



(12)发明专利

(10)授权公告号 CN 105678340 B

(45)授权公告日 2018.12.25

(21)申请号 201610035975.7

(22)申请日 2016.01.20

(65)同一申请的已公布的文献号

申请公布号 CN 105678340 A

(43)申请公布日 2016.06.15

(73)专利权人 福州大学

地址 350108 福建省福州市闽侯县上街镇

大学城学园路2号福州大学新区

(72)发明人 柯道 周铭柯 杜明智

(74)专利代理机构 福州元创专利商标代理有限公司

公司 35100

代理人 蔡学俊

(51)Int.Cl.

G06K 9/62(2006.01)

(56)对比文件

CN 104166859 A,2014.11.26,

CN 104679863 A,2015.06.03,

CN 105184303 A,2015.12.23,

CN 104156736 A,2014.11.19,

US 2003158830 A1,2003.08.21,

审查员 王佩文

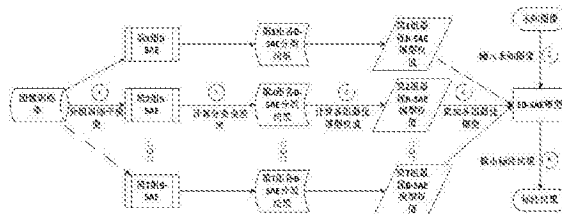
权利要求书3页 说明书7页 附图1页

(54)发明名称

一种基于增强型栈式自动编码器的自动图像标注方法

(57)摘要

本发明涉及一种基于增强型栈式自动编码器的自动图像标注方法:针对深度学习中传统SAE模型难以有效训练有偏数据集的问题,提出一种提升低频标签准确率的平衡栈式自动编码器,较好地改善低频标签的标注效果。然后针对单个B-SAE模型不稳定导致标注效果易随参数改变而发生较大变化的问题,提出一种针对图像标注任务的增强平衡栈式自动编码器,通过分组按序训练、加权累加各组最优B-SAE子模型,取得稳定的标注结果。该方法通过逐层预训练权值并用后向传播算法整体调优,改善了传统浅层模型泛化能力弱、难以收敛到最佳极值点等问题,并在训练过程中加强弱标签样本的训练,提升了整个模型的标注效果,该方法简单灵活,具有较强的实用性。



1. 一种基于增强型栈式自动编码器的自动图像标注方法,其特征在于,按照如下步骤实现:

步骤S1:构建栈式自动编码器模型,在所述栈式自动编码器模型上判别弱标签样本,并添加噪声以增加所述弱标签样本的训练次数,进而构建平衡栈式自动编码器模型;

步骤S2:通过所述平衡栈式自动编码器模型对训练图像分组训练子平衡栈式自动编码器模型,加权累加各组最优子模型得到增强平衡栈式自动编码器模型;

步骤S3:将未知图像输入到所述增强平衡栈式自动编码器模型并输出标注结果;

在所述步骤S1中,还包括如下步骤:

步骤S11:定义编码器 f_{θ} 和解码器 $g_{\theta'}$;所述编码器 f_{θ} 将输入图像 x 转换为隐层表达 h ,解码器 $g_{\theta'}$ 将所述隐层表达 h 重构为与所述输入图像 x 维度一致的向量 x' ;其中, $f_{\theta}(x) = \sigma(W \cdot x + b)$, $\theta = \{W, b\}$, W 为网络权重,满足 $W' = W^T$, b 为偏置向量, $\sigma(x) = \frac{1}{(1 + e^{-x})}$ 为激活函数;

$$g_{\theta'}(h) = \begin{cases} \sigma(W' \cdot h + b'), & x \in [0, 1] \\ W' \cdot h + b' & , x \in \mathbb{R} \end{cases}, \theta' = \{W', b'\};$$

步骤S12:学习一个函数使输出 $x' = g_{\theta'}(f_{\theta}(x))$ 和所述输入图像 x 近似,且定义损失函数为 $L(x, x') = (x - x')^2$,并通过最小化损失函数进行学习: $\theta^*, \theta'^* = \arg \min_{\theta, \theta'} \frac{1}{N} \sum_{i=1}^N L(x_i, g_{\theta'}(f_{\theta}(x_i)))$;

步骤S13:记用于图像标注的SAE模型有 L 层,并用序号 $l \in \{1, \dots, L\}$ 表示;用 h^l 表示第 l 层的输出向量, W^l 和 b^l 表示第 l 层的网络权重和偏置,通过自动编码器对 $\{W^l, b^l\}$, $l \in \{1, \dots, L\}$ 逐层预训练;

