IPLab 智能感知与信息处理实验室

论文分享会

余海军　2025 年 3 月 26日

#弱监督多标签分类

Large Loss Matters in Weakly Supervised Multi-Label Classification

Youngwook Kim[1*]　Jae Myung Kim[2*]　Zeynep Akata[2,3,4]　Jungwoo Lee[1,5]

[1]Seoul National University　[2]University of Tübingen
[3]Max Planck Institute for Intelligent Systems　[4]Max Planck Institute for Informatics　[5]HodooAI Lab

Abstract

Weakly supervised multi-label classification (WSML) task, which is to learn a multi-label classification using partially observed labels per image, is becoming increasingly important due to its huge annotation cost. In this work, we first regard unobserved labels as negative labels, casting the WSML task into noisy multi-label classification. From this point of view, we empirically observe that memorization effect, which was first discovered in a noisy multi-class setting, also occurs in a multi-label setting. That is, the model first learns the representation of clean labels, and then starts memorizing noisy labels. Based on this finding, we propose novel methods for WSML which reject or correct the large loss samples to prevent model from memorizing the noisy label. Without heavy and complex components, our proposed methods outperform previous state-of-the-art WSML methods on several partial label settings including Pascal VOC 2012, MS COCO, NUSWIDE, CUB, and OpenImages V3 datasets. Various analysis also show that our methodology actually works well, validating that treating large loss properly matters in a weakly supervised multi-label classification. Our code is available at https://github.com/snucml/LargeLossMatters.

1. Introduction

Multi-label classification aims to find all existing objects or attributes in a single image. It is gaining attention since the real world is made up of a scene with multiple objects in it [28, 35]. Moreover, some of the single-label datasets, also called multi-class datasets, actually have images containing multiple objects [33, 56]. However, the multi-label classification task has some fundamental difficulties in making a dataset because it requires annotators to label all categories' existence/absence for every image. As the number of categories and images in the dataset increase, annotation cost becomes tremendous [19].

*Equal contribution.

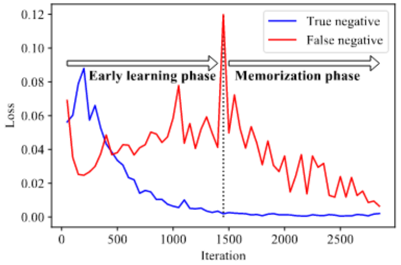


Figure 1. **Memorization in WSML.** When training ResNet-50 model on PASCAL VOC dataset with partial label, we set all unobserved labels as negative. These labels are composed of true negative and false negative. We observe that the model first fits into true negative label (learning), and then fits into false negative (memorization).

To alleviate these issues, weakly supervised learning approach in multi-label classification task (WSML) has been taken into consideration [2, 18, 36, 50]. In a WSML setting, labels are given as a form of partial label, which means only a small amount of categories is annotated per image. This setting reflects the recently released large-scale multi-label datasets [12, 19] which provide only partial label. Thus, it is becoming increasingly important to develop learning strategies with partial labels.

There are two naive approaches to train the model with partial labels. One is to train the model with observed labels only, ignoring the unobserved labels. The other is to assume all unobserved labels are negative and incorporate them into training because majorities of labels are negative in a multi-label setting [32]. As the second one has a limitation that this assumption produces some noise in a label which hampers the model learning, previous works [7, 9, 16, 21] mostly follow the first approach and try to explore the cue of unobserved labels using various techniques such as bootstrapping or regularization. However, these approaches include

论文基本信息

- 论文题目：
  - **Large Loss Matters in Weakly Supervised Multi-Label Classification**
  - 大损失在弱监督多标签分类中的重要性
- 作者信息：首尔国立大学
- 会议分区：CVPR（CCFA）

摘要

Abstract

Weakly supervised multi-label classification (WSML) task, which is to learn a multi-label classification using partially observed labels per image, is becoming increasingly important due to its huge annotation cost. In this work, we first regard unobserved labels as negative labels, casting the WSML task into noisy multi-label classification. From this point of view, we empirically observe that memorization effect, which was first discovered in a noisy multi-class setting, also occurs in a multi-label setting. That is, the model first learns the representation of clean labels, and then starts memorizing noisy labels. Based on this finding, we propose novel methods for WSML which reject or correct the large loss samples to prevent model from memorizing the noisy label. Without heavy and complex components, our proposed methods outperform previous state-of-the-art WSML methods on several partial label settings including Pascal VOC 2012, MS COCO, NUSWIDE, CUB, and OpenImages V3 datasets. Various analysis also show that our methodology actually works well, validating that treating large loss properly matters in a weakly supervised multi-label classification. Our code is available at https://github.com/snucml/LargeLossMatters.

介绍研究领域：弱监督下的多标签分类任务（WSML）

主要贡献一：首次将未发现标签认定为负标签、将任务转化为噪声多标签分类。

主要贡献二：首次将发现了多标签分类中存在“记忆效应”。

基于记忆效应的发现，作者利用拒绝或修正大损失的方法防止模型记住噪声标签。

简单高效的设计，在多个数据集上取得了优质的效果。

引言读解

Multi-label classification aims to find all existing objects or attributes in a single image. It is gaining attention since the real world is made up of a scene with multiple objects in it [28, 35]. Moreover, some of the single-label datasets, also called multi-class datasets, actually have images containing multiple objects [33, 56]. However, the multi-label classification task has some fundamental difficulties in making a dataset because it requires annotators to label all categories' existence/absence for every image. As the number of categories and images in the dataset increase, annotation cost becomes tremendous [19].

To alleviate these issues, weakly supervised learning approach in multi-label classification task (WSML) has been taken into consideration [2, 18, 36, 50]. In a WSML setting, labels are given as a form of partial label, which means only a small amount of categories is annotated per image. This setting reflects the recently released large-scale multi-label datasets [12, 19] which provide only partial label. Thus, it is becoming increasingly important to develop learning strategies with partial labels.

多标签分类问题：因为世界是由多种多样的视觉特征组成的，因此多标签分类受到重视。

甚至现有的单标签数据集，也会存在一个图片实际包含多个标签的现象。

挖坑一：多标签分类任务依赖巨量的标注，费时费力。

为了解决这个问题，弱监督多标签分类任务出现了（WSML）

WSML设定：一张图片只会标注部分标签。

引言读解

There are two naive approaches to train the model with partial labels. One is to train the model with observed labels only, ignoring the unobserved labels. The other is to assume all unobserved labels are negative and incorporate them into training because majorities of labels are negative in a multi-label setting [32]. As the second one has a limitation that this assumption produces some noise in a label which hampers the model learning, previous works [7,9,16,21] mostly follow the first approach and try to explore the cue of unobserved labels using various techniques such as bootstrapping or regularization. However, these approaches include heavy computation or complex optimization pipeline.

