



多标记数据驱动的可变换算子值核

汪振鑫, 陈德刚, 车晓雅

引用本文:

汪振鑫, 陈德刚, 车晓雅. 多标记数据驱动的可变换算子值核[J]. *智能系统学报*, 2026, 21(2): 365-374.

WANG Zhenxin, CHEN Degang, CHE Xiaoya. Transformable operator-valued kernels driven by multi-label datasets[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(2): 365-374.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202503021>

您可能感兴趣的其他文章

基于级联宽度学习的多模态材质识别

Cascade broad learning for multi-modal material recognition

智能系统学报. 2020, 15(4): 787-794 <https://dx.doi.org/10.11992/tis.201908021>

代价敏感数据的多标记特征选择算法

Multi-label feature selection algorithm for cost-sensitive data

智能系统学报. 2019, 14(5): 929-938 <https://dx.doi.org/10.11992/tis.201807027>

结合谱聚类的标记分布学习

Label distribution learning based on spectral clustering

智能系统学报. 2019, 14(5): 966-973 <https://dx.doi.org/10.11992/tis.201809019>

鲁棒的半监督多标签特征选择方法

A robust, semi-supervised, and multi-label feature selection method

智能系统学报. 2019, 14(4): 812-819 <https://dx.doi.org/10.11992/tis.201809017>

弹性网络核极限学习机的多标记学习算法

Multi-label learning algorithm of an elastic net kernel extreme learning machine

智能系统学报. 2019, 14(4): 831-842 <https://dx.doi.org/10.11992/tis.201806005>

基于改进卷积神经网络的多标记分类算法

A multi-label classification algorithm based on an improved convolutional neural network

智能系统学报. 2019, 14(3): 566-574 <https://dx.doi.org/10.11992/tis.201804056>

DOI: 10.11992/tis.202503021

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20250928.1752.004>

多标记数据驱动的可变换算子值核

汪振鑫¹, 陈德刚², 车晓雅²

(1. 华北电力大学 控制与计算机工程学院, 北京 102200; 2. 华北电力大学 数理学院, 北京 102200)

摘要: 算子值核是取值为希尔伯特空间上算子的二元函数, 在机器学习领域中旨在更好地描述多任务学习中不同任务之间的关联性。多标记学习是一种特殊的多任务学习, 本文基于核对齐方法从多标记数据集中学习算子值核并构建多标记学习的预测模型。1) 利用核对齐方法学习样例级特征重要度分布; 2) 基于样例级特征重要度分布构造算子值核, 证明其不仅是偏迹核而且是可变换算子值核, 且其对应核矩阵中的每个分块刻画了样例间标记相关性的交互信息; 3) 设计基于可变换算子值核的多标记学习算法, 在 9 个多标记数据集上与 4 种高性能算法进行对比实验, 结果验证了所提算法的有效性。

关键词: 多标记学习; 算子值核; 可变换算子值核; 偏迹核; 标记相关性; 样例级特征重要度分布; 交互信息; 核对齐

中图分类号: TP181 文献标志码: A 文章编号: 1673-4785(2026)02-0365-10

中文引用格式: 汪振鑫, 陈德刚, 车晓雅. 多标记数据驱动的可变换算子值核 [J]. 智能系统学报, 2026, 21(2): 365-374.

英文引用格式: WANG Zhenxin, CHEN Degang, CHE Xiaoya. Transformable operator-valued kernels driven by multi-label datasets[J]. CAAI transactions on intelligent systems, 2026, 21(2): 365-374.

Transformable operator-valued kernels driven by multi-label datasets

WANG Zhenxin¹, CHEN Degang², CHE Xiaoya²

(1. North China Electric Power University, School of Control and Computer Engineering, Beijing 102200, China; 2. North China Electric Power University, Department of Mathematics and Physics, Beijing 102200, China)

Abstract: An operator-valued kernel is a binary function that takes the value of an operator on Hilbert space, which in the field of machine learning aims to better describe the correlation between different tasks in multi-task learning. Multi-label learning is a special kind of multi-task learning, in this paper, we learn operator-valued kernels from multi-label datasets based on the kernel alignment method and construct a prediction model for multi-label learning. Firstly, we use kernel alignment method to learn the instance-level feature importance distribution; secondly, we construct operator-valued kernel based on the instance-level feature importance distribution, and prove that it is not only partial trace kernel but also transformable operator-valued kernel, and that each block of its corresponding kernel matrix depicts the interaction information of label correlation among the samples; lastly, we design the multi-label learning algorithms based on transformable operator-valued kernel, and conduct comparative experiments with four high-performance algorithms on nine multi-label datasets, the results verify effectiveness of our proposed algorithm.

Keywords: multi-label learning; operator-valued kernel; transformable operator-valued kernel; partial trace kernel; label correlation; instance-level feature importance distribution; interaction information; kernel alignment

算子值核^[1-12]作为数值核的扩展, 其取值为希尔伯特空间上的算子。在机器学习领域, 引入算子值核旨在更好地描述多任务学习中不同任务之间的关联性。在多标记学习^[13-22]这一特定的多任务学习场景中, 这种关联性可以细化为样例之

间标记相关性^[23]的交互信息, 这种交互信息往往与特定的数据集密切相关。因此, 从多标记数据集中学习出能够准确描述样例之间标记相关性交互信息的算子值核具有重要的理论意义和应用价值。

算子值核最早是由 Micchelli 等^[1]引入机器学习领域的, 目前其在机器学习领域内的研究主要围绕理论与应用两个方向展开。在基础理论层

收稿日期: 2025-03-23. 网络出版日期: 2025-09-29.

基金项目: 国家自然科学基金项目 (12571496, 12201213).

通信作者: 陈德刚. E-mail: chengdegang@263.net.

面, Kadri 等^[2]讨论了算子值核的有限维线性组合的学习问题。与此同时, Kadri 等^[3]将监督学习中泛函学习问题转化为再生核希尔伯特空间(reproducing kernel Hilbert space, RKHS)中函数值的设置问题,实现了算子值核理论体系与学习范式的统一。Huusari 等^[4]在偏迹核的框架下对已知算子值核进行了系统的分类,将其分为可分核、可变换核和纠缠核,这一分类为后续研究提供了理论基准。Dinuzzo 等^[5]首先提出了可分离核的学习方法,并且利用学习到的可分离核来解决希尔伯特空间上向量值函数的正则化问题^[6]。Alvarez 等^[7]介绍了两类用于向量值函数的通用核,分别是可分离核和不可分离核。Lim 等^[8]结合可分离核和可变换核,在知识发现领域引入了一种基于算子值核的自回归向量模型,并用于识别动态系统和检索代表其组件之间相互作用的目标网络。Gregorová等^[9]学习了可分离核的和,并提出了一种适应非线性预测函数的学习方法。Audifren 等^[10]考虑了在线学习环境中向量值函数的学习问题,设计了基于算子值核的在线学习框架。值得关注的是,算子值核在多任务学习中的应用价值日益凸显: Minh 等^[11]给出了算子值核在向量值流形正则化和协正则化多视图学习中的应用。Huusari 等^[12]在多视图学习中介绍并学习一类新的可分离算子值核,该算子值核被用来刻画学习视图内和视图间之间的度量,以此来捕捉数据的多模态结构。

多标记学习旨在构建统一模型同时对多个标记进行类别预测,而多任务学习则是构建统一模型同时处理多个相关任务,如分类和回归问题,因而多标记学习是一种特殊的多任务学习。截至目前算子值核还未应用到多标记学习中。如果考虑在多标记学习中应用算子值核,那么算子值核所取值的矩阵应该能够描述样例之间标记相关性的交互信息。为了实现这一目标,本文从多标记数据集中学习可变换算子值核,并设计了有效的多标记学习算法。主要贡献如下:基于核对齐^[24]方法计算样例级特征重要度分布,从而构造算子值核,以刻画样例之间标记相关性的交互信息,并证明该多标记数据驱动^[25-27]的算子值核不但是偏迹核而且是可变换核;设计基于该核的多标记学习算法,并通过实验验证其性能优势。

