

DOI: 10.11992/tis.201803022

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20180522.1345.002.html>

基于宽度学习方法的多模态信息融合

贾晨¹, 刘华平^{2,3}, 续欣莹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. 清华大学 智能技术与系统国家重点实验室, 北京 100084)

摘要: 多模态机器学习通过有效学习各个模态的丰富特征来解决不同模态数据的融合问题。考虑到模态间的差异性, 基于宽度学习方法提出了一个能够学习和融合两种模态特征的框架, 首先利用宽度学习方法分别提取不同模态的抽象特征, 然后将高维特征表示在同一个特征空间进行相关性学习, 并通过非线性融合得到最后的特征表达, 输入分类器进行目标识别。相关实验建立在康奈尔大学抓取数据集和华盛顿大学 RGB-D 数据集上, 实验结果验证了相比于传统的融合方法, 所提出的方法具有更好的稳定性和快速性。

关键词: 宽度学习方法; 多模态融合; 相关性分析; 特征提取; 非线性变换; 目标识别; 神经网络; RGB-D 图像分类

中图分类号: TP391 文献标志码: A 文章编号: 1673-4785(2019)01-0150-08

中文引用格式: 贾晨, 刘华平, 续欣莹, 等. 基于宽度学习方法的多模态信息融合[J]. 智能系统学报, 2019, 14(1): 150-157.

英文引用格式: JIA Chen, LIU Huaping, XU Xinying, et al. Multi-modal information fusion based on broad learning method[J]. CAAI transactions on intelligent systems, 2019, 14(1): 150-157.

Multi-modal information fusion based on broad learning method

JIA Chen¹, LIU Huaping^{2,3}, XU Xinying¹, 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, Tsinghua University, Beijing 100084, China)

Abstract: Multi-modal machine learning solves the fusion problem that arises in data with different modalities by effectively learning their rich characteristics. Considering the differences between various modalities, we propose a framework that can learn and fuse two kinds of modal characteristics based on the broad learning method. This method first extracts different abstract characteristics, then represents the high-dimension features in the same space to determine their correlation. We obtain a final representation of these characteristics by nonlinear fusion and inputs these characteristics into a classifier for target recognition. Relevant experiments are conducted on the Cornell Grasping Dataset and the Washington RGB-D Object Dataset, and our experimental results confirm that, compared with traditional fusion methods, the proposed algorithm has greater stability and rapidity.

Keywords: broad learning method; multi-modal fusion; correlation analysis; feature extraction; nonlinear transformation; object recognition; neural networks; RGB-D images classification

现实世界中数据信息是以包括图像、文本、声音、纹理等多种模态形式存在的。信息和科技智能化发展为机器学习方法提供了硬件条件。近

年来, 以多模态为研究对象的机器学习算法不断涌现, 使得多模态机器学习成为科研人员广泛关注的研究热点。以机器人为例, 越来越多的研究定位于机器人感知系统的设计, 仅靠单模态中典型的视觉图像分割^[1]、物体检测^[2]等任务已经不能完全满足机器人的识别与感知需求。在机器人手势交互领域, 视觉手势与其他模态(表情、方向、

收稿日期: 2018-03-16. 网络出版日期: 2018-05-23.

基金项目: 国家自然科学基金项目(61673238); 国家高技术研究发展计划课题(2015AA042306); 山西省回国留学人员科研资助项目(2015-045, 2016-044).

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

声音、触觉等)相结合能够增强交互的高效性和互补性^[3]。将不同隐含层局部模态信息融合得到的全局特征用于人脸识别,可以多元化检测人脸并解决遮挡问题^[4]。在机器人情感分析方面,有效结合文本和图像特征并对融合语义进行训练,会得到优于单一模态的情感分类效果^[5]。此外,对多模态异构数据进行相关性度量有助于提高视频检索效率^[6]。实现基于道路场景理解的多模态信息(RGB-D、超声、激光雷达等)融合可以提高复杂环境下环境感知和道路识别的准确性与可靠性,有助于机器人实时路径规划和自主定位与导航^[7-10]。

以上研究大多基于时下流行的深度学习算法。考虑到深度结构运行时间长、难以收敛等性质,本文旨在寻求一种简单、高效的机器学习技术,在降低模型训练时间、节约成本的基础上对不同模态信息进行融合学习,提高系统识别的准确率。宽度学习系统(broad learning system, BLS)方法以其强大的数学理论支撑、简单的平层网络结构、快速的增量建模过程等优点,成功地应用于各种分类和回归任务中,尤其在图像分类问题上表现出优越的性能^[11]。因此,本文基于宽度学习方法提出了一种用于解决多模态数据特征融合问题的模型框架,通过对各个模态并行学习、联合映射、串联融合后根据得到的特征实现分类输出。实验结果表明,本文提出的融合方法能有效提高系统的分类性能。

