

标签噪声学习算法综述^①

王晓莉, 薛 丽

(长安大学 信息工程学院, 西安 710064)

通讯作者: 王晓莉, E-mail: 2904850965@qq.com



摘 要: 机器学习中, 训练样本的标签质量严重影响着分类算法的最终效果. 虽然干净的标签产生的效果相对来说比较好, 但是采集和使用时却费时费力. 因此为了节约成本, 同时也为了使模型能够适应于一般情况, 研究人员逐渐开始针对普通类数据进行学习, 即带有标签噪声的数据. 虽然近些年有些许著作专门针对标签噪声进行研究, 但是缺乏对其的全面分析. 基于此, 本文首先对标签噪声进行简要而全面的介绍, 然后对近几年标签噪声的学习算法分别进行显式、隐式两个方面的分析, 并做出总结, 最后对未来标签噪声的研究做出展望.

关键词: 标签噪声; 深度学习; 人工智能; 机器学习; 损失函数

引用格式: 王晓莉, 薛丽. 标签噪声学习算法综述. 计算机系统应用, 2021, 30(1): 10-18. <http://www.c-s-a.org.cn/1003-3254/7776.html>

Review on Label Noise Learning Algorithms

WANG Xiao-Li, XUE Li

(School of Information Engineering, Chang'an University, Xi'an 710064, China)

Abstract: In machine learning, the label quality of the training samples seriously affects the final effect of the classification algorithms. Although the effect of a clean label is relatively good, it takes time and effort to collect and use. Therefore, in order to save costs and make the model adapt to the general situation, researchers gradually began to learn from ordinary data, that is, data with label noise. In recent years, although some works have been devoted to label noise, they lack comprehensive analysis. Based on this, this paper first introduces the label noise briefly and comprehensively, then analyzes the learning algorithms of tag noise in recent years explicitly and implicitly, and summarizes them. Finally, we look forward to the future research on label noise.

Key words: label noise; deep learning; artificial intelligence; machine learning; loss function

目前诸如图像分类、对象检测、语义分割等应用在深度学习方面表现出优异性能的根本原因分为三方面, 首先, 计算机硬件系统设备的不断提升, 尤其在图形处理单元和并行处理方面有明显提高; 其次, 网络参数初始化和训练方法方面其技术的飞跃进步, 但以上两种情况需要以数据作为支撑才可以将其良好的性能表现出来, 即第三个原因, 有类似于 ImageNet 这类清晰注释的大型数据集. 在实际情况中, 收集带有噪声的

数据集很容易, 但收集大规模干净的数据集既昂贵又耗时. 一方面, 专家标记数据对某些研究是必须的, 如, 医学中的相关数据, 由于其专业性, 需要专家进行标注; 另一方面, 对于类似于 JFT300M 这种包含 3 亿张图像的庞大数据集是无法对其进行手动标记. 同时, 机器学习领域中, 使用带有标签的数据训练神经网络进而实现分类、回归或其他目的的学习方式非常常见, 但在相关模型学习中, 如果使用含有大量错误信息的数据,

① 基金项目: 国家自然科学基金面上项目 (61771075); 长安大学中央高校基本科研业务费专项资金 (300102249203)

Foundation item: General Program of National Natural Science Foundation of China (61771075); The Fundamental Research Funds for the Central Universities of China of Chang'an University (300102249203)

收稿时间: 2020-06-05; 修改时间: 2020-06-30, 2020-07-15; 采用时间: 2020-07-23; csa 在线出版时间: 2020-12-31

则会影响最终预测模型的准确性。

因此,越来越多的学者开始针对噪声数据进行研究,其通常分为两类:特征噪声和标签噪声^[1-3]。特征噪声是指训练样本实例特征本身和其真实特征之间的偏差,例如人为地在已有特征上添加高斯噪声。标签噪声是指用于训练的目标标签与相应实例本身的真实标签之间的偏差,例如设定标签时错误设置标签所属类别。这两种噪声类型都会导致系统的性能下降,但相关研究^[4,5]表明标签噪声的危害性更大,如表情识别中,高品质面部表情到低品质微表情的不确定性,会导致标签不一致甚至标签错误,最终降低系统最终的性能;医学图像分析领域中,病变程度和相应位置的不同其病因也不同,稍有偏差的标记都会影响最终诊疗结果;军事领域中,由于外观的相似性,坦克和自行榴弹炮等也容易被错误标注,导致目标识别场景中的错误检测;在语义分割中,由于标注规则的不合理,从而导致训练模型效果不好等,这些都体现出处理标签噪声的重要性。

因此本文主要针对标签噪声讨论。

1 标签噪声的介绍

1.1 标签噪声的来源和类型

标签噪声是一种复杂的现象,其产生的原因有多种,主要分为4类。(1)在标注过程中的信息不足,无法对类别数据进行充分的描述^[6,7];(2)由于数据质量较差导致辨识度降低,使得专家在标注过程中无法正确标注相应标签^[8,9];(3)不同主观性所致,即标签的标注非全由专家进行标注,这是由于可靠的标签是一项耗时且昂贵的任务,因此有些标签是从非专家处获得,且不同的标注人员根据不同角度对标签进行标记得到的结果不完全一致^[10,11];(4)来自数据编码或通信问题也可能导致样本标签出现错误^[12]。

