



(12) 发明专利

(10) 授权公告号 CN 112005299 B

(45) 授权公告日 2023. 12. 22

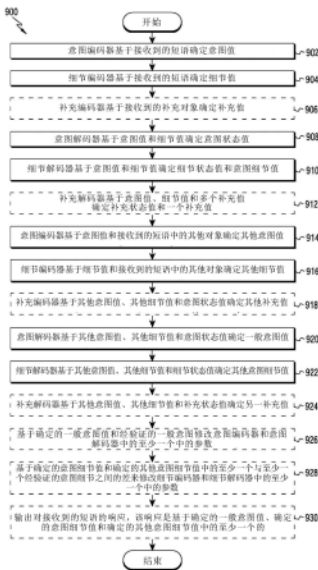
(21) 申请号 201980027126.9
(22) 申请日 2019.04.25
(65) 同一申请的已公布的文献号
 申请公布号 CN 112005299 A
(43) 申请公布日 2020.11.27
(30) 优先权数据
 62/662,610 2018.04.25 US
 16/390,241 2019.04.22 US
(85) PCT国际申请进入国家阶段日
 2020.10.20
(86) PCT国际申请的申请数据
 PCT/KR2019/005000 2019.04.25
(87) PCT国际申请的公布数据
 W02019/209040 EN 2019.10.31
(73) 专利权人 三星电子株式会社
 地址 韩国京畿道
(72) 发明人 王昱 沈逸麟 金红霞
(74) 专利代理机构 北京市立方律师事务所
 11330

(51) Int.Cl.
 G10L 15/18 (2013.01)
 G10L 19/00 (2013.01)
 G06F 40/30 (2020.01)
(56) 对比文件
 CN 105874530 A, 2016.08.17
 CN 106575502 A, 2017.04.19
 CN 107004408 A, 2017.08.01
 CN 107491468 A, 2017.12.19
 CN 107516511 A, 2017.12.26
 CN 107533542 A, 2018.01.02
 CN 107810496 A, 2018.03.16
 CN 107924680 A, 2018.04.17
 CN 1551103 A, 2004.12.01
 US 2012053945 A1, 2012.03.01
 US 2016196820 A1, 2016.07.07
 US 2018067923 A1, 2018.03.08
 (续)
 审查员 朱晓旻

专利代理师 谢玉斌 权利要求书4页 说明书26页 附图10页

(54) 发明名称
 理解自然语言短语的多模型
(57) 摘要

系统基于接收到的短语中的对象确定意图值,并且基于接收到的短语中的对象确定细节值。系统基于意图值和细节值确定意图状态值,并且基于意图值和细节值确定细节状态值和意图细节值。系统基于意图值和接收到的短语中的另一对象来确定其他意图值,并且基于细节值和接收到的短语中的另一对象来确定其他细节值。系统基于其他意图值、其他细节值和意图状态值来确定一般意图值,并且基于其他意图值、其他细节值和细节状态值来确定另一意图细节值。



CN 112005299 B

[接上页]

(56) 对比文件

Grégoire Mesnil等.Using Recurrent Neural Networks for Slot Filling in Spoken Language Understanding.IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND

LANGUAGE PROCESSING.2014,530-539.

Bing Liu等.Attention-Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling.arXiv:1609.01454.2016,686-687.

1. 一种设备,所述设备包括:

一个或多个处理器;以及

非暂时性计算机可读介质,所述非暂时性计算机可读介质存储有多个指令,所述指令在被执行时使所述一个或多个处理器执行以下操作:

意图编码器基于接收到的短语中的对象确定意图值;

细节编码器基于接收到的短语中的对象确定细节值;

意图解码器基于所述意图值和所述细节值确定意图状态值;

细节解码器基于所述意图值和所述细节值确定细节状态值和意图细节值;

所述意图编码器基于所述意图值和接收到的短语中的另一个对象确定其他意图值;

所述细节编码器基于所述细节值和接收到的短语中的所述另一个对象确定其他细节值;

所述意图解码器基于所述其他意图值、所述其他细节值和所述意图状态值确定一般意图值;以及

所述细节解码器基于所述其他意图值、所述其他细节值和所述细节状态值确定另一意图细节值。

2. 根据权利要求1所述的设备,其中,所述多个指令在被执行时使所述一个或多个处理器进一步执行以下至少一个操作:

基于确定的一般意图值与经验证的一般意图之间的差来修改所述意图编码器和所述意图解码器中的至少一个中的参数;或者

基于确定的意图细节值和确定的其他意图细节值中的至少一个与至少一个经验证的意图细节之间的差来修改所述细节编码器和所述细节解码器中的至少一个中的参数。

3. 根据权利要求1所述的设备,其中,所述多个指令在被执行时使所述一个或多个处理器进一步执行以下操作:

输出对接收到的短语的响应,所述响应基于确定的一般意图值、确定的意图细节值和确定的其他意图细节值中的至少一个。

4. 根据权利要求1所述的设备,其中,所述多个指令在被执行时使所述一个或多个处理器进一步执行以下操作:

补充编码器基于接收到的补充对象确定多个补充值;

补充解码器基于所述意图值、所述细节值和所述多个补充值确定补充状态值和一个补充值;

所述补充编码器基于所述意图值、所述细节值和接收到的补充对象确定其他补充值;以及

所述补充解码器基于所述其他意图值、所述其他细节值和所述补充状态值确定另一补充值。

5. 根据权利要求4所述的设备,其中,

所述补充编码器包括图像编码器,

所述多个补充值包括图像值,

接收到的补充对象包括图像对象,

所述补充解码器包括图像解码器,

所述补充状态值包括图像状态值，
所述一个补充值包括图像值，
所述其他补充值包括其他图像值，并且
所述另一补充值包括另一图像值。

6. 根据权利要求4所述的设备，其中，
所述补充编码器包括音频编码器，
所述多个补充值包括音频值，
接收到的补充对象包括音频对象，
所述补充解码器包括音频解码器，
所述补充状态值包括音频状态值，
所述一个补充值包括音频值，
所述其他补充值包括其他音频值，
所述另一补充值包括另一音频值，
所述补充编码器包括建议编码器，
所述多个补充值包括建议值，
接收到的补充对象包括建议对象，
所述补充解码器包括建议解码器，
所述补充状态值包括建议状态值，
所述一个补充值包括建议值，
所述其他补充值包括其他建议值，并且
所述另一补充值包括另一建议值。

7. 一种方法，所述方法包括：

意图编码器基于接收到的短语中的对象确定意图值；
细节编码器基于接收到的短语中的对象确定细节值；
意图解码器基于所述意图值和所述细节值确定意图状态值；
细节解码器基于所述意图值和所述细节值确定细节状态值和意图细节值；
所述意图编码器基于所述意图值和接收到的短语中的另一个对象确定其他意图值；
所述细节编码器基于所述细节值和接收到的短语中的所述另一个对象确定其他细节值；

所述意图解码器基于所述其他意图值、所述其他细节值和所述意图状态值确定一般意图值；以及

所述细节解码器基于所述其他意图值、所述其他细节值和所述细节状态值确定另一意图细节值。

8. 根据权利要求7所述的方法，其中，所述方法还包括以下至少之一：

基于确定的一般意图值与经验证的一般意图之间的差来修改所述意图编码器和所述意图解码器中的至少一个中的参数；或者

基于确定的意图细节值和确定的其他意图细节值中的至少一个与至少一个经验证的意图细节之间的差来修改所述细节编码器和所述细节解码器中的至少一个中的参数。

9. 根据权利要求7所述的方法，其中，所述方法还包括：

输出对接收到的短语的响应,所述响应基于确定的一般意图值、确定的意图细节值和确定的其他意图细节值中的至少一个。

10. 根据权利要求7所述的方法,其中,所述方法还包括:

补充编码器基于接收到的补充对象确定多个补充值;

补充解码器基于所述意图值、所述细节值和所述多个补充值确定补充状态值和一个补充值;

所述补充编码器基于所述意图值、所述细节值和接收到的补充对象确定其他补充值;以及

所述补充解码器基于所述其他意图值、所述其他细节值和所述补充状态值确定另一补充值。

11. 根据权利要求10所述的方法,其中,

所述补充编码器包括图像编码器,

所述多个补充值包括图像值,

接收到的补充对象包括图像对象,

所述补充解码器包括图像解码器,

所述补充状态值包括图像状态值,

所述一个补充值包括图像值,

所述其他补充值包括其他图像值,并且

所述另一补充值包括另一图像值。

12. 根据权利要求10所述的方法,其中,

所述补充编码器包括音频编码器,

所述多个补充值包括音频值,

接收到的补充对象包括音频对象,

所述补充解码器包括音频解码器,

所述补充状态值包括音频状态值,

所述一个补充值包括音频值,

所述其他补充值包括其他音频值,

所述另一补充值包括另一音频值,

所述补充编码器包括建议编码器,

所述多个补充值包括建议值,

接收到的补充对象包括建议对象,

所述补充解码器包括建议解码器,

所述补充状态值包括建议状态值,

所述一个补充值包括建议值,

所述其他补充值包括其他建议值,并且

所述另一补充值包括另一建议值。

13. 一种方法,所述方法包括:

意图编码器基于接收到的短语中的对象确定意图值;

细节编码器基于接收到的短语中的对象确定细节值;

意图解码器基于所述意图值和所述细节值确定意图状态值；

所述意图编码器基于所述意图值和接收到的短语中的另一个对象确定其他意图值；

所述细节编码器基于所述细节值和接收到的短语中的所述另一个对象确定其他细节值；

所述意图解码器基于所述其他意图值、所述其他细节值和所述意图状态值确定一般意图值；以及

基于确定的一般意图值和经验证的一般意图之间的差来修改所述意图编码器和所述意图解码器中的至少一个中的参数。

14. 根据权利要求13所述的方法，其中，所述方法还包括：

细节解码器基于所述意图值和所述细节值确定细节状态值和意图细节值；

所述细节解码器基于所述其他意图值、所述其他细节值和所述细节状态值确定另一意图细节值；以及

基于确定的意图细节值和确定的其他意图细节值中的至少一个与至少一个经验证的意图细节之间的差来修改所述细节编码器和所述细节解码器中的至少一个中的参数。

15. 一种非暂时性计算机可读介质，所述非暂时性计算机可读介质存储有指令，其中，所述指令在被执行时，使一个或多个处理器执行根据权利要求7至14中任一项所述的方法。

理解自然语言短语的多模型

技术领域

[0001] 本公开总体上涉及理解自然语言的系统,并且更具体地,涉及理解自然语言短语的多模型。

背景技术

[0002] 电子设备(例如智能手机或膝上型计算机)可以包括个人数字助理,该个人数字助理可以帮助该设备的用户执行不同的任务,例如设置闹钟、接收天气预报或查看新闻更新。个人数字助理可以使用自然语言理解引擎来理解用户说出或以其他方式输入的短语。用户可以输入不同的短语,例如“Flight from Denver to Philadelphia(从丹佛到费城的航班)”,表达用户查找航班信息的意图。因此,自然语言理解引擎需要执行两项任务:1)了解用户的一般意图,例如查找航班信息,以及2)了解意图细节,例如航班出发城市“Denver(丹佛)”和航班到达城市“Philadelphia(费城)”。

[0003] 许多个人数字助理使用两种基于机器学习/规则的模型来分别执行这两项任务。即使这样的个人数字助理正确执行了一项任务,如果个人数字助理错误地执行了另一项任务,则该个人数字助理也会对用户的输入短语做出错误的响应。例如,如果用户说“Check the price of Apple(查看苹果的价格)”,而个人数字助理将“Check the price of...(查看...的价格)”的一般意图识别为对股票市场价格的请求,并将“...Apple(...苹果)”的意图细节识别为水果的名称,则个人数字助理无法通过尝试提供一种水果的股票市场价格来正确地做出响应。相反,如果个人数字助理将“Check the price of...”的一般意图识别为对食品价格的请求,并且将“...Apple”的意图细节识别为公司名称,则个人数字助理将无法通过尝试提供公司的食品价格来正确地做出响应。

发明内容

[0004] 技术问题

[0005] 本公开是提供理解自然语言短语的多模型。

[0006] 技术方案

[0007] 在实施例中,系统理解自然语言短语。意图编码器基于接收到的短语中的对象确定意图值。细节编码器基于接收到的短语中的对象确定细节值。意图解码器基于所述意图值和所述细节值确定意图状态值。细节解码器基于所述意图值和所述细节值确定细节状态值和意图细节值。所述意图编码器基于所述意图值和所述接收到的短语中的另一个对象确定其他意图值。所述细节编码器基于所述细节值和接收到的短语中的另一个对象确定其他细节值。所述意图解码器基于所述其他意图值、所述其他细节值和所述意图状态值确定一般意图值。所述细节解码器基于所述其他意图值、所述其他细节值和所述细节状态值确定另一意图细节值。

[0008] 在另一实施例中,一种方法理解自然语言短语。该方法包括意图编码器基于接收到的短语中的对象确定意图值,包括细节编码器基于接收到的短语中的对象确定细节值。

该方法还包括意图解码器基于所述意图值和所述细节值确定意图状态值,包括细节解码器基于所述意图值和所述细节值确定细节状态值和意图细节值。该方法还包括所述意图编码器基于所述意图值和所述接收到的短语中的另一个对象确定其他意图值,包括所述细节编码器基于所述细节值和接收到的短语中的另一个对象确定其他细节值。该方法还包括所述意图解码器基于所述其他意图值、所述其他细节值和所述意图状态值确定一般意图值,包括所述细节解码器基于所述其他意图值、所述其他细节值和所述细节状态值确定另一意图细节值。

[0009] 在又一实施例中,一种计算机程序产品包括计算机可读程序代码,该计算机可读程序代码包括理解自然语言短语的指令。该程序代码包括用于意图编码器基于接收的短语中的对象确定意图值以及细节编码器基于接收到的短语中的对象确定细节值的指令。该程序代码还进一步包括用于意图解码器基于意图值和细节值确定意图状态值以及细节解码器基于意图值和细节值确定细节状态值和意图细节值的指令。该程序代码包括用于意图编码器基于意图值和接收到的短语中的另一对象确定其他意图值以及细节编码器基于细节值和接收到的短语中的另一对象确定其他细节值的额外的指令。该程序代码还包括用于意图解码器基于其他意图值、其他细节值和意图状态值确定一般意图值以及细节解码器基于其他意图值、其他细节值和细节状态值确定另一意图细节值的指令。

[0010] 图1描绘了理解自然语言短语的多模型100的示例。在用户说“Buy Apple (购买苹果)”之后,自然语言理解引擎将用户的短语解析为“Buy (购买)”102和“Apple (苹果)”104。在时间步骤1的意图编码器106将“Buy”102的词典编号 x_1 转换为“Buy”的一般意图编号 h_1^1 ,并且在时间步骤1的细节编码器108将“Buy”102的词典编号 x_1 转换为“Buy”意图细节编号 h_1^2 。在时间步骤1的意图解码器110将“Buy”一般意图编号 h_1^1 和“Buy”意图细节编号 h_1^2 转换为“Buy”意图状态编号 s_1^1 ,并且在时间步骤1的细节解码器112将“Buy”一般意图编号 h_1^1 和“Buy”意图细节编号 h_1^2 转换为“Buy”意图状态编号 s_1^2 和确定意图细节标签 $y_1^2=0$ 的意图细节编号 y_1^2 。在时间步骤2的意图编码器114将“Apple”104的词典编号 x_2 和“Buy”一般意图编号 h_1^1 转换为“Buy Apple”一般意图编号 h_2^1 ,并且在时间步骤2的细节编码器116将“Apple”104的词典编号 x_2 和“Buy”意图细节编号的 h_1^2 转换为“Apple”意图细节编号 h_2^2 。在时间步骤2的意图解码器118将“Buy Apple”一般意图编号 h_2^1 、“Apple”意图细节编号 h_2^2 和“Buy”意图状态编号 s_1^1 转换为一般意图值 y_2^1 ,该一般意图值 y_2^1 确定一般意图类别 $y_{\text{意图}}^1$ = “食品订单”或“股票市场购买”,并且每个一般意图类别都包括其相应的概率。在时间步骤2的细节解码器120将“Apple”一般意图编号 h_2^1 和“Apple”意图细节编号 h_2^2 以及“Buy”意图状态编号 s_1^2 转换为确定意图细节标签 y_2^2 = “水果名称”或“公司名称”的意图细节编号 y_2^2 ,并且每个标签都包括其对应的概率。如果确定了意图细节标签 y_2^2 = “水果名称”,则自然语言理解引擎确定一般意图类别 $y_{\text{意图}}^1$ = “食品订单”,并以“您何时要订购苹果这种水果?”的答复对用户进行响应。如果确定了意图细节标签 y_2^2 = “公司名称”,则自然语言理解引擎确定一般意图类别 $y_{\text{意图}}^1$ = “购买股票”,并以“你想购买多少苹果公司的股票”的答复对用户进行响应。

[0011] 提供本发明内容部分仅是为了引入某些概念,而不是确定所要求保护的主题的任

[0012] 技术效果

[0013] 根据本公开的各种实施例,通过理解自然语言短语的多模型,用户的输入短语可以根据其意图被正确地识别。

附图说明

[0014] 附图示出了一个或多个实施例。然而,附图不应被用来仅将本发明限制为示出的实施例。通过阅读以下详细说明并参考附图,各个方面和优点将变得显而易见。

[0015] 图1示出了根据实施例的理解自然语言短语的示例多模型;

[0016] 图2示出了根据实施例的用于理解自然语言短语的多模型的示例系统的框图;

[0017] 图3示出了根据实施例的被训练为理解自然语言短语的示例双模型结构;

[0018] 图4示出了根据实施例的理解自然语言短语的示例双模型结构;

[0019] 图5示出了根据实施例的理解自然语言短语的不包括解码器的示例双模型结构;

[0020] 图6示出了根据实施例的理解自然语言短语的包括图像模型的示例三模型结构;

[0021] 图7示出了根据实施例的理解自然语言短语的包括音频模型的示例三模型结构;

[0022] 图8示出了根据实施例的理解自然语言短语的包括自动建议模型的示例三模型结构;

[0023] 图9是示出根据实施例的用于理解自然语言短语的多模型的方法的流程图;以及

[0024] 图10是示出其中可以实现本主题的示例硬件设备的框图。

具体实施方式

[0025] 以下描述是为了说明一个或多个实施例的一般原理,并不意味着限制本文所保护的发明构思。此外,本文描述的特定特征可以以各种可能的组合和排列中的每一个与其他描述的特征组合使用。除非本文另有明确定义,否则所有术语均应给予其尽可能广泛的解释,包括从说明书中隐含的含义以及本领域技术人员所理解的含义和/或在词典、专论等中定义的含义。

