(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 110678882 A
(43)申请公布日 2020.01.10

(21)申请号 201880036021.5

(22)申请日 2018.10.29

(30)优先权数据
62/578,347 2017.10.27 US

(85)PCT国际申请进入国家阶段日
2019.11.29

(86)PCT国际申请的申请数据
PCT/US2018/058036 2018.10.29

(87)PCT国际申请的公布数据
WO2019/084558 EN 2019.05.02

(71)申请人 谷歌有限责任公司
地址 美国加利福尼亚州

(72)发明人 T.M.奎亚特科夫斯基 A.P.帕里克 S.斯瓦亚姆蒂普塔

(74)专利代理机构 北京市柳沈律师事务所 11105
代理人 金玉洁

(51)Int.Cl.
G06N 3/04(2006.01)
G06N 3/08(2006.01)
G06F 16/30(2006.01)

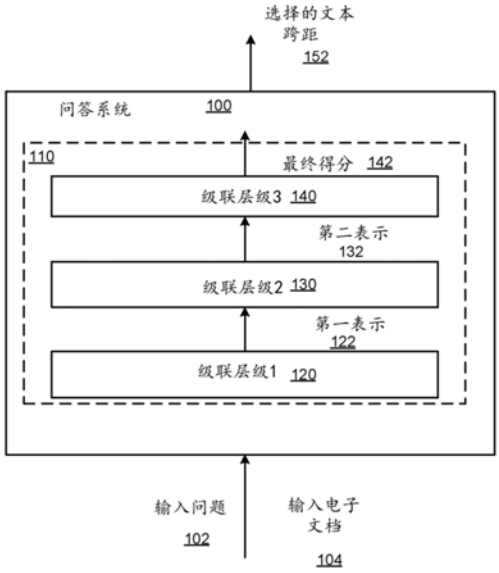
权利要求书2页 说明书9页 附图4页

(54)发明名称

使用机器学习从电子文档选择回答跨距

(57)摘要

一种包括在计算机储存介质上编码的计算机程序的方法、系统和设备，以用于从输入电子文档选择回答输入问题的文本跨距。方法之一包括在输入文档中获得文本跨距的相应第一数值表示；对于文本跨距中的每一个：对于包含文本跨距的分段，确定问题-意识分段向量，对于问题，确定分段-意识问题向量，以及使用第二前馈神经网络处理文本跨距的第一数值表示、问题-意识分段向量和分段-意识问题向量以生成文本跨距的第二数值表示；对于多个文本跨距的每个唯一文本跨距：确定唯一文本跨距的聚合表示，且从聚合表示确定唯一文本跨距的最终得分；以及选择唯一文本跨距。



CN 110678882 A

1.一种计算机实现方法,从输入电子文档选择回答包含多个问题标记的输入问题的文本跨距,所述方法包括:

在所述输入文档中获得多个文本跨距中的每一个的相应的第一数值表示;

对于所述多个文本跨距中的每一个:

对于所述输入文档中包含所述文本跨距的分段,基于所述问题中的所述问题标记和包含所述文本跨距的分段中的分段标记之间的相似度来确定问题-意识分段向量,

对于所述问题,同样基于所述问题中的所述问题标记和包含所述文本跨距的分段中的分段标记之间的相似度来确定针对所述问题的分段-意识问题向量,以及

使用第二前馈神经网络处理所述文本跨距的第一数值表示、所述问题-意识分段向量和所述分段-意识问题向量以生成所述文本跨距的第二数值表示;对于所述多个文本跨距中的每个唯一文本跨距:

从与所述唯一文本跨距对应的所述文本跨距的第二数值表示确定所述唯一文本跨距的聚合表示,且

从所述聚合表示确定所述唯一文本跨距的最终得分,所述最终得分衡量所述唯一文本跨距回答所述问题的良好程度;以及

选择具有最高最终得分的唯一文本跨距作为对所述问题的回答。

2.根据权利要求1所述的方法,还包括:

响应于所述问题,输出所选择的唯一文本跨距。

3.根据权利要求2所述的方法,其中所述问题作为语音输入被接收,并且其中输出所述唯一文本跨距包括:

输出所述文本跨距的口头话语作为对所述问题的响应的一部分。

4.根据权利要求1-3中任一项所述的方法,其中确定所述唯一文本跨距的聚合表示包括:

使用第三前馈神经网络处理与所述唯一文本跨距对应的所述文本跨距中每一个的所述第二数值表示,以生成所述文本跨距中每一个的相应转换的数值表示;以及

通过求和所述转换的数值表示来确定所述聚合表示。

5.根据权利要求1-4中任一项所述的方法,其中确定所述唯一文本跨距的最终得分包括:

使用线性预测层处理所述唯一文本跨距的聚合表示以生成所述最终得分。

6.根据权利要求1-5中任一项所述的方法,其中对于所述输入文档中包含所述文本跨距的分段确定所述问题-意识分段向量包括:

确定每个分段标记的相应的伴随向量,所述伴随向量说明所述分段标记与所述问题标记的相似度;以及

从所述分段标记的伴随向量确定所述问题-意识分段向量。

7.根据权利要求1-6中任一项所述的方法,其中对于所述问题确定所述分段-意识问题向量包括:

确定每个问题标记的相应的伴随向量,所述伴随向量衡量所述问题标记与所述分段标记的相似度;以及

从所述问题标记的伴随向量确定所述分段-意识问题向量。

8.根据权利要求1-7中任一项所述的方法，其中所述第二数值表示是所述第二前馈神经网络中的最后一个隐藏层的输出。

9.根据权利要求1-8中任一项所述的方法，其中获得所述输入文档中的所述多个文本跨距中的每一个的第一数值表示包括，对于每个文本跨距：

基于所述文本跨距中的标记，获得所述文本跨距的初始表示；

基于所述问题标记，获得所述问题的初始表示；以及

从所述文本跨距的初始表示和所述问题的初始表示确定所述文本跨距的问题-跨距表示。

10.根据权利要求9所述的方法，其中获得所述输入文档中的所述多个文本跨距中的每一个的相应的第一数值表示包括，对于每个文本跨距：

基于所述文本跨距的左上下文中的所述标记，获得所述文档中所述文本跨距的左上下文的初始表示；

基于所述文本跨距的右上下文中的所述标记，获得所述文档中所述文本跨距的右上下文的初始表示；

以及

从所述文本跨距的初始表示以及所述左上下文的初始表示和所述右上下文的初始表示确定所述文本跨距的跨距-上下文表示。

11.根据权利要求10所述的方法，其中所述第一表示是所述问题-跨距表示和所述跨距-上下文表示的拼接。

12.根据权利要求9-11中任一项所述的方法，其中所述文本跨距的所述初始表示是在所述文本跨距中所述标记的单词嵌入的袋。

13.根据权利要求9-11中任一项所述的方法，其中所述文本跨距的初始表示是所述文本跨距中所述标记的单词嵌入的袋和问题-单词特征的拼接，所述问题-单词特征指示所述文本跨距是否包含任何所述问题标记。

