(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 114169418 B
(45)授权公告日 2023.12.01

(21)申请号 202111446672.1

(22)申请日 2021.11.30

(65)同一申请的已公布的文献号
申请公布号 CN 114169418 A

(43)申请公布日 2022.03.11

(73)专利权人 北京百度网讯科技有限公司
地址 100085 北京市海淀区上地十街10号百度大厦2层

(72)发明人 骆金昌 王海威 步君昭 陈坤斌 和为

(74)专利代理机构 北京市通商律师事务所 11951
专利代理师 姜莹丽

(51)Int.Cl.
G06F 18/214(2023.01)
G06F 18/241(2023.01)
G06F 16/35(2019.01)
G06F 40/205(2020.01)
G06F 40/30(2020.01)
G06N 3/042(2023.01)
G06N 3/0464(2023.01)
G06N 3/08(2023.01)

(56)对比文件
CN 108268556 A,2018.07.10
CN 102867016 A,2013.01.09
CN 104021233 A,2014.09.03
CN 105260410 A,2016.01.20
CN 110162690 A,2019.08.23
CN 112559749 A,2021.03.26
CN 113434762 A,2021.09.24
CN 113590854 A,2021.11.02
CN 113626704 A,2021.11.09
US 2012084657 A1,2012.04.05
US 2021027146 A1,2021.01.28
US 2021150565 A1,2021.05.20
CN 108804526 A,2018.11.13
屠守中,闫洲,卫玲蔚,朱小燕.异构社交网络用户兴趣挖掘方法.《西安电子科技大学学报》,异构社交网络用户兴趣挖掘方法.2018,正文第83-88段.

审查员 赵迪

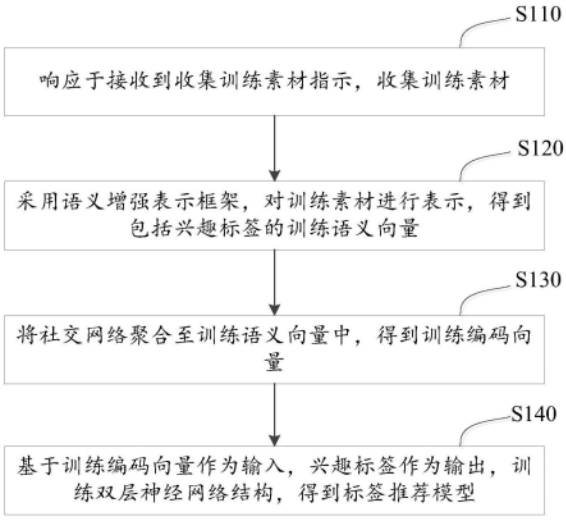
权利要求书3页 说明书11页 附图6页

(54)发明名称

标签推荐模型训练方法及装置、标签获取方法及装置

(57)摘要

本公开提供了一种标签推荐模型训练方法及装置、标签获取方法及装置,涉及数据处理技术领域,尤其涉及深度学习、云服务、内容搜索等技术领域,具体实现方案为:响应于接收到收集训练素材指示,收集训练素材,所述训练素材包含兴趣标签;采用语义增强表示框架,对所述训练素材的特征进行表示,得到包括所述兴趣标签的训练语义向量;将社交网络聚合至所述训练语义向量中,得到训练编码向量;基于所述训练编码向量作为输入,所述兴趣标签作为输出,训练双层神经网络结构,得到标签推荐模型。通过本公开获取的兴趣标签更加精确。



CN 114169418 B

1.一种标签推荐模型训练方法,所述方法包括:

响应于接收到收集训练素材指示,收集训练素材,所述训练素材包含兴趣标签;

采用语义增强表示框架,对所述训练素材的特征进行表示,得到包括所述兴趣标签的训练语义向量;

将社交网络聚合至所述训练语义向量中,得到训练编码向量;

基于所述训练编码向量作为输入,所述兴趣标签作为输出,训练双层神经网络结构,得到标签推荐模型;

其中,所述训练素材包括行为训练素材和业务训练素材;所述采用语义增强表示框架,对所述训练素材的特征进行表示,得到包括兴趣标签的训练语义向量,包括:

基于所述语义增强表示框架,将所述行为训练素材表示为不同长度的训练行为向量,将所述业务训练素材表示为固定长度的训练业务向量;将所述训练行为向量求平均之后,与所述训练业务向量进行融合,得到训练语义向量;

所述将社交网络聚合至所述训练语义向量中,得到训练编码向量,包括:

获取社交网络,并确定社交网络之间的亲密值;将所述亲密值作为矩阵中元素的取值,构建邻接矩阵;在所述邻接矩阵中每行所述元素的权重和为一的条件下,为所述元素分配权重,所述邻接矩阵中对角线元素分配的权重大于其他元素分配的权重;获取所述邻接矩阵中每个元素对应的训练语义向量,基于图卷积网络,计算所述训练语义向量与分配权重之后每个元素取值之间的乘积,得到训练编码向量。

2.根据权利要求1所述的方法,其中,所述基于所述训练编码向量作为输入,所述兴趣标签作为输出,训练双层神经网络结构,得到标签推荐模型,包括:

将所述训练编码向量作为向前网络的输入,训练所述向前网络,得到新的训练编码向量;

将所述新的训练编码向量再次作为全连接网络的输入,训练所述全连接网络,得到训练标签向量;

将所述训练标签向量作为自变量,输出为兴趣标签,得到标签推荐模型。

3.根据权利要求2所述的方法,其中,所述将所述训练标签向量作为自变量,输出为兴趣标签,得到标签推荐模型,包括:

采用激活函数解析所述训练标签向量,得到所述训练标签向量中包含的兴趣标签;

在所述兴趣标签中,确定与所述兴趣标签对应的第一兴趣标签,并计算所述第一兴趣标签在所述兴趣标签中占用的比例,确定标签推荐模型的概率阈值,得到输出标签概率大于或等于所述概率阈值的标签推荐模型。

4.一种标签获取方法,所述方法包括:

响应于接收到获取兴趣标签指示,获取相应的素材;

采用语义增强表示框架,对所述素材的特征进行表示,得到包括兴趣标签的语义向量;

将社交网络聚合至所述语义向量中,得到编码向量;

将所述编码向量输入至预先训练的标签推荐模型中,得到兴趣标签;

其中,所述素材包括行为素材和业务素材;所述采用语义增强表示框架,对所述素材的特征进行表示,得到包括兴趣标签的语义向量,包括:

基于所述语义增强表示框架,将所述行为素材表示为不同长度的行为向量,将所述业

务素材表示为固定长度的业务向量;将所述行为向量求平均之后,与所述业务向量进行融合,得到语义向量;

所述将社交网络聚合至所述语义向量中,得到编码向量,包括:

获取社交网络,并确定社交网络之间的亲密值;将所述亲密值作为矩阵中元素的取值,构建邻接矩阵;在所述邻接矩阵中每行所述元素的权重和为一的条件下,为所述元素分配权重,所述邻接矩阵中对角线元素分配的权重大于其他元素分配的权重;获取所述邻接矩阵中每个元素对应的语义向量,基于图卷积网络,计算所述语义向量与分配权重之后每个元素取值之间的乘积,得到编码向量。

5.根据权利要求4所述的方法,其中,所述将所述编码向量输入至预先训练的标签推荐模型中,得到兴趣标签,包括:

将所述编码向量输入至所述标签推荐模型中的向前网络中,得到新的编码向量;

将所述新的编码向量输入至全连接网络中,得到标签向量;

解析所述标签向量,基于所述标签推荐模型中的概率阈值,输出兴趣标签。

6.根据权利要求5所述的方法,其中,所述解析所述标签向量,基于所述标签推荐模型中的概率阈值,输出兴趣标签,包括:

基于所述标签推荐模型中的激活函数,解析所述标签向量,得到多个标签;

将所述多个标签中出现概率大于或等于概率阈值的标签,确定为兴趣标签。

7.一种标签推荐模型训练装置,所述装置包括:

获取模块,用于响应于接收到收集训练素材指示,收集训练素材,所述训练素材包含兴趣标签;

处理模块,用于采用语义增强表示框架,对所述训练素材的特征进行表示,得到包括所述兴趣标签的训练语义向量;还用于将社交网络聚合至所述训练语义向量中,得到训练编码向量;

训练模块,用于基于所述训练编码向量作为输入,所述兴趣标签作为输出,训练双层神经网络结构,得到标签推荐模型;

其中,所述训练素材包括行为训练素材和业务训练素材;所述处理模块,用于:

基于所述语义增强表示框架,将所述行为训练素材表示为不同长度的训练行为向量,将所述业务训练素材表示为固定长度的训练业务向量;将所述训练行为向量求平均之后,与所述训练业务向量进行融合,得到训练语义向量;

所述处理模块,还用于:

获取社交网络,并确定社交网络之间的亲密值;将所述亲密值作为矩阵中元素的取值,构建邻接矩阵;在所述邻接矩阵中每行所述元素的权重和为一的条件下,为所述元素分配权重,所述邻接矩阵中对角线元素分配的权重大于其他元素分配的权重;获取所述邻接矩阵中每个元素对应的训练语义向量,基于图卷积网络,计算所述训练语义向量与分配权重之后每个元素取值之间的乘积,得到训练编码向量。

