



# GBRAIN:基于颗粒球的鲁棒训练对抗文本标签噪声

Zeli王

重庆邮电大学网络空间大数据智  
能安全教育部重点实验室

中国重庆zlwang@  
cqupt.edu.

cnTuo张

网络空间大重点实验室  
数据智能安全  
教育部, 重庆大学  
邮电  
中国重庆

s210231260@stu.cqupt.edu

.cnShuyin夏\*

网络空间大重点实验室  
数据智能安全  
教育部, 重庆大学  
邮电

中国重庆  
xiasy@cqupt.edu.cn

龙龙的林

计算机与信息学院  
西南大学理学院  
中国重庆

longlonglin@swu.edu.cn

Guoyin王

重庆邮电大学网络空间大数据智  
能安全教育部重点实验室

中国重庆wanggy@  
cqupt.edu.cn

## 摘要

大多数自然语言处理任务依赖大量标记数据来训练优秀的神经网络模型。然而,在标注大规模文本数据集时,不可避免地会引入标签噪声(即错误标签),这会显著降低神经网络模型的性能。为克服这一困境,本文提出了一种新的基于粒球的 $t_{RAIN}$ ing框架GBRAIN,以实现鲁棒的粗粒度表示学习,从而在不同的文本任务中对抗标签噪声。具体来说,考虑到数据集中的大多数样本都是精确标记的,GBRAIN首先提出了一种动态的粒球聚类算法,以无缝融入传统的神经网络模型。该聚类算法的一个显著特征是,它可以自适应地将相似数据的嵌入向量分组到同一个集合中(以下简称为粒球)。来自同一集合的所有样本的嵌入向量和标签将分别由中心向量和粒球的标签进行粗粒度表示。因此,噪声标签可以通过大多数带标签数据的标签进行纠正。此外,我们引入了一种与我们的框架兼容的新的梯度反向传播机制,它可以帮助通过迭代训练优化粗粒度嵌入向量。文本分类和名称实体识别任务的实验结果表明,与最先进的基线相比,我们的建议GBRAIN是确实有效。

## CCS的概念

· 计算方法→自然语言处理。

## 关键字

自然处理语言, 标签噪声, 聚类, 表示学习

## ACM参考格式:

王泽力、张拓、夏淑银、林龙龙、王国银。2024。GBRAIN:基于颗粒球的鲁棒训练对抗文本标签噪声。2024年国际多媒体检索会议论文集(ICMR' 24), 2024年6月10-14日, 泰国普吉岛。ACM, 美国纽约, 9页。https://doi.org/10.1145/3652583.3658084

## 1 介绍

随着计算能力的不断增长,大规模数据集的出现使得深度神经网络(Deep neural Networks, dnn)在各个领域取得了显著的成功。众所周知,深度神经网络的性能严重依赖于大量高质量的非标记数据集。然而,由于人工标注中的人为疏忽或自动标注中的不准确,标注过程将不可避免地引入错误。标签噪声严重拖累了深度神经网络的性能,因为模型学习了不正确的关联和模式。标签噪声的研究受到了学术界和工业界的广泛关注,特别是在计算机视觉(CV)领域[4,11,12,19,24,29,39,44]。

然而,由于文本的不连续性和模糊性,CV中的解决方案不能直接应用于自然语言处理(NLP)任务。很少有研究致力于增强深度神经网络模型对文本标签噪声的恢复能力,主要是通过鲁棒优化[6,25]和噪声滤波[20,26]。前者旨在通过在训练过程中引入对抗性损失[25]或优化目标[6]来增强模型的鲁棒性。后者侧重于识别和消除标签噪声。该方法首先利用模型预测概率[12]、样本不确定性[9]或性能指标[14]来评估标签是否受到噪声的影响,然后用删除或纠正注释

\*Corresponding author

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

潜在噪声[20]。然而，这些方法仍然面临局限性。例如，它们通常处理低维样本，但难以处理在深度神经网络应用中经常出现的高维向量[8,13,40];在过滤大规模标签噪声[41]时，它们可能具有较差的可扩展性。

为了减轻这些限制，我们的目标是提出一个新的框架，可以为NLP任务提供鲁棒性。然而，我们面临着两个重大挑战。如何提高标签噪声下文本表示学习的准确性?表征学习的精度决定了NLP任务的有效性。具体来说，自然语言是复杂的，具有多样化的上下文和语义信息。它具有歧义的词汇和语法结构，可以有多种解释，使其难以学习精确的嵌入表示。此外，识别标签是否对NLP是困难的。在计算机视觉中，可以通过比较生成的和真实的图像来评估准确性。相比之下，文本的含义和标签更加主观，建立准确的标签质量衡量标准是不可能的。因此，在过滤所有标签噪声后进行表示学习并不十分现实。综上所述，确保噪声下表征学习的准确性对于NLP任务来说是非常重要的和具有挑战性的。我们如何保证深度神经网络模型的可扩展性，以有效地处理具有大量标签噪声的庞大数据集?深度学习依赖于大型标注数据集来训练一个优秀的神经网络。如上所述，由于标注过程中的人工错误或工具限制，可能会引入意料之外的标签噪声量。该模型必须是可扩展的，以处理庞大数据集和潜在的大量标签噪声。

本文提出GB<sub>RAIN</sub>，一个粗粒度表示学习框架，用于在带有标签噪声的数据集上训练一个鲁棒的神经模型。针对上述两个挑战，GB<sub>RAIN</sub>实现了一种基于粒球计算(Granular-Ball Computing, GBC)[33]的多粒度表示学习方法。GBC可以在保持数据分布特征的前提下，自适应地对相近的低维对象进行分组。然后它在簇上执行各种计算，以保证鲁棒性[5,34,36,37]。受此启发，GB<sub>RAIN</sub> designs一种用于高维文本嵌入的颗粒球聚类算法。欧几里得距离用于测量向量之间的接近度，以帮助聚类相似的文本。然而，在NLP中，每个嵌入都有数百维的特征向量。即使在两次嵌入之间的某些维度上的向量相距很远，也很容易被所有维度的向量均匀化。为了适应高维向量中的聚类，GB<sub>RAIN</sub> enhances传统的GB聚类方法通过在计算距离时自动为各个维度分配不同的权重。距离越远，维度的权重越大。因此，文本嵌入之间的距离就更加清晰。然后，聚类算法可以基于这些加权欧几里得距离自适应地收集相似文本，产生颗粒球。由于一个球中的所有样本具有相近的语义，因此它们的标签应该是一致的。GB<sub>RAIN</sub> regards球内最普遍的标签作为球内所有样本的标签。然后将一个球内的所有样本替换为一个样本，其嵌入和标签分别为中心向量和球的标签。因此训练集中的标签噪声可以通过球内最精确的标注数据来修正，而粗表示(球的中心向量)也反映了文本

