

DOI:10.13232/j.cnki.jnju.2022.02.011

## 核化的多视角特权协同随机矢量功能链接网络 及其增量学习方法

吴天宇, 王士同\*

(江南大学人工智能与计算机学院, 无锡, 214122)

**摘要:**在许多实际应用场景中, 可以从不同层次、不同角度获取相同对象的特征数据, 如何有效地利用获取的多视角数据是一个值得研究的问题. 和传统的单视角学习相比, 多视角学习在多源数据的应用中显示了一定的优势. 多角度学习 (Multi-View Learning, MVL) 面临的一个重要问题是在满足不同视角互补性的前提下如何保持视角之间的一致性. 针对以上问题, 提出一种新的多视角特权协同核化随机向量功能链接网络 (KMPRVFL) 来有效地解决多视角分类问题, 其基本思想是将冗余视角的额外信息与平均视角上的特权信息相结合来监督当前视角的分类任务, 将多视角数据用核化后加权线性组合成综合第二视角. 同时, 还设计了一种增量学习方法, 可以有效地减少计算量. 在真实数据集上的实验结果表明, 和传统的多视角学习方法相比, KMPRVFL 的能力更强, 其平均测试精度要优于对比算法.

**关键词:**多视角学习, 特权信息, 随机向量函数链接网络, 增量学习

中图分类号: TP181

文献标志码: A

## Kernel Multi-view Privileged Random vector functional link network and its incremental learning method

Wu Tianyu, Wang Shitong\*

(School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, 214122, China)

**Abstract:** In many practical application scenarios, how to effectively use the multi perspective data obtained from different levels and different angles to obtain the feature data of the same object is a problem worthy of study. Compared with traditional single perspective learning, multi perspective learning shows certain advantages in the application of multi-source data. An important problem in multi-view learning (MVL) is how to keep the consistency of perspectives while satisfying the complementarity of different perspectives. To solve problems above, a new kernel multi-view privileged random vector functional link network (KMPRVFL) is proposed to effectively solve the multi view classification problem. The basic idea is to combine the extra information of redundant perspective with the privileged information of average perspective to supervise the classification task of the current perspective. The multi-view data is combined into a comprehensive second view by weighted linear combination after kernel. At the same time, an incremental learning method is designed to effectively reduce the amount of calculation. Experimental results on real datasets show that KMPRVFL is more powerful than traditional multi-view learning methods. The average test accuracy of KMPRVFL algorithm is better than that of comparison methods.

**Key words:** MVL (Multi-View Learning), privileged information, RVFL (Random Vector Functional Link), incremental learning

基金项目: 国家自然科学基金 (61972181)

收稿日期: 2021-09-24

\* 通讯联系人, E-mail: wxjn00@163.com

多视角学习是近年发展起来的一种机器学习的改进方法,它从多个特定集合中进行数据集成.多视角学习在近几年有很大的发展,也面临新的挑战.它主要利用共识原则或互补原则来确保学习方法的有效性,通过探索不同视角的一致性和互补性,多视角学习比单视图学习更有效,并且具有更好的泛化能力<sup>[1]</sup>.多视角学习在计算机辅助诊断<sup>[2]</sup>上已经有所应用,有一定研究价值.

近年来提出了基于不同策略的多视角分类算法,旨在利用多视角之间的相关信息得到更精确、更高效的分类结果,常用的多视角算法有协同正则化算法和协同训练算法<sup>[3]</sup>.实现协同训练型算法的重要前提是存在充分冗余的视角,协同训练类型算法的目标是最大化不同视角之间一致,代表算法有多训练支持向量机(Multi-Training Support Vector Machine, MTSVM)<sup>[4]</sup>.而在协同正则化类算法中,目标函数需要将不同视角之间的分歧最小化,典型的方法有稀疏多视角支持向量机(Sparse Multi View SVMs)<sup>[5]</sup>、多视角拉普拉斯支持向量机(Multi View Laplacian SVMs)<sup>[6]</sup>.这些基于SVM的方法可以有效地解决多视角的应用场景问题,但也具有一定的局限性,它们忽略了视角之间的互补信息.另一方面,支持向量机求解过程中的局限性,如二次规划求解,对计算机内存需求大,迭代速度慢.

现有的多视角学习算法主要体现了多视角学习的共识原理或互补性原理<sup>[1]</sup>.在多视角学习中,共识和互补性原理在指导模型构建中起着重要作用.共识原理的目的是最大化多个不同视角的一致性,改善算法的泛化误差范围;相反,互补原理强调视角之间共享互补信息,目的是更全面地描述数据,提高算法的鲁棒性.Vapnik提出使用特权信息学习(Learning Using Privileged Information, LUPI)<sup>[7]</sup>来解决学习模型中的补充知识. LUPI将人类的教学理念融入机器学习的概念,对于一个特定的任务,比如分类,训练数据不仅是目标任务的监督信息,还有一些额外的特权信息(附加信息).从多视角学习的角度来看,不同的特征视角可以相互提供特权信息,实现互补<sup>[8]</sup>.

核方法(Kernel Methods)将数据映射到更高维的空间,有助于将原本在低维空间中线性不可

分的数据转换为在高维空间中线性可分的数据,可以大大降低分类任务的难度<sup>[9]</sup>,但是核方法往往会让计算量成倍增多.

本文将特权信息和多视角学习概念结合,在核化(Random Vector Functional Link, RVFL)的基础上提出一种多视角特权协同增量核化随机矢量功能链接网络.该网络在平均情况下相互利用冗余视角的附加信息作为特权信息监督当前视角的分类,并以此设计了KM-PRVFL的目标函数,利用解析解对目标函数进行优化,并提出KM-PRVFL的一种增量方法,可以有效地降低计算量.在多个真实数据集上进行实验,实验结果证明KM-PRVFL优于其他算法.

## 1 相关工作

**1.1 核方法** 核方法的使用可以有效地提高支持向量机一类的算法对线性不可分的数据的处理能力.支持向量机的核心思想是最大化数据点与界限之间的距离(margin)<sup>[10]</sup>,由于距离与权重的范数成反比,因此问题就转换为最小化权重的范数,即:

$$\min \frac{1}{2} \|\omega\|^2, s.t. (\omega^T x) y \geq 1 \quad (1)$$

通过拉格朗日乘子法,可将上述最小化问题转换为它的对偶问题:

$$\begin{aligned} \max_a & \left[ \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j \right] \\ s.t. & \alpha_i \geq 0, \sum_i \alpha_i y_i = 0 \end{aligned} \quad (2)$$

经过观察可以发现,该优化问题已经与权重 $\omega$ 无关,也与单个输入数据 $x_i$ 无关,而是需要求解一对输入数据 $x_i, x_j$ 的内积 $\langle x_i x_j \rangle = x_i^T x_j$ .先对数据进行映射,然后用SVM去解决分类问题时,会获得以下表达式:

$$\begin{aligned} \max_a & \left[ \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \phi(x_i)^T \phi(x_j) \right] \\ s.t. & \alpha_i \geq 0, \sum_i \alpha_i y_i = 0 \end{aligned} \quad (3)$$

可见,最终需要求解的东西既不是映射函数 $\phi$ 本身,又不是原始数据的内积,而是映射后数据的内积 $\langle \phi(x_i), \phi(x_j) \rangle$ .能否跳过对映射函数 $\langle \phi(x_i), \phi(x_j) \rangle$ 的求解,直接对式(3)进行求解?如

前文所述,想找到合适的映射 $\phi$ 已经十分困难,再对内积进行求解会使整个问题变得更复杂.

