



## 第3章

# 适应型深度学习模型

深度学习模型不是万能的,在有些场景或应用中会出现局限性,因此需要提高现有深度学习模型的适应性。本章给出了两种适应型深度学习模型:期望与反期望深度学习模型、残缺数据深度学习模型。

### 3.1 期望与反期望深度学习模型

深度学习模型通过输入数据和输出数据对深度学习模型进行训练。如果有的输入数据在不同场景下能产生属性相反的输出数据(这是有可能的,假如还有影响输出结果的其他条件数据没有包括在输入数据中,那么就有可能因为其他条件数据的变化,而产生不同的输出数据),此时就会在训练过程中引起深度学习模型结构的动荡,不利于深度学习模型的训练,也不利于深度学习模型的使用,因为无法得知输出数据的可信度是多少<sup>[11,12]</sup>。

例如,深度学习模型判断一个人是不是男的,输入不同人的的人脸特征,同样脸部特征的人可能是男的,也可能是女的,此时就会在训练过程中引起深度学习模型结构的动荡。

本技术研究并给出了一种新的期望与反期望深度学习方法和神经网络系统,通过输出数据的期望标签和反期望标签构造两个深度学习模型,分别对与期望标签对应的深度学习模型和与反期望标签对应的深度学习模型进行训练,避免了深度学习模型当存在属性相反的输出数据时结构动荡的问题,提高了深度学习模型训练的可靠性。

### 3.1.1 期望与反期望深度学习方法

如图 3-1 所示,本技术方案提供一种期望与反期望深度学习方法,可包括以下步骤:

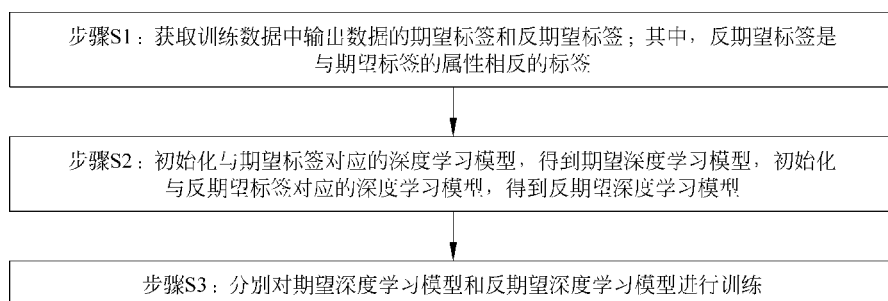


图 3-1 期望与反期望深度学习方法流程

(1) 步骤 S1,获取训练数据中输出数据的期望标签和反期望标签;其中,反期望标签是与期望标签的属性相反的标签。

例如,获取输出数据的期望标签“男人”,反期望标签“女人”。

(2) 步骤 S2,初始化与期望标签对应的深度学习模型,得到期望深度学习模型;初始化与反期望标签对应的深度学习模型,得到反期望深度学习模型。

其中,期望标签对应的深度学习模型的输入格式初始化为训练数据中的输入数据格式;反期望标签对应的深度学习模型的输入格式也初始化为训练数据中的输入数据格式。期望深度学习模型的输出格式初始化为 $0\sim 1$ 之间的数,输出为1时,表示输出为期望标签;输出为0时,表示输出不是期望标签;输出为 $0\sim 1$ 时,表示输出有可能是期望标签。反期望深度学习模型的输出格式初始化为 $0\sim 1$ 的数,输出为1时,表示输出为反期望标签;输出为0时,表示输出不是反期望标签;输出为 $0\sim 1$ 时,表示输出有可能是反期望标签。同时,获取已有同类深度学习模型的配置信息(预设配置信息包括:预设层数、每层预设节点数、各网络连接的预设权值)作为期望标签对应的深度学习模型的配置信息对期望标签对应的深度学习模型进行配置,同时也将该配置信息作为反期望标签对应的深度学习模型的配置信息对反期望标签对应的深度学习模型进行配置。

例如,初始化与期望标签“男人”和反期望标签“女人”对应的两个深度学习模型,称为期望标签“男人”深度学习模型、反期望标签“女人”深度学习模型。

(3) 步骤 S3,分别对期望深度学习模型和反期望深度学习模型进行训练。

在一个技术方案中,本步骤可以采用以下方式实现:

① 步骤 S3-1,通过训练数据中的输入数据分别对期望深度学习模型和反期望深度学习模型进行无监督训练。

例如,通过训练数据中的输入数据人脸图像分别对期望标签“男人”深度学习模型、反期望标签“女人”深度学习模型进行无监督训练。对期望深度学习模型、反期望深度学习模型进行无监督训练,可采用同样的输入数据集合。