步骤S14:执行前馈过程并用后向传播算法调优;所述栈式自动编码器模型的前馈操作表述为: $h^{l+1} = \sigma(W^{l+1}h^l + b^{l+1})$, $l \in \{0, \dots, L-1\}$;所述栈式自动编码器模型的后向传播算法调优表述为: $\theta^* = \arg \min_{\theta} \sum_{i=1}^N L(F_{\theta}(x_i), Y_i)$,其中, $F_{\theta}(x) = \sigma_{\theta_L}(\dots(\sigma_{\theta_1}(x)))$ 是多个自动编码器模型的合成函数,而 θ_l 为参数 $\{W^l, b^l\}$, $l \in \{1, \dots, L\}$,损失函数为 $L(x, y) = (x - y)^2$;

步骤S15:定义约束变量,令向量 $C = (c_1, c_2, \dots, c_M)$, $c_i \in \mathbb{Z}^+$ 表示关键词 y_i 在训练集 P 中出现的次数, $\Pi = \frac{1}{M} \sum_{j=1}^M c_j$ 表示关键词的平均出现次数;向量 $C = (c_1, c_2, \dots, c_M)$ 表示第 i 幅图像 x_i 的每个关键词 Y_i^j , $j \in \{1, 2, \dots, M\}$ 在训练集中出现的次数 $Y_{C,i} = C * Y_i$;从而得到在图像 x_i 中出现次数最低的关键词为 $A_{x_i} = \arg \min_j (Y_{C,i}^j)$;

步骤S16:定义 $\Phi(x)$ 函数,所述栈式自动编码器模型在训练过程中对训练样本进行判断,若输入图像 x 包含低频标签的个数多于 k 个,则对该输入图像 x 添加适当的噪声;定义 $\Gamma(x)$ 函数,对输入图像 x 增加训练强度,若该输入图像 x 所包含标签的出现次数低于预设阈值,则增加训练次数,其中,函数 $\Gamma(x)$ 为:

$$\Gamma(x_i) = \begin{cases} \alpha \cdot \frac{\Pi}{A_{x_i}} = \alpha \cdot \frac{\frac{1}{M} \sum_{j=1}^M c_j}{\arg \min_j (Y_{C_j}^j)}, & A_{x_i} \leq \beta \cdot \Pi, \\ 1, & \text{Others} \end{cases}$$

其中, α 和 β 为常系数, β 用于确定需要加重训练的样本, α 用于控制需要加重训练的样本的训练强度;

函数 $\Phi(x)$ 为:

$$\Phi(x_i) = \begin{cases} \chi \cdot \left(\frac{1}{d} \sum_{j=1}^d x_i^j \right) \cdot \text{Ran}(\cdot), & A_{x_i} \leq \beta \cdot \Pi, \\ x_i, & \text{Others} \end{cases}$$

其中, χ 为常系数, 用于控制噪声添加的强度, d 为图像 x_i 特征的维度, x_i^j 表示图像 x_i 第 j 个维度的值, $\text{Ran}(\cdot)$ 为随机数函数;

步骤 S17: 调整优化等式得到平衡栈式自动编码器模型模型; 将

$$\theta^*, \theta^* = \arg \min_{\theta, \theta^*} \frac{1}{N} \sum_{i=1}^N L(x_i, g_{\theta^*}(f_{\theta}(x_i))) \text{ 调整为 } \theta^*, \theta^* = \arg \min_{\theta, \theta^*} \frac{1}{N} \sum_{i=1}^N \left\{ \frac{1}{\Gamma(x_i)} \sum_{j=1}^{\Gamma(x_i)} L(\Phi(x_i), g_{\theta^*}(f_{\theta}(\Phi(x_i)))) \right\}, \text{ 将}$$

$$\theta^* = \arg \min_{\theta} \sum_{i=1}^N L(F_{\theta}(x_i), Y_i) \text{ 调整为 } \theta^* = \arg \min_{\theta} \sum_{i=1}^N \sum_{j=1}^{\Gamma(x_i)} L(F_{\theta}(\Phi(x_i)), Y_i), \text{ 当模型训练好后, 所述}$$

平衡栈式自动编码器模型的最后一层的输出即为预测图像的关键词的预测分布 D ;

在所述步骤 S2 中, 还包括如下步骤:

步骤 S21: 分组训练子平衡栈式自动编码器模型模型, 将平衡栈式自动编码器模型模型按不同的加噪方式划分不同的组, 每一组内根据不同的隐层神经元个数划分子模型 $B\text{-SAE}_k^t$, t 表示平衡栈式自动编码器模型模型采用第 t 种加噪方式, k 表示第 k 个子 $B\text{-SAE}$ 模型设置的隐层神经元个数;