[0026] 术语“对象”或“位置”通常指表示概念的一个单词、一组单词或一组字符。术语“接收到的短语”或“话语”通常指输入到系统中并且包括至少一个动词和至少一个名词的一组单词。术语“接收到的补充对象”通常指同样输入到系统中并表示概念的一个单词、一组单词或一组字符。术语“图像对象”通常用于指视觉表示。术语“音频对象”通常指输入到系统中的任何类型的声音,例如语音或录音。术语“建议对象”通常指表示所提出的概念的一个单词、一组单词或一组字符。

[0027] 术语“值”通常是指数值、代数术语或标签。术语“一般意图值”通常指与主要目的相关联的数值、代数术语或标签。术语“意图细节值”通常指与目的属性相关联的数值、代数术语或标签。

[0028] 术语“意图值”通常指与目的相关联的数值、代数术语或标签。术语“细节值”通常指与目的属性相关联的数值、代数术语或标签。术语“补充值”通常指附加的数值、代数术语或标签。术语“图像值”通常指与视觉表示相关联的数值、代数术语或标签。术语“音频值”通常指与输入到系统的任何类型的声音相关联的数值、代数术语或标签。术语“建议值”通常指与建议相关联的数值、代数术语或标签。

[0029] 术语“意图状态值”通常指与条件和目的相关联的数值、代数术语或标签。术语“细

节状态值”通常指与目的的条件和属性相关联的数值、代数术语或标签。术语“补充状态值”通常指与条件相关联的附加的数值、代数术语或标签。术语“图像状态值”通常指与条件和视觉表示相关联的数值、代数术语或标签。术语“音频状态值”通常指与条件和输入到系统中的任何类型的声音相关联的数值、代数术语或标签。术语“建议状态值”通常指与条件和建议相关联的数值、代数术语或标签。

[0030] 术语“意图编码器”通常指将与目的相关联的信息或代码从一种格式转换为另一种格式的电路、软件程序或算法。术语“细节编码器”通常指将与目的的属性相关联的信息或代码从一种格式转换为另一种格式的电路、软件程序或算法。术语“补充编码器”通常指将信息或代码从一种格式转换为另一种格式的附加电路、软件程序或算法。术语“图像编码器”通常指将与视觉表示相关联的信息或代码从一种格式转换为另一种格式的电路、软件程序或算法。术语“音频编码器”通常指将与任何类型的声音相关联的信息或代码从一种格式转换为另一种格式的电路、软件程序或算法。术语“建议编码器”通常指将与建议相关联的信息或代码从一种格式转换为另一种格式的电路、软件程序或算法。

[0031] 术语“意图解码器”通常指将与目的相关联的信号或代码从一种格式转换为另一种格式的电路、软件程序或算法。术语“细节解码器”通常指将与目的的属性相关联的信号或代码从一种格式转换为另一种格式的电路、软件程序或算法。术语“补充解码器”通常指将信号或代码从一种格式转换为另一种格式的附加电路、软件程序或算法。术语“图像解码器”通常指将与视觉表示相关联的信号或代码从一种格式转换为另一种格式的电路、软件程序或算法。术语“音频解码器”通常指将与任何类型的声音相关联的信号或代码从一种格式转换为另一种格式的电路、软件程序或算法。术语“建议解码器”通常指将与建议相关联的信号或代码从一种格式转换为另一种格式的电路、软件程序或算法。

[0032] 术语“参数”通常指形成定义系统或设置其操作条件的集合的数值或其他可测量因素。术语“差”通常指量不同的数量,或者一个值与另一个值相减后剩下的余数。术语“经验证的一般意图”通常指已确认的主要目的。术语“经验证的意图细节”通常指确认的目的属性。术语“响应”通常指对某事物的反应。

[0033] 图2示出了根据实施例的用于理解自然语言短语的系统200的框图。如图2所示,系统200可以示出云计算环境,在该云计算环境中数据、应用、服务和其他资源通过共享的数据中心被存储和传递,并且对于终端用户而言表现为单个访问点。系统200还可以代表服务器控制用于不同客户端用户的资源和服务的存储和分配的任何其他类型的分布式计算机网络环境。

[0034] 在实施例中,系统200表示云计算系统,其包括第一客户端202、第二客户端204以及可以由托管公司提供的第一服务器206和第二服务器208。客户端202-204和服务器206-208经由网络210进行通信。尽管图2将第一客户端202描绘为智能手机202并且将第二客户端204描绘为膝上型计算机204,但是客户端202-204中的每一个可以是任何类型的计算机。在实施例中,可以称为自然语言服务器206的第一服务器206包括组件212-234。尽管图2描绘了具有两个客户端202-204、两个服务器206-208和一个网络210的系统200,但是系统200可以包括任何数量的客户端202-204、任何数量的服务器206-208和/或任何数量的网络210。客户端202-204和服务器206-208可以分别与图10所示和以下描述的系统1000基本相似。

[0035] 系统组件212-234(每个组件可以组合成较大的组件和/或划分为较小的组件)包括训练器212、个人数字助理214和自然语言理解引擎216。自然语言理解引擎216包括:意图模型218,该意图模型218包括意图编码器220和意图解码器222;细节模型224,该细节模型224包括细节编码器226和细节解码器228;以及补充模型230,该补充模型230包括补充编码器232和补充解码器234。图2描绘了系统组件212-234完全驻留在自然语言服务器206上,但是系统组件212-234也可以完全驻留在自然语言服务器206上,完全驻留在第二服务器208上,完全驻留在客户端202-204上,或部分驻留在服务器206-208上和部分驻留在客户端202-204上的任何组合中。例如,在自然语言服务器206使用训练器212来训练个人数字助理214中的自然语言理解引擎216时,自然语言服务器206可以将个人数字助理214的副本提供给智能手机202。

[0036] 自然语言理解引擎216可以理解不是自然语言句子而是不完整的句子的自然语言短语,这是因为当与个人数字助理交互时,人们更可能输入的是诸如“flight from Denver to Philadelphia(从丹佛到费城的航班)”这样不完整的句子短语而不是输入诸如“我想要关于从丹佛到费城的航班的信息”这样的句子。此外,如果对自然语言理解引擎216进行充分的训练以便于理解诸如“flight from Denver to Philadelphia(从丹佛到费城的航班)”这样不完整的句子,那么自然语言理解引擎216会受到充分的训练以理解包括完整句子短语的完整句子,例如“我想要有关从丹佛到费城的航班的信息”。

[0037] 图3示出了根据实施例的用于训练以理解自然语言短语的包括解码器的双模型结构300的框图。自然语言服务器206可以具有多个基于神经的编码器-解码器模型,例如一般意图检测任务网络和意图细节检测任务网络,其可以被称作位置标签检测任务网络或位置填充检测任务网络。双向模型结构300包括用于一般意图检测任务网络的双向编码器和用于意图细节检测任务网络的双向编码器。双模型结构300可以使用双向长短期存储器(BLSTM)来实现相应的编码器,并且可以使用长短期存储器(LSTM)来实现相应的解码器。

[0038] 每个编码器读取从接收到的短语前向和反向解析出来的对象序列 $(x_1, x_2, \dots, x_t)$ 中的对象(通常是单词)。然后,用于一般意图检测任务的编码器神经网络生成隐藏状态 h_i^1 ,而用于意图细节检测任务的编码器神经网络生成隐藏状态 h_i^2 。与分别训练一般意图模型和意图细节模型的个人数字助理系统不同,自然语言服务器206通过与其他模型共享每个模型的隐藏状态来一起使用一般意图模型和意图细节模型。例如,来自意图编码器 f_1 的原始隐藏状态 h_i^1 与来自意图细节编码器 f_2 的 h_{i-1}^2 连接,然后被馈送到意图解码器 g_1 。类似地,对于意图细节模型,来自其细节编码器 f_2 的隐藏状态 h_i^2 与来自意图编码器 f_1 的隐藏状态 h_{i-1}^1 连接,然后被馈送到细节解码器 g_2 。因此,BLSTM生成隐藏状态序列 $(h_1^i, h_2^i, \dots, h_t^i)$,其中 $i=1$ 对应于一般意图检测任务网络,而 $i=2$ 对应于意图细节检测任务网络。为了检测一般意图,将意图编码器 f_1 的隐藏状态 h^1 与意图细节编码器 f_2 的隐藏状态 h^2 组合在一起以生成状态 s^1 :

$$[0039] \quad s_t^1 = \phi(s_{t-1}^1, h_{t-1}^1, h_{t-1}^2)$$

$$[0040] \quad y_{\text{意图}}^1 = \arg \max (\hat{y}_t^1) P(\hat{y}_t^1 | s_{t-1}^1, h_{t-1}^1, h_{t-1}^2)$$

[0041] 其中 $\hat{y}_t^1$ 包含最后一个训练阶段 t 的所有一般意图类别的预测概率。

[0042] 类似地,为了检测意图细节,将来自意图细节编码器 f_2 的隐藏状态 h^2 与来自意图编码器 f_1 的隐藏状态 h^1 组合在一起以生成状态 s^2 。然而,如果解决了序列标签问题,则意图细

节解码器 f_2 将在每个训练阶段 t 生成输出 y_t^2 。在每个训练阶段:

$$[0043] \quad s_t^2 = \psi(h_{t-1}^2, h_{t-1}^1, s_{t-1}^2, y_{t-1}^2)$$

$$[0044] \quad y_t^2 = \arg \max(\hat{y}_t^2) P(\hat{y}_t^2 | h_{t-1}^2, h_{t-1}^1, s_{t-1}^2, y_{t-1}^2)$$

[0045] 其中 y_t^2 是在训练阶段 t 检测到的意图细节标签。

[0046] 自然语言服务器206可以使用具有共享隐藏状态参数的自身损失函数来训练多个任务网络。在每个训练迭代中,一般意图检测任务网络和意图细节检测任务网络都从其模型中的先前迭代中生成一组隐藏状态 h_t^1 和 h_t^2 。一般意图检测任务网络读取一批输入数据 x_i 和隐藏状态 h_t^2 ,然后生成一般意图类别 $y_{意图}^1$ 。

[0047] 例如,在系统管理员说“Flight from Denver to Philadelphia(从丹佛到费城的航班)”之后,自然语言理解引擎216将系统管理员的短语解析为“Flight(航班)”302、“from(从)”304、“Denver(丹佛)”306、“to(到)”308以及“Philadelphia(费城)”310。训练阶段1的意图编码器312将“Flight”302的词典编号 x_1 转换为“flight”一般意图编号 h_1^1 ,而训练阶段1的细节编码器314将“Flight”302的词典编号 x_1 转换为“flight”意图细节编号 h_1^2 。诸如“Flight”302之类的对象的词典编号 x_i 是诸如词语“flight”之类的对象的含义的数字表示,其可以基于与该对象相同的短语中的其他词语通过算法被转换为 h_i^1 和 h_i^2 ,这是对象含义的其他数字表示。诸如词语“Flight”302之类的对象的值 x_1 、 x_2 、 h_i^1 、 h_i^2 、 s_1^1 和 s_1^2 可以表示为任意维度(例如,200)的向量。训练阶段1的意图解码器316将“flight”一般意图编号 h_1^1 和“flight”意图细节编号 h_1^2 转换为“flight”意图状态编号 s_1^1 。

[0048] 训练阶段2的意图编码器318将“from”304的词典编号 x_2 和“flight”一般意图编号 h_1^1 和意图细节编号 h_1^2 转换为“flight from(来自...的航班)”一般意图编号 h_2^1 ,并且训练阶段2的细节编码器320将“from”304的词典编号 x_2 、“flight”一般意图编号 h_1^1 和意图细节编号 h_1^2 转换为“from”意图细节编号 h_2^2 。训练阶段2的意图解码器322将“flight from”一般意图编号 h_2^1 、“from”意图细节编号 h_2^2 和“flight”意图状态编号 s_1^1 转换为“flight from”意图状态编号 s_2^1 。

[0049] 训练阶段3的意图编码器324将“Denver”306的词典编号 x_3 、“flight from”一般意图编号 h_2^1 和意图细节编号 h_2^2 转换为“flight from Denver(从丹佛出发的航班)”一般意图编号 h_3^1 ,并且训练阶段3的细节编码器326将“Denver”306的词典编号 x_3 和“from”意图细节编号 h_2^2 转换为“Denver”意图细节编号 h_3^2 。训练阶段3的意图解码器328将“flight from Denver”一般意图编号 h_3^1 、“Denver”意图细节编号 h_3^2 和“from”意图状态编号 s_2^1 转换为“flight from Denver”意图状态编号 s_3^1 。

[0050] 训练阶段4的意图编码器330将“to”308的词典编号 x_4 和“flight from Denver”一般意图编号 h_3^1 和意图细节编号 h_3^2 转换为“flight from Denver to(从丹佛到...的航班)”一般意图编号 h_4^1 ,训练阶段4的细节编码器332将“to”308的词典编号 x_4 和“Denver”意图细节编号 h_3^2 转换为“to”意图细节编号 h_4^2 。训练阶段4的意图解码器334将“flight from Denver to”一般意图编号 h_4^1 、“to”意图细节编号 h_4^2 和“flight from Denver”意图状态编号 s_3^1 转换为“flight from Denver to”意图状态编号 s_4^1 。

[0051] 训练阶段5的意图编码器336将“Philadelphia”310的词典编号 x_5 和“flight from Denver to”的一般意图编号 h_4^1 和意图细节编号 h_4^2 转换为“flight from Denver to Philadelphia”一般意图编号 h_5^1 ,并且训练阶段5的细节编码器338将“Philadelphia”310的

词典编号 x_5 和“to”意图细节编号 h_4^2 转换为“Philadelphia”意图细节编号 h_5^2 。训练阶段5的意图解码器340将“flight from Denver to Philadelphia”一般意图编号 h_5^1 、“Philadelphia”意图细节编号 h_5^2 和“flight from Denver to”意图状态编号 s_4^1 转换为确定一般意图类别 $y_{\text{意图}}^1$ = “flight”的一般意图编号 $\hat{y}_5^1$ ，其中包括一般意图类别的相应概率。训练阶段1-5 (312、318、324、330和336) 的意图编码器是意图模型218中的意图编码器220的示例，而训练阶段1-5 (316、322、328、334和340) 的意图解码器是意图模型218中的意图解码器222的示例。

[0052] 一般意图模型中最后训练阶段的意图解码器通过识别出一般意图编号 $\hat{y}_t^1$ 中的哪一个对应于最大概率来确定接收到的短语的一般意图，从而确定具有一般意图类别 $y_{\text{意图}}^1$ 的相应的一般意图。下面参考图4描述一般意图模型中的意图解码器将一般意图确定为 $y_{\text{意图}}^1$ = “查看水果价格”或“查看股票价格”的示例。

[0053] 自然语言服务器206基于函数 L_1 计算一般意图检测任务网络代价，并在一般意图检测任务网络代价的基础上训练一般意图检测任务网络。使用交叉熵将 L_1 定义为：

$$[0054] \quad L_1 \triangleq -\sum_{i=1}^k \hat{y}_{\text{意图}}^{1,i} \log(y_{\text{意图}}^{1,i})$$

[0055] 其中， k 是一般意图类别类型的数量。

[0056] 在图3的示例中， $y_{\text{意图}}^1$ = “flight”，并且系统管理员或算法验证了短语“flight from Denver to Philadelphia”的一般意图类别“flight”。因此，损失函数或代价函数（数据实例的估计值与真实值之间的差的函数）的计算将导致意图编码器或意图解码器的参数变化最小（如果有的话）。在导致确定的一般意图值与经验证的一般意图值之间存在差异的其他示例中，这种差异可能导致意图编码器和/或意图解码器的参数发生任何数量的改变。

[0057] 接下来，自然语言服务器206将相同批次的数据 x_i 与来自一般意图检测任务网络的隐藏状态 h_t^1 一起提供给意图细节检测任务网络，这使得意图细节检测任务网络能够生成一批训练阶段2的输出 y_t^2 。

[0058] 继续图3所示的示例，训练阶段1的细节解码器342将“flight”一般意图编号 h_1^1 和“flight”意图细节编号 h_1^2 转换为“flight”细节状态编号 s_1^2 和确定出意图细节标签 y_1^2 = “0”的意图细节编号 $\hat{y}_1^2$ 。意图细节模型中的细节解码器通过确定与最大概率相对应的意图细节编号 $\hat{y}_n^2$ 中的一个是否具有大于意图细节概率阈值的概率然后使用意图细节标签 y_n^2 标记相应的意图细节来确定接收到的短语中的对象的意图细节。然而，在多个意图细节编号 $\hat{y}_t^1$ 对应于大于意图细节概率阈值的概率的一些情况下，细节解码器可以用意图细节标签 y_n^2 来标记多个对应的意图细节。下面参考图4来描述细节模型中的细节解码器将意图细节确定并标记为 y_5^2 = “水果名称”或 y_5^2 = “公司名称”的示例。

[0059] 训练阶段2的细节解码器344将“flight from”一般意图编号 h_2^1 、“from”意图细节编号 h_2^2 和“flight”细节状态编号 s_1^2 转换为“from”细节状态编号 s_2^2 和确定意图细节标签 y_2^2 = 0的意图细节编号 $\hat{y}_2^2$ 。训练阶段3的细节解码器346将“flight from Denver”一般意图编号 h_3^1 、“Denver”意图细节编号 h_3^2 和“from”细节状态编号 s_2^2 转换为“Denver”细节状态编号 s_3^2 和确定意图细节标签 $\hat{y}_3^2$ = “出发地（from location）”的意图细节编号 $\hat{y}_3^2$ 。训练阶

段4的细节解码器348将“flight from Denver to”转换为一般意图编号 h_4^1 、“to”意图细节编号 h_4^2 和“Denver”细节状态编号 s_3^2 转换为“to”细节状态编号 s_4^2 和确定细节意图标签 $y_4^2 = 0$ 的意图细节标签 $\hat{y}_4^2$ 。训练阶段5的细节解码器350将“Flight from Denver to Philadelphia”一般意图数字 h_5^1 、“Philadelphia”意图细节数字 h_5^2 和“to”细节状态编号 s_4^2 转换为确定细节意图标签 $y_5^2 =$ “到达地(to location)”的细节意图编号 $\hat{y}_5^2$ 。训练阶段1-5(314、320、326、332和338)的细节编码器是细节模型224中的细节编码器226的示例,训练阶段1-5(342、342、344、346、348和350)的细节解码器的是细节模型224中的细节解码器226的示例。

[0060] 然后,自然语言服务器206基于函数 L_2 计算意图细节检测任务网络代价,并且根据意图细节检测任务网络代价训练意图细节检测任务网络。使用交叉熵将 L_2 定义为:

$$[0061] \quad L_2 \triangleq - \sum_{j=1}^n \sum_{i=1}^m \hat{y}_j^{2,i} \log(y_j^{2,i})$$

