

一种噪声容错弱监督矩阵补全的生存分析方法

陈蕾^{1,2,3} 邵楷¹ 林腾涛¹ 陈兴国^{1,2}

摘要 生存分析旨在预测某个感兴趣事件发生前的延续等待时间,已广泛应用于临床治疗中患者的生存状态分析。然而,受限于研究代价高昂和环境因素的影响,现有的生存分析方法不可避免地面临着高维小样本挑战以及复杂环境所引起的噪声敏感等问题。为了克服上述缺陷,本文提出一类噪声容错弱监督直推式矩阵补全(Weakly supervised transductive matrix completion, WSTMC)生存分析方法。该方法首先将生存分析问题建模为多任务直推式矩阵补全模型,然后引入高斯混合分布拟合真实数据中的复杂噪声以减轻模型的噪声敏感性,同时设计了一类多任务直推式特征选择机制来缓解高维小样本所带来的过拟合缺陷。此外,设计了一类有效的拟期望最大化优化算法用于求解所提出的 WSTMC 模型。最后,5 个微阵列基因表达数据集上的实验结果证实了所提出的 WSTMC 模型优于当前广泛使用的 18 种生存分析方法。

关键词 生存分析,弱监督学习,多任务学习,噪声容错,矩阵补全

引用格式 陈蕾,邵楷,林腾涛,陈兴国.一种噪声容错弱监督矩阵补全的生存分析方法.自动化学报,2021,47(12):2801-2814

DOI 10.16383/j.aas.c190740

Noise-tolerant Weakly Supervised Matrix Completion for Survival Analysis

CHEN Lei^{1,2,3} SHAO Kai¹ LIN Teng-Tao¹ CHEN Xing-Guo^{1,2}

Abstract Survival analysis aims to predict the time of an event of interest, which has been widely applied in survival status prediction of patients in clinical treatment. However, limited to the high cost of research and the influence of environments, the existing survival analysis methods inevitably suffer the over-fitting defects caused by high dimensional small-sample-size and the noise sensitivity caused by complex environments. To address these challenges, we propose a novel noise-tolerant weakly supervised transductive matrix completion (WSTMC) model to predict survival statuses of the censored instances and new instances. Specifically, we first formulate the original survival analysis problem as a multitask matrix completion model. Then the MoG (mixture of Gaussians distribution) model is employed to fit the unknown complex noise, and thus alleviate the noise sensitivity. Meanwhile, we design a novel multitask transductive feature selection mechanism to adaptively select the sharing discriminant features across tasks. Furthermore, an efficient expectation-maximization-like optimization algorithm is designed to solve the proposed WSTMC model. Finally, the experimental results conducted on 5 real microarray gene expression datasets verify that our proposed WSTMC model outperforms 18 widely-used competing methods.

Key words Survival analysis, weakly supervised learning, multi-task learning, noise-tolerant, matrix completion

Citation Chen Lei, Shao Kai, Lin Teng-Tao, Chen Xing-Guo. Noise-tolerant weakly supervised matrix completion for survival analysis. *Acta Automatica Sinica*, 2021, 47(12): 2801-2814

收稿日期 2019-10-26 录用日期 2020-03-11

Manuscript received October 26, 2019; accepted March 11, 2020

国家自然科学基金(61872190, 61772285, 61572263),南京航空航天大学模式分析与机器智能工业和信息化部重点实验室开放基金(TK219016)资助

Supported by National Natural Science Foundation of China (61872190, 61772285, 61572263) and Open Fund of Ministry of Industry and Information Technology (MIIT) Key Laboratory of Pattern Analysis and Machine Intelligence of Nanjing University of Aeronautics and Astronautics (NUAA) (TK219016)

本文责任编辑 吴立刚

Recommended by Associate Editor WU Li-Gang

1. 南京邮电大学计算机学院 南京 210023 2. 南京航空航天大学模式分析与机器智能工业和信息化部重点实验室 南京 211106
3. 江苏省大数据安全与智能处理重点实验室 南京 210023

1. School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023 2. MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing 211106 3. Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing 210023

作为现代统计学的一个重要分支,生存分析旨在建模某个感兴趣事件的发生时间。这些感兴趣事件通常包括临床治疗中患者的生存时间^[1]、机械系统中故障的发生时间^[2]以及客户行为分析中用户的购买时间^[3]等。当前,训练数据缺乏和数据质量不高是生存分析研究面临的两个重要挑战。一方面,由于数据采集的代价高昂,收集到的样本数量往往偏少(这一点在医学领域表现尤为突出),如果再考虑到一些应用领域实例所具有的内在高维特性,那么生存分析问题往往就是一个典型的高维小样本问题。另一方面,由于观测周期有限以及研究对象失访等原因,收集到的数据中还不可避免地存在一些删失实例,也就是说这些实例的感兴趣事件没有在观测周期内发生或者由于跟踪轨迹失效等原因未被

观测到. 特别地, 为描述方便, 对于那些已经观察到感兴趣事件发生的实例, 本文将感兴趣事件的发生时间称为生存时间, 相应的实例称为未删失实例. 而对于其他未观察到感兴趣事件发生的实例, 本文将该实例所耗的观测时间称为删失时间, 相应的实例称为删失实例. 通常右删失是实际应用中最常见的情形, 即删失实例的实际生存时间大于或者等于删失时间. 因此, 不失一般性, 本文将生存分析的研究对象限定为右删失实例.

传统的回归分析模型通常只能将未删失实例作为训练数据, 而未删失实例数量的明显不足很容易导致模型的过拟合, 而另一方面删失实例的合理利用则有助于改善生存分析模型的泛化性能. 因此, 一般不采用传统的回归模型来处理生存分析问题. Cox 比例风险模型^[4]和参数删失回归模型^[5]是生存分析中两类应用最为广泛的模型. 但是, 这两类方法均存在较为严重的缺陷, 所以预测效果并不是很理想. 具体地, Cox 模型是建立在比例风险假设上的生存分析模型, 其假设实例之间的风险比是常数. 因此导致所有实例的生存曲线将呈现相似的形状. 显然, 这种假设在实际应用中是过于严格的, 并且如果数据中存在具有相同生存时间的实例时, Cox 模型还需使用一些近似的方法来处理数据, 这有可能会带来导入偏差 (Inducing bias). 另一方面, 参数删失回归模型的预测性能高度依赖于生存时间分布假设的选取. 然而, 在实际应用中, 存在很多影响感兴趣事件发生的因素, 因此很难选择一个合适的理论分布.

近年来, 随着机器学习理论在各个领域的出色表现, 研究人员开始引入机器学习方法来研究生存分析问题. 相较于传统的生存分析模型, 机器学习能够从有限的中学习更多的信息, 例如数据特征的分布规律及实例特征的抽象表示, 同时还具有出色的函数拟合能力. 尤其是机器学习中的多任务学习方法^[6], 能够帮助模型学习到多个任务之间的共享判别特征, 从而提高模型的泛化能力, 降低新实例的预测错误率. 例如, Li 等^[7]创新性地将生存分析问题建模成预测各个时间间隔生存状态的多任务问题, 并通过对所设计的线性预测器施加 $l_{2,1}$ 范数正则化, 不仅可以筛选出具有跨任务判别能力的共享特征, 还可以缓解高维小样本问题所带来的过拟合缺陷, 他们所提出的 MTLA/ MTLA.V2 模型在生存分析问题上取得了很好的效果. 但是, 这些多任务学习方法依然存在一些缺陷, 即这些模型均未考虑到数据中可能存在的复杂噪声所带来的影响.

为了克服上述缺陷, 本文提出一类噪声容错弱

监督直推式矩阵补全模型 (Weakly supervised transductive matrix completion, WSTMC) 来预测删失实例和新实例的生存时间. 具体地, 基于实例特征潜在的低秩属性, 本文首先将生存分析问题建模为一类多任务直推式矩阵补全 (Multi-task transductive matrix completion, MTMC) 模型^[8]. 受益于直推式学习机制, 该模型不仅可以利用删失实例作为有限训练样本 (非删失实例) 的有效补充, 而且可以在训练阶段同时探索测试样本和训练样本的特征分布, 从而提高模型在测试样本上的泛化能力. 其次, 为了克服 MTMC 模型的噪声敏感性, 本文引入混合高斯分布 (Mixture of Gaussians distribution, MoG) 来拟合实际应用中的未知噪声类型, 其动机在于 MoG 理论上可以拟合任意连续分布^[9]. 进一步, 为了缓解实际应用中高维小样本所带来的过拟合缺陷, 我们设计了一类新颖的多任务直推式特征选择机制, 以期所提出的模型能在去噪后的特征空间自适应地选择出跨任务的共享判别特征, 从而进一步增强模型的泛化能力. 同时, 我们还引入了相邻时间间隔生存状态的先验时序稳定性来指导模型生成软类别标记. 此外, 我们也设计了一类拟期望最大化迭代优化算法来求解所提出的 WSTMC 模型. 针对该模型所涉及的多个超参数, 我们采用了贝叶斯优化方法来自适应地进行选择. 最后, 5 个真实数据集上的实验结果验证了所提出的 WSTMC 模型优于当前广泛使用的 4 类 18 种生存分析方法.

本文结构安排如下: 第 1 节对相关工作进行讨论; 第 2 节介绍一些数学基础及矩阵补全预备知识; 第 3 节为本文主要部分, 详细阐述了所提出的噪声容错弱监督直推式矩阵补全 (WSTMC) 模型及相应的优化算法; 第 4 节是实验部分, 首先介绍了实验数据集和所采用的性能评价指标, 接着阐述了实验方法和对比模型, 最后分析了实验结果; 第 5 节对本文研究内容进行了总结与展望.

