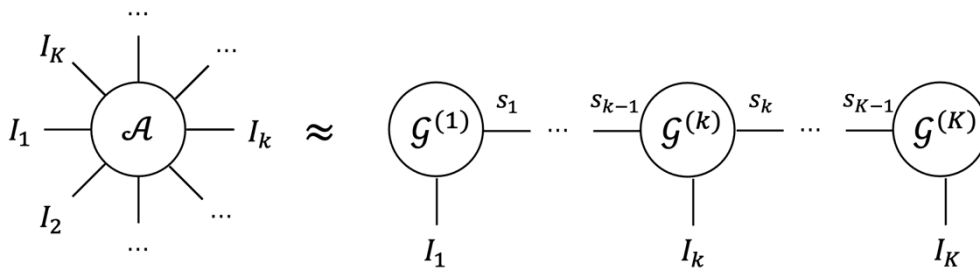


图 2-5 三阶张量的张量链分解

Figure 2-5 Tensor train decomposition of a third-order tensor

对于 K 阶张量 $\mathcal{A} \in \mathbb{R}^{I_1 \times \dots \times I_K}$ ，其 TT 分解的形式的定义如式 (2.9) 所示。其中 $(s_0, s_1, \dots, s_{K-1}, s_K)$ 为长度为 $K + 1$ 的向量，是 TT 分解的秩，称作 TT 秩，其中 $s_0 = s_K = 1$ 。整体记作 $rank_{TT}(\mathcal{A}) = (s_0, s_1, \dots, s_{K-1}, s_K)$ 。将一个复杂的 K 阶张

量 \mathcal{A} 分解为 K 个简单张量的缩并形式，进一步可以写成 K 个矩阵连成的形式，其中 $\mathcal{G}^{(k)}(s_{k-1}, I_k, s_k) = \mathbf{G}^{(k)}[i_k]_{s_{k-1}, s_k}$ 。

$$\begin{aligned} \mathcal{A}[i_1, \dots, i_K] &\approx [[\mathcal{G}^{(1)}, \dots, \mathcal{G}^{(K)}]] \\ &= \sum_{s_0, \dots, s_K} \mathcal{G}^{(1)}(s_0, i_1, s_1) \cdots \mathcal{G}^{(K)}(s_{K-1}, i_K, s_K) \\ &= \mathbf{G}^{(1)}[i_1] \cdots \mathbf{G}^{(K)}[i_K] \end{aligned} \quad (2.9)$$

TT 模型中成分矩阵的求解可以通过 $K - 1$ 次的奇异值分解或者 QR 分解实现。文献^[42]证明了式 (2.10)，描述了如何使用奇异值分解进行 TT 分解。如图 2-6 所示，把 K 阶张量 \mathcal{A} 进行模-1 展开，得到 $\mathbf{A}_{[1]}$ ，对其进行奇异值分解，得到 $\mathbf{U}^{(1)}$ 和 $\mathbf{B}^{(1)}$ ，将 $\mathbf{B}^{(1)}$ 再次进行同样的分解，以此类推，进行 $K - 1$ 次分解，最终可以得到 K 个 TT 成分矩阵。

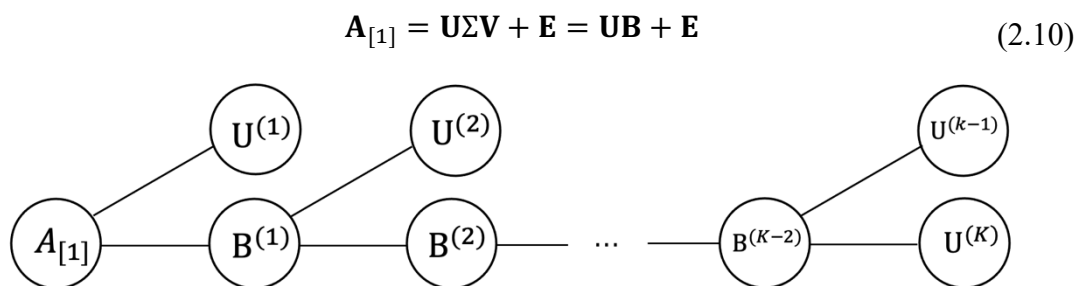


图 2-6 使用 SVD 进行张量链分解
Figure 2-6 Tensor train decomposition using SVD

一个完整的张量链分解可以划分为几个子链，以第 k 个 TT 成分张量为界限，可以将其分成三部分，分别为 $\mathcal{G}^{(<k)}$ 、 $\mathcal{G}^{(k)}$ 和 $\mathcal{G}^{(>k)}$ ，其中 $\mathcal{G}^{(<k)}$ 是指前 $k - 1$ 个 TT 分解成分张量的缩并，而 $\mathcal{G}^{(>k)}$ 是指后 $K - k$ 个 TT 分解成分张量的缩并，其公式定义如式 (2.11) 所示。

$$\begin{aligned} \mathcal{G}^{(<k)} &= [[\mathcal{G}^{(1)}, \dots, \mathcal{G}^{(k-1)}]] \in \mathbb{R}^{I_1 \times \dots \times I_{k-1} \times s_{k-1}} \\ \mathcal{G}^{(>k)} &= [[\mathcal{G}^{(k+1)}, \dots, \mathcal{G}^{(K)}]] \in \mathbb{R}^{s_k \times I_{k+1} \times \dots \times I_K} \end{aligned} \quad (2.11)$$

从矩阵层面用三个子链表达张量 \mathcal{A} 的模- k 展开 $\mathbf{A}_{[k]} \in \mathbb{R}^{I_k \times I_{-k}}$ ，可以得到式 (2.12)。

$$\mathbf{A}_{[k]} = \mathbf{G}_{[2]}^{(k)} \left(\mathbf{G}_{[1]}^{(>k)} \otimes \mathbf{G}_{[k]}^{(<k)} \right) \quad (2.12)$$

2.5 张量岭回归

2.5.1 基于 CP 分解的张量岭回归

基于 CP 分解的张量岭回归由 Lock 在 2018 年提出^[31]，其假设回归系数 \mathcal{B} 具有低 CP 秩结构，利用 CP 模型来低秩近似系数张量 \mathcal{B} 。当其秩为 R 时，回归系数 \mathcal{B} 可以被表示为 $L + M$ 个秩一张量之和，如下式所示。此时，待估参数数量从 $\prod_{i=1}^L P_i \prod_{j=1}^M Q_j$ 降至为 $(\sum_{i=1}^L P_i + \sum_{j=1}^M Q_j) \times R$ ，有效地简化了数据结构，降低了估计成本。对于回归系数的求解，可以使用交替求解的思想，轮流对 $L + M$ 个矩阵进行求解，前 L 维的求解以 $\mathbf{U}^{(l)}$ 为例，后 M 维的求解以 $\mathbf{V}^{(m)}$ 例。

$\mathbf{U}^{(l)}$ 的求解

定义辅助张量 \mathcal{C}_r ，其中 $r \in [1: R]$ ，它是预测变量 \mathcal{X} 与回归系数 \mathcal{B} 除 $\mathbf{U}^{(l)}$ 外的所有剩余成分的收缩张量积，收缩阶数为 $L - 1$ 阶。由于少收缩了 $\mathbf{U}^{(l)}$ 这一维，辅助张量的维度为 $\mathcal{C}_r \in \mathbb{R}^{N \times P_l \times Q_1 \times \dots \times Q_M}$ 。

$$\mathcal{C}_r = \langle \mathcal{X}, \mathbf{U}_r^{(1)} \circ \dots \circ \mathbf{U}_r^{(l-1)} \circ \mathbf{U}_r^{(l+1)} \circ \dots \circ \mathbf{U}_r^{(L)} \circ \mathbf{V}_r^{(1)} \circ \dots \circ \mathbf{V}_r^{(M)} \rangle_{L-1}$$

将辅助张量 \mathcal{C}_r 沿着 P_l 这一维度进行展开，得到矩阵 \mathbf{C}_r ，其维度为 $\mathbf{C}_r \in \mathbb{R}^{NQ \times P_l}$ ，其中 $Q = Q_1 \times \dots \times Q_M$ 。将矩阵 \mathbf{C}_r 按照 r 从 1 到 R 的顺序横向排列，最终可以得到辅助矩阵 $\mathbf{C} = [\mathbf{C}_1 \dots \mathbf{C}_R]$ ，其维度为 $\mathbf{C} \in \mathbb{R}^{NQ \times P_l R}$ 。借助辅助矩阵 \mathbf{C} ，对张量岭回归的优化公式 (3.4) 进行变形，得到新式 (2.13)。

$$\begin{cases} \operatorname{argmin} & \| \operatorname{vec}(\mathcal{Y}) - \mathbf{C} \operatorname{vec}(\mathbf{U}^{(l)}) \|_F^2 + \lambda \| \mathbf{U}^{(1)T} \mathbf{U}^{(1)} \dots \mathbf{V}^{(M)T} \mathbf{V}^{(M)} \|_F^2 \\ \text{s. t.} & \operatorname{rank}(\mathcal{B}) \leq R \end{cases} \quad (2.13)$$

结合张量代数运算法则，可以通过简单运算得到 $\mathbf{U}^{(l)}$ 的解，当 $\lambda \neq 0$ 时，其解如式 (2.14) 所示，其中 \otimes 是 Kronecker 积， \mathbf{I}_{P_l} 是一个单位矩阵，其维度为 $\mathbf{I}_{P_l} \in \mathbb{R}^{P_l \times P_l}$ ，当 $\lambda = 0$ 时，式 (2.13) 退化成为无约束的张量回归。

$$\operatorname{vec}(\mathbf{U}^{(l)}) = (\mathbf{C}^T \mathbf{C} + \lambda \begin{matrix} (\mathbf{U}^{(1)T} \mathbf{U}^{(1)} \dots \mathbf{U}^{(l-1)T} \mathbf{U}^{(l-1)} \\ \cdot \mathbf{U}^{(l+1)T} \mathbf{U}^{(l+1)} \dots \mathbf{V}^{(M)T} \mathbf{V}^{(M)} \\ \otimes \mathbf{I}_{P_l} \end{matrix})^{-1} \mathbf{C}^T \operatorname{vec}(\mathcal{Y}) \quad (2.14)$$

$\mathbf{V}^{(m)}$ 的求解

成分矩阵 $\mathbf{V}^{(m)}$ 的求解与 $\mathbf{U}^{(l)}$ 类似，回归系数 \mathcal{B} 的后 M 模具有位置交换不变性，所以，本节将以 $\mathbf{V}^{(m)}$ 为例说明后 M 模的求解，其余的 $M - 1$ 模是类似的。定义向量 \mathbf{d}_r ，其为预测张量 \mathcal{X} 与除 $\mathbf{V}^{(m)}$ 以外的所有成分矩阵的 L 阶收缩张量积的向量化展开。

$$\mathbf{d}_r = \text{vec}(\langle \mathcal{X}, \mathbf{U}_r^{(1)} \circ \dots \circ \mathbf{U}_r^{(L)} \circ \mathbf{V}_r^{(1)} \circ \dots \circ \mathbf{V}_r^{(m-1)} \circ \mathbf{V}_r^{(m+1)} \circ \dots \circ \mathbf{V}_r^{(M)} \rangle_L)$$

将向量 \mathbf{d}_r 按照 r 从1到 R 的顺序横向排列, 得到辅助矩阵 $\mathbf{D} = [\mathbf{d}_1 \cdots \mathbf{d}_R]$, 其维度为 $\mathbf{D} \in \mathbb{R}^{NQ-m \times R}$, 其中 $Q_{-m} = Q_1 \times \cdots \times Q_{m-1} \times Q_{m+1} \times \cdots \times Q_M$ 。借助辅助矩阵 \mathbf{D} 将张量岭回归的优化公式重写为式 (2.15)。

$$\begin{cases} \text{argmin} & \| \mathbf{Y}_{[m]}^T - \mathbf{D}\mathbf{V}^{(m)T} \|_F^2 + \lambda \| \mathbf{U}^{(1)T} \mathbf{U}^{(1)} \dots \mathbf{V}^{(M)T} \mathbf{V}^{(M)} \|_F^2 \\ \text{s. t.} & \text{rank}(\mathbf{B}) \leq R \end{cases} \quad (2.15)$$

在公式 (2.15) 中, $\mathbf{Y}_{[m]}$ 为张量响应变量 \mathcal{Y} 的模 $-m$ 展开, 其维度为 $\mathbf{Y}_{[m]} \in \mathbb{R}^{Q_m \times NQ-m}$ 。可以通过简单代数运算得到 $\mathbf{V}^{(m)}$ 的解, 当 $\lambda \neq 0$ 时, 带 F 范数约束的解如式 (2.16) 所示, 当 $\lambda = 0$, 退化为无约束的最小二乘解。

$$\mathbf{V}^{(m)T} = (\mathbf{D}^T \mathbf{D} + \lambda (\mathbf{U}^{(1)T} \mathbf{U}^{(1)} \dots \mathbf{V}^{(m-1)T} \mathbf{V}^{(m-1)} \cdot \mathbf{V}^{(m+1)T} \mathbf{V}^{(m+1)} \dots \mathbf{V}^{(M)T} \mathbf{V}^{(M)}))^{-1} \mathbf{D}^T \mathbf{Y}_{[m]}^T \quad (2.16)$$

伪代码算法 1 总结了基于 CP 分解的张量岭回归算法的求解过程。

算法 1 基于 CP 分解的张量岭回归

输入: 自变量 \mathcal{X} , 响应变量 \mathcal{Y} , 惩罚参数 λ , 收敛阈值 convT , 秩 R ;

输出: 张量回归系数 \mathcal{B} ;

- 1: While 两次迭代之间目标函数的绝对差值 > 收敛阈值 Do:
 - 2: For $l = 1, \dots, L$:
 - 3: 使用式 (2.14) 更新 $\mathbf{U}^{(l)}$;
 - 4: End For
 - 5: For $m = 1, \dots, M$:
 - 6: 使用式 (2.16) 更新 $\mathbf{V}^{(m)}$;
 - 7: End For
 - 8: End While
 - 9: 使用更新后的 $\mathbf{U}^{(l)}$ 和 $\mathbf{V}^{(m)}$ 来计算回归系数 \mathcal{B} ;
 - 10: Return 张量回归系数 \mathcal{B} 。
-

2.5.2 基于 TT 分解的张量岭回归

基于 TT 分解的张量岭回归由 Liu 在 2020 年提出^[18], 假设回归系数 \mathbf{B} 具有低 TT 秩结构, 在假设 TT 秩为 $(R_0, R_1, \dots, R_{L+M})$ 时, TT 模型将分解出 $L + M$ 个成分张量。与前述的基于 CP 分解的张量岭回归类似, 对某一维成分进行求解, 交替进行直至对所有的成分张量完成求解。

$$\tilde{\mathbf{Y}}_{[1]} = \begin{bmatrix} \mathbf{Y}_{[1]} \\ \mathbf{0} \end{bmatrix} \quad \tilde{\mathbf{X}}_{[1]} = \begin{bmatrix} \mathbf{X}_{[1]} \\ \sqrt{\lambda} \mathbf{I} \end{bmatrix}$$

为了后续阐述的简便起见, 定义 $\tilde{\mathbf{y}} \in \mathbb{R}^{(N+P) \times Q_1 \times \dots \times Q_M}$ 和 $\tilde{\mathbf{x}} \in \mathbb{R}^{(N+P) \times P_1 \times \dots \times P_L}$, 其模-1 展开 $\tilde{\mathbf{Y}}_{[1]}$ 和 $\tilde{\mathbf{X}}_{[1]}$ 满足上式, 其中 $\mathbf{0}$ 是维度为 $\mathbf{0} \in \mathbb{R}^{P \times Q}$ 的零矩阵, \mathbf{I} 是维度为 $\mathbf{I} \in \mathbb{R}^{P \times P}$ 的单位矩阵, 其中 $P = \prod_{i=1}^L P_i$, $Q = \prod_{j=1}^M Q_j$ 。除此之外, 简写张量 \mathbf{u} 为前 L 维成分张量的缩并, 张量 \mathbf{v} 为后 M 维成分张量的缩并, 如下式所示。

$$\begin{aligned} \mathbf{u} &= [[\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(L)}]] \in \mathbb{R}^{P_1 \times \dots \times P_L \times R_{L-1}} \\ \mathbf{v} &= [[\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(L)}]] \in \mathbb{R}^{R_L \times Q_1 \times \dots \times Q_M} \end{aligned}$$

$\mathbf{u}^{(L)}$ 的求解

首先, 当 $l = L$ 时, 分别对 $\mathbf{u}^{(L)}$ 的第三维 $r = 1, \dots, R_L$ 进行求解, 记此时的变量为 $\mathbf{u}_{::r}^{(L)}$, 针对变量 $\mathbf{u}_{::r}^{(L)}$ 的优化公式简记为 $g(\mathbf{u}_{::r}^{(L)})$, 定义矩阵 $\mathbf{z} = \tilde{\mathbf{Y}}_{[1]} \mathbf{V}^T (\mathbf{V} \mathbf{V}^T)^{-1} \in \mathbb{R}^{(N+P) \times R_L}$, 并将 $\tilde{\mathbf{x}}_{n, \dots}$ 简写记为 $\tilde{\mathbf{x}}_n$, 可以得到式 (2.17)。

$$\begin{aligned} g(\mathbf{u}_{::r}^{(L)}) &= \sum_{n=1}^{N+P} \left(\mathbf{z}_{nr} - \langle \tilde{\mathbf{x}}_n, [[\mathbf{u}^{(1)}, \dots, \mathbf{u}_{::r}^{(L)}]] \rangle_L \right)^2 \\ &= \sum_{n=1}^{N+P} \left(\mathbf{z}_{nr} - \langle (\tilde{\mathbf{x}}_n)_{[L]} (\mathbf{U}_{[L]}^{(<L)})^T, (\mathbf{u}_{::r}^{(L)})^T \rangle_2 \right)^2 \end{aligned} \quad (2.17)$$

式 (2.17) 中, $\mathbf{u}^{(<L)} = [[\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(L-1)}]] \in \mathbb{R}^{P_1 \times \dots \times P_{L-1} \times R_{L-1}}$ 为前 $L - 1$ 个成分张量的缩并, 沿其模- L 展开, 得到矩阵 $\mathbf{U}_{[L]}^{(<L)} \in \mathbb{R}^{R_{L-1} \times P_{-L}}$ 。根据式 (2.17) 分别对 R_L 个 $\mathbf{u}_{::r}^{(L)}$ 使用最小二乘法进行求解即可。

$$\begin{aligned} g(\mathbf{u}^{(l)}) &= \sum_{n=1}^{N+P} \sum_{r=1}^{R_L} \left(\mathbf{z}_{nr} - \langle (\tilde{\mathbf{x}}_n)_{[l]}, \mathbf{U}_{[2]}^{(l)} \left((\mathbf{U}_r)_{[1]}^{(>l)} \otimes \mathbf{U}_{[l]}^{(<l)} \right) \right)_2 \right)^2 \\ &= \sum_{n=1}^{N+P} \sum_{r=1}^{R_L} \left(\mathbf{z}_{nr} - \langle (\tilde{\mathbf{x}}_n)_{[l]} \left((\mathbf{U}_r)_{[1]}^{(>l)} \otimes \mathbf{U}_{[l]}^{(<l)} \right)^T, \mathbf{U}_{[2]}^{(l)} \right)_2 \right)^2 \end{aligned} \quad (2.18)$$

对于 $l \neq L$ 时, 经计算可得式 (2.18), 式中, $\mathbf{U}_{[l]}^{(<l)}$ 为 $\mathbf{u}^{(<l)}$ 的模- l 展开, 其中 $\mathbf{u}^{(<l)} = [[\mathbf{u}^{(1)}, \dots, \mathbf{u}^{(l-1)}]] \in \mathbb{R}^{P_1 \times \dots \times P_{l-1} \times R_{l-1}}$ 为前 $l - 1$ 个成分张量的缩并。除此之外, $(\mathbf{U}_r)_{[1]}^{(>l)}$ 为 $\mathbf{u}_r^{(>l)}$ 的模-1 展开, 其中 $\mathbf{u}_r^{(>l)} = [[\mathbf{u}^{(l+1)}, \dots, \mathbf{u}_{::r}^{(L)}]] \in \mathbb{R}^{R_l \times P_{l+1} \times \dots \times P_L}$ 。结合使用式 (2.17) 和 (2.18) 即可完成对回归系数 \mathbf{B} 的前 L 个成分张量完成求解。

$\mathcal{V}^{(m)}$ 的求解

定义辅助张量 $\mathcal{W} \in \mathbf{R}^{R_L \times Q_1 \times \dots \times Q_M}$, $\mathbf{W}_{[1]} = ((\tilde{\mathbf{X}}_{[1]} \mathbf{U})^T \tilde{\mathbf{X}}_{[1]} \mathbf{U})^{-1} (\tilde{\mathbf{X}}_{[1]} \mathbf{U})^T \tilde{\mathbf{Y}}_{[1]}$ 为其模-1展开, 并将 $\mathcal{W}_{r, \dots}$ 简写为 \mathcal{W}_r , 其中 $r = 1, \dots, R_L$ 。

对张量 $\mathcal{V}^{(1)}$ 第一维 $r = 1, \dots, R_L$ 分别进行求解, 将针对变量 $\mathcal{V}_{r, \dots}^{(1)}$ 的优化公式简记为 $g(\mathcal{V}_{r, \dots}^{(1)})$ 。计算得到式 (2.19), 使用最小二乘法进行求解, 其中, $\mathbf{V}_{[1]}^{(>1)}$ 是 $\mathcal{V}^{(>1)} = [[\mathcal{V}^{(2)}, \dots, \mathcal{V}^{(M)}]] \in \mathbf{R}^{R_{L+1} \times Q_2 \times \dots \times Q_M}$ 的模-1展开。

$$\begin{aligned} g(\mathcal{V}_{r, \dots}^{(1)}) &= \left\| \mathcal{W}_r - [[\mathcal{V}_{r, \dots}^{(1)}, \mathcal{V}^{(2)}, \dots, \mathcal{V}^{(M)}]] \right\|_F^2 \\ &= \left\| (\mathbf{W}_r)_{[1]} - \mathcal{V}_{r, \dots}^{(1)} \mathbf{V}_{[1]}^{(>1)} \right\|_F^2 \end{aligned} \quad (2.19)$$

当 $m \neq 1$ 时, $\mathcal{V}^{(m)}$ 的求解如式 (2.20) 所示, 其中 $\mathbf{V}_{[1]}^{(>m)}$ 是 $\mathcal{V}^{(>m)} = [[\mathcal{V}^{(m+1)}, \dots, \mathcal{V}^{(M)}]] \in \mathbf{R}^{R_{L+m} \times Q_{m+1} \times \dots \times Q_M}$ 的模-1展开, $(\mathbf{V}_r)_{[m]}^{(<m)}$ 是 $\mathcal{V}_r^{(<m)} = [[\mathcal{V}_{r, \dots}^{(1)}, \mathcal{V}^{(2)}, \dots, \mathcal{V}_{m-1}]] \in \mathbf{R}^{Q_1 \times \dots \times Q_m \times R_{L+m-1}}$ 的模- m 展开。使用最小二乘法对 $\mathcal{V}^{(m)}$ 逐个进行求解即可。

$$\begin{aligned} g(\mathcal{V}_{r, \dots}^{(m)}) &= \sum_{r=1}^{R_L} \left\| \mathcal{W}_r - [[\mathcal{V}_{r, \dots}^{(1)}, \mathcal{V}^{(2)}, \dots, \mathcal{V}^{(M)}]] \right\|_F^2 \\ &= \sum_{r=1}^{R_L} \left\| (\mathbf{W}_r)_{[m]} - \mathbf{V}_{[2]}^{(m)} (\mathbf{V}_{[1]}^{(>m)} \otimes (\mathbf{V}_r)_{[m]}^{(<m)}) \right\|_F^2 \end{aligned} \quad (2.20)$$

