

基于混合样本自动数据增强技术的半监督学习方法

许华杰^{1,2} 陈育¹ 杨洋¹ 秦远卓³

1 广西大学计算机与电子信息学院 南宁 530004

2 广西多媒体通信与网络技术重点实验室 南宁 530004

3 广西大学土木建筑工程学院 南宁 530004

(hjxu2009@163.com)

摘要 基于一致性的半监督学习方法通常使用简单的数据增强方法来实现对原始输入和扰动输入的一致性预测。在有标签数据的比例较低的情况下,该方法的效果难以得到保证。将监督学习中一些先进的数据增强方法扩展到半监督学习环境中,是解决该问题的思路之一。基于一致性的半监督学习方法 MixMatch,提出了基于混合样本自动数据增强技术的半监督学习方法 AutoMixMatch,在数据增强阶段采用自动数据增强技术,并在样本混合阶段提出了一种混合样本算法,用于提升对无标签样本的利用效果。通过图像分类方面的实验来测试所提方法的性能,在图像分类基准数据集中,所提方法在3种有标签样本比例下的分类效果均优于对比的几个主流半监督分类方法,验证了所提方法的有效性。此外,所提方法在有标签数据占训练数据比例极低(仅为0.05%)的情况下表现更好,在SVHN数据集上的实验结果表明,所提方法的分类错误率比 MixMatch 低30.17%。

关键词: 半监督学习;一致性;图像分类;自动数据增强;混合样本

中图法分类号 TP391

Semi-supervised Learning Method Based on Automated Mixed Sample Data Augmentation Techniques

XU Hua-jie^{1,2}, CHEN Yu¹, YANG Yang¹ and QIN Yuan-zhuo³

1 College of Computer and Electronic Information, Guangxi University, Nanning 530004, China

2 Guangxi Key Laboratory of Multimedia Communications and Network Technology, Nanning 530004, China

3 College of Civil Engineering and Architecture, Guangxi University, Nanning 530004, China

Abstract Consistency-based semi-supervised learning methods typically use simple data augmentation methods to achieve consistent predictions for both original inputs and perturbed inputs. The effectiveness of this approach is difficult to be guaranteed when the proportion of labeled data is relatively low. Extending some advanced data augmentation method in supervised learning to be used in a semi-supervised learning setting is one of the ideas to solve this problem. Based on the consistency-based semi-supervised learning method MixMatch, a semi-supervised learning method AutoMixMatch based on automated mixed sample data augmentation techniques is proposed, which uses a modified automatic data augmentation technique in the data augmentation phase, and a mixed-sample algorithm is proposed to enhance the utilization of unlabeled samples in the sample mixing phase. The performance of the proposed method is evaluated through image classification experiments. In image classification benchmark datasets, the proposed method outperforms several mainstream semi-supervised classification methods in three labeled sample proportions, which validates the effectiveness of the method. In addition, the proposed method performs better with a very low proportion of labeled data to the training data (only 0.05%), and the classification error rate of the proposed method on the SVHN dataset is 30.17% lower than that of MixMatch.

Keywords Semi-supervised learning, Consistency, Image classification, Automated data augmentation, Mixed sample

到稿日期:2021-01-20 返修日期:2021-07-07

基金项目:广西壮族自治区科技计划项目(2017AB15008);崇左市科技计划项目(FB2018001)

This work was supported by the Science and Technology Plan Project of Guangxi Zhuang Autonomous Region(2017AB15008) and Science and Technology Plan Project of Chongzuo(FB2018001).

通信作者:秦远卓(qinyuanzhuo@st.gxu.edu.cn)

1 引言

在海量的有标签数据的帮助下,深度学习在监督学习任务中取得了很大的进步,但是,一方面有标签数据很难获得,另一方面,对收集到的数据进行标注需要花费大量的人力和计算资源。半监督学习(Semi-Supervised Learning, SSL)^[1]旨在解决监督学习的弱点,在有标签数据有限的条件下使用无标签的数据来提高模型的性能。

