



(12)发明专利申请

(10)申请公布号 CN 110990556 A

(43)申请公布日 2020.04.10

(21)申请号 201911245157.X

G06F 16/36(2019.01)

(22)申请日 2019.12.06

(71)申请人 北京金山数字娱乐科技有限公司  
地址 100085 北京市海淀区小营西路33号  
金山软件大厦2层西区

申请人 成都金山互动娱乐科技有限公司  
北京金山软件有限公司

(72)发明人 刘晓楠 郭昱 汪美玲 李长亮

(74)专利代理机构 北京智信禾专利代理有限公司  
11637

代理人 王治东

(51)Int.Cl.

G06F 16/335(2019.01)

G06F 16/33(2019.01)

G06F 16/338(2019.01)

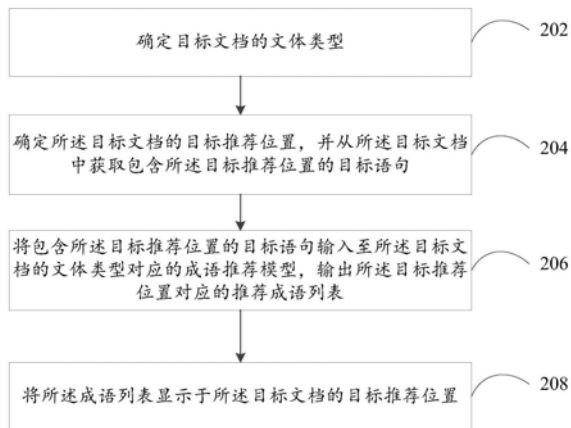
权利要求书2页 说明书12页 附图6页

(54)发明名称

成语推荐方法及装置、成语推荐模型的训练方法及装置

(57)摘要

本申请提供成语推荐方法及装置、成语推荐模型的训练方法及装置,其中成语推荐方法包括:确定目标文档的文体类型;确定目标文档的目标推荐位置,并从目标文档中获取包含目标推荐位置的目标语句;将包含所述目标推荐位置的目标语句输入至目标文档的文体类型对应的成语推荐模型,输出目标推荐位置对应的推荐成语列表,其中,推荐成语列表包括至少一个推荐成语;将成语列表显示于目标文档的目标推荐位置,从而可以根据文体类型生成合适的推荐成语列表,方便用户快捷准确地找到适用于当前文体类型和语境的成语,无需切换到第三方工具、使用户查找成语的过程智能化、大幅度缩短选用成语的路径,并保证推荐成语的准确性。



1. 一种成语推荐方法,其特征在于,包括:
  - 确定目标文档的文体类型;
  - 确定所述目标文档的目标推荐位置,并从所述目标文档中获取包含所述目标推荐位置的目标语句;
  - 将包含所述目标推荐位置的目标语句输入至所述目标文档的文体类型对应的成语推荐模型,输出所述目标推荐位置对应的推荐成语列表,其中,所述推荐成语列表包括至少一个推荐成语;
  - 将所述成语列表显示于所述目标文档的目标推荐位置。
2. 如权利要求1所述的成语推荐方法,其特征在于,从所述目标文档中获取包含所述目标推荐位置的目标语句,包括:
  - 根据设定的窗口大小,确定所述目标推荐位置的上下文语句,并将包含所述目标推荐位置的上下文语句确定为所述目标语句。
3. 如权利要求1所述的成语推荐方法,其特征在于,将包含所述目标推荐位置的目标语句输入至所述目标文档的文体类型对应的成语推荐模型,输出所述目标推荐位置对应的推荐成语列表,包括:
  - 将所述目标推荐位置替换为掩膜标签;
  - 将携带有所述掩膜标签的目标语句输入至所述目标文档的文体类型对应的成语推荐模型,生成语句向量,其中,所述语句向量包括与所述掩膜标签对应的掩膜向量;
  - 所述成语推荐模型对所述语句向量进行处理,得到所述掩膜向量对应的至少一个成语以及每个所述成语对应的分值;
  - 将分值高于设定阈值的成语作为推荐成语,生成并输出所述推荐成语列表。
4. 如权利要求1所述的成语推荐方法,其特征在于,还包括:
  - 在所述目标语句包含其他待推荐位置的情况下,将所述其他待推荐位置生成占位符。
5. 一种成语推荐模型的训练方法,其特征在于,所述方法包括:
  - 获取含有成语的训练文本,并确定所述训练文本的文体类型;
  - 选取每一文体类型的训练文本中含有待推荐成语的语句,将所述待推荐成语所在位置生成空位,生成训练语句;
  - 从预先生成的成语知识图谱中根据所述待推荐成语查找对应的相关成语,并将所述待推荐成语及其相关成语作为答案选项;
  - 将所述训练语句和所述答案选项输入至成语推荐模型进行训练,得到每一文体类型对应的成语推荐模型。
6. 如权利要求5所述的训练方法,其特征在于,所述成语知识图谱通过以下方法生成:
  - 从预设的语料数据库中获取结构化数据,所述结构化数据包括多个成语、多个特征标签、所述成语与特征标签之间的标签关系信息以及多个所述成语之间的近义关系信息;
  - 根据所述结构化数据构建成语知识图谱,以使所述成语知识图谱中包含有互为近义关系的成语以及每个所述成语对应的至少一个特征标签。
7. 如权利要求6所述的训练方法,其特征在于,从预先生成的成语知识图谱中根据所述待推荐成语查找对应的相关成语,包括:
  - 从预先生成的成语知识图谱中根据所述待推荐成语查找对应的至少一个近义关系的

成语和至少一个随机成语。

8. 如权利要求5所述的训练方法,其特征在于,将所述待推荐成语所在位置生成空位,生成训练语句,包括:

将所述待推荐成语对应的所述空位替换为掩膜标签,生成所述训练语句。

9. 如权利要求5或8所述的训练方法,其特征在于,选取每一文体类型的训练文本中含有待推荐成语的语句,将所述待推荐成语所在位置生成空位,生成训练语句,包括:

在所述语句包含至少两个所述待推荐成语的情况下,将任一所述待推荐成语所在位置生成掩膜标签,将其他所述待推荐成语的位置生成占位符,生成每个所述待推荐成语对应的训练语句。

10. 如权利要求8所述的训练方法,其特征在于,将所述训练语句和所述答案选项输入至成语推荐模型进行训练,包括:

将携带有所述掩膜标签的训练语句输入至所述成语推荐模型,生成训练语句向量,其中,所述训练语句向量包括与所述掩膜标签对应的掩膜向量;

使所述成语推荐模型依次将答案选项中的每个成语对应的成语向量添加至掩膜向量处,得到所述答案选项中的每个成语对应的分值;

根据所述答案选项中的每个成语对应的分值对所述成语推荐模型进行训练,直至所述待推荐成语对应的分值大于所述待推荐成语的相关成语对应的分值。

11. 一种成语推荐装置,其特征在于,包括:

文体类型确定模块,被配置为确定目标文档的文体类型;

目标语句获取模块,被配置为确定所述目标文档的目标推荐位置,并从所述目标文档中获取包含所述目标推荐位置的目标语句;

推荐成语输出模块,被配置为将包含所述目标推荐位置的目标语句输入至所述目标文档的文体类型对应的成语推荐模型,输出所述目标推荐位置对应的推荐成语列表,其中,所述推荐成语列表包括至少一个推荐成语;

显示加载模块,被配置为将所述成语列表显示于所述目标文档的目标推荐位置。

12. 一种成语推荐模型的训练装置,其特征在于,包括:

训练文本获取模块,被配置为获取含有成语的训练文本,并确定所述训练文本的文体类型;

训练语句生成模块,被配置为选取每一文体类型的训练文本中含有待推荐成语的语句,将所述待推荐成语所在位置生成空位,生成训练语句;

答案选项确定模块,被配置为从预先生成的成语知识图谱中根据所述待推荐成语查找对应的相关成语,并将所述待推荐成语及其相关成语作为答案选项;

训练模块,被配置为将所述训练语句和所述答案选项输入至成语推荐模型进行训练,得到每一文体类型对应的成语推荐模型。

13. 一种计算设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机指令,其特征在于,所述处理器执行所述指令时实现权利要求1-4或5-10任意一项所述方法的步骤。

