



(12) 发明专利申请

(10) 申请公布号 CN 114238587 A

(43) 申请公布日 2022.03.25

(21) 申请号 202111655536.3

G06V 10/80 (2022.01)

(22) 申请日 2021.12.30

G06K 9/62 (2022.01)

G06N 7/00 (2006.01)

(71) 申请人 中科讯飞互联(北京)信息科技有限公司

地址 100193 北京市海淀区西北旺东路10号院东区5号楼三层311-2

申请人 科大讯飞股份有限公司

(72) 发明人 陈致鹏 崔一鸣 陈志刚

(74) 专利代理机构 北京励诚知识产权代理有限公司 11647

代理人 周慧云

(51) Int. Cl.

G06F 16/33 (2019.01)

G06F 16/332 (2019.01)

G06F 16/56 (2019.01)

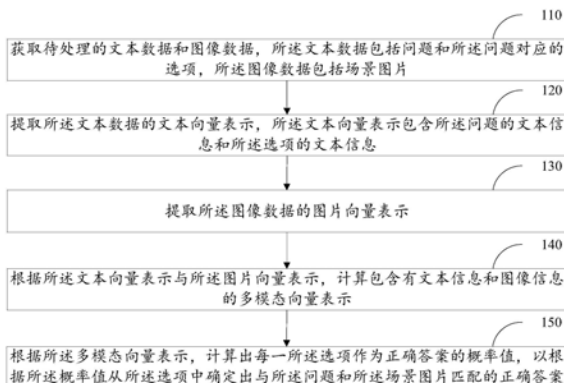
权利要求书2页 说明书14页 附图5页

(54) 发明名称

阅读理解方法、装置、存储介质及计算机设备

(57) 摘要

本申请公开一种阅读理解方法、装置、存储介质及计算机设备。该方法包括：获取待处理的文本数据和图像数据，文本数据包括问题和问题对应的选项，图像数据包括场景图片；提取文本数据的文本向量表示，文本向量表示包含所述问题的文本信息和所述选项的文本信息；提取图像数据的图片向量表示；根据所述文本向量表示与所述图片向量表示，计算包含有文本信息和图像信息的多模态向量表示；根据多模态向量表示，计算出每一选项作为正确答案的概率值，以根据概率值从选项中确定出与问题和场景图片匹配的正确答案，可以实现图片和文本问题输入的多模态阅读理解，且提升了阅读理解的正确性。



1. 一种阅读理解方法,其特征在于,所述方法包括:

获取待处理的文本数据和图像数据,所述文本数据包括问题和所述问题对应的选项,所述图像数据包括场景图片;

提取所述文本数据的文本向量表示,所述文本向量表示包含所述问题的文本信息和所述选项的文本信息;

提取所述图像数据的图片向量表示;

根据所述文本向量表示与所述图片向量表示,计算包含有文本信息和图像信息的多模态向量表示;

根据所述多模态向量表示,计算出每一所述选项作为正确答案的概率值,以根据所述概率值从所述选项中确定出与所述问题和所述场景图片匹配的正确答案。

2. 如权利要求1所述的阅读理解方法,其特征在于,所述提取所述文本数据的文本向量表示,包括:

通过词表将所述文本数据中的每个词转换成每个词在所述词表中对应的序号,并根据所述序号查找所述文本数据的文本向量表示。

3. 如权利要求1所述的阅读理解方法,其特征在于,所述提取所述图像数据的图片向量表示,包括:

根据目标检测模型对所述场景图片进行目标检测以及特征提取,以得到所述图片向量表示,其中,所述图片向量表示包括所述场景图片中每个视觉目标的图像信息向量表示和整个图片的图像信息向量表示。

4. 如权利要求1所述的阅读理解方法,其特征在于,所述根据所述文本向量表示与所述图片向量表示,计算包含文本信息和图像信息的多模态向量表示,包括:

基于自注意力模型对所述文本向量表示与所述图片向量表示进行处理,获得所述问题的文本信息、所述选项的文本信息与所述图像信息之间的全局交互信息;

对所述全局交互信息进行归一化处理,得到第一归一化信息;

根据所述全局交互信息与所述第一归一化信息,确定包含文本信息和图像信息的多模态向量表示。

5. 如权利要求4所述的阅读理解方法,其特征在于,所述将基于自注意力模型对所述文本向量表示与所述图片向量表示进行处理,获得所述问题的文本信息、所述选项的文本信息与所述图像信息之间的全局交互信息,包括:

将根据所述文本向量表示与所述图片向量表示确定的嵌入向量表示输入自注意力模型,根据所述嵌入向量表示与所述嵌入向量表示的转置矩阵之间的乘积,计算匹配矩阵;

根据所述匹配矩阵与所述嵌入向量表示的乘积,确定所述问题的文本信息、所述选项的文本信息与所述图像信息之间的全局交互信息。

6. 如权利要求4所述的阅读理解方法,其特征在于,所述根据所述全局交互信息与所述第一归一化信息,确定包含文本信息和图像信息的多模态向量表示,包括:

将所述全局交互信息与所述第一归一化信息进行相加,以得到第一求和信息;

将所述第一求和信息输入全连接层进行处理后,对所述全连接层的输出结果进行归一化处理,得到第二归一化信息;

将所述第一求和信息与所述第二归一化信息进行相加,得到所述包含文本信息和图像

信息的多模态向量表示。

7. 如权利要求1所述的阅读理解方法,其特征在于,所述方法还包括:

获取位置向量表示和类型向量表示,所述位置向量表示用于标注所述文本数据中每个词的位置,所述类型向量表示用于区分文本类型和图像类型;

所述根据所述文本向量表示与所述图片向量表示,计算包含有文本信息和图像信息的多模态向量表示,包括:

根据所述文本向量表示、所述图片向量表示、所述位置向量表示和所述类型向量表示,计算包含有文本信息和图像信息的多模态向量表示。

8. 如权利要求1-7任一项所述的阅读理解方法,其特征在于,所述根据所述多模态向量表示,计算出每一所述选项作为正确答案的概率值,以根据所述概率值从所述选项中确定出与所述问题和所述场景图片匹配的正确答案,包括:

将所述多模态向量表示进行拆分处理,得到问题选项表示和图像选项表示;

基于交叉注意力模型对所述问题选项表示和所述图像选项表示进行处理,获得关注点向量表示;

根据所述关注点向量表示,计算出每一所述选项作为正确答案的概率值,以根据所述概率值从所述选项中确定出与所述问题和所述场景图片匹配的正确答案。

9. 一种阅读理解装置,其特征在于,所述装置包括:

获取单元,用于获取待处理的文本数据和图像数据,所述文本数据包括问题和所述问题对应的选项,所述图像数据包括场景图片;

第一提取单元,用于提取所述文本数据的文本向量表示,所述文本向量表示包含所述问题的文本信息和所述选项的文本信息;

第二提取单元,用于提取所述图像数据的图片向量表示;

计算单元,用于根据所述文本向量表示与所述图片向量表示,计算包含有文本信息和图像信息的多模态向量表示;

确定单元,用于根据所述多模态向量表示,计算出每一所述选项作为正确答案的概率值,以根据所述概率值从所述选项中确定出与所述问题和所述场景图片匹配的正确答案。

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程序,所述计算机程序适于处理器进行加载,以执行如权利要求1-8任一项所述的阅读理解方法中的步骤。

11. 一种计算机设备,其特征在于,所述计算机设备包括处理器和存储器,所述存储器中存储有计算机程序,所述处理器通过调用所述存储器中存储的所述计算机程序,用于执行权利要求1-8任一项所述的阅读理解方法中的步骤。

阅读理解方法、装置、存储介质及计算机设备

技术领域

[0001] 本申请涉及计算机技术领域,具体涉及一种阅读理解方法、装置、存储介质及计算机设备。

背景技术

[0002] 目前针对多模态的阅读理解方法和系统比较少。基本都是单一模态模型,对数据进行处理,然后抽取相关特征,应用于解决需要多个模态同时来解决的问题。对于多模态的阅读理解,目前常见做法就是通过图像识别(如OCR,人脸识别,人物动作识别等等)提取出图片中的有用信息,作为特征,然后再通过这些特征通过处理文本的模型(如GRU, LSTM, CNN, BERT等等模型)来对问题进行建模,然后结合图像特征和文本特征,通过计算特征向量的相似度从而得出问题的答案。处理多模态的阅读理解问题,现有的方法都需要多个系统协作才能完整的解决多模态的阅读理解问题,需要同时使用图像识别系统和自然语言处理系统对两个信号进行同步处理,再根据两个系统结果,构建第三个系统来统一完成最后答案的计算。目前相关多模态的阅读理解系统在处理多模态的信号输入以及最终完成阅读理解问题时,系统较复杂,且容易丢失相互增加的信号,导致错误率较高,阅读理解的正确性低。

