



(12) 发明专利

(10) 授权公告号 CN 109359768 B

(45) 授权公告日 2022. 08. 19

(21) 申请号 201811115128.7

G06N 3/04 (2006.01)

(22) 申请日 2018.09.25

G06N 3/08 (2006.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 109359768 A

(56) 对比文件

CN 106202805 A, 2016.12.07

CN 108268935 A, 2018.07.10

US 2014257526 A1, 2014.09.11

CN 107992971 A, 2018.05.04

(43) 申请公布日 2019.02.19

(73) 专利权人 联想(北京)有限公司

地址 100085 北京市海淀区上地信息产业
基地创业路6号

审查员 王晓敏

(72) 发明人 杨帆 金宝宝 金继民 张成松

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227

专利代理师 王宝筠

(51) Int. Cl.

G06Q 10/04 (2012.01)

G06Q 50/04 (2012.01)

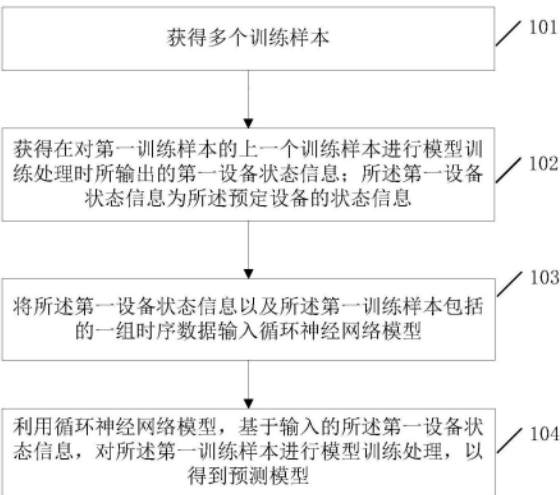
权利要求书4页 说明书23页 附图8页

(54) 发明名称

一种数据处理方法、服务器及电子设备

(57) 摘要

本申请提供一种数据处理方法、服务器及电子设备,可利用循环神经网络训练一预测模型,该预测模型用于对预定设备的目标指标进行预测。在训练预测模型时,除了以训练样本为模型输入,还引入了待预测指标所对应的所述预定设备的设备状态信息,并具体以第一训练样本的上一样本在训练过程中输出的第一设备状态信息作为第一训练样本在进行训练时的状态输入信息,这就使得设备状态信息能够按各样本的训练次序在各样本的训练过程中进行传递,相对地能够使得所得的预测模型达到对设备状态信息的长时记忆效果,后续在利用该预测模型进行预测时,能够结合预测样本的数据及记忆的设备状态信息对目标指标进行预测,从而可使得预测结果具备较高的准确率。



1. 一种数据处理方法,其特征在于,用于训练一预测模型,所述预测模型用于对预定设备的目标指标进行数据预测,所述方法包括:

获得多个训练样本;每个训练样本包括针对所述预定设备的至少一个采集指标所获得的一组时序数据以及该组时序数据对应的目标指标实际值;

针对待处理的第一训练样本,执行以下的预定处理:

获得在对所述第一训练样本的上一个训练样本进行模型训练处理时所输出的第一设备状态信息;所述第一设备状态信息为所述预定设备的状态信息;

将所述第一设备状态信息以及所述第一训练样本包括的一组时序数据输入循环神经网络模型;

利用循环神经网络模型,基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型;

其中,每个训练样本对应的一组时序数据包括所述至少一个采集指标在预定数量的历史时序点中各个历史时序点的指标数据;每个历史时序点的指标数据对应一目标指标实际值;

所述利用循环神经网络模型,基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型,包括:

利用循环神经网络模型,基于所述第一设备状态信息,对所述第一训练样本中第一个时序点的样本数据进行模型训练处理,得到第一训练样本中第一个时序点的样本数据所对应的目标指标预测值及设备状态信息;

若第一训练样本中存在未被处理的时序点样本数据,则利用循环神经网络模型,基于第一训练样本中相邻的上一个时序点样本数据对应得到的设备状态信息,对该未被处理的时序点样本数据进行模型训练处理,得到该未被处理的时序点样本数据所对应的目标指标预测值及设备状态信息;

将第一训练样本中最后一个时序点样本数据对应得到的目标指标预测值及设备状态信息,分别作为第一训练样本的目标指标预测值及第二设备状态信息并输出;所述第二设备状态信息为所述预定设备的状态信息;

基于第一训练样本中的最后一个时序点样本数据对应的目标指标预测值以及目标指标实际值,调整所述循环神经网络模型的模型参数,以得到预测模型。

2. 根据权利要求1所述的方法,其特征在于,所述利用循环神经网络模型,基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型,包括:

对所述第一训练样本的相应样本数据及所述第一设备状态信息进行整合处理,得到第一训练样本对应的整合样本;

利用循环神经网络模型,对所述整合样本进行模型训练处理,得到模型输出的对应于所述第一训练样本的目标指标预测值以及第二设备状态信息;所述第二设备状态信息为所述预定设备的状态信息;

基于第一训练样本的目标指标预测值以及第一训练样本对应的目标指标实际值,调整所述循环神经网络模型的模型参数,以得到预测模型。

3. 根据权利要求1所述的方法,其特征在于,所述获得多个训练样本,包括:

按时间顺序获得所述至少一个采集指标在多个历史时序点的指标数据及每个历史时

序点的指标数据对应的目标指标实际值；

采用一滑动窗口以预定步长对所述多个历史时序点的指标数据及其对应的目标指标实际值进行滑动处理，滑动过程中每一个窗口内对应包括的采集指标的一组时序数据及目标指标实际值作为一个训练样本，得到多个训练样本；所述滑动窗口的时序点长度为所述预定数量。

4. 根据权利要求3所述的方法，其特征在于，在所述采用一滑动窗口以预定步长对所述多个历史时序点的指标数据进行滑动处理之前，还包括：

对所述多个历史时序点的采集指标的指标数据进行筛选处理，得到所述多个历史时序点的符合预定条件的指标数据。

5. 根据权利要求1-4任一项所述的方法，其特征在于，还包括：

利用预测模型对待处理的预测样本进行预测处理，得到所述预测样本对应的目标指标预测值；

所述预测样本对应的目标指标预测值的生成，利用了所述预测模型在训练阶段所涉及的所述预定设备的所有历史状态信息以及在对所述待处理的预测样本处理之前的预测阶段所涉及的所述预定设备的所有历史状态信息。

6. 一种服务器，其特征在于，用于训练一预测模型，所述预测模型用于对预定设备的目标指标进行数据预测，所述服务器包括：

存储器，用于至少存储一组指令集；

处理器，用于调用并执行所述存储器中的所述指令集，通过执行所述指令集进行以下操作：

获得多个训练样本；每个训练样本包括针对所述预定设备的至少一个采集指标所获得的一组时序数据以及该组时序数据对应的目标指标实际值；

针对待处理的第一训练样本，执行以下的预定处理：

获得在对所述第一训练样本的上一个训练样本进行模型训练处理时所输出的第一设备状态信息；所述第一设备状态信息为所述预定设备的状态信息；

将所述第一设备状态信息以及所述第一训练样本包括的一组时序数据输入循环神经网络模型；

利用循环神经网络模型，基于输入的所述第一设备状态信息，对所述第一训练样本进行模型训练处理，以得到预测模型；

其中，所述每个训练样本对应的一组时序数据包括所述至少一个采集指标在预定数量的历史时序点中各个历史时序点的指标数据；每个历史时序点的指标数据对应一目标指标实际值；

所述处理器利用循环神经网络模型，基于输入的所述第一设备状态信息，对所述第一训练样本进行模型训练处理，以得到预测模型，包括：

利用循环神经网络模型，基于所述第一设备状态信息，对所述第一训练样本中第一个时序点的样本数据进行模型训练处理，得到第一训练样本中第一个时序点的样本数据所对应的目标指标预测值及设备状态信息；

若第一训练样本中存在未被处理的时序点样本数据，则利用循环神经网络模型，基于第一训练样本中相邻的上一个时序点样本数据对应得到的设备状态信息，对该未被处理的

时序点样本数据进行模型训练处理,得到该未被处理的时序点样本数据所对应的目标指标预测值及设备状态信息;

将第一训练样本中最后一个时序点样本数据对应得到的目标指标预测值及设备状态信息,分别作为第一训练样本的目标指标预测值及第二设备状态信息并输出;所述第二设备状态信息为所述预定设备的状态信息;

基于第一训练样本中的最后一个时序点样本数据对应的目标指标预测值以及目标指标实际值,调整所述循环神经网络模型的模型参数,以得到预测模型。

7. 根据权利要求6所述的服务器,其特征在于,所述处理器利用循环神经网络模型,基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型,具体包括:

对所述第一训练样本的相应样本数据及所述第一设备状态信息进行整合处理,得到第一训练样本对应的整合样本;

利用循环神经网络模型,对所述整合样本进行模型训练处理,得到模型输出的对应于所述第一训练样本的目标指标预测值以及第二设备状态信息;所述第二设备状态信息为所述预定设备的状态信息;

基于第一训练样本的目标指标预测值以及第一训练样本对应的目标指标实际值,调整所述循环神经网络模型的模型参数,以得到预测模型。

8. 一种电子设备,其特征在于,用于训练一预测模型,所述预测模型用于对预定设备的目标指标进行数据预测,所述电子设备包括:

存储器,用于至少存储一组指令集;

处理器,用于调用并执行所述存储器中的所述指令集,通过执行所述指令集进行以下操作:

获得多个训练样本;每个训练样本包括针对所述预定设备的至少一个采集指标所获得的一组时序数据以及该组时序数据对应的目标指标实际值;

针对待处理的第一训练样本,执行以下的预定处理:

获得在对所述第一训练样本的上一个训练样本进行模型训练处理时所输出的第一设备状态信息;所述第一设备状态信息为所述预定设备的状态信息;

将所述第一设备状态信息以及所述第一训练样本包括的一组时序数据输入循环神经网络模型;

利用循环神经网络模型,基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型;

其中,每个训练样本对应的一组时序数据包括所述至少一个采集指标在预定数量的历史时序点中各个历史时序点的指标数据;每个历史时序点的指标数据对应一目标指标实际值;

所述利用循环神经网络模型,基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型,包括:

利用循环神经网络模型,基于所述第一设备状态信息,对所述第一训练样本中第一个时序点的样本数据进行模型训练处理,得到第一训练样本中第一个时序点的样本数据所对应的目标指标预测值及设备状态信息;

若第一训练样本中存在未被处理的时序点样本数据,则利用循环神经网络模型,基于第一训练样本中相邻的上一个时序点样本数据对应得到的设备状态信息,对该未被处理的时序点样本数据进行模型训练处理,得到该未被处理的时序点样本数据所对应的目标指标预测值及设备状态信息;

将第一训练样本中最后一个时序点样本数据对应得到的目标指标预测值及设备状态信息,分别作为第一训练样本的目标指标预测值及第二设备状态信息并输出;所述第二设备状态信息为所述预定设备的状态信息;

基于第一训练样本中的最后一个时序点样本数据对应的目标指标预测值以及目标指标实际值,调整所述循环神经网络模型的模型参数,以得到预测模型。

一种数据处理方法、服务器及电子设备

技术领域

[0001] 本发明属于大数据分析预测技术领域,尤其涉及一种数据处理方法、服务器及电子设备。

背景技术

[0002] 在以石化行业为代表的流程制造行业中,准确的预测其目标产出物(例如石化行业中的汽油、液化气等)的收率,对于指导生产过程,实现经济效益最大化是非常关键的。

[0003] 由于流程制造行业所涉及的流程繁多,工艺复杂,且各个流程间往往存在复杂的相互影响,从而会导致准确预测其目标产出物的收率较为困难。因此,对于流程制造行业而言,提供一种解决方案以能够有效、高准确率地预测诸如目标产出物收率等指标的指标数据十分必要。

发明内容

[0004] 有鉴于此,本发明的目的在于提供一种数据处理方法、服务器及电子设备,以能够有效、高准确率地预测诸如目标产出物收率等指标的指标数据。

[0005] 为此,本发明公开如下技术方案:

[0006] 一种数据处理方法,用于训练一预测模型,所述预测模型用于对预定设备的目标指标进行数据预测,所述方法包括:

[0007] 获得多个训练样本;每个训练样本包括针对所述预定设备的至少一个采集指标所获得的一组时序数据以及该组时序数据对应的目标指标实际值;

[0008] 针对待处理的第一训练样本,执行以下的预定处理:

[0009] 获得在对所述第一训练样本的上一个训练样本进行模型训练处理时所输出的第一设备状态信息;所述第一设备状态信息为所述预定设备的状态信息;

[0010] 将所述第一设备状态信息以及所述第一训练样本包括的一组时序数据输入循环神经网络模型;

[0011] 利用循环神经网络模型,基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型。

[0012] 上述方法,优选的,所述利用循环神经网络模型,基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型,包括:

[0013] 对所述第一训练样本的相应样本数据及所述第一设备状态信息进行整合处理,得到第一训练样本对应的整合样本;

