



(12) 发明专利申请

(10) 申请公布号 CN 120220643 A

(43) 申请公布日 2025. 06. 27

(21) 申请号 202510214437.3
(22) 申请日 2025.02.26
(71) 申请人 联通沃悦读科技文化有限公司
地址 410011 湖南省长沙市长沙县黄花镇
长沙临空经济示范区人民东路与小康
路交汇处西南角房屋
申请人 联通在线信息科技有限公司
(72) 发明人 李勇刚 何熠 何玄 李隼
(74) 专利代理机构 湖南盈奥知识产权代理事务
所(普通合伙) 43282
专利代理师 龚燕妮 姚瑶

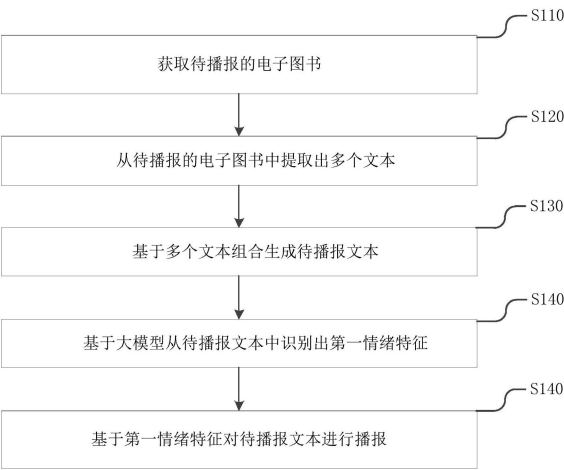
G06F 40/205 (2020.01)
G06F 40/284 (2020.01)
G06F 40/30 (2020.01)
G06F 18/25 (2023.01)
G06N 5/04 (2023.01)
G06N 3/045 (2023.01)
G06N 3/0475 (2023.01)
G06N 3/096 (2023.01)
G06V 30/414 (2022.01)
G06V 30/416 (2022.01)

(51) Int.Cl.
G10L 13/08 (2013.01)
G10L 13/10 (2013.01)
G10L 13/027 (2013.01)

权利要求书1页 说明书10页 附图2页

(54) 发明名称
一种听书的智能播报方法、系统、设备及介
质

(57) 摘要
本申请涉及了一种听书的智能播报方法、系
统、设备及介质,本方法包括获取待播报的电子
图书;从待播报的电子图书中提取出多个文本;
基于多个文本组合生成待播报文本;大模型从待
播报文本中识别出第一情绪特征;基于第一情绪
特征对待播报文本进行播报;通过获取待播报的
电子图书,提取多个文本,组合生成待播报文本,
基于大模型识别情绪特征,并根据情绪特征进行
播报,解决了传统播报方法难以准确表达图书情
绪特征的问题,具有能够准确表达图书情绪特
征,提供个性化和互动性的听书体验,有效提取
和整合图文内容,以及有效组织和梳理长篇幅或
结构复杂的电子图书内容的优点。



1. 一种听书的智能播报方法,其特征在于,所述方法包括:
获取待播报的电子图书;
从所述待播报的电子图书中提取出多个文本;
基于多个文本组合生成待播报文本;
基于大模型从所述待播报文本中识别出第一情绪特征;
基于所述第一情绪特征对待播报文本进行播报。
2. 根据权利要求1所述听书的智能播报方法,其特征在于,所述从所述待播报的电子图书中提取出多个文本,包括:
从所述待播报的电子图书中,确定多个的文字区域;
基于图像转文本技术,识别出多个文字区域中的多个文本。
3. 根据权利要求2所述听书的智能播报方法,其特征在于,所述基于多个文本组合生成待播报文本,包括:
获取所述电子图书的关键字;所述关键字是能够表征所述电子图书的基本信息的词或句子;
根据所述关键字,将所述多个文本整理成所述待播报文本。
4. 根据权利要求3所述听书的智能播报方法,其特征在于,所述大模型为Chat-Gpt或DeepSeek。
5. 根据权利要求4所述听书的智能播报方法,其特征在于,在所述基于大模型从所述待播报文本中识别出第一情绪特征之前,所述方法还包括:
识别用户的体征信息;
基于所述用户的体征信息生成第二情绪特征;
所述基于所述第一情绪特征对待播报文本进行播报,包括:
根据所述第一情绪特征和所述第二情绪特征对所述待播报文本进行播报。
6. 根据权利要求5所述听书的智能播报方法,其特征在于,所述体征信息包括面部特征、心率特征中的任意一种。
7. 根据权利要求6所述听书的智能播报方法,其特征在于,所述第一情绪特征和所述第二情绪特征均包括中性、喜悦、悲伤、愤怒、困倦、厌恶、惊讶或恐惧中的任意一种。
8. 一种听书的智能播报系统,其特征在于,所述系统包括:
图书获取单元,用于获取待播报的电子图书;
文本提取单元,用于从所述待播报的电子图书中提取出多个文本;
文本获取单元,用于基于多个文本组合生成待播报文本;
情绪特征获取单元,用于基于大模型从所述待播报文本中识别出第一情绪特征;
文本播报单元,用于基于所述第一情绪特征对待播报文本进行播报。
9. 一种电子设备,其特征在于,包括:至少一个控制器和用于与所述至少一个控制器通信连接的存储器;所述存储器存储有可被所述至少一个控制器执行的指令,所述指令被所述控制器执行,以使所述控制器执行权利要求1至7任一项所述的听书的智能播报方法。
10. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质存储有计算机可执行指令,所述计算机可执行指令用于使计算机执行权利要求1至7任一项所述的听书的智能播报方法。

