



(12) 发明专利申请

(10) 申请公布号 CN 118715523 A

(43) 申请公布日 2024. 09. 27

(21) 申请号 202380023689.7

(22) 申请日 2023.03.16

(30) 优先权数据

63/320,633 2022.03.16 US

(85) PCT国际申请进入国家阶段日

2024.08.26

(86) PCT国际申请的申请数据

PCT/EP2023/056778 2023.03.16

(87) PCT国际申请的公布数据

W02023/175089 EN 2023.09.21

(71) 申请人 渊慧科技有限公司

地址 英国

(72) 发明人 雅各布·李·米尼克

维拉德米尔·米库力克

马嘉·玛利亚·特雷贝克兹

纳塔涅尔·约翰·麦卡利斯-帕克

杰弗里·伊尔温

(74) 专利代理机构 中原信达知识产权代理有限

责任公司 11219

专利代理师 朴金丹 周亚荣

(51) Int.Cl.

G06F 40/00 (2006.01)

G06N 3/02 (2006.01)

G06F 16/33 (2006.01)

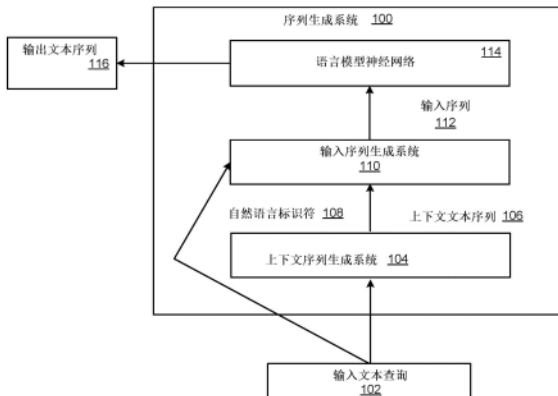
权利要求书3页 说明书19页 附图6页

(54) 发明名称

使用语言模型神经网络生成具有内联证据的输出序列

(57) 摘要

用于使用语言模型神经网络生成输出序列的方法、系统和装置,包括编码在计算机存储介质上的计算机程序。特别地,输出序列包括对输入查询的响应和内联证据,该内联证据包括来自支持响应的上下文文档的引述。



1. 一种由一个或多个计算机执行的方法,所述方法包括:

接收输入文本查询;

获得一个或多个第一上下文文本序列和所述第一上下文文本序列中的每一个的相应自然语言标识符;

生成第一输入序列,所述第一输入序列包括所述输入文本查询、所述一个或多个第一上下文文本序列以及所述一个或多个第一上下文文本序列中的每一个的所述相应自然语言标识符;

使用自回归语言模型神经网络处理所述第一输入文本序列以生成第一输出文本序列,所述第一输出文本序列包括:

(i) 第一输出文本子序列,所述第一输出文本子序列作为对所述输入文本查询的响应;

(ii) 第二输出文本子序列,所述第二输出文本子序列是所述第一上下文文本序列的所述相应自然语言标识符之一,以及

(iii) 第三输出文本子序列,所述第三输出文本子序列是来自由所述第二输出文本子序列中的所述自然语言标识符标识的所述第一上下文文本序列的文本;以及

响应于所述输入文本查询,提供至少所述第一输出文本子序列和所述第三输出文本子序列。

2. 根据权利要求1所述的方法,其中,响应于所述输入文本查询而提供至少所述第一输出文本子序列和所述第一上下文文本序列包括:

响应于所述查询,提供所述第一输出文本子序列、所述第二输出文本子序列和所述第三输出文本子序列。

3. 根据权利要求1或2所述的方法,进一步包括:

从所述第二输出文本子序列确定由所述第二输出文本子序列中的所述自然语言标识符标识的所述第一上下文文本序列的源;以及

响应于所述查询,提供对所述第一上下文文本序列的所述源的引用。

4. 根据任一前述权利要求所述的方法,进一步包括:

获得一个或多个第二上下文文本序列和所述第二上下文文本序列中的每一个的相应自然语言标识符;

生成第二输入序列,所述第二输入序列包括所述输入文本查询、所述一个或多个第二上下文文本序列以及所述一个或多个第二上下文文本序列中的每一个的所述相应自然语言标识符;

使用所述自回归语言模型神经网络处理所述第二输入文本序列以生成第二输出文本序列,所述第二输出文本序列包括:

(i) 第四输出文本子序列,所述第四输出文本子序列作为对所述输入文本查询的响应;

(ii) 第五输出文本子序列,所述第五输出文本子序列是所述第二上下文文本序列的所述相应自然语言标识符之一,以及

(iii) 第六输出文本子序列,所述第六输出文本子序列是来自由所述第五输出文本子序列中的所述自然语言标识符标识的所述第二上下文文本序列的文本;

为包括所述第一输出文本序列和所述第二输出文本序列的集合中的每个输出文本序列生成相应评分;

确定所述第一输出文本序列具有在所述集合中的任何输出文本序列的最高评分;以及
响应于确定所述第一输出文本序列具有所述最高评分,响应于所述输入文本查询,提供至少所述第一输出文本子序列和所述第三输出文本子序列。

5. 根据权利要求4所述的方法,其中,为包括所述第一输出文本序列和所述第二输出文本序列的集合中的每个输出文本序列生成相应评分包括:

使用学习的奖励模型对所述输出文本序列中的每一个进行评分。

6. 根据任一前述权利要求所述的方法,其中,所述第一输出序列包括多个时间步中的每个时间步处的来自词元词表的相应词元,其中,所述自回归神经网络被配置为,对于所述第一输出序列中的每个时间步,以所述第一输入文本序列和在所述第一输出序列中的所述时间步之前的任何时间步处的所述输出序列中的任何词元为条件,为所述词表中的每个词元生成相应评分,并且其中,生成所述第一输出序列包括:

在每个时间步处,使用针对所述时间步由所述神经网络生成的所述词表中的所述词元的所述相应评分来选择所述时间步处的所述词元。

7. 根据权利要求6所述的方法,其中,生成所述第一输出序列包括:

在所述第二输出文本子序列中的所述第一时间步之后的在所述第二输出文本子序列中的每个时间步处:

接收由所述神经网络在所述时间步处生成的所述相应评分;

生成约束评分分布,所述约束评分分布仅向紧接在所述自然语言标识符之一中的所述第二输出文本子序列内已经生成的所述词元之后的词元指派非零评分;以及

在所述时间步处从所述约束评分分布对所述词元进行采样。

8. 根据权利要求7所述的方法,其中,所述第二输出文本子序列之前是所述第一输出文本序列中的一个或多个第一预定语法词元,并且其中,生成所述第一输出序列包括:

在特定时间步处,确定已经在紧接在所述特定时间步之前的一个或多个时间步处选择了所述一个或多个第一预定语法词元,并且作为响应,确定所述特定时间步是所述第二输出文本子序列中的所述第一时间步;

接收由所述神经网络在所述特定时间步处生成的所述相应评分;

响应于确定所述特定时间步是所述第二输出文本子序列中的所述第一时间步,生成约束评分分布,所述约束评分分布仅向作为所述自然语言标识符之一中的所述第一词元的词元指派非零评分;以及

在所述时间步处从所述约束评分分布对所述词元进行采样。

9. 根据任一前述权利要求所述的方法,其中,生成所述第一输出序列包括:

在所述第三输出文本子序列中的所述第一时间步之后的在所述第三输出文本子序列中的每个时间步处:

接收由所述神经网络在所述时间步处生成的所述相应评分;

生成约束评分分布,所述约束评分分布仅向紧接在由所述第二输出文本子序列中的所述自然语言标识符标识的所述第一上下文文本序列中的所述第三输出文本子序列内已经生成的所述词元之后的词元指派非零评分;以及

在所述时间步处从所述约束评分分布对所述词元进行采样。

10. 根据权利要求9所述的方法,其中,所述第三输出文本子序列之前是所述第一输出

文本序列中的一个或多个第二预定语法词元,并且其中,生成所述第一输出序列包括:

在第二特定时间步处,确定已经在紧接在所述第二特定时间步之前的一个或多个时间步处选择了所述一个或多个第二预定语法词元,并且作为响应,确定所述特定时间步是所述第三输出文本子序列中的所述第一时间步;

接收由所述神经网络在所述特定时间步处生成的所述相应评分;

响应于确定所述特定时间步是所述第三输出文本子序列中的所述第一时间步,生成约束评分分布,所述约束评分分布仅向在由所述第二输出文本子序列中的所述自然语言标识符标识的所述第一上下文文本序列中出现的词元指派非零评分;以及

在所述时间步处从所述约束评分分布对所述词元进行采样。

11. 根据任一前述权利要求所述的方法,其中,获得一个或多个第一上下文文本序列和所述第一上下文文本序列中的每一个的相应自然语言标识符包括:

将从所述输入文本查询导出的查询提交给搜索引擎;

响应于所述查询,从所述搜索引擎获得一个或多个上下文文档;以及

从所述一个或多个上下文文档中选择所述一个或多个第一上下文序列。

12. 根据权利要求11所述的方法,其中,所述第一上下文文本序列中的每一个的所述相应自然语言标识符是从中选择所述第一上下文文本序列的所述上下文文档的标题。

13. 根据任一前述权利要求所述的方法,其中,所述神经网络已经通过对语言建模目标的无监督学习而被预训练。

14. 根据任一前述权利要求所述的方法,其中,所述神经网络已经通过监督学习、强化学习或两者进行了微调。

15. 一种系统,包括:

一个或多个计算机;以及

存储指令的一个或多个存储设备,所述指令在由所述一个或多个计算机执行时使所述一个或多个计算机执行根据权利要求1-14中任一项的相应操作。

16. 一个或多个计算机可读存储介质存储指令,所述指令在由一个或多个计算机执行时使所述一个或多个计算机执行根据权利要求1-14中任一项所述的方法的相应操作。

使用语言模型神经网络生成具有内联证据的输出序列

相关申请的交叉引用

[0001] 本申请要求2022年3月16日提交的美国专利申请序列号No. 63/320,633的优先权权益,其全部内容通过引用并入本文。

技术领域

[0002] 本说明书涉及使用神经网络处理输入以生成输出序列。

背景技术

[0003] 神经网络是采用一层或多层非线性单元来预测接收到的输入的输出的机器学习模型。除了输出层之外,一些神经网络还包括一个或更多个隐藏层。每个隐藏层的输出被用作网络中的下一层即另一隐藏层或输出层的输入。网络的每一层根据相应参数集的当前值从接收到的输入生成输出。

发明内容

[0004] 本说明书描述了一种在一个或多个位置中的一个或多个计算机上实现为计算机程序的系统,该系统使用语言模型神经网络生成对接收到的请求的响应。特别地,由系统生成的响应包括(i)对请求的响应和(ii)来自支持响应的一个或多个上下文文本文档的“证据”。证据包括来自上下文文本文档之一的直接引述。

