

解决抽象标签的图像分类的多示例两阶段模型

于全,宋金玉,余晓晗

(陆军工程大学 指挥控制工程学院,江苏 南京 210007)

摘要:经过训练的分类模型可以准确识别出图像中的具体对象,找出“图像中有什么”,但针对诸如“图片描述了什么”的抽象概念标签的图像分类问题研究较少,研究难度也更大。抽象概念标签不属于图像中包含的任何一个具体的对象,而是由许多不同的概念混合在一起,所以直接学习这个抽象标签相当困难。为了解决这类抽象标签的图像分类问题,借助多示例学习方法思路,设计并实现了多示例两阶段模型。该模型由两个阶段构成,第一阶段基于 Yolo 模型修改,实现从图像中快速、精准提取出具体对象,第二阶段构建多层感知机,利用第一阶段模型的结果最终得到图像的分类抽象概念。最后,通过一个具有示范性的实验案例,验证多示例两阶段模型可以利用多示例学习有效解决抽象标签的图像分类问题,展示了多示例两阶段模型的可行性。

关键词:多示例学习;图像分类;抽象标签;Yolo;多层感知机

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2022)06-0068-06

doi:10.3969/j.issn.1673-629X.2022.06.012

Multi-instance Two-stage Model of Solving Image Classification Problem of Abstract Labels

YU Quan, SONG Jin-yu, YU Xiao-han

(School of Command & Control Engineering, Army Engineering University of PLA,
Nanjing 210007, China)

Abstract: The trained classification model can accurately identify the specific objects in the image and find out “what is in the image”. However, there are few researches on the image classification of abstract concept labels such as “what is described in the image”, and the research is more difficult. The abstract concept label does not belong to any one concrete object contained in the image, but is a mixture of many different concepts, so it is quite difficult to learn this abstract label directly. In order to solve the problem of image classification of such abstract labels, with the help of multi-instance learning method, we design and implement a multi-instance two-stage model which consists of two stages. The first-stage is based on the modification of Yolo model to extract the concrete object from the image quickly and accurately, and the second stage is to build multi-layer perceptron and finally obtain the classification abstract concept of images using the results of the first-stage model. Finally, through a demonstrative experimental case, it is verified that the multi-instance two-stage model can effectively solve the image classification problem of abstract labels by using multi-instance learning, and the feasibility of the multi-instance two-stage model is demonstrated.

Key words: multi-instance learning; image classification; abstract labels; Yolo; multilayer perceptron

0 引言

Zhou 等人^[1]提出了一个对抽象标签的图像分类问题,如图 1 中的图像,该图像的标签不再是传统的具体标签,如“巴士”、“人”、“房子”,而是抽象的标签“城镇”。“城镇”不属于图像中包含的任何一个内容,而是对多个具体对象的整合。

这类抽象标签的图像分类问题在实际应用中不少见,如医学领域的 CT 图像癌症识别问题、作战领域的

作战意图识别问题都涉及了抽象的标签。以作战意图识别为例,作战意图具有抽象性,是指挥员根据当前敌方的战场态势图综合推理出的。利用人工智能技术进行作战意图识别,识别模型需要根据战场态势图的输入信息判断出敌方的作战意图(如进攻或防守),传统的技术更多的是理解态势,如敌方兵力和位置,如何在具体的态势理解基础上完成抽象的作战意图识别,相关的研究工作还不多。

收稿日期:2021-07-07

修回日期:2021-11-10

基金项目:国防科技创新项目(19-163-11-LZ-001-003-01)

作者简介:于全(1996-),男,硕士研究生,研究方向为人工智能、数据工程;宋金玉,硕士,教授,CCF 会员(99004M),研究方向为数据工程。

Zhou 等认为可以使用多示例学习方法解决这类问题,并提出了多示例学习算法 MIML^[1],该算法先学习一组子标签(“巴士”、“人”、“房子”),然后在此基础上推导出图像的抽象标签(“城镇”),如图 1 所示。但是在模型验证实验中使用的数据依然是基于老虎、大象、狐狸这类带有具体标签的图像数据,并没有真正将算法应用于像“城镇”这类带有抽象标签的图像数据。

