



(12) 发明专利

(10) 授权公告号 CN 110114764 B

(45) 授权公告日 2023. 09. 15

(21) 申请号 201780081065.5
 (22) 申请日 2017.09.18
 (65) 同一申请的已公布的文献号
 申请公布号 CN 110114764 A
 (43) 申请公布日 2019.08.09
 (85) PCT国际申请进入国家阶段日
 2019.06.27
 (86) PCT国际申请的申请数据
 PCT/CN2017/102141 2017.09.18
 (87) PCT国际申请的公布数据
 W02019/051847 EN 2019.03.21
 (73) 专利权人 微软技术许可有限责任公司
 地址 美国华盛顿州
 (72) 发明人 吴先超
 (74) 专利代理机构 永新专利商标代理有限公司
 72002
 专利代理师 张立达

(51) Int.Cl.
 G06F 16/9032 (2019.01)
 (56) 对比文件
 CN 107077529 A, 2017.08.18
 CN 1655510 A, 2005.08.17
 US 2009077007 A1, 2009.03.19
 US 2014255882 A1, 2014.09.11
 US 2009076903 A1, 2009.03.19
 US 2014122407 A1, 2014.05.01
 US 2015228062 A1, 2015.08.13
 CN 105426436 A, 2016.03.23
 CN 105068661 A, 2015.11.18
 CN 101410866 A, 2009.04.15
 CN 107025297 A, 2017.08.08
 CN 106415586 A, 2017.02.15

审查员 周炬

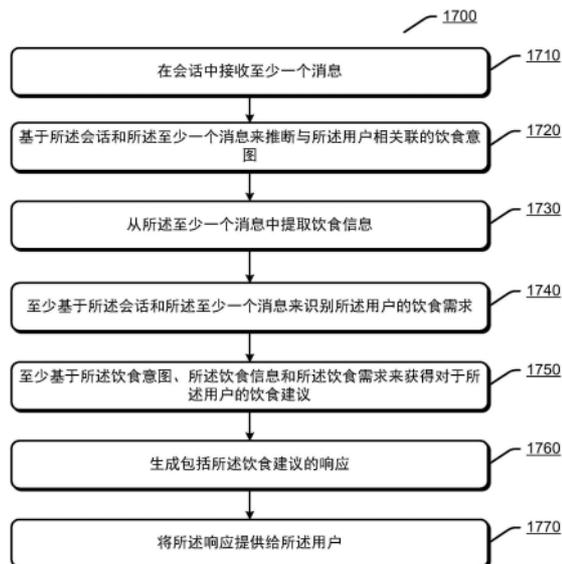
权利要求书3页 说明书23页 附图16页

(54) 发明名称

在会话中提供饮食帮助

(57) 摘要

本公开提供了用于在会话中向用户提供饮食帮助的方法和装置。可以在所述会话中接收至少一个消息,所述会话处于所述用户和电子对话代理之间。可以基于所述会话和所述至少一个消息来推断与所述用户相关联的饮食意图。可以从所述至少一个消息中提取饮食信息。可以至少基于所述会话和所述至少一个消息来识别所述用户的饮食需求。可以至少基于所述饮食意图、所述饮食信息和所述饮食需求来获得对于所述用户的饮食建议。可以生成包括所述饮食建议的响应并将所述响应提供给所述用户。



1. 一种用于在会话中向用户提供饮食帮助的方法,包括:
 - 在所述会话中接收至少一个消息,所述会话处于所述用户和电子对话代理之间;
 - 通过支持向量机(SVM)模型来推断与所述用户相关联的饮食意图,其中,所述SVM模型基于所述会话、所述至少一个消息和所述用户的用户简档来实施饮食意图分类,以从食谱推荐、饮食推荐、产品推荐和饮食知识获取中确定与所述用户相关联的所述饮食意图;
 - 从所述至少一个消息中提取饮食信息;
 - 至少基于所述会话和所述至少一个消息来识别所述用户的饮食需求;
 - 至少基于所述饮食意图、所述饮食信息和所述饮食需求来获得对于所述用户的饮食建议;
 - 生成包括所述饮食建议的响应;以及
 - 将所述响应提供给所述用户。
2. 根据权利要求1所述的方法,其中,
 - 所述饮食意图指示食谱推荐;
 - 所述饮食信息包括至少一种烹饪配料;
 - 所述饮食需求包括烹饪方法偏好和口味偏好中的至少一个;并且
 - 所述饮食建议包括食谱。
3. 根据权利要求2所述的方法,其中,所述识别所述饮食需求包括以下中的至少一个:
 - 从所述会话和所述至少一个消息中提取所述烹饪方法偏好和口味偏好中的至少一个;以及
 - 从所述用户简档中检索所述烹饪方法偏好和口味偏好中的至少一个,所述用户简档是至少基于所述会话和所述至少一个消息来建立的。
4. 根据权利要求2所述的方法,其中,所述获得所述饮食建议包括:
 - 检索至少一个经索引的食谱,所述至少一个经索引的食谱与所述至少一种烹饪配料以及所述烹饪方法偏好和口味偏好中的至少一个相关。
5. 根据权利要求2所述的方法,其中,所述获得所述饮食建议包括:
 - 对所述至少一种烹饪配料进行语义扩展以获得至少一种扩展的烹饪配料;
 - 基于所述至少一种扩展的烹饪配料以及所述烹饪方法偏好和口味偏好中的至少一个来检索至少一个经索引的食谱;以及
 - 基于所述至少一个经索引的食谱来生成食谱,所生成的食谱与所述至少一种烹饪配料以及所述烹饪方法偏好和口味偏好中的至少一个相关。
6. 根据权利要求2所述的方法,还包括:
 - 至少基于所述用户的动作和/或所述至少一种烹饪配料的烹饪状态,指导所述用户遵循所述食谱进行烹饪。
7. 根据权利要求1所述的方法,其中,
 - 所述饮食意图指示饮食推荐和产品推荐中的至少一个;
 - 所述饮食信息包括食物类型、食物名称、菜肴名称和关心的疾病中的至少一个;
 - 所述饮食需求包括所述用户的健康状况;并且
 - 所述饮食建议包括食物名称、菜肴名称和产品信息中的至少一个。
8. 根据权利要求7所述的方法,其中,所述识别所述饮食需求包括以下中的至少一个:

从所述会话和所述至少一个消息中提取所述健康状况;以及
从所述用户简档中检索所述健康状况,所述用户简档是至少基于所述会话和所述至少一个消息来建立的。

9. 根据权利要求7所述的方法,其中,所述获得所述饮食建议包括:

基于以下中的至少一个来对多个候选推荐进行排序:所述用户简档、对所述多个候选推荐的点击率、所述多个候选推荐的关键词竞价信息、与饮食和/或疾病相关的知识图谱、关于所述多个候选推荐的历史情感、以及产品数据集;以及

选择一个或多个排序最高的候选推荐作为所述饮食建议。

10. 根据权利要求1所述的方法,其中,

所述饮食意图指示饮食知识获取;

所述饮食信息包括食物类型、食物名称、菜肴名称和关心的疾病中的至少一个;

所述饮食需求包括所述用户的健康状况;并且

所述饮食建议包括饮食知识。

11. 根据权利要求10所述的方法,其中,所述获得所述饮食建议包括:

至少基于所述饮食意图、所述饮食信息和所述饮食需求,从知识图谱或知识问题-答案对集合中检索所述饮食知识。

12. 根据权利要求1所述的方法,其中,所述至少一个消息包括文本消息、语音消息和图像消息中的至少一个。

13. 一种用于在会话中向用户提供饮食帮助的装置,包括:

消息接收模块,用于在所述会话中接收至少一个消息,所述会话处于所述用户和电子对话代理之间;

饮食意图推断模块,用于通过支持向量机(SVM)模型来推断与所述用户相关联的饮食意图,其中,所述SVM模型基于所述会话、所述至少一个消息和所述用户的用户简档来实施饮食意图分类,以从食谱推荐、饮食推荐、产品推荐和饮食知识获取中确定与所述用户相关联的所述饮食意图;

饮食信息提取模块,用于从所述至少一个消息中提取饮食信息;

饮食需求识别模块,用于至少基于所述会话和所述至少一个消息来识别所述用户的饮食需求;

饮食建议获得模块,用于至少基于所述饮食意图、所述饮食信息和所述饮食需求来获得对于所述用户的饮食建议;

响应生成模块,用于生成包括所述饮食建议的响应;以及

响应提供模块,用于将所述响应提供给所述用户。

14. 根据权利要求13所述的装置,其中,

所述饮食意图指示食谱推荐;

所述饮食信息包括至少一种烹饪配料;

所述饮食需求包括烹饪方法偏好和口味偏好中的至少一个;并且

所述饮食建议包括食谱。

15. 根据权利要求14所述的装置,其中,所述饮食需求识别模块进一步用于以下中的至少一个:

从所述会话和所述至少一个消息中提取所述烹饪方法偏好和口味偏好中的至少一个；
以及

从所述用户简档中检索所述烹饪方法偏好和口味偏好中的至少一个，所述用户简档是至少基于所述会话和所述至少一个消息来建立的。

16. 根据权利要求14所述的装置，其中，所述饮食建议获得模块进一步用于：

对所述至少一种烹饪配料进行语义扩展以获得至少一种扩展的烹饪配料；

基于所述至少一种扩展的烹饪配料以及所述烹饪方法偏好和口味偏好中的至少一个来检索至少一个经索引的食谱；以及

基于所述至少一个经索引的食谱来生成食谱，所生成的食谱与所述至少一种烹饪配料以及所述烹饪方法偏好和口味偏好中的至少一个相关。

17. 根据权利要求14所述的装置，还包括：

烹饪指导模块，用于至少基于所述用户的动作和/或所述至少一种烹饪配料的烹饪状态，指导所述用户遵循所述食谱进行烹饪。

18. 根据权利要求13所述的装置，其中，

所述饮食意图指示饮食推荐和产品推荐中的至少一个；

所述饮食信息包括食物类型、食物名称、菜肴名称和关心的疾病中的至少一个；

所述饮食需求包括所述用户的健康状况；并且

所述饮食建议包括食物名称、菜肴名称和产品信息中的至少一个。

19. 根据权利要求13所述的装置，其中，

所述饮食意图指示饮食知识获取；

所述饮食信息包括食物类型、食物名称、菜肴名称和关心的疾病中的至少一个；

所述饮食需求包括所述用户的健康状况；并且

所述饮食建议包括饮食知识。

20. 一种用于在会话中向用户提供饮食帮助的装置，包括：

一个或多个处理器；以及

存储器，其存储计算机可执行指令，当所述计算机可执行指令被执行时使所述一个或多个处理器：

在所述会话中接收至少一个消息，所述会话处于所述用户和电子对话代理之间；

通过支持向量机 (SVM) 模型来推断与所述用户相关联的饮食意图，其中，所述SVM模型基于所述会话、所述至少一个消息和所述用户的用户简档来实施饮食意图分类，以从食谱推荐、饮食推荐、产品推荐和饮食知识获取中确定与所述用户相关联的所述饮食意图；

从所述至少一个消息中提取饮食信息；

至少基于所述会话和所述至少一个消息来识别所述用户的饮食需求；

至少基于所述饮食意图、所述饮食信息和所述饮食需求来获得对于所述用户的饮食建议；

生成包括所述饮食建议的响应；以及

将所述响应提供给所述用户。

在会话中提供饮食帮助

背景技术

[0001] 人工智能 (AI) 聊天机器人变得越来越流行,并且正在越来越多的场景中得到应用。聊天机器人被设计用于模拟人类对话,并且可以通过文本、语音、图像等与用户聊天。通常,聊天机器人可以扫描由用户输入的消息内的关键词或对消息应用自然语言处理,并向用户提供具有最匹配的关键词或最相似的措辞模式的响应。

发明内容

[0002] 提供本发明内容以便介绍一组概念,这组概念将在以下的具体实施方式中做进一步描述。本发明内容并非旨在标识所保护主题的关键特征或必要特征,也不旨在用于限制所保护主题的范围。

[0003] 本公开的实施例提出了用于在会话中向用户提供饮食帮助的方法和装置。可以在所述会话中接收至少一个消息,所述会话处于所述用户和电子对话代理之间。可以基于所述会话和所述至少一个消息来推断与所述用户相关联的饮食意图。可以从所述至少一个消息中提取饮食信息。可以至少基于所述会话和所述至少一个消息来识别所述用户的饮食需求。可以至少基于所述饮食意图、所述饮食信息和所述饮食需求来获得对于所述用户的饮食建议。可以生成包括所述饮食建议的响应并将所述响应提供给所述用户。

[0004] 应当注意,以上一个或多个方面包括以下详细描述以及权利要求中具体指出的特征。下面的说明书及附图详细提出了所述一个或多个方面的某些说明性特征。这些特征仅仅指示可以实施各个方面的原理的多种方式,并且本公开旨在包括所有这些方面和其等同变换。

附图说明

[0005] 以下将结合附图描述所公开的多个方面,这些附图被提供用以说明而非限制所公开的多个方面。

[0006] 图1示出了根据实施例的部署聊天机器人的示例性网络架构。

[0007] 图2示出了根据实施例的示例性聊天机器人系统。

[0008] 图3示出了根据实施例的示例性用户界面。

[0009] 图4示出了根据实施例的AI助手的示例性硬件结构。

[0010] 图5示出了根据实施例的用于提供饮食帮助的示例性总体过程。

[0011] 图6A和图6B示出了根据实施例的示例性聊天窗口。

[0012] 图7示出了根据实施例的用于建立知识图谱和知识问题-答案(QA)对集合的示例性过程。

[0013] 图8A和图8B示出了根据实施例的示例性依存句法分析。

[0014] 图9A和图9B示出了根据实施例的示例性依存句法分析。

[0015] 图10示出了根据实施例的用于提供饮食和产品推荐的示例性过程。

[0016] 图11示出了根据实施例的示例性聊天窗口。

- [0017] 图12示出了根据实施例的用于生成食谱的示例性过程。
- [0018] 图13示出了根据实施例的示例性文本到图像生成模型。
- [0019] 图14示出了根据实施例的用于视频捕获的示例性过程。
- [0020] 图15A示出了根据实施例的身体关节的示例性表示。
- [0021] 图15B示出了根据实施例的用于身体关节检测的示例性过程。
- [0022] 图16示出了根据实施例的获得用于视频捕获的训练数据的示例性过程。
- [0023] 图17示出了根据实施例的用于在会话中向用户提供饮食帮助的示例性方法的流程图。
- [0024] 图18示出了根据实施例的用于在会话中向用户提供饮食帮助的示例性装置。
- [0025] 图19示出了根据实施例的用于在会话中向用户提供饮食帮助的示例性装置。

具体实施方式

[0026] 现在将参考多种示例性实施方式来讨论本公开。应当理解,这些实施方式的讨论仅仅用于使得本领域技术人员能够更好地理解并从而实施本公开的实施例,而并非教导对本公开的范围的任何限制。

[0027] 在人们的生活中可能存在各种与饮食有关的疑问或问题。例如,人们可能想知道什么样的食物有利于从某些疾病或伤害的短时间恢复,或者有利于长期身体健康或某些疾病的预防。通常,人们可能自己浏览健康或饮食相关网站以找到答案,或通过搜索引擎搜索相关信息,搜索引擎可以对健康或饮食相关网站上的内容进行索引并提供与人们的查询语义匹配的搜索结果。但是,主动浏览或基于搜索引擎的信息检索是耗时的,并且所获得的信息是更为面向一般性的而没有参考人们的个性化信息。此外,例如,人们可能会遇到这样的问题:当烹饪时,他们不知道如何用一些烹饪配料进行烹饪。从饮食或烹饪相关网站获得食谱是耗时的,并且网络上可获得的食谱可能与人们感兴趣的烹饪配料无关。

[0028] 本公开的实施例提出了在用户和电子对话代理之间的会话中向用户提供饮食帮助。电子对话代理可以是例如聊天机器人。通常,聊天机器人可以与用户进行自动会话。在本文中,“会话”可以指两个聊天参与者之间的时间连续对话,并且可以包括对话中的消息和响应,其中,“消息”是指由用户输入的任何信息,例如来自用户的查询、用户对聊天机器人的问题的答案、用户的意见等,“响应”是指由聊天机器人提供的任何信息,例如聊天机器人对用户的问题的答案、聊天机器人的评论等。术语“消息”和术语“查询”也可以互换使用。

[0029] 本公开的实施例可以基于用户的个性化信息,例如用户的健康状况、关心的疾病、烹饪方法偏好、口味偏好等,来提供饮食帮助。饮食帮助可以包括各种饮食建议。在一些方面,聊天机器人向用户提供的饮食帮助可以是与用户关于食物、菜肴(dish)、关心的疾病等的查询有关的饮食知识。在本文中,“食物”可以指人们可以吃或喝的、且不经烹饪或经过有限烹饪过程而获得的东西,而“菜肴”可以指通过烹饪由准备好的食物所形成的可食用品。然而,术语“食物”和术语“菜肴”也可以互换使用。在一些方面,聊天机器人向用户提供的饮食帮助可以是饮食推荐或产品推荐,例如,推荐的食物或菜肴、来自合作伙伴实体的产品信息等。在一些方面,聊天机器人向用户提供的饮食帮助可以是食谱推荐,例如,现有食谱或新生成的食谱。在一些方面,聊天机器人向用户提供的饮食帮助可以是烹饪指导,例如,关于如何根据食谱进行烹饪的指导。

[0030] 图1示出了根据实施例的部署聊天机器人的示例性网络架构100。

[0031] 在图1中,网络110被应用于在终端设备120和聊天机器人服务器130之间进行互连。

[0032] 网络110可以是能够对网络实体进行互连的任何类型的网络。网络110可以是单个网络或各种网络的组合。在覆盖范围方面,网络110可以是局域网(LAN)、广域网(WAN)等。在承载介质方面,网络110可以是有线网络、无线网络等。在数据交换技术方面,网络110可以是电路交换网络、分组交换网络等。