一种听书的智能播报方法、系统、设备及介质

技术领域

[0001] 本申请实施例涉及图书播报技术领域,尤其涉及一种听书的智能播报方法、系统、设备及介质。

背景技术

[0002] 电子图书又称为e-book或数字图书,是随着电子出版、互联网以及现代通信电子技术的发展应运而生的一种新的图书形式。它区别于以纸张为载体的传统出版物,代表人们所阅读的数字化出版物。电子图书的内容主要是以特殊的格式制作而成,可在有线或无线网络上传播,一般由专门的网站组织而成。

[0003] 在电子图书的播报过程中,通常是将电子图书转换成文本格式,然后将文本进行播放,但是现有技术难以准确地表达出图书的情绪。

发明内容

[0004] 以下是对本文详细描述的主题的概述。本概述并非是为了限制权利要求的保护范围。

[0005] 本公开实施例的主要目的在于提出一种听书的智能播报方法、系统、设备及介质,能够解决了传统播报方法难以准确表达图书情绪特征的问题。

[0006] 本申请实施例的第一方面提出了一种听书的智能播报方法,所述方法包括:

[0007] 获取待播报的电子图书;

[0008] 从所述待播报的电子图书中提取出多个文本;

[0009] 基于多个文本组合生成待播报文本;

[0010] 基于大模型从所述待播报文本中识别出第一情绪特征;

[0011] 基于所述第一情绪特征对待播报文本进行播报。

[0012] 本公开实施例提供的一种听书的智能播报方法,至少包括如下有益效果:

[0013] 通过获取待播报的电子图书,提取多个文本,组合生成待播报文本,基于大模型识别情绪特征,并根据情绪特征进行播报,解决了传统播报方法难以准确表达图书情绪特征的问题,具有能够准确表达图书情绪特征,提供个性化和互动性的听书体验,有效提取和整合图文内容,以及有效组织和梳理长篇幅或结构复杂的电子图书内容的优点。

[0014] 在一些实施例中,所述从所述待播报的电子图书中提取出多个文本,包括:

[0015] 从所述待播报的电子图书中,确定多个的文字区域;

[0016] 基于图像转文本技术,识别出多个文字区域中的多个文本。

[0017] 在一些实施例中,所述基于多个文本组合生成待播报文本,包括:

[0018] 获取所述电子图书的关键字;所述关键字是能够表征所述电子图书的基本信息的词或句子;

[0019] 根据所述关键字,将所述多个文本整理成所述待播报文本。

[0020] 在一些实施例中,所述大模型为Chat-Gpt或DeepSeek。

[0021] 在一些实施例中,在所述基于大模型从所述待播报文本中识别出第一情绪特征之前,所述方法还包括:

[0022] 识别用户的体征信息;

[0023] 基于所述用户的体征信息生成第二情绪特征;

[0024] 所述基于所述第一情绪特征对待播报文本进行播报,包括:

[0025] 根据所述第一情绪特征和所述第二情绪特征对所述待播报文本进行播报。

[0026] 在一些实施例中,所述体征信息包括面部特征、心率特征中的任意一种。

[0027] 在一些实施例中,所述第一情绪特征和所述第二情绪特征均包括中性、喜悦、悲伤、愤怒、困倦、厌恶、惊讶或恐惧中的任意一种。

[0028] 本申请实施例的第二方面提出了一种听书的智能播报系统,所述系统包括:

[0029] 图书获取单元,用于获取待播报的电子图书;

[0030] 文本提取单元,用于从所述待播报的电子图书中提取出多个文本;

[0031] 文本获取单元,用于基于多个文本组合生成待播报文本;

[0032] 情绪特征获取单元,用于基于大模型从所述待播报文本中识别出第一情绪特征;

[0033] 文本播报单元,用于基于所述第一情绪特征对待播报文本进行播报。

[0034] 本申请实施例的第三方面提出了一种电子设备,至少一个控制器和用于与所述至少一个控制器通信连接的存储器;所述存储器存储有可被所述至少一个控制器执行的指令,所述指令被所述控制器执行,以使所述控制器执行如第一方面所述的听书的智能播报方法。

[0035] 本申请实施例的第四方面提出了一种计算机可读存储介质,所述计算机可读存储介质存储有计算机可执行指令,所述计算机可执行指令用于使计算机执行如上述的听书的智能播报方法。

[0036] 可以理解的是,上述第二方面至第四方面和相关技术相比存在的有益效果和上述第一方面和相关技术相比存在的有益效果相同,可以参见上述第一方面中的相关描述,在此不再赘述。

