



(12) 发明专利申请

(10) 申请公布号 CN 113361284 A

(43) 申请公布日 2021.09.07

(21) 申请号 202110746931.6

(22) 申请日 2021.06.30

(71) 申请人 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 尚骏远 王硕寰 丁思宇

(74) 专利代理机构 中科专利商标代理有限责任
公司 11021
代理人 庄锦军

(51) Int. Cl.

G06F 40/30 (2020.01)

G06F 40/242 (2020.01)

G06F 16/903 (2019.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

权利要求书2页 说明书8页 附图4页

(54) 发明名称

目标内容的生成方法和装置

(57) 摘要

本公开提供了一种目标内容的生成方法,具体涉及计算机技术领域,尤其涉及人工智能技术领域。具体实现方案包括:利用内容生成模型处理素材内容,得到多个目标内容分段;以及根据多个目标内容分段,生成针对素材内容的目标内容,其中,利用内容生成模型处理素材内容得到多个目标内容分段包括多个处理阶段,多个处理阶段中的每个处理阶段包括:利用内容生成模型处理素材内容,得到候选内容分段和候选内容分段的原始概率分布;根据预定下游数据集和前面的处理阶段输出的目标内容分段,调整原始概率分布,得到调整概率分布;以及根据调整概率分布,从候选内容分段中确定该处理阶段输出的对应目标内容分段。

100



1. 一种目标内容的生成方法,包括:
利用内容生成模型处理素材内容,得到多个目标内容分段;以及
根据多个目标内容分段,生成针对所述素材内容的目标内容,
其中,所述利用内容生成模型处理素材内容,得到多个目标内容分段包括多个处理阶段,所述多个处理阶段中的每个处理阶段包括:
利用内容生成模型处理所述素材内容,得到候选内容分段和候选内容分段的原始概率分布;
根据预定下游数据集和前面的处理阶段输出的目标内容分段,调整所述原始概率分布,得到调整概率分布;以及
根据所述调整概率分布,从所述候选内容分段中确定该处理阶段输出的对应目标内容分段。
2. 根据权利要求1所述的方法,其中,所述根据所述调整概率分布,从所述候选内容分段中确定该处理阶段输出的对应目标内容分段包括:
通过束搜索,从所述候选内容分段中确定对应目标内容分段,所确定的对应目标内容分段的数量等于束宽度。
3. 根据权利要求1所述的方法,其中,所述根据预定下游数据集和前面的处理阶段输出的目标内容分段,调整所述原始概率分布,得到调整概率分布包括:
根据预定下游数据集和前面的处理阶段输出的目标内容分段,确定受限概率分布;以及
利用受限概率分布调整所述候选内容分段的原始概率分布。
4. 根据权利要求3所述的方法,其中,所述根据预定下游数据集和前面的处理阶段输出的目标内容分段,确定受限概率分布包括:
利用所述预定下游数据集构建字典树;以及
根据字典树确定所述受限概率分布。
5. 根据权利要求4所述的方法,其中,所述根据字典树确定所述受限概率分布包括:
通过在字典树中搜索前面的处理阶段输出的目标内容分段,确定所述受限概率分布。
6. 根据权利要求1至5中任一项所述的方法,其中,
所述预定下游数据集包括所述素材内容。
7. 根据权利要求1至5中任一项所述的方法,其中,
所述素材内容包括段落文本和问题,并且所述目标内容包括从所述段落文本得出的针对所述问题的答案。
8. 一种目标内容的生成装置,包括:
处理模块,用于利用内容生成模型处理素材内容,得到多个目标内容分段;以及
生成模块,用于根据多个目标内容分段,生成针对所述素材内容的目标内容,
其中,所述处理模块包括多个处理子模块,每个处理子模块用于:
利用内容生成模型处理所述素材内容,得到候选内容分段和候选内容分段的原始概率分布;
根据预定下游数据集和前面的处理子模块输出的目标内容分段,调整所述原始概率分布,得到调整概率分布;以及

根据所述调整概率分布,从所述候选内容分段中确定该处理子模块输出的对应目标内容分段。

9. 根据权利要求8所述的装置,其中,每个处理子模块包括:

第一确定单元,用于通过束搜索,从所述候选内容分段中确定对应目标内容分段,所确定的对应目标内容分段的数量等于束宽度。

10. 根据权利要求8所述的装置,其中,每个处理子模块包括:

第二确定单元,用于根据预定下游数据集和前面的处理子模块输出的目标内容分段,确定受限概率分布;以及

调整单元,用于利用受限概率分布调整所述候选内容分段的原始概率分布。

11. 根据权利要求10所述的装置,其中,所述第二确定单元包括:

构建子单元,用于利用所述预定下游数据集构建字典树;以及

确定子单元,用于根据字典树确定所述受限概率分布。

12. 根据权利要求11所述的装置,其中,所述确定子单元用于:

通过在字典树中搜索前面的处理子模块输出的目标内容分段,确定所述受限概率分布。

13. 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1-7中任一项所述的方法。

14. 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1-7中任一项所述的方法。

15. 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1-7中任一项所述的方法。

目标内容的生成方法和装置

技术领域

[0001] 本公开涉及计算机技术领域,尤其涉及人工智能技术领域,具体涉及一种目标内容的生成方法和装置。

背景技术

[0002] 生成模型通常具有过大的输出空间,而对于诸如抽取式阅读理解或特定风格内容生成之类的特定任务,其输出空间是有限的。因此,需要有效地限制生成模型的输出空间大小,以生成更加符合特定任务的结果。

发明内容

[0003] 本公开提供了一种目标内容的生成方法和装置。

[0004] 根据本公开的一方面,提供了一种目标内容的生成方法,包括:

[0005] 利用内容生成模型处理素材内容,得到多个目标内容分段;以及

[0006] 根据多个目标内容分段,生成针对所述素材内容的目标内容,

[0007] 其中,所述利用内容生成模型处理素材内容,得到多个目标内容分段包括多个处理阶段,所述多个处理阶段中的每个处理阶段包括:

[0008] 利用内容生成模型处理所述素材内容,得到候选内容分段和候选内容分段的原始概率分布;

[0009] 根据预定下游数据集和前面的处理阶段输出的目标内容分段,调整所述原始概率分布,得到调整概率分布;以及

[0010] 根据所述调整概率分布,从所述候选内容分段中确定该处理阶段输出的对应目标内容分段。

[0011] 根据本公开的另一方面,提供了一种目标内容的生成装置,包括:

[0012] 处理模块,用于利用内容生成模型处理素材内容,得到多个目标内容分段;以及

[0013] 生成模块,用于根据多个目标内容分段,生成针对所述素材内容的目标内容,

[0014] 其中,所述处理模块包括多个处理子模块,每个处理子模块用于:

[0015] 利用内容生成模型处理所述素材内容,得到候选内容分段和候选内容分段的原始概率分布;

[0016] 根据预定下游数据集和前面的处理子模块输出的目标内容分段,调整所述原始概率分布,得到调整概率分布;以及

[0017] 根据所述调整概率分布,从所述候选内容分段中确定该处理子模块输出的对应目标内容分段。

[0018] 根据本公开的另一方面,提供了一种电子设备,包括:

[0019] 至少一个处理器;以及

[0020] 与所述至少一个处理器通信连接的存储器;其中,

[0021] 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根据本公开实施例的方法。

[0022] 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根据本公开实施例的方法。

[0023] 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根据本公开实施例的方法。

[0024] 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。本公开的其他特征将通过以下的说明书而变得容易理解。

附图说明

[0025] 附图用于更好地理解本方案,不构成对本公开的限定。其中:

[0026] 图1是根据本公开实施例的目标内容的生成方法的流程图;

[0027] 图2是示出了根据本公开实施例的目标内容的生成方法的目标内容生成的一个示例的图;

[0028] 图3是示出了根据本公开实施例的用于生成受限概率分布的字典树的一个示例的图;

[0029] 图4是根据本公开实施例的目标内容的生成装置的示意图;以及

[0030] 图5示出了可以用来实施本公开实施例的示例电子设备的示意性框图。

具体实施方式

[0031] 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。

[0032] 可以采用以下几种方法来有效地限制生成模型的输出空间大小,以生成更加符合特定任务的结果。

[0033] 一种方法是在对生成模型进行预训练时,在样本数据前加入风格限定词,使得生成模型学到使风格规范化的格式。这种方法存在的问题是,需要在预训练时构建不同的风格限定词和样本,构建成本较高,并且在完成预训练后模型仅具有固定的、有限的风格。

[0034] 另一种方法是利用小样本学习(few-shot learning)或语境学习(in-context learning)。在原始输入前添加若干个规范化的小样本,通过输入风格规范化的多个样例数据,从而在一定程度上引导模型进行特定风格文本的输出。这种方法存在的问题是可控性较差。

[0035] 另一种方法是针对不同风格的数据进行微调。在微调阶段,使用特定风格的语料库对模型进行训练,以使模型更加适合于特定领域。这种方法存在的问题是需要针对所需的每种不同风格进行微调,成本较高。

[0036] 本公开实现了一种目标内容的生成方法和装置,利用内容生成模型处理素材内容,得到多个目标内容分段,并根据多个目标内容分段,生成针对所述素材内容的目标内容。所述处理包括多个处理阶段。所述多个处理阶段中的每个处理阶段利用内容生成模型处理所述素材内容,得到候选内容分段和候选内容分段的原始概率分布。根据预定下游数

