



(12) 发明专利申请

(10) 申请公布号 CN 112446202 A

(43) 申请公布日 2021.03.05

(21) 申请号 201910759849.X

(22) 申请日 2019.08.16

(71) 申请人 阿里巴巴集团控股有限公司

地址 英属开曼群岛大开曼资本大厦一座四层847号邮箱

(72) 发明人 宋凯嵩 林君 孙常龙 赵露君

(74) 专利代理机构 北京博浩百睿知识产权代理有限公司 11134

代理人 谢湘宁 张文华

(51) Int.Cl.

G06F 40/205 (2020.01)

G06F 40/295 (2020.01)

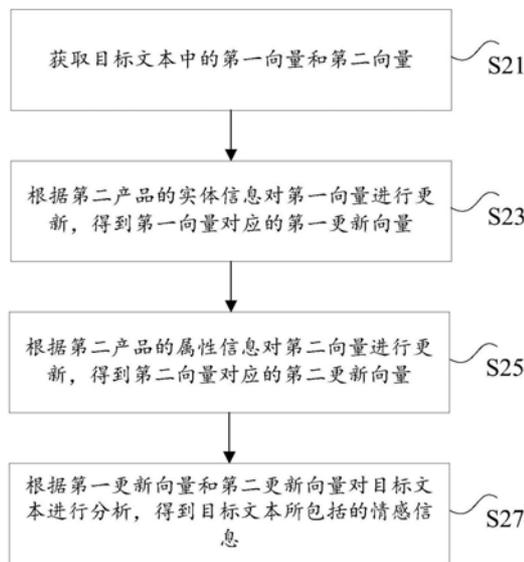
权利要求书5页 说明书20页 附图9页

(54) 发明名称

文本的分析方法和装置

(57) 摘要

本发明公开了一种文本的分析方法和装置。其中,该方法包括:获取目标文本中的第一向量和第二向量,其中,第一向量用于表示第一产品的实体信息,第二向量用于表示第一产品的属性信息,第一产品为目标文本中记载的产品;根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,其中,第二产品为预设文本中记载的产品;根据第二产品的属性信息对第二向量进行更新,得到第二向量对应的第二更新向量;根据第一更新向量和第二更新向量对目标文本进行分析,得到目标文本所包括的情感信息。本发明解决了现有技术中在情感预测模型的训练数据较少的情况下,对文本的情感分析不准确的技术问题。



1. 一种文本的分析方法,其特征在于,包括:

获取目标文本中的第一向量和第二向量,其中,所述第一向量用于表示第一产品的实体信息,所述第二向量用于表示所述第一产品的属性信息,所述第一产品为所述目标文本中记载的产品;

根据第二产品的实体信息对所述第一向量进行更新,得到所述第一向量对应的第一更新向量,其中,所述第二产品为预设文本中记载的产品;

根据第二产品的属性信息对所述第二向量进行更新,得到所述第二向量对应的第二更新向量;

根据所述第一更新向量和所述第二更新向量对所述目标文本进行分析,得到所述目标文本所包括的情感信息。

2. 根据权利要求1所述的方法,其特征在于,获取目标文本中的第一向量和第二向量,包括:

对所述目标文本进行分词,得到多个词语;

确定所述多个词语中用于表示所述实体信息的实体词语和用于表示所述属性信息的属性词语;

对所述实体词语进行向量化处理得到所述第一向量;

对所述属性词语进行向量化处理得到所述第二向量。

3. 根据权利要求1所述的方法,其特征在于,根据第二产品的实体信息对所述第一向量进行更新,得到所述第一向量对应的第一更新向量,包括:根据第二产品的实体信息对所述第一向量进行迭代更新,得到所述第一向量对应的第一更新向量;

其中,根据第二产品的实体信息对所述第一向量进行迭代更新,得到所述第一向量对应的第一更新向量,包括:

确定所述第一向量在此次迭代中对应的输入向量;

根据所述第二产品的实体信息确定所述第一向量在所述第一更新向量中的权重;

根据所述第二产品的实体信息确定所述第一向量在此次迭代中对应的补充向量;

根据所述补充向量和所述权重对此次迭代中对应的所述输入向量进行更新,得到所述第一更新向量。

4. 根据权利要求3所述的方法,其特征在于,确定所述第一向量在此次迭代中对应的输入向量,包括:

根据上一次迭代得到的第一更新向量、上一次迭代得到的第二更新向量和所述目标文本中的词语向量集合,确定上一次迭代中用于表示所述目标文本的上下文信息的上下文向量;

确定上一次迭代得到的第一更新向量与上一次迭代中的所述上下文向量之和为所述第一向量在此次迭代中对应的输入向量。

5. 根据权利要求3所述的方法,其特征在于,在进行首次迭代时,确定所述第一向量在此次迭代中对应的输入向量,包括:

在所述目标文本包括多个第一向量的情况下,确定所述多个第一向量的均值为所述输入向量。

6. 根据权利要求3所述的方法,其特征在于,根据所述第二产品的实体信息确定所述第

一向量在所述第一更新向量中的权重,包括:

获取所述第一向量在训练数据集中出现的次数;

确定训练数据集中的实体向量出现的平均次数;

以自然常数为底,以所述第一向量在训练数据集中出现的次数与所述平均次数比值为指数,确定第一数值;

确定预设值与所述第一数值的比值为第二数值;

确定预设值与所述第二数值之差为所述第一向量在所述第一更新向量中的权重。

7. 根据权利要求3所述的方法,其特征在于,根据所述第二产品的实体信息确定所述第一向量在此次迭代中对应的补充向量,包括:

获取训练数据集中每个实体向量的相似度参数,其中,所述相似度参数用于表示所述实体向量与所述第一向量的相似程度;

获取所述每个实体向量与对应的相似度参数的第一乘积;

确定所述训练数据集中每个实体向量对应的第一乘积之和为所述补充向量。

8. 根据权利要求7所述的方法,其特征在于,获取训练数据集中每个实体向量的相似度参数,包括:

确定所述训练数据集中的实体向量的权重与所述实体向量的第二乘积;

将所述第二乘积与此次迭代的输入向量拼接后通过预设的感知器,得到感知结果;

基于所述感知结果确定所述实体向量对应的相似度参数。

9. 根据权利要求3所述的方法,其特征在于,根据所述补充向量和所述权重对所述此次迭代中对应的输入向量进行更新,得到所述第一更新向量,包括:

获取所述权重与此次迭代的输入向量的第三乘积;

获取预设值与所述权重的差值,并获取所述差值与所述补充向量的第四乘积;

确定所述第三乘积和所述第四乘积之和为所述第一更新向量。

10. 根据权利要求1所述的方法,其特征在于,所述目标文本至少包括如下一种或多种:用户原创内容、对预设对象的评价信息和对所述预设对象的使用说明。

11. 一种文本的分析方法,包括:

显示获取的目标文本,其中,所述目标文本包括了第一产品的实体信息和所述产品的属性信息,所述第一产品为所述目标文本中记载的产品;

获取所述目标文本中的第一向量和第二向量,其中,所述第一向量用于表示所述第一产品的实体信息,所述第二向量用于表示所述第一产品的属性信息;

根据第二产品的实体信息对所述第一向量进行更新,得到所述第一向量对应的第一更新向量,其中,所述第二产品为预设文本中记载的产品;

根据第二产品的属性信息对所述第二向量进行更新,得到所述第二向量对应的第二更新向量;

根据所述第一更新向量和所述第二更新向量对所述目标文本进行分析,输出所述目标文本所包括的情感信息。

12. 根据权利要求11所述的方法,其特征在于,根据第二产品的实体信息对所述第一向量进行更新,得到所述第一向量对应的第一更新向量,包括:根据第二产品的实体信息对所述第一向量进行迭代更新,得到所述第一向量对应的第一更新向量;

其中,根据第二产品的实体信息对所述第一向量进行迭代更新,得到所述第一向量对应的第一更新向量,包括:

确定所述第一向量在此次迭代中对应的输入向量;

根据所述第二产品的实体信息确定所述第一向量在所述第一更新向量中的权重;

根据所述第二产品的实体信息确定所述第一向量在此次迭代中对应的补充向量;

根据所述补充向量和所述权重对此次迭代中对应的所述输入向量进行更新,得到所述第一更新向量。

13. 一种文本的分析方法,其特征在于,包括:

获取第一产品的评价信息,其中,所述评价信息包括了第一产品的实体信息和所述第一产品的属性信息,所述第一产品为所述评价信息所评价的产品;

获取所述评价信息中的第一向量和第二向量,其中,所述第一向量用于表示所述第一产品的实体信息,所述第二向量用于表示所述第一产品的属性信息;

根据第二产品的实体信息对所述第一向量进行更新,得到所述第一向量对应的第一更新向量,其中,所述第二产品为预设文本中记载的产品;

根据第二产品的属性信息对所述第二向量进行更新,得到所述第二向量对应的第二更新向量;

根据所述第一更新向量和所述第二更新向量对所述评价信息进行分析,得到所述评价信息所包括的情感信息。

14. 根据权利要求13所述的方法,其特征在于,根据第二产品的实体信息对所述第一向量进行更新,得到所述第一向量对应的第一更新向量,包括:根据第二产品的实体信息对所述第一向量进行迭代更新,得到所述第一向量对应的第一更新向量;

其中,根据第二产品的实体信息对所述第一向量进行迭代更新,得到所述第一向量对应的第一更新向量,包括:

确定所述第一向量在此次迭代中对应的输入向量;

根据所述第二产品的实体信息确定所述第一向量在所述第一更新向量中的权重;

根据所述第二产品的实体信息确定所述第一向量在此次迭代中对应的补充向量;

根据所述补充向量和所述权重对此次迭代中对应的所述输入向量进行更新,得到所述第一更新向量。

15. 一种文本的分析装置,其特征在于,包括:

获取模块,用于获取目标文本中的第一向量和第二向量,其中,所述第一向量用于表示第一产品的实体信息,所述第二向量用于表示所述第一产品的属性信息,所述第一产品为所述目标文本中记载的产品;

第一更新模块,用于根据第二产品的实体信息对所述第一向量进行更新,得到所述第一向量对应的第一更新向量,其中,所述第二产品为预设文本中记载的产品;

第二更新模块,用于根据第二产品的属性信息对所述第二向量进行更新,得到所述第二向量对应的第二更新向量;

分析模块,用于根据所述第一更新向量和所述第二更新向量对所述目标文本进行分析,得到所述目标文本所包括的情感信息。

16. 根据权利要求15所述的装置,其特征在于,所述第一更新模块包括:第一更新子模

块,用于根据第二产品的实体信息对所述第一向量进行迭代更新,得到所述第一向量对应的第一更新向量;

其中,所述第一更新子模块包括:

第一确定单元,用于确定所述第一向量在此次迭代中对应的输入向量;

第二确定单元,用于根据所述第二产品的实体信息确定所述第一向量在所述第一更新向量中的权重;

第三确定单元,用于根据所述第二产品的实体信息确定所述第一向量在此次迭代中对应的补充向量;

更新单元,用于根据所述补充向量和所述权重对此次迭代中对应的所述输入向量进行更新,得到所述第一更新向量。

17. 根据权利要求16所述的装置,其特征在于,所述第二确定单元包括:

第一获取子单元,用于获取所述第一向量在训练数据集中出现的次数;

第一确定子单元,用于确定训练数据集中的实体向量出现的平均次数;

第二确定子单元,用于以自然常数为底,以所述第一向量在训练数据集中出现的次数与所述平均次数比值为指数,确定第一数值;

第三确定子单元,用于确定预设值与所述第一数值的比值为第二数值;

第四确定子单元,用于确定预设值与所述第二数值之差为所述第一向量在所述第一更新向量中的权重。

18. 根据权利要求16所述的装置,其特征在于,所述第三确定单元包括:

第二获取子单元,用于获取训练数据集中每个实体向量的相似度参数,其中,所述相似度参数用于表示所述实体向量与所述第一向量的相似程度;

第三获取子单元,用于获取所述每个实体向量与对应的相似度参数的第一乘积;

第五确定子单元,用于确定所述训练数据集中每个实体向量对应的第一乘积之和为所述补充向量。

19. 一种存储介质,其特征在于,所述存储介质包括存储的程序,其中,在所述程序运行时控制所述存储介质所在设备执行如下步骤:

获取目标文本中的第一向量和第二向量,其中,所述第一向量用于表示第一产品的实体信息,所述第二向量用于表示所述第一产品的属性信息,所述第一产品为所述目标文本中记载的产品;