[0033] 终端设备120可以是能够连接到网络110、访问网络110上的服务器或网站、处理数据或信号等的任何类型的电子计算设备。例如,终端设备120可以是台式计算机、笔记本电脑、平板电脑、智能电话、AI终端等。尽管在图1中仅示出了一个终端设备,但是应当理解,可以有不同数量的终端设备连接到网络110。

[0034] 在一种实施方式中,终端设备120可以由用户使用。终端设备120可以包括可为用户提供自动聊天服务的聊天机器人客户端122。在一些情况下,聊天机器人客户端122可以与聊天机器人服务器130进行交互。例如,聊天机器人客户端122可以将用户输入的消息传送到聊天机器人服务器130,并且从聊天机器人服务器130接收与消息相关联的响应。然而,应当理解,在其它情况下,聊天机器人客户端122也可以本地地生成对用户输入的消息的响应,而不是与聊天机器人服务器130进行交互。

[0035] 聊天机器人服务器130可以连接到或包含聊天机器人数据库140。聊天机器人数据库140可以包括可由聊天机器人服务器130用于生成响应的信息。

[0036] 应当理解,图1中所示的所有网络实体都是示例性的,根据具体的应用需求,应用场景100中可以涉及任何其它网络实体。

[0037] 图2示出了根据实施例的示例性聊天机器人系统200。

[0038] 聊天机器人系统200可以包括用于呈现聊天窗口的用户界面(UI)210。聊天窗口可以由聊天机器人用于与用户进行交互。

[0039] 聊天机器人系统200可以包括核心处理模块220。核心处理模块220被配置用于通过与聊天机器人系统200的其它模块的协作,在聊天机器人的运行期间提供处理能力。

[0040] 核心处理模块220可以获得由用户在聊天窗口中输入的消息,并将消息存储在消息队列232中。消息可以采用各种多媒体形式,例如文本、语音、图像、视频等。

[0041] 核心处理模块220可以用先进先出的方式处理消息队列232中的消息。核心处理模块220可以调用应用程序接口(API)模块240中的处理单元来处理各种形式的消息。API模块240可以包括文本处理单元242、语音处理单元244、图像处理单元246等。

[0042] 对于文本消息,文本处理单元242可以对文本消息执行文本理解,并且核心处理模块220可以进一步确定文本响应。

[0043] 对于语音消息,语音处理单元244可以对语音消息执行语音到文本转换以获得文本语句,文本处理单元242可以对所获得的文本语句执行文本理解,并且核心处理模块220可以进一步确定文本响应。如果确定以语音提供响应,则语音处理单元244可以对文本响应执行文本到语音转换以生成相应的语音响应。

[0044] 对于图像消息,图像处理单元246可以对图像消息执行图像识别以生成相应的文本,并且核心处理模块220可以进一步确定文本响应。在一些情况下,图像处理单元246也可

以用于基于文本响应来获得图像响应。

[0045] 此外,尽管在图2中未示出,API模块240还可以包括任何其它处理单元。例如,API模块240可以包括视频处理单元,该视频处理单元用于与核心处理模块220协作以处理视频消息并确定响应。

[0046] 核心处理模块220可以通过数据库250来确定响应。数据库250可以包括可由核心处理模块220检索用于确定响应的多个索引项。

[0047] 数据库250可以包括纯聊天索引集251。纯聊天索引集251可以包括索引项,索引项被准备用于聊天机器人和用户之间的自由聊天,并且可以是利用来自例如社交网络的数据来建立的。纯聊天索引集251中的索引项可以采用或不采用问题-答案(QA)对(例如<问题,答案>)的形式。问题-答案对也可以被称为消息-响应对。

[0048] 数据库250可以包括知识图谱252。在本文中,知识图谱252可以指各个领域中的单个知识图谱或多个知识图谱。知识图谱252中的知识信息可以采用元组(tuple)的形式。知识图谱252可以包括医学领域知识图谱,其包括与各种疾病有关的知识信息。医学领域知识图谱可以进一步包括分别由各种疾病形成的子领域话题。知识图谱252可以包括食物领域知识图谱,其包括与各种食物有关的知识信息。知识图谱252还可以包括与疾病和食物两者相关的知识图谱。

[0049] 数据库250可以包括知识QA对集合253。知识QA对集合253可以由知识图谱252形成,并且可以包括QA对形式的知识信息。

[0050] 数据库250可以包括用户简档254。用户简档254可以包括用户的个性化信息。例如,用户简档254可以包括用户的性别、年龄、位置、健康状况、烹饪方法偏好、口味偏好等。

[0051] 数据库250可以包括会话日志255。会话日志255可以包括聊天机器人和用户之间的会话中的对话内容的记录,例如来自用户的消息、聊天机器人的响应等。

[0052] 数据库250可以包括产品数据集256。产品数据集256可以包括各种合作伙伴实体的饮食相关的产品信息。在本文中,“产品”可以指商品或服务,合作伙伴实体可以是聊天机器人可从其获得产品信息的商品的生产者或销售商或者服务的提供者,“产品信息”可以包括商品信息、服务信息、合作伙伴实体信息等。

[0053] 聊天机器人系统200可以包括模块集合260,模块集合260是可以由核心处理模块220操作以生成或获得响应的功能模块的集合。

[0054] 模块集合260可以包括意图分类模块261。意图分类模块261可以被配置为推断或确定用户的饮食意图。在本文中,“饮食意图”可以指在用户和聊天机器人之间的会话中明确或隐含表达的、用户的饮食相关的目的或期望。意图分类模块261可以基于用户的会话日志、当前或最新消息、用户简档等中的至少一个来推断各种饮食意图。饮食意图可以包括指示用户可能想要获得饮食相关的知识的饮食知识获取。饮食意图可以包括饮食推荐,其指示用户可能想要得到与用户的饮食关注所相关的一些饮食推荐。饮食意图可以包括产品推荐,其指示用户可能愿意接收饮食相关的产品信息。饮食意图可以包括食谱推荐,其指示用户需要获得至少基于感兴趣的烹饪配料的食谱。在某些情况下,饮食意图也可能是“无意图”,这指示用户在会话中没有特定的饮食相关意图或者不能推断出饮食相关意图。

[0055] 模块集合260可以包括饮食和产品推荐模块262。饮食和产品推荐模块262可以被配置用于基于知识图谱252、知识QA对集合253、用户简档254、产品数据集256等中的至少一

个来生成或确定对用户的饮食推荐或产品推荐。

[0056] 模块集合260可以包括食谱生成模块263。食谱生成模块263可以被配置用于基于感兴趣的烹饪配料、用户的烹饪方法或口味偏好等来生成食谱。食谱可以是符合用户需求的新生成的食谱。

[0057] 模块集合260可以包括烹饪指导模块264。烹饪指导模块264可以被配置用于指导用户遵循食谱来进行烹饪。例如,该指导可以是通过语音、文本、图像、视频等进行的。

[0058] 模块集合260可以包括响应提供模块265。响应提供模块265可以被配置用于提供或传递对用户的消息的响应。在一些实施方式中,响应可以是由响应提供模块265所生成或确定的。例如,当意图分类模块261没有推断出意图时,响应提供模块265可以从例如纯聊天索引集251中确定响应。例如,在将饮食意图推断为饮食知识获取时,响应提供模块265可以从例如知识图谱252、知识QA对集合253等中来确定响应。在一些实施方式中,由响应提供模块265所提供的响应可以基于来自其它模块的信息。例如,响应提供模块265可以基于来自饮食和产品推荐模块262的食物/菜肴名称或产品信息、来自食谱生成模块263的食谱、来自烹饪指导模块264的烹饪指导等来生成响应。

[0059] 核心处理模块220可以将所确定的响应提供给响应队列或响应高速缓存234。例如,响应高速缓存234可以确保能够以预定义的时间流来显示响应序列。假设对于一个消息,由核心处理模块220确定了不少于两个响应,则对响应的延迟设置可能是必要的。例如,如果用户输入的消息是“你吃早餐了吗?”,则可能确定出两个响应,例如,第一响应是“是的,我吃了面包”,第二响应是“你呢?还觉得饿吗?”。在这种情况下,通过响应高速缓存234,聊天机器人可以确保立即向用户提供第一响应。进而,聊天机器人可以确保以例如1或2秒的时间延迟来提供第二响应,使得第二响应将在第一响应之后的1或2秒被提供给用户。由此,响应高速缓存234可以管理将要发送的响应以及针对每个响应的适当定时。

[0060] 可以将响应队列或响应高速缓存234中的响应进一步传送到UI 210,以便可以在聊天窗口中将响应显示给用户。

[0061] 应当理解,图2中的聊天机器人系统200中所示的所有单元都是示例性的,并且根据具体的应用需求,在聊天机器人系统200中可以省略任何所示的单元并且可以涉及任何其它单元。

[0062] 图3示出了根据实施例的示例性用户界面300。

[0063] 用户界面300被包括在终端设备中,并且可以包括聊天机器人图标310、呈现区域320、控制区域330和输入区域340。聊天机器人图标310可以是表示聊天机器人的照片或图片。呈现区域320显示聊天窗口,该聊天窗口包含用户与聊天机器人之间的会话中的消息和响应。控制区域330包括多个虚拟按钮以由用户用于执行消息输入设置。例如,用户可以通过控制区域330选择进行语音输入、附加图像文件、选择表情符号、进行当前屏幕的截图、启动摄像头等。输入区域340由用户用于输入消息。例如,用户可以通过输入区域340来键入文本。用户界面300还可以包括虚拟按钮350以用于确认发送所输入的消息。如果用户触摸虚拟按钮350,则可以将输入区域340中输入的消息发送到呈现区域320。

[0064] 应当理解,图3中所示的所有单元及其布局都是示例性的。根据具体的应用需求,图3中的用户界面可以省略或添加任何单元,并且图3中的用户界面中的单元的布局也可以以各种方式改变。例如,尽管在呈现区域320中以文本的形式示出消息和响应,但消息和响

应也可以采用语音的形式。从而，聊天机器人和用户可以通过语音聊天。

[0065] 图4示出了根据实施例的AI助手400的示例性硬件结构。AI助手400可以是根据本公开实施例的运行聊天机器人的各种类型的终端设备。在一种实施方式中，可以将AI助手400实施为便携式设备，例如智能电话、智能穿戴产品等，并且可以由用户携带。在一种实施方式中，可以将AI助手400实施为单独的且不可移动的硬件设备，并且被放置在指定的地方处，例如厨房、卧室等。在一种实施方式中，可以将AI助手400集成到例如电脑的其它设备中。在一种实施方式中，AI助手400可以是在几个单独的设备中实施的，其中每个设备执行AI助手400的一部分功能。

[0066] 如图4所示，AI助手400可以包括通信模块410。通信模块410可以使得AI助手400能够基于各种通信技术来接入网络并与其它设备进行通信。例如，通信模块410可以包括用于基于WiFi技术进行通信的WiFi模块412。通信模块410可以包括用于基于蓝牙技术进行通信的蓝牙模块414。虽然未示出，但通信模块410还可以包括用于基于任何其它通信技术进行通信的任何其它模块。

[0067] AI助手400可以包括聊天机器人客户端420。聊天机器人客户端420可以实施聊天机器人的部分或全部功能。从而，AI助手400可以通过聊天机器人客户端420与用户或聊天机器人服务器交互。

[0068] AI助手400可以包括显示器430。显示器430可以由AI助手400用于例如向用户呈现用户界面。

[0069] AI助手400可以包括至少一个处理器440和存储器450。处理器440可以访问存储器450中的数据、执行存储在存储器450中的计算机可执行指令等。例如，当执行计算机可执行指令时，处理器440可以实施聊天机器人客户端420的功能。在一些实施方式中，处理器440可以被配置用于执行根据本公开的实施例的用于提供饮食帮助的方法中涉及的各种过程。

[0070] AI助手400可以包括麦克风460和扬声器470。麦克风460和扬声器470可以用于通过语音与用户交互。此外，AI助手400可以包括至少一个摄像头480。摄像头480可以用于捕获视频或图像，从而，聊天机器人可以检测例如用户的动作或表情以及由摄像头480拍摄的任何其它物品。

[0071] AI助手400可以包括一个或多个控制按钮490。控制按钮490可以是用于控制AI助手400中的模块或功能的物理或虚拟按钮。例如，控制按钮490可以包括用于调大或调小声音的音量控制按钮。

[0072] 应当理解，AI助手400中所示的所有模块都是示例性的，根据实际需求，可以从AI助手400中省略或替换任何模块，并且可以将任何其它模块添加到AI助手400中。例如，在将AI助手400设计用于仅通过语音与用户交互的情况下，甚至可以从AI助手400中省略显示器430。

[0073] 图5示出了根据实施例的用于提供饮食帮助的示例性总体过程500。通过过程500，可以确定用户的饮食意图，并且可以调用相应的饮食帮助过程以获得饮食建议。

[0074] 消息510和会话520可以作为饮食意图分类530的输入。在本文中，消息510可以指用户的一个或多个当前或最新消息，并且会话520可以指在用户和聊天机器人之间的对话的记录，例如用户的会话日志。消息510可以采用文本、语音或图像的形式。由于语音或图像消息可以被转换为相应的文本消息，下面的讨论将以文本消息为例。

[0075] 此外,虽然未示出,但用户的用户简档也可以作为饮食意图分类530的输入。

[0076] 可以执行饮食意图分类530的过程以推断用户的饮食意图。当获得饮食意图时,聊天机器人可以相应地进一步确定对用户的饮食建议。

[0077] 在一种实施方式中,饮食意图可以是饮食知识获取542。例如,在会话期间,如果用户正在咨询糖尿病患者应该怎么吃,则可以将用户的饮食意图推断为获取适合糖尿病患者的食物的饮食知识。从而,聊天机器人可以检索涉及与糖尿病相关的食物类型或食物列表的饮食知识552,以作为对用户的饮食建议。

[0078] 在一种实施方式中,饮食意图可以是饮食推荐544。例如,在会话期间,如果用户正在谈论他已经吃过或将要吃的一些食物或菜肴,或者用户需要对要吃的某些东西的建议,则可以将用户的饮食意图推断为需要饮食推荐。从而,聊天机器人可以确定食物/菜肴名称554,以作为对用户的饮食建议。可以至少考虑用户简档中指示的用户的个性化信息,例如健康状况、口味偏好等,来确定食物/菜肴名称554。例如,如果用户是糖尿病患者,则食物/菜肴名称554可以是“凯撒沙拉”,其适合于糖尿病患者。

[0079] 在一种实施方式中,饮食意图可以是产品推荐546。例如,在会话期间,如果用户对低糖饮料感兴趣或正询问低糖饮料的信息,则可以将用户的饮食意图推断为需要低糖产品的推荐。从而,聊天机器人可以确定产品信息556,以作为对用户的饮食建议。产品信息556可以是一个或多个低糖饮料的名称,例如“零度可乐”、“健怡可乐”等。

[0080] 在一种实施方式中,饮食意图可以是食谱推荐548。例如,在会话期间,如果用户指示他想用香蕉和红薯来烹饪菜肴,则可以将用户的饮食意图推断为需要以香蕉和红薯作为烹饪配料的食谱的推荐。从而,聊天机器人可以检索出或生成食谱558,以作为对用户的饮食建议,其中,食谱558基于香蕉和红薯并且优选地基于用户的偏好。

[0081] 在一种实施方式中,尽管在图5中未示出,饮食意图分类530可以针对用户确定出“无意图”,这指示用户还没有明确的饮食相关意图。在这种情况下,聊天机器人可以以自由聊天的方式向用户提供响应。

[0082] 可以使用支持向量机(SVM)模型来实施饮食意图分类530。SVM模型的输出可以包括一组饮食意图,例如饮食知识获取、饮食推荐、产品推荐、食谱推荐等。应该理解,上述饮食意图是示例性的,并且根据实际需要,可以通过SVM模型确定出更多或更少的饮食意图。

[0083] 在一种实施方式中,可以在SVM模型中使用以下特征来训练<查询,会话日志,用户简档>与一组饮食意图之间的非线性关系,其中,“查询”表示用户的当前消息。

[0084] • 用户简档,包括关于例如性别、年龄、位置、健康状况、烹饪方法偏好、口味偏好等的信息。

[0085] • 词n元项(ngram):查询中词语的一元项(unigram)和二元项(bigram)。例如,查询可以包括“请生成食谱.....”、“给我一个节食的食物建议”、“针对糖尿病,我应该吃什么?”、“我应该吃什么让我变得更强壮?”等。

[0086] • 字符ngram:对于查询中的每个词语,提取字符ngram。例如,在这个模型中使用4元项和5元项。字符ngram对例如中文、日文等亚洲语言是有利的。

[0087] • 词跳元项(skip-gram):对于查询中的所有三元项和4元项,将这些词语中的一个词语替换为例如“*”的符号,以指示存在非连续词语。

[0088] • 布朗(Brown)聚类ngram:Brown聚类用于表示查询中的词语,然后提取一元项和

二元项以作为特征。

[0089] • 词性(POS)标签:将POS标签的存在或不存在用作二元特征。

[0090] • 社交网络相关词语:将查询中的话题标签(hashtag)、表情符号、加长词语和标点符号的数量用作特征。

[0091] • Word2vec聚类ngram:word2vec工具(Mikolov等人,2013)可用于从社交网络数据集中学习100维的词语嵌入。然后,可以采用K-均值算法和词向量的余弦距离来将百万级别的词汇聚类到例如200个类别。这些类别用于表示查询中的归纳的词语。

