

自适应主动半监督学习方法*

李延超¹, 肖甫¹, 陈志¹, 李博²

¹(南京邮电大学 计算机学院 软件学院 网络空间安全学院, 江苏 南京 210023)

²(南京理工大学 计算机科学与工程学院, 江苏 南京 210094)

通讯作者: 李延超, E-mail: yanchao@njupt.edu.cn



摘要: 主动学习从大量无标记样本中挑选样本交给专家标记. 现有的批抽样主动学习算法主要受 3 个限制: (1) 一些主动学习方法基于单选择准则或对数据、模型设定假设, 这类方法很难找到既有不确定性又有代表性的未标记样本; (2) 现有批抽样主动学习方法的性能很大程度上依赖于样本之间相似性度量的准确性, 例如预定义函数或差异性衡量; (3) 噪声标签问题一直影响批抽样主动学习算法的性能. 提出一种基于深度学习批抽样的主动学习方法. 通过深度神经网络生成标记和未标记样本的学习表示和采用标签循环模式, 使得标记样本与未标记样本建立联系, 再回到相同标签的标记样本. 这样同时考虑了样本的不确定性和代表性, 并且算法对噪声标签具有鲁棒性. 在提出的批抽样主动学习方法中, 算法使用的子模块函数确保选择的样本集合具有多样性. 此外, 自适应参数的优化, 使得主动学习算法可以自动平衡样本的不确定性和代表性. 将提出的主动学习方法应用到半监督分类和半监督聚类中, 实验结果表明, 所提出的主动学习方法的性能优于现有的一些先进的方法.

关键词: 主动学习; 半监督学习; 分类; 聚类

中图法分类号: TP181

中文引用格式: 李延超, 肖甫, 陈志, 李博. 自适应主动半监督学习方法. 软件学报, 2020, 31(12): 3808–3822. <http://www.jos.org.cn/1000-9825/5890.htm>

英文引用格式: Li YC, Xiao F, Chen Z, Li B. Adaptive active learning for semi-supervised learning. Ruan Jian Xue Bao/Journal of Software, 2020, 31(12): 3808–3822 (in Chinese). <http://www.jos.org.cn/1000-9825/5890.htm>

Adaptive Active Learning for Semi-supervised Learning

LI Yan-Chao¹, XIAO Fu¹, CHEN Zhi¹, LI Bo²

¹(School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

²(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: Active learning algorithms attempt to overcome the labeling bottleneck by asking queries from a large collection of unlabeled examples. Existing batch mode active learning algorithms suffer from three limitations: (1) the models with assumption on data are hard in finding images that are both informative and representative; (2) the methods that are based on similarity function or optimizing certain diversity measurement may lead to suboptimal performance and produce the selected set with redundant examples; (3) the problem of noise labels has been an obstacle for active learning algorithms. This study proposes a novel batch mode active learning method based on deep learning. The deep neural network generates the representations (embeddings) of labeled and unlabeled examples, and label cycle mode is adopted by connecting the embeddings from labeled examples to those of unlabeled examples and back at the same class, which considers both informativeness and representativeness of examples, as well as being robust to noisy labels. The proposed active learning method is applied to semi-supervised classification and clustering. The submodular function is designed to reduce the redundancy of the selected examples. Moreover, the query criteria of weighting losses are optimized in active learning, which automatically trade off the

* 基金项目: 国家自然科学基金(61932013); 江苏省自然科学基金(BK20200739); 江苏省 333 高层次人才培养工程(BRA2020065)

Foundation item: National Natural Science Foundation of China (61932013); Natural Science Foundation of Jiangsu Province of China (BK20200739); Research Foundation of Jiangsu for 333 High Level Talents Training Project (BRA2020065)

收稿时间: 2019-07-07; 修改时间: 2019-07-28; 采用时间: 2019-09-16

balance of informative and representative examples. Specifically, batch mode active scheme is incorporated into the classification approaches, in which the generalization ability is improved. For semi-supervised clustering, the proposed active scheme for constraints is used to facilitate fast convergence and perform better than unsupervised clustering. To validate the effectiveness of the proposed algorithms, extensive experiments are conducted on diversity benchmark datasets for different tasks, and the experimental results demonstrate consistent and substantial improvements over the state-of-the-art approaches.

Key words: active learning; semi-supervised learning; classification; clustering

儿童在成人的指导下,能够快速学习并认识新的物体.例如:当孩子看过一只猫之后,或许就能够通过头或尾巴的特征识别其他猫.在得到成人更多的指导下,并且看到越来越多的猫后,他们能够更加自信地认出猫.为了在机器上模拟孩子完成类似的任务,系统应该能够获得人工的监督和指导.由于在许多实际应用中,例如图像分类、文档聚类和语音识别中获取大量未标记数据是相当容易的,但是带标签的数据却是相当“昂贵”,因为它们需要人工花大量的时间标注.此外,在计算机辅助医学诊断中,通常可以从常规检查中获得大量的胸部 X 射线图像,但是要求医生在所有图像中标记所有的病灶是费时和困难的.另外一个例子:在人类姿势估计测试中,通常需要很好的训练数据,因为这需要注释者点击每个人的 14 个关节.所以,人体姿势评估方法^[1]建议每分钟一个姿势的合理注释数据.本文的主要目标是:通过“聪明”地选择未标记的样本来标记,而不是随机的选择样本,则需要更少的标记数据去训练模型,从而节省专家宝贵的时间.

主动学习(active learning)^[2]范式在现代许多机器学习问题中发挥了积极作用,其关键组成部分是选择最具信息性或代表性的样本进行标记:信息性衡量样本减少统计模型不确定性的能力,而代表性衡量样本是否能够很好地代表未标记数据的总体输入模式.现有的大多数主动学习方法^[3-5]通过单一查询标准选择未标记样本,这可能会降低模型的学习性能.具体而言,这些方法在选择最具信息性的样本时并未利用未标记数据的结构,导致严重的样本偏倚;而一些方法^[4,6-8]选择最具代表性的样本时并没有考虑未标记数据的分布,导致模型效果不好.此外,一些方法对数据进行假设,但是这种方法很难在图像数据中获得好的效果.另外,学者们也提出了一些结合两种查询选择准则(样本的不确定性和代表性)的主动学习算法^[9-12],但是这些方法缺乏自适应动态平衡两种准则的机制.

值得注意的是:算法每次选择单个样本进行查询,容易导致模型局部最优.此外,如果基于深度神经网络方法的模型,这会使得算法性能低下并且导致严重的过拟合.为了解决这些问题,学者们已经提出了批抽样主动学习(batch mode active learning)方法^[7,13,14],该方法在每次迭代时选择一组样本.但是,许多启发式方法未能考虑样本之间的相互作用,导致选择的样本集具有较高冗余度.例如:基于集群的方法^[15,16]通过将未标记的样本分组到不同的集群中来减少冗余;基于间隔的方法^[4,17]通过采用优化策略来选择 k 个最小间隔样本.此外,基于相似性函数或优化信息度的方法也同样会导致次优的性能,并产生冗余样本.

受关联学习^[18]的启发,图 1 描述了关联学习:如果样本属于同一类,则训练网络产生高相似性的样本表示;从标记的样本表示到未标记的样本表示,并返回可区分的关联循环.

