



(12) 发明专利

(10) 授权公告号 CN 113077051 B

(45) 授权公告日 2024.01.26

(21) 申请号 202110402004.2

G06N 3/0455 (2023.01)

(22) 申请日 2021.04.14

G06N 3/08 (2023.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 113077051 A

(56) 对比文件

CN 109993220 A, 2019.07.09

CN 110472665 A, 2019.11.19

CN 110489545 A, 2019.11.22

CN 111144448 A, 2020.05.12

CN 111966831 A, 2020.11.20

CN 112115826 A, 2020.12.22

DE 102009009904 A1, 2009.10.15

US 2010124378 A1, 2010.05.20

(43) 申请公布日 2021.07.06

(73) 专利权人 南京零洞数智科技有限公司

地址 210004 江苏省南京市建邺区泰山路

151号新丽华中心大厦9层01室

(72) 发明人 黄深能 赵茜 利啟东 佟博

高玮 叶凯亮 胡盼盼

Boyan Zhou et al. BBN: Bilateral-Branch Network With Cumulative Learning for Long-Tailed Visual Recognition. 《2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition》. 2020, 第9716-9725页.

(74) 专利代理机构 广州嘉权专利商标事务所有

限公司 44205

专利代理师 陆志君

审查员 刘洋

(51) Int. Cl.

G06F 18/214 (2023.01)

G06F 18/241 (2023.01)

G06F 18/25 (2023.01)

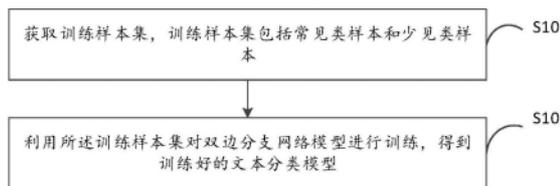
权利要求书2页 说明书9页 附图3页

(54) 发明名称

网络模型训练方法、装置、文本分类模型及网络模型

(57) 摘要

本申请涉及一种网络模型训练方法、装置、文本分类模型及网络模型,属于计算机技术领域。该网络模型训练方法包括:获取训练样本集,所述训练样本集包括常见类样本和少见类样本;利用所述训练样本集对双边分支网络模型进行训练,得到训练好的文本分类模型,其中,所述双边分支网络模型的两个分支网络均包括基于多尺度注意力机制模块的编码层。本申请实施例中,通过基于多尺度注意力机制模块的编码层来构建双边分支网络模型,通过将多个信息头的信息进行融合,来获取丰富的语义信息,从而能使模型尽快收敛,以及提高模型的分类精度。



1. 一种网络模型训练方法,其特征在于,包括:

获取训练样本集,所述训练样本集包括常见类样本和少见类样本;

利用所述训练样本集对双边分支网络模型进行训练,得到训练好的文本分类模型,其中,所述双边分支网络模型的两个分支网络均包括基于多尺度注意力机制模块的编码层;

其中,训练过程中,通过梯度平均模长和训练轮次动态调整两个分支网络对应的自适应权重因子以动态调整两个分支网络的特征融合;

所述双边分支网络模型中的第一分支网络和第二分支网络的自适应权重因子分别为W1和W2;

其中, $W1 = a1 * g1$ ,  $W2 = a2 * g2$ ,  $a1 = 1 - T/2 * Tmax$ ,  $a2 = T/2 * Tmax$ ,  $g1 = \sum_{i=0}^K |gt_i^1 - fc_i| / K$ ,  $g2 = \sum_{i=0}^K |gt_i^2 - fr_i| / K$ , T表示当前训练轮次, Tmax表示训练最大轮次, K为最大样本类别,  $gt_i^1$ 和  $fc_i$ 分别表示当前训练轮次输入所述第一分支网络的样本中的第i类别样本的真实值和预测值,  $gt_i^2$ 和  $fr_i$ 分别表示当前训练轮次输入所述第二分支网络的样本中的第i类别样本的真实值和预测值, g1和g2为梯度平均模长。

2. 根据权利要求1所述的方法,其特征在于,利用所述训练样本集对双边分支网络模型进行训练,包括:

每次迭代训练时,对所述训练样本集中的样本进行随机采样,以及对所述训练样本集中的不同类别样本进行权重采样,在进行权重采样时,类别数量少的样本的采样频率大于类别数量多的样本的采样频率;

将随机采样的N个样本输入所述双边分支网络模型中的第一分支网络中,将权重采样的N个样本输入所述双边分支网络模型中的第二分支网络中,对所述双边分支网络模型进行迭代训练,其中, N为正整数,且小于所述训练样本集中的样本数。

3. 一种文本分类模型,用于处理待分类文本,其特征在于:

所述文本分类模型由如权利要求1-2中任一项所述的网络模型训练方法训练得到。

4. 一种网络模型,其特征在于,包括:

第一分支网络、第二分支网络,所述第一分支网络和所述第二分支网络均包括基于多尺度注意力机制模块的编码层;以及,合并层,用于将所述第一分支网络和所述第二分支网络各自输出的特征向量进行相加,并作为模型的最终预测值输出;

其中,训练过程中,通过梯度平均模长和训练轮次动态调整两个分支网络对应的自适应权重因子以动态调整两个分支网络的特征融合;

双边分支网络模型中的第一分支网络和第二分支网络的自适应权重因子分别为W1和W2;

