

文章编号:1007-5321(2017)04-0001-08

DOI:10.13190/j.jbupt.2017.04.001

模式分类中的特征融合方法

刘渭滨¹, 邹智元¹, 邢薇薇²

(1. 北京交通大学 信息科学研究所, 北京 100044; 2. 北京交通大学 软件学院, 北京 100044)

摘要: 特征融合方法是模式识别领域的一种重要方法. 计算机视觉领域的图像识别问题作为一种特殊的模式分类问题, 仍然存在很多挑战. 特征融合方法能够综合利用多种图像特征, 实现多特征的优势互补, 获得更加鲁棒和准确的识别结果. 笔者基于信息融合理论分析了特征融合方法的原理, 介绍了特征融合方法的研究现状, 讨论了特征融合与3类主流基础理论相结合的方法, 其中基于贝叶斯理论的特征融合算法可以实现多特征的融合决策, 基于稀疏表示理论的特征融合算法能够得到多特征的联合稀疏表示, 基于深度学习理论的特征融合算法能够强化深度神经网络模型的特征学习过程.

关键词: 特征融合; 模式识别; 分类

中图分类号: TP391.4

文献标志码: A

Feature Fusion Methods in Pattern Classification

LIU Wei-bin¹, ZOU Zhi-yuan¹, XING Wei-wei²

(1. Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China;

2. School of Software Engineer, Beijing Jiaotong University, Beijing 100044, China)

Abstract: Feature fusion is an important method in pattern recognition. Image recognition problem in computer vision can be known as a special pattern classification problem, and it still have many challenges. To solve the problem, feature fusion method is able to use multi-feature of image, complement each other's advantages and get more robust and accurate results. Based on information fusion theory, analyze the theory of feature fusion and introduce the development of feature fusion methods. Besides, discuss three methods of feature fusion combining with other basic theories. Bayesian decision theory based feature fusion algorithm implements fusion decision of multi-feature. Sparse representation based feature fusion algorithm can get the joint sparse representation of multi-feature. Deep learning based feature fusion algorithm intensifies feature learning process of deep neural network model.

Key words: feature fusion; pattern recognition; classification

图像识别问题是计算机视觉领域要解决的基础问题, 如果将图像看作一种模式, 那么图像识别问题也是一种特殊的模式分类问题, 包括早期的光学字符识别、生物特征识别中的人脸识别、指纹识别、虹膜识别, 也包括场景识别、视频中的目标识别、动作

识别等. 基于传统的模式分类方法, 这类问题已经得到一定程度的解决, 例如支持向量机、人工神经网络、近邻法等方法, 通过特征提取、分类器训练过程得到分类器, 输出分类结果. 但是, 传统的模式分类方法通常基于人工设计的特征, 经过特征提取算法

收稿日期: 2017-03-03

基金项目: 国家自然科学基金项目(61370127, 61473031, 61472030); 新世纪优秀人才支持计划项目(NCET-13-0659); 中央高校基本科研业务费专项资金项目(2014JBZ004); 北京高等学校青年英才计划项目(YETP0583)

作者简介: 刘渭滨(1973—), 男, 教授, 博士生导师; 邹智元(1991—), 男, 博士生, E-mail: zouzhiyuan@bjtu.edu.cn.

得到原始图像的特征数据,例如颜色特征、SIFT 特征、HOG 特征、HOF 特征、GIST 特征等,而这些特征数据却存在类内方差较小而类间方差较大的问题.因为一种特征通常只对图像部分特性的变化较为敏感,而对其他特性的变化不敏感,所以,当两类图像的差异在某种特征敏感特性上的差异不大时,基于单一特征训练的分类器就无法输出正确的分类.除此之外,图像中复杂的背景噪声也会导致特征数据质量下降,既增加分类器训练的难度,又降低分类的准确性.

解决这种问题的一种思路就是使用特征融合方法,同时提取多种特征进行分类器训练,实现特征互补,降低单一特征固有缺陷的影响.特征融合方法的思想来源于早期的信息融合(information fusion)领域,它的数据主要来源于多种传感器,用于军事领域的多传感器融合.早期信息融合的基础理论主要包括模糊集、证据理论等,后来信息融合在民用领域得到了较快的发展,尤其是在生物安全认证方面^[1],进行多生物特征的融合识别.目前在学术界被公认的信息融合层次划分为分类器级(决策级)、特征级和数据级,处理数据的维度和总量依次升高.在模式识别领域,受到计算机运算能力的限制,早期的研究主要集中在分类器级的融合算法,例如,Kitler 等^[2]提出的基于贝叶斯决策理论的分类器融合框架,为以后的研究奠定了良好的基础.近年来,随着人工智能技术的发展,特征融合成为研究热点,特别是在图像识别问题中得到了广泛的应用.

