



(12) 发明专利申请

(10) 申请公布号 CN 115309873 A

(43) 申请公布日 2022. 11. 08

(21) 申请号 202110500967.6

(22) 申请日 2021.05.08

(71) 申请人 京东科技控股股份有限公司

地址 100176 北京市北京经济技术开发区
科创十一街18号C座2层221室

(72) 发明人 刘航 陈蒙

(74) 专利代理机构 北京清亦华知识产权代理事
务所(普通合伙) 11201

专利代理师 王萌

(51) Int. Cl.

G06F 16/332 (2019.01)

G06F 16/33 (2019.01)

G06F 40/289 (2020.01)

G06F 40/30 (2020.01)

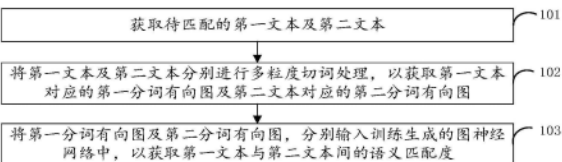
权利要求书6页 说明书22页 附图7页

(54) 发明名称

语义的匹配方法、装置、计算机设备及存储
介质

(57) 摘要

本公开提出一种语义的匹配方法、装置、计算机设备及存储介质,具体涉及人工智能等技术领域。其中,该方法包括:获取待匹配的第一文本及第二文本;将所述第一文本及所述第二文本分别进行多粒度切词处理,以获取所述第一文本对应的第一分词有向图及所述第二文本对应的第二分词有向图;将所述第一分词有向图及所述第二分词有向图,分别输入训练生成的图神经网络中,以获取所述第一文本与所述第二文本间的语义匹配度。由此,通过对包含待匹配文本的多粒度的分词有向图进行处理,以确定两个文本间的语义匹配度,从而提高了语义匹配结果的准确性和可靠性。



1. 一种语义的匹配方法,其特征在于,包括:

获取待匹配的第一文本及第二文本;

将所述第一文本及所述第二文本分别进行多粒度切词处理,以获取所述第一文本对应的第一分词有向图及所述第二文本对应的第二分词有向图;

将所述第一分词有向图及所述第二分词有向图,分别输入训练生成的图神经网络中,以获取所述第一文本与所述第二文本间的语义匹配度。

2. 如权利要求1所述的方法,其特征在于,所述将所述第一文本及所述第二文本分别进行多粒度切词处理,以获取所述第一文本对应的第一分词有向图及所述第二文本对应的第二分词有向图,包括:

将所述第一文本及所述第二文本分别进行多粒度切词处理,以获取所述第一文本对应的多个第一分词有向组及所述第二文本对应的多个第二分词有向组;

将所述多个第一分词有向组及所述多个第二分词有向组中的各分词,分别与预设的词表中各参考分词进行匹配;

在所述预设的词表中包含所述多个第一分词有向组中的各分词的情况下,基于所述多个第一分词有向组中的相同分词,将所述多个第一分词有向组进行融合,以生成所述第一分词有向图;

在所述预设的词表中包含所述多个第二分词有向组中的各分词的情况下,基于所述多个第二分词有向组中的相同分词,将所述多个第二分词有向组进行融合,以生成所述第二分词有向图。

3. 如权利要求2所述的方法,其特征在于,在所述将所述多个第一分词有向组及所述多个第二分词有向组中的各分词,分别与预设的词表中各参考分词进行匹配之后,还包括:

在所述任一第一分词有向组中的任一分词与预设的词表中各参考分词均未匹配的情况下,将所述任一分词进行切词处理,以获取所述任一分词对应的有向词片段;

在所述预设的词表中包括每个所述词片段的情况下,利用所述有向词片段替换所述任一分词,以生成新的第一分词有向组;

基于新的多个第一分词有向组中的相同分词,将所述多个第一分词有向组进行融合,以生成所述第一分词有向图。

4. 如权利要求1-3任一所述的方法,其特征在于,所述将所述第一分词有向图及所述第二分词有向图,分别输入训练生成的图神经网络中,以获取所述第一文本与所述第二文本间的匹配度,包括:

将所述第一分词有向图及所述第二分词有向图,分别输入所述图神经网络的向量表示子网络,以获取所述第一文本对应的第一向量及所述第二文本对应的第二向量;

将所述第一向量及所述第二向量分别输入所述图神经网络的融合子网络,以获取所述第一文本对应的第一语义向量及所述第二文本对应的第二语义向量;

将所述第一语义向量及所述第二语义向量分别输入所述图神经网络的匹配度计算层,以获取所述第一文本与所述第二文本间的语义匹配度。

5. 如权利要求4所述的方法,其特征在于,所述将所述第一分词有向图及所述第二分词有向图,分别输入所述图神经网络的向量表示子网络,以获取所述第一文本对应的第一向量及所述第二文本对应的第二向量,包括:

根据每个所述分词有向图中各个分词间的连接关系,确定每个所述分词对应的正向分词路径及反向分词路径,其中,任一分词对应的正向分词路径为由所述任一分词起至文本的结束符止的各分词组成的路径,任一分词对应的反向分词路径为由文本的起始符起至所述任一分词止的各分词组成的路径;

根据每个所述分词对应的正向分词路径中包含的各个第一分词分别对应的词向量,确定每个所述分词对应的正向向量;

根据每个所述分词对应的反向分词路径中包含的各个第二分词分别对应的词向量,确定每个所述分词对应的反向向量;

将每个所述分词对应的正向向量及每个所述分词对应的反向向量进行融合,以确定每个所述分词对应的聚合向量;

根据所述第一分词有向图中各个分词对应的聚合向量,确定所述第一文本对应的第一向量;

根据所述第二分词有向图中各个分词对应的聚合向量,确定所述第二文本对应的第二向量。

6.如权利要求5所述的方法,其特征在于,所述根据每个所述分词对应的正向分词路径中包含的各个第一分词分别对应的词向量,确定每个所述分词对应的正向向量,包括:

确定每个所述分词与每个所述第一分词间的相似度;

根据所述相似度,确定每个所述第一分词的权重;

根据每个所述第一分词的权重,将各个所述第一分词对应的词向量进行加权求和,以生成每个所述分词对应的正向向量。

7.如权利要求5所述的方法,其特征在于,在所述根据每个所述分词对应的正向分词路径中包含的各个第一分词分别对应的词向量,确定每个所述分词对应的正向向量之前,还包括:

将每个所述分词中的每个字符进行向量映射,以获取每个所述分词中每个字符对应的字符向量;

将每个所述分词中各字符对应的字符向量进行最大池化处理,以生成每个所述分词对应的第一子向量;

将每个所述分词中各字符对应的字符向量进行平均池化处理,以生成每个所述分词对应的第二子向量;

将每个所述分词对应的第一子向量及第二子向量进行融合,以生成每个所述分词对应的词向量。

8.如权利要求4所述的方法,其特征在于,所述将所述第一向量及所述第二向量分别输入所述图神经网络的融合子网络,以获取所述第一文本对应的第一语义向量及所述第二文本对应的第二语义向量,包括:

根据任一分词有向图中每个第三分词分别与另一分词有向图中各个分词间的相似度,确定每个所述第三分词下所述另一分词有向图中各个第四分词对应的权重;

根据所述各个第四分词对应的权重,将所述各个第四分词对应的词向量进行加权求和,确定每个所述第三分词对应的参考向量;

根据所述第一分词有向图中各个第三分词对应的参考向量,确定所述第一分词有向图

对应的第三向量；

根据所述第二分词有向图中各个第三分词对应的参考向量，确定所述第二分词有向图对应的第四向量；

将所述第一向量及所述第三向量进行融合，以生成所述第一文本对应的第一语义向量；

将所述第二向量及所述第四向量进行融合，以生成所述第二文本对应的第二语义向量。

9. 如权利要求8所述的方法，其特征在于，所述将所述第一向量及所述第三向量进行融合，以生成所述第一文本对应的第一语义向量，包括：

将所述第一向量及所述第三向量进行最大池化处理，以生成第一池化向量；

将所述第一向量及所述第三向量进行平均池化处理，以生成第二池化向量；

将所述第一池化向量与所述第二池化向量进行融合，以生成所述第一文本对应的第一语义向量。

10. 如权利要求1-3任一所述的方法，其特征在于，在所述将所述第一分词有向图及所述第二分词有向图，分别输入训练生成的图神经网络中，以获取所述第一文本与所述第二文本间的语义匹配度之前，还包括：

获取训练样本，其中，所述训练样本中包括多个文本对；

将每个所述文本对中的文本分别进行多粒度切词处理，以获取每个所述文本对对应的两个分词有向图；

将每个所述文本对对应的两个分词有向图分别输入教师模型及学生模型，以获取所述教师模型输出的所述文本对的第一语义匹配度、及所述学生模型输出的所述文本对的第二语义匹配度，其中，所述教师模型为已训练完成的模型；

根据所述第一语义匹配度与所述第二语义匹配度的差异，对所述学生模型进行修正，以生成所述图神经网络模型。

11. 一种语义的匹配装置，其特征在于，包括：

第一获取模块，用于获取待匹配的第一文本及第二文本；

处理模块，用于将所述第一文本及所述第二文本分别进行多粒度切词处理，以获取所述第一文本对应的第一分词有向图及所述第二文本对应的第二分词有向图；

第二获取模块，用于将所述第一分词有向图及所述第二分词有向图，分别输入训练生成的图神经网络中，以获取所述第一文本与所述第二文本间的语义匹配度。

12. 如权利要求11所述的装置，其特征在于，所述处理模块，具体用于：

将所述第一文本及所述第二文本分别进行多粒度切词处理，以获取所述第一文本对应的多个第一分词有向组及所述第二文本对应的多个第二分词有向组；

将所述多个第一分词有向组及所述多个第二分词有向组中的各分词，分别与预设的词表中各参考分词进行匹配；

在所述预设的词表中包含所述多个第一分词有向组中的各分词的情况下，基于所述多个第一分词有向组中的相同分词，将所述多个第一分词有向组进行融合，以生成所述第一分词有向图；

在所述预设的词表中包含所述多个第二分词有向组中的各分词的情况下，基于所述多

个第二分词有向组中的相同分词,将所述多个第二分词有向组进行融合,以生成所述第二分词有向图。

13. 如权利要求12所述的装置,其特征在于,所述处理模块,还具体用于:

在所述任一第一分词有向组中的任一分词与预设的词表中各参考分词均未匹配的情况下,将所述任一分词进行切词处理,以获取所述任一分词对应的有向词片段;

在所述预设的词表中包括每个所述词片段的情况下,利用所述有向词片段替换所述任一分词,以生成新的第一分词有向组;

基于新的多个第一分词有向组中的相同分词,将所述多个第一分词有向组进行融合,以生成所述第一分词有向图。

14. 如权利要求11-13任一所述的装置,其特征在于,所述第二获取模块,包括:

第一获取单元,用于将所述第一分词有向图及所述第二分词有向图,分别输入所述图神经网络的向量表示子网络,以获取所述第一文本对应的第一向量及所述第二文本对应的第二向量;

第二获取单元,用于将所述第一向量及所述第二向量分别输入所述图神经网络的融合子网络,以获取所述第一文本对应的第一语义向量及所述第二文本对应的第二语义向量;

第三获取单元,用于将所述第一语义向量及所述第二语义向量分别输入所述图神经网络的匹配度计算层,以获取所述第一文本与所述第二文本间的语义匹配度。

15. 如权利要求14所述的装置,其特征在于,所述第一获取单元,包括:

第一确定子单元,用于根据每个所述分词有向图中各个分词间的连接关系,确定每个所述分词对应的正向分词路径及反向分词路径,其中,任一分词对应的正向分词路径为由所述任一分词起至文本的结束符止的各分词组成的路径,任一分词对应的反向分词路径为由文本的起始符起至所述任一分词止的各分词组成的路径;

第二确定子单元,用于根据每个所述分词对应的正向分词路径中包含的各个第一分词分别对应的词向量,确定每个所述分词对应的正向向量;

第三确定子单元,用于根据每个所述分词对应的反向分词路径中包含的各个第二分词分别对应的词向量,确定每个所述分词对应的反向向量;

融合子单元,用于将每个所述分词对应的正向向量及每个所述分词对应的反向向量进行融合,以确定每个所述分词对应的聚合向量;

