

DOI:10.3969/j.issn.1003-5060.2023.08.021

基于正则化损失的 MeshNet 半监督分类

吕帅君¹, 邢燕¹, 洪沛霖²

(1. 合肥工业大学 数学学院, 安徽 合肥 230601; 2. 安徽中医药大学 医药信息工程学院, 安徽 合肥 230012)

摘要:深度学习在网格分类中的应用越来越受到人们的关注,在网格分类任务中,通常使用交叉熵损失作为损失函数。文章提出一种利用数据的结构相似性和几何一致性的正则化损失,将其加入损失函数中进行优化,可有效提高网格的分类准确率。从实验结果的量化指标来看,提出的正则化损失对于网格半监督分类任务的准确率有很好的提升效果。

关键词:正则化损失;网格分类;半监督学习;神经网络

中图分类号:TP391.411;TP181 **文献标志码:**A **文章编号:**1003-5060(2023)08-1142-05

Semi-supervised MeshNet classification based on regularization loss

LYU Shuaijun¹, XING Yan¹, HONG Peilin²

(1. School of Mathematics, Hefei University of Technology, Hefei 230601, China; 2. School of Medical Information Engineering, Anhui University of Chinese Medicine, Hefei 230012, China)

Abstract: The application of deep learning in mesh classification has attracted increasing attention. In mesh classification tasks, cross entropy loss is usually used as a loss function. In this paper, a regularization loss based on the structure similarity and geometric consistency of data is proposed, which can be directly added to the loss function to improve the classification accuracy of mesh. According to the final quantitative index of experimental results, the proposed regularization loss has a good effect on improving the accuracy of semi-supervised mesh classification task.

Key words: regularization loss; mesh classification; semi-supervised learning; MeshNet

三维形状的获取越来越容易,这对三维物体分类的发展至关重要,目前的主要方法大都依赖卷积神经网络解决三维物体的分类问题。卷积神经网络通过提取物体的各种高维特征对物体进行分类预测。

对于三维形状,有几种比较流行的数据类型:体素^[1]、点云^[2]、多视角图片^[3]和网格^[4]。体素在概念上类似于二维空间中的像素,可以看作数据在三维空间中的最小单位。体素表示法是一种规格化的表示方法,但其分辨率的限制导致很多细节缺失。点云是三维坐标系中点的集合,可以表示物体的外表形状;除了最基本的位置信息以外,

也可以在点云中加入其他信息,如点的色彩信息等。但由于点与点之间没有连接关系,点云表示方法缺少物体的表面信息。多视角图片(简称多视图)是通过不同视角的虚拟摄像机从物体模型中获取的二维图像的集合。多视图通常需要使用比较多的图片来构建完整的三维模型,在固定图片数量的情况下,容易受到物体自遮挡等因素的影响。网格是由一组多边形表面组成的非结构化数据,目前最普遍使用的是三角网格。多边形网格通过一种易于渲染的方式表示三维物体模型,在三维可视化等领域有很大作用。

近年来,随着神经网络的发展和显卡硬件设

收稿日期:2020-12-23;修回日期:2021-03-09

基金项目:国家自然科学基金资助项目(11601115);中央高校基本科研业务费专项资金资助项目(PA2020GDSK0060)

作者简介:吕帅君(1994—),男,河南安阳人,合肥工业大学硕士生;

邢燕(1977—),女,安徽合肥人,博士,合肥工业大学副教授,硕士生导师,通信作者,E-mail:xingyan@hfut.edu.cn.

