

多示例学习下的深度森林架构

任 婕 侯博建 姜 远

(计算机软件新技术国家重点实验室(南京大学) 南京 210023)
(软件新技术与产业化协同创新中心(南京大学) 南京 210023)
(renj@lamda.nju.edu.cn)

Deep Forest for Multiple Instance Learning

Ren Jie, Hou Bojian, and Jiang Yuan

(National Key Laboratory for Novel Software Technology (Nanjing University), Nanjing 210023)
(Collaborative Innovation Center of Novel Software Technology and Industrialization (Nanjing University), Nanjing 210023)

Abstract Multi-instance learning has been applied to various tasks, such as image retrieval, text classification, face recognition, etc. Deep neural network has also been successfully applied to plenty of tasks and problems. MI-Nets are one of the successful applications to multi-instance learning of deep neural network. Although MI-Nets have obtained achievements and the main task they are good at is image task, while on non-image tasks, they show mediocre performance. Over the last two years, deep forest has achieved good performance on non-image tasks and is favored for its less parameters and steady performance compared with deep neural network. Thus it is urgent and necessary to apply deep forest to multi-instance learning. However, due to the limitation of the structure of deep forest, we cannot simply substitute the bag-level forest for each forest of deep forest. Therefore, we need to change the structure of deep forest to achieve our purpose. In this paper, we provide a new structure of deep forest, that is multiple instance deep forest (MIDF). We regard each instance from a bag as a new bag, and thus the distribution output from the middle level can concatenate the original bag to make the cascade structure valid. We can also assure the number of layers of MIDF. Experimental results show that our method has comparable performance with MI-Nets on image task, while on non-image tasks, our method outperforms MI-Nets and other baseline methods.

Key words machine learning; multi-instance learning; deep forest; supervised learning; non-image categorization

摘 要 多示例学习已经广泛地应用到各个领域,如图像检索、文本分类、人脸识别等.而近年来深度神经网络也成功地运用到各个任务和问题上,MI-Nets 是深度神经网络在多示例学习领域一个成功的应用.虽然 MI-Nets 很成功,但其主要在图像相关的任务上表现突出,而在非图像任务比如文本分类任务上的性能并不令人满意.而最近 2 年兴起的深度森林在非图像任务上取得了较好的成绩,并因为其相对于深度神经网络有较少的参数和较稳定的性能而受到青睐.所以用深度森林来提升多示例学习性能

具有可行性.但由于深度森林结构的限制,并不能把组成深度森林的每一个森林都直接替换成包级别的森林,需要修改深度森林的结构来达到目的.提出了一种新的深度森林架构 MIDF.在该架构下,为了使得中间层的输出分布可以和包中的示例拼接成功,拼接时把包里的每个示例都看作是一个包,从而使得级联结构依然有效.另外,还能自动确认深度森林的层数.实验结果表明:该方法在图像任务上的性能与擅长处理图像任务的 MI-Nets 相当;而在文本数据上,该方法取得了比 MI-Nets 和其他基线算法更好的效果.

关键词 机器学习;多示例学习;深度森林;监督学习;非图像分类

中图法分类号 TP18

多示例学习(multiple instance learning, MIL)在药物活性预测调查期间被首次提出^[1].与传统的单示例学习不同,多示例学习的输入是一组标签为正或负的包,而不是一组标签为正或负的示例.在多示例学习中我们不会得到任何示例的标签信息.如今,多示例学习已经广泛地应用于各种领域,如图像分类检索、人脸识别、文本分类、计算机辅助医疗诊断等^[2].在过去的时间内,已经提出了很多有效的 MIL 算法^[3-5].

而近年来,深度神经网络也在各个领域都取得了很好的效果^[6-8].其中 MI-Net with DS 和 MI-Net with RC^[9]是深度神经网络在多示例学习领域的成功应用,它使用了深度学习的技巧:深度监督和剩余连接,在图像领域的数据上取得了优异的成绩.然而在非图像任务上的性能并不像在图像任务上那么优秀.与大部分的深度神经网络模型一样,它也需要人工针对不同的数据集进行结构调整和参数选择.

