(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 109344240 B
(45)授权公告日 2022.11.22

(21)申请号 201811108592.3

(22)申请日 2018.09.21

(65)同一申请的已公布的文献号
申请公布号 CN 109344240 A

(43)申请公布日 2019.02.15

(73)专利权人 联想(北京)有限公司
地址 100085 北京市海淀区上地信息产业基地创业路6号

(72)发明人 杨帆 金继民 金宝宝 张成松

(74)专利代理机构 北京集佳知识产权代理有限公司 11227
专利代理师 王宝筠

(51)Int.Cl.
G06F 16/332(2019.01)

(56)对比文件
CN 108028043 A,2018.05.11
CN 105930452 A,2016.09.07
CN 106934012 A,2017.07.07
CN 107895037 A,2018.04.10

审查员 胡武扬

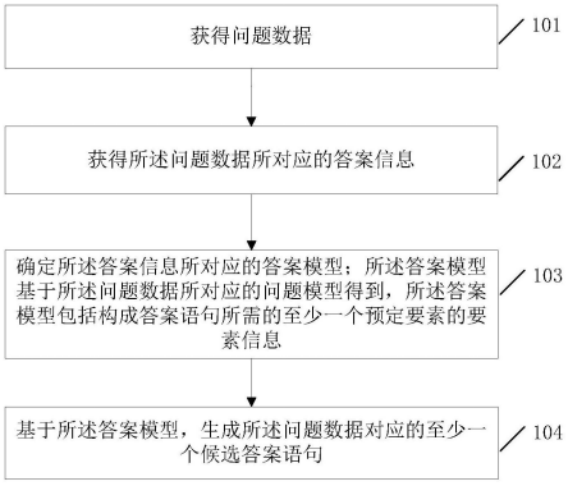
权利要求书3页 说明书23页 附图8页

(54)发明名称

一种数据处理方法、服务器及电子设备

(57)摘要

本申请提供的一种数据处理方法、服务器及电子设备，在获得问题数据以及问题数据所对应的答案信息基础上，基于问题数据所对应的问题模型确定所述答案信息所对应的答案模型，并进而基于所述答案模型生成问题数据对应的至少一个候选答案语句，由于针对问题数据，生成了符合其答案模型的至少一个候选答案语句，从而为答案语句的确定提供了语言表述形式(语句形式)方面的选择空间，能够支持智能问答系统从中选择出更加贴近人类自然语言的答案语句，进而可有效提升智能问答系统在语言表述形式方面的答案质量。



CN 109344240 B

1.一种数据处理方法,其特征在于,包括:

获得问题数据;

获得所述问题数据所对应的答案信息;

确定所述答案信息所对应的答案模型;所述答案模型基于所述问题数据所对应的问题模型得到,所述答案模型包括构成答案语句所需的至少一个预定要素的要素信息;

搜索预定的概率图谱,确定第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置;所述第二SPO模型为所述答案信息所对应的答案模型,所述第二SPO模型中的主体、谓词、宾语为所述答案模型包括的所述要素信息;

其中,所述概率图谱为一预先基于所需业务领域的多个问答语句样本所构建的包括多个节点的有向图,图中的每个节点对应一个词语,任意两个节点之间的边为有向边,有向边所指向的节点的词语依存于有向边所背离的节点的词语,任意两个节点之间的边对应一概率数值,边所对应的概率数值表示边的两个节点的两个词语同时出现的频率与边所背离的节点的词语出现的频率的比值;

基于所述第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置,确定所述第二SPO模型中主体、谓词及宾语的至少一个预定组合顺序在所述概率图谱中所对应的至少一个节点路径;

获得每个节点路径所对应的词序列,每个词序列构成的语句作为所述问题数据对应的一个候选答案语句,得到所述问题数据对应的至少一个候选答案语句。

2.根据权利要求1所述的方法,其特征在于,还包括:

基于预定评价方式,对所述至少一个候选答案语句中的每个候选答案语句进行评价,得到每个候选答案语句的评价结果;

选取评价结果最优的候选答案语句作为所述问题数据的答案语句。

3.根据权利要求1所述的方法,其特征在于,所述获得问题数据所对应的答案信息,确定所述答案信息所对应的答案模型,包括:

提取所述问题数据所对应的第一主体-谓词-宾语SPO模型;第一SPO 模型中的主体或谓词或宾语包括所述问题数据中所携带的疑问点信息;

查询预定的知识库,获得所述知识库中与所述疑问点信息相对应且与所述第一SPO模型中的未包括疑问点信息的部分相匹配的答案点信息;

将所述答案点信息与所述第一SPO模型中的未包括疑问点信息的部分整合为第二SPO模型。

4.根据权利要求3所述的方法,其特征在于,所述获得问题数据对应的答案信息,确定所述答案信息所对应的答案模型,包括:

提取所述问题数据所对应的第一SPO模型;所述第一SPO模型中的主体包括所述问题数据中所携带的疑问点信息;

查询预定的知识库,获得所述知识库中与所述疑问点信息相对应且与所述第一SPO模型中的谓词及宾语相匹配的答案点信息;

将所述答案点信息作为主体,与所述第一SPO模型中的谓词及宾语整合为所述第二SPO模型。

5.根据权利要求1所述的方法,其特征在于,所述至少一个预定组合顺序包括所述第二

SPO模型中主体、谓词及宾语的所有组合顺序。

6.根据权利要求1所述的方法，其特征在于，所述基于所述答案模型，生成所述问题数据对应的至少一个候选答案语句，还包括：

滤除所述至少一个预定组合顺序在所述概率图谱中所对应的节点路径中不符合预置条件的节点路径；

所述预置条件包括：节点路径长度不超出预定长度阈值，和/或，节点路径所对应的词序列在所述概率图谱中的出现概率不低于预定概率阈值。

7.根据权利要求1所述的方法，其特征在于，所述基于预定评价方式，对所述至少一个候选答案语句中的每个候选答案语句进行评价，包括：

提取每个候选答案语句的预定特征的特征信息；所述预定特征包括如下特征中的至少一个：候选答案语句的词序列在所述概率图谱中的出现概率，候选答案语句与问题数据的相似度，候选答案语句中主体、谓词及宾语的组合顺序在所述概率图谱中对应的节点路径长度的概率分布、候选答案语句的词序列中所包括的词语个数；

基于每个候选答案语句的特征信息对每个候选答案语句进行评分，得到每个候选答案语句的评分分值。

8.一种服务器，其特征在于，包括：

存储器，用于至少存储一组指令集；

处理器，用于调用并执行所述存储器中的所述指令集，通过执行所述指令集进行以下操作：

获得问题数据；

获得所述问题数据所对应的答案信息；

确定所述答案信息所对应的答案模型；所述答案模型基于所述问题数据所对应的问题模型得到，所述答案模型包括构成答案语句所需的至少一个预定要素的要素信息；

搜索预定的概率图谱，确定第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置；所述第二SPO模型为所述答案信息所对应的答案模型，所述第二SPO模型中的主体、谓词、宾语为所述答案模型包括的所述要素信息；

其中，所述概率图谱为一预先基于所需业务领域的多个问答语句样本所构建的包括多个节点的有向图，图中的每个节点对应一个词语，任意两个节点之间的边为有向边，有向边所指向的节点的词语依存于有向边所背离的节点的词语，任意两个节点之间的边对应一概率数值，边所对应的概率数值表示边的两个节点的两个词语同时出现的频率与边所背离的节点的词语出现的频率的比值；

基于所述第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置，确定所述第二SPO模型中主体、谓词及宾语的至少一个预定组合顺序在所述概率图谱中所对应的至少一个节点路径；

获得每个节点路径所对应的词序列，每个词序列构成的语句作为所述问题数据对应的一个候选答案语句，得到所述问题数据对应的至少一个候选答案语句。

9.一种电子设备，其特征在于，包括：

存储器，用于至少存储一组指令集；

处理器，用于调用并执行所述存储器中的所述指令集，通过执行所述指令集进行以下

操作：

获得问题数据；

获得所述问题数据所对应的答案信息；

确定所述答案信息所对应的答案模型；所述答案模型基于所述问题数据所对应的问题模型得到，所述答案模型包括构成答案语句所需的至少一个预定要素的要素信息；

搜索预定的概率图谱，确定第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置；所述第二SPO模型为所述答案信息所对应的答案模型，所述第二SPO模型中的主体、谓词、宾语为所述答案模型包括的所述要素信息；

其中，所述概率图谱为一预先基于所需业务领域的多个问答语句样本所构建的包括多个节点的有向图，图中的每个节点对应一个词语，任意两个节点之间的边为有向边，有向边所指向的节点的词语依存于有向边所背离的节点的词语，任意两个节点之间的边对应一概率数值，边所对应的概率数值表示边的两个节点的两个词语同时出现的频率与边所背离的节点的词语出现的频率的比值；

基于所述第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置，确定所述第二SPO模型中主体、谓词及宾语的至少一个预定组合顺序在所述概率图谱中所对应的至少一个节点路径；

获得每个节点路径所对应的词序列，每个词序列构成的语句作为所述问题数据对应的一个候选答案语句，得到所述问题数据对应的至少一个候选答案语句。

一种数据处理方法、服务器及电子设备

技术领域

[0001] 本发明属于基于大数据的数据处理技术领域，尤其涉及一种数据处理方法、服务器及电子设备。

背景技术

[0002] 智能问答系统是在大规模知识处理基础上发展起来的一种面向行业应用的自动服务系统，其为企业与海量用户之间的沟通建立了一种基于自然语言处理的快捷有效的沟通途径。

[0003] 针对用户问题自动给予相匹配的解答答案，是智能问答系统的主要应用形式，在智能问答系统反馈答案时，如果能够将答案内容转换成更加贴近人类自然语言的语句形式，可以显著改善系统的用户体验。然而，目前的智能问答系统更多关注的是回答结果的准确性，较少考虑系统所提供答案的人性化程度，这样势必会影响智能问答系统在语言表述形式方面的答案质量，相对应地影响用户对智能问答系统的使用体验。

发明内容

[0004] 有鉴于此，本发明的目的在于提供一种数据处理方法、服务器及电子设备，用于克服现有智能问答系统存在的上述问题，提升智能问答系统在语言表述形式方面的答案质量。

[0005] 为此，本发明公开如下技术方案：

[0006] 一种数据处理方法，包括：

[0007] 获得问题数据；

[0008] 获得所述问题数据所对应的答案信息；

[0009] 确定所述答案信息所对应的答案模型；所述答案模型基于所述问题数据所对应的问题模型得到，所述答案模型包括构成答案语句所需的至少一个预定要素的要素信息；

[0010] 基于所述答案模型，生成所述问题数据对应的至少一个候选答案语句。

[0011] 上述方法，优选地，还包括：

[0012] 基于预定评价方式，对所述至少一个候选答案语句中的每个候选答案语句进行评价，得到每个候选答案语句的评价结果；

[0013] 选取评价结果最优的候选答案语句作为所述问题数据的答案语句。

[0014] 上述方法，优选地，所述获得问题数据所对应的答案信息，确定所述答案信息所对应的答案模型，包括：

[0015] 提取所述问题数据所对应的第一主体-谓词-宾语SPO模型；所述第一SPO模型中的主体或谓词或宾语包括所述问题数据中所携带的疑问点信息；

[0016] 查询预定的知识库，获得所述知识库中与所述疑问点信息相对应且与所述第一SPO模型中的未包括疑问点信息的部分相匹配的答案点信息；

[0017] 将所述答案点信息与所述第一SPO模型中的未包括疑问点信息的部分整合为第二

SPO模型，所述第二SPO模型为所述答案信息所对应的答案模型，所述第二SPO模型中的主体、谓词、宾语为所述答案模型包括的所述要素信息。

[0018] 上述方法，优选地，所述获得问题数据对应的答案信息，确定所述答案信息所对应的答案模型，包括：

[0019] 提取所述问题数据所对应的第一SPO模型；所述第一SPO模型中的主体包括所述问题数据中所携带的疑问点信息；

[0020] 查询预定的知识库，获得所述知识库中与所述疑问点信息相对应且与所述第一SPO模型中的谓词及宾语相匹配的答案点信息；

[0021] 将所述答案点信息作为主体，与所述第一SPO模型中的谓词及宾语整合为所述第二SPO模型。

[0022] 上述方法，优选地，所述基于所述答案模型，生成所述问题数据对应的至少一个候选答案语句，包括：

[0023] 搜索预定的概率图谱，确定所述第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置；其中，所述概率图谱为一预先基于所需业务领域的多个问答语句样本所构建的包括多个节点的有向图，图中的每个节点对应一个词语，任意两个节点之间的边为有向边，有向边所指向的节点的词语依存于有向边所背离的节点的词语，任意两个节点之间的边对应一概率数值，边所对应的概率数值表示边的两个节点的两个词语同时出现的频率与边所背离的节点的词语出现的频率的比值；

[0024] 基于所述第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置，确定所述第二SPO模型中主体、谓词及宾语的至少一个预定组合顺序在所述概率图谱中所对应的至少一个节点路径；

