(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116186244 A
(43)申请公布日 2023.05.30

(21)申请号 202310117385.9
(22)申请日 2023.01.19
(71)申请人 阿里巴巴达摩院(杭州)科技有限公司
地址 311121 浙江省杭州市余杭区五常街道文一西路969号3幢5层516室
(72)发明人 颜为骧 陈谦 王雯 张庆林
(74)专利代理机构 北京众达德权知识产权代理有限公司 11570
专利代理师 袁媛
(51)Int.Cl.
G06F 16/34(2019.01)
G06F 16/35(2019.01)
G06F 18/214(2023.01)
G06N 3/0464(2023.01)
G06N 3/08(2023.01)

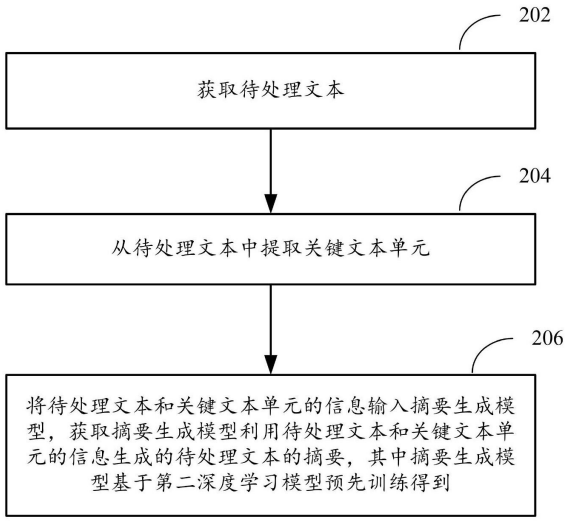
权利要求书3页 说明书16页 附图9页

(54)发明名称

生成文本摘要的方法、训练摘要生成模型的方法及装置

(57)摘要

本申请实施例公开了一种生成文本摘要的方法、训练摘要生成模型的方法及装置，涉及人工智能技术领域。主要技术方案包括：获取待处理文本；从所述待处理文本中提取关键文本单元；将所述待处理文本和所述关键文本单元的信息输入摘要生成模型，获取所述摘要生成模型利用所述待处理文本和所述关键文本单元的信息生成的所述待处理文本的摘要；其中所述摘要生成模型是基于第二深度学习模型预先训练得到的。本申请通过关键文本单元的提取和引入为摘要的生成提供指导，使得摘要生成模型能够聚焦待处理文本中的关键内容，降低噪声的影响，从而提高针对长文本生成摘要的准确性。



CN 116186244 A

1.一种生成文本摘要的方法,其特征在于,所述方法包括:

获取待处理文本;

从所述待处理文本中提取关键文本单元;

将所述待处理文本和所述关键文本单元的信息输入摘要生成模型,获取所述摘要生成模型利用所述待处理文本和所述关键文本单元的信息生成的所述待处理文本的摘要;

其中所述摘要生成模型是基于第二深度学习模型预先训练得到的。

2.根据权利要求1所述的方法,其特征在于,从所述待处理文本中提取关键文本单元包括:

将所述待处理文本输入关键文本提取模型,获取所述关键文本提取模型从所述待处理文本中提取的关键文本单元,其中所述关键文本提取模型基于第一深度学习模型预先训练得到;或者,

利用预设的特征规则从所述待处理文本中提取关键文本单元;或者,

将所述待处理文本输入关键文本提取模型,获取所述关键文本提取模型从所述待处理文本中提取的第一关键文本单元,利用预设的特征规则从所述待处理文本中提取第二关键文本单元,将所述第一关键文本单元和所述第二关键文本单元进行融合,得到关键文本单元。

3.根据权利要求2所述的方法,其特征在于,所述关键文本提取模型包括第一编码网络和分类网络;

所述第一编码网络对所述待处理文本进行编码处理,得到所述待处理文本中各文本单元的特征表示;

所述分类网络利用所述各文本单元的特征表示对各文本单元进行分类,得到各文本单元是否为关键文本单元的分类结果。

4.根据权利要求1所述的方法,其特征在于,所述摘要生成模型包括第二编码网络和解码网络;

所述第二编码网络利用所述关键文本单元的信息对所述待处理文本进行编码处理,得到所述待处理文本中各元素Token的特征表示;

所述解码网络利用所述待处理文本中各Token的特征表示进行解码处理,生成所述待处理文本的摘要。

5.根据权利要求4所述的方法,其特征在于,所述第二编码网络利用所述关键文本单元的信息对所述待处理文本进行编码处理,得到所述待处理文本中各Token的特征表示包括:

所述第二编码网络对待处理文本进行嵌入处理后,得到所述待处理文本中各Token的嵌入特征;

利用所述关键文本单元的信息对所述各Token的嵌入特征进行注意力机制的处理,得到各Token的特征表示,其中所述注意力机制的处理包括:对属于关键文本单元的各Token进行注意力处理时利用所述待处理文本中所有Token的嵌入特征,对不属于关键文本单元的各Token进行注意力处理时利用距离该Token预设窗口距离内的各Token的嵌入特征。

6.根据权利要求1至5中任一项所述的方法,其特征在于,所述方法应用于在线会议场景,所述待处理文本为在线会议的会议记录,所述关键文本单元为关键句,所述摘要为所述会议记录的会议摘要。

7.一种训练摘要生成模型的方法,其特征在于,所述方法包括:

获取多个第二训练样本,所述第二训练样本包括第二文本样本、所述第二文本样本的关键文本单元的信息以及所述第二文本样本的摘要样本;

利用所述多个第二训练样本训练所述摘要生成模型,其中所述摘要生成模型包括第二编码网络和解码网络;

所述第二编码网络利用第二文本样本的关键文本单元的信息对第二文本样本进行编码处理,得到第二文本样本中各Token的特征表示;

所述解码网络利用所述第二文本样本中各Token的特征表示进行解码处理,生成所述第二文本样本的摘要;

所述训练的目标包括:最小化所述解码网络生成的所述第二文本样本的摘要与所述第二文本的摘要样本之间的差异。

8.根据权利要求7所述的方法,其特征在于,所述第二编码网络利用第二文本样本的关键文本单元的信息对第二文本样本进行编码处理,得到第二文本样本中各Token的特征表示包括:

所述第二编码网络获取对第二文本样本进行嵌入处理后,得到的所述第二文本样本中各Token的嵌入特征;

利用所述第二文本样本的关键文本单元的信息对所述各Token的嵌入特征进行注意力机制的处理,得到各Token的特征表示,其中所述注意力机制的处理包括:对属于关键文本单元的各Token进行注意力处理时利用所述第二文本样本中所有Token的嵌入特征,对不属于关键文本单元的各Token进行注意力处理时利用距离该Token预设窗口距离内的各Token的嵌入特征。

9.一种训练关键文本提取模型的方法,其特征在于,所述方法包括:

获取多个第一训练样本,所述第一训练样本包括第一文本样本以及所述第一文本样本被标注的关键文本单元标签;

利用所述多个第一训练样本训练关键文本提取模型,其中所述关键文本提取模型包括第一编码网络和分类网络;

所述第一编码网络对所述第一文本样本进行编码处理,得到所述第一文本样本中各文本单元的特征表示;

所述分类网络利用所述各文本单元的特征表示对各文本单元进行分类,得到各文本单元是否为关键文本单元的分类结果;

所述训练的目标包括:最小化所述分类网络的分类结果与所述第一文本样本被标注的关键文本单元标签之间的差异。

10.一种文本摘要生成装置,其特征在于,所述装置包括:

文本获取单元,被配置为获取待处理文本;

关键提取单元,被配置为从所述待处理文本中提取关键文本单元;

摘要生成单元,被配置为将所述待处理文本和所述关键文本单元的信息输入摘要生成模型,获取所述摘要生成模型利用所述待处理文本和所述关键文本单元的信息生成的所述待处理文本的摘要;其中所述摘要生成模型是基于第二深度学习模型预先训练得到的。

11.一种训练摘要生成模型的装置,其特征在于,所述装置包括:

第二样本获取单元，被配置为获取多个第二训练样本，所述第二训练样本包括第二文本样本、所述第二文本样本的关键文本单元的信息以及所述第二文本样本的摘要样本；

第二模型训练单元，被配置为利用所述多个第二训练样本训练所述摘要生成模型，其中所述摘要生成模型包括第二编码网络和解码网络；所述第二编码网络利用第二文本样本的关键文本单元的信息对第二文本样本进行编码处理，得到第二文本样本中各Token的特征表示；所述解码网络利用所述第二文本样本中各Token的特征表示进行解码处理，生成所述第二文本样本的摘要；所述训练的目标包括：最小化所述解码网络生成的所述第二文本样本的摘要与所述第二文本的摘要样本之间的差异。

12.一种训练关键文本提取模型的装置，其特征在于，所述装置包括：

第一样本获取单元，被配置为获取多个第一训练样本，所述第一训练样本包括第一文本样本以及所述第一文本样本被标注的关键文本单元标签；

第一模型训练单元，被配置为利用所述多个第一训练样本训练关键文本提取模型，其中所述关键文本提取模型包括第一编码网络和分类网络；所述第一编码网络对所述第一文本样本进行编码处理，得到所述第一文本样本中各文本单元的特征表示；所述分类网络利用所述各文本单元的特征表示对各文本单元进行分类，得到各文本单元是否为关键文本单元的分类结果；所述训练的目标包括：最小化所述分类网络的分类结果与所述第一文本样本被标注的关键文本单元标签之间的差异。

13.一种计算机可读存储介质，其上存储有计算机程序，其特征在于，该程序被处理器执行时实现权利要求1至9中任一项所述的方法的步骤。

14.一种电子设备，其特征在于，包括：

一个或多个处理器；以及

与所述一个或多个处理器关联的存储器，所述存储器用于存储程序指令，所述程序指令在被所述一个或多个处理器读取执行时，执行权利要求1至9中任一项所述的方法的步骤。

生成文本摘要的方法、训练摘要生成模型的方法及装置

技术领域

[0001] 本申请涉及人工智能技术领域，特别是涉及一种生成文本摘要的方法、训练摘要生成模型的方法及装置。

背景技术

[0002] 利用人工智能技术将大量文本进行处理，产生简洁、精炼内容的过程就是摘要生成。人们可以通过阅读摘要来把握文本主要内容，节省时间，提高阅读效率。然而在很多场景下需要针对长文本生成摘要，例如针对视频会议、讲座、面试等记录生成摘要。这些场景下的长文本存在持续时间久、关键信息分布稀疏等特点，现有的学术研究方案和工业解决方案均难以针对长文本进行准确地摘要生成。

发明内容

[0003] 有鉴于此，本申请提供了一种生成文本摘要的方法、训练摘要生成模型的方法及装置，用以针对长文本实现准确地摘要生成。

[0004] 本申请提供了如下方案：

[0005] 第一方面，提供了一种生成文本摘要的方法，所述方法包括：

[0006] 获取待处理文本；

[0007] 从所述待处理文本中提取关键文本单元；

[0008] 将所述待处理文本和所述关键文本单元的信息输入摘要生成模型，获取所述摘要生成模型利用所述待处理文本和所述关键文本单元的信息生成的所述待处理文本的摘要；

