

基于词向量融合的遥感场景零样本分类算法

吴 晨^{1,3} 袁昱纬² 王宏伟³ 刘 宇³ 刘思彤⁴ 全吉成³

(中国人民解放军海军航空大学 山东烟台 264001)¹ (91977 部队 北京 102200)²

(中国人民解放军空军航空大学 长春 130022)³ (西安飞行学院 西安 710000)⁴

摘 要 零样本分类算法无须标注要识别的类别样本,因而能大幅度降低实际应用成本,近年来受到广泛关注。遥感场景类别的语义词向量与图像特征空间原型的结构不一致问题,严重影响了遥感场景零样本的分类效果。利用不同词向量间的互补性,文中提出一种基于语义词向量融合的遥感场景零样本分类算法,即耦合式解析字典学习(Coupled Analysis Dictionary Learning, CADL)方法。首先,采用稀疏编码效率较高的解析字典学习方法获取各语义词向量的稀疏系数,以减少冗余信息;然后,将对应的稀疏编码系数串接后作为融合语义词向量表示,并将融合语义词向量线性映射到图像特征空间,与图像特征空间场景类别原型表示进行结构对齐,以降低结构差异性;最后,计算得到要识别的场景类别的图像特征原型,并采用最近邻分类器在图像特征空间完成分类。在 UCM 和 AID 数据集上对多种语义词向量的融合进行定量实验,同时将 RSSCN7 数据集作为已知场景类别的数据集来对两幅实际遥感图像进行定性实验。在 UCM 和 AID 上的定量实验分别获得了最高总体分类准确度 48.40% 和 60.23%,相比于典型零样本分类方法的总体分类准确度分别提升了 4.80% 和 6.98%。对两幅实际遥感图像的定性实验,同样获得了最佳零样本的分类效果。实验结果表明,多种语义词向量融合,可以获得与图像特征空间原型结构更一致的语义词向量,且显著提升了遥感场景零样本分类的准确度。

关键词 遥感场景分类,零样本分类,结构对齐,词向量融合,解析字典学习

中图分类号 TP181 文献标识码 A DOI 10.11896/jsjcx.181202257

Word Vectors Fusion Based Remote Sensing Scenes Zero-shot Classification Algorithm

WU Chen^{1,3} YUAN Yu-wei² WANG Hong-wei³ LIU Yu³ LIU Si-tong⁴ QUAN Ji-cheng³

(Naval Aeronautical University, Yantai, Shandong 264001, China)¹ (The 91977 of PLA, Beijing 102200, China)²

(Aviation University Air Force, Changchun 130022, China)³ (Xi'an Flight Academy of PLA, Xi'an 710000, China)⁴

Abstract Zero-shot classification algorithm does not need to label the sample of unseen classes to be recognized, so it can greatly reduce the cost of practical application, which has attracted wide attention in recent years. The problem of structure difference between word vectors and image feature prototypes seriously affects the zero-shot classification performance of remote sensing scenes. Based on the complementarity among different kinds of word vectors, the remote sensing scenes zero-shot classification algorithm based on word vectors fusion, named coupled analysis dictionary learning method, was proposed. Firstly, the sparse coefficients of different kinds of word vectors are obtained by the more efficient analysis dictionary learning to reduce the redundant information. Then, the sparse coefficients are concatenated and denoted as the fused word vectors, and a structure alignment operation is performed based on the image feature prototypes to reduce structural differences by embedding the fused word vectors into image feature space. Finally, the image feature prototypes of the scene classes unseen are calculated, and the nearest neighbor classifier is employed to complete the classification in the image feature space. Quantitative experiments of the fusion of multiple semantic word vectors were carried out on UCM and AID datasets. At the same time, two real remote sensing images were qualitatively tested with RSSCN7 datasets as the seen dataset. Quantitative experiments obtain the highest overall classification accuracies of 48.40% and 60.23% on UCM and AID, which respectively exceeds the typical comparative methods by 4.80% and 6.98%. In qualitative experiments on two real remote sensing images, the algorithm also obtains the best zero-shot classification performance. The experimental results show that the fused word vectors are more consistent with the prototypes in image feature space, and the zero-shot classification accuracies of remote sensing scenes can be significantly improved.