1 预备知识

多标记分类问题是机器学习中典型的监督学习问题,其中任意样例均同时被高维输入变量与

多重标记变量所描述。

假设 X 是输入空间(即特征空间), Y 是输出空间(即标记空间)。给定多标记数据集 $D = \{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_n, \mathbf{y}_n)\}$ 。对任意多标记样例,输入向量 $\mathbf{x}_i = (x_i^1, x_i^2, \dots, x_i^m)^T \in X, i = 1, 2, \dots, n$ 由特征集合 $A = \{a_1, a_2, \dots, a_m\}$ 特征所表示,其中 m 为特征个数。输出向量 $\mathbf{y}_i = (y_i^1, y_i^2, \dots, y_i^N)^T \in Y, i = 1, 2, \dots, n$ 被标记集合 $L = \{l_1, l_2, \dots, l_N\}$ 中标记变量所描述,其中 N 表示标记个数。任意样例在给定标记下可能输出值为 +1 或 -1,其中 +1 代表样例具有该标记, -1 表示样例不具有该标记。

为评估多标记算法在多标记分类任务中的性能,通常选择以下 5 个常用的评估指标^[28-29]。

1) Hamming Loss (HL) 计算公式为

$$HL(f) = \frac{1}{n} \sum_{i=1}^n \frac{|\mathbf{Y}_i \Delta \mathbf{Z}_i|}{s}$$

式中: \mathbf{Y}_i 表示样例中真实的标记集合, \mathbf{Z}_i 表示样例中预测标记的集合, Δ 表示两个集合的对称差, f 为函数, s 为标记的个数和 n 为训练样例的个数。HL 用来评估一个样例被错分的次数,该值越小,说明算法的性能越好。

2) One-error (OE) 计算公式为

$$OE(f) = \frac{1}{n} \sum_{i=1}^n G\left(\left(\arg \max_{y \in Y} f(\mathbf{x}_i, y)\right) \notin Y_i\right)$$

其中

$$G(l) = \begin{cases} 0, & l \text{ 为真} \\ 1, & l \text{ 为假} \end{cases}$$

式中: n 为训练样例的个数, \mathbf{Y}_i 为样例中真实的标记集合, G 为示性函数, f 为函数。OE 用来评估在输出结果中排序第一的标记不属于真实标记集合的概率,即预测出的最相关标记实际上是样例无关标记的比例。该值越小,说明算法的性能越好。

3) Coverage (CV) 计算公式为

$$CV(f) = \frac{1}{n} \sum_{i=1}^n \max_{y \in Y_i} \text{rank}_f(\mathbf{x}_i, y) - 1$$

式中: n 为训练样例的个数; CV 表示在标记排序中要覆盖所有相关标记需要搜索的深度,该值越小,说明算法的性能越好。

4) Ranking Loss (RL) 计算公式为

$$RL(f) = \frac{1}{n} \sum_{i=1}^n \frac{1}{|\mathbf{Y}_i \Delta \bar{\mathbf{Y}}_i|}$$

式中: n 为训练样例的个数, $\bar{\mathbf{Y}}$ 是 \mathbf{Y}_i 相对于所有类别标记集合 Y 的补集。RL 表示有多少不相关标记排序高于相关标记,即该评价指标衡量的是排序出现错误时相关和不相关标记的比例。该值越小,说明算法的性能越好。

5) Average precision (AVP) 计算公式为

$$I_{AVP} = \frac{1}{n} \sum_{i=1}^n \frac{1}{|Y_i|} \sum_{y \in Y} \frac{|\{y | \text{rank}_f(x_i, y) \leq \text{rank}_f(x_i, \bar{y})\}|}{\text{rank}_f(x_i, \bar{y})}$$

式中: n 为训练样例的个数, Y_i 为样例中真实的标记集合。 I_{AVP} 表示在所有预测结果排序中, 被排序排在相关标记之前的标记依旧是相关标记且属于相关标记集合的概率。该值越大, 说明算法的性能越好。

下面依次给出算子值核, 偏迹核和可变换算子值核的相关定义。假设 X 是输入空间, Y 是希尔伯特空间, $L(Y)$ 是从 Y 到 Y 的线性算子的集合。

定义 1^[4] $L(Y)$ -值核 K 是 $X \times X$ 上的一个函数, $K(\cdot, \cdot): X \times X \rightarrow L(Y)$; 该函数是半正定的, 如果 $K(x, z) = K(z, x)^*$, 其中上标 $*$ 表示伴随算子; 对于任意 $n \in \mathbb{N}$ 并且 $\{(x_i, y_i)_{i=1,2,\dots,n}\} \in X \times Y$, 则

$\sum_{i,j} \langle y_i, K(x_i, x_j) y_j \rangle \geq 0$ 。
在处理多标记学习问题时, Y 是有限维输出空间, $L(Y)$ 上的算子退化为矩阵, 算子值核 $K(x, z)$ 是取值在 $L(Y)$ 上的二元函数, 矩阵 $K(x, z)$ 旨在描述样例 x 和 z 之间标记相关性的交互信息。

文献 [4] 以偏迹核为工具建立了算子值核理论框架。假设 H_1 和 H_2 是可分离的希尔伯特空间。从希尔伯特空间 H_1 到 H_1 的有限迹范数的有界线性算子集合使用 $L(H_1)$ 表示, $H_1 \otimes H_2$ 表示希尔伯特空间 H_1 和 H_2 的张量积。

定义 2^[4] 令 $\{e_i\}$ 是 H_2 中的正交基, 对于 $L(H_1 \otimes H_2)$ 中的算子 A , 其偏迹 $\text{tr}_{H_2} A$ 是 $L(H_1)$ 中的算子, 由如下等价关系定义:

$$\langle x, (\text{tr}_{H_2} A) y \rangle = \sum_i \langle x \otimes e_i, A(y \otimes e_i) \rangle$$

此外, 偏迹算子 $\text{tr}_{H_2} A$ 在 (η, ξ) 位置上的元素正好对应于分块 $A_{\eta\xi}$ 的迹, 即 $[\text{tr}_{H_2} A]_{\eta\xi} = \text{tr}(A_{\eta\xi})$ 。

定义 3^[4] 偏迹核是具有如下形式的算子值核函数 K :

$$K(x, z) = \text{tr}_H (P_{\varphi(x)\varphi(z)})$$

式中: $P_{\varphi(x)\varphi(z)}$ 是 $L(Y \otimes H)$ 上的算子, tr_H 表示数值核的再生核希尔伯特空间 H 上的偏迹运算。

定义 4^[4] 可变换算子值核是一个函数 $K: X \times X \rightarrow \mathbb{R}^{n \times n}$, 写作:

$$K(x, z) = [\hat{k}(T_h(x), T_l(z))]_{h,l=1}^n, \forall x, z \in X$$

式中: $\hat{k}: \hat{X} \times \hat{X} \rightarrow \mathbb{R}$ 是数值核, $\{T_i\}_{i=1}^n$ 是从 X 到 \hat{X} 的映射。

在应用到实际问题时, 数值核 \hat{k} 和映射 $\{T_i\}_{i=1}^n$

一般都需要预设。本文旨在从多标记数据中学习数值核 \hat{k} 和映射 $\{T_i\}_{i=1}^n$ 以此来构造可变换算子值核。

2 学习可变换算子值核

本节先利用核对齐的方法针对单标记问题学习了全局特征对标记的重要度分布。再利用主成分分析 (principal component analysis, PCA)^[30] 提取了单个样例的样例级特征重要度分布。最后, 利用两个样例对应的样例级标记相关性矩阵的乘积作为描述两个样例标记相关性之间交互信息的交互矩阵, 并证明其不仅是偏迹核而且是可变换算子值核。

2.1 基于核对齐的全局特征对标记的重要度分布

本小节利用核对齐的方法针对单标记问题学习全局特征对标记的重要度分布。

令 $D = \{(x_1, y), (x_2, y), \dots, (x_n, y)\}$ 是一个包含 n 个样例的关于标记的单标记数据集, 其中 $x_i = (x_i^1, x_i^2, \dots, x_i^m)^T \in X$ 表示每个样例有 m 维特征, 单标记 $y = (y_1, y_2, \dots, y_n)^T$ 。