[0009] 其中所述摘要生成模型是基于第二深度学习模型预先训练得到的。

[0010] 根据本申请实施例中一可实现的方式，从所述待处理文本中提取关键文本单元包括：

[0011] 将所述待处理文本输入关键文本提取模型，获取所述关键文本提取模型从所述待处理文本中提取的关键文本单元，其中所述关键文本提取模型基于第一深度学习模型预先训练得到；或者，

[0012] 利用预设的特征规则从所述待处理文本中提取关键文本单元；或者，

[0013] 将所述待处理文本输入关键文本提取模型，获取所述关键文本提取模型从所述待处理文本中提取的第一关键文本单元，利用预设的特征规则从所述待处理文本中提取第二关键文本单元，将所述第一关键文本单元和所述第二关键文本单元进行融合，得到关键文本单元。

[0014] 根据本申请实施例中一可实现的方式，所述关键文本提取模型包括第一编码网络和分类网络；

[0015] 所述第一编码网络对所述待处理文本进行编码处理，得到所述待处理文本中各文本单元的特征表示；

[0016] 所述分类网络利用所述各文本单元的特征表示对各文本单元进行分类，得到各文

本单元是否为关键文本单元的分类结果。

[0017] 根据本申请实施例中一可实现的方式,所述摘要生成模型包括第二编码网络和解码网络;

[0018] 所述第二编码网络利用所述关键文本单元的信息对所述待处理文本进行编码处理,得到所述待处理文本中各元素Token的特征表示;

[0019] 所述解码网络利用所述待处理文本中各Token的特征表示进行解码处理,生成所述待处理文本的摘要。

[0020] 根据本申请实施例中一可实现的方式,所述第二编码网络利用所述关键文本单元的信息对所述待处理文本进行编码处理,得到所述待处理文本中各Token的特征表示包括:

[0021] 所述第二编码网络对待处理文本进行嵌入处理后,得到所述待处理文本中各Token的嵌入特征;

[0022] 利用所述关键文本单元的信息对所述各Token的嵌入特征进行注意力机制的处理,得到各Token的特征表示,其中所述注意力机制的处理包括:对属于关键文本单元的各Token进行注意力处理时利用所述待处理文本中所有Token的嵌入特征,对不属于关键文本单元的各Token进行注意力处理时利用距离该Token预设窗口距离内的各Token的嵌入特征。

[0023] 根据本申请实施例中一可实现的方式,所述方法应用于在线会议场景,所述待处理文本为在线会议的会议记录,所述关键文本单元为关键句,所述摘要为所述会议记录的会议摘要。

[0024] 第二方面,提供给了一种训练摘要生成模型的方法,所述方法包括:

[0025] 获取多个第二训练样本,所述第二训练样本包括第二文本样本、所述第二文本样本的关键文本单元的信息以及所述第二文本样本的摘要样本;

[0026] 利用所述多个第二训练样本训练所述摘要生成模型,其中所述摘要生成模型包括第二编码网络和解码网络;

[0027] 所述第二编码网络利用第二文本样本的关键文本单元的信息对第二文本样本进行编码处理,得到第二文本样本中各Token的特征表示;

[0028] 所述解码网络利用所述第二文本样本中各Token的特征表示进行解码处理,生成所述第二文本样本的摘要;

[0029] 所述训练的目标包括:最小化所述解码网络生成的所述第二文本样本的摘要与所述第二文本的摘要样本之间的差异。

[0030] 根据本申请实施例中一可实现的方式,所述第二编码网络利用第二文本样本的关键文本单元的信息对第二文本样本进行编码处理,得到第二文本样本中各Token的特征表示包括:

[0031] 所述第二编码网络获取对第二文本样本进行嵌入处理后,得到的所述第二文本样本中各Token的嵌入特征;

[0032] 利用所述第二文本样本的关键文本单元的信息对所述各Token的嵌入特征进行注意力机制的处理,得到各Token的特征表示,其中所述注意力机制的处理包括:对属于关键文本单元的各Token进行注意力处理时利用所述第二文本样本中所有Token的嵌入特征,对不属于关键文本单元的各Token进行注意力处理时利用距离该Token预设窗口距离内的各

Token的嵌入特征。

[0033] 第三方面，提供了一种训练关键文本提取模型的方法，所述方法包括：

[0034] 获取多个第一训练样本，所述第一训练样本包括第一文本样本以及所述第一文本样本被标注的关键文本单元标签；

[0035] 利用所述多个第一训练样本训练关键文本提取模型，其中所述关键文本提取模型包括第一编码网络和分类网络；

[0036] 所述第一编码网络对所述第一文本样本进行编码处理，得到所述第一文本样本中各文本单元的特征表示；

[0037] 所述分类网络利用所述各文本单元的特征表示对各文本单元进行分类，得到各文本单元是否为关键文本单元的分类结果；

[0038] 所述训练的目标包括：最小化所述分类网络的分类结果与所述第一文本样本被标注的关键文本单元标签之间的差异。

[0039] 第四方面，提供了一种文本摘要生成装置，所述装置包括：

[0040] 文本获取单元，被配置为获取待处理文本；

[0041] 关键提取单元，被配置为从所述待处理文本中提取关键文本单元；

[0042] 摘要生成单元，被配置为将所述待处理文本和所述关键文本单元的信息输入摘要生成模型，获取所述摘要生成模型利用所述待处理文本和所述关键文本单元的信息生成的所述待处理文本的摘要；其中所述摘要生成模型是基于第二深度学习模型预先训练得到的。

[0043] 第五方面，提供了一种训练摘要生成模型的装置，所述装置包括：

[0044] 第二样本获取单元，被配置为获取多个第二训练样本，所述第二训练样本包括第二文本样本、所述第二文本样本的关键文本单元的信息以及所述第二文本样本的摘要样本；

[0045] 第二模型训练单元，被配置为利用所述多个第二训练样本训练所述摘要生成模型，其中所述摘要生成模型包括第二编码网络和解码网络；所述第二编码网络利用第二文本样本的关键文本单元的信息对第二文本样本进行编码处理，得到第二文本样本中各Token的特征表示；所述解码网络利用所述第二文本样本中各Token的特征表示进行解码处理，生成所述第二文本样本的摘要；所述训练的目标包括：最小化所述解码网络生成的所述第二文本样本的摘要与所述第二文本的摘要样本之间的差异。

[0046] 第六方面，提供了一种训练关键文本提取模型的装置，所述装置包括：

[0047] 第一样本获取单元，被配置为获取多个第一训练样本，所述第一训练样本包括第一文本样本以及所述第一文本样本被标注的关键文本单元标签；

[0048] 第一模型训练单元，被配置为利用所述多个第一训练样本训练关键文本提取模型，其中所述关键文本提取模型包括第一编码网络和分类网络；所述第一编码网络对所述第一文本样本进行编码处理，得到所述第一文本样本中各文本单元的特征表示；所述分类网络利用所述各文本单元的特征表示对各文本单元进行分类，得到各文本单元是否为关键文本单元的分类结果；所述训练的目标包括：最小化所述分类网络的分类结果与所述第一文本样本被标注的关键文本单元标签之间的差异。

[0049] 根据第七方面，提供了一种计算机可读存储介质，其上存储有计算机程序，该程序

被处理器执行时实现上述第一方面至第三方面中任一项所述的方法的步骤。

[0050] 根据第八方面，提供了一种电子设备，包括：

[0051] 一个或多个处理器；以及

[0052] 与所述一个或多个处理器关联的存储器，所述存储器用于存储程序指令，所述程序指令在被所述一个或多个处理器读取执行时，执行上述第一方面至第三方面中任一项所述的方法的步骤。

[0053] 根据本申请提供的具体实施例，本申请公开了以下技术效果：

[0054] 1)本申请首先从待处理文本中提取关键文本单元，再由基于深度学习模型的摘要生成模型利用待处理文本和关键文本单元的信息生成待处理文本的摘要。通过关键文本单元的提取和引入为摘要的生成提供指导，使得摘要生成模型能够聚焦待处理文本中的关键内容，降低噪声的影响，从而提高针对长文本生成摘要的准确性。

[0055] 2)本申请中不仅可以采用有监督方式(即通过关键文本提取模型)实现对待处理文本的关键文本单元的提取，还可以进一步结合无监督(即利用预设的特征规则)的方式，对提取的关键文本单元的信息进行增强。

[0056] 3)本申请中摘要生成模型在利用关键文本单元的信息对待处理文本生成摘要时，仅对关键文本单元中的各Token进行全局注意力处理，而对于其他各Token进行局部注意力处理，从而增强了摘要生成模型聚焦关键信息的能力和抗噪声干扰的能力，降低了计算开销，使得摘要生成模型能够顺利针对长文本实现摘要生成。

[0057] 当然，实施本申请的任一产品并不一定需要同时达到以上所述的所有优点。

附图说明

[0058] 为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例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，对于本领域普通技术人员来讲，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。

[0059] 图1为是本申请实施例所适用的系统架构图；

[0060] 图2为本申请实施例提供的文本摘要生成方法的主要流程图；

[0061] 图3为本申请实施例提供的关键文本提取模型的结构原理图；

[0062] 图4为本申请实施例提供的关键文本提取模型的训练方法流程图；

[0063] 图5为本申请实施例提供的摘要生成模型的结构原理图；

[0064] 图6为本申请实施例提供的摘要生成模型的训练方法流程图；

[0065] 图7为本申请实施例提供的在线会议记录的摘要提取的示意图；

[0066] 图8为本申请实施例提供的文本摘要生成装置的示意性框图；

[0067] 图9为本申请实施例提供的训练关键文本提取模型的装置的示意性框图；

[0068] 图10为本申请实施例提供的训练摘要生成模型的装置的示意性框图；

[0069] 图11为本申请实施例提供的电子设备的示意性框图。

具体实施方式

[0070] 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，本领域普通技术人员所获得的所有其他实施例，都属于本申请保护的范围。

[0071] 在本发明实施例中使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本发明。在本发明实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其他含义。

[0072] 应当理解，本文中使用的术语“和/或”仅仅是一种描述关联对象的关联关系，表示可以存在三种关系，例如，A和/或B，可以表示：单独存在A，同时存在A和B，单独存在B这三种情况。另外，本文中字符“/”，一般表示前后关联对象是一种“或”的关系。

[0073] 取决于语境，如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”或“响应于检测”。类似地，取决于语境，短语“如果确定”或“如果检测(陈述的条件或事件)”可以被解释成为“当确定时”或“响应于确定”或“当检测(陈述的条件或事件)时”或“响应于检测(陈述的条件或事件)”。

[0074] 目前基于Transformer(转换)网络的模型在针对短文本生成摘要方面取得了优异的性能。针对长文本摘要的研究只停留在学术界，学术界给出了多种解决方案，但效果均不佳。

[0075] 有鉴于此，本申请提供了一种全新的摘要提取框架。为了方便对本申请的理解，首先对本申请所适用的系统架构进行简单描述。图1示出了可以应用本申请实施例的示例性系统架构，如图1中所示，该系统架构可以包括第一模型训练装置、第二模型训练装置和文本摘要生成装置。