伪代码算法 2 总结了基于 TT 分解的张量岭回归的系数 \mathcal{B} 的求解思路。

算法 2 基于 TT 分解的张量岭回归

输入: 自变量 \mathcal{X} , 响应变量 \mathcal{Y} , 惩罚参数 λ , 收敛阈值 $convT$, 秩 (R_0, \dots, R_{L+M}) ;
输出: 张量回归系数 \mathcal{B} ;

1: While 两次迭代之间目标函数的绝对差值 > 收敛阈值 Do:

2: For $l = 1, \dots, L$:

3: 使用式 (2.17)、(2.18) 更新 $\mathcal{U}^{(l)}$;

4: End For

5: For $m = 1, \dots, M$:

6: 使用式(2.19)、(2.20)更新 $\mathcal{V}^{(m)}$;

7: End For

8: End While

9: 使用更新后的 $\mathcal{U}^{(l)}$ 和 $\mathcal{V}^{(m)}$ 来计算回归系数 \mathcal{B} ;

10: Return 张量回归系数 \mathcal{B} 。

2.6 基于拉普拉斯的贝叶斯张量回归

基于高斯先验的贝叶斯张量回归算法由 Lock 于 2018 年提出^[31], 算法的关键在于假设回归系数 \mathcal{B} 的先验分布服从球形高斯分布^[43], 其为多维高斯分布的一种特例, 即回归系数 \mathcal{B} 的每一个元素都独立同分布于一维高斯分布。若变量 x 服从均值为 μ , 方差为 σ^2 的正态分布, 记为 $x \sim N(\mu, \sigma^2)$, 其密度函数如式 (2.21) 所示。

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (2.21)$$

假设回归系数 \mathcal{B} 的每一个元素都独立同分布于均值为 0, 方差为 σ^2/λ 的正态分布, 经过计算可以得到 $P(\mathcal{B}) \propto \exp(-(\lambda/2\sigma^2) \|\mathcal{B}\|_F^2)$ 。可以发现, 当惩罚参数 λ 越大时, 高斯分布的方差就越小, 数据就越集中在 0 附近, 表明约束越严格。

$$\begin{aligned} P(\mathcal{B}) &= \prod_{i=1}^{PQ} P(b_i) = \prod_{i=1}^{PQ} \frac{\sqrt{\lambda}}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\lambda(b_i)^2}{2\sigma^2}\right) \\ &\propto \prod_{i=1}^{PQ} \exp\left(-\frac{\lambda(b_i)^2}{2\sigma^2}\right) = \exp\left(-\frac{\lambda}{2\sigma^2} \|\mathcal{B}\|_F^2\right) \end{aligned}$$

假设残差项 ε 的每一个元素都服从均值为 0, 方差为 σ^2 的高斯分布, 那么似然函数可以写成式 (2.22) 的形式。

$$P(\mathcal{Y}|\sigma^2, \mathcal{B}, \mathcal{X}) \propto \exp\left(-\frac{1}{2\sigma^2} \|\mathcal{Y} - \langle \mathcal{X}, \mathcal{B} \rangle_L\|_F^2\right) \quad (2.22)$$

结合先验分布和似然函数, 可以得到回归系数 \mathcal{B} 的后验分布形式, 如式 (2.23) 所示, 可以发现其形式与张量正则岭回归的优化公式是一致的。由于先验分布和似然函数都是服从正态分布的, 根据共轭分布的规则, 后验分布也是服从正态分布的, 只不过后验正态分布的参数发生了变化。所以, 可以通过确定后验高斯分布的参数, 直接从后验分布中进行采样。

$$P(\mathcal{B}|\mathcal{Y}, \mathcal{X}, \mathcal{B}, \sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2} (\|\mathcal{Y} - \langle \mathcal{X}, \mathcal{B} \rangle_L\|_F^2 + \lambda \|\mathcal{B}\|_F^2)\right) \quad (2.23)$$

结合 2.5.1 中基于 CP 分解的张量岭回归算法, 可以得出 $\mathbf{U}^{(l)}$ 的后验分布为高斯分布, 如式 (2.24) 所示。

$$P(\text{vec}(\mathbf{U}^{(l)})|\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(l-1)}, \mathbf{U}^{(l+1)}, \dots, \mathbf{U}^{(L)}, \mathbf{V}^{(1)}, \dots, \mathbf{V}^{(M)}, \mathcal{X}, \mathcal{Y}, \sigma^2) = N(\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l) \quad (2.24)$$

$$\begin{cases} \boldsymbol{\mu}_l &= (\mathbf{C}^T \mathbf{C} + \lambda(\mathbf{U}^{(1)T} \mathbf{U}^{(1)} \dots \mathbf{U}^{(l-1)T} \mathbf{U}^{(l-1)} \\ &\quad \cdot \mathbf{U}^{(l+1)T} \mathbf{U}^{(l+1)} \dots \mathbf{V}^{(M)T} \mathbf{V}^{(M)}) \otimes \mathbf{I}_{P_l})^{-1} \mathbf{C}^T \text{vec}(\mathcal{Y}) \\ \boldsymbol{\Sigma}_l &= \sigma^2 (\mathbf{C}^T \mathbf{C} + \lambda(\mathbf{U}^{(1)T} \mathbf{U}^{(1)} \dots \mathbf{U}^{(l-1)T} \mathbf{U}^{(l-1)} \\ &\quad \cdot \mathbf{U}^{(l+1)T} \mathbf{U}^{(l+1)} \dots \mathbf{V}^{(M)T} \mathbf{V}^{(M)}) \otimes \mathbf{I}_{P_l})^{-1} \end{cases}$$

$\mathbf{V}^{(m)}$ 的后验分布同样也是高斯分布，如式 (2.25) 所示。

$$P(\text{vec}(\mathbf{V}^{(m)})|\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(L)}, \mathbf{V}^{(m-1)}, \mathbf{V}^{(m+1)}, \dots, \mathbf{V}^{(M)}, \mathcal{X}, \mathcal{Y}, \sigma^2) = N(\boldsymbol{\mu}_{L+m}, \boldsymbol{\Sigma}_{L+m}) \quad (2.25)$$

$$\begin{cases} \boldsymbol{\mu}_{L+m} &= \mathbf{V}^{(m)T} = (\mathbf{D}^T \mathbf{D} + \lambda(\mathbf{U}^{(1)T} \mathbf{U}^{(1)} \dots \mathbf{V}^{(m-1)T} \mathbf{V}^{(m-1)} \\ &\quad \cdot \mathbf{V}^{(m+1)T} \mathbf{V}^{(m+1)} \dots \mathbf{V}^{(M)T} \mathbf{V}^{(M)}))^{-1} \mathbf{D}^T \mathbf{Y}_{[m]}^T \\ \boldsymbol{\Sigma}_{L+m} &= \sigma^2 \mathbf{V}^{(m)T} = (\mathbf{D}^T \mathbf{D} + \lambda(\mathbf{U}^{(1)T} \mathbf{U}^{(1)} \dots \mathbf{V}^{(m-1)T} \mathbf{V}^{(m-1)} \\ &\quad \cdot \mathbf{V}^{(m+1)T} \mathbf{V}^{(m+1)} \dots \mathbf{V}^{(M)T} \mathbf{V}^{(M)}))^{-1} \otimes \mathbf{I}_{Q_m} \end{cases}$$

根据式 (2.24) 和 (2.25)，可以对系数张量 \mathcal{B} 的 $L + M$ 个成分矩阵逐个使用 Gibbs 方法进行采样。除此之外，由于对 σ^2 的大或小没有明显的偏好，所以，可以为其假设一个无信息先验，也称 Jeffrey 先验， $P(\sigma^2) = \frac{1}{\sigma^2}$ 。

结合似然函数 (2.22)， σ^2 的后验分布为逆高斯分布，如式 (2.26) 所示。

$$P(\sigma^2|\mathcal{X}, \mathcal{Y}, \mathcal{B}) \propto P(\sigma^2)P(\mathcal{Y}|\sigma^2, \mathcal{B}, \mathcal{X}) = IG\left(\frac{NQ}{2}, \frac{\|\mathcal{Y} - \langle \mathcal{X}, \mathcal{B} \rangle_L\|_F^2}{2}\right) \quad (2.26)$$

基于高斯分布的贝叶斯张量回归模型中，Lock 通过对回归系数 \mathcal{B} 假设一个围绕 0 随机波动高斯先验分布来起到稀疏正则的作用，将算法的简要过程总结在算法 4 中。

算法 3 基于高斯先验的贝叶斯张量回归

输入：预测变量 \mathcal{X} ，响应变量 \mathcal{Y} ，测试集 \mathcal{X}_{new} ，采样数 T ，初始 \mathcal{B}_0 ，惩罚参数 λ ，秩 R ；

输出：张量回归系数 \mathcal{B} ，预测值 MCMC 样本集 $\mathcal{Y}_{samples}$ ；

- 1: For $t = 1, \dots, T$:
 - 2: 从式 (2.26) 中采样，得到 $\sigma^{2(t)}$ ；
 - 3: For $l = 1, \dots, L$:
 - 4: 从式 (2.24) 中采样，得到 $\mathbf{U}^{(l)(t)}$ ；
 - 5: End For
 - 6: For $m = 1, \dots, M$:
 - 7: 使用式(2.25)更新 $\mathbf{V}^{(m)(t)}$ ；
 - 8: End For
 - 9: 根据采样所得的 $\mathbf{U}^{(l)(t)}$ 和 $\mathbf{V}^{(m)(t)}$ 来计算 $\mathcal{B}^{(t)}$ ；
 - 10: 根据新输入的样本 \mathcal{X}_{new} 来对每个变量进行预测，得到 $\mathcal{Y}_{samples}^{(t)}$ ；
 - 11: End For
 - 12: Return 张量回归系数 \mathcal{B} ，预测值 MCMC 样本集 $\mathcal{Y}_{samples}$ 。
-

第3章 稀疏张量回归算法

对于张量数据而言，早期的大部分算法都是将数据展开成矩阵或是向量的形式加以拟合，这在一定程度上浪费了张量数据本身的结构信息。近年来，张量数据的广泛存在以及张量分解技术的日益成熟，将经典的回归算法扩展到高阶张量领域可以帮助探究和利用数据本身的信息。

本文中，预测变量和响应变量均为任意阶张量。预测变量 $\mathcal{X} \in \mathbb{R}^{N \times P_1 \times \cdots \times P_L}$ 为 $L + 1$ 阶张量，第一维 N 是指样本个数，其余 L 维是指每个样本特征的维度，响应变量 $\mathcal{Y} \in \mathbb{R}^{N \times Q_1 \times \cdots \times Q_M}$ 是 $M + 1$ 阶张量，其第一维 N 是指样本个数，其余 M 维是指每个样本下响应变量的维度，张量回归模型可以写成式 (3.1) 的形式。其中 \mathcal{E} 是服从高斯分布的随机干扰项，其维度同响应变量 \mathcal{Y} ，回归系数 \mathcal{B} 是 $L + M$ 阶张量 $\mathcal{B} \in \mathbb{R}^{P_1 \times \cdots \times P_L \times Q_1 \times \cdots \times Q_M}$ 。 $\langle \mathcal{X}, \mathcal{B} \rangle_L$ 是自变量 \mathcal{X} 和回归参数 \mathcal{B} 的 L 阶收缩张量积， \mathcal{B} 的前 L 维和自变量 \mathcal{X} 的 L 维特征进行压缩，后 M 维是沿着响应变量 \mathcal{Y} 的 M 维特征的扩展，当 L 和 M 都取 1 时，式 (3.1) 退化成普通多元线性回归 $\mathbf{Y} = \langle \mathbf{X}, \mathbf{B} \rangle_1 + \mathbf{E}$ ，其中 $\langle \mathbf{X}, \mathbf{B} \rangle_1$ 为普通矩阵相乘。

$$\mathcal{Y} = \langle \mathcal{X}, \mathcal{B} \rangle_L + \mathcal{E} \quad (3.1)$$

对于回归系数 \mathcal{B} 的估计，可以通过最小化残差平方和来实现，但在样本数 N 太少或是自变量特征之间具有较大的相关性时，该优化公式是病态的，会导致过拟合，以至于产生的解不稳定。故针对该问题，可以使用正则化技术来避免过拟合，增加系统的稳定性。常见的正则化技术有施加 l_2 惩罚项或 l_1 惩罚项，在张量形式中， l_2 范数也称 F 范数，带有 l_2 惩罚项的回归称为岭回归，带有 l_1 惩罚项的回归称为 Lasso 回归，张量岭回归的优化求解式如式 (3.2) 所示，张量 Lasso 回归如式 (3.3) 所示。

$$\hat{\mathcal{B}} = \operatorname{argmin} \|\mathcal{Y} - \langle \mathcal{X}, \hat{\mathcal{B}} \rangle_L\|_F^2 + \lambda \|\hat{\mathcal{B}}\|_F^2 \quad (3.2)$$

$$\hat{\mathcal{B}} = \operatorname{argmin} \|\mathcal{Y} - \langle \mathcal{X}, \hat{\mathcal{B}} \rangle_L\|_F^2 + \lambda \|\hat{\mathcal{B}}\|_1 \quad (3.3)$$

对于上式，回归系数 \mathcal{B} 共有 $\prod_{i=1}^L P_i \prod_{j=1}^M Q_j$ 个参数值需要估计，而在数据本身为高维张量的情况下，其稀疏特性表明了回归系数 \mathcal{B} 具有低秩结构，利用张量分解技术可以有效降低待估参数的数量。张量分解技术已经较为成熟，经典的方法有 CP 分解、Tucker 分解，近年来引起关注的张量网络分解方法有 TT 分解和 TR 分解等，对系数张量 \mathcal{B} 假设不同的低秩结构，则需要应用不同的算法进行求解。本章将提出基于 CP 分解的张量 Lasso 回归算法，并将其与 2.5 节中的两个高效回归模型进行对比分析。

3.1 基于 CP 分解的张量 Lasso 回归

3.1.1 模型

在张量回归模型中利用 CP 模型来低秩近似系数张量 \mathcal{B} 是一种常用且重要的方法, 当其 CP 秩为 R 时, 回归系数 \mathcal{B} 可以被表示为 $L + M$ 个秩一张量之和, 如下式所示。此时, 待估参数数量从 $\prod_{i=1}^L P_i \prod_{j=1}^M Q_j$ 降至为 $(\sum_{i=1}^L P_i + \sum_{j=1}^M Q_j) \times R$, 有效地简化了数据结构, 降低了估计成本。

$$\mathcal{B} \approx \left[\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(L)}, \mathbf{V}^{(1)}, \dots, \mathbf{V}^{(M)} \right] = \sum_{r=1}^R \mathbf{U}_r^{(1)} \circ \dots \circ \mathbf{U}_r^{(L)} \circ \mathbf{V}_r^{(1)} \circ \dots \circ \mathbf{V}_r^{(M)}$$

在 2.5 节中所述的张量岭回归是施加 l_2 惩罚项的回归模型, 在张量中, l_2 范数也称 F 范数, Lasso 回归则是施加 l_1 惩罚项的回归模型, 相比于岭回归, Lasso 回归的解更加稀疏和稳健。本文提出了基于 CP 分解的张量 Lasso 回归模型, 假设张量回归系数具有低 CP 秩结构。结合 CP 分解的定义, 可以得到新的优化目标公式, 如式 (3.4) 所示。

$$\begin{cases} \operatorname{argmin} & \|\mathcal{Y} - \langle \mathcal{X}, \mathcal{B} \rangle_L\|_F^2 + \lambda \|\mathcal{B}\|_1 \\ \text{s. t.} & \operatorname{rank}(\mathcal{B}) \leq R \end{cases} \quad (3.4)$$

在岭回归模型中, F 范数是平滑的凸函数, 可以通过求导进行优化求解, 但在 Lasso 回归中, l_1 范数虽然是凸函数, 但是却不可导, 故在对回归系数 \mathcal{B} 进行求解时不能通过直接运算得到显式解, 本文通过对优化公式进行变形重构, 结合使用广义交替方向乘子法使用迭代更新的思想加以求解。

3.1.2 模型求解

对于式 (3.4) 的求解, 可以使用交替求解的思想, 轮流对 $L + M$ 个矩阵进行求解。如求解 $\mathbf{U}^{(1)}$ 时, 固定其他 $L + M - 1$ 个成分矩阵 $\mathbf{U}^{(2)}, \dots, \mathbf{U}^{(L)}, \mathbf{V}^{(1)}, \dots, \mathbf{V}^{(M)}$ 不变。具体的求解方法根据回归系数 \mathcal{B} 本身的特征分成了两部分, 第一部分是 \mathcal{B} 的前 L 维, 即 $\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(L)}$, 它们是与预测变量 \mathcal{X} 的 L 维特征进行压缩的部分, 第二部分是 \mathcal{B} 的后 M 维, 即 $\mathbf{V}^{(1)}, \dots, \mathbf{V}^{(M)}$, 它们是响应变量 \mathcal{Y} 的 M 维特征的扩展。本节将以 $\mathbf{U}^{(l)}$ 和 $\mathbf{V}^{(m)}$ 为例分别阐述两部分的求解方法。

$\mathbf{U}^{(l)}$ 的求解

根据张量 CP 分解的定义, 分解而得的成分矩阵之间的先后顺序是可以交换的, 也就是 $\sum_{r=1}^R \mathbf{U}_r^{(1)} \circ \mathbf{U}_r^{(2)} = \sum_{r=1}^R \mathbf{U}_r^{(2)} \circ \mathbf{U}_r^{(1)}$ 。所以, 可以以 $\mathbf{U}^{(l)}$ 为例阐述回归系数 \mathcal{B} 的前 L 模的解法, 其余的 $L - 1$ 模是与之类似的。

定义辅助张量 \mathcal{C}_r ，其中 $r \in [1: R]$ ，它是预测变量 \mathcal{X} 与回归系数 \mathbf{B} 除 $\mathbf{U}^{(l)}$ 外的所有剩余成分的收缩张量积，收缩阶数为 $L - 1$ 阶。由于少收缩了 $\mathbf{U}^{(l)}$ 这一维，辅助张量的维度为 $\mathcal{C}_r \in \mathbb{R}^{N \times P_l \times Q_1 \times \dots \times Q_M}$ 。

$$\mathcal{C}_r = \langle \mathcal{X}, \mathbf{U}_r^{(1)} \circ \dots \circ \mathbf{U}_r^{(l-1)} \circ \mathbf{U}_r^{(l+1)} \circ \dots \circ \mathbf{U}_r^{(L)} \circ \mathbf{V}_r^{(1)} \circ \dots \circ \mathbf{V}_r^{(M)} \rangle_{L-1}$$

将辅助张量 \mathcal{C}_r 沿着 P_l 这一维度进行展开，得到矩阵 \mathbf{C}_r ，其维度为 $\mathbf{C}_r \in \mathbb{R}^{NQ \times P_l}$ ，其中 $Q = Q_1 \times \dots \times Q_M$ 。将矩阵 \mathbf{C}_r 按照 r 从 1 到 R 的顺序横向排列，最终可以得到辅助矩阵 $\mathbf{C} = [\mathbf{C}_1 \dots \mathbf{C}_R]$ ，其维度为 $\mathbf{C} \in \mathbb{R}^{NQ \times P_l R}$ 。

除辅助张量 \mathcal{C}_r 外，再定义辅助张量 \mathcal{F}_r ，如下式所示，其为单位张量 \mathcal{J} 与除 $\mathbf{U}^{(l)}$ 外的所有剩余成分矩阵的 $L - 1$ 阶收缩张量积，其中单位张量 \mathcal{J} 的维度为 $\mathcal{J} \in \mathbb{R}^{P_1 \times \dots \times P_L \times P_1 \times \dots \times P_{l-1} \times P_{l+1} \times \dots \times P_L}$ 。

$$\mathcal{F}_r = \langle \mathcal{J}, \mathbf{U}_r^{(1)} \circ \dots \circ \mathbf{U}_r^{(l-1)} \circ \mathbf{U}_r^{(l+1)} \circ \dots \circ \mathbf{U}_r^{(L)} \circ \mathbf{V}_r^{(1)} \circ \dots \circ \mathbf{V}_r^{(M)} \rangle_{L-1}$$

式中， \mathcal{F}_r 的维度为 $\mathcal{F}_r \in \mathbb{R}^{P_1 \times \dots \times P_L \times P_l \times Q_1 \times \dots \times Q_M}$ ，将 \mathcal{F}_r 沿着 P_l 维展开，可以得到矩阵 \mathbf{F}_r ，其中 $P = \prod_{i=1}^L P_i$ 。将 \mathbf{F}_r 沿着 r 从 1 到 R 的顺序横向排列，可以得到辅助矩阵 $\mathbf{F} = [\mathbf{F}_1 \dots \mathbf{F}_R]$ ，其维度为 $\mathbf{F} \in \mathbb{R}^{PQ \times P_l R}$ 。借助辅助矩阵 \mathbf{C} 和 \mathbf{F} ，张量 Lasso 回归的优化公式可以被写成 (3.5)。

$$\begin{cases} \operatorname{argmin} & \| \operatorname{vec}(\mathcal{Y}) - \mathbf{C} \operatorname{vec}(\mathbf{U}^{(l)}) \|_F^2 + \lambda \| \mathbf{F} \operatorname{vec}(\mathbf{U}^{(l)}) \|_1 \\ \text{s. t.} & \operatorname{rank}(\mathbf{B}) \leq R \end{cases} \quad (3.5)$$

简记 $\mathbf{y} = \operatorname{vec}(\mathcal{Y})$ ， $\boldsymbol{\beta} = \operatorname{vec}(\mathbf{U}^{(l)})$ ，可将式 (3.5) 简写成两部分，其中 $f(\boldsymbol{\beta}) = \| \mathbf{y} - \mathbf{C} \boldsymbol{\beta} \|_F^2$ 是可导的凸函数， $g(\boldsymbol{\gamma}) = \| \mathbf{F} \boldsymbol{\beta} \|_1$ 是不可导的凸函数。通过简化得到式 (3.5) 的新形式，如式 (3.6) 所示。

$$\begin{cases} \operatorname{argmin} & f(\boldsymbol{\beta}) + \lambda g(\boldsymbol{\gamma}) \\ \text{s. t.} & \operatorname{rank}(\mathbf{B}) \leq R \\ & \boldsymbol{\gamma} = \mathbf{F} \boldsymbol{\beta} \end{cases} \quad (3.6)$$

此时，可以使用广义交替方向乘子法 (Aug-ADMM) 对式 (3.6) 进行更新求解。与普通 ADMM 相比，广义 ADMM 通过引入一个增广变量 $\tilde{\boldsymbol{\gamma}} = (\mathbf{W} - \mathbf{C}^T \mathbf{C})^{1/2} \boldsymbol{\beta}$ 来达到提高计算效率的效果。具体的理论推导和证明可以参见文献^[44]。增广变量中的 \mathbf{C} 和 $\boldsymbol{\beta}$ 正如前文所述，其中的 \mathbf{W} 是一个可以被灵活选择的矩阵， $\mathbf{W} = \delta \mathbf{I}$ ， \mathbf{I} 是单位矩阵，而 δ 是不小于 \mathbf{C} 的矩阵范数的一个值。参数的原始更新步骤如下式所示，其中 $\boldsymbol{\alpha}$ 是对偶变量， ρ 是正的惩罚参数。

$$\begin{cases} \boldsymbol{\beta}^{k+1} = \operatorname{argmin} f(\boldsymbol{\beta}) + \frac{\rho}{2} \left(\left\| \mathbf{C}\boldsymbol{\beta} - \boldsymbol{\gamma}^k - \frac{\boldsymbol{\alpha}^k}{\rho} \right\|_F^2 + \left\| (\mathbf{W} - \mathbf{C}^T\mathbf{C})^{-1/2}\boldsymbol{\beta} - \tilde{\boldsymbol{\gamma}}^k + \frac{\tilde{\boldsymbol{\alpha}}^k}{\rho} \right\|_F^2 \right) \\ \boldsymbol{\gamma}^{k+1} = \operatorname{argmin} g(\boldsymbol{\gamma}) + \left\| \frac{\rho}{2}\mathbf{C}\boldsymbol{\beta}^{k+1} - \boldsymbol{\gamma} + \frac{\boldsymbol{\alpha}^k}{\rho} \right\|_F^2 \\ \tilde{\boldsymbol{\gamma}}^{k+1} = (\mathbf{W} - \mathbf{C}^T\mathbf{C})^{1/2}\boldsymbol{\beta}^{k+1} + \frac{\tilde{\boldsymbol{\alpha}}^k}{\rho} \\ \boldsymbol{\alpha}^{k+1} = \boldsymbol{\alpha}^k + \rho(\mathbf{C}\boldsymbol{\beta}^{k+1} - \boldsymbol{\gamma}^{k+1}) \\ \tilde{\boldsymbol{\alpha}}^{k+1} = \tilde{\boldsymbol{\alpha}}^k + \rho((\mathbf{W} - \mathbf{C}^T\mathbf{C})^{1/2}\boldsymbol{\beta}^{k+1} - \tilde{\boldsymbol{\gamma}}^{k+1}) \end{cases}$$