第四确定子单元,用于根据所述第一分词有向图中各个分词对应的聚合向量,确定所述第一文本对应的第一向量;

第五确定子单元,用于根据所述第二分词有向图中各个分词对应的聚合向量,确定所述第二文本对应的第二向量。

16. 如权利要求15所述的装置,其特征在于,所述第二确定子单元,具体用于:

确定每个所述分词与每个所述第一分词间的相似度;

根据所述相似度,确定每个所述第一分词的权重;

根据每个所述第一分词的权重,将各个所述第一分词对应的词向量进行加权求和,以生成每个所述分词对应的正向向量。

17. 如权利要求15所述的装置,其特征在于,所述第一获取单元,还包括:

第一获取子单元,用于将每个所述分词中的每个字符进行向量映射,以获取每个所述

分词中每个字符对应的字符向量；

第一生成子单元,用于将每个所述分词中各字符对应的字符向量进行最大池化处理,以生成每个所述分词对应的第一子向量；

第二生成子单元,用于将每个所述分词中各字符对应的字符向量进行平均池化处理,以生成每个所述分词对应的第二子向量；

第三生成子单元,用于将每个所述分词对应的第一子向量及第二子向量进行融合,以生成每个所述分词对应的词向量。

18.如权利要求14所述的装置,其特征在于,所述第二获取单元,包括:

第六确定子单元,用于根据任一分词有向图中每个第三分词分别与另一分词有向图中各个分词间的相似度,确定每个所述第三分词下所述另一分词有向图中各个第四分词对应的权重；

第七确定子单元,用于根据所述各个第四分词对应的权重,将所述各个第四分词对应的词向量进行加权求和,确定每个所述第三分词对应的参考向量；

第八确定子单元,用于根据所述第一分词有向图中各个第三分词对应的参考向量,确定所述第一分词有向图对应的第三向量；

第九确定子单元,用于根据所述第二分词有向图中各个第三分词对应的参考向量,确定所述第二分词有向图对应的第四向量；

第三生成子单元,用于将所述第一向量及所述第三向量进行融合,以生成所述第一文本对应的第一语义向量；

第四生成子单元,用于将所述第二向量及所述第四向量进行融合,以生成所述第二文本对应的第二语义向量。

19.如权利要求18所述的装置,其特征在于,所述第三生成子单元,具体用于:

将所述第一向量及所述第三向量进行最大池化处理,以生成第一池化向量；

将所述第一向量及所述第三向量进行平均池化处理,以生成第二池化向量；

将所述第一池化向量与所述第二池化向量进行融合,以生成所述第一文本对应的第一语义向量。

20.如权利要求11-13任一所述的装置,其特征在于,

所述第一获取模块,还用于获取训练样本,其中,所述训练样本中包括多个文本对；

所述处理模块,还用于将每个所述文本对中的文本分别进行多粒度切词处理,以获取每个所述文本对对应的两个分词有向图；

所述第二获取模块,还用于将每个所述文本对对应的两个分词有向图分别输入教师模型及学生模型,以获取所述教师模型输出的所述文本对的第一语义匹配度、及所述学生模型输出的所述文本对的第二语义匹配度,其中,所述教师模型为已训练完成的模型；

所述第二获取模块,还用于根据所述第一语义匹配度与所述第二语义匹配度的差异,对所述学生模型进行修正,以生成所述图神经网络模型。

21.一种计算机设备,其特征在于,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时,实现如权利要求1-10中任一所述的语义的匹配方法。

22.一种计算机可读存储介质,存储有计算机程序,其特征在于,所述计算机程序被处

理器执行时实现如权利要求1-10中任一所述的语义的匹配方法。

23.一种计算机程序产品,其特征在于,包括计算机程序,所述计算机程序在被处理器执行时实现如权利要求1-10中任一项所述的语义的匹配方法。

语义的匹配方法、装置、计算机设备及存储介质

技术领域

[0001] 本公开涉及人工智能技术领域,具体涉及一种语义的匹配方法、装置、计算机设备及存储介质。

背景技术

[0002] 随着计算机技术的蓬勃发展,在许多场景中都应用到了文本匹配,比如搜索引擎、智能问答、知识检索、对话系统等等。

[0003] 目前常用的文本匹配方法可以分为基于文本特征的匹配和基于语义特征的匹配。基于文本特征的匹配方法可以直接利用文本的词法信息,该方法简单直接,但没有理解用户问题真实的语义和意图,匹配的准确率较低。

[0004] 后来随着深度学习的发展,基于语义特征的匹配方法得到了迅速发展,如使用语义匹配模型计算文本之间的语义相似度作为匹配分数。语义匹配模型通常采用词粒度或者字粒度的信息进行处理,词粒度的信息容易受分词标准的影响,不同的分词标准可能导致匹配分数有所差异,从而导致匹配的可信度较低。

发明内容

[0005] 本公开旨在至少在一定程度上解决相关技术中的技术问题之一。

[0006] 本公开第一方面实施例提出了一种语义的匹配方法,包括:

[0007] 获取待匹配的第一文本及第二文本;

[0008] 将所述第一文本及所述第二文本分别进行多粒度切词处理,以获取所述第一文本对应的第一分词有向图及所述第二文本对应的第二分词有向图;

[0009] 将所述第一分词有向图及所述第二分词有向图,分别输入训练生成的图神经网络中,以获取所述第一文本与所述第二文本间的语义匹配度。

[0010] 本公开第二方面实施例提出了一种语义的匹配装置,包括:

[0011] 第一获取模块,用于获取待匹配的第一文本及第二文本;

[0012] 处理模块,用于将所述第一文本及所述第二文本分别进行多粒度切词处理,以获取所述第一文本对应的第一分词有向图及所述第二文本对应的第二分词有向图;

[0013] 第二获取模块,用于将所述第一分词有向图及所述第二分词有向图,分别输入训练生成的图神经网络中,以获取所述第一文本与所述第二文本间的语义匹配度。

[0014] 本公开第三方面实施例提出了一种计算机设备,包括:存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,所述处理器执行所述程序时,实现如本公开第一方面实施例提出的语义的匹配方法。

[0015] 本公开第四方面实施例提出了一种非临时性计算机可读存储介质,存储有计算机程序,所述计算机程序被处理器执行时实现如本公开第一方面实施例提出的语义的匹配方法。

[0016] 本公开第五方面实施例提出了一种计算机程序产品,当所述计算机程序产品中的

指令处理器执行时,执行本公开第一方面实施例提出的语义的匹配方法。

[0017] 本公开的提供的语义的匹配方法、装置、计算机设备及存储介质,可以先获取待匹配的第一文本及第二文本,之后将第一文本及第二文本分别进行多粒度切词处理,以获取第一文本对应的第一分词有向图及第二文本对应的第二分词有向图,之后再将第一分词有向图及第二分词有向图,分别输入训练生成的图神经网络中,以获取第一文本与第二文本间的语义匹配度。由此,通过对包含待匹配文本的多粒度的分词有向图进行处理,以确定两个文本间的语义匹配度,从而提高了语义匹配结果的准确性和可靠性。

[0018] 本公开附加的方面和优点将在下面的描述中部分给出,部分将从下面的描述中变得明显,或通过本公开的实践了解到。

附图说明

[0019] 本公开上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解,其中:

[0020] 图1A为本公开一实施例所提供的语义的匹配方法的流程示意图;

[0021] 图1B为本公开一实施例所提供的一种分词有向图的示意图;

[0022] 图2A为本公开另一实施例所提供的语义的匹配方法的流程示意图;

[0023] 图2B为本公开一实施例所提供的另一种分词有向图的示意图;

[0024] 图3A为本公开另一实施例所提供的语义的匹配方法的流程示意图;

[0025] 图3B为本公开一实施例所提供的一种确定语义的匹配度方法的示意图;

[0026] 图4A为本公开另一实施例所提供的语义的匹配方法的流程示意图;

[0027] 图4B为本公开一实施例所提供的另一种分词有向图的示意图;

[0028] 图5A为本公开另一实施例所提供的语义的匹配方法的流程示意图;

[0029] 图5B为本公开一实施例所提供的一种确定语义的匹配度方法的示意图;

[0030] 图6为本公开另一实施例所提供的语义的匹配方法的流程示意图;

[0031] 图7为本公开另一实施例所提供的语义的匹配装置的结构示意图;

[0032] 图8示出了适于用来实现本公开实施方式的示例性计算机设备的框图。

具体实施方式

[0033] 下面详细描述本公开的实施例,所述实施例的示例在附图中示出,其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的,旨在用于解释本公开,而不能理解为对本公开的限制。

[0034] 下面参考附图描述本公开实施例的语义的匹配方法、装置、计算机设备和存储介质。

[0035] 图1A为本公开实施例所提供的语义的匹配方法的流程示意图。

[0036] 本公开实施例以该语义的匹配方法被配置于语义的匹配装置中来举例说明,该语义的匹配装置可以应用于任一计算机设备中,以使该计算机设备可以执行语义的匹配方法功能。

[0037] 其中,计算机设备可以为个人电脑(Personal Computer,简称PC)、云端设备、移动设备等,移动设备例如可以为手机、平板电脑、个人数字助理、穿戴式设备、车载设备等具有

各种操作系统、触摸屏和/或显示屏的硬件设备。

[0038] 如图1A所示,该语义的匹配方法可以包括以下步骤:

[0039] 步骤101,获取待匹配的第一文本及第二文本。

[0040] 其中,第一文本和第二文本,可以为包含任意内容的文本,比如,可以分别为问题语句和应答语句,输入语句和应答语句,搜索文本和搜索结果等等任意形式,本公开对此不做限定。

[0041] 需要说明的是,本公开提供的语义的匹配方法,可以适用于任意进行文本匹配的场景中,比如搜索引擎、智能问答、知识检索、对话系统等等任意场景,本公开对此不做限定。

[0042] 步骤102,将第一文本及第二文本分别进行多粒度切词处理,以获取第一文本对应的第一分词有向图及第二文本对应的第二分词有向图。

[0043] 其中,多粒度切词处理,指的是可以按照不同的切词规则对文本进行切词处理,从而可以得到多个分词结果,多个分词结果可能相同,或者也可能不同,本公开对此不做限定。

[0044] 比如说,文本为“我是一个学生”,按照不同的切词规则对其进行多粒度切词,得到的多个分词结果可以为:我/是/一/个/学/生、我/是/一个/学/生、我/是/一/个/学生、我/是/一个/学生等等,本公开对此不做限定。

[0045] 可以理解的是,对第一文本进行一次切词处理,得到的一组分词中,各个分词之间具有一定的关联,这样的一组有关联的分词即为一个第一分词有向组。使用不同的规则,对第一文本进行多粒度切词处理,可以得到多个第一分词有向组,其可以是不同的,或者也可以为部分相同的等等。相应的,使用不同的分词规则,对第二文本进行多粒度切词处理,可以得到多个第二分词有向组,其可以是完全不同的,或者也可以为部分相同等等,本公开对此不做限定。

[0046] 第一分词有向图,可以由多个第一分词有向组构成,各个第一分词有向组中的各个分词之间,具有一定的关系,可以使用箭头等等来表示各个分词之间的关系等,本公开对此不做限定。

[0047] 可以理解的是,可以将多个第一分词有向组先进行融合,之后再生成第一分词有向图。

[0048] 举例来说,对第一文本进行多粒度切词处理,得到的第一分词有向组分别为:A1/A2/A3/A4、A1A2/A3/A4、A1/A2A3/A4。之后可以将A1/A2/A3/A4、A1A2/A3/A4、A1/A2A3/A4根据各个分词及其关系进行融合,构成第一分词有向图,其示意图可以如图1B所示。

[0049] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中各分词有向组、各分词有向图等限定。

[0050] 相应的,第二分词有向图,可以由多个第二分词有向组构成,可以使用箭头等表示其中各个分词之间的关系。可以理解的是,也可以先将多个第二分词有向组先进行融合,之后再生成第二分词有向图,此处不再赘述。

[0051] 本公开实施例中,通过使用多粒度切词处理,可以尽可能多的获取到第一文本、第二文本的所有分词结果,从而不仅可以降低单粒度切词可能带来的错误影响,而且也可以更加全面准确的了解用户的真实意图,从而为后续处理提供了条件。

