



(12) 发明专利申请

(10) 申请公布号 CN 113947702 A

(43) 申请公布日 2022. 01. 18

(21) 申请号 202111080047.X

(22) 申请日 2021.09.15

(71) 申请人 复旦大学

地址 200433 上海市杨浦区邯郸路220号

(72) 发明人 张立华 杨鼎康 王顺利 卞昊鹏
黄帅

(74) 专利代理机构 上海科盛知识产权代理有限公司 31225

代理人 赵继明

(51) Int. Cl.

G06V 10/80 (2022.01)

G06V 10/764 (2022.01)

G06V 10/82 (2022.01)

G06V 40/16 (2022.01)

G06V 40/10 (2022.01)

G06V 40/20 (2022.01)

G06K 9/62 (2022.01)

G10L 25/30 (2013.01)

G10L 25/63 (2013.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

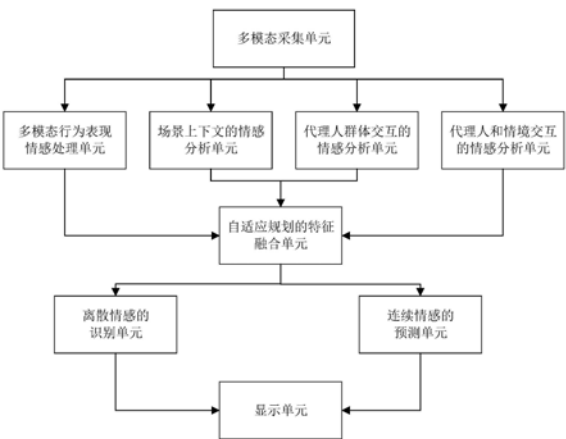
权利要求书2页 说明书9页 附图8页

(54) 发明名称

一种基于情境感知的多模态情感识别方法和系统

(57) 摘要

本发明涉及一种基于情境感知的多模态情感识别方法和系统,系统包括多模态信息采集单元、基于多模态行为表现的情感处理单元、基于场景上下文的情感分析单元、基于代理人群体交互的情感分析单元、基于代理人和情境交互的情感分析单元、基于自适应规划的特征融合单元、基于离散情感的识别单元、基于连续情感的预测单元以及显示模块,多模态数据一部分来源于面部表情、步态和手势信息;另一部分来自情境中的场景上下文、代理人群体互动上下文以及场景与代理人互动的上下文信息。与现有技术相比,本发明有效的解决了真实场景中情感识别效率低下,现有算法情感识别准确率不达标,识别算法鲁棒性和泛化能力受外界因素干扰的问题。



1. 一种基于情境感知的多模态情感识别方法,其特征在于,包括以下步骤:

多模态信息采集步骤:采集用于情感识别的视频和真实世界图像,所述视频中包括待进行情感识别的人类主体和其它代理人;

基于多模态行为表现的情感处理步骤:根据所述视频,提取面部表情特征、面部界标点、人体姿态和人体手势的特征向量,并通过初始特征级联的方式生成外在行为表现模态向量;

基于场景上下文的情感分析步骤:通过为所述真实世界图像和视频各视频帧中的人类主体添加蒙版,获取场景图像,然后进行场景情感语义的特征提取,得到第一情感特征向量;

基于代理人群体交互的情感分析步骤:在所述真实世界图像和视频各视频帧中分别提取人类主体和其它代理人信息,然后提取初始表征特征,将各初始表征特征作为情感节点馈送入图注意力网络中,构建情感关系图;根据该情感关系图,计算不同的其它代理人对人类主体的情感影响强度和程度,通过情感相似系数判断其它代理人交互产生的情感特征向量的权重大小,并与初始表征特征进行加权平均操作,得到第二情感特征向量;

基于代理人 and 情境交互的情感分析步骤:通过为所述真实世界图像和视频各视频帧中的其它代理人添加蒙版,获取场景图像,提取场景的初始特征;根据其它代理人的初始表征特征,建立基本特征图,对所述场景的初始特征和基本特征图进行特征聚合,得到第三情感特征向量;

特征融合步骤:对所述外在行为表现模态向量、第一情感特征向量、第二情感特征向量和第三情感特征向量进行特征融合,得到融合特征向量;

情感识别步骤:根据所述融合特征向量进行情感识别。

2. 根据权利要求1所述的一种基于情境感知的多模态情感识别方法,其特征在于,所述基于场景上下文的情感分析步骤中,进行场景情感语义的特征提取具体为:选取残差神经网络作为主体模型的骨干网络,所述残差神经网络中的多个残差连接块按顺序交替嵌入基于通道和空间的注意力机制模块形成完整的注意力提取网络,将所述场景图像载入该注意力提取网络中进行场景情感语义的特征提取。

3. 根据权利要求2所述的一种基于情境感知的多模态情感识别方法,其特征在于,所述基于通道和空间的注意力机制模块包括通道注意力机制和空间注意力机制,所述通道注意力机制包括:通过全局平均池化推理出一张1D的通道注意力图 $M_{avg} \in \mathbb{R}^{c \times 1 \times 1}$,随后在输出层通过通道级别的乘法进行特征合并;

所述空间注意力机制包括:通过全局最大池化层推理出一张2D的空间注意力图 $M_{max} \in \mathbb{R}^{1 \times w \times h}$,随后在输出层通过通道级别的乘法进行特征合并。

4. 根据权利要求1所述的一种基于情境感知的多模态情感识别方法,其特征在于,所述特征融合步骤中进行特征融合具体为:

从所述外在行为表现模态向量、第一情感特征向量、第二情感特征向量和第三情感特征向量中选取强相关的特征向量和弱相关的特征向量,对于强相关的特征向量通过特征级联操作进行特征融合,对于弱相关的特征向量通过乘性融合方式进行特征融合。

5. 根据权利要求1所述的一种基于情境感知的多模态情感识别方法,其特征在于,所述

情感识别步骤具体包括离散情感识别子步骤和连续情感预测子步骤；

所述离散情感识别子步骤包括：将所述融合特征向量映射到0到1之间，然后对每个输出节点和对应的标签计算交叉熵损失函数，通过计算出每一类可能输出的表情标签的概率来预测得到的表情标签。

6. 根据权利要求5所述的一种基于情境感知的多模态情感识别方法，其特征在于，所述连续情感预测子步骤包括依次进行数据归一化、标签差值求和、误差幅度计算和连续数值预测，所述连续情感预测子步骤通过预先构建并训练好的网络模型实现，该网络模型采用均方误差损失来计算预测数值和目标数值之间的差值的平方和从而进行网络模型的训练。

