

基于不稳定性采样的主动学习方法*

何花, 谢明昆, 黄圣君

(南京航空航天大学 计算机科学与技术学院, 江苏 南京 211106)

摘要:传统的主动学习方法往往仅基于当前的目标模型来挑选样本,而忽略了历史模型所蕴含的对未标注样本预测稳定性的信息。因此,提出基于不稳定性采样的主动学习方法,依据历史模型的预测差异来估计每个未标注样本对提高模型性能的潜在效用。该方法基于历史模型对样本的预测后验概率之间的差异来衡量无标注样本的不稳定性,并挑选最不稳定的样本进行查询。在多个数据集上的大量实验结果验证了方法的有效性。

关键词:主动学习;标注代价;不稳定性;后验概率;熵

中图分类号:TP181 **文献标志码:**A **文章编号:**1001-2486(2022)03-050-07

Active learning method based on instability sampling

HE Hua, XIE Mingkun, HUANG Shengjun

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: Traditional active learning methods select examples by only considering the predictions of the current model. However, these methods neglect the information of the previous trained models, which reflect the stability of the prediction sequence for each unlabeled example during the active learning stage. Thus, a novel active learning method with instability sampling was proposed, which attempted to estimate the potential utility of each unlabeled examples for improving the model performance based on the difference among predictions of the previous models. The proposed method measured the instability of unlabeled example based on the difference between the posterior probabilities predicted by the previous models, and the example with the largest instability was selected to be queried. Extensive experiments were conducted on multiple datasets with diverse classification models. The experimental results validate the effectiveness of the proposed method.

Keywords: active learning; labeling cost; instability; posterior probability; entropy

深度学习已成功地应用于多种实际任务中,如,图像识别^[1-4]、自然语言处理^[5-9]和目标检测^[10-13]等。在现实应用场景中,训练一个有效的深度模型往往依赖于大量已标注样本,而准确标注大规模数据往往耗时耗力且代价高昂。为降低模型对数据的依赖,包括无监督学习^[14],半监督学习^[15-16]以及弱监督学习^[17-18]等领域的学习方法相继提出,并已吸引了大量关注。在这些方法中,主动学习是降低样本标注代价的主要途径之一。

主动学习通过迭代的方式,选择最有价值的样本进行标注并加入训练,旨在以最小的标注代价有效地提高模型性能。在该过程中,如何挑选样本是影响主动学习效果的关键因素。常用的采样策略往往基于样本的不确定性,包括最低置信度采样、边缘采样和熵采样^[19]。不确定性采样策

略的核心思想是用模型对样本的预测后验概率来估计该样本的不确定性。一般而言,模型在未标记样本上的预测概率越均衡,则越难以判断该样本所属的类别,因而,将这个样本加入训练中将有 效地提高模型 的分类性能。以二分类任务为例,熵采样策略通常选择后验概率最接近 0.5 的样本。

现有的主动采样策略都是根据当前模型对样本的预测来挑选样本。然而,这些方法忽略了在主动学习迭代过程中,分类模型对挑选样本的潜在价值。一般而言,在迭代过程中,不同轮次的分类模型对同一样本的识别效果是具有差异的,而这种差异性一定程度上反映了分类模型对该样本的识别能力的不稳定性。图 1 展示了在主动学习过程中,随着迭代轮次的增加,分类模型对 MNIST 数据集中的同一个无标注样本的预测情况。从图

* 收稿日期:2021-06-04

基金项目:新一代人工智能重大资助项目(2020AAA0107000);江苏省自然科学基金资助项目(BK20211517)

作者简介:何花(1995—),女,重庆人,硕士研究生,E-mail:hehua@nuaa.edu.cn;

黄圣君(通信作者),男,教授,博士,博士生导师,E-mail:huangsj@nuaa.edu.cn

中可以看出,在整个迭代过程中,分类模型的预测持续变化,这种变化反映了模型预测该样本的不稳定性,说明该样本是未标注数据集中较难以准确预测的样本,将其加入训练能有效地提高模型的性能。相反,忽略历史模型的这种潜在价值将导致主动学习策略挑选的并不一定是最有价值的样本。因此,在主动采样的过程中,除了考虑当前模型对未标注样本的预测,还应考虑以往模型预测的差异。