半监督学习中,基于一致性的方法假设对无标签的数据施加扰动,模型预测不会发生显著变化。这种做法不需要已知数据的标签,因此可以很好地被应用到无标签数据上。Laine 等^[2]提出的 Π 模型利用 Dropout 进行扰动,以实现基于一致性的半监督学习。Tarvainen 等^[3]提出的 Mean Teacher 使用水平翻转、高斯噪声等方法作为输入数据的扰动方法。上述半监督方法在一定程度上利用了无标签数据,但是随着有标签数据的减少,这种采用简单数据增强或正则化技术的方法的效果往往也会出现大幅度下滑。最近,在监督学习中出现了许多强大的数据增强方法,可以产生多样化、高质量的数据,在图像分类及分割任务中取得了较好的效果。一些研究者开始尝试将这些数据增强方法扩展至半监督的环境中。Verma 等^[4]引入了混合样本数据增强技术 MixUp^[5]对无标签样本进行混合,让模型学习对无标签数据的一致性预测;Xie 等^[6]引入了 RandAugment^[7]对无标签数据进行数据增强,在多个图像及文本分类任务中提高了半监督学习的性能。MixMatch^[8]是一种基于一致性的半监督学习方法,其分为数据增强、标签预测和样本混合 3 个步骤,其在数据增强阶段使用随机水平翻转和裁剪,在标签预测阶段对多次增强后的无标签样本进行预测,取多次结果的平均值,从而得到预测标签,并在样本混合阶段使用 MixUp 方法进行数据混合。

针对目前基于一致性的方法面临的问题以及数据增强技术的发展趋势,本文探讨了对监督学习领域先进的数据增强技术的改进并在半监督学习环境中对其进行了应用。本文提出将用于监督学习领域的自动数据增强方法 AutoAugment (Automated Data Augmentation)^[9]进行改进,并将其应用到 MixMatch 中的数据增强阶段中;之后在混合样本阶段提出了一种基于 CutMix^[10]和 MixUp 的混合样本算法 MixAug (Mixed Sample Data Augmentation),该算法结合两种算法的优点来提升样本混合的效果。

2 相关工作

2.1 基于一致性的半监督学习方法

基于一致性的半监督方法主要用于分类任务,灵感来自聚类假设,其利用无标签数据来强制训练的模型符合聚类假设^[11],即学习到的决策边界必须位于低密度区域。这种方法基于一个简单的事实,即如果对一个无标签的样本施加一个扰动,预测结果不会有很大的变化,因此所有具有相同标签的数据点将集中在密集的数据区域。对数据进行扰动的操作主要有两种:一种是直接将噪声用于无标签的数据,另一种是通过扰动分类器本身,将噪声隐式地应用于数据。Mean Teacher^[3]提出了一种模型参数空间的一致性方法,其目标是在

不需要额外训练的情况下,从学生模型中形成一个更好的教师模型。Fast-SWA^[12]沿用了这一思想,其使用随机权重平均^[13]来改进 SGD。VAT^[14]探索能最大化模型预测变化的扰动。MixMatch^[8]将 3 类(熵最小化、一致性损失和传统正则化)损失函数统一为单一的损失函数,并在图像分类任务中取得了较好的效果。

2.2 监督学习中的数据增强方法

近年来,数据增强在计算机视觉、自然语言处理等领域的应用取得了重大进展。许多先进的数据增强方法可以生成多样化、高质量的图像,其中自动数据增强(AutoAugment)方法以及混合样本数据增强(MixAug)方法在许多图像分类任务上取得了较好的效果。自动数据增强使用自动搜索的方式搜索数据增强策略。AutoAugment^[9]基于强化学习进行自动搜索,它能针对不同的任务自动搜索合适的数据增强策略。RandAugment^[7]去除了单独的搜索阶段,可直接在任何数据集使用,无需根据模型或数据集大小来调整强度。

混合样本数据增强根据一定策略将两个数据样本组合,创建一个混合后的数据样本,这种混合方式也包括样本的标签。MixUp^[5]以一定的比例混合两个样本,生成一个混合样本和对应的混合标签,这种方法混合的图像往往不自然且模糊不清,而 CutMix^[10]弥补了 MixUp 的缺点,通过将一个样本中的部分区域删除,然后把删除的区域覆盖至另一个样本中,从而产生新的混合样本。French 等^[15]在半监督语义分割中应用了 CutMix,但对于半监督分类来说,这种策略的效果尚未得到探讨。

3 基于混合样本自动数据增强技术的半监督学习方法

本文提出的基于混合样本自动数据增强技术的半监督学习方法 AutoMixMatch 是基于 MixMatch 改进而来的,包括数据增强、标签预测和样本混合 3 个步骤。AutoMixMatch 算法的流程如算法 1 所示。