7. 根据权利要求5所述的一种基于情境感知的多模态情感识别方法，其特征在于，所述离散情感识别子步骤中的表情标签包括高兴、惊喜、悲伤、厌恶、兴奋、平和、恐惧和生气；

所述连续情感预测子步骤的输出为情感的VAD连续模型的1到10中的预测值，该VAD连续模型指的是情感的唤醒度、控制度以及愉悦度。

8. 根据权利要求1所述的一种基于情境感知的多模态情感识别方法，其特征在于，所述基于多模态行为表现的情感处理步骤中，通过面部检测器提取面部表情轮廓，随后通过设计的卷积神经网络进行特征提取操作，得到面部表情特征向量；

通过面部检测器提取多个面部界标点，通过卷积神经网络获取并转换为情感特征向量；

通过姿态检测器提取多个人体姿态的坐标点，将其馈送到编码器网络进行特征提取，得到人体姿态的特征向量；

通过姿态检测器提取人体手部表征关键点，利用卷积神经网络获取人体手势的特征向量。

9. 一种采用如权利要求1-8任一所述的一种基于情境感知的多模态情感识别方法的系统，其特征在于，包括：

多模态信息采集单元，被配置为执行所述多模态信息采集步骤；

基于多模态行为表现的情感处理单元，被配置为执行所述基于多模态行为表现的情感处理步骤；

基于场景上下文的情感分析单元，被配置为执行所述基于场景上下文的情感分析步骤；

基于代理人群体交互的情感分析单元，被配置为执行所述基于代理人群体交互的情感分析步骤；

基于代理人和情境交互的情感分析单元，被配置为执行所述基于代理人和情境交互的情感分析步骤；

基于自适应规划的特征融合单元，被配置为执行所述特征融合步骤；

情感识别单元，被配置为执行所述情感识别步骤。

10. 根据权利要求9所述的系统，其特征在于，所述系统还包括显示模块，该显示模块被配置为线束所述情感识别单元的输出结果。

一种基于情境感知的多模态情感识别方法和系统

技术领域

[0001] 本发明涉及情感识别技术领域,尤其是涉及一种基于情境感知的多模态情感识别方法和系统。

背景技术

[0002] 情感识别作为人机情感交互的基础,能够使机器理解人的感性思维,影响着机器智能化的继续发展,成为人机自然交互的关键要素。近年来,基于多模态的情感识别技术越来越受到研究者的关注。该方法受到情感心理学研究的推动,旨在融合诸如面部表情,语音,身体手势以及步态等多种情感信号,通过多样的融合模式以提高情感识别的准确性和精度。

[0003] 情境感知作为计算机视觉领域目前的热点研究,对于在真实场景中理解人类情感具有举足轻重的作用。通常人类所处的情境上下文中蕴含着丰富的语义信息,通过深度学习的方式在不同的情境上下文中感知人类的情感,利用从中抽取的多模态情感特征进行特征级别和决策级别的融合,往往可以获得人类主体之外更多的情感线索,以促进情绪表达和情感理解。

[0004] 真实环境下的情境感知技术被用于多模态的情感识别是极具研究价值的新领域,目前尚未出现相关基于情景感知的多模态情感识别的专利方法,大部分采用传统的使用单一模态的情感识别方法。目前的大多数工作基于深度学习的网络体系结构实现。早期的研究工作旨在结合面部表情的直观情感表达和整体的情境上下文信息实现情感识别;随后的工作在此基础上利用区域提议网络(RPN)从样本中提取上下文元素,将其作为情感图的节点,馈送入图卷积神经网络(GCN)以对上下文信息进行编码,最终基于特征级联的方式实现多模态情感识别。最近的工作中,部分研究者将面部以外的所有信息视为上下文,通过从图像中屏蔽人类的面部表情从而提取上下文情感表达。还有一些著名的工作从心理学角度入手,基于注意力机制学习情境中情感特征的表示后,结合热力图的方式探索多人交互情境下人类主体之间的接近程度和距离,以挖掘群体间的情感上下文信息,从而提高了情感识别的准确率。

[0005] 现有的基于情境感知的多模态情感识别方法,往往仅针对图像和视频中采集的完整上下文语义信息进行编码和特征提取,这些方法的缺点是没有考虑到背景环境中其他主体的情感流露变化对情感识别主体的情感预测的干扰,大致准确率大大降低;同时针对多人交互情境中情感信息的编码方式简单,通常使用图卷积网络或者热力图(heat map)的形式进行建模分析,这类方法将多人之间的高维情感距离视为不变的常量,且只能粗略的度量情感张力变化,导致很难捕获到与主体情感变化有关的情境上下文表征,缺乏合理性。同时,情境中人类主体和环境的交互上下文变化很少被考虑到,这导致大多数多模态情感特征的表达出现冗余和错误,情感识别模型的鲁棒性无法获得保障。

[0006] 目前的多模态形式往往专注于对情境中除去人类主体之外的情感线索提取,而忽略了人类外在情感行为表现特征的利用。例如将面部表情,步态以及手势等与情感信息流

露密切相关的模态信号相结合。此外,传统的基于离散情绪模型的情感定义模式无法科学有效的描述情感变化的本质,导致情感识别结果的评估和分析缺乏有效性。

[0007] 综上所述,开发新型基于情境感知的方法,充分利用面部表情,步态,以及手势等人体外在行为表现模态,结合对人类主体和场景以及人类群体之间情感交互行为进行建模分析的多模态情感识别系统成为了本研究领域技术人员亟待解决的问题。

发明内容

[0008] 本发明的目的就是为了克服上述现有技术存在的缺陷而提供一种充分利用面部表情,步态,以及手势等人体外在行为表现模态,结合对人类主体和场景以及人类群体之间情感交互行为进行建模分析的基于情境感知的多模态情感识别方法和系统。

[0009] 本发明的目的可以通过以下技术方案来实现:

[0010] 一种基于情境感知的多模态情感识别方法,包括以下步骤:

[0011] 多模态信息采集步骤:采集用于情感识别的视频和真实世界图像,所述视频中包括待进行情感识别的人类主体和其它代理人;

[0012] 基于多模态行为表现的情感处理步骤:根据所述视频,提取面部表情特征、面部界标点、人体姿态和人体手势的特征向量,并通过初始特征级联的方式生成外在行为表现模态向量;

[0013] 基于场景上下文的情感分析步骤:通过为所述真实世界图像和视频各视频帧中的人类主体添加蒙版,获取场景图像,然后进行场景情感语义的特征提取,得到第一情感特征向量;

[0014] 基于代理人群体交互的情感分析步骤:在所述真实世界图像和视频各视频帧中分别提取人类主体和其它代理人信息,然后提取初始表征特征,将各初始表征特征作为情感节点馈送入图注意力网络中,构建情感关系图;根据该情感关系图,计算不同的其它代理人对人类主体的情感影响强度和程度,通过情感相似系数判断其它代理人交互产生的情感特征向量的权重大小,并与初始表征特征进行加权平均操作,得到第二情感特征向量;

