



(12) 发明专利

(10) 授权公告号 CN 113434721 B

(45) 授权公告日 2024.07.19

(21) 申请号 202110734783.6

G06V 10/82 (2022.01)

(22) 申请日 2021.06.30

G06V 10/80 (2022.01)

(65) 同一申请的已公布的文献号

G06V 20/62 (2022.01)

申请公布号 CN 113434721 A

G06N 3/0455 (2023.01)

G06N 3/0464 (2023.01)

(43) 申请公布日 2021.09.24

G06N 3/08 (2023.01)

(73) 专利权人 平安科技(深圳)有限公司

(56) 对比文件

地址 518000 广东省深圳市福田区福田街
道福安社区益田路5033号平安金融中
心23楼

CN 112651448 A, 2021.04.13

审查员 杨驭涵

(72) 发明人 刘杰 王健宗 瞿晓阳

(74) 专利代理机构 深圳市明日今典知识产权代

理事务所(普通合伙) 44343

专利代理师 王杰辉 罗燕

(51) Int. Cl.

G06V 10/764 (2022.01)

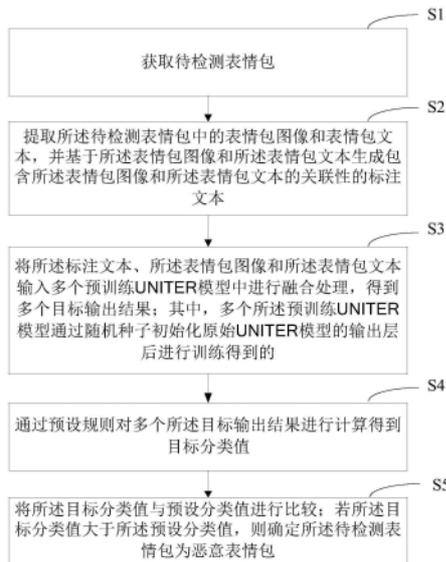
权利要求书2页 说明书9页 附图3页

(54) 发明名称

表情包分类方法、装置、计算机设备和存储
介质

(57) 摘要

本申请涉及人工智能领域,提供了一种表情
包分类方法、装置、计算机设备和存储介质,获取
待检测表情包;提取所述待检测表情包中的表情
包图像和表情包文本,并基于表情包图像和表情
包文本生成包含所述表情包图像和表情包文本
的关联性的标注文本;将标注文本、表情包图像
和表情包文本输入多个预训练UNITER模型中进
行融合处理,得到多个目标输出结果;通过预设
规则对多个目标输出结果进行计算得到目标分
类值;将目标分类值与预设分类值进行比较;目
标分类值大于预设分类值时确定待检测表情
包为恶意表情包。根据本申请提供的表情包分类
方法、装置、计算机设备和存储介质,能够更加准确
的对表情包进行分类。



1. 一种表情包分类方法,其特征在于,包括以下步骤:

获取待检测表情包;

提取所述待检测表情包中的表情包图像和表情包文本,并基于所述表情包图像和所述表情包文本生成包含所述表情包图像和所述表情包文本的关联性的标注文本;

将所述标注文本、所述表情包图像和所述表情包文本输入多个预训练UNITER模型中进行融合处理,得到多个目标输出结果;其中,多个所述预训练UNITER模型通过随机种子初始化原始UNITER模型的输出层后进行训练得到的;

通过预设规则对多个所述目标输出结果进行计算得到目标分类值;

将所述目标分类值与预设分类值进行比较;若所述目标分类值大于所述预设分类值,则确定所述待检测表情包为恶意表情包;

所述将所述标注文本、所述表情包图像和所述表情包文本输入多个预训练UNITER模型中进行融合处理,得到多个目标输出结果的步骤中,任意一个所述预训练UNITER模型得到对应的目标输出结果的过程,包括:

将所述标注文本输入所述预训练UNITER模型中进行处理,得到 $N \times 1$ 的矩阵作为第一标注文本输出结果;

将所述表情包图像文本输入所述预训练UNITER模型中进行处理,得到 $N \times 1$ 的矩阵作为第一表情包图像输出结果;

将所述表情包文本输入所述预训练UNITER模型中进行处理,得到 $N \times 1$ 的矩阵作为第一表情包文本输出结果;

将所述第一标注文本输出结果、所述第一表情包图像输出结果和所述第一表情包文本输出结果进行融合处理,得到所述目标输出结果;

所述将所述第一标注文本输出结果、所述第一表情包图像输出结果和所述第一表情包文本输出结果进行融合处理,得到所述目标输出结果的步骤,包括:

通过公式 $L = (v + y)(x + v)^T$ 对所述第一标注文本输出结果、所述第一表情包图像输出结果和所述第一表情包文本输出结果进行融合处理,得到第一目标结果;其中,所述 v 为所述第一表情包图像输出结果,所述 y 为所述第一表情包文本输出结果,所述 x 为所述第一标注文本输出结果,所述 T 为转置;

将所述第一目标结果经过归一化处理,得到所述目标输出结果;

所述将所述目标分类值与预设分类值进行比较;若所述目标分类值大于所述预设分类值,则确定所述待检测表情包为恶意表情包的步骤,包括:

计算所述标注文本和所述表情包文本的相似度;

将所述相似度与预设阈值进行比较;

若所述相似度小于所述预设阈值,在所述目标分类值上增加预设惩罚值作为新的目标分类值;

将所述新的目标分类值与所述预设分类值进行比较;