其中, $W1 = a1 * g1$ ,  $W2 = a2 * g2$ ,  $a1 = 1 - T/2 * Tmax$ ,  $a2 = T/2 * Tmax$ ,  $g1 = \sum_{i=0}^K |gt_i^1 - fc_i| / K$ ,  $g2 = \sum_{i=0}^K |gt_i^2 - fr_i| / K$ , T表示当前训练轮次, Tmax表示训练最大轮次, K为最大样本类别,  $gt_i^1$ 和  $fc_i$ 分别表示当前训练轮次输入所述第一分支网络的样本中的第i类别样本的真实值和预测值,  $gt_i^2$ 和  $fr_i$ 分别表示当前训练轮次输入所述第二分支网络的样本中的第i类别样本的

真实值和预测值,  $g_1$ 和 $g_2$ 为梯度平均模长。

5. 根据权利要求4所述的模型, 其特征在于, 所述编码层中的多尺度注意力机制模块中不同信息头head对应的超参数权重不同。

6. 一种网络模型训练装置, 其特征在于, 包括:

获取模块, 用于获取训练样本集, 所述训练样本集包括常见类样本和少见类样本;

训练模块, 用于利用所述训练样本集对双边分支网络模型进行训练, 得到训练好的文本分类模型, 其中, 所述双边分支网络模型的两个分支网络均包括基于多尺度注意力机制模块的编码层; 还用于训练过程中, 通过梯度平均模长和训练轮次动态调整两个分支网络对应的自适应权重因子以动态调整两个分支网络的特征融合; 所述双边分支网络模型中的第一分支网络和第二分支网络的自适应权重因子分别为 $W_1$ 和 $W_2$ ; 其中,  $W_1 = a_1 * g_1$ ,  $W_2 = a_2 * g_2$ ,  $a_1 = 1 - T/2 * T_{max}$ ,  $a_2 = T/2 * T_{max}$ ,  $g_1 = \sum_{i=0}^K |gt_i^1 - fc_i| / K$ ,  $g_2 = \sum_{i=0}^K |gt_i^2 - fr_i| / K$ ,  $T$ 表示当前训练轮次,  $T_{max}$ 表示训练最大轮次,  $K$ 为最大样本类别,  $gt_i^1$ 和 $fc_i$ 分别表示当前训练轮次输入所述第一分支网络的样本中的第 $i$ 类别样本的真实值和预测值,  $gt_i^2$ 和 $fr_i$ 分别表示当前训练轮次输入所述第二分支网络的样本中的第 $i$ 类别样本的真实值和预测值,  $g_1$ 和 $g_2$ 为梯度平均模长。

7. 一种电子设备, 其特征在于, 包括:

存储器和处理器, 所述处理器与所述存储器连接;

所述存储器, 用于存储程序;

所述处理器, 用于调用存储于所述存储器中的程序, 以执行如权利要求1-2中任一项所述的方法。

8. 一种存储介质, 其特征在于, 其上存储有计算机程序, 所述计算机程序被处理器运行时, 执行如权利要求1-2中任一项所述的方法。

## 网络模型训练方法、装置、文本分类模型及网络模型

### 技术领域

[0001] 本申请属于计算机技术领域,具体涉及一种网络模型训练方法、装置、文本分类模型及网络模型。

### 背景技术

[0002] 随着产生的文本数量越来越多,纯靠人力去分类显然是不现实。目前文本分类技术主要分两类,一类是基于传统机器学习的分类方法,主要有贝叶斯,支持向量机(Support Vector Machine,SVM)等;另外一类是基于深度学习的分类方法。传统机器学习方法需要的数据量不是很多,也不过多依靠计算机,但是由于提取特征相对简单,模型的泛化能力相对较弱,分类准确性较差。而深度学习对数量依赖比较大,在实际应用中,由于有些数据的采集相对困难,并没有足够的数据进行训练,使得模型的收敛相对较难。另外,有些行业的数据量并不是很均衡,例如,病历文本分类中,一些少见病情对应的文本可能是不够的,这就造成模型对少见类缺乏足够的拟合能力,导致模型的分类精度不高。

### 发明内容

[0003] 鉴于此,本申请的目的在于提供一种网络模型训练方法、装置、文本分类模型及网络模型,以改善现有分类模型的收敛相对较慢以及分类精度不高的问题。

[0004] 本申请的实施例是这样实现的:

[0005] 第一方面,本申请实施例提供了一种网络模型训练方法,包括:获取训练样本集,所述训练样本集包括常见类样本和少见类样本;利用所述训练样本集对双边分支网络模型进行训练,得到训练好的文本分类模型,其中,所述双边分支网络模型的两个分支网络均包括基于多尺度注意力机制模块的编码层。

[0006] 本申请实施例中,通过基于多尺度注意力机制模块的编码层来构建双边分支网络模型,通过将多个信息头的信息进行融合,来获取丰富的语义信息,从而能使模型尽快收敛,以及提高模型的分类精度。

[0007] 结合第一方面实施例的一种可能的实施方式,利用所述训练样本集对双边分支网络模型进行训练,包括:每次迭代训练时,对所述训练样本集中的样本进行随机采样,以及对所述训练样本集中的不同类别样本进行权重采样,在进行权重采样时,类别数量少的样本的采样频率大于类别数量多的样本的采样频率;将随机采样的N个样本输入所述双边分支网络模型中的第一分支网络中,将权重采样的N个样本输入所述双边分支网络模型中的第二分支网络中,对所述双边分支网络模型进行迭代训练,其中,N为正整数,且小于所述训练样本集中的样本数。

[0008] 本申请实施例中,通过在每次迭代训练时,通过对训练样本集中的样本进行随机采样,将获得的N个样本输入第一分支网络中,对第一分支网络进行训练;通过对训练样本集中的不同类别样本进行权重采样,将获得的N个样本输入第二分支网络中,对第二分支网络进行训练,且在进行权重采样时,类别数量越少的样本,采样的频率越高,这样能解决模