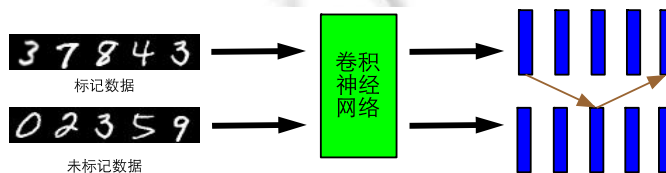


Fig.1 Associative learning

图 1 关联学习

本文提出了主动学习方法,通过深度神经网络生成标记和未标记样本的学习表示和标签循环模式,使得标记样本与未标记样本建立联系再回到相同标签的标记样本,这样同时考虑了样本的信息性(关联概率)和代表性(影响概率).在提出的批抽样主动学习方法中,使用子模函数(submodular function)确保选择的样本集合具有多样性.本文将提出的主动学习方法应用到不同的任务中,即半监督分类和半监督聚类.在分类方法中,将主动学

习方法嵌入到半监督学习算法中,它对准确率驱动型应用任务非常有用^[19],如计算机辅助医学诊断,并且适用于复杂任务,例如人体姿势估计^[1].对于聚类方法,采用样本约束对的主动监督,实现聚类算法的快速收敛,并取得比无监督聚类更好的性能.此外,应用聚类可以发现数据中一些人工发现不了的类别结构.

许多分类模型需要通过获取大量标注的样本数据来训练模型,以取得强泛化性能.例如:标记肺部疾病通常需要专家,但是让医生在所有图像中标记所有病灶是“昂贵”和困难的.这些应用是以性能为驱动的,其目的是提高模型的准确性,并最大限度地减少获取标记数据的成本.根据模型在实际应用中获得的经验,导致模型性能下降主要有3个原因:首先,标签数据不足意味着训练的模型不能反映真实的数据分布和产生强泛化性能;其次,标记的数据有噪音,例如 MNIST 演化数据集;第三,描述标记数据的特征不足以作出所需的判别,例如人的身高特征无助于诊断哪种类型的肺部疾病.此外,基于决策边界方法的错误假设会导致模型选择的样本不具有信息性和代表性.另外,现有批量抽样的主动学习方法选取的样本带有高度冗余性,这些激励我们自适应选择多样化的样本数据.

本文提出的主动学习算法使用深度神经网络方法,学习标记和未标记样本特征表示,进而查询选择具有信息性和代表性的样本,以最小化获得标记数据的成本.而且,提出方法的标签循环机制对噪音标签具有鲁棒性.此外,本文将提出的方法嵌入到不同的任务中,并且性能都得到提升.本文的贡献总结如下.

- (1) 提出了一种结合半监督学习的主动学习方法.采用标签循环模式,使得标记样本与未标记样本建立联系再回到相同标签的标记样本,这样同时考虑了样本的信息性和代表性,并且算法对噪声标签具有鲁棒性;
- (2) 提出了自适应参数的主动学习方法,使得算法自主学习数据的不确定性和代表性权重.此外,子模函数的嵌入确保批抽样所选样本集合的多样性;
- (3) 分别在分类和聚类任务中设计了两种算法,算法实验结果表明:设计的批抽样主动学习方法是有效的,并且算法能够快速收敛;
- (4) 针对不同应用任务,即分类和聚类,算法在基准数据集进行了大量实验.结果表明,提出的方法优于目前先进的算法.

1 主动监督

本节介绍提出的主动学习方法,它同时考虑了样本的不确定性和代表性.此外,本节分析主动学习算法中批处理样本多样性和自适应选择方法,为后面设计的算法提供指导.

1.1 主动选择

主动学习背后的关键思想:如果允许选择模型训练的数据,则机器学习算法可以使用较少的标记训练样本,实现模型更高的准确性.受关联学习的启发,本节对其进行修改并提出了主动学习方法.主动学习方法选择最具信息性(不确定性)和代表性(影响力)的监督实例,交给专家标记.在每次主动学习迭代 $t \in \{0, \dots, T\}$ 中,根据下面定义的监督损失主动选择批量 S_t 中的样本,并添加到已有的标记数据中: $L_t = S_t \cup L_{t-1}$. 提出的主动监督方法同时考虑样本影响和不确定因素,具体而言,不确定性衡量标记和未标记样本之间的相似性.为了防止算法仅在“简单”的例子中进行标记,主动选择考虑了样本的影响力因素,这可以更好地泛化到其他样本,并“传播”更有价值的信息.所以,主动选择的贡献在于样本不确定性和影响力的结合.

- 不确定性

给定一批标记和未标记的样本,将这些样本输入到深度学习模型,例如卷积神经网络(CNN),得到特征向量表示,即 A 和 B .如前所述的关联概率,本文希望最大化相同标签从 A 到 B 并返回 A 的概率.这种特性衡量了样本的不确定信息.

定义 1. 样本表示 A 和 B 的相似性可以定义为

$$M_{ij} = A_i \cdot B_j.$$

其中, A 和 B 表示批数据的矩阵;行表示样本的索引;点积也可以采用其他相似性度量,例如欧式距离.

定义 2. 交叉损失函数输出 A 到 B 中列 M 的概率:

$$P_{ij}^{ab} = P(B_j | A_i) := (\text{softmax}_{\text{cols}}(M))_{ij} = \exp(M_{ij}) / \sum_{j'} \exp(M_{ij'}).$$

所以,从 A_i 到 A_j 的关联概率是

$$P_{ij}^{aba} = (P^{ab} P^{ba})_{ij} = \sum_k P_{ik}^{ab} P_{kj}^{ba}.$$

不确定性的损失定义为标签正确往返 T 的均匀目标分布与关联概率 P^{aba} 之间的交叉熵 L . 只要开始和结束样本的标签都属于同一类,损失函数就会惩罚错误的关联,并鼓励更正分发:

$$\ell_{\text{uncertainty}} = L(T, P^{aba}).$$

其中,带有均匀目标分布:

$$T_{ij} = \begin{cases} 1/|\text{class}(A_i)|, & \text{class}(A_i) = \text{class}(A_j), \\ 0, & \text{otherwise} \end{cases},$$

其中, $|\text{class}(A_i)|$ 是类别 A_i 在 A 中出现的次数. 均匀分布允许循环机制中开始的标签和结束的标签不一致.

- 代表性

未标记的样本可能是低质量的或含有一些噪声. 代表性解释了样本的影响性,并鼓励更好地“影响”其他样本. 为了避免总是在“简单”样本之间建立关联,并降低影响所有样本的复杂度,本文定义均匀目标分布 V 和影响概率 P^{inf} 之间的交叉熵 L :

$$\ell_{\text{influence}} = L(V, P^{\text{inf}}),$$

其中, B 中样本的影响概率为 $P_j^{\text{inf}} = P_{ij}^{ab}$, 均匀目标分布定义为 $V_j = 1/|B|$.

1.2 自适应选择

样本的不确定性和代表性随着标注样本的过程中需要自适应地调节,受多任务学习^[20]权重优化的启发,本文采用自适应优化不确定性和代表性的目标函数学习权重,其目标函数可以表示为

$$L_{\text{total}} = \sum_i \theta_i \ell_i.$$

参数 θ 的选择对主动学习算法性能的影响是很大的;此外,在一些任务中,调节超参数是很花时间的. 所以,本文希望学习优化这个参数来自适应的调节样本的不确定性和代表性的权重. 定义 $f^w(\mathbf{x})$ 作为样本 \mathbf{x} 在神经网络的输出,对于分类任务,定义概率向量为

$$P(y|f^w(\mathbf{x})) = \text{softmax}(f^w(\mathbf{x})).$$

其多目标的似然可以定义为如下:

$$P(y_1, \dots, y_k | f^w(\mathbf{x})) = P(y_1 | f^w(\mathbf{x})) \dots P(y_k | f^w(\mathbf{x})),$$

其中,多目标的输出分别为 y_1, \dots, y_k . 所以,可以得到多目标最小化的优化函数:

$$L = -\log P(y_1, y_2 | f^w(\mathbf{x})) \propto \frac{1}{2\sigma_1^2} \|y_1 - f^w(\mathbf{x})\|^2 + \frac{1}{2\sigma_2^2} \|y_2 - f^w(\mathbf{x})\|^2 + \log \sigma_1 \sigma_2 = \frac{1}{2\sigma_1^2} \ell_1(\mathbf{w}) + \frac{1}{2\sigma_2^2} \ell_2(\mathbf{w}) + \log \sigma_1 \sigma_2.$$

其中, $\ell_1(\mathbf{w}) = \|y_1 - f^w(\mathbf{x})\|^2$ 表示第 1 个输出变量的损失函数,同理表示 $\ell_2(\mathbf{w})$. 我们从数据学习最小化目标函数最后一项的 σ_1 (变量 y_1 的噪声参数) 和 σ_2 (变量 y_2 的噪声参数), 即学习损失函数 $\ell_1(\mathbf{w})$ 和 $\ell_2(\mathbf{w})$ 的权重 θ_1 和 θ_2 . 当 σ_1 增加,其损失函数 $\ell_1(\mathbf{w})$ 的权重减少. 反之亦然.