步骤 S22: 设置初始权值并计算子平衡栈式自动编码器模型模型模型分类误差率, 对训练数据设置权值如下:

$$W = (w_{11}, \dots, w_{1t}, \dots, w_{1N}), \quad w_{ii} = \frac{1}{N}, \quad i = 1, 2, \dots, N,$$

$$\text{计算 } B\text{-SAE}_k^t \text{ 的分类误差率: } e_k^t = \sum_{i=1}^N w_{ii} \cdot \text{Sgn}(B\text{-SAE}_k^t(x_i) \neq Y_i), \text{ 其中, } \text{Sgn}(x) = \begin{cases} 1, & x = \text{true} \\ 0, & x = \text{false} \end{cases},$$

$B\text{-SAE}_k^t(x_i) \neq Y_i$ 表示: 假设图像 x_i 的真实标签集 Y_i 包含 c 个关键词, 并通过模型 $B\text{-SAE}_k^t$ 预测得到标签集 Y_i^* 的个数也为 c 个, 如果 $Y_i = Y_i^*$, 则 $B\text{-SAE}_k^t(x_i) \neq Y_i$ 为 false, 否则为 true;

步骤 S23: 计算平衡栈式自动编码器模型模型权重, 并更新训练数据权值; 根据组内所有子 $B\text{-SAE}_k^t$ 模型的分类误差率, 可以得到该组分类误差率最低的模型 $B\text{-SAE}^t$ 以及对应的分类误差率 e^t , 计算 $B\text{-SAE}^t$ 的权重: $\alpha^t = \frac{1}{2} \log \frac{1-e^t}{e^t}$; 当第 t 组的模型训练完后, 更新训练数据的权值, 以获得下一组模型的权重, 更新训练数据权值的方式如下:

$$W_{t+1} = \{w_{t+1,1}, \dots, w_{t+1,i}, \dots, w_{t+1,N}\}, w_{t+1,i} = \frac{w_i \cdot e^{(-\alpha^t \cdot Y_i \cdot B \cdot SAE^t(x_i))}}{\sum_{i=1}^N w_i \cdot e^{(-\alpha^t \cdot Y_i \cdot B \cdot SAE^t(x_i))}}, i = 1, 2, \dots, N;$$

步骤S24: 加权累加子平衡栈式自动编码器模型模型得到增强平衡栈式自动编码器模型, 当所有组都训练完后, 即得到关键词预测分布: $D = \sum_{t=1}^T \alpha^t \cdot B \cdot SAE^t(x)$ 。

一种基于增强型栈式自动编码器的自动图像标注方法

技术领域

[0001] 本发明涉及模式识别与计算机视觉领域,特别是一种基于增强型栈式自动编码器的自动图像标注方法。

背景技术

[0002] 随着多媒体影像技术快速发展,互联网上图像信息呈爆炸性增长。这些数字图像的应用非常广泛,如商业、新闻媒体、医学、教育等方面。因此,如何帮助用户快速、准确地找到需要的图像成为近年来多媒体研究的热点课题之一。而解决这一课题最重要的技术就是图像检索和自动图像标注技术。

[0003] 自动图像标注是指自动给图像添加若干关键词来表示图像的语义内容。自动图像标注可以利用已标注的图像集,自动学习语义概念空间与视觉特征空间的关系模型,并用此模型标注未知语义的图像。一方面,自动图像标注试图给高层语义特征和底层视觉特征之间建立一座桥梁,因此,它可以一定程度解决大多数基于内容图像检索方法存在的语义鸿沟问题,并且具有较好的客观性。另一方面,自动图像标注可以生成与图像内容相关的文字信息,具有更好的准确性。如果能实现自动图像标注,那么现有的图像检索问题实际上可以转化成较成熟的文本检索问题。因此,图像自动标注技术可以很方便地实现基于关键字的图像检索,符合人们的检索习惯。总的来说,自动图像标注涉及计算机视觉、机器学习、信息检索等多方面的内容,具有很强的研究价值和潜在的商业应用,如图像分类、图像检索、图像理解以及智能图像分析等。