② 步骤 S3-2,从训练数据中获取与期望标签一致的输出标签对应的输入数据,将该输入数据作为输入,将“1”作为预期输出,对期望深度学习模型进行有监督训练,并从训练数据中获取与期望标签及反期望标签都不一致的输出标签对应的输入数据,将该输入数据作为输入,将“0”作为预期输出,对期望深度学习模型进行有监督训练。

可以将训练数据中与期望标签一致的输出标签对应的输入数据作为第一输入数据;从第一输入数据中筛选出第一预设比例(假设为  $P\%$ ,  $P$  为  $0\sim 100$  的实数)的第一输入数据,得到对应的每一第二输入数据(即将第一输入数据中  $P\%$  的数据保留,其余数据清空,得到对应的每一第二输入数据);将每一第二输入数据作为输入,将第一预设比例作为预期输出,对期望深度学习模型进行有监督训练。

之所以不进行从训练数据中获取与反期望标签一致的输出标签对应的输入数据,将该输入数据作为输入,将“0”作为预期输出,对期望深度学习模型进行有监督训练,是因为与反期望标签一致的输出标签所对应的输入数据有可能与期望标签一致的输出标签所对应的输入数据类似(例如,有的男人和女人的脸部特征很相像),从而导致类似的输入数据在同一个深度学习模型中产生不同的输出标签,进而会对期望深度学习模型的训练效果产生负面影响。

例如,从训练数据中获取与期望标签“男人”一致的输出标签(预期输出为“1”)及其对应输入数据人脸图像对期望标签“男人”深度学习模型进行有监督训练;从训练数据中获取与期望标签“男人”及反期望标签“女人”都不一致的输出标签(预期输出为“0”)及其对应输入数据(例如动物头像)对期望标签“男人”深度学习模型进行有监督训练;从训练数据中获取与期望标签“男人”一致的输出标签(预期输出为“1”)及其对应输入数据人脸图像,将输入数据人脸图像中  $60\%$  的部分保留(可随机选择,也可以均匀选择,或按照一定预设方式选择需要保留的部分),其余  $40\%$  的部分从图像中清除(可随机选择,也可以均匀选择,或按照一定预设方式选择需要保留的部分),然后将处理后的输入数据人脸图像作为新输入数据人脸图像,将输出数据“1”变为新输出数据“ $60\%$ ”,对期望标签“男人”深度学习模型进行有监督训练。

③ 步骤 S3-3,从训练数据中获取与反期望标签一致的输出标签对应的输入数据,将该输入数据作为输入,将“1”作为预期输出,对反期望深度学习模型进行有监督训练,并从训练数据中获取与期望标签及反期望标签都不一致的输出标签对应的输入数据,将该输入数据作为输入,将“0”作为预期输出,对反期望深度学习模型进行有监督训练。

可以将训练数据中与反期望标签一致的输出标签对应的输入数据作为第三输入数据;从每一第三输入数据中筛选出第二预设比例(假设为  $P\%$ ,  $P$  为  $0\sim 100$  的实数)的数据,得到对应的每一第四输入数据(即将每一第三输入数据中  $P\%$  的数据保留,其余数据清空,得到对应的每一第四输入数据);将每一第四输入数据作为输入,将对应的第二预设比例作为预期输出,对反期望深度学习模型进行有监督训练。

例如,从训练数据中获取与反期望标签“女人”一致的输出标签(预期输出为“1”)及其对应输入数据人脸图像对反期望标签“女人”深度学习模型进行有监督训练;从训练数据中获取与期望标签“男人”及反期望标签“女人”都不一致的输出标签(预期输出为“0”)及其对应输入数据(例如动物头像)对期望标签“女人”深度学习模型进行有监督训练。

从训练数据中获取与反期望标签“女人”一致的输出标签(预期输出为“1”)及其对应输

入数据人脸图像,将输入数据人脸图像中 60% 的部分保留(可随机选择,也可以均匀选择,或按照一定预设方式选择需要保留的部分),其余 40% 的部分从图像中清除(可随机选择,也可以均匀选择,或按照一定预设方式选择需要保留的部分),然后将处理后的输入数据人脸图像作为新输入数据人脸图像,将输出数据“1”变为新输出数据“60%”,对反期望标签“女人”深度学习模型进行有监督训练。

在分别对期望深度学习模型和反期望深度学习模型进行训练之后,还可以计算输入数据对应的输出属于期望标签及反期望标签的可信度。可以将输入数据输入期望深度学习模型,得到期望深度学习模型的输出数据;将输入数据输入反期望深度学习模型,得到反期望深度学习模型的输出数据;根据期望深度学习模型与反期望深度学习模型的输出数据得到输入数据对应的输出属于期望标签及反期望标签的可信度。