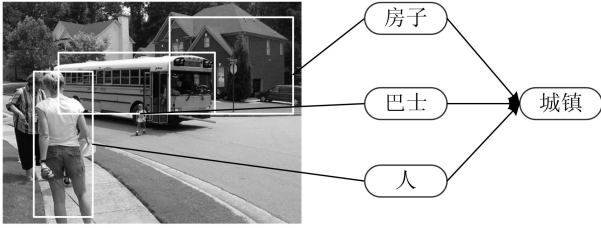


图 1 通过开发子标签推导出“城镇”这个抽象标签

多示例学习 (multi-instance learning, MIL) 最早是由 Dietterich 等人^[2]在研究药物分子活性预测中提出的,是传统监督学习的一种变体。在传统的监督学习中,每个对象都有一个标签来指导算法模型学习,而在多示例学习中训练集由若干个具有标签的包组成,每个包由若干对象组成,这些对象称为示例,只有包有准确的标签信息,示例的标签是未知的。MIL 的目的是从给定标签的包中学习分类器,并对未知标签的包进行分类。多示例学习自提出以来,广受学者重视,开发出了各种应对多示例学习的算法,诸如 Citation-kNN 算法^[3]、SMILE 算法^[4]、MI-NET 算法^[5]、MInD 算法^[6]等,已被应用到药物活性预测^[7]、文本分类^[8-9]、数据挖掘^[10]、图像识别^[11-12]等问题上。

为了实现抽象标签的图像分类的目标,该文借助多示例学习方法思路,基于 Yolo (you only look once)^[13]模型和多层感知机设计并实现了解决带有抽象标签的图像分类问题的算法模型,称为多示例两阶段模型。首先利用 Yolo 模型完成图像中具体对象的提取,然后将提取出的具体对象输入到多层感知机中完成图像抽象概念的预测。通过实验案例表明,多示例两阶段模型能够快速并准确对带有抽象标签的图像进行分类,体现了多示例两阶段模型的可行性。

1 多示例两阶段模型

对于传统的监督学习,每个训练样本都被标记,而多示例学习^[14]是监督学习的一种变体,MIL 的标记对象不再是一个样本,而是一个包 (bag)。每个包是若干个样本组成的集合,这些样本也称为示例 (instances),并且每个包都有一个训练标签(记为正标签或负标签),但包中的示例没有标签。如果包为负标签,则包中所有示例都认为是负示例;如果包为正标

签,则包中至少有一个示例被认为是正标签。MIL 的目的是从给定标签的正包和负包中学习分类器,并将未知标签的包进行分类,划分为正包或负包。

假设有 N 个包 $\{B_1, B_2, \dots, B_N\}$,第 i 个包 B_i 由 a_i 个示例组成 $\{B_{i,1}, B_{i,2}, \dots, B_{i,a_i}\}$,每个示例都是一个 k 维属性值向量。例如第 i 个包的第 j 个示例为 $[B_{i,j,1}, B_{i,j,2}, \dots, B_{i,j,k}]^T$,其中 T 表示向量的转置。 N 个包对应的标签集为 $\{y_i | y_i \in \{0,1\}, i = 1,2, \dots, N\}$,其中 1 代表包为正标签,0 代表包为负标签。

模型框架:

在进行 MIL 的建模时,通常为每类示例构建分类器:

$$\arg \min_{\theta} \sum_{i=1}^N E(\max_j \{f_{ij}(B_{ij}, \theta)\}, y_i) \quad (1)$$

其中, $f_{ij}(\cdot)$ 表示用于示例 B_{ij} 的分类器, θ 表示分类器的参数, $E(\cdot)$ 表示分类器的误差函数, N 表示样本数据个数。上述模型为每一类示例构建了一个分类器 $f_{ij}(\cdot)$,借助取大函数,将每个分类器输出,整合成包的最终输出,即公式中的 $\max_j \{f_{ij}(B_{ij}, \theta)\}$,然后将该输出与对应的标签值 y_i 比较取误差。以往解决图像分类的多示例算法模型都是基于公式(1)构建的,但是使用 \max 函数整合多个示例的输出过于简化,不能很好描述复杂情况。为此,该文考虑构建一个新的分类器 $t(\cdot)$ 替换该 \max 函数,来整合所有示例输出 $f_{i1}(B_{i1}, \theta), f_{i2}(B_{i2}, \theta), \dots, f_{ia_i}(B_{ia_i}, \theta)$ 。其概率模型如下:

$$P(\hat{y} | I) = \sum_Z P(\hat{y}, Z | I) = \sum_Z P(Z | I) P(\hat{y} | Z) \quad (2)$$