型对少见类样本缺乏足够的拟合能力的问题。

[0009] 结合第一方面实施例的一种可能的实施方式,训练过程中,通过梯度平均模长和训练轮次动态调整两个分支网络对应的自适应权重因子,从而动态调整两个分支网络的特征融合。

[0010] 本申请实施例中,通过梯度平均模长和训练轮次动态调整两个分支网络对应的自适应权重因子,从而动态调整两个分支网络的特征融合,能够有效防止过度采样少见类样本数据造成模型对少见类样本数据过拟合,同时也能抑制常见类别数据对模型的影响过大的问题。

[0011] 结合第一方面实施例的一种可能的实施方式,所述双边分支网络模型中的第一分支网络和第二分支网络的自适应权重因子分别为 $W_1$ 和 $W_2$ ;其中, $W_1 = a_1 * g_1$ ,  $W_2 = a_2 * g_2$ ,  $a_1 = 1 - T/2 * T_{max}$ ,  $a_2 = T/2 * T_{max}$ ,  $g_1 = \sum_{i=0}^K |gt_i^1 - fc_i| / K$ ,  $g_2 = \sum_{i=0}^K |gt_i^2 - fr_i| / K$ ,  $T$ 表示当前训练轮次, $T_{max}$ 表示训练最大轮次, $K$ 为最大样本类别, $gt_i^1$ 和 $fc_i$ 分别表示当前训练轮次输入所述第一分支网络的样本中的第 $i$ 类别样本的真实值和预测值, $gt_i^2$ 和 $fr_i$ 分别表示当前训练轮次输入所述第二分支网络的样本中的第 $i$ 类别样本的真实值和预测值。

[0012] 本申请实施例中,在刚开始训练的时候, $T$ 比较小, $a_1$ 初始值接近1,等到训练后期接近0.5。在训练初期第一分支网络占的权重比较大,主要训练第一分支网络,以提高模型的基础特征提取能力;等到了后期第一分支网络的 $a_1$ 值和第二分支网络的 $a_2$ 值相近,此时,主要靠每个支路上的梯度平均模长( $g_1$ 、 $g_2$ )动态调节对应的权重,当第一分支网络上的支路梯度平均模长比较大,则占的权重就大,能够有效防止过度采样少见类样本数据造成模型对少见类样本数据过拟合,同时也能抑制常见类别数据对模型的影响过大的问题。

[0013] 第二方面,本申请实施例还提供了一种文本分类模型,用于处理待分类文本,所述文本分类模型由利用上述第一方面实施例和/或结合第一方面实施例的任一种可能的实施方式提供的网络模型训练方法训练得到。

[0014] 第三方面,本申请实施例还提供了一种网络模型,包括:第一分支网络、第二分支网络以及合并层;所述第一分支网络和所述第二分支网络均包括基于多尺度注意力机制模块的编码层;合并层,用于将所述第一分支网络和所述第二分支网络各自输出的特征向量进行相加,并作为模型的最终预测值输出。

[0015] 结合第三方面实施例的一种可能的实施方式,所述编码层中的多尺度注意力机制模块中不同信息头head对应的超参数权重不同。

[0016] 本申请实施例中,通过将编码层中的多尺度注意力机制模块中不同信息头head对应的超参数权重不同,从而使模型更倾向有重要信息的head,提高模型提取少见类样本的语义信息的能力。

[0017] 第四方面,本申请实施例还提供了一种网络模型训练装置,包括:获取模块以及训练模块;获取模块,用于获取训练样本集,所述训练样本集包括常见类样本和少见类样本;训练模块,用于利用所述训练样本集对双边分支网络模型进行训练,得到训练好的文本分类模型,其中,所述双边分支网络模型的两个分支网络均包括基于多尺度注意力机制模块的编码层。

[0018] 第五方面,本申请实施例还提供了一种电子设备,包括:存储器和处理器,所述处

理器与所述存储器连接;所述存储器,用于存储程序;所述处理器,用于调用存储于所述存储器中的程序,以执行上述第一方面实施例和/或结合第一方面实施例的任一种可能的实施方式提供的方法。

[0019] 第六方面,本申请实施例还提供了一种存储介质,其上存储有计算机程序,所述计算机程序被处理器运行时,执行上述第一方面实施例和/或结合第一方面实施例的任一种可能的实施方式提供的方法。

[0020] 本申请的其他特征和优点将在随后的说明书阐述,并且,部分地从说明书中变得显而易见,或者通过实施本申请实施例而了解。本申请的目的和其他优点可通过在所写的说明书以及附图中所特别指出的结构来实现和获得。

## 附图说明

[0021] 为了更清楚地说明本申请实施例或现有技术中的技术方案,下面将对实施例中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本申请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。通过附图所示,本申请的上述及其它目的、特征和优势将更加清晰。在全部附图中相同的附图标记指示相同的部分。并未刻意按实际尺寸等比例缩放绘制附图,重点在于示出本申请的主旨。