8.根据权利要求7所述的装置,其中,所述训练模块,用于:

将所述训练编码向量作为向前网络的输入,训练所述向前网络,得到新的训练编码向量;

将所述新的训练编码向量再次作为全连接网络的输入,训练所述全连接网络,得到训

练标签向量;

将所述训练标签向量作为自变量,输出为兴趣标签,得到标签推荐模型。

9.根据权利要求8所述的装置,其中,所述训练模块,用于:

采用激活函数解析所述训练标签向量,得到所述训练标签向量中包含的兴趣标签;

在所述兴趣标签中,确定与所述兴趣标签对应的第一兴趣标签,并计算所述第一兴趣标签在所述标签中占用的比例,确定画像模型的概率阈值,得到输出标签概率大于或等于所述概率阈值的标签推荐模型。

10.一种标签获取装置,所述装置包括:

获取模块,用于响应于接收到获取兴趣标签指示,获取相应的素材;

处理模块,用于采用语义增强表示框架,对所述素材的特征进行表示,得到包括兴趣标签的语义向量;还用于将社交网络聚合至所述语义向量中,得到编码向量;

预测模块,用于将所述编码向量输入至预先训练的标签推荐模型中,得到兴趣标签;

其中,所述素材包括行为素材和业务素材;所述处理模块,用于:

基于所述语义增强表示框架,将所述行为素材表示为不同长度的行为向量,将所述业务素材表示为固定长度的业务向量;将所述行为向量求平均之后,与所述业务向量进行融合,得到语义向量;

所述处理模块,还用于:

获取社交网络,并确定社交网络之间的亲密值;将所述亲密值作为矩阵中元素的取值,构建邻接矩阵;在所述邻接矩阵中每行所述元素的权重和为一的条件下,为所述元素分配权重,所述邻接矩阵中对角线元素分配的权重大于其他元素分配的权重;获取所述邻接矩阵中每个元素对应的语义向量,基于图卷积网络,计算所述语义向量与分配权重之后每个元素取值之间的乘积,得到编码向量。

11.根据权利要求10所述的装置,其中,所述预测模块,用于:

将所述编码向量输入至所述标签推荐模型中的向前网络中,得到新的编码向量;

将所述新的编码向量输入至全连接网络中,得到标签向量;

解析所述标签向量,基于所述标签推荐模型中的概率阈值,输出兴趣标签。

12.根据权利要求11所述的装置,其中,所述预测模块,用于:

基于所述标签推荐模型中的激活函数,解析所述标签向量,得到多个标签;

将所述多个标签中出现概率大于或等于概率阈值的标签,确定为兴趣标签。

13.一种电子设备,包括:

至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行权利要求1-3中任一项所述的方法,或以使所述至少一个处理器能够执行权利要求4-6中任一项所述的方法。

14.一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行根据权利要求1-3中任一项所述的方法,或所述计算机指令用于使所述计算机执行根据权利要求4-6中任一项所述的方法。

标签推荐模型训练方法及装置、标签获取方法及装置

技术领域

[0001] 本公开涉及数据处理技术领域，尤其涉及深度学习、云服务、内容搜索等技术领域，具体而言涉及一种标签推荐模型训练方法及装置、标签获取方法及装置。

背景技术

[0002] 兴趣画像包括基于规则、传统模型两种技术方案。属性画像可以是年龄、性别等固定属性，获取简单方便。兴趣画像表示的是兴趣爱好，例如偏好，技能，习惯等方面。两种技术方案的特点是特征，多采用文本表示特征。

发明内容

[0003] 本公开提供了一种标签推荐模型训练方法及装置、标签获取方法及装置。

[0004] 根据本公开的一方面，提供了一种标签推荐模型训练方法，所述方法包括：

[0005] 响应于接收到收集训练素材指示，收集训练素材，所述训练素材包含兴趣标签；采用语义增强表示框架，对所述训练素材的特征进行表示，得到包括所述兴趣标签的训练语义向量；将社交网络聚合至所述训练语义向量中，得到训练编码向量；基于所述训练编码向量作为输入，所述兴趣标签作为输出，训练双层神经网络结构，得到标签推荐模型。

[0006] 根据本公开的第二方面，提供了一种标签获取方法，所述方法包括：

[0007] 响应于接收到获取兴趣标签指示，获取相应的素材；采用语义增强表示框架，对所述素材的特征进行表示，得到包括兴趣标签的语义向量；将社交网络聚合至所述语义向量中，得到编码向量；将所述编码向量输入至预先训练的标签推荐模型中，得到兴趣标签。

[0008] 根据本公开的第三方面，提供了一种标签推荐模型训练装置，所述装置包括：

[0009] 获取模块，用于响应于接收到收集训练素材指示，收集训练素材，所述训练素材包含兴趣标签；处理模块，用于采用语义增强表示框架，对所述训练素材的特征进行表示，得到包括所述兴趣标签的训练语义向量；还用于将社交网络聚合至所述训练语义向量中，得到训练编码向量；训练模块，用于基于所述训练编码向量作为输入，所述兴趣标签作为输出，训练双层神经网络结构，得到标签推荐模型。

[0010] 根据本公开的第四方面，提供了一种标签获取装置，所述装置包括：

[0011] 获取模块，用于响应于接收到获取兴趣标签指示，获取相应的素材；处理模块，用于采用语义增强表示框架，对所述素材的特征进行表示，得到包括兴趣标签的语义向量；还用于将社交网络聚合至所述语义向量中，得到编码向量；预测模块，用于将所述编码向量输入至预先训练的标签推荐模型中，得到兴趣标签。

[0012] 根据本公开的第五方面，提供了一种电子设备，包括：

[0013] 至少一个处理器；以及与所述至少一个处理器通信连接的存储器；其中，所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够执行第一方面或第二方面所述的方法。

[0014] 根据本公开的第六方面，提供了一种存储有计算机指令的非瞬时计算机可读存储

介质,其中,所述计算机指令用于使所述计算机执行根据第一方面或第二方面中所述的方法。

[0015] 根据本公开的第七方面,提供了一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现根据第一方面或第二方面所述的方法。

[0016] 应当理解,本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征,也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

附图说明

[0017] 附图用于更好地理解本方案,不构成对本公开的限定。其中:

[0018] 图1示出了本公开实施例提供的一种标签推荐模型训练方法的流程示意图;

[0019] 图2示出了本公开实施例提供的一种确定训练语义向量方法的流程示意图;

[0020] 图3示出了本公开实施例提供的一种语义向量表示的示意图;

[0021] 图4示出了本公开实施例提供的一种确定训练编码向量方法的流程示意图;

[0022] 图5示出了本公开实施例提供的一种训练模型方法的流程示意图;

[0023] 图6示出了本公开实施例提供的一种神经网络的示意图;

[0024] 图7示出了本公开实施例提供的一种标签推荐模型训练方法的流程示意图;

[0025] 图8示出了本公开实施例提供的一种标签获取方法的流程示意图;

[0026] 图9示出了本公开实施例提供的一种标签推荐模型使用方法的流程示意图;

[0027] 图10示出了本公开实施例提供的一种标签获取方法的流程示意图;

[0028] 图11示出了本公开实施例提供的一种标签推荐模型训练的结构示意图;

[0029] 图12示出了本公开实施例提供的一种标签获取结构示意图;

[0030] 图13是用来实现本公开实施例的电子设备的框图。

具体实施方式

[0031] 以下结合附图对本公开的示范性实施例做出说明,其中包括本公开实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本公开的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0032] 标签在个性化推荐、搜索以及广告点击率预估等多种产品中有着广泛的应用,可以通过兴趣画像获取到准确的兴趣偏好、使用习惯和人口属性等。通过画像可以提升用户的对于产品的体验和收益。

[0033] 一般标签可以分为属性标签和兴趣标签,属性标签用于表征年龄、性别、毕业院校等固定属性。兴趣标签可以包括偏好,拥有技能,习惯等多方面。而兴趣标签不仅应用范围广泛,也体现出千人千面的效果,以提高服务的准确性。

[0034] 但在实际过程中,兴趣爱好是隐式的,一般难以收集或通过规则预测出来,甚至用户自身也很难准确的描述自己的兴趣爱好,在这种情况下,如何准确获取兴趣爱好,以及如何准确获取兴趣标签成为当下的关键问题。

[0035] 相关技术中,在获取兴趣标签的方法采用一般规则或传统模型。例如,在一般规则中,通过人为定义的规则,给用户打上相关的标签,以应用场景为企业办公场景为例,若用