1 相关知识

1.1 多模态机器学习

一般而言,模态是指事件发生或客观物体存在的形态。为了更好地利用人工智能来帮助我们感知和理解世界,就需要解释和推理出多模态数据中有用的信息和特征。多模态机器学习旨在建立一个可以处理和关联多种模态之间交互信息的模型框架,从早期基于视听语音识别的研究到近期新兴的语义和视觉领域应用,多模态机器学习逐渐发展成一个有巨大挖掘潜力和研究价值的交叉学科。

多模态机器学习虽然能够学习不同模态数据的深层特征表达,但是由于模态间的差异大小和影响因子的不同,目前多模态技术还存在以下3个方面的挑战。

1) 特征表达:对于同一种物体来说,不同的模态形式可能包含了同样的信息,因此,如何去除冗余属性,更好地总结和表示出不同模态的互补信息,仍是值得深入研究和探讨的问题。

2) 关联和转换:如何将两个不同模态的信息

关联起来,或者怎样把一种模态映射到另一种模态空间中去,是一个巨大的挑战。

3) 融合与共同学习:在同一个特征映射空间中,将两种模态相互融合,使每个模态对最终的输出都产生相应的影响,即共同学习。

1.2 宽度学习方法

传统的神经网络如BP网络,因其反向传播计算的运行时间长、容易陷入局部最优等缺陷,网络的分类性能往往受初始化区域的影响较大。为了有效解决这些问题,不少研究者致力于寻求简单的单层网络模型,利用广义逆直接求解全局最优来优化网络的效率,因此基于宽度的网络结构逐渐发展起来。典型的方法有单层前馈神经网络(single layer feedforward neural networks, SLFN)^[12]、随机向量功能连接网络(random vector functional link neural network, RVFL)^[13-15]、极限学习机(extreme learning machine, ELM)^[16-18]等。

宽度学习方法^[11]是由澳门大学陈俊龙教授提出的一种基于RVFL平面网络结构的增量学习算法,模型结构如图1所示。与传统RVFL结构不同的是,宽度学习系统的输入权值矩阵不是随机生成,而是通过稀疏自编码方式经编码后,在解码过程中选取了最优权值。宽度学习方法的输入样本经过一次线性变换后将特征表达映射在特征平面上形成特征节点,得到的特征节点再经过激活函数非线性变换生成增强节点。特征节点和增强节点共同连接作为系统的实际输入信号,经由连接矩阵线性输出。与RVFL相同,考虑到经典BP算法的高时间成本和容易陷入局部最优等缺点,宽度学习方法采用岭回归广义逆直接求取输出连接矩阵。

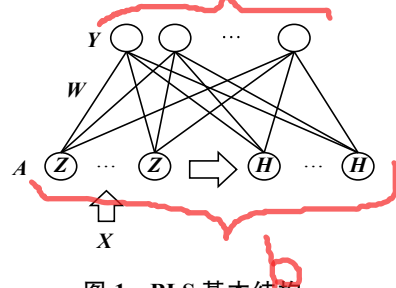


图1 BLS基本结构

Fig. 1 Basic BLS structure

给定输入数据 $X \in \mathbf{R}^{N \times M}$,其中 N 表示输入样本个数, M 表示每个样本向量的特征维数。假设特征节点的个数为 b ,则根据宽度结构可以得到特征平面上的特征如式(1):

$$Z^{N \times b} = X^{N \times M} \cdot W_e^{M \times b} \quad (1)$$

式中 W_e 是由稀疏自编码求得的最优输入权值矩阵。若生成 d 个增强节点,高层特征可表示为

$$H^{N \times d} = \phi(Z^{N \times b} \cdot W_h^{b \times d} + \beta_h^{N \times d}) \quad (2)$$

式中: W_h 和 β_h 分别代表随机矩阵和偏置; $\phi(\cdot)$ 是一个可选的非线性激活函数。将特征节点和增强节点相连得到的合并矩阵作为系统的实际输入, 并假设输出矩阵为 $Y \in \mathbf{R}^{N \times Q}$, 则宽度模型可以由式 (3) 求出:

$$Y^{N \times Q} = A^{N \times (b+d)} \cdot W^{(b+d) \times Q} = [Z^{N \times b} | H^{N \times d}] \cdot W^{(b+d) \times Q} \quad (3)$$

式中: A 代表 BLS 的实际输入矩阵; W 代表输出连接权值矩阵, 且 W 是通过 A^+ 的岭回归近似根据式 (4) 计算得到的:

$$A^+ = \lim_{\lambda \rightarrow 0} (\lambda I + A A^T)^{-1} A^T \quad (4)$$

为了直观地展示出宽度学习系统, 图 2 给出了 BLS 平面结构的立体模型, 详细描述了一个输入样本 X 通过 BLS 网络进行学习的全过程。其中, Z 表示特征节点层, H 表示增强节点层, 输出层为 Y , 网络只需要学习输出矩阵 W 即可。