[0022] 图1示出了本申请实施例提供的一种双边分支网络模型的结构示意图。

[0023] 图2为现有的基于注意力机制模块的编码层的原理示意图。

[0024] 图3a示出了本申请实施例提供的一种基于多尺度注意力机制模块的编码层的原理示意图。

[0025] 图3b示出了本申请实施例提供的又一种基于多尺度注意力机制模块的编码层的原理示意图

[0026] 图3c示出了本申请实施例提供的一种基于自适应权重的多尺度注意力机制模块的编码层的原理示意图。

[0027] 图4示出了本申请实施例提供的一种网络模型训练方法的流程示意图。

[0028] 图5示出了本申请实施例提供的一种文本分类方法的流程示意图。

[0029] 图6示出了本申请实施例提供的一种网络模型训练装置的结构框图。

[0030] 图7示出了本申请实施例提供的一种电子设备的结构示意图。

## 具体实施方式

[0031] 下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行描述。

[0032] 应注意到:相似的标号和字母在下面的附图中表示类似项,因此,一旦某一项在一个附图中被定义,则在随后的附图中不需要对其进行进一步定义和解释。同时,在本申请的描述中诸如“第一”、“第二”等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多

限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0033] 再者,本申请中术语“和/或”,仅仅是一种描述关联对象的关联关系,表示可以存在三种关系,例如,A和/或B,可以表示:单独存在A,同时存在A和B,单独存在B这三种情况。

[0034] 鉴于现有的文本分类方法的准确性不高,对少见类样本缺乏足够的拟合能力的问题,本申请实施例提供了一种基于多尺度注意力机制的语义信息的提取和双边分支网络(Bilateral-Branch Network, BBN)模型的文本分类方法,能够有效提取语义信息加速模型收敛和提高模型的泛化能力,还能抑制长尾效应,提高模型对少见类样本的拟合能力。

[0035] 下面将对本申请涉及的包含多尺度注意力机制模块的双边分支网络模型进行说明。如图1所示,该双边分支网络模型包括:第一分支网络、第二分支网络和合并层。其中,第一分支网络和第二分支网络均包括基于多尺度注意力机制模块的编码层。

[0036] 其中,第一分支网络包括第一主干网络(backbone)、第一编码层(encoder)和第一特征提取层。训练样本X1输入第一主干网络后再经过第一编码层,获得特征向量Fc,特征向量Fc再经过第一特征提取层乘以自适应权重因子W1后,得到特征向量Wc。其中,第一主干网络由多层编码层组成,例如由11层编码层串接而成。第一主干网络和第一编码层均为基于多尺度注意力机制模块的编码层。

[0037] 第二分支网络包括第二主干网络(backbone)、第二编码层(encoder)和第二特征提取层。训练样本X2输入第二主干网络后再经过第二编码层,获得特征向量Fr,特征向量Fr再经过第二特征提取层乘以自适应权重因子W2后,得到特征向量Wr。其中,第二主干网络由多层编码层组成,例如由11层编码层串接而成。第二主干网络和第二编码层均为基于多尺度注意力机制模块的编码层。

[0038] 其中,第一主干网络和第二主干网络共享权重值,第一编码层和第二编码层不共享权重值。第一主干网络和第二主干网络用于从输入样本中提取基础特征,第一编码层和第二编码层用于提取语义特征。本申请实施例中采用基于多尺度注意力机制模块的编码层来搭建双边分支网络模型,能够有效提取语义信息加速模型收敛和提高模型的泛化能力。

[0039] 现有的双边分支网络模型中的编码层中的注意力机制模块(self-Attention)是将上一层的信息头输出的信息进行累加,再输入到下一层中,如图2所示,j层的信息头累加融合到j+1层(其中, $h_0^j$ 表示第j层第0个信息头),该机制只能学习单个信息头之间的关系。所以,self-Attention机制为了获取丰富的语义,需要大量的语料进行长时间的训练,才能使模型收敛。

[0040] 本申请实施例中,构建基于多尺度注意力机制模块(Scale-Aware Self-Attention)的编码层,通过将多个信息头的信息进行融合,来获取丰富的语义信息。如图3a所示,head1的信息来自信息头 $h_0^j$ ,head2的信息来自信息头 $h_0^j$ 、 $h_1^j$ 、 $h_2^j$ ,head3的信息来自信息头 $h_0^j$ 、 $h_1^j$ 、 $h_2^j$ 、 $h_3^j$ 、 $h_4^j$ ,然后再将head1、head2、head3的信息传给下一层的信息头 $h_0^{j+1}$ 。下一层编码层的其他信息头的获取和 $h_0^{j+1}$ 类似,例如 $h_1^{j+1}$ 的head1的信息来自信息头 $h_1^j$ ,head2的信息来自信息头 $h_1^j$ 、 $h_2^j$ 、 $h_3^j$ ,head3的信息来自信息头 $h_0^j$ 、 $h_1^j$ 、 $h_2^j$ 、 $h_3^j$ 、 $h_4^j$ ,如图3b所示。

[0041] 考虑到由于文本的重要信息的分布相对随机,因此本发明给各个head设置不同的超参数权重,让模型去学习该权重值,如图3c所示。这样模型就能根据输入文本学习得到不同head的权重,从而更倾向有重要信息的head。也即编码层中的多尺度注意力机制模块中不同信息头head对应的超参数权重不同,此时该多尺度注意力机制模块为自适应权重的多尺度注意力机制模块(Weighted Scale-Aware Self-Attention)。图3c中可以看出head1的超参数权重为H1,head2的超参数权重为H2,head3的超参数权重为H3,其中,H1、H2、H3的值不同。

[0042] 合并层,用于将第一分支网络和第二分支网络各自输出的特征向量进行相加,并作为模型的最终预测值输出。也即将第一分支网络输出的特征向量 $W_c$ 和第二分支网络输出的特征向量 $W_r$ 进行相加,并作为模型的最终预测值(Loss)输出。