从嘈杂的标签中学习一直以来都是一项长期挑战,在研究其对分类器或分类算法的影响时发现,标签噪声可能受到以下3个因素影响:数据特征、数据的真实标签和标签特性,根据这些因素相互间的依赖关系进行分析可以将标签噪声分为3类^[13,14]:随机标签噪声(Random Classification label Noise, RCN)、类标签噪声(Class-Conditional label Noise, CCN)和实例相关标签噪声(Instance-Dependent label Noise, IDN)。如图1所示,这些模型的依存关系复杂性从左到右依次增加,

其分别表示为错误标签与实例和真实标签均无关、错误标签只与真实标签有关、错误标签只与实例特征有关。其中 x 表示观察到的实例特征, y 表示真实标签,表示相应的噪声标签,随机变量 e 表示该实例特征是否受到标签噪声带来的错误影响。

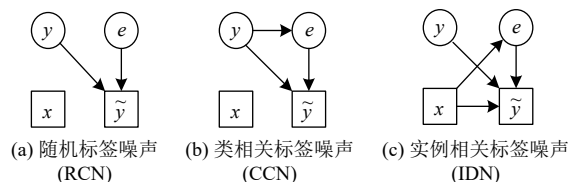


图1 标签噪声类型

1.2 标签噪声的影响

通常情况下,带有标签噪声的数据在现实生活中无处不在,直接使用标签噪声的数据会对实验产生很大的影响,研究发现标签噪声产生的影响大致分为3类:(1)标签噪声下的分类方法导致预测性能降低,例如分类时经常使用的线性分类器、kNN分类器都会受到标签噪声的影响^[15]。同时,标签噪声也会影响常用的决策树、支持向量机等方法。(2)训练特征数和模型复杂性增加,例如由于标签噪声的影响,SVM中支持向量的数量明显增加,导致其模型更加复杂^[7];标签噪声需要进行多次训练才会达到指定的性能指标^[16]。(3)观察到的类别出现频率可能会发生变化,如标签噪声对某种疾病发病率产生的影响^[8]。

2 标签噪声学习算法

近年来,硬件设备的不断提高使得越来越多的研究者开始倾向于在一般情况下的研究,即进行实验时所采集的数据基本都是被污染的、有噪声的。目前通过网络爬虫或众包等方法采集的数据大部分都包含噪声,这些数据显然掩盖了实例特征和其类别之间的真实关系;同时在根据数据创建模型或者基于数据进行决策的学习中,使用存在噪声的数据会在构建分类器的时间、分类的准确性以及分类器大小等方面降低系统性能,也会增加学习所需的样本数量和相应模型的复杂性。因此设计适应于与标签噪声有关的学习算法在研究中具有重要的研究意义和应用价值。

在标签噪声学习算法的有关研究中,不同的研究人员针对不同的分类情况进行研究,有根据数据集噪声嘈杂程度进行区分,有根据噪声类型划分等。与大多

数研究不同, 本文主要侧重于噪声结构的建模方式对其进行区分, 其方式分为显式处理和隐式处理. 目前, 有很多学习算法都对标签数据在嘈杂环境中的学习能力进行了研究^[17], 本文分析有关标签噪声相关处理学习算法如图2所示, 具体分析如下.

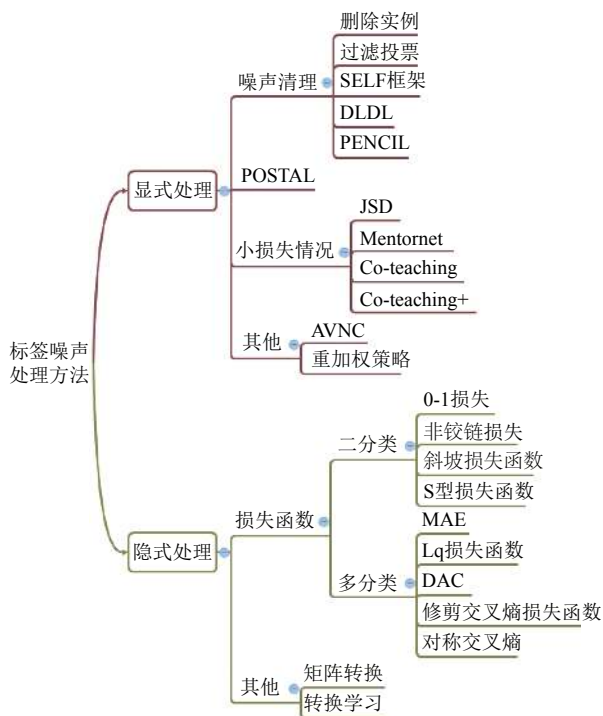


图2 标签噪声处理方法

2.1 显式处理

显式处理即以某种方式对噪声进行建模并操纵该模型的输入流. 通常情况下, 使用标签噪声学习时, 最先想到的是对标签噪声进行清理, 其旨在消除或纠正训练数据中标记错误的示例, 该步骤可以在训练前完成, 也可以和主模型训练同时进行.