这种思路就是寻找核函数(Kernel Function):

$$k(x_i, x_j) = \phi(x_i)^T \phi(x_j) = \langle \phi(x_i), \phi(x_j) \rangle$$

则可以得到判别函数:

$$f(x) = \omega^T \phi(x) + b = \sum_i \alpha_i k(x_i, x_j) + b$$

映射函数本身其实不必显式地写出来,更重要的是,有些情况下映射函数是写不出来的,适合的核函数的存在使得在不知道 $\phi(x_i)$ 和 $\phi(x_j)$ 的具体形式下也可求得. 几种常见的核方法:

$$\text{线性核: } k(x_i, x_j) = x_i^T x_j$$

$$\text{多项式核: } k(x_i, x_j) = (x_i^T x_j)^n, n \geq 1$$

$$\text{高斯核: } k(x_i, x_j) = \exp\left(-\frac{\|x_i, x_j\|^2}{2\sigma^2}\right), \sigma > 0$$

多核学习方法是在训练过程中训练多个基本核函数,对这些不同的核函数及其参数进行组合以获得最优核组合来结合多种核函数的优点,进行更优的特征映射. 按相同的构造方式,有几种类型:若 $k_1, k_2$ 为核函数,则对于任意的正数 $\gamma_1, \gamma_2$ ,其线性组合 $\gamma_1 k_1 + \gamma_2 k_2$ 也是核函数;若 $k_1, k_2$ 为核函数,则核函数的直积 $k_1 \otimes k_2 = k_1(x, z) k_2(x, z)$ 也是核函数;若 $k_1$ 为核函数,则对于任意函数 $g(x), k(x, z) = g(x) k_1(x, z) g(z)$ 也是核函数.

**1.2 多视角学习** 多视角学习在一定程度上和特定的单视角学习相比具有一定的优势. 多视角学习算法通常是在共识或互补原理的指导下建立的,现有的多视角算法可以分三类:协同训练类型算法、协同正则化类型算法和余量一致性类型算法<sup>[2]</sup>. 协同训练类型算法旨在最大化不同视角之间的一致性,如鲁棒协同训练支持向量机(Robust Co-Training SVM)<sup>[11]</sup>. 相比之下,协同正则化类型的算法可以最大程度地减少不同视角之间的分歧<sup>[11]</sup>,例如多视角拉普拉斯支持向量机<sup>[6]</sup>. 余量一致性类型的算法利用多视角分类结果的潜在一致性<sup>[13]</sup>,在最大熵判别(Maximue Entropy Discrimination, MED)<sup>[14]</sup>的基础上实现. 余量一致性类型算法与协同正则化类型算法对多视角的判别或回归函数进行限制不同,它对多视角的余量变

量进行建模,使其尽可能接近即每个输出变量和判别函数都大于每个余量变量,例如MVMED(Multi-View Maximum Entropy Discrimination),以最小化两个视角边距之间的相对熵实现了边距一致性<sup>[15]</sup>. MED-2C(Consensus and Complementarity Based Maximun Entropy Discrimination)以互补子空间将共识和互补的两个原理整合到多视角MED,其泛化性比MVMED更好<sup>[16]</sup>; Tang et al<sup>[17]</sup>的PSVM-2V(Privileged Svm-Based Two-View Classification Mode)是在SVM-2K的基础上结合特权信息并以QP问题迭代求解的算法.

LUPI使用仅在训练期间可用的数据帮助学习模型在测试阶段实现更好的预测结果. 特权信息作为附加特征用来改进特定的分类器, Vipnik and Vashist提出了最早的LUPI算法支持向量机(Support Vector Machine+, SVM+)<sup>[7]</sup>. Tang et al<sup>[18]</sup>将LUPI和多视角结合提出PSVM(Multiview Privileged Support Vector Machines), Li et al<sup>[19]</sup>提出基于LUPI双支持向量机(Robust Capped  $L_1$ -norm Twin Support Vector Machine with Privileged Information, R-CTSVM+)来解决异常值和噪声问题.

这些不同的方法主要利用共识原理或互补原理,应用多视角数据改善模型的通用性能. 本文在平均情况下相互利用冗余视角的附加信息作为特权信息监督当前视角的分类,该方法同时满足共识原理和互补原理,和其他同时利用共识原理和互补原理的方法相比,更关注在平均状况下相互监督的情况,在此基础上设计的目标函数可以利用解析解对目标函数进行优化,使KMPPRVFL的泛化能力更好,训练速度更快.

## 2 算法介绍

**2.1 核化的随机向量函数链接网络** RVFL由于其通用能力和出色的泛化性能<sup>[20]</sup>成为较流行的单层前馈神经网络之一. 为了减少模型的参数并避免选择隐藏节点的数量和隐藏映射函数,同时契合本文多视角和多核的结合应用场景,使用基于核的RVFL. 下面详细介绍本文使用的K-RVFL,其网络结构如图1所示.

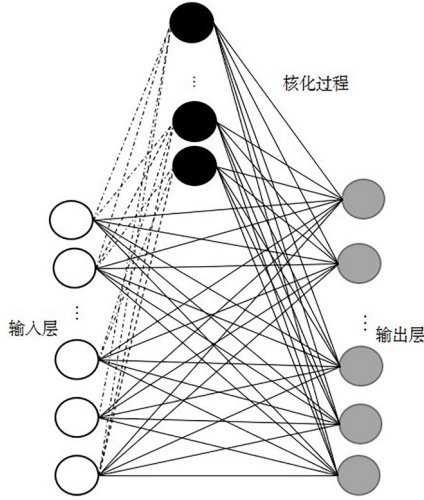


图1 K-RVFL网络的架构

Fig. 1 The architecture of the K-RVFL network

一组具有标记的数据  $\{(x_i, y_i) | x_i \in \mathbb{R}^n, y_i \in \{-1, 1\}, i = 1, \dots, N\}$  并具有  $P$  个增强节点的 RVFL 网络可以表示如下:

$$O = W_{\text{out}}^T \Phi \quad (4)$$

$W_{\text{out}}$  是输出向量权重, 标签  $y_i \in \{-1, 1\}$ ,  $\Phi(x_i)$  表示组合特征向量,  $\Phi = [\Phi_1 \Phi_2]$  是连接输入数据和隐藏层的级联矩阵,  $O$  是网络输出. 为避免产生奇异值, 在计算输出权重时, 在  $\Phi \Phi^T$  或  $\Phi^T \Phi$  对角线上增加一个正值  $\frac{I}{C_R}$ , 可以提高网络的泛化性能.

