



(10) 授权公告号 CN 113597602 B

(45) 授权公告日 2025. 05. 23

(21) 申请号 202080021344.4

(22) 申请日 2020.04.02

(65) 同一申请的已公布的文献号
申请公布号 CN 113597602 A

(43) 申请公布日 2021.11.02

(30) 优先权数据
16/390,845 2019.04.22 US

(85) PCT国际申请进入国家阶段日
2021.09.14

(86) PCT国际申请的申请数据
PCT/IB2020/053124 2020.04.02

(87) PCT国际申请的公布数据
W02020/217121 EN 2020.10.29

(73) 专利权人 国际商业机器公司

地址 美国纽约

(72) 发明人 元中方 殷昆燕 杨远琳 刘桐
李贺

(74) 专利代理机构 北京市中咨律师事务所
11247

专利代理人 于静 刘薇

(51) Int. Cl.
G06F 16/00 (2006.01)

(56) 对比文件
CN 108363690 A, 2018.08.03

审查员 卢洁

权利要求书2页 说明书16页 附图6页

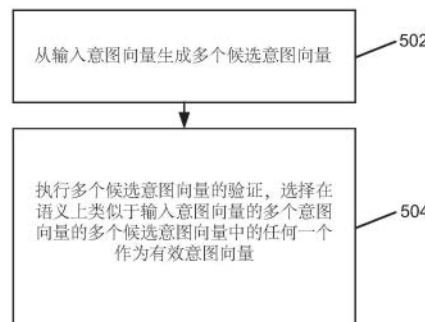
(54) 发明名称

从随机化意图向量邻近度创建意图识别模型

(57) 摘要

从输入意图向量生成多个候选意图向量。执行多个候选意图向量的验证,该验证将语义上与输入意图向量相似的多个候选意图向量中的任意一个选择为有效意图向量。

500



1. 一种计算机实现的方法,包括:
由处理器:
从输入意图向量生成多个候选意图向量;
执行多个候选意图向量的验证,该验证将语义上与输入意图向量相似的多个候选意图向量中的任意一个选择为有效意图向量,以及
与噪声向量并行迭代地处理输入意图向量,使得多个候选意图向量中的每个候选意图向量相对于输入意图向量随机分布。
2. 根据权利要求1所述的计算机实现的方法,其中处理器从输入意图向量生成多个候选意图向量包括:处理器为多个候选意图向量中的每个候选意图向量:
在长短期记忆LSTM模型中并行处理输入意图向量和噪声向量。
3. 根据权利要求1所述的计算机实现的方法,其中处理器执行多个候选意图向量的验证,该验证将语义上与输入意图向量相似的多个候选意图向量中的任意一个选择为有效意图向量包括:处理器为多个候选意图向量中的每个候选意图向量:
确定输入意图向量与相应候选意图向量之间的多维意图向量距离;和
选择在输入意图向量配置的多维意图向量距离内的候选意图向量作为在语义上与输入意图向量相似的有效意图向量。
4. 根据权利要求1所述的计算机实现的方法,还包括处理器:
响应于在验证期间确定生成的候选意图向量在语义上与输入意图向量不相似,执行从验证到生成的候选意图向量生成反馈,调整用于生成候选意图向量的向量生成参数;和
通过使用候选意图向量生成反馈迭代地调整向量生成参数,直到生成的候选意图向量在语义上与输入意图向量相似。
5. 根据权利要求1所述的计算机实现的方法,还包括处理器从语义上与输入意图向量相似的有效意图向量创建意图识别模型。
6. 根据权利要求1所述的计算机实现的方法,还包括处理器生成和处理器执行验证的至少之一作为云环境的服务提供。
7. 一种计算机系统,包括:
存储器;和
至少一个处理器集,被编程为:
在存储器内从输入意图向量生成多个候选意图向量;
执行多个候选意图向量的验证,该验证将语义上与输入意图向量相似的多个候选意图向量中的任意一个选择为有效意图向量;以及
与噪声向量并行迭代地处理输入意图向量,使得多个候选意图向量中的每个候选意图向量相对于输入意图向量随机分布。
8. 根据权利要求7所述的系统,其中编程为在存储器内从输入意图向量生成多个候选意图向量,该至少一个处理器集被编程为为多个候选意图向量中的每个候选意图向量:
在长短期记忆LSTM模型中并行处理输入意图向量和噪声向量。
9. 根据权利要求7所述的系统,其中编程为执行多个候选意图向量的验证,该验证将语义上与输入意图向量相似的多个候选意图向量中的任意一个选择为有效意图向量,该至少一个处理器集被编程为为多个候选意图向量中的每个候选意图向量:

确定输入意图向量与相应候选意图向量之间的多维意图向量距离;和
选择在输入意图向量配置的多维意图向量距离内的候选意图向量作为在语义上与输入意图向量相似的有效意图向量。

10. 根据权利要求7所述的系统,该至少一个处理器集还被编程为:

响应于在验证期间确定生成的候选意图向量在语义上与输入意图向量不相似,执行从验证到生成的候选意图向量生成反馈,调整用于生成候选意图向量的向量生成参数;和

通过使用候选意图向量生成反馈迭代地调整向量生成参数,直到生成的候选意图向量在语义上与输入意图向量相似。

11. 根据权利要求7所述的系统,该至少一个处理器集还被编程为从语义上与输入意图向量相似的有效意图向量创建意图识别模型。

12. 一种计算机程序产品,包括:

一种包含计算机可读程序代码的计算机可读存储介质,其中计算机可读存储介质本身不是瞬态信号,并且计算机可读程序代码在计算机上执行时使计算机:

从输入意图向量生成多个候选意图向量;

执行多个候选意图向量的验证,该验证将语义上与输入意图向量相似的多个候选意图向量中的任意一个选择为有效意图向量;以及

与噪声向量并行迭代地处理输入意图向量,使得多个候选意图向量中的每个候选意图向量相对于输入意图向量随机分布。

13. 根据权利要求12所述的计算机程序产品,其中在使计算机从输入意图向量生成多个候选意图向量的时,计算机可读程序代码在计算机上执行时使计算机为多个候选意图向量中的每个候选意图向量:

在长短期记忆LSTM模型中并行处理输入意图向量和噪声向量。

14. 根据权利要求12所述的计算机程序产品,其中在使计算机执行多个候选意图向量的验证,该验证将语义上与输入意图向量相似的多个候选意图向量中的任意一个选择为有效意图向量时,计算机可读程序代码在计算机上执行时使计算机为多个候选意图向量中的每个候选意图向量:

确定输入意图向量与相应候选意图向量之间的多维意图向量距离;和

选择在输入意图向量配置的多维意图向量距离内的候选意图向量作为在语义上与输入意图向量相似的有效意图向量。

15. 根据权利要求12所述的计算机程序产品,其中计算机可读程序代码在计算机上执行时还使计算机:

响应于在验证期间确定生成的候选意图向量在语义上与输入意图向量不相似,执行从验证到生成的候选意图向量生成反馈,调整用于生成候选意图向量的向量生成参数;和

通过使用候选意图向量生成反馈迭代地调整向量生成参数,直到生成的候选意图向量在语义上与输入意图向量相似。

16. 根据权利要求12所述的计算机程序产品,其中计算机可读程序代码在计算机上执行时还使计算机从语义上与输入意图向量相似的有效意图向量创建意图识别模型。

17. 根据权利要求12所述的计算机程序产品,计算机生成和计算机执行验证的至少之一作为云环境的服务提供。

从随机化意图向量邻近度创建意图识别模型

背景技术

[0001] 本发明的领域涉及认知意图识别应用,例如聊天机器人,其利用意图识别模型推断说出句子或命令的用户的意图。例如,用户可以询问“我在哪里可以找到咖啡?”,认知意图识别应用试图从意图识别模型推断出用户想要得到去咖啡店的路线。

[0002] 然而,由于存在于几乎每种不同语言中并且跨越使用语言的不同区域来表达想法和意图的自然语言(NL)变体的复杂性,意图识别模型的常规创建存在问题。例如,某些传统方法通过“为输入话语中的每个单词”确定“一组对应的语义相似词”来生成“语义相似的句子”。然而,对于任何给定单词,可能存在多个这样的语义上相似的单词(例如,同义词),并且可以使用相当不同的句子或短语来表达相似的含义。这些类型的语言复杂性的结果是可用于表达想法的大量单词和短语组合的排列。作为进一步的结果,需要以高质量句子/短语输入变化的形式的大量训练数据来精确地训练传统的意图识别模型。因此,传统的意图识别模型的创建是从各种不同输入源收集和过滤输入变化的手动密集过程。然后使用语言专家使用同义词作为用于意图识别模型生成和训练的种子文本来生成合理且有意义的句子/短语变体。提供的数据量和生成的短语变体确定了意图识别模型的质量。然而,很少有传统的意图识别模型是使用必要数量的高质量句子/短语输入变化来准备的。因此,传统的意图识别模型创建不充分且训练不充分,因此传统的意图识别模型无法有效地在所有自然语言变体和句子格式中正确识别用户的意图,以在任何给定语言中表达想法和意图。

[0003] 长短期记忆(LSTM)是在基于计算机的深度学习处理中使用的人工递归神经网络(RNN)架构。除了处理单个数据点之外,LSTM还具有允许处理数据序列(例如语音或视频)的反馈连接。

[0004] 使用单词嵌入来形成句子向量以表示单词。通过使用神经网络构建句子向量,该神经网络递归地组合生成模型中的词嵌入,例如递归/递归神经网络或使用一些其他非神经网络算法(例如,doc2vec)。与字嵌入相比,句子向量通常具有相似的形状。