发明内容

[0003] 本申请实施例提供一种阅读理解方法、装置、存储介质及计算机设备,可以实现图片和文本问题输入的多模态阅读理解,提升了阅读理解的正确性。

[0004] 一方面,提供一种阅读理解方法,所述方法包括:

[0005] 获取待处理的文本数据和图像数据,所述文本数据包括问题和所述问题对应的选项,所述图像数据包括场景图片;

[0006] 提取所述文本数据的文本向量表示,所述文本向量表示包含所述问题的文本信息和所述选项的文本信息;

[0007] 提取所述图像数据的图片向量表示;

[0008] 根据所述文本向量表示与所述图片向量表示,计算包含有文本信息和图像信息的多模态向量表示;

[0009] 根据所述多模态向量表示,计算出每一所述选项作为正确答案的概率值,以根据所述概率值从所述选项中确定出与所述问题和所述场景图片匹配的正确答案。

[0010] 可选的,所述提取所述文本数据的文本向量表示,包括:

[0011] 通过词表将所述文本数据中的每个词转换成每个词在所述词表中对应的序号,并根据所述序号查找所述文本数据的文本向量表示。

[0012] 可选的,所述提取所述图像数据的图片向量表示,包括:

[0013] 根据目标检测模型对所述场景图片进行目标检测以及特征提取,以得到所述图片向量表示,其中,所述图片向量表示包括所述场景图片中每个视觉目标的图像信息向量表

示和整个图片的图像信息向量表示。

[0014] 可选的,所述根据所述文本向量表示与所述图片向量表示,计算包含文本信息和图像信息的多模态向量表示,包括:

[0015] 基于自注意力模型对所述文本向量表示与所述图片向量进行处理,获得所述问题的文本信息、所述选项的文本信息与所述图像信息之间的全局交互信息;

[0016] 对所述全局交互信息进行归一化处理,得到第一归一化信息;

[0017] 根据所述全局交互信息与所述第一归一化信息,确定包含文本信息和图像信息的多模态向量表示。

[0018] 可选的,所述将基于自注意力模型对所述文本向量表示与所述图片向量进行处理,获得所述问题的文本信息、所述选项的文本信息与所述图像信息之间的全局交互信息,包括:

[0019] 将根据所述文本向量表示与所述图片向量表示确定的嵌入向量表示输入自注意力模型,根据所述嵌入向量表示与所述嵌入向量表示的转置矩阵之间的乘积,计算匹配矩阵;

[0020] 根据所述匹配矩阵与所述嵌入向量表示的乘积,确定所述问题的文本信息、所述选项的文本信息与所述图像信息之间的全局交互信息。

[0021] 可选的,所述根据所述全局交互信息与所述第一归一化信息,确定包含文本信息和图像信息的多模态向量表示,包括:

[0022] 将所述全局交互信息与所述第一归一化信息进行相加,以得到第一求和信息;

[0023] 将所述第一求和信息输入全连接层进行处理后,对所述全连接层的输出结果进行归一化处理,得到第二归一化信息;

[0024] 将所述第一求和信息与所述第二归一化信息进行相加,得到所述包含文本信息和图像信息的多模态向量表示。

[0025] 可选的,所述方法还包括:

[0026] 获取位置向量表示和类型向量表示,所述位置向量表示用于标注所述文本数据中每个词的位置,所述类型向量表示用于区分文本类型和图像类型;

[0027] 所述根据所述文本向量表示与所述图片向量表示,计算包含有文本信息和图像信息的多模态向量表示,包括:

[0028] 根据所述文本向量表示、所述图片向量表示、所述位置向量表示和所述类型向量表示,计算包含有文本信息和图像信息的多模态向量表示。

[0029] 可选的,所述根据所述多模态向量表示,计算出每一所述选项作为正确答案的概率值,以根据所述概率值从所述选项中确定出与所述问题和所述场景图片匹配的正确答案,包括:

[0030] 将所述多模态向量表示进行拆分处理,得到问题选项表示和图像选项表示;

[0031] 基于交叉注意力模型对所述问题选项表示和所述图像选项表示进行处理,获得关注点向量表示;

[0032] 根据所述关注点向量表示,计算出每一所述选项作为正确答案的概率值,以根据所述概率值从所述选项中确定出与所述问题和所述场景图片匹配的正确答案。

[0033] 另一方面,提供一种阅读理解装置,所述装置包括:

[0034] 获取单元,用于获取待处理的文本数据和图像数据,所述文本数据包括问题和所述问题对应的选项,所述图像数据包括场景图片;

[0035] 第一提取单元,用于提取所述文本数据的文本向量表示,所述文本向量表示包含所述问题的文本信息和所述选项的文本信息;

[0036] 第二提取单元,用于提取所述图像数据的图片向量表示;

[0037] 计算单元,用于根据所述文本向量表示与所述图片向量表示,计算包含有文本信息和图像信息的多模态向量表示;

[0038] 确定单元,用于根据所述多模态向量表示,计算出每一所述选项作为正确答案的概率值,以根据所述概率值从所述选项中确定出与所述问题和所述场景图片匹配的正确答案。

[0039] 另一方面,提供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程序,所述计算机程序适于处理器进行加载,以执行如上任一实施例所述的阅读理解方法中的步骤。

[0040] 另一方面,提供一种计算机设备,所述计算机设备包括处理器和存储器,所述存储器中存储有计算机程序,所述处理器通过调用所述存储器中存储的所述计算机程序,用于执行如上任一实施例所述的阅读理解方法中的步骤。

[0041] 另一方面,提供一种计算机程序产品,包括计算机指令,所述计算机指令被处理器执行时实现如上任一实施例所述的阅读理解方法中的步骤。

[0042] 本申请实施例通过获取待处理的文本数据和图像数据,文本数据包括问题和问题对应的选项,图像数据包括场景图片;提取文本数据的文本向量表示,文本向量表示包含问题的文本信息和选项的文本信息;提取图像数据的图片向量表示;根据文本向量表示与图片向量表示,计算包含有文本信息和图像信息的多模态向量表示;根据多模态向量表示,计算出每一选项作为正确答案的概率值,以根据概率值从选项中确定出与问题和场景图片匹配的正确答案。本申请实施例通过Transformer模型实现图片和文本问题输入的多模态阅读理解,模型通过同时输入图像数据与包含问题和选项的文本数据,通过Transformer模型内部的注意力(attention)机制,计算包含有文本信息和图像信息的多模态向量表示,以将有用的图片信息和文本信息过滤出来,然后根据多模态向量表示选出正确的答案选项,提升了阅读理解的正确性。

附图说明

[0043] 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的一些实施例,对于本领域技术人员来讲,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。

[0044] 图1为本申请实施例提供的阅读理解系统的结构框架图。

[0045] 图2为本申请实施例提供的阅读理解方法的第一流程示意图。

[0046] 图3为本申请实施例提供的阅读理解方法的第一应用场景示意图。

[0047] 图4为本申请实施例提供的阅读理解方法的第二流程示意图。

[0048] 图5为本申请实施例提供的阅读理解方法的第二应用场景示意图。

- [0049] 图6为本申请实施例提供的阅读理解方法的第三流程示意图。
- [0050] 图7为本申请实施例提供的阅读理解方法的第三应用场景示意图。
- [0051] 图8为本申请实施例提供的阅读理解装置的结构示意图。
- [0052] 图9为本申请实施例提供的计算机设备的结构示意图。

具体实施方式

[0053] 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清楚、完整地描述,显然,所描述的实施例仅仅是本申请一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域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,都属于本申请保护的范围。

[0054] 本申请实施例提供一种阅读理解方法、装置、计算机设备和存储介质。具体地,本申请实施例的阅读理解方法可以由计算机设备执行,其中,该计算机设备可以为终端或者服务器等设备。该终端可以为智能手机、平板电脑、笔记本电脑、智能电视、智能音箱、穿戴式智能设备、个人计算机(Personal Computer,PC)等设备,终端还可以包括客户端,该客户端可以是视频客户端、浏览器客户端或即时通信客户端等。服务器可以是独立的物理服务器,也可以是多个物理服务器构成的服务器集群或者分布式系统,还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、内容分发网络(Content Delivery Network,CDN)、以及大数据和人工智能平台等基础云计算服务的云服务器。

[0055] 本申请实施例可应用于人工智能、语音识别、智慧交通等各种场景。

[0056] 首先,在对本申请实施例进行描述的过程中出现的部分名词或者术语作如下解释:

[0057] 人工智能(Artificial Intelligence,AI)是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能,感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说,人工智能是计算机科学的一个综合技术,它企图了解智能的实质,并生产出一种新的能以人类智能相似的方式做出反应的智能机器。人工智能也就是研究各种智能机器的设计原理与实现方法,使机器具有感知、推理与决策的功能。人工智能基础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、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、语音处理技术、自然语言处理技术以及机器学习/深度学习等几大方向。

[0058] 机器学习(Machine Learning,ML)是一门多领域交叉学科,涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。机器学习是人工智能的核心,是使计算机具有智能的根本途径,其应用遍及人工智能的各个领域。机器学习和深度学习通常包括人工神经网络、置信网络、强化学习、迁移学习、归纳学习、式教学习等技术。

[0059] 深度学习(Deep Learning,DL)是机器学习的分支,是一种试图使用包含复杂结构或由多重非线性变换构成的多个处理层对数据进行高层抽象的算法。深度学习是学习训练样本数据的内在规律和表示层次,这些学习过程中获得的信息对文字、图像和声音等数据

的解释有很大的帮助。深度学习的最终目标是让机器能够像人一样具有分析学习能力,能够识别文字、图像和声音等数据。深度学习是一个复杂的机器学习算法,在语音和图像识别方面取得的效果,远远超过先前相关技术。

[0060] 神经网络(Neural Network, NN)是在机器学习和认知科学领域的一种模仿生物神经网络结构和功能的深度学习模型。

[0061] Transformer模型是一种NLP(自然语言处理)经典模型。Transformer模型完全基于注意力来编码输入和计算输出,而不依赖于序列对齐的循环神经网络或卷积神经网络,Transformer模型使用Self-Attention(自注意力)机制,而不采用RNN的顺序结构,使得模型可以并行化训练,而且能够拥有全局信息。

[0062] 目前相关多模态的阅读理解系统在处理多模态的信号输入以及最终完成阅读理解问题时,主要存在以下几点不足的地方:

[0063] 1、对于多模态任务,多个系统进行处理的方法会导致错误累积。例如,若图像系统出现错误,提取了无用的特征信号,会进一步影响后续系统做出正确判断。同样,当处理文本的系统出现错误或者精度下降,那么后续再通过图片信息来回答阅读理解问题时,就会更加容易做出错误的判断。

[0064] 2、多模态信号使用多个单模态系统进行联合处理时,会丢失一些互相增强的信号。因为在抽取特征时,图像系统不知道文本系统需要什么样的特征,文本系统在抽取特征时也无法得到图像的信号输入,所以在最后融合时相关重要的信息可能无法正常获取到。

[0065] 3、采用多个系统来分别处理不同模态的信息时,系统就变得更加复杂,从而系统发生错误概率也会明显提高。因为物理上图像和文本就是分开处理的,所以文本和图像本身能够通过共现关系能够消除的一些歧义的东西就被丢掉了,这导致对于多模态的信号的理解问题的回答能力显著下降。并且阅读理解问题本身就具有非常大的难度,如果不能同时建模两个模态的信息,那么其难度就会进一步加大。

[0066] 随着人工智能技术的发展,单一模态的模型越来越无法满足日常人机交互的需求。为了顺应技术发展的要求,本申请实施例提出一种多模态的阅读理解系统,在解决阅读理解问题的时候,本申请实施例阅读理解的对象由输入文字,变成了复杂场景的图片,通过对图片的理解,然后选择正确答案。随着人工神经网络的发展,单一模型多个模态信息的处理能力,越来越成为人机交互的关键技术,因此本申请实施例提出基于多模态的阅读理解技术,用于帮助机器更好的理解环境信号,并且通过这些信号回答相关的问题。该技术在未来的机器人身上具有非常大的应用潜力。

[0067] 本申请实施例可以用于解决以图片和文本问题输入的多模态阅读理解问题或者是图像的问答问题。可以广泛应用于人机现实场景对话,以及图像的理解和检索。具有非常巨大的应用价值。

[0068] 本申请实施例提出的基于神经网络的深度学习模型的多模态阅读理解解题系统恰好就能很好的克服以上几个问题。首先,本申请实施例对文本数据进行向量化的表示,每个词、每个问题、每个选项、每句话都会被映射到一个特定的空间。然后,选项的表示会与问题及不同模态的图像向量表示进行基于空间关系的关注点计算(attention计算),也就是说问题和选项会通过该计算获取到图像信息中与之最相关的部分。最后,全连接的神经网络会将问题图像与选项之间相关的信息映射到解空间,从而选择出一个正确的选项来

作为答案。本申请实施例系统主要优点：1、具有良好的迁移性。只要更换训练数据就可以迁移到其它模态的阅读理解问题上。2、不需要人工编辑大量的专家知识。模型通过自动从数据中获取到解题的知识。3、使用到的图像问题及选项都使用空间向量表示，具有较强的泛化能力。4、一套系统就能处理图像以及文字模态的输入。极大简化后期应用部署。

[0069] 请参阅图1,图1为本申请实施例提供的阅读理解系统的结构框架图。该阅读理解系统包括一个多模态的Transformer模型。首先文本模态的文本数据的输入,该包括问题和选项。然后是图像模态的图像数据的输入,图像数据包括场景图片。然后是一个多模态的Transformer模型。可以通过Transformer模型计算出与问题和场景图片匹配的选项作为问题的正确答案。其中,Transformer模型通过同时输入的问题、选项和场景图片的信号,通过Transformer模型内部的attention机制,将有用的图片信息和文本信息都过滤出来,计算包含有文本信息和图像信息的多模态向量表示,以将有用的图片信息和文本信息过滤出来,然后根据这些过滤出来的信息,通过全连接,选出正确的答案选项。其中,文本和图像的两个模态的特征表示,是在图片和文本同时输入时学习到的,模型能够准确的得到有用的多模态向量表示信息,最后再准确的判断出这个场景图片是不是与当前问题的答案匹配。由于整个阅读理解系统是直接通过一个模型来完成阅读理解问题的,所以结构比较简洁,性能也比较高效。

[0070] 以下分别进行详细说明。需说明的是,以下实施例的描述顺序不作为对实施例优先顺序的限定。

[0071] 本申请各实施例提供了一种阅读理解方法,该方法可以由终端或服务器执行,也可以由终端和服务器共同执行;本申请实施例以阅读理解方法由服务器执行为例来进行说明。

[0072] 请参阅图2至图7,图2、图4及图6均为本申请实施例提供的阅读理解方法的流程示意图,图3、图5及图7均为本申请实施例提供的阅读理解方法的应用场景示意图。该方法包括:

[0073] 步骤110,获取待处理的文本数据和图像数据,所述文本数据包括问题和所述问题对应的选项,所述图像数据包括场景图片。

[0074] 例如,该文本数据包括问题和该问题对应的选项,其中,选项可以为一个或多个。比如,获取的待处理的问题为“图中谁戴着眼镜?”;比如选项为一个,比如选项为“男人”,比如选项为多个,比如选项为“男人、女人、小孩”等。

[0075] 例如,该图像数据包括场景图片,该场景图片是对应该问题描述的场景提供的图片,比如用户在作答过程中,可以通过观察该场景图片来回答该问题。

[0076] 步骤120,提取所述文本数据的文本向量表示,所述文本向量表示包含所述问题的文本信息和所述选项的文本信息。

[0077] 可选的,所述提取所述文本数据的文本向量表示,包括:

[0078] 通过词表将所述文本数据中的每个词转换成每个词在所述词表中对应的序号,并根据所述序号查找所述文本数据的文本向量表示。

[0079] 首先,可以对文本数据进行向量化的表示,以将每个词、每个问题、每个选项、每句话映射到一个特定的空间,以得到文本数据的文本向量表示。

[0080] 例如,结合图3进行说明,该阅读理解系统可以包括多模态的数据输入处理模块、

多模态的特征抽取模块和选项打分模块,其中,该特征抽取模块可以采用Transformer模型。将输入该数据输入处理模块的TXT格式的文本数据(即TXT格式的问题及选项对应的原始文本),通过词表将原始文本转换成每个词对应词表中的序号(ID),然后将原始文本通过ID查找每个词对于词表的嵌入(embedding)向量表示。例如,问题:“图中谁戴着眼镜?”,选项为男人,通过词表转化为ID[1,4,3,6,7,0,12,87,98,10],因为“图”在词表中的ID是1,所以就转换成1,然后再通过ID找到每个词对应的向量表示(w1,w2,w3,w4,w5,w6,w7,w8,w9,w10),得到词向量序列。然后将这一串词向量序列后续作为Transformer模型的输入参数,可以定义为文本向量表示,该文本向量表示为seq_len乘hid_size的矩阵,其中,seq_len表示文本长度,hid_size表示词向量的大小。