在特征空间中, 每个特征所对应高斯核为 $K_q(x_i, x_j) = \exp \frac{-\|a_q(x_i) - a_q(x_j)\|_2^2}{\sigma^2}$, 则有核矩阵 K_q 。对于 m 个核矩阵 $K_q, q = 1, 2, \dots, m$ 线性凸组合构成的多核矩阵为 $K = \sum_{q=1}^m \mu_q K_q, \mu_q \geq 0$ 。针对单标记数据集, 其理想核为 $K_{\text{ideal}} = yy^T$, 其中 $y \in \{+1, -1\}^n$ 。为刻画特征空间和标记向量之间的相似性, 通过特征空间对应线性组合核 $K = \sum_{q=1}^m \mu_q K_q$ 与标记向量对应理想核 K_{ideal} 之间的对齐值来实现:

$$\max_{\mu} f(\mu) = A(K, K_{\text{ideal}})$$

$$\text{s.t.} \sum_{q=1}^m \mu_q = 1, \mu_q \geq 0.$$

式中: f 为向量值函数, F 为 Frobenius 范数,

$$A(K, K_{\text{ideal}}) = \frac{\langle K, K_{\text{ideal}} \rangle_F}{\|K\|_F \|K_{\text{ideal}}\|_F} = \frac{\left\langle \sum_{q=1}^m \mu_q K_q, yy^T \right\rangle_F}{\left\| \sum_{q=1}^m \mu_q K_q \right\|_F \|yy^T\|_F}$$

由于 yy^T 是固定的, 相应的 $\|K_{\text{ideal}}\|_F$ 是常数。显然该项对求解优化问题没有影响, 则简化为

$$\max_{\mu} f(\mu) = \frac{\left\langle \sum_{q=1}^m \mu_q K_q, yy^T \right\rangle_F}{\left\| \sum_{q=1}^m \mu_q K_q \right\|_F}$$

$$\text{s.t. } \sum_{q=1}^m \mu_q = 1, \quad \mu_q \geq 0$$

进一步,上述优化问题可以转化为求解以下最小化形式的优化问题:

$$\begin{aligned} \min_{\mu} f(\mu) &= \lambda \left\langle \sum_{q=1}^m \mu_q \mathbf{K}_q, \sum_{q=1}^m \mu_q \mathbf{K}_q \right\rangle_F - \\ &\quad \left\langle \sum_{q=1}^m \mu_q \mathbf{K}_q, \mathbf{y}\mathbf{y}^T \right\rangle_F \\ \text{s.t. } \sum_{q=1}^m \mu_q &= 1, \quad \mu_q \geq 0 \end{aligned}$$

式中 λ 是用于平衡两项的参数。因为 $f(\mu) = \sum_{q,h} (\lambda \langle \mathbf{K}_h, \mathbf{K}_q \rangle_F) \cdot \mu_h \mu_q - \sum_{q=1}^m \langle \mathbf{K}_q, \mathbf{y}\mathbf{y}^T \rangle_F \cdot \sum_{q,h} (\lambda \langle \mathbf{K}_h, \mathbf{K}_q \rangle_F) \cdot \mu_h \mu_q = \lambda \left\langle \sum_{h=1}^m \mu_h \mathbf{K}_h, \sum_{q=1}^m \mu_q \mathbf{K}_q \right\rangle_F - \lambda \left\| \sum_{q=1}^m \mu_q \mathbf{K}_q \right\|_F \geq 0$ 。

上述优化问题可参照支持向量机 (support vector machine, SVM) 对偶问题的方法利用二次规划问题求解^[31]。据此可以求得特征空间中全局特征对标记重要度分布,并且此分布随训练数据变化而调整。

2.2 利用主成分分析学习样例级特征重要度分布

全局特征对标记重要度分布反映了整体数据集上特征与标记之间的关联,而样例级特征重要度分布则强调在具体的样例上特征与标记之间的关联。本小节针对单标记问题,利用 PCA 降维提取针对单个样例 \mathbf{x}_i 的特征集合对单个标记的重要度分布,即样例级特征重要度分布,其反映了数据分布的协方差结构提取样本特异性特征。

利用 PCA 降维挖掘样例点 \mathbf{x}_i 处样例级特征重要度分布的具体步骤如下。

1) 根据 2.1 节可得特征对标记 l_α 的全局重要度分布:

$$\boldsymbol{\mu}_\alpha(\mathbf{X}) = (\mu_{\alpha 1}(\mathbf{X}), \mu_{\alpha 2}(\mathbf{X}), \dots, \mu_{\alpha m}(\mathbf{X})), \alpha = 1, 2, \dots, N.$$

2) 针对缺失样例点 \mathbf{x}_i , 利用核对齐方法度量缺失点处特征对标记重要度分布:

$$\boldsymbol{\mu}_\alpha(\mathbf{X} - \mathbf{x}_i) = (\mu_{\alpha 1}(\mathbf{X} - \mathbf{x}_i), \mu_{\alpha 2}(\mathbf{X} - \mathbf{x}_i), \dots, \mu_{\alpha m}(\mathbf{X} - \mathbf{x}_i)).$$

3) 令 $\mathbf{B}(i) = \begin{pmatrix} \boldsymbol{\mu}_\alpha(\mathbf{X}) \\ \boldsymbol{\mu}_\alpha(\mathbf{X} - \mathbf{x}_i) \end{pmatrix} \in \mathbf{R}^{2 \times m}$, 对矩阵 $\mathbf{B}(i)$ 的每一行进行零均值化, 可得新矩阵 $\mathbf{B}^*(i)$ 。然后, 通过求解协方差矩阵 $\mathbf{C}(i) = \mathbf{B}^*(i)(\mathbf{B}^*(i))^T \in \mathbf{R}^{2 \times 2}$ 的特征值及其对应的特征向量。同时, 将其特征向量按对应特征值的大小排列成矩阵, 取第一行向量 $\mathbf{P}(i)$ 与 $\mathbf{B}^*(i)$ 相乘并归一化可得样例级特征重要度分布:

$$\boldsymbol{\mu}_\alpha(\mathbf{x}_i) = (\mu_{\alpha 1}(\mathbf{x}_i), \mu_{\alpha 2}(\mathbf{x}_i), \dots, \mu_{\alpha m}(\mathbf{x}_i)).$$

利用上述样例级特征重要度分布, 可以构建样例级标记相关性矩阵 \mathbf{K}_{x_i} :

$$\mathbf{K}_{x_i} = [\boldsymbol{\mu}_\alpha(\mathbf{x}_i) (\boldsymbol{\mu}_\beta(\mathbf{x}_i))^T]_{N \times N}, \alpha, \beta = 1, 2, \dots, N.$$

该矩阵中每个元素刻画了其所对应的两个标记在样例 \mathbf{x}_i 上的样例级标记相关性。

2.3 基于样例级特征重要度分布的可变换算子值核

在本小节中, 利用两个样例对应的样例级标记相关性矩阵的乘积作为描述两个样例之间标记相关性交互信息的交互矩阵。算子值核 $\mathbf{K}(\mathbf{x}, \mathbf{z})$ 是从 $\mathbf{X} \times \mathbf{X}$ 到 $\mathbf{L}(\mathbf{Y})$ 取值为上述交互矩阵的二元函数, 可以证明 $\mathbf{K}(\mathbf{x}, \mathbf{z})$ 是正定算子值核, 并可以进一步证明 $\mathbf{K}(\mathbf{x}, \mathbf{z})$ 是偏迹核和可变换算子值核。

定义 5 令 $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{K}_{x_i}^T \mathbf{K}_{x_j}$, 称它为样例 \mathbf{x}_i 和 \mathbf{x}_j 之间标记相关性的交互矩阵。

$$\text{有 } \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{K}_{x_i}^T \mathbf{K}_{x_j} =$$

$$\begin{aligned} &\begin{bmatrix} \boldsymbol{\mu}_1(\mathbf{x}_i) (\boldsymbol{\mu}_1(\mathbf{x}_i))^T & \cdots & \boldsymbol{\mu}_1(\mathbf{x}_i) (\boldsymbol{\mu}_N(\mathbf{x}_i))^T \\ \vdots & \ddots & \vdots \\ \boldsymbol{\mu}_N(\mathbf{x}_i) (\boldsymbol{\mu}_1(\mathbf{x}_i))^T & \cdots & \boldsymbol{\mu}_N(\mathbf{x}_i) (\boldsymbol{\mu}_N(\mathbf{x}_i))^T \end{bmatrix} \times \\ &\begin{bmatrix} \boldsymbol{\mu}_1(\mathbf{x}_j) (\boldsymbol{\mu}_1(\mathbf{x}_j))^T & \cdots & \boldsymbol{\mu}_1(\mathbf{x}_j) (\boldsymbol{\mu}_N(\mathbf{x}_j))^T \\ \vdots & \ddots & \vdots \\ \boldsymbol{\mu}_N(\mathbf{x}_j) (\boldsymbol{\mu}_1(\mathbf{x}_j))^T & \cdots & \boldsymbol{\mu}_N(\mathbf{x}_j) (\boldsymbol{\mu}_N(\mathbf{x}_j))^T \end{bmatrix} = \\ &\begin{bmatrix} \boldsymbol{\mu}_1(\mathbf{x}_i) \mathbf{M} (\boldsymbol{\mu}_1(\mathbf{x}_j))^T & \cdots & \boldsymbol{\mu}_1(\mathbf{x}_i) \mathbf{M} (\boldsymbol{\mu}_N(\mathbf{x}_j))^T \\ \vdots & \ddots & \vdots \\ \boldsymbol{\mu}_N(\mathbf{x}_i) \mathbf{M} (\boldsymbol{\mu}_1(\mathbf{x}_j))^T & \cdots & \boldsymbol{\mu}_N(\mathbf{x}_i) \mathbf{M} (\boldsymbol{\mu}_N(\mathbf{x}_j))^T \end{bmatrix} \end{aligned}$$

