

DOI: 10.11992/tis.201703034

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20180328.1649.014.html>

基于超限学习机的非线性典型相关分析及应用

温晓红¹, 刘华平^{2,3}, 阎高伟¹, 孙富春^{2,3}

(1. 太原理工大学电气与动力工程学院, 山西太原 030600; 2. 清华大学计算机科学与技术系, 北京 100084; 3. 智能技术与系统国家重点实验室, 北京 100084)

摘要: 典型相关分析是目前常用的研究两个变量间相关性的统计方法。针对线性典型相关分析难以准确揭示变量之间复杂关系的问题, 提出一种基于超限学习机的非线性典型相关分析多模态特征提取方法。首先, 采用超限学习机分别的对每个模态进行无监督特征学习, 得到抽象的深度特征表示; 然后将这些深度抽象特征通过典型相关分析极大化模态之间的相关性, 同时得到两组相关变量, 实现多模态数据的复杂非线性和高相关性表示。最后在康奈尔大学机器抓取公开数据集上进行实验验证, 结果表明, 所提出的方法与其他相关算法相比, 训练速度得到显著提升。

关键词: 典型相关分析; 超限学习机; 特征提取; 多模态融合; 机器抓取; 目标识别; RGB-D 数据; 神经网络
中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1673-4785(2018)04-0633-07

中文引用格式: 温晓红, 刘华平, 阎高伟, 等. 基于超限学习机的非线性典型相关分析及应用[J]. 智能系统学报, 2018, 13(4): 633-639.

英文引用格式: WEN Xiaohong, LIU Huaping, YAN Gaowei, et al. Nonlinear canonical correlation analysis and application based on extreme learning machine[J]. CAAI transactions on intelligent systems, 2018, 13(4): 633-639.

Nonlinear canonical correlation analysis and application based on extreme learning machine

WEN Xiaohong¹, LIU Huaping^{2,3}, YAN Gaowei¹, SUN Fuchun^{2,3}

(1. College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030600, China; 2. Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China; 3. State Key Laboratory of Intelligent Technology and Systems, Beijing 100084, China)

Abstract: Canonical correlation analysis (CCA) is a statistical technique commonly used to determine the correlativity of two variables. It is difficult to accurately identify the complex underlying relationship between variables using linear CCA, so we propose a nonlinear CCA based on an extreme learning machine (ELM) for multi-modal feature extraction. First, to obtain abstract-depth feature representation, we use the ELM to perform unsupervised feature learning for each modality. Then, we use CCA to maximize the correlation between the nonlinear representations, thereby simultaneously obtaining two groups of related variables, and realize complex nonlinear and high-correlation representations of multi-modality data. Lastly, we conducted an experiment using the Cornell grasping dataset. The results show that, in comparison with other related algorithms, the proposed method significantly increases the training speed.

Keywords: canonical correlation analysis; extreme learning machine; feature extraction; multi-modal fusion; robotic grasping; object recognition; RGB-D data; neural networks

收稿日期: 2017-03-24. 网络出版日期: 2018-03-28.

基金项目: 国家自然科学基金重点项目 (U1613212); 国家高技术研究发展计划项目 (2015AA042306).

通信作者: 刘华平. E-mail: hpliu@tsinghua.edu.cn.

典型相关分析 (canonical correlation analysis, CCA)^[1]是一种经典统计方法, 旨在研究两组随机变量之间的相关性。通过极大化两组变量在线性

投影空间中的相关性而达到特征融合的目的。CCA 可用于对多模态数据的特征提取和融合^[2-4], 减小预测或识别问题中的样本复杂度^[5], 已成功应用于字符与人脸图像识别及多模态检索等任务中^[6]。核典型相关分析 (kernel CCA, KCCA)^[7] 是对 CCA 的扩展, 解决了 CCA 只能进行线性映射的问题。通过隐性的非线性映射将两组特征分别映射到高维特征空间, 借助于核技巧, 以线性的方式提取变量的非线性特征。尽管 KCCA 可以通过核来学习变量之间的非线性关系, 但其需要通过人工选择核函数, 并且当训练集规模较大时, 会造成巨大的计算消耗。

为了解决 KCCA 存在的问题, 文献[8]提出深度典型相关分析 (deep CCA, DCCA), 即用深度神经网络学习灵活的非线性相关表示, 使得到的特征具有更高的相关度。文献[9]提出基于神经网络和自编码 (auto encoder, AE) 的深度典型相关自编码 (deep canonically correlated autoencoders, DCCAE), 用于无标签多视图特征学习。但是这些方法在参数训练过程中, 需要根据梯度下降法多次迭代至误差收敛, 容易陷入局部最优, 且计算消耗大。

另一方面, 超限学习机 (extreme learning machine, ELM) 以结构简单、学习速度快和泛化能力好等优点^[10], 近年来已广泛应用于分类、回归等任务中。伴随着大数据时代的到来, 关于数据潜在信息的挖掘引起了大量关注。然而 ELM 是一种单隐层前馈神经网络, 其浅层的网络结构难以发现数据深层抽象特征, 因此基于 ELM 进行深度学习的算法相继提出。文献[11]提出了 ELM-AE, 首次将 ELM 用于数据的特征表示; 文献[12]提出了基于 ELM 堆栈构成的深度表示模型; 文献[13]将流行正则引入原始 ELM-AE 中, 并提出一种新的深度神经网络; 文献[14]使用深度 ELM 学习图像的非线性结构, 用于图像分类中, 并且在速度和准确率上表现出优越的性能; 文献[15]将深度 ELM 模型用于多模态融合中, 表明了 ELM 在多模态数据应用中的有效性。

