



(21) 申请号 202210421439.6

(22) 申请日 2022.04.21

(30) 优先权数据

102021204040.3 2021.04.22 DE

(71) 申请人 罗伯特·博世有限公司

地址 德国斯图加特

(72) 发明人 C·朔恩 L·高尔霍夫

(74) 专利代理机构 中国专利代理(香港)有限公

司 72001

专利代理师 刘晗曦 刘春元

(51) Int.Cl.

G06N 20/00 (2019.01)

G06N 3/04 (2006.01)

G06N 3/08 (2006.01)

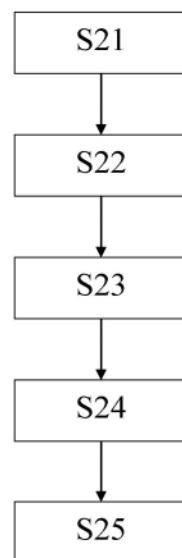
权利要求书2页 说明书8页 附图4页

(54) 发明名称

用于在车辆中创建训练数据的方法、设备和
计算机程序

(57) 摘要

本发明涉及用于探测用于机器学习系统(201)的输入参量是否适合作为用于机器学习系统(201)的附加训练数据或训练数据用于再训练或测试的方法,包括步骤:通过机器学习系统(201)处理所检测的输入参量,其中储存在通过机器学习系统(201)处理输入参量时确定的中间结果。通过异常探测器(202)处理所储存的中间结果,其中该异常探测器输出输出参量,所述输出参量表征分配给所述中间结果的所检测的输入参量是否导致机器学习系统的异常行为。基于网络(202)的输出参量存储/选择网络(201)的输入参量和附加地被定义为相关的输入参量。本发明此外涉及计算机系统、计算机程序和机器可读存储元件,其上存储有计算机程序。



1. 一种用于探测 (S22) 用于机器学习系统 (201) 的输入参量是否适合作为用于再训练的附加训练数据 (201) 或作为用于验证的测试数据的方法, 所述方法包括以下步骤:

通过所述机器学习系统 (201) 处理所检测的输入参量 (21), 其中储存在通过所述机器学习系统 (201) 处理所述输入参量时确定的中间结果 (24);

通过异常探测器 (202) 处理所储存的中间结果 (24) 中的至少一个,

其中所述异常探测器输出输出参量 (26、27), 所述输出参量表征分配给所述中间结果的所检测的输入参量 (21) 是否是关于所述机器学习系统的训练数据的异常数据点或者是触发所述机器学习系统 (201) 的异常行为的关于所述训练数据的正常数据点, 以及

如果所述异常探测器输出存在异常数据点或异常行为, 则将所检测的输入参量 (21) 标记为附加训练数据或测试数据。

2. 根据权利要求1所述的方法, 其中所述异常探测器是神经网络 (202), 其中所述神经网络已经被训练为: 如果所述输入参量 (21) 尚未以相似的形式包含在用于训练所述机器学习系统 (201) 的训练数据中, 则该神经网络 (202) 探测分配给所述中间结果的输入参量 (21) 是否是异常数据点, 和/或所述神经网络 (202) 探测所述机器学习系统是否关于经训练的行为存在异常行为。

3. 根据前述权利要求中任一项所述的方法, 其中所述异常探测器 (202) 在处理所述中间结果时获得附加参量作为输入参量, 所述附加参量是所储存的中间结果 (24) 之一的经压缩的中间结果。

4. 根据前述权利要求中任一项所述的方法, 其中对所储存的中间结果进行归一化。

5. 根据前述权利要求中任一项所述的方法, 其中如果所检测的输入参量已被标记, 则所述所检测的输入参量被添加到所述训练数据集, 并且尤其是存储在时间上紧接着所检测的输入参量而出现的多个其他输入参量并且尤其是添加到训练和测试数据集。

6. 一种用于再训练 (S24) 机器学习系统 (201) 的方法, 其中输入参量 (21) 根据前述权利要求中任一项被探测为合适的训练数据或测试数据, 其中所述机器学习系统 (201) 根据所述训练数据集以补充所标记的输入参量的方式被再训练或被测试。

7. 根据权利要求6所述的方法, 其中所述异常探测器还根据所扩展的训练数据被再训练。

8. 根据权利要求6或7中任一项所述的方法, 其中在再训练之后, 将所述机器学习系统和/或所述异常探测器的参数传输到技术系统 (10), 其中所述技术系统 (10) 根据所述机器学习系统被运行, 并且其中所述技术系统 (10) 利用所传输的参数更新所述机器学习系统 (201)。

9. 一种用于训练 (S21) 异常探测器 (202) 的方法, 所述方法包括以下步骤:

提供第一训练数据集 ($D_{\text{train in}}$) 和第二训练数据集 ($D_{\text{train out}}$),

其中所述第二训练数据集包含训练数据, 所述训练数据触发所述机器学习系统 (201) 的不一致行为, 尤其是所述训练数据并非来自所述第一训练数据集的训练数据所来自的分布;

储存所述机器学习系统的所确定的如下第一中间结果, 其中当通过所述机器学习系统处理了所述第一训练数据集 ($D_{\text{train in}}$) 的训练数据时, 所述机器学习系统 (201) 确定了所述第一中间结果;

将所储存的第一中间结果分别分配给如下标签,所述标签表征所储存的中间结果是“正常的”;