[0005] 高斯噪声是统计噪声的一种形式,其具有等于正态分布的概率密度函数(PDF),可替代地称为高斯分布。高斯噪声在电信和计算机网络等领域使用随机变量建模,以模拟来自影响电信和计算机网络系统的自然源的噪声的影响。影响那些电信和计算机网络系统的这些自然源包括导体中原子的热振动和来自天体的辐射(例如,来自地球,来自太阳等)。

发明内容

[0006] 一种计算机实现的方法包括从输入意图向量生成一组候选意图向量。计算机实现的方法包括执行候选意图向量集合的验证,该验证选择语义上与输入意图向量相似的候选意图向量集合中的任何一个作为有效意图向量。该实施例具有通过依赖较少的句子/短语输入以及通过自主地推导具有增加的意图推断准确度的鲁棒意图识别模型来改进意图识别模型创建和验证的优点。

[0007] 还描述了执行计算机实现的方法的系统 and 使计算机执行计算机实现的方法的计算机程序产品。

[0008] 许可实施例涉及通过与噪声向量并行迭代地处理输入意图向量来生成候选意图向量集合,使得候选意图向量集合中的每个候选意图向量相对于输入意图向量随机分布,与使用传统技术相比,它的优势在于可以更快地创建意图识别模型。

[0009] 许可实施例涉及通过在长短期记忆(LSTM)模型内与噪声向量并行处理输入意图向量来生成候选意图向量集合,具有比使用传统技术更准确的意图识别模型创建的优势。

[0010] 许可实施例涉及通过确定输入意图向量和相应候选意图向量之间的多维意图向量距离并选择在输入意图向量配置的多维意图向量距离内的候选意图向量作为在语义上与输入意图向量相似的有效意图向量,优点是识别在多维向量空间中相对于样本意图向量最接近的候选向量,从而识别出相似的意图向量编码。

[0011] 许可实施例涉及在云环境中作为服务生成和执行验证中的至少一个,具有本文描述的意图识别模型创建技术的快速部署和可服务性的优点。

[0012] 许可实施例涉及从语义上与输入意图向量相似的有效意图向量创建意图识别模型,优点是可从经过验证的意图向量快速创建意图识别模型。

[0013] 许可实施例涉及,响应于在验证期间确定生成的候选意图向量在语义上与输入意图向量不相似,执行从验证到生成的候选意图向量生成反馈,调整用于生成候选意图向量的向量生成参数,并通过使用候选意图向量生成反馈迭代地调整向量生成参数,直到生成的候选意图向量在语义上与输入意图向量相似,优点是增加了计算机生成的高质量候选意图向量的百分比,提高了意图识别模型创建的计算速度,提高了生成的意图识别模型的准确性。

附图说明

[0014] 图1描绘了根据本发明实施例的云计算环境;

[0015] 图2描绘了根据本发明实施例的抽象模型层;

[0016] 图3是根据本主题实施例的用于从随机化意图向量邻近度创建意图识别模型的系统的实现的示例的框图;

[0017] 图4是根据本主题实施例的能够从随机化意图向量邻近度执行意图识别模型创建的核心处理模块的实现的示例的框图;

[0018] 图5是根据本主题实施例的用于从随机化意图向量邻近度创建自动意图识别模型的过程的实现的示例的流程图;

[0019] 图6是根据本发明实施例的用于从随机化意图向量邻近度创建自动意图识别模型的过程的实现的示例的流程图,其示出了附加细节和某些附加/替代操作。

具体实施方式

[0020] 下面列出的实施例代表了使本领域技术人员能够实施本发明并说明实施本发明的最佳模式所必需的信息。在根据附图阅读以下描述后,本领域技术人员将理解本发明的概念并且将认识到本文未特别提及的这些概念的应用。应当理解,这些概念和应用落入本公开和所附权利要求的范围。

[0021] 本文描述的主题提供了来自随机化意图向量邻近度的意图识别模型创建。本技术通过提供包括新形式的自主计算机控制的创建和验证句子/短语向量的技术来解决公认的

意图识别模型创建问题,所述句子/短语向量提高了意图识别模型创建的速度和与传统的意图识别模型有关的意图识别模型功能的准确性。这里描述的技术基于有限的句子/短语输入变化的计算机控制的处理,并且因此相对于传统的意图识别模型,需要更少的输入数据来创建更精确的意图识别模型。

[0022] 本发明的一些实施例可以包括以下特征、特征操作和/或优点中的一个或多个:
(i) 将噪声与句子(样本)向量组合以生成许多候选向量;(ii) 使用具有反馈的长短期记忆(LSTM)模型以促进计算机生成在初始样本向量周围抖动的大量候选向量;(iii) 使用在得到的向量空间中接近的后高斯向量来识别意图相似于编码到初始样本向量中的短语口头短语(一旦也被转换为向量),而不使用传统的基于同义词的向量编码;(iv) 考虑语言使用的变化,包括方言和区域语言的变化;和/或(v) 至少在这方面改进了意图识别,因为传统的基于同义词的向量编码可能不能从使用这些类型的口语/编码词的变化所说的短语中识别意图。

[0023] 本文描述的技术基于以下观察来操作:具有相同或相似含义的编码的意图向量之间的向量距离在多维意图向量空间中非常接近。作为该观察的结果,确定了意图向量之间的距离可以用于表示不同向量编码的句子或短语的含义的相似性,并且通过随机改变意图向量本身,创建相对于传统的输入文本变体更准确的意图识别模型。这里描述的技术完全改变了意图识别模型的创建,并且所得/创建的意图识别模型显著地改进了用户意图的基于计算机的识别。

[0024] 基于使用意图向量之间的距离,本文描述的技术计算机生成大量经验证的编码意图向量以训练和形成意图识别模型。因此,本文描述的技术与使用同义词和单词选择变化的传统手动创建语义相似的句子,随后编码那些相似的句子相比使用较少的输入数据来提供大规模、更快速和更准确的意图识别模型的创建。

[0025] 本文描述的技术处理将句子或短语形式的样本意图文本的一个版本作为输入,其表示要被编码到意图识别模型中的意图。然后,处理使用便于处理平台的数字编码方法将样本意图文本转换并编码成样本意图向量。样本意图文本到样本意图向量的转换可以通过适合于给定实现的任何处理来进行,例如频域编码(例如,快速傅立叶变换等)或其他形式的编码,每个单词或话语以数字形式表示,可用于唯一地识别相应的单词或话语。基于计算机的意图向量生成器从样本意图向量自动生成大量候选意图向量,用于评估相对于样本意图向量的多维向量距离。为了提高准确性,基于计算机的向量验证器将在多维向量距离中相互接近(例如,矩阵维度余弦相似性,其中针对余弦相似性评估的维度的数量等于向量矩阵维度)的有效且有意义的候选意图向量过滤成最终的一组相似意图向量,同时丢弃在与样本意图向量的多维向量距离上不相似(例如,在矩阵维度余弦相似性中不接近)的任何候选意图向量。最后一组相似的意图向量用于训练意图识别模型。这样,本文描述的技术提高了基于计算机的意图识别模型的创建速度和准确性,并且当部署意图识别模型时另外提高了基于计算机的意图识别的速度和准确性。作为进一步的结果,相对于传统形式的意图识别模型创建,本文描述的技术大规模地操作,具有显著降低的数据输入要求。具体地,与传统的基于同义词的方法相比,本文描述的技术利用单个样本输入文本来创建意图识别模型,该模型能够确定表达与样本输入文本的意图相似的用户输入短语的各种变体的意图。

[0026] 为了对样本意图向量进行编码,根据可配置/指定的维度来对样本意图文本短语

进行数字化。出于示例的目的,可以指定三百为维度(维度=300),但是可以适当地使用诸如二百五十六(维度=256)的其他维度用于给定的实现。出于示例的目的,继续以三百为维度(维度=300),可以将样本输入文本的每个单词转换为包括表示特定单词的300个数字值的300维向量。利用四个单词的样本输入文本短语,样本输入文本短语的数字表示产生编码该短语的四个这样的300维向量的矩阵(例如,4×300数字值的矩阵)。

[0027] 然后将得到的数字值矩阵输入“卷积神经网络”模型,并选择最大池化层处理之后的中间结果作为样本意图向量。由此获得的样本意图向量具有在多维意图向量空间中数字地表示的样本意图文本短语的语义。

[0028] 为了进一步详述卷积神经网络算法中的最大池化处理,将整个样本输入文本/短语的表示抽象为单个/结果样本意图向量。该单个向量可以用作语义向量空间中的样本输入文本/短语的语义表示。通常,共同确定卷积核的类型(n)和卷积核的数量(m)($n * m$)。在上面的示例中,结果样本意图向量的维度也是三百维(300),其中类型(n)已设置为三($n = 3$)并且卷积核的数量(m)已设置到一百($m = 100$)。然而,应该注意,卷积核的数量可以替代地设置为四百(400)、五百(500)或适合于给定实现的其他任何值。

[0029] 基于计算机的意图向量生成器通过改变样本意图向量的各个元素,从样本意图向量生成可配置数量的候选意图向量。然后针对相对于样本意图向量的多维向量距离评估/验证候选意图向量。出于示例的目的,要生成和验证的候选意图向量的数量可以被设置为一千零二十四(1,024),尽管要生成和验证的配置的候选意图向量的数量可以被设置适用于给定实现(例如,2,048、4,096等)的任何数量。

[0030] 为了进一步改善所生成的候选意图向量的概率准确性,基于计算机的意图向量生成器利用随机生成的噪声向量,该噪声向量的每个元素对于样本意图向量的每个元素(例如,单词/话语)具有跨频率的高斯(例如,正态)分布。向样本向量添加高斯噪声将每个生成的向量随机分布在样本向量周围的语义空间中。这样,虽然生成的向量不直接对应于真实语言中的词汇,但是生成的向量在语义空间中具有有形的含义,如本文所述可用于制定改进的和更准确的意图识别模型。在该实施例中,基于计算机的意图向量生成器在“长期短期记忆”(LSTM)模型内与随机生成的高斯噪声向量重复并并行地处理样本意图向量,并且针对每次迭代获得后高斯(候选)意图向量。这样,LSTM模型处理产生大量且可配置数量的后高斯候选意图向量。