若所述新的目标分类值大于所述预设分类值,则确定所述待检测表情包为恶意表情包。

2. 根据权利要求1所述的表情包分类方法,其特征在于,所述提取所述待检测表情包中

的表情包图像和表情包文本,并基于所述表情包图像和所述表情包文本生成包含所述表情包图像和所述表情包文本的关联性的标注文本的步骤,包括:

通过pixelLink模型检测所述待检测表情包上的文本区域;

通过fasterRCNN模型在所述文本区域识别所述表情包文本;

通过mask-rcnn模型或yolo v5模型提取所述表情包图像;

将所述表情包文本和所述表情包图像输入multi-head attention模型中进行信息融合,得到所述标注文本。

3. 根据权利要求1所述的表情包分类方法,其特征在于,各所述预训练UNITER模型的输出层的神经元均为N个,所述N大于等于2000。

4. 根据权利要求1所述的表情包分类方法,其特征在于,所述计算所述标注文本和所述表情包文本的相似度的步骤,包括:

通过word2vec模型分别得到所述标注文本和所述表情包文本的词向量;

根据标注文本和所述表情包文本的词向量计算所述标注文本和所述表情包文本的余弦相似度。

5. 根据权利要求1所述的表情包分类方法,其特征在于,所述通过预设规则对多个所述目标输出结果进行计算得到目标分类值的步骤,包括:

根据公式 $S=k_1a_1+k_2a_2+\dots+k_na_n$ 得到所述目标分类值;其中, k_1, k_2, \dots, k_n 为各个所述目标输出结果的预设权重, n 为所述预训练UNITER模型的个数, a_1, a_2, \dots, a_n 为所述目标输出结果, $k_1+k_2+\dots+k_n=1$ 。

6. 一种恶意表情包确定装置,其特征在于,用于实现权利要求1-5任一项所述的表情包分类方法,包括:

获取单元,用于获取待检测表情包;

提取单元,用于提取所述待检测表情包中的表情包图像和表情包文本,并基于所述表情包图像和所述表情包文本生成包含所述表情包图像和所述表情包文本的关联性的标注文本;

处理单元,用于将所述标注文本、所述表情包图像和所述表情包文本输入多个预训练UNITER模型中进行融合处理,得到多个目标输出结果;其中,多个所述预训练UNITER模型通过随机种子初始化原始UNITER模型的输出层后进行训练得到的;

目标分类单元,用于通过预设规则对多个所述目标输出结果进行计算得到目标分类值;

分类单元,用于将所述目标分类值与预设分类值进行比较;若所述目标分类值大于所述预设分类值,则确定所述待检测表情包为恶意表情包。

7. 一种计算机设备,包括存储器和处理器,所述存储器中存储有计算机程序,其特征在于,所述处理器执行所述计算机程序时实现权利要求1至5中任一项所述的表情包分类方法的步骤。

8. 一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1至5中任一项所述的表情包分类方法的步骤。

表情包分类方法、装置、计算机设备和存储介质

技术领域

[0001] 本申请涉及人工智能的技术领域,特别涉及一种表情包分类方法、装置、计算机设备和存储介质。

背景技术

[0002] 表情包在社交媒体上作为一种信息传递媒介,占据着重要的地位。研究人员一直试图使用机器学习的方法来有效识别恶意表情包并加以禁止,但是目前几乎没有模型可以优于人工筛查。此项任务困难的原因在于,表情包分类涉及到多方因素。不同于传统文字或图像分类,表情包的分类需要同时考虑语义信息与图像信息,并对两者之间的组合效应加以理解,这也被称为多模态学习。多模态学习的特点就是不同模块信息的组合可能表达出乎意料的含义。举例说明:中性文字‘看看多少人喜欢你’与中性图片‘一片沙漠’的组合实际上可以传达出恶意信息。

[0003] 现有的基于多模态学习表情包分类模型并没有比纯文本学习更加准确。并且,某些经过特殊处理的表情包数据集,增设了善意反转机制,使得分类难度进一步提高。举例说明:如果将上面例子中的沙漠图片换成一张人群图片,则表情包在文字信息不变的情况下,会由恶意转为善意。由于文字信息并没有改变,而图片的变化就改变了表情包的分类,这将对模型的多模态分析能力提出更高的要求。

发明内容

[0004] 本申请的主要目的为提供一种表情包分类方法、装置、计算机设备和存储介质,旨在解决表情包分类不准确的技术问题。

[0005] 为实现上述目的,本申请提供了一种表情包分类方法,包括以下步骤:

[0006] 获取待检测表情包;

[0007] 提取所述待检测表情包中的表情包图像和表情包文本,并基于所述表情包图像和所述表情包文本生成包含所述表情包图像和所述表情包文本的关联性的标注文本;

[0008] 将所述标注文本、所述表情包图像和所述表情包文本输入多个预训练UNITER模型中进行融合处理,得到多个目标输出结果;其中,多个所述预训练UNITER模型通过随机种子初始化原始UNITER模型的输出层后进行训练得到的;

[0009] 通过预设规则对多个所述目标输出结果进行计算得到目标分类值;

[0010] 将所述目标分类值与预设分类值进行比较;若所述目标分类值大于所述预设分类值,则确定所述待检测表情包为恶意表情包。

[0011] 进一步地,所述提取所述待检测表情包中的表情包图像和表情包文本,并基于所述表情包图像和所述表情包文本生成包含所述表情包图像和所述表情包文本的关联性的标注文本的步骤,包括:

[0012] 通过pixellink模型检测所述待检测表情包上的文本区域;

[0013] 通过fasterRCNN模型在所述文本区域识别所述表情包文本;

