



(12) 发明专利申请

(10) 申请公布号 CN 116257783 A

(43) 申请公布日 2023. 06. 13

(21) 申请号 202310258564.4
(22) 申请日 2023.03.16
(71) 申请人 京东科技信息技术有限公司
地址 100176 北京市大兴区北京经济技术
开发区科创十一街18号院2号楼6层
601
(72) 发明人 赵宇明 宋阳 陈蒙
(74) 专利代理机构 北京律智知识产权代理有限
公司 11438
专利代理师 孙宝海

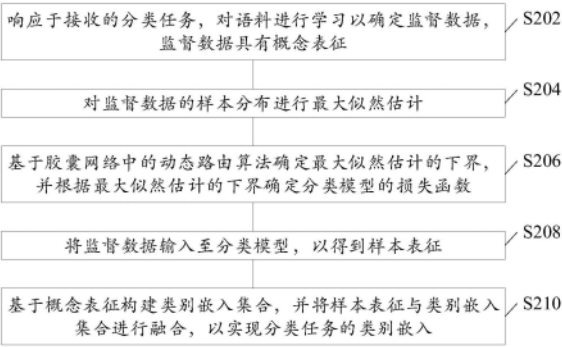
G06N 3/0464 (2023.01)
G06N 3/09 (2023.01)
G06N 3/088 (2023.01)
G06N 20/00 (2019.01)

(51) Int.Cl.
G06F 18/24 (2023.01)
G06F 18/25 (2023.01)
G06F 16/35 (2019.01)
G06F 40/242 (2020.01)

权利要求书2页 说明书13页 附图5页

(54) 发明名称
机器识别场景下的类别嵌入方法、装置、设
备和可读介质

(57) 摘要
本公开提供一种机器识别场景下的类别嵌入方法、装置、设备和可读介质,其中,机器识别场景下的类别嵌入方法包括:响应于接收的分类任务,对语料进行学习以确定监督数据,监督数据具有概念表征;对监督数据的样本分布进行最大似然估计;基于胶囊网络中的动态路由算法确定最大似然估计的下界,并根据最大似然估计的下界确定分类模型的损失函数;将监督数据输入至分类模型,以得到样本表征;基于概念表征构建类别嵌入集合,并将样本表征与类别嵌入集合进行融合,以实现分类任务的类别嵌入。通过本公开实施例,可以提高嵌入类别的泛化性和可靠性,进而有助于提高机器识别的可靠性和准确性。



1. 一种机器识别场景下的类别嵌入方法,其特征在于,包括:
响应于接收的分类任务,对语料进行学习以确定监督数据,所述监督数据具有概念表征;
对所述监督数据的样本分布进行最大似然估计;
基于胶囊网络中的动态路由算法确定所述最大似然估计的下界,并根据所述最大似然估计的下界确定分类模型的损失函数;
将所述监督数据输入至所述分类模型,以得到样本表征;
基于所述概念表征构建类别嵌入集合,并将所述样本表征与所述类别嵌入集合进行融合,以实现所述分类任务的类别嵌入。
2. 如权利要求1所述的机器识别场景下的类别嵌入方法,其特征在于,响应于接收的分类任务,对语料进行学习以确定监督数据包括:
对多个指定领域的语料进行无监督任务学习和/或监督任务学习,以得到样本数据集,所述指定领域包括电商、金融、新闻、法律、计算机、电子、化学和生物钟的至少一种;
根据所述样本数据集确定正例样本和负例样本;
采用所述对比学习算法对所述正例样本和所述负例样本进行对比学习,以确定所述监督数据。
3. 如权利要求2所述的机器识别场景下的类别嵌入方法,其特征在于,根据所述样本数据集确定正例样本和负例样本包括:
若所述样本数据集是经过所述无监督任务学习确定的,则通过两次dropout算法确定所述样本数据集中的正例样本;
将所述样本数据集内除所述正例样本外的其他样本确定为所述负例样本。
4. 如权利要求2所述的机器识别场景下的类别嵌入方法,其特征在于,根据所述样本数据集确定正例样本和负例样本对所述监督数据的样本分布进行最大似然估计包括:
若所述样本数据集是经过所述监督任务学习确定的,则确定所述监督学习任务下的多个标签;
确定同一标签下的多个样本数据互为所述正例样本和所述负例样本。
5. 如权利要求1所述的机器识别场景下的类别嵌入方法,其特征在于,对所述监督数据的样本分布进行最大似然估计包括:
对所述监督数据的对数进行似然估计计算,以确定样本分布;
引入概念变量作为隐变量,并对所述似然估计计算的结果进行最大化。
6. 如权利要求5所述的机器识别场景下的类别嵌入方法,其特征在于,引入概念变量作为隐变量,并对所述似然估计计算的结果进行最大化包括:
引入所述概念变量作为隐变量,并对所述隐变量的样本分布进行推断;
根据推断结果确定所述隐变量分布的聚类中心;
通过所述动态路由算法将所述聚类中心确定为所述胶囊网络中的高层胶囊,并将所述隐变量分布中除聚类中心外的其他聚类样本确定为所述胶囊网络中的底层胶囊;
基于所述胶囊网络的迭代结果对所述似然估计计算的结果进行最大化。
7. 如权利要求1-6中任一项所述的机器识别场景下的类别嵌入方法,其特征在于,基于所述概念表征构建类别嵌入集合,并将所述样本表征与所述类别嵌入集合进行融合,以实

现所述分类任务的类别嵌入包括：

基于所述概念表征构建所述类别嵌入集合；通过所述动态路由算法将所述样本表征集合确定为所述胶囊网络的低层胶囊，将所述类别嵌入集合确定为所述胶囊网络的高层胶囊；

对所述样本表征和所述类别嵌入集合进行融合，以实现所述分类任务的类别嵌入。

8. 一种机器识别场景下的类别嵌入装置，其特征在于，包括：

学习模块，设置为响应于接收的分类任务，对语料进行学习以确定监督数据，所述监督数据具有概念表征；

所述学习模块还设置为，对所述监督数据的样本分布进行最大似然估计；

确定模块，设置为基于胶囊网络中的动态路由算法确定所述最大似然估计的下界，并根据所述最大似然估计的下界确定分类模型的损失函数；

所述学习模块还设置为，将所述监督数据输入至所述分类模型，以得到样本表征；

嵌入模块，设置为基于所述概念表征构建类别嵌入集合，并将所述样本表征与所述类别嵌入集合进行融合，以实现所述分类任务的类别嵌入。

9. 一种电子设备，其特征在于，包括：

存储器；以及

耦合到所述存储器的处理器，所述处理器被配置为基于存储在所述存储器中的指令，执行如权利要求1-7任一项所述的机器识别场景下的类别嵌入方法。

10. 一种计算机可读存储介质，其上存储有程序，该程序被处理器执行时实现如权利要求1-7任一项所述的机器识别场景下的类别嵌入方法。

机器识别场景下的类别嵌入方法、装置、设备和可读介质

技术领域

[0001] 本公开涉及机器学习技术领域,具体而言,涉及一种机器识别场景下的类别嵌入方法、装置、设备和可读介质。

背景技术

[0002] 目前,类别嵌入(label embedding)是一种将类别信息使用向量表示的技术,可用于增强意图识别任务的表现,在很多任务上都表现出了一定的有效性,其关键点在于嵌入内容、嵌入方法。

[0003] 在相关技术中,使用类型相关信息包括层级信息、详细描述和属性短语等作为类别的信息进行表征,或者将类别随机初始化为向量,通过对监督数据的学习得到最终的表征。在借助类别嵌入的意图识别任务中,一个重要的环节就是如果得到类别嵌入向量。

[0004] 但是,现有技术方案中类别嵌入的实现要么依赖于人工的挖掘,要么缺少泛化性。比如,针对一个类别构建其层次信息,或者使用一段文字描述其内涵,都需要人工的参与。又如,若只是使用随机初始化为向量,仅靠监督数据的学习,那么就会依赖监督数据的特性,数据量的大小会影响表征质量,领域会影响表征包含的知识,从而丧失一定的泛化性。