例如,获取输入数据人脸图像,将输入数据人脸图像输入期望深度学习模型,得到期望深度学习模型的输出数据,该输出数据为 0~1 的数,越接近 1,则表明输出为期望标签的概率越大;将输入数据输入反期望深度学习模型,得到反期望深度学习模型的输出数据,该输出数据为 0~1 的数,越接近 1,则表明输出为反期望标签的概率越大。

在计算可信度时,可以将期望深度学习模型的输出数据作为输出属于期望标签的概率记为  $A$ ,将反期望深度学习模型的输出数据作为输出属于反期望标签的概率记为  $B$ ;将输入数据对应的输出属于期望标签或反期望标签的可信度记为  $F$ ,则  $F = (A + B) / 2$ ;输入数据对应的输出属于期望标签的可信度为  $F \times A$ ,属于相反标签的可信度为  $F \times B$ ,属于其他标签的可信度为  $1 - F$ 。

例如,如果这个人是男人,那么这个人是男人的概率是  $F \times A$ ,作为这个人是男人的可信度;是女人的概率是  $F \times B$ ,作为这个人是女人的可信度;属于其他标签的概率为  $1 - F$ ,作为这个人既不是男人又不是女人的可信度。

深度学习模型判断一个人是不是男人,训练数据中输入头像图像,如果训练数据中输出数据是男人,则用“男人”标签对应的深度学习模型进行训练,例如用输出“1”代表是男人,用“0”代表不是男人(例如是动物);如果训练数据中输出数据是女人,则用“女人”标签对应的深度学习模型进行训练,例如用输出用 0~1 的数来代表是女人的可能性大小,用 1 代表是女人,用 0 代表不是女人(例如是动物);如果训练数据中输出数据是跟男人、女人无关的,例如动物,则“男人”标签对应的深度学习模型和“女人”标签对应的深度学习模型同时进行训练。

在使用时,输入一个数据到“男人”和“女人”标签对应的深度学习模型同时计算,如果“男人”标签对应的深度学习模型输出 0.8;如果“女人”标签对应的深度学习模型输出 0.5,则输出是“男人”或“女人”的可信度是  $(0.8 + 0.5) / 2 = 0.65 = 65\%$ ,输出不是“男人”或“女人”的可信度是  $1 - 65\% = 35\%$ ,输出是“男人”的可信度是  $65\% \times 0.8 = 52\%$ ,输出是“女人”的可信度是  $65\% \times 0.5 = 32.5\%$ ,相对而言是“男人”的可信度更高,所以判断为“男人”。

### 3.1.2 期望与反期望深度学习系统

如图 3-2 所示,本技术方案还提供一种神经网络系统,可包括:

(1) 标签获取模块 1,用于获取训练数据中输出数据的期望标签和反期望标签;其中,

反期望标签是与期望标签的属性相反的标签。

例如,获取输出数据的期望标签“男人”和反期望标签“女人”。

(2) 初始化模块 2,用于初始化与期望标签对应的深度学习模型,得到期望深度学习模型,初始化与反期望标签对应的深度学习模型,得到反期望深度学习模型。

其中,期望标签对应的深度学习模型的输入格式初始化为训练数据中的输入数据格式;反期望标签对应的深度学习模型的输入格式也初始化为训练数据中的输入数据格式。期望深度学习模型的输出格式初始化为 0~1 的数,输出为“1”时,表示输出为期望标签;输出为“0”时,表示输出不是期望标签;输出为 0~1 的数时,表示输出有可能是期望标签。反期望深度学习模型的输出格式初始化为 0~1 的数,输出为“1”时,表示输出为反期望标签,输出为“0”时,表示输出不是反期望标签;输出为 0~1 的数时,表示输出有可能是反期望标签。同时,获取已有同类深度学习模型的配置信息(预设配置信息包括:预设层数、每层预设节点数、各网络连接的预设权值)作为期望标签对应的深度学习模型的配置信息对期望标签对应的深度学习模型进行配置,同时也将该配置信息作为反期望标签对应的深度学习模型的配置信息对反期望标签对应的深度学习模型进行配置。

例如,初始化与期望标签“男人”和反期望标签“女人”对应的两个深度学习模型,称为期望标签“男人”深度学习模型、反期望标签“女人”深度学习模型。

(3) 训练模块 3,用于分别对期望深度学习模型和反期望深度学习模型进行训练。

本技术方案通过输出数据的期望标签和反期望标签构造两个深度学习模型,如果训练数据中输出数据与期望标签一致,则对与期望标签对应的深度学习模型进行训练;如果训练数据中输出数据与反期望标签一致,则对与反期望标签对应的深度学习模型进行训练;如果训练数据中输出数据与属性无关(当然也与反期望标签无关),则同时对期望标签及反期望标签对应的两个深度学习模型进行训练。训练得到两个深度学习模型,在应用中将输入数据同时输入这两个深度学习模型,得到两个输出数据,然后综合这两个输出数据得到输出数据属于期望标签及反期望标签的可信度。