[0015] 基于代理人和情境交互的情感分析步骤:通过为所述真实世界图像和视频各视频帧中的其它代理人添加蒙版,获取场景图像,提取场景的初始特征;根据其它代理人的初始表征特征,建立基本特征图,对所述场景的初始特征和基本特征图进行特征聚合,得到第三情感特征向量;

[0016] 特征融合步骤:对所述外在行为表现模态向量、第一情感特征向量、第二情感特征向量和第三情感特征向量进行特征融合,得到融合特征向量;

[0017] 情感识别步骤:根据所述融合特征向量进行情感识别。

[0018] 进一步地,所述基于场景上下文的情感分析步骤中,进行场景情感语义的特征提取具体为:选取残差神经网络作为主体模型的骨干网络,所述残差神经网络中的多个残差连接块按顺序交替嵌入基于通道和空间的注意力机制模块形成完整的注意力提取网络,将所述场景图像载入该注意力提取网络中进行场景情感语义的特征提取。

[0019] 进一步地,所述基于通道和空间的注意力机制模块包括通道注意力机制和空间注意力机制,所述通道注意力机制包括:通过全局平均池化推理出一张1D的通道注意力图 $M_{avg} \in \mathbb{R}^{c \times 1 \times 1}$,随后在输出层通过通道级别的乘法进行特征合并;

[0020] 所述空间注意力机制包括:通过全局最大池化层推理出一张2D的空间注意力图 $M_{\max} \in \mathbb{R}^{1 \times w \times h}$,随后在输出层通过通道级别的乘法进行特征合并。

[0021] 进一步地,所述特征融合步骤中进行特征融合具体为:

[0022] 从所述外在行为表现模态向量、第一情感特征向量、第二情感特征向量和第三情感特征向量中选取强相关的特征向量和弱相关的特征向量,对于强相关的特征向量通过特征级联操作进行特征融合,对于弱相关的特征向量通过乘性融合方式进行特征融合。

[0023] 进一步地,所述情感识别步骤具体包括离散情感识别子步骤和连续情感预测子步骤;

[0024] 所述离散情感识别子步骤包括:将所述融合特征向量映射到0到1之间,然后对每个输出节点和对应的标签计算交叉熵损失函数,通过计算出每一类可能输出的表情标签的概率来预测得到的表情标签。

[0025] 进一步地,所述连续情感预测子步骤包括依次进行数据归一化、标签差值求和、误差幅度计算和连续数值预测,所述连续情感预测子步骤通过预先构建并训练好的网络模型实现,该网络模型采用均方误差损失来计算预测数值和目标数值之间的差值的平方和从而进行网络模型的训练。

[0026] 进一步地,所述离散情感识别子步骤中的表情标签包括高兴、惊喜、悲伤、厌恶、兴奋、平和、恐惧和生气;

[0027] 所述连续情感预测子步骤的输出为情感的VAD连续模型的1到10中的预测值,该VAD连续模型指的是情感的唤醒度、控制度以及愉悦度。

[0028] 进一步地,所述基于多模态行为表现的情感处理步骤中,通过面部检测器提取面部表情轮廓,随后通过设计的卷积神经网络进行特征提取操作,得到面部表情特征向量;

[0029] 通过面部检测器提取多个面部界标点,通过卷积神经网络获取并转换为情感特征向量;

[0030] 通过姿态检测器提取多个人体姿态的坐标点,将其馈送到编码器网络进行特征提取,得到人体姿态的特征向量;

[0031] 通过姿态检测器提取人体手部表征关键点,利用卷积神经网络获取人体手势的特征向量。

[0032] 本发明还提供一种采用如上所述的一种基于情境感知的多模态情感识别方法的系统,包括:

[0033] 多模态信息采集单元,被配置为执行所述多模态信息采集步骤;

[0034] 基于多模态行为表现的情感处理单元,被配置为执行所述基于多模态行为表现的情感处理步骤;

[0035] 基于场景上下文的情感分析单元,被配置为执行所述基于场景上下文的情感分析步骤;

[0036] 基于代理人群体交互的情感分析单元,被配置为执行所述基于代理人群体交互的情感分析步骤;

[0037] 基于代理人和情境交互的情感分析单元,被配置为执行所述基于代理人和情境交互的情感分析步骤;

[0038] 基于自适应规划的特征融合单元,被配置为执行所述特征融合步骤;

[0039] 情感识别单元,被配置为执行所述情感识别步骤。

[0040] 进一步地,所述系统还包括显示模块,该显示模块被配置为线束所述情感识别单元的输出结果。

[0041] 与现有技术相比,本发明具有以下优点:

[0042] (1) 区别于传统的多模态情感识别方法,本发明提出了基于情境感知的情感理解和推理模式,试图通过情感识别主体之外的多模态语义辅助进行情感判别和分析。具体的,场景上下文的情感分析单元通过提取真实世界中蕴含的事物和背景环境中的上下文情感语义信息,从而强化情感外在表征,提高情感识别的判别能力;代理人群体交互的情感分析单元通过分析情感识别主体和周围其他代理人之间的情感传递关系,利用先进的图注意力神经网络分析不同代理人之间的情感强度,以辅助和增强识别主体的情感表征能力;代理人 and 情境交互的情感分析单元致力于挖掘其他代理人在场景中进行社会活动所触发的隐藏情感状态,通过特征聚合的形式以完备识别主体的情感表达空间。

[0043] (2) 基于多模态行为表现的情感处理单元充分结合了情感识别主体的面部表情信息,面部关键点信息、手势信号以及姿态信号等外在情感信息,所获取的基于外在行为表现的多模态信息有效的解决了生活中的遮挡和传感器噪声导致的部分模态信息缺失和异常时,所带来得系统性能下降问题。同时,面部分析中同时采用面部表情信息和面部关键点信息,通过先进的卷积神经网络进行特征提取和融合,最大化强化了外在情感的表征能力。

[0044] (3) 区别于传统特征融合中单一的特征拼接方式,本发明提出的自适应规划融合单元,可以有效的动态规划不同模态的融合策略,自适应的考虑异构模态之间的差异性和相关性,通过乘性融合和特征级联智能选择的方式,充分挖掘潜在的情感特征以及进一步加强显式情感特征的分类和预测能力。

[0045] (4) 本发明提出的基于情感分类和预测的多任务学习模式,能够有效的揭示情感的演化和流露过程,此方法重新定义了情感识别规则,不像是传统模式一样仅仅考虑离散情感的分类,而是结合高维情感空间的离散情感节点和连续空间变化中的情感状态,两者之间的共同训练和学习,有效的促进了多模态情感识别模型的可靠性和准确性。