最近几年,深度森林^[10]进入了人们的视野.这是一种基于非可微模块构建的深度模型,具有比深度神经网络少得多的超参数,同时其模型复杂性可以通过数据相关的方式自动确定.实验表明,其性能对于超参数设置非常稳健,因此在大多数情况下,即使面对来自不同领域的不同数据,也可以通过使用相同的默认设置获得出色的性能.此外由于决策树的特性,深度森林在非图像的领域上也有着不错的效果.

因此在非图像任务上利用深度森林架构来解决多示例学习问题或许是一种更好的选择. Amores^[2]指出包级别的算法通常比示例级的算法效果要好.但由于深度森林结构的限制,组成深度森林的每一个森林并不能直接被替换成包级别的森林,所以并不能把包级别的思想直接套用到深度森林框架上.这就使得我们需要提出一种新的深度森林结构来解决多示例学习问题.

本文提出了一种新的深度森林架构(multiple instance deep forest, MIDF),使用了包级的算法思想.在拼接时我们把包里的每个示例都看做是一个包,从而使得中间层的输出分布可以和包中的示例拼接成功,让级联结构依然有效.此外我们的架构由 2 个级联结构组成,一个使用 k 折交叉验证来帮助自动确定深度森林的层数,另一个根据所求的层数来计算最终的结果.实验结果表明:我们的方法在图像任务上的性能与擅长处理图像任务的 MI-Net with DS 和 MI-Net with RC 相当;而在文本数据上,我们的算法取得了比它们和其他基线算法更好的效果.

1 背景知识

本节主要介绍多示例学习问题的定义,并回顾一个经典的基于包级别的多示例决策树算法 ID3-MI^[11].

多示例学习(MIL)问题最初在药物活性预测中提出.现在多示例学习已经被广泛应用于许多领域,并成为机器学习中的一个重要问题.在大量的多媒体数据中都包含着多示例(multiple instance, MI)结构,例如文本数据中的文章可以被分为多个段落,图像数据可以被划分成多个局部区域,基因表达数据包含着多个基因.多示例学习能够有效地处理这些多示例(MI)数据.

多示例学习(MIL)是一种弱监督学习(weakly supervised learning, WSL).在 MIL 中,每一个训练样本都是一个由许多示例组成的包.对于二分类问题,如果一个包中存在至少一个标签为正的示例,那么我们称这样的包是一个正包.如果一个包中的示例标签都为负,那么我们称这样的包是一个负包.尽管我们可以得到包的标签,但是所有示例的标签都是未知的.多示例学习的目标就是要在这样的假设

下,训练出一个学习器能够很好地预测出包的标签.用形式化的语言来说:假设 \mathcal{X} 是一个示例空间, \mathcal{Y} 是一个类别标签的集合.给定一个数据集 $\{(X_1, y_1), (X_2, y_2), \dots, (X_i, y_i), \dots, (X_N, y_N)\}$, 其中 $X_i = \{\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(i)}, \dots, \mathbf{x}_j^{(i)}, \dots, \mathbf{x}_{n_i}^{(i)}\} \subseteq \mathcal{X}$ 表示一个包, $y_i \in \mathcal{Y} = \{-1, +1\}$ 表示 X_i 的标签,多示例学习的任务就是去寻找一个能够分类未知标签包的函数 $f: 2^{\mathcal{X}} \rightarrow \{-1, +1\}$.

目前已经有许多算法被用来解决 MIL 问题.根据 Amores 的调查, MIL 算法可以分为 3 个种类: 示例空间 (instance space, IS) 范式、包空间 (bag space, BS) 范式和嵌入空间 (embedded space, ES) 范式.简而言之,对于 IS 范式,判别信息被认为是在示例级别上的,而在 BS 范式中,判别信息是在包级别上的.ES 范式中的多示例学习算法明确或隐含地将每个包映射到单个特征向量上,该特征向量总结了关于整个包的相关信息.