[0014] 利用循环神经网络模型,对所述整合样本进行模型训练处理,得到模型输出的对应于所述第一训练样本的目标指标预测值以及第二设备状态信息;所述第二设备状态信息为所述预定设备的状态信息;

[0015] 基于第一训练样本的目标指标预测值以及第一训练样本对应的目标指标实际值,调整所述循环神经网络模型的模型参数,以得到预测模型。

[0016] 上述方法,优选的,每个训练样本对应的一组时序数据包括所述至少一个采集指标在预定数量的历史时序点中各个历史时序点的指标数据;每个历史时序点的指标数据对应一目标指标实际值;

[0017] 所述利用循环神经网络模型,基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型,包括:

[0018] 利用循环神经网络模型,基于所述第一设备状态信息,对所述第一训练样本中第一个时序点的样本数据进行模型训练处理,得到第一训练样本中第一个时序点的样本数据所对应的目标指标预测值及设备状态信息;

[0019] 若第一训练样本中存在未被处理的时序点样本数据,则利用循环神经网络模型,基于第一训练样本中相邻的上一个时序点样本数据对应得到的设备状态信息,对该未被处理的时序点样本数据进行模型训练处理,得到该未被处理的时序点样本数据所对应的目标指标预测值及设备状态信息;

[0020] 将第一训练样本中最后一个时序点样本数据对应得到的目标指标预测值及设备状态信息,分别作为第一训练样本的目标指标预测值及第二设备状态信息并输出;所述第二设备状态信息为所述预定设备的状态信息;

[0021] 基于第一训练样本中的最后一个时序点样本数据对应的目标指标预测值以及目标指标实际值,调整所述循环神经网络模型的模型参数,以得到预测模型。

[0022] 上述方法,优选的,所述获得多个训练样本,包括:

[0023] 按时间顺序获得所述至少一个采集指标在多个历史时序点的指标数据及每个历史时序点的指标数据对应的目标指标实际值;

[0024] 采用一滑动窗口以预定步长对所述多个历史时序点的指标数据及其对应的目标指标实际值进行滑动处理,滑动过程中每一个窗口内对应包括的采集指标的一组时序数据及目标指标实际值作为一个训练样本,得到多个训练样本;所述滑动窗口的时序点长度为所述预定数量。

[0025] 上述方法,优选的,在所述采用一滑动窗口以预定步长对所述多个历史时序点的指标数据进行滑动处理之前,还包括:

[0026] 对所述多个历史时序点的采集指标的指标数据进行筛选处理,得到所述多个历史时序点的符合预定条件的指标数据。

[0027] 上述方法,优选的,还包括:

[0028] 利用预测模型对待处理的预测样本进行预测处理,得到所述预测样本对应的目标指标预测值;

[0029] 所述预测样本对应的目标指标预测值的生成,利用了所述预测模型在训练阶段所涉及的所述预定设备的所有历史状态信息以及在对所述待处理的预测样本处理之前的预测阶段所涉及的所述预定设备的所有历史状态信息。

[0030] 一种服务器,用于训练一预测模型,所述预测模型用于对预定设备的目标指标进行数据预测,所述服务器包括:

[0031] 存储器,用于至少存储一组指令集;

[0032] 处理器,用于调用并执行所述存储器中的所述指令集,通过执行所述指令集进行以下操作:

[0033] 获得多个训练样本;每个训练样本包括针对所述预定设备的至少一个采集指标所获得的一组时序数据以及该组时序数据对应的目标指标实际值;

[0034] 针对待处理的第一训练样本,执行以下的预定处理:

[0035] 获得在对所述第一训练样本的上一个训练样本进行模型训练处理时所输出的第一设备状态信息;所述第一设备状态信息为所述预定设备的状态信息;

[0036] 将所述第一设备状态信息以及所述第一训练样本包括的一组时序数据输入循环神经网络模型;

[0037] 利用循环神经网络模型,基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型。

[0038] 上述服务器,优选的,所述处理器利用循环神经网络模型,基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型,具体包括:

[0039] 对所述第一训练样本的相应样本数据及所述第一设备状态信息进行整合处理,得到第一训练样本对应的整合样本;

[0040] 利用循环神经网络模型,对所述整合样本进行模型训练处理,得到模型输出的对应于所述第一训练样本的目标指标预测值以及第二设备状态信息;所述第二设备状态信息为所述预定设备的状态信息;

[0041] 基于第一训练样本的目标指标预测值以及第一训练样本对应的目标指标实际值,调整所述循环神经网络模型的模型参数,以得到预测模型。

[0042] 上述服务器,优选的,所述每个训练样本对应的一组时序数据包括所述至少一个采集指标在预定数量的历史时序点中各个历史时序点的指标数据;每个历史时序点的指标数据对应一目标指标实际值;

[0043] 所述处理器利用循环神经网络模型,基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型,包括:

[0044] 利用循环神经网络模型,基于所述第一设备状态信息,对所述第一训练样本中第一个时序点的样本数据进行模型训练处理,得到第一训练样本中第一个时序点的样本数据所对应的目标指标预测值及设备状态信息;

[0045] 若第一训练样本中存在未被处理的时序点样本数据,则利用循环神经网络模型,基于第一训练样本中相邻的上一个时序点样本数据对应得到的设备状态信息,对该未被处理的时序点样本数据进行模型训练处理,得到该未被处理的时序点样本数据所对应的目标指标预测值及设备状态信息;

[0046] 将第一训练样本中最后一个时序点样本数据对应得到的目标指标预测值及设备状态信息,分别作为第一训练样本的目标指标预测值及第二设备状态信息并输出;所述第二设备状态信息为所述预定设备的状态信息;

[0047] 基于第一训练样本中的最后一个时序点样本数据对应的目标指标预测值以及目标指标实际值,调整所述循环神经网络模型的模型参数,以得到预测模型。

[0048] 一种电子设备,用于训练一预测模型,所述预测模型用于对预定设备的目标指标进行数据预测,所述电子设备包括:

[0049] 存储器,用于至少存储一组指令集;

[0050] 处理器,用于调用并执行所述存储器中的所述指令集,通过执行所述指令集进行

以下操作：

[0051] 获得多个训练样本；每个训练样本包括针对所述预定设备的至少一个采集指标所获得的一组时序数据以及该组时序数据对应的目标指标实际值；

[0052] 针对待处理的第一训练样本，执行以下的预定处理：

[0053] 获得在对所述第一训练样本的上一个训练样本进行模型训练处理时所输出的第一设备状态信息；所述第一设备状态信息为所述预定设备的状态信息；

[0054] 将所述第一设备状态信息以及所述第一训练样本包括的一组时序数据输入循环神经网络模型；

[0055] 利用循环神经网络模型，基于输入的所述第一设备状态信息，对所述第一训练样本进行模型训练处理，以得到预测模型。

[0056] 根据以上方案可知，本申请提供的数据处理方法、服务器及电子设备，可利用循环神经网络训练一预测模型，该预测模型用于对预定设备的目标指标进行数据预测。在训练预测模型时，除了以训练样本的样本数据为模型输入，还引入了待预测指标所对应设备（即所述预定设备）的设备状态信息，并具体以第一训练样本的上一样本在训练过程中输出的第一设备状态信息作为第一训练样本在进行模型训练时的状态输入信息，这就使得设备状态信息能够按各样本的训练次序在各样本的训练过程中进行传递，相对应地能够使得训练所得的预测模型达到对训练过程中所涉及的设备状态信息的长时记忆效果，后续在利用该预测模型进行数据预测时，能够结合预测样本的样本数据以及记忆的设备状态信息对目标指标进行预测，从而可使得最终的预测结果与设备状态更加相符、具备较高的准确率。

附图说明

[0057] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的实施例，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，还可以根据提供的附图获得其他的附图。

[0058] 图1是本申请提供的一种数据处理方法实施例一的流程图；

[0059] 图2是本申请提供的一种数据处理方法实施例二的流程图；

[0060] 图3是本申请实施例二提供的获得训练样本集的处理过程流程图；

[0061] 图4是循环神经网络的基本模型结构示意图；

[0062] 图5是本申请实施例二提供的长时记忆神经网络的基本模型结构示意图；

[0063] 图6是本申请实施例三提供的获得训练样本集的处理过程流程图；

[0064] 图7是本申请提供的一种数据处理方法实施例四的流程图；

[0065] 图8是本申请实施例四提供的基于构建预测模型并利用预测模型对目标指标进行预测的整体逻辑框架示意图；

[0066] 图9是本申请提供的一种服务器实施例五的结构示意图；

[0067] 图10是本申请提供的一种电子设备实施例九的结构示意图。

具体实施方式

[0068] 为了引用和清楚起见，下文中使用的技术名词、简写或缩写总结解释如下：

[0069] 收率：也称作反应收率，一般用于化学及工业生产，是指在化学反应或相关的化学工业生产中，投入单位数量原料获得的实际生产的产品产量与理论计算的产品产量的比值。同样的一个化学反应在不同的压力、温度下会有不同的收率。

[0070] 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，都属于本发明保护的范围。

[0071] 在以石化行业为代表的流程制造行业中，准确的预测其目标产出物的收率，对于指导生产过程，实现经济效益最大化是非常关键的。发明人经研究发现，实际生产环境中某一时刻的目标产出物收率会受历史各时刻的设备（即生产出目标产出物的过程中所使用的设备）状态的影响，如某时刻汽油收率会受历史各时刻反应再生系统的设备状态（可以将反应再生系统理解为一个整体设备，其设备状态可基于反应塔的相应温度值、再生器的相应温度值、再生器的相应风量值等来体现）的影响等，这就需要在对目标产出物的收率进行预测时，能够记忆历史各时刻的设备状态信息，进而基于记忆的设备状态信息进行收率预测。

[0072] 然而，现有的预测方案均无法达到该效果或未考虑历史各时刻设备状态信息对目标产出物收率的影响，鉴于此，本申请提供了一种数据处理方法、服务器及电子设备，以通过对待预测指标所对应设备的设备状态信息进行长时记忆进而基于长时记忆的设备状态信息对目标指标进行数据预测，来实现高准确率地预测诸如目标产出物收率等目标指标的指标数据。以下将通过多个实施例对本申请的数据处理方法、服务器及电子设备进行说明。

[0073] 参考图1，是本申请提供的一种数据处理方法实施例一的流程图，该数据处理方法用于训练一预测模型，所述预测模型用于对预定设备的目标指标进行数据预测，该数据处理方法可应用于能够提供数据处理服务的本地/网络端服务器或服务器集群中，或者还可以应用于智能手机、平板电脑、台式机、笔记本、一体机等各类型终端设备中。如图1所示，本实施例中，所述数据处理方法包括如下步骤：

[0074] 步骤101、获得多个训练样本；每个训练样本包括针对所述预定设备的至少一个采集指标所获得的一组时序数据以及该组时序数据对应的目标指标实际值。

[0075] 本申请将主要以流程制造行业中目标产出物（如石化行业中的汽油、液化气等）的收率预测所需的预测模型为例，对本申请方案进行阐述。

[0076] 所述获得多个训练样本，即指获得训练样本集，以预测模型所需预测的目标指标为汽油、液化气等产出物的收率为例，训练样本集可以包括反应再生系统在各历史时刻的反应塔温度、再生器温度、再生器风量等一系列采集指标的指标数据（这些指标数据能够体现反应再生系统整体设备的设备运行状态），其中，每个训练样本可包括各采集指标的一组时序数据，例如，每个训练样本具体包括反应再生系统在预定数量的时序点所对应的各采集指标的指标数据等，该预定数量的时序点所对应的各采集指标的指标数据即构成该训练样本所对应的一组采集指标时序数据。

[0077] 除了各采集指标的指标数据，训练样本中还应包括各采集指标的指标数据所对应的目标指标实际值，也就是说，训练样本应同时包括采集指标的指标数据及其对应的目标指标实际值，以目标指标为汽油收率为例，所述训练样本则具体可以包括反应塔温度、再生器温度、再生器风量等一系列采集指标的指标数据及这些指标数据对应的汽油实际收率

值。

[0078] 针对待处理的第一训练样本,执行以下的预定处理:

[0079] 步骤102、获得在对所述第一训练样本的上一个训练样本进行模型训练处理时所输出的第一设备状态信息;所述第一设备状态信息为所述预定设备的状态信息。

[0080] 所述第一训练样本,是指所获得的训练样本集中除第一个参与模型训练的样本之外的其他样本中的任意一个训练样本,也就是说,所述第一训练样本在训练样本集中至少对应一个在其前参与模型训练的训练样本。

[0081] 所述第一设备状态信息具体是指在对所述第一训练样本的上一个训练样本进行模型训练处理时模型所输出的设备状态信息,该设备状态信息为待预测的目标指标所对应设备的状态信息,以目标指标为汽油收率为例,则该设备状态信息为反应再生系统所对应的整体设备的设备运行状态信息。