$$\Phi_1 = \begin{bmatrix} x_{11} & \cdots & x_{1M} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NM} \end{bmatrix} \quad (5)$$

$$\Phi_2 = \begin{bmatrix} K(a_1 \cdot x_1 + b_1) & \cdots & K(a_p \cdot x_1 + b_1) \\ \vdots & \ddots & \vdots \\ K(a_1 \cdot x_N + b_1) & \cdots & K(a_p \cdot x_N + b_p) \end{bmatrix} \quad (6)$$

$$f(x) = \phi(x_i) \Phi^T \left( \frac{I}{C_R} + \Phi \Phi^T \right)^{-1} Y = \quad (7)$$

$$\phi(x_i) \Phi \left( \frac{I}{C_R} + \Phi_1 \Phi_1^T + \Phi_2 \Phi_2^T \right)^{-1} Y$$

定义其核矩阵:

$$K_1 = \Phi_1 \Phi_1^T: K_{1i,j} = K_1(x_i, x_j)$$

$$K_2 = \Phi_2 \Phi_2^T: K_{2i,j} = K_2(x_i, x_j)$$

$K_1$  是线性核,  $K_2$  是如高斯核这样的非线性核, 将其代入式(7)得到:

$$f_{ke}(x) = \begin{bmatrix} K_1(x, x_1) + K_2(x, x_1) \\ \vdots \\ K_1(x, x_n) + K_2(x, x_n) \end{bmatrix}^{-1} \times \left( \frac{I}{C_R} + K_1 + K_2 \right)^{-1} Y \quad (8)$$

从式(8)可以直接通过 Moore-Penrose 伪逆或者岭回归直接计算权重  $W_{\text{out}}$ ,  $I$  是一个单位矩阵,  $C_R$  权衡参数.

**2.2 多视角特权协同核化随机矢量功能链接网络**  
在多视角学习中普遍接受的假设是每个特征视角都可以单独提供一个信息分类器, 从不同特征视角构建的分类器在预测时往往是一致的. KM-PRVFL 的目标是训练一个决策函数  $f(x)$  满足多视角分类的一致性和互补性, 如图 2 所示.

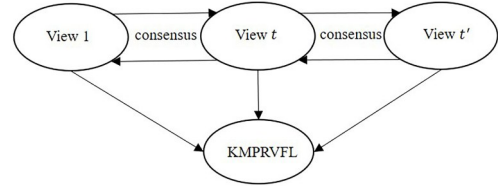


图2 KM-PRVFL 结构示意图

Fig. 2 Illustration of the construction of KM-PRVFL

Bartlett<sup>[21]</sup>证明在具有小经验误差的前馈神经网络中, 权重的范数越小, 泛化性能好. 首先, KM-PRVFL 要最小化各个视角权重  $\|w\|_2^2$ , 然后要最小化各个视角的误差和, 构造非负松弛函数  $J_1$ . 各个视角的  $J_1$  如式(9)所示. 为满足视角之间的一致性和互补性, 视角  $t$  和视角  $t'$  相互提供特权信息相互监督, 结合 LUPI 概念, 在平均情况下各个视角之间约束表示为  $J_2$ , 最后得到目标函数  $J$ :

$$J_1 = \sum_{i=1}^M \sum_{j=1}^N (y_i - \Phi(x_i^j) w_i)^2 \quad (9)$$

$$J_2 = \sum_{1 \leq t \neq t' \leq M} \left( E_{t'} \left( \sum_{i=1}^N \xi_{ti}^2 \geq y_i \Phi(x_i^t) w_A \right) + E_{t'} \left( \sum_{i=1}^N \xi_{ti}^2 \geq y_i \Phi(x_i^{t'}) \right) \right) \quad (10)$$

$$J = \sum_{i=1}^M \|w_i\|_2^2 + J_1 + J_2 \quad (11)$$

因为式(10)不连续并且优化起来较复杂, 考虑使用一个替代(Surrogate)函数<sup>[22]</sup>来优化这个问题,  $\xi_{ti}^2$  和  $\xi_{ti}^2$  非负根据原函数性质选择相应形式的



替代函数如下:

若  $p(x) \geq q(x) \geq 0$ , 则:

$$1(\bullet) = \begin{cases} 1, \bullet \geq 0 \\ 0, \text{otherwise} \end{cases} \quad (12) \quad 1(p(x) \geq q(x)) = \begin{cases} 1, p(x) \geq q(x) \\ 0, \text{otherwise} \end{cases} \quad (13)$$

可以用  $(1 - (p(x) - q(x)))^2$  替代式(10), 有:

$$J_2 = \sum_{i=1}^N (1 - (y_i - \Phi(x_i') \mathbf{w}_i - y_i \Phi(x_i') \mathbf{w}_{i'}))^2 + \sum_{i=1}^N (1 - (y_i - \Phi(x_i') \mathbf{w}_{i'} - y_i \Phi(x_i') \mathbf{w}_i))^2 \quad (14)$$

由此可以得到目标函数:

$$\min_{\mathbf{w}_i} \frac{C_A}{2MN} \sum_{1 \leq i \neq i' \leq M} \left( \sum_{i=1}^N (1 - (y_i - \Phi(x_i') \mathbf{w}_i - y_i \Phi(x_i') \mathbf{w}_{i'}))^2 + \eta \sum_{i=1}^N (1 - (y_i - \Phi(x_i') \mathbf{w}_{i'} - y_i \Phi(x_i') \mathbf{w}_i))^2 \right) + \frac{1}{2M} \sum_{i=1}^M \|\mathbf{w}_i\|_2^2 + \frac{C}{2MN} \sum_{i=1}^M \sum_{i'=1}^N (y_i - \Phi(x_i') \mathbf{w}_i)^2 \quad (15)$$

为了进一步说明 KMPRVFL 的机理, 给出如下详注:

(1)  $\sum_{i=1}^M \|\mathbf{w}_i\|_2^2$  分别是各视角对应的输出权重.

(2)  $\sum_{i=1}^M \sum_{i'=1}^N (y_i - \Phi(x_i') \mathbf{w}_i)^2$  分别为各个视角对应的非负松弛变量,  $C$  为非负折衷参数, 它们分别确保

每个视角的分类能力.

