(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 117407507 B
(45)授权公告日 2024.12.03

(21)申请号 202311416014.7
(22)申请日 2023.10.27
(65)同一申请的已公布的文献号
申请公布号 CN 117407507 A
(43)申请公布日 2024.01.16
(73)专利权人 北京百度网讯科技有限公司
地址 100085 北京市海淀区上地十街10号百度大厦2层
(72)发明人 赵一麟 沈智勇 陆勤 龚建
(74)专利代理机构 北京品源专利代理有限公司 11332
专利代理师 杨志欣
(51)Int.Cl.
G06F 16/332(2019.01)
G06F 16/33(2019.01)
G06F 16/338(2019.01)
G06F 16/34(2019.01)
(56)对比文件
CN 116821307 A,2023.09.29
审查员 解明惠

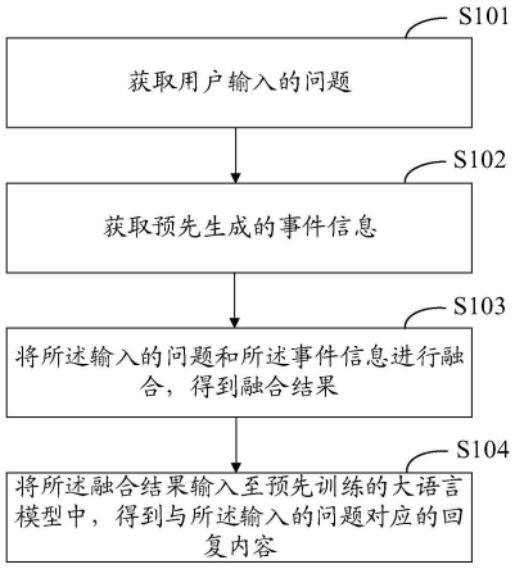
权利要求书3页 说明书13页 附图6页

(54)发明名称

基于大语言模型的事件处理方法、装置、设备及介质

(57)摘要

本公开提供了一种基于大语言模型的事件处理方法、装置、设备及介质,涉及大模型领域,具体涉及人工智能、大语言模型和智慧城市领域。具体实现方案为:获取用户输入的问题;获取预先生成的事件信息;将所述输入的问题和所述事件信息进行融合,得到融合结果;将所述融合结果输入至预先训练的大语言模型中,得到与所述输入的问题对应的回复内容。本公开实施例可以提高问题回复速度。



CN 117407507 B

1.一种基于大语言模型的事件处理方法,包括:

获取用户输入的问题;

获取预先生成的事件信息;

将所述输入的问题和所述事件信息进行融合,得到融合结果;

将所述融合结果输入至预先训练的大语言模型中,得到与所述输入的问题对应的回复内容;

所述获取预先生成的事件信息,包括:

获取预先设定的场景事件类型,场景事件类型通过所述用户根据需要解决的问题预先输入;

在采集的多个图像中,筛选与所述场景事件类型对应的场景图像;

在所述场景图像中,检测与所述场景事件类型对应的目标框;

获取所述场景图像的采集信息;

根据所述场景事件类型、所述场景图像的采集信息、所述目标框,生成事件信息;

在生成的多个事件信息中获取预先生成的事件信息。

2.根据权利要求1所述的方法,其中,所述在所述场景图像中,检测与所述场景事件类型对应的目标框,包括:

对所述场景事件类型的文本进行特征提取,得到场景文本特征;

对所述场景图像进行特征提取,得到多个场景图像特征;

在所述多个场景图像特征中,筛选与所述场景文本特征相似的目标图像特征;

根据所述目标图像特征,检测与所述场景事件类型对应的备选框;

计算所述场景事件类型对应的备选框与所述场景文本特征之间的相似度;

根据所述场景事件类型对应的备选框与所述场景文本特征之间的相似度,计算所述场景事件类型对应的备选框的置信度;

根据所述场景事件类型对应的备选框的置信度,在所述场景事件类型对应的备选框中,确定与所述场景事件类型对应的目标框。

3.根据权利要求1所述的方法,其中,所述将所述输入的问题和所述事件信息进行融合,包括:

对获取的多个事件信息进行筛选,得到与所述问题对应的目标信息;

将所述输入的问题与所述目标信息进行融合;所述目标信息的数量为至少一个。

4.根据权利要求1所述的方法,其中,所述将所述输入的问题和所述事件信息进行融合,得到融合结果,包括:

获取标准提示信息,所述标准提示信息中包括所述事件信息的槽位和问题槽位;

将所述事件信息填充到所述事件信息的槽位处;

将所述问题填充到所述问题槽位处,得到融合结果。

5.根据权利要求1所述的方法,还包括:

获取实时视频流以及所述实时视频流的采集信息;

从所述实时视频流中,获取多个采集的图像;

根据所述实时视频流的采集信息,确定各所述采集的图像的采集信息;其中,所述采集信息包括采集设备标识、采集设备位置和采集时间。

6.根据权利要求5所述的方法,其中,所述采集设备包括城市范围内的交通监控摄像头;所述采集设备位置包括所述交通监控摄像头拍摄位置。

7.根据权利要求1所述的方法,其中,所述预先训练的大语言模型,通过对处理通用任务的大语言模型经过指令调优确定,所述处理通用任务的大语言模型通过无监督训练、标注有监督微调和人类反馈强化学习训练得到。

8.根据权利要求1所述的方法,还包括:

对所述问题进行解析,得到关键词语;

将所述关键词语与所述问题对应的回复内容进行对应存储,以在接收到与关键词语对应的其他问题时,获取所述关键词语对应的回复内容进行反馈。

9.一种基于大语言模型的事件处理装置,包括:

输入问题获取模块,用于获取用户输入的问题;

事件信息获取模块,用于获取预先生成的事件信息;

事件问题融合模块,用于将所述输入的问题和所述事件信息进行融合,得到融合结果;

回复处理模块,用于将所述融合结果输入至预先训练的大语言模型中,得到与所述输入的问题对应的回复内容;

所述事件信息获取模块,包括:

场景事件类型获取单元,用于获取预先设定的场景事件类型,场景事件类型通过所述用户根据需要解决的问题预先输入;

场景图像筛选单元,用于在采集的多个图像中,筛选与所述场景事件类型对应的场景图像;

目标框检测单元,用于在所述场景图像中,检测与所述场景事件类型对应的目标框;

采集信息获取单元,用于获取所述场景图像的采集信息;

事件信息生成单元,用于根据所述场景事件类型、所述场景图像的采集信息和所述目标框,生成事件信息;

事件信息获取单元,用于在生成的多个事件信息中获取预先生成的事件信息。

10.根据权利要求9所述的装置,其中,所述目标框检测单元,包括:

文本特征提取子单元,用于对所述场景事件类型的文本进行特征提取,得到场景文本特征;

图像特征提取子单元,用于对所述场景图像进行特征提取,得到多个场景图像特征;

图像特征筛选子单元,用于在所述多个场景图像特征中,筛选与所述场景文本特征相似的目标图像特征;

备选框子单元,用于根据所述目标图像特征,检测与所述场景事件类型对应的备选框;

相似计算子单元,用于计算所述场景事件类型对应的备选框与所述场景文本特征之间的相似度;

置信度计算子单元,用于根据所述场景事件类型对应的备选框与所述场景文本特征之间的相似度,计算所述场景事件类型对应的备选框的置信度;

目标框检测子单元,用于根据所述场景事件类型对应的备选框的置信度,在所述场景事件类型对应的备选框中,确定与所述场景事件类型对应的目标框。

11.根据权利要求9所述的装置,其中,所述事件问题融合模块,包括:

事件信息筛选单元,用于对获取的多个事件信息进行筛选,得到与所述问题对应的目标信息;

目标信息融合单元,用于将所述输入的问题与所述目标信息进行融合;所述目标信息的数量为至少一个。

12.根据权利要求9所述的装置,其中,所述事件问题融合模块,包括:

提示信息获取单元,用于获取标准提示信息,所述标准提示信息中包括所述事件信息的槽位和问题槽位;

事件信息融合单元,用于将所述事件信息填充到所述事件信息的槽位处;

问题融合单元,用于将所述问题填充到所述问题槽位处,得到融合结果。

13.根据权利要求9所述的装置,还包括:

实时视频获取模块,用于获取实时视频流以及所述实时视频流的采集信息;

图像获取模块,用于从所述实时视频流中,获取多个采集的图像;

采集信息获取模块,用于根据所述实时视频流的采集信息,确定各所述采集的图像的采集信息;其中,所述采集信息包括采集设备标识、采集设备位置和采集时间。

14.根据权利要求13所述的装置,其中,所述采集设备包括城市范围内的交通监控摄像头;所述采集设备位置包括所述交通监控摄像头拍摄位置。

15.根据权利要求9述的装置,其中,所述预先训练的大语言模型,通过对处理通用任务的大语言模型经过指令调优确定,所述处理通用任务的大语言模型通过无监督训练、标注有监督微调和人类反馈强化学习训练得到。

16.根据权利要求9所述的装置,还包括:

关键词语提取模块,用于对所述问题进行解析,得到关键词语;

回复存储模块,用于将所述关键词语与所述问题对应的回复内容进行对应存储,以在接收到与关键词语对应的其他问题时,获取所述关键词语对应的回复内容进行反馈。

17. 一种电子设备,包括:

至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行权利要求1-8中任一项所述的基于大语言模型的事件处理方法。

18.一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行根据权利要求1-8中任一项所述的基于大语言模型的事件处理方法。

19.一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现根据权利要求1-8中任一项所述的基于大语言模型的事件处理方法。

基于大语言模型的事件处理方法、装置、设备及介质

技术领域

[0001] 本公开涉及大模型领域,具体涉及人工智能、大语言模型和人机交互领域,尤其涉及一种基于大语言模型的事件处理方法、装置、设备及介质。

背景技术

[0002] 近年来,智能系统借由对城市运行态势进行全感知、全互连、全分析、全分析、全响应和全应用,提升城市治理精细化和智能化水平。

[0003] 智能系统可以实现城市核心要素的连接、实现公共资源高效调配,以及城市事件精准处置。

发明内容

[0004] 本公开提供了一种基于大语言模型的事件处理方法、装置、设备及介质。

[0005] 根据本公开的一方面,提供了一种基于大语言模型的事件处理方法,包括:

[0006] 获取用户输入的问题;

[0007] 获取预先生成的事件信息;

[0008] 将所述输入的问题和所述事件信息进行融合,得到融合结果;

[0009] 将所述融合结果输入至预先训练的大语言模型中,得到与所述输入的问题对应的回复内容。

[0010] 根据本公开的一方面,提供了一种基于大语言模型的事件处理装置,包括:

[0011] 输入问题获取模块,用于获取用户输入的问题;

[0012] 事件信息获取模块,用于获取预先生成的事件信息;

[0013] 事件问题融合模块,用于将所述输入的问题和所述事件信息进行融合,得到融合结果;

[0014] 回复处理模块,用于将所述融合结果输入至预先训练的大语言模型中,得到与所述输入的问题对应的回复内容。

[0015] 根据本公开的另一方面,提供了一种电子设备,包括:

[0016] 至少一个处理器;以及

[0017] 与所述至少一个处理器通信连接的存储器;其中,

[0018] 所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行本公开任一实施例所述的基于大语言模型的事件处理方法。

[0019] 根据本公开的另一方面,提供了一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行本公开任一实施例所述的基于大语言模型的事件处理方法。

[0020] 根据本公开的另一方面,提供了一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现本公开任一实施例所述的基于大语言模型的事件处理方法。

[0021] 本公开实施例可以提高问题回复速度。

[0022] 应当理解,本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征,也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

附图说明

[0023] 附图用于更好地理解本方案,不构成对本公开的限定。其中:

[0024] 图1是根据本公开实施例公开的一种基于大语言模型的事件处理方法的流程图;

[0025] 图2是根据本公开实施例公开的另一种基于大语言模型的事件处理方法的流程图;

[0026] 图3是根据本公开实施例公开的另一种基于大语言模型的事件处理方法的流程图;

[0027] 图4是根据本公开实施例公开的一种事件信息生成方法的场景图;

[0028] 图5是根据本公开实施例公开的一种基于大语言模型的训练方法的场景图;

[0029] 图6是根据本公开实施例公开的一种回复内容的示意图;

[0030] 图7是根据本公开实施例公开的基于大语言模型的事件处理装置的结构示意图;

[0031] 图8是根据本公开实施例公开的基于大语言模型的事件处理方法的电子设备的框图。

具体实施方式

[0032] 以下结合附图对本公开的示范性实施例做出说明,其中包括本公开实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本公开的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0033] 图1是根据本公开实施例公开的一种基于大语言模型的事件处理方法的流程图,本实施例可以适用于大语言模型基于采集的事件进行回复的情况。本实施例方法可以由基于大语言模型的事件处理装置来执行,该装置可采用软件和/或硬件的方式实现,并具体配置于具有一定数据运算能力的电子设备中,该电子设备可以是客户端设备或服务器设备,客户端设备例如手机、平板电脑、车载终端和台式电脑等。

[0034] S101、获取用户输入的问题。

[0035] 问题可以是指用户在人机交互过程中输入的自然语言内容。其中,用户可以以文本、图像、语音或视频等至少一个类型的方式输入数据,对直接输入的数据进行识别,得到问题文本。客户端可以接收用户输入数据,并发送至本公开实施例的电子设备,该电子设备根据输入数据获取问题文本。可选的,用户输入的问题为问题文本。

[0036] S102、获取预先生成的事件信息。

[0037] 事件信息可以是指采集到已经存在或已经发生的真实事件的信息。事件可以是指在一定区域范围内发生的事件。例如,在某个城市内发生的事件。示例性的,事件可以是异常事件也可以是正常事件。又如,事件可以是交通事件、环保事件、经营事件或城管事件等。事件可以是某个领域的特定事件,也可以是多个领域的综合性事件。

[0038] 可以对事件进行采集,并汇总。其中,采集的事件信息的数量为至少一个。事件可

以是现实世界中感知到的数据,基于感知到的事件信息,进行后续处理,可以提高事件处理的实时性。需要说明的是,事件的采集、处理和展示等操作均经过合法合规授权。

[0039] S103、将所述输入的问题和所述事件信息进行融合,得到融合结果。

[0040] 融合结果作为最终输入到大语言模型中的内容。融合结果包括问题和事件信息,其中,包括的事件信息用于丰富可获取到的问题的上下文,可以作为问题的补充内容。

[0041] S104、将所述融合结果输入至预先训练的大语言模型中,得到与所述输入的问题对应的回复内容。

[0042] 大语言模型用于对自然语言内容进行理解,并生成对应的回复内容。可选的,大语言模型为问答模型,大语言模型用于对问题内容进行解答,并生成回复内容反馈给用户。将问题以及事件信息共同输入到大语言模型中,以使得大语言模型对问题以及上下文的事件信息进行理解,可以从事件信息理解到更加准确的内容,从而更加准确生成回复内容。

[0043] 在一个具体的例子中,事件信息包括:当天的城市A的交通拥堵事件,问题为:当天城市A的交通处理策略。回复内容为:当天城市A的某个时间内拥堵路段的交通资源调配策略,以及剩余时间内剩余路段的交通资源调配策略。

[0044] 根据本公开的技术方案,通过获取预先生成的事件信息,并与用户输入的问题进行融合,得到融合结果,将融合结果输入到大语言模型中进行问题理解生成,得到问题对应的回复内容,可以增加问题关联的上下文,并汇总事件,基于汇总的事件进行问题回复,可以提高回复内容的预测准确性,尤其是针对问题文本的应用场景,可以提高回复文本的处理准确性。

[0045] 图2是根据本公开实施例公开的另一种基于大语言模型的事件处理方法的流程图,基于上述技术方案进一步优化与扩展,并可以与上述各个可选实施方式进行结合。所述第一中间状态,包括下述至少一项:动作、动作描述内容、动作推理过程、中间结果和动作标识;其中,所述动作是所述大语言模型在预设的多个动作中筛选得到。

[0046] S201、获取预先设定的场景事件类型。

[0047] 场景事件类型用于区分采集的事件。通常场景事件类型可以采用关键词句进行表示。场景事件可以是典型事件,例如,城市管理或城市运行过程中发生的事件。示例性的,场景事件类型为交通事故,又如,场景事件类型为停车场停车,又如,场景事件类型为内涝现象。