14.一种系统，包括一个或多个计算机和储存指令的一个或多个储存装置，所述指令在由所述一个或多个计算机执行时使得所述一个或多个计算机执行权利要求1-13中任一项中的相应方法的所述操作。

15.一种储存指令的计算机储存介质，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执行权利要求1-13中任一项中的相应方法的所述操作。

使用机器学习从电子文档选择回答跨距

[0001] 相关申请的交叉引用

[0002] 本申请要求在2017年10月27日提交的美国专利申请No.62/578,347的优先权，其全部内容通过引用由此并入本文。

背景技术

[0003] 本说明书涉及使用如神经网络的机器学习模型来处理电子文档。

[0004] 电子文档可以是各种文档的任何一种，其可以用电子形式保存且可以由用户在计算机上观看，如网页、文字处理文档、文本文档、电子表格等。

[0005] 神经网络是机器学习模型，其采用非线性单元的一个或多个层以针对接收的输入预测输出。一些神经网络除了输出层之外包括一个或多个隐藏层。每个隐藏层的输出用作向网络中的下一层(即，下一个隐藏层或输出层)的输入。网络的每个层根据参数的相应集合的当前值从接收的输入生成输出。

[0006] 一些神经网络是循环神经网络。循环神经网络是接收输入序列且从输入序列生成输出序列的神经网络。特别地，循环神经网络可以将来自先前时间步骤的网络的内部状态的一些或全部用于计算当前时间步骤时的输出。

发明内容

[0007] 本说明书描述了在一个或多个位置中的一个或多个计算机上实现为计算机程序的系统，其从输入电子文档选择回答输入问题的文本跨距，该输入问题包括多个问题标记。

[0008] 本说明书中所描述的主题可以在特定实施例中实现，以便实现如下优点中的一个或多个。

[0009] 通过采用以级联进行组合的轻量级(即有计算效率的)模型来找到对输入问题的回答，描述的系统可以在输入文档中有效地定位回答输入问题的文本。特别地，描述的系统可以胜任更加复杂、较低计算效率的架构。因此，描述的系统可以有效地回答接收的问题，而与常规方法相比消耗更少的计算资源，例如更少的内存和更低的处理能力，这当系统在资源受限环境中(例如在移动装置上)实现时是特别有利的。特别是，尽管与以前的先进系统(例如使用计算密集型循环神经网络的系统)相比消耗更加少的计算资源，但是该系统可以对许多问答任务获得最新的结果以处理文档标记、问题标记或两者。

[0010] 下面在所附附图和描述中提出在本说明书中的主题的一个或多个实施例的细节。通过说明书、附图和权利要求书，主题的其他特征、方面和潜在优势将变得显而易见。

附图说明

[0011] 图1A示出了示例的问答系统。

[0012] 图1B示出了级联的机器学习系统的示例架构。

[0013] 图2是训练级联的机器学习系统的示例过程的流程图。

[0014] 图3是从输入文档选择回答跨距的示例过程的流程图。

[0015] 在各附图中的相同的附图标记和命名指示相同的元件。

具体实施方式

[0016] 本说明书总体上描述从电子文档选择回答接收的问题的文本跨距(text span)的系统。文本跨距是电子文档中的一个或多个连续单词的序列。

[0017] 一旦系统已经选择文本跨距作为对问题的回答，系统(或其他系统)可以输出选择的文本跨距作为对问题的响应的一部分。

[0018] 例如，输入问题可以已经作为语音查询提交，并且系统可以提供选择的文本跨距的口头话语作为对查询的响应的一部分。作为特定示例，使用语音输入与用户交互的移动装置、智能扬声器、或其他计算装置可以如在数据通信网络上接收由用户所说的语音查询并向系统传输接收的查询。然后系统可以标识可能包含对接收的查询的回答的候选电子文档，使用本说明书中描述的技术从文档选择文本跨距，并且然后向计算装置传输文本跨距作为对语音查询的响应的一部分，即作为表示文本跨距的口头话语的数据或者作为转换成计算装置处的语音的文本。在一些情况下，用户可以明确地或隐含地标识候选文档。例如，如果用户在使用计算装置观看给定的文档时已提交语音查询，则系统可以标识给定的文档作为候选电子文档。在一些其它情形下，外部系统(如互联网搜索引擎)响应于查询而标识候选电子文档，且向系统提供候选电子文档。

[0019] 作为其他示例，系统可以接收作为文本查询的问题，并且可以提供文本跨距以在用户装置上演示作为对文本查询的响应的一部分。例如，互联网搜索引擎可以接收文本查询，并且可以由互联网搜索引擎包括由系统标识的文本跨距作为搜索查询的响应的一部分，如作为内容的格式表示以及互联网搜索引擎标识的搜索结果作为对查询的响应。

[0020] 图1A示出了示例的问答系统100。问答系统100是在一个或多个位置中的一个或多个计算机上实现为计算机程序的系统的示例，其中实现下文所描述的系统、组件和技术。

[0021] 如上文所描述，系统100接收输入问题102和输入电子文档104，并且从该系统已经确定的电子文档104标识文本跨距152，以提供对输入问题102的回答。特别地，输入问题102和电子文档104的两者都被标记化(tokenized)，即使得输入问题102和电子文档104二者的文本表示为标记(token)的相应集合。标记可以例如是选自可能标记的词汇表中的单词、短语或其他元组(n-gram)。