[0043] 为了解决长尾效应(某些类别的数据会比较少,而其他类别的数据比较多),造成类别质检的数量失衡,模型主要学习到常见类类别,对少见类类别欠拟合的问题。本申请实施例中,在模型训练过程中通过梯度平均模长和训练轮次动态调整两个分支对应的自适应权重因子,从而动态调整两个分支的特征融合,实现特征的均衡化。为了便于理解,下面将结合图4对本申请实施例提供的网络模型训练方法进行说明。该模型训练方法包括:

[0044] 步骤S101:获取训练样本集,训练样本集包括常见类样本和少见类样本。

[0045] 其中,获取训练样本集的过程包括:获取包含至少两种类别的样本,且至少两种类别的样本中常见类样本和少见类样本的数量比为 $M:1$ , $M$ 为大于等于20的正整数。例如,该训练样本集中样本的数量有10万条(并不限于此),其中少见类样本类别和常见类样本类别之间的数量比为1:20。

[0046] 步骤S102:利用所述训练样本集对双边分支网络模型进行训练,得到训练好的文本分类模型。

[0047] 在获取到训练样本集后,利用获取的训练样本集对双边分支网络模型进行训练,便可得到训练好的文本分类模型。其中,该双边分支网络模型即为上述所述的双边分支网络模型,此处不再对模型结构进行说明。

[0048] 其中,利用训练样本集对双边分支网络模型进行训练的过程包括:每次迭代训练时,对训练样本集中的样本进行随机采样,以及对训练样本集中的不同类别样本进行权重采样,在进行权重采样时,类别数量少的样本的采样频率大于类别数量多的样本的采样频率(也即类别数量越少的样本,采样的频率越高),将随机采样的 $N$ 个样本输入双边分支网络模型中的第一分支网络中,将权重采样的 $N$ 个样本输入所述双边分支网络模型中的第二分支网络中,对双边分支网络模型进行迭代训练,其中, $N$ 为正整数,且小于训练样本集中的样本数。结合上述的模型结构进行说明,也即在对第一分支网络进行训练的训练样本 $X_1$ 为对训练样本集中的样本进行随机采样获得的 $N$ 个样本,而对第二分支网络进行训练的训练样本 $X_2$ 为对训练样本集中的不同类别样本进行权重采样而获得的 $N$ 个样本。且在进行权重采样时,类别数量越少的样本,采样的频率越高,这样能解决模型对少见类样本缺乏足够的拟合能力的问题。

[0049] 为了防止过度采样少见类样本数据造成模型对少见类样本数据过拟合,同时也能抑制常见类别数据对模型的影响过大的问题,本申请实施例中,训练过程中,通过梯度平均模长和训练轮次动态调整两个分支网络对应的自适应权重因子,从而动态调整两个分支网

络的特征融合。

[0050] 其中,双边分支网络模型中的第一分支网络和第二分支网络的自适应权重因子分别为 $W1$ 和 $W2$ 。其中, $W1=a1*g1$ , $W2=a2*g2$ , $a1=1-T/2*T \max$ , $a2=T/2*T \max$ ,

$g1 = \sum_{i=0}^K |gt_i^1 - fc_i|/K$ ,  $g2 = \sum_{i=0}^K |gt_i^2 - fr_i|/K$ ,  $T$ 表示当前训练轮次, $T_{\max}$ 表示训练最大轮次(也即迭代总次数), $K$ 为最大样本类别, $gt_i^1$ 和 $fc_i$ 分别表示当前训练轮次输入第一分支网络的样本中的第 $i$ 类别样本的真实值和预测值, $gt_i^2$ 和 $fr_i$ 分别表示当前训练轮次输入第二分支网络的样本中的第 $i$ 类别样本的真实值和预测值。为了便于理解,举例进行说明,假设训练最大轮次为20,则 $T_{\max}=20$ ,若当前为第10次迭代,则当前训练轮次 $T=10$ , $K$ 为最大样本类别,假设共有10类样本类别,则 $K=10$ 。其中,上式中的 $g1$ 和 $g2$ 分别用于求取第一分支网络和第二分支网络的梯度平均模长的。

[0051] 通过上述公式可以看出,在刚开始训练的时候, $T$ 比较小, $a1$ 初始值接近1,等到训练后期接近0.5。在训练初期第一分支网络占的权重比较大,主要训练第一分支网络,以提高模型的基础特征提取能力。等到了后期第一分支网络的 $a1$ 值和第二分支网络的 $a2$ 值相近,此时,主要靠每个支路上的梯度平均模长( $g1$ 、 $g2$ )动态调节对应的权重。例如,当第一分支网络上的支路梯度平均模长比较大,则占的权重就大,能够有效防止过度采样少见类样本数据造成模型对少见类样本数据过拟合,同时也能抑制常见类别数据对模型的影响过大的问题。

[0052] 本申请通过基于自适应权重的多尺度注意力机制模块(Weighted Scale-Aware Self-Attention)来构建多注意力机制编码层;然后用该编码层搭建双边分支网络模型,并将通过随机采样的训练样本和通过权重采样的训练样本分别输入模型两个分支进行训练,训练过程中通过梯度平均模长和训练轮次动态调整两个分支网络对应的自适应权重因子,从而动态调整两个分支网络的特征融合,实现特征的均衡化;最后便可利用训练好的模型进行文本分类。