14. 一种计算机可读存储介质,其存储有计算机指令,其特征在于,该指令被处理器执行时实现权利要求1-4或5-10任意一项所述方法的步骤。

## 成语推荐方法及装置、成语推荐模型的训练方法及装置

### 技术领域

[0001] 本申请涉及计算机技术领域,特别涉及成语推荐方法及装置、成语推荐模型的训练方法及装置、计算设备和计算机可读存储介质。

### 背景技术

[0002] 在文字撰写过程中,经常会用到成语的使用。现有的网络成语词典,例如百度百科、辞海、汉典,主要提供了成语的读音、释义、出处、近义词、反义词等信息,通常采用关系型数据库组织和存储。在此基础上,网络成语词典可为用户提供的推荐适合当前上下文语境的成语的方式为:通过用户自己对于上下文的理解,对于特定成语进行搜索,进而查看成语及其相关成语的信息,然后用户对于成语的释义及关系进行了解和判断,选择出合适的成语。

[0003] 在写作过程中,作者已经撰写部分文章或是要修改已完成文章的情况下,当用户想在文章中的某一位置添加合适的成语时,必须对于当前文体的上下文语境进行判断,然后切换到第三方的搜索、字典等工具。这些工具目前主要支持针对输入的成语返回相关成语信息,需要用户对于这些成语信息进行一一辨别,这就要求用户对工具返回的成语进行较多的甄别、筛选,极大破坏了用户关于文档写作思路的连续性。

### 发明内容

[0004] 有鉴于此,本申请实施例提供了一种成语推荐方法及装置、成语推荐模型的训练方法及装置、计算设备和计算机可读存储介质,以解决现有技术中存在的技术缺陷。

[0005] 本申请实施例提供了一种成语推荐方法,包括:

[0006] 确定目标文档的文体类型;

[0007] 确定所述目标文档的目标推荐位置,并从所述目标文档中获取包含所述目标推荐位置的目标语句;

[0008] 将包含所述目标推荐位置的目标语句输入至所述目标文档的文体类型对应的成语推荐模型,输出所述目标推荐位置对应的推荐成语列表,其中,所述推荐成语列表包括至少一个推荐成语;

[0009] 将所述成语列表显示于所述目标文档的目标推荐位置。

[0010] 可选地,从所述目标文档中获取包含所述目标推荐位置的目标语句,包括:

[0011] 根据设定的窗口大小,确定所述目标推荐位置的上下文语句,并将包含所述目标推荐位置的上下文语句确定为所述目标语句。

[0012] 可选地,将包含所述目标推荐位置的目标语句输入至所述目标文档的文体类型对应的成语推荐模型,输出所述目标推荐位置对应的推荐成语列表,包括:

[0013] 将所述目标推荐位置替换为掩膜标签;

[0014] 将携带有所述掩膜标签的目标语句输入至所述目标文档的文体类型对应的成语推荐模型,生成语句向量,其中,所述语句向量包括与所述掩膜标签对应的掩膜向量;

- [0015] 所述成语推荐模型对所述语句向量进行处理,得到所述掩膜向量对应的至少一个成语以及每个所述成语对应的分值;
- [0016] 将分值高于设定阈值的成语作为推荐成语,生成并输出所述推荐成语列表。
- [0017] 可选地,所述成语推荐方法还包括:在所述目标语句包含其他待推荐位置的情况下,将所述其他待推荐位置生成占位符。
- [0018] 本申请实施例提供了一种成语推荐模型的训练方法,所述方法包括:
- [0019] 获取含有成语的训练文本,并确定所述训练文本的文体类型;
- [0020] 选取每一文体类型的训练文本中含有待推荐成语的语句,将所述待推荐成语所在位置生成空位,生成训练语句;
- [0021] 从预先生成的成语知识图谱中根据所述待推荐成语查找对应的相关成语,并将所述待推荐成语及其相关成语作为答案选项;
- [0022] 将所述训练语句和所述答案选项输入至成语推荐模型进行训练,得到每一文体类型对应的成语推荐模型。
- [0023] 可选地,所述成语知识图谱通过以下方法生成:
- [0024] 从预设的语料数据库中获取结构化数据,所述结构化数据包括多个成语、多个特征标签、所述成语与特征标签之间的标签关系信息以及多个所述成语之间的近义关系信息;
- [0025] 根据所述结构化数据构建成语知识图谱,以使所述成语知识图谱中包含有互为近义关系的成语以及每个所述成语对应的至少一个特征标签。
- [0026] 可选地,从预先生成的成语知识图谱中根据所述待推荐成语查找对应的相关成语,包括:从预先生成的成语知识图谱中根据所述待推荐成语查找对应的至少一个近义关系的成语和至少一个随机成语。
- [0027] 可选地,将所述待推荐成语所在位置生成空位,生成训练语句,包括:
- [0028] 将所述待推荐成语对应的所述空位替换为掩膜标签,生成所述训练语句。
- [0029] 可选地,选取每一文体类型的训练文本中含有待推荐成语的语句,将所述待推荐成语所在位置生成空位,生成训练语句,包括:
- [0030] 在所述语句包含至少两个所述待推荐成语的情况下,将任一所述待推荐成语所在位置生成掩膜标签,将其他所述待推荐成语的位置生成占位符,生成每个所述待推荐成语对应的训练语句。
- [0031] 可选地,将所述训练语句和所述答案选项输入至成语推荐模型进行训练,包括:将携带有所述掩膜标签的训练语句输入至所述成语推荐模型,生成训练语句向量,其中,所述训练语句向量包括与所述掩膜标签对应的掩膜向量;
- [0032] 使所述成语推荐模型依次将答案选项中的每个成语对应的成语向量添加至掩膜向量处,得到所述答案选项中的每个成语对应的分值;
- [0033] 根据所述答案选项中的每个成语对应的分值对所述成语推荐模型进行训练,直至所述待推荐成语对应的分值大于所述待推荐成语的相关成语对应的分值。
- [0034] 本申请实施例公开了一种成语推荐装置,包括:
- [0035] 文体类型确定模块,被配置为确定目标文档的文体类型;
- [0036] 目标语句获取模块,被配置为确定所述目标文档的目标推荐位置,并从所述目标

文档中获取包含所述目标推荐位置的目标语句；

[0037] 推荐成语输出模块,被配置为将包含所述目标推荐位置的目标语句输入至所述目标文档的文体类型对应的成语推荐模型,输出所述目标推荐位置对应的推荐成语列表,其中,所述推荐成语列表包括至少一个推荐成语；

[0038] 显示加载模块,被配置为将所述成语列表显示于所述目标文档的目标推荐位置。

[0039] 本申请实施例公开了一种成语推荐模型的训练装置,包括：

[0040] 训练文本获取模块,被配置为获取含有成语的训练文本,并确定所述训练文本的文体类型；

[0041] 训练语句生成模块,被配置为选取每一文体类型的训练文本中含有待推荐成语的语句,将所述待推荐成语所在位置生成空位,生成训练语句；

[0042] 答案选项确定模块,被配置为从预先生成的成语知识图谱中根据所述待推荐成语查找对应的相关成语,并将所述待推荐成语及其相关成语作为答案选项；

[0043] 训练模块,被配置为将所述训练语句和所述答案选项输入至成语推荐模型进行训练,得到每一文体类型对应的成语推荐模型。

[0044] 本申请实施例公开了一种计算设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机指令,所述处理器执行所述指令时实现如上所述的成语推荐方法或成语推荐模型的训练方法的步骤。

[0045] 本申请实施例公开了一种计算机可读存储介质,其存储有计算机指令,该指令被处理器执行时实现如上所述的成语推荐方法或成语推荐模型的训练方法的步骤。

[0046] 本申请提供的成语推荐方法及装置,在需要获得推荐成语的情况下,确定目标文档的文体类型和目标推荐位置,并从目标文档中获取包含目标推荐位置的目标语句,将目标语句输入至成语推荐模型便得到目标待推荐位置对应的推荐成语列表,从而可以根据文体类型生成合适的推荐成语列表,使用户根据推荐成语列表选择合适的成语并使用,方便用户快捷准确地找到适用于当前文体类型和语境的成语,无需切换到第三方工具、也无需对推荐成语可行性进行判断,使用户查找成语的过程智能化、大幅度缩短选用成语的路径,并保证推荐成语的准确性。