的典型特征。当在这样的环境下进行训练时，模型的鲁棒性会更强。因此，第一个挑战得到了缓解。对于第二个挑战，因为球的中心向量将以粒状球的形式呈现所有样本，训练数据集将急剧下降。因此，GB<sub>RAIN</sub> will具有极佳的可扩展性。

GB<sub>RAIN</sub> consists的三个主要部分。首先，将标注好的数据输入GB<sub>RAIN</sub>的文本编码器，映射为高维特征向量。其次，颗粒球聚类层自适应地聚类相近的特征向量，形成具有相似上下文的特征球，并为每个球生成中心向量和标签。最后，将所有球的中心向量和标签视为粗表示的实例和标签，用于训练神经网络。实验在不同数据集的多个DNN模型上进行。我们评估GB<sub>RAIN</sub> on两种流行的NLP任务，包括命名实体识别(NER)和文本分类。对于NER任务，GB<sub>RAIN</sub> framework在预训练的屏蔽语言模型RoBERTa上进行评估。对于文本分类任务，GB<sub>RAIN</sub> is在BiLSTM、GRU和TextCNN等传统序列模型上进行了评估。此外，我们比较了GB<sub>RAIN</sub> with最新技术，即BOND [20]，SCDL[42]和花名册[25]。实验结果表明，GB<sub>RAIN</sub> presents具有良好的鲁棒性，在标签噪声下优于现有工作。简而言之，我们强调的主要贡献如下：

- 我们提出了一种有效的高维向量颗粒球聚类算法，该算法可以嵌入到不同的DNN模型中，在保持数据特征的情况下对相似的文本进行聚类。
- 我们提出了一个基于粗粒度表示的鲁棒训练框架，以帮助DNN模型对抗NLP中的标签噪声。
- 对流行的NLP任务的广泛评估表明，我们的提议比现有的工作提供了更好的鲁棒性。所发布的代码和数据集用于复制:<https://github.com/zhangtuo723/GBRAIN>。

## 2 相关工作

### 2.1 应对CV中标签噪声的对策

Sukhbaatar等[29]在神经网络中加入了一个额外的噪声层，以调整输出以匹配噪声分布。Manwani et al.[24]证明，使用损失函数0-1的风险最小化表现出噪声容忍特性，而平方误差损失仅容忍均匀噪声。Jindal et al.[18]通过加入对标签噪声统计进行建模的SoftMax层来增强标准的深度网络。Zhuang等人[44]提出了一种鲁棒的、弱监督的深度学习框架来对抗网络图像中的噪声标签。Wang等人[31]开发了一种对称交叉熵学习方法，利用与鲁棒噪声相对应的逆交叉熵对称地提高交叉熵。一些研究人员使用噪声过滤方法来减少标签噪声对模型的影响。Han et al.[12]引入了一种联合学习的噪声记忆方法，该方法涉及两个网络在小批量数据上协同学习，有效地过滤噪声样本。Yao et al.[39]提出了一种名为“搜索探索”的学习方法，与[12]中的手动设置相比，自适应地确定遗忘率。Azadi et al.[1]引入了一种辅助图像正则化技术，自动使用训练样本之间的相互上下文信息来鼓励模型

以选择可靠的样本来提高学习。Jiang et al.[17]通过提供课程(样本权重方案)来关注具有潜在正确标签的样本来监督基础深度网络(即StudentNet)的训练。Jie et al.[15]建议定期改变学习率,使模型在过拟合和欠拟合之间振荡,以检测噪声标签样本。Jindal et al.[19]引入了一个非线性处理层,对不正确标签的数据进行建模,防止模型过度拟合噪声。

## 2.2 NLP中标签噪声的对策

在NLP中,为了提高模型对标签噪声的鲁棒性,大多数研究者总是将鲁棒优化和噪声滤波相结合。Zhang et al.[42]提出自合作降噪学习,训练两个师生网络,每个网络通过自降噪使用可靠的标签,通过协同降噪探索不可靠的an-notation。Liang et al.[20]首先使用适应NER任务的预训练语言模型来提高精度,然后采用远程标签去除和自训练来增强鲁棒性。孟等人[25]提出了一种噪声-鲁棒学习方案,包括一个新的损失函数和一个噪声标签删除过程,训练网络模型对数据进行鲁棒性标记。Qiao et al.[27]提出了一种名为SelfMix的方法,通过利用高斯混合模型、半监督学习、dropout机制和文本级mixup训练策略,有效地解决了文本分类任务中的标签噪声。Cheng et al.[6]首先在训练集上进行热身过程,根据模型对不同噪声场景的适应性动态终止。然后使用具有各种泛化策略的混合训练阶段逐步纠正错误标记的实例。Garg et al.[9]引入了一个双成分beta混合模型,在使用去噪损失训练分类器和噪声模型之前,为每个样本分配具有干净或噪声标签的概率分数。

## 2.3 颗粒球计算

Wang将颗粒计算与science杂志[3]上的研究相结合,提出了颗粒认知计算[30]。在此基础上,Xia[33]提出了颗粒球计算(granular-ball computing, GBC),即根据数据的分布生成多个覆盖整个数据集的颗粒球(granular-ball, GB)。每个GB由一个中心和一个半径来描述,其中中心是球中所有物体向量的平均值,半径是GB中所有物体到其中心的最大或平均距离。所有取代原始对象的GBs都被用作下游任务的输入,实现鲁棒性。Cheng等人[5]提出引入颗粒球表示,并提出了一种名为GB-DP(基于颗粒球的密度峰值)的新型聚类算法,旨在解决大规模数据聚类的挑战。Xie等人,[37]提出了一种基于粒球的谱聚类算法,以减少大规模数据中的时间和资源成本。与传统方法相比,该算法效率更高,鲁棒性更强。Xia等[35]提出了颗粒球邻域粗糙集(GBNRS)方法,解决了邻域粗糙集因邻域半径固定而导致的效率和有效性问题。通过自适应生成不同邻域,GBNRS显著提高了通用性和灵活性。

## 3 我们提出的解决方案:gbrain

图1展示了我们框架的概述。GBRAINtakes作为带标签噪声的训练数据集的输入。数据集中的每个样本都是一个带标签的句子,不一定正确。然后输入的文本首先由一个编码器处理,这个编码器是一个表示网络。这些输入被映射到一个高维表示空间,称为特征向量。随后,GBC层根据特征向量聚类接近输入,构建多粒度语料库(即颗粒球)。所有样本的表示和标签将基于粒球重新表达,在保留数据特征的同时纠正标签噪声。被表示的实例被进一步送入一个分类器,该分类器依赖输出概率分布来预测它们的标签。然后可以根据预测和重新表示的标签计算一个损失值,这反映了模型在训练数据上的性能。通过迭代训练中的反向传播和减少损失,可以对模型进行优化,并弄清楚粗粒度表示和正确标签之间的关系。因此,可以缓解训练中标签噪声对模型性能的不利影响。模型在推理过程中可以更具有一般性和鲁棒性。如上所述,GBRAINincludes有三个模块:编码器、gbc层和分类器。我们声称GBRAINis是一个通用框架,因此编码器模块可以是任何可以将输入文本表示为特征向量的DNN模型,如LSTM、CNN和BERT。类似地,分类器可以是任何分类器网络。我们的框架的核心是GBC层,主要由输入重表示(3.1节)和表示优化(3.2节)组成。我们将在接下来的小节中详细介绍它们。

