(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 101926646 B
(45)授权公告日 2012.11.28

(21)申请号 201010224927.5
(22)申请日 2004.07.01
(30)优先权数据
60/483,961 2003.07.01 US
(62)分案原申请数据
200480018683.8 2004.07.01
(73)专利权人 卡迪尔马格成像公司
地址 美国纽约
(72)发明人 卡斯滕·斯特尼克尔
马克·J·恩布克斯
伯勒斯洛·K·希曼斯基
(74)专利代理机构 永新专利商标代理有限公司 72002
代理人 刘炳胜 王英
(51)Int.Cl.
A61B 5/04(2006.01)
G06K 9/00(2006.01)

(56)对比文件
US 6210344 B1,2001.04.03,
US 2002/0165837 A1,2002.11.07,
US 4958638 ,1990.09.25,

审查员 孔祥云

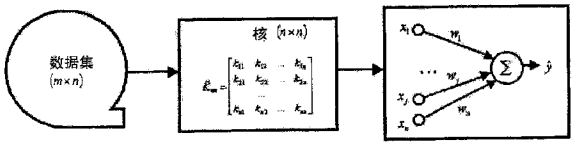
权利要求书 2 页 说明书 15 页 附图 4 页

(54)发明名称

使用机器学习来进行心磁图分类

(57)摘要

本发明公开了在心磁描记术中使用机器学习来进行模式识别，所述心磁描记术用于测量心脏电生理活动所发射的磁场。使用直接核方法来将异常 MCG 心脏图形与正常的图形加以分离。对于无监督学习来说，引入了基于直接核的自组织映射。对于有监督学习来说，使用了直接核部分最小二乘以及（直接）核岭回归。然后把这些结果与经典支持向量机以及核部分最小二乘相比较。由此，公开了一种用于分类心动描记数据的设备以及相关联的方法，包括把核变换应用于从用来检测电磁心脏活动的传感器获取的测数据，产生变换数据，然后使用机器学习来分类已变换数据。