到稿日期:2018-12-04 返修日期:2019-04-04 本文受国家青年自然科学基金(61301233)资助。

吴 晨(1991—),男,博士生,主要研究方向为遥感图像场景分类, E-mail: wuchen_research@aliyun.com; 袁昱纬(1988—),男,博士,工程师,主要研究方向为遥感图像智能识别; 王宏伟(1984—),男,博士生,讲师,主要研究方向为遥感图像智能识别; 刘 宇(1968—),男,博士,副教授,主要研究方向为遥感图像解译; 刘思彤(1990—),女,硕士,讲师,主要研究方向为遥感图像处理; 全吉成(1960—),男,博士,教授,主要研究方向为三维虚拟现实, E-mail: jicheng_quan@126.com(通信作者)。

Keywords Remote sensing scenes classification, Zero-shot classification, Structure alignment, Word vectors fusion, Analysis dictionary learning

1 引言

遥感技术的进步和广泛应用,催生了基于场景的遥感图像分类方法^[1]。作为具有某种概念语义的遥感图像块,场景已经成为海量遥感图像分类的基本单元,使针对大规模遥感图像的快速分析成为可能。然而,现有遥感图像场景分类主要采用监督学习方法,需要标注大量样本,严重阻碍了遥感场景分类算法的实际应用^[2]。

作为一种特殊的无监督学习方法,零样本分类(Zero-Shot Classification, ZSC)算法使用类别名称的语义词向量,通过迁移已知(seen)类别标注样本中的知识,来推断新的未见过(unseen)类别样本的分布中心(即类别的图像特征空间中的原型表示),从而实现新类别样本的分类。由于不需要标注 unseen 类样本, ZSC 近年来受到了广泛关注^[3-9]。Xian 等^[3]提出的 LatEm 算法在兼容函数学习过程中引入隐式变量模型,进行细粒度 ZSC。针对映射函数泛化能力不足的问题, Wang 等^[4]提出了关系知识迁移(RKT)算法。Zhang 等^[5]提出联合隐式相似性嵌入(JLSE)方法。Zhang 等^[6]提出的语义相似性嵌入(SSE)方法将源域或目标域数据视为由训练类组合而成。Wang 等^[7]提出的双向隐式嵌入(BiDiLEL)方法,将图像特征和语义特征分别映射到公共空间。Li 等^[8]提出的双视觉语义映射(DMaP)方法,利用了语义空间流形和视觉语义映射迁移能力之间的关系。Zhao 等^[9]提出利用直推式框架(MDP)来估计 unseen 样本特征分布的特点。

语义词向量^[10-11]是采用自然语言训练模型,在大规模文本语料集上通过无监督学习得到实体单词的高维向量表示,每个类别名称均有唯一对应的语义词向量,从而可提供不同类别之间的距离关系。在 ZSC 中,采用类别名称的语义词向量提供类别间的距离关系,以辅助推断图像特征空间 unseen 类别的原型表示。因此,语义词向量能否反映图像特征空间的类间距离关系,是 ZSC 方法的关键。现有 ZSC 方法均是针对单领域细粒度类别的分类,然而,遥感场景类别涉及不同领域,这对词向量提出了更高的要求,即需要真实反映场景类别间的距离关系。单词词向量受训练语料、训练模型的限制,难以适应多领域的遥感场景类别的情形。

近几年,随着自然语言处理技术的进步,已能便捷地获取不同训练模型(如 Word2Vector^[12], Glove^[13]等)和不同训练语料(如 Wikipedia, Common Crawl 等)的语义词向量。这些语义词向量具有一定的互补性,通过融合可获得与图像特征空间中场景类别距离结构更一致的语义词向量,从而提升遥感场景零样本分类的准确度。因此,本文分别对来自不同训练模型、不同训练语料的语义词向量进行融合,以利用不同词向量之间的互补性,从而提升遥感场景 ZSC 的性能。

语义词向量中存在一定的冗余信息,影响了场景类间距离结构信息的有效表达,因此需要对语义词向量进行稀疏编码,以减少冗余信息。鉴于解析字典学习方法具有优越的稀疏编码能力,本文采用解析字典学习方法,提出一种基于词向量融合的遥感场景 ZSC 方法,即耦合式解析字典学习方法。

为了保留不同词向量各自的特点,融合不同词向量的互补性, CADL 将稀疏系数串接,从而形成新的语义词向量。为了降低不同空间中距离结构的差异性,将新词向量线性变换到图像特征空间,并与其中的场景类别原型对齐,从而降低两种空间距离结构的差异性; unseen 类的图像特征空间原型,可由 unseen 类融合词向量从融合词向量空间传递到图像特征空间而得到,即具有耦合性。借助耦合性获得遥感场景 unseen 类的图像特征空间原型表示后,采用最近邻分类器完成 unseen 类样本的分类,因此本文方法称为耦合式解析字典学习方法。

2 解析字典学习

字典学习分为两类,即合成性字典学习(Synthesis Dictionary Learning, SDL)和解析性字典学习(Analysis Dictionary Learning, ADL)。SDL 认为输入特征可以由字典和相应的稀疏系数重建得到;而 ADL 将字典视为滤波器,并将字典应用到输入特征上,从而获得特征的稀疏系数。ADL 的理念类似于傅里叶变换和小波变换,但是 ADL 的字典是在数据上通过学习得到的,因而能较好地适应数据的特点。