其中矩阵 $\mathbf{M} = \sum_{c=1}^N (\boldsymbol{\mu}_c(\mathbf{x}_i))^T \boldsymbol{\mu}_c(\mathbf{x}_j)$, $\boldsymbol{\mu}_\alpha(\mathbf{x}_i), \boldsymbol{\mu}_\beta(\mathbf{x}_i), \alpha, \beta = 1, 2, \dots, N$ 为样例级特征重要度分布。矩阵 $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$ 利用样例级特征重要度分布描述了两两标记相关性矩阵之间的关联关系, 从而刻画了样例之间标记相关性的交互信息。

定理 1 $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$ 是半正定算子值核。

证明 由定义 2.1 可知

$$\sum_{c=1}^N (\boldsymbol{\mu}_c(\mathbf{x}_i))^T \boldsymbol{\mu}_c(\mathbf{x}_j) =$$

$$\begin{bmatrix} \sum_{c=1}^N \mu_{c1}(\mathbf{x}_i) \mu_{c1}(\mathbf{x}_j) & \cdots & \sum_{c=1}^N \mu_{c1}(\mathbf{x}_i) \mu_{cm}(\mathbf{x}_j) \\ \vdots & \ddots & \vdots \\ \sum_{c=1}^N \mu_{cm}(\mathbf{x}_i) \mu_{c1}(\mathbf{x}_j) & \cdots & \sum_{c=1}^N \mu_{cm}(\mathbf{x}_i) \mu_{cm}(\mathbf{x}_j) \end{bmatrix}$$

令向量

$$\mathbf{S}^d(\mathbf{x}_i) = [\mu_{d1}(\mathbf{x}_i), \mu_{d2}(\mathbf{x}_i), \dots, \mu_{dN}(\mathbf{x}_i)]^T, d = 1, 2, \dots, m,$$

那么矩阵

$$\sum_{c=1}^N (\mu_c(\mathbf{x}_i))^T \mu_c(\mathbf{x}_j) =$$

$$\begin{bmatrix} (\mathbf{S}^1(\mathbf{x}_i))^T \mathbf{S}^1(\mathbf{x}_j) & \cdots & (\mathbf{S}^1(\mathbf{x}_i))^T \mathbf{S}^m(\mathbf{x}_j) \\ \vdots & \ddots & \vdots \\ (\mathbf{S}^m(\mathbf{x}_i))^T \mathbf{S}^1(\mathbf{x}_j) & \cdots & (\mathbf{S}^m(\mathbf{x}_i))^T \mathbf{S}^m(\mathbf{x}_j) \end{bmatrix}$$

因此, 简化矩阵 $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$ 中的每个元素

$$\mu_\alpha(\mathbf{x}_i) \sum_{c=1}^N (\mu_c(\mathbf{x}_i))^T \mu_c(\mathbf{x}_j) (\mu_\beta(\mathbf{x}_j))^T \text{ 可得}$$

$$\begin{bmatrix} \mu_{\alpha 1}(\mathbf{x}_i) \mu_{\alpha 2}(\mathbf{x}_i) & \cdots & \mu_{\alpha m}(\mathbf{x}_i) \\ \mathbf{S}^1(\mathbf{x}_j) \mathbf{S}^2(\mathbf{x}_j) & \cdots & \mathbf{S}^m(\mathbf{x}_j) \end{bmatrix} \begin{bmatrix} (\mathbf{S}^1(\mathbf{x}_i))^T \\ (\mathbf{S}^2(\mathbf{x}_i))^T \\ \vdots \\ (\mathbf{S}^m(\mathbf{x}_i))^T \end{bmatrix} = \begin{bmatrix} \mu_{\beta 1}(\mathbf{x}_j) \\ \mu_{\beta 2}(\mathbf{x}_j) \\ \vdots \\ \mu_{\beta m}(\mathbf{x}_j) \end{bmatrix}$$

$$\begin{bmatrix} \sum_{d=1}^m \mu_{\alpha d}(\mathbf{x}_i) \mu_{1d}(\mathbf{x}_i) \sum_{d=1}^m \mu_{\alpha d}(\mathbf{x}_i) \mu_{2d}(\mathbf{x}_i) \cdots \\ \sum_{d=1}^m \mu_{\alpha d}(\mathbf{x}_i) \mu_{Nd}(\mathbf{x}_i) \end{bmatrix}$$

$$\begin{bmatrix} \sum_{d=1}^m \mu_{\beta d}(\mathbf{x}_j) \mu_{1d}(\mathbf{x}_j) \\ \vdots \\ \sum_{d=1}^m \mu_{\beta d}(\mathbf{x}_j) \mu_{Nd}(\mathbf{x}_j) \end{bmatrix} = \phi_\alpha(\mathbf{x}_i) (\phi_\beta(\mathbf{x}_j))^T$$

其中 $\phi_\alpha(\mathbf{x}_i) =$

$$\begin{bmatrix} \sum_{d=1}^m \mu_{\alpha d}(\mathbf{x}_i) \mu_{1d}(\mathbf{x}_i) \sum_{d=1}^m \mu_{\alpha d}(\mathbf{x}_i) \mu_{2d}(\mathbf{x}_i) \cdots \sum_{d=1}^m \mu_{\alpha d}(\mathbf{x}_i) \mu_{Nd}(\mathbf{x}_i) \end{bmatrix}$$

并且 $\alpha, \beta = 1, 2, \dots, N$ 。综上所述, $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) =$

$$\begin{bmatrix} \phi_1(\mathbf{x}_i) \\ \phi_2(\mathbf{x}_i) \\ \vdots \\ \phi_N(\mathbf{x}_i) \end{bmatrix} \begin{bmatrix} (\phi_1(\mathbf{x}_j))^T & (\phi_2(\mathbf{x}_j))^T & \cdots & (\phi_N(\mathbf{x}_j))^T \end{bmatrix}.$$

由算子值核的定义可知:

$$\sum_{i,j} \langle \mathbf{y}_i, \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) \mathbf{y}_j \rangle = \sum_{i,j} \mathbf{y}_i^T \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) \mathbf{y}_j =$$

$$\sum_{i,j} \mathbf{y}_i^T \begin{bmatrix} \phi_1(\mathbf{x}_i) \\ \phi_2(\mathbf{x}_i) \\ \vdots \\ \phi_N(\mathbf{x}_i) \end{bmatrix} \begin{bmatrix} (\phi_1(\mathbf{x}_j))^T & (\phi_2(\mathbf{x}_j))^T & \cdots & (\phi_N(\mathbf{x}_j))^T \end{bmatrix} \mathbf{y}_j = \sum_{i,j} \langle \phi(\mathbf{x}_i) \mathbf{y}_i, \phi(\mathbf{x}_j) \mathbf{y}_j \rangle =$$

$$\left\| \sum_i [\phi_1(\mathbf{x}_i)^T \ \phi_2(\mathbf{x}_i)^T \ \cdots \ \phi_N(\mathbf{x}_i)^T] \mathbf{y}_i \right\|^2 \geq 0$$

其中 $\phi(\mathbf{x}_i) = [\phi_1(\mathbf{x}_i)^T \ \phi_2(\mathbf{x}_i)^T \ \cdots \ \phi_N(\mathbf{x}_i)^T]$, 则 $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$ 是半正定的核。