CN 101926646 B

1. 一种自动识别有意义的特征并且形成用于分类心磁描记数据的专家规则的方法，包括：

将小波变换应用于从检测由病人的心脏活动产生的电磁场的传感器获取的检测数据，以产生小波域数据；

将核变换应用于所述小波域数据，从而得到变换数据；以及

采用机器学习，由所述变换数据来识别所述有意义的特征并且形成所述专家规则。

2. 权利要求 1 所述的方法，其中，所述核变换满足 Mercer 条件。

3. 权利要求 1 所述的方法，所述核变换包括径向基函数。

4. 权利要求 1 所述的方法，所述应用核变换的步骤包括：

把所述变换数据分配至神经网络的第一隐层；

应用训练数据描述符作为所述神经网络的所述第一隐层的权重；并且

数值计算所述神经网络的第二隐层的权重。

5. 权利要求 1 所述的方法，还包括：

采用直接核部分最小二乘 (DK-PLS) 机器学习来分类所述变换数据。

6. 如权利要求 1 所述的方法，所述把上述检测数据转换为小波域数据包括：

把 Daubechies 小波变换应用于上述检测数据。

7. 权利要求 1 所述的方法，还包括：

从上述小波域数据中选择用于改进心磁描记数据的所述分类的诸特征。

8. 权利要求 1 所述的方法，还包括：

归一化上述检测数据。

9. 权利要求 8 所述的方法，所述归一化上述检测数据还包括：

对上述检测数据进行马氏缩放。

10. 权利要求 1 所述的方法，还包括：

使所述核变换的核居中。

11. 一种用于自动识别有意义的特征并且形成用于分类心动描记数据的专家规则的设备，包括：

用于将小波变换应用于从检测由病人的心脏活动产生的电磁场的传感器获取的检测数据，以产生小波域数据的模块；

用于将核变换应用于所述小波域数据，从而得到变换数据的模块；以及

用于采用机器学习，由所述变换数据来识别所述有意义的特征并且形成所述专家规则的模块。

12. 权利要求 11 所述的设备，其中所述核变换满足 Mercer 条件。

13. 权利要求 11 所述的设备，所述核变换包括径向基函数。

14. 权利要求 11 所述的设备，用于应用核变换的所述模块还包括：

用于把所述变换数据分配至神经网络的第一隐层的模块；

用于应用训练数据描述符作为所述神经网络的上述第一隐层的权重的模块；以及

用于数值计算所述神经网络的第二隐层的权重的模块。

15. 权利要求 11 所述的设备，还包括：

用于采用直接核部分最小二乘 (DK-PLS) 机器学习来分类所述变换数据的模块。

16. 权利要求 11 所述的设备，用于把上述检测数据转换到小波域数据的所述模块包括：

用于把 Daubechies 小波变换应用于上述检测数据的模块。

17. 权利要求 11 所述的设备，还包括：

用于从所述小波域数据中选择用于改进心磁描记数据的所述分类的诸特征的模块。

18. 权利要求 11 所述的设备，还包括：

用于归一化上述检测数据的模块。

19. 权利要求 18 所述的设备，用于归一化上述检测数据的所述模块包括：

用于对上述检测数据进行马氏缩放的模块。

20. 权利要求 11 所述的设备，还包括：

用于使所述核变换的核居中的模块。

使用机器学习来进行心磁图分类

[0001] 本申请是 2004 年 7 月 1 日提交的申请号为 200480018683.8、名称为‘使用机器学习来进行心磁图分类’的分案申请。

背景技术

[0002] 虽然心磁描记法 (MCG) 早在二十世纪六十年代就被引入作为可能的诊断工具，但是花费了近三十年的时间才成功地证明了它的临床价值。现在，它代表新近出现的、全世界医院中内科医师采用的心脏病学技术之一。MCG 方法的临床应用明显受益于现代多通道传感器技术、成熟的软件以及硬件中的最新改善，所述硬件允许使用所述设备而无需磁屏蔽室。

[0003] MCG 研究是快速的、安全的并且是完全非侵入性的。因此，它为患者提供了极大的便利。目前，许多组织从事建立参考数据库和数据标准化的工作。在多种临床应用中，MCG 已经提供了临床上有益的结果。例如，MCG 可以诊断并且定位急性心肌梗塞、区分具有或不具有恶性心室心律不齐易感性的心肌梗塞患者、检测心室肥厚和心脏移植之后的排异性、定位心室预激励的位置和各类心律不齐以及揭示胎儿心律不齐和传导紊乱 [40]。另外，最近已经研究了 MCG 的其它几种临床应用 :心肌病的检测和危险分类（扩张、肥大、心律不齐、糖尿病）、自发心室纤维性颤动之后的危险分类、心肌耐久性的检测和定位以及胎儿生长和神经系统完整性的后续检查。某些研究已经明显地表明 :MCG 对复极（化）过程中（例如在心肌梗塞之后或者在遗传性长 QT 综合症 [42] 中）的变化非常灵敏。在 [41] 中可以找到 MCG 应用以及目前使用的分析技术的最相关的综述。

[0004] 然而，具有重大挑战性的是减少或者消除因 MCG 数据的人为解释所引入的偏离，并且明显改善基于机器的分类性能和推广质量，并同时使计算机处理时间与实时诊断相容。

[0005] 当把人工智能（机器学习）应用于测量数据时，始终执行三个基本步骤 :1、数据测量，2、测量数据的预处理，3、自适应分类器的训练。将这种基本方案编入 EKG/ECG 数据或者其它生物数据的专利包括美国专利 5,092,343 ;5,280,792 ;5,465,308 ;5,680,866 ;5,819,007 ;6,128,608 ;6,248,063 ;6,443,889 ;6,572,560 ;6,714,925 以及 6,728,691。

[0006] 使用人工智能进行 MCG 场图分析十分局限于数据。把人工智能应用于分析生物磁性信号的一篇参考文献是第 5,417,211 号美国专利，公开了一种用来对活体内发生的电生理活动所生成的场图进行分类的方法，包括如下步骤 :使用多通道测量设备来测量作为对象体外电生理活动结果而出现的场图、生成对应于所测场图的特征向量、把所述特征向量提供给自适应分类器以及利用训练场图来训练自适应分类器，其中所述训练场图已经由可定位电生理活动代用模型生成。所述方法还包括步骤 :在自适应分类器的输出端为每个场图生成概率值，该概率值表示能够由所选可定位代用模型来生成每一个场图的概率。类似于上面引用的 EKG/ECG 参考文献，该文献论述了机器学习对测量数据的普遍适用性，但是它没有提及改善分类性能以及推广质量的细节。