- [0014] 通过mask-rcnn模型或yolo v5模型提取所述表情包图像；
- [0015] 将所述表情包文本和所述表情包图像输入multi-head attention模型中进行信息融合,得到所述标注文本。
- [0016] 进一步地,各所述预训练UNITER模型的输出层的神经元均为N个,所述N大于等于2000,所述将所述标注文本、所述表情包图像和所述表情包文本输入多个预训练UNITER模型中进行融合处理,得到多个目标输出结果的步骤中,任意一个所述预训练UNITER模型得到对应的目标输出结果的过程,包括:
- [0017] 将所述标注文本输入所述预训练UNITER模型中进行处理,得到N*1的矩阵作为第一标注文本输出结果；
- [0018] 将所述表情包图像文本输入所述预训练UNITER模型中进行处理,得到N*1的矩阵作为第一表情包图像输出结果；
- [0019] 将所述表情包文本输入所述预训练UNITER模型中进行处理,得到N*1的矩阵作为第一表情包文本输出结果；
- [0020] 将所述第一标注文本输出结果、所述第一表情包图像输出结果和所述第一表情包文本输出结果进行融合处理,得到所述目标输出结果。
- [0021] 进一步地,所述将所述第一标注文本输出结果、所述第一表情包图像输出结果和所述第一表情包文本输出结果进行融合处理,得到所述目标输出结果的步骤,包括:
- [0022] 通过公式 $L = (v + y)(x + v)^T$ 对所述第一标注文本输出结果、所述第一表情包图像输出结果和所述第一表情包文本输出结果进行融合处理,得到第一目标结果;其中,所述v为所述第一表情包图像输出结果,所述y为所述第一表情包文本输出结果,所述x为所述第一标注文本输出结果,所述T为转置；
- [0023] 将所述第一目标结果经过归一化处理,得到所述目标输出结果。
- [0024] 进一步地,所述将所述目标分类值与预设分类值进行比较;若所述目标分类值大于所述预设分类值,则确定所述待检测表情包为恶意表情包的步骤,包括:
- [0025] 计算所述标注文本和所述表情包文本的相似度；
- [0026] 将所述相似度与预设阈值进行比较；
- [0027] 若所述相似度小于所述预设阈值,在所述目标分类值上增加预设惩罚值作为新的目标分类值；
- [0028] 将所述新的目标分类值与所述预设分类值进行比较；
- [0029] 若所述新的目标分类值大于所述预设分类值,则确定所述待检测表情包为恶意表情包。
- [0030] 进一步地,所述计算所述标注文本和所述表情包文本的相似度的步骤,包括:
- [0031] 通过word2vec模型分别得到所述标注文本和所述表情包文本的词向量；
- [0032] 根据标注文本和所述表情包文本的词向量计算所述标注文本和所述表情包文本的余弦相似度。
- [0033] 进一步地,所述通过预设规则对多个所述目标输出结果进行计算得到目标分类值的步骤,包括:
- [0034] 根据公式 $S=k_1a_1+k_2a_2+\dots+k_na_n$ 得到所述目标分类值;其中,所述k为各个所述目标输出结果的预设权重,n为所述预训练UNITER模型的个数,所述a为所述目标输出结果,

$k_1+k_2+\dots+k_n=1$ 。

[0035] 本申请还提供一种恶意表情包确定装置,包括:

[0036] 获取单元,用于获取待检测表情包;

[0037] 提取单元,用于提取所述待检测表情包中的表情包图像和表情包文本,并基于所述表情包图像和所述表情包文本生成包含所述表情包图像和所述表情包文本的关联性的标注文本;

[0038] 处理单元,用于将所述标注文本、所述表情包图像和所述表情包文本输入多个预训练UNITER模型中进行融合处理,得到多个目标输出结果;其中,多个所述预训练UNITER模型通过随机种子初始化原始UNITER模型的输出层后进行训练得到的;

[0039] 目标分类单元,用于通过预设规则对多个所述目标输出结果进行计算得到目标分类值;

[0040] 分类单元,用于将所述目标分类值与预设分类值进行比较;若所述目标分类值大于所述预设分类值,则确定所述待检测表情包为恶意表情包。

[0041] 本申请还提供一种计算机设备,包括存储器和处理器,所述存储器中存储有计算机程序,所述处理器执行所述计算机程序时实现上述任一项所述的表情包分类方法的步骤。

[0042] 本申请还提供一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现上述任一项所述的表情包分类方法的步骤。

[0043] 本申请提供的表情包分类方法、装置、计算机设备和存储介质,预训练UNITER模型可以应对成分复杂的表情包信息,在预训练UNITER模型中引入表情包图像、表情包文本和标注文本得到目标输出结果,可使模型更有效地利用图像信息和文本信息,使其表现优于纯依赖文本或图像模型,同时根据多个目标输出结果得到目标分类值,进而根据目标分类值去对待检测表情包进行分类,能够更加准确对表情包进行分类。

附图说明

[0044] 图1 是本申请一实施例中表情包分类方法步骤示意图;

[0045] 图2 是本申请一实施例中表情包分类装置结构框图;

[0046] 图3 为本申请一实施例的计算机设备的结构示意图。

[0047] 本申请目的的实现、功能特点及优点将结合实施例,参照附图做进一步说明。

具体实施方式

[0048] 为了使本申请的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本申请进行进一步详细说明。应当理解,此处描述的具体实施例仅仅用以解释本申请,并不用于限定本申请。

[0049] 参照图1,本申请一实施例提供了一种表情包分类方法,包括以下步骤:

[0050] 步骤S1,获取待检测表情包;

[0051] 步骤S2,提取所述待检测表情包中的表情包图像和表情包文本,并基于所述表情包图像和所述表情包文本生成包含所述表情包图像和所述表情包文本的关联性的标注文本;