根据第二产品的实体信息对所述第一向量进行更新,得到所述第一向量对应的第一更新向量,其中,所述第二产品为预设文本中记载的产品;

根据第二产品的属性信息对所述第二向量进行更新,得到所述第二向量对应的第二更新向量;

根据所述第一更新向量和所述第二更新向量对所述目标文本进行分析,得到所述目标文本所包括的情感信息。

20. 一种处理器,其特征在于,所述处理器用于运行程序,其中,所述程序运行时执行如下步骤:

获取目标文本中的第一向量和第二向量,其中,所述第一向量用于表示第一产品的实体信息,所述第二向量用于表示所述第一产品的属性信息,所述第一产品为所述目标文本

中记载的产品；

根据第二产品的实体信息对所述第一向量进行更新,得到所述第一向量对应的第一更新向量,其中,所述第二产品为预设文本中记载的产品；

根据第二产品的属性信息对所述第二向量进行更新,得到所述第二向量对应的第二更新向量；

根据所述第一更新向量和所述第二更新向量对所述目标文本进行分析,得到所述目标文本所包括的情感信息。

## 文本的分析方法和装置

### 技术领域

[0001] 本发明涉及信息处理领域,具体而言,涉及一种文本的分析方法和装置。

### 背景技术

[0002] 情感分析(Sentiment analysis),也即倾向性分析,是对带有感情色彩的主观性文本进行分析、处理、归纳和推理的过程,其根据文本所表达的含义和情感信息,将文本分为褒义或贬义这两种或多种类型。

[0003] 情感分析的通常是为了找出说话者或作者对某个话题上的观点的态度,这个态度即为该说话者或作者的情感状态。互联网上产生了大量的用户参与的、对于诸如任务、时间、产品等有价值的评论信息,这些评论信息表达了人们的各种情感色彩和情感倾向,如喜、怒、哀、乐和批评、赞扬,基于此,潜在的用户就可以通过浏览这些具有主观色彩的评论来了解大众舆论对某一事件或产品的看法。

[0004] 在目前的情感分析方法中,各种类型的对象信息已经被成功应用。典型的对象包括实体(例如产品名称)和方面(例如产品属性)。然而在真实的场景下,常常会出现目标对象的训练数据不足情况,在这种情况下可能给对象的表示学习带来困难,进而降低预测性能的情况,这种现象称为“冷启动”问题。

[0005] 针对现有技术中在情感预测模型的训练数据较少的情况下,对文本的情感分析不准确的问题,目前尚未提出有效的解决方案。

### 发明内容

[0006] 本发明实施例提供了一种文本的分析方法和装置,以至少解决现有技术中在情感预测模型的训练数据较少的情况下,对文本的情感分析不准确的技术问题。

[0007] 根据本发明实施例的一个方面,提供了一种文本的分析方法,包括:获取目标文本中的第一向量和第二向量,其中,第一向量用于表示第一产品的实体信息,第二向量用于表示第一产品的属性信息,第一产品为目标文本中记载的产品;根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,其中,第二产品为预设文本中记载的产品;根据第二产品的属性信息对第二向量进行更新,得到第二向量对应的第二更新向量;根据第一更新向量和第二更新向量对目标文本进行分析,得到目标文本所包括的情感信息。

[0008] 根据本发明实施例的另一方面,还提供了一种文本的分析方法,包括:显示获取的目标文本,其中,目标文本包括了第一产品的实体信息和产品的属性信息,第一产品为目标文本中记载的产品;获取目标文本中的第一向量和第二向量,其中,第一向量用于表示第一产品的实体信息,第二向量用于表示第一产品的属性信息;根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,其中,第二产品为预设文本中记载的产品;根据第二产品的属性信息对第二向量进行更新,得到第二向量对应的第二更新向量;根据第一更新向量和第二更新向量对目标文本进行分析,输出目标文本所包括的情感信

息。

[0009] 根据本发明实施例的另一方面,还提供了一种文本的分析方法,包括:获取第一产品的评价信息,其中,评价信息包括了第一产品的实体信息和第一产品的属性信息,第一产品为评价信息所评价的产品;获取评价信息中的第一向量和第二向量,其中,第一向量用于表示第一产品的实体信息,第二向量用于表示第一产品的属性信息;根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,其中,第二产品为预设文本中记载的产品;根据第二产品的属性信息对第二向量进行更新,得到第二向量对应的第二更新向量;根据第一更新向量和第二更新向量对评价信息进行分析,得到评价信息所包括的情感信息。

[0010] 根据本发明实施例的另一方面,还提供了一种存储介质,存储介质包括存储的程序,其中,在程序运行时控制存储介质所在设备执行如下步骤:获取目标文本中的第一向量和第二向量,其中,第一向量用于表示第一产品的实体信息,第二向量用于表示第一产品的属性信息,第一产品为目标文本中记载的产品;根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,其中,第二产品为预设文本中记载的产品;根据第二产品的属性信息对第二向量进行更新,得到第二向量对应的第二更新向量;根据第一更新向量和第二更新向量对目标文本进行分析,得到目标文本所包括的情感信息。

[0011] 根据本发明实施例的另一方面,还提供了一种处理器,处理器用于运行程序,其中,程序运行时执行如下步骤:获取目标文本中的第一向量和第二向量,其中,第一向量用于表示第一产品的实体信息,第二向量用于表示第一产品的属性信息,第一产品为目标文本中记载的产品;根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,其中,第二产品为预设文本中记载的产品;根据第二产品的属性信息对第二向量进行更新,得到第二向量对应的第二更新向量;根据第一更新向量和第二更新向量对目标文本进行分析,得到目标文本所包括的情感信息。

[0012] 在本发明实施例中,获取目标文本中用于表示实体信息的第一向量和用于表示属性信息的第二向量,并基于其他产品的实体信息和属性信息分别对第一向量和第二向量进行更新,最后基于更新的结果对目标文本进行情感分析,得到目标文本的情感分析结果。上述方案考虑了冷启动问题对于情感分析的影响,并提出了可扩展的框架来解决情感分析任务,通过引入其他产品的实体信息和属性信息来增强对目标文本中的实体和属性的表示学习,进而能够在训练数据较少的情况下提高情感分析的准确度,解决了现有技术中在情感预测模型的训练数据较少的情况下,对文本的情感分析不准确的技术问题。

## 附图说明

[0013] 此处所说明的附图用来提供对本发明的进一步理解,构成本申请的一部分,本发明的示意性实施例及其说明用于解释本发明,并不构成对本发明的不当限定。在附图中:

[0014] 图1a示出了一种用于实现文本的分析方法的计算机设备(或移动设备)的硬件结构框图;

[0015] 图1b示出一种对文本进行情感分析的结果框图。

[0016] 图2是根据本申请实施例1的一种文本的分析方法的流程图;

[0017] 图3是根据本申请实施例1的一种可选的情感分析的示意图;

- [0018] 图4是根据本申请实施例2的一种文本的分析方法的流程图；  
[0019] 图5是根据本申请实施例3的一种文本的分析方法的流程图；  
[0020] 图6是根据本申请实施例4的一种文本的分析装置的示意图；  
[0021] 图7是根据本申请实施例5的一种文本的分析装置的示意图；  
[0022] 图8是根据本申请实施例6的一种文本的分析装置的示意图；以及  
[0023] 图9是根据本申请实施例7的一种计算机设备的结构框图。

### 具体实施方式

[0024] 为了使本技术领域的人员更好地理解本发明方案，下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分的实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都应当属于本发明保护的范围。

[0025] 需要说明的是，本发明的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本发明的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

#### [0026] 实施例1

[0027] 根据本发明实施例，还提供了一种文本的分析方法的实施例，需要说明的是，在附图的流程图示出的步骤可以在诸如一组计算机可执行指令的计算机系统中执行，并且，虽然在流程图中示出了逻辑顺序，但是在某些情况下，可以以不同于此处的顺序执行所示出或描述的步骤。

[0028] 本申请实施例一所提供的方法实施例可以在移动终端、计算机设备或者类似的运算装置中执行。图1a示出了一种用于实现文本的分析方法的计算机设备(或移动设备)的硬件结构框图。如图1a所示，计算机设备10(或移动设备10)可以包括一个或多个(图中采用102a、102b，……，102n来示出)处理器102(处理器102可以包括但不限于微处理器MCU或可编程逻辑器件FPGA等的处理装置)、用于存储数据的存储器104、以及用于通信功能的传输模块106。除此以外，还可以包括：显示器、输入/输出接口(I/O接口)、通用总线(BUS)端口(可以作为I/O接口的端口中的一个端口被包括)、网络接口、电源和/或相机。本领域普通技术人员可以理解，图1a所示的结构仅为示意，其并不对上述电子装置的结构造成限定。例如，计算机设备10还可包括比图1a中所示更多或者更少的组件，或者具有与图1a所示不同的配置。

[0029] 应当注意到的是上述一个或多个处理器102和/或其他数据处理电路在本文中通常可以被称为“数据处理电路”。该数据处理电路可以全部或部分的体现为软件、硬件、固件或其他任意组合。此外，数据处理电路可为单个独立的处理模块，或全部或部分的结合到计算机设备10(或移动设备)中的其他元件中的任意一个内。如本申请实施例中所涉及到的，

该数据处理电路作为一种处理器控制(例如与接口连接的可变电阻终端路径的选择)。

[0030] 存储器104可用于存储应用软件的程序以及模块,如本发明实施例中的文本的分析方法对应的程序指令/数据存储装置,处理器102通过运行存储在存储器104内的软件程序以及模块,从而执行各种功能应用以及数据处理,即实现上述的文本的分析方法。存储器104可包括高速随机存储器,还可包括非易失性存储器,如一个或者多个磁性存储装置、闪存、或者其他非易失性固态存储器。在一些实例中,存储器104可进一步包括相对于处理器102远程设置的存储器,这些远程存储器可以通过网络连接至计算机设备10。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0031] 传输模块106用于经由一个网络接收或者发送数据。上述的网络具体实例可包括计算机设备10的通信供应商提供的无线网络。在一个实例中,传输模块106包括一个网络适配器(Network Interface Controller, NIC),其可通过基站与其他网络设备相连从而可与互联网进行通讯。在一个实例中,传输模块106可以为射频(Radio Frequency, RF)模块,其用于通过无线方式与互联网进行通讯。

[0032] 显示器可以例如触摸屏式的液晶显示器(LCD),该液晶显示器可使得用户能够与计算机设备10(或移动设备)的用户界面进行交互。

[0033] 此处需要说明的是,在一些可选实施例中,上述图1a所示的计算机设备(或移动设备)可以包括硬件元件(包括电路)、软件元件(包括存储在计算机可读介质上的计算机代码)、或硬件元件和软件元件两者的结合。应当指出的是,图1a仅为特定具体实例的一个实例,并且旨在示出可存在于上述计算机设备(或移动设备)中的部件的类型。

[0034] 在上述运行环境下,本申请提供了如图2所示的文本的分析方法。图2是根据本申请实施例1的一种文本的分析方法的流程图。

[0035] 步骤S21,获取目标文本中的第一向量和第二向量,其中,第一向量用于表示第一产品的实体信息,第二向量用于表示第一产品的属性信息,第一产品为目标文本中记载的产品。

[0036] 具体的,上述目标文本可以是论坛、购物网站、电商平台中的评论。例如,目标文本可以是购物网站中用户对产品的评价,汽车论坛中用户对汽车的评价等,点评网站中用户对餐厅的评价等。

[0037] 第一产品为目标文本所评价的对象。目标文本具有实体信息和属性信息,实体信息用于表示第一产品的名称或品牌,属性信息用于表示目标文本中第一产品本身或第一产品的品牌在某种属性上的特点。上述第一向量为实体信息所对应的向量,第二向量为属性信息所对应的向量。在一种可选的实施例中,可以在目标文本中确定用于表示实体信息的实体词语和用于表示属性信息的属性词语,通过对实体词语和属性词语进行向量化处理,得到目标文本的第一向量和第二向量。

[0038] 以图1b为例对文本的情感分析进行说明,结合图1b所示,在一种可选的实施例中,以电商平台的场景为例,目标文本可以是平台中用户对实体的评论,一条评论中,评论的主体即可以为该文本的实体,例如,对于评论1“我就喜欢XXX,非常干爽,但是有点贵”,该文本中的实体即为“XXX”,属性信息即为“干爽度”和“价格”。该示例中只有一个实体,还在一些可选的实施例中,目标文本中可以包括多个实体,例如,对于评论2“我就喜欢XXX,非常干爽,但是有点贵。YY不够好,但是便宜”,该文本中的实体即为“XXX”和“YY”,属性信息即为