算法 1 AutoMixMatch

```
Input: labeled samples  $x_l$  with their one-hot label  $l$ , unlabeled samples  $x_u$ , sharpening temperature  $s$ , number of augmentations  $k$ 
1.  $\hat{x}_1 = \text{Augment}(x_l)$ 
2. for  $i=1, \dots, k-1$  do
3.  $\hat{x}_{u,i} = \text{Augment}(x_u)$ 
4. end for
5.  $g = \text{Average}(P(y|\{x_u, \hat{x}_{u,i}\}; \theta))$ 
6.  $g = \text{Sharpen}(g, s)$ 
7.  $LS = \{(\hat{x}_1, l)\}$ 
8.  $UnS = \{(\{x_u, \hat{x}_{u,i}\}, g); i \in (1, \dots, k-1)\}$ 
9.  $LS', UnS' = \text{MixAug}(LS, UnS)$ 
10. return  $LS', UnS'$ 
```

给定一组有标签样本(包括其对应的标签)和一组无标签样本,AutoMixMatch 可产生经过处理的有标签样本 LS' 以及经过处理的无标签样本 UnS' (包括其预测标签),其算法流程可分为以下 3 个步骤。

(1)数据增强,对应于算法 1 中的第 1 行至第 4 行。记 k

为数据增强的次数,对有标签样本 x_l 和无标签样本 x_u 分别进行1次和 $k-1$ 次数据增强,得到增强后的有标签样本 \hat{x}_l 和增强后的无标签样本 $\hat{x}_{u,i}$ 。与 MixMatch 只使用常规数据增强方法相比,AutoMixMatch 改进并使用了监督学习领域较先进的自动数据增强方法 AutoAugment。对有标签样本进行数据增强是监督学习中常用的做法,用于增强数据的多样性,而对无标签样本进行多次数据增强是为了后续计算无标签样本的预测标签。

(2) 标签预测,对应于算法1中的第5行至第6行。使用神经网络预测步骤1得到的 $k-1$ 个增强后的无标签样本和原始的无标签样本,得到 k 个预测分布。对于得到的所有无标签样本,模型预测结果应保持一致,因此可以计算这些样本预测结果的平均值,以降低预测结果的不确定性。 k 个预测分布的平均值的计算方式如式(1)所示:

$$Average(P) := \frac{1}{k} \sum_{i=1}^k P(y | \{x_u, \hat{x}_{u,i}\}; \theta) \quad (1)$$

其中,模型 P 代表对一个参数为 θ 的输入产生一个类标签分布 y 。之后对无标签样本的平均预测分布使用锐化函数^[8],以减小平均预测分布的熵,即减小平均预测分布的不确定程度。经过这一步,得到无标签样本的预测标签 g ,将其作为无标签样本的标签。为无标签样本生成预测标签是基于一致性的半监督方法的常用做法。

(3) 样本混合,对应于算法1中的第7行至第9行。将有标签样本及其标签 LS 和无标签样本及其预测标签 UnS 组成一个集合,将集合打乱后随机抽取样本,通过本文提出的混合样本算法 MixAug 混合有标签样本和无标签样本。在模型训练过程中,通过有标签样本和无标签样本的混合,模型可以从无标签样本中学习到有标签样本的信息,因为有标签样本的标签是确定的,所以可以通过样本混合提高对无标签样本的标签预测的准确性。

输入数据经过3个步骤后,使用式(2)所示的损失函数来计算模型的损失:

$$Loss = Loss_l + \lambda Loss_{u_l} \quad (2)$$

其中, $Loss_l$ 为有标签样本的标签和模型预测之间的损失, $Loss_{u_l}$ 为无标签样本的预测标签和模型预测之间的损失, λ 为无标签样本的损失权重。通过梯度下降法最小化损失函数 $Loss$,不断迭代直至模型收敛。初始时,神经网络对训练样本的学习不充分,对无标签样本的预测标签的计算结果可能不准确,这时权重 λ 的值较低,避免了在预测时有较大偏差的无标签样本的干扰。随着训练的进行,逐渐调高权重 λ ,以充分利用无标签样本。

下文重点对 AutoMixMatch 中的核心内容,即步骤(1)“数据增强”中改进的自动数据增强方法和步骤(3)“样本混合”中提出的混合样本算法 MixAug 进行叙述。

3.1 自动数据增强

因为自动数据增强方法 AutoAugment 应用于监督学习领域在许多基准图像分类数据集中取得了较好的效果且易于实现,所以本文使用 AutoAugment 对有标签样本和无标签样本进行增强,其增强效果如图1所示(由于所采用的数据集中图片的点阵只有 32×32 ,因此视觉上分辨率有限)。Auto-