[0062] 其中,m是意图细节标签类型的数量,n是短语中对象(词语)的数量。

[0063] 在图3的示例中, $y_1^2 =$ “0”、 $y_3^2 =$ “出发地(from position)”、 $y_5^2 =$ “到达地”,并且系统管理员或算法仅验证短语“flight from Denver to Philadelphia”的意图细节标签“出发地”和“到达地”。因此,损失函数或代价函数的计算将导致细节编码器和/或细节解码器的参数发生变化,从而降低生成意图细节标签 $y_1^2 =$ “0”的概率。在导致确定的意图细节值与经验证的意图细节值之间没有差异的其他示例中,细节编码器和/或细节解码器的参数中的变化量(如果有的话)最小。

[0064] 由于针对两个不同的任务,保持两个单独的代价函数很重要,所以自然语言服务器206使用异步训练。自然语言服务器206过滤了两个任务之间的负面影响,从而克服了一种模型的结构限制。只能通过共享两个模型的隐藏状态来学习两个任务之间的交叉影响,这两个模型使用两个代价函数进行单独训练。通过减少每个任务的错误并学习有用的共享信息,可以提高所有任务的性能。

[0065] 图4示出了根据实施例的用于理解自然语言短语的包括解码器的双模型结构400的框图。在用户说“Check the price of Apple(查看该苹果的价格)”之后,自然语言理解引擎216将用户的短语解析为“Check(查看)”402、“the(该)”404、“price(价格)”406、“of(的)”408和“apple(苹果)”410。在时间步骤1的意图编码器412将“Check”402的词典编号 x_1 转换为“check”一般意图编号 h_1^1 ,并且时间步骤1的细节编码器414将“Check”402的词典编号 x_1 转换为“check”意图细节编号 h_1^2 。在时间步骤1的意图解码器416将“check”意图编号 h_1^1 和“check”意图细节编号 h_1^2 转换为“check”意图状态编号 s_1^1 ,并且在时间步骤1的细节解码器418将“check”一般意图编号 h_1^1 和“check”意图细节编号 h_1^2 转换为“check”细节状态编号 s_1^2 和确定意图细节标签 $y_1^2 = 0$ 的意图细节编号 $\hat{y}_1^2$ 。

[0066] 在时间步骤2的意图编码器420将“the”404的词典编号 x_2 、“check”一般意图编号 h_1^1 和意图细节编号 h_1^2 转换为“check the(查看该)”一般意图编号 h_2^1 ,在时间步骤2的细节编码器422将“the”404的词典编号 x_2 、“check”意图细节编号 h_1^2 和一般意图编号 h_1^1 转换为“the”意图细节编号 h_2^2 。在时间步骤2的意图解码器424将“check the”一般意图编号 h_2^1 、“the”意图细节编号 h_2^2 和“check”意图状态编号 s_1^1 转换为“check the”意图状态序号 s_2^1 ,并且在时间步骤2的细节解码器426将“check the”一般意图编号 h_2^1 、“the”意图细节编号 h_2^2

和“check”细节状态编号 s_1^2 转换为“the”细节状态编号 s_2^2 和确定意图细节标签 $y_2^2=0$ 意图细节编号 $\hat{y}_2^2$ 。

[0067] 在时间步骤3的意图编码器428将“price”406的词典编号 x_3 、“check the”一般意图编号 h_2^1 和意图细节编号 h_2^2 转换为“check the price (查看该价格)”一般意图编号 h_3^1 ，并且在时间步骤3的细节编码器430将“price”406的词典编号 x_3 、“the”意图细节编号 h_2^2 和一般意图编号 h_2^1 转换为“price”意图细节编号 h_3^2 。在时间步骤3的意图解码器432将“check the price”一般意图编号 h_3^1 、“price”意图细节编号 h_3^2 和“check the price”意图状态编号 s_2^1 转换为“check the price”意图状态编号 s_3^1 ，在时间步骤3的细节解码器434将“check the price”一般意图编号 h_3^1 、“price”意图细节编号 h_3^2 和“the”细节状态编号 s_2^2 转换为“price”细节状态编号 s_3^2 和确定意图细节标签 $y_3^2=0$ 的意图细节编号 $\hat{y}_3^2$ 。

[0068] 时间步骤4的意图编码器436将“of (的)”408的词典编号 x_4 、“check the price”一般意图编号 h_3^1 、以及意图细节编号 h_3^2 转换为“check the price of (查看该…的价格)”一般意图编号 h_4^1 ，在时间步骤4的细节编码器438将“of”408的词典编号 x_4 、“price”意图细节编号 h_3^2 和一般意图编号 h_3^1 转换为“of”意图细节编号 h_4^2 。在时间步骤4的意图解码器440将“check the price of”一般意图编号 h_4^1 、“of”意图细节编号 h_4^2 和“check the price of”意图状态编号 s_3^1 转换为“check the price of”意图状态编号 h_4^1 ，在时间步骤4的细节解码器442将“check the price of”一般意图编号 h_4^1 、“of”意图细节编号 h_4^2 和“price”细节状态编号 s_3^2 转换为“of”细节状态编号 s_4^2 和确定意图细节标签 $y_4^2=0$ 的意图细节编号 $\hat{y}_4^2$ 。

[0069] 在时间步骤5的意图编码器444将“Apple”410的词典编号 x_5 、“check the price of”一般意图编号 h_4^1 和意图细节编号 h_4^2 转换为“check the price of Apple”一般意图编号 h_5^1 ，并且在时间步骤5的细节编码器446将“Apple”310的词典编号 x_5 、“of”意图细节编号 h_4^2 和一般意图编号 h_4^1 转换为“Apple”意图细节编号 h_5^2 。在时间步骤5的意图解码器448将“check the price of Apple”一般意图编号 h_5^1 、“Apple”意图细节编号 h_5^2 和“check the price of”意图状态编号 h_4^1 转换为一般意图编号 $\hat{y}_5^1$ ，该一般意图编号 $\hat{y}_5^1$ 确定一般意图类别 $y_{\text{意图}}^1$ = “check stock price (查看股票价格)”或“查看水果价格”，其中包括每个一般意图类别的对应概率，并且在时间步骤5的细节解码器450将“check the price of Apple”一般意图编号 h_5^1 、“Apple”意图细节编号 h_5^2 和“of”细节状态编号 s_4^2 转换为意图细节标记 $\hat{y}_5^2$ ，该意图细节标记 $\hat{y}_5^2$ 确定意图细节标签 y_5^2 = “公司名称”或“水果名称”，其包括每个标签的相应概率。然后，自然语言理解引擎216可以通过以下答复对用户做出响应：“您想知道苹果这种水果的价格还是苹果公司的股票价格？”。与使用两个不相关的模型来实现用于理解自然语言短语的单独的一般意图任务和意图细节任务的典型的个人数字助理相反，双模型结构400同时执行两个任务，这通过在两个任务网络之间共享有用的信息来提高每个任务的性能。因此，双模型结构400使个人数字助理214比典型的个人数字助理更好地理解模糊的自然语言短语。

[0070] 图5示出了根据实施例的用于理解自然语言短语的不包括解码器的双模型结构500的框图。在该双模型结构500中，不存在以前的双模型结构300和400中那样的解码器。用于一般意图检测任务的编码器在最后一个时间步骤n仅生成一个检测到的一般意图类别

$y_{\text{意图}}^1$, 其中 n 等于接收到的短语中的对象(词语)的数量。状态值 y_t^1 和一般意图类别 $y_{\text{意图}}^1$ 生成成为:

$$[0071] \quad h_t^1 = \varphi(h_{t-1}^1, h_{t-1}^2)$$

$$[0072] \quad y_{\text{意图}}^1 = \arg \max(\hat{y}_n^1) P(\hat{y}_n^1 | h_{n-1}^1, h_{n-1}^2)$$

[0073] 对于意图细节检测任务,编码器的基本结构与用于一般意图检测任务的编码器相似,其不同之处在于,在每个时间步骤 t 都生成一个意图细节标签 y_t^2 。意图细节编码器还从两个编码器中获得隐藏状态 h_{t-1}^1 和 h_{t-1}^2 以及意图细节标签 y_{t-1}^2 ,以生成下一个状态值 h_t^2 和意图细节标签 y_t^2 。这些在数学上表示为:

$$[0074] \quad h_t^2 = \psi(h_{t-1}^2, h_{t-1}^1, y_{t-1}^2)$$

$$[0075] \quad y_t^2 = \arg \max(\hat{y}_t^2) P(\hat{y}_t^2 | h_{t-1}^2, h_{t-1}^1, y_{t-1}^2)$$

[0076] 例如,图5示出了根据实施例的用于理解自然语言短语的不包括解码器的双模型结构500的框图。在用户说“flight from Denver to Philadelphia”之后,自然语言理解引擎216将用户的短语解析为“Flight”502、“from”504、“Denver”506、“to”508和“Philadelphia”510。在时间步骤1的意图编码器512将“Flight”502的词典编号 x_1 转换为“flight”一般意图编号 h_1^1 ,并且在时间步骤1的细节编码器514将“Flight”502的词典编号 x_1 转换为“flight”意图细节编号 h_1^2 。在时间步骤1的意图编码器512将“flight”一般意图编号 h_1^1 和“flight”意图细节编号 h_1^2 转换为“flight”一般意图状态编号 $\hat{y}_1^1$,并且在时间步骤1的细节编码器514将“flight”一般意图编号 h_1^1 和“flight”意图细节编号 h_1^2 转换为确定意图细节标签 $y_1^2=0$ 的“flight”意图细节状态编号 $\hat{y}_1^2$ 。

[0077] 在时间步骤2的意图编码器516将“from”504的词典编号 x_2 和“flight”一般意图状态编号 $\hat{y}_1^1$ 转换为“flight from”一般意图编号 h_2^1 ,并且在时间步骤2的细节编码器518将“from”504的词典编号 x_2 和“flight”意图细节状态编号 $\hat{y}_1^2$ 转换为“from”意图细节编号 h_2^2 。在时间步骤2的意图编码器516将“flight from”一般意图编号 h_2^1 和“from”意图细节编号 h_2^2 转换为“flight from”一般意图状态编号 $\hat{y}_2^1$,并且在时间步骤2的细节编码器518将“flight from”一般意图编号 h_2^1 和“from”意图细节编号 h_2^2 转换为确定意图细节标签 $y_2^2=0$ 的“from”意图细节状态编号 $\hat{y}_2^2$ 。

[0078] 在时间步骤3的意图编码器520将“Denver”506的词典编号 x_3 和“flight from”一般意图状态编号 $\hat{y}_2^1$ 转换为“flight from Denver(从丹佛出发的航班)”一般意图编号 h_3^1 ,并且在时间步骤3的细节编码器522将“Denver”506的词典编号 x_3 和“froms”意图细节状态编号 $\hat{y}_2^2$ 转换为“Denver”意图细节编号 h_3^2 。在时间步骤3的意图编码器520将“flight from Denver”一般意图编号 h_3^1 和“Denver”意图细节编号 h_3^2 转换为“flight from Denver”一般意图状态编号 $\hat{y}_3^1$,并且在时间步骤3的细节编码器522将“flight from Denver”一般意图编号 h_3^1 和“Denver”意图细节编号 h_3^2 转换为确定意图细节标签 $y_3^2=$ “出发地”的“Denver”细节状态编号 $\hat{y}_3^2$ 。

[0079] 在时间步骤4的意图编码器524将“to”508的词典编号 x_4 和“flight from Denver”一般意图状态编号 $\hat{y}_3^1$ 转换为“flight from Denver to”一般意图编号 h_4^1 ,并且在时间步骤4的细节编码器526将“to”508的词典编号 x_4 和“Denver”意图细节编号 h_3^2 转换为意图细节编号 h_4^2 。在时间步骤4的意图编码器524将“flight from Denver to”一般意图编号 h_4^1 和“to”意图细节编号 h_4^2 转换为“flight from Denver to”一般意图状态编号 $\hat{y}_4^1$,并且在时间步骤4的细节编码器526将“flight from Denver to”一般意图编号 $\hat{y}_4^1$ 和“to”意图细节编号 h_4^2 转换为确定意图细节标签 $y_4^2=0$ 的“to”意图细节状态编号 $\hat{y}_4^2$ 。

[0080] 时间步骤5的意图编码器528将“Philadelphia”510的词典编号 x_5 和“flight from Denver to”一般意图状态编号 $\hat{y}_4^1$ 转换为“flight from Denver to Philadelphia”一般意图编号 h_5^1 ,并且在时间步骤5的细节编码器530将“Philadelphia”510的词典编号 x_5 和“to”意图细节状态编号 $\hat{y}_4^2$ 转换为“Philadelphia”意图细节编号 h_5^2 。时间步骤5的意图编码器528将“flight from Denver to Philadelphia”一般意图编号 h_5^1 和“Philadelphia”意图细节编号 h_5^2 转换为确定一般意图类别 $y_{\text{意图}}^1=\text{“flight”}$ 的一般意图状态编号 $\hat{y}_5^1$ 。在时间步骤5的细节编码器530将“flight from Denver to Philadelphia”一般意图编号 h_5^1 和“Philadelphia”意图细节编号 h_5^2 转换为确定意图细节标签 $y_5^2=\text{“到达地”}$ 的细节状态编号 $\hat{y}_5^2$ 。

[0081] 先前的示例是基于双模型结构-具有共享隐藏状态信息的被异步训练用于两个相关任务的两个模型。该概念可以进一步扩展到针对一个任务或多个相关任务异步训练多个模型的情况。除了一般意图模型和意图细节模型之外,自然语言服务器206可以通过与一般意图模型和意图细节模型共享其隐藏状态来添加另一种模型以根据给定的训练数据进行学习。类似地,通过连接三个模型的隐藏状态,一般意图模型和意图细节模型的隐藏状态也可以与新模型共享。如上所述,这三个模型也可以异步训练。

[0082] 例如,基于视觉的查询回答任务的主要目的是找到基于给定图像的问题的答案。典型模型将图像特征和问题特征同时作为输入,并将它们一起训练。自然语言理解引擎214使用不同的异步训练后的模型获取多种类型的数据,并找到它们之间的内部交叉影响。例如,问题“how many surfers(有多少个冲浪者)?”与给定的图像(该图像描绘了在海滩上的四个人,其中一个人拿着冲浪板,另一个人在冲浪时站在冲浪板上)有较强的相关性。自然语言理解引擎216通过识别图像中的“how many surfers”来回应该问题,这可以由使用不同的方法的注意力图来显示,例如通过用冲浪者1标签、冲浪者2标签和冲浪者3标签给图片中的三个被识别为冲浪者的人加标签。自然语言理解引擎216可以提供注意力图,标记目标人群并正确回答,这是典型自然语言理解引擎所无法实现的。

[0083] 图6示出了根据实施例的用于理解自然语言短语的包括图像模型的三模型结构600的框图。在用户说“how many surfers?”并提供了海滩上四个人的图像之后,自然语言理解引擎216将用户的短语解析为“How(有)”602、“many(多少)”604和“surfers(冲浪者)”606,并将图像划分为颜色层608、纹理层610和对比层612。时间步骤1的意图编码器614将“How”602的词典编号 x_1 转换为“how”一般意图编号 h_1^1 ,时间步骤1的细节编码器616将“How”

602的词典编号 x_1 转换为“how”意图细节编号 h_1^2 ,并且时间步骤1的图像编码器618将颜色层608的词典编号 z_1 转换为颜色图像编号 h_1^3 。时间步骤1的意图解码器620将“how”一般意图编号 h_1^1 、“how”意图细节编号 h_1^2 和颜色图像编号 h_1^3 转换为“how”意图状态编号 s_1^1 。在时间步骤1的细节解码器622将“how”一般意图编号 h_1^1 、“how”意图细节编号 h_1^2 和颜色图像编号 h_1^3 转换为“how”细节状态编号 s_1^2 和确定意图细节标签 $y_1^2=0$ 的意图细节编号 $\hat{y}_1^2$ 。在时间步骤1的图像解码器622将“how”一般意图编号 h_1^1 、“how”意图细节编号 h_1^2 和颜色图像编号 h_1^3 转换为“how”图像状态编号 s_1^3 和用冲浪者1标签标记的注意力图 $1y_1^3$ 。

[0084] 在时间步骤2的意图编码器626将“how”604的词典编号 x_2 和“how”一般意图编号 h_1^1 转换为“how many (有多少)”一般意图编号 h_2^1 ,在时间步骤2的细节编码器628将“many”604的词典编号 x_2 和“how”意图细节编号 h_1^2 转换为“many”意图细节编号 h_2^2 ,并且在时间步骤2的图像编码器630将纹理层610的词典编号 z_2 和颜色图像编号 h_1^3 转换为纹理图像编号 h_2^3 。在时间步骤2的意图解码器632将“how many”一般意图编号 h_2^1 、“many”意图细节编号 h_2^2 、纹理图像编号 h_2^3 和“how”意图状态编号 s_1^1 转换为“how many”意图状态编号 s_2^1 。在时间步骤2的细节解码器634将“how many”一般意图编号 h_2^1 、“many”意图细节编号 h_2^2 、纹理图像编号 h_2^3 和“how”细节状态编号 s_1^2 转换为“many”细节状态编号 s_2^2 和确定意图细节标签 $y_2^2=0$ 的“many”意图细节编号 $\hat{y}_2^2$ 。在时间步骤2的图像解码器636将“how many”一般意图编号 h_2^1 、“many”意图细节编号 h_2^2 、纹理图像编号 h_2^3 和“how”图像状态编号 s_1^3 转换为“many”图像状态编号 s_2^3 和用冲浪者2标签标记的纹理注意力图 $2y_2^3$ 。

