

# 融合一致性正则与流形正则的半监督深度学习算法

王杰<sup>1,2</sup>, 张松岩<sup>1,2</sup>, 梁吉业<sup>1,2</sup>

1. 山西大学计算机与信息技术学院, 山西 太原 030006;
2. 计算智能与中文信息处理教育部重点实验室, 山西 太原 030006

## 摘要

半监督学习已被广泛应用于大数据分析。目前, 基于一致性正则的方法是半监督深度学习的研究热点之一。然而这类方法没有考虑数据的流形结构, 可能会导致部分相近的样本得到差异很大的输出, 进而导致分类器性能下降。针对这个问题, 提出了一种融合一致性正则与流形正则的半监督深度学习算法。该算法在对模型施加一致性约束的同时, 对样本构图并加入平滑性损失, 实现了每个样本点局部邻域的平滑以及邻近(相连)样本点之间的平滑, 从而提高半监督深度学习算法的泛化性能。在多个图像和文本数据集上的实验结果表明, 与其他的半监督深度学习算法相比, 所提算法更有效。

## 关键词

半监督深度学习; 一致性正则; 流形正则; 平滑性约束

中图分类号: TP183

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2022027

## *A semi-supervised deep learning algorithm combining consistency regularization and manifold regularization*

WANG Jie<sup>1,2</sup>, ZHANG Songyan<sup>1,2</sup>, LIANG Jiye<sup>1,2</sup>

1. School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China
2. Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan 030006, China

## Abstract

Semi-supervised learning has been widely used in big data analysis. Currently, one of the hot research topics in semi-supervised deep learning is consistency-based methods. However, such methods do not take into account the manifold structure of the data, which may cause a portion of similar samples to get very different outputs, resulting in degraded classifier performance. To address this problem, a semi-supervised deep learning algorithm that combines consistency regularization with manifold regularization was proposed. The algorithm imposed a consistency constraint on the model while constructing a graph and adding a smoothing loss to achieve smoothing within the local neighborhood of each

sample point and between adjacent (connected) sample points, thus improving the generalization performance of the semi-supervised learning algorithm. The results on several image and text datasets show that the proposed algorithm is more effective compared with other semi-supervised deep learning algorithms.

### Key words

semi-supervised deep learning, consistency regularization, manifold regularization, smoothness constraint

## 0 引言

随着互联网和信息产业的飞速发展,人们采集与获取数据的能力大大提高,信息量以前所未有的速度增长。世界已进入大数据时代<sup>[1]</sup>,这些大数据蕴含着巨大的价值,对于社会、经济、科学等各个方面都具有重要的战略意义<sup>[2-5]</sup>,为人们更深入地感知、认识和控制物理世界提供了前所未有的丰富信息。大数据时代的到来引发了深度学习的研究热潮并取得了巨大的成功,但训练一个深度网络模型往往需要大量具有高质量标记的训练样本<sup>[6-7]</sup>。对于许多深度学习任务来说,获取大量用于训练的有标记样本的成本是极其昂贵的,且需要耗费大量的时间,与此同时,无标记样本的获取相对容易且廉价。因此,如何利用大量无标记样本来辅助提高学习方法的泛化性能,已成为一个重要研究问题。

为了应对这一问题,半监督学习(semi-supervised learning, SSL)应运而生,其目的是通过在模型训练中引入无标记样本来解决传统监督学习在训练样本不足时性能差的问题<sup>[8]</sup>。近年来,随着深度学习的兴起,半监督深度学习取得了很多显著的成果并受到越来越多的关注,其中基于一致性正则的方法<sup>[9-14]</sup>是半监督深度

学习<sup>[15]</sup>研究中的热点问题之一。一致性是指模型对扰动后训练样本的预测结果应与原预测结果保持一致。由于这类方法并不依赖于样本的真实标记,因此可以使用大量的无标记数据。一致性正则鼓励预测函数对样本的邻域具有光滑性,使得样本点局部的预测是平滑的,这种具有局部平滑性的模型更容易推广。然而,基于一致性正则的方法仅仅考虑模型对样本的邻域具有光滑性,没有考虑数据流形结构,可能会使得一部分相近的样本得到差异很大的输出,导致分类器性能下降。如图1(a)所示,尽管内外圈的样本点均局部平滑,但外围点中出现了两处低密度空白区域,这样分类面可能会位于该低密度区域,使得外圈中右侧的样本错分,造成分类性能下降。图1中红色和黑色的实心点为两类有标记样本,空心点为无标记样本,紫色的虚线圈为样本点邻域的一致性表示,蓝色、绿色和黄色的实线为可能的分类面。

为了应对上述问题,本文提出一种融合一致性正则与流形正则的半监督深度算法SmoothMatch。流形正则<sup>[16-20]</sup>是基于半监督学习中常见的流形假设,即假设数据分布在流形上,邻近的样本拥有相似的输出值。这里的邻近程度常用相似程度来刻画。流形是在局部与欧氏空间同胚的空间,换言之,它在局部具有欧氏空间的性质,能用欧氏距离进行距