附图说明

[0046] 图1为本发明实施例中提供的一种基于情境感知的多模态情感识别系统原理框图;

[0047] 图2为本发明实施例中提供的基于多模态行为表现的情感处理单元原理框图;

[0048] 图3为本发明实施例中提供的基于场景上下文的情感分析单元原理框图;

[0049] 图4为本发明实施例中提供的基于代理人群体交互的情感分析单元原理框图;

[0050] 图5为本发明实施例中提供的基于代理人 and 情境交互的情感分析单元原理框图;

[0051] 图6为本发明实施例中提供的基于自适应规划的特征融合单元原理框图;

[0052] 图7为本发明实施例中提供的基于离散情感的识别单元原理框图;

[0053] 图8为本发明实施例中提供的基于连续情感的预测单元原理框图。

具体实施方式

[0054] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合本发明实施例

中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本发明一部分实施例,而不是全部的实施例。通常在此处附图中描述和示出的本发明实施例的组件可以以各种不同的配置来布置和设计。

[0055] 因此,以下对在附图中提供的本发明的实施例的详细描述并非旨在限制要求保护的本发明的范围,而是仅仅表示本发明的选定实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0056] 实施例1

[0057] 本实施例提供一种基于情境感知的多模态情感识别方法,包括以下步骤:

[0058] 多模态信息采集步骤:采集用于情感识别的视频和真实世界图像,视频中包括待进行情感识别的人类主体和其它代理人;

[0059] 基于多模态行为表现的情感处理步骤:根据视频,提取面部表情特征、面部界标点、人体姿态和人体手势的特征向量,并通过初始特征级联的方式生成外在行为表现模态向量;

[0060] 基于场景上下文的情感分析步骤:通过为真实世界图像和视频各视频帧中的人类主体添加蒙版,获取场景图像,然后进行场景情感语义的特征提取,得到第一情感特征向量;

[0061] 基于代理人群体交互的情感分析步骤:在真实世界图像和视频各视频帧中分别提取人类主体和其它代理人信息,然后提取初始表征特征,将各初始表征特征作为情感节点馈送入图注意力网络中,构建情感关系图;根据该情感关系图,计算不同的其它代理人对人类主体的情感影响强度和程度,通过情感相似系数判断其它代理人交互产生的情感特征向量的权重大小,并与初始表征特征进行加权平均操作,得到第二情感特征向量;

[0062] 基于代理人 and 情境交互的情感分析步骤:通过为真实世界图像和视频各视频帧中的其它代理人添加蒙版,获取场景图像,提取场景的初始特征;根据其它代理人的初始表征特征,建立基本特征图,对场景的初始特征和基本特征图进行特征聚合,得到第三情感特征向量;

[0063] 特征融合步骤:对外在行为表现模态向量、第一情感特征向量、第二情感特征向量和第三情感特征向量进行特征融合,得到融合特征向量;

[0064] 情感识别步骤:根据融合特征向量进行情感识别。

[0065] 下面对各步骤进行详细描述。

[0066] 1.1、基于多模态行为表现的情感处理步骤

[0067] 基于多模态行为表现的情感处理步骤中,通过面部检测器提取面部表情轮廓,随后通过设计的卷积神经网络进行特征提取操作,得到面部表情特征向量;

[0068] 通过面部检测器提取多个面部界标点,通过卷积神经网络获取并转换为情感特征向量;

[0069] 通过姿态检测器提取多个人体姿态的坐标点,将其馈送到编码器网络进行特征提取,得到人体姿态的特征向量;

[0070] 通过姿态检测器提取人体手部表征关键点,利用卷积神经网络获取人体手势的特征向量。

[0071] 1.2、基于场景上下文的情感分析步骤

[0072] 基于场景上下文的情感分析步骤中,进行场景情感语义的特征提取具体为:选取残差神经网络作为主体模型的骨干网络,残差神经网络中的多个残差连接块按顺序交替嵌入基于通道和空间的注意力机制模块形成完整的注意力提取网络,将场景图像载入该注意力提取网络中进行场景情感语义的特征提取。

[0073] 基于通道和空间的注意力机制模块包括通道注意力机制和空间注意力机制,通道注意力机制包括:通过全局平均池化推理出一张1D的通道注意力图 $M_{avg} \in \mathbb{R}^{c \times 1 \times 1}$,随后在输出层通过通道级别的乘法进行特征合并;

[0074] 空间注意力机制包括:通过全局最大池化层推理出一张2D的空间注意力图 $M_{max} \in \mathbb{R}^{1 \times w \times h}$,随后在输出层通过通道级别的乘法进行特征合并。

[0075] 1.3、特征融合步骤

[0076] 特征融合步骤中进行特征融合具体为:

[0077] 从外在行为表现模态向量、第一情感特征向量、第二情感特征向量和第三情感特征向量中选取强相关的特征向量和弱相关的特征向量,对于强相关的特征向量通过特征级联操作进行特征融合,对于弱相关的特征向量通过乘性融合方式进行特征融合。

[0078] 1.4、情感识别步骤

[0079] 情感识别步骤具体包括离散情感识别子步骤和连续情感预测子步骤;

[0080] 离散情感识别子步骤包括:将融合特征向量映射到0到1之间,然后对每个输出节点和对应的标签计算交叉熵损失函数,通过计算出每一类可能输出的表情标签的概率来预测得到的表情标签。

[0081] 连续情感预测子步骤包括依次进行数据归一化、标签差值求和、误差幅度计算和连续数值预测,连续情感预测子步骤通过预先构建并训练好的网络模型实现,该网络模型采用均方误差损失来计算预测数值和目标数值之间的差值的平方和从而进行网络模型的训练。

[0082] 离散情感识别子步骤中的表情标签包括高兴、惊喜、悲伤、厌恶、兴奋、平和、恐惧和生气;

[0083] 连续情感预测子步骤的输出为情感的VAD连续模型的1到10中的预测值,该VAD连续模型指的是情感的唤醒度、控制度以及愉悦度。

[0084] 本实施例还提供一种采用如上的一种基于情境感知的多模态情感识别方法的系统,包括:

[0085] 多模态信息采集单元,被配置为执行多模态信息采集步骤;

[0086] 基于多模态行为表现的情感处理单元,被配置为执行基于多模态行为表现的情感处理步骤;

[0087] 基于场景上下文的情感分析单元,被配置为执行基于场景上下文的情感分析步骤;

[0088] 基于代理人群体交互的情感分析单元,被配置为执行基于代理人群体交互的情感分析步骤;