1 相关工作

现有的生存分析方法大致可分为如下几类: 1) Cox 比例风险模型; 2) 参数删失回归模型; 3) 线性模型; 4) 多任务学习模型. 其中, Cox 比例风险模型是生存分析中应用最为广泛的模型之一. 在统计和数据挖掘领域, 该模型吸引了众多研究人员的关注与兴趣. 近年来, 由于各种数据采集和数据分析技术的发展, 高维数据在各类实际应用包括生存分析领域频繁出现, 例如医学领域中微阵列基因表达数据往往包含数千维特征. 为了缓解高维数据带来的过拟合问题, 研究人员在传统的 Cox 模型中加入

各种正则化项来帮助模型选择合适的特征以降低特征维度, 从而提出了 LASSO-COX (Least absolute shrinkage and selection operator Cox) 模型^[10]、EN-COX (Elastic-net Cox) 模型^[11] 和 KEN-COX (Kernel elastic-net Cox) 模型^[12]. 它们分别在传统模型中加入了 l_1 范数正则化项、弹性网正则化项和核化的弹性网正则化项. 这种方式扩展了 Cox 比例风险模型的应用场景, 但并没有解决该模型中比例风险假设在实际应用中过于严苛而可能带来的欠拟合问题. 参数删失回归模型是生存分析中另一类常用的重要模型^[5, 13-15]. 该模型假设实例的生存时间或者对数生存时间满足某种特殊的数据分布. 其中较常见的假设包括指数分布、威布尔分布 (Weibull distribution)、对数 (Logistic) 分布和极值分布^[13] 等. 此外, 一些线性模型还进一步假设生存时间或对数生存时间与特征之间存在线性关联关系. 这类线性模型可以被看成一类特殊的参数删失回归模型. 例如, 结合高斯分布的 Tobit^[14] 模型以及引入 Kaplan-Meier 估计的 Buckley-James 回归模型^[15]. 在实际应用中, 如何选择出合适的分布限制了参数删失回归模型进一步提高生存分析的预测效果.

近年来, 为了松弛传统统计模型中所依赖的严苛假设, 即比例风险假设和生存时间分布假设等, 一些研究人员开始采用机器学习领域的多任务学习模型来研究生存分析问题. 例如, Li 等^[7] 提出了一种 $l_{2,1}$ 范数正则化的多任务学习模型 MTLSA/MTLSA-V2, 该模型利用 $l_{2,1}$ 范数来共享任务之间的相关性, 同时解决高维小样本数据带来的过拟合问题; 此外还在模型中加入了相邻时间间隔内生存状态稳定的先验条件来约束模型生成标记. 但是, 该方法仅仅预设数据分布包含单一类型的高斯噪声, 忽略了特征矩阵中可能存在的未知复杂噪声类型, 因此存在一定程度的噪声敏感性缺陷. 受现有模型尤其是近年来多任务学习生存分析模型的启发, 本文提出了一种噪声容错的弱监督多任务学习 (WSTMC) 模型来处理生存分析问题. 相较于以往的模型, 所提出的 WSTMC 模型不仅有效利用了特征矩阵的内在低秩属性, 还能够通过所引入的 MoG 噪声模型来平滑数据分布中存在的复杂噪声. 此外, 不同于传统的多任务学习方法, 我们还在去噪后的特征空间引入了多任务直推式特征选择机制, 帮助模型选择出跨任务共享的判别特征, 这在一定程度上有利于缓解高维小样本数据固有的过拟合问题.

2 预备知识

2.1 数学基础

定义 1. 瘦型奇异值分解^[16]. 设矩阵 X 是一个

秩为 r 的 $m \times d$ 维矩阵, 则存在两个单位正交矩阵 $U \in \mathbf{R}^{m \times r}$ 和 $V \in \mathbf{R}^{d \times r}$, 以及对角阵 $\Sigma = \text{diag}\{\sigma_i | 1 \leq i \leq r\}$, 其中奇异值 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ 且满足

$$X = U\Sigma V^T \quad (1)$$

式 (1) 称为矩阵 X 的瘦型奇异值分解.

定义 2. 矩阵范数^[16]. 设秩为 r 的矩阵 $X \in \mathbf{R}^{m \times d}$ 存在如式 (1) 所示的瘦型奇异值分解, 则

1) 矩阵 X 的核范数定义为

$$\|X\|_* = \sum_{i=1}^r |\sigma_i| \quad (2)$$

2) 矩阵 X 的 Frobenius 范数定义为

$$\|X\|_F = \left(\sum_{i=1}^m \sum_{j=1}^d X_{ij}^2 \right)^{\frac{1}{2}} \quad (3)$$

3) 矩阵 X 的 $l_{2,1}$ 范数定义为

$$\|X\|_{2,1} = \sum_{i=1}^m \left(\sum_{j=1}^d X_{ij}^2 \right)^{\frac{1}{2}} \quad (4)$$

定义 3. 近邻算子^[17]. 设 $F(X)$ 为矩阵 X 的实值凸函数, 则对于任意矩阵 M 及 $\tau > 0$, 函数 $F(X)$ 的近邻算子定义为

$$\text{prox}_{\tau F(X)}(M) = \arg \min_X \tau F(X) + \frac{1}{2} \|X - M\|_F^2 \quad (5)$$

定义 4. 阈值收缩算子^[17]. 假设 $\tau > 0$, 阈值收缩算子 $\text{prox}_{\tau \|X\|_1}(M)$ 定义为

$$\text{prox}_{\tau \|X\|_1}(M) = S_\tau(M) \quad (6)$$

其中, $S_\tau(M)$ 定义为

$$[S_\tau(M)]_{ij} = \text{sign}(M_{ij}) \times \max(0, |M_{ij}| - \tau) \quad (7)$$

其中, $\text{sign}(\cdot)$ 为符号函数.

定义 5. 核范数近邻算子^[17]. 设秩为 r 的矩阵 M 的奇异值分解为 $U\Sigma V^T$, 对任意的 $\tau > 0$, 核范数近邻算子 $\text{prox}_{\tau \|X\|_*}(M)$ 定义为

$$D_\tau(M) = U S_\tau(\Sigma) V^T \quad (8)$$

定义 6. $l_{2,1}$ 范数近邻算子^[18]. 对于矩阵 $M \in \mathbf{R}^{m \times d}$ 和任意的 $\tau > 0$, 其相应的 $l_{2,1}$ 范数近邻算子 $\text{prox}_{\tau \|X\|_{2,1}}(M)$ 定义为

$$\text{prox}_{\tau \|X\|_{2,1}}(M) = J_\tau(M) \quad (9)$$

其中, $J_\tau(M)$ 定义为

$$(J_\tau(M))_{(i)} = \frac{\max\{0, \|M_{(i)}\|_2 - \tau\}}{\|M_{(i)}\|_2} M_{(i)} \quad (10)$$

其中, $i = 1, 2, \dots, m$, $M_{(i)}$ 表示矩阵 M 的第 i 行, $\|\cdot\|_2$ 表示向量的 l_2 范数.

定理 1. 近邻前向后向分裂 (Proximal forward backward splitting, PFBS) 算法^[18]. 假设 F_1, F_2 是两个下半连续的凸函数, F_2 在 $\mathbf{R}^{m \times d}$ 中可微且对某个常数 $\beta > 0$ 具有 β -Lipschitz 连续梯度, 即 $\|\nabla F_2(U) - \nabla F_2(V)\|_F \leq \beta\|U - V\|_F$, 则凸优化问题

$$\min_X (F_1(X) + F_2(X)) \quad (11)$$

有如下性质:

1) 如果 $F_1 + F_2$ 是强制的, 即 $\lim_{\|X\|_F \rightarrow +\infty} (F_1(X) + F_2(X)) = +\infty$, 则该凸优化问题至少存在一个解;

2) 如果 $F_1 + F_2$ 是严格凸的, 则该凸优化问题至多存在一个解;

3) 如果 F_1 和 F_2 满足上述两个条件, 则凸优化问题存在唯一解, 且对于任意的初始值 X^0 以及 $0 < \delta < 2/\beta$ 用如下方法生成的迭代序列 X^{k+1} 收敛到凸优化问题的唯一解

$$X^{k+1} = \text{prox}_{\delta F_1(X)}(X^k - \delta \nabla F_2(X^k)) \quad (12)$$

2.2 矩阵补全

矩阵补全可视为压缩感知理论^[19-21] 从一维向量空间向二维矩阵空间的一种自然推广, 旨在研究如何在数据不完整的情形下对数据的缺失信息进行填补. 标准的矩阵补全问题可以建模为如下形式的秩最小化约束优化模型^[22]:

$$\begin{aligned} \min_X \text{rank}(X) \\ \text{s.t. } P_\Omega(M) = P_\Omega(X) \end{aligned} \quad (13)$$

其中, X 为待补全的目标矩阵, M 为部分元素已知的采样矩阵, Ω 表示采样元素的索引集合, $P_\Omega(\cdot)$ 为正交投影算子, 表示当 $(i, j) \in \Omega$ 时, M_{ij} 为采样元素, 即

$$[P_\Omega(M)]_{ij} = \begin{cases} M_{ij}, & \text{如果 } (i, j) \in \Omega \\ 0, & \text{否则} \end{cases} \quad (14)$$

Fazel^[23] 证明了矩阵核范数是秩函数在矩阵谱范数意义下单位球上的最佳凸逼近. 因此, 类似于压缩感知理论中常用的将向量 ℓ_0 范数松弛为向量 ℓ_1 范数的技巧, 为了使标准的矩阵补全问题易于求解, 一个自然的想法就是利用凸核范数代替非凸秩函数, 也就是将原先的秩最小化问题松弛为如下形式的核范数最小化模型^[22]:

$$\begin{aligned} \min_X \mu \|X\|_* \\ \text{s.t. } P_\Omega(M) = P_\Omega(X) \end{aligned} \quad (15)$$