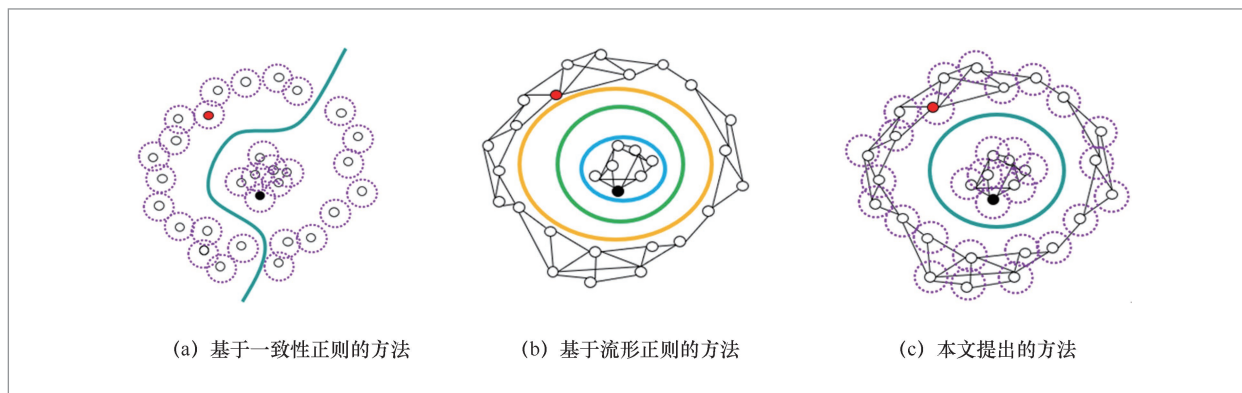


图1 基于一致性正则、流形正则的半监督学习方法以及本文提出方法的示意图

离计算。针对每个样本点，基于欧氏距离找出其邻近样本点，然后建立一个邻近连接图，图中邻近点之间存在连接，而非邻近点之间不存在连接。这类方法将有标记样本和无标记样本映射为图，以近似刻画数据的流形分布，从而可以充分利用数据中蕴含的流形结构信息。但这类方法只能保证在构造的图上保持局部相似性，即相连的样本有相似的输出。当样本扰动方向不是沿着图所近似的流形时，预测函数对扰动的样本的预测可能会发生很大的改变，即模型不能为未见到的样本点提供合理的输出。如图1(b)所示，黄色、蓝色分类面分别靠近外圈和内圈的数据点，这会导致学习器的泛化性能降低。本文提出的算法综合了两类方法的优点，如图1(c)所示，本文提出的算法不仅考虑了每个样本点的局部预测平滑，也考虑了真实的邻近（相连）样本具有平滑性，使得模型充分利用了数据的结构（流形）信息，进而可以将分类边界推向合理的低密度区域，有效地提高半监督深度学习算法的性能。在多个图像和文本标准数据集上进行实验，与相关算法/模型相比，本文所提算法SmoothMatch的性能有明显提高。

## 1 相关工作

### 1.1 半监督学习

半监督学习是近20年发展起来的一类新型机器学习方法，目前半监督分类算法可以大致分为如下几类：基于支持向量机（support vector machine, SVM）的半监督算法<sup>[21]</sup>、基于协同训练的算法<sup>[22-23]</sup>、基于生成式的算法<sup>[24-25]</sup>、基于图的半监督算法<sup>[16-20]</sup>，以及半监督深度学习算法<sup>[15]</sup>。基于一致性正则的半监督学习算法<sup>[9-14]</sup>是半监督深度学习算法中一类非常重要的学习范式。除此之外，半监督学习算法还包括半监督聚类<sup>[26]</sup>和半监督回归<sup>[27]</sup>等算法。下面着重回顾与本文相关的基于图的半监督学习算法和基于一致性正则的半监督学习算法。

### 1.2 基于图的半监督学习算法

基于图的半监督学习利用有标记和无标记样本之间的联系得到一个关于样本

空间的图结构,然后利用这个图结构将标记从有标记样本“传播”到无标记样本。如Zhu X J等人<sup>[16]</sup>提出了一种基于高斯随机场模型的半监督学习算法,该算法将有标记和无标记的数据表示为一个加权图,边上的权重表示数据之间的相似性。然后,学习问题被表述为图上的高斯随机场;Belkin M等人<sup>[17]</sup>引入图拉普拉斯正则化,将直推式图半监督学习拓展到归纳式,可以对训练集中没有出现过的样本进行分类,使得模型具备一定的通用性和泛化能力;Bai L等人<sup>[18]</sup>提出了一种新的标记传播算法,该算法将标记的成对关系作为约束条件,建立有约束的标记传播。Wang J等人<sup>[19]</sup>提出了一种基于图神经网络的半监督分类算法,该算法融合多个图神经网络的结果,在保证有标记节点分类正确的同时,利用大量无标记节点的伪标记信息最大化多个图神经网络的多样性,从而提升图半监督学习的性能。Liang J Y等人<sup>[20]</sup>提出了一种自适应构图的方法,将构图和标记推理集成到统一的优化框架中,实现二者的相互指导和动态提升,从而实现鲁棒的图半监督学习。