[0089] 基于代理人和情境交互的情感分析单元,被配置为执行基于代理人和情境交互的

情感分析步骤;

[0090] 基于自适应规划的特征融合单元,被配置为执行特征融合步骤;

[0091] 情感识别单元,被配置为执行情感识别步骤;

[0092] 具体的,本实施例中情感识别单元包括基于离散情感的识别单元和基于连续情感的预测单元。

[0093] 优选地,系统还包括显示模块,该显示模块被配置为线束情感识别单元的输出结果。

[0094] 具体的,多模态数据一方面来源于人体外在表现的面部、姿态以及手势信息,另一方面来自预处理过程中获取的来自图像或者视频中的完整场景信息、所有代理人交互信息以及场景和代理人交互信息。随后在不同的情感处理和分析单元通过不同的神经网络和处理技术提取到对应的不同情感语音特征。在自适应规划的特征融合单元中,为了抵抗多模态数据采集单元中产生的信号干扰和冗余信息,该单元采取乘性和级联策略结合的自适应方式进行特征融合,以确保多模态情感特征的完整性和有效性。随后馈送到离散情感的识别单元利用多标签分类损失训练网络后输出情感类别。具体的,情感类别包括高兴、惊喜、悲伤、厌恶、兴奋、平和、恐惧和生气。在连续情感预测单元中通过均方误差损失训练网络后,输出情感的VAD连续模型的1到10中的预测值。具体的,VAD模型指的是情感的唤醒度、控制度以及愉悦度,其衡量了情感空间在连续状态上的变化情况,能够更加形象生动的刻画情感强度和表征情感差异。随后,通过显示单元可以呈现出离散情感分析和连续情感预测的结果。

[0095] 下面对该系统各单元的具体实施过程进行详细描述。

[0096] 2.1、基于多模态行为表现的情感处理单元

[0097] 图2为本实施例提供的基于多模态行为表现的情感处理单元原理框图,该单元包括四个子特征提取单元,分别是面部表情提取单元、面部关键点提取单元、情感姿态提取单元以及情感手势提取单元。针对面部表情提取单元,首先通过OpenFace面部检测器提取面部表情轮廓,得到尺寸为 224×224 的面部图像,随后通过设计的五层卷积神经网络进行特征提取操作。该神经网络头部包含五层卷积层以及相连接的批量归一化层和ReLU激活层,尾部包含两层最大池化层进行特征降维和缩放,最终获得 $26 \times 1 \times 1$ 大小的特征向量。针对面部关键点提取单元,首先通过OpenFace面部检测器提取68个面部界标点,随后转换成 $136 \times 1 \times 1$ 大小的初始特征向量。该向量通过两层一维的卷积层以及相连接的批量归一化层和ReLU激活层,最后经过一层全连接层获取 $26 \times 1 \times 1$ 大小的特征向量。针对情感姿态提取单元,通过AlphaPose姿态检测器提取26个人体姿态的坐标点,将其转换成 $26 \times 2 \times 1$ 的特征向量馈送到编码器网络进行特征提取,最终获得 $26 \times 1 \times 1$ 大小的特征向量。针对情感手势提取单元,通过AlphaPose姿态检测器提取人体手部表征关键点,利用三层转置卷积层以及相连接的批量归一化层和ReLU激活层获得 $512 \times 1 \times 1$ 的特征向量,随后再通过三层卷积层获取 $26 \times 1 \times 1$ 的大小的特征向量。随后通过初始特征级联的方式获取 $104 \times 1 \times 1$ 大小的外在行为表现模态向量进行模态输出。

[0098] 2.2、基于场景上下文的情感分析单元

[0099] 图3为本实施例提供的基于场景上下文的情感分析单元原理框图,单元主要包括数据预处理、基于通道的注意力特征提取、基于空间的注意力特征提取、热力图生成以及特

征提取和输出几个阶段。具体的,在数据预处理阶段,通过为输入的真实世界图像和利用视频截取的视频帧表示为中的待进行情感识别的人类代理人主体添加蒙版,以获取只保留场景信息的场景图像,随后选取残差神经网络的ResNet-18作为主体模型的骨干网络,在其中所包含的8个残差连接块中按顺序交替嵌入基于通道和空间的注意力机制模块形成完整的注意力提取网络进行场景情感语义的特征提取。针对通道注意力机制,通过全局平均池化推理出一张1D的通道注意力图,随后在输出层通过通道级别的乘法进行特征合并;针对空间注意力机制,通过全局最大池化层推理出一张2D的空间注意力图,随后在输出层通过通道级别的乘法进行特征合并。在注意力机制的帮助下,使得该系统可以更加聚焦于关注和所对应代理人相关的情感线索,进一步生成注意力热图。在注意力热图中,本实施例可以将权重较大的场景语义进行可视化,并进行标注,便于分析场景信息和情感的关联程度和耦合关系。最终讲进行特征提取和得到 $26*1*1$ 大小的情感特征向量。

[0100] 2.3、基于代理人群体交互的情感分析单元

[0101] 图4为本实施例提供的基于代理人群体交互的情感分析单元原理框图,其主要包括数据预处理、特征预提取、情感关系图构建以及情感特征输出四个步骤。具体的,数据预处理阶段,通过在输入的真实世界图像和利用视频截取的视频帧表示为I中分别提取待情感识别的人类主体 I_{agent} 和其他代理人信息 I_p ,通过深度残差网络ResNet-18首先提取两方的初始特征分别记为 f_{agent} 和 f_p ,随后将不同的特征作为情感节点馈送入图注意力网络(Graph Attention Network)中构建情感关系图。之后考虑到不同的其他代理人对识别主体的情感影响强度和程度不同,为此计算它们之间的情感相似系数记为 $n_{ij} = \alpha([Wf_{agent} || Wf_p])$ 。其中W表示权重参数, $\alpha(.)$ 表示特征映射关系以及 $||$ 表示为连接操作。为了增强情感传递的关系学习,本实施例同时使用多头注意机制实现域节点特征的融合,即在相邻的两个特征节点之间进行三次情感关系系数的计算。最终基于得到的不同权重影响下的其他代理人特征和原识别主体特征做加权平均操作,输出 $26*1*1$ 大小的情感特征向量 h_3 。

[0102] 2.4、基于代理人和情境交互的情感分析单元

[0103] 图5为本实施例提供的基于代理人和情境交互的情感分析单元原理框图。该单元旨在探索其他代理人和场景交互过程中产生的情感线索对识别主体代理人产生的情感影响。其中包括数据预处理、代理人特征图构建、特征聚合以及模态输出四个步骤。具体的,通过为输入的真实世界图像和利用视频截取的视频帧表示为I中的所有其他代理人添加蒙版得到仅保留场景信息的图像 I_s ,通过深度残差网络ResNet-18首先提取场景的初始特征 f_s ,随后利用代理人群体交互的情感分析单元中得到的 f_{agent} 馈送入两层的图卷积神经网络建立基本特征图。接下来利用长短期记忆网络实现 f_s 和 f_{agent} 的特征聚合,得到 $26*1*1$ 大小的情感特征向量 h_4 。