[0052] 步骤103,将第一分词有向图及第二分词有向图,分别输入训练生成的图神经网络中,以获取第一文本与第二文本间的语义匹配度。

[0053] 其中,图神经网络,可以为已经提前训练好的网络模型,将第一分词有向图及第二分词有向图输入到该图神经网络中,经过图神经网络中各网络层的逐层处理,可以输出第一分词有向图与第二分词有向图之间的匹配度,也即第一文本与第二文本之间的语义匹配度。

[0054] 本公开实施例中,第一分词有向图、第二分词有向图是根据多粒度切词处理得到的分词结果所构成的,其可以更加全面、准确的包含第一文本及第二文本的信息,从而使用该第一分词有向图、第二分词有向图所确定出的语义匹配度,准确性、可靠性更高。

[0055] 本公开实施例,可以先获取待匹配的第一文本及第二文本,之后将第一文本及第二文本分别进行多粒度切词处理,以获取第一文本对应的第一分词有向图及第二文本对应的第二分词有向图,之后再分别输入训练生成的图神经网络中,以获取第一文本与第二文本间的语义匹配度。由此,通过对包含待匹配文本的多粒度的分词有向图进行处理,以确定两个文本间的语义匹配度,从而提高了语义匹配结果的准确性和可靠性。

[0056] 上述实施例,通过将获取到的第一文本、第二文本分别进行多粒度切词处理,得到各自对应的第一分词有向图、第二分词有向图,之后将其分别输入训练生成的图神经网络中,即可确定出第一文本、第二文本之间的语义匹配度。在一种可能的实现方式中,可以将多个第一分词有向组及多个第二分词有向组中的各分词分别与预设的词表中各参考分词进行匹配,根据匹配结果生成第一分词有向图、第二分词有向图,下面结合图2A对上述过程进行详细说明。

[0057] 图2A为本公开实施例所提供的语义的匹配方法的流程示意图,如图2A所示,该语义的匹配方法可以包括以下步骤:

[0058] 步骤201,获取待匹配的第一文本及第二文本。

[0059] 步骤202,将第一文本及第二文本分别进行多粒度切词处理,以获取第一文本对应的多个第一分词有向组及第二文本对应的多个第二分词有向组。

[0060] 步骤203,将多个第一分词有向组及多个第二分词有向组中的各分词,分别与预设的词表中各参考分词进行匹配。

[0061] 其中,预设的词表,可以为对历史数据进行处理后,得到的包含各个粒度分词的词表。

[0062] 可选的,本公开实施例中,还可以根据需要,为不同的应用场景,预先设置符合该场景的预设的词表。比如说,可以为电商领域、银行领域等任意应用场景设置所对应的词表,之后,在将第一文本及第二进行匹配时,即可根据该文本所在的应用场景,选取对应的预设的词表。本公开对此不做限定。

[0063] 需要说明的是,本公开的技术方案中,所涉及的历史数据的获取、存储和应用等,均符合相关法律法规的规定,且不违背公序良俗。

[0064] 另外,同一分词,在不同的词表中,可能有相同的分词组合结果,或者也可能有不同的分词组合结果等等,本公开对此不做限定。

[0065] 可以理解的是,可以将第一分词有向组中的各个分词,与预设的词表中各个参考

分词进行匹配。

[0066] 比如说,第一分词有向组1中包含的各个分词为:A1/A2A3,将第一分词有向组1中的各个分词依次与预设的词表中各参考分词进行匹配,预设的词表中与第一分词有向组1相关的参考分词的组合有:A1/A2/A3、A1/A2A3、A1A2/A3等,可以看出预设的词表中包含第一分词有向组1中的各个分词。

[0067] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中,第一分词有向组、预设的词表等的限定。

[0068] 步骤204,在预设的词表中包含多个第一分词有向组中的各分词的情况下,基于多个第一分词有向组中的相同分词,将多个第一分词有向组进行融合,以生成第一分词有向图。

[0069] 步骤205,在预设的词表中包含多个第二分词有向组中的各分词的情况下,基于多个第二分词有向组中的相同分词,将多个第二分词有向组进行融合,以生成第二分词有向图。

[0070] 举例来说,各第一分词有向组分别为:A1/A2/A3、A1/A2A3、A1A2/A3,预设的词表中包含各第一分词有向组中的各个分词:A1/A2/A3、A1/A2A3、A1A2/A3,之后可以根据各第一分词有向组中的相同分词,将其进行融合,生成第一分词有向图。各第二分词有向组分别为:B1/B2/B3/B4、B1/B2B3/B4,预设的词表中包含各第二分词有向组中的各个分词:B1/B2/B3/B4、B1/B2B3/B4,之后可以根据各第二分词有向组中的相同分词,将其进行融合,生成第二分词有向图。生成的第一分词有向图、第二分词有向图可以如图2B所示。

[0071] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中第一分词有向组中各分词、预设的词表等的限定。

[0072] 本公开实施例中,通过基于第一分词有向组中的相同分词、第二分词有向组中的相同分词,可以分别将第一分词有向组、第二分词有向组进行融合,从而可以生成更加准确、可靠的第一分词有向图、第二分词有向图,并为后续处理提供了条件。

[0073] 步骤206,在任一第一分词有向组中的任一分词与预设的词表中各参考分词均未匹配的情况下,将任一分词进行切词处理,以获取任一分词对应的有向词片段。

[0074] 其中,根据多粒度切词处理得到的第一分词有向组,可能有多种情况,从而在预设的词表中可能不包含任一第一分词有向组中的任一分词。此时,为了得到准确的分词结果,提高语义匹配的准确度,可以将该任一分词继续进行切词处理,以便得到该任一分词对应的更短有向词片段。

[0075] 步骤207,在预设的词表中包括每个词片段的情况下,利用有向词片段替换任一分词,以生成新的第一分词有向组。

[0076] 举例来说,第一分词有向组为:A1/A2A3,其中任一分词A2A3,与预设的词表中的各参考分词均不匹配,可以将A2A3继续进行切词处理,从而获取到与其对应的有向词片段:A2/A3。之后可以使用A2/A3替换第一分词有向组中的A2A3,新的第一分词有向组即为:A1/A2/A3,之后可以使用新的第一分词有向组进行融合。

[0077] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中任一分词、预设的词表、有向词片段等的限定。

[0078] 可以理解的是,预设的词表中可以包含字符粒度的各个分词,若预设的词表中仍

未包括某个有向词片段的情况下,可以继续对该有向词片段进行切词处理,直到预设的词表中包含各个有向词片段。

[0079] 步骤208,基于新的多个第一分词有向组中的相同分词,将多个第一分词有向组进行融合,以生成第一分词有向图。

[0080] 其中,将多个第一分词有向组进行融合,以生成第一分词有向图的过程可以参照本公开其它各实施例,此处不再赘述。

[0081] 本公开实施例中,通过利用有向词片段替换未匹配的任一分词,可以生成新的第一分词有向组,并通过将多个第一分词有向组进行融合,可以得到第一分词有向图,该第一分词有向图可以更加全面、准确的包含第一文本的信息,从而该第一分词有向图的准确性和可靠性更高。

[0082] 需要说明的是,若第二分词有向组中的任一分词与预设的词表中各参考分词均未匹配时,可以参照上述第一分词有向组中的任一分词与预设的词表中各参考分词均未匹配的实现过程,此处不再赘述。

[0083] 步骤209,将第一分词有向图及第二分词有向图,分别输入训练生成的图神经网络中,以获取第一文本与第二文本间的语义匹配度。

[0084] 本公开实施例,先将获取到的第一文本及第二文本分别进行多粒度切词处理,可以得到对应的多个第一分词有向组及多个第二分词有向组,之后可以将各个分词有向组中的各分词,分别与预设的词表中各参考分词进行匹配、融合,以生成第一分词有向图、第二分词有向图。还可以在任一第一分词有向组中的任一分词与预设的词表中各参考分词均未匹配的情况下,将任一分词进行切词处理,并利用得到的有向词片段替换该任一分词,以生成新的第一分词有向组、第一分词有向图,之后再利用图神经网络可以获取第一文本与第二文本间的语义匹配度。由此,通过对待匹配的文本进行多粒度切词,并与预设的词表进行匹配,生成包含待匹配文本的多粒度的分词有向图,之后再对分词有向图进行处理,以确定两个文本间的语义匹配度,从而提高了语义匹配结果的准确性和可靠性。

[0085] 上述实施例,通过将第一文本、第二文本进行多粒度切词处理,之后将多个第一分词有向组及多个第二分词有向组中的各分词分别与预设的词表中各参考分词进行匹配,根据匹配结果生成第一分词有向图、第二分词有向图,之后再利用图神经网络,即可确定出第一文本与第二文本间的语义匹配度。在一种可能的实现方式中,将第一分词有向图及第二分词有向图,分别输入训练生成的图神经网络中,确定第一文本与第二文本间的语义匹配度时,还可以先确定出第一文本、第二文本分别对应的第一向量、第二向量,之后再继续进行处理,以确定出第一文本与第二文本间的语义匹配度,下面结合图3对上述过程进行进一步说明。

[0086] 图3A为本公开实施例所提供的语义的匹配方法的流程示意图,如图3所示,该语义的匹配方法可以包括以下步骤:

[0087] 步骤301,获取待匹配的第一文本及第二文本。

[0088] 步骤302,将第一文本及第二文本分别进行多粒度切词处理,以获取第一文本对应的第一分词有向图及第二文本对应的第二分词有向图。

[0089] 步骤303,将第一分词有向图及第二分词有向图,分别输入图神经网络的向量表示子网络,以获取第一文本对应的第一向量及第二文本对应的第二向量。

[0090] 其中,向量表示子网络,可以为图神经网络中的一个子网络,其中可以包括一个网络层,或者也可以包括多个网络层,本公开对此不做限定。

[0091] 另外,图神经网络,为已经提前训练好的网络模型,相应的,其中的向量表示子网络也为已提前训练好的子网络。

[0092] 可以理解的是,将第一分词有向图及第二分词有向图输入到该向量表示子网络中,可以得到第一分词有向图对应的第一向量、第二分词有向图对应的第二向量,也即第一文本对应的第一向量、第二文本对应的第二向量。

[0093] 步骤304,将第一向量及第二向量分别输入图神经网络的融合子网络,以获取第一文本对应的第一语义向量及第二文本对应的第二语义向量。

[0094] 其中,融合子网络,也可以为图神经网络中的一个子网络,其中可以包括一个网络层,或者也可以包括多个网络层,本公开对此不做限定。

[0095] 另外,图神经网络,为已经提前训练好的网络模型,相应的,其中的融合子网络也为已提前训练好的子网络。

[0096] 可以理解的是,将第一向量及第二向量输入到该融合子网络中,可以得到对应的第一语义向量、第二语义向量,也即第一文本对应的第一语义向量、第二文本对应的第二语义向量。

[0097] 步骤305,将第一语义向量及第二语义向量分别输入图神经网络的匹配度计算层,以获取第一文本与第二文本间的语义匹配度。

[0098] 其中,匹配度计算层,可以为图神经网络中的一个网络层,可以用来对输入其中的第一语义向量及第二语义向量进行处理,从而可以得到二者之间的语义匹配度,也即第一文本与第二文本间的语义匹配度。

[0099] 本公开实施例中,通过使用训练好的图神经网络中,可以快速、准确的确定出两个文本之间的语义匹配度,不仅提高了处理的效率,而且提高了准确性。

[0100] 下面以图3B所示的示意图为例,对本公开提供的语义的匹配方法中,确定语义匹配度的方法进行说明。

[0101] 如图3B所示,可以将获取到的待匹配的第一文本及第二文本,分别进行多粒度切词处理,从而可以确定给出分别对应的第一分词有向图及第二分词有向图。之后可以将第一分词有向图及第二分词有向图分别输入图神经网络中的向量表示子网络,从而可以输出对应的第一向量及第二向量。之后再将二者分别输入图神经网络中的融合子网络,以确定出分别对应的第一语义向量及第二语义向量,之后将两个语义向量输入匹配度计算层,即可输出第一文本与第二文本间的语义匹配度。由此,可以快速、准确的确定出两个文本之间的语义匹配度,不仅提高了处理的效率,而且提高了准确性。