### 3.1 输入粒球重表示

该模块旨在纠正噪声标签的同时,保留基于聚类相似样本的数据特征。为了保持数据的原始特征,该模块将通过粗粒度表示学习重新表示输入。因此,这个模块的主要任务是对样本进行聚类和重新表示。

对于聚类任务,样本已经映射成高维特征向量,经过文本编码器的处理,如图1所示。传统的聚类方法要么局限于低维数据,要么因为预先定义了集合的大小而不够正确。基于粒球计算,我们针对高维数据设计了一种无需配置集合大小的自适应聚类方法。整个过程如算法1所示。具体来说,该算法取一批特征向量(即一些输入文本的嵌入),记为D,以及一个用户自定义的纯度阈值Tas输入。该算法的输出是覆盖输入数据集中所有样本的所有gb值,以及这些gb值的标签和中心向量。GB是定义3.1中相似输入样本的集合。纯度是评价GB质量的一个度量,见定义3.2。该算法首先将队列Q初始化为空集。Q维护着一个中间集合列表,其纯度没有达到阈值T,等待进一步优化。随后,它将输入数据集D添加到空队列q中。接下来,算法迭代处理队列中的每个集合Q<sub>i</sub>in,以创建从第3行到第11行令人满意的颗粒球。对于每个集合,算法首先通过函数计算其纯度值

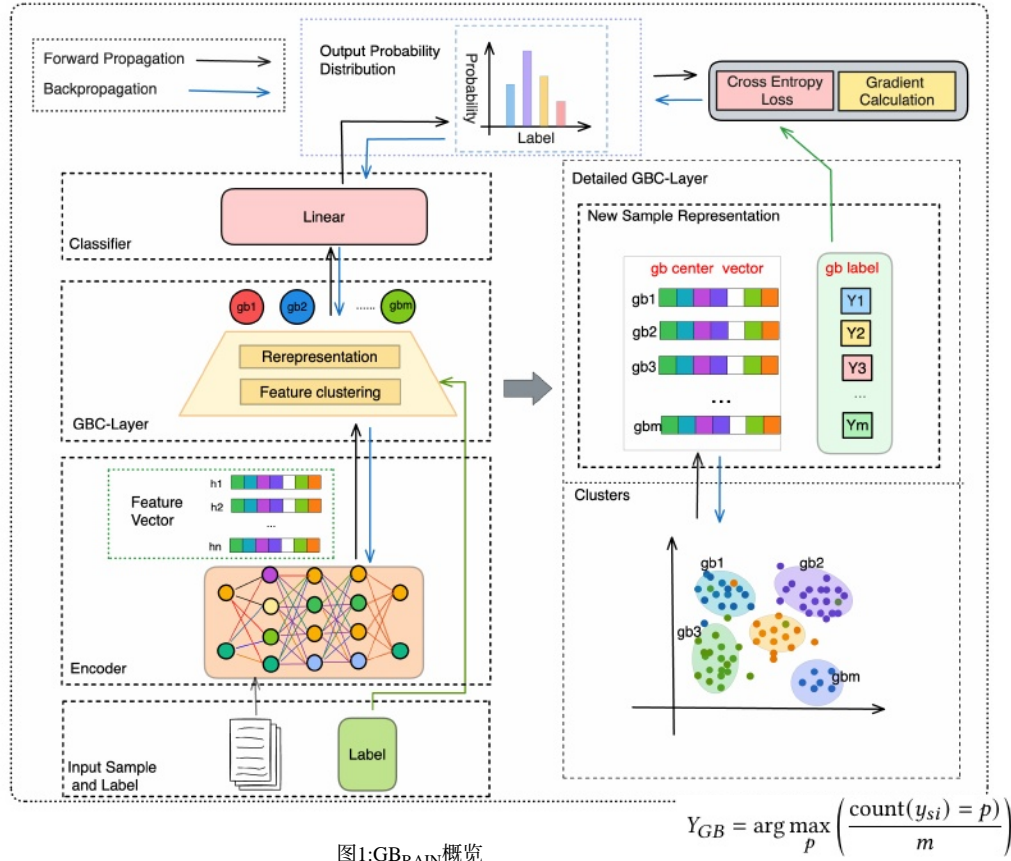


图1:GBRAIN概览。

GETPURITY() (第6行), 通过计算定义3.2中呈现的 $Q_1$ as中样本最多的一组样本数量与全部样本数量的比值。如果纯度值小于阈值, 函数SPLITSET()会将 $Q_1$ 进一步分割为 $m$ 个子集(第8行), 其纯度值都满足阈值 $T$ 。否则, 当前的集合 $Q_1$ satisfies颗粒球的要求, 并将记录在输出GBS中。最后, 当算法收敛时, 几乎所有的输入样本都在某个粒球中, 所有粒球的并集近似等于整个输入。这个函数的细节如算法1中的第17~31行所示。SPLITSET()函数迭代地分割每个不满意的簇 $Q_i$ in其等待队列 $S$ , 直到所有簇的纯度满足阈值 $T$ 。具体来说, 对于每个簇, 它随机选择 $K$ 个样本作为新的簇中心。 $K$ 等于标签类型的数量。 $Q_i$ will中的一个样本被分配到一个新的聚类 $Q_{C_i}$ , 如果 $Q_{C_i}$ is的中心向量比其他中心离样本最近(第23-26行)。在 $Q_i$ are中的所有样本重组为新的簇后, 该函数将根据其包含的样本更新每个簇的中心向量。整个聚类过程是自适应的, 无需预先定义最终聚类的数量。

定义3.1(粒状球)给定一批样品 $D = \{(X_1 Y_1)(X_2 Y_2) \dots (X_N Y_N)\}$ ,  $X_i$ is一个样本的特征向量,  $Y_i$ is样本的标签,

$$\text{Granular Ball} = \{(x_{s1}, y_{s1}), \dots, (x_{sm}, y_{sm})\} \quad (1)$$

$$C_{GB} = \frac{1}{N} \sum_{i=1}^N x_{si} \quad (3)$$

在哪里  $(X_{Si}, Y_{Si}) \in D$ , 即  $\text{GRANULARBALL} \subseteq D$ ,  $Mis$  大小的 $\text{GRANULARBALL}$ , 和所有特征向量 $x_{Si}$ are pre-defined 内近距离。 $C_{GB}$  and  $Y_{GB}$  are 当前颗粒球的中心点向量和标签。

