



(12) 发明专利

(10) 授权公告号 CN 110750626 B

(45) 授权公告日 2022.05.06

(21) 申请号 201810739259.6

G06F 40/289 (2020.01)

(22) 申请日 2018.07.06

(56) 对比文件

(65) 同一申请的已公布的文献号  
申请公布号 CN 110750626 A

CN 104050256 A, 2014.09.17

CN 106056207 A, 2016.10.26

CN 108197191 A, 2018.06.22

(43) 申请公布日 2020.02.04

CN 106970907 A, 2017.07.21

(73) 专利权人 中国移动通信有限公司研究院  
地址 100032 北京市西城区金融大街29号  
19层

CN 105512228 A, 2016.04.20

CN 107015962 A, 2017.08.04

CN 108009287 A, 2018.05.08

专利权人 中国移动通信集团有限公司

CN 106383853 A, 2017.02.08

US 2017097618 A1, 2017.04.06

(72) 发明人 张大朋 冯俊兰 黄毅

US 2016110921 A1, 2016.04.21

(74) 专利代理机构 北京同达信恒知识产权代理  
有限公司 11291

唐静笑等. 用户查询意图的层次化识别方法. 《现代图书情报技术》. 2014, 第36-42页.

专利代理师 郭润湘

审查员 唐文俊

(51) Int. Cl.

G06F 16/332 (2019.01)

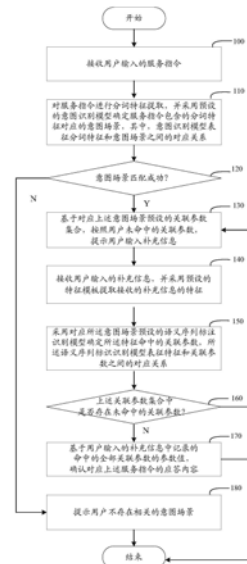
权利要求书3页 说明书12页 附图2页

(54) 发明名称

一种基于场景的任务驱动的多轮对话方法及系统

(57) 摘要

本发明涉及本发明涉及计算机技术,特别涉及一种基于场景的任务驱动的多轮对话方法及系统。用以提高服务系统的应答内容的准确性,以及提升系统处理效率。该方法为:便将服务指令进行了场景化处理,从而可以按照具体的意图场景,以对话方式提示用户不断输入服务指令的补充信息,直到明确用户的最终目的,并生成最终的应答内容,显然,采用这种多轮对话方式可以有效避免用户反复调整服务指令,并缩短了应答内容的确认时长,同时也提高了应答内容的确认精确性,进而有效提高了服务系统的处理效率以及服务质量。



1. 一种基于场景的任务驱动的多轮对话方法,其特征在于,包括:

对用户输入的服务指令进行分词特征提取,并采用预设的意图识别模型确定所述服务指令包含的分词特征对应的意图场景,所述意图识别模型表征分词特征和意图场景之间的对应关系;

基于对应所述意图场景设置的关联参数集合,逐步提示用户输入补充信息,直到确定用户输入的补充信息命中所述关联参数集合中全部关联参数为止;其中,每接收到一次用户输入的补充信息,采用预设的特征模板提取所述补充信息的特征,并采用对应所述意图场景预设的语义序列标注识别模型,根据输出结果判定所述特征体现了所述关联参数集合中的关联参数的参数值时,确定所述特征命中的关联参数,所述语义序列标注识别模型表征特征和关联参数的之间的对应关系;

基于用户输入的补充信息中记录的命中的全部关联参数的参数值,确认对应所述服务指令的应答内容。

2. 如权利要求1所述的方法,其特征在于,在预处理阶段,进一步执行以下操作:

对第一样本数据集合进行分词特征提取,并确定各个分词特征对应的预设的意图场景;

针对所述第一样本数据集合中各个样本数据包含的各个分词特征进行词频-逆文档词频统计,获得统计结果;

基于所述统计结果,去除停用词,参考各个分词特征与意图场景之间的对应关系,利用预设的监督学习模型机进行训练,获得相应的意图识别模型。

3. 如权利要求1所述的方法,其特征在于,在预处理阶段,针对预设的各个意图场景分别进一步执行以下操作:

对一个意图场景对应的第二样本数据集合中各个样本数据进行分词标注、词性标注和关联参数标注,获得标注结果;

基于所述标注结果,确定所述各个样本数据包含的各个分词特征对应的关联参数集合和各个关联参数的参数值;

获取对应所述第二样本数据集合中各个样本数据预设的特征模板,所述特征模板表征了样本数据包含的各个分词的独立特征以及各个分词之间的关联特征;

基于获得的各个样本数据包含的各个分词特征对应的关联参数集合和各个关联参数的参数值,以及各个样本数据对应的特征模板,使用预设的概率模型进行训练,获得所述一个意图场景对应的语义序列标注识别模型。

4. 如权利要求1所述的方法,其特征在于,采用预设的特征模板提取所述补充信息的特征,包括:

采用预设的特征模型对所述补充信息进行分词处理、词性标注,并确定各个分词的独立特征和各个分词之间的关联特征。

5. 如权利要求1-4任一项所述的方法,其特征在于,进一步包括:

在所述意图场景的处理过程中,或者,在所述意图场景结束之后,若基于用户输入的补充信息识别出新的意图场景,则基于对应所述新的意图场景设置的关联参数集合,逐步提示用户继续输入新的补充信息,直到确定用户输入的新的补充信息命中所述新的意图场景对应的关联参数集合中全部关联参数为止,再基于用户输入的新的补充信息记录的命中的

全部关联参数的参数值,确认新的应答内容。

6. 一种基于场景的任务驱动的多轮对话系统,其特征在于,包括:

第一处理单元,用于对用户输入的服务指令进行分词特征提取,并采用预设的意图识别模型确定所述服务指令包含的分词特征对应的意图场景,所述意图识别模型表征分词特征和意图场景之间的对应关系;

第二处理单元,用于基于对应所述意图场景设置的关联参数集合,逐步提示用户输入补充信息,直到确定用户输入的补充信息命中所述关联参数集合中全部关联参数为止;其中,每接收到一次用户输入的补充信息,采用预设的特征模板提取所述补充信息的特征,并采用对应所述意图场景预设的语义序列标注识别模型,根据输出结果判定所述特征体现了所述关联参数集合中的关联参数的参数值时,确定所述特征命中的关联参数,所述语义序列标注识别模型表征特征和关联参数的之间的对应关系;

