

文本生成方法、装置、电子设备和计算机可读介质

技术领域

本公开的实施例涉及计算机技术领域，具体涉及文本生成方法、装置、电子设备和计算机可读介质。

背景技术

随着计算机技术和深度学习技术的迅猛发展，在自然语言处理的各个研究方向都取得了突破性的进展，如机器翻译、文本生成(比如生成摘要)。现有的文本生成模型在解码过程中，往往使用了 Beam Search (集束搜索) 算法，用于贪心式地在较大的求解空间中求解，该算法是文本生成模型的解码模块中常用的一种算法。

然而，当采用上述方式生成文本时，经常会存在如下技术问题：

现有的集束搜索在解码过程中，生成文本主题的效率较低及准确度较低。

发明内容

本公开的内容部分用于以简要的形式介绍构思，这些构思将在后面的具体实施方式部分被详细描述。本公开的内容部分并不旨在标识要求保护的技术方案的关键特征或必要特征，也不旨在用于限制所要求的保护的技术方案的范围。本公开的一些实施例提出了文本生成方法、装置、电子设备和计算机可读介质，来解决以上背景技术部分提到的技术问题中的一项或多项。

第一方面，本公开的一些实施例提供了一种文本生成方法，该方法包括：基于待处理文本，确定解码词信息组集，其中，上述待处理文本用于描述指定物体；对上述解码词信息组集进行去重处理以生成候选词信息组集；从上述候选词信息组集中的每个候选词信息组选择满足目标条件的候选词信息作为目标词信息，得到目标词信息集；对于上述目标词信息集中的每个目标词信息，响应于确定上述目标词信

息满足收敛条件，基于预设词表，将上述目标词信息和上述目标词信息所对应的历史目标词信息序列进行拼接以生成目标文本。

可选地，基于待处理文本，确定解码词信息组集，包括：将上述待处理文本输入至文本编码器中以生成编码隐层向量；将上述编码隐层向量输入至解码器中以生成解码词信息组集。

可选地，解码词信息组集中的解码词信息包括：解码词和上述解码词对应的解码词概率值；以及上述对上述解码词信息组集进行去重处理以生成候选词信息组集，包括：对于上述解码词信息组集中的每个解码词信息组，根据上述解码词，将上述解码词信息组划分为重复解码词信息组和单一解码词信息组；从上述重复解码词信息组中选择满足预设条件的重复解码词信息作为目标重复解码词信息；将上述目标重复解码词信息和上述单一解码词信息组进行拼接以生成上述候选词信息组。

可选地，对于上述解码词信息组集中的每个解码词信息组，根据上述解码词，将上述解码词信息组划分为重复解码词信息组和单一解码词信息组，包括：对于上述解码词信息组集中的每个解码词信息组，响应于上述解码词信息组中存在包含该解码词信息的解码词的其他解码词信息，将该解码词信息和上述其他解码词信息放入重复解码词信息组，否则，将该解码词信息放入单一解码词信息组。

可选地，候选词信息组集中的候选词信息包括：候选词和上述候选词对应的候选词概率值；以及上述从上述候选词信息组集中的每个候选词信息组选择满足目标条件的候选词信息作为目标词信息，包括：对于上述候选词信息组集中的每个候选词信息组，按照上述候选词概率值由大到小的顺序从上述候选词信息组中选择至少一个候选词信息；将上述至少一个候选词信息中概率值最大的候选词信息放入历史目标词信息集，并确定上述概率值最大的候选词信息的至少一个初始目标词信息，并基于上述至少一个初始目标词信息生成对应该候选词信息的目标词信息。上述至少一个初始目标词信息中的初始目标词信息包括：初始目标词和对应上述初始目标词的初始目标词概率值。

可选地，基于上述至少一个初始目标词信息生成对应该候选词信

息的目标词信息，包括：响应于初始目标词概率值最大的初始目标词信息对应的初始目标词属于上述历史目标词信息集，确定上述初始目标词概率值最大的初始目标词信息与上述历史目标词信息集中对应的历史目标词信息之间的重复词概率差；响应于上述重复词概率差大于该候选词信息对应的其他初始目标词信息的初始目标词概率值，将初始目标词概率值最大的初始目标词信息设置为目标词信息，并将上述目标词信息放入上述历史目标词信息集。

可选地，基于上述至少一个初始目标词信息生成对应该候选词信息的目标词信息，包括：响应于初始目标词概率值最大的初始目标词信息对应的初始目标词属于上述历史目标词信息集，确定上述初始目标词概率值最大的初始目标词信息与上述历史目标词信息集中对应的历史目标词信息之间的重复词概率差；响应于上述重复词概率差小于等于该候选词信息对应的其他初始目标词信息的初始目标词概率值，将其他初始目标词信息中的初始目标词概率值最大的初始目标词信息设置为目标词信息，并将上述目标词信息放入上述历史目标词信息集。

可选地，基于上述至少一个初始目标词信息生成对应该候选词信息的目标词信息，包括：响应于初始目标词概率值最大的初始目标词信息对应的初始目标词不属于上述历史目标词信息集，将初始目标词概率值最大的初始目标词信息设置为目标词信息，并将上述目标词信息放入上述历史目标词信息集。

可选地，对于上述目标词信息集中的每个目标词信息，响应于确定上述目标词信息满足收敛条件，基于预设词表，将上述目标词信息和上述目标词信息所对应的历史目标词信息序列进行拼接以生成目标文本，包括：对于上述目标词信息集中的每个目标词信息，响应于确定上述目标词信息满足预设收敛条件，将上述目标词信息与上述目标词信息所对应的历史目标词信息序列进行拼接以生成拼接文本；从上述预设词表选择与上述拼接文本相匹配的词作为连接词；对上述连接词和上述拼接文本进行组合以生成目标文本。

第二方面，本公开的一些实施例提供了一种文本生成装置，装置包括：确定单元，被配置成基于待处理文本，确定解码词信息组集，

其中，上述待处理文本用于描述指定物体；去重单元，被配置成对上述解码词信息组集进行去重处理以生成候选词信息组集；选择单元，被配置成从上述候选词信息组集中的每个候选词信息组选择满足目标条件的候选词信息作为目标词信息，得到目标词信息集；拼接单元，被配置成对于上述目标词信息集中的每个目标词信息，响应于确定上述目标词信息满足收敛条件，基于预设词表，将上述目标词信息和上述目标词信息所对应的历史目标词信息序列进行拼接以生成目标文本。

可选地，确定单元被进一步配置成：将上述待处理文本输入至文本编码器中以生成编码隐层向量；将上述编码隐层向量输入至解码器中以生成解码词信息组集。

可选地，解码词信息组集中的解码词信息包括：解码词和上述解码词对应的解码词概率值；以及去重单元被进一步配置成：对于上述解码词信息组集中的每个解码词信息组，根据上述解码词，将上述解码词信息组划分为重复解码词信息组和单一解码词信息组；从上述重复解码词信息组中选择满足预设条件的重复解码词信息作为目标重复解码词信息；将上述目标重复解码词信息和上述单一解码词信息组进行拼接以生成上述候选词信息组。