[0005] 例如,系统可以提供用户与访问文档语料库的信息检索系统之间的接口。该接口允许系统利用信息检索系统来提供更可靠的信息,特别是可验证地正确的信息。

[0006] 在一个方面,一种方法包括:接收输入文本查询;获得一个或多个第一上下文文本序列和第一上下文文本序列中的每一个的相应自然语言标识符;生成第一输入序列,该第一输入序列包括输入文本查询、一个或多个第一上下文文本序列和一个或多个第一上下文文本序列中的每一个的相应自然语言标识符;使用自回归语言模型神经网络处理第一输入文本序列以生成第一输出文本序列,该第一输出文本序列包括:(i)第一输出文本子序列,该第一输出文本子序列是对输入文本查询的响应;(ii)第二输出文本子序列,该第二输出文本子序列是第一上下文文本序列的相应自然语言标识符之一,以及(iii)第三输出文本子序列,该第三输出文本子序列是来自由第二输出文本子序列中的自然语言标识符标识的第一上下文文本序列的文本;以及响应于输入文本查询,提供至少第一输出文本子序列和第三输出文本子序列。

[0007] 在一些实施方式中,响应于输入文本查询提供至少第一输出文本子序列和第一上下文文本序列包括响应于查询提供第一输出文本子序列、第二输出文本子序列和第三输出文本子序列。

[0008] 在一些实施方式中,该方法进一步包括从第二输出文本子序列确定由第二输出文本子序列中的自然语言标识符标识的第一上下文文本序列的源;以及响应于查询,提供对第一上下文文本序列的源的引用。

[0009] 在一些实施方式中,该方法进一步包括获得一个或多个第二上下文文本序列和每个第二上下文文本序列的相应自然语言标识符;生成包括输入文本查询、一个或多个第二上下文文本序列的第二输入序列,以及一个或多个第二上下文文本序列中的每一个的相应自然语言标识符;使用自回归语言模型神经网络处理第二输入文本序列以生成第二输出文本序列,第二输出文本序列包括:(i) 第四输出文本子序列,该第四输出文本子序列作为对输入文本查询的响应;(ii) 第五输出文本子序列,该第五输出文本子序列是第二上下文文本序列的相应自然语言标识符之一,以及(iii) 第六输出文本子序列,该第六输出文本子序列是来自由第五输出文本子序列中的自然语言标识符标识的第二上下文文本序列的文本;为包括第一输出文本序列和第二输出文本序列的集合中的每个输出文本序列生成相应评分;确定第一输出文本序列具有集合中的任何输出文本序列的最高评分;以及响应于确定第一输出文本序列具有最高评分,响应于输入文本查询,提供至少第一输出文本子序列和第三输出文本子序列。

[0010] 在一些实施方式中,为包括第一输出文本序列和第二输出文本序列的集合中的每个输出文本序列生成相应评分包括:使用学习的奖励模型对每个输出文本序列进行评分。

[0011] 在一些实施方式中,第一输出序列包括多个时间步中的每个时间步处的来自词元词表的相应词元,其中,自回归神经网络被配置为,对于第一输出序列中的每个时间步(例如,对于多个时间步中的每个时间步;与当前时间步相对应的词元可以方便地被称为“当前词元”),以第一输入文本序列和第一输出序列中的时间步之前的任何时间步处的输出序列中的任何词元(输出序列中的在当前词元之前的任何词元)为条件,为词表中的每个词元生成相应评分,并且其中生成第一输出序列包括:在每个时间步处,使用由神经网络为时间步生成的词表中的词元的相应评分来选择时间步处的词元(当前词元)。

[0012] 在一些实施方式中,第二输出文本子序列的词元还与(第二)多个时间步中的对应时间步相对应。生成第一输出序列包括:在第二输出文本子序列中的第一时间步之后的第二输出文本子序列中的每个时间步(第二多个时间步中的每一个)处:接收由神经网络在该时间步处生成的相应评分;生成约束评分分布,该约束评分分布仅向紧接在自然语言标识符之一中的第二输出文本子序列内已经生成的词元之后的词元指派非零评分;以及在该时间步处从约束评分分布对词元进行采样。

[0013] 在一些实施方式中,第二输出文本子序列之前是第一输出文本序列中的一个或多个第一预定语法词元,并且生成第一输出序列包括:在特定时间步,确定已经在紧接在特定时间步之前的一个或多个时间步处选择了一个或多个第一预定语法词元,并且作为响应,确定特定时间步是第二输出文本子序列中的第一时间步;接收由神经网络在特定时间步处生成的相应评分;响应于确定特定时间步是第二输出文本子序列中的第一时间步,生成约束评分分布,该约束评分分布仅向作为自然语言标识符之一中的第一词元的词元指派非零评分;以及根据约束评分分布在该时间步处对词元进行采样。

[0014] 在一些实施方式中,第三输出文本子序列的词元还与(第三)多个时间步中的对应时间步相对应。生成第一输出序列包括:在第三输出文本子序列中的第一时间步之后的第三输出文本子序列中的每个时间步(即,第三多个时间步中的每一个)处:接收由神经网络在时间步处生成的相应评分;生成约束评分分布,该约束评分分布仅向紧接在由第二输出文本子序列中的自然语言标识符标识的第一上下文文本序列中的第三输出文本子序列内

已经生成的词元之后的词元指派非零评分;以及在时间步处从约束评分分布中对词元进行采样。

[0015] 在一些实施方式中,第三输出文本子序列之前是第一输出文本序列中的一个或多个第二预定语法词元,并且生成第一输出序列包括:在第二特定时间步处,确定已经在紧接在第二特定时间步之前的一个或多个时间步处选择了一个或多个第二预定语法词元,并且作为响应,确定特定时间步是第三输出文本子序列中的第一时间步;接收由神经网络在特定时间步生成的相应评分;响应于确定特定时间步是第三输出文本子序列中的第一时间步,生成约束评分分布,该约束评分分布仅向在由第二输出文本子序列中的自然语言标识符标识的第一上下文文本序列中出现的词元指派非零评分;以及在时间步处从约束评分分布中对词元进行采样。

[0016] 在一些实施方式中,获得一个或多个第一上下文文本序列和第一上下文文本序列中的每一个的相应自然语言标识符包括:将从输入文本查询导出的查询提交给搜索引擎;响应于查询,从搜索引擎获得一个或多个上下文文档;以及从一个或多个上下文文档中选择一个或多个第一上下文序列。

[0017] 在一些实施方式中,第一上下文文本序列中的每一个的相应自然语言标识符是从中选择第一上下文文本序列的上下文文档的标题。

[0018] 在一些实施方式中,神经网络已经通过对语言建模目标的无监督学习进行了预训练。

[0019] 在一些实施方式中,神经网络已经通过监督学习、强化学习或两者进行了微调。

[0020] 本说明书中描述的主题可以在特定实施例实现,以便实现以下优点中的一个或多个。

[0021] 本说明书中描述的系统提供了一种用于访问生成语言模型神经网络的用户接口,该生成语言模型神经网络生成对接收到的请求的响应。特别地,生成语言模型(LM)对于回答关于世界的问题越来越有用。然而,默认情况下,LM生成用户必须选择盲目地接受或验证自己的无基础声明(ungrounded claim)。

[0022] 本说明书描述了帮助用户通过生成声明以及支持证据来评估由LM生成的响应的技术。特别地,该证据采取从一个或多个文本数据库检索的较长上下文文档中提取的逐字引述的形式。文档可以由因特网搜索引擎或任何其他合适的信息检索系统检索。因此,本系统提供了用户与信息检索系统之间的用户界面,并且其增强了使用信息检索系统获得的信息的可靠性和可验证性。

[0023] 为了用生成方法确保引述是“逐字的”,本说明书描述了语言模型在从文档引述时使用的特殊语法,并且在一些情况下,基于该语法,将语言模型的输出约束为来自检索到的文档的精确引述。这可以确保语言模型准确地从上下文文档引述,即使该模型是在不需要从输入引述的目标上预训练的。

[0024] 此外,被实现为神经网络的大规模语言模型可以在包括问答的一系列自然语言处理任务上产生令人印象深刻的结果。然而,一些这些模型——特别是基于Transformer的模型——的实施方式可以具有多于十亿的参数,并且可能需要大量的计算资源、功率和时间来处理网络输入以生成网络输出。有时,这样的模型可以具有多于100亿或多于1000亿的参数。如果这样的模型被大规模使用以服务大量用户请求,则将消耗显著的能量。

[0025] 当神经网络在例如移动设备的数字助理设备上实现时,出现附加考虑,该数字助理设备在计算系统中实现,该计算系统包括通过诸如因特网的数据通信网络与数字助理设备通信的后端组件,特别是数据服务器。然后,需要优化数字助理设备和后端组件之间的计算负载。该需求对于大规模语言模型可能特别严重,因为与典型在移动设备上发现的那些内存和计算要求相比它的大量的内存和计算要求。

[0026] 本文描述的技术解决了这些问题。在一些实施方式中,所描述的技术促进减少的计算负载和改进的负载分布,特别是当大规模语言模型被实现为多任务和并行处理计算机系统神经网络、跨多个站点分布并通过数据通信网络互连时。

[0027] 在一些实施方式中,所描述的技术使得能够在网络中的本地移动计算设备和后端服务器之间有益地分配计算负载。更具体地,在实施方式中,通过基于问题在表示从因特网搜索获得的文档的上下文上调节语言模型神经网络,使得能够使用较小的语言模型神经网络,这有助于在具有有限内存和计算资源的移动设备上实现神经网络。

[0028] 此外,使用本说明书中描述的技术,系统可以利用搜索引擎结果来使用搜索引擎结果中包括的最新信息生成关于输入文本的预测。一些现有系统使用预训练的神经网络而不访问这样的搜索引擎结果来生成预测,因此预测可能不太可靠,因为神经网络只能对训练期间神经网络可用的信息进行编码;也就是说,这些预测可能依赖于陈旧的信息,因此不正确或至少过时。因此,使用本说明书中描述的技术,系统可以生成更准确和适时的预测。

[0029] 此外,一些现有系统必须重复地重新训练神经网络以确保神经网络对最新信息进行编码。因为本说明书中描述的系统可以重复访问新的搜索引擎结果,所以系统不需要重新训练神经网络,从而节省了显著的计算资源。

[0030] 使用本说明书中描述的技术,系统可以响应于处理搜索引擎查询,使用在由搜索引擎提供的多个不同文档中编码的信息来生成对输入文本的预测。多个不同文档可以各自包括与预测相关的相应不同信息。因此,由系统生成的预测可以比使用单个文档生成的预测更准确。

[0031] 在附图和下面的描述中阐述了本说明书的主题的一个或多个实施例的细节。

[0032] 根据说明书、附图和权利要求,主题的其他特征、方面和优点将变得显而易见。

附图说明

[0033] 图1是示例序列生成系统的图。

[0034] 图2是用于生成输出序列的示例过程的流程图。

[0035] 图3是用于选择候选输出序列的示例过程的流程图。

[0036] 图4示出了向用户呈现输出序列的示例用户界面。

[0037] 图5示出了训练语言模型神经网络的示例。

[0038] 图6示出了用于对生成的样本进行评级的示例用户界面。

[0039] 各个附图中相同的附图词元和标号表示相同的元件。

具体实施方式

[0040] 图1示出了示例序列生成系统100。序列生成系统100是被实现为一个或多个位置中的一个或多个计算机上的计算机程序的系统的示例,其中可以实现下面描述的系统、组