储存所述机器学习系统的所确定的如下第二中间结果,其中当通过所述机器学习系统处理了所述第二训练数据集合($D_{\text{train out}}$)的训练数据时,所述机器学习系统(201)确定了所述第二中间结果;

将所储存的第二中间结果分别分配给如下标签,所述标签表征所储存的中间结果是“非正常的”;

训练(S21)所述异常探测器,使得所述异常探测器根据所述第一和第二中间结果确定其所分配的标签。

10. 根据权利要求9所述的方法,其中所述第一和第二训练数据集合基本上是相同大小的。

11. 根据权利要求8至10中任一项所述的方法,其中根据所储存的中间结果确定归一化参数,其中所述中间结果根据所述归一化参数被归一化,并且然后被用于所述异常探测器的输入参量。

12. 异常探测器用于收集数据的用途,所述数据适用于再训练和测试机器学习系统。

13. 一种计算机程序,所述计算机程序包括指令,当在计算机上执行所述指令时,所述指令促使所述计算机执行根据前述权利要求1至11中任一项所述的方法。

14. 一种机器可读存储元件(15),其上存储有根据权利要求13所述的计算机程序。

15. 一种设备(14),所述设备被设立用于执行根据权利要求1至11中任一项所述的方法。

用于在车辆中创建训练数据的方法、设备和计算机程序

技术领域

[0001] 本发明涉及一种用于改善针对训练数据以及测试数据的数据检测的方法,该方法可以以有意义的方式有助于机器学习系统的训练或再训练。本发明同样涉及分别被设立用于执行该方法的设备和计算机程序。

背景技术

[0002] 机器学习系统,例如神经网络(英语:Deep Neural Network(深度神经网络),DNN)为了自身的训练然而也为了自身的评估而需要大量数据。在此,期望的行为、诸如分类精度和鲁棒性不仅取决于数据量,而且取决于数据的多样性及其代表性。然而,在许多应用中,难以描述所需要的数据或详细说明所需要的数据。

[0003] 然而,为了改善数据检测而有必要了解(缺失的)相关数据特征和数据内容,相关数据在这里指的是:导致DNN的所需要的行为(例如DNN的高性能和鲁棒行为)的训练数据和测试数据,其中所述测试数据可以评估所需要的行为。

[0004] 为了解决该问题,存在一些方案以使得有针对性地找到合适的的数据,所述合适的的数据可以有意义地补充数据量。

[0005] 简而言之,这迄今为止如此被解决,使得持久地收集数据、传输到服务器并且手动或也以机器方式进行选择。然而,不利的是:在该方法中必须传输巨大数据量,所述巨大的数据量引起高传输成本并且还必须被存储,这引起高存储成本。此外,该方法在许多国家中与数据保护相矛盾。

[0006] DE 10 2018 207 220公开一种用于在借助于另一神经网络在计算单元上运行待监控的神经网络时探测计算单元或存储器的计算误差或功能故障的方法。另一神经网络获得待监控的神经网络的中间结果,并且根据其中间结果确定在运行待监控的神经网络时是否出现了计算误差或功能故障。

[0007] 本发明的任务和优点

本发明的任务在于以自动化的方式并且以低的计算耗费识别相关数据并且仅收集(erheben)所述相关数据。

[0008] 本发明通过提供监控机器学习系统的小型探测器、尤其是小型神经网络的方式来解决该问题。探测器能够在运行机器学习系统时根据机器学习系统的中间结果来识别通过机器学习系统的所处理的数据是否适合用于有意义地补充数据量。

[0009] 在此情况下优点是:因此需要较少的存储/传输资源。另一优点是小型神经网络已经可以利用非常少数量的参数被运行,因此该神经网络可以在不扩展硬件的情况下与待监控的机器学习系统并行地被运行。

[0010] 此外,本发明具有以下优点:不经由数据点的分布之间的距离而进行探测,由此变得可能的是,探测训练数据的分布内的如下相关数据点,例如因为机器学习系统没有学会经由该数据点而进行泛化,因此所述相关数据点导致机器学习系统的错误结论(Aussage)。因此,本发明导致明显更可靠的数据采集。

[0011] 发明公开

在第一方面,提出一种用于借助于异常探测器探测用于机器学习系统的输入参量是否适合作为用于再训练的相关训练数据并且尤其是作为相关测试数据的以计算机实现的方法。换言之,如果输入参量可以作出以下贡献,即在利用该训练数据进行再训练时改善或稳定化机器学习系统的所需要的或期望的工作能力(Leistungsfähigkeit)或利用相关测试数据验证和/或证实所述工作能力,则所输入参量是相关训练数据。该工作能力可以是机器学习系统的预测精度或安全相关能力。机器学习系统优选地被设立用于基于计算机的视觉,尤其是用于分类、(对象)探测或语义分割。

[0012] 该方法可选地以借助于传感器、尤其是用于检测图像的传感器检测输入参量的步骤开始。之后是通过机器学习系统处理所检测的输入参量。储存在通过机器学习系统处理输入参量时确定的中间结果。处理也可以被称为传播。中间结果被理解为机器学习系统的中间结果,所述中间结果由机器学习系统确定并且由所述机器学习系统进一步使用来最终确定机器学习系统的输出参量。如果机器学习系统是神经网络,则中间结果可以被理解为神经网络的隐藏层(英语:hidden layer)的输出参量。