### 1.3 基于一致性正则的半监督学习算法

基于一致性正则的方法分为基于样本扰动的方法与基于模型扰动的方法。虽然二者在具体实现上有诸多不同,但目的都是最小化模型预测的一致性损失。

基于样本扰动的方法将原样本和扰动后的样本输入同一个模型中,然后最小化二者预测的不一致性。该类方法依赖于数据增广技术。为了产生高质量的扰动样本,近年来研究者提出了大量数据增广技术,如Miyato T等人<sup>[12]</sup>提出了虚拟对抗训练(virtual adversarial training, VAT)模型,其主要思想是找到使模型输出偏差最大的方向,然后在这个方向上对输入产生扰

动;Verma V等人<sup>[13]</sup>提出了插值一致性训练(interpolation consistency training, ICT)模型,该模型将一个样本点沿另一个样本点的方向做扰动,模型对两个样本点间插值的预测结果应与模型对两个样本点预测结果的插值一致;谷歌的Berthelot D等人<sup>[9]</sup>融合多种数据增广技术提出了MixMatch算法,达到非常低的分类错误率。

基于模型扰动的方法将训练样本输入结构相同但参数不同的两个或多个网络中,实现模型预测的一致性。其中,Laine S等人<sup>[10]</sup>提出了 $\Pi$ 模型和temporal ensembling模型。 $\Pi$ 模型将训练样本输入两个结构相同的网络,两个网络使用随机Dropout技术产生不同的网络参数,最后最小化两个网络的预测结果,从而达到一致性正则化的目的;temporal ensembling模型首先计算训练样本在前几个训练轮次(epoch)中预测的平均值,然后最小化该平均值与当前epoch的预测值,利用多个epoch的预测来实现一致性。Tarvainen A等人<sup>[11]</sup>提出了mean teacher模型,与temporal ensembling对前几轮的预测进行平均不同,mean teacher对前几轮的模型参数进行平均,并最小化该模型与当前模型的预测值,从而实现模型扰动的一致性。

然而上述基于一致性正则的方法仅仅计算样本邻域内的一致性,并没有考虑数据点之间的连接,这样可能会缺失样本数据结构中的信息。因此,笔者将基于一致性正则的方法与基于流形正则的方法结合,提出一种融合一致性正则与流形正则的半监督深度学习算法SmoothMatch。

## 2 融合一致性正则与流形正则的半监督深度学习算法

在详细介绍算法之前,首先介绍算法

中用到的部分变量: 假设数据集 $D$ 里有 $N$ 个样本, 其中 $\mathcal{L} = \{(x_i, y_i)\}_{i=1}^L$ 为有标记样本集合, 标记 $y_i \in \mathcal{Y} = \{1, 2, \dots, K\}$ , 共 $K$ 个类别;  $\mathcal{U} = \{x_i\}_{i=L+1}^N$ 为无标记样本集合。

本文提出一种融合一致性正则与流形正则的半监督深度学习算法SmoothMatch, 其不仅对样本局部区域的扰动施加平滑约束, 同时考虑了样本点之间的结构信息。算法的总体损失函数如下:

$$\ell = \ell_e + \lambda_c \ell_c + \lambda_s \ell_s \quad (1)$$

总体损失主要由3项构成: ①对于有标记样本, 比较模型的预测结果与样本的真实标记, 计算交叉熵损失 $\ell_e$ ; ②对于无标记样本, 采用数据增广技术 $\text{Augment}(x_u)$ 计算一致性损失 $\ell_c$ ; ③从有标记样本和无标记样本中抽取样本, 将这些样本进行特征空间映射并构图, 最后计算平滑性损失 $\ell_s$ 。 $\lambda_c$ 与 $\lambda_s$ 为防止某一项损失过大或过小而平衡3项损失的权值参数。算法整体框架如图2所示。

## 2.1 样本邻域内的一致性损失

本节使用ICT模型<sup>[13]</sup>中的数据增广方法Mixup来计算一致性损失。Mixup的基本计算式如下:

$$\text{Mix}_\lambda(a, b) = \lambda a + (1 - \lambda)b \quad (2)$$

其中,  $\lambda$ 为服从 $\beta$ 分布的权值参数,  $\text{Mix}_\lambda(a, b)$ 为 $a$ 和 $b$ 之间的插值。

给定一个小批量(mini-batch)数据集中的任意两个样本点 $x_i$ 与 $x_j$ 以及模型预测结果 $f(x_i, \theta)$ 和 $f(x_j, \theta)$ , 根据式(2)可以得到这两个样本点间的插值 $\hat{x} = \text{Mix}_\lambda(x_i, x_j)$ , 则模型对该插值的预测结果为 $f(\hat{x}, \theta) = f(\text{Mix}_\lambda(x_i, x_j), \theta)$ , 同时可以得到模型对样本点 $x_i$ 与 $x_j$ 预测结果的插值 $\hat{y} = \text{Mix}_\lambda(f(x_i, \theta), f(x_j, \theta))$ , 一致性损失便是要求 $f(\hat{x}, \theta)$ 与 $\hat{y}$ 保持一致。因此, 在一个mini-batch数据集内的一致性损失为:

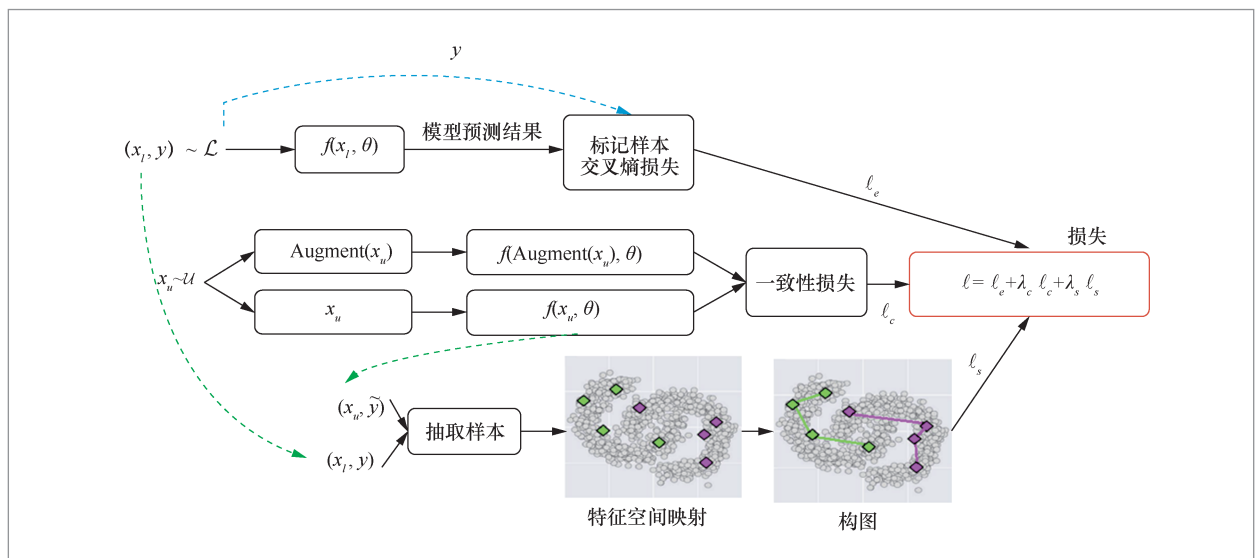


图2 算法整体框架

$$\ell_c(\theta, \mathcal{L}, \mathcal{U}) = \frac{2}{|B||B-1|} \sum_{x_i, x_j \in B} d(f(\hat{x}, \theta), \hat{y}) \quad (3)$$

其中,  $d(\cdot, \cdot)$  用于度量分类器对插值样本的标记与标记插值之间的预测结果差异, 本文采用深度学习中非常常见的交叉熵损失  $d(f(\hat{x}, \theta), \hat{y}) = -\log[f(\hat{x}, \theta)]\hat{y}$  来度量  $f(\hat{x}, \theta)$  与  $\hat{y}$  的一致性,  $B$  为一个 mini-batch 数据集。

## 2.2 样本间的平滑性损失

在大数据环境下, 刻画样本间平滑性损失的流形正则项面临如下挑战。①大多数现有的构图方法是对输入空间的距离度量, 但该方法有很大的局限性。例如对于图像样本, 其输入是多通道像素值, 然而像素距离并不能很好地反映样本间的语义相似性。②传统基于图的算法计算整个数据集的邻接矩阵并在此基础上构图, 耗费的时间过长、空间复杂度过高。③传统方法构建的是静态固定的图, 因此不能利用分类器提取的知识进行图的动态调整。

为了解决这些问题, 本文提出在 mini-batch 数据集内构图并计算平滑性损失的方法。在样本的标记空间计算样本相似度, 并采用动态构图的方式, 随着学习不断更新图结构, 从而更好地指导学习器学习样本间的平滑。

### (1) 构图与邻接矩阵的计算

在每一个 mini-batch 数据集内, 用其中的数据构造  $K$  近邻 ( $K$ -nearest neighbor,  $KNN$ ) 图, 与传统构图不同的是, 这里使用样本的标记空间度量样本间的相似度, 权值矩阵计算如下:

$$W_{ij} = \exp\left(-\frac{\|f(x_i), f(x_j)\|_2^2}{2\delta^2}\right) \quad (4)$$

其中,  $\delta > 0$  是指定的高斯函数带宽参数,  $f(x_i)$  是模型对样本的预测。

### (2) 平滑性损失计算

给定邻接矩阵  $W$  与样本特征表达后, 平滑性损失如下:

$$\ell_s(\theta, \mathcal{L}, \mathcal{U}, W) = \frac{1}{2} \sum_{x_i, x_j \in B} W_{ij} \|h(x_i) - h(x_j)\|^2 \quad (5)$$

其中,  $h: \mathcal{X} \rightarrow \mathbb{R}^p$  为输入空间到网络倒数第二层的映射。

## 2.3 算法实现细节

SmoothMatch 的整体损失函数如下:

$$\begin{aligned} \ell = & -\frac{1}{|B \cap \mathcal{L}|} \sum_{x_i \in (B \cap \mathcal{L})} \log[f(x_i; \theta)]y_i + \\ & w_c(t)\lambda_c \frac{2}{|B||B-1|} \sum_{x_i, x_j \in B} d(f(\hat{x}, \theta), \hat{y}) + \\ & w_s(t)\lambda_s \frac{1}{2} \sum_{x_i, x_j \in B} W_{ij} \|h(x_i) - h(x_j)\|^2 \end{aligned} \quad (6)$$