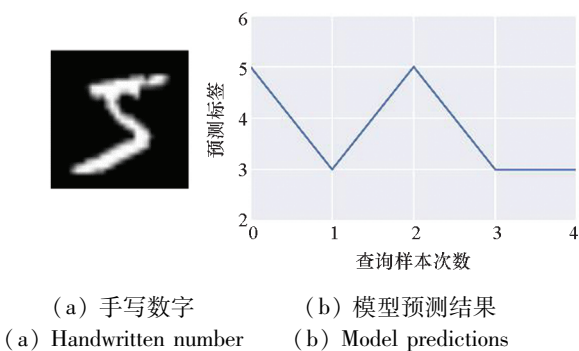


图1 历史模型对同一无标注样本的不稳定性预测

Fig. 1 Instability prediction of the same unlabeled instance by history models

本文提出基于不稳定性采样的主动学习方法,根据模型在整个学习过程中对无标注样本的预测差异来衡量未标注样本对提升模型性能的潜在效用。具体而言,在每一轮的主动查询中,对于每一个无标注样本,计算最近的 N 个模型对其预测后验概率的差异用以衡量其不稳定性,并选择最不稳定的样本进行标注。为了验证提出方法的有效性,分别使用传统和深度模型在 9 个数据集上进行了实验。实验结果表明,基于不稳定性主动采样方法能有效地提高模型的泛化性能。

1 相关工作

主动学习有选择地从标注者处查询最有价值的信息,旨在以最少的查询训练一个有效的模型。主动学习的关键任务是设计一个合适的策略,使所查询的信息对改进目标模型最有帮助。在传统设置下已提出了许多主动学习方法^[19]。其中一些方法挑选信息量最大的样本进行查询^[20-30],

信息量越大的样本对提升模型性能越有价值。而信息量可以用不同的标准来衡量,如不确定性^[21-27]、泛化误差减少量^[28-30]等,该类方法只考虑了模型对样本的需求,有可能导致挑选的样本分布与数据集真实分布存在差异。另外一些方法则查询具有代表性样本的标签^[31-33],其代表性可以根据聚类结构^[27,31]或者密度^[32]来估计,该

类方法挑选的是最能够代表样本分布的样本,忽略了模型本身对样本分类性能的信息。

目前主流的信息量与代表性相结合的方式可分为三类。第一类为串行结合方式,依次使用每个挑选策略来过滤“低价值”样本。常用做法为先从无标注样本集中挑选最有信息量的一批样本,然后使用聚类算法对这一批样本进行聚类,得到的聚类中心即为待查询样本^[34]。第二类为概率选择方式,在每轮主动学习迭代中,依据概率参数决定当前轮迭代使用的采样策略^[35]。第三类为并行结合方式,是目前最流行的主动学习策略结合方式,使用不同采样策略标准的加权求和或者多目标优化方法,计算混合得分,根据分数对未标注样本进行排序,挑选得分最高的一批样本^[36-41]。例如,Huang等从基于边际采样的主动学习策略推导出一个最小最大框架的目标^[41],放松未标注样本标签变量后,得到仅含最大化的目标函数,利用该目标得分挑选既具有信息量,又具有代表性的样本进行标注^[38]。Wang等将经验风险最小化推广到主动学习^[39],对导出的目标函数进行交替优化,其中该目标函数包括一个估计不确定性的项用以挑选具有较高信息量的样本,和一个估计有标记数据集与整个数据集之间分布差异的项使挑选出的样本具有代表性。Tang等在目标函数中除了考虑信息量和代表性,还结合了自步学习^[41],利用来自网络的不确定性和后验概率分布差异信息,挑选最简单的、最具信息量和代表性的样本进行标注。

此外,随着深度学习的发展,衍生出了许多其他采样标准,部分研究将主动学习策略视为从无标记样本到一个排序得分的映射,并利用深度模型去拟合这种映射。如,Liu等结合模仿学习(DAGGER)的框架直接学习一个采样策略,其中模仿学习所使用的专家策略为每轮主动采样策略的贪心选择^[43]。Yoo等利用预测损失模块拟合样本到目标模型损失值的映射,认为损失值越大的样本,对目标模型的提升性能越大^[44]。在此基础上,Li等将目标函数损失转换为排序,通过最小化排序损失来学习一个主动学习策略^[45]。该类方法中主动学习策略的学习依赖于大量训练数据,在少样本数据集中易过拟合。

虽然上述方法都试图估计一个样本对于改进模型的潜在价值,但都只用当前模型来对无标注样本进行评估,忽略了历史分类模型是否蕴藏挑选最具潜在价值的样本的能力。在不同迭代周期中,目标模型对同一样本的识别效果是变化的,如