[0022] 当接收电子文档104时，系统100在文档中标识候选文本跨距。例如，系统100可以在文档中标识一个或多个标记的每个可能连续序列作为候选文本跨距，该候选文本跨距包括少于阈值数目的标记。

[0023] 因为相同候选文本跨距可以在整个电子文档中多次发生，所以系统100还从文档中的候选文本跨距标识唯一文本跨距(unique text span)的集合，即使得在唯一文本跨距的集合中的任何文本跨距都不对应于唯一文本跨距中的其他任何文本跨距。作为一个示例，如果两个文本跨距是在彼此的阈值编辑距离内，则系统100可以认为一个文本跨距对应于另一个文本跨距。作为另一个示例，如果两个文本跨距由命名实体识别系统确定为指代的相同实体，则系统100可以认为它们对应。

[0024] 系统100然后使用级联的机器学习系统110(即具有级联模型架构的机器学习系统)，以从唯一文本跨距的集合中选择文本跨距作为回答该输入问题的文本跨距152。

[0025] 级联的模型结构具有三个层级的机器学习模型:层级1 120,层级2 130,以及层级3 140。架构被称为“级联”,因为级联的每个层中的(多个)模型将级联的前一个层中的(多个)模型的输出作为输入接收。级联的最终层(即层3)中的(多个)模型从前一个层(即层2)中的模型的输出生成机器学习系统110的最终预测。

[0026] 更具体地,级联的层级1对问题的简单特征和候选文本跨距进行操作,以生成每个文本跨距的相应的第一数值表示122。数值表示是数值的有序集合,例如浮点值或量化浮点值的向量、矩阵、或者更高阶张量。

[0027] 特别地,层级1中的(多个)模型仅对来自预先训练的标记嵌入的字典的嵌入(embedding)以及可选地指示给定的跨距是否包含来自问题的标记的二进制问题词特征进行操作。嵌入是固定维度空间中的数值的向量。因为嵌入已经被预先训练,嵌入在固定维度空间中的位置反映了它们表示的标记之间的相似度,例如语义相似度。作为一个示例,在固定维度空间中,单词“国王(king)”的嵌入可以比单词“兵卒(pawn)”的嵌入更接近单词“王后(queen)”的嵌入。可以由系统100使用的这种预先训练的嵌入的示例包括word2vec嵌入和GloVe嵌入。

[0028] 级联的层2中的模型使用由层级1生成的第一数值表示122以及注意机制,该注意机制对于每个候选跨距将问题标记与含有候选跨距的文档分段中的标记(例如电子文档中的包含候选跨距的句子、段落或标记的其它组)对齐(align),以对于每个候选回答跨距生成相应第二数值表示132。

[0029] 层3中的模型接收候选文本跨距的第二数值表示132并且从在文档中多次提及(即在整个文档中多次发生)的所有候选回答跨距聚集信息,以便为每个唯一文本跨距确定相应最终得分142。给定的唯一文本跨距的最终得分142衡量唯一文本跨距回答问题的良好程度。

[0030] 参考图1B和3下面将更详细地描述级联的机器学习系统110的操作。

[0031] 系统100然后基于最终得分从唯一文本跨距选择文本跨距152。例如,系统100可以选择具有最高最终得分的唯一文本跨距作为对问题的回答。

[0032] 为了允许级联的机器学习系统110有效地对回答跨距打分,即,使得由级联的层3生成的最终得分可以用于准确地标识对输入问题的回答,系统100在包括标记的(labeled)训练示例的训练数据上训练级联中的机器学习模型。换言之,每个标记的训练示例包括用标识正确文本跨距(即来自电子文档的最佳回答问题的文本跨距)的数据标记的问题-电子文档对。参考图1B和图2在下文更详细地描述在该训练数据上训练级联中的机器学习模型。

[0033] 图1B示出了级联的机器学习系统110的示例架构。

[0034] 如图1B中所示的,级联的层级1包括共同生成第一数值表示的两个模型:跨距+短上下文模型160和问题+跨距模型170。

[0035] 对于任何给定的文本跨距,模型160对以下进行操作:(i)输入文档中文本跨距的左上下文(context)的初始表示154,(ii)文本跨距的初始表示156,以及(iii)输入文档中的文本跨距的右上下文的初始表示158,以生成文本跨距的跨距-上下文表示162作为输出。

[0036] 基于文本跨距中的标记的预先训练的嵌入生成文本跨距的初始表示。在一些实现方式中,文本跨距的初始表示是文本跨距中标记的单词嵌入的袋(bag of words embedding),即文本跨距中标记的嵌入的平均值。在一些其它的实现方式中,文本跨距的初

始表示是文本跨距中标记的单词嵌入的袋和指示文本跨距是否包括任何问题标记的问题-单词特征的拼接(concatenation)。问题-单词特征可以是二进制特征,例如当文本跨距包括一个或多个问题标记时其值为1,并且当文本跨距不包括任何问题标记时其值为0。

[0037] 左上下文的初始表示是文本跨距的左上下文中的标记的单词嵌入的袋,即输入文档中文本跨距的最左边的K标记的嵌入的平均值。

[0038] 相似地,右上下文的初始表示是文本跨距的右上下文中的标记的单词嵌入的袋,即输入文档中文本跨距的最右边的K标记的嵌入的平均值。

[0039] 为了生成文本跨距的跨距-上下文表示,模型160使用前馈神经网络来处理以下的拼接:(i)输入文档中的文本跨距的左上下文的初始表示,(ii)文本跨距的初始表示,以及(iii)输入文档中的文本跨距的右上下文的初始表示。在一些实现方式中,神经网络是具有修正的线性单元(rectified linear unit,ReLU)激活的两层前馈神经网络。特别地,在这些实现方式中,由前馈神经网络执行的操作以从输入x生成表示h可以表达为:

[0040] h=ffnn(x)

[0041] =ReLU{U{ReLU{Vx+a}}+b},

[0042] 其中U和V参数矩阵以及a和b是前馈网络的参数偏差。

[0043] 虽然在推断期间没有使用,在训练期间,模型160还配置为生成文本跨距的得分(比如最终得分),其衡量唯一文本跨距回答问题的良好程度(如图1B所示,作为对损失项l_2的输入)。特别地,模型160可以通过由将向量映射到单个值的线性预测层处理文本跨距的跨距-上下文表示162生成得分。特别地,由线性预测层执行的操作以从输入表示h生成值φ可以表达为:

[0044]
$$\varphi = \text{linear}(\mathbf{h}) = \mathbf{w}^{\mathrm{T}}\mathbf{h} + \mathbf{z},$$

[0045] 其中w和z是线性预测层的参数。

[0046] 下文将详细地描述为了训练使用由模型160生成的得分。

[0047] 对于任何给定的文本跨距,模型170对(i)文本跨距156的初始表示和(ii)问题的初始表示164进行操作,以生成文本跨距的问题-跨距表示172。

[0048] 特别地,模型170首先基于每个问题标记的嵌入生成问题标记的每一个的权重。

[0049] 模型170可以通过首先向问题标记的嵌入施加其他前馈神经网络(即施FFNN操作)来生成问题标记的权重,以生成问题标记的初始表示,并且然后向问题标记的初始表示施加另一个线性预测层。

[0050] 模型170然后可以通过计算问题标记的嵌入的加权平均值来生成问题标记的初始表示,其中每个问题标记的嵌入由计算的权重的归一化版本进行加权。

[0051] 一旦已经生成问题的初始表示,通过向文本跨距的初始表示和问题的初始表示的拼接施加其他前馈神经网络(即施加前文所描述的FFNN),模型170生成文本跨距的问题-跨距表示。

[0052] 虽然在推断期间没有使用,在训练期间,模型170还配置为生成文本跨距的得分(比如最终得分),其衡量唯一文本跨距回答问题的良好程度。特别地,模型170可以通过由其他线性预测层处理文本跨距的问题-跨距表示来生成得分。

[0053] 文本跨距的第一数值表示122是问题-跨距表示和跨距-上下文表示的拼接并且设置为向级联的层级2的输出。

[0054] 级联的层级2包括上下文注意模型180，其对于给定的问题跨距对第一数值表示122进行操作以生成问题跨距的第二数值表示132。

[0055] 对于给定的文本跨距，模型180(i)对于输入文档中包含文本跨距的分段，基于问题中的问题标记和包含文本跨距的分段中的分段标记之间的相似度来生成问题-意识分段向量166，并且(ii)对于问题，同样基于问题中的问题标记和包含文本跨距的分段中的分段标记之间的相似度来生成针对问题的分段-意识问题向量168。

[0056] 为了生成这两个向量，模型180衡量每对问题和分段嵌入之间的相似度，即生成每个问题嵌入和每个分段嵌入之间的相应相似度得分。为了生成给定的问题嵌入q_i-分段嵌入d_j对的相似度得分η_{ij}，模型180执行以下操作：

[0057] $\eta_{ij}=\mathrm{ffnn}(q_i)^{T}\mathrm{ffnn}(d_j)$。

[0058] 为了生成对于输入文档中包含文本跨距的分段的问题-意识分段向量，模型180然后为每个分段标记确定相应的伴随向量(attended vector)，该伴随向量说明如由相似度得分反映的分段标记与问题标记的相似度，并且从分段标记的伴随向量确定问题-意识分段向量。

[0059] 为了确定针对问题的分段-意识问题向量，模型180为每个问题标记确定相应的伴随向量，该伴随向量衡量如由相似度得分反映的问题标记与分段标记的相似度，并且从问题标记的伴随向量确定分段-意识问题向量。

[0060] 特别地，为了生成问题-意识分段向量，每个原始分段嵌入向量及其对应的伴随向量被拼接且通行穿过其他前馈网络，将由该网络生成的表示求和以获得问题-意识分段向量。相似地，每个原始问题嵌入向量及其对应的伴随向量被拼接且穿过该前馈网络，将由该网络生成的表示求和以获得分段-意识问题向量。

[0061] 模型180然后使用其他前馈神经网络处理第一文本跨距的第一数值表示、问题-意识分段向量、分段-意识问题向量、以及可选地问题-跨距特征之间的拼接，以生成文本跨距的第二数值表示。

[0062] 虽然在推断期间没有使用，在训练期间，模型180还配置为生成文本跨距的得分(比如最终得分)，其衡量唯一文本跨距回答问题的良好程度。特别地，模型180可以通过由其他线性预测层处理文本跨距的第二数值表示来生成得分。

[0063] 层级3包括聚集多次提及(aggregating multiple mentions)模型190，该模型190接收候选回答跨距的第二数值表示132，并且基于第二数值表示132从整个文档中出现多次的所有候选回答跨距聚集信息。

[0064] 特别地，对于每个唯一问题跨距，模型190使用其他前馈神经网络来处理与唯一文本跨距对应的每个文本跨距的第二数值表示，以生成每个文本跨距的相应转换的数值表示。模型190然后通过对与唯一文本跨距对应的候选文本跨距的转换的数值表示进行求和来确定唯一文本跨距的聚合表示。

[0065] 模型190然后通过由其他线性预测层处理唯一文本跨距的聚合表示来生成唯一文本跨距的最终得分142。

[0066] 尽管采用模型160-190的各种前馈神经网络和各种线性投射层的架构总体上是相

同的，但是每个前馈神经网络和线性投射层总体上具有与其他神经网络或投射层不同的参数值。为了确定这些参数值，系统100在训练数据上训练级联的机器学习系统110。

[0067] 图2是训练级联的神经网络系统的示例过程200的流程图。为了方便起见，过程200将被描述为由位于一个或多个位置上的一个或多个计算机的系统执行。例如，适当编程的问答系统(如图1的问答系统100)可以执行过程200。

[0068] 系统可以对多个训练示例重复地执行过程200以重复地更新级联的神经网络系统的参数值。

[0069] 系统获得训练示例(步骤202)。训练示例包括训练问题和训练文档，并且标识来自训练文档的最佳回答问题的正确单词跨距。

[0070] 系统使用级联的神经网络系统处理训练问题和训练文档，以生成(i)与正确单词跨距对应的唯一单词跨距的最终得分，以及(ii)针对训练文档中正确单词跨距的每个提及的模型160-180中的每一个模型的得分(步骤204)。

[0071] 特别地，如上文所描述，尽管训练后只有最终得分用于选择输入问题的最佳回答，但是在训练期间，模型160-180的每一个配置为生成训练文档中每个候选单词跨距的相应得分。