户在的工作的周报中多次提到“深度学习”,则打上“深度学习”的兴趣标签;若用户的主要工作是产品设计和规划,即分配“产品经理(Product Manager,PM)”标签等。在基于传统模型获取用户的兴趣标签时,基于传统模型的方法往往把标签预测转成文本的多分类任务。例如,收集用户的素材,其中素材可以是用户在办公场景下的工作内容,以及与工作内容相关的素材或者文档等。,从而在工作内容、与工作内容相关的素材或者文档中得到用户的特征。需要说明的是,上述工作内容均是在用户允许并同意的情况下获取的。之后应用极致梯度提升(eXtreme Gradient Boosting,XGBoost)、支持向量机(Support Vector Machine,SVM)等分类模型进行分类,其中,每个类别都可以是兴趣标签。

[0036] 如上述实施方式,若采用规则的方法,需要消耗很多的人力资源来总结规则。而且一般只能整理出简单的规则,实现不了隐式的映射,例如,当用户的特征具有文本分类、信息检索与数据挖掘的常用加权技术(Term Frequency-Inverse Document Frequency,TF-IDF)、ONE-HOT编码表示等关键词,则可以确定该用户对“自然语言处理”比较感兴趣,但是却很难总结出特这与标签的映射规则。随着信息的不断变化以及时间的推移,用户的兴趣可能会改变,此时通过规则的方法往往不具有时效性,所以效果变差。

[0037] 若采用传统模型获取用户兴趣画像,虽然可以给员工打上兴趣标签,但效果往往不佳。原因如下:

[0038] (1)传统模型冷启动问题严重,从而导致用户兴趣画像预测失败。其中冷启动问题是指用户的素材比较缺乏,导致特征表达能力不足,传统模型的效果较差。甚至存在部分的用户完全收集不到素材的情况,此时传统模型将完全无法预测。

[0039] (2)在传统模型中,一般采用one-hot编码或语言模型word2vec表示用户特征。但是该类语言表示模型的技术只能捕获浅层的语义信息,往往会导致模型泛化能力不足。

[0040] (3)在传统模型中,传统模型仅采用用户的自身特征作为输入,未包括社交网络等额外的信息。并且由于训练数据集收集比较困难,导致训练数据集一般很小,传统模型在这两种因素下,容易过拟合。

[0041] 基于上述涉及相关技术中存在的不足,在本公开提供一种获取方法,通过用户的社交网络和图神经网络技术实现用户的兴趣画像精准构建。从而确定出可以准确获取到兴趣画像的模型。

[0042] 下述实施例将结合附图对本公开进行说明。

[0043] 图1示出了本公开实施例提供的一种标签推荐模型训练方法的流程示意图,如图1中所示,该方法可以包括:

[0044] 步骤S110:响应于接收到收集训练素材指示,收集训练素材。

[0045] 在本公开实施例中,需要说明的是训练素材为历史数据,训练素材中还包含兴趣标签。在本公开中收集的训练素材可以是与用户相关的素材,当然也可以是其他的素材,在此不做具体限定。

[0046] 在本公开实施例中,训练素材可以是点击/收藏/阅读过的文章。本公开中,可以从知识推荐产品和搜索产品的行为日志中,收集行为训练素材。还可以在办公过程中撰写/编辑的相关文章,收集业务训练素材。在办公过程中撰写/编辑的相关文章可以是周报、晋升材料、项目总结、需求文档等。其中业务训练素材可以是与业务相关的信息,例如,工作中提交的代码分布(C++90%,Python 10%)。

[0047] 通过多渠道收集素材,可以获取隐式反馈(即,行为训练素材),例如日志。还可以获取真实置信的素材,例如办公素材。还可以获取业务训练素材,从而全面的获取素材,有效地保证素材的覆盖率和精确性,有效地解决素材缺乏的问题;以便后续精准的表示素材具有的特征。

[0048] 步骤S120:采用语义增强表示框架,对训练素材进行表示,得到包括兴趣标签的训练语义向量。

[0049] 在本公开实施例中,语义增强表示框架为基于知识增强的持续学习语义理解框架(Enhanced Representation from kNowledge IntEgration,ERNIE)。基于ERNIE对训练素材进行语义表示。得到包括兴趣标签的训练语义向量。

[0050] 需要说明的是,该框架将大数据预训练与多源丰富知识相结合,通过持续学习技术,不断吸收海量文本数据中词汇、结构、语义等方面的知识,实现模型效果不断进化。

[0051] 步骤S130:将社交网络聚合至训练语义向量中,得到训练编码向量。

[0052] 在本公开实施例中,获取社交网络关系,社交关系可以是朋友,在网络中朋友还可以称为邻居。将社交网络关系聚合至训练语义向量中,加强训练语义向量,得到训练编码向量。

[0053] 步骤S140:基于训练编码向量作为输入,兴趣标签作为输出,训练双层神经网络结构,得到标签推荐模型。

[0054] 在本公开实施例中,神经网络可以是深度神经网络(Deep Neural Networks,DNN),也可以是其他神经网络。在本公开中,以神经网络为DNN为例,构建了双层DNN结构。

[0055] 将训练编码向量作为双层DNN结构的输入,兴趣标签作为双层DNN结构的输出,训练双层神经网络结构,得到标签推荐模型。

[0056] 通过本公开实施例提供的标签推荐模型训练方法,采用ERNIE对训练素材进行语义表示,可以使得训练素材具有的特征表示的更加精确。通过训练双层神经网络结构,可以增加素材的覆盖率,从而提升获取兴趣标签的精确度。

[0057] 本公开下述实施例将对采用语义增强表示框架,对所述训练素材进行表示,得到包括兴趣标签的训练语义向量进行说明。

[0058] 图2示出了本公开实施例提供的一种确定训练语义向量方法的流程示意图,如图2中所示,该方法可以包括:

[0059] 步骤S210:基于语义增强表示框架,将行为训练素材表示为不同长度的训练行为向量,将业务训练素材表示为固定长度的训练业务向量。

[0060] 如上述实施例,本公开中训练素材包括行为训练素材和业务训练素材。

[0061] 在本公开实施例中,将行为训练素材表示到具有区分性语义向量中,例如,将与兴趣相似的行为训练素材采用距离相对较小的语义向量进行表示,与兴趣不相似的行为训练素材采用距离相对较大的语义向量进行表示,得到不同长度的训练行为向量。将其他训练素材,表示为固定长度的训练业务向量,例如,业务训练素材。通过ERNIE对业务训练素进行语义表示,例如代码分布[0.9,0.1,...],其中,向量的维数等于编程语言数量,项目中可以设定为10。

[0062] 步骤S220:将训练行为向量求平均之后,与训练业务向量进行融合,得到训练语义向量。

[0063] 在本公开实施例中,不同长度的训练行为向量求平均之后与训练业务向量拼接,得到训练语义向量。

[0064] 示例性的,图3示出了本公开实施例提供的一种语义向量表示的示意图,如图3中所示,将点击的标题,搜索的日志,周报等通过输入成,编码成,再到汇集层,聚合之后输出层输出语义向量,并用代码表示出来。

[0065] 通过本公开实施例将训练行为向量和训练业务向量进行拼接,得到的最终的训练语义向量具有固定的、合理的长度,有利于提高神经网络模型的泛化能力。

[0066] 基于兴趣与具有社交关系的其他兴趣相似的构思,对社交网络进行编码。例如,喜欢游戏的用户,则会有一些同样喜欢游戏的其他用户,他们之间有社交关系。在确定的语义向量的基础上进行编码,得到编码向量。本公开下述实施例将对将社交网络聚合至训练语义向量中,得到训练编码向量进行说明。

[0067] 图4示出了本公开实施例提供的一种确定训练编码向量方法的流程示意图,如图4中所示,该方法可以包括:

[0068] 步骤S310:获取社交网络,并确定社交网络之间的亲密值。

[0069] 在本公开实施例中,社交网络可以是用户之间的社交情况,例如用户之间的互动情况。根据用户之间的社交网络计算用户之间的亲密值,在本公开中亲密值也可以称为亲密度。其中,亲密值的取值范围可以是(0~1.0)。例如采用公式score=(sigmoid(近期沟通天数)+sigmoid(近期沟通次数))/2.0。

[0070] 步骤S320:将亲密值作为矩阵中元素的取值,构建邻接矩阵。

[0071] 在本公开实施例中,示例性的,以用户作为矩阵的元素,根据计算的用户之间的亲密值,以每一行表示一个用户,每一列表示与该用户具有社交联系的其他用户,将亲密值作为矩阵中元素的取值,构建邻接矩阵,并将邻接矩阵表示为A。

[0072] 步骤S330:在邻接矩阵中每行所述元素的权重和为一的条件下,为元素分配权重。

[0073] 其中,邻接矩阵中对角线元素分配的权重大于其他元素分配的权重。

[0074] 在本公开实施例中,基于自身的信息,对邻接矩阵的对角线设定为较大的权重,例如5~10.0。最后通过以下公式,把邻接矩阵的权重归一化,使每一行的和为1。

[0075] $$\widetilde{D}_{ii} = \sum_j \widetilde{A}_{ij}$$

[0076] $$\widehat{A} = \widetilde{D}^{-\frac{1}{2}}\widetilde{A}\widetilde{D}^{-\frac{1}{2}}$$