[0047] 另外,本申请的成语推荐模型的训练方法及装置,通过选取每一文体类型的训练文本中含有待推荐成语的语句生成训练语句,从成语知识图谱中根据待推荐成语查找对应的相关成语,并将待推荐成语及其相关成语作为答案选项,然后将训练语句和答案选项输入至成语推荐模型进行训练,从而可以获得每种文体类型对应的成语推荐模型,有利于成语推荐模型更加准确地给出成语推荐结果。

[0048] 另外,本申请的成语推荐模型的训练方法及装置,从预先生成的成语知识图谱中根据所述待推荐成语查找对应的相关成语,并将所述待推荐成语及其相关成语作为答案选项来训练成语推荐模型,从而可以通过基于特征标签的成语知识图谱支持用户可以从多侧面获取待推荐成语的信息,得到更准确的成语推荐结果。

## 附图说明

[0049] 图1是本申请一实施例的成语推荐方法的示意框图；

[0050] 图2是本申请一实施例的成语推荐方法的流程示意图；

- [0051] 图3是本申请一实施例的成语推荐方法中生成推荐成语列表的流程示意图；
- [0052] 图4是本申请另一实施例的成语推荐模型的训练方法示意图；
- [0053] 图5是本申请另一实施例的成语推荐装置的模块示意图；
- [0054] 图6是本申请另一实施例的成语推荐模型的训练装置的模块示意图；
- [0055] 图7是本申请另一实施例的计算设备的结构示意图。

## 具体实施方式

[0056] 在下面的描述中阐述了很多具体细节以便于充分理解本申请。但是本申请能够以很多不同于在此描述的其它方式来实施，本领域技术人员可以在不违背本申请内涵的情况下做类似推广，因此本申请不受下面公开的具体实施的限制。

[0057] 在本说明书一个或多个实施例中使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本说明书一个或多个实施例。在本说明书一个或多个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。还应当理解，本说明书一个或多个实施例中使用的术语“和/或”是指并包含一个或多个相关联的列出项目的任何或所有可能组合。

[0058] 应当理解，尽管在本说明书一个或多个实施例中可能采用术语第一、第二等来描述各种信息，但这些信息不应限于这些术语。这些术语仅用来将同一类型的信息彼此区分开。例如，在不脱离本说明书一个或多个实施例范围的情况下，第一也可以被称为第二，类似地，第二也可以被称为第一。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”。

[0059] 首先，对本发明一个或多个实施例涉及的名词术语进行解释。

[0060] 知识图谱：知识图谱旨在描述真实世界中存在的各种实体或概念及其关系，其构成一张巨大的语义网络图，节点表示实体或概念，边则由属性或关系构成。

[0061] 文体类型：文章的类型，如小说、散文、新闻等。

[0062] BERT (Bidirectional Encoder Representation from Transformers, 双向注意力神经网络) 模型：一种双向注意力神经网络模型。BERT模型可以通过左、右两侧上下文来预测当前词和通过当前句子预测下一个句子。BERT模型的目标是利用大规模无标注语料训练、获得文本的包含丰富语义信息的文本的语义表示，然后将文本的语义表示在特定NLP任务中作微调，最终应用于该NLP任务。

[0063] 掩膜标签：Mask，即BERT模型中对候选词所处位置进行遮盖的标签。

[0064] 在本申请中，提供了一种成语推荐方法及装置、成语推荐模型的训练方法及装置、计算设备和计算机可读存储介质，在下面的实施例中逐一进行详细说明。

[0065] 本实施例公开了一种成语推荐方法参见图1和图2。其中，图1示出了本实施例的成语推荐方法的示意框图。如图所示，对于目标文档，包括n个目标语句。对于成语推荐模型，包括A、B、C、D四个模型。其中，与该目标文档的文体类型对应的模型为成语推荐模型B。

[0066] 在成语推荐过程中，将目标语句1~n依次输入成语推荐模型B，得到目标语句1对应的成语推荐列表{推荐成语11, 推荐成语12, ……推荐成语1m}，目标语句2对应的成语推荐列表{推荐成语21, 推荐成语22, ……推荐成语2m}，……目标语句n对应的成语推荐列表{推荐成语n1, 推荐成语n2, ……推荐成语nm}。

[0067] 详细的步骤参见图2。图2是示出了根据本申请一实施例的成语推荐方法的示意性流程图,包括步骤202至步骤208。

[0068] 202、确定目标文档的文体类型。

[0069] 其中,文体类型可以为多种,例如小说、散文、诗歌等等。

[0070] 文体的不同,会导致使用成语的不同。例如对于散文,是以记人、叙事、写景或状物为主要内容,以叙述、描写为主要表达方式的一种文体;对于小说,以塑造人物形象为中心,通过故事情节叙述和环境描写反映社会生活,小说必须具备生动的人物形象、完整的情节和人物活动的具体环境这三个要素;诗歌是一种主情的文学体裁,它以抒情的方式,高度凝练,集中地反映社会生活,用丰富的想象、富有节奏感、韵律美的语言和分行排列的形式来抒发思想情感。

[0071] 例如在具体应用过程中,以一个目标语句为例,如果只根据上下文来判断,该成语可以用于该目标语句中,但是再进一步地根据文体类型来判断,该成语可以适用于小说,但是却不一定适用于诗歌。本实施例中通过引入文体类型的判断,可以在后续的成语推荐过程中实现更精准的成语推荐。

[0072] 在一种使用场景下,用户在撰写散文的过程中,如用户想要在某一位置使用恰当的成语,用户可以选择“散文”文体,并将针对“散文”文体语料训练好的成语推荐模型进行预加载。

[0073] 本实施例中,成语推荐模型可以为多个,例如包括对应散文的成语推荐模型、对应小说的成语推荐模型、对应诗歌的成语推荐模型等。

[0074] 以下面目标文档的一段语句为例:

[0075] 我喜欢冬,不是因为它的美,因为任何美都不一定永恒,就像(昙花一现);我喜欢冬,不是因为它的洁白,因为任何白都不一定(完美无瑕),就像白色的墙。我喜欢冬,是因为我知道只要冬天来了,春天就不远了。只要你在最寒冷的时候在坚持一下,一下就好,你就能看见迎春花开。所以,请同学们相信:只要冬天来了,春天就不远了!

[0076] 确定该目标文档的文体类型为“散文”,并将针对“散文”文体语料训练好的成语推荐模型进行预加载。

[0077] 204、确定所述目标文档的目标推荐位置,并从所述目标文档中获取包含所述目标推荐位置的目标语句。

[0078] 其中,目标推荐位置的确定方式有多种,例如通过文档中的输入光标位置确定目标推荐位置,或者根据用户的输入命令来确定文档中的目标推荐位置。

[0079] 另外,本实施例中,从目标文档中获取包含目标推荐位置的目标语句,包括多种方式。

[0080] 在一种具体方式下:根据设定的窗口大小,确定目标推荐位置的上下文语句,并将目标推荐位置的上下文语句确定为目标语句。

[0081] 在另一种具体方式下:根据标点符号,确定包含目标推荐位置的目标语句。

[0082] 其中,以根据设定窗口确定目标语句为例,窗口大小可以根据实际需求而设置,例如目标文档的每行文字为35,设置窗口大小为35\*1、35\*2等等。

[0083] 当前,包含目标推荐位置的目标语句也可以仅包括从目标推荐位置开始的下文,例如目标推荐位置为段首的情形;或者仅包括以目标推荐位置结束的上文,例如目标推荐



位置为段尾的情形。为了更加准确地判断出目标语句的语义,应当尽量选择包括目标推荐位置的上下文语句。

[0084] 仍以上述散文目标文档为例,由目标文档可见,需要推荐成语的推荐位置一共有两处,参见上述目标文档的加括号处。括号内为添加于该推荐位置的最优推荐成语。

[0085] 设置窗口大小为35\*1,那么,对于第一个推荐位置,目标语句为“我喜欢冬,不是因为它的美,因为任何美都不一定永恒,就像(\*\*\*\*)” ;对于第二个推荐位置,目标语句为“我喜欢冬,不是因为它的洁白,因为任何白都不一定(\*\*\*\*),就像白色的”。

[0086] 206、将包含所述目标推荐位置的目标语句输入至所述目标文档的文体类型对应的成语推荐模型,输出所述目标推荐位置对应的推荐成语列表。

[0087] 其中,所述推荐成语列表包括至少一个推荐成语。

[0088] 具体地,参见图3,步骤206包括下述步骤2062~2068:

[0089] 2062、将所述目标推荐位置替换为掩膜标签。

[0090] 本实施例中,掩膜标签(Mask)即成语推荐模型对目标推荐位置进行遮盖的标签。

[0091] 在确定目标推荐位置后,将该位置替换为掩膜标签[mask],以标识此处需要添加推荐成语。

[0092] 2064、将携带有所述掩膜标签的目标语句输入至所述目标文档的文体类型对应的成语推荐模型,生成语句向量。

[0093] 其中,所述语句向量包括与所述掩膜标签对应的掩膜向量。

[0094] 并且,成语推荐模型可以为多种,本实施例优选成语推荐模型为BERT模型。并且,每种文体类型对应于一个预先训练好的成语推荐模型,并且保存于后台,在接收到加载命令的情况下被加载。

[0095] 2066、所述成语推荐模型对所述语句向量进行处理,得到所述掩膜向量对应的至少一个成语以及每个所述成语对应的分值。

[0096] 例如对于目标推荐位置1,成语推荐模型生成的成语包括{成语1,成语2,成语3,成语4},对应的得分分别为{0.8,0.5,0.3,0.2}。

[0097] 2068、将分值高于设定阈值的成语作为推荐成语,生成并输出所述推荐成语列表。

[0098] 本实施例中,阈值可以根据实际需求而设置,例如设置为0.25。那么对于得分分别为{0.8,0.5,0.3,0.2}的{成语1,成语2,成语3,成语4},得到的推荐成语列表为{成语1,成语2,成语3}。

[0099] 以上述步骤202中的段落为例,将目标语句“我喜欢冬,不是因为它的美,因为任何美都不一定永恒,就像(\*\*\*\*)”生成“我喜欢冬,不是因为它的美,因为任何美都不一定永恒,就像[mask]”输入至散文对应的成语推荐模型,输出目标推荐位置对应的推荐成语列表:{好景不长,昙花一现,旷世难逢}。

[0100] 将目标语句“我喜欢冬,不是因为它的洁白,因为任何白都不一定(\*\*\*\*),就像白色的”生成“我喜欢冬,不是因为它的洁白,因为任何白都不一定[mask],就像白色的”,然后输入至散文对应的成语推荐模型,输出目标推荐位置对应的推荐成语列表:{完美无瑕,十全十美,天衣无缝}。

[0101] 当然,为本领域技术人员所知晓的,在本步骤2068中,一种可替代的处理方法包括:将分值最高的成语作为推荐成语输出。此种方式下,返回的推荐成语列表仅包括一个推

荐成语,也即成语推荐模型得到的与目标文档的上下文语境最贴切的成语。

[0102] 可选地,对于一个目标语句包含其他待推荐位置的情况下,将其他待推荐位置生成占位符,以保证每次处理的过程中,只有一处待推荐位置需要被添加推荐成语。

[0103] 以语句“在误解面前,要尽量把原原委委讲清楚,不(矢口狡辩)(委罪于人)、更不能(矢口否认)(委曲求全),澄清事实,是解脱自己的最佳选择!”为例:

[0104] 对于第一次的成语推荐任务,生成的目标语句为“在误解面前,要尽量把原原委委讲清楚,不[mask][unk],更不能[unk][unk],澄清事实,是解脱自己的最佳选择”。成语推荐模型得到的成语推荐列表为{矢口狡辩,唯唯诺诺,小心翼翼},得分分别为{0.8,0.6,0.5},则掩膜标签处对应的推荐成语为“矢口狡辩”。

[0105] 对于第二次的成语推荐任务,生成的目标语句为“在误解面前,要尽量把原原委委讲清楚,不矢口狡辩[mask],更不能[unk][unk],澄清事实,是解脱自己的最佳选择”。成语推荐模型得到的成语推荐列表为{委罪于人,得寸进尺,小心翼翼},得分分别为{0.8,0.5,0.3},则掩膜标签处对应的推荐成语为“委罪于人”。

[0106] 对于第三次的成语推荐任务,生成的目标语句为“在误解面前,要尽量把原原委委讲清楚,不矢口狡辩委罪于人,更不能[mask][unk],澄清事实,是解脱自己的最佳选择”。成语推荐模型得到的成语推荐列表为{颠倒黑白,矢口否认,指鹿为马},得分分别为{0.2,0.8,0.5},则掩膜标签处对应的推荐成语为“矢口否认”。

[0107] 对于第四次的成语推荐任务,生成的目标语句为“在误解面前,要尽量把原原委委讲清楚,不矢口狡辩委罪于人,更不能矢口否认[mask],澄清事实,是解脱自己的最佳选择”。成语推荐模型得到的成语推荐列表为{委曲求全,矢口否认,指鹿为马},得分分别为{0.6,0.2,0.5},则掩膜标签处对应的推荐成语为“委曲求全”。

[0108] 208、将所述成语列表显示于所述目标文档的目标推荐位置。

[0109] 本步骤中,显示的方式可以为多种,在一种具体使用场景下,可以在目标推荐位置的下方显示出一个可供用户通过点选命令进行选择的列表,以显示本实施例中的成语列表。

[0110] 在另一种具体使用场景下,也可以通过弹出对话框的形式来显示本实施例中的成语列表。

[0111] 在又一种具体使用场景下,也可以将分值最高的成语高亮显示于目标推荐位置,并在收到“确定”命令后将该成语加载于目标推荐位置中。

[0112] 本申请提供的成语推荐方法,在需要获得推荐成语的情况下,确定目标文档的文体类型和目标推荐位置,并从目标文档中获取包含目标推荐位置的目标语句,将目标语句输入至成语推荐模型便得到目标待推荐位置对应的推荐成语列表,可以根据文体类型生成合适的推荐成语列表,使用户根据推荐成语列表选择合适的成语并使用,方便用户快捷准确地找到适用于当前文体类型和语境的成语,无需切换到第三方工具、也无需对推荐成语可行性进行判断,使用户查找成语的过程智能化、大幅度缩短选用成语的路径,并保证推荐成语的准确性。

[0113] 参见图4,本申请一实施例还提供一种成语推荐模型的训练方法,包括下述步骤402~408:

[0114] 402、获取含有成语的训练文本,并确定所述训练文本的文体类型。

[0115] 其中,文体类型可以为多种,例如小说、散文、诗歌等等。本实施例中,可以对训练文本添加文体类型的标签。

[0116] 404、选取每一文体类型的训练文本中含有待推荐成语的语句,将所述待推荐成语所在位置生成空位,生成训练语句。

[0117] 具体地,步骤404包括:将待推荐成语对应的空位替换为掩膜标签,生成所述训练语句。

[0118] 可选地,对于语句包含至少两个待推荐成语的情况,步骤404包括:将任一待推荐成语所在位置生成掩膜标签,将其他待推荐成语的位置生成占位符,生成每个待推荐成语对应的训练语句。

[0119] 以训练文本“如果你曾经历过(天寒地冻)的冬天,那么你就会有(春色满园)的春天!如果你有着(坚韧不拔)的信念,那么春天一定不会遥远;如果你正在(勤勤恳恳)的付出,那么总有一天你会拥有(春华秋实)”为例,生成的训练语句包括:

[0120] 1)、如果你曾经历过[mask]的冬天,那么你就会有[unk]的春天!

[0121] 2)如果你曾经历过[unk]的冬天,那么你就会有[mask]的春天!