“干爽度”和“价格”。

[0039] 图3是根据本申请实施例1的一种可选的情感分析的示意图,结合图3所示,可以从实体记忆模块中获取上述第一向量 $v_{e1}$ 、 $v_{e2}$ 、 $v_{e3}$ …… $v_{eN}$ ,并从属性记忆模块中获取上述第二向量 $v_{a1}$ 、 $v_{a2}$ 、 $v_{a3}$ …… $v_{aM}$ 。

[0040] 步骤S23,根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,其中,第二产品为预设文本中记载的产品。

[0041] 具体的,上述预设文本可以为用于作为训练样本的文本,在一种可选的实施例中,可以基于预设文本构建测试数据集,然后基于测试数据集确定第二产品,其中,第二产品可以为测试数据集中,除第一产品之外的其他产品。

[0042] 步骤S25,根据第二产品的属性信息对第二向量进行更新,得到第二向量对应的第二更新向量。

[0043] 上述步骤S23和S25用于根据其他的产品的向量,对目标文本中的第一产品的向量进行更新,从而借助其他产品的向量对第一产品的第一向量和第二向量进行进一步增强,也即向第一向量和第二向量中引入了额外的信息,进而缓解了由于目标文本中的第一产品相关的训练数据不充足所导致的对目标文本的情感分析不准确的问题。

[0044] 在一种可选的实施例中,可以在训练数据集中确定需要额外引入的信息,再通过需要额外引入的信息对第一向量和第二向量进行更新。例如,可以在训练数据集中查找与第一向量相似度高于预设值的实体向量,并在训练数据集中查找与第二向量相似度高于预设值的属性向量,然后将查找到的实体向量叠加在第一向量中,得到第一更新向量,并将查找到的属性向量叠加在第二向量中,得到第二更新向量。在第一向量上叠加查找到的实体向量的方式可以是,根据预设的权重值,将第一向量和查找到的实体向量进行加权,从而得到第一更新向量,在第二向量上叠加查找到的属性向量的方式可以相同。

[0045] 步骤S27,根据第一更新向量和第二更新向量对目标文本进行分析,得到目标文本所包括的情感信息。

[0046] 上述情感信息可以用于表示用户对实体所表示的品牌或产品在某个属性上的态度,从而得到品牌或产品的缺陷,以便商家进行改进。情感信息可以包括正向、负向和中立。在一种可选的实施例中,可以使用分类器softmax来实现情感信息的分析,以分类器softmax为例,softmax输出了文本属于不同情感信息的相对概率,最终确定相对概率最大的情感信息为目标文本所属的情感信息。

[0047] 具体的,可以通过多视角情感分析模型对目标文本进行处理后,将处理结果进行输出,目标文本所包括的情感信息用于表示该目标文本中的产品在预设属性上的情感倾向。

[0048] 仍在图1b的示例中,以目标文本为评论1“我就喜欢XXX,非常干爽,但是有点贵”为例,得到的情感信息,以及情感分析结果包括[XXX,正向]和[XXX,负向],其中,[XXX,正向]用于表示该文本在干爽度这一属性上的评价是正向的,[XXX,负向]用于表示该文本在价格这一属性上的评价是负向的。再例如,以目标文本为评论2“我就喜欢XXX,非常干爽,但是有点贵.YY不够好,但是便宜”为例,得到的情感信息包括[(XXX,干爽度),正向]、[(XXX,价格),负向]、[(YY,干爽度),负向]、[(YY,价格),正向]。

[0049] 由此可见,基于上述对文本的分析方法,可以对目标文本进行多方面的情感分析,

以目标文本为商品的评价信息为例,可以得到用户对商品价格、质量、售后等多方面的情感信息,从而可以指导商家对商品本身进行改进,或对商品的销售策略进行改进,进而推动商家对商品的销售。

[0050] 本申请上述实施例获取目标文本中用于表示实体信息的第一向量和用于表示属性信息的第二向量,并基于其他产品的实体信息和属性信息分别对第一向量和第二向量进行更新,最后基于更新的结果对目标文本进行情感分析,得到目标文本的情感分析结果。上述方案考虑了冷启动问题对于情感分析的影响,并提出了可扩展的框架来解决情感分析任务,通过引入其他产品的实体信息和属性信息来增强对目标文本中的实体和属性的表示学习,进而能够在训练数据较少的情况下提高情感分析的准确度,解决了现有技术中在情感预测模型的训练数据较少的情况下,对文本的情感分析不准确的技术问题。

[0051] 进一步地,在通过引入额外信息来解决“冷启动”这一问题上,还可以引入目标文本所包括的产品的附加属性描述信息等,但这类信息的获取具有一定的难度,即使获取到了这些信息,情感分析模型也难以对这些信息进行泛化和使用。而本申请的上述方案引入的额外信息是其他产品的实体信息或属性信息,因此具有易获取的特点,且也便于模型的使用。

[0052] 作为一种可选的实施例,获取目标文本中的第一向量和第二向量,包括:对目标文本进行分词,得到多个词语;确定多个词语中用于表示实体信息的实体词语和用于表示属性信息的属性词语;对实体词语进行向量化处理得到第一向量;对属性词语进行向量化处理得到第二向量。

[0053] 具体的,上述步骤可以由数据采集和预处理模块执行,数据采集模块直接采集评论作为目标文本,并由预处理模块对目标文本进行预处理。

[0054] 上述分词处理用于将目标文本划分为多个词语,可以通过bigram的分词方式对目标文本进行分词,在对目标文本进行分词之后,还可以进行词根还原,以将具有词形变化的词语还原为词根,然后还可以通过噪音剔除从分词结果中剔除特殊字符、特殊符号等噪声词语。

[0055] 在得到实体词语和属性词语后,可以通过word2vec等方式对实体词语和向量词语进行向量化处理,从而得到上述的第一向量和第二向量。

[0056] 作为一种可选的实施例,根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,包括:根据第二产品的实体信息对第一向量进行迭代更新,得到第一向量对应的第一更新向量;其中,根据第二产品的实体信息对第一向量进行迭代更新,得到第一向量对应的第一更新向量,包括:确定第一向量在此次迭代中对应的输入向量;根据第二产品的实体信息确定第一向量在第一更新向量中的权重;根据第二产品的实体信息确定第一向量在此次迭代中对应的补充向量;根据补充向量和权重对此次迭代中对应的输入向量进行更新,得到第一更新向量。

[0057] 在上述步骤的迭代过程中,每次迭代时的输入向量并不一定相同。在一种可选的实施例中,在首次迭代时,第一向量对应的输入向量可以是第一向量的初始向量,在非首次迭代时,第一向量对应的输入向量可以通过上一次迭代得到的第一更新向量确定。

[0058] 实体信息可以包括第二产品的实体向量的信息,例如,实体向量在训练数据集中出现的次数、频率等。补充向量是其他产品,即第二产品中获取的向量,用来增强第一向量

在“冷启动”时的表现。

[0059] 上述步骤中,最终的第一更新向量可以通过输入向量和补充向量共同得到,在相加过程中,加入加权的处理。由于第一更新向量中仅包括输入向量这一部分和补充向量这一部分,因此当第一向量在更新向量的权重为 $q$ 时,补充向量的权重为 $(1-q)$ 。在一种可选的实施例中,将第一向量对应的输入向量和补充向量根据对应的权重进行加权,即可得到第一更新向量。

[0060] 值得注意的是,上述方案实际是通过注意力机制来寻找与目标文本中的对象(实体或属性)相似的对象集合,并通过相似对象的表示来增强“冷启动”对象的表现学习。上述注意力机制包括如下两层:(1)第一层在确定第一向量在第一更新向量中的权重的过程中实现,第二层在确定补充向量的过程中实现。

[0061] 需要说明的是,在一次迭代结束后,需要判断是否达到预设的迭代次数,如果已经到达了预设的迭代次数,则根据输出的第一更新向量和第二更新向量进行情感预测,否则需要进行下一次迭代。

[0062] 结合图3所示,实体和属性具有共享模块,通过共享模块中的判断模块来判断迭代是否停止,如果停止,则输出更新后得到的第一更新向量和第二更新向量至情感预测模块进行情感预测,否则将此次迭代的结果输入至上下文表示模块,以进行下一次迭代。

[0063] 还需要说明的是,根据第二产品的属性信息对第二向量进行更新与上述根据第二产品的实体信息对第一向量进行更新的方式相同,此处不再赘述。

[0064] 作为一种可选的实施例,确定第一向量在此次迭代中对应的输入向量,包括:根据上一次迭代得到的第一更新向量、上一次迭代得到的第二更新向量和目标文本中的词语向量集合,确定上一次迭代中用于表示目标文本的上下文信息的上下文向量;确定上一次迭代得到的第一更新向量与上一次迭代中的上下文向量之和为第一向量在此次迭代中对应的输入向量。

[0065] 上述步骤通过上一次迭代得到的结果获取此次迭代时,第一向量对应的输入向量。

[0066] 在一种可选的实施例中,上述上下文向量 $c = \text{ContextEncoding}(v_e, v_a, \{w_s\})$ ,其中, $v_e$ 用于表示第一更新向量, $v_a$ 用于表示第二更新向量, $w_s$ 用于表示目标文本中的词语对应的词语向量集合。当前迭代过程中所使用的上下文向量为上一次迭代得到的上下文信息,即 $c^{(l-1)}$ , $c^{(l-1)} = \text{ContextEncoding}(v_e^{(l-1)}, v_a^{(l-1)}, \{w_s\})$ ,其中, $v_e^{(l-1)}$ 用于表示上一次迭代得到的第一更新向量, $v_a^{(l-1)}$ 用于表示上一次迭代得到的第二更新向量。

[0067] 通过上述步骤得到上一次迭代中的上下文信息后,将上一次迭代得到的第一更新向量和上一次迭代中的上下文向量相加,即可得到此次迭代中,第一向量的输入向量。

[0068] 仍在上述实施例中,此次迭代过程中,第一向量对应的输入向量为 $v_e^{(l)} = v_e^{(l-1)} + c^{(l-1)}$ 。

[0069] 结合图3所示,如果不停止循环,则将上一次迭代得到的第一更新向量 $v_e^{(l-1)}$ 和上下文表示模块输出的上一次迭代中的上下文向量 $c^{(l-1)}$ 通过 $\Sigma$ 相加,即可得到此次迭代过程中第一向量对应的输入向量,然后将给输入向量输入至实体更新模块中的融合门机制(图3

中的符号G用于表示该融合门机制)中,以进行第一向量的再一次更新迭代。

[0070] 作为一种可选的实施例,在进行首次迭代时,确定第一向量在此次迭代中对应的输入向量,包括:在目标文本包括多个第一向量的情况下,确定多个第一向量的均值为输入向量。

[0071] 具体的,在上述步骤中,在目标文本包括一个第一向量的情况,将目标文本所包括的一个第一向量作为首次迭代中的输入向量,在目标文本包括多个第一向量的情况下,确定多个第一向量的均值为首次迭代中的输入向量。

[0072] 在一种可选的实施例中,可以将目标文本所包含的词语对应的词向量进行平均来初始化表示。例如,属性词语“食品质量”可通过将分词后获得的词语“食品”和“质量”对应的词向量相加再平均来获得该第一向量对应的输入向量。

[0073] 作为一种可选的实施例,根据第二产品的实体信息确定第一向量在第一更新向量中的权重,包括:获取第一向量在训练数据集中出现的次数;确定训练数据集中的实体向量出现的平均次数;以自然常数为底,以第一向量在训练数据集中出现的次数与平均次数比值为指数,确定第一数值;确定预设值与第一数值的比值为第二数值;确定预设值与第二数值之差为第一向量在第一更新向量中的权重。

[0074] 具体的,上述预设值可以为1。

[0075] 需要说明的是,上述步骤实现了第一层注意力机制,即频率的注意力机制。结合图3所示,可以由实体更新模块中的频率注意力机制来实现。相似的,属性更新模块中也具有作用相同的频率注意力机制。

[0076] 在选择相似对象之前,首先要选择具有准确向量表示的对象。而在训练数据集中频繁出现的对象,通常拥有足够数量的评论数据,因此这些对象通常可以被更准确地表示为 $v_e$ 和 $v_a$ 。为了从实体记忆模块E和属性记忆模块A中尽可能选择这些表示良好的实体向量和属性向量,因此可以借助实体和属性的出现频率这一统计特征来表征E和A中每个 $v_e$ 和 $v_a$ 的不同重要性。