其中, I 表示输入变量,即包的输入, \hat{y} 表示预测变量,即包的正负预测, Z 表示中间变量,对应示例分类器的输出 $f_{i1}(B_{i1}, \theta), f_{i2}(B_{i2}, \theta), \dots, f_{ia_i}(B_{ia_i}, \theta)$ 。式(2)表示通过包的输入 I 来预测包的正负,可以分解为两个步骤:首先根据包的输入特征预测该包是否包含某些示例,然后基于该包所包含的示例预测正包还是负包。因此,可如下构造多示例学习的两阶段模型:

$$\arg \min_{\theta_1} \sum_{i=1}^N \sum_{j=1}^{a_i} E_1(f_{ij}(B_{ij}, \theta_1), z_j) \quad (3)$$

$$\arg \min_{\theta_2} \sum_{i=1}^N E_2(t_i(\{f_{ij}(B_{ij}, \theta_1)\}, \theta_2), y_i) \quad (4)$$

其中,公式(3)和公式(4)分别表示第一、二阶段模型, $f_{ij}(\cdot)$ 和 $t(\cdot)$ 分别表示分类器 1 (第一阶段模型) 和分类器 2 (第二阶段模型), θ_1 和 θ_2 分别是分类器 1 和分类器 2 的参数, $E_1(\cdot)$ 和 $E_2(\cdot)$ 分别是分类器 1 和分类器 2 的误差函数。在第一阶段模型中,输入 N 个包到分类器 1 计算每个示例的输出 $f_{ij}(B_{ij}, \theta_1)$,然后根据这些输出与对应的标签值 z_j 的误差,进而优化参数 θ_1 ;

在第二阶段模型中,将分类器 1 的输出 $f_{i1}(B_{i1}, \theta_1)$, $f_{i2}(B_{i2}, \theta_1), \dots, f_{ia_i}(B_{ia_i}, \theta_1)$ 放入分类器 2 计算包 B_i 的最终预测结果,即公式中的 $t_i(\{f_{ij}(B_{ij}, \theta_1)\}, \theta_2)$, 通过对标签值 y_i 优化参数 θ_2 。

多示例两阶段模型框架如图 2 所示,分类器 1 和分类器 2 各自训练、共同预测。基于包中具体包含哪些示例的标签,训练分类器 1 的参数。例如计算机视觉领域已经有比较成熟的方法用于预测一张图片中的多个对象,可以借助已有方法完成分类器 1 的训练。在完成分类器 1 训练的基础上,利用分类器 1 和包的标签对分类器 2 进行训练。成功训练后,可用于多示例样本的预测。