应答单元,用于基于用户输入的补充信息中记录的命中的全部关联参数的参数值,确认对应所述服务指令的应答内容。

7. 如权利要求6所述的系统,其特征在于,在预处理阶段,所述第一处理单元进一步用于:

对第一样本数据集合进行分词特征提取,并确定各个分词特征对应的预设的意图场景;

针对所述第一样本数据集合中各个样本数据包含的各个分词特征进行词频-逆文档词频统计,获得统计结果;

基于所述统计结果,去除停用词,参考各个分词特征与意图场景之间的对应关系,利用预设的机器学习模型机进行训练,获得相应的意图识别模型。

8. 如权利要求6所述的系统,其特征在于,在预处理阶段,所述第二处理单元进一步用于:针对预设的各个意图场景分别执行以下操作:

对一个意图场景对应的第二样本数据集合中各个样本数据进行分词标注、词性标注和关联参数标注,获得标注结果;

基于所述标注结果,确定所述各个样本数据包含的各个分词特征对应的关联参数集合和各个关联参数的参数值;

获取对应所述第二样本数据集合中各个样本数据预设的特征模板,所述特征模板表征了样本数据包含的各个分词的独立特征以及各个分词之间的关联特征;

基于获得的各个样本数据包含的各个分词特征对应的关联参数集合和各个关联参数的参数值,以及各个样本数据对应的特征模板,使用预设的概率模型进行训练,获得所述一个意图场景对应的语义序列标注识别模型。

9. 如权利要求6所述的系统,其特征在于,采用预设的特征模板提取所述补充信息的特征时,所述第二处理单元用于:

采用预设的特征模型对所述补充信息进行分词处理、词性标注,并确定各个分词的独立特征和各个分词之间的关联特征。

10. 如权利要求6-9任一项所述的系统,其特征在于,所述第一处理单元进一步用于:

在所述意图场景的处理过程中,或者,在所述意图场景结束之后,基于用户输入的补充信息识别出新的意图场景;

所述第二处理单元进一步用于：

基于对应所述新的意图场景设置的关联参数集合，逐步提示用户继续输入新的补充信息，直到确定用户输入的新的补充信息命中所述新的意图场景对应的关联参数集合中全部关联参数为止；

所述应答单元进一步用于：

基于用户输入的新的补充信息记录的命中的全部关联参数的参数值，确认新的应答内容。

11. 一种存储介质，其特征在于，存储有用于实现基于场景的任务驱动的多轮对话方法的程序，所述程序被处理器运行时，执行以下步骤：

对用户输入的服务指令进行分词特征提取，并采用预设的意图识别模型确定所述服务指令包含的分词特征对应的意图场景，所述意图识别模型表征分词特征和意图场景之间的对应关系；

基于对应所述意图场景设置的关联参数集合，逐步提示用户输入补充信息，直到确定用户输入的补充信息命中所述关联参数集合中全部关联参数为止；其中，每接收到一次用户输入的补充信息，采用预设的特征模板提取所述补充信息的特征，并采用对应所述意图场景预设的语义序列标注识别模型确定所述特征命中的关联参数，所述语义序列标注识别模型表征特征和关联参数的之间的对应关系；

基于用户输入的补充信息中记录的命中的全部关联参数的参数值，确认对应所述服务指令的应答内容。

12. 一种通信装置，其特征在于，包括一个或多个处理器；以及

一个或多个计算机可读介质，所述可读介质上存储有指令，所述指令被所述一个或多个处理器执行时，使得所述装置执行如权利要求1至5中任一项所述的方法。

## 一种基于场景的任务驱动的多轮对话方法及系统

### 技术领域

[0001] 本发明涉及计算机技术,特别涉及一种基于场景的任务驱动的多轮对话方法及系统。

### 背景技术

[0002] 目前,随着智能技术的迅速发展,各类服务行业也逐渐引入智能技术也提高服务效率和服务质量。

[0003] 现有技术下,用户在使用服务系统时,当用户输入服务指令后,服务系统通常会直接基于服务指令包含的关键词来确认应答内容,并触发相应的服务。

[0004] 然而,大多数情况下,用户输入的服务指令往往并不精确,只是模糊地反应了用户的使用需求,因此,服务系统基于服务指令所确定的应答内容也不精确,这样,用户便需要反复调整并重新输入服务指令,直至服务系统确定正确的应答内容为止。

[0005] 显然,现有的技术方案会耗用户大量的时间用于调整服务指令,同时,服务系统反复确定应答内容也会在一定程度上造成系统资源的浪费,同时也降低了系统处理效率。

[0006] 有鉴于此,需要重新设计一种技术方案以克服上述缺陷。

### 发明内容

[0007] 本发明实施例提供一种基于场景的任务驱动的多轮对话方法及系统,用以提高服务系统的应答内容的准确性,以及提升系统处理效率。

[0008] 本发明实施例提供的具体技术方案如下:

[0009] 一种基于场景的任务驱动的多轮对话方法,包括:

[0010] 对用户输入的服务指令进行分词特征提取,并采用预设的意图识别模型确定所述服务指令包含的分词特征对应的意图场景,所述意图识别模型表征分词特征和意图场景之间的对应关系;

[0011] 基于对应所述意图场景设置的关联参数集合,逐步提示用户输入补充信息,直到确定用户输入的补充信息命中所述关联参数集合中全部关联参数为止;其中,每接收到一次用户输入的补充信息,采用预设的特征模板提取所述补充信息的特征,并采用对应所述意图场景预设的语义序列标注识别模型确定所述特征命中的关联参数,所述语义序列标注识别模型表征特征和关联参数的之间的对应关系;

[0012] 基于用户输入的补充信息中记录的命中的全部关联参数的参数值,确认对应所述服务指令的应答内容。

[0013] 可选的,在预处理阶段,进一步执行以下操作:

[0014] 对第一样本数据集合进行分词特征提取,并确定各个分词特征对应的预设的意图场景;

[0015] 针对所述第一样本数据集合中各个样本数据包含的各个分词特征进行词频-逆文档词频统计,获得统计结果;

[0016] 基于所述统计结果,去除停用词,参考各个分词特征与意图场景之间的对应关系,利用预设的机器学习模型机进行训练,获得相应的意图识别模型。

[0017] 可选的,在预处理阶段,针对预设的各个意图场景分别进一步执行以下操作:

[0018] 对一个意图场景对应的第二样本数据集合中各个样本数据进行分词标注、词性标注和关联参数标注,获得标注结果;

[0019] 基于所述标注结果,确定所述各个样本数据包含的各个分词特征对应的关联参数集合和各个关联参数的参数值;

[0020] 获取对应所述第二样本数据集合中各个样本数据预设的特征模板,所述特征模板表征了样本数据包含的各个分词特征的独立特征以及各个分词特征之间的关联特征;

[0021] 基于获得的各个样本数据包含的各个分词特征对应的关联参数集合和各个关联参数的参数值,以及各个样本数据对应的特征模板,使用预设的概率模型进行训练,获得所述一个意图场景对应的语义序列标注识别模型。

[0022] 可选的,采用预设的特征模板提取所述补充信息的特征,包括:

[0023] 采用预设的特征模型对所述补充信息进行分词处理、词性标注,并确定各个分词的独立特征和各个分词之间的关联特征。

[0024] 可选的,采用对应所述意图场景预设的语义序列标注识别模型确定所述特征命中的关联参数,包括:

[0025] 将所述特征输入所述对应所述意图场景预设的语义序列标注识别模型,根据输出结果判定所述特征体现了所述关联参数集合中的关联参数的参数值时,确定所述特征命中所述关联参数。

[0026] 可选的,进一步包括:

[0027] 在所述意图场景的处理过程中,或者,在所述意图场景结束之后,若基于用户输入的补充信息识别出新的意图场景,则基于对应所述新的意图场景设置的关联参数集合,逐步提示用户继续输入新的补充信息,直到确定用户输入的新的补充信息命中所述新的意图场景对应的关联参数集合中全部关联参数为止,再基于用户输入的新的补充信息记录的命中的全部关联参数的参数值,确认新的应答内容。

[0028] 一种基于场景的任务驱动的多轮对话系统,包括:

[0029] 第一处理单元,用于对用户输入的服务指令进行分词特征提取,并采用预设的意图识别模型确定所述服务指令包含的分词特征对应的意图场景,所述意图识别模型表征分词特征和意图场景之间的对应关系;

[0030] 第二处理单元,用于基于对应所述意图场景设置的关联参数集合,逐步提示用户输入补充信息,直到确定用户输入的补充信息命中所述关联参数集合中全部关联参数为止;其中,每接收到一次用户输入的补充信息,采用预设的特征模板提取所述补充信息的特征,并采用对应所述意图场景预设的语义序列标注识别模型确定所述特征命中的关联参数,所述语义序列标识识别模型表征特征和关联参数的之间的对应关系;

[0031] 应答单元,用于基于用户输入的补充信息中记录的命中的全部关联参数的参数值,确认对应所述服务指令的应答内容。

[0032] 可选的,在预处理阶段,所述第一处理单元进一步用于:

[0033] 对第一样本数据集合进行分词特征提取,并确定各个分词特征对应的预设的意图

场景；

[0034] 针对所述第一样本数据集合中各个样本数据包含的各个分词特征进行词频-逆文档词频统计,获得统计结果；

[0035] 基于所述统计结果,去除停用词,参考各个分词特征与意图场景之间的对应关系,利用预设的机器学习模型机进行训练,获得相应的意图识别模型。

[0036] 可选的,在预处理阶段,所述第二处理单元进一步用于:针对预设的各个意图场景分别执行以下操作:

[0037] 对一个意图场景对应的第二样本数据集合中各个样本数据进行分词标注、词性标注和关联参数标注,获得标注结果；

[0038] 基于所述标注结果,确定所述各个样本数据包含的各个分词特征对应的关联参数集合和各个关联参数的参数值；

[0039] 获取对应所述第二样本数据集合中各个样本数据预设的特征模板,所述特征模板表征了样本数据包含的各个分词的独立特征以及各个分词之间的关联特征；

[0040] 基于获得的各个样本数据包含的各个分词特征对应的关联参数集合和各个关联参数的参数值,以及各个样本数据对应的特征模板,使用预设的概率模型进行训练,获得所述一个意图场景对应的语义序列标注识别模型。

[0041] 可选的,采用预设的特征模板提取所述补充信息的特征时,所述第二处理单元用于:

[0042] 采用预设的特征模型对所述补充信息进行分词处理、词性标注,并确定各个分词的独立特征和各个分词之间的关联特征。

[0043] 可选的,采用对应所述意图场景预设的语义序列标注识别模型确定所述特征命中的关联参数时,所述第二处理单元用于:

[0044] 将所述特征输入所述对应所述意图场景预设的语义序列标注识别模型,根据输出结果判定所述特征体现了所述关联参数集合中的关联参数的参数值时,确定所述特征命中所述关联参数。

[0045] 可选的,所述第一处理单元进一步用于:

[0046] 在所述意图场景的处理过程中,或者,在所述意图场景结束之后,基于用户输入的补充信息识别出新的意图场景；

[0047] 所述第二处理单元进一步用于:

[0048] 基于对应所述新的意图场景设置的关联参数集合,逐步提示用户继续输入新的补充信息,直到确定用户输入的新的补充信息命中所述新的意图场景对应的关联参数集合中全部关联参数为止；

[0049] 所述生成单元进一步用于:

[0050] 基于用户输入的新的补充信息记录的命中的全部关联参数的参数值,确认新的应答内容。

[0051] 一种存储介质,存储有用于实现基于场景的任务驱动的多轮对话方法的程序,所述程序被处理器运行时,执行以下步骤:

[0052] 对用户输入的服务指令进行分词特征提取,并采用预设的意图识别模型确定所述服务指令包含的分词特征对应的意图场景,所述意图识别模型表征分词特征和意图场景之

间的对应关系；

[0053] 基于对应所述意图场景设置的关联参数集合,逐步提示用户输入补充信息,直到确定用户输入的补充信息命中所述关联参数集合中全部关联参数为止;其中,每接收到一次用户输入的补充信息,采用预设的特征模板提取所述补充信息的特征,并采用对应所述意图场景预设的语义序列标注识别模型确定所述特征命中的关联参数,所述语义序列标注识别模型表征特征和关联参数的之间的对应关系;