[0025] 获得每个节点路径所对应的词序列，每个词序列构成的语句作为所述问题数据对应的一个候选答案语句，得到所述问题数据对应的至少一个候选答案语句。

[0026] 上述方法，优选地，所述至少一个预定组合顺序包括所述第二SPO模型中主体、谓词及宾语的所有组合顺序。

[0027] 上述方法，优选地，所述基于所述答案模型，生成所述问题数据对应的至少一个候选答案语句，还包括：

[0028] 滤除所述至少一个预定组合顺序在所述概率图谱中所对应的节点路径中不符合预置条件的节点路径；

[0029] 所述预置条件包括：节点路径长度不超出预定长度阈值，和/或，节点路径所对应的词序列在所述概率图谱中的出现概率不低于预定概率阈值。

[0030] 上述方法，优选地，所述基于预定评价方式，对所述至少一个候选答案语句中的每个候选答案语句进行评价，包括：

[0031] 提取每个候选答案语句的预定特征的特征信息；所述预定特征包括如下特征中的至少一个：候选答案语句的词序列在所述概率图谱中的出现概率，候选答案语句与问题数据的相似度，候选答案语句中主体、谓词及宾语的组合顺序在所述概率图谱中对应的节点路径长度的概率分布、候选答案语句的词序列中所包括的词语个数；

[0032] 基于每个候选答案语句的特征信息对每个候选答案语句进行评分，得到每个候选答案语句的评分分值。

[0033] 一种服务器，包括：

[0034] 存储器，用于至少存储一组指令集；

[0035] 处理器，用于调用并执行所述存储器中的所述指令集，通过执行所述指令集进行以下操作：

[0036] 获得问题数据；

[0037] 获得所述问题数据所对应的答案信息；

[0038] 确定所述答案信息所对应的答案模型；所述答案模型基于所述问题数据所对应的问题模型得到，所述答案模型包括构成答案语句所需的至少一个预定要素的要素信息；

[0039] 基于所述答案模型，生成所述问题数据对应的至少一个候选答案语句。

[0040] 一种电子设备，包括：

[0041] 存储器，用于至少存储一组指令集；

[0042] 处理器，用于调用并执行所述存储器中的所述指令集，通过执行所述指令集进行以下操作：

[0043] 获得问题数据；

[0044] 获得所述问题数据所对应的答案信息；

[0045] 确定所述答案信息所对应的答案模型；所述答案模型基于所述问题数据所对应的问题模型得到，所述答案模型包括构成答案语句所需的至少一个预定要素的要素信息；

[0046] 基于所述答案模型，生成所述问题数据对应的至少一个候选答案语句。

[0047] 根据以上方案可知，本申请提供的数据处理方法、服务器及电子设备，在获得问题数据以及问题数据所对应的答案信息基础上，基于问题数据所对应的问题模型确定所述答案信息所对应的答案模型，并进而基于所述答案模型生成问题数据对应的至少一个候选答案语句，由于针对问题数据，生成了符合其答案模型的至少一个候选答案语句，从而为答案语句的确定提供了语言表述形式(语句形式)方面的选择空间，能够支持智能问答系统从中选择出更加贴近人类自然语言的答案语句，进而可有效提升智能问答系统在语言表述形式方面的答案质量。

附图说明

[0048] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述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发明的实施例，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，还可以根据提供的附图获得其他的附图。

[0049] 图1是本申请提供的一种数据处理方法实施例一的流程图；

[0050] 图2是本申请提供的一种数据处理方法实施例二的流程图；

[0051] 图3是本申请实施例二提供的概率图谱的一示例图；

[0052] 图4(a)是本申请实施例二提供的节点间边的一种示例图；

[0053] 图4(b)是本申请实施例二提供的节点间边的另一种示例图；

[0054] 图5是本申请实施例二提供的在概率图谱中定位的第二SPO模型中主体、谓词及宾语的节点位置的示例图；

[0055] 图6是本申请实施例二提供的数据处理方法的处理逻辑框架示意图；

[0056] 图7是本申请提供的一种数据处理方法实施例三的流程图；

[0057] 图8是本申请提供的一种数据处理方法实施例四的流程图；

[0058] 图9是本申请提供的一种服务器实施例五的结构示意图；

[0059] 图10是本申请提供的一种电子设备实施例九的结构示意图。

具体实施方式

[0060] 下面将结合本发明实施例中的附图，对本发明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发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0061] 为了提升智能问答系统在语言表述形式方面的答案质量，使得所确定出的答案语句更加贴近人类自然语言，本申请提供了一种数据处理方法、服务器及电子设备，以下将通过多个实施例对本申请的数据处理方法、服务器及电子设备进行说明。

[0062] 参考图1，是本申请提供的一种数据处理方法实施例一的流程图，该数据处理方法可应用于能够提供数据处理服务的本地/网络端服务器或服务器集群中，或者还可以应用于智能手机、平板电脑、台式机、笔记本、一体机等各类型终端设备中。如图1所示，本实施例中，所述数据处理方法包括如下步骤：

[0063] 步骤101、获得问题数据。

[0064] 所述问题数据可以是但不限于用户在智能问答场景中基于其实际需求提交至智能问答系统的问题语句，如用户在其智能手机、笔记本等终端设备中通过智能问答APP(Application，应用程序)的应用界面或网页提交至智能问答系统的问题语句等。

[0065] 相对应地，本步骤中获得问题数据，则可以是本申请方法的执行主体(如用户的终端设备或提供数据处理服务的服务器/服务器集群等)基于智能问答系统所获得的用户提交的问题语句，如用户提交的“什么药可以治疗感冒?”等。其中，若所述执行主体为用户的终端设备本身，则终端设备可直接基于其相应问题输入界面所提供的问题输入/提交功能获得用户所输入或提交的问题语句，若所述执行主体为后台服务器/服务器集群，如本地/网络端的服务器/服务器集群等，则后台服务器/服务器集群可基于局域网、互联网或移动网络等网络的通信功能获得用户在其终端设备所提交的问题语句。

[0066] 需要说明的是，实际应用中，所述用户可以是自然人用户或者基于人工智能的非自然人用户，这里，并不对此进行限定。

[0067] 步骤102、获得所述问题数据所对应的答案信息。

[0068] 通常来说，问题数据中会携带疑问点信息，以反映用户的提问需求，如上述的问题语句“什么药可以治疗感冒?”即携带了“什么(药)”这一疑问点信息。

[0069] 所述答案信息可以是与所述问题数据中携带的疑问点信息相对应且与疑问点信息之外的其他部分相匹配的、能够解决问题的答案点信息，如针对上述问题语句“什么药可以治疗感冒?”中的疑问点信息“什么(药)”，则所述答案信息相对应地可以为能够治疗感冒的答案点信息“阿莫西林”。

[0070] 本步骤具体可通过对问题数据进行问句解析，确定出其中的疑问点信息，进而基于相应业务知识(如预先制定的涵盖了本领域或多个领域业务知识的知识库等)确定出与

所述疑问点信息相对应且与疑问点信息之外的其他部分相匹配的、能够解决问题的答案点信息。

[0071] 其中，所述疑问点信息一般来说是一些能够反映用户提问需求的疑问词，如“什么”、“哪些”、“哪个”、“哪儿”、“几个”、“如何”等等。

[0072] 步骤103、确定所述答案信息所对应的答案模型；所述答案模型基于所述问题数据所对应的问题模型得到，所述答案模型包括构成答案语句所需的至少一个预定要素的要素信息。

[0073] 对于待解答的问题数据，本申请在获得其答案信息后(如上述的答案点信息)，并不直接为其生成相对应的答案语句，而是确定答案信息所对应的答案模型，该答案模型包括构成答案语句所需的至少一个预定要素的要素信息，可将其形象地理解为一包括了答案语句的各要素信息的框架模型。

[0074] 容易理解的是，所述答案模型中应至少包括所述对应于疑问点信息且与问题语句中疑问点信息之外的其他部分相匹配的、能够解决问题的答案点信息。

[0075] 其中，所述答案模型具体可基于所述答案信息中的答案点信息以及问题数据所对应的问题模型确定出，所述问题模型相对应地包括构成问题语句所需的至少一个预定要素的要素信息，该问题模型同样可相应形象地理解为一包括了问题语句的各要素信息的框架模型，且所述问题模型应至少包括疑问点信息。

[0076] 具体地，发明人经研究发现，实际应用中，对于待解答的问题数据，一般来说，可通过将其问题语句中所包括的疑问点信息(如疑问词“什么(药)”)替换为所确定出的答案点信息来得到所对应的答案语句，也即，将所确定出的答案点信息与问题语句中非答案点信息的部分进行拼接来得出问题语句所对应的答案语句，如针对上述的问题语句“什么药可以治疗感冒?”，可通过将疑问点信息“什么(药)”替换为答案点信息“阿莫西林”，来得到其对应的答案语句“阿莫西林可以治疗感冒”。基于此，从模型角度来说，问题语句所对应的答案语句的答案模型可以与问题语句的问题模型具备相类似的模型结构，不同之处仅在于答案模型相比于问题模型来说，将问题模型中的疑问点信息替换为了答案点信息。

[0077] 鉴于此，本步骤中，具体可对问题语句进行问句解析，提取其问题模型，进而通过将问题模型中的疑问点信息替换为答案点信息，而模型的其他部分维持不变来得到问题语句的答案语句所对应的答案模型。

[0078] 步骤104、基于所述答案模型，生成所述问题数据对应的至少一个候选答案语句。

[0079] 如前文所述，所述答案模型为一包括了答案语句的各要素信息的框架模型，但其并不足以构成一完整的问题语句，为了支持智能问答系统能够确定出更加贴近人类自然语言的答案语句，本步骤基于所述答案模型，生成所述问题数据对应的至少一个候选答案语句。

[0080] 容易理解的是，每个候选答案语句都是符合所述答案模型的模型要求的，也即，每个候选答案语句都包括了构成答案语句所需的各个要素信息，从而，从准确度方面来说，各个候选答案语句无明显差别，都涵盖了解决问题所需的答案点信息能够解决问题，区别主要在于语言表述形式(语句形式)方面的不同。

[0081] 实际应用中，可根据业务需求，通过线上(即从网络上收集)和/或线下方式收集本领域或多个领域的高质量问答语句，来预先构建语料库(知识库)，所构建的预料库(知识

库)可尽可能基于特定领域或多个领域的业务特点,体现特定领域或多个领域的自然语言语句描述方式、风格,以使得基于该预料库(知识库)所确定出的答案语句更加贴近人类自然语言。

[0082] 在构建语料库(知识库)的基础上,可基于预料库(知识库)生成符合所述答案模型的模型要求的至少一个候选答案语句,以使得为问题数据所对应的答案语句的确定提供语言表述形式方面的选择空间,后续,可基于相应的选择策略/选择机制从所述至少一个候选答案语句中选择其中之一作为所述问题数据最终的答案语句。

[0083] 根据以上方案可知,本实施例提供的数据处理方法,在获得问题数据以及问题数据所对应的答案信息基础上,基于问题数据所对应的问题模型确定所述答案信息所对应的答案模型,并进而基于所述答案模型生成问题数据对应的至少一个候选答案语句,由于针对问题数据,生成了符合其答案模型的至少一个候选答案语句,从而为答案语句的确定提供了语言表述形式方面的选择空间,能够支持智能问答系统从中选择出更加贴近人类自然语言的答案语句,进而可有效提升智能问答系统在语言表述形式方面的答案质量。

[0084] 以下通过另一实施例继续对上述实施例中的数据处理方法进行进一步详述,参考图2,是本申请提供的一种数据处理方法实施例二的流程图,本实施例二中,所述数据处理方法可以通过如下的处理过程实现:

[0085] 步骤201、获得问题数据。

[0086] 所述问题数据可以是但不限于用户在智能问答场景中基于其实际需求提交至智能问答系统的问题语句,如用户在其智能手机、笔记本等终端设备中通过智能问答APP的应用界面或网页提交至智能问答系统的问题语句等。

[0087] 相对应地,本步骤中获得问题数据,则可以是本申请方法的执行主体(如用户的终端设备或提供数据处理服务的服务器/服务器集群等)基于智能问答系统所获得的用户提交的问题语句,如用户提交的“什么药可以治疗感冒?”等。其中,若所述执行主体为用户的终端设备本身,则终端设备可直接基于其相应问题输入界面所提供的问题输入/提交功能获得用户所输入或提交的问题语句,若所述执行主体为后台服务器/服务器集群,如本地/网络端的服务器或服务器集群等,则后台服务器/服务器集群可基于局域网、互联网或移动网络等网络的通信功能获得用户在其终端设备所提交的问题语句。

[0088] 需要说明的是,实际应用中,所述用户可以是自然人用户或者基于人工智能的非自然人用户,这里,并不对此进行限定。

[0089] 步骤202、提取所述问题数据所对应的第一SPO(Subject-Predicate-Object,主体-谓词-宾语)模型;所述第一SPO模型中的主体或谓词或宾语包括所述问题数据中所携带的疑问点信息。

[0090] 其中,可首先分别基于相应的实体识别技术及关系提取技术,识别问题数据的问题语句中所包括的实体和关系。具体地,对于实体识别,目前已存在很多可用算法,例如CRF(conditional random field,条件随机场)、HMM(Hidden Markov Model,隐马尔可夫模型)等算法,从而,可选取这些算法中的其中一种对问题语句中的实体进行识别,或者亦可以通过模式匹配的方式进行实体识别,这里不局限于一种技术。所述关系是指句子中实体与实体间的谓词关系,从而问题语句中关系的识别可通过对问题语句进行谓词识别来实现,而对于谓词,则具体可通过对问题语句进行词法分析及句法分析,并在词法分析及句法分析