件和技术。

[0041] 序列生成系统100充当访问一个或多个文本数据库(未示出)的信息检索系统的用户界面,或者为在与序列生成系统100分离但与其通信的用户计算机上实现的用户界面提供功能。文本数据库共同形成文档语料库。文档语料库可以例如是网页和可通过因特网访问的其他文档。可替代地,文档语料库可以例如是例如科学发布者或其他组织的专有文本数据库的一部分。序列生成系统100使用上下文序列生成系统104、输入序列生成系统110和语言模型神经网络114来处理来自用户的输入文本查询102,以生成输出序列116。

[0042] 输入文本查询102可以是由用户通过用户计算机提交给系统100的查询、由通过用户计算机提交给系统100的问题、或需要来自系统100的响应的不同请求。在一些情况下,系统从用户计算机接收查询作为文本。在一些其他情况下,系统从用户接收自然语言语音查询,并且通过将语音识别引擎应用于语音来将语音转换为输入文本查询102。输入文本查询102可以以由用户计算机的麦克风捕获的声音(语音)信号的形式接收,其由语音识别引擎——即语音到文本转换器——转换以形成输入文本查询102。可替代地,可以通过使用用户计算机的数据输入设备键入来输入。

[0043] 一旦系统100接收到输入文本查询102,上下文序列生成系统104就获得一个或多个第一上下文文本序列106和第一上下文文本序列106中的每一个的相应自然语言标识符108。

[0044] 例如,可以从相应的上下文文档中提取每个上下文文本序列106,并且标识符108可以是上下文文档的标题。作为另一示例,可以从同一上下文文档中提取上下文文本序列106中的一些或全部,并且标识符108可以是其提取上下文文本序列的文档的部分的章节头部或其他标识符。

[0045] 下面参考图2更详细地描述获得上下文序列。

[0046] 然后,输入序列生成系统110生成第一输入序列112,该第一输入序列112包括输入文本查询102、一个或多个第一上下文文本序列106以及一个或多个第一上下文文本序列中的每一个的相应自然语言标识符108。

[0047] 例如,第一输入序列112可以包括根据预定输入语法布置的查询102、上下文文本序列106和标识符108。在一些情况下,第一输入序列112还可以包括其他文本,例如,一个或多个自然语言“提示”、分离输入序列的各种元素的一个或多个分隔符词元或两者。自然语言提示是输入-输出对的示例,其中输入是可以提供的输入的示例,并且输出是应该生成的输出的示例。下面将更详细地描述提示。

[0048] 然后,序列生成系统100使用自回归语言模型神经网络114处理第一输入序列112以生成第一输出文本序列116。

[0049] 输出序列116包括(i)作为对输入文本查询102的响应的第一输出文本子序列,(ii)作为第一上下文文本序列106的相应自然语言标识符108之一的第二输出文本子序列,以及(iii)作为来自第二输出文本子序列中的自然语言标识符标识的第一上下文文本序列的文本的第三输出文本子序列。

[0050] 特别地,(i)、(ii)和(iii)根据预定的输出语法布置在输出序列内。下文参考图3更详细地描述预定语法的一个示例。

[0051] 然后,序列生成系统100响应于输入文本查询102提供至少第一输出文本子序列和

第三输出文本子序列。因此,系统100提供对输入文本查询102的文本响应和来自上下文文本序列106之一的文本作为文本响应的支持证据。

[0052] 在一些实施方式中,序列生成系统100响应于输入查询102生成多个候选输出序列116。

[0053] 在这些实施方式中,系统100还为每个候选输出序列生成相应的评分,并且仅响应于用户查询而提供来自评分最高的候选输出序列的文本。

[0054] 在这些实施方式中的一些中,如果没有候选者具有超过阈值的评分,则系统100代替地发出对用户查询的默认文本响应,例如,“I don't know(我不知道)”或“I am not sure(我不确定)”。

[0055] 下面参考图3描述对候选输出序列进行评分。

[0056] 语言模型神经网络114可以是任何适当的语言模型神经网络,其接收由从词表中选择的文本词元组成的输入序列,并且自回归地生成由来自词表的文本词元组成的输出序列。例如,语言模型神经网络114可以是基于Transformer的语言模型神经网络或基于递归神经网络的语言模型。

[0057] 词表中的词元可以是任何适当的文本词元,例如词、词片、标点符号等,其表示一种或多种自然语言的文本元素,以及可选地,在文本语料库中找到的数字和其他文本符号。通常,输入文本查询102、自然语言标识符108和/或上下文文本序列106也是从词表中选择的词元序列。

[0058] 语言模型神经网络114被称为自回归神经网络,因为神经网络114通过以当前输入序列为条件生成输出序列中的每个特定词元来自回归地生成词元的输出序列,该当前输入序列包括在输出序列中的特定文本词元之前的任何词元,即,已经针对输出序列中的在特定词元的特定位置之前的任何先前位置生成的词元,以及为输出序列提供上下文的上下文输入。

[0059] 例如,当在输出序列中的任何给定位置处生成词元时,当前输入序列可以包括输入序列和在输出序列中的给定位置之前的任何先前位置处的输出序列的词元。作为特定示例,当前输入序列可以包括输入序列,之后是在输出序列中给定位置之前的任何先前位置处的词元。可选地,在当前输入序列内,输入序列和来自输出序列的词元可以由当前输入序列内的一个或多个预定词元——即,来自词表的一个或多个词元的指定集合——分开。也就是说,在输入序列和来自输出序列的词元之间可以存在一个或多个预定词元。

[0060] 更具体地,为了在输出序列内的特定位置处生成特定词元,神经网络114可以处理当前输入序列以生成例如概率分布的评分分布,该评分分布向词元词表中的每个词元指派例如相应概率的相应评分。然后,神经网络114可以使用评分分布从词表中选择词元作为特定词元。例如,神经网络114可以贪婪地选择评分最高的词元,或者可以例如使用核采样或另一采样技术从分布中对词元进行采样。

[0061] 作为特定示例,语言模型神经网络114可以是基于自回归变换的神经网络,其包括(i)多个注意力块,每个注意力块应用自注意力操作,以及(ii)输出子网络,其处理最后一个注意力块的输出以生成评分分布。

[0062] 神经网络114可以具有各种基于Transformer的神经网络架构中的任何一种。这样的架构的示例包括以下各项中描述的那些架构:J. Hoffmann, S. Borgeaud, A. Mensch,

E. Buchatskaya, T. Cai, E. Rutherford, D. d. L. Casas, L. A. Hendricks, J. Welbl, A. Clark, et al. Training compute-optimal large language models(训练计算最优的大语言模型), arXiv preprint arXiv:2203.15556, 2022; J.W. Rae, S. Borgeaud, T. Cai, K. Millican, J. Hoffmann, H. F. Song, J. Aslanides, S. Henderson, R. Ring, S. Young, E. Rutherford, T. Hennigan, J. Menick, A. Cassirer, R. Powell, G. van den Driessche, L. A. Hendricks, M. Rauh, P. Huang, A. Glaese, J. Welbl, S. Dathathri, S. Huang, J. Uesato, J. Mellor, I. Higgins, A. Creswell, N. McAleese, A.Wu, E. Elsen, S. M. Jayakumar, E. Buchatskaya, D. Budden, E. Sutherland, K. Simonyan, M. Paganini, L. Sifre, L. Martens, X. L. Li, A. Kuncoro, A. Nematzadeh, E. Gribovskaya, D. Donato, A. Lazaridou, A. Mensch, J. Lespiau, M. Tsimpoukelli, N. Grigorev, D. Fritz, T. Sottiaux, M. Pajarskas, T. Pohlen, Z. Gong, D. Toyama, C. de Masson d'Autume, Y. Li, T. Terzi, V. Mikulik, I. Babuschkin, A. Clark, D. de Las Casas, A. Guy, C. Jones, J. Bradbury, M. Johnson, B. A. Hechtman, L. Weidinger, I. Gabriel, W. S. Isaac, E. Lockhart, S. Osindero, L. Rimell, C. Dyer, O. Vinyals, K. Ayoub, J. Stanway, L. Bennett, D. Hassabis, K. Kavukcuoglu, and G. Irving. Scaling language models: Methods, analysis & insights from training gopher(扩展语言模型:来自训练gopher的方法、分析和见解). CoRR, abs/2112.11446, 2021; Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer(利用统一的文本到文本transformer探索迁移学习的极限). arXiv preprint arXiv:1910.10683, 2019; Daniel Adiwardana, Minh-Thang Luong, David R. So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, and Quoc V. Le. Towards a human-like open-domain chatbot(迈向类似人类的开放域聊天机器人). CoRR, abs/2001.09977, 2020; and Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners(语言模型是少样本学习者). arXiv preprint arXiv:2005.14165, 2020.

[0063] 然而,通常,基于Transformer的神经网络包括注意力块序列,并且在给定输入序列的处理期间,序列中的每个注意力块接收给定输入序列中的每个输入词元的相应输入隐藏状态。然后,注意力块至少部分地通过应用自注意力来更新隐藏状态中的每一个,以生成输入词元中的每一个的相应输出隐藏状态。第一注意力块的输入隐藏状态是输入序列中的输入词元的嵌入,并且每个后续注意力块的输入隐藏状态是由先前注意力块生成的输出隐藏状态。

[0064] 在该示例中,输出子网络处理由输入序列中的最后输入词元的序列中的最后注意力块生成的输出隐藏状态,以生成评分分布。

[0065] 通常,因为神经网络114是自回归的,所以系统100可以使用相同的神经网络114来

响应于相同的请求而生成多个不同的候选输出序列,例如,通过使用来自神经网络114生成的评分分布的波束搜索解码,使用采样和排名解码策略,通过使用用于对通过神经网络114的不同运行进行采样的伪随机数生成器的不同随机种子,或者使用利用神经网络114的自回归性质的另一解码策略。

[0066] 在一些实施方式中,语言模型114被预训练,即,在不需要响应于用户问题提供证据的语言建模任务上训练,并且系统100使神经网络114通过输入序列中的自然语言提示根据预定语法生成输出序列。

[0067] 例如,系统100或另一训练系统在语言建模任务上预训练语言模型神经网络114,例如,需要在给定当前文本词元序列的情况下预测训练数据中当前序列之后的下一个词元的任务。作为特定示例,语言模型神经网络114可以在大文本数据集上的最大似然目标上预训练,例如,从因特网或另一文本语料库公开可用的文本。

[0068] 在一些其他实施方式中,在预训练之后,系统100例如通过监督学习、强化学习或两者对确实需要根据语法生成输出序列的目标微调语言模型114。这在下面参考图5更详细地描述。

[0069] 在这些实施方式中的一些中,系统100仍然在推理时,即在训练之后,在语言模型114的输入中包括一个或多个自然语言提示。

[0070] 如上所述,自然语言提示是输入-输出对的示例,其中输入是可以提供的输入的示例,并且输出是应该生成的输出的示例。因此,每个提示将包括示例查询的示例输入序列、一个或多个上下文序列的示例集合以及根据预定输入语法布置的一个或多个上下文序列的相应标识符。每个提示还将包括根据输出语法布置作为对示例查询的响应的示例第一输出文本子序列、作为示例上下文文本序列之一的相应自然语言标识符之一的示例第二输出文本子序列、以及作为来自示例第二输出文本子序列中的自然语言标识符标识的示例上下文文本序列的文本的示例第三输出文本子序列。可选地,输入序列还可以包括来自将每个提示分开的词表的一个或多个词元以及将最终提示与用户查询分开的一个或多个词元。