[0031] 继续该实施例,基于计算机的意图向量验证器接收所生成的后高斯候选意图向量,并且将每个后高斯候选意图向量输入到意图识别算法中。通过首先测量样本输入向量和相应的后高斯候选意图向量之间的差异以获得向量距离作为以误差的形式测量相对于样本输入向量的损失或变化(例如,15%损失),将给定的后高斯候选意图向量视为接受作为意图识别模型的有效/最终意图向量。然后可以反转损失的测量以获得关于相应的后高斯候选意图向量的意图与样本意图向量的相似性的置信水平。出于示例的目的,零点八五(0.85)的多维向量距离导致百分之八十五(85%)的置信水平。此外,可以使用百分之八十五(85%)或更高的置信水平(例如,接近1.0的多维向量距离)来将相应的后高斯候选意图向量视为表示与样本有效的相似意图。尽管应该理解,任何其他置信水平可以适用于给定的实现。

[0032] 从意图向量验证器到意图向量生成器和LSTM模型的意图向量生成反馈循环用于

根据确定的误差/损失测量来训练意图向量生成器。在整个反馈过程中迭代地更新LSTM参数,以改进所生成的每个候选意图向量的意图向量生成以及随时间的附加候选意图向量处理。重复该迭代和反馈训练过程,直到由意图向量生成器生成的输出后高斯候选意图向量一致地通过意图向量验证器并且被认为是意图识别模型的有效/最终意图向量。

[0033] 因此,本文描述的处理以编程方式自我调整意图向量生成以确保更高百分比的有效意图向量由计算机控制和自动意图向量生成处理产生。然后使用这个更高百分比的有效意图向量来快速创建和训练同时具有高度置信度的意图识别模型,该意图识别模型在部署时准确地确定用户对表示与样本输入文本的意图相似的各种短语的用户意图。

[0034] 具体地,在部署的意图识别模型中以编程方式生成和验证的大量有效/最终意图向量被认为形成意图向量空间或随机分布,其统计地跨越围绕原始样本意图向量的多维向量空间区域。这样,当部署的意图识别模型的给定用户说出短语时,该短语被数字化,并且如果用户说出的短语的数字化表示接近或位于由该组有效/最终意图向量创建的多维向量空间区域内,则用户说出的短语被映射回样本意图向量,并进而映射回原始样本意图文本以确定用户的意图。应当注意,由于如上所述表示相似意图的意图向量的邻近度,即使在用户说出的特定短语未在意图识别模型中被编码为最终意图向量的情况下,也可以快速且实时地确定用户意图的准确解释。这样,这里描述的技术扩展了可以确定意图的短语集合,而不需要像传统技术识别单个短语所要求的短语的每个可能变化的单独编码。以这种方式,本文描述的技术提高了具有有限的句子/短语输入变化的部署的意图识别模型的准确性,并且因此被认为是相对于传统意图识别模型技术的显着改进。

[0035] 现在将解释用于描述本发明的实施例的一些术语。“自然语言”在此被定义为一个或多个地区的人所说的可识别的口头语言或方言,并且可以包括诸如中文、英语、法语、德语和其他语言正式语言以及这些语言的方言变体。“卷积神经网络”在此被定义为能够被执行以将口述文本转换成一个或多个数据元素序列的计算模型。数据元素数字地表示口述文本的短语或句子作为多维矩阵,并且在数字计算设备内以矩阵形式处理。“意图向量元素”在此被定义为已经被数字编码成多维向量的一个口语单词或话语,该多维向量表示和区分用于作为数字数据元素处理的口语单词或话语。“意图向量”在本文中可选地称为“句子向量”,并且因此被定义为将语音的句子或短语编码成多维矩阵并且作为如本文所述的数据的矩阵/单元处理的意图向量元素序列。“样本意图向量”在本文中可替代地称为“样本句子向量”,并且因此被定义为通过将生成模型内的字嵌入递归地组合成用作如本文所述的意图/句子向量生成和验证过程的初始/开始输入的向量格式而从样本文本生成的意图向量。“高斯噪声”在此被定义为相对于平均值或平均值集合在频率上具有标准/正态分布的噪声。“高斯噪声意图向量”在此被定义为数据元素矩阵,每个数据元素表示相对于在特定意图向量内表示的一组元素的高斯噪声的一个分布。“后高斯意图向量”在本文中可选地称为“候选意图向量”,并且因此被定义为从意图向量和高斯噪声意图向量的并行处理输出的编码数据元素的结果矩阵,其中每个所得到的编码数据元素根据由高斯噪声意图向量的数据元素序列内表示的高斯噪声提供的正态分布对意图向量的相应/成对的意图向量元素的统计变化进行编码。“长短期记忆”(LSTM)在此被定义为能够将意图向量和高斯噪声意图向量并行处理为单个输入单元并且提供后高斯意图向量的输出的计算机处理网络。“多维向量距离”在此被定义为样本意图向量和根据各个意图向量的意图向量矩阵的维度数量计算的

候选意图向量之间的向量距离。“矩阵-维度余弦相似性”在此被定义为多维向量距离的形式,其识别在各个意图向量的意图向量矩阵的维度数量上计算的余弦相似度。“经验证的意图向量”在本文中被定义为已经被验证为在给定样本句子向量的配置的多维向量距离内的候选意图向量,从该给定样本句子向量生成相应的后高斯/候选意图向量。“验证的意图识别模型”在本文中定义为如本文定义和描述的使用自动生成和验证的意图向量定义的意图识别模型。

[0036] 本文描述的技术通过从输入意图向量生成一组候选意图向量,并且执行候选意图向量集合的验证来操作,该候选意图向量集合选择在语义上与输入意图向量相似的任何候选意图向量的集合作为有效意图向量。该实施例具有通过依赖较少的句子/短语输入来改进意图识别模型创建和验证的优点,并且通过自主地导出具有增加的意图推断准确性的鲁棒意图识别模型。

[0037] 现在将描述本主题的某些实施例的某些优点。通过与噪声向量并行地迭代地处理输入意图向量来执行候选意图向量集合的生成的实施例,使得候选意图向量集合的每个候选意图向量相对于输入意图向量随机分布,这与使用传统技术相比,优点是能够更快的创建意图识别模型。通过与长短期记忆 (LSTM) 模型内的噪声向量并行处理输入意图向量来执行候选意图向量集合的生成的实施例,与使用传统技术相比,优点是能够更准确的创建意图识别模型。执行候选意图向量集合的验证的实施例涉及确定输入意图向量与相应候选意图向量之间的多维意图向量距离并选择在输入意图向量的配置的多维意图向量内的候选意图作为在语义上与输入意图向量相似的有效意图向量,具有识别相对于样本意图向量在多维向量空间邻近度的候选向量进而相似意图向量编码的优点。涉及在云环境中生成和执行验证作为服务中的至少一个的实施例具有快速部署和意图识别模型创建的可服务性的优点。涉及从语义上与输入意图向量相似的有效意图向量创建意图识别模型的实施例具有从经验证的意图向量快速创建意图识别模型的优点。涉及执行从验证到生成的候选意图向量生成反馈而调整用于生成候选意图向量的向量生成参数的实施例,响应于在验证期间确定生成的候选意图向量在语义上不与输入意图向量相似,并且通过使用候选意图向量生成反馈迭代地调整向量生成参数,直到得到的生成的候选意图向量在语义上与输入意图向量相似,具有增加计算机生成的高质量候选意图向量的百分比的优点,这提高了意图识别模型创建的计算速度并提高了所得意图识别模型的准确性。

[0038] 应当注意,本主题的概念源于对与意图识别模型创建相关联的某些限制的识别。例如,有人观察到,创建一个强大的意图识别模型需要以经过验证的高质量句子/短语输入的形式提供大量数据,但是负责创建意图识别模型的人员通常无法获得经过验证的高质量句子/短语输入。进一步观察到,由于有限的可用输入数据,意图识别模型相对于从用户的陈述/问题相对于用户的实际意图推断的意图具有高错误率。根据这些观察结果确定,需要新的计算技术来改进意图识别模型的创建和验证,其依赖于较少的句子/短语输入,并且相对于传统形式的意图识别,自主地推导出具有增加的意图推理精度的鲁棒意图识别模型创建。本文描述的本主题通过从随机化意图向量邻近度提供意图识别模型创建来改进意图识别模型创建和验证,如上文和下文更详细地描述。这样,通过使用本技术获得了改进的意图识别模型创建和验证。

[0039] 可以实时执行根据本文描述的随机化意图向量邻近度的意图识别模型创建,以允

许快速创建和验证意图识别模型。出于本说明书的目的,实时应包括足够短持续时间的任何时间帧,以便为所述主题的用户可接受的信息处理提供合理的响应时间。另外,术语“实时”应包括通常称为“近实时”的内容-通常意味着足够短持续时间的任何时间帧,以便为所描述主题的用户可接受的按需信息处理提供合理的响应时间(例如,在一秒的一部分内或在几秒内)。这些术语虽然难以精确定义,但是本领域技术人员可以很好地理解。

[0040] 下面的表1示出了基于输入文本/短语“我无法呼吸,这是慢性鼻炎的症状吗?”的编码的300维样本意图向量的一部分的一个示例。下表1通过使用省略号点来减少示例的长度省略许多中间值。然而,应该理解,形成表1的示例的编码的300维样本意图向量包括三百(300)个元素,因为如上所述并且出于示例的目的,已经设置了类型(n)在用于编码样本意图向量的卷积神经网络算法中,将三个(n=3)和卷积核的数量(m)设置为一百(m=100)。

[0041]