虽然 SDL 方法应用广泛,但其计算效率不高。而 ADL 通常具有闭式解,加上良好的编码能力,因此计算效率较高^[14-15]。ADL 的基本公式为:

$$\arg \min_{\Omega, Z} \frac{1}{2} \| Z - \Omega X \|_F^2 \quad (1)$$

$$\text{s. t. } \Omega \in \Gamma, \| z_i \|_0 \leq T_0$$

其中, $X = [x_1, \dots, x_n] \in \mathbb{R}^{m \times n}$ 为 n 个输入样本组成的特征矩阵, $x_i \in \mathbb{R}^m$ 为第 i 个样本; Z 为 X 的稀疏系数,其样本稀疏性采用 l_0 范数及参数 T_0 来实现; Ω 为解析字典; Γ 是为了避免出现平凡解而对 Ω 施加的 $\log\text{-det}$ 限制条件^[16]。

3 耦合式解析字典学习

CADL 采用 ADL 方法获得各词向量的稀疏系数,串接稀疏系数后得到融合词向量表示,并将其与图像特征空间类别原型结构对齐。由于 seen 类和 unseen 类之间没有重叠,因此下文将分别阐述。

3.1 seen 类词向量融合

seen 类词向量融合的目标函数如式(2)所示,其中第一项为 ADL 项,旨在提取词向量的稀疏系数;第二项为结构对齐项,旨在将 seen 类新语义词向量与图像特征空间 seen 类原型结构对齐。图像特征空间类别的原型是指每个类别样本的均值中心,其与语义词向量表示的是相同的场景类别,因此两者之间是一一对应关系。但由于词向量空间和图像特征空间的维数不同,因此需要进行跨空间的对应,即类别原型结构对齐。本文采用结构对齐项,将融合词向量线性变换到图像特征空间,并最小化对齐误差。该方法可有效建模词向量空间与图像特征空间之间的映射关系,实现两种空间之间的结构对齐,缓解遥感场景类别的语义词向量与图像特征空间原型的结构不一致问题。

$$\xi_s = \min_{\Omega_i, Z_i^s, W} \frac{1}{2} \sum_{i=1}^M \|Z_i^s - \Omega_i C_i^s\|_F^2 + \|WZ_c^s - P^s\|_F^2 \quad (2)$$

其中, $C_i^s \in \mathbb{R}^{d_i \times n_s}$ 为 seen 类的第 i 种词向量(共 M 种不同词向量)矩阵, d_i 为对应词向量的维度, n_s 为 seen 类数; $\Omega_i \in \mathbb{R}^{d_i \times d_i}$ 为第 i 种词向量的解析字典; $P^s \in \mathbb{R}^{q \times n_s}$ 为 seen 类场景在图像特征空间中的原型, q 为图像特征维数; $W \in \mathbb{R}^{q \times D}$ 为线性变换矩阵, 其中 $D = \sum_{i=1}^M d_i$ 为串接后新词向量的维数; $Z_i^s \in \mathbb{R}^{d_i \times n_s}$ 为 seen 类的第 i 种词向量的稀疏系数。

式(3)给出串接后的 seen 类新词向量表示 Z_c^s :

$$Z_c^s = \begin{bmatrix} Z_1^s \\ \dots \\ Z_M^s \end{bmatrix} \in \mathbb{R}^{D \times n_s} \quad (3)$$

3.2 unseen 类词向量融合

unseen 类词向量融合同样包含 ADL 项和结构对齐项, 目标函数 ξ_u 为:

$$\xi_u = \min_{P^u, \Omega_i, Z_i^u, W} \frac{1}{2} \sum_{i=1}^M \|Z_i^u - \Omega_i C_i^u\|_F^2 + \|WZ_c^u - P^u\|_F^2 \quad (4)$$

其中, $C_i^u \in \mathbb{R}^{d_i \times n_u}$ 为 unseen 类的第 i 种词向量; $P^u \in \mathbb{R}^{q \times n_u}$ 为遥感场景 unseen 类在图像特征空间的原型表示; $Z_i^u \in \mathbb{R}^{d_i \times n_u}$ 为 C_i^u 的稀疏系数; Z_c^u 为串接后的 unseen 类新词向量。 P^u 未知, 需要在融合及结构对齐过程中求解。获得 P^u 后, 即可采用最近邻分类器对 unseen 样本进行分类。

3.3 词向量融合的目标函数

CADL 方法的总体目标函数为:

$$\xi = \xi_s + \xi_u + R(W) \quad (5)$$

其中, 为了增强映射矩阵 W 求解过程的鲁棒性, 加入正则项 $R(W)$, 即: $R(W) = \|W\|_F^2$ 。由于式(5)非凸, 因此不能直接对其进行优化求解, 但可以采用逐个优化的方式进行求解。初始化 Z_i^s 和 Z_i^u 为 one-hot 向量矩阵, 则式(5)的求解过程如下。