据集和前面的处理阶段输出的目标内容分段,调整所述原始概率分布,得到调整概率分布,并且根据所述调整概率分布,从所述候选内容分段中确定该处理阶段输出的对应目标内容分段。通过这种方式,可以利用预定下游数据集来有效地限制生成模型的输出空间大小,以生成更加符合特定任务的结果。这种方式无需像现有技术那样构建不同的风格限定词和样本,也无需进行微调,成本较低,并且可控性良好。

[0037] 图1是根据本公开实施例的目标内容的生成方法100的流程图。下面参考图1对根据本公开实施例的目标内容的生成方法100进行说明。

[0038] 在步骤S110,利用内容生成模型处理素材内容,得到多个目标内容分段。

[0039] 内容生成模型是对素材内容进行处理,从而生成针对素材内容的内容的生成式模型。内容生成模型可以是用于生成数据的任意模型,例如,朴素贝叶斯模型、混合高斯模型、马尔可夫随机场模型、生成对抗网络(GAN)、变分自编码器(VAE)、生成式的长短期记忆网络(LSTM)、BERT模型、ERNIE模型等。

[0040] 每个目标内容分段可以是目标内容的一个组成部分。例如,当目标内容是一段文本内容时,每个目标内容分段可以是该段文本内容中的一个或多个字符。

[0041] 在步骤S120,根据多个目标内容分段,生成针对所述素材内容的目标内容。

[0042] 可以根据在步骤S110中得到的多个目标内容分段,生成针对所述素材内容的完整的目标内容。例如,可以将多个目标内容分段以一定顺序组合在一起,从而生成完整的目标内容。以目标内容是一段文本内容为例,可以将步骤S110中得到的每个目标内容分段(即字符)依次组合在一起,从而生成完整的一段文本内容。在一个实施例中,素材内容可以是一段文本内容和一个问题,针对素材内容的目标内容可以从这段文本内容生成的针对这个问题的答案。该实施例的一个示例是抽取式阅读理解,即,从阅读理解的文章中抽取出部分内容,作为阅读理解的问题的答案。在另一个实施例中,素材内容可以是多个关键词,针对素材内容的目标内容可以从这些关键词生成的满足某种风格的诗歌、对联或歌词。在又一个实施例中,素材内容可以是一组音符与和弦,针对素材内容的目标内容可以从这组音符与和弦生成的一段音乐。

[0043] 在步骤S110中,利用内容生成模型处理素材内容,得到多个目标内容分段可以包括多个处理阶段,多个处理阶段中的每个处理阶段包括以下处理:利用内容生成模型处理素材内容,得到候选内容分段和候选内容分段的原始概率分布;根据预定下游数据集和前面的处理阶段输出的目标内容分段,调整原始概率分布,得到调整概率分布;以及根据调整概率分布,从候选内容分段中确定该处理阶段输出的对应目标内容分段。

[0044] 内容生成模型可以负责生成目标内容分段。在一个实施例中,目标内容可以是包括N个字符的文本内容,处理步骤可以相应地包括N个处理阶段,第一个处理阶段可以生成第1个字符,第二个处理阶段可以生成第2个字符,……,第N个处理阶段可以生成第N个字符。第1个字符、第2个字符、……、第N个字符组合起来可以构成完整的目标内容。N是大于1的整数。

[0045] 在第一个处理阶段至第N个处理阶段中的每一个处理阶段中,可以利用内容生成模型处理所述素材内容,得到候选内容分段和候选内容分段的原始概率分布。例如,当素材内容是一段英语的文字内容时,每一个处理阶段可以利用内容生成模型处理素材内容,得到多个英文字母作为候选内容分段,并得到这多个字母中的每个字母在该处理阶段的概率

分布,以作为原始概率分布。对素材内容的处理例如可以包括对素材内容的结构、特征、各个组成部分之间的关系等进行分析等处理,本公开对此不作特殊限制。

[0046] 每一个处理阶段可以基于候选内容分段的概率分布,从这些候选内容分段中选择一个或多个最优的候选内容分段,以作为该处理阶段输出的一个或多个对应目标内容分段。

[0047] 为了限制输出空间大小,可以基于预定下游数据集调整原始概率分布,得到调整概率分布,并根据该调整概率分布来从候选内容分段中选择出对应的目标内容分段。

[0048] 预定下游数据集是预先确定的一个数据集,用来构建受限空间,以将生成模型的输出限制在特定空间中。下游数据集是指该数据集是在模型生成输出时发挥调整作用,相对于预训练数据集而言是处于下游的。例如,预定下游数据集中可以包括更有可能或更希望出现在目标内容中的内容分段,可以将与预定下游数据集中的数据相对应的候选内容分段的概率分布调整得更大,使得这些候选内容分段更有可能被相应的处理阶段选中。通过这种方式,可以构建受限空间,使得输出的目标内容分段被限制在该受限空间中。受限空间是指由更有可能或更希望出现在目标内容中的内容分段构成的空间。预定下游数据集可以与前面的处理阶段输出的目标内容分段(即,当前已生成的全部目标内容分段)相结合,从而确定当前处理阶段中的调整概率分布。预定下游数据集能够用于构建受限空间,而前面的处理阶段输出的目标内容分段使得生成具有前序依赖的特性。