0.06217742	-0.09291203	-0.02840577	0.00374683	-0.10506968	-0.13434719
-0.15676261	0.04667582	0.19579822	0.04030076	-0.21957162	0.04010129
0.07424975	0.01369099	0.07016561	0.14306846	-0.08841907	-0.28088206
-0.23514667	-0.01831863	-0.00475325	-0.0811624	-0.37781501	-0.17635982
-0.00429389	0.05205948	-0.03414214	-0.11769973	-0.02660859	0.16670637
0.07024501	0.25753206	0.01059075	-0.12026082	-0.22058713	0.06767957
...
0.10361771	0.09943719	0.32700372	-0.0283169	0.12805387	0.01565604

[0042] 表1 300维样本意图向量的部分

[0043] 基于上面的表1的示例,还应当注意,根据本文描述的技术生成的多个候选意图向量每个也将被表示为300维候选意图向量,其中相应候选意图向量的各个元素可以相对于样本意图向量的相应单个元素随机改变。上面的表1的示例提供了关于如本文所述的样本意图向量和候选意图向量的形成的足够细节,并且为了简洁起见,省略了意图向量的数字编码图案的附加示例。如上所述,应用多维向量距离计算以确定样本意图向量与每个生成的候选意图向量之间的相对向量空间距离,以执行本文所述的验证。

[0044] 下面将进一步提供算法处理和计算效率的其他细节。本说明书的以下部分提供了可以在其中实现本技术的高级计算平台的示例,接着是根据本文描述的随机化意图向量邻近度创建意图识别模型的进一步细节。

[0045] 首先应当理解,尽管本公开包括关于云计算的详细描述,但其中记载的技术方案的实现却不限于云计算环境,而是能够结合现在已知或以后开发的任何其它类型的计算环境而实现。

[0046] 云计算是一种服务交付模式,用于对共享的可配置计算资源池进行方便、按需的网络访问。可配置计算资源是能够以最小的管理成本或与提供者进行最少的交互就能快速部署和释放的资源,例如可以是网络、网络带宽、服务器、处理、内存、存储、应用、虚拟机和服务。这种云模式可以包括至少五个特征、至少三个服务模型和至少四个部署模型。

[0047] 特征包括:

[0048] 按需自助式服务:云的消费者在无需与服务提供者进行人为交互的情况下能够单方面自动地按需部署诸如服务器时间和网络存储等的计算能力

[0049] 广泛的网络接入:计算能力可以通过标准机制在网络上获取,这种标准机制促进

了通过不同种类的瘦客户机平台或厚客户机平台(例如移动电话、膝上型电脑、个人数字助理PDA)对云的使用。

[0050] 提供者的计算资源被归入资源池并通过多租户(multi-tenant)模式服务于多重消费者,其中按需将不同的实体资源和虚拟资源动态地分配和再分配。一般情况下,消费者不能控制或甚至并不知晓所提供的资源的确切位置,但可以在较高抽象程度上指定位置(例如国家、州或数据中心),因此具有位置无关性。

[0051] 迅速弹性:能够迅速、有弹性地(有时是自动地)部署计算能力,以实现快速扩展,并且能迅速释放来快速缩小。在消费者看来,用于部署的可用计算能力往往显得是无限的,并能在任意时候都能获取任意数量的计算能力。

[0052] 可测量的服务:云系统通过利用适于服务类型(例如存储、处理、带宽和活跃用户帐号)的某种抽象程度的计量能力,自动地控制和优化资源效用。可以监测、控制和报告资源使用情况,为服务提供者和消费者双方提供透明度。

[0053] 服务模型如下:

[0054] 软件即服务(SaaS):向消费者提供的能力是使用提供者在云基础架构上运行的应用。可以通过诸如网络浏览器的瘦客户机接口(例如基于网络的电子邮件)从各种客户机设备访问应用。除了有限的特定于用户的应用配置设置外,消费者既不管理也不控制包括网络、服务器、操作系统、存储、乃至单个应用能力等的底层云基础架构。

[0055] 平台即服务(PaaS):向消费者提供的能力是在云基础架构上部署消费者创建或获得的应用,这些应用利用提供者支持的程序设计语言和工具创建。消费者既不管理也不控制包括网络、服务器、操作系统或存储的底层云基础架构,但对其部署的应用具有控制权,对应用托管环境配置可能也具有控制权。

[0056] 基础架构即服务(IaaS):向消费者提供的能力是消费者能够在其中部署并运行包括操作系统和应用的任意软件的处理、存储、网络和其他基础计算资源。消费者既不管理也不控制底层的云基础架构,但是对操作系统、存储和其部署的应用具有控制权,对选择的网络组件(例如主机防火墙)可能具有有限的控制权。

[0057] 部署模型如下:

[0058] 私有云:云基础架构单独为某个组织运行。云基础架构可以由该组织或第三方管理并且可以存在于该组织内部或外部。

[0059] 共同体云:云基础架构被若干组织共享并支持有共同利害关系(例如任务使命、安全要求、政策和合规考虑)的特定共同体。共同体云可以由共同体内的多个组织或第三方管理并且可以存在于该共同体内部或外部。

[0060] 公共云:云基础架构向公众或大型产业群提供并由出售云服务的组织拥有。

[0061] 混合云:云基础架构由两个或更多部署模型的云(私有云、共同体云或公共云)组成,这些云依然是独特的实体,但是通过使数据和应用能够移植的标准化技术或私有技术(例如用于云之间的负载平衡的云突发流量分担技术)绑定在一起。

[0062] 云计算环境是面向服务的,特点集中在无状态性、低耦合性、模块性和语意的互操作性。云计算的核心是包含互连节点网络的基础架构。

[0063] 现在参考图1,其中显示了根据本发明一个实施例的示例性的云计算环境50。如图所示,云计算环境50包括云计算消费者使用的本地计算设备可以与其相通信的一个或者多

个云计算节点10,本地计算设备例如可以是个人数字助理(PDA)或移动电话54A,台式电脑54B、笔记本电脑54C和/或汽车计算机系统54N。云计算节点10之间可以相互通信。可以在包括但不限于如上所述的私有云、共同体云、公共云或混合云或者它们的组合的一个或者多个网络中将云计算节点10进行物理或虚拟分组(图中未显示)。这允许云的消费者无需在本地计算设备上维护资源,云计算环境50就能提供基础架构即服务、平台即服务和/或软件即服务。应当理解,图1显示的各类计算设备54A-N仅仅是示意性的,云计算节点10以及云计算环境50可以与任意类型网络上和/或网络可寻址连接的任意类型的计算设备(例如使用网络浏览器)通信。

[0064] 现在参考图2,其中显示了根据本发明一个实施例的云计算环境50(图1)提供的一组功能抽象层。首先应当理解,图2所示的组件、层以及功能者仅是示意性的,本发明的实施例不限于此。如图2所示,提供下列层和对应功能:

[0065] 硬件和软件层60包括硬件和软件组件。硬件组件的例子包括:主机61;基于RISC(精简指令集计算机)体系结构的服务器62;服务器63;刀片服务器64;存储设备65;网络和网络组件66。软件组件的例子包括:网络应用服务器软件67以及数据库软件68。

[0066] 虚拟层70提供一个抽象层,该层可以提供下列虚拟实体的例子:虚拟服务器71、虚拟存储72、虚拟网络73(包括虚拟私有网络)、虚拟应用和操作系统74,以及虚拟客户端75。

[0067] 在一个示例中,管理层80可以提供下述功能:资源供应功能81:提供用于在云计算环境中执行任务的计算资源和其它资源的动态获取;计量和定价功能82:在云计算环境内对资源的使用进行成本跟踪,并为此提供帐单和发票。在一个例子中,该资源可以包括应用软件许可。安全功能:为云的消费者和任务提供身份认证,为数据和其它资源提供保护。用户门户功能83:为消费者和系统管理员提供对云计算环境的访问。服务水平管理功能84:提供云计算资源的分配和管理,以满足必需的服务水平。服务水平协议(SLA)计划和履行功能85:为根据SLA预测的对云计算资源未来需求提供预先安排和供应。

[0068] 工作负载层90提供可以利用云计算环境的功能的示例。可以从该层提供的工作负载和功能的示例包括:映射和导航91;软件开发和生命周期管理92;虚拟课堂教育交付93;数据分析处理94;交易处理95;和从随机化意图向量邻近度创建意图识别模型。

[0069] 在以上示例中,云计算环境示出了可以在其上部署和管理虚拟代理的几种类型的计算设备54A-N。关于替代实施方案,下面的图3和4是针对此类替代方案的。应当理解,对于给定的实施方式,各种替代方案可以与上述实施方式选项组合或替代。

[0070] 图3是用于从随机化意图向量邻近度创建意图识别模型的系统100的实现的示例的框图。计算设备_1 102到计算设备_N 104经由网络106与若干其他设备通信。计算设备_1 102到计算设备_N 104表示利用用于用户界面和设备控制处理的一个或多个意图识别模型的用户设备。然而,应该注意,计算设备_1 102到计算设备_N 104可以附加地/替代地从本文描述的随机化意图向量邻近度实现自动意图识别模型创建。作为另外的替代方案,意图识别模型(IRM)服务器108可以通过根据本文所述的随机化向量邻近度的自动意图识别模型创建生成和验证一个或多个意图识别模型,为计算设备_1 102到计算设备_N 104提供服务。适合于给定的实现,IRM服务器108或计算设备_1 102到计算设备_N 104中的一个或多个可以在本地或在意图识别模型(IRM)数据库110内存储一个或多个经验证的意图识别模型,以在系统100内访问和使用。

[0071] 鉴于上述实现方案,本技术可以在云计算平台内、在用户计算设备上、在服务器设备级实现,或者通过适合于给定的这些平台和设备的组合来实现。存在用于实现本主题的各种可能性,并且所有这些可能性都被认为在本主题的范围之内。