[0071] 另外,在一些实施方式中,系统100在选择要包括在输出序列中的词元时执行“约束采样”。这可以确保神经网络114的输出遵循语法并且序列在内部一致,即,确保证据是来自输出序列中的自然语言标识符108标识的上下文文本序列106的直接引述。

[0072] 在系统100生成多个候选输出序列的情况下,约束采样防止系统必须对无效或不一致的输出序列进行评分,并且极大减少了需要生成以确保高质量输出的候选的数量,从而大大改进了系统100的计算效率,即减少了系统100消耗的计算资源的量。

[0073] 下面参考图3更详细地描述约束采样的示例。

[0074] 图2是用于在给定输入查询的情况下生成输出序列的示例过程200的流程图。为了方便起见,过程200将被描述为由位于一个或多个位置的一个或多个计算机的系统执行。例如,根据本说明书适当编程的序列生成系统,例如图1中描绘的序列生成系统100,可以执行过程200。

[0075] 系统例如使用用户界面从用户接收输入文本查询(步骤202)。

[0076] 系统获得一个或多个第一上下文文本序列和第一上下文文本序列中的每一个的相应自然语言标识符(步骤204)。

[0077] 例如,系统可以通过向搜索引擎提交从输入文本查询导出的搜索查询来获得一个

或多个上下文序列和每个第一上下文序列的相应自然语言标识符。搜索引擎具有对文档语料库的访问,并且被配置为基于搜索查询来搜索文档语料库。例如,搜索查询可以是与输入文本查询相同的文本,或者可以由系统修改,例如,以添加同义词、纠正打字或拼写错误等。

[0078] 然后,系统可以响应于搜索查询从搜索引擎获得一个或多个文档。一个或多个文档可以由搜索引擎例如根据与接收到的搜索查询的质量和相关性来排名。

[0079] 然后,系统可以例如通过选择一个或多个排名最高的搜索结果来从一个或多个上下文文档中选择一个或多个第一上下文序列。该系统还将相应的自然语言标识符与每个第一上下文序列相关联。

[0080] 在一些实施方式中,搜索引擎还提供来自对应上下文文档的摘录作为标识对应上下文文档的搜索结果的一部分。在这些实施方式中的一些中,系统可以通过从对应的上下文文档中提取摘录和摘录周围的文本来生成给定文档的上下文序列。例如,系统可以使用摘录提取摘录文本,以便说明文档长度被改变以及文档长度通常超过语言模型最大上下文窗口大小(如下所述)。

[0081] 因此,特别是在一次呈现多个文档时的少样本提示(few shot prompting)的情况下,系统可能需要限制在给定输入序列内的文档内容上花费的词元的数量。所以,系统可以通过使用如上所述的摘录对文档进行截断。例如,系统可以使用摘录来将给定文档截断为最大词元长度片段,使得该片段包含相关搜索摘录。

[0082] 在一些实施方式中,系统可以确保截断的片段从句子或段落的开头开始。

[0083] 作为特定示例,在训练时,系统可以随机选择这样的起始位置以增加输入的种类。在推理时,系统可以允许在摘录片段的开始之前的最大数量,例如250、500或1000个字符,并标识在该范围内开始的第一句子,并使用该第一句子作为截断片段的开始。

[0084] 搜索引擎可以是可以由系统访问并且搜索例如网页、书籍或其他文档的任何适当的文档语料库的任何适当的搜索引擎。例如,搜索引擎可以是搜索并返回引用因特网上可用的文档的结果的因特网搜索引擎。作为另一示例,搜索引擎可以是搜索私有文档语料库——例如在内部网络上可用或在一个或多个数据库的合集中存储的文档——的不同搜索引擎。

[0085] 例如,第一上下文文本序列中的每一个的相应自然语言标识符可以是从中选择第一上下文文本序列的上下文文档的标题。

[0086] 系统生成第一输入序列,该第一输入序列包括输入文本查询、一个或多个第一上下文文本序列以及一个或多个第一上下文文本序列中的每一个的相应自然语言标识符(步骤206)。

[0087] 系统使用自回归语言模型神经网络处理第一输入文本序列以生成第一输出文本序列(步骤208)。

[0088] 第一输出文本序列包括作为对输入文本查询的响应的第一输出文本子序列、作为第一上下文文本序列的相应自然语言标识符之一的第二输出文本子序列、以及作为来自第二输出文本子序列中的自然语言标识符标识的第一上下文文本序列的文本的第三输出文本子序列。

[0089] 响应于输入文本查询,系统(例如,向用户)提供至少第一输出文本子序列和第三输出文本子序列(步骤210)。

[0090] 系统可以响应于查询提供第一输出文本子序列、第三输出文本子序列以及可选地提供第二输出文本子序列。

[0091] 另外,在一些实施方式中,系统可以从第二输出文本子序列确定由第二输出文本子序列中的自然语言标识符标识的第一上下文文本序列的源,并响应于查询提供对第一上下文文本序列的源的引用。例如,系统可以提供引用作为链接到第一上下文文本序列的源——例如网页——的超链接。

[0092] 下面参考图4描述由系统生成的输出序列的示例呈现。

[0093] 如上所述,在一些实施方式中,系统生成多个候选输出序列的集合(其包括第一输出文本序列),并且对于每个候选输出序列,生成相应的评分,并且仅响应于确定第一输出文本序列在任何候选输出序列中具有最高评分而提供第一输出序列。

[0094] 例如,当处理第一输入文本序列时,系统可以通过从由语言模型神经网络生成的输出中采样不同的候选输出序列来生成集合中的候选输出序列中的至少一些。

[0095] 另外,在一些实施方式中,系统可以生成比可以适合语言模型神经网络的“上下文窗口”更多的上下文序列。也就是说,语言模型神经网络可能例如由于内存约束或由于其中神经网络被训练的框架,仅能够处理不多于最大字符数的输入序列。在一些实施方式中,包括所有上下文序列的自然语言标识符和词元可以超过该最大数量。在这些实施方式中,系统生成多个不同的输入序列,每个输入序列包括上下文序列的相应子集。

[0096] 换句话说,除了第一上下文文本序列之外,系统还可以获得一个或多个第二上下文文本序列和第二上下文文本序列中的每一个的相应自然语言标识符,并且生成第二输入序列,该第二输入序列包括输入文本查询、一个或多个第二上下文文本序列和一个或多个第二上下文文本序列中的每一个的相应自然语言标识符。然后,系统可以使用自回归语言模型神经网络处理第二输入文本序列以生成第二输出文本序列,该第二输出文本序列包括:(i)作为对输入文本查询的响应的第四输出文本子序列;(ii)作为第二上下文文本序列的相应自然语言标识符之一的第五输出文本子序列,以及(iii)作为来自第五输出文本子序列中的自然语言标识符标识的第二上下文文本序列的文本的第六输出文本子序列。

[0097] 然后,系统针对集合——例如包括第一输出文本序列和第二输出文本序列的集合——中的每个候选输出文本序列生成相应评分,并且确定第一输出文本序列具有集合中的任何输出文本序列的最高评分。在一些情况下,这可以通过使用学习的奖励模型对输出文本序列中的每一个进行评分来完成。下面参考图3描述使用学习的奖励模型对候选输出序列进行评分。

[0098] 然后,系统可以响应于确定第一输出文本序列具有最高评分,响应于输入文本查询,提供至少第一输出文本子序列和第三输出文本子序列。

[0099] 图3示出了当系统响应于给定文本查询生成多个候选输出序列时序列生成系统的操作的示例。

[0100] 如图3的示例中所示,系统100例如从用户计算机接收问题302。

[0101] 系统100执行因特网搜索304以标识与问题302最相关的前k个文档。通常,k是大于1的整数,例如5、10、20或100。例如,系统100可以将问题302或从问题302导出的查询提供给因特网搜索引擎,并且从因特网搜索引擎获得标识前k个文档的搜索结果。

[0102] 然后,系统使用生成器306来生成到语言模型神经网络114的一个或多个输入序

列,并使用语言模型神经网络114对N个候选输出序列进行采样308。在一些实施方式中,候选输出序列的数量N大于文档的数量k。

[0103] 例如,生成器306可以生成包括来自所有k个文档的上下文的单个输入序列,然后使用语言模型神经网络114多次处理单个输入序列以对N个候选输出序列进行采样。

[0104] 作为另一示例,生成器306可以生成多个输入序列,每个输入序列包括来自k个文档的相应子集的上下文,然后使用语言模型神经网络114多次处理多个输入序列中的每一个,以对N个候选输出序列进行采样。

[0105] 作为另一示例,生成器306可以生成多个输入序列,每个输入序列包括来自k个文档中的相应一个文档的上下文,然后使用语言模型神经网络114处理多个输入序列中的每一个。

[0106] 在上述任一示例中,可以以轮询次序对多个输入序列进行采样,直到已经对N个候选输出序列进行采样。

[0107] 在一些实施方式中,N可以是k的倍数。在其他实施方式中,N可以是k不可分割的。

[0108] 然后,系统100对N个候选输出序列中的每一个执行奖励模型评分310。

[0109] 也就是说,系统100使用学习的奖励模型向N个候选输出序列中的每一个指派相应的评分。

[0110] 学习的奖励模型310是例如另一语言模型神经网络的模型,其接收由神经网络114生成的输入文本查询和响应以及引述作为输入,并且生成表示响应和引述的质量的评分作为输出。例如,评分可以表示用户相对于对由神经网络114生成的相同查询的其他响应(和伴随的引述)更喜欢该响应(和引述)的似然性。

[0111] 下面参考图5描述训练奖励模型。

[0112] 然后,系统100选择“最佳”样本312,即,根据学习的奖励模型从N个序列中选择具有最高评分的候选输出序列,作为最终输出序列。

[0113] 在一些实施方式中,如果没有候选者具有超过阈值的评分,则系统100代替地发出对用户查询的默认文本响应,例如,“我不知道”或“我不确定”。

[0114] 如上所述,每个候选输出序列包括(i)作为对输入文本查询的响应的第一输出文本子序列,(ii)作为第一上下文文本序列的相应自然语言标识符之一的第二输出文本子序列,以及(iii)作为来自第二输出文本子序列中的自然语言标识符标识的第一上下文文本序列的文本的第三输出文本子序列。

[0115] 特别地,(i)、(ii)和(iii)根据预定的输出语法布置在输出序列内。

[0116] 如图3的示例所示,输出语法是:

[0117] %<Claim>% (Document title) % [Quote from document] % 其中,“%<”、“>%”、“)”、“% [”和“] %”是模板词元,即,插入子序列之前和之后的预定语法词元,“Claim(声明)”是第一输出文本子序列的占位符,“Document title(文档标题)”是第二输出文本子序列的占位符,并且“Quote from document(来自文档的引述)”是第三输出文本子序列的占位符。

[0118] 然而,可以使用将“Claim”占位符、“Document title”占位符和“Quote from document”占位符放置在输出序列内的预定位置的各种语法中的任何语法。

[0119] 在一些实施方式中并且如上所述,系统使用约束采样对N个候选中的每一个进行

采样,以确保每个候选满足语法,即,包括来自自由序列中的自然语言标识符标识的上下文序列的精确引述。

[0120] 也就是说,如上所述,生成器306通过以下方式对给定候选输出序列进行采样:对于输出序列中的每个时间步,以第一输入文本序列和输出序列中的任何词元为条件,在第一输出序列中的时间步之前的任何时间步处,为词表中的每个词元生成相应评分,并且在每个时间步处,使用由神经网络为时间步生成的词表中的词元的相应评分来选择在时间步处的词元。

