



(12) 发明专利申请

(10) 申请公布号 CN 117355848 A

(43) 申请公布日 2024. 01. 05

(21) 申请号 202280036401.5

(22) 申请日 2022.05.17

(30) 优先权数据

102021205274.6 2021.05.21 DE

(85) PCT国际申请进入国家阶段日

2023.11.20

(86) PCT国际申请的申请数据

PCT/EP2022/063345 2022.05.17

(87) PCT国际申请的公布数据

W02022/243329 EN 2022.11.24

(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 M·坎德米尔

(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 刘艺诗 吕传奇

(51) Int.Cl.

G06N 20/00 (2006.01)

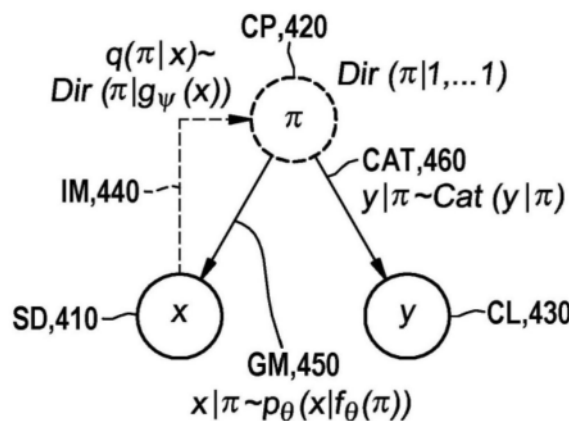
权利要求书3页 说明书18页 附图9页

(54) 发明名称

计算机控制系统的安全控制/监测

(57) 摘要

本发明涉及一种对传感器数据进行分类以用于控制和/或监测计算机控制系统的计算机实现方法(700)。分类模型包括推理模型,推理模型基于传感器数据来确定针对相应多个类别的类别概率的狄利克雷分布的相应浓度参数。分类模型进一步包括生成模型,生成模型基于类别概率,根据分类模型的训练数据集来确定传感器数据的概率分布的参数。通过确定传感器数据根据生成模型基于浓度参数被生成的概率,根据推理模型的浓度参数被用于异常检测。相同的浓度参数被用于确定类别概率。



1. 一种对传感器数据进行分类以用于控制和/或监测计算机控制系统的计算机实现方法 (700), 所述方法包括:

- 获得 (710) 表示分类模型的模型数据, 其中分类模型被配置成将传感器数据分类到来自多个类别的集合的类别中, 其中分类模型包括:

- 经训练的推理模型, 其被配置成基于传感器数据来确定针对相应多个类别的类别概率的狄利克雷分布的相应浓度参数,

- 经训练的生成模型, 其被配置成基于类别概率, 根据分类模型的训练数据集来确定传感器数据的概率分布的参数;

- 获得 (720) 传感器数据, 其中传感器数据表示计算机控制系统和/或其环境的一个或多个物理量的测量;

- 将推理模型应用 (730) 到传感器数据以获得浓度参数;

- 通过确定传感器数据根据生成模型基于浓度参数被生成的概率来确定 (740) 指示传感器数据与训练数据集的对应性的分布外值;

- 至少如果分布外值指示 (750) 充分的对应性, 则从浓度参数确定 (760) 类别概率值, 其中类别概率值指示传感器数据属于来自多个类别的集合的类别的概率, 并且输出 (770) 类别概率值以用于控制和/或监测。

2. 根据权利要求1所述的方法 (700), 其中传感器数据包括所述一个或多个物理量的测量的时间序列, 其中生成模型包括递归模型, 所述递归模型被配置成基于在前时间点处的概率分布的参数来确定在一时间点处的所述一个或多个物理量的值的概率分布的参数。

3. 根据权利要求2所述的方法 (700), 进一步包括基于所确定的类别概率值来控制车辆的车道保持支持系统, 其中传感器数据包括车辆和/或车辆的环境中的交通参与者的位置信息, 并且其中相应类别表示车辆的驾驶员的相应驾驶员任务。

4. 根据权利要求3所述的方法 (700), 进一步包括使用生成模型来预测所述时间序列之后的一个或多个未来时间点处的所述一个或多个物理量的未来值, 并且输出所预测的未来值以用于控制和/或监测。

5. 根据权利要求1所述的方法 (700), 其中传感器数据表示计算机控制系统和/或其环境的所捕获的图像。

6. 根据权利要求5所述的方法 (700), 其中分类模型是语义分割模型, 所述语义分割模型被配置成将所述图像的相应图像部分分类到相应类别中。

7. 根据权利要求6所述的方法 (700), 其中生成模型被配置成确定针对相应图像部分的相应概率分布的参数, 并且其中所述方法包括确定针对相应图像部分的相应分布外值。

8. 根据任一项前述权利要求所述的方法 (700), 其中浓度参数被限制为大于或等于一。

9. 根据任一项前述权利要求所述的方法 (700), 包括: 如果分布外值指示充分的对应性和/或类别概率值指示充分的置信度, 则使用常规控制模块来确定用于控制计算机控制系统的控制数据, 并且否则使用回退控制模块来确定所述控制数据; 以及基于所述控制数据来控制计算机控制系统。

10. 根据任一项前述权利要求所述的方法 (700), 进一步包括如果分布外值指示非对应性和/或类别概率值指示不充分的置信度, 则存储传感器数据以供未来使用, 并且否则丢弃传感器数据。

11. 根据任一项前述权利要求所述的方法 (700), 其中推理模型附加地被给予上下文变量的值作为输入, 并且其中所述方法进一步包括从上下文实例集合中确定上下文变量的值, 其中上下文实例包括传感器数据以及可选的对应的目标类别。

12. 一种训练分类模型以用于控制和/或监测计算机控制系统的计算机实现方法 (600), 其中分类模型被配置成将传感器数据分类到来自多个类别的集合的类别中, 所述方法包括:

- 获得 (610) 包括多个训练实例的训练数据集;
- 访问 (620) 表示分类模型的模型数据, 其中分类模型包括:
 - 可训练的推理模型, 其被配置成基于传感器数据来确定针对相应多个类别的类别概率的狄利克雷分布的相应浓度参数,
 - 可训练的生成模型, 其被配置成基于类别概率来确定传感器数据的概率分布的参数;
- 通过以下方式来训练 (630) 分类模型:
 - 从训练数据集中选择 (640) 训练实例, 其中所述训练实例包括传感器数据和多个类别的集合中的对应目标类别;
 - 将推理模型应用 (650) 到传感器数据以获得浓度参数;
 - 导出 (660) 用于所述训练实例的训练信号, 其中训练信号基于传感器数据基于浓度参数被分类到目标类别中的概率, 并且其中训练信号进一步基于传感器数据根据生成模型基于浓度参数被生成的概率;
 - 基于训练信号来更新 (670) 推理模型和/或生成模型的参数。

13. 一种用于对传感器数据进行分类以用于控制和/或监测计算机控制系统的系统 (200, 300), 所述系统包括:

- 数据接口 (220), 用于访问表示分类模型的模型数据 (040), 其中分类模型被配置成将传感器数据分类到来自多个类别的集合的类别中, 其中分类模型包括:
 - 经训练的推理模型, 其被配置成基于传感器数据来确定针对相应多个类别的类别概率的狄利克雷分布的相应浓度参数,
 - 经训练的生成模型, 其被配置成基于类别概率, 根据分类模型的训练数据集来确定传感器数据的概率分布的参数;
- 传感器接口 (260), 用于获得传感器数据 (224), 其中传感器数据表示计算机控制系统 (200, 300) 和/或其环境 (082, 083) 的一个或多个物理量的测量;
- 处理器子系统 (240), 其被配置成:
 - 将推理模型应用到传感器数据以获得浓度参数;
 - 通过确定传感器数据根据生成模型基于浓度参数被生成的概率来确定指示传感器数据与训练数据集的对应性的分布外值;
 - 至少如果分布外值指示充分的对应性, 则从浓度参数确定类别概率值, 其中类别概率值指示传感器数据属于来自多个类别的集合的类别的概率, 并且输出类别概率值以用于控制和/或监测。

14. 一种用于训练分类模型以用于控制和/或监测计算机控制系统的系统 (100), 其中分类模型被配置成将传感器数据分类到来自多个类别的集合的类别中, 所述系统包括:

- 数据接口 (120), 用于访问: 包括多个训练实例的训练数据集 (030); 以及表示分类模

型的模型数据(040),其中分类模型包括:

- 可训练的推理模型,其被配置成基于传感器数据来确定针对相应多个类别的类别概率的狄利克雷分布的相应浓度参数,
- 可训练的生成模型,其被配置成基于类别概率来确定传感器数据的概率分布的参数;
- 处理器子系统(140),被配置成通过以下方式来训练分类模型:
 - 从训练数据集中选择训练实例,其中所述训练实例包括传感器数据和多个类别的集合中的对应目标类别;
 - 将推理模型应用到传感器数据以获得浓度参数;
 - 导出用于所述训练实例的训练信号,其中训练信号基于传感器数据基于浓度参数被分类到目标类别中的概率,并且其中训练信号进一步基于传感器数据根据生成模型基于浓度参数被生成的概率;
 - 基于训练信号来更新推理模型和/或生成模型的参数。

15. 一种计算机可读介质(800),包括暂时性或非暂时性数据(810),表示

- 指令,当由处理器系统执行时,所述指令使得处理器系统执行根据权利要求1-11和/或权利要求12中任一项的计算机实现方法;和/或
- 表示分类模型的模型数据,其中分类模型包括:可训练的推理模型,其被配置成基于传感器数据来确定针对相应多个类别的类别概率的狄利克雷分布的相应浓度参数;以及可训练的生成模型,其被配置成基于类别概率,根据分类模型的训练数据集来确定传感器数据的概率分布的参数。

计算机控制系统的安全控制/监测

技术领域

[0001] 本发明涉及一种对传感器数据进行分类以用于控制和/或监测计算机控制系统的计算机实现方法,并且涉及一种对应的系统。本发明进一步涉及一种训练分类模型以用于控制和/或监测计算机控制系统的计算机实现方法,并且涉及一种对应的系统。本发明进一步涉及一种包括指令和/或模型数据的计算机可读介质。

背景技术

[0002] 汽车和其他车辆正在基于传感器数据的分类做出越来越多的自主决策。例如,在当今市场上可获得的车辆中,车道保持支持系统使用传感器数据,诸如车辆环境的图像或从这种图像中提取的特征,来识别驾驶员任务,例如保持车道或改变车道。车道保持支持系统然后可以使用所识别的任务以适当的方式辅助驾驶员。在未来的车辆中,越来越多的任务、以及还有如果它们被不正确地执行则会涉及越来越多风险量的任务将基于所分类的传感器数据被自主地执行。这对于许多其他计算机控制系统可能也是如此,该系统包括机器人系统、家用电器、制造机器、个人助理、访问控制系统、无人机、纳米机器人和加热控制系统。

[0003] 传感器数据的误分类可能导致错误的自动化决策,该错误的自动化决策具有潜在的危险后果。特别地,如果分类模型被应用到传感器数据输入,该传感器数据输入并不充分对应于该模型在其上被训练的数据,例如表示训练数据集中未遇到过的非常见交通情形的传感器数据,或者来自有缺陷的测量的传感器数据,则可能发生这种误分类。出于该原因,检测分布外(“OOD”)样本是重要的,该分布外(“OOD”)样本即不充分对应于分类模型在其上被训练的训练数据集的传感器数据输入。例如,如果传感器数据被检测为OOD,则有可能警告人类,切换到回退机制,和/或通过分类模型向输出指派较低的重要性。

[0004] 在M.Sensoy等人的论文“Evidential Deep Learning to Quantify Classification Uncertainty”(可在<https://arxiv.org/abs/1806.01768>获得,并且通过引用并入本文)中,提出了多类别分类模型的预测不确定性的显式建模。多类别分类问题的预测器是狄利克雷分布,其参数由神经网络的连续输出来设置。基于这些参数,可以确定类别标签和不确定性两者。

