



(12)发明专利申请

(10)申请公布号 CN 107249434 A

(43)申请公布日 2017.10.13

(21)申请号 201680009843.5

(22)申请日 2016.02.05

(30)优先权数据

62/115,330 2015.02.12 US

(85)PCT国际申请进入国家阶段日

2017.08.11

(86)PCT国际申请的申请数据

PCT/IB2016/050592 2016.02.05

(87)PCT国际申请的公布数据

W02016/128866 EN 2016.08.18

(71)申请人 皇家飞利浦有限公司

地址 荷兰艾恩德霍芬

(72)发明人 B·康罗伊 L·J·埃谢曼

C·波特斯 M·徐

(74)专利代理机构 永新专利商标代理有限公司
72002

代理人 李光颖 王英

(51)Int.Cl.

A61B 5/00(2006.01)

G06F 17/18(2006.01)

G06F 19/00(2011.01)

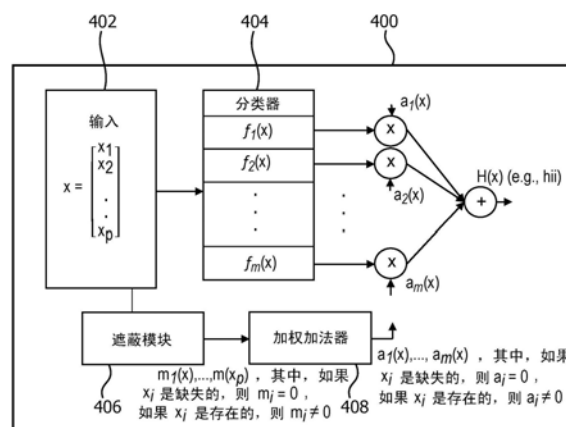
权利要求书2页 说明书12页 附图5页

(54)发明名称

鲁棒分类器

(57)摘要

在本文中所描述的各种实施例涉及用于鲁棒分类的方法和装置。许多真实世界的数据集受缺失和不完整的数据影响。通过基于哪个(些)特征是缺失的或不完整的将权重分配给数据集的特定特征,本发明的实施例能够提供针对缺失的数据的鲁棒性和恢复性。



1. 一种用于鲁棒分类的装置,所述装置包括:
接收器,其接收数据,接收到的数据具有来自特征集合的至少一个特征;
存储器,其被配置为存储计算机可读指令,所述计算机可读指令用于执行鲁棒分类;以及
处理器,其与所述存储器和所述接收器通信,其中,对用于执行鲁棒分类的所述计算机可读指令的运行令所述处理器实施:
至少一个低维分类器,其与所述接收器通信,其中,所述至少一个低维分类器中的每个与来自所述特征集合的相应的输入特征相关联,并且基于所述输入特征在所述接收到的数据中的存在来提供输出;以及
加权加法器,其与所述至少一个低维分类器中的每个以及所述接收器通信,其中,所述加权加法器提供所述至少一个低维分类器的所述输出的加权和。
2. 根据权利要求1所述的装置,其中,所述接收到的数据包括至少一个训练范例。
3. 根据权利要求2所述的装置,其中,使用接收到的至少一个训练范例来训练所述至少一个低维分类器中的每个。
4. 根据权利要求3所述的装置,其中,所述加权加法器响应于所述接收到的至少一个训练范例而调节被分配给所述至少一个低维分类器中的每个的所述输出的权重。
5. 根据权利要求1所述的装置,其中,所述加权加法器的输出是血液动力学不稳定性评分。
6. 根据权利要求1所述的装置,其中,所述加权加法器将为零的权重分配给关联特征不存在于所述接收到的数据中的分类器,并且在其他情况下将非零的权重分配给关联特征存在于所述接收到的数据中的分类器。
7. 根据权利要求1所述的装置,其中,所述低维分类器中的至少一个被配置为在其关联特征不存在于所述接收到的数据中时输出零值。
8. 根据权利要求7所述的装置,其中,所述加权加法器响应于通过所述低维分类器中的至少一个的零值的所述输出而调节被分配给输出非零值的所述低维分类器中的至少一个的所述输出的所述权重。
9. 一种由处理器执行的方法,所述处理器运行被存储在存储器上的用于执行鲁棒分类的指令,所述方法包括:
经由接收器接收数据,所述数据具有来自特征集合的至少一个特征;
向接收到的数据应用至少一个低维分类器,所述至少一个低维分类器中的每个与所述特征集合的相应的输入特征相关联,所述至少一个低维分类器的应用基于所述输入特征在所述接收到的数据中的存在来产生输出;并且
基于所述至少一个低维分类器的所述输出使用加权加法器来输出加权和。
10. 根据权利要求9所述的方法,其中,接收所述数据包括接收至少一个训练范例。
11. 根据权利要求10所述的方法,其中,使用接收到的至少一个训练范例来训练所述至少一个低维分类器中的每个。
12. 根据权利要求11所述的方法,其中,响应于所述接收到的至少一个训练范例而调节被分配给所述至少一个低维分类器中的每个的所述输出的所述权重。
13. 根据权利要求11所述的方法,其中,加权和输出是血液动力学不稳定性评分。

14. 根据权利要求9所述的方法, 其中, 输出分类器输出的加权和包括将为零的权重分配给关联特征不存在于所述接收到的数据中的分类器, 并且在其他情况下将非零的权重分配给关联特征存在于所述接收到的数据中的分类器。

15. 根据权利要求9所述的方法, 其中, 所述低维分类器中的至少一个被配置为在其关联特征不存在于所述接收到的数据中时输出零值。

16. 根据权利要求15所述的方法, 其中, 响应于通过所述低维分类器中的至少一个的零值的所述输出而调节被分配给输出非零值的所述低维分类器中的至少一个的所述输出的所述权重。

17. 一种包含计算机可执行指令的非瞬态计算机可读介质, 所述计算机可执行指令用于由处理器运行, 以用于执行用于鲁棒分类的方法, 所述介质包括:

用于经由接收器接收数据的计算机可执行指令, 所述数据具有来自特征集合的至少一个特征;

用于向接收到的数据应用至少一个低维分类器的计算机可执行指令, 所述至少一个低维分类器中的每个与所述特征集合的相应的输入特征相关联, 所述至少一个低维分类器的应用基于所述输入特征在所述接收到的数据中的存在来产生输出; 以及

用于基于所述至少一个低维分类器的所述输出使用加权加法器输出加权重的计算机可执行指令。

18. 根据权利要求17所述的介质, 其中, 用于输出分类器输出的加权重的所述指令包括用于将为零的权重分配给相关联特征不存在于所述接收到的数据中的分类器并且在其他情况下将非零的权重分配给关联特征存在于所述接收到的数据中的分类器的指令。

19. 根据权利要求17所述的介质, 其中, 所述低维分类器中的至少一个被配置为在其关联特征不存在于所述接收到的数据中时输出零值。

20. 根据权利要求17所述的介质, 其中, 响应于通过所述低维分类器中的至少一个的零值的所述输出而调节被分配给所述低维分类器中的至少一个的所述输出的所述权重。

鲁棒分类器

[0001] 相关申请的交叉引用

[0002] 本申请要求享有于2015年2月12日提交的待决美国临时申请No.62/115330的权益,通过引用将其整体公开并本文入,如同出于所有目的在本文中对其整体进行了阐述。

技术领域

[0003] 在此所描述的各种实施总体涉及用于鲁棒分类的方法和装置,并且更具体地,但并非排他地,涉及使用可能具有缺失或不完整的数据的数据集的分类。

背景技术

[0004] 数据集常常受缺失或不完整的数据的影响。数据可能出于各种原因而从具体数据集中缺失,诸如由于人为误差或出故障的数据采集设备造成的不充分的监测。基于这样的数据集的结论或分类因此可能是不准确的,而不管是出于缺失或不完整的数据的原因。

[0005] 然而,已经存在用于处理缺失或不完整的数据的特定技术。一种技术被称为数据填补。数据填补涉及计算针对缺失数据的近似。这种技术可以涉及构建针对给定数据特征的缺失值的估计器。估计值可以根据数据集中的其他值导出。例如,估计值例如可以是总体均值、中值、众数等。

[0006] 数据填补是易于实施的,但是不是非常准确,并且能够简单地向数据中注入噪声。额外地,比数据集的均值或中值的计算更为复杂的值估计可以要求关于特征之间的关系的先验知识和/或复杂的估计技术,以便有效。