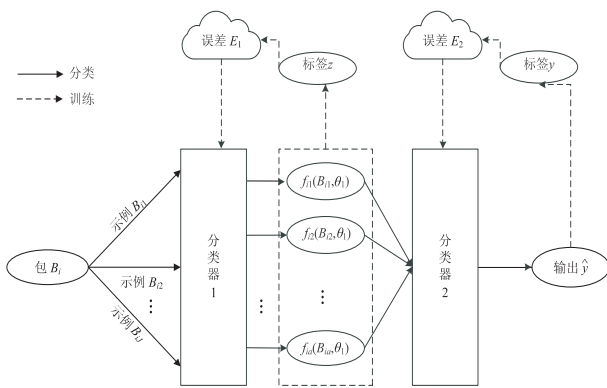


图 2 多示例两阶段模型框架

2 图像分类的多示例两阶段模型

针对图像抽象概念分类问题,该文设计了第一阶段使用 Yolo^[13] 模型,第二阶段使用多层感知机 (MLP) 的多示例两阶段模型,如图 3 所示。

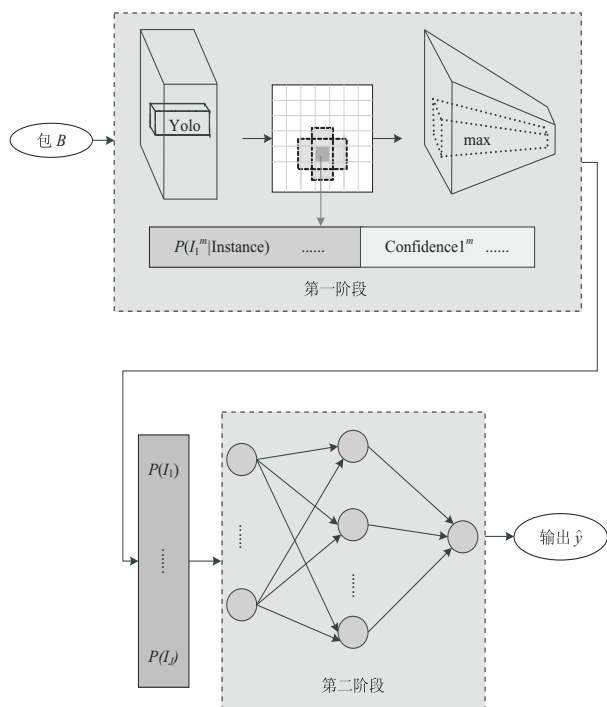


图 3 多示例两阶段模型

Yolo 是基于深度神经网络的对象识别和定位算法,被应用到计算机视觉领域目标检测与分类中^[15-17]。输入一张图像,可以输出图像中所包含的所有对象,以及每个对象在图像中对应的位置,具有很高的实用性。使用 Yolo 作为第一阶段模型,可以很好地实现对图像中各个子对象的识别,并给出子对象的置信度(存在概率)。同时 Yolo 模型很成熟,可靠性高,可借鉴实例多。非常适合作为图像抽象概念分类的第一阶段模型。

图像输入到第一阶段的 Yolo 模型,得到所有示例在该图像中存在的概率 $P(I_1), P(I_2), \dots, P(I_J)$ 。使用 MLP 完成从存在概率 $P(I_1), P(I_2), \dots, P(I_J)$ 到包的类别 $\{\hat{y} | \hat{y} \in \{0,1\}\}$ 的预测。

2.1 第一阶段模型

第一阶段使用 Yolo 模型的作用是对包中示例进行初步分类,得到包中所有示例存在的概率,这需要在 Yolo 模型的基础上对其输出方式进行修改。

Yolo 模型的输入图像被划分为 7×7 的网格,其中一个网格的输出为 30 维向量,那么一张图像通过 Yolo 模型会得到 $7 \times 7 \times 30$ 的张量。其中,一个网格输出的 30 维向量共由三部分组成:一是 20 个对象的概率(20 种不同的对象包括人、鸟、猫、汽车等),所以这 20 个值分别表示该网格位置存在这 20 种对象的概率,记作 $P(O_j | \text{Object})$,表示该网格存在一个对象 Object,且该对象是 O_j 的概率是 $P(O_j | \text{Object})$;二是 2 个边框的位置,每个边框由 4 个数值表示其位置;三是 2 个边框的置信度,一个边框的置信度表示该边框是否包含对象且位置准确的程度,第 s 个边框的置信度记作 Confidence_s 。根据得到的 $7 \times 7 \times 30$ 的输出张量, Yolo 模型利用非极大值抑制算法 (non-maximal suppression, NMS) 输出其中最有可能的那些对象及其所在位置。

下面根据第二阶段模型的需要对第一阶段 Yolo 模型的输出方式进行必要的修改。在 Yolo 模型输出的 $7 \times 7 \times 30$ 张量中,有 49 组针对 J 个示例的条件概率数据:

$$\begin{aligned}
 & [[P(I_1^1 | \text{Instance}), P(I_2^1 | \text{Instance}), \dots, \\
 & \quad P(I_J^1 | \text{Instance})] \\
 & [P(I_1^2 | \text{Instance}), P(I_2^2 | \text{Instance}), \dots, \\
 & \quad P(I_J^2 | \text{Instance})] \\
 & \quad \dots \\
 & [P(I_1^{49} | \text{Instance}), P(I_2^{49} | \text{Instance}), \dots, \\
 & \quad P(I_J^{49} | \text{Instance})]]
 \end{aligned}$$