附图说明

[0037] 为了更清楚地说明本申请实施例中的技术方案,下面将对实施例或相关技术描述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实施例的一些实施例,对于本领域普通技术人员来说,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0038] 图1是本申请提供的一种听书的智能播报方法的示意图

[0039] 图2是本申请实施例提出的一种听书的智能播报系统的示意图;

[0040] 图3是本申请实施例提出的电子设备的示意图。

具体实施方式

[0041] 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,此处所描述的具体实施例仅用以解释本申请,并不用于限定本申请。

[0042] 需要说明的是,虽然在装置示意图中进行了功能模块划分,在流程图中示出了逻辑顺序,但是在某些情况下,可以以不同于装置中的模块划分,或流程图中的顺序执行所示出或描述的步骤。说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。

[0043] 除非另有定义,本文所使用的所有的技术和科学术语和属于本申请的技术领域的技术人员通常理解的含义相同。本文中所使用的术语只是为了描述本申请实施例的目的,不是旨在限制本申请。

[0044] 电子图书作为一种新兴的数字化出版物形式,已经成为现代阅读和信息获取的重要载体。然而,在电子图书的播报过程中,存在着难以准确表达图书情绪的技术问题。这一问题主要源于当前技术将电子图书简单转换为文本格式后进行播放的方法,无法有效捕捉和传递原文中蕴含的情感特征。

[0045] 具体来说,在电子图书播报系统中,文本到语音的转换过程通常采用标准化的语音合成技术。这种技术虽然能够准确地将文字转换为语音,但往往缺乏对文本内容情感特征的识别和表达能力。例如,当播报一段描述悲伤场景的文本时,系统可能会以平淡或不恰当的语气进行播报,这就导致了原文所要传达的情感无法被准确地呈现给听众。此外,由于电子图书的内容往往包含复杂的情节和丰富的情感变化,单一的语音合成方法难以适应这种多样性,从而影响了用户的听书体验。

[0046] 由此,如果不能有效解决这一技术问题,将会对电子图书播报系统的实用性和用户体验产生显著的负面影响。首先,听众可能无法准确理解作者想要传达的情感,导致对文学作品的误解或欣赏度降低。其次,缺乏情感表达的播报方式可能会使听书过程变得单调乏味,降低用户的参与度和持续使用意愿。此外,对于某些特定类型的电子图书,如儿童读物或情感类小说,准确的情感表达尤为重要,不能解决这一问题将严重限制这些类型电子图书的播报效果。因此,开发一种能够准确识别和表达电子图书情感特征的智能播报方法,对于提升电子图书播报系统的技术水平和用户体验具有重要意义。

[0047] 针对电子图书播报过程中难以准确表达图书情绪的技术问题,本申请进行了深入的思考和探索。

[0048] 首先,考虑到电子图书内容的复杂性和情感表达的多样性,需要一种能够全面理解文本内容的方法。其中,大模型技术因其强大的自然语言处理能力成为一个可能的解决方案。大模型能够理解文本的语义和上下文,从而有可能识别出文本中蕴含的情绪特征。

[0049] 具体来说,本申请提出了使用大模型从待播报文本中识别情绪特征的方法。这种方法的优势在于,大模型可以基于海量的训练数据,对文本进行深度语义分析,从而准确捕捉文本中的情感信息。例如,当遇到描述悲伤场景的文本时,大模型能够识别出其中的悲伤情绪,为后续的播报提供情感指导。

[0050] 然而,仅仅识别出情绪特征还不足以解决问题。如何将识别出的情绪特征应用到实际的播报过程中,成为下一个需要解决的问题。由此,本申请进一步提出了基于识别出的情绪特征对待播报文本进行播报的方法。这种方法可以根据不同的情绪特征,调整播报的语气、语速和音调,从而更准确地表达文本中的情感。

[0051] 如图1,为了实现这一方法,本申请设计了一个完整的处理流程。

[0052] 步骤S110,获取待播报的电子图书;

[0053] 步骤S120,从待播报的电子图书中提取出多个文本;

[0054] 步骤S130,基于多个文本组合生成待播报文本;

[0055] 步骤S140,基于大模型从待播报文本中识别出第一情绪特征;

[0056] 步骤S150,基于第一情绪特征对待播报文本进行播报。

[0057] 这种方法通过先确定文字区域,再利用图像转文本技术进行识别,可以更准确地从电子图书中提取出文本内容。这种方法特别适用于包含图文混排的电子图书,能够有效地将文字内容与图片等非文字内容区分开来。

[0058] 在实际应用中,确定文字区域可以采用多种方法。例如,可以使用图像分割算法,将电子图书页面划分为不同的区域,然后通过特征分析识别出文字区域。或者,可以利用机器学习模型,训练一个能够自动识别文字区域的分类器。

[0059] 基于图像转文本技术识别文本的过程中,可以采用光学字符识别(OCR)技术。OCR技术可以将图像中的文字转换为可编辑的文本格式。具体实现时,可以选择适合电子图书特点的OCR算法,如针对不同字体、不同语言的专门算法,以提高识别的准确性。