1 特征融合原理

特征融合在信息融合中属于中间层次的融合,信息融合理论是特征融合的基础理论.信息融合是对多源异构数据进行综合处理,从而达到联合决策的目的.其实,人类对事物的认知过程就是对多源信息的融合过程^[3],人们每时每刻都在通过视觉、听觉、触觉、嗅觉、味觉等多种感官接收各种各样的信息,这些信息转化成了神经脉冲信号传输到大脑皮层进行综合决策.实际应用中,由于多源数据的结构和含义不同,在处理时不能一概而论,必须经过融合算法处理.在计算机视频和模式识别领域,使用特征融合方法解决图像识别问题,就是基于信息融合思想,通过融合算法使用图像多角度、多层次的特征(图像的颜色、形状、梯度)、空间特征和时间特征、全局特征和局部特征等,来实现更智能的图像信

息处理.

随着学术研究的推进,信息融合的数据来源更加多样,应用的领域也更加广泛,特别是 CPU、GPU 计算性能的显著提升,大规模分布式计算、并行计算模型进一步发展,人类能够对大数据进行快速处理,这使得信息融合的研究从分类器级推进到了特征级和数据级的层面.如果将图像识别问题看作模式分类问题,那么使用特征融合方法主要基于 2 个基本的经验性假设:1) 融合多特征通常比单一特征具有更好的分类性能;2) 进行融合的多种特征之间相关性较小.前者是应用该方法的出发点和基础思想,后者对于图像多特征的选择具有指导意义.应当从多个角度选择相关性尽量小的特征参与特征融合,才能发挥该方法的优点,实现特征互补.特征融合方法直接作用在特征上面,它的优点在于可以直接利用已有的特征提取算法提取特征,相对于重新设计特征和特征提取算法,其花费的代价更低.

模式分类问题的解决过程一般包括数据获取、预处理、特征提取、分类器设计与训练和分类决策等步骤,如图 1 所示,信息融合的 3 个层次恰好可以与这个过程对应.特征保留了必要的、显著的信息,既降低原始数据的冗余性,减少数据噪声,又比分类器决策结果有更充分的数据信息,数据量和数据维度适中,因此在这个层次上进行融合是目前最优的选择.引入特征融合方法的模式分类问题可以表述为,定义模式空间 $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ 和样本集合 X ,其中 ω_k 表示一个模式类, X 中每个元素 X_j 是一个样本的特征集合,其中 $j \in [1, P]$, P 为样本总数;特征集合 $X_j = \{X_j^1, X_j^2, \dots, X_j^D\}$ 表示样本有 D 种特征,样本特征向量 $\mathbf{X}_j^i = \{x_{j1}^i, x_{j2}^i, \dots, x_{j n_i}^i\} \in X_j$, $x_{j n_i}^i$ 表示样本第 j 种 i 特征的一个维度, n_i 是这种特征的维度, $N = \sum_{i=1}^D n_i$, 为样本特征总体维度;基于特征融合的

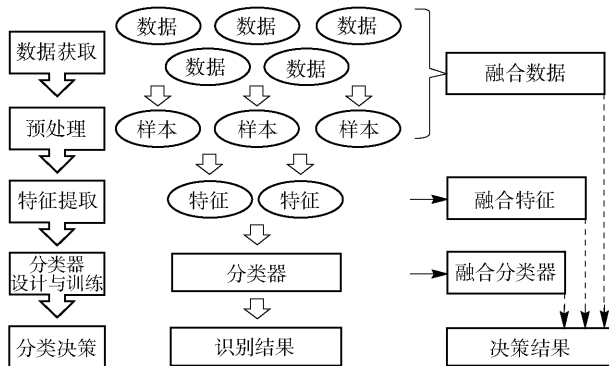


图 1 融合层次与模式分类

模式分类问题就是得到样本特征集合 X_j 到 ω_k 的对应关系,记为 $f: X_j \rightarrow \omega_k$.