[0082] 步骤103、将所述第一设备状态信息以及所述第一训练样本包括的一组时序数据输入循环神经网络模型。

[0083] 与传统的模型训练过程中仅以训练样本的样本数据为模型输入相区别,本申请是将第一训练样本的样本数据,以及在对第一训练样本的上一个训练样本进行模型训练处理时所输出的所述第一设备状态信息均作为模型的输入信息,即,这两种信息均参与针对第一训练样本的模型训练过程。

[0084] 本申请具体基于循环神经网络进行模型训练,从而模型训练过程中,可将待处理的第一训练样本的样本数据以及所获得的所述第一设备状态信息输入当前训练时刻所得的循环神经网络模型中,以进行模型训练。

[0085] 步骤104、利用循环神经网络模型,基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型。

[0086] 在获得所输入的第一训练样本的样本数据以及所述第一设备状态信息后,循环神经网络模型基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型。

[0087] 由于循环神经网络模型基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,从而在对第一训练样本进行模型训练处理时,引入了设备状态信息,考虑了设备状态信息对预测结果的影响,从而,可使得训练所得的预测模型在预测准确率方面具备较好的性能。

[0088] 这里,需要说明的是,对于训练样本集中第一个参与模型训练的训练样本而言,由于不存在在其之前参与训练的样本,从而相对应地无法获得其上一个训练样本在训练过程中输出的第一设备状态信息,针对该情况,可基于人工设置等方式,为其在进行模型训练时所需输入的设备状态信息赋予一初始值,该初始值可由所属领域的技术人员根据业务经验设定,例如将其设置为0或非0的其他数值等。

[0089] 根据以上方案可知,本实施例提供的数据处理方法,在利用循环神经网络训练预测模型时,除了以训练样本的样本数据为模型输入,还引入了待预测指标所对应设备(即所述预定设备)的设备状态信息,并具体以第一训练样本的上一样本在训练过程中输出的第一设备状态信息作为第一训练样本在进行模型训练时的状态输入信息,这就使得设备状态信息能够按各样本的训练次序在各样本的训练过程中进行传递,相对应地能够使得训练所

得的预测模型达到对训练过程中所涉及的设备状态信息的长时记忆效果,后续在利用该预测模型进行数据预测时,能够结合预测样本的样本数据以及记忆的设备状态信息对目标指标进行预测,从而可使得最终的预测结果与设备状态更加相符、具备较高的准确率。

[0090] 以下通过另一实施例继续对上述实施例中的数据处理方法进行进一步详述,参考图2,是本申请提供的一种数据处理方法实施例二的流程图,本实施例二中,所述数据处理方法可以通过如下的处理过程实现:

[0091] 步骤201、获得多个训练样本;每个训练样本包括针对所述预定设备的至少一个采集指标所获得的一组时序数据以及该组时序数据对应的目标指标实际值。

[0092] 本实施例中,每个训练样本对应的一组时序数据包括所述至少一个采集指标在预定数量的历史时序点中各个历史时序点的指标数据;每个历史时序点的指标数据对应一目标指标实际值。

[0093] 参考图3,本步骤可通过以下的处理过程来获得多个训练样本,也即获得训练样本集:

[0094] 步骤301、按时间顺序获得所述至少一个采集指标在多个历史时序点的指标数据及每个历史时序点的指标数据对应的目标指标实际值;

[0095] 步骤302、采用一滑动窗口以预定步长对所述多个历史时序点的指标数据及其对应的目标指标实际值进行滑动处理,滑动过程中每一个窗口内对应包括的采集指标的一组时序数据及目标指标实际值作为一个训练样本,得到多个训练样本;所述滑动窗口的时序点长度为所述预定数量。

[0096] 以下以最终构建的预测模型所需预测的目标指标为流程制造行业的目标产出物收率为例,对上述的训练样本集获取过程进行详述。

[0097] 具体以汽油收率为例,可从目标石化企业的数据库(如IP21实时数据库)中按时间顺序获取反应塔温度、再生器温度、再生器风量等各采集指标在多个历史时序点的指标数据,以及每个历史时序点的采集指标数据所对应的汽油收率;优选地,在获得多个历史时序点的采集指标数据及其对应的目标产出物收率时,任意相邻时序点间的时间间隔相同,当然,实际应用中并不以此为限。

[0098] 假设从数据库中采集了 M ($M > 1$, 且 M 为自然数)个历史时序点的采集指标数据及其对应的目标指标实际值,其中采集指标的个数为 n ($n \geq 1$, 且 n 为自然数),则具体可将采集的 M 个时序点的采集指标数据及其对应的目标指标实际值表示为以下的表1:

[0099] 表1

[0100]

时间	X1	X2	...	Xn	Y
T1	x1_1	X2_1	...	Xn_1	Y_1
T2	x1_2	X2_2	...	Xn_2	Y_2
T3	x1_3	X2_3	...	Xn_3	Y_3
T4	x1_4	X2_4	...	Xn_4	Y_4
...
TM	x1_M	X2_M		Xn_M	Y_M

[0101] 在上述的表1中, T_i ($1 \leq i \leq M$, i 为自然数)表示第 i 个时序点对应的时间, X_j ($1 \leq j \leq n$, j 为自然数)表示第 j 个采集指标, X_{j_i} 表示第 j 个采集指标在第 i 个时序点的指标数据,

Y_i 表示第 i 个时序点的采集指标数据所对应的目标指标(如汽油收率)实际值。

[0102] 由于本申请中每个训练样本均包括采集指标的一组时序数据(同时也是循环神经网络所要求的样本形式),从而在获得上述多个历史时序点的采集指标数据及其对应的目标指标实际值后,需继续对其进行时序化处理。

[0103] 具体地,可采用一滑动窗口以预定步长对所述至少一个采集指标在多个历史时序点的指标数据及其对应的目标指标实际值进行滑动处理,滑动过程中每一个窗口内对应包括的采集指标的一组时序数据及其对应的目标指标实际值作为一个训练样本,得到多个训练样本;所述滑动窗口的时序点长度为所述预定数量。

[0104] 参考以下的表2,假设所述滑动窗口的时序点长度为3,滑动步长为1,则如表2所示,在滑动窗口所处的第一个窗口内包括T1-T3这3个时序点,从而T1-T3这3个时序点的采集样本数据及其对应的目标指标实际值构成第一个训练样本Sample1,相对应的,在滑动窗口所处的第二个窗口内包括T2-T4这3个时序点,从而T2-T4这3个时序点的采集样本数据及其对应的目标指标实际值构成第二个训练样本Sample2,……,在滑动窗口所处的最后一个窗口内包括T(M-2)-TM这3个时序点,从而T(M-2)-TM这3个时序点的采集样本数据及其对应的目标指标实际值构成第(M-2)个训练样本Sample(M-2),从而最终得到包括(M-2)个训练样本的训练样本集:Train_Set=[Sample1,Sample2,...,Sample(M-2)]。其中,训练样本中所包括的采集指标的指标数据用于作为样本的特征,而目标指标实际值则用于作为样本的标签。

[0105] 表2

时间 ϕ	X1 ϕ	X2 ϕ	... ϕ	Xn ϕ	Y ϕ
T1 ϕ	x1_1 ϕ	X2_1 ϕ	... ϕ	Xn_1 ϕ	Y_1 ϕ
T2 ϕ	x1_2 ϕ	X2_2 ϕ	... ϕ	Xn_2 ϕ	Y_2 ϕ
T3 ϕ	x1_3 ϕ	X2_3 ϕ	... ϕ	Xn_3 ϕ	Y_3 ϕ
T4 ϕ	x1_4 ϕ	X2_4 ϕ	... ϕ	Xn_4 ϕ	Y_4 ϕ
... ϕ	... ϕ	... ϕ	... ϕ	... ϕ	... ϕ
TM ϕ	x1_M ϕ	X2_M ϕ	ϕ	Xn_M ϕ	Y_M ϕ

[0107] 针对待处理的第一训练样本,执行以下的预定处理:

[0108] 步骤202、获得在对所述第一训练样本的上一个训练样本进行模型训练处理时所输出的第一设备状态信息;所述第一设备状态信息为所述预定设备的状态信息。

[0109] 本申请利用循环神经网络来训练预测模型,参考图4示出的循环神经网络的基本模型结构,其中,该图4具体以循环神经网络包括3个时序单元(下文简称时序单元为CELL,循环神经网络包括3个时序单元相对应地要求其训练样本包括3个时序点的样本数据)为例,示出了其所对应的模型结构,图中的三个CELL具体是同一个CELL在不同时序上的展开,训练该“循环神经网络”以得到预测模型的过程,就是不断调整所述CELL的CELL参数(模型参数)的过程。

[0110] 现有技术中,针对每个训练样本,如图4所示,对该训练样本中第一个时序点样本数据的处理,都是将第一个时序点的样本数据以及人工设定的初始状态信息(如图4中的

S0,该值一般设定为0)共同作为模型的输入,即将这两种信息共同传入CELL以得到第一个时序点的目标指标输出值(对应图4中的V1_t1)及状态信息输出值(对应图4中的S1);而之后的时序点样本数据的训练,则是以该之后时序点的样本数据及其上一个相邻时序点的状态信息输出值为模型输入,将这两种信息共同传入CELL来得到该时序点的目标指标输出值及状态输出值,以图4中训练样本的第三个时序点为例,其是将第三个时序点的样本数据R1_t3以及第二时序点对应输出的状态信息S2共同传入CELL,并得到该第三个时序点的目标指标输出值V1_t3,及状态输出值S3,直至当前训练样本中最后一个时序点样本数据的训练过程结束,会基于最后一个时序点样本数据对应的目标指标输出值及目标指标实际值来调整模型参数,即调整所述CELL的CELL参数。

[0111] 需要说明的是,现有技术中,训练样本的最后一个时序点对应的状态输出值(如图4中的S3)在模型训练过程中不起任何作用,会将其直接丢弃。在当前训练样本的训练过程结束之后,会继续下一个训练样本的训练过程,其中对于每一个训练样本都需要人工设定其第一个时序点样本数据对应所需的状态输入值(一般将其初始化为0)。

[0112] 现有循环神经网络的上述特点,使得循环神经网络仅能在样本内部针对同一样本包括的各个时序点样本数据进行状态信息记忆(如目标指标所对应设备的设备状态等),无法达到长时记忆效果,为解决该问题,本申请提出了在训练样本间进行状态信息传递的技术构思。

[0113] 如图5所示,对于当前样本的最后一个时序点所对应的状态输出值(如图5中R1_t3对应的S3),本申请并不将其丢弃,而是将其作为对下一个训练样本中第一个时序点样本数据进行训练时的状态输入信息,将其与所述下一个训练样本中第一个时序点的样本数据共同输入模型。

[0114] 鉴于此,本步骤中,获得在对所述第一训练样本的上一个训练样本进行模型训练处理时所输出的第一设备状态信息,具体是指获得在对所述第一训练样本的上一个训练样本进行模型训练处理时所述上一个训练样本的最后一个时序点的样本数据对应输出的设备状态信息。例如,假设所述第一训练样本为图5中的sample2,则所述第一状态信息即为图5中的sample1的最后一个时序点的样本数据R1_t3对应输出的设备状态信息S3。

[0115] 步骤203、将所述第一设备状态信息以及所述第一训练样本包括的一组时序数据输入循环神经网络模型。

[0116] 本步骤中将所述第一设备状态信息以及所述第一训练样本包括的一组时序数据输入循环神经网络模型,具体可以包括将第一训练样本的上一个训练样本的最后一个时序点样本数据对应输出的设备状态信息(即所述第一设备状态信息),以及所述第一训练样本的第一个时序点的样本数据共同输入循环神经网络模型,以实现在对第一训练样本进行训练处理时,首先对第一训练样本的第一个时序点的样本数据进行模型训练处理。

[0117] 步骤204、利用循环神经网络模型,基于所述第一设备状态信息,对所述第一训练样本中第一个时序点的样本数据进行模型训练处理,得到第一训练样本中第一个时序点的样本数据所对应的目标指标预测值及设备状态信息。

[0118] 循环神经网络模型在对第一训练样本的第一个时序点的样本数据进行模型训练处理时,具体基于所述第一设备状态信息(即所述上一个训练样本的最后一个时序点的样本数据在训练时对应输出的设备状态信息),对所述第一训练样本中第一个时序点的样本

数据进行模型训练处理,并得到第一训练样本中第一个时序点的样本数据所对应的目标指标预测值及设备状态信息。

[0119] 仍以图5为例,假设第一训练样本为图5中的所述sample2,则sample2的第一个时序点的样本数据R2_t1所对应的第一设备状态信息S0'即为sample1的最后一个时序点的样本数据在训练时对应输出的设备状态信息S3,从而对于sample2,会将所述S3及sample2的第一个时序点的样本数据R2_t1共同输入神经网络模型的CELL,进而神经网络模型的CELL基于S3对第一个时序点的样本数据R2_t1进行模型训练处理,并输出该第一个时序点的样本数据R2_t1所对应的目标指标预测值V2_t2及设备状态信息S1',所输出的设备状态信息S1'用于作为下一个时序点的样本数据的状态输入信息。