[0004] 根据现有自动图像标注方法的主要实现特点,可以分为两类:基于概率统计的标注方法和基于机器学习的标注方法。基于概率统计的方法虽然可以很方便的扩展到大数据集,但总体标注效果不够理想。基于机器学习的方法,一旦模型训练完毕,就可以进行快速标注,而当前多数分类、回归等学习方法为浅层结构算法,针对复杂分类问题其泛化能力受到一定制约。近年来,深度学习作为机器学习的革新算法,广泛用于目标识别、图像分类、语音识别及其它领域,但鲜有在图像标注问题上得到应用。由于深度学习可以训练深层、复杂的模型,在处理大数据问题上有很大优势。DBN和CNN这两个模型在标签较少、特征简单、特征完整的识别任务中可以取得较好效果,而图像标注问题标签众多、图像特征多样且复杂,且现实图像中还存在大量各类文本、网址、二维码以及图像水印等噪声问题,极大地影响了DBN和CNN的应用效果。而SAE网络,更加注重特征间的近似表达,容易调整模型将复杂的输入表达为理想的输出并应用于特定情形,因此,本专利选用SAE模型解决图像标注问题。

发明内容

[0005] 本发明的目的在于提供一种基于增强型栈式自动编码器的自动图像标注方法,以克服现有技术中存在的缺陷,解决针对多对象多标签的自动图像标注问题。

[0006] 为实现上述目的,本发明的技术方案是:一种基于增强型栈式自动编码器的自动图像标注方法,按照如下步骤实现:

[0007] 步骤S1:构建栈式自动编码器模型,在所述栈式自动编码器模型上判别弱标签样本,并添加噪声以增加所述弱标签样本的训练次数,进而构建平衡栈式自动编码器模型;

[0008] 步骤S2:通过所述平衡栈式自动编码器模型对训练图像分组训练子平衡栈式自动编码器模型,加权累加各组最优子模型得到增强平衡栈式自动编码器模型;

[0009] 步骤S3:将未知图像输入到所述增强平衡栈式自动编码器模型并输出标注结果。

[0010] 在本发明一实施例中,在所述步骤S1中,还包括如下步骤:

[0011] 步骤S11:定义编码器 f_θ 和解码器 $g_{\theta'}$;所述编码器 f_θ 将输入图像 x 转换为隐层表达 h ,解码器 $g_{\theta'}$ 将所述隐层表达 h 重构为与所述输入图像 x 维度一致的向量 x' ;其中, $f_\theta(x) = \sigma(W \cdot x + b)$, $\theta = \{W, b\}$, W 为网络权重,满足 $W' = W^T$, b 为偏置向量, $\sigma(x) = \frac{1}{(1 + e^{-x})}$ 为激活函数; $g_{\theta'}(h) = \begin{cases} \sigma(W' \cdot h + b'), & x \in [0, 1] \\ W' \cdot h + b', & x \in \mathbb{R} \end{cases}$, $\theta' = \{W', b'\}$;

[0012] 步骤S12:学习一个函数使输出 $x' = g_{\theta'}(f_\theta(x))$ 和所述输入图像 x 近似,且定义损失函数为 $L(x, x') = (x - x')^2$,并通过最小化损失函数进行学习:

$$\theta^*, \theta'^* = \arg \min_{\theta, \theta'} \frac{1}{N} \sum_{i=1}^N L(x_i, g_{\theta'}(f_\theta(x_i)));$$

[0013] 步骤S13:记用于图像标注的SAE模型有 L 层,并用序号 $l \in \{1, \dots, L\}$ 表示;用 h^l 表示第 l 层的输出向量, W^l 和 b^l 表示第 l 层的网络权重和偏置,通过自动编码器对 $\{W^l, b^l\}$, $l \in \{1, \dots, L\}$ 逐层预训练;

[0014] 步骤S14:执行前馈过程并用后向传播算法调优;所述栈式自动编码器模型的前馈操作表述为: $h^{l+1} = \sigma(W^{l+1}h^l + b^{l+1})$, $l \in \{0, \dots, L-1\}$;所述栈式自动编码器模型的后向传播

算法调优表述为: $\theta^* = \arg \min_{\theta} \sum_{i=1}^N L(F_\theta(x_i), Y_i)$,其中, $F_\theta(x) = \sigma_{\theta_L}(\dots(\sigma_{\theta_1}(x)))$ 是多个自动编码器模型的合成函数,而 θ_l 为参数 $\{W^l, b^l\}$, $l \in \{1, \dots, L\}$,损失函数为 $L(x, y) = (x - y)^2$;

[0015] 步骤S15:定义约束变量,令向量 $C = (c_1, c_2, \dots, c_M)$, $c_j \in \mathbb{Z}^+$ 表示关键词 y_i 在训练

集 P 中出现的次数, $\bar{C} = \frac{1}{M} \sum_{j=1}^M c_j$ 表示关键词的平均出现次数;向量 $C = (c_1, c_2, \dots, c_M)$ 表示

第 i 幅图像 x_i 的每个关键词 Y_i^j , $j \in \{1, 2, \dots, M\}$ 在训练集中出现的次数 $Y_{C,i} = C * Y_i$;从而得到在图像 x_i 中出现次数最低的关键词为 $A_{x_i} = \arg \min_j (Y_{C,i}^j)$;