[0048] 场景事件类型可以是用户根据需要解决的问题预先输入的。可以是第一用户输入场景事件类型,第二用户输入问题。或者可以是同一用户输入场景事件类型和问题。通常基于要解决的问题可以预先定义场景事件类型。示例性的,用户想要知道城市交通情况,可以针对性设定场景事件类型为交通拥堵、交通限行路段或交通事故等,根据城市实时生成的数据中,采集并生成与场景事件类型相关的事件信息。以便根据场景事件类型相关的事件信息,进行提问。

[0049] S202、在采集的多个图像中,筛选与所述场景事件类型对应的场景图像。

[0050] 采集的多个图像中,存在与场景事件类型无关的图像,其中,无关的图像中不存在与场景事件类型相关的内容。如前例,场景事件类型为交通事故,采集的图像包括小动物的图像、两车相撞的图像以及园区监控图像,其中,两车相撞的图像为交通事故对应的场景图像,其他图像为与交通事故无关的图像。

[0051] 可以对多个图像进行描述内容识别,确定各图像对应的描述信息,将描述信息与场景事件类型相同或相似的图像,确定为场景图像。又如,可以将场景事件类型的文本与各图像输入到预先训练的深度学习模型中进行内容匹配,得到场景事件类型对应的场景图像。此外,还有其他情况,对此不具体限定。

[0052] 可选的,所述基于大语言模型的事件处理方法,还包括:获取实时视频流以及所述实时视频流的采集信息;从所述实时视频流中,获取多个采集的图像;根据所述实时视频流的采集信息,确定各所述采集的图像的采集信息;其中,所述采集信息包括采集设备标识、采集设备位置和采集时间。

[0053] 实时视频流可以是指实时采集的视频流。对实时视频流进行截取或采样,得到多个采集的图像。根据实时视频流的采集信息,确定图像的采集时间。实时视频流的采集信息可以包括采集实时视频流的采集设备的标识、采集设备位置和采集时间,其中,图像和实时视频流采集设备的标识相同,采集设备位置相同。实时视频流记录录制每一帧的时间,可以根据图像在实时视频流中的次序或时间确定图像的采集时间。

[0054] 通过从实时视频流中获取采集的图像,以及根据实时视频流的采集信息,确定采集的图像的采集信息,可以提高事件的实时性,从而提高事件感知速度。

[0055] 可选的,所述采集设备包括城市范围内的交通监控摄像头;所述采集设备位置包括所述交通监控摄像头拍摄位置。

[0056] 交通监控摄像头用于监控道路异常。采集设备位置可以包括交通监控摄像头的所在位置和/或拍摄位置等。示例性的,采集设备位置可以是路口A。在一个具体的例子中,问题可以是找到交通拥堵最高发的路段。

[0057] 通过将应用场景设置于交通监控事件,可以汇总城市场景内的交通事件,并在此基础上对交通监控事件的问题进行回复,可以对城市实时交通事件进行提取和总结,减少城市交通事件的搜索和筛选难度,并提高城市交通事件的问题回复准确性。

[0058] S203、在所述场景图像中,检测与所述场景事件类型对应的目标框。

[0059] 目标框可以是指场景事件类型对应的图像内容区域。目标框内的图像内容与场景事件类型相同或相似,例如,场景事件类型为不文明养宠物,目标框内的图像可以是道路上未栓绳的宠物狗的图像。场景图像是筛选得到与场景事件类型相关的图像,场景图像中存在目标框与场景事件类型对应。

[0060] 可选的,所述在所述场景图像中,检测与所述场景事件类型对应的目标框,包括:对所述场景事件类型的文本进行特征提取,得到场景文本特征;对所述场景图像进行特征提取,得到多个场景图像特征;在所述多个场景图像特征中,筛选与所述场景文本特征相似的目标图像特征;根据所述目标图像特征,检测与所述场景事件类型对应的备选框;计算所述场景事件类型对应的备选框与所述场景文本特征之间的相似度;根据所述场景事件类型对应的备选框与所述场景文本特征之间的相似度,计算所述场景事件类型对应的备选框的置信度;根据所述场景事件类型对应的备选框的置信度,在所述场景事件类型对应的备选框中,确定与所述场景事件类型对应的目标框。

[0061] 场景事件类型的文本可以是指该文本的内容为场景事件类型,对场景事件类型的文本进行特征提取,得到场景文本特征,示例性的,可以采用BERT(Bidirectional Encoder Representation from Transformers,基于转换器模型的双向编码器表示)作为文本编码

器,对场景事件类型的文本进行特征提取,得到场景文本特征。对场景图像进行特征提取,得到多个场景图像特征,示例性的,可以采用Vision Transformer(视觉转换器模型),作为图像编码器,对场景图像进行特征提取,得到多个场景图像特征。

[0062] 场景文本特征和场景图像特征之间的相似度用于表示文本特征代表的内容和图像特征的内容之间的相似程度。可以将场景文本特征和场景图像特征输入到Cross Modal Query Selection(跨模态查询选择网络)中,得到场景文本特征和各场景图像特征之间的相似度。根据相似度,筛选最相似的多个场景图像特征确定为目标图像特征。例如,900个场景图像特征中,筛选相似度最高的300个场景图像特征,确定为目标图像特征。将目标图像特征输入到解码器中,得到备选框,具体是得到备选框位置,通常采用备选框的矩形四个顶点的坐标表示备选框位置。针对每个备选框,计算场景文本特征与备选框之间的相似度,并作为备选框的置信度,示例性的,可以根据确定备选框的场景图像特征与该场景文本特征之间的相似度,作为场景文本特征与备选框之间的相似度。

[0063] 通过针对预先设定的任意场景类型的文本,建立图像与文本之间的匹配能力,实现开放性词汇目标感知,根据输入图像和待感知的文本描述,在图像中找到与文本描述对应的区域,框选出该区域边界四点坐标,实现开放词汇目标检测,从而可以接受任意文本描述作为输入,实现将目标检测的类别数扩展至近似无穷,丰富场景类型的内容,增加应用场景。

[0064] S204、获取所述场景图像的采集信息。

[0065] 场景图像的采集信息可以包括下述至少一项:指采集该场景图像的时间、位置、方式、拍摄内容和拍摄范围等。在获取采集的多个图像时,同时可以获取到每个图像的采集信息。通常在图像采集时,记录该图像的采集信息。

[0066] S205、根据所述场景事件类型、所述场景图像的采集信息和所述目标框,生成事件信息。

[0067] 生成的事件信息属于场景事件类型,场景图像的采集信息用于确定事件信息所代表的事件的发生的内容、状态和过程等。目标框用于确定事件信息的依据,即表明事件信息真实存在。通过场景事件类型、采集信息、目标框,确定事件信息,可以从多维度记录该事件关联的信息,丰富事件的内容,还可以根据目标框进行溯源,便于定位事件的发生过程和发生位置。

[0068] S206、获取用户输入的问题。

[0069] S207、在生成的多个事件信息中获取预先生成的事件信息。

[0070] 预先生成多个事件信息。针对多个事件信息,可以获取全部生成的事件信息,或者筛选部分生成的事件信息,其中,筛选条件可以是时间,也可以是相似度。若未生成事件信息,获取预先生成的事件信息可以为空。其中,S201-S205与S206的顺序可以调换。通常S201-S205执行完成之后,执行S206。

[0071] S208、将所述输入的问题和所述事件信息进行融合,得到融合结果。

[0072] 可选的,所述将所述输入的问题和所述事件信息进行融合,得到融合结果,包括:获取标准提示信息,所述标准提示信息中包括所述事件信息的槽位和问题槽位;将所述事件信息填充到所述事件信息的槽位处;将所述问题填充到所述问题槽位处,得到融合结果。

[0073] 标准提示信息(prompt)是向模型输入的一段文本提示,其中包含用户要查询的信

息或问题的关键词和上下文,以便模型更好地理解用户的意图并给出更准确的回复。标准提示信息用于添加事件信息和问题,形成大语言模型的输入数据。事件信息的槽位用于填充事件信息,问题槽位用于填充问题。融合结果可以是指标准提示信息填充事件信息和问题之后得到的文本。