本技术方案神经网络系统与期望和反期望深度学习方法一一对应,上述期望与反期望深度学习方法的技术方案阐述的技术特征及其有益效果均适用于神经网络系统的技术方案中。

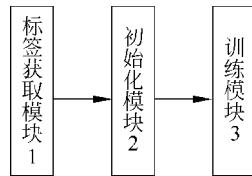


图 3-2 深度学习系统的结构示意图

## 3.2 残缺深度学习模型

人工神经网络(Artificial Neural Network, ANN)也简称为神经网络或连接模型,它是一种模仿动物神经网络行为特征,进行分布式并行信息处理的算法数学模型。这种网络依靠系统的复杂程度,通过调整内部大量节点之间的相互连接关系,从而达到处理信息的目的。

传统技术中,神经网络包括用于深度学习的神经网络都是通过输入数据和输出数据对神经网络进行训练,然后在使用阶段将输入数据输入已经训练好的神经网络,从而得到输出

数据。但是,在实际使用过程中,输入数据往往都不完整,因此基于传统技术的神经网络无法根据不完整的输入数据得到输出数据<sup>[13,14]</sup>。

本技术方案研究并给出了一种新型的残缺数据深度学习模型方法、系统,方法包括:获取预先训练好的深度学习模型;从待预测数据中获取输入数据;根据输入数据和预先训练好的深度学习模型的输入变量确定至少一个缺少数据的输入变量;将对缺少数据的输入变量进行赋值获得的数据,与从待预测数据中获取的输入数据进行组合,生成多组完整输入数据;将多组完整输入数据输入到预先训练好的深度学习模型中,计算得到多组输出数据。通过这种方式,在输入数据缺少数据的情况下,仍然能够通过神经网络输出预测的输出变量,为研究工作提供数据支持,减少了输入数据检测和采集的成本。

### 3.2.1 残缺深度学习模型方法

#### 1. 残缺深度学习模型的基本方法

如图 3-3 所示,本技术方案提供了一种深度学习模型方法,该方法以应用于服务器进行举例说明。包括:

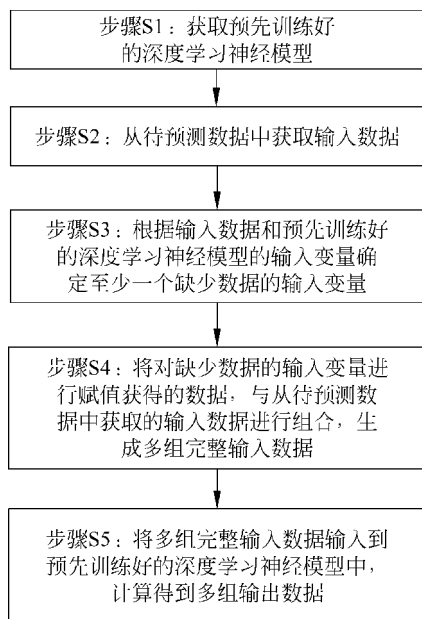


图 3-3 深度学习模型方法的流程示意图

- (1) 步骤 S1,获取预先训练好的深度学习模型。
- (2) 步骤 S2,从待预测数据中获取输入数据。
- (3) 步骤 S3,根据输入数据和预先训练好的深度学习模型的输入变量确定至少一个缺少数据的输入变量。

预先训练好的深度学习模型有  $N$  个输入变量。待预测数据中有其中  $M$  个输入变量对应的数据,其余  $N-M$  个输入变量没有对应的数据,则这  $N-M$  个输入变量是缺少数据的

输入变量；其中， $M$ 、 $N$  为自然数。

例如，深度学习模型输入变量为：身高、体重、血压、血液检查结果、尿液检查结果；深度学习模型输出是糖尿病类型（例如，用 0、1、2、3 来表示不同类型的糖尿病）。

待测试的不完整输入数据包括血压数据、血液检查结果数据、尿液检查结果数据，缺少数据输入变量为身高、体重。

(4) 步骤 S4，将对缺少数据的输入变量进行赋值获得的数据，与从待预测数据中获取的输入数据进行组合，生成多组完整输入数据。

通过知识库获取缺少数据的输入变量中每一缺少数据的输入变量的取值范围。具体包括 3 种方式：

第 1 种方式：从输入变量范围知识库中检索出缺少数据输入变量对应的预设取值范围；其中，输入变量范围知识库预先存有所有输入变量对应的预设取值范围。

第 2 种方式：提示并接受用户输入缺少数据输入变量的预设取值范围。

第 3 种方式：从每一训练输入数据中获取缺少数据输入变量的值，统计获取到的缺少数据输入变量的最大值和最小值，将最小值到最大值的范围作为缺少数据输入变量的预设取值范围。