然而, 尽管标准的矩阵补全理论在数据表示、信息重建以及图像恢复等领域取得了较大成功, 但是仍然无法解决弱监督情形下的多标记分类问题. 为此, Goldberg 等^[8] 于 2010 年提出了一类新颖的

多任务直推式矩阵补全 (MTMC) 模型并将其成功应用于多标记图像分类问题. 这类 MTMC 模型主要建立在两个假设之上. 首先, 假设包含 m 个 d 维样本的特征矩阵 X 和 t 个任务的矩阵 $Y \in \{0, 1, ?\}^{m \times t}$ 之间存在线性依赖关系, 也就是说, 存在隐含的权重矩阵 $W^0 \in \mathbf{R}^{d \times t}$ 使得 $Y = XW^0$. 其次, 假设特征矩阵 X 满足内在低秩属性. 基于上述假设, 容易推断出特征-任务堆叠矩阵 $[X, Y]$ 也满足低秩性. 因此标准的 MTMC 模型可建模如下:

$$\min_Z \left(\mu \|Z\|_* + \frac{1}{2} \|Z_X - X\|_F^2 + \beta \sum_{(i,j) \in \Omega_Y} C_y(Z_{i(d+j)}, Y_{ij}) \right) \quad (16)$$

其中, $Z = [Z_X, Z_Y]$ 表示待求解的特征-任务堆叠矩阵, Z_X 表示待求解的去噪特征子矩阵, Z_Y 表示待求解的软标记子矩阵, Ω_Y 表示任务矩阵 Y 中已知标记的索引集合, $C_y(\cdot)$ 表示标记损失函数 (通常为 logistic 损失函数或平方损失函数).

3 面向生存分析的 WSTMC 模型

3.1 问题阐述

在生存分析问题中, 对于第 i 个实例, 如果该实例未删失, 则可以观察到它的生存时间 $S_time(i)$, 反之, 可以得到该实例的删失时间 $C_time(i)$. 为了方便表达, 定义观察时间 $O_time(i)$ 如下:

$$O_time(i) = \begin{cases} S_time(i), & \text{若 } \delta_i = 0 \\ C_time(i), & \text{若 } \delta_i = 1 \end{cases} \quad (17)$$

其中, $\delta_i \in \{0, 1\}$ 表示删失状态指示符, $\delta_i = 1$ 表示该实例为删失实例, $\delta_i = 0$ 表示该实例为未删失实例. 通常, 可以采用三元组 $(x_i, O_time(i), \delta_i)$ 来表示生存分析问题中的实例, 其中下标 i 表示实例编号.

本文目标是通过合理建模特征 x_i 和生存时间 $S_time(i)$ 之间的关系, 从而能够依据实例特征, 准确预测该实例的生存时间. 然而, 由于存在删失数据, 传统的分类和回归方法并不适合生存分析问题. 受 Li 等^[7] 的启发, 为了更好地利用删失实例, 本文将生存分析问题建模为多任务学习模型. 具体地, 先将连续的时间进行离散化, 也就是将整个研究周期离散化为若干个相等的时间间隔, 通过考察各个时间间隔上实例的生存状态从而间接得到整个研究周期内实例的生存状态. 由此每个实例 $(x_i, O_time(i), \delta_i)$ 可转换为三元组 $(x_i, O_num(i), \delta_i)$, 其中 $O_num(i)$ 表示第 i 个实例的观察时间所对应的时间间隔数. 此外, 第 i 个实例在第 j 个时间间隔上的生存状态用

$Y_{ij} \in Y$ 来表示. 如果在该时间间隔内感兴趣事件未发生则 $Y_{ij} = 1$, 已发生则 $Y_{ij} = 0$, 状态未知则使用 “?” 表示. 通过上述方式, 生存分析问题转化为了多任务学习问题, 预测每个时间间隔上的生存状态就是一个学习任务. 值得注意的是, 对于每个数据集选择最大的 $O_num(i)$ 作为多任务学习的任务总数 t . 图 1 图示了如何将生存分析问题表述为弱监督多任务学习问题. 从图 1 可以看出, 对于第 i 个实例, 如果实例是未删失的, 则任务矩阵 Y 中相应行的生存状态从开始到第 $O_num(i)$ 个时间间隔都记为 “1”, 之后都记为 “0”. 如果实例是删失的, 则任务矩阵 Y 中相应行的生存状态从开始到第 $O_num(i)$ 个时间间隔都记为 “1”, 之后都记为 “?”.

综上所述, 可以通过先将生存分析问题阐述为弱监督多任务问题, 然后引入交叉熵损失函数将其建模为如下形式的改进 MTMC 模型

$$\min_Z \left(\mu \|Z\|_* + \frac{1}{2} \|Z_X - X\|_F^2 + \beta R(Z_Y, Y) \right) \quad (18)$$

其中, $R(Z_Y, Y) = \sum_{(i,j) \in \Omega_Y} C_y(\sigma(Z_{(i+d)j}), Y_{ij})$, $C_y(u, v) = u \log v + (1 - u) \log(1 - v)$ 是交叉熵损失函数, $\sigma(Z_{ij}) = 1/(1 + e^{-cZ_{ij}})$ 是 logistic 函数, c 是大于 0 的常数.

3.2 建模

然而, 直接将生存分析问题阐述为上述多任务

直推式矩阵补全 (MTMC) 问题仍然存在以下几点不足: 1) 噪声容错性差. MTMC 模型采用 Frobenius 范数拟合特征噪声, 已有研究表明 Frobenius 范数是高斯分布的最佳逼近^[21], 然而生存分析数据所面临的噪声通常并不是单一的高斯分布且其实际分布类型也往往是未知的^[24]. 2) 高维小样本过拟合. 在实际的生存分析中, 由于研究代价高昂和观测时间的限制, 能收集到的观测实例往往数量偏少. 而受益于数据收集和检测技术的发展, 所能获取的实例特征则越来越丰富, 这加剧了高维小样本问题所带来的模型过拟合风险. 3) 先验信息融合性不强. MTMC 模型忽略了数据固有的先验信息, 在生存分析中, 相邻时间间隔内生存状态存在时序稳定性, 这种先验时序信息的合理融合通常能有效提升模型的预测性能.

为此, 针对噪声容错性差的缺陷, 我们考虑引入 MoG 模型来拟合一般的未知复杂噪声, 其动机来源于已有研究表明 MoG 分布能在理论上逼近任意连续分布, 这种噪声建模思想已经在一些典型的机器学习及计算机视觉任务中得到了很好的应用^[24-25]. 具体地, 不妨假设特征矩阵中的每一个元素均由两部分组成, 即

$$X_{ij} = X_{ij}^o + E_{ij} \quad (19)$$

其中, X_{ij}, X_{ij}^o, E_{ij} 分别表示第 i 个样本的第 j 个观测特征、真实潜在特征及相应的噪声. 假设噪声 E_{ij}

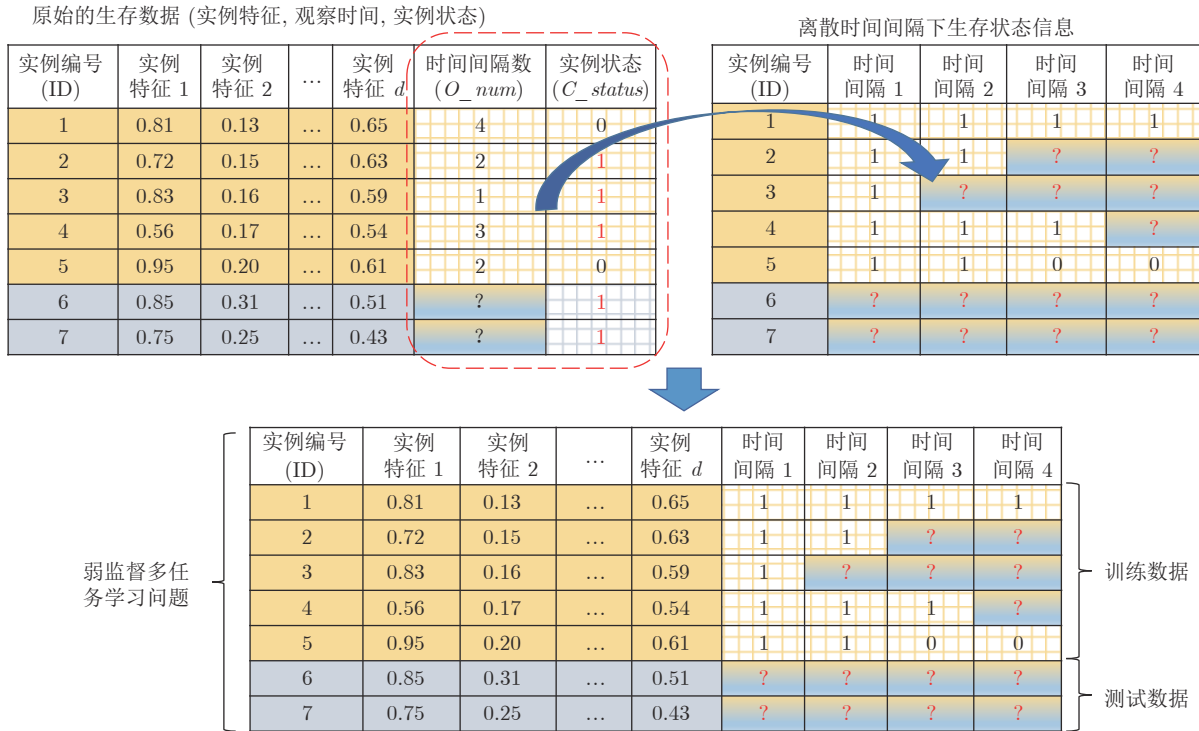


图 1 生存分析问题建模为弱监督多任务学习问题的图示

Fig.1 Illustration of formulating the survival analysis problem as a weakly supervised multi-task learning problem