[0077] 上述第二产品表示的是测试数据集中,除目标文本中的第一产品之外的其他产品,在该实施例中,第二产品的实体信息用于表示,第二产品所对应的实体词语的信息,例如,实体词语对应的实体向量出现的次数、频率等。

[0078] 具体的,上述训练数据集是预先构建的。在一种可选的实施例中,在获取了样本数据后,将样本数据划分为两部分,其中一部分作为测试数据集,另一部分作为训练数据集。

[0079] 确定训练数据集中实体向量出现的平均次数,可以首先获取每个实体向量在训练数据集中出现的次数以及训练数据集中实体向量的个数,再使用训练数据集中实体向量出现的总次数除以实体向量的个数,即可得到训练数据集中的实体向量出现的平均次数。

[0080] 在一种可选的实施例中,第二产品的实体信息为第二产品的实体向量在训练数据集中出现的次数。可以通过如下公式表示上述获取第一更新向量中的权重的过程:

$q_e = 1 - 1/e^{\frac{f(e)}{Avg(\{e\})}}$ ,其中, $q_e$ 用于表示第一向量的权重, $f(e)$ 用于表示第一向量的训练数据集中出现的次数, $Avg(\{e\})$ 用于表示训练数据集中实体向量出现的平均次数。

[0081] 仍在上述实施例中,可以以相同的方式得到第二更新向量的权重为 $q_a = 1 - 1/e^{\frac{f(a)}{Avg(\{a\})}}$ 。

[0082] 作为一种可选的实施例,根据第二产品的实体信息确定第一向量在此次迭代中对应的补充向量,包括:获取训练数据集中每个实体向量的相似度参数,其中,相似度参数用于表示实体向量与第一向量的相似程度;获取每个实体向量与对应的相似度参数的第一乘积;确定训练数据集中每个实体向量对应的第一乘积之和为补充向量。

[0083] 上述方案用于实现对象注意力层,相似对象在上下文中的用法常常是相似的,包括相似的上下文词语使用,相似的产品属性和特点等。因此可以从具有良好表示的对象集中自动寻找相似的子集,可以通过注意力机制来自动获得。结合图3所示,该步骤由实体更新模块中的融合门机制G来执行。

[0084] 上述步骤中,获取了每个实体向量与第一向量的相似度参数,并将实体向量与其所对应的相似度参数相乘,再将每个实体向量与其对应的相似度参数相乘的结果进行此相加,得到此次迭代使用的补充向量。在该方案中,实体向量与第一向量的相似度参数可以认为是实体向量的权重,也即,与第一向量相似度越高的实体向量,具有更高的权重,通过该方式,实现了相似度的注意力层。

[0085] 在一种可选的实施例中,假设有了第1-1层的输出向量:实体向量 $v_e^{(l-1)}$ 和上下文向量 $c^{(l-1)}$ ,可以获得1层的实体向量表示 $v_e^{(l)} = v_e^{(l-1)} + c^{(l-1)}$ ,同理有方面向量表示 $v_a^{(l)} = v_a^{(l-1)} + c^{(l-1)}$ 。该基于相似度的注意力机制可以通过如下方式实现:首先,通过softmax函数来计算 $E_i$ 的重要性 $\beta_i$ ,该重要性 $\beta_i$ 即用于表示实体向量与第一向量的相似度参数,然后通过 $\beta_i$ 加权 $E_i$ 即可获得实体补充向量 $v_{ce}^{(l)} = \sum \beta_i * E_i$ 。同理也可以获得方面补充向量 $v_{ca}^{(l)} = \sum \alpha_i * A_i$ 。

[0086] 作为一种可选的实施例,获取训练数据集中每个实体向量的相似度参数,包括:确定训练数据集中的实体向量的权重与实体向量的第二乘积;将第二乘积与此次迭代的输入向量拼接后通过预设的感知器,得到感知结果;基于感知结果确定实体向量对应的相似度参数。

[0087] 在一种可选的实施例中,频率注意力层的输出 $q_i * E_i$ ,并与实体向量 $v_e^{(l)}$ 拼接后通过一个单层的感知器MLP (Multi-Layer Perception,多层感知器);然后通过softmax函数根据感知层的输出结果来确定实体向量 $E_i$ 与第一向量的相似度参数。;

[0088] 作为一种可选的实施例,根据补充向量和权重对此次迭代中对应的输入向量进行更新,得到第一更新向量,包括:获取权重与此次迭代的输入向量的第三乘积;获取预设值与权重的差值,并获取差值与补充向量的第四乘积;确定第三乘积和第四乘积之和为第一更新向量。

[0089] 在上述步骤中,1与第一向量的权重的差值即为补充向量的权重,因此上述步骤将输入向量和补充向量分别与对应的权重进行加权,得到的第一更新向量。

[0090] 在一种可选的实施例中,可以通过如下公式表示第一更新向量: $v_e^{(l)} = q_e * v_e^{(l)} + (1 - q_e) * v_{ce}^{(l)}$ ;其中,1用于表示预设的迭代的次数, $v_e^{(l)}$ 用于表示第一更新向量, $q_e$ 用于表示第一向量的权重,(1- $q_e$ )用于表示补充向量的权重, $v_e^{(l)}$ 用于表示此次迭代的输入向量,

$v_{ce}^{(l)}$ 用于此次迭代的补充向量。

[0091] 同理的,第二更新向量可以通过如下公示表示: $v_a^{(l)} = q_a * v_a^{(l)} + (1 - q_a) * v_{ca}^{(l)}$ ;

其中,1用于表示预设的迭代的次数, $v_a^{(l)}$ 用于表示第二更新向量, $q_a$ 用于表示第二向量的权重,(1- $q_a$ )用于表示补充向量的权重, $v_a^{(l)}$ 用于表示此次迭代的输入向量, $v_{ca}^{(l)}$ 用于表示此次迭代的补充向量。

[0092] 在一种可选的实施例中,上述目标文本至少包括如下一种或多种:用户原创内容、对预设对象的评价信息和对所述预设对象的使用说明。

[0093] 具体的,上述用户原创内容即为UGC(User Generated Content),即为用户生成的内容,例如用户在社交网络上进行的视频分享、短文分享等内容。对用户原创内容进行分析得到其情感信息,有利于应用程序等产品基于每个用户的不同情感信息,为其推送不同的待推送内容,从而使得用户对应用程序等产品的粘性更高。

[0094] 上述对预设对象的评价信息除了可以是论坛、购物网站、电商平台中对商品的评论之外,还可以是用户的测评信息,例如用户对新电子产品的测评等。对商品的测评信息通常较为专业和公正,对其他的大量用户起着指导性作用,因此对商品的测评信息进行分析,具有更大的指导性意义。

[0095] 上述对预设对象的使用说明可以是用户对商品的使用心得,商品的使用心得可以是好物分享类应用程序中的信息,该类使用心得与普通的商品评论相比,篇幅较长,能够包含对商品较为全面的评价,且该类信息通常由在网络中具有一定影响能力的用户所发出,因此对该类信息进行分析,也对商品的销售具有非常大的指导意义。

[0096] 需要说明的是,对于前述的各方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本发明并不受所描述的动作顺序的限制,因为依据本发明,某些步骤可以采用其他顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施例,所涉及的动作和模块并不一定是本发明所必须的。

[0097] 通过以上的实施方式的描述,本领域的技术人员可以清楚地了解到根据上述实施例的方法可借助软件加必需的通用硬件平台的方式来实现,当然也可以通过硬件,但很多情况下前者是更佳的实施方式。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质(如ROM/RAM、磁碟、光盘)中,包括若干指令用以使得一台终端设备(可以是手机,计算机,服务器,或者网络设备)执行本发明各个实施例的方法。

[0098] 实施例2

[0099] 根据本发明实施例,还提供了一种文本的分析方法的实施例,图4是根据本申请实施例2的一种文本的分析方法的流程图,结合图4所示,该方法包括如下步骤:

[0100] 步骤S41,显示获取的目标文本,其中,目标文本包括了第一产品的实体信息和产品的属性信息,第一产品为目标文本中记载的产品。

[0101] 具体的,上述目标文本可以是论坛、购物网站、电商平台中的评论。例如,目标文本可以是购物网站中用户对产品的评价,汽车论坛中用户对汽车的评价等,点评网站中用户

对餐厅的评价等。

[0102] 步骤S43,获取目标文本中的第一向量和第二向量,其中,第一向量用于表示第一产品的实体信息,第二向量用于表示第一产品的属性信息。

[0103] 上述第一产品为目标文本所评价的对象。目标文本具有实体信息和属性信息,实体信息用于表示第一产品的名称或品牌,属性信息用于表示目标文本中第一产品本身或第一产品的品牌在某种属性上的特点。上述第一向量为实体信息所对应的向量,第二向量为属性信息所对应的向量。在一种可选的实施例中,可以在目标文本中确定用于表示实体信息的实体词语和用于表示属性信息的属性词语,通过对实体词语和属性词语进行向量化处理,得到目标文本的第一向量和第二向量。

[0104] 图3是根据本申请实施例的一种可选的情感分析的示意图,结合图3所示,可以从实体记忆模块中获取上述第一向量 $v_{e1}$ 、 $v_{e2}$ 、 $v_{e3}$ …… $v_{eN}$ ,并从属性记忆模块中获取上述第二向量 $v_{a1}$ 、 $v_{a2}$ 、 $v_{a3}$ …… $v_{aN}$ 。

[0105] 步骤S45,根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,其中,第二产品为预设文本中记载的产品。

[0106] 步骤S47,根据第二产品的属性信息对第二向量进行更新,得到第二向量对应的第二更新向量。

[0107] 上述步骤S45和S47用于根据其他的产品的向量,对目标文本中的第一产品的向量进行更新,从而借助其他产品的向量对第一产品的第一向量和第二向量进行进一步增强,也即向第一向量和第二向量中引入了额外的信息,进而缓解了由于目标文本中的第一产品相关的训练数据不充足所导致的对目标文本的情感分析不准确的问题。

[0108] 在一种可选的实施例中,可以在训练数据集中确定需要额外引入的信息,再通过需要额外引入的信息对第一向量和第二向量进行更新。例如,可以在训练数据集中查找与第一向量相似度高于预设值的实体向量,并在训练数据集中查找与第二向量相似度高于预设值的属性向量,然后将查找到的实体向量叠加在第一向量中,得到第一更新向量,并将查找到的属性向量叠加在第二向量中,得到第二更新向量。在第一向量上叠加查找到的实体向量的方式可以是,根据预设的权重值,将第一向量和查找到的实体向量进行加权,从而得到第一更新向量,在第二向量上叠加查找到的属性向量的方式可以相同。

[0109] 步骤S49,根据第一更新向量和第二更新向量对目标文本进行分析,输出目标文本所包括的情感信息。

[0110] 上述情感信息可以用于表示用户对实体所表示的品牌或产品在某个属性上的态度,从而得到品牌或产品的缺陷,以便商家进行改进。情感信息可以包括正向、负向和中立。在一种可选的实施例中,可以使用分类器softmax来实现情感信息的分析,以分类器softmax为例,softmax输出了文本属于不同情感信息的相对概率,最终确定相对概率最大的情感信息为目标文本所属的情感信息。

[0111] 具体的,可以通过多视角情感分析模型对目标文本进行处理后,将处理结果进行输出,目标文本所包括的情感信息用于表示该目标文本中的产品在预设属性上的情感倾向。

[0112] 本申请上述实施例获取目标文本中用于表示实体信息的第一向量和用于表示属性信息的第二向量,并基于其他产品的实体信息和属性信息分别对第一向量和第二向量进

行更新,最后基于更新的结果对目标文本进行情感分析,得到目标文本的情感分析结果。上述方案考虑了冷启动问题对于情感分析的影响,并提出了可扩展的框架来解决情感分析任务,通过引入其他产品的实体信息和属性信息来增强对目标文本中的实体和属性的表示学习,进而能够在训练数据较少的情况下提高情感分析的准确度,解决了现有技术中在情感预测模型的训练数据较少的情况下,对文本的情感分析不准确的技术问题。