[0007] 用于处理缺失或不完整数据的另一种技术是首先建立分类器的大的集合,其中的每个分类器取决于特征的不同子集。然后,基于当前输入中的测量特征的模式来识别最为合适的分类器,并且做出对其进行评价以对总体输入模式中的缺失条目的预测。

[0008] 尽管该技术避免了上述数据填补的问题,但是其受其自身的缺点的影响。针对一个,特征的可能子集的数量是特征的总数的组合。额外地,快速存储关于所有可能的特征子集(或者甚至其部分)的分类器随着总特征的数据的增加而变得不可行。

[0009] 因此,存在对于用于处理具有缺失或不完整数据的数据集的方法和装置的需求,其克服了上文所提到的缺点。

发明内容

[0010] 提供该简要说明来以简化的形式来介绍对原理的选择,其在下文中的详细说明部分中进一步详细描述。该简要说明并非旨在识别所请求保护的主题的关键特征或实质特征,也并非旨在被用作确定所请求保护的主题的范围中的辅助。

[0011] 各种实施例涉及用于鲁棒分类的装置。所述装置包括:接收器,其接收数据,接收到的数据具有来自特征集合的至少一个特征;存储器,其被配置为存储计算机可读指令,所述计算机可读指令用于执行鲁棒分类;以及处理器,其与所述存储器和所述接收器通信,其中,对用于执行鲁棒分类的所述计算机可读指令的运行令所述处理器实施:至少一个低维

分类器,其与所述接收器通信,其中,所述至少一个低维分类器中的每个与来自所述特征集合的相应的输入特征相关联,并且基于所述输入特征在所述接收到的数据中的存在来提供输出;以及加权加法器,其与所述接收器以及所述至少一个低维分类器中的每个通信,其中,所述加权加法器提供所述至少一个低维分类器的所述输出的加权和。

[0012] 在装置的一些实施例中,接收到的数据包括至少一个训练范例。在装置的一些实施例中,使用接收到的至少一个训练范例来训练所述至少一个低维分类器中的每个。在装置的一些实施例中,所述加权加法器响应于接收到的至少一个训练范例而调节被分配给所述至少一个低维分类器中的每个的输出的权重。

[0013] 在装置的一些实施例中,接收到的数据是从包括生命体征和实验室测量结果的组中选择的。在装置的一些实施例中,所述加权加法器响应于在接收到的数据中的至少一个特征的存在或不存在而调节所述权重。

[0014] 在装置的一些实施例中,所述加权加法器的输出是血液动力学不稳定性评分。

[0015] 在装置的一些实施例中,所述加权加法器将零的权重分配给关联特征不存在于接收到的数据中的分类器,并且在其他情况下将非零的权重分配给关联特征存在于接收到的数据中的分类器。

[0016] 在装置的一些实施例中,所述低维分类器中的至少一个被配置为在其关联特征不存在于接收到的数据中时输出零值。在装置的一些实施例中,所述加权加法器响应于通过所述低维分类器中的至少一个的零值的输出而调节被分配给输出非零值的低维分类器中的至少一个的输出的权重。

[0017] 根据本公开的另一方面,各种实施例涉及用于鲁棒分类的方法。所述方法包括:经由接收器接收数据,所述数据具有来自特征集合的至少一个特征;向接收到的数据应用至少一个低维分类器,所述至少一个低维分类器中的每个与所述特征集合的相应的输入特征相关联,所述至少一个低维分类器的应用基于所述输入特征在所述接收到的数据中的存在来产生输出;并且基于所述至少一个低维分类器的所述输出使用加权加法器来输出加权和。。

[0018] 在方法的一些实施例中,接收所述数据包括接收至少一个训练范例。在一些实施例中,使用接收到的至少一个训练范例来训练所述至少一个低维分类器中的每个。在一些实施例中,响应于接收到的至少一个训练范例来调节被分配给所述至少一个低维分类器中的每个的输出的权重。

[0019] 在方法的一些实施例中,接收到的数据是从包括生命体征和实验室测量结果的组中选择的。在一些实施例中,响应于至少一个特征在接收到的数据中的存在或不存在而调节分配的权重。在特定实施例中,加权和输出是血液动力学不稳定性评分。

[0020] 在方法的一些实施例中,输出分类器输出的加权和包括将零的权重分配给关联特征不存在于接收到的数据中的分类器,并且在其他情况下将非零的权重分配给关联特征存在于接收到的数据中的分类器。

[0021] 在方法的一些实施例中,所述低维分类器中的至少一个被配置为在其关联特征不存在于接收到的数据中时输出零值。

[0022] 在方法的一些实施例中,响应于通过所述低维分类器中的至少一个的零值的输出而调节被分配给输出非零值的所述低维分类器中的至少一个的输出的权重。

[0023] 根据本公开的又一方面,各种实施例涉及包含计算机可执行指令的计算机解读介质,所述计算机可执行指令用于执行用于鲁棒分类的方法。所述介质包括:用于经由接收器接收数据的计算机可执行指令,所述数据具有来自特征集合的至少一个特征;用于向接收到的数据应用至少一个低维分类器的计算机可执行指令,所述至少一个低维分类器中的每个与所述特征集合的相应的输入特征相关联,所述至少一个低维分类器的应用基于所述输入特征在所述接收到的数据中的存在来产生输出;以及用于基于所述至少一个低维分类器的所述输出使用加权加法器输出加权值的计算机可执行指令。

[0024] 在介质的一些实施例中,用于输出分类器输出的加权值的指令包括如下指令:其用于将为零的权重分配给相关联特征不存在于接收到的数据中的分类器并且在其他情况下将非零的权重分配给关联特征存在于接收到的数据中的分类器。

[0025] 在介质的一些实施例中,所述低维分类器中的至少一个被配置为在其关联特征不存在于接收到的数据中时输出零值。

[0026] 在介质的一些实施例中,响应于通过所述低维分类器中的至少一个的零值的输出而调节被分配给所述低维分类器中的至少一个的输出的权重。

[0027] 根据对下文详细描述的阅读以及对关联附图的查看,表征当前非限制实施例的这些和其他特征和优点将是显而易见的。应当理解,先前的一般性描述和下文的详细描述两者仅仅是解释性的,而并非对所请求保护的非限制性实施例的限制。

附图说明

[0028] 附图并非旨在按照比例进行绘制。在附图中,在各图中所图示的每个相同或接近相同的部件可以由相似的附图标记来表示。出于清楚的目的,可以不在每个附图中标记每个部件。现在将通过范例参考附图来描述各种实施例,在附图中:

[0029] 图1图示了根据一个实施例的鲁棒分类装置;

[0030] 图2示意性图示了可以结合图1的鲁棒分类装置100使用的预测性训练模块;

[0031] 图3图示了一表格,其示出了可以在使用图1的鲁棒分类装置100计算血液动力学评分中考虑的示范性特征;

[0032] 图4示意性图示了可以结合图1的鲁棒分类装置100使用的实时预测模块;

[0033] 图5描绘了根据一个实施例的用于鲁棒分类的方法的流程图;并且

[0034] 图6呈现了根据一个实施例的用于对数据集的鲁棒分类的系统。

具体实施方式

[0035] 下文参考附图更为完全地描述各种实施例,所述附图形成其部分,并且所述附图示出了特定示范性实施例。然而,本公开的想法可以以许多不同的形式来实施,而不应当被解读为限制到在本文中所阐述的实施例;相反,这些实施例被提供作为完全和完整公开的部分,从而向本领域技术人员传达本公开的想法、技术和实施方式的范围。实施例可以被实践为方法、系统或装置。因此,实施例可以采用硬件实施方式、完全软件实施方式或组合软件和硬件方面的实施方式的形式。因此,下文的详细描述并非是以限制性的意义采取。

[0036] 在本说明书中对“一个实施例”或“实施例”的引用意味着结合实施例所描述的具体特征、结构或特性被包括在根据本公开的至少一个范例实施方式或技术中。在说明书中