[0074] 在一个具体的例子中,标准提示信息可以是:你现在是市长的助理,以下是城市内今天发生的异常事件:{事件信息},请你处理{问题}。

[0075] 通过标准提示信息适配不同问题和事件信息,使得大语言模型根据提示模板快速理解问题以及事件信息,生成问题对应的回复内容,省略大语言模型针对不同问题理解不同场景的语义才能实现生成相应回复内容的训练步骤,快速实现大语言模型的调优过程,降低大语言模型调优更新的人力成本,同时可以实时根据问题快速获取对应的提示模板,提高大语言模型的更新实时性,增加大语言模型与问题之间的普适性和通用性。

[0076] S209、将所述融合结果输入至预先训练的大语言模型中,得到与所述输入的问题对应的回复内容。

[0077] 根据本公开的技术方案,通过采集多个图像,并从中晒寻从与场景事件类型对应的场景图像,以及在场景图像中检测场景事件类型对应的目标框,实现图像事件的获取,根据场景图像的采集信息、场景事件类型和目标框,生成事件信息,可以真实实时生成事件,提高事件的获取效率和真实性。

[0078] 图3是根据本公开实施例公开的另一种基于大语言模型的事件处理方法的流程图,基于上述技术方案进一步优化与扩展,并可以与上述各个可选实施方式进行结合。所述将所述输入的问题和所述事件信息进行融合,具体化为:对获取的多个事件信息进行筛选,得到与所述问题对应的目标信息;将所述输入的问题与所述目标信息进行融合;所述目标信息的数量为至少一个。

[0079] S301、获取用户输入的问题。

[0080] S302、获取预先生成的事件信息。

[0081] S303、对获取的多个事件信息进行筛选,得到与所述问题对应的目标信息。

[0082] 目标信息为与问题有关的事件信息。示例性的,问题为询问洪涝情况,事件信息为不文明养宠物的事件。该事件信息与问题无关。可以获取到多个事件信息,该多个事件信息中可以存在与问题无关的事件信息,可以对事件信息进行筛选,仅保留与目标信息有关的内容,可以减少冗余输入数据。

[0083] S304、将所述输入的问题与所述目标信息进行融合,得到融合结果;所述目标信息的数量为至少一个。

[0084] 筛选得到的目标信息的数量可以有多个。将全部的目标信息与输入问题进行融合,可以丰富输入数据。

[0085] 具体的,标准提示信息用于添加目标信息和问题,形成大语言模型的输入数据。事件信息的槽位用于填充目标信息,问题槽位用于填充问题。融合结果可以是指标准提示信息填充目标信息和问题之后得到的文本。

[0086] 如前例,标准提示信息可以是:你现在是市长的助理,以下是城市内今天发生的异常事件:{目标信息},请你处理{问题}。

[0087] S305、将所述融合结果输入至预先训练的大语言模型中,得到与所述输入的问题

对应的回复内容。

[0088] 可选的,所述预先训练的大语言模型,通过对处理通用任务的大语言模型经过指令调优确定,所述处理通用任务的大语言模型通过无监督训练、标注有监督微调和人类反馈强化学习训练得到。

[0089] 处理通用任务的大语言模型可以是不区分具体应用领域和具体业务场景任务,可以实现通用性的理解生成功能的大语言模型。处理通用任务的大语言模型可以是预训练模型,还可以具体经过微调以更好适配具体的应用场景。

[0090] 针对得到处理通用任务的大语言模型来说,其中,无监督预训练,该阶段通常基于海量互联网中文语料,将其中的若干词汇替换成[MASK],然后基于上下文预测被[MASK]掉的位置的原文。经过无监督预训练的语言模型通常具有比较好的WordEmbedding(词语嵌入式)以及上下文建模能力。有监督微调主要在上一无监督预训练阶段的基础上加入有标注数据进行微调,以提升模型输出的有用性和无害性。

[0091] 人类反馈强化学习(Reinforcement Learning from Human Feedback,RLHF)是根据人类反馈来对模型的响应进行排序标注(如,根据人类偏好选择文本简介)。然后,用这些带标注的响应来训练偏好模型,该模型用于返回RL(强化学习)优化器的标量奖励。最后,通过强化学习训练对话代理来模拟偏好模型。可以将大语言模型的输出,输入到预先训练的偏好模型中,得到输出结果的得分,并根据得分不断调整大语言模型的参数,以使得大语言模型的输出更加符合人类偏好的表达形式,和更准确的语法等。

[0092] 指令调优(Instrction tuning)用于提供更加明确的指令,以使大语言模型去理解并输出正确的回复内容。指令调优主要以指令的形式指导模型执行特定操作或完成任务。Instruction可以是计算机程序或脚本,也可以是用户编写的指导性文本。Instruction的目的是告诉模型如何处理数据或执行某个操作,而不是简单地提供上下文或任务相关信息。

[0093] 例如,输入Input:2023年X月X日在XX路口发生了道路拥堵事件,这起事件的地点是什么?标签Target:西直门路口。

[0094] 经过instruction tuning的模型将更擅长执行跟训练任务相似的下游任务,而不需要对每个样本(example)进行单独训练。

[0095] 通过多层训练过程,提高大语言模型的理解能力和生成能力,从而提高大语言模型的输出准确性。

[0096] 可选的,所述基于大语言模型的事件处理方法,还包括:对所述问题进行解析,得到关键词语;将所述关键词语与所述问题对应的回复内容进行对应存储,以在接收到与关键词语对应的其他问题时,获取所述关键词语对应的回复内容进行反馈。

[0097] 从问题中提取关键词语。关键词语用于描述该问题的内容。将关键词语与回复内容进行对应存储,可以根据该关键词语查询到回复内容。示例性的,可以采用将关键词语作为键key,将回复内容作为值value,将二者以键值形式进行存储。后续接收到相似的问题,可以直接根据关键词语查询到对应的回复内容,提高回复速度。

[0098] 通过解析问题得到关键词语,并将关键词语与该问题对应的回复内容进行对应存储,减少冗余计算操作,可以加快回复速度。

[0099] 此外,大语言模型还可以以多轮对话形式进行问答,将多轮对话的内容添加到标

准提示信息中,增加输入的内容,并输入到大语言模型中,提高大语言模型的输入的准确性,以及输出的准确性。

[0100] 根据本公开的技术方案,通过基于问题对事件信息进行筛选,得到目标信息,并将筛选得到的多个目标信息与输入的问题进行融合,作为输入数据,输入到大语言模型中,可以减少冗余输入数据,提高数据处理速度,同时可以增加问题相关的上下文内容,增加输入的代表性,提高大语言模型的输出准确性。

[0101] 图4-图6是根据本公开实施例公开的一种基于大语言模型的事件处理方法的场景图。本公开实施例主要划分四个阶段:(1)开放词汇目标感知(2)提示词汇总(3)大语言模型(4)问答解析。

[0102] (1)开放词汇目标感知的目的是通过建立图像和文本的匹配能力,根据采集的图像和待感知的场景事件类型的文本描述,在图像中找到与文本描述对应的区域,框选出该区域边界四点坐标。与传统的固定类别目标检测不同的是,开放词汇目标检测可以接受任意文本描述作为输入,因此可以将目标检测的类别数扩展至近似无穷。

[0103] 如图4所示,可以采用BERT作为文本编码器,对场景事件类型的文本进行特征提取,得到场景文本特征。可以采用Vision Transformer作为图像编码器,对场景图像进行特征提取,得到多个场景图像特征。将场景文本特征和场景图像特征输入到Cross Modal Query Selection,得到场景文本特征和各场景图像特征之间的相似度。从900个场景图像特征中,筛选相似度最高的300个场景图像特征,确定为目标图像特征。将目标图像特征输入到多模态解码器中,得到备选框,具体是得到备选框位置,通常采用备选框的矩形四个顶点的坐标表示备选框位置。针对每个备选框,计算场景文本特征与备选框之间的相似度,并作为备选框的置信度,可以根据确定备选框的场景图像特征与该场景文本特征之间的相似度,作为场景文本特征与备选框之间的相似度。根据备选框的置信度,筛选目标框。

