

控制与决策

Control and Decision

基于分解排序的多维分类器链算法

李二超, 杨宏强

引用本文:

李二超, 杨宏强. 基于分解排序的多维分类器链算法[J]. 控制与决策, 2025, 40(7): 2223-2232.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2024.1344>

您可能感兴趣的其他文章

Articles you may be interested in

面向全局搜索的自适应领导者樽海鞘群算法

Global search-oriented adaptive leader salp swarm algorithm

控制与决策. 2021, 36(9): 2152-2160 <https://doi.org/10.13195/j.kzyjc.2020.0090>

行人重识别中度量学习方法研究进展

A survey on metric learning in person re-identification

控制与决策. 2021, 36(7): 1547-1557 <https://doi.org/10.13195/j.kzyjc.2020.0801>

基于地标特征和元学习方法推荐最适用优化算法

Recommending best suitable metaheuristic based on landmarking feature and meta-learning approach

控制与决策. 2021, 36(5): 1223-1231 <https://doi.org/10.13195/j.kzyjc.2019.0993>

基于相异性度量选取初始聚类中心改进的K-means聚类算法

Improved K-means clustering algorithm for selecting initial clustering centers based on dissimilarity measure

控制与决策. 2021, 36(12): 3083-3090 <https://doi.org/10.13195/j.kzyjc.2020.0554>

l_p -范数约束下MKL-OC-ELM的装备故障检测

MKL and OC-ELM fault detection based on l_p -norm constraint

控制与决策. 2021, 36(10): 2379-2388 <https://doi.org/10.13195/j.kzyjc.2020.0443>

基于分解排序的多维分类器链算法

李二超[†], 杨宏强

(兰州理工大学 电气工程与信息工程学院, 兰州 730050)

摘要: 多维分类问题中的类别变量具有复杂的依赖关系, 这对分类性能有着重要影响. 分类器链算法能够有效地建模这些依赖关系, 但由于标签顺序选择和错误传播问题, 性能提升受限. 为此, 提出一种基于分解排序的多维分类器链算法. 首先, 通过一对一分解规则将多维分类问题转化为多个二类分类问题, 以降低问题的复杂度; 其次, 将标签顺序建模为线性排序问题, 并利用遗传算法进行优化, 确保标签排序的合理性; 最后, 通过控制特征空间策略减弱前序分类器的错误预测对后续分类器的负面影响, 从而提高算法的鲁棒性. 在 10 个真实的多维分类数据集上进行的对比实验表明, 所提出的算法在泛化性能上优于当前先进的多维分类算法, 同时具有较低的计算复杂度.

关键词: 机器学习; 多维分类; 分类器链; 一对一分解; 标签顺序; 特征空间

中图分类号: TP181 文献标志码: A

DOI: 10.13195/j.kzyjc.2024.1344

引用格式: 李二超, 杨宏强. 基于分解排序的多维分类器链算法 [J]. 控制与决策, 2025, 40(7): 2223-2232.

Decomposition and ranking-based classifier chain for multi-dimensional classification

LI Er-chao[†], YANG Hong-qiang

(College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China)

Abstract: Classification performance is significantly affected by dependencies between class variables in multi-dimensional classification. These dependencies are effectively modeled by classifier chain algorithms. However, their performance is constrained by issues such as label order selection and error propagation. To address these limitations, this paper introduces a decomposition and ranking-based classifier chain for multi-dimensional classification algorithm. Initially, the multi-dimensional classification problem is simplified into binary classification problems using a one-vs-one strategy, which reduces complexity. Subsequently, the label order is treated as a linear ordering problem and optimized with the genetic algorithm to determine the optimal sequence. Finally, a feature space control strategy is proposed to minimize the impact of early classification errors on subsequent classifiers. Experiments conducted on 10 real-world datasets demonstrate that the proposed algorithm outperforms the state-of-the-art methods while also exhibiting lower computational complexity.

Keywords: machine learning; multi-dimensional classification; classifier chain; one-vs-one decomposition; label ordering; feature space

0 引言

多类分类 (MCC) 是机器学习^[1] 领域中一种重要的学习范式, 其核心假设是每个数据对象仅与一个类别变量相关联. 然而, 在许多实际应用中, 对象的语义具有多维特征. 在服装分类中, 每件衣物都具有多个维度的特征, 例如款式 (包括 T 恤、衬衫、外套等)、颜色 (包括蓝色、白色、黑色等)、材质 (包括棉、丝绸、聚酯纤维等). 显然, 单一类别变量的多类

分类难以完整刻画对象的复杂语义. 为了更有效地描述包含多维语义特征的对象, 一种合理的解决方案是将每个对象与多个类别变量关联. 研究者们将此类学习范式称为多维分类 (MDC)^[2].

在多维分类中, 每个对象由一个特征向量表示, 并在多个类别变量上具有对应的标签, 每个类别变量构成一个类别空间, 从不同的维度描述对象的语义信息. 实际上, 多维分类的需求广泛存在于文本分

收稿日期: 2024-11-18; 录用日期: 2025-01-07.

基金项目: 国家自然科学基金项目 (62063019, 62306131).

责任编委: 巩敦卫.

[†]通信作者. E-mail: lecstarr@163.com.

类^[3-4]、计算机视觉^[5-6]、生物信息学^[7]、生态信息学^[8]和资源分配^[9]等实际应用中。

为了解决多维分类问题,已有研究提出了多种方法,其中较为直接的方法包括 Binary Relevance (BR)^[10]和 Class Powerset (CP)^[11]。BR方法将多维分类问题转化为多个独立的多类分类问题,其缺点在于忽略了类别变量之间潜在的依赖关系;CP方法则将整个输出空间的每一种类别组合视为一个新的类别,将多维分类问题转化为多类分类问题。然而CP方法计算复杂度较高,且在训练集中未出现的类别组合难以预测,可能会导致过拟合。这两种方法的局限性突显出考虑类别变量之间依赖关系的重要性。现有的一些建模类别变量之间依赖关系的方法包括:将原类别空间划分为若干超类(SC),并将关联性强的类别空间合并构成新的类别空间^[2];在类别变量与特征变量之间构建一个有向无环图(DAG)^[12-13];利用协方差正则化技术和标签编码技术捕捉类别变量间的依赖关系^[14-16]。

近年来,分类器链(CC)方法受到越来越多研究者的关注,并已成功应用于多标签分类(MLC)^[17]。CC方法为每个标签训练一个二类分类器,链中后续分类器的预测基于前序分类器的输出,以此建立标签之间的依赖关系。CC方法结构简单,计算效率高,已被证明是多标签分类最有效的解决方法之一^[18]。然而,直接将CC方法应用于多维分类时,分类效果难以与最先进的多维分类方法相比,其局限可能源于:1)输出变量的特性不同,MDC涉及多类型输出,而MLC为二类型;2)链序随机性,不同的链序可能影响模型的泛化性能;3)链上前序分类器的错误预测可能会传递给后续分类器,导致错误传播。尽管已有研究尝试解决输出变量特性问题,但由于采用随机链序集成策略,导致计算复杂度较高^[19]。

针对上述问题,本文提出基于分解排序的多维分类器链算法(DRCC)。本文的主要贡献如下:

1) 利用一对一分解策略将多维分类任务转化为多个二类分类任务,从而降低问题的复杂度,提高分类效率。

2) 提出将二类分类器的链序问题建模为线性排序问题,并利用遗传算法优化链序结构,解决传统分类器链方法中的标签顺序选择问题。

3) 提出基于分类器预测准确率的特征空间优化策略,对于低准确率的分器,限制其预测标签参与特征更新,有效降低错误传播的影响,提高算法的鲁棒性。

本文的剩余部分结构安排如下:第1节简要介

绍相关工作,第2节介绍DRCC的技术细节,第3节展示对比实验结果并进行分析,第4节对全文进行总结与展望。

1 相关工作

1.1 多维分类的定义

假设 $\mathcal{X} = \mathbb{R}^d$ 表示 d 维特征构成的输入空间, $\mathcal{Y} = C_1 \times C_2 \times \dots \times C_q$ 表示由 q 个类别空间的笛卡尔乘积构成的输出空间。其中,每个类空间 $C_j (1 \leq j \leq q)$ 由 K_j 个可能的类标签构成,即 $C_j = \{c_1^j, c_2^j, \dots, c_{K_j}^j\}$ 。给定多维分类数据集 $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i) | 1 \leq i \leq m\}$,有 m 个训练样例,对于每个训练样例 $(\mathbf{x}_i, \mathbf{y}_i) \in \mathcal{D}$, $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{id}]^T \in \mathcal{X}$ 表示 d 维特征向量, $\mathbf{y}_i = [y_{i1}, y_{i2}, \dots, y_{iq}]^T \in \mathcal{Y}$ 表示与 \mathbf{x}_i 相关联的 q 维类别向量。其中, y_{ij} 表示从类别空间 C_j 中取一个可能的类标签,即 $y_{ij} \in C_j$ 。多维分类的任务是从数据集 \mathcal{D} 中学习一个预测模型 $f: \mathcal{X} \mapsto \mathcal{Y}$,该模型可以预测未见示例 \mathbf{x}_* 的类别向量 $f(\mathbf{x}_*) \in \mathcal{Y}$ 。表1进一步总结了这些符号。

表1 多维分类定义中的符号

符号	描述
m	训练样例个数
d	输入空间的特征个数
q	输出空间的类别空间个数(即维度个数)
K_j	第 $j (1 \leq j \leq q)$ 个类别空间包含的类标签个数
\mathcal{X}	d 维输入(特征)空间,即 $\mathcal{X} = \mathbb{R}^d$
\mathcal{Y}	输出空间,其中 $\mathcal{Y} = C_1 \times C_2 \times \dots \times C_q$
C_j	第 j 个类别空间,其中 $C_j = \{c_1^j, c_2^j, \dots, c_{K_j}^j\}$
c_a^j	C_j 中的第 $a (1 \leq a \leq K_j)$ 个类标签
\mathcal{D}	多维分类训练集,其中 $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i) 1 \leq i \leq m\}$
\mathbf{x}_i	第 i 个训练样例的特征向量
\mathbf{y}_i	第 i 个训练样例的类别向量
\mathbf{x}_*	未见示例,其中 $\mathbf{x}_* \in \mathcal{X}$
f	多维分类预测模型: $\mathcal{X} \mapsto \mathcal{Y}$, 其中 $f(\mathbf{x}_*) \in \mathcal{Y}$

1.2 分类器链

分类器链(CC)算法最初是在多标签分类问题中提出的^[17],它通过设定标签的排列顺序,将各标签按此顺序构成链式结构,使后续标签可以利用前序标签的预测结果。CC算法可以看作是BR算法^[10]的改进版本,相较于BR方法,CC算法有效地建立了标签之间的依赖关系,在多标签分类中能够取得更好的性能。然而,CC算法的实现需要预先设定标签的排列顺序,而合理的标签顺序往往难以预先确定。在多标签分类问题中,标签的数量庞大,排列组合的可能性极多,导致确定标签顺序十分困难,且排序不

恰当时可能会严重影响模型的分类性能。

针对这一问题, 研究者们提出了多种 CC 的改进方法来优化标签顺序, 主要可分为两大类。

1) 单链策略. 这类方法主要集中在寻找最佳的标签顺序, 旨在解决 CC 算法的标签顺序随机性带来的问题. Dembczyński 等^[20] 提出了概率分类器链算法 (PCC), 它基于估计标签的联合分布来提升性能, 但该算法计算复杂度较高; Liu 等^[21] 提出基于动态规划的分类器链算法 (CC-DP), 利用动态规划和贪婪搜索来优化标签顺序, 但是计算复杂度显著增加; Zaragoza 等^[22] 设计的贝叶斯分类器链算法 (BCC) 采用贝叶斯网络建模标签间的树形依赖结构, 同样面临着计算复杂度的问题, 且其对贝叶斯网络结构的假设可能限制了其在某些数据分布上的适用性; Xie 等^[23] 设计了基于条件熵的分类器链算法 (EbCC), 该算法使用条件熵确定标签排序, 将高熵标签置于低熵标签之后, 从而提升分类性能并降低计算成本; Ding 等^[24] 提出基于关联规则的分类器链算法 (ARECC) 借助关联规则构建标签依赖图, 通过拓扑排序得到线性标签序列, 有效提高了分类性能. 此外, 亦有研究尝试利用蒙特卡罗技术^[25] 来优化标签顺序。

2) 多链策略. 这类方法构建多条分类器链, 每条链使用不同的标签顺序, 并通过集成策略来增强模型的鲁棒性. 集成分类器链 (ECC)^[17] 是一种典型的多链策略方法, 它通过生成多个标签顺序并集成各条链的结果, 从而提升模型性能. 尽管 ECC 在一定程度上缓解了标签顺序的选择带来的性能限制, 但在标签数目庞大的情况下, 排列组合的数量迅速增加, 这使得标签顺序的选择依然充满挑战. Freitas 等^[26] 提出了一种基于决策模板的 ECC 方法, 为每个标签生成决策模板进行判定, 这种方法适合大规模样本数据集, 但依然未能解决 ECC 算法在计算复杂度上的普遍问题。

受针对多维分类的基于分解的分类器链方法 (DCC)^[19] 的启发, 本文基于单链策略提出一种改进的多维分类器链算法. 该算法通过合理排序标签对, 充分利用标签对之间的关联性, 提升分类性能。

2 DRCC 方法

2.1 一对一分解

为了通过构建二类分类器链来解决多维分类问题, 需要将多维分类问题转化为多个二类分类问题^[19]. 现有研究已经证明了一对一 (OvO) 分解是将多类分类问题转化为二类分类的最优方案^[27], 因此本文提出的 DRCC 方法在多维分类的每一维度上通过一对

一分解将其转化成二类分类问题. 易知多维分类问题一共可以分解为 $N = \sum_{j=1}^q C_{K_j}^2$ 个二类分类问题. 假设第 n ($1 \leq n \leq N$) 个分解得到的二类分类问题对应的数据集为 $\mathcal{D}_b^n = \{(\mathbf{x}_j^n, l_j^n) | j \in \mathcal{I}_+^n \cup \mathcal{I}_-^n\}$. 其中: $\mathbf{x}_j^n \in \mathcal{X}$; \mathcal{I}_+^n 和 \mathcal{I}_-^n 是训练样例的索引集, 分别对应于类标签对 (A^n, B^n) 中的 A^n 和 B^n , 当 A^n 为相关标签时 $l_j^n = +1$, B^n 为相关标签时 $l_j^n = -1$ ^[19]. 图 1 直观地展示了一对一分解的过程, 标签为 0 表示该样例不用于训练相应的二类分类器。

input	C_1	C_2	C_3	input	(a,b)	(a,c)	(b,c)	(a,β)	(a,γ)	(β,γ)	(I,II)
x_1	a	α	I	x_1	1	1	0	1	1	0	1
x_2	b	γ	I	x_2	-1	0	1	0	-1	-1	1
x_3	c	γ	II	x_3	0	-1	-1	0	-1	-1	-1
x_4	b	β	I	x_4	-1	0	1	-1	0	1	1
x_5	a	β	II	x_5	1	1	0	-1	0	1	-1

图1 一对一分解示意图

2.2 优化标签对顺序

2.2.1 条件熵和互信息

在信息论中, 熵是接收的每条消息中包含信息的平均量. 消息表示来自分布或数据流中的事件、样本或特征. 熵越高, 能传递的信息越多; 熵越低, 能传递的信息越少. 设一个离散随机变量 X 可以从集合 $\{x_1, x_2, \dots, x_n\}$ 中取值, 则 X 的熵为

$$H(X) = - \sum_{i=1}^n p(x_i) \log_2 p(x_i), \quad (1)$$

其中 $p(x_i)$ 表示 X 取值为 x_i 的概率。

设离散随机变量 X 和 Y 分别从集合 $\{x_1, x_2, \dots, x_n\}$ 和 $\{y_1, y_2, \dots, y_m\}$ 中取值, 则 X 和 Y 的联合熵为

$$H(X, Y) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) \log_2 p(x_i, y_j), \quad (2)$$