噪声清理中最简单的方式是直接删除分类器中分类错误的训练实例, 也有将异常值影响较大的实例或看起来比较可疑的实例删除^[18], 但这些方式会造成数据严重缺失, 最终导致算法的准确性降低. 因此, 研究者正在找寻更有效的方式来清理标签噪声. 有学者使用过滤投票^[19]的方式进行噪声清理, 简单理解为: 当一个集合中的所有(或几乎所有)学习者都同意删除该实例时, 该实例将被删除, 此方式能够很好地解决数据严重缺失问题, 但也容易误删某些正确的实例.

文献[20]则提出了一种自集成标签过滤框架

(Self-Ensemble Label Filtering, SELF), 该框架采取渐进式过滤, 过滤策略基于标签和最大似然预测之间的一致性来确定潜在的标签, 当模型预测相应的标签具有最高可能性的正确类别时, 才会将该标签用于训练, 经过测试发现该框架在训练时间足够长的情况下也不易过拟合, 同时该技术可以和不同的半监督损耗兼容, 但其在噪声比较高的情况下表现不佳.

文献[21]将传统的SLR (Single-Label Recognition) 和MLR (Multi-Label Recognition) 问题转换成LDL (Label Distribution Learning) 问题, 提出了端到端的学习框架DLDL (Deep Label Distribution Learning), 其主要利用深层ConvNet在特征学习和分类器学习中的标签歧义性来学习标签分布, ConvNet最后一个全连接层激活函数的概率分布为:

$$\hat{y}_j = \frac{\exp(x_j)}{\sum_r \exp(x_r)}; \quad x = \varphi(X; \theta) \quad (1)$$

最终的标签分布 y 可以根据学习到的 θ 得到, 实验结果表明在训练集很小的情况下, DLDL 也有助于防止网络过度拟合, 但 DLDL 中标签分布固定, 无法更新.

为解决标签分布无法更新问题, 文献[22]则提出了名为PENCIL (Probabilistic End-to-end Noise Correction for Learning with noisy labels) 的端到端框架, 其独立于骨干网络结构, 通过端到端方式在学习网络参数和标签中维护并更新标签分布来校正噪声标签, 在迭代过程中, 损失函数(2)用于更新操作, 其中 y^d 为噪声标签,

$$L = \frac{1}{c} L_c(f(x; \theta), Y^d) + \alpha L_o(\hat{Y}, Y^d) + \frac{\beta}{c} L_e(f(x; \theta)) \quad (2)$$

其中, α 和 β 为超参. 研究表明该框架在不需要辅助的干净数据集或有关噪声的先验信息的条件下也容易部署, 但该方法在噪声较高时该方法会失效.

文献[23]则提出的新颖框架中引入了一种名为AVNC (Adaptive Voting Noise Correction) 的技术来精确识别和校正潜在的噪声标签, 该框架的核心是噪声处理. 主要包括过滤和校正两个步骤. 首先在过滤阶段, 经过 M 轮 K 折交叉验证后对式(3)进行降序排列,

$$r^{(i)} = B_c^{(i)} + \left(1 - B_e^{(i)} \left/ \sum_{i'} B_e^{(i')} \right. \right) \quad (3)$$

并根据 $B_c^{(i)} = B_e^{(i)} + 1$ 来确定噪声, 即将数据集分为噪声和干净数据集, 然后通过校正步骤尝试预测噪声数据

集中的正确标签. 该 AVNC 根据 K 折交叉验证预测实例生成的高质量的数据集来构建模型, 从而对噪声进行校正. 但 ANVC 仅针对二分类标签进行研究, 在多数情况下其噪声校正性能会很低.

对此, 文献 [24] 使用重加权策略对称噪声标签进行研究, 并将该策略应用到 SVM 和 MLR 两个常规的分类器, 如 IWMLR 的分类概率分别为:

$$P(Y^k = 1|X, W) = \frac{\exp(W_k^T X)}{\sum_{l=1}^m \exp(W_l^T X)} \quad (4)$$

但该方法主要适用于 RCN 模型下的对称噪声标签, 并不适用于非对称噪声, 且其权重并不容易准确估计.

有学者则将人类学习方式融入到模型学习中, 即先对简单任务学习, 然后逐步学习较困难的任务, 从而建立一个可靠的模型. 如 Bengio 等人 [25] 在 2009 年提出了课程学习 (Curriculum Learning, CL) 框架, Kumar 等人 [26] 在 2010 年提出一种类似的学习方式, 即自步学习 (Self-Paced Learning, SPL), 但是两者都没有应用到标签噪声中, 无法解决实际应用中的问题. 有研究将 CL 思想融入到嘈杂标签学习中, 即从干净的标签慢慢转移到噪声标签的学习.

