



(12) 发明专利

(10) 授权公告号 CN 114896376 B

(45) 授权公告日 2025. 05. 23

(21) 申请号 202210325699.3

G06F 40/35 (2020.01)

(22) 申请日 2022.03.29

G06F 18/214 (2023.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 114896376 A

(56) 对比文件

CN 109815326 A, 2019.05.28

CN 113590769 A, 2021.11.02

(43) 申请公布日 2022.08.12

审查员 陈锋

(73) 专利权人 京东科技信息技术有限公司

地址 100176 北京市大兴区北京经济技术
开发区科创十一街18号院2号楼6层
601

(72) 发明人 王泽勋 陈蒙 冯明超

(74) 专利代理机构 北京路浩知识产权代理有限
公司 11002

专利代理师 王治东

(51) Int. Cl.

G06F 16/3329 (2025.01)

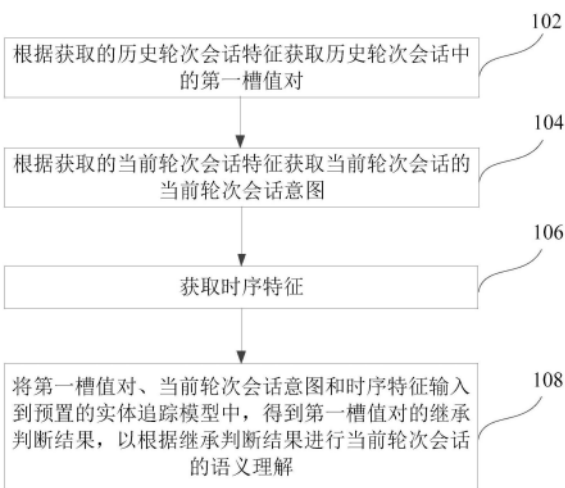
权利要求书2页 说明书9页 附图3页

(54) 发明名称

实体追踪方法、装置、电子设备和存储介质

(57) 摘要

本公开提供一种实体追踪方法、装置、电子设备和存储介质,所述方法包括:根据获取的历史轮次会话特征获取历史轮次会话中的第一槽值对;根据获取的当前轮次会话特征获取当前轮次会话的当前轮次会话意图;获取时序特征,所述时序特征根据所述历史轮次会话和所述当前轮次会话的时间差得到;将所述第一槽值对、所述当前轮次会话意图和所述时序特征输入到预置的实体追踪模型中,得到所述第一槽值对的继承判断结果,以根据所述继承判断结果进行当前轮次会话的语义理解,其中,所述继承判断结果表征所述第一槽值对是否需要被继承到所述当前轮次会话中。本公开的技术方案可以提高实体追踪的准确性。



1. 一种实体追踪方法,其特征在于,所述方法包括:

根据获取的历史轮次会话特征获取历史轮次会话中的第一槽值对;

根据获取的当前轮次会话特征获取当前轮次会话的当前轮次会话意图;

获取时序特征,所述时序特征根据所述历史轮次会话和所述当前轮次会话的时间差得到;

将所述第一槽值对、所述当前轮次会话意图和所述时序特征输入到预置的实体追踪模型中,得到所述第一槽值对的继承判断结果,以根据所述继承判断结果进行当前轮次会话的语义理解,其中,所述继承判断结果表征所述第一槽值对是否需要被继承到所述当前轮次会话中;

所述根据获取的历史轮次会话特征获取历史轮次会话中的第一槽值对之前,所述方法还包括:

对所述当前轮次会话的第二用户文本进行文本编码,将其转化为文本向量,得到所述当前轮次会话特征。

2. 根据权利要求1所述的方法,其特征在于,所述实体追踪模型的训练方法包括:

获取样本会话数据集,所述样本会话数据集中的样本会话数据至少包括两个轮次会话的会话数据,所述样本会话数据包括经过标注的以下训练标签:前轮次会话中的所述第一槽值对、后轮次会话的会话意图,以及所述前轮次会话和所述后轮次会话的时间差;

根据所述样本会话数据集和所述训练标签对初始实体追踪模型进行训练,得到所述实体追踪模型。

3. 根据权利要求1所述的方法,其特征在于,所述获取时序特征,包括:

获取所述历史轮次会话和所述当前轮次会话的时间差;

获取预先确定的时间桶;

根据时间桶获取所述时间差对应的时序特征,所述时序特征表征所述时间差对应的时间桶中的时间区域。

4. 根据权利要求1所述的方法,其特征在于,所述根据获取的历史轮次会话特征获取历史轮次会话中的第一槽值对之前,所述方法还包括:

对所述历史轮次会话的第一用户文本和系统回复文本进行文本编码,得到所述历史轮次会话特征。

5. 根据权利要求4所述的方法,其特征在于,所述根据获取的当前轮次会话特征获取当前轮次会话的当前轮次会话意图,包括:

根据所述当前轮次会话特征得到所述第二用户文本的意图词向量;

采用独热编码方式对意图词向量进行编码,得到所述当前轮次会话意图。

6. 根据权利要求4所述的方法,其特征在于,所述根据获取的历史轮次会话特征获取历史轮次会话中的第一槽值对,包括:

采用预置的词向量模型根据所述历史轮次会话特征获取所述第一用户文本的槽值信息;