Augment 对每个样本执行一个策略,每个策略包括两个与图像有关的数据增强操作。每个操作附带两个参数,一个是幅度范围参数,另一个是概率参数。幅度范围参数用于量化数据增强操作的幅度,如图像对比度的幅度、图像反转的幅度等,概率参数给出应用此数据增强操作的概率。

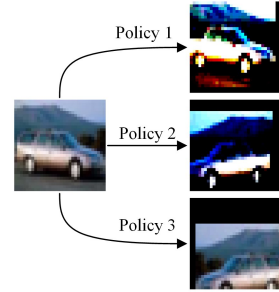


图1 AutoAugment 增强后的结果

Fig. 1 Augmented results by AutoAugment

MixMatch 对无标签样本进行 k 次数据增强,得到 k 个无标签样本,然后用它们来生成预测标签^[8]。本文方法在应用 AutoAugment 的同时对该步骤做了改进,因为 AutoAugment 对每幅图片都执行不同的增强策略,有较强的随机性,如果进行预测的无标签样本全部都是经过 AutoAugment 数据增强后的样本,则最终的预测分布可能会偏离无标签样本的实际的类别分布,进而影响生成的预测标签的准确性。为了使得到的结果更稳定,减小预测偏差,对无标签样本进行 $k-1$ 次数据增强,同时保留1组原始的无标签数据。在实验中发现,当 $k=2$ 时,训练模型的效果会更好。无标签样本的损失函数如式(3)所示:

$$Loss_{u_l} = \sum_{x, g \in UnS} H(g, P(y | x; \theta)) \quad (3)$$

其中, H 为距离函数,代表两个分布之间的距离,参考 MixMatch 的做法,本文也选用 L2 损失作为无标签样本损失的距离函数。

3.2 混合样本数据增强算法 MixAug

MixMatch 中使用的 MixUp 算法为混合样本数据增强方法中的一种,这种方法混合出的样本较为模糊,而且不自然^[8]。同样是混合样本数据增强算法的 CutMix,使用二元掩码将样本中的方形区域插入到另一个数据点中,使学习到的模型保留对自然数据的良好认知。CutMix 的计算方式如下,对于两个样本 (x_1, l_1) 和 (x_2, l_2) ,CutMix 生成了一个新的混合样本 (x', l') ,如式(4)和式(5)所示^[10]:

$$x' = Mx_1 + (1-M)x_2 \quad (4)$$

$$l' = \lambda_M l_1 + (1-\lambda_M)l_2 \quad (5)$$

其中, M 为二元掩码,用于表示图像中被裁剪和保留的区域;混合比 λ_M 从 β 分布中取样,然后根据 M 的值进行调整。混合比 λ_M 控制二元掩码 M 中裁剪区域的生成大小,根据 CutMix 文献中的定义^[10],裁剪区域占图像全部区域的比例等于 $1-\lambda_M$ 。当生成的裁剪区域越过图像边缘时,超出边缘的无效部分就会被移除,此时实际的裁剪区域小于生成的裁剪区域,进而使得实际的 λ_M 小于采样得到的 λ_M 。需要计算 M 中实际裁剪区域占图像全部区域的比例,然后用1减去这一比例得到实际的 λ_M 值。CutMix 的混合效果如图2所示(图片的点阵

只有 $32 * 32$, 因此分辨率有限)。

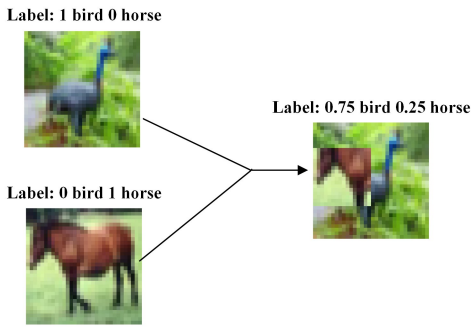


图2 CutMix 数据混合的结果
Fig. 2 Mixed results by CutMix

CutMix 需要计算裁剪的区域, 因此其计算量略高于 MixUp, 本文不使用完全替换的方法, 而是结合 MixUp 和 CutMix, 提出了一种混合样本算法 MixAug, 用于混合有标签和无标签样本。给定一组样本(其中包含有标签样本和无标签样本以及这些样本所对应的标签, 对于无标签样本来说是预测标签), MixAug 可对这些样本进行混合, 同时样本对应的标签组合形成混合标签作为混合生成样本的标签, 其算法的流程如算法 2 所示。