其中,  $w_c(t)$  和  $w_s(t)$  是随 epoch 迭代线性上升的权值函数。算法流程如下。

**输入:** 有标记样本集合  $\mathcal{L}$ , 无标记样本集合  $\mathcal{U}$ , 随 epoch 迭代线性上升的权值函数  $w_c(t)$  和  $w_s(t)$ , 模型  $f(x; \theta)$ , 损失平衡项  $\lambda_c$  与  $\lambda_s$ , 模型迭代次数 numepochs

**输出:** 更新后的模型参数  $\theta$

- 1: 初始化: 随机初始化模型参数  $\theta$
- 2: repeat
- 3: for each mini-batch  $B$  Do
- 4: 计算有标记样本的交叉熵损失

- 5: 由式(3)计算 mini-batch 内数据的一致性损失
- 6: 由式(4)计算权值矩阵  $W$
- 7: 由式(5)计算 mini-batch 内数据的平滑性损失
- 8: 由式(6)得到 3 项损失的加权求和, 根据损失更新模型参数
- 9: end for
- 10: until 迭代次数 > numepochs

### 3 实验设计与分析

为了测试 SmoothMatch 算法的有效性, 在 3 个图像数据集以及两个英文文本数据集进行实验, 图像数据集分别为 CIFAR-10、CIFAR-100 和 SVHN, 文本数据集分别为 IMDB 和 Yahoo! Answers。数据集介绍见表 1。

本文数据遵循参考文献[28]的划分方法, 对于图像数据集, CIFAR-10 与 CIFAR-100 分别包括 45 000 个训练样本、5 000 个验证样本和 10 000 个测试样本, SVHN 数据集包括 65 932 个训练样本、7 325 个验证样本和 26 032 个测试样本; 对于文本数据集, IMDB 数据

集包括 63 000 个训练样本、7 000 个验证样本和 25 000 个测试样本, Yahoo! Answers 数据集包括 45 000 个训练样本、5 000 个验证样本和 60 000 个测试样本。为了测试算法在半监督学习环境下的性能, 标准做法是将大部分训练样本视为无标记数据, 只随机抽取并使用小部分有标记数据。

#### 3.1 基线方法

对于图像数据集, 对比  $\Pi$  模型<sup>[10]</sup>、mean teacher 模型<sup>[11]</sup>、VAT 模型<sup>[12]</sup>、MixMatch 算法<sup>[9]</sup>以及本文的 SmoothMatch 算法的实验结果。为了确保对比实验一致, 实验使用 Wide ResNet-28 模型, 模型结构与详细说明参照参考文献[28], 学习率衰减值为 0.999, 权值衰减值为 0.02。

对于文本数据集, 对比 Xie Q Z 等人<sup>[14]</sup>提出的一致性算法 UDA, 以及预训练的 BERT 模型<sup>[29]</sup>。对于英文文本的数据增广, 使用德语作为中间语言的回译过程, 即将一个英文样本先翻译为德语, 再翻译回英文样本。

关于超参数  $\lambda_c$  与  $\lambda_s$ , 根据参考文献[28]的建议将 CIFAR-10、CIFAR-100 和 SVHN 的  $\lambda_c$  分别固定为 75、150 和 250, 并将 IMDB 和 Yahoo! Answers 的  $\lambda_c$  固定为 100。各个数据集的  $\lambda_s$  从集合  $\{0, \lambda_c / 1000, \lambda_c / 100, \lambda_c / 10, \lambda_c / 3, 2\lambda_c / 3, \lambda_c\}$

表 1 数据集介绍

| 数据集            | 训练集    | 验证集   | 测试集    | 特征维度    | 类别  |
|----------------|--------|-------|--------|---------|-----|
| CIFAR-10       | 45 000 | 5 000 | 10 000 | 32×32×3 | 10  |
| CIFAR-100      | 45 000 | 5 000 | 10 000 | 32×32×3 | 100 |
| SVHN           | 65 932 | 7 325 | 26 032 | 32×32×3 | 10  |
| IMDB           | 63 000 | 7 000 | 25 000 | 128     | 2   |
| Yahoo! Answers | 45 000 | 5 000 | 60 000 | 128     | 10  |

中遍历取值,并用验证集交叉验证取得最优值。模型的迭代次数由验证集损失确定,即当验证集的损失在一定迭代次数(本文为50次)内变化不大时,停止模型的迭代。

### 3.2 实验结果与分析

对于CIFAR-10和SVHN数据集,在250、500和1 000这3个不同数量标记样本上评估5种算法/模型的错误率,结果见表2和表3。对于CIFAR-100数据集,使用10 000个有标记样本对5种算法模型进行实验,结果见表4。

表2 5种算法/模型在CIFAR-10数据集上不同标记样本数下的错误率

| 算法/模型        | 250个   | 500个   | 1 000个 |
|--------------|--------|--------|--------|
| II           | 53.02% | 41.82% | 31.53% |
| mean teacher | 47.32% | 42.01% | 17.32% |
| VAT          | 36.03% | 26.11% | 18.68% |
| MixMatch     | 17.60% | 13.77% | 11.16% |
| SmoothMatch  | 14.40% | 12.99% | 10.22% |