[0053] 为了便于说明本申请改进后的模型的性能提升,下面将普通注意力机制(Self-Attention)、多尺度注意力机制(Scale-Aware Self-Attention)、自适应权重的多尺度注意力机制(Weighted Scale-Aware Self-Attention)以及本申请的包含自适应权重的多尺度注意力机制的双边分支网络模型(Weighted Scale-Aware Self-Attention+BBN)的模型收敛次数以及准确率进行对比。通过获取包含至少两种类别的样本,且至少两种类别的样本中常见类样本和少见类样本的数量比为20:1的训练样本集(包括训练集和测试集),分别对普通注意力机制(Self-Attention)、多尺度注意力机制(Scale-Aware Self-Attention)、自适应权重的多尺度注意力机制(Weighted Scale-Aware Self-Attention)以及本申请的包含自适应权重的多尺度注意力机制的双边分支网络模型(Weighted Scale-Aware Self-Attention+BBN)进行训练和测试,其训练和测试实验结果如表1可知。

[0054] 表1

模型	准确率	少见类样本准确率	模型收敛需要训练次数
Self-Attention	72%	40%	10
Scale-Aware Self-Attention	85%	51%	5
[0055] Weighted Scale-Aware Self-Attention	89%	58%	3
Weighted Scale-Aware Self-Attention+BBN	92%	81%	4

[0056] 通过表1可知,基于自适应权重的多尺度注意力机制的双边分支网络模型能够有效的提高模型的语义提取能力,而且加速模型的收敛和提高分类的准确性。

[0057] 基于同样的发明构思,本申请实施例还提供了一种文本分类模型,用于处理待分类文本,该用于处理待分类文本由图4所述的网络模型训练方法训练得到。

[0058] 基于同样的发明构思,下面将结合图5,对本申请实施例提供的文本分类方法进行说明。

[0059] 步骤S201:获取待分类文本。

[0060] 步骤S202:利用上述的网络模型训练方法训练的文本分类模型对所述待分类文本进行处理,得到分类结果。

[0061] 当需要对待分类文本进行分类时,利用上述的网络模型训练方法训练的文本分类模型(包含多尺度注意力机制模块的双边分支网络模型)对待分类文本进行处理,便可得到分类结果。

[0062] 其中,在进行文本预测时,使用双边分支网络模型中的第二分支网络对待分类文本进行预测和分类。

[0063] 其中,训练文本分类模型的方法,以及文本分类模型的具体结果请参照前述相同部分即可。

[0064] 基于同样的发明构思,本申请实施例还提供了一种网络模型训练装置100,如图6所示。该网络模型训练装置100包括:获取模块110、训练模块120。

[0065] 获取模块110,用于获取训练样本集,所述训练样本集包括常见类样本和少见类样本。

[0066] 训练模块120,用于利用所述训练样本集对双边分支网络模型进行训练,得到训练好的文本分类模型,其中,所述双边分支网络模型的两个分支网络均包括基于多尺度注意力机制模块的编码层。

[0067] 在本申请实施例中,训练模块120,具体用于:每次迭代训练时,对所述训练样本集中的样本进行随机采样,以及对所述训练样本集中的不同类别样本进行权重采样,在进行权重采样时,类别数量越少的样本,采样的频率越高;将随机采样的N个样本输入所述双边分支网络模型中的第一分支网络中,将权重采样的N个样本输入所述双边分支网络模型中的第二分支网络中,对所述双边分支网络模型进行迭代训练,其中,N为正整数,且小于所述训练样本集中的样本数。

[0068] 可选地,在本申请实施例中,训练过程中,通过梯度平均模长和训练轮次动态调整两个分支网络对应的自适应权重因子,从而动态调整两个分支网络的特征融合。

[0069] 可选地,在本申请实施例中,所述双边分支网络模型中的第一分支网络和第二分支网络的自适应权重因子分别为 $W1$ 和 $W2$ ;其中, $W1=a1*g1$ , $W2=a2*g2$ , $a1=1-T/2*Tmax$ , $a2=T/2*Tmax$ , $g1=\sum_{i=0}^K|gt_i^1-fc_i|/K$ , $g2=\sum_{i=0}^K|gt_i^2-fr_i|/K$ , $T$ 表示当前训练轮次, $Tmax$ 表示训练最大轮次, $K$ 为最大样本类别, $gt_i^1$ 和 $fc_i$ 分别表示当前训练轮次输入所述第一分支网络的样本中的第 $i$ 类别样本的真实值和预测值, $gt_i^2$ 和 $fr_i$ 分别表示当前训练轮次输入所述第二分支网络的样本中的第 $i$ 类别样本的真实值和预测值。

[0070] 本申请实施例所提供的网络模型训练装置100,其实现原理及产生的技术效果和前述方法实施例相同,为简要描述,装置实施例部分未提及之处,可参考前述方法实施例中相应内容。

[0071] 基于同样的发明构思,如图7所示,图7示出了本申请实施例提供的一种电子设备200的结构框图。所述电子设备200包括:收发器210、存储器220、通讯总线230以及处理器240。

[0072] 所述收发器210、所述存储器220、处理器240各元件相互之间直接或间接地电性连接,以实现数据的传输或交互。例如,这些元件相互之间可通过一条或多条通讯总线230或信号线实现电性连接。其中,收发器210用于收发数据。存储器220用于存储计算机程序,如存储有图6中所示的软件功能模块,即网络模型训练装置100。其中,网络模型训练装置100包括至少一个可以软件或固件(firmware)的形式存储于所述存储器220中或固化在所述电子设备200的操作系统(operating system,OS)中的软件功能模块。所述处理器240,用于执行存储器220中存储的可执行模块,例如网络模型训练装置100包括的软件功能模块或计算机程序。例如,处理器240,用于获取训练样本集,所述训练样本集包括常见类样本和少见类样本;利用所述训练样本集对双边分支网络模型进行训练,得到训练好的文本分类模型,其中,所述双边分支网络模型的两个分支网络均包括基于多尺度注意力机制模块的编码层。

