



(12) 发明专利

(10) 授权公告号 CN 109313719 B

(45) 授权公告日 2022.03.22

(21) 申请号 201780008760.9

(22) 申请日 2017.03.17

(65) 同一申请的已公布的文献号
申请公布号 CN 109313719 A

(43) 申请公布日 2019.02.05

(30) 优先权数据
62/310,513 2016.03.18 US(85) PCT国际申请进入国家阶段日
2018.07.27(86) PCT国际申请的申请数据
PCT/US2017/023047 2017.03.17(87) PCT国际申请的公布数据
WO2017/161320 EN 2017.09.21(73) 专利权人 谷歌有限责任公司
地址 美国加利福尼亚州

(72) 发明人 张远 戴维·约瑟夫·魏斯

(74) 专利代理机构 中原信达知识产权代理有限
责任公司 11219

代理人 李宝泉 周亚荣

(51) Int.Cl.

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

(56) 对比文件

CN 102866989 A, 2013.01.09

CN 102662931 A, 2012.09.12

CN 104049755 A, 2014.09.17

CN 104575501 A, 2015.04.29

US 8935151 B1, 2015.01.13

US 7962329 B1, 2011.06.14

US 2015161996 A1, 2015.06.11

WO 9839714 A1, 1998.09.11

US 2015220833 A1, 2015.08.06

US 2007016398 A1, 2007.01.18

WO 2007090033 A2, 2007.08.09

Ronan Collobert et al., Natural
Language Processing (almost) from
Scratch.《Journal of Machine Learning
Research》.2011,第12卷45.

(续)

审查员 黎成超

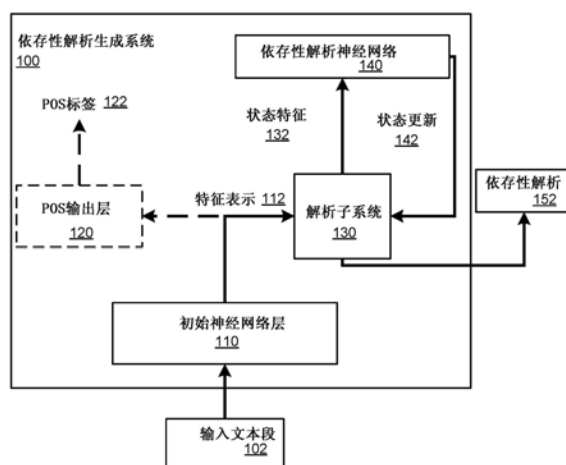
权利要求书3页 说明书8页 附图3页

(54) 发明名称

使用神经网络生成文本段的依存性解析

(57) 摘要

方法、系统和装置包括被编码在计算机存储介质上的计算机程序,用于生成针对输入文本段的依存性解析,其可以作为输入被提供到自然语言处理系统。所述系统中的一个包括第一神经网络,包括:一个或多个初始神经网络层,被配置成针对输入文本序列中的每个标记:接收针对该标记的特征;并且共同地处理该特征以生成用于在确定所述输入文本序列中的该标记的词性中使用的该特征的替选表示;以及依存性解析神经网络,被配置成:处理由所述一个或多个初始神经网络层所生成的、针对所述输入文本序列中的该标记的该特征的该替选表示以生成所述输入文本序列的依存性解析。



[转续页]

[接上页]

(56) 对比文件

Morante R et al..Dependency Parsing and Semantic Role Labeling as a Single Task.《the 7th International Conference on Recent Advances in Natural Language Processing (RANLP-2009)》.2009,6.

Danqi Chen et al..A Fast and Accurate Dependency Parser using Neural Networks.《Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)》.2014,11.

1. 一种由一个或多个计算机实现的系统,所述系统包括:
第一神经网络,包括:
一个或多个初始神经网络层,被配置成对于输入文本序列中的每个标记:
接收针对该标记的特征输入;以及
共同地处理针对该标记的特征输入以生成用于在确定所述输入文本序列中的该标记的词性中使用的该特征输入的备选表示;以及
依存性解析神经网络,被配置成:
处理由所述一个或多个初始神经网络层所生成的、针对所述输入文本序列中的标记的特征输入的备选表示以生成所述输入文本序列的依存性解析,其中,所述依存性解析神经网络是基于状态的依存性解析神经网络,并且其中,所述依存性解析神经网络被配置成通过处理针对当前解析器状态中的特定位置中的标记的特征输入的备选表示而通过更新所述当前解析器状态来生成所述依存性解析,而无需在生成所述依存性解析中直接地使用所述输入文本序列中的标记的词性。
2. 根据权利要求1所述的系统,其中,所述第一神经网络还包括:
输出层,被配置成对于每个标记:
接收该特征输入的该备选表示并且生成多个词性标签中的每个词性标签的相应得分,该相应得分表示对应的词性是所述输入文本序列中的该标记的词性的可能性。
3. 根据权利要求1所述的系统,其中,所述依存性解析神经网络被配置成接收包括针对所述当前解析器状态中的特定位置中的标记的特征输入的备选表示的状态特征输入并且处理该状态特征输入以生成多个解析器状态更新中的每个解析器状态更新的相应得分。
4. 根据权利要求1-3中的任一项所述的系统,进一步包括自然语言处理系统,被配置成:
使用所生成的依存性解析作为输入来分析所述输入文本序列;以及
基于所分析的输入文本序列,使得执行响应于所述输入文本序列的操作。
5. 根据权利要求1所述的系统,其中,所述当前解析器状态将堆栈或缓冲区中的相应位置指派给所述输入文本序列中的所述标记中的至少一些,并标识所述输入文本序列中的所述标记之间的依存性关系。
6. 根据权利要求5所述的系统,其中,所述输入文本序列中的所述标记之间的依存性关系每一个包括依存性三元组,所述依存性三元组标识来自所述输入文本序列的中心标记、来自所述输入文本序列的依存标记以及指定所述中心标记和所述依存标记之间的关系类型的标签。
7. 根据权利要求1所述的系统,其中,给定标记的所述特征输入包括所述给定标记的特征和围绕所述给定标记的窗口中的一个或多个标记的特征。
8. 一个或多个编码有指令的非暂时性计算机可读存储介质,所述指令在由一个或多个计算机执行时,使得所述一个或多个计算机实现神经网络系统,所述神经网络系统包括:
第一神经网络,包括:
一个或多个初始神经网络层,被配置成对于输入文本序列中的每个标记:
接收针对该标记的特征输入;以及
共同地处理针对该标记的特征输入以生成用于在确定所述输入文本序列中的该标记