为了解决基于传统深度神经网络的非线性 CCA 训练速度缓慢、无法快速收敛等问题, 本文将 ELM-AE 引入 CCA 框架, 用于多模态数据的特征提取。首先利用 ELM-AE 分别对每个模态进行多层无监督特征提取, 然后极大化深层特征的相关性。实验结果表明, 相比于线性 CCA 和 DCCA, 本文方法在取得高相关度的同时, 还显著提升了算法的快速性。将该算法在康奈尔大学机器抓取

数据集上进行验证, 实验结果表明该方法有效提高了抓取点识别率。

1 相关工作

1.1 典型相关分析

给定 N 对已中心化的样本 $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_N]$ 与 $\mathbf{Y} = [\mathbf{y}_1 \ \mathbf{y}_2 \ \cdots \ \mathbf{y}_N]$, CCA 的基本思想是找到一对投影向量 \mathbf{a} 和 \mathbf{b} , 满足 $\mathbf{u} = \mathbf{a}^T \mathbf{X}$ 和 $\mathbf{v} = \mathbf{b}^T \mathbf{Y}$ 之间具有最大的相关系数。然后通过对典型相关变量 \mathbf{u} 和 \mathbf{v} 的相关关系研究, 代替原来两组变量之间的相关关系研究。

\mathbf{u} 和 \mathbf{v} 的相关系数为

$$\rho = \text{corr}(\mathbf{a}^T \mathbf{X}, \mathbf{b}^T \mathbf{Y}) = \frac{\mathbf{a}^T \boldsymbol{\Sigma}_{\mathbf{XY}} \mathbf{b}}{\sqrt{(\mathbf{a}^T \boldsymbol{\Sigma}_{\mathbf{XX}} \mathbf{a})(\mathbf{b}^T \boldsymbol{\Sigma}_{\mathbf{YY}} \mathbf{b})}} \quad (1)$$

式中: $\boldsymbol{\Sigma}_{\mathbf{XX}}$ 和 $\boldsymbol{\Sigma}_{\mathbf{YY}}$ 分别表示特征集 \mathbf{X} 和 \mathbf{Y} 的协方差矩阵, $\boldsymbol{\Sigma}_{\mathbf{XY}}$ 表示 \mathbf{X} 和 \mathbf{Y} 的互协方差矩阵。

典型相关分析就是在约束条件 $\mathbf{a}^T \boldsymbol{\Sigma}_{\mathbf{XX}} \mathbf{a} = 1$, $\mathbf{b}^T \boldsymbol{\Sigma}_{\mathbf{YY}} \mathbf{b} = 1$ 下求 \mathbf{a} 和 \mathbf{b} 的值, 使 ρ 达到最大, 此时目标函数为 $\max \rho = \mathbf{a}^T \boldsymbol{\Sigma}_{\mathbf{XY}} \mathbf{b}$ 。

定义矩阵 $\mathbf{T} = \boldsymbol{\Sigma}_{\mathbf{XX}}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{XY}} \boldsymbol{\Sigma}_{\mathbf{YY}}^{-1/2}$, 对 \mathbf{T} 进行奇异值分解, 得到矩阵 \mathbf{T} 的前 k 个左奇异向量 $\boldsymbol{\alpha}_k$ 和右奇异向量 $\boldsymbol{\eta}_k$, 则最终优化目标值为 $(\hat{\mathbf{a}}, \hat{\mathbf{b}}) = (\boldsymbol{\alpha}_k \boldsymbol{\Sigma}_{\mathbf{XX}}^{-1/2}, \boldsymbol{\eta}_k \boldsymbol{\Sigma}_{\mathbf{YY}}^{-1/2})$ 。

为了防止训练过拟合, 在协方差计算中加入正则项 $\gamma > 0$:

$$\boldsymbol{\Sigma}_{\mathbf{XX}} = \frac{1}{N-1} \mathbf{X} \mathbf{X}^T + \gamma \mathbf{I} \quad (2)$$

1.2 深度典型相关分析

Andrew 等^[8]对 CCA 进行扩展改进, 提出基于深度神经网络的 CCA, 即 DCCA。给定模态 \mathbf{X} 和 \mathbf{Y} , 两个深度神经网络模型 $f = f(\mathbf{X}; \mathbf{W}_f)$ 和 $g = g(\mathbf{Y}; \mathbf{W}_g)$ 分别用来学习 \mathbf{X} 和 \mathbf{Y} 的非线性结构, 其中 $\mathbf{W} = [\mathbf{W}_f; \mathbf{W}_g]$ 为神经网络模型参数。然后通过 CCA 将提取的特征 f 和 g 相关性极大化:

$$\begin{aligned} & \max_{\mathbf{w}_f, \mathbf{w}_g, \mathbf{u}, \mathbf{v}} \frac{1}{N} \text{tr}(\mathbf{U}^T f(\mathbf{X}; \mathbf{W}_f) g(\mathbf{Y}; \mathbf{W}_g)^T \mathbf{V}) \\ & \text{s.t. } \mathbf{U}^T \left(\frac{1}{N} f(\mathbf{X}) f(\mathbf{X})^T + r_x \mathbf{I} \right) \mathbf{U} = \mathbf{I} \\ & \mathbf{V}^T \left(\frac{1}{N} g(\mathbf{Y}) g(\mathbf{Y})^T + r_y \mathbf{I} \right) \mathbf{V} = \mathbf{I} \\ & \mathbf{u}_i^T f(\mathbf{X}) g(\mathbf{Y})^T \mathbf{v}_j = 0, \quad i \neq j \end{aligned} \quad (3)$$

式中: $\mathbf{U} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \ \mathbf{u}_L]$, $\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_L]$ 是 CCA 对 DNN 输出特征的映射矩阵, $r_x > 0, r_y > 0$ 是样本协方差估计的正则化参数。

令 \mathbf{H}_X 和 \mathbf{H}_Y 分别为通过深度模型 f 和 g 得到的最终特征表示, 中心化后得到 $\hat{\mathbf{H}}_X = \mathbf{H}_X - \mathbf{H}_X/N$, 同理得到 $\hat{\mathbf{H}}_Y$, 并且计算两者的协方差矩阵 $\hat{\boldsymbol{\Sigma}}_{\mathbf{XY}}$ 、 $\hat{\boldsymbol{\Sigma}}_{\mathbf{XX}}$ 、 $\hat{\boldsymbol{\Sigma}}_{\mathbf{YY}}$ 。根据 1.1 节中对 CCA 的描述, 对 \mathbf{H}_X 和 \mathbf{H}_Y 进

行相关分析, 得到映射矩阵 U 和 V 。假设 T 的奇异值分解矩阵为 $T = UDV^T$, 则目标函数对 H_X 和 H_Y 的梯度计算为

$$\frac{\partial \text{corr}(H_X, H_Y)}{\partial H_X} = \frac{1}{N-1} (2\nabla_{XX} H_X + \nabla_{XY} H_Y) \quad (4)$$

式中: $\nabla_{XY} = \widehat{\Sigma}_{XX}^{-1/2} U V^T \widehat{\Sigma}_{YY}^{-1/2}$, $\nabla_{XX} = -\frac{1}{2} \widehat{\Sigma}_{XX}^{-1/2} U D U^T \widehat{\Sigma}_{XX}^{-1/2}$

根据梯度下降法对网络参数 $W = [W_f; W_g]$ 进行优化。 $\tilde{f}(X) = U^T f(X)$ 和 $\tilde{g}(Y) = V^T g(Y)$ 即为最终 X 和 Y 的投影矩阵, 用于测试样本的映射。

与 CCA 不同, DCCA 没有固定的解, 其参数需要通过梯度下降法优化, Andrew 等采用批处理算法 L-BFGS 对参数进行调整。王等^[16]随后提出采用随机梯度下降 (stochastic gradient descent, SGD) 对 DCCA 的权值进行优化。本文选择与后者提出的方法进行比较。

1.3 超限学习机

超限学习机是一类针对单隐含层前馈神经网络的学习算法。典型的单隐层前馈神经网络结构如图 1 所示, 由输入层、隐含层和输出层组成。

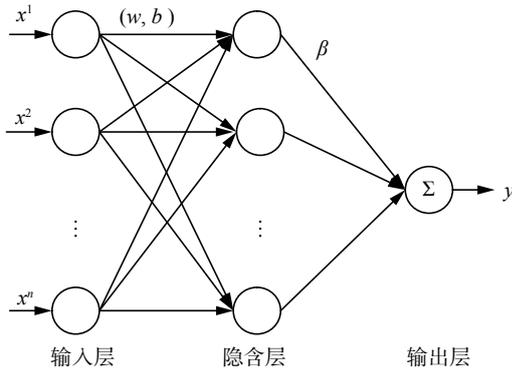


图 1 ELM 网络结构

Fig. 1 The model structure of ELM

给定 N 个训练样本, $X \in \mathbb{R}^{N \times n}$, $t \in \mathbb{R}^{N \times 1}$, n 为输入的维度。第 i 个样本的网络输出可表示为

$$y_i = \sum_{j=1}^h g(x_i w_j^T + b_j) \beta_j \quad (5)$$

式中: y_i 是网络的实际输出, h 是隐含层节点数, $w_j \in \mathbb{R}^n$ 是输入权值, b_j 是隐含层神经元的偏置, β_j 是和是输出层的权值, $g(x)$ 为隐含层的激活函数, 取正切 sigmoid 函数。式 (5) 可表示为矩阵形式:

$$H\beta = T \quad (6)$$

式中: H 为神经网络的隐含层输出矩阵, T 是网络期望输出值。

在训练过程中, 对 w 和 b 随机初始化, 且保持不变, 而隐含层与输出层的连接权值 β 可以通过求解以下方程组获得:

$$\widehat{\beta} = H^+ T \quad (7)$$

式中: H^+ 为隐含层输出矩阵 H 的 Moore-Penrose 广义逆。

2 基于超限学习机的非线性典型相关分析

本文提出基于 ELM 的非线性典型相关分析 (ELM-CCA), 通过对两个模态分别进行复杂的非线性转换, 最终使学习到的特征最大相关, 其模型结构如图 2 所示。给定样本 $\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, N 为样本个数, $x_i \in \mathbb{R}^p, y_i \in \mathbb{R}^q$, 因此可以得到两个模态的样本矩阵, 分别为 $X \in \mathbb{R}^{N \times p}$ 和 $Y \in \mathbb{R}^{N \times q}$, 传统的 CCA 对 X 和 Y 进行线性变换, 使变换后的两组数据相关性最大。实际中很多问题需要非线性变换来挖掘事物间潜在的非线性关系。而 ELM-CCA 算法可以实现非线性相关学习。

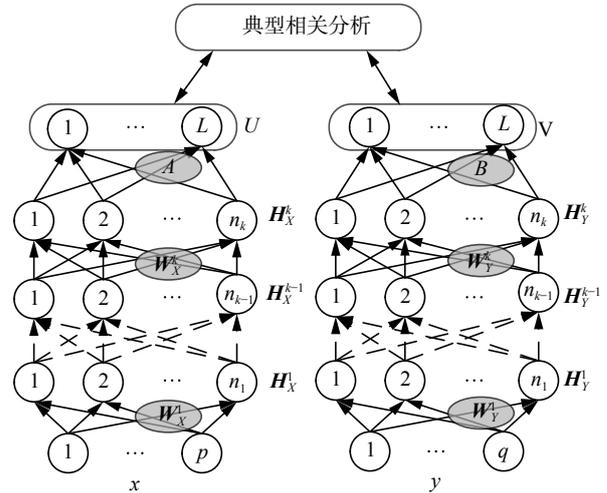


图 2 ELM-CCA 模型结构

Fig. 2 The model structure of ELM-CCA

ELM-CCA 结构分为两部分: 分别对每个模态进行多层非线性转换; 极大化对两个模态多层非线性映射后特征的相关性。对于 $X \in \mathbb{R}^{N \times p}$ 中的其中一个样本 $x \in \mathbb{R}^p$, 设有 k 个隐含层, 则第 1 个隐含层的输出为

$$H_X^1 = f(W_X^1 x + b_X^1) \quad (8)$$

式中: $W_X^1 \in \mathbb{R}^{n_1 \times p}$ 是连接输入与第 1 个隐含层的权值, $b_X^1 \in \mathbb{R}^{n_1}$ 是第 1 层隐含层的偏置向量, f 是隐含层的激活函数。将上一层的输出作为输入, 依次可得到第 k 层的输出:

$$H_X^k = f(W_X^k H_X^{k-1} + b_X^k) \quad (9)$$

将模态 X 最终得到的隐含层输出简记为 H_X , 对于模态 Y , 通过相同的方式, 得到最后一层隐含层的输出 H_Y 。

对 H_X 和 H_Y 进行典型相关分析, 分别得到对应的线性转换矩阵 $A \in \mathbb{R}^{L \times n_k}$ 和 $B \in \mathbb{R}^{L \times n_k}$, 最后的输出

层分别为

$$U = AH_X \quad (10)$$

$$V = BH_Y \quad (11)$$

通过深度 ELM-CCA 的训练, 得到 X 和 Y 的非线性映射表示 U 和 V , 并且 U 和 V 的相关度最大。

2.1 无监督非线性映射

采用 ELM-AE 逐层训练得到每层之间的连接权值, 即每层权值的训练都作为一个独立的 ELM。在独立组件 ELM-AE 中, 令其输出 $y=x$, 通过最小化输出的重构误差来训练权值。

对于第 k 层权值 W_X^k , 构造 ELM-AE 模型, 如图 3 所示。随机给定正交化输入权值 w_X^k 和隐含层偏置 b_X^k , 则其隐含层输出为

$$H_X^k = f(w_X^k H_{X^{k-1}} + b_X^k)(w_X^k)^T w_X^k = I, b_X^k{}^T b_X^k = 1 \quad (12)$$

式中: $w_X^k \in R^{n_{k-1} \times n_k}$ 是在输入层和隐含层节点之间正交的随机权值, $b_X^k \in R^{n_k \times 1}$ 是正交的随机偏置。

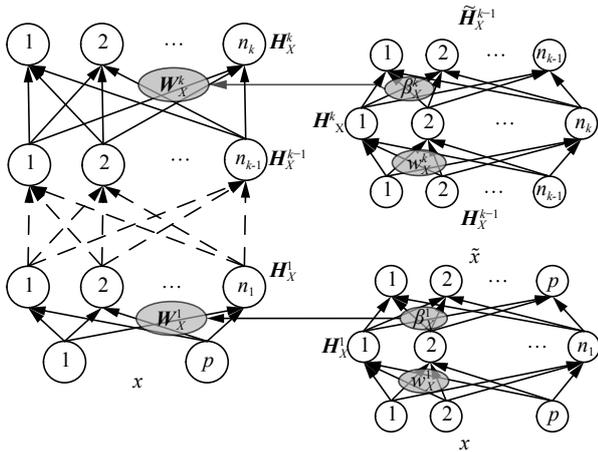


图 3 ELM-AE 无监督非线性映射

Fig. 3 Unsupervised nonlinear learning of ELM-AE

ELM-AE 的输出权值 β 负责从特征空间到输入数据的学习转换。可以根据以下式子计算输出权值 β :

$$\beta_X^k = \left(\frac{I}{C} + (H_X^k)^T H_X^k \right)^{-1} (H_X^k)^T H_X^{k-1} \quad (13)$$