[0122] 3)、如果你有着[mask]的信念,那么春天一定不会遥远。

[0123] 4)、如果你正在[mask]的付出,那么总有一天你会拥有[unk]。

[0124] 5)、如果你正在[unk]的付出,那么总有一天你会拥有[mask]。

[0125] 406、从预先生成的成语知识图谱中根据所述待推荐成语查找对应的相关成语,并将所述待推荐成语及其相关成语作为答案选项。

[0126] 具体地,成语知识图谱通过以下方法S4062~S4064生成:

[0127] S4062、从预设的语料数据库中获取结构化数据,所述结构化数据包括多个成语、多个特征标签、所述成语与特征标签之间的标签关系信息以及多个所述成语之间的近义关系信息。

[0128] S4064、根据所述结构化数据构建成语知识图谱,以使所述成语知识图谱中包含有互为近义关系的成语以及每个所述成语对应的至少一个特征标签。

[0129] 其中,每个成语对应的特征标签有多种,例如拼音、释义、出处、朝代、成语故事等信息。

[0130] 例如,对于目标成语“沉鱼落雁”和其对应的特征标签“容貌”,系统根据预设的推荐条件在预设的语料数据库中进行检索获取结构化数据,以获取“容貌”对应的近义词“冰肌玉骨”、“明眸皓齿”和“出水芙蓉”,并将上述成语作为目标成语“沉鱼落雁”对应的近义词成语。

[0131] 在构建的所述成语知识图谱中存在近义词关系,假设目标成语B和成语实体C是所述成语知识图谱中的成语,并且目标成语B和成语实体C之间的关系为“近义词S”则可以表示为三元组(B,S,C)的联结关系,即在满足所述推荐条件下目标成语B和成语实体C在目标文档中可以进行替换,例如,将目标文档为“小芳的容貌可谓是沉鱼落雁”中的“沉鱼落雁”替换为近义词“明眸皓齿”,并不影响所述目标文档的整体表达;此外,在目标成语B与成语实体C互为近义词的情况下,目标成语B与成语实体C之间拥有至少一个相同的所述特征标签r,则同样可以以特征标签r为关系表示为近义词三元组(B,r,C)的联结关系,例如,成语“沉鱼落雁”与成语“闭月羞花”都是用于形容特征标签“容貌”的成语,则成语“沉鱼落雁”、

成语“明眸皓齿”与特征标签“容貌”之间可以表示为三元组(沉鱼落雁,容貌,明眸皓齿)的联结关系。

[0132] 另外,步骤406中,从预先生成的成语知识图谱中根据所述待推荐成语查找对应的成语,包括:从预先生成的成语知识图谱中根据待推荐成语查找对应的至少一个近义关系的成语和至少一个随机成语,从而作为包括正确选项和错误选项的答案选项,以训练成语推荐模型。

[0133] 需要说明的是,对于近义关系的成语的数量不宜过多,过多的情况下会导致模型的训练结果变差。所以,近义关系的成语一般为1~2个;对于其他随机成语,需要无条件随机且不重复即可。

[0134] 以上述示例为例,若待推荐成语为“沉鱼落雁”,则挑选“明眸皓齿”、“虎背熊腰”、“出水芙蓉”作为答案选项。

[0135] 408、将所述训练语句和所述答案选项输入至成语推荐模型进行训练,得到每一文体类型对应的成语推荐模型。

[0136] 具体地,步骤408的训练过程包括:

[0137] S4082、将携带有所述掩膜标签的训练语句输入至所述成语推荐模型,生成训练语句向量,其中,所述训练语句向量包括与所述掩膜标签对应的掩膜向量。

[0138] 以目标语句“小芳的容貌可谓(沉鱼落雁)”为例,得到的训练语句“小芳的容貌可谓[mask]”,答案选项为{沉鱼落雁、明眸皓齿、虎背熊腰、出水芙蓉}。将该训练语句输入至成语推荐模型,生成包括掩膜向量的训练语句向量。

[0139] S4084、使所述成语推荐模型依次将答案选项中的每个成语对应的成语向量添加至掩膜向量处,得到所述答案选项中的每个成语对应的分值。

[0140] S4086、根据所述答案选项中的每个成语对应的分值对所述成语推荐模型进行训练,直至所述待推荐成语对应的分值大于所述待推荐成语的相关成语对应的分值。

[0141] 仍以目标语句“小芳的容貌可谓是[mask]”为例,依次将{沉鱼落雁、明眸皓齿、虎背熊腰、出水芙蓉}对应的成语向量添加至掩膜向量处,得到所述答案选项中的每个成语对应的分值分别为{0.7,0.7,0.6,0.7}。

[0142] 对成语推荐模型进行训练,直至答案选项中的每个成语对应的分值分别为{0.8,0.6,0.1,0.5},则认为此次训练达到训练目的。

[0143] 另外,本申请的成语推荐模型的训练方法,通过选取每一文体类型的训练文本中含有待推荐成语的语句生成训练语句,从成语知识图谱中根据待推荐成语查找对应的成语,并将待推荐成语及其相关成语作为答案选项,然后将训练语句和答案选项输入至成语推荐模型进行训练,从而可以获得每种文体类型对应的成语推荐模型,有利于成语推荐模型更加准确地给出成语推荐结果。

[0144] 另外,本申请的成语推荐模型的训练方法,从预先生成的成语知识图谱中根据所述待推荐成语查找对应的成语,并将所述待推荐成语及其相关成语作为答案选项来训练成语推荐模型,从而可以通过基于特征标签的成语知识图谱支持用户可以从多侧面获取待推荐成语的信息,得到更准确的成语推荐结果。

[0145] 本申请一实施例还提供一种成语推荐装置,参见图5,包括:

[0146] 文体类型确定模块502,被配置为确定目标文档的文体类型;

[0147] 目标语句获取模块504,被配置为确定所述目标文档的目标推荐位置,并从所述目标文档中获取包含所述目标推荐位置的目标语句;

[0148] 推荐成语输出模块506,被配置为将包含所述目标推荐位置的目标语句输入至所述目标文档的文体类型对应的成语推荐模型,输出所述目标推荐位置对应的推荐成语列表,其中,所述推荐成语列表包括至少一个推荐成语;

[0149] 显示加载模块508,被配置为将所述成语列表显示于所述目标文档的目标推荐位置。

[0150] 可选地,目标语句获取模块504具体被配置为:根据设定的窗口大小,确定所述目标推荐位置的上下文语句,并将包含所述目标推荐位置的上下文语句确定为所述目标语句。

[0151] 可选地,推荐成语输出模块506具体被配置为:将所述目标推荐位置替换为掩膜标签;

[0152] 将携带有所述掩膜标签的目标语句输入至所述目标文档的文体类型对应的成语推荐模型,生成语句向量,其中,所述语句向量包括与所述掩膜标签对应的掩膜向量;

[0153] 所述成语推荐模型对所述语句向量进行处理,得到所述掩膜向量对应的至少一个成语以及每个所述成语对应的分值;

[0154] 将分值高于设定阈值的成语作为推荐成语,生成并输出所述推荐成语列表

[0155] 可选地,所述装置还包括:目标语句处理模块,被配置为在所述目标语句包含其他待推荐位置的情况下,将所述其他待推荐位置生成占位符。

[0156] 本申请提供的成语推荐装置,在需要获得推荐成语的情况下,确定目标文档的文体类型和目标推荐位置,并从目标文档中获取包含目标推荐位置的目标语句,将目标语句输入至成语推荐模型得到目标待推荐位置对应的推荐成语列表,从而可以根据文体类型生成合适的推荐成语列表,使用户根据推荐成语列表选择合适的成语并使用,方便用户快捷准确地找到适用于当前文体类型和语境的成语,无需切换到第三方工具、也无需对推荐成语可行性进行判断,使用户查找成语的过程智能化、大幅度缩短选用成语的路径,并保证推荐成语的准确性。

[0157] 上述为本实施例的一种成语推荐装置的示意性方案。需要说明的是,该装置的技术方案与上述的成语推荐方法的技术方案属于同一构思,装置的技术方案未详细描述的细节内容,均可以参见上述成语推荐方法的技术方案的描述。

[0158] 本申请一实施例还提供一种成语推荐模型的训练装置,参见图6,包括:

[0159] 训练文本获取模块602,被配置为获取含有成语的训练文本,并确定所述训练文本的文体类型;

[0160] 训练语句生成模块604,被配置为选取每一文体类型的训练文本中含有待推荐成语的语句,将所述待推荐成语所在位置生成空位,生成训练语句;

[0161] 答案选项确定模块606,被配置为从预先生成的成语知识图谱中根据所述待推荐成语查找对应的相关成语,并将所述待推荐成语及其相关成语作为答案选项;

[0162] 训练模块608,被配置为将所述训练语句和所述答案选项输入至成语推荐模型进行训练,得到每一文体类型对应的成语推荐模型。

[0163] 可选地,所述装置还包括成语知识图谱生成模块,被配置为:

[0164] 从预设的语料数据库中获取结构化数据,所述结构化数据包括多个成语、多个特征标签、所述成语与特征标签之间的标签关系信息以及多个所述成语之间的近义关系信息;