ID3-MI 算法就是一种包级多实例学习的算法,它遵循的是一种常见的分而治之的决策树方式.众所周知,决策树算法有着 2 个重要组成部分:1) 如何选择分割树节点的属性;2) 如何使用树进行预测.因此 ID3-MI 算法使用与标准决策树相同的方法来进行预测,也就是使用未知标签的包所落入的叶子节点的标签来确定该包的标签.显然 ID3-MI 算法的关键是如何定义一种衡量标准用来选择决策树的划分点.

对于传统的熵 (entropy),不妨假设数据集 D 包含 p 个正例和 n 个负例,那么依据类别得到的 D 的熵可以写作:

$$Info(D) = -\frac{p}{p+n} \lg\left(\frac{p}{p+n}\right) - \frac{n}{p+n} \lg\left(\frac{n}{p+n}\right). \quad (1)$$

如果选择在属性 V 上划分,并将 D 划分为 $\{D_1, D_2, \dots, D_l\}$,那么此时的熵可以写作:

$$Info(D, V) = \sum_{i=1}^l \frac{|D_i|}{|D|} Info(D_i). \quad (2)$$

此时的信息增益可以写作:

$$Gain(D, V) = Info(D) - Info(D, V) = Info(D) - \sum_{i=1}^l \frac{|D_i|}{|D|} Info(D_i). \quad (3)$$

对于多示例学习,我们只需要考虑如何把计算正负例的个数转换成计算正负包的个数.假设 $\pi(\mathcal{X})$ 和 $\nu(\mathcal{X})$ 分别表示在数据集 \mathcal{X} 中出现的正包和负包

的个数.那么使用 V 来划分数据集 D 所得到的熵可以写作:

$$Info_{multi}(D) = -\frac{\pi(D)}{\pi(D) + \nu(D)} \lg\left(\frac{\pi(D)}{\pi(D) + \nu(D)}\right) - \frac{\nu(D)}{\pi(D) + \nu(D)} \lg\left(\frac{\nu(D)}{\pi(D) + \nu(D)}\right), \quad (4)$$

$$Info_{multi}(D, V) = \sum_{i=1}^l \frac{\pi(D_i) + \nu(D_i)}{\pi(D) + \nu(D)} Info_{multi}(D_i). \quad (5)$$

使用相同思路可以得到多示例学习下的信息增益:

$$Gain_{multi}(D, V) = Info_{multi}(D) - Info_{multi}(D, V) = Info_{multi}(D) - \sum_{i=1}^l \frac{\pi(D_i) + \nu(D_i)}{\pi(D) + \nu(D)} Info_{multi}(D_i). \quad (6)$$

算法 1 通过伪代码的形式给出了包级多实例决策树 ID3-MI 生成决策树过程的详细描述:

算法 1. *BuildID3-MI*($\mathcal{X}, \mathcal{Y}, node, depth$).

- ① if $depth > \max_depth$ 或者 $\forall y_i \in \mathcal{Y}$ 标签相同 then
- ② 将节点 $node$ 置为叶子节点;
- ③ return;
- ④ end if
- ⑤ $featureId, threshold \leftarrow FindBestSplit(\mathcal{X}, y)$;
- ⑥ 将左孩子节点中的包初始化 $\mathcal{X}_{Left} \leftarrow \emptyset$, $\mathcal{Y}_{Left} \leftarrow \emptyset$;
- ⑦ 将右孩子节点中的包初始化 $\mathcal{X}_{Right} \leftarrow \emptyset$, $\mathcal{Y}_{Right} \leftarrow \emptyset$;
- ⑧ for bag $X_i \in \mathcal{X}$ do
- ⑨ if $\exists instance \mathbf{x}_j^{(i)} \in X_i$ 且 $\mathbf{x}_j^{(i), featureID} \leq threshold$ then
- ⑩ $\mathcal{X}_{Left} \leftarrow \mathcal{X}_{Left} \cup X_i$;
- ⑪ $\mathcal{Y}_{Left} \leftarrow \mathcal{Y}_{Left} \cup y_i$;
- ⑫ else
- ⑬ $\mathcal{X}_{Right} \leftarrow \mathcal{X}_{Right} \cup X_i$;
- ⑭ $\mathcal{Y}_{Right} \leftarrow \mathcal{Y}_{Right} \cup y_i$;
- ⑮ end if
- ⑯ end for
- ⑰ if $\mathcal{X}_{Left} \neq \emptyset$ and $\mathcal{X}_{Right} \neq \emptyset$ then
- ⑱ $node.left \leftarrow DecisionNode(node, node.depth + 1)$;
- ⑲ $node.right \leftarrow DecisionNode(node, node.depth + 1)$;