果能够量化这种变化信息,就能挑选出模型识别效果最不稳定的样本,选择这类样本进行标注将对提升目标模型泛化性能提供更多的有效信息。

2 不稳定性指标定义

本章节首先在 2.1 节中介绍问题设定以及基于不稳定性的主动学习框架,然后在 2.2 节中提出不稳定性采样策略。

2.1 问题设定和学习框架

假设 $L = \{(x_i, y_i)\}_{i=1}^{n_l}$ 表示含有 n_l 个样本的已标注数据集, $U = \{x_j\}_{j=n_l+1}^{n_l+n_u}$ 表示含有 n_u 个样本的无标注数据集,其中 $n_l \ll n_u$,且 $n = n_l + n_u$ 。在主动学习第 t 轮迭代中,主动学习采样策略从 U 中挑选 b 个样本 $Q = \{x_q\}_{q=1}^b$ 进行标注,并加入已标注集。 M_t 表示第 t 轮迭代训练得到的模型。在不稳定性采样中,用 $P_{j,k}$ 表示模型 M_k 对样本 x_j 的预测向量,其对应的熵记为 $H_{j,k}$ 。表 1 汇总了本文使用的数学符号。

表 1 数学符号
Tab.1 Math notations

符号	注释
L	含有 n_l 个样本的已标注数据集
U	含有 n_u 个样本的无标注数据集
t	主动学习迭代轮次
b	每轮主动学习迭代挑选的无标注样本数
M_t	第 t 轮迭代训练得到的目标模型
$P_{j,k}$	模型 M_k 对无标注样本 x_j 的预测向量
$H_{j,k}$	模型 M_k 对无标注样本 x_j 的预测向量的熵
$\hat{H}_{j,k}$	相邻模型熵增 ($H_{j,k} > H_{j,k-1}$) 的指示函数取值
N	不稳定性采样保存的最近历史目标模型个数
s_j	样本 x_j 的不稳定性取值

图 2 展示了基于不稳定性采样的主动学习框架。除在第一轮主动学习迭代中,采用随机采样挑选样本,此后的每一轮迭代,都使用离当前轮次最近的 N 个历史分类模型 $\{M_{t-1}, M_{t-2}, \dots, M_{t-N}\}$ 对每一个无标注样本 x_j 进行预测,得到 N 个后验概率。在此基础上,使用不稳定性采样来估计每个无标注样本的不稳定性,并挑选最不稳定的样本进行标注。

2.2 不稳定性采样策略

正如前文提到的,在主动学习的第 t 轮迭代中,以往的模型,即 $\{M_{t-1}, M_{t-2}, \dots, M_{t-N}\}$ 对无标

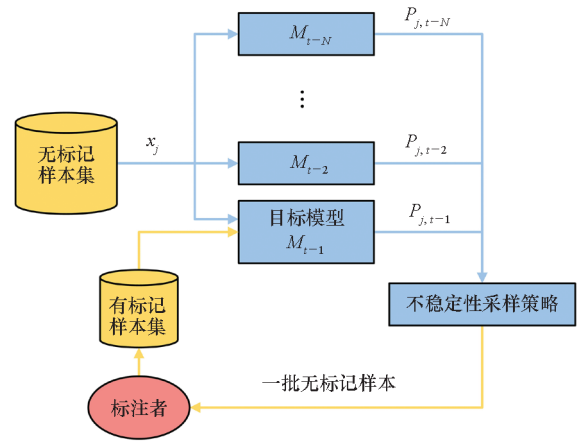


图 2 学习框架

Fig. 2 Learning framework

注样本 x_j 预测结果不稳定,这表明目标模型对该样本的识别能力不足。预测越不稳定,该样本越难以被有效地识别。因而,应尽可能挑选最不稳定的样本进行查询。

首先介绍本文使用的度量模型对未标注样本识别能力的指标。现有主动学习往往用信息熵来衡量模型预测的不确定性,普遍认为模型越难以判断样本所属类别,该模型的识别能力就越低。

$$H_{j,k} = \sum_i P_{j,k}^i \log_2 \frac{1}{P_{j,k}^i} \quad (1)$$

上式的含义是模型 M_k 对样本 x_j 的预测向量的熵。计算 N 个历史模型对同一个无标注样本的后验概率的熵,旨在根据识别能力对相邻两轮迭代的目标模型进行排序。为此,定义:

$$\hat{H}_{j,t-k} = \begin{cases} 1, & H_{j,t-k} > H_{j,t-k-1} \\ 0, & H_{j,t-k} \leq H_{j,t-k-1} \end{cases} \quad (2)$$

当相邻两轮迭代中,模型的信息熵增加,说明模型对样本的识别能力变弱,此时变弱的程度即为本文提出的不稳定性程度的度量;相反,模型的信息熵减少,说明模型识别能力趋于稳定。