[0060] 在某些情况下,电子图书可能包含复杂的版面布局或特殊字体。为了应对这种情况,本申请的方法可以进一步结合版面分析技术。版面分析可以帮助理解文字区域的布局结构,从而更准确地提取文本内容。例如,可以识别出标题、正文、页眉页脚等不同类型的文字区域,并根据其特点采用不同的处理策略。

[0061] 此外,为了提高文本提取的效率和准确性,本申请的方法还可以采用并行处理技术。例如,可以将电子图书的不同页面或不同区域同时进行处理,充分利用计算资源,加快文本提取的速度。

[0062] 作为一种具体的实施例,假设有一本包含300页的电子图书需要进行文本提取。首先,使用图像分割算法将每一页划分为若干区域。然后,应用一个预先训练好的深度学习模型来识别这些区域中的文字区域。识别出文字区域后,使用专门针对中文优化的OCR算法来识别文字内容。在这个过程中,系统会同时处理多个页面,例如每次处理10页。对于识别出的文本,系统还会进行后处理,如去除多余的空格、修正常见的识别错误等。最后,系统将所有识别出的文本按照原书的顺序组合起来,形成完整的文本内容。

[0063] 与现有技术相比,本申请提出的方法具有以下优势:首先,通过先确定文字区域再进行文本识别的两步法,提高了文本提取的准确性,特别是对于复杂版面的电子图书;其次,采用图像转文本技术,使得方法适用于各种格式的电子图书,不局限于特定的文件格式;最后,通过并行处理等技术手段,提高了文本提取的效率。这些优势使得本申请的方法能够更好地解决电子图书文本提取的问题,为后续的智能播报提供了高质量的文本基础。

[0064] 在上述的一些实施例中,本申请在实施的过程中,还存在如何更有效地组合多个文本以生成待播报文本的问题。

[0065] 对此,本申请进一步提出了基于多个文本组合生成待播报文本的方法,包括获取电子图书的关键字,关键字是能够表征电子图书的基本信息的词或句子,然后根据关键字将多个文本整理成待播报文本。

[0066] 这种方法通过引入关键字的概念,为文本组合提供了一个有效的指导方向。关键字作为电子图书基本信息的代表,可以帮助系统更准确地理解和组织文本内容。

[0067] 具体来说,首先获取电子图书的关键字。这些关键字可以是表示电子图书主题、人

物、情节或主要概念的词语或短句。例如,对于一本科幻小说,关键字可能包括“太空探索”、“人工智能”、“时间旅行”等。

[0068] 获取关键字的方法可以有多种。一种方式是通过自然语言处理技术从电子图书的目录、摘要或正文中提取。另一种方式是利用元数据信息,如图书的标签或分类信息。还可以结合人工智能技术,使用机器学习模型来识别和提取关键字。

[0069] 接下来,系统会根据这些关键字来整理多个文本。这个过程可能包括以下几个步骤:

[0070] 1. 文本相关性分析:系统会评估每个文本片段与关键字的相关程度。这可以通过计算文本中关键字的出现频率、位置等因素来实现。

[0071] 2. 文本排序:根据相关性分析的结果,系统会对文本进行排序,将与关键字最相关的文本放在前面。

[0072] 3. 文本组合:系统会按照排序结果将文本片段组合在一起,形成一个连贯的待播报文本。在这个过程中,可能需要添加一些过渡语句,以确保文本的流畅性。

[0073] 4. 内容优化:系统可能会进行进一步的优化,如去除重复内容、调整段落顺序等,以提高待播报文本的质量。

[0074] 通过这种方法,本申请能够更有效地组织和整理电子图书中的文本内容,生成一个结构合理、重点突出的待播报文本。这不仅提高了文本的连贯性和可读性,还能确保播报内容更好地反映电子图书的核心信息。

[0075] 作为一种优选的实施方式,可以考虑以下具体实例:

[0076] 假设有一本名为《人工智能与未来社会》的电子图书。系统首先通过分析图书的目录和摘要,提取出关键字如“人工智能”、“机器学习”、“社会影响”、“伦理问题”等。

[0077] 然后,系统从电子图书中提取出多个文本片段。例如:

[0078] 片段1:“人工智能技术正在快速发展,影响着我们生活的方方面面。”

[0079] 片段2:“机器学习算法能够从大量数据中学习模式和规律。”

[0080] 片段3:“随着AI的普及,我们需要考虑其对就业市场的影响。”

[0081] 片段4:“人工智能在医疗诊断领域展现出巨大潜力。”

[0082] 系统会根据关键字对这些片段进行相关性分析和排序。在这个例子中,片段1和片段2与关键字“人工智能”和“机器学习”高度相关,可能会被排在前面。片段3涉及到“社会影响”,也会被纳入待播报文本。片段4虽然与人工智能有关,但可能与主要关键字的相关性较低,可能会被排在后面或被省略。

[0083] 最后,系统会将这些片段组合成一个连贯的待播报文本,可能如下:

[0084] “人工智能技术正在快速发展,影响着我们生活的方方面面。其中,机器学习算法能够从大量数据中学习模式和规律,这是人工智能发展的关键技术之一。随着AI的普及,我们需要考虑其对就业市场的影响,这是人工智能发展过程中不可忽视的社会问题。”