的基础上，通过融合词法和句法特征来实现谓词识别。

[0091] 例如，对于问题语句“什么药可以治疗感冒?”通过对其进行实体及关系识别，可以获知其包括“什么药”(本质上是以疑问词形式指代的实体)及“感冒”两个实体，包括“治疗”这一谓词关系。

[0092] 需要说明的是，通常来说，问题数据的问题语句中会携带疑问点信息，以反映用户的提问需求，如上述的问题语句“什么药可以治疗感冒?”即携带了“什么(药)”这一疑问点信息，问题语句“阿莫西林如何治疗感冒”则携带了“如何(治疗)”这一疑问点信息。

[0093] 问题语句中的疑问点信息一般来说是上文所述的“什么”、“哪些”、“哪个”、“哪儿”、“几个”或“如何”等疑问词，其通常包括在问题语句的实体或者关系中，从而，可通过对识别出的实体或关系进一步进行疑问词识别，来获知问题语句中的疑问点信息。仍以上述的问题语句“什么药可以治疗感冒?”为例，在识别出其两个实体“什么药”、“感冒”以及一个关系“治疗”的基础上，可进一步通过对实体及关系进行疑问词识别来获知实体“什么药”中包括疑问词“什么(药)”，当然，还可以直接通过对问题语句进行疑问词识别来获知其疑问点信息(而并非在实体、关系识别的基础上进行疑问词识别)，本实施例并不对此进行限定。

[0094] 其中，由于疑问词数量有限，具体实施中可以通过模式匹配的方式对问题语句中的疑问词进行识别。

[0095] 在实体及关系识别的基础上，可通过继续对问题语句进行问句句法解析，来获知问题语句中的实体及关系属于主体(Subject)、谓词(Predicate)、宾语(Object)中的哪一种，进而提取出问题语句所对应的第一SPO模型，其中，所述第一SPO模型中的主体或谓词或宾语包括所述问题数据中所携带的疑问点信息。

[0096] 如对于问题语句“什么药可以治疗感冒?”，通过上述处理过程，可以提取出其第一SPO模型：“(什么)药(S)-治疗(P)-感冒(O)”，在该示例中，主体中包括疑问词“什么(药)”；对于问题语句“阿司匹林如何治疗感冒”，通过上述处理过程，可以提取出其第一SPO模型：“阿司匹林-(如何)治疗-感冒”，在该示例中，谓词中包括疑问词“如何(治疗)”。

[0097] 步骤203、查询预定的知识库，获得所述知识库中与所述疑问点信息相对应且与所述第一SPO模型中的未包括疑问点信息的部分相匹配的答案点信息。

[0098] 所述预定的知识库，可以是但不限于基于所收集的某一特定业务领域或多个业务领域的业务知识所构建的知识图谱，具体地，所述知识图谱可以为一预先基于所述某一特定业务领域或多个业务领域的能够反映业务知识的一系列问答语句样本所构建的有向图，该有向图包括多个节点，图中的每个节点对应一个词语(实体词语)，任意两个节点之间的边为有向边，有向边所指向的节点的词语依存于有向边所背离的节点的词语，任意两个节点之间的边对应一谓词关系，用于表示边的两个节点所对应的两个词语间的关系。

[0099] 本步骤具体可将所述知识图谱所提供的有向图作为搜索空间，通过查询所述知识图谱，来获得所述知识图谱中与所述疑问点信息相对应且与所述第一SPO模型中的未包括疑问点信息的部分相匹配的答案点信息。

[0100] 如对于问题语句“什么药可以治疗感冒?”的第一SPO模型“(什么)药-治疗-感冒”，通过查询所述知识图谱，可获得与疑问点信息“什么(药)”相对应且与“治疗-感冒”相匹配的答案点信息“阿司匹林”。

[0101] 步骤204、将所述答案点信息与所述第一SPO模型中的未包括疑问点信息的部分整

合为第二SPO模型，所述第二SPO模型为所述答案信息所对应的答案模型，所述第二SPO模型中的主体、谓词、宾语为所述答案模型包括的所述要素信息。

[0102] 具体地，可通过将所获得的答案点信息替代问题语句的第一SPO模型中的疑问点信息，实现答案点信息与所述第一SPO模型中的未包括疑问点信息的部分的整合，以此得到问题语句所对应的答案语句的第二SPO模型。

[0103] 仍以上述的问题语句“什么药可以治疗感冒?”为例，可将答案点信息“阿司匹林”替代其第一SPO模型“(什么)药-治疗-感冒”中的“什么(药)”，来得到其问题语句的第二SPO模型“阿司匹林-治疗-感冒”。

[0104] 其中，该第二SPO模型中包括构成问题语句所需的各个要素(主体、谓词、宾语)的要素信息，但其并不足以构成一完整的、符合人类自然语言的问题语句。

[0105] 步骤205、搜索预定的概率图谱，确定所述第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置。

[0106] 其中，所述概率图谱为一预先基于某一特定业务领域或多个业务领域的一系列问答语句样本所构建的包括多个节点的有向图，图中的每个节点对应一个词语(可以是实体词语、谓词词语、形容词词语、复合型词语等各类型词语)，任意两个节点之间的边为有向边，有向边所指向的节点的词语依存于有向边所背离的节点的词语，任意两个节点之间的边对应一概率数值，边所对应的概率数值表示边的两个节点的两个词语同时出现的频率与边所背离的节点的词语出现的频率的比值。参考图3，为本实施例提供的概率图谱的一示意图。

[0107] 具体地，可通过以下处理过程来构建上述的概率图谱：

[0108] 对收集的一系列问答语句样本进行分词，分词之后句子以词序列的形式存在，句子所表达的语义蕴含在词语之间的依存关系之中，对于每一个句子，依据其词序列的先后关系，通过边(即有向图的边)将词序列进行串联得到该句子的语义路径，并将各个句子的语义路径进行合并关联，得到整个语料的语义图，其中语义图中的每个节点为一个词语，而词语与词语在句子中存在的先后关系则体现为语义图中边的指向关系，也即，边的方向体现了边的两个节点所对应的词语在句子中的先后顺序，该先后顺序也体现了该两个词语在句子中的依存关系，其中顺序在前的词语在有向边的发出端，顺序在后的词语在有向边的箭头端，且有向边所指向的节点的词语依存于有向边所背离的节点的词语，在构建出语义图的基础上，进一步对语义图中各节点的词语进行词频统计及概率计算来为每个有向边赋予相对应的概率值(边箭头端词语和发出端词语共现的频数与边发出端词语频数的比值)，从而得到所述概率图谱。

[0109] 需要说明的是，在概率图谱中边是有方向的，不具有对称性，因此对于两个邻近的节点，节点之间可能存在两条边，每个节点可能存在多个入边和出边，例如，如图4(a)及图4(b)所示，在图4(a)中，节点A到节点B存在一条有向边，则节点A与节点B之间存在依存关系，且节点B依存于节点A，在图4(b)中，节点A与节点B之间存在两条有向边，则节点A与节点B之间存在依存关系，对应于从节点A指向节点B的有向边，节点B依存于节点A，而对应于从节点B指向节点A的有向边，则节点A依存于节点B。

[0110] 对于上述步骤中所获得第二SPO模型，可通过搜索该概率图谱，来确定所述第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置，参考图5所示的在

概率图谱中基于搜索来定位第二SPO模型中主体、谓词及宾语的节点位置的示例图，在该示例中，通过对概率图谱进行搜索，确定出第二SPO模型中的主体、谓词及宾语分别位于概率图谱有向图中的word8、word9及word10这些节点位置处。

[0111] 步骤206、基于所述第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置，确定所述第二SPO模型中主体、谓词及宾语的至少一个预定组合顺序在所述概率图谱中所对应的至少一个节点路径。

[0112] 所述至少一个预定组合顺序可以包括所述第二SPO模型中主体、谓词及宾语的所有组合顺序，如SPO，SOP、PSO、POS、OSP和OPS等。或者还可以仅包括所述第二SPO模型中主体、谓词及宾语的某一特定顺序，如SPO等，本实施例对此不作限定。

[0113] 以SPO这一组合顺序为例，附图5中，所述第二SPO模型中主体、谓词及宾语的SPO组合顺序在所述概率图谱中所对应的节点路径包括：

[0114] word8→word5→word9→word7→word10；

[0115] word8→word3→word9→word7→word10；

[0116] word8→word1→word3→word9→word7→word10。

[0117] 步骤207、获得每个节点路径所对应的词序列，每个词序列构成的语句作为所述问题数据对应的一个候选答案语句，得到所述问题数据对应的至少一个候选答案语句。

[0118] 由于概率图谱中每个节点均对应一个词语，从而可基于概率图谱，获得每个上述节点路径所对应的词序列，其中，每个词序列构成的语句作为所述问题数据对应的一个候选答案语句，从而得到所述问题数据对应的至少一个候选答案语句。

[0119] 其中，每个候选答案语句都包括了构成答案语句所需的各要素的要素信息，相对应地都能够解决问题语句所体现的问题，区别仅在于或语言表述形式方面的不同，例如，对于问题语句“什么药可以治疗感冒?”，假设通过上述的节点搜索定位及节点路径上的词序列获取等处理，可以得到“阿司匹林-可以-治疗-感冒”、“阿司匹林-治疗-感冒”、“治疗-感冒-用-阿司匹林”“感冒-治疗-药-阿司匹林”等多个词序列，则相对应地可得到该问题语句的多个候选答案语句：“阿司匹林可以治疗感冒”、“阿司匹林治疗感冒”、“治疗感冒用阿司匹林”、“感冒治疗药阿司匹林”，每个候选答案语句都能够解决问题语句所体现的问题，区别仅在于语言表述形式方面的不同。

[0120] 通过生成问题数据的至少一个候选答案语句，可以为问题数据所对应的答案语句的确定提供语言表述形式方面的选择空间，后续，可基于相应的选择策略/选择机制(如基于相关概率计算的选择策略、基于节点路径长度控制的选择策略等)从所述至少一个候选答案语句中选择其中之一作为所述问题数据最终的答案语句。

[0121] 该实施例所对应的整个处理过程的处理逻辑框架具体可参考图6所示。

[0122] 本实施例通过搜索概率图谱，为答案模型生成了符合其模型要求的至少一个候选答案语句，从而为答案语句的确定提供了语言表述形式方面的选择空间，能够支持智能问答系统从中选择出更加贴近人类自然语言的答案语句，进而，可有效提升智能问答系统在语言表述形式方面的答案质量。

[0123] 参考图7，是本申请提供的一种数据处理方法实施例三的流程图，本实施例中，如图7所示，所述数据处理方法在所述步骤206之后，还可以包括以下处理步骤：

[0124] 步骤206’：滤除所述至少一个预定组合顺序在所述概率图谱中所对应的节点路径

中不符合预置条件的节点路径；

[0125] 其中，所述预置条件可以包括：节点路径长度不超出预定长度阈值，和/或，节点路径所对应的词序列在所述概率图谱中的出现概率不低于预定概率阈值。

[0126] 节点路径长度是指节点路径上所包括的节点的数量。

[0127] 节点路径所对应的词序列在所述概率图谱中的出现概率，可通过以下的计算式计算得出：

[0128] $P(seq)=\rho(w_n|w_{n-1})\cdots\rho(w_i|w_{i-1})\cdots\rho(w_2|w_1)\rho(w_1)$；

[0129] 其中，seq表示节点路径所对应的词序列$w_1,w_2,\ldots,w_n$，w_i（$1\leq i\leq n$，i为自然数）表示seq的第i个词语，P(seq)表示seq在所述概率图谱中出现的概率，$\rho(w_1)$表示seq中首个词w_1在概率图谱中出现的概率（$\rho(w_1)=w_1$的词频/概率图谱中各个词的词频的累计值），$\rho(w_i|w_{i-1})$表示词语在w_{i-1}条件下w_i的概率，即构建的概率图谱中从节点w_{i-1}到节点w_i的边上的概率。

[0130] 一般来说，若节点路径长度过长（超出预定长度阈值），会使得所确定出的候选答案语句过于复杂，于用户而言则答案语句不够简洁；若节点路径所对应的词序列在所述概率图谱中的出现概率过低（低于预定概率阈值），则考虑该节点路径对应的语句在日常生活或所属业务领域内不常使用，相对应地会认为该语句于用户而言并不太贴近人类的自然语言。鉴于此，在基于节点路径对应的词序列确定问题语句的候选答案语句之前，可首先基于上述的预置条件，将路径长度超出预定长度阈值的节点路径和/或所对应的词序列在所述概率图谱中的出现概率不低于预定概率阈值的节点路径滤除掉，以使得所保留的各节点路径所对应的语句于用户而言不会过于复杂和/或过于偏门（不贴仅人类自然语言）。

[0131] 具体实施中，所述阈值条件不必局限于本实施例所提供的上述内容，可以由技术人员或者用户根据其实际需求进行设定，例如，还可以将所述预置条件设定为节点路径长度不低于另一预定的长度阈值，以避免因节点路径过短而导致所对应的答案语句过于简洁，进而会因缺乏相应的修饰词/衔接词而使得答案语句太过生硬不够贴近人类自然语言。

[0132] 在对所述至少一个预定组合顺序在所述概率图谱中所对应的至少一个节点路径进行过滤处理的基础上，可继续针对过滤处理后所保留的节点路径，进行候选答案语句的生成。