[0085] 在时间步骤3的意图编码器638将“surfers”606的词典编号 x_3 和“how many”一般意图编号 h_2^1 转换为“how many surfers”一般意图编号 h_3^1 ,在时间步骤3的细节编码器640将“surfers”606的词典编号 x_3 和“many”意图细节编号 h_2^2 转换为“surfers”意图细节编号 h_3^2 ,并且在时间步骤3的图像编码器642将对比层612的词典数字 z_3 和“many”图像编号 h_2^3 转换为对比图像编号 h_3^3 。在时间步骤3的意图解码器644将“how many surfers”一般意图编号 h_3^1 、“surfers”意图细节编号 h_3^2 、对比图像编号 h_3^3 和“many”意图状态编号 s_2^1 转换为确定一般意图类别 $y_{意图}^1$ = “计数”的“how many surfers”一般意图状态编号 $\hat{y}_3^1$ 。在时间步骤3的细节解码器646将“how many surfers”一般意图编号 h_3^1 、“surfers”意图细节编号 h_3^2 、对比图像编号 h_3^3 和“many”细节状态编号 s_2^2 转换为确定意图细节标签 y_3^2 = “surfers”的“surfers”意图细节编号 $\hat{y}_3^2$ 。在时间步骤3的图像解码器648将“how many surfers”一般意图编号 h_3^1 、“surfers”意图细节编号 h_3^2 、对比图像编号 h_3^3 和纹理图像编号 s_2^3 转换为用冲浪者1标签和冲浪者2标签标注的对比注意力图 $3y_3^3$ 。时间步骤1-3的图像编码器(618、630和642)是补充模型230中的补充编码器232的示例,时间步骤1-3的图像解码器(624、636和648)是补充模型230中的补充解码器234的示例。然后,自然语言理解引擎216可以在用户提供的图像中以“2个冲浪者”的计数来回复用户,并可能通过将注意力图1-3彼此叠加来使用注意力图1-3来识别被标记的冲浪者1和冲浪者2的位置。尽管图6描绘了三模型结构内的图像模型,但是该图像模型可以作为一个模型并入双模型结构中,而另一模型则将一般意图模型和意图细节模型组合到文本模型中。与典型的个人数字助理所使用的模型相反,该图像模型将图像和自然语言短语同时作为输入并一起训练。三模型结构600可以使用不同模型获取这些类型的异步训练后的数据,并找到它们之间的内部交叉影响,从而提供更准确的结果。

[0086] 图7示出了根据实施例的用于理解自然语言短语的包括音频记录模型的三模型结构700的框图。对于此示例,系统用户希望收听四个小时时长的在此期间讨论了诉讼的音频记录的每个部分,而不必收听整个记录。在用户说“when is litigation discussed(什么时候讨论了诉讼)?”之后,自然语言理解引擎216将用户的短语解析为“When(什么时候)”702、“is(是)”704、“litigation(诉讼)”706和“discussed(讨论了)”708,并将音频记录解析为音频记录的第一个小时710、音频记录的第二个小时712、音频记录的第三个小时714和音频记录的第四个小时716。在时间步骤1的意图编码器718将“When”702的词典编号 x_1 转换为“when”一般意图编号 h_1^1 ,在时间步骤1的细节编码器720将“When”702的词典编号 x_1 转换为“when”意图细节编号 h_1^2 ,并且在时间步骤1的音频编码器722将音频记录的第一个小时710的音频编号 z_1 转换为第一个小时的音频编号 h_1^3 。在时间步骤1的意图解码器724将“when”一般意图编号 h_1^1 、“when”意图细节编号 h_1^2 和第一个小时的音频编号 h_1^3 转换为“when”意图状态编号 s_1^1 。在时间步骤1的细节解码器726将“when”一般意图编号 h_1^1 、“when”意图细节编号 h_1^2 和第一个小时的音频编号 h_1^3 转换为“when”细节状态编号 s_1^2 和确定意图细节标签 $y_1^2=0$ 的意图细节编号 $\hat{y}_1^2$ 。在时间步骤1的音频解码器728将“when”一般意图编号 h_1^1 、“when”意图细节编号 h_1^2 和第一个小时的音频编号 h_1^3 转换为第一个小时的音频状态编号 s_1^3 和确定第一个小时的音频标签 $y_1^3=0$ 的第一个小时的音频编号 $\hat{y}_1^3$ 。

[0087] 时间步骤2的意图编码器730将“is”704的词典编号 x_2 和“when”一般意图编号 h_1^1 转换为“when is(什么时候是)”一般意图编号 h_2^1 ,时间步骤2的细节编码器732将“is”704的词典编号 x_2 和“when”意图细节 h_1^2 变为“is”意图细节编号 h_2^2 ,并且在时间步骤2的音频编码器734将音频记录的第二个小时712的音频编号 z_2 和第一个小时的音频编号 h_1^3 转换为第二个小时的音频编号 h_2^3 。在时间步骤2的意图解码器736将“when is”一般意图编号 h_2^1 、“is”意图细节编号 h_2^2 、第二个小时的音频编号 h_2^3 和“when”意图状态编号 s_1^1 转换为“when is”意图状态编号 s_2^1 。在时间步骤2的细节解码器738将“when is”一般意图编号 h_2^1 、“is”意图细节编号 h_2^2 、第二个小时的音频编号 h_2^3 和“when”细节状态编号 s_1^2 转换为“is”细节状态编号 s_2^2 和确定意图细节标签 $y_2^2=0$ 的意图细节编号 $\hat{y}_2^2$ 。在时间步骤2的音频解码器740将“when is”一般意图编号 h_2^1 、“is”意图细节编号 h_2^2 、第二个小时的音频状态编号 s_2^3 和第一个小时的音频状态编号 s_1^3 转换为第二个小时的音频状态编号 s_2^3 和确定意图细节标签 $y_2^3=0$ 的意图细节编号 $\hat{y}_2^3$ 。

[0088] 在时间步骤3的意图编码器742将“litigation”706的词典编号 x_3 和“when is”一般意图编号 h_2^1 转换为“when is litigation(什么时候是诉讼)”一般意图编号 h_3^1 ,在时间步骤3的细节编码器744将“litigation”706的词典编号 x_3 和“is”意图细节 h_2^2 转换为“litigation”意图细节编号 h_3^2 ,并且在时间步骤3的音频编码器746将音频记录的第三个小时714的音频编号 z_3 和第二个小时的音频编号 h_2^3 转换为第三个小时的音频编号 h_3^3 。时间步骤3的意图解码器748将“when is litigation”一般意图编号 h_3^1 、“litigation”意图细节编号 h_3^2 、第三个小时的音频编号 h_3^3 和“when is litigation”意图状态编号 s_2^1 转换为“when is litigation”意图状态编号 s_3^1 。在时间步骤3的细节解码器750将“when is litigation”一般意图编号 h_3^1 、“litigation”意图细节编号 h_3^2 、第三个小时的音频编号 h_3^3 和“is”细节状态编号 s_2^2 转换为“litigation”意图状态编号 s_3^2 和确定意图细节标签 $y_3^2=$ “话题(topic)”

的意图细节编号 $\hat{y}_3^2$ 。在时间步骤3的音频解码器752将“when is litigation”一般意图编号 h_3^1 、“litigation”意图细节编号 h_3^2 、第三个小时的音频编号 h_3^3 和第二个小时的音频状态编号 s_2^3 转换为第三个小时音频状态编号 s_3^3 和确定第三个小时的音频标签 $y_3^3=0$ 的第三个小时的音频编号 $\hat{y}_3^3$ 。

[0089] 在时间步骤4的意图编码器754将“discussed”708的词典编号 x_4 和“when is litigation”的一般意图编号 h_3^1 转换为“when is litigation discussed”一般意图编号 h_4^1 ，在时间步骤4的细节编码器756将“discussed”708的词典编号 x_4 和“litigation”意图细节编号 h_3^2 转换为“discussed”意图细节编号 h_4^2 。在时间步骤4的音频编码器758将音频记录的第四个小时716的音频编号 z_4 和第三个小时的音频编号 h_3^3 转换为第四个小时的音频编号 h_4^3 。在时间步骤4的意图解码器760将“when is litigation discussed”一般意图编号 h_4^1 、“discussed”意图细节编号 h_4^2 、第四个小时的音频编号 h_4^3 和“when is litigation”意图状态编号 s_3^1 转换为确定一般意图类别 $y_{意图}^1$ = “时间”的“when is litigation discussed”一般意图编号 $\hat{y}_4^1$ 。在时间步骤4的细节解码器762将“when is litigation discussed”一般意图编号 h_4^1 、“discussed”意图细节编号 h_4^2 、第四个小时的音频编号 h_4^3 和“litigation”细节状态编号 s_3^2 转换为确定意图细节标签 $y_4^2=0$ 的“litigation”意图细节编号 $\hat{y}_4^2$ 。在时间步骤4的音频解码器764将“when is litigation discussed”一般意图编号 h_4^1 、“discussed”意图细节编号 h_4^2 、第四个小时的音频编号 h_4^3 和第三个小时的音频状态编号 s_3^3 转换为确定第四个小时的音频标签 $y_4^3=$ “3:19到3:38的第四个小时的音频编号 $\hat{y}_4^3$ ”。然后，自然语言理解引擎216可以用“音频记录包括从3:19到3:38的诉讼讨论”的回复来回复用户，从而为用户节省了本是用来听这3个小时19分钟录音的三小时以上的时间。尽管图7描绘了三模型结构中的音频模型，但是音频模型可以作为一个模型包含在双模型结构中，而另一模型将一般意图模型和意图细节模型组合到文本模型中。

[0090] 图8示出了根据实施例的用于理解自然语言短语的包括语言模型的三模型结构800的框图。在用户输入文本“Flight from Denver to Philadelphia”之后，自然语言理解引擎216将用户的短语解析为“Flight”802、“from”804、“Denver”806、“to”808和“Philadelphia”810，并使用自动建议的语言模型从基于先前词语的时间步骤1的词汇集812、时间步骤2的词汇集814、时间步骤3的词汇集816、时间步骤4的词汇集818、时间步骤5的词汇集816中选择预测出的词语。在时间步骤1的意图编码器822将“Flight”802的词典编号 x_1 转换为“flight”一般意图编号 h_1^1 ，在时间步骤1的细节编码器824将“Flight”802的词典编号 x_1 转换为“flight”意图细节编号 h_1^2 ，在时间步骤1的建议编码器826将词汇集812的词典编号 z_1 转换为第一词汇编号 h_1^3 。在时间步骤1的意图解码器828将“flight”一般意图编号 h_1^1 、“flight”意图细节编号 h_1^2 和第一词汇编号 h_1^3 转换为“flight”意图状态编号 s_1^1 。在时间步骤1的细节解码器830将“flight”一般意图编号 h_1^1 、“flight”意图细节编号 h_1^2 和第一词汇编号 h_1^3 转换为“flight”细节状态编号 s_1^2 和确定意图细节标签 $y_1^2=0$ 的“flight”意图细节编号 $\hat{y}_1^2$ 。在时间步骤1的建议解码器832将“flight”一般意图编号 h_1^1 、“flight”意图细节编号 h_1^2 和第一词汇编号 h_1^3 转换为第一词汇建议状态编号 s_1^3 和确定第一个词汇标签 $y_1^3=$ “机票”的第一词汇编号数 $\hat{y}_1^3$ 。

[0091] 在时间步骤2的意图编码器834将“from”804的词典编号 x_2 和“flight”一般意图编号 h_1^1 转换为“flight from(来自…的航班)”的一般意图编号 h_2^1 ,在时间步骤2的细节编码器836将“from”804的词典编号 x_2 和“flight”一般意图编号 h_1^2 转换为“from”意图细节编号 h_2^2 ,并且在时间步骤2的建议编码器838将时间步骤2的词汇集814的词典编号 z_2 和第一词汇编号 h_1^3 转换为第二词汇编号 h_2^3 。在时间步骤2的意图解码器840将“flight from”一般意图编号 h_2^1 、“from”意图细节编号 h_2^2 、第二词汇编号 h_2^3 和“flight”意图状态编号 s_1^1 转换为“flight from”意图状态编号 s_2^1 。在时间步骤2的细节解码器842将“flight from”一般意图编号 h_2^1 、“from”意图细节编号 h_2^2 、第二词汇编号 h_2^3 和“flight”细节状态编号 s_1^2 转换为“from”细节状态编号 s_2^2 和确定意图细节标签 $y_2^2=0$ 的“from”意图细节编号 $\hat{y}_2^2$ 。在时间步骤2的建议解码器844将“flight from”一般意图编号 h_2^1 、“from”意图细节编号 h_2^2 、第二词汇编号 h_2^3 和第一词汇状态编号 s_1^3 转换为第二词汇状态编号 s_2^3 和确定第二词汇标签 $y_2^3=$ “状态”的第二词汇编号 $\hat{y}_2^3$ 。

[0092] 在时间步骤3的意图编码器846将“Denver”806的词典编号 x_3 和“flight from”一般意图编号 h_2^1 转换为“flight from Denver”一般意图编号 h_3^1 ,在时间步骤3的细节编码器848将“Denver”306的词典编号 x_3 和“from”意图细节编号 h_2^2 转换为“Denver”意图细节编号 h_3^2 ,并且在时间步骤3的建议编码器850将时间步骤3的词汇集816的词典编号 z_3 和第二词汇编号 h_2^3 转换为第三词汇编号 h_3^3 。在时间步骤3的意图解码器852将“flight from Denver”一般意图编号 h_3^1 、“Denver”意图细节编号 h_3^2 、第三词汇编号 h_3^3 和“from”意图状态编号 s_2^1 转换为“flight from Denver”意图状态编号 s_3^1 。在时间步骤3的细节解码器854将“flight from Denver”一般意图编号 h_3^1 、“Denver”意图细节编号 h_3^2 、第三词汇编号 h_3^3 和“from”细节状态编号 s_2^2 转换为“Denver”细节状态编号 s_3^2 和确定意图细节标签 $y_3^2=$ “出发地”的“Denver”意图细节编号 $\hat{y}_3^2$ 。在时间步骤3的建议解码器856将“flight from Denver”一般意图编号 h_3^1 、“Denver”意图细节编号 h_3^2 、第三词汇编号 h_3^3 和第二词汇状态编号 s_2^3 转换为第三词汇状态编号 s_3^3 和确定第三词汇标签 $y_3^3=$ “时间”的第三词汇编号 $\hat{y}_3^3$ 。

[0093] 在时间步骤4的意图编码器858将“to”808的词典编号 x_4 和“flight from Denver”一般意图编号 h_3^1 转换为“flight from Denver to”一般意图编号 h_4^1 ,在时间步骤4的细节编码器860将“to”808的词典编号 x_4 和“Denver”意图细节编号 h_3^2 转换为“to”意图细节编号 h_4^2 ,并且在时间步骤4的建议编码器862将时间步骤4的词汇集818的词典编号 z_4 和第三词汇编号 h_3^3 转换为第四词汇编号 h_4^3 。在时间步骤4的意图解码器864将“flight from Denver”一般意图编号 h_4^1 、“to”意图细节编号 h_4^2 、第四词汇编号 h_4^3 和“flight from Denver”意图状态编号 s_3^1 转换为“flight from Denver to”意图状态编号 s_4^1 。在时间步骤4的细节解码器866将“flight from Denver to”一般意图编号 h_4^1 、“to”意图细节编号 h_4^2 、第四词汇编号 h_4^3 和“flight”细节状态编号 s_3^2 转换为“to”细节状态编号 s_4^2 和确定意图细节标签 $y_4^2=0$ 的“to”意图细节编号 $\hat{y}_4^2$ 。在时间步骤4的建议解码器868将“flight from Denver to”一般意图编号 h_4^1 、“to”意图细节编号 h_4^2 、第四词汇编号 h_4^3 和第三词汇状态编号 s_3^3 转换为第四词汇状态编号 s_4^3 和确定第四词汇标签 $y_4^3=$ “登机口”的第四词汇编号 $\hat{y}_4^3$ 。

[0094] 在时间步骤5的意图编码器870将“Philadelphia”810的词典编号 x_5 和“flight

from Denver to”一般意图编号 h_4^1 转换为“Flight from Denver to Philadelphia”一般意图编号 h_5^1 ,在时间步骤5的细节编码器872将“Philadelphia”810的词典编号 x_5 和“to”意图细节编号 h_4^2 转换为“Philadelphia”意图细节编号 h_5^2 ,并且在时间步骤5的建议编码器874将时间步骤5的词汇集820的词典编号 z_5 和第四词汇编号 h_4^3 转换为第五词汇编号 h_5^3 。在时间步骤5的意图解码器876将“Flight from Denver to Philadelphia”一般意图编号 h_5^1 、“Philadelphia”意图细节编号 h_5^2 、第五词汇编号 h_5^3 和“flight from Denver to”意图状态编号 s_4^1 转换为确定一般意图类别 $y_{\text{意图}}^1$ = “flight”的“Flight from Denver to Philadelphia”一般意图编号 $\hat{y}_5^1$ 。时间步骤5的细节解码器878将“Flight from Denver to Philadelphia”一般意图编号 h_5^1 、“Philadelphia”意图细节编号 h_5^2 、第五词汇编号 h_5^3 和“to”细节状态编号 s_4^2 转换为确定了意图细节标签 y_5^2 = “到达地”的“Philadelphia”意图细节编号 $\hat{y}_5^2$ 。在时间步骤5的建议解码器880将“Flight from Denver to Philadelphia”一般意图编号 h_5^1 、“Philadelphia”意图细节编号 h_5^2 、第五词汇编号 h_5^3 和第四词汇状态编号 s_4^3 转换为确定第五词汇标签 y_5^3 = 0的第五词汇编号 $\hat{y}_5^3$ 。然后,自然语言理解引擎216可以通过自动建议其他文本输入选项 y_1^3 = “机票”、 y_2^3 = “状态”、 y_3^3 = “时间”和 y_4^3 = “登机口”来响应用户输入文本“Flight from Denver to Philadelphia”。尽管图8描绘了三模型结构中的自动建议模型,但是可以将自动建议作为一个模型并入双模型结构中,而另一模型则将一般意图模型和意图细节模型组合到文本模型中。

[0095] 自然语言服务器206可以将附加信号用于基于视觉的提问回答任务。例如,自然语言服务器206可以使用图像模型、文本问题模型和语音信号模型,这使人们能够通过输入文本和讲话两者来针对给定图像提出问题。通过共享它们的隐藏状态信息,这三个模型可以通过利用来自三个不同数据源的重要信息来同时工作,以提高基于视觉的查询回答任务的性能。

[0096] 图9是示出根据实施例的用于理解自然语言短语的多模型的方法的流程图。流程图900示出了作为图2的客户端202-204和/或服务器206-208中和/或之间所涉及的某些步骤的流程图框示出的方法动作。

[0097] 框902:意图编码器基于接收到短语中的对象确定意图值。系统将词语的词典编号从输入的短语转换为一般意图编号。例如但不限于,这可以包括在训练阶段4的意图编码器330将“to”308的词典编号 x_4 和“flight from Denver”一般意图编号 h_3^1 转换为“flight from Denver to”一般意图编号 h_4^1 。基于图4所示的示例模型,在时间步骤4的意图编码器436将“of”408的词典编号 x_4 和“check the price”一般意图编号 h_3^1 转换为“check the price of”一般意图编号 h_4^1 。

[0098] 基于图6所示的示例模型,在时间步骤2的意图编码器626将“many”604的词典编号 x_2 和“how”一般意图编号 h_1^1 转换为“how many”一般意图编号 h_2^1 。基于图7所示的示例模型,在时间步骤3的意图编码器742将“litigation”706的词典编号 x_3 和“When is”一般意图编号 h_2^1 转换为“When is litigation”一般意图编号 h_3^1 。基于图8所示的示例模型,在时间步骤4的意图编码器858将“to”808的词典编号 x_4 和“Flight from Denver”一般意图编号 h_3^1 转换为“Flight from Denver to”一般意图编号 h_4^1 。