算法 2 MixAug

Input: A batch (B size) of labeled and unlabeled samples with their corresponding (real or guessed) labels (data, label); Beta distribution parameters α_m and α_c

1. $\lambda_m \sim \text{Beta}(\alpha_m, \alpha_m)$
2. $\lambda_c \sim \text{Beta}(\alpha_c, \alpha_c)$
3. $\lambda_m = \text{Max}(\lambda_m, 1 - \lambda_m)$
4. $\lambda_c = \text{Max}(\lambda_c, 1 - \lambda_c)$
5. for $i=0, \dots, B$ do
6. $p \sim \text{Uniform}(0, 1)$
7. if $p < \text{mix_prob}$ then
8. $\text{mix_data} = \text{CutMix}(\text{data}, \text{Shuffle}(\text{data}), \lambda_c)$
9. else then
10. $\text{mix_data} = \text{MixUp}(\text{data}, \text{Shuffle}(\text{data}), \lambda_m)$
11. end if
12. end for
13. $\lambda_{\text{aug}} = \text{Adjust}(\lambda_m, \lambda_c)$
14. $\text{mix_label} = \text{label} * \lambda_{\text{aug}} + \text{Shuffle}(\text{label}) * (1 - \lambda_{\text{aug}})$
15. return $\text{mix_data}, \text{mix_label}$

算法 2 中, 首先从 Beta 分布中采样 MixUp 的混合比 λ_m 以及 CutMix 的混合比 λ_c , 之后使用 Max 操作修正混合比, 以保证使用混合比后最终混合后的样本更接近原始样本。此外, MixAug 中对原始的 CutMix 算法做了一些修改, 不在同一个 batch 中采样一个固定的混合比, 而是从 Beta 分布中采样不同的混合比值, 使得同一个 batch 中每张图片的裁剪区域都不同, 这样做可以使在一个 batch 中混合样本的结果更具多样性。通过采样与修正计算得出, 混合比对应于算法 2 的第 1 行至第 4 行。第 5 行至第 12 行是 MixAug 核心的混合操作, 对于 batch 中的每一幅图像, 会在指定的概率下应用 CutMix 或 MixUp 操作, 将其与打乱后的 batch 中的图像进行混合, 如果对此图像应用了 MixUp, 则记录下此图像在

batch 中的索引。算法 2 中第 13 行的 Adjust 函数的逻辑如下: 定义混合比 λ_{aug} 等于 CutMix 的混合比 λ_c , 再根据应用 MixUp 的索引, 取 MixUp 的混合比 λ_m 中对应索引的值来替换 λ_{aug} 中相同索引的值, 得到最终的混合比 λ_{aug} 。之后用该混合比来计算混合标签, 对应算法 2 中的第 14 行。

4 实验及其结果分析

实验部分将本文提出的基于混合样本自动数据增强技术的半监督学习方法应用于图像分类任务, 通过图像分类方面的实验测试所提方法的性能。本文方法的目标是在有标签样本占比较低的情况下追求较低的错误率, 由于篇幅所限, 实验部分主要为验证是否达到此目标而设计。第一组实验是将本文方法与当前主流的半监督学习方法进行分类错误率的比较; 第二组实验是测试更低的有标签样本比例对本文方法的影响; 第三组实验为消融实验, 测试和比较所提方法中的每一部分对整体方法所起的作用。

4.1 实验数据集

本文在 CIFAR-10, SVHN 两个图像分类的基准数据集上对所提方法进行了有效性验证。这两个数据集的训练数据均为有标签样本, 原本用于监督学习, 但可以通过忽略部分有标签样本的标签(即把它们当成无标签样本)的方式将这些数据集用于测试基于半监督学习的图像分类的效果。

CIFAR-10 共有 6 万张分辨率为 $32 * 32$ 的 RGB 图像, 其中包含 5 万张训练图像和 1 万张测试图像。数据集包含 10 个自然界和生活中常见的物体类别, 分别为飞机、汽车、鸟类、猫、鹿、狗、蛙、马、船和卡车, 每类有 6000 张图像。SVHN 数据集包含在真实街景中出现的门牌号码图像, 由 73 257 张训练图像和 26 032 张测试图像组成, 分为 10 个类, 图像均为 $32 * 32$ 分辨率的 RGB 图像。