[0133] 本实施例通过基于预置条件对所述至少一个预定组合顺序在所述概率图谱中所对应的至少一个节点路径进行过滤，可以提升所得的候选答案语句在语言表述形式方面的质量，同时也可以降低后续对候选答案语句进行选择处理时的处理工作量，相对应地可提高从各候选答案语句中选择出最终答案语句的处理效率。

[0134] 参考图8，是本申请提供的一种数据处理方法实施例四的流程图，本实施例中，如图8所示，所述数据处理方法，还可以包括以下处理步骤：

[0135] 步骤105、基于预定评价方式，对所述至少一个候选答案语句中的每个候选答案语句进行评价，得到每个候选答案语句的评价结果。

[0136] 具体地，可通过提取每个候选答案语句的预定特征的特征信息，并基于每个候选答案语句的特征信息对每个候选答案语句进行评分，来得到每个候选答案语句的评价结果。

[0137] 其中，所述预定特征可以包括但不限于如下特征中的至少一个：

[0138] 特征1:候选答案语句的词序列在所述概率图谱中的出现概率;

[0139] 特征2:候选答案语句与问题数据的相似度;

[0140] 特征3:候选答案语句中主体、谓词及宾语的组合顺序在所述概率图谱中对应的节点路径长度的概率分布;

[0141] 特征4:候选答案语句的词序列中所包括的词语个数。

[0142] 其中,对于上述特征1,即候选答案语句的词序列在所述概率图谱中的出现概率,在上一实施例中已对其计算方式进行了介绍(即上一实施例中节点路径所对应的词序列在所述概率图谱中的出现概率),具体可参见上一实施例的描述,这里,对其计算方式不再进行赘述。其中,该特征1的取值越大,表示所对应的候选答案语句在日常生活或所属业务领域内使用越频繁,从而可以认为其在语言表述上较为贴近人类自然语言,反之,该特征1的取值越小,则表示所对应的候选答案语句在日常生活或所属业务领域内越不常使用,相对应地可以认为该语句于用户而言在语言表述上并不太贴近人类的自然语言。

[0143] 对于上述特征2,即候选答案语句与问题数据的相似度,本质上是指候选答案语句的句子文本与问题数据的句子文本间的文本相似度,而文本相似度的计算目前有很多可用算法,如最小公共子序列、字符串编辑距离、向量相似度等一系列算法,因此,可采用但不局限于上述算法中的任意一种来计算候选答案语句与问题数据的相似度数值。

[0144] 一般来说,候选答案语句与问题数据的相似度数值与候选答案语句的答案质量呈正向关系,即,若候选答案语句与问题数据的相似度数值较高,一般会认为该候选答案语句的答案质量较高。

[0145] 上述特征3,即候选答案语句中主体、谓词及宾语的组合顺序在所述概率图谱中对应的节点路径长度的概率分布,可利用以下的计算式计算得到:

[0146] $PP(seq)=\rho(sp)\rho(po)$;

[0147] 其中,$\rho(sp)$表示主体(Subject)到谓词(Predicate)的路径的路径长度的出现概率,$\rho(po)$表示谓词(Predicate)到宾语(Object)的路径的路径长度的出现概率。

[0148] $\rho(sp)$和$\rho(po)$的计算方式类似,本实施例仅举例说明$\rho(sp)$的计算过程,其中,$\rho(sp)$可通过如下的计算过程获得:

[0149] 在概率图谱中,由S到P可能存在多条路径,路径的长度可能存在多种取值,每种取值都可以得到一个概率值,例如假设路径长度取值为[10,15,20],每一种长度的频数为[3,4,3],则可知从S到P共有10条路径,则由S到P的的各路径的路径长度[10,15,20]的出现概率对应地为[0.3,0.4,0.3]。从而,如果某一候选语句中S到P的路径长度为15,则其$\rho(sp)$取值为0.4。

[0150] $\rho(po)$的计算方式与$\rho(sp)$计算方式类似,具体可参考$\rho(sp)$的上述计算方式,这里,针对$\rho(po)$的计算过程不再进行详述。

[0151] 上述特征3的取值越大,则其所对应的候选答案语句的路径在所有候选答案语句的路径中的占比越大,相应地可加大选择该候选答案语句的概率。

[0152] 上述特征4,可通过累计候选答案语句的词序列中所包括的词语的个数得到,其中,对于词序列中所包括的相同词语(同一词语在同一词序列中多次出现)不对其进行合并处理,按其实际出现次数进行累计。该特征4能够表示所对应的候选答案语句的精简程度,其值越小则候选答案语句的句子越精简。

[0153] 在针对每一个候选答案语句,提取出其上述特征(可以是以上四种特征中的任意一种或多种)的基础上,可基于所提取的特征信息对该候选答案语句进行评分。

[0154] 其中,具体实施中,作为一种可能的实施方式,可采用预先构建的评分模型,基于所提取的特征信息,来对候选答案语句进行评分,其中模型的输入为所提取的候选答案语句的各个特征取值(一般来说实际输入为对其进行向量化后的向量值),输出为候选答案语句的得分值。

[0155] 所述评分模型可预先通过对多条已标注了特征信息与得分情况的已有语句进行模型训练得到,模型的训练可以使用Lasso、决策树、随机森林、支持向量机等中任意一种算法,并不局限于某一个算法。

[0156] 作为另一种可能的实施方式,还可以根据经验对各特征设定相应权重,并基于所设定的权重通过对各特征的特征取值进行加权计算来得到候选答案语句的评分分值,当然还可以是其他的可行方式,本实施例不对候选答案语句的评分方式进行限定。

[0157] 骤106、选取评价结果最优的候选答案语句作为所述问题数据的答案语句。

[0158] 候选答案语句的评价结果可反映其优劣程度,也即反映其在语言表述形式方面是否较为贴近人类自然语言,从而,可基于候选答案语句的评价结果,来从各候选答案语句中择优选择其中之一作为问题数据的最终答案语句。

[0159] 根据上文所述可知,具体可通过候选答案语句的评分分值来体现其评价结果,从而,可根据候选答案语句的得分情况来进行最终答案语句的选取,其中,若候选答案语句的评分分值越高表示候选答案语句越优良,则可从各候选答案语句中选取出得分分值最高的候选答案语句,作为问题数据最终的答案语句并反馈给用户。

[0160] 本实施例通过利用相应评价机制,从问题数据所对应的至少一个候选答案语句中选取出评价结果最优的候选答案语句作为问题数据的最终答案语句,实现了从不同语言表述形式的候选答案语句选择空间中选取出了更加贴近人类自然语言的答案语句,可有效提升智能问答系统在语言表述形式方面的答案质量。

[0161] 本申请还提供一种服务器,所述服务器可以是本地/网络端单独部署的服务器,或者本地/网络端的服务器集群中的服务器,参考图9,是本申请提供的一种服务器实施例五的结构示意图,所述服务器包括:

[0162] 存储器901,用于至少存储一组指令集。

[0163] 所述存储器901可以是具备数据存储功能的各类型存储器件,如ROM(Read Only Memory,只读存储器),FLASH,光盘,硬盘等,所存储的所述指令集用于指示处理器902执行如下文所述的相应数据处理操作,该指令集可以以程序形式存储于所述存储器901中。

[0164] 处理器902,用于调用并执行所述存储器中的所述指令集,通过执行所述指令集进行以下操作:

[0165] 获得问题数据;

[0166] 获得所述问题数据所对应的答案信息;

[0167] 确定所述答案信息所对应的答案模型;所述答案模型基于所述问题数据所对应的问题模型得到,所述答案模型包括构成答案语句所需的至少一个预定要素的要素信息;

[0168] 基于所述答案模型,生成所述问题数据对应的至少一个候选答案语句。

[0169] 所述问题数据可以是但不限于用户在智能问答场景中基于其实际需求提交至智

能问答系统的问题语句，如用户在其智能手机、笔记本等终端设备中通过智能问答APP(Application，应用程序)的应用界面或网页提交至智能问答系统的问题语句等。

[0170] 相对应地，所述获得问题数据，则可以是服务器基于智能问答系统所获得的用户提交的问题语句，如用户提交的“什么药可以治疗感冒?”等。其中，本地/网络端的服务器/服务器集群中的服务器，具体可基于局域网、互联网或移动网络等网络的通信功能获得用户在其终端设备所提交的问题语句。

[0171] 需要说明的是，实际应用中，所述用户可以是自然人用户或者基于人工智能的非自然人用户，这里，并不对此进行限定。

[0172] 通常来说，问题数据中会携带疑问点信息，以反映用户的提问需求，如上述的问题语句“什么药可以治疗感冒?”即携带了“什么(药)”这一疑问点信息。

[0173] 所述答案信息可以是与所述问题数据中携带的疑问点信息相对应且与疑问点信息之外的其他部分相匹配的、能够解决问题的答案点信息，如针对上述问题语句“什么药可以治疗感冒?”中的疑问点信息“什么(药)”，则所述答案信息相对应地可以为能够治疗感冒的答案点信息“阿莫西林”。

[0174] 本步骤具体可通过对问题数据进行问句解析，确定出其中的疑问点信息，进而基于相应业务知识(如预先制定的涵盖了本领域或多个领域业务知识的知识库等)确定出与所述疑问点信息相对应且与疑问点信息之外的其他部分相匹配的、能够解决问题的答案点信息。

[0175] 其中，所述疑问点信息一般来说是一些能够反映用户提问需求的疑问词，如“什么”、“哪些”、“哪个”、“哪儿”、“几个”、“如何”等等。

[0176] 对于待解答的问题数据，本申请在获得其答案信息后(如上述的答案点信息)，并不直接为其生成相对应的答案语句，而是确定答案信息所对应的答案模型，该答案模型包括构成答案语句所需的至少一个预定要素的要素信息，可将其形象地理解为一包括了答案语句的各要素信息的框架模型。

[0177] 容易理解的是，所述答案模型中应至少包括所述对应于疑问点信息且与问题语句中疑问点信息之外的其他部分相匹配的、能够解决问题的答案点信息。

[0178] 其中，所述答案模型具体可基于所述答案信息中的答案点信息以及问题数据所对应的问题模型确定出，所述问题模型相对应地包括构成问题语句所需的至少一个预定要素的要素信息，该问题模型同样可相应形象地理解为一包括了问题语句的各要素信息的框架模型，且所述问题模型应至少包括疑问点信息。

[0179] 具体地，发明人经研究发现，实际应用中，对于待解答的问题数据，一般来说，可通过将其问题语句中所包括的疑问点信息(如疑问词“什么(药)”)替换为所确定出的答案点信息来得到所对应的答案语句，也即，将所确定出的答案点信息与问题语句中非答案点信息的部分进行拼接来得出问题语句所对应的答案语句，如针对上述的问题语句“什么药可以治疗感冒?”，可通过将疑问点信息“什么(药)”替换为答案点信息“阿莫西林”，来得到其对应的答案语句“阿莫西林可以治疗感冒”。基于此，从模型角度来说，问题语句所对应的答案语句的答案模型可以与问题语句的问题模型具备相类似的模型结构，不同之处仅在于答案模型相比于问题模型来说，将问题模型中的疑问点信息替换为了答案点信息。

[0180] 鉴于此，本步骤中，具体可对问题语句进行问句解析，提取其问题模型，进而通过

将问题模型中的疑问点信息替换为答案点信息，而模型的其他部分维持不变来得到问题语句的答案语句所对应的答案模型。

[0181] 如前文所述，所述答案模型为一包括了答案语句的各要素信息的框架模型，但其并不足以构成一完整的问题语句，为了支持智能问答系统能够确定出更加贴近人类自然语言的答案语句，本步骤基于所述答案模型，生成所述问题数据对应的至少一个候选答案语句。

[0182] 容易理解的是，每个候选答案语句都是符合所述答案模型的模型要求的，也即，每个候选答案语句都包括了构成答案语句所需的各个要素信息，从而，从准确度方面来说，各个候选答案语句无明显差别，都涵盖了解决问题所需的答案点信息能够解决问题，区别主要在于语言表述形式(语句形式)方面的不同。

[0183] 实际应用中，可根据业务需求，通过线上(即从网络上收集)和/或线下方式收集本领域或多个领域的高质量问答语句，来预先构建语料库(知识库)，所构建的预料库(知识库)可尽可能基于特定领域或多个领域的业务特点，体现特定领域或多个领域的自然语言语句描述方式、风格，以使得基于该预料库(知识库)所确定出的答案语句更加贴近人类自然语言。

[0184] 在构建语料库(知识库)的基础上，可基于预料库(知识库)生成符合所述答案模型的模型要求的至少一个候选答案语句，以使得为问题数据所对应的答案语句的确定提供语言表述形式方面的选择空间，后续，可基于相应的选择策略/选择机制从所述至少一个候选答案语句中选择其中之一作为所述问题数据最终的答案语句。

[0185] 根据以上方案可知，本实施例提供的服务器，在获得问题数据以及问题数据所对应的答案信息基础上，基于问题数据所对应的问题模型确定所述答案信息所对应的答案模型，并进而基于所述答案模型生成问题数据对应的至少一个候选答案语句，由于针对问题数据，生成了符合其答案模型的至少一个候选答案语句，从而为答案语句的确定提供了语言表述形式方面的选择空间，能够支持智能问答系统从中选择出更加贴近人类自然语言的答案语句，进而可有效提升智能问答系统在语言表述形式方面的答案质量。