[0013] 异常探测器优选地处理神经网络的倒数第二层的至少一个输出参量作为中间参量,因为这些中间参量具有其在前的层的信息。最后的层是不与任何其他后续层连接的那个层。因此,倒数第二层是与最后的层连接的紧接在前的层。

[0014] 令人惊讶地,来自不同的层并且尤其是来自机器学习系统的前面部分(也称为特征提取器(Feature Extraktor))的层的中间结果对于异常探测器而言导致特别好的结果。

[0015] 紧接着通过异常探测器处理所储存的中间结果中的至少一个。异常探测器被设立用于输出输出参量,所述输出参量表征异常探测器是否探测到中间结果中的不一致性或异常,以下也称为数据异常。如果给定关于机器学习系统的训练数据的异常数据点,则可能存在所谓的数据异常。换句话说,如果通过异常探测器处理的中间结果基本上与在基于其训练数据集训练机器学习系统时确定的中间参量不相似或根本不一致,则可能存在数据异常。基本上可以理解为:中间结果由于修改而彼此不同,其中这些中间参量在通过机器学习系统进一步传播时产生等效(äquivalent)输出参量,即例如机器学习系统将给这些中间结果或其所分配的输入参量分配了相同分类。

[0016] 因此,异常探测器的输出参量可以表征:分配给中间结果的所检测的输入参量是否基本上以相似的形式已包含在机器学习系统的训练数据集中,其中异常探测器在基于其训练数据集利用来自该训练数据集的正常训练数据来训练机器学习系统时已见到所述相似的形式。因此可以说,异常探测器被构造用于识别中间参量是否表示关于训练数据分布的异常,例如来自于所述训练数据也来自于的分布,或者由训练数据定义该分布。

[0017] 此外,异常探测器可以被构造用于探测机器学习系统的数据异常和/或行为异常。如果关于训练数据的正常数据点触发机器学习系统的异常行为,则可能存在行为异常。正常数据点可以被理解为因此在训练数据中出现或将要出现的有规则的数据点。异常行为可以以如下方式表现出来:机器学习系统并非表现得像是在训练期间已学习的那样,即,并非表现得像是对于来自该数据点所属的分布中的正常数据点应该表现得那样。这具有以下优点,即可以找到如下数据点,机器学习系统针对所述数据点已经构成了错误行为、例如学习了错误关联。因此,有可能验证机器学习系统是否实际上学习了根据对象的形状对对象进

行分类或者例如根据对象颜色而以错误的方式对此进行分类。

[0018] 随后接着是：如果异常探测器输出中间参量尚未包含在训练数据中，即机器学习系统与经训练的行为相比显示出不一致的行为，则将所检测的输入参量标记为附加的相关训练数据或测试数据。即使当所确定的输出参量大于预先给定的阈值时，也可以执行标记。例如，标记可以通过标志 (Flag) 进行。然后可以使用这种标记作为用于对输入参量和/或围绕相关输入参量的所定义的输入参量区间进行数据存储的触发器。

[0019] 提出：异常探测器是神经网络，并且神经网络被训练为使得该神经网络探测所分配的中间参量的所检测的输入参量是否已包含在用于训练机器学习系统的训练数据中，尤其是导致机器学习系统的非惯常行为。训练神经网络被理解为：通过改变参数的值来优化至少与神经网络的所述参数有关的成本函数。

[0020] 此外提出，异常探测器在处理中间参量时获得附加参量作为输入参量，所述附加参量是经压缩的中间参量。例如，可以通过将中间参量的多个/所有元素相加来实现压缩。这种基于压缩的附加输入参量具有多个优点。首先，由此实现相对于用于表示函数激活的输入图像变换（例如旋转和缩放）而言的一定程度的不变性 (Invarianz)。因为这些变换自然发生在移动应用中并且这些变换不应被视为异常，所以这是值得期望的。其次，累加导致比较低维的特征激活表示。这减少所需要的参数并且因此减小异常探测器的大小。此外，在机器学习系统和异常探测器之间的数据传递被减少，这尤其是对于其中异常探测器在分开的安全监控硬件上运行的系统架构是有用的。

[0021] 此外提出，对所储存的中间结果进行归一化。优选地，对中间结果进行归一化，其方式是减去平均值并且除以对于中间参量的每个所获得的元素而言的标准偏差。为了避免除以零，可以将偏移量（例如 10^{-8} ）加到标准偏差。在训练时根据训练数据确定归一化参数。

[0022] 此外提出，如果所检测的输入参量已被标记或与异常探测器的输出参量有关，则根据使用目的将所述输入参量添加到训练或测试数据集。

[0023] 此外提出，根据前述权利要求中任一项将输入参量探测为合适的训练数据，其中机器学习系统根据训练数据集以补充所标记的输入参量的方式被再训练。再训练可以被理解为将机器学习系统的已经优化的参数重新优化，尤其是针对所补充的训练数据集而重新优化。

[0024] 此外提出，异常探测器也根据经扩展的训练数据被再训练，尤其是异常探测器利用如下中间参量被再训练，其中经再训练的机器学习系统在处理所补充的训练数据时输出所述中间参量。