- ⑳ $BuildID3-MI(\mathcal{X}_{Left}, \mathcal{Y}_{Left}, node.left);$
- ㉑ $BuildID3-MI(\mathcal{X}_{Right}, \mathcal{Y}_{Right}, node.right);$
- ㉒ end if

2 多示例深度森林

多示例深度森林同时具有多示例森林和深度森林的优势,但是简单地将两者结合起来并不可行,我们需要更有效的结合方式.

2.1 多示例森林

本节介绍 2 种多示例森林算法 MIRF 和 BLRT^[12],它们分别受到随机森林算法和极根随机树算法的启发.

随机森林(random forests, RFs)最初由 Amit 等人提出^[13]并由 Breiman^[14]扩展.在随机森林中,每一棵树的训练集都是通过对原始训练集的有放回采样(自助采样法 bootstrap)得到的.除此之外,在建树时对每个节点划分的选择也不再是对所有的属性计算一个最优化分,而是在属性集的一个随机子集中选择最优化分属性.如果我们假设所有的样本属性总数都是 d ,那么随机森林通常会挑出 \sqrt{d} 个属性用于选择最优化分.多示例随机森林(MIRF)使用了和随机森林类似的集成方法,它们之间的区别是 MIRF 集成的是包级多示例决策树 ID3-MI 而非传统的决策树.

在极根随机树算法(ExtraTrees)^[15]中,使用了更进一步的随机化步骤.与随机森林一样,极根随机树也是在候选特征的随机子集上来寻找划分,但不同的是极根随机树并非在每个属性上计算最优划分点,而是在这些属性上随机选择划分点.换言之这个值是在属性的经验范围内均匀随机选取的.在所有的随机划分点中,极根随机树会选择其中最优的作

为该节点的划分点.BLRT 受到了极根随机树的影响,并将传统的 ExtraTree 像 ID3-MI 那样推广成了包级多示例 ExtraTree.

2.2 多示例深度森林

多示例深森林(MIDF)受到了深度森林的启发,它也具有级联结构,级联的每一层都是从其前一层接收特征,并将结果发送给下一层作为输入. MIDF 的每一层都是 MIRF 和 BLRT 这 2 种不同多示例森林的集成,使用不同种类的森林集成可以增加 MIDF 算法的多样性.

在深度森林中,如果将示例提供给级联结构的某一层,则这一层中所有的森林都会预测出一个类别分布的估计值,并将这个分布估计值转化为类别向量,如图 1 所示.之后我们将类别向量与原始特征向量拼接起来,作为输入传递给下一层.但在 MIDF 中,得到该层中的每个森林所给出的包级预测后,我们不能简单地将长度为 2 的类别向量与大小为 $n_i \times d$ 的原始特征矩阵直接拼接起来,因为他们在维度上并不相同.