的各种位置中的短语“在一个实施例中”的出现不必都参考相同的实施例。

[0037] 可以在被存储在计算机存储器内的非瞬态信号上的操作的符号表示方面呈现下文的描述的一些部分。这些描述和表示被数据处理领域的技术人员用于将其工作的实质有效传达给本领域中的其他技术人员。这样的操作通常要求对物理量的物理操纵。通常,尽管不是必要的,但是这些量采取能够被存储、传输、组合、比较和以其他方式操纵的电学、磁性或光学信号的形式。有时方便的是,主要出于通常使用的原因,将这些信号称为位、值、元素、符号、字符、术语、数字等。此外,有时也方便的是,将要求在不失一般性的情况下对物理量的物理操纵的步骤的特定布置称为模块或代码设备。

[0038] 然而,所有这些和相似的术语都要与适当的物理量相关联并且仅仅是应用于这些量的方便的标签。除非明确陈述,否则,如根据下文的讨论明显的,将意识到,贯穿说明书,利用诸如“处理”或“计算”或“运算”或“确定”或“显示”等的术语进行的讨论指的是计算机系统或类似电子计算设备的动作和处理,其对被表示为计算机系统存储器或寄存器或其他这样信息存储、传输或显示设备内的物理(电子)量的数据进行操纵和变换。本公开的各部分包括过程和指令,其可以被实现在软件、固件或硬件中,并且当被实现在软件中时,可以被下载以驻留在由各种操作系统所使用的不同平台上或者从其操作。

[0039] 本公开也涉及用于执行在本文的操作的装置。该装置可以被专门构造用于所要求的目的,或者其可以包括由被存储在计算机中的计算机程序选择性地激活或重新配置的通用计算机。这样的计算机程序可以被存储在计算机可读存储介质中,所述计算机可读存储介质诸如是,但不限于:任何类型的盘,包括软盘、光盘、CD-ROM、磁-光盘、只读存储器(ROM)、随机存取存储器(RAM)、EPROM、EEPROM、磁性或光学卡、专用集成电路(ASIC)或者适于存储电子指令的任何类型的介质,并且每个可以被耦合到计算机系统总线。此外,在说明书中所提到的计算机可以包括单个处理器,或者可以是采用多处理器设计的架构以获得增加的计算能力。

[0040] 本文中所呈现的过程和显示不固有地涉及任何具体计算机或其他装置。各种通用系统也可以利用根据在本文中的教导的程序来使用,或者可以证明方便构造更专用的装置来执行一个或多个方法步骤。在下文的描述中讨论了针对各种这些系统的结构。另外,可以使用足以实现本公开的技术和实施方式的任何具体编程语言。各种编程语言可以被用于实施如在本文中所讨论的本公开。

[0041] 另外,在说明书中所使用的语言已经主要出于可读性和指导性的目的被选择,并且可能并非被选择用于描绘或限制所公开的主题。因此,本公开旨在是在本文中所讨论的想法的范围的图示而非限制。

[0042] 图1示意性图示了根据示范性实施例的鲁棒分类器100。在该实施例中,鲁棒分类器100可以包括接收器102、至少一个低维分类器设备104_{1,2,3...n}以及加权加法器106。各种实施例可以被用在各种应用中,亦即,可以被用在依赖于数据的任何应用中。例如,各种实施例可以被用在医疗保健、金融、市场分析、运动分析、零售等中。

[0043] 在操作中,接收器元件102接收来自至少一个数据源(未示出)的至少一个输入特征。如果鲁棒分类器100被实施在医疗保健设施中,例如,数据源可以是用于收集来自患者的信息(即,特征)的传感器设备。该信息可以涉及患者的心率(HR)(在这种情况下,对应的传感器设备将是心率传感器)、血压、体重、代谢、年龄等。

[0044] 尽管图示了四个输入特征,但是预期可以使用超过或少于四个输入特征,并且输入特征的数量可以取决于应用。所述数据源可以经由任何类型的硬线或无线连接将所述特征传递到接收器102。

[0045] 来自所述数据源中的每个数据源的数据可以与至少一个特征相关联。特征可以是性状、特性、或者描述对象(例如,患者)的变量。每个特征然后可以被传递到指定的低维分类器设备104。

[0046] 每个分类器设备104然后可以输出预测评分,或者在其相关特征中的任何缺失的情况下为零(0)。这些分类器设备104能够独立于彼此训练,或者使用诸如提升的多变量学习技术来联合地训练。

[0047] 所述评分然后可以被传递到加权加法器106。加权加法器106也可以与接收器设备102进行通信。加权加法器106可以向每个特征分配权重,并且由加法器106分配的权重可以取决于特定特征在输入数据中的存在或不存在。加权加法器106然后可以输出加权加法器输出的总和作为针对模型的预测,从而提供鲁棒数据分析以及分类,尽管是缺失的或不完整数据。

[0048] 各种实施例涉及两个阶段:(1)训练;以及(2)使用经训练的模型的实时预测。训练阶段还能够被分解成(a)对低维分类器的训练;以及(b)对动态系统模型的训练。在各种实施例中,这两个阶段可以由相同的设备来执行,而在其他实施例中,第一设备可以执行训练阶段,而第二设备执行实时预测阶段。例如,训练设备可以处理训练集以创建经训练的模型,其然后可以被分布在被部署在各个位置之中的许多实时预测设备之中。

[0049] 各种实施例处理缺失或不完整数据,而没有数据填补。这使得在本文中所描述的特征可应用于宽范围的问题。各种实施例训练低维分类器的集合并且将其组合为有力的系统。在系统中的分类器的加权是动态的,并且根据当前输入中的测量的特征而变化,其因此增加对缺失的数据的鲁棒性和可靠性。

[0050] 图2图示了根据一个实施例的训练模块200。训练模块200可以利用图1的鲁棒分类装置100来实施或者以其他方式与图1的鲁棒分类装置100结合使用。所述训练阶段可以接收训练数据集202中的n个标记的训练范例 $((x)^{(1)}, (y)^{(1)}), \dots, ((x)^{(n)}, (y)^{(n)})$ 作为输入。每个输入 $x^{(i)}$ 是p维输入模式,并且其第i个训练范例的第j个输入元素可以由 $x_j^{(i)}$ 来表达; $y^{(i)}$ 是关联类别标记或其他输出值,其被测量或以其他方式被确定以针对第i个训练范例的输入集合进行校正。缺失的值是允许的,即, $x_j^{(i)}$ 被假设为针对一些特征j和训练范例i是缺失的。目标是构建模型,所述模型根据来自输入的新的集合(与训练集合 $x^{(i)}$ 中的那些相一致)来预测新的输出值(与训练集合 $y^{(i)}$ 中的那些相一致)。

[0051] 为了完成该目标,使用机器学习方法,诸如,例如逻辑回归或线性回归,来训练多个单变量或低维分类器。(一个或多个)低维分类器训练模块204中的每个训练m个低维分类器的集合, $f_1(x), f_2(x), \dots, f_m(x)$,其中,每个分类器接受一个或多个输入特征,并且提供输出。经训练的分类器的数量m通常是在数据集中的特征的数量p的量级上。例如,当每个分类器 $f(x)$ 是仅一个输入特征的函数,那么 $m=p$ 。每个分类器输出预测评分,或者在其相关特征中的任意特征缺失的情况下为0。这些分类器 $f_1(x), f_2(x), \dots, f_m(x)$ 能够独立于彼此训练,或者使用诸如提升的多变量训练技术来联合地训练。

[0052] 在本文中所描述的实施例的特征可以在各种应用中实施。亦即,实施例可以适于

几乎任何类型的数据相关的应用,其可以受缺失或不完整数据影响。

[0053] 在医疗保健设施中,关于特定患者生命体征和/或实验室值的数据可以由于患者监测缺陷而缺失或不完整的。例如,患者监测设备可能发生故障,护理提供者可能忘记获得这样的信息,或者对这样的信息的采集可能不是患者的处置计划的部分。然而,当存在关于特定患者生命体征和/或实验室值的数据时,这些特征能够是在对患者状态的预测方面具有高度辨别力的。因此,期望具有在预测性模型中利用稀疏测量的或不完整的特征(即,缺失数据)的技术。

[0054] 在一个实施例中,例如,在本文中所描述的实施例的特征可以被用于预测患者的血液动力学不稳定性。心血管系统的根本目的是为了保证身体组织的足够的灌注和供氧,以维持正常、健康的组织和器官功能。在正常情况下,在健康生理系统中,存在多个反馈回路和补偿机制,有助于维持适当的血压和心输出,以实现端部器官的充分灌注。