[0104] (2)提示词汇总模块

[0105] 将开放词汇感知的检测结果(上述目标框位置、采集时间、采集设备标识和采集设备位置等)转化为大语言模型输入的自然语言形式。原始检测结果主要由一系列键值对构成,如:

[0106] {

[0107] camera_id:123123123,

[0108] camera_name:XX路口,

[0109] event:车辆逆行,

[0110] bbox:[123,234,345,456],

[0111] time:X-X-X,

[0112] }

[0113] 设定提示词模板用于将检测结果中的关键词组装进自然语言输入,如上述键值对所对应的提示词模板可以为:

[0114] “{time},在位置{camera_name}发生了事件{event}”

[0115] 在每次事件发生后,则按照以上模板填充为自然语言,得到事件信息,并暂存,用于后续输入大语言模型进行汇总。

[0116] (3)大语言模型

[0117] 大语言模型使用基于中文语料训练的理解生成模型,大语言模型的预训练任务分为三个阶段:

[0118] (1)无监督预训练,如图5所示,该阶段通常基于海量互联网中文语料,将其中的若干词汇替换成[MASK],然后基于上下文预测被[MASK]掉的位置的原文。经过无监督预训练的语言模型通常具有比较好的WordEmbedding以及上下文建模能力。

[0119] (2)有监督微调

[0120] 有监督微调主要在上一阶段的基础上加入有标注数据进行微调,以提升模型输出的有用性和无害性。

[0121] (3)人类反馈强化学习(RLHF)

[0122] 人类反馈强化学习(Reinforcement Learning from Human Feedback,RLHF)是根据人类反馈来对模型的响应进行排序标注(如,根据人类偏好选择文本简介)。然后,用这些带标注的响应来训练偏好模型,该模型用于返回RL(强化学习)优化器的标量奖励。最后,通过强化学习训练对话代理来模拟偏好模型。可以将大语言模型的输出,输入到预先训练的偏好模型中,得到输出结果的得分,并根据得分不断调整大语言模型的参数,以使得大语言模型的输出更加符合人类偏好的表达形式,和更准确的语法等。

[0123] 经过上述流程,得到了一个处理通用任务的大语言模型,然后需要基于Instruction Tuning的方法在业务目标上进行微调。

[0124] 指令调优主要以指令的形式指导模型执行特定操作或完成任务。Instruction可以是计算机程序或脚本,也可以是用户编写的指导性文本。Instruction的目的是告诉模型如何处理数据或执行某个操作,而不是简单地提供上下文或任务相关信息。

[0125] 例如,输入Input:2023年X月X日在XX路口发生了道路拥堵事件,这起事件的地点是什么?标签Target:西直门路口。

[0126] (4)问答解析

[0127] 问答解析可以提供与用户交互的界面,主要功能是将产出的结构化结果(事件信息)以及用户输入的其他query输入大语言模型,并进行结果解析。

[0128] 本方案会预置一系列query模板,用于周期性输出默认内容。

[0129] 例如,你现在是市长的助理,以下是城市内今天发生的异常事件:{事件信息},请你总结今天的高发事件与异常事件,形成一份报告。回复内容可以参考如图6所示。

[0130] 此外该模块接收用户的开放式query,输入大语言模型以多轮对话的形式进行问答,可进一步支撑用户的开放式需求。

[0131] 基于预置代码对模型输出结果进行结构化解析,提取关键字进行键值对存储。当用户输入query时,进行关键字筛选,与存储的键进行匹配,若配对成功则输出对应的值。

[0132] 本公开自动化地感知与汇总城市场景内的异常事件,并可以基于相关内容进行对话。相较于传统的感知事件并在业务系统中陈列展示的方案,本公开实施例可以高效抽取一些开放性问题,如“找到交通拥堵最高发的路段”,“八月比七月的城市内涝现象是否得到了改善”等。

[0133] 根据本公开的实施例,图7是本公开实施例中的基于大语言模型的事件处理装置的结构图,本公开实施例适用于大语言模型基于采集的事件进行回复的情况。该装置采用软件和/或硬件实现,并具体配置于具备一定数据运算能力的电子设备中。

[0134] 如图7所示的一种基于大语言模型的事件处理装置700,包括:输入问题获取模块701、事件信息获取模块702、事件问题融合模块703和回复处理模块704。其中,

[0135] 输入问题获取模块701,用于获取用户输入的问题;

[0136] 事件信息获取模块702,用于获取预先生成的事件信息;

[0137] 事件问题融合模块703,用于将所述输入的问题和所述事件信息进行融合,得到融合结果;

[0138] 回复处理模块704,用于将所述融合结果输入至预先训练的大语言模型中,得到与所述输入的问题对应的回复内容。

[0139] 根据本公开的技术方案,通过获取预先生成的事件信息,并与用户输入的问题进行融合,得到融合结果,将融合结果输入到大语言模型中进行问题理解生成,得到问题对应的回复内容,可以增加问题关联的上下文,并汇总事件,基于汇总的事件进行问题回复,可以提高回复内容的预测准确性,尤其是针对问题文本的应用场景,可以提高回复文本的处理准确性。

[0140] 进一步的,所述事件信息获取模块,包括:场景事件类型获取单元,用于获取预先设定的场景事件类型;场景图像筛选单元,用于在采集的多个图像中,筛选与所述场景事件类型对应的场景图像;目标框检测单元,用于在所述场景图像中,检测与所述场景事件类型对应的目标框;采集信息获取单元,用于获取所述场景图像的采集信息;事件信息生成单元,用于根据所述场景事件类型、所述场景图像的采集信息和所述目标框,生成事件信息;事件信息获取单元,用于在生成的多个事件信息中获取预先生成的事件信息。

[0141] 进一步的,所述目标框检测单元,包括:文本特征提取子单元,用于对所述场景事件类型的文本进行特征提取,得到场景文本特征;图像特征提取子单元,用于对所述场景图像进行特征提取,得到多个场景图像特征;图像特征筛选子单元,用于在所述多个场景图像特征中,筛选与所述场景文本特征相似的目标图像特征;备选框子单元,用于根据所述目标图像特征,检测与所述场景事件类型对应的备选框;相似计算子单元,用于计算所述场景事件类型对应的备选框与所述场景文本特征之间的相似度;置信度计算子单元,用于根据所述场景事件类型对应的备选框与所述场景文本特征之间的相似度,计算所述场景事件类型对应的备选框的置信度;目标框检测子单元,用于根据所述场景事件类型对应的备选框的置信度,在所述场景事件类型对应的备选框中,确定与所述场景事件类型对应的目标框。

[0142] 进一步的,所述事件问题融合模块703,包括:事件信息筛选单元,用于对获取的多个事件信息进行筛选,得到与所述问题对应的目标信息;目标信息融合单元,用于将所述输入的问题与所述目标信息进行融合;所述目标信息的数量为至少一个。

[0143] 进一步的,所述事件问题融合模块703,包括:提示信息获取单元,用于获取标准提示信息,所述标准提示信息中包括所述事件信息的槽位和问题槽位;事件信息融合单元,用于将所述事件信息填充到所述事件信息的槽位处;问题融合单元,用于将所述问题填充到所述问题槽位处,得到融合结果。

[0144] 进一步的,基于大语言模型的事件处理装置,还包括:实时视频获取模块,用于获取实时视频流以及所述实时视频流的采集信息;图像获取模块,用于从所述实时视频流中,获取多个采集的图像;采集信息获取模块,用于根据所述实时视频流的采集信息,确定各所述采集的图像的采集信息;其中,所述采集信息包括采集设备标识、采集设备位置和采集时

间。

[0145] 进一步的,所述采集设备包括城市范围内的交通监控摄像头;所述采集设备位置包括所述交通监控摄像头拍摄位置。

[0146] 进一步的,所述预先训练的大语言模型,通过对处理通用任务的大语言模型经过指令调优确定,所述处理通用任务的大语言模型通过无监督训练、标注有监督微调和人类反馈强化学习训练得到。

[0147] 进一步的,基于大语言模型的事件处理装置,还包括:关键词语提取模块,用于对所述问题进行解析,得到关键词语;回复存储模块,用于将所述关键词语与所述问题对应的回复内容进行对应存储,以在接收到与关键词语对应的其他问题时,获取所述关键词语对应的回复内容进行反馈。