式中: C 是正则项参数, H_X^k 是 ELM-AE 的隐含层输出, H_X^{k-1} 是它的输入数据。

因此, 第 $k-1$ 层到第 k 层的连接权值 W_X^k 为 $W_X^k = (\beta_X^k)^T$ 。

2.2 极大化非线性映射的相关性

X 和 Y 分别经过多层非线性映射, 得到特征 $H_X \in R^{n_k \times N}$ 和 $H_Y \in R^{n_k \times N}$, 对 H_X 和 H_Y 执行线性 CCA:

$$(A, B) = \arg \max_{A, B} \frac{A^T \Sigma_{XX} B}{\sqrt{A^T \Sigma_{XY} A} \sqrt{B^T \Sigma_{YY} B}} \quad (14)$$

由于式 (14) 中 A 或 B 等比例变化时, 目标函数的值不变, 因此 CCA 定义了下式约束条件 $A^T \Sigma_{XY} A = I, B^T \Sigma_{YY} B = I$, 此时, 目标函数等价于:

$$(A, B) = \max A^T \Sigma_{XY} B \quad (15)$$

上述优化的具体过程如下, 将 $H_X \in R^{n_k \times N}$ 中心化: $\hat{H}_X = H_X - H_X/N$ 。

计算 \hat{H}_X 的协方差矩阵:

$$\Sigma_{XX} = \frac{1}{N-1} \hat{H}_X \hat{H}_X^T + \gamma_X I \quad (16)$$

式中: γ_X 为正则项, I 为单位矩阵。同样的方式中心化 $H_Y \in R^{n_k \times N}$, 并计算协方差矩阵 Σ_{YY} 。

计算 H_X 和 H_Y 的交叉协方差矩阵:

$$\Sigma_{XY} = \frac{1}{N-1} \hat{H}_X \hat{H}_Y^T \quad (17)$$

对矩阵 $T = \Sigma_{XX}^{-1/2} \Sigma_{XY} \Sigma_{YY}^{-1/2}$ 奇异值分解, 得到左奇异矩阵 α_L 和右奇异矩阵 η_L , 则可得到优化目标值:

$$\hat{A} = \alpha_L \Sigma_{XX}^{-1/2}, \hat{B} = \eta_L \Sigma_{YY}^{-1/2} \quad (18)$$

经过矩阵 A 和 B 转换, 使最终输出 U 和 V 实现最大相关, 从而实现对 X 和 Y 的非线性典型相关分析学习。

ELM-CCA 与 DCCA 具有相同的模型结构, 二者均通过深度神经网络学习特征的非线性表示, 然后对得到的深度特征进行线性 CCA 求解, 最终将非线性模态相关性转化为线性相关分析。不同的是, 在神经网络参数的优化过程中, DCCA 首先由式 (4) 计算目标函数对深度特征的梯度, 然后根据反向传播算法优化每层的权值。而在 ELM-CCA 中, 每层之间的权值是通过前向逐层训练, 由式 (13) 直接计算得到。相比 DCCA, 不需要反向微调, 极大地提高了训练速度。

3 实验结果

为了验证本文提出的 ELM-CCA 算法在多模态特征提取中的有效性, 我们在康奈尔大学抓取数据集^[17]上开展了实验对比。机器抓取是指机器人根据传感器采集到的信息来推断夹持器放置的位置, 抓取过程涉及感知、规划、识别和控制等问题。在实施抓取操作之前, 机器人首先需要对物体可抓取部分做出准确的识别, 因此本文将提出的算法应用于机器抓取识别任务中。

康奈尔大学抓取数据集包括由机器人视觉感知部位 RGB-D 相机采集到的一系列图像, 图 4 给出了部分图像样本。相比传统的 2-D 图像, 采用 RGB-D 相机能够在获取彩色图像的同时得到每个像素的深度信息, 提高了机器人抓取的成功率^[18]。文献^[19]通过对物体深度信息进行处理, 实现了目标抓取任务。近年来, 相关文献的研究也表明了多模态特征融合的必要性^[20], 文献^[21]提出采用多特征光学遥感图像提高目标的分类识别性能; 文献^[22]指出 RGB-D 融合的多应用场合。因

此本文将对图像的颜色 RGB 模态和深度信息模态进行研究, 寻找其相关信息。



图 4 康奈尔大学数据集样本

Fig. 4 Example objects from the Cornell grasping dataset

3.1 实验数据

实验数据集包含 885 张 RGB 图像, 这些图像来自于对 240 个不同物体的多角度拍摄。每张图片被多个已标签矩形框标记, 共记 8 019 个, 如图 5 所示, 粗线对应夹持器抓取的位置。其中正矩形表示可抓取, 负矩形表示当前状态不可抓取。