其他矩阵可以直接计算

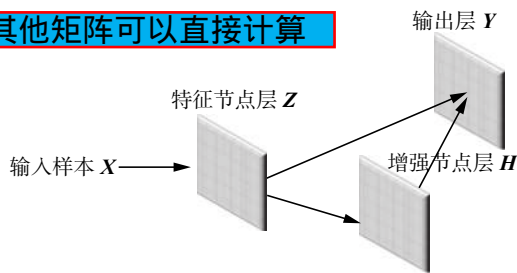


图2 BLS的3D模型
Fig. 2 3D BLS model

1.3 典型相关性分析

典型相关性分析 (CCA)^[19] 是对两个数据集进行联合与降维的经典算法。对于两个没有显性关联的数据集, CCA 可以把不同模态的数据点映射到同一个特征空间来构造关联规则。图 3 为 CCA 的基本匹配过程, 通过将两种模态的数据样本进行连接, 就可以得到相关特征子空间内一一对应的特征表达。

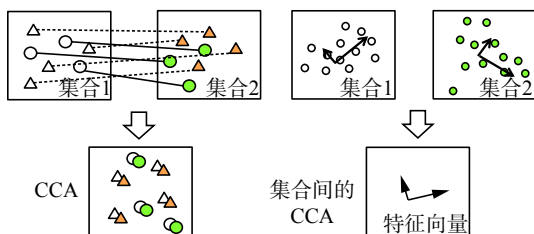


图3 CCA的匹配方式
Fig. 3 CCA match way

根据 CCA 的关联方式, 假设两个分别属于不同模态, 样本个数一一对应的 n 维数据集 X 和 Y , 其中 $X = \{x_1, x_2, \dots, x_n\}$, $Y = \{y_1, y_2, \dots, y_n\}$ 。通过映射基向量 u 和 v 进行线性变换, 重新为每个集合中的数据点寻找一个新的坐标空间, 并得到度量

模态间相关性的参数 ρ , 即

$$\rho = \max_{u,v} \frac{u^T \Sigma_{xy} v}{\sqrt{u^T \Sigma_{xx} u} \sqrt{v^T \Sigma_{yy} v}} \quad (5)$$

式中 Σ_{xy} 、 Σ_{xx} 和 Σ_{yy} 代表类间和类内协方差矩阵, 即

$$\begin{aligned} \Sigma_{xy} &= E[xy^T] = \frac{1}{n} \sum_{i=1}^n x_i y_i^T \\ \Sigma_{xx} &= E[xx^T] = \frac{1}{n} \sum_{i=1}^n x_i x_i^T \\ \Sigma_{yy} &= E[yy^T] = \frac{1}{n} \sum_{i=1}^n y_i y_i^T \end{aligned} \quad (6)$$

实际应用中, 式 (5) 表示的优化问题可以转化为特征值求解问题进行相关计算。

2 基于双宽度学习的典型相关性分析

为了实现多模态信息的融合问题, 本文对宽度学习方法进行了一定程度的创新和改进, 并提出了一个可以有效学习两种模态特征的融合算法框架。

2.1 双宽度学习结构

双宽度学习结构是一个以 BLS 为基本单元的双宽度学习 (double broad learning, DBL) 框架, 如图 4 所示, 此结构主要由两个 BLS 单元组成, 用于处理不同模态的融合问题。图中 Z 和 H 分别表示特征节点和增强节点, 下标 C 、 D 、 R 代表彩色通道、深度通道和融合节点层。当两种模态数据输入系统时, 模型训练的过程可分为以下 3 个部分:

- 1) 利用宽度学习单元分别提取每个模态的高维特征, 主要包括特征节点映射特征和增强节点映射特征;
- 2) 在融合节点层将两个模态的特征映射经过一个非线性变换抽象融合起来;
- 3) 通过学习输出权值矩阵, 利用岭回归广义逆直接求取全局最优解得到输出类别属性。

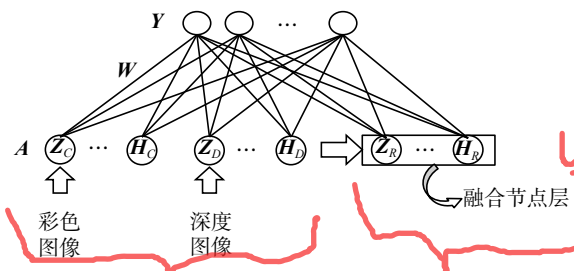


图4 DBL基本结构
Fig. 4 Basic DBL structure

2.1.1 特征提取

假设 DBL 模型的输入样本个数为 N , 设 DBL 的特征节点和增强节点个数分别为 N_1 和 N_2 , 则彩色图像的特征表达式为

$$A_c = [Z_1 | H_1] \quad (7)$$

式中由一个 BLS 单元生成、代表彩色图像的特征节点和增强节点分别可以表示为

$$\begin{aligned} Z_1 &= \{z_i | z_i \in \mathbf{R}^{N_1}\}_{i=1}^N \\ H_1 &= \{h_j | h_j \in \mathbf{R}^{N_2}\}_{j=1}^N \end{aligned} \quad (8)$$