[0007] 在所有情况中，用于确定成功的两个关键措施是分类性能以及推广质量。对非预

处理数据进行训练会导致很差的分类结果，而所谓的过度训练则防碍自适应分类器推广到实际数据的正常识别。

[0008] 成功的关键在于数据的最优预处理，迄今为止，本文引用的任何参考文献还没有实现这一点。最重要的是识别能确定被探查数据集的所属分类的所有特征。识别那些特征既不是显而易见的也不是微不足道的。此外，这些特征可以根据生物系统的不同以及测量数据类型的不同而有所不同。因此，大部分基于人工智能的过程都在如何执行预处理方面存在差别。

[0009] 正如此处将要详细公开的那样，为机器学习而使用核变换以及小波变换对数据进行预处理为成功的机器学习方法提供了基础，就准确分类、推广质量以及处理速度而言，所述方法明显改进了现有技术。这些内容在本文引用的任何现有技术中均没有被公开或者建议。

发明内容

[0010] 本文公开了在心磁描记术中使用机器学习进行图形识别，所述心磁描记术测量心脏的电生理活动所发射的磁场。使用直接核方法来将异常 MCG 心脏图形与正常图形加以分离。对于无监督学习来说，引入了基于直接核的自组织映射。对于有监督学习来说，使用了直接核部分最小二乘以及（直接）核岭回归。然后把这些结果与经典支持向量机以及核部分最小二乘相比较。在测试以前，根据该训练数据的有效子集来调整用于这些方法的超参数。研究内容还包括使用局部、垂直、水平以及二维（全局）马氏 (Mahanalobis，马哈拉诺比斯）缩放、小波变换以及通过滤波进行变量选择来进行预处理。对所有三种方法都类似的结果令人鼓舞，它超过了已训练专家所达到的分类质量。

[0011] 本文公开了一种用于分类心动描记数据的设备以及相关方法，包括把核变换应用于从用来检测电磁心脏活动的传感器所获取的检测数据，产生变换数据，然后使用机器学习来分类已变换数据。

附图说明

[0012] 在所附权利要求书中阐明了被认为具有新颖性的本发明的特征。然而参照下述结合附图进行的描述可以理解本发明以及其进一步的目的和优势，所述附图汇总如下：

[0013] 图 1 举例说明了在包括 6×6 栅格的 36 个通道中经由一个心动周期所收集的已过滤平均瞬态 MCG 曲线。

[0014] 图 2 是说明正确图形和错误图形数目（基于 36 个测试数据的阴性和阳性情况）以及心磁图数据的执行时间的图表。支持向量机库 (SVMLib) 和核部分最小二乘 (K-PLS) 采用时间域，而其余的方法采用 D-4 小波域。

[0015] 图 3 是说明用于为心磁图数据创建预测模型的不同方法的质量测度的图表。

[0016] 图 4 是对小波变换数据的基于 K-PLS 的 35 种测试病例的误差曲线图。

[0017] 图 5 是表示假阳性和假阴性之间的可能权衡的接收器操作员特性 (ROC) 曲线。

[0018] 图 6 是基于（左）直接核主元分析 (DK-PCA) 和（右）直接核 PLS(DK-PLS) 的 73 个训练数据的投影图。患病的情况被为实心圆圈。没有画出测试数据。

[0019] 图 7 说明在环绕模式中，基于 9×18 直接核自组织映射 (DK-SOM) 显示在自组织映

象上的测试数据。

[0020] 图 8 说明了对测试数据集使用不同技术所得的局部缺血预测结果。

[0021] 图 9 是以直接核方法作为数据预处理步骤的操作示意图。

[0022] 图 10 是说明用于直接核方法的、采用核居中来进行数据预处理的流程图。

[0023] 图 11 是自组织映射 (SOM) 的典型结构。

[0024] 图 12 是此处公开的、用于心动描记数据自动分类的技术的列表。

具体实施方式

[0025] 此公开内容描述了在心磁描记术 (MCG) 中使用直接核方法和支持向量机进行图形识别，所述心磁描记术用于测量人类心脏的电生理活动所发射的磁场。用于 MCG 的、基于 SQUID 的测量设备目前正处于独立开发阶段，其中所述设备可被用于普通的医院房间（不需特别地屏蔽电磁干扰）。所述系统的操作是计算机控制的，并且在很大程度上是自动的。专用软件被用来进行精确的 24 位控制和获取数据，然后进行滤波、求均值、电 / 磁活动定位、心脏电流重构和诊断评分推导。