[0052] 步骤S3,将所述标注文本、所述表情包图像和所述表情包文本输入多个预训练UNITER模型中进行融合处理,得到多个目标输出结果;其中,多个所述预训练UNITER模型通过随机种子初始化原始UNITER模型的输出层后进行训练得到的;

[0053] 步骤S4,通过预设规则对多个所述目标输出结果进行计算得到目标分类值;

[0054] 步骤S5,将所述目标分类值与预设分类值进行比较;若所述目标分类值大于所述预设分类值,则确定所述待检测表情包为恶意表情包。

[0055] 本实施例中,如上述步骤S1所述,本实施例中获取的待检测表情包为同时带有文字和图像的表情包。在其他实施例中仅有文字或图像的表情包可采用其他的直接基于文字和图像进行分类的方法进行分类。

[0056] 如上述步骤S2所述,通过预设规则将表情包文本和表情包图像从待检测表情包中提取出来,同时还应得到标注文本,所述标注文本是学习到表情包文本和表情包图像之间的关联意义的一种分类标签形成的。

[0057] 如上述步骤S3-S4所述,UNITER模型是将图像和文本视为同一数据流的两个部分,被拼接后一起传入同一个Transformer中,直接进行多模态融合。即UNITER模型的输入是表情包图像RoI(Region of Interest)序列与表情包文本Token序列的简单拼接。RoI是特征抽取+位置信息,但位置信息包含了RoI的高度、宽度和面积,是一个7维向量,Token的处理遵从BERT(Bidirectional Encoder Representations from Transformer)模型的设置。为了让同一个Transformer能够处理表情包图像和表情包文本两种模态,将两种模态的表示投影到同一个表示空间中,所以在原本的RoI和Token表示之上又加了一个线性层。

[0058] 预训练UNITER模型均通过原始UNITER模型训练而成的,具体的,训练过程如下:获取训练集,训练集中包括有若干个训练表情包,如1万个表情包,每个表情包均具有正确的表情包训练文本、表情包训练图像、训练标注文本。在训练原始UNITER模型时,用随机种子初始化原始UNITER模型的输出层的参数,从训练集中随机获取多个表情包进入原始UNITER模型中进行迭代训练,如一次输入100个训练表情包进行迭代训练,1万个训练表情包可进行100次迭代训练,100次迭代训练即完成一次EPOCH的训练,完成20次EPOCH的训练即训练完成,得到一个预训练UNITER模型。其他的预训练UNITER模型经过同样的训练流程进行训练,由于通过随机种子初始化原始UNITER模型的输出层的参数,使得经过同一训练集训练出的预训练UNITER模型是不一样的。

[0059] 通过将标注文本、表情包文本和表情包图像分别输入多个预训练UNITER模型中进行处理,得到对应的输出结果,使得目标输出结果经过三个输入确定的,不会过分的依赖表情包图像,输出结果平衡了标注文本和表情包文本,使得基于目标输出结果去对待检测表情包进行分类能够更加准确。具体的,可计算多个目标输出结果的平均值得到目标分类值。根据目标分类值去确定待检测表情包是否为恶意表情包。将目标分类值与预设分类值进行比较,预设分类值可由用户根据经验自行进行设置,如预设分类值设置为0.5,当目标分类值大于0.5时,表明待检测表情包为恶意表情包。

[0060] 本实施例中,预训练UNITER模型可以应对成分复杂的表情包信息,在预训练UNITER模型中引入表情包图像、表情包文本和标注文本,可使模型更有效地利用图像信息和文本信息,使其表现优于纯依赖文本或图像的模式,能够更加准确的确定恶意表情包。

[0061] 在一实施例中,所述提取所述待检测表情包中的表情包图像和表情包文本,并基

于所述表情包图像和所述表情包文本生成包含所述表情包图像和所述表情包文本的关联性的标注文本的步骤S2,包括:

[0062] 步骤S21,通过pixelLink模型检测所述待检测表情包上的文本区域;

[0063] 步骤S22,通过fasterRCNN模型在所述文本区域识别所述表情包文本;

[0064] 步骤S23,通过mask-rcnn模型或yolo v5模型提取所述表情包图像;

[0065] 步骤S24,将所述表情包文本和所述表情包图像输入multi-head attention模型中进行信息融合,得到所述标注文本。

[0066] 本实施例中,如上述步骤S21所述,pixelLink(像素连接)模型主要是基于CNN网络,做某个像素(pixel)的文本/非文本的分类预测,以及该像素的8个邻域方向是否存在连接(link)的分类预测。然后基于OpenCV的minAreaRect(最小外接矩形)这种基于连通域的操作,获取不同大小的文本连通域,再进行噪声滤除操作,最后通过“并查集”(disjoint-set data structure)并出最终的文本边界框。通过pixelLink模型能够准备的检测到待检测表情包上的表情包文本所在的表情包区域。

[0067] 如上述步骤S22所述,fasterRCNN模型将特征抽取(feature extraction)、proposal提取、bounding box regression(rect refine)、classification都整合在了一个网络中,使得整个模型的综合性能有较大提高,能够快速识别出表情包文本。

[0068] 如上述步骤S23-S24所述,通过mask-rcnn模型或yolo v5模型对待检测表情包中的表情包图像进行特征提取,如山水、物品、颜色等等。将提取到表情包文本和表情包图像输入到multi-head attention(多头注意力)模型中进行信息融合,得到标注文本,multi-head attention模型能够注意到不同的方面,如表情包图像与表情包文本之间的关联含义,根据multi-head attention模型能够准确的对待检测表情包进行准确分类,获取相应的标签,形成标注文本。