[0005] 需要说明的是,在所述背景技术部分公开的信息仅用于加强对本公开的背景的理解,因此可以包括不构成对本领域普通技术人员已知的现有技术的信息。

发明内容

[0006] 本公开的目的在于提供一种机器识别场景下的、装置、电子设备和可读介质,用于至少在一定程度上克服由于相关技术的限制和缺陷而导致的类别嵌入的泛化性差的问题。

[0007] 根据本公开实施例的第一方面,提供一种机器识别场景下的类别嵌入方法,包括:响应于接收的分类任务,对语料进行学习以确定监督数据,所述监督数据具有概念表征;对所述监督数据的样本分布进行最大似然估计;基于胶囊网络中的动态路由算法确定所述最大似然估计的下界,并根据所述最大似然估计的下界确定分类模型的损失函数;将所述监督数据输入至所述分类模型,以得到样本表征;基于所述概念表征构建类别嵌入集合,并将所述样本表征与所述类别嵌入集合进行融合,以实现所述分类任务的类别嵌入。

[0008] 在本公开的一种示例性实施例中,响应于接收的分类任务,对语料进行学习以确定监督数据包括:对多个指定领域的语料进行无监督任务学习和/或监督任务学习,以得到样本数据集,所述指定领域包括电商、金融、新闻、法律、计算机、电子、化学和生物钟的至少一种;根据所述样本数据集确定正例样本和负例样本;采用所述对比学习算法对所述正例样本和所述负例样本进行对比学习,以确定所述监督数据。

[0009] 在本公开的一种示例性实施例中,根据所述样本数据集确定正例样本和负例样本包括:若所述样本数据集是经过所述无监督任务学习确定的,则通过两次dropout算法确定所述样本数据集中的正例样本;将所述样本数据集内除所述正例样本外的其他样本确定为所述负例样本。

[0010] 在本公开的一种示例性实施例中,根据所述样本数据集确定正例样本和负例样本对所述监督数据的样本分布进行最大似然估计包括:若所述样本数据集是经过所述监督任务学习确定的,则确定所述监督学习任务下的多个标签;确定同一标签下的多个样本数据互为所述正例样本和所述负例样本。

[0011] 在本公开的一种示例性实施例中,对所述监督数据的样本分布进行最大似然估计包括:对所述监督数据的对数进行似然估计计算;引入概念变量作为隐变量,并对所述似然估计计算的结果进行最大化。

[0012] 在本公开的一种示例性实施例中,引入概念变量作为隐变量,并对所述似然估计计算的结果进行最大化包括:引入概念变量作为隐变量,并对所述隐变量的样本分布进行推断;根据推断结果确定所述隐变量分布的聚类中心;通过动态路由算法将所述聚类中心确定为所述胶囊网络中的高层胶囊,并将所述隐变量分布中除聚类中心外的其他聚类样本确定为所述胶囊网络中的底层胶囊;基于所述胶囊网络的迭代结果对所述似然估计计算的结果进行最大化。

[0013] 在本公开的一种示例性实施例中,基于所述概念表征构建类别嵌入集合,并将所述样本表征与所述类别嵌入集合进行融合,以实现所述分类任务的类别嵌入包括:基于所述概念表征构建所述类别嵌入集合;通过所述动态路由算法将所述样本表征集合确定为所述胶囊网络的低层胶囊,将所述类别嵌入集合确定为所述胶囊网络的高层胶囊;对所述样本表征和所述类别嵌入集合进行融合,以实现所述分类任务的类别嵌入。

[0014] 根据本公开实施例的第二方面,提供一种机器识别场景下的类别嵌入装置,包括:学习模块,设置为响应于接收的分类任务,对语料进行学习以确定监督数据,所述监督数据具有概念表征;所述学习模块还设置为,对所述监督数据的样本分布进行最大似然估计;确定模块,设置为基于胶囊网络中的动态路由算法确定所述最大似然估计的下界,并根据所述最大似然估计的下界确定分类模型的损失函数;所述学习模块还设置为,将所述监督数据输入至所述分类模型,以得到样本表征;嵌入模块,设置为基于所述概念表征构建类别嵌入集合,并将所述样本表征与所述类别嵌入集合进行融合,以实现所述分类任务的类别嵌入。

[0015] 根据本公开的第三方面,提供一种电子设备,包括:存储器;以及耦合到所述存储器的处理器,所述处理器被配置为基于存储在所述存储器中的指令,执行如所述任意一项所述的方法。

[0016] 根据本公开的第四方面,提供一种计算机可读存储介质,其上存储有程序,该程序被处理器执行时实现如所述任意一项所述的机器识别场景下的类别嵌入方法。

[0017] 本公开实施例,通过对语料进行学习以确定监督数据,并基于所述概念表征构建类别嵌入集合,并将所述样本表征与所述类别嵌入集合进行融合,以实现所述分类任务的类别嵌入,提高了类别嵌入的表征质量和泛化性,进而提升了机器识别的泛化性。

[0018] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本公开。

附图说明

[0019] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本公开的实施例

例,并与说明书一起用于解释本公开的原理。显而易见地,下面描述中的附图仅仅是本公开的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0020] 图1示出了可以应用本公开实施例的机器识别场景下的类别嵌入方案的示例性系统架构的示意图;

[0021] 图2是本公开示例性实施例中一种机器识别场景下的类别嵌入方法的流程图;

[0022] 图3是本公开示例性实施例中另一种机器识别场景下的类别嵌入方法的流程图;

[0023] 图4是本公开示例性实施例中另一种机器识别场景下的类别嵌入方法的流程图;

[0024] 图5是本公开示例性实施例中另一种机器识别场景下的类别嵌入方法的流程图;

[0025] 图6是本公开示例性实施例中另一种机器识别场景下的类别嵌入方法的流程图;

[0026] 图7是本公开示例性实施例中另一种机器识别场景下的类别嵌入方法的流程图;

[0027] 图8是本公开示例性实施例中另一种机器识别场景下的类别嵌入方法的流程图;

[0028] 图9是本公开示例性实施例中一种机器识别场景下的类别嵌入方案的流程图;

[0029] 图10是本公开示例性实施例中一种机器识别场景下的类别嵌入方案的示意框图;

[0030] 图11是本公开示例性实施例中一种机器识别场景下的类别嵌入装置的方框图;

[0031] 图12是本公开示例性实施例中一种电子设备的方框图。

具体实施方式

[0032] 现在将参考附图更全面地描述示例实施方式。然而,示例实施方式能够以多种形式实施,且不应被理解为限于在此阐述的范例;相反,提供这些实施方式使得本公开将更加全面和完整,并将示例实施方式的构思全面地传达给本领域的技术人员。所描述的特征、结构或特性可以以任何合适的方式结合在一个或更多实施方式中。在下面的描述中,提供许多具体细节从而给出对本公开的实施方式的充分理解。然而,本领域技术人员将意识到,可以实践本公开的技术方案而省略所述特定细节中的一个或更多,或者可以采用其它的方法、组元、装置、步骤等。在其它情况下,不详细示出或描述公知技术方案以避免喧宾夺主而使得本公开的各方面变得模糊。

[0033] 此外,附图仅为本公开的示意性图解,图中相同的附图标记表示相同或类似的部分,因而将省略对它们的重复描述。附图中所示的一些方框图是功能实体,不一定必须与物理或逻辑上独立的实体相对应。可以采用软件形式来实现这些功能实体,或在一个或多个硬件模块或集成电路中实现这些功能实体,或在不同网络和/或处理器装置和/或微控制器装置中实现这些功能实体。

