

# 拉普拉斯阶梯网络<sup>\*</sup>

胡聪<sup>1</sup>, 吴小俊<sup>1</sup>, 舒振球<sup>2,3</sup>, 陈素根<sup>4</sup>



<sup>1</sup>(江南大学 人工智能与计算机学院, 江苏 无锡 214122)

<sup>2</sup>(江南大学 物联网工程学院, 江苏 无锡 214122)

<sup>3</sup>(江苏理工学院 计算机工程学院, 江苏 常州 213001)

<sup>4</sup>(安庆师范大学 数学与计算科学学院, 安徽 安庆 246133)

通讯作者: 吴小俊, E-mail: xiaojun\_wu\_jnu@163.com

**摘要:** 阶梯网络不仅是一种基于深度学习的特征提取器,而且能够应用于半监督学习中.深度学习在实现了复杂函数逼近的同时,也缓解了多层神经网络易陷入局部最小化的问题.传统的自编码、玻尔兹曼机等方法易忽略高维数据的低维流形结构信息,使用这些方法往往会获得无意义的特征表示,这些特征不能有效地嵌入到后续的预测或识别任务中.从流形学习的角度出发,提出一种基于阶梯网络的深度表示学习方法,即拉普拉斯阶梯网络 LLN (Laplacian ladder network).拉普拉斯阶梯网络在训练的过程中不仅对每一编码层嵌入噪声并进行重构,而且在各重构层引入图拉普拉斯约束,将流形结构嵌入到多层特征学习中,以提高特征提取的鲁棒性和判别性.在有限的有标签数据情况下,拉普拉斯阶梯网络将监督学习损失和非监督损失融合到了统一的框架进行半监督学习.在标准手写数据数据集 MNIST 和物体识别数据集 CIFAR-10 上进行了实验,结果表明,相对于阶梯网络和其他半监督方法,拉普拉斯阶梯网络都得到了更好的分类效果,是一种有效的半监督学习算法.

**关键词:** 阶梯网络;流形正则化;图拉普拉斯;深度自编码;半监督学习

**中图法分类号:** TP181

中文引用格式: 胡聪,吴小俊,舒振球,陈素根.拉普拉斯阶梯网络.软件学报,2020,31(5):1525-1535. <http://www.jos.org.cn/1000-9825/5680.htm>

英文引用格式: Hu C, Wu XJ, Shu ZQ, Chen SG. Laplacian ladder networks. Ruan Jian Xue Bao/Journal of Software, 2020,31(5): 1525-1535 (in Chinese). <http://www.jos.org.cn/1000-9825/5680.htm>

## Laplacian Ladder Networks

HU Cong<sup>1</sup>, WU Xiao-Jun<sup>1</sup>, SHU Zhen-Qiu<sup>2,3</sup>, CHEN Su-Gen<sup>4</sup>

<sup>1</sup>(School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China)

<sup>2</sup>(School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China)

<sup>3</sup>(School of Computer Engineering, Jiangsu University of Technology, Changzhou 213001, China)

<sup>4</sup>(School of Mathematics and Computational Science, Anqing Normal University, Anqing 246133, China)

**Abstract:** Ladder networks is not only an effective deep learning-based feature extractor, but also can be applied on semi-supervised learning. Deep learning has the advantage of approximating the complicated function and alleviating the optimization difficulty associated with deep models. Autoencoders and restricted Boltzmann machines ignore the manifold information of high-dimensional data and usually

\* 基金项目: 国家自然科学基金(61373055, 61672265, 61603159, 61702012, U1836218); 教育部 111 引智计划(B12018); 江苏省自然科学基金(BK20160293); 安徽省高等学校优秀青年人才支持计划(gxyq2017026)

Foundation item: National Natural Science Foundation of China (61373055, 61672265, 61603159, 61702012, U1836218); 111 Project of Ministry of Education of China (B12018); Natural Science Foundation of Jiangsu Province (BK20160293); Outstanding Young Talent Support Project of Higher Educations of Anhui Province (gxyq2017026)

收稿时间: 2018-05-03; 修改时间: 2018-06-16; 采用时间: 2018-09-23

achieve unmeaning features which are very difficult to use in the subsequent tasks, such as prediction and recognition. From the perspective of manifold learning, a novel deep representation method Laplacian ladder networks (LLN) is proposed, which is based on ladder networks (LN). When training LLN, LLN reconstructs noisy input and encoder layers, and adds graph Laplacian constrains to learn hierarchical representations for improving the robustness and discrimination of system. Under the condition of limited labeled data, LLN fuses the supervised learning and unsupervised learning to training in a semi-supervised manner. This study performs the experiments on the MNIST and CIFAR-10 datasets. Experimental results show that the proposed method LLN achieves superior performance compared with LN and other semi-supervised methods, and it is an effective semi-supervised method.

**Key words:** ladder network; manifold regularization; graph Laplacian; deep autoencoder; semi-supervised learning

近年来,深度学习促进了人工智能各领域的发展,特别是以卷积神经网络和长短时记忆网络为主的监督学习方法.它们在物体识别<sup>[1-3]</sup>、语音识别<sup>[4,5]</sup>、图像标题生成<sup>[6-8]</sup>、机器翻译<sup>[9,10]</sup>、医学检测<sup>[11]</sup>等方面取得了成功的应用.尽管这些方法取得了巨大成功,但是这些方法存在一个共同的缺点就是需要大量带标签的数据.而在实际生活中,得到这些带标签数据成本昂贵且非常耗时.半监督学习能将未标签数据的学习融入到监督学习的框架下,能够取得比单纯利用有标签数据学习更好的学习效果.因此,半监督学习成为机器学习研究者的一个重点研究方向.