[0081] 步骤130,提取所述图像数据的图片向量表示。

[0082] 可选的,所述提取所述图像数据的图片向量表示,包括:

[0083] 根据目标检测模型对所述场景图片进行目标检测以及特征提取,以得到所述图片向量表示,其中,所述图片向量表示包括所述场景图片中每个视觉目标的图像信息向量表示和整个图片的图像信息向量表示。

[0084] 例如,将输入该数据输入处理模块的图像数据,首先通过目标检测模型(Fast RCNN)抽取物理的数值信息,然后通过模型学习,得到一个能够正确抽取物理信息的Fast RCNN模型。Fast RCNN(Fast Regions with CNN features)一种快速的基于区域的卷积神经网络方法,用于目标检测。

[0085] 例如,结合图3进行说明,输入包含有场景图片的图像数据,首先通过Fast-RCNN对图像数据中的场景图片进行目标检测以及特征抽取,得到图片中每个视觉目标的图像信息向量表示和整个图片的图像信息向量表示。其中,该视觉目标为对应于文本数据中需要关注的对象,比如文本数据中提到“图中谁”,那么就需要关注场景图片中的人物,比如图中的男人和女人。其中,整个图片的图像信息向量表示会分别应用到每个文本字上面,而单个视觉目标的图像信息向量表示会对应到专门表示图像信息的文本向量上,也就是图3示出的img文本向量。最终图像数据作为Transformer模型的输入参数也是一个seq_len乘hid_size的矩阵,可定义为图片向量表示。

[0086] 步骤140,根据所述文本向量表示与所述图片向量表示,计算包含有文本信息和图像信息的多模态向量表示。

[0087] 例如,将包含图像、问题及选项的Embedding向量表示输入到Transformer模型中,计算出每个选项与图片、问题之间的多模态向量表示。多模态向量表示是一个既包含文本信息又包含图像信息的向量。

[0088] 可选的,所述方法还包括:获取位置向量表示和类型向量表示,所述位置向量表示用于标注所述文本数据中每个词的位置,所述类型向量表示用于区分文本类型和图像类型;

[0089] 所述根据所述文本向量表示与所述图片向量表示,计算包含有文本信息和图像信息的多模态向量表示,包括:

[0090] 根据所述文本向量表示、所述图片向量表示、所述位置向量表示和所述类型向量表示,计算包含有文本信息和图像信息的多模态向量表示。

[0091] 其中,位置向量表示用于标注文本数据中每个词的位置,位置向量表示的大小为

seq_len乘hid_size的矩阵。类型向量表示用于区分文本类型和图像类型,类型向量表示的大小为seq_len乘hid_size的矩阵,比如分文本类型表示为0,图像类型表示为1。

[0092] 例如,如图3所示,最终输入Transformer模型的输入参数可以包括由文本向量表示加上图片向量表示加上位置向量表示和类型向量表示构成的嵌入向量表示Embedding,记为E。例如,将包含图像、问题及选项的Embedding向量表示输入到Transformer模型中,计算出每个选项与图片、问题之间的多模态向量表示。多模态向量表示是一个既包含文本信息又包含图像信息的向量,即选项和问题中的每个的向量表示与图像的表达通过transformer模型内部的计算,通过各组共现的方式抽取的最优的特征向量,最后将这个多模态向量表示连接一个全连接计算出当前问题的正确答案。其中,全连接层最后输出每个选项成为正确答案的概率值,最终将这些概率值进行归一化作为每个选项作为正确答案的概率。

[0093] 例如,结合图3进行说明,可以通过根据多模态的特征抽取模块对文本向量表示与图片向量表示进行处理,计算包含有文本信息和图像信息的多模态向量表示,其中,该特征抽取模块可以采用Transformer模型。该特征抽取模块的主要功能是计算融合了文本信息和图像信息的多模态向量表示。如图3所示,通过Transformer模型处理后最终会得到一个融合了文本信息和图像信息的多模态向量表示,而这个多模态向量表示可以用于计算问题答案与当前场景图片是否匹配的概率值。如图3所示,计算的就是“男人”这个选项作为正确答案的概率值。该特征抽取模块主要是为了计算图像、问题与选项之间的关联关系,并且输出匹配的特征矩阵,该特征矩阵可以作为包含有文本信息和图像信息的多模态向量表示。

[0094] 可选的,如图4所示,步骤140可通过步骤141至步骤143来实现,具体为:

[0095] 步骤141,基于自注意力模型对所述文本向量表示与所述图片向量进行处理,获得所述问题的文本信息、所述选项的文本信息与所述图像信息之间的全局交互信息。

[0096] 可选的,所述将基于自注意力模型对所述文本向量表示与所述图片向量进行处理,获得所述问题的文本信息、所述选项的文本信息与所述图像信息之间的全局交互信息,包括:将根据所述文本向量表示与所述图片向量表示确定的嵌入向量表示输入自注意力模型,根据所述嵌入向量表示与所述嵌入向量表示的转置矩阵之间的乘积,计算匹配矩阵;根据所述匹配矩阵与所述嵌入向量表示的乘积,确定所述问题的文本信息、所述选项的文本信息与所述图像信息之间的全局交互信息。

[0097] 其中,全局交互信息的长度维度与嵌入向量表示的长度维度相同。

[0098] 例如,请参阅图5,特征抽取模块的输入为:图像、问题及选项的向量表示,其大小为seq_len乘hid_size的矩阵。例如,该图像、问题及选项的向量表示可以为文本向量表示加上图片向量表示构成的Embedding向量表示,记为E。该图像、问题及选项的向量表示也可以为文本向量表示加上图片向量表示加上位置向量表示和类型向量表示,构成的Embedding向量表示,记为E。

[0099] 特征抽取模块的输出为:融合了图像信息和所有文本信息的多模态向量表示,其大小为seq_len乘hid_size的矩阵。

[0100] 例如,请参阅图5,通过特征抽取模块内部的自注意力(self_attention)模型计算匹配矩阵,输入为seq_len乘hid_size的embedding向量表示E,该embedding向量表示即为图像、问题及选项的向量表示;输出为全局交互信息 H_s , H_s 的大小为seq_len乘hid_size。具

体计算过程中, self_attention即自己和自己计算注意力(attention)表示, E矩阵乘 E^T , 得到匹配矩阵M, 其中, 匹配矩阵M的大小为seq_len乘seq_len, 然后M矩阵乘E得到 H_s , H_s 的大小为seq_len乘hid_size。其中, E^T 是E矩阵的转置矩阵。

[0101] 步骤142, 对所述全局交互信息进行归一化处理, 得到第一归一化信息。

[0102] 例如, 请参阅图5, 对self_attention模型输出的全局交互信息 H_s 进行归一化, 即norm, 得到第一归一化信息 H_n , 第一归一化信息 H_n 的大小为seq_len乘hid_size, 归一化不会影响矩阵大小。其中, 全局交互信息的长度维度与嵌入向量表示的长度维度相同, 全局交互信息的词向量大小与嵌入向量表示的词向量大小相同。

[0103] 步骤143, 根据所述全局交互信息与所述第一归一化信息, 确定包含文本信息和图像信息的多模态向量表示。

[0104] 可选的, 所述根据所述全局交互信息与所述第一归一化信息, 确定包含文本信息和图像信息的多模态向量表示, 包括: 将所述全局交互信息与所述第一归一化信息进行相加, 以得到第一求和信息; 将所述第一求和信息输入全连接层进行处理后, 对所述全连接层的输出结果进行归一化处理, 得到第二归一化信息; 将所述第一求和信息与所述第二归一化信息进行相加, 得到所述包含文本信息和图像信息的多模态向量表示。

[0105] 例如, 请参阅图5, 全连接层的输入为 H_s+H_n , 即将全局交互信息 H_s 与第一归一化信息 H_n 进行相加得到第一求和信息, 然后将将第一求和信息输入全连接层进行处理, 其中全连接层的输出结果的大小为seq_len乘hid_size。然后再对全连接层的输出结果进行归一化处理(norm), 得到第二归一化信息, 并再次将得到第二归一化信息与第一求和信息相加, 得到多模态向量表示 H_{mn} 。由于 H_{mn} 与输入的E矩阵大小一样, self_attention模型这里可以叠加多层, 一般可以设置为12层或者24层。