例如，身高的取值范围是 0.1~2m，体重的取值范围是 2~100kg。

知识库中存储有多组数据，其中包括多组预先训练好的深度学习模型输入数据。可通过知识库中存储的数据获取缺少数据输入数据的取值范围。

对缺少数据输入变量在对应的预设取值范围内进行  $N$  次赋值，具体包括两种实现方法：

第 1 种实现方法：对缺少数据输入变量在对应的预设取值范围内随机地取  $N$  个值对缺少数据输入变量进行  $N$  次赋值。

第 2 种实现方法：从所有训练输入数据中随机抽取  $N$  个训练输入数据，然后提取其中该输入变量对应的数据对缺少数据输入变量进行  $N$  次随机赋值。

其中，预设次数  $N$  的确定方法包括：

接受用户输入预设次数  $N$  的值，如果用户没有输入，则将训练输入数据的数量作为  $N$  的值。

例如，对缺少数据输入变量身高、体重在对应的预设取值范围 0.1~2m、2~100kg 内复制其他样本中该输入变量数据进行 100 次随机赋值，因为在很多样本中是具备身高、体重数据的，所以通过获取已有其他样本中这些输入变量的数据是可行的：

身高 0.5m，体重 10kg；

身高 1.5m，体重 50kg；

……

身高 1.7m，体重 55kg；

将该缺少数据输入变量身高、体重对应的 100 份输入数据与待测试的不完整输入数据“血压数据  $C$ 、血液检查结果数据  $D$ 、尿液检查结果数据  $E$ ”组合得到 100 份完整的输入数据：

身高 0.5m，体重 10kg，血压数据  $C$ ，血液检查结果数据  $D$ ，尿液检查结果数据  $E$ ；

身高 1.5m，体重 50kg，血压数据  $C$ ，血液检查结果数据  $D$ ，尿液检查结果数据  $E$ ；

……

身高 1.7m, 体重 55kg, 血压数据  $C$ , 血液检查结果数据  $D$ , 尿液检查结果数据  $E$ 。

通过知识库中存储的数据获取缺少输入数据的取值范围后, 在此取值范围内取预定数量的数据, 将选取的数据赋值给缺少数据的输入变量。预定数量是预先设置的数量值, 可根据不同的输入变量设定不同的预定数量。选择  $M$  个数量的数据对缺少数据的输入变量进行赋值, 就能得到  $M$  个缺少数据的输入变量的输入数据, 再将填入了数据的缺少数据的输入变量与从待预测数据中获取的输入数据进行组合, 生成多组完整的输入数据。

(5) 步骤 S5, 将多组完整输入数据输入到预先训练好的深度学习模型中, 计算得到多组输出数据。

将缺少数据的输入变量缺少数据的输入数据补充完整后, 与从待预测数据中获取的输入数据进行组合, 生成多组完整的输入数据。再将多组完整的输入数据输入到预先训练好的深度学习模型中, 则能够获得多个输出数据。

例如, 将 100 份完整的输入数据输入已训练的深度学习模型, 得到多份输出:

身高 0.5m, 体重 10kg, 血压  $C$ , 血液检查结果  $D$ , 尿液检查结果  $E$ ; 该份完整的输入数据输入已训练的深度学习模型得到输出  $Y_1$ ;

身高 1.5m, 体重 50kg, 血压  $C$ , 血液检查结果  $D$ , 尿液检查结果  $E$ ; 该份完整的输入数据输入已训练的深度学习模型得到输出  $Y_2$ ;

.....

身高 1.7m, 体重 55kg, 血压  $C$ , 血液检查结果  $D$ , 尿液检查结果  $E$ ; 该份完整的输入数据输入已训练的深度学习模型得到输出  $Y_{100}$ 。

通过这种方式, 在输入数据缺少数据的情况下, 仍然能够通过神经网络输出预测的输出变量, 为研究工作提供数据支持。

## 2. 残缺深度学习模型的详细方法

在一个技术方案中, 将对缺少数据的输入变量进行赋值获得的数据, 与从待预测数据中获取的输入数据进行组合, 生成多组完整输入数据, 包括: 从输入变量范围知识库中检索出缺少数据输入变量对应的预设取值范围, 其中, 输入变量范围知识库预先存有所有输入变量对应的预设取值范围; 或从每一训练输入数据中获取缺少数据输入变量的值, 统计获取缺少数据输入变量的最大值和最小值, 将从最小值到最大值之间值的范围作为缺少数据输入变量的预设取值范围; 或提示并接受用户输入缺少数据输入变量的预设取值范围, 对缺少数据输入变量在对应的预设取值范围内随机地生成预定数量个值对缺少数据的输入变量进行预定数量次赋值, 其中, 预定数量由用户输入或预先设置。