[0072] 网络106包括适合于预期目的的任何形式的互连,包括分别为能够互连各个设备的诸如内联网或因特网的私有或公共网络、直接模块间互连、拨号、无线或任何其他互连机制。

[0073] IRM服务器108包括能够经由诸如网络106之类的网络提供数据以供诸如计算设备_1 102之类的设备(例如计算设备_1 110)消费的任何设备。因此,IRM服务器108可以包括网络、服务器、应用程序服务器或其他数据服务器设备。

[0074] IRM数据库110包括关系数据库、对象数据库或任何其他存储类型的设备。这样,IRM数据库110可以适当地实现给定的实现。

[0075] 图4是能够从随机化意图向量邻近度执行意图识别模型创建的核心处理模块200的实现的示例的框图。适合于给定的实现,核心处理模块200可以与计算设备_1 102到计算设备_N 104,或与IRM服务器108,与云计算环境50内的设备相关联。这样,本文一般性地描述了核心处理模块200,但是应当理解,核心处理模块200内的组件的实现的许多变化是可能的,并且所有这些变化都在本主题的范围之内。此外,适合于给定的实现,核心处理模块200可以实现为嵌入式处理设备,其具有专门设计用于执行本文所述的处理的电路。

[0076] 核心处理模块200可以从与每个实现相关联的随机化意图向量邻近度提供意图识别模型创建的不同和补充处理。这样,对于以下任何示例,应当理解,关于结合另一设备描述的任何设备(例如,发送/发送等)描述的功能的任何方面应被理解为同时描述其他相应设备的功能(例如,接收/接收等)。

[0077] 中央处理单元(CPU)202(“处理器”或“应用程序专用”处理器)提供在核心处理模块200内执行计算机指令执行、计算和其他能力的硬件。显示器204提供视觉信息到核心处理模块200的用户和输入设备206为用户提供输入能力。

[0078] 显示器204包括任何显示装置,例如阴极射线管(CRT)、液晶显示器(LCD)、发光二极管(LED)、电子墨水显示器、投影、触摸屏或其他显示元件或面板。输入设备206包括计算机键盘、小键盘、鼠标、笔、操纵杆、触摸屏、语音命令处理单元或用户可以通过其与显示器204上的信息交互并响应信息的任何其他类型的输入设备。

[0079] 应当注意,显示器204和输入设备206可以是用于某些实现/设备的核心处理模块200的可选组件,或者可以远离相应设备定位并且由位于其中的与各个设备通信的另一计算设备托管。因此,核心处理模块200可以作为完全自动化的嵌入式设备操作而无需直接的用户可配置性或反馈。然而,适合于给定的实现,核心处理模块200还可以分别经由显示器204和输入设备206提供用户反馈和可配置性。

[0080] 适合于给定实现,通信模块208提供硬件、协议栈处理和互连功能,其允许核心处理模块200与系统100内或云计算环境50内的其他模块通信。适用于给定的实现,通信模块208包括可用于提供互连能力的任何电、协议和协议转换能力。这样,通信模块208表示能够与其他设备进行通信的通信设备。适合于给定的实现,通信模块208还可以包括一个或多个无线通信能力。

[0081] 存储器210包括意图向量处理和存储区域212,其在核心处理模块200内存储意图

向量处理信息。如下面将更详细描述,存储在意图向量处理和存储区域内的意图向量处理信息212用于基于有限的句子/短语输入变化快速(实时)生成高质量的意图识别模型。这样,意图向量处理和存储区域212存储样本意图向量、高斯噪声向量、一组或多组后高斯/候选意图向量,以及经验证的处于给定样本意图向量的配置的多维向量距离内的验证意图向量。可以在存储器210内执行用于创建一个或多个经验证的意图识别模型的处理。经验证的意图识别模型存储区域214为从经验证的意图向量生成的一个或多个经验证的意图识别模型提供存储。可以在经验证的意图识别模型存储区域214内部署/分发或使用经验证的意图识别模型。适合于给定的实现,可以在经验证的意图识别模型存储区域214内与经验证的意图识别模型的创建和验证相关联地存储许多其他形式的中间信息。

[0082] 应当理解,存储器210包括适合于预期目的、适当地分布或定位的易失性和非易失性存储器的任何组合,并且可以包括为了便于说明目的而在本示例中未示出的其他存储器部分。例如,存储器210可以包括代码存储区域、操作系统存储区域、代码执行区域和数据区域,而不脱离本主题的范围。

[0083] 还示出了意图向量创建和邻近度验证模块216。意图向量创建和邻近度验证模块216包括意图向量生成器模块218,提供随机化后高斯/候选意图向量生成。意图向量创建和邻近度验证模块216还包括意图向量验证器模块220,为核心处理模块200提供对所生成的候选意图向量到样本意图向量的邻近度的验证,如上文和下文更详细地描述的那样。意图向量创建和邻近度验证模块216通过执行样本意图向量编码、候选意图向量生成、候选意图向量的验证、验证的意图识别的生成以及适用于给定实现的模型和其他相关处理,从核心处理模块200的随机化向量邻近度实现自动意图识别模型创建。

[0084] 还应注意,意图向量创建和邻近度验证模块216可以形成所描述的其他电路的一部分而不脱离本主题的范围。意图向量创建和邻近度验证模块216可以形成中断服务例程(ISR)的一部分、操作系统的一部分或应用程序的一部分而不脱离本主题的范围。意图向量创建和邻近度验证模块216还可以包括具有电路的嵌入式设备,该电路专门设计用于执行适合于给定实现的本文描述的处理。

[0085] 再次在图4中示出与核心处理模块200相关联的IRM数据库110。这样,适合于给定的实现,IRM数据库110可以在不使用网络连接的情况下可操作地耦合到核心处理模块200,并且可以与核心处理模块200相关联地使用,以创建或使用一个或多个经验证的意图识别模型。

[0086] CPU 202、显示器204、输入设备206、通信模块208、存储器210、意图向量创建和邻近度验证模块216以及IRM数据库110经由互连222互连。互连222包括系统总线、网络或能够为各个组件提供用于相应目的的适当互连的任何其他互连。

[0087] 尽管为了便于说明和描述的目的,图4中示出的不同模块被示为组件级模块,但应注意,这些模块包括用于执行以下操作的任何硬件、编程处理器和存储器。如上所述并且在下面更详细地描述了各个模块的功能。例如,模块可以包括专用集成电路(ASIC)、处理器、天线和/或分立集成电路和组件形式的附加控制器电路,用于执行与各个模块相关联的通信和电控制活动。此外,模块可以包括适当的中断级、堆栈级和应用级模块。此外,模块可以包括用于存储、执行和数据处理的任何存储器组件,用于执行与各个模块相关联的处理活动。模块还可以形成所描述的其他电路的一部分,或者可以在不脱离本主题的范围的情况

下进行组合。

[0088] 另外,虽然核心处理模块200被示出并且具有所描述的某些组件,但是其他模块和组件可以与核心处理模块200相关联而不脱离本主题的范围。另外,应当注意,虽然为了便于说明目的将核心处理模块200描述为单个设备,但是核心处理模块200内的组件可以通过网络共同定位或分布和互连,而不会偏离本主题的范围。用于核心处理模块200的组件的许多其他可能的布置是可能的,并且所有这些布置都被认为在本主题的范围。还应该理解,尽管出于示例的目的将IRM数据库110示出为单独的组件,但是存储在IRM数据库110内的信息也可以/可选地存储在存储器210内,而不脱离本主题的范围。因此,核心处理模块200可以采用许多形式并且可以与许多平台相关联。

[0089] 下面描述的图5至图6表示可以由诸如核心处理模块200之类的设备执行的示例过程,以从与本主题相关联的随机化意图向量邻近度执行自动意图识别模型创建。关于示例过程的许多其他变型是可能的,并且所有变体都被认为在本主题的范围。示例过程可以由与这些设备相关联的模块执行,诸如意图向量创建和邻近度验证模块216和/或由CPU 202执行。应该注意,为了便于说明的目的,在下面描述的示例过程中没有示出超时过程和其他错误控制过程。然而,应理解,所有这些程序都被认为是在本主题的范围。此外,可以组合所描述的过程,可以改变所描述的处理的序列,并且可以添加或去除附加处理而不脱离本主题的范围。

[0090] 图5是用于从随机化意图向量邻近度创建自动意图识别模型的过程500的实现的示例的流程图。过程500表示执行本文描述的基于计算机的验证的意图识别模型创建的计算机实现的方法。在框502,过程500从输入意图向量生成多个候选意图向量。在框504,过程500执行多个候选意图向量的验证,选择在语义上与输入意图向量相似的多个候选意图向量中的任何一个作为有效意图向量。

[0091] 图6是用于从随机化意图向量邻近度创建自动意图识别模型的过程600的实现的示例的流程图,其示出了附加细节和某些附加/替代操作。过程600表示执行本文描述的基于计算机的验证的意图识别模型创建的计算机实现的方法。在决策点602,过程600确定是否生成意图识别模型(IRM)。响应于确定生成意图识别模型,过程600在框604接收样本文本。样本文本是要编码的短语或句子,并且为其生成意图识别模型。在框606,过程600将样本文本编码为样本意图向量。如上所述,样本意图向量从接收的样本文本生成为多维矩阵,其中每个单词对应于维度向量(例如,维度=300)。作为维度向量的多维矩阵的样本意图向量用作意图向量生成和验证处理的输入。样本意图向量包括表示样本文本的数据元素序列。这样,样本意图向量将数字化单词/话语的序列表示为编码和可区分元素的序列。

[0092] 在框608,过程600基于样本意图向量生成高斯噪声向量。如上所述,高斯噪声向量包括数据元素序列,每个数据元素表示相对于形成样本意图向量的各个元素的高斯噪声分布,并且还是噪声元素的矩阵。这样,在样本意图向量的元素和高斯噪声向量的高斯噪声分布元素之间存在映射。