定理 2 $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$ 是偏迹核。

证明 由定义 5 可知 $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) =$

$$\begin{bmatrix} \mu_1(\mathbf{x}_i) \mathbf{M}(\mu_1(\mathbf{x}_j))^T & \cdots & \mu_1(\mathbf{x}_i) \mathbf{M}(\mu_N(\mathbf{x}_j))^T \\ \vdots & \ddots & \vdots \\ \mu_N(\mathbf{x}_i) \mathbf{M}(\mu_1(\mathbf{x}_j))^T & \cdots & \mu_N(\mathbf{x}_i) \mathbf{M}(\mu_N(\mathbf{x}_j))^T \end{bmatrix},$$

其中矩阵 $\mathbf{M} = \sum_{c=1}^N (\mu_c(\mathbf{x}_i))^T \mu_c(\mathbf{x}_j)$ 。

算子值核 $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$ 中每个元素 $\mu_\alpha(\mathbf{x}_i) \mathbf{M}(\mu_\beta(\mathbf{x}_j))^T$ 可写成迹的形式, 具体为

$$\mu_\alpha(\mathbf{x}_i) \mathbf{M}(\mu_\beta(\mathbf{x}_j))^T = \text{tr} \left([(\mu_\beta(\mathbf{x}_j))^T \mu_\alpha(\mathbf{x}_i)] \cdot \mathbf{M} \right) \quad \alpha, \beta = 1, 2, \dots, N$$

则算子值核 $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$ 可以改写成

$$\begin{bmatrix} \text{tr}(\mathbf{V}_{11} \mathbf{M}) & \cdots & \text{tr}(\mathbf{V}_{N1} \mathbf{M}) \\ \vdots & \ddots & \vdots \\ \text{tr}(\mathbf{V}_{1N} \mathbf{M}) & \cdots & \text{tr}(\mathbf{V}_{NN} \mathbf{M}) \end{bmatrix},$$

其中矩阵

$$\mathbf{V}_{\alpha\beta} = (\mu_\beta(\mathbf{x}_j))^T \mu_\alpha(\mathbf{x}_i), \quad \alpha, \beta = 1, 2, \dots, N$$

$$\mathbf{M} = \sum_{c=1}^N (\mu_c(\mathbf{x}_i))^T \mu_c(\mathbf{x}_j)$$

根据偏迹核的定义,

$$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \text{tr} \begin{bmatrix} \mathbf{V}_{11} \mathbf{M} & \cdots & \mathbf{V}_{N1} \mathbf{M} \\ \vdots & \ddots & \vdots \\ \mathbf{V}_{1N} \mathbf{M} & \cdots & \mathbf{V}_{NN} \mathbf{M} \end{bmatrix} = \begin{bmatrix} \text{tr}(\mathbf{V}_{11} \mathbf{M}) & \cdots & \text{tr}(\mathbf{V}_{N1} \mathbf{M}) \\ \vdots & \ddots & \vdots \\ \text{tr}(\mathbf{V}_{1N} \mathbf{M}) & \cdots & \text{tr}(\mathbf{V}_{NN} \mathbf{M}) \end{bmatrix}$$

则算子值核 $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$ 是偏迹核。

定理 3 $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$ 是可变换算子值核。

证明 由定理 1 可知, 算子值核 $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$ 可以

$$\text{写作: } \begin{bmatrix} \phi_1(\mathbf{x}_i) \\ \phi_2(\mathbf{x}_i) \\ \vdots \\ \phi_N(\mathbf{x}_i) \end{bmatrix} \begin{bmatrix} (\phi_1(\mathbf{x}_j))^T & (\phi_2(\mathbf{x}_j))^T & \cdots & (\phi_N(\mathbf{x}_j))^T \end{bmatrix}.$$

由可变换算子值核的定义可知 $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$ 满足条件,

$\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = [\langle \mathbf{T}_h(\mathbf{x}_i), \mathbf{T}_l(\mathbf{x}_j) \rangle]_{h,l=1}^N$ 。其中映射 $\mathbf{T}_h(\mathbf{x}_i) = \phi_h(\mathbf{x}_i)$ 和 $\mathbf{T}_l(\mathbf{x}_j) = \phi_l(\mathbf{x}_j)$, 并且 $h, l = 1, 2, \dots, N$ 和 $i, j = 1, 2, \dots, n$ 。综上所述, $\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)$ 是可变换算子值核。

3 基于可变换算子值核的学习算法

文献 [32-33] 中提出了求预测函数 $\mathbf{g}(\mathbf{x})$ 的算法, 本节基于可变换算子值核将其运用到多标记学习中。对给定数据集可以求得算子值核矩阵 \mathbf{K} , 针对新进样例 \mathbf{x}_τ 预测函数具有如下形式^[33]:

$$\mathbf{g}(\mathbf{x}_\tau) = \mathbf{K}^T(\mathbf{x}_\tau, \mathbf{x}_\tau) \cdot \mathbf{k}_\tau^T (\mathbf{K} + n\lambda \mathbf{I})^{-1} \mathbf{Y}$$

其中

$$Y = [y_1, y_2, \dots, y_n]^T \in L(Y)^n$$

$$K = [K(x_i, x_j)]_{i,j=1}^n \in L(Y)^{n \times n}$$

$$k_{x_\tau}^T = [K(x_1, x_\tau), K(x_2, x_\tau), \dots, K(x_n, x_\tau)] \in L(Y)^{1 \times n}$$

在 $k_{x_\tau}^T$ 中 $K(x_i, x_\tau) = K_{x_i}^T K_{x_\tau}$, $i = 1, 2, \dots, n$ 。

当计算 $k_{x_\tau}^T$ 和算子 $K(x_\tau, x_\tau)$ 时, 需依赖于新进样例 x_τ 的样例级特征重要度分布构建标记相关性矩阵 k_{x_τ} 。因此, 对于任意新进样例 $x_\tau \in X$, 首先根据岭回归算法度量其特征对标记 $l_\alpha, \alpha = 1, 2, \dots, N$ 样例级特征重要度分布

$$\mu_\alpha(x_\tau) = (\mu_{\alpha 1}(x_\tau), \mu_{\alpha 2}(x_\tau), \dots, \mu_{\alpha n}(x_\tau)),$$

目标函数为

$$f^\lambda = \arg \min_{f \in H} \frac{1}{n} \|f(x_i) - z_i\|^2 + \lambda \|W\|^2. \quad (1)$$

式中预测函数 $f(x_i) = W \cdot x_i$ 。对于新进样例 x_τ , 通过式 (1), 可以求解关于特征对标记的权重矩阵 $W \in \mathbf{R}^{m \times m}$ 。 λ 是正则化系数, $Z = [z_1, z_2, \dots, z_n]^T \in \mathbf{R}^{m \times n}$, 其中 $z_i = \mu_\alpha(x_i), i = 1, 2, \dots, n$ 表示特征对标记 l_α 的样例级特征重要度分布。该优化问题求解过程为: 在特征空间中, 定义输入数据矩阵 F , 其第 i 列为 x_i 。在标记空间中, 定义目标矩阵 Z , 其第 i 列为 z_i 。分别计算输入数据的自相关矩阵 $FF^T \in \mathbf{R}^{m \times m}$ 和目标与输入的互相关矩阵 $ZF^T \in \mathbf{R}^{m \times m}$ 。则最优权重矩阵为 $W = ZF^T (FF^T + n\lambda I)^{-1}$ 。其中 I 是单位矩阵, $n\lambda$ 是缩放后的正则化系数。最后, 预测新进点 x_τ 的样例级特征重要度分布:

$$f(x_\tau) = W \cdot x_\tau = \mu_\alpha(x_\tau), \quad \alpha = 1, 2, \dots, N$$

并对其进行归一化处理。则新进样例点的样例级标记相关性矩阵表示为

$$K_{x_\tau} = [(\mu_\alpha(x_\tau))^T \mu_\beta(x_\tau)]_{N \times N}, \quad \alpha, \beta = 1, 2, \dots, N.$$