其中 $p(x_i, y_j)$ 表示 X 取值为 x_i 时 Y 取值为 y_j 的概率。

条件熵衡量的是一个随机变量在给定另一个随机变量的值时的不确定性, 离散随机变量 X 在另一个离散随机变量 Y 给定时的条件熵为

$$H(X|Y) = H(X, Y) - H(Y). \quad (3)$$

条件熵是非对称的, 即 $H(X|Y) \neq H(Y|X)$. 当 $H(X|Y) > H(Y|X)$ 时, 表示 X 已知时 Y 的不确定性小于 Y 已知时 X 的不确定性, 在分类器链算法中, 不确定性较高的标签应该放在链的尾端, 以减小错误传播的影响。

2.2.2 线性排序问题

线性排序问题 (LOP)^[28] 的任务是寻找一个使给定的目标函数最优化的排列. 形式上, 可以得到一个由标签对之间的条件熵构成的矩阵 $A_{N \times N}$, 目标函数

的值为上三角矩阵中所有元素的和. 改变矩阵行和相应列的顺序会得到不同的上三角矩阵和, LOP 搜索行和相应列的顺序, 使这个和达到最大.

在本文中, N 个标签对有 $N!$ 种可能的排列顺序, 当 N 值较小时, 计算所有排序的目标函数值是可以实现的, 但当 N 值较大时, LOP 是一个 NP 难问题. 遗传算法 (GA) 是求解这类问题的一种很好的选择^[29], 最终求解得到的最优标签对顺序记为 $\Phi(N) = \{\varphi(1), \varphi(2), \dots, \varphi(N)\}$. LOP 用到的条件熵矩阵如表 2 所示, 其中 L_N 表示多维分类问题分解的 N 个二类分类问题对应的类标签对.

表2 条件熵矩阵

	L_1	L_2	\dots	L_N
L_1	0	$H(L_2 L_1)$	\dots	$H(L_N L_1)$
L_2	$H(L_1 L_2)$	0	\dots	$H(L_N L_2)$
\vdots	\vdots	\vdots	\ddots	\vdots
L_N	$H(L_1 L_N)$	$H(L_2 L_N)$	\dots	0

2.2.3 遗传算法

遗传算法 (GA) 通过避免在搜索空间中进行穷举搜索降低了计算复杂度. GA 包括以下 6 个步骤:

1) 初始化种群. 将问题的解进行编码形成初始种群, 每个个体代表一个潜在解. 本文采用排序编码, 并随机生成一个由 50 个个体组成的种群^[30]. 对于本文提出的标签对排序问题, 每个个体的染色体表示为标签对顺序的一个排列. 例如, 假设有 5 个标签对, 分别编码为 1、2、3、4、5. 种群中某个个体的染色体为 [1, 3, 4, 5, 2], 表示按此标签对的顺序训练二类分类器.

2) 适应度评估. 定义一个适应度函数衡量每个个体的质量, 适应度越高表示越接近问题的最优解. 本文的适应度函数定义为上三角条件熵矩阵的元素和, 使之达到最大.

3) 亲本选择. 基于适应度值挑选个体作为下一代亲本. 本文使用基于精英适应度的选择策略, 精英个体设置为 5, 种群剩余个体使用轮盘赌选择.

4) 交叉. 以指定的交叉概率, 在选定的亲本个体之间交换遗传物质以产生新的后代. 本文使用顺序交叉, 交叉概率为 0.9. 顺序交叉的操作如图 2 所示. 值得注意的是, 在排序编码和顺序交叉的模式下, 由于每个基因在交叉过程中保持唯一性, 交叉后的个体不会产生不可行解.

5) 变异. 变异会随机改变个体的基因, 为种群引入多样性并有助于避免陷入局部最优解. 本文使

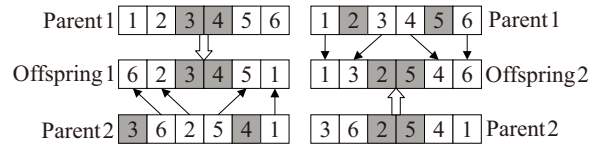


图2 顺序交叉操作

用交换变异, 变异概率为 0.01. 交换变异操作如图 3 所示.

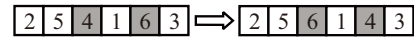


图3 交换变异操作

6) 重复步骤 2) ~ 步骤 5), 直到满足终止条件. GA 算法的流程如图 4 所示.

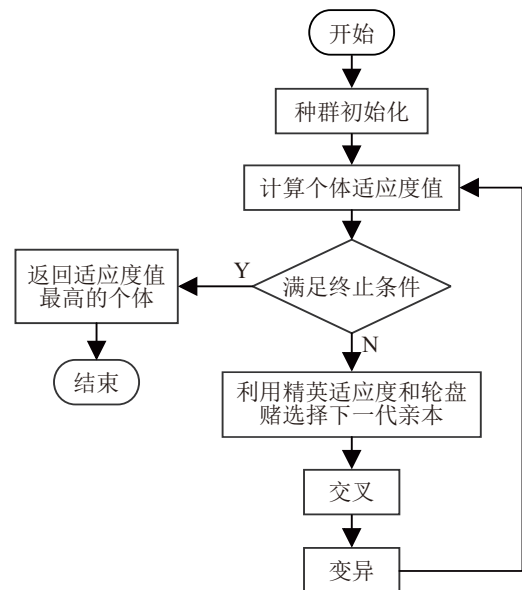


图4 GA 算法流程

2.3 控制特征空间

本文通过 LOP 和 GA 找到最优的二类分类器链序后, 在以下数据集上为每个标签对分别训练二类分类器 $h_{\varphi(n)}$:

$$\mathcal{D}^{\varphi(n)} = \{(\mathbf{x}_j^{\varphi(n)}, l_j^{\varphi(n)}) | j \in \mathcal{I}_+^{\varphi(n)} \cup \mathcal{I}_-^{\varphi(n)}\},$$

$$\text{where } \mathbf{x}_j^{\varphi(n)} = [\mathbf{x}_j; \hat{l}_j^{\varphi(1)}; \dots; \hat{l}_j^{\varphi(n-1)}]. \quad (4)$$

其中: $\hat{l}_j^{\varphi(n)} = h_{\varphi(n)}(\mathbf{x}_j^{\varphi(n)})$, $h_{\varphi(n)} = \mathcal{B}(\mathcal{D}^{\varphi(n)})$, \mathcal{B} 表示采用的二类分类算法.

由于每个标签对的训练数据集不同, 当前二类分类器使用的样本在训练链上前面的二类分类器时可能未使用, 导致训练样本的特征空间中可能缺少预测标签. 本文采用“训练-预测”策略解决这一问题^[19]. 具体来说, 使用已训练好的二类分类器对多维分类训练集中的所有样本进行预测, 以确保训练链上后续二类分类器使用的样本的特征空间中均添加了预测标签.