1.3 批抽样选择

本节介绍批量抽样的方法,由此需要注意两个问题:查询哪些样本以及如何在小批量设置中查询.

(1) 选取哪些样本?

主动学习应选择最具信息性和代表性的样本进行查询. 正如第 1.1 节所讨论的那样,根据动态信息函数,选取一些具有很高不确定性和代表性的样本 U . 具体来说,本文假设 U^i 中的每个元素 u , 输出是分值:

$$\rho_1^i, \dots, \rho_k^i.$$

我们平衡所选集合中的类别,这意味着集合大小为 K 的 S 中,每个类别标签的比例与整个数据 N 相同,即:对 K 中任意 $y \in Y$,使得 $|S \cap N^y| = KN^y/N$;否则,将 KN^y/N 舍入到最接近的整数.通过加入平衡和约束条件 $|S|=K$,可以得到所有 $y \in Y$ 都满足 $m_y(S) = S \cap N^y = KN^y/N$.

(2) 怎么选?

在小批量主动学习设置中,每轮选取小批量数据样本进行标记.例如:批量抽样想要获得 $Q=KT$ 标记的样本,其中, K 是每轮迭代选择的样本数, T 是轮数.给定一组标记的样本 L ,在每一轮 t 中,算法给每个样本 $u \in U \setminus L$ 打上分值 $\rho_u^t > 0$.在前面的步骤下,受文献[21]方法的启发,本文定义提出方法的子模函数,它考虑了数据对数似然函数集 $f: 2^N \rightarrow \mathbb{R}$,将每个子集 $S \subseteq N$ 映射到整个集合 N 上的对数似然分值:

$$f(S) = \sum_{i \in N} \log p(\mathbf{x}^i | y^i; \theta(S)) + \sum_{i \in N} \log p(y^i; \theta(S)),$$

其中, $p(\mathbf{x}_i | y^i; \theta(S))$ 和 $p(y^i; \theta(S))$ 分别是样本 $i \in N$ 带有参数 $\theta(S)$ 的生成式和先验式似然.子数据集选择的思想:选择一个小样本集 S ,使得 $f(S)$ 最大化.本文把先验表达为 $p(y^i; \theta(S)) = m_{y^i}(S) / |S|$,其中: $m_y(S) = \sum_{i \in S} \mathbb{I}\{y^i = y\}$, $p(\mathbf{x}_i | y^i; \theta(S))$ 中带有标签 y^i ,并由最接近样本 i 的样本 j 决定.例如 $j \in \arg \max_{s \in S \cap N^{y^i}} w(i, s)$,其中, $w(i, j) = d - \|\mathbf{x}^i - \mathbf{x}^j\|_2^2$ 定义样本

对 i 和 j 的相似性, $d = \max_{n \in N, n' \in N} \|\mathbf{x}^n - \mathbf{x}^{n'}\|_2^2$ 是最大化的样本距离.生成式似然函数可表达为

$$p(\mathbf{x}^i | y^i; \theta(S)) = ce^{-\|\mathbf{x}^i - \mathbf{x}^j\|_2^2} = ce^{w(i, j) - d} = c'e^{w(i, j)} = c' \exp \left(\max_{s \in S \cap N^{y^i}} w(i, s) \right).$$

其中, c 和 c' 都是常数.所以, $\log p(\mathbf{x}^i | y^i; \theta(S)) = \log c' + \max_{s \in S \cap N^{y^i}} w(i, s)$ 可以表示为

$$f(S) = \sum_{y \in Y} \sum_{i \in N^y} \max_{s \in S \cap N^{y^i}} w(i, s) + \sum_{y \in Y} m_y(N) \log m_y(S) - |N| \log |S| + C.$$

子模函数的第 1 项是:

$$f(S) = \sum_{y \in Y} \sum_{i \in N^y} \max_{s \in S \cap N^{y^i}} w(i, s).$$

因为 $m_y(S)$ 是平衡的,所以第 2 项也是常数. $M(N, I)$ 是使用分区的分区拟阵(partition matroid) $N_{y \in Y}^y$,其中,分区的数据 $B(M)$ 可以定义为 $B(M) = S \subseteq N: |S \cap N^y| = K|N^y|/|N|, \forall y \in Y$.因此,平衡大小为 K 的数据集 S 等价于 S 并且是分区 M 中的一部分,即 $S \in B(M)$.同理,第 2 项和第 3 项也是常数.因此,带约束的子模最大化问题可以表示为

$$\max_{S \in B(M)} f(S).$$

由于不知道样本 U' 的标签,根据关联概率,使用最可能的预测 \hat{y}_u ;然后,在数据集 $U' (2^{U'} \rightarrow \mathbb{R}_+)$ 使用子模块对象 f ;最后,解决以下的优化问题:

$$\max_{|S|=K, S \subseteq U'} f(S).$$

求解方式引用了文献[21]的方法.自适应批抽样主动学习的伪代码见算法 1.

算法 1. 自适应批抽样主动学习(AdaActive).

输入: $U, T, K, \{\beta_t\}_{t=1}^T$, 初始标记样本 L ;

输出: 标记样本 L .

1. **repeat**

2. 用标记样本 L 训练模型,并产生监督分值 ρ^t ;

3. $U^t \in \arg \max_{u \in U \setminus L: U^t = \beta_t} \sum_{u \in U} \rho_u^t$;

4. 获得最优可能的类别并当作假设类别 $\{\hat{y}_u\}_{u \in U^t}$ 和样本 U^t ;

5. 在假设类别 $\{\hat{y}_u\}_{u \in U^t}$ 和样本 U^t 上实例化 $\hat{f}_t: 2^{U^t} \rightarrow \mathbb{R}_+$;

6. 解决 $L' \in \arg \max_{|S|=K: S \subseteq U^t \setminus L} f(S)$;

7. $L=L\cup L'$.
8. **until** $t>T$.

2 自适应的主动半监督学习

2.1 半监督分类

通过主动而不是随机地选择样本,将需要较少的标记样本训练模型,从而节省标记专家宝贵的时间.在准确性驱动的研究项目中,人们可以通过积极和迭代地与领域专家合作,更好地节省他们的时间,提高机器学习系统的性能.本文将提出的主动学习方法应用到先进的半监督学习 MT 模型^[22]和 kNN 模型上.以下是主动学习应用到半监督分类的伪代码:

算法 2. 半监督分类.

输入: v 时刻选择的最佳子集,超参数 β ,模型 M ,无标记样本池的大小 N^U ,初始训练集 N^L ;
输出:训练好的模型 M .