[0093] 在框610,过程600配置长期短期存储器(LSTM)参数以供LSTM模型用于生成候选意图向量。如上所述,LSTM模型通过并行处理相应的样本意图向量和高斯噪声向量来生成候选意图向量。LSTM处理的输出产生一组候选意图向量。

[0094] 过程600开始迭代处理以生成一组候选意图向量。在框612,过程600再次通过用高

斯噪声向量并行处理样本意图向量来生成候选意图向量。根据高斯噪声向量的相应元素随机处理作为矩阵的样本意图向量的每个元素,以创建唯一的候选意图向量。

[0095] 然后,过程600开始候选意图向量验证处理。在框614,过程600利用意图识别算法处理候选意图向量。在框616,过程600确定候选意图向量相对于样本意图向量的向量距离和置信水平。通过从样本意图向量测量候选意图向量的多维向量距离来确定候选意图向量的向量距离和置信度,例如通过确定两个多维向量/矩阵的矩阵-维度余弦相似性。通过使用任何合适的向量距离算法来建立和配置用于接受候选意图向量作为有效意图向量的可接受的多维向量距离或阈值。出于示例而非限制的目的,可以利用多维向量距离或其他距离测量技术,并且阈值多维向量距离为零点八五(0.85),或者表示为八十五的置信水平。可以配置百分比(85%),使得具有满足或超过百分之八十五的置信水平(85%)的多维向量距离的任何候选意图向量被认为是有效的意图向量,而任何候选意图向量具有低于百分之八十五(85%)的置信水平的多维向量距离不被认为是有效的意图向量。应当注意,许多用于测量向量距离的技术是可能的,并且所有这些技术都被认为是在本主题的范围之内。

[0096] 然后,过程600开始基于滤波器的处理,以确定候选意图向量相对于样本意图向量的匹配质量。在决策点618,过程600根据候选意图向量的损失和置信水平的测量来确定候选意图向量是否通过验证。如上所述,如果候选意图向量接近多维意图向量空间中的样本意图向量(例如,矩阵维度余弦相似度大于或等于0.85),则候选意图向量被认为是有效向量从而通过验证)。

[0097] 响应于在决策点618确定候选意图向量未通过验证(例如,对于矩阵维度余弦相似度低于0.85),过程600开始迭代反馈处理以调整用于生成候选意图向量的LSTM模型的候选意图向量生成参数。在框620,过程600丢弃未通过验证的候选意图向量。在框622,过程600更新用于从样本意图向量和高斯噪声向量生成候选意图向量的LSTM参数。该过程迭代并返回到框612以使用更新的LSTM参数生成新的候选意图向量,并如上所述进行迭代。

[0098] 返回到决策点618的描述,响应于确定候选意图向量通过验证(例如,对于矩阵维度余弦相似度大于或等于0.85),过程600通过在框624将候选意图向量添加到意图识别模型来将候选意图向量分配作为最终意图识别模型(IRM)向量。在框626,过程600存储在创建经验证的意图向量创建期间使用的LSTM参数以及用最终的IRM向量确定的置信水平。

[0099] 在决策点628,过程600确定是否迭代以生成并验证另一候选意图向量。响应于确定生成并验证另一候选发明向量,过程600返回到框610以配置新的LSTM参数以使LSTM模型基于样本意图向量和高斯噪声向量生成新的/唯一的候选意图向量。这样,过程600迭代地调整LSTM参数以在意图识别模型内快速生成多个经验证的意图向量。

[0100] 过程600如上所述迭代以生成并验证候选意图向量,直到已将足够的最终IRM向量添加到意图识别模型,以在部署期间跨越意图识别模型的各种自动生成的意图向量实现意图识别。响应于在决策点628确定不生成并验证另一候选意图向量,过程600将最终IRM向量和映射存储到接收和编码的样本文本,作为在框630设置的意图识别模型向量。在框632,适当地用于给定的实现,过程600将所生成的意图识别模型部署到IRM数据库110或者。过程600返回到决策点602并如上所述进行迭代。

[0101] 这样,过程600接收样本文本并将其编码为样本意图向量,并使用已经参数化以调整高斯噪声向量的元素对样本意图向量的元素的效果的LSTM模型与高斯噪声向量一起并

行处理样本样本意图向量。过程600迭代地生成并验证反馈回路内的候选意图向量,该反馈回路调整LSTM参数直到产生高质量(近距离/最小距离)最终IRM向量。过程600迭代地创建最终IRM向量,直到足够数量的最终IRM向量跨越原始样本文本的自然(随机)变化。过程600将经验证的一组最终IRM向量存储为意图识别模型,并部署计算机生成的意图识别模型。

[0102] 本发明的一些实施例以下列方式中的一种或多种改进计算机技术:(i)用于大规模意图识别模型生成的较低输入数据要求;(ii)使用较少的数据更快速地生成意图识别模型;(iii)从生成的意图识别模型中更准确地部署意图识别。

[0103] 本发明不是抽象的,因为它特别涉及计算机操作和/或硬件,其原因包括:(i)计算机控制的处理,其使用比传统方法更少的输入来创建更准确的意图识别模型;(ii)计算机控制的最终/验证意图向量分布,以编程方式改变样本意图向量,从而改进意图识别模型的创建;(iii)计算机控制的处理,提高计算机在从用户的口头短语中识别用户意图的准确性。

[0104] 如以上结合图1至图6所描述的,示例系统和过程提供了从随机化意图向量邻近度创建意图识别模型。与从随机化意图向量邻近度创建意图识别模型相关联的许多其他变化和附加活动是可能的,并且所有这些都认为在本主题的范围内。

[0105] 本发明可以是系统、方法和/或计算机程序产品。计算机程序产品可以包括计算机可读存储介质,其上载有用于使处理器实现本发明的各个方面的计算机可读程序指令。

[0106] 计算机可读存储介质可以是保持和存储由指令执行设备使用的指令的有形设备。计算机可读存储介质例如可以是--但不限于--电存储设备、磁存储设备、光存储设备、电磁存储设备、半导体存储设备或者上述的任意合适的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括:便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、静态随机存取存储器(SRAM)、便携式压缩盘只读存储器(CD-ROM)、数字多功能盘(DVD)、记忆棒、软盘、机械编码设备、例如其上存储有指令的打孔卡或凹槽内凸起结构、以及上述的任意合适的组合。这里所使用的计算机可读存储介质不被解释为瞬时信号本身,诸如无线电波或者其他自由传播的电磁波、通过波导或其他传输媒介传播的电磁波(例如,通过光纤电缆的光脉冲)、或者通过电线传输的电信号。

[0107] 这里所描述的计算机可读程序指令可以从计算机可读存储介质下载到各个计算/处理设备,或者通过网络、例如因特网、局域网、广域网和/或无线网下载到外部计算机或外部存储设备。网络可以包括铜传输电缆、光纤传输、无线传输、路由器、防火墙、交换机、网关计算机和/或边缘服务器。每个计算/处理设备中的网络适配卡或者网络接口从网络接收计算机可读程序指令,并转发该计算机可读程序指令,以供存储在各个计算/处理设备中的计算机可读存储介质中。

[0108] 用于执行本发明操作的计算机程序指令可以是汇编指令、指令集架构(ISA)指令、机器指令、机器相关指令、微代码、固件指令、状态设置数据、集成电路配置数据或者以一种或多种编程语言的任意组合编写的源代码或目标代码,所述编程语言包括面向对象的编程语言—诸如Smalltalk、C++等,以及过程式编程语言—诸如“C”语言或类似的编程语言。计算机可读程序指令可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算

机或服务器上执行。在涉及远程计算机的情形中,远程计算机可以通过任意种类的网络-包括局域网(LAN)或广域网(WAN)-连接到用户计算机,或者,可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。在一些实施例中,通过利用计算机可读程序指令的状态信息来个性化定制电子电路,例如可编程逻辑电路、现场可编程门阵列(FPGA)或可编程逻辑阵列(PLA),该电子电路可以执行计算机可读程序指令,从而实现本发明的各个方面。

[0109] 这里参照根据本发明实施例的方法、装置(系统)和计算机程序产品的流程图和/或框图描述了本发明的各个方面。应当理解,流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合,都可以由计算机可读程序指令实现。

[0110] 这些计算机可读程序指令可以提供给通用计算机、专用计算机或其它可编程数据处理装置的处理器,从而生产出一种机器,使得这些指令在通过计算机或其它可编程数据处理装置的处理器执行时,产生了实现流程图和/或框图中的一个或多个方框中规定的功能/动作的装置。也可以把这些计算机可读程序指令存储在计算机可读存储介质中,这些指令使得计算机、可编程数据处理装置和/或其他设备以特定方式工作,从而,存储有指令的计算机可读介质则包括一个制品,其包括实现流程图和/或框图中的一个或多个方框中规定的功能/动作的各个方面的指令。

[0111] 也可以把计算机可读程序指令加载到计算机、其它可编程数据处理装置、或其它设备上,使得在计算机、其它可编程数据处理装置或其它设备上执行一系列操作步骤,以产生计算机实现的过程,从而使得在计算机、其它可编程数据处理装置、或其它设备上执行的指令实现流程图和/或框图中的一个或多个方框中规定的功能/动作。

[0112] 附图中的流程图和框图显示了根据本发明的多个实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上,流程图或框图中的每个方框可以代表一个模块、程序段或指令的一部分,所述模块、程序段或指令的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。在有些作为替换的实现中,方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如,两个连续的方框实际上可以基本并行地执行,它们有时也可以按相反的顺序执行,这依所涉及的功能而定。也要注意的,框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合,可以用执行规定的功能或动作的专用的基于硬件的系统来实现,或者可以用专用硬件与计算机指令的组合来实现。