[0121] 当采用约束采样时,系统根据输出序列将采样约束为仅采样将是有效的接下来的词元的词元。

[0122] 例如,当生成第二输出文本子序列时,并且在第二输出文本子序列中的第一时间步之后的在第二输出文本子序列中的每个时间步处,生成器306可以接收由神经网络在该时间步处生成的相应评分,并且生成约束评分分布,该约束评分分布仅向在自然语言标识符中的一个(或多个)中紧接在第二输出文本子序列内已经生成的词元之后的词元指派非零评分,然后在该时间步处从约束评分分布而不是从接收到的评分分布对词元进行采样。也就是说,系统约束采样以仅向如果被追加到已经为第二输出文本子序列选择的词元则会产生对应输入序列中的一个或多个自然语言标识符的有效前缀的词元指派非零评分。

[0123] 作为另一示例,在一些情况下,第二输出文本子序列之前是第一输出文本序列中的一个或多个第一预定语法词元。例如,在图3的示例中,在输出语法中第二输出文本子序列之前是词元“>% (”。

[0124] 在这些情况下,使用约束采样生成输出序列包括在特定时间步处,确定已经在紧接在特定时间步之前的一个或多个时间步处选择了一个或多个第一预定语法词元,并且作为响应,确定特定时间步是第二输出文本子序列中的第一时间步。例如,系统可以确定词元“>% (”已经被采样,并且作为响应,确定下一时间步是第二子序列中的第一时间步。

[0125] 在该示例中,系统可以接收由神经网络在特定时间步处生成的相应评分,并且响应于确定特定时间步是第二输出文本子序列中的第一时间步,生成约束评分分布,该约束评分分布仅向作为对应输入序列中的自然语言标识符之一中的第一词元的词元指派非零评分,并且在该时间步处从约束评分分布对词元进行采样。也就是说,系统约束采样以仅向作为对应输入序列中的一个或多个自然语言标识符的第一词元的词元指派非零评分。

[0126] 作为另一示例,当使用约束采样时,在第三输出文本子序列中的第一时间步之后的第三输出文本子序列中的每个时间步处,系统可以接收由神经网络在该时间步处生成的相应评分,并生成约束评分分布,该约束评分分布仅向紧接在由第二输出文本子序列中的自然语言标识符标识的第一上下文文本序列中的第三输出文本子序列内已经生成的词元之后的词元指派非零评分。然后,系统在来自约束评分分布的时间步处对词元进行采样。也就是说,系统约束采样以仅向如果被追加到已经为第三输出文本子序列选择的词元则会产生与由第二输出文本子序列中的自然语言标识符标识的第一上下文文本序列中的子序列的直接匹配的词元指派非零评分。因此,系统确保第三输出文本子序列是来自由第二输出文本子序列中的自然语言标识符标识的上下文文档的直接引述。

[0127] 作为又一示例,在一些情况下,第三输出文本子序列之前是第一输出文本序列中的一个或多个第二预定语法词元。例如,在图3的示例中,在输出语法中第二输出文本子序

列之前是词元“%)%[”。

[0128] 在这些情况下,当使用约束采样时,在第二特定时间步处,系统确定已经在紧接在第二特定时间步之前的一个或多个时间步处选择了一个或多个第二预定语法词元,并且作为响应,确定特定时间步是第三输出文本子序列中的第一时间步。然后,在接收到由神经网络在特定时间步处生成的相应评分时,系统生成约束评分分布,该约束评分分布仅向在由第二输出文本子序列中的自然语言标识符标识的第一上下文文本序列中出现的词元指派非零评分,并在该时间步处从约束评分分布中对词元进行采样。

[0129] 然后,系统100提供来自最佳样本312的文本中的至少一些以呈现给用户。例如,系统100可以在用户界面中渲染314最佳样本312的呈现。

[0130] 如图3所示,呈现可以包括“声明”的文本,即第一子序列的文本、来自支持“声明”的上下文文档的引述,即第三子序列的文本,以及可选地,来自第二子序列的文档标识符。

[0131] 图4示出了向用户呈现输出序列的用户界面400的示例。

[0132] 在图4的示例中,用户已提交查询402“What kind of animal is Scooby Doo?(史酷比是哪种动物)?”

[0133] 作为响应,系统100已经生成了包括三个子序列的输出序列:(i)“A Great Dane dog(一只大丹犬)”,(ii)“Wikipedia Page: Scooby Doo(维基百科页面:史酷比)”,以及(iii)来自维基百科页面的具有标题“Scooby Doo”的引述。

[0134] 然后,响应于用户查询402,系统在用户界面400中呈现第一子序列404、第二子序列406和第三子序列408。

[0135] 另外,系统已经将第一子序列404显示为链接到第三子序列408的源的超链接,即,链接到Scooby Doo的维基百科页面,即,链接到标题为“Wikipedia Page: Scooby Doo”的网页。在用户界面400中包括超链接允许用户导航到由第二子序列指示的源,以例如验证引述的准确性或获得关于响应的附加上下文。

[0136] 图5示出了训练语言模型神经网络114的示例。

[0137] 如图5所示,系统获得502预训练的语言模型。

[0138] 例如,如上所述,语言模型可能已经在大的文本文档语料库上的语言建模目标上被训练。

[0139] 在获得502预训练的语言模型之后,系统生成样本504并经由人类评估对生成的样本进行评级。

[0140] 例如,为了获得每个评级,系统可以向评级者用户呈现问题和两个候选答案,例如,使用具有少样本提示的预训练语言模型生成的两个样本。每个候选答案可以被划分成“声明”部分和“支持证据”部分,例如,如上面参考图4所示。

[0141] 然后,系统可以从评级者用户获得输入,该输入指定任一声明是否是对问题的合理响应,声明是否由伴随的引述证据支持,以及评级者用户偏好哪个答案。对问题的合理响应是对问题的合理切题响应。所支持的响应是所提供的证据足以验证响应的有效性的响应。

[0142] 图6中示出了可用于从用户获得输入的用户界面的一个示例。

[0143] 也就是说,图6示出了用于对生成的样本进行评级的示例用户界面600,例如,其可以接收用于生成的样本的人类评估的输入。

[0144] 如图6所示,向用户呈现查询602和对查询602的两个候选响应604和606。每个候选响应604和606包括对查询的响应、来自响应的支持证据、以及支持证据的源的标识符。

[0145] 对于每个候选响应604和606,用户界面呈现对应的选择元素608和610,其允许用户提交指示对应的候选响应是否是合理答案(或指示用户不确定)的输入,并且提交指示对应的支持证据是否支持对应的候选响应(或指示用户不确定)的输入。

[0146] 选择元素608和610还各自允许用户提交指示对应的候选响应604或606是(在两个候选响应中)对查询602的优选响应的输入。

[0147] 用户界面600还可以允许用户提交指示两个响应“平局(tied)”的输入或提交关于样本的评论。

[0148] 回到图5的描述,然后,系统使用评级样本来执行监督微调(SFT) 506,其中系统通过监督学习在评级样本上训练语言模型。

[0149] 也就是说,对于用于SFT的每个样本,给定样本中的问题和包括具有支持证据的文本的上下文序列的上下文序列的集合,系统训练语言模型以产生声明和样本中的支持证据。

[0150] 可选地,当执行SFT时,系统可以仅使用被评级为合理的并且被支持用于监督微调的样本。

[0151] 作为特定示例,系统可以如下在SFT期间生成给定样本的输入序列。

[0152] 对于一定比例的样本,例如,对于数据的1/3或1/2,系统仅使用上下文中的单个文档,从其提取支持证据的相同文档,强制支持证据存在于上下文序列内。

[0153] 对于采样器的剩余部分,系统在上下文中使用n个文档,例如,其中,n是在1和例如5、10或15的固定数之间随机抽取的。类似地,系统强制目标文档和支持证据引述存在于上下文序列中。对于上下文序列中的文档的其余部分,系统可以使用例如由搜索引擎提供的问题的前n-1个搜索结果。

[0154] 系统可以截断上下文文档中的每一个,使得输入序列的总词元长度不超过基于语言模型的上下文窗口的固定数字。该词元长度允许可以在提示中包括的文档之间随机划分,使得语言模型看到来自同一输入序列内的不同上下文文档的不同大小的上下文序列。当将给定上下文文档截断到其最大允许长度时,系统可以确保每个文档包含摘录,如上所述。

[0155] 可选地,在执行监督微调(SFT) 506之后,系统可以使用SFT模型来生成再次经由人类评估评级的附加样本。

[0156] 然后,系统在生成的样本——例如原始生成的样本或原始生成的样本和使用SFT模型生成的附加样本——上训练奖励模型(RM) 508。

[0157] 如上所述,学习的奖励模型是模型,例如另一种语言模型神经网络,其接收输入文本查询和由神经网络114生成的响应和引述作为输入,并生成表示响应和引述的质量的评分作为输出。

[0158] 例如,给定查询和响应字符串,系统可以将奖励模型训练为分类器,该分类器预测指示给定对中的哪个示例是优选的二进制变量。也就是说,给定由奖励模型针对该对中的两个示例生成的评分,系统可以计算该对中的第一示例是优选的概率。例如,系统可以使用交叉熵目标来训练奖励模型,该交叉熵目标使用用户偏好作为真实值并且使用计算的概率

作为预测。

[0159] 可选地,在训练期间,奖励模型还将对中的响应的二进制支持和合理判断预测为辅助损失。因此,在这些情况下,最终损失是例如成对偏好预测损失和辅助预测损失的平均值或加权平均值的组合。

[0160] 在一些实施方式中,系统可以用制造(“合成”)比较的集合来增强RM训练集。例如,系统可以从事实检查数据集的支持和反驳声明生成伪造比较。这样的数据集的一个示例是FEVER数据集(Thorne等人,2018)。包括这些伪造的比较可以提供非提取的附加的问题回答的分布外模式,并且可以使奖励模型更好地验证证据的支持性。例如FEVER数据集的这样的数据集的示例可以包含通过更改从源文本中提取的句子而生成的声明。然后将这些声明分类为支持的(Supported)、反驳的(Refuted)或不足够的(NotEnough),并且用相关联的证据标记。为了将这样的声明转换成具有答案比较的问题的示例,系统可以使用各种技术中的任何一种。现在将描述技术类型的一些示例。

[0161] 类型A:系统可以通过直接模板操作从声明生成问题(例如,‘{claim}?’ , ‘Is it true that {claim}?’ , ‘Is it correct to say that {claim}?’ , ‘{claim}. Do you agree?’ (‘{声明}?’ , ‘说那个{声明}是正确的吗?’ , ‘{声明}.你同意吗?’)。示例比较肯定答案,如‘Yes(是)’ , ‘This is correct(这是正确的)’ , ‘It is true(是)’ 与支持引述组合,并且否定答案与相同引述组合。如果支持原始声明,则肯定答案被标记为优选的、支持的和合理的。否则,负一被标记为优选支持和合理。