为了解决拼接问题,我们将所有的森林都按包级进行训练(不妨假设每层由 2 个 MIDF 和 2 个 BLRT 组成).之后对于任意包 $X_i \subseteq \mathcal{X}$,将其中的每个示例 $\mathbf{x}_j^{(i)} \in X_i$ 都看作一个包,这样会得到一个新的包集合 \mathcal{X}' ,并把它提供给当前层的森林.森林会给出一个大小为 $2 \times \sum_{i=1}^N n_i$ 的类别向量的集合.此时训练集中的每个示例都可以与一个长度为 2 的类别向量进行拼接.在做好拼接之后,我们重新按照 \mathcal{X} 将所有的示例划分成包,得到新的集合 \mathcal{X}'' ,并把它作为输入传递给下一层. \mathcal{X}'' 中的每个示例 $\mathbf{x}_j^{(i)} \in X_i$ 都有个 $d+8$ 个属性.图 2 的级联结构中详细地给出了具体的拼接过程.

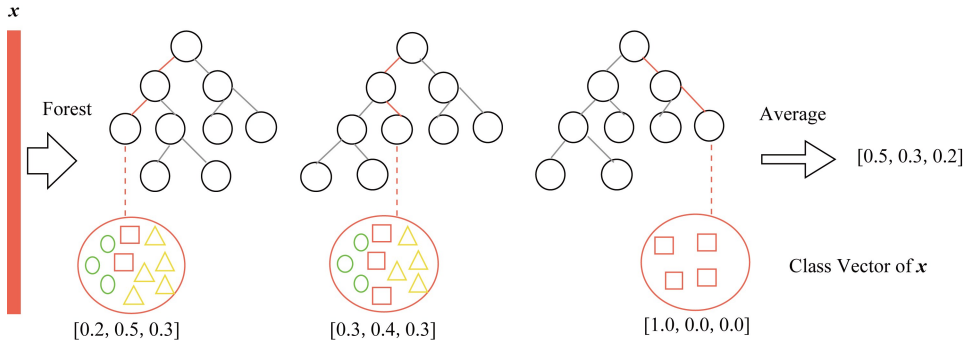


Fig. 1 The illustration of class vector generation

图 1 类向量生成示意图

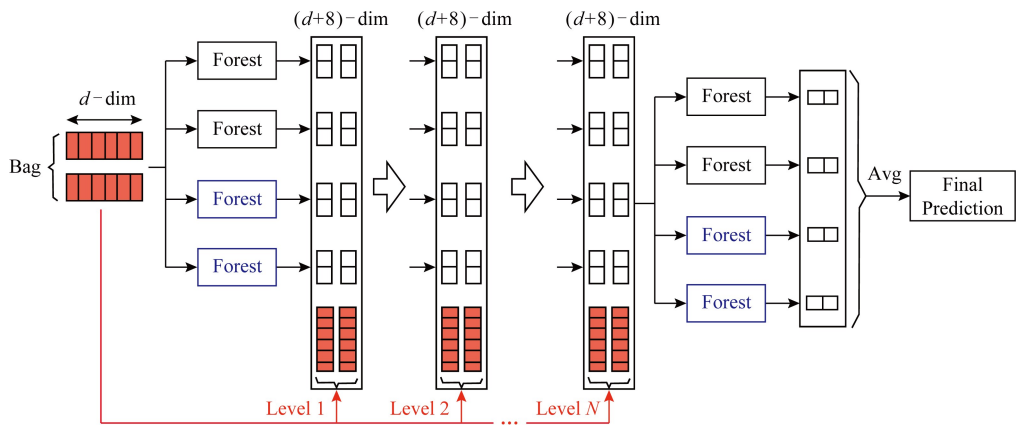


Fig. 2 The illustration of the cascade structure of multiple instance deep forest
图2 多示例深度森林级联示意图

本文算法由 2 个级联结构组成: 1) 用于确定 MIDF 的层数; 2) 用于计算最终预测结果. 第 1 个级联中的每个森林所预测出的类别向量都是通过 k 折交叉验证得到的, 这里我们通常取 $k=3$. 详细地, 每个示例都会被用作训练数据 $k-1$ 次, 产生 $k-1$ 个类别向量, 然后对其取平均值以产生用于传递给下一层的增强特征. 此外, 在扩展一个新层之后, 整个级联的性能将在验证集上进行评估, 如果没有显著的性能增益, 那么训练过程将会终止. 因此, 我们可以通过第 1 个级联结构来自动地确定级联的层数. 第 2 个级联的每一层则使用了完整的训练数据, 其训练的层数由第 1 个级联结构计算给出, 并得到最终的预测结果.