[0049] 如上所述,根据本公开实施例的目标内容的生成方法100可以利用预定下游数据集来构建受限空间,以将每个处理阶段输出的目标内容分段限制在受限空间中,从而限制生成模型输出空间的大小。

[0050] 每一个处理阶段在从候选内容分段中确定该处理阶段输出的对应目标内容分段时,可以采用例如贪心搜索算法、束搜索算法等来搜索期望的候选内容分段。贪心搜索算法可以选择输出概率值最大(即最优)的一个候选内容分段,而束搜索(beam search)算法可以选择输出概率值排名最靠前的W个候选内容分段,W即为束宽度(beam width)。束搜索算法在W=1时会变成贪心搜索算法。束搜索算法虽然不能找到最优解,但能够搜索出W个次优解,能够用比贪心搜索算法更快的速度找到最接近正确的解。

[0051] 在采用束搜索算法的实施例中,利用预定下游数据集在束搜索的基础上进行了改进,从而限制生成模型空间大小,生成满足可控风格的内容。

[0052] 图2是示出了根据本公开实施例的目标内容的生成方法的目标内容生成的一个示例的图。下面参考图2对根据本公开实施例的目标内容的生成方法100的目标内容生成的一个示例进行详细说明。

[0053] 为了便于清楚描述,图2示出的是一个简化的示例。以素材内容是一段文本内容和一个问题为例,目标内容生成根据这段文本内容生成针对该问题的答案文本,文本由英文字母组成。假设目标内容生成包括两个处理阶段,在第一个处理阶段中,向内容生成模型输入命令201START,表示开始执行生成模型的处理。第一个处理阶段对素材内容进行处理后得到的候选内容分段和它们的原始概率分布202为A(0.4)、B(0.1)、C(0.3)、D(0.2)。作为一个示例,第一个处理阶段可以将素材内容输入到BERT(Bidirectional Encoder Representations from Transformer)模型或ERNIE模型(BERT模型的一种改进模型)等中,BERT模型或ERNIE模型等基于该输入来输出候选内容分段和它们的原始概率分布202。BERT

模型或ERNIE模型均是自然语言模型,它们利用大规模语料进行预训练,获得文本的语义表示,并在特定的自然语言处理任务中进行微调,以应用于该任务。以目标内容生成采用束宽度为2的束搜索算法为例,如果根据原始概率分布来确定对应目标内容分段,则束搜索算法将输出概率较大的A(0.4)和C(0.3)。但为了限制输出空间大小,由受限模块203基于预定下游数据集204和第一个处理阶段的输入205START确定受限概率分布206,即A(0.0)、B(0.6)、C(0.3)、D(0.1),将受限概率分布206与原始概率分布202相加,得到调整概率分布A(0.4)、B(0.7)、C(0.6)、D(0.3)。基于此,束搜索算法将输出概率较大的B(0.7)和C(0.6),而不是根据原始概率分布得到的A和C,作为第一个处理阶段所确定的目标内容分段207。

[0054] 在第二个处理阶段中,针对第一个处理阶段所确定的目标内容分段207中的第一个目标内容分段B,对素材内容进行处理后得到的候选内容分段和它们的原始概率分布208为A(0.1)、B(0.3)、C(0.5)、D(0.1)。第二个处理阶段可以与第一个处理阶段同样地采用例如BERT模型或ERNIE模型等内容生成模型。由受限模块209基于预定下游数据集204和第二个处理阶段的输入210中的“B”确定受限概率分布211即A(0.6)、B(0.2)、C(0.1)、D(0.1),将受限概率分布211加到原始概率分布208上,得到调整概率分布A(0.7)、B(0.5)、C(0.6)、D(0.2)。假设目标内容生成所生成的目标内容的最大数量为2,则束搜索算法将输出概率最大的A(0.7),作为第二个处理阶段针对“B”所确定的目标内容分段。

[0055] 类似地,在第二个处理阶段中,针对第一个处理阶段所确定的目标内容分段207中的第二个目标内容分段C,对素材内容进行处理后得到的候选内容分段和它们的原始概率分布212为A(0.6)、B(0.1)、C(0.2)、D(0.1)。由受限模块209基于预定下游数据集204和第二个处理阶段的输入210中的“C”确定受限概率分布213即A(0.5)、B(0.2)、C(0)、D(0.3),将受限概率分布213加到原始概率分布212上,得到调整概率分布A(1.1)、B(0.3)、C(0.2)、D(0.4)。基于此,束搜索算法将输出概率最大的A(1.1),作为第二个处理阶段针对“C”所确定的目标内容分段。