[0120] 循环神经网络在基于输入的所述第一设备状态信息,对第一训练样本进行模型训练处理时,会通过所述第一训练样本的相应样本数据及所述第一设备状态信息进行整合处理,来得到第一训练样本对应的整合样本;进而通过对所述整合样本进行模型训练处理,得到模型输出的对应于所述第一训练样本的目标指标预测值以及第二设备状态信息;所述第二设备状态信息为所述预定设备的状态信息,用于作为下一个训练样本的状态输入信息。

[0121] 具体地,对于第一训练样本的第一个时序点的样本数据来说,循环神经网络相应地会将该第一训练样本的上一个样本的最后一个时序点对应的状态输出信息(即所述第一设备状态信息),与该时序点的样本数据进行整合,并对整合后所得的整合样本进行模型训练处理。

[0122] 步骤205、若第一训练样本中存在未被处理的时序点样本数据,则利用循环神经网络模型,基于第一训练样本中相邻的上一个时序点样本数据对应得到的设备状态信息,对该未被处理的时序点样本数据进行模型训练处理,得到该未被处理的时序点样本数据所对应的目标指标预测值及设备状态信息。

[0123] 在对第一训练样本的第一个时序点的样本数据进行训练处理的基础上,若第一训练样本中还存在未被处理的其他时序点样本数据,则循环神经网络会将该未被处理时序点的上一个时序点的状态输出信息(为一设备状态信息)与该时序点的样本数据进行整合,并对整合后所得的整合样本进行模型训练处理,直至完成对该第一训练样本的最后一个时序点样本数据的训练处理后,对该第一训练样本的训练过程结束。

[0124] 步骤206、将第一训练样本中最后一个时序点样本数据对应得到的目标指标预测值及设备状态信息,分别作为第一训练样本的目标指标预测值及第二设备状态信息并输出;所述第二设备状态信息为所述预定设备的状态信息。

[0125] 且具体地,在完成对该第一训练样本的最后一个时序点样本数据的训练处理后,循环神经网络会将第一训练样本中最后一个时序点样本数据对应得到的目标指标预测值及设备状态信息,分别作为第一训练样本的目标指标预测值及第二设备状态信息并输出。

[0126] 以使得后续基于该第一训练样本的目标指标预测值通过反向传播来实现对模型参数(CELL参数)的调整,所述第二设备状态信息则具体用于作为所述第一训练样本的下一个样本中第一个时序点的状态输入信息,参与其下一个样本中第一个时序点的样本数据的训练。

[0127] 步骤207、基于第一训练样本中的最后一个时序点样本数据对应的目标指标预测

值以及目标指标实际值,调整所述循环神经网络模型的模型参数,以得到预测模型。

[0128] 在基于第一训练样本的目标指标预测值通过反向传播来实现对模型参数 (CELL参数) 的调整时,具体可基于第一训练样本中最后一个时序点样本数据对应的目标指标预测值以及该最后一个时序点样本数据对应的目标指标实际值,对循环神经网络模型的模型参数进行调整,也即对循环神经网络模型中CELL的CELL参数进行调整。

[0129] 具体地,循环神经网络模型中CELL的CELL参数可以包括对应于所输入的设备状态信息的状态权重,以及对应于所输入的样本特征 (每一个采集指标值可作为训练样本的一个样本特征) 的特征权重,从而对循环神经网络模型的训练本质上是指对其CELL中状态权重和特征权重的训练。

[0130] 在完成对训练样本集中所有样本的训练处理后,可得到最终的能够对目标设备的目标指标 (如汽油收率) 进行数据预测的预测模型。

[0131] 利用本申请方案,当在对训练样本集中的样本进行训练时,仅需针对参与训练的第一个样本的第一个时序点样本数据,为其设定一个状态输入 (即需输入的设备状态信息) 的初始值,该初始值可由所属领域的技术人员根据业务经验设定,例如将其设置为0或非0的其他数值等。在后续的训练过程中,每一个样本内部及不同样本之间均按时序对设备状态信息 (对目标指标的影响) 进行了传递,不需再针对其他任何样本进行其状态输入值的初始化设定。

[0132] 这里,需要说明的是,虽然从形式上来说,在对第一训练样本进行训练时,仅将在对第一训练样本的上一个训练样本进行模型训练处理时所输出的第一设备状态信息 (而非第一训练样本之前的所有样本对应的设备状态信息),以及第一训练样本的样本数据作为模型输入,共同参与了对第一训练样本的训练过程,然而,从模型训练的整体角度而言,每一个第一训练样本在训练时都引入了其上一个训练样本在训练过程中所输出的设备状态信息,并通过对上一个训练样本的状态输出信息与当前样本的整合样本进行处理,来得到当前样本的目标指标预测值及状态输出,这就使得训练模型过程中所涉及的各个历史设备状态信息得以传递,相应地,使得循环神经网络达到了对历史训练过程中所涉及的所有设备状态信息的长时记忆效果,从而,本质上来说,本申请通过训练所得的所述预测模型,为在循环神经网络基础上所得的一长时记忆神经网络模型、

[0133] 在利用本实施例构建的预测模型进行数据预测时,能够结合预测样本的样本数据以及长时记忆的设备状态信息对目标指标进行数据预测,可使得最终的预测结果与设备状态更加相符,进而可使得预测结果具备较高的准确率。

[0134] 在接下来的实施三中,继续对上一实施例中获得多个训练样本 (即获得训练样本集) 的部分进行进一步说明,本实施例中,参考图6示出的获得多个训练样本的流程示意图,在获得多个训练样本时,还可以包括以下步骤:

[0135] 步骤601、对所述多个历史时序点的采集指标的指标数据进行筛选处理,得到所述多个历史时序点的符合预定条件的指标数据。

[0136] 实际应用中,在所获得的采集指标的指标数据中,可能存在某些采集指标与待预测的目标指标高相关、而另一些采集指标则与待预测的目标指标低相关或不相关的情况,例如,在汽油收率预测中,反应塔温度及再生塔温度这些采集指标均与汽油收率高相关,而再生器风量等采集指标则与汽油收率低相关,针对该情况,可以对所采集的多个历史时序

点的指标数据进行筛选处理,以滤除掉其中的与目标指标低相关和/或不相关的采集指标的指标数据,从而使得所保留的各个采集指标的指标数据均与目标指标间具备较高的相关性。

[0137] 在对所述多个历史时序点的指标数据进行筛选处理后,可继续基于筛选处理后所得的指标数据进行窗口滑动处理,以得到时序数据形式的各个训练样本。

[0138] 本实施例通过对采集的多个历史时序点的指标数据进行筛选处理,可有效滤除掉其中的与目标指标低相关和/或不相关的采集指标数据,从而可使得所得的训练样本数据能够保持与目标指标的高相关性,进而可使得模型训练过程更为有效。

[0139] 参考图7,是本申请提供的一种数据处理方法实施例四的流程图,本实施例中,所述数据处理方法还可以包括以下步骤:

[0140] 步骤105、利用预测模型对待处理的预测样本进行预测处理,得到所述预测样本对应的目标指标预测值;

[0141] 所述预测样本对应的目标指标预测值的生成,利用了所述预测模型在训练阶段所涉及的所述预定设备的所有历史状态信息以及在对所述待处理的预测样本处理之前的预测阶段所涉及的所述预定设备的所有历史状态信息。

[0142] 在基于训练样本集进行模型训练,从而得到相对应的预测模型后,可利用所得的预测模型对待处理的预测样本进行预测处理,以得到所述预测样本所对应的目标指标预测值。

[0143] 参考图8示出的基于构建预测模型并利用预测模型对目标指标进行预测的整体逻辑框架示意图,对目标指标的数据预测可以通过以下的处理过程实现:

[0144] 1) 预测样本的准备:

[0145] 对于待预测时刻(作为一个时序点),可获得该时刻对应的各采集指标的指标数据,并同时获得该时刻之前相邻的 $(p-1)$ 个历史时序点处各采集指标的指标数据,将该待预测时刻对应的各采集指标的指标数据以及其之前相邻的 $(p-1)$ 个历史时序点处各采集指标的指标数据,按时序整合为当前待预测的预测样本,其中, P 为训练样本/预测样本的一组时序数据所对应的时序点个数,该数值也即在通过窗口滑动方式获得训练样本时滑动窗口的窗口长度。

[0146] 2) 预测目标指标的指标数据:

[0147] 对于当前待处理的预测样本,可将该预测样本的第一个时序点的样本数据以及所述预测模型最近输出的设备状态信息共同作为预测模型的输入,预测模型在获得这两种输入信息后,会对两种输入信息进行整合,并对整合后所得的整合样本进行预测处理,从而输出该第一个时序点的样本数据对应的目标指标预测值及设备状态信息;

[0148] 其中,对于利用预测模型预测的第一个预测样本来说,所述预测模型最近输出的设备状态信息(作为该预测样本预测时的输入状态信息)为最后一个训练样本的最后一个时序点对应输出的设备状态信息,而对于其他预测样本,所述预测模型最近输出的设备状态信息则为其上一个预测样本的最后一个时序点对应输出的设备状态信息。

[0149] 在对待处理的预测样本的第一个时序点的样本数据进行预测处理的基础上,若该预测样本中还存在未被处理的其他时序点样本数据,则预测模型会将该未被处理时序点的上一个时序点的状态输出值与该未被处理时序点的样本数据进行整合,并对整合后所得的

整合样本进行模型训练处理,直至完成对该预测样本的最后一个时序点样本数据的预测处理后,对该预测样本的预测过程结束。

[0150] 其中,该预测样本所对应的目标指标预测值,即为该预测样本的最后一个时序点样本数据对应的目标指标预测值,而该预测样本中其他时序点样本数据对应的目标指标预测值则丢弃,且模型对应于该预测样本的最后一个时序点样本数据所输出的设备状态信息,用于作为下一个预测样本的第一个时序点样本数据的输入状态信息,也即,在预测过程中每一个预测样本内部及不同预测样本之间同样均按时序对设备状态信息进行了传递,不需针对任意的预测样本进行其输入状态信息的初始化设定。

[0151] 这就使得预测样本对应的目标指标预测值的生成,不仅利用了所述预测模型在训练阶段所涉及的所述预定设备的所有历史状态信息,还利用了在对所述待处理的预测样本处理之前的预测阶段所涉及的所述预定设备的所有历史状态信息。从而,可使得最终的预测结果与设备状态更加符合,进而可使得预测结果具备较高的准确率。

[0152] 与上述的数据处理方法相对应,本申请还提供一种服务器,所述服务器可以是本地/网络端单独部署的服务器,或者本地/网络端的服务器集群中的服务器,参考图9,是本申请提供的一种服务器实施例五的结构示意图,所述服务器包括:

[0153] 存储器901,用于至少存储一组指令集。

[0154] 所述存储器901可以是具备数据存储功能的各类型存储器件,如ROM(Read Only Memory,只读存储器),FLASH,光盘,硬盘等,所存储的所述指令集用于指示处理器902执行如下文所述的相应数据处理操作,该指令集可以以程序形式存储于所述存储器901中。

[0155] 处理器902,用于调用并执行所述存储器中的所述指令集,通过执行所述指令集进行以下操作:

[0156] 获得多个训练样本;每个训练样本包括针对所述预定设备的至少一个采集指标所获得的一组时序数据以及该组时序数据对应的目标指标实际值;

[0157] 针对待处理的第一训练样本,执行以下的预定处理:

[0158] 获得在对所述第一训练样本的上一个训练样本进行模型训练处理时所输出的第一设备状态信息;所述第一设备状态信息为所述预定设备的状态信息;

[0159] 将所述第一设备状态信息以及所述第一训练样本包括的一组时序数据输入循环神经网络模型;

[0160] 利用循环神经网络模型,基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型。

[0161] 本申请将主要以流程制造行业中目标产出物(如石化行业中的汽油、液化气等)的收率预测所需的预测模型为例,对本申请方案进行阐述。

[0162] 所述获得多个训练样本,即指获得训练样本集,以预测模型所需预测的目标指标为汽油、液化气等产出物的收率为例,训练样本集可以包括反应再生系统在各历史时刻的反应塔温度、再生器温度、再生器风量等一系列采集指标的指标数据(这些指标数据能够体现反应再生系统整体设备的设备运行状态),其中,每个训练样本可包括各采集指标的一组时序数据,例如,每个训练样本具体包括反应再生系统在预定数量的时序点所对应的各采集指标的指标数据等,该预定数量的时序点所对应的各采集指标的指标数据即构成该训练样本所对应的一组采集指标时序数据。

[0163] 除了各采集指标的指标数据,训练样本中还应包括各采集指标的指标数据所对应的目标指标实际值,也就是说,训练样本应同时包括采集指标的指标数据及其对应的目标指标实际值,以目标指标为汽油收率为例,所述训练样本则具体可以包括反应塔温度、再生器温度、再生器风量等一系列采集指标的指标数据及这些指标数据对应的汽油实际收率值。