[0034] 图1示出了可以应用本公开实施例的机器识别场景下的类别嵌入方案的示例性系统架构的示意图。

[0035] 如图1所示,系统架构100可以包括终端设备101、102、103中的一种或多种,网络104和服务器105。网络104用以在终端设备101、102、103和服务器105之间提供通信链路的介质。网络104可以包括各种连接类型,例如有线、无线通信链路或者光纤电缆等等。

[0036] 应该理解,图1中的终端设备、网络和服务器的数目仅仅是示意性的。根据实现需要,可以具有任意数目的终端设备、网络和服务器的。比如服务器105可以是多个服务器组成的服务器集群等。

[0037] 用户可以使用终端设备101、102、103通过网络104与服务器105交互,以接收或发送消息等。终端设备101、102、103可以是具有显示屏的各种电子设备,包括但不限于智能手机、平板电脑、便携式计算机和台式计算机等等。

[0038] 在一些实施例中,本公开实施例所提供的机器识别场景下的一般由服务器105执行,相应地,机器识别场景下的类别嵌入装置一般设置于终端设备103(也可以是终端设备101或102)中。在另一些实施例中,某些终端可以具有与服务器设备相似的功能从而执行本方法。

[0039] 下面结合附图对本公开示例实施方式进行详细说明。

[0040] 图2是本公开示例性实施例中机器识别场景下的类别嵌入方法的流程图。

[0041] 参考图2,机器识别场景下的类别嵌入方法可以包括:

[0042] 步骤S202,响应于接收的分类任务,对语料进行学习以确定监督数据,所述监督数据具有概念表征。

[0043] 步骤S204,对所述监督数据的样本分布进行最大似然估计。

[0044] 步骤S206,基于胶囊网络中的动态路由算法确定所述最大似然估计的下界,并根据所述最大似然估计的下界确定分类模型的损失函数。

[0045] 步骤S208,将所述监督数据输入至所述分类模型,以得到样本表征。

[0046] 步骤S210,基于所述概念表征构建类别嵌入集合,并将所述样本表征与所述类别嵌入集合进行融合,以实现所述分类任务的类别嵌入。

[0047] 本公开实施例,通过对语料进行学习以确定监督数据,并基于所述概念表征构建类别嵌入集合,并将所述样本表征与所述类别嵌入集合进行融合,以实现所述分类任务的类别嵌入,提高了类别嵌入的表征质量和泛化性,进而提升了机器识别的泛化性。

[0048] 下面,对机器识别场景下的类别嵌入方法的各步骤进行详细说明。

[0049] 在本公开的一种示例性实施例中,如图3所示,响应于接收的分类任务,对语料进行学习以确定监督数据包括:

[0050] 步骤S302,对多个指定领域的语料进行无监督任务学习和/或监督任务学习,以得到样本数据集,所述指定领域包括电商、金融、新闻、法律、计算机、电子、化学和生物钟的至少一种。

[0051] 步骤S304,根据所述样本数据集确定正例样本和负例样本。

[0052] 步骤S306,采用所述对比学习算法对所述正例样本和所述负例样本进行对比学习,以确定所述监督数据。

[0053] 在所述实施例中,基于无监督任务学习产生正例样本时,预料输入至模型得到自身编码,然后调整dropout rate,再次将自身送入模型得到编码,此为自身的正例。

[0054] 在所述实施例中,基于监督任务学习产生正例样本时,如果一个batch内有4个样本A、B、C、D,因为样本均设置标签,对于A,如果只有C和自身标签一致,那么C就是正例,同样的,C的正例也是A。

[0055] 在本公开的一种示例性实施例中,如图4所示,根据所述样本数据集确定正例样本和负例样本包括:

[0056] 步骤S402,若所述样本数据集是经过所述无监督任务学习确定的,则通过两次dropout算法确定所述样本数据集中的正例样本。

[0057] 步骤S404,将所述样本数据集内除所述正例样本外的其他样本确定为所述负例样本。

[0058] 在所述实施例中,如果样本数据集中的数据是无监督数据,在产生正例样本时,可采用现有比较成熟的方法,比如两次dropout,在产生负例样本时,将每一个batch内的样本,除本身和正例外,均作为负例对待。

[0059] 其中,Dropout可以作为训练深度神经网络的一种trick供选择。在每个训练批次中,通过忽略一半的特征检测器(让一半的隐层节点值为0),可以明显地减少过拟合现象。这种方式可以减少特征检测器(隐层节点)间的相互作用,检测器相互作用是指某些检测器依赖其他检测器才能发挥作用。

[0060] 在本公开的一种示例性实施例中,如图5所示,根据所述样本数据集确定正例样本和负例样本对所述监督数据的样本分布进行最大似然估计包括:

[0061] 步骤S502,若所述样本数据集是经过所述监督任务学习确定的,则确定所述监督学习任务下的多个标签。

[0062] 步骤S504,确定同一标签下的多个样本数据互为所述正例样本和所述负例样本。

[0063] 在所述实施例中,如果样本数据集的数据是由监督数据组成,在产生正例样本时,可以将同一标签下的样本互为正例,否则,作为负例对待。具体而言,随机采样一个batch size大小为B的样本集合 $S = \{s_1, \dots, s_B\}$,然后送入预训练语言模型,得到输出表征 $Z = \{z_1, \dots, z_B\}$,其中 $z_i \in \mathbb{R}^d$,是bert最后一层输出在token维度的平均。无监督数据情况下,需要调整dropout比率,再次将S送入预训练模型,得到输出表征 $\hat{Z} = \{\hat{z}_1, \dots, \hat{z}_B\}$,其中 $\hat{z}_i \in \mathbb{R}^d, 1 \leq i \leq B$ 。对于每一条样本表征 z_i , $\hat{z}_i$ 将作为其正例,而 $\{z_j, \hat{z}_j\}_{j \neq i}$ 将作为其负例,在监督数据情况下,对于每一条样本表征 z_i , $\{z_j\}_{y_j=y_i}$ 作为其正例, $\{z_j\}_{y_j \neq y_i}$ 作为其负例。

[0064] 在本公开的一种示例性实施例中,如图6所示,对所述监督数据的样本分布进行最大似然估计包括:

[0065] 步骤S602,对所述监督数据的对数进行似然估计计算,以确定样本分布。

[0066] 步骤S604,引入概念变量作为隐变量,并对所述似然估计计算的结果进行最大化。

[0067] 在所述实施例中,引入概念并对其进行表征,并将概念作为隐变量进行推断,目的是在找出样本通用的概念知识的条件下,能够保证对样本分布进行极大似然估计。在机器识别场景下的类别嵌入方案中,由于概率密度函数包含了一个随机变量的全部信息;样本数据可能不多;特征向量x的维度可能很大等原因,类条件概率的估计(非常难),要直接估计类条件概率的密度函数很难。而本公开的实施例通过极大似然估计,把估计完全未知的概率密度转化为估计参数,就将概率密度估计问题转化为参数估计问题,从而简化了类条件概率的计算过程。

[0068] 在本公开的一种示例性实施例中,如图7所示,引入概念变量作为隐变量对所述似然估计计算的结果进行最大化包括:

[0069] 步骤S702,引入概念变量作为隐变量,并对所述隐变量的样本分布进行推断。

[0070] 步骤S704,根据推断结果确定所述隐变量分布的聚类中心。

[0071] 步骤S706,通过动态路由算法将所述聚类中心确定为所述胶囊网络中的高层胶囊,并将所述隐变量分布中除聚类中心外的其他聚类样本确定为所述胶囊网络中的底层胶

囊。

[0072] 步骤S708,基于所述胶囊网络的迭代结果对所述似然估计计算的结果进行最大化。

[0073] 在所述实施例中,可以借助聚类算法将C定义为K个聚类中心点集合,也即得到 $C = \{c_1, \dots, c_K\}$,为了聚类的稳定性,待聚类样本数量尽可能大些,可以选择50倍batch size的数据量,输入当前模型,此处仅做正向计算,不需要梯度更新,聚类算法仍旧选择胶囊网络中的动态路由算法,将中心点看做高层胶囊,待聚类样本看做底层胶囊。