可选地，去重单元被进一步配置成：对于上述解码词信息组集中的每个解码词信息组，响应于上述解码词信息组中存在包含该解码词信息的解码词的其他解码词信息，将该解码词信息和上述其他解码词信息放入重复解码词信息组，否则，将该解码词信息放入单一解码词信息组。

可选地，候选词信息组集中的候选词信息包括：候选词和上述候选词对应的候选词概率值；以及选择单元被进一步配置成：对于上述候选词信息组集中的每个候选词信息组，按照上述候选词概率值由大到小的顺序从上述候选词信息组中选择至少一个候选词信息；将上述至少一个候选词信息中概率值最大的候选词信息放入历史目标词信息集，并确定上述概率值最大的候选词信息的至少一个初始目标词信息，并基于上述至少一个初始目标词信息生成对应该候选词信息的目标词

信息。上述至少一个初始目标词信息中的初始目标词信息包括：初始目标词和对应上述初始目标词的初始目标词概率值。

可选地，选择单元被进一步配置成：响应于初始目标词概率值最大的初始目标词信息对应的初始目标词属于上述历史目标词信息集，确定上述初始目标词概率值最大的初始目标词信息与上述历史目标词信息集中对应的历史目标词信息之间的重复词概率差；响应于上述重复词概率差大于该候选词信息对应的其他初始目标词信息的初始目标词概率值，将初始目标词概率值最大的初始目标词信息设置为目标词信息，并将上述目标词信息放入上述历史目标词信息集。

可选地，选择单元被进一步配置成：响应于初始目标词概率值最大的初始目标词信息对应的初始目标词属于上述历史目标词信息集，确定上述初始目标词概率值最大的初始目标词信息与上述历史目标词信息集中对应的历史目标词信息之间的重复词概率差；响应于上述重复词概率差小于等于该候选词信息对应的其他初始目标词信息的初始目标词概率值，将其他初始目标词信息中的初始目标词概率值最大的初始目标词信息设置为目标词信息，并将上述目标词信息放入上述历史目标词信息集。

可选地，选择单元被进一步配置成：响应于初始目标词概率值最大的初始目标词信息对应的初始目标词不属于上述历史目标词信息集，将初始目标词概率值最大的初始目标词信息设置为目标词信息，并将上述目标词信息放入上述历史目标词信息集。

可选地，拼接单元被进一步配置成：对于上述目标词信息集中的每个目标词信息，响应于确定上述目标词信息满足预设收敛条件，将上述目标词信息与上述目标词信息所对应的历史目标词信息序列进行拼接以生成拼接文本；从上述预设词表选择与上述拼接文本相匹配的词作为连接词；对上述连接词和上述拼接文本进行组合以生成目标文本。

第三方面，本公开的一些实施例提供了一种电子设备，包括：一个或多个处理器；存储装置，其上存储有一个或多个程序，当一个或多个程序被一个或多个处理器执行，使得一个或多个处理器实现上述

第一方面任一实现方式所描述的方法。

第四方面，本公开的一些实施例提供了一种计算机可读介质，其上存储有计算机程序，其中，程序被处理器执行时实现上述第一方面任一实现方式所描述的方法。

本公开的上述各个实施例具有如下有益效果：通过本公开的一些实施例的文本生成方法提高了生成文本主题的效率及准确度。具体来说，造成生成文本主题的效率较低及准确度较低的原因在于：对于相同重复的文本没有进行去重处理，并将重复文本对应的重复信息参与到后续计算中，从而导致所生成的文本存在较多的重复文本，造成生成文本主题的效率较低及准确度较低。基于此，本公开的一些实施例的文本方法，首先，基于待处理文本，确定解码词信息组集。由此，所生成的解码词信息组集为后续进行文本去重以及生成目标文本提供数据支撑。其次，对上述解码词信息组集进行去重处理以生成候选词信息组集。由于上述解码词信息组集中的每个解码词信息组中往往存在具有相同解码词的解码词信息，通过将上述具有相同解码词的解码词信息进行去重处理，从而提高了生成文本主题的效率。再次，从上述候选词信息组集中的每个候选词信息组选择满足目标条件的候选词信息作为目标词信息，得到目标词信息集。从候选词信息组集中选择能够引起较强注意力的目标词信息集。最后，对于目标词信息集中的每个目标词信息，响应于确定目标词信息满足收敛条件，基于预设词表，将目标词信息和目标词信息所对应的历史目标词信息序列进行文本拼接以生成目标文本。上述预设词表为口语词汇所组成的词表。将上述预设词表与兼具较高注意力和流畅性的目标词信息集进行组合，从而提高了生成文本主题的准确度。

附图说明

结合附图并参考以下具体实施方式，本公开各实施例的上述和其他特征、优点及方面将变得更加明显。贯穿附图中，相同或相似的附图标记表示相同或相似的元素。应当理解附图是示意性的，元件和元素不一定按照比例绘制。

图 1 是根据本公开的一些实施例的文本生成方法的一个应用场景示意图；

图 2 是根据本公开的文本生成方法的一些实施例的流程图；

图 3 是根据本公开的文本生成方法的另一些实施例的流程图；

图 4 是根据本公开的文本生成方法的另一些实施例的应用场景图；

图 5 是根据本公开的文本生成装置的一些实施例的结构示意图；

图 6 是适于用来实现本公开的一些实施例的电子设备的结构示意图。

具体实施方式

下面将参照附图更详细地描述本公开的实施例。虽然附图中显示了本公开的某些实施例，然而应当理解的是，本公开可以通过各种形式来实现，而且不应该被解释为限于这里阐述的实施例。相反，提供这些实施例是为了更加透彻和完整地理解本公开。应当理解的是，本公开的附图及实施例仅用于示例性作用，并非用于限制本公开的保护范围。