2 特征融合的研究现状

特征融合早期应用在生物特征识别方面,瑞士 IDIAP 研究所和格里菲斯大学的研究人员 Sanderson 和 Pailwal^[1]在 2004 年发布的研究报告中总结了信息融合在个体生物认证方面的理论和技术,并且将信息融合划分为先映射融合(pre-mapping fusion)和后映射融合(post-mapping fusion),即根据信息融合发生在分类器(classifier)处理之前还是之后来划分.通常文献中提到的数据级融合(data level fusion)和特征级融合(feature level fusion)属于前者,分值级融合(score level fusion)和分类器级融合(classifier level fusion)属于后者,分类级融合也被认为是决策级融合(decision level fusion).虽然在学术研究中,不同级别的融合概念早就被提出来,但是受到计算机处理能力的限制,直接进行特征级甚至数据级的信息融合往往不容易.因此,早期的研究主要集中于分值级融合和分类器级融合. Kittler^[2]提出的基于贝叶斯决策理论的分类器融合框架,总结了乘法规则、加法规则、最大规则、最小规则、中位数规则、多数票规则等经典的融合规则,是分类器级融合良好的基础工作,也为特征融合奠定了基础. Kuncheva^[4]在这个基础上进行了理论研究,对基于贝叶斯的融合方法的分类错误率进行了建模分析,论证了融合方法的优势.而 Jain^[5]在 2005 发表的论文中归纳总结了多模生物特征识别系统中各种分值规则化的方法,成为分值级融合的经典文献,受到研究人员的广泛关注,之后发展出了基于变换、基于分类器、基于概率密度的分值级融合等各个分支.在此基础上, Nandakumar 等^[6]使用高斯混合模型改进了原有的单纯基于高斯分布的分值级融合方法,重新定义了似然比融合规则(likelihood ratio fusion rule),在人脸、指纹、虹膜、声音等生物特征识别中取得了良好的效果.随着学术研究的推进,特征融合的优势逐渐体现出来, Ma 等^[7]在线性融合模型的基础上分别提出了分类器级和特征级的线性依赖模型,通过增加依赖项来解决求解联合概率密度时独立同分布条件的限制问题,是对基于贝叶斯理论框架的融合方法的发展.与稀疏表示理论相结合, Shekhar 等^[8]提出基于联合稀疏表示的特征级融合方法,也应用于多模生物特征的识别中,通过优化一

个正则的最小二乘问题求解多模生物特征的联合稀疏表示矩阵. Yan 等^[9]引入了稀疏的思想并结合特征融合方法,在原有的降秩回归模型(RRR, reduced-rank regression)基础上进行改进,提出了稀疏核降秩回归模型(SKRRR, sparse kernel reduced-rank regression),将人说话时的语音特征和面部表情特征的进行融合,实现双模的表情识别. Bai 等^[10]使用特征融合的方法解决图像检索中的重排序(re-ranking)问题,提出一种新的稀疏情景激活的图像特征. Zhang L 等^[11]提出半监督的多特征学习方法,使用全局标签一致分类器实现多特征的共享学习.视频作为增加了时序信息的图像数据,特征融合方法也被应用到解决基于视频的目标识别和动作识别问题中,应对视频数据本身所具有的复杂性.视频数据中图像的空间结构信息也通常被视为提高视频识别性能的有效信息,而 SIFT、HOG、HOF 等特征提取方法在提取图像特征之后,会丢失空间结构信息, Yu 等^[12]提出了一种基于局部能量特征的二元表示方法,能够保持原始图像的结构信息,进而提高视频识别的准确性. Demirkus 等^[13]通过与多特征融合的方法相结合,提出了一种层级式的时空概率图模型来提取二元的人脸特征,解决视频中的人脸识别问题. Hu 等^[14]在处理携带有景深信息的视频数据中,提出了一种联合学习模型,实现多通道异构特征的融合,解决动作识别问题.与融合思想相结合的融合深度学习方法也被应用在视频目标识别的任务中, Feichtenhofer 等^[15]在文献[16]的基础上,改进了针对卷积神经网络的双流融合方法,总结了在深度卷积神经网络中的时间融合算法和空间融合算法,从而解决视频中动作识别的问题. Zhang X 等^[17]将特征融合思想引入深度学习模型中,提出了一种深度融合的学习架构,用于学习多媒体视频中高层语义特征.

以上介绍的特征融合算法的研究,分析其本质可以分为的 3 类:第 1 类是基于贝叶斯决策理论的算法,它的基础是统计推断和估计理论,这类算法目的在于求解联合概率密度;第 2 类是基于稀疏表示理论的算法,这类算法通常与最优化理论相结合,将特征融合问题转化为优化问题,进而求解联合稀疏表示来达到融合目的;第 3 类是基于深度学习理论的算法,这种算法的主要思想是在深度卷积神经网络中加入特征融合过程,以改进网络模型.