3 实验测试

在本节中, 我们分别在多个多示例学习基准数据集上进行实验, 包括分子活动、图像识别以及文本分类数据集. 并在这些数据集上对比多示例深度森林算法 (MIDF)、Wang 等人提出的多示例神经网络算法 (MI-Net with DS 和 MI-Net with RC) 以及 MILES, miGraph, ID3-MI, MIRF 等多示例学习算法的效果.

3.1 任务及其数据集

我们分别在 3 类任务上进行实验:

1) 药物激活预测. Musk 数据集是 1997 年由 Dietterich 等人发布的关于预测分子是否具有麝香味的数据集. 由于每个分子都可以被描述为它能够折叠成的不同形状 (构象异构体), 我们将每个分子对应为一个包, 而包中的每一个示例则对应该分子

的不同构象. 其中, 不同构象负责分子的不同性质, 也就是它的气味. 在实验里, 如果至少包含一种能够让分子产生麝香味的构象, 我们就将这个分子 (包) 标为正类, 如果没有任何一种构象具有这种性质, 我们就把这个分子标为负类.

2) 自动图像标注. 在这一类任务中有 3 个比较有名的数据集, 分别是 Andrews 等人于 2002 年提出的 Fox, Tiger 和 Elephant 数据集^[3]. 在这类任务中, 我们把图像本身作为包, 图像的不同区域作为包中的示例. 对于每个类别, 如果图像包含该类动物, 我们就把这个图像看作为正包, 如果图像仅包含其他动物 (来自其他类别的动物, 不仅仅是 Fox, Tiger 和 Elephant), 那我们将其看作为负包.

3) 文本分类. 我们从一个名为 20 Newsgroups 的语料库中生成了 20 个文本分类数据集, 对于 20 个新闻类别中的每一个都生成了 50 个正包和 50 个负包. 其中每个正包都包含了从目标类别中随机抽取的 3% 的示例, 并从其他类别中随机均匀采样剩余示例, 每个负包则只包含从其他类别中均匀采样的示例, 并不包含本类别的数据. 在所生成的包中的每个示例均拥有 80 维的特征.

3.2 实验设计

我们设计实验用来验证本文的多示例深度森林算法能够与深度神经网络以及其他算法在多个数据集上效果相当, 并且在某些数据集上会取得更好的效果, 同时我们的算法还能够更加容易地进行超参数的调整. 在实验中, 对于所有的数据集, 我们都使用相同的级联结构 (参数): 每一层由 2 个包级多示例极跟随机树 (BLRT) 和 2 个包级多示例随机森林

(MIRF)构成,其中每个森林均包含 50 棵树.

对于深度神经网络,我们使用 Wang 等人在 2018 年提出的 MI-Net with DS 和 MI-Net with RC 算法,这些算法都需要设定不同的学习率、权重衰减和动量值来获得在不同数据集上的高性能.因此,我们进行实验时在验证集上验证不同参数的效果,选择最佳的参数,并在训练集上重新训练得到最终的预测结果.

对于其余的多示例算法我们也使用了和 MI-Net with DS 和 MI-Net with RC 一样的方法来确定它们在不同数据集上的最优参数,并使用该参数得到最终的预测结果.

3.3 实验结果

我们通过 10 次 10 折交叉验证来比较 ID3-MI、MIRF 以及 MIDF 算法的性能(运行 10 次 10 折交叉验证,每次采取不同的随机数据划分).表 1 和表 2 展示了测试准确率的比较.