[0085] 这种方法相比于简单地按照文本在书中的顺序进行组合,能够更好地突出电子图书的核心内容,使得生成的待播报文本更加精炼和有重点。同时,通过关键字的引导,可以确保不同章节或段落之间的内容能够更加连贯,提高了听书体验的质量。

[0086] 此外,这种方法还具有很强的灵活性和可扩展性。例如,可以根据用户的兴趣或听书的目的来调整关键字的权重,从而生成更符合用户需求的待播报文本。也可以结合自然

语言处理技术,进一步优化文本的组合逻辑,如考虑语义连贯性、情感一致性等因素。

[0087] 与现有技术相比,本申请的方法在文本组合方面具有明显优势。传统方法可能只是简单地按照文本在书中的顺序进行组合,或者使用一些基本的规则来选择文本。而本申请通过引入关键字概念,为文本组合提供了一个更加智能和有针对性的方法。这不仅提高了待播报文本的质量,还能更好地保留和传达电子图书的核心信息,从而显著改善了用户的听书体验。

[0088] 在上述的一些实施例中,本申请在实施的过程中,还存在无法准确识别待播报文本中的情绪特征的问题。

[0089] 对此,本申请进一步提出了大模型为Chat-Gpt或DeepSeek的技术方案。

[0090] 本申请提出的听书的智能播报方法包括以下步骤:获取待播报的电子图书;从待播报的电子图书中提取出多个文本;基于多个文本组合生成待播报文本;基于大模型从待播报文本中识别出第一情绪特征;基于第一情绪特征对待播报文本进行播报。其中,从待播报的电子图书中提取出多个文本包括:从待播报的电子图书中,确定多个的文字区域;基于图像转文本技术,识别出多个文字区域中的多个文本。基于多个文本组合生成待播报文本包括:获取电子图书的关键字,关键字是能够表征电子图书的基本信息的词或句子;根据关键字,将多个文本整理成待播报文本。

[0091] 本申请通过使用Chat-Gpt或DeepSeek作为大模型来识别待播报文本中的情绪特征,可以更准确地捕捉文本中的情感信息。这两种大模型都具有强大的自然语言处理能力,能够理解文本的上下文和语义,从而更精确地识别出文本中蕴含的情绪。

[0092] 具体来说,Chat-Gpt是一种基于transformer架构的大规模语言模型,它通过预训练大量文本数据,学习了丰富的语言知识和上下文理解能力。在识别情绪特征时,Chat-Gpt可以分析文本的语言结构、词语选择和语气,从而推断出文本所表达的情绪。

[0093] DeepSeek则是另一种先进的自然语言处理模型,它同样具有强大的文本理解和分析能力。DeepSeek可以通过深度学习算法,从文本中提取出情感相关的特征,并将这些特征映射到不同的情绪类别。

[0094] 使用这两种大模型中的任一种,都可以显著提高情绪特征识别的准确性。例如,当遇到一段描述悲伤场景的文本时,大模型可以通过分析文本中的关键词、句子结构和整体语境,准确地识别出文本中的悲伤情绪。这种准确的情绪识别为后续的播报提供了重要基础,使得播报能够更好地传达文本的情感内容。

[0095] 进一步地,由于Chat-Gpt和DeepSeek都是持续更新和优化的模型,它们的性能会随着时间的推移而不断提升。这意味着本申请的方法可以随着大模型的进步而自动获得性能提升,无需频繁地更新系统。

[0096] 作为一种优选的实施方式,可以选择使用Chat-Gpt作为大模型。具体实施时,首先将待播报文本输入到Chat-Gpt模型中。然后,通过设置适当的提示(prompt),要求Chat-Gpt分析文本的情绪特征。例如,可以使用如下提示:“请分析以下文本的情绪特征,并给出主要的情绪类型。”接着,Chat-Gpt会输出分析结果,包括识别出的主要情绪类型(如喜悦、悲伤、愤怒等)以及相应的置信度。最后,系统根据这个分析结果,选择合适的语音合成参数(如语调、语速、音量等)来进行播报,从而更好地表达文本的情感内容。

[0097] 与现有技术相比,本申请使用Chat-Gpt或DeepSeek作为大模型来识别情绪特征具

有显著优势。传统的情绪识别方法通常依赖于预定义的情感词典或简单的规则,难以准确捕捉复杂的情感表达。而本申请利用大模型的强大语言理解能力,可以更全面地分析文本的语义和上下文,从而实现更准确的情绪识别。这种方法不仅能够识别出基本的情绪类型,还能捕捉到更细腻的情感变化,为听书体验带来质的提升。

[0098] 在上述的一些实施例中,本申请在实施的过程中,还存在难以准确识别用户情绪状态的问题。

[0099] 对此,本申请进一步提出了体征信息包括面部特征、心率特征中的任意一种的技术方案。

[0100] 本申请的技术方案通过引入用户体征信息,包括面部特征和心率特征,来更准确地识别用户的情绪状态。这种方法结合了生理和心理学的知识,可以提供更全面和客观的情绪评估。

[0101] 具体来说,本申请的技术方案可以通过以下方式实现:

[0102] 1. 面部特征识别:使用计算机视觉技术,如深度学习算法,来分析用户的面部表情。这可能包括检测眉毛位置、嘴角弧度、眼睛开合程度等微小变化。例如,微笑可能表示喜悦,皱眉可能表示困惑或不满。

[0103] 2. 心率特征分析:通过可穿戴设备或非接触式传感器来监测用户的心率变化。心率的快慢和变化模式可以反映用户的情绪状态。例如,心率加快可能表示兴奋或紧张,而心率平稳可能表示平静或专注。

[0104] 3. 数据融合:将面部特征和心率特征的数据进行综合分析,使用机器学习算法来建立更准确的情绪识别模型。这种多模态方法可以提高情绪识别的准确性和鲁棒性。

[0105] 4. 实时反馈:基于识别到的情绪状态,系统可以动态调整播报内容的语气、速度和音调,以更好地匹配用户的情绪需求。

[0106] 通过这种方法,本申请的技术方案可以更精确地捕捉用户的情绪变化,从而提供更个性化和适应性的听书体验。例如,当检测到用户心率较快且面部表情紧张时,系统可能会选择更平缓的语调来播报内容,以帮助用户放松。

[0107] 作为一种优选的实施方式,可以使用红外摄像头来捕捉用户的面部热图,结合面部表情分析算法,来更准确地识别面部特征。同时,可以使用光电容积描记法 (PPG) 技术的传感器来非接触式地监测用户的心率变化。这些技术的结合可以在不干扰用户正常听书体验的情况下,持续收集有关用户情绪状态的数据。

[0108] 具体实施例如下:

[0109] 在一个智能听书设备中,集成了高清摄像头和红外传感器。当用户开始使用听书功能时,设备首先会进行用户授权确认。得到授权后,设备开始采集用户的面部图像和心率数据。面部图像通过深度学习模型进行分析,提取关键面部特征点,如眉毛、眼睛、嘴角等的位置和形状变化。同时,心率数据通过傅里叶变换等信号处理技术,提取心率变异性 (HRV) 指标。

[0110] 这些数据被输入到一个预先训练好的多模态情绪识别模型中。该模型采用了注意力机制和长短时记忆网络 (LSTM),能够综合考虑面部特征和心率特征的时序变化。模型输出用户当前的情绪状态,如中性、喜悦、悲伤等。

[0111] 基于识别出的情绪状态,听书系统会实时调整播报参数。例如,当识别到用户情绪

为悲伤时,系统可能会选择更温和的语音合成音色,并略微降低播报速度,以给予用户更多理解和感受内容的时间。

[0112] 与现有技术相比,本申请的技术方案具有以下优势:

[0113] 1.多模态融合:通过结合面部特征和心率特征,提供了比单一模态更全面和准确的情绪识别。

[0114] 2.非侵入性:采用非接触式传感技术,避免了传统情绪识别方法可能带来的不适感。

[0115] 3.实时性:能够持续监测和分析用户情绪状态,实现动态调整,提供更加个性化的听书体验。

[0116] 4.适应性强:该方法可以适应不同用户的个体差异,通过持续学习来提高识别准确率。

[0117] 通过这种方式,本申请的技术方案有效解决了现有技术中难以准确捕捉用户情绪变化的问题,为智能听书系统提供了更加人性化和智能化的解决方案。

[0118] 如图2,本申请的一个实施例,提供了听书的智能播报系统,所述系统包括:

[0119] 图书获取单元1100用于获取待播报的电子图书;

[0120] 文本提取单元1200用于从待播报的电子图书中提取出多个文本;

[0121] 文本获取单元1300用于基于多个文本组合生成待播报文本;

[0122] 情绪特征获取单元1400用于基于大模型从待播报文本中识别出第一情绪特征;

[0123] 文本播报单元1500用于基于第一情绪特征对待播报文本进行播报。

[0124] 需要注意的是,本实施例提供的一种听书的智能播报系统与上述的一种听书的智能播报方法是基于相同的发明构思,因此上述的一种听书的智能播报方法的相关内容同样适用于的一种听书的智能播报系统内容,因此,此处不再赘述。

[0125] 如图3,本申请实施例还提供了一种电子设备,本电子设备包括:

[0126] 至少一个存储器;

[0127] 至少一个处理器;

[0128] 至少一个程序;

[0129] 程序被存储在存储器中,处理器执行至少一个程序以实现本公开实施上述的一种听书的智能播报方法。

[0130] 该电子设备可以为包括手机、平板电脑、个人数字助理(Personal Digital Assistant,PDA)、车载电脑等任意智能终端。

[0131] 下面对本申请实施例的电子设备进行详细介绍。

[0132] 处理器1600,可以采用通用的中央处理器(Central Processing Unit,CPU)、微处理器、应用专用集成电路(Application Specific Integrated Circuit,ASIC)、或者一个或多个集成电路等方式实现,用于执行相关程序,以实现本发明实施例所提供的技术方案;

[0133] 存储器1700,可以采用只读存储器(Read Only Memory,ROM)、静态存储设备、动态存储设备或者随机存取存储器(Random Access Memory,RAM)等形式实现。存储器1700可以存储操作系统和其他应用程序,在通过软件或者固件来实现本说明书实施例所提供的技术方案时,相关的程序代码保存在存储器1700中,并由处理器1600来调用执行本发明实施例的一种听书的智能播报方法。