[0148] 上述基于大语言模型的事件处理装置可执行本公开任意实施例所提供的基于大语言模型的事件处理方法,具备执行基于大语言模型的事件处理方法相应的功能模块和有益效果。

[0149] 本公开的技术方案中,所涉及的用户个人信息的收集、存储、使用、加工、传输、提供和公开等处理,均符合相关法律法规的规定,且不违背公序良俗。

[0150] 根据本公开的实施例,本公开还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

[0151] 图8示出了可以用来实施本公开的实施例的示例电子设备800的示意性区域图。电子设备旨在表示各种形式的数字计算机,诸如,膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置,诸如,个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例,并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0152] 如图8所示,设备800包括计算单元801,其可以根据存储在只读存储器(ROM)802中的计算机程序或者从存储单元808加载到随机访问存储器(RAM)803中的计算机程序,来执行各种适当的动作和处理。在RAM 803中,还可存储设备800指令所需的各种程序和数据。计算单元801、ROM 802以及RAM 803通过总线804彼此相连。输入/输出(I/O)接口805也连接至总线804。

[0153] 设备800中的多个部件连接至I/O接口805,包括:输入单元806,例如键盘、鼠标等;输出单元807,例如各种类型的显示器、扬声器等;存储单元808,例如磁盘、光盘等;以及通信单元809,例如网卡、调制解调器、无线通信收发机等。通信单元809允许设备800通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0154] 计算单元801可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元801的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI)计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元801执行上文所描述的各个方法和处理,例如基于大语言模型的事件处理方法。例如,在一些实施例中,基于大语言模型的事件处理方法可被实现为计算机软件程序,其被有形地包含于机器可读介质,例如存储单元808。在一些实施例中,计算机程序的部分或者全部可以经由ROM 802和/或通信单元809而被载入

和/或安装到设备800上。当计算机程序加载到RAM 803并由计算单元801执行时,可以执行上文描述的基于大语言模型的事件处理方法的一个或多个步骤。备选地,在其他实施例中,计算单元801可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行基于大语言模型的事件处理方法。

[0155] 本文中以上描述的系统和技术的各种实施方式可以在数字电子电路系统、集成电路系统、现场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准对象(ASSP)、芯片上系统的系统(SOC)、复杂可编程逻辑设备(CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0156] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器,使得程序代码当由处理器或控制器执行时使流程图和/或区域图中所规定的功能/指令被实施。程序代码可以完全在机器上执行、部分地在机器上执行,作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0157] 在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0158] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,该计算机具有:用于向用户显示信息的显示装置(例如,CRT(阴极射线管)或者LCD(液晶显示器)监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0159] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术的实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:局域网(LAN)、广域网(WAN)区块链网络和互联网。

[0160] 计算机系统可以包括客户端和服务器。客户端和服务器一般远离彼此并且通常通

过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务器的关系。服务器可以是云服务器,又称为云计算服务器或云主机,是云计算服务体系中的一项主机产品,以解决了传统物理主机与VPS服务中,存在的管理难度大,业务扩展性弱的缺陷。服务器也可以为分布式系统的服务器,或者是结合了区块链的服务器。

[0161] 人工智能是研究使计算机来模拟人的某些思维过程和智能行为(如学习、推理、思考、规划等)的学科,既有硬件层面的技术也有软件层面的技术。人工智能硬件技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理等技术;人工智能软件技术主要包括计算机视觉技术、语音识别技术、自然语言处理技术及机器学习/深度学习技术、大数据处理技术、知识图谱技术等几大方向。

[0162] 云计算(cloud computing),指的是通过网络接入弹性可扩展的共享物理或虚拟资源池,资源可以包括服务器、指令系统、网络、软件、应用和存储设备等,并可以按需、自服务的方式对资源进行部署和管理的技术体系。通过云计算技术,可以为人工智能、区块链等技术应用、模型训练提供高效强大的数据处理能力。

[0163] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例如,本发公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本公开提供的技术方案所期望的结果,本文在此不进行限制。

[0164] 上述具体实施方式,并不构成对本公开保护范围的限制。本领域技术人员应该明白的是,根据设计要求和其他因素,可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等,均应包含在本公开保护范围之内。

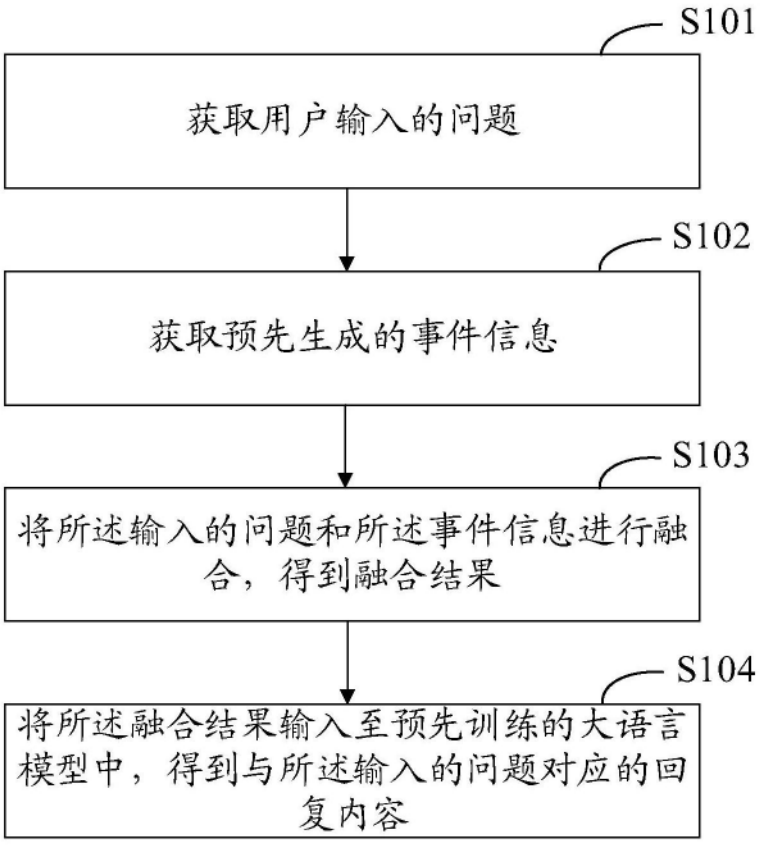

S101
获取用户输入的问题
S102
获取预先生成的事件信息
S103
将所述输入的问题和所述事件信息进行融合，得到融合结果
S104
将所述融合结果输入至预先训练的大语言模型中，得到与所述输入的问题对应的回复内容