的词性中使用的该特征输入的备选表示;以及

依存性解析神经网络,被配置成:

处理由所述一个或多个初始神经网络层所生成的、针对所述输入文本序列中的标记的特征输入的备选表示以生成所述输入文本序列的依存性解析,其中,所述依存性解析神经网络是基于状态的依存性解析神经网络,并且其中,所述依存性解析神经网络被配置成通过处理针对当前解析器状态中的特定位置中的标记的特征输入的备选表示而通过更新所述当前解析器状态来生成所述依存性解析,而无需在生成所述依存性解析中直接地使用所述输入文本序列中的标记的词性。

9. 一种用于训练第一神经网络和依存性解析神经网络的方法,

其中,所述第一神经网络包括一个或多个初始神经网络层以及输出层,所述一个或多个初始神经网络层被配置成对于输入文本序列中的每个标记:接收针对该标记的特征输入,以及共同地处理针对该标记的特征输入以生成用于在确定所述输入文本序列中的该标记的词性中使用的该特征输入的备选表示,所述输出层被配置成对于每个标记:接收该特征输入的该备选表示并且生成多个词性标签中的每个词性标签的相应得分,该相应得分表示对应的词性是所述输入文本序列中的该标记的词性的可能性,

并且其中,所述依存性解析神经网络被配置成:

处理由所述一个或多个初始神经网络层所生成的、针对所述输入文本序列中的标记的特征输入的备选表示以生成所述输入文本序列的依存性解析,其中,所述依存性解析神经网络是基于状态的依存性解析神经网络,并且其中,所述依存性解析神经网络被配置成通过处理针对当前解析器状态中的特定位置中的标记的特征输入的备选表示而通过更新所述当前解析器状态来生成所述依存性解析,而无需在生成所述依存性解析中直接地使用所述输入文本序列中的标记的词性,所述方法包括:

获得包括训练文本序列的第一训练数据,并且对于每个训练文本序列,获得该训练文本序列中的每个标记的相应的词性标签;

获得包括训练文本序列的第二训练数据,并且对于每个训练文本序列,获得定义该训练文本序列的依存性解析的解析器状态集合;以及

在所述第一训练数据和所述第二训练数据上训练所述第一神经网络和所述依存性解析神经网络,包括迭代执行以下操作:

在来自所述第一训练数据的训练示例上训练所述第一神经网络以更新所述第一神经网络的参数值;以及

训练所述依存性解析神经网络和所述第一神经网络的所述初始神经网络层包括:

针对来自所述第二训练数据的训练示例来确定解析损耗;以及

通过所述依存性解析神经网络和所述初始神经网络层反向传播所述解析损耗以更新所述依存性解析神经网络和所述初始神经网络层的参数值。

10. 根据权利要求9所述的方法,其中,所述迭代执行包括:

在执行对所述第一神经网络的训练与在来自所述第一训练数据和所述第二训练数据的不同的训练示例上训练所述解析神经网络和所述初始神经网络层之间重复地交替。

11. 根据权利要求9所述的方法,其中,获得所述第二训练数据包括:

对于所述第二训练数据中的训练文本序列中的每个训练文本序列:

获得该训练文本序列的依存性解析;以及

展开该依存性解析以确定定义该训练文本序列的依存性解析的解析器状态集合。

12. 根据权利要求9-11中的任一项所述的方法,其中,在所述第一训练数据和所述第二训练数据上训练所述第一神经网络和所述依存性解析神经网络还包括:

在迭代地执行所述训练操作之前对所述第一神经网络进行预训练。

13. 一种计算机实现的方法,所述方法包括:

使用第一神经网络来处理输入文本序列中的标记,所述第一神经网络包括:

一个或多个初始神经网络层,被配置成对于输入文本序列中的每个标记:

接收针对该标记的特征输入,以及

共同地处理针对该标记的特征输入以生成用于在确定所述输入文本序列中的该标记的词性中使用的该特征输入的备选表示;以及

使用依存性解析神经网络来处理所述备选表示,所述依存性解析神经网络被配置成:

处理由所述一个或多个初始神经网络层所生成的、针对所述输入文本序列中的标记的特征输入的备选表示以生成所述输入文本序列的依存性解析,其中,所述依存性解析神经网络是基于状态的依存性解析神经网络,并且其中,所述依存性解析神经网络被配置成通过处理针对当前解析器状态中的特定位置中的标记的特征输入的备选表示而通过更新所述当前解析器状态来生成所述依存性解析,而无需在生成所述依存性解析中直接地使用所述输入文本序列中的标记的词性。

14. 根据权利要求13所述的方法,其中,所述第一神经网络还包括:

输出层,被配置成对于每个标记:

接收该特征输入的该备选表示并且生成多个词性标签中的每个词性标签的相应得分,该相应得分表示对应的词性是所述输入文本序列中的该标记的词性的可能性。

15. 根据权利要求13所述的方法,其中,所述依存性解析神经网络被配置成接收包括针对所述当前解析器状态中的特定位置中的标记的特征输入的备选表示的状态特征输入并且处理该状态特征输入以生成多个解析器状态更新中的每个解析器状态更新的相应得分。

16. 根据权利要求13所述的方法,进一步包括使用自然语言处理系统来处理所述输入文本序列,所述自然语言处理系统被配置成:

使用所生成的依存性解析作为输入来分析所述输入文本序列;以及

基于所分析的输入文本序列,使得执行响应于所述输入文本序列的操作。

17. 根据权利要求13所述的方法,其中,所述当前解析器状态将堆栈或缓冲区中的相应位置指派给所述输入文本序列中的所述标记中的至少一些,并标识所述输入文本序列中的所述标记之间的依存性关系。