定义3.2. 给定一个GB, 它由 $k$ 类样本组成, 即  $\bigcup_{i=1}^K C_i = GB$  和  $\tilde{N}_{i=1}^K C_i =$ , 其中 $C_i$  represents 为GB中的第 $i$ 类样本。纯度 $T_{GB}$  of GB 可表示为

$$T_{GB} = \frac{\max |c_i|}{|GB|}, i = 1, 2, \dots, k \quad (4)$$

$| \cdot |$  表示集合中的样本数量。

如上所述, 与现有工作相比, 我们的聚类方法更鲁棒, 与高维特征空间兼容。与 $k$ -means等传统聚类方法需要预先定义集合数量不同, GB聚类算法根据纯度和标签类型的数量, 自适应地决定是否拆分一个集合以及拆分为多少个子集, 使算法能够在不限制集合数量的情况下自动划分和聚类。因此, 我们的方法是灵活的

**Algorithm 1:** The clustering algorithm of  $GB_{RAIN}$ 


---

**Data:**  $\mathcal{D}$ : a batch of samples,  $T$ : the threshold of purity  
**Result:**  $GBs$ : all granular balls that cover the whole input data, together with their center vector and label

```

1 Initializing:  $Q = \emptyset, GBs_{init} = \emptyset$ 
2  $Q = Q \cup \mathcal{D}$ 
3 while  $Q$  is not empty do
4   for  $Q_i$  in  $Q$  do
5      $Q \leftarrow Q - Q_i$ 
6      $T_{Q_i} \leftarrow getPurity(Q_i)$ 
7     if  $T_{Q_i} < T$  then
8        $[Q_{i-child1}, \dots, Q_{i-childm}] \leftarrow splitSet(Q_i)$ 
9        $Q = Q \cup [Q_{i-child1}, \dots, Q_{i-childm}]$ 
10    else
11       $GBs_{init} = GBs_{init} \cup Q_i$ 
12 for  $gb$  in  $GBs_{init}$  do
13    $C_{gb} = getCenter(gb)$  (using Eq 3)
14    $Lable_{gb} = getLabel(gb)$  (using Eq 2)
15    $GBs = GBs \cup gb(C_{gb}, Lable_{gb})$ 
16 return  $GBs$ 
17 Function  $splitSet(Q_i)$ :
18  $S = Q_i$ 
19 repeat
20   for each cluster  $Q \in S$  and  $T_Q < T$  do
21      $S = S - Q$ 
22     Randomly select  $K$  data points from  $Q_i$  as new
       cluster centers:  $C_1, C_2, \dots, C_K$ 
23     for each data point  $x_i \in Q_i$  do
24       Calculate the distance to each  $C_j$  using Eq 6.
25        $d_{ij} = Dis(x_i, C_j)$  for  $j = 1, 2, \dots, K$ 
26       Assign  $x_i$  to the nearest cluster:
          $c_j = \arg \min_j d_{ij}, Q_{c_j} \leftarrow x_i$ 
27     for each new cluster  $Q_{C_j}$  do
28       update its center:
          $C_j = \frac{1}{|\{x_m | C_{x_m} = C_j\}|} \sum_{x_m \in C_j} x_m$ 
29      $S = S \cup Q_{C_j}$ 
30 until Convergence;
31 return  $S$ 

```

---

足以创建具有上下文相似的多粒度样本集，这有助于生成的聚类更好地拟合数据特征分布，从而产生更鲁棒的聚类。这一特性赋予了GB聚类更广泛的适用性，特别是在集群数量未知的场景中。此外，传统的聚类方法并不一定适用于高维空间。这是因为传统的距离计算可能面临“维数灾难”的挑战，即在高维数据中，距离计算变得复杂。我们的方法通过引入加权距离的概念来解决这个问题。它在距离计算时为不同维度分配不同的权重，如定义3.3所示，提高差异较大维度的影响。这

种策略更好地兼容高维空间，增强了在此类环境中的性能和效率。总的来说，在这样的多粒度聚类方法的支持下，我们的输入重表示将更加鲁棒和精确。

定义3.3. 给定两个高维向量 $x(X_1, X_2, \dots, X_N)$ 和 $y(Y_1, Y_2, \dots, Y_N)$ ，两个向量之间的距离计算为

$$d = x - y = ((x_1 - y_1), (x_2 - y_2), \dots, (x_n - y_n)) \quad (5)$$

$$Dis(\mathbf{x}, \mathbf{y}) = \|(\alpha_1, \alpha_2, \dots, \alpha_n) \mathbf{d}^T\|_2 \quad (6)$$

其中 $n$ 是向量的维数， $d$ 是维数为 $N$ 的向量保持两者之间的差 $x$ 和 $y$ ， $Dis_0$ 表示加权距离， $\alpha_l$  represent 具有不同权重的超参数，如果 $D_l > D_j$ ，则 $\alpha_l > \alpha_j$ 。

至于重新表示任务，所有样本在聚类过程后都被划分为颗粒球。根据GB聚类原理，一个球中的所有样本是接近的，语义相似的，所以它们应该有相同的标签。我们使用球内样本最多的标签作为所有样本和球的标签，如式(2)所示。因此，标签噪声可以通过大多数准确标注的数据来纠正。此外，我们使用球的中心向量作为球内每个样本的新特征向量。中心向量是按式(3)计算的GB内所有样本向量的平均值。这种嵌入重表示的原因主要有两点。首先，如果GB包含一些有噪声的数据点，取球内样本的平均值有助于减轻噪声的负面影响。其次，聚类过程不够准确，即一个粒球内的所有样本不一定具有相同的标签。如果样本保持原来的嵌入，但有了新的重新表示的标签，就会创建异常的训练样本，这些样本的嵌入与它们的标签不对应。然而，中心向量在一个颗粒球中包含了样本的所有特征，并且可以通过对所有样本进行平均来降低误差。因此，我们使用中心向量作为样本的新嵌入。总而言之，这个模块在训练过程中将数据组织成多个粒度级别。它将具有相似特征的样本聚合成粒球，最终降低了标签噪声对模型训练的影响。

### 3.2 表示优化

通过第3.1节中的重新表示过程，我们将所有样本表示为新的嵌入和标签。然而，这些表示需要在多轮训练中进一步优化。为了实现这一目标，我们引入了颗粒球计算(GBC)层，并将其集成到模型训练过程的前向和后向传播中。GBC层的前向传播对应于输入重表示模块的工作流(第3.1节)。该模块为样本获取粗粒度的特征向量和相应的标签，将它们输入分类器以训练模型。设 $X$ 表示一批输入向量，其中每个向量表示为 $x_i$ 。通过GBC层后，生成包含这些输入的多个颗粒球。GBRAIN<sup>network</sup>最终将一个样本表示为其从属颗粒球的中心向量。