[0186] 在接下来的实施例六中，继续对上述服务器中处理器902的数据处理功能进行进一步详述。本实施例中，所述处理器902具体可通过执行以下处理实现其数据处理功能：

[0187] 获得问题数据；

[0188] 提取所述问题数据所对应的第一SPO(Subject-Predicate-Object，主体-谓词-宾语)模型；所述第一SPO模型中的主体或谓词或宾语包括所述问题数据中所携带的疑问点信息；

[0189] 查询预定的知识库，获得所述知识库中与所述疑问点信息相对应且与所述第一SPO模型中的未包括疑问点信息的部分相匹配的答案点信息；

[0190] 将所述答案点信息与所述第一SPO模型中的未包括疑问点信息的部分整合为第二SPO模型，所述第二SPO模型为所述答案信息所对应的答案模型，所述第二SPO模型中的主体、谓词、宾语为所述答案模型包括的所述要素信息；

[0191] 搜索预定的概率图谱，确定所述第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置；

[0192] 基于所述第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点

位置，确定所述第二SPO模型中主体、谓词及宾语的至少一个预定组合顺序在所述概率图谱中所对应的至少一个节点路径；

[0193] 获得每个节点路径所对应的词序列，每个词序列构成的语句作为所述问题数据对应的一个候选答案语句，得到所述问题数据对应的至少一个候选答案语句。

[0194] 所述问题数据可以是但不限于用户在智能问答场景中基于其实际需求提交至智能问答系统的问题语句，如用户在其智能手机、笔记本等终端设备中通过智能问答APP的应用界面或网页提交至智能问答系统的问题语句等。

[0195] 相对应地，所述获得问题数据，则可以是服务器基于智能问答系统所获得的用户提交的问题语句，如用户提交的“什么药可以治疗感冒?”等。其中，本地/网络端的服务器/服务器集群中的服务器，具体可基于局域网、互联网或移动网络等网络的通信功能获得用户在其终端设备所提交的问题语句。

[0196] 需要说明的是，实际应用中，所述用户可以是自然人用户或者基于人工智能的非自然人用户，这里，并不对此进行限定。

[0197] 其中，可首先分别基于相应的实体识别技术及关系提取技术，识别问题数据的问题语句中所包括的实体和关系。具体地，对于实体识别，目前已存在很多可用算法，例如CRF(conditional random field，条件随机场)、HMM(Hidden Markov Model，隐马尔可夫模型)等算法，从而，可选取这些算法中的其中一种对问题语句中的实体进行识别，或者亦可以通过模式匹配的方式进行实体识别，这里不局限于一种技术。所述关系是指句子中实体与实体间的谓词关系，从而问题语句中关系的识别可通过对问题语句进行谓词识别来实现，而对于谓词，则具体可通过对问题语句进行词法分析及句法分析，并在词法分析及句法分析的基础上，通过融合词法和句法特征来实现谓词识别。

[0198] 例如，对于问题语句“什么药可以治疗感冒?”通过对其进行实体及关系识别，可以获知其包括“什么药”(本质上是以疑问词形式指代的实体)及“感冒”两个实体，包括“治疗”这一谓词关系。

[0199] 需要说明的是，通常来说，问题数据的问题语句中会携带疑问点信息，以反映用户的提问需求，如上述的问题语句“什么药可以治疗感冒?”即携带了“什么(药)”这一疑问点信息，问题语句“阿莫西林如何治疗感冒”则携带了“如何(治疗)”这一疑问点信息。

[0200] 问题语句中的疑问点信息一般来说是上文所述的“什么”、“哪些”、“哪个”、“哪儿”、“几个”或“如何”等疑问词，其通常包括在问题语句的实体或者关系中，从而，可通过对识别出的实体或关系进一步进行疑问词识别，来获知问题语句中的疑问点信息。仍以上述的问题语句“什么药可以治疗感冒?”为例，在识别出其两个实体“什么药”、“感冒”以及一个关系“治疗”的基础上，可进一步通过对实体及关系进行疑问词识别来获知实体“什么药”中包括疑问词“什么(药)”，当然，还可以直接通过对问题语句进行疑问词识别来获知其疑问点信息(而并非在实体、关系识别的基础上进行疑问词识别)，本实施例并不对此进行限定。

[0201] 其中，由于疑问词数量有限，具体实施中可以通过模式匹配的方式对问题语句中的疑问词进行识别。

[0202] 在实体及关系识别的基础上，可通过继续对问题语句进行问句句法解析，来获知问题语句中的实体及关系属于主体(Subject)、谓词(Predicate)、宾语(Object)中的哪一种，进而提取出问题语句所对应的第一SPO模型，其中，所述第一SPO模型中的主体或谓词或

宾语包括所述问题数据中所携带的疑问点信息。

[0203] 如对于问题语句“什么药可以治疗感冒?”,通过上述处理过程,可以提取出其第一SPO模型:“(什么)药(S)-治疗(P)-感冒(O)”,在该示例中,主体中包括疑问词“什么(药)”;对于问题语句“阿司匹林如何治疗感冒”,通过上述处理过程,可以提取出其第一SPO模型:“阿司匹林-(如何)治疗-感冒”,在该示例中,谓词中包括疑问词“如何(治疗)”。

[0204] 所述预定的知识库,可以是但不限于基于所收集的某一特定业务领域或多个业务领域的业务知识所构建的知识图谱,具体地,所述知识图谱可以为一预先基于所述某一特定业务领域或多个业务领域的能够反映业务知识的一系列问答语句样本所构建的有向图,该有向图包括多个节点,图中的每个节点对应一个词语(实体词语),任意两个节点之间的边为有向边,有向边所指向的节点的词语依存于有向边所背离的节点的词语,任意两个节点之间的边对应一谓词关系,用于表示边的两个节点所对应的两个词语间的关系。

[0205] 具体可将所述知识图谱所提供的有向图作为搜索空间,通过查询所述知识图谱,来获得所述知识图谱中与所述疑问点信息相对应且与所述第一SPO模型中的未包括疑问点信息的部分相匹配的答案点信息。

[0206] 如对于问题语句“什么药可以治疗感冒?”的第一SPO模型“(什么)药-治疗-感冒”,通过查询所述知识图谱,可获得与疑问点信息“什么(药)”相对应且与“治疗-感冒”相匹配的答案点信息“阿司匹林”。

[0207] 具体地,可通过将所获得的答案点信息替代问题语句的第一SPO模型中的疑问点信息,实现答案点信息与所述第一SPO模型中的未包括疑问点信息的部分的整合,以此得到问题语句所对应的答案语句的第二SPO模型。

[0208] 仍以上述的问题语句“什么药可以治疗感冒?”为例,可将答案点信息“阿司匹林”替代其第一SPO模型“(什么)药-治疗-感冒”中的“什么(药)”,来得到其问题语句的第二SPO模型“阿司匹林-治疗-感冒”。

[0209] 其中,该第二SPO模型中包括构成问题语句所需的各个要素(主体、谓词、宾语)的要素信息,但其并不足以构成一完整的、符合人类自然语言的问题语句。

[0210] 其中,所述概率图谱为一预先基于某一特定业务领域或多个业务领域的一系列问答语句样本所构建的包括多个节点的有向图,图中的每个节点对应一个词语(可以是实体词语、谓词词语、形容词词语、复合型词语等各类型词语),任意两个节点之间的边为有向边,有向边所指向的节点的词语依存于有向边所背离的节点的词语,任意两个节点之间的边对应一概率数值,边所对应的概率数值表示边的两个节点的两个词语同时出现的频率与边所背离的节点的词语出现的频率的比值。参考图3,为本实施例提供的概率图谱的一示意图。

[0211] 具体地,可通过以下处理过程来构建上述的概率图谱:

[0212] 对收集的一系列问答语句样本进行分词,分词之后句子以词序列的形式存在,句子所表达的语义蕴含在词语之间的依存关系之中,对于每一个句子,依据其词序列的先后关系,通过边(即有向图的边)将词序列进行串联得到该句子的语义路径,并将各个句子的语义路径进行合并关联,得到整个语料的语义图,其中语义图中的每个节点为一个词语,而词语与词语在句子中存在的先后关系则体现为语义图中边的指向关系,也即,边的方向体现了边的两个节点所对应的词语在句子中的先后顺序,该先后顺序也体现了该两个词语在

句子中的依存关系，其中顺序在前的词语在有向边的发出端，顺序在后的词语在有向边的箭头端，且有向边所指向的节点的词语依存于有向边所背离的节点的词语，在构建出语义图的基础上，进一步对语义图中各节点的词语进行词频统计及概率计算来为每个有向边赋予相对应的概率值(边箭头端词语和发出端词语共现的频数与边发出端词语频数的比值)，从而得到所述概率图谱。

[0213] 需要说明的是，在概率图谱中边是有方向的，不具有对称性，因此对于两个邻近的节点，节点之间可能存在两条边，每个节点可能存在多个入边和出边，例如，如图4(a)及图4(b)所示，在图4(a)中，节点A到节点B存在一条有向边，则节点A与节点B之间存在依存关系，且节点B依存于节点A，在图4(b)中，节点A与节点B之间存在两条有向边，则节点A与节点B之间存在依存关系，对应于从节点A指向节点B的有向边，节点B依存于节点A，而对应于从节点B指向节点A的有向边，则节点A依存于节点B。

[0214] 对于所获得第二SPO模型，可通过搜索该概率图谱，来确定所述第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置，参考图5所示的在概率图谱中基于搜索来定位第二SPO模型中主体、谓词及宾语的节点位置的示例图，在该示例中，通过对概率图谱进行搜索，确定出第二SPO模型中的主体、谓词及宾语分别位于概率图谱有向图中的word8、word9及word10这些节点位置处。

[0215] 所述至少一个预定组合顺序可以包括所述第二SPO模型中主体、谓词及宾语的所有组合顺序，如SPO，SOP、PSO、POS、OSP和OPS等。或者还可以仅包括所述第二SPO模型中主体、谓词及宾语的某一特定顺序，如SPO等，本实施例对此不作限定。

[0216] 以SPO这一组合顺序为例，附图5中，所述第二SPO模型中主体、谓词及宾语的SPO组合顺序在所述概率图谱中所对应的节点路径包括：

[0217] word8→word5→word9→word7→word10；

[0218] word8→word3→word9→word7→word10；

[0219] word8→word1→word3→word9→word7→word10。

[0220] 由于概率图谱中每个节点均对应一个词语，从而可基于概率图谱，获得每个上述节点路径所对应的词序列，其中，每个词序列构成的语句作为所述问题数据对应的一个候选答案语句，从而得到所述问题数据对应的至少一个候选答案语句。

[0221] 其中，每个候选答案语句都包括了构成答案语句所需的各要素的要素信息，相对应地都能够解决问题语句所体现的问题，区别仅在于或语言表述形式方面的不同，例如，对于问题语句“什么药可以治疗感冒?”，假设通过上述的节点搜索定位及节点路径上的词序列获取等处理，可以得到“阿司匹林-可以-治疗-感冒”、“阿司匹林-治疗-感冒”、“治疗-感冒-用-阿司匹林”“感冒-治疗-药-阿司匹林”等多个词序列，则相对应地可得到该问题语句的多个候选答案语句：“阿司匹林可以治疗感冒”、“阿司匹林治疗感冒”、“治疗感冒用阿司匹林”、“感冒治疗药阿司匹林”，每个候选答案语句都能够解决问题语句所体现的问题，区别仅在于语言表述形式方面的不同。

[0222] 通过生成问题数据的至少一个候选答案语句，可以为问题数据所对应的答案语句的确定提供语言表述形式方面的选择空间，后续，可基于相应的选择策略/选择机制(如基于相关概率计算的选择策略、基于节点路径长度控制的选择策略等)从所述至少一个候选答案语句中选择其中之一作为所述问题数据最终的答案语句。

[0223] 本实施例通过搜索概率图谱，为答案模型生成了符合其模型要求的至少一个候选答案语句，从而为答案语句的确定提供了语言表述形式方面的选择空间，能够支持智能问答系统从中选择出更加贴近人类自然语言的答案语句，进而，可有效提升智能问答系统在语言表述形式方面的答案质量。

[0224] 在接下来的实施七中，所述服务器中的处理器902还可以通过调用存储器901中指令集中的相应指令，在获得至少一个节点路径后执行以下操作：

[0225] 滤除所述至少一个预定组合顺序在所述概率图谱中所对应的节点路径中不符合预置条件的节点路径。

[0226] 其中，所述预置条件可以包括：节点路径长度不超出预定长度阈值，和/或，节点路径所对应的词序列在所述概率图谱中的出现概率不低于预定概率阈值。

[0227] 节点路径长度是指节点路径上所包括的节点的数量。

[0228] 节点路径所对应的词序列在所述概率图谱中的出现概率，可通过以下的计算式计算得出：

[0229] $P(seq)=\rho(w_n|w_{n-1})\cdots\rho(w_i|w_{i-1})\cdots\rho(w_2|w_1)\rho(w_1)$；

[0230] 其中，seq表示节点路径所对应的词序列$w_1,w_2,...,w_n$，w_i（$1\leqslant i\leqslant n$，i为自然数）表示seq的第i个词语，P(seq)表示seq在所述概率图谱中出现的概率，$\rho(w_1)$表示seq中首个词w_1在概率图谱中出现的概率（$\rho(w_1)=w_1$的词频/概率图谱中各个词的词频的累计值），$\rho(w_i|w_{i-1})$表示词语在w_{i-1}条件下w_i的概率，即构建的概率图谱中从节点w_{i-1}到节点w_i的边上的概率。