[0072] 系统通过确定损失函数相对于参数的梯度来确定对级联的机器学习系统的参数的更新(步骤206)。如图1B的示例中可看出，损失函数l包括各取决于由模型160-190的对应的一个生成的得分的项l_1、l_2、l_3、和l_4。特别地，损失函数包括，对于模型160-180中的每一个而言，取决于分配到训练文档中正确单词跨距的提及的得分的相应损失项，以及对于模型190而言，取决于由模型190分配给与正确单词跨距对应的唯一单词跨距的最终得分的损失项。

[0073] 特别地，损失函数可以是在所有子模型160-190下正确回答跨距的总负对数似然。例如，损失函数可以表达为：

[0074] $$-\sum_{k=1}^{3}\lambda_k \log \sum_{\hat{\mathbf{s}}\in \mathbf{S}^*} p^{(k)}(\hat{\mathbf{s}}|\mathbf{q},\mathbf{d}) - \lambda_4 \log \sum_{\hat{\mathbf{u}}\in \mathbf{S}^*} p^{(4)}(\hat{\mathbf{u}}|\mathbf{q},\mathbf{d})$$

[0075] 其中每个λ是超参数，使得λ增加1，S*是训练文档中正确回答跨距的所有提及的集合，$p^{(k)}(s|q,d)$是由模型160-180的第k个模型在集合S*中的提及s的得分，并且$p^{(4)}(u|q,d)$是由模型190分配到唯一回答跨距的最终得分。

[0076] 系统可以使用机器学习训练技术(如反向传播)确定相对于每个参数的梯度，并且然后可以通过向梯度施加更新规则(如ADAM更新规则、rmsprop更新规则、或随机梯度下降学习率)而从梯度确定更新。

[0077] 图3是响应于问题从电子文档选择文本跨距的示例过程300的流程图。为了方便起见，过程300将被描述为由位于一个或多个位置上的一个或多个计算机的系统执行。例如，适当编程的问答系统(如图1的问答系统100)可以执行过程100。

[0078] 系统在输入文档中获得多个文本跨距中的每一个的相应第一数值表示(步骤302)。例如，系统可以使用如上文所描述的级联的机器学习系统的层级1来生成相应的第一数值表示。

[0079] 对于多个文本跨距中的每一个，系统确定相应第二数值表示(步骤304)。例如，系

统可以使用如上文所描述的级联的机器学习系统的层级2来生成相应的第二数值表示。特别地，对于多个文本跨距中的每一个，系统可以：对于输入文档中包含文本跨距的分段，基于问题中的问题标记和包含文本跨距的分段中的分段标记之间的相似度来确定问题-意识分段向量，且对于问题，同样基于问题中的问题标记和包含文本跨距的分段中的分段标记之间的相似度来确定针对问题的分段-意识问题向量，并且使用第二前馈神经网络处理文本跨距的第一数值表示、问题-意识分段向量和分段-意识问题向量以生成文本跨距的第二数值表示。

[0080] 系统对于多个文本跨距中的每个唯一文本跨距，从与唯一文本跨距对应的文本跨距的第二数值表示确定唯一文本跨距的聚合表示(步骤306)，并且从聚合表示确定唯一文本跨距的最终得分，该最终得分衡量唯一文本跨距回答问题的良好程度(步骤308)。

[0081] 系统选择具有最高最终得分的唯一文本跨距作为对问题的回答(步骤310)。

[0082] 本说明书将术语“配置”用于与系统和计算机程序组件关联。对于配置为执行特定操作或动作的一个或多个计算机的系统，意味着该系统在其上已经安装操作中使得系统执行操作或动作的软件、固件、硬件或其组合。对于配置为执行特定操作或动作的一个或多个计算机程序，意味着该一个或多个程序包括指令，该指令在由数据处理设备执行时使得该设备执行操作或动作。

[0083] 本说明书中所描述的主题的实施例和功能操作可以被实现在数字电子电路中或者在有形实施的计算机软件或固件中、计算机硬件中，包含本说明书中所公开的结构以及其结构的等同物，或者它们中的一个或多个的组合。本说明书中所描述的主题的实施例可以被实现为(多个)有形非易失性存储介质上所编码的一个或多个计算机程序，即计算机程序指令的一个或多个模块，用于由数据处理设备执行或者控制数据处理设备的操作。计算机储存介质可以是计算机可读储存装置、计算机可读储存衬底、随机或串行存取存储器阵列或装置、或者它们中一个或多个的组合。替代地或额外地，程序指令可以被编码在例如机器生成的电信号、光信号或电磁信号的人工生成的传播信号上，生成该信号来编码信息以传输到合适的接收器设备用于由数据处理设备执行。

[0084] 术语“数据处理设备”是指数据处理硬件并且涵盖处理数据的所有类型的设备、装置和机器，包括例如可编程处理器、计算机或多个处理器或计算机。设备还可以是，或还包括专用逻辑电路，例如FPGA(现场可编程门阵列)或ASIC(专用集成电路)。除了硬件，该设备可选地可以包括代码，该代码创建用于计算机程序的执行环境，例如，构成处理器固件、协议栈、数据库管理系统、操作系统、或其一个或多个的组合的代码。

[0085] 计算机程序(也称为或描述为程序、软件、软件应用、app、模块、软件模块、脚本或代码)可以以任何形式的编程语言来写入，包含编译的或解释的语言、或者声明性语言或进程语言，并且计算机程序可以部署为任何形式，包含作为单独的程序或作为模块、组件、子例程或合适于在计算环境中使用的其他单元。程序可以但不必对应与文件系统中的文件。程序可以被存储在保存其他程序或数据的部分文件中，例如在标记语言文档中存储的一个或多个脚本、在专用于讨论中的程序的单个文件中、或在多个协同文件中，例如存储一个或多个模块、子程序或部分代码的文件。计算机程序可以部署为在一个计算机上或者在多个计算机上执行，该多个计算机位于一个站点处或者分布跨越多个站点并由数据通信网络互连。

[0086] 在本说明书中，术语“数据库”宽泛地用于指代数据的任何收集：数据不需要以任何特定方式进行结构化，或完全不需要进行结构化，并且可以将数据储存在一个或多个位置中的储存装置上。因此，例如，索引数据库可以包括数据的多个收集，其中每一个可以被不同地组织和存取。