发明内容

[0005] 尽管已知的多类别分类器可以用于确定传感器数据是否是OOD,例如它是否表示传感器误读或非常见情形,但是它如下缺点:它不提供经校准的域内不确定性。域内不确定性是重要的,因为不是OOD的传感器数据可能仍然难以分类。例如,传感器数据可能表示驾驶员是保持车道还是切换车道之间的边缘情况。此外,这种类型的不确定性为了确定如何使用分类器输出来进行控制或监测是重要的,因为它还确定了控制器可能对给定分类的依赖有多么强。

[0006] 特别合期望的是如果这种域内不确定性可以以输入数据属于给定类别的概率的

形式被确定,该给定类别被校准(例如被训练)使得从该模型针对其提供了70%概率的输入集合中,该输入集合的近似70%确实属于该给定类别等。例如,具有这种经校准的不确定性值允许比较针对不同传感器输入和/或根据相同模型的不同版本所计算的不确定性信息。

[0007] 将合期望的是,提供用于对传感器数据进行分类以用于控制和/或监测计算机控制系统的技术,除了确定传感器数据是否在分布外,该技术还提供了更好的经校准的域内不确定性。

[0008] 根据本发明的第一方面,提供了一种用于对传感器数据进行分类以用于控制和/或监测计算机控制系统的计算机实现方法和对应的系统,分别如权利要求1和13所定义。根据本发明的进一步方面,提供了一种用于训练分类模型以用于控制和/或监测计算机控制系统的计算机实现方法和对应的系统,分别如权利要求12和14所定义。根据本发明的一方面,描述了一种计算机可读介质,如权利要求15所定义。

[0009] 各种方面涉及分类模型,该分类模型将传感器数据分类到来自多个类别的集合的类别中。例如,分类模型可以输出传感器数据的一个总体分类,或者分类模型可以包括针对传感器数据的相应部分的到该类别集合中的多个相应分类。例如,对于时间序列传感器数据,可以确定针对相应时间点的相应类别。对于图像数据,分类模型可以确定针对相应图像部分(例如,像素)的相应类别,例如分类模型可以是图像分割模型。

[0010] 本文中描述的分类模型可以包括经训练的推理模型和经训练的生成模型两者。给定传感器数据,推理模型可以针对传感器数据可以被分类到的相应类别确定类别概率的狄利克雷分布的浓度参数。有趣的是,这些浓度参数可以用作输入,以使用生成模型来确定分布外值、以及确定针对相应类别的类别概率值(至少如果分布外值指示传感器数据与训练数据集的充分对应性)两者。给定类别概率,生成模型可以被训练成根据训练数据集来确定传感器数据的概率分布的参数。因此,基于所确定的浓度参数,可以确定传感器数据根据生成模型而生成的概率,并且该概率可以充当指示传感器数据与分类模型的训练数据集的对应性的分布外值。此外,可以从浓度参数确定传感器数据属于特定类别的概率。

[0011] 有趣的是,发明人发现,使用浓度参数来确定分布外值和确定类别概率两者得到了更好的经校准的类别概率。也就是说,当训练分类模型时,可以有效地迫使分类模型以如下这种方式确定浓度参数,使得它们不仅允许预测正确的类别,而且它们还允许传感器数据被生成模型足够准确地再现。发明人发现,这充当迫使该模型确定更好地计及了关于传感器数据的不确定性的浓度参数的有效机制。

[0012] 例如,在“Evidential Deep Learning to Quantify Classification Uncertainty”中,不存在这种机制来提供经校准的类别概率。事实上,该模型可以通过模型拟合和正则化的组合来训练。模型拟合鼓励该模型以最大确定性来输出训练数据的实际标签。正则化鼓励该模型指示它对于分布外值是不确定的。然而,该模型拟合或该正则化都没有提供鼓励该模型针对分布内数据提供经校准的概率值的机制。相比之下,如本文中所描述,通过使用也用作去往生成模型以用于OOD检测的输入的浓度参数提供了这种机制。实际上,发明人发现,根据所提供的技术导出的类别概率值在校准方面比从“Evidential Deep Learning to Quantify Classification Uncertainty”的狄利克雷分布导出的类别概率值准确得多。

[0013] 有趣的是,由于分布外值是使用生成模型确定的——该生成模型被训练成从浓度

参数再现传感器数据,因此可以使得能够在不需要OOD样本的情况下训练分类模型。这是有利的,因为根据定义难以获得OOD样本的代表性集合。在OOD样本的显式集合上训练模型导致模型可能无法检测与OOD训练集不相似的OOD案例。因此,仅通过域内样本来学习其域对于OOD检测的准确性是有利的。例如,在替代方法中,分类模型可以被训练成输出针对训练数据集的域内样本的正确类别,并且输出针对训练数据集的分布外样本的最大不确定性。这种训练方法具有需要OOD样本的缺点,并且也没有提供一种机制来提供经校准的域内概率值。

[0014] 相比之下,通过在域内样本上训练生成模型并且使用该模型来确定分布外值,如本文中提出的,使得能够仅在域内样本上进行训练,并且可以确定经校准的域内概率值。特别地,分类模型可以在包括传感器数据输入和对应目标类别的训练数据集上被训练。可以通过将推理模型应用到传感器数据以获得浓度参数,并且基于浓度参数导出训练信号,来训练该模型。然后,训练信号可以用于更新推理模型和/或生成模型的参数。

[0015] 有趣的是,浓度参数可以用于训练分类和OOD检测两者。为此,训练信号可以包括对证据性分类器训练的贡献和对密度估计的贡献。因此,该模型可以基于浓度参数来确定类别概率值,该浓度参数也被训练成让生成模型再现传感器数据,从而得到经改进的校准。

[0016] 特别地,对于证据性分类器训练,训练信号可以基于传感器数据基于浓度参数被分类到目标类别中的概率。浓度参数可以定义针对相应类别的类别概率上的概率分布,并且这些类别概率又可以定义传感器数据被分类到目标类别中的概率。训练信号可以被配置成最大化针对目标类别的该概率。因此,可以鼓励该模型提供传感器数据的正确分类。

[0017] 对于密度估计,训练信号可以进一步基于传感器数据根据生成模型基于浓度参数被生成的概率。如所讨论的,浓度参数可以定义针对相应类别的类别概率上的概率分布。基于这些类别概率,生成模型可以确定传感器数据上的概率分布的参数。训练信号可以被配置成最大化训练数据集的传感器数据根据该概率分布而出现的概率。因此,该模型可以以如下这种方式来训练,使得针对根据训练数据集的传感器数据,应用推理模型、接着是生成模型,导致了所述传感器数据比不根据训练数据集的传感器数据有相对更高的概率被再现。因此,该概率可以用作分布外值。这得到了更准确的分布外值,但是同时也提供了类别概率值的校准,因为这些是从也在密度估计中使用的相同浓度参数中确定的。

[0018] 可选地,应用分类模型可以涉及输出分布外值。例如,可以通过将传感器数据根据生成模型被生成的概率与阈值进行比较来获得分布外值。针对一个或多个类别(例如针对一个或多个最可能的类别、针对一个或多个请求的类别、或针对所有类别)的类别概率值也可以被输出,例如,始终被输出,或者仅在分布外值指示充分的对应性时才被输出。

[0019] 输出(例如分布外值和/或类别概率)可以用于控制和/或监测计算机控制系统。例如,计算机控制系统可以是机器人、车辆、家用电器、电动工具、制造机器、个人助理或访问控制系统。这些是系统的示例,所述系统基于从该系统和/或其环境测量的传感器数据被控制,并且因此具有如下问题:即,它们可能被应用在分类器模型没有针对其被训练的情形中,或者可能获得有缺陷的传感器测量。因此,可以有利地应用本文中提供的技术。可以被监测的示例系统包括监控系统或医学(成像)系统。

[0020] 原则上,还有可能将所提供的技术用于除了控制和/或监测计算机控制系统之外的使用分布外值和/或类别概率值的应用。因此,所提供的方法可以是对传感器数据(例如

图像)本身进行分类的方法,并且所提供的系统也可以是用于对传感器数据(例如图像)本身进行分类的系统,而没有必要将分类结果用于控制和/或监测。

[0021] 一般来说,本文中提供的技术可以应用到许多不同种类的传感器数据,例如图像数据、音频数据和各种其他类型的物理量测量,诸如温度、压力等。本文中提供了各种示例。特别地,如本文中提供的多个相应分类模型可以被应用到多个相应类型的传感器数据,从而获得具有相互可比较的类别概率值的分类。原则上不需要对传感器数据的应用,例如分类模型可以针对其需要OOD检测和/或类别概率确定的任何模型输入进行操作。

[0022] 可选地,传感器数据可以包括一个或多个物理量的测量的时间序列,例如至少2个、至少5个或至少10个时间点。在实践中,时间序列的分类常常是在控制和/或监测计算机控制系统时的重要输入。时间序列的分类可以包括确定针对整个时间序列的类别和/或确定针对时间序列的相应时间点的相应类别。可以使用本身已知的适当推理模型。生成模型可以包括递归模型,该递归模型被配置成基于时间序列的在前时间点处的概率分布的参数来确定时间序列的一时间点处的一个或多个物理量的值的概率分布的参数。例如,递归模型可以接收时间序列的总体分类作为总体输入,或者可以接收相应时间点处的类别概率作为用于确定该时间点处的概率分布参数的输入。

[0023] 可选地,分类模型可以用于车辆的车道保持支持系统中。传感器数据可以包括(例如从安装在车辆上的一个或多个相机的图像数据中提取的)车辆和/或车辆环境中的交通参与者的位置信息。该位置可以用于将传感器数据分类到传感器数据所表示的不同驾驶员任务中。特别地,驾驶员任务可以包括表示车辆保持其车道的类别以及表示车辆改变其车道(例如,向左改变或向右改变)的一个或多个类别。

[0024] 可选地,包括递归模型的生成模型可以用于预测时间序列之后的一个或多个未来时间点处的一个或多个物理量的未来值,该未来时间点例如位于未来的、或至少针对其还没有可用测量的一个或多个时间点。这些所预测的未来值然后可以用于控制和/或监测计算机控制系统。例如,通过在时间上进一步推出(rolling out)生成模型,可以在时间点 t 处使用递归生成模型来预测最接近的车辆在一个或多个未来时间点 $t+1$ 、 $t+2$ 处的位置。未来预测可以例如由自动驾驶系统的规划和控制模块来使用,以生成针对速度和/或转向角的控制。

[0025] 可选地,传感器数据可以表示计算机控制系统和/或其环境的所捕获图像。例如,传感器数据可以是视频数据、雷达数据、激光雷达数据、超声数据、运动数据或热图像数据。作为推理和生成模型,可以使用对图像进行操作的各种机器可学习模型,例如卷积神经网络、并且特别是全卷积神经网络,如本领域本身已知的。因此,分类模型可以是图像分类模型。例如,图像分类模型可以将图像数据分类到对应于该图像中存在多个相应对象中的哪一个的多个相应类别中。图像分类模型也可以输出图像中是否存在一个或多个相应对象,例如,图像分类模型可以是对象检测模型。图像分类模型还可以将相应图像部分(例如,个体像素)分类到相应类别中,例如,图像分类模型可以是语义分割模型。例如,图像分类模型的多个类别可以包括交通标志、道路表面、行人、车辆和/或检测到的车辆的驾驶员任务。

[0026] 可选地,当分类器模型是语义分割模型时,生成模型可以被配置成确定针对相应图像部分的相应概率分布的参数。因此,生成模型可以用于确定针对相应图像部分的相应分布外值。分布外值可以用于突出显示其标签不确定的图像区域。作为结果,例如,这些突

出显示的图像区域可以被在计算机控制系统的控制和/或监测中排除,或者可以针对这种区域优先考虑其他模态。

[0027] 可选地,确定分布外值可以包括从狄利克雷分布采样针对多个类别的类别概率;从类别概率来确定传感器数据的概率分布的参数;以及根据概率分布的参数来确定传感器数据被采样的概率。这允许高效地将分布外值确定为传感器数据根据生成模型被生成的概率的近似,例如,如果概率不能用封闭形式的表达式来表达,也可以使用该近似。