考虑到第二阶段模型需要输入有关图像的包中是否存在某个示例的数据,从以下两个步骤对 Yolo 模型进行修改:第一步是将示例的条件概率 $P(I_j^m |$

Instance) 转化为示例的概率 $P(I_j^m)$, 第二步是合并 49 组示例的概率, 其中 m 表示图像被划分的第 m 个网格, 具体操作如下。

首先利用公式(5)求得在第 m 个网格中第 j 个示例 I_j^m 存在的概率 $P(I_j^m)$ 。

$$P(I_j^m) = \max_{s=1}^2 \sum (P(I_j^m | \text{Instance}) \times \text{Confidence}_s^m) \quad (5)$$

将第 m 个网格中的第 j 个示例 I_j^m 存在的条件概率 $P(I_j^m | \text{Instance})$ 与 2 个边框置信度 Confidence_s^m 分别相乘得到了 2 个示例 I_j^m 的概率 $P(I_{j,1}^m), P(I_{j,2}^m)$, 由于只需要知道第 m 个网格中存在示例 I_j^m 的概率, 所以利用 \max 函数选择更大的概率表示第 m 个网格中存在示例 I_j^m 的概率 $P(I_j^m)$ 。

基于公式(5)逐个计算在第 m 个网格中 J 个示例存在的概率 $P(I_1^m), P(I_2^m), \dots, P(I_J^m)$ 。这样 49 个网格一共得到 49 组示例存在的概率, 也就是说对于每一个示例都有 49 个概率, 其中每个概率都可以用来表示包中包含该示例的概率, 所以再次利用 \max 函数选择 49 个概率中最大的概率作为包中包含示例的概率 $P(I_j)$, 如式(6)所示。

$$P(I_j) = \max_{m=1}^{49} P(I_j^m) \quad (6)$$

其中, m 表示第 m 个网格。基于上述方法计算包中包含各示例的概率 $P(I_1), P(I_2), \dots, P(I_J)$, 计算过程如图 4 所示。

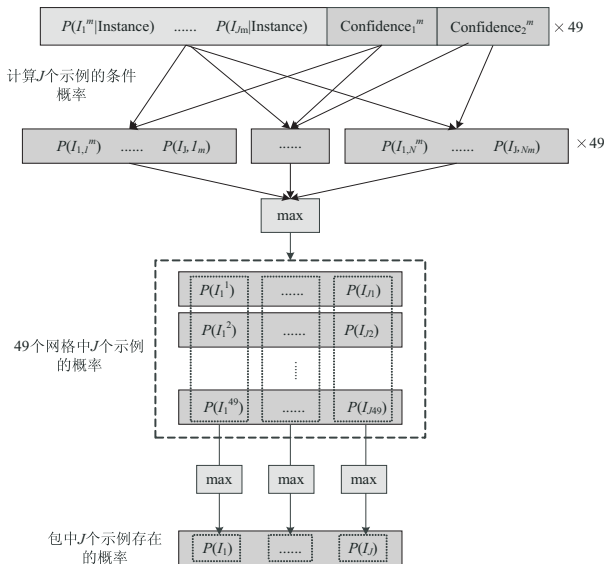


图 4 第一阶段模型的输出

2.2 第二阶段模型

在设计实现第一阶段模型后, 完成对包的初步分类得到包中所有示例在该包中存在的概率 $P(I_1), P(I_2), \dots, P(I_J)$, 本节将设计第二阶段模型基于第一阶段计算出的概率用于包的分类。

第二阶段模型由一个三层结构的多层感知机

(MLP) 构成, 最底层是输入层, 输入层有 J 个神经元, 分别用于输入 J 个概率 $P(I_1), P(I_2), \dots, P(I_J)$ 。中间是单隐藏层, 其中神经元有 K 个, 最后是输出层, 输出层只有一个神经元, 对应两阶段模型最终输出 y , 即图像的抽象标签, 层与层之间是全连接的。给定训练数据 $\{(B_1, y_1), \dots, (B_N, y_N) | B_i \in R^{1 \times J}, y_i \in \{0, 1\}\}$, 对如式(7)的第二阶段模型进行训练:

$$\hat{y} = g_n(\mathbf{W}^{(2)}(g_i(\mathbf{W}^{(1)}\mathbf{B} + b^{(1)})) + b^{(2)}) \quad (7)$$