- [0134] 输入/输出接口1800,用于实现信息输入及输出;
- [0135] 通信接口1900,用于实现本设备与其他设备的通信交互,可以通过有线方式(例如USB、网线等)实现通信,也可以通过无线方式(例如移动网络、WIFI、蓝牙等)实现通信;
- [0136] 总线2000,在设备的各个组件(例如处理器1600、存储器1700、输入/输出接口1800和通信接口1900)之间传输信息;
- [0137] 其中处理器1600、存储器1700、输入/输出接口1800和通信接口1900通过总线2000实现彼此之间在设备内部的通信连接。
- [0138] 本发明实施例还提供了一种存储介质,该存储介质是计算机可读存储介质,该计算机可读存储介质存储有计算机可执行指令,该计算机可执行指令用于使计算机执行上述一种听书的智能播报方法。
- [0139] 存储器作为一种非暂态计算机可读存储介质,可用于存储非暂态软件程序以及非暂态性计算机可执行程序。此外,存储器可以包括高速随机存取存储器,还可以包括非暂态存储器,例如至少一个磁盘存储器件、闪存器件、或其他非暂态固态存储器件。
- [0140] 在一些实施方式中,存储器可选包括相对于处理器远程设置的存储器,这些远程存储器可以通过网络连接至该处理器。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。
- [0141] 本发明描述的实施例是为了更加清楚的说明本发明实施例的技术方案,并不构成对于本发明实施例提供的技术方案的限定,本领域技术人员可知,随着技术的演变和新应用场景的出现,本发明实施例提供的技术方案对于类似的技术问题,同样适用。
- [0142] 本领域技术人员可以理解的是,图中示出的技术方案并不构成对本发明实施例的限定,可以包括比图示更多或更少的步骤,或者组合某些步骤,或者不同的步骤。
- [0143] 以上所描述的装置实施例仅仅是示意性的,其中作为分离部件说明的单元可以是或者也可以不是物理上分开的,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。
- [0144] 本领域普通技术人员可以理解,上文中所公开方法中的全部或某些步骤、系统、设备中的功能模块/单元可以被实施为软件、固件、硬件及其适当的组合。
- [0145] 本申请的说明书及上述附图中的术语“第一”、“第二”、“第三”、“第四”等(如果存在)是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数字在适当情况下可以互换,以便这里描述的本申请的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,例如,包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。
- [0146] 应当理解,在本申请中,“至少一个(项)”是指一个或者多个,“多个”是指两个或两个以上。“和/或”,用于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:只存在A,只存在B以及同时存在A和B三种情况,其中A,B可以是单数或者复数。字符“/”一般表示前后关联对象是一种“或”的关系。“以下至少一项(个)”或其类似表达,是指这些项中的任意组合,包括单项(个)或复数项(个)的任意组合。例如,a,b或c中的至少一项(个),可以表示:a,b,c,“a和b”,“a和c”,“b和c”,或“a和b和c”,其中a,b,c可以是单个,也可

以是多个。

[0147] 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,可以通过其它的方式实现。例如,以上所描述的装置实施例仅仅是示意性的,例如,单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,装置或单元的间接耦合或通信连接,可以是电性,机械或其它的形式。

[0148] 作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0149] 另外,在本申请各个实施例中的各功能单元可以集成在一个处理单元中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

[0150] 集成的单元如果以软件功能单元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全部或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,包括多指令用以使得一台电子设备(可以是个人计算机,服务器,或者网络设备等)执行本申请各个实施例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的介质。

[0151] 以上是对本申请实施例的较佳实施进行了具体说明,但本申请实施例并不局限于上述实施方式,熟悉本领域的技术人员在不违背本申请实施例精神的前提下还可作出种种的等同变形或替换,这些等同的变形或替换均包含在本申请实施例权利要求所限定的范围内。