本文所设计算法的具体步骤在算法 1 中。算法 1 的空间复杂度主要由算子值核矩阵来决定。下面将本文所设计的算法的空间复杂度与其他算法进行比较。ML-KNN 算法 (multi-label K-nearest neighbors algorithm)^[13] 的空间复杂度为 $O(n \cdot (m+N) + n \cdot N \cdot k)$, 其中 n 为样本数量, m 为特征向量的维数, N 为标记数量, k 为相近邻样本数量。MDDM 算法 (multilabel dimensionality reduction via dependence maximization)^[14] 的空间复杂度为 $O(N^2 + m^2 + mN + mp)$, 其中 m 为原始特征维数, p 为降维后的目标维数, N 为标记数量。RMFRS 算法 (feature selection for multi-label learning based on kernelized fuzzy rough sets)^[15] 的空间复杂度为 $O(n^2 + mp)$, 其中 m 为特征向量的维数, n 为样本数量, p 为降维后的目标维数。PMFS 算法 (Pareto-based feature selection algorithm for multi-la-

bel classification)^[16] 的空间复杂度为 $O(n^2 N^2 + n^2)$, 其中 n 为样本数量, N 为标记数量。本文所提算法的整体空间复杂度为 $O(n^2 N^2 + m^2)$, 其中 n 为样本数量, N 为标记数量, m 为特征向量的维数。此算法的整体空间复杂度由两部分组成: 1) 算子值核的空间复杂度为 $O(N^2)$; 2) 核岭回归方法的空间复杂度为 $O(n^2 N^2)$ 。并且此算法的整体空间复杂度不如 ML-KNN、MDDM、RMFRS 算法, 优于 PMFS 算法。此算法的空间复杂度较高是核岭回归问题导致的。而本文所设计的算子值核的空间复杂度与 MDDM 算法相同, 优于其他算法的空间复杂度。

算法 1 可变换算子核学习算法 (TRANS-OVK 算法)

输入 多标记数据集 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$; $X = (x_1, x_2, \dots, x_n)$ 表示特征空间, $Y = (y_1, y_2, \dots, y_n)$ 表示标记空间; $x_i = (x_i^1, x_i^2, \dots, x_i^m)^T$ 表示特征空间中第 i 个样例所对应的向量, $y_i = (y_i^1, y_i^2, \dots, y_i^N)^T$ 表示标记空间中第 i 个样例所对应的向量。

输出 预测函数 $g(x_\tau)$ 。

- 1) 计算组合核矩阵 $K = \sum_{q=1}^m \alpha_q K_q$, $\alpha_q \geq 0$ 和理想核 $K_{ideal} = yy^T$;
- 2) 基于核对齐, 由数据标记结构动态优化生成全局特征重要度分布 $\mu_\alpha = (\mu_{\alpha 1}, \mu_{\alpha 2}, \dots, \mu_{\alpha m})$;
- 3) 计算缺失点特征对标记的重要度分布 $\mu_\alpha(X - x_i) = (\mu_{\alpha 1}(X - x_i), \mu_{\alpha 2}(X - x_i), \dots, \mu_{\alpha m}(X - x_i))$;
- 4) 利用 PCA 降维提取样例级特征重要度分布 $\mu_\alpha(x_i) = (\mu_{\alpha 1}(x_i), \mu_{\alpha 2}(x_i), \dots, \mu_{\alpha m}(x_i))$, 其反映了数据分布的协方差结构自动提取样本特异性特征;
- 5) 基于样例级重要度分布构造算子值核 $K(x_i, x_j) = K_{x_i}^T K_{x_j}$, 其中样例级标记相关性矩阵 $K_{x_i} = [\mu_\alpha(x_i)^T \mu_\beta(x_i)]_{N \times N}$, 并且算子值核矩阵 $K = [K(x_i, x_j)]_{i,j=1}^n$ 。算子值核 $K(x_i, x_j)$ 完全依赖样本局部特征重要度, 实现样本自适应的核学习;
- 6) 针对新进点 x_τ , 通过求解岭回归问题, 计算关于特征对标记的权重矩阵 W ;
- 7) 预测新进点 x_τ 的样例级特征重要度分布 $\mu_\alpha(x_\tau) = W \cdot x_\tau, \alpha = 1, 2, \dots, N$;
- 8) 计算新进点 x_τ 的可变换算子值核 $K(x_\tau, x_\tau) = K_{x_\tau}^T K_{x_\tau}$, 其中 $K_{x_\tau} = [(\mu_\alpha(x_\tau))^T \mu_\beta(x_\tau)]_{N \times N}$;
- 9) 计算 $k_{x_\tau}^T = [K(x_1, x_\tau), \dots, K(x_n, x_\tau)]$, 其中 $K(x_i, x_\tau) = K_{x_i}^T K_{x_\tau}, i = 1, 2, \dots, n$;
- 10) 通过求解核岭回归问题, 得到预测结果 $g(x_\tau) = K^T(x_\tau, x_\tau) \cdot k_{x_\tau}^T (K + n\lambda I)^{-1} Y$ 。

4 实验验证

本章根据第 1 章所提 5 项评价指标, 给出 4 种高性能对比算法和本文提出算法在 9 个数据集上的数值实验结果。

4.1 实验设置

多标记分类算法旨在从多标记训练数据集中挖掘信息, 构建目标函数, 然后预测任何测试样例的多维语义输出。为验证多标记学习算法的分类有效性, 实验选取了 9 个数据集。分别为 birds、business、cal500、emotion、genbase、image、reuters、scene、yeast。其中, birds 数据集是声音的集合, cal500, emotion 代表音乐中的数据集, genbase、image 和 scene 是图像信息组成的数据集, reuters 和 business 是文本分类数据集, yeast 是一种与基因功能的分类有关的生物数据集。每个数据集的详细情况总结在表 1 中。上述所有多标记数据集均来自开源数据库 Mulan Library。本文首次将算子值核应用到多标记学习领域中, 目前尚无其他与算子值核相关的多标记学习算法, 因此选择 4 种具有代表性的多标记学习算法进行比较, 包括 ML-KNN^[13]、MDDM^[14]、RMFRS^[15] 及 PMFS^[16]。

本文采用标准化实验流程: 对于每个多标记数据集, 通过 0-1 归一化方法对输入变量进行规范化处理。在实验设计方面, 每个数据集均进行多次重复实验以验证分类结果的稳定性。训练数据集和测试数据集按 4:1 进行随机划分, 其中占比为 80% 的训练样本用于模型训练, 剩余占比 20%

的测试集用于评估模型性能。最后, 实验结果通过计算多次独立实验测量指标的算术平均值及其标准差进行呈现, 以此确保实验结论的统计可靠性。

表 1 多标记数据集
Table 1 Multi-label data set

数据集	样例数量	特征数量	标记数量
birds	645	260	20
business	5 000	438	30
cal500	502	68	174
emotion	593	72	6
genbase	662	1 185	27
image	2 000	135	5
reuters	2 000	243	7
scene	2 407	294	6
yeast	2 417	103	14

4.2 实验结果与分析

从表 2 可以看出 TRANS-OVK (transformable operator-valued Kernel) 算法在多个数据集上都呈现出分类性能优势, 尤其在数据集 birds、business、emotion、genbase、image、scene、yeast 上, 以 AVP 为评价指标, 是优于其他算法的。并且 TRANS-OVK 算法和 PMFS 算法在数据集 business、emotion 和 genbase 上分类性能结果相近。在数据集 cal500 上, ML-KNN 算法明显比 TRANS-OVK 算法和其他 3 个对比算法表现更为优异, 并且 PMFS 算法在数据集 reuters 上展现出的性能是优于其他算法的。

表 2 在评价指标 AVP(↑) 下 TRANS-OVK 算法与 4 个对比算法的实验结果比较
Table 2 Comparison between TRANS-OVK algorithm and four comparison algorithms on AVP(↑)

数据集	MDDM	ML-KNN	RMFRS	PMFS	TRANS-OVK
birds	0.708 8±0.021 3	0.581 6±0.004 8	0.678 7±0.036 3	0.707 1±0.032 7	0.723 7±0.013 9
business	0.871 6±0.007 1	0.862 7±0.000 3	0.873 0±0.006 3	0.875 6±0.004 7	0.877 1±0.004 2
cal500	0.485 7±0.004 9	0.498 1±0.003 0	0.482 8±0.004 4	0.493 1±0.003 9	0.490 3±0.005 6
emotion	0.744 0±0.017 2	0.583 1±0.012 3	0.749 9±0.050 9	0.754 0±0.016 6	0.756 2±0.035 8
genbase	0.751 1±0.152 0	0.751 3±0.147 1	0.456 1±0.055 3	0.984 3±0.031 3	0.987 1±0.029 5
image	0.697 8±0.042 4	0.518 1±0.008 8	0.657 4±0.054 7	0.692 1±0.055 8	0.697 9±0.035 7
reuters	0.826 0±0.041 6	0.601 2±0.002 0	0.814 5±0.041 7	0.842 0±0.035 6	0.825 4±0.026 9
scene	0.744 8±0.079 6	0.318 1±0.046 0	0.589 7±0.151 9	0.741 6±0.093 0	0.752 2±0.024 2
yeast	0.744 6±0.014 1	0.710 2±0.001 2	0.733 2±0.015 9	0.748 7±0.010 6	0.754 3±0.032 8