另外还需要说明的是，为了便于描述，附图中仅示出了与有关发明相关的部分。在不冲突的情况下，本公开中的实施例及实施例中的特征可以相互组合。

需要注意，本公开中提及的“第一”、“第二”等概念仅用于对不同的装置、模块或单元进行区分，并非用于限定这些装置、模块或单元所执行的功能的顺序或者相互依存关系。

需要注意，本公开中提及的“一个”、“多个”的修饰是示意性而非限制性的，本领域技术人员应当理解，除非在上下文另有明确指出，否则应该理解为“一个或多个”。

本公开实施方式中的多个装置之间所交互的消息或者信息的名称仅用于说明性的目的，而并不是用于对这些消息或信息的范围进行限制。

下面将参考附图并结合实施例来详细说明本公开。

图 1 是本公开的一些实施例的文本生成方法的一个应用场景的示意图。

在图 1 的应用场景中，首先，计算设备 101 可以基于待处理文本 102（例如，“洗衣机”，“洗衣机”，“呵护”，“呵护”，“衣服”，“衣服”，“美丽”，“美丽”，“清洗”），确定解码词信息组集 103（例如：[“洗衣机”：0.03，“洗衣机”：0.07，“呵护”：0.02，“呵护”：0.01，“衣服”：0.4，“衣服”：0.1，“美丽”：0.1，“美丽”：0.01，“清洗”：0.04]，[“洗衣机”：0.01，“洗衣机”：0.04，“呵护”：0.06，“呵护”：0.02，“衣服”：0.04，“衣服”：0.07，“美丽”：0.01，“美丽”：0.5，“清洗”：0.01]）。其中，上述待处理文本 102 用于描述指定物体。然后，计算设备 101 可以对上述解码词信息组集 103 进行去重处理以生成候选词信息组集 104（例如：[“洗衣机”：0.07，“呵护”：0.02，“衣服”：0.4，“美丽”：0.1，“清洗”：0.04]，[“洗衣机”：0.04，“呵护”：0.06，“衣服”：0.07，“美丽”：0.5，“清洗”：0.01]）。之后，计算设备 101 可以从上述候选词信息组集 104 中的每个候选词信息组选择满足目标条件的候选词信息作为目标词信息，得到目标词信息集 105（例如：“衣服”：0.4，“美丽”：0.5）。最后，计算设备 101 可以对于上述目标词信息集 105 中的每个目标词信息，响应于确定上述目标词信息满足收敛条件，基于预设词表 106（例如：“您的”，“更”，“能够”，“有效地”...），将上述目标词信息和上述目标词信息所对应的历史目标词信息序列 107（例如：[“洗衣机”：0.4，“清洗”：0.3]，[“呵护”：0.5，“衣服”：0.6]）进行拼接以生成目标文本 108 例如（[“洗衣机能够有效地洗衣服”]，[“呵护您的衣服更美丽”]）。

需要说明的是，上述计算设备 101 可以是硬件，也可以是软件。当计算设备为硬件时，可以实现成多个服务器或终端设备组成的分布式集群，也可以实现成单个服务器或单个终端设备。当计算设备体现为软件时，可以安装在上述所列举的硬件设备中。其可以实现成例如用来提供分布式服务的多个软件或软件模块，也可以实现成单个软件或软件模块。在此不做具体限定。

应该理解，图 1 中的计算设备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计算设备。

继续参考图 2，示出了根据本公开的文本生成方法的一些实施例的流程 200。该文本生成方法，包括以下步骤：

步骤 201，基于待处理文本，确定解码词信息组集。

在一些实施例中，文本生成方法的执行主体（例如图 1 所示的计算设备 101）可以基于待处理文本，确定解码词信息组集，其中，上述待处理文本可以是等待被处理的文本。上述待处理文本可以是用于描述指定物体的词汇。上述解码词信息组集可以通过将上述预处理文本输入至预先训练的文本生成模型中，所生成的各个时间步中的任意一个时间步的解码词信息组集。上述预先训练的文本生成模型可以是 Seq2Seq（Sequence to sequence，序列对序列）模型。上述预先训练的文本生成模型也可以是 SeqGAN（Sequence Generative Adversarial Nets with Policy Gradient，序列对抗网络）模型。上述解码词信息组集中的解码词信息可以包括：解码词和上述解码词对应的解码词概率。

作为示例，上述待处理文本可以是[“洗衣机”、“洗衣机”、“呵护”、“呵护”、“衣服”、“衣服”、“美丽”、“美丽”、“清洗”]。将上述待处理文本输入至预先训练的文本生成模型中，所生成的解码词信息组集可以是[[“洗衣机”：0.03、“洗衣机”：0.07、“呵护”：0.02、“呵护”：0.01、“衣服”：0.4、“衣服”：0.1、“美丽”：0.1、“美丽”：0.01、“清洗”：0.04]，[“洗衣机”：0.01、“洗衣机”：0.04、“呵护”：0.06、“呵护”：0.02、“衣服”：0.04、“衣服”：0.07、“美丽”：0.01、“美丽”：0.5、“清洗”：0.01]]。

步骤 202，对解码词信息组集进行去重处理以生成候选词信息组集。

在一些实施例中，上述文本生成方法的执行主体对解码词信息组集进行去重处理以生成候选词信息组集。其中，基于上述解码词，对于上述解码词信息组集中的每个解码词信息组，将上述解码词信息组中具有相同解码词的解码词信息进行去重处理，只保留相同解码词信息中概率值最大的解码词信息，并将不具有相同解码词的解码词信息

与保留相同解码词信息中概率值最大的解码词信息进行组合以生成候选词信息组。

作为示例，上述解码词信息组集可以是[[“洗衣机”：0.03、“洗衣机”：0.07、“呵护”：0.02、“呵护”：0.01、“衣服”：0.4、“衣服”：0.1、“美丽”：0.1、“美丽”：0.01、“清洗”：0.04]，[“洗衣机”：0.01、“洗衣机”：0.04、“呵护”：0.06、“呵护”：0.02、“衣服”：0.04、“衣服”：0.07、“美丽”：0.01、“美丽”：0.5、“清洗”：0.01]]。对上述解码词信息组集进行去重处理以生成候选词信息组集可以是[[“洗衣机”：0.07、“呵护”：0.02、“衣服”：0.4、“美丽”：0.1、“清洗”：0.04]，[“洗衣机”：0.04、“呵护”：0.06、“衣服”：0.07、“美丽”：0.5、“清洗”：0.01]]。

步骤 203，从候选词信息组集中的每个候选词信息组选择满足目标条件的候选词信息作为目标词信息，得到目标词信息集。

在一些实施例中，上述文本生成方法的执行主体可以从候选词信息组集中的每个候选词信息组选择满足目标条件的候选词信息作为目标词信息，得到目标词信息集。其中，上述目标条件可以是上述候选信息可以是上述候选词信息组中具有最大概率值的候选词信息。

作为示例，上述候选词信息组集可以是[[“洗衣机”：0.07、“呵护”：0.02、“衣服”：0.4、“美丽”：0.1、“清洗”：0.04]，[“洗衣机”：0.04、“呵护”：0.06、“衣服”：0.07、“美丽”：0.5、“清洗”：0.01]]。则从候选词信息组集中的每个候选词信息组选择满足目标条件的候选词信息作为目标词信息，得到目标词信息集可以是[[“衣服”：0.4]，[“美丽”：0.5]]。

步骤 204，对于目标词信息集中的每个目标词信息，响应于确定目标词信息满足收敛条件，基于预设词表，将目标词信息和目标词信息所对应的历史目标词信息序列进行文本拼接以生成目标文本。

在一些实施例中，上述文本生成方法的执行主体可以对于目标词信息集中的每个目标词信息，响应于确定目标词信息满足收敛条件，将目标词信息所对应的目标词和目标词信息所对应的历史目标词信息序列所对应的历史目标词序列进行文本拼接以生成拼接文本。并确定上述预设词表中与上述拼接文本相匹配的词，之后将所匹配的词与上

述拼接文本进行组合，生成目标文本。其中，上述目标词信息满足收敛条件可以是上述目标词信息为预设时间步长的最后一个时间步所对应的目标词信息。上述预设词表可以是口语词汇所组成的词表。上述历史目标词信息序列中历史目标词信息可以是与上述目标词信息所对应的预设时间步长中的每个时间步所生成的历史目标词信息。