[0162] 类型B:系统可以使用少样本提示的、预训练的语言模型神经网络将声明转换为问题。例如,声明Roman Atwood is a content creator.(罗曼·阿特伍德是内容创建者。)可以转换为Who is Roman Atwood?(谁是罗曼·阿特伍德?)。作为已经被转换为问题的声明的比较,系统可以使用一个答案作为来自数据集(具有支持引述)的对应声明,并且使用经由模板化产生的声明的直接否定(例如,‘It is not true that {claim}({声明}不为真)’)作为另一答案。如果支持原始声明,则包含声明的答案被标记为优选的、支持的和合理的。否则,否定声明被标记为优选。作为另一示例,如果支持原始声明,则系统可以使用原始声明作为一个答案并且使用随机生成的声明作为比较,其中原始声明被标记为优选的、支持的和合理的。

[0163] 如上所述,系统然后可以在采样时使用奖励模型来向候选输出序列指派评分。

[0164] 在训练RM 508之后,系统可以使用经训练的奖励模型来通过强化学习510进一步微调SFT模型。也就是说,系统使用奖励模型通过训练模型以最大化由经训练的RM 508预测的预期奖励来执行来自人类偏好的强化学习(RLHP)技术。

[0165] 可选地,系统然后可以使用进一步微调的模型来生成用于人类评估的附加样本,并且通过SFT或RL或两者来重新微调模型,以重新训练RM或两者。也就是说,系统可以执行所描述的训练循环的多于一次迭代以进一步微调语言模型、进一步微调奖励模型或两者。

[0166] 另外,虽然图5的示例描述了系统使用SFT和RL两者来微调语言模型,但是在一些情况下,系统仅使用SFT或RL而不是两者。例如,当使用奖励模型进行重新排名时,使用仅通过SFT或RL(而不是两者)微调的模型可以提高性能,使得奖励模型被提供用于更多样的样本以进行重新排名。

[0167] 现在如下是可以由语言模型神经网络采用的自注意力的描述。

[0168] 如上所述,自注意力块是包括注意力机制的神经网络层,该注意力机制对自注意力块输入(或从层输入导出的输入)进行操作以生成自注意力块输出。自注意力机制可以被因果地掩蔽,使得输入序列中的任何给定位置不关注输入序列中给定位置之后的任何位置(例如,使用来自该位置的数据)。存在许多不同的可能注意力机制。包括注意力机制的自注意力层的一些示例在以下中描述:Vaswani et al. “Attention is all you need(注意力就是你所需要的)”, 31st Conference on Neural Information Processing Systems(神经信息处理系统会议)(NIPS 2017), Long Beach, CA, USA; Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer(利用统一的文本到文本transformer探索迁移学习的极限). arXiv preprint arXiv:1910.10683, 2019; Daniel Adiwardana, Minh-Thang Luong, David R. So, Jamie Hall, Noah Fiedel, Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, and Quoc V. Le. Towards a human-like open-domain chatbot(迈向类似人类的开放域聊天机器人). CoRR, abs/2001.09977, 2020; and Tom B Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners(语言模型是少样本学习者). arXiv preprint arXiv:2005.14165, 2020。

[0169] 通常,注意力机制将查询和键值对集合映射到输出,其中查询、键和值都是向量。输出被计算为值的加权和,其中指派给每个值的权重由查询与对应键的兼容性函数——例如点积或缩放点积——计算。

[0170] 通常,自注意力机制被配置为关联同一序列中的不同位置,以确定序列的变换版本作为输出。例如,注意力层输入可以包括输入序列的每个元素的向量。这些向量向自注意力机制提供输入,并由自注意力机制用于确定注意力层输出的相同序列的新表示,其类似地包括输入序列的每个元素的向量。自注意力机制的输出可以用作注意力层输出,或者它可以由前馈层、跳过连接或归一化操作中的一个或多个来处理以提供注意力层输出。

[0171] 在一些实施方式中,注意力机制被配置为将例如由矩阵 $\mathbf{W}^Q$ 定义的查询变换、例如由矩阵 $\mathbf{W}^K$ 定义的键变换和例如由矩阵 $\mathbf{W}^V$ 定义的值变换中的每一个应用于作为注意力层的输入数据 $\mathbf{X}$ 的注意力层输入,以导出包括针对输入序列中的每个向量的相应查询的查询矩阵 $\mathbf{Q} = \mathbf{XW}^Q$ 、包括针对输入序列中的每个向量的相应键的键矩阵 $\mathbf{K} = \mathbf{XW}^K$ 和包括针对输入序列中的每个向量的相应值的值矩阵 $\mathbf{V} = \mathbf{XW}^V$,其用于确定输出的注意力序列。例如,注意力机制可以通过将每个查询向量应用于每个键向量以确定每个值向量的相应权重,然后使用相应权重来组合值向量以确定输入序列的每个元素的自注意力层输出而应用的点积注意力机制。自注意力层输出可以通过缩放因子——例如通过查询和键的维度的平方根——来缩放,以实现缩放的点积注意力。因此,例如,注意力机制的输出可以被确定为 $\text{softmax}\left(\frac{\mathbf{QK}^T}{\sqrt{d}}\right)\mathbf{V}$,其中 d 是键(和值)向量的维度。在另一实施方式中,注意力机制包括使用具有隐藏层的前馈网络来计算兼容性函数的“加性注意力”机制。注意力机制的输出可以由一个或多个完全连接的前馈神经网络层进一步处理。

[0172] 注意力机制可以实现多头注意力,即,它可以并行应用多个不同的注意力机制。然后,如果必要,这些的输出可以与应用学习的线性变换来被组合,例如被连结,以减少到原始维度。

[0173] 本说明书结合系统和计算机程序组件使用术语“配置”。对于要被配置为执行特定操作或动作的一个或多个计算机的系统,意味着系统已经在其上安装了软件、固件、硬件或它们的组合,其在操作中使系统执行操作或动作。对于要被配置为执行特定操作或动作的一个或多个计算机程序,意味着一个或多个程序包括当由数据处理装置执行时使装置执行操作或动作的指令。

[0174] 本说明书中描述的主题和功能操作的实施例可以在数字电子电路、有形体现的计算机软件或固件、计算机硬件,包括本说明书中公开的结构及其结构等同物,或它们中的一个或多个的组合中实现。本说明书中描述的主题的实施例可以被实现为一个或多个计算机程序,例如,在有形非暂时性存储介质上编码的计算机程序指令的一个或多个模块,用于由数据处理装置执行或控制数据处理装置的操作。计算机存储介质可以是机器可读存储设备、机器可读存储基板、随机或串行存取存储器设备、或它们中的一个或多个的组合。可替代地或附加地,程序指令可以被编码在人工生成的传播信号上,例如在机器生成的电、光或电磁信号上,该传播信号被生成以对信息进行编码以传输到合适的接收器装置以供数据处理装置执行。

[0175] 术语“数据处理装置”是指数据处理硬件,并且涵盖用于处理数据的所有种类的装置、设备和机器,包括例如可编程处理器、计算机或多个处理器或计算机。该装置还可以是或进一步包括专用逻辑电路,例如FPGA(现场可编程门阵列)或ASIC(专用集成电路)。除了硬件之外,该装置还可以可选地包括为计算机程序创建执行环境的代码,例如,构成处理器固件、协议栈、数据库管理系统、操作系统或它们中的一个或多个的组的代码。

[0176] 也可以被称为或描述为程序、软件、软件应用、app、模块、软件模块、脚本或代码的计算机程序可以以任何形式的编程语言编写,包括编译或解释语言,或者声明性或过程性语言;并且它可以以任何形式部署,包括作为独立程序或作为模块、组件、子例程或适合在计算环境中使用的其他单元。程序可以但不必对应于文件系统中的文件。程序可以存储在保存其他程序或数据的文件的一部分中,例如存储在标记语言文档中的一个或多个脚本,存储在专用于所讨论的程序的单个文件中,或者存储在多个协调文件中,例如在存储一个或多个模块、子程序或代码部分的文件中。计算机程序可以被部署为在位于一个站点或跨多个站点分布并通过数据通信网络互连的一个计算机或多个计算机上执行。

[0177] 在本说明书中,术语“数据库”广泛地用于指代任何数据合集:数据不需要以任何特定方式结构化,或者根本不需要结构化,并且它可以存储在一个或多个位置中的存储设备上。因此,例如,索引数据库可以包括多个数据合集,每个数据合集可以被不同地组织和访问。

[0178] 类似地,在本说明书中,术语“引擎”被广泛地用于指代被编程为执行一个或多个特定功能的基于软件的系统、子系统或过程。通常,引擎将被实现为安装在一个或多个位置中的一个或多个计算机上的一个或多个软件模块或组件。在一些情况下,一个或多个计算机将专用于特定引擎;在其他情况下,可以在同一计算机或多个计算机上安装和运行多个引擎。

[0179] 本说明书中描述的过程和逻辑流程可以由执行一个或多个计算机程序的一个或多个可编程计算机执行,以通过对输入数据进行操作并生成输出来执行功能。过程和逻辑流程还可以由例如FPGA或ASIC的专用逻辑电路或由专用逻辑电路和一个或多个编程计算机的组合来执行。

[0180] 适合于执行计算机程序的计算机可以基于通用或专用微处理器或两者,或任何其他种类的中央处理单元。通常,中央处理单元将从只读存储器或随机存取存储器或两者接收指令和数据。计算机的基本元件是用于执行或实行指令的中央处理单元和用于存储指令和数据的一个或多个存储器设备。中央处理单元和存储器可以由专用逻辑电路补充或并入专用逻辑电路中。通常,计算机还将包括用于存储数据的一个或多个大容量存储设备,例如磁盘、磁光盘或光盘,或者可操作地耦合以从用于存储数据的一个或多个大容量存储设备接收数据或将数据传送到用于存储数据的一个或多个大容量存储设备或两者。然而,计算机不需要具有这样的设备。此外,计算机可以嵌入在另一设备中,例如移动电话、个人数字助理(PDA)、移动音频或视频播放器、游戏控制台、全球定位系统(GPS)接收器或便携式存储设备,例如通用串行总线(USB)闪存驱动器,仅举几例。

[0181] 适合于存储计算机程序指令和数据的计算机可读介质包括所有形式的非易失性存储器、介质和存储器设备,包括例如半导体存储器设备,例如EPROM、EEPROM和闪存设备;磁盘,例如内部硬盘或可移动盘;磁光盘;以及CD ROM和DVD-ROM盘。

[0182] 为了提供与用户的交互,本说明书中描述的主题的实施例可以在具有用于向用户显示信息的显示设备——例如CRT(阴极射线管)或LCD(液晶显示器)监视器——以及用户可以通过其向计算机提供输入的键盘和定点设备——例如鼠标或轨迹球——的计算机上实现。也可以使用其他种类的设备来提供与用户的交互;例如,提供给用户的反馈可以是任何形式的感觉反馈,例如视觉反馈、听觉反馈或触觉反馈;并且可以以任何形式接收来自用户的输入,包括声学、语音或触觉输入。此外,计算机可以通过向由用户使用的设备发送文档和从由用户使用的设备接收文档来与用户交互;例如,通过响应于从web浏览器接收的请求向用户设备上的web浏览器发送网页。此外,计算机可以通过向个人设备——例如正在运行消息传送应用的智能电话——发送文本消息或其他形式的消息并作为回报从用户接收响应消息来与用户交互。