18. 根据权利要求13所述的方法,其中,所述输入文本序列中的所述标记之间的依存性关系每一个包括依存性三元组,所述依存性三元组标识来自所述输入文本序列的中心标记、来自所述输入文本序列的依存标记以及指定所述中心标记和所述依存标记之间的关系类型的标签。

19. 根据权利要求13所述的方法,其中,给定标记的所述特征输入包括所述给定标记的特征和围绕所述给定标记的窗口中的一个或多个标记的特征。

使用神经网络生成文本段的依存性解析

技术领域

[0001] 本说明书涉及使用神经网络生成文本段的依存性解析(dependency parse)。

背景技术

[0002] 神经网络是采用非线性单元的一个或多个层来针对接收到的输入预测输出的机器学习模型。除输出层之外,一些神经网络包括一个或多个隐藏层。每个隐藏层的输出被用作网络中的下一层(即,下一隐藏层或者输出层)的输入。网络的每个层根据相应的参数集的当前值从接收到的输入生成输出。

发明内容

[0003] 自然语言处理是涉及计算机与人类(自然)语言之间的交互的计算机科学的领域。其具有包括被使用来提供人机自然语言接口中的许多应用。这样的接口可以允许人类仅使用其语音控制设备,例如,在“智能环境”中控制。

[0004] 仅基于口述命令或者话语来理解人类的意图可能是可能难以准确地执行的数据密集型任务。如此,理解口述命令的意图常常是不可能的,并且因此,尽管在试图理解命令时耗费了计算资源,意图的操作(例如,以特定的方式控制特定设备)可能仍未被正确地执行或者完全未被确实,。

[0005] 因此,改进自然语言处理的可靠性以便减少这样的情形发生是期望的。增加自然语言处理的速度和/或减少计算资源的使用也可以是期望的。

[0006] 能够被使用在自然语言处理中的一个输入是依存性解析,其一般而言,定义文本段中的词之间的依存性。即,依存性解析定义文本段中的哪些词依赖于文本段中的哪些其它词,并且可选地定义依存性中的每个依存性的类型。本说明书描述被实现为一个或多个位置中的一个或多个计算机上的计算机程序的系统能够如何以经改进的方式生成针对输入文本段的依存性解析,从而潜在地改进关于输入文本段的自然语言处理的性能。

[0007] 在本说明书中所描述的主题的特定实施例能够实现以便实现以下优点中的一个或多个。通过在生成文本段的依存性解析中非直接地使用POS标签,在预测文本段的依存性解析中的POS标签错误的串接被减少并且依存性解析能够准确地被预测。通过不要求词汇化特征由依存性解析器神经网络使用并且通过作为替代重新使用由初始神经网络层所生成的表示,生成依存性解析的模型具有更小的大小(即,是更紧凑的),并且能够比使用词汇化特征的替选方案更快地生成依存性解析。如将理解到,通过更迅速地和/或更准确地生成依存性解析,基于输入文本段和依存性解析来由自然语言处理系统所确定的响应性操作(例如,智能家居中的设备的控制)还能够更迅速地和/或准确地被执行。另外,通过使用允许跨多个任务(即,跨POS标签和依存性解析)的流水线向下反向传播的堆栈的连续形式,模型的组件能够有效率地并且迅速地训练。

[0008] 在附图和以下描述中阐述本说明书中所描述的主题的一个或多个实施例的细节。主题的其它特征、方面和优点将根据说明书、附图和权利要求书而变得显而易见。

附图说明

- [0009] 图1示出了示例依存性解析生成系统。
- [0010] 图2是用于生成用于输入文本段的依存性解析的示例过程的流程图。
- [0011] 图3是用于更新当前解析状态的示例过程的流程图。
- [0012] 图4是用于训练POS标签神经网络和依存性解析神经网络的示例过程的流程图。
- [0013] 各附图中的相同附图标记和标号指示相同元件。

具体实施方式

- [0014] 图1示出了示例依存性解析生成系统100。依存性解析生成系统100是被实现为一个或多个位置中的一个或多个计算机上的计算机程序的系统的示例,其中,下文所描述的系统、组件和技术能够被实现。
- [0015] 依存性解析生成系统100接收输入文本段102并且生成每个输入文本段的相应的依存性解析152。每个输入文本段是标记(token)的序列——即特定自然语言的词的序列并且可选地是特定自然语言的词和标点符号的序列。例如,段能够是句子、句子片段或者另一多词序列。
- [0016] 针对给定文本段的依存性解析是定义该段中的标记之间的依存性关系的数据。特别地,依存性解析标识段中的中心词(head word)和词,以及可选地标识该段中依赖于中心词的标点符号。
- [0017] 可选地,依存性解析还能够包括对于依赖于给定中心词的每个词而言,指定中心词与依存词(dependent work)之间的关系的类型的标签,例如,依存词是否是中心词的形容词补语、中心词的副词修饰、中心词的直接宾语等等。在http://nlp.stanford.edu/software/dependencies_manual.pdf.中描述了能够在依存性解析中的标记之间指定的示例关系类型集。
- [0018] 在一些情况下,依存性解析是三元组的集合,其中每个三元组标识来自段的中心词、来自段的依存词、和指定中心词与依存词之间的关系的类型的标签。
- [0019] 依存性解析生成系统100包括一个或多个初始神经网络层110、解析子系统130和依存性解析神经网络140。在训练期间并且可选地在训练之后,依存性解析生成系统100还包括POS(词性)输出层120。
- [0020] 作为生成针对输入文本段102的依存性解析的一部分,依存性解析生成系统100使用初始神经网络层110处理针对输入文本段102中的每个标记的相应的特征输入。
- [0021] 通常地,针对给定标记的特征输入包括给定标记的特征和输入文本段中围绕给定标记的一个或多个标记(即,输入文本段中在给定标记的窗口内的标记)的特征。
- [0022] 标记的特征是表征标记的向量或者其它数值集合。例如,标记的特征能够包括标识标记(即,标识词或者标点符号)的独热编码特征向量、指示标记是否包括符号(例如,连字符、数字或者标点)的特征向量、指示标记是否是包括前缀或者后缀的词的标志特征向量、指示标记是否是大写的词的标志特征向量等等。
- [0023] 在一些实施方式中,窗口是针对每个特征的相同固定大小。在其它实施方式中,不同的特征具有不同的窗口大小。例如,针对标识标记的独热编码特征向量的窗口大小能够包括距给定标记三个标记内的标记,同时指示大写的标志特征向量能够包括距给定标记一个标