在测试阶段前, 需考察每个二类分类器的预测

性能, 找出预测准确率低的二类分类器, 本文将这些二类分类器称为“易错分类器”. 在测试集上进行最终预测时, 易错分类器的预测标签不会添加到后续二类分类器的测试样本的特征空间中, 以此减轻错误传播的影响. 设每个二类分类器在预测多维分类训练集所有样本的准确率为 t_i , 预测准确率小于阈值 α 的二类分类器划为易错分类器, 本文选取 $\alpha = 0.1$, 易错分类器集合记为 \mathcal{E} , 有

$$\mathcal{E} = \{L_i | t_i < \alpha, 1 \leq i \leq N\}. \quad (5)$$

在对未见示例 \mathbf{x}_* 进行预测时, 对于第 $\varphi(n)$ ($1 \leq n \leq N$)个分解的二类分类问题, 其预测的类标签由下式确定:

$$\begin{aligned} \hat{l}_*^{\varphi(n)} &= h_{\varphi(n)}(\mathbf{x}_*^{\varphi(n)}), \\ \text{where } \mathbf{x}_*^{\varphi(n)} &= [\mathbf{x}_*; \mathbf{l}_*^{\varphi(n-1)-\mathcal{E}}], \end{aligned} \quad (6)$$

其中 $\mathbf{l}_*^{\varphi(n-1)-\mathcal{E}}$ 表示 $[\hat{l}_*^{\varphi(1)}; \hat{l}_*^{\varphi(2)}; \dots; \hat{l}_*^{\varphi(n-1)}]$ 中剔除易错分类器的预测标签之后的部分. 得到二类分类的预测类标签向量 $\mathbf{l}_* = [\hat{l}_*^1; \hat{l}_*^2; \dots; \hat{l}_*^N]^T$ 后, 可以通过 OvO 解码规则得到未见示例 \mathbf{x}_* 的预测类别向量 $\mathbf{y}_* = [y_{*1}, y_{*2}, \dots, y_{*q}]^T$. 算法 1 总结了 DRCC 方法的工作流程.

算法 1 DRCC 方法.

输入: 多维分类数据集 \mathcal{D} , 采用的二类分类算法 \mathcal{B} , 未见示例 \mathbf{x}_* ;

输出: \mathbf{x}_* 的类别预测向量 \mathbf{y}_* .

step 1: 通过 OvO 分解规则将多维分类问题转化为 $N = \sum_{j=1}^q C_{K_j}^2$ 个二类分类问题;

step 2: 利用分解后的二类分类对应的标签对生成表 2 所示的条件熵矩阵;

step 3: 通过 LOP 和 GA 得到最优标签对顺序 $\Phi(N)$;

step 4: 根据式 (4) 构建二类分类训练集 $\mathcal{D}^{\varphi(n)}$;

step 5: 训练二类分类器 $h_{\varphi(n)} = \mathcal{B}(\mathcal{D}^{\varphi(n)})$;

step 6: 得到二类分类器链;

step 7: 通过预测步骤得到二类分类器的准确率 t_i ;

step 8: 根据式 (5) 得到易错分类器集合 \mathcal{E} ;

step 9: 对未见示例 \mathbf{x}_* 的类别标签进行预测时, 在特征空间中剔除易错分类器 \mathcal{E} 的预测标签;

step 10: 根据式 (6) 确定未见示例 \mathbf{x}_* 的类别标签 $\hat{l}_*^{\varphi(n)}$;

step 11: 根据 $\mathbf{l}_* = [\hat{l}_*^1; \hat{l}_*^2; \dots; \hat{l}_*^N]^T$ 通过 OvO 解码规则确定类别向量 $\mathbf{y}_* = [y_{*1}, y_{*2}, \dots, y_{*q}]^T$.

3 实验

3.1 实验设置

3.1.1 数据集

本文的对比实验共使用 10 个真实的多维分类数据集, 并在表 3 汇总了这些数据集的详细信息, 包括样本个数 (Example)、维度数 (Dim.)、每个维度包含的类标签个数 (Label/Dim.) 以及特征维度 (Feature). 其中, 对于“Label/Dim.”, 如果数据集在所有维度上的类标签数量相同, 则标注为该值; 否则, 每个维度的标签数量按顺序列出. n 和 x 分别表示数字特征和名义特征.

表3 实验数据集的特点

Data Set	Example	Dim.	Label/Dim.	Feature
Edm	154	2	3	$16n$
Flare1	323	3	3, 4, 2	$10x$
Oes97	334	16	3	$263n$
Jura	359	2	4, 5	$9n$
Oes10	403	16	3	$298n$
WQplants	1060	7	4	$16n$
WQanimals	1060	7	4	$16n$
WaterQuality	1060	14	4	$16n$
BeLaE	1930	5	5	$1n, 44x$
Default	28779	4	2, 7, 4, 2	$14n, 6x$

3.1.2 评价指标

本文选取 3 种评价指标来衡量不同对比方法的性能表现, 包括汉明分值 (HS)、精确匹配 (EM) 和亚精确匹配 (SEM). 设 $T = \{(\mathbf{x}_i, \mathbf{y}_i) | 1 \leq i \leq s\}$ 表示多维分类测试集, $f: \mathcal{X} \mapsto \mathcal{Y}$ 表示待评估的多维分类模型; 对于测试样本 \mathbf{x}_i , 令 $\mathbf{y}_i = [y_{i1}, y_{i2}, \dots, y_{iq}]^T \in \mathcal{Y}$ 表示其真实的类别向量, $\hat{\mathbf{y}}_i = [\hat{y}_{i1}, \hat{y}_{i2}, \dots, \hat{y}_{iq}]^T \in \mathcal{Y}$ 表示模型预测的类别向量. 定义样本 \mathbf{x}_i 被预测正确的维度个数记为 $\rho^{(i)} = \sum_{j=1}^q [y_{ij} = \hat{y}_{ij}]$, 其中 $[\cdot]$ 表示当 \cdot 成立时 $[\cdot]$ 返回值为 1, 当 \cdot 不成立时 $[\cdot]$ 返回值为 0. 3 种评价指标的计算公式定义如下:

汉明分值

$$\text{HS}(f) = \frac{1}{s} \sum_{i=1}^s \frac{1}{q} \cdot \rho^{(i)}, \quad (7)$$

汉明分值表示样本的类别空间被预测正确个数的平均值;

精确匹配

$$\text{EM}(f) = \frac{1}{s} \sum_{i=1}^s [\rho^{(i)} = q], \quad (8)$$

精确匹配度量所有类别空间均被预测正确的样本所占的比例;

亚精确匹配

$$\text{SEM}(f) = \frac{1}{s} \sum_{i=1}^s [\rho^{(i)} \geq q - 1], \quad (9)$$

亚精确匹配度量至少有 $q - 1$ 个类别空间被预测正确的样本所占的比例。

需要注意的是, 以上 3 种评价指标均为指标数值越大代表模型泛化性能越好。在本文的对比实验中, 采用十折交叉验证的方法来对模型进行评估, 并计算出各评价指标的均值和方差, 用来比较各对比方法的优劣。

3.1.3 对比方法

本文将 DRCC 与 7 种先进的 MDC 方法进行比较。

EDCC (基于分解策略的分类器链集成方法)^[19]: 该方法通过分解策略将多维分类任务转化为多个二类分类任务, 使用 10 个随机生成的链顺序构建了 10 个 DCC 分类器, 并通过多数投票将各分类器的预测结果合并。

BR^[10]: 将多维分类任务转化为多个独立的多类分类任务, 针对每个维度单独训练多类分类器。

CP^[11]: 将输出空间的每个类别组合视为一个新的类别, 然后将多维分类问题转化为新的多类分类问题。

EBCC (基于贝叶斯分类器链的集成方法)^[22]: 该方法在简化的贝叶斯框架下生成多个确定的链顺序, 利用这些链顺序构建了多个 BCC 分类器, 并采用多数投票将各分类器的预测结果结合。

ESC (超类分类器的集成方法)^[2]: 该方法将类别变量划分为若干组, 仅考虑同一组中类别变量之间的依赖关系, 其余的类别依赖关系可以通过后续方案进一步考虑。

MDKNN (基于实例的多维分类方法)^[31]: 该方法基于每对类别空间的 K 近邻计数统计, 进行最大后验推断, 并通过参考经验 K 近邻准确度, 确定未见实例在各类别空间中的类别标签。

DLEM (分解标签编码方法)^[32]: 该方法通过分解策略将多维分类的输出空间转换为三元标签空间, 仅考虑同一类别空间内的标签依赖关系, 其余依赖关系通过特征流形结构进一步处理。