[0092] • 查询中的、也被包括在疾病词典中的词语及其数量。

[0093] • 查询中的、也被包括在食物词典中的词语及其数量。

[0094] 应该理解,用于SVM模型的以上讨论的特征是说明性的而不是限制性的,并且根据实际需要,SVM模型可以使用更多或更少的特征。

[0095] 图6A示出了根据实施例的示例性聊天窗口610。聊天窗口610示出了根据实施例的用于在会话中提供包括饮食知识、食物/菜肴名称和产品信息饮食建议的示例性过程。

[0096] 当在会话中从用户接收到消息“嗨!”时,聊天机器人可基于纯聊天索引集给出响应“早上好!”。

[0097] 当在会话中接收到消息“你知道糖尿病患者应该怎么吃?”时,聊天机器人可以至少基于该会话和该消息来确定用户的饮食意图是需要糖尿病的饮食知识。聊天机器人可以从该消息中提取饮食信息,例如关心的疾病“糖尿病”。同时,聊天机器人可能已经从用户简档中知道该用户的饮食需求,即,该用户的健康状况与糖尿病相关。然后,至少基于饮食意图、饮食信息和饮食需求,聊天机器人可以获得饮食知识以作为饮食建议,例如“少油少盐”、“确定的进食时间和数量”、“控制吃无糖蛋糕”等。聊天机器人可以向用户提供包括该饮食建议的相应响应。

[0098] 当用户进一步输入消息“有什么推荐的食物?”时,聊天机器人至少可以基于该会话和该消息来确定用户的饮食意图是获取糖尿病患者的饮食和/或产品推荐。聊天机器人可以从该消息和先前的消息中提取饮食信息,例如,关心的疾病“糖尿病”和食物类型“为糖尿病患者推荐的食物”。同时,该用户的饮食需求指示该用户的健康状况与糖尿病相关。然后,至少基于饮食意图、饮食信息和饮食需求,聊天机器人可以获得食物名称或产品信息以作为饮食建议,例如“南瓜(14)”、“苦瓜(13)”、“粗粮(10)”、“黄瓜(9)”等,其中,食物名称后面的数字表示相应食物被所有用户给出正面反馈的频率。聊天机器人可以向用户提供包括该饮食建议的相应响应。该响应也可以包括产品销售信息的链接。例如,如果用户点击或将光标移动到“粗粮(10)”上,则可显示相应的销售信息,例如“商店A”、“每千克5美元”、“地址:xxxx”等。

[0099] 图6B示出了根据实施例的示例性聊天窗口620。聊天窗口620示出了根据实施例的用于在会话中提供包括食物/菜肴名称和产品信息饮食建议的示例性过程。

[0100] 当在会话中从用户接收到消息“我想吃一些炸薯条”时,聊天机器人可以至少基于该会话和该消息来确定用户的饮食意图是需要饮食推荐。聊天机器人可以从该消息中提取饮食信息,例如食物名称“炸薯条”。同时,聊天机器人可能已经从用户简档中知道该用户的饮食需求,即,该用户的健康状况与糖尿病相关。聊天机器人可以基于例如知识图谱来确定“炸薯条”对于糖尿病患者而言不是合适的食物。因此,至少基于饮食意图、饮食信息和饮食

需求,聊天机器人可以获得关于不推荐用户吃炸薯条的饮食建议。例如,聊天机器人可以向用户提供包括该饮食建议的相应响应,例如“太糟糕了!你仍是个糖尿病患者!”。

[0101] 当从用户接收到消息“那么我午餐应该吃什么?”时,聊天机器人可以至少基于该会话和该消息来确定用户的饮食意图是获取饮食推荐。聊天机器人可以从该消息中提取饮食信息,例如食物类型“午餐”。同时,用户的饮食需求指示该用户的健康状况与糖尿病相关。然后,至少基于饮食意图、饮食信息和饮食需求,聊天机器人可以获得菜肴名称以作为饮食建议,例如“凯撒沙拉”,其中,“凯撒沙拉”是适合于糖尿病患者的菜肴。然后,聊天机器人可以向用户提供包括该饮食建议的相应响应,例如“你可以吃凯撒沙拉”。

[0102] 当用户进一步输入消息“有什么推荐的餐馆供应凯撒沙拉么?”时,聊天机器人可以至少基于该会话和该消息确定用户的饮食意图是获取关于凯撒沙拉的产品推荐。聊天机器人可以从该消息中提取饮食信息,例如菜肴名称“凯撒沙拉”。同时,该用户的饮食需求指示该用户的健康状况与糖尿病相关。然后,至少基于饮食意图、饮食信息和饮食需求以及用户简档中指示的用户的位置,聊天机器人可以获得产品信息以作为饮食建议,例如“餐馆A”、“100米远”等。聊天机器人可以向用户提供包括该饮食建议的相应响应,例如“餐馆A。它距你仅100米远”。

[0103] 图7示出了根据实施例的用于建立知识图谱和知识QA对集合的示例性过程700。

[0104] 可以将食物词典702和疾病词典704中的至少一个用于为搜索引擎706提供搜索关键词。食物词典702包括各种食物或菜肴的名称,以下讨论以“食物”为例。疾病词典704包括各种疾病的名称。应该理解,可以使用任何其它类型的词典来提供搜索关键词。

[0105] 可以将食物词典702和疾病词典704中的每个条目提供给搜索引擎706。搜索引擎706可以通过将该条目用作搜索关键词来从网络708抓取相关网页710。附加地或可替换地,网页710也可以由合作伙伴实体712提供。合作伙伴实体712可以是例如可提供网页710或相关数据的生产者、销售者、餐馆等。网页710可能包含至少一种类型的食物、至少一种疾病、或这两者,因此网页710可以被视为包括关于食物和/或疾病的知识。

[0106] 在714处,可以确定网页是否采用QA对样式,例如采用<问题,答案>的形式。如果是,则可以将网页中的QA对添加到知识QA对集合716中。如果不是,即网页采用纯文本形式,则可以在718处对纯文本执行依存句法分析。可以通过718处的依存句法分析来识别纯文本中语句的句法结构,然后,可以在720处从语句的依存树中提取知识元组。知识元组可进一步形成知识图谱722。

[0107] 图8A和图8B示出了根据实施例的示例性依存句法分析。图8A示出了对语句“Non-starchy vegetables such as leafy greens, broccoli, cauliflower, asparagus, and carrots are low in carbohydrates and high in fiber and other nutrients. (非淀粉类蔬菜,例如绿叶蔬菜、西兰花、菜花、芦笋和胡萝卜,碳水化合物含量低且纤维和其它营养素含量高)”的示例性依存句法分析810,图8B示出了对语句“Eating non-starchy vegetables is a good opinion for diabetes mellitus. (吃非淀粉类蔬菜对于糖尿病是好的意见)”的示例性依存句法分析820。通过对语句执行依存句法分析810和820来获得依存树。应该理解,此处可以使用各种依存句法分析技术来执行依存句法分析。

[0108] 可以从语句的依存树中提取知识元组。提取过程可以遵循链接谓词与论元以及链接实体与句法关系的依存弧。

[0109] 假设食物词典702包含实体“non-starchy vegetables (非淀粉类蔬菜)”、“leafy greens (绿叶蔬菜)”、“broccoli (西兰花)”、“cauliflower (菜花)”、“asparagus (芦笋)”和“carrots (胡萝卜)”,并且疾病词典704包含实体“diabetes mellitus (糖尿病)”。然后,可以通过图8A和图8B所示的以下依存弧来链接食物“non-starchy vegetables”和疾病“diabetes mellitus”: $\langle \text{non-starchy vegetables, good opinion, nsubj} \rangle$ 其中“nsubj”表示依存弧的名词性主语;以及 $\langle \text{good opinion, diabetes mellitus, nmod:for} \rangle$ 其中“nmod:for”表示由“good opinion”和“diabetes mellitus”之间的关键词“for”所引导的名词性修饰关系。然后,可以通过将这两个依存弧组合在一起来获得以下元组: $\langle \text{non-starchy vegetables, diabetes mellitus, good opinion} \rangle$ 。

[0110] 以上是采用 $\langle \text{实体, 实体, 关系} \rangle$ 形式的示例性元组,例如, $\langle \text{食物, 疾病, 关系} \rangle$ 。通过这种方式,可以收集纯文本中出现的食物和疾病的各种可能的关系。

[0111] 此外,也可以提取两个都是食物或都是疾病的实体的关系。以食物实体为例,在图8A中示出了以下依存弧: $\langle \text{non-starchy vegetables, leafy greens, nmod:such_as} \rangle$,其中“nmod”表示由关键短语“such as (例如)”引导的名词性修饰关系;以及 $\langle \text{leafy greens, broccoli, conj:and} \rangle$,其中“conj”表示连接并且由关键词“and (和)”所引导。如果“nmod:such_as”依存类型被映射为更一般化的“is_a (是_一种)”关系,则可以获得以下“is_a”关系的元组: $\langle \text{leafy greens, non-starchy vegetables, is a} \rangle$,其意指“leafy greens is a non-starchy vegetable (绿叶蔬菜是一种非淀粉类蔬菜)”;以及 $\langle \text{broccoli, non-starchy vegetables, is a} \rangle$,其意指“broccoli is a non-starchy vegetable (西兰花是一种非淀粉类蔬菜)”。通过这种方式,可以收集纯文本中出现的食物的可能关系。类似地,也可以收集纯文本中出现的疾病的可能关系。

[0112] 所提取的元组可以用于形成或更新知识图谱722。知识图谱722可以是包括与各种疾病有关的知识信息的医学领域知识图谱。知识图谱722可以是包括与各种食物有关的知识信息的食物领域知识图谱。知识图谱722也可以是与疾病和食物两者相关的知识图谱。

[0113] 在一种实施方式中,可以通过链接食物领域知识图谱和医学领域知识图谱来获得与疾病和食物两者相关的知识图谱。该链接可以基于以下启发式规则中的至少一个:

[0114] • 通过使用两个词典而挖掘的知识元组进行链接:例如,关于元组 $\langle \text{non-starchy vegetables, diabetes mellitus, good opinion} \rangle$ 和 $\langle \text{non-starchy vegetables, leafy greens, nmod:such_as} \rangle$,可以基于相同或相似的词语“non-starchy vegetables”而链接这两个元组。

[0115] • 通过在一个语句、一个段落、或者甚至一个文档中出现的一种疾病和一种食物的共现(co-occurrence)频率进行链接。可以基于这些共现来分配不同的相似性分数,使得可以利用这些相似性分数来链接疾病和食物。

[0116] • 通过潜在语义相似性分数进行链接。可以使用疾病领域和食物领域网页来训练联合word2vec模型,然后可以将其用于通过对一对 $\langle \text{疾病, 食物} \rangle$ 的向量表示的点积来计算该对 $\langle \text{疾病, 食物} \rangle$ 的“潜在语义分数”。例如,假设存在元组 $\langle \text{leafy greens, diabetes mellitus, good opinion} \rangle$,并且通过word2vec模型确定“leafy greens”和“broccoli”具有较高的语义相似性,从而可以获得扩展元组 $\langle \text{broccoli, diabetes mellitus, good opinion} \rangle$ 。

[0117] 在724处,可以将知识图谱722中的元组变换为QA对。以元组 $\langle \text{diabetes mellitus,}$

pumpkin,suitale food>(<糖尿病,南瓜,适合的食物>)为例,这个元组描述了作为疾病名称的实体“diabetes mellitus”与作为食物名称的实体“pumpkin”之间的关系“suitale food”。这个元组可以变换为以下问题-答案对:

[0118] 问题=什么是适合糖尿病的食物?答案=南瓜。

[0119] 问题=南瓜是适合糖尿病的食物吗?答案=是的,它是。

[0120] 问题=南瓜是适合何种疾病的食物?答案=糖尿病。

[0121] 以这种方式,可以将一个元组自动变换为采用自然语言的多个QA对。这些自然语言风格的QA对可以用于向用户提供自然语言风格的响应。可以将从知识元组变换的QA对添加到知识QA对集合716中。

[0122] 在给定了用户的查询的情况下,可以使用学习排序(LTR)模型来对知识QA对集合716中的候选答案进行排序。在一些实施方式中,可以采用潜在语义特征来在密集向量空间中对该查询与候选<问题,答案>对进行比较。

[0123] 在一种实施方式中,可以在LTR模型中采用“依存弧匹配”分数。可以对该查询和每个候选<问题,答案>对中的问题或答案两者执行依存句法分析,然后可以比较该查询和该问题或答案的依存弧以获得相似性分数。

[0124] 给定一个查询“Any recommended food for diabetes mellitus?(对于糖尿病有什么推荐的食物?)”,图9A示出了根据实施例的对该查询的示例性依存句法分析910。此外,给定一个候选<问题,答案>对中的问题“Eating what is a good opinion for diabetes mellitus?(吃什么对于糖尿病是好的意见?)”,图9B示出了根据实施例的对该问题的示例性依存句法分析920。

[0125] 该查询和该问题共享相同的依存弧,例如“diabetes-compound-mellitus”,“for-case-mellitus”等,它们是显式的词语和弧的匹配。此外,在该查询和该问题之间也共享相似的依存弧,其是隐式的密集向量空间匹配,例如,该查询中的“food-nmod:for-mellitus”类似于该问题中的“opinion-nmod:for-mellitus”,该查询中的“recommended-amod-food”类似于该问题中的“good-amod-opinion”等。通过将依存弧投射为密集空间向量,然后通过例如余弦函数来计算两个向量的相似性分数,可以在潜在向量空间中执行依存弧的比较。

[0126] 在一种实施方式中,LTR模型可以采用梯度提升决策树(GBDT)算法来对针对查询的候选QA对进行排序,可以在GBDT算法中利用以下特征。

[0127] • 查询和候选问题的依存树之间的隐式/显式依存弧相似性分数。显式依存弧相似性分数是基于依存弧的文本串级别的比较而获得的。隐式依存弧相似性分数是基于依存弧的密集空间向量级别的比较而获得的。

[0128] • 查询和候选答案的依存树之间的隐式/显式依存弧相似性分数。

[0129] • 该用户和其他用户对候选答案的正面反馈的频率。

[0130] • 关于查询q和候选问题Q的用于信息检索的语言模型:给定查询q和候选问题Q,该特征通过以下方程来衡量q和Q之间的相关性:

[0131]
$$P(q|Q) = \prod_{w \in q} [(1-\lambda)P_{ml}(w|Q) + \lambda P_{ml}(w|C)] \quad \text{方程(1)}$$

[0132] 其中, $P_{ml}(w|Q)$ 是从Q估计出词语w的最大似然性, $P_{ml}(w|C)$ 是平滑项,该平滑项被计算为大规模语料库C中的最大似然估计。此处,语料库C可以是知识QA对集合。平滑项避免了零概率,该零概率源于在候选问题Q中出现但没有在查询q中出现的那些词语。 λ 是作为似

然性和平滑项之间的权衡的参数,其中, $\lambda \in (0, 1)$ 。当查询和候选问题之间有多个词语重叠时,该特征的效果较好。

[0133] • 关于查询 q 和候选答案 A 的用于信息检索的语言模型。在这个语言模型中,通过用候选答案 A 代替方程(1)中的 Q ,也使用方程(1)来计算查询和候选答案之间的相似性分数。

[0134] • 关于查询 q 和候选问题 Q 的基于翻译的语言模型。该特征可以从例如问题-答案对中学习词语到词语和/或短语到短语的翻译概率,并且可以将所学习的信息并入最大似然性中。

[0135] 给定查询 q 和候选问题 Q ,基于翻译的语言模型可以被定义为:

$$P_{\text{trb}}(q|Q) = \prod_{w \in q} [(1-\lambda)P_{\text{mx}}(w|Q) + \lambda P_{\text{ml}}(w|C)] \quad \text{方程(2)}$$

$$P_{\text{mx}}(w|Q) = \alpha P_{\text{ml}}(w|Q) + \beta P_{\text{tr}}(w|Q) \quad \text{方程(3)}$$

$$P_{\text{tr}}(w|Q) = \sum_{v \in Q} P_{\text{tp}}(w|v)P_{\text{ml}}(v|Q) \quad \text{方程(4)}$$

[0139] 此处, λ 、 α 和 β 是满足 $\lambda \in (0, 1)$ 且 $\alpha + \beta = 1$ 的参数。 $P_{\text{tp}}(w|v)$ 是从 Q 中的词语 v 到 q 中的词语 w 的翻译概率。

[0140] • 关于查询 q 和候选答案 A 的基于翻译的语言模型。在这个语言模型中,通过在方程(2) - (4)中用候选答案 A 代替 Q ,也使用方程(2) - (4)来计算查询和候选答案之间的相似性分数。

[0141] • 查询与候选问题之间的、词语或字符级别的编辑距离。

[0142] • 查询与候选问题之间的最大子序列比。

[0143] • 将门控递归单元(GRU:gated-recurrent unit)用于编码的递归神经网络(RNN)。该编码将词语序列、短语序列或依存弧序列投射到密集向量空间中,其是该序列的潜在语义表示。可以将查询和候选问题或答案分别提供给RNN-GRU层以获得相应的密集向量,然后可以计算两个密集向量之间的相似性分数。

[0144] 应该理解,GBDT算法中的所有上述特征都是示例性的,在各种实施方式中可以在GBDT算法中采用更多或更少的特征。

[0145] 图10示出了根据实施例的用于提供饮食和产品推荐的示例性过程1000。

[0146] 在1002处,可以在会话中接收至少一个消息。

[0147] 在1004处,可以以上面结合图5中的530处的操作所讨论的方式来执行饮食意图分类。在图10的场景中,可以将用户的饮食意图确定为饮食推荐或产品推荐。

[0148] 在1008处,可以将饮食和产品推荐模型用于提供饮食推荐或产品推荐,例如食物/菜肴名称、产品信息等。饮食和产品推荐模型可以对多个饮食或产品推荐进行排序,并向用户提供候选推荐列表1020。