记内的标记。

[0024] 初始神经网络层110是前馈神经网络层,其共同地被配置成对于输入段中的每个标记而言,处理针对该标记的特征输入以生成针对该标记的特征表示112。针对标记的特征表示是针对标记的特征输入的替选表示(即,表示针对标记的特征输入的向量或者其它数值集合)。

[0025] 例如,初始神经网络层110可以包括嵌入层,其后有一个或多个完全连接的隐藏层。

[0026] 嵌入层被配置成接收特征输入中的通常是稀疏的特征以确定针对每个特征的相应的嵌入向量,并且组合(例如,串接)嵌入向量以生成特征的组合嵌入。

[0027] 一个或多个完全连接的隐藏层被配置成处理该组合嵌入以生成针对标记的特征表示(即,通过将一个或多个非线性变换应用到组合嵌入来生成)。

[0028] 解析子系统130接收针对由初始神经网络层110所生成的输入文本段102中的标记的特征表示112并且使用该特征表示112来生成对于依存性解析神经网络140的输入。

[0029] 依存性解析神经网络140是前馈神经网络,其被配置成接收状态特征132(即,由解析子系统130所维持的解析状态的当前配置的特征),并且生成定义对当前解析状态的状态更新142的输出。在一些实施方式中,依存性解析神经网络140包括嵌入层、一个或多个完全连接的隐藏层和输出层。

[0030] 在依存性解析的生成期间的任何给定时间处,由解析子系统130所维持的解析状态将堆栈或者缓冲区中的相应位置指派给输入序列中的标记中的一些或全部并且标识迄今为止所生成的输入段的依存性三元组。在以下文献中更详细地描述了解析状态的示例配置:Danqi Chen and Christopher Manning, A Fast and Accurate Dependency Parser Using Neural Networks. In Proceedings of EMNLP 2014。

[0031] 给定配置的状态特征是多个特征表示的有序组合(例如,矩阵或者串接向量),其中有序组合中的每个位置对应于解析状态的堆栈或者缓冲区中的不同的位置。

[0032] 在一些实施方式中,为了生成针对解析状态的当前配置的状态特征,解析子系统130选择以下来作为有序组合中的每个位置处的特征表示:解析状态的当前配置中的对应的位置中的标记的特征表示。因此,有序组合中的每个位置包括来自解析状态的当前配置中的对应的位置的特征表示。作为一个简化示例,如果有序组合中的位置是[buffer_0; stack_0; stack_1]并且当前配置包括缓冲区中的位置0处的标记1、堆栈中的位置0处的标记2和堆栈中的位置1处的标记3,则用于当前配置的状态特征将是[针对标记1的特征表示、针对标记2的特征表示、针对标记3的特征表示]。如果与有序组合中的位置中的一个位置相对应的位置是空的,则解析子系统130能够将指命的空值特征表示添加到有序组合。

[0033] 在一些实施方式中,有序组合中的每个位置在解析状态的当前配置中具有对应的位置并且在该有序组合位置处的特征是以下的组合(例如,串接):处于对应的配置位置处的标记的特征表示和与根据迄今为止所生成的依存性三元组的对应的配置位置处的标记有关的标记(例如,子标记和同胞标记)的特征表示。如果对应的配置位置处的标记没有具有迄今为止所生成的三元组中的特定关系的标记,则系统能够使用指命的空值特征表示。

[0034] 在一些实施方式中,有序组合还包括迄今为止所生成的依存性三元组的标签的数值表示。

[0035] 分析子系统130和依存性解析神经网络140通过重复地更新解析状态来从特征表示112生成依存性解析152。参考图2和图3下文更详细地描述了生成依存性解析并且更新解析状态。

[0036] POS输出层120被配置成处理输入文本段中的给定标记的特征表示112以生成输出,该输出定义给定标记的POS标签122,该POS标签122定义针对给定标记的词性指派。特别地,POS输出层120是softmax或者其它种类的神经网络输出层,其被配置成处理特征表示112以生成预定POS标签集中的每个POS标签的相应得分。每个得分表示对应的词性是输入文本序列中的标记的词性的可能性。

[0037] 因此,特征表示112能够被用于以下二者:生成输入段102中的标记的POS标签(即,通过选择最高得分POS标签作为标记的POS标签)并且生成输入段102的依存性解析152。然而,依存性解析神经网络140在生成依存性解析152时(即,通过仅使用特征表示112)未直接地使用标记的POS标签。

[0038] 在一些实施方式中,一旦依存性解析152被生成,依存性解析生成系统100便将依存性解析152作为输入提供到另一系统以用于进一步处理。例如,依存性解析生成系统100能够将依存性解析152作为输入提供到自然语言处理系统,其分析输入文本段102,并且引起响应性操作的执行。例如,自然语言处理系统可以被用于促进人机自然语言(例如,语音)接口。在这样的示例中,响应性操作可以包括控制智能环境中的设备。可以由自然语言处理系统执行的操作的其它示例包括情感分析、问题回答、概要、或者另一自然语言处理任务。这些操作能够通过例如智能扬声器或者其它自动化智能个人助理服务被执行例如以作为处理用户请求的一部分。

[0039] 取代或者补充将依存性解析152作为输入提供到另一系统,依存性解析生成系统100能够存储依存性解析或者提供指定依存性解析的数据以供向依存性解析生成系统100的用户呈现。