1) 固定 Z_i^u, W, P^u, Z_i^s , 更新 Ω_i , 此时的总体目标函数为:

$$\begin{aligned} \min_{\Omega_i, Z_i^s, Z_i^u} & \sum_{i=1}^M \|Z_i^s - \Omega_i C_i^s\|_F^2 + \sum_{i=1}^M \|Z_i^u - \Omega_i C_i^u\|_F^2 \\ & = \min_{\Omega_i, Z_i^s, Z_i^u} \sum_{i=1}^M \|[Z_i^s, Z_i^u] - \Omega_i [C_i^s, C_i^u]\|_F^2 \\ & = \min_{\Omega_i, Z_i^s, Z_i^u} \sum_{i=1}^M \|Z_i - \Omega_i C_i\|_F^2 \end{aligned} \quad (6)$$

其中, $Z_i = [Z_i^s, Z_i^u]$, $C_i = [C_i^s, C_i^u]$ 。为避免出现平凡解, 引入对字典 Ω_i 的 log-det 条件 $R(\Omega_i) = \|\Omega_i\|_F^2 - \log|\det \Omega_i|$ 。因此, 更新 Ω_i 的目标函数为:

$$\min_{\Omega_i, Z_i^s, Z_i^u} \sum_{i=1}^M \|Z_i - \Omega_i C_i\|_F^2 + R(\Omega_i) \quad (7)$$

式(7)的求解步骤^[16]为: 首先, 将 $C_i C_i^T + I$ 矩阵分解为 LL^T ; 其次, 对 $L^{-1} C_i Z_i^T$ 进行 SVD 分解, 得到 $Q\Sigma R^T$; 最后, 得到字典 $\Omega_i = 0.5R(\Sigma + (\Sigma^2 + 2I)^{1/2})Q^T L^{-1}$ 。

2) 固定 $Z_i^u, \Omega_i, P^u, Z_i^s$, 更新 W , 此时的总体目标函数为:

$$\min_W \|WZ_c^s - P^s\|_F^2 + \|W\|_F^2 \quad (8)$$

通过对 W 求导, 得到更新 W 的解析式为:

$$W = P^s (Z_c^s)^T (Z_c^s (Z_c^s)^T + I)^{-1} \quad (9)$$

3) 固定 $Z_i^u, \Omega_i, W, Z_i^s$, 更新 Z^u , 此时的总体目标函数为

$\min_{P^u} \|WZ_c^u - P^u\|_F^2$, 因此更新 unseen 类在图像特征空间的原

型表示 P^u 为 WZ_c^u 。

4) 固定 Ω_i, W, P^u, Z_i^s , 更新 Z_c^s 和 Z_c^u , 此时的总体目标函数为:

$$\begin{aligned} \min_{Z_c^s, Z_c^u} & \|WZ_c^s - P^s\|_F^2 + \|WZ_c^u - P^u\|_F^2 \\ & = \min_{Z_c^s, Z_c^u} \|W[Z_c^s, Z_c^u] - [P^s, P^u]\|_F^2 \end{aligned} \quad (10)$$

通过分别对 Z_c^s 和 Z_c^u 求导, 得到: $Z_c^s = (W^T W)^{-1} W^T P^s$, $Z_c^u = (W^T W)^{-1} W^T P^u$ 。

5) 稀疏化 Z_c^s 和 Z_c^u , 更新 Z_i^s 和 Z_i^u 。从 Z_c^s 和 Z_c^u 中提取 Z_i^s 和 Z_i^u , 并基于门限参数 T_0 , 采用式(11)稀疏化 Z_i^s 和 Z_i^u 。

$$(Z_i^s)_{jk} = \begin{cases} 0, & (Z_i^s)_{jk} < T_0 \\ (Z_i^s)_{jk}, & \text{otherwise} \end{cases} \quad (11)$$

3.4 CADL 算法的步骤

基于词向量融合的遥感场景零样本分类(CADL)算法的步骤如算法1所示。

算法1 CADL

输入: seen 类样本均值(类原型) P^s , M 种不同的词向量(其中 seen 类词向量为 C_i^s , unseen 类词向量为 C_i^u , $i=1, 2, \dots, M$), unseen 类样本 $X^u \in \mathbb{R}^{q \times N_u}$, 最大迭代次数 Iter_N