[0056] 最终,该目标内容生成示例将输出“BA”和“CA”作为目标内容。

[0057] 在该示例中,考虑到作为目标内容的答案一般是从作为素材内容的文本内容和问题中抽取出来的,因此可以认为每个目标内容分段均包括在素材内容中。在这种情况下,预定下游数据集可以包括素材内容,这样可以增大与素材内容中包括的字母相对应的候选字母的概率值。当然,预定下游数据集不限于此,而可以是能够限制输出空间的任何数据集。例如,在要生成特定风格的诗歌时,预定下游数据集可以包括特定的诗歌集合。

[0058] 在该示例中,如上所述,由受限模块基于预定下游数据集和当前阶段的输入(即,前面的处理阶段输出的目标内容分段)确定受限概率分布,以调整原始概率分布。也就是说,受限概率分布是针对候选内容分段确定的概率分布,用来调整候选内容分段的原始概率分布,以限制生成模型输出空间大小。利用受限概率分布,可以容易地调整候选内容分段的原始概率分布,从而更有效地限制生成模型输出空间大小。

[0059] 需要注意,虽然在该示例中将受限概率分布与原始概率分布相加而得到调整概率分布,但本公开不限于此,只要能够基于预定下游数据集和当前阶段的输入来调整原始概率分布,可以采用本领域技术人员可以想到的任何调整方式。

[0060] 在一个实施例中,受限模块可以利用字典树来实现。通过利用字典树实现受限模块,能够使受限概率分布具有前序依赖的特性,并且无需进行模型微调。字典树又称Trie

树,是一种专门用于字符串匹配的树形结构,能够高效地在—组字符串中搜索期望字符串,与红黑树、散列表类似。当然,受限模块并不限于字典树,而是可以用本领域技术人员能够想到的各种方式来实现,本公开对此不作特殊限制。

[0061] 图3示出了根据本公开实施例的用于生成受限概率分布的字典树的一个示例。下面参考图3对根据本公开实施例的用于生成受限概率分布的字典树的一个示例进行说明。

[0062] 基于预定下游数据集中包括的多个字符串,可以构建图3所示的字典树300。例如,图3所示的字典树300中,最上面的一个节点301为根节点,表示字符串的开始,根节点下面的第一层中的节点302“B”(0.6)、节点303“C”(0.3)、节点304“D”(0.1)表示预定下游数据集中的字符串的第一个字符是“B”、“C”、“D”中之一,并且基于预定下游数据集中的字母出现的频率,第一个字符是“B”的概率为0.6,第一个字符是“C”的概率为0.3,第一个字符是“D”的概率为0.1,第一个字符是“A”的概率为0。这样,在根据本公开实施例的目标内容的生成方法的第一个处理阶段中,基于当前阶段的输入START,可以从图3的字典树300中得到受限概率分布A(0.0)、B(0.6)、C(0.3)、D(0.1)。START对应于根节点,它们不对应于任何实际的内容字符,而是表示起始点。

[0063] 图3所示的字典树300中,对于根节点下面的第一层中的节点302“B”,该节点下面一层的节点305“A”(0.6)、节点306“B”(0.2)、节点307“C”(0.1)、节点308“D”(0.1)表示在预定下游数据集中的以“B”开头的那些字符串中,第二个字符是“A”的概率为0.6,第二个字符是“B”的概率为0.2,第二个字符是“C”的概率为0.1,第二个字符是“D”的概率为0.1。这样,在根据本公开实施例的目标内容的生成方法的第二个处理阶段中,基于当前阶段的输入“B”,可以从图3的字典树300中得到受限概率分布A(0.6)、B(0.2)、C(0.1)、D(0.1)。

[0064] 类似地,图3所示的字典树300中,对于根节点下面的第一层中的节点303“C”,该节点下面一层的节点309“A”(0.5)、节点310“B”(0.2)、节点311“D”(0.3)表示在预定下游数据集中的以“C”开头的那些字符串中,第二个字符是“A”的概率为0.5,第二个字符是“B”的概率为0.2,第二个字符是“C”的概率为0,第二个字符是“D”的概率为0.3。在图2的第二个处理阶段中,基于当前阶段的输入“C”,可以从图3的字典树300中得到受限概率分布A(0.5)、B(0.2)、C(0)、D(0.3)。

[0065] 如上所述,在用字典树实现受限模块时,首先利用预定下游数据集构建字典树,然后在每个处理阶段中,通过在字典树中搜索前面的处理阶段输出的目标内容分段,来确定受限概率分布。通过这种方式,能够快速而高效地确定具有前序依赖特性的受限概率分布。