4.2 实验设置

为了比较的公平性, 本文使用与 MixMatch 相同的“Wide ResNet-28-2”模型, 实现和训练过程也遵循 MixMatch 文献中的做法, 学习率为 0.002, batch 大小为 64, 权重衰减为 0.02, 使用 Adam 优化器^[8]。在 MixMatch 文献中, 所有数据集的迭代次数均为 1024, 但在本文的实验中, 可以用更少的迭代次数, 以达到模型收敛, 因此实验迭代次数在 CIFAR-10 数据集上设置为 850, 在 SVHN 数据集上设置为 300。一次实验结束后, 在最后 100 次迭代中, 计算所有 20 次连续迭代的错误率的中位数, 然后选择这些中位数的最大值作为本次实验的最终结果, 这种做法的目的在于, 获得稳定的结果用于比较。硬件方面, 所有实验均在具有 4 颗 NVIDIA GeForce GTX 1080Ti GPU 的服务器上进行。

在 AutoMixMatch 中应用自动数据增强时, 对不同的数据集有不同的操作: 在 CIFAR-10 数据集的实验中, 首先应用改进的 AutoAugment, 然后使用常规的数据增强方法, 如零填充、随机裁剪和水平翻转; 对于 SVHN 数据集的实验, 只应用改进的 AutoAugment, 不使用其他的数据增强方法, 因为类似水平翻转这样的数据增强方法会影响深度学习模型对门牌数字的识别。由于 AutoAugment 在策略搜索过程中需要花费大量的时间, 因此直接使用 AutoAugment 文献在 CI-

FAR-10 和 SVHN 上公开的数据增强策略, 跳过搜索过程。

4.3 实验结果比较及分析

4.3.1 CIFAR-10 数据集上的半监督学习图像分类比较

本文在 CIFAR-10 数据集上将本文方法与当前 4 个主流的半监督学习分类方法 Π -Model, Mean Teacher, VAT 和 MixMatch 进行比较, 评价的指标为图像分类的错误率, 对比方法所对应的实验结果均来源于 MixMatch 文献。使用 3 个不同规模的有标签样本 250, 1000 以及 4000 来评估所有的方法, 3 个有标签样本的规模占 CIFAR-10 数据集全部训练样本的比例分别为 0.5%, 2%, 8%。对于每个规模的有标签样本, 从训练集中随机选取 5 组有标签样本进行 5 次实验, 记录 5 次实验错误率的平均值, 并将其作为每个规模的有标签样本所对应的实验结果。对于 MixAug 算法中所用到的两个混合数据增强方法的超参数, α_m 设置为 0.75, α_c 设置为 1.25。MixAug 的 mix_prob 参数设置为 0.5。

在 CIFAR-10 数据集上的实验结果如图 3 所示。由图可知, 在有标签样本占比较大, 如 8% 的情况下, 与 MixMatch 相比, AutoMixMatch 的错误率降低了 0.61%, 与其他 3 种方法中效果最好的 VAT 相比, AutoMixMatch 的错误率从 11.05% 下降为 5.63%, 后者为前者的约 1/2; 在有标签样本占比较小, 如 0.5% 的情况下, 与 MixMatch 相比, AutoMixMatch 的错误率降低了 1.17%, 与 VAT 相比, AutoMixMatch 的错误率从 36.03% 下降为 9.91%, 后者仅为前者的约 1/4。由上述结果可以得出以下结论: 1) 在有标签样本占比为 0.5%, 2% 和 8% 的 3 种情况下, MixMatch 和 AutoMixMatch 取得的分类错误率远低于其他 3 种方法, 而本文提出的 AutoMixMatch 的分类错误率比 MixMatch 更低; 2) 随着有标签样本占比的提高, 所有方法对应的错误率逐渐降低; 3) 在不同的有标签样本占比下, AutoMixMatch 相比其他方法的提升效果有所不同, 有标签样本的占比越小, 所提方法的提升效果越明显。造成上述结果的可能原因主要是, AutoMixMatch 采用改进的 AutoAugment 自动数据增强方法提高了数据的多样性, 并且后续的混合样本增强算法 MixAug 进一步对增强后的数据进行了混合, 丰富了训练时学习到的样本信息。

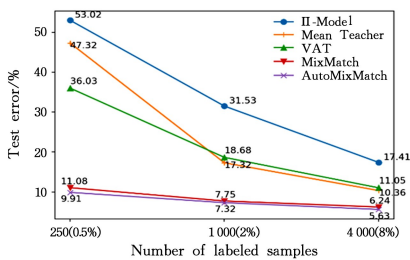


图 3 在 CIFAR-10 数据集上的错误率

Fig. 3 Error rates on CIFAR-10 dataset

表 1 列出了两种方法在训练过程中的错误率值及其变化情况。在前 100 次迭代中, AutoMixMatch 的错误率始终比 MixMatch 低约 10%。在训练到 500 次迭代时, AutoMixMatch 的错误率为 10.73%, 已经低于图 3 中 MixMatch 报告的 11.08% 的错误率(经过 1024 次迭代), 由此可见,