[0028] 可选地,浓度参数可以被限制为大于或等于一,例如,推理模型可以被定义为使得它仅能够输出大于或等于一的值。虽然原则上没有必要通过限制推理模型来强制这一点,但是使浓度参数大于一确保了参数位于狄利克雷分布的更规则部分,从而改进了训练期间的数值稳定性。

[0029] 可选地,(一个或多个)所确定的分布外值和/或类别概率值可以用于在是使用常规控制模块还是回退控制模块之间决定。如果分布外值指示充分的对应性和/或针对类别的类别概率值指示针对到该类别中的分类的足够置信度,则常规控制模块可以用于确定控制数据。否则,可以使用回退控制模块来确定控制数据。然后,可以基于控制数据来控制计算机控制系统。因为所提供的技术提供了更准确的OOD检测和更准确的类别概率,所以通过在需要的情况下切换到回退控制模块改进了安全性,同时通过在确定这可能时使用常规控制模块改进了控制。

[0030] 可选地,如果分布外值指示非对应性,则可以存储传感器数据以供未来使用。否则,传感器数据可以被丢弃。这具有如下优点:它使得能够更高效地收集相关训练数据。例如,可以在所收集的数据上训练另外的机器学习模型。自主车辆和其他类型的计算机控制系统通常收集大量数据,并且存储所有收集的数据和/或传输所有收集的数据以供训练系统使用常常是不可行的。因此,选择预期是相关的训练数据是重要的。例如,传感器数据可以是用于自动驾驶感知的车辆传感器数据。这种传感器数据可以通过在自由形式驾驶期间与真实世界交互来获得。数据可能不断地流动到传感器中,但可能没有足够的空间来存储所有数据。所确定的类别概率值和/或所确定的分布外值可以用于确定是否保留传感器数据,例如,通过对存储传感器数据的预期信息增益进行信息论(information-theoretic)计算。传感器数据可以存储在车辆本身处,但是也可以从传感器发送到中央服务器并存储在那里。因此,可以实现智能数据选择,并且可以降低数据收集成本。

[0031] 可选地,生成模型可以被给予上下文变量的值作为输入。上下文变量可以确定关于类别概率的先验分布,例如浓度参数。当应用该模型时,可以从上下文实例的集合中确定上下文变量的值,每个上下文实例包括传感器数据和对应的目标类别。这提供了更灵活的分类型模型,该分类型模型可以容易地动态适配于不同的境况。例如,上下文实例可以是计算机控制系统的物理量的最近测量。例如,在车道保持支持系统中,上下文实例可以在由上下文实例表示的时间跨度上有效地表示驾驶员意图。还有可能使用更长时间跨度的测量和/或来自其他计算机控制系统的测量,使得例如自主或半自主车辆中的分类型模型可以适配于不同类型的天气、不同类型的道路、不同的国家等。

[0032] 本领域技术人员将领会,本发明的两个或更多个上述实施例、实现方式和/或可选方面可以以任何被认为有用的方式被组合。

[0033] 基于本说明书,本领域技术人员可以实行对任何系统和/或任何计算机可读介质

的修改和变型,这些修改和变化对应于对应的计算机实现方法的所描述的修改和变型。

附图说明

[0034] 本发明的这些和其他方面将通过参考实施例并且参考附图而清楚并且将参考实施例并且参考附图被进一步阐述,这些实施例在下面的描述中以示例的方式被描述,在附图中:

[0035] 图1示出了用于训练分类模型的系统;

[0036] 图2示出了用于对传感器数据进行分类的系统;

[0037] 图3示出了用于对传感器数据进行分类以用于控制(半)自主车辆的系统;

[0038] 图4a示出了不包括生成模型的分类型模型的详细示例;

[0039] 图4b示出了包括生成模型的分类型模型的详细示例;

[0040] 图4c示出了包括生成模型并且使用上下文变量的分类型模型的详细示例;

[0041] 图5a示出了如何使用包括生成模型的分类型模型对传感器数据进行分类的详细示例;

[0042] 图5b示出了如何训练包括生成模型的分类型模型的详细示例;

[0043] 图6示出了训练分类模型的计算机实现方法;

[0044] 图7示出了对传感器数据进行分类的计算机实现方法;

[0045] 图8示出了包括数据的计算机可读介质。

[0046] 应该注意的是,各图纯粹是图示性的,并且没有按比例绘制。在各图中,与已经描述的元素相对应的元素可以具有相同的附图标记。

具体实施方式

[0047] 图1示出了用于训练分类模型以用于控制和/或监测计算机控制系统的系统100。分类模型可以被配置成将传感器数据分类到来自多个类别的集合的类别中。

[0048] 系统100可以包括数据接口120。数据接口120可以用于访问训练数据集030。训练数据集030可以包括多个训练实例,例如,至少1000个、至少10000个或至少100000个。训练实例可以被标记为各自包括传感器数据和来自多个类别的集合的对应目标类别的训练实例。训练数据集030通常不包含被指定为分布外的训练实例,例如,集合030可以是分布内训练数据的集合。

[0049] 数据接口120可以用于进一步访问模型数据040。模型数据040可以表示分类模型。分类模型可以包括可训练的推理模型。推理模型可以被配置成基于传感器数据来确定针对相应多个类别的类别概率的狄利克雷分布的相应浓度参数。分类模型可以进一步包括可训练的生成模型。生成模型可以被配置成基于类别概率来确定传感器数据的概率分布的参数。例如,模型数据040可以包括推理模型和/或生成模型的可训练参数。例如,推理模型的可训练参数的数量可以是至少1000个、至少10000个或至少1000000个。例如,生成模型的可训练参数的数量可以是至少1000个、至少10000个或至少1000000个。模型数据040可以用于根据本文中描述的方法、例如通过图2的系统200或图3的系统300来控制/或监测计算机控制系统。

[0050] 例如,也如图1中所图示,输入接口可以由数据存储接口120构成,数据存储接口

120可以从数据存储装置021访问数据030、040。例如,数据存储接口120可以是存储器接口或永久存储接口,例如硬盘或SSD接口,但是也可以是个人、局域或广域网接口,诸如蓝牙、Zigbee或Wi-Fi接口或以太网或光纤接口。数据存储装置021可以是系统100的内部数据存储装置,诸如硬盘驱动器或SSD,但是也可以是外部数据存储装置,例如网络可访问的数据存储装置。在一些实施例中,数据030、040可以各自从不同的数据存储装置中被访问,例如,经由数据存储接口120的不同子系统。每个子系统可以属于如上针对数据存储接口120所描述的类型。

[0051] 系统100可以进一步包括处理器子系统140,处理器子系统140可以被配置成在系统100的操作期间训练分类模型040。

[0052] 为了训练该模型,处理器子系统140可以从训练数据集030中选择训练实例。训练实例可以包括传感器数据和多个类别的集合中的对应目标类别。处理器子系统140可以进一步将推理模型应用到传感器数据,以获得浓度参数。处理器子系统140可以进一步导出用于训练实例的训练信号。训练信号可以基于传感器数据基于浓度参数被分类到目标类别中的概率。训练信号可以进一步基于传感器数据根据生成模型基于浓度参数被生成的概率。处理器子系统140可以基于训练信号来更新推理模型和/或生成模型的参数040。

[0053] 系统100可以进一步包括输出接口,用于输出表示经学习(或“经训练”)模型的训练数据040。例如,也如图1中所图示,输出接口可以由数据接口120构成,在这些实施例中,所述接口是输入/输出(“IO”)接口,经由该输入/输出(“IO”)接口,经训练模型数据可以被存储在数据存储装置021中。例如,定义“未训练”分类模型的模型数据040可以在训练期间或之后至少部分地被经训练模型的模型数据所替换,因为该模型的参数、诸如神经网络的权重和其他类型的参数可以被适配成反映在训练数据030上的训练。这也在图1中由相同的附图标记040示出,该附图标记040指代数据存储装置021上的经训练和未训练模型数据两者。在其他实施例中,经训练模型数据可以与定义了“未训练”动态模型的模型数据分离地存储。在一些实施例中,输出接口可以与数据存储接口120分离,但是一般而言可以属于如上针对数据存储接口120所描述的类型。

[0054] 图2示出了用于对传感器数据224进行分类以用于控制和/或监测计算机控制系统的系统200。

[0055] 系统200可以包括数据接口220,用于访问用于访问表示分类模型的模型数据040。分类模型可以被配置成将传感器数据分类到来自多个类别的集合的类别中。分类模型可以包括经训练的推理模型。推理模型可以被配置成基于传感器数据来确定针对相应多个类别的类别概率的狄利克雷分布的相应浓度参数。分类模型可以进一步包括经训练的生成模型。生成模型可以被配置成基于类别概率根据分类模型的训练数据集来确定传感器数据的概率分布的参数。模型数据可以包括生成模型和/或推理模型的训练参数。模型数据040可能先前已经如本文中描述的那样被训练,例如通过图1的系统100或如其他地方所描述。除了应用模型之外,系统200还可以训练该模型,例如,系统200可以与图1的系统组合。

[0056] 例如,也如图2中所图示,数据接口可以由数据存储接口220构成,数据存储接口220可以从包含数据040的数据存储装置022访问数据040。一般而言,数据接口220和数据存储装置022可以属于与参考图1针对数据接口120和数据存储装置021所描述的类型相同的类型。存储装置022可以是系统200中的一部分,但是也可以是外部的。数据存储装置022还

可以可选地包括传感器数据。

[0057] 系统200可以进一步包括处理器子系统240,处理器子系统240可以被配置成在系统200的操作期间将推理模型应用到传感器数据以获得浓度参数。处理器子系统240可以进一步被配置成确定指示传感器数据与训练数据集的对应性的分布外值。可以通过确定传感器数据根据生成模型基于浓度参数被生成的概率来确定分布外值。处理器子系统240可以进一步被配置成至少如果分布外值指示充分的对应性(例如,小于或大于对应性阈值),则从浓度参数确定类别概率值。类别概率值可以指示传感器数据属于来自多个类别的集合的类别的概率。处理器子系统240可以进一步被配置成输出类别概率值,以用于控制和/或监测计算机控制系统。

[0058] 将领会的是,与图1的处理器子系统140一样,相同的考虑和实现选项适用于处理器子系统240。将进一步领会,与图1的系统100一样,相同的考虑和实现选项一般而言可以适用于系统200,除非另有指出。

[0059] 系统200可以包括传感器接口260,用于例如从传感器072获得传感器数据224。传感器可以布置在环境082中,但是也可以远离环境082布置,例如如果可以远程地测量(一个或多个)量的话。传感器072可以但不需要是系统200的一部分。传感器072可以具有任何合适形式,诸如图像传感器、激光雷达传感器、雷达传感器、压力传感器、容器温度传感器等。在一些实施例中,传感器数据072可以感测不同物理量的测量,因为它可以从感测不同物理量的两个或更多个不同传感器来获得。传感器数据接口260可以具有在类型上对应于传感器类型的任何合适的形式,包括但不限于低级通信接口,例如基于I2C或SPI数据通信,或者如上针对数据接口220所描述的类型的数据存储接口。

[0060] 在一些实施例中,系统200可以包括致动器接口280,用于向环境082中的致动器092提供控制数据226。这种控制数据226可以由处理器子系统240生成,以基于所确定的分布外值和/或(一个或多个)类别概率值来控制致动器。致动器可以是系统200的一部分,例如,系统200本身可以是被控制的系统。例如,致动器可以是电动、液压、气动、热、磁和/或机械致动器。具体但非限制性的示例包括电动机、电活性聚合物、液压缸、压电致动器、气动致动器、伺服机构、螺线管、步进电机等。这种类型的控制参考图3针对(半)自主车辆来描述。