[0055] 然而,在生理学上脆弱的患者之中,像通常在重症监护单元(ICU)中遇到的那些患者中,这些机制会变为折衷的;诸如败血症、出血和急性心脏衰竭的重大疾病过程可能导致对这些控制功能的显著损害以允许血液动力学恶化。因此,在这样的情况下,ICU临床医师常常面临通过透彻理解无数ICU数据并且与适当的介入(静脉内流体、血液制品以及药理学制剂的形式)起反应以帮助患者维持适当的心脏输出和灌注来优化血液动力学的挑战。对血液动力学不稳定发作的早期检测以及对适当的校正性介入的即时发起能够显著地改进患者结果。

[0056] ICU临床医师被呈现大量的生理学数据,其包括周期性并且频繁地采样的测量结果(例如,逐秒、逐分钟、5分钟或者15分钟,取决于具体设备配置),诸如心率和呼吸率,以及非周期性测量结果,诸如无创血压和实验室研究。他们必须针对在其护理下的每位患者采集、汇总和解读这些丰富的数据源。对逼近的血液动力学不稳定的解读和反应在存在过量数据、频繁故障警报以及频繁中断的工作流程的情况下会是尤其困难的任务。

[0057] 本申请的特征因此可以帮助临床医师在血液动力学恶化的早期阶段或者甚至在发作之前来检测血液动力学恶化,从而使得他们能够将注意力引导至可能从其受益最多的那些患者。鲁棒分类装置可以通过有意义地组合在当前ICU环境中可用的数据来实现这种情况。这是对先前的技术的改进,其中,其利用可能通常不测量但当测量时会非常重要的信息。重要的是要强调,这种技术并非旨在替换现有的故障保护机制,诸如当血压下降到临界水平(例如,收缩血压<90mmHg)时发出声音的警报;相反,其旨在通过将注意力引导至先导状况来扩展这些警报,所述先导状况可能要求介入,由此为临床医师给出更多时间来采取校正性动作。

[0058] 在该实施例中,所述鲁棒分类装置的输出可以是血液动力学不稳定性评分。该评分可以被用于准确地预测患者的血液动力学不稳定,即使特定信息可能是缺失或不完整的。尽管先前的讨论针对血液动力学不稳定性风险评分,但是本申请的特征可以在任何其他类型的应用(在临床决策支持的内部和外部两者)中实施,并且具体地,在可能受缺失和不完整数据影响的应用中实施。

[0059] 图3例如示出了列举可以在计算血液动力学风险评分以及其测量频率(即,具有针对该特征的至少一个测量结果的患者的百分比)中使用的示范性特征的表格300。在这种情况下,每个输入特征 $x^{(i)}$ 可以对应于在表格300中所列举的特征,并且 $y^{(i)}$ 是患者状态标签

(其在这种情况下可以是表示患者的稳定性的风险评分,诸如从0到1的评分)。

[0060] 图3的表格300图示了数据可能常常是不完整的,并且特征可能不存在于每个特征模式中或者不是在每个特征模式中测量的。例如,特定的特征,诸如乳酸和中心静脉压力,均可能缺失80%或更多的样本。

[0061] 继续在图2中的图示,在获得(或者训练,例如,从低维分类器训练模块204生成)分类器 $f_1(x)$ 、 $f_2(x)$ 、 $\dots$ 、 $f_m(x)$ 之后,动态系综学习模块206向 m 个低维分类器中的每个分配权重以形成系综。被分配给每个分类器 $f(x)$ 的权重是动态的,即,在输入中的个体测量的特征的存在或不存在的参数化函数。所分配的加权调制每个个体分类器对汇总的风险评分的效果的强度。

[0062] 例如,使 $H(x) = f_1(x_1) + f_2(x_2) + f_3(x_3)$,其中,三个单变量分类器是近似线性相关的,使得:

$$[0063] \quad \beta_1 f_1(x_1) + \beta_2 f_2(x_2) + \beta_3 f_3(x_3) \approx 0$$

[0064] 并且 β_1 、 β_2 和 β_3 是标量。亦即,尽管所述特征自身并未被假设为相关的,其单变量分类器预测被假设为相关的。

[0065] 目标是忠实地再现以上在缺失数据(即,如果不存在特征)的情况下对 $H(x)$ 的预测。例如,假定给定的输入模式是缺失针对 x_1 ($x_1 = \phi$)的值,其使得 $f_1(x_1)$ 放弃提供输出。给定 $f_2(x_2)$ 和 $f_3(x_3)$,我们能够通过对其进行再现来考虑 $f_1(x_1)$,使得:

$$[0066] \quad H(\mathbf{x} = (\phi, x_2, x_3)) \approx \left(1 - \frac{\beta_2}{\beta_1}\right) f_2(x_2) + \left(1 - \frac{\beta_3}{\beta_1}\right) f_3(x_3)$$

[0067] 如果 x_2 或 x_3 缺失,则能够导出相似的等式。

[0068] 接下来的步骤是识别加权函数 $a_1(x)$ 、 $\dots$ 、 $a_m(x)$,以形成系综分类器 $H(x)$:

$$[0069] \quad H(x) = a_1(x) f_1(x) + \dots + a_m(x) f_m(x)$$

[0070] 加权函数 $a(x)$ 是通过动态系综学习模块206来提供的,并且其具体值取决于在输入中存在哪些特征。亦即:

$$[0071] \quad a_i = s_{i0} + s_{i1}m(x_1) + \dots + s_{ip}m(x_p)$$

[0072] 其中, s_{ij} 是系数,而 $m(x_j)$ 是指示器掩膜,其得到二值向量 $\mathbf{m} = (m_1, m_2, \dots, m_p)$,其中,当第 j 个特征缺失时, $m_j = 0$,或者当第 j 个特征被测量时, $m_j = 1$ 。组合以上等式,系综分类器能够被写为:

$$[0073] \quad H(\mathbf{x}) = \sum_{i=1}^m s_{i0} f_i(\mathbf{x}) + \sum_{i=1}^m \sum_{j=1}^p s_{ij} f_i(\mathbf{x}) m(x_j)$$

[0074] 根据该等式,分类器 $H(x)$ 在系数中是线性的, s_{ij} , $i=1, \dots, m$, $j=0, 1, \dots, p$,因此,这些系数可以使用诸如逻辑回归的标准线性分类技术来学习,以预测关于训练集合的标签。具体而言,为了从 $H(x^{(1)})$ 、 $H(x^{(2)})$ 、 $\dots$ 、 $H(x^{(n)})$ 预测 $y^{(1)}$ 、 $y^{(2)}$ 、 $\dots$ 、 $y^{(n)}$ 。

[0075] 最终的预测模型是由 $H(x)$ 来定义的,其是从 m 个低维分类器 $f_1(x)$ 、 $f_2(x)$ 、 $\dots$ 、 $f_m(x)$ 以及加权函数 $a_1(x)$ 、 $\dots$ 、 $a_m(x)$ 构造的,其是通过系数 s_{ij} , $i=1, \dots, m$, $j=0, 1, \dots, p$ 来参数化的。在血液动力学评分应用中,例如,系数 s_{ij} , $i=1, \dots, m$, $j=0, \dots, p$ 可以在部署之前在训练数据上的次级离线阶段中学习。

[0076] 因此,每个分类器的权重是动态的,并且取决于特征的测量模式。这允许最终的分

[0077] 一旦已经构造所述分类器,其被实时地应用以执行在稀疏或不完整数据集上的分类。图4图示了实时预测模块400的一个实施例。实时预测模块400可以利用图1的鲁棒分类装置100来实施或者以其他方式与其结合使用。一旦已经训练了预测模型,实时预测模块400能够评价可能包括缺失的数据的新的输入。在该阶段中,预测评分通过计算模型预测 $H(x)$ 而被分配给新的输入 x 。

[0078] 输入 x 402可以包括一系列 p 特征 $x_1, x_2, \dots, x_p$ 。如果鲁棒分类装置被用于计算患者的血液动力学评分中,如上文所讨论的,则这些特征可以包括在图3的表格300中所列举的特征中的任意。接下来,每个低维分类器 $f_1(x), f_2(x), \dots, f_m(x)$ 404被应用到输入 x ,其中,每个分类器 $f_1(x), f_2(x), \dots, f_m(x)$ 404与输入特征 $x_1, x_2, \dots, x_p$ 相关联。在一个或多个分类器404可能为多变量(例如,低维但并非单变量)的实施例中,这样的分类器可以与所述输入特征中的超过一个输入特征相关联。