半监督学习是机器学习的热点之一,根据半监督学习不同学习范式,目前的半监督学习算法大致可划分为 3 类:基于生成式方法的算法<sup>[12]</sup>、基于协同训练的算法<sup>[13,14]</sup>和基于图正则化框架的算法<sup>[15]</sup>.流形正则化(manifold regularization,简称 MR)是一种常用的图正则化框架<sup>[16]</sup>.半监督学习最早出现于 20 世纪 70 年代,当时大多数是基于生成式方法的研究.直推式支持向量机(TSVM)<sup>[17]</sup>也是一种半监督方法,和传统的支持向量机一样,它的学习目的也是最大化决策边界和数据的间隔<sup>[18]</sup>.直推式支持向量机尝试将无标签数据也考虑进来达到同样的学习目的.但是这种支持向量机半监督学习方式较难学习,因为这是一个非凸问题.2006 年,Hinton 等人首次提出深度学习的概念,提出了多层自动编码器深层结构<sup>[19]</sup>.后来,深度置信网(DBN)<sup>[20]</sup>和堆叠自动编码器(SAE)<sup>[21]</sup>也相继被提出来.近年来,大多数的学习任务都利用深度学习得到了非常好的表现,而半监督学习作为机器学习的其中一项任务也得到了长足的发展.深度学习源自以自编码和受限波尔兹曼机为基础的非监督学习方法,以自编码为基础表示学习研究中,大部分学习采用两步法的学习过程:第 1 阶段为无监督的特征学习过程,第 2 阶段为有监督的微调.这种学习范式成为最初的半监督深度学习途径.在第 1 阶段,无监督特征学习初始化模型参数;在第 2 阶段的监督学习过程中,再对这些参数进行微调.由于对无监督的特征表示进行了不同约束而发生特征变化,有监督学习会降低无监督学习的学习效果.同时,第一阶段特征学习并未用到任何监督学习的信息,削弱了第 2 步监督学习时目标信息对于非监督学习的启示性.随着深度学习的不断发展,近些年来还有其他一些深度学习的半监督方法取得了较好的效果,如 EmbedCNN<sup>[22]</sup>、S3C<sup>[23]</sup>等.变分自编码(VAE)<sup>[24]</sup>是一种深度生成模型,它将变分推断运用到自动编码器中,并在半监督学习中得到了较好的效果.SWVAE<sup>[25]</sup>是另外一种基于卷积自动编码器的深度学习模型,它将监督学习、非监督学习和半监督学习融入到一个统一的框架之中.2015 年,阶梯网络(ladder network,简称 LN)<sup>[26]</sup>被提出来,LN 将监督学习和非监督学习融入到一个统一的框架中同时进行学习,这样的半监督学习方式能够有效提高监督信息的利用效率和非监督特征学习的判别性.

大多数以自编码为基础结构的深度学习方法之所以能取得成功,是因为这些网络将高维数据的流形结构信息考虑进来,比如降噪自编码(DAE)<sup>[21]</sup>、图正则化深度神经网络(GR-DNN)<sup>[27]</sup>.本文将流形正则化框架引入到阶梯网络梯网络的学习中,用流形正则化约束各层特征学习,使得各层特征具有局部保持特性.拉普拉斯阶梯网络(LLN)把标记样本和未标记样本一起训练,并用于半监督模式识别.在对含噪声数据进行重构的同时,通过提取数据的图拉普拉斯结构信息,不仅对重构数据进行流形约束,而且对各解码层进行流形约束,有效地提高了特征提取的判别能力,从而提高了 LLN 模型半监督学习的泛化性能.

本文研究并提出了拉普拉斯阶梯网络算法.第 1 节简述相关工作,包括传统的降噪自动编码器和阶梯网络算法.第 2 节介绍我们提出的拉普拉斯阶梯网络.第 3 节用标准手写数字数据集 MNIST 与物体检测数据集 CIFAR-10 进行测试,并在基于多层传感器和卷积结构的两种拉普拉斯阶梯网络上实验,与其他半监督方法进行

对比分析.实验结果表明,在少量样本有标签的情况下,与传统的方法及阶梯网络相比,拉普拉斯阶梯网络在分类精度上都有所提高.第4节是结论和未来工作展望.

## 1 相关工作

### 1.1 降噪自动编码器

自动编码器(autoencoders,简称 AE)是通过训练来达到重构数据自身的一种神经网络,由编码器和解码器两部分组成<sup>[28]</sup>:编码器通过函数  $f$  将原始数据  $x$  映射到隐含层  $h$ ;解码器正好相反,用解码函数  $g$  将隐含层  $h$  作为输入,得到数据重构  $\hat{x}$ ,重构数据具有和原始数据一样的结构,用公式表达如下:

$$h=f(W_e x+b_e) \tag{1}$$

$$\hat{x}=g(W_d h+b_d) \tag{2}$$

其中, $W_e, b_e$  分别是编码器的权重和偏置; $W_d, b_d$  分别是解码器的权重和偏置; $f, g$  分别为编码器和解码器的激活函数,在自动编码器训练的过程中,它使得重构数据  $\hat{x}$  与原始输入  $x$  尽量逼近,从而表明隐含层  $h$  学到了原始数据另外一种潜在的表达方式.因此,在给定一个数据集  $X=\{x_1, x_2, \dots, x_N\}$ ,自动编码器通过最小化以下函数来优化模型参数  $\theta=\{W_e, b_e, W_d, b_d\}$ :

$$\arg \min_{\theta} \frac{1}{N} \sum_{i=1}^N C(x_i, g(f(x_i))) \tag{3}$$

其中  $f, g$  分别为编码器和解码器的激活函数; $C$  为重构函数,一般使用平方误差函数或者交叉熵损失函数.