[0102] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中语义匹配方法的方式等的限定。

[0103] 本公开实施例,首先将获取到的第一文本及第二文本分别进行多粒度切词处理,可以获取到对应的第一分词有向图及第二分词有向图,之后再将第一分词有向图及第二分词有向图,分别输入图神经网络的向量表示子网络,以获取第一文本对应的第一向量及第二文本对应的第二向量,再将第一向量及第二向量分别输入图神经网络的融合子网络,以获取第一文本对应的第一语义向量及第二文本对应的第二语义向量,之后将第一语义向量

及第二语义向量分别输入图神经网络的匹配度计算层,以获取第一文本与第二文本间的语义匹配度。由此,通过利用第一分词有向图、第二分词有向图以及图神经网络中的各个子网络,可以快速、准确的确定出待匹配文本间的语义匹配度,有效提高了处理效率和准确性,进一步提高了系统的性能。

[0104] 上述实施例中,通过将第一分词有向图及第二分词有向图,分别输入图神经网络的向量表示子网络,以获取第一文本对应的第一向量及第二文本对应的第二向量。在一种可能的实现方式中,还可以先确定出各个分词分别对应的正向向量、反向向量,之后再对其进行融合得到每个分词对应的聚合向量,从而确定出第一文本、第二文本分别对应的第一向量、第二向量,下面结合图4A对上述过程进行进一步说明。

[0105] 图4A为本公开实施例所提供的语义的匹配方法的流程示意图,如图4A所示,该语义的匹配方法可以包括以下步骤:

[0106] 步骤401,获取待匹配的第一文本及第二文本。

[0107] 步骤402,将第一文本及第二文本分别进行多粒度切词处理,以获取第一文本对应的第一分词有向图及第二文本对应的第二分词有向图。

[0108] 步骤403,根据每个分词有向图中各个分词间的连接关系,确定每个分词对应的正向分词路径及反向分词路径。

[0109] 其中,任一分词对应的正向分词路径为由任一分词起至文本的结束符止的各分词组成的路径,任一分词对应的反向分词路径为由文本的起始符起至任一分词止的各分词组成的路径。

[0110] 另外,各个分词有向图中,各个分词间的连接关系不同,任一分词可能有一条正向分词路径,或者也可能有多条正向分词路径,或者也可能不存在正向分词路径,任一分词,可能有一条或多条反向分词路径,或者也可能不存在反向路径等等,本公开对此不做限定。

[0111] 举例来说,在如图4B所示的第一分词有向图中,各个第一分词有向组分别为:A1/A2/A3、A1/A2A3、A1A2/A3。其中,分词A1对应的正向分词路径为:A1→A2→A3,A1→A2A3,A1不存在对应的反向分词路径。分词A2对应的正向分词路径为:A2→A3,分词A2对应的反向分词路径为:A1→A2。分词A3无对应的正向分词路径,其对应的反向分词路径为:A1→A2→A3。分词A1A2对应的正向分词路径为A3,不存在反向分词路径;分词A2A3不存在正向分词路径,其对应的反向分词路径为:A1→A2A3。

[0112] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中各分词对应的正向路径、反向路径等的限定。

[0113] 步骤404,根据每个分词对应的正向分词路径中包含的各个第一分词分别对应的词向量,确定每个分词对应的正向向量。

[0114] 其中,每个第一分词中可以包括一个字符,或者也可以包括多个字符,本公开对此不做限定。

[0115] 另外,若某个第一分词包括一个字符,可以直接将该字符对应的字符向量确定为该第一分词对应的词向量,本公开对此不做限定。

[0116] 或者,若某个第一分词中包括多个字符时,为了保证准确性,可以对该第一分词中的每个字符都进行处理,得到该第一分词对应的词向量。

[0117] 比如说,可以将每个第一分词中的每个字符进行向量映射,以获取每个第一分词

中每个字符对应的字符向量,之后,将每个第一分词中各字符对应的字符向量进行最大池化处理,以生成每个第一分词对应的第一子向量,并将每个第一分词中各字符对应的字符向量进行平均池化处理,以生成每个第一分词对应的第二子向量,再将每个第一分词对应的第一子向量及第二子向量进行融合,以生成每个第一分词对应的词向量。

[0118] 其中,可以使用字符向量表对每个第一分词中的每个字符进行映射,或者也可以使用向量表示子网络中的网络层对每个第一分词中的每个字符进行处理,从而确定出每个字符对应的字符向量等等,本公开对此不做限定。

[0119] 另外,可以利用图神经网络中的向量表示子网络,对每个第一分词中各字符对应的字符向量进行最大池化处理及平均池化处理,以得到每个第一分词对应的第一子向量及第二子向量。

[0120] 可以理解的是,将每个第一分词对应的第一子向量及第二子向量进行融合时,可以有多种融合方式。

[0121] 比如说,可以将每个第一分词对应的第一子向量与第二子向量先进行拼接,之后可以使用网络模型对其进行处理,比如使用前馈神经网络等等,从而可以得到每个第一分词对应的词向量。

[0122] 或者,也可以先对每个第一分词对应的第一子向量、第二子向量分别设置权重,之后按照各个子向量对应的权重对其进行调整,再将调整之后的第一子向量、第二子向量拼接,从而可以得到每个第一分词对应的词向量。

[0123] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中,将第一子向量及第二子向量进行融合时的方式等的限定。

[0124] 举例来说,对于由k个字符构成的第一分词A,其可以记为 $A = (c_i + c_{i+1}, \dots, c_{i+k-1})$,其中,各个字符的字符向量可以由e表示。确定第一分词A的词向量过程如下:先利用图神经网络中的向量表示子网络,对第一分词A中各字符对应的字符向量e进行最大池化处理及平均池化处理,以得到对应的第一子向量 A_{mean} 及第二子向量 A_{max} ,将二者进行拼接之后,可以使用前馈神经网络对其进行处理,从而得到第一分词A对应的词向量 node_A 。具体计算公式如下所示:

$$[0125] \quad A_{\text{max}} = \text{pooling}_{\text{max}}(e_i, e_{i+1}, \dots, e_{i+k-1}) \quad (1)$$

$$[0126] \quad A_{\text{mean}} = \text{pooling}_{\text{mean}}(e_i, e_{i+1}, \dots, e_{i+k-1}) \quad (2)$$

$$[0127] \quad \text{node}_A = \text{FFN}([A_{\text{mean}}, A_{\text{max}}]) \quad (3)$$

[0128] 其中,公式(1)中的max代表向量表示子网络的最大池化操作,公式(2)中的mean代表向量表示子网络的平均池化操作,公式(3)中的 $[A_{\text{mean}}, A_{\text{max}}]$ 表示向量 A_{mean} 和向量的 A_{max} 的拼接操作,FFN代表前馈神经网络。

[0129] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中第一分词及其表示形式、第一子向量及第二子向量进行融合时的方式等的限定。

[0130] 另外,确定每个分词对应的正向向量的方式有多种。

[0131] 比如,可以将各个分词对应的正向分词路径中包含的各个第一分词分别对应的词向量相加,得到的和即为各个分词对应的正向向量。

[0132] 举例来说,分词C3,对应的正向分词路径为: $C3 \rightarrow C4 \rightarrow C5$ 、 $C3 \rightarrow C4C5$,正向路径中包含的第一分词C4对应的词向量为: C_4 ,第一分词C5对应的词向量为: C_5 ,第一分词C4C5对应的

词向量为： C_{45} 。分词C3对应的正向向量可以表示为： $C_3 = C_4 + C_5 + C_{45}$ 。

[0133] 或者，也可以将各个分词对应的正向分词路径中包含的各个第一分词分别对应的词向量进行拼接，得到的拼接结果即为各个分词对应的正向向量。

[0134] 比如说，C3，其对应的正向分词路径为： $C_3 \rightarrow C_4 \rightarrow C_5$ 、 $C_3 \rightarrow C_4C_5$ ，第一分词C4对应的词向量为： C_4 ，第一分词C5对应的词向量为： C_5 ，第一分词C4C5对应的词向量为： C_{45} 。则分词C3对应的正向向量可以表示为： $C_3 = [C_4C_5C_{45}]$ 。

[0135] 或者，也可以先确定每个分词与每个第一分词间的相似度，之后根据相似度，确定每个第一分词的权重，再根据每个第一分词的权重，将各个第一分词对应的词向量进行加权求和，以生成每个分词对应的正向向量。

[0136] 比如说，对于分词C3，其对应的正向分词路径为： $C_3 \rightarrow C_4 \rightarrow C_5$ 、 $C_3 \rightarrow C_4C_5$ ，可以先确定出C3与C4间的相似度、C3与C5间的相似度、C3与C4C5之间的相似度，之后可以再根据各个相似度，可以确定出C4、C5、C4C5分别对应的权重： W_4 、 W_5 、 W_{45} 。又第一分词C4对应的词向量为： C_4 ，第一分词C5对应的词向量为： C_5 ，第一分词C4C5对应的词向量为： C_{45} 。分词C3对应的正向向量可以表示为： $C_3 = W_4 * C_4 + W_5 * C_5 + W_{45} * C_{45}$ 。

[0137] 或者，对于分词w，也可以使用如下所示的公式(4)计算其对应的正向向量：

$$[0138] \quad node_w^{fw} = \sum_{node_A \in N_{fw}(node_w)} \alpha_{wA} * node_A \quad (4)$$

[0139] 其中， $N_{fw}(node_w)$ 代表正向路径中，除分词w之外的各个第一分词， $node_A$ 代表正向路径中任一第一分词A对应的词向量， α_{wA} 代表分词w与正向路径中任一第一分词A的权重。

[0140] 需要说明的是，上述示例只是举例说明，不能作为对本公开实施例中，确定各个分词对应的正向向量的方式等的限定。

[0141] 步骤405，根据每个分词对应的反向分词路径中包含的各个第二分词分别对应的词向量，确定每个分词对应的反向向量。

[0142] 其中，每个第二分词中可以包括一个字符，或者也可以包括多个字符，本公开对此不做限定。

[0143] 另外，若某个第二分词包括一个字符，可以直接可以将该字符对应的字符向量确定为该第二分词对应的词向量，本公开对此不做限定。

[0144] 或者，若某个第二分词中包括多个字符时，为了保证准确性，可以对该第二分词中的每个字符都进行处理，得到该第二分词对应的词向量。

[0145] 比如说，可以将每个第二分词中的每个字符进行向量映射，以获取每个第二分词中每个字符对应的字符向量，之后，将每个第二分词中各字符对应的字符向量进行最大池化处理，以生成每个第二分词对应的第一子向量，并将每个第二分词中各字符对应的字符向量进行平均池化处理，以生成每个第二分词对应的第二子向量，再将每个第二分词对应的第一子向量及第二子向量进行融合，以生成每个第二分词对应的词向量。

[0146] 需要说明的是，确定各个第二分词分别对应的词向量的过程，可以参照确定各个第一分词分别对应的词向量的实现形式，此处不再赘述。

[0147] 另外，确定每个分词对应的反向向量的方式有多种。

[0148] 比如，可以将各个分词对应的反向分词路径中包含的各个第二分词分别对应的词

向量相加,得到的和即为各个分词对应的反向向量。

[0149] 举例来说,分词C3,其对应的反向分词路径为:C1→C2→C3、C1C2→C3,已经确定出第二分词C1对应的词向量为: C_1 ,第二分词C2对应的词向量为: C_2 ,第一分词C1C2对应的词向量为: C_{12} 。分词C3对应的反向向量可以表示为: $C_3 = C_1 + C_2 + C_{12}$ 。

[0150] 或者,也可以将各个分词对应的反向分词路径中包含的各个第二分词分别对应的词向量进行拼接,得到的拼接结果即为各个分词对应的反向向量。

[0151] 比如说,对于分词C3,其对应的反向分词路径为:C1→C2→C3、C1C2→C3,已经确定出第二分词C1对应的词向量为: C_1 ,第二分词C2对应的词向量为: C_2 ,第一分词C1C2对应的词向量为: C_{12} 。分词C3对应的反向向量可以表示为: $C_3 = C_1 C_2 C_{12}$ 。

[0152] 或者,也可以先确定每个分词与每个第二分词间的相似度,之后根据相似度,确定每个第二分词的权重,再根据每个第二分词的权重,将各个第二分词对应的词向量进行加权求和,以生成每个分词对应的反向向量。