作为示例，上述目标词信息集可以是[[“衣服”：0.4]，[“美丽”：0.5]]。可以对于目标词信息集中的每个目标词信息，响应于确定目标词信息为预设时间步长的最后一个时间步所对应的目标词信息。上述预设时间步长可以是3。上述目标词信息可以是[“衣服”：0.4]。上述目标词信息所对应的历史目标词信息序列可以是[[“洗衣机”：0.4]，[“清洗”：0.3]]。上述目标词信息可以是[“美丽”：0.5]。上述目标词信息所对应的历史目标词信息序列可以是[[“呵护”：0.5]，[“衣服”：0.6]]。上述预设词表可以是[“您的”，“更”，“能够”，“有效地”]。则最后文本拼接后的所生成的目标文本集可以是[[“洗衣机能够有效地清洗衣服”]，[“呵护您的衣服更美丽”]]。

本公开的上述各个实施例具有如下有益效果：通过本公开的一些实施例的文本生成方法提高了生成文本主题的效率及准确度。具体来说，造成生成文本主题的效率较低及准确度较低的原因在于：对于相同重复的文本没有进行去重处理，并将重复文本对应的重复信息参与到后续计算中，从而导致所生成的文本存在较多的重复文本，造成生成文本主题的效率较低及准确度较低。基于此，本公开的一些实施例的文本方法，首先，基于待处理文本，确定解码词信息组集。由此，所生成的解码词信息组集为后续进行文本去重以及生成目标文本提供数据支撑。其次，对上述解码词信息组集进行去重处理以生成候选词信息组集。由于上述解码词信息组集中的每个解码词信息组中往往存在具有相同解码词的解码词信息，通过将上述具有相同解码词的解码词信息进行去重处理，从而提高了生成文本主题的效率。再次，从上述候选词信息组集中的每个候选词信息组选择满足目标条件的候选词信息作为目标词信息，得到目标词信息集。从候选词信息组集中选择能够引起较强注意力的目标词信息集。最后，对于目标词信息集中的

每个目标词信息，响应于确定目标词信息满足收敛条件，基于预设词表，将目标词信息和目标词信息所对应的历史目标词信息序列进行文本拼接以生成目标文本。上述预设词表为口语词汇所组成的词表。将上述预设词表与兼具较高注意力和流畅性的目标词信息集进行组合，从而提高了生成文本主题的准确度。

继续参考图 3，图 3 示出了根据本公开的文本生成方法的另一些实施例的流程 300。该文本生成方法，包括以下步骤：

步骤 301，基于待处理文本，确定解码词信息组集。

在一些实施例中，上述执行主体基于待处理文本，确定解码词信息组集，可以包括以下步骤：

第一步，将上述待处理文本输入至文本编码器中以生成编码隐层向量。

在一些实施例中，上述执行主体可以将上述待处理文本输入至文本编码器中以生成编码隐层向量。上述文本编码器可以用于将待处理文本压缩至低纬度文本向量。上述编码隐层向量可以为经过特征压缩后的词向量。上述文本编码器可以是由各种不同类型的神经网络模型构成的编码器。上述文本编码器可以是 Auto-Encoder（自动编码器）。上述文本编码器也可以是 seq2seq (Sequence to Sequence, 序列对序列) 模型中的编码器。

第二步，将上述编码隐层向量输入至解码器中以生成解码词信息组集。

在一些实施例中，上述执行主体可以将上述编码隐层向量输入至解码器中以生成解码词信息组集。上述解码器可以用于将上述编码隐层向量输入至预先训练的神经网络模型进行词解码以生成解码词信息组集。上述预先训练的神经网络模型可以是 RNN (Recurrent Neural Network, 循环神经网络)。上述预先训练的神经网络模型也可以是 CNN (Convolutional Neural Networks, 卷积神经网络)。上述解码词信息组集中的解码词信息可以包括解码词和上述解码词对应的解码词概率。

作为示例,上述待处理文本可以是[“洗衣机”、“洗衣机”、“呵护”、“呵护”、“衣服”、“衣服”、“美丽”、“美丽”、“清洗”]。上述所生成的解码词信息组集可以是[[“洗衣机”:0.03、“洗衣机”:0.07、“呵护”:0.02、“呵护”:0.01、“衣服”:0.04、“衣服”:0.01、“美丽”:0.1、“美丽”:0.01、“清洗”:0.4], [“洗衣机”:0.01、“洗衣机”:0.04、“呵护”:0.06、“呵护”:0.02、“衣服”:0.4、“衣服”:0.07、“美丽”:0.01、“美丽”:0.05、“清洗”:0.04]]。

步骤 302,对于解码词信息组集中的每个解码词信息组,根据解码词,将解码词信息组划分为重复解码词信息组和单一解码词信息组。

在一些实施例中,上述执行主体可以对于上述解码词信息组集中的每个解码词信息组,响应于上述解码词信息组中存在包含该解码词信息的解码词的其他解码词信息,将该解码词信息和上述其他解码词信息放入重复解码词信息组,否则,将该解码词信息放入单一解码词信息组。上述重复解码词信息组中的重复解码词信息可以是具有相同解码词的解码词信息。上述单一解码词信息组中的单一解码词信息可以是不具有相同解码词的解码词信息。

作为示例,上述所生成的解码词信息组集可以是[[“洗衣机”:0.03、“洗衣机”:0.07、“呵护”:0.02、“呵护”:0.01、“衣服”:0.04、“衣服”:0.01、“美丽”:0.1、“美丽”:0.01、“清洗”:0.4], [“洗衣机”:0.01、“洗衣机”:0.04、“呵护”:0.06、“呵护”:0.02、“衣服”:0.4、“衣服”:0.07、“美丽”:0.01、“美丽”:0.05、“清洗”:0.04]]。

对于上述解码词信息组集中的其中一个解码词信息:[“洗衣机”:0.03、“洗衣机”:0.07、“呵护”:0.02、“呵护”:0.01、“衣服”:0.04、“衣服”:0.01、“美丽”:0.1、“美丽”:0.01、“清洗”:0.4]对应的重复解码词信息组和单一解码词信息组分别可以是[[“洗衣机”:0.03、“洗衣机”:0.07], [“呵护”:0.02、“呵护”:0.01], [“衣服”:0.04、“衣服”:0.01], [“美丽”:0.1、“美丽”:0.01]]和[“清洗”:0.4]。

作为又一示例,上述所生成的解码词信息组集可以是[[“洗衣机”:0.03、“洗衣机”:0.07、“呵护”:0.02、“呵护”:0.01、“衣服”:0.04、“衣服”:0.01、“美丽”:0.1、“美丽”:0.01、“清洗”:0.4], [“洗衣

机”:0.01、“洗衣机”: 0.04、“呵护”: 0.06、“呵护”: 0.02、“衣服”: 0.4、“衣服”: 0.07、“美丽”: 0.01、“美丽”: 0.05、“清洗”: 0.04]]。