使用后验概率分布的差异来衡量模型识别能力变弱的大小。常用于衡量两个分布差异的度量方式有 KL 散度、JS 散度和 Wasserstein 距离。

$$D_{KL}(P_1, P_2) = \sum_i P_1^i \ln \frac{P_1^i}{P_2^i} \quad (3)$$

$$D_{JS}(P_1, P_2) = \frac{1}{2} D_{KL}\left(P_1, \frac{P_1 + P_2}{2}\right) + \frac{1}{2} D_{KL}\left(P_2, \frac{P_1 + P_2}{2}\right) \quad (4)$$

$$D_W(P_1, P_2) = \inf_{\gamma \in \Pi(P_1, P_2)} E_{(x,y) \sim \gamma} (\|x - y\|) \quad (5)$$

通过实验发现,使用上述几种度量方式的实验性能相当,因此,统一采用 Wasserstein 距离来

度量模型识别能力变弱的程度。

基于信息熵和分布差异,引入 s_j 衡量无标注样本 x_j 的不稳定性。使用同一未标注样本的信息熵来对相邻模型识别效果进行排序,熵增时,累计模型性能变化大小。具体的,在第 t 轮迭代过程中,首先计算离当前轮次最近的前 N 个目标模型 $\{M_{t-1}, M_{t-2}, \dots, M_{t-N}\}$ 对无标注样本 x_j 的预测向量 $\{P_{j,t-1}, P_{j,t-2}, \dots, P_{j,t-N}\}$ 。然后,通过式(6)计算 x_j 的不稳定性 s_j 。

$$s_j = \sum_{k=1}^{N-1} \widehat{H}_{j,t-k} D(P_{j,t-k}, P_{j,t-k-1}) \quad (6)$$

其中, $D(P_1, P_2)$ 代表 Wasserstein 距离。

根据式(6)得到每一个无标注样本的不稳定性后,选择值最大的 b 个样本进行标注并加入训练。算法1中总结了不稳定性采样的计算框架。

算法1 基于不稳定性采样的主动学习方法

Alg.1 Active learning based on instability sampling

已知:无标注数据集 U , 批大小 b , 历史模型 $\{M_{t-1}, M_{t-2}, \dots, M_{t-N}\}$

1. for x_j in U
2. 根据历史模型 $\{M_{t-1}, M_{t-2}, \dots, M_{t-N}\}$ 得到样本 x_j 的后验概率 $\{P_{j,t-1}, P_{j,t-2}, \dots, P_{j,t-N}\}$;
3. 根据式(1)计算熵 $\{H_{j,t-1}, H_{j,t-2}, \dots, H_{j,t-N}\}$;
4. 根据式(6)计算不稳定性 s_j ;
5. end for
6. 挑选不稳定性排名前 b 个的样本作为查询样本集

3 实验

本节主要介绍实验设置以及使用不同模型在多个数据集上的实验结果。

3.1 实验设置

在9个数据集上进行实验,表2展示了所用数据集的样本数量,特征维度以及类别数目。为了进一步验证所提出方法在传统模型和深度模型上的有效性,使用不同的基分类模型,包括逻辑斯蒂回归(logistic regression, LR)模型、LeNet-5和ResNet18。所有实验使用的都是未经过预训练的初始化模型。

将数据集划分为70%的训练样本和30%的测试样本。在传统模型上,训练集随机采样5%的样本来初始化有标注样本集,在每轮主动学习迭代中,通过采样策略挑选 $b=1$ 个未标注样本进行标注并加入有标注集,总标注预算为200。在深度模型上,初始的已标注训练样本占整个训练

表2 实验所用数据集

Tab.2 Datasets used in the experiments

数据集	样本数量	特征维度	类别数目
iris	150	4	3
wine	178	13	3
new-thyroid	215	5	3
clean1	476	168	2
breastcancer	569	30	2
digits	1 797	64	10
mnist	70 000	28 × 28	10
fashionmnist	70 000	28 × 28	10
svhn	99 289	3 × 32 × 32	10

集的0.5%,总标注预算为500;svhn除外,其随机采样1%的样本来初始已标注集,总标注预算为2000。不同的标注预算依据实验最终的收敛情况而定,以便于观察主动学习采样策略性能。深度模型在每轮主动学习迭代中挑选 $b=10$ 个样本进行标注。