同理, 另一个 BLS 单元生成的深度图像总特征以及特征节点、增强节点分别为

$$\begin{aligned} A_d &= [Z_2 | H_2] \\ Z_2 &= \{z_i | z_i \in \mathbf{R}^{N_1}\}_{i=1}^N \\ H_2 &= \{h_j | h_j \in \mathbf{R}^{N_2}\}_{j=1}^N \end{aligned} \quad (9)$$

由于 BLS 生成特征节点过程中采用了稀疏编码, 在特征提取过程中自动去除了冗余信息, 减小了训练的计算复杂度, 根据多模态机器学习概念, 仅需要考虑后期不同模态融合过程中的信息互补即可。为了更好地学习彩色和深度这两种模态的共同特征, 在输入融合节点层之前, 需要将彩色图像和深度图像进行混合, 使得两种模态的图像映射到同一个样本空间。考虑到神经网络的学习特性, DBL 仅简单地将两种模态的特征并联起来, 作为最后提取的总特征和融合空间的净输入:

$$F^{N \times 2(N_1 + N_2)} = [A_c^{N \times (N_1 + N_2)} | A_d^{N \times (N_1 + N_2)}] \quad (10)$$

2.1.2 特征的融合与分类

对于特征融合部分, 由 2.1.1 节的介绍和图 5 所示立体模型可知, 彩色图像模态和深度图像模态混合的总特征 F 就是 DBL 融合节点层的新输入。值得注意的是, 为了更好地将两种图像模态的信息结合起来, 本文引入融合节点映射层, 借鉴传统神经网络的非线性拟合特点将不同的模态特征抽象融合, 最后利用输出矩阵进行快速有效地学习, 以提高系统的分类性能。假设融合节点个数为 N_3 , 则融合节点层输出为

$$T^{N \times N_3} = \phi(F^{N \times 2(N_1 + N_2)} \cdot W_t^{2(N_1 + N_2) \times N_3} + b_t^{N \times N_3}) \quad (11)$$

式中选取的 $\phi(\cdot)$ 是一个 S 型非线性激活函数。

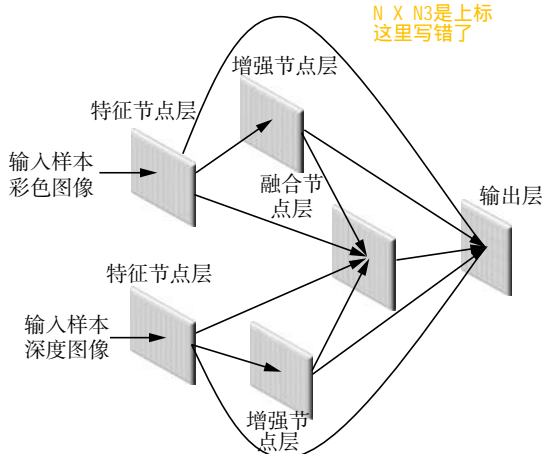


图 5 DBL 的 3D 模型

Fig. 5 3D DBL model

图中融合节点层融合节点层是直接由式 (11) 计算得出, 再进行 DBL

基于 1.2 节对 BLS 的描述, 真正意义上的宽度学习应该是将特征节点、增强节点、融合节点作为共同的特征, 按照不同的权值共同作用于输出的网络。因此, DBL 模型结构的输出连接矩阵 W 包含了两种模态的特征节点层、增强节点层和最后一层融合节点层的总权值, 可以很容易地由式 (12) 所示的岭回归广义逆求得:

$$W^{[2(N_1 + N_2) + N_3] \times O} = [F^{N \times 2(N_1 + N_2)} | T^{N \times N_3}]^+ Y^{N \times O} \quad (12)$$

式中 Y 是由样本标签组成的 O 维 one-hot 形式的期望输出矩阵。

2.2 基于双宽度学习的典型相关性分析

图 6 表示基于双宽度学习的典型相关性分析 (DBL-CCA) 结构的多模态信息融合和学习过程。以 DBL 为基础, 把 2 种模态经过特征节点和增强节点粗提取的样本特征统一映射到 CCA 生成的特征子空间上进行学习、匹配和降维, 然后再由融合节点直接非线性化融合, 连接到输出层进行分类输出。