服从独立的同 MoG 分布, 即

$$p(E_{ij}; \pi, \sigma) = \sum_{k=1}^K \pi_k \mathcal{N}(E_{ij}; 0, \sigma_k^2) \quad (20)$$

其中, $\pi = \{\pi_1, \pi_2, \dots, \pi_K\}$, $\sigma = \{\sigma_1, \sigma_2, \dots, \sigma_K\}$, $\mathcal{N}(E_{ij}; 0, \sigma_k^2)$ 表示均值为 0、方差为 σ_k^2 的高斯分布, K 表示混合高斯分布中独立的高斯分量个数, π_k 表示第 k 个高斯分量的混合比例且满足 $\sum_{k=1}^K \pi_k = 1, \pi_k \geq 0$. 由于假设每一个特征受到的噪声污染是相互独立的, 因此, 定义如下似然函数

$$P(E; \pi, \sigma) = \prod_{i,j} p(E_{ij}; \pi, \sigma) \quad (21)$$

进一步可得到 log 似然函数

$$L(E, \pi, \sigma) = \sum_{i,j} \log \prod_{i,j} p(E_{ij}; \pi, \sigma) \quad (22)$$

我们的目标是通过最大化该 log 似然函数解析出特征矩阵 X 中存在的未知复杂噪声. 为此, 结合式 (18), 易得到如下所示的噪声容错 MTMC 模型

$$\begin{aligned} \min_{Z, E, \pi, \sigma} & \left(\mu \|Z\|_* - \frac{1}{2} L(E, \pi, \sigma) + \beta R(Z_Y, Y) \right) \\ \text{s.t. } & E = X - Z_X, \sum_{k=1}^K \pi_k = 1, \pi_k \geq 0 \end{aligned} \quad (23)$$

其次, 针对实际应用中高维小样本所带来的过拟合缺陷, 我们设计了一种新颖的多任务直推式特征选择机制来自适应地选择跨任务之间的共享判别特征, 从而一定程度上降低原始高维样本的特征维度, 进而缓解模型的过拟合缺陷. 为此, 首先考虑将 MTMC 模型中关于实例特征与标记之间的隐式线性依赖假设 $Z_Y = Z_X W$ 以正则化项的形式显式加入目标函数 (23), 即

$$\begin{aligned} \min_{Z, W, E, \pi, \sigma} & \left(\mu \|Z\|_* - \frac{\alpha}{2} L(E, \pi, \sigma) + \beta R(Z_Y, Y) + \right. \\ & \left. \frac{1}{2} \|Z_Y - Z_X W\|_F^2 \right) \\ \text{s.t. } & E = X - Z_X, \sum_{k=1}^K \pi_k = 1, \pi_k \geq 0 \end{aligned} \quad (24)$$

其中, $W \in \mathbf{R}^{d \times t}$ 表示显式的线性预测器. 然后, 继续考虑在上述模型 (24) 的基础上引入如下的多任务直推式特征选择机制

$$\min_W \left(\lambda \|W\|_{2,1} + \frac{1}{2} \|Z_Y - Z_X W\|_F^2 \right) \quad (25)$$

其中, $\|W\|_{2,1}$ 项用来约束线性预测器 W 保持行稀疏, 从而学习到跨任务之间的共享判别特征. 特别地, 我们注意到这里采用的特征选择明显区别于一

般的特征选择机制, 一方面我们是在去噪特征空间中实施特征选择, 另一方面所有的实例 (包括删失实例、未删失实例以及测试实例) 均参与了特征选择.

最后, 针对先验信息融合性不强的缺陷, 进一步考虑引入如下的 Toeplitz 矩阵 $S \in \mathbf{R}^{t \times (t-1)}$ 并以 $\|Z_Y S\|_F^2$ 正则化项的形式来诱导软标记矩阵 Z_Y 的生成

$$S = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & -1 \end{bmatrix} \quad (26)$$

为此, 最终构建出噪声容错弱监督直推式矩阵补全 (WSTMC) 模型为

$$\begin{aligned} \min_{Z, W, E, \pi, \sigma} & \left(\mu \|Z\|_* - \frac{\alpha}{2} L(E, \pi, \sigma) + \beta R(Z_Y, Y) + \right. \\ & \left. \lambda \|W\|_{2,1} + \frac{1}{2} \|Z_Y - Z_X W\|_F^2 + \frac{\gamma}{2} \|Z_Y S\|_F^2 \right) \\ \text{s.t. } & E = X - Z_X, \sum_{k=1}^K \pi_k = 1, \pi_k \geq 0 \end{aligned} \quad (27)$$

3.3 优化

由于所提出的 WSTMC 模型涉及到最大似然估计项 $L(E, \pi, \sigma)$ 的求解, 受到期望最大化方法 (Expectational maximization, EM) 的启发, 我们也考虑引入隐含变量 $h_{ijk} \in \{0, 1\}$ ($\sum_{k=1}^K h_{ijk} = 1$) 来表征噪声 E_{ij} 是否属于混合高斯分布的第 k 个分量, 从而得到完全数据的 log 似然函数

$$\begin{aligned} L(E, H, \pi, \sigma) &= \log P(E, H; \pi, \sigma) = \\ & \sum_{i,j} \sum_k \log \pi_k^{h_{ijk}} \mathcal{N}(E_{ij}; 0, \sigma_k^2)^{h_{ijk}} \end{aligned} \quad (28)$$

然后, 采用 EM 方法的思想求解所提出的 WSTMC 模型, 为此, 我们令 $\theta = (Z, W, E, \pi, \sigma)$, 首先通过 E 步求出隐含变量 h_{ijk} 的期望 r_{ijk} (这里 r_{ijk} 表示在当前模型参数下第 (i, j) 个观测数据 E_{ij} 来自第 k 个分量的概率, 也即分量 k 对观测数据 E_{ij} 的响应度):

E-step (求解 r_{ijk}).

$$r_{ijk} = \frac{\tilde{r}_{ijk}}{\sum_{k=1}^K \tilde{r}_{ijk}} \quad (29)$$

其中, $\tilde{r}_{ijk} = \pi_k \mathcal{N}(X_{ij} - (Z_X)_{ij} | 0, \sigma_k^2)$.

接下来, 定义函数 $Q(\theta, \theta^{(t)})$ 如下:

$$Q(\theta, \theta^{(t)}) = E_H \left\{ \frac{\alpha}{2} (\log P(E, H; \theta) | E, \theta^{(t)}) - \lambda \|W\|_{2,1} - \mu \|Z\|_* - \beta R(Z_Y, Y) - \frac{\gamma}{2} \|Z_Y S\|_F^2 - \frac{1}{2} \|Z_Y - Z_X W\|_F^2 \right\} \quad (30)$$

根据 EM 方法的思想, 接下来可以通过使函数 $Q(\theta, \theta^{(t)})$ 极大化的方式来求解参数 θ , 即

$$\theta^{(t+1)} = \arg \max_{\theta} Q(\theta, \theta^{(t)}) \quad (31)$$

此时, 可采用交替求解方法, 因此有:

M-step1 (求解 π_k).

$$\pi_k = \frac{\sum_{i,j} r_{ijk}}{\sum_k \sum_{i,j} r_{ijk}} \quad (32)$$

M-step2 (求解 σ_k^2).

$$\sigma_k^2 = \frac{\sum_{i,j} r_{ijk} (X_{ij} - (Z_X)_{ij})^2}{\sum_{i,j} r_{ijk}} \quad (33)$$

M-step3 (求解 Z, W).

$$(Z^*, W^*) = \arg \min_{Z, W} \left(\mu \|Z\|_* + \beta R(Z_Y, Y) + \frac{\alpha}{2} \|B \odot (X - Z_X)\|_F^2 + \lambda \|W\|_{2,1} + \frac{1}{2} \|Z_Y - Z_X W\|_F^2 + \frac{\gamma}{2} \|Z_Y S\|_F^2 \right) \quad (34)$$

其中, $B_{ij} = (\sum_{k=1}^K r_{ijk} / 2\sigma_k^2)^{\frac{1}{2}}$ 且 \odot 表示逐元素的哈达马乘积. 我们采用新近提出的块近邻线性算法 (Block prox-linear, BPL)^[26] 来求解子问题 (34). 该算法的关键步骤在于迭代优化下面两个变量 (l 表示迭代次数):

$$\begin{cases} Z_l = \arg \min_Z \left(\mu \|Z\|_* + \beta R(Z_Y, Y) + \frac{\alpha}{2} \|B \odot (X - Z_X)\|_F^2 + \frac{1}{2} \|Z_Y - Z_X W_{l-1}\|_F^2 + \frac{\gamma}{2} \|Z_Y S\|_F^2 \right) & (35a) \\ W_l = \arg \min_W \left(\frac{1}{2} \|(Z_Y)_l - (Z_X)_l W\|_F^2 + \lambda \|W\|_{2,1} \right) & (35b) \end{cases}$$

具体地, 对于变量 Z , 根据定理 1 可由如下方式求解:

M-step3.1 (求解 Z).

$$Z_l = D_{\mu\tau_Z^l} (Z_{l-1} - \tau_Z^l \nabla G(Z_{l-1})) \quad (36)$$

其中, τ_Z^l 表示第 l 次迭代的梯度步长, $\nabla G(Z)$ 代表函数 $G(Z)$ 的梯度. $G(Z)$ 定义为

$$G(Z) = \beta R(Z_Y, Y) + \frac{\alpha}{2} \|B \odot (X - Z_X)\|_F^2 + \frac{1}{2} \|Z_Y - Z_X W_{l-1}\|_F^2 + \frac{\gamma}{2} \|Z_Y S\|_F^2 \quad (37)$$

对于变量 W , 同样根据定理 1 可由如下方式求解:

M-step3.2 (求解 W).

$$W_l = J_{\lambda\tau_W^l} (W_{l-1} - \tau_W^l \nabla M(W_{l-1})) \quad (38)$$

其中, τ_W^l 表示第 l 次迭代的梯度步长, $\nabla M(W)$ 代表函数 $M(W)$ 的梯度. $M(W)$ 定义为

$$M(W) = \frac{1}{2} \|(Z_Y)_l - (Z_X)_l W\|_F^2 \quad (39)$$

基于上述分析, 所提出的求解 WSTMC 模型的拟期望最大化优化算法可概述为算法 1.