实验初始学习率设为0.01,批量大小为64,在mnist和fashionmnist数据集上进行的实验每50次迭代更新学习率为原来的10%,svhn数据集每20次迭代更新学习率为原来的90%。重复地进行5次实验,计算每轮主动学习迭代中目标模型的平均准确率,并绘制平均准确率随查询样本数的变化曲线,曲线提升得越快,说明采样策略性能越高。

为了验证方法的有效性,将不稳定性采样方法与下列方法进行对比:

- 1) 随机采样:对未标注样本进行随机采样。
- 2) 最低置信度采样:根据模型对未标注样本的预测后验概率,挑选一批置信度最低的样本进行标注。
- 3) 最大熵采样:基于模型的预测后验概率计算每一个未标注样本的信息熵,并挑选一批信息熵最大的样本。

上述所有比对方法都基于ALiPy^[46]库提供的接口,使用默认参数进行对比实验。

3.2 实验结果与分析

实验结果如图3所示,其中图3(a)~(f)是传统模型的实验效果,图3(g)~(i)是深度模型的实验结果。从图中可以看出:不稳定性采样方法在大多数情况下都实现了最佳性能;不稳定性采样几乎在所有情况下都显著优于基准方法随机

采样;不稳定性采样几乎在所有情况下都与不确定性采样方法(最低置信度采样和最大熵采样)表现相当或优于他们。这些实验结果证明了不稳定性采样能有效地挑选对模型最有用的样本,并提升主动学习性能;同时说明了考虑历史模型预测不稳定性比仅基于当前模型挑选样本带来的潜在效用大。

3.3 模型数量的影响

本小节进一步研究历史模型数量 N 对实验结果的影响。图 4 为模型数量对比实验结果,

分别设置 $N=2,3,5$,并展示出性能曲线。通过观察发现,当 $N=5$ 时,不稳定性采样方法的性能比 $N=2$ 和 $N=3$ 的性能差。原因可能为本文使用距当前主动学习轮次最近的前 N 个历史模型进行实验,随着主动学习迭代轮次的增加,前几轮训练得到的模型性能较弱,这些模型预测的后验概率准确率较低,计算得到的数据的不稳定都较高,使得筛选出的数据可能不是“预期的高质量数据”,最终导致随着 N 增大,本文方法的效果下降。