备的升级,深度学习在识别、分类、数据挖掘、机器翻译、自然语言处理、语音处理、计算机视觉等领域都取得了许多成果,解决了很多复杂的模式识别难题,甚至在某些领域的识别能力超过人类。卷积神经网络在三维网格处理上的应用也越来越多。文献[4]提出了一种学习网格信息的卷积神经网络,它将面作为最小单元,定义了共享边的面之间的连接,使得网络能够通过逐面处理解决网格数据的复杂性和不规则性;文献[5]提出了一种基于边的能够简化网格模型用于三维网格分类和分割任务的卷积神经网络。

深度学习的成功得益于几十年积累的海量数据,但是对于很多特定任务的标签,数据并不充足,而且获取也相当困难,因此研究半监督、弱监督和无监督学习很必要。弱监督学习^[6]是已知数据和其一一对应的弱标签,训练一个智能算法,将输入数据映射到一组更强的标签的过程。半监督学习^[7]是指已知部分数据对应的标签,另一部分数据的标签未知,训练一个智能算法,学习已知标签和未知标签的数据,将输入数据映射到标签的过程。

受到弱监督的图像分割任务^[8-9]的启发,本文针对半监督三维网格数据的分类问题,在文献[4]的全监督 MeshNet 网络的基础上,利用网格数据的结构相似性和几何一致性,在一般的交叉熵损失函数中加入正则化损失。这一针对网格分类任务的正则化损失的设计,有效提高了网格分类的准确率。

1 本文方法

三维形状的网络数据包括顶点、边和面,边由 2 个顶点确定,封闭的边组成面。文献[4]将面看作最小处理单元,并将面的特征分为空间特征和结构特征。空间特征用三角形的中心点坐标表示;结构特征包括三角形中心点到 3 个顶点的角向量、三角形的单位法向量和三角形邻域面的索引,如图 1 所示。

文献[4]是在全监督的情况下对数据集进行分类,本文针对标签数据不足的半监督情况,增加正则化损失,提高分类准确率。为了测试本文提出的正则化损失在不同标注数据比例下的效率,对 10%、30%、50%、70%、90% 的标注数据均进行了实验。实验结果表明,增加的正则化损失可在各种标注比例的数据集上训练所得的模型获得更高的分类准确率。