3 特征融合算法

在模式识别领域对特征融合方法的研究包括基于贝叶斯理论的特征融合、基于稀疏表示理论的特征融合算法等. 近年来,随着深度学习理论的发展,基于深度学习理论的特征融合算法也逐渐成为研究热点. 前两者是在提取样本多特征的基础上提出融合算法,重点在于对多特征进行综合处理得到融合特征;后者是将融合的思想引入深度学习理论中,在深度神经网络模型中加入融合过程,从而对网络模型进行优化.

3.1 基于贝叶斯理论的特征融合算法

基于贝叶斯理论的特征融合算法是在分类器融合算法基础上发展起来的. 贝叶斯决策理论是一种经典的模式识别方法. 已知模型空间 Ω 包含 C 种模式类,记为 $\Omega = (\omega_1, \dots, \omega_C)$,未知样本 x 由 N 维实数特征组成,记为 $x = [x_1, x_2, \dots, x_N]^T$. 根据最小错误率的贝叶斯决策理论,若将样本 x 分为第 j 类,则该类就是在已知样本 x 条件下后验概率最大的模式类,这个决策过程表示为

$$\begin{aligned} x &\rightarrow \omega_j, \\ \text{if } F(\omega_j) &= \max_{k=1, \dots, C} P(\omega_k | x) \end{aligned} \quad (1)$$

其中: $P(\omega_k | x)$ 表示第 k 类的后验概率, $k \in [1, C]$.

3.1.1 分类器融合的乘法规则与加法规则

在式(1)所示的决策过程中,如果将 x 看作分类器的输出结果,就可以得到基于贝叶斯理论的分类器融合算法. 假设有 M 个分类器,对于给定模式类每个分类器的输出结果都通过不同度量向量进行表示,用 x_i 表示第 i 个分类器输出结果的度量向量. 这里为了区分,用 Z 表示一个未知样本,那么将样本 Z 进行分类的决策过程可以表示为

$$\begin{aligned} Z &\rightarrow \omega_j, \\ \text{if } F(\omega_j) &= \max_{k=1, \dots, C} P(\omega_k | x_1, \dots, x_M) \end{aligned} \quad (2)$$

其中: $P(\omega_k | x_1, \dots, x_M)$ 表示在已知 M 个分类器输出结果条件下第 k 类的后验概率, $k \in [1, C]$. 在此基础上,引入分类器独立性假设并结合式(3),就可以将联合类条件表示为各个分类器的类条件概率相乘的形式,得到分类器融合的乘法规则,用式(4)表示,见文献[2]. 等式(3)右侧分子中 $p(x_1, \dots, x_M | \omega_k)$ 表示分类器的联合类条件概率.

$$P(\omega_k | x_1, \dots, x_M) = \frac{p(x_1, \dots, x_M | \omega_k) p(\omega_k)}{p(x_1, \dots, x_M)} \quad (3)$$

$$Z \rightarrow \omega_j,$$

$$\text{if } F(\omega_j) = \max_{k=1, \dots, C} P^{-(M-1)}(\omega_k) \prod_{i=1}^M P(\omega_k | x_i) \quad (4)$$

其中 $P(\omega_k)$ 表示第 k 类的先验概率. 分类器独立性假设各个分类器的输出结果所服从的概率分布是相互独立的. 为了解决第 k 类后验概率 $P(\omega_k | x_i)$ 为 0 时导致的分类错误问题,在乘法规则的基础上,再引入先验概率和后验概率近似相等假设,可以推导出分类器融合的加法规则,表示为

$$Z \rightarrow \omega_j,$$

$$\text{if } F(\omega_j) = \max_{k=1} \left[(1-M)P(\omega_k) + \sum_{i=1}^M P(\omega_k | x_i) \right] \quad (5)$$

其中:后验概率 $P(\omega_k | x_i)$ 采用相加的形式替代了式(4)中相乘的形式,可以解决后验概率为 0 时的分类错误问题. 先验概率和后验概率近似相等假设认为在各个分类器条件下的后验概率不会显著偏离该类的先验概率,表示为

$$P(\omega_k | x_i) = P(\omega_k) (1 + \delta_{ki})$$

其中 δ_{ki} 为一个极小值.