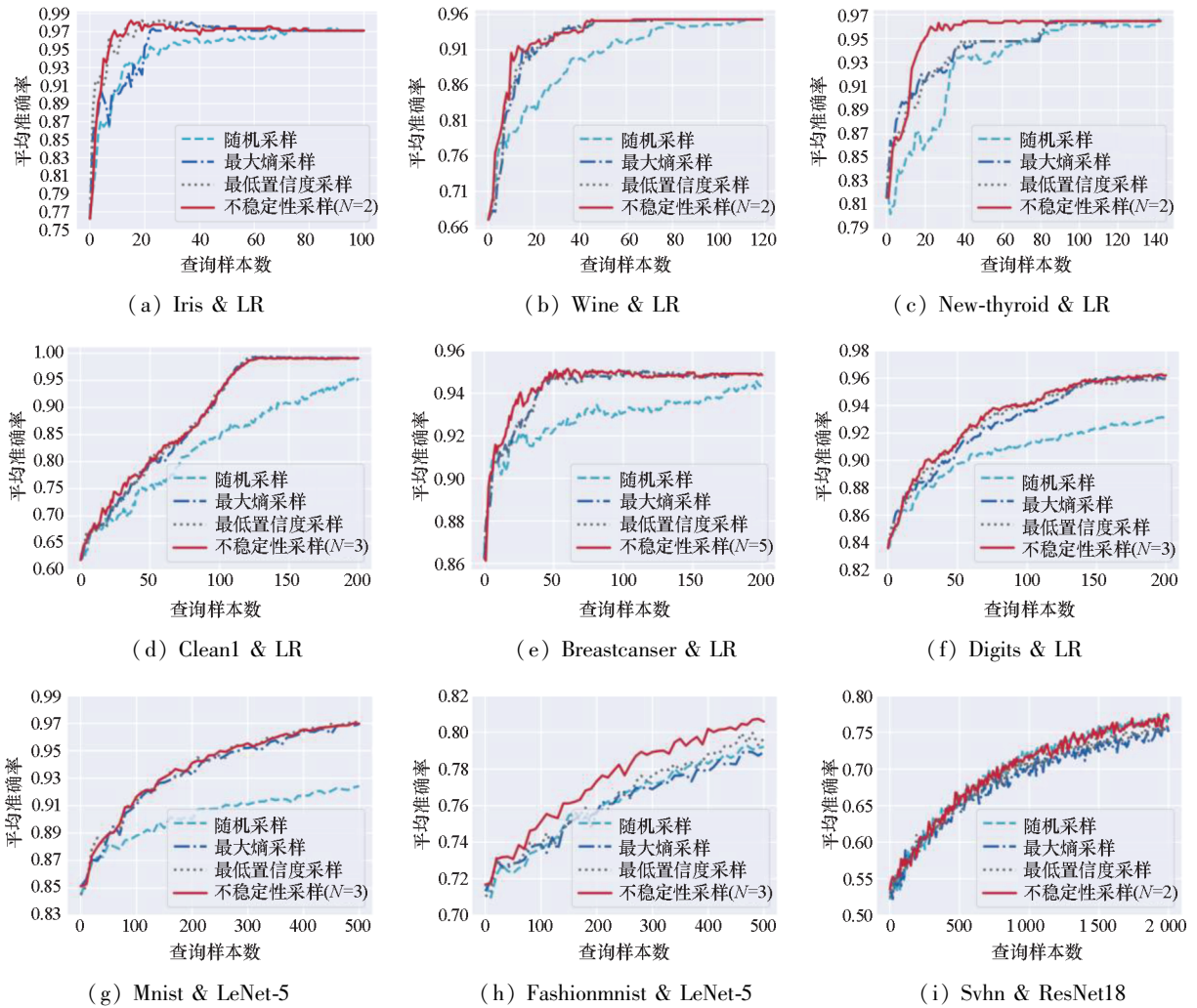
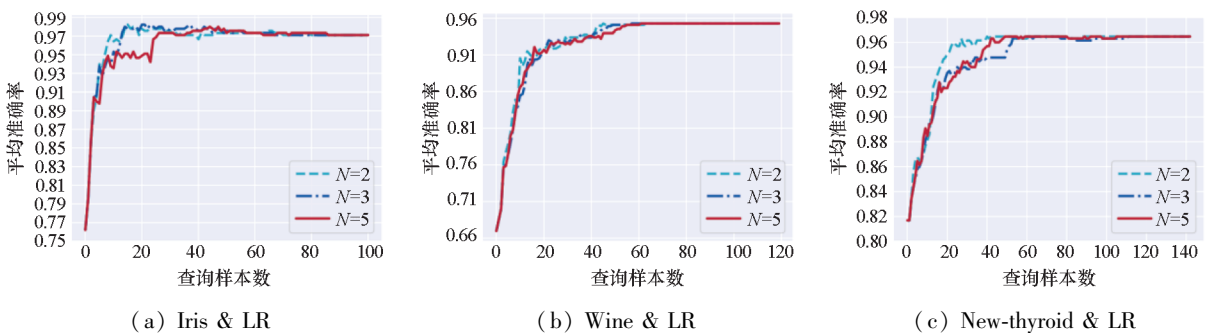