[0153] 比如说,对于分词C3,其对应的反向分词路径为C1→C2→C3、C1C2→C3,可以先确定出C1与C3间的相似度、C2与C3间的相似度、C1C2与C3之间的相似度,之后可以再根据各个相似度,确定出C1、C2、C1C2分别对应的权重: W_1 、 W_2 、 W_{12} 。又已经确定出第二分词C1对应的词向量为: C_1 ,第二分词C2对应的词向量为: C_2 ,第一分词C1C2对应的词向量为: C_{12} 。分词C3对应的反向向量可以表示为: $C_3 = W_1 * C_1 + W_2 * C_2 + W_{12} * C_{12}$ 。

[0154] 或者,对于分词w,也可以使用如下所示的公式(5)计算其对应的反向向量:

$$[0155] \quad node_w^{bw} = \sum_{node_B \in N_{bw}(node_w)} \alpha_{wB} * node_B \quad (5)$$

[0156] 其中, $N_{bw}(node_w)$ 代表反向路径中,除分词w之外的各个第二分词, $node_B$ 代表反向路径中任一第二分词B对应的词向量, α_{wB} 代表分词w与反向路径中任一第二分词B的权重。

[0157] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中,确定各个分词对应的反向向量的方式等的限定。

[0158] 步骤406,将每个分词对应的正向向量及每个分词对应的反向向量进行融合,以确定每个分词对应的聚合向量。

[0159] 其中,将每个分词对应的正向向量及每个分词对应的反向向量进行融合时,可能有多种融合方式。

[0160] 举例来说,可以将各个分词对应的正向向量、反向向量直接相加进行融合,从而得到的结果即为各个分词对应的聚合向量。

[0161] 比如,可以使用如下所示的公式(6),表示分词w对应的聚合向量:

$$[0162] \quad node_w^{self} = node_w^{fw} + node_w^{bw} \quad (6)$$

[0163] 其中, $node_w^{fw}$ 代表分词w对应的正向向量, $node_w^{bw}$ 代表分词w对应的反向向量。

[0164] 或者,也可以将各个分词对应的正向向量、反向向量进行拼接,从而得到的拼接结果即为各个分词对应的聚合向量。

[0165] 或者,也可以给各个分词对应的正向向量、反向向量分别设置权重,从而得到的拼

接结果即为各个分词对应的聚合向量。需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中确定分词对应的聚合向量的方式等的限定。

[0166] 步骤407,根据第一分词有向图中各个分词对应的聚合向量,确定第一文本对应的第一向量。

[0167] 其中,确定第一文本对应的第一向量可以有多种方式。

[0168] 比如,可以将第一分词有向图中各个分词对应的聚合向量相加,得到的结果即为第一文本对应的第一向量。

[0169] 或者,也可以将第一分词有向图中各个分词对应的聚合向量进行拼接,得到的拼接结果即为第一文本对应的第一向量。

[0170] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中确定第一文本对应的第一向量等的限定。

[0171] 步骤408,根据第二分词有向图中各个分词对应的聚合向量,确定第二文本对应的第二向量。

[0172] 其中,确定第二文本对应的第二向量的具体实现方式,可以参照上述确定第一文本对应的第一向量的说明,此处不再赘述。

[0173] 步骤409,将第一向量及第二向量分别输入图神经网络的融合子网络,以获取第一文本对应的第一语义向量及第二文本对应的第二语义向量。

[0174] 步骤410,将第一语义向量及第二语义向量分别输入图神经网络的匹配度计算层,以获取第一文本与第二文本间的语义匹配度。

[0175] 本公开实施例中,首先将获取到的第一文本及第二文本分别进行多粒度切词处理,可以获取到对应的第一分词有向图及第二分词有向图,再根据每个分词有向图中各个分词间的连接关系,确定每个分词对应的正向分词路径及反向分词路径,之后可以根据每个分词对应的正向分词路径、反向分词路径,确定出每个分词对应的正向向量、反向向量,并将二者进行融合,得到每个分词对应的聚合向量,之后根据各分词有向图中各个分词对应的聚合向量,可以确定第一文本对应的第一向量、第二文本对应的第二向量,再利用图神经网络确定出第一文本与第二文本间的语义匹配度。由此,通过对包含待匹配文本的多粒度的分词有向图进行处理,以确定两个文本间的语义匹配度,从而提高了语义匹配结果的准确性和可靠性。

[0176] 可以理解的是,在利用图神经网络中的融合子网络确定第一语义向量、第二语义向量时,为了提高准确性,可以进一步考虑两个分词有向图中各个分词之间的联系,比如可以先确定出任一分词有向图中每个第三分词分别与另一分词有向图中各个分词间的相似度,之后再确定出每个第三分词下另一分词有向图中各个第四分词对应的权重,之后根据权重再进行相应的处理,下面结合图5对上述过程进行详细说明。

[0177] 图5A为本公开实施例所提供的语义的匹配方法的流程示意图,如图5所示,该语义的匹配方法可以包括以下步骤:

[0178] 步骤501,获取待匹配的第一文本及第二文本。

[0179] 步骤502,将第一文本及第二文本分别进行多粒度切词处理,以获取第一文本对应的第一分词有向图及第二文本对应的第二分词有向图。

[0180] 步骤503,将第一分词有向图及第二分词有向图,分别输入图神经网络的向量表示

子网络,以获取第一文本对应的第一向量及第二文本对应的第二向量。

[0181] 步骤504,根据任一分词有向图中每个第三分词分别与另一分词有向图中各个分词间的相似度,确定每个第三分词下另一分词有向图中各个第四分词对应的权重。

[0182] 比如说,任一分词有向图中的第三分词:C1,另一分词有向图中的各个分词分别为:D1、D2、D3、D4,可以分别确定出C1与D1之间的相似度、C1与D2之间的相似度、C1与D3之间的相似度、C1与D4之间的相似度。

[0183] 其中,可以使用余弦距离公式确定出各个分词间的相似度,或者也可以使用欧氏距离公式、曼哈顿距离公式等等来确定各个分词间的相似度,本公开对此不做限定。

[0184] 之后,根据C1与D1之间的相似度、C1与D2之间的相似度、C1与D3之间的相似度、C1与D4之间的相似度,可以确定出D1、D2、D3、D4各自对于C1的权重: W_{1C1} 、 W_{2C1} 、 W_{3C1} 、 W_{4C1} 。

[0185] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中各第三分词、第四分词及各个第四分词对应的权重等的限定。

[0186] 步骤505,根据各个第四分词对应的权重,将各个第四分词对应的词向量进行加权求和,确定每个第三分词对应的参考向量。

[0187] 比如说,各个第四分词D1、D2、D3、D4各自对应的词向量分别为: D_1 、 D_2 、 D_3 、 D_4 ,各个第四分词对于第三分词C1的权重分别为: W_{1C1} 、 W_{2C1} 、 W_{3C1} 、 W_{4C1} ,C1对应的参考向量可以表示为: $C_{01}=D_1*W_{1C1}+D_2*W_{2C1}+D_3*W_{3C1}+D_4*W_{4C1}$ 。

[0188] 或者,也可以使用如下所示的公式(7)表示第三分词X对应的参考向量:

$$[0189] \quad node_X^{cross} = \sum_{node_Y \in N'} \alpha_{XY} * node_Y \quad (7)$$

[0190] 其中, N' 代表另一分词有向图中,第三分词X下的各个第四分词Y, $node_Y$ 代表任一第四分词Y对应的词向量, α_{XY} 代表任一第四分词Y对于第三分词X的权重。

[0191] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中,确定每个第三分词对应的参考向量等的限定。

[0192] 步骤506,根据第一分词有向图中各个第三分词对应的参考向量,确定第一分词有向图对应的第三向量。

[0193] 其中,确定第一分词有向图对应的第三向量可以有多种方式。

[0194] 比如,可以将第一分词有向图中各个第三分词对应的参考向量相加,得到的结果即为第一分词有向图对应的第三向量。

[0195] 或者,也可以将第一分词有向图中各个第三分词对应的参考向量进行拼接,得到的拼接结果即为第一分词有向图对应的第三向量。

[0196] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中确定第一文本对应的第一向量等的限定。

[0197] 步骤507,根据第二分词有向图中各个第三分词对应的参考向量,确定第二分词有向图对应的第四向量。

[0198] 其中,确定第二分词有向图对应的第四向量可以有多种方式。

[0199] 比如,可以将第二分词有向图中各个第三分词对应的参考向量相加,得到的结果即为第二分词有向图对应的第四向量。

[0200] 或者,也可以将第二分词有向图中各个第三分词对应的参考向量进行拼接,得到的拼接结果即为第二分词有向图对应的第四向量。

[0201] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中确定第一文本对应的第一向量等的限定。

[0202] 步骤508,将第一向量及第三向量进行融合,以生成第一文本对应的第一语义向量。

[0203] 其中,将第一向量与第三向量进行融合时,有多种方式。

[0204] 比如说,可以将第一向量与第三向量相加,得到的结果即为第一文本对应的第一语义向量。

[0205] 或者,也可以将第一向量与第三向量进行拼接,得到的拼接结果即为第一文本对应的第一语义向量。

[0206] 或者,还可以将第一向量及第三向量进行最大池化处理,以生成第一池化向量,之后将第一向量及第三向量进行平均池化处理,以生成第二池化向量,再将第一池化向量与第二池化向量进行融合,以生成第一文本对应的第一语义向量。

[0207] 其中,可以利用图神经网络中的融合子网络,对第一向量及第三向量进行最大池化处理及平均池化处理,以得到第一池化向量及第二池化向量。

[0208] 另外,将第一池化向量与第二池化向量进行融合时,有多种方式。

[0209] 比如可以直接将第一池化向量和第二池化向量相加,所得结果即为第一文本对应的第一语义向量。

[0210] 或者,也可以将第一池化向量和第二池化向量分别赋予对应的权重,之后将各池化向量与各自对应的权重相乘之后再相加,所得结果即为第一文本对应的第一语义向量。

[0211] 或者,还可以将第一池化向量和第二池化向量直接进行拼接,所得结果即为第一文本对应的第一语义向量。

[0212] 比如说,第一向量记为 $Node_w^{self}$, 第三向量记为 $Node_w^{cross}$, 可以先将第一向量 $Node_w^{self}$ 与第三向量 $Node_w^{cross}$ 进行拼接,之后将拼接后的 $[Node_w^{self}, Node_w^{cross}]$ 进行最大池化处理、平均池化处理,得到对应的第一池化向量max、第二池化向量mean,再将max与mean进行拼接,即可得到第一语义向量 g_1 。具体计算公式如下所示:

$$[0213] \quad max = pooling_{max} \left([Node_w^{self}, Node_w^{cross}] \right) \quad (8)$$

$$[0214] \quad mean = pooling_{mean} \left([Node_w^{self}, Node_w^{cross}] \right) \quad (9)$$

$$[0215] \quad g_1 = [mean, max] \quad (10)$$

[0216] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中将第一向量及第三向量进行融合的方式、第一池化向量和第二池化向量进行融合的方式等的限定。

[0217] 步骤509,将第二向量及第四向量进行融合,以生成第二文本对应的第二语义向量。

[0218] 其中,将两个向量进行融合的具体实现方式,可以参照步骤508,此处不再赘述。

[0219] 步骤510,将第一语义向量及第二语义向量分别输入图神经网络的匹配度计算层,以获取第一文本与第二文本间的语义匹配度。

[0220] 其中,根据第一语义向量、第二语义向量确定第一文本与第二文本间的语义匹配度时,有多种方式。

[0221] 比如,可以使用距离公式,如余弦距离公式、曼哈顿距离公式、欧氏距离公式等等,确定出第一语义向量、第二语义向量间的匹配度,也即第一文本与第二文本间的语义匹配度。

[0222] 或者,也可以先对第一语义向量和第二语义向量进行叉乘和绝对差值计算,之后可以将第一语义向量、第二语义向量、第一语义向量和第二语义向量叉乘的结果、第一语义向量和第二语义向量绝对差值的结果输入到网络模型中,输出的结果即为第一文本与第二文本间的语义匹配度。

[0223] 比如,可以先对第一语义向量 g_1 和第二语义向量 g_2 ,进行叉乘和绝对差值计算,得到叉乘向量 $|g_1 \odot g_2|$ 、绝对差值向量 $|g_1 - g_2|$,之后可以将 g_1 、 g_2 、 $|g_1 \odot g_2|$ 、 $|g_1 - g_2|$ 进行拼接,再使用前馈神经网络FFN对其处理,即可确定出第一文本与第二文本间的语义匹配度score,具体计算公式如下所示:

[0224] $score = FFN([g_1, g_2, g_1 \odot g_2, |g_1 - g_2|])$ (11)

[0225] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中确定第一文本与第二文本间的语义匹配度的方式等的限定。

[0226] 本公开实施例中,充分考虑到两个分词有向图中各个分词之间的关系,由此所确定出的第一语义向量、第二语义向量也更加准确、全面,从而确定出的语义匹配度也更加准确、可靠。

[0227] 本公开提供的语义的匹配方法,也可以应用到检索式对话系统中,下面以图5B所示的示意图为例对其进行说明。

[0228] 如图5B所示,在获取到用户输入的问题之后,可以先对其进行预处理,比如进行噪声文本的过滤等等,之后可以对其进行多粒度切词处理,从而确定出对应的第一分词有向图。从知识库中获取到与用户问题待匹配的答案,对其进行预处理以及多粒度切词处理,从而确定出对应的第二分词有向图。之后可以使用训练生成的图神经网络,将第一分词有向图和第二分词有向图进行语义匹配。再根据语义匹配度对语义匹配结果进行排序,将语义匹配度最高的答案确定为应答结果进行输出,从而可以对用户的问题做出准确的答复,提高了语义匹配的准确性,进而可以给予用户良好的体验。

[0229] 需要说明的是,上述示例只是举例说明,不能作为对本公开实施例中将本公开提供的语义的匹配方法应用到检索式对话系统的方式、内容等的限定。

[0230] 本公开实施例,首先将获取到的第一文本及第二文本分别进行多粒度切词处理,可以获取到对应的第一分词有向图及第二分词有向图,再利用向量表示子网络,可以获取到对应的第一向量及第二向量,再根据任一分词有向图中每个第三分词分别与另一分词有向图中各个分词间的相似度,确定每个第三分词下另一分词有向图中各个第四分词对应的权重,之后可以通过对各个第四分词对应的词向量进行加权求和,确定每个第三分词对应的参考向量,进而确定第一分词有向图对应的第三向量、第二分词有向图对应的第四向量,将第一向量及第三向量进行融合,以生成第一文本对应的第一语义向量,将第二向量及第

四向量进行融合,以生成第二文本对应的第二语义向量,之后再利用匹配度计算层,可以获取第一文本与第二文本间的语义匹配度。由此,通过对包含待匹配文本的多粒度的分词有向图进行处理,以确定两个文本间的语义匹配度,从而提高了语义匹配结果的准确性和可靠性。

[0231] 可以理解的是,图神经网络模型,可以通过训练生成,下面结合图6对上述图神经网络的训练过程进行详细说明。

[0232] 图6为本公开实施例所提供的语义的匹配方法的流程示意图,如图6所示,该语义的匹配方法可以包括以下步骤:

[0233] 步骤601,获取训练样本,其中,训练样本中包括多个文本对。

[0234] 其中,训练样本,可以为包含任意内容的文本样本。文本对,可以为一组包含任意内容的文本。比如,任一文本对可以为问题语句和应答语句、输入语句和应答语句、搜索文本和搜索结果等等任意形式,本公开对此不做限定。

[0235] 步骤602,将每个文本对中的文本分别进行多粒度切词处理,以获取每个文本对对应的两个分词有向图。

[0236] 其中,可以按照不同的切词规则对每个文本对中的文本分别进行多粒度切词处理,从而可以得到切词处理后的多种分词结果,之后根据各分词之间的联系,可以构成该文本对对应的分词有向图。

[0237] 步骤603,将每个文本对对应的两个分词有向图分别输入教师模型及学生模型,以获取教师模型输出的文本对的第一语义匹配度、及学生模型输出的文本对的第二语义匹配度,其中,教师模型为已训练完成的模型。

[0238] 其中,教师模型,可以为已训练完成的模型,比如可以使用深层预训练模型等等作为教师模型,本公开对此不做限定。

[0239] 另外,学生模型,可以为还未训练完成的模型,其可以根据教师模型进行相应的调整等等,本公开对此不做限定。

[0240] 可以理解的是,本公开实施例中的教师模型可以具备提取深层语义特征的能力,可以对文本对进行更为准确的匹配,从而其输出文本对的第一语义匹配度更为准确、可靠。

[0241] 步骤604,根据第一语义匹配度与第二语义匹配度的差异,对学生模型进行修正,以生成图神经网络模型。

[0242] 其中,可以先确定出第一语义匹配度与第二语义匹配度之间的差异,之后根据该差异,对学生模型使用反向传播进行逐层反向修正,以使学生模型的学习能力得到提高,具有更为接近教师网络的性能。

[0243] 本公开实施例中,通过使用教师模型对学生模型进行修正,可以使学生模型具有和教师模型更加相似的特征,具有更好的学习能力和特征表达,从而可以进一步提升学生模型的性能。进而,使用该训练完成的学生模型作为图神经网络模型对待匹配文本进行匹配,不仅提高了文本匹配的效率和保证了匹配结果的精确性和可靠性。

[0244] 本公开实施例,先获取训练样本,其中,训练样本中包括多个文本对,之后将每个文本对中的文本分别进行多粒度切词处理,以获取每个文本对对应的两个分词有向图,再将每个文本对对应的两个分词有向图分别输入教师模型及学生模型,以获取教师模型输出的文本对的第一语义匹配度、及学生模型输出的所述文本对的第二语义匹配度,其中,教师

模型为已训练完成的模型,之后再根据第一语义匹配度与第二语义匹配度的差异,对学生模型进行修正,以生成图神经网络模型。由此,通过基于训练样本中的各个文本对及对应的分词有向图,利用教师模型对学生模型进行修正,从而不仅可以提高学生网络的性能,以生成性能更好的图神经网络模型,而且为提高语义匹配的准确性和效率,提供了条件。

[0245] 为了实现上述实施例,本公开还提出一种语义的匹配装置。

[0246] 图7为本公开实施例所提供的语义的匹配装置的结构示意图。

[0247] 如图7所示,该语义的匹配装置100可以包括:第一获取模块110、处理模块120、第二获取模块130。

[0248] 其中,第一获取模块110,用于获取待匹配的第一文本及第二文本。

[0249] 处理模块120,用于将所述第一文本及所述第二文本分别进行多粒度切词处理,以获取所述第一文本对应的第一分词有向图及所述第二文本对应的第二分词有向图。

[0250] 第二获取模块130,用于将所述第一分词有向图及所述第二分词有向图,分别输入训练生成的图神经网络中,以获取所述第一文本与所述第二文本间的语义匹配度。

[0251] 在一种可能的实现方式中,所述处理模块120,具体用于将所述第一文本及所述第二文本分别进行多粒度切词处理,以获取所述第一文本对应的多个第一分词有向组及所述第二文本对应的多个第二分词有向组;将所述多个第一分词有向组及所述多个第二分词有向组中的各分词,分别与预设的词表中各参考分词进行匹配;在所述预设的词表中包含所述多个第一分词有向组中的各分词的情况下,基于所述多个第一分词有向组中的相同分词,将所述多个第一分词有向组进行融合,以生成所述第一分词有向图;在所述预设的词表中包含所述多个第二分词有向组中的各分词的情况下,基于所述多个第二分词有向组中的相同分词,将所述多个第二分词有向组进行融合,以生成所述第二分词有向图。

[0252] 在一种可能的实现方式中,所述处理模块120,还具体用于在所述任一第一分词有向组中的任一分词与预设的词表中各参考分词均未匹配的情况下,将所述任一分词进行切词处理,以获取所述任一分词对应的有向词片段;在所述预设的词表中包括每个所述词片段的情况下,利用所述有向词片段替换所述任一分词,以生成新的第一分词有向组;基于新的多个第一分词有向组中的相同分词,将所述多个第一分词有向组进行融合,以生成所述第一分词有向图。

[0253] 在一种可能的实现方式中,所述第二获取模块120,包括:

[0254] 第一获取单元,用于将所述第一分词有向图及所述第二分词有向图,分别输入所述图神经网络的向量表示子网络,以获取所述第一文本对应的第一向量及所述第二文本对应的第二向量;

[0255] 第二获取单元,用于将所述第一向量及所述第二向量分别输入所述图神经网络的融合子网络,以获取所述第一文本对应的第一语义向量及所述第二文本对应的第二语义向量;

[0256] 第三获取单元,用于将所述第一语义向量及所述第二语义向量分别输入所述图神经网络的匹配度计算层,以获取所述第一文本与所述第二文本间的语义匹配度。

[0257] 在一种可能的实现方式中,所述第一获取单元,包括:

[0258] 第一确定子单元,用于根据每个所述分词有向图中各个分词间的连接关系,确定每个所述分词对应的正向分词路径及反向分词路径,其中,任一分词对应的正向分词路径

为由所述任一分词起至文本的结束符止的各分词组成的路径,任一分词对应的反向分词路径为由文本的起始符起至所述任一分词止的各分词组成的路径;

[0259] 第二确定子单元,用于根据每个所述分词对应的正向分词路径中包含的各个第一分词分别对应的词向量,确定每个所述分词对应的正向向量;

[0260] 第三确定子单元,用于根据每个所述分词对应的反向分词路径中包含的各个第二分词分别对应的词向量,确定每个所述分词对应的反向向量;

[0261] 融合子单元,用于将每个所述分词对应的正向向量及每个所述分词对应的反向向量进行融合,以确定每个所述分词对应的聚合向量;

[0262] 第四确定子单元,用于根据所述第一分词有向图中各个分词对应的聚合向量,确定所述第一文本对应的第一向量;

[0263] 第五确定子单元,用于根据所述第二分词有向图中各个分词对应的聚合向量,确定所述第二文本对应的第二向量。

[0264] 在一种可能的实现方式中,所述第二确定子单元,具体用于确定每个所述分词与每个所述第一分词间的相似度;根据所述相似度,确定每个所述第一分词的权重;根据每个所述第一分词的权重,将各个所述第一分词对应的词向量进行加权求和,以生成每个所述分词对应的正向向量。

[0265] 在一种可能的实现方式中,所述第一获取单元,还包括:

[0266] 第一获取子单元,用于将每个所述分词中的每个字符进行向量映射,以获取每个所述分词中每个字符对应的字符向量;

[0267] 第一生成子单元,用于将每个所述分词中各字符对应的字符向量进行最大池化处理,以生成每个所述分词对应的第一子向量;

[0268] 第二生成子单元,用于将每个所述分词中各字符对应的字符向量进行平均池化处理,以生成每个所述分词对应的第二子向量;

[0269] 第三生成子单元,用于将每个所述分词对应的第一子向量及第二子向量进行融合,以生成每个所述分词对应的词向量。

[0270] 在一种可能的实现方式中,所述第二获取单元,包括:

[0271] 第六确定子单元,用于根据任一分词有向图中每个第三分词分别与另一分词有向图中各个分词间的相似度,确定每个所述第三分词下所述另一分词有向图中各个第四分词对应的权重;

[0272] 第七确定子单元,用于根据所述各个第四分词对应的权重,将所述各个第四分词对应的词向量进行加权求和,确定每个所述第三分词对应的参考向量;

[0273] 第八确定子单元,用于根据所述第一分词有向图中各个第三分词对应的参考向量,确定所述第一分词有向图对应的第三向量;

[0274] 第九确定子单元,用于根据所述第二分词有向图中各个第三分词对应的参考向量,确定所述第二分词有向图对应的第四向量;

[0275] 第三生成子单元,用于将所述第一向量及所述第三向量进行融合,以生成所述第一文本对应的第一语义向量;

[0276] 第四生成子单元,用于将所述第二向量及所述第四向量进行融合,以生成所述第二文本对应的第二语义向量。

[0277] 在一种可能的实现方式中,所述第三生成子单元,具体用于:将所述第一向量及所述第三向量进行最大池化处理,以生成第一池化向量;将所述第一向量及所述第三向量进行平均池化处理,以生成第二池化向量;将所述第一池化向量与所述第二池化向量进行融合,以生成所述第一文本对应的第一语义向量。