[0113] 进一步地,在通过引入额外信息来解决“冷启动”这一问题上,还可以引入目标文本所包括的产品的附加属性描述信息等,但这类信息的获取具有一定的难度,即使获取到了这些信息,情感分析模型也难以对这些信息进行泛化和使用。而本申请的上述方案引入的额外信息是其他产品的实体信息或属性信息,因此具有易获取的特点,且也便于模型的使用。

[0114] 作为一种可选的实施例,根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,包括:根据第二产品的实体信息对第一向量进行迭代更新,得到第一向量对应的第一更新向量;

[0115] 其中,根据第二产品的实体信息对第一向量进行迭代更新,得到第一向量对应的第一更新向量,包括:确定第一向量在此次迭代中对应的输入向量;根据第二产品的实体信息确定第一向量在第一更新向量中的权重;根据第二产品的实体信息确定第一向量在此次迭代中对应的补充向量;根据补充向量和权重对此次迭代中对应的输入向量进行更新,得到第一更新向量。

[0116] 在上述步骤的迭代过程中,每次迭代时的输入向量并不一定相同。在一种可选的实施例中,在首次迭代时,第一向量对应的输入向量可以是第一向量的初始向量,在非首次迭代时,第一向量对应的输入向量可以通过上一次迭代得到的第一更新向量确定。

[0117] 实体信息可以包括第二产品的实体向量的信息,例如,实体向量在训练数据集中出现的次数、频率等。补充向量是其他产品,即第二产品中获取的向量,用来增强第一向量在“冷启动”时的表现。

[0118] 上述步骤中,最终的第一更新向量可以通过输入向量和补充向量共同得到,在相加过程中,加入加权的处理。由于第一更新向量中仅包括输入向量这一部分和补充向量这一部分,因此当第一向量在更新向量的权重为 $q$ 时,补充向量的权重为 $(1-q)$ 。在一种可选的实施例中,将第一向量对应的输入向量和补充向量根据对应的权重进行加权,即可得到第一更新向量。

[0119] 值得注意的是,上述方案实际是通过注意力机制来寻找与目标文本中的对象(实体或属性)相似的对象集合,并通过相似对象的表示来增强“冷启动”对象的表现学习。上述注意力机制包括如下两层:(1)第一层在确定第一向量在第一更新向量中的权重的过程中实现,第二层在确定补充向量的过程中实现。

[0120] 需要说明的是,在一次迭代结束后,需要判断是否达到预设的迭代次数,如果已经到达了预设的迭代次数,则根据输出的第一更新向量和第二更新向量进行情感预测,否则需要进行下一次迭代。

[0121] 结合图3所示,实体和属性具有共享模块,通过共享模块中的判断模块来判断迭代是否停止,如果停止,则输出更新后得到的第一更新向量和第二更新向量至情感预测模块进行情感预测,否则将此次迭代的结果输入至上下文表示模块,以进行下一次迭代。

[0122] 还需要说明的是,根据第二产品的属性信息对第二向量进行更新与上述根据第二产品的实体信息对第一向量进行更新的方式相同,此处不再赘述。

[0123] 本实施例中的文本的分析方法还包括实施例1中的其他步骤,此处也不再赘述。

[0124] 实施例3

[0125] 根据本发明实施例,还提供了一种文本的分析方法的实施例,图5是根据本申请实施例3的一种文本的分析方法的流程图,结合图5所示,该方法包括如下步骤:

[0126] 步骤S51,获取第一产品的评价信息,其中,评价信息包括了第一产品的实体信息和第一产品的属性信息,第一产品为评价信息所评价的产品。

[0127] 具体的,上述评价信息可以是论坛、购物网站、电商平台中的评论。例如,评价信息可以是购物网站中用户对产品的评价,汽车论坛中用户对汽车的评价等,点评网站中用户对餐厅的评价等。

[0128] 步骤S53,获取评价信息中的第一向量和第二向量,其中,第一向量用于表示第一产品的实体信息,第二向量用于表示第一产品的属性信息。

[0129] 上述第一产品为评价信息所评价的对象。评价信息具有实体信息和属性信息,实体信息用于表示第一产品的名称或品牌,属性信息用于表示评价信息中第一产品本身或第一产品的品牌在某种属性上的特点。上述第一向量为实体信息所对应的向量,第二向量为属性信息所对应的向量。在一种可选的实施例中,可以在评价信息中确定用于表示实体信息的实体词语和用于表示属性信息的属性词语,通过对实体词语和属性词语进行向量化处理,得到评价信息的第一向量和第二向量。

[0130] 图3是根据本申请实施例的一种可选的情感分析的示意图,结合图3所示,可以从实体记忆模块中获取上述第一向量 $v_{e1}$ 、 $v_{e2}$ 、 $v_{e3}$ …… $v_{eN}$ ,并从属性记忆模块中获取上述第二向量 $v_{a1}$ 、 $v_{a2}$ 、 $v_{a3}$ …… $v_{aN}$ 。

[0131] 步骤S55,根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,其中,第二产品为预设文本中记载的产品。

[0132] 步骤S57,根据第二产品的属性信息对第二向量进行更新,得到第二向量对应的第二更新向量。

[0133] 上述步骤S55和S57用于根据其他的产品的向量,对评价信息中的第一产品的向量进行更新,从而借助其他产品的向量对第一产品的第一向量和第二向量进行进一步增强,也即向第一向量和第二向量中引入了额外的信息,进而缓解了由于评价信息中的第一产品相关的训练数据不充足所导致的对评价信息的情感分析不准确的问题。

[0134] 在一种可选的实施例中,可以在训练数据集中确定需要额外引入的信息,再通过需要额外引入的信息对第一向量和第二向量进行更新。例如,可以在训练数据集中查找与第一向量相似度高于预设值的实体向量,并在训练数据集中查找与第二向量相似度高于预设值的属性向量,然后将查找到的实体向量叠加在第一向量中,得到第一更新向量,并将查找到的属性向量叠加在第二向量中,得到第二更新向量。在第一向量上叠加查找到的实体向量的方式可以是,根据预设的权重值,将第一向量和查找到的实体向量进行加权,从而得到第一更新向量,在第二向量上叠加查找到的属性向量的方式可以相同。

[0135] 步骤S59,根据第一更新向量和第二更新向量对评价信息进行分析,得到评价信息所包括的情感信息。

[0136] 上述情感信息可以用于表示用户对实体所表示的品牌或产品在某个属性上的态度,从而得到品牌或产品的缺陷,以便商家进行改进。情感信息可以包括正向、负向和中立。在一种可选的实施例中,可以使用分类器softmax来实现情感信息的分析,以分类器softmax为例,softmax输出了文本属于不同情感信息的相对概率,最终确定相对概率最大的情感信息为评价信息所属的情感信息。

[0137] 具体的,可以通过多视角情感分析模型对评价信息进行处理后,将处理结果进行输出,评价信息所包括的情感信息用于表示该评价信息中的产品在预设属性上的情感倾向。

[0138] 本申请上述实施例获取评价信息中用于表示实体信息的第一向量和用于表示属性信息的第二向量,并基于其他产品的实体信息和属性信息分别对第一向量和第二向量进行更新,最后基于更新的结果对评价信息进行情感分析,得到评价信息的情感分析结果。上述方案考虑了冷启动问题对于情感分析的影响,并提出了可扩展的框架来解决情感分析任务,通过引入其他产品的实体信息和属性信息来增强对评价信息中的实体和属性的表示学习,进而能够在训练数据较少的情况下提高情感分析的准确度,解决了现有技术中在情感预测模型的训练数据较少的情况下,对文本的情感分析不准确的技术问题。

[0139] 进一步地,在通过引入额外信息来解决“冷启动”这一问题上,还可以引入评价信息所包括的产品的附加属性描述信息等,但这类信息的获取具有一定的难度,即使获取到了这些信息,情感分析模型也难以对这些信息进行泛化和使用。而本申请的上述方案引入的额外信息是其他产品的实体信息或属性信息,因此具有易获取的特点,且也便于模型的使用。

[0140] 作为一种可选的实施例,根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,包括:根据第二产品的实体信息对第一向量进行迭代更新,得到第一向量对应的第一更新向量;

[0141] 其中,根据第二产品的实体信息对第一向量进行迭代更新,得到第一向量对应的第一更新向量,包括:确定第一向量在此次迭代中对应的输入向量;根据第二产品的实体信息确定第一向量在第一更新向量中的权重;根据第二产品的实体信息确定第一向量在此次迭代中对应的补充向量;根据补充向量和权重对此次迭代中对应的输入向量进行更新,得到第一更新向量。

[0142] 在上述步骤的迭代过程中,每次迭代时的输入向量并不一定相同。在一种可选的实施例中,在首次迭代时,第一向量对应的输入向量可以是第一向量的初始向量,在非首次迭代时,第一向量对应的输入向量可以通过上一次迭代得到的第一更新向量确定。

[0143] 实体信息可以包括第二产品的实体向量的信息,例如,实体向量在训练数据集中出现的次数、频率等。补充向量是其他产品,即第二产品中获取的向量,用来增强第一向量在“冷启动”时的表现。

[0144] 上述步骤中,最终的第一更新向量可以通过输入向量和补充向量共同得到,在相加过程中,加入加权的处理。由于第一更新向量中仅包括输入向量这一部分和补充向量这一部分,因此当第一向量在更新向量的权重为 $q$ 时,补充向量的权重为 $(1-q)$ 。在一种可选的实施例中,将第一向量对应的输入向量和补充向量根据对应的权重进行加权,即可得到第一更新向量。

[0145] 值得注意的是,上述方案实际是通过注意力机制来寻找与评价信息中的对象(实体或属性)相似的对象集合,并通过相似对象的表示来增强“冷启动”对象的表现学习。上述注意力机制包括如下两层:(1)第一层在确定第一向量在第一更新向量中的权重的过程中实现,第二层在确定补充向量的过程中实现。

[0146] 需要说明的是,在一次迭代结束后,需要判断是否达到预设的迭代次数,如果已经到达了预设的迭代次数,则根据输出的第一更新向量和第二更新向量进行情感预测,否则需要进行下一次迭代。

[0147] 结合图3所示,实体和属性具有共享模块,通过共享模块中的判断模块来判断迭代是否停止,如果停止,则输出更新后得到的第一更新向量和第二更新向量至情感预测模块进行情感预测,否则将此次迭代的结果输入至上下文表示模块,以进行下一次迭代。

[0148] 还需要说明的是,根据第二产品的属性信息对第二向量进行更新与上述根据第二产品的实体信息对第一向量进行更新的方式相同,此处不再赘述。

[0149] 本实施例中的文本的分析方法还包括实施例1中的其他步骤,此处也不再赘述。

[0150] 实施例4

[0151] 根据本发明实施例,还提供了一种用于实施上述实施例1中的文本的分析方法的文本的分析装置,图6是根据本申请实施例4的一种文本的分析装置的示意图,如图6所示,该装置600包括:

[0152] 获取模块602,用于获取目标文本中的第一向量和第二向量,其中,第一向量用于表示第一产品的实体信息,第二向量用于表示第一产品的属性信息,第一产品为目标文本中记载的产品。

[0153] 第一更新模块604,用于根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,其中,第二产品为预设文本中记载的产品。

[0154] 第二更新模块606,用于根据第二产品的属性信息对第二向量进行更新,得到第二向量对应的第二更新向量。

[0155] 分析模块608,用于根据第一更新向量和第二更新向量对目标文本进行分析,得到目标文本所包括的情感信息。

[0156] 此处需要说明的是,上述获取模块602、第一更新模块604、第二更新模块606和分析模块608对应于实施例1中的步骤S21至步骤S27,四个模块与对应的步骤所实现的实例和应用场景相同,但不限于上述实施例一所公开的内容。需要说明的是,上述模块作为装置的一部分可以运行在实施例一提供的计算机设备10中。

[0157] 作为一种可选的实施例,获取模块包括:分词子模块,用于对目标文本进行分词,得到多个词语;确定子模块,用于确定多个词语中用于表示实体信息的实体词语和用于表示属性信息的属性词语;第一处理子模块,用于对实体词语进行向量化处理得到第一向量;第二处理子模块,用于对属性词语进行向量化处理得到第二向量。

[0158] 作为一种可选的实施例,第一更新模块包括:第一更新子模块,用于根据第二产品的实体信息对第一向量进行迭代更新,得到第一向量对应的第一更新向量;

[0159] 其中,第一更新子模块包括:第一确定单元,用于确定第一向量在此次迭代中对应的输入向量;第二确定单元,用于根据第二产品的实体信息确定第一向量在第一更新向量中的权重;第三确定单元,用于根据第二产品的实体信息确定第一向量在此次迭代中对应