表3 5种算法/模型在SVHN数据集上不同标记样本数下的错误率

| 算法/模型        | 250个   | 500个   | 1 000个 |
|--------------|--------|--------|--------|
| II           | 17.65% | 11.44% | 8.60%  |
| mean teacher | 5.85%  | 5.45%  | 5.21%  |
| VAT          | 8.41%  | 7.44%  | 5.98%  |
| MixMatch     | 5.58%  | 5.46%  | 4.45%  |
| SmoothMatch  | 5.11%  | 5.04%  | 4.23%  |

表4 5种算法/模型在CIFAR-100数据集10 000个标记样本下的错误率

| 算法/模型        | 错误率    |
|--------------|--------|
| II           | 39.19% |
| mean teacher | 37.17% |
| VAT          | 35.42% |
| MixMatch     | 32.88% |
| SmoothMatch  | 32.23% |

由表2~表4可以得到如下结论。

①SmoothMatch算法在3个图像数据集的各种标记样本数量下的准确性优于其他方法,例如,在CIFAR-10数据集上,在仅有250个有标记样本下SmoothMatch算法的错误率为14.40%,而相同条件下的MixMatch算法错误率为17.60%;在SVHN数据集上,当有250个有标记样本时,SmoothMatch明显优于II模型,这说明了融合一致性正则和流形正则的有效性。②随着有标记样本数量的增加,上述几种方法的错误率均降低,这是因为有标记样本越多,所提供的监督信息越多,分类器能更好地拟合数据。特别地,在CIFAR-10数据集上,mean teacher模型的错误率有明显的降低,这说明该方法对有标记数据的依赖性很强,而本文所提算法由于可以充分利用数据的流形结构,能够使相似(相连)的样本有相似的输出,从而降低了对有标记样本的依赖性,可以在有标记样本较少的情况下达到不错的效果。

此外,笔者还验证了SmoothMatch算法在文本分类任务上的表现。IMDB和Yahoo! Answer数据集上的实验结果见表5和表6。

由表5和表6可以得到,SmoothMatch算法在两个文本数据集上均优于UDA算法和BERT模型。特别地,在IMDB数据集上,SmoothMatch算法在有标记样本数为20时的错误率为12.27%,明显优于BERT在有标记样本数为100时的结果,这说明了所提算法的优越性。

图像数据集和文本数据集上的实验结果表明,融合一致性正则和流形正则的方法在考虑样本局部预测平滑的同时,充分利用了数据的流形结构,使得相似的样本有相似的输出,提高了模型的泛化性能。

### 3.3 参数分析

为了进一步分析一致性正则和流形正则对模型的影响,在3个图像数据集上对比了SmoothMatch算法在不同 $\lambda_s/\lambda_c$ 比值下的错误率。3个数据集上的实验结果如图3所示。

可以看出,当 $\lambda_s/\lambda_c$ 为0,即只利用一致性损失优化目标函数时,其错误率均处于较高水平。但随着平滑性损失权重 $\lambda_s$ 的提高,错误率逐渐降低,到 $\lambda_s/\lambda_c$ 为0.1时到达最优。由此可见,平滑性损失的加入使得模型对同一类样本的低维表示更集中,

表5 3种算法/模型在IMDB数据集上不同标记样本数下的错误率

| 算法/模型       | 20个    | 100个   |
|-------------|--------|--------|
| BERT        | 32.50% | 19.59% |
| UDA         | 13.70% | 9.50%  |
| SmoothMatch | 12.27% | 8.96%  |

表6 3种算法/模型在Yahoo! Answer数据集上不同标记样本数下的错误率

| 算法/模型       | 100个   | 500个   |
|-------------|--------|--------|
| BERT        | 48.90% | 37.56% |
| UDA         | 41.55% | 33.83% |
| SmoothMatch | 40.53% | 32.78% |