图 3 性能对比实验结果

Fig. 3 Performance comparison of experimental results



(a) Iris & LR

(b) Wine & LR

(c) New-thyroid & LR

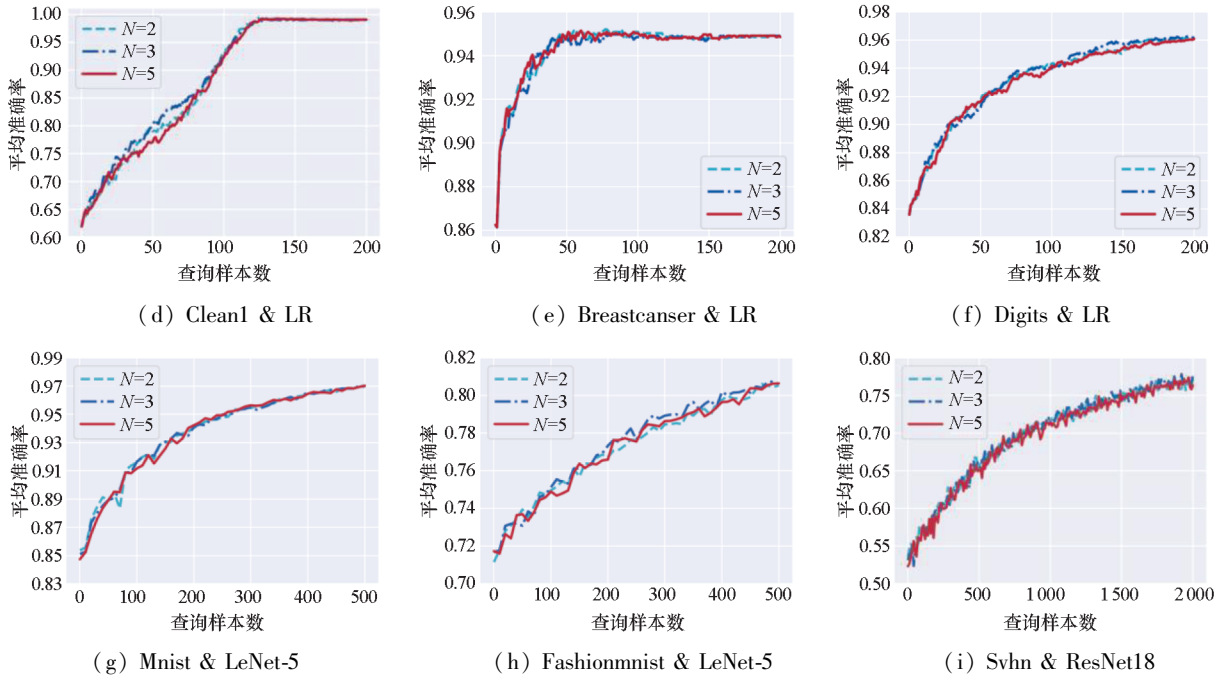


图4 模型数量对比实验结果

Fig. 4 Comparative experimental results on the different number of models

4 结论

本文提出了一种基于不稳定性采样的主动学习方法,通过以往模型的预测差异来衡量无标注样本的信息量。该方法考虑了以往模型在同一无标注样本上的预测的不稳定性,使挑选的样本对改善模型性能具有较高的潜在价值。为了充分验证方法的有效性,使用传统和深度模型在多个数据集上进行实验。实验结果表明,不稳定采样方法优于经典的基于不确定性的主动学习方法。未来,打算将不稳定采样方法用于其他实际任务中,如目标检测任务。

参考文献 (References)

- [1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84–90.
- [2] SERMANET P, EIGEN D, ZHANG X, et al. Overfeat: integrated recognition, localization and detection using convolutional networks[EB/OL]. (2014–02–24) [2021–06–01]. <https://arxiv.org/abs/1312.6229>.
- [3] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]//*Proceedings of the 3rd International Conference on Learning Representations*, 2015.
- [4] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [5] ZHAO W, PENG H Y, EGER S, et al. Towards scalable and reliable capsule networks for challenging NLP applications[C]//*Proceedings of the 57th Annual Meeting of*

- the Association for Computational Linguistics, 2019.
- [6] STRUBELL E, GANESH A, MCCALLUM A. Energy and policy considerations for deep learning in NLP [C]//*Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019.
- [7] LIU H, YIN Q Y, WANG W Y. Towards explainable NLP: a generative explanation framework for text classification[C]//*Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019.
- [8] LI X Y, SUN X F, MENG Y X, et al. Dice loss for data-imbalanced NLP tasks[EB/OL]. (2020–08–29) [2021–06–01]. <https://arxiv.org/abs/1911.02855>.
- [9] FLEK L. Returning the N to NLP: towards contextually personalized classification models [C]//*Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020: 7828–7838.
- [10] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 580–587.
- [11] ERHAN D, SZEGEDY C, TOSHEV A, et al. Scalable object detection using deep neural networks [C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 2155–2162.
- [12] GIRSHICK R. Fast R-CNN [C]//*Proceedings of IEEE International Conference on Computer Vision*, 2015: 1440–1448.
- [13] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149.
- [14] CLARK D G, LIVEZEY J A, BOUCHARD K E. Unsupervised discovery of temporal structure in noisy data with dynamical components analysis[C]//*Proceedings of the 33rd Conference on Neural Information Processing Systems*, 2019.