[0183] 用于实现机器学习模型的数据处理装置还可以包括例如专用硬件加速器单元,用于处理机器学习训练或生产的常见和计算密集型部分,例如推理、工作负载。

[0184] 可以使用例如TensorFlow框架或Jax框架的机器学习框架来实现和部署机器学习模型。

[0185] 本说明书中描述的主题的实施例可以在计算系统中实现,该计算系统包括后端组件,例如作为数据服务器,或者包括中间件组件,例如应用服务器,或者包括前端组件,例如具有用户可以通过其与本说明书中描述的主题的实施方式进行交互的图形用户界面、web浏览器或应用的客户端计算机,或者一个或多个这样的后端、中间件或前端组件的任何组合。系统的组件可以通过例如通信网络的任何形式或介质的数字数据通信互连。通信网络的示例包括局域网(LAN)和广域网(WAN),例如因特网。

[0186] 计算系统可以包括客户端和服务端。客户端和服务端通常彼此远离,并且典型地通过通信网络进行交互。客户端和服务端的关系是借助在相应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而产生的。在一些实施例中,服务器将例如HTML页面的数据传输到用户设备,例如,用于向与充当客户端的设备交互的用户显示数据和从用户接收用户输入。可以在服务器处从设备接收在用户设备处生成的数据,例如用户交互的结果。

[0187] 虽然本说明书包含许多特定的实施方式细节,但是这些不应被解释为对任何发明的范围或可能要求保护的范围的限制,而是作为可能特定于特定发明的特定实施例的特征的描述。在本说明书中在分开的实施例的上下文中描述的某些特征也可以在单个实施例中组合实现。相反,在单个实施例的上下文中描述的各种特征也可以单独地或以任何合适的子组合在多个实施例中实现。此外,尽管特征可以在上面被描述为以某些组合起作用并且甚至最初被如此要求保护,但是在一些情况下,来自所要求保护的组合的一个或多个特征可以从组合中删除,并且所要求保护的组合可以针对子组合或子组合的变型。

[0188] 类似地,虽然在附图中描绘了操作并且在权利要求中以特定次序叙述了操作,但这不应被理解为要求这样的操作以所示的特定次序或以顺序次序执行,或者要求执行所有示出的操作以实现期望的结果。在某些情况下,多任务和并行处理可能是有利的。此外,上述实施例中的各种系统模块和组件的分离不应被理解为在所有实施例中都需要这样的分离,并且应当理解,所描述的程序组件和系统通常可以一起集成在单个软件产品中或封装到多个软件产品中。

[0189] 已经描述了主题的特定实施例。其他实施例在所附权利要求的范围内。例如,权利要求中记载的动作可以以不同的次序执行,并且仍然实现期望的结果。作为一个示例,附图中描绘的过程不一定需要所示的特定次序或顺序次序来实现期望的结果。在一些情况下,多任务和并行处理可能是有利的。