当待预测数据中缺少一个或多个输入变量的数据时, 则需要从知识库存储的对应数据中选取预定数量的数据对缺少数据的输入变量进行赋值。预定数量可以是用户通过输入设备输入的, 当用户未输入时, 预定数量可以根据预设比例计算。具体而言, 先根据缺少数据的输入变量从知识库中查找缺少数据输入数据数量, 例如缺少数据的输入变量为体重, 那么就从知识库中查找体重这一项的数据, 当体重这一项的数据有 120 项时, 那么知识库中缺少数据的输入变量为体重的对应数据数量则为 120。再获取到预设比例为  $2/3$ , 那么预定数量则等于 120 乘以  $2/3$ , 即预定数量为 80。当计算出来的预定数量不是整数时, 则可以通过四舍五入或取整的方式取整数值。预设比例的设定可由研究人员根据研究需求而定。预设比例的设定较灵活, 可根据实际项目需求进行调整, 根据预设比例与总体数据数量进行预设数



量的确定,也能保证选取数据的数量,提高输出数据的总体准确性。

在一个技术方案中,在将多组完整输入数据输入到预先训练好的深度学习模型中,计算得到多组输出数据之后,统计多组输出数据中各类输出数据所占的比例;将比例最高的一类输出数据作为预测结果;将预测结果及各类输出数据及其相应比例输出给用户。

在得到多个输出数据之后,需要分别统计各类输出数据所占的比例,再将比例最高的输出数据作为预测结果。

例如,输入变量为体重、身高、血压、血液检查结果,输出变量为糖尿病严重程度等,当输入多组体重、身高、血压、血液检查结果的数据后,能获得多组输出的关于糖尿病严重程度的数据。当输出的糖尿病严重程度数据为“0”的有10个,为“1”的有30个,为“2”的有15个时,则将占比例最高的为“1”的数据作为预测结果。其中,数据为“0”代表没有糖尿病,数据为“1”代表有轻微糖尿病,数据为“2”代表有严重糖尿病。具体的数据输出可由研究人员进行自定义设置。

例如,根据 $Y_1, Y_2, \dots, Y_{100}$ 的值进行统计,其中输出为0的有10个,为“1”的有30个,为“2”的有60个。其中“0”代表没有糖尿病,“1”代表有轻微糖尿病,“2”代表有严重糖尿病,则输出为“0”(代表没有糖尿病)的比例是10%,输出为“1”(代表轻微糖尿病)的比例是30%,输出为“2”(代表严重糖尿病)的比例是60%。

将多个输出数据及其对应比例输出给用户,作为不完整输入数据的输出数据;将最高比例对应的输出数据输出给用户,作为不完整输入数据的最大概率输出数据。

例如,将输出为“0”(代表没有糖尿病)的比例是10%,输出为“1”(代表轻微糖尿病)的比例是30%,输出为“2”(代表严重糖尿病)的比例是60%输出给用户,作为不完整输入数据血压 $C$ 、血液检查结果 $D$ 、尿液检查结果 $E$ 的输出。

在医疗行业,虽然有诊疗神经网络模型,但输入变量很多,例如糖尿病神经网络模型,输出是无糖尿病、 $X$ 类糖尿病、 $Y$ 类糖尿病等(实际中 $X$ 、 $Y$ 有具体名称),输入为血压、身高、体重、血液检查结果(实际中包括多项具体指标)、尿液检查结果(实际中包括多项具体指标)。但患者刚开始来的时候,可能其中血液没有检查,那么在现有技术中就无法使用该模型了,而采用本技术方案可以通过已有的患者输入数据,大概地判别各种疾病的可能性,进行预诊和分诊,假如通过现有残缺的数据就足以判断出想要的结果,例如,现有数据输入后,发现输出的结果只有无糖尿病这一类,那么就无须再多做检查,反而节省了输入数据检测和采集的成本;或者看到输出为“2”(代表严重糖尿病)的比例是60%,说明患严重糖尿病的可能性最大,则需要进一步做更深入的检查。

当输出数据有多个时,则需要对输出结果进行筛选,将比例较高的数据作为预测输出数据,更能确保预测数据的准确性。

在一个技术方案中,将对缺少数据的输入变量进行赋值获得的数据,与待预测数据中获取的输入数据进行组合,生成多组完整输入数据,包括:从样本数据库中随机选取预定数量的训练输入数据,然后提取这些训练输入数据中该输入变量对应的数据对缺少数据输入变量进行预定数量次赋值;其中,预定数量由用户输入或将小于等于样本数据库中输入数据的数量的 $P\%$ 作为预定数量, $P$ 小于等于100,样本数据库中的输入数据都是完整的输入数据。