根据获取的槽位信息和所述槽值信息生成所述第一槽值对。

7. 根据权利要求1所述的方法,其特征在于,所述得到所述第一槽值对的继承判断结果之后,所述方法还包括:

根据所述继承判断结果、所述历史轮次会话特征和所述当前轮次会话特征获取当前轮次会话的语义理解结果。

8. 一种实体追踪装置,其特征在於,所述装置包括:

槽值对获取单元,用于根据获取的历史轮次会话特征获取历史轮次会话中的第一槽值对;

意图获取单元,用于根据获取的当前轮次会话特征获取当前轮次会话的当前轮次会话意图;

时序获取单元,用于获取时序特征,所述时序特征根据所述历史轮次会话和所述当前轮次会话的时间差得到;

追踪单元,用于将所述第一槽值对、所述当前轮次会话意图和所述时序特征输入到预置的实体追踪模型中,得到所述第一槽值对的继承判断结果,以根据所述继承判断结果进行当前轮次会话的语义理解,其中,所述继承判断结果表征所述第一槽值对是否需要被继承到所述当前轮次会话中;

还包括编码单元,用于对所述当前轮次会话的第二用户文本进行文本编码,将其转化为文本向量,得到所述当前轮次会话特征。

9. 一种电子设备,包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序,其特征在於,所述处理器执行所述程序时实现如权利要求1至7任一项所述的方法的步骤。

10. 一种非暂态计算机可读存储介质,其上存储有计算机程序,其特征在於,所述计算机程序被处理器执行时实现如权利要求1至7任一项所述的方法的步骤。

实体追踪方法、装置、电子设备和存储介质

技术领域

[0001] 本公开涉及计算机技术领域,尤其涉及一种实体追踪方法、装置、电子设备和非暂态计算机可读存储介质。

背景技术

[0002] 当前,对话系统在各个领域越来越引起人们的重视,深度学习技术的不断进步极大地推动着对话系统的发展。

[0003] 在进行多轮对话的对话管理时,需要根据对话历史信息,决定此刻对用户的回复内容。对话状态追踪表示根据历史对话状态以及之前系统的动作,追踪当前对话状态。对话状态追踪中的槽值对信息追踪,是对历史对话信息和当前用户信息进行建模,即通过上下文信息和当前信息判断槽位是否更新。

[0004] 在槽值对信息追踪过程中,一些方案直接将历史信息拼接作为上下文特征,这样会忽略了历史信息的时序信息。另一些方案会考虑话术顺序,将对话的前后关系作为时序信息,以当前对话和开始对话的轮次距离作为时序特征作为参数来进行槽值对信息追踪,但是仍然不能准确地模拟对话真实情况。

发明内容

[0005] 本公开提供一种实体追踪方法、装置、电子设备和非暂态计算机可读存储介质,用以解决现有技术中不能准确模拟对话真实情况的问题,提高实体追踪的准确性。

[0006] 本公开提供一种实体追踪方法,包括:根据获取的历史轮次会话特征获取历史轮次会话中的第一槽值对;根据获取的当前轮次会话特征获取当前轮次会话的当前轮次会话意图;获取时序特征,所述时序特征根据所述历史轮次会话和所述当前轮次会话的时间差得到;将所述第一槽值对、所述当前轮次会话意图和所述时序特征输入到预置的实体追踪模型中,得到所述第一槽值对的继承判断结果,以根据所述继承判断结果进行当前轮次会话的语义理解,其中,所述继承判断结果表征所述第一槽值对是否需要被继承到所述当前轮次会话中。

[0007] 根据本公开提供的一种实体追踪方法,所述实体追踪模型的训练方法包括:获取样本会话数据集,所述样本会话数据集中的样本会话数据至少包括两个轮次会话;所述样本会话数据集中的样本会话数据至少包括两个轮次会话的会话数据,所述样本会话数据包括经过标注的以下训练标签:前轮次会话中的槽值对、后轮次会话的会话意图,以及所述前轮次会话和所述后轮次会话的时间差;根据所述样本会话数据集和所述训练标签对初始实体追踪模型进行训练,得到所述实体追踪模型。

[0008] 根据本公开提供的一种实体追踪方法,所述获取时序特征,包括:获取所述历史轮次会话和所述当前轮次会话的时间差;获取预先确定的时间桶;根据时间桶获取所述时间差对应的时序特征,所述时序特征表征所述时间差对应的时间桶中的时间区域。

[0009] 根据本公开提供的一种实体追踪方法,所述根据获取的历史轮次会话特征获取历

史轮次会话中的第一槽值对之前,所述方法还包括:对所述历史轮次会话的第一用户文本和系统回复文本进行文本编码,得到所述历史轮次会话特征;对所述当前轮次会话的第二用户文本进行文本编码,得到所述当前轮次会话特征。

[0010] 根据本公开提供的一种实体追踪方法,所述根据获取的当前轮次会话特征获取当前轮次会话的当前轮次会话意图,包括:根据所述当前轮次会话特征得到所述第二用户文本的意图词向量;采用独热编码方式对意图词向量进行编码,得到所述当前轮次会话意图。

[0011] 根据本公开提供的一种实体追踪方法,所述根据获取的历史轮次会话特征获取历史轮次会话中的第一槽值对,包括:采用预置的词向量模型根据所述历史轮次会话特征获取所述第一用户文本的槽值信息;根据获取的槽位信息和所述槽值信息生成所述第一槽值对。