的补充向量;更新单元,用于根据补充向量和权重对此次迭代中对应的输入向量进行更新,得到第一更新向量。

[0160] 作为一种可选的实施例,第一确定单元包括:第六确定子单元,用于根据上一次迭代得到的第一更新向量、上一次迭代得到的第二更新向量和目标文本中的词语向量集合,确定上一次迭代中用于表示目标文本的上下文信息的上下文向量;第七确定子单元,用于确定上一次迭代得到的第一更新向量与上一次迭代中的上下文向量之和为第一向量在此次迭代中对应的输入向量。

[0161] 作为一种可选的实施例,在进行首次迭代时,第一确定单元包括:第八确定子单元,用于在目标文本包括多个第一向量的情况下,确定多个第一向量的均值为输入向量。

[0162] 作为一种可选的实施例,第二确定单元包括:第一获取子单元,用于获取第一向量在训练数据集中出现的次数;第一确定子单元,用于确定训练数据集中的实体向量出现的平均次数;第二确定子单元,用于以自然常数为底,以第一向量在训练数据集中出现的次数与平均次数比值为指数,确定第一数值;第三确定子单元,用于确定预设值与第一数值的比值为第二数值;第四确定子单元,用于确定预设值与第二数值之差为第一向量在第一更新向量中的权重。

[0163] 作为一种可选的实施例,第三确定单元包括:第二获取子单元,用于获取训练数据集中每个实体向量的相似度参数,其中,相似度参数用于表示实体向量与第一向量的相似程度;第三获取子单元,用于获取每个实体向量与对应的相似度参数的第一乘积;第五确定子单元,用于确定训练数据集中每个实体向量对应的第一乘积之和为补充向量。

[0164] 作为一种可选的实施例,第二获取子单元包括:第八确定子单元,用于确定训练数据集中的实体向量的权重与实体向量的第二乘积;感知子单元,用于将第二乘积与此次迭代的输入向量拼接后通过预设的感知器,得到感知结果;第九确定子单元,用于基于感知结果确定实体向量对应的相似度参数。

[0165] 作为一种可选的实施例,更新单元包括:第四获取子单元,用于获取权重与此次迭代的输入向量的第三乘积;第五获取子单元,用于获取预设值与权重的差值,并获取差值与补充向量的第四乘积;第十确定子单元,用于确定第三乘积和第四乘积之和为第一更新向量。

[0166] 作为一种可选的实施例,目标文本至少包括如下一种或多种:用户原创内容、对预设对象的评价信息和对预设对象的使用说明。

[0167] 实施例5

[0168] 根据本发明实施例,还提供了一种用于实施上述实施例2中的文本的分析方法的文本的分析装置,图7是根据本申请实施例5的一种文本的分析装置的示意图,如图7所示,该装置700包括:

[0169] 显示模块702,用于显示获取的目标文本,其中,目标文本包括了第一产品的实体信息和产品的属性信息,第一产品为目标文本中记载的产品。

[0170] 获取模块704,用于获取目标文本中的第一向量和第二向量,其中,第一向量用于表示第一产品的实体信息,第二向量用于表示第一产品的属性信息。

[0171] 第一更新模块706,用于根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,其中,第二产品为预设文本中记载的产品。

[0172] 第二更新模块708,用于根据第二产品的属性信息对第二向量进行更新,得到第二向量对应的第二更新向量。

[0173] 分析模块7010,用于根据第一更新向量和第二更新向量对目标文本进行分析,输出目标文本所包括的情感信息。

[0174] 此处需要说明的是,上述显示模块702、获取模块704、第一更新模块706、第二更新模块708和第二更新模块708对应于实施例2中的步骤S41至步骤S49,五个模块与对应的步骤所实现的实例和应用场景相同,但不限于上述实施例一所公开的内容。需要说明的是,上述模块作为装置的一部分可以运行在实施例一提供的计算机设备10中。

[0175] 作为一种可选的实施例,第一更新模块包括:更新迭代模块,用于根据第二产品的实体信息对第一向量进行迭代更新,得到第一向量对应的第一更新向量;

[0176] 更新迭代模块包括:第一确定子模块,用于确定第一向量在此次迭代中对应的输入向量;第二确定子模块,用于根据第二产品的实体信息确定第一向量在第一更新向量中的权重;第三确定子模块,用于根据第二产品的实体信息确定第一向量在此次迭代中对应的补充向量;更新子模块,用于根据补充向量和权重对此次迭代中对应的输入向量进行更新,得到第一更新向量。

[0177] 实施例6

[0178] 根据本发明实施例,还提供了一种用于实施上述实施例3中的文本的分析方法的文本的分析装置,图8是根据本申请实施例6的一种文本的分析装置的示意图,如图8所示,该装置800包括:

[0179] 第一获取模块802,用于获取第一产品的评价信息,其中,评价信息包括了第一产品的实体信息和第一产品的属性信息,第一产品为评价信息所评价的产品。

[0180] 第二获取模块804,用于获取评价信息中的第一向量和第二向量,其中,第一向量用于表示第一产品的实体信息,第二向量用于表示第一产品的属性信息。

[0181] 第一更新模块806,用于根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,其中,第二产品为预设文本中记载的产品。

[0182] 第二更新模块808,用于根据第二产品的属性信息对第二向量进行更新,得到第二向量对应的第二更新向量。

[0183] 分析模块8010,用于根据第一更新向量和第二更新向量对评价信息进行分析,得到评价信息所包括的情感信息。

[0184] 作为一种可选的实施例,第一更新模块包括:更新迭代模块,用于根据第二产品的实体信息对第一向量进行迭代更新,得到第一向量对应的第一更新向量;

[0185] 更新迭代模块包括:第一确定子模块,用于确定第一向量在此次迭代中对应的输入向量;第二确定子模块,用于根据第二产品的实体信息确定第一向量在第一更新向量中的权重;第三确定子模块,用于根据第二产品的实体信息确定第一向量在此次迭代中对应的补充向量;更新子模块,用于根据补充向量和权重对此次迭代中对应的输入向量进行更新,得到第一更新向量。

[0186] 实施例7

[0187] 本发明的实施例可以提供一种计算机设备,该计算机设备可以是计算机设备群中的任意一个计算机设备。可选地,在本实施例中,上述计算机设备也可以替换为移动终端等

终端设备。

[0188] 可选地,在本实施例中,上述计算机设备可以位于计算机网络的多个网络设备中的至少一个网络设备。

[0189] 在本实施例中,上述计算机设备可以执行文本的分析方法中以下步骤的程序代码:获取目标文本中的第一向量和第二向量,其中,第一向量用于表示第一产品的实体信息,第二向量用于表示第一产品的属性信息,第一产品为目标文本中记载的产品;根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,其中,第二产品为预设文本中记载的产品;根据第二产品的属性信息对第二向量进行更新,得到第二向量对应的第二更新向量;根据第一更新向量和第二更新向量对目标文本进行分析,得到目标文本所包括的情感信息。

[0190] 可选地,图9是根据本申请实施例7的一种计算机设备的结构框图。如图9示,该计算机设备A可以包括:一个或多个(图中仅示出一个)处理器902、存储器904、以及外设接口906。

[0191] 其中,存储器可用于存储软件程序以及模块,如本发明实施例中的文本的分析方法和装置对应的程序指令/模块,处理器通过运行存储在存储器内的软件程序以及模块,从而执行各种功能应用以及数据处理,即实现上述的文本的分析方法。存储器可包括高速随机存储器,还可以包括非易失性存储器,如一个或者多个磁性存储装置、闪存、或者其他非易失性固态存储器。在一些实例中,存储器可进一步包括相对于处理器远程设置的存储器,这些远程存储器可以通过网络连接至终端A。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

[0192] 处理器可以通过传输装置调用存储器存储的信息及应用程序,以执行下述步骤:获取目标文本中的第一向量和第二向量,其中,第一向量用于表示第一产品的实体信息,第二向量用于表示第一产品的属性信息,第一产品为目标文本中记载的产品;根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,其中,第二产品为预设文本中记载的产品;根据第二产品的属性信息对第二向量进行更新,得到第二向量对应的第二更新向量;根据第一更新向量和第二更新向量对目标文本进行分析,得到目标文本所包括的情感信息。

[0193] 可选的,上述处理器还可以执行如下步骤的程序代码:对目标文本进行分词,得到多个词语;确定多个词语中用于表示实体信息的实体词语和用于表示属性信息的属性词语;对实体词语进行向量化处理得到第一向量;对属性词语进行向量化处理得到第二向量。

[0194] 可选的,上述处理器还可以执行如下步骤的程序代码:确定第一向量在此次迭代中对应的输入向量;根据第二产品的实体信息确定第一向量在第一更新向量中的权重;根据第二产品的实体信息确定第一向量在此次迭代中对应的补充向量;根据补充向量和权重对此次迭代中对应的输入向量进行更新,得到第一更新向量。

[0195] 可选的,上述处理器还可以执行如下步骤的程序代码:根据上一次迭代得到的第一更新向量、上一次迭代得到的第二更新向量和目标文本中的词语向量集合,确定上一次迭代中用于表示目标文本的上下文信息的上下文向量;确定上一次迭代得到的第一更新向量与上一次迭代中的上下文向量之和为第一向量在此次迭代中对应的输入向量。

[0196] 可选的,上述处理器还可以执行如下步骤的程序代码:在进行首次迭代时,在目标

文本包括多个第一向量的情况下,确定多个第一向量的均值为输入向量。

[0197] 可选的,上述处理器还可以执行如下步骤的程序代码:获取第一向量在训练数据集中出现的次数;确定训练数据集中的实体向量出现的平均次数;以自然常数为底,以第一向量在训练数据集中出现的次数与平均次数比值为指数,确定第一数值;确定预设值与第一数值的比值为第二数值;确定预设值与第二数值之差为第一向量在第一更新向量中的权重。

[0198] 可选的,上述处理器还可以执行如下步骤的程序代码:获取训练数据集中每个实体向量的相似度参数,其中,相似度参数用于表示实体向量与第一向量的相似程度;获取每个实体向量与对应的相似度参数的第一乘积;确定训练数据集中每个实体向量对应的第一乘积之和为补充向量。

[0199] 可选的,上述处理器还可以执行如下步骤的程序代码:确定训练数据集中的实体向量的权重与实体向量的第二乘积;将第二乘积与此次迭代的输入向量拼接后通过预设的感知器,得到感知结果;基于感知结果确定实体向量对应的相似度参数。

[0200] 可选的,上述处理器还可以执行如下步骤的程序代码:获取权重与此次迭代的输入向量的第三乘积;获取预设值与权重的差值,并获取差值与补充向量的第四乘积;确定第三乘积和第四乘积之和为第一更新向量。

[0201] 可选的,目标文本至少包括如下一一种或多种:用户原创内容、对预设对象的评价信息和对预设对象的使用说明。

[0202] 采用本发明实施例,提供了一种文本的情感分析的方案。通过获取目标文本中用于表示实体信息的第一向量和用于表示属性信息的第二向量,并基于其他产品的实体信息和属性信息分别对第一向量和第二向量进行更新,最后基于更新的结果对目标文本进行情感分析,得到目标文本的情感分析结果。上述方案考虑了冷启动问题对于情感分析的影响,并提出了可扩展的框架来解决情感分析任务,通过引入其他产品的实体信息和属性信息来增强对目标文本中的实体和属性的表示学习,进而能够在训练数据较少的情况下提高情感分析的准确度,解决了现有技术中在情感预测模型的训练数据较少的情况下,对文本的情感分析不准确的技术问题。

[0203] 本领域普通技术人员可以理解,图9所示的结构仅为示意,计算机设备也可以是智能手机(如Android手机、iOS手机等)、平板电脑、掌上电脑以及移动互联网设备(Mobile Internet Devices, MID)、PAD等终端设备。图9其并不对上述电子装置的结构造成限定。例如,计算机设备90还可包括比图9中所示更多或者更少的组件(如网络接口、显示装置等),或者具有与图9所示不同的配置。

[0204] 本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通程序来指令终端设备相关的硬件来完成,该程序可以存储于一计算机可读存储介质中,存储介质可以包括:闪存盘、只读存储器(Read-Only Memory, ROM)、随机存取器(Random Access Memory, RAM)、磁盘或光盘等。