AutoMixMatch 可以用更少的迭代次数达到更低的错误率。造成这一差别的原因可能是, 相比 MixMatch 使用的简单数据增强方法, AutoMixMatch 采用改进的 AutoAugment 对每个样本执行不同的数据增强策略, 使产生的图像数据的多样性得到显著提高, 相当于使用了更多的训练样本, 在训练过程中, 可以用较少的迭代次数使错误率下降到较低水平。

表 1 MixMatch 和 AutoMixMatch 的错误率比较

Table 1 Comparison of the error rates between MixMatch and

AutoMixMatch

(单位: %)

Epochs\Methods	MixMatch	AutoMixMatch (ours)
10	40.45	31.76
50	29.79	17.92
100	25.35	14.77
500	13.37	10.73

4.3.2 在有标签样本比例更低情况下的比较

由于本文提出的半监督学习方法使用了先进的多样化数据增强技术, 使得方法本身对有标签样本数据占比的要求极低, 最后一组实验尝试在极低的有标签样本比例的情况下评估所提方法的性能。为了实现极低的有标签样本比例且限于文章篇幅, 仅选用包含数据量更大的 SVHN 数据集, 在其中仅设置 40 个样本作为有标签样本, 即 10 类数据中每类只有 4 个有标签样本, 有标签样本占总训练样本的比例仅为 0.05%。在前文比较的几种主流的半监督分类方法对应的文献中, 均未有在如此低的有标签样本占比下进行实验的报道; 且根据图 3 所示的 CIFAR-10 错误率, 在有标签样本占比较低的情况下, 除了 MixMatch 外, 其他方法的错误率均相对较高, 在 SVHN 数据集上的情况类似。基于以上两方面的原因, 本组实验只选择 MixMatch 与本文方法进行对比, 表 2 列出了 MixMatch 和 AutoMixMatch 的比较结果。因为在有标签样本数极低的情况下, 随机选取的不同样本集之间的错误率的差异相对较大, 因此除了错误率的平均值外, 表 2 还详细列出了 5 次随机选取的有标签样本集的实验结果。与 MixMatch 相比, AutoMixMatch 有明显的改进, 平均错误率从 42.20% 下降到了 12.03%, 降低了 30.17%。造成这种结果的原因可能是, MixMatch 使用随机裁剪等简单的增强方法, 在有标签样本数极少的情况下, 产生的增强样本的多样性和质量都十分有限, 而 AutoMixMatch 采用改进的 AutoAugment, 使得产生的样本的多样性和质量都得到显著提升, 从而提升了模型的泛化能力。

表 2 SVHN 数据集在更低的有标签样本比例上的错误率

Table 2 Error rates on SVHN dataset with lower labeled ratio

(单位: %)

Runs\Methods	MixMatch	AutoMixMatch (ours)
1	53.92	20.67
2	38.99	10.78
3	27.74	7.77
4	33.15	10.24
5	57.22	10.71
Average	42.20	12.03

4.3.3 消融实验

AutoMixMatch 中采用了改进的 AutoAugment 以及所

提出的混合样本算法 MixAug,为了更好地说明 Auto-MixMatch 和其各个部分的有效性,本组实验检测所采用的这两种方法对整体方法各自的贡献。在 CIFAR-10 上使用 250 个有标签样本进行实验。

消融实验结果如表 3 所列,其中“*”表示该错误率值从 MixMatch 文献中获得。对于 CIFAR-10 数据集,在 Auto-MixMatch 中移除改进的 AutoAugment 的情况下,错误率从 10.21% 增加到 11.51%;在移除 MixAug 的情况下,错误率从 10.21% 增加到 10.74%;同时采用改进的 AutoAugment 和 MixAug 的情况下,错误率为 10.21%,比单独采用两种方法的错误率均更低。消融实验的结果说明,单独使用改进的 AutoAugment 或 MixAug 都可以提升方法的性能,并且二者结合使用比单独使用其中任何一种方法表现都更好,验证了改进的 AutoAugment 和 MixAug 在方法中起到的作用。