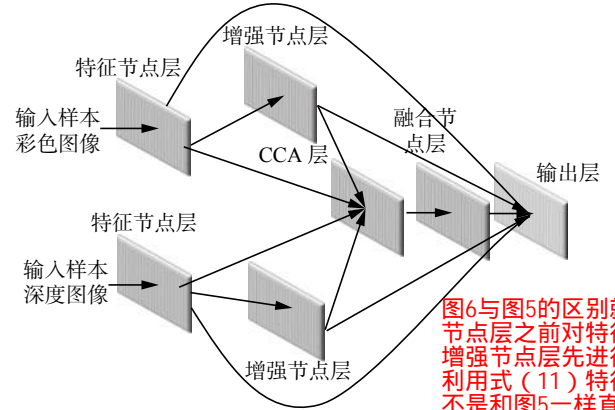


图 6 DBL-CCA 模型

Fig. 6 DBL-CCA model

图6与图5的区别就是在融合节点层之前对特征节点层和增强节点层先进行CCA, 再利用式 (11) 特征融合。而不是和图5一样直接就利用式 (11) 给出融合节点层, 再进行DBL

因此, CCA 在整个融合体系中承担着实现多模态机器学习关联和转换任务的作用, 同时 CCA 的降维特性为简化模型提供了便利。由 DBL 结构可知, CCA 层的净输入为式 (10) 求出的特征 F , 其前 L 个最大主成分输出 $P^{N \times 2L}$ 即为融合节点层的实际输入。因此, 式 (11) 变为

$$T^{N \times N_3} = \phi(P^{N \times 2L} \cdot W_t^{2L \times N_3} + b_t^{N \times N_3}) \quad (13)$$

3 实验分析

3.1 康奈尔大学抓取数据集

康奈尔大学抓取数据集^[20]是判断给定物体是否可被机器人抓取的开源数据集, 包含 240 个待判断的物体, 共 885 幅图像。每个图像划分为若干个抓取矩形框, 分别代表不同的抓取位置。根据给定标签 (正表示可抓取, 负表示不可抓取) 可

知,这是一个二分类的任务。已知矩形框共有 8 019 个,按照 8:2 的随机划分方式可得到 6 415 个训练样本和 1 604 个测试样本,每个样本均包含尺寸为 24×24 的 3 通道彩色图像和单通道深度图像 2 种模态。

图 7 为抓取数据集的 20 个简单样例。图 8 展示了实验采用的 4 种分类模型,每个模型中的矩形框均代表机器人夹板的抓取操作矩形,矩形框的宽边为夹板的初始位置,矩形框的长边表示机器人夹板的运行轨迹。



图 7 抓取数据集样例

Fig. 7 Grasp dataset examples

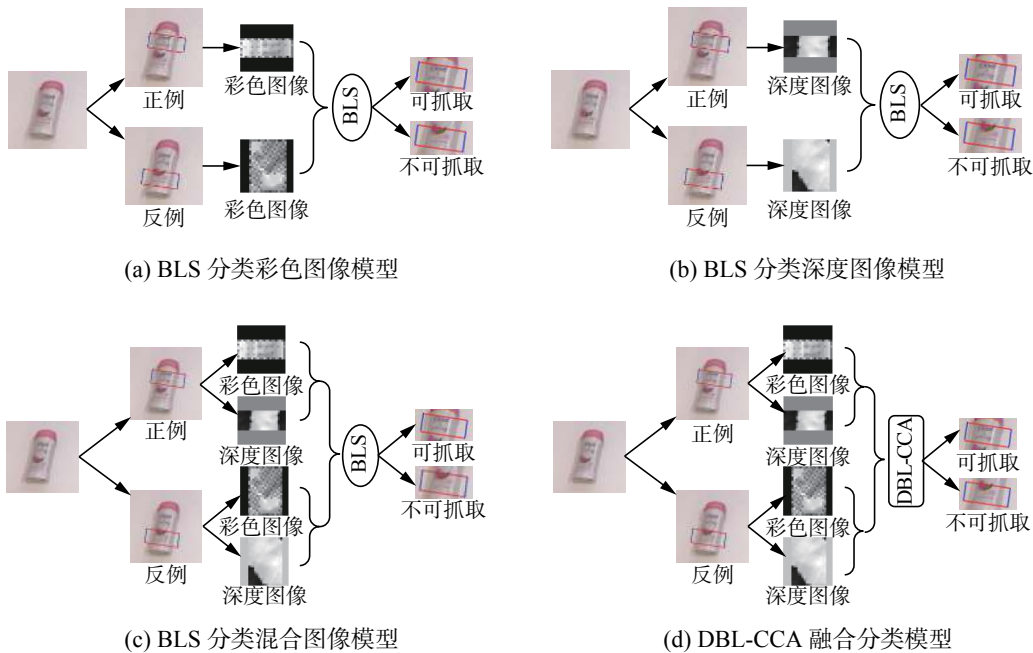


图 8 实验采用的模型

Fig. 8 Models used in our experiments

DBL-CCA 算法先通过特征提取及融合处理彩色图像和深度图像,之后输入分类器对物体是否可被机器人抓取作出分类判别。为了检验本文模型的融合效果,对比模型采用 BLS 分别对彩色图像、深度图像这两种单一模态以及两种图像的简单混合模态进行学习和分类。