[0076] 其中，第一模型训练装置在获取第一训练数据后，可以采用本申请实施例提供的方法进行模型训练，得到关键文本提取模型。

[0077] 第二模型训练装置在获取第二训练数据后，可以采用本申请实施例提供的方法进行模型训练，得到摘要生成模型。

[0078] 上述第一模型训练装置和第二模型训练装置可以采用离线方式建立摘要生成模型。

[0079] 文本摘要生成装置利用摘要生成模型针对输入的待处理文本生成摘要，在摘要生成过程中还可以进一步结合关键文本提取模型进行关键文本单元的提取，进而利用提取的关键文本单元针对待处理文本来生成摘要。文本摘要生成装置可以在线进行摘要的生成，也可以采用离线的方式进行摘要的生成。上述第一模型训练装置和关键文本提取模型非必须，也可以采用其他方式进行关键文本单元的提取，具体将在后续实施例中详述。

[0080] 第一模型训练装置、第二模型训练装置和文本摘要生成装置可以分别设置为独立的服务器，也可以设置于同一个服务器或服务器群组，还可以设置于独立的或者同一云服务器。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，以解决传统物理主机与虚拟专用服务器(VPs，VirtualPrivateServer)服务中存在的管理难度大，服务扩展性弱的缺陷。第一模型训练装置、第二模型训练装置和文本摘要生成装置还可以设置于具有较强计算能力的计算机终端。

[0081] 应该理解，图1中的第一模型训练装置、第二模型训练装置、文本摘要生成装置、关键文本提取模型和摘要生成模型的数目仅仅是示意性的。根据实现需要，可以具有任意数

目的第一模型训练装置、第二模型训练装置、文本摘要生成装置、关键文本提取模型和摘要生成模型。

[0082] 图2为本申请实施例提供的文本摘要生成方法的主要流程图，该方法可以由图1所示系统中的文本摘要生成装置执行。如图2中所示，该方法可以包括以下步骤：

[0083] 步骤202：获取待处理文本。

[0084] 步骤204：从待处理文本中提取关键文本单元。

[0085] 步骤206：将待处理文本和关键文本单元的信息输入摘要生成模型，获取摘要生成模型利用待处理文本和关键文本单元的信息生成的待处理文本的摘要，其中摘要生成模型基于第二深度学习模型预先训练得到。

[0086] 由上述流程可以看出，本申请首先从待处理文本中提取关键文本单元，然后由基于深度学习模型的摘要生成模型利用待处理文本和关键文本单元的信息生成待处理文本的摘要。通过关键文本单元的提取和引入为摘要的生成提供指导，从而提高生成摘要的准确性。

[0087] 需要说明的是，本公开中涉及的“第一”、“第二”等限定并不具备大小、顺序和数量等方面的限制，仅仅用以在名称上加以区分。例如“第一深度学习模型”和“第二深度学习模型”用以区分两个深度学习模型，再例如“第一关键文本单元”和“第二关键文本单元”用以区分两个关键文本单元，再例如“第一编码网络”和“第二编码网络”用以区分两个编码网络。

[0088] 下面结合实施例分别对上述流程中的各步骤进行详细描述。首先对上述步骤202即“获取待处理文本”进行详细描述。

[0089] 本申请实施例中涉及的待处理文本指的是需要从中提取摘要的文本。在本申请实施例中，该待处理文本可以是长文本，也可以是短文本。也就是说，本申请实施例提供的方式对待处理文本的长度没有限制，不仅能够实现短文本的摘要提取，也可以实现长文本的摘要提取。其中“长文本”和“短文本”是一个相对的概念，具体的标准可以自定义。例如可以将少于512个字符的文本称为短文本，将多于或等于512个字符的文本称为长文本。

[0090] 待处理文本可以从存储文本的数据库中获取，也可以从生成该待处理文本的系统中获取。例如，在线会议服务可以针对用户的会议内容生成会议记录，本申请实施例中的文本摘要生成装置可以从在线会议服务端获取该会议记录。在线会议服务可以将会议记录持久化至数据库中，本申请实施例中的文本摘要生成装置可以在获取到摘要生成指令后，从该数据库中获取会议记录作为待处理文本。

[0091] 下面对上述步骤204即“从待处理文本中提取关键文本单元”进行详细描述。

[0092] 本步骤中涉及的关键文本单元可以是段落、关键句、关键短语等文本单元。作为其中一种较为优选的方式，后续实施例中均以关键句作为关键文本单元为例进行描述。从待处理文本中提取的关键文本单元可以是一个，也可以是多个。例如执行本步骤后从待处理文本中提取出m个关键句，m为正整数。

[0093] 本步骤中提取关键文本单元的方式可以采用但不限于以下三种方式：

[0094] 第一种方式：利用深度学习模型提取关键文本单元。即将待处理文本输入关键文本提取模型，获取关键文本提取模型从待处理文本中提取的关键文本单元。

[0095] 其中，上述关键文本提取模型是基于第一深度学习模型预先训练得到的。关键文

本提取模型的结构可以如图3中所示,包括第一编码网络(Encoder)和分类网络。

[0096] 第一编码网络对待处理文本进行编码处理,得到待处理文本中各文本单元的特征表示。

[0097] 该第一编码网络可以采用Transformer网络实现,例如可以基于预训练语言模型实现,预训练语言模型可以采用诸如BERT(BidirectionalEncoderRepresentationfromTransformers,基于转换的双向编码表示)模型、GPT(GenerativePre-Training,生成式预训练)模型、XLNet(一种通过排列语言模型实现双向上下文信息的自回归模型)等。

[0098] 其中,第一编码网络可以首先对待处理文本中的各Token进行Embedding(嵌入)处理,然后基于Embedding的结果进行编码,得到各Token的特征表示。然后再利用各文本单元中所包含各Token的特征表示分别得到各文本单元的特征表示。例如图3中所示,待处理文本包括n个句子,对于各句子,可以将一个句子中各Token的特征表示进行拼接得到句子级别的特征表示。各Token可以包括字符、起始符和分隔符等。

[0099] 上述Embedding处理可以包括:词Embedding、位置Embedding、句Embedding。词Embedding,即将各Token进行词向量编码,得到词向量表示。位置Embedding,即将各Token在待预测文本序列中的位置进行编码,得到位置的表示。例如可以依据各Token在文本序列中的位置依次编号为0、1、2、3、4、5和6等。句Embedding是将各Token所在的句子信息进行编码,得到所属句子的表示。例如将位于第1个句子的各Token编码为0,将位于第2个句子的各Token编码为1,等等。

[0100] 分类网络利用各文本单元的特征表示对各文本单元进行分类,得到各文本单元是否为关键文本单元的分类结果。分类网络实际上是一个二分类网络,例如利用各句子的特征表示对句子进行分类,分类结果为:关键句或非关键句。经过分类网络对各句子进行分类后,就能够得到待处理文本中的关键句的信息,即哪个或哪些句子是关键句,例如图3中所示,输出关键句i,j,…,k等共m个关键句。

[0101] 第二种方式:采用无监督的方式进行关键文本单元的提取,即利用预设的特征规则从待处理文本中提取关键文本单元。

[0102] 作为其中一种可实现的方式,可以基于预设的特征规则对待处理文本中的各句子分别进行打分,得到各句子的评分值,将评分值满足预设要求的句子作为关键句。

[0103] 例如,可以统计句子中包含各词语的TF(termfrequency,词频)-IDF(inversedocument frequency,逆文档率),基于各词语的TF-IDF对该句子进行打分。再例如,可以采用TextRank(文本排序)算法基于句子之间的相似程度来对各句子进行打分,其中TextRank算法将文本中的句子类比成PageRank算法中的网页,构建句子之间的图关系,通过与PageRank类似的迭代计算得到句子的重要度排名。鉴于TextRank算法为目前已有的算法,在此不做详述。

[0104] 第三种方式:结合深度学习模型和无监督方式进行关键文本单元的提取,即将上述第一种方式和第二种方式进行结合,既利用了深度学习模型对关键信息的识别能力,又引入了无监督方法对关键信息进行增强。

[0105] 作为其中一种可实现的方式,可以将待处理文本输入关键文本提取模型,获取关键文本提取模型从待处理文本中提取的第一关键文本单元,利用预设的特征规则从待处理文本中提取第二关键文本单元,将第一关键文本单元和第二关键文本单元进行融合,得到

关键文本单元。

[0106] 也就是说，将上述第一种方式提取的关键文本单元作为第一关键文本单元，将上述第二种方式提取的关键文本单元作为第二关键文本单元。在第一关键文本单元和第二关键文本单元进行融合时，可以将两种方式提取的关键文本单元取交集或者并集等，得到最终的关键文本单元。例如，第一种方式获取了四个关键句：句3、句20、句32，第二种方式获取了三个关键句：句5、句20、句32，则可以进行取并集的处理，得到句3、句5、句20和句32作为关键句。或者，也可以进行取交集的处理，得到句20和句32作为关键句。

[0107] 上面已经提及，第一种方式和第三种方式中涉及的关键文本提取模型是基于第一深度学习模型预先训练得到的，模型结构如图3中所示。下面对关键文本提取模型的训练过程进行介绍。如图4中所示，关键文本提取模型可以采用如下步骤训练得到：

[0108] 步骤402：获取多个第一训练样本，第一训练样本包括第一文本样本以及第一文本样本被标注的关键文本单元标签。

[0109] 在训练文本提取模型时，可以选取一些文本作为第一文本样本。在对第一文本样本进行关键文本单元标签的标注时，可以采用人工方式进行标注。由于人工标注的方式效率较低，可以采用一些方式自动确定第一文本样本的关键文本单元的信息。

[0110] 作为其中一种可实现的方式，可以将已知摘要的文本作为第一文本样本，即将一些具备摘要的文本作为第一文本样本。然后通过计算第一文本样本中各文本单元与第一文本样本的摘要之间的相似度来确定关键文本单元。例如通过GreedySearch(贪婪搜索)的方式，从第一文本样本中搜索出与摘要的相似度最高的m个句子作为关键句，m为预设的正整数。

[0111] 作为另一种可实现的方式，可以采用无监督的方式进行关键文本单元的提取，例如基于预设的特征规则对待处理文本中的各句子分别进行打分，得到各句子的评分值，将评分值满足预设要求的句子作为关键句。例如，可以统计句子中包含各词语的TF-IDF，基于各词语的TF-IDF对该句子进行打分。再例如，可以采用TextRank算法基于句子之间的相似程度来对各句子进行打分。

[0112] 步骤404：利用多个第一训练样本训练关键文本提取模型，其中关键文本提取模型包括第一编码网络和分类网络，训练的目标包括：最小化分类网络的分类结果与第一文本样本被标注的关键文本单元标签之间的差异。

[0113] 关键文本提取模型的结构同样如图3中所示，包括第一编码网络和分类网络。将第一训练样本中的第一文本样本作为关键文本提取模型的输入，将第一文本样本被标注的关键文本单元标签作为关键文本提取模型的目标输出。

[0114] 第一编码网络对第一文本样本进行编码处理，得到第一文本样本中各文本单元的特征表示。

[0115] 其中，第一编码网络可以首先对第一文本样本中的各Token进行Embedding处理，然后基于Embedding的结果进行编码，得到各Token的特征表示。然后再利用各文本单元中所包含各Token的特征表示分别得到各文本单元的特征表示。例如，第一文本样本包括n个句子，对于各句子，可以将一个句子中各Token的特征表示进行拼接得到句子级别的特征表示。各Token可以包括字符、起始符和分隔符等。

[0116] 分类网络利用各文本单元的特征表示对各文本单元进行分类，得到各文本单元是

否为关键文本单元的分类结果。分类网络实际上是一个二分类网络，例如利用各句子的特征表示对句子进行分类，分类结果为：关键句或非关键句。经过分类网络对各句子进行分类后，就能够得到第一文本样本中的关键句的信息。