[0077] 式中,i表示邻接矩阵中的行,j表示邻接矩阵中的列,$\widehat{A}$表示邻接矩阵,$\widetilde{D}_{ii}$表示亲密度。其中,编码向量用$\widehat{A}X$表示。$\widehat{A}$表示编码向量,X表示向量矩阵。

[0078] 步骤S340:获取邻接矩阵中每个元素对应的训练语义向量,基于图卷积网络,计算训练语义向量与分配权重之后每个元素取值之间的乘积,得到训练编码向量。

[0079] 在本公开实施例中,在构建的邻接矩阵的基础上,基于图卷积网络(Graph Convolutional Networks),根据邻接矩阵中每个亲密值以及分配的权重,计算训练语义向量与分配权重之后每个元素取值之间的乘积,确定训练编码向量。

[0080] 在本公开中,通过对邻接矩阵的对角线设定为较大的权重,可以使得编码后形成的向量和更多地偏向于用户的信息。并且社交关系进行编码,缓解决了模型冷启用问题,甚

至可以捕获没有收集素材的特征。

[0081] 下述实施例将对基于训练编码向量作为输入,兴趣标签作为输出,训练双层神经网络结构,得到标签推荐模型进行说明。

[0082] 图5示出了本公开实施例提供的一种训练模型方法的流程示意图,如图5中所示,该方法可以包括:

[0083] 步骤S410:将训练编码向量作为向前网络的输入,训练向前网络,得到新的训练编码向量。

[0084] 在本公开实施例中,本公开采用了Relu作为向前网络的激动函数,表示为$\mathrm{ReLU}(\hat{A}XW^0)$,$W^0$表示神经网络的全连接矩阵,为神经网络的参数,输出的新的训练编码向量为聚合之后的训练编码向量。

[0085] 在本公开一示例性实施例中,图6示出了本公开实施例提供的一种神经网络的示意图。如图6中所示,图中A,B,C,D,E,F表示不同的用户。其中,用户A的具有社交关系的是用户B和用户C。与用户B具有社交关系的是用户A、用户E和用户D。与用户C具有社交关系的是用户A和用户F。以用户A为目标用户为例,根据社交关系对用户A,以及与用户A具有社交关系的用户B的训练编码向量和用户C的训练编码向量进行第一次聚合之后,得到聚合之后用户A的训练编码向量,以及与用户A具有社交关系的用户B的训练编码向量和用户C的训练编码向量。

[0086] 步骤S420:将新的训练编码向量再次作为全连接网络的输入,训练全连接网络,得到训练标签向量。

[0087] 在本公开实施例中,将新的训练编码向量$\mathrm{ReLU}(\hat{A}XW^0)$作为第二层全连接网络的输入,公式表示为$(AVW^1)$,输出的训练标签向量记为$\hat{A}\ \mathrm{ReLU}(\hat{A}XW^0)W^1$。

[0088] 如图6所示,将得到聚合之后用户A的训练编码向量,以及与用户A具有社交关系的用户B的训练编码向量和用户C的训练编码向量再次输入到DNN全连接网络W^1中,得到新的用户训练编码向量,为便于描述,本公开将$\mathrm{ReLU}(\hat{A}XW^0)$记为V。将再次将用户A,用户B和用户C$\mathrm{ReLU}(\hat{A}XW^0)$聚合之后训练编码向量,作为双层神经网络中第二层神经网络的输入,再次输入至神经网络的全连接网络中,公式表示为(AVW^1),得到标签向量$\hat{A}\ \mathrm{ReLU}(\hat{A}XW^0)W^1$,即图6中的Y。

[0089] 其中,需要理解的是,聚合之后的编码向量是一个多维向量,例如,100维的向量,映射100个标签。也就是每个维度代表一个标签。

[0090] 本公开采用两层神经网络结构,通过用户社交关系增加用户的素材,扩大用户素材收集的范围,从而避免过拟合的问题。

[0091] 步骤S430:将训练标签向量作为自变量,输出为兴趣标签,得到标签推荐模型。

[0092] 在本公开实施例中,通过作用于训练标签向量的函数,解析训练标签向量,输出训练兴趣标签。通过计算训练兴趣标签与实际具有的兴趣标签的关系,确定标签推荐模型。

[0093] 图7示出了本公开实施例提供的一种标签推荐模型训练方法的流程示意图,如图7中所示,该方法可以包括:

[0094] 步骤S510:采用激活函数解析训练标签向量,得到训练标签向量中包含的兴趣标签。

[0095] 在本公开实施例中,确定作用于训练标签向量的激活函数,其中,激活函数可以是sigmoid函数。将得到的训练标签向量,作为激活函数的自变量,通过激活函数对训练标签向量进行解析,得到多个标签,即多个训练兴趣标签。

[0096] 步骤S520:在兴趣标签中,确定与兴趣标签对应的第一兴趣标签,并计算第一兴趣标签在兴趣标签中占用的比例,确定标签推荐模型的概率阈值,得到输出标签概率大于或等于概率阈值的标签推荐模型。

[0097] 在本公开实施例中,多个标签中,计算每个标签出现的次数占所以标签出现次数的概率。并计算出与兴趣标签对应的第一兴趣标签出现的次数占所以标签出现次数的概率,从而确定标签推荐模型的概率阈值,得到输出的标签概率大于或等于概率阈值的标签推荐模型。

[0098] 基于相同/相似的构思,本公开还提供一种标签获取方法。

[0099] 图8示出了本公开实施例提供的一种标签获取方法的流程示意图,如图8中所示,该方法可以包括:

[0100] 步骤S610:响应于接收到获取兴趣标签指示,获取相应的素材。

[0101] 在本公开实施例中,若接收到获取兴趣标签指示,获取与该指示对应的素材,如上述实施例,素材包括行为素材和业务素材。

[0102] 步骤S620:采用语义增强表示框架,对素材的特征进行表示,得到包括兴趣标签的语义向量。

[0103] 在本公开实施例中,采用语义增强表示框架对获取的该行为素材和业务素材进行表示,得到包括兴趣标签的行为向量和业务向量。

[0104] 步骤S630:将社交网络聚合至语义向量中,得到编码向量。

[0105] 在本公开实施例中,将行为向量和业务向量通过上述实施例提供的方法进行融合,采用图卷积网络对与具有社交关系语义向量进行编码。根据图卷积网络定义,编码向量可以表征用户,则用户的编码向量=Σ亲密度*员工朋友向量,即$\widehat{A}X$,X表示用户的向量矩阵,一行为一名用户。

[0106] 将得到的语义向量,通过得到的邻接矩阵融入该语义向量中,得到该编码向量。

[0107] 步骤S640:将编码向量输入至预先训练的标签推荐模型中,得到兴趣标签。

[0108] 在本公开实施例中,将得到的编码向量输入至训练完成的标签推荐模型中,标签推荐模型输出该兴趣标签,即得到用户的兴趣标签。

[0109] 通过本公开提供的标签获取方法,可以准确的获取用户的兴趣标签,从而可以准确推荐相关物料。

[0110] 在本公开中,使用标签推荐模型的步骤可参见如下实施例。

[0111] 图9示出了本公开实施例提供的一种标签推荐模型使用方法的流程示意图,如图9中所示,该方法可以包括:

[0112] 步骤S710:将编码向量输入至标签推荐模型中的向前网络中,得到新的编码向量。

[0113] 在本公开实施例中,利用确定训练编码向量的方法,得到编码向量,将编码向量输入至标签推荐模型中的向前网络中,通过该层模型的全连接网络,得到新的编码向量。

[0114] 步骤S720:将新的编码向量输入至全连接网络中,得到标签向量。

[0115] 在本公开实施例中,将新的编码向量输入至标签推荐模型中第二层的全连接网络中,得到标签向量。

[0116] 示例性的,该标签向量中包括用户的特征,例如,深度学习、架构技术、云计算、自然语言处理等特征。

[0117] 步骤S730:解析标签向量,基于标签推荐模型中的概率阈值,输出兴趣标签。

[0118] 在本公开实施例中,通过sigmoid作为激活函数,解析标签向量。通过标签向量中具有的特征,得到与特征对应的兴趣标签,从而在得到的兴趣标签中,确定用户具有的兴趣标签。

[0119] 示例性的,多个特征可以对应一个兴趣标签,例如,具有的文本分类、TF-IDF、ONE-HOT特征都可以对应“自然语言处理”标签。

[0120] 下述实施例将对解析所述标签向量,基于所述标签推荐模型中的概率阈值,输出兴趣标签进行说明。

[0121] 图10示出了本公开实施例提供的一种标签获取方法的流程示意图,如图10中所示,该方法可以包括:

[0122] 步骤S810:基于标签推荐模型中的激活函数,解析标签向量,得到多个标签。

[0123] 根据上述实施例可知,标签向量表示为$\hat{A}\,\mathrm{ReLU}(\hat{A}XW^0)W^1$。其中,解析函数为Z=sigmoid(R),即,