[0165] 根据所述结构化数据构建成语知识图谱,以使所述成语知识图谱中包含有互为近义关系的成语以及每个所述成语对应的至少一个特征标签。

[0166] 可选地,答案选项确定模块606,具体被配置为:从预先生成的成语知识图谱中根据所述待推荐成语查找对应的至少一个近义关系的成语和至少一个随机成语。

[0167] 可选地,训练语句生成模块604,具体被配置为:将所述待推荐成语对应的所述空位替换为掩膜标签,生成所述训练语句。

[0168] 可选地,训练语句生成模块604,具体被配置为:在所述语句包含至少两个所述待推荐成语的情况下,将任一所述待推荐成语所在位置生成掩膜标签,将其他所述待推荐成语的位置生成占位符,生成每个待推荐成语对应的训练语句。

[0169] 可选地,训练模块608,具体被配置为:

[0170] 将携带有所述掩膜标签的训练语句输入至所述成语推荐模型,生成训练语句向量,其中,所述训练语句向量包括与所述掩膜标签对应的掩膜向量;

[0171] 使所述成语推荐模型依次将答案选项中的每个成语对应的成语向量添加至掩膜向量处,得到所述答案选项中的每个成语对应的分值;

[0172] 根据所述答案选项中的每个成语对应的分值对所述成语推荐模型进行训练,直至所述待推荐成语对应的分值大于所述待推荐成语的相关成语对应的分值。

[0173] 本申请的成语推荐模型的训练装置,通过选取每一文体类型的训练文本中含有待推荐成语的语句生成训练语句,从成语知识图谱中根据待推荐成语查找对应的相关成语,并将待推荐成语及其相关成语作为答案选项,然后将训练语句和答案选项输入至成语推荐模型进行训练,从而可以获得每种文体类型对应的成语推荐模型,有利于成语推荐模型更加准确地给出成语推荐结果。

[0174] 上述为本实施例的一种成语推荐模型的训练装置的示意性方案。需要说明的是,该装置的技术方案与上述的成语推荐模型的训练方法的技术方案属于同一构思,装置的技术方案未详细描述的细节内容,均可以参见上述成语推荐模型的训练方法的技术方案的描述。

[0175] 本申请一实施例还提供一种计算机可读存储介质,其存储有计算机指令,该指令被处理器执行时实现如前所述成语推荐方法或成语推荐模型的训练方法的步骤。

[0176] 上述为本实施例的一种计算机可读存储介质的示意性方案。需要说明的是,该存储介质的技术方案与上述的成语推荐方法或成语推荐模型的训练方法的技术方案属于同一构思,存储介质的技术方案未详细描述的细节内容,均可以参见上述成语推荐方法或成语推荐模型的训练方法的技术方案的描述。

[0177] 所述计算机指令包括计算机程序代码,所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括:能够携带所述计算机程序代码的任何实体或装置、记录介质、U盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、电载波信号、电信信号以及软件分发介质等。需要说明的是,所述计算机可读介质包含的内

容可以根据司法管辖区内立法和专利实践的要求进行适当的增减,例如在某些司法管辖区,根据立法和专利实践,计算机可读介质不包括电载波信号和电信信号。

[0178] 本申请一实施例还公开了一种计算设备,参见图7,包括存储器710、处理器720及存储在存储器710上并可在处理器上运行的计算机指令,所述处理器720执行所述指令时实现如上所述的成语推荐方法或成语推荐模型的训练方法的步骤。

[0179] 图7是示出了根据本说明书一实施例的计算设备700的结构框图。该计算设备700的部件包括但不限于存储器710和处理器720。处理器720与存储器710通过总线730相连接,数据库750用于保存数据。

[0180] 计算设备700还包括接入设备740,接入设备740使得计算设备700能够经由一个或多个网络760通信。这些网络的示例包括公用交换电话网(PSTN)、局域网(LAN)、广域网(WAN)、个域网(PAN)或诸如因特网的通信网络的组合。接入设备740可以包括有线或无线的任何类型的网络接口(例如,网络接口卡(NIC))中的一个或多个,诸如IEEE802.11无线局域网(WLAN)无线接口、全球微波互联接入(Wi-MAX)接口、以太网接口、通用串行总线(USB)接口、蜂窝网络接口、蓝牙接口、近场通信(NFC)接口,等等。

[0181] 在本说明书的一个实施例中,计算设备700的上述部件以及图7中未示出的其他部件也可以彼此相连接,例如通过总线。应当理解,图7所示的计算设备结构框图仅仅是出于示例的目的,而不是对本说明书范围的限制。本领域技术人员可以根据需要,增添或替换其他部件。

[0182] 计算设备700可以是任何类型的静止或移动计算设备,包括移动计算机或移动计算设备(例如,平板计算机、个人数字助理、膝上型计算机、笔记本计算机、上网本等)、移动电话(例如,智能手机)、可佩戴的计算设备(例如,智能手表、智能眼镜等)或其他类型的移动设备,或者诸如台式计算机或PC的静止计算设备。计算设备700还可以是移动式或静止式的服务器。

[0183] 需要说明的是,对于前述的各方法实施例,为了简便描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本申请并不受所描述的动作顺序的限制,因为依据本申请,某些步骤可以采用其它顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施例,所涉及的动作和模块并不一定是本申请所必须的。

[0184] 在上述实施例中,对各个实施例的描述都各有侧重,某个实施例中未详述的部分,可以参见其它实施例的相关描述。

[0185] 以上公开的本申请优选实施例只是用于帮助阐述本申请。可选实施例并没有详尽叙述所有的细节,也不限制该发明仅为所述的具体实施方式。显然,根据本说明书的内容,可作很多的修改和变化。本说明书选取并具体描述这些实施例,是为了更好地解释本申请的原理和实际应用,从而使所属技术领域技术人员能很好地理解和利用本申请。本申请仅受权利要求书及其全部范围和等效物的限制。