[0205] 实施例8

[0206] 本发明的实施例还提供了一种存储介质。可选地,在本实施例中,上述存储介质可以用于保存上述实施例一所提供的文本的分析方法所执行的程序代码。

[0207] 可选地,在本实施例中,上述存储介质可以位于计算机网络中计算机设备群中的

任意一个计算机设备中,或者位于移动终端群中的任意一个移动终端中。

[0208] 可选地,在本实施例中,存储介质被设置为存储用于执行以下步骤的程序代码:获取目标文本中的第一向量和第二向量,其中,第一向量用于表示第一产品的实体信息,第二向量用于表示第一产品的属性信息,第一产品为目标文本中记载的产品;根据第二产品的实体信息对第一向量进行更新,得到第一向量对应的第一更新向量,其中,第二产品为预设文本中记载的产品;根据第二产品的属性信息对第二向量进行更新,得到第二向量对应的第二更新向量;根据第一更新向量和第二更新向量对目标文本进行分析,得到目标文本所包括的情感信息。

[0209] 上述本发明实施例序号仅仅为了描述,不代表实施例的优劣。

[0210] 在本发明的上述实施例中,对各个实施例的描述都各有侧重,某个实施例中沒有详述的部分,可以参见其他实施例的相关描述。

[0211] 在本申请所提供的几个实施例中,应该理解到,所揭露的技术内容,可通过其它的方式实现。其中,以上所描述的装置实施例仅仅是示意性的,例如所述单元的划分,仅仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如多个单元或组件可以结合或者可以集成到另一个系统,或一些特征可以忽略,或不执行。另一点,所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口,单元或模块的间接耦合或通信连接,可以是电性或其它的形式。

[0212] 所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

[0213] 另外,在本发明各个实施例中的各功能单元可以集成在一个处理单元中,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单元中。上述集成的单元既可以采用硬件的形式实现,也可以采用软件功能单元的形式实现。

[0214] 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用时,可以存储在一个计算机可读取存储介质中。基于这样的理解,本发明的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干指令用以使得一台计算机设备(可为个人计算机、服务器或者网络设备)执行本发明各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

[0215] 以上所述仅是本发明的优选实施方式,应当指出,对于本技术领域的普通技术人员来说,在不脱离本发明原理的前提下,还可以做出若干改进和润饰,这些改进和润饰也应视为本发明的保护范围。

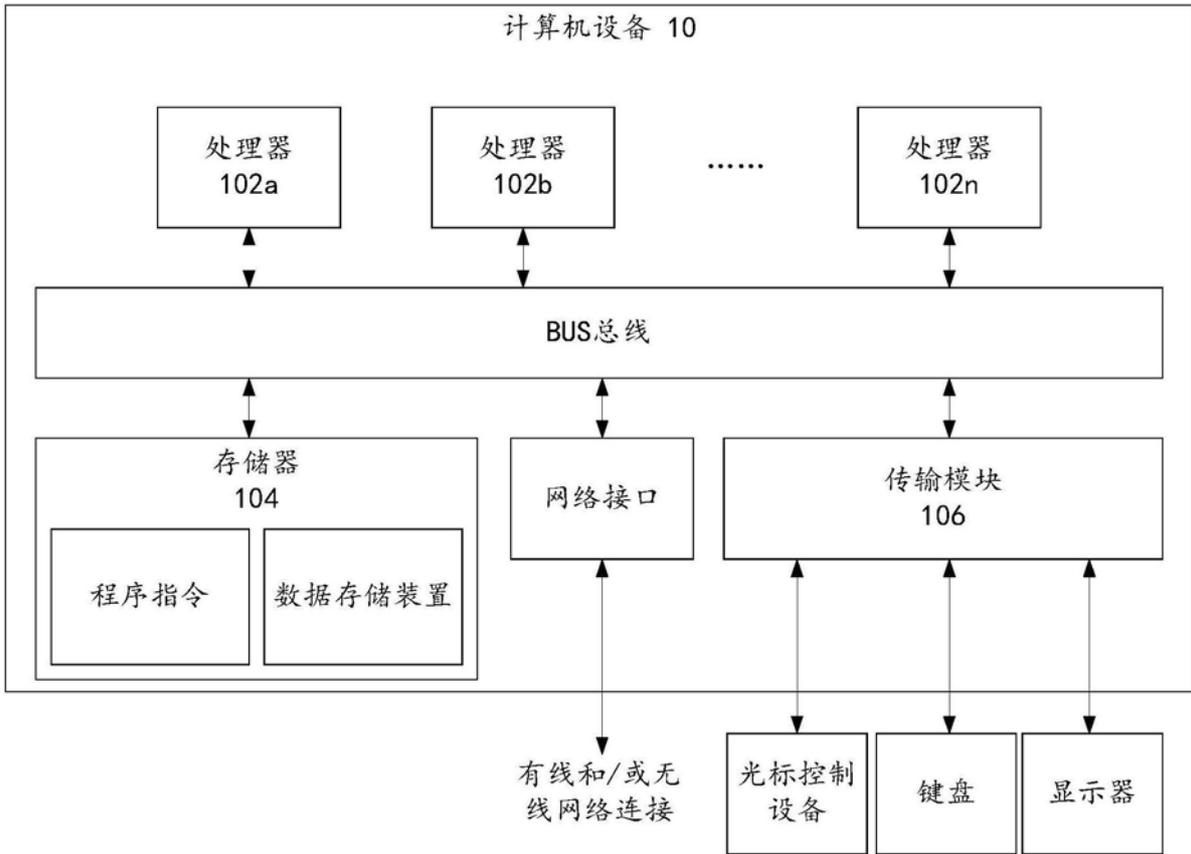


图1a

基于属性的情感分析	多实体基于属性的情感分析
<p><b>评论 1:</b> “我就喜欢 XXX，非常干爽，但是有点贵。”</p> <p><b>评论 1 分析结果:</b> 实体: XXX 属性: 干爽度、价格</p> <p><b>情感分析结果:</b> [XXX, 正向]、[XXX, 负向]</p>	<p><b>评论 2:</b> “我就喜欢 XXX，非常干爽，但是有点贵。YY 不够好，但是便宜。”</p> <p><b>评论 2 分析结果:</b> 实体: XXX、YY 属性: 干爽度、价格</p> <p><b>情感分析结果:</b> [(XXX, 干爽度), 正向]、[(XXX, 价格), 负向]、 [(YY, 干爽度), 负向]、[(YY, 价格), 正向]</p>