[0104] 2.5、基于自适应规划的特征融合单元

[0105] 图6为本实施例提供的基于自适应规划的特征融合单元原理框图。具体的,该单元会自适应的根据上述情感分析和识别单元得到的多模态特征的特点进行特征筛选和融合。当输入的特征是基于面部表情、面部关键点等强相关的特征向量时,为保证情感特征空间的完备性和强表达能力,系统会自动的进行特征级联操作进行特征融合;当输入的特征是基于代理人之间传递的情感信息和场景的情感语义信息等弱相关的特征向量时,为了充分挖掘和增强情感信息的表征能力,系统会自动执行乘性融合方式,通过保持原维度不变的

形式,进行特征渗透和互补,最大限度弥补特征之间的差异性。最终将融合后的特征进行输出。

[0106] 2.6、基于离散情感的识别单元

[0107] 图7为本实施例提供的基于离散情感的识别单元原理框图,在该单元中,首先对得到的融合后情感特征向量进行归一化处理,随后利用Sigmoid函数将输出的特征值映射到0到1之间,然后对每个输出节点和对应的标签计算交叉熵损失函数,通过计算出每一类可能输出的表情标签的概率,且所有八种概率之和加起来为1。随后选取最大概率的标签输出,即为系统预测为得到的表情标签。

[0108] 2.7、基于连续情感的预测单元

[0109] 图8为本实施例提供的基于连续情感的预测单元原理框图,其主要包含数据归一化、标签差值求和、误差幅度计算以及最终的连续数值预测几个步骤。具体的,该单元采用均方误差损失来计算预测数值和目标数值之间的差值的平方和。由于本实施例中数据收集的可靠性,因此对预测局外点具有鲁棒性。所以模型会致力减小局外点造成的误差,从而使模型的整体表现提高。

[0110] 均方误差损失进行优化可以得到所有观测的均值,为此针对神经网络训练过程中,它的梯度在极值点处会有很大的跃变,误差幅度计算在极值点有着良好的特性。同时加上动态调整学习率的策略,均方误差的梯度随着损失函数的减小而减小,这一特性使得在最后的训练过程中能得到更精确的结果。

[0111] 通过离散情感的识别单元和连续情感的预测单元获得情感信息后,在显示单元中通过可视化的方式进一步的实现结果展示和呈现。

[0112] 本实施例展示的基于情境感知的多模态情感识别系统首次引入情境感知中的场景语义、代理人交互以及代理人与场景交互中蕴含的情感特征,有效的推动了真实世界下的多模态情感识别的发展。同时,通过充分采集情感识别主体的面部表情信息,面部关键点信息、手势信号以及姿态信号等外在情感信息,结合不同识别单元对异构模态信息进行预处理和特征提取,进一步使用自适应规划的特征融合方式实现了多模态特征的有机融合。最终基于情感分类和预测的多任务学习模式,极大的加强了多模态情感识别的精度,提高了模型的泛化性能和准确率。本实施例所提出的方法,可以提供完整有效的情感判别特征空间,为后续的人类自然情感理解以及开放世界的情感表征提供可靠保障。

[0113] 以上详细描述了本发明的较佳具体实施例。应当理解,本领域的普通技术人员无需创造性劳动就可以根据本发明的构思做出诸多修改和变化。因此,凡本技术领域技术人员依本发明的构思在现有技术的基础上通过逻辑分析、推理或者有限的实验可以得到的技术方案,皆应在由权利要求书所确定的保护范围内。