[0124] $$Z=\mathrm{sigmoid}(\hat{A}\,\mathrm{ReLU}(\hat{A}XW^0)W^1)$$

[0125] 其中,Z表示预测的兴趣标签,从而得到多个标签。

[0126] 步骤S820:将多个标签中出现概率大于或等于概率阈值的标签,确定为兴趣标签。

[0127] 在本公开实施例中,在得到的兴趣标签中,计算每个兴趣标签出现次数占所有兴趣标签出现次数的概率,将概率大于或等于概率阈值的兴趣标签,确定为用户具有的兴趣标签。

[0128] 例如概率阈值为0.5,则解析的维度结果中,大于0.5的预测值则确定为用户具有的兴趣标签。

[0129] 在本公开实施例中,可以应用于多种不同的场景,尤其适用于企业内部的知识管理,例如可以是企业的办公场景。本公开以企业的办公场景为例,但不限于该场景。

[0130] 在企业的办公场景中,兴趣可以分为技能、业务、职称三类标签。技能即知识分类体系,例如深度学习、架构技术、云计算、自然语言处理等;业务指员工参与的产品或者项目,例如应用A、应用B等;职称标签也称为序列,代表用户的角色,具体可分为研发工程师(Research and Development engineer,RD)、品质保证(Quality Assurance,QA)、PM、操作员或管理员(Operator,OP)等等。本公开中的目标是每一位用户预测准确的兴趣画像,例如用户A的标签为:路径规划、地图技术、RD、等。

[0131] 通过本公开提出的方法,还可以应用到公司内部的知识推荐和搜索产品中,实现千人千面的推荐效果和精准的搜人效果。首先,在知识推荐产品,借助用户画像的兴趣标签,可以准确地了解到用户的偏好,从而可向用户推荐感兴趣的文章和视频;相比于仅基于人口属性的标签,兴趣标签描述的范围更加广泛,更能体现用户个人的喜好,所以推荐效果

更好。由于用户与产品/项目进行了关联,在搜索产品/项目时,可以直接返回相关人的结构化信息,让用户更快速地获取到相关人信息,降低了搜索成本。实现精确的用户画像预测,有利于提升下游产品的体验,例如推荐和搜索。

[0132] 基于与图1中所示的方法相同的原理,图11示出了本公开实施例提供的一种标签推荐模型训练的结构示意图,如图11所示,该装置100可以包括:

[0133] 获取模块101,用于响应于接收到收集用户训练素材指示,收集训练素材。处理模块102,用于采用语义增强表示框架,对训练素材进行表示,得到包括兴趣标签的训练语义向量。还用于将社交网络聚合至训练语义向量中,得到训练编码向量。训练模块103,用于基于训练编码向量作为输入,兴趣标签作为输出,训练双层神经网络结构,得到标签推荐模型。

[0134] 在本公开实施例中,训练素材包括行为训练素材和业务训练素材。

[0135] 处理模块102,用于基于语义增强表示框架,将行为训练素材表示为不同长度的训练行为向量,将业务训练素材表示为固定长度的训练业务向量。将训练行为向量求平均之后,与训练业务向量进行融合,得到训练语义向量。

[0136] 处理模块102,用于获取社交网络,并确定社交网络之间的亲密值。将亲密值作为矩阵中元素的取值,构建邻接矩阵。在邻接矩阵中每行元素的权重和为一的条件下,为元素分配权重,邻接矩阵中对角线元素分配的权重大于其他元素分配的权重。获取邻接矩阵中每个元素对应的训练语义向量,基于图卷积网络,计算训练语义向量与分配权重之后每个元素取值之间的乘积,得到训练编码向量。

[0137] 训练模块103,用于将训练编码向量作为向前网络的输入,训练向前网络,得到新的训练编码向量。将新的训练编码向量再次作为全连接网络的输入,训练全连接网络,得到训练标签向量。将训练标签向量作为自变量,输出为兴趣标签,得到标签推荐模型。

[0138] 训练模块103,还用于采用激活函数解析训练标签向量,得到训练标签向量中包含的标签。在标签中,确定与兴趣标签对应的第一兴趣标签,并计算第一兴趣标签在标签中占用的比例,确定标签推荐模型的概率阈值,得到输出标签概率大于或等于概率阈值的标签推荐模型。

[0139] 基于与图8中所示的方法相同的原理,图12示出了本公开实施例提供的一种标签获取结构示意图,如图12所示,该标签获取装置200可以包括:

[0140] 获取模块201,用于响应于接收到获取兴趣标签指示,获取相应的素材。处理模块202,用于采用语义增强表示框架,对素材的特征标签表示,得到包括兴趣标签的语义向量。还用于将社交网络聚合至语义向量中,得到编码向量。预测模块203,用于将编码向量输入至预先训练的标签推荐模型中,得到兴趣标签。

[0141] 处理模块202,用于将编码向量输入至画像模型中的向前网络中,得到新的编码向量。将新的编码向量输入至全连接网络中,得到标签向量。解析标签向量,基于画像模型中的概率阈值,输出兴趣标签。

[0142] 预测模块203,用于基于画像模型中的激活函数,解析标签向量,得到多个标签。将多个标签中出现概率大于或等于概率阈值的标签,确定为兴趣标签。

[0143] 本公开的技术方案中,所涉及的用户个人信息的获取,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。

[0144] 根据本公开的实施例,本公开还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

[0145] 图13示出了可以用来实施本公开的实施例的示例电子设备300的示意性框图。电子设备旨在表示各种形式的数字计算机,诸如,膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置,诸如,个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅仅作为示例,并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0146] 如图13所示,设备300包括计算单元301,其可以根据存储在只读存储器(ROM)302中的计算机程序或者从存储单元308加载到随机访问存储器(RAM)303中的计算机程序,来执行各种适当的动作和处理。在RAM 303中,还可存储设备300操作所需的各种程序和数据。计算单元301、ROM 302以及RAM 303通过总线304彼此相连。输入/输出(I/O)接口305也连接至总线304。

[0147] 设备300中的多个部件连接至I/O接口305,包括:输入单元306,例如键盘、鼠标等;输出单元307,例如各种类型的显示器、扬声器等;存储单元308,例如磁盘、光盘等;以及通信单元309,例如网卡、调制解调器、无线通信收发机等。通信单元309允许设备300通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0148] 计算单元301可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元301的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI)计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元301执行上文所描述的各个方法和处理,例如标签推荐模型训练方法和标签获取方法。例如,在一些实施例中,标签推荐模型训练方法和标签获取方法可被实现为计算机软件程序,其被有形地包含于机器可读介质,例如存储单元308。在一些实施例中,计算机程序的部分或者全部可以经由ROM 302和/或通信单元309而被载入和/或安装到设备300上。当计算机程序加载到RAM 303并由计算单元301执行时,可以执行上文描述的标签推荐模型训练方法和标签获取方法的一个或多个步骤。备选地,在其他实施例中,计算单元301可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行标签推荐模型训练方法和标签获取方法。

[0149] 本文中以上描述的系统和技术的各种实施方式可以在数字电子电路系统、集成电路系统、现场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、芯片上系统的系统(SOC)、负载可编程逻辑设备(CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0150] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器,使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的

功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行,作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0151] 在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0152] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,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)或者LCD(液晶显示器)监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以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传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0153] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术的实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:局域网(LAN)、广域网(WAN)和互联网。

[0154] 计算机系统可以包括客户端和服务器。客户端和服务器一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务器的关系。服务器可以是云服务器,也可以为分布式系统的服务器,或者是结合了区块链的服务器。

[0155] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例如,本公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本公开公开的技术方案所期望的结果,本文在此不进行限制。

[0156] 上述具体实施方式,并不构成对本公开保护范围的限制。本领域技术人员应该明白的是,根据设计要求和其他因素,可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等,均应包含在本公开保护范围之内。