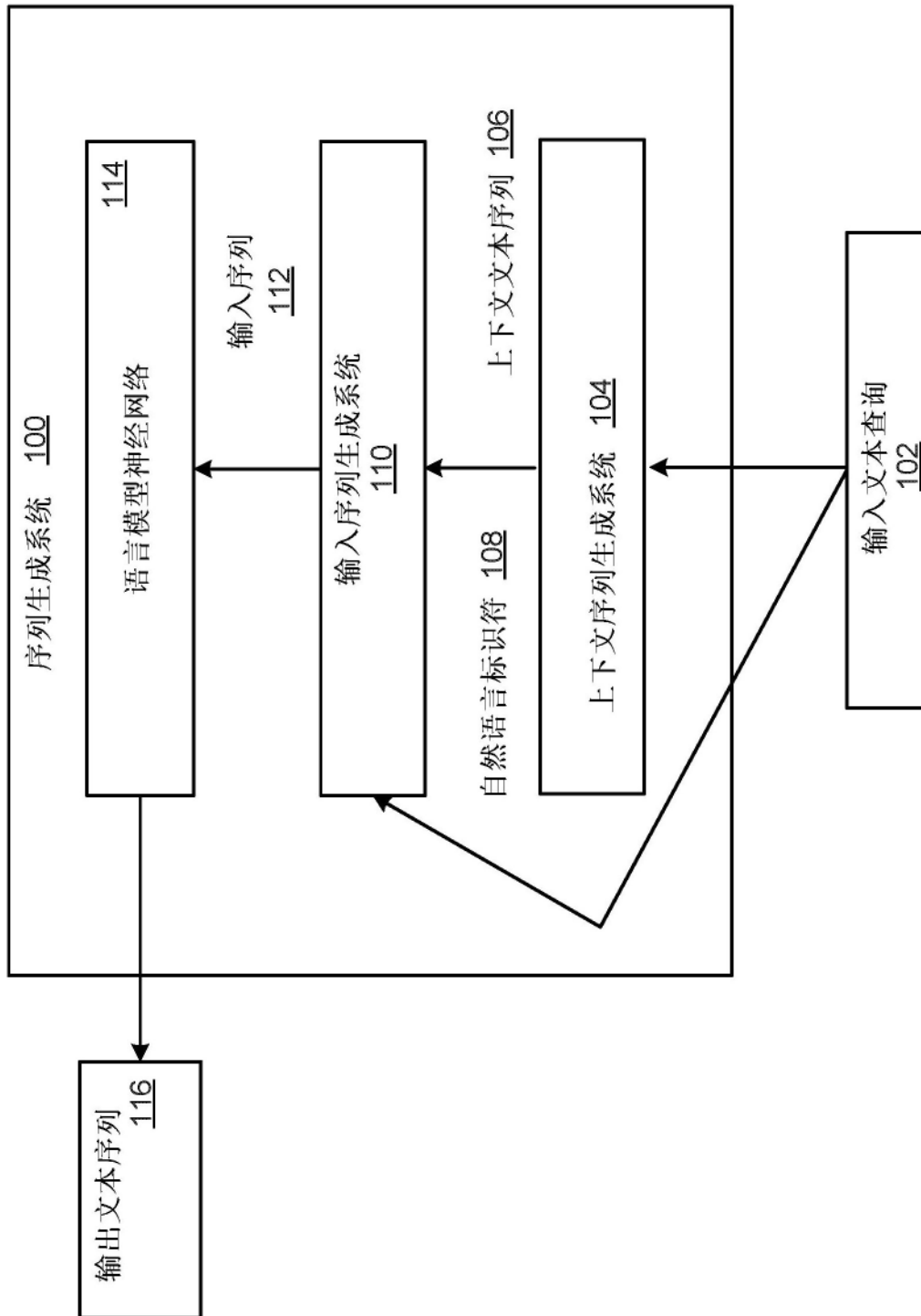


图1

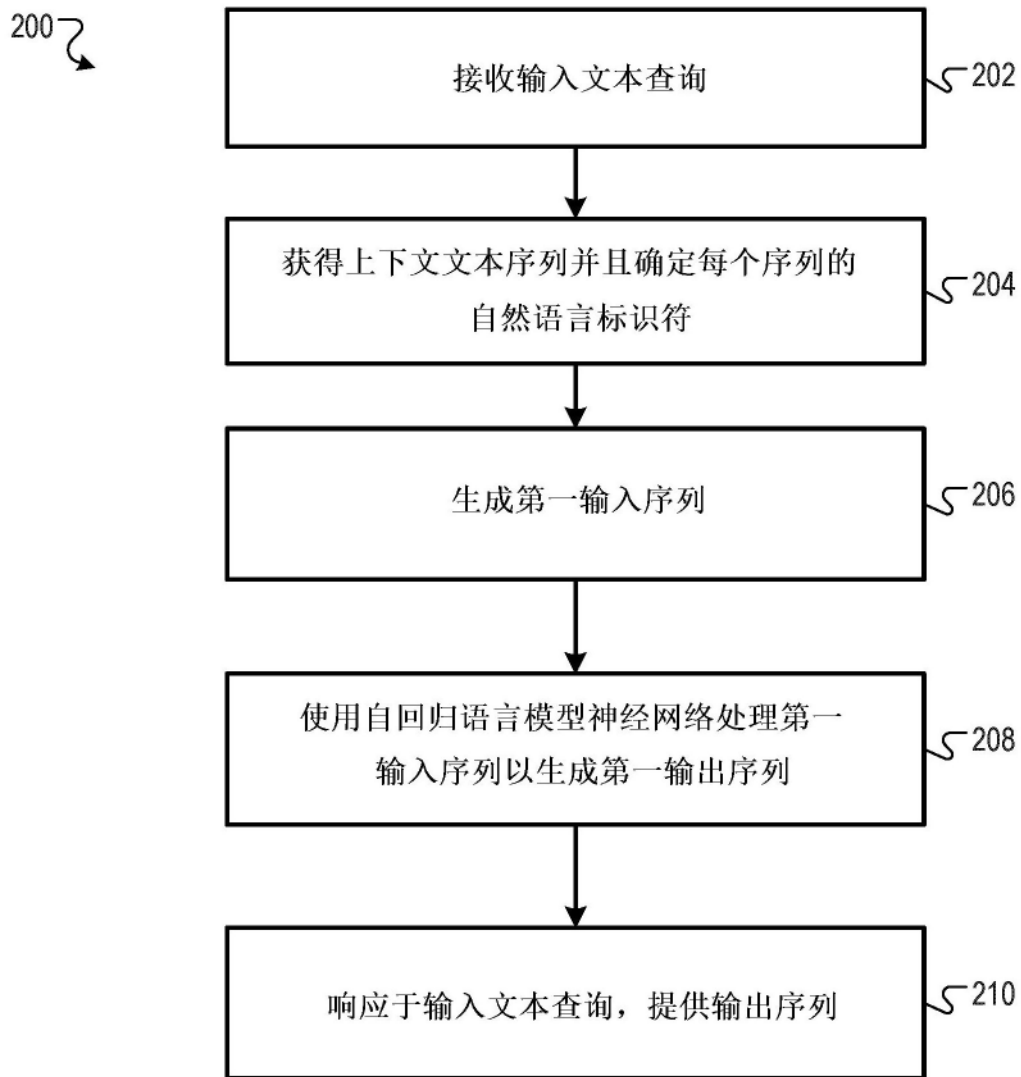


图2

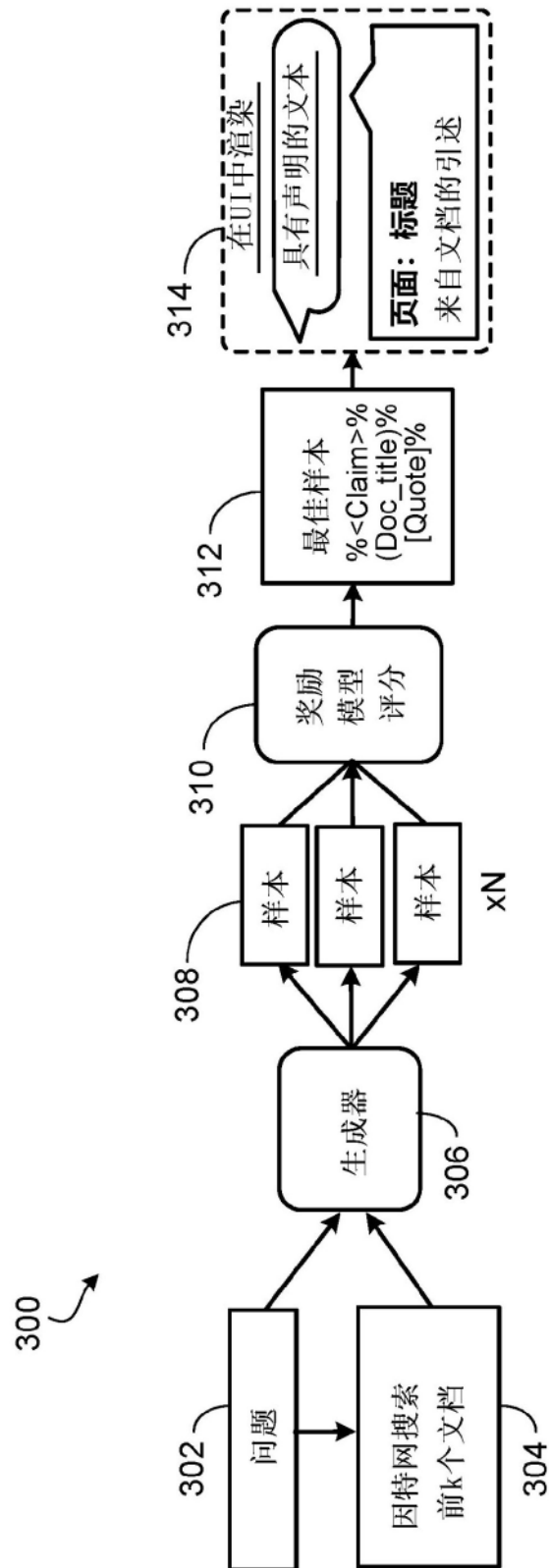


图3

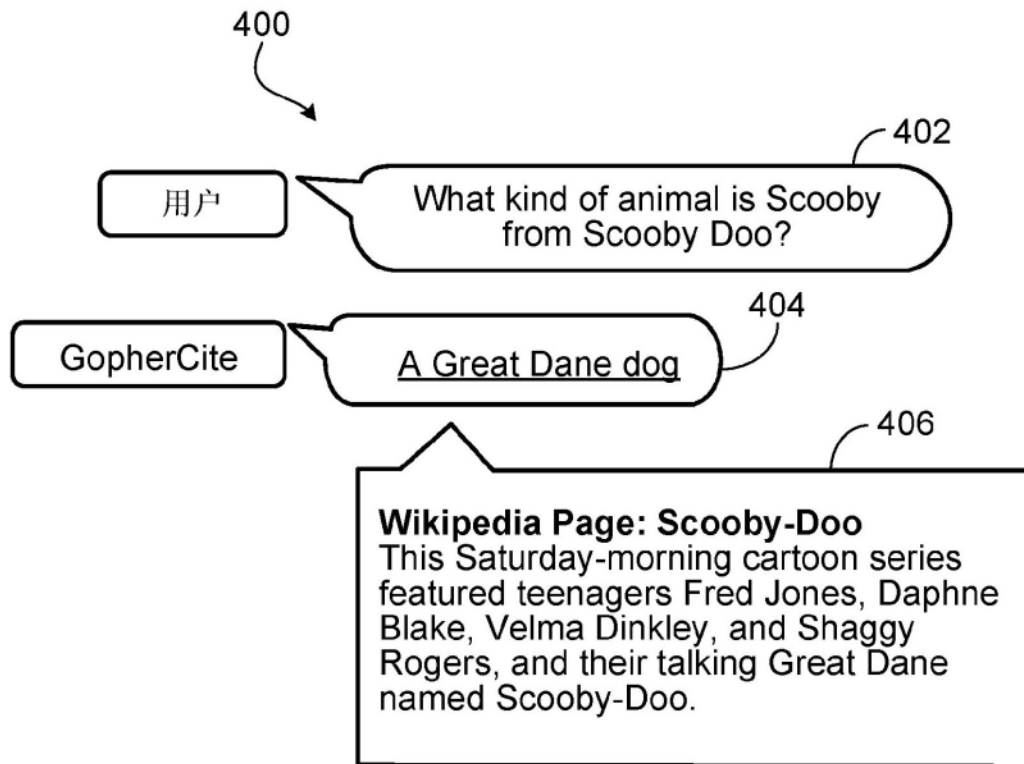


图4

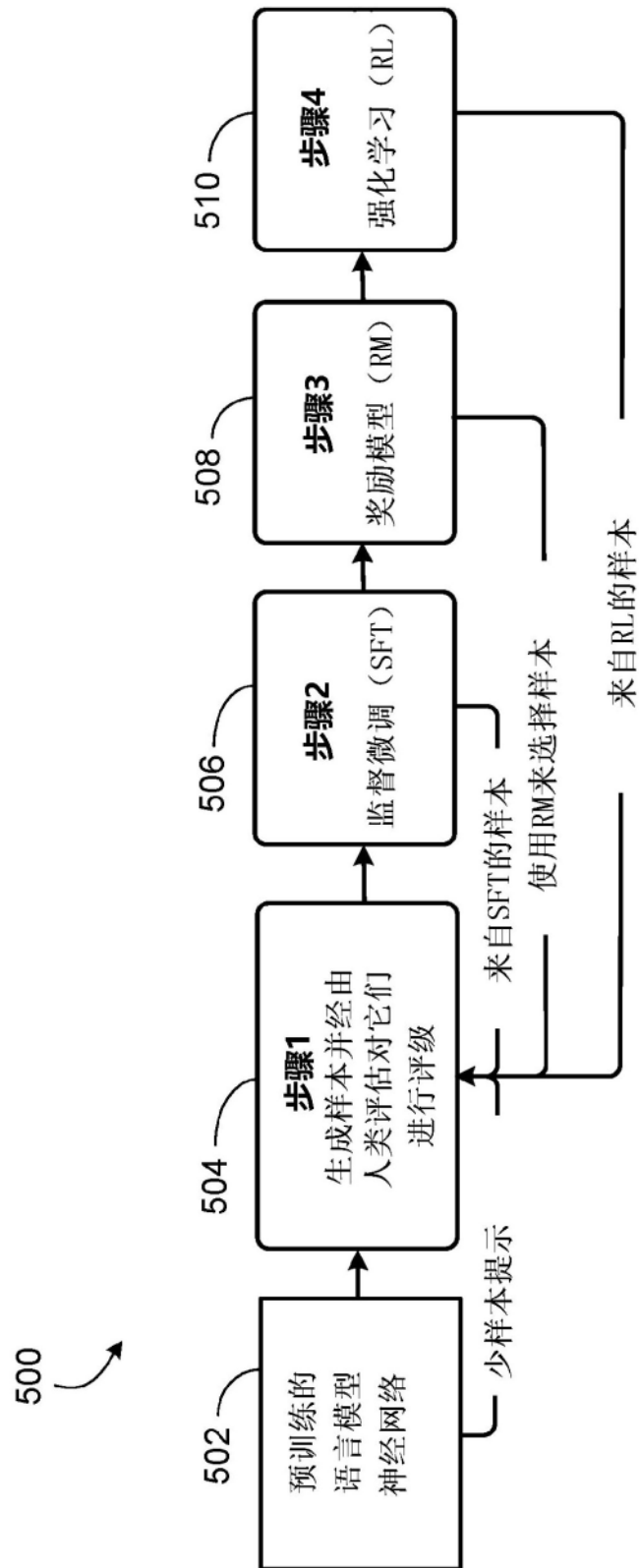


图5

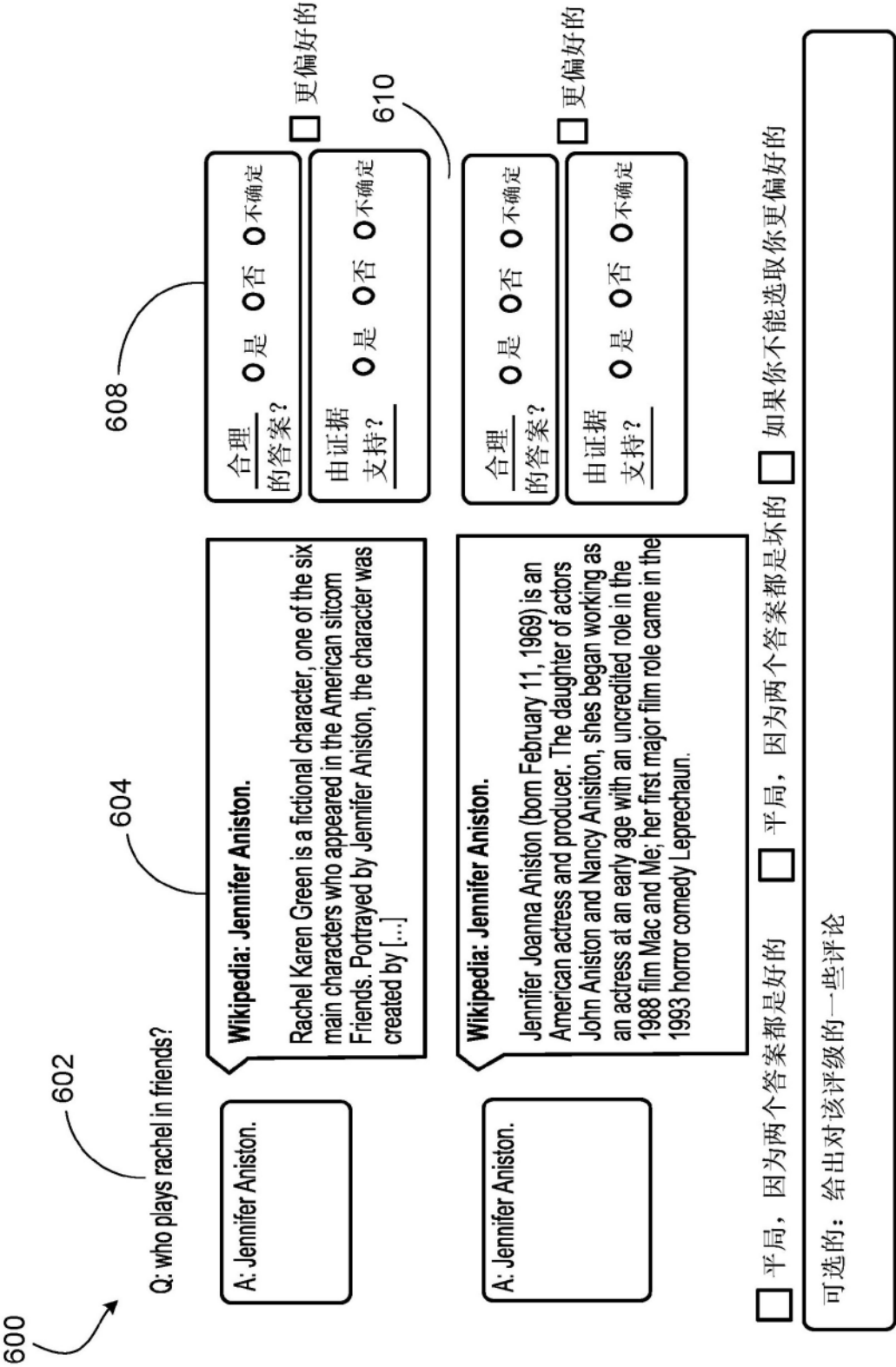


图6