[0054] 基于用户输入的补充信息中记录的命中的全部关联参数的参数值,确认对应所述服务指令的应答内容。

[0055] 一种通信装置,包括一个或多个处理器;以及

[0056] 一个或多个计算机可读介质,所述可读介质上存储有指令,所述指令被所述一个或多个处理器执行时,使得所述装置执行上述任一项所述的方法。

[0057] 本发明实施例中,识别出用户的服务指令对应的意图场景,并按照对应所述意图场景设置的关联参数集合,逐步提示用户输入补充信息,直到确定用户输入的补充信息命中所述关联参数集合中全部关联参数为止,最后,再基于用户输入的补充信息中记录的命中的全部关联参数的参数值,确认对应所述服务指令的应答内容。这样,便将服务指令进行了场景化处理,从而可以按照具体的意图场景,以对话方式提示用户不断输入服务指令的补充信息,直到明确用户的最终目的,并生成最终的应答内容,显然,采用这种多轮对话方式可以有效避免用户反复调整服务指令,并缩短了应答内容的确认时长,同时也提高了应答内容的确认精确性,进而有效提高了服务系统的处理效率以及服务质量。

## 附图说明

[0058] 图1为本发明实施例中应答内容确认流程示意图;

[0059] 图2为本发明实施例中服务系统功能结构示意图。

## 具体实施方式

[0060] 为了提高服务系统的应答内容的准确性,以及提升系统处理效率,本发明实施例中,对用户输入的服务指令进行意图分析,确定意图场景,然后,对服务指令进行分词特征提取,再采用对应上述意图场景下预设的语义序列标注识别模型,对服务指令包含的分词特征进行解析,确定服务指令在上述意图场景对应的关联参数集合中未命中的关联参数,并按照所述未命中的关联参数,逐步向用户进行提示,直到确定用户输入的补充信息命中所述关联参数集合中全部关联参数为止,最后,基于命中的全部关联参数,确认对应所述服务指定的应答内容。

[0061] 下面结合附图对本发明优选的实施方式作出进一步详细说明。

[0062] 本发明实施例中,使用Flask网络应用框架,采用自然语言处理工具(如,jieba)对样本数据进行分词特征提取以及词性标注,从而分析出每一个样本数据包含的各个分词特征,然后,基于用于意图识别的第一样本数据集合进行意图识别模型训练,以及基于用于序列标注的第二样本数据集合进行语义序列标注识别模型训练,这样,在后续流程中,可以采用意图识别模型训练来识别用户的服务指令对应的意图场景,采用语义序列标注识别模型来识别用户输入的信息所命中的关联参数。



[0063] 具体的,本发明实施例中,在预处理阶段,服务系统会对知识库进行整理,将各个场景按照意图进行分类,划分出若干意图场景,并对对应意图场景设置相关的描述信息。

[0064] 可选的,描述信息可以包括意图场景名称、关联参数集合(如,可以采用参数列表形式记录)、应答模板;其中,

[0065] 关联参数集合中可以至少包括参数名、是否必要、提示语三个组成部分;用于与待处理序列中进行比较,确定待处理序列标注命中的关联参数(如,用户输入的补充信息所表征的特征,体现了某一意图场景对应的关联参数集合中某一关联参数的参数值,即为“命中”)。

[0066] 应答模板,用于在最终向用户反馈最终确定的应答内容。

[0067] 例如,假设一意图场景名称为:“运输货物”,在这一意图场景下,最终的应答模板为:“您的货物将从{{parameters[“从”]}}运到{{parameters[“到”]}}”,其中,“从”、“到”就是参数名,均设置为“必要”;而“从”对应的提示语为“您的始发地是哪里?”;以及“到”对应的提示语为“您的目的地是哪里?”。

[0068] 又例如,假设又一意图场景名称为:“预约转人工”,在这一意图场景下,

[0069] 最终的应答模板为“您预留的电话号码是{{parameters[“电话号码”]}}”,咨询的业务是{{parameters[“业务”]}}”,稍后将客服人员主动联系您,请保持电话的畅通。”其中,“电话号码”,“业务”就是参数名,均设置为必要;而“电话号码”对应的提示语为“请您留下电话号码”,“业务”对应的提示语为“请留下您要咨询的问题”。

[0070] 进一步地,在预处理阶段,需要收集样本数据进行模型训练,较佳的,可以分为用于进行意图识别训练的样本数据集合(以下简称为集合1)和用于语义序列标注识别训练的样本数据集合(以下简称为集合2)。

[0071] 第一方面,可以先对集合1进行分词特征提取,并确定各个分词特征对应的预设的意图场景。由于是样本数据,因此,集合1中包含的各个分词特征所对应的意图场景是已知的,因而,经过分词特征提取的集合1可以用于进行意图识别模型训练,

[0072] 接着,可以将集合1中各个语名样本数据包含的各个分词特征进行词频-逆文档(tf-idf)词频统计,获得统计结果,再基于所述统计结果去除停用词,以及参考各个分词特征与意图场景之间的对应关系,利用预设的监督学习模型(如,scikit-learn中的线性支持向量机、线性核函数支持向量机等等)进行训练,获得相应的意图识别模型,并进行保存。

[0073] 在训练过程,可以利用一些自动优化工具进行自动寻参,自动训练。

[0074] 在预测识别时,也可以设置置信度,来做进一步分析意图场景是否属于采用意图识别模型识别出的类。

[0075] 第二方面,可以再针对集合2中所有的样本数据进行分词特征提取以及词性识别,然后针对已经进行了分词特征提取、词性识别的样本数据中的待回答内容,进一步采用开始包含不包含(Begin,Inside,Outside,BIO)格式进行标注,其中,标注的基本格式为:分词标注,词性标注,BIO特征标注(即关联参数标注),标注结果存入MongoDB数据库中。当然,采用BIO格式进行标注仅为本发明实施例中的一种举例,实际应用中还可以采用其他方式进行标注,在此不再赘述。

[0076] 例如,假设“运输货物”这一意图场景对应的一个样本数据为:您的货物从北京市西城区运到上海市港口;则BIO特征标注示例如下:

[0077] [“您”,“r”,“0”][“的”,“uj”,“0”],[“货物”,“n”,“0”],[“将”,“d”,“0”],[“从”,“p”,“0”],[“北京市”,“ns”,“B-从”],[“西城区”,“ns”,“I-从”],[“运到”,“v”,“0”],[“上海市”,“ns”,“B-到”],[“港口”,“n”,“I-到”]]。