(3) 为方便讨论, 以视角  $i$  和视角  $i'$  为例, 在其中引入非负松弛变量 (16a) 和 (16b), 在平均情况下让两个视角之间相互监督, 从而保证它们之间的一致性和互补性.  $C_A$  是一个非负惩罚参数.

$$\sum_{i=1}^N (1 - (y_i - \Phi(x_i') \mathbf{w}_i - y_i \Phi(x_i') \mathbf{w}_{i'}))^2 \quad (16a)$$

$$\eta \sum_{i=1}^N (1 - (y_i - \Phi(x_i') \mathbf{w}_{i'} - y_i \Phi(x_i') \mathbf{w}_i))^2 \quad (16b)$$

为了更好地解释非负松弛变量, 保证它们之间的一致性和互补性的作用, 将 (16a) 和 (16b) 展开得到 (17a) 和 (17b):

$$\sum_{i=1}^N (1 - (y_i - \Phi(x_i') \mathbf{w}_i - y_i \Phi(x_i') \mathbf{w}_{i'}))^2 - \sum_{i=1}^N 2(y_i - \Phi(x_i') \mathbf{w}_i - y_i \Phi(x_i') \mathbf{w}_{i'}) \quad (17a)$$

$$\eta \sum_{i=1}^N (1 - (y_i - \Phi(x_i') \mathbf{w}_{i'} - y_i \Phi(x_i') \mathbf{w}_i))^2 - \eta \sum_{i=1}^N 2(y_i - \Phi(x_i') \mathbf{w}_{i'} - y_i \Phi(x_i') \mathbf{w}_i) \quad (17b)$$

对 (17a) 和 (17b) 中第二项进行变换得到 (18a) 和 (18b), 然后进一步展开得到 (19a) 和 (19b):

$$\sum_{i=1}^N (y_i - (1 + y_i) \Phi(x_i') \mathbf{w}_i + y_i (\Phi(x_i') \mathbf{w}_i - \Phi(x_i') \mathbf{w}_{i'}))^2 \quad (18a)$$

$$\sum_{i=1}^N (y_i - (1 + y_i) \Phi(x_i') \mathbf{w}_{i'} + y_i (\Phi(x_i') \mathbf{w}_{i'} - \Phi(x_i') \mathbf{w}_i))^2 \quad (18b)$$

$$\frac{C_A}{2NM} \sum_{i=1}^N \left( (y_i - (1 + y_i) \Phi(x_i') \mathbf{w}_i)^2 + (y_i (\Phi(x_i') \mathbf{w}_i - \Phi(x_i') \mathbf{w}_{i'}))^2 \right) + \quad (19a)$$

$$2(y_i^2 - y_i(1 + y_i) \Phi(x_i') \mathbf{w}_i) \times (\Phi(x_i') \mathbf{w}_i - \Phi(x_i') \mathbf{w}_{i'}) \\ \frac{\eta C_A}{2NM} \sum_{i=1}^N \left( (y_i - (1 + y_i) \Phi(x_i') \mathbf{w}_{i'})^2 + (y_i (\Phi(x_i') \mathbf{w}_{i'} - \Phi(x_i') \mathbf{w}_i))^2 \right) + \quad (19b) \\ 2(y_i^2 - y_i(1 + y_i) \Phi(x_i') \mathbf{w}_{i'}) \times (\Phi(x_i') \mathbf{w}_{i'} - \Phi(x_i') \mathbf{w}_i)$$

一致性约束项: (19a) 和 (19b) 第三项相加, 为方便展示, 假设系数  $\eta$  和  $C_A$  为 1. 可以得到  $(\Phi(x_i') \mathbf{w}_i - \Phi(x_i') \mathbf{w}_{i'})^2$  平均一致性约束项, 此项通过标签的相关性约束预测变量之间的一致性, 满足共识原理. 在 (19a) 和 (19b) 中第四项相加此处的隐式约束包含同符号的约束.

在(17a)和(17b)中,第三项在视角 $t'$ 误差项 $y_i - \Phi(x_i^{t'})w_{t'}$ 中含有视角 $t$ 的附加信息 $y_i \Phi(x_i^t)w_t$ ,反之亦然.表示视角互相提供附加信息补充丰富两者,在标签的相关性中引入视角间的矫正功能,实现互补原理,允许在一定参数范围内两视角的存在分歧提高模型的泛化能力.

在(19a)和(19b)中第一项是一个凸函数(convex) $\Phi(x_i^t)w_t, \Phi(x_i^{t'})w_{t'}$ 和 $y_i$ 正相关,此隐式项提升了整体目标的泛化能力.

对式(15)中求解 $w_t$ 后,在联合视角上建立分类器,根据具体情况对新样本的标签进行预测.

在多个视角的应用场景下,如若单独将每个视角都分别相互对应计算式(14),计算量会比较

依据式(15),

$$\min_{w_t} \frac{C_A}{M} \sum_{1 \leq t \neq t' \leq M} \left( \|1 - (Y - k_t w_t - y_i k_{t'} w_{t'})\|_2^2 + \eta \|1 - (Y - k_{t'} w_{t'} - y_i k_t w_t)\|_2^2 \right) + \frac{1}{2M} \sum_{t=1}^M \|w_t\|_2^2 + \frac{C}{M} \sum_{t=1}^M \|Y - k_t w_t\|_2^2 \quad (20)$$

构造拉格朗日函数 $L(w_t, w_{t'})$ ,对 $w_t, w_{t'}$ 求导.

$$\frac{\partial L}{\partial w_t} = w_t + (k_t^T k_t w_t - k_t^T Y) + \eta C_A (k_t^T Y^T - k_t^T Y^T Y + k_t^T Y^T k_{t'} w_{t'} + k_t^T Y^T Y k_t w_t) + C_A (k_t^T - k_t^T Y + k_t^T k_t w_t + k_t^T Y k_t w_t) = 0 \quad (21)$$

$$\frac{\partial L}{\partial w_{t'}} = C_A (k_{t'}^T Y^T - k_{t'}^T Y^T Y + k_{t'}^T Y^T k_t w_t + k_{t'}^T Y^T Y k_{t'} w_{t'}) + \eta C_A (k_{t'}^T - k_{t'}^T Y + k_{t'}^T k_{t'} w_{t'} + k_{t'}^T Y k_{t'} w_{t'}) = 0 \quad (22)$$

构造以 $\begin{pmatrix} w_t \\ w_{t'} \end{pmatrix}$ 为目标的函数,将式(16)(17)相加后写成如下形式:

$$\left[ \begin{pmatrix} I_{n \times n} & 0_{n \times m} \\ 0_{m \times n} & 0_{m \times m} \end{pmatrix} + \begin{pmatrix} P_t & 0_{n \times m} \\ 0_{m \times n} & P_{t'} \end{pmatrix} \right] \begin{pmatrix} w_t \\ w_{t'} \end{pmatrix} = \begin{pmatrix} k_t^T & 0_{n \times N} \\ 0_{m \times N} & k_{t'}^T \end{pmatrix} \begin{pmatrix} Q_t \\ Q_{t'} \end{pmatrix} \quad (23)$$

$$\begin{pmatrix} w_t \\ w_{t'} \end{pmatrix} = \left[ \begin{pmatrix} I_{n \times n} & 0_{n \times m} \\ 0_{m \times n} & \eta I_{m \times m} \end{pmatrix} + \begin{pmatrix} P_t & 0_{n \times m} \\ 0_{m \times n} & P_{t'} \end{pmatrix} \right]^{-1} \times \begin{pmatrix} k_t^T & 0_{n \times N} \\ 0_{m \times N} & k_{t'}^T \end{pmatrix} \begin{pmatrix} Q_t \\ Q_{t'} \end{pmatrix} \quad (24)$$