算法 1. 所提出的求解 WSTMC 模型的拟期望最大化优化算法

输入. 特征矩阵 X , 任务矩阵 Y , 以及超参数 $\alpha, \beta, \gamma, \lambda, \mu, c, K$.

输出. 线性预测器 W^{opt} , 特征-任务矩阵 Z^{opt} .

1: 初始化线性预测器 W_0 为全 0 矩阵, 初始化特征-任务矩阵 Z_0 为 $[X, Y]$ 的秩 1 逼近矩阵 (未知元素设为 0).

2: **While** not converged **do**

3: (E-Step for r):

更新 r_{ijk} by 式 (29);

4: (M-step for π, σ):

更新 π_k, σ_k by 式 (32) 和式 (33);

5: (M-step for Z, W):

更新 Z, W by 式 (36) 和式 (38);

6: **End While**;

7: 输出 $W^{\text{opt}} \leftarrow W, Z^{\text{opt}} \leftarrow Z$.

3.4 时间复杂度分析

求解 WSTMC 模型的算法 1 是一个迭代优化算法, 迭代过程由 E-step 和 M-step 组成. 假设算法 1 的外循环迭代次数为 N_{EM} , 简单分析可知 E-step 更新 r_{ijk} 所需的时间复杂度为 $O(Kmd)$; 基于 E-step 的计算结果更新 M-step 的 π, σ 所需的时间复杂度均为 $O(Kmd)$; 基于 r 和 σ 计算中间变量矩阵 B 所需的时间复杂度为 $O(Kmd)$; 接下来, 在更新 M-step 的 Z, W 时又涉及到两个需要迭代求解的子问题 (35a) 和 (35b), 此时不妨假设求解这两个子问题的内部循环迭代次数为 N_{BPL} , 那么对于 M-step 的子问题 (35a), 求解 Z 的时间复杂度为 $O(mdt + m(d+t)\min\{m, (d+t)\})$, 对于 M-step 的另一个子问题 (35b), 求解 W 的时间复杂度则为 $O(mdt + md^2)$. 综上所述, 算法 1 的时间复杂度为 $O(N_{\text{EM}}(Kmd + N_{\text{BPL}}(m(d+t)\min\{m, (d+t)\} + mdt + md^2)))$. 为简单起见, 考虑到实际问题中样本的特征维度 d 通常大于任务个数 t , 所以求解模型 WSTMC

的算法时间复杂度可以简化为 $O(N_{EM}N_{BPL}(m^2d+md^2))$. 在本文第 4 节的实验部分, 我们比较了 WSTMC 模型和其他 6 种同属于多任务学习范型的生存分析模型, 对于这 6 种相关模型的详细介绍及实验结果参见第 4.4 节及第 4.5 节. 表 1 给出了 WSTMC 及其他 6 种相关模型的时间复杂度 (为简单起见, 这里给出的时间复杂度均假设样本的特征维度 d 大于任务个数 t). 从表 1 可以看出, 相较于其他 6 种多任务学习方法, 本文所提出的 WSTMC 模型具有最高的算法时间复杂度. 然而, 实验中发现, 每一步并不要求出子问题的精确解, 实际上, 只需更新 Z 与 W 各一次得到子问题的一个近似解, 已足以使算法最终获得与精确求解子问题时相当的性能.

表 1 WSTMC 及其他相关模型的时间复杂度比较
Table 1 Time complexity comparison of the proposed WSTMC and the other related models

模型	时间复杂度
Multi-LASSO ^[27]	$O(dtm_{tr})$
Multi- $\ell_{2,1}$ ^[27]	$O(dtm_{tr})$
MTLSA ^[7]	$O(Ndtm_{tr})$
MTLSA.V2 ^[7]	$O(Ndtm_{tr})$
MTMC ^[8]	$O(Nmd \min\{m, d\})$
NLMC ^[28]	$O(Nmdt)$
WSTMC	$O(N_{EM}N_{BPL}(md^2 + m^2d))$

注: m 表示样本数 (包括训练样本和测试样本); m_{tr} 表示训练样本数; d 表示样本特征维数; t 表示任务数; N 表示迭代次数.

4 实验

在本节中, 首先介绍实验数据集和评价指标, 然后介绍所采用的实验方法和生存分析比较模型, 最后报告实验结果并展开实验分析.

4.1 实验数据集

本文使用了 5 个公开的癌症生存分析数据集, 具体包括: NSBCD (Norwegian/Stanford Breast Cancer Data)^[29]、DBCD (Dutch Breast Cancer Data)^[30]、Lung (Gene Expression Profiles of Lung Adenocarcinoma)^[31]、DLBCL (Diffuse Large B-Cell Lymphoma)^[32] 和 MCL (Mantle Cell Lymphoma)^[33] 数据集. 这些数据集可以分别从 http://user.it.uu.se/~liuya_610/download.html 和 <http://llmmp.nih.gov/MCL/> 下载. 表 2 概述了这些数据集的相关信息, 其中 “#Instances” 列表示数据集所包含的实例数 (包括删失实例和未删失实例), “#Features” 列表示相应数据集中实例的特征数, “#Censored” 列表示数据集所含的删失实例数, “#La-

bels” 列表示每个数据集对应的任务个数 (即所划分的时间间隔数, 其中 NSBCD 和 Lung 数据集以 “月” 作为时间间隔单位; DBCD、MCL 和 DLBCL 数据集以 “年” 作为时间间隔单位). 此外, 为了表明所采用的数据集是否为高维小样本问题, 我们还在 “#Ratios” 列记录了每个数据集中实例数与特征数之间的比例, 通常认为当样本个数比特征维数低一个数量级时即为高维小样本问题, 按照这个标准, 容易发现表 2 中后四个数据集都属于高维小样本问题.

表 2 实验所用数据集概述
Table 2 Details of datasets used in this study

Dataset	#Instances	#Features	#Censored	#Labels	#Ratios
NSBCD	115	549	77	188	0.2094
DBCD	295	4919	216	18	0.0599
Lung	86	7129	62	110	0.0120
DLBCL	240	7399	102	21	0.0137
MCL	92	8810	28	14	0.0104

4.2 评价指标

由于数据集中存在删失数据, 传统回归模型常用的评价指标在生存分析中并不适用. 类似于文献 [7], 本文也选用了 C-index 和 Weighted average AUC 两个指标来评估生存分析性能. 其中 C-index 指标侧重评估模型在所有任务上的整体回归性能, 而 Weighted average AUC 指标则注重评估模型在各个任务上的平均分类性能. 两个指标的具体定义如下:

1) C-index: 该指标旨在通过考虑不同事件的相对风险来评估预测值和实际值之间的差异. 例如, 考虑一对二元组变量 (T_i^*, O_i) 和 (T_j^*, O_j) , 其中 T_i^* 表示第 i 个实例的预测存活时间, O_i 表示第 i 个实例的观察时间. 首先定义一致性概率为

$$c = \Pr(T_j^* > T_i^* | O_j \geq O_i) \quad (40)$$

根据定义, 对于可以直接预测生存时间的模型来说, C-index 指标计算如下:

$$c = \frac{1}{num} \sum_{\delta_i=1} \sum_j I(T_j^* > T_i^*) \quad (41)$$

其中,

$$num = \sum_{\delta_i=1} \sum_j I(O_j > O_i) \quad (42)$$

其中, $I(\cdot)$ 为指示函数. 对于多任务类型的生存分析模型, 可以通过判断事件是否发生 (以设定阈值的方式) 从而计算出生存时间. 由于阈值的选取具有主观性, 因此根据 Li 等^[7] 的建议, 多任务模型中 C-

index 可以通过如下方式计算:

$$c = \frac{1}{num} \sum_{\delta_i=1} \sum_{T_j^* > T_i^*} I(S_j > S_i) \quad (43)$$

其中, S_i 表示第 i 个实例在所有任务上评分之和, 即

$$S_i = \sum_{j=\{1, \dots, t\}} y_{ij} \quad (44)$$

2) Weighted average AUC: 该指标旨在评估模型的整体分类性能, 即评估模型能否准确预测出实例在某时间间隔上的生存状态. 将 1 当做事件未发生的标记, 0 当做事件已发生的标记, 则每个时间段就可以看成一个分类任务, 整个任务矩阵就相当于 t 组分类任务. 具体地, Weighted average AUC 指标可以定义如下:

$$AUC_{avg} = \frac{\sum_{i=1}^t AUC^{(i)} n_c^{(i)}}{\sum_{i=1}^t n_c^{(i)}} \quad (45)$$

其中, $AUC^{(i)}$ 表示第 i 个任务的 AUC 值, $n_c^{(i)}$ 表示第 i 个任务上已知生存状态的实例数目.

4.3 实验方法

为了确定所提出的 WSTMC 模型的最佳参数 $(\gamma, \beta, \lambda, \mu, \alpha, c, K)$, 本文采用交叉验证的方式来评估实验结果. 具体来说, 对于实例数目超过 150 的生存数据集, 我们采用 5 折交叉验证, 其他的数据集则采用 3 折交叉验证. 为了公平起见, 其他对比方法均采用上述的交叉验证方式.