[0078] 其中,以[“北京市”,“ns”,“B-从”]为例,“北京市”是分词,“ns”是词性,“B-从”是关联参数。

[0079] 又例如:假设“预约转人工”这一意图场景对应的一个样本数据为:您预留的电话号码是12312341234,咨询的业务为话费,稍后将有客服人员主动联系您,则BIO特征标注示例如下:

[0080] [‘您’,‘zg’,‘0’],[‘预留’,‘v’,‘0’],[‘的’,‘uj’,‘0’],[‘电话号码’,‘n’,‘0’],[‘是’,‘p’,‘0’],[‘12312341234’,‘m’,‘B-电话号码’],[‘,’,’x’,‘0’],[‘咨询’,‘vn’,‘0’],[‘的’,‘uj’,‘0’],[‘业务’,‘n’,‘0’],[‘为’,‘p’,‘0’],[‘话费’,‘n’,‘B-业务’],[‘,’,’x’,‘0’],[‘稍后’,‘d’,‘0’],[‘将’,‘d’,‘0’],[‘有’,‘v’,‘0’],[‘客服’,‘n’,‘0’],[‘人员’,‘n’,‘0’],[‘主动’,‘b’,‘0’],[‘联系’,‘n’,‘0’],[‘您’,‘zg’,‘0’],[‘,’,’x’,‘0’],[‘请’,‘v’,‘0’],[‘保持’,‘v’,‘0’],[‘电话’,‘n’,‘0’],[‘的’,‘uj’,‘0’],[‘畅通’,‘a’,‘0’],[‘.’,’x’,‘0’]]。

[0081] 其中,以[‘话费’,‘n’,‘B-业务’]为例,“话费”是分词,“n”是词性,“B-业务”是BIO特征。

[0082] 经过标注的集合2可以用于进行语义序列标注识别模型训练。

[0083] 在进行语义序列标注识别模型训练时,是针对不同的意图场景分别进行的。

[0084] 以意图场景x为例,获得意图场景x在集合2中对应的经过标注的样本数据集合(以下简称集合X)后,

[0085] 具体为:先设计每个样本数据的特征模板,特征模板表征了样本数据包含的各个分词特征的独立特征以及各个分词特征之间的关联特征。

[0086] 然后基于各个样本数据的标注结果(即分词标注、词性标注、关联参数标注),基于所述标注结果,可以确定所述各个样本数据包含的各个分词特征对应的关联参数集合,其中,关联参数集合中进一步也包括了各个关联参数的参数值;

[0087] 接着,获取对应所述第二样本数据集合中各个样本数据预设的特征模板,所述特征模板表征了样本数据包含的各个分词特征的独立特征以及各个分词特征之间的关联特征;

[0088] 最后,基于获得的各个样本数据包含的各个分词特征对应的关联参数集合以及各个关联参数的参数值,以及各个样本数据对应的特征模板,使用预设的概率模型[如,条件随机声(CRF)、基于深度表示的条件随机场(如Bi-LSTM-CRF)等等]来进行训练,获得意图场景x对应的语义序列标注识别模型,并进行保存。

[0089] 可选的,特征模板的设计遵守如下规则:

[0090] 第一:首词主要记录以下维度的信息:当前词、当前词是否为字符、当前词是否为数字、当前词的词性、首词(begining of a sentence,BOS)、后一词、后一词是否为字符、后一词是否为数字、后一词的词性。

[0091] 第二:中间词主要记录以下维度的信息,当前词、当前词是否是字符、当前是否是数字、当前词的词性、前一词、前一词是否为字符、前一词是否为数字、前一词的词性、后一

词、后一词是否为字符、后一词是否为数字、后一词的词性。

[0092] 第三:尾词主要记录以下维度的信息:当前词、当前词是否为字符、当前词是否为数字、当前词的词性、前一词、前一词是否为字符、前一词是否为数字、前一词的词性、尾词(ending of a sentence,EOS)。

[0093] 以此样本数据为例:

[0094] [[“我”,“r”,“0”],[“的”,“uj”,“0”],[“电话”,“n”,“0”],[“是”,“v”,“0”],[“12347856589”,“m”,“B-电话号码”],[“”,“x”,“0”],[“我”,“r”,“0”],[“咨询”,“vn”,“0”],[“的”,“uj”,“0”],[“业务”,“n”,“0”],[“是”,“v”,“0”],[“话费”,“n”,“B-业务”]]],

[0095] 得到特征模板为:

[0096] [['word=我', 'word.isalpha=True', 'word.isdigit=False', 'postag=r', 'BOS', '+1:word=的', '+1:word.isalpha=True', '+1:word.isdigit=False', '+1:postag=uj'], ['word=的', 'word.isalpha=True', 'word.isdigit=False', 'postag=uj', '-1:word=我', '-1:word.isalpha=True', '-1:word.isdigit=False', '-1:postag=r', '+1:word=电话', '+1:word.isalpha=True', '+1:word.isdigit=False', '+1:postag=n'], …………… ['word=是', 'word.isalpha=True', 'word.isdigit=False', 'postag=v', '-1:word=业务', '-1:word.isalpha=True', '-1:word.isdigit=False', '-1:postag=n', '+1:word=话费', '+1:word.isalpha=True', '+1:word.isdigit=False', '+1:postag=n'], ['word=话费', 'word.isalpha=True', 'word.isdigit=False', 'postag=n', '-1:word=是', '-1:word.isalpha=True', '-1:word.isdigit=False', '-1:postag=v', 'EOS']]。

[0097] 简化为类标签数据(即特征)记录为:['0', '0', '0', '0', 'B-电话号码', '0', '0', '0', '0', '0', '0', 'B-业务']。

[0098] 将集合x中的所有样本数据均通过相应的特征模板提取上述形式的类标签数据,然后汇总送入pycrfsuite.Trainer函数进行训练,获得最终的意图场景x对应的语义序列标注识别模型。

[0099] 将各个意图场景对应的语义序列标注识别模型均训练完毕后进行保存。

[0100] 在后续预测识别时,只需将待识别语句按照特征模板的设置规则特征化后,调用相应的语义序列标注识别模型,即可预测出语句的BIO标注序列,而通过BIO标注序列即可以看出待识别语句命中了哪些关联参数,即体现了哪些关联参数的参数值。