在反向传播阶段，我们设计了一种方法来确保梯度信息从



分类器到每个颗粒球, 基于卷积神经网络(CNN)中平均池化的反向传播。具体来说, 该模型需要计算损失函数的偏导数 $\frac{\partial \mathcal{L}}{\partial \mathbf{x}_i}$  concerning 每个输入向量 $\mathbf{x}_i$ , 即梯度, 帮助优化模型参数。但在我们的场景中, 每个样本的representation已经重新的中心向量 $\mathbf{c}_j$  所属granular-ballGB $\mathbf{J}_j$ , 反向传播必须首先通过球, 然后达到单样本。为了实现这个传播, 我们直接计算损失函数的梯度 $\frac{\partial \mathcal{L}}{\partial \mathbf{c}_j}$  concerning granular-balls的中心向量, 并把他们作为包含样品的梯度定义3.4所示。换句话说, 对于样本属于granular-ballGB $\mathbf{J}_j$ , 他们的梯度是由失去关于中心的梯度向量 $\frac{\partial \mathcal{L}}{\partial \mathbf{c}_j}$  of GB $\mathbf{J}_j$ 。这个过程应用于整个训练过程中的所有粒球和样本。这确保了反向传播中的梯度从粒球传递到每个单个样本, 类似于平均池化反向传播中将残差复制到每个样本。与之最大的区别是, 我们没有使用粒球梯度的平均值(即 $(\frac{\partial \mathcal{L}}{\partial \mathbf{c}_j})/|\mathbf{GB}_j|$ )作为每个样本的梯度。相反, 我们

使用加权策略。具体来说, 我们的梯度计算是根据颗粒球中当前的样本数量(gb)分配不同的权重( $\gamma$ )。我们设置了一个阈值, 当前设置为3; 如果一个gb的样本数量小于3, 那么 $\gamma$ 设置为0, 否则,  $\gamma$ 设置为1。

定义3.4。给定一个输入向量 $\mathbf{x}_i$ 和一个带有中心向量 $\mathbf{c}_j$ 的颗粒球GB $\mathbf{J}_j$ , 计算得到的损失函数 $\mathcal{L}$  concerning  $\mathbf{x}_i$ 的梯度为

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}_i} = \gamma \frac{\partial \mathcal{L}}{\partial \mathbf{c}_j}, \text{ for } \mathbf{x}_i \in gb_j, \gamma = (0||1)$$

## 4 实验

我们进行了充分的实验, 以验证我们所提出的方法在标签噪声上的有效性。此外, 为了衡量通用性, 评估了所提出的框架GBRAINon常见的NLP分类任务(即文本分类和命名实体识别)。我们将在接下来的小节中详细介绍对这两个任务的评估细节。

### 4.1 关于文本分类的评价

本节将评估GBRAINon文本分类任务的有效性。为了使评估充分和令人信服, 我们在两个数据集上评估了六个DNN模型。详细情况如下。

4.1.1 实验设置。数据集。在两个流行的数据集上进行了实验。AGNews数据集[43]由12万个训练样本和7600个测试样本组成, 来自2000多个新闻媒体。每个样本都是短文本, 标有四种类型中的一种, 表明其类别或主题。IMDB数据集[23].情感分析的基准, 包括50000年的电影评论从互联网电影数据库(IMDB)[23].分为25000份训练和测试样本。每个评论贴上一个二进制的情绪标签, 表明积极或消极的情绪。模型。GBRAIN是一个通用框架, 可以与任何DNN模型集成, 以提高当前模

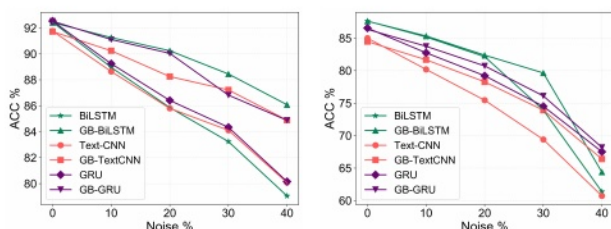
型的鲁棒性。因此, 我们应用GBRAINto三种不同的文本分类模型, BiLSTM[40]、GRU[8]和Text-CNN[13]来评估其对标签噪声的有效性。我们使用GBRAINframeworkGB-BiLSTM、GB-GRU和GB-Text-CNN来称呼这些模型。因此, 我们在这次评估中得到了6个模型。对于BiLSTM和GRU, 我们使用了一个包含128个隐藏单元的双层网络。在Text-CNN的情况下, 我们使用三个卷积层, 内核大小分别为3,4,5, 单词向量的维数随机初始化为100。我们设定最大序列长度为400令牌保持序列信息, 因为最长的样品只包含500令牌。环境配置。训练时学习率为 $1e-3$ 。实验在两块NVIDIA GeForce RTX 3090 gpu上运行。

4.1.2 实现细节: 为了验证GBRAIN在噪声环境中的有效性, 首先通过有意修改两种训练集中的标签来引入随机噪声。我们实现了一个Python脚本, 自动修改一定比例的数据集的标签。百分比范围从0%到40%, 这也反映了相关修改数据集中的噪声水平。在这样的情况下, 我们可以全面评估我们的建议在不同水平的噪声环境下的性能, 为其在处理标签错误和不确定性方面的鲁棒性提供见解。其次, 我们将我们的框架与原始评估模型集成。具体来说, 我们在训练过程中在原始模型(BiLSTM、GRU和Text-CNN)的编码器和分类器之间插入GBRAIN的GBC层, 得到GB-BiLSTM、GB-GRU和GB-Text-CNN模型。同时, 我们将原始模型作为基线, 进行详细的比较研究。在准备好数据集和模型之后, 我们开始在带有标签噪声的数据集上训练这6个模型。当模型的准确率保持稳定时, 这个过程就停止了。在推理过程中, 我们评估了6个模型在干净数据集和有不同程度噪声的数据集上的性能。我们使用测试集上的准确率来表示它们的性能, 其计算方法为正确预测样本的数量除以所有样本的数量。此外, 为了更深入地研究影响模型性能的因素, 我们在不同的噪声水平下监控整个实验过程中失败预测的概率。具体来说, 每次模型在100批训练样本上完成训练时, 都会在测试数据集上对其性能进行评估。这个过程不断重复, 直到测试性能变得稳定。这种评估只在Agnews数据集中具有20%和40%噪声水平的Bi-LSTM基模型上进行评估, 这在实践中更具有代表性。我们用错误率来代表不同模型的性能, 以100%和准确率之间的差值来计算。