[0026] 对 MCG 记录的解释还遗留有挑战性问题。因此，此公开内容考虑了用于自动解释 MCG 测量结果，从而使用于分析的人工输入最少的方法。测试集中于检测局部缺血，这是在许多可能导致心脏病发作的常见心脏疾病中出现的一种情况，在美国这是死亡的主要原因，但这仅仅是示例性的，而不是限制性的。

[0027] 在科学上，此公开内容考虑了两类分离问题（患病的心脏与健康的心脏），其中描述符（数据点）的数目超出数据集的数目。因此，此公开内容现在集中致力于解决该问题的两个任务。

[0028] 第一个有待回答的问题是所述问题究竟是线性还是非线性，因为这将确定能够解决所述问题的可能候选函数的类别（称为“假设”或者“机器学习技术”）。我们的目标在于保持分离过程自身的线性，如果预处理中有非线性，则对非线性进行编码。后者可以通过在执行实际机器学习之前把（非线性）核变换应用于所述数据来实现（我们把对核变换数据进行操作的技术称为“非线性”技术）。因此，如果所述数据包含非线性，那么与非线性技术相比，纯粹的线性方法将显示出很差的性能。

[0029] 第二个目的在于寻求（或者开发）实际解决该分离问题的机器学习技术。此处的焦点主要不在于得到最优的解决方案，而是得到对所述数据执行得同样好的一类技术。这有助于建立对所选模型以及它们的推广能力的信心（一个假设能对不在训练集中的数据正确分类的能力被称为其“推广性”）。很容易开发对所述训练数据能进行最优操作但是无法预测未见数据的模型（所述现象经常被称为过度训练）。然而，建立（并且调整）只基于少数数据就能进行良好预测的模型是非常困难的。

[0030] 我们将从论述数据获取和预处理开始。特别是，我们讨论对于不同的学习方法，哪种预处理适合。此后，我们核心结果 :不同机器学习技术对于我们的问题的性能的比较，以及评估预测质量和调整参数选择的方法。此后，我们讨论特征选择。

[0031] 数据获取和预处理

[0032] MCG 数据是在躯干上方的 36 个部位通过在彼此相邻的位置中进行四次顺序测量来获取的。在每个位置中，九个传感器使用 1000 赫兹的采样速率在 90 秒内测量心脏磁场，

从而产生 36 个单独时间序列。对于缺血诊断来说，需要 0.5 赫兹至 20 赫兹的带宽，所以采用了使用六次贝塞耳滤波器特性的 100 赫兹硬件低通滤波器，继之采用利用相同特性、但更高次的 20 赫兹附加数字低通滤波器。为了消除剩余的随机噪声分量，使用心动周期的 R 峰最大值作为触发点来平均整个时间序列。对于自动分类来说，我们采用来自心动周期 J 点和 T 峰 [5] 之间的时窗的数据，其中根据所测数据内插了 32 个均匀间隔的点的值。所述训练数据包括 73 个病例，这些病例易于由已训练专家目测分类。该测试对一组 36 个病例进行，所述病例包括其心磁图对执行目测分类的已训练专家产生误导或迷惑的患者。

[0033] 在该情况下，通过首先从每个信号中减掉偏移值来预处理数据。然后，我们研究对于我们的多变量时间序列信号最有效的预处理，包括局部、垂直、水平和二维（全局）马氏缩放以及小波变换。一个重要的考虑是保护数据局部性，这是通过对每个信号应用 Daubechies（多布西）-4 小波变换 [3] 来实现的。之所以选择它，是由于在每个内插的时间信号中的数据集 (32) 相对较小。只有用于观察输入端的数据局部性的 SOM 和 K-PLS 方法不要求这种变换。接下来，我们对数据进行马氏缩放，首先对所有 36 个信号进行，然后垂直进行（对所有信号，但基于 SOM 的方法除外）。图 1 中显示了被内插至 ST 段 [5] 中 32 个等间隔点并且在对每一个单个信号进行马氏缩放之后的 36 个信号的典型数据集。