[0106] 例如, 多模态向量表示 H_{mn} 最后通过直接输出output, output的大小为seq_len乘hid_size的矩阵, 该多模态向量表示为融合了图像信息和所有文本信息的多模态向量表示。其中, 多模态向量表示的长度维度与嵌入向量表示的长度维度相同。

[0107] 通过该特征抽取模块, 计算得到融合了选项与图像、问题的多模态向量表示, 该多模态向量表示包含文本和图像之间的高度抽象的语义匹配关系, 为后续模块根据匹配信息选择出正确答案提供了丰富的信息。同时也将以往只是通过简单的字符级别的文本匹配转换到了向量空间之间的匹配。该特征抽取模块使得文本匹配上升到了语义空间级别。

[0108] 步骤150, 根据所述多模态向量表示, 计算出每一所述选项作为正确答案的概率值, 以根据所述概率值从所述选项中确定出与所述问题和所述场景图片匹配的正确答案。

[0109] 例如, 可以将特征抽取模块的输出的output输出结果(融合了选项与图像、问题的多模态向量表示), 通过一个选项打分模块进行处理, 计算出当前选项作为正确答案的概率值。

[0110] 可选的, 如图6所示, 步骤150可通过步骤151至步骤153来实现, 具体为:

[0111] 步骤151, 将所述多模态向量表示进行拆分处理, 得到问题选项表示和图像选项表示。

[0112] 例如, 请参阅图7, 输入的融合了图像和文本的多模态向量表示, 通过矩阵拆分, 得到问题选项表示 H_{qc} 和图像选项表示 H_{pc} 。Output的大小为seq_len乘hid_size的矩阵, 问题选项为seq1_len乘hid_size的矩阵, 图像选项表示为sseq2_len乘hid_size的矩阵, 其中,

$seq = seq1 + seq2$ 。

[0113] 步骤152,基于交叉注意力模型对所述问题选项表示和所述图像选项表示进行处理,获得关注点向量表示。

[0114] 例如,请参阅图7,基于交叉注意力(cross attention)模型,使用问题选项表示 H_{qc} 从图像选项表示 H_{pc} 上计算一个关注带你(attend)向量表示出来。 H_{qc} 和 H_{pc} 分别接一个全连接,得到 H'_{qc} 和 H'_{pc} ,然后将 H'_{qc} 矩阵乘以 H'_{pc} 的转置,得到attention表示,attention表示的矩阵大小为 $seq1 \times seq2$ 。Attention表示再矩阵乘 H_{pc} ,得到attend向量表示 H_{att} , H_{att} 的矩阵大小为 $seq1 \times hid_size$ 。该attention表示为感兴趣的内容或兴趣点,比如图3示例的问题和场景图片,获得的关注点向量表示可以包含有“图片中的男人戴着眼镜”和“图片中的女人没有戴眼镜”的内容。

[0115] 步骤153,根据所述关注点向量表示,计算出每一所述选项作为正确答案的概率值,以根据所述概率值从所述选项中确定出与所述问题和所述场景图片匹配的正确答案。

[0116] 例如,请参阅图7,最终计算选项作为答案的概率值。Attend向量表示在 $seq1$ 这个维度求和,得到向量 V ,向量 V 的大小为 $1 \times hid_size$ 的向量。然后再接一个全连接,将向量 V 变成一个得分(score),该得分可以表示选项作为正确答案的概率值。

[0117] 其中,通过选项打分模块计算得分的时候,再次将选项、问题以及图片进行区分对待,并且再次通过attention的方式,最终得到当前选项作为正确答案的概率值,如果一个问题有四个选项,通过对比四个选项的不同得分,就可以确定出正确答案。例如,将概率值最高(即得分最高)的选项,确定为与问题和场景图片匹配的正确答案。

[0118] 在解题时,整个多模态的阅读理解系统通过多模态的数据输入处理模块、多模态的特征抽取模块和选项打分模块依次处理,得到问题的正确答案。其中,在使用该阅读理解系统之前,还可以提供足够多的多模态的阅读理解数据,对该阅读理解系统进行模型学习。通过学习过后,整个阅读理解系统就可以自动对特定场景(比如中高考英文)的阅读理解题进行解答。

[0119] 本申请实施例可以基于transformer的多模态阅读理解技术,具有多模态的多种模态信息抽取能力,可以根据多模态向量表示计算问题的正确答案。

[0120] 本申请实施例可以采用的端到端的深度神经网络技术,相比传统的基于规则和专家知识的系统,本阅读理解系统极大的简化整个解题流程,去掉繁杂的规则以及知识特征提取的过程。只需要提供系统模型训练的数据,就可以让本系统自动学习到解题相关的知识和规则。本阅读理解系统可以快速迁移到其它类似的任务和领域上。本阅读理解系统采用了模块化得设计方案,不同模块功能专一,可以快速迁移到相关的系统上。例如选项打分模块可以直接应用到其它需要计算向量相似度的任务上。并且整个系统可以直接通过更换训练数据直接应用到多模态的阅读理解任务上。

[0121] 上述所有的技术方案,可以采用任意结合形成本申请的可选实施例,在此不再一一赘述。

[0122] 本申请实施例通过获取待处理的文本数据和图像数据,文本数据包括问题和问题对应的选项,图像数据包括场景图片;提取文本数据的文本向量表示,文本向量表示包含问题的文本信息和选项的文本信息;提取图像数据的图片向量表示;根据文本向量表示与图片向量表示,计算包含有文本信息和图像信息的多模态向量表示;根据多模态向量表示,计

算出每一选项作为正确答案的概率值,以根据概率值从选项中确定出与问题和场景图片匹配的正确答案。本申请实施例通过Transformer模型实现图片和文本问题输入的多模态阅读理解,模型通过同时输入图像数据与包含问题和选项的文本数据,通过Transformer模型内部的注意力(attention)机制,计算包含有文本信息和图像信息的多模态向量表示,以将有用的图片信息和文本信息过滤出来,然后根据多模态向量表示选出正确的答案选项,提升了阅读理解的正确性。

[0123] 为便于更好的实施本申请实施例的阅读理解方法,本申请实施例还提供一种阅读理解装置。请参阅图8,图8为本申请实施例提供的阅读理解装置的结构示意图。其中,该阅读理解装置200可以包括:

[0124] 获取单元201,用于获取待处理的文本数据和图像数据,所述文本数据包括问题和所述问题对应的选项,所述图像数据包括场景图片;

[0125] 第一提取单元202,用于提取所述文本数据的文本向量表示,所述文本向量表示包含所述问题的文本信息和所述选项的文本信息;

[0126] 第二提取单元203,用于提取所述图像数据的图片向量表示;

[0127] 计算单元204,用于根据所述文本向量表示与所述图片向量表示,计算包含有文本信息和图像信息的多模态向量表示;

[0128] 确定单元205,用于根据所述多模态向量表示,计算出每一所述选项作为正确答案的概率值,以根据所述概率值从所述选项中确定出与所述问题和所述场景图片匹配的正确答案。

[0129] 可选的,所述第一提取单元202,可以用于通过词表将所述文本数据中的每个词转换成每个词在所述词表中对应的序号,并根据所述序号查找所述文本数据的文本向量表示。

[0130] 可选的,所述第二提取单元203,可以用于根据目标检测模型对所述场景图片进行目标检测以及特征提取,以得到所述图片向量表示,其中,所述图片向量表示包括所述场景图片中每个视觉目标的图像信息向量表示和整个图片的图像信息向量表示。

[0131] 可选的,所述计算单元204,可以具体用于:基于自注意力模型对所述文本向量表示与所述图片向量表示进行处理,获得所述问题的文本信息、所述选项的文本信息与所述图像信息之间的全局交互信息;对所述全局交互信息进行归一化处理,得到第一归一化信息;根据所述全局交互信息与所述第一归一化信息,确定包含文本信息和图像信息的多模态向量表示。

[0132] 可选的,所述计算单元204,在将基于自注意力模型对所述文本向量表示与所述图片向量表示进行处理,获得所述问题的文本信息、所述选项的文本信息与所述图像信息之间的全局交互信息时,可以用于:将根据所述文本向量表示与所述图片向量表示确定的嵌入向量表示输入自注意力模型,根据所述嵌入向量表示与所述嵌入向量表示的转置矩阵之间的乘积,计算匹配矩阵;根据所述匹配矩阵与所述嵌入向量表示的乘积,确定所述问题的文本信息、所述选项的文本信息与所述图像信息之间的全局交互信息。

[0133] 可选的,所述计算单元204,在根据所述全局交互信息与所述第一归一化信息,确定包含文本信息和图像信息的多模态向量表示时,可以用于:将所述全局交互信息与所述第一归一化信息进行相加,以得到第一求和信息;将所述第一求和信息输入全连接层进行