另一方面, 传统上一般使用网格搜索来寻找合适的参数, 但对于本文所提出的 WSTMC 模型来说, 这将非常耗时. 因此, 需要一种更加有效的超参数选择策略. 本文采用了基于贝叶斯优化^[34] 的自动调参方法来选择最佳的超参数. 该方法可以帮助模型选择出具有更高概率提升预测效果的超参数. 具体来说, 贝叶斯优化假设存在一个未知的函数 $\psi = CV_WSTMC(P)$, 其中 $P \in \mathbf{R}^7$ 表示一个超参数向量, ψ 表示 WSTMC 模型对于每组输入的参数向量进行交叉验证后取得的预测结果. 显然, 最终目的是寻找这个函数的最大值以及对应的超参数向量. 首先, 贝叶斯优化假设未知函数的结果是通过高斯过程 (Gaussian processes, GP) 采样得到的. 因此, 可以基于历史记录 (先前的超参数值及其相应的交叉验证的预测结果) 计算出超参数的后验概率分布. 然后, 重复迭代下面 3 个步骤, 直到满足停止标准 (迭代到最大次数, 或者连续 10 次迭代 ψ 值都无法得到提升).

步骤 1. 根据最大化采集函数选择出下一个最

具有“潜力”的评估点 (超参数);

步骤 2. 根据选择出的参数评估未知函数的预测值;

步骤 3. 将当前的记录加入到历史记录中, 并更新高斯过程. 值得注意的是, 本文没有选择 PI (Probability of improvement) 作为采集函数, 而是使用 EI (Expected improvement) 作为采集函数来选择下一次试验的超参数. 虽然 PI 策略能够选择相对当前预测结果提升概率最大的评估点, 但它仅仅反映了提升的概率, 并没有反映提升量的大小. 与之相比, EI 策略两者都能考虑到, 还能进一步处理局部和全局之间的关系^[35]. 此外, 协方差函数也存在很多种选择方案^[34]. 本文采用了自动相关性确定 (Automatic relevance determination, ARD) 作为高斯过程的协方差函数^[36], 它有助于有效地识别并去除参数向量中不相关的维度.

4.4 对比模型

为了验证所提出的 WSTMC 模型的有效性, 我们将其与 18 种广泛使用的生存分析模型进行了比较. 这些模型可归纳为如下 4 类: 基于 Cox 的模型、参数删失回归模型、线性模型和基于多任务的模型. 表 3 概述了这 18 种比较模型和所提出的 WSTMC 模型的相关特点, 主要从噪声容错机制、直推式学习机制、时序稳定性机制、特征选择机制和多任务学习机制等 5 个方面进行了比较. 具体地, 对于 Cox 模型, 我们选择了传统 Cox 模型, LASSO-COX 模型^[10], EN-Cox 模型^[11], Cox- $\ell_{2,1}$ 模型和 Cox-Trace 模型^[37]. 对于参数模型, 选择了 4 种基于不同分布假设的模型, 即 Weibull、Logistic、Log-logistic 和 Log-Gaussian. 此外, 还与 3 种线性模型进行了比较, 包括普通的最小二乘 (Ordinary least square, OLS) 模型^[38]、Tobit 模型^[14] 和 RWRSS (Regularized weighted residual sum-of-squares) 模型^[35]. 对于基于多任务的模型, 选择了 Multi-LASSO 模型、Multi- $\ell_{2,1}$ 模型^[27]、MTMC 模型^[8]、MTLSA/MTLSA.V2^[7] 模型和 NLMC^[28] 模型. 值得注意的是, OLS 模型、Multi-LASSO 模型和 Multi- $\ell_{2,1}$ 模型均无法处理删失实例. 因此, 它们只能基于未删失实例进行学习. 为公平起见, 比较模型所涉及的所有参数均采用了和 WSTMC 模型相同的交叉验证策略和参数调整方法.

4.5 实验结果

首先, 采用 C-index 指标来评估生存分析模型整体回归性能. 表 4 报告了不同方法在 5 种真实数据集上的实验结果. 对于每种数据集, 我们用粗体

表 3 对比模型的特征比较
Table 3 Comparison of characteristics for the competing models

	噪声容错性	直推式学习机制	时序稳定性	自适应特征选择	多任务学习机制
COX	×	×	√	×	×
LASSO-COX	×	×	√	√	×
EN-COX	×	×	√	√	×
Cox- $l_{2,1}$	×	×	√	√	×
Cox-Trace	×	×	√	×	×
Logistic	×	×	√	×	×
Weibull	×	×	√	×	×
Log-gaussian	×	×	√	×	×
Log-logistic	×	×	√	×	×
OLS	×	×	×	×	×
Tobit	×	×	√	×	×
RWRSS	×	×	√	√	×
Multi-LASSO	×	×	×	×	√
Multi- $l_{2,1}$	×	×	×	√	√
MTLSA	×	×	√	√	√
MTLSA.V2	×	×	√	√	√
NLMC	×	√	×	×	√
MTMC	×	√	×	×	√
WSTMC	√	√	√	√	√

将最优 C-index 值突出显示, 也同时用粗体加下划线标示出第二优和第三优的 C-index 值. 从表 4 可以看到, WSTMC 模型在所有数据集上均取得了最好的结果. 此外, 分析实验结果还可以发现排名前三的结果大多是由基于多任务学习的模型所获得, 这表明通过将原始的生存分析问题转化为预测各个时间间隔内生存状态的多任务学习问题, 可以有效利用任务之间共享的判别特征来提高预测性能. 同时, 我们也注意到 MTLA 和 MTLA.V2 模型的性能明显优于 Multi-Lasso 和 Multi- $l_{2,1}$, 这反映出能够在训练阶段利用所有实例 (删失实例和未删失实例) 的模型比只能够利用单一的未删失实例的模型效果更好. 除此之外, 对比 WSTMC 模型以及 MTMC 和 NLMC 模型, 由于后两种矩阵补全模型无法自适应地选择出跨任务共享的判别特征, 也没能利用时序稳定性这一先验知识, 所以效果始终差于我们所提出的模型. 另外, 我们也发现, 相较于 MTLA 和 MTLA.V2 模型, WSTMC 模型引入混合高斯分布来拟合数据分布中未知的复杂噪声, 有利于消除噪声对特征选择的影响, 从而进一步降低了预测错误率.

接着, 我们采用 Weighted average AUC 指标来评估生存分析模型在各个任务中的平均分类性能. 表 5 报告了不同方法在 5 种真实数据集上的实

验结果. 同样地, 对于每种数据集, 最佳结果使用粗体突出显示, 其他前两名的结果使用粗体加下划线突出显示. 从表 5 可以看出, 我们的模型除了在 DLBCL 数据集上效果略差外, 在其他数据集上始终优于其他对比方法. 综合表 4 和表 5 的实验结果, 我们可以看到参数删失回归模型在两个评价指标中都没有取得任何前三名的结果. 这表明常用的数据分布无法满足实际需求, 也就是说难以选择出合适的理论分布来建模实际问题. 相比之下, Cox- $l_{2,1}$ 模型、MTLSA 模型、MTMC 模型和 WSTMC 模型都在两个评价指标中获得过前三名. 一方面, 这验证了 $l_{2,1}$ 范数和直推式学习机制有利于缓解高维小样本带来的过拟合问题. 另一方面, 也证实了相较于传统的生存分析模型, 多任务学习方法, 尤其是有效地利用了先验时序稳定机制的多任务学习方法能够很好地解决生存分析问题.

4.6 消融实验

进一步, 为了验证 WSTMC 模型中 MoG 噪声容错机制、多任务直推式特征选择机制以及时序稳定性机制的有效性, 我们也比较了 3 种 WSTMC 的消融模型, 即: 未引入 MoG 噪声容错机制的 WSTMC-nM 模型、未引入多任务直推式特征选择机制的 WSTMC-nF 模型和未引入时序稳定性机制

表 4 所提出的 WSTMC 模型和其他比较模型在 C-index 指标上的性能比较 (标准差)
Table 4 Comparison of the WSTMC and competing models using C-index (standard deviations)

		NSBCD	Lung	DBCD	DLBCL	MCL
COX based	COX	0.4411 (0.0589)	0.5158 (0.1333)	0.5539 (0.1233)	0.4553 (0.0718)	0.5773 (0.0591)
	LASSO-COX	0.5910 (0.1086)	0.6698 (0.0910)	0.6880 (0.0429)	0.6344 (0.0421)	0.6824 (0.0701)
	EN-COX	0.6046 (0.1000)	0.6652 (0.0702)	0.7214 (0.0306)	0.6488 (0.0394)	0.6734 (0.0733)
	Cox- $l_{2,1}$	0.7453 (0.0742)	0.7470 (0.0450)	0.7548 (0.0640)	0.6499 (0.0474)	0.7229 (0.0379)
	Cox-Trace	0.7550 (0.0737)	0.7348 (0.0431)	0.6946 (0.0576)	0.6478 (0.0387)	0.7127 (0.0902)
Parametric models	Logistic	0.3787 (0.0195)	0.5714 (0.0596)	0.4908 (0.0872)	0.4840 (0.0496)	0.4827 (0.0682)
	Weibull	0.3045 (0.1528)	0.4287 (0.1023)	0.4555 (0.1046)	0.2507 (0.0627)	0.4735 (0.0747)
	Log-gaussian	0.4435 (0.0539)	0.4122 (0.0754)	0.4875 (0.0553)	0.3167 (0.0914)	0.2564 (0.0715)
	Log-logistic	0.2378 (0.0500)	0.5924 (0.0655)	0.5257 (0.0232)	0.4246 (0.1243)	0.4802 (0.0724)
Linear models	OLS	0.6333 (0.1108)	0.5743 (0.0658)	0.5690 (0.0744)	0.5024 (0.1023)	0.5007 (0.1059)
	Tobit	0.3733 (0.0214)	0.4689 (0.1358)	0.4869 (0.0762)	0.4969 (0.0527)	0.4591 (0.0322)
	RWRSS	0.6766 (0.1277)	0.6969 (0.0430)	0.7216 (0.0446)	0.6265 (0.0657)	0.7118 (0.0737)
Multi-task based	Multi-LASSO	0.6117 (0.1493)	0.4410 (0.1655)	0.6256 (0.0749)	0.6104 (0.0512)	0.6539 (0.0140)
	Multi- $l_{2,1}$	0.6100 (0.1700)	0.5248 (0.1130)	0.6899 (0.0720)	0.6115 (0.0512)	0.6912 (0.0602)
	MTLSA.V2	0.6858 (0.0834)	0.6769 (0.0271)	0.7515 (0.0625)	0.6545 (0.0600)	0.7079 (0.0963)
	MTLSA	0.6820 (0.0446)	0.6327 (0.0753)	0.7581 (0.0304)	0.6527 (0.0713)	0.7274 (0.1257)
	NLMC	0.6827 (0.1415)	0.6939 (0.1500)	0.7563 (0.0565)	0.6178 (0.0702)	0.7232 (0.1035)
	MTMC	0.7620 (0.0576)	0.6958 (0.0217)	0.4292 (0.0660)	0.6611 (0.0491)	0.7223 (0.0284)
	WSTMC	0.7970 (0.0135)	0.8153 (0.0992)	0.7705 (0.0562)	0.6810 (0.0571)	0.7336 (0.0697)