文献 [27] 利用 CL 优势的同时, 结合随机梯度下降 (Stochastic Gradient Descent, SGD) 具有减少时间和降低存储成本的优点, 首次两者更新过程结合在一起并提出了渐进式随机学习 (Progressive Stochastic Learning, POSTAL), 主要通过设定的动态阈值 $D_{th} = D_{th} - \mu \sqrt{T}$ 来控制更新区域, 该更新方式在初始时期只能在可靠标签上建立健壮模型, 随后逐渐降低该动态阈值并更新嘈杂标签, 但是其课程通常是预定义的, 忽略了反馈信息, 同时对于小批量的 SGD 训练非常困难.

为此, 基于深度神经网络的记忆效应, 有研究专门针对小批量损失情况下的实例进行训练来处理嘈杂标签. 例如, 文献 [28] 则提出了 Mentornet 网络在 mini-batch 上根据相应的反馈信息来更新课程, 该网络能够克服损坏标签的过度拟合, 但容易累积错误信息; 文献 [29] 提出的 Co-teaching 以对称方式交叉训练两个网络, 即将在小批量数据中过滤的噪声样本传递给对等网络来更新参数, 解决了 Mentornet 的累积错误问题, 但训练时间太长, 两个网络容易达到收敛共识; 文献 [30] 则在 Co-teaching 的基础上进行改进, 提出的 Co-teaching+ 采用分歧更新策略来训练网络, 能够明显提高 DNN 抵御噪声标签的鲁棒性, 但当数据集噪声率极高时, 却很

难适用于小批量训练. mini-batch 下相应的网络训练结构如图 3 所示.

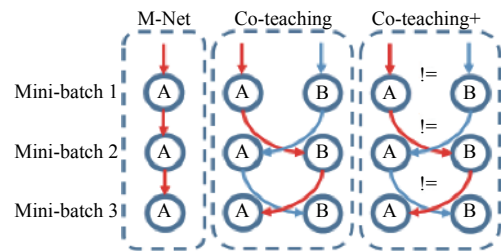


图 3 mini-batch 相关网络

2.2 隐式处理

隐式处理即不对噪声进行显式建模, 而是设计更加通用的算法来消除噪声的负面影响. 使用固有噪声容忍方法能够实现噪声的鲁棒性, 该方法主要在风险最小化的框架内进行研究, 且风险最小化的稳健性主要取决于损失函数. 大量研究也表明, 损失函数的处理会使标签噪声在训练时鲁棒性更好.

在二分类问题中, 0-1 损失对于对称或均匀标签噪声体现出很好的鲁棒性 [31], 通常非铰链损失函数 (Unhinged Loss) [32,33]、斜坡损失函数 (Ramp Loss) [31] 和 S 型损失函数 (Sigmoidal Loss) [31] 也应用于二分类研究中. 但仅对二分类情况下的标签噪声进行研究是无法满足实际情况下的大多数问题. 因此有学者提出鲁棒风险最小化方法, 用于通过估计标签损坏概率来学习多类分类的神经网络.

文献 [34] 针对多类分类中常见的损失函数进行研究, 定义了分类器 f 的最小化风险损失函数满足:

$$R_L^n(f) = E_{D_{\eta}} \left[L(f(x), \hat{y}_x) \right] = E_{x, \hat{y}_x} \left[L(f(x), \hat{y}_x) \right] \quad (5)$$

然后通过反向传播用随机梯度下降方法进行学习, 并对比基于分类交叉熵 (Categorical Cross Entropy, CCE), 均方误差 (Mean Square Error, MSE) 和平均绝对值误差 (Mean Absolute Error, MAE) 在不同类型标签下的实验, 结果表明 MAE 具有很好的鲁棒性, 但是单独的 MAE 会增加训练的难度, 特别是在复杂数据集上, 会直接导致最终的性能下降.

CCE 是一种非对称且无界的损耗函数, 收敛速度快, 但对标签噪声比较敏感. 因此, 文献 [35] 提出了截断的 L_q 损失函数 (Truncated L_q Loss), 如式 (6):

$$L_{\text{trunc}}(f(x), e_j) = \begin{cases} L_q(k) & \text{if } f_j(x) \leq k \\ L_q(f(x), e_j) & \text{if } f_j(x) > k \end{cases} \quad (6)$$

为了凸显 MAE 的噪声鲁棒性和 CCE 的隐式加权的优势,使用负 Box-Cox 变换作为损失函数,即在继承 MAE 鲁棒性的同时也继承了 CCE 收敛速度快的特点,但其对阈值 k 有严格的要求,同时在求全局最小值时,修剪步骤计算过于复杂,并且在相似度高的数据集上会错误标记非常相似的类。

也有单独针对交叉熵 (Cross Entropy, CE) 损失函数进行改进,如文献 [36] 提出一种允许在训练过程中可弃权的损失函数,深层弃置分类器损失函数 (Deep Abstaining Classifier, DAC) 如式 (7), 式中 p_{k+1} 表示弃权可能性, α 表示相应惩罚. 该策略允许模型放弃对某些数据点的预测,同时继续学习和改善未弃权样本的分类性能. 该方法易于实现,且只需更改损失函数就可和现有任意 DNN 架构一起使用,但若随机标签中存在不明显噪声时,该方法并不适用。

$$L(x_j) = (1 - p_{k+1}) \left(- \sum_{i=1}^k t_i \log \frac{p_i}{1 - p_{k+1}} \right) + \alpha \log \frac{1}{1 - p_{k+1}} \quad (7)$$