We hypothesize that if label noise can be handled properly, the second approach could be a good starting point because it has the advantage of incorporating many true negative labels into model training. Therefore, we try to look at the WSML problem from the perspective of noisy label learning.

有俩种传统的方法解决WSML场景的问题。
方法一：只使用已经发现的标签进行训练。

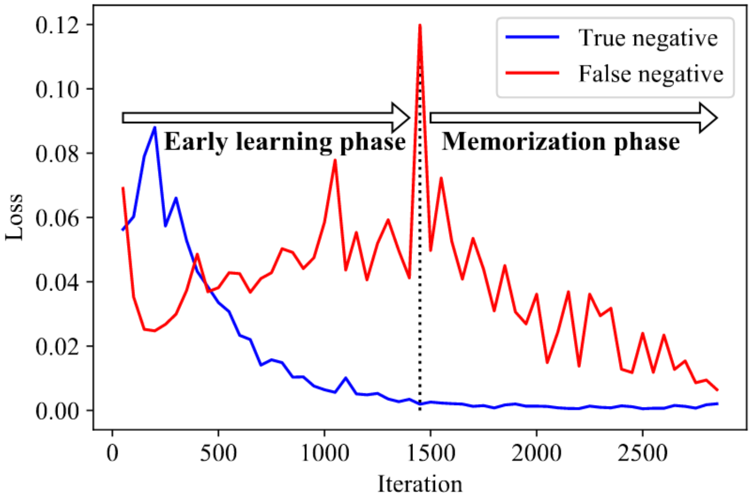
方法二：假定未发现的标签是负标签进行训练，因为多标签设定下大多是负标签。

挖坑二：方法一前沿的技术依赖巨大算力和具有较大优化难度，方法二又会引入噪声。

如果能合适地解决方法二的噪声问题，是一个更好的起点：它利用了更多真阴性标签进行训练。

引言读解

Our key observation is about the memorization effect [1] in a noisy label learning literature. It is known that when training a model with a noisy label, the model fits into clean labels first and then starts memorizing noisy labels. Although previous work showed the memorization effect only in a noisy *multi-class* classification scenario, we found for the first time that this same effect also happens in a noisy *multi-label* classification scenario. As shown in Figure 1, during training, the loss value from the clean label (true negative) decreases from the beginning while the loss from the noisy label (false negative) decreases from the middle.



顺着上段：如何合适地解决噪声标签问题继续往下说。

本文核心贡献是发现“多标签学习的记忆效应”，此前记忆效应只出现在多分类任务中。

对于真阴性的标签，其训练时损失一直保持下降；对于假阴性的标签，损失先上升后下降。

- 模型会在前期记住干净的标签。
- 模型会在中后期才会记住噪声的标签。

引言读解

Based on this finding, we borrow the idea from noisy multi-class literature [13, 17, 23] which selectively trains the model with samples having small loss and adapt this idea into a multi-label scenario. Specifically, by assigning the unknown labels as negative in a WSML setting, label noise appears in the form of false negative. Then we develop the three different schemes to prevent false negative labels from being memorized into the multi-label classification model by rejecting or correcting large loss samples during training.

一段话总结技术。

基于这一发现，我们借鉴了**噪声多类别**领域文献中的思想，即通过选择性地训练具有**小损失**的样本，并将这一思想应用于多标签情境。具体来说，通过在WSML设置中将未知标签视为负标签，**标签噪声以假阴性的形式**出现。然后，我们开发了三种不同的方案，通过在训练过程中拒绝或修正大损失样本，防止假阴性标签被记忆到多标签分类模型中。

引言读解

Our method is light and simple, yet effective. It involves negligible computation overhead and does not require complex optimization for model training. Nonetheless, our method surpasses the weakly supervised multi-label classification performance compared to the state-of-the-art methods in Pascal VOC 2012 [10], MS COCO [24], NUSWIDE [6], CUB [42], and OpenImages V3 [19] datasets. Moreover, while some existing methods are only effective in specific partial label setting [7, 9, 16], our method is broadly applicable in both artificially created and real partial label datasets. Finally, we provide some analysis about the reason why our methods work well from various perspectives.

To sum up, our contributions are as follows;

1) We empirically show for the first time that the memorization effect occurs during noisy multi-label classification.

2) We propose a novel scheme for weakly supervised multi-label classification that explicitly utilizes a learning technique with noisy label.

3) Although light and simple, our proposed method achieves state-of-the-art classification performance on various partial label datasets.