[0099] 框904:在生成一般意图编号时,细节编码器基于接收到的短语中的对象确定细节

值。系统将词语的词典编号从输入的短语转换为意图细节编号。例如但不限于,这可以包括在训练阶段4的细节编码器332将“to”308的词典编号 x_4 和“Denver”意图细节编号 h_3^2 转换为“to”意图细节编号 h_4^2 。基于图4所示的示例模型,在时间步骤4的细节编码器438将“of”408的词典编号 x_4 和“price”意图细节编号 h_3^2 转换为“of”意图细节编号 h_4^2 。

[0100] 基于图6所示的示例模型,在时间步骤2的细节编码器628将“many”604的词典编号 x_2 和“how”意图细节编号 h_1^2 转换为“many”意图细节编号 h_2^2 。基于图7所示的示例模型,在时间步骤3的细节编码器744将“litigation”706的词典编号 x_3 和“is”意图细节编号 h_2^2 转换为“litigation”意图细节编号 h_3^2 。基于图8所示的示例模型,在时间步骤4的细节编码器860将“to”808的词典编号 x_4 和“Denver”意图细节编号 h_3^2 转换为“to”意图细节编号 h_4^2 。

[0101] 框906:在生成一般意图和意图细节的编号的过程中,补充编码器可选地基于接收到的补充对象确定多个补充值。系统将另一个对象的词典编号转换为其他编号,例如图像、音频或建议编号。在实施例中,这可以包括在时间步骤2的图像编码器630将纹理层610的词典编号 z_2 和颜色图像编号 h_1^3 转换为纹理图像编号 h_2^3 。

[0102] 基于图7所示的示例模型,在时间步骤3的音频编码器746将记录的第三个小时714的音频编号 z_3 和第二个小时的音频编号 h_2^3 转换为第三个小时的音频编号 h_3^3 。基于图8所示的示例模型,在时间步骤4的建议编码器862将时间步骤4的词汇集818的词典编号 z_4 和第三词汇编号 h_3^3 转换为第四词汇编号 h_4^3 。

[0103] 框908:在生成了意图值和细节值之后,意图解码器基于意图值和细节值确定意图状态值。系统将隐藏状态转换为一般意图状态编号。例如但不限于,这可以包括在训练阶段4的意图解码器334将“flight from Denver to”一般意图编号 h_4^1 、“to”意图细节编号 h_4^2 和“flight from Denver”意图状态编号 s_3^1 转换为“flight from Denver to”意图状态编号 s_4^1 。基于图4所示的示例模型,在时间步骤4的意图解码器440将“check the price of”一般意图编号 h_4^1 、“of”意图细节编号 h_4^2 和“check the price”意图状态编号 s_3^1 转换为“check the price of”意图状态编号 s_4^1 。

[0104] 基于图6所示的示例模型,时间步骤2的意图解码器632将“how many”一般意图编号 h_2^1 、“many”意图细节编号 h_2^2 、纹理图像编号 h_2^3 和“how”意图状态编号 s_1^1 转换为“how many”意图状态编号 s_2^1 。基于图7所示的示例模型,在时间步骤3的意图解码器748将“when is litigation”一般意图编号 h_3^1 、“litigation”意图细节编号 h_3^2 、第三个小时的音频编号 h_3^3 和“when is”意图状态编号 s_2^1 转换为“when is litigation”意图状态编号 s_3^1 。基于图8所示的示例模型,在时间步骤4的意图解码器864将“flight from Denver to”一般意图编号 h_4^1 、“to”意图细节编号 h_4^2 、第四词汇编号 h_4^3 和“flight from Denver”意图状态编号 s_3^1 转换为“flight from Denver to”意图状态编号 s_4^1 。

[0105] 框910:在生成意图值和细节值之后,细节解码器基于意图值和细节值确定细节状态值和意图细节值。系统将隐藏状态转换为意图细节状态编号和意图细节标签。例如但不限于,这可以包括训练阶段4的细节解码器348将“flight from Denver to”一般意图编号 h_4^1 、“to”意图细节编号 h_4^2 和“flight”细节状态编号 s_3^2 转换为“to”细节状态编号 s_4^2 和确定意图细节标签 $y_4^2=0$ 的意图细节编号 $\hat{y}_4^2$ 。基于图4所示的示例模型,在时间步骤4的细节解码器442将“check the price of”一般意图编号 h_4^1 、“of”意图细节编号 h_4^2 和“price”细节状态编号 s_3^2 转换为“of”细节状态编号 s_4^2 和确定意图细节标签 $y_4^2=0$ 的“of”意图细节编号

$\hat{y}_4^2$ 。

[0106] 基于图6所示的示例模型,在时间步骤2的细节解码器634将“how many”一般意图编号 h_2^1 、“many”意图细节编号 h_2^2 、纹理图像编号 h_2^3 和“how”细节状态编号 s_1^2 转换为“many”细节状态编号 s_2^2 和具有意图细节标签 $y_2^2=0$ 的“many”意图细节编号 $\hat{y}_2^2$ 。基于图7所示的示例模型,在时间步骤3的细节解码器750将“when is litigation”一般意图编号 h_3^1 、“litigation”意图细节编号 h_3^2 、第三个小时的音频编号 h_3^3 和“is”细节状态编号 s_2^2 转换为“litigation”意图状态编号 s_3^2 和确定意图细节标签 $y_3^2=$ “话题”的“litigation”意图细节编号 $\hat{y}_3^2$ 。基于图8所描绘的示例模型,时间步骤4的细节解码器866将“flight from Denver to”一般意图编号 h_4^1 、“to”意图细节编号 h_4^2 、第四词汇编号 h_4^3 和“flight”细节状态编号 s_3^2 转换为“to”细节状态编号 s_4^2 和确定意图细节标签 $y_4^2=0$ 的“to”意图细节编号 $\hat{y}_4^2$ 。

[0107] 框912:在生成意图值和细节值之后,补充解码器可选地基于意图值、细节值和多个补充值确定补充状态值和一个补充值。系统将隐藏状态转换为状态编号和其他模型的标签,例如图像、音频或建议模型。在实施例中,这可以包括在时间步骤的2图像解码器636将“how many”一般意图编号 h_2^1 、“many”意图细节编号 h_2^2 、纹理图像编号 h_2^3 和“how”图像状态编号 s_1^2 转换为“many”图像状态编号 s_2^3 和包括冲浪者2标签的纹理注意力图 $2y_2^3$ 。

[0108] 基于图7所示的示例模型,在时间步骤3的音频解码器752将“when is litigation”一般意图编号 h_3^1 、“litigation”意图细节编号 h_3^2 、第三个小时的音频编号 h_3^3 和第二个小时的音频状态编号 s_2^3 转换为第三个小时的音频状态编号 s_3^3 以及确定第三个小时的音频标签 $y_3^3=0$ 的第三个小时的音频编号 $\hat{y}_3^3$ 。基于图8所示的示例模型,在时间步骤4的建议解码器868将“flight from Denver to”一般意图编号 h_4^1 、“to”意图细节编号 h_4^2 、第四词汇编号 h_4^3 和第三词汇状态编号 s_3^3 转换为第四词汇状态编号 s_4^3 和确定第四词汇标签 $y_4^3=$ “机票”的第四词汇编号 $\hat{y}_4^3$ 。

[0109] 框914:在对一些值进行解码之后,意图编码器基于意图值和接收到的短语中的另一个对象确定其他意图值。系统将词语的词典编号从输入的短语转换为一般意图编号。例如但不限于,这可以包括在训练阶段5的意图编码器336将“Philadelphia”310的词典编号 x_5 和“flight from Denver to”一般意图编号 h_4^1 转换为“flight from Denver to Philadelphia”一般意图编号 h_5^1 。基于图4所示的示例模型,在时间步骤5的意图编码器444将“Apple”410的词典编号 x_5 和“check the price of”一般意图编号 h_4^1 转换为“check the price of Apple”一般意图编号 h_5^1 。

[0110] 基于图6所示的示例模型,在时间步骤3的意图编码器638将“surfers”606的词典编号 x_3 和“how many”一般意图编号 h_2^1 转换为“how many surfers”一般意图编号 h_3^1 。基于图7所示的示例模型,在时间步骤4的意图编码器752将“discussed”708的词典编号 x_4 和“when is litigation”的一般意图编号 h_3^1 转换为“when is litigation discussed”一般意图编号 h_4^1 。基于图8所示的示例模型,在时间步骤5的意图编码器870将“Philadelphia”810的词典编号 x_5 和“flight from Denver to”一般意图编号 h_4^1 转换为“flight from Denver to Philadelphia”一般意图编号 h_5^1 。

[0111] 框916:在解码了一些值之后,细节编码器基于细节值和接收到的短语中的另一对

象确定其他细节值。系统将词语的词典编号从输入的短语转换为意图细节编号。例如但不限于,这可以包括在训练阶段5的细节编码器338将“Philadelphia”310的词典编号 x_5^1 和“to”意图细节编号 h_4^2 转换为“Philadelphia”意图细节编号 h_5^2 。基于图4所示的示例模型,在时间步骤5的细节编码器446将“Apple”310的词典编号 x_5 和“of”意图细节编号 h_4^2 转换为“Apple”意图细节编号 h_5^2 。

[0112] 基于图6所示的示例模型,在时间步骤3的细节编码器640将“surfers”606的词典编号 x_3 和“many”意图细节编号 h_2^2 转换为“surfers”意图细节编号 h_3^2 。基于图7所示的示例模型,在时间步骤4的细节编码器754将“discussed”708的词典编号 x_4 和“litigation”意图细节编号 h_3^2 转换为“discussed”意图细节编号 h_4^2 。基于图8所示的示例模型,在时间步骤5的细节编码器872将“Philadelphia”810的词典编号 x_5 和“to”意图细节编号 h_4^2 转换为“Philadelphia”意图细节编号 h_5^2 。

[0113] 框918:在对一些值进行解码之后,补充编码器可选地基于其他意图值、其他细节值和接收到的补充对象确定其他补充值。系统将另一个对象的词典编号转换为其他编号,例如图像、音频或建议编号。在实施例中,这可以包括在时间步骤3的图像编码器642将对比层612的词典编号 z_3 和“many”图像编号 h_2^3 转换为对比图像编号 h_3^3 。

[0114] 基于图7所示的示例模型,在时间步骤4的音频编码器756将记录的第四个小时716的音频编号 z_4 和第三个小时的音频编号 h_3^3 转换为第四个小时的音频编号 h_4^3 。基于图8所示的示例模型,在时间步骤5的建议编码器874将时间步骤5的词汇集820的词典编号 z_5 和第四词汇编号 h_4^3 转换为第五词汇编号 h_5^3 。

[0115] 框920:在生成了意图值和细节值之后,意图解码器基于其他意图值、其他细节值和意图状态值确定一般意图值。系统识别该一般意图。例如但不限于,这可以包括在训练阶段5的意图解码器340将“flight from Denver to Philadelphia”一般意图编号 h_5^1 、“Philadelphia”意图细节编号 h_5^2 和“flight from Denver to”意图状态编号 s_4^1 转换为确定一般意图类别 $y_{\text{意图}}^1 = \text{“flight”}$ 的一般意图编号 $\hat{y}_5^1$ 。基于图4所示的示例模型,在时间步骤5的意图解码器448将“check the price of Apple”一般意图编号 h_5^1 、“Apple”意图细节编号 h_5^2 以及“check the price of”的意图状态编号 s_4^1 转换为一般意图编号 $\hat{y}_5^1$,该一般意图编号 $\hat{y}_5^1$ 确定一般意图类别 $y_{\text{意图}}^1 = \text{“查看股票价格”}$ 或“查看水果价格”。

[0116] 基于图6所示的示例模型,在时间步骤3的意图解码器644将“how many surfers”一般意图编号 h_3^1 、“surfers”意图细节编号 h_3^2 、对比图像编号 h_3^3 和“many”意图状态编号 s_2^1 转换为确定了一般意图类别 $y_{\text{意图}}^1 = \text{“计数”}$ 的“how many surfers”一般意图状态编号 $\hat{y}_3^1$ 。基于图7所示的示例模型,在时间步骤4的意图解码器758将“when is litigation discussed”一般意图编号 h_4^1 、“discussed”意图细节编号 h_4^2 、第四个小时的音频编号 h_4^3 和“when is litigation”意图状态编号 s_3^1 转换为确定一般意图类别 $y_{\text{意图}}^1 = \text{“时间”}$ 的“when is litigation discussed”一般意图编号 $\hat{y}_4^1$ 。基于图8所示的示例模型,在时间步骤5的意图解码器876将“flight from Denver to Philadelphia”一般意图编号 h_5^1 、“Philadelphia”意图细节编号 h_5^2 、第五词汇编号 h_5^3 和“flight from Denver to”意图状态编号 s_4^1 转换为确定一般意图类别 $y_{\text{意图}}^1 = \text{“flight”}$ 的“flight from Denver to

Philadelphia”一般意图编号 $\hat{y}_5^1$ 。

[0117] 框922:在生成意图值和细节值之后,细节解码器基于其他意图值、其他细节值和细节状态值确定另一意图细节值。系统将隐藏状态转换为意图细节状态编号和意图细节标签。例如但不限于,这可以包括训练阶段5处的细节解码器350将“flight from Denver to Philadelphia”一般意图编号 h_5^1 、“Philadelphia”意图细节编号 h_5^2 和“to”细节状态编号 s_4^2 转换为确定意图细节标签 y_5^2 =“到达地”的意图细节编号 $\hat{y}_5^2$ 。基于图4所示的示例模型,在时间步骤5的细节解码器450将“check the price of Apple”一般意图编号 h_5^1 、“Apple”意图细节编号 h_5^2 、和“of”细节状态编号 s_4^2 转换为意图细节编号 $\hat{y}_5^2$,该意图细节编号 $\hat{y}_5^2$ 确定意图细节标签 y_5^2 =“公司名称”或“水果名称”。

[0118] 基于图6所示的示例模型,在时间步骤3的细节解码器646将“how many surfers”一般意图编号 h_3^1 、“surfers”意图细节编号 h_3^2 、对比图像编号 h_3^3 和“many”细节状态编号 s_2^2 转换为“surfers”意图细节编号 $\hat{y}_3^2$,该意图细节编号 $\hat{y}_3^2$ 确定意图细节标签 y_3^2 =“surfers”。根据图7所示的示例模型,时间步骤5的细节解码器760将“when is litigation discussed”一般意图编号 h_4^1 、“discussed”意图细节编号 h_4^2 、第四个小时的音频编号 h_4^3 和“litigation”细节状态编号 s_3^2 转换为确定意图细节标签 y_4^2 =0的“discussed”意图细节编号 $\hat{y}_4^2$ 。基于图8所示的示例模型,在时间步骤5的细节解码器878将“flight from Denver to Philadelphia”一般意图编号 h_5^1 、“Philadelphia”意图细节编号 h_5^2 、第五词汇编号 h_5^3 和“to”细节状态编号 s_4^2 转换为确定意图细节标签 y_5^2 =“到达地”的“Philadelphia”意图细节编号 $\hat{y}_5^2$ 。

[0119] 框924:除了确定一般意图外,补充解码器还可以根据其他意图值、其他细节值和补充状态值确定另一补充值。系统将隐藏状态转换为另一个模型的标签,例如图像、音频记录或自动建议。在实施例中,这可以包括在时间步骤3的图像解码器648将“how many surfers”一般意图编号 h_3^1 、“surfers”意图细节编号 h_3^2 、对比图像编号 h_3^3 和纹理图像编号 s_2^3 转换为包含冲浪者1标签和冲浪者2标签的对比注意力图 $3y_3^3$ 。

[0120] 基于图7所示的示例模型,在时间步骤4的音频解码器762将“when is litigation discussed”一般意图编号 h_4^1 、“discussed”意图细节编号 h_4^2 、第四个小时的音频编号 h_4^3 和第三个小时的音频状态编号 s_3^3 转换为确定第四个小时的音频标签 y_4^3 =“3:19至3:38”的第四个小时的音频编号 $\hat{y}_4^3$ 。基于图8所示的示例模型,在时间步骤5的建议解码器880将“flight from Denver to Philadelphia”一般意图编号 h_5^1 、“Philadelphia”意图细节编号 h_5^2 、第五词汇编号 h_5^3 和第四词汇状态编号 s_4^3 转换为确定第五词汇标签 y_5^3 =0的第五词汇编号 $\hat{y}_5^3$ 。

[0121] 框926:在确定了一般意图值之后,基于确定的一般意图值和经验证的一般意图之间的差,可选地在意图编码器和/或意图解码器中修改参数。系统基于一般意图类别的正确性训练一般意图模型。例如但不限于,这可以包括训练器212基于 $y_{\text{意图}}^1$ =“flight”来计算损失函数,以及系统管理员验证短语“flight from Denver to Philadelphia”的一般意图类别“flight”。因此,训练器212对意图编码器或意图解码器的参数进行最小的改变(如果有

的话)。

[0122] 框928:在确定了各种意图细节值之后,可选地在细节编码器和/或细节解码器中基于确定的意图细节值和/或确定的其他意图细节值与至少一个经验证的意图细节之间的差来修改参数。系统基于意图细节标签的正确性来训练意图细节模型。例如但不限于,这可以包括训练器212基于 $y_1^2 = \text{"0"}$ 、 $y_3^2 = \text{"出发地"}$ 、 $y_5^2 = \text{"到达地"}$ 来计算损失函数,以及系统管理员仅验证短语“flight from Denver to Philadelphia”的意图细节标签“出发地”和“到达地”。因此,训练器212改变细节编码器和/或细节解码器中的参数,以减小生成意图细节标签 $y_1^2 = \text{"0"}$ 的概率。

[0123] 框930:在多个模型理解了接收到的短语之后,可选地输出对接收到的短语的响应,该响应是基于确定的一般意图值、确定的意图细节值和/或确定的其他意图细节值。系统输出对多个模型所理解的自然语言短语的响应。在实施例中,这可以包括自然语言理解引擎216以回答“您什么时候要订购苹果这种水果(When do you want to order delivery of the fruit apple)?”或“您想购买多少苹果公司的股票(How much stock do you want to purchase in the company Apple)?”来回复用户。

[0124] 基于图6所示的示例模型,自然语言理解引擎216通过用户提供的图像中的“2个冲浪者”的计数来响应用户,并使用注意力图1-3识别标注了冲浪者1和2的位置。基于图7所示的示例模型,自然语言理解引擎216以“音频记录包括从3:19到3:38的诉讼讨论”的答复对用户做出响应,从而为用户节省了本是用来听这3个小时19分钟录音的三小时以上的时间。基于图8所示的示例模型,自然语言理解引擎216通过自动建议其他文本输入选项 $y_1^3 = \text{"机票"}$ 、 $y_2^3 = \text{"状态"}$ 、 $y_3^3 = \text{"时间"}$ 、 $y_4^3 = \text{"登机口"}$ 来响应用户输入的文本“flight from Denver to Philadelphia”。