基于统计理论的降噪自动编码器是一种改进的自动编码器,其核心思想是:编码器将含有一定统计特性的噪声加入输入数据,成为腐坏数据  $\tilde{x}$ ,然后进行编码得到隐含层码字  $h=f(\tilde{x})$ ;解码器在对隐含层进行解码过程中,估计出噪声分布,消除背景噪声,从而重构出干净的不含噪声的原始数据.降噪自动编码器的代价函数如下.

$$\arg \min_{\theta} \frac{1}{N} \sum_{i=1}^N C(x_i, g(f(\tilde{x}_i))) \tag{4}$$

### 1.2 阶梯网络

深度自动编码器是阶梯网络的基础结构,区别于一般的单层自动编码器和由单层自动编码器堆叠起来的栈式自动编码器,深度自动编码由一个深度编码器和一个深度解码器组成,多层自动编码网络公式如下.

$$h^{(r)}=f^{(r)}(W_e^{(r)} h^{(r-1)}+b_e^{(r)}) \tag{5}$$

$$\hat{h}^{(r)}=f^{(r)}(W_d^{(r)} \hat{h}^{(r+1)}+b_d^{(r)}) \tag{6}$$

其中  $f^{(r)}$  和  $g^{(r)}$  分别为编码器和解码器第  $r$  层的激活函数,  $W_e^{(r)}$  和  $b_e^{(r)}$  分别是编码器第  $r$  层的权重和偏置,  $W_d^{(r)}$  和  $b_d^{(r)}$  分别是解码器第  $r$  层的权重和偏置.深度自动编码器的示意图如图 1 所示.

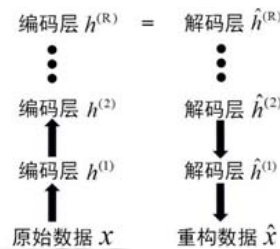


Fig.1 Architecture of deep autoencoders

图 1 深度自动编码器结构

阶梯网络<sup>[26]</sup>是由 Rasmus 等人提出的一种改进的非监督深度表示学习模型.与深度自动编码器不同的是,阶梯网络每一编码层都有一条横向连接线连接到对应的解码层,而且阶梯网络在编码层的每一层都引入噪声,类似于降噪自动编码器.不过,降噪自动编码器只对输入层加入噪声,而阶梯网络对每一层都进行加噪处理.阶

梯网络的损失函数也不同于传统的降噪自动编码器,阶梯网络不仅要求模型能从含噪声的数据中重构出原始干净的数据,而且要求特征也能从含噪声的特征层中重构出干净的特征.阶梯网络的损失函数如下所示.

$$\arg \min_{\theta} \frac{1}{N} \sum_{r=1}^R \sum_{i=1}^N \lambda^{(r)} C^{(r)}(z_i^{(r)}, \hat{z}_i^{(r)}) \quad (7)$$

其中,输入层  $x=z^{(0)}$ ;  $C^{(r)}$ 表示第  $r$  层的重构误差,一般用最小平方误差来衡量各层的重构误差; $\lambda^{(r)}$ 表示第  $r$  层重构误差的调节权重.

### 2 拉普拉斯阶梯网络

拉普拉斯阶梯网络在半监督学习的过程中,将监督学习和非监督学习组合在统一的框架中进行学习,它还利用编码器和解码器在不同层级的横向快捷连接,提高深层网络提取数据抽象不变性特征的能力.拉普拉斯阶梯网络结构示意图如图 2 所示.