[0117] 在训练关键文本提取模型时采用的训练的目标包括：最小化分类网络的分类结果与第一文本样本被标注的关键文本单元标签之间的差异。可以依据上述训练目标构造损失函数，在每一轮迭代中利用损失函数的取值，采用诸如梯度下降等方式更新模型参数，直至满足预设的训练结束条件。其中训练结束条件可以包括诸如损失函数的取值小于或等于预设的损失函数阈值，迭代次数达到预设的次数阈值等。

[0118] 继续参见图2，下面结合实施例对图2中的步骤206即“将待处理文本和关键文本单元的信息输入摘要生成模型，获取摘要生成模型利用待处理文本和关键文本单元的信息生成的待处理文本的摘要”进行详细描述。

[0119] 本步骤中摘要生成模型是基于关键文本单元的信息生成待处理文本的摘要，其结构可以如图5中所示，包括第二编码网络(Encoder)和解码网络(Decoder)，即采用了一种Encoder-Decoder框架来实现序列到序列的预测。

[0120] 第二编码网络利用关键文本单元的信息对待处理文本进行编码处理，得到待处理文本中各元素Token的特征表示。

[0121] 具体地，第二编码网络对待处理文本进行Embedding处理后，得到待处理文本中各Token的嵌入特征。然后利用关键文本单元的信息对各Token的嵌入特征进行注意力机制的处理，得到各Token的特征表示。其中注意力机制的处理可以包括：对属于关键文本单元的各Token进行全局注意力(GlobalAttention)处理，对不属于关键文本单元的各Token进行局部注意力(LocalAttention)处理。

[0122] 上述Embedding处理可以包括：词Embedding、位置Embedding、句Embedding。词Embedding，即将各Token进行词向量编码，得到词向量表示。位置Embedding，即将各Token在待预测文本序列中的位置进行编码，得到位置的表示。例如可以依据各Token在文本序列中的位置依次编号为0、1、2、3、4、5和6等。句Embedding是将各Token所在的句子信息进行编码，得到所属句子的表示。例如将位于第1个句子的各Token编码为0，将位于第2个句子的各Token编码为1，等等。需要说明的是，上述Embedding处理也可以仅包括词Embedding和位置Embedding，也可以仅包括词Embedding和句Embedding。

[0123] 另外，传统的自注意力处理是将所有的Token均进行全局注意力处理，考虑所有隐状态，将他们合并成一个注意力矩阵，然后乘以权重矩阵。输入量的增加也会带来注意力矩阵大小的增加，因此会产生大量的计算量并受到内存的限制而通常将输入序列的长度限制于512个字符，无法实现长文本的摘要提取。

[0124] 本申请实施例中在对各Token的嵌入特征进行注意力机制的处理时，不再采用传统的自注意力处理，而是仅针对属于关键文本单元的各Token进行全局注意力处理，对属于关键文本单元的各Token进行注意力处理时利用待处理文本中所有Token的嵌入特征，即针对属于关键文本单元的Token，计算待处理文本中所有Token对该属于关键文本单元的Token的注意力信息。而针对不属于关键文本单元的其他各Token进行局部注意力处理，对不属于关键文本单元的各Token进行注意力处理时利用距离该Token预设窗口距离内的各Token的嵌入特征，即针对不属于关键文本单元的Token，仅计算待处理文本中距离该Token

预设窗口距离内的各Token对该不属于关键文本单元的Token的注意力信息。这种方式能够帮助模型获得关注关键信息的能力，降低模型受到关键信息稀疏的长文本中噪音的影响。并且相比较传统对所有Token都进行全局注意力的方式，大大降低了计算开销，为摘要生成服务的应用落地提供了强有力的支持。

[0125] 所谓全局注意力处理指的是计算token的特征表示时，需要考虑输入序列(即整个待处理文本)的所有Token的隐状态。局部注意力处理指的是计算token的特征表示时，仅需要考虑输入序列隐状态的一个子集，该子集通常是以当前时间步位置为中心的预设长度的窗口内对应的子序列的隐状态。窗口的预设长度可以采用经验值或实验值，通常会取512个字符之内的一个正整数。

[0126] 假设步骤204提取出关键句i,j,…,k，对于关键句i,j,…,k中各Token进行全局注意力处理，对于其他句子中的各Token进行局部注意力处理，最终得到待处理文本中各Token的特征表示。

[0127] 解码网络利用待处理文本中各Token的特征表示进行解码处理，生成待处理文本的摘要。

[0128] 下面对摘要生成模型的训练过程进行介绍。如图6中所示，摘要生成模型可以采用如下步骤训练得到：

[0129] 步骤602：获取多个第二训练样本，第二训练样本包括第二文本样本、第二文本样本的关键文本单元的信息以及第二文本样本的摘要样本。

[0130] 在训练摘要生成模型时，可以选取一些已具有摘要的文本作为第二文本样本，具有的摘要作为摘要样本。也可以选取一些文本作为第二文本样本，然后人工为第二文本样本生成摘要作为摘要样本。

[0131] 第二文本样本的关键文本单元的信息可以采用人工方式进行标注。由于人工标注的方式效率较低，可以采用一些方式自动确定第二文本样本的关键文本单元的信息。

[0132] 作为其中一种可实现的方式，可以通过计算第二文本样本中各文本单元与第二文本样本的摘要之间的相似度来确定关键文本单元。例如通过GreedySearch的方式，从第二文本样本中搜索出与摘要的相似度最高的m个句子作为关键句，m为预设的正整数。

[0133] 作为另一种可实现的方式，可以利用已经训练得到的关键文本提取模型从第二文本样本中提取关键文本单元。

[0134] 作为再一种可实现的方式，可以采用无监督的方式从第二文本样本中提取关键文本单元，例如基于预设的特征规则对待处理文本中的各句子分别进行打分，得到各句子的评分值，将评分值满足预设要求的句子作为关键句。例如，可以统计句子中包含各词语的TF-IDF，基于各词语的TF-IDF对该句子进行打分。再例如，可以采用TextRank算法基于句子之间的相似程度来对各句子进行打分。

[0135] 作为再一种可实现的方式，可以将上述至少两种方式提取的关键文本单元进行融合，得到最终提取关键文本单元的结果。其中的融合处理可以是取交集或者并集等。

[0136] 步骤604：利用多个第二训练样本训练摘要生成模型，其中摘要生成模型包括第二编码网络和解码网络；训练的目标包括：最小化解码网络生成的第二文本样本的摘要与第二文本的摘要样本之间的差异。

[0137] 摘要生成模型的结构可以参见图5中所示，第二编码网络利用第二文本样本的关

键文本单元的信息对第二文本样本进行编码处理，得到第二文本样本中各Token的特征表示。

[0138] 具体地，第二编码网络获取对第二文本样本进行嵌入处理后，得到的第二文本样本中各Token的嵌入特征；利用第二文本样本的关键文本单元的信息对各Token的嵌入特征进行注意力机制的处理，得到各Token的特征表示。其中注意力机制的处理包括：对属于关键文本单元的各Token进行全局注意力处理，即对属于关键文本单元的各Token进行注意力处理时利用第二文本样本中所有Token的嵌入特征；对不属于关键文本单元的各Token进行局部注意力处理，即对不属于关键文本单元的各Token进行注意力处理时利用第二文本样本中距离该Token预设窗口距离内的各Token的嵌入特征。

[0139] 解码网络利用第二文本样本中各Token的特征表示进行解码处理，生成第二文本样本的摘要。

[0140] 在训练摘要生成模型时采用的训练的目标包括：最小化解码网络生成的摘要与第二文本样本的摘要样本之间的差异。可以依据上述训练目标构造损失函数，在每一轮迭代中利用损失函数的取值，采用诸如梯度下降等方式更新模型参数，直至满足预设的训练结束条件。其中训练结束条件可以包括诸如损失函数的取值小于或等于预设的损失函数阈值，迭代次数达到预设的次数阈值等。

[0141] 本申请实施例提供的上述方法可以应用于多种应用场景，包括但不限于：

[0142] 应用场景1、在线会议记录的摘要提取

[0143] 随着全球市场化持续增长，在线会议服务越来越多地被使用，用户可以使用在线会议软件进行网络会议、参与课堂教学、足不出户远程面试、创办讲座和论坛等。用户还可以一边进行视频会议，一边进行信息的记录和传输。

[0144] 当在线会议服务对会议过程中的语音进行语音识别后生成在线会议的会议记录。该在线会议记录中除了语音识别的结果之外，还可以包括用户在会议界面中输入的文字信息、文档信息等等。为了方便用户了解会议的主要内容，需要针对会议记录生成会议摘要。由于会议、讲座、面试等场景存在持续时间久、关键信息分布稀疏、口语化程度较高的特点，因此如图7中所示，可以采用本申请实施例所提供的方法从在线会议服务的服务器端获取会议记录作为待处理文本。然后一方面利用关键文本提取模型从待处理文本中提取关键句，另一方面利用无监督的方式即采用预设的特征规则从待处理文本中提取关键句，将两方面提取的关键句进行融合，得到待处理文本的关键句。再将待处理文本和关键句的信息输入摘要生成模型，得到待处理文本即会议记录的摘要。

[0145] 会议记录的摘要可以自动发送给该在线会议的各参与者。也可以将该会议记录与摘要关联存储至数据库，以便后续响应于用户的请求将该会议记录的摘要发送给用户，等等。

[0146] 通过本申请实施例提供的方式，在针对会议记录生成摘要时引入了从会议记录中提取的关键句的信息，从而为摘要的生成提供指导，使得摘要生成模型能够聚焦会议记录中的关键内容，降低噪声的影响。经过实验论证，采用本申请实施例提供的方式针对诸如在线会议这种持续时间久、关键信息分布稀疏且口语化的长文档上，能够生成准确的摘要，表现出了较优的性能。

[0147] 在进行关键句的提取时，结合有监督和无监督两种方式，对提取的关键句信息进

行增强。另外，摘要生成模型在利用关键句的信息对待处理文本生成摘要时，仅对关键句中的各Token进行全局注意力处理，而对于其他各Token进行局部注意力处理，从而降低了计算开销，使得模型能够顺利针对长文本实现摘要生成。

[0148] 应用场景2、论文、新闻等的摘要提取

[0149] 随着互联网的广泛使用，互联网上的文档数据呈爆炸式的增长，出现了大量的论文、新闻等文档，并在一些数据平台上提供这些诸如论文、新闻等文档的阅读。为了方便用户快速了解这些文档的内容，通常需要针对这些文档生成摘要并在平台上提供摘要供用户阅读和参考。在这种应用场景下，也可以采用本申请实施例中提供的方式，将诸如论文、新闻等文档作为待处理文本生成摘要。具体过程在此不做赘述。

[0150] 上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中，多任务处理和并行处理也是可以的或者可能是有利的。

[0151] 根据另一方面的实施例，提供了一种文本摘要生成装置。图8示出根据一个实施例的该文本摘要生成装置的示意性框图。如图8所示，该装置800包括：文本获取单元801、关键提取单元802和摘要生成单元803，其中各组成单元的主要功能如下：

[0152] 文本获取单元801，被配置为获取待处理文本。

[0153] 关键提取单元802，被配置为从待处理文本中提取关键文本单元。

[0154] 摘要生成单元803，被配置为将待处理文本和关键文本单元的信息输入摘要生成模型，获取摘要生成模型利用待处理文本和关键文本单元的信息生成的待处理文本的摘要；其中摘要生成模型是基于第二深度学习模型预先训练得到的。