[0061] 在其他实施例(图2中未示出)中,系统200可以包括呈现设备(诸如显示器、光源、扬声器、振动马达等)的输出接口,该输出接口可以用于生成感官可感知的输出信号,该输出信号可以基于分布外值和/或(一个或多个)类别概率值来生成。感官可感知输出信号可以直接指示这些输出,但是也可以表示所导出的感官可感知输出信号,例如用于物理系统的引导、导航或其他类型的控制。

[0062] 一般而言,本说明书中描述的每个系统(包括但不限于图1的系统100和图2的系统200)可以体现为单个设备或装置(诸如工作站或服务器)或者体现在单个设备或装置中。该设备可以是嵌入式设备。该设备或装置可以包括一个或多个执行适当软件的微处理器。例如,相应系统的处理器子系统可以由单个中央处理单元(CPU)来体现,但是也可以由这种CPU和/或其他类型的处理单元的组合或系统来体现。软件可能已经被下载和/或存储在对应的存储器中,例如诸如RAM之类的易失性存储器、或诸如闪存之类的非易失性存储器。替代地,相应系统的处理器子系统可以以可编程逻辑的形式在该设备或装置中实现,例如作为现场可编程门阵列(FPGA)。一般而言,相应系统的每个功能单元可以以电路的形式实现。

相应系统也可以以分布式方式实现,例如,涉及不同的设备或装置,诸如分布式本地或基于云的服务器。在一些实施例中,系统200可以是车辆、机器人或类似物理实体的一部分,和/或可以表示被配置成控制物理实体的控制系统。

[0063] 图3示出了上述各项的示例,其中用于对传感器数据进行分类的系统(例如图2的系统200)被示出为在环境083中操作的(半)自主车辆62的控制系统300的一部分。车辆62可以是自主的,因为它可以包括自动驾驶系统或驾驶辅助系统,其中后者也被称为半自主系统。

[0064] 自主车辆62可以例如并入分类系统作为车道保持支持系统300的一部分。分类系统可以接收车辆62和/或车辆环境083中的交通参与者的位置信息的时间序列作为输入,其例如表示公路场景并且表征车辆随时间(例如最近五秒内)的行为。例如,如图中所示,位置信息可以从由安装在车辆上的相机22取得的车辆62的环境图像中提取。

[0065] 分类系统可以将该时间序列传感器数据分类到表示车辆驾驶员的相应驾驶员任务(例如保持车道或改变车道)的相应类别中。为了分析这种时间序列数据,推理模型可以包括递归模型,该递归模型基于相应时间点的输入数据输出分类,这本身是已知的。同样,用于进行OOD检测的生成模型可以是递归模型,该递归模型可以被配置成基于时间序列的在前时间点处的概率分布的参数来确定时间序列的一时间点处的一个或多个物理量的值的概率分布的参数。可以使用现有类型的递归模型。

[0066] 车道保持支持系统300可以使用从浓度参数导出的经校准的、因此更可靠的类别概率值,来对预测进行一致性评估。基于一致性评估,该系统可以决定是使用车道保持支持系统的常规控制模块还是切换到回退控制系统来控制例如车辆的车轮42。例如,如果预测器报告中等水平的不确定性,则可以触发安全回退功能性,诸如车道保持支持算法,该算法不太舒适,但是高度可解释的,并且因此使用起来更安全。例如,如果预测器报告高的不确定性,和/或如果分布外值指示传感器数据是分布外的,则车道保持支持系统300可以将控制交还给人类驾驶员。

[0067] 虽然在该图中没有具体示出,但是分类系统还可以被配置成:如果分布外值指示非对应性和/或类别概率值指示不充分的置信度,则存储传感器数据以供未来使用。例如,该系统可以存储传感器数据本身,或者将其传输用于外部存储。在其他情况下,传感器数据可以在使用之后被丢弃。以这种方式,自主车辆62可以收集用于训练机器学习模型的相关训练数据,以用于控制(半)自主车辆。通过仅存储相关传感器数据,可以更高效地使用车辆62处的有限存储容量和/或用于传输到外部存储装置的有限传输容量。

[0068] 作为切换到回退控制的另一个示例,还有可能使用如本文中描述的分类模型以用于车辆62中的异常检测。在该示例中,分类模型可以是语义分割模型,该语义分割模型被配置成将由相机22取得的图像的相应图像部分分类到相应类别中。基于语义分割,控制系统300可以计算所有行人的深度信息,计算这些行人周围的轨迹,并且控制车辆62足够紧密地跟随该轨迹,使得车辆不会撞到行人。同样在这种情况下,如果针对相应图像部分的分布外值指示不充分的对应性和/或针对相应图像部分的类别概率值指示不充分的置信度,则语义分割可以被认为不足够可靠,并且可以使用安全回退控制系统。尽管(半)自主车辆62被用作示例,但是将理解的是,用于避开其环境中的人类的相同技术可以应用到任何移动机器人中。

[0069] 图4a示出了不包括生成模型的分类型模型的详细但非限制性的示例。该图示出了板块图,其中实线圆圈指示可观察的随机变量,并且虚线圆圈指示从可观察的随机变量中推理出的隐随机变量。

[0070] 该图中示出的是证据性分类器,也称为先验网络,例如,根据“Evidential Deep Learning to Quantify Classification Uncertainty”的设计。该图示出了传感器数据 x SD 410,该数据用于导出类别概率 π CP 420,从该类别概率,确定输出变量 y CL。这种模型可以被描述为如下的层级式概率模型:

$$[0071] \quad \pi|x \sim \text{Dir}(\pi|g_{\psi}(x)),$$

$$[0072] \quad y|\pi \sim \text{Cat}(y|\pi),$$

[0073] 其中 $\text{Dir}(\cdot)$ 是狄利克雷分布, $g_{\psi}(x)$ 是将给定输入 x SD(例如,观察到的图像)映射到狄利克雷浓度参数(也称为强度参数)的函数,并且 $\text{Cat}(\cdot)$ 是在给定来自先验的样本 π CP的情况下在输出变量 y CL上指派概率质量的分类分布。

[0074] 这种证据性模型可以通过最小化关于 ψ (例如如果 $g_{\psi}(\cdot)$ 被选择作为神经网络,则为权重)的下述损失来训练:

$$[0075] \quad L_1(\theta) = E_{\pi \sim \text{Dir}(\pi|g_{\psi}(x))}[\|y - \pi\|_2^2] + \beta \cdot \text{KL}(\text{Dir}(\pi|g_{\psi}(x))||\text{Dir}(\pi|1, \dots, 1))$$

[0076] 其中 $\text{KL}(\cdot || \cdot)$ 是两个分布之间的Kullback-Leibler(KL)散度,并且 $\beta > 0$ 是退火因子,该退火因子调整KL项对该损失的正则化有多强。

[0077] 也不包括一般模型的类似分类模型可以通过根据以下损失函数在域内样本集合和OOD样本集合上进行训练来获得:

$$[0078] \quad L_2(\theta, \gamma) = \text{KL}(\text{Dir}(\pi|g_{\psi}(x))||\text{Dir}(\gamma \cdot y)) + \beta \cdot \text{KL}(\text{Dir}(\pi|g_{\psi}(x_{\text{ood}}))||\text{Dir}(\pi|1, \dots, 1))$$

[0079] 其中 x_{ood} 是来自真实OOD数据集的样本,例如如果任务是手写数字识别,则是鸟的图像,并且 $\gamma > 0$ 是针对域内样本 x 的独热编码真实类别标签 y 上的放大因子。

[0080] 可以观察到,这两个损失函数 L_1 、 L_2 是以下更一般模式的特殊情况:

$$[0081] \quad L(\theta) = \text{Div}(\text{Dir}(\pi|g_{\psi}(x)), p_{\text{true}}(y)) + \beta \cdot \text{KL}(\text{Dir}(\pi|g_{\psi}(x))||\text{Dir}(\pi|1, \dots, 1))$$

[0082] 其中 $\text{Div}(\cdot, \cdot)$ 是两个分布之间的散度量。对于 L_2 ,它是KL散度。对于 L_1 , $p_{\text{true}}(y)$ 被选择作为少上的点质量,因此 $\text{Div}(\cdot, \cdot)$ 则是得分函数,具体地说,它是Brier得分,而对于 L_2 ,它是另一个狄利克雷分布,其强度是从真实标签 y 生成的。对于 L_1 , $g_{\psi}(x)$ 在正则化KL项中被馈送有域内数据,而 L_2 向其馈送真实OOD数据以用于正则化。

[0083] 可以观察到,根据上述模式的损失函数(例如不包括生成模型)一般而言不足以实现如下两者:(i)关于域内样本的预测不确定性的经校准的量化、以及(ii)域外OOD样本的检测,而不需要在训练时间期间看到示例OOD观察。这是因为遵循上述结构的损失函数缺乏评估给定输入样本 x 是否在目标域内的机制,而没有从真实OOD样本获得显式监督。实际上,上述损失结构鼓励该模型以非常高的确定性针对域内样本提供模型输出。因此,发明人发现,作为后果,准确的类别概率值可能是无法从这种模型中导出的。

[0084] 图4b示出了包括生成模型的分类型模型的详细但非限制性的示例。如图4a中那样,该模型被示出为板块图。

[0085] 该图中的分类模型包括推理模型IM,推理模型IM被配置成基于输入传感器数据 x SD 410来确定狄利克雷分布的相应浓度参数 $h_{\psi}(x)$ 。狄利克雷分布是针对传感器数据SD要

被分类到的相应类别的类别概率 π CP 420的概率分布。基于类别概率CP 420,传感器数据SD的分类CL 430可以根据分类分布CAT 460来定义。可能类别的集合是有限的。例如,类别的数量可以是2个,至多或至少5个,或者至多或至少10个。

[0086] 分类模型进一步包括生成模型GM 450,生成模型GM 450被配置成:基于类别概率CP,根据分类模型在其上被训练的训练数据集来确定传感器数据SD的概率分布的参数 $f_\theta(\pi)$ 。基于类别概率CP,或者基于生成它们的浓度参数,可以通过确定传感器数据SD根据生成模型GM被生成的概率来确定指示传感器数据SD与训练数据集的对应性的分布外值。

[0087] 因此,使用生成模型GM的密度估计可以与使用推理模型IM的证据性分类器训练同时执行,以学习评估给定样本SD的域相关性,而不需要显式监督。在数学上,该模型可以由以下生成设计来表示:

$$[0088] \quad \pi \sim \text{Dir}(\pi | 1, \dots, 1)$$

$$[0089] \quad x | \pi \sim p_\theta(x | f_\theta(\pi))$$

$$[0090] \quad y | \pi \sim \text{Cat}(y | \pi)$$

[0091] 其中 $p_\theta(x | f_\theta(\pi))$ 是输入域 x 上的似然函数,换句话说,是传感器数据的概率分布。例如,概率分布可以是正态分布、分类分布、伯努利分布等,例如:

$$[0092] \quad x | \pi \sim N(x | f_\theta^1(\pi), f_\theta^2(\pi)).$$

[0093] 如图中所示,可以训练推理模型IM,使得根据推理模型IM给出的浓度参数从狄利克雷分布采样的类别概率CP提供针对训练数据集的类别概率的变分推理近似。在该图中,推理模型IM的虚线表示变量对近似后验分布 $q_\psi(\pi_n | x_n)$ 的依赖性。

[0094] 特别地,给定包括 N 个输入-输出对的训练数据集 $D = \{(x_1, y_1), \dots, (x_N, y_N)\}$,例如图像 x_n 或其他类型的传感器数据及其对应的标签 y_n ,推理模型可以被训练成提供对隐证据变量上的难以处理的后验分布 $p(\pi_1, \dots, \pi_N | D)$ 的准确近似。该目标的训练损失可以使用变分推理来导出,例如通过最小化

$$[0095] \quad \text{KL}(q_\psi(\pi_1, \dots, \pi_N | x) || p(\pi_1, \dots, \pi_N | D))$$

[0096] 其中

$$[0097] \quad q_\psi(\pi_1, \dots, \pi_N | D) = \prod_{n=1}^N \underbrace{\text{Dir}(\pi_n | g_\psi(x_n) + 1)}_{q_\psi(\pi_n | x_n)}$$