[0125] 尽管图9描绘了以特定顺序出现的框902-930,但是框902-930可以以另一顺序出现。在其他实施方式中,框902-930中的每个框也可以与其他框组合执行和/或一些框可以划分为不同的框集合。在其他实施方式中,可以省略框902-930中的至少一个。例如,可以省略框906、912、918、924、926、928和930。

[0126] 根据本公开的实施例,一种方法包括:意图编码器基于接收到的短语中的对象确定意图值;细节编码器基于接收到的短语中的对象确定细节值;意图解码器基于所述意图值和所述细节值确定意图状态值;所述意图编码器基于所述意图值和接收到的短语中的另一对象确定其他意图值;所述细节编码器基于所述细节值和接收到的短语中的另一对象确定其他细节值;所述意图解码器基于所述其他意图值、所述其他细节值和所述意图状态值确定一般意图值;以及所述细节解码器基于所述其他意图值、所述其他细节值和所述细节状态值确定另一意图细节值。

[0127] 根据本公开的实施例,所述方法还包括:基于确定的一般意图值与经验证的一般意图之间的差来修改所述意图编码器和所述意图解码器中的至少一个中的参数;或者基于确定的意图细节值和确定的其他意图细节值中的至少一个与至少一个经验证的意图细节之间的差来修改所述细节编码器和所述细节解码器中的至少一个中的参数。

[0128] 根据本公开的实施例,该方法还包括:输出对接收到的短语的响应,所述响应基于确定的一般意图值、确定的意图细节值和确定的其他意图细节值中的至少一个。

[0129] 根据本公开的实施例,该方法还包括:补充编码器基于接收到的补充对象确定多

个补充值；补充解码器基于所述意图值、所述细节值和所述多个补充值确定补充状态值和一个补充值；所述补充编码器基于所述意图值、所述细节值和接收到的补充对象确定其他补充值；以及所述补充解码器基于所述其他意图值、所述其他细节值和所述补充状态值来确定另一补充值。

[0130] 根据本公开的实施例，所述补充编码器包括图像编码器，所述多个补充值包括图像值，接收到的补充对象包括图像对象，所述补充解码器包括图像解码器，所述补充状态值包括图像状态值，所述一个补充值包括图像值，所述其他补充值包括其他图像值，并且所述另一补充值包括另一图像值。

[0131] 根据本公开的实施例，所述补充编码器之一包括音频编码器，所述多个补充值包括音频值，接收到的补充对象包括音频对象，所述补充解码器包括音频解码器，所述补充状态值包括音频状态值，所述一个补充值包括音频值，所述其他补充值包括其他音频值，所述另一补充值包括另一音频值，所述补充编码器包括建议编码器，所述多个补充值包括建议值，接收到的补充对象包括建议对象，所述补充解码器包括建议解码器，所述补充状态值包括建议状态值，所述一个补充值包括建议值，所述其他补充值包括其他建议值，并且所述另一补充值包括另一建议值。

[0132] 根据本公开的实施例，一种方法包括：由意图编码器基于接收到的短语中的对象来确定意图值；细节编码器基于接收到的短语中的对象确定细节值；意图解码器基于所述意图值和所述细节值确定意图状态值；所述意图编码器基于所述意图值和接收到的短语中的另一个对象确定其他意图值；所述细节编码器基于所述细节值和接收到的短语中的另一个对象确定其他细节值；所述意图解码器基于所述其他意图值、所述其他细节值和所述意图状态值确定一般意图值；基于确定的一般意图值和经验证的一般意图之间的差来修改所述意图编码器和所述意图解码器中的至少一个中的参数。

[0133] 根据本公开的实施例，所述方法还包括：细节解码器基于所述意图值和所述细节值确定细节状态值和意图细节值；所述细节解码器基于所述其他意图值、所述其他细节值和所述细节状态值确定另一意图细节值；以及基于确定的意图细节值和确定的其他意图细节值中的至少一个与至少一个经验证的意图细节之间的差来修改所述细节编码器和所述细节解码器中的至少一个中的参数。

[0134] 将描述可以在其中实现主题的示例性硬件设备。本领域普通技术人员将理解，图10所示的元件可以根据系统实现而变化。参考图10，用于实现本文公开的主题的示例性系统包括硬件设备1000，其包括处理器1002、存储器1004、存储装置1006、数据输入模块1008、显示适配器1010、通信接口1012、以及将元件1004-1012耦接到处理器1002的总线1014。

[0135] 总线1014可以包括任何类型的总线架构。示例包括存储器总线、外围总线、局部总线等。处理器1002是指令执行机器、装置或设备，并且可以包括微处理器、数字信号处理器、图形处理单元、专用集成电路(ASIC)、现场可编程门阵列(FPGA)等。处理器1002可以配置为执行存储在存储器1004和/或存储装置1006中和/或通过数据输入模块1008接收的程序指令。在本公开的实施例中，处理器1002可以包括或控制意图解码器(例如，意图编码器220)、意图解码器(例如，意图解码器22)、细节编码器(例如，细节编码器226)、细节解码器(例如，细节解码器228)、补充编码器(例如，补充编码器232)和/或补充解码器(例如，补充解码器234)。

[0136] 存储器1004可以包括只读存储器(ROM) 1016和随机存取存储器(RAM) 1018。存储器1004可以被配置为在设备1000的操作期间存储程序指令和数据。在各个实施例中,存储器1004可以包括多种存储技术中的任何一种,例如静态随机存取存储器(SRAM)或动态RAM(DRAM),例如包括诸如双数据速率同步DRAM(DDR SDRAM)、纠错码同步DRAM(ECC SDRAM)或RAMBUS DRAM(RDRAM)。存储器1004还可包括诸如非易失性闪存RAM(NVRAM)或ROM的非易失性存储技术。在一些实施例中,可以预期的是,存储器1004可以包括诸如前述的技术以及未特别提及的其他技术的组合。当主题在计算机系统中实现时,包含帮助在计算机系统内的元件之间(例如在启动期间)传递信息的基本例程的基本输入/输出系统(BIOS) 1020被存储在ROM 1016中。

[0137] 存储装置1006可以包括:闪存数据存储设备,用于从闪存读取和写入到闪存;硬盘驱动器,用于从硬盘读取和写入到硬盘;磁盘驱动器,用于从磁盘驱动器读取或写入到磁盘驱动器;和/或用于从写入可移动光盘(例如CD ROM,DVD或其他光学介质)读取或写入到可移动光盘的光盘驱动器。驱动器及其相关的计算机可读介质为硬件设备1000提供了计算机可读指令、数据结构、程序模块和其他数据的非易失性存储。

[0138] 注意,本文描述的方法可以体现在存储在计算机可读介质中的可执行指令中,以供指令执行机器、装置或设备(例如基于计算机或包含处理器的机器、装置或设备)使用或与其结合使用。本领域技术人员将认识到,对于一些实施例,可以使用其他类型的计算机可读介质,其可以存储可由计算机访问的数据,例如磁带、闪存卡、数字视频光盘、伯努利墨盒等。在示例性操作环境中,RAM、ROM等也可以被使用。如这里所使用的,“计算机可读介质”可以包括任何适当的介质中的一个或多个,用于以电子、磁性、光学和电磁格式中的一种或多种存储计算机程序的可执行指令,从而使得指令执行机器、系统、装置或设备可以从计算机可读介质读取(或获取)指令并执行用于执行所描述的方法的指令。常规示例性计算机可读介质的非详尽列表包括:便携式计算机软盘;RAM;ROM;可擦可编程只读存储器(EPROM或闪存);光学存储设备,包括便携式光盘(CD)、便携式数字视频光盘(DVD)、高清DVD(HD-DVD™)和蓝光光盘;等等。

[0139] 多个程序模块可以存储在存储装置1006、ROM 1016或RAM 1018上,包括操作系统1022、一个或更多个应用程序1024、程序数据1026和其他程序模块1028。用户可以将命令和信息通过数据输入模块1008输入到硬件设备1000中。数据输入模块1008可以包括诸如键盘、触摸屏、定点设备等的机构。其他外部输入设备(未显示)通过外部数据输入接口1030连接到硬件设备1000。例如但不限于,外部输入设备可以包括麦克风、操纵杆、游戏手柄、碟形卫星天线、扫描仪等。在一些实施例中,外部输入设备可以包括视频或音频输入设备,诸如摄像机、静态照相机等。数据输入模块1008可以被配置为从设备1000的一个或更多个用户接收输入并将输入通过总线1014传送到处理器1002和/或存储器1004。

[0140] 显示器1032也经由显示器适配器1010连接到总线1014。显示器1032可以被配置为向一个或更多个用户显示设备1000的输出。在一些实施例中,例如触摸屏的给定设备可以充当数据输入模块1008和显示器1032两者。外部显示设备还可以经由外部显示接口1034连接到总线1014。未示出的输出设备(例如,扬声器和打印机)可以连接到硬件设备1000。

[0141] 硬件设备1000可以使用经由通信接口1012到一个或更多个远程节点(未示出)的逻辑连接在联网环境中操作。远程节点可以是另一台计算机、服务器、路由器、对等设备或

其他公共网络节点,并且通常包括上述相对于硬件设备1000的许多或所有元件。通信接口1012可以与无线网络和/或有线网络接口。无线网络的示例包括例如蓝牙网络、无线个人局域网、无线802.11局域网(LAN)和/或无线网络(例如,蜂窝、PCS或GSM网络)。有线网络的示例包括例如LAN、光纤网络、有线个人局域网、电话网络和/或广域网(WAN)。这样的联网环境在企业内部网、因特网、办公室、企业范围的计算机网络等中是常见的。在一些实施例中,通信接口1012可以包括被配置为支持存储器1004与其他设备之间的直接存储器访问(DMA)传输的逻辑。

[0142] 在联网环境中,相对于硬件设备1000或其部分描述的程序模块可以存储在远程存储设备中,例如存储在服务器上。将理解的是,可以使用在硬件设备1000和其他设备之间建立通信链路的其他硬件和/或软件。

[0143] 应当理解,图10所示的硬件设备1000的布置仅仅是一种可能的实现方式,并且其他布置也是可能的。还应该理解的是,由权利要求书所限定的,下面描述的并且在各种框图中示出的各种系统组件(和装置)表示被配置为执行本文描述的功能的逻辑组件。例如,这些系统组件(和装置)中的一个或多个可以全部或部分地通过硬件设备1000的布置中所示的至少一些组件来实现。

[0144] 根据本发明的一个实施例,一种系统包括一个或多个处理器;存储多个指令的非暂时性计算机可读介质,所述指令在执行时使所述一个或多个处理器执行以下操作:意图编码器基于接收到的短语中的对象确定意图值;细节编码器基于接收到的短语中的对象确定细节值;意图解码器基于所述意图值和所述细节值确定意图状态值;细节解码器基于所述意图值和所述细节值确定细节状态值和意图细节值;所述意图编码器基于所述意图值和所述接收到的短语中的另一个对象确定其他意图值;所述细节编码器基于所述细节值和所述接收到的短语中的另一个对象确定其他细节值;所述意图解码器基于所述其他意图值、所述其他细节值和所述意图状态值确定一般意图值;以及所述细节解码器基于所述其他意图值、所述其他细节值和所述细节状态值确定另一意图细节值。

[0145] 根据本公开的实施例,所述多个指令在被执行时使所述一个或多个处理器进一步执行以下至少一个操作:基于确定的一般意图值与经验证的一般意图之间的差来修改所述意图编码器和所述意图解码器中的至少一个中的参数;或者基于确定的意图细节值和确定的其他意图细节值中的至少一个与至少一个经验证的意图细节之间的差来修改所述细节编码器和所述细节解码器中的至少一个中的参数。

[0146] 根据本公开的实施例,所述多个指令在被执行时使所述一个或多个处理器进一步执行以下操作:输出对接收到的短语的响应,所述响应基于确定的一般意图值、确定的意图细节值和确定的其他意图细节值中的至少一个。

[0147] 根据本公开的实施例,所述多个指令在被执行时将进一步使所述一个或多个处理器执行以下操作:补充编码器基于接收到的补充对象确定多个补充值;补充解码器基于所述意图值、所述细节值和所述多个补充值确定补充状态值和一个补充值;所述补充编码器基于所述意图值、所述细节值和接收到的补充对象确定其他补充值;以及所述补充解码器基于所述其他意图值、所述其他细节值和所述补充状态值来确定另一补充值。

[0148] 根据本公开的实施例,所述补充编码器包括图像编码器,所述多个补充值包括图像值,接收到的补充对象包括图像对象,所述补充解码器包括图像解码器,所述补充状态值

包括图像状态值,所述一个补充值包括图像值,所述其他补充值包括其他图像值,并且所述另一补充值包括另一图像值。

[0149] 根据本公开的实施例所述补充编码器之一包括音频编码器,所述多个补充值包括音频值,接收到的补充对象包括音频对象,所述补充解码器包括音频解码器,所述补充状态值包括音频状态值,所述一个补充值包括音频值,所述其他补充值包括其他音频值,所述另一补充值包括另一音频值,所述补充编码器包括建议编码器,所述多个补充值包括建议值,接收到的补充对象包括建议对象,所述补充解码器包括建议解码器,所述补充状态值包括建议状态值,所述一个补充值包括建议值,所述其他补充值包括其他建议值,并且所述另一补充值包括另一建议值。

[0150] 根据本公开的实施例,一种计算机程序产品包括当从非暂时性计算机可读介质检索时由一个或多个处理器执行的计算机可读程序代码。该程序代码包括指令以便:意图编码器基于接收到的短语中的对象确定意图值;细节编码器基于接收到的短语中的对象确定细节值;意图解码器基于所述意图值和所述细节值确定意图状态值;细节解码器基于所述意图值和所述细节值确定细节状态值和意图细节值;所述意图编码器基于所述意图值和所述接收到的短语中的另一个对象确定其他意图值;所述细节编码器基于所述细节值和所述接收到的短语中的另一个对象确定其他细节值;所述意图解码器基于所述其他意图值、所述其他细节值和所述意图状态值确定一般意图值;以及所述细节解码器基于所述其他意图值、所述其他细节值和所述细节状态值确定另一意图细节值。

[0151] 根据本公开的实施例,所述程序代码还包括指令以便:基于确定的一般意图值与经验证的一般意图之间的差来修改所述意图编码器和所述意图解码器中的至少一个中的参数;或者基于确定的意图细节值和确定的其他意图细节值中的至少一个与至少一个经验证的意图细节之间的差来修改所述细节编码器和所述细节解码器中的至少一个中的参数。

[0152] 根据本公开的实施例,程序代码还包括指令以便:输出对接收到的短语的响应,所述响应基于确定的一般意图值、确定的意图细节值和确定的其他意图细节值中的至少一个。

[0153] 根据本公开的实施例,程序代码还包括指令以便:补充编码器基于接收到的补充对象确定多个补充值;补充解码器基于所述意图值、所述细节值和所述多个补充值确定补充状态值和一个补充值;所述补充编码器基于所述意图值、所述细节值和接收到的补充对象确定其他补充值;以及所述补充解码器基于所述其他意图值、所述其他细节值和所述补充状态值来确定另一补充值。

[0154] 根据本公开的实施例,所述补充编码器包括图像编码器,所述多个补充值包括图像值,接收到的补充对象包括图像对象,所述补充解码器包括图像解码器,所述补充状态值包括图像状态值,所述一个补充值包括图像值,所述其他补充值包括其他图像值,并且所述另一补充值包括另一图像值。

[0155] 根据本公开的实施例,所述补充编码器之一包括音频编码器,所述多个补充值包括音频值,接收到的补充对象包括音频对象,所述补充解码器包括音频解码器,所述补充状态值包括音频状态值,所述一个补充值包括音频值,所述其他补充值包括其他音频值,所述另一补充值包括另一音频值,所述补充编码器包括建议编码器,所述多个补充值包括建议值,接收到的补充对象包括建议对象,所述补充解码器包括建议解码器,所述补充状态值包

括建议状态值,所述一个补充值包括建议值,所述其他补充值包括其他建议值,并且所述另一补充值包括另一建议值。

[0156] 另外,尽管这些组件中的至少一个至少部分地被实现为电子硬件组件,并因此构成机器,但是其他组件可以以软件、硬件、或软件和硬件的组合来实现。更具体地,由权利要求书限定的至少一个组件至少部分地被实现为电子硬件组件,诸如指令执行机器(例如,基于处理器或包含处理器的机器)和/或专用电路或电路(例如,相互连接以执行特定功能的离散逻辑门),如图10所示。

[0157] 其他组件可以用软件、硬件、或软件和硬件的组合来实现。而且,可以组合这些其他组件中的一些或全部,可以完全省略其中的一些,并且可以添加附加组件,同时仍实现本文所述的功能。因此,本文描述的主题可以以许多不同的变形来体现,并且所有这样的变体都被认为在所要求保护的范围内。

[0158] 在上面的描述中,除非另外指出,否则将参考由一个或多个设备执行的动作为和符号表示来描述主题。这样,可以理解,有时被称为计算机执行的这种动作和操作包括处理单元对结构化形式的数据的操纵。该操作转换数据或将其保存在计算机的存储系统中的位置,以本领域技术人员熟知的方式重新配置或以其他方式改变设备的操作。维护数据的数据结构是存储器的物理位置,具有由数据格式定义的特定属性。然而,尽管在上下文中描述了主题,但是这并不意味着是限制性的,因为本领域技术人员将理解,下文中描述的各种动作和操作也可以以硬件来实现。

[0159] 为了促进对上述主题的理解,根据动作序列描述了许多方面。由权利要求书限定的这些方面中的至少一个是由电子硬件组件执行的。例如,将认识到,各种动作可以由专用电路或电路系统,由一个或多个处理器执行的程序指令,或两者的组合来执行。本文对任何动作序列的描述并不旨在暗示必须遵循为执行该序列而描述的特定顺序。除非本文另外指出或与上下文明显矛盾,否则本文描述的所有方法可以以任何合适的顺序执行。

[0160] 尽管已经通过示例的方式并且根据特定实施例描述了一种或多种实施方式,但是应当理解,一种或多种实施方式不限于所公开的实施方式。相反,本发明意图涵盖对本领域技术人员显而易见的各种修改和类似布置。因此,所附权利要求书的范围应被赋予最宽泛的解释,以涵盖所有这样的修改和类似的布置。