图1b

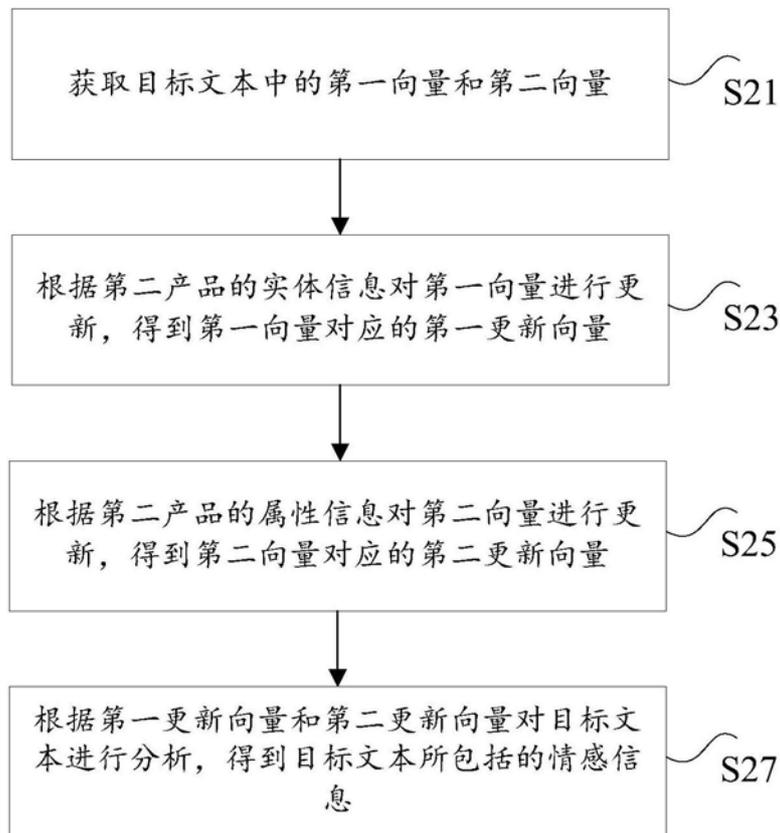


图2

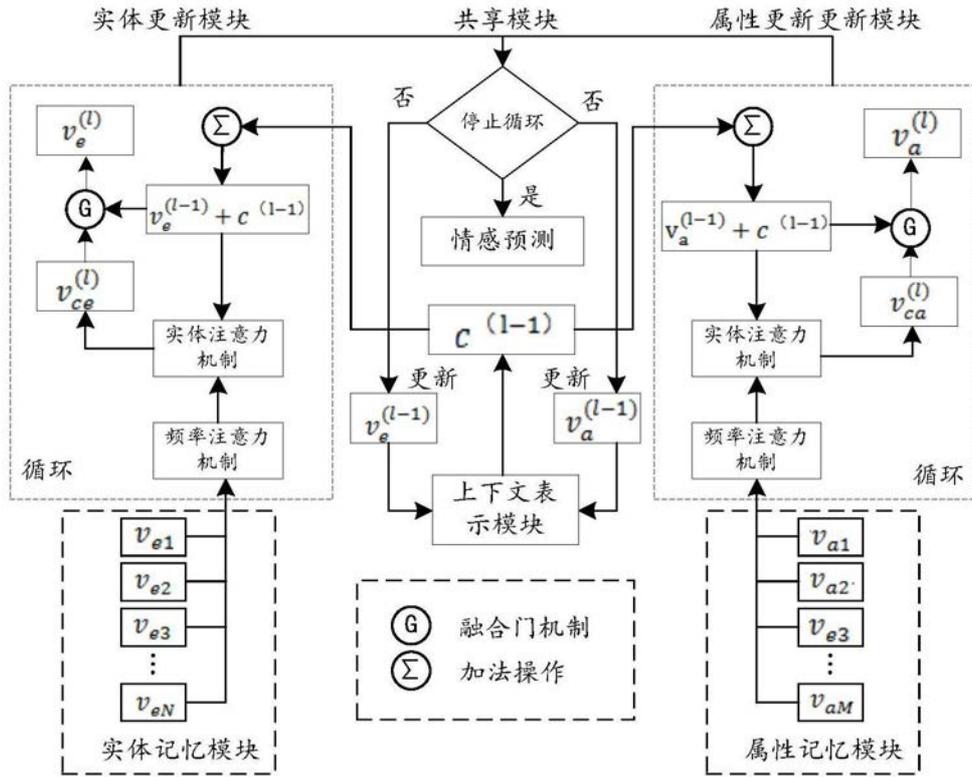


图3

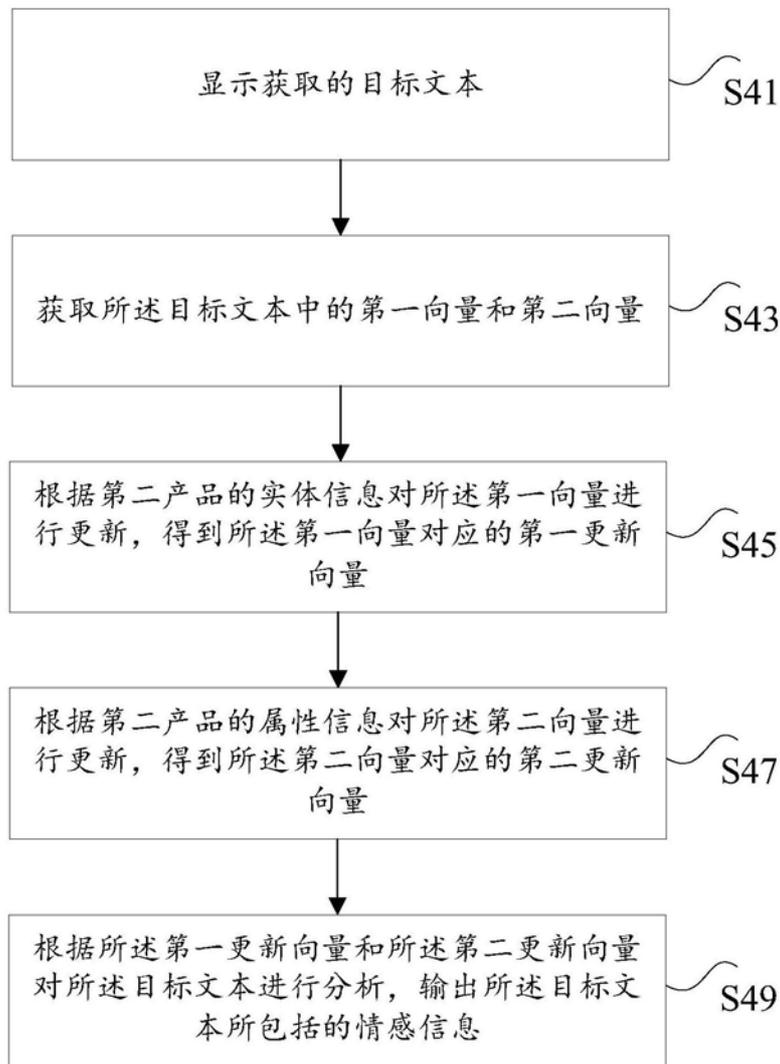


图4

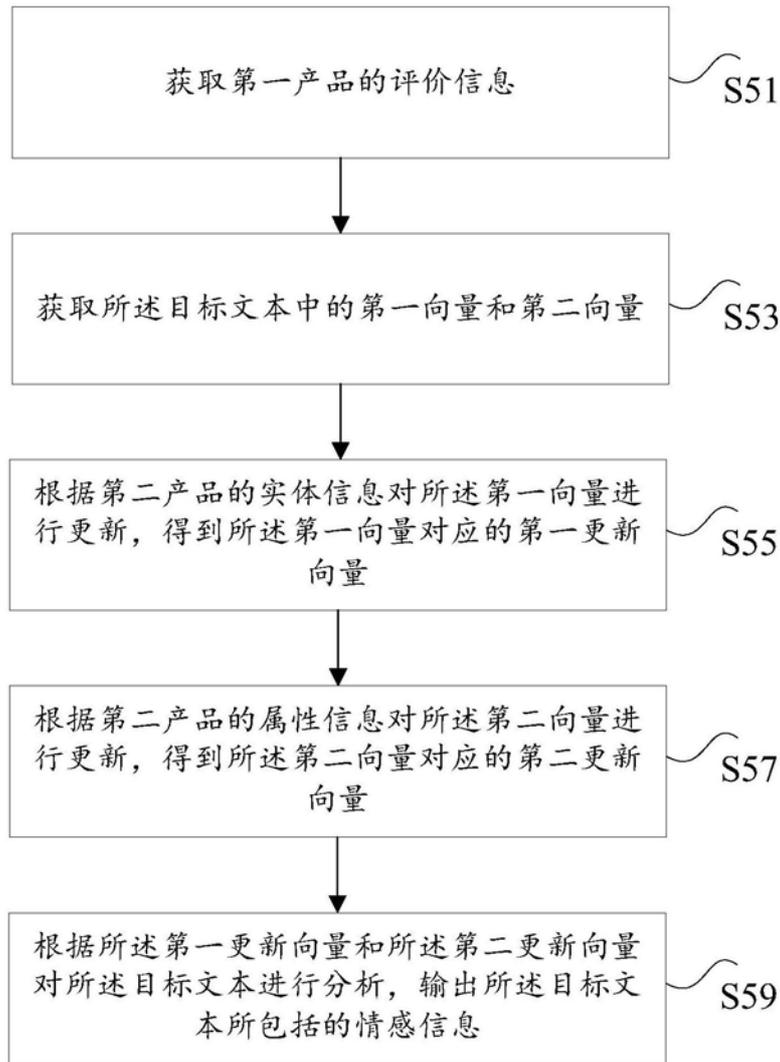


图5

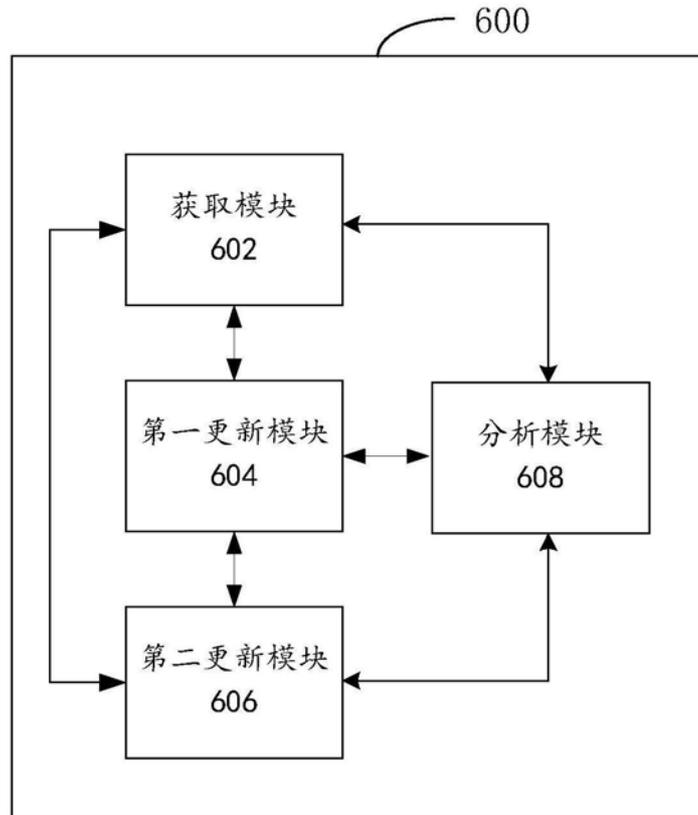


图6

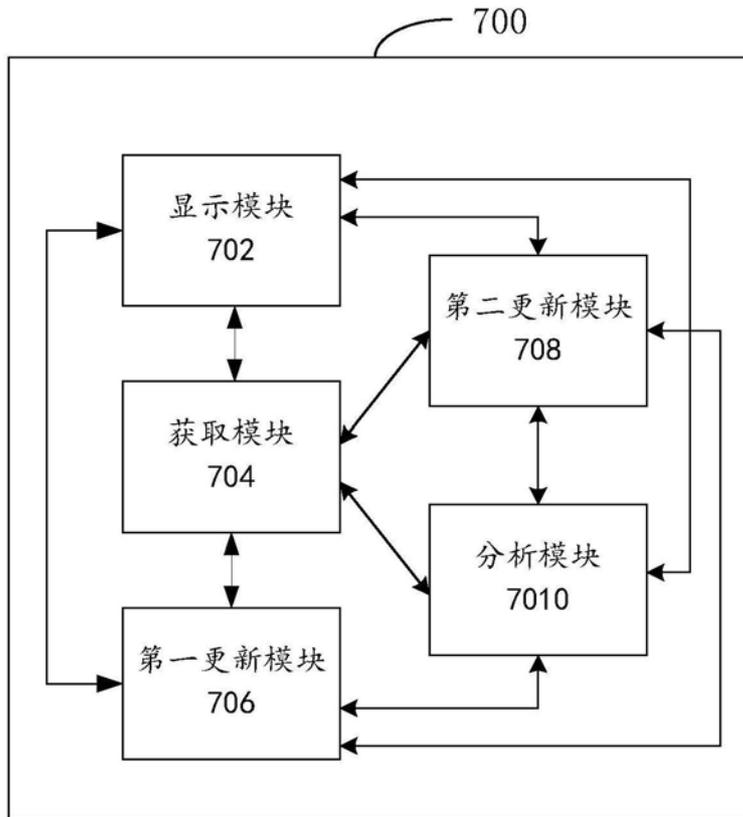


图7

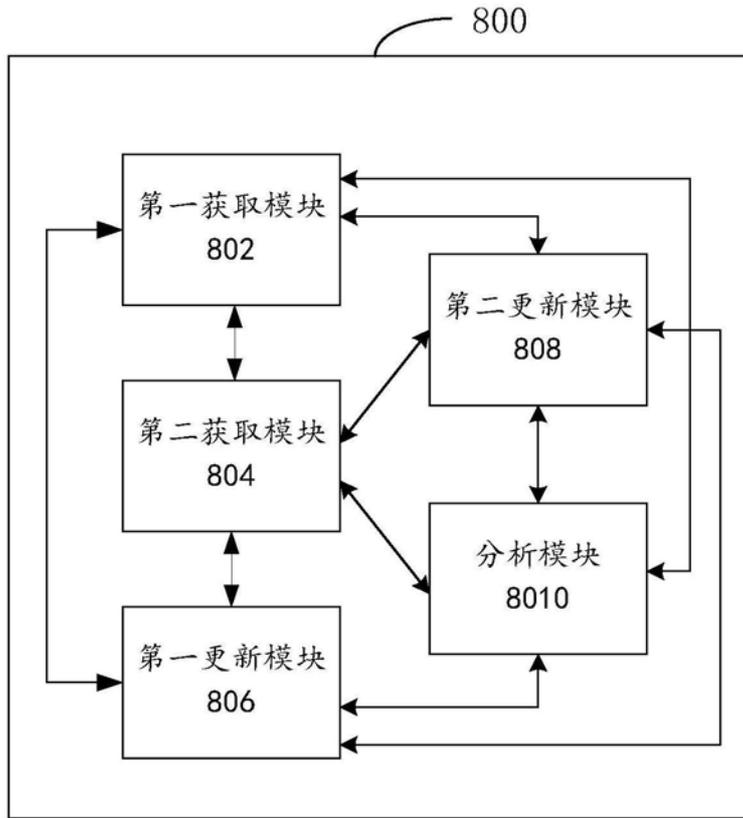


图8

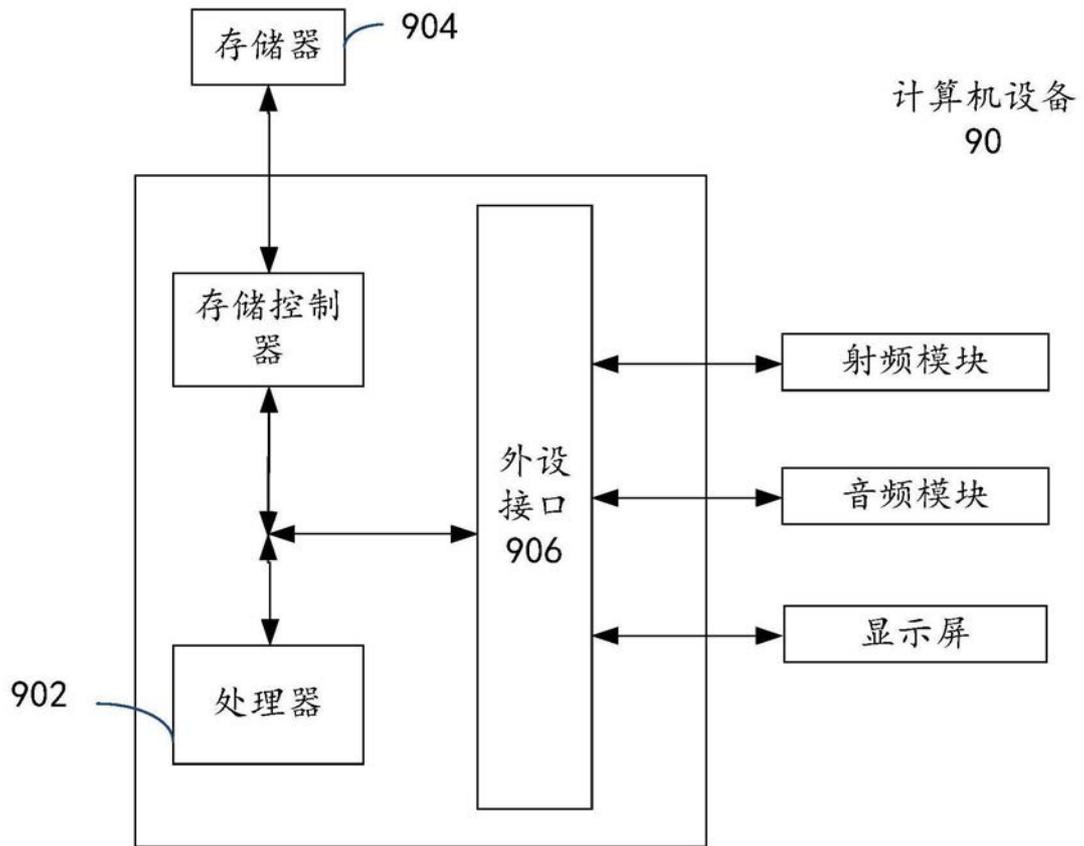


图9