[0012] 根据本公开提供的一种实体追踪方法,所述得到所述第一槽值对的继承判断结果之后,所述方法还包括:根据所述继承判断结果、所述历史轮次会话特征和所述当前轮次会话特征获取当前轮次会话的语义理解结果。

[0013] 本公开还提供一种实体追踪装置,包括:槽值对获取单元,用于根据获取的历史轮次会话特征获取历史轮次会话中的第一槽值对;意图获取单元,用于根据获取的当前轮次会话特征获取当前轮次会话的当前轮次会话意图;时序获取单元,用于获取时序特征,所述时序特征根据所述历史轮次会话和所述当前轮次会话的时间差得到;追踪单元,用于将所述第一槽值对、所述当前轮次会话意图和所述时序特征输入到预置的实体追踪模型中,得到所述第一槽值对的继承判断结果,以根据所述继承判断结果进行当前轮次会话的语义理解,其中,所述继承判断结果表征所述第一槽值对是否需要被继承到所述当前轮次会话中。

[0014] 根据本公开提供的一种实体追踪装置,所述装置还包括训练单元,用于对实体追踪模型进行训练,所述训练单元用于:获取样本会话数据集,所述样本会话数据集中的样本会话数据至少包括两个轮次会话的会话数据,所述样本会话数据包括经过标注的以下训练标签:前轮次会话中的所述第一槽值对、后轮次会话的会话意图,以及所述前轮次会话和所述后轮次会话的时间差;根据所述样本会话数据集和所述训练标签对初始实体追踪模型进行训练,得到所述实体追踪模型。

[0015] 根据本公开提供的一种实体追踪装置,所述时序获取单元还用于,所述获取时序特征,包括:获取所述历史轮次会话和所述当前轮次会话的时间差;获取预先确定的时间桶;根据时间桶获取所述时间差对应的时序特征,所述时序特征表征所述时间差对应的时间桶中的时间区域。

[0016] 根据本公开提供的一种实体追踪装置,所述装置还包括编码单元,用于:对所述历史轮次会话的第一用户文本和系统回复文本进行文本编码,得到所述历史轮次会话特征;对所述当前轮次会话的第二用户文本进行文本编码,得到所述当前轮次会话特征。

[0017] 根据本公开提供的一种实体追踪装置,所述意图获取单元还用于:根据所述当前轮次会话特征得到所述第二用户文本的意图词向量;采用One-Hot编码方式对意图词向量进行编码,得到所述当前轮次会话意图。

[0018] 根据本公开提供的一种实体追踪装置,所述槽值对获取单元,还用于:采用预置的词向量模型根据所述历史轮次会话特征获取所述第一用户文本的槽值信息;根据获取的槽位信息和所述槽值信息生成所述第一槽值对。

[0019] 根据本公开提供的一种实体追踪装置,所述装置还包括语义获取单元,用于:根据所述继承判断结果、所述历史轮次会话特征和所述当前轮次会话特征获取当前轮次会话的语义理解结果。

[0020] 本公开还提供一种电子设备,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程序,所述处理器执行所述程序时实现如上述任一种所述实体追踪方法的步骤。

[0021] 本公开还提供一种非暂态计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现如上述任一种所述实体追踪方法的步骤。

[0022] 本公开提供的实体追踪方法、装置、电子设备和非暂态计算机可读存储介质,通过历史轮次会话与当前轮次会话间的时间差来进行对话管理,并结合历史轮会话和当前轮次会话实现对第一槽值对信息的追踪,可以提高实体追踪的准确性,从而更好地理解用户的语义并给出较为准确的系统回复。

附图说明

[0023] 为了更清楚地说明本公开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一简单地介绍,显而易见地,下面描述中的附图是本公开的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0024] 图1本公开提供的实体追踪方法的流程示意图之一;

[0025] 图2本公开提供的实体追踪方法的流程示意图之二;

[0026] 图3是本公开提供的实体追踪模型的训练方法的流程示意图;

[0027] 图4是本公开提供的实体追踪装置的结构示意图;

[0028] 图5是本公开提供的电子设备的结构示意图。

具体实施方式

[0029] 为使本公开的目的、技术方案和优点更加清楚,下面将结合本公开中的附图,对本公开中的技术方案进行清楚、完整地描述,显然,所描述的实施例是本公开一部分实施例,而不是全部的实施例。基于本公开中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本公开保护的范围。

[0030] 在本公开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,而非旨在限制本公开一个或多个实施例。在本公开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,本公开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。

[0031] 应当理解,尽管在本公开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,在不脱离本公开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”。

[0032] 在对话系统中,用户和系统的交互往往需要多轮交互来达到用户的目的。例如,在订票系统中,系统需要通过多次和用户的交互来确认用户出发地,目的地,出发时间等信息。

[0033] 对话系统的对话管理主要作用是根据自然语言理解的结果来更新系统的状态,并生成相应的系统动作,对话管理的主要内容包括对话状态追踪和对话策略学习。自然语言理解主要作用是对用户输入的句子或者语音识别的结果进行处理,提取用户的对话意图以及用户所传递的信息。多轮对话中,对话状态的正确管理,可以使得用户和系统之间的对话更真实、更合理。