[0066] 图4是根据本公开实施例的目标内容的生成装置400的示意图。下面参考图4对根据本公开实施例的目标内容的生成装置进行说明。目标内容的生成装置400包括处理模块410和生成模块420。

[0067] 处理模块410被配置为利用内容生成模型处理素材内容,得到多个目标内容分段。

[0068] 生成模块420被配置为根据多个目标内容分段,生成针对所述素材内容的目标内容。

[0069] 处理模块410包括多个处理子模块。每个处理子模块用于:利用内容生成模型处理素材内容,得到候选内容分段和候选内容分段的原始概率分布;根据预定下游数据集和前面的处理阶段输出的目标内容分段,调整原始概率分布,得到调整概率分布;以及根据调整概率分布,从候选内容分段中确定该处理阶段输出的对应目标内容分段。

[0070] 根据该目标内容的生成装置400,同样能够通过利用预定下游数据集来有效地限制生成模型的输出空间大小,以生成更加符合特定任务的结果。

[0071] 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,通过利用预定下游数据集来有效地限制生成模型的输出空间大小,以生成更加符合特定任务的结果。本公开的实施例无需像现有技术那样构建不同的风格限定词和样本,也无需进行微调,成本较低,并且可控性良好。

[0072] 图5示出了可以用来实施本公开的实施例的示例电子设备500的示意性框图。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。

[0073] 如图5所示,设备500包括计算单元501,其可以根据存储在只读存储器 (ROM) 502中的计算机程序或者从存储单元508加载到随机访问存储器 (RAM) 503中的计算机程序,来执行各种适当的动作和处理。在RAM 503中,还可存储设备500操作所需的各种程序和数据。计算单元501、ROM 502以及RAM 503通过总线504彼此相连。输入/输出 (I/O) 接口505也连接至总线504。

[0074] 设备500中的多个部件连接至I/O接口505,包括:输入单元506,例如键盘、鼠标等;输出单元507,例如各种类型的显示器、扬声器等;存储单元508,例如磁盘、光盘等;以及通信单元509,例如网卡、调制解调器、无线通信收发机等。通信单元509允许设备500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。

[0075] 计算单元501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501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501执行上文所描述的各个方法和处理,例如由上述路侧计算设备、交通提示设备的处理器或远处处理器执行的方法和处理。例如,在一些实施例中,这些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508。在一些实施例中,计算机程序的部分或者全部可以经由ROM 502和/或通信单元509而被载入和/或安装到设备500上。当计算机程序加载到RAM 503并由计算单元501执行时,可以执行上文描述的分布式系统的测试方法的一个或多个步骤。备选地,在其他实施例中,计算单元501可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行上述方法。设备500例如可以是分布式系统的控制中心,或者是位于分布式系统内部或外部的任意设备。设备500不限于上述示例,只要能够实现上述测试方法即可。

[0076] 本文中以上描述的系统和技术和各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列 (FPGA)、专用集成电路 (ASIC)、专用标准产品 (ASSP)、芯片上系统的系统 (SOC)、负载可编程逻辑设备 (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。

[0077] 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/或框图中所规定的功能/操作被实施。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。

[0078] 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。

[0079] 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。

[0080] 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)和互联网。

[0081] 计算机系统可以包括客户端和服务端。客户端和服务端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端的关系。服务器可以是云服务器,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。

[0082] 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公开的技术方案所期望的结果,本文在此不进行限制。