[0040] 依存性解析生成系统100能够训练初始神经网络层110和依存性解析神经网络140来确定初始神经网络层110和依存性解析神经网络140的参数的训练值。参考图4下文更详细地描述了训练初始神经网络层和依存性解析神经网络。

[0041] 图2是用于生成针对输入文本序列的依存性解析的示例过程200的流程图。为了方便起见,过程200将被描述为由被定位在一个或多个位置中的一个或多个计算机的系统执行。例如,适当地编程的依存性解析生成系统(例如,图1的依存性解析生成系统100)能够执行过程200。

[0042] 该系统获得输入文本段(步骤202)。输入文本段包括根据输入顺序布置的多个标记,即词和可选的即词和标点符号。

[0043] 该系统通过一个或多个初始神经网络层处理针对输入文本段中的标记中的每个标记的特征输入(步骤204)。如上文所描述的,初始神经网络层被配置成处理给定标记的特征输入以生成特征表示——即特征的备选表示,用于在确定标记的词性中使用——即用于由POS标签输出层处理,以及用于在生成输入文本段的依存性解析中使用。

[0044] 系统使用依存性解析神经网络处理备选表示来生成输入文本段的依存性解析(步骤206)。

[0045] 特别地,系统初始化解析状态并且使用依存性解析神经网络重复地更新解析状态

来生成依存性解析。例如,初始解析状态能够包括堆栈中的预定ROOT标记、在缓冲区中的根据该标记在输入段中的位置来布置的输入段中的标记,并且没有已经生成的依存性三元组。系统重复地更新解析状态直到解析状态是终止状态并且然后将终止状态中的依存性三元组视为针对输入文本段的依存性解析。例如,终止状态能够包括空缓冲区和仅堆栈中的ROOT标记。

[0046] 在下面参考图3更详细地描述了更新解析状态。

[0047] 图3是用于更新解析状态的示例过程300的流程图。为了方便起见,过程300将被描述为由被定位在一个或多个位置中的一个或多个计算机的系统执行。例如,适当地编程的依存性解析生成系统(例如,图1的依存性解析生成系统100)能够执行过程300。

[0048] 系统使用当前解析状态来生成状态特征(步骤302)。特别地,系统通过将针对处于解析状态的当前配置中的对应的位置处的标记的特征表示添加到有序组合中的每个位置,生成特征表示的有序组合。

[0049] 系统使用依存性解析神经网络来处理状态特征(步骤304)。依存性解析神经网络被配置成处理状态特征以生成定义对当前解析状态的更新的输出。

[0050] 更特别地,依存性解析神经网络生成潜在解析状态更新集中的每个潜在更新的相应得分。解析状态更新包括从缓冲区移除标记并且将其推送到堆栈上的更新,并且对于依存词与中心词之间的每个可能标签而言:(i)将堆栈上的第二标记标签为依存于第一标记的相应的左弧更新,将对应的标签指派到依存性,并且从堆栈移除第二标记和(ii)将堆栈上的第一标记标签为依存于第二标记的相应的右弧更新,将对应的标签指派到依存性,并且从堆栈移除第一标记。例如,可能的标签能够是在http://nlp.stanford.edu/software/dependencies_manual.pdf中所描述的关系类型集中的关系类型中的一些或全部。

[0051] 系统使用输出来更新当前解析状态(步骤306)。特别地,系统使用得分从来可能的更新中选择更新——即通过选择具有最高得分的更新或者通过根据对应的得分来从可能的更新中进行采样来进行选择,并且将更新应用到当前解析状态以生成更新的解析状态。

[0052] 如果更新的解析状态是终止解析状态,则系统选择由更新的解析状态所定义的依存性解析来作为针对文本段的最后依存性解析。

[0053] 如果更新的解析状态不是预定终止解析状态,则系统重复过程300来利用更新的解析状态代替当前解析状态。

[0054] 图4是用于训练POS标签神经网络和依存性解析神经网络的示例过程400的流程图。为了方便起见,过程400将被描述为由被定位在一个或多个位置中的一个或多个计算机的系统执行。例如,适当地编程的依存性解析生成系统(例如,图1的依存性解析生成系统100)能够执行过程400。

[0055] 系统执行过程400以训练POS标签神经网络——即包括图1的一个或多个初始神经网络层和POS输出层的神经网络和依存性解析神经网络——即图1的依存性解析神经网络)。

[0056] 系统获得第一训练数据,其包括训练文本序列,并且对于每个训练文本序列而言,获得针对该训练文本序列中的每个标记的相应的POS标签(步骤402)。

[0057] 系统获得第二训练数据,其包括训练文本序列,并且对于每个训练文本序列而言,获得定义该训练文本序列的依存性解析的解析器状态的集合(步骤404)。

[0058] 即,对于第二训练数据中的给定训练文本序列而言,系统获得产生依存性解析的(解析器状态,动作)对的集合。每个(解析器状态,动作)对中的动作是状态更新,其应当被应用到该对中的解析器状态以便针对训练文本序列被生成正确依存性解析。

[0059] 在一些实施方式中,系统接收(解析器状态,动作)对。在其它实施方式中,系统接收针对序列的依存性解析并且展开依存性解析以确定针对序列的(解析器状态,动作)对。

[0060] 然后,系统在第一训练数据和第二训练数据上训练POS神经网络和依存性解析神经网络以确定初始神经网络层、POS输出层(如果输出层具有可训练的参数)和依存性解析神经网络的参数的训练值(步骤306)。

[0061] 特别地,系统通过迭代地执行以下操作来训练网络:(i)在来自第一训练数据的训练示例上训练POS神经网络以更新POS神经网络的参数的值;和(ii)在来自第二训练数据的训练示例上训练解析神经网络和初始神经网络层。