[0098] 其中推理模型 $g_\psi(x_n)$ 可以通过适当的 K 维可训练函数来实现。例如,推理模型可以被定义为提供正值。如该示例中所示,由推理模型提供的浓度参数 $g_\psi(x_n) + 1$ 可以被限制为大于或等于一,这改进了训练期间的数值稳定性。到至少一的限制不是必需的,但是输出通常至少被限制为非负的。作为示例, $g_\psi(\cdot)$ 可以是具有权重 ψ 的神经网络,例如在其输出层处具有软加算子。

[0099] 有趣的是,该模型可以被视为证据性学习和变分自编码器(VAE)之间的混合形式。针对该原因,该模型可以被称为证据性变分自编码器。从证据性学习的角度和从VAE的角度两者来看,该模型都提供了惊人的优点。从证据性学习模型的角度来看,引入VAE特性惊人地实现了域内校准。从将模型用作VAE的角度来看,在如VAE的这种正常使用中,使用狄利克雷分布将不会被预期是有益的,例如,在该设置中,标准正态分布是更典型的选择。而且,用作VAE的模型通常具有相对大的隐空间维度,例如,至少50个或至少100个。在VAE设置中,减

少该维度将不会被预期是有益的。取而代之,在实施例中,隐空间维度对应于分类器的类别的数量,并且因此例如小于50个,甚至例如至多5个或至多10个。惊人的是,发明人发现具有相对小维度的狄利克雷分布的这种设计使得能够与证据性学习组合,并且帮助显著改进预测器的域内校准性能。

[0100] 在上述示例中,应用了摊销(amortization),例如,针对推理模型的近似后验分布 $q_{\psi}(\pi_n | x_n)$ 被定义为依赖于观察D。该模型也可以以未摊销的方式来定义,例如,使用后验分布 $q_{\psi}(\pi_n)$ 。

[0101] 图4c示出了包括生成模型并且使用上下文变量的分类模型的详细但非限制性的示例。该示例是图4b的示例的扩展。特别地,该图示出了传感器数据SD 410;推理模型IM' 441;生成模型GM 450;类别概率CP 420;分类分布CAT 460;和分类CL 430,类似于图4b中的那些。

[0102] 在该示例中,推理模型IM' 被附加地给予上下文变量z CV 470的值作为输入。在使用分类模型期间,从上下文实例的集合确定上下文变量CV的值,其中上下文实例包括上下文传感器数据 x_j CSD 411,并且可选地还包括对应的上下文目标类别 y_j CCL 431。

[0103] 特别地,如图中所示,可以通过如下方式来确定上下文变量CV的值:将经训练的上下文模型(例如, $h_{\phi}(x_j, y_j)$)应用到相应的上下文实例CSD、CCL,以获得相应的上下文参数贡献;将相应的上下文参数贡献聚集成上下文参数集合,例如 $[r_a, r_b] = 1/C \sum_{j=1}^C h_{\phi}(x_j, y_j)$,或者使用另一个聚集函数;以及通过对根据上下文参数集合的上下文概率分布(例如 $z | x_c, y_c \sim G(z_k | r_a, r_b)$)进行采样来确定上下文变量的值。

[0104] 例如,上下文变量可以是实值或离散的,例如,从2个、至多或至少5个、或至多或至少10个可能的值中选择。作为将上下文集合 $(x_1, y_1), \dots, (x_k, y_k)$ 映射到上下文变量CV的上下文模型,例如,可以使用推理模型IM', 忽略输出 y_k ,或者可以使用上下文编码器网络 $m(x_k, y_k)$,该网络映射输入-输出对并且与推理和生成模型联合地被训练。

[0105] 例如,可以使用以下数学模型来分别定义上下文变量CV、类别概率CP、分类CL和传感器数据SD(对于传感器数据SD示出了正态分布,但是也有可能取而代之使用另一种概率分布):

$$[0106] \quad [r_a, r_b] = \frac{1}{C} \sum_{j=1}^C h_{\phi}(x_j, y_j) \quad z | X_C, Y_C \sim \prod_{k=1}^K G(z_k | r_a, r_b) \quad \Pi \sim \prod_n^N \text{Dir}(\pi_n | z)$$

$$[0107] \quad Y_P | \Pi \sim \prod_n^N \text{Cat}(y_n | \pi_n) \quad X_P | \Pi \sim \prod_n^N N(x_n | f_{\theta}(\pi_n), g_{\psi}(\pi_n))$$

[0108] 在该示例中,给定传感器数据SD和上下文变量CV,类别概率CP的近似后验分布可以被定义为:

$$[0109] \quad q(\pi | x, z) = \text{Dir}(\pi | u_{\lambda}(z, x))$$

[0110] 基于该模型的类别CL的预测可以被定义为:

$$[0111] \quad p(y_* | x_*, D) = \int \int p(y_* | \pi) q(\pi | x_*, z) p(z | X_C, Y_C) d\pi dz.$$

[0112] 一般地,分类模型可以包括一个或多个上下文变量CV。例如,这种上下文变量的数量可以是至多或至少5个、或者至多或至少10个。上下文变量CV的使用可以允许分类模型有效地确定新示出的样本是否与其之前已经见过的相似,例如,针对相似的样本提供比不熟悉的样本更低的不确定性。然而,如果没有上下文信息CSD、CCL可用,还有可能应用使用上下文变量的模型,例如通过使用狄利克雷分布上的先验分布来代替它。

[0113] 图5a示出了如何使用分类器模型对传感器数据进行分类的详细但非限制性的示例,该分类器模型包括生成模型,例如关于图4b描述的模型或关于图4c描述的模型。

[0114] 该图中示出的是模型输入 x^* SD 510,例如,表示要被控制或监测的计算机控制系统和/或其环境的一个或多个物理量的测量。例如,传感器数据SD可以是图像、一个或多个物理量的测量的时间序列等。

[0115] 给定传感器数据SD,推理模型 g_{ψ} IM 541可以用于获得针对传感器数据要被分类到的相应多个类别的类别概率的狄利克雷分布的浓度参数 CP_i 542。推理模型可以由具有自由参数 ψ 的函数 $g_{\psi}(\cdot)$ 来定义,该函数 $g_{\psi}(\cdot)$ 从输入域映射到K维正实数, $\alpha = g_{\psi}(x)$,其中K是类别的数量。该函数可在计算机上实现,例如具有权重 θ 的神经网络。有趣的是,所确定的浓度参数可以用于确定分布外值、以及确定类别概率值两者。

[0116] 指示传感器数据SD属于来自多个类别的集合的类别的概率的类别概率值 CP_i 530可以通过借助于来自类别概率的分类分布来定义类别概率值 CP_i 而从浓度参数确定的Prob 560,该类别概率是根据由浓度参数 CP_i 定义的狄利克雷分布生成的。

[0117] 具体地,对于给定测试输入 x^* ,其标签 y^* 上的预测分布可以以封闭形式被计算为:

$$\begin{aligned} Pr(y^* = k|x^*) &= \int p(y^* = k|\pi) q_{\hat{\pi}}(\pi|x = x^*) d\pi \\ &= \int Cat(y^*|\pi) Dir(\pi|h_{\hat{\pi}}(x = x^*)) d\pi \end{aligned}$$

[0118]

$$= \frac{\Gamma(1 + g_{\hat{\pi}}^k(x^*))}{\Gamma(g_{\hat{\pi}}^k(x^*))} \frac{\Gamma(\sum_{j=1}^K g_{\hat{\pi}}^j(x^*))}{\Gamma(1 + \sum_{k=1}^K g_{\hat{\pi}}^j(x^*))}$$

[0119] 其中 $\Gamma(\cdot)$ 是伽玛函数, $g_{\hat{\pi}}^j(x)$ 是预测器 $g_{\psi}(\cdot)$ 的第j个输出通道,K是分类问题中的类别数量,并且k是针对 y^* 的所查询类别。

[0120] 当使用从上下文实例集合 X_c, Y_c (其中目标类别 Y_c 是可选的)确定的上下文变量 z 时,类别概率值可以被类似地计算为:

$$p(y_*|x_*, D) = \iint p(y_*|\pi) q(\pi|x_*, z) p(z|X_c, Y_c) d\pi dz.$$

[0122] 通过选择具有最高概率质量的类别,可以将输入分类到特定类别中:

$$[0123] \quad k_{pred}^* = \operatorname{argmax}_{k \in \{1, \dots, K\}} Pr(y^* = k|x = x^*),$$

[0124] 其中 k_{pred}^* 是针对输入 x^* 的所预测类别。对于下游使用,诸如拒绝采取动作或采取安全默认动作,可以使用所预测类别和/或其他类别的类别概率 $Pr(y^* = k_{pred}^*|x^*)$,例如通过对其应用阈值。有趣的是,要注意,可以高效地执行分类,并且特别地,当执行分类时,不需要使用生成模型。仍然,通过将生成模型与推理模型一起训练,可以从推理模型中获得更好校准的类别概率值。

[0125] 有趣的是,浓度参数 CP_i 也可以用于OOD检测,换句话说,离群值检测。可以确定分布外值OODV 551,其指示传感器数据SD与分类模型在其上被训练的训练数据集的对应性。分布外值可以被确定为传感器数据SD根据生成模型 f_{θ} 基于浓度参数 CP_i 被生成的概率,换句话说,通过计算传感器数据SD的输入重建似然。在数学上,这可以表示如下:

$$[0126] \quad p_{\theta}(x = x^*) = \int p_{\theta}(x = x^* | f_{\theta}(\pi)) q_{\psi}(\pi | x = x^*) d\pi.$$

[0127] 同样,该数学表达式可以根据需要被适配成考虑上下文变量。

[0128] 具体地,如图中所图示,可以通过对来自狄利克雷分布(例如 $q_{\psi}(\pi | x^*)$)的多个类别的类别概率 $\pi_{CP\ 520}$ 进行采样,并且然后关于重建概率 $p_{\theta}(x = x^* | f_{\theta}(\pi))$ 对它们进行评估SamG 550,从而以蒙特卡罗方式来确定分布外值OODV。

[0129] 这可以涉及应用生成模型 $f_{\theta}(\pi)$ 来确定传感器数据的概率分布 p_{θ} 的参数。生成模型可以由具有自由参数 θ 的函数 $f_{\theta}(\cdot)$ 来定义,该函数从K维单形域映射到输入域,例如 $x = f_{\theta}(\pi)$ 。该函数可在计算机上实现,例如具有权重 θ 的神经网络或其他已知技术。

[0130] 基于所确定的参数,可以确定传感器数据SD根据那些概率被生成的概率。该计算可以基于从 $p_{\theta}(x | \lambda)$ 的采样,例如使用无偏采样器。概率分布可以是正态分布,该正态分布例如具有由生成模型输出的相应平均值和可选地还有相应标准差。

[0131] 分布外值OODV可以是所确定的概率,或者可以通过阈值化从该概率获得,例如,如果基于手边应用的安全性要求来选择针对阈值 $\epsilon > 0$ 的 $u_{ood} = p_{\theta}(x = x^*) < \epsilon$,则样本 x^* 可以被表征为OOD或离群值。要注意的是,概率 $p_{\theta}(x = x^*)$ 通常不是像类别概率 CP_i 那样的经校准的概率,例如,例如,可能不准确地指示传感器数据是否为OOD的概率。仍然,概率 $p_{\theta}(x = x^*)$ 可以提供OOD性的相对度量,并且因此是有用的。

[0132] 取决于应用,可以输出针对一个或多个类别的分布外值OODV和/或类别概率 CP_i 。例如,可以仅在传感器数据SD未被确定为在分布外的情况下才确定类别概率,但是还有可能计算它们,而不管传感器数据是否为OOD。例如,可能的预测输出可以包括所预测的类别 k_{pred}^* ,并且可选地包括针对该类别的类别概率值;以及关于输入样本是否为OOD的二元判定 u_{ood} 。