[0034] MCG 数据分类的预测建模

[0035] 机器学习的目的在于将某些智能决策基础委托给计算机。在其当前形式中，机器学习的重要部分是开发鲁棒的分类、回归工具和特征选择方法。

[0036] 在心脏诊断环境中，机器学习的最终目的是要能够识别有意义的特征，所述特征可以解释所述模型并且能够表述具有透明度的专家规则。

[0037] 机器学习的关键因素是防止过度训练。Tikhonov（基霍诺夫）调整的概念在机器学习中是达到这个目的的一种十分有效的概念。机器学习中的第二个问题是需要构造可靠的非线性方法。支持向量机 (SVM) 以及其它基于核的方法（诸如核主元分析、核岭回归以及部分最小二乘）都是把非线性及调整合并到机器学习方法中去的有效方法。

[0038] 在机器学习中，当前具有挑战性的问题是特征多于数据的大型问题、在数据中存在许多不确定因素以及噪音的问题以及具有混合模式的无序多类分类问题。

[0039] 对正确预处理的需要极度取决于领域的选择，但是研究不同的预处理方法，并将领域的专业知识合并到这个预处理阶段是使机器学习方法可行的关键因素。

[0040] 我们的目的在于不仅识别“最佳”的机器学习方法，而且识别对数据能够执行得同样好的一类技术。因此，我们考虑 SVM，这是机器学习领域的一种主要工具。我们也使用可能比 SVM 更易于调整或者更易于硬件实现、但是预期其性能可以与 SVM 媲美的其它基于核的方法。

[0041] 成功的机器学习的一个关键在于数据的预处理。许多不同的预处理情况是值得考虑的。我们如下区分四类预处理：

[0042] 1、归一化：为了使得数据能够被加以比较需要这样做。这通常指的是数据被按比例缩放并去偏。然而，这里人们有许多选择。

[0043] 2、信息定位：所谓定位，我们指的是应用能够重排数据的变换来使包含绝大部分信息的系数被首先呈现。一个突出的例子是小波变换，它甚至能保持信息的局部性。

[0044] 3、特征选择：这通常对已经变换的数据执行。它指的是不包含信息或者包含很少

信息的系数被剔除以便减少输入域的维数。这对于加速学习尤其有益。

[0045] 4、核变换：所述核变换是使回归模型非线性的一种简明的方式。核是包含数据集的相似性测度的矩阵：或者是指该数据集的自身数据之间的相似性，或者是指与其它数据之间的相似性（例如，支持向量[2]）。很明显，这样做提供了用于改进心动图分类的方法的多种组合。

[0046] 首先讨论归一化。这是机器学习中使所有描述符居中并且使它们具有单位方差的常用方法。相同的过程然后也被应用于响应。居中及方差归一化的这个过程被称为马氏缩放。虽然马氏缩放不是对所述数据进行预处理的唯一方式，但是，要进行能够很好适用于全体数据的预处理，这大概是最普遍的并且最鲁棒的方式。如果我们把特征向量表示为$\vec{z}$，那么马氏缩放将产生被重新缩放的特征向量$\vec{z}'$，并且可以被总结为：

[0047] $$\vec{z}' = \frac{\vec{z} - \bar{z}}{std(\vec{z})} \qquad (1)$$

[0048] 其中$\bar{z}$表示平均值，而$\mathrm{std}(\vec{z})$表示属性$\vec{z}$的标准差。

[0049] 当36个时间序列被（沿时间轴）分别单独缩放时，我们将称其为“水平马氏缩放”，当一个时刻的所有36个点都被缩放时，则称其为“垂直马氏缩放”，而当所有32个时刻的36个点都被缩放时，则称其为“全局马氏缩放”。

[0050] 接下来我们讨论定位。应用小波变换[10]一方面有助于定位信息的“热点”，另一方面有助于定位对所述信号没有贡献的“冷区”。小波变换比傅里叶变换更加合适的属性在于单个小波函数是空间定位的。傅里叶正弦和余弦函数却不是如此。小波变换不具有单独的基函数集，但具有可能基函数的无穷集。