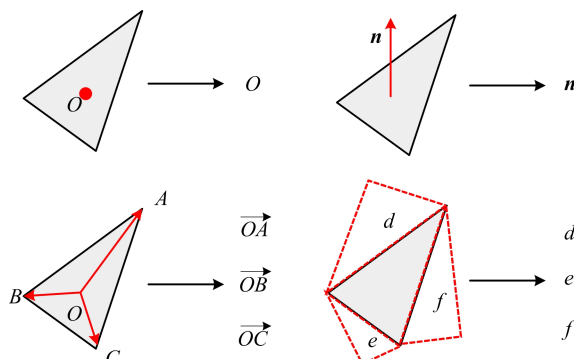


图 1 三角面的各种特征

1.1 网络结构

本文采用 MeshNet 网络,网络架构如图 2 所示。

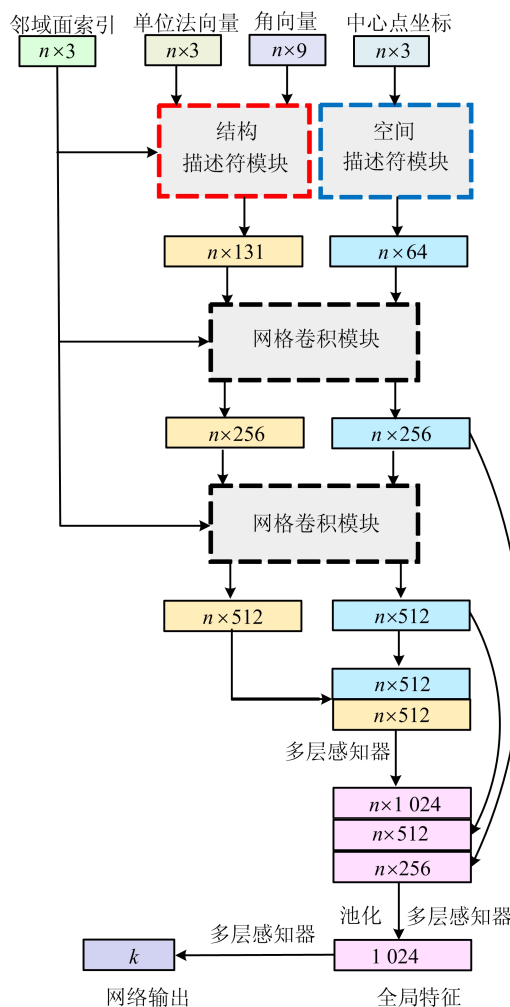


图 2 MeshNet 网络架构

网络的输入是三角形面相关的一系列初始值,如三角形中心点坐标、单位法向量、角向量、邻域面索引(图 1)。经过空间描述符模块和结构描述符模块产生初始的空间特征和结构特征。随后这

些特征与邻域面索引被聚合到网格卷积模块,并输入到一个池化函数中,输出用于后续任务的全局特征。

与空间位置相关的初始值是中心点坐标。在空间描述符模块中,对每一个中心点坐标简单地应用一个共享的多层感知器,输出初始的空间特征,如图 3 所示。



图 3 空间描述符模块

结构描述符模块比较复杂,包括面的旋转卷积和核相关操作。面旋转卷积(face rotation convolution, FRC)考虑面的形状信息,获取的是面的内部结构,而面核相关(face kernel correlation, FK)考虑面的周围信息,获取的是面的外部结构。

FRC 模块以面的角向量作为输入,假设面的角向量分别为 $\mathbf{v}_1, \mathbf{v}_2$ 和 \mathbf{v}_3 , 则 FRC 模块的输出定义如下:

$$g\left(\frac{1}{3}[f(\mathbf{v}_1, \mathbf{v}_2) + f(\mathbf{v}_2, \mathbf{v}_3) + f(\mathbf{v}_3, \mathbf{v}_1)]\right) \quad (1)$$

其中, $f(\cdot, \cdot, \cdot): \mathbf{R}^3 \times \mathbf{R}^3 \rightarrow \mathbf{R}^{K_1}$ 和 $g(\cdot): \mathbf{R}^{K_1} \rightarrow \mathbf{R}^{K_2}$ 是由待学习的权值形成的函数, K_1, K_2 分别表示函数 $f(\cdot, \cdot, \cdot)$ 和 $g(\cdot)$ 输出的维度。

FKC 模块将面的单位法向量和邻域面索引作为输入,计算面和可学习的核之间的相关性。第 i 个面和第 k 个核之间的核相关定义如下:

$$K_c(i, k) = \frac{1}{|N_i| |M_k|} \sum_{n \in N_i} \sum_{m \in M_k} K_\sigma(\mathbf{n}, \mathbf{m}) \quad (2)$$

其中: $|N_i|$ 为第 i 个面和它的邻域面的法向集合的元素个数; $|M_k|$ 为第 k 个核中法向集合的元素个数;核函数 $K_\sigma(\cdot, \cdot, \cdot): \mathbf{R}^3 \times \mathbf{R}^3 \rightarrow \mathbf{R}$ 度量 2 个向量的亲和度,实验选用高斯核函数。核相关可以挖掘网格中的局部结构^[10-11]。

网格卷积模块包括空间特征与结构特征的联合部分和结构特征与邻域特征的聚合部分。空间特征与结构特征的联合是将它们串联起来并使用 1 个多层感知器来生成新的特征。结构特征与邻域特征的聚合是先将结构特征与 3 个邻域面特征分别串联,并经过 1 个共享的多层感知器处理,再对结果应用最大池化得到新的特征。