注: 加粗表示本行最优结果。

从表 3 可以看出 MDDM 算法在数据集 birds 和 scene 上, 以 RL 为评价指标, 所呈现出的分类效果是优于其他算法的, 而 PMFS 算法在数据集

emotion 上的表现是最好的。在其余数据集上, 本文所提出的 TRANS-OVK 算法明显优于其他 4 个对比算法。

表 3 在评价指标 RL(\downarrow) 下 TRANS-OVK 算法与 4 个对比算法的实验结果比较
Table 3 Comparison between TRANS-OVK algorithm and four comparison algorithms on RL(\downarrow)

数据集	MDDM	ML-KNN	RMFRS	PMFS	TRANS-OVK
birds	0.1152±0.0079	0.1861±0.0038	0.1363±0.0241	0.1217±0.0206	0.1153±0.0136
business	0.0427±0.0029	0.0487±0.0003	0.0418±0.0030	0.0410±0.0024	0.0401±0.0013
cal500	0.1853±0.0016	0.1813±0.0007	0.1860±0.0018	0.1819±0.0018	0.1810±0.0014
emotion	0.2276±0.0184	0.4265±0.0171	0.2229±0.0640	0.2031±0.0205	0.2173±0.0395
genbase	0.0617±0.0371	0.0580±0.0337	0.1547±0.0187	0.0083±0.0068	0.0078±0.0044
image	0.2609±0.0445	0.4622±0.0096	0.3016±0.0560	0.2681±0.0592	0.2546±0.0436
reuters	0.1155±0.0278	0.2791±0.0027	0.1238±0.0327	0.1059±0.0261	0.1123±0.0239
scene	0.1399±0.0674	0.6529±0.0734	0.2856±0.1457	0.1542±0.0792	0.1565±0.0697
yeast	0.1804±0.0095	0.2042±0.0010	0.1876±0.0119	0.1808±0.0084	0.1734±0.0114

注: 加粗表示本行最优结果。

从表 4 可以看出 MDDM 算法数据集在 birds 和 scene 上, 以 CV 评价指标, 展现出的效果是最优的。在其余数据集上, 本文所提出 TRANS-OVK 算法明显优于其他 4 个对比算法。

表 4 在评价指标 CV(\downarrow) 下 TRANS-OVK 算法与 4 个对比算法的实验结果比较
Table 4 Comparison between TRANS-OVK algorithm and four comparison algorithms on CV(\downarrow)

数据集	MDDM	ML-KNN	RMFRS	PMFS	TRANS-OVK
birds	3.3770±0.1749	4.8792±0.0798	3.8138±0.5010	3.5624±0.4262	3.5121±0.2160
business	2.3547±0.1105	2.5912±0.0134	2.3374±0.1073	2.3152±0.0869	2.2491±0.0502
cal500	132.1324±0.7145	128.9072±0.8342	132.4331±0.9138	127.5946±0.8853	126.6102±0.9114
emotion	2.1899±0.0936	3.2241±0.0920	2.1488±0.3141	2.0410±0.1136	2.0311±0.2382
genbase	2.3537±1.1237	1.9435±0.9042	4.9291±0.5734	0.6510±0.1924	0.6383±0.2169
image	1.2885±0.1689	2.0978±0.0394	1.4456±0.2168	1.3359±0.2248	1.2396±0.2005
reuters	0.8684±0.1638	1.8229±0.0168	0.9197±0.1958	0.8145±0.1564	0.8982±0.2319
scene	0.7333±0.3369	3.3339±0.3667	1.4602±0.7281	0.8667±0.3958	0.8164±0.2965
yeast	6.4695±0.1433	6.7680±0.0313	6.5256±0.1443	6.5807±0.1168	6.2864±0.1498

注: 加粗表示本行最优结果。

从表 5 可以看出 TRANS-OVK 算法在数据集 birds、business、genbase、image、reuters、scene 上, 以 OE 为评价指标是优于其他算法的。仅在数据集 cal500 和 yeast 上, TRANS-OVK 算法不如 PMFS 算法。同时, 在数据集 emotion 上, TRANS-OVK 算法不如 RMFRS 算法。

表 5 在评价指标 OE(\downarrow) 下 TRANS-OVK 算法与 4 个对比算法的实验结果比较
Table 5 Comparison between TRANS-OVK algorithm and four comparison algorithms on OE(\downarrow)

数据集	MDDM	ML-KNN	RMFRS	PMFS	TRANS-OVK
birds	0.3688±0.0419	0.5698±0.0061	0.3893±0.0409	0.3629±0.0468	0.3397±0.0323
business	0.1265±0.0071	0.1336±0.0003	0.1246±0.0060	0.1245±0.0052	0.1213±0.0045
cal500	0.1289±0.0051	0.0965±0.0040	0.1286±0.0052	0.0900±0.0067	0.0935±0.0049
emotion	0.3445±0.0291	0.5495±0.0279	0.3391±0.0684	0.3739±0.0217	0.3434±0.0323
genbase	0.3358±0.2089	0.3446±0.2023	0.7122±0.0698	0.0139±0.0433	0.0155±0.0564
image	0.4705±0.0649	0.7153±0.0170	0.5285±0.0817	0.4701±0.0820	0.4605±0.0799
reuters	0.2627±0.0687	0.6224±0.0030	0.2783±0.0611	0.2329±0.0533	0.2279±0.0590
scene	0.4376±0.1121	0.9123±0.0420	0.6334±0.1914	0.4285±0.1260	0.4193±0.1203
yeast	0.2345±0.0152	0.2443±0.0016	0.2482±0.0112	0.2340±0.0079	0.2345±0.0115

注: 加粗表示本行最优结果。

从表 6 可以看出 ML-KNN 算法在数据集 cal500 上的结果优于其他算法。在数据集 image 上, PMFS 算法比 TRANS-OVK 算法和其他 3 个对比算法效

果显著。在数据集 yeast 上, MDDM 算法结果最优。在其余数据集上, 本文所提出 TRANS-OVK 算法明显优于其他 4 个对比算法。

表 6 在评价指标 HL(\downarrow) 下 TRANS-OVK 算法与 4 个对比算法的实验结果比较
Table 6 Comparison between TRANS-OVK algorithm and four comparison algorithms on HL(\downarrow)

数据集	MDDM	ML-KNN	RMFRS	PMFS	TRANS-OVK
birds	0.060 8±0.003 3	0.078 2±0.002 0	0.058 5±0.002 4	0.060 2±0.005 4	0.057 3±0.002 1
business	0.027 7±0.000 8	0.028 4±0.000 0	0.027 5±0.000 7	0.027 5±0.000 5	0.027 2±0.000 0
cal500	0.137 1±0.000 9	0.129 9±0.000 8	0.137 6±0.001 2	0.131 2±0.001 1	0.131 1±0.001 2
emotion	0.241 7±0.012 8	0.325 3±0.007 8	0.242 2±0.029 4	0.238 4±0.014 4	0.233 8±0.020 7
genbase	0.028 6±0.015 8	0.029 2±0.015 5	0.047 0±0.004 1	0.006 2±0.002 5	0.004 6±0.004 1
image	0.220 1±0.017 2	0.246 7±0.000 7	0.227 8±0.014 0	0.210 2±0.018 8	0.212 1±0.019 2
reuters	0.085 2±0.019 7	0.166 5±0.001	0.090 1±0.013 2	0.076 8±0.011 9	0.072 1±0.010 9
scene	0.149 3±0.020 7	0.191 5±0.019 8	0.169 5±0.024 2	0.147 6±0.021 8	0.143 4±0.021 6
yeast	0.201 2±0.009 7	0.229 1±0.000 8	0.208 6±0.011 0	0.204 5±0.007 2	0.203 9±0.001 6

注: 加粗表示本行最优结果。

表 2~6 分别给出了在 AVP、HL、OE、CV 和 RL 5 个评价指标下, 5 个多标记算法在 9 个多标记数据集下的实验结果比较。综上分析, 可知本文所提 TRANS-OVK 算法的分类性能在绝大部分数据集上明显优于其他 4 个多标记对比算法。