对于上述解码词信息组集中的另外一个解码词信息: [“洗衣机”:0.01、“洗衣机”: 0.04、“呵护”: 0.06、“呵护”: 0.02、“衣服”: 0.4、“衣服”: 0.07、“美丽”: 0.01、“美丽”: 0.05、“清洗”: 0.04]]对应的重复解码词信息组和单一解码词信息组分别可以是[[“洗衣机”: 0.01、“洗衣机”: 0.04]、[“呵护”: 0.06、“呵护”: 0.02]、[“衣服”: 0.4、“衣服”: 0.07]、“美丽”: 0.01、“美丽”: 0.05]]和[“清洗”: 0.04]。

步骤 303, 从重复解码词信息组中选择满足预设条件的重复解码词信息作为目标重复解码词信息。

在一些实施例中, 上述执行主体可以从重复解码词信息组中选择满足预设条件的重复解码词信息作为目标重复解码词信息。其中, 上述预设条件可以是上述重复解码词信息是上述重复解码词信息组中解码词概率值最大的重复解码词信息。对于上述重复解码词信息组中的每个重复解码词信息, 按照解码词概率值由高到低进行排序, 得到的重复解码词信息序列, 并从重复解码词信息序列选取最大解码词概率值的重复解码词信息。

作为示例, 上述重复解码词信息组可以是[[“洗衣机”: 0.03、“洗衣机”: 0.07], [“呵护”: 0.02、“呵护”: 0.01], [“衣服”: 0.04、“衣服”: 0.01], [“美丽”: 0.1、“美丽”: 0.01]]。则所生成的目标重复解码词信息可以是[[“洗衣机”: 0.07], [“呵护”: 0.02], [“衣服”: 0.04], [“美丽”: 0.1]]。

作为又一示例, 上述重复解码词信息组可以是[[“洗衣机”: 0.01、“洗衣机”: 0.04]、[“呵护”: 0.06、“呵护”: 0.02]、[“衣服”: 0.4、“衣服”: 0.07]、“美丽”: 0.01、“美丽”: 0.05]]。则所生成的目标重复解码词信息可以是[[“洗衣机”: 0.04]、“呵护”: 0.06]、“衣服”: 0.4]、“美丽”: 0.05]]。

步骤 304, 将目标重复解码词信息和单一解码词信息组进行拼接以生成候选词信息组。

在一些实施例中, 上述执行主体可以将目标重复解码词信息和单

一解码词信息组进行拼接以生成候选词信息组。上述拼接处理可以根据解码词信息组，将上述目标重复解码词信息和上述目标重复解码词信息对应的单一解码词信息组进行组合以生成候选词信息组。

作为示例，上述目标重复解码词信息可以是[[“洗衣机”：0.07]，[“呵护”：0.02]，[“衣服”：0.04]，[“美丽”：0.1]]。与上述目标重复解码词信息对应的单一解码词信息组可以是[“清洗”：0.04]。则将两组信息组合所生成的候选词信息组可以是[[“洗衣机”：0.07]，[“呵护”：0.02]，[“衣服”：0.04]，[“美丽”：0.1]，[“清洗”：0.4]]。

作为又一示例，上述目标重复解码词信息可以是[[“洗衣机”：0.04]、[“呵护”：0.06]、[“衣服”：0.4]、[“美丽”：0.05]]。与上述目标重复解码词信息对应的单一解码词信息组可以是[“清洗”：0.01]。则将两组信息组合所生成的候选词信息组可以是[[“洗衣机”：0.04]、[“呵护”：0.06]、[“衣服”：0.4]、[“美丽”：0.05]，[“清洗”：0.01]]。

步骤 305，对于候选词信息组集中的每个候选词信息组，按照候选词概率值由大到小的顺序从候选词信息组中选择至少一个候选词信息。

在一些实施例中，上述执行主体可以对于候选词信息组集中的每个候选词信息组，按照候选词概率值由大到小的顺序从候选词信息组中选择至少一个候选词信息，其中，候选词信息包括：候选词和对应候选词的候选词概率值。

作为示例，上述候选词信息组可以是[[“洗衣机”：0.07]，[“呵护”：0.02]，[“衣服”：0.04]，[“美丽”：0.1]，[“清洗”：0.4]]。从上述候选词信息组中按照候选词概率值由大到小的顺序从候选词信息组中选择至少一个候选词信息可以是[“美丽”：0.1，“清洗”：0.4]。另一个候选词信息组可以是[[“洗衣机”：0.04]、[“呵护”：0.06]、[“衣服”：0.4]、[“美丽”：0.05]，[“清洗”：0.01]]。从上述候选词信息组中按照候选词概率值由大到小的顺序从候选词信息组中选择至少一个候选词信息可以是[“呵护”：0.06，“衣服”：0.4]。

步骤 306，将上述至少一个候选词信息中概率值最大的候选词信息放入历史目标词信息集，并确定上述概率值最大的候选词信息的至

少一个初始目标词信息，并基于上述至少一个初始目标词信息生成对应应该候选词信息的目标词信息。

在一些实施例中，上述执行主体可以首先对于至少一个候选词信息中的候选词信息，将至少一个候选词信息中概率值最大的候选词信息放入历史目标词信息集。其次，确定概率值最大的候选词信息的至少一个初始目标词信息。最后，基于至少一个初始目标词信息生成对应应该候选词信息的目标词信息。其中，上述至少一个初始目标词信息中的初始目标词信息包括：初始目标词和对应上述初始目标词的初始目标词概率值。从至少一个初始目标词信息选取初始目标词概率值最大的初始目标词信息作为目标词信息。

作为示例，上述至少一个候选词信息可以是[“美丽”：0.1，“清洗”：0.4]。上述至少一个候选词信息也可以是[“呵护”：0.06，“衣服”：0.4]。概率值最大的候选词信息可以是[“清洗”：0.4]或[“衣服”：0.4]。将上述候选词信息放入历史目标词信息集可以是[[“清洗”：0.4]，[“衣服”：0.4]]。确定上述候选词信息中的初始目标词信息。上述候选词信息可以是[“清洗”：0.4]。与上述候选词信息对应的初始目标词信息可以是[“衣服”：0.7，“美丽”：0.1]，且所生成与上述候选词信息对应的目标词信息可以是[“衣服”：0.7]。上述候选词信息可以是[“衣服”：0.4]，而与上述候选词信息对应的初始目标词信息可以是[“衣服”：0.07，“美丽”：0.5]。且所生成与上述候选词信息对应的目标词信息可以是[“美丽”：0.5]。具体的，生成目标词信息的应用场景如图4所示。

在一些实施例的一些可选的实现方式中，上述执行主体首先可以响应于初始目标词概率值最大的初始目标词信息对应的初始目标词属于上述历史目标词信息集，确定上述初始目标词概率值最大的初始目标词信息与上述历史目标词信息集中对应的历史目标词信息之间的重复词概率差。其次，上述执行主体可以响应于上述重复词概率差大于该候选词信息对应的其他初始目标词信息的初始目标词概率值，将初始目标词概率值最大的初始目标词信息设置为目标词信息，并将上述目标词信息放入上述历史目标词信息集。