[0101] 在训练好意图识别模型和语义序列标识识别模型后,便可以开始进行实际预测了。具体的,参阅图1所示,本发明实施例中,服务系统确定应答内容的具体流程如下:

[0102] 步骤100:服务系统接收用户输入的服务指令。

[0103] 例如:用户初始输入的服务指令为:我需要包裹邮递。

[0104] 步骤110:服务系统对服务指令进行分词特征提取,并采用预设的意图识别模型确定服务指令包含的分词特征对应的意图场景,其中,意图识别模型表征分词特征和意图场景之间的对应关系。

[0105] 例如,将“我需要包裹邮递”进行分词特征提取后,获得分词“包裹邮递”的分词特征,将获得的分词特征输入意图识别模型后,确定匹配的意图场景为“运输货物”。

[0106] 步骤120:服务系统判断意图场景是否匹配成功?若是,则执行步骤130;否则,执行

步骤180。

[0107] 步骤130:服务系统基于对应上述意图场景预设的关联参数集合,按照用户未命中的关联参数,提示用户输入补充信息。

[0108] 例如:假设对应意图场景“运输货物”预设的关联参数为“始发地”和“目的地”,而在初始阶段,在用户尚未输入任何补充信息时,命中的关联参数为0,此时,服务系统会先获取对应“始发地”设置的提示语“始发地是哪里?”,并将上述提示语呈现给用户,提示用户输入补充信息。

[0109] 步骤140:服务系统接收用户输入的补充信息,采用预设的特征模板提取接收的补充信息的特征。

[0110] 具体的,可以按照前述实施例中提及的预设的特征模板,对用户输入的补充信息进行解析,提取出相应的特征,在此不再赘述。

[0111] 步骤150:服务系统采用对应上述意图场景预设的语义序列标注识别模型确定所述特征命中的关联参数,所述语义序列标识识别模型表征特征和关联参数之间的对应关系。

[0112] 具体的,可以将所述特征输入所述对应所述意图场景预设的语义序列标注识别模型,根据输出结果判定所述特征体现了所述关联参数集合中的关联参数的参数值时,确定所述特征命中所述关联参数。

[0113] 例如:假设用户第一次输入的补充信息为“始发地是北京”,则按照上述特征模板进行特征提取后,经过语义序列标识识别模型进行匹配,从而识别出“北京”是关联参数“始发地”的参数值,则说明用户第一次输入的补充信息命中了关联参数“始发地”。

[0114] 步骤160:服务系统判断上述关联参数集合中是否存在未命中的关联参数?若是,则返回步骤130;否则,执行步骤170。

[0115] 例如:进一步地,当用户第一次输入补充信息后,服务系统判定仍存在未命中的关联参数为“目的地”,则服务系统会获取对应“目的地”设置的提示语“目的地是哪里?”,并将上述提示语呈现给用户,提示用户再次输入补充信息。

[0116] 而假设用户再次输入的补充信息为“目的地是上海”,则按照上述特征模板进行特征提取后,经过语义序列标识识别模型进行匹配,从而识别出“上海”是关联参数“目的地”的参数值,则说明用户再次输入的补充信息命中了关联参数“目的地”。

[0117] 步骤170:服务系统基于用户输入的补充信息中记录的命中的全部关联参数的参数值,确认对应上述服务指令的应答内容。

[0118] 例如,服务系统确认“始发地”和“目的地”两个关联参数均命中后,会基于两个命中的关联参数的参数值“北京”和“上海”,来补充填写对应意图场景“运输货物”预设的应答模板,从而生成最终的应答内容,即“您的货物将从北京运到上海”,服务系统会将上述应答内容反馈给用户,同时结束本次对话。

[0119] 步骤180:服务系统提示用户不存在相关的意图场景,并结束当前流程。

[0120] 基于上述实施例,实际应用中,服务系统还可以在其他意思场景下采用相同方式来确认用户的应答内容。

[0121] 例如:

[0122] 假设用户输入的服务指令为“人工服务”,则服务系统调预先训练的意图识别模型

确认“人工服务”的分词特征与意图场景“转人工咨询”匹配成功。

[0123] 接着,服务系统会获取对应“转人工咨询”预先训练好的语义序列标注识别模型标注来确认命中的关联参数。

[0124] 初始阶段,BIO标识为['0','0','0'],未命中任何关联参数;因此,服务系统按照第一个关联参数“电话号码”向用户呈现提示语“请您留下电话号码”;假设用户输入的补充信息为“我的电话号码是12314785966”,则服务系统会通过语义序列标注识别模型识别出“12314785966”命中了关联参数“电话号码”。

[0125] 进一步地,服务系统再次确认仍存在未命中的关联参数,则服务系统按照第二个关联参数“咨询问题”,向用户呈现提示语:“请留下您要咨询的问题”;

[0126] 假设用户再次输入的补充信息为:咨询流量,则服务器会通过语义序列标注识别模型识别出“流量”命中了关联参数“咨询问题”。

[0127] 最后,服务系统确认所有关联参数均命中后,则会获取对应“转人工咨询”预设的应答模板,将命中的关联参数的参数值填写至应答模板,从而最终确认应答内容:“您预留的电话号码是12314785966,咨询的业务是流量,稍后将有客服人员主动联系您,请保持电话的畅通。”。

[0128] 进一步地,基于上述实施例,假设在“转人工咨询”意图场景处理过程中,或者,在“转人工咨询”意图场景处理结束之后,用户未经提示,继续输入了新的补充信息“我对你们的服务非常满意”,那么,服务系统会识别出相应的新的意图场景为“评价”,并将上述补充信息标注为:[["我","r","0"],["对","p","0"],["你们","r","0"],["的","uj","0"],["服务","vn","0"],["非常","d","B-是"],["满意","v","I-是"]],这里的“是”是必要参数;然而,服务系统采用对应“评价”这一意图场景预设的语义序列标识识别模型,识别出命中的关联参数是“评价是”,对应的应答模板是“您对我们的评价是……”,并将关联参数的参数值“非常满意”填写入应答模板,从而生成最终的应答内容“您对我们的评价是非常满意”。当然,针对“评价”这一意图场景进行处理时,服务系统也可以提示用户按照预设的关联参数集合,不断输入新的补充信息,以完善应答模板,上述过程仅以输入一次新的补充信息为例进行说明,在此不再赘述。