4.1.3 结果与分析: 有效性的实验结果如表1所示。当训练数据集是干净的噪音水平0%, 第二个和第七列所示的表1, 我们的建议的性能低于cor-responding原始模型。例如, 在两个数据集上, GB-BiLSTM的准确率都低于BiLSTM; 同样, GB-TextCNN低于Text-CNN, GB-GRU低于GRU。由于在批处理中对向量进行聚类生成颗粒球(Granular-Balls, GBs), 一个GB内的多个样本被一个中心向量代替。这种替换往往会降低样本的多样性, 导致

表1:文本分类任务的实验结果。亮点代表了最佳性能(%)。

Datasets	Agnews					IMDB					
	Noise(%)	0	10	20	30	40	0	10	20	30	40
BiLSTM		92.45	88.96	85.86	83.24	79.05	<b>87.63</b>	85.18	82.15	73.94	61.36
Text-CNN		91.76	88.64	85.80	84.11	81.08	85.00	80.19	75.44	69.40	60.71
GRU		<b>92.53</b>	89.24	86.41	84.34	80.16	86.55	82.74	79.21	74.51	67.51
GB-BiLSTM(ours)		92.41	<b>91.26</b>	90.24	<b>88.45</b>	<b>86.08</b>	87.61	<b>85.33</b>	<b>82.38</b>	<b>79.64</b>	64.38
GB-TextCNN(ours)		91.72	90.24	88.25	87.22	84.89	84.47	81.67	78.27	73.91	66.40
GB-GRU(ours)		92.52	91.11	<b>90.04</b>	86.82	84.88	86.53	83.75	80.72	76.11	<b>68.18</b>

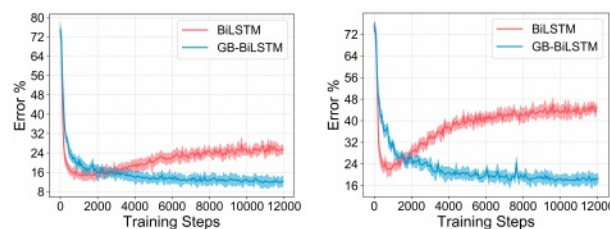


(a) Accuracy rate in Agnews dataset (b) Accurate rate in IMDB dataset

图2:不同噪声水平环境下模型准确率的实验结果

模型性能略有下降。当噪声应用于训练数据集时,我们基于gbc的方法总是优于原始模型。随着噪声水平的增加,所有模型的性能都在持续下降,但原始模型和我们的建议之间的性能差距总体上变得更大,如图2所示。图2展示了数据集不同噪声水平下模型精度变化的趋势,其中不同的颜色表示不同的基础模型,而相同的颜色表示相同的基础模型。曲线上的模式用于区分使用相同基础模型的不同方案。这表明我们的方法有了显著的改进GB<sub>RAIN</sub>to地址标签噪声。具体在高噪声环境下,GB<sub>RAIN</sub>is的表现更加突出。正是因为我们的模型通过在噪声环境下的粗粒度表示学习来增强鲁棒性和泛化性,从而使GB<sub>RAIN</sub>can更好地适应真实输入的噪声。

此外,图3给出了在训练过程中观察模型性能的实验结果。如图所示,在不同噪声水平下,原始模型BiLSTM(红线)在开始时的性能优于我们改进的模型GB-BiLSTM(蓝线),但在大约2000步之后,其性能开始越来越差于GB-BiLSTM。当训练过程收敛时,我们的模型的表现比原始模型要好得多。对比图3(a)和图3(b),在40%噪声环境下,GB-BiLSTM与BiLSTM的性能差异更为显著。这种现象表明BiLSTM过拟合训练数据集。由于噪声训练样本的影响,模型的决策边界偏离了正常的数据特征分布。当这样的模型在正常测试样本上运行时,其预测可能会出错。相反,GB-BiLSTM减小了噪声样本的影响



(a) 20% noise in the training dataset (b) 40% noise in the training dataset

图3:在各种噪声水平环境下对模型性能随训练步骤变化的实验观察

通过基于粒球的粗粒度表示学习。即异常样本的嵌入和标注可以通过基于粒球的大多数正确训练本来优化。在这种情况下,GB-BiLSTM可以避免噪声样本的过拟合,对测试样本具有良好的鲁棒性和性能。总而言之,评估证明了我们的框架在文本分类上的有效性。

## 4.2 命名实体识别评价

命名实体识别(Named Entity Recognition, NER)是自然语言处理中的一项序列标注任务,其重点是从文本序列中识别实体。为了有效地执行这项任务,通常需要大规模的标注语料库。然而,由于获取和标注这类语料库的成本较高,通常使用远程监督的命名实体识别,基于外部知识库和词典,为未标注的数据集生成远程标签。尽管这种方法的优点在减少注释成本,知识库的局限性总是介绍标签噪音。因此,我们应用我们的框架G<sub>RAIN</sub>to远程监督NER任务来评估有效性。在RoBERTa模型上进行评价。此外,我们将我们的提议与BOND[20]、SCDL[42]和名册[25]这三种技术状态进行了比较。详情如下。

4.2.1实验设置。数据集。评估在三个流行的公共数据集上进行。CONLL2003数据集[28]是一个著名的开放域NER数据集,来自CONLL共享任务[7],包含来自1,393篇英语新闻文章的注释数据,具有四种主要实体类型,包括人员、地点、组织和杂项实体。Twitter数据集[10]是来自WNUT 2016 NER共享任务[38]的开放域NER数据集,包含来自2,400条tweet的注释数据,其中34,000条tweet

表2:命名实体识别任务上的实验结果。

Method	CoNLL2003			WikiGold			Twitter		
	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall
Distant Match	0.714	81.13	63.75	0.478	47.9	47.63	0.358	40.34	32.22
RoBERTa	0.759	82.29	70.47	0.525	47.67	58.59	0.464	50.97	42.66
BOND	0.814	82.05	80.92	0.600	53.44	68.58	0.480	53.16	43.76
SCDL	0.836	87.96	79.82	0.640	62.25	66.12	0.510	<b>59.87</b>	44.57
RoSTER	0.854	<b>85.90</b>	84.90	0.678	64.90	71.00	-	-	-
GB-RoBERTa(ours)	<b>0.866</b>	85.50	<b>87.70</b>	<b>0.724</b>	<b>69.60</b>	<b>75.40</b>	<b>0.538</b>	53.50	<b>54.20</b>