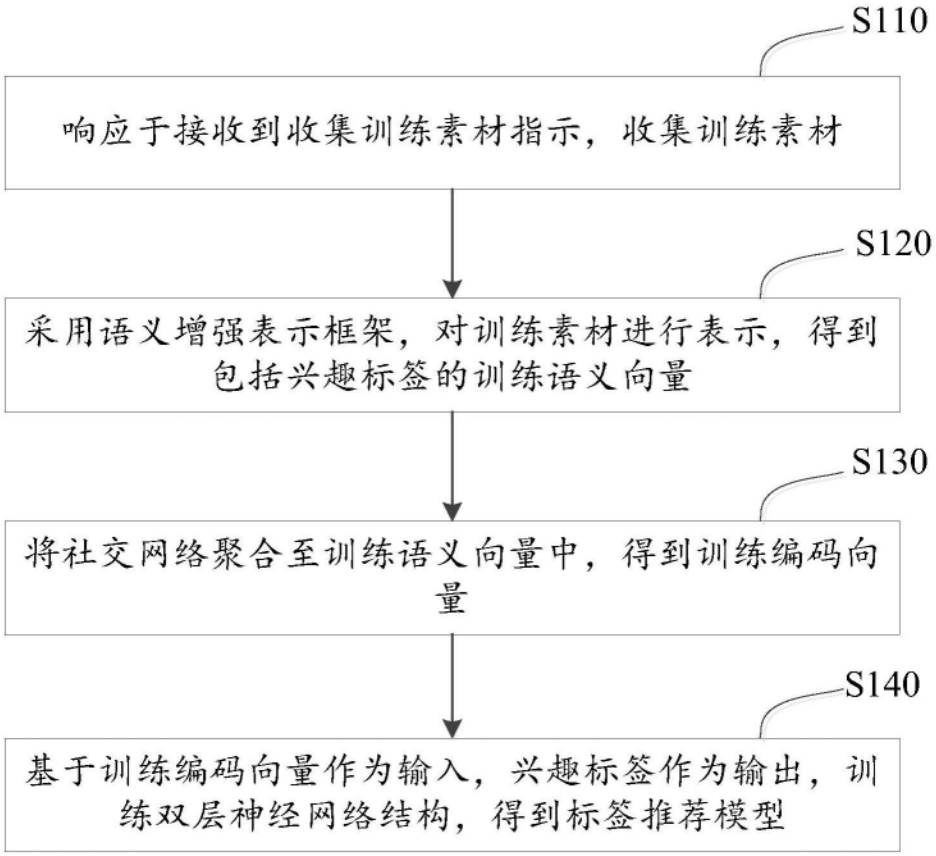

S110
响应于接收到收集训练素材指示，收集训练素材
S120
采用语义增强表示框架，对训练素材进行表示，得到包括兴趣标签的训练语义向量
S130
将社交网络聚合至训练语义向量中，得到训练编码向量
S140
基于训练编码向量作为输入，兴趣标签作为输出，训练双层神经网络结构，得到标签推荐模型

图1

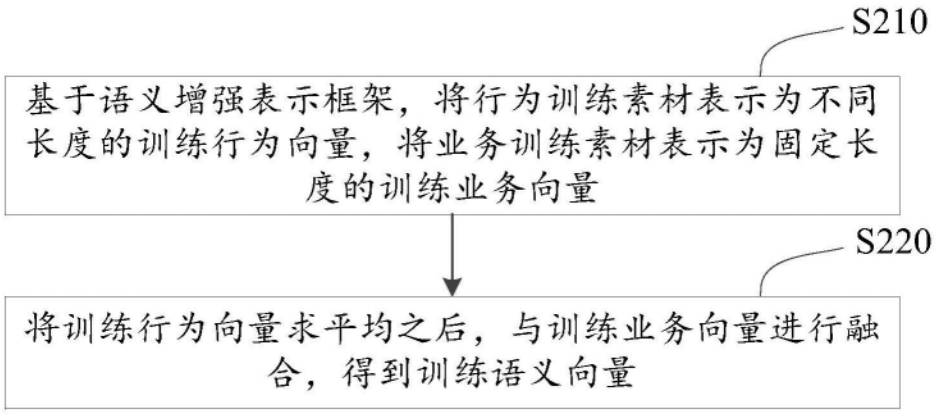

S210
基于语义增强表示框架，将行为训练素材表示为不同长度的训练行为向量，将业务训练素材表示为固定长度的训练业务向量
S220
将训练行为向量求平均之后，与训练业务向量进行融合，得到训练语义向量

图2

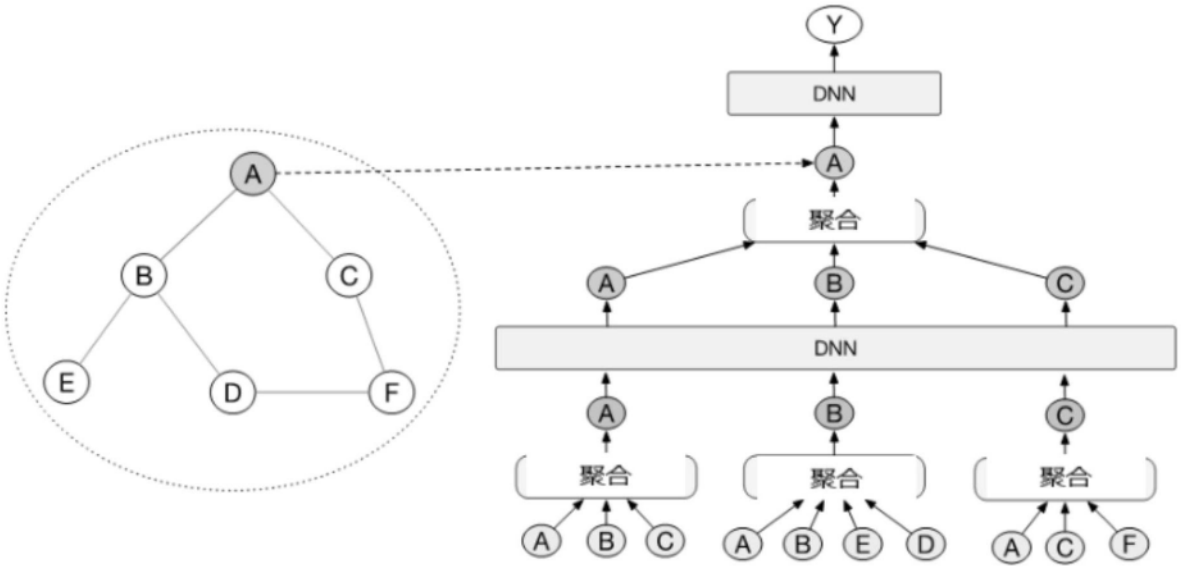

Y
DNN
A
聚合
A
B
C
DNN
A
B
C
聚合
聚合
聚合
A
B
C
A
B
E
D
A
C
F
A
B
C
D
E
F

图3

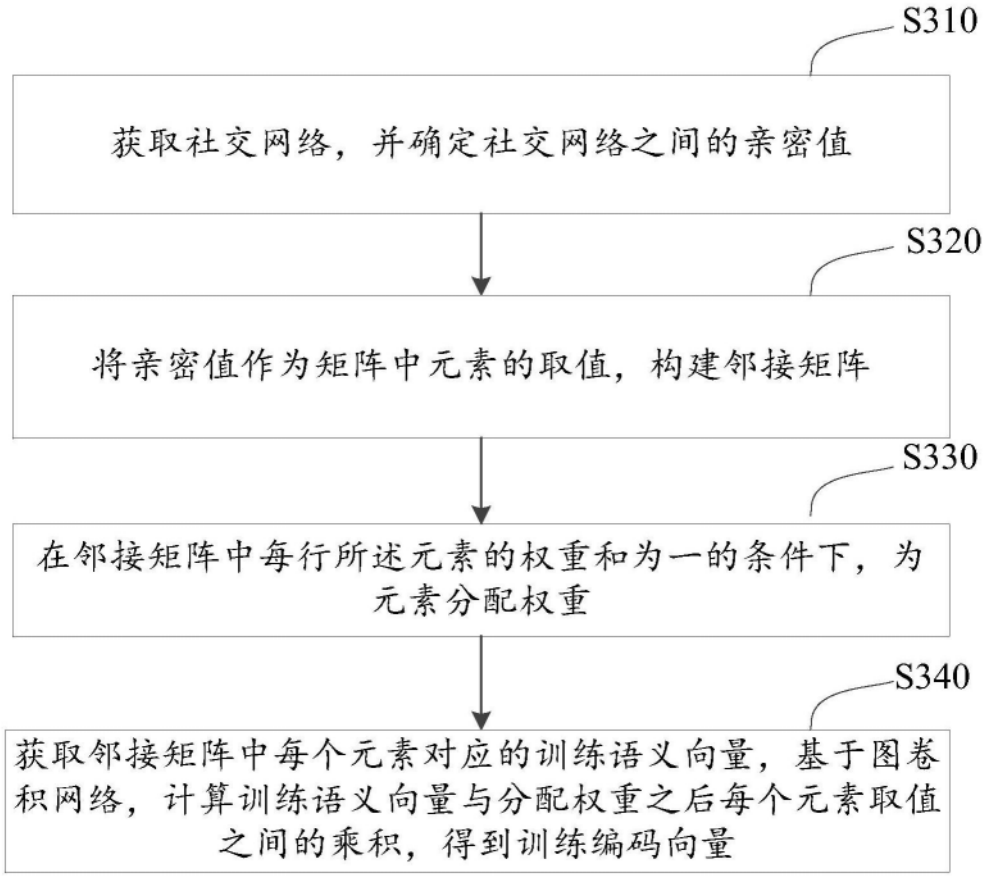

S310
获取社交网络，并确定社交网络之间的亲密值
S320
将亲密值作为矩阵中元素的取值，构建邻接矩阵
S330
在邻接矩阵中每行所述元素的权重和为一的条件下，为元素分配权重
S340
获取邻接矩阵中每个元素对应的训练语义向量，基于图卷积网络，计算训练语义向量与分配权重之后每个元素取值之间的乘积，得到训练编码向量