[0129] 由此可见,意图场景之间可以采用连接方式处理,也可以采用嵌套方式处理,简言之,在一个意图场景处理过程中,或者,在一个意图场景处理结束之后,若服务系统基于用户输入的补充信息识别出新的意图场景,则服务系统基于对应所述新的意图场景设置的关联参数集合,逐步提示用户继续输入新的补充信息,直到确定用户输入的新的补充信息命中所述新的意图场景对应的关联参数集合中全部关联参数为止,再基于用户输入的新的补充信息记录的命中的全部关联参数的参数值,确认新的应答内容。

[0130] 当然,意图场景是嵌套出现的,则在处理完新的意图场景后,还需要返回上一个意图场景继续处理,在此不再赘述。

[0131] 基于上述实施例,参阅图2所示,本发明实施例中,提供一种基于场景的任务驱动的多轮对话系统(即,上述服务系统),至少包括第一处理单元21、第二处理单元22和应答单元23,其中,

[0132] 第一处理单元21,用于对用户输入的服务指令进行分词特征提取,并采用预设的意图识别模型确定所述服务指令包含的分词特征对应的意图场景,所述意图识别模型表征

分词特征和意图场景之间的对应关系；

[0133] 第二处理单元22,用于基于对应所述意图场景设置的关联参数集合,逐步提示用户输入补充信息,直到确定用户输入的补充信息命中所述关联参数集合中全部关联参数为止;其中,每接收到一次用户输入的补充信息,采用预设的特征模板提取所述补充信息的特征,并采用对应所述意图场景预设的语义序列标注识别模型确定所述特征命中的关联参数,所述语义序列标识识别模型表征特征和关联参数的之间的对应关系;

[0134] 应答单元23,用于基于用户输入的补充信息中记录的命中的全部关联参数的参数值,确认对应所述服务指令的应答内容。

[0135] 上述第一处理单元21、第二处理单元22和应答单元23,可以各自为独自装置,可以是集成在同一装置中的功能单元,具体视具体应用环境而定,在此不再赘述。

[0136] 可选的,在预处理阶段,所述第一处理单元21进一步用于:

[0137] 对第一样本数据集合进行分词特征提取,并确定各个分词特征对应的预设的意图场景;

[0138] 针对所述第一样本数据集合中各个样本数据包含的各个分词特征进行词频-逆文档词频统计,获得统计结果;

[0139] 基于所述统计结果,去除停用词,参考各个分词特征与意图场景之间的对应关系,利用预设的监督学习模型机进行训练,获得相应的意图识别模型。

[0140] 可选的,在预处理阶段,所述第二处理单元22进一步用于:针对预设的各个意图场景分别执行以下操作:

[0141] 对一个意图场景对应的第二样本数据集合中各个样本数据进行分词标注、词性标注和关联参数标注,获得标注结果;

[0142] 基于所述标注结果,确定所述各个样本数据包含的各个分词特征对应的关联参数集合和各个关联参数的参数值;

[0143] 获取对应所述第二样本数据集合中各个样本数据预设的特征模板,所述特征模板表征了样本数据包含的各个分词的独立特征以及各个分词之间的关联特征;

[0144] 基于获得的各个样本数据包含的各个分词特征对应的关联参数集合和各个关联参数的参数值,以及各个样本数据对应的特征模板,使用预设的概率模型进行训练,获得所述一个意图场景对应的语义序列标注识别模型。

[0145] 可选的,采用预设的特征模板提取所述补充信息的特征时,所述第二处理单元22用于:

[0146] 采用预设的特征模型对所述补充信息进行分词处理、词性标注,并确定各个分词的独立特征和各个分词之间的关联特征。

[0147] 可选的,采用对应所述意图场景预设的语义序列标注识别模型确定所述特征命中的关联参数时,所述第二处理单元22用于:

[0148] 将所述特征输入所述对应所述意图场景预设的语义序列标注识别模型,根据输出结果判定所述特征体现了所述关联参数集合中的关联参数的参数值时,确定所述特征命中所述关联参数。

[0149] 可选的,所述第一处理单元21进一步用于:

[0150] 在所述意图场景的处理过程中,或者,在所述意图场景结束之后,基于用户输入的

补充信息识别出新的意图场景；

[0151] 所述第二处理单元22进一步用于：

[0152] 基于对应所述新的意图场景设置的关联参数集合，逐步提示用户继续输入新的补充信息，直到确定用户输入的新的补充信息命中所述新的意图场景对应的关联参数集合中全部关联参数为止；

[0153] 所述生成单元23进一步用于：

[0154] 基于用户输入的新的补充信息记录的命中的全部关联参数的参数值，确认新的应答内容。

[0155] 一种存储介质，存储有用于实现基于场景的任务驱动的多轮对话方法的程序，所述程序被处理器运行时，执行以下步骤：

[0156] 对用户输入的服务指令进行分词特征提取，并采用预设的意图识别模型确定所述服务指令包含的分词特征对应的意图场景，所述意图识别模型表征分词特征和意图场景之间的对应关系；

[0157] 基于对应所述意图场景设置的关联参数集合，逐步提示用户输入补充信息，直到确定用户输入的补充信息命中所述关联参数集合中全部关联参数为止；其中，每接收到一次用户输入的补充信息，采用预设的特征模板提取所述补充信息的特征，并采用对应所述意图场景预设的语义序列标注识别模型确定所述特征命中的关联参数，所述语义序列标注识别模型表征特征和关联参数的之间的对应关系；

[0158] 基于用户输入的补充信息中记录的命中的全部关联参数的参数值，确认对应所述服务指令的应答内容。

[0159] 一种通信装置，包括一个或多个处理器；以及

[0160] 一个或多个计算机可读介质，所述可读介质上存储有指令，所述指令被所述一个或多个处理器执行时，使得所述装置执行上述任一项所述的方法。

[0161] 综上所述，本发明实施例中，识别出用户的服务指令对应的意图场景，并按照对应所述意图场景设置的关联参数集合，逐步提示用户输入补充信息，直到确定用户输入的补充信息命中所述关联参数集合中全部关联参数为止，最后，再基于用户输入的补充信息中记录的命中的全部关联参数的参数值，确认对应所述服务指令的应答内容。这样，便将服务指令进行了场景化处理，从而可以按照具体的意图场景，以对话方式提示用户不断输入服务指令的补充信息，直到明确用户的最终目的，并生成最终的应答内容，显然，采用这种多轮对话方式可以有效避免用户反复调整服务指令，并缩短了应答内容的确认时长，同时也提高了应答内容的确认精确性，进而有效提高了服务系统的处理效率以及服务质量。