[0025] 应该注意的是，可以重复用于收集所标记的输入参量的步骤，直至存在足够多的输入参量，并且然后才执行再训练。

[0026] 此外提出，在再训练之后将机器学习系统和/或异常探测器的参数传输到技术系统，其中所述技术系统根据机器学习系统被运行，并且其中所述技术系统利用所传输的参数更新所述机器学习系统。

[0027] 在本发明的另一方面，提出一种用于训练异常探测器的以计算机实现的方法。在这里异常探测器被设立用于探测：用于机器学习系统的分配给中间参量的输入参量是否适合作为进一步的训练数据或测试数据。该方法包括以下步骤：

首先,提供第一训练数据集合($D_{\text{train in}}$)和第二训练数据集合($D_{\text{train out}}$)。例如,这可以通过以下方式进行,将所提供的训练数据划分为不变的第一训练数据集合和具有分布外(OOD)训练数据的第二集合。第二集合的训练数据可以例如通过操纵(如高斯噪声、椒盐(Salt and Peper)、运动模糊等)成OOD训练数据来实现。通常,不来自第一训练数据集合的训练数据所来自的分布(在这里:分布外)的第二训练数据集合包含所述数据。分布被理解为描述可能的训练数据的虚构(fiktiv)分布。如果从虚构分布中抽取训练数据,则产生不同训练数据,所述不同训练数据鉴于一定的特性方面彼此相似。例如,分布可以描述所有可能的猫图像,其中在从分布中抽取时产生各种各样的猫图像。因为狗图像不属于猫图像的分布,因此狗图像会属于第二集合。

[0028] 紧接着之后是储存机器学习系统的所确定的如下中间结果,当通过所述机器学习系统处理了所述第一训练数据集合($D_{\text{train in}}$)的训练数据时,所述机器学习系统确定了所述中间结果。然后将所储存的中间结果分配给如下标签,所述标签表征所储存的中间结果是“正常的”(即“分布内数据”)。

[0029] 然后对于第二集合执行所述最后两个步骤。也即储存机器学习系统的所确定的如下其他中间结果,当通过所述机器学习系统处理了所述第二训练数据集合($D_{\text{train out}}$)的训练数据时,所述机器学习系统确定了所述其他中间结果。以及将所储存的其他中间结果分配给如下标签,所述标签表征所储存的中间参量是“非正常的”(即“分布外数据”)。随后,异常探测器被训练为使得所述异常探测器根据所有所储存的中间结果确定其所分配的标签。

[0030] 提出:第一和第二训练数据集合基本上是相同大小的。因此存在保持平衡的(ausbalanciert)训练数据。在此情况下有利的是,异常探测器学习更好地在异常与非异常之间进行区分。此外,这也导致异常探测器的更好的结论质量。基本上可以意味着,集合的训练数据的数量相差最大10%,优选相差最大5%并且特别优选地相差最大1%。

[0031] 此外提出,具有OOD训练数据的第二集合包含不同类型的OOD。此外提出,除了中间结果之外,附加地分别储存其他参量,所述其他参量是相应中间结果的压缩并且异常探测器获得这些经压缩的中间结果中的至少一个作为附加输入参量。

[0032] 机器学习系统可以根据所检测的输入参量确定控制参量。控制参量可以被用于控制技术系统的执行器。例如,技术系统可以是(至少部分自主的)车辆、机器人、工具、工作机器或诸如无人机之类的飞行物。

[0033] 在另一方面,提出一种计算机程序。计算机程序被设立用于执行以上提到的方法之一。该计算机程序包括命令,当计算机程序在计算机上运行时,所述命令促使计算机执行具有其所有步骤的这些所提到的方法之一。此外提出一种机器可读存储模块,其上存储有计算机程序。此外,提出一种设备,所述设备被设立用于执行所述方法之一。

附图说明

[0034] 上面提到的方面的实施例在所附附图中示出并且在随后的描述中更详细地予以阐述。在此:

图1示出车辆的示意图;

图2示出第一和第二神经网络的示意图;

图3示出用于探测合适的其他训练数据的方法的一种实施方式的示意图;

图4示出用于训练第一和/或第二神经网络的设备的一种实施方式的示意图。

具体实施方式

[0035] 图1示出车辆(10)的示意图。在另一实施例中,车辆(10)可以是服务、装配或固定生产机器人,可替代地可以是自主飞行器,例如无人机。至少车辆(10)可以包括检测单元(11)。检测单元(11)例如可以是摄像机,所述摄像机检测车辆(10)的周围环境。如雷达或激光雷达之类的其他传感器也是可设想的。检测单元(11)可以与经训练的第一神经网络(201)连接。经训练的第一神经网络(201)根据所提供的输入参量并且根据经训练的第一神经网络(201)的多个参数确定输出参量,其中所述输入参量例如由检测单元(11)提供。输出参量可以被转发给执行器控制单元(13)。执行器控制单元(13)根据经训练的第一神经网络(201)的输出参量控制执行器。在该实施例中,执行器可以是车辆(10)的发动机。

[0036] 经训练的第一神经网络(201)借助于经训练的第二神经网络(202)被监控。如果例如经训练的第一神经网络(201)的输入参量通过该神经网络被传播,所述输入参量例如在训练数据中是代表性不足的(unterrepräsentiert)或根本不包含在训练数据中,尤其是这导致网络(201)的不一致的行为,则可以借助于经训练的第二神经网络(202)探测这一点。必要时可以通过执行器控制单元(13)考虑所探测的功能故障并且相应地操控执行器。