- [15] SOHN K, BERTHELOT D, LI C L, et al. Fixmatch: simplifying semi-supervised learning with consistency and confidence[C]//Proceedings of the 34th Conference on Neural Information Processing Systems, 2020.
- [16] CHEN J, SHAH V, KYRILLIDIS A. Negative sampling in semi-supervised learning [C]//Proceedings of the 37th International Conference on Machine Learning, 2020.
- [17] LOCATELLO F, POOLE B, RÄTSCH G, et al. Weakly-supervised disentanglement without compromises [C]//Proceedings of the 37th International Conference on Machine Learning, 2020: 6348–6359.
- [18] LI X Y, KAN M N, SHAN S G, et al. Weakly supervised object detection with segmentation collaboration [C]//Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV), 2019: 9735–9744.
- [19] SETTLES B. Active learning literature survey[R]. Madison: University of Wisconsin-Madison, 2010.
- [20] BALCAN M F, BRODER A, ZHANG T. Margin based active learning[C]//Proceedings of the 20th Annual Conference on Learning Theory, 2007.
- [21] GAL Y, ISLAM R, GHAHRAMANI Z. Deep Bayesian active learning with image data [C]//Proceedings of the 34th International Conference on Machine Learning, 2017.
- [22] YAN Y F, HUANG S J. Cost-effective active learning for hierarchical multi-label classification[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018: 2962–2968.
- [23] SIDDIQUI Y, VALENTIN J, NIESSNER M. ViewAL: active learning with viewpoint entropy for semantic segmentation[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 9430–9440.
- [24] WANG K Z, ZHANG D Y, LI Y, et al. Cost-effective active learning for deep image classification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(12): 2591–2600.
- [25] SEUNG H S, OPPER M, SOMPOLINSKY H. Query by committee[C]//Proceedings of the 5th Annual Workshop on Computational Learning Theory, 1992: 287–294.
- [26] KRAWCZYK B, CANO A. Adaptive ensemble active learning for drifting data stream mining[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence, 2019: 2763–2771.
- [27] CORTES C, DESALVO G, GENTILE C, et al. Adaptive region-based active learning [C]//Proceedings of the 37th International Conference on Machine Learning, 2020.
- [28] FREYTAG A, RODNER E, DENZLER J. Selecting influential examples: active learning with expected model output changes [C]//Proceedings of the 13th European Conference on Computer Vision, 2014.
- [29] KÄDING C, RODNER E, FREYTAG A, et al. Active and continuous exploration with deep neural networks and expected model output changes [C]//Proceedings of the 30th Conference on Neural Information Processing Systems, 2016.
- [30] KRAWCZYK B, CANO A. Adaptive ensemble active learning for drifting data stream mining[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence, 2019: 2763–2771.
- [31] CHATTOPADHYAY R, WANG Z, FAN W, et al. Batch mode active sampling based on marginal probability distribution matching [C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2012: 741–749.
- [32] HAJMOHAMMADI M S, IBRAHIM R, SELAMAT A, et al. Combination of active learning and self-training for cross-lingual sentiment classification with density analysis of unlabelled samples [J]. Information Sciences, 2015, 317: 67–77.
- [33] KIRSCH A, VAN AMERSFOORT J, GAL Y. BatchBALD: efficient and diverse batch acquisition for deep Bayesian active learning[C]//Proceedings of the 33rd Conference on Neural Information Processing Systems, 2019.
- [34] WU D R. Pool-based sequential active learning for regression[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(5): 1348–1359.
- [35] HSU W N, LIN H T. Active learning by learning [C]//Proceedings of the 29th AAAI Conference on Artificial Intelligence, 2015: 2659–2665.
- [36] HUANG S J, ZHOU Z H. Active query driven by uncertainty and diversity for incremental multi-label learning [C]//Proceedings of IEEE 13th International Conference on Data Mining, 2013.
- [37] YANG Y, MA Z G, NIE F P, et al. Multi-class active learning by uncertainty sampling with diversity maximization [J]. International Journal of Computer Vision, 2015, 113: 113–127.
- [38] HUANG S J, JIN R, ZHOU Z H. Active learning by querying informative and representative examples [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(10): 1936–1949.
- [39] WANG Z, YE J P. Querying discriminative and representative samples for batch mode active learning [J]. ACM Transactions on Knowledge Discovery from Data, 2015, 9(3): 1–23.
- [40] YANG L, ZHANG Y Z, CHEN J X, et al. Suggestive annotation: a deep active learning framework for biomedical image segmentation [C]//Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2017: 359–407.
- [41] TANG Y P, HUANG S J. Self-paced active learning: query the right thing at the right time[J]. Proceedings of the 33rd AAAI Conference on Artificial Intelligence, 2019, 33: 5117–5124.
- [42] HOI S C H, JIN R, ZHU J K, et al. Semi-supervised SVM batch mode active learning for image retrieval [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008.
- [43] LIU M, BUNTINE W, HAFFARI G. Learning how to actively learn: a deep imitation learning approach [C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018.
- [44] YOO D, KWEON I S. Learning loss for active learning [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 93–102.
- [45] LI M H, LIU X L, VAN DE WEIJER J, et al. Learning to rank for active learning: a listwise approach [C]//Proceedings of the 25th International Conference on Pattern Recognition (ICPR), 2021: 5587–5594.
- [46] TANG Y P, LI G X, HUANG S J. ALiPy: active learning in python[EB/OL]. (2019–01–12) [2021–06–01]. <https://arxiv.org/pdf/1901.03802.pdf>.