模型的鲁棒性。而若继续加大 $\lambda_s$ ,则弱化了模型的一致性正则,使得样本点局部邻

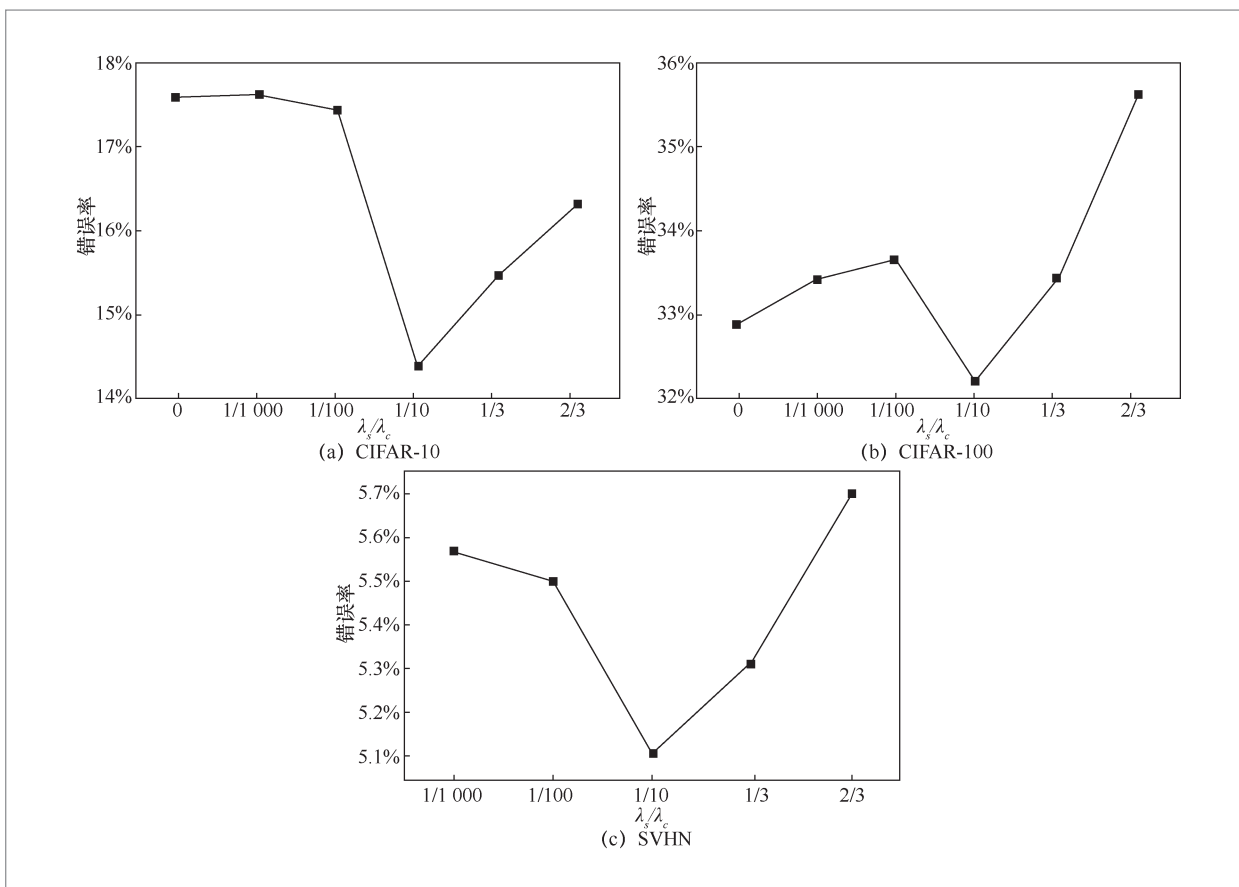


图3 3个图像数据集上SmoothMatch算法在不同 $\lambda_s/\lambda_c$ 比值下的错误率

域的预测不平滑,导致错误率逐渐提高,模型的性能降低。这说明了一致性正则与流形正则的合理结合确实能够提高算法的性能。

## 4 结束语

本文针对基于一致性正则的半监督深度学习算法可能会使得一部分相近的样本得到差异很大的输出,进而导致学习器性能退化的问题,提出了一种融合一致性正则与流形正则的半监督深度学习算法。该算法在对模型施加一致性约束的同时,对样本构图并加入平滑性损失,实现了每个样本点局部邻域的平滑以及样本点之间的平滑,从而提高半监督学习算法的泛化性能。在多个图像和文本数据集上的实验结果表明,融合一致性正则与流形正则的半监督深度学习算法获得了更优的性能。

## 参考文献:

- [1] 李国杰. 对大数据的再认识[J]. 大数据, 2015, 1(1): 1-9.  
LI G J. Further understanding of big data[J]. Big Data Research, 2015, 1(1): 1-9.
- [2] 梁吉业, 冯晨娇, 宋鹏. 大数据相关分析综述[J]. 计算机学报, 2016, 39(1): 1-18.  
LIANG J Y, FENG C J, SONG P. A survey on correlation analysis of big data[J]. Chinese Journal of Computers, 2016, 39(1): 1-18.
- [3] 胡湘红, 彭衡, 杨灿, 等. 基因大数据的集成分析[J]. 大数据, 2019, 5(4): 67-88.  
HU X H, PENG H, YANG C, et al. Integrative analysis for big data in genomics[J]. Big Data Research, 2019, 5(4): 67-88.
- [4] 杨孟辉, 杜小勇. 政府大数据治理: 政府管理的新形态[J]. 大数据, 2020, 6(2): 3-18.  
YANG M H, DU X Y. Big data governance in governments: a new form of the government administration[J]. Big Data Research, 2020, 6(2): 3-18.
- [5] 刘晓波, 蒋阳升, 唐优华. 综合交通大数据应用技术创新平台[J]. 大数据, 2018, 4(6): 78-84.  
LIU X B, JIANG Y S, TANG Y H. Innovation platform of integrated transportation big data application technology[J]. Big Data Research, 2018, 4(6): 78-84.
- [6] CAO F L, YAO K X, LIANG J Y. Deconvolutional neural network for image super-resolution[J]. Neural Networks, 2020, 132: 394-404.
- [7] YAO K X, CAO F L, LEUNG Y, et al. Deep neural network compression through interpretability-based filter pruning[J]. Pattern Recognition, 2021, 119: 108056.
- [8] 刘建伟, 刘媛, 罗雄麟. 半监督学习方法[J]. 计算机学报, 2015, 38(8): 1592-1617.  
LIU J W, LIU Y, LUO X L. Semi-supervised learning methods[J]. Chinese Journal of Computers, 2015, 38(8): 1592-1617.
- [9] BERTHELOT D, CARLINI N, GOODFELLOW I, et al. MixMatch: a holistic approach to semi-supervised learning[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Cambridge: The MIT Press, 2019: 5049-5059.
- [10] LAINE S, AILA T M. Temporal ensembling for semi-supervised learning[C]//Proceedings of the 5th International Conference on Learning Representations. [S.l.:s.n.], 2017.
- [11] TARVAINEN A, VALPOLA H. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results[C]//Proceedings of the 31st Conference on Neural Information Processing Systems. Cambridge: The MIT Press, 2017: 1195-1204.
- [12] MIYATO T, MAEDA S I, KOYAMA M, et al. Virtual adversarial training: a