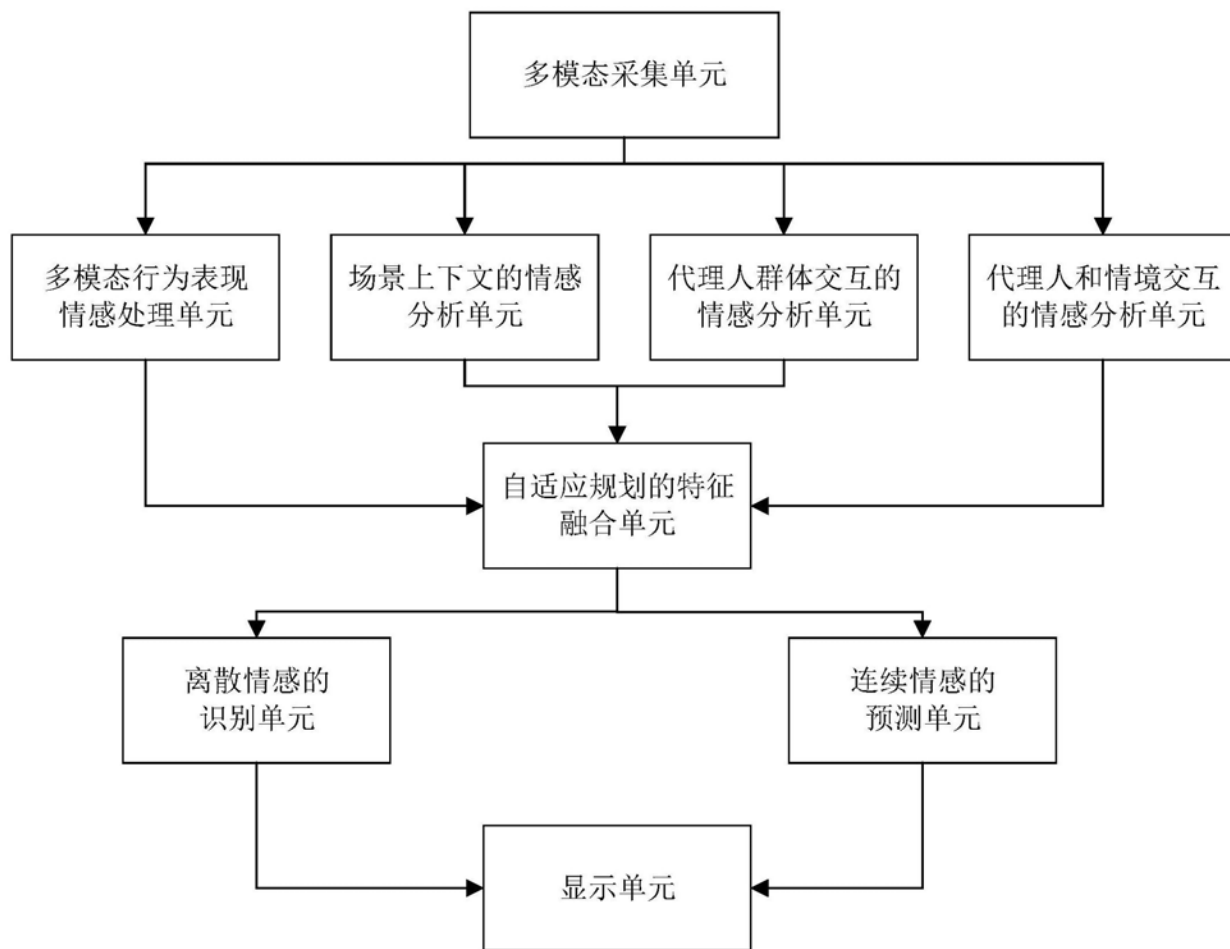


图1

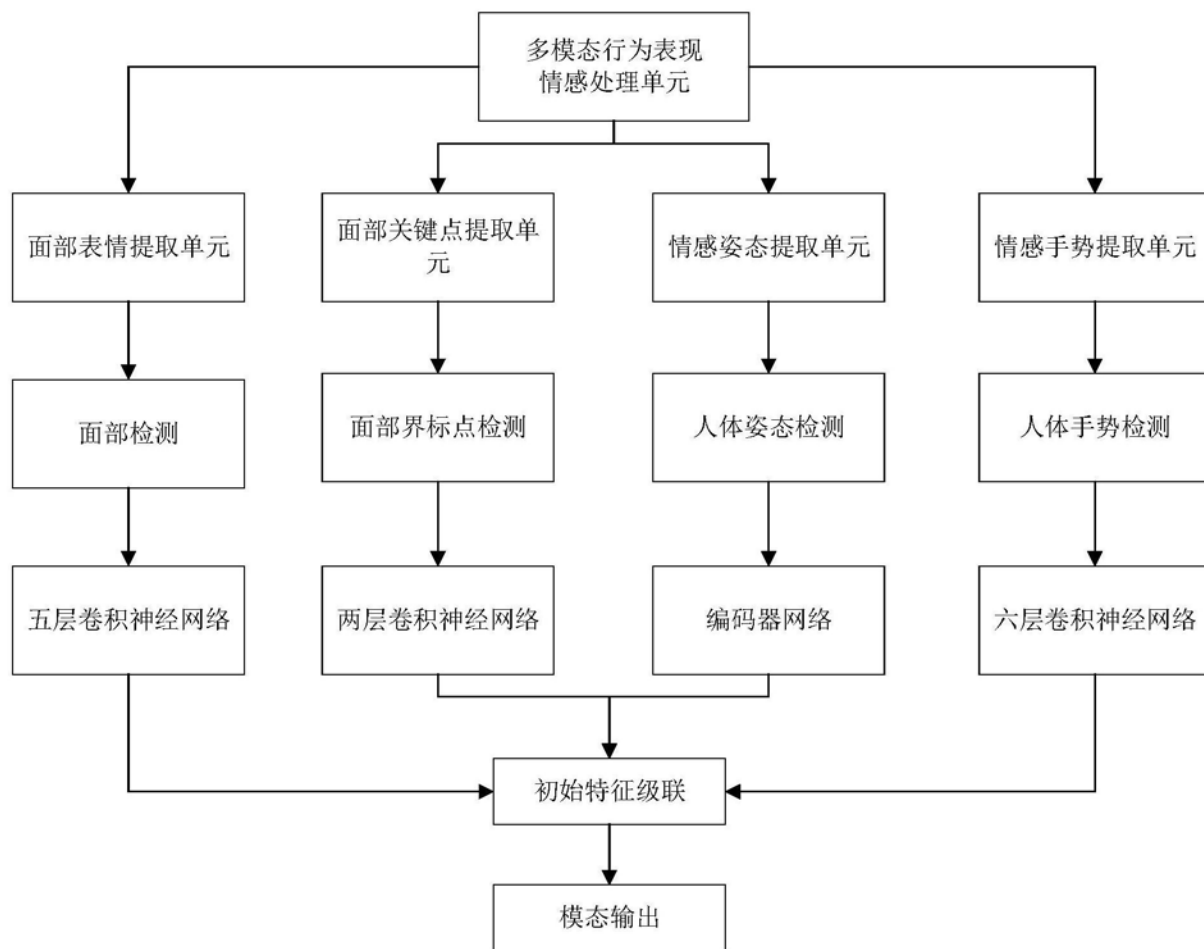


图2

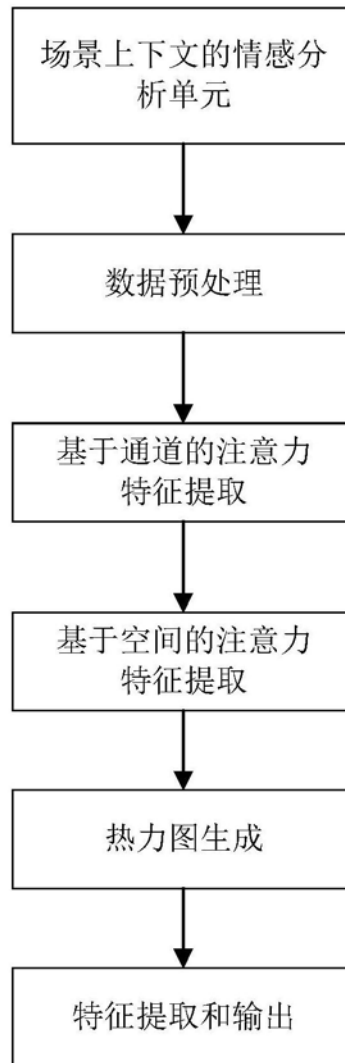


图3

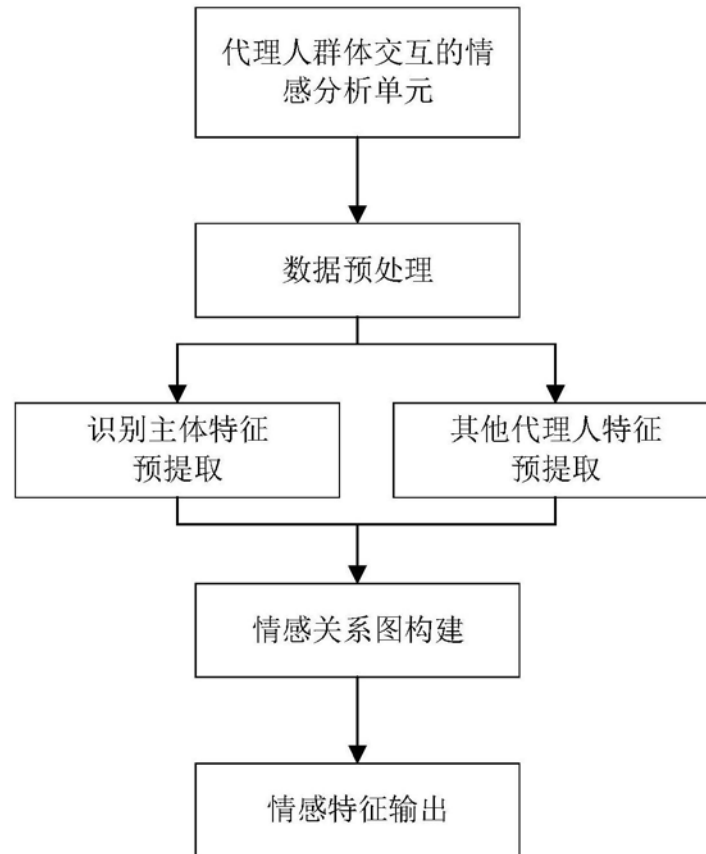


图4