[0074] 在本公开的一种示例性实施例中,如图8所示,基于所述概念表征构建类别嵌入集合,并将所述样本表征与所述类别嵌入集合进行融合,以实现所述分类任务的类别嵌入包括:

[0075] 步骤S802,基于所述概念表征构建所述类别嵌入集合;。

[0076] 步骤S804,通过所述动态路由算法将所述样本表征集合确定为所述胶囊网络的低层胶囊,将所述类别嵌入集合确定为所述胶囊网络的高层胶囊。

[0077] 步骤S806,对所述样本表征和所述类别嵌入集合进行融合,以实现所述分类任务的类别嵌入。

[0078] 结合图9和图10所示,对本公开限定的类别嵌入框架1000的一阶段1002和二阶段1004进行具体说明:

[0079] 步骤一S902,采集一阶段语料:

[0080] 1) 在一阶段,主要使用无监督任务学习,因此对于训练语料没有特殊要求,为了达到形成通用概念知识的目的,会采集多领域的文本数据进行学习。为了平衡训练成本和通用性,选取中文语义匹配、中文语言推理、中文文本分类等任务,电商、金融、新闻、法律等领域已有的数据集作为一阶段语料。

[0081] 2) 汇总所述语料,得到一阶段数据集,表示为 $\{q_1, \dots, q_N\}$,其中 q_i 是单条样本,N表示样本数量。

[0082] 步骤二S904,进行一阶段训练:

[0083] 首先,构建词典,由于采用的预训练语言模型bert,因此沿用bert自带的词典,大小为 $|V|$;

[0084] 其次,对于一个样本 q_i ,将其中的每一个字,经过wordpiece分词后,转换成词典中对应的id值,得到一个id序列, $s = \{x_1, \dots, x_M\}$,其中M为 q_i 分词后的长度, x_M 为取值在 $[0, |V|)$ 之间的正整数,然后对其进行补零操作,设置最大长度为T,得到最终样本序列, $s = \{x_1, \dots, x_T\}$;

[0085] 再次,为了构建对比学习任务,需要产生正负例对,如果数据是无监督数据,对于产生正例,可采用现有比较成熟的方法,比如两次dropout,对于产生负例,将每一个batch内的样本,除本身和正例外,均作为负例对待;如果数据是由监督数据组成,对于产生正例,可以将同一标签下的样本互为正例,否则,作为负例对待。

[0086] 具体而言,随机采样一个batch size大小为B的样本集合 $S = \{s_1, \dots, s_B\}$,然后送入预训练语言模型,得到输出表征 $Z = \{z_1, \dots, z_B\}$,其中 $z_i \in \mathbb{R}^d$,是bert最后一层输出在token维度的平均。无监督数据情况下,需要调整dropout比率,再次将S送入预训练模型,得到输出表征 $\hat{Z} = \{\hat{z}_1, \dots, \hat{z}_B\}$,其中 $\hat{z}_i \in \mathbb{R}^d$ 。对于每一条样本表征 z_i , $\hat{z}_i$ 将作为其正例,而

$\{z_j, \hat{z}_j\}_{j \neq i}$ 将作为其负例；监督数据情况下，对于每一条样本表征 z_i ， $\{z_j\}_{y_j=y_i}$ 作为其正例， $\{z_j\}_{y_j \neq y_i}$ 作为其负例。

[0087] 进而利用所述正负例，进行表征的对比学习，无监督数据下损失函数为 $\mathcal{L}_{self}$ 如表达式(1)所示，监督数据下损失函数为 $\mathcal{L}_{sup}$ 如表达式(2)所示：

$$[0088] \quad \mathcal{L}_{self} = E_{z_i \sim Data} \left[-\log \frac{\exp(\text{sim}(z_i, \hat{z}_i)/\tau)}{\sum_j^B \exp\left(\frac{\text{sim}(z_i, \hat{z}_{j \neq i})}{\tau}\right) + \exp\left(\frac{\text{sim}(z_i, z_{i \neq j})}{\tau}\right)} \right], \text{ 表达式 (1)}$$

$$[0089] \quad \mathcal{L}_{sup} = E_{z_i \sim Data} \left[\sum_k -\log \frac{\exp(\text{sim}(z_i, z_{y_k=y_i, k \neq i})/\tau)}{\sum_j^B \exp(\text{sim}(z_i, z_{j \neq i})/\tau)} \right], \text{ 表达式 (2)}$$

[0090] 进一步地，构建概念形成任务，即引入概念并对其表征。将概念作为隐变量进行推断，目的是在找出样本通用的概念知识的条件下，能够保证对样本分布进行极大似然估计。

[0091] 进一步地，使用 z_i 表示样本随机变量，其样本表征为上一步中的 z_i ，即观测数据，使用 c_k 表示概念随机变量，即隐变量，对于样本的对数似然估计如表达式(3)所示：

$$[0092] \quad E_{z \sim Data} [\log p(z)], \text{ 表达式 (3)}$$

[0093] 引入概念隐变量后如表达式(4)和如表达式(5)所示示，

$$[0094] \quad p(z) = \sum_{c \in C} p(z, c), \text{ 表达式 (4)}$$

$$[0095] \quad E_{z \sim Data} [\log p(z)] = E_{z \sim Data} \left[\sum_{c \in C} \log p(z, c) \right], \text{ 表达式 (5)}$$

[0096] 由于 $p(z, c)$ 未知，并不能得到解析形式，引入分布 $q(c)$ ，上式改写后如表达式(6)所示：

$$[0097] \quad \begin{aligned} E_{z \sim Data} \left[\log \left(\sum_{c \in C} p(z, c) \right) \right] &= E_{z \sim Data} \left[\log \left(\sum_{c \in C} q(c) \frac{p(z, c)}{q(c)} \right) \right] \\ &\geq E_{z \sim Data} \left[\sum_{c \in C} q(c) \log \left(\frac{p(z, c)}{q(c)} \right) \right], \text{ 表达式 (6)} \end{aligned}$$

[0098] 不等式通过詹森不等式得到，为了保证等号的成立，需要满足 $\frac{p(z, c)}{q(c)}$ 对于不同的 c 都相等成立，即， $\frac{p(z, c)}{q(c)} = U$ ， U 为常数，两边同时对 c 求积分后如表达式(7)所示：

$$[0099] \quad \sum_c p(z, c) = \sum_c U \cdot q(c), \text{ 表达式 (7)}$$

[0100] 因此，得到 $U = p(z)$ ，故确定 $q(c) = \frac{p(z, c)}{p(z)} = p(c|z)$ 。

[0101] 所述推导表明通过引入概念变量作为隐变量,通过不断最大化一个下界,来最大化对数似然。

[0102] 在本公开的一个实施例中,借助EM(Expectation Maximization Algorithm,又译期望最大化算法)完成所述过程,首先是隐变量分布的推断,即 $p(c|z)$,由于希望C能代表通用概念,因此借助聚类算法,将C定义为K个聚类中心点集合,也就是说 $C = \{c_1, \dots, c_K\}$ 。为了聚类的稳定性,待聚类样本数量尽可能大些,选择50倍batch size的数据量,输入当前模型,此处仅做正向计算,不需要梯度更新。

[0103] 在本公开的一个实施例中,聚类算法仍旧选择胶囊网络中的动态路由算法,将中心点看做高层胶囊,待聚类样本看做底层胶囊。经过r次迭代后,可以将第r次迭代,低层胶囊在高层胶囊上的 $h = \text{softmax}(b)$ 归一化权重作为 $p(c|z)$,因此,需要优化的对数似然函数下界,忽略常数项后如表达式(8)所示:

$$[0104] \quad E_{z \sim \text{Data}} \left[\sum_{c \in C} q(c) \log p(z, c) \right] = E_{z \sim \text{Data}} \left[\sum_{c \in C} h \log p(z, c) \right], \text{ 表达式 (8)}$$