1.2 正则化损失

深度学习中,正则化项有防止过拟合、提高泛化能力的作用。当模型参数过多时,网络会发生

过拟合,即对训练数据拟合得很好的模型通常无法很好地拟合测试数据。一般地,过于复杂的模型会引起过拟合,此时模型泛化能力比较低;相反,模型相对简单时,模型拟合能力较低,而泛化能力相对较高。引入正则化项是利用人们对问题的认识和对解的估计的先验知识。先验知识增加了对模型的限制,可以降低模型的方差,提高模型的稳定性,但也会使模型的偏置增大。偏置与方差之间的权衡是机器学习的基本主题之一。

本文利用“特征相似的网格更加可能属于同一类别”这样的先验知识提高分类准确率。网络学习得到的特征向量包含网格可用于分类的各种结构特征和空间几何特征,利用特征向量差值的 L2 范数的平方来度量不同网格之间的差异。正则化损失定义如下:

$$R(\mathbf{y}, \mathbf{f}) = \sum_{i,j} [1_{\|\mathbf{y}_i - \mathbf{y}_j\| < \delta} \|\mathbf{f}_i - \mathbf{f}_j\|^2 + 1_{\|\mathbf{y}_i - \mathbf{y}_j\| \geq \delta} \exp(-\|\mathbf{f}_i - \mathbf{f}_j\|^2)] \quad (3)$$

其中: \mathbf{y} 为预测的分类概率向量; \mathbf{f} 为网络学到的网格特征向量; \mathbf{f}_i 和 \mathbf{f}_j 分别表示从网络模型得到的第 i 个和第 j 个网格的特征向量; \mathbf{y}_i 和 \mathbf{y}_j 分别表示第 i 个和第 j 个网格的网络预测输出(属于不同类别的概率向量); $1_{\|\mathbf{y}_i - \mathbf{y}_j\| < \delta}$ 、 $1_{\|\mathbf{y}_i - \mathbf{y}_j\| \geq \delta}$ 是指示函数,分别表示条件 $\|\mathbf{y}_i - \mathbf{y}_j\| < \delta$ 和 $\|\mathbf{y}_i - \mathbf{y}_j\| \geq \delta$ 为真时等于 1, 否则等于 0; δ 为区分是否属于同一类别的阈值。

模型最终的损失函数如下:

$$l = D(\mathbf{y}, \mathbf{g}) + \lambda R(\mathbf{y}, \mathbf{f}) \quad (4)$$

其中: $D(\mathbf{y}, \mathbf{g})$ 为数据项,是预测结果 \mathbf{y} 与真实标签 \mathbf{g} 之间的交叉熵损失; $R(\mathbf{y}, \mathbf{f})$ 为式(3)定义的正则化项; λ 为平衡因子,用来平衡数据项和正则化项, λ 越大,正则化项的作用越大,不同的 λ 值会导致不同的分类结果。根据经验,实验中设置 $\lambda = 5 \times 10^{-6}$ 。

1.3 模型的实验设置

本文模型使用的数据集来自文献[1]的 ModelNet40 数据集,数据集包括来自 40 个类别的 12 311 个网格模型,其中 9 843 个用于训练,2 468 个用于测试。为了统一输入数据格式,每个网格被简化或扩充为 1 024 个面的网格,再被平移到网格的几何中心并归一化到单位球内。训练过程中,通过一个均值为 0、标准差为 0.01 的高斯噪声抖动顶点位置来增强数据。

模型训练采用的优化算法为随机梯度下降法,初始学习率为 0.01,动量为 0.9,学习率的衰

减率为 0.000 5。设置最大 epoch 次数为 150,批大小为 64。

本文实验在 Python3.6, Intel(R) Core(TM) i9-9900K CPU @3.60 GHz×16, RAM 32 GiB 和 Nvidia RTX 2080Ti GPU 环境中运行的。

2 实验结果与讨论

2.1 数值指标

关于分类的数值指标,本文选用常见的准确率(accuracy)和平均精度均值(mean average precision, mAP)。

首先引入 4 个符号:

T_P : 预测正确的正例样本个数;

T_N : 预测正确的负例样本个数;

F_P : 标签为负例,被错分为正例的样本个数;