图1

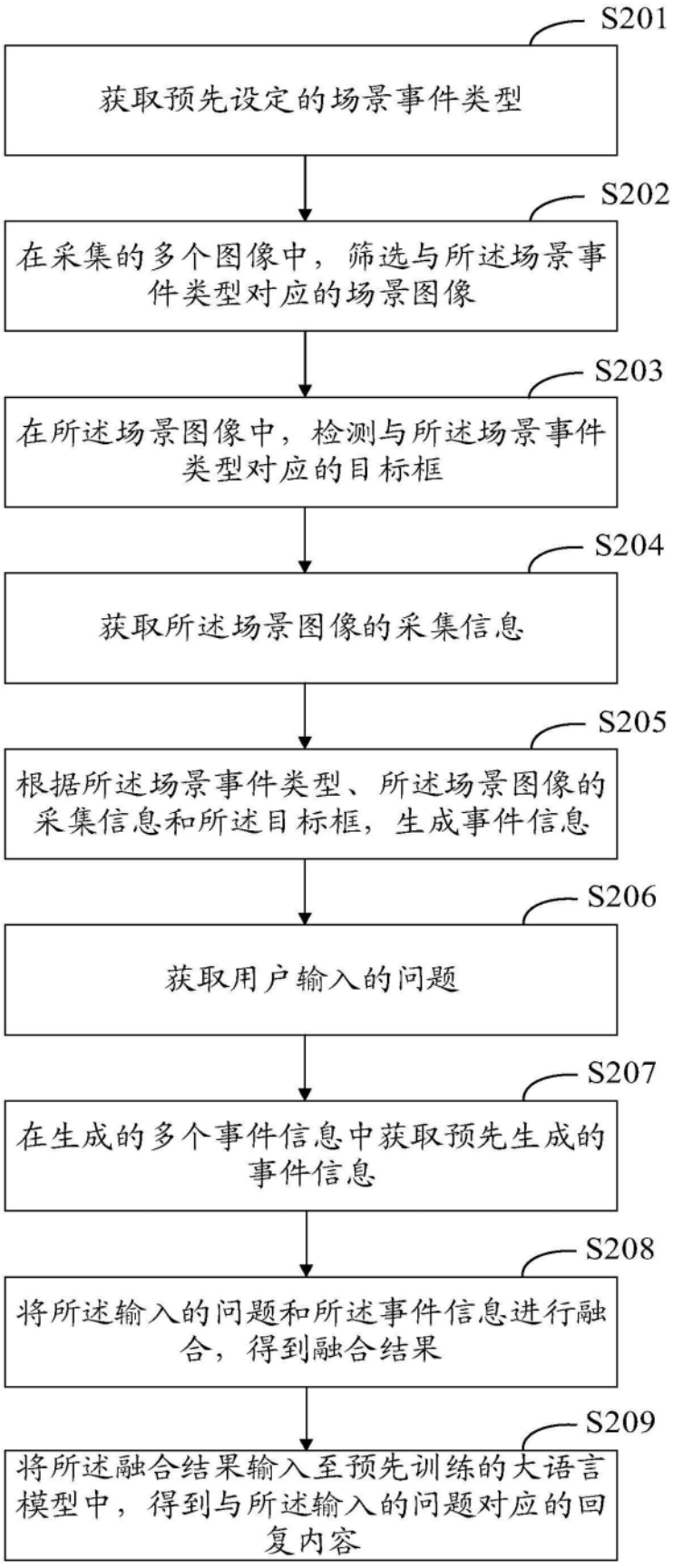

S201
获取预先设定的场景事件类型
S202
在采集的多个图像中，筛选与所述场景事件类型对应的场景图像
S203
在所述场景图像中，检测与所述场景事件类型对应的目标框
S204
获取所述场景图像的采集信息
S205
根据所述场景事件类型、所述场景图像的采集信息和所述目标框，生成事件信息
S206
获取用户输入的问题
S207
在生成的多个事件信息中获取预先生成的事件信息
S208
将所述输入的问题和所述事件信息进行融合，得到融合结果
S209
将所述融合结果输入至预先训练的大语言模型中，得到与所述输入的问题对应的回复内容

图2

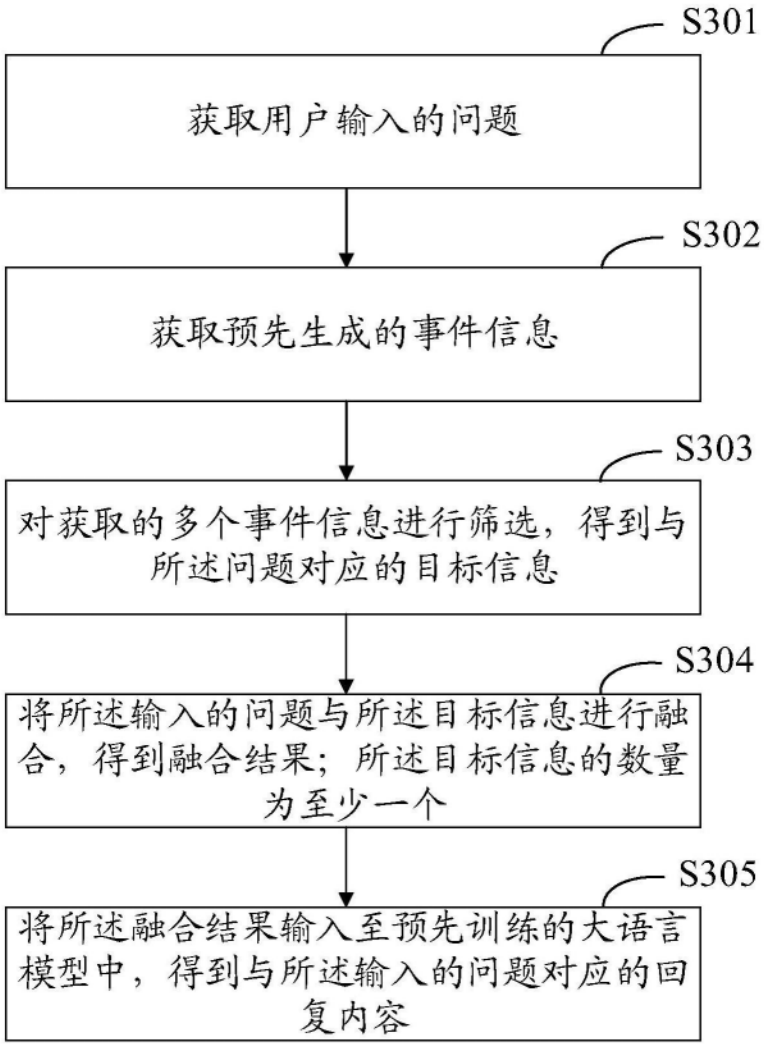

S301
获取用户输入的问题
S302
获取预先生成的事件信息
S303
对获取的多个事件信息进行筛选，得到与所述问题对应的目标信息
S304
将所述输入的问题与所述目标信息进行融合，得到融合结果；所述目标信息的数量为至少一个
S305
将所述融合结果输入至预先训练的大语言模型中，得到与所述输入的问题对应的回复内容