[0133] 在各种实施例中,传感器数据SD可以包括一个或多个物理量的测量的时间序列。在这种情况下,推理模型IM和/或生成模型SamG可以包括递归模型,诸如递归模型,诸如递归神经网络。例如,递归神经网络可以是基于门控递归单元(GRU)或基于长短期记忆(LSTM)的神经网络。基于狄利克雷浓度参数 CP_i 的配置,各种动力学模态是可能的。例如,由推理模型IM确定的浓度参数 CP_i 可以针对多个时间点来生成每个时间点的离散类别标签,使得分类模型预报包括连续和/或离散变量的时间序列,例如,以执行序列到序列的分类。狄利克雷变量也可以被约束到单个预报中的单个值,在这种情况下,分类模型可以执行总体时间序列的分类。

[0134] 同样,生成模型SamG可以使用递归模型来确定概率分布的参数,以用于在相应时间点处生成传感器数据。例如,该模型可以基于先前时间点的参数和给定时间点的(一个或多个)浓度参数来确定给定时间点的(一个或多个)参数。有趣的是,这不仅允许确定总体时间序列的分布外值OODV,而且还允许确定针对相应时间点的相应分布外值,例如,以在使用相应时间点的分类时考虑在内。

[0135] 在各种实施例中,传感器数据SD可以包括图像数据。同样在这种情况下,推理模型

IM可以执行总体图像的分类,而且还可以执行相应图像区域的分类,例如,分类模型可以是语义分割模型。在后一种情况下,生成模型可以是图像到图像变换模型,该变换模型确定针对相应图像部分的相应概率分布的参数。类似于时间序列的情况,这允许不仅确定总体分布外值,而且还确定针对相应图像部分的相应分布外值。例如,生成模型可以确定针对相应图像部分的相应平均值以及可选地标准差。合适的图像到图像变换模型本身在本领域中是已知的,并且可以容易地应用。

[0136] 图5b示出了如何训练分类器模型的详细但非限制性的示例,该分类器模型包括生成模型,例如关于图4b描述的模型或关于图4c描述的模型。参考图5a描述的这种模型的各种选项在这里也适用。

[0137] 训练可以基于包括多个训练实例的训练数据集。训练实例被标记。出于说明性目的,示出了包括传感器数据 x_n SD 511和对应的目标类 y_n TC 531的训练实例。训练数据集可以被标示为 $D = \{(x_1, y_1), \dots, (x_N, y_N)\}$, 其中 x_n 来自输入域,并且 $y_n \in \{1, \dots, K\}$ 针对类别计数K。

[0138] 作为训练的一部分,推理模型 g_ψ IM 541可以被应用到传感器数据SD,以获得针对相应多个类别的类别概率 π CP 520的狄利克雷分布Dir 543的浓度参数 CP_i 542。基于浓度参数 CP_i ,可以导出针对推理模型IM的推理的损失以及用于使用这些浓度参数来重建传感器数据的损失两者。有趣的是,发明人发现,因为浓度参数IM用于重建以及推理,所以改进了推理的类别概率的校准。

[0139] 具体地,如图中所示,训练信号可以包括损失项LOSS2 582——所确定的Prob 560,其基于传感器数据SD基于浓度参数被分类到目标类别TC中的概率。该损失项可以用于学习准确的标签预测器。例如,该损失项可以实现为 $E_{q_\psi(\pi_n|x_n)}[\log \text{Cat}(y_n|\pi_n)]$ 。该损失项在这里被示出为基于所采样的类别概率CP,但是类似于图5a,也可以在没有采样的情况下从浓度参数 CP_i 直接计算。

[0140] 示出了另外的损失项LOSS1 581——即,所确定的SamG 550,其基于传感器数据SD根据生成模型 f_θ 基于浓度参数CP被生成的概率。该项可以通过表达式 $E_{q_\psi(\pi_n|x_n)}[\log p_\theta(x_n|f_\theta(\pi_n))]$ 来实现,并且可以将 x_n 的隐表示——其是类别指派的离散概率分布——映射到可观察的输入域。该项可以学习从类别指派概率CP生成样本。因此,该项可以鼓励该模型提供符合数据生成过程的不确定性估计。有趣的是,这在没有将决策边界平坦化或使用外部样本的情况下是可能的。

[0141] 如图中所示,训练信号还可以可选地包括第三损失项LOSS3 583——所确定的DIV 580,以提供所确定的类别概率CP的正则化,例如,通过惩罚从表示最大不确定性的类别概率的发散。该项可以被实现为例如 $KL(q_\psi(\pi_n|x_n) || \text{Dir}(\pi|1, \dots, 1))$ 。使用该正则化项,从平坦狄利克雷分布所表征的“我不知道”状态的发散——该发散不用于解释数据——可能被惩罚。

[0142] 例如,上述三个损失可以被组合到以下证据下限(ELBO)损失函数中,以分别关于生成模型和推理模型的可训练参数 $\{\theta, \psi\}$ 被最大化:

$$[0143] \quad L_{EVAE}(\theta, \psi) = \sum_{n=1}^N \{ E_{q_{\psi}(\pi_n|x_n)} [\log p_{\theta}(x_n|f_{\theta}(\pi_n))] + E_{q_{\psi}(\pi_n|x_n)} [\log \text{Cat}(y_n|\pi_n)] \\ - KL(q_{\psi}(\pi_n|x_n) \| \text{Dir}(\pi|1, \dots, 1)) \}$$

[0144] 第一项可以学习拟合域内数据;第二项可以学习域内分布,例如不确定性;并且第三项可以提供正则化。

[0145] 上述损失函数可以被适配成使用上下文变量,例如,如参考图4c和图5a所讨论的,如下:

$$[0146] \quad \underset{\theta, \phi, \psi, \lambda}{\operatorname{argmax}} \sum_n^N E_{p(z|x_c, y_c)} \{ E_{q(\pi_n|z, x_n)} [\log \text{Cat}(y_n|\pi_n)] + E_{q(\pi_n|z, x_n)} [\log N(x_n|f_{\theta}(\pi_n), g_{\phi}(\pi_n))] \\ - KL(q(\pi_n|z, x_n) \| p(\pi_n)) + KL(\text{Dir}(\pi|u_{\lambda}(z, f_{\theta}(1, \dots, 1))) \| \text{Dir}(\pi|1, \dots, 1)) \}$$

[0147] 例如,在训练期间,小批量内随机选择的训练观察子集可以用作上下文实例集合。

[0148] 基于训练信号,例如损失LOSS1、LOSS2以及可选地还有LOSS3的组合,推理模型的参数PAR_ψ501和/或生成模型的参数PAR_θ502可以在训练操作Train 590中被更新。例如,可以以迭代的方式执行训练,其中在每次迭代中,更新参数集合中的一个或两者。用于确定上下文变量的可训练参数(如果使用的话)也可以被更新。

[0149] 训练Train可以利用无约束优化器来执行,该无约束优化器诸如Adam,在Kingma和Ba的“Adam:A Method for Stochastic Optimization”(可在<https://arxiv.org/abs/1412.6980>获得并且通过引入并入本文)中被公开;随机梯度下降;或者RMSProp。如所已知的,这种优化方法可以是启发式的和/或达到局部最优。训练可以在逐实例的基础上执行,或者批量地(例如至多或至少64个或至多或至少256个实例)执行。训练信号可以使用可自动微分的库来实现,该库例如PyTorch或TensorFlow。使用训练Train,可以确定针对生成模型 $f_{\theta}(\cdot)$ 和推理模型 $g_{\psi}(\cdot)$ 的自由参数的学习值 $(\hat{\psi}, \hat{\theta})$ 。

[0150] 在许多情况下,损失项LOSS1和LOSS3可以通过计算它们相应的解析解来实现。在一些情况下,针对损失项LOSS2的解析解也可能是可用的。一种替代方案是使用可微分的蒙特卡罗采样方法,如用于对类别概率进行采样的采样操作SamD 543以及用于对传感器数据的概率分布的参数进行采样的采样操作SamG所图示。例如,可以使用已知深度学习库中可用的“r样本”(“r”意味着可重新参数化)方法。

[0151] 图6示出了训练分类模型的计算机实现方法600的框图。该模型可以用于控制和/或监测计算机控制系统。分类模型可以被配置成将传感器数据分类到来自多个类别的集合的类别中。方法600可以对应于图1的系统100的操作。然而,这不是限制,因为方法600也可以使用另一个系统、装置或设备来执行。

[0152] 方法600可以包括:在题为“获得训练数据”的操作中,获得610包括多个训练实例的训练数据集。

[0153] 方法600可以包括:在题为“访问模型数据”的操作中,访问620表示分类模型的模型数据。分类模型可以包括可训练的推理模型。推理模型可以被配置成基于传感器数据来确定针对相应多个类别的类别概率的狄利克雷分布的相应浓度参数。分类模型可以进一步

包括可训练的生成模型。生成模型可以被配置成基于类别概率来确定传感器数据的概率分布的参数。

[0154] 方法600可以包括:在题为“训练模型”的操作中,训练630分类模型。分类的训练可以包括:在题为“选择实例”的操作中,从训练数据集中选择640训练实例。训练实例可以包括传感器数据和多个类别的集合中的对应目标类别。分类的训练可以包括:在题为“应用推理模型”的操作中,将推理模型应用650到传感器数据以获得浓度参数。分类的训练可以包括:在题为“导出训练信号”的操作中,导出660用于训练实例的训练信号。训练信号可以基于传感器数据基于浓度参数被分类到目标类别中的概率。训练信号可以进一步基于传感器数据根据生成模型基于浓度参数被生成的概率。方法600可以包括:在题为“更新模型”的操作中,基于训练信号来更新670推理模型和/或生成模型的参数。

[0155] 图7示出了对传感器数据进行分类以用于控制和/或监测计算机控制系统的计算机实现方法700的框图。方法700可以对应于图2的系统200或图3的系统300的操作。然而,这不是限制,因为方法700也可以使用另一个系统、装置或设备来执行。

[0156] 方法700可以包括:在题为“获得模型数据”的操作中,获得710表示分类模型的模型数据。例如,分类模型可能已经根据本文中描述的方法被预先训练。分类模型可以被配置成将传感器数据分类到来自多个类别的集合的类别中。分类模型可以包括经训练的推理模型。推理模型可以被配置成基于传感器数据来确定针对相应多个类别的类别概率的狄利克雷分布的相应浓度参数。分类模型可以进一步包括经训练的生成模型。生成模型可以被配置成基于类别概率来确定传感器数据的概率分布的参数。

[0157] 方法700可以包括:在题为“获得传感器数据”的操作中,获得720传感器数据。传感器数据可以表示计算机控制系统和/或其环境的一个或多个物理量的测量。

[0158] 方法700可以包括:在题为“应用推理模型”的操作中,将推理模型应用730到传感器数据以获得浓度参数。

[0159] 方法700可以包括:在题为“检测00D”的操作中,确定740指示传感器数据与训练数据集的对应性的分布外值。确定740可以通过确定传感器数据根据生成模型基于浓度参数被生成的概率来执行。

[0160] 方法700可以包括:在题为“确定类别概率”的操作中,从浓度参数来确定760类别概率值。类别概率值可以指示传感器数据属于来自多个类别的集合的类别的概率。方法700可以进一步包括:在题为“输出类别概率”的操作中,输出770类别概率值以用于控制和/或监测。可以有条件地执行确定760和输出770,例如,至少如果在标记为“充分的对应性?”的判定中分布外值指示750充分的对应性的话。

[0161] 将领会的是,一般而言,图6的方法600和图7的方法700的操作可以以任何合适的次序执行,例如连续地、同时地或其组合,在适用的情况下,受制于例如因输入/输出关系而使特定次序成为必要。也可以组合一些或所有方法,例如,应用经训练的模型的方法700可以随后被应用到根据方法600训练的该经训练的模型。

[0162] (一个或多个)方法可以在计算机上实现为计算机实现方法、专用硬件、或两者的组合。同样如图8中所图示,用于计算机的指令、例如可执行代码可以存储在计算机可读介质800上,例如,以机器可读物理标记的序列810的形式和/或作为具有不同电气(例如,磁性)或光学属性或值的元素的序列。可执行代码可以以暂时性或非暂时性的方式来存储。计