其中, 矩阵 $\mathbf{W}^{(1)} \in R^{J \times K}$ 和 $\mathbf{W}^{(2)} \in R^{K \times 1}$ 分别表示输入层与隐藏层之间和隐藏层与输出层之间的权重参数, $b^{(1)} \in R^{1 \times K}$ 和 $b^{(2)} \in R^{1 \times 1}$ 分别表示添加到隐藏层和输出层的偏置, $g_i(\cdot)$ 为激活函数。因为第二阶段模型属于一个二分类的分类器, 所以 $g_n(\cdot)$ 使用 logistic 函数 $f(x) = 1/(1 + e^{-x})$ 获得 0 到 1 之间的输出值, 在进行预测时设置阈值 0.5, 将大于或等于 0.5 的输出包的预测结果为 1, 否则为 0。分类损失函数定义为:

$$\text{Loss}(\hat{y}, y) = -y \ln \hat{y} - (1 - y) \ln(1 - \hat{y}) \quad (8)$$

反向传播该损失对权重参数进行训练。在完成第二阶段模型的训练后, 训练好的第二阶段模型即可与第一阶段模型一起用于对包分类, 即将包分别输入到两个阶段的模型进行计算, 最终预测出包的类别 \hat{y} 。

3 实验

本节通过实验说明利用多示例两阶段模型解决抽象标签的图像分类问题的过程。

3.1 实验方案

对于带有抽象标签的图像分类问题, 目前还没有一个标准测试集, 该文基于图像数据集 VOC 2012, 引入抽象标签设计图像分类实验。

首先看一下 VOC 2012 数据集, 该数据集中训练集中有 2 501 张图像用于模型训练, 测试集中有 4 952 张图像用于模型测试, 数据集中图像的标签一共分为 20 类(人、鸟、猫、汽车等), 其中每张图像标签不只有一个。但是这 20 类标签都属于图像中包含的具体标签, 无法直接作为抽象标签用于实验。所以该文根据数据集特点利用专家的知识经验和经验构造出规则库生成图像的抽象标签“室内/室外”, 其中“室内”作为多示例数据的正包, “室外”作为多示例数据的负包, 同时将 20 类标签则作为包中的 20 个示例, 用于实验。数据集标注结果如下: 在 VOC 2012 训练集的 2 501 张图像中, 得到 330 张标签为“室内”的图像, 2 171 张标签为“室外”的图像, 用于多示例两阶段模型的训练, VOC 2012 测试集的 4 952 张图像中, 得到 651 张标签为“室内”的图像, 4 301 张标签为“室外”的图像, 用于多示例两阶段模型的测试。

下面根据得到的标注数据结合多示例两阶段模型

进行图像分类的实验。

3.2 参数设置

实验分为训练和测试两个阶段,在训练阶段中将提出的第一阶段模型和第二阶段模型分开进行训练,其中第一阶段模型使用 Yolo 模型在 VOC 2012 训练集上集训练好的模型^[13],第二阶段模型中多层感知机的超参数通过调试最终设置如下:输入层的单元数为 20、隐藏层的单元数为 4、输出层单元数为 1、迭代次数为 100、学习率为 0.05、输出的阈值为 0.5。在测试阶段,使用训练好的第一阶段模型和第二阶段模型共同测试。

3.3 实验结果

实验结果从两部分给出:一是在实验数据集上的分析;二是通过一个实验案例进行展示。

3.3.1 实验数据分析

在训练集上,利用 python 对多示例两阶段模型进行仿真,并在测试集上对训练好的多示例两阶段模型进行测试,选择损失值和准确率分别对训练过程和测试过程进行分析,其中损失值 Loss 通过公式(8)计算得到,准确率 Acc 通过下式计算得到。

$$Acc = \frac{n(\hat{y}=y)}{N} \quad (9)$$

其中, \hat{y} 为多示例两阶段模型的输出, y 为图像的标签, N 是输入到多示例两阶段模型的图像总数, $n(\hat{y}=y)$ 为将图像输入到多示例两阶段中计算得到的输出与标签相等的图像数。