令牌和与Twitter中常见命名实体相关的十种实体类型。wikgold数据集[2]由一组文章组成， 这些文章是从2008年英文维基百科转储[32]中随机选择的。该数据集包含40,000个令牌， 并使用维基百科文章中的四种类型的实体(人员、位置、组织和杂项)进行手动注释。模型。我们选择RoBERTa [21]作为基础模型， 一个在Hugging Face[16]中通过远程监督标记的数据上进行微调的预训练模型。我们应用GB<sub>RAIN</sub> toRoBERTa得到我们的模型GB-RoBERTa。 我们还将GB-RoBERTa与遥远匹配[20]、BOND[20]、SCDL[42]和名册[25]这四个基线进行了比较。遥远匹配是实现具有远程监督的NER的基础模型。BOND首先在通过远程监督和早期停止标记的数据上训练RoBERTa模型。随后， 它采用教师-学生框架进行迭代训练， 以增强模型的性能。SCDL以一种互惠的方式训练两个师生网络来迭代地改进噪声标签。名册首先使用广义交叉熵损失函数和标签噪声去除， 然后利用上下文增强的语言模型来训练网络模型的泛化。环境配置。CoNLL2003、Twitter和wikgold的最大序列长度分别设置为150、100和120个令牌。训练过程中， 三个数据集的批处理大小为128， 预训练模型和分类器的学习率分别为1e-6和1e-4。AdamW

选择[22]作为优化器。实验在配备了NVIDIA GeForce RTX 3090 gpu的服务器上进行。

4.2.2实现细节:为了评估我们的建议， 我们首先集成GB<sub>RAIN</sub>with基本NER模型(称为RoBERTa)[21]， 通过插入GB<sub>RAIN</sub>between预训练的屏蔽语言模型和NER分类器的GBC层， 获得我们的增强模型GB-RoBERTa。我们没有评估遥远匹配模型， 因为它的结果是直接从Liang等人[20]的工作中提取出来的， 这些结果是基于对遥远标记结果和地面真值的比较而计算出来的。在训练过程中， 为了模拟远程监督NER， 我们在三个不带注释标签的数据集上训练了五个模型， 即RoBERTa、BOND、SCDL、RoSTER和GB-RoBERTa。相反， 我们使用[20]工具自动为训练样本生成标签， 它是基于将一个远程知识库与样本内容匹配来获取远程标签的。在自动标注过程中， 标签噪声会被引入到训练数据集中。我们使用测试数据集上模型的F1分数、精度和召回率来显示模型性能。

4.2.3结果与分析:实验结果如表2所示。比较遥远匹配和基本模型RoBERTa， RoBERTa在这三个指标上都表现得更好。这证实了远程监督NER引入了许多噪声， 这证明了我们的噪声训练环境。然而， 与四种改进的模型相比， RoBERTa在所有数据集中的表现都更差。这说明标签噪声对DNN模型性能影响严重， 可能导致模型过拟合标签噪声， 泛化能力下降。缓解标签噪声的研究迫在眉睫。当将我们的GB-RoBERTa解决方案与现有工作进行比较时， GB-RoBERTa在所有数据集上都具有最佳的F1分数和召回率。然而， 考虑到Twitter数据集CONLL2003的精度， 它的表现更差。特别是， GB-RoBERTa的召回率明显增加。总的来说， 这些现象表明我们的建议在噪声文本中捕获实体有很大的有效性， 可以正确识别实体， 同时减少假阴性。我们的框架也总是优于现有的工作。GB-RoBERTa的优势可能来源于我们的鲁棒训练框架中的多粒度表示方法。总而言之， 这些评估说明了我们的框架在远程监督NER中的有效性。

5 结论

在本文中， 我们提出了一种新的基于颗粒球的t<sub>RAIN</sub>ing框架， 命名为GBRAIN， 以减轻标签噪声对神经网络模型性能的影响。GB<sub>RAIN</sub>is的主要优势是它能够粗粒度地学习所有样本的嵌入向量， 然后自适应地纠正它们的标签。在此基础上， GB<sub>RAIN</sub>also引入了一种新的反向传播机制——通过迭代训练进一步细化粗粒度的嵌入向量。在实体识别和文本分类任务上的实验结果表明， 与最先进的基线相比， 我们提出的解决方案显著降低了标签噪声对传统神经网络模型的影响。

6 确认

国家自然科学基金资助项目(62222601、62176033、62221005、61936001)， 重庆市教委重点合作项目;重庆市自然科学基金项目(cstc2022ycjh-bgzxm0128、cstc2021ycjh-bgzxm0013、CSTB2023NSCQ-JQX0034、CSTB2022NSCQ-MSX1588)。