[0034] 对话状态可以理解作为一种表示用户意图和槽值对的数据结构。多轮对话系统中,对话信息是实体追踪的重要特征,但是仅仅用话术顺序建模并不能完全模拟对话真实情况。比如用户第一句话是“到成都需要开多久的车”,第二句是“今天的天气如何”,可能会出现两种情况:情况一,第二句话是在第一句话过后10秒钟说的,那第二句话中的槽值对应该继承第一句中的“地址—成都”,也就是说此时用户表述的就是“成都今天的天气”。情况二,第二句话是在第一句话1分钟过后说的,那此时用户很可能并不是想知道“成都今天的天气”,而是“所在位置今天的天气”。

[0035] 在第二种情况中,继承上文“地址—成都”的槽值对,显然不符合对话的真实情况,会直接给用户带来错误的答案。此时仅根据对话顺序来处理对话,不能够完全覆盖真实场景,一旦错误确认对话状态,便直接给对话带来错误或冗余,造成不好的用户体验。

[0036] 为解决这些问题,本公开实施例提供一种实体追踪方法、装置、电子设备和非暂态计算机可读存储介质。

[0037] 下面结合附图对本公开实施例实施方式进行详细说明。

[0038] 如图1所示的是本公开实施例的实体追踪方法的流程图。本公开实施例提供的方法可以由任意具备计算机处理能力的电子设备执行,例如终端或服务器。如图1所示,该实体追踪方法包括:

[0039] 步骤102,根据获取的历史轮次会话特征获取历史轮次会话中的第一槽值对。

[0040] 具体地,历史轮次会话指的是当前轮次会话之前的会话。历史轮次会话特征包括在历史轮次会话中的用户提问文本和系统回复文本的特征。第一槽值对包括槽位信息和槽值信息,第一槽值对根据用户提问文本得到,用于表征会话状态。

[0041] 步骤104,根据获取的当前轮次会话特征获取当前轮次会话的当前轮次会话意图。

[0042] 具体地,当前轮次会话特征包括在当前轮次会话中的用户提问文本的特征。会话意图根据用户提问文本得到,用于表征会话状态。

[0043] 步骤106,获取时序特征,时序特征根据历史轮次会话和当前轮次会话的时间差得到。

[0044] 具体地,系统根据会话日志中的不同轮次会话的用户会话起始时间可以得到不同轮次会话的时间差,进而可以以时间桶的方式对该时间差进行编码,得到时序特征。

[0045] 步骤108,将第一槽值对、当前轮次会话意图和时序特征输入到预置的实体追踪模型中,得到第一槽值对的继承判断结果,以根据继承判断结果进行当前轮次会话的语义理解,其中,继承判断结果表征第一槽值对是否需要被继承到当前轮次会话中。

[0046] 具体地,实体追踪模型可以给出第一槽值对是否需要被继承的判断结果。如果需

要继承,则应该结合第一槽值对中的信息对当前轮次会话进行语义理解。

[0047] 本公开实施例的技术方案,提出基于时间间隔的实体追踪方法,通过参考不同轮次会话的时间差进行对话管理。具体地,本公开实施例的技术方案以会话之间的时间间隔作为时序特征,将其和历史轮次会话特征、当前轮次会话特征融合,训练一个实体追踪模型,决策是否对上文实体进行继承。该方案可以准确地判断用户目的,使人机多轮对话更符合人人对话。

[0048] 本公开实施例的技术方案基于实体追踪模型进行继承判断,因此需要预先训练得到实体追踪模型。实体追踪模型可以为一种深度学习模型,且并不局限于此。

[0049] 如图3所示,本公开实施例的实体追踪模型的训练方法包括:

[0050] 步骤302,获取样本会话数据集,样本会话数据集中的样本会话数据至少包括两个轮次会话的会话数据,样本会话数据包括经过标注的以下训练标签:前轮次会话中的第一槽值对、后轮次会话的会话意图,以及前轮次会话和后轮次会话的时间差。

[0051] 步骤304,根据样本会话数据集和训练标签对初始实体追踪模型进行训练,得到实体追踪模型。

[0052] 具体地,训练实体追踪模型时,可以获取多轮会话系统中带时间的会话日志,然后进行人工标注。在人工标注过程中,需要标注出每一轮用户话术对应的会话状态,包含用户意图和槽值对信息。通过标注结果可以知道对于前文提及的槽值信息,当前轮次是否会继承,构造出训练实体追踪模型需要的数据。

[0053] 对初始实体追踪模型进行训练的过程中,通过构造损失函数,可以通过损失函数值调整实体追踪模型,直到该损失函数收敛,完成对实体追踪模型的训练。

[0054] 在步骤102之前,实体追踪方法还包括:对历史轮次会话的第一用户文本和系统回复文本进行文本编码,得到历史轮次会话特征;对当前轮次会话的第二用户文本进行文本编码,得到当前轮次会话特征。

[0055] 如图2所示,虚线框中文本201和文本202是历史轮次会话信息,包含第一用户文本和系统回复文本,对历史轮次会话信息进行文本编码可以将其转化为文本向量,即作为上下文特征的历史轮次会话特征。文本203是当前轮次的第二用户文本,同样对第二用户文本进行文本编码可以将其转化为文本向量,即当前轮次会话特征。

[0056] 在本公开实施例中,文本编码器可以为LSTM(Long Short Term Memory networks,长短期记忆网络)或者BERT(Bidirectional Encoder Representations from Transformers,基于转换器的双向编码表征),且并不局限于此。