[0037] 车辆(10)此外包括计算单元(14)和机器可读存储元件(15)。计算机程序可以存储在存储元件(15)上,所述计算机程序包括指令,当在计算单元(14)上执行所述指令时,所述指令导致计算单元(14)执行根据本发明的方法。也可设想的是,分别可以包括计算机程序的下载产品或人工生成的信号在车辆(10)的接收器处接收之后促使计算单元(14)执行根据本发明的方法。

[0038] 在另一实施例中,执行器控制单元(13)包括释放系统(Freigabesystem)。释放系统根据经训练的第一神经网络(201)的输出参量来决定:对象、例如所检测的机器人或所检测的人员是否可以访问一个区域。执行器可以优选地根据释放系统的决定而被操控。

[0039] 在一种替代实施例中,车辆(10)可以是工具或工作机器。可以借助于经训练的第一神经网络对工件的材料进行分类。在此情况下,执行器例如可以是运行磨头的发动机。

[0040] 在另一实施方式中,在测量系统中使用经训练的第一神经网络(201),所述测量系统在图中未示出。该测量系统与根据图1的车辆(10)在如下方面不同:该测量系统不包括执行器控制单元(13)。测量系统可以存储经训练的第一神经网络(201)的输出参量或者例如借助于视觉或听觉表示来表示所述输出参量,而不是将所述输出参量转发给执行器控制单元(13)。

[0041] 还可设想的是,在测量系统的一种改进方案中,检测单元(11)检测人类身体或动物身体或其一部分的图像。例如,这可以借助于光学信号、借助于超声信号或借助于MRT/CT方法来进行。在该改进方案中,测量系统可以包括经训练的第一神经网络(201),所述经训练的第一神经网络被训练以根据输入参量输出分类,例如基于该输入参量而可能存在哪种症状。在此情况下,经训练的第二神经网络(202)监控经训练的第一神经网络(201)。

[0042] 这两个经训练的神经网络(201、202)和其布线在图1中示意性地示出。

[0043] 经训练的第一神经网络(201)包括多个层,所述层分别具有多个神经元,所述神经元与在前的和后续的层的神经元连接。经训练的第一神经网络(201)的第一层获得输入参

量(21),所述输入参量在经训练的第一神经网络(201)的第一层中被处理。第一层的结果被转发给后续层,所述后续层作为输入参量获得该结果,并且根据该结果确定输出参量。输出参量随后被转发给后续层。一直执行输入参量沿经训练的第一神经网络(201)的所述逐层处理(英语:propagation(传播)),直至经训练的第一神经网络(201)的最后一层已经确定其输出参量(22)。

[0044] 经训练的第二神经网络(202)获得经训练的第一神经网络(201)的这些层中的至少一个层的至少一个输出参量作为输入参量(24),所述输入参量首先如开头所述的那样被预处理,并且然后用作输入参量,并且随后根据该输入参量(24)确定输出参量(26)。该输出参量(26)优选地表征:通过第一神经网络传播的输入参量(21)是否是异常,即该输入参量(21)是否已包含在用于训练第一神经网络的训练数据中,或者是否代表性不足等等,或触发网络(201)中的期望的/已知的行为。

[0045] 经训练的第二神经网络(202)的输入参量(24)可以例如借助于至少一个连接(25)被提供给经训练的第二神经网络(202)。

[0046] 在另一实施例中,经训练的第一神经网络(201)的这些层之一的至少一个输出参量可以包括较高维向量,所述较高维向量的各个元素以求和的方式作为经训练的第二神经网络(202)的附加输入参量被提供。可设想的是,使用类似的信息压缩方法,使得经训练的第二神经网络(202)的输入参量(24)更紧凑(kompakter)。

[0047] 附加地,经训练的第二神经网络(202)可以根据所提供的输入参量(24)确定第二输出参量(27)。第二输出参量(27)可以与经训练的第一神经网络(201)的输出参量(22)一样地表征经训练的第一神经网络(201)的输入参量,尤其是对所述输入参量进行分类。

[0048] 在另一实施方式中,第一或第二输出参量(26、27)可以仅仅是触发信号,所述触发信号例如触发经处理的数据点的数据检测。可替代地,数据检测可以经多个经处理的数据点而延伸。例如,可以在所定义的时间间隔之后或根据所定义的采集中止准则被中断,以便这样记录例如数据点序列。采集中止准则可以例如是:当第二神经网络的所考虑的输出参量下降到阈值以下时。

[0049] 在另一实施方式中,至少一个输出参量(26、27)是标量(Skalar),所述标量可以具有来自区间[0;1]中的值并且表征异常概率或表征类别、诸如“异常”和“非异常”。可替代地,输出参量可以在多个类别之间进行区分,例如是否存在针对特定类别的异常,例如由于不同的天气条件、罕见或危险的情形引起的异常。标量还可以表征输入参量中的“感兴趣区域(Region of Interest)”。在这里可以给标量的每个值分配输入参量的片段(Ausschnitt)。

[0050] 第二神经网络(202)要么可以输出这些输出参量之一,要么可替代地该网络可以共同输出这些输出参量中的多个或这些输出参量中的每一个。一般而言,可以对于所有其他实施方式和应用使用以这种方式构型的第二神经网络(202),其中如果输出参量中的要么至少一个要么多个输出参量超过预先给定的阈值,则触发数据采集。可以针对不同的输出参量分开定义所述阈值。