[0278] 在一种可能的实现方式中,所述第一获取模块110,还用于获取训练样本,其中,所述训练样本中包括多个文本对。

[0279] 所述处理模块120,还用于将每个所述文本对中的文本分别进行多粒度切词处理,以获取每个所述文本对对应的两个分词有向图。

[0280] 所述第二获取模块130,还用于将每个所述文本对对应的两个分词有向图分别输入教师模型及学生模型,以获取所述教师模型输出的所述文本对的第一语义匹配度、及所述学生模型输出的所述文本对的第二语义匹配度,其中,所述教师模型为已训练完成的模型。

[0281] 所述第二获取模块130,还用于根据所述第一语义匹配度与所述第二语义匹配度的差异,对所述学生模型进行修正,以生成所述图神经网络模型。

[0282] 本公开实施例中的上述各模块的功能及具体实现原理,可参照上述各方法实施例,此处不再赘述。

[0283] 本公开实施例的语义的匹配装置,可以先获取待匹配的第一文本及第二文本,之后将第一文本及第二文本分别进行多粒度切词处理,以获取第一文本对应的第一分词有向图及第二文本对应的第二分词有向图,之后再再将第一分词有向图及第二分词有向图,分别输入训练生成的图神经网络中,以获取第一文本与第二文本间的语义匹配度。由此,通过对包含待匹配文本的多粒度的分词有向图进行处理,以确定两个文本间的语义匹配度,从而提高了语义匹配结果的准确性和可靠性。

[0284] 为了实现上述实施例,本公开还提出一种计算机设备,包括:存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,处理器执行程序时,实现如本公开前述实施例提出的语义的匹配方法。

[0285] 为了实现上述实施例,本公开还提出一种非临时性计算机可读存储介质,存储有计算机程序,计算机程序被处理器执行时实现如本公开前述实施例提出的语义的匹配方法。

[0286] 为了实现上述实施例,本公开还提出一种计算机程序产品,当计算机程序产品中的指令处理器执行时,执行如本公开前述实施例提出的语义的匹配方法。

[0287] 图8示出了适于用来实现本公开实施方式的示例性计算机设备的框图。图8显示的计算机设备12仅仅是一个示例,不应对本公开实施例的功能和使用范围带来任何限制。

[0288] 如图8所示,计算机设备12以通用计算设备的形式表现。计算机设备12的组件可以包括但不限于:一个或者多个处理器或者处理单元16,系统存储器28,连接不同系统组件(包括系统存储器28和处理单元16)的总线18。

[0289] 总线18表示几类总线结构中的一种或多种,包括存储器总线或者存储器控制器,外围总线,图形加速端口,处理器或者使用多种总线结构中的任意总线结构的局域总线。举例来说,这些体系结构包括但不限于工业标准体系结构(Industry Standard Architecture;以下简称:ISA)总线,微通道体系结构(Micro Channel Architecture;以下

简称:MAC)总线,增强型ISA总线、视频电子标准协会(Video Electronics Standards Association;以下简称:VESA)局域总线以及外围组件互连(Peripheral Component Interconnection;以下简称:PCI)总线。

[0290] 计算机设备12典型地包括多种计算机系统可读介质。这些介质可以是任何能够被计算机设备12访问的可用介质,包括易失性和非易失性介质,可移动的和不可移动的介质。

[0291] 存储器28可以包括易失性存储器形式的计算机系统可读介质,例如随机存取存储器(Random Access Memory;以下简称:RAM)30和/或高速缓存存储器32。计算机设备12可以进一步包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例,存储系统34可以用于读写不可移动的、非易失性磁介质(图8未显示,通常称为“硬盘驱动器”)。尽管图8中未示出,可以提供用于对可移动非易失性磁盘(例如“软盘”)读写的磁盘驱动器,以及对可移动非易失性光盘(例如:光盘只读存储器(Compact Disc Read Only Memory;以下简称:CD-ROM)、数字多功能只读光盘(Digital Video Disc Read Only Memory;以下简称:DVD-ROM)或者其它光介质)读写的光盘驱动器。在这些情况下,每个驱动器可以通过一个或者多个数据介质接口与总线18相连。存储器28可以包括至少一个程序产品,该程序产品具有一组(例如至少一个)程序模块,这些程序模块被配置以执行本公开各实施例的功能。

[0292] 具有一组(至少一个)程序模块42的程序/实用工具40,可以存储在例如存储器28中,这样的程序模块42包括但不限于操作系统、一个或者多个应用程序、其它程序模块以及程序数据,这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块42通常执行本公开所描述的实施例中的功能和/或方法。

[0293] 计算机设备12也可以与一个或多个外部设备14(例如键盘、指向设备、显示器24等)通信,还可与一个或者多个使得用户能与该计算机设备12交互的设备通信,和/或与使得该计算机设备12能与一个或多个其它计算设备进行通信的任何设备(例如网卡,调制解调器等等)通信。这种通信可以通过输入/输出(I/O)接口22进行。并且,计算机设备12还可以通过网络适配器20与一个或者多个网络(例如局域网(Local Area Network;以下简称:LAN),广域网(Wide Area Network;以下简称:WAN)和/或公共网络,例如因特网)通信。如图示,网络适配器20通过总线18与计算机设备12的其它模块通信。应当明白,尽管图中未示出,可以结合计算机设备12使用其它硬件和/或软件模块,包括但不限于:微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID系统、磁带驱动器以及数据备份存储系统等。

[0294] 处理单元16通过运行存储在系统存储器28中的程序,从而执行各种功能应用以及数据处理,例如实现前述实施例中提及的方法。

[0295] 本公开的技术方案,可以先获取待匹配的第一文本及第二文本,之后将第一文本及第二文本分别进行多粒度切词处理,以获取第一文本对应的第一分词有向图及第二文本对应的第二分词有向图,之后再再将第一分词有向图及第二分词有向图,分别输入训练生成的图神经网络中,以获取第一文本与第二文本间的语义匹配度。由此,通过对包含待匹配文本的多粒度的分词有向图进行处理,以确定两个文本间的语义匹配度,从而提高了语义匹配结果的准确性和可靠性。

[0296] 在本说明书的描述中,参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特

点包含于本公开的至少一个实施例或示例中。在本说明书中,对上述术语的示意性表述不必针对的是相同的实施例或示例。而且,描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外,在不相互矛盾的情况下,本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[0297] 此外,术语“第一”、“第二”仅用于描述目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。在本公开的描述中,“多个”的含义是至少两个,例如两个,三个等,除非另有明确具体的限定。

[0298] 流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为,表示包括一个或更多个用于实现定制逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分,并且本公开的优选实施方式的范围包括另外的实现,其中可以不按所示出或讨论的顺序,包括根据所涉及的功能按基本同时的方式或按相反的顺序,来执行功能,这应被本公开的实施例所属技术领域的技术人员所理解。

[0299] 在流程图中表示或在此以其他方式描述的逻辑和/或步骤,例如,可以被认为用于实现逻辑功能的可执行指令的定序列表,可以具体实现在任何计算机可读介质中,以供指令执行系统、装置或设备(如基于计算机的系统、包括处理器的系统或其他可以从指令执行系统、装置或设备取指令并执行指令的系统)使用,或结合这些指令执行系统、装置或设备而使用。就本说明书而言,“计算机可读介质”可以是任何可以包含、存储、通信、传播或传输程序以供指令执行系统、装置或设备或结合这些指令执行系统、装置或设备而使用的装置。计算机可读介质的更具体的示例(非穷尽性列表)包括以下:具有一个或多个布线的电连接部(电子装置),便携式计算机盘盒(磁装置),随机存取存储器(RAM),只读存储器(ROM),可擦除可编程只读存储器(EPROM或闪速存储器),光纤装置,以及便携式光盘只读存储器(CDROM)。另外,计算机可读介质甚至可以是可在其上打印所述程序的纸或其他合适的介质,因为可以例如通过对纸或其他介质进行光学扫描,接着进行编辑、解译或必要时以其他合适方式进行处理来以电子方式获得所述程序,然后将其存储在计算机存储器中。

[0300] 应当理解,本公开的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中,多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。如,如果用硬件来实现和在另一实施方式中一样,可用本领域公知的下列技术中的任一项或他们的组合来实现:具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路,具有合适的组合逻辑门电路的专用集成电路,可编程门阵列(PGA),现场可编程门阵列(FPGA)等。

[0301] 本技术领域的普通技术人员可以理解实现上述实施例方法携带的全部或部分步骤是可以通程序来指令相关的硬件完成,所述的程序可以存储于一种计算机可读存储介质中,该程序在执行时,包括方法实施例的步骤之一或其组合。

[0302] 此外,在本公开各个实施例中的各功能单元可以集成在一个处理模块中,也可以是各个单元单独物理存在,也可以两个或两个以上单元集成在一个模块中。上述集成的模块既可以采用硬件的形式实现,也可以采用软件功能模块的形式实现。所述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用,也可以存储在一个计算机

可读取存储介质中。

[0303] 上述提到的存储介质可以是只读存储器,磁盘或光盘等。尽管上面已经示出和描述了本公开的实施例,可以理解的是,上述实施例是示例性的,不能理解为对本公开的限制,本领域的普通技术人员在本公开的范围内可以对上述实施例进行变化、修改、替换和变型。