[0057] 在步骤102中,采用预置的词向量模型根据历史轮次会话特征获取第一用户文本的槽值信息;根据获取的槽位信息和槽值信息生成第一槽值对。

[0058] 如图2所示,候选槽值对204包含槽位信息:地址和槽值信息:成都,采用槽值编码器对齐进行槽编码,可以得到槽值对向量。

[0059] 槽值编码器可以实现槽位和槽值的词向量的拼接,将第一用户文本输入词向量模型word2vec可以得到槽值。word2vec是一种用来产生词向量的模型,可以进行无监督文本聚类产生词向量。

[0060] 在步骤104中,根据当前轮次会话特征得到第二用户文本的意图词向量;采用One-Hot编码方式对意图词向量进行编码,得到当前轮次会话意图。

[0061] 可以看出,第二用户文本203的意图是询问天气,将当前意图205采用意图编码器进行编码可以得到意图向量。采用One-Hot编码的方式进行意图编码,可以构建意图长度的一维全零向量,然后将对应意图位置的值设置为1,此处是将天气位置的值设置为1。

[0062] One-Hot (独热) 编码又称为一位有效编码,主要是采用N位状态寄存器来对N个状态进行编码,每个状态都由独立的寄存器位,并且在任意时候只有一位有效。One-Hot编码是分类变量作为二进制向量的表示。

[0063] 以一个询问天气的对话为例,比如用户输入:“今天深圳的天气怎么样?”,此时用户所表达的是查询天气,可以认为查询天气就是一种意图,那具体查询哪里的天气,哪一天的天气?在这里用户传递出以下信息,地点=深圳,日期=今天,这个地点和日期就是槽位信息,对应地,深圳和今天就是槽值信息。如果在一个对话系统中含有多种意图和槽值,进行意图识别属于文本分类的任务,进行槽值填充属于序列标注的任务。

[0064] 在步骤106中,获取历史轮次会话和当前轮次会话的时间差;获取预先确定的时间桶;根据时间桶获取时间差对应的时序特征,时序特征表征时间差对应的的时间桶中的时间区域。

[0065] 具体地,时间差206表示两轮用户会话的时间差,因为系统回复较快,所以时间差是针对用户的不同轮次的会话的时间差,然后采用时间编码器将时间差转化为时间向量。时间编码可以以分桶的形式表示,即设置桶的信息,如0~10s、10s~30s、30s~60s、>60s这些桶,然后将实际值放入这些桶中,时间向量便只有一维,表示该时间差所处的区间。

[0066] 如图2所示的整个实体追踪流程包含三个部分,第一部分是构建会话的上下文特征,包含历史轮次的用户文本、用户实体、系统回复。第二部分是构建当前轮次会话特征,包含当前轮次的用户文本、用户意图等;最后是构建时序特征,主要是会话的时间差。

[0067] 得到这些特征向量后,可以将其拼接后输入到全连接神经网络层207中,最后得到预测结果,即对于候选槽值对是否需要继承。

[0068] 在步骤108中,根据继承判断结果、历史轮次会话特征和当前轮次会话特征获取当前轮次会话的语义理解结果。

[0069] 在一种实施例,历史轮次会话的用户文本为:到上海需要开多久的车,系统会话文本为:需要两个小时,第二槽值对为地址-上海,当前轮次会话的用户文本为:今天的天气如何。如果第二槽值对对应的继承判断结果为否,则当前轮次会话的用户文本的语义理解结果为:本地今天的天气如何。如果第二槽值对对应的继承判断结果为是,则当前轮次会话的用户文本的语义理解结果为:上海今天的天气如何。

[0070] 根据当前轮次会话的用户文本的语义理解结果,对话系统可以产生相应的系统回答话语,来反馈给用户,例如,本地今天晴朗,或者上海今天有小雨。

[0071] 对当前轮次会话的用户文本进行语义理解时,所继承的槽值对可以不局限于一个。如果继承的槽值对为两个或多个,槽值对可以来自于相同轮次的历史会话,也可以来自不同轮次的历史会话。

[0072] 在一个例子中,针对历史会话文本:今天深圳的天气怎么样,具有两个槽值对,地点-深圳,日期-今天。当前会话文本:交通状况怎么样?当前会话意图为:交通。两次会话的时间间隔为60s。采用本公开实施例的技术方案。可以得到两个槽值对均需要被继承,也就是说,当前会话的语义理解结果为:深圳今天的交通状况怎么样?

[0073] 本公开提供的实体追踪方法,通过历史轮次会话与当前轮次会话间的时间差来进行对话管理,并结合历史轮会话和当前轮次会话实现对槽值对信息的追踪,可以提高实体追踪的准确性,从而更好地理解用户的语义并给出较为准确的系统回复。

[0074] 下面对本公开提供的实体追踪装置进行描述,下文描述的实体追踪装置与上文描述的实体追踪方法可相互对应参照。

[0075] 如图4所示,本公开实施例的一种实体追踪装置,该装置包括:

[0076] 槽值对获取单元402,可以用于根据获取的历史轮次会话特征获取历史轮次会话中的第一槽值对。

[0077] 意图获取单元404,可以用于根据获取的当前轮次会话特征获取当前轮次会话的当前轮次会话意图。

[0078] 时序获取单元406,可以用于获取时序特征,时序特征根据历史轮次会话和当前轮次会话的时间差得到。

[0079] 追踪单元408,可以用于将第一槽值对、当前轮次会话意图和时序特征输入到预置的实体追踪模型中,得到第一槽值对的继承判断结果,以根据继承判断结果进行当前轮次会话的语义理解,其中,继承判断结果表征第一槽值对是否需要被继承到当前轮次会话中。