[0079] 掩蔽模块406然后可以确定针对 p 特征中的每个的适当的掩膜 $m(x_1), \dots, m(x_p)$ 。注意到,如果第 j 个特征缺失,则 $m(x_j) = 0$,并且如果第 j 个特征存在,则 $m(x_j) = 1$ 。如先前所提到的,加权加法器408能够通过使用如下等式来分配权重:

$$[0080] \quad a_i = S_{i0} + S_{i1}m(x_1) + \dots + S_{ip}m(x_p)$$

[0081] 加权加法器408向分类器 $f_1(x), f_2(x), \dots, f_m(x)$ 的输出应用权重 $a_1(x), \dots, a_m(x)$,以输出单个评分 $H(x)$ 。在血液动力学评分实施例中,通过下式表示单个血液动力学不稳定指示器(hii)风险评分:

$$[0082] \quad hii(x) = \sum_{i=1}^p a_i(x) f_i(x_i)$$

[0083] 因为这是针对hii的组成单变量分类器的加权平均,所述加权可以取决于所述输入(图3的哪些特征存在于/不存在于输入 x)的测量模式。可以基于哪个(些)特征缺失来对特定特征进行更多或更少强烈地加权。

[0084] 加权加法器408可以调节特征的权重以本质地替换缺失的特征,如果当其他特征未被测量时“替换”特征被经常测量。作为在血液动力学不稳定性评分应用中的范例,血红蛋白和血细胞比容是高度相关的,并且存在一个被测量而其他未被测量的少数情况一因此,它们具有替换彼此的很少的值。因此,加权加法器408可以偏置于选择或者非常频繁测量的或者完全针对缺失特征测量的“替换”特征。

[0085] 存在可以在血液动力学不稳定性评分应用中开发的该类型的多个关系。例如,参考图3,仅在大约17%的患者中测量乳酸。在这些情况下,被调节以考虑缺失的乳酸特征的主导特征是动脉 $PaCO_2$ (其在45%的患者中被测量)。动脉 $PaCO_2$ 被强烈地加权,因为乳酸和动脉 $PaCO_2$ 两者与缺氧呼吸有关。

[0086] 作为另一范例,碳酸氢盐(HCO_3)测量存在于仅48%的范例中。在这些情况中,通过调节二氧化碳预测器(当缺失碳酸氢盐时,其可用于96%的范例)的权重来考虑缺失的碳酸氢盐量度。因为血液中的大部分二氧化碳以碳酸氢盐的形式存在,因此相应地对二氧化碳进行加权。

[0087] 所输出的hii评分可以在非归一化的尺度上,其中,大的正值可以指示血液动力学不稳定的高风险,并且大的负值可以指示血液动力学不稳定的低风险。为了产生更多的直觉,针对临床医师或其他感兴趣部分的用户友好的评分,所述评分可以通过例如使其经过S形函数来归一化,例如:

[0088]
$$HII(x) = \frac{1}{1 + \exp(-hii(x))}$$

[0089] 这产生了范围[0,1]中的归一化的评分HII(x),其中,接近1的值可以指示血液动力学不稳定的高风险,并且更接近0的值可以指示血液动力学不稳定的较低风险。该评分可以经由图形显示器被呈现给临床医师或其他类型的医务人员。类似地,HII评分可以与颜色相关联以进一步强调风险的水平(例如,绿色=低等风险,黄色=中等风险,红色=高等风险等)。

[0090] 图5描绘了根据一个实施例的用于鲁棒分类的方法500的流程图。步骤502涉及经由接收器接收具有至少一个特征的数据。所述接收器可以通过任何硬线或无线连接来接收数据。该数据可以涉及患者的健康,例如,并且可以由医务人员输入,和/或从健康监测系统或设备来接收。

[0091] 在预测性训练阶段中,接收到的数据包括至少一个训练范例。亦即,使用所述训练范例来训练至少一个低维分类器。

[0092] 步骤504涉及将至少一个低维分类器应用到接收到的数据。这导致基于接收到的数据的输出,其中,所述低维分类器中的每个与接收到的数据中的特征中的一个相关联。

[0093] 步骤506涉及确定针对特征的每个集合的适当的掩膜。如先前所陈述的,如果第j个特征缺失,则 $m(x_j) = 0$,并且如果第j个特征存在,则 $m(x_j) = 1$ 。该步骤可以由掩蔽模块406来执行。

[0094] 步骤508涉及基于所述掩膜将权重应用到所述分类器。这些权重取决于掩膜,并且亦即,测量输入中的哪些特征。该步骤可以由加权加法器408来执行。

[0095] 步骤510涉及输出至少一个分类器的输出的加权和。如先前所陈述的,所述加权和部分取决于哪些特征不存在以及哪些特征存在。被分配给每个分类器的输出的权重也取决于训练范例。在血液动力学不稳定性评分应用中,该输出可以是指示患者是否处于血液动力学不稳定的风险中的血液动力学风险评分。

[0096] 图6图示了硬件系统600的范例,所述硬件系统用于实施可以参与在本文中所描述的各种系统的各种设备。硬件600可以实施用户移动设备或支撑体移动设备。如在图6中所示的,硬件600包括一条或多条系统总线610,其连接处理器620、缓存/系统存储器630、用户接口640、通信接口650和存储设备660。将理解,图6仅仅是示范性的并且在一些方面中构成抽象内容,并且硬件600的部件的实际组织可以变化并且可以比所图示的更为复杂。

[0097] 处理器620可以是能够运行被存储在存储器630或存储设备660中的指令或者以其他方式来处理数据的任何硬件设备。这样一来,处理器620可以包括微处理器、现场可编程门阵列(FPGA)、专用集成电路(ASIC)或其他相似设备。在一些实施例中,诸如依赖于一个或多个ASIC的那些实施例中,如经由软件部分提供的所描述的功能可以替代地被硬接线到ASIC的操作中,并且这样一来,可以忽略关联软件。

[0098] 缓存/系统存储器630可以包括各种存储器,诸如,例如L1、L2或L3缓存或系统存储器。这样一来,存储器630可以包括静态随机存取存储器(SRAM)、动态RAM(DRAM)、闪速存储器、只读存储器(ROM)或其他相似的存储器设备。

[0099] 用户接口640可以包括一个或多个设备,其用于实现与诸如施予者的用户的通信。例如,用户接口640可以包括显示器、鼠标、键盘、触摸屏、按钮、相机、麦克风、振动器、触觉引擎等。在一些实施例中,用户接口640可以包括命令线接口或图形用户接口,其可以经由

通信接口650被呈现给远程终端。

[0100] 通信接口650可以包括一个或多个设备,其用于实现与其他硬件设备的通信。例如,通信接口650可以包括网络接口卡(NIC),其被配置为根据以太网协议来通信。另外,通信接口650可以实施TCP/IP堆栈,其用于根据TCP/IP协议的通信。用于通信接口650的各种备选或额外的硬件或配置将是显而易见的。在一些实施例中,通信接口650可以包括NFC、蓝牙或其他短程无线接口。用于通信接口650的各种备选或额外的硬件或配置将是显而易见的。

[0101] 存储设备660可以包括一个或多个机器可读存储介质,诸如只读存储器(ROM)、随机存取存储器(RAM)、磁盘存储介质、光学存储介质、闪速存储器设备、或者相似的存储介质。在各种实施例中,存储设备660可以存储用于由处理器620运行的指令,或者可以利用处理器620操作的数据。例如,存储设备660可以存储操作系统661,其用于控制硬件600的各种基本操作。

[0102] 操作系统661可以包括先前所描述的各种部件。这些部件例如可以是低维分类器模块662,其例如可以类似于图2的低维分类器学习模块204。这些部件可以训练 m 个低维分类器的集合, $f_1(x)$ 、 $f_2(x)$ 、 $\dots$ 、 $f_m(x)$,其中,每个分类器接受一个或多个输入特征并且提供输出。

[0103] 操作系统661的另一部件可以是动态系综学习模块663。图6的动态系综学习模块663可以类似于图2的动态系综学习模块206,例如。动态系综学习模块663可以将权重分配至 m 个低维分类器中的每个以形成系综。如先前所陈述的,被分配给每个分类器的权重 $f(x)$ 是动态的,并且可以调制每个个体分类器在汇总的风险评分上的效应的强度。