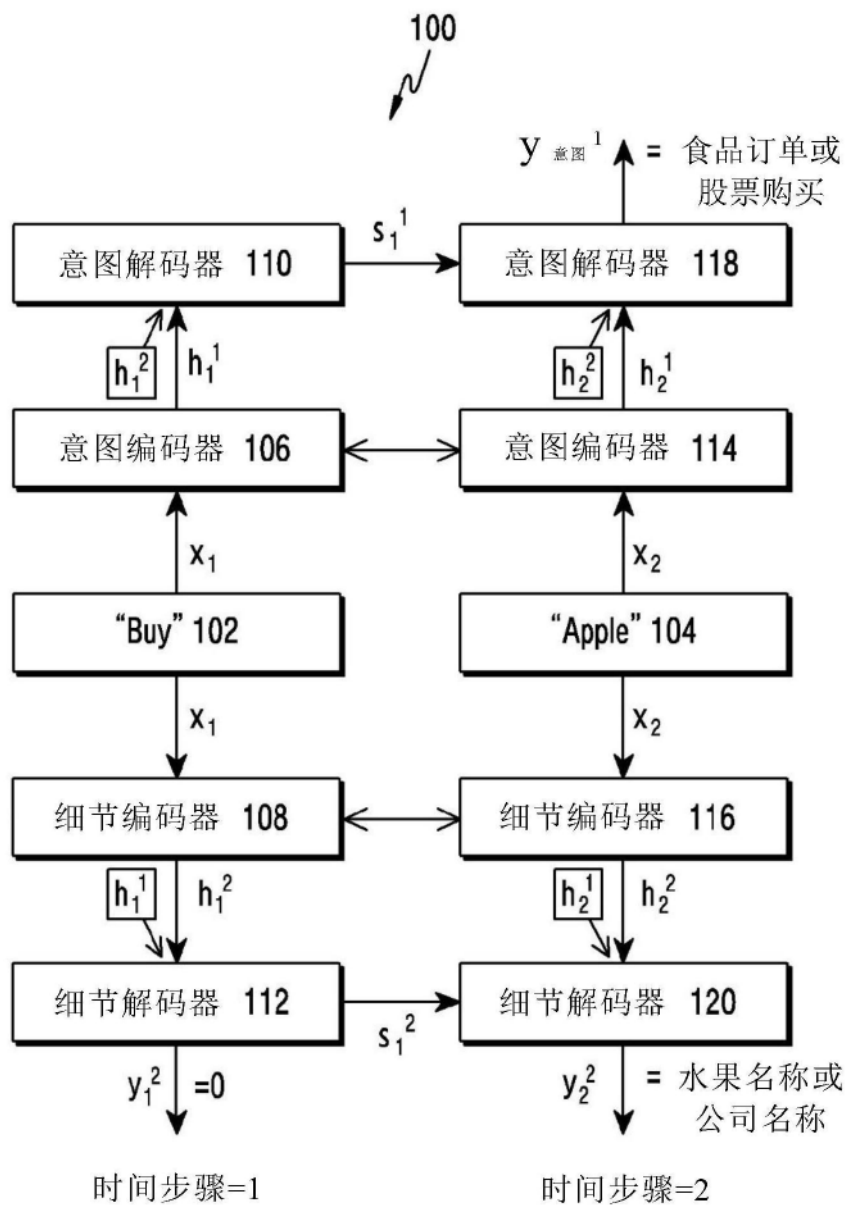


图1

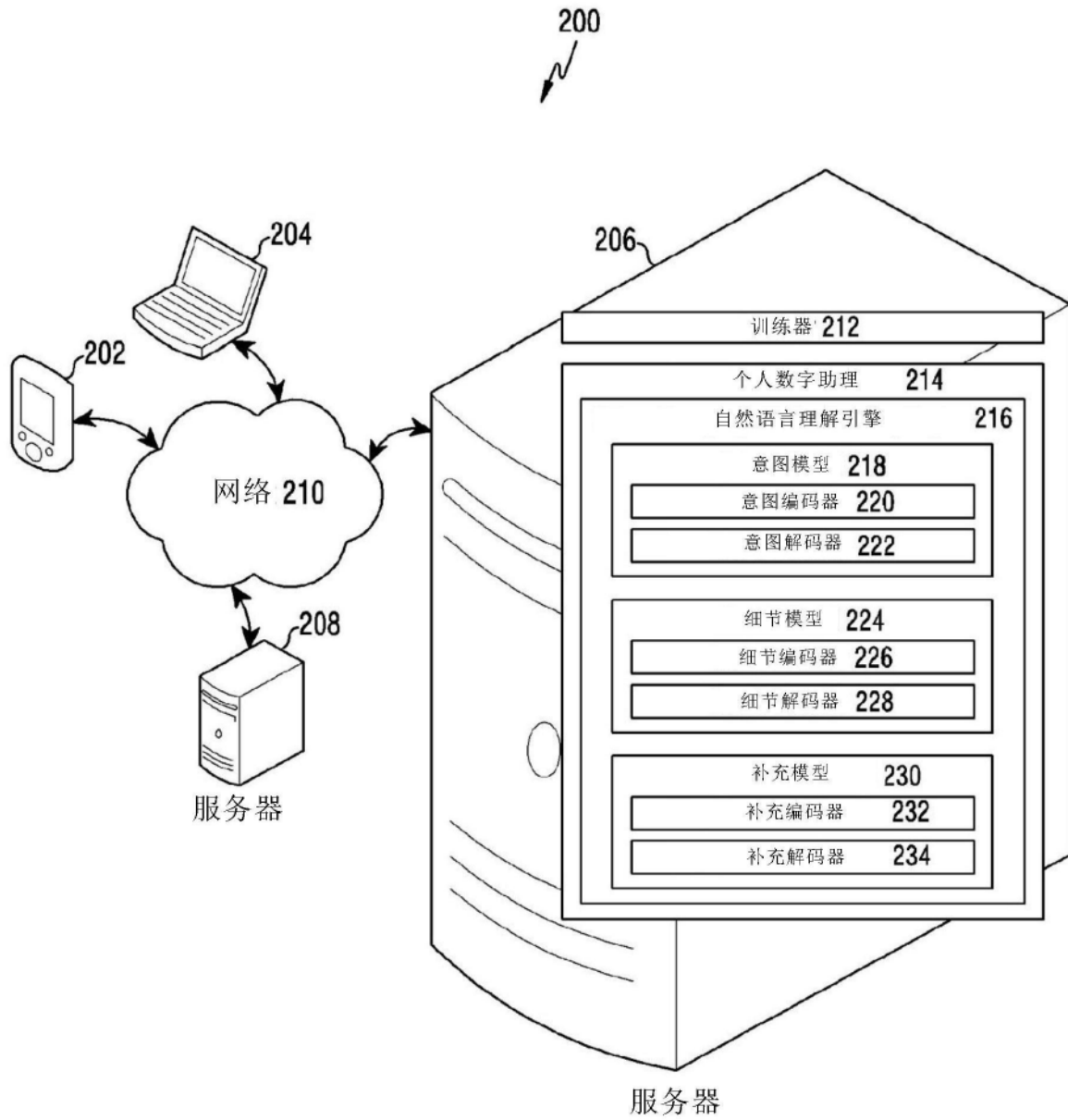


图2

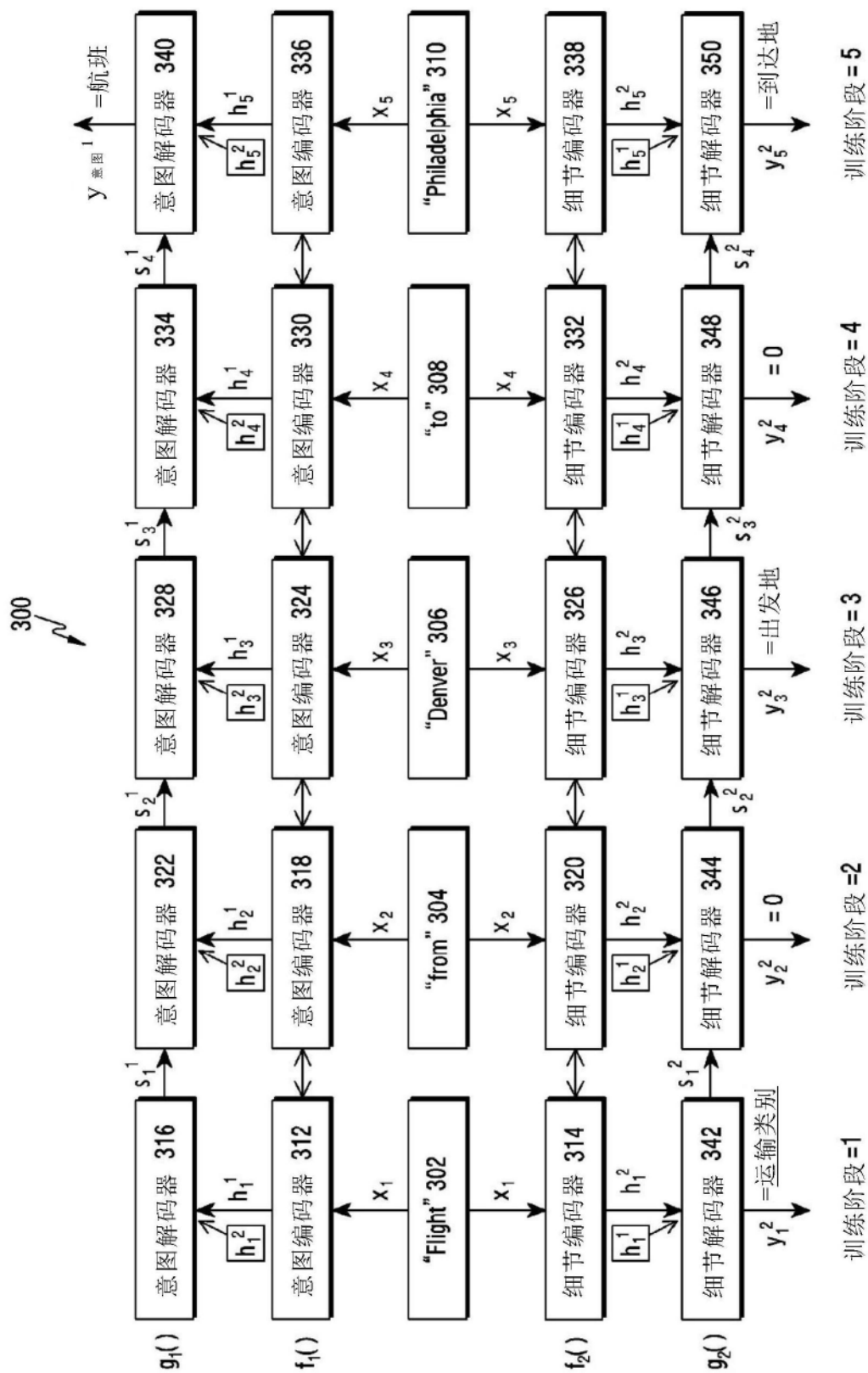


图3

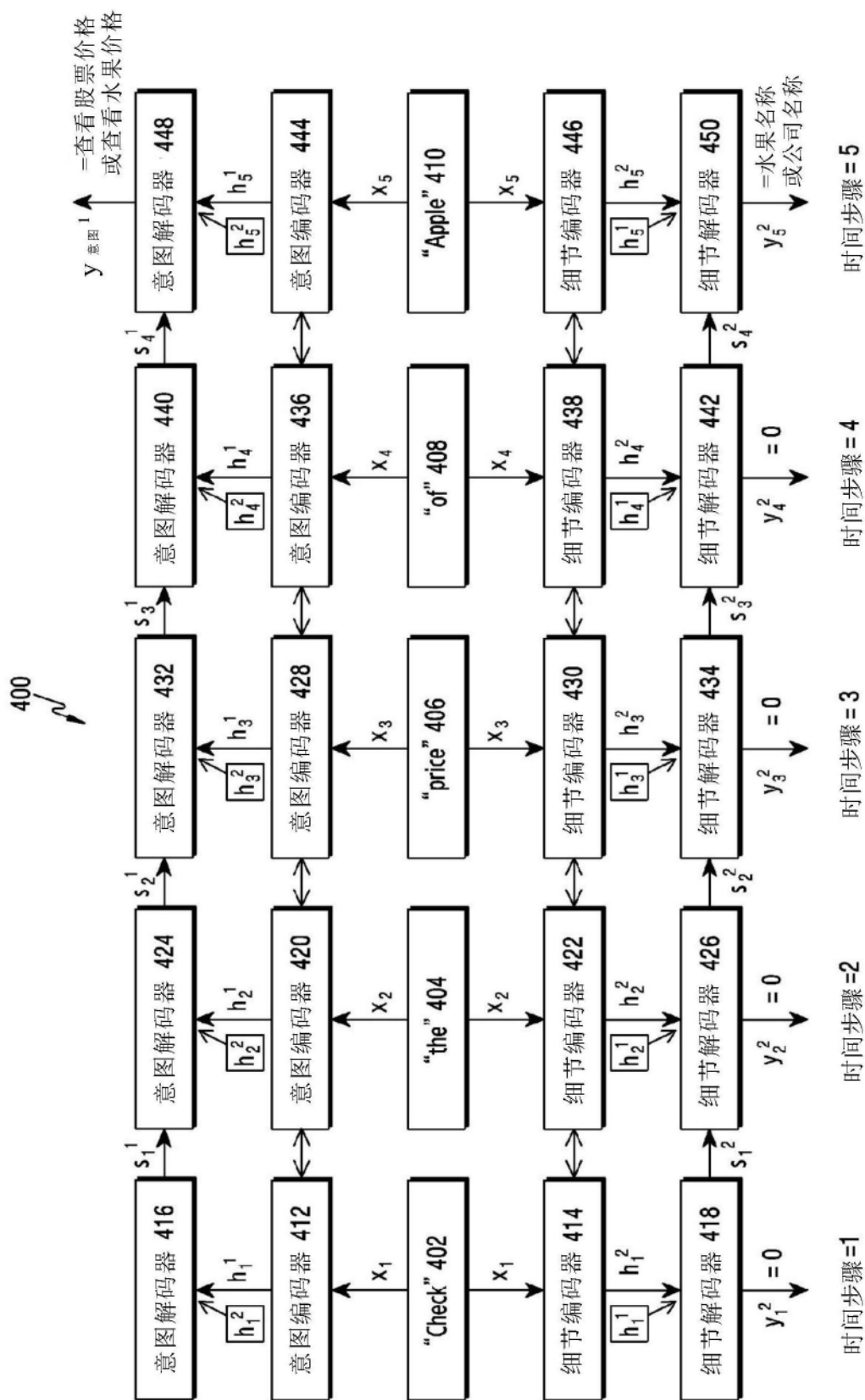


图4

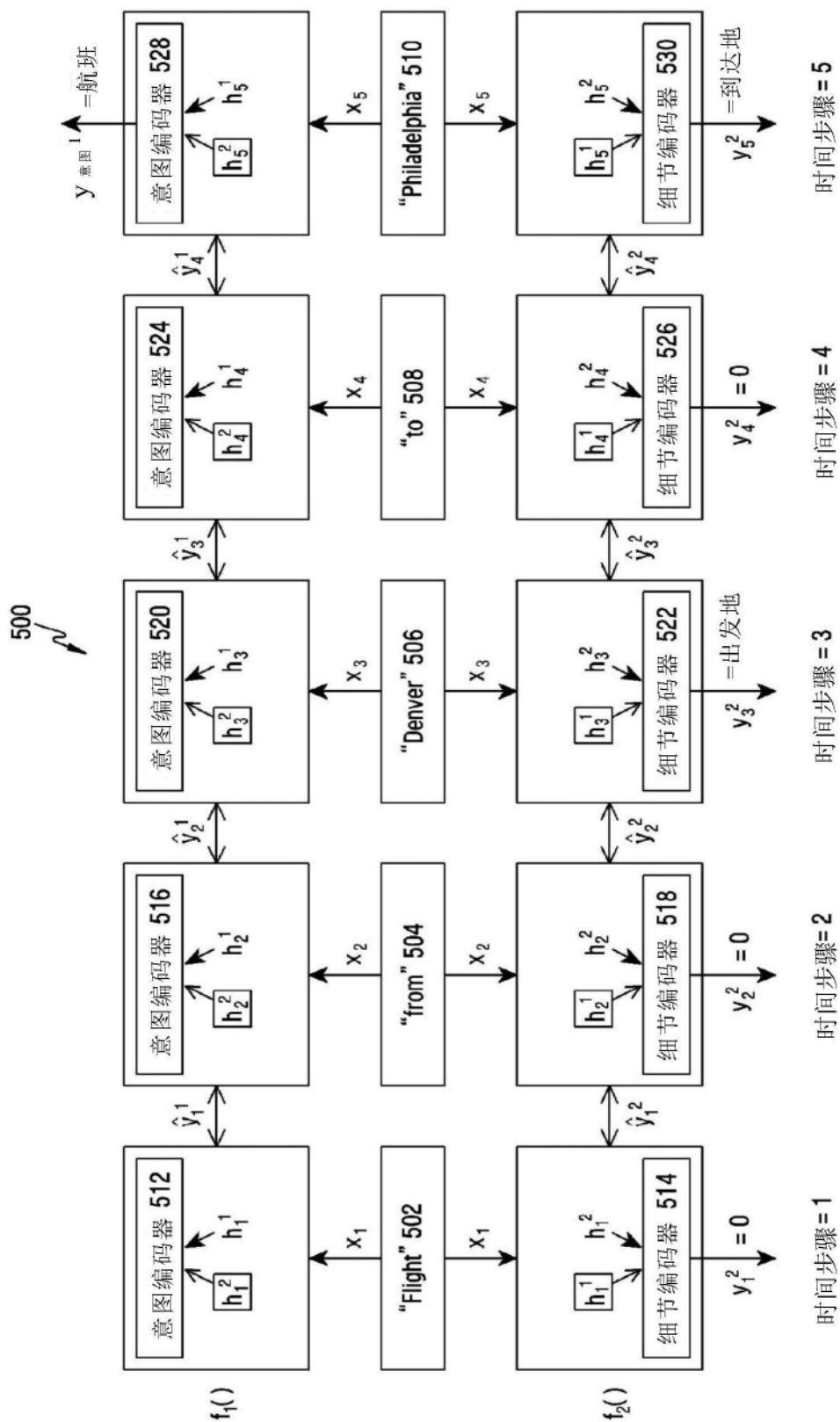


图5