[0231] 一般来说，若节点路径长度过长（超出预定长度阈值），会使得所确定出的候选答案语句过于复杂，于用户而言则答案语句不够简洁；若节点路径所对应的词序列在所述概率图谱中的出现概率过低（低于预定概率阈值），则考虑该节点路径对应的语句在日常生活或所属业务领域内不常使用，相对应地会认为该语句于用户而言并不太贴近人类的自然语言。鉴于此，在基于节点路径对应的词序列确定问题语句的候选答案语句之前，可首先基于上述的预置条件，将路径长度超出预定长度阈值的节点路径和/或所对应的词序列在所述概率图谱中的出现概率不低于预定概率阈值的节点路径滤除掉，以使得所保留的各节点路径所对应的语句于用户而言不会过于复杂和/或过于偏门（不贴仅人类自然语言）。

[0232] 具体实施中，所述阈值条件不必局限于本实施例所提供的上述内容，可以由技术人员或者用户根据其实际需求进行设定，例如，还可以将所述预置条件设定为节点路径长度不低于另一预定的长度阈值，以避免因节点路径过短而导致所对应的答案语句过于简洁，进而会因缺乏相应的修饰词/衔接词而使得答案语句太过生硬不够贴近人类自然语言。

[0233] 在对所述至少一个预定组合顺序在所述概率图谱中所对应的至少一个节点路径进行过滤处理的基础上，可继续针对过滤处理后所保留的节点路径，进行候选答案语句的生成。

[0234] 本实施例通过基于预置条件对所述至少一个预定组合顺序在所述概率图谱中所对应的至少一个节点路径进行过滤，可以提升所得的候选答案语句在语言表述形式方面的质量，同时也可以降低后续对候选答案语句进行选择处理时的处理工作量，相对应地可提高从各候选答案语句中选择出最终答案语句的处理效率。

[0235] 在接下来的实施八中，所述服务器中的处理器902还可以通过调用存储器901中指

令集中的相应指令，执行以下操作：

[0236] 基于预定评价方式，对所述至少一个候选答案语句中的每个候选答案语句进行评价，得到每个候选答案语句的评价结果；

[0237] 选取评价结果最优的候选答案语句作为所述问题数据的答案语句。

[0238] 具体地，可通过提取每个候选答案语句的预定特征的特征信息，并基于每个候选答案语句的特征信息对每个候选答案语句进行评分，来得到每个候选答案语句的评价结果。

[0239] 其中，所述预定特征可以包括但不限于如下特征中的至少一个：

[0240] 特征1：候选答案语句的词序列在所述概率图谱中的出现概率；

[0241] 特征2：候选答案语句与问题数据的相似度；

[0242] 特征3：候选答案语句中主体、谓词及宾语的组合顺序在所述概率图谱中对应的节点路径长度的概率分布；

[0243] 特征4：候选答案语句的词序列中所包括的词语个数。

[0244] 其中，对于上述特征1，即候选答案语句的词序列在所述概率图谱中的出现概率，在上一实施例中已对其计算方式进行了介绍(即上一实施例中节点路径所对应的词序列在所述概率图谱中的出现概率)，具体可参见上一实施例的描述，这里，对其计算方式不再进行赘述。其中，该特征1的取值越大，表示所对应的候选答案语句在日常生活或所属业务领域内使用越频繁，从而可以认为其在语言表述上较为贴近人类自然语言，反之，该特征1的取值越小，则表示所对应的候选答案语句在日常生活或所属业务领域内越不常使用，相对应地可以认为该语句于用户而言在语言表述上并不太贴近人类的自然语言。

[0245] 对于上述特征2，即候选答案语句与问题数据的相似度，本质上是指候选答案语句的句子文本与问题数据的句子文本间的文本相似度，而文本相似度的计算目前有很多可用算法，如最小公共子序列、字符串编辑距离、向量相似度等一系列算法，因此，可采用但不局限于上述算法中的任意一种来计算候选答案语句与问题数据的相似度数值。

[0246] 一般来说，候选答案语句与问题数据的相似度数值与候选答案语句的答案质量呈正向关系，即，若候选答案语句与问题数据的相似度数值较高，一般会认为该候选答案语句的答案质量较高。

[0247] 上述特征3，即候选答案语句中主体、谓词及宾语的组合顺序在所述概率图谱中对应的节点路径长度的概率分布，可利用以下的计算式计算得到：

[0248] PP(seq)＝ρ(sp)ρ(po)；

[0249] 其中，ρ(sp)表示主体(Subject)到谓词(Predicate)的路径的路径长度的出现概率，ρ(po)表示谓词(Predicate)到宾语(Object)的路径的路径长度的出现概率。

[0250] ρ(sp)和ρ(po)的计算方式类似，本实施例仅举例说明ρ(sp)的计算过程，其中，ρ(sp)可通过如下的计算过程获得：

[0251] 在概率图谱中，由S到P可能存在多条路径，路径的长度可能存在多种取值，每种取值都可以得到一个概率值，例如假设路径长度取值为[10,15,20]，每一种长度的频数为[3,4,3]，则可知从S到P共有10条路径，则由S到P的的各路径的路径长度[10,15,20]的出现概率对应地为[0.3,0.4,0.3]。从而，如果某一候选语句中S到P的路径长度为15，则其ρ(sp)取值为0.4。

[0252] ρ(po)的计算方式与ρ(sp)计算方式类似，具体可参考ρ(sp)的上述计算方式，这里，针对ρ(po)的计算过程不再进行详述。

[0253] 上述特征3的取值越大，则其所对应的候选答案语句的路径在所有候选答案语句的路径中的占比越大，相应地可加大选择该候选答案语句的概率。

[0254] 上述特征4，可通过累计候选答案语句的词序列中所包括的词语的个数得到，其中，对于词序列中所包括的相同词语(同一词语在同一词序列中多次出现)不对其进行合并处理，按其实际出现次数进行累计。该特征4能够表示所对应的候选答案语句的精简程度，其值越小则候选答案语句的句子越精简。

[0255] 在针对每一个候选答案语句，提取出其上述特征(可以是以上四种特征中的任意一种或多种)的基础上，可基于所提取的特征信息对该候选答案语句进行评分。

[0256] 其中，具体实施中，作为一种可能的实施方式，可采用预先构建的评分模型，基于所提取的特征信息，来对候选答案语句进行评分，其中模型的输入为所提取的候选答案语句的各个特征取值(一般来说实际输入为对其进行向量化后的向量值)，输出为候选答案语句的得分值。

[0257] 所述评分模型可预先通过对多条已标注了特征信息与得分情况的已有语句进行模型训练得到，模型的训练可以使用Lasso、决策树、随机森林、支持向量机等中任意一种算法，并不局限于某一个算法。

[0258] 作为另一种可能的实施方式，还可以根据经验对各特征设定相应权重，并基于所设定的权重通过对各特征的特征取值进行加权计算来得到候选答案语句的评分分值，当然还可以是其他的可行方式，本实施例不对候选答案语句的评分方式进行限定。

[0259] 候选答案语句的评价结果可反映其优劣程度，也即反映其在语言表述形式方面是否较为贴近人类自然语言，从而，可基于候选答案语句的评价结果，来从各候选答案语句中择优选择其中之一作为问题数据的最终答案语句。

[0260] 根据上文所述可知，具体可通过候选答案语句的评分分值来体现其评价结果，从而，可根据候选答案语句的得分情况来进行最终答案语句的选取，其中，若候选答案语句的评分分值越高表示候选答案语句越优良，则可从各候选答案语句中选取出得分分值最高的候选答案语句，作为问题数据最终的答案语句并反馈给用户。

[0261] 本实施例通过利用相应评价机制，从问题数据所对应的至少一个候选答案语句中选取出评价结果最优的候选答案语句作为问题数据的最终答案语句，实现了从不同语言表述形式的候选答案语句选择空间中选取出了更加贴近人类自然语言的答案语句，可有效提升智能问答系统在语言表述形式方面的答案质量。

[0262] 本申请还提供一种电子设备，所述电子设备可以是智能手机、平板电脑、台式机、笔记本、一体机等各类型终端设备。参考图10，是本申请提供的一种电子设备实施例九的结构示意图，所述电子设备包括：

[0263] 存储器1001，用于至少存储一组指令集。

[0264] 所述存储器1001可以是具备数据存储功能的各类型存储器件，如ROM，FLASH，光盘，硬盘等，所存储的所述指令集用于指示处理器1002执行如下文所述的相应数据处理操作，该指令集可以以程序形式存储于所述存储器1001中。

[0265] 处理器1002，用于调用并执行所述存储器中的所述指令集，通过执行所述指令集

进行以下操作：

[0266] 获得问题数据；

[0267] 获得所述问题数据所对应的答案信息；

[0268] 确定所述答案信息所对应的答案模型；所述答案模型基于所述问题数据所对应的问题模型得到，所述答案模型包括构成答案语句所需的至少一个预定要素的要素信息；

[0269] 基于所述答案模型，生成所述问题数据对应的至少一个候选答案语句。

[0270] 所述问题数据可以是但不限于用户在智能问答场景中基于其实际需求提交至智能问答系统的问题语句，如用户在其智能手机、笔记本等终端设备中通过智能问答APP(Application，应用程序)的应用界面或网页提交至智能问答系统的问题语句等。

[0271] 相对应地，所述获得问题数据，则可以是电子设备如上述的各类型终端设备直接基于其相应问题输入界面所提供的问题输入/提交功能获得用户所输入或提交的问题语句。

[0272] 需要说明的是，实际应用中，所述用户可以是自然人用户或者基于人工智能的非自然人用户，这里，并不对此进行限定。

[0273] 通常来说，问题数据中会携带疑问点信息，以反映用户的提问需求，如上述的问题语句“什么药可以治疗感冒?”即携带了“什么(药)”这一疑问点信息。

[0274] 所述答案信息可以是与所述问题数据中携带的疑问点信息相对应且与疑问点信息之外的其他部分相匹配的、能够解决问题的答案点信息，如针对上述问题语句“什么药可以治疗感冒?”中的疑问点信息“什么(药)”，则所述答案信息相对应地可以为能够治疗感冒的答案点信息“阿莫西林”。

[0275] 本步骤具体可通过对问题数据进行问句解析，确定出其中的疑问点信息，进而基于相应业务知识(如预先制定的涵盖了本领域或多个领域业务知识的知识库等)确定出与所述疑问点信息相对应且与疑问点信息之外的其他部分相匹配的、能够解决问题的答案点信息。

[0276] 其中，所述疑问点信息一般来说是一些能够反映用户提问需求的疑问词，如“什么”、“哪些”、“哪个”、“哪儿”、“几个”、“如何”等等。

[0277] 对于待解答的问题数据，本申请在获得其答案信息后(如上述的答案点信息)，并不直接为其生成相对应的答案语句，而是确定答案信息所对应的答案模型，该答案模型包括构成答案语句所需的至少一个预定要素的要素信息，可将其形象地理解为一包括了答案语句的各要素信息的框架模型。

[0278] 容易理解的是，所述答案模型中应至少包括所述对应于疑问点信息且与问题语句中疑问点信息之外的其他部分相匹配的、能够解决问题的答案点信息。

[0279] 其中，所述答案模型具体可基于所述答案信息中的答案点信息以及问题数据所对应的问题模型确定出，所述问题模型相对应地包括构成问题语句所需的至少一个预定要素的要素信息，该问题模型同样可相应形象地理解为一包括了问题语句的各要素信息的框架模型，且所述问题模型应至少包括疑问点信息。

[0280] 具体地，发明人经研究发现，实际应用中，对于待解答的问题数据，一般来说，可通过将其问题语句中所包括的疑问点信息(如疑问词“什么(药)”)替换为所确定出的答案点信息来得到所对应的答案语句，也即，将所确定出的答案点信息与问题语句中非答案点信

息的部分进行拼接来得出问题语句所对应的答案语句，如针对上述的问题语句“什么药可以治疗感冒?”，可通过将疑问点信息“什么(药)”替换为答案点信息“阿莫西林”，来得到其对应的答案语句“阿莫西林可以治疗感冒”。基于此，从模型角度来说，问题语句所对应的答案语句的答案模型可以与问题语句的问题模型具备相类似的模型结构，不同之处仅在于答案模型相比于问题模型来说，将问题模型中的疑问点信息替换为了答案点信息。

[0281] 鉴于此，本步骤中，具体可对问题语句进行问句解析，提取其问题模型，进而通过将问题模型中的疑问点信息替换为答案点信息，而模型的其他部分维持不变来得到问题语句的答案语句所对应的答案模型。

[0282] 如前文所述，所述答案模型为一包括了答案语句的各要素信息的框架模型，但其并不足以构成一完整的问题语句，为了支持智能问答系统能够确定出更加贴近人类自然语言的答案语句，本步骤基于所述答案模型，生成所述问题数据对应的至少一个候选答案语句。

[0283] 容易理解的是，每个候选答案语句都是符合所述答案模型的模型要求的，也即，每个候选答案语句都包括了构成答案语句所需的各个要素信息，从而，从准确度方面来说，各个候选答案语句无明显差别，都涵盖了解决问题所需的答案点信息能够解决问题，区别主要在于语言表述形式(语句形式)方面的不同。