本文中的所有方法 (DLEM 不需要基分类器) 都使用由开源软件包 LIBLINEAR^[33] 实现的逻辑回归 (LR) 作为基分类器。

3.2 实验结果

DRCC 和所有对比算法在 10 个真实的 MDC 数

据集上基于上述 3 种评价指标的实验结果如表 4 所示, 在每个数据集上的最佳性能以粗体显示, 表 4 中还进一步给出了平均排名 (Avg. Rank) 的结果, 以显示这些方法的综合分类性能。此外, 本文使用显著度为 0.05 的威尔科克森符号秩检验^[34] 来显示在 10 个数据集上 DRCC 方法是否可以取得在统计上优于 7 种对比方法的泛化性能, 表 5 总结了相应的统计结果。

根据给出的实验结果, 可以得到如下观察:

1) 在 10 个 MDC 数据集上, DRCC 在 Hamming Score 上的平均排名为 1.8, 在 Exact Match 上的平均排名为 2.1, 在 Sub-Exact Match 上的平均排名为 1.3。在所有评价指标中, DRCC 在 60% 的情况下排名第 1, 在 16.7% 的情况下排名第 2, 在 16.7% 的情况下排名第 3, 优于其他算法。

2) 如表 5 所示, DRCC 方法在 Sub-Exact Match 评价指标上取得统计上明显优于其他对比方法的泛化性能。

3) 如表 4 和表 5 所示, DRCC 方法的整体泛化性能优于 EDCC 方法, 充分显示了 DRCC 方法所提出的优化链序策略的有效性。

4) 值得注意的是, DRCC 方法在所有 3 个评价指标上都取得优于 BR 方法的泛化性能, 这个实验结果清晰地验证了考虑类别变量之间依赖关系的必要性。

5) 对比方法 CP 可以看作通过类似幂集变换操作优化 Exact Match 这个评价指标, ESC 方法通过将类别变量划分为若干组, 以减轻 CP 方法在训练集上的过拟合问题, 值得注意的是, DRCC 方法并未针对任何评价指标进行调整, 仍然可以在 Exact Match 指标上取得与 CP 方法持平 (tie), 优于 (win) ESC 方法的统计检验结果。

6) MDKNN 方法采用了两级策略来考虑类别依赖关系, 其中, 在第 1 级考虑了成对依赖关系, 而在第 2 级通过动态选择最优成对预测结果进行集成, 间接优化全局预测。实验结果显示, 尽管 DRCC 方法在 30 种配置 (10 个数据集 \times 3 种评价指标) 中有 22 种优于 MDKNN, 但也存在较多劣于 MDKNN 的情况。因此, 探索融合两级策略和链式结构以生成更好的学习模型是一个未来可研究的方向。

7) DLEM 方法利用特征空间中的流形结构对分解后的标签信息进行增强, 通过优化标签信息, 构建高效的预测模型。实验结果显示, 在 Edm 和 Jura 数据集上, DLEM 方法明显优于 DRCC 方法, 而在其他数据集上, DRCC 与 DLEM 持平或明显占优。因此,

表4 各算法在3种评价指标上的实验结果

Data Set	Hamming Score							
	DRCC	EDCC	BR	CP	EBCC	ESC	MDKNN	DLEM
Edm	0.744 ± 0.071(4)	0.721 ± 0.067(6)	0.721 ± 0.072(6)	0.737 ± 0.072(5)	0.784 ± 0.083(2)	0.708 ± 0.077(8)	0.757 ± 0.102(3)	0.787 ± 0.063(1)
Flare1	0.925 ± 0.034(1)	0.923 ± 0.037(3)	0.925 ± 0.034(1)	0.923 ± 0.033(3)	0.909 ± 0.033(7)	0.922 ± 0.033(5)	0.922 ± 0.035(5)	0.897 ± 0.037(8)
Oes97	0.746 ± 0.021(2)	0.747 ± 0.022(1)	0.736 ± 0.028(6)	0.523 ± 0.049(8)	0.737 ± 0.027(5)	0.732 ± 0.024(7)	0.739 ± 0.025(3)	0.738 ± 0.027(4)
Jura	0.625 ± 0.053(3)	0.616 ± 0.058(4)	0.605 ± 0.067(6)	0.592 ± 0.071(8)	0.610 ± 0.070(5)	0.593 ± 0.068(7)	0.652 ± 0.071(2)	0.720 ± 0.060(1)
Oes10	0.800 ± 0.015(3)	0.797 ± 0.018(4)	0.796 ± 0.017(5)	0.687 ± 0.037(8)	0.795 ± 0.020(6)	0.779 ± 0.022(7)	0.803 ± 0.013(2)	0.805 ± 0.016(1)
WQplants	0.662 ± 0.015(1)	0.659 ± 0.015(3)	0.658 ± 0.014(4)	0.649 ± 0.016(7)	0.655 ± 0.015(5)	0.653 ± 0.016(6)	0.660 ± 0.015(2)	0.632 ± 0.024(8)
WQanimals	0.637 ± 0.011(1)	0.635 ± 0.012(2)	0.631 ± 0.013(4)	0.628 ± 0.013(7)	0.629 ± 0.014(6)	0.631 ± 0.014(4)	0.635 ± 0.015(2)	0.606 ± 0.010(8)
WaterQuality	0.647 ± 0.014(1)	0.647 ± 0.012(1)	0.644 ± 0.011(4)	0.625 ± 0.011(7)	0.644 ± 0.012(4)	0.642 ± 0.014(6)	0.646 ± 0.011(3)	0.619 ± 0.013(8)
BeLaE	0.451 ± 0.013(1)	0.449 ± 0.016(2)	0.427 ± 0.017(3)	0.383 ± 0.023(8)	0.424 ± 0.018(4)	0.420 ± 0.022(5)	0.405 ± 0.020(7)	0.412 ± 0.025(6)
Default	0.672 ± 0.004(1)	0.671 ± 0.004(3)	0.669 ± 0.003(4)	0.669 ± 0.004(4)	0.669 ± 0.003(4)	0.672 ± 0.004(1)	0.655 ± 0.002(8)	0.663 ± 0.002(7)
Avg.Rank	1.8	2.9	4.3	6.5	4.8	5.6	3.7	5.2

Data Set	Exact Match							
	DRCC	EDCC	BR	CP	EBCC	ESC	MDKNN	DLEM
Edm	0.520 ± 0.154(5)	0.475 ± 0.127(7)	0.474 ± 0.139(8)	0.578 ± 0.106(3)	0.525 ± 0.116(4)	0.486 ± 0.139(6)	0.611 ± 0.170(2)	0.638 ± 0.080(1)
Flare1	0.821 ± 0.075(1)	0.818 ± 0.081(4)	0.821 ± 0.075(1)	0.817 ± 0.068(5)	0.786 ± 0.061(7)	0.821 ± 0.073(1)	0.814 ± 0.083(6)	0.762 ± 0.078(8)
Oes97	0.063 ± 0.056(1)	0.057 ± 0.048(4)	0.060 ± 0.045(2)	0.054 ± 0.046(6)	0.054 ± 0.044(6)	0.054 ± 0.044(6)	0.057 ± 0.043(4)	0.060 ± 0.049(2)
Jura	0.382 ± 0.095(3)	0.382 ± 0.105(3)	0.359 ± 0.107(6)	0.357 ± 0.108(7)	0.371 ± 0.112(5)	0.351 ± 0.096(8)	0.454 ± 0.104(2)	0.540 ± 0.080(1)
Oes10	0.097 ± 0.043(3)	0.099 ± 0.045(1)	0.084 ± 0.033(7)	0.082 ± 0.049(8)	0.097 ± 0.045(3)	0.099 ± 0.037(1)	0.092 ± 0.053(6)	0.094 ± 0.045(5)
WQplants	0.098 ± 0.032(1)	0.094 ± 0.036(3)	0.092 ± 0.033(7)	0.093 ± 0.031(5)	0.094 ± 0.034(3)	0.093 ± 0.036(5)	0.095 ± 0.027(2)	0.085 ± 0.028(8)
WQanimals	0.067 ± 0.020(1)	0.065 ± 0.014(2)	0.058 ± 0.017(5)	0.065 ± 0.018(2)	0.058 ± 0.018(5)	0.064 ± 0.019(4)	0.058 ± 0.019(7)	0.047 ± 0.015(8)
WaterQuality	0.008 ± 0.007(2)	0.006 ± 0.008(4)	0.005 ± 0.008(5)	0.000 ± 0.000(8)	0.004 ± 0.007(7)	0.005 ± 0.008(5)	0.008 ± 0.005(2)	0.009 ± 0.006(1)
BeLaE	0.034 ± 0.014(1)	0.033 ± 0.013(2)	0.021 ± 0.008(8)	0.026 ± 0.014(6)	0.031 ± 0.010(3)	0.027 ± 0.009(4)	0.022 ± 0.008(7)	0.027 ± 0.014(4)
Default	0.186 ± 0.007(3)	0.185 ± 0.008(4)	0.181 ± 0.007(6)	0.194 ± 0.008(1)	0.184 ± 0.007(5)	0.187 ± 0.007(2)	0.177 ± 0.006(8)	0.181 ± 0.007(6)
Avg.Rank	2.1	3.4	5.5	5.1	4.8	4.2	4.6	4.4