[0051] 第二神经网络(202)的输入参量(24)是第一神经网络(201)的至少一个中间结果(以下也称为中间参量)。然而,也可设想的是,该输入参量(24)包括直至所有的中间结果。例如,这些中间结果可以被组合成张量。应该注意的是,于是也可以根据该张量的维度来设

计第二神经网络(202)的输入。

[0052] 如果第一神经网络(201)具有(2D)卷积层,所述卷积层通常被使用在图像分类中,则层输出由多个(2D)中间结果图(英语:feature map(特征图))组成,所述中间结果图对应于该层的不同的过滤核(Filterkernen)。这些中间结果图可以直接或以经压缩的方式被添加到输入参量(24)。

[0053] 在一种优选实施方式中,除了每个中间结果图之外,附加地添加通过对相应中间结果图的所有值求和得出的单个值。

[0054] 在一种特别优选的实施方式中,第二神经网络(202)是小型前向神经网络,所述小型前向神经网络根据其输入参量(24)确定是存在异常还是非异常。

[0055] 第二神经网络(202)可以通过其他模型、例如经典/统计方法来代替。该主要优点从如下方面而得出:使用最重要的DNN的中间结果作为输入以及鉴于结果关于在训练第一神经网络(201)时存在的中间结果而言的差异来对所述结果进行分析。

[0056] 图3示出用于检测其他相关训练数据和可选地根据所检测的其他训练数据进行下游的再训练的方法的示意图。

[0057] 该方法能够以步骤S21开始。在该步骤中,对第二神经网络(202)进行训练。为此假设第一神经网络(201)已经被训练。

[0058] 可以使用具有二元交叉熵损失(英语:binary cross-entropy loss)的按标准的监督式训练和具有标准超参数的ADAM优化器(Optimierer)用于训练。

[0059] 第二神经网络(202)可以被训练用于“分布外(out-of-distribution)”(OOD)识别。为此提供训练数据,所述训练数据包括来自第一训练数据集($D_{\text{train in}}$)的分布内训练数据以及分布外训练数据集($D_{\text{train out}}$),其中第一神经网络(201)已基于所述第一训练数据集被训练。典型地,并非所有可能类型的OOD数据在设计时间点都是已知的。因此, $D_{\text{train out}}$ 应该被选择为使得第二神经网络(202)以概括化方式(verallgemeinernd)学习不同的OOD数据。

[0060] 来自所述两个训练数据集($D_{\text{train in}}$, $D_{\text{train out}}$)的训练数据通过第一神经网络(201)被传播,并且在此出现的中间结果被记录并且用二元标签标记,使得将其分类为异常的或非异常的。这些二元标签根据从哪个训练数据集中提取了相应训练数据而被分配。

[0061] 应该注意的是,用于评估第二神经网络的检验数据可以与训练数据类似地被创建。还应该注意的是,上述方法既可以被应用于学习数据异常,而且可以针对通过异常探测器引起的行为异常而被应用,其中分布外训练数据集包含相应地被操纵的训练数据,使得所述被操纵的训练数据引起相应的异常。

[0062] 附加地或可替代地,可以交替地训练这两个神经网络(201、202),其中第一神经网络首先优选地仅利用训练数据 $D_{\text{train in}}$ 被训练。

[0063] 在步骤S21的训练结束之后,接着是步骤S22。在此,探测异常。在例如车辆(10)中的第一神经网络(201)的运行中,第二神经网络(202)获得第一神经网络(201)的中间结果。

[0064] 如果输出参量(26、27)大于预先给定的阈值,或者如果输出“异常”作为类别,于是则通过第二神经网络(201)探测到异常。

[0065] 如果在步骤S22中探测到了异常,则在随后的步骤S23中,将如下输入参量添加到训练数据和/或测试数据(优选地包括标签在内),其中根据所述输入参量的中间结果通过

第二神经网络探测到了异常。在随后的可选步骤中,该输入参量被加标签。

[0066] 该输入参量然后优选地被传送给中央服务器并且在那里作为其他训练数据被添加到用于第一神经网络(201)的训练数据。这些数据也可以被用作测试数据,因为这些数据可包含“极端状况(Corner Cases)”。因此,极端状况可能包含导致网络(201)中的不期望的行为的异常。中间步骤是可能的,其方式是,所收集的第一数据在测试数据集中被评估为“极端状况”,并且最后再次基于其而再训练异常探测器,以便于是在第二数据采集时有针对性地收集训练数据。借助于这些数据,机器学习系统的所需要或期望的工作能力应该被测试或者在利用该训练数据再训练时被改善或稳定化。该工作能力可以是机器学习系统的预测精度或安全相关的能力。

[0067] 紧接着是步骤S24。在这里,根据如下训练数据集执行对第一神经网络(201)的再训练,其中所述训练数据集在上一步骤中已被扩展。有利地,在这里也根据该训练数据集并且也根据经再训练的第一神经网络的新记录的中间参量来再训练第二神经网络(202)。

[0068] 可选地,可以升级车辆中的经再训练的第一和/或第二神经网络(201、202),即在步骤S24之后将改变的参数传送给车辆,所述车辆紧接着用所传送的参数代替其神经网络的参数。