3.1.2 基于线性特征依赖模型的特征融合算法

由于分类器独立性假设与先验概率和后验概率近似相等假设都有相应的成立条件,在各种应用情景中不一定完全成立,尤其是分类器独立性假设在许多分类问题中都难以成立. 因此, Ma 等^[7] 在此基础上进一步研究了线性依赖模型,并且针对特征级融合提出了线性特征依赖模型 (LFDM, linear feature dependency modeling),实现了特征级融合. 该模型通过增加线性依赖项 $\alpha_i^k P(\omega_k) \delta_i^k$ 来对分类器间的依赖关系进行建模,其中 α_i^k 表示第 i 个分类器的依赖权重, δ_i^k 是一个极小值. 式(4)的后验概率可以表示为式(6)的形式.

$$P(\omega_k | x_1, \dots, x_M) = \frac{P_0}{P(\omega_k)^{M-1}} \prod_{i=1}^M (P(\omega_k | x_i) + \alpha_i^k P(\omega_k) \delta_i^k), i \in [1, M] \quad (6)$$

$$P_0 = \frac{\prod_{i=1}^M p(x_i)}{p(x_1, \dots, x_M)} \quad (7)$$

这里需要将 $x_i, i \in [1, M]$ 仍看作样本的 N_i 维特征向量,表示为 $x_i = [x_{i1}, x_{i2}, \dots, x_{iN_i}]^T$. 若进一步将 x_i 向量的每个维度都看作一个分类器,则式(6)的后验概率可以表示为

本,那么测试样本总数为

$$d = \sum_{i=1}^D d_i$$

矩阵中的每一列是一个测试样本,如图3所示.

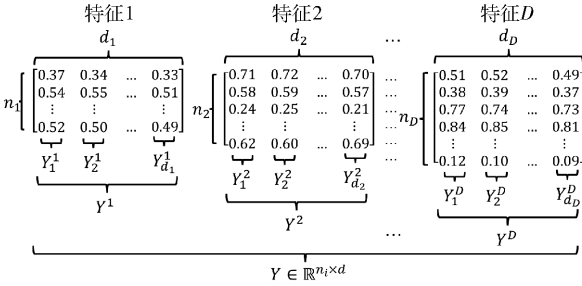


图3 测试样本

根据稀疏表示理论,可以将测试样本矩阵 \mathbf{Y}^i 表示成字典 \mathbf{X}^i 与稀疏矩阵 \mathbf{F}^i 相乘的形式,记为

$$\mathbf{Y}^i = \mathbf{X}^i \mathbf{F}^i + \mathbf{N}^i, i = 1, \dots, D \quad (12)$$

其中 \mathbf{N}^i 是一个稀疏噪声矩阵.

因此,对于每种特征都有一个稀疏矩阵 \mathbf{F}^i 与之对应,各个稀疏矩阵组合在一起可以得到联合稀疏表示矩阵 $\mathbf{F} = [\mathbf{F}^1, \mathbf{F}^2, \dots, \mathbf{F}^D] \in \mathbb{R}^{p \times d}$. 矩阵 \mathbf{F} 可通过求解式(13)的最小二乘优化问题得到. 对 \mathbf{F} 施加 l_1/l_q 范数的正则化, λ 是预设参数,如式(14)所示,其中 γ^k 是矩阵 \mathbf{F} 中的第 k 行. 该优化问题可以通过 ADMM (alternating direction method of multipliers) 算法求解^[8].

$$\hat{\mathbf{F}} = \arg \min_{\mathbf{F}} \frac{1}{2} \sum_{i=1}^D \|\mathbf{Y}^i - \mathbf{X}^i \mathbf{F}^i\|_F^2 + \lambda \|\mathbf{F}\|_{1,q} \quad (13)$$

$$\|\mathbf{F}\|_{1,q} = \sum_{k=1}^p \|\gamma^k\|_q \quad (14)$$

得到矩阵 \mathbf{F} 之后,通过式(15)的分类判别函数可以对未知样本 \mathbf{Y} 进行分类,输入 D 种特征,输出该样本的模式类标签.

$$\text{identity}(\mathbf{Y}) = \arg \min_j \sum_{i=1}^D \|\mathbf{Y}^i - \mathbf{X}^i \delta_j^i(\mathbf{F}^i)\|_F^2 \quad (15)$$

其中 δ_j^i 是一个矩阵指示器函数,它对稀疏矩阵 \mathbf{F}^i 中的值起到了选择的作用,即保持矩阵 \mathbf{F}^i 与第 j 类相关行的值,而将其他行全部设置为 0.