[0113] 此处使用的术语仅用于描述特定实施例的目的,并不旨在限制本发明。如本文所用,单数形式“一个”和“这个”、“那个”也旨在包括复数形式,除非上下文另有明确指示。将进一步理解,术语“包含”和/或“包括”,当在本说明书中使用,指定所述特征、整数、步骤、操作、元素和/或组件的存在,但不排除存在或添加一个或多个其他特征、整数、步骤、操作、元素、组件和/或它们的组。

[0114] 权利要求中的所有装置或步骤加功能元件的相应结构、材料、动作和等效物旨在包括用于与如具体要求保护的其他要求保护的元件组合执行功能的任何结构、材料或动作。本发明的描述是为了说明和描述的目的而呈现的,但并不旨在穷举或限制所公开形式的本发明。在不脱离本发明的范围的情况下,基于本文的教导,对本领域普通技术人员而言,许多修改和变化将是显而易见的。描述该主题是为了解释本发明的原理和实际应用,并使本领域的其他普通技术人员能够理解本发明的各种实施例以及适合于预期的特定用途

的各种修改。

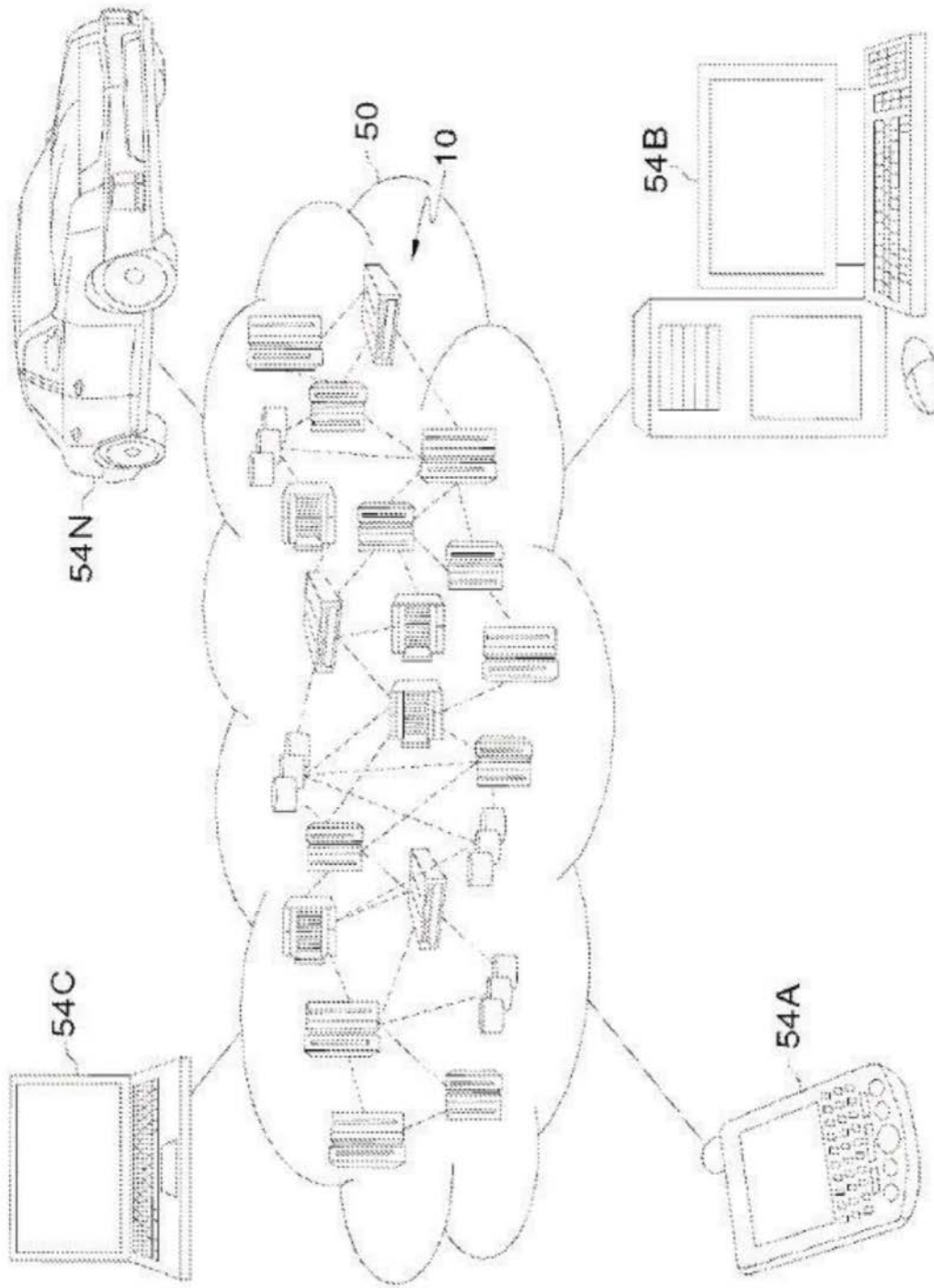


图1

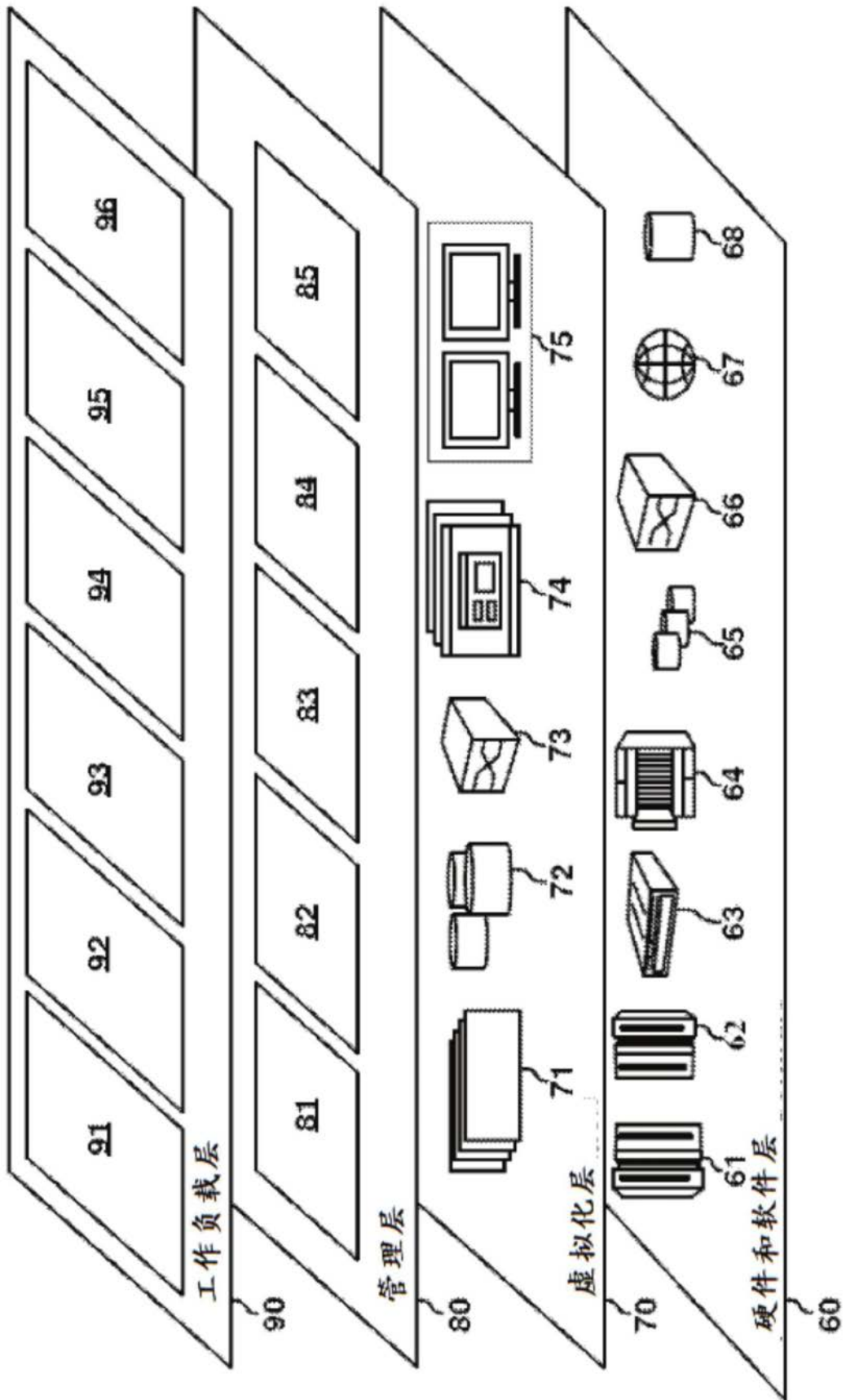
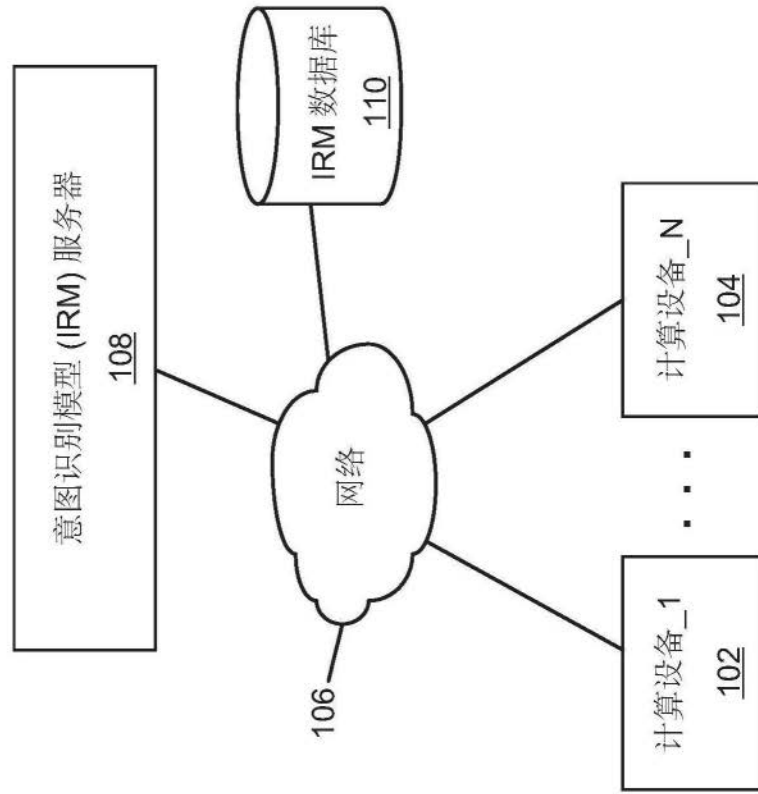
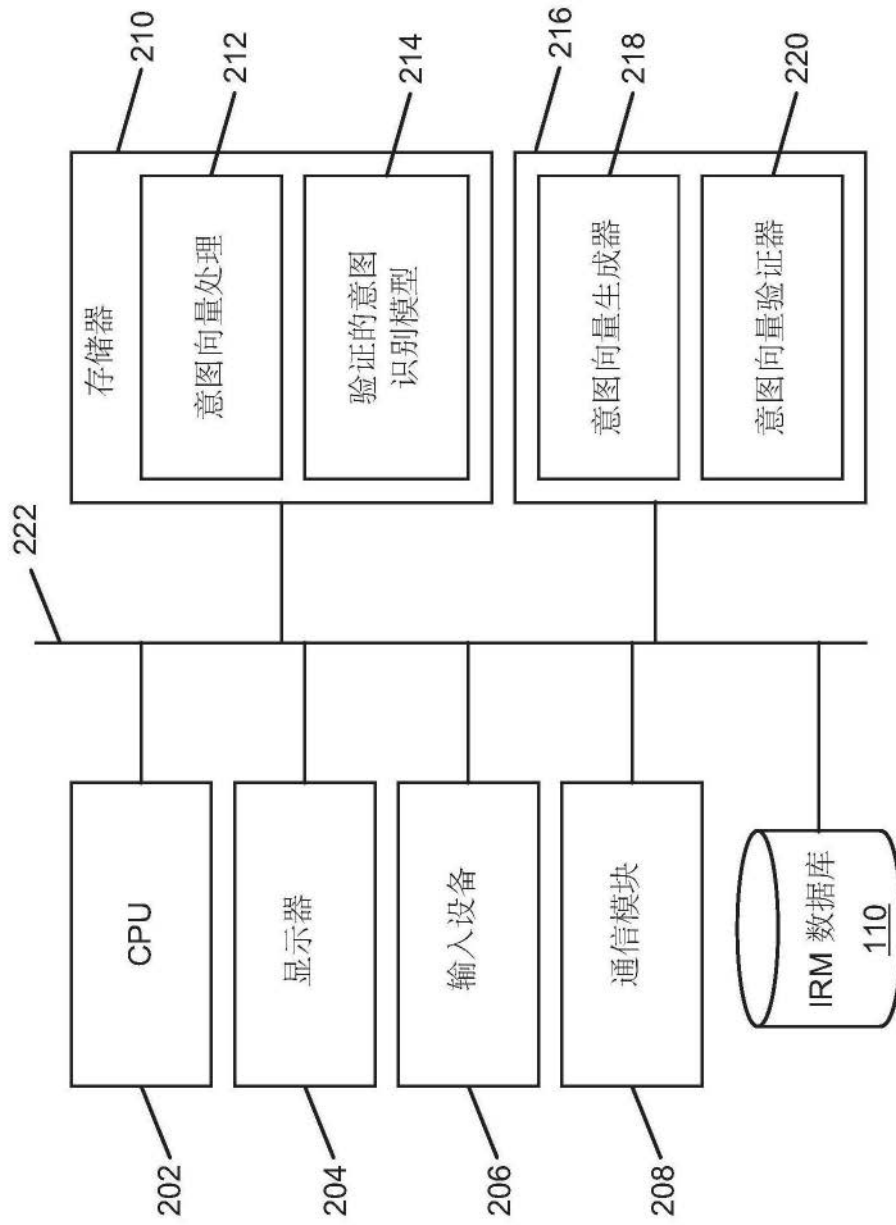