$$P_t = C_A (\eta k_t^T Y^T Y k_t + k_t^T Y^T k_t) + (1 + C_A) k_t^T k_t + C_A k_t^T Y^T k_t \quad (25)$$

$$P_{t'} = C_A (k_{t'}^T Y^T Y k_{t'} + \eta k_{t'}^T Y^T k_{t'} + k_{t'}^T Y k_{t'}) + \eta C_A k_{t'}^T k_{t'} \quad (26)$$

$$Q_t = Y - C_A (I - Y) + \eta C_A (Y^T Y - Y^T) \quad (27)$$

$$Q_{t'} = \eta C_A (Y - I) - C_A (Y^T - Y^T Y) \quad (28)$$

在二分类中通过以下方法确定测试样品的预测标签:

$$\hat{y} = \frac{1}{M} \sum_{i=1}^M (f_t(x_{\text{test}}) + f_{t'}(x_{\text{test}})) \quad (29)$$

基于上述优化结果的实现如算法1所示.

---

#### Algorithm1 KMPRVFL

---

**INPUT:** 训练数据 $\{(x_i^t, y_i)\}_{i=1}^l, x_i^t \in \mathbb{R}^n, y_i \in \{-1, 1\}, t = 1, \dots, M$

大,应考虑在实际应用时候将除当前视角 $t$ 的其余视角通过多核方式进行组合. $k_t = (x_i^t, x_j^t) =$

$\exp\left(-\frac{1}{2\sigma^2} \|x_i^t - x_j^t\|^2\right) + x_i^{tT} x_j^t, k_{t'}$ 为其余视角核

化的线性组合,即 $k_{t'} = \frac{1}{M} \sum_{1 \leq t \neq u \leq M} k_u$ .

**2.3 目标函数优化** 可以通过固定 $w_t$ 对 $w_{t'}$ 求逆、再固定 $w_{t'}$ 对 $w_t$ 求逆相互迭代,或者使用梯度下降法来一步步迭代求解出合适的 $w_t$ .这样求解一般情况下,其解不保证是全局最优解并且靠近极小值时收敛速度减慢.将 $w_t$ 构造在一起,将式(25)(26)(27)(28)代入式(23)直接得到 $w_t$ ,就得到了目标函数的解析解,比梯度法速度更快.

指定适合的参数 $\eta, C_A, C_1, C$

**OUTPUT:** 输出权重 $w_1, \dots, w_t, w_{t'}, \dots, w_{t'}$

---

Step 1. 将数据核化后得到 $k_1, \dots, k_t$ .

Step 2. 代入式(25)(26)(27)(28),计算相应的

$P_t, P_{t'}, Q_t, Q_{t'}$ .

Step 3. 将  $P_t, P_{t'}, Q_t, Q_{t'}$  代入式 (24), 得到包含  $w_1, \dots, w_t, w_{t'}, \dots, w_{t'}$  的矩阵  $A$ .

Step 4. 从  $A$  按相对应的维度输出权重向量  $w_1, \dots, w_t, w_{t'}, \dots, w_{t'}$ .

讨论 KMPRVFL 的时间复杂度, 对时间复杂度大于  $O(N^2 d)$  的进行分析. Step 1 是多个核的组合, 算法的时间复杂度和输入样本量  $N$  有关, 为  $O(N^2)$ ; Step 2,  $N$  阶矩阵逆的时间复杂度为  $O(N^3)$ , 其中,  $N$  是样本数量,  $d$  是矩阵特征维数. 则 KMPRVFL 的复杂度为  $O(N^2 d + N^3 + N^3)$ . 一般情况下  $d \ll N$  且  $n \ll N$ , 所以综合来看, 算法的时间复杂度为  $O(N^3)$ .

**2.4 多视角特权协同核化随机矢量功能链接网络的增量学习方法** 在实际应用中, 训练数据集不是一次性获取的, 当一些新的训练数据加入网络时就需要动态更新网络的输出权重. 传统的神经网络结构通常是重新训练整个训练集, 这样会对已经计算过的数据进行重新训练, 这个过程将浪费大量的时间. 而 KMPRVFL 网络只需更新新增样本及其对应的映射特征节点和隐藏节点的权重就可以适应新的训练集, 不仅减少了时间的消耗, 还能更好地满足实际应用. 详细过程如下: 构建一个可以更新且足够典型的训练样本簇, 增量学习时对这个典型的样本簇进行学习. 首先定义样本间的距离, 有欧氏距离、曼哈顿距离、切比雪夫距离等. 在此, 以欧氏距离为例:

$$D_u(k^a, k^b) = \sqrt{\sum_{i=1}^n (k_i^a - k_i^b)^2} \quad (30)$$

$D_u$  为两个样本簇之间的距离. 首先将新增的样本簇进行测试, 保留分类正确的样本, 计算得到  $D_t^w$  并降序排序, 同时计算得到的  $D_u^{\text{old}}$  也进行降序排序, 按照一定比例对  $\Phi^{\text{old}}$  进行更新得到  $\Phi^{\text{new}}$ , 将其代入计算得到  $w^{\text{new}}$ , 如算法 2 所示.

---

**Algorithm 2 KMPRVFL2**

---

**INPUT:** 增量训练数据  $\{(x_i^t, y_i)\}_{i=1}^t = \{((x_i^t, 1), y_i)\}_{i=1}^t$ ,  $x_i^t \in \mathbb{R}^n, y_i \in \{-1, 1\}, t = 1, \dots, M$

之前训练好的输出权重  $w_1, \dots, w_t, w_{t'}, \dots, w_{t'}$  和对应的分类器

**OUTPUT:** 输出增量权重  $w_1^{\text{new}}, \dots, w_t^{\text{new}}, w_{t'}^{\text{new}}, \dots, w_{t'}^{\text{new}}$

---

Step 1. 将增量训练数据依据式 (29) 进行测试得到获胜簇.

Step 2. 计算原训练簇  $\Phi^{\text{old}}$  的距离并以距离降序排列.

Step 3. 将获胜簇的数据每次选择  $i$  个替换进  $\Phi^{\text{old}}$  并计算  $D_u$ , 得到  $D_{u\text{Max}}$  时候的  $\Phi^{\text{new}}$ .

Step 4. 将  $\Phi^{\text{new}}$  进行核化得到  $k_1^{\text{new}}, \dots, k_t^{\text{new}}$ .

Step 5. 代入式 (25) (26) (27) (28), 计算相应的  $P_t, P_{t'}, Q_t, Q_{t'}$ .

Step 6. 将  $P_t, P_{t'}, Q_t, Q_{t'}$  代入式 (24), 得到包含  $w_1, \dots, w_t, w_{t'}, \dots, w_{t'}$  的矩阵  $A^{\text{new}}$ .