表 1 给出了在药物激活预测以及自动图像标注任务上的实验结果,表 2 给出了在文本分类任务上的实验结果,其中用黑体加粗来标注得到的最好结果.

Table 1 Detailed Characteristics of Datasets

表 1 数据集的具体信息

Datasets	Attribute	Bag			Instance
		Positive	Negative	Total	
Musk1	166	47	45	92	476
Musk2	166	39	63	102	6 598
Elephant	230	100	100	200	1 220
Tiger	230	100	100	200	1 391
Fox	230	100	100	200	1 320

Table 2 Average Prediction Accuracy of Different Methods for Drug and Image Datasets

表 2 药物和图片标注数据集上不同算法的平均预测准确率

Algorithms	Musk1	MUSK2	Elephant	Fox	Tiger
MILES	0.84	0.80	0.82	0.61	0.80
miGraph	0.88	0.84	0.80	0.60	0.79
MI-Net with RC	0.88	0.85	0.84	0.60	0.81
MI-Net with DS	0.89	0.84	0.85	0.59	0.81
ID3-MI	0.77	0.71	0.67	0.56	0.69
MIRF	0.87	0.80	0.76	0.54	0.72
MIDF	0.89	0.81	0.84	0.58	0.79

Note: The best results are in bold.

Table 3 Average Prediction Accuracy of Different Methods for Text Categorization Datasets

表 3 文本分类数据集上不同算法的平均预测准确率

Datasets	MILES	miGraph	MI-Net with RC	MI-Net with Ds	ID3-MI	MIRF	MIDF
alt.atheism	0.65	0.74	0.77	0.68	0.67	0.75	0.79
comp.graphics	0.64	0.80	0.79	0.74	0.73	0.81	0.81
comp.os.ms-windows.misc	0.57	0.57	0.64	0.58	0.56	0.60	0.61
comp.sys.ibm.pc.hardware	0.63	0.57	0.58	0.55	0.66	0.74	0.73
comp.sys.mac.hardware	0.59	0.73	0.74	0.68	0.67	0.75	0.79
comp.windows.x	0.74	0.77	0.70	0.56	0.73	0.78	0.77
misc.forsale	0.54	0.65	0.58	0.57	0.53	0.60	0.64
rec.autos	0.61	0.71	0.62	0.66	0.67	0.70	0.72
rec.motorcycles	0.65	0.78	0.84	0.80	0.76	0.79	0.81
rec.sport.baseball	0.71	0.74	0.79	0.70	0.69	0.78	0.82
rec.sport.hockey	0.78	0.79	0.74	0.60	0.71	0.77	0.80
sci.crypt	0.74	0.66	0.70	0.66	0.67	0.74	0.76
sci.electronics.mat	0.70	0.86	0.84	0.78	0.84	0.90	0.91
sci.med	0.70	0.81	0.77	0.69	0.64	0.71	0.71
sci.space	0.61	0.67	0.73	0.67	0.63	0.69	0.74
soc.religion.christian	0.61	0.71	0.71	0.65	0.61	0.74	0.74
talk.politics.guns	0.70	0.62	0.72	0.63	0.65	0.69	0.71
talk.politics.mideast	0.71	0.62	0.64	0.64	0.73	0.79	0.80
talk.politics.misc	0.56	0.61	0.62	0.64	0.58	0.60	0.64
talk.religion.misc	0.60	0.64	0.63	0.62	0.59	0.67	0.68

Note: The best results are in bold.

MIDF 在这些数据集上均展现了较好的效果,尤其是在处理文本分类任务时,我们在 20 个数据集 中的 12 个数据集上取得了最优效果,并在其他 8 个 数据集上同最优结果相当.同时在图像数据上我们 同示例级的神经网络算法效果也相差不大.此外尽 管其他算法在某些数据集上也展现了不错的效果, 但他们往往需要针对不同数据集使用不同的超参 数,这类超参数的选择十分困难且耗时.而我们的算 法只需要使用相同的超参就可以得到很好的结果.