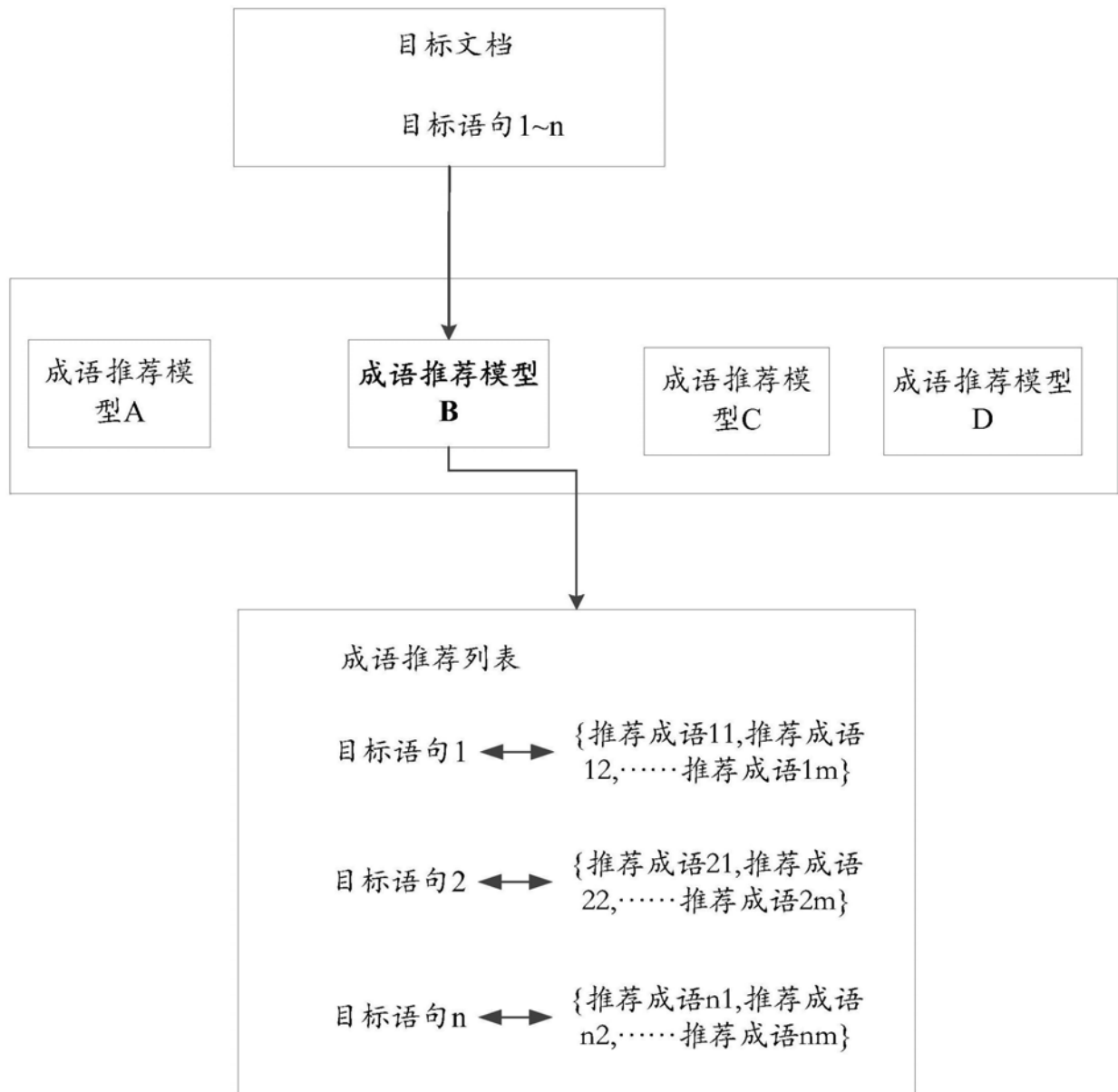


图1



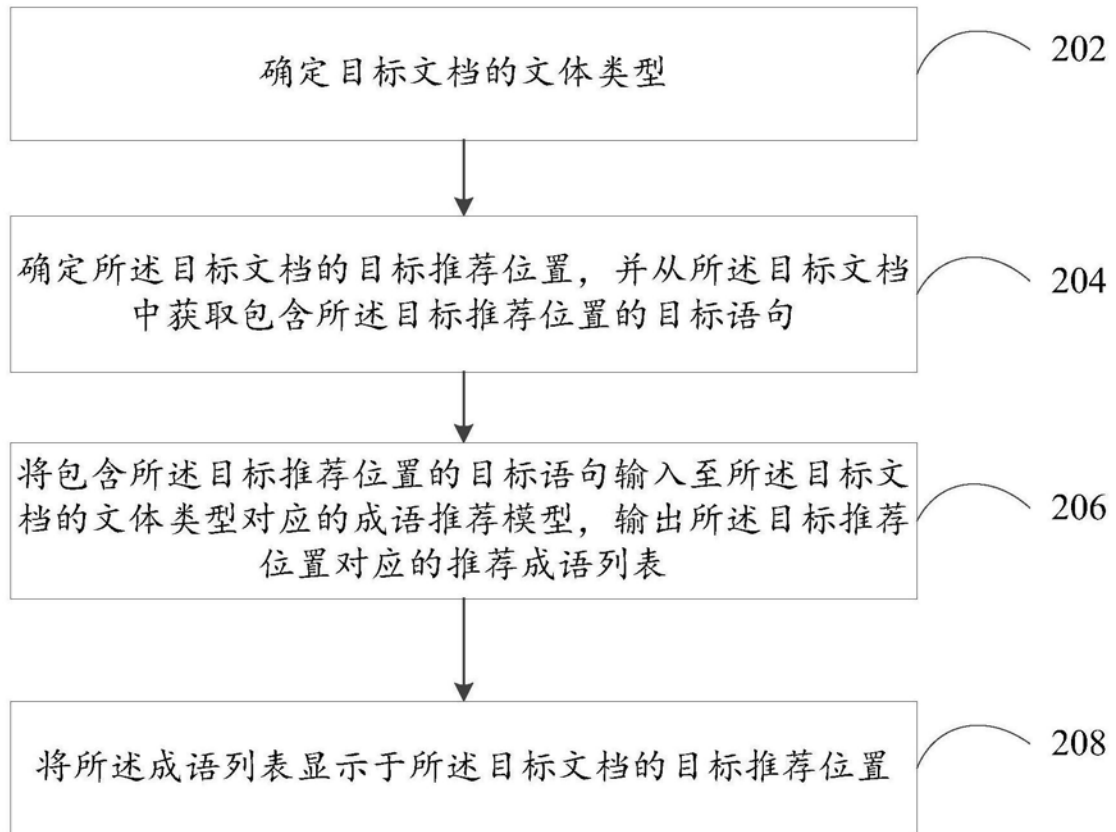


图2

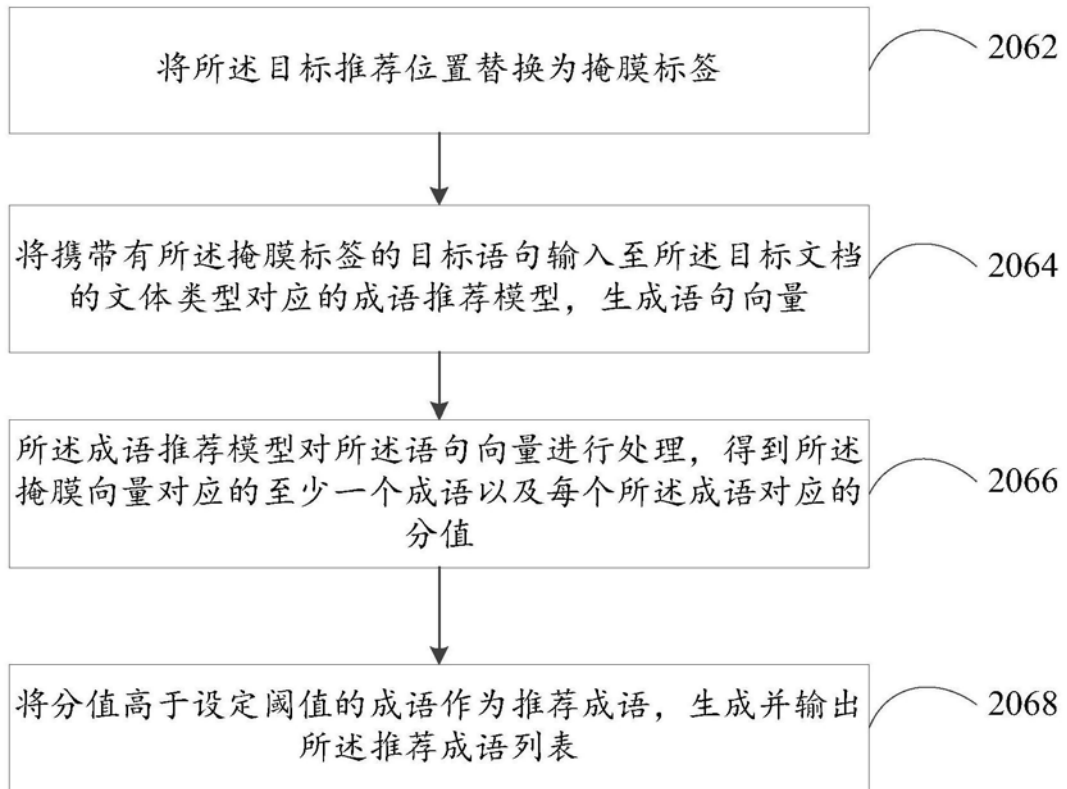


图3