[0105] 其中,动态路由算法的代码如下:

```

“1: for all capsule i in layer l and capsule j in layer
(l+1):  $b_{ij} \leftarrow 0$ .
2: for r iterations do
3:   for all capsule i in layer l:  $e_i \leftarrow \text{softmax}(b_i)$ 
4:   for all capsule j in layer (l+1):  $s_j \leftarrow \sum_i e_{ij} u_{j|i}$ 
5:   for all capsule j in layer (l+1):  $v_j \leftarrow \text{squash}(s_j)$ 
6:   for all capsule I in layer l and capsule j in layer
(l+1):  $b_{ij} \leftarrow b_{ij} + u_{j|i} v_j$ .
7: return  $v_j, c_i$ ”
```

[0107] 接下来进行M步, $p(z, c) = p(z|c) p(c)$,可以假设先验 $p(c)$ 是均匀分布 $U(0, K)$,问题的关键是 $p(z|c)$,获得其表示形式后,便可以通过梯度下降来最大化对数似然下界。此时,可以假设, $p(z|c)$ 是高斯分布,其均值即为c,方差为 σ_c^2 ,如表达式(9)所示:

$$[0108] \quad p(z|c) = \frac{\exp(-(z-c)^2/2\sigma_c^2)}{\tilde{Z}}, \text{ 表达式 (9)}$$

[0109] 其中 $\tilde{Z}$ 是归一化因子,可表示为 $\sum_{c \in C} \exp(-(z-c)^2/2\sigma_c^2)$ 。将z、c归一化后,完整的下界如表达式(10)所示:

$$\begin{aligned}
& E_{z \sim \text{Data}} \left[\sum_{c \in C} \text{hlog} p(z, c) \right] = E_{z \sim \text{Data}} \left[\sum_{c \in C} \text{hlog} \frac{\exp \left(-\frac{(z - c)^2}{2\sigma_c^2} \right)}{\tilde{Z}} \right] \\
& = E_{z \sim \text{Data}} \left[\sum_{c \in C} \text{hlog} \frac{\exp \left(\frac{z \cdot c}{\sigma_c^2} - \frac{1}{\sigma_c^2} \right)}{\tilde{Z}} \right], \text{ 表达式 (10)}
\end{aligned}$$

[0111] 其中, σ_c^2 是可以描述当前类中心点紧密程度的参数, 此处设置为经验值。因此, 概念形成任务的最终损失函数如表达式 (11) 所示:

$$\mathcal{L}_{con} = E_{z \sim \text{Data}} \left[\frac{1}{K} \sum_{c \in C} -\text{hlog} \frac{\exp \left(\frac{z \cdot c}{\sigma_c^2} \right)}{\sum_{c' \in C} \exp (z \cdot c' / \sigma_{c'}^2)} \right], \text{ 表达式 (11)}$$

[0113] 步骤三步骤S906, 采集二阶段语料:

[0114] 二阶段即下游任务阶段, 训练语料即为对应任务的监督数据。表示为 $\{(a_1, y_1), \dots, (a_N, y_N)\}$, 其中 a_i 是文本, y_i 为对应标签, N 表示样本数量, 类别数量为 G 。

[0115] 步骤四S908, 进行二阶段训练:

[0116] 构建词典, 由于采用的预训练语言模型 b_{ert} , 因此沿用 b_{ert} 自带的词典, 大小为 $|V|$;

[0117] 对于一个样本 a_i , 将其中的每一个字, 经过 wordpiece 分词后, 转换成词典中对应的 id 值, 得到一个 id 序列, $s = \{x_1, \dots, x_M\}$, 其中 M 为 a_i 分词后的长度, x_M 为取值在 $[0, |V|)$ 之间的正整数, 然后对其进行补零操作, 设置最大长度为 T , 得到最终样本序列, $s = \{x_1, \dots, x_T\}$

[0118] 将样本送入模型中, 得到样本的表征, $z \in \mathbb{R}^{dxT}$ 。

[0119] 利用一阶段训练得到的概念表征构建类别嵌入集合 $L \in \mathbb{R}^{dxG}$ 。一阶段可以得到概念表征集合 $C \in \mathbb{R}^{dxK}$, 将该集合视作低层胶囊, 将类别嵌入视作高层胶囊, 仍旧通过动态路由算法, 得到类别嵌入 $L \in \mathbb{R}^{dxG}$ 。

[0120] 将样本表征与类别嵌入进行融合, 此处选用简单的融合方式, 即交叉注意力机制, 得到最终的融合表征 Z^o 如下:

$$[0121] \quad H = C^T L;$$

$$[0122] \quad \hat{C} = L \text{softmax}(H^T);$$

$$[0123] \quad \hat{L} = C \text{softmax}(H);$$

$$[0124] \quad Z^o = \text{Concat}(L + \hat{L}, C + \hat{C}).$$

[0125] 分类器选取线性分类器, 其权重矩阵是 $W \in \mathbb{R}^{2d \times G}$, 那么最终的概率 $\hat{y} \in \mathbb{R}^G$ 如表达式 (12) 所示:

$$[0126] \quad \hat{y} = \text{softmax}(WZ^o), \text{ 表达式 (12)}$$

[0127] 使用交叉熵作为损失函数如表达式(13)所示:

[0128] $L = CE(\hat{y}, y)$, 表达式 (13)

[0129] 其中, CE 表示交叉熵函数, y 为真实标签, 并使用基于Adam的优化器进行梯度下降训练。

[0130] 对应于所述方法实施例, 本公开还提供一种机器识别场景下的类别嵌入装置, 可以用于执行所述方法实施例。

[0131] 图11是本公开示例性实施例中一种机器识别场景下的类别嵌入装置的方框图。

[0132] 参考图11, 机器识别场景下的类别嵌入装置1100可以包括:

[0133] 学习模块1102, 设置为响应于接收的分类任务, 对语料进行学习以确定监督数据, 所述监督数据具有概念表征。

[0134] 所述学习模块1102还设置为, 对所述监督数据的样本分布进行最大似然估计。

[0135] 确定模块1104, 设置为基于胶囊网络中的动态路由算法确定所述最大似然估计的下界, 并根据所述最大似然估计的下界确定分类模型的损失函数。

[0136] 所述学习模块1102还设置为, 将所述监督数据输入至所述分类模型, 以得到样本表征。

[0137] 嵌入模块1106, 设置为基于所述概念表征构建类别嵌入集合, 并将所述样本表征与所述类别嵌入集合进行融合, 以实现所述分类任务的类别嵌入。

[0138] 在本公开的一种示例性实施例中, 学习模块1102还设置为: 对多个指定领域的语料进行无监督任务学习和/或监督任务学习, 以得到样本数据集, 所述指定领域包括电商、金融、新闻、法律、计算机、电子、化学和生物钟的至少一种; 根据所述样本数据集确定正例样本和负例样本; 采用所述对比学习算法对所述正例样本和所述负例样本进行对比学习, 以确定所述监督数据。

[0139] 在本公开的一种示例性实施例中, 学习模块1102还设置为: 若所述样本数据集是经过所述无监督任务学习确定的, 则通过两次dropout算法确定所述样本数据集中的正例样本; 将所述样本数据集内除所述正例样本外的其他样本确定为所述负例样本。

[0140] 在本公开的一种示例性实施例中, 学习模块1102还设置为: 若所述样本数据集是经过所述监督任务学习确定的, 则确定所述监督学习任务下的多个标签; 确定同一标签下的多个样本数据互为所述正例样本和所述负例样本。

[0141] 在本公开的一种示例性实施例中, 机器识别场景下的类别嵌入装置1100还设置为: 对所述监督数据的对数进行似然估计计算; 引入概念变量作为隐变量, 并对所述似然估计计算的结果进行最大化。