总结方法的优势。

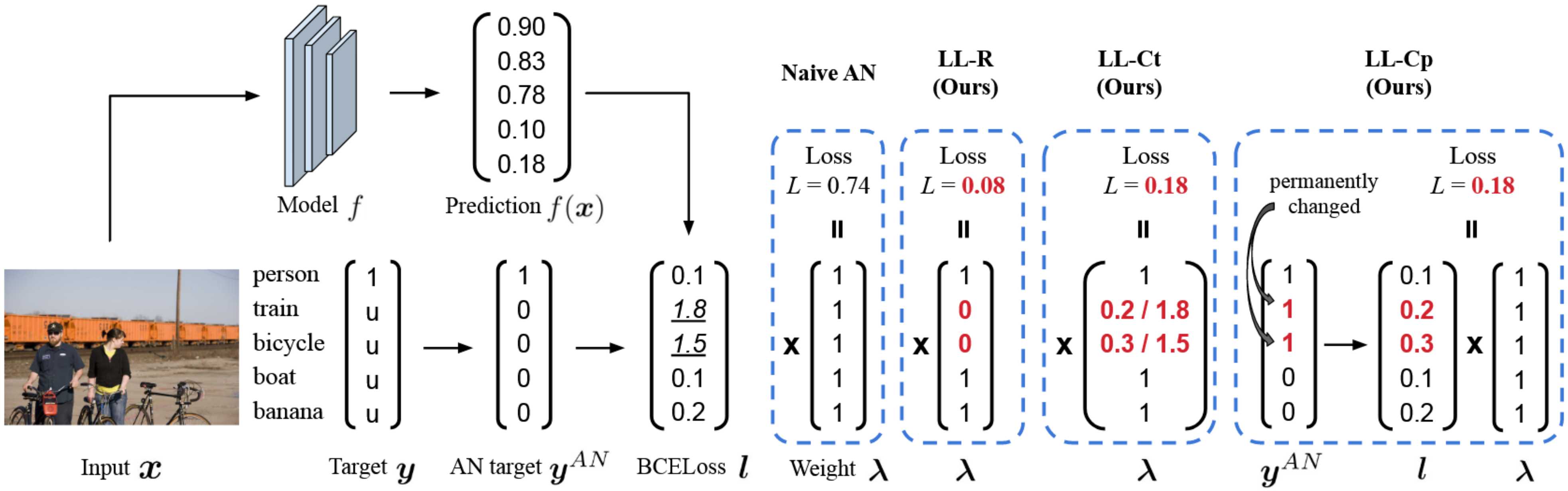
我们的方法轻量、简单并且有效，既不需要巨大的算力、也不需要复杂的优化。

尽管如此，我们的方法在公开数据集弱多标签分类任务的表现超过了现有技术。

我们的贡献可以总结如下：

- 首次发现多标签分类中的记忆效应
- 将弱监督多标签任务转化为噪声多标签任务
- 模型轻量且高效，超越了SOTA。

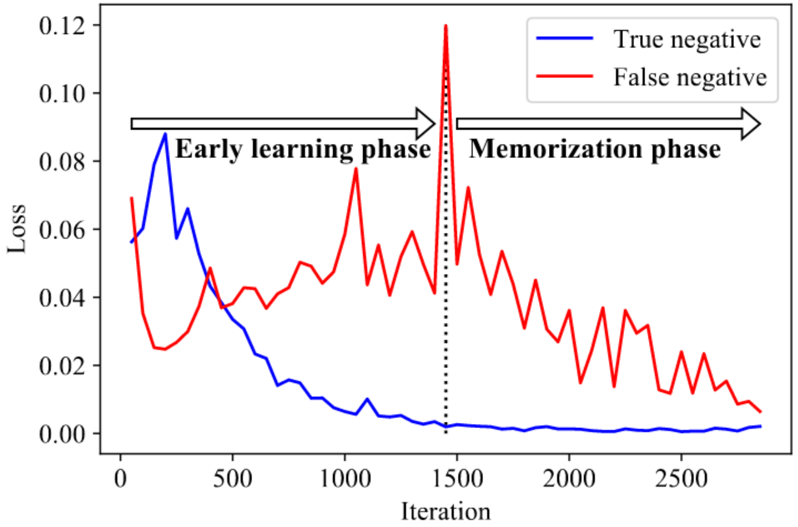
方法学-算法流程图



对于一个样本x，它的标注的标签为 Targrt y，我们首先假定y中所有**缺失的标签**都是负标签进而获得y^{AN}。然后我们通过一个预测器f来预测y^{AN}，得到$g(x)$,之后计算它们的交叉熵$l = \mathrm{BCE}(f(x), y^{AN})$。接着我们通过三种修正方法来修正$l$，分别是 **大损失标签拒绝(LL-R)**、**大损失标签临时修正(LL-Ct)**、**大损失标签修正(LL-Cp)**。假设有N个训练样本，每个样本有K个标签。

方法学-记忆效应

不同标签损失统计



最大损失出现在不同阶段的标签数量统计

Highest loss phase	Pascal VOC (%)			MS COCO (%)		
	TP	TN	FN	TP	TN	FN
Warmup	**88.3**	**90.7**	23.8	**64.0**	**82.6**	17.3
Regular	11.7	9.3	**72.2**	36.0	17.4	**82.7**