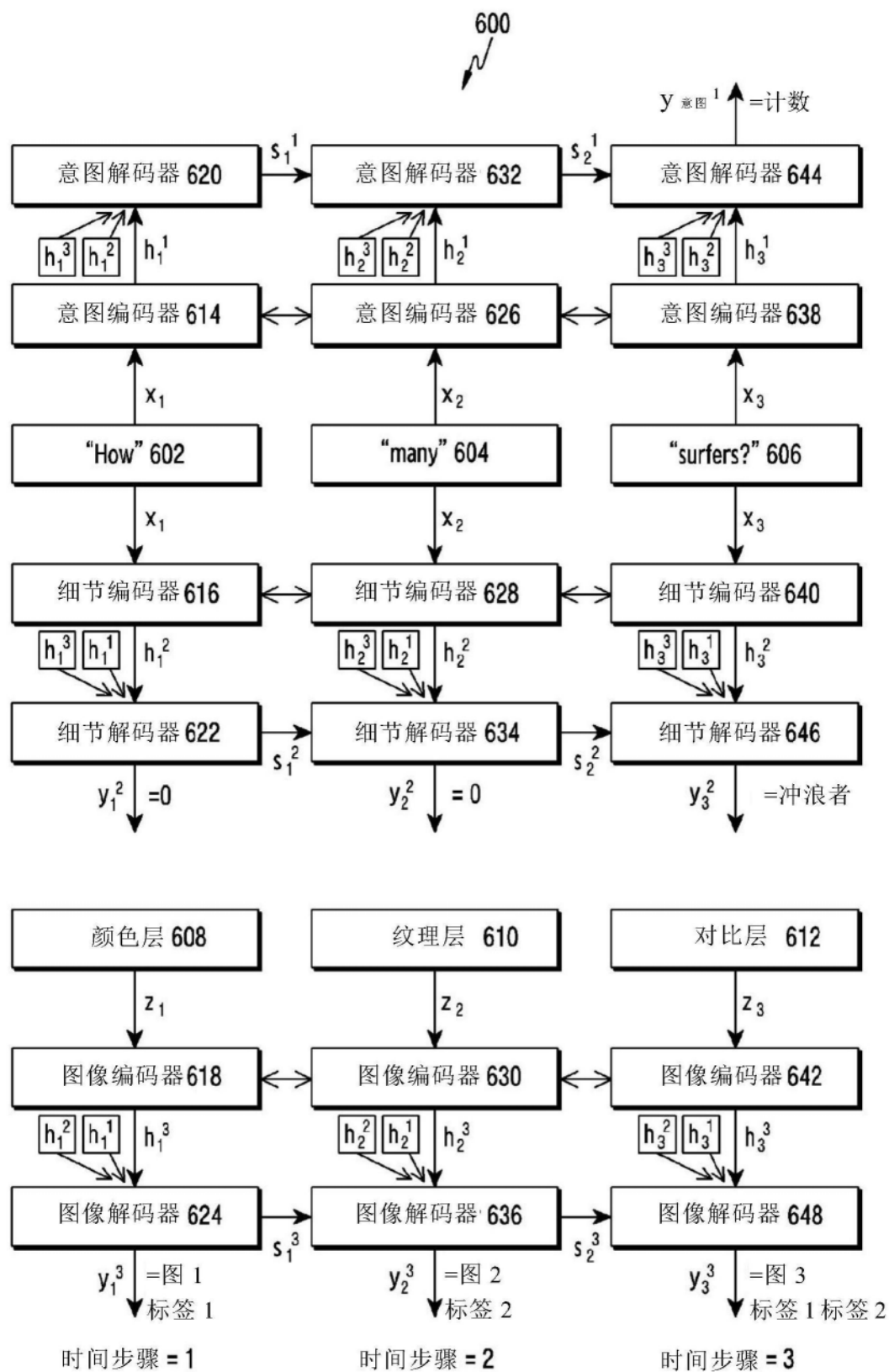


图6

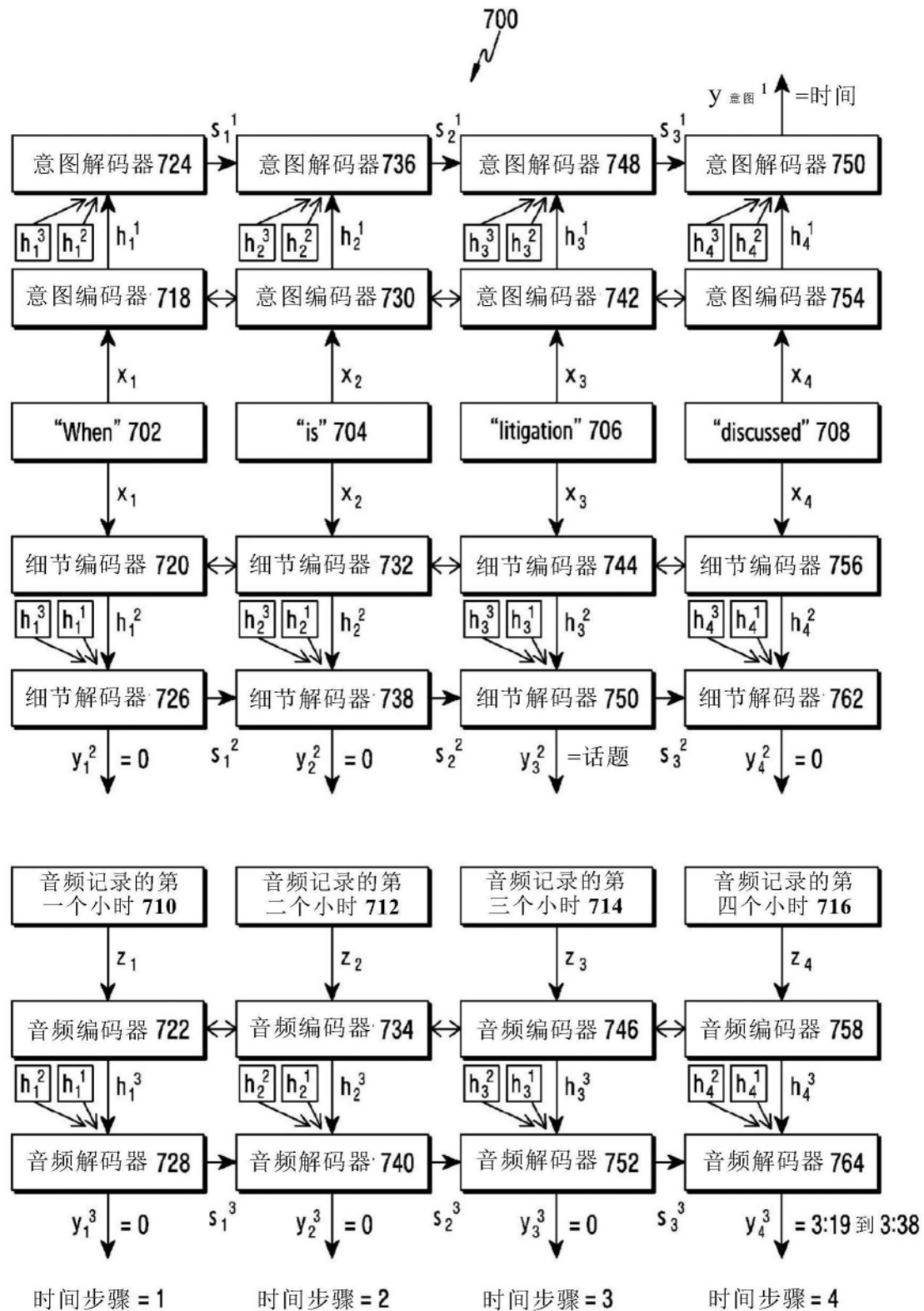


图7

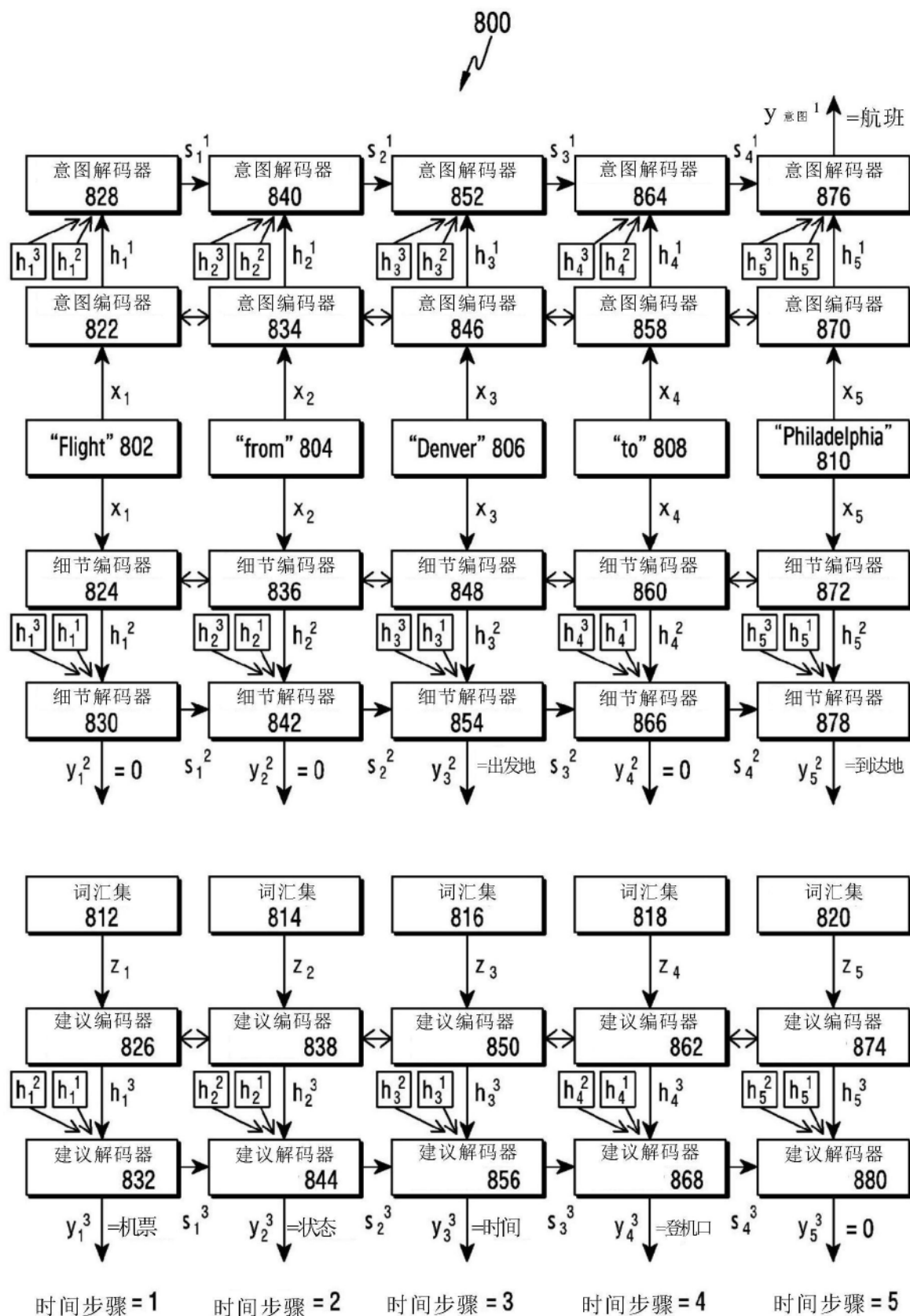


图8

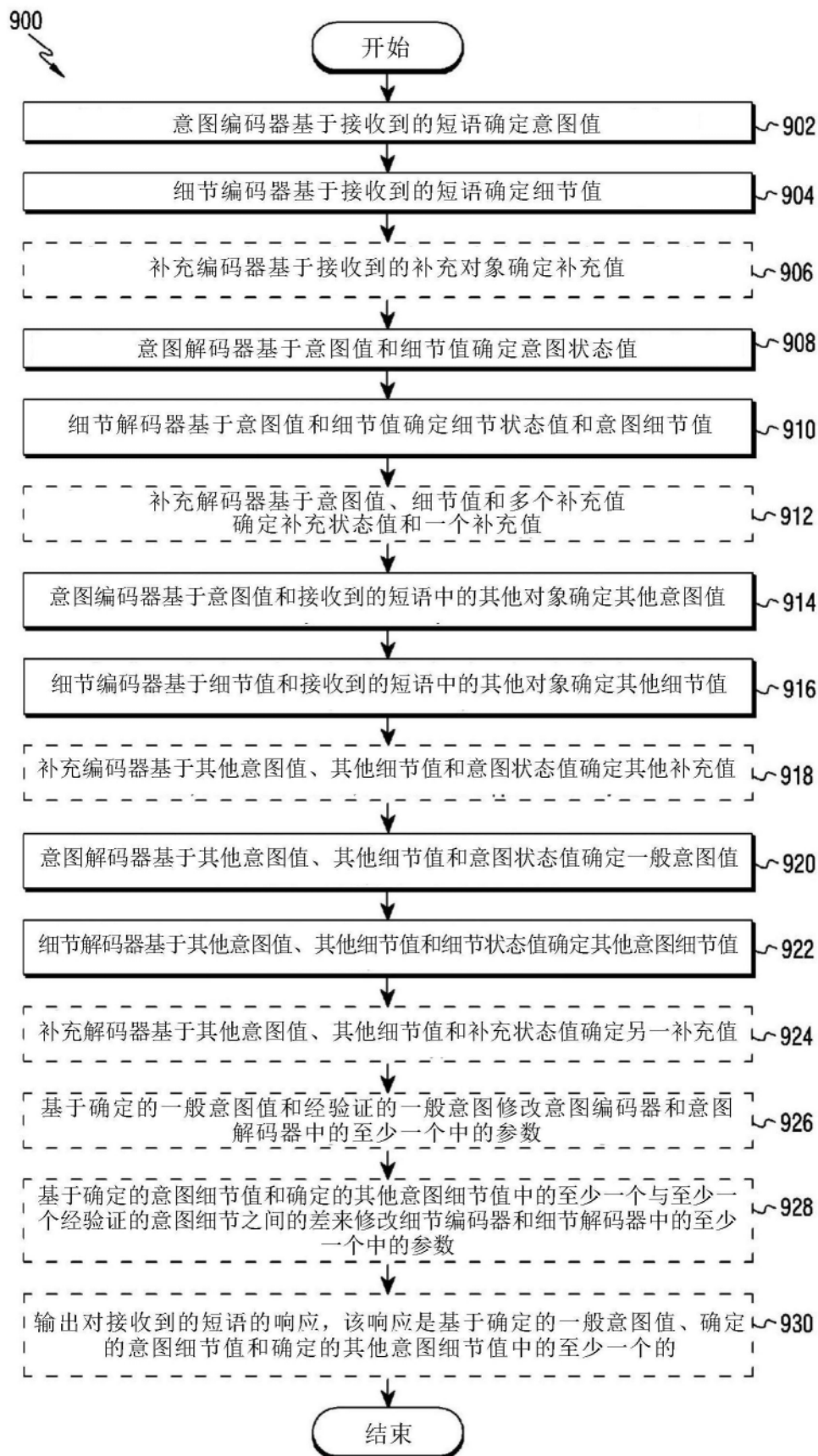


图9

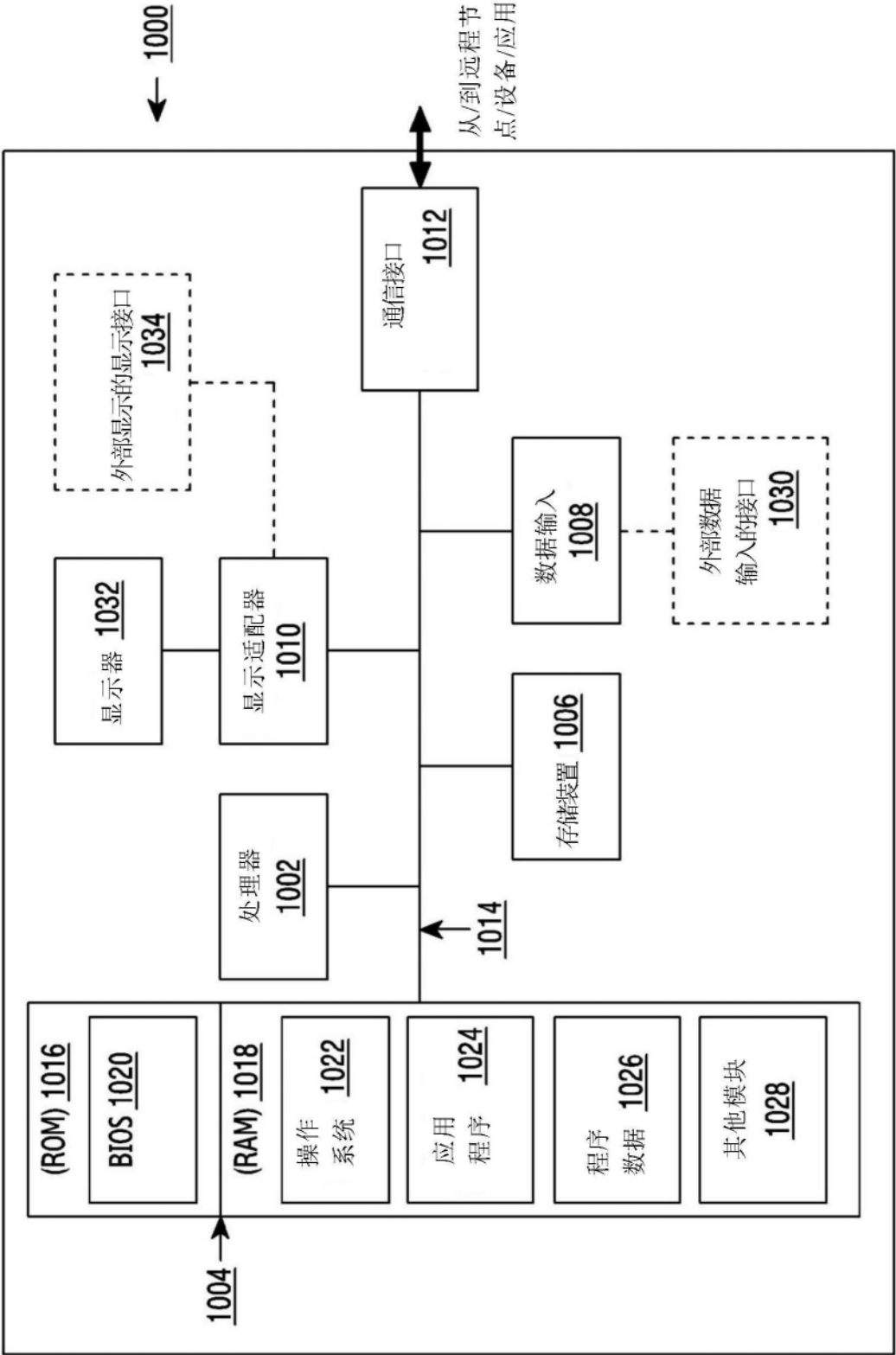


图10