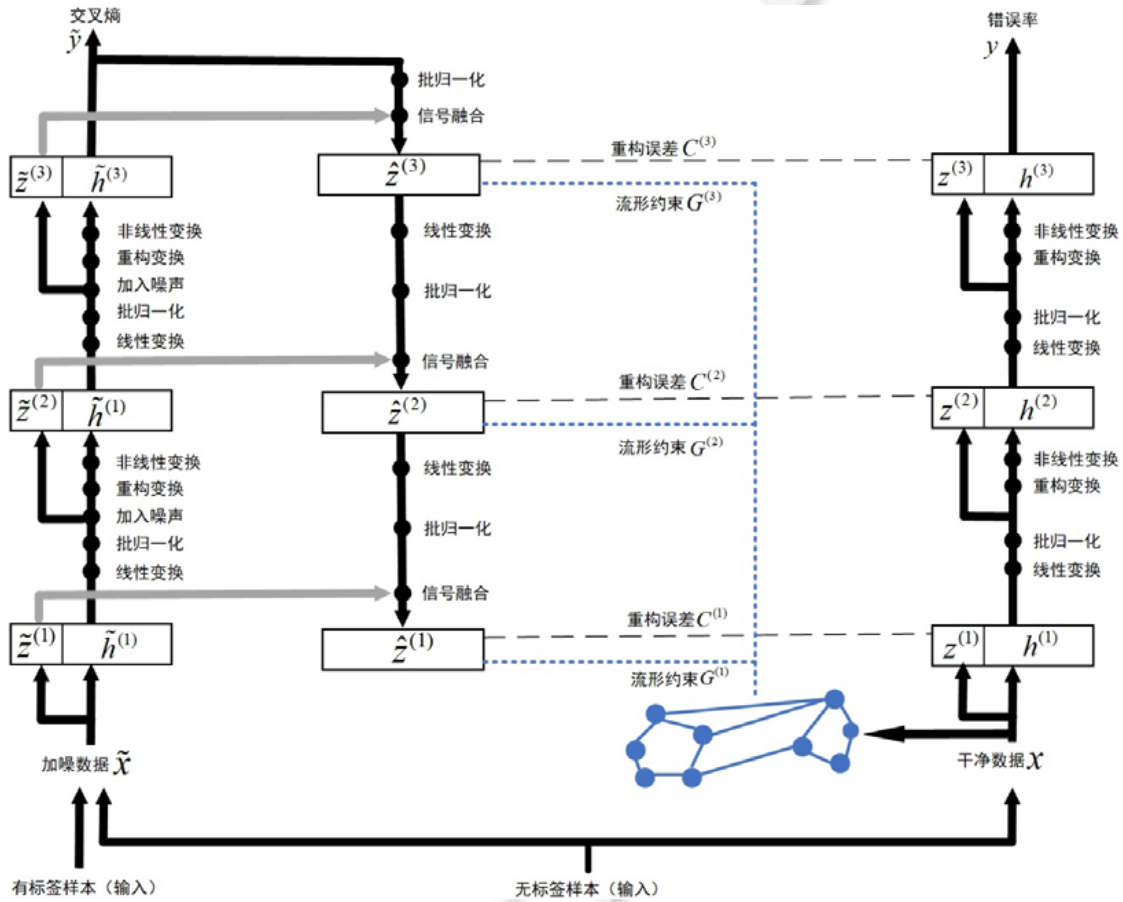


Fig.2 Architecture of Laplacian ladder networks

图 2 拉普拉斯阶梯网络结构

基于图算法的正则化框架是当前半监督学习研究的热点之一,它利用了流形假设的思想.它把标记和未标记的样本当作数据图的顶点,顶点的边当作权重.如果两个样本  $x_i$  和  $x_j$  接近,那么它们的特征  $z_i$  和  $z_j$  也应该彼此接近.为了实现这个假设,流形正则化构架使用以下代价函数最小化.

$$G^{(r)} = \sum_{i,j} v_{ij} \| \hat{z}_i^{(r)} - \hat{z}_j^{(r)} \|^2 = \text{tr}(\hat{Z}^T (D - V) \hat{Z}) = \text{tr}(\hat{Z}^T L \hat{Z}) \quad (8)$$

其中,特征矩阵  $Z = [\tilde{z}, \tilde{z}_j]^T$ ,  $D$  是对角矩阵,  $L$  是拉普拉斯矩阵,亲和矩阵(近邻图)  $V = [v_{ij}]_{N \times N}$ .

$$v_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{\sigma}}, & x_j \in N(x_i) \\ 0, & x_j \notin N(x_i) \end{cases} \quad (9)$$

阶梯网络忽略了流形结构信息,拉普拉斯阶梯网络显式地加入了流形约束到阶梯网络中,对解码器进行显式的流形约束.目标函数如下.

$$\arg \min_{\theta} \frac{1}{N} \sum_{r=1}^R \sum_{i=1}^N \lambda^{(r)} C^{(r)}(z_i^{(r)}, \hat{z}_i^{(r)}) + \frac{1}{N} \sum_{r=1}^R \sum_{i=1}^N \sum_{j=1}^N \beta^{(r)} G^{(r)}(\hat{z}_i^{(r)}, \hat{z}_j^{(r)}, X) \quad (10)$$

其中,  $R$  代表拉普拉斯阶梯网络由下而上的总层数,  $N$  表示训练样本集规模,  $\tilde{h}^{(r)}$  代表第  $r$  层含有噪声的编码特征,  $\hat{z}^{(r)}$  表示第  $r$  层的解码特征.

从图 2 中可以看出,拉普拉斯阶梯网络可以分为含噪声数据编码器、干净数据编码器和降噪解码器这 3 部分.这 3 部分表示如下.

$$\tilde{x}, \tilde{z}^{(1)}, \dots, \tilde{z}^{(R)}, \tilde{y} \quad (11)$$

$$x, z^{(1)}, \dots, z^{(R)}, y \quad (12)$$

$$\hat{x}, \hat{z}^{(1)}, \dots, \hat{z}^{(R)} \quad (13)$$

其中,  $x, y, \tilde{y}$  分别表示输入、干净数据的输出和含有噪声数据的输出.在训练的过程中,交叉熵函数利用含噪声的输出  $\tilde{y}$  作监督学习的损失;而在测试的过程,利用干净的数据输出  $y$  作测试精度的评估计算.  $z, \tilde{z}^{(r)}, \hat{z}^{(r)}$  分别表示第  $r$  层的干净数据的隐含层特征、含噪声的隐含层特征及重建特征.

在编码器的前向传播的过程中,批归一化(BN)技术得到应用:

$$\tilde{z}_{pre}^{(r)} = W^{(r)} \tilde{h}^{(r-1)} \quad (14)$$

$$\mu^{(r)} = \text{mean}(\tilde{z}_{pre}^{(r)}) \quad (15)$$

$$\sigma^{(r)} = \text{stdv}(\tilde{z}_{pre}^{(r)}) \quad (16)$$

$$\tilde{z}^{(r)} = \frac{\tilde{z}_{pre}^{(r)} - \mu^{(r)}}{\sigma^{(r)}} + N(0, \sigma^2) \quad (17)$$

$$\tilde{h}^{(r)} = \phi(\gamma^{(r)}(\tilde{z}^{(r)} + \beta^{(r)})) \quad (18)$$

其中,  $\phi(\cdot)$  表示非线性激活函数.上述公式代表含噪声数据的编码过程,如果去掉随机噪声  $N(0, \sigma^2)$ ,且将含噪声的输出  $\tilde{h}, \tilde{z}$  用  $h, z$  代替,我们将得到干净数据的编码.

在解码的过程中,与传统的深度自编码不同,阶梯网络的解码层不仅有层间的纵向连接,而且各解码层和对应的编码层有横向连接,对应的解码通过融合函数  $g$  输出,详细的解码过程如下.

$$u_{pre}^{(r+1)} = V^{(r)} \tilde{z}^{(r+1)} \quad (19)$$

$$\mu^{(r+1)} = \text{mean}(u_{pre}^{(r+1)}) \quad (20)$$

$$\sigma^{(r+1)} = \text{stdv}(u_{pre}^{(r+1)}) \quad (21)$$

$$u^{(r+1)} = \frac{u_{pre}^{(r+1)} - \mu^{(r+1)}}{\sigma^{(r+1)}} \quad (22)$$

$$\hat{z}^{(r)} = g(\tilde{z}^{(r)}, u^{(r+1)}) \quad (23)$$

其中,  $V^{(r)}$  表示连接  $r+1$  和  $r$  层的权重.融合函数  $g$  通过纵向连接  $u^{(r+1)}$  和横向连接  $\tilde{z}^{(r)}$  得到解码输出  $\hat{z}^{(r)}$ ,形式如下.