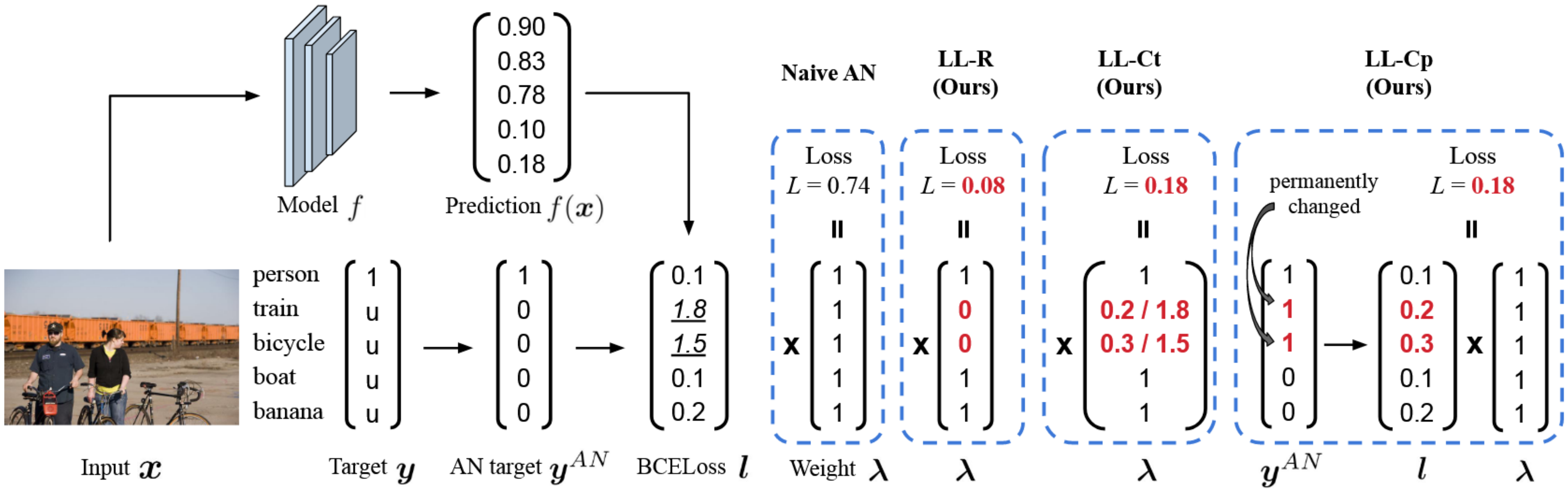
TP：真阳性　　TN：真阴性　　TN：假阴性（噪声）

Warmup：预热训练阶段　　Regular：常规训练阶段

期:
- TN标签损失正常下降；FN标签损失先上升
- TP和TN标签损失通常在预热阶段损失到达最大值。

模型训练中后期:
- FN标签损失开始下降
- FN标签损失通常在常规阶段达到最大值。

方法学-算法流程图



对于一个样本$\boldsymbol{x}$，它的标注的标签为 Targrt $\boldsymbol{y}$，我们首先假定$\boldsymbol{y}$中所有**缺失的标签**都是负标签进而获得$\boldsymbol{y}^{AN}$。然后我们通过一个预测器$\boldsymbol{f}$来预测$\boldsymbol{y}^{AN}$，得到$g(\boldsymbol{x})$,之后计算它们的交叉熵$\boldsymbol{l} = \mathrm{BCE}(f(\boldsymbol{x}), \boldsymbol{y}^{AN})$。接着我们通过三种修正方法来修正$\boldsymbol{l}$，分别是**大损失标签拒绝(LL-R)**、**大损失标签临时修正(LL-Ct)**、**大损失标签修正(LL-Cp)**。假设有N个训练样本，每个样本有K个标签。

方法学-LLR

Large loss rejection： 大损失拒绝法

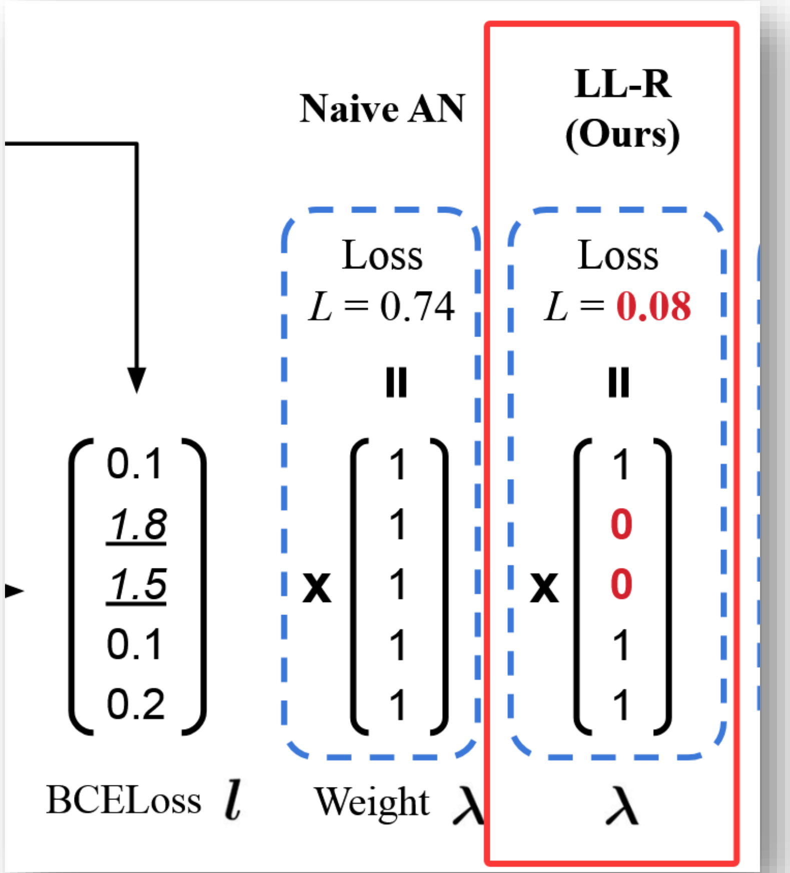
$$L=\frac{1}{|\mathcal{D}'|}\sum_{(\boldsymbol{x},\boldsymbol{y}^{AN})\in\mathcal{D}'}\frac{1}{K}\sum_{i=1}^{K} l_i\times\lambda_i\,.$$

直接拒绝大损失的标签，即将大损失标签的权值设置为0。

$$\lambda_i=\begin{cases}0, & i\in\mathcal{S}^u \text{ and } l_i>R(t)\\ 1, & \text{otherwise},\end{cases}$$

其中阈值 $R(t)$ 是前 $[(t-1)\cdot\Delta_{rel}]\%$ 大的标签损失值（未标注的标签中）Δ_{rel} 是超参数。 t 是训练轮次。

第一轮不会拒绝任何标签、随着轮次进行，拒绝标签更多。



方法学-*LL-Ct*

Large loss correction (temporary). ：大损失修正法（暂时修正）

$$L = \frac{1}{|\mathcal{D}'|} \sum_{(\boldsymbol{x}, \boldsymbol{y}^{AN}) \in \mathcal{D}'} \frac{1}{K} \sum_{i=1}^{K} l_i \times \lambda_i \,.$$

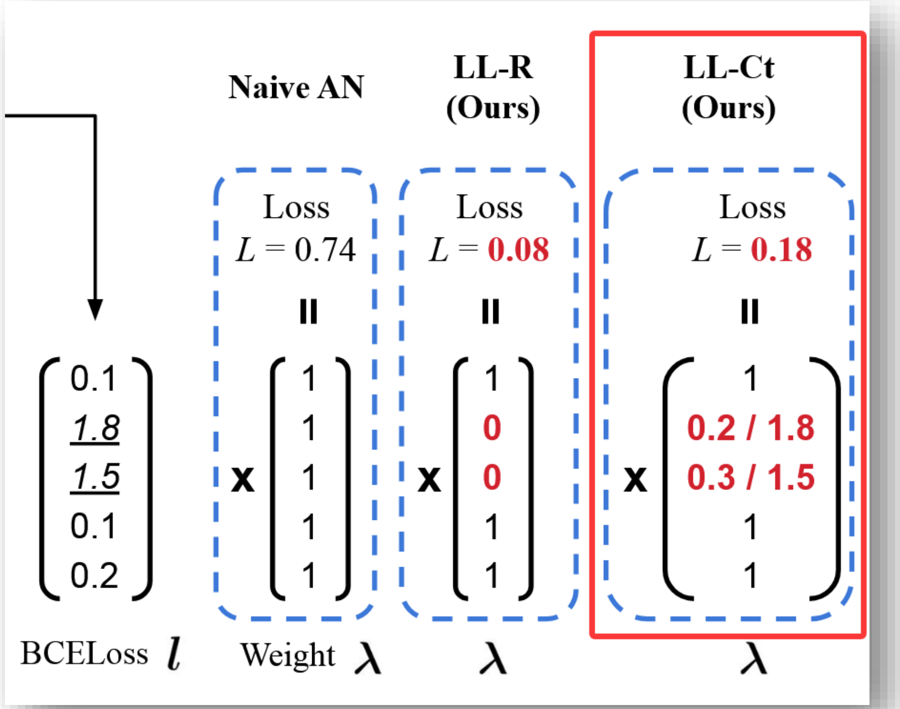
直接修正大损失的标签，即将大损失标签的标签暂时置1。

$$\lambda_i = \begin{cases} \frac{\log f(\boldsymbol{x})_i}{\log(1 - f(\boldsymbol{x})_i)}, & i \in \mathcal{S}^u \text{ and } l_i > R(t) \\ 1, & \text{otherwise}\,, \end{cases}$$