[0142] 在本公开的一种示例性实施例中, 机器识别场景下的类别嵌入装置1100还设置为: 引入概念变量作为隐变量, 并对所述隐变量的样本分布进行推断; 根据推断结果确定所述隐变量分布的聚类中心; 通过动态路由算法将所述聚类中心确定为所述胶囊网络中的高层胶囊, 并将所述隐变量分布中除聚类中心外的其他聚类样本确定为所述胶囊网络中的底层胶囊; 基于所述胶囊网络的迭代结果对所述似然估计计算的结果进行最大化。

[0143] 在本公开的一种示例性实施例中, 所述嵌入模块1106还配置为: 基于所述概念表征构建类别嵌入集合, 并将所述样本表征与所述类别嵌入集合进行融合, 以实现所述分类任务的类别嵌入包括: 基于所述概念表征构建所述类别嵌入集合; 通过所述动态路由算法

将所述样本表征集合确定为所述胶囊网络的低层胶囊,将所述类别嵌入集合确定为所述胶囊网络的高层胶囊;对所述样本表征和所述类别嵌入集合进行融合,以实现所述分类任务的类别嵌入。

[0144] 由于机器识别场景下的类别嵌入装置1100的各功能已在其对应的方法实施例中予以详细说明,本公开于此不再赘述。

[0145] 应当注意,尽管在上文详细描述中提及了用于动作执行的设备的若干模块或者单元,但是这种划分并非强制性的。实际上,根据本公开的实施方式,上文描述的两个或更多模块或者单元的特征和功能可以在一个模块或者单元中具体化。反之,上文描述的一个模块或者单元的特征和功能可以进一步划分为由多个模块或者单元来具体化。

[0146] 在本公开的示例性实施例中,还提供了一种能够实现所述方法的电子设备。

[0147] 所属技术领域的技术人员能够理解,本公开的各个方面可以实现为系统、方法或程序产品。因此,本公开的各个方面可以具体实现为以下形式,即:完全的硬件实施方式、完全的软件实施方式(包括固件、微代码等),或硬件和软件方面结合的实施方式,这里可以统称为“电路”、“模块”或“系统”。

[0148] 下面参照图12来描述根据本公开的这种实施方式的电子设备1200。图12显示的电子设备1200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。

[0149] 如图12所示,电子设备1200以通用计算设备的形式表现。电子设备1200的组件可以包括但不限于:所述至少一个处理单元1210、所述至少一个存储单元1220、连接不同系统组件(包括存储单元1220和处理单元1210)的总线1230。

[0150] 其中,所述存储单元存储有程序代码,所述程序代码可以被所述处理单元1210执行,使得所述处理单元1210执行本说明书所述“示例性方法”部分中描述的根据本公开各种示例性实施方式的步骤。例如,所述处理单元1210可以执行如本公开实施例所示的方法。

[0151] 存储单元1220可以包括易失性存储单元形式的可读介质,例如随机存取存储单元(RAM) 12201和/或高速缓存存储单元12202,还可以进一步包括只读存储单元(ROM) 12203。

[0152] 存储单元1220还可以包括具有一组(至少一个)程序模块12205的程序/实用工具12204,这样的程序模块12205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。

[0153] 总线1230可以为表示几类总线结构中的一种或多种,包括存储单元总线或者存储单元控制器、外围总线、图形加速端口、处理单元或者使用多种总线结构中的任意总线结构的局域总线。

[0154] 电子设备1200也可以与一个或多个外部设备1240(例如键盘、指向设备、蓝牙设备等)通信,还可与一个或者多个使得用户能与该电子设备1200交互的设备通信,和/或与使得该电子设备1200能与一个或多个其它计算设备进行通信的任何设备(例如路由器、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1250进行。并且,电子设备1200还可以通过网络适配器1260与一个或者多个网络(例如局域网(LAN),广域网(WAN)和/或公共网络,例如因特网)通信。如图所示,网络适配器1260通过总线1230与电子设备1200的其它模块通信。应当明白,尽管图中未示出,可以结合电子设备1200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。

[0155] 通过以上的实施方式的描述,本领域的技术人员易于理解,这里描述的示例实施方式可以通过软件实现,也可以通过软件结合必要的硬件的方式来实现。因此,根据本公开实施方式的技术方案可以以软件产品的形式体现出来,该软件产品可以存储在一个非易失性存储介质(可以是CD-ROM,U盘,移动硬盘等)中或网络上,包括若干指令以使得一台计算设备(可以是个人计算机、服务器、终端装置、或者网络设备等)执行根据本公开实施方式的方法。

[0156] 在本公开的示例性实施例中,还提供了一种计算机可读存储介质,其上存储有能够实现本说明书所述方法的程序产品。在一些可能的实施方式中,本公开的各个方面还可以实现为一种程序产品的形式,其包括程序代码,当所述程序产品在终端设备上运行时,所述程序代码用于使所述终端设备执行本说明书所述“示例性方法”部分中描述的根据本公开各种示例性实施方式的步骤。

[0157] 根据本公开的实施方式的用于实现所述方法的程序产品可以采用便携式紧凑盘只读存储器(CD-ROM)并包括程序代码,并可以在终端设备,例如个人电脑上运行。然而,本公开的程序产品不限于此,在本文件中,可读存储介质可以是任何包含或存储程序的有形介质,该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

[0158] 所述程序产品可以采用一个或多个可读介质的任意组合。可读介质可以是可读信号介质或者可读存储介质。可读存储介质例如可以为但不限于电、磁、光、电磁、红外线、或半导体的系统、装置或器件,或者任意以上的组合。可读存储介质的更具体的例子(非穷举的列表)包括:具有一个或多个导线的电连接、便携式盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、光纤、便携式紧凑盘只读存储器(CD-ROM)、光存储器件、磁存储器件、或者所述的任意合适的组合。

[0159] 计算机可读信号介质可以包括在基带中或者作为载波一部分传播的数据信号,其中承载了可读程序代码。这种传播的数据信号可以采用多种形式,包括但不限于电磁信号、光信号或所述的任意合适的组合。可读信号介质还可以是可读存储介质以外的任何可读介质,该可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

[0160] 可读介质上包含的程序代码可以用任何适当的介质传输,包括但不限于无线、有线、光缆、RF等等,或者所述的任意合适的组合。

[0161] 可以以一种或多种程序设计语言的任意组合来编写用于执行本公开操作的程序代码,所述程序设计语言包括面向对象的程序设计语言—诸如Java、C++等,对所述监督数据的样本分布进行最大似然估计包括常规的过程式程序设计语言—诸如“C”语言或类似的设计语言。程序代码可以完全地在用户计算设备上执行、部分地在用户设备上执行、作为一个独立的软件包执行、部分在用户计算设备上部分在远程计算设备上执行、或者完全在远程计算设备或服务器上执行。在涉及远程计算设备的情形中,远程计算设备可以通过任意种类的网络,包括局域网(LAN)或广域网(WAN),连接到用户计算设备,或者,可以连接到外部计算设备(例如利用因特网服务提供商来通过因特网连接)。

[0162] 此外,所述附图仅是根据本公开示例性实施例的方法所包括的处理的示意性说明,而不是限制目的。易于理解,所述附图所示的处理并不表明或限制这些处理的时间顺序。另外,也易于理解,这些处理可以是例如在多个模块中同步或异步执行的。

[0163] 本领域技术人员在考虑说明书及实践这里公开的发明后,将容易想到本公开的其他实施方案。本申请旨在涵盖本公开的任何变型、用途或者适应性变化,这些变型、用途或者适应性变化遵循本公开的一般性原理并包括本公开未公开的本技术领域中的公知常识或惯用技术手段。说明书和实施例仅被视为示例性的,本公开的真正范围和构思由权利要求指出。