[0155] 作为其中一种可实现的方式，关键提取单元802可以具体被配置为：将待处理文本输入关键文本提取模型，获取关键文本提取模型从待处理文本中提取的关键文本单元，其中关键文本提取模型基于第一深度学习模型预先训练得到。

[0156] 作为另一种可实现的方式，关键提取单元802可以具体被配置为：利用预设的特征规则从待处理文本中提取关键文本单元。

[0157] 作为再一种可实现的方式，关键提取单元802可以具体被配置为：将待处理文本输入关键文本提取模型，获取关键文本提取模型从待处理文本中提取的第一关键文本单元，利用预设的特征规则从待处理文本中提取第二关键文本单元，将第一关键文本单元和第二关键文本单元进行融合，得到关键文本单元。

[0158] 作为其中一种可实现的方式，关键文本提取模型可以包括第一编码网络和分类网络。

[0159] 第一编码网络对待处理文本进行编码处理，得到待处理文本中各文本单元的特征表示。

[0160] 其中，第一编码网络可以首先对待处理文本中的各Token进行Embedding处理，然后基于Embedding的结果进行编码，得到各Token的特征表示。然后再利用各文本单元中所包含各Token的特征表示分别得到各文本单元的特征表示。例如，待处理文本包括n个句子，对于各句子，可以将一个句子中各Token的特征表示进行拼接得到句子级别的特征表示。各

Token可以包括字符、起始符和分隔符等。

[0161] 分类网络利用各文本单元的特征表示对各文本单元进行分类，得到各文本单元是否为关键文本单元的分类结果。

[0162] 上述关键文本提取模型的结构和原理可以参见上述方法实施例中的相关记载，在此不做赘述。

[0163] 作为其中一种可实现的方式，摘要生成模型可以包括第二编码网络和解码网络。

[0164] 第二编码网络利用关键文本单元的信息对待处理文本进行编码处理，得到待处理文本中各元素Token的特征表示。

[0165] 其中，第二编码网络对待处理文本进行嵌入处理后，得到待处理文本中各Token的嵌入特征；利用关键文本单元的信息对各Token的嵌入特征进行注意力机制的处理，得到各Token的特征表示，其中注意力机制的处理包括：对属于关键文本单元的各Token进行注意力处理时利用待处理文本中所有Token的嵌入特征，对不属于关键文本单元的各Token进行注意力处理时利用距离该Token预设窗口距离内的各Token的嵌入特征。

[0166] 解码网络利用待处理文本中各Token的特征表示进行解码处理，生成待处理文本的摘要。

[0167] 上述摘要生成模型的结构和原理可以参见上述方法实施例中的相关记载，在此不做赘述。

[0168] 图9示出根据一个实施例的训练关键文本提取模型的装置的示意性框图。如图9所示，该装置900包括：第一样本获取单元901和第一模型训练单元902，其中各组成单元的主要功能如下：

[0169] 第一样本获取单元901，被配置为获取多个第一训练样本，第一训练样本包括第一文本样本以及第一文本样本被标注的关键文本单元标签。

[0170] 在训练文本提取模型时，可以选取一些文本作为第一文本样本。在对第一文本样本进行关键文本单元标签的标注时，可以采用人工方式进行标注。由于人工标注的方式效率较低，可以采用一些方式自动确定第一文本样本的关键文本单元的信息。

[0171] 作为其中一种可实现的方式，可以将已知摘要的文本作为第一文本样本，即将一些具备摘要的文本作为第一文本样本。然后通过计算第一文本样本中各文本单元与第一文本样本的摘要之间的相似度来确定关键文本单元。例如通过GreedySearch的方式，从第一文本样本中搜索出与摘要的相似度最高的m个句子作为关键句，m为预设的正整数。

[0172] 作为另一种可实现的方式，可以采用无监督的方式进行关键文本单元的提取，例如基于预设的特征规则对待处理文本中的各句子分别进行打分，得到各句子的评分值，将评分值满足预设要求的句子作为关键句。例如，可以统计句子中包含各词语的TF-IDF，基于各词语的TF-IDF对该句子进行打分。再例如，可以采用TextRank算法基于句子之间的相似程度来对各句子进行打分。

[0173] 第一模型训练单元902，被配置为利用多个第一训练样本训练关键文本提取模型，其中关键文本提取模型包括第一编码网络和分类网络；第一编码网络对第一文本样本进行编码处理，得到第一文本样本中各文本单元的特征表示；分类网络利用各文本单元的特征表示对各文本单元进行分类，得到各文本单元是否为关键文本单元的分类结果；训练的目标包括：最小化分类网络的分类结果与第一文本样本被标注的关键文本单元标签之间的差

异。

[0174] 其中,第一编码网络可以首先对第一文本样本中的各Token进行Embedding处理,然后基于Embedding的结果进行编码,得到各Token的特征表示。然后再利用各文本单元中所包含各Token的特征表示分别得到各文本单元的特征表示。例如,第一文本样本包括n个句子,对于各句子,可以将一个句子中各Token的特征表示进行拼接得到句子级别的特征表示。各Token可以包括字符、起始符和分隔符等。

[0175] 分类网络实际上是一个二分类网络,例如利用各句子的特征表示对句子进行分类,分类结果为:关键句或非关键句。经过分类网络对各句子进行分类后,就能够得到第一文本样本中的关键句的信息。

[0176] 第一模型训练单元902在训练关键文本提取模型时采用的训练的目标包括:最小化分类网络的分类结果与第一文本样本被标注的关键文本单元标签之间的差异。可以依据上述训练目标构造损失函数,在每一轮迭代中利用损失函数的取值,采用诸如梯度下降等方式更新模型参数,直至满足预设的训练结束条件。其中训练结束条件可以包括诸如损失函数的取值小于或等于预设的损失函数阈值,迭代次数达到预设的次数阈值等。

[0177] 图10示出根据一个实施例的训练摘要生成模型的装置的示意性框图。如图9所示,该装置1000包括:第二样本获取单元1001和第二模型训练单元1002。其中各组成单元的主要功能如下:

[0178] 第二样本获取单元1001,被配置为获取多个第二训练样本,第二训练样本包括第二文本样本、第二文本样本的关键文本单元的信息以及第二文本样本的摘要样本。

[0179] 在训练摘要生成模型时,可以选取一些已具有摘要的文本作为第二文本样本,具有的摘要作为摘要样本。也可以选取一些文本作文作为第二文本样本,然后人工为第二文本样本生成摘要作为摘要样本。

[0180] 第二文本样本的关键文本单元的信息可以采用人工方式进行标注。由于人工标注的方式效率较低,可以采用一些方式自动确定第二文本样本的关键文本单元的信息。

[0181] 作为其中一种可实现的方式,可以通过计算第二文本样本中各文本单元与第二文本样本的摘要之间的相似度来确定关键文本单元。例如通过GreedySearch的方式,从第二文本样本中搜索出与摘要的相似度最高的m个句子作为关键句,m为预设的正整数。

[0182] 作为另一种可实现的方式,可以利用已经训练得到的关键文本提取模型从第二文本样本中提取关键文本单元。

[0183] 作为再一种可实现的方式,可以采用无监督的方式从第二文本样本中提取关键文本单元,例如基于预设的特征规则对待处理文本中的各句子分别进行打分,得到各句子的评分值,将评分值满足预设要求的句子作为关键句。例如,可以统计句子中包含各词语的TF-IDF,基于各词语的TF-IDF对该句子进行打分。再例如,可以采用TextRank算法基于句子之间的相似程度来对各句子进行打分。

[0184] 作为再一种可实现的方式,可以将上述至少两种方式提取的关键文本单元进行融合,得到最终提取关键文本单元的结果。其中的融合处理可以是取交集或者并集等。

[0185] 第二模型训练单元1002,被配置为利用多个第二训练样本训练摘要生成模型,其中摘要生成模型包括第二编码网络和解码网络;第二编码网络利用第二文本样本的关键文本单元的信息对第二文本样本进行编码处理,得到第二文本样本中各Token的特征表示;解

码网络利用第二文本样本中各Token的特征表示进行解码处理，生成第二文本样本的摘要；训练的目标包括：最小化解码网络生成的第二文本样本的摘要与第二文本的摘要样本之间的差异。

[0186] 其中，第二编码网络获取对第二文本样本进行嵌入处理后，得到的第二文本样本中各Token的嵌入特征；利用第二文本样本的关键文本单元的信息对各Token的嵌入特征进行注意力机制的处理，得到各Token的特征表示，其中注意力机制的处理包括：对属于关键文本单元的各Token进行注意力处理时利用第二文本样本中所有Token的嵌入特征，对不属于关键文本单元的各Token进行注意力处理时利用距离该Token预设窗口距离内的各Token的嵌入特征。

[0187] 第二模型训练单元1002在训练摘要生成模型时采用的训练的目标包括：最小化解码网络生成的摘要与第二文本样本的摘要样本之间的差异。可以依据上述训练目标构造损失函数，在每一轮迭代中利用损失函数的取值，采用诸如梯度下降等方式更新模型参数，直至满足预设的训练结束条件。其中训练结束条件可以包括诸如损失函数的取值小于或等于预设的损失函数阈值，迭代次数达到预设的次数阈值等。

[0188] 本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于装置实施例而言，由于其基本相似于方法实施例，所以描述得比较简单，相关之处参见方法实施例的部分说明即可。以上所描述的装置实施例仅仅是示意性的，其中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模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性劳动的情况下，即可以理解并实施。

[0189] 需要说明的是，本申请所涉及的用户信息(包括但不限于用户设备信息、用户个人信息等)和数据(包括但不限于用于分析的数据、存储的数据、展示的数据等)，均为经用户授权或者经过各方充分授权的信息和数据，并且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准，并提供有相应的操作入口，供用户选择授权或者拒绝。

[0190] 另外，本申请实施例还提供了一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现前述方法实施例中任一项所述的方法的步骤。

[0191] 以及一种电子设备，包括：

[0192] 一个或多个处理器；以及

[0193] 与所述一个或多个处理器关联的存储器，所述存储器用于存储程序指令，所述程序指令在被所述一个或多个处理器读取执行时，执行前述方法实施例中任一项所述的方法的步骤。

[0194] 本申请还提供了一种计算机程序产品，包括计算机程序，该计算机程序在被处理器执行时实现前述方法实施例中任一项所述的方法的步骤。

[0195] 其中，图11示例性的展示出了电子设备的架构，具体可以包括处理器1110，视频显示适配器1111，磁盘驱动器1112，输入/输出接口1113，网络接口1114，以及存储器1120。上述处理器1110、视频显示适配器1111、磁盘驱动器1112、输入/输出接口1113、网络接口

1114,与存储器1120之间可以通过通信总线1130进行通信连接。

[0196] 其中,处理器1110可以采用通用的CPU、微处理器、应用专用集成电路(Application SpecificIntegratedCircuit,ASIC)、或者一个或多个集成电路等方式实现,用于执行相关程序,以实现本申请所提供的技术方案。