Step 7. 从  $A^{\text{new}}$  按相对应的维度输出权重向量  $w_1^{\text{new}}, \dots, w_t^{\text{new}}, w_{t'}^{\text{new}}, \dots, w_{t'}^{\text{new}}$ .

---

### 3 实验评估

**3.1 数据集** 为了检验 KMPRVFL 的性能, 对一些真实数据集进行评估. 为了保证实验的真实准确, 每种方法都进行五折交叉验证, 计算平均结果和标准差作为最终结果, 采用常用的准确率 (Accuracy) 作为衡量指标. 所有模拟实验都在同一个环境下完成, Windows10 2004 系统, 在 Inter i7-9750 2.60 GHz 六核 CPU, 32 GB RAM 的计算机上搭建 Matlab 2016b 环境中进行.

NUS-wide<sup>[23]</sup> 是由新加坡国立大学的媒体实验室创建的网络图像数据集, 包括 269648 张图像, 共有 5018 个唯一标签, 从这些图像中提取六种类型的低级特征, 包括 64-D 颜色直方图、144-D 颜色相关图、73-D 边缘方向直方图、128-D 小波纹理、225-D 逐块颜色矩和基于 SIFT 描述的 500-D 特征, 可用于评估的 81 类分类场景. 实验使用其中的单目标图集, 选择 225-D 逐块颜色矩、Normalized\_CM55 和 73-D 边缘方向直方图 Normalized\_EDH.

AwA2<sup>[24]</sup>: 包含 50 种动物的 30475 张图像, 图像数据是 2016 年从公共资源收集的. 每张图像具有六个预先提取的特征表示. 在二分类实验中使用 (Speeded Up Robust Features, SURF) 特征 2000-D、(Histogram of Oriented Gradient, HOG) 特征 252-D、(Color Histogram features CHF) 特征 2600-D、(Local Self-Similarity features LSS) 特征 2000-D.

数据集的信息如表 1 所示. 为了进行更基本

的比较,把数据集拆分成多个二分类,方便与对比算法进行比较. NUS-wide 数据集类别排序按照首字母顺序. 第二个实验直接以首字母排序后从

第 9 个到第 51 个,每间隔六组成一组二分类数据集. AwA2 按作者给出的说明选择一部分作为数据集.

表 1 实验中用到的数据集

Table1 Characteristic of datasets used in experiments

数据集	数量	类别数	特征 1	特征 2	特征 3	特征 4
NUS-wide	4089	7	CM55 (225)	WT (128)	EDH (73)	COPR (144)
AwA2	5560	6	SURF (2000)	HOG(252)	CHF (2600)	ISS (2000)

**3.2 实验设置** 实验的对比算法:KRVFL,为带权重参数  $C$  的用岭回归求解的 RVFL,选择四个视角分别运行一次,取四个视角中最好的结果记录,相当于一组消融实验;MED-2C 方法,将共识性和互补性原则集成到 MED 框架中以进行多视角分类;PSVM-2V,在 SVM-2K 的基础上结合特权信息,并以 QP 问题迭代求解.

为了获得所有方法的最佳参数,实施五重交叉验证,运行 10 次求平均值. MED-2C 和 PSVM-2V 的高斯 RBF 核函数的核参数  $\gamma$  选自  $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10^1, 10^2, 10^3, 10^4, 10^5\}$ , 依据原作者的设置方法在实验中将算法中两个视角

的内核参数设置为相同值. 在集合  $\{10^{-5}, \dots, 10^5\}$  上调整 PSVM-2V, MED-2C, RVFL 中的参数  $C$  等. 在 KMPRVFL 中,设置  $C_1 = C_2 = C$ ,并在相同的核化参数范围内选取,即  $\{10^{-5}, \dots, 10^5\}$ .

**3.3 实验结果和分析** 比较 KMPRVFL 和所有对比测试方法的性能. 讨论 KMPRVFL 的参数敏感性,在隐节点参数固定后,精度随  $\eta, C_A, C$  变化的选择参数的部分结果,如图 3 所示. 由图可见,在 buildings vs elk 组实验中,  $\eta=0.1$  时在  $C=10^{-5}$  处精度达到极大值,  $C$  或  $C_A$  都较小时模型的鲁棒性最佳. 另外,在达到有效的  $C$  后,  $\eta$  的变化幅度不大,  $\eta=0.01$  时达到极大值.

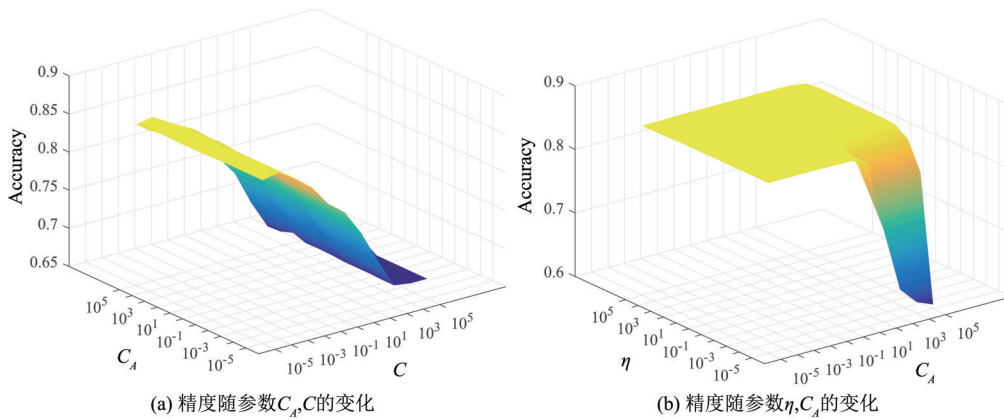


图 3 不同参数下 KMPRVFL 在 NUS-wide 数据集上的性能变化

Fig.3 Performance of KMPRVFL with different parameters in dataset NUS-wide

表 2 和表 3 分别列出了来自 NUS-wide 和 AwA 的 28 个数据集的二分类结果. 本文方法在 NUS-wide 数据集上比单个视角 KRVFL 平均高出约 6%, 比 MED-2C 平均高出约 5%, 比 PSVM-2V 平均高出约 4%; 在 fox vs plants 组数据集略低于 PSVM-2V, 但是两者非常接近.