[0073] 其中,存储器220可以是,但不限于,随机存取存储器(Random Access Memory, RAM),只读存储器(Read Only Memory,ROM),可编程只读存储器(Programmable Read-Only Memory,PROM),可擦除只读存储器(Erasable Programmable Read-Only Memory,EPR0M),电可擦除只读存储器(Electric Erasable Programmable Read-Only Memory,EEPROM)等。

[0074] 处理器240可能是一种集成电路芯片,具有信号的处理能力。上述的处理器可以是通用处理器,包括中央处理器(Central Processing Unit,CPU)、网络处理器(Network Processor,NP)等;还可以是数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现场可编程门阵列(Field Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本申请实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器240也可以是任何常规的处理器等。

[0075] 其中,上述的电子设备200,包括但不限于计算机、服务器等。

[0076] 本申请实施例还提供了一种非易失性计算机可读取存储介质(以下简称存储介质),该存储介质上存储有计算机程序,该计算机程序被计算机如上述的电子设备200运行时,执行上述所示的网络模型训练方法,或者上述的文本分类方法。

[0077] 需要说明的是,本说明书中的各个实施例均采用递进的方式描述,每个实施例重

点说明的都是与其他实施例的不同之处,各个实施例之间相同相似的部分互相参见即可。

[0078] 在本申请所提供的几个实施例中,应该理解到,所揭露的装置和方法,也可以通过其它的方式实现。以上所描述的装置实施例仅仅是示意性的,例如,附图中的流程图和框图显示了根据本申请的多个实施例的装置、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或代码的一部分,所述模块、程序段或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意,在有些作为替换的实现方式中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个连续的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意的是,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0079] 另外,在本申请各个实施例中的各功能模块可以集成在一起形成一个独立的部分,也可以是各个模块单独存在,也可以两个或两个以上模块集成形成一个独立的部分。

[0080] 所述功能如果以软件功能模块的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可以是个人计算机,笔记本电脑,服务器,或者电子设备等)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(Read-Only Memory, ROM)、随机存取存储器(Random Access Memory, RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

[0081] 以上所述,仅为本申请的具体实施方式,但本申请的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到变化或替换,都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应所述以权利要求的保护范围为准。