文献 [37] 则对 CE 在标签噪声下容易产生过拟合问题进行分析,受对称 KL 散度 (Symmetric KL-Divergence) 启发提出的反向交叉熵 (Reverse Cross Entropy, RCE) 能够很好地容忍噪声,同时将 RCE 和 CE 进行结合提出了对称交叉熵 (Symmetric cross entropy Learning, SL), 如式 (8) 中, α 和 β 为解耦超参, α 用于解决 CE 中的过拟合问题,而 β 对应于 RCE 的鲁棒性. SL 使用 RCE 的噪声鲁棒性特点对称地增强了 CE 中的学习不足和过度拟合问题,该方法在手动损坏的嘈杂标签上有很好的性能,但对真实数据下比较相似的类容易混淆。

$$l_{sl} = \alpha l_{ce} + \beta l_{rce} \quad (8)$$

文献 [38] 则针对 CCE 进行研究,提出在网络误差计算中进行修整,用更可靠的误差度量代替 CCE,即基于修剪绝对值准则的修剪分类交叉熵 (Trimmed Categorical Cross-Entropy, TCCE), 损失函数定义为:

$$E_{TCCE} = -\frac{1}{h} \sum_{i=1}^h q_{i:N} \quad (9)$$

其中, $q_{1:N} \leq \dots \leq q_{N:N}$ 是每个观测值的有序损耗,即:

$$q_i = \sum_{c=1}^C (p_{ic} \log(v_{ic})) \quad (10)$$

该方法能够在较大噪声水平下产生很好的性能,但在噪声过高时,性能没有 CCE 好,且无法获得干净

数据标签的最高准确度。

有研究使用噪声转换矩阵处理噪声标签,即该矩阵定义了一个类别转换为另一个类别的可能性,文献 [39] 使用的前向和后向两种校正方法即采用的该策略,该损失校正与应用程序和网络体系结构无关,但现实中的噪声过渡矩阵并不是一直可用的,而且也很难得到准确的估计。

同时上述方式并没有针对像本文所提及的 IDN 模型下的标签噪声进行研究,这类噪声函数通常会设置有严格的要求,如给实例错误标签设定统一边界. 这些苛刻的要求一般都很难应用到实际情况中,因此一些研究通常不对标签噪声的生成过程做出假设,而隐式地处理实例相关的标签噪声。

对深层神经网络的可传递性观察发现,使用转移学习可以很好地避免过拟合,即在较大数据集上预训练模型的同时在较小数据集上进行微调,该策略必须保证两任务之间存在很大相关性,不能过于松散. 文献 [40] 对一般情况下的噪声标签进行研究发现,只在干净标签上微调的效果并没有同时在干净和嘈杂的标签上效果好,因此引入一个含有少量干净和大量嘈杂标签的通用框架 CNN, 同时使用标签噪声类型进行建模,如图 4 所示 (其中 θ_1 和 θ_2 分别表示 CNN 中的参数集, y 和 z 分别表示真实标签和标签噪声类型, x 和 \bar{y} 分别表示观察到的图像和嘈杂标签), 最后将其集成到端到端的深度学习系统中. 该方法在处理一般情况下标签噪声的同时也可以处理多类问题,但在只有嘈杂标签情况下训练网络时,该模型会出现漂移现象。

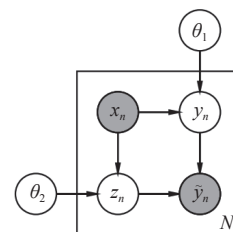


图 4 标签噪声概率图模型

综上所述,显式和隐式两种分类方法对标签噪声的处理都可以达到很好的效果. 由于噪声行为本身的特殊性,显式处理对噪声本身进行建模,并在训练过程中使用建模后的信息来获得更好的性能. 而隐式处理则是针对标签噪声的鲁棒性进行研究,并不是直接对噪声行为本身进行改进. 如表 1 所示总结了每个算法的优缺点。