[0062] 来自第一训练数据的训练示例是针对来自第一训练数据中的训练序列中的一个的标记的特征输入和针对该标记的对应的POS标签。

[0063] 来自第二训练数据的训练示例是与第二训练数据中的(解析状态,动作)对相对应的状态特征和来自该(解析状态,动作)对的动作。

[0064] 系统能够从对应的训练数据中的训练示例中随机地选择用于操作(i)和(ii)的每个迭代的训练示例。

[0065] 为了迭代地执行操作(i)和(ii),系统能够重复地在执行操作(i)与执行操作(ii)之间交替。例如,系统能够在以第一数目的训练回合(epoch)执行操作(i)与以第二数目的训练回合执行操作(ii)之间交替。在一些实施方式中,第一数目等于第二数目。在一些其它实施方式中,第一数目与第二数目不同。系统能够在以预定数目的迭代执行操作之间交替,直到训练数据中的所有训练示例已经被采样,或者直到某个其它终止准则被满足。

[0066] 为了在来自第二训练数据的训练示例上训练解析神经网络和初始神经网络层,系统确定训练示例的解析损耗的梯度并且通过依存性解析神经网络和初始神经网络层反向传播该梯度以更新依存性解析神经网络和初始神经网络层的参数的值——即使用利用反向传播训练技术的随机梯度下降。

[0067] 特别地,系统根据以下来确定解析损耗:(i)由解析神经网络通过处理训练示例中的状态特征所生成的得分与(ii)将训练示例中的动作识别为待应用到状态的状态更新的得分集合之间的误差。

[0068] 为了在来自第一训练数据的训练示例上训练POS神经网络,系统确定POS损耗的梯度并且通过POS输出层和初始神经网络层反向传播该梯度以更新初始神经网络层的参数的值——即使用利用反向传播训练技术的相同随机梯度下降或者利用反向传播训练技术的不同的随机梯度下降)。

[0069] 特别地,系统根据以下来确定POS损耗:(i)由POS神经网络通过处理针对训练示例中的标记的特征输入所生成的得分与(ii)指示训练示例中的POS标签是针对标记的正确POS标签的得分集合之间的误差。

[0070] 在一些实施方式中,系统在迭代地执行训练操作之前对POS神经网络进行预训练。即,系统通过在开始执行操作(ii)之前重复地执行操作(i)来对POS神经网络进行预训练。

[0071] 能够在数字电子电路中、在有形实现的计算机软件或固件中、在计算机硬件

中——包括本说明书中所公开的结构和其结构等同物、或者其中的一个或多个的组合中实现本说明书中所描述的主题和功能操作的实施例。

[0072] 本说明书中所描述的主题的实施例能够实现为一个或多个计算机程序(即,被编码在有形非暂态程序载体上用于由数据处理装置执行或者控制数据处理装置的操作的计算机程序指令的一个或多个模块)。替选地或者附加地,程序指令能够被编码在人工生成的传播信号上,例如,机器生成的电气、光学或电磁信号,其被生成以编码用于传输到合适的接收器装置供数据处理装置执行的信息。计算机存储介质能够是机器可读存储设备、机器可读存储基底、随机或串行存取存储器设备、或它们中的一个或多个的组合。

[0073] 术语“数据处理装置”涵盖用于处理数据的所有种类的装置、设备和机器,以示例的方式包括可编程处理器、计算机或多个处理器或计算机。装置能够包括专用逻辑电路(例如,FPGA(现场可编程门阵列)或ASIC(专用集成电路))。除硬件之外,装置还能够包括创建用于讨论中的计算机程序的执行环境的代码,例如构成处理器固件、协议栈、数据库管理系统、操作系统、或者它们中的一个或多个的组的代码。

[0074] 计算机程序(其还可以被称为或被描述为程序、软件、软件应用、模块、软件模块、脚本或代码)能够以任何形式的编程语言编写,包括编译或者解释语言或者声明性或者过程性语言,并且其能够以任何形式部署,包括作为单独程序或者作为模块、组件、子例程或适于使用在计算环境中的其它单元。计算机程序可以但是不需要对应于文件系统中的文件。程序可以被存储在保持其它程序或数据(例如,被存储在标记语言文档中的一个或多个脚本)的文件的一部分中、在专用于讨论中的程序的单个文件中或在多个协调文件中(例如,存储一个或多个模块、子程序或代码的部分的文件)。计算机程序可以被部署以在被定位在一个地点处的一个计算机上或在跨多个地点分布并且由通信网络相互连接的多个计算机上被执行。

[0075] 本说明书中所描述的过程和逻辑流程可以通过一个或多个可编程处理器执行,其通过对输入数据进行操作并且生成输出执行一个或多个计算机程序以执行动作。过程和逻辑流程还能够由专用逻辑电路(例如,FPGA(现场可编程门阵列)或ASIC(专用集成电路))来执行,并且装置还能够被实现为专用逻辑电路(例如,FPGA(现场可编程门阵列)或ASIC(专用集成电路))。

[0076] 通过示例,适于计算机程序的执行的计算机包括能够基于通用微处理器或者专用微处理器或者二者,或者任何其它种类的中央处理单元。一般地,中央处理单元将从只读存储器或者随机存取存储器或者二者接收指令和数据。计算机的基本元件是用于执行或者运行指令的中央处理单元和用于存储指令和数据的一个或多个存储器设备。通常,计算机将还包括用于存储数据的一个或多个海量存储设备(例如,磁性、磁光盘或者光盘)或操作性地耦合以从其接收数据、向其传送数据到其或者两者。然而,计算机不需要具有这样的设备。而且,计算机能够被嵌入在另一设备中(例如,移动电话、个人数字助理(PDA)、移动音频或视频播放器、游戏控制台、全球定位系统(GPS)接收器)或者便携式存储设备(例如,通用串行总线(USB)闪存驱动器),仅举几例。

[0077] 适于存储计算机程序指令和数据的计算机可读介质包括所有形式的不挥发的性存储器、介质和存储器设备,通过示例包括半导体存储器设备(例如,EPROM、EEPROM和闪存存储器设备);磁盘(例如,内部硬盘或者可移除磁盘);磁光盘;以及CD ROM和DVD-ROM光盘。