[0284] 实际应用中，可根据业务需求，通过线上(即从网络上收集)和/或线下方式收集本领域或多个领域的高质量问答语句，来预先构建语料库(知识库)，所构建的预料库(知识库)可尽可能基于特定领域或多个领域的业务特点，体现特定领域或多个领域的自然语言语句描述方式、风格，以使得基于该预料库(知识库)所确定出的答案语句更加贴近人类自然语言。

[0285] 在构建语料库(知识库)的基础上，可基于预料库(知识库)生成符合所述答案模型的模型要求的至少一个候选答案语句，以使得为问题数据所对应的答案语句的确定提供语言表述形式方面的选择空间，后续，可基于相应的选择策略/选择机制从所述至少一个候选答案语句中选择其中之一作为所述问题数据最终的答案语句。

[0286] 根据以上方案可知，本实施例提供的电子设备，在获得问题数据以及问题数据所对应的答案信息基础上，基于问题数据所对应的问题模型确定所述答案信息所对应的答案模型，并进而基于所述答案模型生成问题数据对应的至少一个候选答案语句，由于针对问题数据，生成了符合其答案模型的至少一个候选答案语句，从而为答案语句的确定提供了语言表述形式方面的选择空间，能够支持智能问答系统从中选择出更加贴近人类自然语言的答案语句，进而可有效提升智能问答系统在语言表述形式方面的答案质量。

[0287] 需要说明的是，本说明书中的各个实施例均采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似的部分互相参见即可。

[0288] 为了描述的方便，描述以上系统或装置时以功能分为各种模块或单元分别描述。当然，在实施本申请时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

[0289] 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软件产品的形式体现出来，该计算机软件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。

[0290] 最后,还需要说明的是,在本文中,诸如第一、第二、第三和第四等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0291] 以上所述仅是本发明的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也应视为本发明的保护范围。

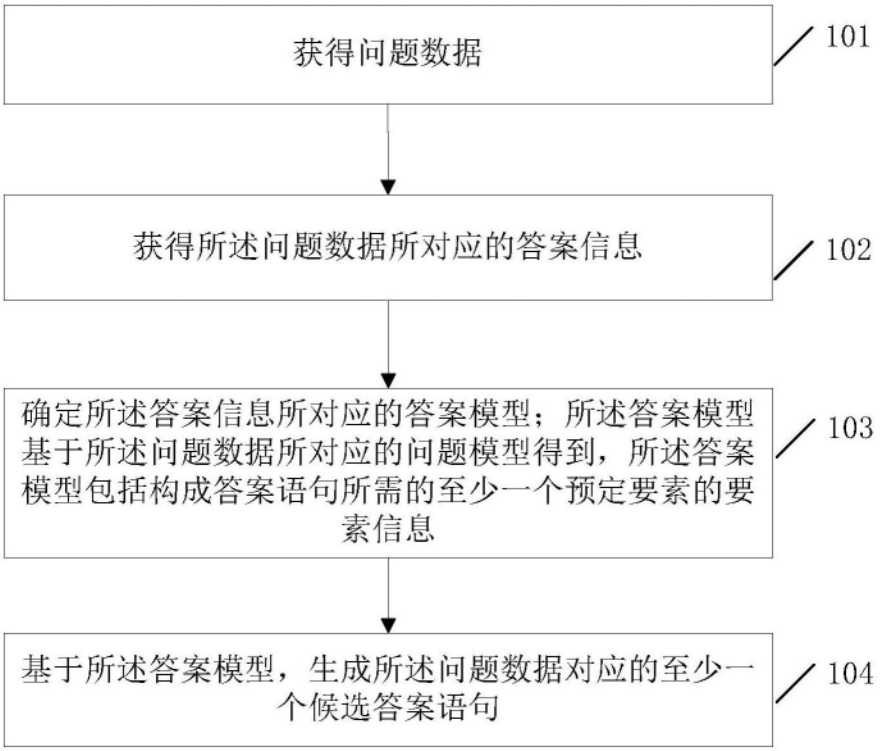

获得问题数据
101
获得所述问题数据所对应的答案信息
102
确定所述答案信息所对应的答案模型；所述答案模型基于所述问题数据所对应的问题模型得到，所述答案模型包括构成答案语句所需的至少一个预定要素的要素信息
103
基于所述答案模型，生成所述问题数据对应的至少一个候选答案语句
104

图1

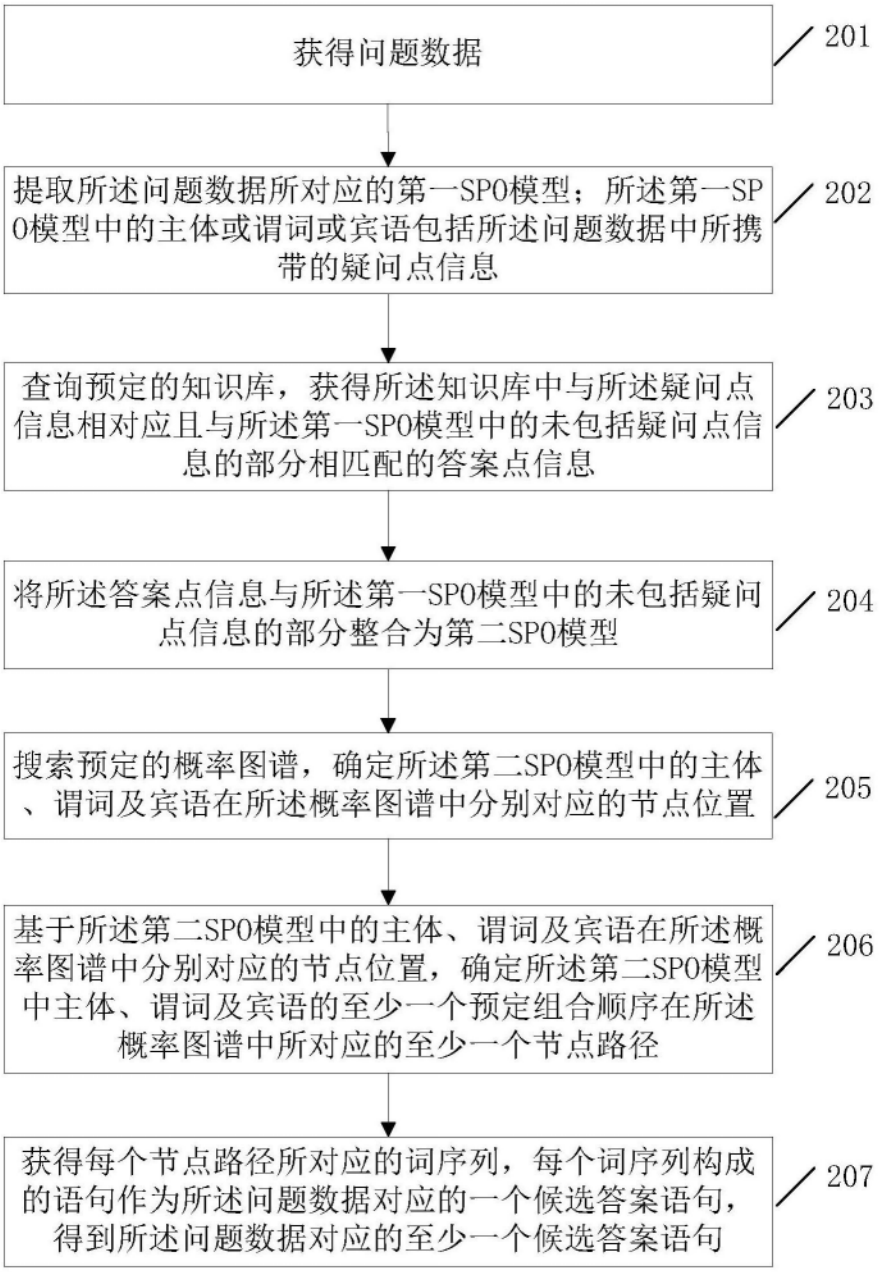

获得问题数据
201
提取所述问题数据所对应的第一SPO模型；所述第一SPO模型中的主体或谓词或宾语包括所述问题数据中所携带的疑问点信息
202
查询预定的知识库，获得所述知识库中与所述疑问点信息相对应且与所述第一SPO模型中的未包括疑问点信息的部分相匹配的答案点信息
203
将所述答案点信息与所述第一SPO模型中的未包括疑问点信息的部分整合为第二SPO模型
204
搜索预定的概率图谱，确定所述第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置
205
基于所述第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置，确定所述第二SPO模型中主体、谓词及宾语的至少一个预定组合顺序在所述概率图谱中所对应的至少一个节点路径
206
获得每个节点路径所对应的词序列，每个词序列构成的语句作为所述问题数据对应的一个候选答案语句，得到所述问题数据对应的至少一个候选答案语句
207

图2

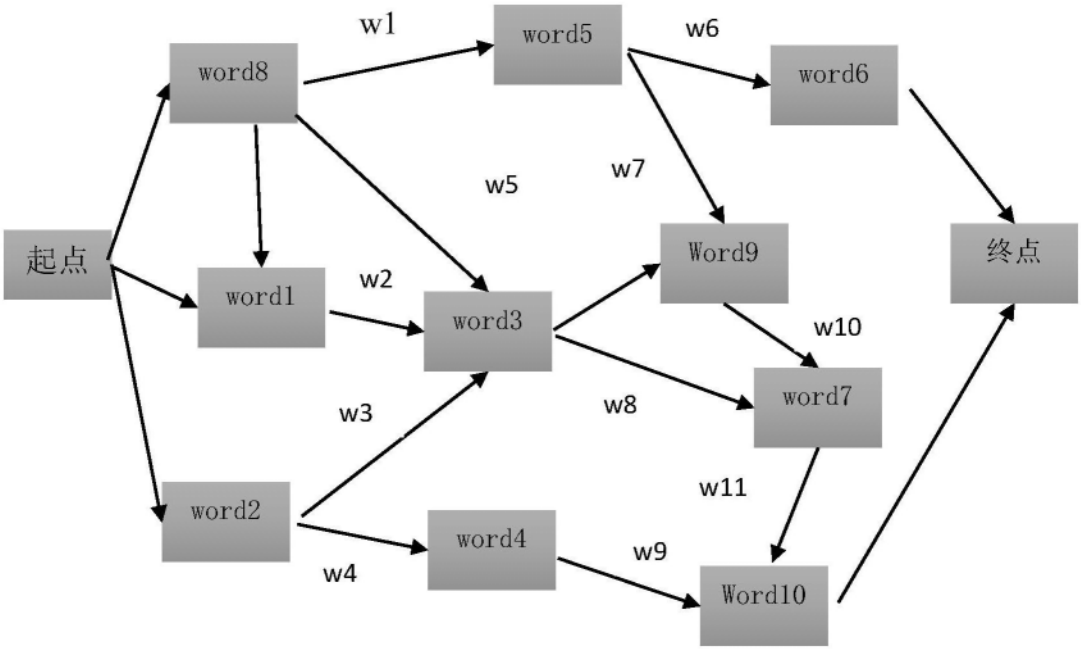

w1
word5
w6
word8
word6
w5
w7
起点
word1
w2
word3
Word9
终点
w10
w3
w8
word7
w11
word2
word4
w9
w4
Word10

图3

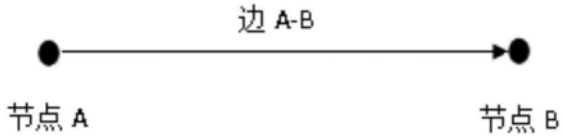

边 A-B
节点 A
节点 B

图4(a)


边 A-B
节点 A
边 B-A
节点 B

图4(b)

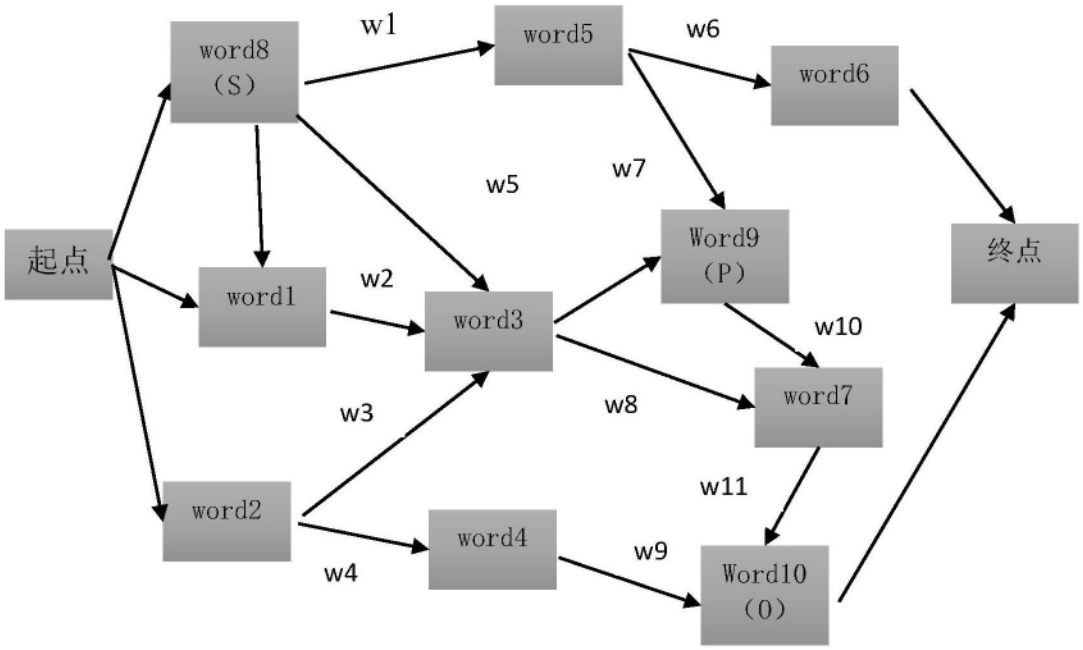

起点
word8
(S)
word5
word6
word1
word3
Word9
(P)
终点
word7
word2
word4
Word10
(0)
w1
w2
w3
w4
w5
w6
w7
w8
w9
w10
w11