在一个技术方案中,在从待预测数据中获取输入变量之前,对深度学习模型进行初始

化；获取训练数据；通过训练数据对初始化后的深度学习模型进行训练得到预先训练好的深度学习模型。

深度学习模型在根据输入数据预测输出数据之前，需要先获取深度学习模型并对其进行初始化，再获取训练数据，通过训练数据对初始化后的深度学习模型进行训练，训练后则能获得训练好的深度学习模型，训练好的深度学习模型才能用于对输出数据的预测操作。深度学习模型可以自行创建，也可以为了节约创建时间，直接从深度学习模型库中获取匹配度较高的深度学习模型，再对获取到的深度学习模型进行初始化，初始化后的深度学习模型才能与当前研究项目需要的输入变量与输出变量更契合，而对深度学习模型训练后才能使得深度学习模型更准确地预测输出数据。

在一个技术方案中，对深度学习模型进行初始化，包括：根据样本数据库中训练数据的字段名确定样本输入变量与样本输出变量；从深度学习模型库中选取输入变量和输出变量与样本输入变量和样本输出变量匹配的深度学习模型；对选取的深度学习模型进行初始化。

在知识库中存储有多种类型的数据，每个类型的数据有对应的字段。例如，存储的数据中包括有体重、身高、血压这三种类型的数据，那么对应的字段可设为体重、身高、血压。再根据研究项目需求，从知识库中选择需要用到的字段，将选取的字段确认为样本输入变量与样本输出变量。例如，当项目需要研究根据身高、体重、血压、血液检测结果等数据预测该用户是否患有糖尿病和患病严重程度时，可从知识库中选择体重、血压、血液检测结果的字段作为样本输入变量，并选择糖尿病类型作为样本输出变量，则项目研究需要的深度学习模型的输入变量为体重、血压、血液检测结果，而输出变量则为糖尿病类型。

当确定了样本输入变量与样本输出变量后，可根据样本输入变量与样本输出变量从深度学习模型库中选取相匹配的深度学习模型。由于深度学习模型库中的神经网络模型数量较多，在选取与实际需求对应的神经网络模型时，应该先根据研究需求从知识库中选择对应的字段将项目需要的神经网络模型的输入变量与输出变量设定，然后再根据设定的输入变量与输出变量从深度学习模型库中查找与实际项目需求相匹配的神经网络模型。这样能减少选择神经网络模型的操作步骤，节约时间成本。

在一个技术方案中，从深度学习模型库中选取输入变量和输出变量与样本输入变量和样本输出变量匹配的深度学习模型，采用如下公式计算深度学习模型库中输入变量和输出变量分别与样本输入变量和样本输出变量之间的匹配度：

$$\text{匹配度} = \text{深度学习模型库中输入变量与样本输入变量的匹配度} \times u\% + \text{深度学习模型库中输出变量与样本输出变量的匹配度} \times (1-u\%)$$

其中， $u$  为预设权重。从匹配度中选取匹配度最大的深度学习模型返回给用户；将用户确认或修改后的深度学习模型作为预置深度学习模型。

当根据样本输入变量和样本输出变量从深度学习模型库中选择匹配的深度学习模型时，可将深度学习模型库中的多个深度学习模型的输入变量和输出变量分别与样本输入变量和样本输出变量进行对比匹配，再将两者的匹配度综合，选择综合匹配度最高的。具体而言，可以设置一个预定权重，例如设定权重为 50 时，则说明研究人员认为输入变量与输出变量的重要性是一致的。将深度学习模型库中的多个深度学习模型的输入变量与样本输入变量一一对比，获取输入变量与样本输入变量的匹配度。再将深度学习模型库中的多个深度

学习模型的输出变量与样本输出变量一一对比,获取输出变量与样本输出变量的匹配度。

例如,神经网络 A 的输入变量与样本输入变量的匹配度为  $4/5$ ,神经网络 A 的输出变量与样本输出变量的匹配度为  $3/4$ ,那么神经网络 A 的综合匹配度为  $4/5 \times 50\% + 3/4 \times (1 - 50\%) = 0.775$ 。当将深度学习模型库中的多个深度学习模型都一一匹配完毕后,统计每个深度学习模型与样本输入变量和样本输出变量的综合匹配度,并将匹配度最高的神经网络作为预置深度学习模型。因此,若是其他神经网络的综合匹配度低于  $0.775$  时,则选择匹配度最高的神经网络 A 作为预置深度学习模型。

将样本输入变量与样本输出变量均作为挑选标准,能在更大程度上保证“挑选”出的深度学习模型与实际项目需求的契合度,更能节约研究的时间成本。

在一个技术方案中,对初始化后的深度学习模型进行训练得到预先训练好的深度学习模型,包括:获取初始化后的深度学习模型中的输入变量与输出变量;从知识库中获取输入数据,将其输入到深度学习模型中进行无监督训练;从知识库中获取输入数据及输出数据,将输入数据作为初始化后的深度学习模型的输入,将输出数据作为初始化后的深度学习模型的预期输出,对初始化后的深度学习模型进行有监督训练。将每个训练输入数据作为深度学习模型的输入,对深度学习模型进行无监督训练;将每个带标签的训练输入数据及其对应的预期标签作为深度学习模型的输入和预期输出,对深度学习模型进行有监督训练。