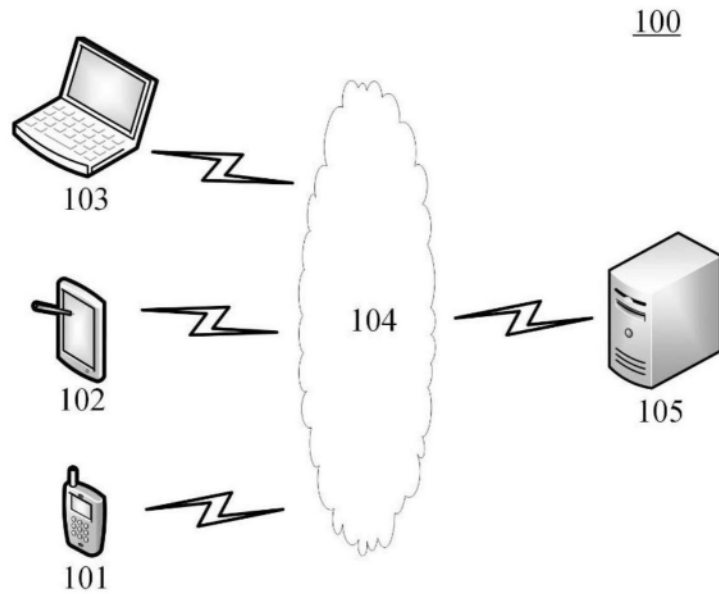


图1

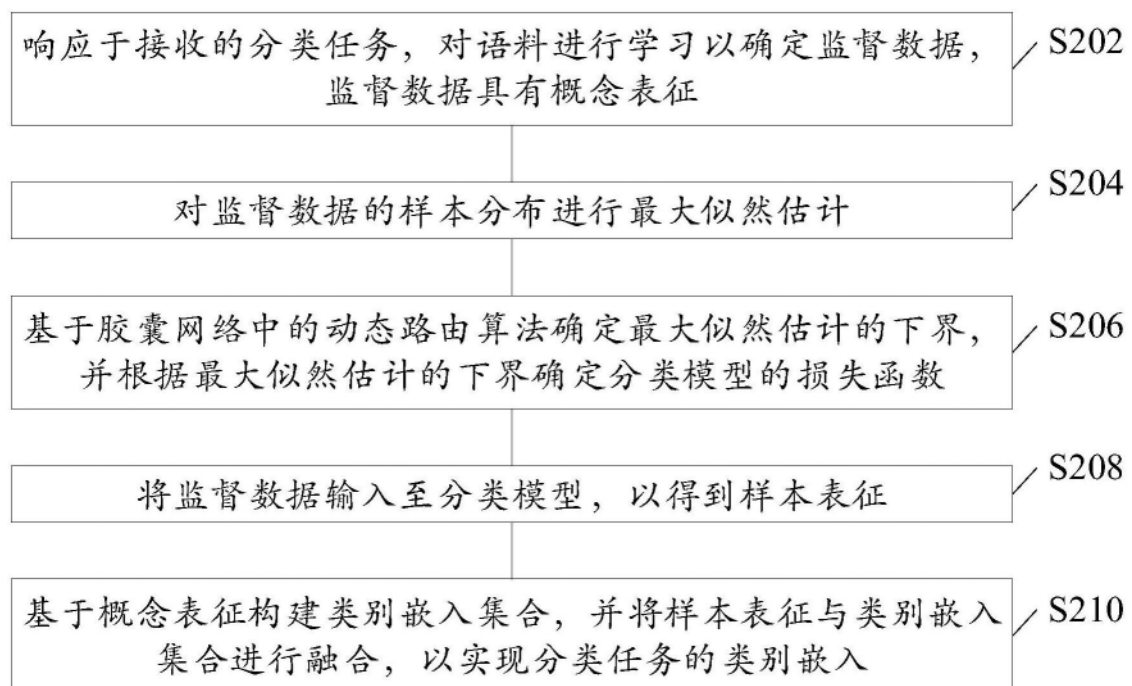


图2

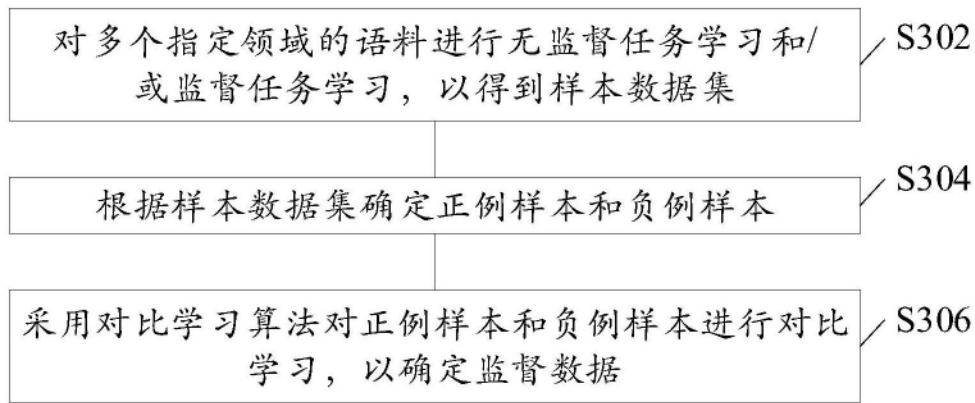


图3

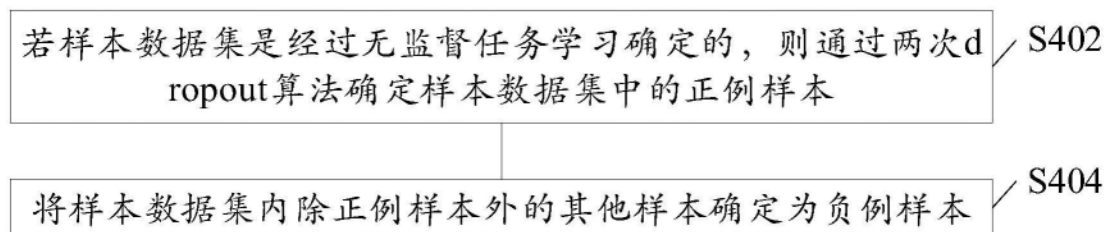


图4

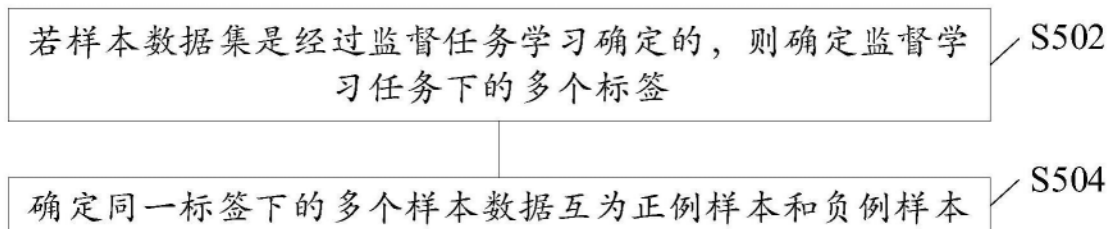


图5