## 参考文献

- [1] Samaneh Azadi, Jiashi Feng, Stefanie Jegelka, and Trevor Darrell. 2015. Aux-iliary image regularization for deep cnns with noisy labels. *arXiv preprint arXiv:1511.07069* (2015).
- [2] Dominic Balasuriya, Nicky Ringland, Joel Nothman, Tara Murphy, and James R Curran. 2009. Named entity recognition in wikipedia. In *Proceedings of the 2009 workshop on the people's web meets NLP: Collaboratively constructed semantic resources (People's Web)*. 10–18.
- [3] Lin Chen. 1982. Topological structure in visual perception. *Science* 218, 4573 (1982), 699–700.
- [4] Robert S Chen, Brendan Lucier, Yaron Singer, and Vasilis Syrgkanis. 2017. Ro-bust optimization for non-convex objectives. *Advances in Neural Information Processing Systems* 30 (2017).
- [5] Dongdong Cheng, Ya Li, Shuyin Xia, Guoyin Wang, Jinlong Huang, and Sulan Zhang. 2023. A fast granular-ball-based density peaks clustering algorithm for large-scale data. *IEEE Transactions on Neural Networks and Learning Systems* (2023).
- [6] Shaohuan Cheng, Wenyu Chen, Fu Mingsheng, Xuanting Xie, and Hong Qu. 2023. Adaptive Textual Label Noise Learning based on Pre-trained Models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*. 3174–3188.
- [7] CoNLL. 2003. Conference on Natural Language Learning. <https://www.conll.org/>.
- [8] Rahul Dey and Fathi M Salem. 2017. Gate-variants of gated recurrent unit (GRU) neural networks. In *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*. IEEE, 1597–1600.
- [9] Siddhant Garg, Goutham Ramakrishnan, and Varun Thumbe. 2021. Towards robustness to label noise in text classification via noise modeling. In *Proceedings of The 30th ACM International Conference On Information & Knowledge Management*. 3024–3028.
- [10] Frédéric Godin, Baptist Vandersmissen, Wesley De Neve, and Rik Van de Walle. 2015. Multimedia lab@ acl wnut ner shared task: Named entity recognition for twitter microposts using distributed word representations. In *Proceedings of the workshop on noisy user-generated text*. 146–153.
- [11] Sheng Guo, Weilin Huang, Haozhi Zhang, Chenfan Zhuang, Dengke Dong, Matthew R Scott, and Dinglong Huang. 2018. Curriculumnet: Weakly supervised learning from large-scale web images. In *Proceedings of the European conference on computer vision (ECCV)*. 135–150.
- [12] Bo Han, Quanming Yao, Xingrui Yu, Gang Niu, Miao Xu, Weihua Hu, Ivor Tsang, and Masashi Sugiyama. 2018. Co-teaching: Robust training of deep neural net-works with extremely noisy labels. *Advances in neural information processing systems* 31 (2018).
- [13] Tong He, Weilin Huang, Yu Qiao, and Jian Yao. 2016. Text-attentional convo-lutional neural network for scene text detection. *IEEE transactions on image processing* 25, 6 (2016), 2529–2541.
- [14] Dan Hendrycks and Kevin Gimpel. 2016. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136* (2016).
- [15] Jinchi Huang, Lie Qu, Rongfei Jia, and Binqiang Zhao. 2019. O2U-Net: a simple noisy label detection approach for deep neural networks. in *2019 IEEE International Conference on Computer Vision (ICCV)*, Vol. 80. 3325–3333.
- [16] Hugging Face. 2024. Transformers: State-of-the-art Natural Language Processing. <https://huggingface.co/> (2024).
- [17] Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li-Jia Li, and Li Fei-Fei. 2018. Mentor-net: Learning data-driven curriculum for very deep neural networks on corrupted labels. In *International conference on machine learning*. PMLR, 2304–2313.
- [18] Ishan Jindal, Matthew Nokleby, and Xuewen Chen. 2016. Learning deep networks from noisy labels with dropout regularization. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*. IEEE, 967–972.
- [19] Ishan Jindal, Daniel Pressel, Brian Lester, and Matthew Nokleby. 2019. An effective label noise model for dnn text classification. *arXiv preprint arXiv:1903.07507* (2019).
- [20] Chen Liang, Yue Yu, Haoming Jiang, Siawpeng Er, Ruijia Wang, Tuo Zhao, and Chao Zhang. 2020. Bond: Bert-assisted open-domain named entity recognition with distant supervision. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*. 1054–1064.
- [21] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).
- [22] Ilya Loshchilov and Frank Hutter. 2018. Fixing weight decay regularization in adam. (2018).
- [23] Andrew Maas, Raymond E Daly, Peter T Pham, Dan Huang, Andrew Y Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*. 142–150.
- [24] Naresh Manwani and PS Sastry. 2013. Noise tolerance under risk minimization. *IEEE transactions on cybernetics* 43, 3 (2013), 1146–1151.
- [25] Yu Meng, Yunyi Zhang, Jiaxin Huang, Xuan Wang, Yu Zhang, Heng Ji, and Jiawei Han. 2021. Distantly-supervised named entity recognition with noise-robust learning and language model augmented self-training. *arXiv preprint arXiv:2109.05003* (2021).
- [26] Giorgio Patrini, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. 2017. Making deep neural networks robust to label noise: A loss correction approach. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1944–1952.
- [27] Dan Qiao, Chenchen Dai, Yuyang Ding, Juntao Li, Qiang Chen, Wenliang Chen, and Min Zhang. 2022. SelfMix: Robust Learning Against Textual Label Noise with Self-Mixup Training. *arXiv preprint arXiv:2210.04525* (2022).
- [28] Erik F Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. *arXiv preprint cs/0306050* (2003).
- [29] Sainbayar Sukhbaatar, Joan Bruna, Manohar Paluri, Lubomir Bourdev, and Rob Fergus. 2014. Training convolutional networks with noisy labels. *arXiv preprint arXiv:1406.2080* (2014).
- [30] Guoyin Wang. 2017. DGCC: data-driven granular cognitive computing. *Granular Computing* 2, 4 (2017), 343–355.
- [31] Yisen Wang, Xingjun Ma, Zaiyi Chen, Yuan Luo, Jinfeng Yi, and James Bailey. 2019. Symmetric cross entropy for robust learning with noisy labels. In *Proceedings of the IEEE/CVF international conference on computer vision*. 322–330.
- [32] Wikipedia. 2024. Title of the Article. <https://en.wikipedia.org/> Accessed January 31, 2024.
- [33] Shuyin Xia, Yunsheng Liu, Xin Ding, Guoyin Wang, Hong Yu, and Yuoguo Luo. 2019. Granular ball computing classifiers for efficient, scalable and robust learning. *Information Sciences* 483 (2019), 136–152.
- [34] Shuyin Xia, Daowan Peng, Deyu Meng, Changqing Zhang, Guoyin Wang, Elisabeth Gien, Wei Wei, and Zizhong Chen. 2020. A fast adaptive k-means with no bounds. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).
- [35] Shuyin Xia, Hao Zhang, Wenhua Li, Guoyin Wang, Elisabeth Gien, and Zizhong Chen. 2020. GBNRS: A novel rough set algorithm for fast adaptive attribute reduction in classification. *IEEE Transactions on Knowledge and Data Engineering* 34, 3 (2020), 1231–1242.
- [36] Shuyin Xia, Shaoyuan Zheng, Guoyin Wang, Xinbo Gao, and Binggui Wang. 2021. Granular ball sampling for noisy label classification or imbalanced classification. *IEEE Transactions on Neural Networks and Learning Systems* (2021).
- [37] Jiang Xie, Weiyu Kong, Shuyin Xia, Guoyin Wang, and Xinbo Gao. 2023. An Efficient Spectral Clustering Algorithm Based on Granular-Ball. *IEEE Transactions on Knowledge and Data Engineering* (2023).
- [38] Wei Xu, Bo Han, and Alan Ritter. 2015. Proceedings of the workshop on noisy user-generated text. In *Proceedings of the Workshop on Noisy User-generated Text*.
- [39] Quanming Yao, Hansi Yang, Bo Han, Gang Niu, and James Tin-Yau Kwok. 2020. Searching to exploit memorization effect in learning with noisy labels. In *International Conference on Machine Learning*. PMLR, 10789–10798.
- [40] Yong Yu, Xiaosheng Si, Changhua Hu, and Jianxun Zhang. 2019. A review of recurrent neural networks: LSTM cells and network architectures. *Neural computation* 31, 7 (2019), 1235–1270.
- [41] Qingjun Yuan, Gaopeng Gou, Yanbei Zhu, Yuefei Zhu, Gang Xiong, and Yongjuan Wang. 2023. MCRe: A Unified Framework for Handling Malicious Traffic with Noise Labels Based on Multidimensional Constraint Representation. *IEEE Transactions on Information Forensics and Security* (2023).
- [42] Xinghua Zhang, Bowen Yu, Tingwen Liu, Zhenyu Zhang, Jiawei Sheng, Mengge Xue, and Hongbo Xu. 2021. Improving distantly-supervised named entity recognition with self-collaborative denoising learning. *arXiv preprint arXiv:2110.04429* (2021).
- [43] Xiang Zhang, Junbo Zhao, and Yann LeCun. 2015. Character-level convolutional networks for text classification. *Advances in neural information processing systems* 28 (2015).
- [44] Bohan Zhuang, Lingqiao Liu, Yao Li, Chunhua Shen, and Ian Reid. 2017. Attend in groups: a weakly-supervised deep learning framework for learning from web data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1878–1887.