[0069] 在一实施例中,各所述预训练UNITER模型的输出层的神经元均为N个,所述N大于等于2000,所述将所述标注文本、所述表情包图像和所述表情包文本输入多个预训练UNITER模型中进行融合处理,得到多个目标输出结果的步骤S3中,任意一个所述预训练UNITER模型得到对应的目标输出结果的过程,包括:

[0070] 将所述标注文本输入所述预训练UNITER模型中进行处理,得到N*1的矩阵作为第一标注文本输出结果;

[0071] 将所述表情包图像输入所述预训练UNITER模型中进行处理,得到N*1的矩阵作为第一表情包图像输出结果;

[0072] 将所述表情包文本输入所述预训练UNITER模型中进行处理,得到N*1的矩阵作为第一表情包文本输出结果;

[0073] 步骤S34,将所述第一标注文本输出结果、所述第一表情包图像输出结果和所述第一表情包文本输出结果进行融合处理,得到所述目标输出结果。

[0074] 本实施例中,将原始UNITER模型的输出层的神经元修改为N个再进行训练,N大于等于2000个,每一个维度均代表某一方面的信息,例如某个局部信息、山水等等,分别处理标注文本、表情包文本和表情包图像,处理得到的第一标注文本输出结果、第一表情包图像输出结果和第一表情包文本输出结果均为一个N*1的矩阵。再将三个矩阵进行融合处理,使得目标输出结果能够学习到三个方面,从而更加准确的对待检测表情包进行识别。多个的

预训练UNITER模型按照第一预训练UNITER模型、第二预训练UNITER模型、第三预训练UNITER模型的规律依次进行命名,第一预训练UNITER模型得到第一目标输出结果、第二预训练UNITER模型得到第二目标输出结果,依次进行命名。

[0075] 在一实施例中,所述将所述第一标注文本输出结果、所述第一表情包图像输出结果和所述第一表情包文本输出结果进行融合处理,得到所述目标输出结果的步骤S34,包括:

[0076] 步骤S341,通过公式 $L = (v + y)(x + v)^T$ 对所述第一标注文本输出结果、所述第一表情包图像输出结果和所述第一表情包文本输出结果进行融合处理,得到第一目标结果;其中,所述 v 为所述第一表情包图像输出结果,所述 y 为所述第一表情包文本输出结果,所述 x 为所述第一标注文本输出结果,所述 T 为转置;

[0077] 步骤S342,将所述第一目标结果经过归一化处理,得到所述目标输出结果。

[0078] 本实施例中,为了使对待检测表情包的识别结果不会过分依赖于图像结果,本申请平衡了两个文本结果,通过公式 $L = (v + y)(x + v)^T$ 对表情包图像和表情包文本及生成的标注文本进行拟合,随后对 L 进行归一化得到最后的目标输出结果为一个2分类的值,即0和1。同样的,其他的目标输出结果等根据同样的处理得到。

[0079] 在一实施例中,所述将所述目标分类值与预设分类值进行比较;若所述目标分类值大于所述预设分类值,则确定所述待检测表情包为恶意表情包的步骤S4,包括:

[0080] 步骤S4A,计算所述标注文本和所述表情包的相似度;

[0081] 步骤S4B,将所述相似度与预设阈值进行比较;

[0082] 步骤S4C,若所述相似度小于所述预设阈值,在所述目标分类值上增加预设惩罚值作为新的目标分类值;

[0083] 步骤S4D,将所述新的目标分类值与所述预设分类值进行比较;

[0084] 步骤S4E,若所述新的目标分类值大于所述预设分类值,则确定所述待检测表情包为恶意表情包。

[0085] 本实施例中,根据生活经验可知,当表情包的文本信息与图片信息含义完全相反时,则有很高的概率此表情包是恶意的。因此,在得到目标分类值之后,计算标注文本和表情包文本的相似度,并将相似度与预设阈值进行比较,遇着阈值自行进行设置,如预设阈值设置为-0.8,仅当小于-0.8时才能确认文本信息和图片信息的含义基本相反,-1则代表完全相反。当相似度小于-0.8时,在目标分类值上增加一个预设惩罚值,如预设惩罚值设置为0.5,将增加后的结果作为新的目标分类值去与愈合分类值进行比较,当新的目标分类值大于预设分类值时,则确定待检测表情包为恶意表情包。预设惩罚值等于0.5是一个强机制,当它触发时,待检测表情包应该被直接分类为恶意。唯一的例外是当目标分类值是0时,表示其中第一输出结果、第二输出结果和第三输出结果都是0,新的目标分类值 $0+0.5 \leq 0.5$,确定该待检测表情包为非恶意表情包。本实施例中,在目标分类值的结果上再附加预设惩罚值,能够更加有效地应对善意反转等复杂情况。

[0086] 在一实施例中,所述计算所述标注文本和所述表情包的相似度的步骤S4A,包括:

[0087] 步骤S4A1,通过word2vec模型分别得到所述标注文本和所述表情包的词向

量;

[0088] 步骤S4A2,根据标注文本和所述表情包文本的词向量计算所述标注文本和所述表情包文本的余弦相似度。

[0089] 本实施例中,使用word2vec分别得出标注文本与表情包文本的词向量(vector representations),然后计算他们的余弦相似度,余弦相似度用向量空间中两个向量夹角的余弦值作为衡量两个个体间差异的大小。余弦值越接近1,就表明夹角越接近0度,也就是两个向量越相似。在其他实施例中,可计算标注文本和表情包文本的欧几里得距离、曼哈顿距离等作为两者之间的相似度。