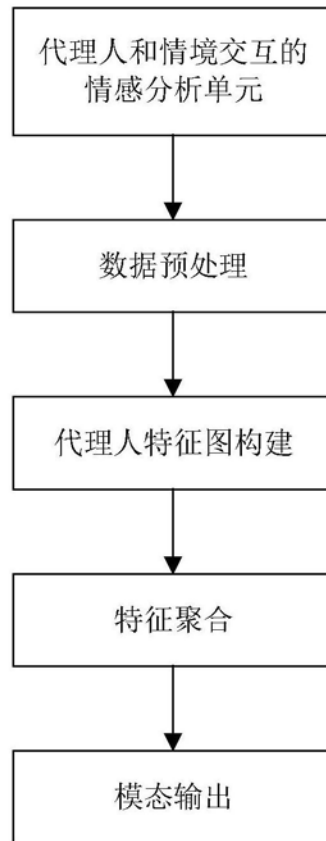


图5

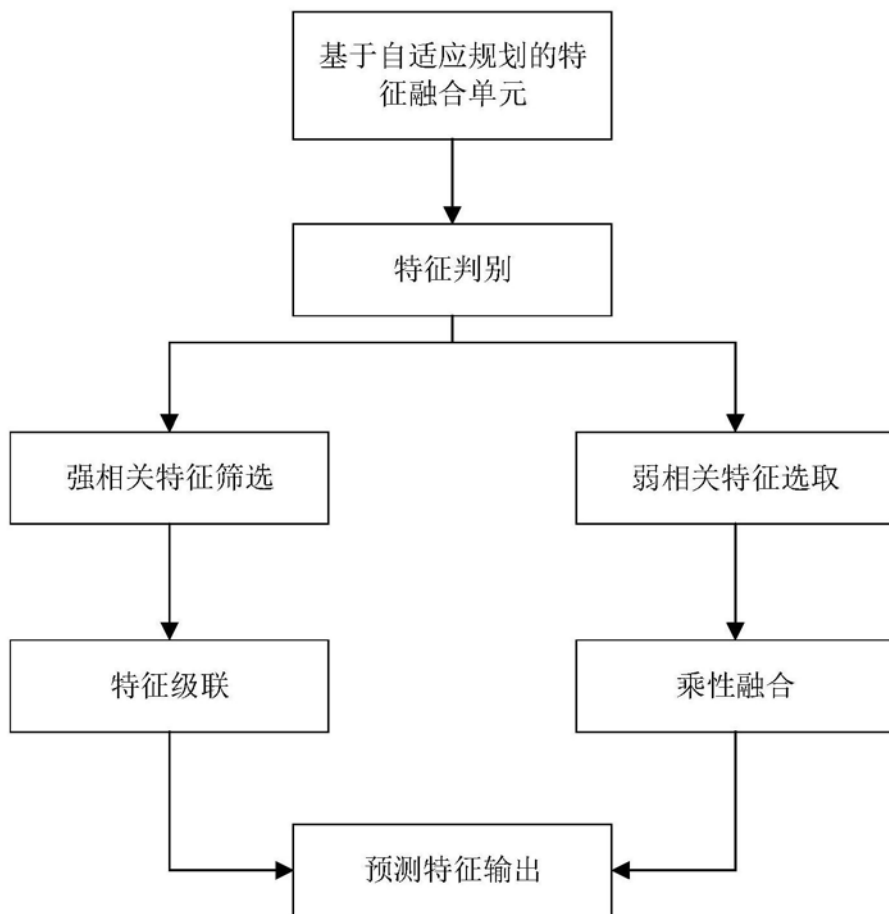


图6

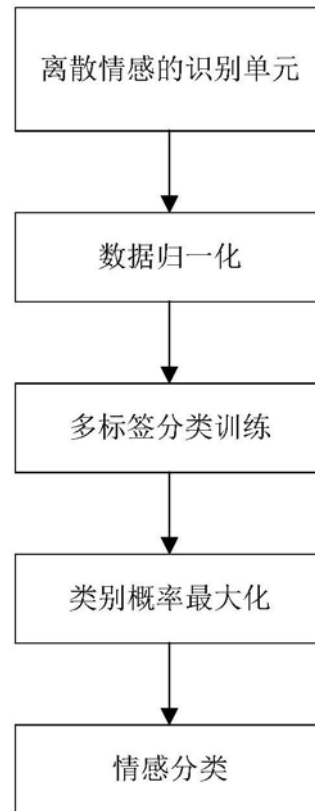


图7

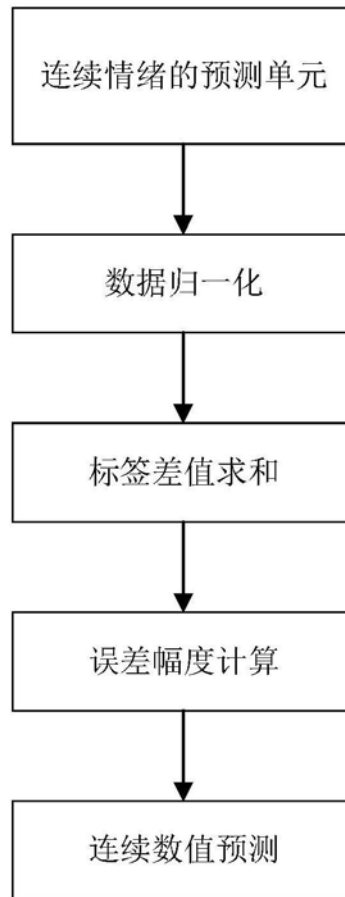


图8