表1 显示处理和隐式处理算法优缺点总结

处理方式	方法	不同方法优缺点	
显示处理	清理噪声	文献[18]处理方式简单,但数据容易缺失;文献[19]解决数据严重缺失问题,但也容易误删某些正确的实例;文献[20]可以和不同的半监督损耗兼容,却在噪声比较高情况下表现不好;文献[21]训练集很小的情况下可防止网络过度拟合,但标签分布固定无法更新;文献[22]易部署,噪声较高时会失效.	
	ANVC	文献[23]适用于二分类,多分类下性能低	
	重加权	文献[24]适用对称噪声标签,权重并不易估计	
	样本选择	文献[27]减少时间、降低存储成本,但易忽略反馈信息,且对小批量SGD训练困难	
隐式处理	小损失	文献[28]克服了损坏标签的过度拟合情况,但容易累积错误信息;文献[29]解决了累积错误问题,但训练时间太长,两个网络容易达到收敛共识;文献[30]鲁棒性好,但在噪声率极高时,难适用于小批量训练.	
	0-1损失 ^[31] 、非铰链损失 ^[32,33] 、斜坡和S型损失 ^[33]	适用于二分类,不适用于多分类	
	损失函数	DAC ^[36] 、SL ^[37]	文献[36]易于实现,可以和现有任意DNN架构一起使用,但若随机标签中存在不明显噪声,则该方法并不适用;文献[37]在手动损坏的嘈杂标签上有很好的性能,但真实数据下比较相似的类容易混淆.
		CCE、TCCE ^[38]	CCE收敛速度快,但对标签噪声比较敏感;文献[38]较大噪声水平下性能好,但噪声过高时,其性能没有CCE好,且无法获得干净数据标签的最高准确度.
		MAE ^[34] 、截断Lq ^[35]	文献[34]鲁棒性好,但数据集复杂时性能下降;文献[35]鲁棒性好收敛速度快,但修剪时计算过于复杂,且会错误标记非常相似的类.
		转移学习	很好地避免过拟合,但必须保证两任务之间存在很大的相关性,不能过于松散.
其他	前后向校正 ^[39]	文献[39]该方式与应用程序和网络体系结构无关,但现实中非一直可用的,而且难以准确估计.	
	通用框架 ^[40]	文献[40]可以处理一般情况下的标签噪声的同时也可以处理多类分类问题,但在只有嘈杂标签时,模型会漂移.	

3 展望与进一步挑战

3.1 展望

3.1.1 数量趋势

本文针对2015~2019年顶级会议上的论文进行调研,统计并分析和标签噪声相关论文数如图5所示.

(1) 图5(a)中可以很明显地看出,有关标签噪声的各类会议在2015年屈指可数,但这之后其呈现出快速增长的趋势,虽然每年都存在小幅度波动,但丝毫不影响其整体性发展.

(2) 图5(b)中可以发现,2015~2019年中关于噪声的研究文献明显增长,特别是在2019年,其研究文献已经达到2015年的4倍,可以预计,未来对标签噪声的研究会越来越多,亦有可能呈现爆发性增长,其将会成为人工智能领域的一个热门研究对象.

(3) 上述顶级会议论文中,针对标签噪声的学习不仅仅是关于理论知识的研究,同时也包含其实际应用中的研究,理论与应用两方面的研究从侧面体现了关于标签噪声学习的重要性.

3.1.2 热点趋势

处理标签噪声是一个开放性问题,本文根据其特点和实际应用性将其热点趋势大致分为4种情况:

(1) 多角度性: 参阅文献可以看出,针对标签噪声研究方法多种多样,但大部分研究主要针对以下几方面,如: 处理噪声敏感的标签; 类的相似性过高下的误判; 训练过程中的过拟合; 不同噪声比下的误差过大; 不同网络间优势相结合等等. 显而易见,从算法不同角度及其本身进行剖析一直是该领域研究热点.

(2) 通用性: 有文献设计出与应用程序和网络体系都无关的方法,也有设计出可以和任意DNN架构一起使用的方法,这些设计从侧面反映出其方法对环境的非依赖性,同时方法的通用性会进一步减少研究成本,由此可推断出学者会针对此方面做进一步研究.

(3) 适用性: 上述研究基本针对人工合成噪声数据,其准确率虽然不错,但并未对真实数据研究,为增强噪声数据的适用性,可以预测未来将逐渐针对如网络爬虫、众包等真实数据或更复杂的环境进行研究.

(4) 易于实现性: 虽然现有框架下的研究对标签噪声处理取得了不错效果, 但若应用到实际情况, 其效果不尽其然, 因此如何设计出能够有效处理标签噪声且易于实现的方法必将成为研究热点。

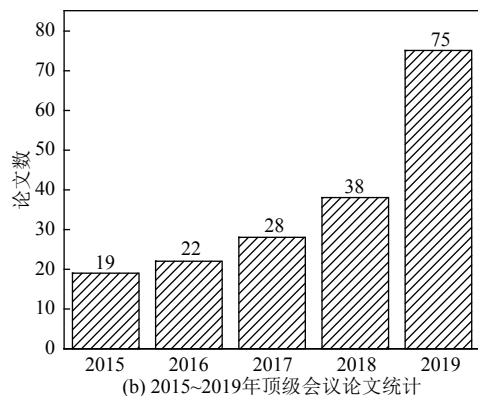
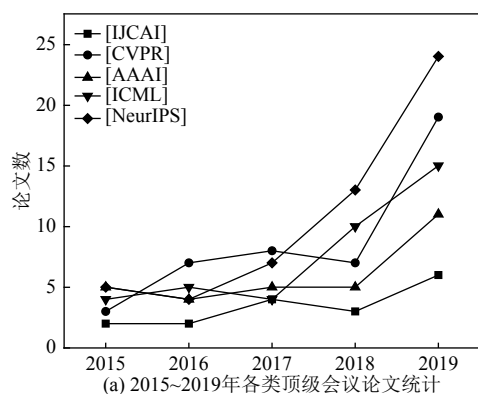