3.2.2 鲁棒的基于联合稀疏表示的特征融合算法

对于图像中较严重的噪声,例如遮挡问题,可通过在式(12)中增加噪声项来增强模型的鲁棒性,如式(16)所示. 在遮挡问题中, $\mathbf{B}^i \mathbf{A}^i$ 是背景噪声项,

\mathbf{B}^i 是基础遮挡样本, \mathbf{A}^i 是遮挡噪声稀疏矩阵. 在得到 \mathbf{A}^i 后,可以通过 $\mathbf{Y}^i - \mathbf{B}^i \mathbf{A}^i$ 来去除遮挡噪声. 对于每种特征都有一个遮挡噪声稀疏矩阵与之对应,它们的联合稀疏表示矩阵记为 $\mathbf{A} = [\mathbf{A}^1, \mathbf{A}^2, \dots, \mathbf{A}^D]$.

$$\mathbf{Y}^i = \mathbf{X}^i \mathbf{F}^i + \mathbf{B}^i \mathbf{A}^i + \mathbf{N}^i, i = 1, 2, \dots, D \quad (16)$$

增加了背景噪声项后,需要同时求解矩阵 \mathbf{F} 和矩阵 \mathbf{A} ,该问题可以表示为式(17),其中对施加 l_1 范数的正则化, λ_1 和 λ_2 是预设参数. 在得到两个联合稀疏表示矩阵 \mathbf{F} 和 \mathbf{A} 后,通过式(18)的分类判别函数可以对未知样本 \mathbf{Y} 进行分类.

$$\hat{\mathbf{F}}, \hat{\mathbf{A}} = \arg \min_{\mathbf{F}, \mathbf{A}} \frac{1}{2} \sum_{i=1}^D \|\mathbf{Y}^i - \mathbf{X}^i \mathbf{F}^i - \mathbf{B}^i \mathbf{A}^i\|_F^2 + \lambda_1 \|\mathbf{F}\|_{1,q} + \lambda_2 \|\mathbf{A}\|_1 \quad (17)$$

$$\text{identity}(\mathbf{Y}) = \arg \min_j \sum_{i=1}^D \|\mathbf{Y}^i - \mathbf{X}^i \delta_j^i(\mathbf{F}^i) - \mathbf{B}^i \mathbf{A}^i\|_F^2 \quad (18)$$

3.3 基于深度学习理论的特征融合算法

深度学习理论是在人工神经网络的基础上发展起来的机器学习理论,在多层神经网络中加入了更多隐层单元,得到了深度神经网络模型. 其中深度卷积神经网络模型是该理论中的重要模型之一. 使用该模型可以进行有监督地学习,将特征提取过程与分类器训练过程整合在一起,实现端到端的机器学习. 基于深度学习理论的特征融合算法是将特征融合的思想引入深度神经网络模型,使用多特征输入到模型中进行训练,在模型中选择 2 个隐层进行特征融合.

Simonyan 等^[16]首先提出了一种使用双流架构的深度卷积神经网络模型,可解决视频中的动作识别问题. 该模型分别建立一个空间流卷积神经网络和一个时间流卷积神经网络进行独立训练,在最终的 Softmax 分类输出层将这 2 个网络进行融合,属于分类器级的融合. 而 Feichtenhofer 等^[15]则在这个基础上改进了网络融合方法,提出了空间特征融合方法和时间特征融合算法,不仅可以在 Softmax 层进行融合,还可以在卷积层之后的 ReLU 层进行融合,实现特征级的融合.

3.3.1 空间特征融合算法

空间特征融合算法可以对卷积层输出的 2 个特征图 (feature map) 进行融合,得到融合后的特征图,从而将 2 个深卷积神经网络模型连接在一起,这个连接点就是融合点. 因此,引入特征融合方法之后,

2个深卷积神经网络模型在融合点之前分别进行特征学习,并在融合点将独立学习的特征进行融合,最后开始共同学习. 融合函数定义为

$$f: \mathbf{x}_t^a + \mathbf{x}_t^b \rightarrow \mathbf{y}_t \quad (19)$$

其中: \mathbf{x}_t^a 和 \mathbf{x}_t^b 表示 t 时刻的视频帧分别经过卷积运算得到的空间特征图, \mathbf{y}_t 表示融合空间特征图, $\mathbf{x}_a^b, \mathbf{x}_b^a \in \mathbb{R}^{HWD}$ 而 H, W 和 D 分别表示特征图的长度、宽度和通道数量. 融合函数包含加性融合函数、最大融合函数、级联融合函数、卷积融合函数和双线性融合函数等.