[0087] 相似地，在本说明书中，术语“引擎”宽泛地用于指代基于软件的系统、子系统、或过程，其编程为执行一个或多个具体功能。总体上，引擎将实现为在一个或多个位置中的一个或多个计算机上安装的一个或多个软件模块或组件。在一些情况下，一个或多个计算机将专用于特定引擎，在其他情况下，可以在同一个计算机或多个计算机上安装并运行多个引擎。

[0088] 可以由执行一个或多个计算机程序的一个或多个可编程计算机来进行在本说明书中所描述的过程和逻辑流，以通过在输入数据上操作并且产生输出来进行功能。还可以由例如FPGA和ASIC的专用逻辑电路，或是由专用逻辑电路以及一个或多个编程计算机来执行处理和逻辑流。

[0089] 适合于执行计算机程序的计算机可以基于通用或专用微处理器或两者，或者任何其他类型的中央处理单元。总体上，中央处理单元将从只读存储器或随机存取存储器或两者接收指令和数据。计算机的基本元件是进行或执行指令的中央处理单元以及存储指令和数据的一个或多个存储器装置。中央处理单元和存储器可以由专用逻辑电路补充，或者合并在专用逻辑电路中。通常，计算机还将包含用于存储数据的一个或多个海量存储装置(例如磁、磁光盘或光盘)，或者可操作地耦合以从海量存储装置(例如磁、磁光盘或光盘)接收数据或者将数据传输到海量存储装置(例如磁、磁光盘或光盘)，或者以上两者。但是，计算机不必具有这样的装置。此外，计算机可以被嵌入另一个设备中，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)接收器或便携式储存装置(例如通用串行总线(USB)闪速驱动器)、仅例举一些。

[0090] 适用于储存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非易失性存储器、介质和存储器装置，作为示例包含半导体存储器装置(例如EPROM、EEPROM和闪速存储装置)、磁盘(例如内部硬盘或可移除磁盘)、磁光盘以及CD ROM和DVD-ROM磁盘。

[0091] 为了提供与用户的交互，可以计算机上实现本发明的实施例，该计算机具有用于向用户显示信息的显示装置——例如，CRT(阴极射线管)或LCD(液晶显示器)监控器，以及键盘和指点器——例如鼠标或轨迹球，用户可以通过该键盘和指示器向计算机提供输入。其他类型的装置还可以用于提供与用户的交互；例如向用户所提供的反馈可以是任何形式的传感反馈，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用包括声音、语音或触觉输入的任何形式来接收。此外，计算机可以通过发送文档到由用户使用的装置或者从由用户使用的装置接收文档来与用户交互，例如在响应于从网络浏览器所接收的请求的情况下通过发送网页到用户装置的网络浏览器。另外，计算机可以通过将文本消息或其它形式的消息发送到个人装置(例如正在运行消息应用的智能手机)并且从用户接收响应消息作为回答来与用户交互。

[0092] 实现机器学习模型的数据处理设备还可以包括例如处理机器学习训练或生产(即推断)负载的共同和计算密集部分的专用硬件加速器单元。

[0093] 机器学习模型可以使用机器学习框架来实现并部署，例如TensorFlow框架、微软

认知工具包(Microsoft Cognitive Toolkit)框架、Apache Singa框架或Apache MXNet框架。

[0094] 本说明书中所描述的主体的实施例可以在计算系统中来实现，该计算系统包括例如作为数据服务器的后端组件、或者包含例如应用服务器的中间件组件、或者包含例如具有图形用户界面、web浏览器或app的客户端计算机的前端组件，用户可以通过该图形用户界面、web浏览器或app与本说明书中所描述的主题的实现方式或者一个或多个这样的后端、中间件或前端组件的任意组合交互。系统的组件可以通过任何形式或者例如通信网络的数字数据通信的介质进行互连。通信网络的示例包括局域网(LAN)和例如互联网的广域网(WAN)。

[0095] 计算系统可以包含客户端和服务器。客户端和服务器总体上彼此远离，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彼此之间具有客户端-服务器关系的计算机程序而出现。在一些实施例中，服务器将例如HTML页面的数据传输到客户端装置，例如出于将数据显示到客户端装置并接收来自与用作客户端的装置交互的用户的用户输入的目的。可以在服务器处从装置接收用户装置处生成的数据，例如用户交互的结果。

[0096] 尽管本说明书包含许多具体实现方式细节，但这些不应当解释为对任何发明的范围或者要求保护的范围的限制，而是专用于对特定发明的特定实施例的特征进行描述。在本说明书中所描述的在单独实施例的上下文中的某些特征还可以在单个实施例中组合地实现。相反地，在单个实施例的上下文中所描述的各种特征还可以分别在多个实施例中来实现或者以各种合适的子组合来实现。此外，尽管特征可以如上文描述为以某些组合起作用并且甚至最初同样地要求，但是在某些情况下来自所要求保护的组合的一个或多个特征可以从组合中去除，并且所要求保护的组合可以针对子组合或子组合的变化。

[0097] 类似地，尽管以特定顺序在附图中描绘并在权利要求中叙述操作，但这不应当理解为要求按所示的特定顺序或连续的顺序执行这样的操作或者执行所有图示的操作，以实现期望的结果。在某些环境下，多任务处理和并行处理可以是有利的。此外，如上所描述的实施例中的各种系统组件的分离不应被理解为在所有实施例中需要这样的分离，并且应当理解的是，所描述的程序组件和系统总体上可以集成在单个软件产品中或者打包到多个软件产品中。

[0098] 已经描述主题的特定实施例。其他实施例在所附权利要求的范围内。例如，权利要求中所述的行为可以以不同的顺序进行，并且仍然实现期望的结果。作为一个示例，所附附图中所描绘的步骤不是必须按所示出的特定顺序或先后顺序，以实现期望的结果。在一些情况下，多任务和并行处理可以是有利的。