如上权值设计等效于短暂将标签置1计算损失，证明如下：

$$\begin{aligned} l_i \times \lambda_i &= \mathrm{BCELoss}(f(\boldsymbol{x})_i, y_i^{AN} = 0) \times \lambda_i \\ &= -\log(1 - f(\boldsymbol{x})_i) \times \lambda_i \\ &= -\log f(\boldsymbol{x})_i \\ &= \mathrm{BCELoss}(f(\boldsymbol{x})_i, 1)\,. \end{aligned}$$

$R(t)$ 设计与LLR相同。



方法学-*LL-Cp*

Large loss correction (permanent)：大损失修正法（永久修正）

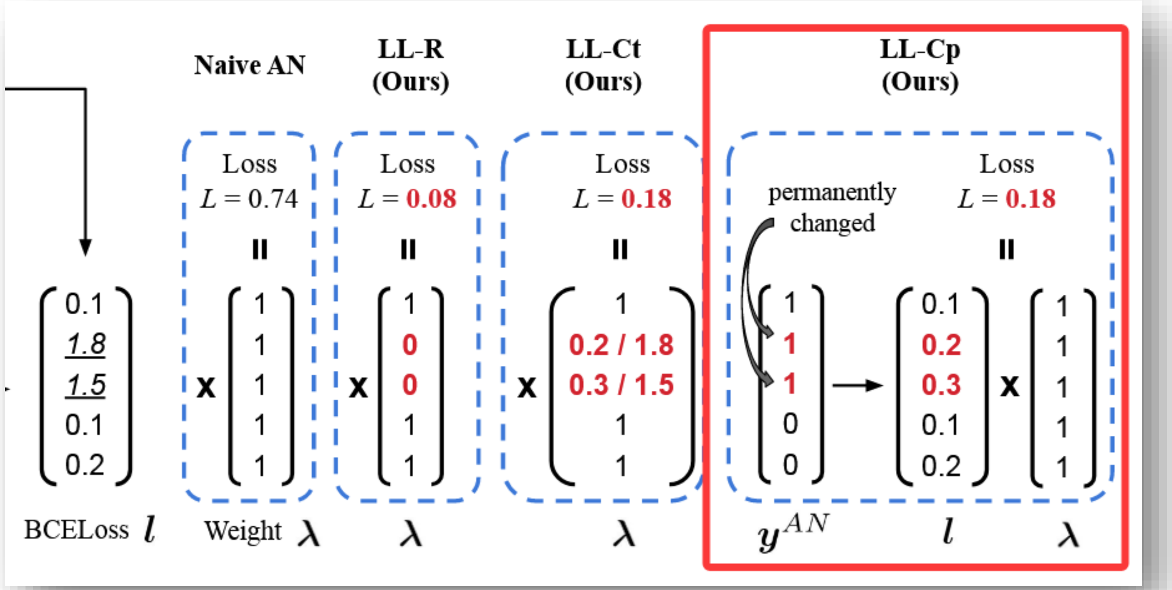
$$L = \frac{1}{|\mathcal{D}'|} \sum_{(\boldsymbol{x}, \boldsymbol{y}^{AN}) \in \mathcal{D}'} \frac{1}{K} \sum_{i=1}^{K} l_i \times \lambda_i \,.$$

直接修正大损失的标签，即将大损失标签的标签永久置1。

$$y_i^{AN} = \begin{cases} 1, & i \in \mathcal{S}^u \text{ and } l_i > R(t) \\ unchanged, & \text{otherwise}\,, \end{cases}$$

$R(t)$ 是前 $\Delta_{rel}\%$ 大的标签损失值，其是固定值。

之后轮次该标签将永久为1参与损失计算。



数据集-人工创建不完整数据集

数据集说明：将现有的多标签数据集中的部分标签进行删除，从而构建不完整多标签数据集。具体来说，我们从现有的多标签数据集中只随机保留一个**正标签**，其余标签全部删除。所使用的数据集包括 `VOC 2012`、`MS COCO 2014`、`NUSWIDE` 和 `CUB`。其中 `CUB` 是一个鸟类分类的数据集，**注意**这里我们不用于预测鸟类分类任务，而是用于预测鸟类的多标签特征例如颜色、形状等。

数据处理：本文具体对数据进行了如下处理：

1. 使用相同的**种子数**来随机删除标签，以确保每个数据集的不完整多标签数据集是相同的。
2. 测试模型选用在 `ImageNet` 上预训练的 `ResNet-50` 模型。
3. 模型训练的Batch Size设置为 16。
4. 训练时采取**随机水平翻转**进行数据增强。
5. 所有输入图像的大小都调整为 448x448。
6. 设定俩种训练策略，一种是 `LinearInit`，另一种是 `End-to-end`,其中LinearInit是在第一轮训练时，只训练模型最后一层，其余层的参数保持不变；End-to-end是直接训练整个模型。

对比方法：本文将提出的三种方法的俩种训练模式分别与以下方法进行对比：`Full label`：直接使用完整标签训练模型；`Naive AN`、`Weak AN(WAN)`、Label Smoothing with AN `LSAN`、和 `ROLE`。以上均是**第二种解决流派**的前沿方法；由于实验设定与第一种解决流派的前沿方法不同，因此不进行对比。

数据集-真实的不完整数据集

数据集说明：这里使用了 OpenImages V3 数据集，该数据集包含了**3.4M**的训练图像、**42K**的验证图像和**125K**的测试图像，并且有**5000**个类别，且只有不到1%的标签是被标注的。