加性融合函数 $\mathbf{y}^{\text{sum}} = \mathbf{f}^{\text{sum}}(\mathbf{x}^a, \mathbf{x}^b)$, 是对2个特征图对应位置元素的值进行相加, 如式(20)所示, 融合特征图的通道数不变, 其中 $i \in [1, H], j \in [1, W], d \in [1, D]$.

$$\mathbf{y}_{i,j,d}^{\text{sum}} = \mathbf{x}_{i,j,d}^a + \mathbf{x}_{i,j,d}^b \quad (20)$$

最大融合函数 $\mathbf{y}^{\text{max}} = \mathbf{f}^{\text{max}}(\mathbf{x}^a, \mathbf{x}^b)$ 与加性融合函数相似, 是将2个特征图对应的位置元素值较大的一个作为融合结果, 即

$$\mathbf{y}_{i,j,d}^{\text{max}} = \max\{\mathbf{x}_{i,j,d}^a, \mathbf{x}_{i,j,d}^b\} \quad (21)$$

级联融合函数 $\mathbf{y}^{\text{cat}} = \mathbf{f}^{\text{cat}}(\mathbf{x}^a, \mathbf{x}^b)$ 与前两者不同, 它保留了2个特征图的结果, 并将融合后特征图的通道数变为原始特征图的两倍, 如

$$\mathbf{y}_{i,j,2d}^{\text{cat}} = \mathbf{x}_{i,j,d}^a, \mathbf{y}_{i,j,2d-1}^{\text{cat}} = \mathbf{x}_{i,j,d}^b \quad (22)$$

其中 $\mathbf{y} \in \mathbb{R}^{H \times W \times 2D}$.

卷积融合函数 $\mathbf{y}^{\text{conv}} = \mathbf{f}^{\text{conv}}(\mathbf{x}^a, \mathbf{x}^b)$, 是将级联融合函数的融合结果与滤波器 f 进行卷积运算, 并且引入偏差值 b , 从而实现融合特征图的降维处理. 表示为

$$\mathbf{y}^{\text{conv}} = \mathbf{y}^{\text{cat}} \mathbf{f} + \mathbf{b} \quad (23)$$

其中 $\mathbf{f} \in \mathbb{R}^{1 \times 1 \times 2D \times D}, \mathbf{b} \in \mathbb{R}^D$.

双线性融合函数 $\mathbf{y}^{\text{bil}} = \mathbf{f}^{\text{bil}}(\mathbf{x}^a, \mathbf{x}^b)$, 是对2个特征图对应的位置元素进行外积运算后求和, 融合特征图的通道数是原始特征图通道数的平方, 表示为

$$\mathbf{y}^{\text{bil}} = \sum_{i=1}^H \sum_{j=1}^W \mathbf{x}_{i,j}^{aT} \otimes \mathbf{x}_{i,j}^b \quad (24)$$

其中 $\mathbf{y}^{\text{bil}} \in \mathbb{R}^{D^2}$. 这种融合函数常被用在 ReLU 层, 能够对2个特征图对应通道进行融合.

3.3.2 时间特征融合算法

时间特征融合算法主要作用在深度卷积神经网络的池化层. 在进行池化处理之前, 需要将空间特征图按照时序进行堆叠, 使得输入池化层的特征图维度提高, 表示为 $\mathbf{x}_{1,\dots,T} \in \mathbb{R}^{HWT D}$. 文献[15]提出了3D池化方法和3D卷积+3D池化方法, 直接保留了

时序信息.

3D池化方法是对深度卷积神经网络中通常使用的2D池化方法的直接扩展, 使用HWT的3D池化立方代替原始池化层的计算方法, 对输入的特征图 x 进行池化运算, 从而将池化处理从空域扩展到时域. 3D池化立方可以使用最大池化方法, 保留立方区域中的最大值作为池化结果. 这种方法是对特征图的单通道进行处理, 不进行跨通道的池化运算.