- regularization method for supervised and semi-supervised learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(8): 1979–1993.
- [13] VERMA V, KANNALA J, LAMB A, et al. Interpolation consistency training for semi-supervised learning[J]. Neural Networks, 2022, 145: 90–106.
- [14] XIE Q Z, DAI Z H, HOVY E, et al. Unsupervised data augmentation for consistency training[C]//Proceedings of the 30th Advances in Neural Information Processing Systems. Cambridge: The MIT Press, 2020: 6256–6268.
- [15] VAN ENGELEN J E, HOOS H H. A survey on semi-supervised learning[J]. Machine Learning, 2020, 109(2): 373–440.
- [16] ZHU X J, GHAHRAMANI Z, LAFFERTY J D. Semi-supervised learning using gaussian fields and harmonic functions[C]//Proceedings of the 20th International Conference on Machine Learning. Palo Alto: AAAI Press, 2003: 912–919.
- [17] BELKIN M, NIYOGI P, SINDHWANI V. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples[J]. Journal of Machine Learning Research, 2006, 7: 2399–2434.
- [18] BAI L, WANG J B, LIANG J Y, et al. New label propagation algorithm with pairwise constraints[J]. Pattern Recognition, 2020, 106: 107411.
- [19] WANG J, LIANG J Q, CUI J B, et al. Semi-supervised learning with mixed-order graph convolutional networks[J]. Information Sciences, 2021, 573: 171–181.
- [20] LIANG J Y, CUI J B, WANG J, et al. Graph-based semi-supervised learning via improving the quality of the graph dynamically[J]. Machine Learning, 2021, 110(6): 1345–1388.
- [21] JOACHIMS T. Transductive inference for text classification using support vector machines[C]//Proceedings of the 16th International Conference on Machine Learning. Bled: Morgan Kaufmann, 1999: 200–209.
- [22] BLUM A, MITCHELL T. Combining labeled and unlabeled data with co-training[C]//Proceedings of the 11th Annual Conference on Computational Learning Theory. New York: ACM Press, 1998: 92–100.
- [23] WANG W, ZHOU Z H. A new analysis of co-training[C]//Proceedings of the 27th International Conference on Machine Learning. Haifa: Omnipress, 2010: 1135–1142.
- [24] LOOG M. Contrastive pessimistic likelihood estimation for semi-supervised classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(3): 462–475.
- [25] SHAHSHAHANI B M, LANDGREBE D A. The effect of unlabeled samples in reducing the small sample size problem and mitigating the Hughes phenomenon[J]. IEEE Transactions on Geoscience and Remote Sensing, 1994, 32(5): 1087–1095.
- [26] BAI L, LIANG J Y, CAO F Y. Semi-supervised clustering with constraints of different types from multiple information sources[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(9): 3247–3258.
- [27] LI Y F, ZHA H W, ZHOU Z H. Learning safe prediction for semi-supervised regression[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2017: 2217–2223.
- [28] OLIVER A, ODENA A, RAFFEL C, et al. Realistic evaluation of deep semi-supervised learning algorithms[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Cambridge: The MIT Press, 2018: 3239–3250.
- [29] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep

bidirectional transformers for language understanding[C]//Proceedings of the 13th Conference of the North American Chapter of the Association

for Computational Linguistics: Human Language Technologies. Minneapolis: Association for Computational Linguistics, 2019: 4171-4186.

#### 作者简介



王杰(1990-),男,山西大学计算机与信息技术学院博士生,主要研究方向为数据挖掘和机器学习。



张松岩(1996-),男,山西大学计算机与信息技术学院硕士生,主要研究方向为数据挖掘和机器学习。



梁吉业(1962-),男,博士,山西大学计算机与信息技术学院教授,中国计算机学会会士,主要研究方向为数据挖掘与机器学习,在国内外重要期刊和会议上发表学术论文200余篇。

收稿日期: 2021-08-11

通信作者: 梁吉业, ljiy@sxu.edu.cn

基金项目: 国家自然科学基金资助项目(No.61976184); 山西省重点研发计划项目(No.201903D121162); 山西省1331工程项目

**Foundation Items:** The National Natural Science Foundation of China (No.61976184), The Key Project of Research and Development Plan of Shanxi Province (No.201903D121162), The 1331 Engineering Project of Shanxi Province