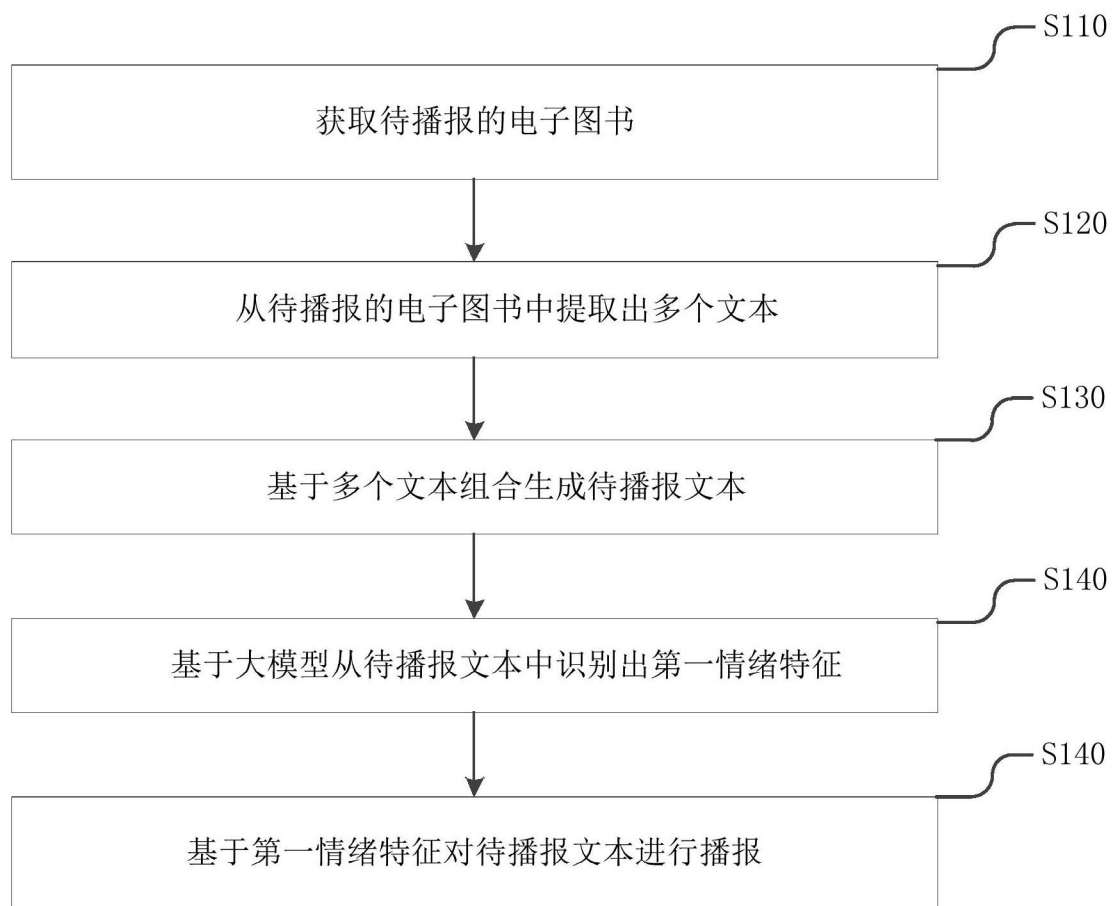


图1

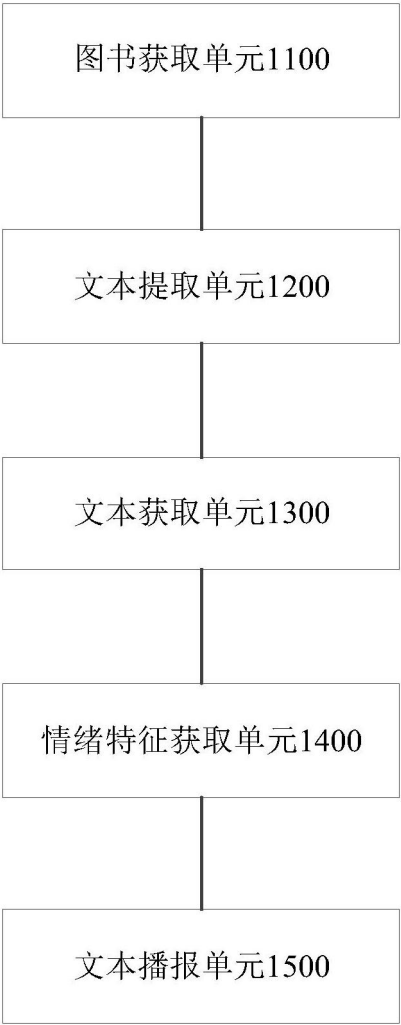


图2

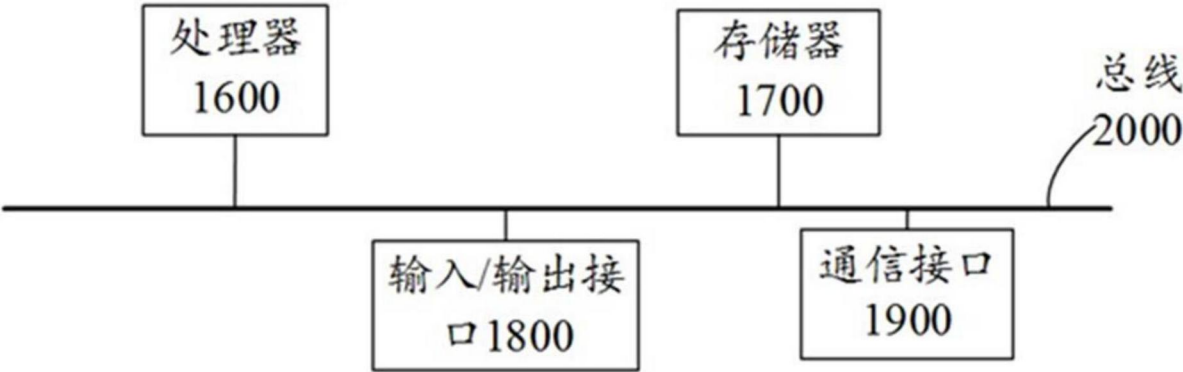


图3