图3

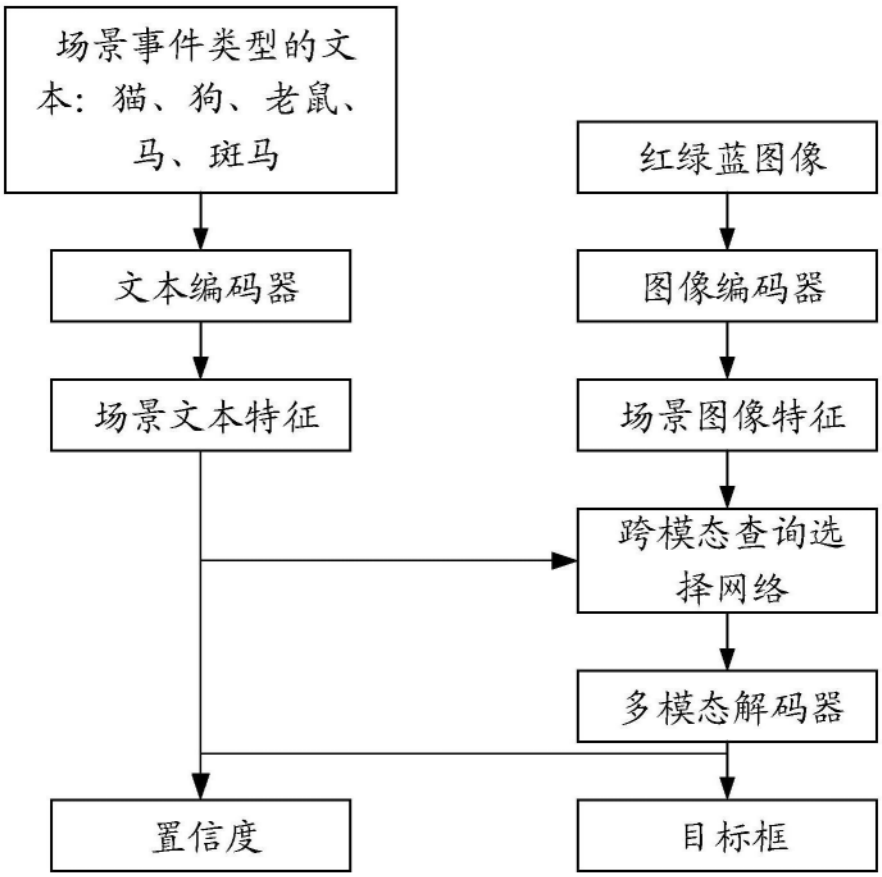

场景事件类型的文本：猫、狗、老鼠、马、斑马
红绿蓝图像
文本编码器
图像编码器
场景文本特征
场景图像特征
跨模态查询选择网络
多模态解码器
置信度
目标框

图4

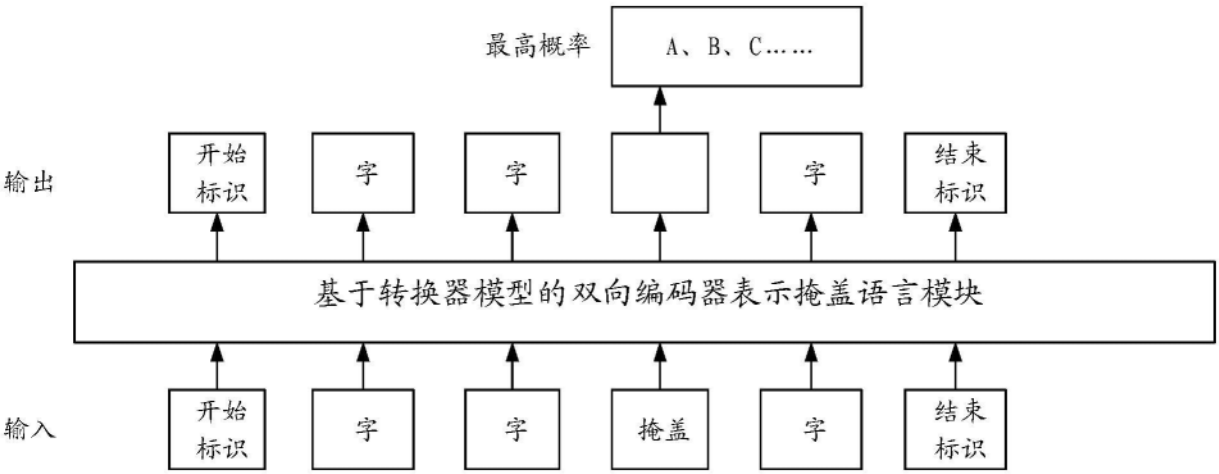

最高概率
A、B、C……
输出
开始标识
字
字
字
结束标识
基于转换器模型的双向编码器表示掩盖语言模块
输入
开始标识
字
字
掩盖
字
结束标识

图5

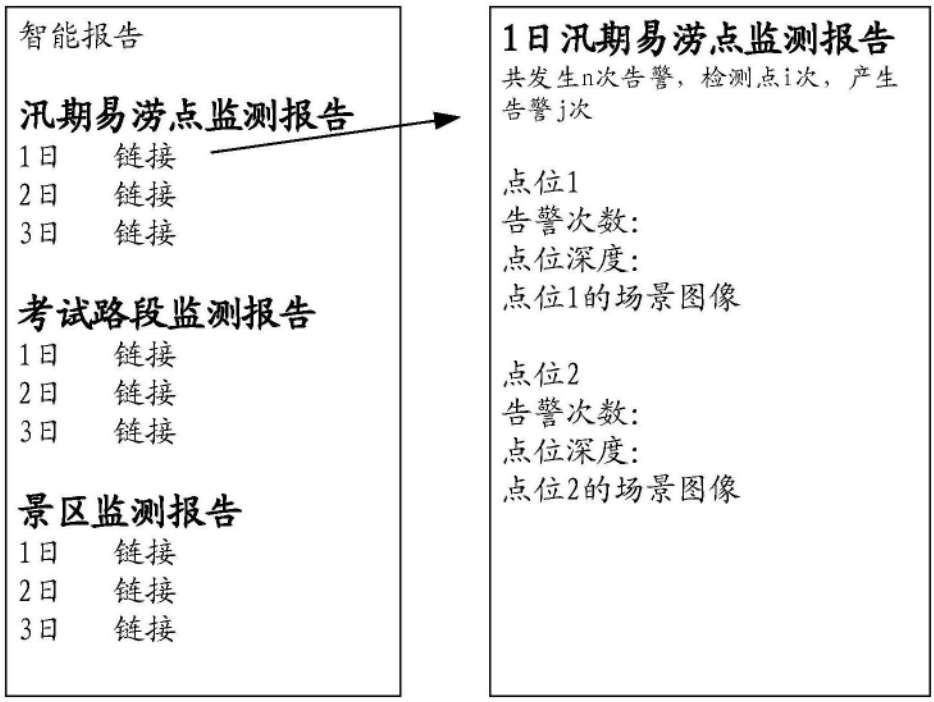

智能报告
汛期易涝点监测报告
1日 链接
2日 链接
3日 链接
考试路段监测报告
1日 链接
2日 链接
3日 链接
景区监测报告
1日 链接
2日 链接
3日 链接
1日汛期易涝点监测报告
共发生n次告警，检测点i次，产生告警j次
点位1
告警次数:
点位深度:
点位1的场景图像
点位2
告警次数:
点位深度:
点位2的场景图像

图6

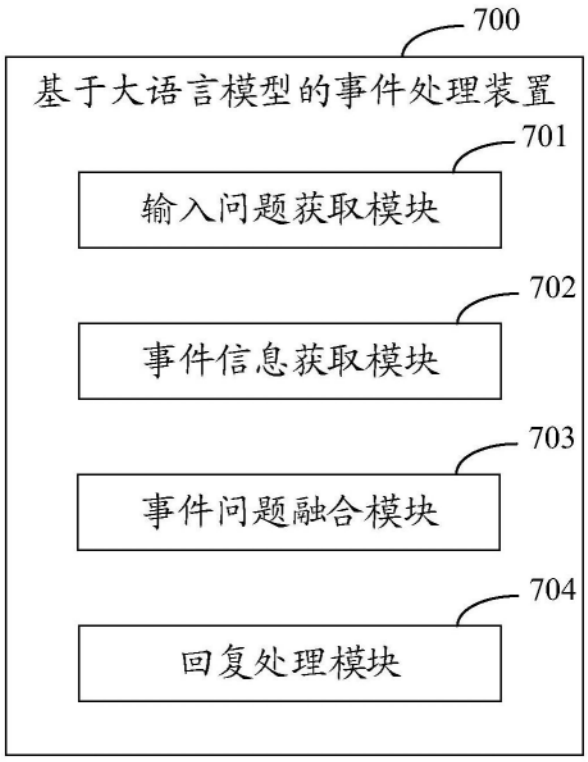

700
基于大语言模型的事件处理装置
701
输入问题获取模块
702
事件信息获取模块
703
事件问题融合模块
704
回复处理模块

图7

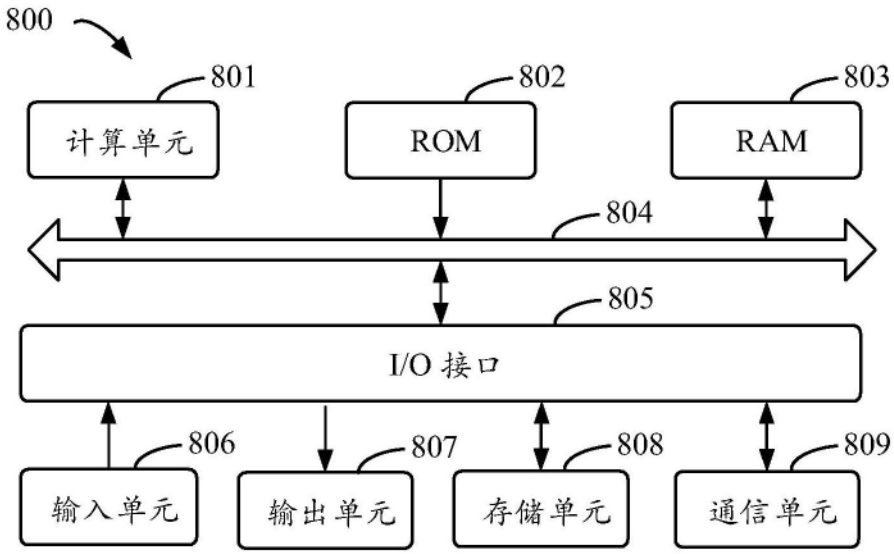

800
801
计算单元
802
ROM
803
RAM
804
805
I/O 接口
806
输入单元
807
输出单元
808
存储单元
809
通信单元

图8