图5 顶级会议相关文献统计

3.2 标签噪声处理的进一步挑战

虽然标签噪声的研究在理论层面和工程领域都有丰硕的研究成果, 但其在实际应用中仍存在很多问题。

(1) 虽然标签噪声清理方法能够用去除或纠正错误标签的实例, 但这些方法很难将信息丰富的示例与有害的错误贴标签的示例明显区分, 同时过度清洗也可能会将分类器的性能降低。

(2) 大多数损失函数对标签噪声并不是完全鲁棒且有些对异常值敏感, 对于处理方式过于复杂的损失函数, 容易产生过拟合现象。

(3) 现有方法对标签噪声的相关研究假设了一定的不实际的前提条件, 同时有很少文献针对一般情况下的标签噪声进行研究, 如大多数研究并不是基于网络爬虫等进行研究, 无法将其一般化于各类数据情况下的标签噪声。

(4) 多分类情况下, 标签噪声在破坏原始干净数据分布的同时会导致标签浮动噪声问题, 其容易限制模型的泛化能力; 且多特征数据中有些标签相关性能比较微弱, 需要仔细辨别。

(5) 有关标签噪声的应用问题, 在不同场景下应对策略也有所不同, 需要针对标签噪声的场景适用性问题作出进一步研究。

4 总结

本文根据噪声结构的建模方式对标签噪声的处理方法从显式和隐式两大方面做出系统性梳理和总结。从众多文献可以看出, 在训练前直接删除数据会造成数据严重缺失, 目前大部分研究主要使用“训练-清洗数据-再训练”此种迭代方式进行研究, 且逐渐倾向于多网络方式结合、监督和非监督技术结合或损失函数相关处理方面等; 但大部分研究都有相应的特定场景, 且最终效果并不乐观, 和实际情况下不同场景的应用有很大差距。因此需要研究出一种具有通用性好、适用性强并且易于实现的方法来处理不同场景下的标签噪声。

参考文献

- Quinlan JR. Induction of decision trees. *Machine Learning*, 1986, 1(1): 81-106.
- Zhu XQ, Wu XD. Class noise vs. attribute noise: A quantitative study. *Artificial Intelligence Review*, 2004, 22(3): 177-210. [doi: 10.1007/s10462-004-0751-8]
- Wu XD. *Knowledge Acquisition from Databases*. Norwood, New Jersey: Ablex, 1995.
- Sáez JA, Galar M, Luengo J, *et al.* Analyzing the presence of noise in multi-class problems: Alleviating its influence with the One-vs-One decomposition. *Knowledge and Information Systems*, 2014, 38(1): 179-206. [doi: 10.1007/s10115-012-0570-1]
- Mirylenka K, Giannakopoulos G, Do L M, *et al.* On classifier behavior in the presence of mislabeling noise. *Data Mining and Knowledge Discovery*, 2017, 31(3): 661-701. [doi: 10.1007/s10618-016-0484-8]
- Brazdil P, Clark P. *Machine Learning, Meta-Reasoning and Logics—Learning from Imperfect Data*. Boston: Springer, 1990: 207-232.
- Brodley CE, Friedl MA. Identifying mislabeled training data. *Journal of Artificial Intelligence Research*, 2011, 11(1): 131-167.