[0104] 例如,(一个或多个)低维分类器664可能类似于图4的低维分类器404。每个分类器664 $f_1(x)$ 、 $f_2(x)$ 、 $\dots$ 、 $f_m(x)$ 与输入特征 x_1 、 x_2 、 $\dots$ 、 x_p 相关联。在一个或多个分类器664可能为多变量(例如,低维但并非单变量)的实施例中,这样的分类器可以与所述输入特征的超过一个输入特征相关联。

[0105] 例如,掩蔽模块665可以类似于图4的掩蔽模块406。如先前所陈述的,掩蔽模块665可以确定针对 p 特征中的每个特征的适当的掩膜 $m(x_1)$ 、 $\dots$ 、 $m(x_p)$ 。注意到,如果第 j 个特征缺失,则 $m(x_j)=0$,并且如果第 j 个特征存在,则 $m(x_j)=1$ 。

[0106] 例如,加权加法器666可以类似于图4的加权加法器408。如先前所陈述的,加权加法器666向分类器 $f_1(x)$ 、 $f_2(x)$ 、 $\dots$ 、 $f_m(x)$ 的输出应用权重 $a_1(x)$ 、 $\dots$ 、 $a_m(x)$,以输出单个评分 $H(x)$ 。这是组成单变量分类器的加权平均,并且加权可以取决于对所述输入的测量模式。可以基于哪个(些)特征缺失更强烈或更不强烈地对特定特征进行加权。

[0107] 尽管在本文中已经描述和图示了若干实施例,但是本领域技术人员将容易设想用于执行功能和/或获得在本文中所描述的结果和/或优点中的一个或多个的各种其他手段和/或结构,这样的变化和/或修改中的每个被认为处在本发明的范围之内。更一般而言,本领域技术人员将容易意识到,在本文中所描述的所有参数、尺度、材料和配置旨在是示范性的,并且实际参数、尺度、材料和/或配置将取决于一个或多个具体应用,针对所述应用,使用本发明的教导。本领域技术人员将认识到,或者能够使用不超过常规实验来确认,在本文中所描述的本发明的具体实施例的许多等价物。因此,应当理解,前述实施例仅通过范例的方式呈现,并且其在权利要求及其等价物的范围之内,本发明可以以具体描述和请求保护

的之外的方式来实践。本发明涉及在本文中所描述的每个个体特征、系统、物品、材料和/或方法。另外,两个或更多个这样的特征、系统、物品、材料和/或方法的任意组合,如果这样的特征、系统、物品、材料和/或方法不是相互抵触的,被包括在本发明的范围之内。

[0108] 在本文中在说明书和在权利要求书中所使用的词语“一”和“一个”,除非明确指示相反,应当被理解为意味着“至少一个”。

[0109] 在本文中在说明书和在权利要求书中所使用的术语“和/或”应当被理解为意味着这样结合的元件中的“任一个或两者”,例如,在一些情况中结合地呈现的并且在其他情况中分类地呈现的元件。可以任选地存在由“和/或”条款具体指定的元件之外的其他元件,不管是否与具体识别的那些元件相关或不相关,除非明确指示是相反的。因此,作为非限制性范例,对“A和/或B”的引用,当被用在诸如“包括”的开放式语言结合使用是,在一个实施例中,能够指代A没有B(任选地包括B之外的其他元件);在另一实施例中,指代B没有A(任选地包括A之外的元件);在又一实施例,指代A和B两者(任选地包括其他元件);等等。

[0110] 如在本文中在说明书和在权利要求书中所使用的,“或”应当被理解为具有与如上文所定义的“和/或”相同的含义。例如,当分隔列表中的项时,“或”或“和/或”应当被解读为包容性的,例如,对包括至少一个,但是也包括超过一个,多个元件或元件的列表,并且任选地,额外未列出的项。仅明确被指示相反的术语,诸如“仅其中之一”或者“精确地其中之一”,或者,当在权利要求中使用,“由……构成”将指代多个元件或元件的列表中的精确一个元件的包括。一般而言,如在本文中所使用的术语“或者”仅应当被解读为指示独有的备选(例如,“一个或另一个而非两者”),当依照独有的,诸如“任一个”、“其中之一”、“仅其中之一”或者“精确地其中之一”在前时。“本质上由……构成”,当被用在权利要求中时,应当具有其在专利法的领域中的普通含义。

[0111] 在本文中在说明书和权利要求书中所使用的,关于一个或多个元件的列表的术语“至少一个”应当被理解为意指从元件的列表中的元件的任意一个或多个中选择的至少一个元件,但是不必包括元件的列表内具体列举的每个和每一个元件中的至少一个,并且不排除元件的列表中的元件的任意组合。该定义也允许元件可以任选地存在,而非在元件的列表内具体识别的元件,术语“至少一个”涉及所述元件,不管是否与具体识别的那些元件相关或不相关。因此,作为非限制性范例,“A和B的至少一个”(或等同地,“A或B中的至少一个”,或等同地“A和/或B中的至少一个”)可以在一个实施例中指代至少一个(任选地包括多于一个)A,而不存在B(并且任选地包括除B之外的元件);在另一实施例中,指代至少一个(任选地包括多于一个)B,而不存在A(并且任选地包括除A之外的元件);在又一实施例中,指代至少一个(任选地包括多于一个)A,以及至少一个(任选地包括多于一个)B(并且任选地包括其他元件);等等。

[0112] 在权利要求中,以及在上述说明书中,所有过渡短语,诸如“包括”、“包含”、“承载”、“具有”、“含有”、“涉及”、“保持”等要被理解为是开放式的,例如,意味着包括但不限于。

[0113] 仅过渡性短语“由……构成”以及“本质上由……构成”应当分别是封闭的或半封闭的过渡性短语,如在专利审查流程的美国专利局手册第2111.03节阐述的。

[0114] 在权利要求中对修饰权利要求元素的诸如“第一”、“第二”、“第三”等的顺序术语的使用自身并不暗示任何优先性、优先级、或者一个权利要求元素相对于另一个的顺序或

者执行方法的动作的时间顺序,但是仅仅被用作标记以区分具有特定名称的一个权利要求元素与具有相同名称的另一元素(但用于顺序术语的使用),以区分权利要求元素。

[0115] 还应当理解,除非明确指示相反,在在本文中所请求保护的包括超过一个步骤或动作的任意方法中,所述方法的步骤或动作的顺序不必受限于所记载的方法的步骤或动作的顺序。