作为示例，上述初始目标词概率值最大的初始目标词信息对应的

初始目标词可以是“衣服”。上述历史目标词信息集可以是[[“清洗”: 0.4], [“衣服”: 0.4]]。上述初始目标词概率值最大的初始目标词信息对应的初始目标词的初始目标词概率值可以是 0.7。则确定上述初始目标词概率值最大的初始目标词信息与上述历史目标词信息集中对应的历史目标词信息之间的重复词概率差可以是 0.3。上述其他初始目标词信息可以是 [“美丽”: 0.1]。响应于上述重复词概率差大于该候选词信息对应的其他初始目标词信息的初始目标词概率值，将初始目标词概率值最大的初始目标词信息[“衣服”: 0.7]设置为目标词信息，并将上述目标词信息放入上述历史目标词信息集[“清洗”: 0.4, “衣服”: 0.7,]。

在一些实施例的一些可选的实现方式中，上述执行主体首先响应于初始目标词概率值最大的初始目标词信息对应的初始目标词属于上述历史目标词信息集，确定上述初始目标词概率值最大的初始目标词信息与上述历史目标词信息集中对应的历史目标词信息之间的重复词概率差。其次，响应于上述重复词概率差小于等于该候选词信息对应的其他初始目标词信息的初始目标词概率值，将其他初始目标词信息中的初始目标词概率值最大的初始目标词信息设置为目标词信息，并将上述目标词信息放入上述历史目标词信息集。

在一些实施例的一些可选的实现方式中，上述执行主体首先响应于初始目标词概率值最大的初始目标词信息对应的初始目标词不属于上述历史目标词信息集，将初始目标词概率值最大的初始目标词信息设置为目标词信息，并将上述目标词信息放入上述历史目标词信息集。

步骤 307，对于目标词信息集中的每个目标词信息，响应于确定目标词信息满足收敛条件，基于预设词表，将目标词信息和目标词信息所对应的历史目标词信息序列进行拼接以生成目标文本。

在一些实施例中，上述执行主体对于目标词信息集中的每个目标词信息，响应于确定目标词信息满足收敛条件，基于预设词表，将目标词信息和目标词信息所对应的历史目标词信息序列进行拼接以生成目标文本，可以包括以下步骤：

第一步，对于上述目标词信息集中的每个目标词信息，响应于确定上述目标词信息满足预设收敛条件，将上述目标词信息与上述目标

词信息所对应的历史目标词信息序列进行拼接以生成拼接文本。

作为示例，上述目标词信息可以是[“衣服”: 0.7]。上述目标词信息所对应的历史目标词信息序列可以是[“洗衣机”: 0.3, “清洗”: 0.4]。将上述目标词信息与上述目标词信息所对应的历史目标词信息序列进行拼接以生成拼接文本可以是[“洗衣机清洗衣服”]。

第二步，从上述预设词表选择与上述拼接文本相匹配的词作为连接词。

其中，用各种机器学习的方法，从上述预设词表选择与上述拼接文本相匹配的词作为连接词。

作为示例，上述预设词表可以[“您的”、“能够”、“有效地”]。所选择的连接词可以是[“能够”、“有效地”]。

第三步，对上述连接词和上述拼接文本进行组合以生成目标文本。

作为示例，对上述连接词和上述拼接文本进行组合以生成目标文本可以是[“洗衣机能够有效地清洗衣服”]。

从图 3 可以看出，与图 2 对应的一些实施例的描述相比，图 3 对应的一些实施例中的信息处理方法的流程 300。首先，基于待处理文本，确定解码词信息组集。由此，所生成的解码词信息组集为后续进行文本去重以及生成目标文本提供数据支撑。其次，对上述解码词信息组集进行去重处理以生成候选词信息组集。由于上述解码词信息组集中的每个解码词信息组中往往存在具有相同解码词的解码词信息，通过将上述具有相同解码词的解码词信息进行去重处理，从而提高了生成文本主题的效率。再次，从上述候选词信息组集中的每个候选词信息组中选择满足目标条件的候选词信息作为目标词信息，得到目标词信息集。其中，目标词信息所对应的目标词即使在之前被解码过，但若该目标词信息在当前时间步能够引起极强的注意力，仍然会被解码作为目标词信息。由此，从使所生成的目标词信息为实际生活中较为流畅的预测词。最后，对于目标词信息集中的每个目标词信息，响应于确定目标词信息满足收敛条件，基于预设词表，将目标词信息和目标词信息所对应的历史目标词信息序列进行文本拼接以生成目标文本。上述预设词表为口语词汇所组成的词表。将上述预设词表与兼具

较高注意力和流畅性的目标词信息集进行组合，不仅提高了生成文本主题的准确度，而且所生成的目标文本兼具较高注意力和流畅性的特征。

进一步参考图 5，作为对上述各图所示方法的实现，本公开提供了一种文本生成装置的一些实施例，这些装置实施例与图 2 所示的那些方法实施例相对应，该装置具体可以应用于各种电子设备中。

如图 5 所示，一些实施例的文本生成装置 500 包括：确定单元 501，被配置成基于待处理文本，确定解码词信息组集，其中，上述待处理文本用于描述指定物体；去重单元 502，被配置成对上述解码词信息组集进行去重处理以生成候选词信息组集；选择单元 503，被配置成从上述候选词信息组集中的每个候选词信息组选择满足目标条件的候选词信息作为目标词信息，得到目标词信息集；拼接单元 504，被配置成对于上述目标词信息集中的每个目标词信息，响应于确定上述目标词信息满足收敛条件，基于预设词表，将上述目标词信息和上述目标词信息所对应的历史目标词信息序列进行拼接以生成目标文本。

可以理解的是，该装置 500 中记载的诸单元与参考图 2 描述的方法中的各个步骤相对应。由此，上文针对方法描述的操作、特征以及产生的有益效果同样适用于装置 500 及其中包含的单元，在此不再赘述。

在一些实施例的可选的实现方式中，确定单元 501 被进一步配置成：将上述待处理文本输入至文本编码器中以生成编码隐层向量；将上述编码隐层向量输入至解码器中以生成解码词信息组集。

在一些实施例的可选的实现方式中，解码词信息组集中的解码词信息包括：解码词和上述解码词对应的解码词概率值；以及去重单元 502 被进一步配置成：对于上述解码词信息组集中的每个解码词信息组，根据上述解码词，将上述解码词信息组划分为重复解码词信息组和单一解码词信息组；从上述重复解码词信息组中选择满足预设条件的重复解码词信息作为目标重复解码词信息；将上述目标重复解码词信息和上述单一解码词信息组进行拼接以生成上述候选词信息组。

在一些实施例的可选的实现方式中，去重单元 502 被进一步配置成：对于上述解码词信息组集中的每个解码词信息组，响应于上述解码词信息组中存在包含该解码词信息的解码词的其他解码词信息，将该解码词信息和上述其他解码词信息放入重复解码词信息组，否则，将该解码词信息放入单一解码词信息组。