$$g(\tilde{z}^{(r)}, u^{(r+1)}) = b_0 + w_{0z} \odot \tilde{z}^{(r)} + w_{0u} \odot u^{(r+1)} + w_{0zu} \odot \tilde{z}^{(r)} \odot u^{(r+1)} + w_{\sigma} \odot \text{Sigmoid}(b_1 + w_{1z} \odot \tilde{z}^{(r)} + w_{1u} \odot u^{(r+1)} + w_{1zu} \odot \tilde{z}^{(r)} \odot u^{(r+1)}) \quad (24)$$

其中,  $\odot$  是点乘操作.  $w_{\{0,1\}z}, w_{\{0,1\}u}, w_{\{0,1\}zu}, b_{\{0,1\}}, w_{\sigma}$  的初始分别为 1,0,0,0,1.最终,重建函数  $C$  如下所示.

$$C^{(r)}(z^{(r)}, \hat{z}^{(r)}) = \left\| \frac{\hat{z}^{(r)} - u^{(r)}}{\sigma^{(r)}} - z^{(r)} \right\|^2 \quad (25)$$

非监督学习损失  $L_{unsup}$  由各层的重构误差  $C^{(r)}$  和流形正则项  $G^{(r)}$  组合而成, 而有监督损失是通过对  $M$  个有标签数据进行学习得到的, 在含噪声数据编码器的最顶层接入分类器, 对有标签数据进行分类误差计算, 得到监督学习的损失函数:

$$L_{sup} = -\sum_{i=1}^M \log P(\tilde{y}_i = y_i^* | x_i) \quad (26)$$

在对拉普拉斯阶梯网络的训练过程中, 监督学习利用有标签数据来提升模型的判别性, 重构误差项通过衡量干净数据的编码与含噪声数据的解码之间的差异提高模型的鲁棒性, 图正则化项通过局部保持约束数据在解码过程中仍保持与输入数据相同的低维流形结构. 最后, 拉普拉斯阶梯网络的详细公式如下.

$$\arg \min_{\theta} -\sum_{i=1}^M \log P(\tilde{y}_i = y_i^* | x_i) + \sum_{r=1}^R \sum_{i=1}^N \lambda^{(r)} C^{(r)}(z_i^{(r)}, \hat{z}_i^{(r)}) + \sum_{r=1}^R \sum_{i=1}^N \sum_{j=1}^N \beta^{(r)} G^{(r)}(\hat{z}_i^{(r)}, \hat{z}_j^{(r)}, v_{ij}) \quad (27)$$

$\lambda^{(r)}$  与  $\beta^{(r)}$  代表第  $r$  层重构误差和流形约束的调节参数. 拉普拉斯阶梯网络的损失函数  $L_{LLN}$  由有标签数据的监督损失  $L_{sup}$ 、无标签数据的重构误差  $C$  和拉普拉斯正则约束  $G$  这 3 部分组成. 利用公式(27)得到损失函数后, 我们利用传统的梯度下降算法对模型参数  $\theta$  进行偏导计算.

$$\frac{\partial L_{LLN}}{\partial \theta} = \frac{\partial L_{sup}}{\partial \theta} + \frac{\partial C}{\partial \theta} + \frac{\partial G}{\partial \theta} \quad (28)$$

拉普拉斯阶梯网络算法的具体流程如算法 1 所示.

**算法 1.** 拉普拉斯阶梯网络(LLN)算法.

输入:

- 有标签样本集  $\{(x_1, y_1^*), (x_2, y_2^*), (x_3, y_3^*), \dots, (x_M, y_M^*)\}$ ;
- 无标签样本集  $\{x_{M+1}, x_{M+2}, x_{M+3}, \dots, x_{M+N}\}$ ;
- 近邻个数  $K$ ;
- 各层重构误差项和拉普拉斯正则约束项调节参数  $\lambda^{(r)}, \beta^{(r)}$ ;
- 学习速率  $u$ .

输出: 编码器权值  $\theta$ .

Step 1. 加入噪声的数据编码和分类

$$\tilde{h}^{(0)} \leftarrow \tilde{z}^{(0)} \leftarrow x_i + noise$$

**for**  $r=1$  to  $R$  **do**

$$\tilde{z}^{(r)} \leftarrow batchnorm(W^{(r)}\tilde{h}^{(r-1)}) + noise$$

$$\tilde{h}^{(r)} \leftarrow activation(\gamma^{(r)} \odot (\tilde{z}^{(r)} + \beta^{(r)}))$$

**end for**

$$P(\tilde{y} | x) \leftarrow \tilde{h}^{(R)}$$

Step 2. 干净数据编码

$$h^{(0)} \leftarrow z^{(0)} \leftarrow x_i$$

**for**  $r=1$  to  $R$  **do**

$$z^{(r)} \leftarrow batchnorm(W^{(r)}h^{(r-1)})$$

$$h^{(r)} \leftarrow activation(\gamma^{(r)} \odot (z^{(r)} + \beta^{(r)}))$$

**end for**

Step 3. 对加入噪声的数据进行解码和降噪

**for**  $r=R$  to  $0$  **do**

**if**  $r=R$  **then**

$$u^{(R)} \leftarrow batchnorm(\tilde{h}^{(R)})$$

```

else
     $u^{(r)} \leftarrow \text{batchnorm}(V^{(r+1)}\hat{z}^{(r+1)})$ 
end if
 $\forall i: \hat{z}_i^{(r)} \leftarrow g(\hat{z}_i^{(r)}, u_i^{(r)})$ 
 $\forall i: \hat{z}_{i, BN}^{(r)} \leftarrow \frac{\hat{z}_i^{(r)} - u_i^{(r)}}{\sigma_i^{(r)}}$ 
    
```

end for

Step 4. 计算拉普拉斯阶梯网络的损失函数

用  $K$  近邻算法构建原始数据的亲和矩阵  $V=[v_{ij}]_{N \times N}$ .