[0164] 所述第一训练样本,是指所获得的训练样本集中除第一个参与模型训练的样本之外的其他样本中的任意一个训练样本,也就是说,所述第一训练样本在训练样本集中至少对应一个在其前参与模型训练的训练样本。

[0165] 所述第一设备状态信息具体是指在对所述第一训练样本的上一个训练样本进行模型训练处理时模型所输出的设备状态信息,该设备状态信息为待预测的目标指标所对应设备的状态信息,以目标指标为汽油收率为例,则该设备状态信息为反应再生系统所对应的整体设备的设备运行状态信息。

[0166] 与传统的模型训练过程中仅以训练样本的样本数据为模型输入相区别,本申请是将第一训练样本的样本数据,以及在对第一训练样本的上一个训练样本进行模型训练处理时所输出的所述第一设备状态信息均作为模型的输入信息,即,这两种信息均参与针对第一训练样本的模型训练过程。

[0167] 本申请具体基于循环神经网络进行模型训练,从而模型训练过程中,可将待处理的第一训练样本的样本数据以及所获得的所述第一设备状态信息输入当前训练时刻所得的循环神经网络模型中,以进行模型训练。

[0168] 在获得所输入的第一训练样本的样本数据以及所述第一设备状态信息后,循环神经网络模型基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型。

[0169] 由于循环神经网络模型基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,从而在对第一训练样本进行模型训练处理时,引入了设备状态信息,考虑了设备状态信息对预测结果的影响,从而,可使得训练所得的预测模型在预测准确率方面具备较好的性能。

[0170] 这里,需要说明的是,对于训练样本集中第一个参与模型训练的训练样本而言,由于不存在在其之前参与训练的样本,从而相对应地无法获得其上一个训练样本在训练过程中输出的第一设备状态信息,针对该情况,可基于人工设置等方式,为其在进行模型训练时所需输入的设备状态信息赋予一初始值,该初始值可由所属领域的技术人员根据业务经验设定,例如将其设置为0或非0的其他数值等。

[0171] 根据以上方案可知,本实施例提供的服务器,在利用循环神经网络训练预测模型时,除了以训练样本的样本数据为模型输入,还引入了待预测指标所对应设备(即所述预定设备)的设备状态信息,并具体以第一训练样本的上一样本在训练过程中输出的第一设备状态信息作为第一训练样本在进行模型训练时的状态输入信息,这就使得设备状态信息能够按各样本的训练次序在各样本的训练过程中进行传递,相对应地能够使得训练所得的预测模型达到对训练过程中所涉及的设备状态信息的长时记忆效果,后续在利用该预测模型进行数据预测时,能够结合预测样本的样本数据以及记忆的设备状态信息对目标指标进行预测,从而可使得最终的预测结果与设备状态更加相符、具备较高的准确率。

[0172] 在接下来的实施例六中,继续对上述服务器中处理器902的数据处理功能进行进一步详述。本实施例中,所述处理器902具体可通过执行以下处理实现其数据处理功能:

[0173] 获得多个训练样本;每个训练样本包括针对所述预定设备的至少一个采集指标所获得的一组时序数据以及该组时序数据对应的目标指标实际值;

[0174] 获得在对所述第一训练样本的上一个训练样本进行模型训练处理时所输出的第一设备状态信息;所述第一设备状态信息为所述预定设备的状态信息;

[0175] 将所述第一设备状态信息以及所述第一训练样本包括的一组时序数据输入循环神经网络模型;

[0176] 利用循环神经网络模型,基于所述第一设备状态信息,对所述第一训练样本中第一个时序点的样本数据进行模型训练处理,得到第一训练样本中第一个时序点的样本数据所对应的目标指标预测值及设备状态信息;

[0177] 若第一训练样本中存在未被处理的时序点样本数据,则利用循环神经网络模型,基于第一训练样本中相邻的上一个时序点样本数据对应得到的设备状态信息,对该未被处理的时序点样本数据进行模型训练处理,得到该未被处理的时序点样本数据所对应的目标指标预测值及设备状态信息;

[0178] 将第一训练样本中最后一个时序点样本数据对应得到的目标指标预测值及设备状态信息,分别作为第一训练样本的目标指标预测值及第二设备状态信息并输出所述第二设备状态信息;所述第二设备状态信息为所述预定设备的状态信息;

[0179] 基于第一训练样本中的最后一个时序点样本数据对应的目标指标预测值以及目标指标实际值,调整所述循环神经网络模型的模型参数,以得到预测模型。

[0180] 本实施例中,每个训练样本对应的一组时序数据包括所述至少一个采集指标在预定数量的历史时序点中各个历史时序点的指标数据;每个历史时序点的指标数据对应一目标指标实际值。

[0181] 具体地可通过以下的处理过程来获得多个训练样本,也即获得训练样本集:

[0182] 1) 按时间顺序获得所述至少一个采集指标在多个历史时序点的指标数据及每个历史时序点的指标数据对应的目标指标实际值;

[0183] 2) 采用一滑动窗口以预定步长对所述多个历史时序点的指标数据及其对应的目标指标实际值进行滑动处理,滑动过程中每一个窗口内对应包括的采集指标的一组时序数据及目标指标实际值作为一个训练样本,得到多个训练样本;所述滑动窗口的时序点长度为所述预定数量。

[0184] 以下以最终构建的预测模型所需预测的目标指标为流程制造行业的目标产出物收率为例,对上述的训练样本集获取过程进行详述。

[0185] 具体以汽油收率为例,可从目标石化企业的数据库(如IP21实时数据库)中按时间顺序获取反应塔温度、再生器温度、再生器风量等各采集指标在多个历史时序点的指标数据,以及每个历史时序点的采集指标数据所对应的汽油收率;优选地,在获得多个历史时序点的采集指标数据及其对应的目标产出物收率时,任意相邻时序点间的时间间隔相同,当然,实际应用中并不以此为限。

[0186] 假设从数据库中采集了 M ($M > 1$, 且 M 为自然数)个历史时序点的采集指标数据及其对应的目标指标实际值,其中采集指标的个数为 n ($n \geq 1$, 且 n 为自然数),则具体可将采集的

M个时序点的采集指标数据及其对应的目标指标实际值表示为以下的表1:

[0187] 表1

[0188]	时间	X1	X2	...	Xn	Y
	T1	x1_1	X2_1	...	Xn_1	Y_1
	T2	x1_2	X2_2	...	Xn_2	Y_2
	T3	x1_3	X2_3	...	Xn_3	Y_3
	T4	x1_4	X2_4	...	Xn_4	Y_4
[0189]
	TM	x1_M	X2_M		Xn_M	Y_M

[0190] 在上述的表1中, T_i ($1 \leq i \leq M$, i 为自然数) 表示第 i 个时序点对应的的时间, X_j ($1 \leq j \leq n$, j 为自然数) 表示第 j 个采集指标, X_{j_i} 表示第 j 个采集指标在第 i 个时序点的指标数据, Y_i 表示第 i 个时序点的采集指标数据所对应的目标指标(如汽油收率)实际值。

[0191] 由于本申请中每个训练样本均包括采集指标的一组时序数据(同时也是循环神经网络所要求的样本形式), 从而在获得上述多个历史时序点的采集指标数据及其对应的目标指标实际值后, 需继续对其进行时序化处理。

[0192] 具体地, 可采用一滑动窗口以预定步长对所述至少一个采集指标在多个历史时序点的指标数据及其对应的目标指标实际值进行滑动处理, 滑动过程中每一个窗口内对应包括的采集指标的一组时序数据及其对应的目标指标实际值作为一个训练样本, 得到多个训练样本; 所述滑动窗口的时序点长度为所述预定数量。

[0193] 参考以下的表2, 假设所述滑动窗口的时序点长度为3, 滑动步长为1, 则如表2所示, 在滑动窗口所处的第一个窗口内包括T1-T3这3个时序点, 从而T1-T3这3个时序点的采集样本数据及其对应的目标指标实际值构成第一个训练样本Sample1, 相对应的, 在滑动窗口所处的第二个窗口内包括T2-T4这3个时序点, 从而T2-T4这3个时序点的采集样本数据及其对应的目标指标实际值构成第二个训练样本Sample2, ……在滑动窗口所处的最后一个窗口内包括T(M-2)-TM这3个时序点, 从而T(M-2)-TM这3个时序点的采集样本数据及其对应的目标指标实际值构成第(M-2)个训练样本Sample(M-2), 从而最终得到包括(M-2)个训练样本的训练样本集: $\text{Train_Set} = [\text{Sample1}, \text{Sample2}, \dots, \text{Sample}(M-2)]$ 。其中, 训练样本中所包括的采集指标的指标数据用于作为样本的特征, 而目标指标实际值则用于作为样本的标签。

[0194] 表2

[0195]

时间 ^⓪	X1 ^⓪	X2 ^⓪	... ^⓪	Xn ^⓪	Y ^⓪
T1 ^⓪	x1_1 ^⓪	X2_1 ^⓪	... ^⓪	Xn_1 ^⓪	Y_1 ^⓪
T2 ^⓪	x1_2 ^⓪	X2_2 ^⓪	... ^⓪	Xn_2 ^⓪	Y_2 ^⓪
T3 ^⓪	x1_3 ^⓪	X2_3 ^⓪	... ^⓪	Xn_3 ^⓪	Y_3 ^⓪
T4 ^⓪	x1_4 ^⓪	X2_4 ^⓪	... ^⓪	Xn_4 ^⓪	Y_4 ^⓪
... ^⓪	... ^⓪	... ^⓪	... ^⓪	... ^⓪	... ^⓪
TM ^⓪	x1_M ^⓪	X2_M ^⓪	^⓪	Xn_M ^⓪	Y_M ^⓪

[0196] 针对待处理的第一训练样本,执行以下的预定处理:

[0197] 本申请利用循环神经网络来训练预测模型,参考图4示出的循环神经网络的基本模型结构,其中,该图4具体以循环神经网络包括3个时序单元(下文简称时序单元为CELL,循环神经网络包括3个时序单元相对应地要求其训练样本包括3个时序点的样本数据)为例,示出了其所对应的模型结构,图中的三个CELL具体是同一个CELL在不同时序上的展开,训练该“循环神经网络”以得到预测模型的过程,就是不断调整所述CELL的CELL参数(模型参数)的过程。

[0198] 现有技术中,针对每个训练样本,如图4所示,对该训练样本中第一个时序点样本数据的处理,都是将第一个时序点的样本数据以及人工设定的初始状态信息(如图4中的S0,该值一般设定为0)共同作为模型的输入,即将这两种信息共同传入CELL以得到第一个时序点的目标指标输出值(对应图4中的V1_t1)及状态信息输出值(对应图4中的S1);而之后的时序点样本数据的训练,则是以该之后时序点的样本数据及其上一个相邻时序点的状态信息输出值为模型输入,将这两种信息共同传入CELL来得到该时序点的目标指标输出值及状态输出值,以图4中训练样本的第三个时序点为例,其是将第三个时序点的样本数据R1_t3以及第二时序点对应输出的状态信息S2共同传入CELL,并得到该第三个时序点的目标指标输出值V1_t3,及状态输出值S3,直至当前训练样本中最后一个时序点样本数据的训练过程结束,会基于最后一个时序点样本数据对应的目标指标输出值及目标指标实际值来调整模型参数,即调整所述CELL的CELL参数。

[0199] 需要说明的是,现有技术中,训练样本的最后一个时序点对应的状态输出值(如图4中的S3)在模型训练过程中不起任何作用,会将其直接丢弃。在当前训练样本的训练过程结束之后,会继续下一个训练样本的训练过程,其中对于每一个训练样本都需要人工设定其第一个时序点样本数据对应所需的状态输入值(一般将其初始化为0)。

[0200] 现有循环神经网络的上述特点,使得循环神经网络仅能在样本内部针对同一样本包括的各个时序点样本数据进行状态信息记忆(如目标指标所对应设备的设备状态等),无法达到长时记忆效果,为解决该问题,本申请提出了在训练样本间进行状态信息传递的技术构思。

[0201] 如图5所示,对于当前样本的最后一个时序点所对应的状态输出值(如图5中R1_t3对应的S3),本申请并不将其丢弃,而是将其作为对下一个训练样本中第一个时序点样本数据进行训练时的状态输入信息,将其与所述下一个训练样本中第一个时序点的样本数据共同输入模型。

[0202] 鉴于此,获得在对所述第一训练样本的上一个训练样本进行模型训练处理时所输

出的第一设备状态信息,具体是指获得在对所述第一训练样本的上一个训练样本进行模型训练处理时所述上一个训练样本的最后一个时序点的样本数据对应输出的设备状态信息。例如,假设所述第一训练样本为图5中的sample2,则所述第一状态信息即为图5中的sample1的最后一个时序点的样本数据R1_t3对应输出的设备状态信息S3。

[0203] 将所述第一设备状态信息以及所述第一训练样本包括的一组时序数据输入循环神经网络模型,具体可以包括将第一训练样本的上一个训练样本的最后一个时序点样本数据对应输出的设备状态信息(即所述第一设备状态信息),以及所述第一训练样本的第一个时序点的样本数据共同输入循环神经网络模型,以实现在对第一训练样本进行训练处理时,首先对第一训练样本的第一个时序点的样本数据进行模型训练处理。