的 WSTMC-nT 模型. 具体地, 我们通过禁用 WSTMC 模型中 MoG 噪声容错机制、多任务直推式特征选择机制及时序稳定性机制中的任意一种来验证这些机制在模型中所起到的实际效果, 表 6 报告了两种评价指标 C-index 和 Weighted average AUC 上 5 种相关模型 MTMC、WSTMC-nM、WSTMC-nT、WSTMC-nF 以及 WSTMC 的性能. 实验结果表明: 1) 相比于传统的 MTMC 模型, 三种消融模型 WSTMC-nM、WSTMC-nF 和 WSTMC-nT 在两种评价指标 C-index 和 Weighted average AUC 上都无一例外地优于 MTMC, 这表明引入 MoG 噪声容错机制、多任务直推式特征选择

机制以及时序稳定性机制是有助于提高传统 MTMC 模型性能的; 2) 相比于同时引入 3 种机制的 WSTMC 模型, 消融模型 WSTMC-nM、WSTMC-nF 和 WSTMC-nT 在两种评价指标 C-index 和 Weighted average AUC 上性能均弱于 WSTMC, 进一步验证了 3 种机制的同时引入优于任意两种机制的组合引入, 从而实验上表明我们所提出的 WSTMC 模型中 3 种机制的引入是合理有效的.

5 总结与讨论

针对生存分析领域通常面临的高维小样本和噪声敏感问题, 本文提出了一类新颖的噪声容错弱监

表 5 所提出的 WSTMC 模型和其他比较模型在 Weighted average AUC 指标上的性能比较 (标准差)
Table 5 Comparison of the WSTMC and competing models using Weighted average AUC (standard deviations)

		NSBCD	Lung	DBCD	DLBCL	MCL
COX based	Cox	0.4611 (0.1893)	0.5464 (0.1632)	0.5334 (0.1620)	0.4480 (0.1079)	0.4695 (0.1701)
	LASSO-COX	0.5986 (0.1589)	0.7499 (0.1780)	0.7068 (0.0292)	0.7104 (0.0533)	0.7401 (0.0166)
	EN-COX	0.6479 (0.0970)	0.7540 (0.1398)	0.7494 (0.0189)	0.7260 (0.0618)	0.7350 (0.0025)
	Cox- $l_{2,1}$	0.7752 (0.0450)	0.8079 (0.0462)	0.7545 (0.0365)	0.7157 (0.0795)	0.8215 (0.0737)
	Cox-Trace	0.6729 (0.0883)	0.7074 (0.0455)	0.7078 (0.0465)	0.6768 (0.0903)	0.7197 (0.0209)
Parametric models	Logistic	0.4597 (0.1742)	0.6301 (0.0924)	0.4840 (0.1086)	0.5011 (0.0489)	0.2986 (0.0501)
	Weibull	0.4575 (0.2622)	0.4379 (0.1018)	0.4707 (0.0809)	0.4320 (0.1080)	0.3240 (0.0484)
	Log-gaussian	0.4992 (0.2378)	0.4182 (0.0680)	0.4742 (0.0763)	0.4270 (0.0977)	0.4457 (0.0161)
	Log-logistic	0.3304 (0.1057)	0.5822 (0.1544)	0.5302 (0.0298)	0.4712 (0.0627)	0.2983 (0.0505)
Linear models	OLS	0.6599 (0.1042)	0.5677 (0.1120)	0.5998 (0.1096)	0.4934 (0.1952)	0.5594 (0.1191)
	Tobit	0.4567 (0.1812)	0.4708 (0.1422)	0.4668 (0.1021)	0.5243 (0.0691)	0.5074 (0.0283)
	RWRSS	0.7016 (0.1369)	0.6821 (0.0840)	0.6928 (0.0183)	0.5622 (0.0127)	0.7056 (0.1367)
Multi-task based	Multi-LASSO	0.6495 (0.1226)	0.4410 (0.1655)	0.6402 (0.0572)	0.5876 (0.1047)	0.6079 (0.0696)
	Multi- $l_{2,1}$	0.6501 (0.1314)	0.5589 (0.1486)	0.7125 (0.0775)	0.6001 (0.0528)	0.6476 (0.0653)
	MTLSA.V2	0.6822 (0.0576)	0.8076 (0.0559)	0.7569 (0.0645)	0.7405 (0.0719)	0.7639 (0.0651)
	MTLSA	0.7032 (0.0427)	0.7169 (0.0964)	0.8003 (0.0425)	0.7385 (0.0638)	0.8095 (0.0367)
	NLMC	0.5724 (0.0705)	0.5842 (0.0994)	0.6212 (0.0687)	0.6130 (0.0657)	0.7175 (0.0664)
	MTMC	0.8206 (0.0929)	0.6035 (0.1422)	0.4334 (0.0506)	0.6989 (0.0351)	0.8255 (0.0729)
	WSTMC	0.8662 (0.0788)	0.8629 (0.0519)	0.8007 (0.0549)	0.7064 (0.0563)	0.8430 (0.0767)

表 6 在两种评价指标 C-index 和 Weighted average AUC 上的消融性实验性能比较 (标准差)
Table 6 Comparison of the ablation experiments using C-index and Weighted average AUC (standard deviations)

		NSBCD	Lung	DBCD	DLBCL	MCL
C-index	MTMC	0.7620 (0.0576)	0.6958 (0.0217)	0.4292 (0.0660)	0.6611 (0.0491)	0.7223 (0.0284)
	WSTMC-nM	0.7633 (0.0406)	0.7053 (0.1566)	0.6847 (0.0454)	0.6661 (0.0670)	0.7241 (0.1023)
	WSTMC-nT	0.7642 (0.0650)	0.7345 (0.0767)	0.7270 (0.0422)	0.6659 (0.0529)	0.7234 (0.0729)
	WSTMC-nF	0.7664 (0.0164)	0.7293 (0.1086)	0.7123 (0.0586)	0.6641 (0.0497)	0.7273 (0.0934)
	WSTMC	0.7970 (0.0135)	0.8153 (0.0992)	0.7705 (0.0562)	0.6810 (0.0571)	0.7336 (0.0697)

表 6 在两种评价指标 C-index 和 Weighted average AUC 上的消融性实验性能比较 (标准差) (续表)

Table 6 Comparison of the ablation experiments using C-index and Weighted average AUC (standard deviations) (continued)

	NSBCD	Lung	DBCD	DLBCL	MCL
MTMC	0.8206 (0.0929)	0.6035 (0.1422)	0.4334 (0.0506)	0.6989 (0.0351)	0.8255 (0.0729)
WSTMC-nM	0.8547 (0.0441)	0.6674 (0.0777)	0.6353 (0.0467)	0.6994 (0.0526)	0.8256 (0.1488)
Weighted average AUC	0.8557 (0.0447)	0.7676 (0.0531)	0.6061 (0.0726)	0.6998 (0.0535)	0.8334 (0.1075)
WSTMC-nT	0.8421 (0.0915)	0.7420 (0.0433)	0.6560 (0.0435)	0.7053 (0.0467)	0.8268 (0.0230)
WSTMC-nF	0.8662 (0.0788)	0.8629 (0.0519)	0.8007 (0.0549)	0.7064 (0.0563)	0.8430 (0.0767)

督直推式矩阵补全 (WSTMC) 模型. 首先将原始的生存分析问题建模成一类传统的多任务直推式矩阵补全 (MTMC) 模型, 然后引入高斯混合分布来拟合数据中未知的复杂噪声. 同时为了缓解高维小样本所带来的过拟合缺陷, 我们还设计了一类适用于去噪特征空间的多任务直推式特征选择机制, 以期筛选出跨任务共享的判别特征, 从而提高模型的泛化性能. 相较于传统的 MTMC 模型, 我们提出的 WSTMC 模型一定程度上克服了 MTMC 模型噪声容错性差、泛化性能不足以及先验信息融合性不强的缺陷. 最后, 多个真实数据集上的实验结果证实了所提出的 WSTMC 模型优于其他广泛使用的生存分析方法.

然而, 我们的方法也仍然存在改进的空间, 比如无论是传统的 MTMC 模型还是我们的 WSTMC 模型都是基于实例特征与实例标记的线性依赖假设, 这种假设与实际情形未必相符, 如果能将我们的模型推广至非线性假设情形, 将极有可能提升生存分析的预测性能. 我们将在后续研究中尝试采用核化的方法来放宽已有模型的线性依赖假设, 并同时尝试引入其他类型的非线性依赖实现机制.