[0197] 存储器1120可以采用ROM(ReadOnlyMemory,只读存储器)、RAM(RandomAccess Memory,随机存取存储器)、静态存储设备,动态存储设备等形式实现。存储器1120可以存储用于控制电子设备1100运行的操作系统1121,用于控制电子设备1100的低级别操作的基本输入输出系统(BIOS)1122。另外,还可以存储网页浏览器1123,数据存储管理系统1124,以及文本摘要生成装置1125等等。上述文本摘要生成装置1125就可以是本申请实施例中具体实现前述各步骤操作的应用程序。总之,在通过软件或者固件来实现本申请所提供的技术方案时,相关的程序代码保存在存储器1120中,并由处理器1110来调用执行。

[0198] 输入/输出接口1113用于连接输入/输出模块,以实现信息输入及输出。输入输出/模块可以作为组件配置在设备中(图中未示出),也可以外接于设备以提供相应功能。其中输入设备可以包括键盘、鼠标、触摸屏、麦克风、各类传感器等,输出设备可以包括显示器、扬声器、振动器、指示灯等。

[0199] 网络接口1114用于连接通信模块(图中未示出),以实现本设备与其他设备的通信交互。其中通信模块可以通过有线方式(例如USB、网线等)实现通信,也可以通过无线方式(例如移动网络、WIFI、蓝牙等)实现通信。

[0200] 总线1130包括一通路,在设备的各个组件(例如处理器1110、视频显示适配器1111、磁盘驱动器1112、输入/输出接口1113、网络接口1114,与存储器1120)之间传输信息。

[0201] 需要说明的是,尽管上述设备仅示出了处理器1110、视频显示适配器1111、磁盘驱动器1112、输入/输出接口1113、网络接口1114,存储器1120,总线1130等,但是在具体实施过程中,该设备还可以包括实现正常运行所必需的其他组件。此外,本领域的技术人员可以理解的是,上述设备中也可以仅包含实现本申请方案所必需的组件,而不必包含图中所示的全部组件。

[0202] 通过以上的实施方式的描述可知,本领域的技术人员可以清楚地了解到本申请可借助软件加必需的通用硬件平台的方式来实现。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可以以计算机程序产品的形式体现出来,该计算机程序产品可以存储在存储介质中,如ROM/RAM、磁碟、光盘等,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,或者网络设备等)执行本申请各个实施例或者实施例的某些部分所述的方法。

[0203] 以上对本申请所提供的技术方案进行了详细介绍,本文中应用了具体个例对本申请的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本申请的方法及其核心思想;同时,对于本领域的一般技术人员,依据本申请的思想,在具体实施方式及应用范围上均会有改变之处。综上所述,本说明书内容不应理解为对本申请的限制。

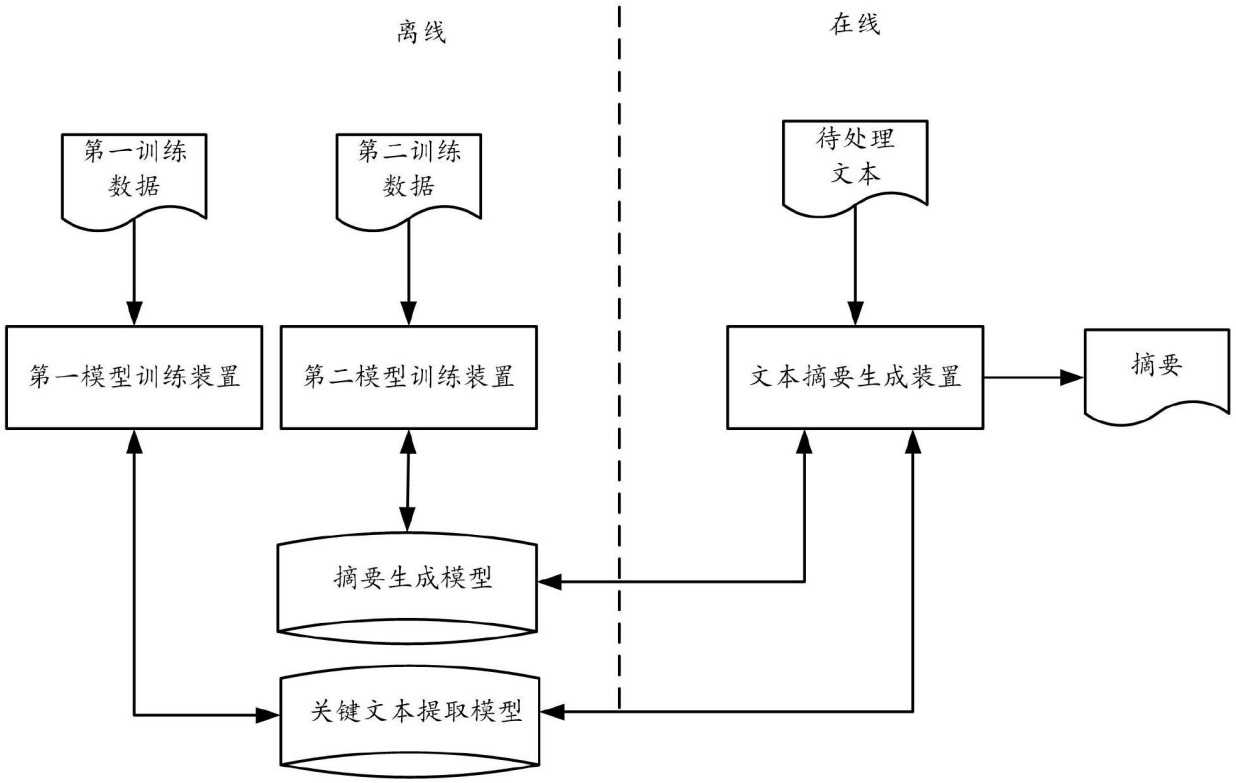

离线
在线
第一训练数据
第二训练数据
待处理文本
第一模型训练装置
第二模型训练装置
文本摘要生成装置
摘要
摘要生成模型
关键文本提取模型

图1

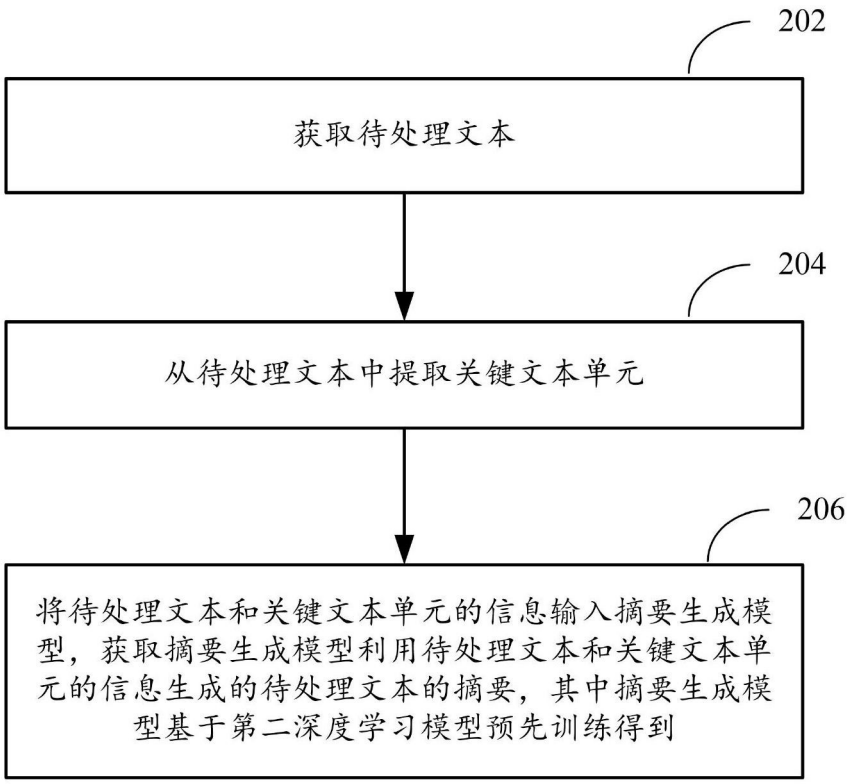

202
获取待处理文本
204
从待处理文本中提取关键文本单元
206
将待处理文本和关键文本单元的信息输入摘要生成模型，获取摘要生成模型利用待处理文本和关键文本单元的信息生成的待处理文本的摘要，其中摘要生成模型基于第二深度学习模型预先训练得到