[0090] 在一实施例中,所述通过预设规则对多个所述目标输出结果进行计算得到目标分类值的步骤,包括:

[0091] 根据公式 $S=k_1a_1+k_2a_2+\dots+k_na_n$ 得到所述目标分类值;其中,所述k为各个所述目标输出结果的预设权重,n为所述预训练UNITER模型的个数,所述a为所述目标输出结果, $k_1+k_2+\dots+k_n=1$ 。

[0092] 本实施例中,对每个目标输出结果都设置有一个预设权重,使得目标分类值能够侧重不同的目标输出结果。如预训练UNITER模型包括3个,分别得到第一目标输出结果 a_1 、第二目标输出结果 a_2 和第三目标输出结果 a_3 ,分别对第一输出结果、第二输出结果和第三输出结果设置了一个预设权重 k_1 、 k_2 、 k_3 ,根据公式 $S=k_1a_1+k_2a_2+k_3a_3$ 得到所述目标分类值,根据权重侧重不同的输出结果,能够更加准确的得到目标分类结果。

[0093] 参见图2,本申请一实施例提供了一种恶意表情包确定装置,包括:

[0094] 获取单元10,用于获取待检测表情包;

[0095] 提取单元20,用于提取所述待检测表情包中的表情包图像和表情包文本,并基于所述表情包图像和所述表情包文本生成包含所述表情包图像和所述表情包文本的关联性的标注文本;

[0096] 处理单元30,用于将所述标注文本、所述表情包图像和所述表情包文本输入多个预训练UNITER模型中进行融合处理,得到多个目标输出结果;其中,多个所述预训练UNITER模型通过随机种子初始化原始UNITER模型的输出层后进行训练得到的;

[0097] 目标分类单元40,用于通过预设规则对多个所述目标输出结果进行计算得到目标分类值;

[0098] 分类单元50,用于将所述目标分类值与预设分类值进行比较;若所述目标分类值大于所述预设分类值,则确定所述待检测表情包为恶意表情包。

[0099] 在一实施例中,所述提取单元20,包括:

[0100] 检测子单元,用于通过pixelLink模型检测所述待检测表情包上的文本区域;

[0101] 识别子单元,用于通过fasterRCNN模型在所述文本区域识别所述表情包文本;

[0102] 提取子单元,用于通过mask-rcnn模型或yolo v5模型提取所述表情包图像;

[0103] 融合子单元,用于将所述表情包文本和所述表情包图像输入multi-head attention模型中进行信息融合,得到所述标注文本。

[0104] 在一实施例中,所述预训练UNITER模型的输出层的神经元为N个,所述N大于等于2000,所述处理单元30中,包括:

[0105] 第一处理子单元,用于将所述标注文本输入所述预训练UNITER模型中进行处理,

得到 $N \times 1$ 的矩阵作为第一标注文本输出结果；

[0106] 第二处理子单元,用于将所述表情包图像文本输入所述预训练UNITER模型中进行处理,得到 $N \times 1$ 的矩阵作为第一表情包图像输出结果；

[0107] 第三处理子单元,用于将所述表情包文本输入所述预训练UNITER模型中进行处理,得到 $N \times 1$ 的矩阵作为第一表情包文本输出结果；

[0108] 融合处理子单元,用于将所述第一标注文本输出结果、所述第一表情包图像输出结果和所述第一表情包文本输出结果进行融合处理,得到所述目标输出结果。

[0109] 在一实施例中,所述融合处理子单元,包括:

[0110] 融合处理模块,用于通过公式 $L = (v + y)(x + v)^T$ 对所述第一标注文本输出结果、所述第一表情包图像输出结果和所述第一表情包文本输出结果进行融合处理,得到第一目标结果;其中,所述 v 为所述第一表情包图像输出结果,所述 y 为所述第一表情包文本输出结果,所述 x 为所述第一标注文本输出结果,所述 T 为转置;

[0111] 归一化处理模块,用于将所述第一目标结果经过归一化处理,得到所述目标输出结果。

[0112] 在一实施例中,所述分类单元50,包括:

[0113] 计算子单元,用于计算所述标注文本和所述表情包文本的相似度;

[0114] 第一比较子单元,用于将所述相似度与预设阈值进行比较;

[0115] 增加子单元,用于若所述相似度小于所述预设阈值,在所述目标分类值上增加预设惩罚值作为新的目标分类值;

[0116] 第二比较子单元,用于将所述新的目标分类值与所述预设分类值进行比较;

[0117] 确定子单元,用于若所述新的目标分类值大于所述预设分类值,则确定所述待检测表情包为恶意表情包。

[0118] 在一实施例中,所述计算子单元,包括:

[0119] 词向量模块,用于通过word2vec模型分别得到所述标注文本和所述表情包文本的词向量;

[0120] 计算模块,用于根据标注文本和所述表情包文本的词向量计算所述标注文本和所述表情包文本的余弦相似度。

[0121] 在一实施例中,所述目标分类单元40,包括:

[0122] 目标分类子单元,用于根据公式 $S = k_1 a_1 + k_2 a_2 + \dots + k_n a_n$ 得到所述目标分类值;其中,所述 k 为各个所述目标输出结果的预设权重, n 为所述预训练UNITER模型的个数,所述 a 为所述目标输出结果, $k_1 + k_2 + \dots + k_n = 1$ 。