[0204] 循环神经网络模型在对第一训练样本的第一个时序点的样本数据进行模型训练处理时,具体基于所述第一设备状态信息(即所述上一个训练样本的最后一个时序点的样本数据在训练时对应输出的设备状态信息),对所述第一训练样本中第一个时序点的样本数据进行模型训练处理,并得到第一训练样本中第一个时序点的样本数据所对应的目标指标预测值及设备状态信息。

[0205] 仍以图5为例,假设第一训练样本为图5中的所述sample2,则sample2的第一个时序点的样本数据R2_t1所对应的第一设备状态信息S0'即为sample1的最后一个时序点的样本数据在训练时对应输出的设备状态信息S3,从而对于sample2,会将所述S3及sample2的第一个时序点的样本数据R2_t1共同输入神经网络模型的CELL,进而神经网络模型的CELL基于S3对第一个时序点的样本数据R2_t1进行模型训练处理,并输出该第一个时序点的样本数据R2_t1所对应的目标指标预测值V2_t2及设备状态信息S1',所输出的设备状态信息S1'用于作为下一个时序点的样本数据的状态输入信息。

[0206] 循环神经网络在基于输入的所述第一设备状态信息,对第一训练样本进行模型训练处理时,会通过所述第一训练样本的相应样本数据及所述第一设备状态信息进行整合处理,来得到第一训练样本对应的整合样本;进而通过对所述整合样本进行模型训练处理,得到模型输出的对应于所述第一训练样本的目标指标预测值以及第二设备状态信息;所述第二设备状态信息为所述预定设备的状态信息,用于作为下一个训练样本的状态输入信息。

[0207] 具体地,对于第一训练样本的第一个时序点的样本数据来说,循环神经网络相应地会将该第一训练样本的上一个样本的最后一个时序点对应的状态输出信息(即所述第一设备状态信息),与该时序点的样本数据进行整合,并对整合后所得的整合样本进行模型训练处理。

[0208] 在对第一训练样本的第一个时序点的样本数据进行训练处理的基础上,若第一训练样本中还存在未被处理的其他时序点样本数据,则循环神经网络会将该未被处理时序点的上一个时序点的状态输出信息(为一设备状态信息)与该时序点的样本数据进行整合,并对整合后所得的整合样本进行模型训练处理,直至完成对该第一训练样本的最后一个时序点样本数据的训练处理后,对该第一训练样本的训练过程结束。

[0209] 且具体地,在完成对该第一训练样本的最后一个时序点样本数据的训练处理后,循环神经网络会将第一训练样本中最后一个时序点样本数据对应得到的目标指标预测值及设备状态信息,分别作为第一训练样本的目标指标预测值及第二设备状态信息并输出。

[0210] 以使得后续基于该第一训练样本的目标指标预测值通过反向传播来实现对模型参数(CELL参数)的调整,所述第二设备状态信息则具体用于作为所述第一训练样本的下一个样本中第一个时序点的状态输入信息,参与其下一个样本中第一个时序点的样本数据的训练。

[0211] 在基于第一训练样本的目标指标预测值通过反向传播来实现对模型参数(CELL参数)的调整时,具体可基于第一训练样本中最后一个时序点样本数据对应的目标指标预测值以及该最后一个时序点样本数据对应的目标指标实际值,对循环神经网络模型的模型参数进行调整,也即对循环神经网络模型中CELL的CELL参数进行调整。

[0212] 具体地,循环神经网络模型中CELL的CELL参数可以包括对应于所输入的设备状态信息的状态权重,以及对应于所输入的样本特征(每一个采集指标值可作为训练样本的一个样本特征)的特征权重,从而对循环神经网络模型的训练本质上是指对其CELL中状态权重和特征权重的训练。

[0213] 在完成对训练样本集中所有样本的训练处理后,可得到最终的能够对目标设备的目标指标(如汽油收率)进行数据预测的预测模型。

[0214] 利用本申请方案,当在对训练样本集中的样本进行训练时,仅需针对参与训练的样本的第一个时序点样本数据,为其设定一个状态输入(即需输入的设备状态信息)的初始值,该初始值可由所属领域的技术人员根据业务经验设定,例如将其设置为0或非0的其他数值等。在后续的训练过程中,每一个样本内部及不同样本之间均按时序对设备状态信息(对目标指标的影响)进行了传递,不需再针对其他任何样本进行其状态输入值的初始化设定。

[0215] 这里,需要说明的是,虽然从形式上来说,在对第一训练样本进行训练时,仅将在对第一训练样本的上一个训练样本进行模型训练处理时所输出的第一设备状态信息(而非第一训练样本之前的所有样本对应的设备状态信息),以及第一训练样本的样本数据作为模型输入,共同参与了对第一训练样本的训练过程,然而,从模型训练的整体角度而言,每一个第一训练样本在训练时都引入了其上一个训练样本在训练过程中所输出的设备状态信息,并通过对上一个训练样本的状态输出信息与当前样本的整合样本进行处理,来得到当前样本的目标指标预测值及状态输出,这就使得训练模型过程中所涉及的各个历史设备状态信息得以传递,相应地,使得循环神经网络达到了对历史训练过程中所涉及的所有设备状态信息的长时记忆效果,从而,本质上来说,本申请通过训练所得的所述预测模型,为在循环神经网络基础上所得的一长时记忆神经网络模型、

[0216] 在利用本实施例构建的预测模型进行数据预测时,能够结合预测样本的样本数据以及长时记忆的设备状态信息对目标指标进行数据预测,可使得最终的预测结果与设备状态更加相符,进而可使得预测结果具备较高的准确率。

[0217] 在接下来的实施七中,继续对上一实施例中获得多个训练样本(即获得训练样本集)的部分进行进一步说明,本实施例中,在获得多个训练样本时,所述服务器中的处理器902还可以通过调用存储器901中指令集中的相应指令,执行以下操作:

[0218] 对所述多个历史时序点的指标数据进行筛选处理,得到所述多个历史时序点的符合预定条件的指标数据。

[0219] 实际应用中,在所获得的采集指标的指标数据中,可能存在某些采集指标与待预

测的目标指标高相关、而另一些采集指标则与待预测的目标指标低相关或不相关的情况,例如,在汽油收率预测中,反应塔温度及再生塔温度这些采集指标均与汽油收率高相关,而再生器风量等采集指标则与汽油收率低相关,针对该情况,可以对所采集的多个历史时序点的指标数据进行筛选处理,以滤除掉其中的与目标指标低相关和/或不相关的采集指标的指标数据,从而使得所保留的各个采集指标的指标数据均与目标指标间具备较高的相关性。

[0220] 在对所述多个历史时序点的指标数据进行筛选处理后,可继续基于筛选处理后所得的指标数据进行窗口滑动处理,以得到时序数据形式的各个训练样本。

[0221] 本实施例通过对采集的多个历史时序点的指标数据进行筛选处理,可有效滤除掉其中的与目标指标低相关和/或不相关的采集指标数据,从而可使得所得的训练样本数据能够保持与目标指标的高相关性,进而可使得模型训练过程更为有效。

[0222] 在接下来的实施八中,所述服务器中的处理器902还可以通过调用存储器901中指令集中的相应指令,执行以下操作:

[0223] 利用预测模型对待处理的预测样本进行预测处理,得到所述预测样本对应的目标指标预测值。

[0224] 所述预测样本对应的目标指标预测值的生成,利用了所述预测模型在训练阶段所涉及的所述预定设备的所有历史状态信息以及在对所述待处理的预测样本处理之前的预测阶段所涉及的所述预定设备的所有历史状态信息。

[0225] 在基于训练样本集进行模型训练,从而得到相对应的预测模型后,可利用所得的预测模型对待处理的预测样本进行预测处理,以得到所述预测样本所对应的目标指标预测值。

[0226] 参考图8示出的基于构建预测模型并利用预测模型对目标指标进行预测的整体逻辑框架示意图,对目标指标的数据预测可以通过以下的处理过程实现:

[0227] 1) 预测样本的准备:

[0228] 对于待预测时刻(作为一个时序点),可获得该时刻对应的各采集指标的指标数据,并同时获得该时刻之前相邻的 $(p-1)$ 个历史时序点处各采集指标的指标数据,将该待预测时刻对应的各采集指标的指标数据以及其之前相邻的 $(p-1)$ 个历史时序点处各采集指标的指标数据,按时序整合为当前待预测的预测样本,其中, P 为训练样本/预测样本的一组时序数据所对应的时序点个数,该数值也即在通过窗口滑动方式获得训练样本时滑动窗口的窗口长度。

[0229] 2) 预测目标指标的指标数据:

[0230] 对于当前待处理的预测样本,可将该预测样本的第一个时序点的样本数据以及所述预测模型最近输出的设备状态信息共同作为预测模型的输入,预测模型在获得这两种输入信息后,会对两种输入信息进行整合,并对整合后所得的整合样本进行预测处理,从而输出该第一个时序点的样本数据对应的目标指标预测值及设备状态信息;

[0231] 其中,对于利用预测模型预测的第一个预测样本来说,所述预测模型最近输出的设备状态信息(作为该预测样本预测时的输入状态信息)为最后一个训练样本的最后一个时序点对应输出的设备状态信息,而对于其他预测样本,所述预测模型最近输出的设备状态信息则为其上一个预测样本的最后一个时序点对应输出的设备状态信息。

[0232] 在对待处理的预测样本的第一个时序点的样本数据进行预测处理的基础上,若该预测样本中还存在未被处理的其他时序点样本数据,则预测模型会将该未被处理时序点的上一个时序点的状态输出值与该未被处理时序点的样本数据进行整合,并对整合后所得的整合样本进行模型训练处理,直至完成对该预测样本的最后一个时序点样本数据的预测处理后,对该预测样本的预测过程结束。

[0233] 其中,该预测样本所对应的目标指标预测值,即为该预测样本的最后一个时序点样本数据对应的目标指标预测值,而该预测样本中其他时序点样本数据对应的目标指标预测值则丢弃,且模型对应于该预测样本的最后一个时序点样本数据所输出的设备状态信息,用于作为下一个预测样本的第一个时序点样本数据的输入状态信息,也即,在预测过程中每一个预测样本内部及不同预测样本之间同样均按时序对设备状态信息进行了传递,不需针对任意的预测样本进行其输入状态信息的初始化设定。

[0234] 这就使得预测样本对应的目标指标预测值的生成,不仅利用了所述预测模型在训练阶段所涉及的所述预定设备的所有历史状态信息,还利用了在对所述待处理的预测样本处理之前的预测阶段所涉及的所述预定设备的所有历史状态信息。从而,可使得最终的预测结果与设备状态更加符合,进而可使得预测结果具备较高的准确率。

[0235] 与上述的数据处理方法相对应,本申请还提供一种电子设备,所述电子设备可以但不限于是智能手机、平板电脑、台式机、笔记本、一体机等各类型终端设备,参考图10,是本申请提供的一种电子设备实施例九的结构示意图,所述电子设备包括:

[0236] 存储器1001,用于至少存储一组指令集。

[0237] 所述存储器1001可以是具备数据存储功能的各类型存储器件,如ROM(Read Only Memory,只读存储器),FLASH,光盘,硬盘等,所存储的所述指令集用于指示处理器1002执行如下文所述的相应数据处理操作,该指令集可以以程序形式存储于所述存储器1001中。

[0238] 处理器1002,用于调用并执行所述存储器中的所述指令集,通过执行所述指令集进行以下操作:

[0239] 获得多个训练样本;每个训练样本包括针对所述预定设备的至少一个采集指标所获得的一组时序数据以及该组时序数据对应的目标指标实际值;

[0240] 针对待处理的第一训练样本,执行以下的预定处理:

[0241] 获得在对所述第一训练样本的上一个训练样本进行模型训练处理时所输出的第一设备状态信息;所述第一设备状态信息为所述预定设备的状态信息;

[0242] 将所述第一设备状态信息以及所述第一训练样本包括的一组时序数据输入循环神经网络模型;

[0243] 利用循环神经网络模型,基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型。

[0244] 本申请将主要以流程制造行业中目标产出物(如石化行业中的汽油、液化气等)的收率预测所需的预测模型为例,对本申请方案进行阐述。

[0245] 所述获得多个训练样本,即指获得训练样本集,以预测模型所需预测的目标指标为汽油、液化气等产出物的收率为例,训练样本集可以包括反应再生系统在各历史时刻的反应塔温度、再生器温度、再生器风量等一系列采集指标的指标数据(这些指标数据能够体现反应再生系统整体设备的设备运行状态),其中,每个训练样本可包括各采集指标的一组

时序数据,例如,每个训练样本具体包括反应再生系统在预定数量的时序点所对应的各采集指标的指标数据等,该预定数量的时序点所对应的各采集指标的指标数据即构成该训练样本所对应的一组采集指标时序数据。