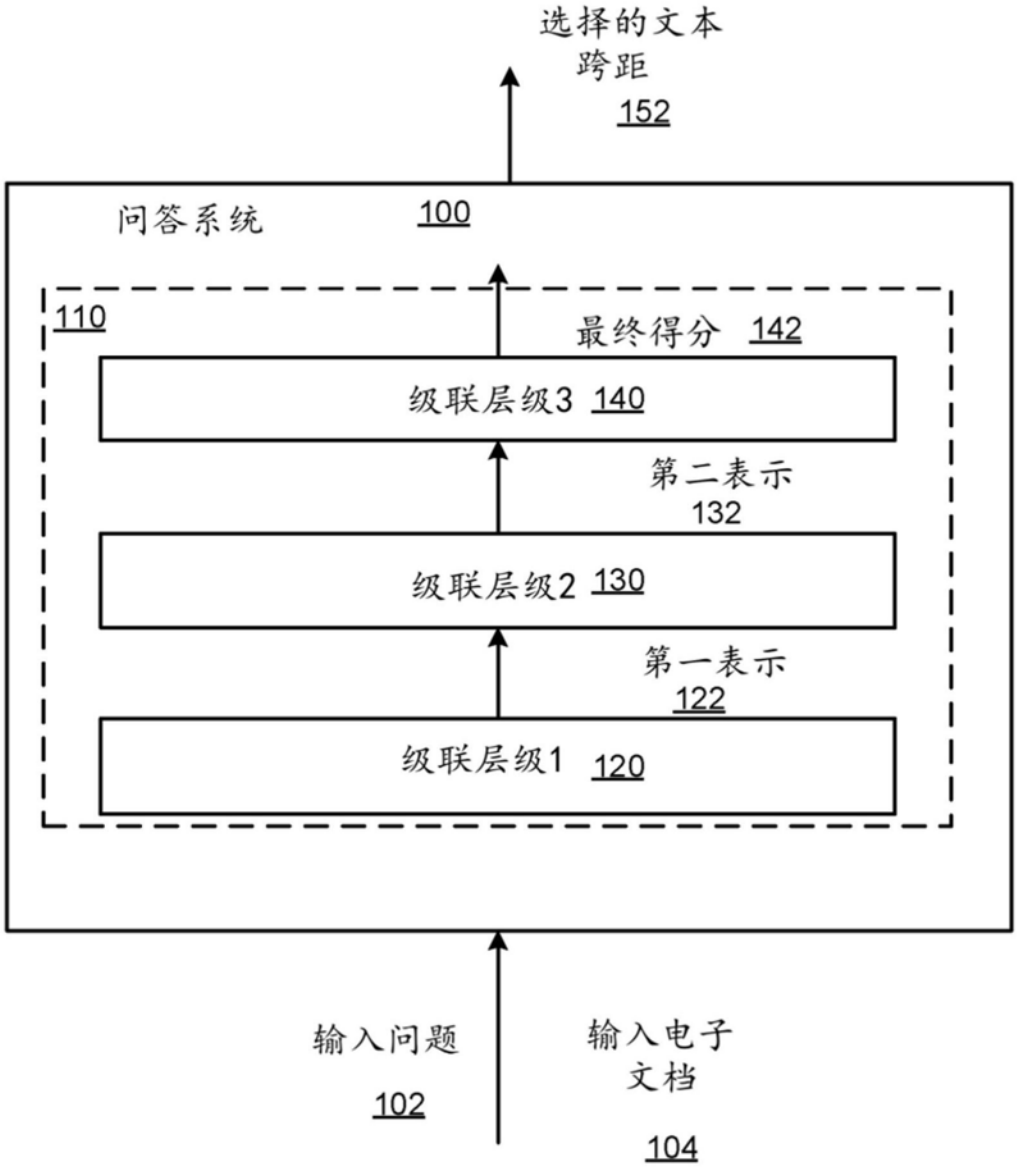

选择的文本
跨距
152
问答系统
100
110
最终得分 142
级联层级3 140
第二表示
132
级联层级2 130
第一表示
122
级联层级1 120
输入问题
102
输入电子
文档
104