图2



100

图3



200

图4

500

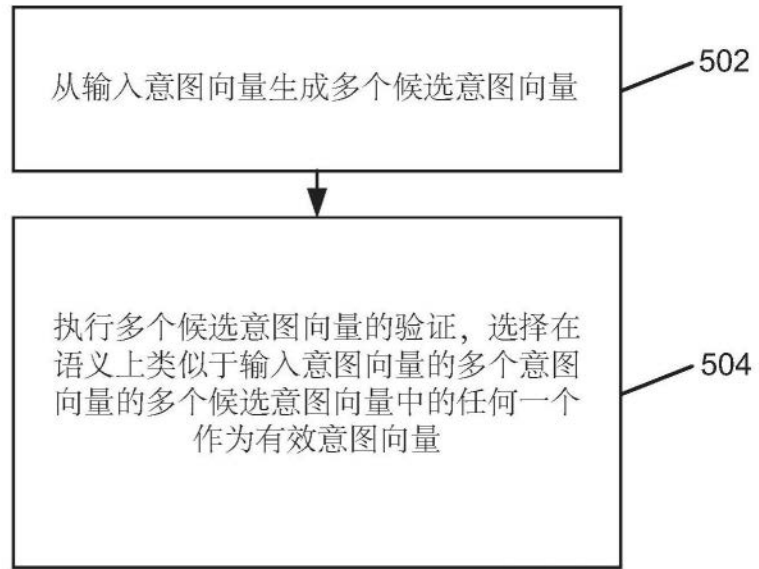


图5

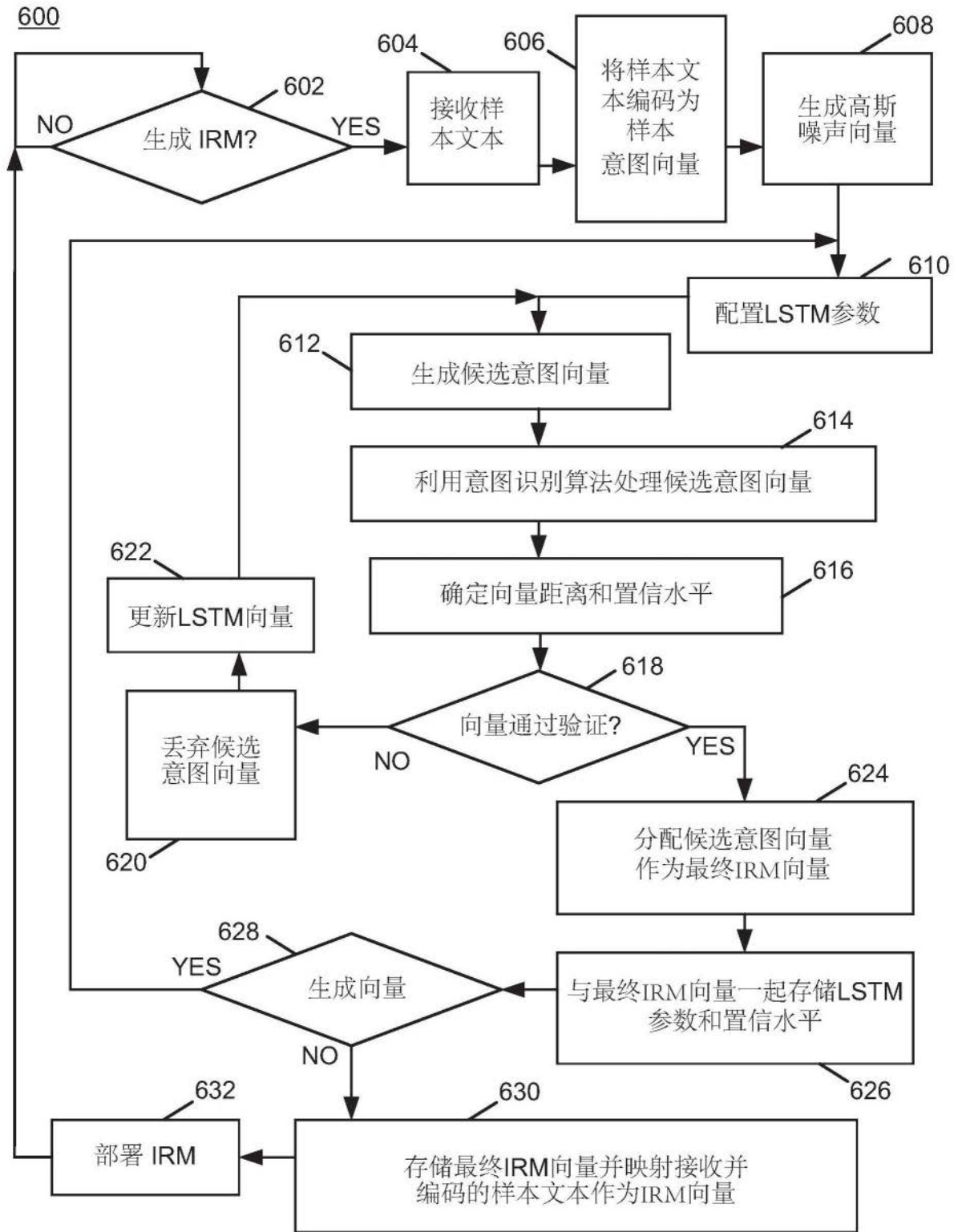


图6