[0080] 本公开实施例的技术方案,提出基于时间间隔的实体追踪方法,通过参考不同轮次会话的时间差进行对话管理。具体地,本公开实施例的技术方案以会话之间的时间间隔作为时序特征,将其和历史轮次会话特征、当前轮次会话特征融合,训练一个实体追踪模型,决策是否对上文实体进行继承。该方案可以准确地判断用户目的,使人机多轮对话更符合人人对话。

[0081] 在本公开实施例中,实体追踪装置还可以包括训练单元,用于对实体追踪模型进行训练,训练单元用于:获取样本会话数据集,样本会话数据集中的样本会话数据至少包括两个轮次会话的会话数据,样本会话数据包括经过标注的以下训练标签:前轮次会话中的第一槽值对、后轮次会话的会话意图,以及前轮次会话和后轮次会话的时间差;根据样本会话数据集和训练标签对初始实体追踪模型进行训练,得到实体追踪模型。

[0082] 在本公开实施例中,时序获取单元还可以用于,获取时序特征,包括:获取历史轮次会话和当前轮次会话的时间差;获取预先确定的时间桶;根据时间桶获取时间差对应的时序特征,时序特征表征时间差对应的时间桶中的时间区域。

[0083] 在本公开实施例中,实体追踪装置还可以包括编码单元,用于:对历史轮次会话的第一用户文本和系统回复文本进行文本编码,得到历史轮次会话特征;对当前轮次会话的第二用户文本进行文本编码,得到当前轮次会话特征。

[0084] 在本公开实施例中,意图获取单元可以还用于:根据当前轮次会话特征得到第二用户文本的意图词向量;采用One-Hot编码方式对意图词向量进行编码,得到当前轮次会话意图。

[0085] 在本公开实施例中,槽值对获取单元,还可以用于:采用预置的词向量模型根据历史轮次会话特征获取第一用户文本的槽值信息;根据获取的槽位信息和槽值信息生成第一槽值对。

[0086] 在本公开实施例中,实体追踪装置还可以包括语义获取单元,用于:根据继承判断结果、历史轮次会话特征和当前轮次会话特征获取当前轮次会话的语义理解结果。

[0087] 由于本公开的示例实施例的实体追踪装置的各个功能模块与上述实体追踪方法的示例实施例的步骤对应,因此对于本公开装置实施例中未披露的细节,请参照本公开上述的实体追踪方法的实施例。

[0088] 本公开提供的实体追踪装置,通过历史轮次会话与当前轮次会话间的时间差来进行对话管理,并结合历史轮会话和当前轮次会话实现对槽值对信息的追踪,可以提高实体追踪的准确性,从而更好地理解用户的语义并给出较为准确的系统回复。

[0089] 图5示例了一种电子设备的实体结构示意图,如图5所示,该电子设备可以包括:处理器(processor)510、通信接口(Communications Interface)520、存储器(memory)530和通信总线540,其中,处理器510,通信接口520,存储器530通过通信总线540完成相互间的通信。处理器510可以调用存储器530中的逻辑指令,以执行实体追踪方法,该方法包括:根据获取的历史轮次会话特征获取历史轮次会话中的第一槽值对;根据获取的当前轮次会话特征获取当前轮次会话的当前轮次会话意图;获取时序特征,所述时序特征根据所述历史轮次会话和所述当前轮次会话的时间差得到;将所述第一槽值对、所述当前轮次会话意图和所述时序特征输入到预置的实体追踪模型中,得到所述第一槽值对的继承判断结果,以根据所述继承判断结果进行当前轮次会话的语义理解,其中,所述继承判断结果表征所述第一槽值对是否需要被继承到所述当前轮次会话中。

[0090] 此外,上述的存储器530中的逻辑指令可以通过软件功能单元的形式实现并作为独立的产品销售或使用时,可以存储在一个计算机可读取存储介质中。基于这样的理解,本公开的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,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)执行本公开各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(ROM, Read-Only Memory)、随机存取存储器(RAM, Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0091] 另一方面,本公开还提供一种计算机程序产品,所述计算机程序产品包括存储在非暂态计算机可读存储介质上的计算机程序,所述计算机程序包括程序指令,当所述程序指令被计算机执行时,计算机能够执行上述各方法所提供的实体追踪方法,该方法包括:根据获取的历史轮次会话特征获取历史轮次会话中的第一槽值对;根据获取的当前轮次会话特征获取当前轮次会话的当前轮次会话意图;获取时序特征,所述时序特征根据所述历史轮次会话和所述当前轮次会话的时间差得到;将所述第一槽值对、所述当前轮次会话意图和所述时序特征输入到预置的实体追踪模型中,得到所述第一槽值对的继承判断结果,以根据所述继承判断结果进行当前轮次会话的语义理解,其中,所述继承判断结果表征所述第一槽值对是否需要被继承到所述当前轮次会话中。