Data Set	Sub-Exact Match							
	DRCC	EDCC	BR	CP	EBCC	ESC	MDKNN	DLEM
Edm	0.968 ± 0.055(1)	0.968 ± 0.047(1)	0.968 ± 0.045(1)	0.896 ± 0.071(8)	0.942 ± 0.056(4)	0.929 ± 0.065(6)	0.902 ± 0.071(7)	0.936 ± 0.052(5)
Flare1	0.957 ± 0.039(1)	0.954 ± 0.039(3)	0.957 ± 0.039(1)	0.954 ± 0.039(3)	0.944 ± 0.046(7)	0.951 ± 0.036(6)	0.954 ± 0.037(3)	0.938 ± 0.044(8)
Oes97	0.135 ± 0.075(1)	0.123 ± 0.076(3)	0.123 ± 0.065(3)	0.084 ± 0.062(8)	0.120 ± 0.068(5)	0.111 ± 0.060(6)	0.126 ± 0.072(2)	0.111 ± 0.068(6)
Jura	0.869 ± 0.043(2)	0.850 ± 0.044(3)	0.850 ± 0.049(3)	0.827 ± 0.053(8)	0.850 ± 0.051(3)	0.836 ± 0.059(7)	0.850 ± 0.082(3)	0.900 ± 0.085(1)
Oes10	0.223 ± 0.074(1)	0.201 ± 0.064(4)	0.198 ± 0.055(6)	0.169 ± 0.042(8)	0.211 ± 0.073(3)	0.201 ± 0.056(4)	0.213 ± 0.067(2)	0.191 ± 0.073(7)
WQplants	0.292 ± 0.040(2)	0.287 ± 0.046(3)	0.286 ± 0.044(4)	0.285 ± 0.052(5)	0.281 ± 0.044(7)	0.282 ± 0.049(6)	0.294 ± 0.043(1)	0.241 ± 0.044(8)
WQanimals	0.242 ± 0.025(1)	0.235 ± 0.033(3)	0.229 ± 0.030(6)	0.232 ± 0.032(4)	0.227 ± 0.032(7)	0.231 ± 0.029(5)	0.238 ± 0.031(2)	0.193 ± 0.023(8)
WaterQuality	0.051 ± 0.021(2)	0.047 ± 0.023(4)	0.047 ± 0.019(4)	0.034 ± 0.017(8)	0.045 ± 0.019(6)	0.048 ± 0.019(3)	0.056 ± 0.014(1)	0.044 ± 0.017(7)
BeLaE	0.161 ± 0.022(1)	0.157 ± 0.022(2)	0.134 ± 0.025(5)	0.117 ± 0.019(8)	0.146 ± 0.018(3)	0.128 ± 0.024(6)	0.120 ± 0.029(7)	0.136 ± 0.032(4)
Default	0.604 ± 0.007(1)	0.604 ± 0.008(1)	0.601 ± 0.006(4)	0.594 ± 0.008(6)	0.600 ± 0.006(5)	0.604 ± 0.008(1)	0.568 ± 0.006(8)	0.586 ± 0.005(7)
Avg.Rank	1.3	2.7	3.7	6.6	5.0	5.0	3.6	6.1

表5 DRCC与各对比方法之间的威尔科克森符号秩检验(显著度为0.05)

Evaluation Metrics	DRCC against						
	EDCC	BR	CP	EBCC	ESC	MDKNN	DLEM
Hamming Score	win[1.17e-02]	win[3.91e-03]	win[1.95e-03]	tie[7.81e-02]	win[3.91e-03]	tie[5.10e-01]	tie[6.25e-01]
Exact Match	win[2.73e-02]	win[3.91e-03]	tie[2.32e-01]	win[3.52e-02]	win[1.56e-02]	tie[5.47e-01]	tie[6.33e-01]
Sub-Exact Match	win[7.81e-03]	win[7.81e-03]	win[1.95e-03]	win[1.95e-03]	win[3.91e-03]	win[1.95e-02]	win[3.91e-02]

根据不同的数据集采用对应的MDC方法可能是一种合理的决策。

3.3 消融实验

为了验证DRCC方法设计的有效性,本节将DRCC与其他两个退化版本进行对比,进行消融实验(Ablation Study)。退化版本DeV1使用10个随机的链序来构建分类器,通过多数投票规则组合预测结

果;退化版本DeV2不考虑易错分类器的影响,使用优化后的单链策略构建分类器。

DRCC与两个退化版本的对比实验结果如图5所示。此外,本文仍然使用显著度为0.05的威尔科克森符号秩检验^[34]来检测在10个数据集上DRCC、DeV1和DeV2方法的泛化性能之间的统计关系,其中表6总结了相应的统计结果。