[0016] 步骤S16:定义 $\Phi(x)$ 函数,所述栈式自动编码器模型在训练过程中对训练样本进行判断,若输入图像 x 包含低频标签的个数多于 k 个,则对该输入图像 x 添加适当的噪声;定义 $\Gamma(x)$ 函数,对输入图像 x 增加训练强度,若该输入图像 x 所包含标签的出现次数低于预设

阈值(一般取 $\bar{C} = \frac{1}{M} \sum_{j=1}^M c_j$),则增加训练次数,其中,函数 $\Gamma(x)$ 为:

$$[0017] \quad \Gamma(x_i) = \begin{cases} \alpha \cdot \frac{\Pi}{A_{x_i}} = \alpha \cdot \frac{\frac{1}{M} \sum_{j=1}^M c_j}{\arg \min_j (Y_{C,i}^j)}, & A_{x_i} \leq \beta \cdot \Pi, \\ 1, & \text{Others} \end{cases}$$

[0018] 其中, α 和 β 为常系数, β 用于确定需要加重训练的样本, α 用于控制需要加重训练的样本的训练强度;

[0019] 函数 $\Phi(x)$ 为:

$$[0020] \quad \Phi(x_i) = \begin{cases} x \cdot \left(\frac{1}{d} \sum_{j=1}^d x_i^j \right) \cdot \text{Ran}(\cdot), & A_{x_i} \leq \beta \cdot \Pi, \\ x_i, & \text{Others} \end{cases}$$

[0021] 其中, x 为常系数,用于控制噪声添加的强度, d 为图像 x_i 特征的维度, x_i^j 表示图像 x_i 第 j 个维度的值, $\text{Ran}(\cdot)$ 为随机数函数;

[0022] 步骤S17:调整优化等式得到平衡栈式自动编码器模型模型;将

$$\theta^*, \theta^* = \arg \min_{\theta, \theta^*} \frac{1}{N} \sum_{i=1}^N L(x_i, g_{\theta}(f_{\theta}(x_i))) \text{ 调整为 } \theta^*, \theta^* = \arg \min_{\theta, \theta^*} \frac{1}{N} \sum_{i=1}^N \left\{ \frac{1}{\Gamma(x_i)} \sum_{j=1}^{\Gamma(x_i)} L(\Phi(x_i), g_{\theta}(f_{\theta}(\Phi(x_i)))) \right\},$$

$$\theta^* = \arg \min_{\theta} \sum_{i=1}^N L(F_{\theta}(x_i), Y_i) \text{ 调整为 } \theta^* = \arg \min_{\theta} \sum_{i=1}^N \sum_{j=1}^{\Gamma(x_i)} L(F_{\theta}(\Phi(x_i)), Y_i),$$

当模型训练好后,所述平衡栈式自动编码器模型的最后一层的输出即为预测图像的关键词的预测分布D。

[0023] 在本发明一实施例中,在所述步骤S2中,还包括如下步骤:

[0024] 步骤S21:分组训练子平衡栈式自动编码器模型模型,将平衡栈式自动编码器模型模型按不同的加噪方式划分不同的组,每一组内根据不同的隐层神经元个数划分子模型 $B\text{-}SAE_k^t$, t 表示平衡栈式自动编码器模型模型采用第 t 种加噪方式, k 表示第 k 个子 $B\text{-}SAE$ 模型设置的隐层神经元个数;

[0025] 步骤S22:设置初始权值并计算子平衡栈式自动编码器模型模型模型分类误差率,对训练数据设置权值如下:

$$[0026] \quad W = (w_{11}, \dots, w_{1i}, \dots, w_{1N}), w_{1i} = \frac{1}{N}, i = 1, 2, \dots, N,$$

[0027] 计算 $B\text{-}SAE_k^t$ 的分类误差率: $e_k^t = \sum_{i=1}^N w_{1i} \cdot \text{Sgn}(B\text{-}SAE_k^t(x_i) \neq Y_i)$, 其中,

$$\text{Sgn}(x) = \begin{cases} 1, & x = \text{true} \\ 0, & x = \text{false} \end{cases}, B\text{-}SAE_k^t(x_i) \neq Y_i \text{ 表示: 假设图像 } x_i \text{ 的真实标签集 } Y_i \text{ 包含 } c \text{ 个关键词,}$$

并通过模型 $B\text{-}SAE_k^t$ 预测得到标签集 Y_i^* 的个数也为 c 个,如果 $Y_i = Y_i^*$,则 $B\text{-}SAE_k^t(x_i) \neq Y_i$ 为false,否则为true;