- 8 Dawid AP, Skene AM. Maximum likelihood estimation of observer error-rates using the EM algorithm. *Journal of the Royal Statistical Society: Series C*, 1979, 28(1): 20–28.
- 9 Hickey RJ. Noise modelling and evaluating learning from examples. *Artificial Intelligence*, 1996, 82(1–2): 157–179.
- 10 Marcus MP, Marcinkiewicz MA, Santorini B. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 2002, 19(2): 313–330.
- 11 Snow R, O'Connor B, Jurafsky D, *et al.* Cheap and fast-but is it good?: Evaluating non-expert annotations for natural language tasks. *Proceedings of 2008 Conference on Empirical Methods in Natural Language Processing*. Honolulu, HI, USA. 2008. 254–263.
- 12 Sculley D, Cormack GV. Filtering email spam in the presence of noisy user feedback. *Proceedings of the 5th Conference on Email and Anti-Spam*. Mountain View, CA, USA. 2008.
- 13 Frénay B, Verleysen M. Classification in the presence of label noise: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, 25(5): 845–869. [doi: [10.1109/TNNLS.2013.2292894](https://doi.org/10.1109/TNNLS.2013.2292894)]
- 14 Frénay B, Kabán A. A comprehensive introduction to label noise. *Proceedings of 2014 European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. Bruges, Belgium. 2014. 667–676.
- 15 Okamoto S, Nobuhiro Y. An average-case analysis of the k-nearest neighbor classifier for noisy domains. *Proceedings of the 15th International Joint Conference on Artificial Intelligence*. San Francisco, CA, USA. 1997. 238–245.
- 16 Aslam JA, Decatur SE. On the sample complexity of noise-tolerant learning. *Information Processing Letters*, 1996, 57(4): 189–195. [doi: [10.1016/0020-0190\(96\)00006-3](https://doi.org/10.1016/0020-0190(96)00006-3)]
- 17 Lee KH, He XD, Zhang L, *et al.* CleanNet: Transfer learning for scalable image classifier training with label noise. *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA. 2018. 5447–5456.
- 18 Gamberger D, Lavrac N. Conditions for Occam's razor applicability and noise elimination. *Proceedings of the 9th European Conference on Machine Learning*. Czech Republic. 1997. 108–123.
- 19 Brodley CE, Friedl MA. Identifying and eliminating mislabeled training instances. *Proceedings of the 13th National Conference on Artificial Intelligence*. Portland, ON, USA. 1996. 799–805.
- 20 Nguyen DT, Mummadi CK, Ngo TPN, *et al.* SELF: Learning to filter noisy labels with self-ensembling. *arXiv: 1910.01842v1*, 2019.
- 21 Gao BB, Xing C, Xie CW, *et al.* Deep label distribution learning with label ambiguity. *IEEE Transactions on Image Processing*, 2016, 26(6): 2825–2838.
- 22 Yi K, Wu JX. Probabilistic end-to-end noise correction for learning with noisy labels. *Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, CA, USA. 2019. 7017–7025.
- 23 Zhang J, Sheng VS, Li T, *et al.* Improving crowdsourced label quality using noise correction. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(5): 1675–1688. [doi: [10.1109/TNNLS.2017.2677468](https://doi.org/10.1109/TNNLS.2017.2677468)]
- 24 Wang RX, Liu TL, Tao DC. Multiclass learning with partially corrupted labels. *Proceedings of the IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(6): 2568–2580. [doi: [10.1109/TNNLS.2017.2699783](https://doi.org/10.1109/TNNLS.2017.2699783)]
- 25 Bengio Y, Louradour J, Collobert R, *et al.* Curriculum learning. *Proceedings of the 26th Annual International Conference on Machine Learning*. Montreal, QC, Canada. 2009. 41–48.
- 26 Kumar MP, Packer B, Koller D. Self-paced learning for latent variable models. *Proceedings of the 23rd International Conference on Neural Information Processing Systems*. Vancouver, BC, Canada. 2010. 1189–1197.
- 27 Han B, Tsang IW, Chen L, *et al.* Progressive stochastic learning for noisy labels. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(10): 5136–5148. [doi: [10.1109/TNNLS.2018.2792062](https://doi.org/10.1109/TNNLS.2018.2792062)]
- 28 Jiang L, Zhou ZY, Leung T, *et al.* MentorNet: Learning data-driven curriculum for very deep neural networks on corrupted labels. *International Conference on Machine Learning*. Stockholm, Sweden. 2018. 2309–2318.
- 29 Han B, Yao QM, Yu XR, *et al.* Co-teaching: Robust training of deep neural networks with extremely noisy labels. *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Montreal, QC, Canada. 2018. 8527–8537.
- 30 Yu XR, Han B, Yao JC, *et al.* How does disagreement help generalization against label corruption? *arXiv: 1901.04215*, 2019.
- 31 Ghosh A, Manwani N, Sastry PS. Making risk minimization tolerant to label noise. *Neurocomputing*, 2015, 160: 93–107. [doi: [10.1016/j.neucom.2014.09.081](https://doi.org/10.1016/j.neucom.2014.09.081)]
- 32 Van Rooyen B, Menon AK, Williamson RC. Learning with symmetric label noise: The importance of being unhinged.

- Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, QC, Canada. 2015. 10–18.
- 33 Natarajan N, Dhillon IS, Ravikumar PK, *et al.* Learning with noisy labels. Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA. 2013. 1196–1204.
- 34 Ghosh A, Kumar H, Sastry PS. Robust loss functions under label noise for deep neural networks. Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, CA, USA. 2017.
- 35 Zhang ZL, Sabuncu MR. Generalized cross entropy loss for training deep neural networks with noisy labels. Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montreal, QC, Canada. 2018. 8778–8788.
- 36 Thulasidasan S, Bhattacharya T, Bilmes J, *et al.* Combating label noise in deep learning using abstention. Proceedings of the 36th International Conference on Machine Learning. Long Beach, CA, USA. 2019. 6234–6243.
- 37 Wang YS, Ma XJ, Chen ZY, *et al.* Symmetric cross entropy for robust learning with noisy labels. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Republic of Korea. 2019. 322–330.
- 38 Rusiecki A. Trimmed robust loss function for training deep neural networks with label noise. Proceedings of the 18th International Conference on Artificial Intelligence and Soft Computing. Zakopane, Poland. 2019. 215–222.
- 39 Patrini G, Rozza A, Krishna Menon A, *et al.* Making deep neural networks robust to label noise: A loss correction approach. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 1944–1952.
- 40 Xiao T, Xia T, Yang Y, *et al.* Learning from massive noisy labeled data for image classification. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 2691–2699.