每个矩形框对应一个样本, 同时对这些矩形框内的图像提取颜色信息和深度信息, 分别对应算法中的 X 和 Y 。颜色特征为三通道 24×24 像素的 RGB 图像, 即 $X \in R^{1728}$ 。深度特征包含单通道图像深度信息, 即 $Y \in R^{576}$ 。图 5 右所示为抓取识别任务的实验流程。

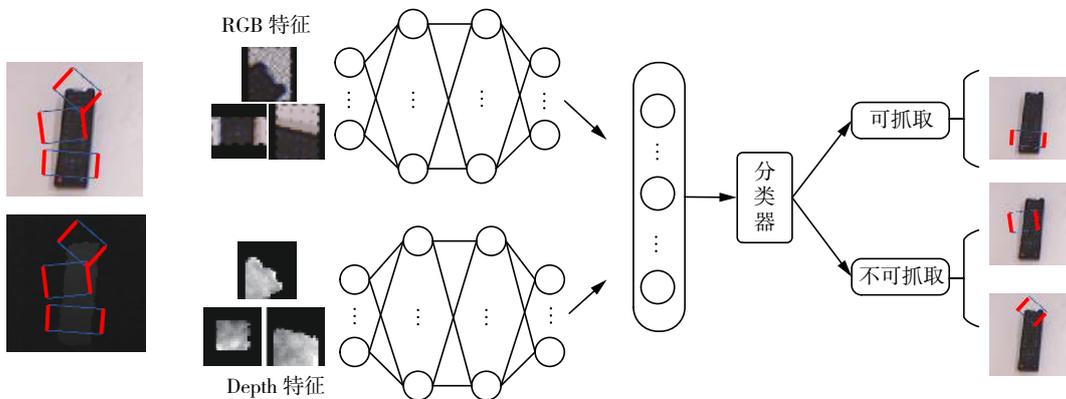


图 5 抓取识别任务流程

Fig. 5 The process of grasping recognition tasks

3.2 实验结果

为了体现本文算法的优越性, 将实验结果与线性 CCA、分数阶嵌入典型相关分析 (fractional-order embedding canonical correlation analysis, FECCA)^[4]和 DCCA 进行比较。分别在相关性、时间和抓取识别任务 3 个方面验证了 ELM-CCA 的优势。

CCA 是对随机变量线性映射的统计方法, FECCA 使用分子阶思想, 对传统 CCA 类内和类间协方差进行重新估计, 以减小训练样本噪声造成的样本协方差偏离真实数据的问题。DCCA 和 ELM-CCA 都属于非线性映射方法。上述方法的目标都是极大化两个随机变量之间的相关系数。因此, 首先对以上算法学习的特征进行相关度的比较。图 6 为不同特征下, 4 种方法在测试集下得到的前 100 个最相关典型变量的相关系数之和。从图中可以看出, ELM-CCA 学习到的特征相关度始终高于其他算法。尤其在低维时表现突出, 当输出维度为 100 时, 采用 ELM-CCA 得到的相关系数之和相比前两者优势最大。输出维度较高时, 相关度基本趋于一致。

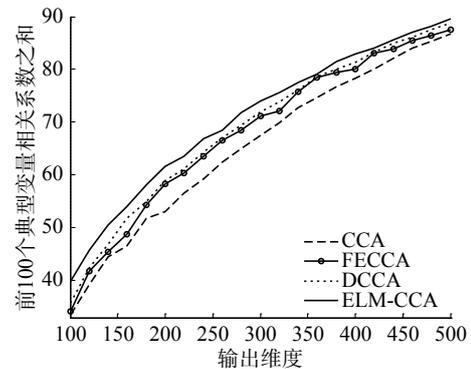


图 6 前 100 个典型变量相关度随映射维度变化曲线

Fig. 6 The correlation curve of the top 100 canonical variables with the increase of output dimensions

图 7 对 DCCA 和 ELM-CCA 在不同网络结构下, 提取特征至 100 维时, 进行了时间和相关度的比较, 横坐标表示隐含层的规模。从图 7(a) 可以看出, 随着网络层数和隐含层节点数增加, 算法消耗的时间在显著增长。并且在每个网络结构下, DCCA 算法消耗的时间都远远高于 ELM-CCA。当隐含层层数低于 4 层, 且每层节点数在 2 000 以内时, ELM-CCA 的训练时间可保持在 10 s 之内。而 DCCA 在隐含层结构为 1 000-1 000 时, 训练时间已经达到 466 s, 可以看出 ELM-CCA 在

训练速度上取得了很大的优势。图 7(b) 为在各网络结构映射下得到的特征相关度, ELM-CCA 学习到的特征相关度均高于 DCCA, 随着网络层数增多, 相关度逐渐增加, 当隐含层层数为 3 层时, 得到的结果最好。因此选择该网络结构提取多模态特征, 用于机器人抓取点的识别任务。