图5

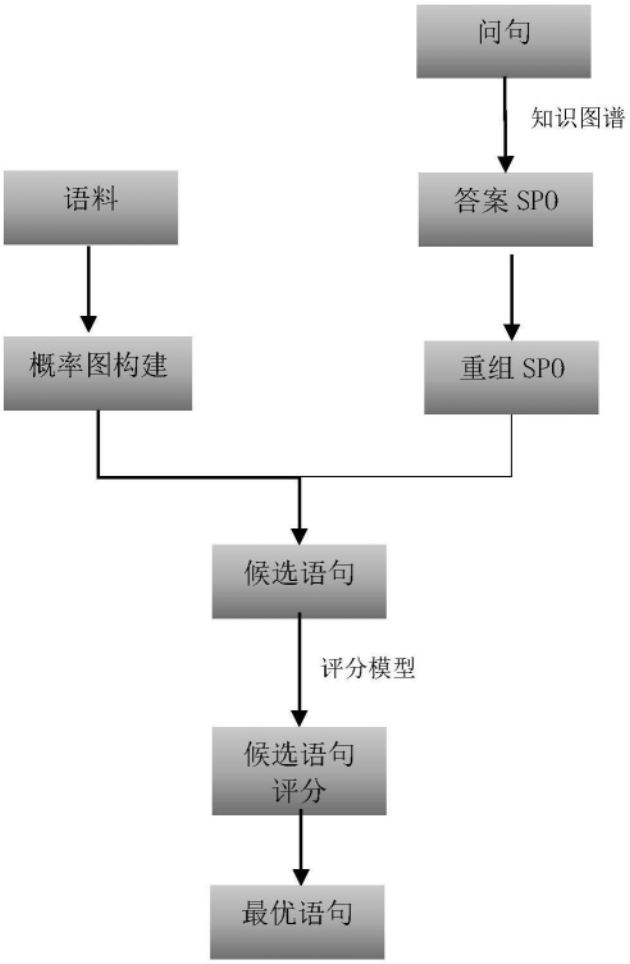


图6

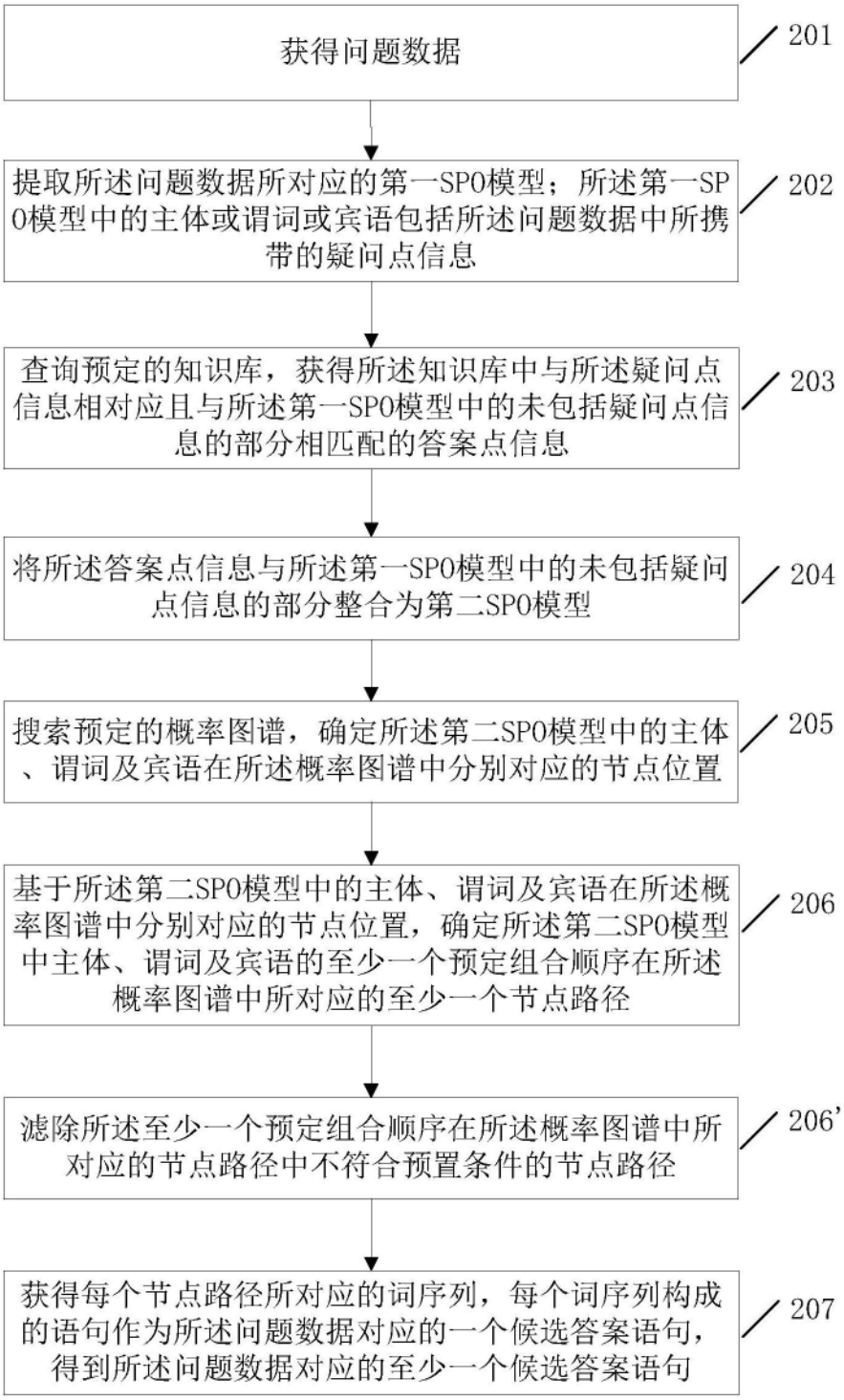

获得问题数据
201
提取所述问题数据所对应的第一SPO模型；所述第一SPO模型中的主体或谓词或宾语包括所述问题数据中所携带的疑问点信息
202
查询预定的知识库，获得所述知识库中与所述疑问点信息相对应且与所述第一SPO模型中的未包括疑问点信息的部分相匹配的答案点信息
203
将所述答案点信息与所述第一SPO模型中的未包括疑问点信息的部分整合为第二SPO模型
204
搜索预定的概率图谱，确定所述第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置
205
基于所述第二SPO模型中的主体、谓词及宾语在所述概率图谱中分别对应的节点位置，确定所述第二SPO模型中主体、谓词及宾语的至少一个预定组合顺序在所述概率图谱中所对应的至少一个节点路径
206
滤除所述至少一个预定组合顺序在所述概率图谱中所对应的节点路径中不符合预置条件的节点路径
206'
获得每个节点路径所对应的词序列，每个词序列构成的语句作为所述问题数据对应的一个候选答案语句，得到所述问题数据对应的至少一个候选答案语句
207

图7

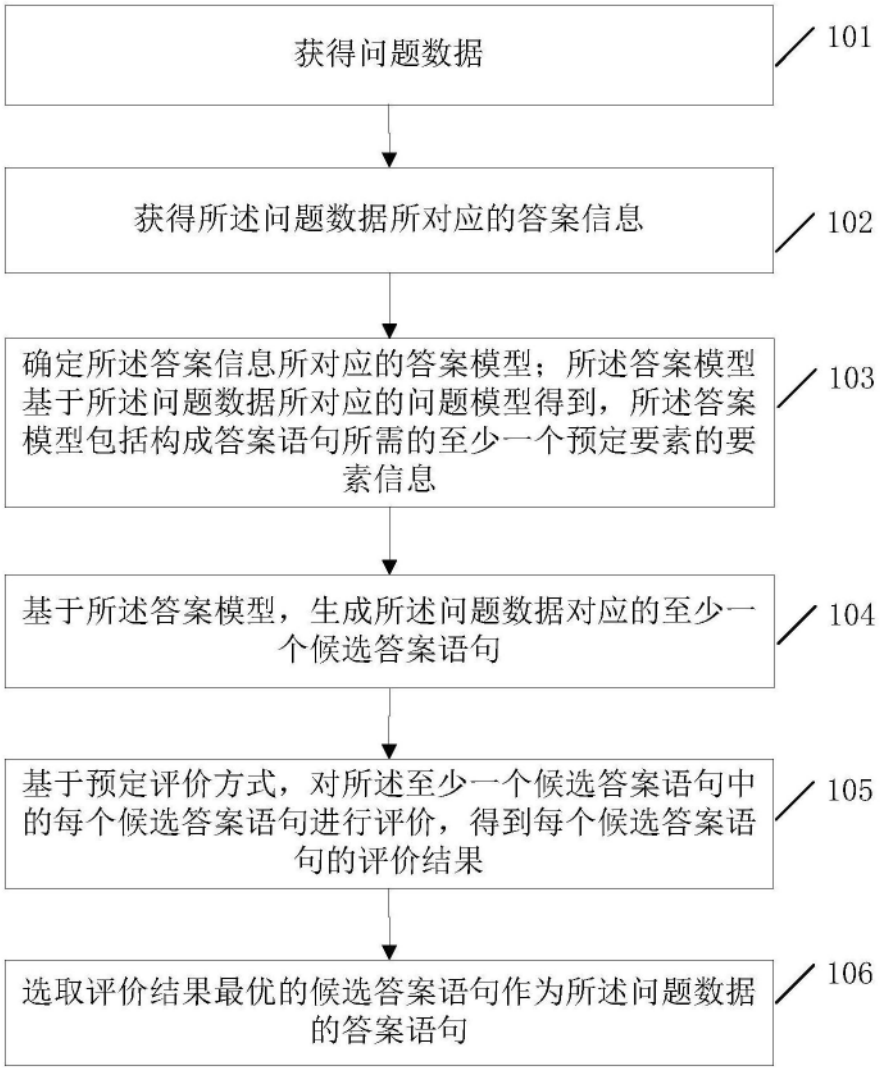

获得问题数据
101
获得所述问题数据所对应的答案信息
102
确定所述答案信息所对应的答案模型；所述答案模型基于所述问题数据所对应的问题模型得到，所述答案模型包括构成答案语句所需的至少一个预定要素的要素信息
103
基于所述答案模型，生成所述问题数据对应的至少一个候选答案语句
104
基于预定评价方式，对所述至少一个候选答案语句中的每个候选答案语句进行评价，得到每个候选答案语句的评价结果
105
选取评价结果最优的候选答案语句作为所述问题数据的答案语句
106

图8

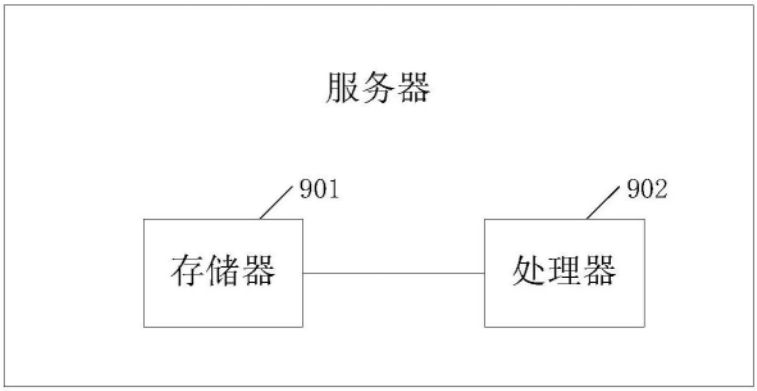

服务器
901
902
存储器
处理器

图9

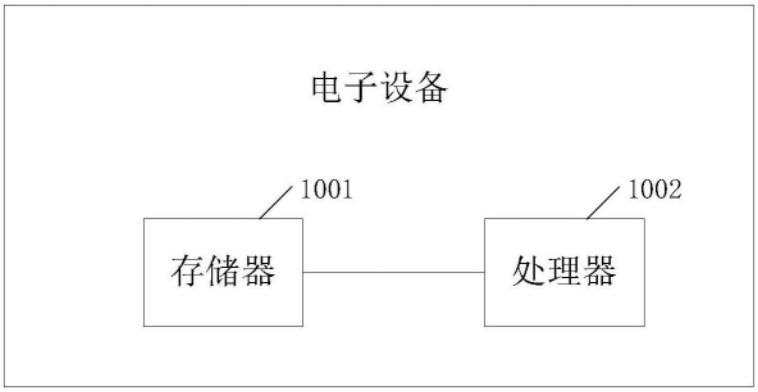


图10