图1

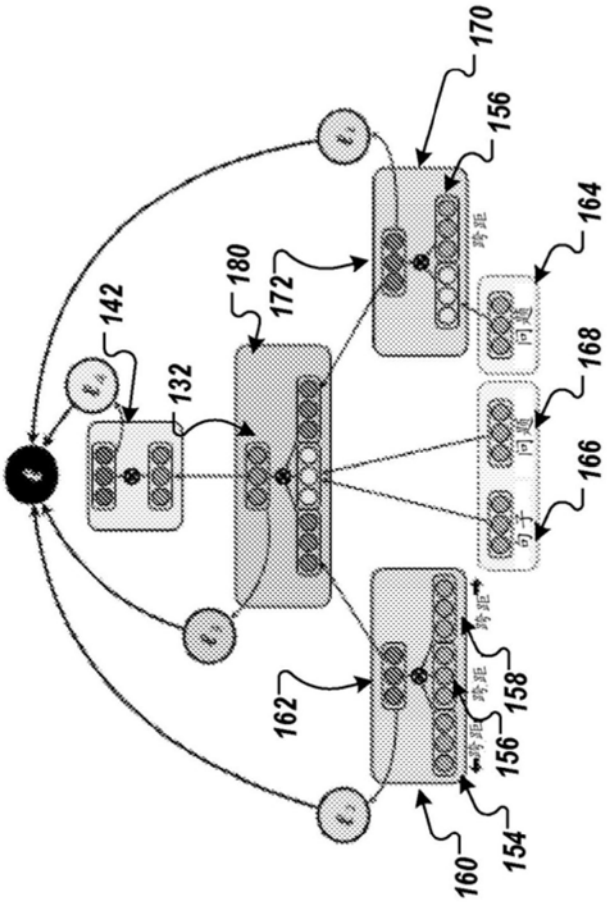


图1B

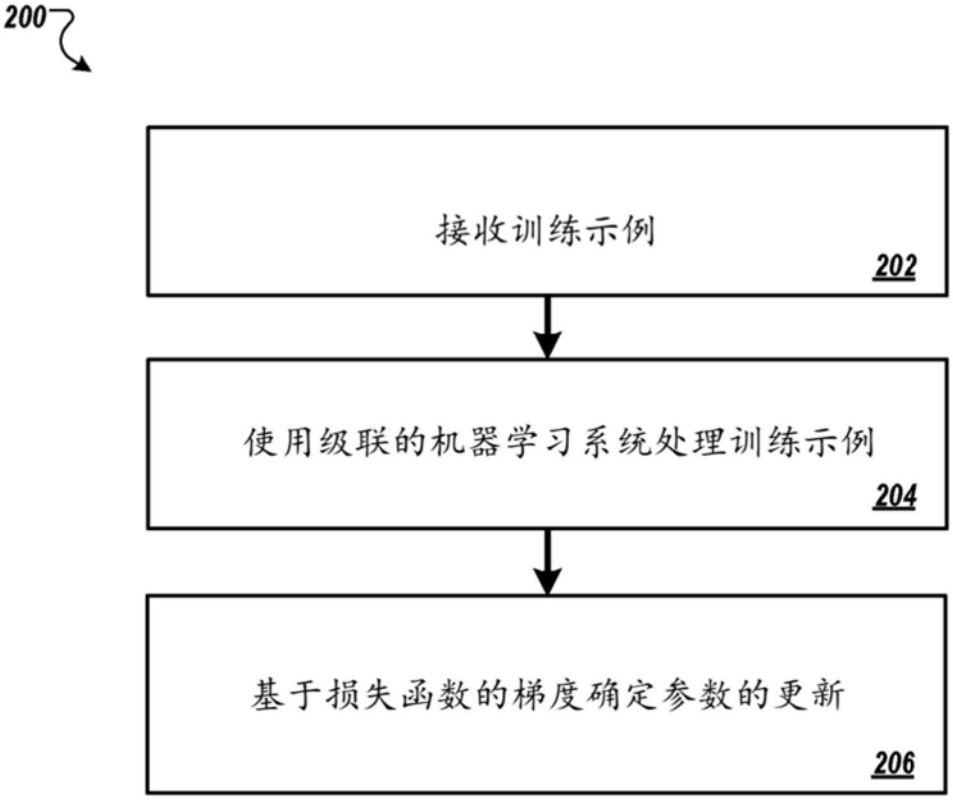

200
接收训练示例
202
使用级联的机器学习系统处理训练示例
204
基于损失函数的梯度确定参数的更新
206

图2

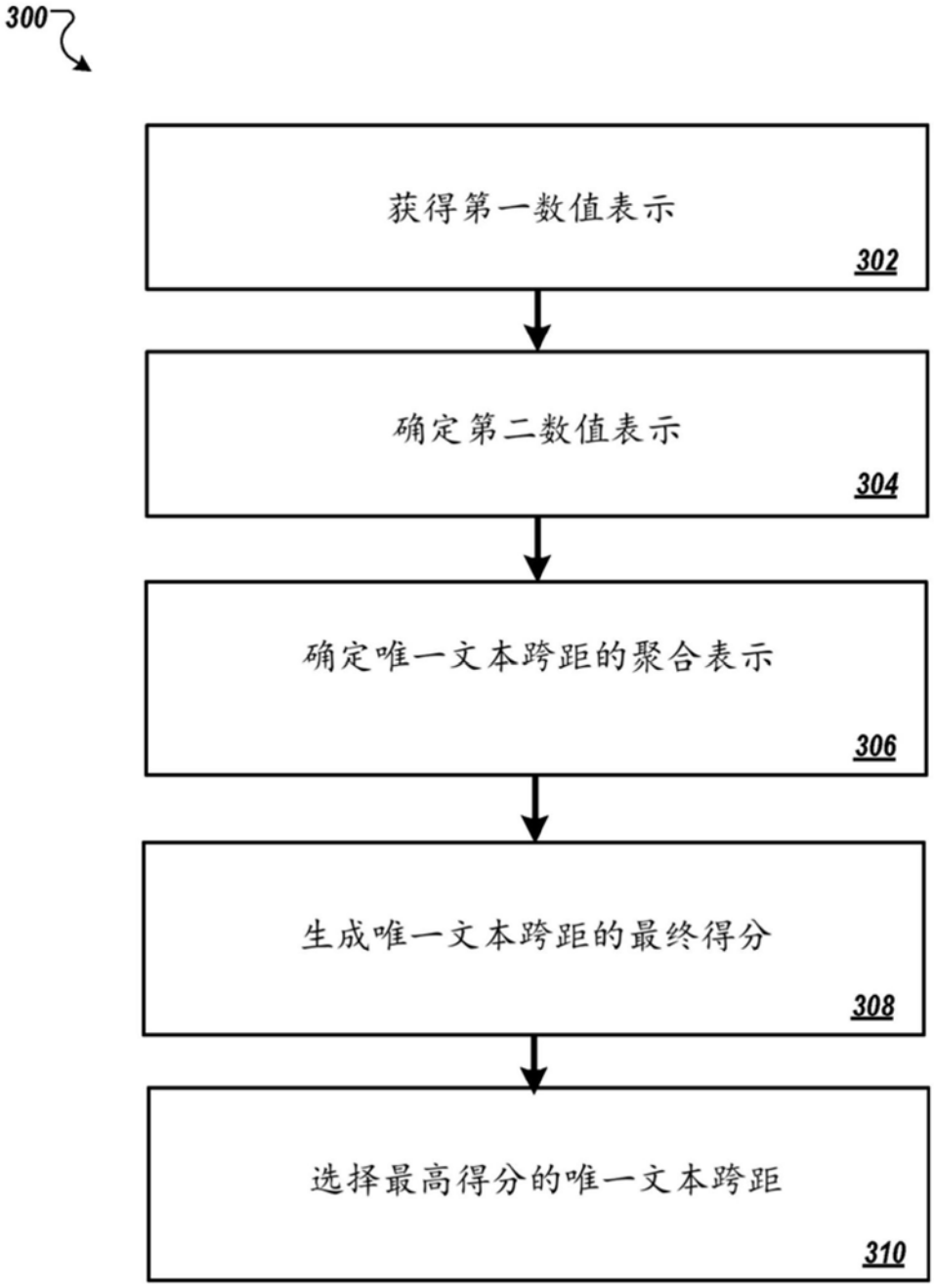

300
获得第一数值表示
302
确定第二数值表示
304
确定唯一文本跨距的聚合表示
306
生成唯一文本跨距的最终得分
308
选择最高得分的唯一文本跨距
310

图3