[0028] 步骤S23:计算平衡栈式自动编码器模型模型权重,并更新训练数据权值;根据组内所有子 $B\text{-}SAE_k^t$ 模型模型分类误差率,可以得到该组分类误差率最低的模型 $B\text{-}SAE^t$ 以及对应

的分类误差率 e^t , 计算B-SAE^t的权重: $\alpha^t = \frac{1}{2} \log \frac{1-e^t}{e^t}$; 当第t组的模型训练完后, 更新训练数据的权值, 以获得下一组模型的权重, 更新训练数据权值的方式如下:

$$[0029] \quad W_{t+1} = \{w_{t+1,1}, \dots, w_{t+1,i}, \dots, w_{t+1,N}\}, \quad w_{t+1,i} = \frac{w_i \cdot e^{(-\alpha^t \cdot Y_i \cdot B \cdot SAE^t(x_i))}}{\sum_{i=1}^N w_i \cdot e^{(-\alpha^t \cdot Y_i \cdot B \cdot SAE^t(x_i))}}, i = 1, 2, \dots, N;$$

[0030] 步骤S24: 加权累加子平衡栈式自动编码器模型得到增强平衡栈式自动编码器模型, 当所有组都训练完后, 即得到关键词预测分布: $D = \sum_{i=1}^T \alpha^i \cdot B \cdot SAE^i(x)$ 。

[0031] 相较于现有技术, 本发明具有以下有益效果: 本发明所提出的一种基于增强型栈式自动编码器的自动图像标注方法, 利用SAE深度神经网络强大的特征表达能力, 基于对自动图像标注、多标签分类和栈式自动编码器的认识, 提出针对图像数据集标签不平衡、难以有效训练大图像数据等问题的增强型栈式自动编码器的自动图像标注方法, 最后得到一种深层的、复杂的自动图像标注模型, 特别是一种基于增强型栈式自动编码器的自动图像标注方法。该方法简单, 实现灵活, 实用性较强。

附图说明

[0032] 图1为本发明中基于增强型栈式自动编码器的自动图像标注方法的流程图。

具体实施方式

[0033] 下面结合附图, 对本发明的技术方案进行具体说明。

[0034] 本发明提出了一种基于增强型栈式自动编码器的自动图像标注方法, 首先针对深度学习传统SAE (Stacked Auto-Encoder, SAE) 模型难以有效训练有偏数据集的问题, 提出一种提升低频标签准确率的平衡栈式自动编码器 (Balance Stacked Auto-Encoder, B-SAE), 较好地改善低频标签的标注效果。然后针对单个B-SAE模型不稳定 (模型复杂、参数较多) 导致标注效果易随参数改变而发生较大变化的问题, 提出一种针对图像标注任务的增强平衡栈式自动编码器 (Enhanced Balance Stacked Auto-Encoder, EB-SAE), 通过分组按序训练、加权累加各组最优B-SAE子模型, 取得稳定的标注结果。具体步骤如下:

[0035] S1: 先构建SAE模型, 接着在SAE模型上判别弱标签样本并添加噪声增加弱标签样本的训练次数, 以此构建B-SAE模型;

[0036] S2: 利用步骤S1得到B-SAE模型对训练图像分组训练子B-SAE模型, 加权累加各组最优子模型得到EB-SAE模型, 如图1所示;

[0037] S3: 将未知图像输入到步骤S2得到的EB-SAE模型并输出标注结果。

[0038] 进一步的, 在本实施例中, 在步骤S1中按照以下步骤实现构建B-SAE模型:

[0039] 步骤S11: 定义编码器 f_θ 和解码器 $g_{\theta'}$, 编码器 f_θ 将输入图像 x 转换为隐层表达 h , 解码器 $g_{\theta'}$ 将 h 重构为和 x 维度一致的向量 x' 。 $f_\theta(x) = \sigma(W \cdot x + b)$, 其中, $\theta = \{W, b\}$, W 为网络权重, 满足 $W' = W^T$, b 为偏置向量, $\sigma(x) = \frac{1}{(1 + e^{-x})}$ 为激活函数。

$$g_{\theta'}(h) = \begin{cases} \sigma(W' \cdot h + b'), & x \in [0,1] \\ W' \cdot h + b' & , x \in \mathbb{R} \end{cases}, \text{其中}, \theta' = \{W', b'\}.$$