图4

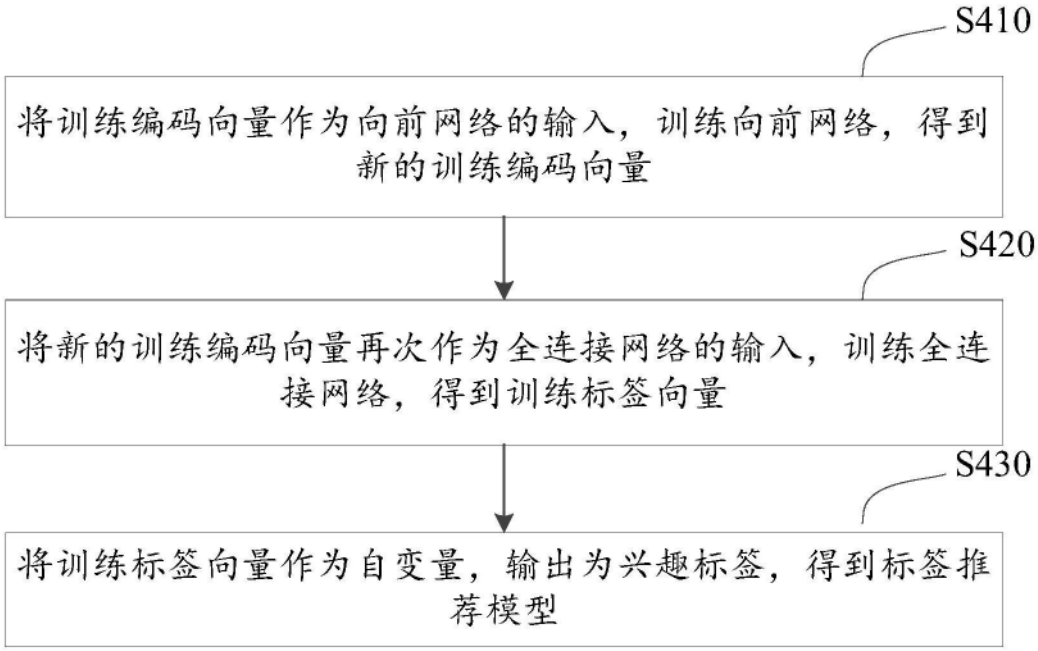

S410
将训练编码向量作为向前网络的输入，训练向前网络，得到新的训练编码向量
S420
将新的训练编码向量再次作为全连接网络的输入，训练全连接网络，得到训练标签向量
S430
将训练标签向量作为自变量，输出为兴趣标签，得到标签推荐模型

图5

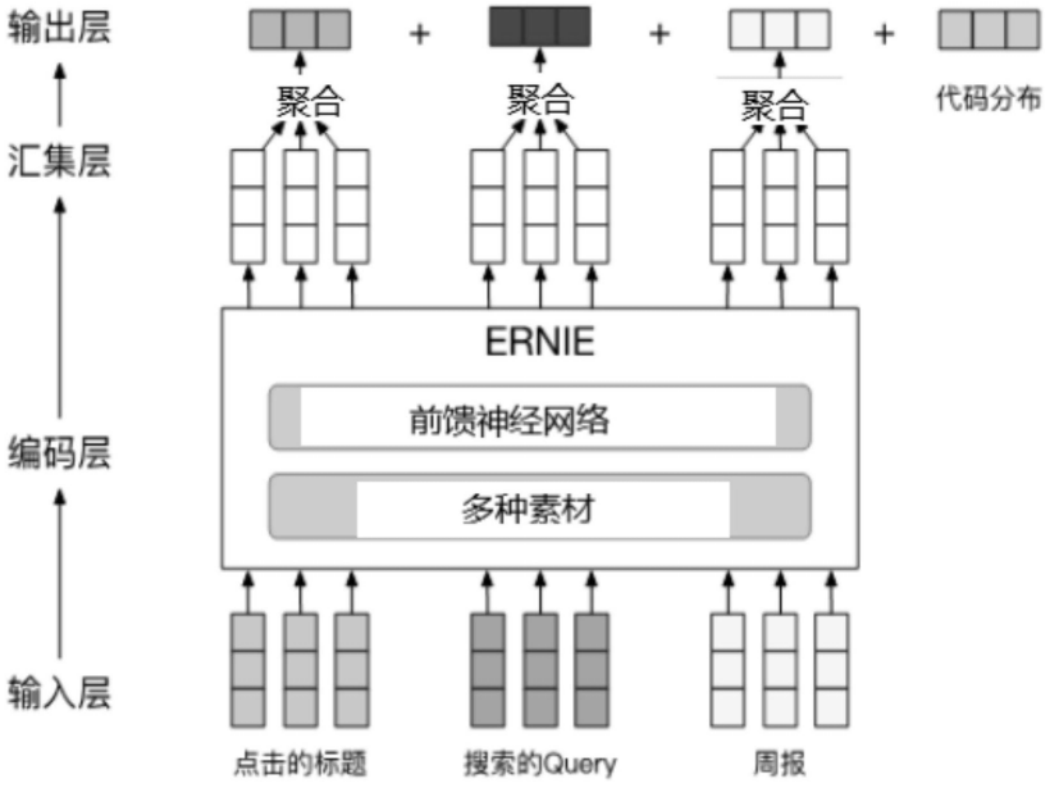

输出层
汇集层
编码层
输入层
聚合
聚合
聚合
代码分布
ERNIE
前馈神经网络
多种素材
点击的标题
搜索的Query
周报

图6

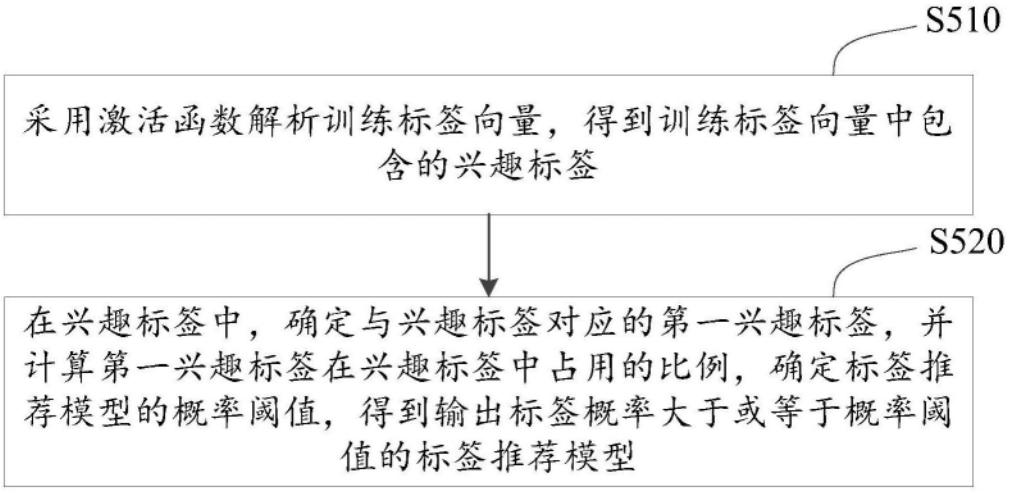

S510
采用激活函数解析训练标签向量，得到训练标签向量中包含的兴趣标签
S520
在兴趣标签中，确定与兴趣标签对应的第一兴趣标签，并计算第一兴趣标签在兴趣标签中占用的比例，确定标签推荐模型的概率阈值，得到输出标签概率大于或等于概率阈值的标签推荐模型

图7

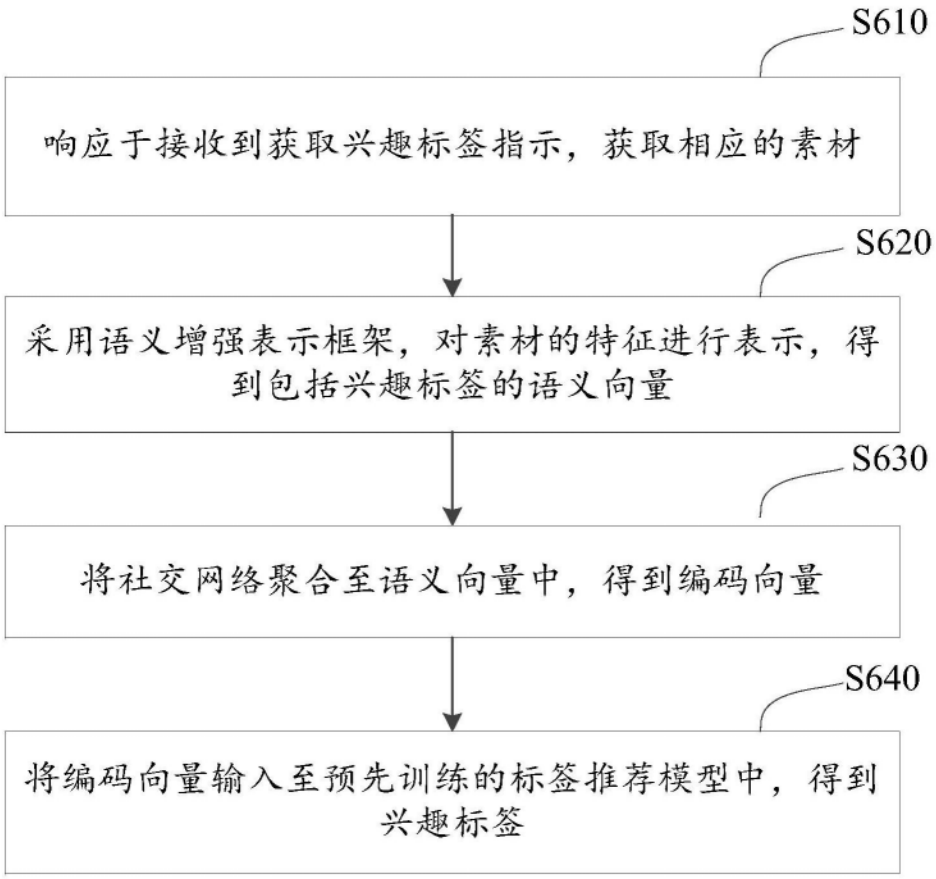

S610
响应于接收到获取兴趣标签指示，获取相应的素材
S620
采用语义增强表示框架，对素材的特征进行表示，得到包括兴趣标签的语义向量
S630
将社交网络聚合至语义向量中，得到编码向量
S640
将编码向量输入至预先训练的标签推荐模型中，得到兴趣标签