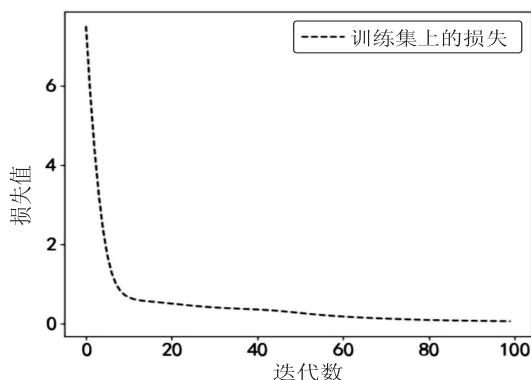


图 5 多示例两阶段模型在训练集上的损失随迭代次数的变化

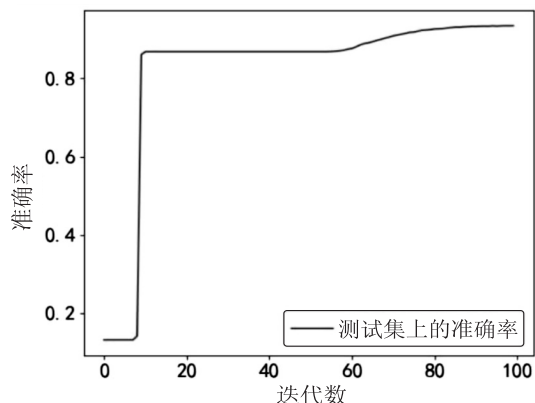


图 6 多示例两阶段模型在测试集上的准确率随迭代次数的变化

图 5 和图 6 分别显示了多示例两阶段模型在训练集上的损失和测试集上的准确率随迭代次数的变化曲线,表 1 显示了多示例两阶段模型迭代结束后在训练集上的损失和测试集上的准确率以及测试集上平均运行时间。

表 1 多示例两阶段模型迭代结束后在训练集上的损失、测试集上的准确率和测试集上的平均运行时间

指标	值
训练集上的损失	0.024
测试集上的准确率	0.935
测试集上的平均运行时间	1.1 s

从图 5 和表 1 可以看出,随着迭代次数的增加,模型不断向着损失函数减小方向不断进化在 80 代左右,损失函数逐渐收敛至最小值 0.024,说明设计实现的两阶段模型能够在训练集上成功进行训练。从图 6 和表 1 中可以看出,在 80 代左右模型在测试集上的准确率收敛至 0.935,且对一张图像分类平均需要 1.1 s,说明提出的多示例两阶段模型能够以较好的预测效果和较快速度解决抽象标签的图像分类问题。

3.3.2 实验案例

为了体现多示例两阶段模型的可行性,给出一张图像的分类案例,如下:首先从测试集中选择一张标签为“室内”的图像。将该图像作为输入,通过训练好的第一阶段模型计算得到输出,如表 2 和图 7 所示。

表 2 第一阶段模型的输出

示例	飞机	自行车	鸟	船	瓶子	公共汽车	汽车	猫	椅子	牛
准确率	0.001	0.014	0.011	0.004	0.004	0.011	0.014	0.019	0.386	0.009
示例	餐桌	狗	马	摩托车	人	盆栽	羊	沙发	火车	电视/显示器
准确率	0.003	0.003	0.002	0.014	0.044	0.006	0.001	0.128	0.001	0.664

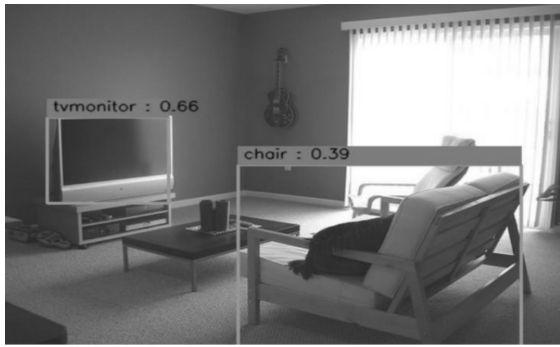


图7 第一阶段模型的输出

表2得到了图像(包)中包含的20个对象(示例)的概率,其中,该图像包含椅子的概率为0.386、包含沙发的概率为0.128、包含电视/显示器的概率为0.664,其余17个对象都不足0.1。然后将这20个概率通过训练好的第二阶段模型计算,输出为1,也就是说将该图像分类为“室内”(正包),即与该图像的标签“室内”一样,对该图像完成正确分类。