计算机可读介质的示例包括存储器设备、光学存储设备、集成电路、服务器、在线软件等。图8示出了光盘800。替代地,计算机可读介质800可以包括表示模型数据的暂时性或非暂时性数据810,该模型数据表示分类模型,其例如是根据本文中描述的方法训练的和/或根据如本文中描述的方法使用。分类模型可以包括可训练的推理模型,该推理模型被配置成基于传感器数据来确定针对相应多个类别的类别概率的狄利克雷分布的相应浓度参数。分类模型可以进一步包括可训练的生成模型,该生成模型被配置成基于类别概率,根据分类模型的训练数据集来确定传感器数据的概率分布的参数。

[0163] 示例、实施例或可选特征——无论是否被指示为非限制性的——都不应被理解为限制所要求保护的本发明。

[0164] 应当注意,上述实施例说明而非限制了本发明,并且本领域技术人员将能够在不脱离所附权利要求的范围的情况下设计许多替代实施例。在权利要求中,置于括号之间的任何附图标记都不应被解释为对权利要求进行限制。动词“包括”及其变形的使用不排除权利要求中所陈述的那些之外的元素或阶段的存在。元素前面的冠词“一(a或an)”不排除多个这种元素的存在。诸如“……中的至少一个”的表达当在元素列表或元素组之前时,表示从该列表或组中选择所有元素或元素的任何子集。例如,表述“A、B和C中的至少一个”应当被理解为包括仅A、仅B、仅C、A和B两者、A和C两者、B和C两者、或者A、B和C全部。本发明可以借助于包括若干个不同元素的硬件以及借助于适当编程的计算机来实现。在列举了若干个装置的设备权利要求中,这些装置中的若干个可以由同一个硬件项来体现。在相互不同的从属权利要求中记载了某些措施的仅有事实并不指示这些措施的组合不能够被有利地使用。