F_N : 标签为正例,被错分为负例的样本个数。

准确率(A)是正确分类(即预测与标签一致)的样本数与样本总数之比。对于某一类别,计算公式如下:

$$A = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \quad (5)$$

其中:分子是预测正确的样本数,包括预测正确的正样本数 T_P 和预测正确的负样本数 T_N ;分母为样本总数,包括预测正确的样本数($T_P + T_N$)和预测错误的样本数($F_P + F_N$)。

平均精度均值(用 m_{AP} 表示)指标需要先定义查准率和召回率:

$$P = T_P / (T_P + F_P) \quad (6)$$

$$R = T_P / (T_P + F_N) \quad (7)$$

其中: P 为查准率(precision),表示真实的正例占预测为正例的样本比例,它针对预测为正例的结果; R 为召回率(recall),表示被正确预测的正例占总正例的比例,它针对正例标签的样本。

平均精度(average precision, AP)的定义如下:

$$A_P = \int_0^1 P(R) dR \quad (8)$$

其中, $P(R)$ 为系统在召回率为 R 时的查准率。

因为实验中的样本一般为离散量,所以求解 A_P 的积分一般用离散求和形式替代,具体表达式为:

$$A_P = \sum_{k=1}^N \max_{\bar{k} \geq k} P(\bar{k}) \Delta R(k) \quad (9)$$

其中: N 为检测样本数; $\Delta R(k)$ 为样本序列下标从 $k-1$ 变化到 k 时召回率的变化量; $\max_{\bar{k} \geq k} P(\bar{k})$ 为

相应区间上 P 的最大值。

m_{AP} 可表示 K 个类别的 A_P 均值或 K 次查询的 A_P 均值,定义如下:

$$m_{AP} = \frac{1}{K} \sum_{j=1}^K A_P(j) \quad (10)$$

2.2 对比实验

实验中,将本文方法与以往有代表性的方法作对比,结果见表 1 所列。本文方法使用了与 MeshNet 相同的实验参数设置,其他对比方法沿用了其论文中最高性能的参数设置,表 1 中的数据有些直接摘自其发表的论文。由于 MeshNet 网络架构充分利用了网格的空间特征和结构特征,加上精心设计的先验损失,本文所提的分类方法在全监督分类性能上比以往方法都有所提升,既远远超出传统的基于网格的 SPH 方法^[12]和基于多视图的 LFD 方法^[13],又能与基于体素的 3DShapeNets^[1]、基于点云的 PointNet^[2]、基于多视图的 MVCNN^[3]和基于网格的 MeshNet^[4]等深度学习的方法相媲美。

表 1 ModelNet40 数据集上的分类性能比较 %

方法	A	m_{AP}
SPH	68.20	33.30
LFD	75.50	40.90
3DShapeNets	77.30	49.20
MVCNN	90.10	79.50
PointNet	89.20	—
MeshNet	89.25	85.29
本文方法	90.38	85.69

为进一步验证所提的正则化损失的有效性,本文对 ModelNet40 数据集选择不同的标注比例进行半监督实验,实验结果的 2 种数值指标分别见表 2 和表 3 所列。

表 2 ModelNet40 数据集上半监督分类的准确率 %

标注数据比例	未加正则化	加正则化
10	21.14	34.66
30	74.80	83.24
50	85.35	86.16
70	86.69	86.93
90	88.80	88.88

表 3 ModelNet40 数据集上半监督分类的 m_{AP} 指标 %

标注数据比例	未加正则化	加正则化
10	26.32	39.83
30	68.35	76.08
50	78.92	79.44
70	81.66	82.43
90	84.44	84.79