[0083] 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。

100

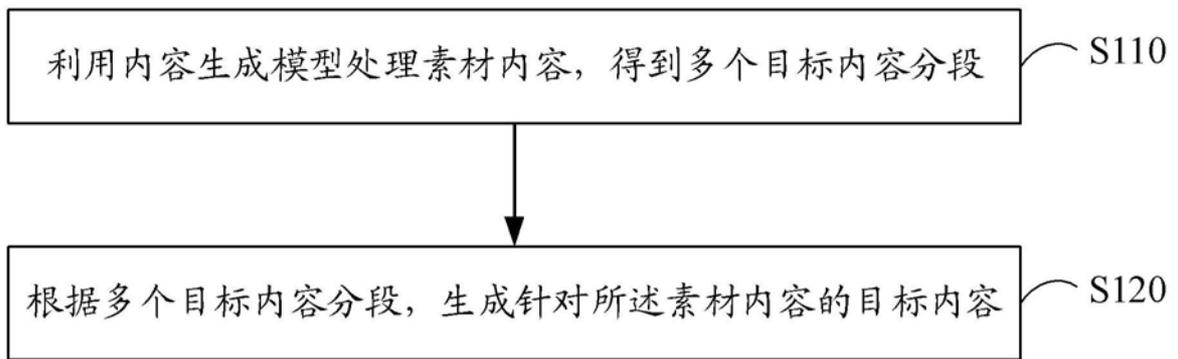


图1

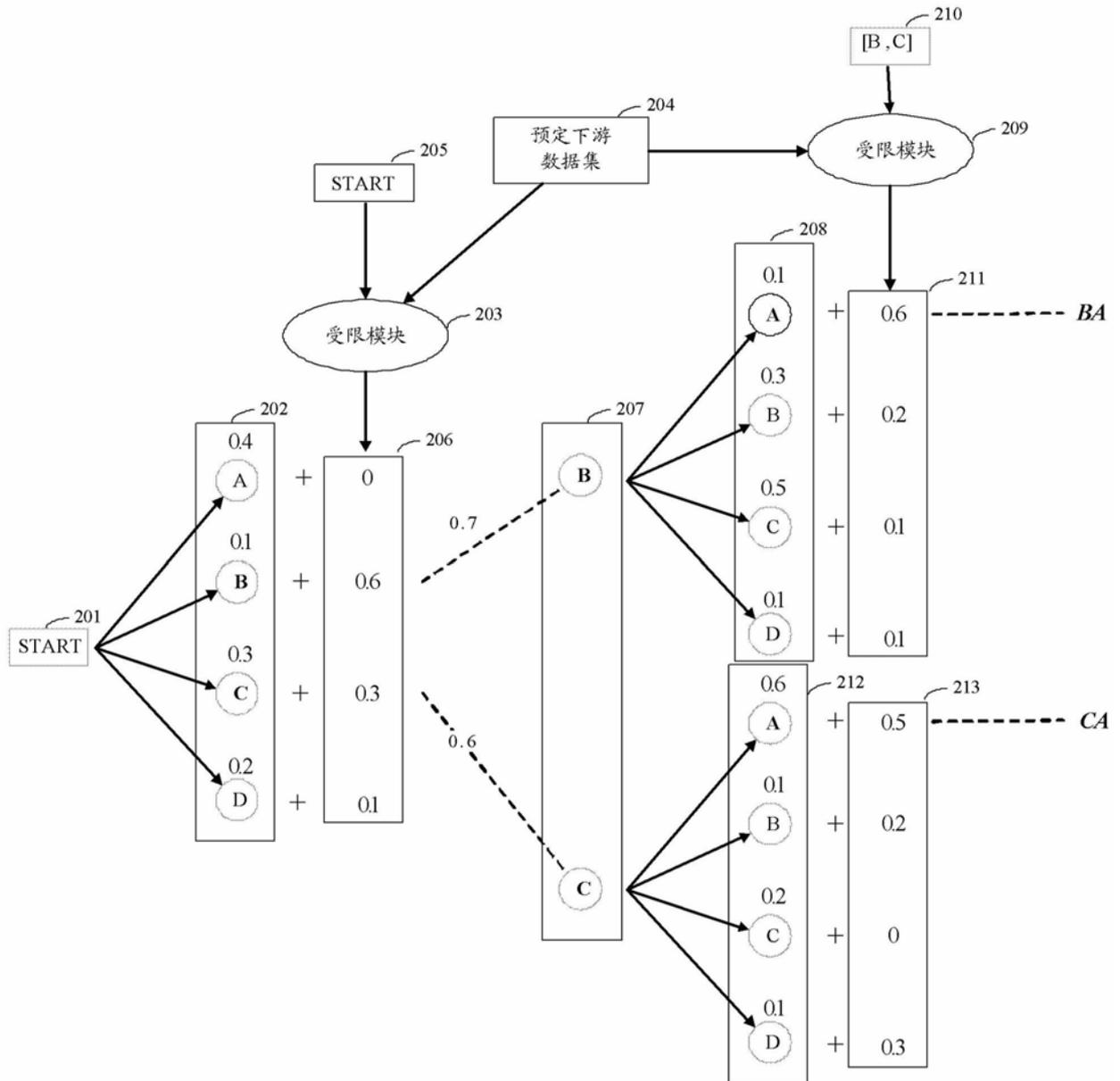