References

- Li Y, Wang L, Zhou J Y, Ye J P. Multi-task learning based survival analysis for multi-source block-wise missing data. *Neurocomputing*, 2019, **364**(3): 95–107
- Emmert-Streib F, Dehmer M. Introduction to survival analysis in practice. *Machine Learning and Knowledge Extraction*, 2019, **1**(3): 1013–1038
- Yang W S, Huang T, Zeng J L, Tang Y, Chen L J, Michra S, Liu Y E. Purchase prediction in free online games via survival analysis. In: Proceedings of the 2019 IEEE International Conference on Big Data. Los Angeles, USA: IEEE, 2019. 4444–4449
- Efron B. The efficiency of Cox's likelihood function for censored data. *Journal of American Statistical Association*, 1977, **72**(359): 557–565
- Crowther M J, Lambert P C. A general framework for parametric survival analysis. *Statistics in Medicine*, 2014, **33**(30): 5280–5297
- Liu Hui-Ting, Leng Xin-Yang, Wang Li-Li, Zhao Peng. A joint embedded multi-label classification algorithm. *Acta Automatica Sinica*, 2019, **45**(10): 1969–1982 (刘慧婷, 冷新杨, 王利利, 赵鹏. 联合嵌入式多标签分类算法. 自动化学报, 2019, **45**(10): 1969–1982)
- Li Y, Wang J P, Ye J P, Reddy C K. A multi-task learning formulation for survival analysis. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA: ACM, 2016. 1715–1724
- Goldberg A B, Zhu X J, Recht B, Xu J M, Nowak R. Transduction with matrix completion: Three birds with one stone. In: Proceedings of the 23rd International Conference on Neural Information Processing Systems. Vancouver, British Columbia, Canada: MIT Press, 2010. 757–765
- Maz Y, Schmidt G. On approximate approximations using Gaussian kernels. *IMA Journal of Numerical Analysis*, 1996, **16**(1): 13–29
- Tibshirani R. The LASSO method for variable selection in the Cox model. *Statistics in Medicine*, 1997, **16**(4): 385–395
- Simon N, Hastie T, Tibshirani R. Regularization paths for Cox's proportional hazards model via coordinate descent. *Journal of Statistical Software*, 2011, **39**(5): 1–13
- Vinzamuri B, Reddy C K. Cox regression with correlation based regularization for electronic health records. In: Proceedings of the 13th IEEE International Conference on Data Mining, Dallas, USA: IEEE, 2013. 757–766
- Wang P, Li Y, Reddy C K. Machine learning for survival analysis: A survey. *ACM Computing Surveys*, 2019, **51**(6): 1–36
- Tobin J. Estimation of relationships for limited dependent variables. *Econometrica*, 1958, **26**(1): 24–36
- Buckley J, James I. Linear regression with censored data. *Biometrika*, 1979, **66**(3): 429–436
- Boyd S. Convex optimization. *IEEE Transactions on Automatic Control*, 2006, **51**(11): 1859–1859
- Chen Lei, Yang Geng, Chen Zheng-Yu, Xiao Fu, Chen Song-Can. Linearized Bregman iteration algorithm for matrix completion with structural noise. *Chinese Journal of Computers*, 2015, **38**(7): 1357–1371 (陈蕾, 杨庚, 陈正宇, 肖甫, 陈松灿. 基于线性Bregman迭代的结构化噪声矩阵补全算法. 计算机学报, 2015, **38**(7): 1357–1371)
- Chen Lei, Yang Geng, Chen Zheng-Yu, Xiao Fu, Xu Jian. Web services QoS prediction via matrix completion with structural noise. *Journal on Communications*, 2015, **36**(6): 49–59 (陈蕾, 杨庚, 陈正宇, 肖甫, 许建. 基于结构化噪声矩阵补全的Web服务QoS预测. 通信学报, 2015, **36**(6): 49–59)
- Lian Qiu-Sheng, Fu Li-Peng, Chen Shu-Zhen, Shi Bao-Shun. A

- compressed sensing algorithm based on multiscale residual reconstruction network. *Acta Automatica Sinica*, 2019, **45**(11): 2082–2091
(练秋生, 富利鹏, 陈书贞, 石保顺. 基于多尺度残差网络的压缩感知重构算法. *自动化学报*, 2019, **45**(11): 2082–2091)
- 20 Wang Chuan-Yun, Qin Shi-Yin. Background modeling of infrared image in dynamic scene with Gaussian mixture model in compressed sensing domain. *Acta Automatica Sinica*, 2018, **44**(7): 1212–1226
(王传云, 秦世引. 动态场景红外图像的压缩感知域高斯混合背景建模. *自动化学报*, 2018, **44**(7): 1212–1226)
- 21 Liu Zhou-Zhou, Li Shi-Ning, Wang Hao, Zhang Qian-Yun. A compressed sensing reconstruction algorithm based on elastic collision and gradient pursuit strategy for WSNs. *Acta Automatica Sinica*, 2020, **46**(1): 178–192
(刘洲洲, 李士宁, 王皓, 张倩昀. 联合弹性碰撞与梯度追踪的WSNs压缩感知重构. *自动化学报*, 2020, **46**(1): 178–192)
- 22 Candès E J, Recht B. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 2009, **9**(6): 717–772
- 23 Fazel M. Matrix rank minimization with applications [Ph.D. dissertation], Stanford University, USA, 2002.
- 24 Cao X Y, Xu Z B, Meng D Y. Spectral-spatial hyperspectral image classification via robust low-rank feature extraction and Markov random field. *Remote Sensing*, 2019, **11**(13): 1–18
- 25 Chen X A, Han Z, Wang Y, Zhao Q, Meng D Y, Lin L, Tang Y D. A generalized model for robust tensor factorization with noise modeling by mixture of Gaussians. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, **29**(11): 5380–5393
- 26 Xu Y Y, Yin W T. A globally convergent algorithm for nonconvex optimization based on block coordinate update. *Journal of Scientific Computing*, 2017, **72**(2): 700–734
- 27 Zhou J Y, Chen J H, Ye J P. MALSAR: Multi-task learning via structural regularization [Online], available: <http://jiayuzhou.github.io/MALSAR/>, March 17, 2020
- 28 Alameda X, Ricci E, Yan Y, Sebe N. Recognizing emotions from abstract paintings using non-linear matrix completion. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 5240–5248
- 29 Sørlie T, Tibshirani R, Parker J. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences of the United States of America*, 2003. **100**(14): 8418–8423
- 30 Li Y, Xu K S, Reddy C K. Regularized parametric regression for high-dimensional survival analysis. In: Proceedings of the 2016 SIAM International Conference on Data Mining. Miami, USA: SIAM, 2016. 765–773
- 31 Beer D G, Kardia S, Huang C C. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature Medicine*, 2002, **8**(8): 816–824
- 32 Van Houwelingen H C, Bruinsma T, Wessels L F. Cross-validated Cox regression on microarray gene expression data. *Statistics in Medicine*, 2006, **25**(18): 3201–3216
- 33 Rosenwald A, Wright G, Wiestner A. The proliferation gene expression signature is a quantitative integrator of oncogenic events that predicts survival in mantle cell lymphoma. *Cancer Cell*, 2003, **3**(2): 185–197
- 34 Wang D, Wang C C, Xiao J H, Xiao Z, Chen W W, Havyarimana V. Bayesian optimization of support vector machine for regression prediction of short-term traffic flow. *Intelligent Data Analysis*, 2019, **23**(2): 481–497
- 35 Thung K H, Yap P T, Adeli E, Lee S W, Shen D G. Conversion and time-to-conversion predictions of mild cognitive impairment using low-rank affinity pursuit denoising and matrix completion. *Medical image analysis*, 2018, **45**(2): 68–82
- 36 Cui Jia-Xu, Yang Bo. Survey on Bayesian optimization methodology and applications. *Ruan Jian Xue Bao/Journal of Software*, 2018, **29**(10): 3068–3090
(崔佳旭, 杨博. 贝叶斯优化方法和应用综述. *软件学报*, 2018, **29**(10): 3068–3090)
- 37 Wang L, Li Y, Zhou J Y, Zhu D X, Ye J P. Multi-task survival analysis. In: Proceedings of the 2017 IEEE International Conference on Data Mining (ICDM). New Orleans, USA: IEEE, 2017. 485–494
- 38 Faraway J J. Practical regression and ANOVA using R [Online], available: <https://people.bath.ac.uk/jjf23/book/practical.pdf>, March 17, 2020



陈蕾 南京邮电大学教授. 主要研究方向为大规模机器学习, 基于医学影像的脑疾病分析. 本文通信作者.

E-mail: chenlei@njupt.edu.cn

(CHEN Lei Professor at Nanjing University of Posts and Telecommunications. His research interest

covers large-scale machine learning, and brain diseases analysis based on medical imaging. Corresponding author of this paper.)



邵楷 南京邮电大学计算机学院硕士研究生. 主要研究方向为机器学习.

E-mail: sk1017041210@163.com

(SHAO Kai Master student at Nanjing University of Posts and Telecommunications. His main research interest is machine learning.)



林腾涛 南京邮电大学计算机学院硕士研究生. 主要研究方向为机器学习.

E-mail: ltt1995711@126.com

(LIN Teng-Tao Master student at Nanjing University of Posts and Telecommunications. His main research interest is machine learning.)



陈兴国 南京邮电大学讲师. 主要研究方向为机器学习, 强化学习, 智能游戏. E-mail: chenxg@njupt.edu.cn

(CHEN Xing-Guo Lecturer at Nanjing University of Posts and Telecommunications. His research interest covers machine learning, reinforcement learning, and intelligent games.)