计算有标签数据的监督损失  $L_{sup}: L_{sup} = -\sum_{i=1}^M \log P(\tilde{y}_i = y_i^* | x_i)$ .

计算各层的重构误差  $C^{(r)}: C^{(r)}(z^{(r)}, \hat{z}^{(r)}) = \left\| \frac{\hat{z}^{(r)} - u^{(r)}}{\sigma^{(r)}} - z^{(r)} \right\|^2$ .

计算各层的拉普拉斯正则约束  $G^{(r)}: G^{(r)} = \sum_{i,j} v_{ij} \| \hat{z}_i^{(r)} - \hat{z}_j^{(r)} \|^2 = \text{tr}(\hat{Z}^T (D - V) \hat{Z}) = \text{tr}(\hat{Z}^T L \hat{Z})$ .

计算拉普拉斯阶梯网络总损失  $L_{LLN}$ :

$$\arg \min_{\theta} -\sum_{i=1}^M \log P(\tilde{y}_i = y_i^* | x_i) + \sum_{r=1}^R \sum_{i=1}^N \lambda^{(r)} C^{(r)}(\hat{z}_i^{(r)}, \hat{z}_i^{(r)}) + \sum_{r=1}^R \sum_{i=1}^N \sum_{j=1}^N \beta^{(r)} G^{(r)}(\hat{z}_i^{(r)}, \hat{z}_j^{(r)}, v_{ij}).$$

Step 5. 误差反向传播及更新参数  $\theta$

计算参数  $\theta$  的偏导数:  $\frac{\partial L_{LLN}}{\partial \theta} = \frac{\partial L_{sup}}{\partial \theta} + \frac{\partial C}{\partial \theta} + \frac{\partial G}{\partial \theta}$ .

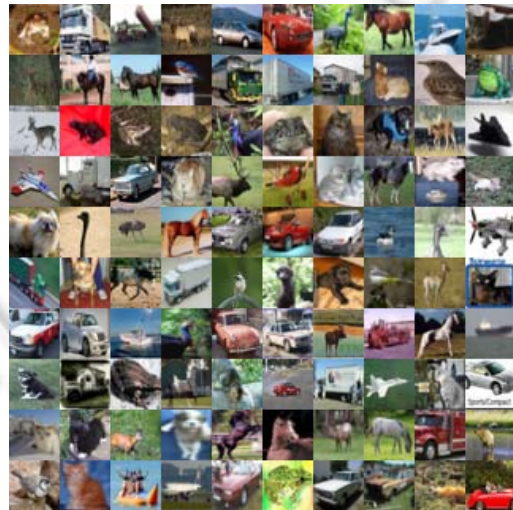
利用梯度下降方法更新编码器权值  $\theta: \theta \leftarrow \theta - u \times \frac{\partial L_{LLN}}{\partial \theta}$ .

### 3 实验与分析

为了测试所提出算法的性能,我们用 MNIST<sup>[29]</sup>和 CIFAR-10<sup>[30]</sup>数据集进行测试,两个数据集的部分样本如图 3 所示.



(a) MNIST 数据集部分样本



(b) CIFAR-10 数据集部分样本

Fig.3 Some samples of MNIST (a) and CIFAR-10 (b) datasets

图 3 MNIST(a)和 CIFAR-10(b)数据集的部分样本

本文对两种网络结构的拉普拉斯阶梯网络进行了实验,包括多层感知机和卷积结构的拉普拉斯阶梯网络.一共使用了 3 个深层网络模型,包括一个 7 层的感知器模型和两个卷积网络模型.本文对 MNIST 和 CIFAR-10 所使用的两个不同的卷积网络模型描述见表 1.在对比实验中,本文的方法与其他传统的半监督学习方法的测试结果进行了比较,参与对比的方法如下.

- 支持向量机(SVM)<sup>[18]</sup>
- 直推式支持向量机(TSVM)<sup>[17]</sup>
- 变分自编码(VAE)<sup>[24]</sup>
- EmbedCNN<sup>[22]</sup>
- SWWAE<sup>[25]</sup>
- S3C<sup>[23]</sup>
- 图正则化深度神经网络(GR-DNN)<sup>[27]</sup>
- 阶梯网络(LN)<sup>[26]</sup>

**Table 1** Model parameters of convolutional Laplacian ladder networks (Conv-LLN)

**表 1** 卷积拉普拉斯阶梯网络(Conv-LLN)模型参数

Conv-LLN (for MNIST)	Conv-LLN (for CIFAR-10)
Input: 28×28×1	Input: 32×32×3
5×5 conv.32, ReLU	3×3conv. 96, BN, LeakyReLU
–	3×3conv. 96, BN, LeakyReLU
–	3×3conv. 96, BN, LeakyReLU
2×2max-pooling stride 2, BN	2×2max-pooling stride 2, BN
3×3conv. 64, BN, ReLU	3×3conv. 192, BN, LeakyReLU
3×3conv. 64, BN, ReLU	3×3conv. 192, BN, LeakyReLU
–	3×3conv. 192, BN, LeakyReLU
2×2max-pooling stride 2, BN	2×2max-pooling stride 2, BN
3×3conv. 128, BN, ReLU	3×3conv. 192, BN, LeakyReLU
–	1×1 conv. 192, BN, LeakyReLU
1×1 conv. 10, BN, ReLU	1×1 conv. 10, BN, LeakyReLU
Global meanpool, BN	Global meanpool, BN
Fully connected 10, BN	–
10-way softmax	10-way softmax

### 3.1 MNIST

为了验证拉普拉斯阶梯网络的半监督学习效果,我们分别利用多层感知器和卷积两种深度网络结构,对拉普拉斯阶梯网络进行了实验.第 1 组实验,我们在 MNIST 数据集进行测试,MNIST 数据集包含 70 000 个数据,60 000 个训练样本和 10 000 个测试样本,所有样本的大小皆为 28×28 的灰度图像.我们将 MNIST 数据集的 60 000 个训练样本随机拆分出 10 000 个样本作为验证集,其他的 50 000 个样本组成训练集.