通过近端映射，上式可以进一步简化为式 (3.7)，仅需更新 $\boldsymbol{\beta}$ 和 $\boldsymbol{\alpha}$ 即可，具体的简化原理本文不作赘述，可参见文献^[44]。

$$\begin{cases} \boldsymbol{\beta}^{k+1} = \operatorname{argmin} f(\boldsymbol{\beta}) + (2\boldsymbol{\alpha}^k - \boldsymbol{\alpha}^{k-1})^T \mathbf{C}\boldsymbol{\beta} + \frac{\rho}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}^k)^T \mathbf{W}(\boldsymbol{\beta} - \boldsymbol{\beta}^k) \\ \boldsymbol{\alpha}^{k+1} = \operatorname{prox}_{\rho g^*(\cdot)}(\boldsymbol{\alpha}^k + \rho \mathbf{C}\boldsymbol{\beta}^{k+1}) \end{cases} \quad (3.7)$$

$\mathbf{V}^{(m)}$ 的求解

成分矩阵 $\mathbf{V}^{(m)}$ 的求解与 $\mathbf{U}^{(l)}$ 类似，回归系数 \mathbf{B} 的后 M 模具有位置交换不变性，所以，本节将以 $\mathbf{V}^{(m)}$ 为例说明后 M 模的求解，其余的 $M - 1$ 模是类似的。定义向量 \mathbf{d}_r ，其为预测张量 \mathcal{X} 与除 $\mathbf{V}^{(m)}$ 以外的所有成分矩阵的 L 阶收缩张量积的向量化展开。

$$\mathbf{d}_r = \operatorname{vec}(\langle \mathcal{X}, \mathbf{U}_r^{(1)} \circ \dots \circ \mathbf{U}_r^{(L)} \circ \mathbf{V}_r^{(1)} \circ \dots \circ \mathbf{V}_r^{(m-1)} \circ \mathbf{V}_r^{(m+1)} \circ \dots \circ \mathbf{V}_r^{(M)} \rangle_L)$$

将向量 \mathbf{d}_r 按照 r 从 1 到 R 的顺序横向排列，得到辅助矩阵 $\mathbf{D} = [\mathbf{d}_1 \cdots \mathbf{d}_R]$ ，其维度为 $\mathbf{D} \in \mathbb{R}^{N_{Q-m} \times R}$ ，其中 $N_{Q-m} = Q_1 \times \cdots \times Q_{m-1} \times Q_{m+1} \times \cdots \times Q_M$ 。

除辅助矩阵 \mathbf{D} 外，定义辅助矩阵 \mathbf{G} ， $\mathbf{G} \in \mathbb{R}^{P_{Q-m} \times R}$ ，其第 r 列 \mathbf{g}_r 为除 $\mathbf{V}^{(m)}$ 以外所有的成分矩阵第 r 列内积的向量展开，如下式所示。沿着 CP 秩 $r = 1, \dots, R$ 将 \mathbf{g}_r 横向排列得到矩阵 $\mathbf{G} = [\mathbf{g}_1, \dots, \mathbf{g}_R]$ 。

$$\mathbf{g}_r = \operatorname{vec}(\mathbf{U}_r^{(1)} \circ \dots \circ \mathbf{U}_r^{(L)} \circ \mathbf{V}_r^{(1)} \circ \dots \circ \mathbf{V}_r^{(m-1)} \circ \mathbf{V}_r^{(m+1)} \circ \dots \circ \mathbf{V}_r^{(M)})$$

借助所构造的辅助矩阵 \mathbf{D} 和 \mathbf{G} ，可以将式 (3.4) 改写为式 (3.8)，逐列使用上述的广义 ADMM 算法求解即可。

$$\begin{cases} \operatorname{argmin} & \| \mathbf{Y}_{[m]}^T - \mathbf{D}\mathbf{V}^{(m)T} \|_F^2 + \lambda \| \mathbf{G}\mathbf{V}^{(m)T} \|_1 \\ \text{s. t.} & \operatorname{rank}(\mathcal{B}) \leq R \end{cases} \quad (3.8)$$

伪代码算法 3 总结了基于 CP 分解的张量 Lasso 回归模型的求解步骤。

算法 4 基于 CP 分解的张量 Lasso 回归

输入：自变量 \mathcal{X} ，响应变量 \mathbf{y} ，惩罚参数 λ ，迭代次数 $Iter$ ，秩 R ；

输出：张量回归系数 \mathcal{B} ；

- 1: For $t = 1, \dots, Iter$:
 - 2: For $l = 1, \dots, L$:
 - 3: 使用式 (3.7) 更新 $\mathbf{U}^{(l)}$;
 - 4: End For
 - 5: For $m = 1, \dots, M$:
 - 6: 使用式 (3.8) 更新 $\mathbf{V}^{(m)}$;
 - 7: End For
 - 8: End For
 - 9: 使用更新后的 $\mathbf{U}^{(l)}$ 和 $\mathbf{V}^{(m)}$ 来计算回归系数 \mathcal{B} ;
 - 10: Return 张量回归系数 \mathcal{B} 。
-

3.2 模拟实验

为了测试以上三种张量回归方法的表现，使用随机生成的数据来进行拟合测试。在不同的条件下，考察使用三阶预测变量 \mathcal{X} 对三阶响应变量 \mathbf{y} 的预测能力，使用训练集进行模型训练和参数学习，使用测试集考察模型的表现，实验均在 Intel(R) Core(TM) i5-8257U CPU @ 1.40GHz 的处理器上进行，所使用的软件为 R 4.0.3。

本节将从三个方面来设计模拟实验进行对比分析，分别是样本量、信噪比，和数据的维度。首先，当样本量 N 较大时，模型通常可以取得比较好的拟合效果，而当样本量较少时，则可能会出现过拟合或是欠拟合的情况，故选取样本量 N 分别为 50 和 200 进行实验，以考察在不同样本量下模型的性能表现。其次，信噪比水平也会影响模型的拟合效果，当信噪比大时，有效信息多，能够比较轻易取得较好的拟合效果，而当信噪比较少时，干扰信息多，拟合能力受到挑战，故选取信噪比水平 SNR 为 1 和 10 进行实验。最后，考察样本数据维度的影

响，当自变量和响应变量具有不同的维度时，拟合能力可能会有所差异，本文也在实验中考察这一差异，选取了三种情况，分别是 \mathbf{y} 的维度大于 \mathbf{x} 、 \mathbf{x} 的维度大于 \mathbf{y} ，和 \mathbf{x} 与 \mathbf{y} 的维度相近。

为了保证结果能代表模型真实的表现，不使用某一次的拟合结果来说明问题，而是将对每一种情况生成十次不同的数据，使用十次建模拟合的平均值来代表模型的表现，对于上述三个方面下的十二种情况，共计一百二十个模拟数据集，数据生成的具体过程如下所示。

1. 从正态分布 $N(0,1)$ 中生成自变量 $\mathbf{X} \in \mathbb{R}^{N \times P_1 \times P_2}$ 和残差 $\boldsymbol{\varepsilon} \in \mathbb{R}^{N \times Q_1 \times Q_2}$;
2. 从正态分布 $N(0,1)$ 中生成 $\mathbf{U}^{(l)} \in \mathbb{R}^{P_l \times R}$ ，其中 $l = 1, \dots, L$ ；
从正态分布 $N(0,1)$ 中生成 $\mathbf{V}^{(m)} \in \mathbb{R}^{Q_m \times R}$ ，其中 $m = 1, \dots, M$ ；
3. 对回归系数 \mathbf{B} 进行变换， $\mathbf{B} = c \left[[\mathbf{U}^{(1)}, \dots, \mathbf{U}^{(L)}, \mathbf{V}^{(1)}, \dots, \mathbf{V}^{(M)}] \right]$ ，使其满足下式，最后得到响应变量 $\mathbf{y} = \langle \mathbf{X}, \mathbf{B} \rangle_L + \boldsymbol{\varepsilon}$ 。

$$\frac{\|\langle \mathbf{X}, \mathbf{B} \rangle\|_F^2}{\|\boldsymbol{\varepsilon}\|_F^2} = SNR$$

对于拟合效果的判定，本文选择了三个衡量指标，分别是相关系数 cor 、 Q^2 和均方根误差 $RMSE$ 。其中 cor 和 Q^2 是正向指标，数值越大越好，均方根误差 $RMSE$ 是负向指标，数值越小越好。

$$\begin{aligned} cor &= cor(\mathbf{y}, \hat{\mathbf{y}}) \\ Q^2 &= 1 - \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|_F^2}{\|\mathbf{y}\|_F^2} \\ RMSE &= \sqrt{\frac{1}{NQ} (\mathbf{y} - \hat{\mathbf{y}})^2} \end{aligned}$$

3.4.1 训练集样本量的影响

为了使数据具有代表意义，每一个数值均由 60 次 (3 个维度 \times 2 个信噪比水平 \times 10 次重复模拟) 模拟平均得出。如图 3-1 所示，当样本量较大时，三个模型总体表现较好，当样本量较小时，三个模型的性能均有所下滑，但基于 CP 分解的 Lasso 回归下滑幅度最少。

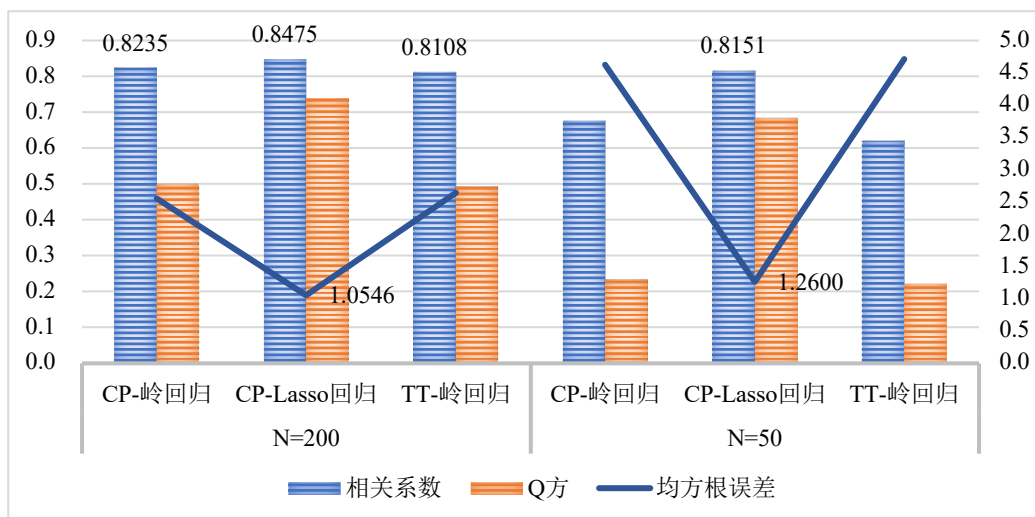


图 3-1 不同样本量下预测任务的预测性能

Figure 3-1 Prediction performance of prediction tasks under different sample sizes

如表 3-1 所示, 当样本量为 200 时, 基于 CP 分解的 Lasso 回归取得了最好的预测效果, Q 方达到 73.95%, 其次是基于 CP 分解的岭回归, Q 方为 49.97%, 最后是基于 TT 分解的岭回归模型, 相关系数为 0.8108, Q 方为 49.31%。当样本量为 50 时, 基于 CP 分解的张量 Lasso 回归与两个岭回归模型拉开了显著差距, 相关系数依旧保持在 0.8 以上, 而两个岭回归的模型均不足 0.7。实验表明, 本文提出的基于 CP 分解的 Lasso 回归具有更强的学习能力, 但是由于使用了交替求解而非直接进行矩阵计算导致在训练时间上高于两个岭回归模型。

表 3-1 不同样本量下预测任务的预测性能

Table 3-1 Prediction performance of prediction tasks under different sample sizes

模型	样本量	相关系数	Q 方	均方根误差	运行时间 (s)
CP-岭回归	200	0.8235	0.4997	2.5508	0.6996
CP-Lasso 回归		0.8475	0.7395	1.0546	2.5139
TT-岭回归		0.8108	0.4931	2.6413	0.4896
CP-岭回归	50	0.6760	0.2354	4.6212	0.6782
CP-Lasso 回归		0.8151	0.6842	1.2600	2.8770
TT-岭回归		0.6192	0.2213	4.7125	0.7141

3.4.2 信噪比的影响

每一个数值均由 60 次 (3 个维度 \times 2 个样本量 \times 10 次重复模拟) 模拟平均得出。如图 3-2 所示, 当信噪比水平较高时, 三个模型表现整体较好, Q 方最高达到了 97.45%, 当信噪比水平较低时, 模型表现相对一般, Q 方最高仅为 44.91%。

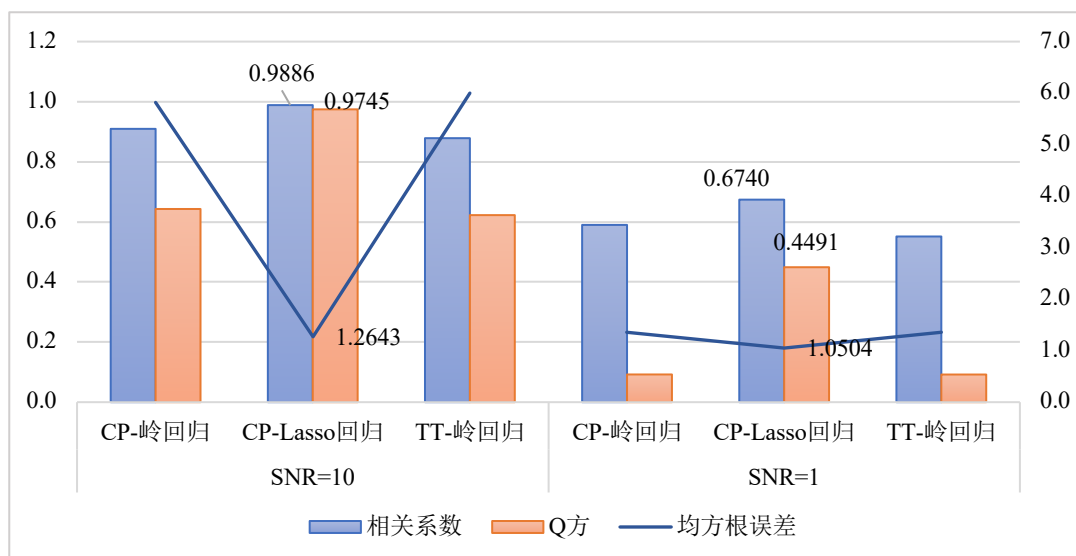


图 3-2 不同信噪比下预测任务的预测性能

Figure 3-2 Prediction performance of prediction tasks under different signal-to-noise ratios

具体数值如表 3-2 所示，当信噪比为 10 时，三个模型都表现出了非常强的预测能力，尤其是基于 CP 分解的张量 Lasso 回归，相关系数 cor 达到 0.9886， Q^2 达到了 97.45%，但是运行时间也最长，平均为 2.60 秒，模型拟合效果位居其次的是基于 CP 分解的张量岭回归，其相关系数为 0.9096， Q 方为 64.29%，运行时间为 0.79 秒。当信噪比为 1 时，三个模型的拟合效果都受到了限制，但 CP-Lasso 回归依旧具有相对较强的学习能力，在测试集上的相关系数为 0.6740，与其余两个岭回归模型拉开了差距。

表 3-2 不同信噪比下预测任务的预测性能

Table 3-2 Prediction performance of prediction tasks under different signal-to-noise ratios

模型	信噪比	相关系数	Q 方	均方根误差	运行时间 (s)
CP-岭回归	10	0.9096	0.6429	5.8182	0.7864
CP-Lasso 回归		0.9886	0.9745	1.2643	2.5992
TT-岭回归		0.8786	0.6225	5.9999	0.5363
CP-岭回归	1	0.5899	0.0922	1.3537	0.5915
CP-Lasso 回归		0.6740	0.4491	1.0504	2.7916
TT-岭回归		0.5514	0.0920	1.3539	0.6674

3.4.3 数据维度的影响

每一个数值均由 40 次 (2 个样本量 \times 2 个信噪比水平 \times 10 次重复模拟) 模拟平均得出。如图 3-3 所示，不同的数据维度上三个模型的整体表现有一定差异。

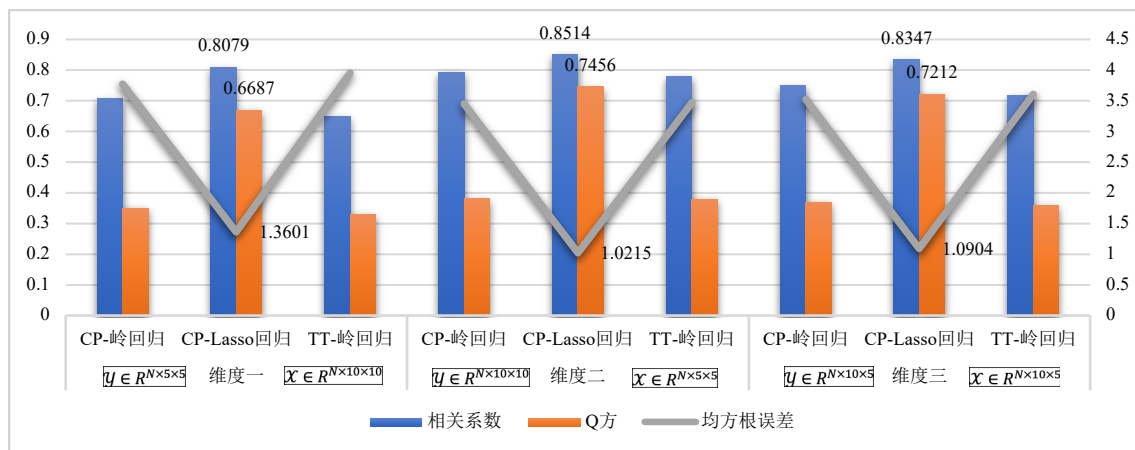


图 3-3 不同数据维度下预测任务的预测性能

Figure 3-3 Prediction performance of prediction tasks under different data dimensions

具体数值如表 3-3 所示，维度二中，响应变量 $y \in \mathbb{R}^{N \times 10 \times 10}$ 的维度较大而预测变量 $X \in \mathbb{R}^{N \times 5 \times 5}$ 的维度较小，三个模型整体表现得最好，最高的相关系数达到了 0.8514，最低的也达到了 0.7783。维度一中，响应变量 $y \in \mathbb{R}^{N \times 5 \times 5}$ 的维度较小且预测变量 $X \in \mathbb{R}^{N \times 10 \times 10}$ 的维度较大，预测效果整体低于其他两个数据维度，对比维度二降低了 5%-13% 不等。维度三中，数据的大小介于上述两者之间 $y \in \mathbb{R}^{N \times 10 \times 5}$ ， $X \in \mathbb{R}^{N \times 10 \times 5}$ ，拟合效果也介于两者之间。结论表明，对于本文所阐述的三种张量回归模型而言，响应变量的维度越大，预测效果越好。

表 3-3 不同数据维度下预测任务的预测性能

Table 3-3 Prediction performance of prediction tasks under different data dimensions

模型	维度	相关系数	Q方	均方根误差	运行时间(s)
CP-岭回归	$X \in \mathbb{R}^{N \times 10 \times 10}$ $y \in \mathbb{R}^{N \times 5 \times 5}$	0.7074	0.3504	3.7779	0.8436
CP-Lasso 回归		0.8079	0.6687	1.3601	2.3574
TT-岭回归		0.6488	0.3301	3.9505	0.8527
CP-岭回归	$X \in \mathbb{R}^{N \times 5 \times 5}$ $y \in \mathbb{R}^{N \times 10 \times 10}$	0.7931	0.3820	3.4523	0.5072
CP-Lasso 回归		0.8514	0.7456	1.0215	2.9753
TT-岭回归		0.7783	0.3812	3.4746	0.3049
CP-岭回归	$X \in \mathbb{R}^{N \times 10 \times 5}$ $y \in \mathbb{R}^{N \times 10 \times 5}$	0.7488	0.3702	3.5277	0.7159
CP-Lasso 回归		0.8347	0.7212	1.0904	2.7536
TT-岭回归		0.7179	0.3603	3.6056	0.6480

通过上述三个方面的对比，可以发现在相同的参数之下，CP-Lasso 回归比张量岭回归具有更强的学习能力和泛化能力，整体的拟合效果均有一定幅度的提升，尤其是在小样本和低信噪比等严苛的条件下。但是，可以发现张量 Lasso

回归也存在一定短板，即运行时间高于张量岭回归模型，但这仅仅是在最理想的模拟数据之上，在后续真实的数据集上也将对模型的运行时间加以考察。

3.3 案例分析

本节用一个真实的数据集来进行案例分析，并在案例中考察三个模型的拟合预测效果，进行比较说明。数据集选自 UCI 机器学习数据库中的空气质量数据集 (Air Quality)，该数据集由 De Vito 等人于 2016 年公开共享^[45]。记录了 2004 年 3 月至 2005 年 2 月期间分小时的空气质量相关指标的数据，共计 9358 条记录。由于传感器所具有的长期稳定性、自身选择性和时间空间影响的问题，会导致测量所得的浓度与实际的气体浓度之间存在差距，有研究表明分别对气体浓度进行单变量校准时，效果往往不那么理想，将多个变量协同考虑时，取得了较为可观的效果^[46]。所以将上述的模型算法应用在此案例中，考察模型在真实数据集上的表现。

表 3-4 空气质量数据集特征说明

Table 3-4 Description of the characteristics of the Air Quality dataset

变量类型	变量名称	物理单位	时间单位
预测变量	传感器 - CO 浓度	mg/m ³	每小时
	传感器 - NMHC 浓度	microg/m ³	每小时
	传感器 - Nox 浓度	ppb	每小时
	传感器 - NO2 浓度	microg/m ³	每小时
	传感器 - O3 浓度	microg/m ³	每小时
	温度	°C	-
	相对湿度	%	-
	绝对湿度	-	-
响应变量	真实 - CO 浓度	mg/m ³	每小时
	真实 - NMHC 浓度	microg/m ³	每小时
	真实 - Benzene 浓度	microg/m ³	每小时
	真实 - Nox 浓度	ppb	每小时
	真实 - NO2 浓度	microg/m ³	每小时

先对数据进行处理，将其还原为维度为 (天 × 小时 × 变量) 的张量形式，并将传感器所得的温度、湿度以及气体浓度等作为预测变量，将实际的气体浓度作为响应变量，进行拟合预测。训练集和测试集按照 7: 3 的比例进行划分，使用模型在测试集上的指标数据来衡量模型的表现，另外为了使得模型之间具有可比性，参数统一选择为：回归系数 B 的秩为 $R=3$ (在 CP 模型中 CP 秩为 R ，在