数据处理：本文具体对数据进行了如下处理：

1. 测试模型选用**ImageNet-pretrained ResNet-101**模型。
2. 模型训练的Batch Size设置为 288，使用4个GPU。
3. 所有输入图像的大小都调整为 224x224。
4. 训练时采取**随机水平翻转**进行数据增强。
5. 由于不同类别对应的图像数据的量是不同的，作者将数据集分为了5组，每一组有1000个类别，其中Group 1是数据量最小，Group 5是数据量最大。

对比方法：本文将提出的三种方法的俩种训练模式分别与以下方法进行对比：第一种流派技术 **Naive IU**、**Curriculum labeling**、**IMCL**。第二种流派技术 **Naive AN**、**WAN**、**LSAN**。

SOTA对比实验

- 人工创建的不完整数据集
- 真实存在的不完整数据集

SOTA对比实验-人工创建的不完整数据集

Method	End-to-end				LinearInit.			
	VOC	COCO	NUSWIDE	CUB	VOC	COCO	NUSWIDE	CUB
Full label	90.2	78.0	54.5	32.9	91.1	77.2	54.9	34.0
Naive AN	85.1	64.1	42.0	19.1	86.9	68.7	47.6	20.9
WAN [7, 27]	86.5	64.8	46.3	20.3	87.1	68.0	47.5	21.1
LSAN [7, 37]	86.7	66.9	44.9	17.9	86.5	69.2	50.5	16.6
EPR [7]	85.5	63.3	46.0	20.0	84.9	66.8	48.1	21.2
ROLE [7]	87.9	66.3	43.1	15.0	88.2	69.0	**51.0**	16.8
LL-R (Ours)	**89.2**	**71.0**	47.4	19.5	**89.4**	**71.9**	49.1	21.5
LL-Ct (Ours)	89.0	70.5	48.0	**20.4**	89.3	71.6	49.6	**21.8**
LL-Cp (Ours)	88.4	70.7	**48.3**	20.1	88.3	71.0	49.4	21.4

使用完整标签训练的模型的结果在第二行，以显示 WSML 的上限。“End-to-end” 表示模型的整个权重从一开始就进行了微调，而 “LinearInit.” 表示主干在前几个 epoch 中被冻结。LL-Ct 在 8 种设置中的 7 种情况下优于所有基线方法，而 LL-R 和 LL-Cp 在 8 种设置中的 6 种情况下优于所有基线方法。

SOTA对比实验-真实存在的不完整标签数据集

第一类方法：只使用已经观察的标签训练

第二类方法：将未观察的标签转化为0

Method	G1	G2	G3	G4	G5	All Gs
Naive IU	69.5	70.3	74.8	79.2	85.5	75.9
Curriculum [9]	70.4	71.3	76.2	80.5	86.8	77.1
IMCL [16]	71.0	72.6	77.6	81.8	87.3	78.1
Naive AN	77.1	78.7	81.5	84.1	88.8	82.0
WAN [7, 27]	71.8	72.8	76.3	79.7	84.7	77.0
LSAN [7, 37]	68.4	69.3	73.7	77.9	85.6	75.0
LL-R (Ours)	77.4	79.1	82.0	84.5	89.5	82.5
LL-Ct (Ours)	77.7	79.3	82.1	84.7	89.4	**82.6**
LL-Cp (Ours)	77.6	79.1	81.9	84.6	89.4	82.5

因为每张图像的观察到的平均类别数远小于完整类别的数量，这阻碍了模型在仅使用有限数量的观察到的标签进行训练时推广到看不见的数据。

5000 个类别根据已知该类别标签的训练图像数量按升序排序，然后从 Group1 到 Group5 按顺序分组，所有组具有相同的大小。所有 G 对应于所有类别的集合。我们观察到 LL-Ct 的性能最好，其次是 LL-Cp 和 LL-R。