处理后,对所述全连接层的输出结果进行归一化处理,得到第二归一化信息;将所述第一求和信息与所述第二归一化信息进行相加,得到所述包含文本信息和图像信息的多模态向量表示。

[0134] 可选的,所述获取单元201,还可以用于获取位置向量表示和类型向量表示,所述位置向量表示用于标注所述文本数据中每个词的位置,所述类型向量表示用于区分文本类型和图像类型;

[0135] 所述计算单元204,可以用于根据所述文本向量表示、所述图片向量表示、所述位置向量表示和所述类型向量表示,计算包含有文本信息和图像信息的多模态向量表示。

[0136] 可选的,所述确定单元205,可以具体用于:将所述多模态向量表示进行拆分处理,得到问题选项表示和图像选项表示;基于交叉注意力模型对所述问题选项表示和所述图像选项表示进行处理,获得关注点向量表示;根据所述关注点向量表示,计算出每一所述选项作为正确答案的概率值,以根据所述概率值从所述选项中确定出与所述问题和所述场景图片匹配的正确答案。

[0137] 需要说明的是,本申请实施例中的阅读理解装置200中各模块的功能可对应参考上述各方法实施例中任意实施例的具体实现方式,这里不再赘述。

[0138] 上述阅读理解装置中的各个单元可全部或部分通过软件、硬件及其组合来实现。上述各个单元可以以硬件形式内嵌于或独立于计算机设备中的处理器中,也可以以软件形式存储于计算机设备中的存储器中,以便于处理器调用执行上述各个单元对应的操作。

[0139] 阅读理解装置200例如可以集成在具备存储器并安装有处理器而具有运算能力的终端或服务器中,或者该阅读理解装置200为该终端或服务器。该终端可以为智能手机、平板电脑、笔记本电脑、智能电视、智能音箱、穿戴式智能设备、个人计算机(Personal Computer,PC)等设备,终端还可以包括客户端,该客户端可以是视频客户端、浏览器客户端或即时通信客户端等。服务器可以是独立的物理服务器,也可以是多个物理服务器构成的服务器集群或者分布式系统,还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、内容分发网络(Content Delivery Network,CDN)、以及大数据和人工智能平台等基础云计算服务的云服务器。

[0140] 图9为本申请实施例提供的计算机设备的结构示意图,如图9所示,计算机设备300可以包括:通信接口301,存储器302,处理器303和通信总线304。通信接口301,存储器302,处理器303通过通信总线304实现相互间的通信。通信接口301用于装置300与外部设备进行数据通信。存储器302可用于存储软件程序以及模块,处理器303通过运行存储在存储器302的软件程序以及模块,例如前述方法实施例中的相应操作的软件程序。

[0141] 可选的,该处理器303可以调用存储在存储器302的软件程序以及模块执行如下操作:获取待处理的文本数据和图像数据,所述文本数据包括问题和所述问题对应的选项,所述图像数据包括场景图片;提取所述文本数据的文本向量表示,所述文本向量表示包含所述问题的文本信息和所述选项的文本信息;提取所述图像数据的图片向量表示;根据所述文本向量表示与所述图片向量表示,计算包含有文本信息和图像信息的多模态向量表示;根据所述多模态向量表示,计算出每一所述选项作为正确答案的概率值,以根据所述概率值从所述选项中确定出与所述问题和所述场景图片匹配的正确答案。

[0142] 可选的,该计算机设备300为该终端或服务器。该终端可以为智能手机、平板电脑、

笔记本电脑、智能电视、智能音箱、穿戴式智能设备、个人计算机等设备。该服务器可以是独立的物理服务器,也可以是多个物理服务器构成的服务器集群或者分布式系统,还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、CDN、以及大数据和人工智能平台等基础云计算服务的云服务器。

[0143] 可选的,本申请还提供了一种计算机设备,包括存储器和处理器,存储器中存储有计算机程序,该处理器执行计算机程序时实现上述各方法实施例中的步骤。

[0144] 本申请还提供了一种计算机可读存储介质,用于存储计算机程序。该计算机可读存储介质可应用于计算机设备,并且该计算机程序使得计算机设备执行本申请实施例中的阅读理解方法中的相应流程,为了简洁,在此不再赘述。

[0145] 本申请还提供了一种计算机程序产品,该计算机程序产品包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得计算机设备执行本申请实施例中的阅读理解方法中的相应流程,为了简洁,在此不再赘述。

[0146] 本申请还提供了一种计算机程序,该计算机程序包括计算机指令,计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得计算机设备执行本申请实施例中的阅读理解方法中的相应流程,为了简洁,在此不再赘述。

[0147] 应理解,本申请实施例的处理器可能是一种集成电路芯片,具有信号的处理能力。在实现过程中,上述方法实施例的各步骤可以通过处理器中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器可以是通用处理器、数字信号处理器(Digital Signal Processor,DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现成可编程门阵列(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也可以是任何常规的处理器等。结合本申请实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器,闪存、只读存储器,可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器,处理器读取存储器中的信息,结合其硬件完成上述方法的步骤。

[0148] 可以理解,本申请实施例中的存储器可以是易失性存储器或非易失性存储器,或可包括易失性和非易失性存储器两者。其中,非易失性存储器可以是只读存储器(Read-Only Memory,ROM)、可编程只读存储器(Programmable ROM,PROM)、可擦除可编程只读存储器(Erasable PROM,EPROM)、电可擦除可编程只读存储器(Electrically EPROM,EEPROM)或闪存。易失性存储器可以是随机存取存储器(Random Access Memory,RAM),其用作外部高速缓存。通过示例性但不是限制性说明,许多形式的RAM可用,例如静态随机存取存储器(Static RAM,SRAM)、动态随机存取存储器(Dynamic RAM,DRAM)、同步动态随机存取存储器(Synchronous DRAM,SDRAM)、双倍数据速率同步动态随机存取存储器(Double Data Rate SDRAM,DDR SDRAM)、增强型同步动态随机存取存储器(Enhanced SDRAM,ESDRAM)、同步连接动态随机存取存储器(Synchlink DRAM,SLDRAM)和直接内存总线随机存取存储器(Direct

Rambus RAM, DR RAM)。应注意,本文描述的系统和方法的存储器旨在包括但不限于这些和任意其它适合类型的存储器。

[0149] 应理解,上述存储器为示例性但不是限制性说明,例如,本申请实施例中的存储器还可以是静态随机存取存储器(static RAM,SRAM)、动态随机存取存储器(dynamic RAM,DRAM)、同步动态随机存取存储器(synchronous DRAM,SDRAM)、双倍数据速率同步动态随机存取存储器(double data rate SDRAM,DDR SDRAM)、增强型同步动态随机存取存储器(enhanced SDRAM,ESDRAM)、同步连接动态随机存取存储器(synch link DRAM,SLDRAM)以及直接内存总线随机存取存储器(Direct Rambus RAM,DR RAM)等等。也就是说,本申请实施例中的存储器旨在包括但不限于这些和任意其它适合类型的存储器。

[0150] 本领域普通技术人员可以意识到,结合本文中所公开的实施例描述的各示例的单元及算法步骤,能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能,但是这种实现不应认为超出本申请的范围。

[0151] 所属领域的技术人员可以清楚地了解到,为描述的方便和简洁,上述描述的系统、装置和单元的具体工作过程,可以参考前述方法实施例中的对应过程,在此不再赘述。

[0152] 在本申请所提供的几个实施例中,应该理解到,所揭露的系统、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,所述单元的划分,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。

[0153] 所述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。

[0154] 另外,在本申请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。

[0155] 所述功能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,服务器)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、ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0156] 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。