图2

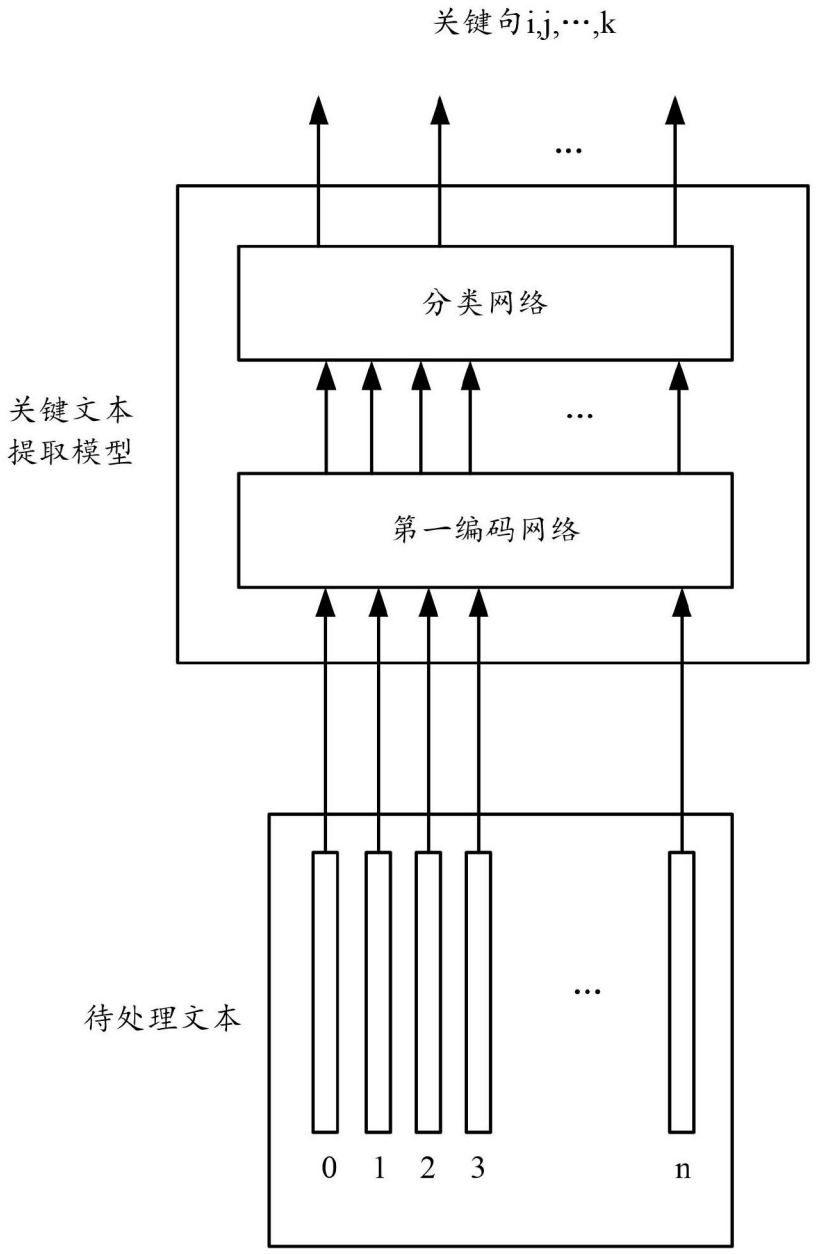


图3

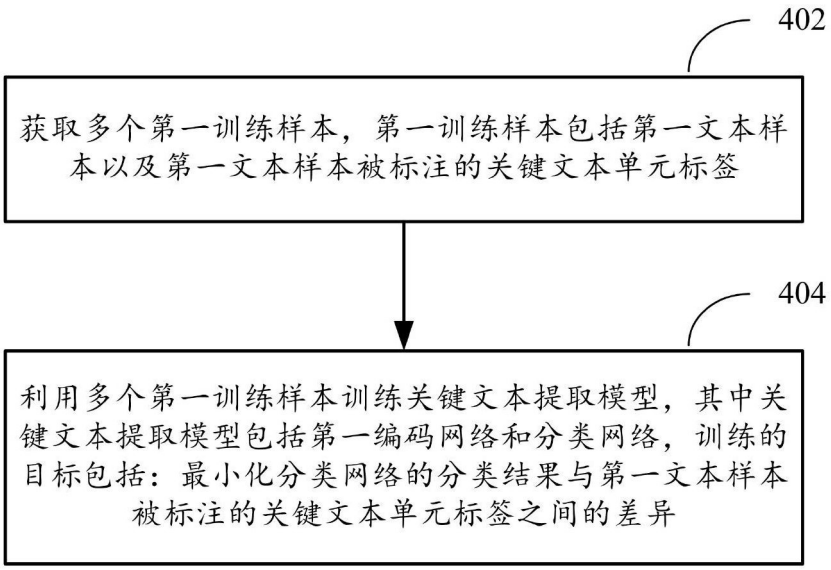

402
获取多个第一训练样本，第一训练样本包括第一文本样本以及第一文本样本被标注的关键文本单元标签
404
利用多个第一训练样本训练关键文本提取模型，其中关键文本提取模型包括第一编码网络和分类网络，训练的目标包括：最小化分类网络的分类结果与第一文本样本被标注的关键文本单元标签之间的差异

图4

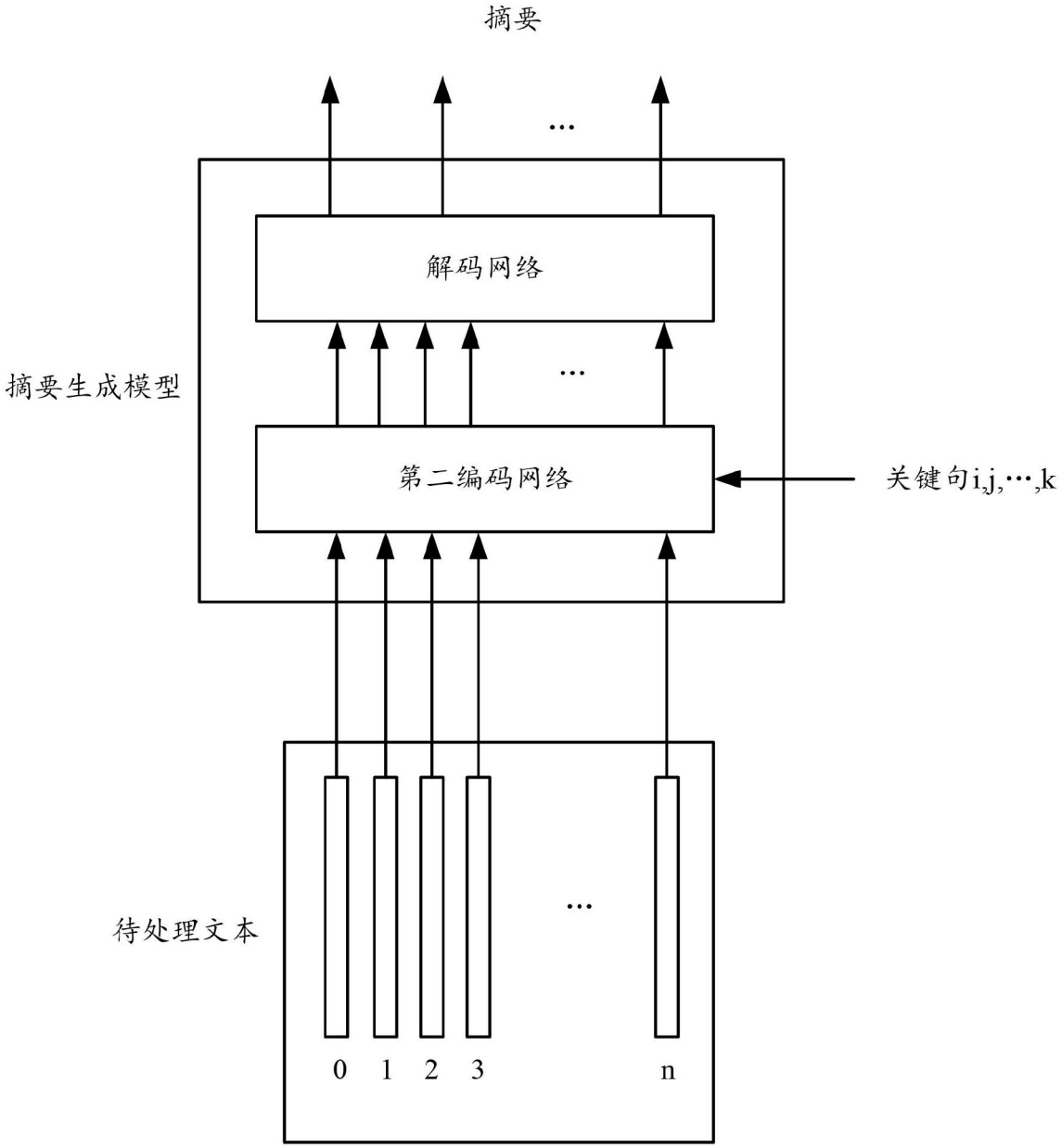


图5

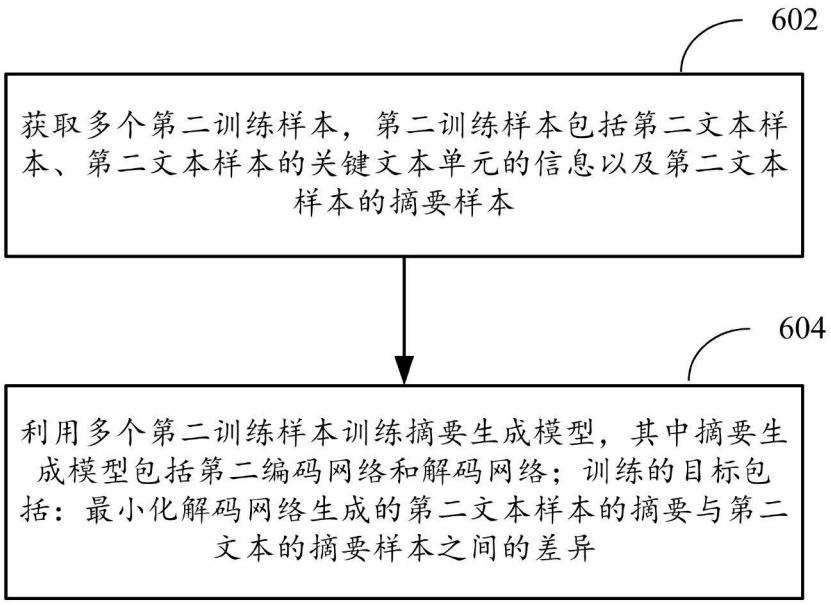

602
获取多个第二训练样本，第二训练样本包括第二文本样本、第二文本样本的关键文本单元的信息以及第二文本样本的摘要样本
604
利用多个第二训练样本训练摘要生成模型，其中摘要生成模型包括第二编码网络和解码网络；训练的目标包括：最小化解码网络生成的第二文本样本的摘要与第二文本的摘要样本之间的差异