对于上述提出的 4 种模型结构,为了提高各种模型性能的可比性,实验均设置了相同的特征节点 (100) 和增强节点 (400) 个数、收敛因子 (0.5) 和正则化因子 (2^{-30}) 大小等参数,并将 DBL-CCA 中 CCA 层的输出特征维度固定在 200。重复 5 次随机试验得到的平均测试精度、训练时间和测试时间如表 1 所示。从表中可以看出,只用 BLS 对彩色图像进行学习比只用深度图像学习得到的分类精度高、时间长;混合两种模态能够提升分类性能,但特征维度的增加消耗了一些训练时间。DBL-CCA 融合算法的分类精度比以上

3 种方法都要高,运行时间也非常迅速,整体而言占有相当大的优势。

表 1 不同图像模态的识别性能

Table 1 Different image modalities' recognition performances

图像模态	识别率/%	训练时间/s	测试时间/s
彩色	91.60	29.02	4.55
深度	90.90	1.08	0.40
混合	92.60	35.03	6.60
融合	94.80	3.52	0.54

在融合彩色图像和深度图像的过程中,CCA 层的输出特征维度会对融合结果产生一定程度的影响,图 9 比较了 DBL-CCA 与 CCA、Cluster-CCA^[21]等算法在不同输出维度下的测试精度。可以看出,在 CCA 层输出特征维度为 25 时,DBL-

CCA 和 Cluster-CCA 的拐点比 CCA 相对明显。随着维度的增加, DBL-CCA 呈上升趋势, 在 200 维处达到最大值 (95.01%); Cluster-CCA 曲线有小幅下降趋势; CCA 精度曲线在 100~200 维之间有小范围的波动。总体来说, DBL-CCA 和 Cluster-CCA 算法相较于 CCA 而言曲线略平缓。因此, 实验结果可以证明 DBL-CCA 无论是从收敛速度、上升趋势还是从分类精度上来看都占有极大的优势, 选取该结构作为多模态图像特征的融合方法具备充分的高效性和稳定性。

为了比较不同参数对模型性能的作用和影响, 实验还对 DBL-CCA 的节点数目和相关因子进行了参数敏感度分析。图 10(a) 中的 N_2 和 N_3 分别代表在 {100, 200, 300, 400, 500, 600, 700, 800} 中变化的增强节点和融合节点数目。增强节点和融合节点对图像识别精度产生的影响并不显著, 说明 DBL-CCA 融合框架具有很好的鲁棒性, 能够过滤

噪声干扰。图 10(b) 为不同相关因子 (r_1 和 r_2) 对模型分类性能的影响程度, 实验选取的参数范围为 $10^{-6} \sim 10^6$ 。从图 10 可以观察到, 相比于节点数目而言, 相关因子对于识别率的影响较大, 有更加明显的局部极大值和极小值区域。

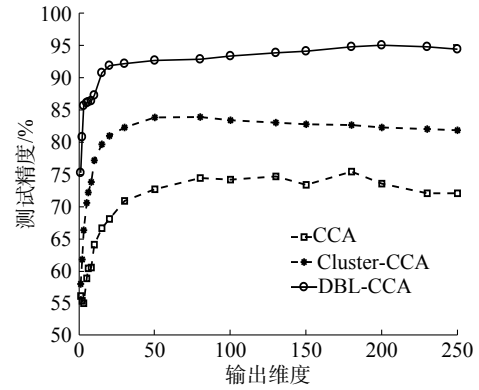
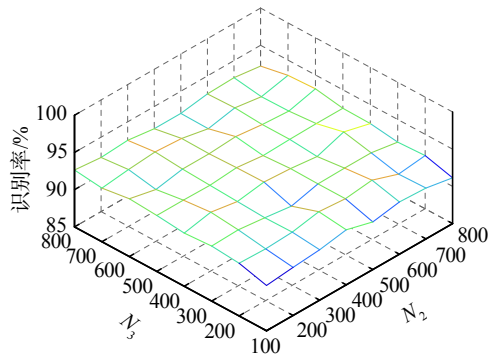
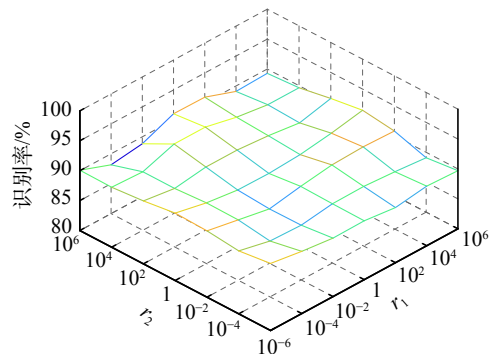