[0069] 可设想的是,步骤S22和S23多次相继地被执行,直至存在足够多的新训练数据或测试数据。在结束步骤S25之后,该方法可以在步骤S22再次开始。此外可设想的是,在执行步骤S24和/或S25期间,步骤S22和S23并行地进行。

[0070] 图4示出用于训练神经网络(201、202)、尤其是用于执行步骤S21和/或S24的设备(40)的示意图。该设备(40)包括训练模块(41)和待训练模块(42)。该待训练模块(42)包含根据图2的两个神经网络。用于训练神经网络(201、202)的设备(40)根据神经网络(201、202)的输出参量以及优选地利用可预先给定的训练数据训练神经网络(201、202)。优选地,探测网络(202)与已经训练完成的机器学习系统(201)分开地被训练。在训练期间,对储存在存储器(43)中的分别待训练的机器学习系统(201、202)的参数进行适配。

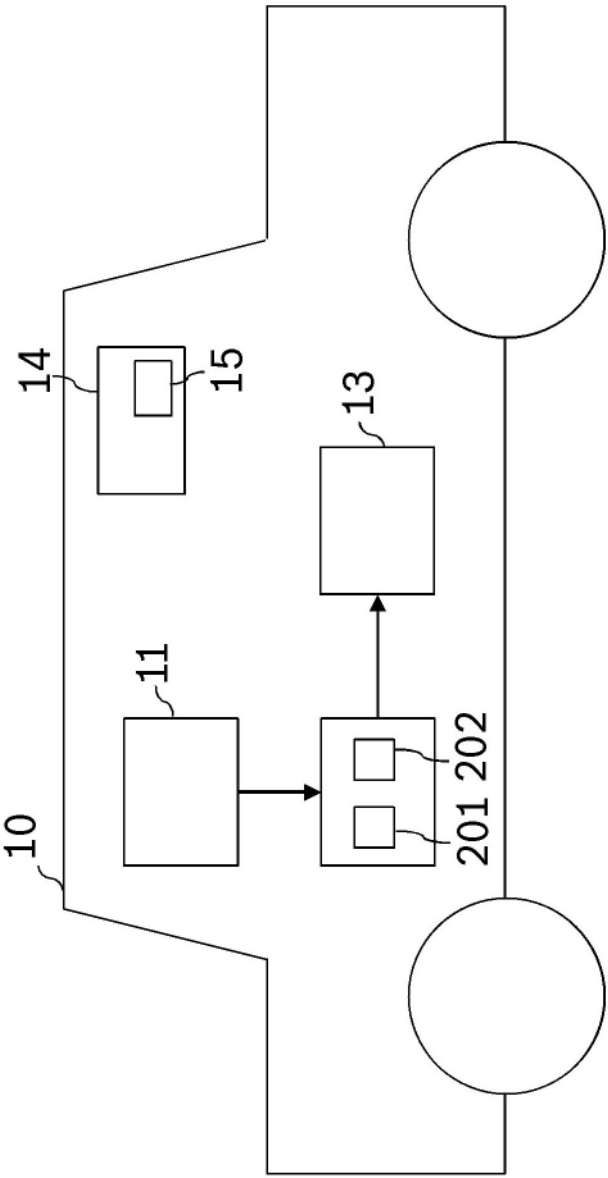


图 1

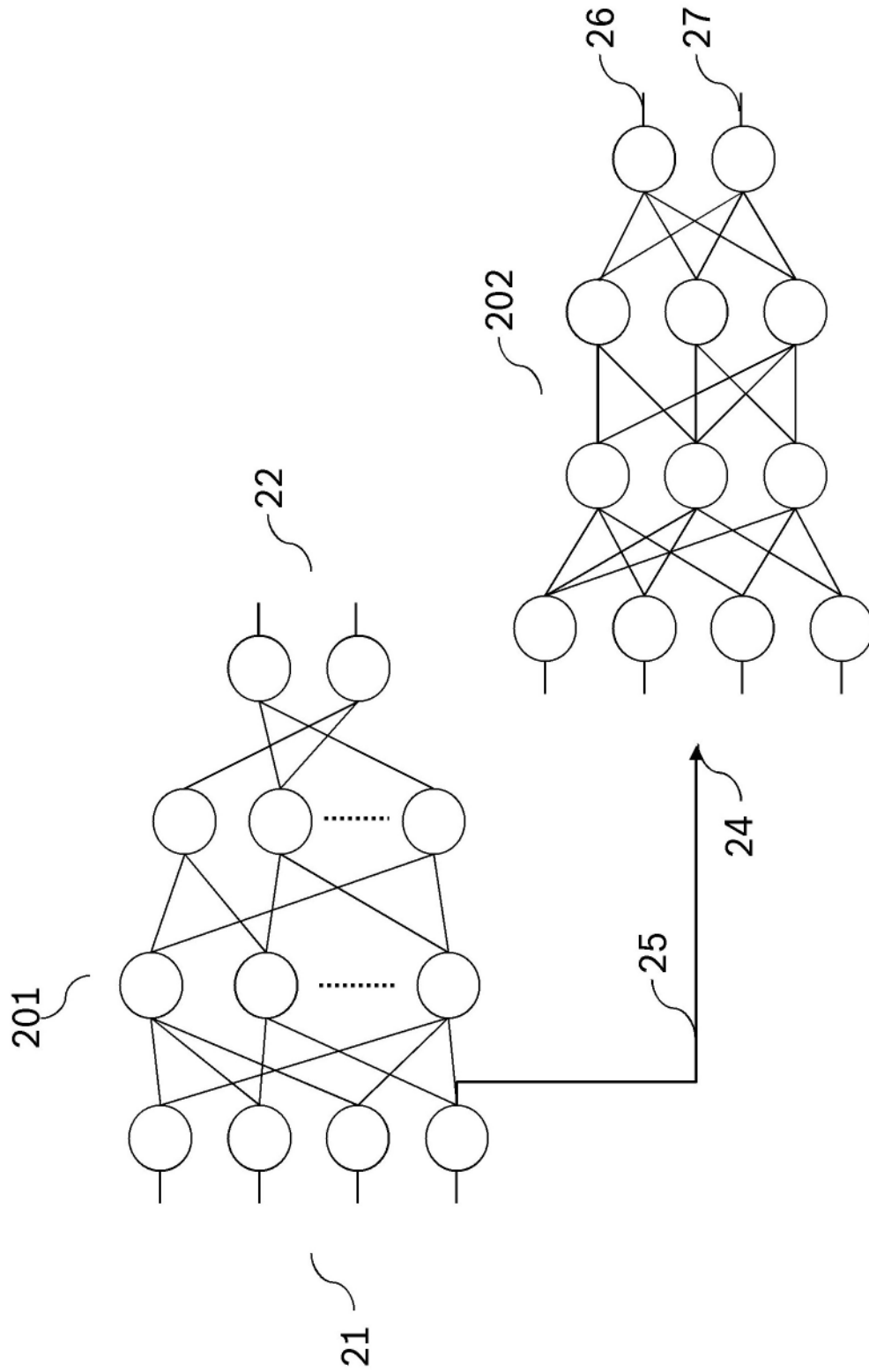


图 2

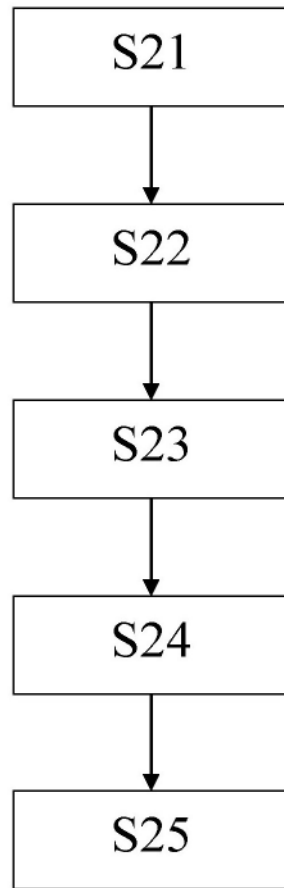


图 3

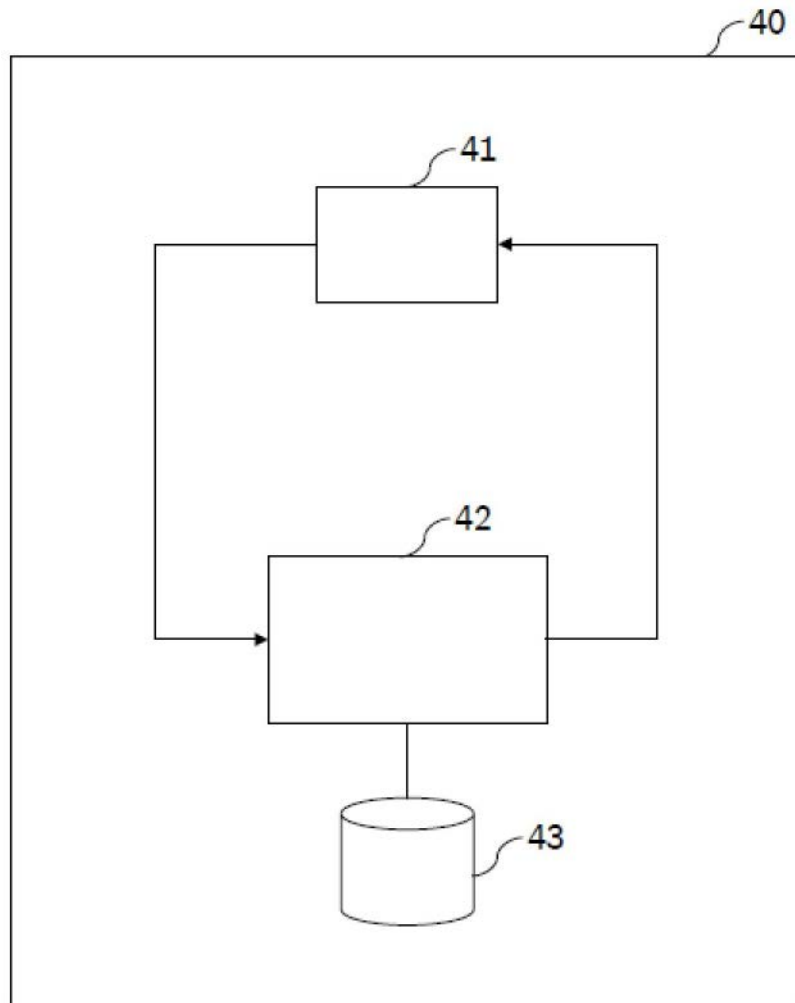


图 4