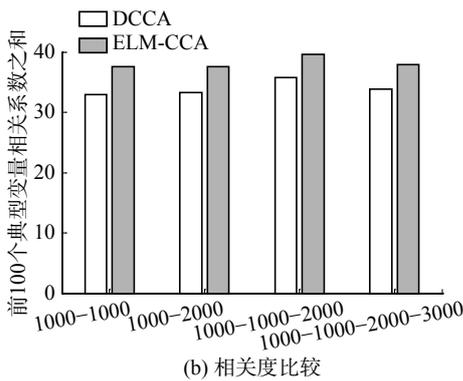
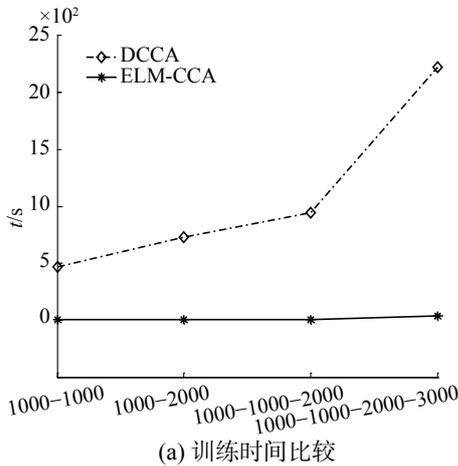


图 7 不同网络结构下 DCCA 和 ELM-CCA 性能比较
Fig. 7 The performance comparison of DCCA and ELM-CCA with different network structures

通过对原始 RGB 和深度特征的非线性映射, 可以得到两个模态特征的最终表示。我们分别将原始特征与经过 CCA、DCCA 和 ELM-CCA 算法提取后的特征输入到分类器中, 通过得到的识别准确率比较不同算法的学习性能。为了使结果更有说服力, 我们在不同的分类器上进行了实验, 选择了基于统计理论的 SVM 和基于神经网络的 ELM。此外, 我们还比较了使用单模态特征 (即 RGB 和 Depth) 和多模态特征融合 (即 RGBD) 对识别率的影响。

表 1 和表 2 分别给出了使用 SVM 分类器和 ELM 分类器得到的分类结果。从表中可以看出, 尽管使用了不同的分类器, 得到的识别率趋势是相同的, 经过特征提取后的识别率要高于直接使用原始特征进行识别的结果。而 DCCA 和 ELM-CCA 经过非线性的学习后, 体现出了比线性 CCA 和 FECCA 的优势, 其中 FECCA 通过对类内

和类间协方差重新估计, 相比传统 CCA, 性能得到了改善。ELM-CCA 在经过快速的特征学习后, 相比 DCCA, 仍然取得了满意的识别率。图 8 所示为不同输出维度下, 采用 ELM 分类器在 RGBD 特征输入下的识别结果。在低维时, 识别率随着输出维度的增加显著提升, 在输出维度达到 20 维时基本稳定, 并且在不同维度下, ELM-CCA 始终体现了较好的识别效果。

表 1 在 SVM 分类器上得到的识别率
Table 1 The recognition rate obtained by the SVM classifier %

模态	原始特征	CCA	FECCA	DCCA	ELM-CCA
RGB	86.97	87.31	87.47	91.02	91.27
Depth	83.30	83.32	83.60	89.53	91.08
RGBD	89.83	90.22	90.27	94.33	94.70

表 2 在 ELM 分类器上得到的识别率
Table 2 The recognition rate obtained by the ELM classifier %

模态	原始特征	CCA	FECCA	DCCA	ELM-CCA
RGB	86.93	90.13	90.77	91.52	91.90
Depth	87.26	83.06	87.59	89.59	90.63
RGBD	88.99	92.27	93.02	93.20	95.14

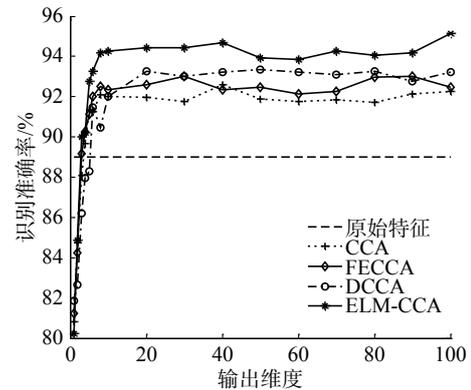


图 8 抓取点识别率随输出维度的变化
Fig. 8 The recognition rate of grasping point along with the increasing of output dimension

4 结束语

本文提出一种新的 ELM-CCA 非线性典型相关分析方法, 并应用于机器人抓取点的识别任务中。对机器人采集到的原始图像提取 RGB 模态和深度模态, 首先采用 ELM 分别的对每个模态进行无监督特征学习, 得到非线性特征表示, 然后将学习到的特征通过 CCA 极大化模态之间的相关性。该方法在保证了识别率较高的情况下还体现了强大的快速性。在实际应用中, 为机器人实现快速准确的抓取操作奠定了基础。

参考文献:

[1] HOTELLING H. Relations between two sets of variates[J].