#### 3.1.1 多层感知机

本文首先利用一个 784-1000-5000-250-250-250-10 的由下向上的 7 层全连接多层感知机模型,来测试拉普拉斯阶梯网络的半监督分类能力.通过交叉验证法和大量的实验,确定了模型的超参数.第 1 层对应的超参 $\lambda^{(l)}$ 和 $\beta^{(l)}$ 对应的值分别为[10e2 10e0 10e-2 10e-2 10e-2 10e-2 10e-2]和[10e-2 10e-4 10e-6 10e-6 10e-6 10e-6 10e-6],在训练过程中,我们随机挑选出 100 个含标签数据作监督损失的计算,且所挑选的各类的样本数量一致,所有的训练数据去除标签后做非监督损失的计算.结果见表 2.为了更好地评估拉普拉斯阶梯网络的稳定性,我们进行了 10 组不同的模型初始化参数,最后对 10 次的测试结果进行了平均.

从表 2 中可以看出,在有标签数据量  $N$  仅为 100 的情况下,基于多层传感器模型的拉普拉斯阶梯网络在 MNIST 数据集上得到了最好的实验结果.将流形约束加入到模型训练的方法取得了较好的效果,如图正则化深度神经网络(GR-DNN)和拉普拉斯阶梯网络(LLN).拉普拉斯阶梯网络对模型的每一层特征进行了流形约束,更加充分地利用了局部结构信息,得到了更好的半监督分类精度.利用流形约束的拉普拉斯阶梯网络比非流形约



束的阶梯网络具有更好的类别判别能力,且分类精度与其他半监督学习方法相比也有较大的优势,模型的稳定性也有所提升.

**Table 2** Performance comparison of MLP-Laplacian ladder networks and other methods on MNIST dataset

表 2 各算法与多层感知机拉普拉斯阶梯网络在 MNIST 数据集上的性能比较

方法	测试误差(% ,N=100)
支持向量机(SVM)	23.44
直推式支持向量机(TSVM)	16.81
变分自编码(VAE+SVM)	11.82±0.25
变分自编码(M1+M2)	3.33±0.14
图正则化深度神经网络(GR-DNN)	0.95±0.22
阶梯网络(LN)	1.06±0.37
拉普拉斯阶梯网络(LLN)*	<b>0.91±0.21</b>

### 3.1.2 卷积网络

在第 2 个实验中,本文仍然随机从每类数字中挑选出 10 个样本,组成 100 个带标签样本集进行监督损失的计算,对卷积结构的拉普拉斯阶梯网络进行了测试.在此实验中,卷积拉普拉斯阶梯网络的结构描述如表 1 的第 1 列所示,其测试结果见表 3.从实验结果可以看出,在手写数字体图像的识别任务中,卷积拉普拉斯阶梯网络仍然具有比卷积阶梯网络和其他基于卷积结构的半监督深度学习模型更好的分类能力.由于卷积网络对图像信息提取处理的优势,相对于多层感知器拉普拉斯阶梯网络,卷积拉普拉斯阶梯网络对的手写数字图像识别的效果又得到了进一步的提高.

**Table 3** Performance comparison of Conv-Laplacian ladder networks and other methods on MNIST dataset

表 3 各算法与卷积拉普拉斯阶梯网络在 MNIST 数据集上的性能比较

方法	测试误差(% ,N=100)
SWWAE	9.17
EmbedCNN	7.75
卷积阶梯网络(Conv-LN)	0.89±0.50
卷积拉普拉斯阶梯网络(Conv-LLN)*	<b>0.78±0.36</b>

## 3.2 CIFAR-10

为了进一步测试拉普拉斯阶梯网络的泛化能力,我们在标准物体图像识别数据集 CIFAR-10 进行了第 3 个实验.CIFAR-10 数据集包含 10 种不同类别的共计 60 000 个样本,每个样本都是 32×32 的 RGB 彩色图片.其中包含 50 000 张带标签样本和 10 000 张测试样本.为了进行半监督学习的实验,我们随机从每类挑选出 400 个样本,组成一个 4 000 个带标签样本集作监督损失的计算,所有训练数据去除标签作非监督损失的计算.

从表 4 中可以看出,在有标签数据量  $N$  仅为 4 000 的情况下,基于卷积网络的拉普拉斯阶梯网络在物体识别数据集 CIFAR-10 上得到了最好的半监督分类结果,进一步验证了拉普拉斯阶梯网络的局部保持特性有助于提高对图像的判别能力.

**Table 4** Performance comparison of Conv-Laplacian ladder networks and other methods on CIFAR-10 dataset

表 4 各算法与卷积 LLN 在 CIFAR-10 数据集上的性能比较

方法	测试误差(% ,N=4000)
S3C	31.9
卷积阶梯网络(Conv-LN)	20.40±0.47
卷积拉普拉斯阶梯网络(Conv-LLN)*	<b>18.97±0.40</b>

## 4 结束语

本文提出了一种改进的深度表示学习方法——拉普拉斯阶梯网络,展示了如何用非监督特征学习来提高多层感知机和卷积神经网络的学习性能,从而有效解决半监督学习的问题.阶梯网络是一种优秀的半监督学习

方法,在此基础上,本文显式地引入了流形正则化框架对阶梯网络的特征学习进行约束.拉普拉斯阶梯网络不仅能够将含噪声的数据和特征重构出干净的原始数据和特征,而且显式地对网络各解码层进行了流形约束,使得网络各层具有局部保持特性和判别性,从而有效地提高了模型的泛化能力.对基于多层感知机和卷积结构的拉普拉斯阶梯网络进行了测试,在手写数字数据集 MNIST 和物体识别数据集 CIFAR-10 上进行了半监督学习实验,都取得了比阶梯网络和其他传统深度和非深度半监督学习方法更好的分类效果.实验结果表明,拉普拉斯阶梯网络是有效的半监督学习算法.但是拉普拉斯阶梯网络还有一些不足之处,如深度模型选择和超参选择困难等问题.这也是我们今后需要进一步研究的工作.