3D卷积+3D池化方法则是在3D池化方法基础上, 再加入一个滤波器 $f \in \mathbb{R}^{H'W'T'D'}$ 和偏差值 b , 对输入的特征图 $\mathbf{x}_{1,\dots,T}$ 进行卷积运算, 如

$$\mathbf{y}_{1,\dots,T} = \mathbf{x}_{1,\dots,T} * \mathbf{f} + \mathbf{b} \quad (25)$$

计算得到的 $\mathbf{y}_{1,\dots,T}$ 再进行3D池化处理. 使用滤波器 f 可以在一个局部的时空邻域内对图像特征的组合, 通过使用 $H'W'T'D'$ 的卷积核进行卷积运算, 增加模型的连接权重数量.

在处理视频数据时, 时间特征融合算法通常与空间特征融合算法同时使用, 通过利用视频中的时序信息, 提高基于视频的动作识别或者目标识别的准确率.

4 结束语

随着人工智能研究的深入, 特别是在解决图像识别问题中, 特征融合方法受到广泛关注, 无论是对特征融合方法本身的研究, 还是结合特征融合方法改进现有方法, 都成为目前研究的热点. 笔者总结了近年来模式识别领域中特征融合方法的研究进展, 基于信息融合理论介绍了特征融合方法的基本原理和发展现状, 介绍和分析了基于叶贝斯理论的特征融合算法、基于稀疏表示理论的特征融合算法和基于深度学习理论的特征融合算法.

特征融合方法可以综合利用多种图像特征实现特征互补, 在解决图像识别问题中, 相对于单一特征而言, 基于特征融合的图像识别方法具有更好的鲁棒性. 笔者介绍的3种特征融合算法分别建立在不同的理论基础上, 体现了特征融合思想的普适性. 目前, 对特征融合方法的研究主要集中在融合算法、融合模型等方面, 在融合特征选择、基础理论等方面也会有进一步的研究.

参考文献:

- [1] Sanderson C, Paliwal K K. Identity verification using speech and face information[J]. Digital Signal Process-

- ing, 2004, 14(5): 449-480.
- [2] Kittler J, Hatef M, Duin R P W, et al. On combining classifiers[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(3): 226-239.
 - [3] 韩崇昭, 朱洪艳, 段战胜等. 多源信息融合[M]. 北京: 清华大学出版社, 2010.
 - [4] Kuncheva L I. A theoretical study on six classifier fusion strategies[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(2): 281-286.
 - [5] Jain A, Nandakumar K, Ross A. Score normalization in multimodal biometric systems [J]. Pattern Recognition, 2005, 38(12): 2270-2285.
 - [6] Nandakumar K, Chen Y, Dass S C, et al. Likelihood ratio-based biometric score fusion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 30(2): 342.
 - [7] Ma A J, Yuen P C, Lai J H. Linear dependency modeling for classifier fusion and feature combination [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(5): 1135-1148.
 - [8] Shekhar S, Patel V M, Nasrabadi N M, et al. Joint sparse representation for robust multimodal biometrics recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(1): 113-126.
 - [9] Yan J, Zheng W, Xu Q, et al. Sparse kernel reduced-rank regression for bimodal emotion recognition from facial expression and speech[J]. IEEE Transactions on Multimedia, 2016, 18(7): 1319-1329.
 - [10] Bai S, Bai X. Sparse contextual activation for efficient visual re-ranking[J]. IEEE Transactions on Image Processing, 2016, 25(3): 1056-1069.
 - [11] Zhang L, Zhang D. Visual understanding via multi-feature shared learning with global consistency[J]. IEEE Transactions on Multimedia, 2016, 18(2): 247-259.
 - [12] Yu M, Liu L, Shao L. Structure-preserving binary representations for RGB-D action recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(8): 1651-1664.
 - [13] Demirkus M, Precup D, Clark J, et al. Hierarchical spatio-temporal probabilistic graphical model with multiple feature fusion for binary facial attribute classification in real-world face videos[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(6): 1185-1203.
 - [14] Hu J F, Zheng W S, Lai J, et al. Jointly learning heterogeneous features for RGB-D activity recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016 (99): 1-1.
 - [15] Feichtenhofer C, Pinz A, Zisserman A. Convolutional two-stream network fusion for video action recognition [C]// CVPR 2016. Las Vegas: IEEE Computer Society, 2016: 1933-1941.
 - [16] Simonyan K, Zisserman A. Two-stream convolutional networks for action recognition in videos[J]. Advances in Neural Information Processing Systems, 2014, 1(4): 568-576.
 - [17] Zhang X, Zhang H, Zhang Y, et al. Deep fusion of multiple semantic cues for complex event recognition [J]. IEEE Transactions on Image Processing, 2016, 25(3): 1033-1046.