1. **repeat**
2. $V\leftarrow AdaActive(N^U, v, N^L, \beta)$;
3. $N^L\leftarrow N^L\cup V$;
4. $N^U\leftarrow N^U\setminus V$;
5. $M\leftarrow M(N^L)$;
6. **until** 达到预算.

注意到,MT 和 kNN 模型的选择并不重要,因为提出的方法可以与任何监督和半监督学习算法一起使用.在每次迭代中,算法选择最佳子集样本进行主动查询,然后作为训练集进行模型训练.总之,所选样本具有以下 3 种属性:(1) 样本具有信息性和代表性;(2) 所选批次集具有低冗余度;(3) 批量集具有高度多样性.显然,这 3 个属性是提出批抽样主动学习的理想属性.我们在第 3.2 小节中给出实验分析算法的性能.

2.2 半监督聚类

半监督聚类(semi-supervised clustering,简称 SSC)旨在借助用户提供的辅助信息,提高聚类模型的性能.学者们提出了很多种不同的监督方式,其中最有力度的策略是确定样本对的关系.我们定义了每一个“必须链接”(must-link)对 $(x_i, x_j)\in M$,表示 x_i 和 x_j 在同一个簇中.相似地,每一个“无法链接”(cannot-link)对 $(x_i, x_j)\in C$ 表示 x_i 和 x_j 不在同一个簇中.例如:对于文档聚类,获取“必须链接”和“无法链接”约束需要用户扫描相关文档,并确定它们之间的关系,这是可行的,但是时间成本很高.我们研究的问题是:如何有效地选择成对查询,进而提高聚类模型的性能.

我们对链接约束使用主动学习方法,以实现快速收敛,并且取得比无监督聚类更好地效果.这里给出主动学习应用到半监督聚类的形式化定义.给定样本集 $D=\{x_1, \dots, x_n\}$,其中每个样本 $x_i\in\mathbb{R}^d$.假设簇的数目为 c .在主动学习方法设置中,信息可以通过下面这种查询方式获取:“样本 x_i 和 x_j 属于同一类吗?”,样本对 (x_i, x_j) 表示为一个查询,这个查询的答案为 $l_{ij}\in\{M, C\}$.特别地,如果 $y_i=y_j(y_i\neq y_j)$,标签返回“M”(“C”).在每次迭代中,基于数据集 D 和当前约束集合 C 选择一个或多个查询.“必须链接”和“无法链接”约束满足以下性质.

- $(x_i, x_j, M)\wedge(x_i, x_k, M)\Rightarrow(x_j, x_k, M)$;
- $(x_i, x_j, M)\wedge(x_i, x_k, C)\Rightarrow(x_j, x_k, C)$.

此外,我们使用组(group)的概念来区分不同的类标签.其主要思想:如果确认组中一个样本,那么就可以推断出组中所有的样本都有成对关系.给定约束集 C ,可以确定 k 个组 $\Sigma=\{\Sigma_1, \dots, \Sigma_k\}$,其中, $k\leq c$, c 是类别的总数.组 Σ_i 中的每个样本都包含“必须链接”,并且与其他组有“无法链接”对.如果不存在“无法链接”对,则可以推断出一个新的组.从另一个视角可以将组看作标记样本,即相同(不同)组中的样本具有相同(不同)的标签.为了使用少量的查询,获取大量样本之间的约束,采用主动学习策略,以逐步增加组中的样本,减少查询样本比对的过程.

在提出的算法 3 中,首先将约束转换为组中的样本.算法在每次迭代中,首先在数据集 D 和约束 C 上执行半监督聚类方法,然后根据半监督聚类的结果、数据集 D 和组,使用主动学习方法选择最具信息性和代表性的样本.对于所选集中的每个样本,在每个组中按概率递减顺序输出 x_i ,并查询组 Σ_i 中的最高概率数据样本.最后,更新约束并将其加入到组中,算法运行直到没有可加入的“必须链接”,重复此过程,达到算法收敛的条件.

算法 3. 半监督聚类.

输入:样本数据 N ,类别总数 c , v 时刻选择的最佳子集,超参数 β .

初始化: $C=\emptyset$; $\Sigma=\{\Sigma_1,\dots,\Sigma_k\}$.

输出: $SSC(N,C)$.

1. **repeat**
2. $\pi=SSC(N,C)$;
3. $V=AdaActive(N,\pi,\Sigma,v,\beta)$;
4. **for each** x_i **in** V **do**
5. **for each** $\Sigma_j \in \Sigma$ **in decreasing order of probability** $P(x_i \in \Sigma_j)$ **do**
6. 查询 x_i 与集合 $x_i \in \Sigma_j$ 中任意一个样本的关系;
7. 根据返回的答案更新 C ;
8. **if** (x_i, x_k, M) **then**
9. $\Sigma_j = \Sigma_j \cup \{x_i\}$;
10. **break**;
11. **end if**
12. **end for**
13. **if no must-link is achieved then**
14. $l++$;
15. $\Sigma_l = x_i$;
16. $\Sigma = \Sigma \cup \Sigma_l$;
17. **end if**
18. **end for**
19. **until** 达到预算.

3 实验分析

本节首先介绍在不同任务中使用的数据集,然后将提出的算法在半监督分类和半监督聚类任务中与一些基准方法进行比较.最后,我们研究了算法参数对分类和聚类任务的影响.

3.1 数据集

MNIST 数据集总共包含 70 000 个手写数字样本,每个样本(0~9)像素大小为 28×28 .同时,本文还采用了 CIFAR-10 数据集,其中包含 10 个类别共 60 000 个 32×32 的彩色图像,其中,每个类别 6 000 张图像.官方数据中有 50 000 个训练图像和 10 000 个测试图像.

3.2 分类任务

本文在 MNIST 和 CIFAR-10 数据集上进行了分类算法实验.在下面的实验中,数据分为 3 个部分:基础训练集、未标记的池集和测试集.训练数据用作初始标记数据,以训练初始分类器;而测试数据用来测试算法性能,并评估不同的主动学习算法;池集具有大量的未标记数据,以选择最具信息性和代表性的样本.在每个数据集上,实验通过改变训练集、池集和测试集的样本,实验重复 10 次并记录算法的性能.第 3.2.2 节中记录了各种基准和相关主动学习方法的分类性能.

3.2.1 对比方法和实验设置

为了验证提出方法的有效性,将它与基准和一些相关的最新方法进行比较.

- 1) RAND:随机数据选择作为基准方法;
- 2) S-MARGIN^[3]:简单间隔算法,旨在查询接近边界的数据样本;
- 3) QUIRE^[10]:基于每个样本的信息性和代表性,计算样本不确定性;
- 4) MMC^[23]:使用模型变化最大化的思想来查询影响力的样本;
- 5) FASS^[21]:结合子模函数和深度学习查询高信息度的样本;
- 6) EE-BMAL^[14]:将深度相似性度量引入到批抽样的主动学习中;
- 7) ASCENT^[12]:基于深度学习特征抽取的方法选取高影响力的样本.

3.2.2 分类实验

在本节中,本文将记录每轮主动查询的准确性.对于 MNIST 数据集,算法在每轮中主动查询 100 个样本;对于 CIFAR-10 数据集,算法每轮中主动查询 0.5k 个样本.图中每条曲线显示算法随机独立运行 10 次后的平均性能.实验显示,提出的 AdaActive 方法始终比基准方法表现的更好.实验证明了 AdaActive 方法是有效的.在图 2 中,RAND 方法在 MNIST 数据集上和 MMC 方法在 CIFAR-10 数据集上,算法性能得到提升,但是提升的程度没有提出的 AdaActive 方法大.例如:经过 10 轮主动查询后,AdaActive 方法在 MNIST 数据集上的分类准确率为 99.37%,性能超过所有的对比方法.此外,RAND,S-MARGIN 或 MMC 方法在增加查询的过程中有时会降低性能.例如,S-MARGIN 和 MMC 方法分别在 MNIST 和 CIFAR-10 数据集上性能会下降.类似的结果如图 3 所示.