输出: unseen 类样本均值(类原型) P^u , 对 X^u 中样本赋予类别标签

- 步骤1 初始化 Z_i^s, Z_i^u 为 one-hot 向量矩阵;
- 步骤2 根据式(7)更新 Ω_i ;
- 步骤3 根据式(9)更新 W ;
- 步骤4 更新 P^u 为 WZ_c^u ;
- 步骤5 根据式(10)和式(11)更新 Z_c^s, Z_c^u 和 Z_i^s, Z_i^u ;
- 步骤6 判断是否达到最大迭代次数 Iter_N , 若是, 则执行步骤7, 否则循环执行步骤2-步骤6;
- 步骤7 利用优化得到的 unseen 类原型 P^u , 采用最近邻分类器, 确定 unseen 样本 X^u 的类别标签。

4 实验及结果分析

4.1 数据集及实验设置

实验采用3种遥感场景数据集: UC-Merced(UCM)数据集^[17]、航空图像数据集(Aerial Image Dataset, AID)^[18]及RSSCN7数据集^[19]。其中, UCM数据集和AID用于定量实验, RSSCN7用于定性实验, 即在定性实验中RSSCN7场景图像作为seen类样本, 以测试实际遥感图像上unseen类场景的ZSC效果。UCM数据集有21类场景, 共2100张图像, 图像大小为 256×256 像素, 部分样本如图1所示; AID共有30类场景, 共10000张场景图像, 图像大小为 600×600 像素, 部分样本如图2所示。RSSCN7共2800张遥感场景图像, 分为7个类别, 图像大小为 400×400 像素, 部分样本如图3所示。

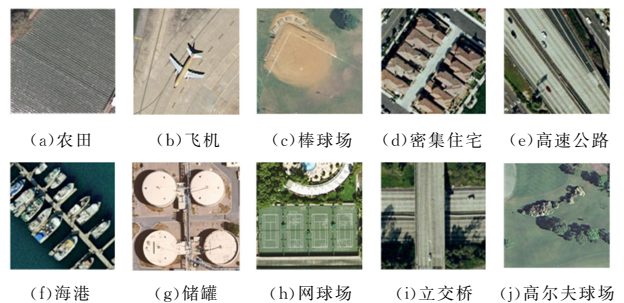


图1 UCM 数据集集中的部分样本

Fig. 1 Partial of samples in UCM dataset

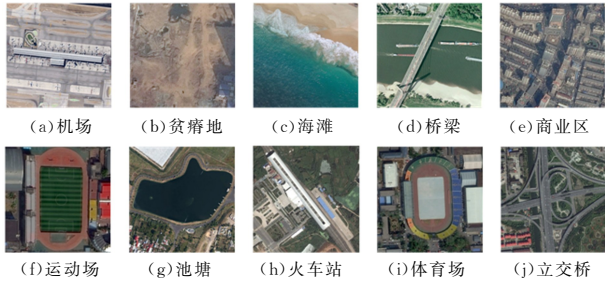


图 2 AID 若干类的样本
Fig. 2 Samples of several AID classes

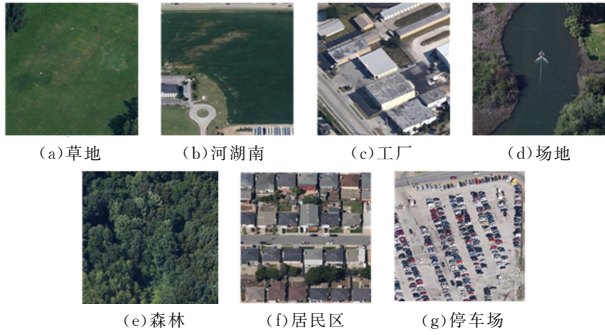


图 3 RSSCN7 数据集类的样本
Fig. 3 Sample of RSSCN7 dataset class

实验采用卷积网络模型 GoogLeNet^[20] 的全连接层输出作为场景图像特征。词向量融合分为不同训练模型、不同语料词向量融合。其中,不同训练模型的词向量融合实验涉及两种训练模型,即 Glove(gl)和 Word2Vec(wv),这两种词向量均在 Wikipedia 语料上训练得到。不同语料词向量融合实验采用两种训练语料,即 Wikipedia(Wiki)和 Common Crawl(Crawl),且均采用 Glove 模型对其进行训练。令 $Iter_N$ 为 40,定量实验采用总体分类准确度(Overall Accuracy, OA)作为评价指标,其值为重复 10 次实验的 ZSC 准确度的均值。对于 UCM 和 AID 场景集,按场景类别将其划分为 seen 类和 unseen 类数据,本文的 seen 类和 unseen 类划分情况分别为 16,5 和 25,5,即分别采用 16 类场景和 25 类场景作为 seen 类数据,余下的 5 类场景作为要进行分类识别的 unseen 类数据。将式(11)中的稀疏化门限参数 T_0 设置为 0.01,表示仅保留大于或等于 T_0 的元素值(须先对元素值进行归一化预处理)。