[0040] 步骤S12:学习一个函数使输出 $x' = g_{\theta'}(f_{\theta}(x))$ 和 x 近似,定义损失函数为 $L(x, x') = (x - x')^2$,则该模型可通过最小化损失函数进行学习:

$$\theta^*, \theta'^* = \arg \min_{\theta, \theta'} \frac{1}{N} \sum_{i=1}^N L(x_i, g_{\theta'}(f_{\theta}(x_i))).$$

[0041] 步骤S13:执行前馈过程并用后向传播算法调优,假设用于图像标注的SAE模型有L层,用序号 $l \in \{1, \dots, L\}$ 表示.用 h^l 表示第l层的输出向量 ($h^0 = x$ 表示输入, h^L 表示输出). W^l 和 b^l 表示第l层的网络权重和偏置.根据前面所述, $\{W^l, b^l\}, l \in \{1, \dots, L\}$ 使用AE逐层预训练.SAE的前馈操作可以表述为: $h^{l+1} = \sigma(W^{l+1}h^l + b^{l+1}), l \in \{0, \dots, L-1\}$,整个模型用后向传

播算法调优: $\theta^* = \arg \min_{\theta} \sum_{i=1}^N L(F_{\theta}(x_i), Y_i)$, 其中, $F_{\theta}(x) = \sigma_{\theta_L}(\dots(\sigma_{\theta_1}(x)))$ 是多个AE模型的合成函数,而 θ_l 为参数 $\{W^l, b^l\}, l \in \{1, \dots, L\}$,损失函数定义为 $L(x, y) = (x - y)^2$.

[0042] 步骤S14:定义约束变量,令向量 $C = (c_1, c_2, \dots, c_M), c_j \in \mathbb{Z}^+$ 表示关键词 y_j 在训练集

P中出现的次数, $\Pi = \frac{1}{M} \sum_{j=1}^M c_j$ 表示关键词的平均出现次数.这样,我们可以得到一个向量,

表示第i幅图像 x_i 的每个关键词 $Y_i^j, j \in \{1, 2, \dots, M\}$ 在训练集中出现的次数 $Y_{C,i} = C * Y_i$ (*表示两个向量对应点相乘得到一个新向量).从而得到在图像 x_i 中出现次数最低的关键词为 $A_{x_i} = \arg \min_j (Y_{C,i}^j)$.

[0043] 步骤S15:定义 $\Phi(x)$ 函数,让模型在训练过程中对训练样本进行判断,若样本 x ,也即输入图像 x ,包含低频标签的个数多于 k 个,则对该样本添加适当的噪声.定义 $\Gamma(x)$ 函数,对样本 x 增加训练强度,若该样本所包含标签的出现次数低于一定阈值,则增加它的训练次

数,在本实施例中,该阈值一般取 $\Pi = \frac{1}{M} \sum_{j=1}^M c_j$.

$$[0044] \quad F(x_i) = \begin{cases} \alpha \cdot \frac{\Pi}{A_{x_i}} = \alpha \cdot \frac{\frac{1}{M} \sum_{j=1}^M c_j}{\arg \min_j (Y_{C,i}^j)}, & A_{x_i} \leq \beta \cdot \Pi, \\ 1 & , Others \end{cases}$$

[0045] 其中, α 和 β 为常系数, β 用于确定哪些样本需要加重训练, α 用于控制需要加重训练的样本的训练强度.

$$[0046] \quad \Phi(x_i) = \begin{cases} x \cdot \left(\frac{1}{d} \sum_{j=1}^d x_i^j \right) \cdot \text{Ran}(\cdot), & A_{x_i} \leq \beta \cdot \Pi, \\ x_i & , Others \end{cases}$$

[0047] 其中, x 为常系数,用于控制噪声添加的强度, d 为图像 x_i 特征的维度, x_i^j 表示图像

x_i 第j个维度的值, $\text{Ran}(\cdot)$ 为随机数函数, 比如, $\text{Ran}(\cdot)$ 可取服从(0, 1) 高斯分布的随机函数或取值为0到1的均匀分布的随机函数。

[0048] 步骤S16: 调整优化等式得到B-SAE模型, $\theta^*, \theta^* = \arg \min_{\theta, \theta^*} \frac{1}{N} \sum_{i=1}^N L(x_i, g_{\theta^*}(f_{\theta}(x_i)))$ 调整为

$$\theta^*, \theta^* = \arg \min_{\theta, \theta^*} \frac{1}{N} \sum_{i=1}^N \left\{ \frac{1}{\Gamma(x_i)} \sum_{j=1}^{\Gamma(x_i)} L(\Phi(x_i), g_{\theta^*}(f_{\theta}(\Phi(x_i)))) \right\}, \theta^* = \arg \min_{\theta} \sum_{i=1}^N L(F_{\theta}(x_i), Y_i) \text{ 调整为}$$

$$\theta^* = \arg \min_{\theta} \sum_{i=1}^N \sum_{j=1}^{\Gamma(x_i)} L(F_{\theta}(\Phi(x_i)), Y_i), \text{ 当模型训练好后, B-SAE的最后一层的输出即为预测图}$$