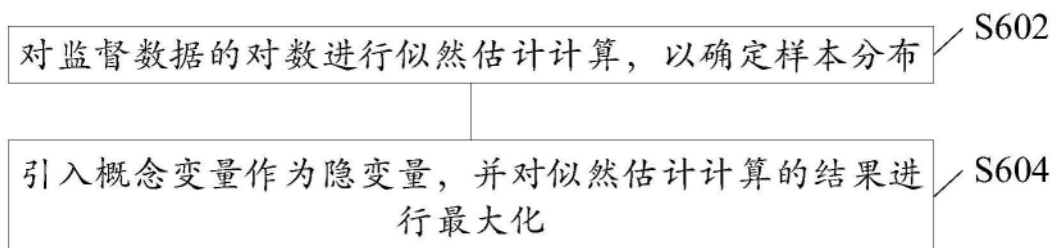


图6



图7

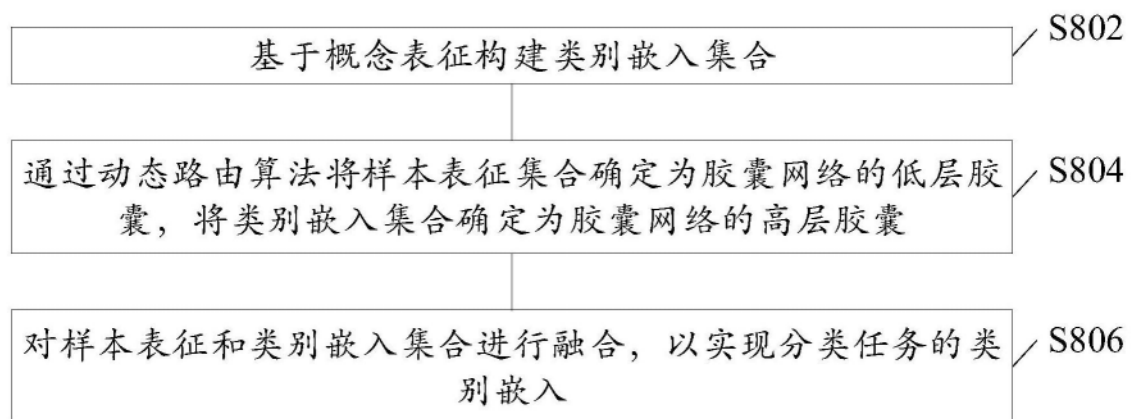


图8



图9

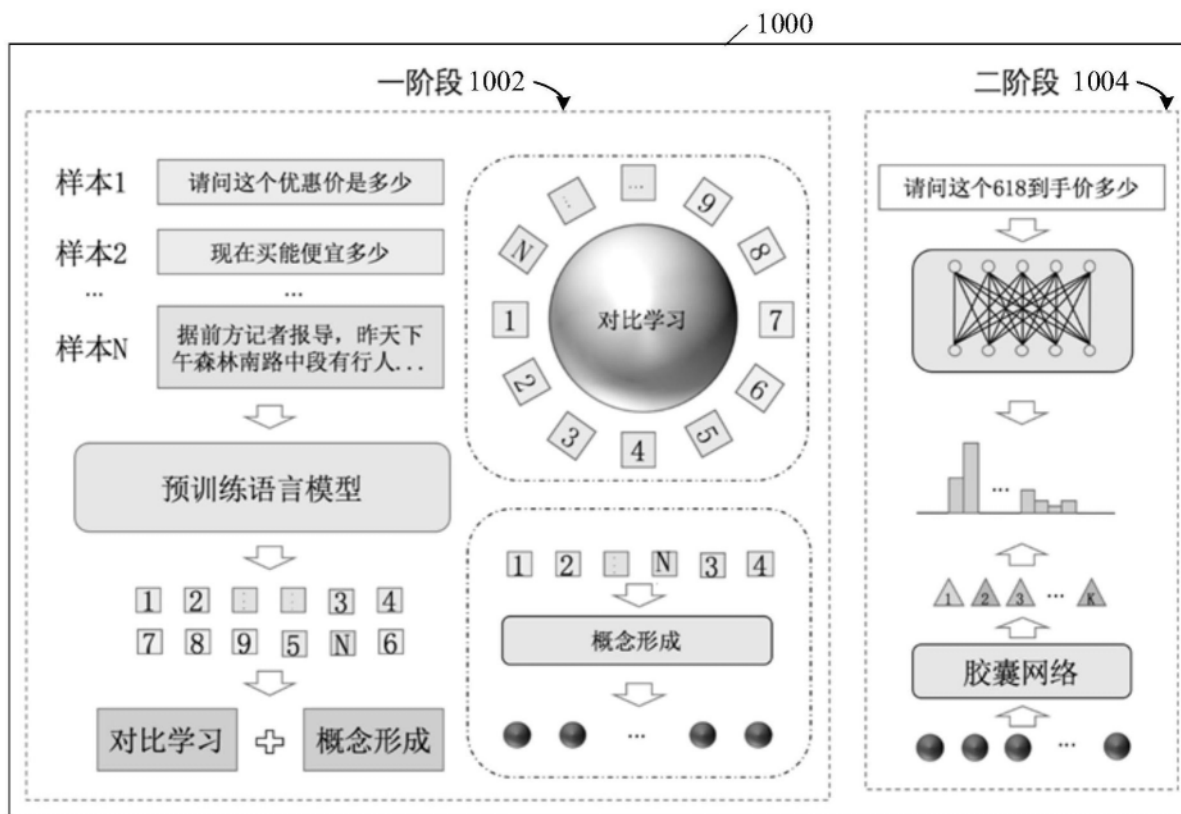


图10

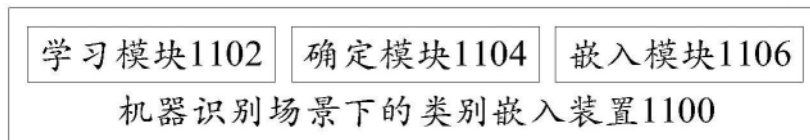


图11

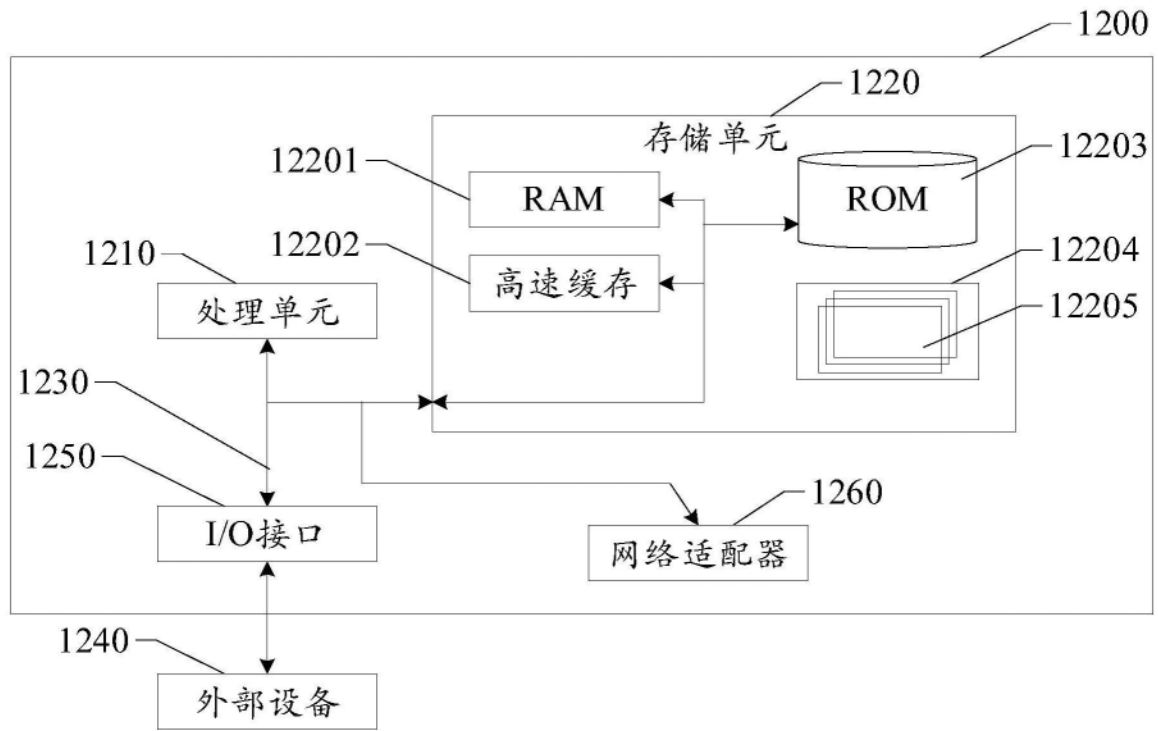


图12