[0149] 除了将饮食意图分类的输出用作饮食和产品推荐模型的输入之外,饮食和产品推荐模型还可以将用户点击模型1006的输出用作输入。

[0150] 用户点击模型可以预测针对来自用户的查询的候选推荐的点击率(CTR)。用户点击模型的输入可以包括例如用户ID、会话日志、历史点击记录、用户简档、当前查询等,并且用户点击模型的输出是候选推荐的CTR的预测。

[0151] 如图10所示,从“结束”方框1026到用户点击模型1006连接有箭头以传递反馈,因此,可以在线更新用户对候选推荐的点击行为。用户点击模型可以存储用户的历史点击记

录,并且在给定的时间间隔期间(例如每天、每周等)重新训练自己。

[0152] 在一种实施方式中,可以使用以下特征来训练逻辑回归模型以实现用户点击模型。逻辑回归模型是有优势的,因为其容易被训练或重新训练并且具有快速的解码速度。

[0153] • 用户简档,包括关于例如性别、年龄、位置、健康状况、烹饪方法偏好、口味偏好等的信息。

[0154] • 用户的历史点击记录:记录的示例性格式是<用户ID,候选推荐,点击计数>。

[0155] • 词n元项:查询中词语的一元项和二元项。

[0156] • 字符n元项:对于查询中的每个词语,提取字符n元项。例如,在这个模型中使用4元项和5元项。

[0157] • 词跳元项:对于查询中的所有三元项和4元项,将这些词语中的一个词语替换为例如“*”的符号,以指示存在非连续词语。

[0158] • 布朗聚类n元项:布朗聚类用于表示查询中的词语,然后提取一元项和二元项以作为特征。

[0159] • 词性(POS)标签:将POS标签的存在或不存在用作二元特征。

[0160] • 社交网络相关词语:将查询中的话题标签、表情符号、加长词语和标点符号的数量用作特征。

[0161] • Word2vec聚类n元项:word2vec工具(Mikolov等人,2013)可用于从社交网络数据集中学习100维的词语嵌入。然后,可以采用K-均值算法和词向量的L2距离来将百万级别的词汇聚类到例如200个类别。这些类别用于表示查询中的归纳的词语。

[0162] 应该理解,用于逻辑回归模型的上述特征是说明性的而不是限制性的,并且根据实际需要,逻辑回归模型可以使用更多或更少的特征。

[0163] 对饮食和产品推荐模型的输入还可以包括由用户简档模型输出的用户简档1010。用户简档模型可用于根据用户的历史查询来预测用户简档中的性别、年龄和位置信息。训练数据可以是人为创建的。

[0164] 对于性别分类模型,输入是<用户id,查询>并且输出是“男性”或“女性”的标签。可以利用查询中的多个线索,例如,“我妻子不做任何家务活”可能是已婚男子说的,“我丈夫最近很忙”更通常是已婚女子说的等。

[0165] 对于年龄预测模型,输入是<用户id,查询>并且输出是例如“10+”、“20+”、“30+”、“40+”、“50+”或“60+”的标签,其中,“10+”表示年龄在10至20之间,“20+”表示年龄在20至30之间,“30+”表示年龄在30至40之间,等等。年龄预测模型可以基于查询确定年龄信息。例如,如果用户在会话中说“我是高中学生”,则可以确定用户的年龄是“10+”。如果用户在会话中说“我已经退休了”,则可以确定用户很可能是“60+”。

[0166] 对于位置检测模型,输入是<用户id,查询>并且输出可以是位置的至少一个标签。位置检测模型可以基于查询来确定位置信息。例如,如果用户在会话中说“你有什么关于上野(Ueno)附近工作午餐的餐馆的建议吗?”,则可以确定用户在东京的上野附近工作。

[0167] 可以基于训练数据<用户id,查询,目标标签>来训练上述三个分类或检测模型。用于训练的特征可以包括:

[0168] • 查询的目标产品类别或公司:例如,女性可能对化妆品及其相关公司的领域有更强烈的倾向。

[0169] • 用户历史查询中包括的疾病关键词:例如,男性和女性可能具有共同的疾病以及性别敏感的疾病,例如女性的乳房相关疾病和男性的前列腺相关疾病。

[0170] • 用户提到最多的位置相关词语的列表:可将现有的位置词典用于检测用户查询中与位置有关的词语。

[0171] • 词n元项:查询中词语的一元项和二元项。

[0172] • 字符n元项:对于查询中的每个词语,提取字符n元项。例如,在这个模型中使用4元项和5元项。

[0173] • 词跳元项:对于查询中的所有三元项和4元项,将这些词语中的一个词语替换为例如“*”的符号,以指示存在非连续词语。

[0174] • 布朗聚类n元项:布朗聚类用于表示查询中的词语,然后提取一元项和二元项以作为特征。

[0175] • 词性(POS)标签:将POS标签的存在或不存在用作二元特征。

[0176] • 社交网络相关词语:将查询中的话题标签、表情符号、加长词语和标点符号的数量用作特征。

[0177] • Word2vec聚类n元项:word2vec工具(Mikolov等人,2013)可用于从社交网络数据集中学习100维的词语嵌入。然后,可以采用K-均值算法和词向量的L2距离来将百万级别的词汇聚类到例如200个类别。这些类别用于表示查询中的归纳的词语。

[0178] 可以使用上述示例性特征来训练多级支持向量机(SVM)模型。这三个模型可以共享相似的特征模板。在一种实施方式中,对于用户的健康状况或关心的疾病,可以直接使用现有的疾病词典来匹配用户的历史查询,然后可以记录健康状况或疾病及其被提及的频率。此外,用户简档1010可以包括用户的烹饪方法偏好或口味偏好,其也可以是从查询中提取的。

[0179] 对饮食和产品推荐模型的输入还可以包括知识图谱1012。知识图谱1012可以根据图7中的过程700来建立的。

[0180] 对饮食和产品推荐模型的输入还可以包括产品数据集1014和关键词竞价数据1016。产品数据集1014和关键词竞价数据1016可以从合作伙伴实体1018获得的。产品数据集1014可以包括各种产品相关信息,例如产品列表、服务列表、品牌、价格、促销、生产者名称、销售地址等。关键词竞价数据1016可以包括由合作伙伴实体定制的关键词、类别、合作伙伴实体愿意竞价的价格等。

[0181] 虽然未示出,但是对饮食和产品推荐模型的输入还可以包括通过情感分析确定的用户关于候选推荐的历史情感。例如,如果在会话日志中记录有查询“我喜欢吃水果”,这将表明用户对水果的正面意见。

[0182] 可以通过使用LTR模型构建饮食和产品推荐模型。在来自用户的特定查询下,饮食和产品推荐模型可以对可能的推荐进行排序。可以训练GBDT算法以用于排序。在一种实施方式中,可以将以下特征用于GBDT算法:

[0183] • 饮食意图分类的输出

[0184] • 用户简档,包括关于例如性别、年龄、位置、健康状况、烹饪方法偏好、口味偏好等的信息。

[0185] • 用户点击模型的输出:例如,候选推荐的CTR预测。

- [0186] • 关键词竞价数据
- [0187] • 词n元项:查询中词语的一元项和二元项。
- [0188] • 字符n元项:对于查询中的每个词语,提取字符n元项。例如,在这个模型中使用4元项和5元项。
- [0189] • 词跳元项:对于查询中的所有三元项和4元项,将这些词语中的一个词语替换为例如“*”的符号,以指示存在非连续词语。
- [0190] • 布朗聚类n元项:布朗聚类用于表示查询中的词语,然后提取一元项和二元项以作为特征。
- [0191] • 词性(POS)标签:将POS标签的存在或不存在用作二元特征。
- [0192] • 社交网络相关词语:将查询中的话题标签、表情符号、加长词语和标点符号的数量用作特征。
- [0193] • Word2vec聚类n元项:word2vec工具(Mikolov等人,2013)可用于从社交网络数据集中学习100维的词语嵌入。然后,可以采用K-均值算法和词向量的L2距离来将百万级别的词汇聚类到例如200个类别。这些类别用于表示查询中的归纳的词语。
- [0194] • 查询中的话题词语与候选推荐中的词语相比的知识图谱的距离:例如,图8B中的“diabetes mellitus”和“non-starchy vegetables”可以具有1的距离,因为这两个条目可以在一个元组中链接在一起。
- [0195] • 用户关于候选推荐的历史情感:例如,如果在会话日志中存在查询“我喜欢吃水果”,并且当前候选推荐具有材料“苹果”,则该候选推荐可能有相对较高的机会被推荐。
- [0196] 应该理解,GBDT算法的上述特征是说明性的而不是限制性的,并且根据实际需要,GBDT算法可以使用更多或更少的特征。
- [0197] 如上所述,可以由饮食和产品推荐模型输出候选推荐列表1020。
- [0198] 首先,可以向用户提供排序最高的候选推荐。然后,可以在1022处确定是否从用户接收到正面反馈。在本文中,“正面反馈”可以指,例如,来自用户的指示用户喜欢该推荐的消息、用户对候选推荐的点击操作等。如果接收到正面反馈,则过程1000将在1026处结束。如果从用户接收到负面反馈,例如消息“我不喜欢它。换另一个”,则可以在1024处检查列表中是否存在任何其它剩余的候选推荐。如果在1024处为是,则向用户提供下一个排序最高的候选推荐,而如果在1024处为否,则过程1000将在1026处结束。
- [0199] 作为替代方式,如果在1024处为否,则可以向用户通知没有针对该查询的合适的候选推荐。此外,可以基于这样的结果来生成反馈并将其提供给供应产品的合作伙伴实体,使得合作伙伴实体可以获知存在改进其产品列表或开发新产品的可能的需求。
- [0200] 在一种实施方式中,可以将用户对相应候选推荐的点击动作传送到用户点击模型1006,并且相应地,用户点击模型1006可以更新用户的历史点击记录。历史点击记录中的条目的示例可以是来自用户的对候选推荐的正面反馈的数量或者负面反馈的数量。
- [0201] 图11示出了根据实施例的示例性聊天窗口1100。聊天窗口1100示出了根据实施例的用于在会话中提供包括食谱的饮食建议的示例性过程。
- [0202] 当在会话中接收到消息“我想用香蕉和红薯烹饪”时,聊天机器人可以至少基于该会话和该消息来确定用户的饮食意图是获取与香蕉和红薯相关的食谱推荐。聊天机器人可以从该消息中提取饮食信息,例如烹饪配料“香蕉”和“红薯”。聊天机器人可以尝试检索出

在数据库中被索引的现有食谱以提供饮食建议,但是在该示例中,不存在与烹饪配料“香蕉”和“红薯”两者都相关的经索引的食谱。因此,聊天机器人可以向用户提供响应,例如“听起来很棒。我检索了,但没有找到一起使用这两种材料的任何现有的菜肴。你想让我为你生成一个新菜肴吗?”

[0203] 当在会话中接收到消息“太好了。我更喜欢甜味,给我个建议”时,聊天机器人可以至少基于该会话和该消息来确定用户的饮食意图是获取与香蕉和红薯有关的食谱推荐。聊天机器人可以进一步至少基于该会话和该消息来识别饮食需求,例如口味偏好“甜味”。应该理解,在一些情况下,聊天机器人还可以从用户的用户简档中识别饮食需求,该用户简档可以包括用户的口味偏好或烹饪方法偏好。然后,至少基于饮食意图、先前提取的饮食信息“香蕉”和“红薯”以及饮食需求,聊天机器人可以获得作为饮食建议的食谱,其中,该食谱是聊天机器人用烹饪配料“香蕉”和“红薯”以口味偏好“甜味”来新生成的。例如,图11示出了由聊天机器人生成的响应中的食谱,其包括8个烹饪步骤。此外,聊天机器人可以进一步向用户提供基于该食谱所生成的菜肴的图像。

[0204] 当在会话中接收到消息“我喜欢这个,在我开始烹饪时指导我”时,聊天机器人可以开始指导用户遵循该食谱进行烹饪。聊天机器人可以提供响应,例如,“乐意效劳,现在让我们开始第一步:.....我将在你处理这些材料时观察”。同时,聊天机器人可以通过至少一个摄像头来监视用户的动作和烹饪配料的烹饪状态。

[0205] 当识别出用户停止了动作并且根据第一步准备好了烹饪配料时,聊天机器人可以与用户确认并开始指导第二步。这样,聊天机器人可以最终指导用户基于该食谱来烹饪出菜肴。

[0206] 图12示出了根据实施例的用于生成食谱的示例性过程1200。

[0207] 在1202处,可以获得当前会话和/或至少一个最近的消息,例如如图11所示的“我想用香蕉和红薯烹饪”、“你想让我为你生成一个新菜肴吗?”、“太好了。我更喜欢甜味,给我个建议”等。

[0208] 在1204处,可以以上面结合图5中的530处的操作所讨论的方式来执行饮食意图分类。在图12的场景下,可以将用户的饮食意图确定为食谱推荐。

[0209] 在1206处,可以从所述当前会话和/或所述至少一个消息中提取关键词。关键词可以包括例如烹饪配料“香蕉”和“红薯”的饮食信息以及例如口味偏好“甜味”的饮食需求。

[0210] 在1208处,可以对所提取的关键词执行语义关键词扩展。在一种实施方式中,可以采用预先建立的话题知识图谱来用于关键词扩展。例如,可以将关键词“香蕉”语义扩展为“拔丝香蕉”、“焦糖香蕉”等,并且可以将关键词“红薯”语义扩展为“蜜汁红薯”等。

[0211] 在1210处,可以基于扩展的关键词来检索现有的或经索引的食谱。例如,可以检索出针对“焦糖香蕉”的现有食谱1212和针对“蜜汁红薯”的现有食谱1214。聊天机器人可以预先从网络收集各种现有食谱。

[0212] 在1216处,可以组合所检索出的现有食谱1212和1214以生成新食谱1218。

[0213] 在一种实施方式中,食谱中的步骤可以分成几个阶段,例如,包括:阶段1“配料准备和预处理”;阶段2“对主要配料进行烹饪”,其中主要配料可以是例如香蕉和红薯;阶段3“对辅助配料进行烹饪”,其中辅助配料可以是例如糖;阶段4“对主要配料和辅助配料两者进行烹饪”;阶段5“后处理”;等等。上述阶段是示例性的,根据实际需求,可以为食谱划分更

多或更少的阶段。

[0214] 食谱1212中的步骤1至2与食谱1214中的步骤1可以对应于阶段1,食谱1212中的步骤3至4与食谱1214中的步骤2至3可以对应于阶段2,食谱1212中的步骤5与食谱1214中的步骤4可以对应于阶段3,食谱1212中的步骤6至7与食谱1214中的步骤5至6可以对应于阶段4,食谱1212中的步骤8与食谱1214中的步骤7可以对应于阶段5。

[0215] 可以将支持向量机(SVM)用于训练“阶段”分类模型,阶段分类模型可以使用例如从网络收集的食谱的三元项字符和二元项词语特征。此外,可以将这5个阶段人为地注释到食谱以构建训练数据。

[0216] 在一种实施方式中,可以将不同食谱中的类似步骤链接在一起。如果除了命名实体以外,两个步骤的语句非常相似,则可以进行将这些命名实体链接在一起的简单组合,以便获得确切地指示一起处理这些命名实体的新语句。例如,食谱1212中的步骤6与食谱1214中的步骤5是相似的,除了不同的命名实体“香蕉块”和“红薯块”之外。这两个步骤可以组合在一起以作为新食谱1218中的步骤6,该步骤6指示一起处理“香蕉块”和“红薯块”。

[0217] 在一种实施方式中,可以从食谱中识别不同的步骤或阶段。当一个步骤专用于一个食谱时,例如食谱1212中的步骤2,可以将该步骤简单地插入到新食谱的相应阶段中,例如新食谱1218的阶段1“配料准备和预处理”。

[0218] 图13示出了根据实施例的示例性文本到图像生成模型1300。模型1300可以用于基于相应的食谱来生成菜肴的图像。可以将所生成的图像提供给用户,使得用户可以对聊天机器人所建议的食谱具有直观的认知。

[0219] 在模型1300的整个生成过程中可以采用生成对抗网络(GAN)结构。GAN结构包括两个生成器和两个鉴别器,表示为阶段I生成器、阶段I鉴别器、阶段II生成器和阶段II鉴别器。阶段I生成器通过根据给定文本草绘对象的大致形状和基本颜色、然后根据随机噪声向量描绘背景颜色,来绘制低分辨率图像。然后,阶段II生成器通过再次以阶段I的结果和文本输入作为条件,来生成具有照片般逼真细节的高分辨率图像。

[0220] 在传统的GAN中,训练过程类似于采用以下目标函数进行优化的双人最小-最大(min-max)游戏:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}} [\log D(x)] + \mathbb{E}_{z \sim p_z} [\log(1 - D(G(z)))], \quad \text{方程 (5)}$$

[0222] 其中,D和G分别表示鉴别器和生成器。G被优化为通过生成使鉴别器D难以与实际图像区分的图像来再现真实数据分布 p_{data} 。同时,D被优化为区分基于 p_{data} 的实际图像与由G生成的合成图像。

[0223] 在方程(5)中,x是遵循真实数据分布 p_{data} 的实际图像,z是从例如均匀分布或一维/多维高斯分布的分布 p_z 中采样的噪声向量。D(x)是在给定x作为D的输入的情况下D输出的概率,D(G(z))是D分配给由G生成的z的概率。D的任务是使得V(D,G)中的分数最大化,V(D,G)用于“让实际数据尽可能接近1,并让合成数据尽可能接近0”。此外,G的任务是尽可能地“欺骗”D,即,使D的益处最小化。