当从深度学习模型库中选择了某一个深度学习模型后,会先对该深度学习模型进行初始化,得到初始化后的深度学习模型,初始化后的深度学习模型对应有输入变量与输出变量。然后对初始化后的深度学习模型进行训练,训练分为无监督训练和有监督训练。一般情况下,先进行无监督训练,再进行有监督训练。无监督训练是指从知识库中获取输入数据,将输入数据输入到初始化后的深度学习模型中进行无监督训练。无监督训练不需要预期输出,而有监督训练则需要预期输出。

例如,当输入变量为体重、身高、血压、血液检测结果、尿液检测结果时,从知识库中获取多组体重、身高、血压、血液检测结果、尿液检测结果字段对应的数据作为输入变量对应的数据,可以对初始化后的深度学习模型进行无监督训练。

在无监督训练后,再对深度学习模型进行有监督训练。进行有监督训练时,会将深度学习模型的输入数据与预期输出均提供。例如,当体重、身高、血压、血液检测结果、尿液检测结果为输入变量,糖尿病类型为输出变量时,从知识库中获取体重、身高、血压、血液检测结果、尿液检测结果字段对应的多组数据作为输入变量对应的数据,将糖尿病类型字段对应的数据作为输出变量对应的数据,对始化后的深度学习模型进行有监督训练。

用户样本数据中存储有多组用户数据。例如,每组用户数据中的数据类型包括有每个用户的体重、身高、血压、血液检测结果、尿液检测结果以及糖尿病类型数据。但是,并不是每组用户数据都是完整的,可能存在有的用户数据中缺少糖尿病类型数据,这部分缺少糖尿病类型数据的用户数据可以用于对深度学习模型进行无监督训练,避免了数据的浪费。由于无监督训练先训练了深度学习模型的特征提取能力,在进行了无监督训练后再进行有监督训练也能提升深度学习模型的训练效果,提升训练后的深度学习模型预测输出数据的准确度。

### 3.2.2 残缺深度学习模型系统

图 3-4 提供了一种深度学习模型装置,包括:

(1) 数据获取模块 1,用于获取预先训练好的深度学习模型,从待预测数据中获取输入数据。

(2) 缺少数据的输入变量确认模块 2,用于根据输入数据和预先训练好的深度学习模型的输入变量确定至少一个缺少数据的输入变量。

(3) 取值范围获取模块 3,用于将对缺少数据的输入变量进行赋值获得的数据,与从待预测数据中获取的输入数据进行组合,生成多组完整输入数据。

在一个技术方案中,取值范围获取模块 3 还用于从输入变量范围知识库中检索出缺少数据输入变量对应的预设取值范围,其中,输入变量范围知识库预先存有所有输入变量对应的预设取值范围;或从所有训练输入数据中获取缺少数据输入变量的值,统计获取到的缺少数据输入变量的最大值和最小值,将从最小值到最大值之间值的范围作为缺少数据输入变量的预设取值范围;或提示并接受用户输入缺少数据输入变量的预设取值范围。

在对应的预设取值范围内随机生成预定数量个值对缺少数据的输入变量进行预定数量次赋值;其中,预定数量由用户输入或预先设置。

在一个技术方案中,取值范围获取模块 3 还用于从样本数据库中随机选取预定数量个训练输入数据,然后提取其中该输入变量对应的数据对缺少数据输入变量进行预定数量次赋值;其中,预定数量由用户输入或将小于等于样本数据库中输入数据的数量的  $P\%$  作为预定数量, $P$  小于等于 100,样本数据库中的输入数据都是完整的输入数据。

(4) 数据输入模块 4,用于将多组完整输入数据输入到预先训练好的深度学习模型中,计算得到多组输出数据。

在一个技术方案中,数据输入模块 4 还用于统计多组输出数据中各类输出数据所占的比例;将比例最高的一类输出数据作为预测结果;将预测结果及各类输出数据及其相应比例输出给用户。

(5) 在一个技术方案中,上述深度学习模型装置还包括训练模块 5,用于对深度学习模型进行初始化,获取训练数据,通过训练数据对初始化后的深度学习模型进行训练得到预先训练好的深度学习模型。

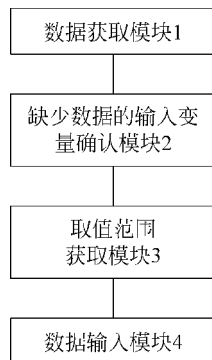


图 3-4 深度学习模型方法的装置结构框图