像的关键词的预测分布D。

[0049] 进一步的, 在本实施例中, 步骤S2中按照以下步骤实现训练EB-SAE模型:

[0050] 步骤S21: 分组训练子B-SAE模型, B-SAE模型按不同的加噪方式划分不同的组, 每一组内根据不同的隐层神经元个数划分子模型 $B-SAE_k^t$, t表示模型B-SAE采用第t种加噪方式, k表示第k个子B-SAE模型设置的隐层神经元个数。

[0051] 步骤S22: 设置初始权值并计算子B-SAE模型分类误差率, 对训练数据设置权值如下:

$$[0052] \quad W = (w_{11}, \dots, w_{1i}, \dots, w_{1N}), w_{1i} = \frac{1}{N}, i = 1, 2, \dots, N,$$

[0053] 这样, $B-SAE_k^t$ 的分类误差率可以这样计算: $e_k^t = \sum_{i=1}^N w_{1i} \cdot \text{Sgn}(B-SAE_k^t(x_i) \neq Y_i)$, 其

中, $\text{Sgn}(x) = \begin{cases} 1, & x = \text{true} \\ 0, & x = \text{false} \end{cases}$, $B-SAE_k^t(x_i) \neq Y_i$ 表示的意思为, 假设图像 x_i 的真实标签集 Y_i 包含

c个关键词, 并通过模型 $B-SAE_k^t$ 预测得到标签集 Y_i^* 的个数也为c个, 如果 $Y_i = Y_i^*$, 则 $B-SAE_k^t(x_i) \neq Y_i$ 为false, 否则为true。

[0054] 步骤S23: 计算B-SAE模型权重并更新训练数据权值, 根据组内所有子 $B-SAE_k^t$ 模型 的分类误差率, 可以得到该组分类误差率最低的模型 $B-SAE^t$ 和对应的分类误差率 e^t , $B-SAE^t$ 的权重可这样计算: $\alpha^t = \frac{1}{2} \log \frac{1-e^t}{e^t}$ 。当第t组的模型训练完后, 需要更新训练数据的权值, 以便更好的获得下一组模型的权重, 更新训练数据权值的方式如下:

$$[0055] \quad W_{t+1} = \{w_{t+1,1}, \dots, w_{t+1,i}, \dots, w_{t+1,N}\}, w_{t+1,i} = \frac{w_{1i} \cdot e^{(-\alpha^t \cdot Y_i \cdot B-SAE^t(x_i))}}{\sum_{i=1}^N w_{1i} \cdot e^{(-\alpha^t \cdot Y_i \cdot B-SAE^t(x_i))}}, i = 1, 2, \dots, N,$$

[0056] 步骤S24: 加权累加子B-SAE模型得到EB-SAE模型, 当所有组都训练完后, 就可以得到关键词预测分布: $D = \sum_{t=1}^T \alpha^t \cdot B-SAE^t(x)$ 。

[0057] 以上是本发明的较佳实施例, 凡依本发明技术方案所作的改变, 所产生的功能作

用未超出本发明技术方案的范围时,均属于本发明的保护范围。

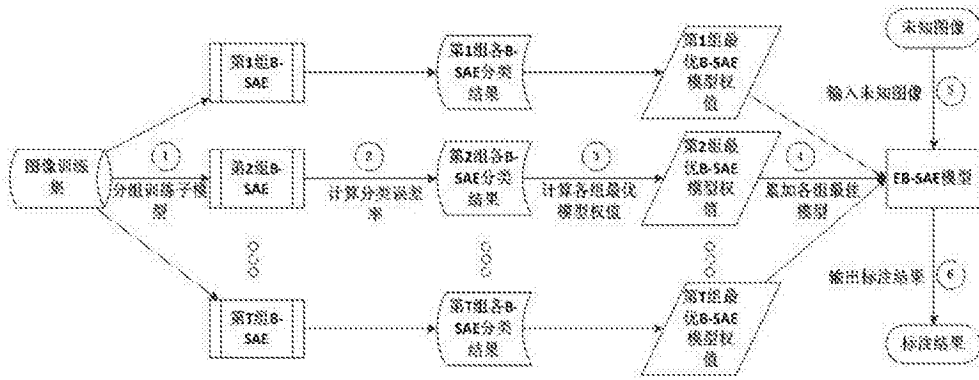


图1