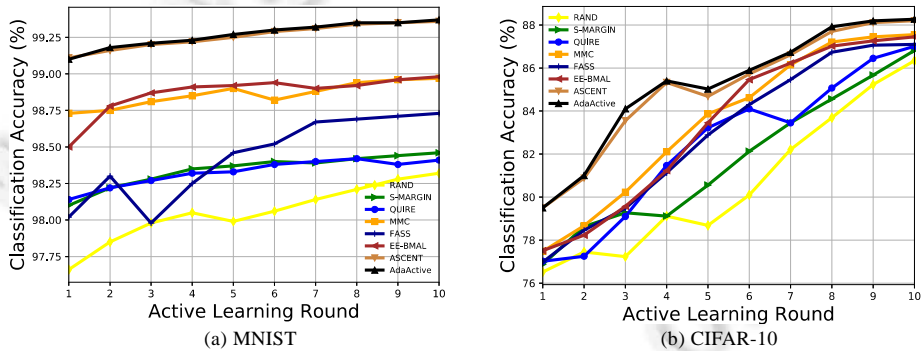


Fig.2 Comparison results of different active learning algorithms by taking MT model as the base classification classifier on the datasets

图 2 各算法在使用 MT 模型作为元学习器的分类结果

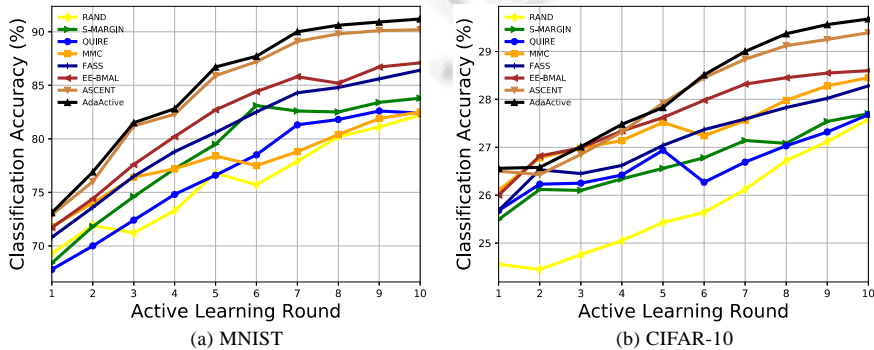


Fig.3 Comparison results of different active learning algorithms by taking best kNN as the base classification model on the datasets

图 3 各算法在使用 kNN 作为元学习器的分类结果

随着查询数量的增加,学习器模型训练效果越来越好,从而使得提出的方法具有更显著的性能优势.此外,使用 k NN 的分类精度的绝对值通常低于使用 MT 模型(采用卷积神经网络)的结果,但仍然比基准方法表现得更好.这些结果符合本文关于不同级别分类难度的表述,并证明 AdaActive 方法可以在困难分类任务中很好地工作.实验中发现一个有趣的现象:在主动学习开始的几轮中,准确度的提升并不明显.这是因为未经训练的网络已经能够基于所产生的样本表示进行一些初步关联.但是,网络会产生许多错误的关联,并且只会受到无监督小批量 B 中少数样本的影响,这些样本与监督小批量 A 中的样本最相似.随着训练的加深,样本表示之间的关联会变得更好.此外,样本代表性的目标函数设计确保 B 中的所有样本受到相同概率的选择.这种分类实验表明:通过选择“正确”的样本,可以提高模型的性能,也突出了主动学习的重要作用.

3.3 聚类任务

本节将评估聚类任务中主动学习与其他基准方法的表现.算法在 MNIST 数据集上进行了测试.为了评估聚类算法的有效性,本文使用归一化互信息(normalized mutual information,简称 NMI)和 F -measure 作为实验中的评估标准.本文使用众所周知的半监督聚类算法 MPCCKmeans^[24]作为元学习器,它在度量学习和聚类步骤之间交替.值得注意的是:选择 MPCCKmeans 算法并不重要,因为提出的算法可以与任何基于约束的聚类模型一起使用.在实验中,MPCCKmeans 模型的最大迭代次数设置为 220.对于 MNIST 数据集,算法每轮选择 15 个成对查询;基于数据集的实际类标签回答查询.为了获得可靠的结论,每个实验都会重复 10 次,并根据评估指标记录模型性能.第 3.3.2 小节中记录了各种基准和相关方法的 NMI 和 F -measure 值.

3.3.1 对比方法和实验设置

为了验证提出方法的有效性,将它与基准和一些相关的最新方法进行比较.

- 1) RAND:选择当前约束集合 C 中未包含或推导的随机成对查询;
- 2) Min-Max^[25]:一种基于邻域的方法,首先遍历构建 c (群集总数)个不相交的邻域;然后使用基于距离的 Min-Max 标准,选择要查询的样本,逐渐扩展样本的邻域;
- 3) QUIRE^[10]:根据每个样本的信息性和代表性来计算样本不确定性;
- 4) NPU^[26]:通过选择信息样本并查询它们与邻居的关系,扩展约束对的数量;
- 5) FASS^[21]:结合子模函数和深度学习查询高信息度的样本;
- 6) URASC^[27]:根据不确定性减少原则来选择聚类成对约束;
- 7) ASCENT^[12]:基于深度学习特征抽取的方法选取高影响力的样本.

3.3.2 聚类实验

本节记录算法在 MNIST 数据集上不同主动查询次数的 NMI 值.实验结果显示在图 4 中,每条曲线显示了方法随机独立运行 10 次后的平均性能

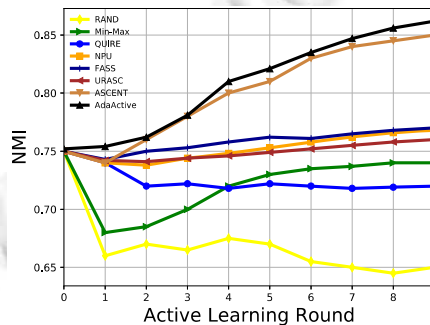


Fig.4 The NMI of different methods on the MNIST dataset

图 4 各方法在 MNIST 数据集上的 NMI 值

可以观察到,提出的 AdaActive 方法始终比基准方法表现的更好.相比之下,NPU 和 URASC 方法通常能够

在增加查询数量时提升性能,但没有提出的方法更加有效.例如:经过 10 轮主动查询后,AdaActive 方法的 *NMI* 值为 0.862,性能超过所有的对比方法.此外,RAND 方法有时会降低模型的性能,这也可以在之前的分类结果中发现相似的现象.这进一步说明了选择正确约束对的重要性.值得注意的是:在算法训练的早期阶段,提出方法的性能接近于其他方法;但随着主动查询数量的增加,性能变得更加强大.这是因为随着训练的加深,样本表示之间的关联会变得更好.另外,样本代表性的目标函数设计确保无标记数据中的所有样本受到相同概率的选择.随着主动查询数量的增加,聚类模型越来越鲁棒,从而凸显了提出的方法的优势.在实际中,实验发现算法有时候会提前收敛.例如,AdaActive 方法在 MNIST 数据集上查询 120 次就已经收敛.实验结果表明,主动学习可以提高聚类算法的性能.