4 结束语

本文提出了一个新的深度森林框架 MIDF 来 解决多示例学习问题.在这种新的框架下,拼接时会 把包中的每个示例都看做是一个包,从而使得中间 层的输出分布可以和包中的示例拼接成功,进而使 得级联结构依然有效.另外,我们的架构由 2 个级联 结构组成,其中一个使用 k 折交叉验证来帮助自动 确定深度森林的层数,另一个根据所求的层数来计 算最终的结果.实验结果表明:我们的方法在图像任 务上的性能并不比擅长处理图像任务的 MI-Nets 差,而在在文本数据上,本文方法取得了比 MI-Nets 更好的效果.

参 考 文 献

[1] Dietterich T G, Lathrop R H, Lozano-Pérez T. Solving the multiple instance problem with axis-parallel rectangles [J]. Artificial Intelligence, 1997, 89(1/2): 31-71

[2] Amores J. Multiple instance classification: Review, taxonomy and comparative study [J]. Artificial Intelligence, 2013, 201: 81-105

[3] Andrews S, Tsochantaridis I, Hofmann T. Support vector machines for multiple-instance learning [C] //Proc of the Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2003: 577-584

[4] Zhou Zhihua, Sun Yuyin, Li Yufeng. Multi-instance learning by treating instances as non-iid samples [C] //Proc of the 26th Annual Int Conf on Machine Learning. New York: ACM, 2009: 1249-1256

[5] Chen Yinxin, Bi Jinbo, Wang J Z. MILES: Multiple-instance learning via embedded instance selection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(12): 1931-1947

[6] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C] // Proc of the Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2012: 1097-1105

[7] Zeng Daojian, Liu Kang, Chen Yubo, et al. Distant supervision for relation extraction via piecewise convolutional neural networks [C] //Proc of the 2015 Conf on Empirical Methods in Natural Language Processing. Stroudsburg, PA: ACL, 2015: 1753-1762

[8] Goodfellow I, Bengio Y, Courville A. Deep Learning [M]. Cambridge, MA: MIT Press, 2016

[9] Wang Xinggang, Yan Yongluan, Tang Peng, et al. Revisiting multiple instance neural networks [J]. Pattern Recognition, 2018, 74(C): 15-24

[10] Zhou Zhihua, Feng Ji. Deep forest: Towards an alternative to deep neural networks [C] //Proc of the 26th Int Joint Conf on Artificial Intelligence. San Francisco, CA: Morgan Kaufmann, 2017: 3553-3559

[11] Chevalere Y, Zucker J D. Solving multiple-instance and multiple-part learning problems with decision trees and rule sets. Application to the mutagenesis problem [C] //Proc of the Conf of the Canadian Society for Computational Studies of Intelligence. Berlin: Springer, 2001: 204-214

[12] Komárek T, Somol P. Multiple instance learning with bag-level randomized trees [C] //Proc of the Joint European Conf on Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2018: 259-272

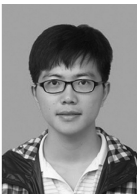
[13] Amit Y, Geman D. Shape quantization and recognition with randomized trees [J]. Neural Computation, 1997, 9(7): 1545-1588

[14] Breiman L. Random forests [J]. Machine Learning, 2001, 45(1): 5-32

[15] Geurts P, Ernst D, Wehenkel L. Extremely randomized trees [J]. Machine Learning, 2006, 63(1): 3-42



Ren Jie, born in 1995. Master in Nanjing University. Her main research interests include artificial intelligence, machine learning and data mining.



Hou Bojian, born in 1992. PhD candidate in Nanjing University. His main research interests include artificial intelligence, machine learning and data mining. (houbj@lamda.nju.edu.cn)



Jiang Yuan, born in 1976. PhD, professor in the Department of Computer Science and Technology at Nanjing University. Her main research interests include artificial intelligence, machine learning, and data mining. (jiangy@lamda.nju.edu.cn)