综上所述,在绝大多数情况下, KMPRVFL 以最高的平均准确度获得了最佳性能, 优于其他算法. 对于绝大多数数据集, KMPRVFL 的准确性比 PSVM-2V 更高, 也比自身单独视角的性能更好, 这进一步证明 KMPRVFL 本身可以按照互补性原理, 充分利用两个视角作为特权信息, 并遵



表2 KMPRVFL和对比算法在NUS-wide数据集上的性能

Table 2 Binary classification performance of KMPRVFL and other algorithms on NUS-wide dataset

	Datasets-A	Datasets-B	KMPRVFL		KRVFL		MED-2C		PSVM-2V	
			Accuracy	STD	Accuracy	STD	Accuracy	STD	Accuracy	STD
1	buildings	computer	<b>83.86%</b>	0.021	71.93%	0.015	77.17%	0.015	78.36%	0.012
2	buildings	elk	<b>86.57%</b>	0.022	81.01%	0.017	81.04%	0.017	82.23%	0.015
3	buildings	fox	<b>90.64%</b>	0.013	82.28%	0.035	83.41%	0.035	84.30%	0.029
4	buildings	horses	<b>85.35%</b>	0.030	76.42%	0.011	78.87%	0.011	79.82%	0.015
5	buildings	moon	<b>83.59%</b>	0.023	75.16%	0.009	81.65%	0.009	78.52%	0.020
6	buildings	plants	<b>85.98%</b>	0.015	79.19%	0.013	81.34%	0.013	82.25%	0.007
7	buildings	road	<b>72.68%</b>	0.013	63.93%	0.026	62.08%	0.026	65.22%	0.027
8	computer	elk	<b>86.12%</b>	0.033	77.21%	0.025	77.47%	0.025	78.66%	0.025
9	computer	fox	<b>85.41%</b>	0.031	76.39%	0.030	75.18%	0.030	76.92%	0.024
10	computer	horses	<b>89.04%</b>	0.022	78.11%	0.012	81.52%	0.012	82.90%	0.016
11	computer	moon	<b>83.10%</b>	0.028	76.45%	0.026	74.64%	0.026	77.99%	0.029
12	computer	plants	<b>87.00%</b>	0.012	80.68%	0.014	79.90%	0.014	81.34%	0.013
13	computer	road	<b>81.36%</b>	0.021	74.51%	0.034	70.62%	0.034	76.17%	0.033
14	elk	fox	<b>75.45%</b>	0.030	68.95%	0.023	68.57%	0.023	70.65%	0.026
15	elk	horses	<b>78.88%</b>	0.021	76.67%	0.023	75.73%	0.023	76.74%	0.024
16	elk	moon	<b>87.77%</b>	0.029	82.68%	0.013	83.98%	0.013	84.53%	0.025
17	elk	plants	<b>84.77%</b>	0.033	82.21%	0.009	83.75%	0.009	83.09%	0.012
18	elk	road	<b>82.60%</b>	0.022	78.90%	0.032	79.24%	0.032	79.28%	0.032
19	fox	horses	<b>83.80%</b>	0.021	76.65%	0.018	77.68%	0.018	79.99%	0.019
20	fox	moon	<b>86.22%</b>	0.023	81.94%	0.044	82.25%	0.044	83.92%	0.045
21	fox	plants	79.83%	0.057	79.73%	0.014	73.88%	0.014	<b>82.31%</b>	0.004
22	fox	road	<b>84.16%</b>	0.033	78.81%	0.033	81.67%	0.033	79.65%	0.031
23	horses	moon	<b>88.18%</b>	0.033	82.56%	0.028	83.76%	0.028	84.59%	0.032
24	horses	plants	<b>89.15%</b>	0.011	85.05%	0.012	85.12%	0.012	86.73%	0.015
25	horses	road	<b>80.14%</b>	0.012	73.94%	0.035	74.57%	0.035	75.67%	0.039
26	moon	plants	<b>86.90%</b>	0.028	81.11%	0.033	82.40%	0.033	83.03%	0.027
27	moon	road	<b>81.34%</b>	0.027	74.71%	0.015	75.23%	0.015	76.67%	0.029
28	plants	road	<b>82.13%</b>	0.017	76.43%	0.031	77.68%	0.030	78.97%	0.030
Average			<b>84.00%</b>	0.023	77.63%	0.018	78.23%	0.221	79.66%	0.024

循共识性原理添加正则化项,以实现更好的分类性能.

选取 NUS-wide 数据集的 elk vs buildings, elk vs horses, horses vs road 三组做增量实验,首先选取 50% 的数据作为原始样本簇,然后每次增加 10% 作为增量样本,实验的结果如图 4 所示. 由图可见,增量学习方式可以有效地提高 KMPRVFL 的精度.

## 4 结语

本文提出一种核化的多视角特权协同随机矢量功能链接网络(KMPRVFL),可以有效地解决多视角分类任务. 在平均情况下,KMPRVFL 利用冗余视角的附加信息作为特权信息监督当前视角的分类. 该方法的目标函数可以求出解析解,然后使用伪逆方法快速求解. 在真实的多视角数据集上进行实验,和 MED-2C, PSVM-2V 相比, KMPRVFL 可以实现更好的泛化性能,实验结果

表 3 KMPRVFL 和对比算法在 Awa2 数据集上的性能

Table 3 Classification performance of KMPRVFL and other algorithms on Awa2 dataset

	Datasets-A	Datasets-B	KMPRVFL		KRVFL		MED-2C		PSVM-2V	
			Accuracy	STD	Accuracy	STD	Accuracy	STD	Accuracy	STD
1	chimps	panda	<b>91.01%</b>	0.009	87.14%	0.027	86.93%	0.027	90.92%	0.023
2	chimps	leopard	<b>92.50%</b>	0.01	86.84%	0.018	82.80%	0.04	87.61%	0.043
3	chimps	cat	<b>92.18%</b>	0.023	86.65%	0.034	82.07%	0.049	86.52%	0.067
4	chimps	pig	<b>89.15%</b>	0.05	82.09%	0.05	82.86%	0.033	83.81%	0.056
5	chimps	hippo	<b>89.05%</b>	0.014	86.84%	0.022	82.82%	0.043	84.45%	0.063
6	chimps	rat	<b>89.78%</b>	0.023	82.36%	0.041	75.12%	0.035	81.16%	0.082
7	chimps	seal	<b>91.90%</b>	0.012	81.46%	0.035	83.07%	0.026	87.76%	0.016
8	panda	leopard	<b>92.83%</b>	0.016	88.84%	0.014	84.31%	0.026	90.46%	0.012
9	panda	cat	<b>95.05%</b>	0.015	90.04%	0.026	88.03%	0.02	89.53%	0.067
10	panda	pig	<b>90.53%</b>	0.014	83.62%	0.025	78.83%	0.042	83.46%	0.035
11	panda	hippo	<b>93.89%</b>	0.013	87.69%	0.018	87.47%	0.009	91.13%	0.05
12	panda	rat	<b>91.75%</b>	0.02	85.92%	0.043	82.25%	0.025	86.57%	0.064
13	panda	seal	<b>92.62%</b>	0.024	89.10%	0.017	86.89%	0.033	89.63%	0.026
14	leopard	cat	<b>91.74%</b>	0.015	86.94%	0.021	86.15%	0.033	90.69%	0.045
15	leopard	pig	<b>85.56%</b>	0.037	82.62%	0.028	78.53%	0.027	84.11%	0.057
16	leopard	hippo	<b>91.13%</b>	0.01	83.10%	0.028	82.54%	0.037	88.49%	0.092
17	leopard	rat	<b>87.58%</b>	0.014	84.53%	0.033	80.17%	0.04	85.21%	0.082
18	leopard	seal	<b>90.78%</b>	0.019	86.47%	0.027	87.83%	0.045	88.00%	0.091
19	cat	pig	<b>84.87%</b>	0.037	79.27%	0.044	73.92%	0.061	76.80%	0.055
20	cat	hippo	<b>89.57%</b>	0.016	87.08%	0.029	85.07%	0.037	86.29%	0.074
21	cat	rat	<b>77.01%</b>	0.029	74.17%	0.04	62.40%	0.025	68.46%	0.043
22	cat	seal	<b>86.52%</b>	0.023	76.55%	0.044	82.60%	0.038	83.68%	0.06
23	pig	hippo	<b>82.20%</b>	0.019	77.21%	0.055	71.42%	0.037	74.49%	0.066
24	pig	rat	<b>73.18%</b>	0.029	69.21%	0.032	70.52%	0.025	<b>74.31%</b>	0.08
25	pig	seal	<b>84.59%</b>	0.021	78.27%	0.045	71.86%	0.037	77.08%	0.092
26	hippo	rat	<b>85.17%</b>	0.02	82.92%	0.019	72.94%	0.035	75.23%	0.056
27	hippo	seal	<b>81.41%</b>	0.026	70.45%	0.034	67.08%	0.03	69.48%	0.053
28	rat	seal	<b>80.59%</b>	0.036	77.32%	0.042	71.96%	0.02	75.69%	0.017
Average			<b>87.41%</b>	0.019	82.58%	0.031	79.78%	0.034	83.59%	0.054