处理器和存储器能够由专用逻辑电路补充或者并入专用逻辑电路中。

[0078] 为了提供与用户的交互,本说明书中所描述的主题的实施例能够被实现在计算机上,所述计算机具有用于将信息显示给用户的显示设备(例如,CRT(阴极射线管)或LCD(液晶显示器)监视器)和用户通过其能够向计算机发送输入的键盘和指示设备(例如,鼠标或轨迹球)。其他种类的设备也可以被用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的感官反馈,例如视觉反馈、听觉反馈或者触觉反馈;并且来自用户的输入可以以任何形式接收,包括声音、语音或者触觉输入。另外,计算机能够通过将文档发送到由用户所使用的设备并且从其接收文档与用户交互;例如,通过响应于从web浏览器所接收到的请求,将网页发送到用户的客户端设备上的web浏览器。

[0079] 在本说明书中所描述的主题的实施例能够被实现在计算系统中,其包括后端组件(例如,作为数据服务器),或者其包括中间件组件(例如,应用服务器),或者其包括前端组件(例如,具有用户通过其能够与本说明书中所描述的主题的实施方式交互的图形用户接口或网络浏览器的客户端计算机),或者一个或多个这样的后端、中间件或前端组件的任何组合。系统的组件能够以任何形式或数字数据通信例如通信网络的介质相互连接。通信网络的示例包括局域网(“LAN”)和广域网(“WAN”)(例如,因特网)。

[0080] 计算系统能够包括客户端和服务器。客户端和服务器通常远离彼此并且典型地通过通信网络交互。客户端和服务器的关系借助于在相应的计算机上运行并且具有彼此的客户端-服务器关系的计算机程序生成。

[0081] 虽然本说明书包含许多特定实施方式细节,但是这些不应该被解释为对任何发明或可以要求保护的内容的范围的限制,而是可以解释为特定于本发明的特定实施例的特征的描述。在本说明书中在分离的实施例的语境中所描述的某些特征还能够组合被实现在单个实施例中。相反地,在单个实施例中的语境中所描述的各种特征还能够分离地或者以任何适合的子组合被实现在多个实施例中。而且,尽管特征可以上文描述为在某些组合中作用并且甚至如此初始地要求保护,但是在一些情况下,可以从组合去除所要求保护的组合的一个或多个特征,并且所要求保护的组合可以涉及子组合或子组合的变型。

[0082] 类似地,虽然操作以特定次序在附图中描绘,但是这不应该被理解为要求这样的操作以所示的特定次序或者以顺序次序执行,或者全部所图示的操作被执行以实现期望的结果。在某些情况下,多任务和并行处理可以是有利的。而且,上文所描述的实施例中的各种系统模块和组件的分离不应该被理解为要求所有实施例中的这样的分离,并且应该理解的是,所描述的程序组件和系统可以一般地一起集成在单个软件产品或者封装到多个软件产品中。

[0083] 已经描述本主题的特定实施例。其他实施例在所附权利要求书的范围内。例如,权利要求书中所记载的动作能够以不同的次序执行并且仍然实现期望的结果。作为一个示例,附图中所描绘的过程不必要求所示的特定次序或顺序次序来实现期望的结果。在某些实施方式中,多任务和并行处理可以是有利的。