表 3 在 CIFAR-10 数据集上的消融研究

Table 3 Ablation study on CIFAR-10 dataset

Ablation	CIFAR-10
AutoMixMatch	10.21
AutoMixMatch without MixAug	10.74
AutoMixMatch without modified AutoAugment	11.51
MixMatch	11.80*

结束语 半监督学习中的基于一致性的方法引起了很多学者的关注,针对常规数据增强方法面对极低的有标签样本比例性能大幅下滑的问题,本文提出了半监督方法 MixMatch 的改进方法 AutoMixMatch,将两种监督学习中的数据增强方法扩展至半监督环境中,利用数据增强产生多样性、高质量的样本,进而更好地学习增强后的无标签样本之间的一致性。在图像分类基准数据集 CIFAR-10 和 SVHN 上评估所提方法的性能,实验结果表明,AutoMixMatch 在不同的有标签样本比例下的分类错误率均低于 MixMatch,并且需要的迭代次数更少;特别是在当前文献未有报道的有标签样本比例极低的情况下,即在 SVHN 数据集中设置有标签样本占比只有 0.05% 时,本文方法仍可获得较好的分类效果,错误率仅为 12.03%,比 MixMatch 的错误率降低了 30.17%。

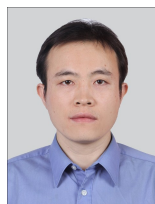
本文方法的计算成本仍较高,在未来的工作中,需要探索如何降低计算成本,并尝试将其他更优秀的数据增强方法应用于半监督学习。此外,将本文方法扩展到其他分类任务也是一个有潜力的方向。

参考文献

- [1] CHAPPELLE O, SCHOLKOPF B, ZIEN A. Semi-supervised learning (chappelle, o. et al., eds.; 2006) [book reviews] [J]. IEEE Transactions on Neural Networks, 2009, 20(3): 542.
- [2] LAINE S, AILA T. Temporal Ensembling for Semi-Supervised Learning [C] // Proceedings of the International Conference on Learning Representations (ICLR). 2017.
- [3] TARVAINEN A, VALPOLA H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results [C] // Advances in Neural Information Processing Systems. 2017: 1195-1204.
- [4] VERMA V, LAMB A, KANNALA J, et al. Interpolation Con-

sistency Training for Semi-supervised Learning [C] // Proceedings of the 28th International Joint Conference on Artificial Intelligence. AAAI Press, 2019: 3635-3641.

- [5] ZHANG H, CISSE M, DAUPHIN Y N, et al. mixup: Beyond Empirical Risk Minimization [J]. arXiv: 1710. 09412, 2017.
- [6] XIE Q, DAI Z, HOVY E, et al. Unsupervised Data Augmentation for Consistency Training [J]. arXiv: 1904. 12848, 2019.
- [7] CUBUK E D, ZOPH B, SHLENS J, et al. RandAugment: Practical data augmentation with no separate search [J]. arXiv: 1909. 13719, 2019.
- [8] BERTHELOT D, CARLINI N, GOODFELLOW I, et al. MixMatch: A Holistic Approach to Semi-Supervised Learning [J]. arXiv: 1905. 02249, 2019.
- [9] CUBUK E D, ZOPH B, MANE D, et al. AutoAugment: Learning Augmentation Strategies From Data [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 113-123.
- [10] YUN S, HAN D, OH S J, et al. CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features [C] // International Conference on Computer Vision (ICCV). 2019.
- [11] QIN Y, DING S F. Survey of Semi-supervised Clustering [J]. Computer Science, 2019, 46(9): 15-21.
- [12] ATHIWARATKUN B, FINZI M, IZMAILOV P, et al. There Are Many Consistent Explanations of Unlabeled Data: Why You Should Average [C] // Proceedings of the International Conference on Learning Representations (ICLR). 2019.
- [13] IZMAILOV P, PODOPRIKHIN D, GARISOPOV T, et al. Averaging weights leads to wider optima and better generalization [J]. arXiv: 1803. 05407, 2018.
- [14] MIYATO T, MAEDA S, 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, 2018, 41(8): 1979-1993.
- [15] FRENCH G, AILA T, LAINE S, et al. Semi-supervised semantic segmentation needs strong, high-dimensional perturbations [J]. arXiv: 1906. 01916, 2019.



XU Hua-jie, born in 1974, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include artificial intelligence, acoustic signal recognition and computer vision.



QIN Yuan-zhuo, born in 1996, doctoral candidate. His main research interests include artificial intelligence and computer vision and their applications in engineering.