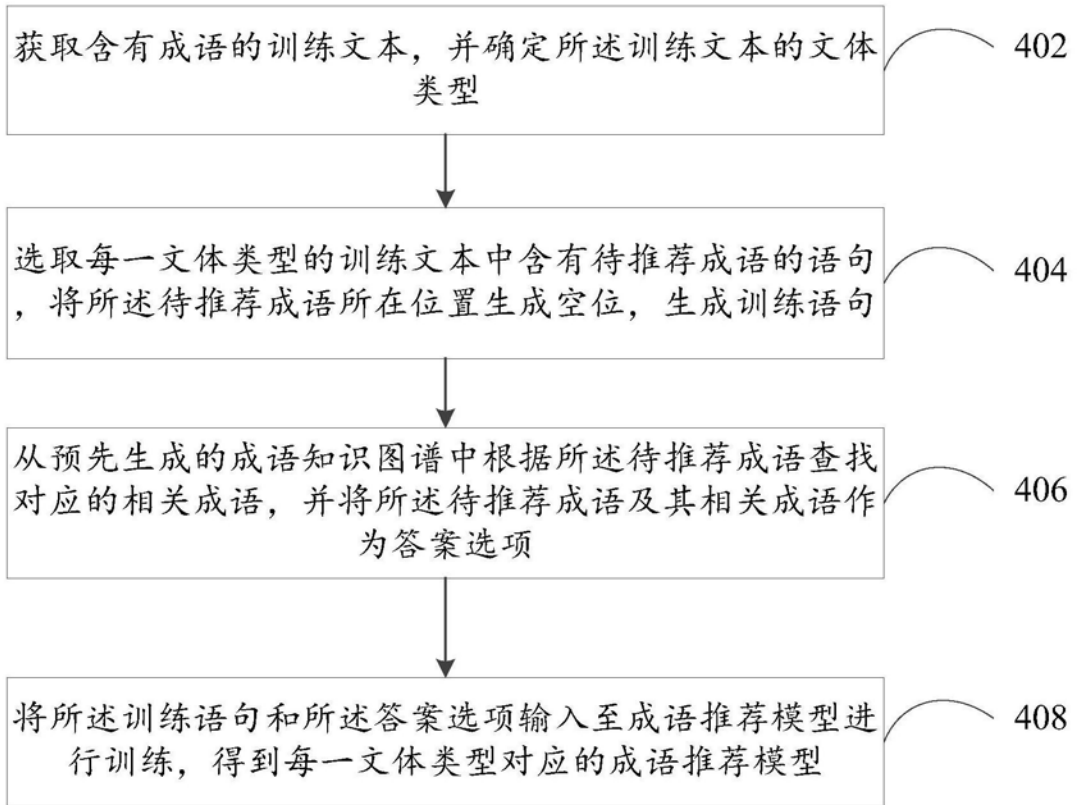


图4

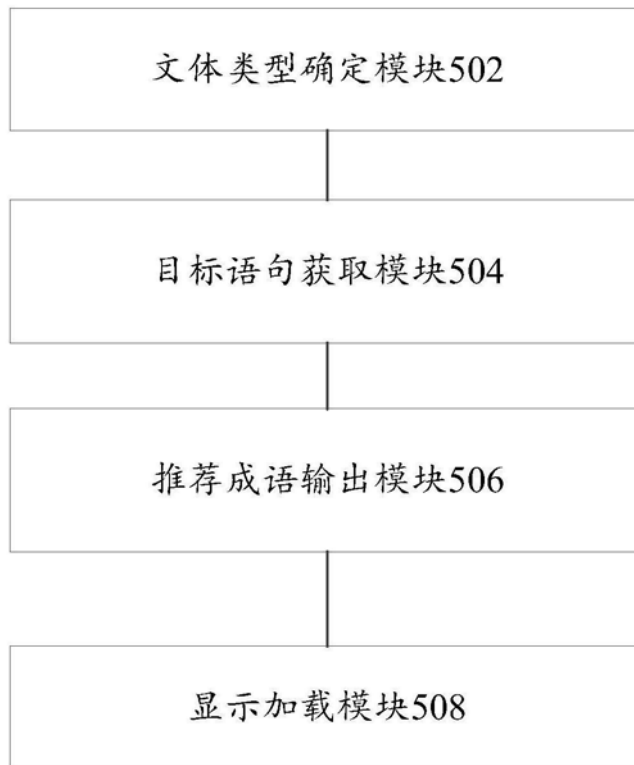


图5

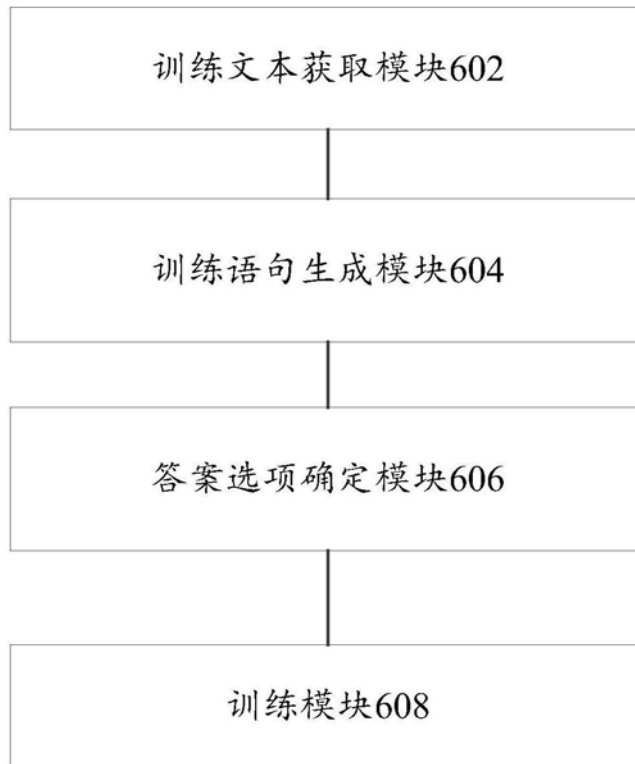


图6

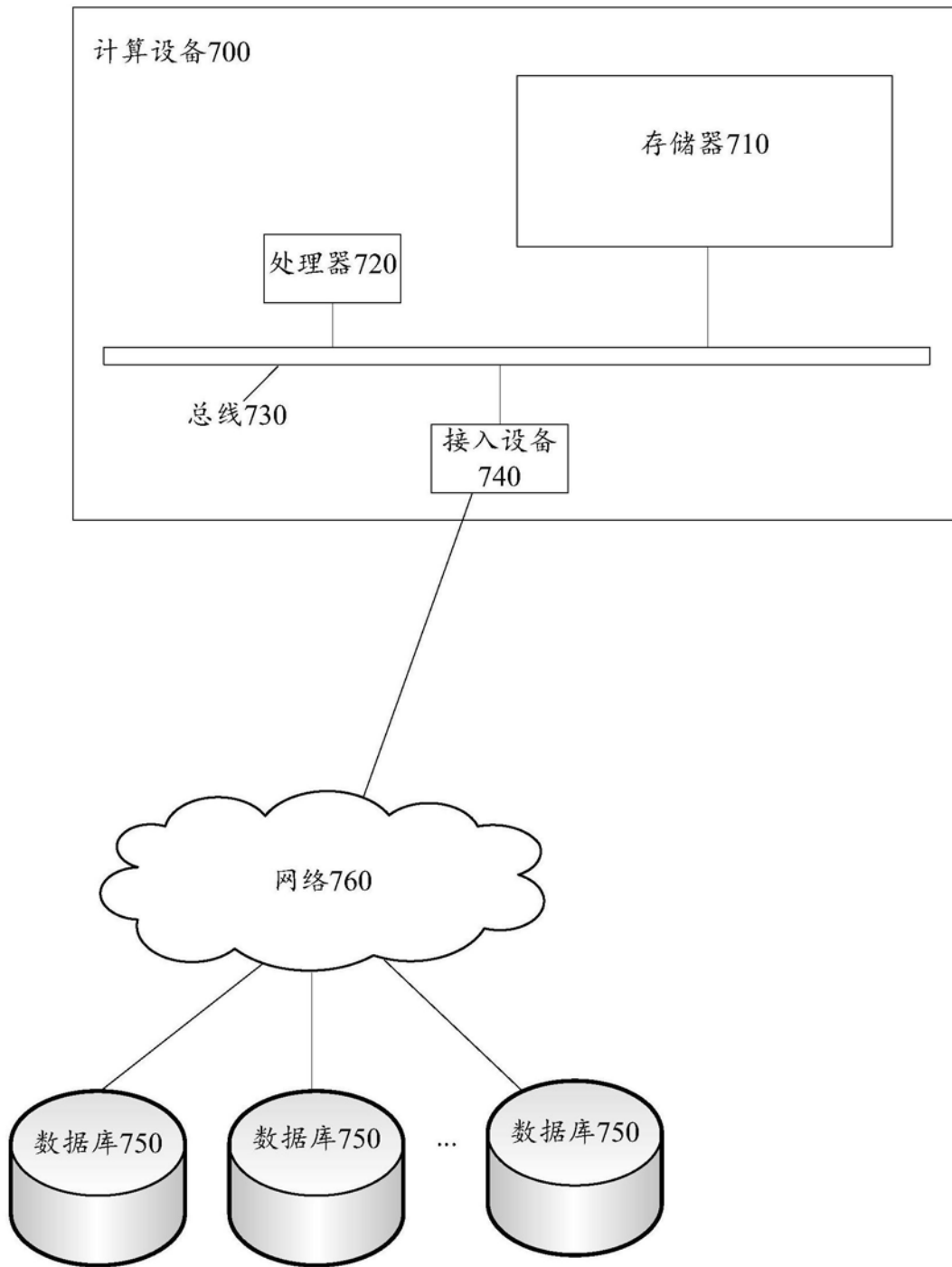


图7