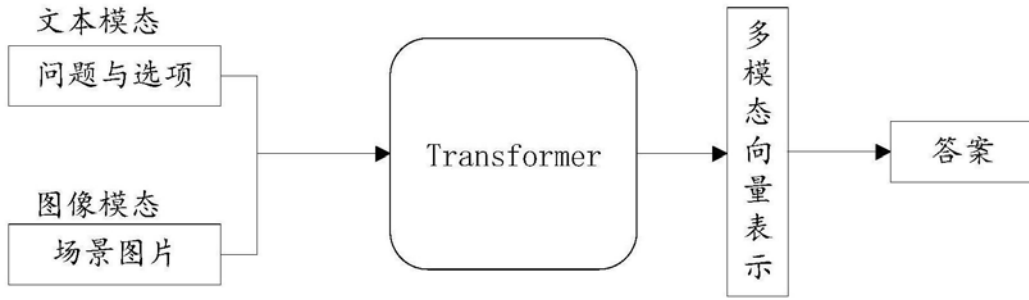


图1

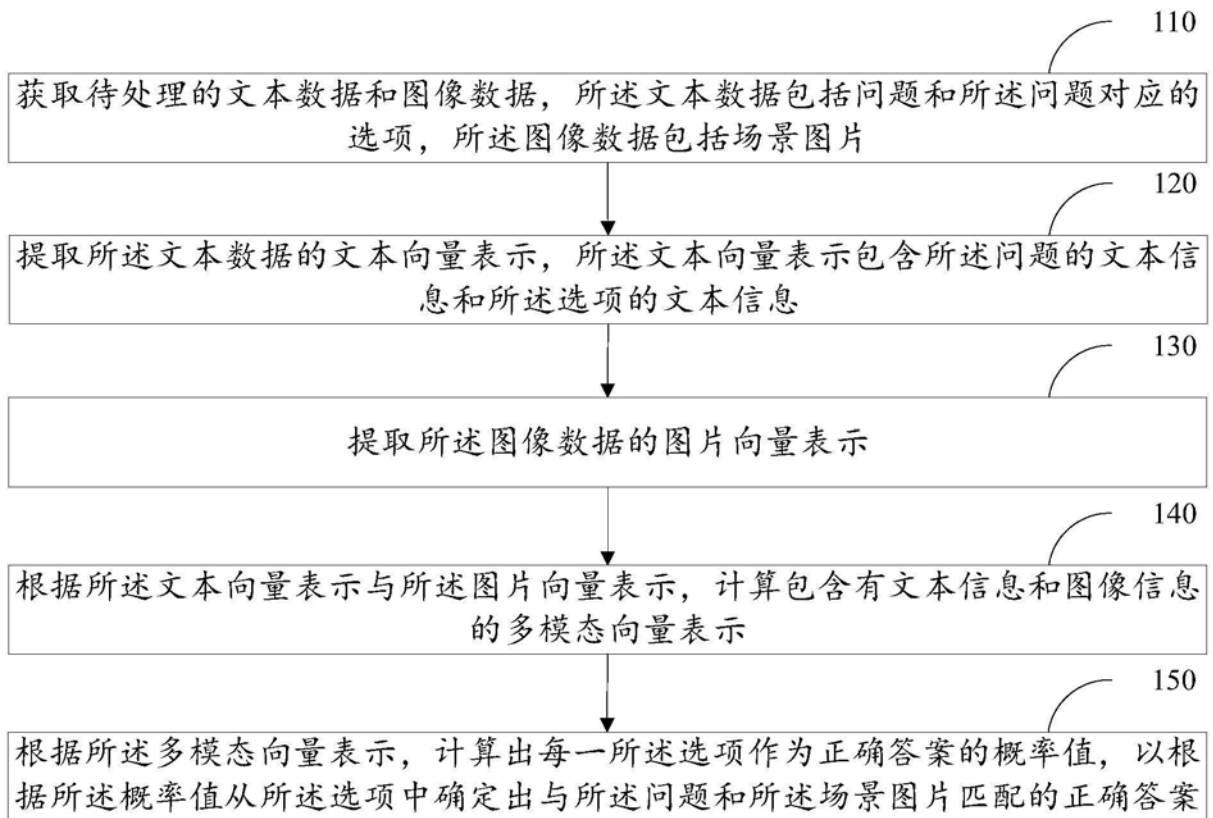


图2

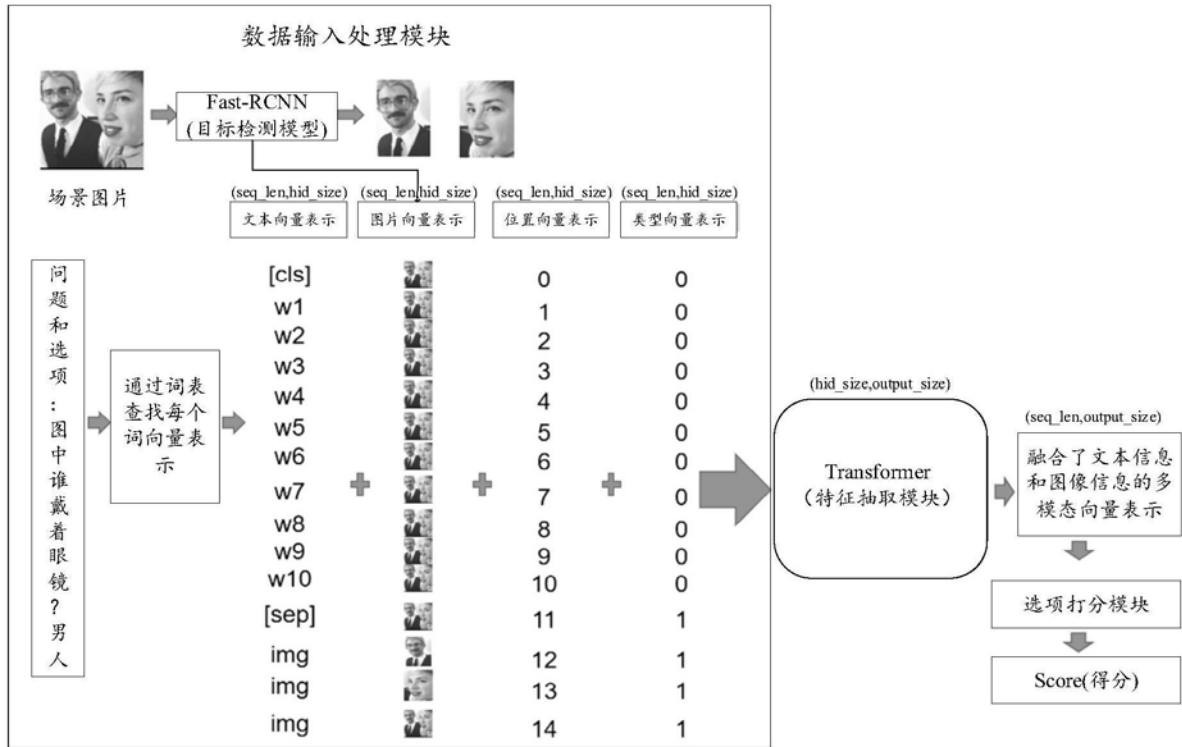


图3

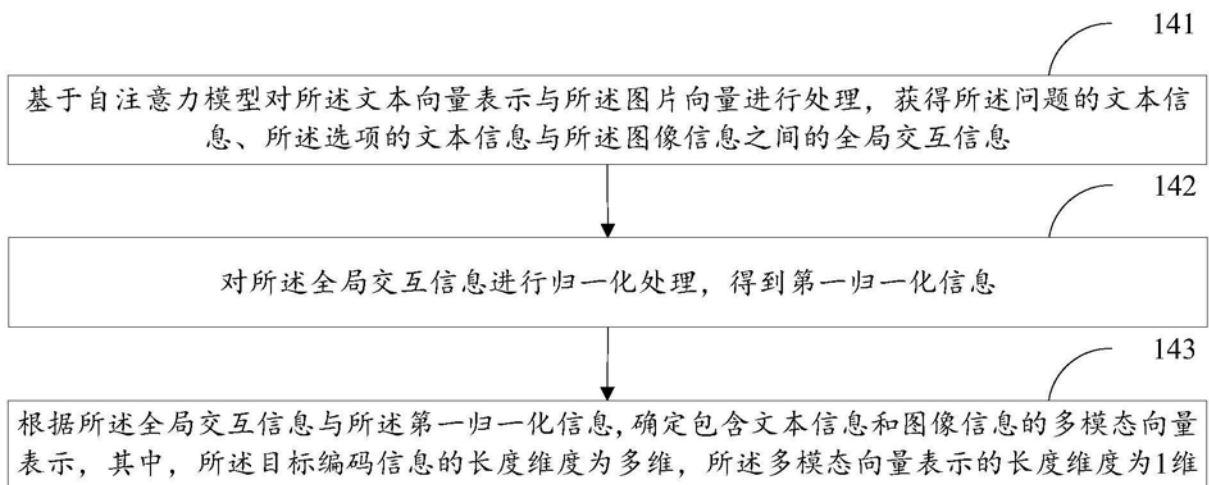


图4

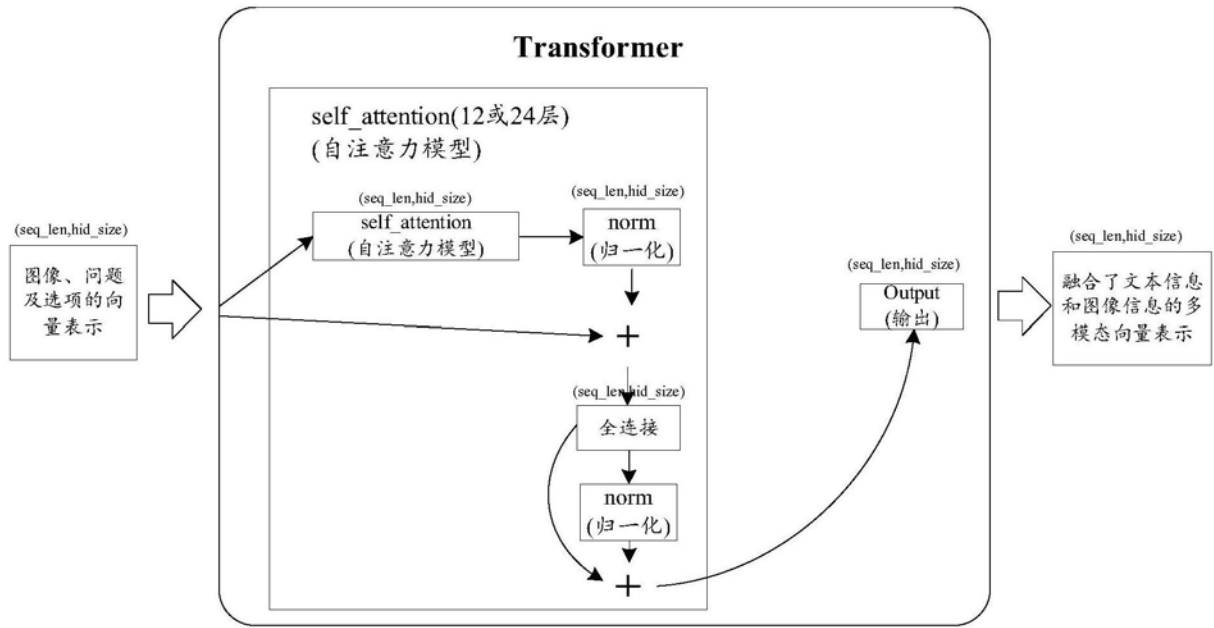