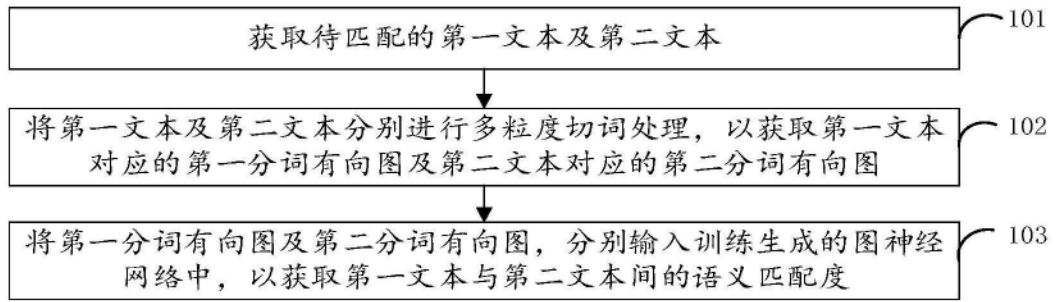


图1A

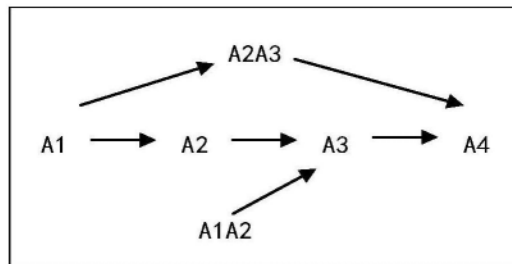


图1B

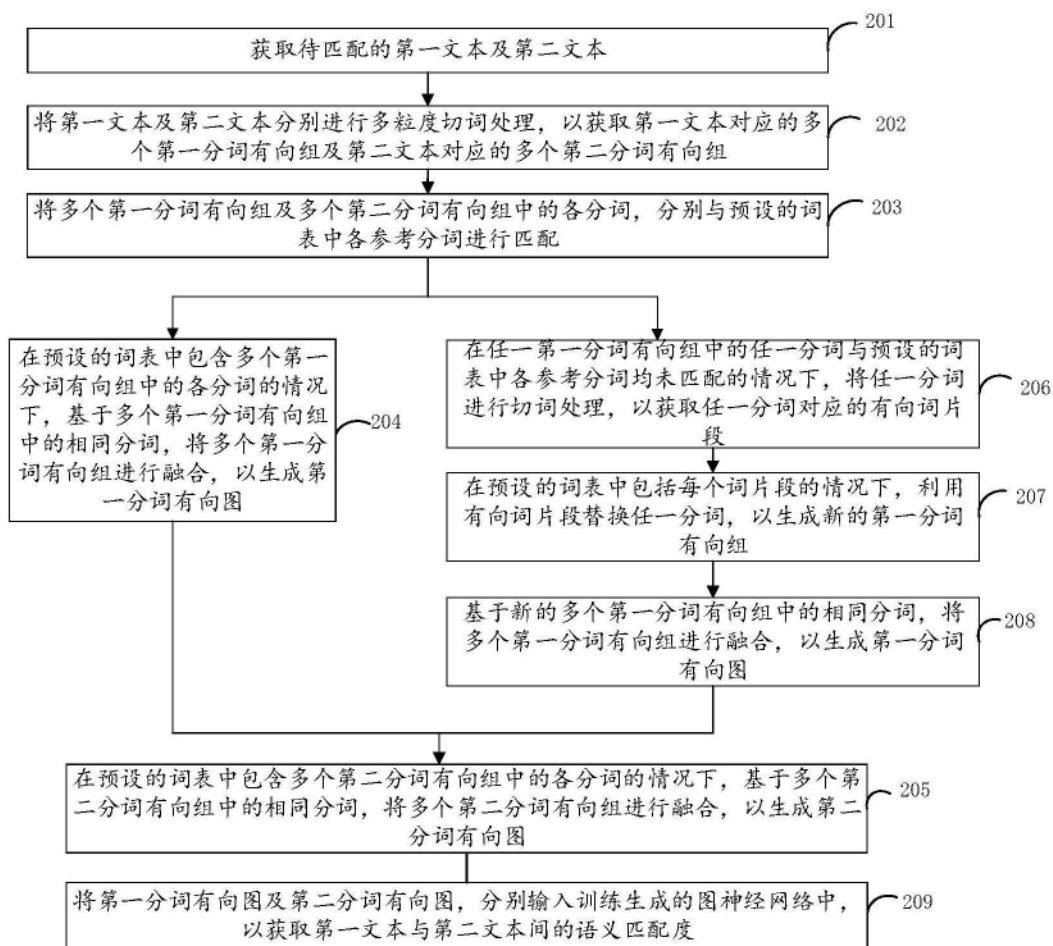


图2A

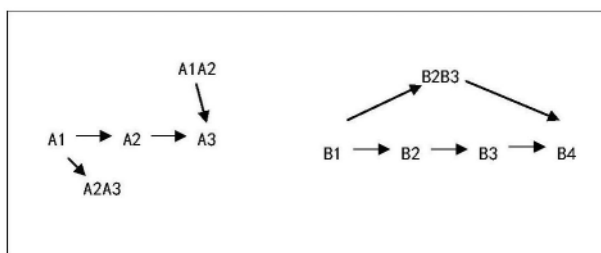


图2B

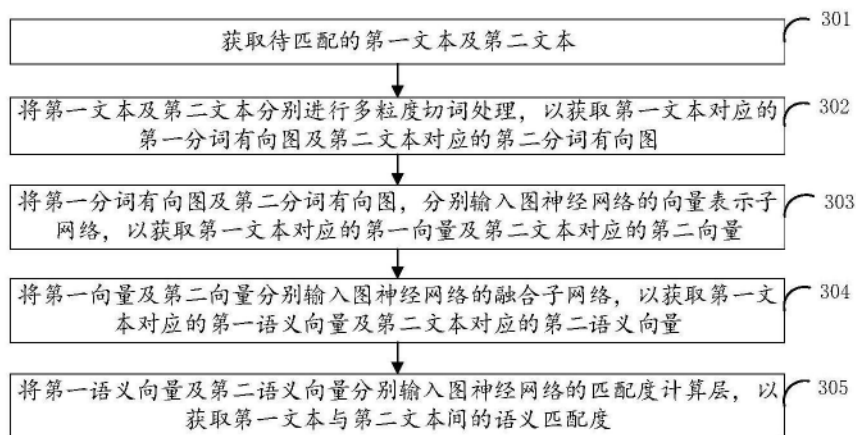


图3A

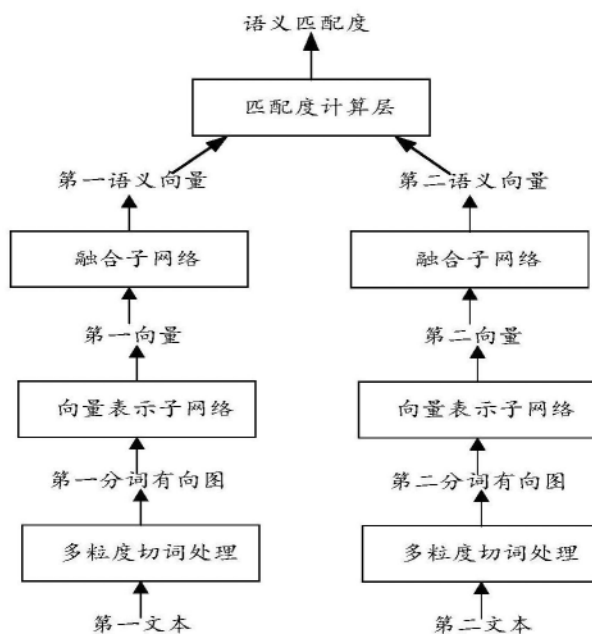


图3B

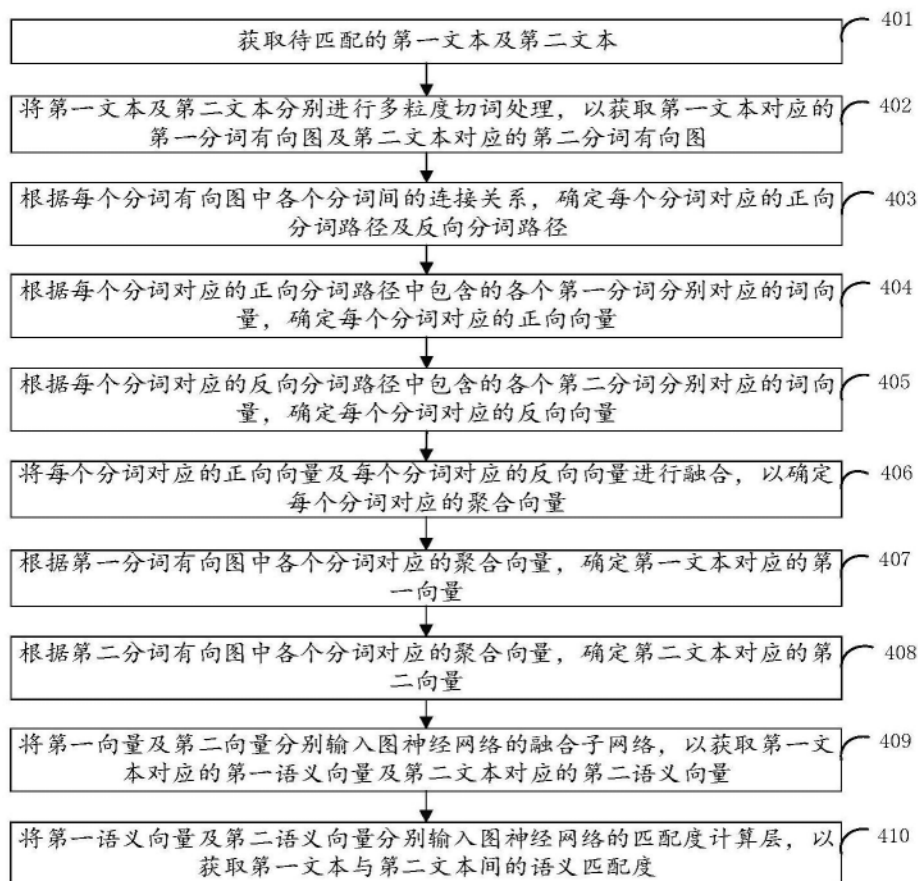


图4A

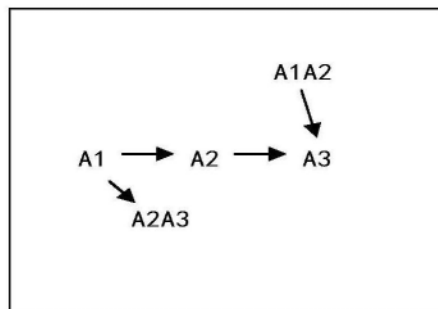


图4B



图5A

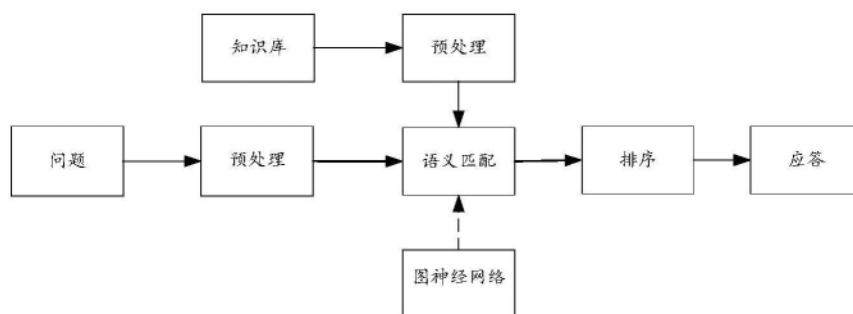


图5B

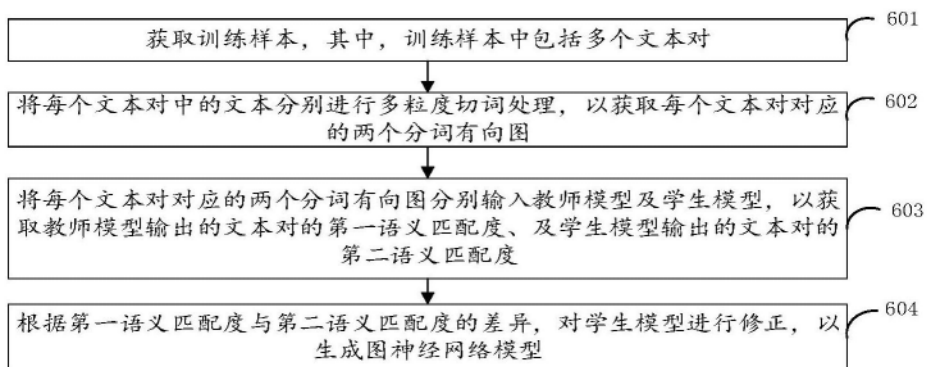


图6

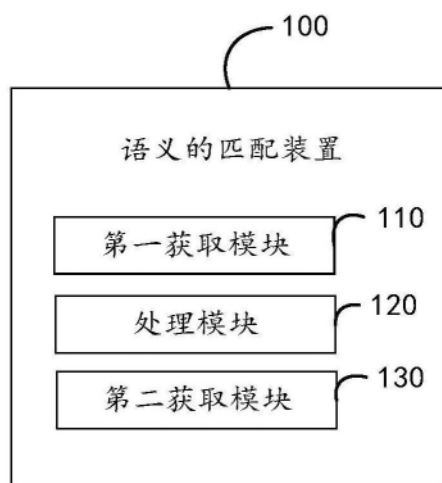


图7

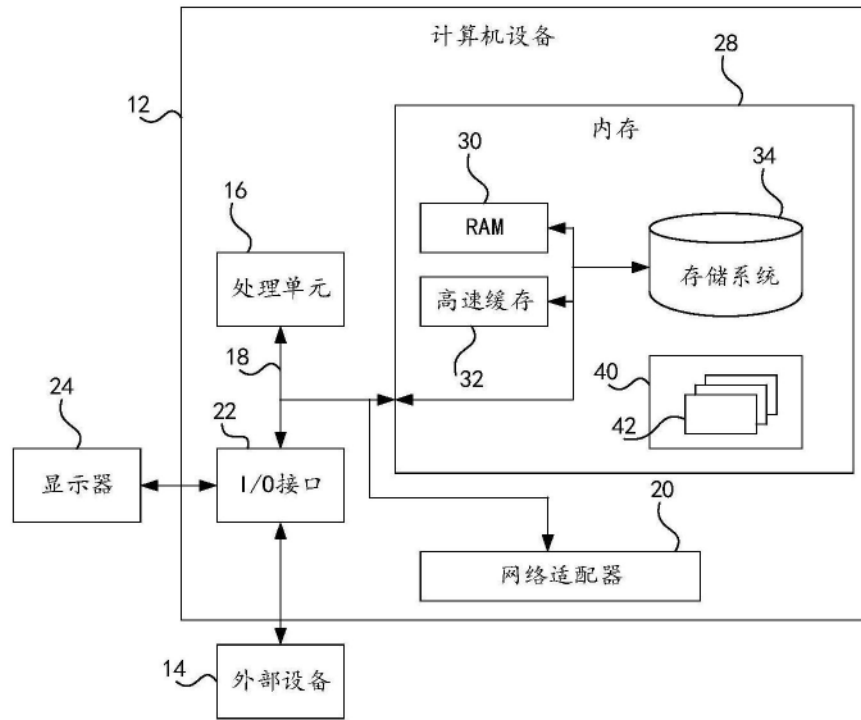


图8