- tion in flickering of three-coupled candle flames[J]. Scientific Reports, 2016, 6: 36145.
- [6] MANOJ K, PAWAR S A, SUJITH R I. Experimental evidence of amplitude death and phase-flip bifurcation between in-phase and anti-phase synchronization[J]. Scientific Reports, 2018, 8: 11626.
- [7] CHEN T, GUO X, JIA J, et al. Frequency and phase characteristics of candle flame oscillation[J]. Scientific Reports, 2019, 9: 342.
- [8] YANG T, XIA X, ZHANG P. Vortex-dynamical interpretation of anti-phase and in-phase flickering of dual buoyant diffusion flames [J]. Physical Review Fluids, 2019, 4(5): 053202.
- [9] 朱慧君, 陈轲. 卡门涡街的原理及其流体领域的应用[J]. 工程技术, 2017(3): 286-287.
- [10] 林杏全, 陈武秀. 纹影法及其应用[J]. 华中师院学报(自然科学版), 1980(1): 45-51.

(责任编辑 胡亚敏)

(上接第 1145 页)

同时,表 2 和表 3 也给出了消融实验的结果用以对比。从表 2 和表 3 可以看出,增加的正则化项在不同比例的标注数据下都提高了分类准确率。

3 结 论

本文针对半监督三维网格数据的分类问题提出一种基于正则化损失的 MeshNet 半监督分类方法,利用数据的结构相似性和几何相似性提高了分类准确率,多种量化指标都优于仅用交叉熵损失分类。本文所提的分类方法不仅适用于网格分类问题,而且在网格分割、网格去噪、图像处理、语音处理等领域均可以使用。在后续的工作中,将探讨超参数的设定,如式(3)中 δ 的选取,并探索设计更好的与任务相适应的先验项来提高网络性能。

[参 考 文 献]

- [1] WU Z, SONG S, KHOSLA A, et al. 3D ShapeNets: a deep representation for volumetric shapes[C]//28th IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 1912-1920.
- [2] QI C R, SU H, MO K, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]//30th IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 77-85.
- [3] SU H, MAJI S, KALOGERAKIS E, et al. Multi-view convolutional neural networks for 3D shape recognition[C]//15th IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 945-953.
- [4] FENG Y, FENG Y, YOU H, et al. MeshNet: mesh neural network for 3D shape representation[C]//33th AAAI Conference on Artificial Intelligence. Honolulu: ASSOC Advancement Artificial Intelligence, 2019: 8279-8286.
- [5] HANOCCA R, HERTZ A, FISH N, et al. MeshCNN: A network with an edge[J]. ACM Transactions on Graphics, 2019, 38(4): 1-12.
- [6] 刘永胜. 基于深度神经网络的弱监督学习方法在图像领域的研究[D]. 成都: 电子科技大学, 2020.
- [7] 王丹. 基于无监督与半监督框架的医学图像分类关键技术研究[D]. 长春: 吉林大学, 2020.
- [8] TANG M, DJELOUAH A, PERAZZI F, et al. Normalized cut loss for weakly-supervised CNN segmentation[C]//31th IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 1818-1827.
- [9] TANG M, PERAZZI F, DJELOUAH A, et al. On regularized losses for weakly-supervised CNN segmentation[C]//15th European Conference on Computer Vision. Munich: Springer, 2018: 524-540.
- [10] SHEN Y, FENG C, YANG Y, et al. Mining point cloud local structures by kernel correlation and graph pooling [C]//31th IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 4548-4557.
- [11] TSIN Y, KANADE T. A correlation-based approach to robust point set registration[C]//8th European Conference on Computer Vision. Berlin: Springer-Verlag, 2004: 558-569.
- [12] KAZHDAN M, FUNKHOUSER T, RUSINKIEWICZ S. Rotation invariant spherical harmonic representation of 3D shape descriptors [C]//Eurographics/ACM SIGGRAPH Symposium on Geometry Processing. Goslar: ACM, 2003: 156-164.
- [13] CHEN D, TIAN X, SHEN Y, et al. On visual similarity based 3D model retrieval[J]. Computer Graphics Forum, 2003, 22(3): 223-232.

(责任编辑 朱晓临)