消融实验

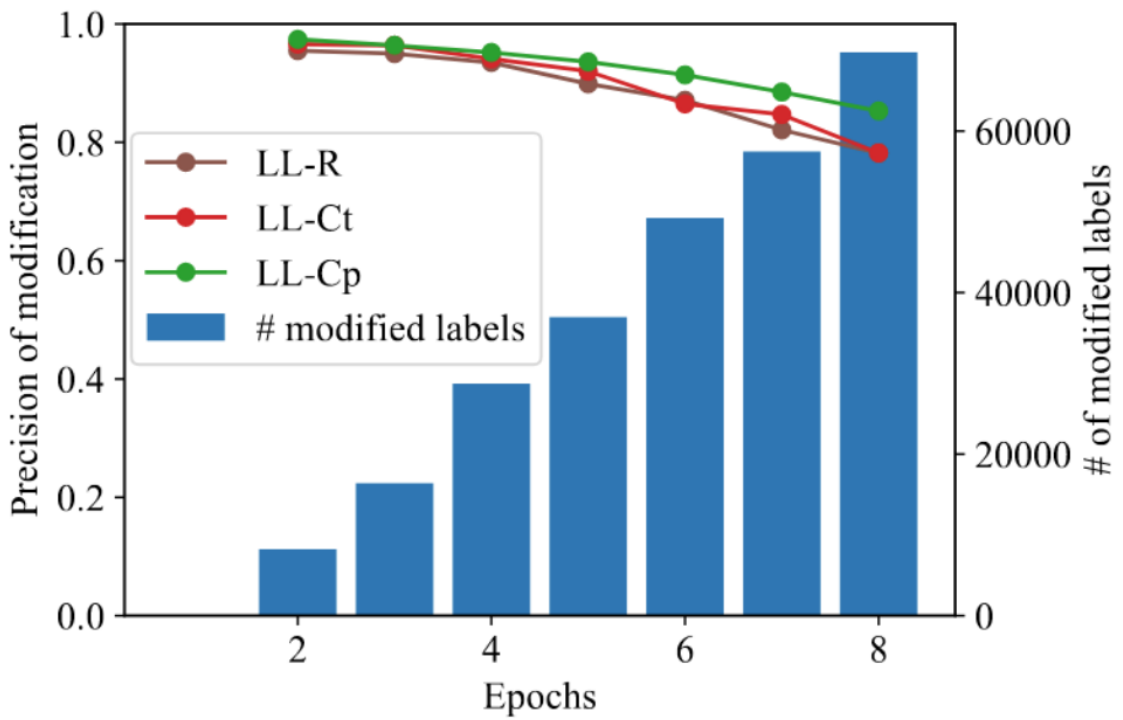
- 损失调整的有效性
- 超参数 $\Delta_{rel}\%$ 有效性

消融实验-损失修正的有效性

作者定义了损失修正有效性的指标：**修正准确率（Precision of modification）**

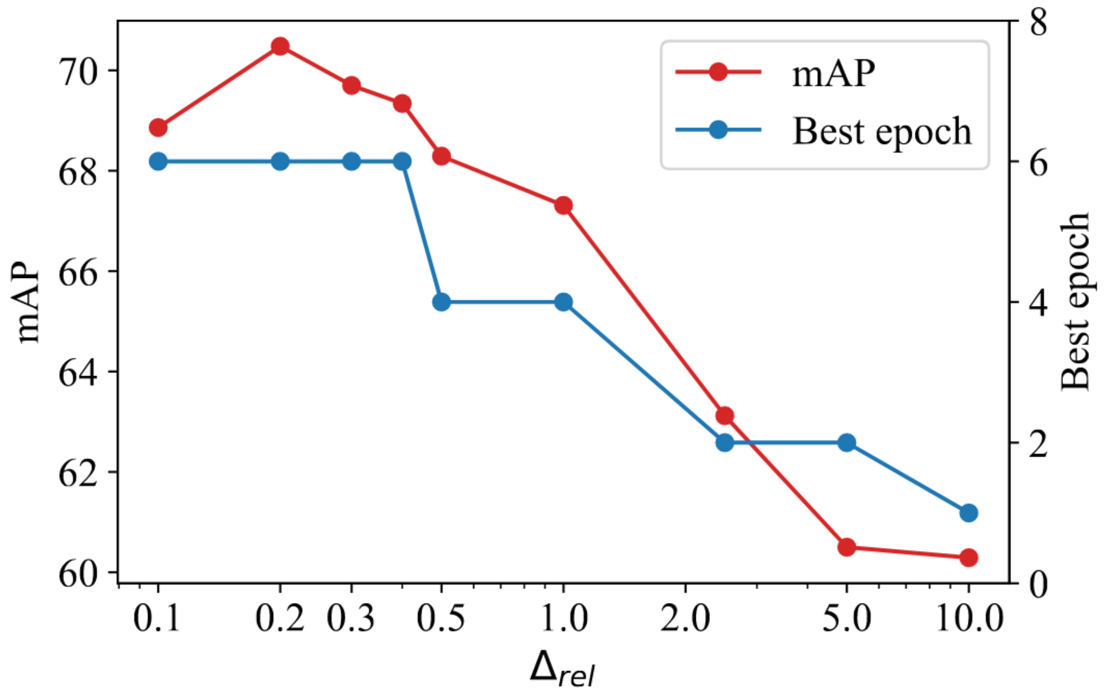
$$\text{修正准确率} = \frac{\text{成功修正或拒绝的标签数量}}{\text{模型所有修正或拒绝的标签数量}}$$

- 我们的方法以高准确的修正率有效地修正了假阴性标签。
- 随着epoch的增加，准确率下降，因为模型会逐渐记住错误的标签。



消融实验-超参数的有效性

- mAP情况下最有效的 Δ_{rel} 是 0.2，较小的拒绝率和较大的拒绝率都会导致mAP下降。
- 较小的拒绝率导致模型不能及时拒绝或修正假阴性的标签，导致性能下降。
- 较大的拒绝率导致模型错误地拒绝或修正真阴性的标签。



定性实验-可视化

LL-Ct:　Given：给出的正标签；箭头：随着训练模型判断的正标签；GT：图片真实的正标签

Given : fire hydrant
➡ fire hydrant, car
➡ fire hydrant, car, person, bicycle

GT : fire hydrant, car, person, bicycle

Given : banana
➡ banana, orange
➡ banana, orange, bowl

GT : banana, orange, bowl

Given : vase
➡ vase, person
➡ vase, person, chair
➡ vase, person, chair, dining table

GT : vase, person, chair, dining table, bottle, wine glass

Given : truck
➡ truck, car
➡ truck, car, person

GT : truck, boat, motorcycle

- 我们展示了 LL-Ct 正确修改未注释的真值标签的三种情况，以及第四列的失败情况。
- 尽管没有给出所有的真阳性标签，模型可以逐渐将未注释的GT类别更正为阳性。
- 我们还观察到LL-Ct在一个epoch进行临时修正的标签，再之后的epoch仍会继续修正该标签。
- 我们还在最右侧报告了我们方法的失败案例，其中模型将汽车混淆为卡车，这是一个类似的类别，并误解了不存在的类别的人。

可解释性实验

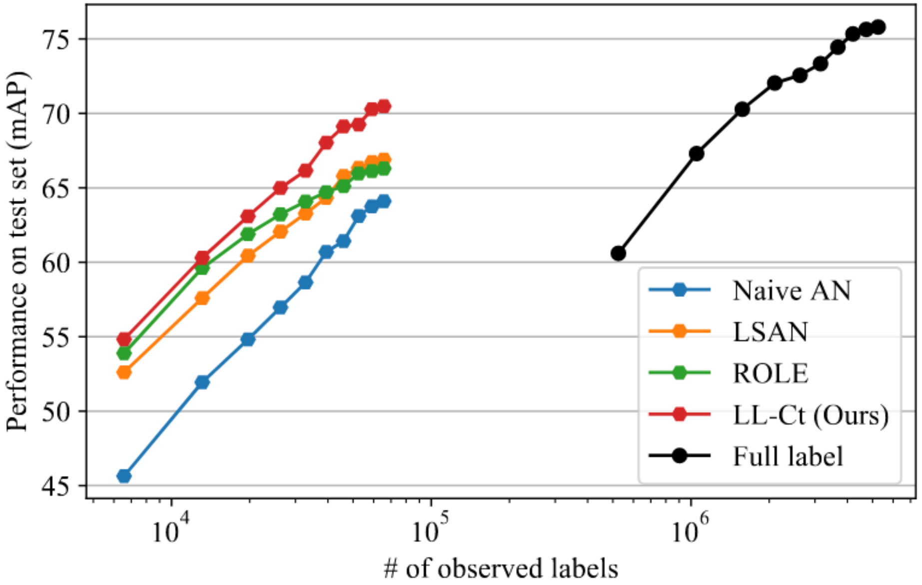
- 为了研究模型表现优异是否与其对数据更好的理解相关，作者探讨了模型的解释与人类推理过程的关系。
- 作者使用类激活映射（Class Activation Mapping，简称CAM）来表示模型的解释，而ground truth物体（即图像中的实际物体）则被视为人类的解释。
- Pointing Game: 为了衡量模型的解释与人类的解释对齐程度，作者使用了指向游戏度量。在该度量中，当模型的CAM的最大值对应的像素点位于物体的边界框内时，算作“Hit”命中；如果不在边界框内，则算作“Miss”漏检。
- 对于每个类别，作者统计所有测试数据中的“Hit”和“Miss”数量，并计算每个类别的平均“Hit”率（即Hit的数量与Hit+Miss总和的比值），最终以百分比形式报告。