图2

300

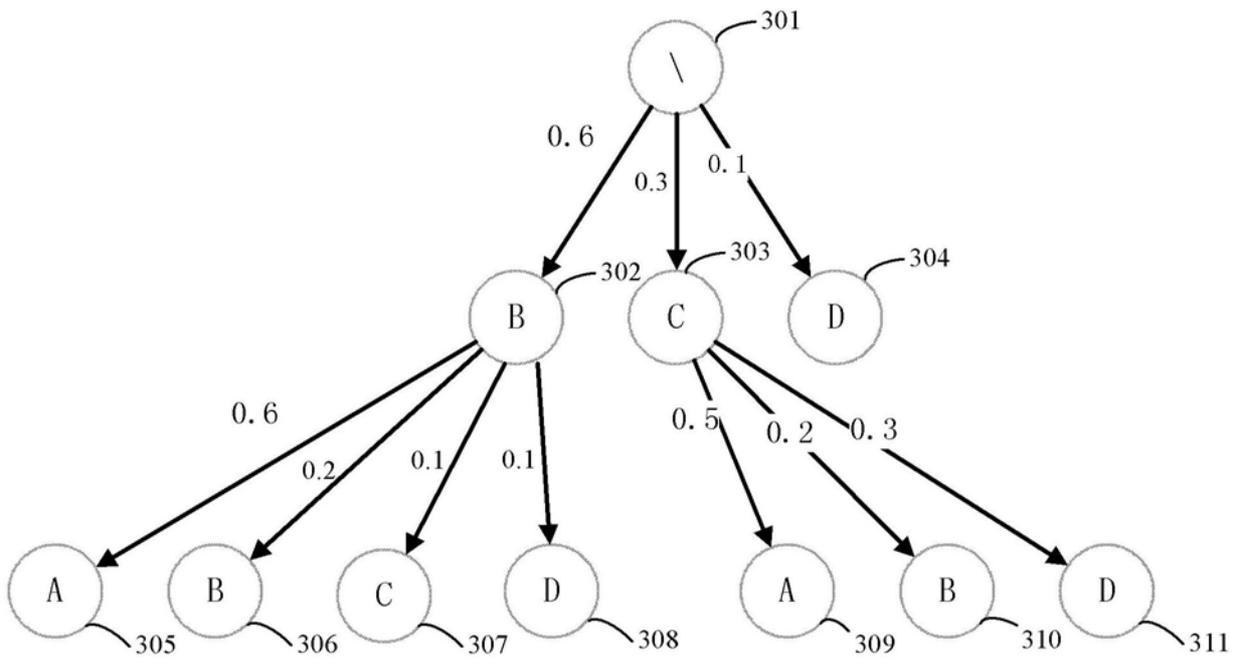


图3

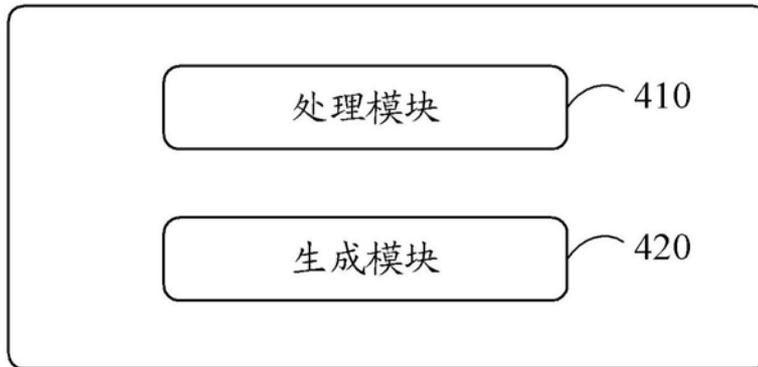


图4

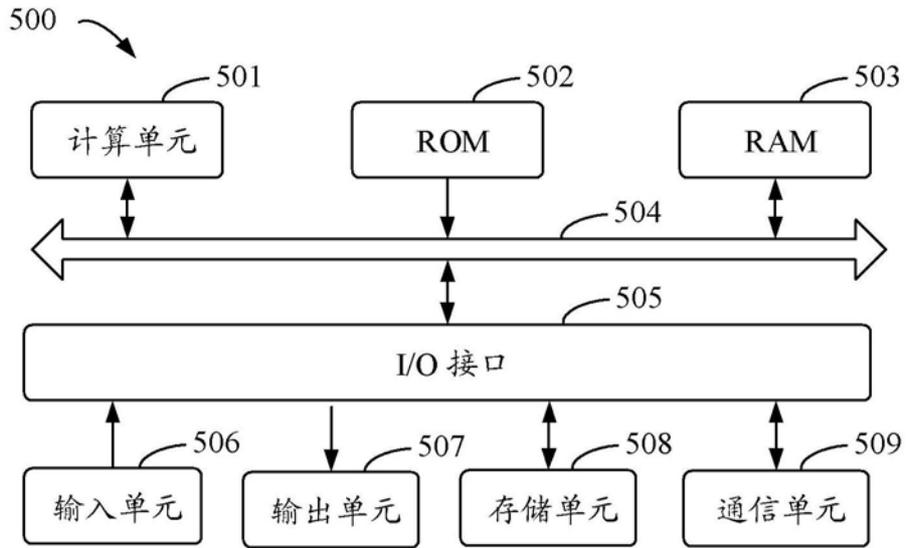


图5