4.2 定量实验结果及分析

在 UCM 数据集和 AID 上进行定量实验,并从结构对齐、算法收敛性、融合效果以及与典型 ZSC 方法的对比这 4 个方面来分别阐述分析。

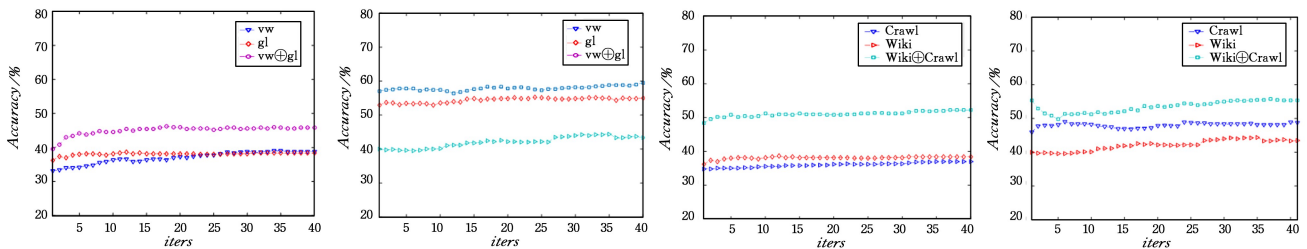


图 6 UCM 数据集和 AID 上 CADL 算法的收敛情况

Fig. 6 Convergence of CADL algorithm on UCM datasets and AID

4.2.1 结构对齐的效果分析

ZSC 的本质是借助语义词向量提供的类间距离关系,将图像特征空间中的类别原型迁移至 unseen 类,获得 unseen 类的图像特征空间原型表示,最后利用该原型表示对 unseen 样本进行分类。而本文结构对齐项的实质作用就是降低两种空间类别间距离的不一致性。因此,这里定义语义词向量空间与图像特征空间的类别距离结构差异度(DM)为:

$$DM = (\sum_{i,j} [d(c_i, c_j) - d(p_i, p_j)]^2)^{1/2} \quad (12)$$

其中, $d(c_i, c_j)$ 表示第 i, j 类别词向量 c_i 和 c_j 的余弦距离, $d(p_i, p_j)$ 表示第 i, j 类别图像特征空间类原型 p_i 和 p_j 的余弦距离。DM 值越大,两个空间的类间距离结构越不一致;DM 值越小,两个空间的类间距离结构越一致。图 4 和图 5 分别给出不同训练模型、不同训练语料词向量融合前后的 DM 变化情况。符号 \oplus 表示经过本文 CADL($M=2$)算法融合,符号 \otimes 表示直接串接的词向量。

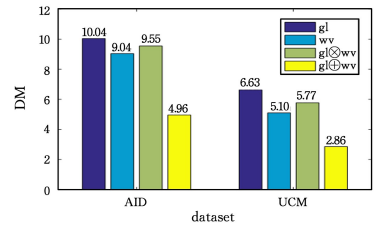


图 4 不同模型词向量融合的结构对齐效果

Fig. 4 Structure alignment effect of different model word vectors fusion

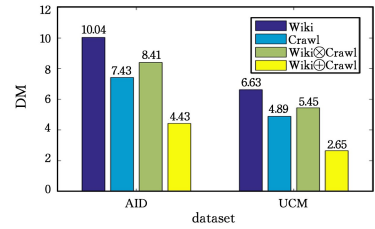


图 5 不同语料词向量融合的结构对齐效果

Fig. 5 Structure alignment effect of different corpora word vectors fusion

可以看出,相比未融合的单词向量及直接串接的词向量,本文 CADL 融合算法得到的词向量具有最小的 DM 值,表明 CADL 的结构对齐效果显著优于直接串接以及未融合的词向量的效果。

4.2.2 算法收敛分析

CADL 在 UCM 和 AID 上的运行情况如图 6 所示。可以看出,CADL 算法的准确度均能很快且稳定的收敛;而且,对于来自不同训练模型、不同训练语料的词向量,CADL 算法的 ZSC 准确度均显著高于未融合的词向量。

4.2.3 词向量融合效果分析

在 UCM 和 AID 上,不同训练模型词向量和不同训练语料词向量的融合效果如表 1 所列。可以看出,融合后的词向量在两种数据集上的 OA 值均得到显著提升。其中,训练模型 gl 与 wv 的融合词向量在 UCM 数据集上的 OA 为 45.30%,比单独的 gl 和 wv 词向量分别提升了 7.26%和 6.71%。

表 1 不同训练模型词向量与不同训练语料词向量的融合效果

Table 1 Fusion effect of different models and different corpora
(单位:%)