Method	VOC	COCO
Naive AN	78.9	46.4
WAN [7, 27]	79.8	47.7
LSAN [7, 37]	79.5	49.1
EPR [7]	80.2	48.1
ROLE [7]	82.5	51.5
LL-R (Ours)	**83.7**	54.0
LL-Ct (Ours)	**83.7**	**54.1**
LL-Cp (Ours)	83.5	53.3

Table 4. **Pointing Game.**

小数据训练实验

- COCO按其10%，20%...100%的数据量进行实验。
- 尽管10%数据量的全监督标签数相当于8倍的100%数据量的弱监督标签数，LL-Ct的性能已经和全监督方法相当。
- LL-Ct在所有观察到的标注量的mAP均优于其他弱监督方法。



总结

- 本文提出了大型损失修改方案，这些方案拒绝或纠正在训练带有部分标记注释的多标签分类模型期间出现的大损失样本。
- 本文创新地将弱监督的多标签分类任务转化为噪声的多标签分类任务。
- 本文首次提出记忆效应也发生在嘈杂的多标签分类场景中。
- 虽然不包括沉重和复杂的组件，本文的方案成功地防止了多标签分类模型记住假阴性标签，在各种部分标记的多标签数据集上实现了最先进的性能。

好词好句

- 1) we empirically observe that memorization effect, which was first discovered in a noisy multi-class setting, also occurs in a multi-label setting.
  - 通过实验证明…，第一次发现…
- 2) To alleviate these issues, weakly supervised learning approach in multi-label classification task (WSML) has been taken into consideration.
  - To alleviate these issues +被动句：为了解决这些问题，…被提出。

好词好句

- 3) As the number of categories and images in the dataset increase, annotation cost becomes tremendous.
  - 巨大的标注开销
- 4) There are two naive approaches to train the model with partial labels. One is to train the model with observed labels only, ignoring the unobserved labels. The other is to assume all unobserved labels are negative and incorporate them into training because majorities of labels are negative in a multilabel setting
  - 俩种传统方法流派去解决…
  - 其中一种是…另外一种是…

好词好句

- 5) Our method is light and simple, yet effective. It involves negligible computation overhead and does not require complex optimization for model training.
  - negligible computation：可忽略不计的计算量
- 6) we borrow the idea from noisy multi-class literature which selectively trains the model with samples having small loss and adapt this idea into a multi-label scenario.
  - 描述借用别人的idea的方法
  - Borrow the idea from
  - Adapt this idea into

好词好句

- 7) Our method is light and simple, yet effective. It involves negligible computation overhead and does not require complex optimization for model training.
  - negligible computation：可忽略不计的计算量
- 8) Concisely, we regard the class activation mapping (CAM) [58] as the model's explanation and the ground truth object as the human's explanation.
  - 简而言之，可以替代specifically。

好词好句

- 9) Our method is light and simple, yet effective. It involves negligible computation overhead and does not require complex optimization for model training.
  - 我们的方法虽然轻量，但效果显著
- 10) Our methodology makes one step progress towards dealing with noisy multi-label classification.
  - 我们的方法在…问题上迈出了一步。

观点论据

- 1) Multi-label classification aims to find all existing objects or attributes in a single image. It is gaining attention since the real world is made up of a scene with multiple objects in it Moreover, some of the single-label datasets, also called multi-class datasets, actually have images containing multiple objects.
- 多标签分类的动机
    - 多标签分类旨在在单张图像中找到所有存在的物体或属性。由于现实世界是由包含多个物体的场景组成的，因此这一任务越来越受到关注。此外，一些单标签数据集，也被称为多类别数据集，实际上包含了多个物体的图像。

观点论据

- 2) However, the multi-label classification task has some fundamental difficulties in making a dataset because it requires annotators to label all categories' existence/absence for every image. As the number of categories and images in the dataset increase, annotation cost becomes tremendous.
- **多标签分类挖坑-标注成本大**
  - 然而，多标签分类任务在构建数据集时存在一些基本困难，因为它要求标注人员为每张图像标注所有类别的存在/缺失。随着数据集中类别和图像数量的增加，标注成本变得非常巨大。

观点论据

- 3) Our key observation is about the memorization effect in a noisy label learning literature. It is known that when training a model with a noisy label, the model fits into clean labels first and then starts memorizing noisy labels. Although previous work showed the memorization effect only in a noisy multi-class classification scenario, we found for the first time that this same effect also happens in a noisy multi-label classification scenario.
- **记忆效应描述**
  - 我们主要的发现是关于噪声标签学习中的记忆效应。众所周知，在使用噪声标签训练模型时，模型首先会适应干净标签，然后才开始记忆噪声标签。尽管之前的研究仅在噪声多类分类场景中展示了记忆效应，但我们首次发现这种效应也出现在噪声多标签分类场景中。

观点论据

- 4) There are two naive approaches to train the model with partial labels. One is to train the model with observed labels only, ignoring the unobserved labels. The other is to assume all unobserved labels are negative and incorporate them into training because majorities of labels are negative in a multilabel setting .
- **解决遗漏标签的方法流派**
  - 使用部分标签训练模型有两种简单的方法。 一种是只用观察到的标签来训练模型，忽略未观察到的标签。 另一种是假设所有未观察到的标签都是负的，并将其纳入训练，因为在多标签设置中，大部分标签都是负的。

观点论据

- 5) We hypothesize that if label noise can be handled properly, the second approach could be a good starting point because it has the advantage of incorporating many true negative labels into model training. Therefore, we try to look at the WSML problem from the perspective of noisy label learning.
- 假定负标签-动机
    - 我们假设如果能妥善处理标签噪声，第二种方法（假定负标签）可能是一个很好的起点，因为它的优势在于能将许多真实的负标签纳入模型训练。 因此，我们尝试从噪声标签学习的角度来看待 WSML 问题。