[0224] 容易通过附加额外的条件变量c来扩展G和D,产生G(x,c)和D(x,c)。该公式允许G生成以变量c为条件的图像。例如,图12中的食谱1218的文本输入是一种类型的c。

[0225] 如图13所示,条件文本描述 t 首先由编码器编码,产生文本嵌入 ϕ_t 。可以将GRU和细粒度的文本到情感网络用于获得该文本嵌入 $\phi_t = [\text{GRU}_t, \text{情感}_t]$ 。此处的一个难点在于以文本嵌入向量为条件的潜在空间通常具有高维度。由于有限的数量,其通常会导致潜在数据流形中的不连续性,这对学习生成器是有害的。如图13所示,可以将条件增强模块用于为生成器产生更多的条件变量。即,从独立的高斯分布 $N(\mu(\phi_t), \Sigma(\phi_t))$ 中采样出更多的潜在变量,其中,平均 $\mu(\phi_t)$ 和对角线协方差矩阵 $\Sigma(\phi_t)$ 是文本嵌入 ϕ_t 的函数。这将在给定少量图像-文本对的情况下导致更多训练对,因为引入了更多的语句,这些语句共享与原始文本输入相比的相似的语义含义。假设原始文本输入是“那个女孩大约4岁,日美混血,粉红色衣服且甜美的微笑”,所引入的更多语句可以包括例如“那个4岁女孩是日美混血,穿着粉红色衬衫并且在笑”的语句。可以将标准高斯分布 $N(0, 1)$ 用于强化在条件流形上的平滑性以避免过度拟合。此处可使用KL发散度来计算:

$$[0226] \quad D_{\text{KL}}(N(\mu(\phi_t), \Sigma(\phi_t)) \parallel N(0, I)) \quad \text{方程(6)}$$

[0227] “条件增强”的输出是 c ,其被发送到阶段I生成器。通常,阶段I GAN通过使 L_{D_0} 最大化并使 L_{G_0} 最小化来训练 D_0 和 G_0 :

$$[0228] \quad \mathcal{L}_{D_0} = \mathbb{E}_{(I_0, t) \sim p_{\text{data}}} [\log D_0(I_0, \varphi_t)] + \mathbb{E}_{z \sim p_z, t \sim p_{\text{data}}} [\log(1 - D_0(G_0(z, c_0), \varphi_t))],$$

$$[0229] \quad \mathcal{L}_{G_0} = \mathbb{E}_{z \sim p_z, t \sim p_{\text{data}}} [\log(1 - D_0(G_0(z, c_0), \varphi_t))] + \lambda D_{\text{KL}}(N(\mu_0(\varphi_t), \Sigma_0(\varphi_t)) \parallel N(0, I)), \quad \text{方程(7)}$$

[0230] 此处,文本描述 t 和实际图像 I_0 来自真实数据分布 p_{data} , I_0 是图像的图像向量和情感向量的组合,类似于 $[I_{\text{图像}}, \text{情感向量}]$ 。如上所述, z 是噪声向量。 λ 是控制 L_{G_0} 中两个项之间平衡的正则化参数。例如,可以将 $\lambda=1$ 用于该模型。从 $N(\mu_0(\phi_t), \Sigma_0(\phi_t))$ 采样出高斯条件变量 c_0 以反映文本描述。

[0231] 对于阶段I生成器,将情感文本嵌入 ϕ_t 馈送到全连接层中以针对高斯分布 $N(\mu_0(\phi_t), \Sigma_0(\phi_t))$ 生成 μ_0 和 σ_0 ,其中, σ_0 是 Σ_0 的对角线中的值。向量 c_0 是通过 $c_0 = \mu_0 + \sigma_0 * e$ 计算的,其中,*是按元素相乘, e 遵循 $N(0, 1)$ 。然后,将 c_0 与 N_z 维噪声向量级联以通过一系列上采样组块生成 $W_0 \times H_0$ 图像。

[0232] 对于鉴别器,文本嵌入 ϕ_t 首先被利用全连接层而压缩到 N_d 维,然后进行空间复制以形成 $M_d \times M_d \times N_d$ 张量。同时,图像被馈送通过一系列下采样组块,直到其具有 $M_d \times M_d$ 空间维度。然后,将图像过滤映射沿着通道维度与文本张量级联。得到的张量被进一步馈送到 1×1 卷积层,以在图像和文本上联合地学习特征。最后,使用具有一个节点的全连接层来产生决策分数。

[0233] 阶段II GAN可以以与上述阶段I GAN类似的方式构建。阶段II GAN构建在阶段I GAN之上,以生成照片般逼真的高分辨率图像。阶段II GAN不仅以前一阶段生成的低分辨率图像为条件,还再次以文本嵌入为条件,以纠正阶段I结果中的缺陷,并促使模型提取文本中先前可能忽略的信息以生成更多照片般逼真的细节。

[0234] 以低分辨率样本 s_0 和高斯潜在变量 c 为条件,通过交替地使 L_D 最大化和使 L_G 最小化来训练阶段II GAN中的鉴别器 D 和生成器 G :

$$\begin{aligned}
 [0235] \quad \mathcal{L}_D &= \mathbb{E}_{(I,t) \sim p_{data}} [\log D(I, \varphi_t)] + \\
 &\quad \mathbb{E}_{s_0 \sim p_{G_0}, t \sim p_{data}} [\log(1 - D(G(s_0, c), \varphi_t))], \\
 [0236] \quad \mathcal{L}_G &= \mathbb{E}_{s_0 \sim p_{G_0}, t \sim p_{data}} [\log(1 - D(G(s_0, c), \varphi_t))] + \\
 &\quad \lambda D_{KL}(\mathcal{N}(\mu(\varphi_t), \Sigma(\varphi_t)) \parallel \mathcal{N}(\mathbf{0}, I)),
 \end{aligned}$$

方程 (8)

[0237] 此处, $s_0 = G_0(z, c_0)$ 是由阶段 I GAN 生成的。图 13 中的“残余组块”使用与现有 ResNet 相同的残余功能和组块。

[0238] 对模型 1300 的一种扩展是考虑用于从粗到精的文本到图像生成的 GAN 序列。借助更多的 GAN, 可以预期利用 GAN 组块来关注更多细节并生成更高分辨率的图像。例如, 在图 13 中, 可以以与从阶段 I 到阶段 II 的扩展相类似的方式来扩展阶段 III GAN 和更多阶段 GAN。

[0239] 根据本公开的实施例, 聊天机器人可以对烹饪视频流执行视频捕获以获得相应的文本和语音。例如, 通过视频捕获, 可以生成文本集合以描述烹饪视频中正在播放的内容, 例如视频中人的动作的文本描述、视频中配料的烹饪状态的文本描述等。最后, 该文本集合可以形成可作为饮食建议提供给用户的食谱。此外, 可以通过各种文本到语音技术, 基于该文本集合生成语音, 并且因此可以针对由该文本集合形成的食谱来准备语音指导。烹饪视频流可以是在线视频流或离线准备的视频文件。

[0240] 图 14 示出了根据实施例的用于视频捕获的示例性过程 1400。视频捕获可以基于视频到文本类型的编码-解码算法, 其中, 可以将 2D/3D 卷积神经网络 (CNN) 作为编码部分, 并且可以将 GRU 类型的 RNN 网络作为解码部分。

[0241] 在 1402 处, 可以执行视频到向量编码过程, 通过该过程可以获得输入视频的向量表示。在一种实施方式中, 输入视频的帧可以是根据 2D CNN (例如 AlexNet、GoogleNet、VGG、ResNet 等) 来编码的, 并且输入视频的片段可以由 3D CNN 中的卷积 3D 滤波器编码的。可以组合 2D CNN 和 3D CNN 的输出以获得输入视频的向量表示。

[0242] 在 1404 处, 可以执行身体关节检测, 以识别用户在输入视频中的动作。图 15A 示出了根据实施例的身体关节的示例性表示 1510。基本上, 存在 14 个重要的身体关节来呈现或识别完整的身体形态。在图 15A 中示出了总共 14 个读取点, 用于表示实际身体部位的身体关节, 例如头顶、头的底部、左肩点和右肩点、两个肘部、两个手臂、两个臀部点、两个膝部和两个脚踝。应该理解, 14 个身体关节不一定全部出现在给定照片中。例如, 可能在半身照片中找到少于 14 个身体关节, 并且对于身体的侧面轮廓, 大多数对称身体关节是相同的, 例如两个肩部点变成一个。

[0243] 可以训练深度学习模型以从视频流中自动检测 14 个身体关节。图 15B 示出了根据实施例的用于身体关节检测的示例性过程 1520。在 1522 处, 可以输入原始视频帧。在 1524 处, 可以采用现有的 N-最优 (N-Best) 方法来为每个单帧生成多种姿态假设的集合。在 1526 处, 根据在 1524 处生成的 N-最优姿态假设, 针对每个帧中的每个身体部位生成实际身体部位假设, 并且传播到相邻帧。在 1528 处, 将实际身体部位组合为抽象身体部位, 并且将实际身体部位假设也组合为抽象身体部位假设, 以便去除帧内简单循环。在 1530 处, 根据在 1528 处生成的抽象身体部位假设, 针对包括单个身体部位和联接身体部位的抽象身体部位来生成轨迹段 (tracklet)。在 1532 处, 进行基于树的姿态估计。例如, 可以构建姿态假设图, 其中每个节点是对应于抽象身体部位的径迹段, 并且可以通过从姿态假设图中为身体部位选择

最佳假设来获得最佳姿态估计。

[0244] 在1406处,可以对1402处生成的输入视频的向量表示和1404处的身体关节检测的向量表示进行组合以获得级联向量。

[0245] 可以通过关注机制1408来将该级联向量提供给GRU-RNN解码器1410。GRU-RNN解码器1410可以被配置用于将该级联向量解码为文本1412。从而,实现了从输入视频到文本的转换。

[0246] 在一种实施方式中,过程1400可以进一步包括通过各种文本到语音技术来将文本1412变换为语音1414。

[0247] 图16示出了根据实施例的获得用于视频捕获的训练数据的示例性过程1600。

[0248] 在1602处,可以获得烹饪相关的视频和相应的语音。例如,可以从烹饪网站、电视节目等收集视频和语音。然后,可以从在1602处获得的视频和相应的语音中分离出视频部分1604和语音部分1606。

[0249] 对于视频部分1604,可以在1608处执行身体关节检测,以便识别视频部分中的身体关节并获得相邻身体关节之间的连接线。

[0250] 对于语音部分1606,可以执行语音到文本转换以获得相应的文本1612。

[0251] 在1614处,可以用输入视频、在1610处识别的身体关节、以及文本1612来获得训练数据,并且可以以(视频,身体关节,文本)的形式表示训练数据。该训练数据可以用于训练根据本公开实施例的视频捕获过程。

[0252] 如上所述,根据本公开实施例的聊天机器人可以指导用户遵循食谱来进行烹饪。在一个方面,聊天机器人可以在烹饪指导中采用结合图15A和图15B讨论的身体关节检测方法。例如,在将聊天机器人集成在诸如面向家庭的智能语音盒的终端设备中时,当用户在聊天机器人的指导下进行烹饪时,可以使用终端设备的摄像头来实时拍摄视频,并且通过身体关节检测,可以捕获并识别用户在视频中的动作。因此,聊天机器人可以获知用户是否遵循食谱以适当的方式进行烹饪。在另一方面,聊天机器人可以从摄像头拍摄的视频中识别配料的烹饪状态,例如,识别准备好的配料的大小、配料是否已被烹饪好等。在另一个方面,在烹饪指导期间,聊天机器人可以基于食谱来生成语音,或者如果语音1414对应于食谱,则直接使用图14中的语音1414。

[0253] 图17示出了根据实施例的用于在会话中向用户提供饮食帮助的示例性方法1700的流程图。

[0254] 在1710处,可以在会话中接收至少一个消息,该会话处于所述用户和电子对话代理之间。

[0255] 在1720处,可以基于所述会话和所述至少一个消息来推断与所述用户相关联的饮食意图。

[0256] 在1730处,可以从所述至少一个消息中提取饮食信息。

[0257] 在1740处,可以至少基于所述会话和所述至少一个消息来识别所述用户的饮食需求。

[0258] 在1750处,可以至少基于所述饮食意图、所述饮食信息和所述饮食需求来获得对于所述用户的饮食建议。

[0259] 在1760处,可以生成包括所述饮食建议的响应。

[0260] 在1770处,可以将所述响应提供给所述用户。

[0261] 在一种实施方式中,所述饮食意图可以指示食谱推荐;所述饮食信息可以包括至少一种烹饪配料;所述饮食需求可以包括烹饪方法偏好和口味偏好中的至少一个;并且所述饮食建议可以包括食谱。

[0262] 所述识别所述饮食需求可以包括以下中的至少一个:从所述会话和所述至少一个消息中提取所述烹饪方法偏好和口味偏好中的至少一个;以及从用户简档中检索所述烹饪方法偏好和口味偏好中的至少一个,所述用户简档是至少基于所述会话和所述至少一个消息来建立的。

[0263] 所述获得所述饮食建议可以包括:检索至少一个经索引的食谱,所述至少一个经索引的食谱与所述至少一种烹饪配料以及所述烹饪方法偏好和口味偏好中的至少一个相关。

[0264] 所述获得所述饮食建议可以包括:对所述至少一种烹饪配料进行语义扩展以获得至少一种扩展的烹饪配料;基于所述至少一种扩展的烹饪配料以及所述烹饪方法偏好和口味偏好中的至少一个来检索至少一个经索引的食谱;以及基于所述至少一个经索引的食谱来生成食谱,所生成的食谱与所述至少一种烹饪配料以及所述烹饪方法偏好和口味偏好中的至少一个相关。

[0265] 方法1700还可以包括:至少基于所述用户的动作和/或所述至少一种烹饪配料的烹饪状态,指导所述用户遵循所述食谱进行烹饪。

[0266] 在一种实施方式中,所述饮食意图可以指示饮食推荐和产品推荐中的至少一个;所述饮食信息可以包括食物类型、食物名称、菜肴名称和关心的疾病中的至少一个;所述饮食需求可以包括所述用户的健康状况;并且所述饮食建议可以包括食物名称、菜肴名称和产品信息中的至少一个。

[0267] 所述识别所述饮食需求可以包括以下中的至少一个:从所述会话和所述至少一个消息中提取所述健康状况;以及从用户简档中检索所述健康状况,所述用户简档是至少基于所述会话和所述至少一个消息来建立的。

[0268] 所述获得所述饮食建议可以包括:基于用户简档、对多个候选推荐的点击率、所述多个候选推荐的关键词竞价信息、与饮食和/或疾病相关的知识图谱、关于所述多个候选推荐的历史情感以及产品数据集中的至少一个,来对所述多个候选推荐进行排序;以及选择一个或多个排序最高的候选推荐作为所述饮食建议。

[0269] 在一种实施方式中,所述饮食意图可以指示饮食知识获取;所述饮食信息可以包括食物类型、食物名称、菜肴名称和关心的疾病中的至少一个;所述饮食需求可以包括所述用户的健康状况;所述饮食建议可以包括饮食知识。

[0270] 所述获得所述饮食建议可以包括:至少基于所述饮食意图、所述饮食信息和所述饮食需求,从知识图谱或知识问题-答案对集合中检索所述饮食知识。

[0271] 在一种实施方式中,所述至少一个消息可以包括文本消息、语音消息和图像消息中的至少一个。

[0272] 应当理解,方法1700还可以包括根据上述本公开实施例的用于在会话中向用户提供饮食帮助的任何步骤/过程。

[0273] 图18示出了根据实施例的用于在会话中向用户提供饮食帮助的示例性装置1800。

[0274] 装置1800可以包括：消息接收模块1810，用于在所述会话中接收至少一个消息，该会话处于所述用户和电子对话代理之间；饮食意图推断模块1820，用于基于所述会话和所述至少一个消息来推断与所述用户相关联的饮食意图；饮食信息提取模块1830，用于从所述至少一个消息中提取饮食信息；饮食需求识别模块1840，用于至少基于所述会话和所述至少一个消息来识别所述用户的饮食需求；饮食建议获得模块1850，用于至少基于所述饮食意图、所述饮食信息和所述饮食需求来获得对于所述用户的饮食建议；响应生成模块1860，用于生成包括所述饮食建议的响应；以及响应提供模块1870，用于将所述响应提供给所述用户。

[0275] 在一种实施方式中，所述饮食意图可以指示食谱推荐；所述饮食信息可以包括至少一种烹饪配料；所述饮食需求可以包括烹饪方法偏好和口味偏好中的至少一个；并且所述饮食建议可以包括食谱。

[0276] 饮食需求识别模块1840可以进一步用于以下中的至少一个：从所述会话和所述至少一个消息中提取所述烹饪方法偏好和口味偏好中的至少一个；以及从用户简档中检索所述烹饪方法偏好和口味偏好中的至少一个，所述用户简档是至少基于所述会话和所述至少一个消息来建立的。

[0277] 饮食建议获得模块1850可以进一步用于：对所述至少一种烹饪配料进行语义扩展以获得至少一种扩展的烹饪配料；基于所述至少一种扩展的烹饪配料以及所述烹饪方法偏好和口味偏好中的至少一个来检索至少一个经索引的食谱；以及基于所述至少一个经索引的食谱来生成食谱，所生成的食谱与所述至少一种烹饪配料以及所述烹饪方法偏好和口味偏好中的至少一个相关。

[0278] 装置1800还可以包括：烹饪指导模块，用于至少基于所述用户的动作和/或所述至少一种烹饪配料的烹饪状态，指导所述用户遵循所述食谱进行烹饪。

[0279] 在一种实施方式中，所述饮食意图可以指示饮食推荐和产品推荐中的至少一个；所述饮食信息可以包括食物类型、食物名称、菜肴名称和关心的疾病中的至少一个；所述饮食需求可以包括所述用户的健康状况；并且所述饮食建议可以包括食物名称、菜肴名称和产品信息中的至少一个。