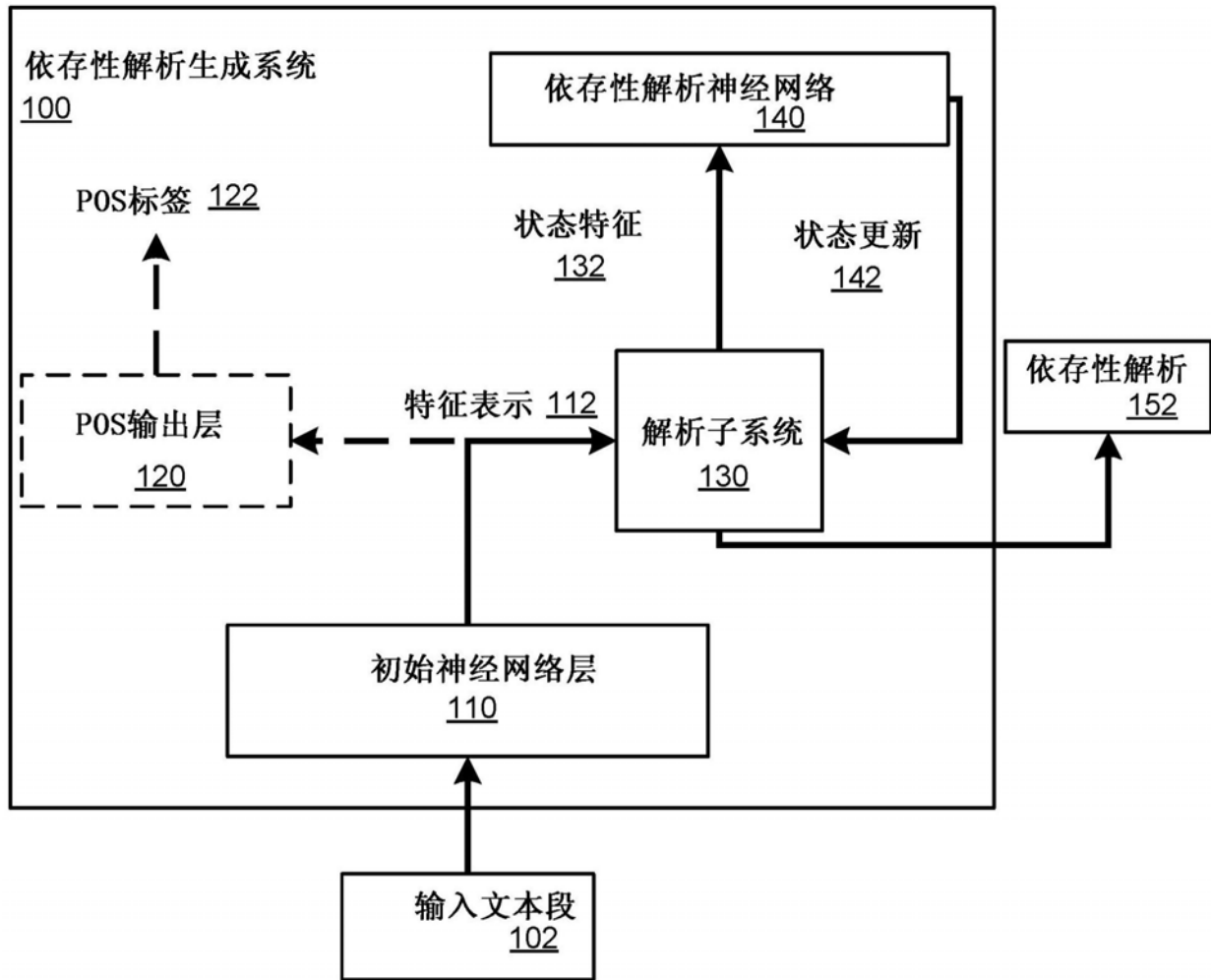


图1

200

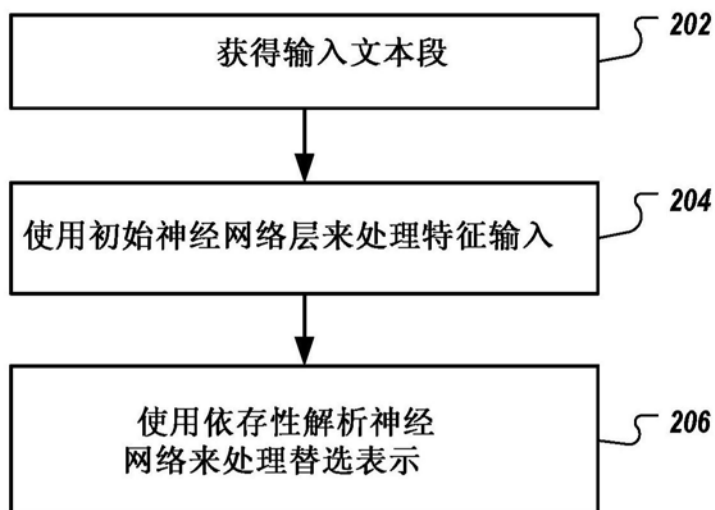


图2

300

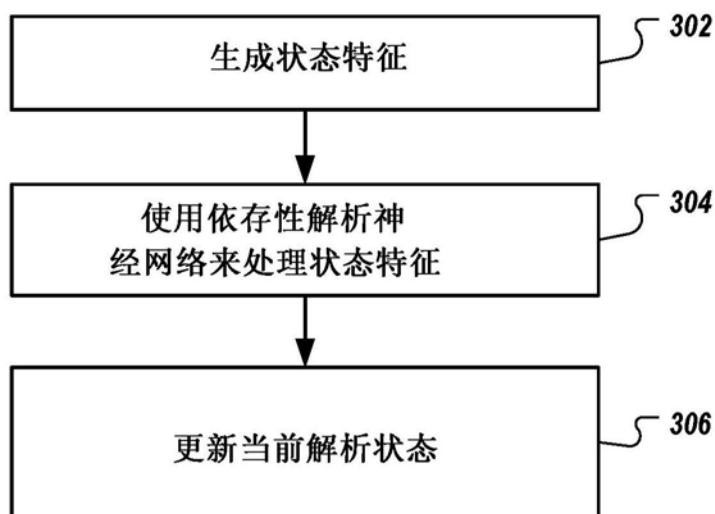


图3

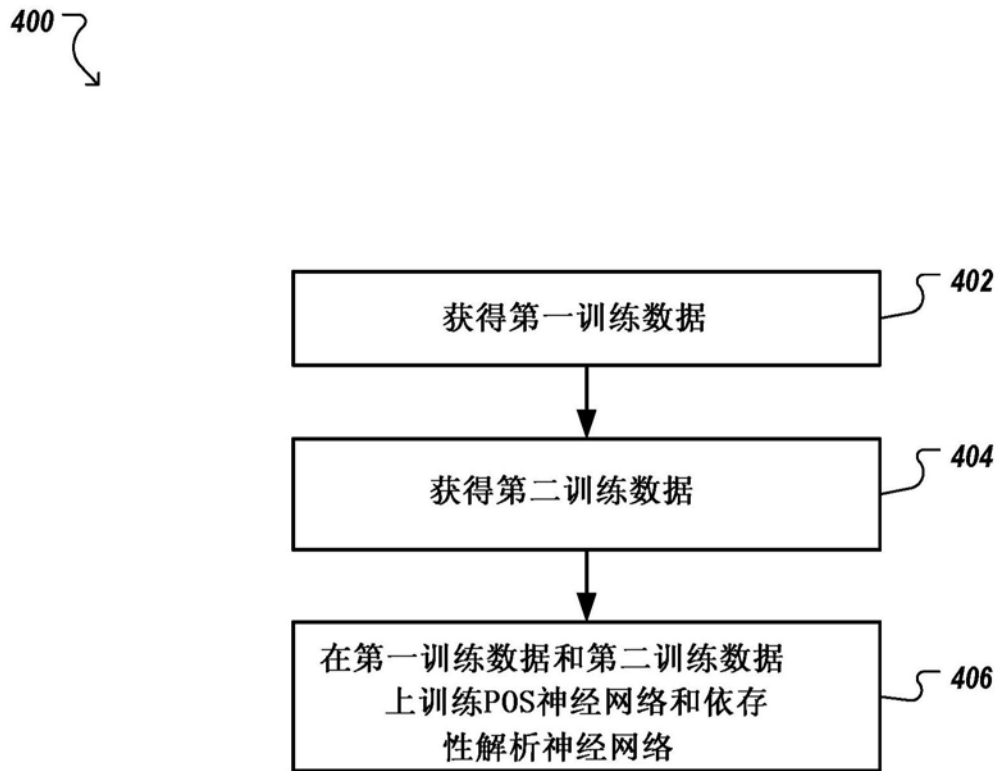


图4