在一些实施例的可选的实现方式中，候选词信息组集中的候选词信息包括：候选词和上述候选词对应的候选词概率值；以及选择单元 503 被进一步配置成：对于上述候选词信息组集中的每个候选词信息组，按照上述候选词概率值由大到小的顺序从上述候选词信息组中选择至少一个候选词信息；将上述至少一个候选词信息中概率值最大的候选词信息放入历史目标词信息集，并确定上述概率值最大的候选词信息的至少一个初始目标词信息，并基于上述至少一个初始目标词信息生成对应该候选词信息的目标词信息。上述至少一个初始目标词信息中的初始目标词信息包括：初始目标词和对应上述初始目标词的初始目标词概率值。

在一些实施例的可选的实现方式中，选择单元 503 被进一步配置成：响应于初始目标词概率值最大的初始目标词信息对应的初始目标词属于上述历史目标词信息集，确定上述初始目标词概率值最大的初始目标词信息与上述历史目标词信息集中对应的历史目标词信息之间的重复词概率差；响应于上述重复词概率差大于该候选词信息对应的其他初始目标词信息的初始目标词概率值，将初始目标词概率值最大的初始目标词信息设置为目标词信息，并将上述目标词信息放入上述历史目标词信息集。

在一些实施例的可选的实现方式中，选择单元 503 被进一步配置成：响应于初始目标词概率值最大的初始目标词信息对应的初始目标词属于上述历史目标词信息集，确定上述初始目标词概率值最大的初始目标词信息与上述历史目标词信息集中对应的历史目标词信息之间的重复词概率差；响应于上述重复词概率差小于等于该候选词信息对应的其他初始目标词信息的初始目标词概率值，将其他初始目标词信息中的初始目标词概率值最大的初始目标词信息设置为目标词信息，

并将上述目标词信息放入上述历史目标词信息集。

在一些实施例的可选的实现方式中，选择单元 503 被进一步配置成：响应于初始目标词概率值最大的初始目标词信息对应的初始目标词不属于上述历史目标词信息集，将初始目标词概率值最大的初始目标词信息设置为目标词信息，并将上述目标词信息放入上述历史目标词信息集。

在一些实施例的可选的实现方式中，拼接单元 504 被进一步配置成：对于上述目标词信息集中的每个目标词信息，响应于确定上述目标词信息满足预设收敛条件，将上述目标词信息与上述目标词信息所对应的历史目标词信息序列进行拼接以生成拼接文本；从上述预设词表选择与上述拼接文本相匹配的词作为连接词；对上述连接词和上述拼接文本进行组合以生成目标文本。

下面参考图 6，其示出了适于用来实现本公开的一些实施例的电子设备（例如图 1 中的计算设备 101）600 的结构示意图。图 6 示出的电子设备仅仅是一个示例，不应对本公开的实施例的功能和使用范围带来任何限制。

如图 6 所示，电子设备 600 可以包括处理装置（例如中央处理器、图形处理器等）601，其可以根据存储在只读存储器（ROM）602 中的程序或者从存储装置 608 加载到随机访问存储器（RAM）603 中的程序而执行各种适当的动作和处理。在 RAM 603 中，还存储有电子设备 600 操作所需的各种程序和数据。处理装置 601、ROM 602 以及 RAM 603 通过总线 604 彼此相连。输入/输出（I/O）接口 605 也连接至总线 604。

通常，以下装置可以连接至 I/O 接口 605：包括例如触摸屏、触摸板、键盘、鼠标、摄像头、麦克风、加速度计、陀螺仪等的输入装置 606；包括例如液晶显示器（LCD）、扬声器、振动器等的输出装置 607；包括例如磁带、硬盘等的存储装置 608；以及通信装置 609。通信装置 609 可以允许电子设备 600 与其他设备进行无线或有线通信以交换数据。虽然图 6 示出了具有各种装置的电子设备 600，但是应理解的是，

并不要求实施或具备所有示出的装置。可以替代地实施或具备更多或更少的装置。图 6 中示出的每个方框可以代表一个装置，也可以根据需要代表多个装置。

特别地，根据本公开的一些实施例，上文参考流程图描述的过程可以被实现为计算机软件程序。例如，本公开的一些实施例包括一种计算机程序产品，其包括承载在计算机可读介质上的计算机程序，该计算机程序包含用于执行流程图所示的方法的程序代码。在这样的一些实施例中，该计算机程序可以通过通信装置 609 从网络上被下载和安装，或者从存储装置 608 被安装，或者从 ROM 602 被安装。在该计算机程序被处理装置 601 执行时，执行本公开的一些实施例的方法中限定的上述功能。

需要说明的是，本公开的一些实施例中记载的计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质或者是上述两者的任意组合。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子可以包括但不限于：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机访问存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑磁盘只读存储器（CD-ROM）、光存储器件、磁存储器件、或者上述的任意合适的组合。在本公开的一些实施例中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。而在本公开的一些实施例中，计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括但不限于电磁信号、光信号或上述的任意合适的组合。计算机可读信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读信号介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括但不限于：电线、光缆、RF（射频）等

等，或者上述的任意合适的组合。

在一些实施方式中，客户端、服务器可以利用诸如 HTTP (HyperText Transfer Protocol, 超文本传输协议) 之类的任何当前已知或未来研发的网络协议进行通信，并且可以与任意形式或介质的数字数据通信（例如，通信网络）互连。通信网络的示例包括局域网（“LAN”），广域网（“WAN”），网际网（例如，互联网）以及端对端网络（例如，ad hoc 端对端网络），以及任何当前已知或未来研发的网络。

上述计算机可读介质可以是上述装置中所包含的；也可以是单独存在，而未装配入该电子设备中。上述计算机可读介质承载有一个或者多个程序，当上述一个或者多个程序被该电子设备执行时，使得该电子设备：基于待处理文本，确定解码词信息组集，其中，上述待处理文本用于描述指定物体；对上述解码词信息组集进行去重处理以生成候选词信息组集；从上述候选词信息组集中的每个候选词信息组选择满足目标条件的候选词信息作为目标词信息，得到目标词信息集；对于上述目标词信息集中的每个目标词信息，响应于确定上述目标词信息满足收敛条件，基于预设词表，将上述目标词信息和上述目标词信息所对应的历史目标词信息序列进行拼接以生成目标文本。

可以以一种或多种程序设计语言或其组合来编写用于执行本公开的一些实施例的操作的计算机程序代码，上述程序设计语言包括面向对象的程序设计语言—诸如 Java、Smalltalk、C++，还包括常规的进程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网（LAN）或广域网（WAN）——连接到用户计算机，或者，可以连接到外部计算机（例如利用因特网服务提供商来通过因特网连接）。

附图中的流程图和框图，图示了按照本公开各种实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这

上，流程图或框图中的每个方框可以代表一个模块、程序段、或代码的一部分，该模块、程序段、或代码的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。也应当注意，在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个接连地表示的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或操作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。

描述于本公开的一些实施例中的单元可以通过软件的方式实现，也可以通过硬件的方式来实现。所描述的单元也可以设置在处理器中，例如，可以描述为：一种处理器包括确定单元、去重单元、选择单元和拼接单元。其中，这些单元的名称在某种情况下并不构成对该单元本身的限定，例如，去重单元还可以被描述为“对上述解码词信息组集进行去重处理以生成候选词信息组集的单元”。