- Biometrika, 1936, 28(3/4): 321–377.
- [2] SARGIN M E, YEMEZ Y, ERZIN E, et al. Audiovisual synchronization and fusion using canonical correlation analysis[J]. IEEE transactions on multimedia, 2007, 9(7): 1396–1403.
- [3] 侯书东, 孙权森. 稀疏保持典型相关分析及在特征融合中的应用[J]. 自动化学报, 2012, 38(4): 659–665.
HOU Shudong, SUN Quansen. Sparsity preserving canonical correlation analysis with application in feature fusion [J]. Acta automatica sinica, 2012, 38(4): 659–665.
- [4] YUAN Yunhao, SUN Quansen, GE Hongwei. Fractional-order embedding canonical correlation analysis and its applications to multi-view dimensionality reduction and recognition[J]. Pattern recognition, 2014, 47(3): 1411–1424.
- [5] XING Xianglei, WANG Kejun, YAN Tao, et al. Complete canonical correlation analysis with application to multi-view gait recognition[J]. Pattern recognition, 2016, 50: 107–117.
- [6] RASIWASIA N, PEREIRA J C, COVIELLO E, et al. A new approach to cross-modal multimedia retrieval[C]//Proceedings of the 18th ACM International Conference on Multimedia. New York, NY, USA, 2010: 251–260.
- [7] HARDOON D R, SZEDMAK S, SHAW-TAYLOR J. Canonical correlation analysis: an overview with application to learning methods[J]. Neural computation, 2004, 16(12): 2639–2664.
- [8] ANDREW G, ARORA R, BILMES J, et al. Deep canonical correlation analysis[C]//Proceedings of the 30th International Conference on Machine Learning. Atlanta, Georgia, USA, 2013: III-1247-III-1255.
- [9] WANG Weiran, ARORA R, LIVESCU K, et al. On deep multi-view representation learning[C]//Proceedings of the 32nd International Conference on Machine Learning. Lille, France, 2015: 1083–1092.
- [10] HUANG Guang, ZHU Qinyu, SIEW C K. Extreme learning machine: theory and applications[J]. Neurocomputing, 2006, 70(1/2/3): 489–501.
- [11] CAMBRIA E, HUANG Guangbin, KASUN L L C, et al. Extreme learning machines: trends and controversies[J]. IEEE intelligent systems, 2013, 28(6): 30–59.
- [12] YU Wenchao, ZHUANG Fuzhen, HE Qing, et al. Learning deep representations via extreme learning machines [J]. Neurocomputing, 2015, 149: 308–315.
- [13] SUN Kai, ZHANG Jiangshe, ZHANG Chunxia, et al. Generalized extreme learning machine autoencoder and a new deep neural network[J]. Neurocomputing, 2017, 230: 374–381.
- [14] UZAIR M, SHAFAIT F, GHANEM B, et al. Representation learning with deep extreme learning machines for efficient image set classification[M]//MACINTYRE J. Neural Computing and Applications. London: Springer, 2015: 1–13.
- [15] WEI Jie, LIU Huaping, YAN Gaowei, et al. Robotic grasping recognition using multi-modal deep extreme learning machine[J]. Multidimensional systems and signal processing, 2017, 28(3): 817–833.
- [16] WANG Weiran, ARORA R, LIVESCU K, et al. Unsupervised learning of acoustic features via deep canonical correlation analysis[C]//Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing. Brisbane, Australia, 2015: 4590–4594.
- [17] LENZ I, LEE H, SAXENA A. Deep learning for detecting robotic grasps[J]. The international journal of robotics research, 2015, 34(4/5): 705–724.
- [18] JIANG Yun, MOSESON S, SAXENA A. Efficient grasping from RGBD images: learning using a new rectangle representation[C]//Proceedings of 2011 IEEE International Conference on Robotics and Automation. Shanghai, China, 2011: 3304–3311.
- [19] 韩峥, 刘华平, 黄文炳, 等. 基于 Kinect 的机械臂目标抓取[J]. 智能系统学报, 2013, 8(2): 149–155.
HAN Zheng, LIU Huaping, HUANG Wenbing, et al. Kinect-based object grasping by manipulator[J]. CAAI transactions on intelligent systems, 2013, 8(2): 149–155.
- [20] 温有福, 贾彩燕, 陈智能. 一种多模态融合的网络视频相关性度量方法[J]. 智能系统学报, 2016, 11(3): 359–365.
WEN Youfu, JIA Caiyan, CHEN Zhineng. A multi-modal fusion approach for measuring web video relatedness [J]. CAAI transactions on intelligent systems, 2016, 11(3): 359–365.
- [21] 姬晓飞, 秦宁丽, 刘洋. 多特征的光学遥感图像多目标识别算法[J]. 智能系统学报, 2016, 11(5): 655–662.
JI Xiaofei, QIN Ningli, LIU Yang. Research on multi-feature based multi-target recognition algorithm for optical remote sensing image[J]. CAAI transactions on intelligent systems, 2016, 11(5): 655–662.
- [22] CAMPLANI M, PAIEMENT A, MIRMEHDI M, et al. Multiple human tracking in RGB-D data: a survey[J]. arXiv: 1606.04450, 2016.

作者简介:



温晓红,女,1993年生,硕士研究生,主要研究方向为智能控制、模式识别、多模态融合。



刘华平,男,1976年生,副教授,博士生导师,主要研究方向为机器人感知、学习与控制,多模态信息融合。



阎高伟,男,1970年生,教授,主要研究方向为复杂工业控制系统、智能控制理论及其应用、机器学习与软测量建模。