TT 模型中 TT 秩为秩向量均为 R)，惩罚参数 $\lambda=10$ ，CP-Lasso 回归的迭代次数为 30 次，模型的拟合结果如表 3-5 所示。

表 3-5 三个惩罚回归模型在空气质量数据集的预测性能

Table 3-5 Prediction performance of three penalized regression models in the Air Quality dataset

模型	相关系数	Q 方	均方根误差	运行时间(s)
CP-岭回归	0.7695	0.5877	122.91	21.08
CP-Lasso 回归	0.7783	0.6033	120.55	24.42
TT-岭回归	0.7612	0.5658	126.12	32.85

观察表 3-5 中模型在空气质量数据集上的表现，可以发现三个模型都取得了相对可观的校准效果，尤其是基于 CP 分解的张量 Lasso 回归模型，在相关系数、Q 方和均方根误差上取得了较好的效果，其次是基于 CP 分解的张量岭回归模型。基于 TT 分解的张量岭回归模型表现相对而言不那么理想，这可能是由于此空气质量数据集的内部数据结构更加接近 CP 模型。在运行时间的表现上，基于 CP 分解的张量 Lasso 回归还算理想，并没有与岭回归模型拉开过大差距。

3.4 本章小结

本章介绍了三个频率学派下的带惩罚项的张量回归模型，并通过以上对三个模型在模拟数据集和真实数据集上的对比分析，得到了以下四个结论：首先，张量回归模型整体在响应变量维度大的情况下具有更强的学习能力和拟合效果；其次，在大样本和高信噪比的情景下，模型之间的差异不显著，均能取得较为可观的拟合结果；再次，在小样本和低信噪比等严苛的条件下，Lasso 回归具有比岭回归更强的学习能力；最后，张量 Lasso 回归的训练时间总体而言高于张量岭回归模型，但也取决于数据结构和分解模型秩的选择，在空气质量数据集上的气体浓度校准任务上，在相同的秩下，与岭回归模型的运行时间差异不大。

第4章 贝叶斯张量回归算法

前述的三种张量正则回归是频率学派下的点估计方法，认为参数是一个具体的数值，固定不变，进而基于极大似然法进行求解输出点估计值。而在贝叶斯学派下，参数被认为是一个变量，有自己的先验分布函数，可以结合参数的先验分布和样本数据信息两部分来计算参数的后验分布，使用最大后验法进行参数的估计，可以通过其后验样本的分布来量化预测值的不确定性。贝叶斯模型的基本形式如式 (4.1) 所示，其中， $P(\theta)$ 是参数 θ 的先验概率分布，是指在得到数据之前对参数 θ 的认识和理解， $P(data|\theta)$ 是样本似然函数，其中包含了参数 θ 的信息，而 $P(\theta|data)$ 是参数 θ 的后验概率分布，是指结合了先验知识和样本数据知识两方面的信息之后对参数 θ 的认识和理解。

$$P(\theta|data) = \frac{P(\theta)P(data|\theta)}{P(data)} \propto P(\theta)P(data|\theta) \quad (4.1)$$

有了后验概率分布之后，就可以对参数有进一步的认识，但是后验分布通常比较复杂，无法轻而易举地获取到一些诸如均值之类的统计特性，所以在计算机技术发展日益成熟的今天，随机模拟的方法可以用来解决这个问题。随机模拟中最为广泛使用的是马尔科夫链蒙特卡洛算法，也称 MCMC 算法，可以从一个复杂分布中进行采样，当样本量充分时，根据大数定理，样本均值会逼近总体均值。

$$E(\theta) = \int_{\theta} \theta P(\theta|data) d\theta \approx \frac{1}{N} \sum_{i=1}^N \theta_i$$

MCMC 方法将马尔科夫链引入蒙特卡洛模拟中，其思路是构造一条马尔科夫链，使其平稳分布为参数的后验分布，从平稳分布中进行采样，使用采样得到的样本进行蒙特卡洛积分。常用的 MCMC 方法有 Metropolis-Hasting 算法 (M-H 算法) 和 Gibbs 算法。Gibbs 算法是 M-H 算法的一种特殊形式，适用于对高维数据进行采样，对于 K 维张量 $\mathcal{A} \in \mathbb{R}^{P_1 \times \dots \times P_K}$ 采样的步骤如下。

1. 输入：密度函数 $P(\mathcal{A})$ ；收敛步数 m ；采样数量 n ；
2. 进行 $m + n$ 次采样；
3. 在每次采样中，依次对 K 个维度的条件分布进行采样，形成一个样本；
4. 输出：舍弃前 m 个样本，得到 n 个样本合集 $\mathcal{A}^{(m+1)}, \dots, \mathcal{A}^{(m+n)}$ 。

根据对张量回归模型中的回归系数 \mathbf{B} 假定不同的先验分布，本章将介绍两个贝叶斯张量回归模型，分别是基于高斯先验的贝叶斯张量回归和基于拉普拉斯的贝叶斯张量回归。

4.1 基于拉普拉斯先验的贝叶斯张量回归

在 2.6 节中阐述的基于高斯先验的贝叶斯张量回归算法关键在于假设回归系数 \mathbf{B} 的先验分布服从球形高斯分布，如图 4-1 高斯概率密度图所示，假设系数 \mathbf{B} 服从均值为 0 的高斯分布即表明 \mathbf{B} 的元素应当分布在 0 的附近，这与频率学派下加入正则项来进行稀疏约束有异曲同工之妙。

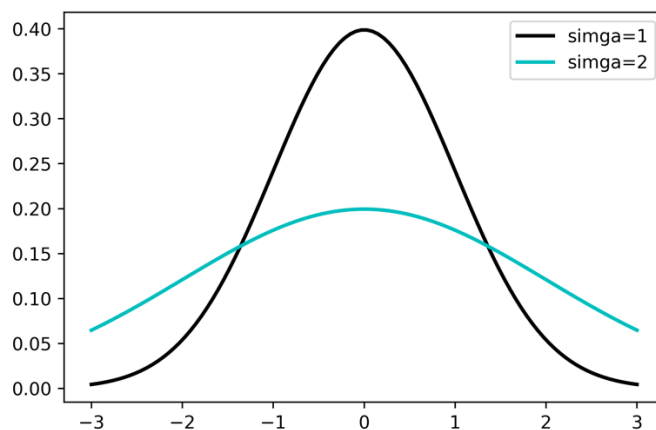


图 4-1 均值为 0 不同方差的高斯分布

Figure 4-1 Gaussian distributions with mean 0 under different variances

基于高斯分布的贝叶斯张量回归是一个半贝叶斯框架的，仅仅对回归系数 \mathbf{B} 和随机干扰项的方差设置了先验，在本节提出的基于拉普拉斯先验的张量贝叶斯回归中，加入了对惩罚参数 λ 的先验，并以此先验为基础，实现进一步自动秩确定的效果。

$$P(x) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right) \quad (4.2)$$

若 x 服从位置参数为 μ ，尺度参数为 b 的一维拉普拉斯分布，记为 $x \sim La(\mu, b)$ ，其密度函数如式 (4.2) 所示。但对于多维拉普拉斯分布的定义有多种不同的形式，没有统一的定论^[47-49]。结合研究目的，本文根据回归系数 \mathbf{B} 的每一个元素都独立同分布于 0 均值的一维拉普拉斯分布来定义多维拉普拉斯分布，其形式是 Eltoft 提出的多维拉普拉斯分布^[47]的一种特例，即各个元素之间相互独立。如图 4-2 所示，拉普拉斯的分布相较于高斯分布而言，在均值 0 附近更尖更陡，也就表明在相同的惩罚参数之下，具有拉普拉斯先验的参数稀疏程度更大，约束更为严格。

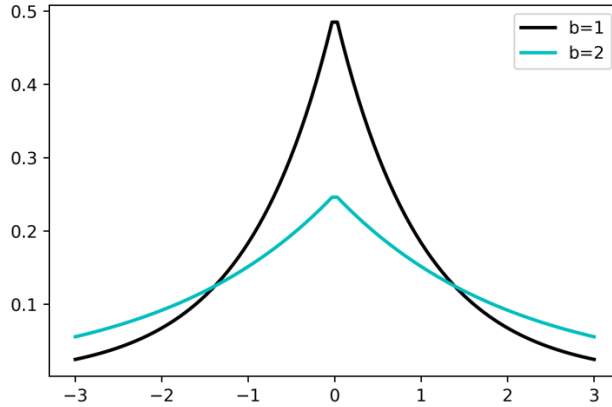


图 4-2 参数位置为 0 不同尺度参数的拉普拉斯分布

Figure 4-2 Laplace distributions of parameters at different scales with parameter position 0

4.1.1 先验分布

由于直接为回归系数 \mathbf{B} 设置拉普拉斯先验会使得后续的采样过程比较困难，所以可以通过引入辅助变量 $\boldsymbol{\tau} = (\tau_1, \dots, \tau_R)$ ，以更为简单和更为可行的方式达到拉普拉斯分布约束的效果，其中 τ_1, \dots, τ_R 均为非负实数。

$$P(\tau_r | \lambda_r^2) = \frac{\lambda_r^2}{2} \exp\left(-\frac{\lambda_r^2}{2} \tau_r\right) = Ga\left(1, \frac{\lambda_r^2}{2}\right) \quad (4.3)$$

τ_r 的条件先验是一个指数分布，指数分布是伽马分布的一个特例，为了后续书写的简便性，记 τ_r 的条件先验服从形状参数为 1，逆尺度参数为 $\lambda_r^2/2$ 的伽马分布。若 x 服从于形状参数为 a ，逆尺度参数为 b 的伽马分布，记为 $x \sim Ga(a, b)$ ，其密度函数如式 (4.4) 所示。

$$Ga(a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx) \quad (4.4)$$

在 $\boldsymbol{\tau}$ 的条件下，假设条件先验 $P(\text{vec}(\mathbf{U}^{(l)}) | \boldsymbol{\tau})$ 服从高斯分布 $N(0, \tilde{\mathbf{I}}_{\boldsymbol{\tau}})$ ，其中 $\tilde{\mathbf{I}}_{\boldsymbol{\tau}} \in \mathbb{R}^{RP_l \times RP_l} = \text{diag}(\text{diag}(\tau_1), \dots, \text{diag}(\tau_R))$ 是一个对角矩阵。

$$\tilde{\mathbf{I}}_{\boldsymbol{\tau}} = \begin{pmatrix} \tau_1 & & & & & \\ & \ddots & & & & \\ & & \tau_1 & & & \\ & & & \ddots & & \\ & & & & \tau_R & \\ & & & & & \ddots \\ & & & & & & \tau_R \end{pmatrix}$$

结合 $P(\tau_r|\lambda_r^2)$ 和 $P(\text{vec}(\mathbf{U}^{(l)})|\boldsymbol{\tau})$ 的式子, 可以得出 $\text{vec}(\mathbf{U}^{(l)})|\boldsymbol{\lambda}$ 条件先验是服从于拉普拉斯分布的, 具体的推导公式如式 (4.5) 所示。为简化公式, 记 $\text{vec}(\mathbf{U}^{(l)})$ 为 $\boldsymbol{\beta}^l$, 其中, $\tilde{\mathbf{I}}_\lambda = \text{diag}(\text{diag}(\lambda_1), \dots, \text{diag}(\lambda_R)) \in \mathbb{R}^{RP_l \times RP_l}$, 其定义类似于上述的 $\tilde{\mathbf{I}}_\tau$ 。

$$\begin{aligned}
 P(\boldsymbol{\beta}^l|\boldsymbol{\lambda}^2) &= P(\boldsymbol{\beta}^l|\boldsymbol{\tau})P(\boldsymbol{\tau}|\boldsymbol{\lambda}^2) = \prod_{r=1}^R \prod_{p=1}^{P_l} P(\mathbf{U}_{pr}^{(l)}|\tau_r)P(\tau_r|\lambda_r^2) \\
 &= \prod_{r=1}^R \prod_{p=1}^{P_l} N(0, \tau_r)Ga\left(1, \frac{\lambda_r^2}{2}\right) \\
 &= \prod_{r=1}^R \prod_{p=1}^{P_l} \frac{1}{\sqrt{2\pi\tau_r}} \exp\left(-\frac{\mathbf{U}_{pr}^{(l)2}}{2\tau_r}\right) \frac{\lambda_r^2}{2} \exp\left(-\frac{\lambda_r^2}{2}\tau_r\right) \\
 &= \prod_{r=1}^R \prod_{p=1}^{P_l} \int_0^\infty \frac{1}{\sqrt{2\pi\tau_r}} \exp\left(-\frac{\mathbf{U}_{pr}^{(l)2}}{2\tau_r}\right) \frac{\lambda_r^2}{2} \exp\left(-\frac{\lambda_r^2}{2}\tau_r\right) d\tau_r \\
 &= \prod_{r=1}^R \prod_{p=1}^{P_l} \frac{\lambda_r}{2} \exp\left(\lambda_r \|\mathbf{U}_{pr}^{(l)}\|\right) \propto \prod_{r=1}^R \exp\left(\lambda_r \|\mathbf{U}_r^{(l)}\|_1\right) \\
 &\propto \exp\left(\|\tilde{\mathbf{I}}_\lambda \boldsymbol{\beta}^l\|_1\right)
 \end{aligned} \tag{4.5}$$

在式(4.5)中, 使用了正态分布的尺度混合, 其定义如式(4.6)所示, 将高斯分布和指数分布混合成了拉普拉斯分布的形式, 具体内容可参见^[50]。

$$\int_0^\infty \frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{z^2}{2s}\right) \frac{a^2}{2} \exp\left(-\frac{a^2}{2}s\right) ds = \frac{a}{2} \exp(-a|z|) \tag{4.6}$$

对于回归系数 \mathbf{B} 的条件先验分布已经构造完成, 记为多维拉普拉斯分布。可以发现, 当 $\lambda_1, \dots, \lambda_R$ 都取一样的值 λ 时, 即对成分矩阵 $\mathbf{U}^{(l)}$ 的 R 个列进行无差别的惩罚, $P(\boldsymbol{\beta}^l|\boldsymbol{\lambda}^2) \propto \exp(\lambda \|\boldsymbol{\beta}^l\|_1)$, 进而可以得到 $P(\mathbf{B}|\boldsymbol{\lambda}^2) \propto \exp(\lambda \|\mathbf{B}\|_1)$, 其与前章所述的 CP-Lasso 回归施加 l_1 范数具有一致的形式, 但对于此种情况不再过多讨论, 在本节中, 希望对 CP 分解的成分张量的不同的列上施加不同的惩罚力度, 以达到自动秩确定的效果。

$$P(\boldsymbol{\beta}^l|\boldsymbol{\lambda}^2) \propto \exp(\|\tilde{\mathbf{I}}_\lambda \boldsymbol{\beta}^l\|_1)$$

除了回归系数 \mathbf{B} 和辅助参数 $\boldsymbol{\tau}$ 外, 也为惩罚参数 $\boldsymbol{\lambda}^2$ 和随机波动项 σ^2 设置先验分布。由于惩罚参数本身就具有正实数的现实意义, 它代表对回归系数 \mathbf{B} 的约束程度, 所以, 可以假设 $\boldsymbol{\lambda}^2 = (\lambda_1^2, \dots, \lambda_R^2)$ 的每一项 λ_r^2 都独立同分布于伽马分布 $Ga(a_r, b_r)$, 可以得到其先验分布函数如式 (4.7) 所示。

$$\begin{aligned}
 P(\boldsymbol{\lambda}^2) &= \prod_{r=1}^R P(\lambda_r^2) = \prod_{r=1}^R Ga(a_r, b_r) \\
 &= \prod_{r=1}^R \frac{b_r^{a_r}}{\Gamma(a_r)} \lambda_r^{2(a_r-1)} \exp(-b_r \lambda_r^2)
 \end{aligned} \tag{4.7}$$

由于对随机波动项的方差没有明显的偏好，与基于高斯分布的贝叶斯张量回归类似，也为 σ^2 设置 Jeffery 先验， $P(\sigma^2) \propto 1/\sigma^2$ 。综上，已经为除了 CP 模型的秩 R 外的所有参数都设置了先验分布，将该贝叶斯模型框架总结于式 (4.8)。

$$P(\mathcal{Y}) = P(\mathcal{Y}|\mathcal{X}, \mathcal{B}, \sigma^2) \cdot P(\mathcal{B}|\boldsymbol{\tau}) \cdot P(\boldsymbol{\tau}|\boldsymbol{\lambda}^2) \cdot P(\boldsymbol{\lambda}^2) \cdot P(\sigma^2) \quad (4.8)$$

4.1.2 后验分布

在残差项服从均值为 $\mathbf{0}$ ，方差为 σ^2 时，样本似然函数如式 (4.9) 所示。结合似然函数，本节将一一对参数的后验分布形式进行计算推演。

$$P(\mathcal{Y}|\sigma^2, \mathcal{B}, \mathcal{X}) \propto \exp\left(-\frac{1}{2\sigma^2} \|\mathcal{Y} - \langle \mathcal{X}, \mathcal{B} \rangle_L\|_F^2\right) \quad (4.9)$$

\mathcal{B} 的后验分布

在对回归系数 \mathcal{B} 的后验分布求解的过程中，简记 $\mathbf{U}^{(l)}$ 为 $\boldsymbol{\beta}^l$ ，且 $\boldsymbol{\beta}^{-l}$ 为除 $\mathbf{U}^{(l)}$ 以外的所有 CP 分解成分矩阵。 $\boldsymbol{\beta}^l$ 的后验分布函数正比于其先验分布函数与似然函数之积，并且与 $\boldsymbol{\beta}^l$ 无关的项都可以被视为常数项暂时忽略。

$$\begin{aligned} P(\boldsymbol{\beta}^l|\mathcal{X}, \mathcal{Y}, \sigma^2, \boldsymbol{\beta}^{-l}) &= P(\boldsymbol{\beta}^l|\boldsymbol{\lambda}^2)P(\text{vec}(\mathcal{Y})|\mathbf{C}, \boldsymbol{\beta}^l, \sigma^2) \\ &= P(\boldsymbol{\beta}^l|\boldsymbol{\tau})P(\boldsymbol{\tau}|\boldsymbol{\lambda}^2)P(\text{vec}(\mathcal{Y})|\mathbf{C}, \boldsymbol{\beta}^l, \sigma^2) \\ &\propto P(\boldsymbol{\beta}^l|\boldsymbol{\tau})P(\text{vec}(\mathcal{Y})|\mathbf{C}, \boldsymbol{\beta}^l, \sigma^2) \\ &\propto N(\mathbf{0}, \tilde{\mathbf{I}}_{\boldsymbol{\tau}})N(\mathbf{C}\boldsymbol{\beta}^l, \mathbf{I}_{\sigma}) \\ &\propto N(\hat{\boldsymbol{\mu}}_l, \hat{\boldsymbol{\Sigma}}_l) \end{aligned} \quad (4.10)$$

在式 (4.10) 中， $\mathbf{I}_{\sigma} = \text{diag}(\sigma^2) \in \mathbb{R}^{RP_l \times RP_l}$ 是一个对角矩阵，由于正态分布共轭于正态分布，所以 $P(\boldsymbol{\beta}^l|\mathcal{X}, \mathcal{Y}, \sigma^2, \boldsymbol{\beta}^{-l})$ 仍然是正态分布，只是均值和协方差发生了变化，故只需要从 $N(\hat{\boldsymbol{\mu}}_l, \hat{\boldsymbol{\Sigma}}_l)$ 中进行采样即可， $\hat{\boldsymbol{\mu}}_l$ 和 $\hat{\boldsymbol{\Sigma}}_l$ 可以通过代数运算得出，如 (4.11) 所示。

$$\begin{cases} \hat{\boldsymbol{\mu}}_l &= (\mathbf{C}^T \mathbf{C} + \sigma^2 \tilde{\mathbf{I}}_{\boldsymbol{\tau}}^{-1})^{-1} \mathbf{C}^T \text{vec}(\mathcal{Y}) \\ \hat{\boldsymbol{\Sigma}}_l &= (\sigma^{-2} \mathbf{C}^T \mathbf{C} + \tilde{\mathbf{I}}_{\boldsymbol{\tau}})^{-1} \end{cases} \quad (4.11)$$

同理，对于 $\mathbf{V}^{(m)}$ 的后验分布的求解与 $\mathbf{U}^{(l)}$ 类似，其中 \mathbf{I}_m 是 Q_m 阶单位矩阵。

$$\begin{aligned}
 P(\boldsymbol{\beta}^m | \mathcal{X}, \mathcal{Y}, \sigma^2, \boldsymbol{\beta}^{-m}) &= P(\boldsymbol{\beta}^m | \lambda^2) P(\mathbf{Y}_{[m]} | \mathbf{D}, \boldsymbol{\beta}^{(m)}, \sigma^2) \\
 &\propto N(\mathbf{0}, \tilde{\mathbf{I}}_\tau) N((\mathbf{D} \otimes \mathbf{I}_m) \boldsymbol{\beta}^m, \mathbf{I}_\sigma) \\
 &\propto N(\hat{\boldsymbol{\mu}}_m, \hat{\boldsymbol{\Sigma}}_m)
 \end{aligned} \tag{4.12}$$

结合前述基于 CP 分解的张量 Lasso 回归模型，记 $\tilde{\mathbf{D}} = \mathbf{D} \otimes \mathbf{I}_m$ ，可以得到 $\boldsymbol{\beta}^m$ 后验分布的参数。至此，结合使用式 (4.10) 和式 (4.12) 即可从回归系数 \mathbf{B} 的后验分布中进行采样。

$$\begin{cases} \hat{\boldsymbol{\mu}}_m &= (\tilde{\mathbf{D}}^T \tilde{\mathbf{D}} + \sigma^2 \tilde{\mathbf{I}}_\tau^{-1})^{-1} \tilde{\mathbf{D}}^T \text{vec}(\mathbf{Y}_{[m]}) \\ \hat{\boldsymbol{\Sigma}}_m &= (\sigma^{-2} \tilde{\mathbf{D}}^T \tilde{\mathbf{D}} + \tilde{\mathbf{I}}_\tau)^{-1} \end{cases} \tag{4.13}$$

σ^2 的后验分布

σ^2 的先验是 Jeffery 无信息先验，结合似然函数，不难得出 σ^2 的后验分布服从于逆伽马分布 (IG)。

$$\pi(\sigma^2 | \mathcal{Y}, \mathcal{X}, \mathbf{B}) = \pi(\sigma^2) \pi(\mathcal{Y} | \mathcal{X}, \mathbf{B}, \sigma^2) \propto IG\left(\frac{NQ}{2}, \frac{\|\mathcal{Y} - \langle \mathcal{X}, \mathbf{B} \rangle_L\|_F^2}{2}\right) \tag{4.14}$$

若 x 服从于形状参数 a 和逆尺度参数 b 的逆伽马分布，记为 $x \sim IG(a, b)$ ，其密度函数如下所示。

$$f(x) = \frac{b^a}{\Gamma(a)} x^{-(1+a)} \exp\left(-\frac{b}{x}\right)$$

λ^2 的后验分布

如式 (4.8) 所示，与 $\boldsymbol{\lambda}$ 的有关的项仅有 $P(\boldsymbol{\lambda})$ 和 $P(\boldsymbol{\tau} | \boldsymbol{\lambda})$ ，其他无关的项可以被视为常数，暂不考虑。 λ_r^2 的先验是一个伽马分布， $\tau_r | \lambda_r^2$ 也是一个伽马分布，通过简单运算可得 λ_r^2 的后验分布也是伽马分布，具体如下式所示。

$$\begin{aligned}
 P(\lambda^2 | \boldsymbol{\tau}) &= \prod_{r=1}^R P(\lambda_r^2 | \tau_r) = \prod_{r=1}^R P(\lambda_r^2) \pi(\tau_r | \lambda_r^2) \\
 &= \prod_{r=1}^R Ga(a_r, b_r) \cdot Ga\left(1, \frac{\lambda_r^2}{2}\right) \\
 &\propto \prod_{r=1}^R \lambda_r^{2(a_r-1)} \exp(-b_r \lambda_r^2) \cdot \frac{\lambda_r^2}{2} \exp\left(-\frac{\lambda_r^2}{2} \tau_r\right) \\
 &\propto \prod_{r=1}^R Ga\left(a_r + 1, b_r + \frac{\tau_r}{2}\right)
 \end{aligned} \tag{4.15}$$