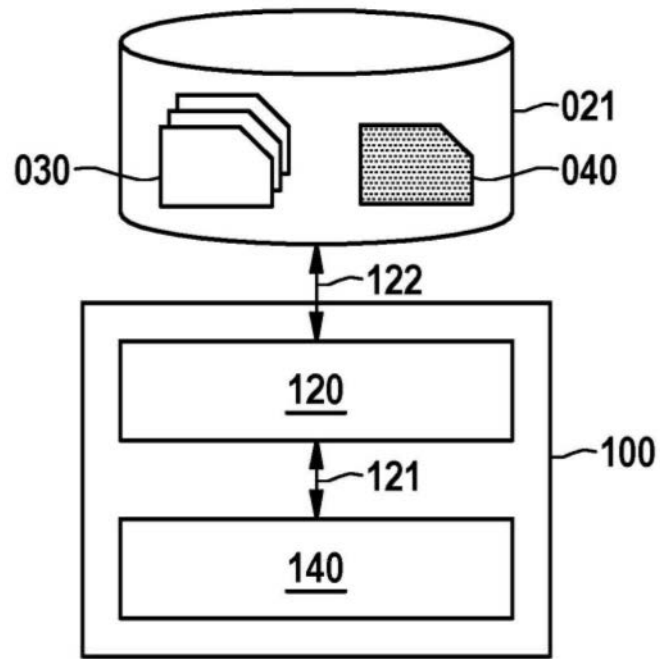


图1

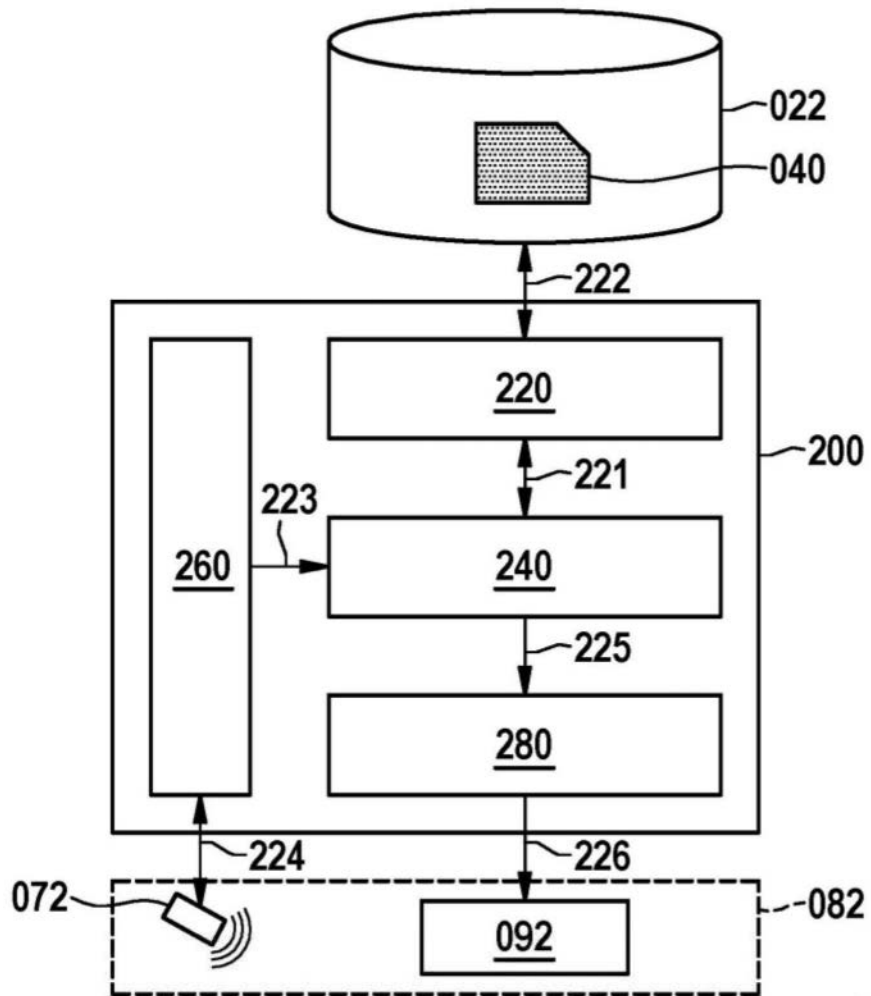


图2

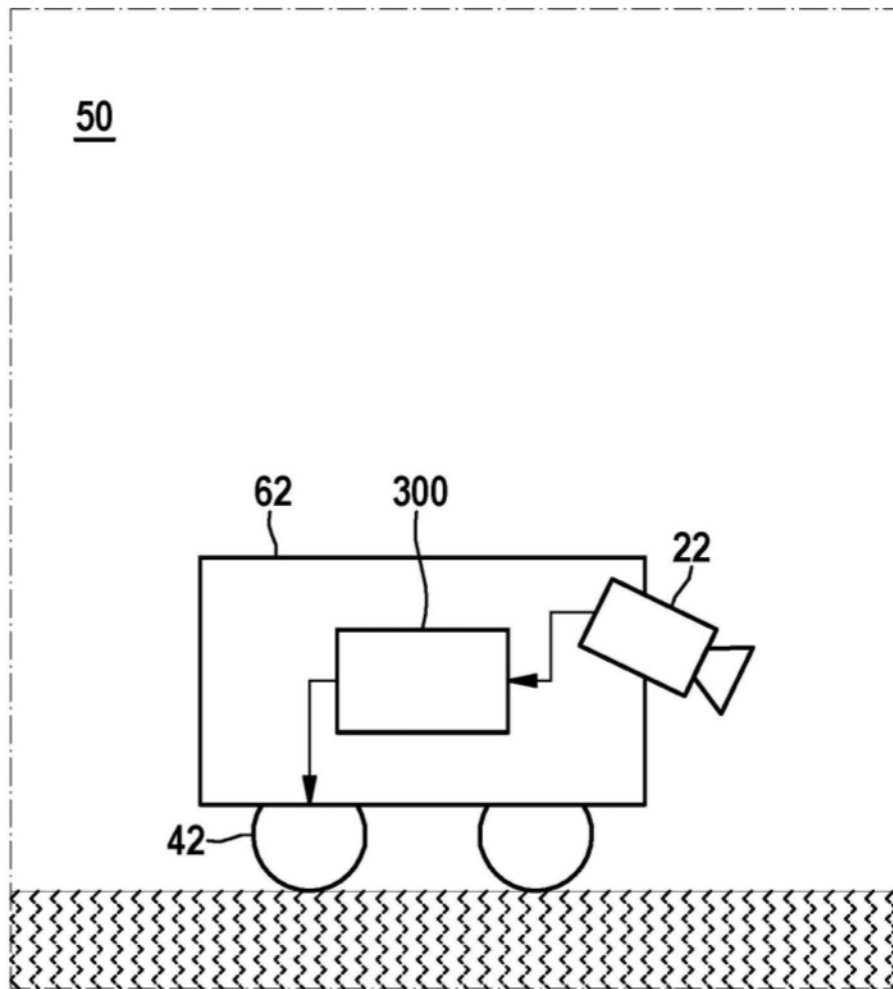


图3

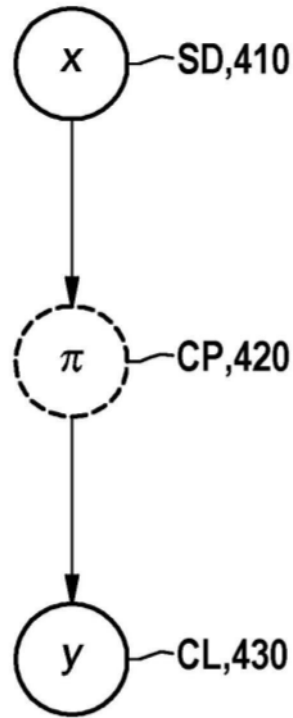


图4a

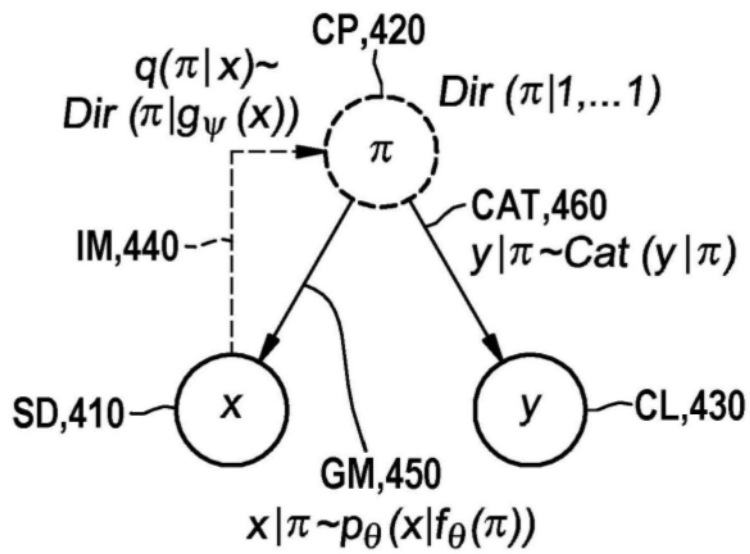


图4b

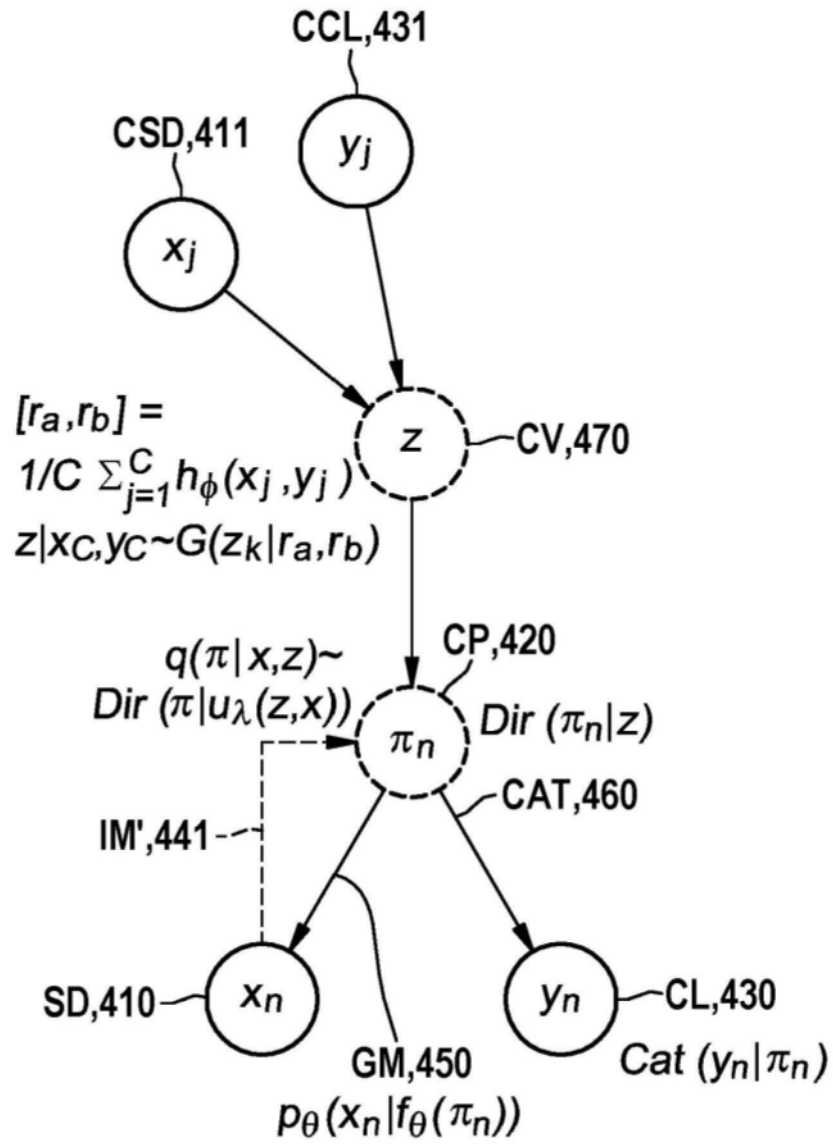


图4c

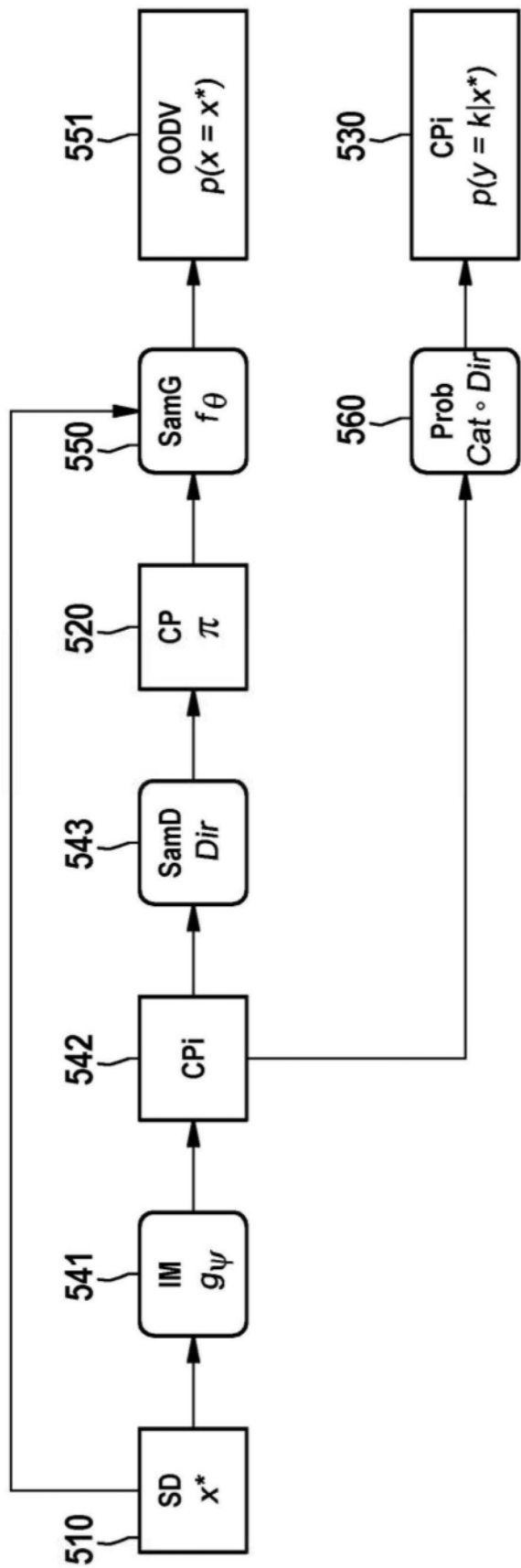


图5a

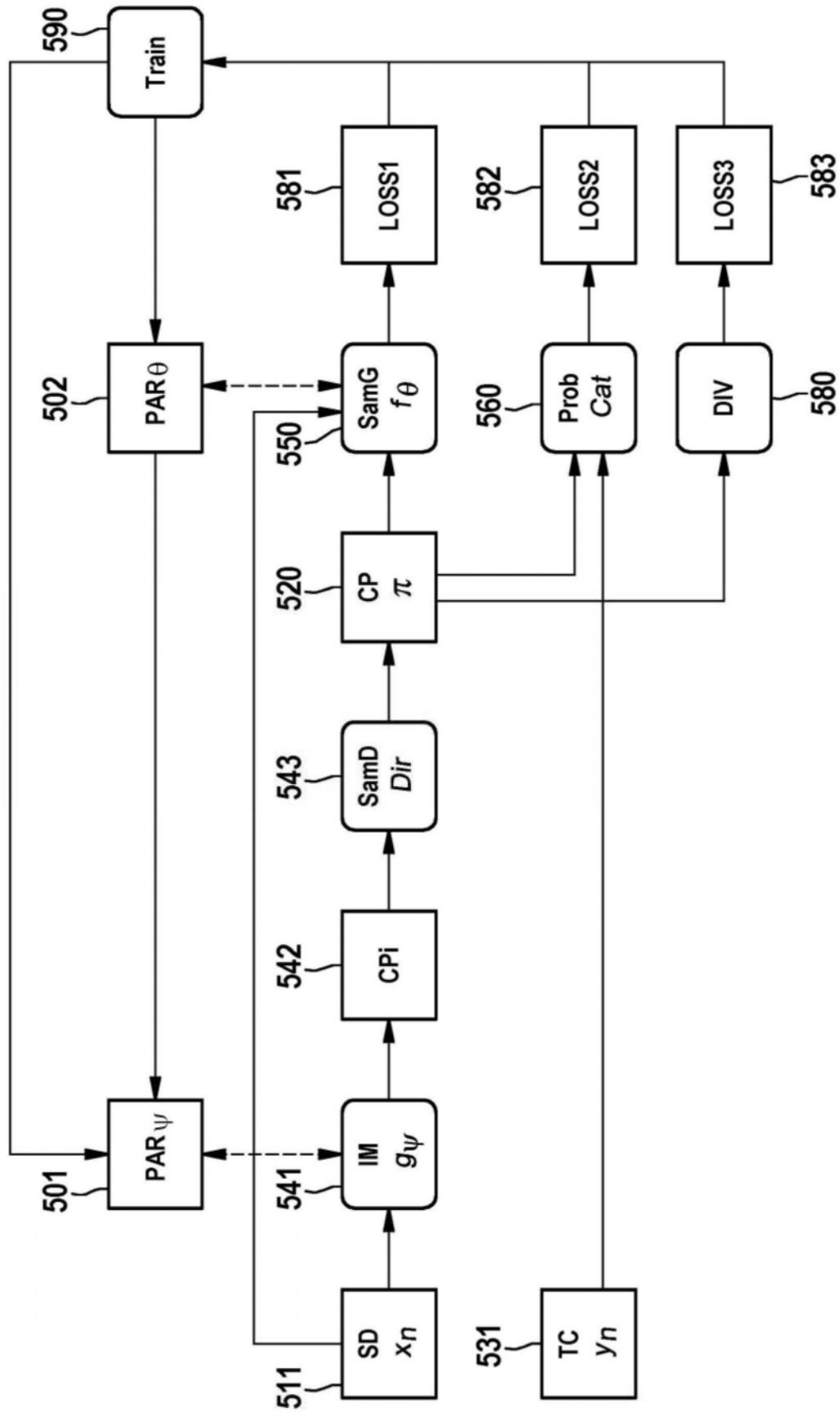


图5b

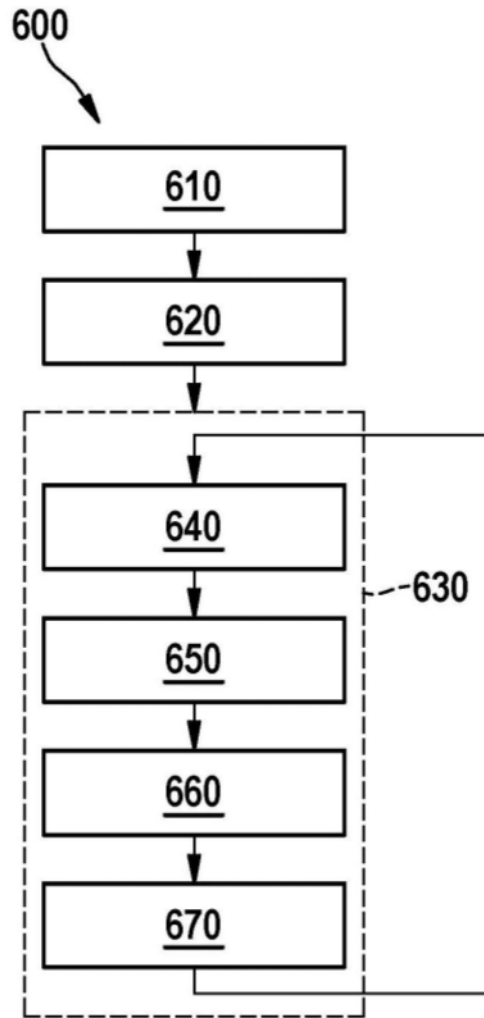


图6

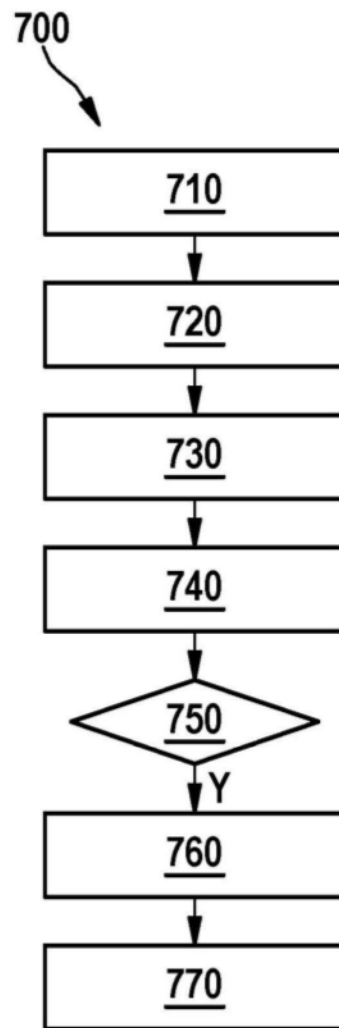


图7

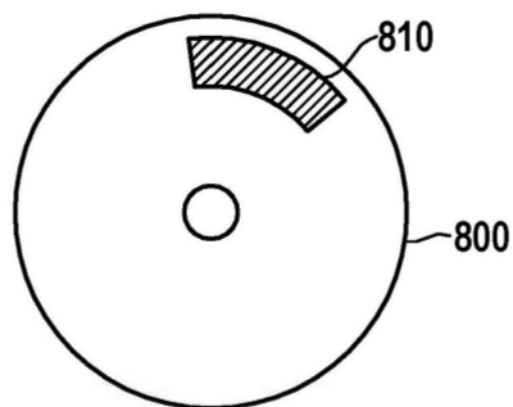


图8