此外,实验还观察了 RAND,Min-Max,QUIRE,NPU,FASS,URASC,ASCENT 以及 AdaActive 方法在 MNIST 数据集上的 *F-measure* 表现.对于每次查询,采用粗体突出显示表现最佳的方法.值得注意的是,非随机方法的性能通常在最初的几次运行中没有显示出统计学上的显著差异.QUIRE 和 Min-Max 方法取得了比随机方法更好的效果,然而提升的程度没有提出的 AdaActive 方法高.与 RAND 方法相比,NPU 方法的提升是不错的,在 MNIST 数据集上查询 100 次之后的平均性能提升 0.1.URASC 方法的表现更为有效,但提升程度低于 AdaActive 方法.更重要的是:与使用 *NMI* 进行评估的结果相似,提出的 AdaActive 方法随着查询数量的增加,算法效果逐渐变好.因为随着查询的增加,标记和未标记样本之间的关联变得更稳定.就实验效果而言,AdaActive 方法在实验中获得最优的表现.表 1 列出了 10 次实验中的标准偏差的平均聚类精度.实验总体趋势可以看出:随着查询数量的增加,聚类算法性能变得越来越好,从而凸显了提出的主动学习方法具有更显著的性能优势.

Table 1 The F-Measure (Mean+_std) of different methods on the MNIST dataset

表 1 各方法在 MNIST 数据集上的 F-Measure (Mean+_std)值

数据集	方法	不同数目的主动查询								
		30	45	60	75	90	105	120	135	150
MNIST	RAND	0.76±0.10	0.77±0.15	0.77±0.10	0.76±0.10	0.75±0.10	0.75±0.13	0.75±0.11	0.75±0.10	0.75±0.11
	Min-Max	0.81±0.08	0.80±0.10	0.79±0.09	0.79±0.07	0.80±0.09	0.82±0.06	0.84±0.00	0.84±0.00	0.84±0.00
	QUIRE	0.77±0.10	0.78±0.08	0.78±0.06	0.80±0.06	0.79±0.01	0.81±0.10	0.82±0.01	0.83±0.11	0.83±0.02
	NPU	0.81±0.07	0.82±0.05	0.83±0.06	0.83±0.03	0.84±0.03	0.84±0.01	0.85±0.00	0.85±0.00	0.85±0.00
	FASS	0.81±0.09	0.83±0.17	0.86±0.01	0.86±0.03	0.87±0.01	0.87±0.02	0.87±0.03	0.89±0.01	0.89±0.02
	URASC	0.81±0.08	0.83±0.15	0.85±0.01	0.86±0.05	0.86±0.01	0.87±0.03	0.87±0.02	0.88±0.05	0.88±0.03
	ASCENT	0.82±0.10	0.85±0.01	0.87±0.01	0.88±0.03	0.88±0.02	0.89±0.02	0.89±0.01	0.90±0.01	0.91±0.01
	AdaActive	0.82±0.10	0.85±0.02	0.87±0.01	0.88±0.02	0.89±0.01	0.89±0.02	0.90±0.01	0.90±0.01	0.92±0.01

3.4 参数影响

本节研究超参数 β 如何影响提出算法的性能.实验在 MNIST 数据集上进行,且设置与之前的实验相同.在图 5 中,实验展示了 AdaActive 方法中 β 值的不同选择. β 同时也表示候选集的大小,并且权衡了主动学习和子模函数目标的比重.

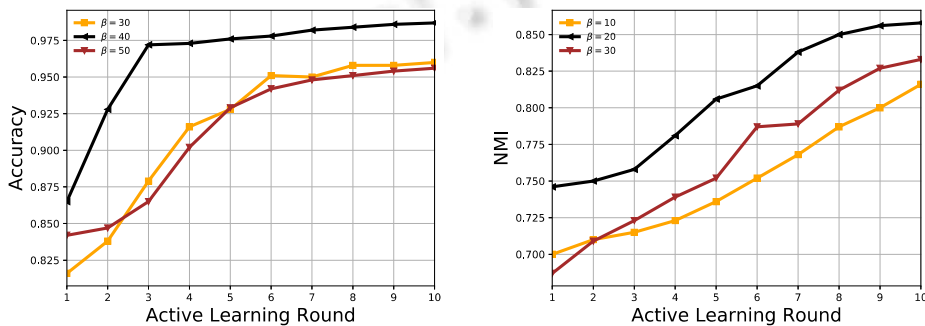


Fig.5 The effect of β on classification and clustering tasks

图 5 参数 β 对分类和聚类任务的影响

通常,在分类任务中,算法对 β 值在 10~30 之间的选择不敏感;在聚类任务中,算法对 β 值在 40~90 的选择也不敏感.在实验中,我们发现选择更少或更多候选集是没有好处的.实验观察到:分类和聚类算法随着 β 值的不同选择而变化,其中,分类的最佳性能为 $\beta=40$,而聚类的最佳性能为 $\beta=20$.该结果突出了候选集大小选择的重要性,并为分类和聚类实验提供了一些指导.

4 相关工作

近年来,学者们已经研究了各种类型的主动学习方法并用于分类和聚类任务中.其中,许多主动分类和聚类算法已经被成功应用到图像分类^[28]、文档聚类^[29]和计量生物学^[30].例如: Xu 等人^[4]利用了未标记样本的聚类信息和分类边界; Donmez 等人^[9]建立在聚类方法^[15]之上,并通过样本的不确定性和密度动态平衡查询选择; Wang 等人^[7]使用批抽样主动学习方法,选择与未标记数据具有区别性和相似分布的样本; Cai 等人^[23]利用模型变化最大化的思想来查询最有影响的样本.此外,还有些方法引入了方差最大化准则^[31,32]和基于损失效率的主动学习方法^[33,34].另外,学者们还提出了一些将多实例或多标签学习与主动学习相结合的方法^[31,35,36],并开发了半监督主动学习模型^[17,37].对于聚类任务,这些方法^[25,38,39]对样本对的约束使用主动学习来进行聚类. Xiong 等人^[27]提出了一种在线框架,用于主动半监督谱聚类.该方法基于模型不确定性降低原理,在聚类进行时选择成对约束. Yu 等人^[40]通过低秩矩阵提出一种有效的三向聚类(three-way clustering)方法,该方法可以提高多视图高维数据的聚类性能. Eriksson 等人^[41]提出一种基于少量样本相似约束对的主动层次聚类算法.此外,主动学习方法还被应用到基于密度的聚类算法^[42]中.然而,一些方法通过单一查询标准选择样本,其并未利用未标记数据的结构^[3-5]或没有考虑未标记数据的分布^[4,6-8],使得模型的学习性能较差;另外,传统的一些方法很难在图像数据中获得较好的效果.

在主动学习方法中,学者们已经开发了很多批抽样的主动学习算法^[28,43].通常,批抽样主动学习方法的关键思想是减少批抽样中所选样本之间的冗余.例如: Hoi 等人^[28]选择一批具有最小 Fisher 信息的样本来减少冗余; Guo 等人^[43]提出了一种判别式的批抽样主动学习方法,该算法将批抽样选择形式化为连续优化问题;在文献^[5,44]中,他们通过 SVM 的超平面选择批样本; Chattopadhyay 等人^[6]通过最小化训练数据和未标记数据之间的数据分布差异来选择样本; Chakraborty 等人^[45]提出一种新的动态批抽样主动学习框架,它通过将批量大小和选择标准集成到单个优化函数中来自适应地确定批量大小.此外,期望模型变化策略^[46]被用于批抽样主动学习方法中.基于深度相似性的方法^[44]和子模函数^[21]也被用于批抽样主动学习方法中.另外,元学习算法(meta-learning)^[47]也被用于学习主动学习算法中,该算法选择最佳的未标记样本集来标记.综上所述,主动学习方法是机器学习领域不可或缺的一部分.