数据集	不同训练模型词向量相融合			不同训练语料词向量相融合		
	gl	wv	gl ⊕ wv	Wiki	Crawl	Wiki ⊕ Crawl
UCM	38.04	38.59	45.30	38.04	36.36	46.04
AID	42.67	53.70	58.42	42.67	48.09	54.35

4.2.4 与典型 ZSC 方法的比较

本节将 CADL 算法与典型 ZSC 方法进行对比,以验证其提升 ZSC 准确度的效果。表 2 中涉及 3 种语义词向量的融

表 2 CADL 及对比算法的 OA 值

Table 2 OA value of method and relative methods

(单位:%)

算法	UCM						AID					
	S ₁	S ₂	S ₃	S ₁ +S ₂	S ₂ +S ₃	S ₁ +S ₂ +S ₃	S ₁	S ₂	S ₃	S ₁ +S ₂	S ₂ +S ₃	S ₁ +S ₂ +S ₃
LatEm	18.80	20.40	19.80	33.00	23.00	20.80	15.90	22.65	23.81	18.71	28.17	21.62
RKT	40.00	39.80	44.60	40.20	43.60	43.60	48.92	48.03	48.15	48.92	50.13	53.25
DMaP	38.20	39.60	41.60	40.80	42.00	40.20	39.24	43.44	38.54	46.67	45.22	44.97
BiDiLEL	28.51	33.48	39.20	40.40	40.00	41.00	32.91	42.55	32.40	47.85	50.44	49.63
JLSE	37.20	34.20	45.60	37.60	34.80	38.00	36.11	34.97	42.30	35.99	43.50	45.54
CADL	36.36	38.59	38.04	46.20	45.30	48.40	48.09	53.70	42.67	58.55	58.42	60.23

4.3 定性实验结果及分析

为了定性分析 CADL 方法的实际遥感场景 ZSC 效果,以 RSSCN7 数据集作为 seen 类样本,对 2 幅高分辨率遥感图像 I 和 II(空间分辨率均为 0.3 m)进行 ZSC 分类。选择 ocean, airport 和 runway 作为 unseen 类。将 S₁+S₂+S₃ 得到的词向量用于定性实验。步骤为:首先,用单类别 SVM 判断遥感

场景样本是否属于 seen 类;然后,将不属于 seen 类的样本视为 unseen 类样本,采用 CADL 方法进行 ZSC。遥感图像 I 的大小为 17920×10752 像素,场景大小设定为 256×256 像素。本文及对比算法在遥感图像 I 上的 ZSC 效果如图 7 所示。可以看出,CADL 算法对 unseen 类场景的分类效果优于对比算法,其中 airport 类的场景分类效果更明显。

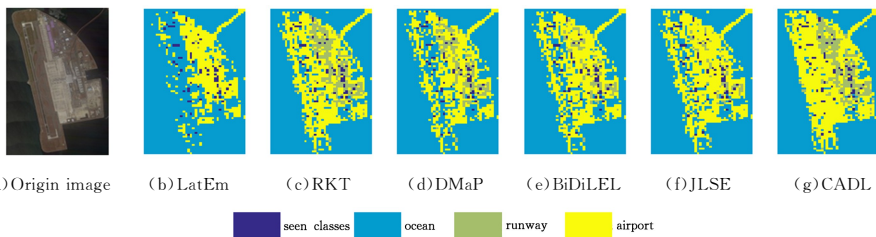


图 7 测试遥感图像 I 的场景零样本分类效果

Fig. 7 Zero-shot classification results of test remote sensing image I

遥感图像 II 的大小为 25344×29952 像素,场景大小设定为 256×256 像素。CADL 及对比算法在遥感图像 II 上的 ZSC 效果如图 8 所示。