[0280] 在一种实施方式中，所述饮食意图可以指示饮食知识获取；所述饮食信息可以包括食物类型、食物名称、菜肴名称和关心的疾病中的至少一个；所述饮食需求可以包括所述用户的健康状况；所述饮食建议可以包括饮食知识。

[0281] 此外，装置1800还可以包括根据上述本公开实施例的被配置用于在会话中向用户提供饮食帮助的任何其它模块。

[0282] 图19示出了根据实施例的用于在会话中向用户提供饮食帮助的示例性装置1900。

[0283] 装置1900可以包括一个或多个处理器1910以及存储了计算机可执行指令的存储器1920。当执行所述计算机可执行指令时，一个或多个处理器1910可以：在所述会话中接收至少一个消息，该会话处于所述用户和电子对话代理之间；基于所述会话和所述至少一个消息来推断与所述用户相关联的饮食意图；从所述至少一个消息中提取饮食信息；至少基于所述会话和所述至少一个消息来识别所述用户的饮食需求；至少基于所述饮食意图、所述饮食信息和所述饮食需求来获得对于所述用户的饮食建议；生成包括所述饮食建议的响应；以及将所述响应提供给所述用户。

[0284] 本公开的实施例可以实施在非暂时性计算机可读介质中。该非暂时性计算机可读介质可以包括指令,当所述指令被执行时,使得一个或多个处理器执行根据上述本公开实施例的用于在会话中向用户提供饮食帮助的方法的任何操作。

[0285] 应当理解,以上描述的方法中的所有操作都仅仅是示例性的,本公开并不限制于方法中的任何操作或这些操作的顺序,而是应当涵盖在相同或相似构思下的所有其它等同变换。

[0286] 还应当理解,以上描述的装置中的所有模块都可以通过各种方式来实施。这些模块可以被实施为硬件、软件、或其组合。此外,这些模块中的任何模块可以在功能上被进一步划分成子模块或组合在一起。

[0287] 已经结合各种装置和方法描述了处理器。这些处理器可以使用电子硬件、计算机软件或其任意组合来实施。这些处理器是实施为硬件还是软件将取决于具体的应用以及施加在系统上的总体设计约束。作为示例,本公开中给出的处理器、处理器的任意部分、或者处理器的任意组合可以实施为微处理器、微控制器、数字信号处理器(DSP)、现场可编程门阵列(FPGA)、可编程逻辑器件(PLD)、状态机、门逻辑、分立硬件电路、以及配置用于执行在本公开中描述的各种功能的其它适合的部件。本公开给出的处理器、处理器的任意部分、或者处理器的任意组合的功能可以实施为由微处理器、微控制器、DSP或其它适合的平台所执行的软件。

[0288] 软件应当被广泛地视为表示指令、指令集、代码、代码段、程序代码、程序、子程序、软件模块、应用、软件应用、软件包、例程、子例程、对象、运行线程、过程、函数等。软件可以驻留在计算机可读介质中。计算机可读介质可以包括例如存储器,存储器可以例如为磁性存储设备(如,硬盘、软盘、磁条)、光盘、智能卡、闪存设备、随机存取存储器(RAM)、只读存储器(ROM)、可编程ROM(PROM)、可擦除PROM(EPROM)、电可擦除PROM(EEPROM)、寄存器或者可移动盘。尽管在本公开给出的多个方面中将存储器示出为是与处理器分离的,但是存储器也可以位于处理器内部(如,缓存或寄存器)。

[0289] 以上描述被提供用于使得本领域任何技术人员可以实施本文所描述的各个方面。这些方面的各种修改对于本领域技术人员是显而易见的,本文限定的一般性原理可以应用于其它方面。因此,权利要求并非旨在被局限于本文示出的方面。关于本领域技术人员已知或即将获知的、对本公开所描述各个方面的元素的所有结构和功能上的等同变换,都将通过引用而明确地包含到本文中,并且旨在由权利要求所覆盖。