[0123] 在本实施例中,上述各个单元、子单元、模块的具体实现请参照上述方法实施例中所述,在此不再进行赘述。

[0124] 参照图3,本申请实施例中还提供一种计算机设备,该计算机设备可以是服务器,其内部结构可以如图3所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中,该计算机设计的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机程序和数据库。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该计算机设备的数据库用于存储待检测表情包等。该计算机设备的网络接口用于与外部的

终端通过网络连接通信。该计算机程序被处理器执行时以实现一种表情包分类方法。

[0125] 本领域技术人员可以理解,图3中示出的结构,仅仅是与本申请方案相关的部分结构的框图,并不构成对本申请方案所应用于其上的计算机设备的限定。

[0126] 本申请一实施例还提供一种计算机可读存储介质,其上存储有计算机程序,计算机程序被处理器执行时实现一种表情包分类方法。

[0127] 综上所述,为本申请实施例中提供的表情包分类方法、装置、计算机设备和存储介质,获取待检测表情包;提取所述待检测表情包中的表情包图像和表情包文本,并基于所述表情包图像和所述表情包文本生成包含所述表情包图像和所述表情包文本的关联性的标注文本;将所述标注文本、所述表情包图像和所述表情包文本输入多个预训练UNITER模型中进行融合处理,得到多个目标输出结果;其中,多个所述预训练UNITER模型通过随机种子初始化原始UNITER模型的输出层后进行训练得到的;通过预设规则对多个所述目标输出结果进行计算得到目标分类值;将所述目标分类值与预设分类值进行比较;若所述目标分类值大于所述预设分类值,则确定所述待检测表情包为恶意表情包。本申请提供的表情包分类方法、装置、计算机设备和存储介质,预训练UNITER模型可以应对成分复杂的表情包信息,在预训练UNITER模型中引入表情包图像、表情包文本和标注文本,可使模型更有效地利用图像信息和文本信息,使其表现优于纯依赖文本或图像模型,能够更加准确的对表情包进行分类。

[0128] 本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程,是可以通过计算机程序来指令相关的硬件来完成,所述的计算机程序可存储于一非易失性计算机可读存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本申请所提供的和实施例中所使用的对存储器、存储、数据库或其它介质的任何引用,均可包括非易失性和/或易失性存储器。非易失性存储器可以包括只读存储器(ROM)、可编程ROM(PROM)、电可编程ROM(EPROM)、电可擦除可编程ROM(EEPROM)或闪存。易失性存储器可包括随机存取存储器(RAM)或者外部高速缓冲存储器。作为说明而非局限,RAM通过多种形式可得,诸如静态RAM(SRAM)、动态RAM(DRAM)、同步DRAM(SDRAM)、双速据率SDRAM(SSRSDRAM)、增强型SDRAM(ESDRAM)、同步链路(Synchlink)DRAM(SLDRAM)、存储器总线(Rambus)直接RAM(RDRAM)、直接存储器总线动态RAM(DRDRAM)、以及存储器总线动态RAM(RDRAM)等。

[0129] 需要说明的是,在本文中,术语“包括”、“包含”或者其任何其它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、装置、物品或者方法不仅包括那些要素,而且还包括没有明确列出的其它要素,或者是还包括为这种过程、装置、物品或者方法所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括该要素的过程、装置、物品或者方法中还存在另外的相同要素。

[0130] 以上所述仅为本申请的优选实施例,并非因此限制本申请的专利范围,凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换,或直接或间接运用在其它相关的技术领域,均同理包括在本申请的专利保护范围内。