可以看出,词向量融合算法 CADL 的 ZSC 效果总体优于参与对比的典型 ZSC 方法,其中 ocean 类场景的分类效果尤其明显。

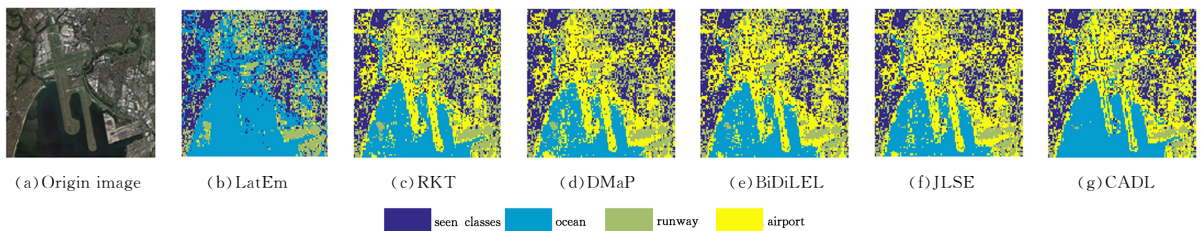


图 8 测试遥感图像 II 的场景零样本分类效果

Fig. 8 Zero-shot classification results of test remote sensing image II

结束语 针对词向量空间与图像特征空间中类原型结构不一致的问题,本文提出了基于词向量融合的遥感场景零样本分类算法。算法将词向量融合与结构对齐结合起来,利用耦合性计算得到图像特征空间中的 unseen 类原型表示,最后通过定量和定性实验验证了算法在不同训练模型、不同训练语料词向量融合方面的有效性。所提算法具有以下特点: 1) 利用解析字典学习方法提取词向量的稀疏系数; 2) 为降低结构差异性,通过串接稀疏系数得到新的词向量,将其线性映射到图像特征空间并与对应的原型对齐。实验结果表明,与典型 ZSC 方法相比,算法在缩小距离结构差异、提升总体分类准确度方面都有更优表现,说明算法能够利用不同词向量的互补性,提升 ZSC 分类的准确度。由于实验中的词向量均是在一般语料集上训练得到的,因此词向量的遥感领域的专业性不足,下一步将采用遥感领域的语料集训练词向量,以增强词向量的遥感专业性,降低结构不一致性,进而提升遥感场景零样本分类的效果。

参 考 文 献

- [1] CHEN S, TIAN Y L. Pyramid of spatial relations for scene-level land use classification[J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2014, 53(4): 1947-1957.
- [2] LI A, LU Z, WANG L, et al. Zero-shot scene classification for high spatial resolution remote sensing images[J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2017, 55(7): 4157-4167.
- [3] XIAN Y Q, AKATA Z, SHARMA G, et al. Latent embeddings for zero-shot classification[C]// *Proceedings of Computer Vision and Pattern Recognition*. LA USA: IEEE, 2016: 69-77.
- [4] WANG D, LI Y, LIN Y, et al. Relational Knowledge Transfer for Zero-Shot Learning[C]// *Proceedings of AAAI*. CA USA: 2016, 2-7.
- [5] ZHANG Z, SALIGRAMA V. Zero-shot learning via joint latent similarity embedding[C]// *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. LA USA: IEEE, 2016: 6034-6042.
- [6] ZHANG Z, SALIGRAMA V. Zero-shot learning via semantic similarity embedding[C]// *Proceedings of the IEEE International Conference on Computer Vision*. Boston USA: IEEE, 2015: 4166-4174.
- [7] WANG Q, CHEN K. Zero-shot visual recognition via bidirectional latent embedding[J]. *International Journal of Computer Vision*, 2017, 124(3): 356-383.
- [8] LI Y, WANG D, HU H, et al. Zero-shot recognition using dual visual-semantic mapping paths[J]. *arXiv:1703.05002*, 2017.
- [9] ZHAO B, WU B, WU T, et al. Zero-shot learning posed as a missing data problem[J]. *arXiv:1612.00560*, 2016.
- [10] LAMPERT C H, NICKISCH H, HARMELING S. Attribute-based classification for zero-shot visual object categorization[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2014, 36(3): 453-465.
- [11] SOCHER R, GANJOO M, MANNING C D, et al. Zero-shot learning through cross-modal transfer[C]// *Proceedings of Advances in Neural Information Processing Systems*. Lake Tahoe USA, ACM 2013: 935-943.
- [12] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. *arXiv:1301.3781*, 2013.
- [13] SOCHER R, MANNING C D. Glove: Global vectors for word representation[C]// *Empirical Methods in Natural Language Processing (EMNLP)*. 2014: 1532-1543.
- [14] YANG M, CHANG H Y, LUO W X. Discriminative analysis-synthesis dictionary learning for image classification[J]. *Neurocomputing*, 2017, 219: 404-411.
- [15] WANG J J, GUO Y Q, GUO J, et al. Synthesis linear classifier-based analysis dictionary learning for pattern classification [J]. *Neurocomputing*, 2017, 238: 103-113.
- [16] RAVISHANKAR S, BRESLER Y. L0 sparsifying transform learning with efficient optimal updates and convergence guarantees[J]. *IEEE Transactions on Signal Processing*, 2015, 63(9): 2389-2404.
- [17] YANG Y, NEWSAM S. Bag-of-visual-words and spatial extensions for land-use classification[C]// *Proceedings of the 18th SIGSPATIAL International Conference on Advances In Geographic Information Systems*. ACM, 2010: 270-279.
- [18] XIA G S, HU J, HU F, et al. AID: A benchmark dataset for performance evaluation of aerial scene classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(7): 3965-3981.
- [19] ZOU Q, NI L, ZHANG T, et al. Deep learning based feature selection for remote sensing scene classification[J]. *IEEE Geoscience & Remote Sensing Letters*, 2015, 12(11): 2321-2325.
- [20] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. *arXiv:1409.1556*, 2014.