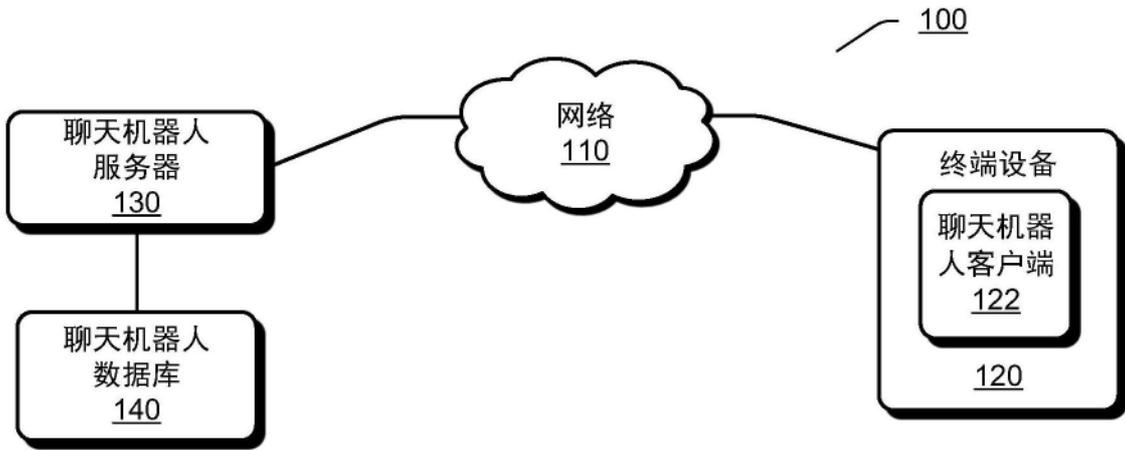


图1

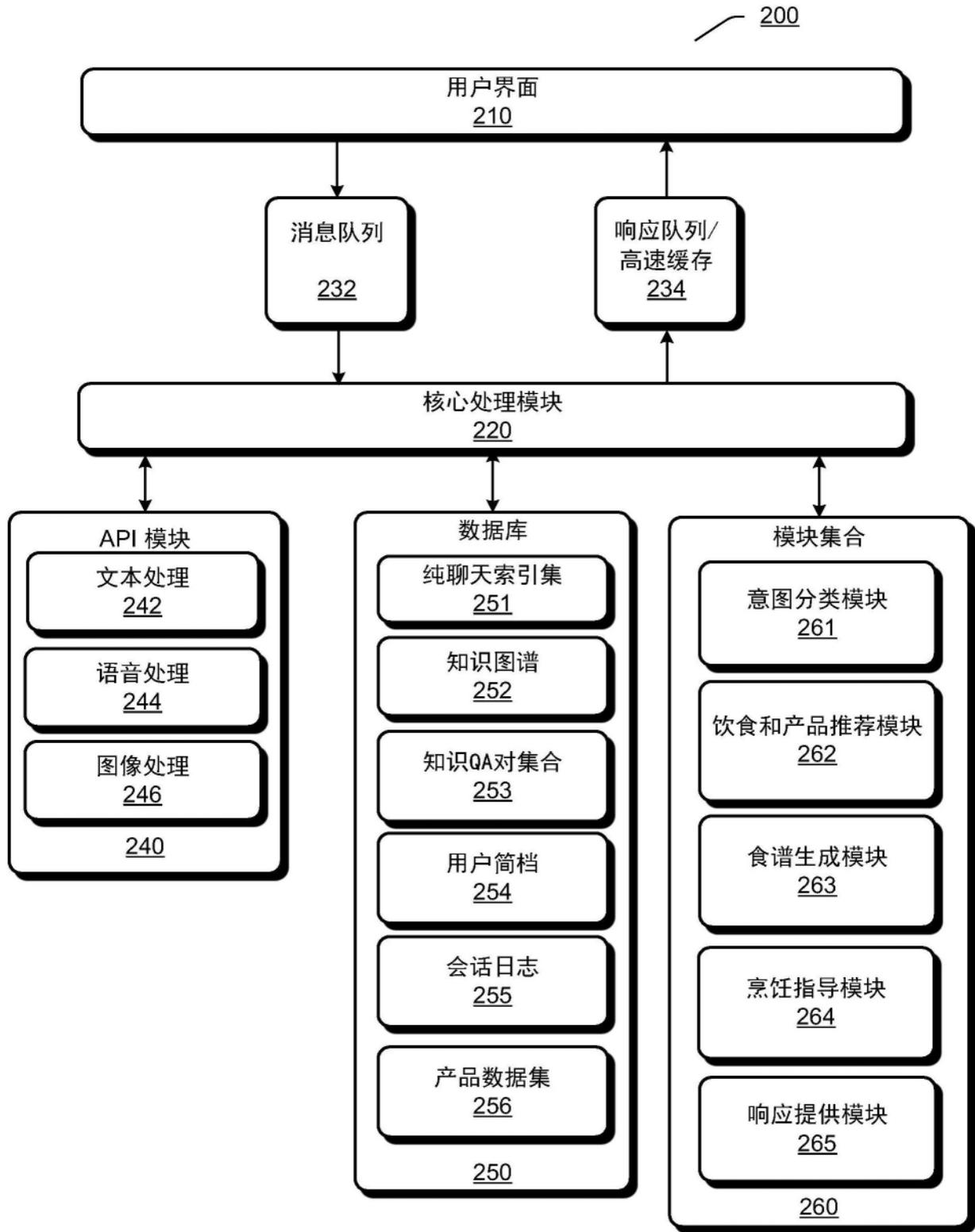


图2

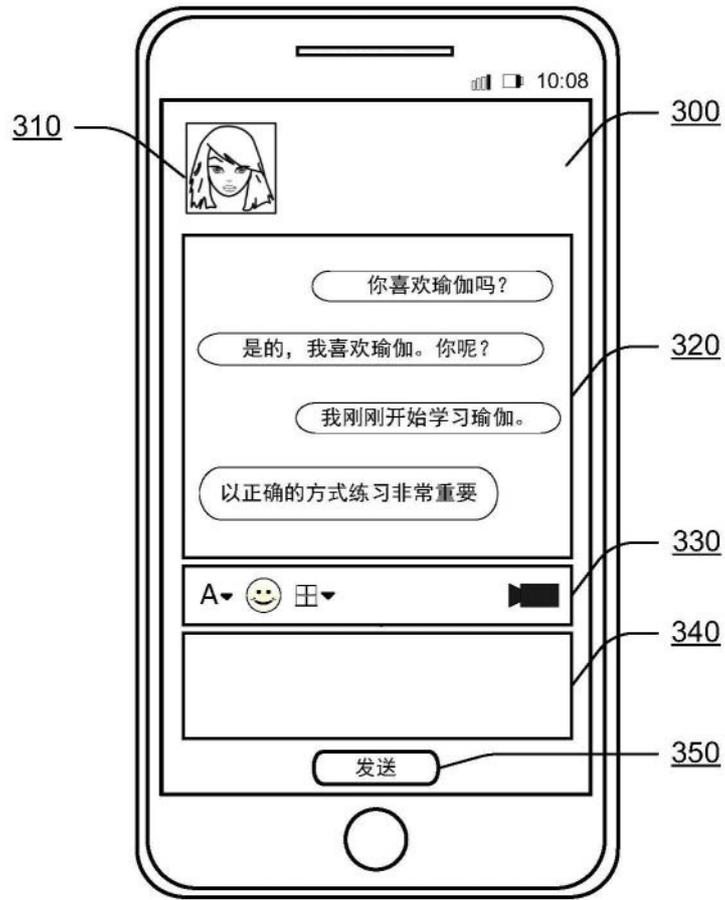


图3

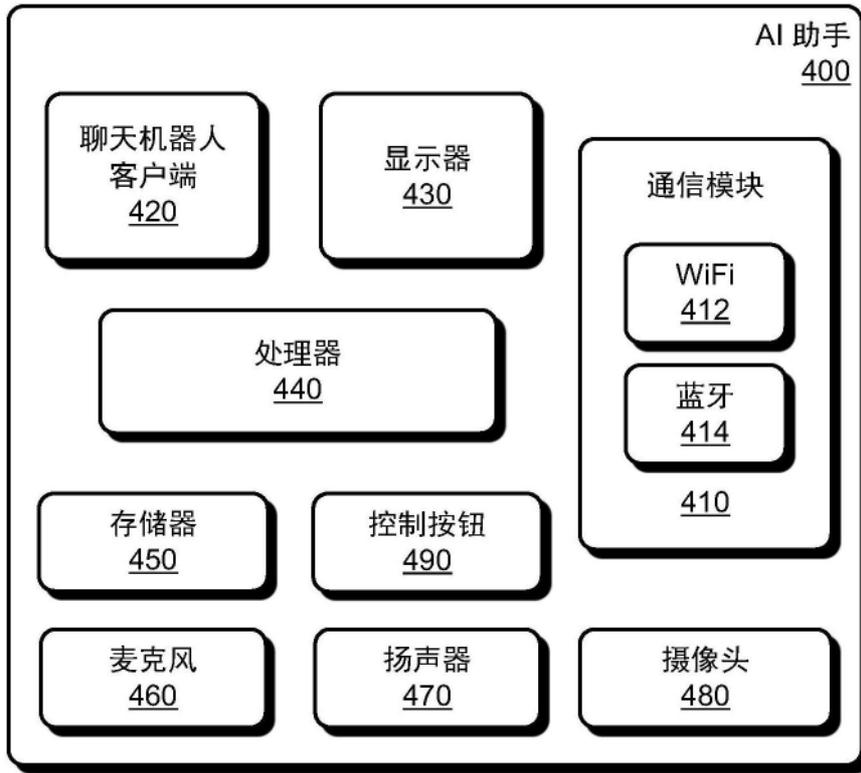


图4

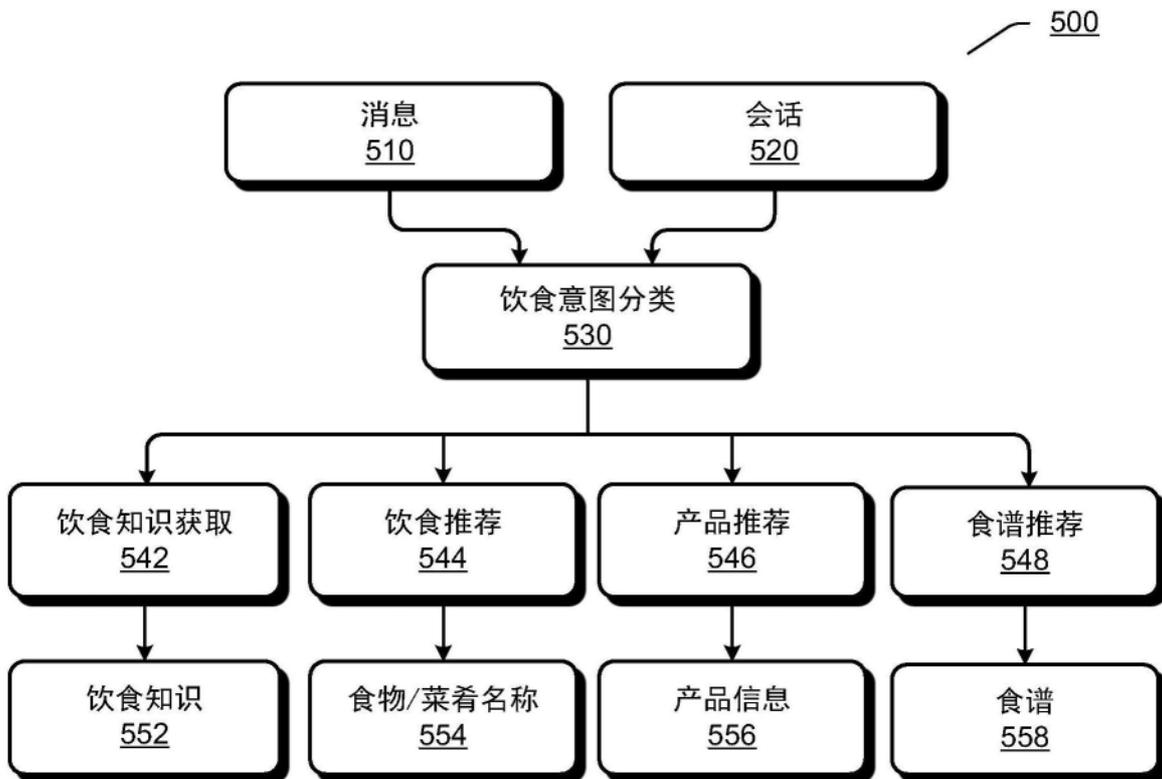


图5

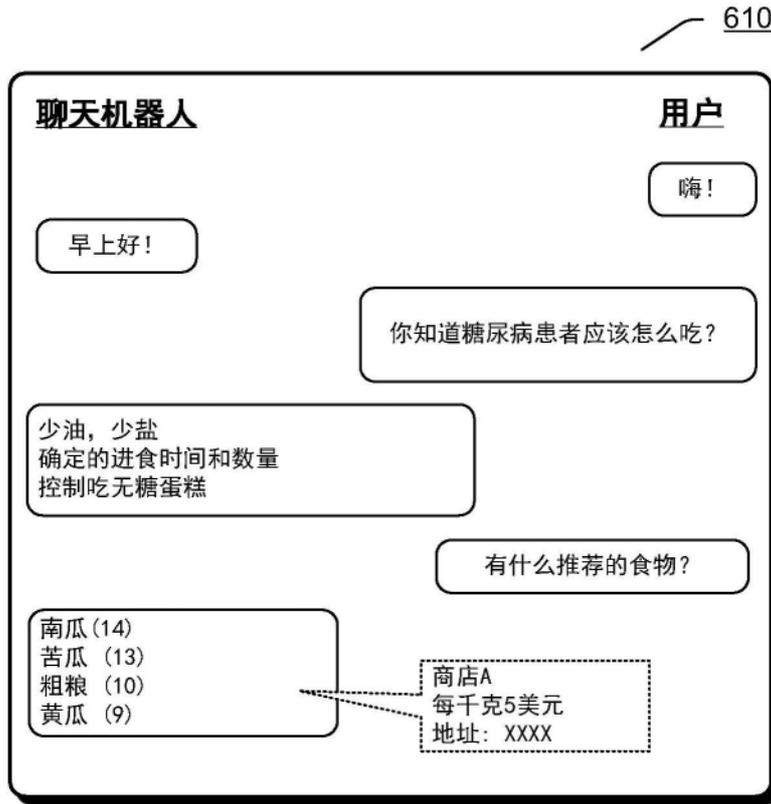


图6A

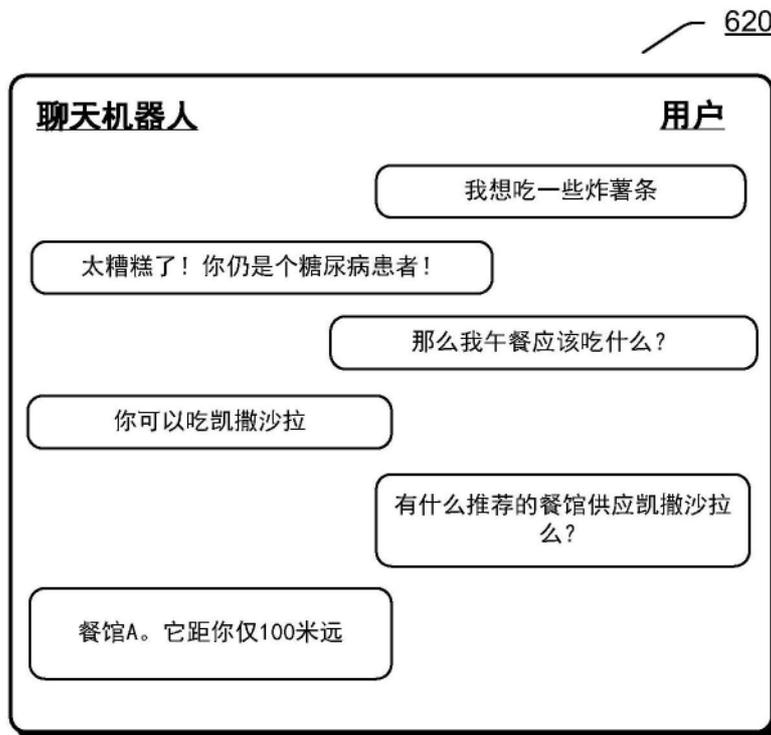


图6B

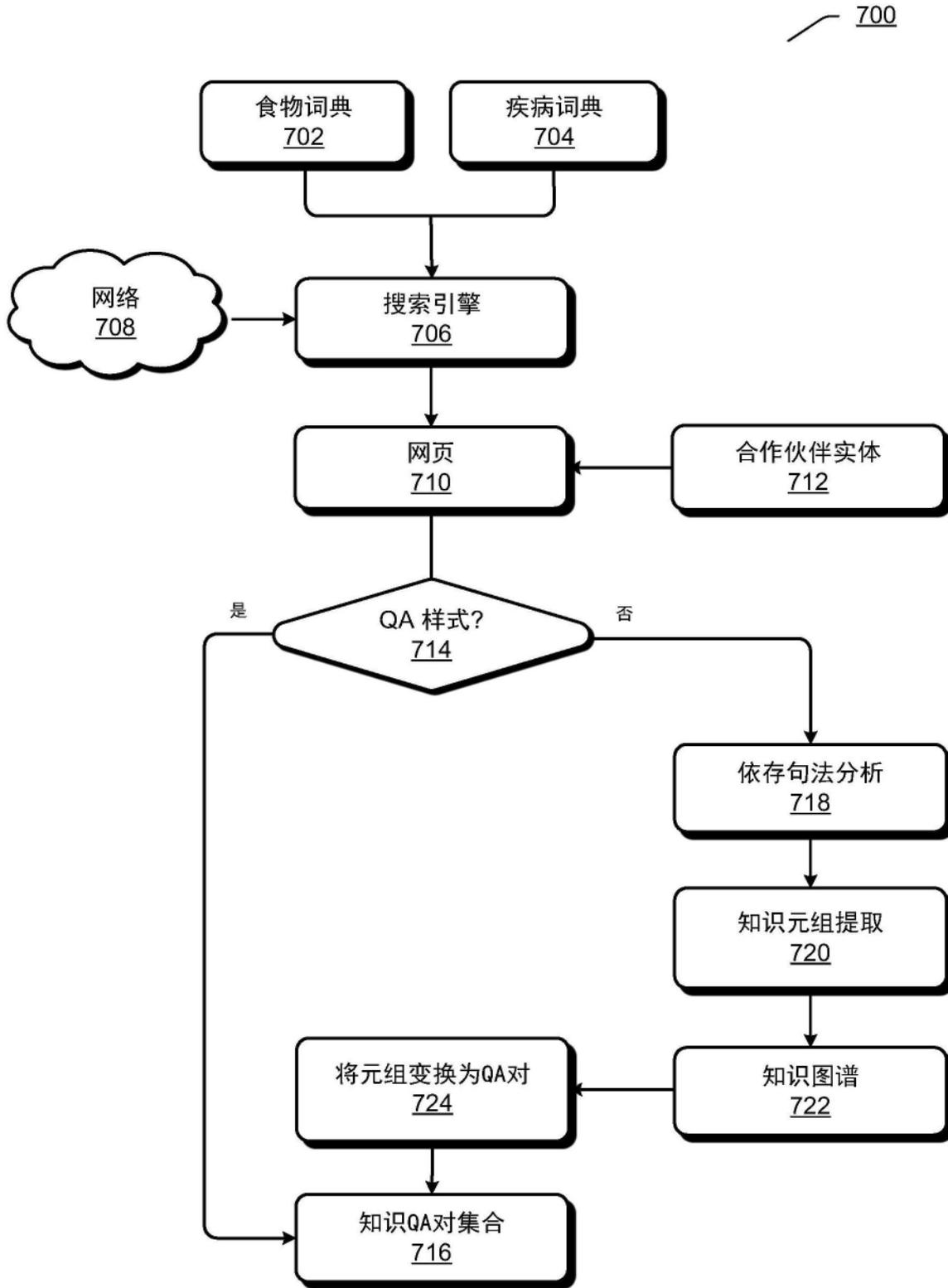


图7

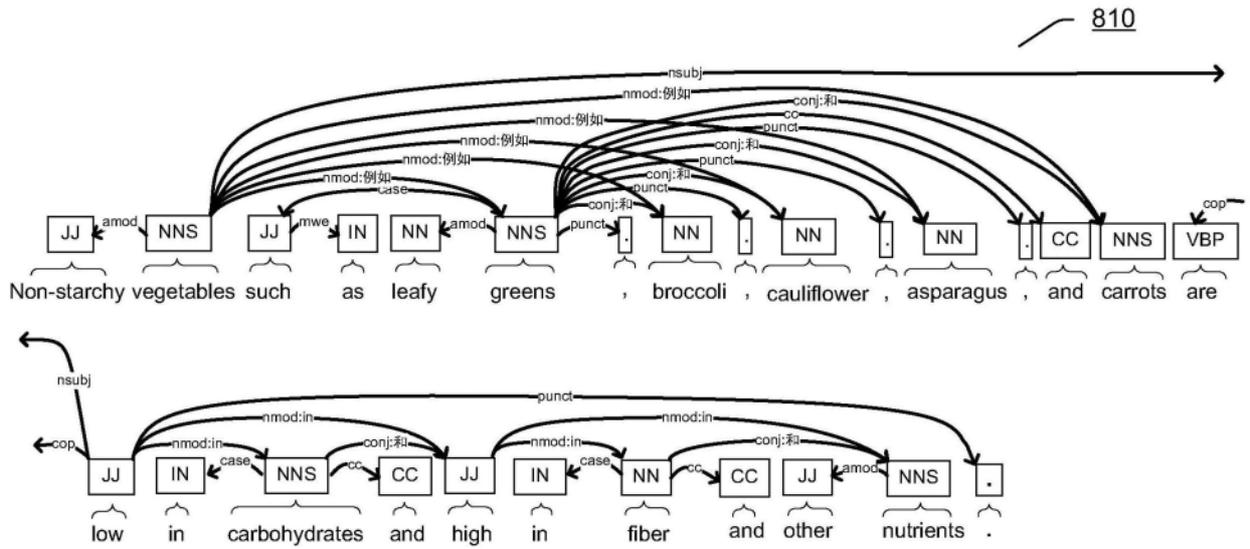


图8A

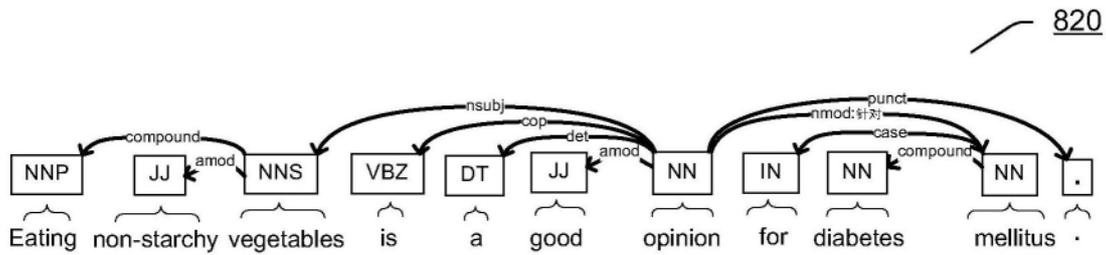


图8B

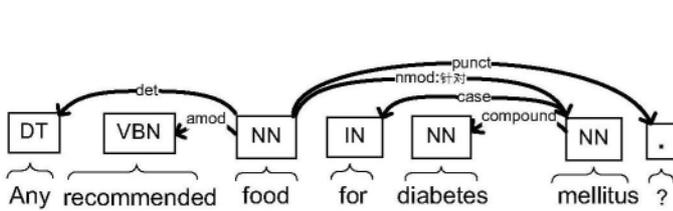


图9A

920

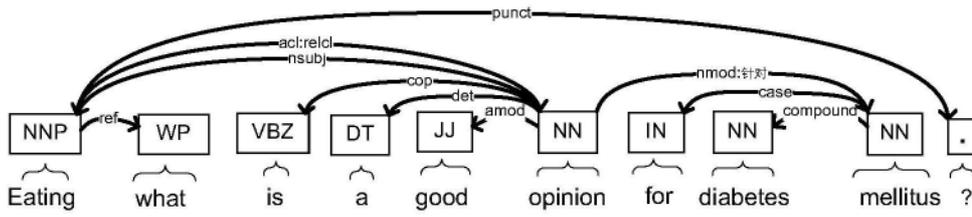


图9B

1000

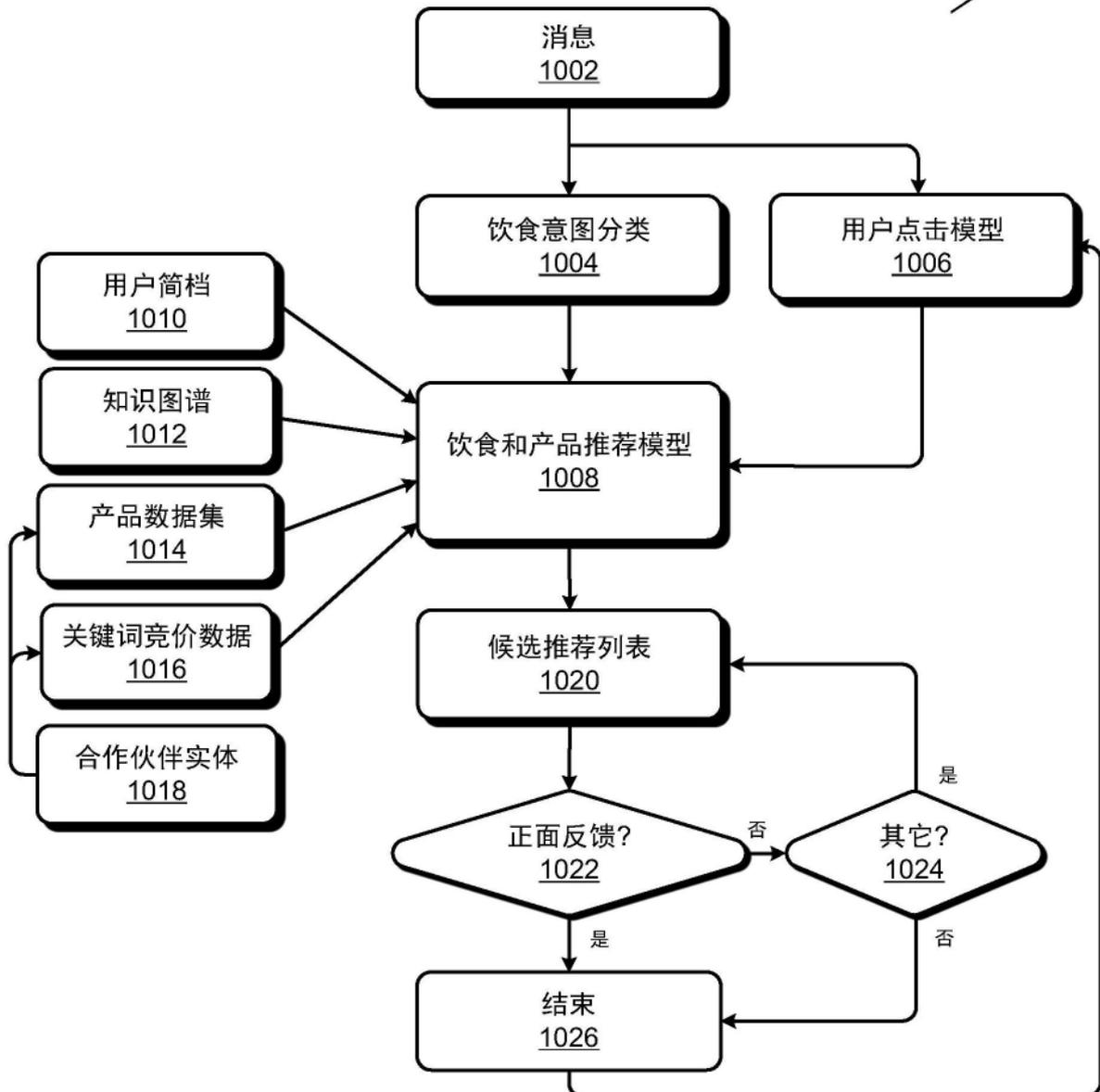


图10

1100

聊天机器人
用户

我想用香蕉和红薯烹饪

听起来很棒。我检索了，但没有找到一起使用这两种材料的任何现有的菜肴。你想让我为你生成一个新菜肴吗？

太好了。我更喜欢甜味，给我个建议

这是食谱：
 1. 将香蕉剥皮并切成2cm左右的块；将红薯去皮并切成3cm左右的块，除去红薯外面的水；
 2. 将香蕉块浸入鸡蛋和淀粉混合浆中；
 3. 在油锅中炸至颜色变为金黄；
 4. 取出并淋油；
 5. 锅中放油加热并放入糖；
 6. 当糖完全融化并准备好生“丝”时，将香蕉块和红薯块放入锅中；
 7. 较充分地搅打以便这些块完全裹上糖；
 8. 取出材料并撒上芝麻。
 这是自动生成的图像：



我喜欢这个，在我开始烹饪时指导我。

乐意效劳，现在让我们开始第一步：
 将香蕉剥皮并切成2cm左右的块；将红薯去皮并切成3cm左右的块，除去红薯外面的水；
 我将在你处理这些材料时观察。

太好了

[然后用户开始执行第1步的动作]

[几分钟后，当用户停止他的/她的动作时]