$\boldsymbol{\tau}$ 的后验分布

$\boldsymbol{\tau}$ 的条件先验服从伽马分布，而 τ_r 仅仅影响 $\mathbf{U}_{(l)}$ 的第 r 列，所以 $\boldsymbol{\tau}$ 的后验仅取决于惩罚参数 $\boldsymbol{\lambda}$ 和 CP 分解的成分矩阵 $\mathbf{U}_{\cdot r}^{(l)}$ 。经过计算，可以得出 τ_r 的后验分布服从于逆高斯分布。

$$\begin{aligned}
 P(\boldsymbol{\tau} | \boldsymbol{\lambda}, \mathbf{U}^{(l)}) &= \prod_{r=1}^R P(\tau_r | \mathbf{U}_{\cdot r}^{(l)}, \lambda_r^2) = \prod_{r=1}^R P(\tau_r | \lambda_r^2) P(\mathbf{U}_{\cdot r}^{(l)} | \tau_r) \\
 &= \prod_{r=1}^R \left(P(\tau_r | \lambda_r^2) \prod_{i=1}^{P_l} P(\mathbf{U}_{ir}^{(l)} | \tau_r) \right) \\
 &= \prod_{r=1}^R \left(Ga \left(1, \frac{\lambda_r^2}{2} \right) \prod_{i=1}^{P_l} N(0, \tau_r) \right) \tag{4.16} \\
 &\propto \prod_{r=1}^R \left(\tau_r^{-\frac{P_l}{2}} \exp \left(-\frac{\lambda_r^2}{2} \tau_r \right) \exp \left(-\frac{\sum_{i=1}^{P_l} \mathbf{U}_{ir}^{(l)}}{2\tau_r} \right) \right) \\
 &\propto \prod_{r=1}^R GIGauss \left(1 - \frac{P_l}{2}, \lambda_r^2, \sum_{i=1}^{P_l} \mathbf{U}_{ir}^{(l)} \right)
 \end{aligned}$$

若 x 服从广义逆高斯分布，记作 $x \sim GIGauss(p, a, b)$ ，其密度函数如下所示。

$$f(x) = \frac{(a/b)^{\frac{p}{2}}}{2K_p(\sqrt{ab})} x^{p-1} \exp \left(-\frac{1}{2} \left(\frac{b}{x} + ax \right) \right)$$

至此，基于拉普拉斯的贝叶斯张量回归模型的推理框架已全部阐述完毕，分别在图 4-3 贝叶斯推理框架图和在伪代码算法 5 中进行了总结。

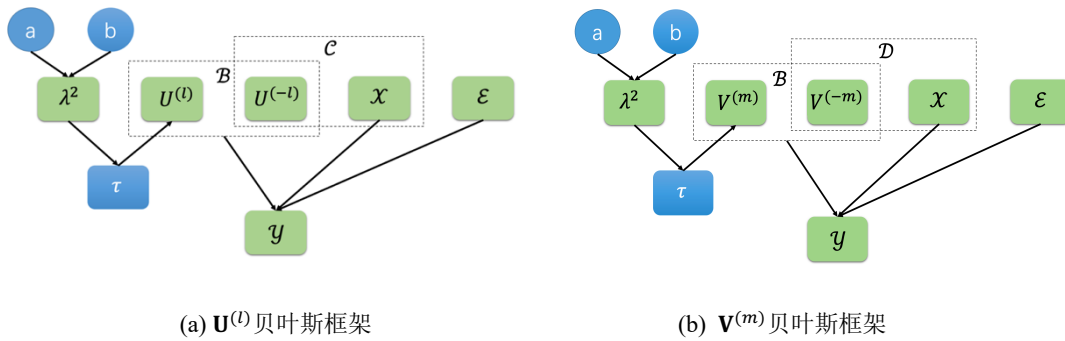


图 4-3 贝叶斯推理框架

Figure 4-3 Bayesian inference framework

算法 5 基于拉普拉斯先验的贝叶斯张量回归

输入：预测变量 \mathcal{X} , 响应变量 \mathcal{Y} , 测试集 \mathcal{X}_{new} , MCMC 采样数 T , 初始值 \mathcal{B}_0, λ_0 , 秩 R ;

输出：张量回归系数 \mathcal{B} , 预测值 MCMC 样本集 $\mathcal{Y}_{samples}$

- 1: 设置 $\mathcal{B}^{(0)} = \mathcal{B}_0, \tau^{(0)} = \lambda^{(0)}$, 其中 \mathcal{B}_0 是前述张量 Lasso 回归的解
 - 2: For $t = 1, \dots, T$:
 - 3: 从(4.14)中采样得到 $\sigma^{2(t)}$
 - 4: For $l = 1, \dots, L$:
 - 5: 从式(4.10)采样得到 $\mathbf{U}^{(l)(t)}$;
 - 6: For $r = 1, \dots, R$:
 - 7: 从式(4.15)采样得到 $\lambda_r^{2(t)}$; 从式(4.16)采样得到 $\tau_r^{(t)}$;
 - 8: End For
 - 9: End For
 - 10: For $m = 1, \dots, M$:
 - 11: 从式(4.12)采样得到 $\mathbf{V}^{(m)(t)}$;
 - 12: For $r = 1, \dots, R$:
 - 13: 从式(4.15)采样得到 $\lambda_r^{2(t)}$; 从式(4.16)采样得到 $\tau_r^{(t)}$;
 - 14: End For
 - 15: End For
 - 16: 根据采样所得的 $\mathbf{U}^{(l)(t)}$ 和 $\mathbf{V}^{(m)(t)}$ 来计算 $\mathcal{B}^{(t)}$;
 - 17: 根据新输入的新样本 \mathcal{X}_{new} 来对每个变量进行预测, 得到 $\mathcal{Y}_{samples}^{(t)}$
 - 18: End For
 - 19: Return 张量回归系数 \mathcal{B} , 预测值 MCMC 样本集 $\mathcal{Y}_{samples}$
-

4.2 模拟实验

对于贝叶斯张量回归模型的模拟实验, 本节除了从样本量、信噪比和数据维度三个方面加以考察外, 还增加对 MCMC 采样样本量 (T) 大小的比较, 数据的生成方式与前述张量正则回归相同。并使用张量正则回归的解作为 MCMC 采样中回归系数 \mathcal{B} 的初始值, 可以减少马尔科夫链的燃烧 (burn-in) 过程, 有效减少采样次数, 降低运行时间。对于模型的评价指标, 选取了样本 95%置信区间覆盖率, 其定义为落在 95%置信区间之内的真实的样本占总样本的百分比。95%置信区间覆盖长度则是指样本的 97.5%分位数与 2.5%分位数之差。

4.2.1 MCMC 采样样本量的影响

表 4-1 中展示了不同的 MCMC 采样样本量对模型性能的影响，每一个数值均由 120 次 (3 个维度×2 个信噪比水平×2 个训练样本量×10 次重复模拟) 模拟数据平均得出，具有代表意义。当 MCMC 样本量为 100 时，两个模型的平均 95% 置信区间覆盖率为 93.24%，而当 MCMC 样本量为 1000 时，平均覆盖率为 94.72%。所以可以认为当 MCMC 样本量较大时，贝叶斯张量回归模型表现得较好。分别观察两个模型的评价指标，可以发现不论是在较小的 MCMC 样本量 ($T = 100$)，还是在较大的 MCMC 样本量 ($T = 1000$) 下，基于拉普拉斯分布的张量回归模型在覆盖率指标上的表现优于基于高斯分布的张量回归模型，但覆盖长度略长于高斯张量回归模型。

表 4-1 不同 MCMC 样本量下的贝叶斯模型的性能

Table 4-1 Performance of Bayesian models with different MCMC sample sizes

MCMC 样本量	T=100		T=1000	
	覆盖率	覆盖长度	覆盖率	覆盖长度
高斯-张量回归	0.9272	1.59	0.9439	1.64
拉普拉斯-张量回归	0.9367	1.63	0.9504	1.68

4.2.2 训练集样本量的影响

表 4-2 比较了不同的训练集样本量对模型性能的影响，每一个数值都是 120 次 (3 个维度×2 个信噪比水平×2 个 MCMC 样本量×10 次重复模拟) 模拟结果的平均，当训练集为大样本 $N=200$ 时，两个模型的平均覆盖率为 94.07%，覆盖长度为 1.58，而当训练集为小样本 $N=50$ 时，平均覆盖率降低为 93.84%，覆盖长度为 1.70。基于拉普拉斯先验的贝叶斯张量回归总体优于基于高斯先验的贝叶斯张量回归模型，尤其是在训练数据集为小样本时，覆盖率达到 94.61%。

表 4-2 不同训练集样本量下的贝叶斯模型的性能

Table 4-2 Performance of Bayesian models with different training set sample sizes

训练样本量	N=200		N=50	
	覆盖率	覆盖长度	覆盖率	覆盖长度
高斯-张量回归	0.9404	1.57	0.9307	1.66
拉普拉斯-张量回归	0.9410	1.58	0.9461	1.74

4.2.3 信噪比的影响

表 4-3 展示了在不同的信噪比水平下模型的性能表现，每个数值均由 120 次 (3 个维度×2 个 MCMC 样本量×2 个训练数据集样本量×10 次重复模拟) 模拟结果平均而得。当数据集的信噪比低 ($SNR = 1$) 时，贝叶斯张量回归模型的平均覆盖率为 94.05%，平均覆盖长度为 2.83，而当数据集的信噪比高 ($SNR = 10$) 时，模型的平均覆盖率为 93.86%，平均覆盖率为 0.44。可以发现，当信噪比水平高时，95%置信区间的覆盖长度显著降低，MCMC 样本分布更加集中。

表 4-3 不同信噪比下的贝叶斯模型的性能

Table 4-3 Performance of Bayesian models at different signal-to-noise ratios

信噪比	SNR=1		SNR=10	
指标	覆盖率	覆盖长度	覆盖率	覆盖长度
高斯-张量回归	0.9399	2.79	0.9312	0.44
拉普拉斯-张量回归	0.9411	2.87	0.9460	0.44

4.2.4 数据维度的影响

表 4-4 展示了在三个不同的数据维度下贝叶斯张量回归模型的表现，每个数值均由 80 次 (2 个 MCMC 样本量×2 个训练数据集样本量×2 个信噪比水平×10 次重复模拟) 模型结果平均而得，具有代表意义。当响应变量 $\mathbf{y} \in \mathbb{R}^{N \times 10 \times 10}$ 的维度高而自变量 $\mathbf{X} \in \mathbb{R}^{N \times 5 \times 5}$ 的维度较低时，模型整体表现得最好，95%置信区间覆盖率平均达到 94.32%，覆盖长度平均为 1.60，两个指标均优于其他两个维度。而当响应变量 $\mathbf{y} \in \mathbb{R}^{N \times 5 \times 5}$ 的维度较低且自变量 $\mathbf{X} \in \mathbb{R}^{N \times 10 \times 10}$ 的维度较高时，模拟拟合效果略低于其他两个维度，95%置信区间覆盖率平均降低至 93.52%。当数据的维度介于两者之间时 $\mathbf{y} \in \mathbb{R}^{N \times 10 \times 5}$ ， $\mathbf{y} \in \mathbb{R}^{N \times 10 \times 5}$ ，拟合效果也介于两者之间。根据结果而言，响应变量的维度越高，越容易取得较好的拟合效果，这与前述的惩罚张量回归模型是一致的。

表 4-4 数据维度对模型性能的影响

Table 4-4 Performance of Bayesian models with different data dimensions

维度	$\mathbf{X} \in \mathbb{R}^{N \times 10 \times 10}$		$\mathbf{X} \in \mathbb{R}^{N \times 5 \times 5}$		$\mathbf{X} \in \mathbb{R}^{N \times 10 \times 5}$	
	$\mathbf{y} \in \mathbb{R}^{N \times 5 \times 5}$		$\mathbf{y} \in \mathbb{R}^{N \times 10 \times 10}$		$\mathbf{y} \in \mathbb{R}^{N \times 10 \times 5}$	
指标	覆盖率	覆盖长度	覆盖率	覆盖长度	覆盖率	覆盖长度
高斯-张量回归	0.9258	1.65	0.9417	1.59	0.9393	1.60
拉普拉斯-张量回归	0.9446	1.73	0.9446	1.61	0.9414	1.64

通过在模拟数据集上的对比实验可以得到以下四个结论：第一，当 MCMC 样本量越大时，越能展现出贝叶斯回归模型的性能；第二，当训练集样本量越大时，95%置信区间的覆盖长度有平均低于较小的训练样本量，拉普拉斯-张量回归对于样本量不敏感，即使是在小样本的情况也取得了较好的拟合效果；第三，数据信噪比越高，95%置信区间覆盖的长度越短，即大量的样本值都集中在一个很短的区间上，说明预测准确的把握大；第四，在响应变量维度越大时，整体贝叶斯张量回归模型表现相对较好，这与前述第三章的结论是一致的。

4.3 案例分析

接上章张量正则回归算法的案例，已经在频率学派的背景下，使用正则化技术来对回归系数进行稀疏约束，使其更稳健、泛化能力更好。在本章中，详细阐述了贝叶斯学派下的张量回归算法，它能够通过重复采样，对每个样本预测值给出后验样本分布直方图，使我们能够对每个样本的预测值的分布有更为宏观的认识，对预测的不确定性有进一步把握。

在此案例中，前章已经对五个真实的响应变量进行了预测，在本章中，将所介绍的两个贝叶斯张量回归模型应用在该数据集上，考察响应变量预测值的后验分布情况，惩罚参数和 CP 秩的初始值的选取与前章一致，并将前章的回归系数 B 的估计值结果作为贝叶斯回归的初始值进行采样。

模型的评价指标如表 4-5 所示，在 MCMC 采样样本数为 1000 的条件下，基于拉普拉斯分布的贝叶斯张量回归模型在 95%置信区间上的覆盖率和覆盖长度均优于基于高斯分布的贝叶斯张量回归模型。高斯-张量回归模型在长度为 3.62 的区间长度上涵盖了 94.78% 的真实样本，而拉普拉斯-张量回归模型在长度为 3.58 的区间长度上涵盖了 95.25% 的真实样本，即在更短的区间内集中的更多的真实样本。结论表明，本文提出的基于拉普拉斯分布的贝叶斯张量回归模型具有实际的应用价值。

表 4-5 两个贝叶斯张量回归模型在空气质量数据集上的表现

Table 4-5 Performance of the two Bayesian tensor regression models on the Air Quality dataset

指标	覆盖率	覆盖长度	惩罚参数	CP 秩	MCMC 样本数
高斯-张量回归	0.9478	3.62	10	3	1000
拉普拉斯-张量回归	0.9525	3.58	10	3	1000

为了可视化地使用贝叶斯回归模型对预测值的不确定性进行量化，本节同时绘制出回归系数基于高斯分布和基于拉普拉斯先验分布的贝叶斯张量回归的后验样本分布直方图。选取了三个样本的三个预测值，如图4-4所示，其中子图(a-c)为基于高斯分布的贝叶斯回归模型所产生的样本，子图(d-f)为基于拉普拉斯的贝叶斯张量回归模型所产生的样本，蓝色的竖直线为该样本的真实值。总体而言，直方图都较好地反映了样本的分布结果，具体来看，对于响应变量1和响应变量3的拟合效果较好，两种方法下MCMC样本都集中在真实值的附近随机波动，真实值几乎都落在了频率最高的区间。相对不够理想的是响应变量2，拉普拉斯-张量回归相比于高斯-张量回归的分布更接近真实值，预测效果更好。

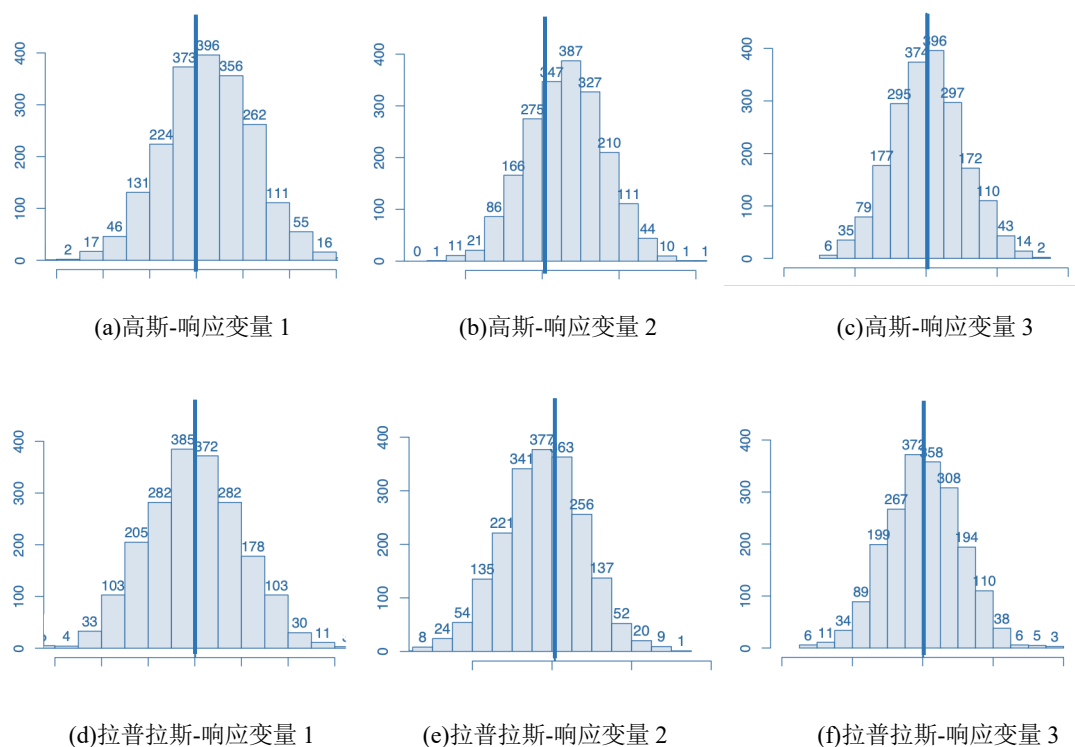


图 4-4 贝叶斯张量回归后验样本直方图

Figure 4-4 Histograms of Bayesian tensor regression posterior samples

4.4 本章小结

在本章中，介绍了两个贝叶斯张量回归模型，并在模拟数据集和真实数据集上进行了实验对比分析，可以得出以下四个结论：一是 MCMC 采样样本量不宜取得过小，采样样本量较大时才能更加合理地反映出变量的分布情况；二是信噪比水平对模型的影响很显著，当信噪比大时，95%置信区间的收敛长度显著缩短；三是张量回归模型在响应变量的维度较大时表现更好；四是在模拟数据集和真实数据集中，基于拉普拉斯分布的贝叶斯张量回归模型在性能表现上

均较基于高斯分布的贝叶斯张量回归模型有一定幅度的提升，说明本文提出的拉普拉斯-张量回归模型具备实际应用的价值。

第5章 基于张量回归算法的用户行为预测

根据上文对于张量回归模型的论述和比较,已经对模型的算法结构有了比较完整清晰的认识,在本节中,将把本文所提出的基于CP分解的张量Lasso回归和基于拉普拉斯的贝叶斯张量回归模型综合应用到一个实际的电商数据集中,论述如何使用用户的特征和商品的特性来对用户的行为进行预测分析。

数据取自飞猪平台上的用户行为数据集。飞猪是阿里巴巴集团旗下出行平台,为用户提供大规模的旅游相关产品,例如机票、车票、酒店、旅行团等等。作为中国被广泛使用的在线出行平台之一,凭借着平台上提供的多样性产品组合,沉淀了较大规模的用户在线行为数据,但是由于数据的低频特点导致的稀疏问题给常规的建模带来了挑战,故本文选用该数据集进行用户行为预测分析。

该数据集包含三张数据表,分别为用户表、物品表和用户行为表,其字段如表5-1所示,覆盖了五百万个用户和二十万个商品,共计两亿多条记录。每位用户选取了五方面的特征,分别是年龄、性别、职业、常居城市和人群标签,每个商品也涵盖了三方面的特征,分别是商品类目、商品城市和商品标签。另外由于数据进行了脱敏处理,数据的标签值并不具有实际含义。

表 5-1 飞猪平台用户商品行为数据集
Table 5-1 User behaviors on products in Fliggy

数据表	字段名称	字段说明
用户表	用户 ID	共计 5655138 位用户
	年龄	共计 109 个年龄分类
	性别	共计 3 个性别类别
	职业	共计 16 个职业类别
	常居城市	共计 356 个城市类别
	人群标签	每个用户具有 4 个标签
商品表	商品 ID	共计 273188 件商品
	商品类目	共计 87 个商品类别
	商品城市	共计 398 个城市类别
	商品标签	每个商品具有 2 个标签 (大部分缺失)
用户行为表	行为 ID	共计 224542017 条记录
	用户 ID	共计 5657890 个用户
	商品 ID	共计 273302 件商品
	行为类型	点击、收藏、加购、购买

5.1 描述性统计分析

本文首先对数据集进行初步的描述性统计分析，以便对数据本身具有更加全面和清晰的认识，为后续的数据处理和建模提供思路 and 依据。本节将分别从用户行为、用户特征和商品特征三方面进行探索分析。

5.1.1 用户行为

用户行为共有四种，分别为“点击”、“收藏”、“加购”和“购买”，其分布非常不均匀，如图 5-1 所示，用户行为主要集中在“点击”行为上，其占比达到 89%，其次是加购行为，占比为 6%，收藏和购买行为占比分别为 3% 和 2%。在用户规模为五百万、商品规模为二十万的数据前提下，去重的用户-商品-行为记录仅为 1.6 亿条，数据的密度不到 1%。

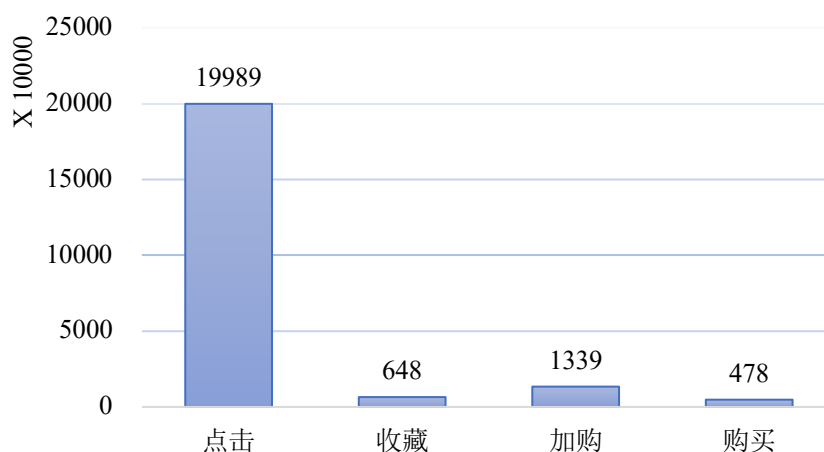


图 5-1 用户行为数在行为类型上的分布

Figure 5-1 Distribution of user behaviors by behavior type

5.1.2 用户特征

本小节从用户角度来探索用户的特征分布情况。用户特征囊括五个方面，分别为年龄、性别、职业、常居城市和人群标签，将分别对这五个方面进行探索分析。

对于用户在年龄标签上的分布如图 5-2 所示 (由于数据进行了脱敏处理，标签值并不代表实际的年龄)，可以发现用户都集中分布在部分的年龄标签上，头部 10% 的年龄标签上覆盖了 50% 以上的总用户，具有马太效应，呈现两极分布的特征。根据现实经验，有出行需求的用户基本集中在成年后和退休前，尤其集中在中青年阶段，经计算，头部 40% 的年龄标签下可以覆盖 98% 的用户数，数据分布与现实情况是相对吻合的。