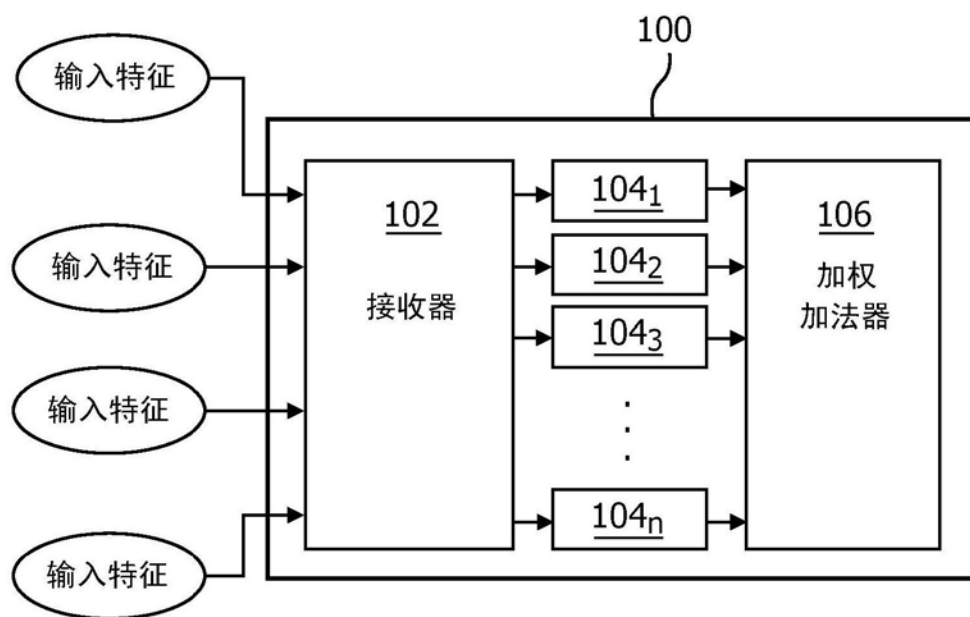


图1

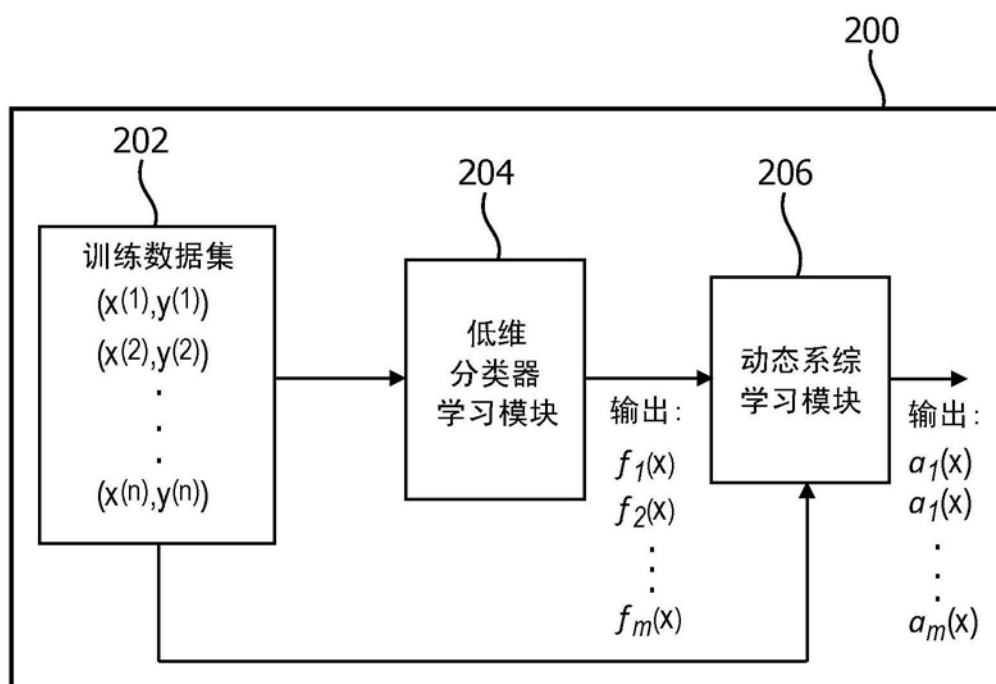


图2

300

动脉血气体		有创生命	
动脉pH	45%	有创平均血压 (iBPMean)	22%
碳酸氢盐 (HCO ₃)	48%	有创收缩血压 (iBPSys)	23%
动脉PaCO ₂	45%	有创舒张血压 (iBPDia)	23%
SaO ₂	43%	有创休克指数 (ISI)	24%
动脉血碱剩余	29%	中央静脉压力 (CVP)	14%
通气机参数		无创生命/人口统计学	
PF比率	27%	无创平均血压 (nBP Mean)	99%
FiO ₂ 设置	30%	无创收缩血压 (nBPSys)	99%
平均气道压力 (MAP)	8%	无创舒张血压 (nBPDia)	99%
峰值Insp压力 (PIP)	10%	心率	100%
		无创休克指数 (NSI)	98%
		年龄	98%
		温度 (T)	21%
基础代谢面板		综合代谢面板	
二氧化碳 (CO ₂)	96%	丙氨酸转氨酶 (ALT)	71%
氯化物	97%	白蛋白	71%
血液尿素氮 (BUN)	97%	碱性磷酸酶 (ALP)	70%
肌酸酐	97%	天冬氨酸转氨酶 (AST)	72%
钾	97%	总胆红素	73%
钠	97%	总蛋白质	67%
葡萄糖	97%		
钙	96%		
完全血液计数		完全血液计数分布	
WBC-白血球	96%	束带	9%
RBC	96%	嗜碱粒细胞 (Basos)	40%
白细胞比容	97%	嗜酸性细胞 (Eos)	40%
血红蛋白	97%	淋巴细胞 (Lymphs)	41%
血小板	96%	单核细胞 (Monos)	40%
		中性粒细胞 (Polys)	40%
		中性粒细胞对单核细胞比率 (NLR)	40%
额外的测试			
淀粉酶	15%	乳酸脱氢酶 (LHD)	14%
CPK	44%	镁	57%
CPK MB	38%	部分促凝血酶原激酶时间 (PTT)	64%
电离钙	19%	凝血时间 (INR)	69%
乳酸盐	17%	甘油三酸酯	20%