[0246] 除了各采集指标的指标数据,训练样本中还应包括各采集指标的指标数据所对应的目标指标实际值,也就是说,训练样本应同时包括采集指标的指标数据及其对应的目标指标实际值,以目标指标为汽油收率为例,所述训练样本则具体可以包括反应塔温度、再生器温度、再生器风量等一系列采集指标的指标数据及这些指标数据对应的汽油实际收率值。

[0247] 所述第一训练样本,是指所获得的训练样本集中除第一个参与模型训练的样本之外的其他样本中的任意一个训练样本,也就是说,所述第一训练样本在训练样本集中至少对应一个在其前参与模型训练的训练样本。

[0248] 所述第一设备状态信息具体是指在对所述第一训练样本的上一个训练样本进行模型训练处理时模型所输出的设备状态信息,该设备状态信息为待预测的目标指标所对应设备的状态信息,以目标指标为汽油收率为例,则该设备状态信息为反应再生系统所对应的整体设备的设备运行状态信息。

[0249] 与传统的模型训练过程中仅以训练样本的样本数据为模型输入相区别,本申请是将第一训练样本的样本数据,以及在对第一训练样本的上一个训练样本进行模型训练处理时所输出的所述第一设备状态信息均作为模型的输入信息,即,这两种信息均参与针对第一训练样本的模型训练过程。

[0250] 本申请具体基于循环神经网络进行模型训练,从而模型训练过程中,可将待处理的第一训练样本的样本数据以及所获得的所述第一设备状态信息输入当前训练时刻所得的循环神经网络模型中,以进行模型训练。

[0251] 在获得所输入的第一训练样本的样本数据以及所述第一设备状态信息后,循环神经网络模型基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,以得到预测模型。

[0252] 由于循环神经网络模型基于输入的所述第一设备状态信息,对所述第一训练样本进行模型训练处理,从而在对第一训练样本进行模型训练处理时,引入了设备状态信息,考虑了设备状态信息对预测结果的影响,从而,可使得训练所得的预测模型在预测准确率方面具备较好的性能。

[0253] 这里,需要说明的是,对于训练样本集中第一个参与模型训练的训练样本而言,由于不存在在其之前参与训练的样本,从而相对应地无法获得其上一个训练样本在训练过程中输出的第一设备状态信息,针对该情况,可基于人工设置等方式,为其在进行模型训练时所需输入的设备状态信息赋予一初始值,该初始值可由所属领域的技术人员根据业务经验设定,例如将其设置为0或非0的其他数值等。

[0254] 根据以上方案可知,本实施例提供的电子设备,在利用循环神经网络训练预测模型时,除了以训练样本的样本数据为模型输入,还引入了待预测指标所对应设备(即所述预定设备)的设备状态信息,并具体以第一训练样本的上一样本在训练过程中输出的第一设备状态信息作为第一训练样本在进行模型训练时的状态输入信息,这就使得设备状态信息能够按各样本的训练次序在各样本的训练过程中进行传递,相对应地能够使得训练所得的

预测模型达到对训练过程中所涉及的设备状态信息的长时记忆效果,后续在利用该预测模型进行数据预测时,能够结合预测样本的样本数据以及记忆的设备状态信息对目标指标进行预测,从而可使得最终的预测结果与设备状态更加相符、具备较高的准确率。

[0255] 需要说明的是,本说明书中的各个实施例均采用递进的方式描述,每个实施例重点说明的都是与其他实施例的不同之处,各个实施例之间相同相似的部分互相参见即可。

[0256] 为了描述的方便,描述以上系统或装置时以功能分为各种模块或单元分别描述。当然,在实施本申请时可以把各单元的功能在同一个或多个软件和/或硬件中实现。

[0257] 通过以上的实施方式的描述可知,本领域的技术人员可以清楚地了解到本申请可借助软件加必需的通用硬件平台的方式来实现。基于这样的理解,本申请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,该计算机软件产品可以存储在存储介质中,如ROM/RAM、磁碟、光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本申请各个实施例或者实施例的某些部分所述的方法。

[0258] 最后,还需要说明的是,在本文中,诸如第一、第二、第三和第四等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来,而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0259] 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。

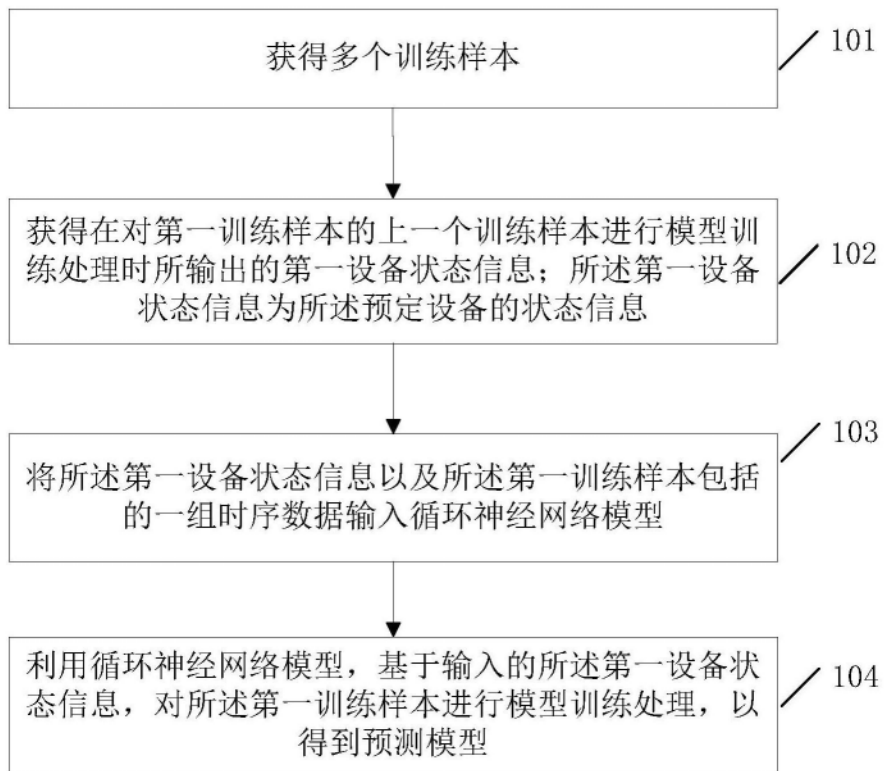


图1

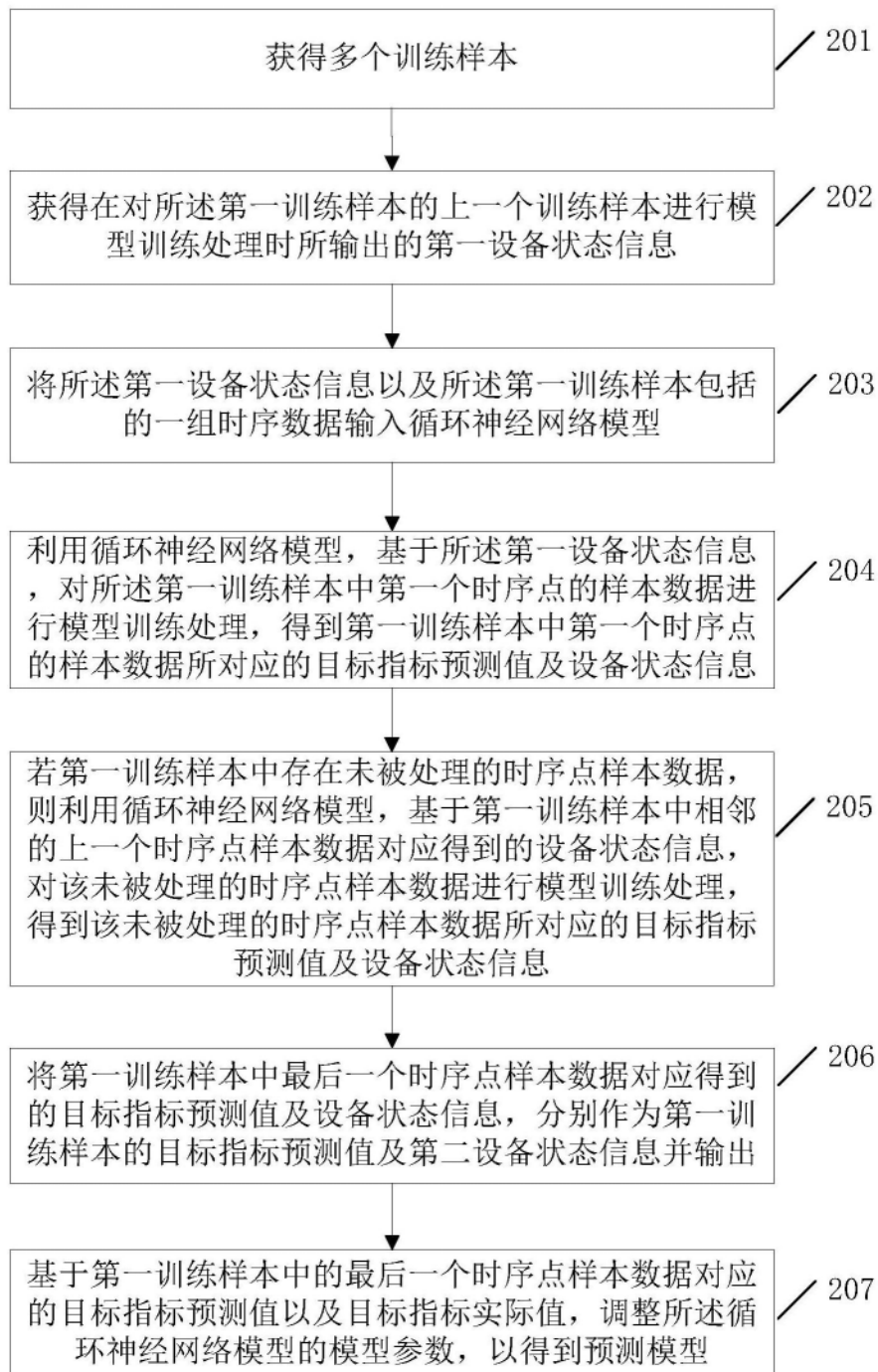


图2

按时间顺序获得所述至少一个采集指标在多个历史时序点的指标数据及每个历史时序点的指标数据对应的目标指标实际值 301

采用一滑动窗口以预定步长对所述多个历史时序点的指标数据及其对应的目标指标实际值进行滑动处理，滑动过程中每一个窗口内对应包括的采集指标的一组时序数据及目标指标实际值作为一个训练样本，得到多个训练样本 302