本文中以上描述的功能可以至少部分地由一个或多个硬件逻辑部件来执行。例如，非限制性地，可以使用的示范类型的硬件逻辑部件包括：现场可编程门阵列（FPGA）、专用集成电路（ASIC）、专用标准产品（ASSP）、片上系统（SOC）、复杂可编程逻辑设备（CPLD）等等。

以上描述仅为本公开的一些较佳实施例以及对所运用技术原理的说明。本领域技术人员应当理解，本公开的实施例中所涉及的发明范围，并不限于上述技术特征的特定组合而成的技术方案，同时也应涵盖在不脱离上述发明构思的情况下，由上述技术特征或其等同特征进行任意组合而形成的其它技术方案。例如上述特征与本公开的实施例中公开的（但不限于）具有类似功能的技术特征进行互相替换而形成的技术方案。

权 利 要 求 书

1、一种文本生成方法，包括：

基于待处理文本，确定解码词信息组集，其中，所述待处理文本用于描述指定物体；

对所述解码词信息组集进行去重处理以生成候选词信息组集；

从所述候选词信息组集中的每个候选词信息组选择满足目标条件的候选词信息作为目标词信息，得到目标词信息集；

对于所述目标词信息集中的每个目标词信息，响应于确定所述目标词信息满足收敛条件，基于预设词表，将所述目标词信息和所述目标词信息所对应的历史目标词信息序列进行拼接以生成目标文本。

2、根据权利要求1所述的方法，其中，所述基于待处理文本，确定解码词信息组集，包括：

将所述待处理文本输入至文本编码器中以生成编码隐层向量；

将所述编码隐层向量输入至解码器中以生成解码词信息组集。

3、根据权利要求1或2所述的方法，其中，所述解码词信息组集中的解码词信息包括：解码词和所述解码词对应的解码词概率值；以及

所述对所述解码词信息组集进行去重处理以生成候选词信息组集，包括：

对于所述解码词信息组集中的每个解码词信息组，根据所述解码词，将所述解码词信息组划分为重复解码词信息组和单一解码词信息组；

从所述重复解码词信息组中选择满足预设条件的重复解码词信息作为目标重复解码词信息；

将所述目标重复解码词信息和所述单一解码词信息组进行拼接以生成所述候选词信息组。

4、根据权利要求3所述的方法，其中，所述对于所述解码词信息组集中的每个解码词信息组，根据所述解码词，将所述解码词信息组划分为重复解码词信息组和单一解码词信息组，包括：

对于所述解码词信息组集中的每个解码词信息组，响应于所述解码词信息组中存在包含该解码词信息的解码词的其他解码词信息，将该解码词信息和所述其他解码词信息放入重复解码词信息组，否则，将该解码词信息放入单一解码词信息组。

5、根据权利要求1-4之一所述的方法，其中，所述候选词信息组集中的候选词信息包括：候选词和所述候选词对应的候选词概率值；以及

所述从所述候选词信息组集中的每个候选词信息组选择满足目标条件的候选词信息作为目标词信息，包括：

对于所述候选词信息组集中的每个候选词信息组，按照所述候选词概率值由大到小的顺序从所述候选词信息组中选择至少一个候选词信息；

将所述至少一个候选词信息中概率值最大的候选词信息放入历史目标词信息集，并确定所述概率值最大的候选词信息的至少一个初始目标词信息，并基于所述至少一个初始目标词信息生成对应该候选词信息的目标词信息。所述至少一个初始目标词信息中的初始目标词信息包括：初始目标词和对应所述初始目标词的初始目标词概率值。

6、根据权利要求5所述的方法，其中，所述基于所述至少一个初始目标词信息生成对应该候选词信息的目标词信息，包括：

响应于初始目标词概率值最大的初始目标词信息对应的初始目标词属于所述历史目标词信息集，确定所述初始目标词概率值最大的初始目标词信息与所述历史目标词信息集中对应的历史目标词信息之间的重复词概率差；

响应于所述重复词概率差大于该候选词信息对应的其他初始目标词信息的初始目标词概率值，将初始目标词概率值最大的初始目标词

信息设置为目标词信息，并将所述目标词信息放入所述历史目标词信息集。

7、根据权利要求5所述的方法，其中，所述基于所述至少一个初始目标词信息生成对应该候选词信息的目标词信息，包括：

响应于初始目标词概率值最大的初始目标词信息对应的初始目标词属于所述历史目标词信息集，确定所述初始目标词概率值最大的初始目标词信息与所述历史目标词信息集中对应的历史目标词信息之间的重复词概率差；

响应于所述重复词概率差小于等于该候选词信息对应的其他初始目标词信息的初始目标词概率值，将其他初始目标词信息中的初始目标词概率值最大的初始目标词信息设置为目标词信息，并将所述目标词信息放入所述历史目标词信息集。

8、根据权利要求5-7之一所述的方法，其中，所述基于所述至少一个初始目标词信息生成对应该候选词信息的目标词信息，包括：

响应于初始目标词概率值最大的初始目标词信息对应的初始目标词不属于所述历史目标词信息集，将初始目标词概率值最大的初始目标词信息设置为目标词信息，并将所述目标词信息放入所述历史目标词信息集。

9、根据权利要求1-8之一所述的方法，其中，所述对于所述目标词信息集中的每个目标词信息，响应于确定所述目标词信息满足收敛条件，基于预设词表，将所述目标词信息和所述目标词信息所对应的历史目标词信息序列进行拼接以生成目标文本，包括：

对于所述目标词信息集中的每个目标词信息，响应于确定所述目标词信息满足预设收敛条件，将所述目标词信息与所述目标词信息所对应的历史目标词信息序列进行拼接以生成拼接文本；

从所述预设词表选择与所述拼接文本相匹配的词作为连接词；

对所述连接词和所述拼接文本进行组合以生成目标文本。

10、一种文本生成装置，包括：

确定单元，被配置成基于待处理文本，确定解码词信息组集，其中，所述待处理文本用于描述指定物体；

去重单元，被配置成对所述解码词信息组集进行去重处理以生成候选词信息组集；

选择单元，被配置成从所述候选词信息组集中的每个候选词信息组选择满足目标条件的候选词信息作为目标词信息，得到目标词信息集；

拼接单元，被配置成对于所述目标词信息集中的每个目标词信息，响应于确定所述目标词信息满足收敛条件，基于预设词表，将所述目标词信息和所述目标词信息所对应的历史目标词信息序列进行拼接以生成目标文本。

11、一种电子设备，包括：

一个或多个处理器；

存储装置，其上存储有一个或多个程序；

当所述一个或多个程序被所述一个或多个处理器执行，使得所述一个或多个处理器实现如权利要求 1-9 中任一所述的方法。

12、一种计算机可读介质，其上存储有计算机程序，其中，所述程序被处理器执行时实现如权利要求 1-9 任一所述的方法。