图5

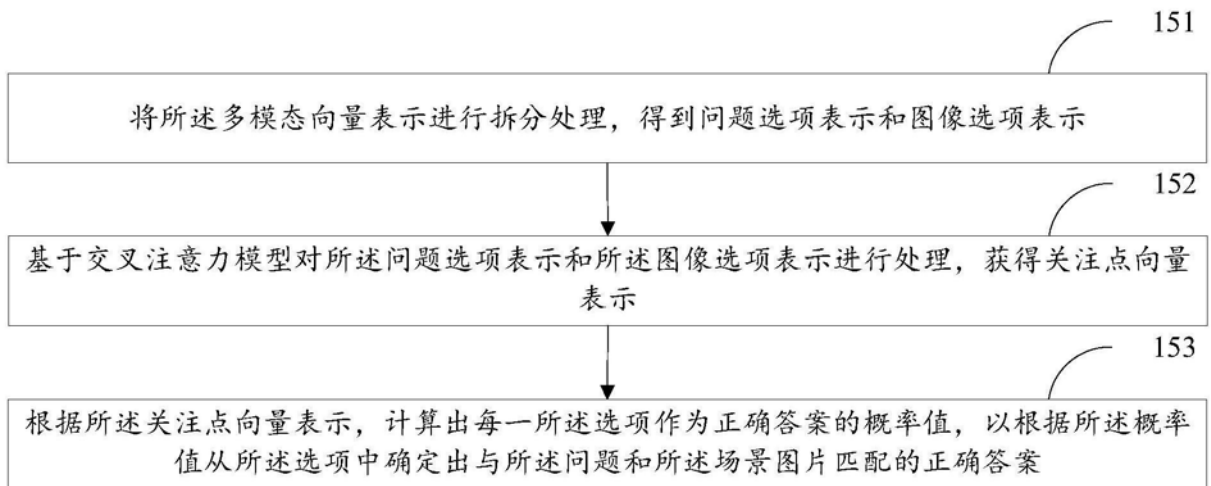


图6

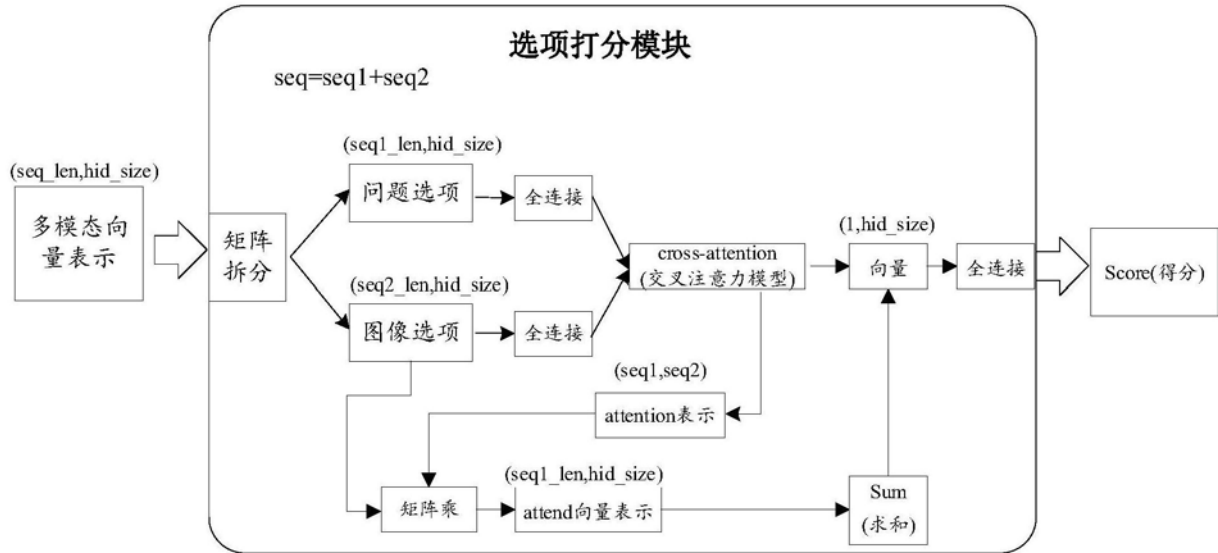


图7

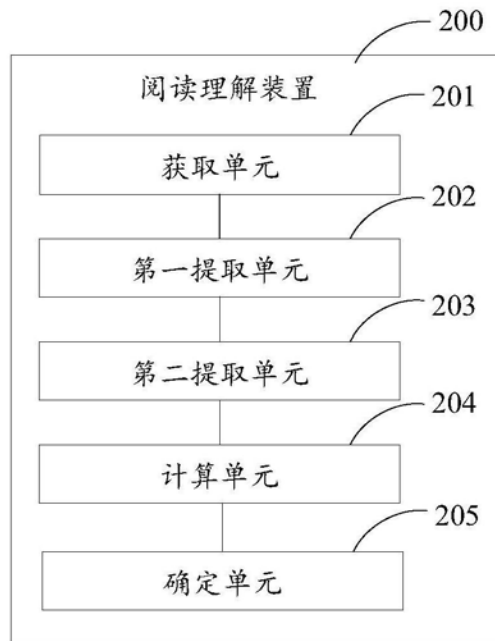


图8

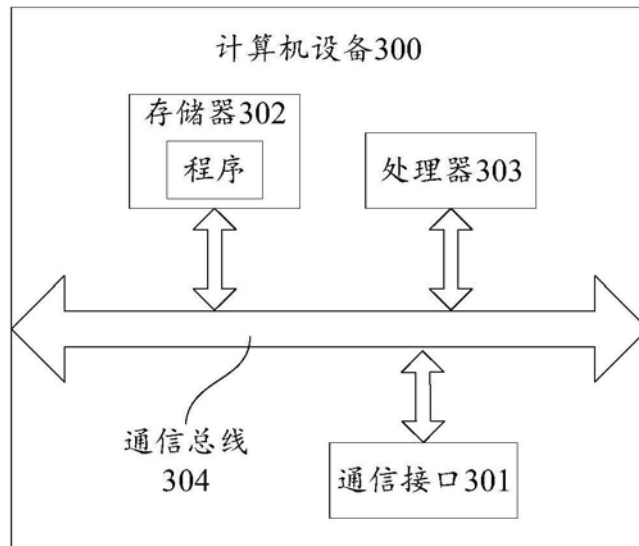


图9