图8

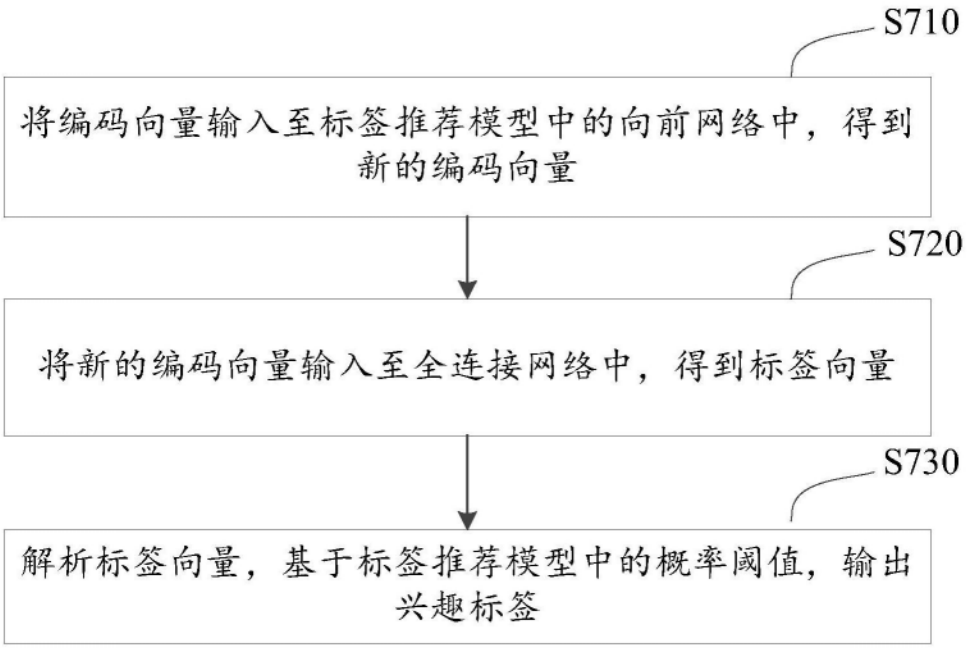

S710
将编码向量输入至标签推荐模型中的向前网络中，得到新的编码向量
S720
将新的编码向量输入至全连接网络中，得到标签向量
S730
解析标签向量，基于标签推荐模型中的概率阈值，输出兴趣标签

图9

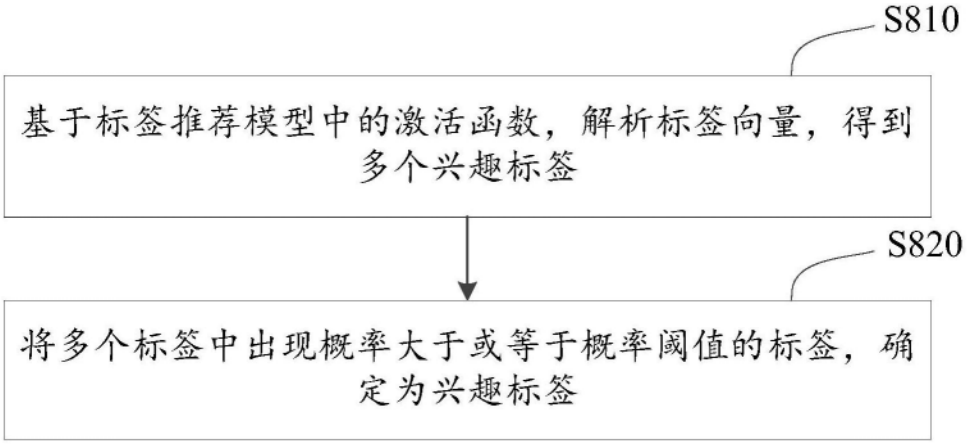

S810
基于标签推荐模型中的激活函数，解析标签向量，得到多个兴趣标签
S820
将多个标签中出现概率大于或等于概率阈值的标签，确定为兴趣标签

图10

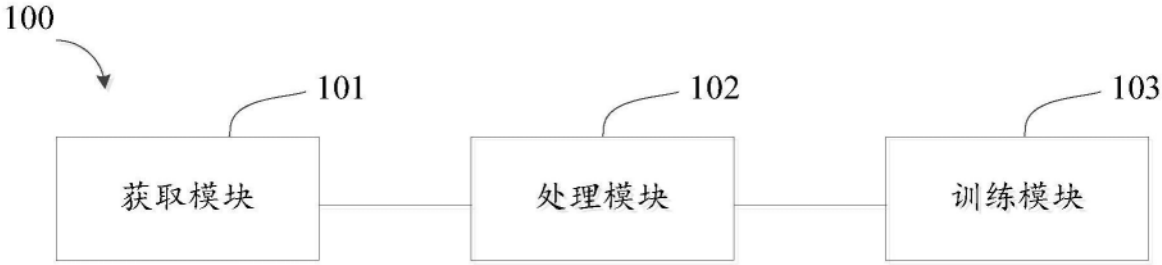

100
101
102
103
获取模块
处理模块
训练模块

图11

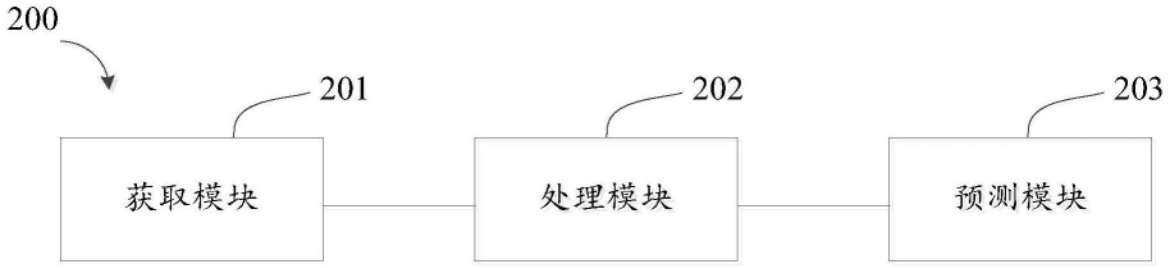

200
201
202
203
获取模块
处理模块
预测模块

图12

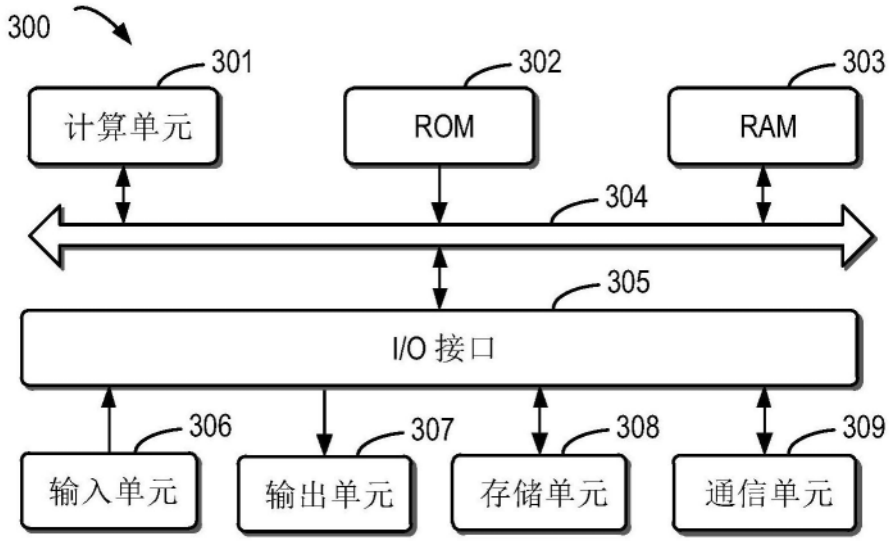

300
301
302
303
计算单元
ROM
RAM
304
305
I/O 接口
306
307
308
309
输入单元
输出单元
存储单元
通信单元

图13