图 9 不同算法的测试精度

Fig. 9 Testing accuracies of various algorithms



(a) 神经元节点数对识别率的影响



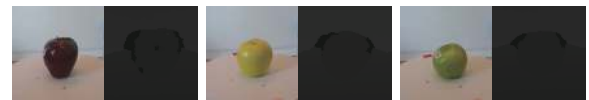
(b) 相关因子对识别率的影响

图 10 不同参数对识别率的影响

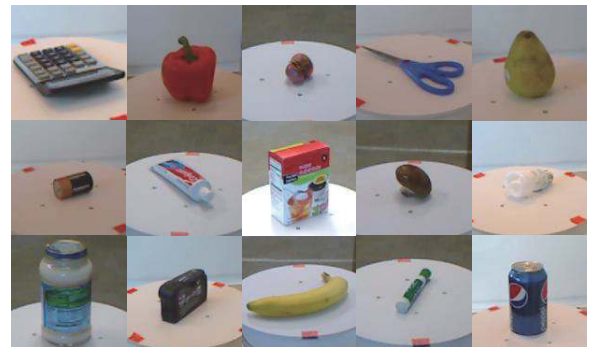
Fig. 10 Recognition rates in different parameters

3.2 华盛顿大学 RGB-D 数据集

为了验证本文提出方法的泛化性能, 采用华盛顿大学 RGB-D 数据集^[22]作进一步的拓展实验。华盛顿大学 RGB-D 数据集由 41 877 个 RGB-D (彩色和深度) 图像组成, 包含了具有 51 种不同类别的共 300 个实例物体。图 11(a) 表示“苹果”这种类别的 3 个具体不同实例的彩色图像及其对应的深度图像, 图 11(b) 为部分样本的彩色图像。该数据集的采集方法是: 将每个物体放置在一个转盘上, 并使用 Kinect 的 3D 相机对转盘的一个完整旋转周期进行录制。每个物体均包含 3 个视频序列, 分别记录在不同高度的摄像机上。实验按照文献^[22]的划分方式选取 45° 角拍摄的图像生成测试集, 由此可以得到用于物体识别的 28 009 个训练样本和 13 868 个测试样本。



(a) 苹果类别的 3 个实例



(b) 数据集中不同样本的彩色图像

图 11 RGB-D 数据集示例

Fig. 11 Several RGB-D dataset examples

在训练模型之前,将彩色模态和深度模态的图像分别调整成大小为 24×24 的 3 通道图像和单通道图像,然后拉伸成 1 728 维和 576 维的行向量。在此基础上根据划分标签类型的不同,可以得到 300 个实例分类和 50 个类别分类的 2 种实验结果。

表 2 说明了 DBL-CCA 融合算法能够有效学习彩色与深度图像的高维非线性特征,其分类识别的最优准确率明显高于单模态和线性混合模型。表 3 为 DBL-CCA 模型在 2 种分类方式下 5 次随机试验的平均训练和测试时间,相比于一般的深度结构,此方法可以有效缩短模型的时间成本,降低空间的计算复杂度。

表 2 不同图像模态的识别准确率
Table 2 Recognition accuracy rates of various image modalities %

图像模态	彩色	深度	混合	融合
300 分类	60.0	29.0	65.2	80.1
50 分类	68.3	60.6	75.4	85.0

表 3 DBL-CCA 在不同分类方式下的消耗时间
Table 3 Time consumptions of DBL-CCA in different methods of classification s

时间	训练时间	测试时间
300 分类	51.2	7.3
50 分类	26.4	6.5

综上所述,本文提出的 DBL-CCA 能够有效提取不同模态图像的丰富信息,完成融合学习和分类任务,并在训练时间、测试精度以及参数影响等方面表现出非常大的优势,使系统保持良好识别性能的同时,兼备了较高的快速性、准确性和鲁棒性。

4 结束语

为了解决实际生活经常遇到的多模态融合问题,本文提出一种新颖的 DBL-CCA 融合方法,通过融合学习各个输入模态的丰富特征对物体属性进行识别。考虑到训练时间、分类精度等方面的制约,DBL-CCA 算法采用了宽度学习方法这种基于平层网络的架构,并在康奈尔大学和华盛顿大学的多模态图像公开数据集上验证了算法的高效性。在具体的研究与应用中,此方法能够帮助机器人完成快速准确的抓取操作;对于复杂环境下可能存在的不同噪声干扰,算法能否持续保持良好的泛化性能,是一个值得深入研究的问题。

参考文献:

- [1] 雷俊,王立辉,何芸倩,等.适用于机器人视觉的图像分割方法[J].系统工程与电子技术,2017,39(7):1653-1659.
LEI Jun, WANG Lihui, HE Yunqian, et al. Image segmentation method for robot vision[J]. Systems engineering and electronics, 2017, 39(7): 1653-1659.
- [2] 毛玉仁,郭松,郑阳明,等.基于似物性判别的视觉目标检测方法[J].传感器与微系统,2017,36(11):147-150.
MAO Yuren, GUO Song, ZHENG Yangming, et al. Visual object detection method based on objectness estimation[J]. Transducer and microsystem technologies, 2017, 36(11): 147-150.
- [3] 齐静,徐坤,丁希仑.机器人视觉手势交互技术研究进展[J].机器人,2017,39(4):565-584.
QI Jing, XU Kun, DING Xilun. Vision-based hand gesture recognition for human-robot interaction: a review[J]. Robot, 2017, 39(4): 565-584.
- [4] 王成济,罗志明,钟准,等.一种多层特征融合的人脸检测方法[J].智能系统学报,2018,13(1):138-146.
WANG Chengji, LUO Zhiming, ZHONG Zhun, et al. Face detection method fusing multi-layer features[J]. CAAI transactions on intelligent systems, 2018, 13(1): 138-146.
- [5] 吴钟强,张耀文,商琳.基于语义特征的多视图情感分类方法[J].智能系统学报,2017,12(5):745-751.
WU Zhongqiang, ZHANG Yaowen, SHANG Lin. Multi-view sentiment classification of microblogs based on semantic features[J]. CAAI transactions on intelligent systems, 2017, 12(5): 745-751.
- [6] 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.
- [7] 吴宗胜,傅卫平,韩改宁.基于深度卷积神经网络的道路场景理解[J].计算机工程与应用,2017,53(22):8-15.
WU Zongsheng, FU Weiping, HAN Gaining. Road scene understanding based on deep convolutional neural network[J]. Computer engineering and applications, 2017, 53(22): 8-15.
- [8] 吴宗胜,傅卫平.移动机器人全局路径规划的模拟退火-教与学优化算法[J].机械科学与技术,2016,35(5):678-685.
WU Zongsheng, FU Weiping. SA and teaching-learning-based optimization algorithm for mobile robots global path planning[J]. Mechanical science and technology for aerospace engineering, 2016, 35(5): 678-685.

- [9] 张文, 刘勇, 张超凡, 等. 基于方向 A*算法的温室机器人实时路径规划[J]. 农业机械学报, 2017, 48(7): 22–28.
ZHANG Wen, LIU Yong, ZHANG Chaofan, et al. Real-time path planning of greenhouse robot based on directional A* algorithm[J]. Transactions of the Chinese society for agricultural machinery, 2017, 48(7): 22–28.
- [10] 张文, 刘勇, 张超凡, 等. 基于语义建图的室内机器人实时场景分类[J]. 传感器与微系统, 2017, 36(8): 18–21, 28.
ZHANG Wen, LIU Yong, ZHANG Chaofan, et al. Real-time scene category of indoor robot based on semantic mapping[J]. Transducer and microsystem technologies, 2017, 36(8): 18–21, 28.
- [11] CHEN C L P, LIU Zhulin. Broad learning system: an effective and efficient incremental learning system without the need for deep architecture[J]. *IEEE transactions on neural networks and learning systems*, 2018, 29(1): 10–24.
- [12] HUANG Guangbin, BABRI H A. Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions[J]. *IEEE transactions on neural networks*, 1998, 9(1): 224–229.
- [13] PAO Y H, TAKEFUJI Y. Functional-link net computing: theory, system architecture, and functionalities[J]. *Computer*, 1992, 25(5): 76–79.
- [14] PAO Y H, PARK G H, SOBAJIC D J. Learning and generalization characteristics of the random vector functional-link net[J]. *Neurocomputing*, 1994, 6(2): 163–180.
- [15] IGELNIK B, PAO Y H. Stochastic choice of basis functions in adaptive function approximation and the functional-link net[J]. *IEEE transactions on neural networks*, 1995, 6(6): 1320–1329.
- [16] HUANG Guangbin, ZHU Qinyu, SIEW C K. Extreme learning machine: theory and applications[J]. *Neurocomputing*, 2006, 70(1/2/3): 489–501.
- [17] HUANG Guangbin, CHEN Lei. Convex incremental extreme learning machine[J]. *Neurocomputing*, 2007, 70(16/17/18): 3056–3062.
- [18] HUANG Guangbin, BAI Zuo, KASUN L L C, et al. Local receptive fields based extreme learning machine[J]. *IEEE computational intelligence magazine*, 2015, 10(2): 18–29.
- [19] HOTELLING H. Relations between two sets of variates[J]. *Biometrika*, 1936, 28(3/4): 321–377.
- [20] 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.
- [21] RASIWASIA N, MAHAJAN D, MAHADEVAN V, et al. Cluster canonical correlation analysis[C]//Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics. Reykjavik, Iceland, 2014: 823–831.
- [22] LAI K, BO Leifeng, REN Xiaofeng, et al. A large-scale hierarchical multi-view RGB-D object dataset[C]//Proceedings of IEEE International Conference on Robotics and Automation. Shanghai, China, 2011: 1817–1824.

作者简介:



贾晨, 女, 1992 年生, 硕士研究生, 中国计算机协会会员, 主要研究方向为智能控制、模式识别、机器视觉、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。



续欣莹, 男, 1979 年生, 副教授, 主要研究方向为粗糙集、粒计算、数据挖掘、计算机视觉。