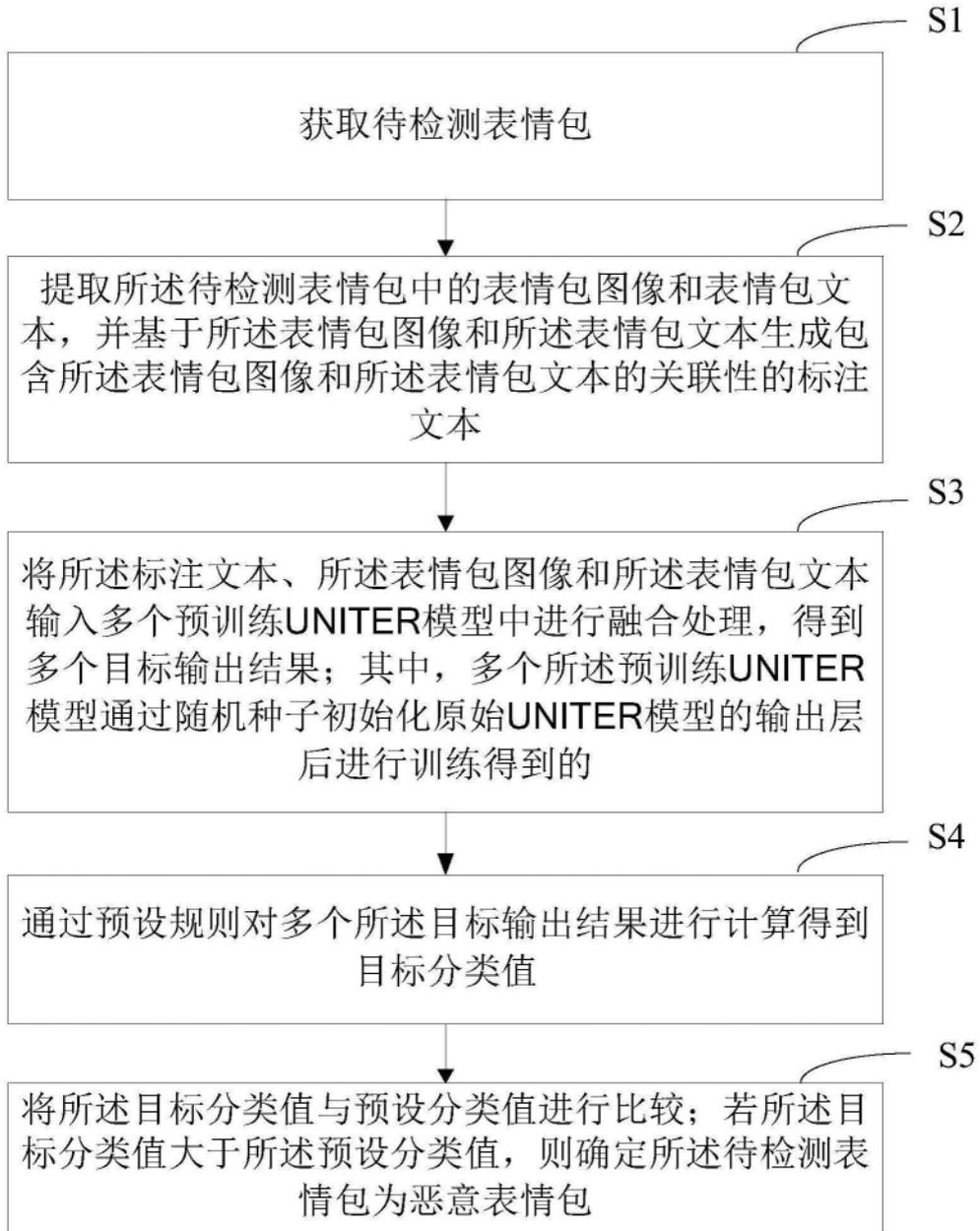


图1

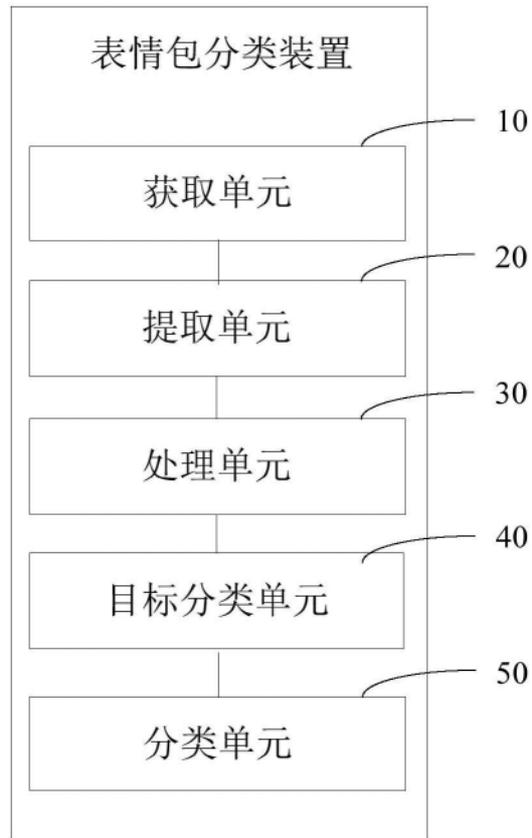


图2

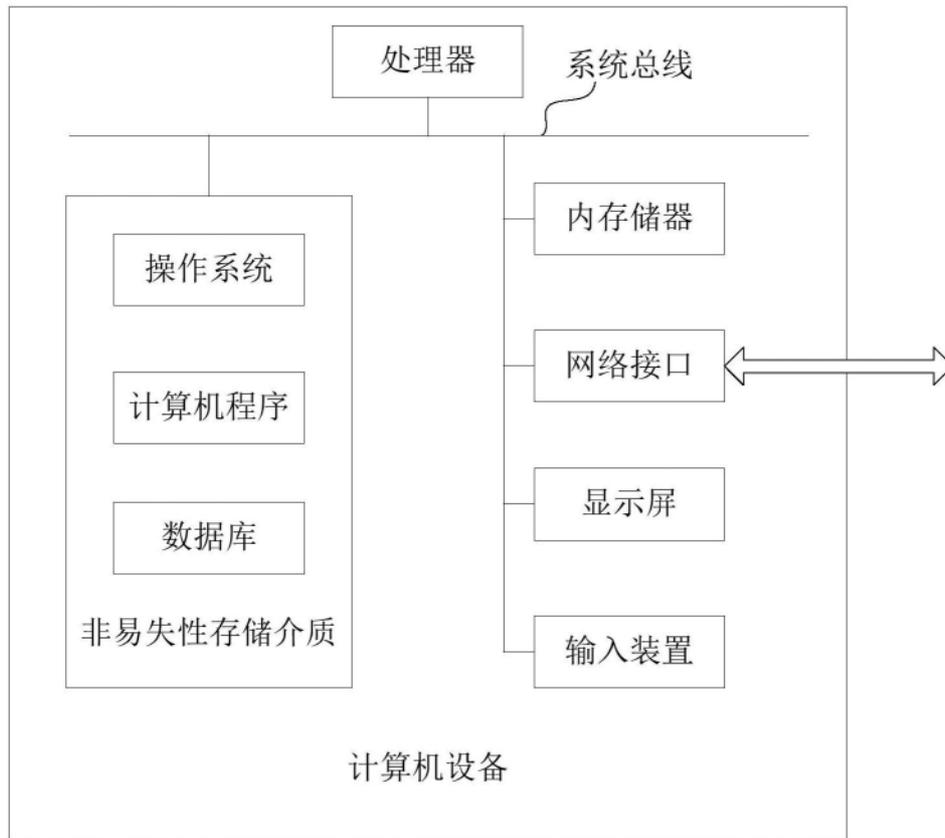


图3