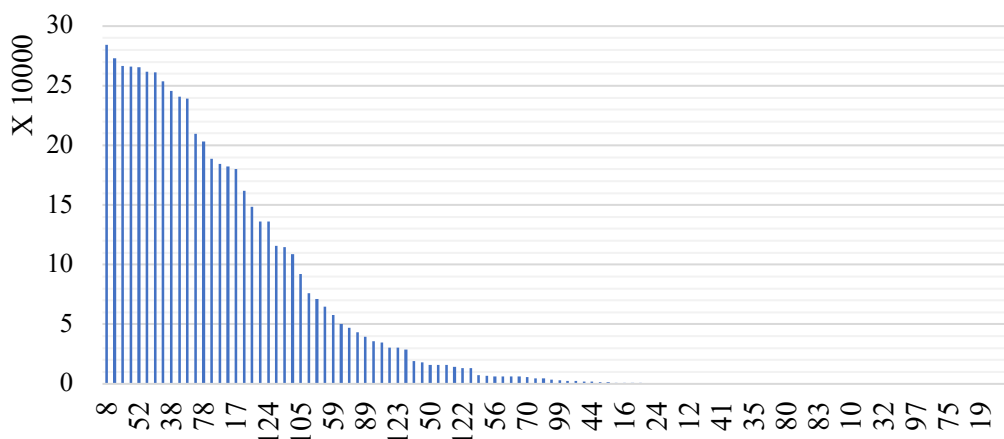


图 5-2 用户数在年龄上的分布

Figure 5-2 Distribution of users by age

用户数量在性别标签上的分布占比如图 5-3 所示，三个标签类别分别为男、女和未知，其中标签 1 的占比最大为 59%，其次是标签 2，用户占比为 38%。除标签 3 外，相对比较均匀，不存在数据集中分布在某一个类别上。后续对数据的处理上，考虑对性别变量进行哑变量编码，设置两个虚拟变量来表示三个性别特征。

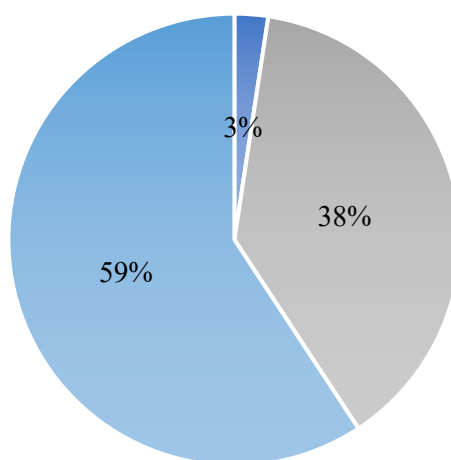


图 5-3 用户数量在性别标签下的分布

Figure 5-3 Distribution of users under the gender

用户数量在职业上的分布如图 5-4 所示，职业标签为 3、4、5 和 6 的用户数量最多，均在五十万以上，这四类职业的用户总数占到用户总数的 68%。用户职业未知的数量为 41 万，占总用户数的 11% 左右。各个标签上的用户数量都相对比较合理，在后续的数据处理中，考虑直接对用户职业变量进行哑变量编码，可以设置 15 个虚拟变量来表示 16 个职业类别 (包含未知)。

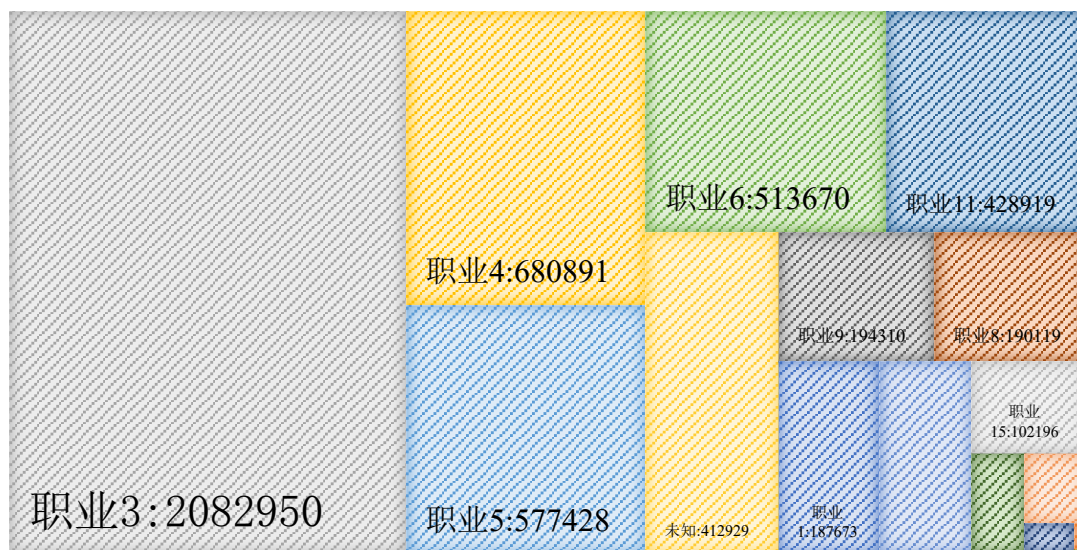


图 5-4 用户数量在职业标签下的分布

Figure 5-4 Distribution of users under the occupation

用户数量在常居城市上的分布图如图 5-5 所示，城市类别数多达 356 个，除部分十余个常居城市的用户数比较多以外，大部分常居城市的用户均在十万个以下。考虑到不同常居城市的用户可能会对商品种类有不同的偏好，本文也将常居城市作为自变量用于预测用户行为。

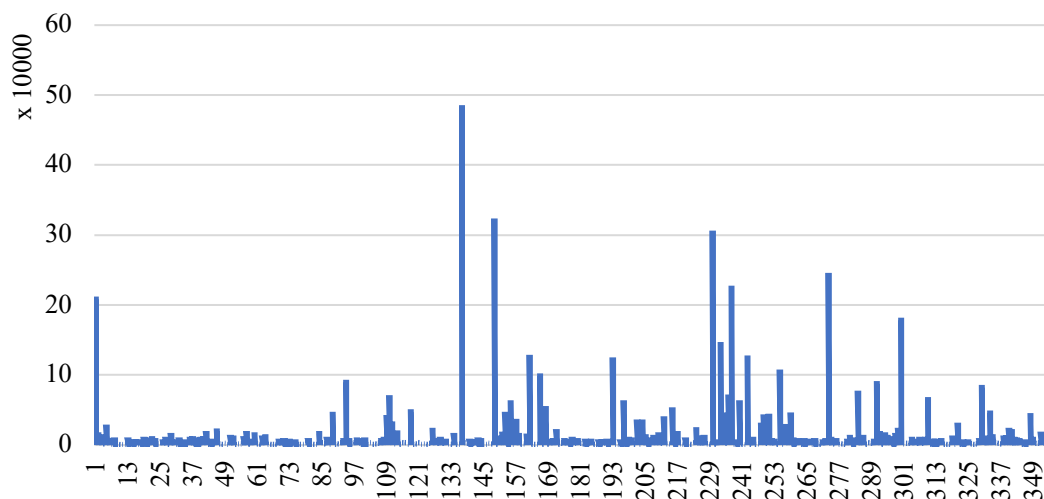


图 5-5: 用户数量在常居城市下的分布

Figure 5-5 Distribution of users by city of residence

用户人群标签共计 22 种，如图 5-6 所示，不同人群标签的用户数量有一定差距，如标签 20 和标签 15 用户数多，分别达到了 380 万和 298 万，但就总体而言，几乎每个人群类别上都有一定数量的用户。人群标签是基于某一特征或者

多种特性对用户进行聚类的结果，所以本文也将其列入自变量进行拟合预测。对于此标签变量，也采用哑变量编码的方式进行处理。

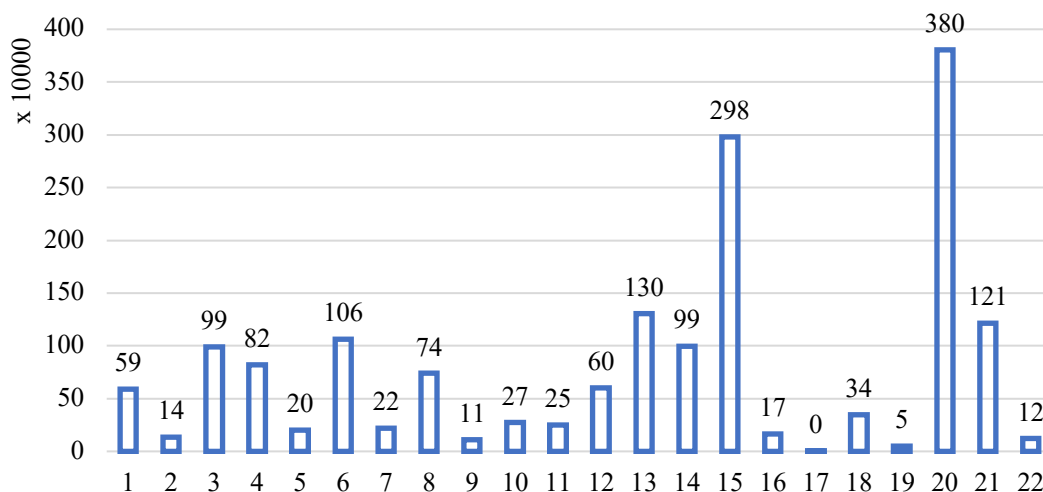


图 5-6 用户数量在人群标签下的分布

Figure 5-6 Distribution of users under the label

经过以上的可视化展示及分析，已经对用户特征的数据有了初步直观的认识，类似地，也对物品数据集的三个特征进行初步分析，以确定后续对物品数据进行预处理的方法。

5.1.3 商品特征

物品表中共含 273188 件物品，首先考察不同商品类别上的物品数量分布，如图 5-7 所示，可以发现有两类商品类型数量特别多，均达到六万件以上，有十种左右的商品类别上的物品数处于中等水平，而其余的几十种商品类型上的数量都非常少。

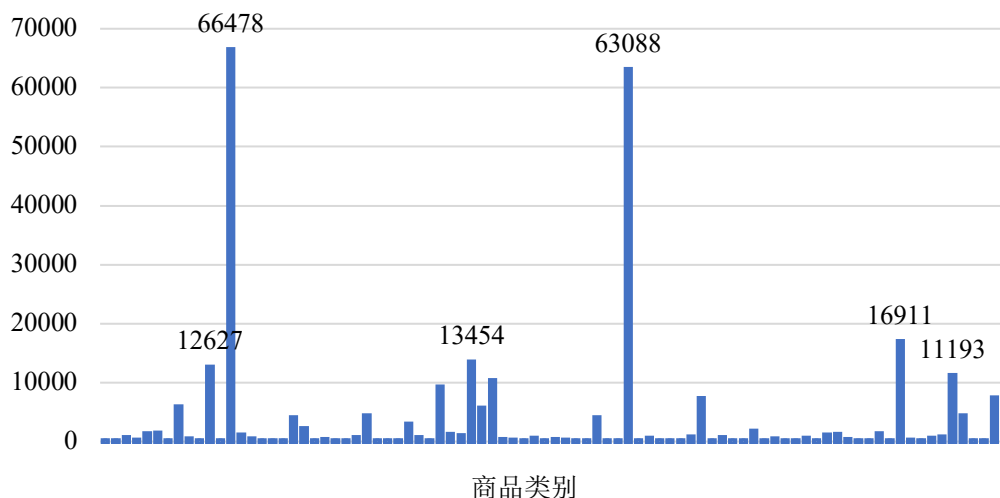


图 5-7 商品在商品类型下的分布

Figure 5-7 Distribution of commodities by commodity type

物品表中的商品遍布在 398 个城市，如图 5-8 商品数量在商品所属城市上的分布图所示，绝大部分的商品都集中在部分城市。有 9 个城市的商品数据达到了 10000 个以上，大部分城市的商品都集中在 5000 个以下。

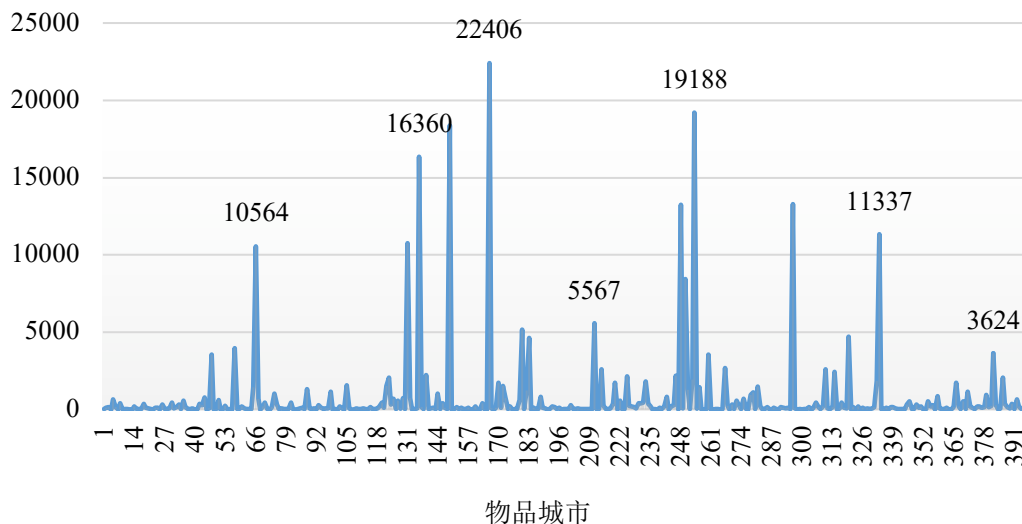


图 5-8: 商品在商品城市下的分布

Figure 5-8 Distribution of commodities under commodity cities

物品表中二十余万商品大部分商品仅有两千多条记录含有玩法标签，在未缺失标签的两千条记录中，标签种类多达 475 个，存在很多一个标签下仅含一个商品的情况。如表 5-2 所示，大部分的玩法标签下所含的物品个数在 10 个以下。若不对数据进行哑变量编码将无法正确反映标签的实际含义，若进行哑变量编码，将使得我们的模型过分复杂和冗余，综合考虑，该变量将不纳入模型的训练之中。

表 5-2 商品标签分布

Table 5-2 Distribution of commodities labels

标签下所含物品范围	标签个数
该标签下含超 1000 个物品	8
该标签下含 100-1000 个物品	54
该标签下含 10-100 个物品	181
该标签下含 10 个以下物品	232

5.2 基于张量回归模型的用户行为预测分析

在推荐系统中对用户行为进行建模预测分析，其目的是希望能够进一步地挖掘用户与商品之间的联系，服务于冷启动、精准营销等关键环节。一方面可

以从用户角度出发，把用户感兴趣的商品推送给用户，以提高用户的满意度以及用户对于平台的价值，另一方面可以从商品角度出发，把商品在合适的时间精准地推送给感兴趣的用户，以达到尽可能高的点击率和转化率等业务关键指标。根据上节描述性统计分析，已经对用户的特征、物品的特征和用户的行为数据有了初步的认识，并针对不同的特征确定了不同的处理方案，主要进行了标签重构和哑变量编码。由于原始数据集的数据量过于庞大，并且第三、四两章的对比实验中验证了本文提出的模型在相对较少的训练样本情况下具有较强的预测性能，所以在本节的建模分析中将从原始数据集中随机抽取部分行为记录进行预测分析。共抽取了 10000 余条用户行为记录，将其转换为用户-商品-行为的数据形式，共计有 230 位不同的用户和 300 件不同的商品。

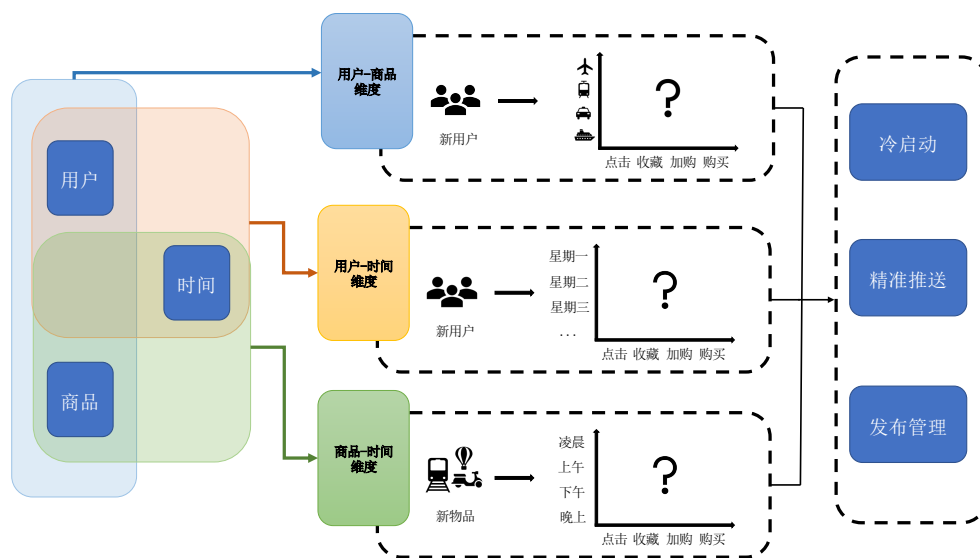


图 5-9: 用户行为预测建模框架

Figure 5-9: Modeling framework for user behavior prediction

用户与商品之间的交互行为错综复杂，可供分析的视角也很丰富，碍于篇幅，本节以三个预测分析角度为例，分别使用用户特征和物品特征对用户行为进行预测，说明张量回归模型在用户行为数据集上的应用价值。建模框架如图 5-9 所示，首先，从用户-商品角度出发，通过用户本身携带的特征信息对用户会在哪些类型上的商品发生行为和发生行为的次数进行预测，以确定用户对不同类型商品的偏好程度；其次，从用户-时间角度出发，考察用户在时间维度上的行为偏好，预测用户在一个星期中每天的行为数；最后，从商品-时间角度出发，预测商品在不同时段被点击、收藏、加购和购买的行为次数，为商品的发布规则提供更多的参考。

5.2.1 基于用户-商品维度的用户行为预测分析

本节从用户-商品的角度出发，基于用户特征对不同类别的商品进行用户行为预测分析，预测变量是进行了独热编码后的 230 位用户的年龄、性别、职业、常居城市和人群标签等特征信息，响应变量是这 230 位用户在 37 个不同种类的商品上的四种行为数据，分别为点击、收藏、加购和购买。将这 230 位用户进行划分，使用 150 位用户的特征及用户行为作为训练集进行参数学习，再输入 80 位新用户的特征作为测试集对模型进行评价，建立的回归模型如下所示。

$$Y_{user_itemtype_behave} = \langle X_{user}, B_{user_itemtype} \rangle + \epsilon$$

首先，使用基于 CP 分解的张量 Lasso 回归模型进行拟合，使用网格搜索的方法在不同的参数组合下测试模型的表现，以选取最合适的参数。如图 5-10 所示，绘制了各个参数组合下的模型在测试集上的相关系数热力图，颜色越深代表模型效果越好。在相同的惩罚力度之下，具有较高的 CP 秩的模型具有更强的拟合能力，可以发现当 CP 秩为 5 时，模型在各个惩罚水平下都取得了相对可观的拟合效果，认为真实的 CP 秩是接近于 5 的。当惩罚力度超过 10 之后，模型整体的性能受到了抑制。

R=1	0.2047	0.5522	0.0220	0.2269	0.5462	0.5443	0.3007	0.1786
R=2	0.3377	0.6355	0.3059	0.5477	0.2559	0.3524	0.2456	0.1790
R=3	0.6287	0.5521	0.5505	0.5478	0.5435	0.3327	0.2269	0.1791
R=4	0.6347	0.5489	0.6269	0.3524	0.5719	0.2882	0.5295	0.1788
R=5	0.7288	0.6313	0.5467	0.5739	0.5460	0.4397	0.5293	0.2924
R=6	0.6532	0.6325	0.5555	0.3010	0.3524	0.5442	0.3720	0.1840
R=7	0.6705	0.6384	0.6240	0.5439	0.5416	0.2047	0.5289	0.1791
R=8	0.6894	0.6381	0.5554	0.6211	0.5461	0.5442	0.2046	0.1791
R=9	0.6783	0.7203	0.2964	0.3814	0.5418	0.6166	0.2870	0.1890
R=10	0.6926	0.6313	0.6261	0.6227	0.6730	0.5442	0.4674	0.1788
	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 10$	$\lambda = 50$	$\lambda = 100$

图 5-10 模型在各个参数组合下的相关系数热力图

Figure 5-10 The correlation coefficient heat map of the model under each parameter combination

根据前述网格搜索法所选取的参数 ($R = 5, \lambda = 1$) 进行训练，在测试集上的性能表现如表 5-3 所示，预测而得的响应变量和真实的响应变量之间的相关系数为 0.7288，整体拟合的 Q 方为 51.82%，均方根误差为 0.4761。

表 5-3 CP-Lasso 回归模型在用户行为数据集上的参数和拟合结果

Table 5-3 Parameters and fitting results of CP-Lasso regression model on user behavior dataset

模型参数	CP 秩	惩罚参数	迭代次数
CP-Lasso 回归	5	1	30
评价参数	相关系数	Q 方	均方根误差
CP-Lasso 回归	0.7288	0.5182	0.4761

整体而言,从评价指标上来看,模型取得了较好的拟合效果,但这三个评价指标相对宏观和空泛,本节将借助分类任务中的常用指标进行辅助评价。将用户行为次数转换为二分类变量,用户行为次数大于0的记录为正样本,否则为负样本。准确率是指所有样本中被正确预测的比例,包含被正确预测的正样本(TP)和被正确预测的负样本(TN),反映了模型预测的准确能力;精确率是指预测为正的样本中实际为真的比例,反映了模型预测正样本的精度;召回率是指实际为真的样本中被模型预测出来的比例,反映了模型从商品集中返回正样本的能力。在本案例中,从用户的感受来看,模型的精确率具有意义,代表了推荐给用户的商品中用户满意的比,而从平台和商家的角度出发,模型的召回率具有更大的意义,即商品库中用户感兴趣的商品有多大比例被正确推送了。

$$\left\{ \begin{array}{l} \text{准确率} = \text{预测正确的样本} / \text{所有样本} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \\ \text{精确率} = \text{真实的正样本} / \text{预测的正样本} = \text{TP} / (\text{TP} + \text{FP}) \\ \text{召回率} = \text{真实的正样本} / \text{实际的正样本} = \text{TP} / (\text{TP} + \text{FN}) \\ \text{F1 得分} = (2 \times \text{精确率} \times \text{召回率}) / (\text{精确率} + \text{召回率}) \end{array} \right.$$

如表 5-4 所示,80 位新用户 37 个商品类型上的 4 种用户行为预测准确率为 98.11%,由于数据的稀疏程度较高,正样本的比例较小,所以精确率和召回率相比于准确率会低很多,但在稀疏度为 98.89%的数据中,精确率为 27.88%的同时达到了 43.94%的召回率,取得了较为理想的结果。

表 5-4 用户-商品维度用户行为预测分类评价指标

用户-商品模型	数据稀疏度	准确率	精确率	召回率	F1 得分
指标	98.89%	98.11%	27.88%	43.94%	34.12%

但指标数字带给的理解是有限的,取出前 30 位用户和前 10 个商品类别的“点击”行为预测结果来可视化的展示模型预测的效能。如图 5-11 所示,深绿色的方块代表该用户在该类商品上有点击行为且经模型预测也有点击行为(TP)。浅绿色方块代表该用户在该类商品上没有点击行为且经模型预测也没有点击行为(TN)。此两类样本均为预测正确的样本,黄色方块代表未被正确预测的样本,其中包含虚假的正样本(FP)和虚假的负样本(FN)。