现有批抽样主动学习方法的性能很大程度上依赖于样本之间的相似性度量的准确性,例如预定义函数或差异性衡量.这在一些图像和医学影像应用中具有局限性.此外,噪声标签问题一直是批抽样主动学习算法的障碍.本文提出的方法使用深度神经网络来学习标记和未标记样本特征表示,同时考虑了样本的信息性和代表性.在批模式主动学习中,使用的子模函数确保了选择样本集合具有多样性.而且,提出方法的标签循环机制对噪声标签具有鲁棒性.更重要的是:自适应的方法可以动态地调节样本的不确定性和代表性的权重,使得抽样的样本保持动态的信息度.

5 总结

本文提出了一种自适应的主动学习方法.从深度神经网络生成的特征表示中学习关联和访问概率损失,使提出的方法同时考虑样本不确定性和影响性.批抽样中子模函数的设计,为算法提供了一些指导.自适应参数的优化,使得主动学习算法可以自动平衡样本的信息性和代表性.此外,本文提出了一个框架,将主动学习引入到半监督分类和半监督聚类中.所提出的算法是一种可以嵌入现有方法的通用算法.针对不同任务,即分类和聚类,算法在多种基准数据集上进行了广泛的实验,实验结果证明,提出的方法全面优于现有相关的方法.在未来的工作中,我们将终身强化学习和 AutoML(automatic machine learning)^[48,49]推广到提出的方法中.另外,我

们计划开发一个系统,集成众包学习、主动学习、半监督学习和可视化工具。

6 模型网络结构和参数

6.1 MNIST数据集上的网络结构

完全连接(fully connected)是一个全面连接的层,具有 n 个输出单位.卷积和完全连接层具有指数线性单位(exponential linear units,简称 ELU)激活函数和额外的 L2 权重正则化项,权值为 10^{-4} .最后完全连接的层,用于生成样本表示。

6.2 CIFAR-10数据集上的网络结构

对于 CIFAR-10 数据集的网络结构,本文稍微修改表 2 的网络结构,见表 3。

Table 2 Active supervision architecture on MNIST dataset

表 2 MNIST 数据集上的网络结构和参数

网络层	超参数
输入	28×28 像素图像
卷积层	32 卷积核 3×3,步长为 1
卷积层	32 卷积核 3×3,步长为 1
池化层	最大池化 2×2
卷积层	64 卷积核 3×3,步长为 1
卷积层	64 卷积核 3×3,步长为 1
池化层	最大池化 2×2
卷积层	128 卷积核 3×3,步长为 1
卷积层	128 卷积核 3×3,步长为 1
池化层	最大池化 2×2
全连接	生成样本表示(embedding)

Table 3 Active supervision architecture on CIFAR-10 and variations of MNIST datasets

表 3 CIFAR-10 数据集上的网络结构和参数

网络层	超参数
输入	32×32 像素图像
卷积层	32 卷积核 3×3,步长为 1
卷积层	64 卷积核 3×3,步长为 1
池化层	最大池化 3×3
卷积层	64 卷积核 3×3,步长为 1
卷积层	128 卷积核 3×3,步长为 1
池化层	最大池化 2×2
卷积层	128 卷积核 3×3,步长为 1
卷积层	256 卷积核 3×3,步长为 1
池化层	最大池化 2×2
全连接	生成样本表示(embedding)

6.3 Mean Teacher模型的网络结构

表 4 描述了 Mean Teacher 模型的卷积网络结构.平均批量标准化(mean-only batch normalization)、权重标准化(weight normalization)用于卷积和 softmax 层.Leaky ReLu($\alpha=0.1$)作为每个卷积层的非线性激活函数.分类损失定义为“学生”模型 softmax 输出和独热(one-hot)标签之间的交叉熵;一致性损失定义为“学生”模型和“教师”模型 softmax 输出之间的均方误差.小批量的大小设置为 100.MeanTeacher 采用 Adam 优化器训练网络,学习率设置为 0.003,参数 $\beta_1=0.9$, $\beta_2=0.999$,以及 $\epsilon=10^{-8}$.在 Mean Teacher 模型中,“教师”模型的参数在每次训练步骤之后使用带有 $\alpha=0.999$ 的 EMA(exponential moving average)更新。

Table 4 The architecture of the convolutional network of Mean Teacher model**表 4** Mean Teacher 模型的网络结构和参数

网络层	超参数
输入	28(32)×28(32)像素图像
转化(translation)	随机 $\{\Delta x, \Delta y\} \sim [-2, 2]$
水平反转(horizontal flip)	随机 $p=0.5$
高斯噪声	$\sigma=0.15$
卷积层	128 卷积核 3×3, 填充为 0(same padding)
卷积层	128 卷积核 3×3, 填充为 0(same padding)
卷积层	128 卷积核 3×3, 填充为 0(same padding)
池化层	最大池化 2×2
Dropout	$p=0.5$
卷积层	256 卷积核 3×3, 填充为 0(same padding)
卷积层	256 卷积核 3×3, 填充为 0(same padding)
卷积层	256 卷积核 3×3, 填充为 0(same padding)
池化层	最大池化 2×2
Dropout	$p=0.5$
卷积层	512 卷积核 3×3, “丢弃”(valid padding)
卷积层	256 卷积核 1×1, 填充为 0(same padding)
卷积层	128 卷积核 1×1, 填充为 0(same padding)
池化层	平均池化(6×6→1×1)
softmax	全连接

References:

- [1] Andriluka M, Pishchulin L, Gehler P, Schiele B. 2D human pose estimation: New benchmark and state of the art analysis. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2014. 3686–3693.
- [2] Settles B. Active learning literature survey. Technical Report 1648. University of Wisconsin-Madison, 2009.
- [3] Tong S, Koller D. Support vector machine active learning with applications to text classification. Journal of Machine Learning Research, 2001,2(Nov.):45–66.
- [4] Xu Z, Yu K, Tresp V, Xu X, Wang J. Representative sampling for text classification using support vector machines. In: Proc. of the European Conf. on Information Retrieval. Berlin, Heidelberg: Springer-Verlag, 2003. 393–407.
- [5] Brinker K. Incorporating diversity in active learning with support vector machines. In: Proc. of the Int'l Conf. on Machine Learning. 2003. 59–66.
- [6] Chattopadhyay R, Wang Z, Fan W, Davidson I, Panchanathan S, Ye J. Batch mode active sampling based on marginal probability distribution matching. ACM Trans. on Knowledge Discovery from Data, 2013,7(3):1–25.
- [7] Wang Z, Ye J. Querying discriminative and representative samples for batch mode active learning. ACM Trans. on Knowledge Discovery from Data, 2015,9(3):1–23.
- [8] Zhang Y, Lease M, Wallace BC. Active discriminative text representation learning. In: Proc. of the AAAI Conf. on Artificial Intelligence. 2017. 3386–3392.
- [9] Donmez P, Carbonell JG, Bennett PN. Dual strategy active learning. In: Proc. of the European Conf. on Machine Learning. Berlin, Heidelberg: Springer-Verlag, 2007. 116–127.
- [10] Huang SJ, Jin R, Zhou ZH. Active learning by querying informative and representative examples. In: Proc. of the Advances in Neural Information Processing Systems. 2010. 892–900.
- [11] Du B, Wang Z, Zhang L, Zhang L, Liu W, Shen J, Tao D. Exploring representativeness and informativeness for active learning. IEEE Trans. on Cybernetics, 2017,47(1):14–26.
- [12] Li YC, Wang YL, Yu DJ, Hu P, Zhao RX. ASCENT: Active supervision for Semi-Supervised learning. IEEE Trans. on Knowledge and Data Engineering, 2020,32(5):868–882.
- [13] Hoi SC, Jin R, Zhu J, Lyu MR. Semi-Supervised SVM batch mode active learning with applications to image retrieval. ACM Trans. on Information Systems, 2009,27(3):1–29.
- [14] Yin C, Qian B, Cao S, Li X, Wei J, Zheng Q, Davidson I. Deep similarity-based batch mode active learning with exploration-exploitation. In: Proc. of the IEEE Int'l Conf. on Data Mining. 2017. 575–584.