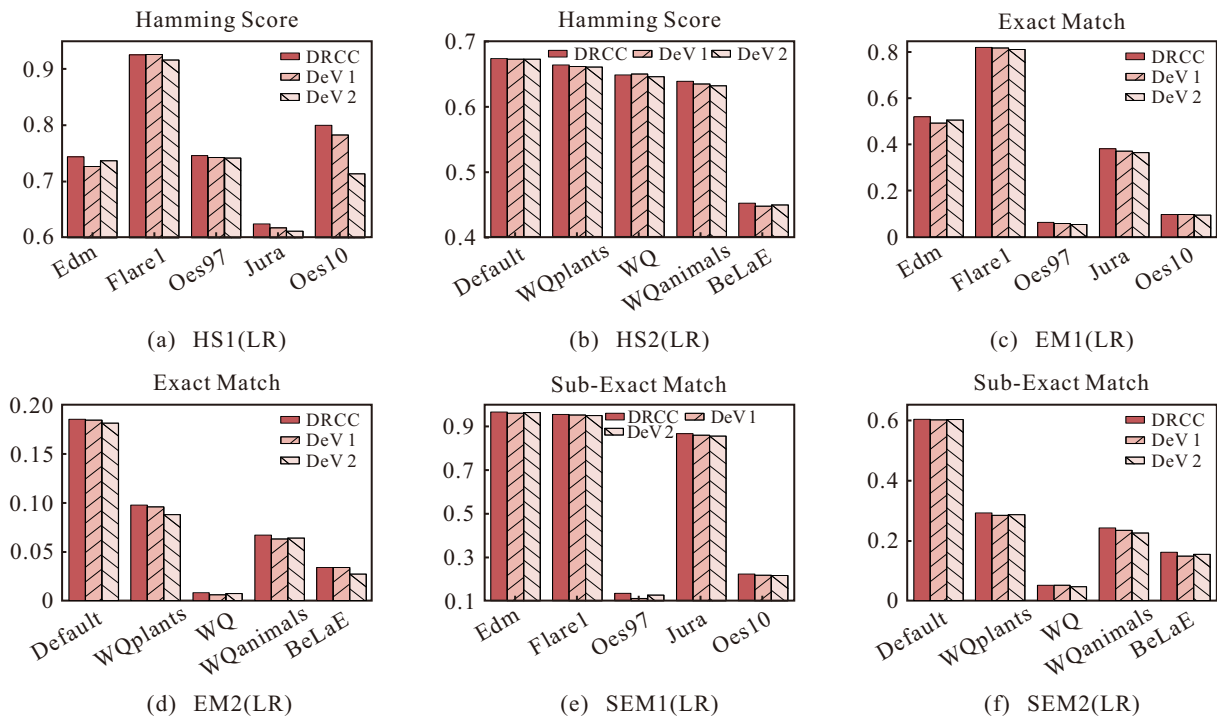


图5 DRCC 和退化版本的泛化性能对比

表6 DRCC 与退化版本的威尔科克森符号秩检验

Evaluation Metrics	DRCC vs DeV1	DRCC vs DeV2
Hamming Score	win[1.56e-02]	win[1.95e-03]
Exact Match	win[7.81e-03]	win[5.86e-03]
Sub-Exact Match	win[3.91e-03]	win[1.95e-03]

实验结果表明, DRCC 比 DeV1 和 DeV2 具有更好的泛化性能, 在所有的 30 种配置 (10 个数据集 \times 3 种评价指标) 中, DRCC 方法在 (25/30) 配置和 (30/30) 配置下的实验结果优于 DeV1 和 DeV2. 从图 5 中还可以看到, DeV1 在 Flare1 和 WQ 数据集上的汉明分值略高于 DRCC 方法. 这种现象的一个可能的原因是, DRCC 采用优化的单链策略在 Flare1

和 WQ 数据集上产生了过拟合的问题, 而 DeV1 采用多条随机链序构成一个集成分类模型, 这种集成策略对少量错误预测具有更好的鲁棒性. 值得注意的是, 本文所提出的 DRCC 方法在大多数配置下都优于 DeV1, 在所有配置下都优于 DeV2, 这可以表明本文所提出的优化单链策略和控制特征空间策略的有效性.

3.4 参数敏感性分析

本节研究 DRCC 方法的参数敏感性, 即二分类器的阈值 α 选取不同值时 DRCC 方法的分类性能如何变化. 图 6 给出了在数据集 Oes10 和 BeLaE 上,

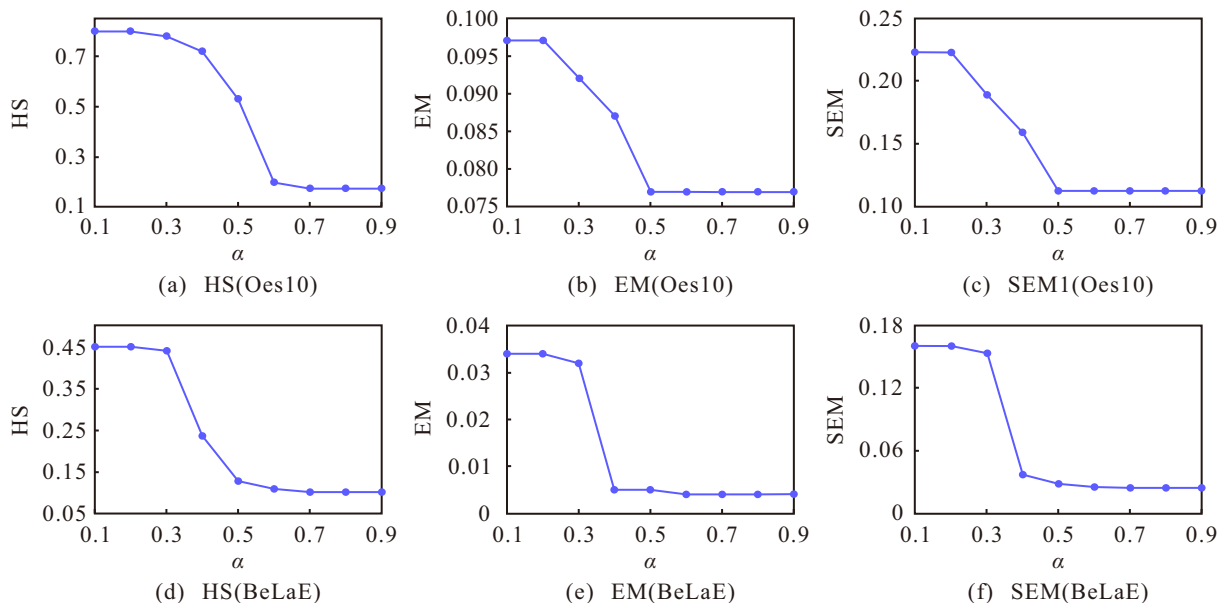


图6 当 α 在 $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ 范围内变化时 DRCC 的泛化性能

当 α 在 $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ 范围内变化时 DRCC 的泛化性能的变化情况. 如图 6 所示, 当 $\alpha = 0.1$ 时 DRCC 方法在各评价指标上都可以取得较好的泛化性能. 因此, 本文将参数 α 固定为 0.1.

3.5 时间复杂度分析

DRCC 算法的时间复杂度主要由 GA 的运行时间 T_{GA} 和二类分类器的训练时间 T_{train} 两部分组成. 其中, T_{GA} 包括适应度计算、选择操作、交叉操作和变异操作所需的计算开销. 设 G 为最大迭代次数, P 为种群大小, N 为分解得到的标签对数量, 则 T_{GA} 的复杂度为 $\mathcal{O}(G \cdot P \cdot N^2) + \mathcal{O}(P^2) + \mathcal{O}(P \cdot N) + \mathcal{O}(P)$, 对于 T_{train} , 设 $\mathcal{F}(n_1, n_2, n_3)$ 为分类算法 \mathcal{L} 的训练复杂度, 其中 n_1, n_2, n_3 分别表示样本个数、特征个数和类别标签个数, 则 T_{train} 的复杂度为 $\mathcal{O}(q \cdot K^2 \cdot \mathcal{F}(m, d + q \cdot K^2, 2))$. 其中: q 表示维度个数; K 表示所有类别空间包含标记个数的最大值, 即 $K = \max_{1 \leq j \leq q} K_j$. 故 DRCC 方法的时间复杂度为 $\mathcal{O}(G \cdot P \cdot N^2) + \mathcal{O}(P^2) + \mathcal{O}(P \cdot N) + \mathcal{O}(P) + \mathcal{O}(q \cdot K^2 \cdot \mathcal{F}(m, d + q \cdot K^2, 2))$.

表 7 给出了每个数据集上 DRCC 方法及 7 种对比方法在十折交叉验证实验中所消耗的全部时间(单位: s). 可以看出, BR 是效率最高的方法, 一般情况下, DRCC 比 ESC 耗时更短, 与 EDCC 和 DLEM 在大部分数据集上耗时相当. 在 Default 这种大型数据集上, 除 BR、DRCC 和 EBCC 外, 其他方法的运行时间都急剧增加.

表7 各算法在每个数据集上的时间成本

Data Set	DRCC	EDCC	BR	CP	EBCC	ESC	MDKNN	DLEM
Edm	2	2	1	1	2	5	1	2
Flare1	3	1	1	1	5	12	1	1
Oes97	19	22	2	12	35	42	3	2
Jura	5	1	1	1	2	9	1	2
Oes10	20	30	3	18	48	55	3	3
WQplants	18	9	1	5	3	16	4	16
WQanimals	15	9	1	5	3	18	4	15
WaterQuality	39	27	1	9	12	32	10	16
BeLaE	21	29	2	41	8	29	6	92
Default	28	208	9	78	40	111	326	444306

4 结论

本文提出了一种基于分解排序的多维分类器链方法 (DRCC), DRCC 通过 OvO 分解规则将多维分类问题分解为多个二类分类问题. 为了解决分类器链中标签对顺序选择问题, DRCC 引入了 LOP 并通过 GA 算法求解, 确定了分类器链中合适的标签顺序. 此外, DRCC 在最终预测阶段通过控制特征空间

来减轻错误传播的影响. 在 10 个多维分类数据集上的实验结果表明, DRCC 优于几种先进的多维分类方法.