在前 30 位用户和前 10 个商品类型的交互行为中,点击行为的准确率为 78%,预测的精确率为 25.76%,模型的召回率为 50%,即经过张量回归模型所推荐的商品中有约四分之一被实际点击了,而在用户所有会实际点击的商品中,有一

半被筛选出来了，总体而言，可以认为基于 CP 分解的张量 Lasso 回归模型在该任务下取得了较好的效果。

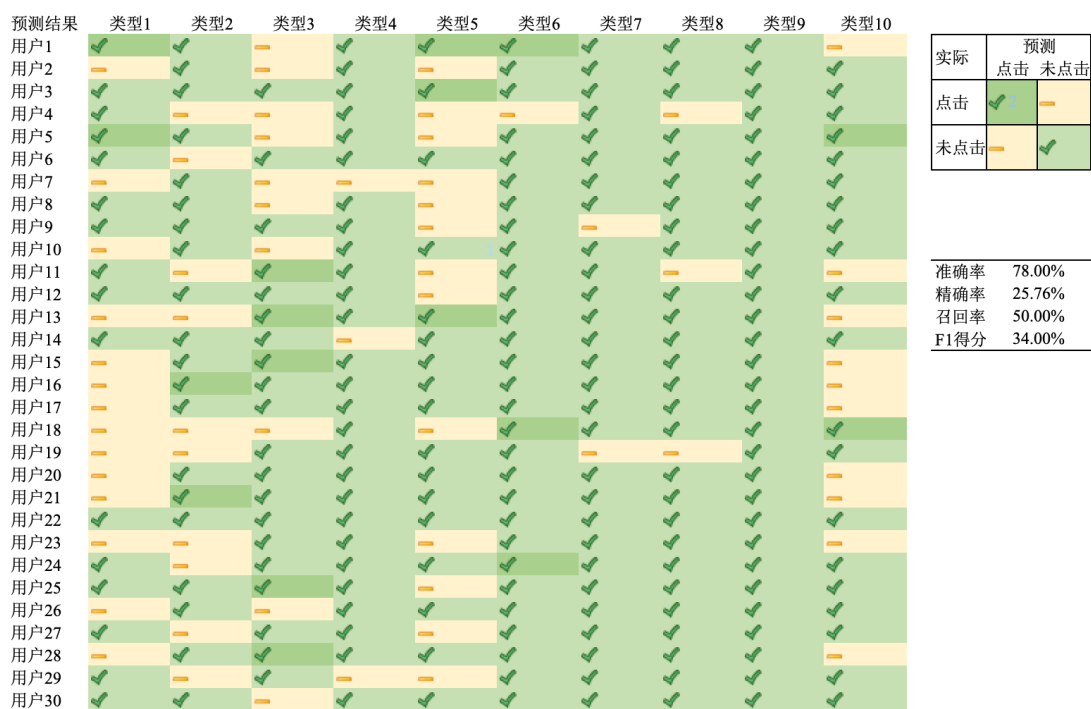


图 5-11 前 30 位用户对前 10 类商品的预测准确分析

Figure 5-11 Analysis of the prediction of the top 10 categories of commodities by the top 30 users

其次，从贝叶斯回归的角度对模型预测结果的不确定性进行量化。使用基于拉普拉斯分布的贝叶斯张量回归模型对训练数据进行拟合，CP 秩和惩罚参数的选择等同于前述的 CP-Lasso 回归模型，拟合结果如表 5-5 所示，在 MCMC 样本量为 1000 的情况下，95%置信区间的覆盖率达到了 99.05%，其覆盖长度为 2.83。覆盖率之所以高出 95%很多，主要是这样用户行为数据在商品类别下的分布非常稀疏，绝大部分都为 0，80 个用户在 37 个类别的行为数据的稀疏度达到了 98.89%。

表 5-5 拉普拉斯-张量回归模型在用户行为数据集上的参数和拟合结果

Table 5-5 Parameters and fitting results of the Laplace-tensor regression model

模型	95%置信区间 覆盖率	95%置信区间 覆盖长度	MCMC 样本量	参数
拉普拉斯-张量回归	0.9905	2.83	1000	a=1, b=1

可以通过 MCMC 后验样本的直方图来可视化地考察预测值的不确定性，深蓝色的竖直线代表样本的实际值，横轴为响应变量的取值，蓝色的柱状长度代表样本落在该区间的个数，即纵坐标代表频数。

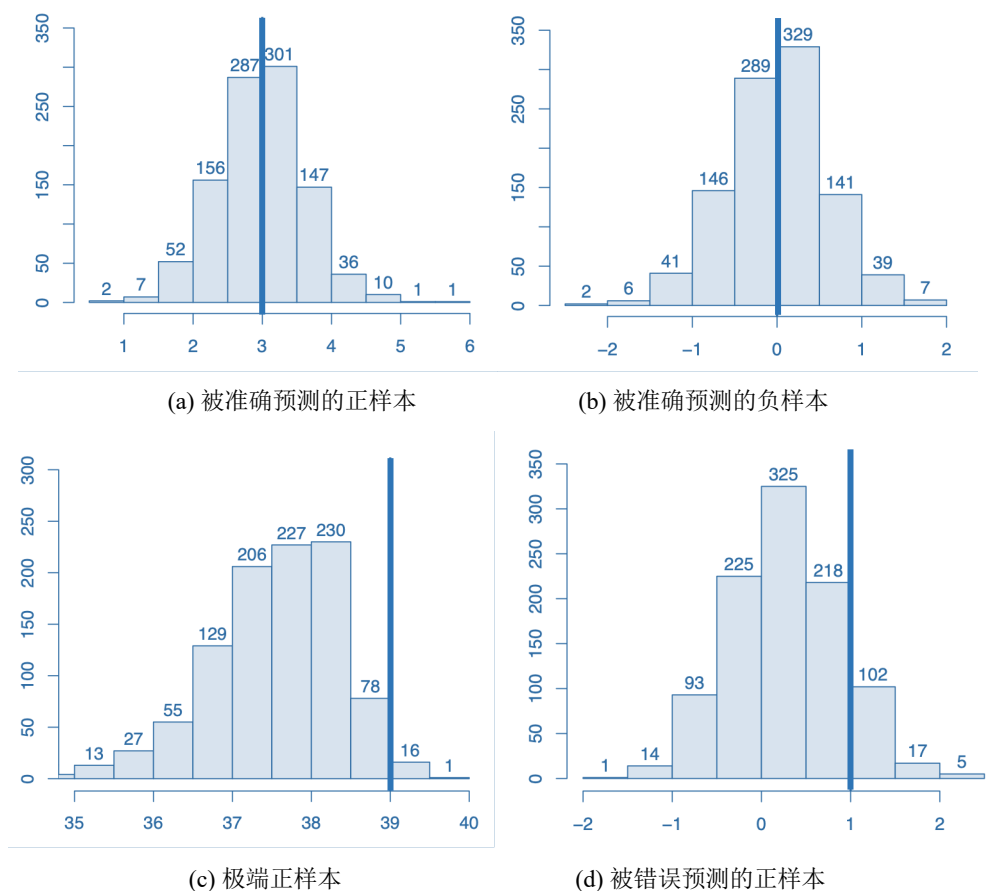


图 5-12 贝叶斯模型后验分布直方图

Figure 5-12 Histogram of the posterior distribution of the Bayesian model

分别挑选了四种情况加以说明，第一种如图 5-12 (a) 所示，为被准确预测的正样本，实际值为 3，说明用户 1 对类型 5 的商品点击了 3 次，通过本文贝叶斯回归模型得出的后验分布很好地预测了其分布特征，后验样本围绕 3 附近随机波动，虽然对于具体的点击次数无法断定其就是 3，但总体都分布在 0 的右侧，可以认为该用户对该品类商品进行点击行为的概率是非常大的。其次是图 5-12 (b)，为被正确预测的负样本，其真实情况是没有被点击，通过后验分布可以发现 MCMC 样本主要围绕在 0 附近波动，样本值集中落在 $(-0.5, 0.5)$ 之间，可以认为其进行点击行为的可能性较小。图 5-12 (c) 是极端情况下被正确预测的正样本，该用户实际对该品类的商品点击了 39 次，后验样本集中分布在 $(36, 39)$ 之间，虽然没有能够准确预测出其点击的具体次数，但其差距很小，也可以认为该模型较好地预测出了用户的点击行为。图 5-12 (d) 为被错误预测的正样本，该用户对该样本实际有 1 次点击行为，模型预测值四舍五入为 0 次，但就分布本身而言，

MCMC 样本集中分布在 $(-0.5,1)$ 之间，整体而言还是更偏向于 0 的右侧，倾向于认为可能会存在点击行为。

5.2.2 基于用户-时间维度的用户行为预测分析

除了上节中对用户在不同类别商品上的行为进行预测外，也可以综合考虑时间维度的影响。对于商家而言，用户行为上携带的时间信息是十分重要的，这可能会是商家上新产品、推广产品的关键考虑因素之一。本节将以用户在一周内每天的行为次数分布为例，建立预测模型，关于月、日、小时等时间维度也可以做类似的考察，与此模型类似，不做额外赘述。

$$Y_{user_week_behave} = \langle \mathbf{X}_{user}, \mathbf{B}_{user_week} \rangle + \varepsilon$$

首先，根据上式建立基于 CP 分解的张量 Lasso 回归模型，预测变量是 230 位用户的特征信息，响应变量是 230 位用户在一周 7 天内的行为次数。为了保证模型具有较好的泛化性能，使用 150 位用户的特征信息及行为次数作为训练集进行模型训练，使用剩下 80 位用户的特征信息及行为次数作为测试集，进行模型评价。

使用网格搜索的方法进行参数选择，如下图 5-13 所示，模型在 CP 秩 R 为 6 且惩罚参数 λ 为 1 的组合下取得了最好的效果，预测值与实际值之间的相关系数达到了 0.7079。从整体出发，CP 秩越大，越倾向于取得较好的预测效能，但当惩罚力度较大时，模型整体的预测能力受到限制，此时，CP 秩对模型的影响已经作用不大。

R=1	0.4851	0.4817	0.4795	0.4795	0.4783	0.4590	0.2792
R=2	0.5986	0.4816	0.5438	0.4794	0.4780	0.4588	0.2922
R=3	0.4851	0.5888	0.5428	0.4796	0.5220	0.4585	0.2791
R=4	0.5957	0.5496	0.4802	0.4789	0.4781	0.4588	0.2791
R=5	0.6761	0.5498	0.5737	0.4792	0.4776	0.4588	0.2913
R=6	0.7079	0.5883	0.5438	0.4790	0.4782	0.4588	0.2792
R=7	0.6772	0.5839	0.4798	0.5656	0.4784	0.4588	0.2790
R=8	0.5952	0.5855	0.6081	0.5279	0.5218	0.4589	0.2791
R=9	0.6777	0.5497	0.5436	0.5294	0.4781	0.4589	0.2791
R=10	0.6935	0.5496	0.4793	0.5280	0.5220	0.4589	0.2789
	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 10$	$\lambda = 50$

图 5-13 模型在各个参数组合下的相关系数热力图

Figure 5-13 The correlation coefficient heat map of the model under each parameter combination

模型选取了参数 CP 秩 $R = 6$ 且惩罚参数 $\lambda = 1$ 的组合，如表 5-6 模型拟合结果所示，预测值与实际值的相关系数为 0.7079，模型整体的 Q 方为 51.56%，均

方根误差为 1.0587，可以认为模型较好的拟合了数据，并在测试集上取得了有一定说服力的预测结果。

表 5-6 CP-Lasso 回归模型在用户行为数据集上的参数和拟合结果

Table 5-6 Parameters and fitting results of CP-Lasso regression model on user behavior dataset

模型参数	CP 秩	惩罚参数	迭代次数
CP-Lasso 回归	6	1	30
评价参数	相关系数	Q 方	均方根误差
CP-Lasso 回归	0.7079	0.5156	1.0587

结合分类任务的评价指标来辅助评价模型预测用户行为是否发生的准确率，如表 5-7 所示，80 个用户在一周 7 天内的 4 种行为次数的准确率为 87.90%，综合的 F1 得分为 43.42%。在数据稀疏度为 92.14%的前提下，精确率为 34.32%，召回率为 59.09%，可以认为模型的预测性能较好。实际上，用户有发生行为的样本数有 176 个，模型在 2240 条样本中预测出了其中的 104 个，超一半以上。

表 5-7 用户-时间维度用户行为预测分类评价指标

Table 5-7 Evaluation indicators of user behavior prediction under task of classification

用户-时间模型	数据稀疏度	准确率	精确率	召回率	F1 得分
指标	92.14%	87.90%	34.32%	59.09%	43.42%

为了可视化地展示模型的预测结果，选取了前 30 位用户在一周七天内的“点击”行为预测结果进行展示。如图 5-14 所示，深蓝色方块为真实的正样本 (TP)，浅蓝色方块为真实的负样本 (TF)，淡黄色方块为错误预测的样本，其中包括虚假的正样本 (FP) 和虚假的负样本 (FN)。就所展示 210 个样本总体而言，预测的准确率为 60.95%，精确率为 36.54%，召回率为 70.37%。这表明预测为正类的样本中有 36% 被实际点击了，而实际的点击行为有 70% 被预测出来了。召回率高于精确率表明模型偏向于认为用户会有点击行为，预测的点击行为要比实际的点击行为要多，这也是相对而言比较合理的，在实际的业务中，需要尽可能地挖掘和学习到用户潜在的需求，除了预测正样本和实际正样本完全匹配的完美情况外，预测的正样本能够在一定程度上略大于实际的正样本可以帮助平台和商家更好地去挖掘用户潜在的习惯。就具体样本而言，用户 20 在七天内的点击行为全部被准确预测，其次还有用户 2、用户 3 和用户 4 等七天内有六天的行为都被正确预测，表现较差的有用户 25，一周内只有两天预测正确。

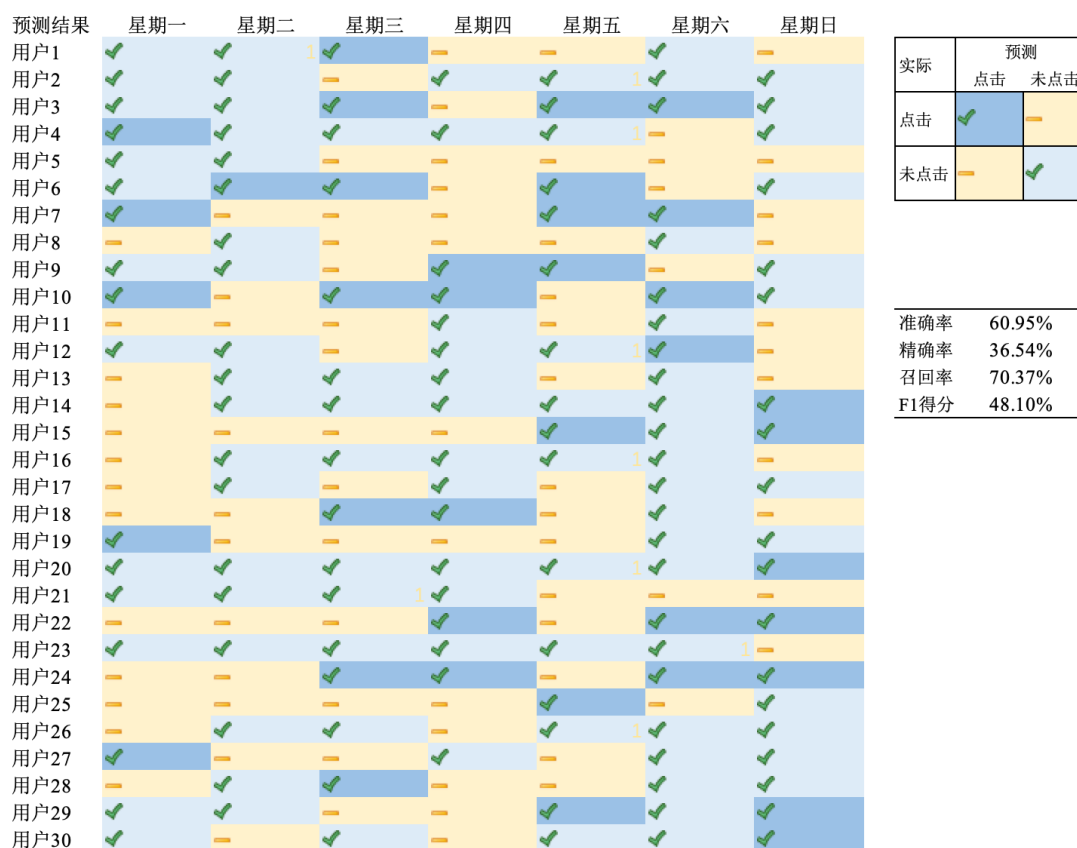


图 5-14 前 30 位用户在一周七天内的行为预测图

Figure 5-14 Analysis of the prediction of the top 30 users within seven days of a week

其次，可以通过贝叶斯张量回归模型对预测结果的不确定性进行量化。使用上述 CP-Lasso 回归模型给出的回归系数的估计作为贝叶斯张量回归模型采样的初始值，采集对每一个用户在每一个商品类别上的行为采集 1000 个贝叶斯后验样本，CP 秩和惩罚力度 λ 的初始值的选取同 CP-Lasso 回归。在该案例中，基于拉普拉斯先验的贝叶斯张量回归模型的评价指标如表 5-8 所示，在 1000 个 MCMC 样本量下 95%置信区间的覆盖率为 96.52%，覆盖长度为 3.08。

表 5-8 拉普拉斯-张量回归模型在用户行为数据集上的参数和拟合结果

Table 5-8 Parameters and fitting results of the Laplace-tensor regression model

模型	95%置信区间覆盖率	95%置信区间覆盖长度	MCMC 样本量	参数
高斯-张量回归	0.9652	3.08	1000	a=1, b=1

绘制部分样本的后验直方图，直观地考察贝叶斯模型的解释效能。如下图 5-15 所示，图 5-15 (a) 是在预测为正样本且实际也为正样本的示例，深蓝色的粗竖线为实际的点击值，其值为 1，说明该用户在该天实际点击了 1 次，蓝色的柱

状为贝叶斯回归模型产生的后验样本的直方图，直方图上的数字代表该区间内出现样本的频数，可以发现样本集中在 (0,2) 之间，占到总样本的约 65%。图 5-15 (b) 为负样本的实例，用户实际点击 0 次，即未点击，1000 个 MCMC 后验样本的分布也都集中在 0 的附近且更偏向于 0 的左侧，说明该用户在该天倾向于不点击。

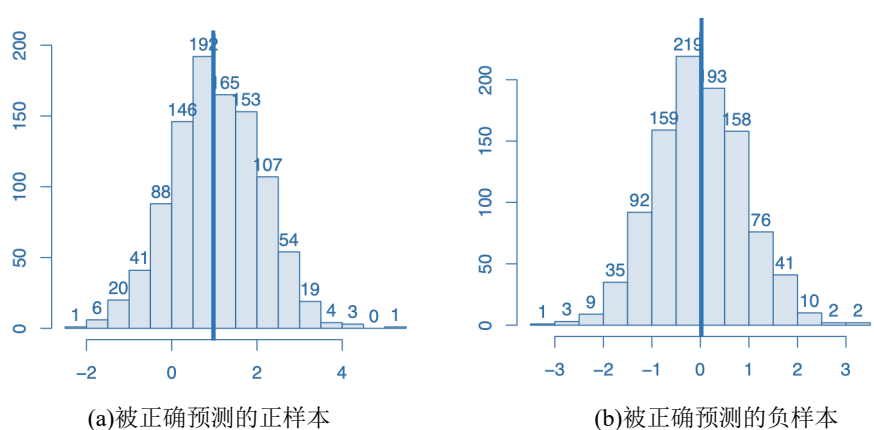


图 5-15 贝叶斯回归模型后验样本直方图

Figure 5-15 Histogram of posterior samples of Bayesian regression model

5.2.3 基于商品-时间维度的用户行为预测分析

前两小节均是基于用户的特征信息对用户行为进行预测，本节将从商品的角度出发，使用商品的特征对用户行为进行预测分析。本节将考察在不同时段中商品被点击、收藏、加购和购买的四种行为的次数。当然，通过商品特征预测不同月份、星期的行为亦或是不同职业、年龄下的用户行为均是可行的，可分析的角度很多，大体上是类似的。

首先，建立基于 CP 分解张量 Lasso 回归模型，预测变量是 300 个商品的特征，如商品类别、商品城市等，响应变量是用户在不同时间段上的行为数据。为了更加简洁并且符合现实意义，我们将 24 个小时合并成 4 个时间段，分别是凌晨 (0-7)、上午 (8-11)、下午 (12-17) 和晚上 (18-23)。响应变量被重塑为 300 个用户在 4 个时间段内的 4 种行为次数。把 300 个商品拆分成 200 个训练集和 100 个测试集，分别进行模型训练和模型评价。

$$Y_{item_hour_behave} = \langle \mathbf{X}_{item}, \mathbf{B}_{item_hour} \rangle + \varepsilon$$

使用网格搜索法对模型的参数进行选择，在十个不同 CP 秩和七个不同的惩罚力度水平下的模型在测试集上的相关系数如图 5-16 所示，取得最好拟合效果的参数组合是 $(\lambda = 1, R = 6)$ ，其相关系数达到了 0.6171。另外，可以发现在该数据集下，模型对于 CP 秩的大小不敏感，在相同的惩罚力度下，各个 CP 秩之

间的模型拟合效果相差不大，但是当惩罚力度取得过大时 (如 $\lambda = 50, \lambda = 100$)，模型的拟合效果受到限制。

R=1	0.5574	0.5563	0.5534	0.5500	0.5423	0.5378	0.5524	0.3400
R=2	0.5872	0.5907	0.5535	0.5482	0.5424	0.5376	0.5269	0.3764
R=3	0.5986	0.6034	0.5533	0.5483	0.5419	0.5378	0.5270	0.1324
R=4	0.6059	0.5589	0.5534	0.5486	0.5426	0.5607	0.5271	0.3763
R=5	0.6093	0.5983	0.5885	0.5485	0.5426	0.5371	0.5234	0.3729
R=6	0.6171	0.5968	0.5877	0.5739	0.5635	0.5376	0.5269	0.3771
R=7	0.6068	0.5896	0.5868	0.5488	0.5414	0.5369	0.5269	0.5202
R=8	0.5965	0.6121	0.5850	0.5745	0.5430	0.5380	0.1141	0.3748
R=9	0.6147	0.5894	0.5868	0.5486	0.5417	0.5375	0.5269	0.3771
R=10	0.5977	0.5896	0.5842	0.5489	0.5634	0.5371	0.5269	0.3682
	$\lambda = 1$	$\lambda = 2$	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 10$	$\lambda = 50$	$\lambda = 100$

图 5-16 基于商品特征预测在不同时段内的用户行为模型参数组合

Figure 5-16 Parameter combinations of user behavior prediction based on product features

选取 CP 秩 $R = 2$ 和惩罚参数 $\lambda = 1$ 的组合，模型的评价指标如表 5-9 所示，整体的相关系数为 0.6171，Q 方为 40.96%，均方根误差为 1.2774，虽然整体模型拟合得也相对不错，但是模型的表现不如前两个基于用户特征的预测模型，这可能有两个原因：一是物品的特征维度少，可供解释的预测变量少，无法充分地解释用户行为；二是物品本身的特征很难对用户行为起到关键解释作用。