4 结束语

基于多示例学习利用 Yolo 模型和多层感知机创新性设计并实现了多示例两阶段模型,解决了抽象标签的图像分类问题。第一阶段模型继承了 Yolo 对图像识别方面快速、准确的优势,引入 max 函数设计多示例学习算法完成对包的初步分类,得到包中所有示例在该包中存在的概率。第二阶段模型则借助成本低、效率高的多层感知机来完成包中示例到包抽象类别的推导。最后,通过将多示例两阶段模型应用于图像分类领域的一个带有抽象标签的实验案例,验证了该模型在解决更复杂的图像分类问题时具有一定的可行性,成功拓展了多示例学习的应用领域。

目前,提出的多示例两阶段模型还存在不足,一是第二阶段中使用的是比较简单的多层感知机对于图像抽象标签的推导不具有可解释性;二是多示例两阶段模型只是在图像数据集上进行验证,还没有应用于实际场景中。下一步的目标是利用模糊树、模糊决策引入专家知识来设计第二阶段模型,使模型具有可解释性,同时将多示例两阶段模型应用于作战意图识别这样的实际应用场景中。

参考文献:

[1] ZHOU Z H, ZHANG M L, HUANG S J, et al. Multi-instance multi-label learning[J]. Artificial Intelligence, 2012, 176(1): 2291-2320.

[2] DIETTERICH T G, LATHROP R H, LOZANO-PEREZ T. Solving the multiple instance problem with axis-parallel rectangles[J]. Artificial Intelligence, 1997, 89(1): 31-71.

[3] WANG J, ZUCKER J D. Solving the multiple-instance problem: a lazy learning approach[C]//Proceedings of the 17th international conference on machine learning. San Francisco: Morgan Kaufmann Publishers, 2000: 1119-1125.

[4] XIAO Y, LIU B, HAO Z, et al. A similarity-based classification framework for multiple-instance learning[J]. IEEE Transactions on Cybernetics, 2017, 44(4): 500-515.

[5] WANG X, YAN Y, PENG T, et al. Revisiting multiple instance neural networks[J]. Pattern Recognition, 2018, 74: 15-24.

[6] CHEPLYGINA V, TAXA D M, LOOG M. Multiple instance learning with bag dissimilarities[J]. Pattern Recognition, 2015, 48(1): 264-275.

[7] HUANG X, CHEN S C, SHYU M L. An open multiple instance learning framework and its application in drug activity prediction problems[C]//Third IEEE symposium on bioinformatics and bioengineering. Bethesda: IEEE, 2003: 53-59.

[8] TIM L, MD B. Textual analysis in finance and accounting: a survey[J]. Journal of Accounting Research, 2016, 54(4): 1187-1230.

[9] ANDREWS S, TSOCHANTARIDIS I, HOFMANN T. Support vector machines for multiple-instance learning[C]//Advances in neural information processing systems. Vancouver and Whistler: MIT, 2003: 561-568.

[10] 钱景辉, 窦立阳, 李荣雨. 一种基于多示例学习的局部离群点检测算法[J]. 信息与控制, 2016, 45(4): 471-478.

[11] 王玉, 申铨京, 陈海鹏. 基于改进的 Fisher 准则的多示例学习视频人脸识别算法[J]. 自动化学报, 2018, 44(12): 2179-2187.

[12] LI W X, VASCONCELOS N. Multiple instance learning for soft bags via top instances[C]//2015 IEEE conference on computer vision and pattern recognition. Boston, MA, USA: IEEE, 2015: 4277-4285.

[13] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016: 779-788.

[14] ZHOU Z H, ZHANG M L. Neural networks for multi-instance learning[C]//Proceedings of the international conference on intelligent information technology. Beijing: [s. n.], 2002: 455-459.

[15] 徐融, 邱晓晖. 一种改进的 YOLO V3 目标检测方法[J]. 计算机技术与发展, 2020, 30(7): 30-33.

[16] PESTANA D, MIRANDA P R, LOPES J D, et al. A full featured configurable accelerator for object detection with YOLO[J]. IEEE Access, 2021, 99: 75864-75877.

[17] XIANG N, CAO Z, WANG Y, et al. A real-time vehicle traffic light detection algorithm based on modified YOLOv3[C]//2021 IEEE 4th international conference on electronics technology. Chengdu: IEEE, 2021: 844-850.