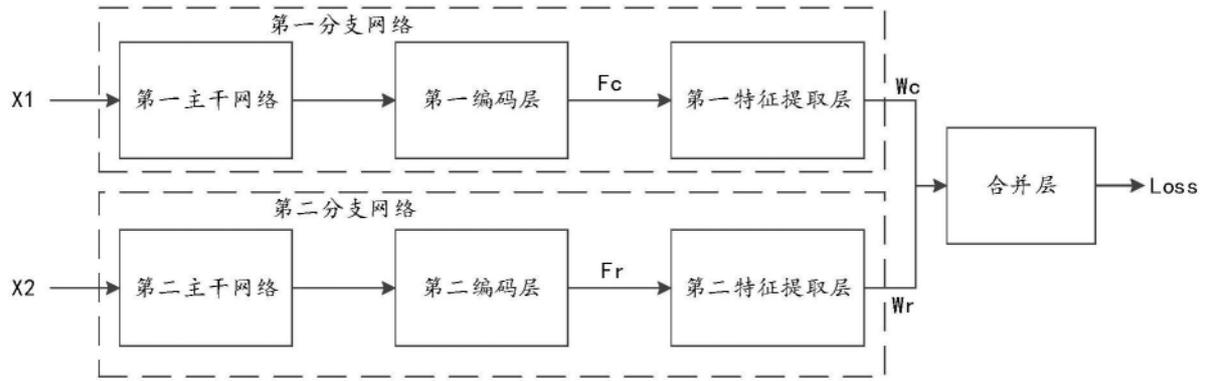


图1

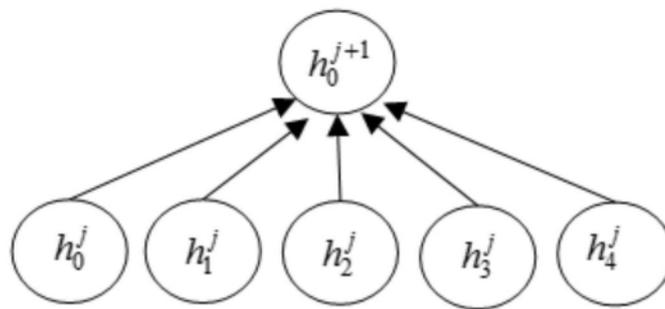


图2

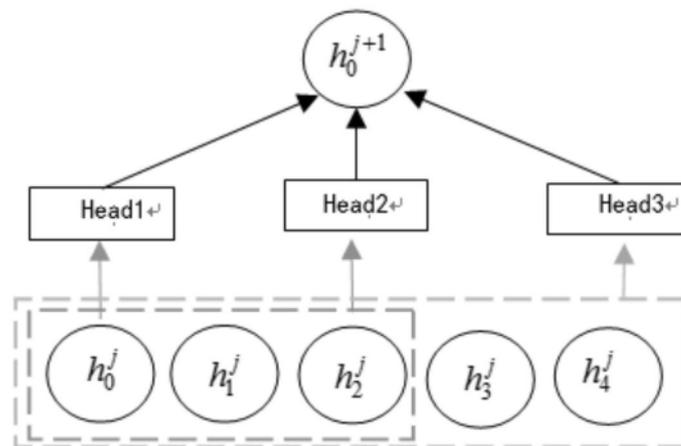


图3a

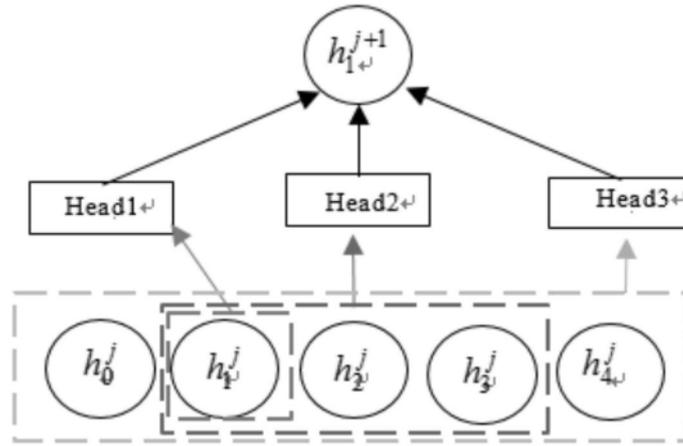


图3b

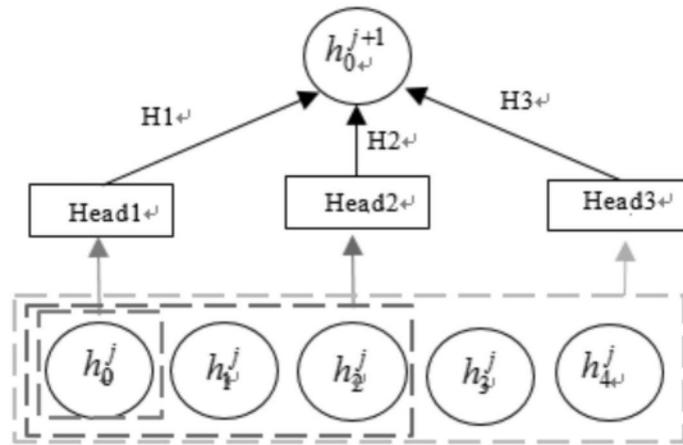


图3c

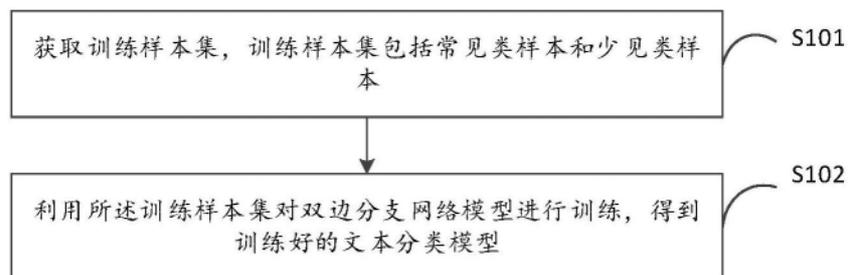


图4

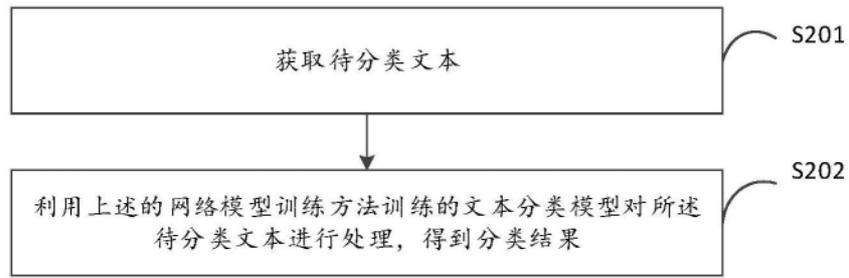


图5

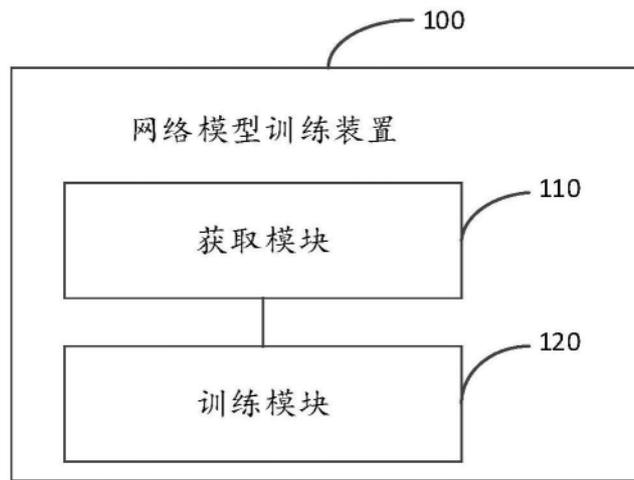


图6

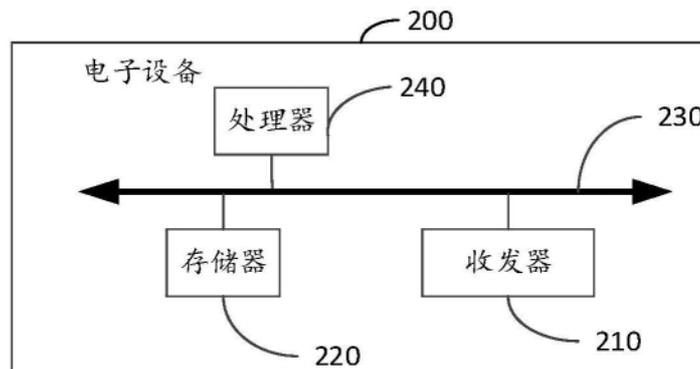


图7