#### References:

- [1] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Proc. of the Advances in Neural Information Processing Systems. 2012. 1097–1105.
- [2] Liao YY, *et al.* Place classification with a graph regularized deep neural network. IEEE Trans. on Cognitive and Developmental Systems, 2017,9(4):304–315.
- [3] Szegedy C, Liu W, Jia Y, *et al.* Going deeper with convolutions. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 1–9.
- [4] Zhao DB, Chen YR, Lv L. Deep reinforcement learning with visual attention for vehicle classification. IEEE Trans. on Cognitive and Developmental Systems, 2017,9(4):356–367.
- [5] Graves A, Jaitly N. Towards end-to-end speech recognition with recurrent neural networks. In: Proc. of the 31st Int'l Conf. on Machine Learning (ICML 2014). 2014. 1764–1772.
- [6] Hinton G, Deng L, Yu D, *et al.* Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine, 2012,29(6):82–97.
- [7] Vinyals O, Toshev A, Bengio S, *et al.* Show and tell: A neural image caption generator. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 3156–3164.
- [8] Fang H, Gupta S, Iandola F, *et al.* From captions to visual concepts and back. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 1473–1482.
- [9] Kiros R, Salakhutdinov R, Zemel RS. Unifying visual-semantic embeddings with multimodal neural language models. arXiv preprint arXiv:1411.2539, 2014.
- [10] Cho K, Van Merriënboer B, Gulcehre C, *et al.* Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078, 2014.
- [11] Sutskever I, Vinyals O, Le QV. Sequence to sequence learning with neural networks. In: Proc. of the Advances in Neural Information Processing Systems. 2014. 3104–3112.
- [12] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society Series B (Methodological), 1977,39(1):1–22.
- [13] Blum A, Mitchell T. Combining labeled and unlabeled data with co-training. In: Proc. of the Int'l Conf. on Computational Learning Theory. 1998. 92–100.
- [14] Zhou ZH, Li M. Tri-training: Exploiting unlabeled data using three classifiers. IEEE Trans. on Knowledge and Data Engineering, 2005,17(11):1529–1541.
- [15] Belkin M, Niyogi P, Sindhvani V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. The Journal of Machine Learning Research, 2006,7(11):2399–2434.
- [16] Ding SF, Zhang N, Shi ZZ. Laplacian multi layer extreme learning machine. Ruan Jian XueBao/Journal of Software, 2017,28(10): 2599–2610 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5128.htm> [doi: 10.13328/j.cnki.jos.005128]
- [17] Joachims T. Transductive inference for text classification using support vector machines. In: Proc. of the ICML, Vol.99. 1999. 200–209.
- [18] Chen SG, Wu XJ. Improved projection twin support vector machine. Acta Electronica Sinica, 2017,45(2):408–416 (in Chinese with English abstract).
- [19] Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. Science, 2006,313(5786):504–507.

- [20] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006,18(7):1527–1554.
- [21] Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *The Journal of Machine Learning Research*, 2010,11(12):3371–3408.
- [22] Weston J, Ratle F, Mobahi H, *et al.* Deep learning via semi-supervised embedding. In: *Proc. of the Neural Networks: Tricks of the Trade*. Berlin, Heidelberg: Springer-Verlag, 2012. 639–655.
- [23] Goodfellow I, Courville A, Bengio Y. Large-scale feature learning with spike-and-slab sparse coding. arXiv preprint arXiv:1206.6407, 2012.
- [24] Kingma DP, Welling M. Auto-encoding variational Bayes. arXiv preprint arXiv:1312.6114, 2013.
- [25] Zhao J, Mathieu M, Goroshin R, *et al.* Stacked what-where auto-encoders. arXiv preprint arXiv:1506.02351, 2015.
- [26] Rasmus A, Berglund M, Honkala M, *et al.* Semi-supervised learning with ladder networks. In: *Advances in Neural Information Processing Systems*. 2015. 3546–3554.
- [27] Yang S, Li L, Wang S, *et al.* A graph regularized deep neural network for unsupervised image representation learning. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2017. 1203–1211.
- [28] Hu C, Wu XJ. Autoencoders with drop strategy. In: *Advances in Brain Inspired Cognitive Systems: Proc. of the 8th Int'l Conf. on BICS 2016*. Beijing: Springer Int'l Publishing, 2016. 80–89.
- [29] LeCun Y, Bottou L, Bengio Y, *et al.* Gradient-based learning applied to document recognition. *Proc. of the IEEE*, 1998,86(11): 2278–2324.
- [30] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images. 2009. Technical Report, TR-2009, Toronto: University of Toronto, 2009. <https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf>.

#### 附中文参考文献:

- [16] 丁世飞,张楠,史忠植.拉普拉斯多层级速学习机.软件学报,2017,28(10):2599–2610. <http://www.jos.org.cn/1000-9825/5128.htm> [doi: 10.13328/j.cnki.jos.005128]
- [18] 陈素根,吴小俊.改进的投影孪生支持向量机.电子学报,2017,45(2):408–416.



胡聪(1987—),男,湖北随州人,博士,讲师,主要研究领域为人工智能,模式识别,计算机视觉.



舒振球(1985—),男,博士,副教授,CCF 专业会员,主要研究领域为模式识别.



吴小俊(1967—),男,博士,教授,博士生导师,CCF 专业会员,主要研究领域为人工智能,模式识别,计算机视觉.



陈素根(1982—),男,博士,副教授,主要研究领域为模式识别与智能系统.