看起来你完成了第1步，这些块切的大小正好。

谢谢。下面做什么？

第二步是：
 将香蕉块浸入鸡蛋和淀粉混合浆中；

图11

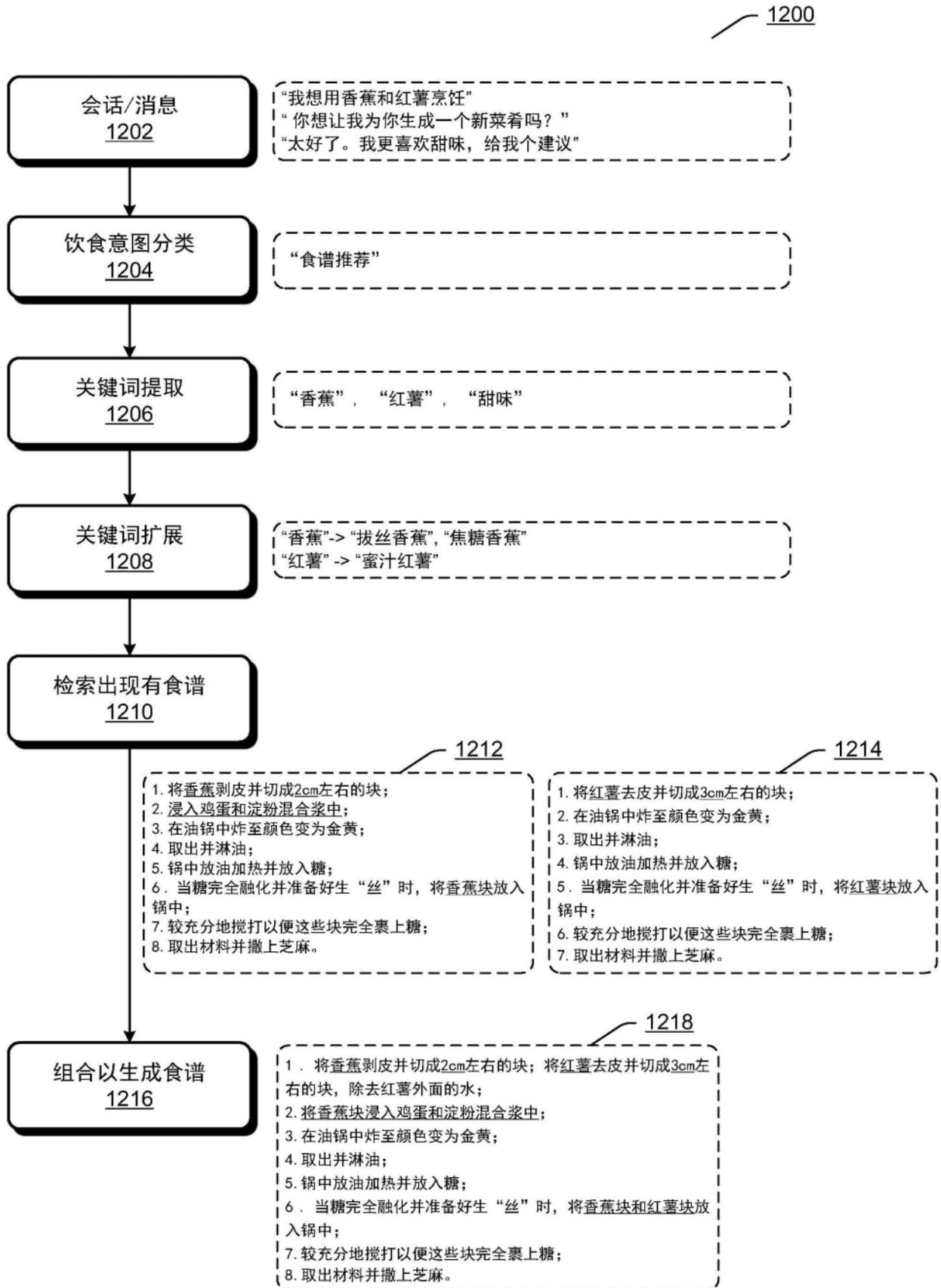


图12

1300

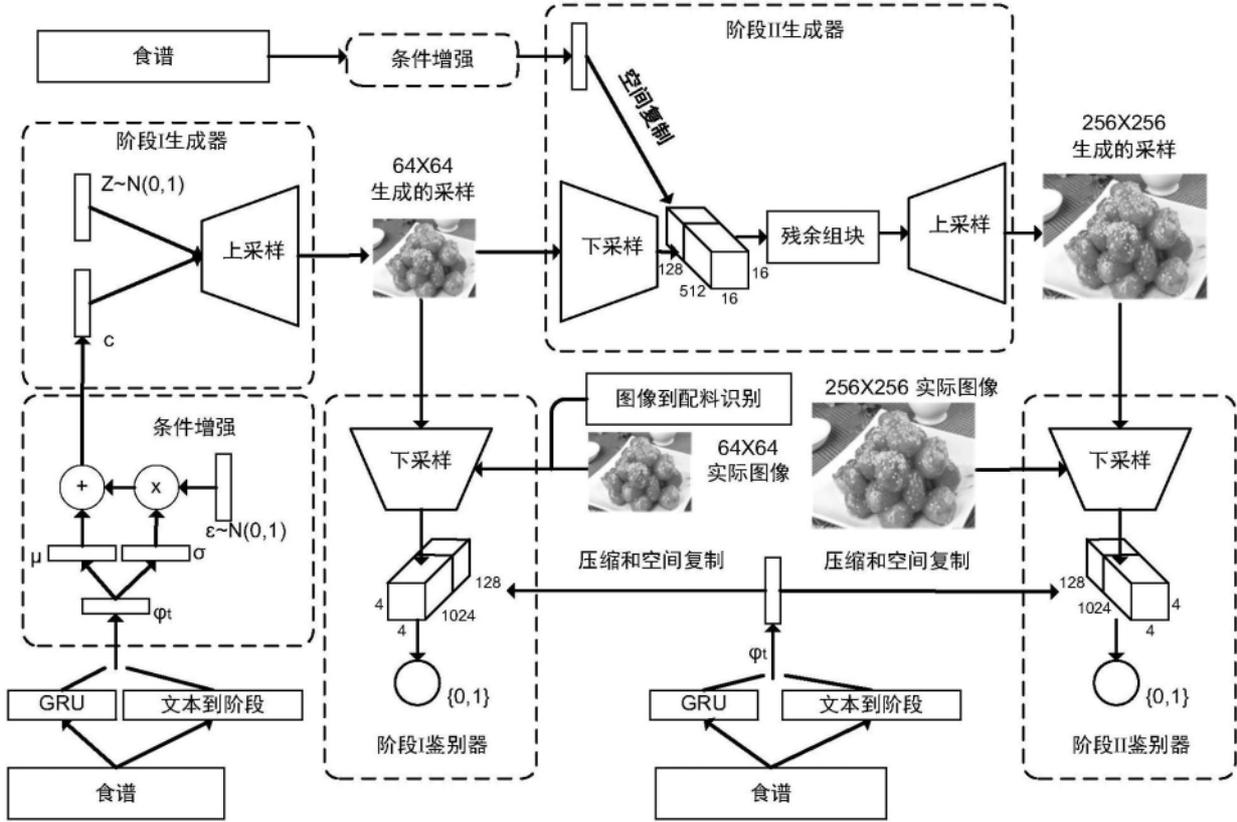


图13

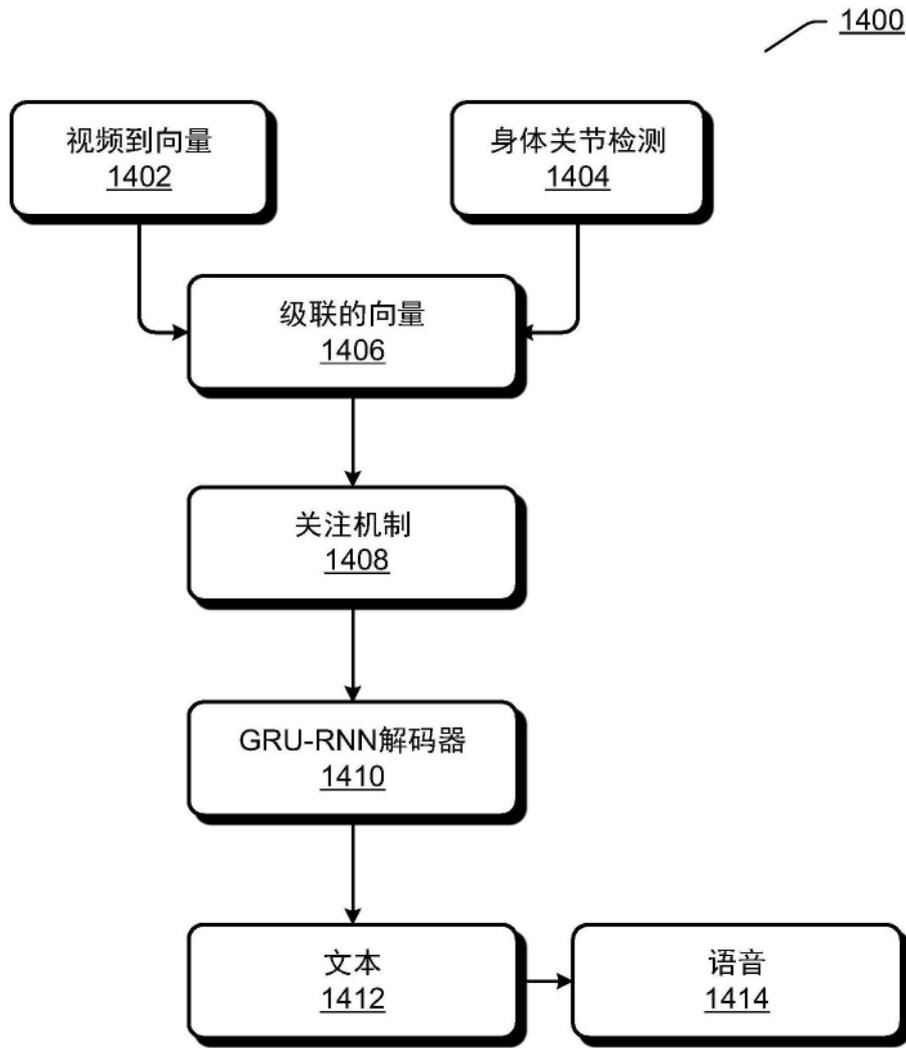


图14

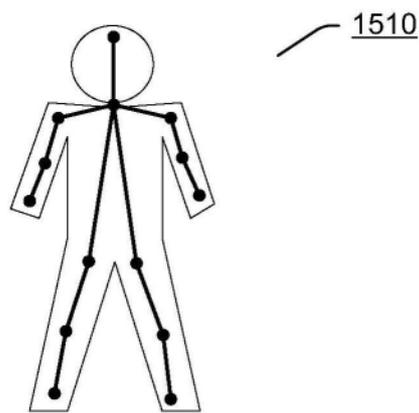


图15A

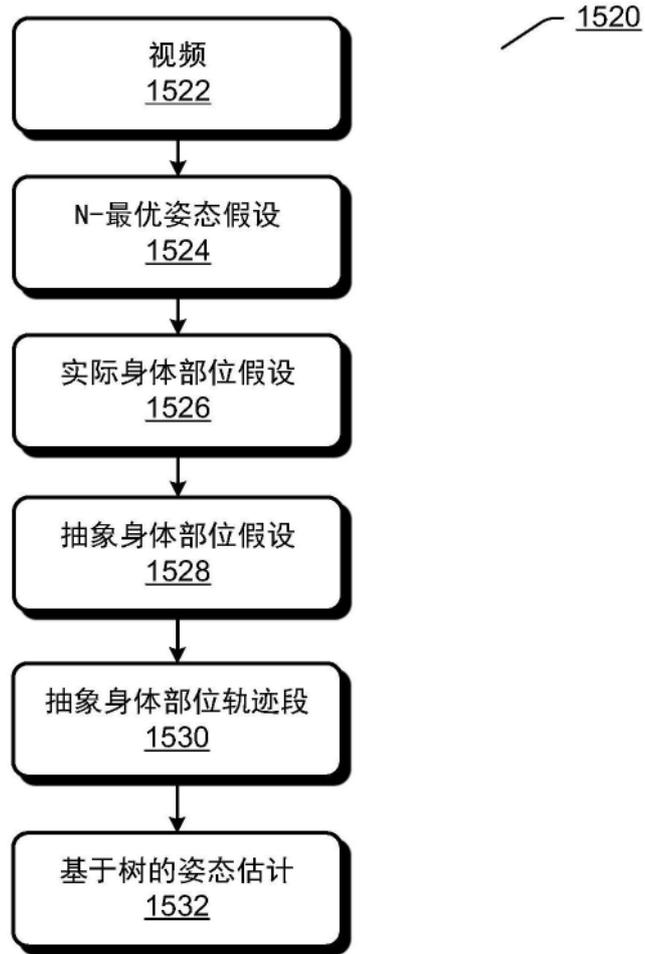


图15B

1600

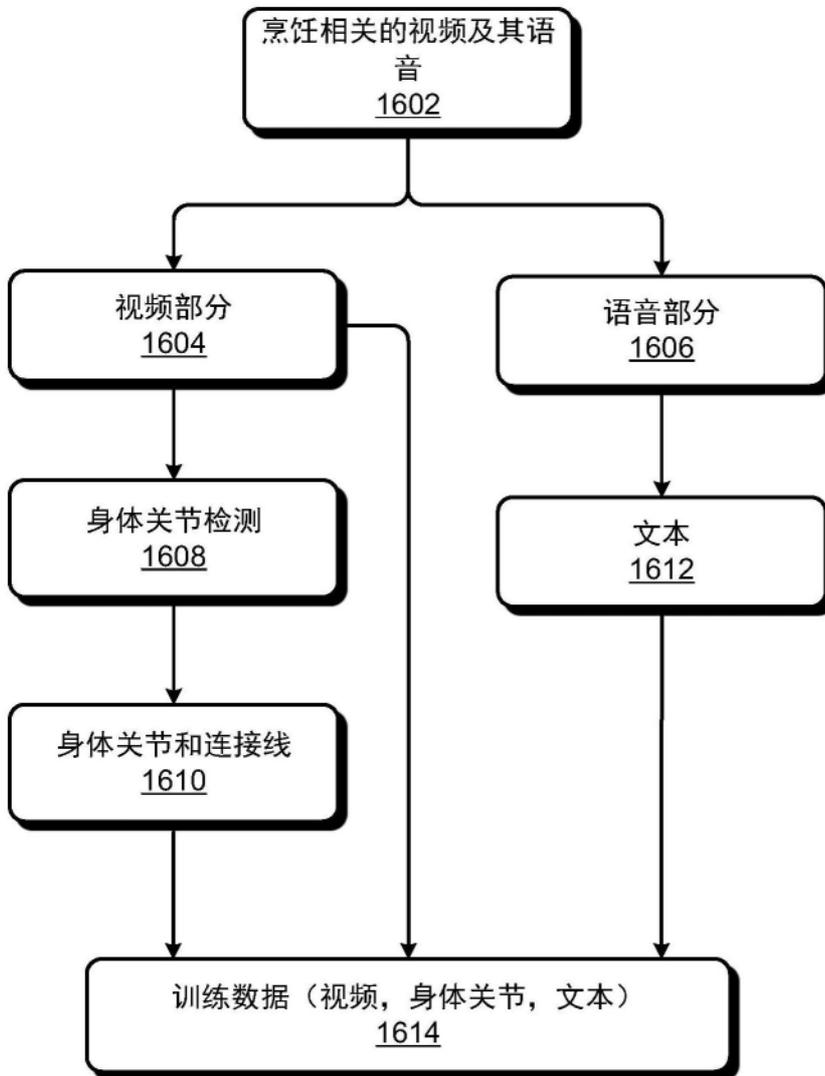


图16

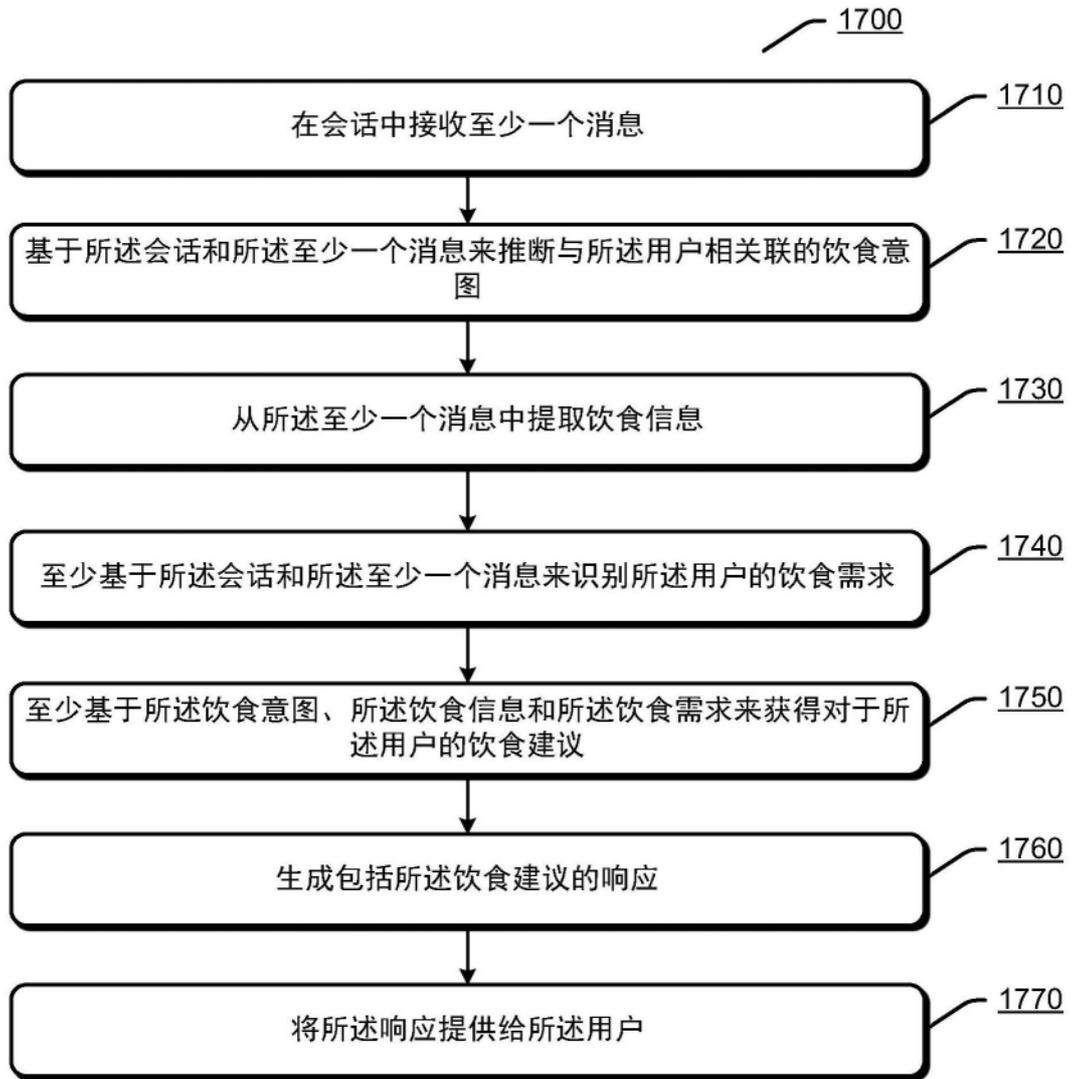


图17



图18



图19