- [15] Nguyen HT, Smeulders A. Active learning using pre-clustering. In: Proc. of the Int'l Conf. on Machine Learning. 2004. 79–88.
- [16] Urner R, Wulff S, Ben-David S. Plal: Cluster-based active learning. In: Proc. of the Annual Conf. on Learning Theory. 2013. 376–397.
- [17] Calma A, Reitmaier T, Sick B. Semi-Supervised active learning for support vector machines: A novel approach that exploits structure information in data. *Information Sciences*, 2018,456:13–33.
- [18] Haeusser P, Mordvintsev A, Cremers D. Learning by association-a versatile semi-supervised training method for neural networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 172–181.
- [19] Zhou ZH. Disagreement-based semi-supervised learning. *Acta Automatica Sinica*, 2013,39(11):1871–1878 (in Chinese with English abstract).
- [20] Kendall A, Gal Y, Cipolla R. Multi-Task learning using uncertainty to weigh losses for scene geometry and semantics. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2018. 7482–7491.
- [21] Wei K, Iyer R, Bilmes J. Submodularity in data subset selection and active learning. In: Proc. of the Int'l Conf. on Machine Learning. 2015. 1954–1963.
- [22] Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In: Proc. of the Advances in Neural Information Processing Systems. 2017. 1195–1204.
- [23] Cai W, Zhang Y, Zhang Y, Zhou S, Wang W, Chen Z, Ding C. Active learning for classification with maximum model change. *ACM Trans. on Information Systems*, 2017,36(2):1–15.
- [24] Bilenko M, Basu S, Mooney RJ. Integrating constraints and metric learning in semi-supervised clustering. In: Proc. of the Int'l Conf. on Machine Learning. 2004. 11–19.
- [25] Mallapragada PK, Jin R, Jain AK. Active query selection for semi-supervised clustering. In: Proc. of the IEEE Int'l Conf. on Pattern Recognition. 2008. 1–4.
- [26] Xiong S, Azimi J, Fern XZ. Active learning of constraints for semi-supervised clustering. *IEEE Trans. on Knowledge and Data Engineering*, 2014,26(1):43–54.
- [27] Xiong C, Johnson DM, Corso JJ. Active clustering with model-based uncertainty reduction. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2017,39(1):5–17.
- [28] Hoi SC, Jin R, Zhu J, Lyu MR. Batch mode active learning and its application to medical image classification. In: Proc. of the Int'l Conf. on Machine Learning. 2006. 417–424.
- [29] Huang A, Milne D, Frank E, Witten IH. Clustering documents with active learning using Wikipedia. In: Proc. of the Int'l Conf. on Data Mining. 2008. 839–844.
- [30] Kutsuna N, Higaki T, Matsunaga S, Otsuki T, Yamaguchi M, Fujii H, Hasezawa S. Active learning framework with iterative clustering for bioimage classification. *Nature Communications*, 2012,3:1032.
- [31] Du B, Wang Z, Zhang L, Zhang L, Tao D. Robust and discriminative labeling for multi-label active learning based on maximum correntropy criterion. *IEEE Trans. on Image Processing*, 2017,26(4):1694–1707.
- [32] Yang Y, Loog M. A variance maximization criterion for active learning. *Pattern Recognition*, 2018,78:358–370.
- [33] Wang K, Zhang D, Li Y, Zhang R, Lin L. Cost-Effective active learning for deep image classification. *IEEE Trans. on Circuits and Systems for Video Technology*, 2017,27(12):2591–2600.
- [34] Lin L, Wang K, Meng D, Zuo W, Zhang L. Active self-paced learning for cost-effective and progressive face identification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2018,40(1):7–19.
- [35] Melendez J, van Ginneken B, Maduskar P, Philipsen RH, Ayles H, Sánchez CI. On combining multiple-instance learning and active learning for computer-aided detection of tuberculosis. *IEEE Trans. on Medical Imaging*, 2016,35(4):1013–1024.
- [36] Wang Z, Fang X, Tang X, Wu C. Multi-Class active learning by integrating uncertainty and diversity. *IEEE Access*, 2018,22: 794–803.
- [37] Wang Z, Du B, Zhang L, Zhang L, Jia X. A novel semi-supervised active learning algorithm for hyperspectral image classification. *IEEE Trans. on Geoscience and Remote Sensing*, 2017,55(6):3071–3083.
- [38] Basu S, Banerjee A, Mooney RJ. Active semi-supervision for pairwise constrained clustering. In: Proc. of the SIAM Int'l Conf. on Data Mining. 2004. 333–344.

- [39] Van Craenendonck T, Blockeel H. Constraint-Based clustering selection. *Machine Learning*, 2017,106(9-10):1497–1521.
- [40] Yu H, Wang X, Wang G, Zeng X. An active three-way clustering method via low-rank matrices for multi-view data. *Information Science*, 2018,576–582.
- [41] Eriksson B, Dasarathy G, Singh A, Nowak R. Active clustering: Robust and efficient hierarchical clustering using adaptively selected similarities. In: *Proc. of the Int'l Conf. on Artificial Intelligence and Statistics*. 2011. 260–268.
- [42] Mai ST, He X, Hubig N, Plant C, Bohm C. Active density-based clustering. In: *Proc. of the Int'l Conf. on Data Mining*. 2013. 508–517.
- [43] Guo Y, Schuurmans D. Discriminative batch mode active learning. In: *Proc. of the Advances in Neural Information Processing Systems*. 2008. 593–600.
- [44] Schohn G, Cohn D. Less is more: Active learning with support vector machines. In: *Proc. of the Int'l Conf. on Machine Learning*. 2000. 839–846.
- [45] Chakraborty S, Balasubramanian V, Panchanathan S. Adaptive batch mode active learning. *IEEE Trans. on Neural Networks and Learning Systems*, 2015,26(8):1747–1760.
- [46] Cai W, Zhang M, Zhang Y. Batch mode active learning for regression with expected model change. *IEEE Trans. on Neural Networks and Learning Systems*, 2017,28(7):1668–1681.
- [47] Ravi S, Larochelle H. Meta-Learning for batch mode active learning. In: *Proc. of the Int'l Conf. on Learning Representations*. 2018. 137–149.
- [48] Li YF, Wang H, Wei T, Tu WW. Towards automated semi-supervised learning. In: *Proc. of the AAAI Conf. on Artificial Intelligence*. Honolulu, HI, 2019. 4237–4244.
- [49] Huang SJ. 主动学习年度进展. Dalian: 大连理工大学, 2018 (in Chinese).

附中文参考文献:

- [19] 周志华. 基于分歧的半监督学习. *自动化学报*, 2013,39(11):1871–1878.
- [49] 黄圣君. 主动学习年度进展. 大连: 大连理工大学, 2018.



李延超(1990—),男,博士,讲师,主要研究领域为人工智能,大数据管理.



陈志(1978—),男,博士,教授,CCF 专业会员,主要研究领域为软件工程,无线传感网,物联网,数据挖掘.



肖甫(1980—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为传感网,物联网.



李博(1979—),男,高级工程师,主要研究领域为自然语言处理,知识图谱.