5 结束语

本文通过利用核对齐方法, 从多标记数据集中学习了算子值核, 并基于此核构建了有效的多标记学习预测模型。研究中, 首先通过核对齐方法得到了样例级特征重要度分布, 以此为基础构造了算子值核。然后, 证明该算子值核不但是偏迹核而且是可变换算子值核, 其对应的核矩阵中的每个分块都能有效刻画样例间标记相关性的交互信息。最后, 基于可变换算子值核设计了多标记学习算法, 并在 9 个多标记数据集上与 4 种高性能算法进行了对比实验。实验结果表明, 所提出的算法在性能上显著优于对比算法, 验证了其在多标记学习中的有效性和优越性。未来将从算法设计上优化现有算法。

参考文献:

- [1] BROUARD C, SHEN Huibin, DÚHRKOP K, et al. Fast metabolite identification with input output kernel regression[J]. *Bioinformatics*, 2016, 32(12): i28–i36.
- [2] KADRI H, RAKOTOMAMONJY A, PREUX P, et al. Multiple operator-valued kernel learning[J]. *Advances in neural information processing systems*, 2012, 25(3): 2429–2437.
- [3] KADRI H, DUFLOS E, PREUX P, et al. Operator-valued kernels for learning from functional response data[J]. *Journal of machine learning research*, 2016, 17(20): 1–54.
- [4] HUUSARI R, KADRI H. Entangled kernels-beyond separability[J]. *Journal of machine learning research*, 2021, 22(24): 1–40.
- [5] DINUZZO F, FUKUMIZU K, HSU C, et al. Low-rank output kernels[J]. *Journal of machine learning research*, 2011, 20: 181–196.
- [6] DINUZZO F, ONG C S, PILLONETTO G, et al. Learning output kernels with block coordinate descent[C]//*Proceedings of the 28th International Conference on Machine Learning*. Bellevue: PMLR, 2011: 49–56.
- [7] ÁLVAREZ M A, ROSASCO L, LAWRENCE N D. Kernels for vector-valued functions: a review[J]. *Foundations and trends® in machine learning*, 2012, 4(3): 195–266.
- [8] LIM N, D'ALCHÉ-BUC F, AULIAC C, et al. Operator-valued kernel-based vector autoregressive models for network inference[J]. *Machine learning*, 2015, 99(3): 489–513.
- [9] GREGOROVÁ M, KALOUSIS A, MARCHAND-MAILLET S. Forecasting and granger modelling with non-linear dynamical dependencies[C]//*Machine Learning and Knowledge Discovery in Databases*. Cham: Springer International Publishing, 2017: 544–558.
- [10] AUDIFFREN J, KADRI H. Online learning with multiple operator-valued kernels[EB/OL]. (2013–11–01)[2025–03–23]. <https://arxiv.org/abs/1311.0222>.
- [11] MINH H Q, BAZZANI L, MURINO V. A unifying framework in vector-valued reproducing kernel Hilbert spaces for manifold regularization and co-regularized multi-view learning[J]. *Journal of machine learning research*, 2016, 17(25): 1–72.

- [12] HUUSARI R, KADRI H, CAPPONI C. Multi-view metric learning in vector-valued kernel spaces[C]//International Conference on Artificial Intelligence and Statistics. Playa Blanca: PMLR, 2018: 415–424.
- [13] ZHANG Minling, ZHOU Zhihua. ML-KNN: a lazy learning approach to multi-label learning[J]. *Pattern recognition*, 2007, 40(7): 2038–2048.
- [14] ZHANG Yin, ZHOU Zhihua. Multilabel dimensionality reduction via dependence maximization[J]. *ACM transactions on knowledge discovery from data*, 2010, 4(3): 1–21.
- [15] LI Yuwen, LIN Yaojin, LIU Jinghua, et al. Feature selection for multi-label learning based on kernelized fuzzy rough sets[J]. *Neurocomputing*, 2018, 318: 271–286.
- [16] HASHEMI A, BAGHER DOWLATSHAHI M, NEZAMABADI-POUR H. An efficient Pareto-based feature selection algorithm for multi-label classification[J]. *Information sciences*, 2021, 581: 428–447.
- [17] HASHEMI A, DOWLATSHAHI M B. MLCR: a fast multi-label feature selection method based on K-means and L2-norm[C]//2020 25th International Computer Conference, Computer Society of Iran. Tehran: IEEE, 2020: 1–7.
- [18] LIN Yaojin, HE Zhuoxin, GUO Lei, et al. Multi-label feature selection via positive or negative correlation[J]. *IEEE transactions on emerging topics in computational intelligence*, 2024, 8(1): 401–415.
- [19] AFDHAL D, ANANTA K W, HARTONO W S. Adverse drug reactions prediction using multi-label linear discriminant analysis and multi-label learning[C]//2020 International Conference on Advanced Computer Science and Information Systems. Depok: IEEE, 2020: 69–76.
- [20] MULIMANI M, MESAROS A. Class-incremental learning for multi-label audio classification[C]//2024 IEEE International Conference on Acoustics, Speech and Signal Processing. Seoul: IEEE, 2024: 916–920.
- [21] CHEN Jiayao, LI Shaoyuan. Class-aware learning for imbalanced multi-label classification[C]//2023 IEEE 5th International Conference on Civil Aviation Safety and Information Technology. Dali: IEEE, 2023: 903–907.
- [22] GE Yuhang, HU Xuegang, LI Peipei, et al. Multi-label learning with data self-augmentation[C]//Neural Information Processing. Singapore: Springer Nature Singapore, 2023: 336–347.
- [23] CHE Xiaoya, CHEN Degang, MI Jusheng. Learning instance-level label correlation distribution for multilabel classification with fuzzy rough sets[J]. *IEEE transactions on fuzzy systems*, 2023, 31(8): 2871–2884.
- [24] CRISTIANINI N, SHAWE-TAYLOR J, ELISSEEFF A, et al. On kernel-target alignment[J]. *Advances in neural information processing systems*, 2002: 367–374.
- [25] DOMINGOS P. A few useful things to know about machine learning[J]. *Communications of the ACM*, 2012, 55(10): 78–87.
- [26] ZHANG Junping, WANG Feiyue, WANG Kunfeng, et al. Data-driven intelligent transportation systems: a survey [J]. *IEEE transactions on intelligent transportation systems*, 2011, 12(4): 1624–1639.
- [27] AMRI M M, ABED S A. The data-driven future of healthcare: a review[J]. *Mesopotamian journal of big data*, 2023, 2023: 68–74.
- [28] SCHAPIRE R E, SINGER Y. BoosTexter: a boosting-based system for text categorization[J]. *Machine learning*, 2000, 39(2): 135–168.
- [29] WU Xizhu, ZHOU Zhihua. A unified view of multi-label performance measures[C]//International Conference on Machine Learning. Sydney: PMLR, 2017: 3780–3788.
- [30] SALEM N, HUSSEIN S. Data dimensional reduction and principal components analysis[J]. *Procedia computer science*, 2019, 163: 292–299.
- [31] CHEN Linlin, CHEN Degang, WANG Hui. Alignment based kernel selection for multi-label learning[J]. *Neural processing letters*, 2019, 49(3): 1157–1177.
- [32] CARMELI C, DE VITO E, TOIGO A. Vector valued reproducing kernel Hilbert spaces of integrable functions and mercer theorem[J]. *Analysis and applications*, 2006, 4(4): 377–408.
- [33] SZABÓ Z, SRIPERUMBUDUR B K, PÓCZOS B, et al. Learning theory for distribution regression[J]. *Journal of machine learning research*, 2016, 17(152): 1–40.

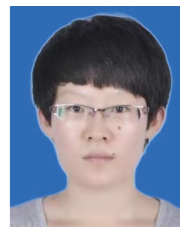
作者简介:



汪振鑫, 博士研究生, 主要研究方向为机器学习。发表学术论文 2 篇。
E-mail: 465095864@qq.com。



陈德刚, 教授, 博士生导师, 主要研究方向为机器学习和数据挖掘。发表学术论文 150 余篇。E-mail: chengdegang@263.net。



车晓雅, 讲师, 主要研究方向为机器学习与数据挖掘。发表学术论文 10 篇。E-mail: chexiaoya@163.com。