本文仅从标签顺序和错误传播的角度研究针对多维分类的分类器链方法, 未来可以进一步探索研究多维分类器链方法的其他问题. 例如, 分类器链方法的计算复杂度会随着标签数量的增加而增加, 可以通过并行化策略加速分类器链的训练过程.

参考文献 (References)

- [1] Zhou Z H. Machine learning[M]. Singapore: Springer Singapore, 2021: 2-20.
- [2] Read J, Bielza C, Larrañaga P. Multi-dimensional classification with super-classes[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2014, 26(7): 1720-1733.
- [3] Shatkay H, Pan F X, Rzhetsky A, et al. Multi-dimensional classification of biomedical text: Toward automated, practical provision of high-utility text to diverse users[J]. *Bioinformatics*, 2008, 24(18): 2086-2093.
- [4] Lertnattee V, Theeramunkong T. Multi-dimensional text classification for drug information[J]. *IEEE Transactions on Information Technology in Biomedicine*, 2004, 8(3): 306-312.
- [5] 王浩宇, 张欣然, 王雪松, 等. 基于自监督增强特征的直推式零样本图像分类[J]. *控制与决策*, 2024, 39(5): 1707-1717.
(Wang H Y, Zhang X R, Wang X S, et al. Transductive zero-shot image classification based on self-supervised enhancement feature[J]. *Control and Decision*, 2024, 39(5): 1707-1717.)
- [6] 张峰, 李希城, 董春茹, 等. 基于深度情感唤醒网络的多模态情感分析与情绪识别[J]. *控制与决策*, 2022, 37(11): 2984-2992.
(Zhang F, Li X C, Dong C R, et al. Deep emotional arousal network for multimodal sentiment analysis and emotion recognition[J]. *Control and Decision*, 2022, 37(11): 2984-2992.)
- [7] Borchani H, Bielza C, Toro C, et al. Predicting human immunodeficiency virus inhibitors using multi-dimensional Bayesian network classifiers[J]. *Artificial Intelligence in Medicine*, 2013, 57(3): 219-229.
- [8] Verma S P, Uscanga-Junco O A, Diaz-González L. A statistically coherent robust multi-dimensional classification scheme for water[J]. *Science of the Total Environment*, 2021, 750: 141704.
- [9] Al Muktedir A H, Miyazawa T, Martinez-Julia P, et al. Multi-target classification based automatic virtual resource allocation scheme[J]. *IEICE Transactions on Information and Systems*, 2019, 102(5): 898-909.
- [10] Zhang M L, Li Y K, Liu X Y, et al. Binary relevance for multi-label learning: An overview[J]. *Frontiers of Computer Science*, 2018, 12(2): 191-202.
- [11] Tsoumakas G, Katakis I, Vlahavas I. Random k -labelsets for multilabel classification[J]. *IEEE Transactions on Knowledge and Data Engineering*,

- 2011, 23(7): 1079-1089.
- [12] Bielza C, Li G, Larrañaga P. Multi-dimensional classification with Bayesian networks[J]. *International Journal of Approximate Reasoning*, 2011, 52(6): 705-727.
- [13] Gil-Begue S, Bielza C, Larrañaga P. Multi-dimensional Bayesian network classifiers: A survey[J]. *Artificial Intelligence Review*, 2021, 54(1): 519-559.
- [14] Jia B B, Zhang M L. Maximum margin multi-dimensional classification[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(12): 7185-7198.
- [15] Jia B B, Zhang M L. Multi-dimensional classification via stacked dependency exploitation[J]. *Science China Information Sciences*, 2020, 63(12): 222102.
- [16] Jia B B, Zhang M L. Multi-dimensional classification via sparse label encoding[C]. Proceedings of the 38th International Conference on Machine Learning. Virtual Event, 2021: 4917-4926.
- [17] Read J, Pfahringer B, Holmes G, et al. Classifier chains for multi-label classification[J]. *Machine Learning*, 2011, 85(3): 333-359.
- [18] Madjarov G, Kocev D, Gjorgjevikj D, et al. An extensive experimental comparison of methods for multi-label learning[J]. *Pattern Recognition*, 2012, 45(9): 3084-3104.
- [19] Jia B B, Zhang M L. Decomposition-based classifier chains for multi-dimensional classification[J]. *IEEE Transactions on Artificial Intelligence*, 2022, 3(2): 176-191.
- [20] Dembczyński K, Cheng W W, Hüllermeier E, et al. Bayes optimal multilabel classification via probabilistic classifier chains[C]. Proceedings of the 27th International Conference on International Conference on Machine Learning. Haifa, New York: ACM, 2010: 279-286.
- [21] Liu W W, Tsang I W, Liu W W, et al. On the optimality of classifier chain for multi-label classification[C]. Proceedings of the 29th International Conference on Neural Information Processing Systems — Volume 1. Montreal, New York: ACM, 2015: 712-720.
- [22] Zaragoza J H, Sucar L E, Morales E F, et al. Bayesian chain classifiers for multidimensional classification[C]. Proceedings of the Twenty-Second international joint conference on Artificial Intelligence. Barcelona, Catalonia, Spain: AAAI Press, 2011: 2192-2197.
- [23] Xie J, Yu L, Zhu L, et al. Conditional entropy based classifier chains for multi-label classification[J]. *Neurocomputing*, 2019, 335: 185-194.
- [24] Ding J M, Zhou S J, Li R X, et al. Association rules-based classifier chains method[J]. *IEEE Access*, 2022, 10: 18210-18221.
- [25] Read J, Martino L, Luengo D. Efficient Monte Carlo optimization for multi-label classifier chains[C]. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 2002.
- [26] Freitas Rocha V, Varejão F M, Segatto M E V. Ensemble of classifier chains and decision templates for multi-label classification[J]. *Knowledge and Information Systems*, 2022, 64(3): 643-663.
- [27] Duan K B, Keerthi S S. Which is the best multiclass SVM method? An empirical study[M]. Multiple Classifier Systems. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005: 278-285.
- [28] Huang G, Lim A. Designing a hybrid genetic algorithm for the linear ordering problem[C]. Proceedings of the 2003 international conference on Genetic and evolutionary computation. Chicago IL USA: Springer Berlin Heidelberg, 2003: 1053-1064.
- [29] Bao Y Y, Liu Y, Wang J S, et al. Genetic algorithm based on grid maps for solving robot path planning problem[J]. *Engineering Letters*, 2023, 31(4): 1635-1648.
- [30] Mishra N K, Singh P K. Linear ordering problem based classifier chain using genetic algorithm for multi-label classification[J]. *Applied Soft Computing*, 2022, 117: 108395.
- [31] Jia B B, Zhang M L. MD-KNN: An instance-based approach for multi-dimensional classification[C]. 2020 25th International Conference on Pattern Recognition. Milan, 2021: 126-133.
- [32] Jia B B, Zhang M L. Multi-dimensional classification via decomposed label encoding[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(2): 1844-1856.
- [33] Fan R E, Chang K W, Hsieh C J, et al. LIBLINEAR: A library for large linear classification[J]. *Journal of Machine Learning Research*, 2008, 9: 1871-1874.
- [34] Demšar J. Statistical comparisons of classifiers over multiple data sets[J]. *Journal of Machine Learning Research*, 2006, 7: 1-30.

作者简介

李二超 (1979-), 男, 教授, 博士生导师, 主要研究方向为人工智能、进化计算、机器人控制, E-mail: lecstarr@163.com;

杨宏强 (2001-), 男, 硕士生, 主要研究方向为多维分类, E-mail: 1357136128@qq.com.