表 5-9 CP-Lasso 回归模型在用户行为数据集上的参数和拟合结果

Table 5-9 Parameters and fitting results of the CP-Lasso regression model

模型参数	CP 秩	惩罚参数	迭代次数
CP-Lasso 回归	6	1	30
评价参数	相关系数	Q 方	均方根误差
CP-Lasso 回归	0.6171	0.4096	1.2774

结合分类任务评价指标辅助评价模型表现，如表 5-10 所示，数据整体的稀疏度为 89.38%，预测准确率为 87.38%，预测出的正样本中精确率为 43.55%，实际的所有正样本中召回率为 63.53%，综合的 F1 得分为 51.67%。可以发现，虽然在具体的预测值上，基于商品特征的模型表现不如基于用户的模型，但是在二分类任务中，在四个评价指标上均高于前两个模型，这可能是由于整体的稀疏度较低，实际的正样本比较多，可以获得较好的二分类表现。

表 5-10 用户-时间维度用户行为预测分类评价指标

Table 5-10 Evaluation indicators of user behavior prediction under task of classification

商品-时间模型	数据稀疏度	准确率	精确率	召回率	F1 得分
指标	89.38%	87.38%	43.55%	63.53%	51.67%

将前 30 位用户在不同时段上的“点击”行为预测结果可视化地展示出来，如图 5-17 所示，其中深绿色为被正确预测的正样本 (TP)，浅绿色为被正确预测的负样本 (TN)，这两种代表被正确预测的样本，用绿色的打钩符号表示。用黄色横杠表示的是被错误预测的样本，其中包括虚假的正样本 (FP) 和虚假的负样本 (FN)。对于这 30 位用户点击行为预测的准确率为 52.50%，精确率为 38.89%，召回率为 68.29%，综合的 F1 得分为 49.56%。宏观而言，用户偏向于在下午和晚上的时间段进行浏览点击。具体而言，其中商品 3、商品 9 和商品 15 等预测得较好，四个时段的行为都被准确预测了，而如商品 6 和商品 18 预测得较差，四个时段中仅有一个被预测正确了。

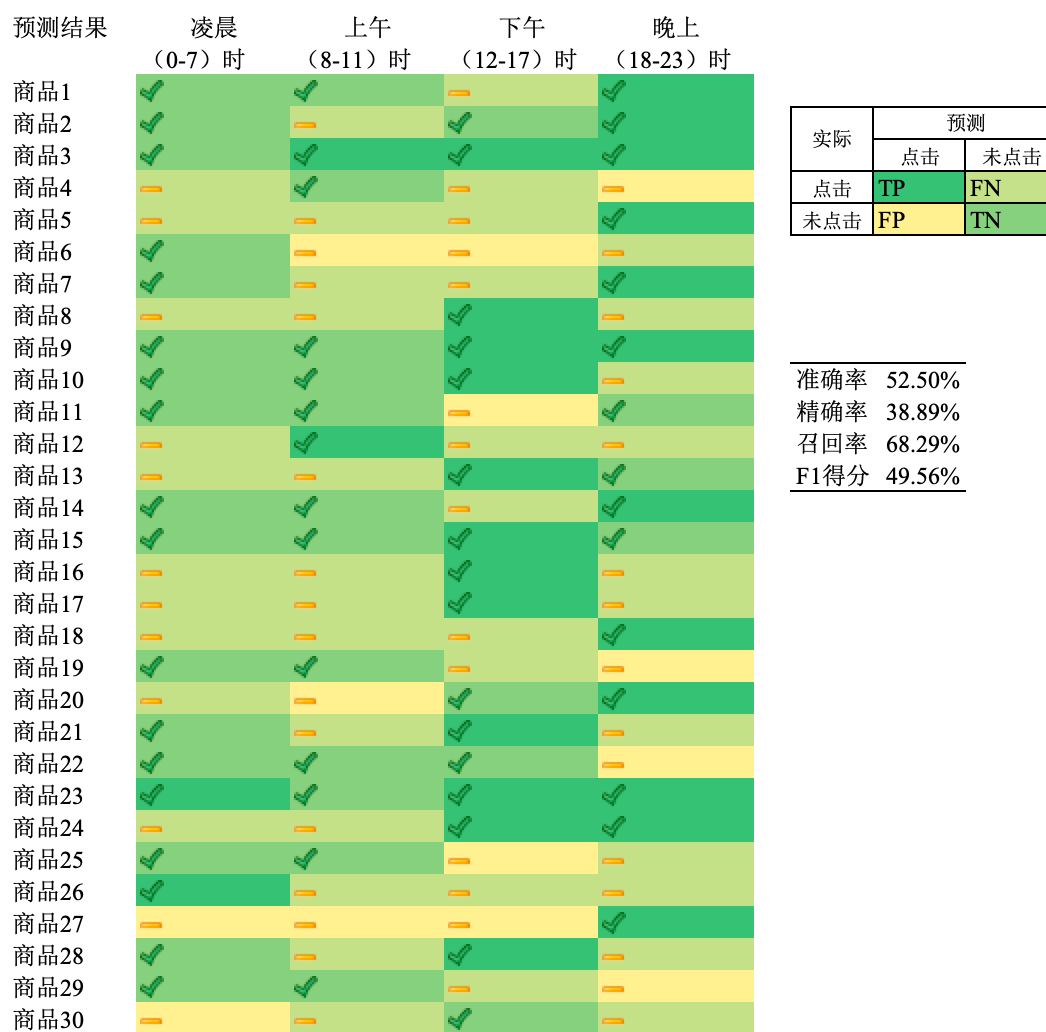


图 5-17 前 30 个商品在不同时段内的行为预测图

Figure 5-17 Prediction of the behavior of the top 30 commodities in different time periods

其次，从贝叶斯角度对所预测的用户行为进行不确定性量化，在上述分析的基础上建立基于拉普拉斯分布的贝叶斯张量回归模型，其中 CP 秩与惩罚力度

的初始值的选择同 CP-Lasso 回归。模型的评价指标如表 5-11 所示，在 1000 个 MCMC 后验样本下，95%置信区间的覆盖率为 95.69%，覆盖长度为 3.41。

表 5-11 拉普拉斯-张量回归模型在用户行为数据集上的参数和拟合结果

Table 5-11 Parameters and fitting results of the Laplace-tensor regression model

模型	95%置信区间 覆盖率	95%置信区间 覆盖长度	MCMC 样本量	参数
拉普拉斯-张量回归	0.9569	3.41	1000	a=1, b=1

绘制预测值的后验样本直方图，将不确定性量化的结果可视化地展示出来。如图 5-18 所示，图 5-18 (a) 为被正确预测的负样本，1000 个后验样本中有一半以上分布在 0 的附近，偏向于认为该商品不会被点击；图 5-18 (b) 和 (c) 均为被正确预测的正样本，其中图 5-18 (b) 的后验样本集中分布在 (0-1) 之间，倾向于认为该商品会被点击一次，而实际也被点击一次，5-18 (c) 虽然也是被正确预测的正样本，但预测的点击次数与实际存在点击偏差，根据后验直方图，倾向于认为该商品在该时段会被点击两次，而实际情况是被点击了三次，尽管有略微差距，但总体而言模型的拟合较为理想，能反映出样本的不确定性程度。

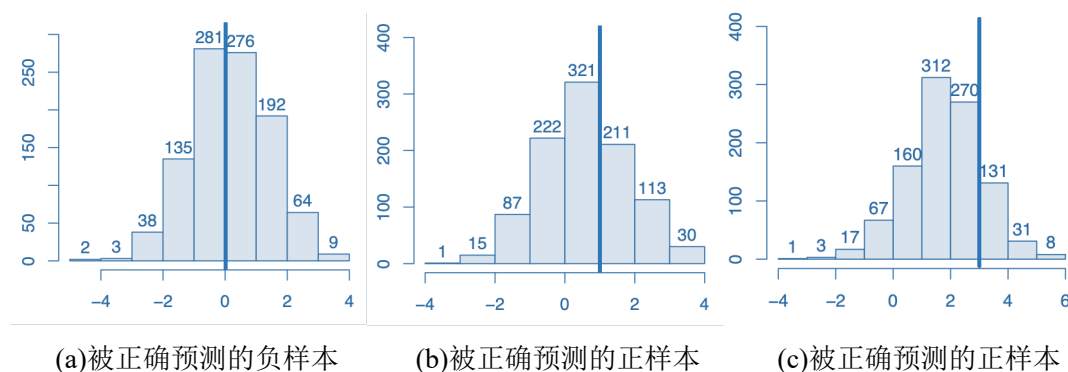


图 5-18 贝叶斯回归模型后验样本直方图

Figure 5-18 Histogram of posterior samples of Bayesian regression model

5.3 本章小结

在本章，使用了阿里天池公布的飞猪平台的用户行为数据集进行预测分析。首先对数据进行了描述性统计分析，对数据有一个初步认识和理解并为后续数据清洗和预处理做准备，其次使用了本文提出的基于 CP 分解的 Lasso 回归分别从三个角度使用前述模型对用户行为进行预测，并在贝叶斯统计的视角下对预测值进行了不确定性量化，均取得了较为理想的预测效果。

具体三个模型的建模思路及结论如下：

首先,从用户-商品的维度出发,建立了基于用户特征对用户行为的预测模型。预测值与实际值的相关系数为 0.7288,在稀疏度为 98.89%的数据下,整体预测的准确率达到 98.11%,召回率 43.94%,并可视化地展示了前三十位用户在前十类商品点击行为上的预测结果,可以发现大部分类别上的商品的用户行为可以很好地被所建立的模型预测出来,尤其是在类型 6、类型 7 和类型 8 上。此外,建立了贝叶斯张量回归模型,在长度为 2.83 的区间上覆盖了 99.05%的样本,绘制后验样本直方图,直观地考察了预测值的不确定性。结论表明张量回归模型具有较强的用户行为次数的预测能力。

其次,从用户-时间的维度出发,基于用户特征来对用户在一周七天内行为次数进行预测分析。数据稀疏度为 92.14%,预测值与实际值的相关系数为 0.7079,预测准确率为 87.90%,召回率为 59.09%,在建立的贝叶斯张量回归模型中,长度为 3.08 的区间上覆盖了 96.52%的样本。结论表明模型具备较强地定位出用户可能会发生行为的时间的能力。

最后,从商品-时间的维度出发,使用商品特征对分时段的用户行为进行预测。数据稀疏度为 89.38%,预测的准确率为 87.38%,精确率和召回率分别为 43.55%和 63.53%,并可视化地展示了前三十个商品的预测结果。在建立的贝叶斯张量回归模型中,长度为 3.41 的区间上覆盖了 95.65%的样本,结论表明商品特征与用户发生行为的时段之间存在一定程度的联系。

总体而言,本章以三个模型为例,论证了基于 CP 分解的张量 Lasso 回归模型在用户行为数据集上展现出的较强的行为次数预测能力与基于拉普拉斯的贝叶斯张量回归模型表现出的不确定性量化能力,表明本文提出的模型具有实际应用价值。

第6章 总结与展望

6.1 总结

在高维数据日益普遍的时代背景下, 本文从频率学派和贝叶斯学派的角度分别研究了预测变量和响应变量均为任意阶张量的高维数据回归模型, 并对高维数据进行了预测分析。本文的主要研究内容及结论如下:

(1) 梳理了在频率学派下两个已有的张量回归模型, 分别为基于 CP 分解的张量岭回归和基于 TT 分解的张量岭回归, 并在此基础上提出了基于 CP 分解的张量 Lasso 回归模型, 其通过对回归系数施加 l_1 惩罚项, 使其解更加稀疏和稳健, 可以有效解决维数灾难、结构信息丢失和过度拟合等问题。在随机生成的模拟数据集和真实的空气质量数据集的对比分析中, 本文提出的 CP-Lasso 回归在相关系数、Q 方和均方根误差等指标上均较两个岭回归模型有一定幅度的提升。

(2) 在贝叶斯推理框架下, 阐述了基于高斯先验的贝叶斯张量回归模型, 并结合本文对稀疏数据的研究, 建立了基于回归系数具有拉普拉斯先验的贝叶斯张量回归模型, 通过对以回归系数为主的各个参数进行合理的先验分布假设实现对参数的自动选择, 有效地降低人为调参的困难。在数据集上论证了模型的表现, 尤其是在真实的空气质量数据集上, 拉普拉斯-贝叶斯张量回归在更短的区间内达到了更高的覆盖率。

(3) 将本文所提出的回归模型应用到飞猪平台用户数据的预测分析当中, 从用户、物品、时间三个维度综合考虑, 建立了基于 CP 分解的张量 Lasso 回归模型对用户行为进行预测分析, 在稀疏度很高的数据上, 取得了较好的预测效果, 最高的预测准确率达到 98.89%。并利用贝叶斯张量回归模型对用户行为次数的预测值进行不确定性量化分析, 模型采样所得的后验样本很好地反映了变量的分布情况。

6.2 展望

本文对于预测变量和响应变量均为任意维度张量数据的回归模型进行了梳理研究, 分别在频率学派和贝叶斯学派下进行预测分析和不确定性量化, 同时在模拟数据集和真实数据集上进行了对比分析, 并在用户行为数据集上进行预测分析, 取得了较好的效果。但是, 仍然还有可以进一步研究的方向:

(1) 在模型方面, 虽然本文提出的张量 Lasso 回归模型在拟合优度和预测精度上取得了一定程度的提升, 但在运行时间上平均高于两个岭回归模型, 后续可以对算法结构进行优化, 以减少运行时间的差距。

(2) 在第五章的案例分析中, 为了利用本文提出的模型在相对较小样本下预测的优势, 只选取了 10000 余条行为记录进行了预测分析, 共计 230 位用户和 300 个商品, 后续可以尝试在更大规模的数据上进行预测分析, 或许能取得更好的预测结果。

参考文献

- [1] Cao X, Yang F, Zheng J, et al. Aberrant Structure MRI in Parkinson's Disease and Comorbidity with Depression Based on Multinomial Tensor Regression Analysis[J]. *Journal of Personalized Medicine*, 2022, 12(1): 89.
- [2] Li C, Zhang H. Tensor quantile regression with application to association between neuroimages and human intelligence[J]. *The Annals of Applied Statistics*, 2021, 15(3): 1455-1477.
- [3] Poythress J C, Ahn J, Park C. Low-rank, Orthogonally Decomposable Tensor Regression with Application to Visual Stimulus Decoding of fMRI Data[J]. *Journal of Computational and Graphical Statistics*, 2021: 1-14.
- [4] Liu J, Wu Z, Xiao L, et al. Generalized tensor regression for hyperspectral image classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 58(2): 1244-1258.
- [5] 李美琪. 基于张量学习的高光谱变化检测算法研究[D]. 西安电子科技大学, 2021. DOI:10.27389/d.cnki.gxadu.2021.001523.
- [6] 胡旷伋. 基于张量学习的判别式相关滤波目标跟踪算法研究[D]. 南京邮电大学, 2021. DOI:10.27251/d.cnki.gnjdc.2021.001053.
- [7] 刘娇. 人脸识别的增量稀疏张量回归算法研究[D]. 华南理工大学, 2016.
- [8] He K, Peng Y, Liu S, et al. Regularized negative label relaxation least squares regression for face recognition[J]. *Neural Processing Letters*, 2020, 51(3): 2629-2647.
- [9] Li Q, Li X, Wu B, et al. Retweeting prediction based on social hotspots and dynamic tensor decomposition[J]. *IEICE Transactions on Information and Systems*, 2018, 101(5): 1380-1392.
- [10] 李也. 张量聚类 and 回归建模及其在消费行为分析上的应用研究[D]. 上海交通大学, 2020. DOI:10.27307/d.cnki.gsytu.2020.002768.
- [11] 王丹. 复杂关联数据的张量模型与应用研究[D]. 厦门大学, 2017.
- [12] Chen B, Guan J, Li Z. Unsupervised Feature Selection via Graph Regularized Nonnegative CP Decomposition[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2022 (01): 1-1.
- [13] Yin M, Phan H, Zang X, et al. Batude: Budget-aware neural network compression based on tucker decomposition[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. 2022.
- [14] Gopalan G, Wikle C K. A Higher-Order Singular Value Decomposition Tensor Emulator for Spatiotemporal Simulators[J]. *Journal of Agricultural, Biological and Environmental Statistics*, 2022, 27(1): 22-45.

- [15] Li C, Sun Z. Evolutionary topology search for tensor network decomposition[C]//International Conference on Machine Learning. PMLR, 2020: 5947-5957.
- [16] Ou-Yang L, Zhang X F, Yan H. Sparse regularized low-rank tensor regression with applications in genomic data analysis[J]. *Pattern Recognition*, 2020, 107: 107516.
- [17] Hitchcock F L. The expression of a tensor or a polyadic as a sum of products[J]. *Journal of Mathematics and Physics*, 1927, 6(1-4): 164-189.
- [18] Tucker L R. Implications of factor analysis of three-way matrices for measurement of change[J]. *Problems in measuring change*, 1963, 15(122-137): 3.
- [19] 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.
- [20] Harshman R A. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis[J]. 1970, 16:1-84.
- [21] Kiers H A L. Towards a standardized notation and terminology in multiway analysis[J]. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 2000, 14(3): 105-122.
- [22] Mocks J. Topographic components model for event-related potentials and some biophysical considerations[J]. *IEEE transactions on biomedical engineering*, 1988, 35(6): 482-484.
- [23] Oseledets I V. Tensor-train decomposition[J]. *SIAM Journal on Scientific Computing*, 2011, 33(5): 2295-2317.
- [24] Perez-Garcia D, Verstraete F, Wolf M M, et al. Matrix product state representations[J]. *Quantum Information & Computation*, 2007, 7(5): 401-430.
- [25] Zhao Q, Zhou G, Xie S, et al. Tensor ring decomposition[J]. arXiv preprint arXiv:1606.05535, 2016.
- [26] Zhao Q, Sugiyama M, Yuan L, et al. Learning efficient tensor representations with ring-structured networks[C]//ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2019: 8608-8612.
- [27] Zhou H, Li L, Zhu H. Tensor regression with applications in neuroimaging data analysis[J]. *Journal of the American Statistical Association*, 2013, 108(502): 540-552.
- [28] Li X, Xu D, Zhou H, et al. Tucker tensor regression and neuroimaging analysis[J]. *Statistics in Biosciences*, 2018, 10(3): 520-545.
- [29] Zhang J, Liu Y, Jiang J. Tensor learning and automated rank selection for regression-based video classification[J]. *Multimedia Tools and Applications*, 2018, 77(22): 29213-29230.
- [30] Kossaiji J, Lipton Z C, Kolbeinsson A, et al. Tensor Regression Networks[J]. *Journal of Machine Learning Research*, 2020, 21(123): 1-21.

- [31] Lock E F. Tensor-on-tensor regression[J]. *Journal of Computational and Graphical Statistics*, 2018, 27(3): 638-647.
- [32] Gahrooei M R, Yan H, Paynabar K, et al. Multiple tensor-on-tensor regression: An approach for modeling processes with heterogeneous sources of data[J]. *Technometrics*, 2021, 63(2): 147-159.
- [33] Liu Y, Liu J, Zhu C. Low-rank tensor train coefficient array estimation for tensor-on-tensor regression[J]. *IEEE transactions on neural networks and learning systems*, 2020, 31(12): 5402-5411.
- [34] Liu J, Zhu C, Long Z, et al. Low-rank tensor ring learning for multi-linear regression[J]. *Pattern Recognition*, 2021, 113: 107753.
- [35] Hoerl A E, Kennard R W. Ridge regression: applications to nonorthogonal problems[J]. *Technometrics*, 1970, 12(1): 69-82.
- [36] Hans C. Bayesian lasso regression[J]. *Biometrika*, 2009, 96(4): 835-845.
- [37] Yang Y, Yang Y. Hybrid prediction method for wind speed combining ensemble empirical mode decomposition and bayesian ridge regression[J]. *IEEE Access*, 2020, 8: 71206-71218.
- [38] Yamaguchi Y, Yoshida S, Misumi T, et al. Multiple imputation for longitudinal data using Bayesian lasso imputation model[J]. *Statistics in Medicine*, 2022, 41(6): 1042-1058.
- [39] Ge T, Chen C Y, Ni Y, et al. Polygenic prediction via Bayesian regression and continuous shrinkage priors[J]. *Nature communications*, 2019, 10(1): 1-10.
- [40] Guhaniyogi R, Spencer D. Bayesian tensor response regression with an application to brain activation studies[J]. *Bayesian Analysis*, 2021, 16(4): 1221-1249.
- [41] Hoff P D. Multilinear tensor regression for longitudinal relational data[J]. *The annals of applied statistics*, 2015, 9(3): 1169.
- [42] Oseledets I V. Tensor-train decomposition[J]. *SIAM Journal on Scientific Computing*, 2011, 33(5): 2295-2317.
- [43] Mnih A, Salakhutdinov R R. Probabilistic matrix factorization[J]. *Advances in neural information processing systems*, 2008, 20: 1257-1264.
- [44] Zhu Y. An augmented ADMM algorithm with application to the generalized lasso problem[J]. *Journal of Computational and Graphical Statistics*, 2017, 26(1): 195-204.
- [45] De Vito S, Massera E, Piga M, et al. On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario[J]. *Sensors and Actuators B: Chemical*, 2008, 129(2): 750-757.

- [46] De Vito S, Piga M, Martinotto L, et al. CO, NO₂ and NO_x urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization[J]. Sensors and Actuators B: Chemical, 2009, 143(1): 182-191.
- [47] Eltoft T, Kim T, Lee T W. On the multivariate Laplace distribution[J]. IEEE Signal Processing Letters, 2006, 13(5): 300-303.
- [48] Wainwright M J, Simoncelli E. Scale mixtures of Gaussians and the statistics of natural images[J]. Advances in neural information processing systems, 1999, 12: 855-861.
- [49] Fernandez C, Osiewalski J, Steel M F J. Modeling and inference with ν -spherical distributions[J]. Journal of the American Statistical Association, 1995, 90(432): 1331-1340.
- [50] Park T, Casella G. The bayesian lasso[J]. Journal of the American Statistical Association, 2008, 103(482): 681-686.

攻读硕士学位期间所做的工作

科研项目：

- [1] 西南大学基本科研业务费专项资金重点项目，高维数据稀疏编码网络及应用（编号：XDJK2020B033），2020.01-2022.12.
- [2] 重庆市北碚区统计局，第四次全国经济普查研究重点课题，北碚区文旅与大健康服务业发展策略研究（编号：No.F2020198），2020.08-2020.11.

致谢

逝者如斯夫，不舍昼夜。很感谢在西大这两年的时光，感谢遇见的人和事，在各方的支持与帮助下完成了这篇学位论文，在此，我希望能诚挚地表达这份谢意：

首先，感谢我的指导老师袁老师平日里的帮助、督促和关心，亦师亦友。在困顿时为我们解惑，在懒散时督促我们用功，在沮丧时给予我们宽慰，课余时间带我们爬山运动，带我们品尝美食，时常告诉我们既要用心科研，也要热爱生活，得师如此，夫复何求；感谢我的同门和室友，一起学习、一起生活、一起努力，我们探讨学术，彼此宽慰，很开心我们并肩走了这一段路，望友谊长存。

其次，感谢学校耗费巨资购置的学术资源让我能够见识到广阔的天地，学习到先进的知识和技能。科研论文库让我窥见知识的无穷，大量的电子图书使我可以便捷地学习、快速地成长，同时也感谢学校学院的工作和所有任课老师的传道受业！

再次，感谢父母给予我的无条件的爱和包容，从不吝啬赞美与鼓励，让我可以勇敢且没有后顾之忧地去追逐我的理想。

最后，感谢我自己对统计的热爱和坚持，感谢弘文图书馆日日夜夜的陪伴。

去日不可追，来日尤可期，怀着感激的心情告别，怀着紧张的心情迎接。我相信所积累的知识可以让我在专业领域有所贡献，我相信所具备面对困难的能力可以让我一次又一次跌倒又站起来。

落笔之际，惟愿早日战胜新冠，愿我国繁荣富强。