图3

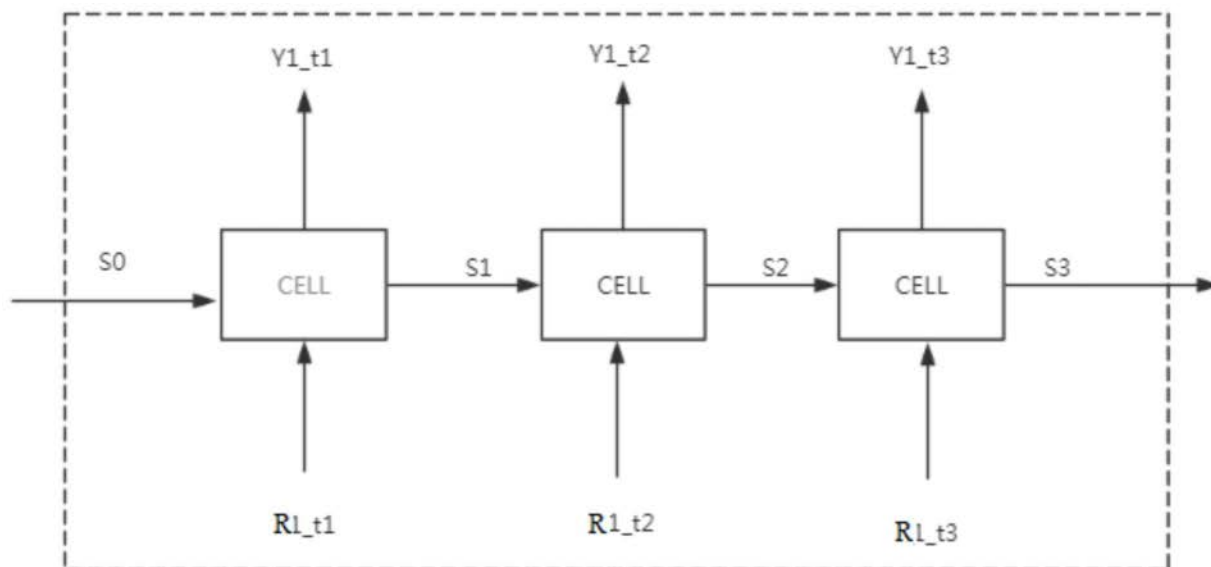


图4

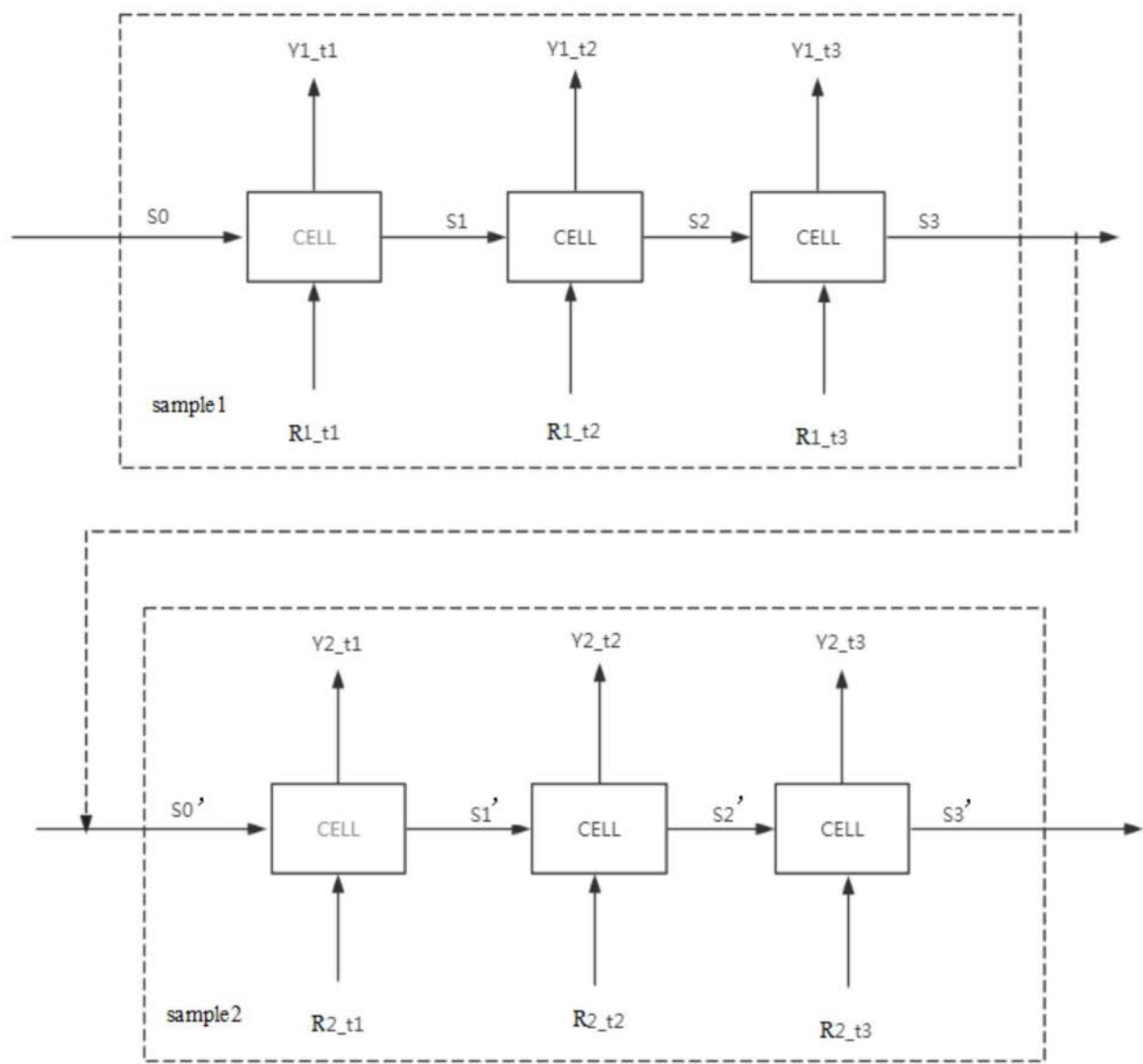


图5

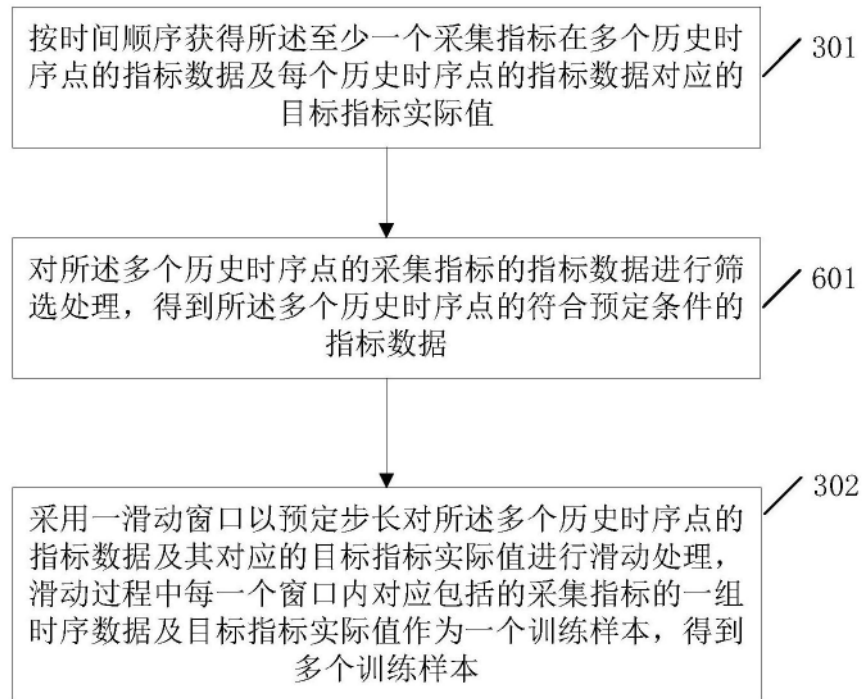


图6

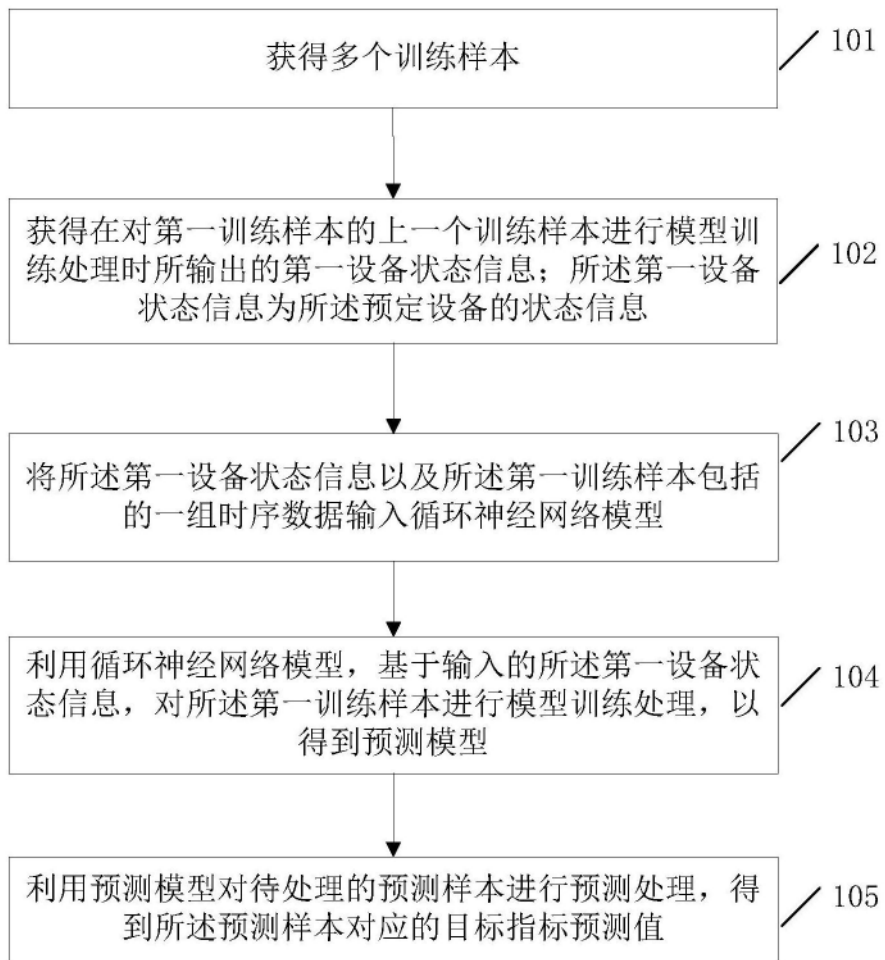


图7

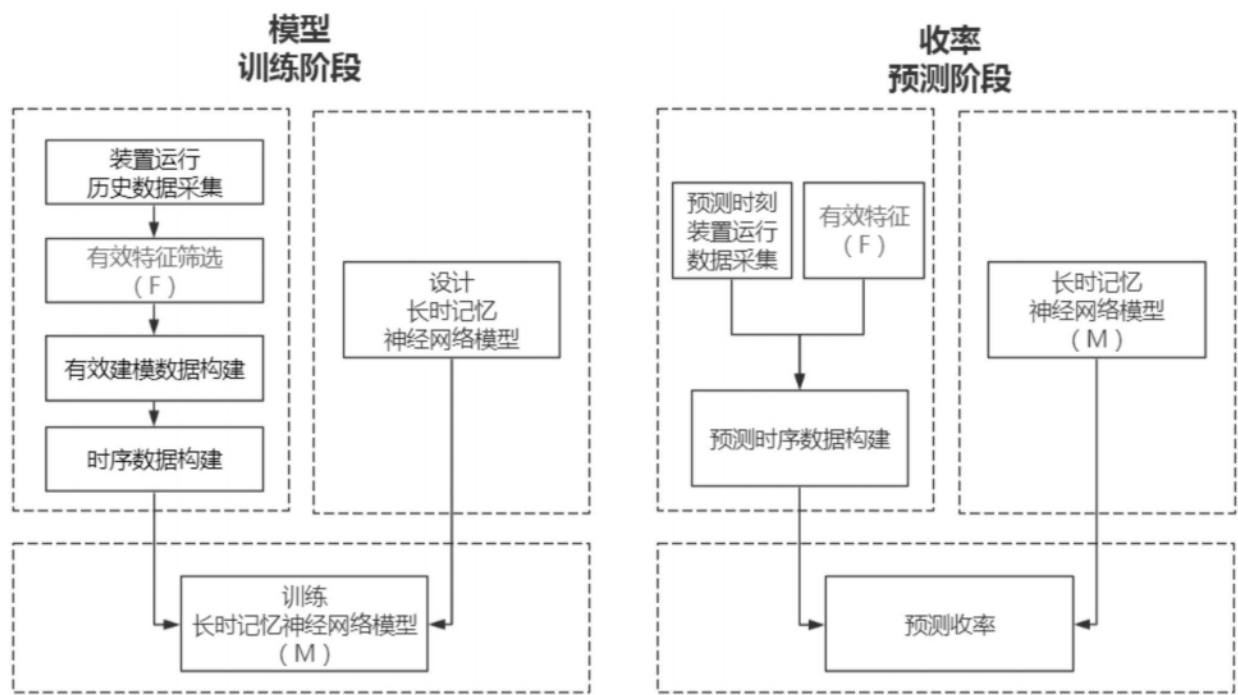


图8

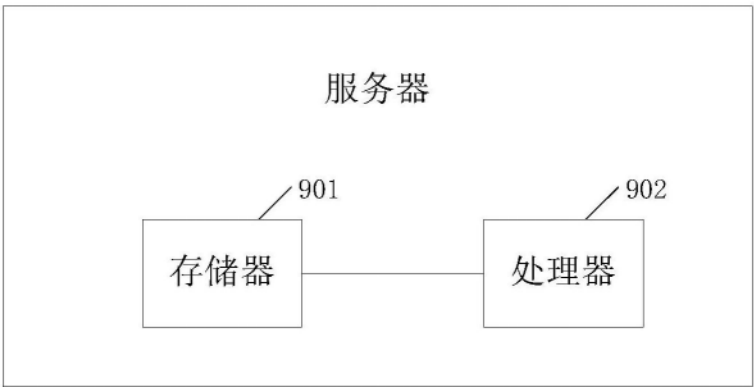


图9

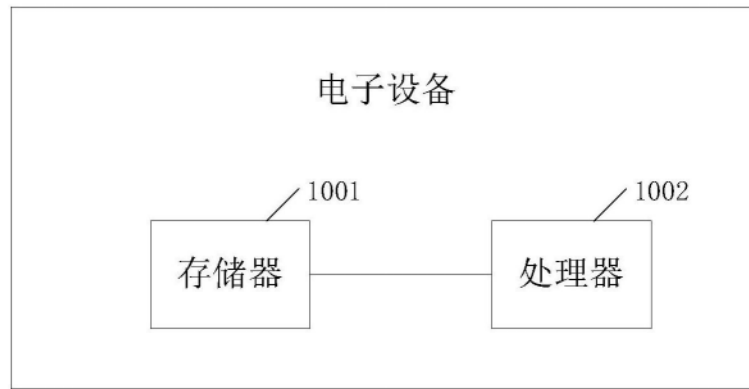


图10