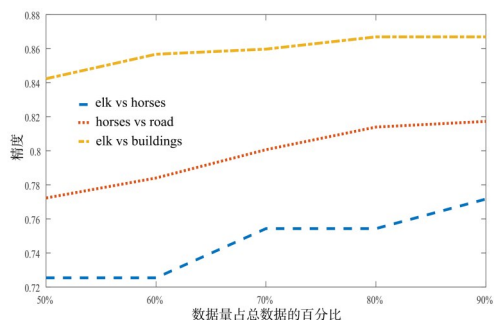


图 4 输入样本增量学习的精度折线图

Fig. 4 Accuracy curves produced by incremental learning of input data

证明了该方法的有效性. 对于增量的应用场景, 还提出一种 KMPRVFL 的增量方法, 同样在实验中也证明它可以有效地在增量任务中提高精度.

#### 参考文献

- [1] Yang Y, Wang H. Multi-view clustering: A survey. Big Data Mining and Analytics, 2018, 1(2): 83—107.
- [2] Chen Y F, Li D D, Zhang X, et al. Computer aided diagnosis of thyroid nodules based on the devised small - datasets multi - view ensemble learning. Medical Image Analysis, 2021, 67: 101819.

- [3] 洪思思,曹辰捷,王喆,等. 基于矩阵的AdaBoost多视角学习. 南京大学学报(自然科学),2018,54(6):1152—1160. (Hong S S, Cao C J, Wang J. Matrix-based multi-view learning with AdaBoost. Journal of Nanjing University (Natural Science), 2018, 54(6): 1152—1160.)
- [4] Li J, Allinson N, Tao D C, et al. Multi-training support vector machine for image retrieval. IEEE Transactions on Image Processing, 2006, 15(11): 3597—3601.
- [5] Sun S L, Shawe-Taylor J. Sparse semi-supervised learning using conjugate functions. Journal of Machine Learning Research, 2010(11): 2423—2455.
- [6] Sun S L. Multi-view Laplacian support vector machines//Proceedings of the 7<sup>th</sup> International Conference on Advanced Data Mining and Applications. Springer Berlin Heidelberg, 2011: 209—222.
- [7] Vapnik V, Vashist A. A new learning paradigm: Learning using privileged information. Neural Networks, 2009, 22(5—6): 544—557.
- [8] Vapnik V, Izmailov R. Learning using privileged information: Similarity control and knowledge transfer. The Journal of Machine Learning Research, 2015, 16(1): 2023—2049.
- [9] Evers L, Messow C M. Sparse kernel methods for high-dimensional survival data. Bioinformatics, 2008, 24(14): 1632—1638.
- [10] Camoriano R. Large-scale kernel methods and applications to lifelong robot learning. 2019, arXiv: 1912.05629.
- [11] Sun S L, Jin F. Robust co-training. International Journal of Pattern Recognition and Artificial Intelligence, 2011, 25(7): 1113—1126.
- [12] Cheng J, Wang K Q. Multi-view sampling for relevance feedback in image retrieval//18<sup>th</sup> International Conference on Pattern Recognition. Hong Kong, China: IEEE, 2006, 2: 881—884.
- [13] 张丹丹,邓赵红,蒋亦樟,等. 基于划分融合与视角加权的极大熵聚类算法. 计算机工程, 2016, 10(4): 554—564. (Zhang D D, Deng Z H, Wang S T. Maximum entropy clustering algorithm for multi-view data. Journal of Frontiers of Computer Science and Technology, 2016, 10(4): 554—564.)
- [14] Valente F, Wellekens C. Maximum entropy discrimination (MED) feature subset selection for speech recognition//2003 IEEE Workshop on Automatic Speech Recognition and Understanding. St Thomas, VI, USA: IEEE, 2003: 327—332.
- [15] Sun S L, Chao G Q. Multi-view maximum entropy discrimination//Proceedings of the 23<sup>rd</sup> International Joint Conference on Artificial Intelligence. Beijing, China: AAAI Press, 2013: 1706—1712.
- [16] Chao G Q, Sun S L. Consensus and complementarity based maximum entropy discrimination for multi-view classification. Information Sciences, 2016(367—368): 296—310.
- [17] Tang J J, Tian Y J, Liu X H, et al. Improved multi-view privileged support vector machine. Neural Networks, 2018(106): 96—109.
- [18] Tang J J, Tian Y J, Zhang P, et al. Multiview privileged support vector machines. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(8): 3463—3477.
- [19] Li Y M, Sun H J, Yan W Z, et al. R-CTSVM+: Robust capped  $L_1$ -norm twin support vector machine with privileged information. Information Sciences, 2021(574): 12—32.
- [20] Pao Y H, Park G H, Sobajic D J. Learning and generalization characteristics of the random vector Functional-link net. Neurocomputing, 1994, 6(2): 163—180.
- [21] Bartlett P L. The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. IEEE Transactions on Information Theory, March 1998, 44(2): 525—536.
- [22] Queipo N V, Haftka R T, Shyy W, et al. Surrogate-based analysis and optimization. Progress in Aerospace Sciences, 2005, 41(1): 1—28.
- [23] Chua T S, Tang J H, Hong R C, et al. Nus-wide: A real-world web image database from national university of Singapore//Proceedings of the 8<sup>th</sup> ACM International Conference on Image and Video Retrieval. Santorini Island, Greece: ACM, 2009: 1—9.
- [24] Xian Y Q, Lampert C H, Schiele B, et al. Zero-shot learning: A comprehensive evaluation of the good, the bad and the ugly. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(9): 2251—2265.

(责任编辑 杨可盛)