[0092] 又一方面,本公开还提供一种非暂态计算机可读存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现以执行上述各提供的实体追踪方法,该方法包括:根据获取的历史轮次会话特征获取历史轮次会话中的第一槽值对;根据获取的当前轮次会话特征获取当前轮次会话的当前轮次会话意图;获取时序特征,所述时序特征根据所述历史轮次会话和所述当前轮次会话的时间差得到;将所述第一槽值对、所述当前轮次会话意图和所述时序特征输入到预置的实体追踪模型中,得到所述第一槽值对的继承判断结果,以

根据所述继承判断结果进行当前轮次会话的语义理解,其中,所述继承判断结果表征所述第一槽值对是否需要被继承到所述当前轮次会话中。

[0093] 以上所描述的装置实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下,即可以理解并实施。

[0094] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到各实施方式可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件。基于这样的理解,上述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计算机可读存储介质中,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)执行各个实施例或者实施例的某些部分所述的方法。

[0095] 最后应说明的是:以上实施例仅用以说明本公开的技术方案,而非对其限制;尽管参照前述实施例对本公开进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本公开各实施例技术方案的精神和范围。

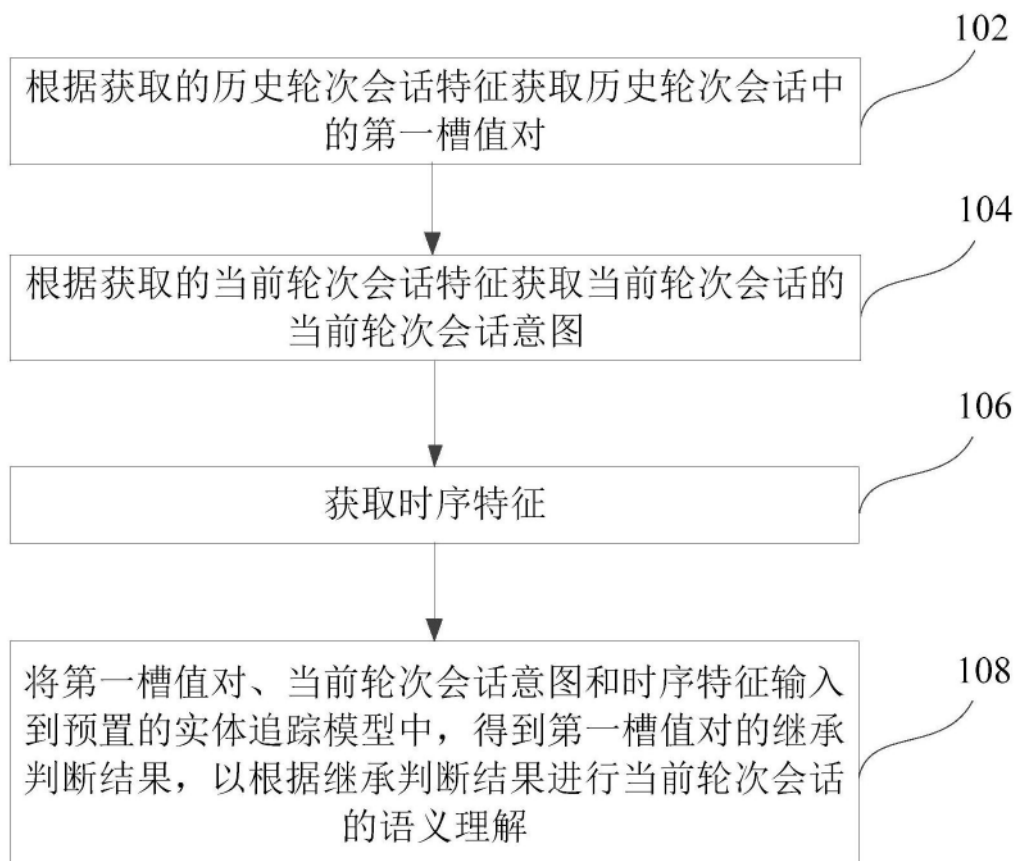


图1

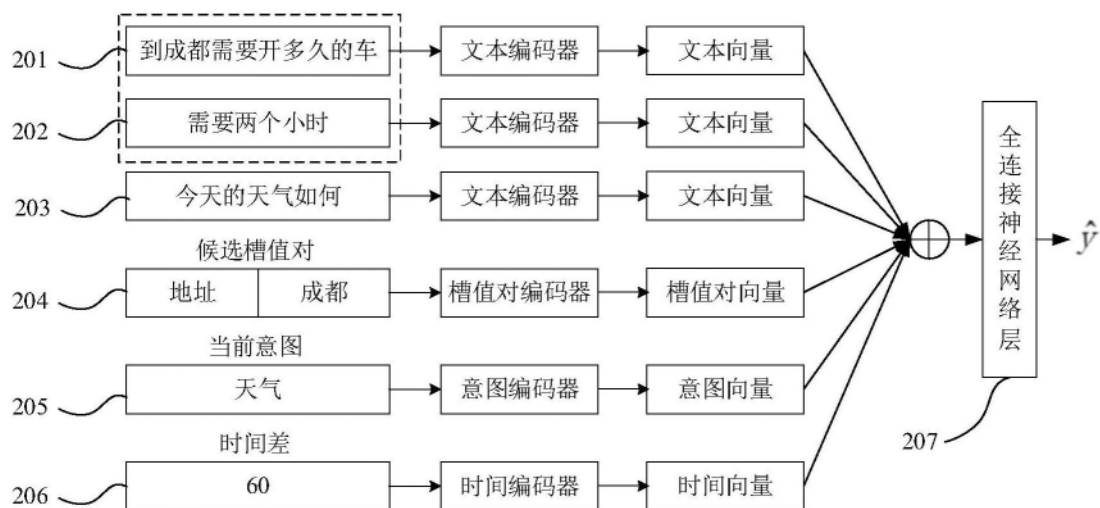


图2

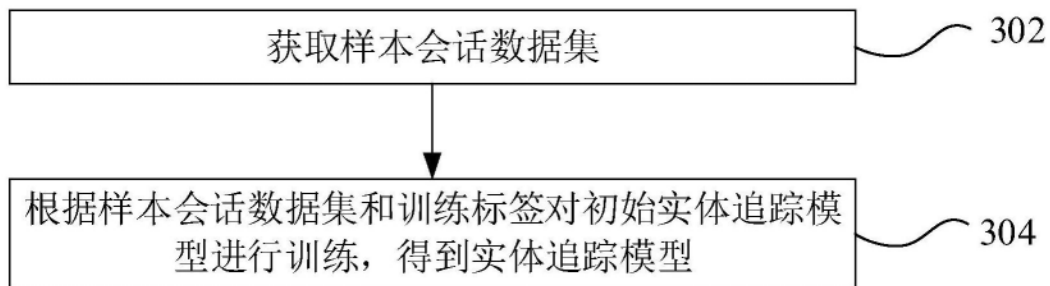


图3

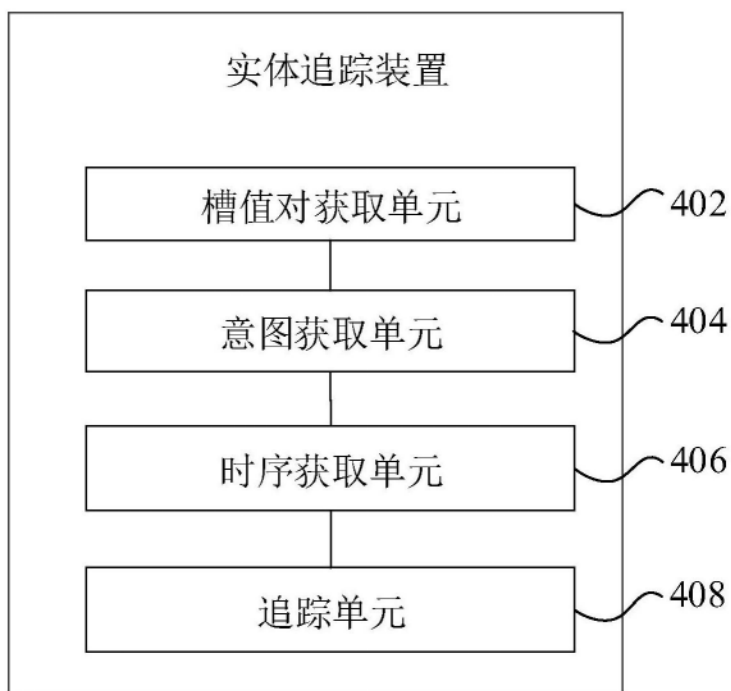


图4

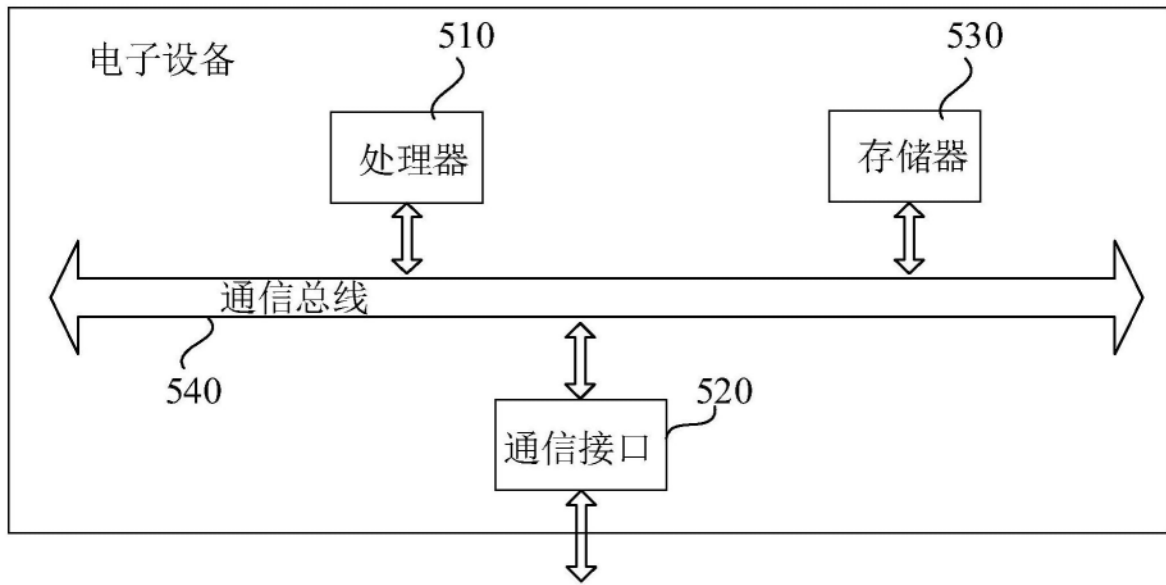


图5