图3

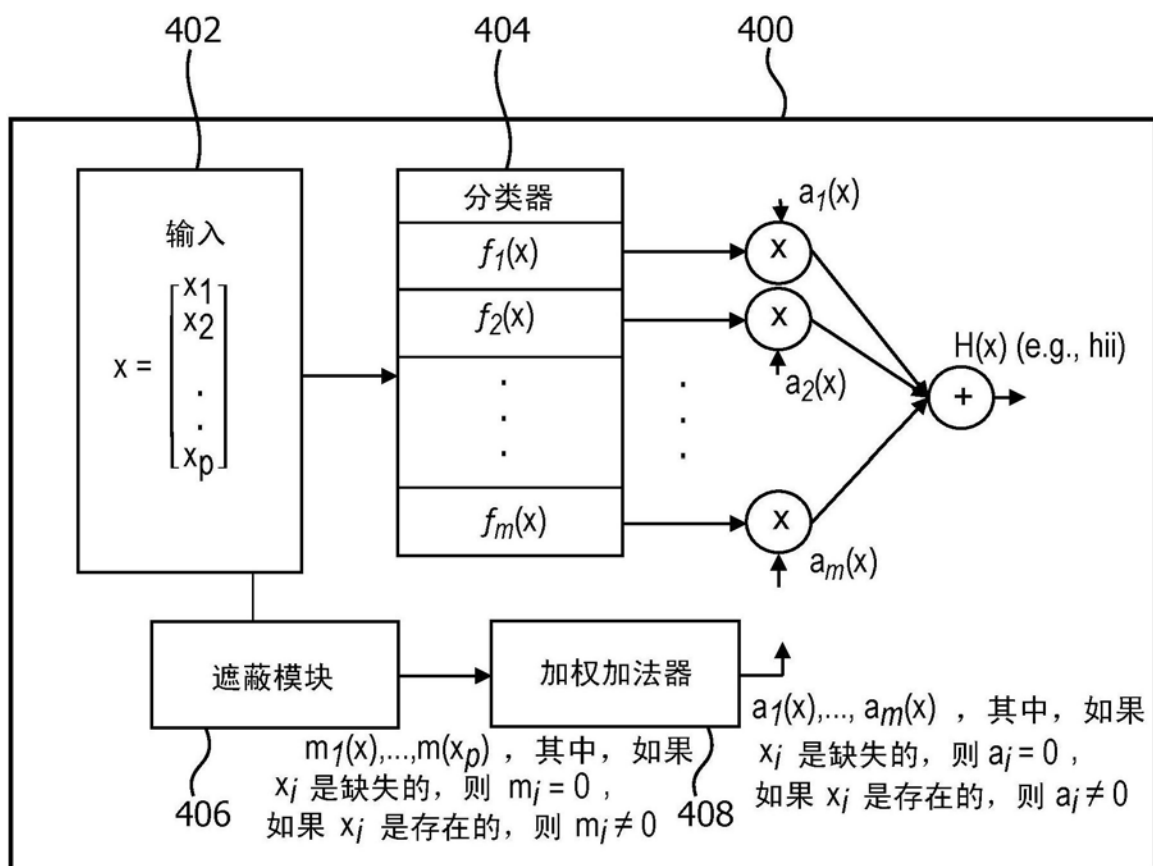


图4

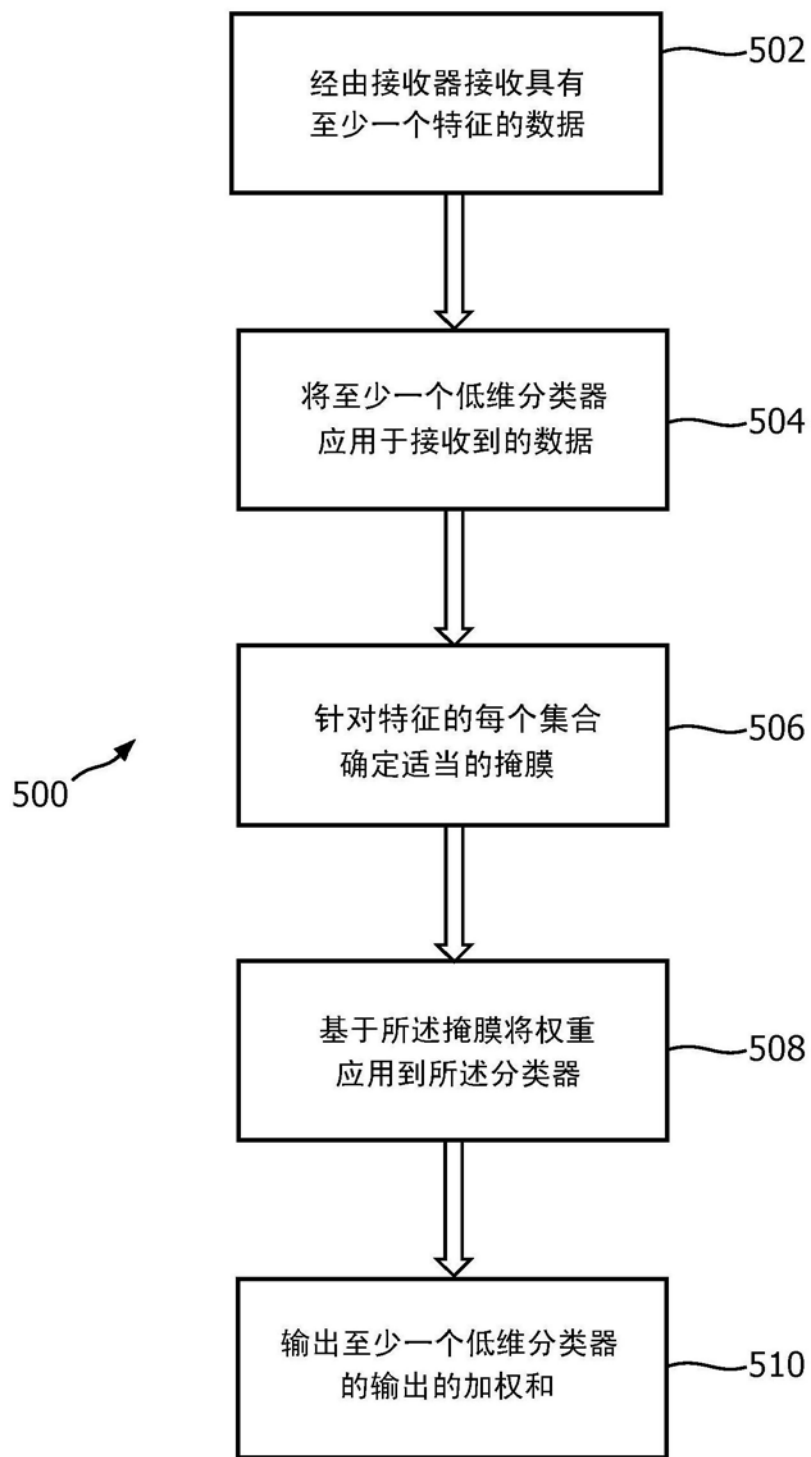


图5

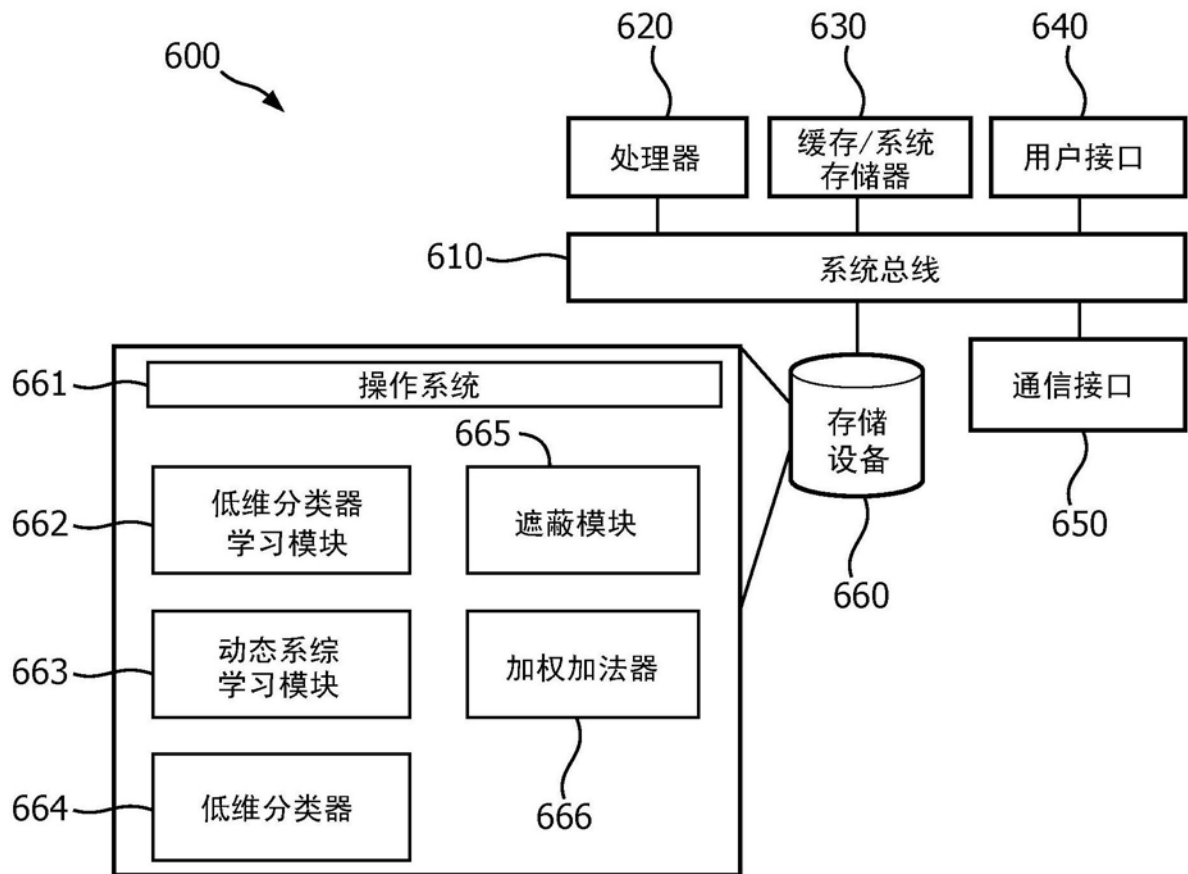


图6

专利名称(译)	鲁棒分类器		
公开(公告)号	CN107249434A	公开(公告)日	2017-10-13
申请号	CN201680009843.5	申请日	2016-02-05
[标]申请(专利权)人(译)	皇家飞利浦电子股份有限公司		
申请(专利权)人(译)	皇家飞利浦有限公司		
当前申请(专利权)人(译)	皇家飞利浦有限公司		
[标]发明人	B康罗伊 LJ埃谢曼 C 波特斯 M徐		
发明人	B·康罗伊 L·J·埃谢曼 C·波特斯 M·徐		
IPC分类号	A61B5/00 G06F17/18 G06F19/00		
CPC分类号	A61B5/7267 A61B5/7275 G06F17/18 G06N20/00 G16H50/20 G06F16/285 G06N5/04 A61B5/02028 G06F7/08 A61B2505/03		
代理人(译)	李光颖 王英		
优先权	62/115330 2015-02-12 US		
外部链接	Espacenet SIPO		

摘要(译)

在本文中所描述的各种实施例涉及用于鲁棒分类的方法和装置。许多真实世界的数据集受缺失和不完整的数据影响。通过基于哪个(些)特征是缺失的或不完整的将权重分配给数据集的特定特征，本发明的实施例能够提供针对缺失的数据的鲁棒性和恢复性。

$$Hx = (x_1, x_2, x_3) \approx \left(1 - \frac{\beta_1}{\beta_2}\right) f(x_1) + \left(1 - \frac{\beta_3}{\beta_2}\right) f(x_2)$$