图6

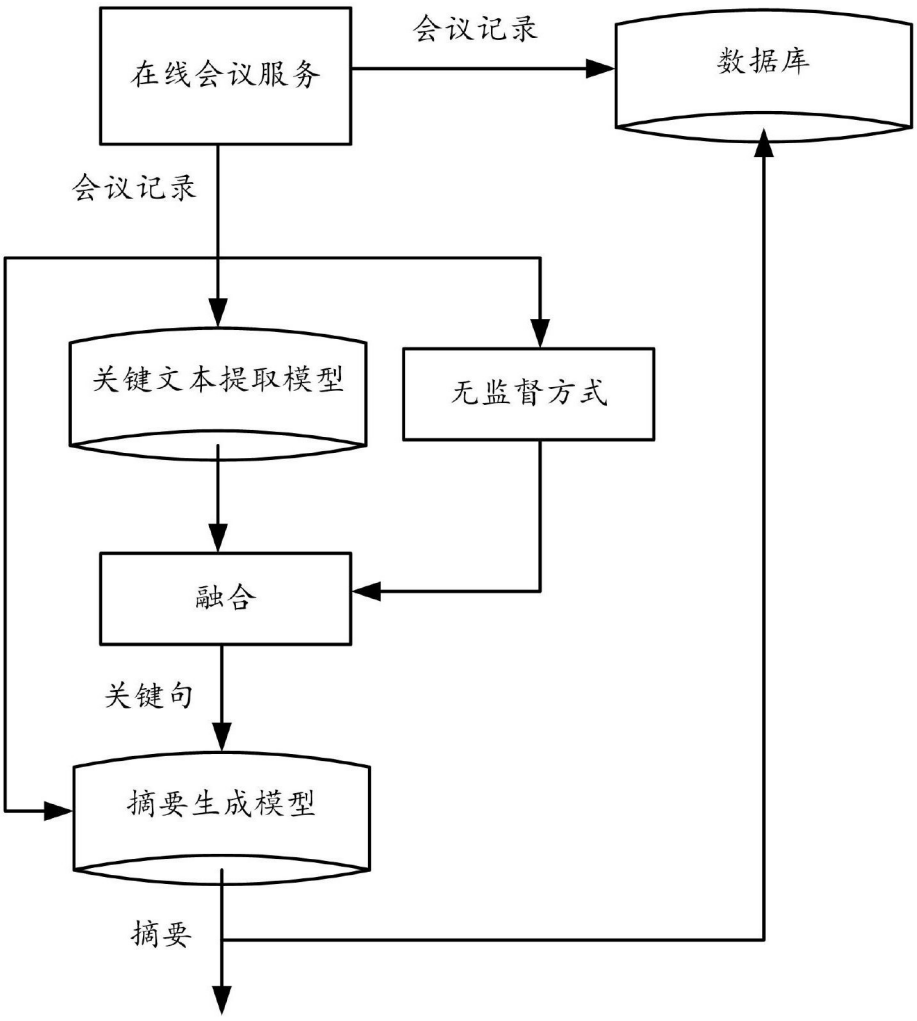


图7

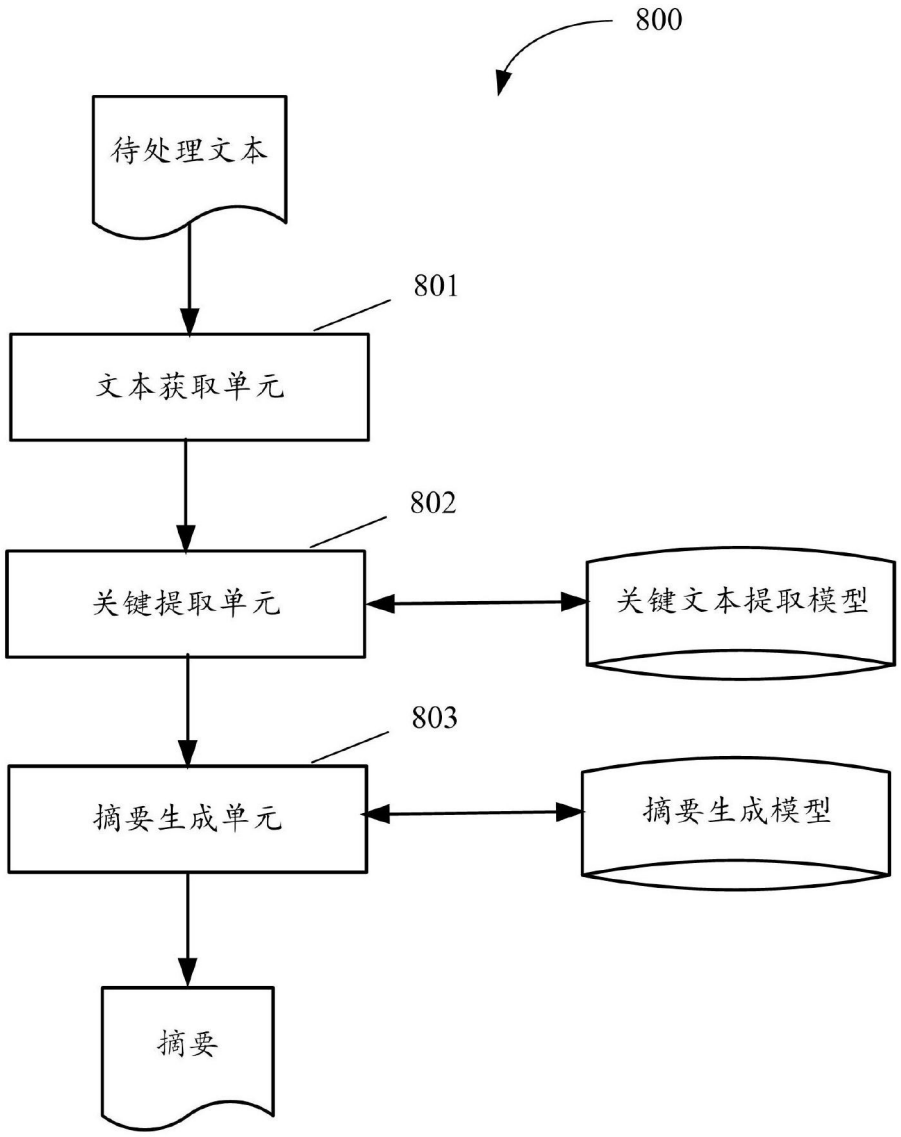

800
待处理文本
801
文本获取单元
802
关键提取单元
关键文本提取模型
803
摘要生成单元
摘要生成模型
摘要

图8

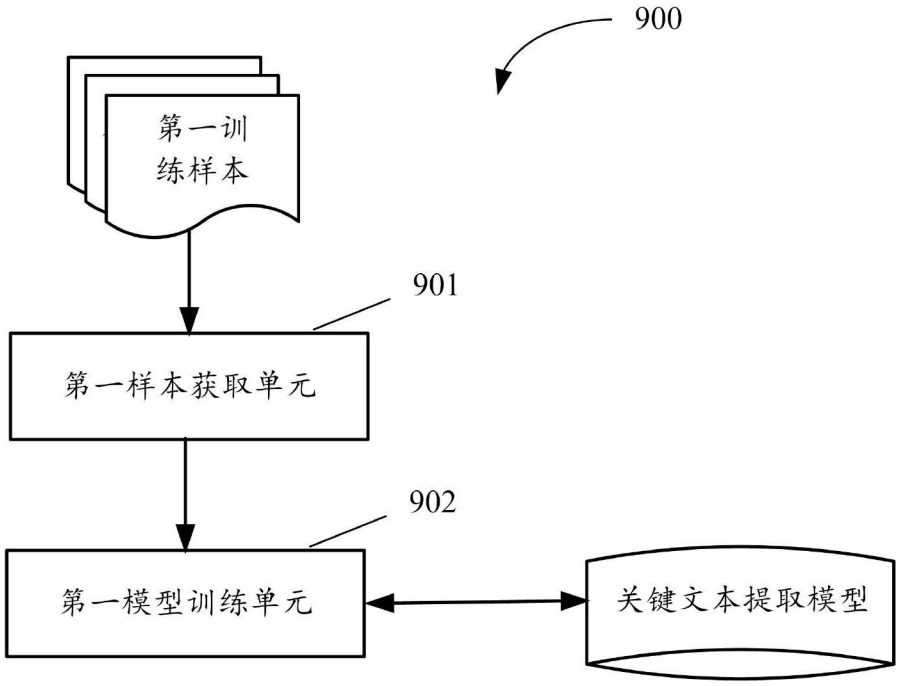

900
第一训练样本
901
第一样本获取单元
902
第一模型训练单元
关键文本提取模型

图9

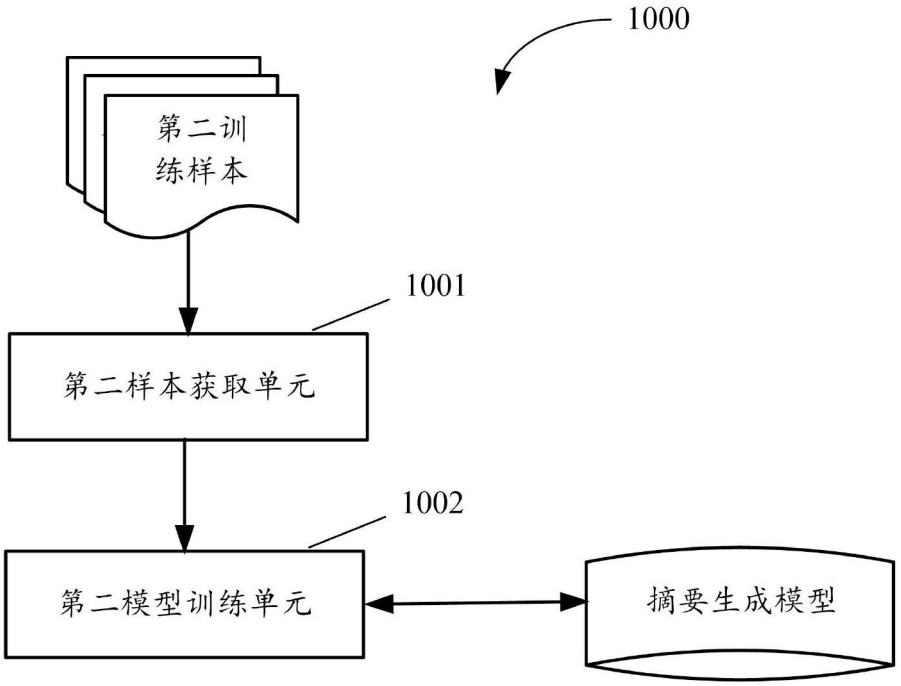

1000
第二训练样本
1001
第二样本获取单元
1002
第二模型训练单元
摘要生成模型

图10

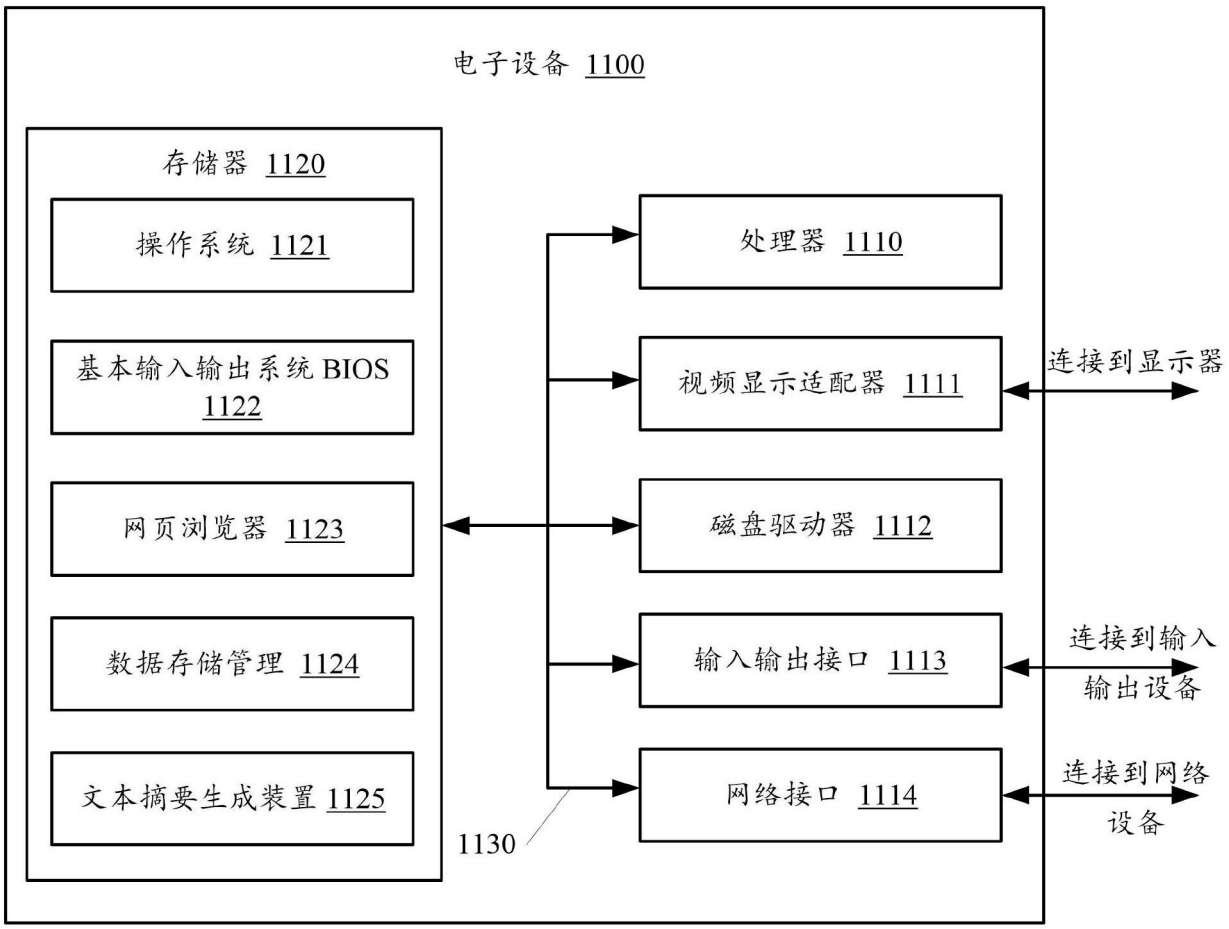

电子设备 1100
存储器 1120
操作系统 1121
基本输入输出系统 BIOS 1122
网页浏览器 1123
数据存储管理 1124
文本摘要生成装置 1125
1130
处理器 1110
视频显示适配器 1111
磁盘驱动器 1112
输入输出接口 1113
网络接口 1114
连接到显示器
连接到输入输出设备
连接到网络设备

图11