[0162] 本领域内的技术人员应明白，本发明的实施例可提供为方法、系统、或计算机程序产品。因此，本发明可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质（包括但不限于磁盘存储器、CD-ROM、光学存储器等）上实施的计算机程序产品的形式。

[0163] 本发明是参照根据本发明实施例的方法、设备（系统）、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序

指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0164] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制品,该指令装置实现在流程图一个流程或多个流程和这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0165] 尽管已描述了本发明的优选实施例,但本领域内的技术人员一旦得知了基本创造性概念,则可对这些实施例作出另外的变更和修改。所以,所附权利要求意欲解释为包括优选实施例以及落入本发明范围/或方框图一个方框或多个方框中指定的功能。

[0166] 显然,本领域的技术人员可以对本发明实施例进行各种改动和变型而不脱离本发明实施例的精神和范围。这样,倘若本发明实施例的这些修改和变型属于本发明权利要求及其等同技术的范围之内,则本发明也意图包含这些改动和变型在内。



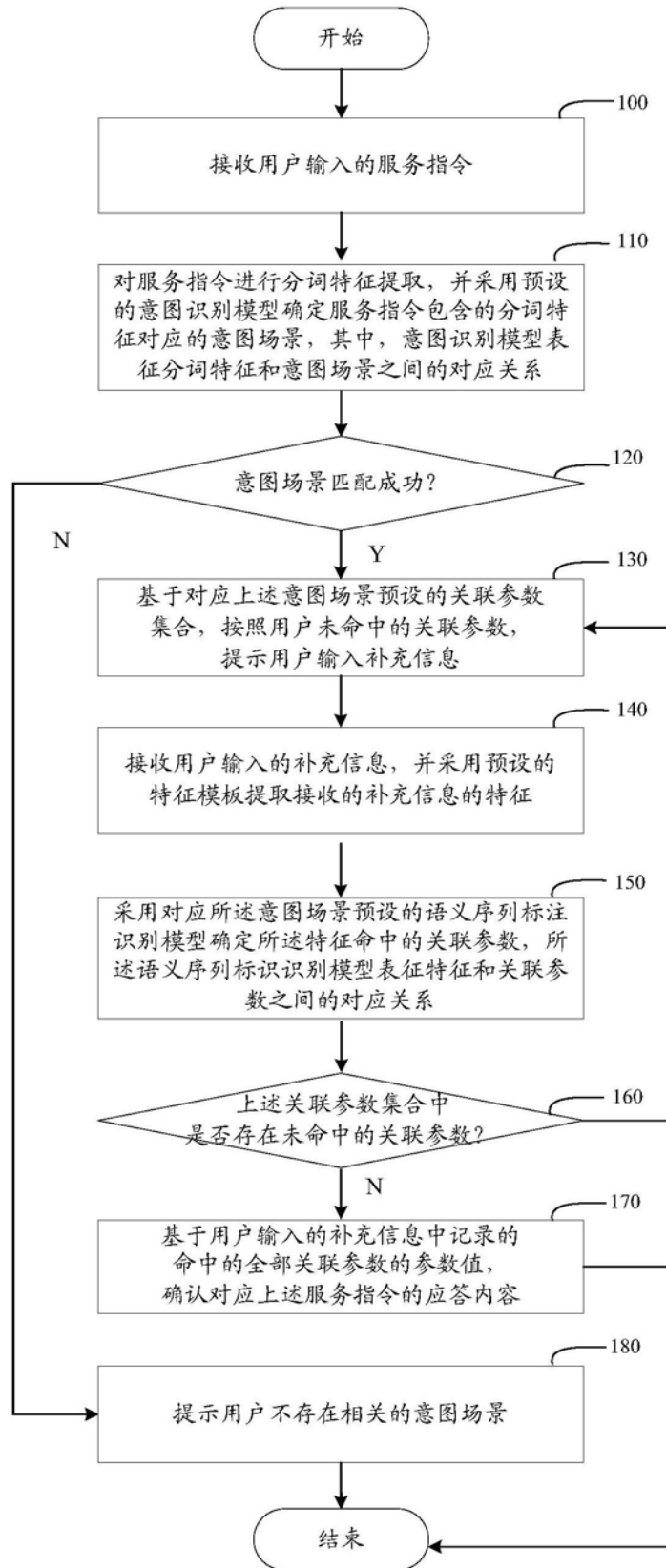


图1

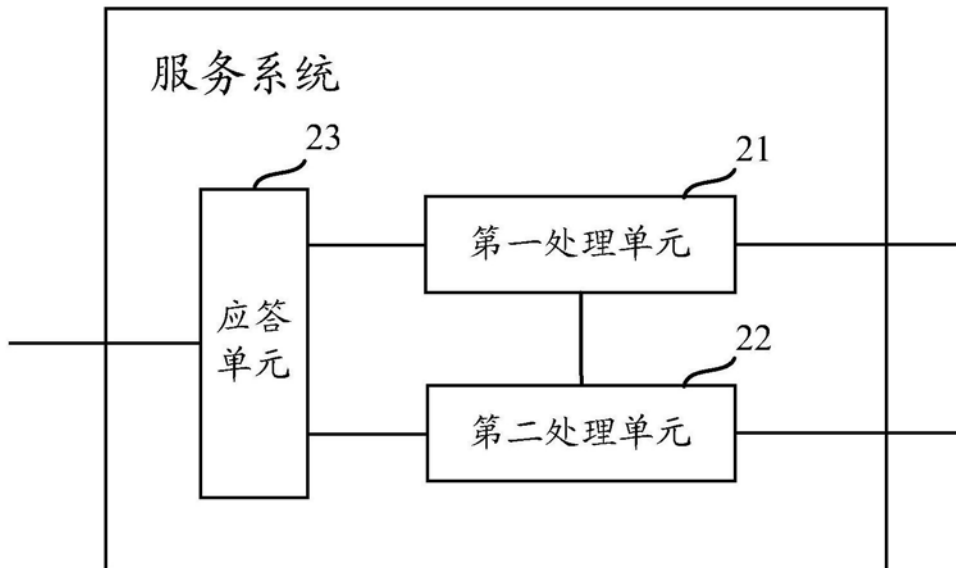


图2