[0051] “母函数”或者“分析小波”$\varphi(x)$的扩展和平移定义了一个正交基，亦称小波基：

[0052] $$\Phi(s,l) = 2^{\frac{-s}{2}}\Phi(2^{-s}x - l) \qquad (2)$$

[0053] 变量s和l是整数，用于缩放和扩张母函数Φ(x)以便生成小波，诸如Daubechies小波系列。缩放指数s表明小波的宽度，而位置指数l给出其位置。应当注意的是，所述母函数通过2的幂来重新缩放或者“扩展”，并且按照整数平移。之所以对小波基特别感兴趣是因为缩放和扩展所引起的自相似性。一旦我们了解了所述的母函数，那么我们就可以理解与所述基有关的一切。

[0054] 为了以不同的分辩率覆盖我们的数据域，把所述分析小波用于缩放公式：

[0055] $$W(x) = \sum_{k=-1}^{N-2} (-1)^k c_{k+1} \Phi(2x + k) \qquad (3)$$

[0056] 其中W(x)是母函数Φ(x)的缩放函数，而c_k是小波系数。所述小波系数必须满足如下形式的线性二次型约束：

[0057] $$\sum_{k=0}^{N-1} c_k = 2, \sum_{k=0}^{N-1} c_k c_{k+2l} = 2\delta_{l,0} \qquad (4)$$

[0058] 其中δ是δ函数而l是位置指数。

[0059] 小波最有用的特征之一是科学家可以很容易用它来选择适用于给定问题的给定小波系统的定义系数。在Daubechies的论文[3]中，她开发了能够很好地表示多项式特性的特定小波系统族。对于MCG时间序列来说，所谓的“Daubechies 4”小波表现出了最优性

能。

[0060] 我们现在转而讨论核变换。核变换及其调整是改进心动图分类能力的重要成分。因此，我们将比较详细地解释这一概念，并且强调在应用核变换时通常被忽略的某些主要问题。

[0061] 核变换是使回归模型非线性的一种简明方式。核变换至少可以追溯到二十世纪初期，当时希尔伯特把核引入数学文献。核是包含数据集的相似性测度的矩阵：或者是指该数据集自身的数据之间的相似性，或者是指与其它数据之间的相似性。核的典型用途是作为主元分析中的相关矩阵，其中特征核包含了属性或者特征之间的线性相似性测度。在支持向量机中，所述核的各元是数据之间而不是特征之间的相似性测度，并且这些相似性测度通常是非线性的。存在许多可能的非线性相似性测度，但是为了便于进行数学处理，所述核必须满足某些条件，既所谓的 Mercer（默塞尔）条件 [2、11、15]。

[0062] $$\vec{\vec{K}}_{nn}=\begin{bmatrix} k_{11} & k_{12} & \cdots & k_{1n} \\ k_{21} & k_{22} & \cdots & k_{2n} \\ & \cdots & & \\ k_{n1} & k_{n2} & \cdots & k_{nn} \end{bmatrix} \qquad (5)$$

[0063] 上述表达式引入了 n 个数据的数据核矩阵$\vec{\vec{K}}_{nm}$的通用结构。所述核矩阵是一个对称矩阵，其中每个元均包含两个数据向量之间的（线性或者非线性）相似性。存在许多不同的可能方法来定义相似性尺度，譬如作为线性相似性测度的点积以及作为非线性相似性测度的径向基函数核或 RBF 核。所述 RBF 核是最广泛使用的非线性核，它的元被定义为：

[0064] $$k_{ij} \equiv e^{-\frac{\left\|\vec{x}_j-\vec{x}_l\right\|_2}{2\sigma^2}} \qquad (6)$$

[0065] 应当注意的是，在上面的核定义中，所述核元包含数据点之间以负指数表示的欧几里得距离，这是相异性（而不是相似性）测度。所述负指数还包含自由参数 σ，这是 RBF 核的 Parzen（帕尔逊）窗口宽度。用于选择 Parzen 窗口的正确选择通常通过对外部有效集合进行附加调整来确定，这种调整被称为超调整。σ 的精确选择不是关键，通常存在某个使模型质量稳定的相对较宽的 σ 选择范围。

[0066] 这里把核变换作为数据变换应用于独立的预处理阶段。我们实际上利用非线性数据核来替代所述数据，并且应用传统的线性预测模型。我们引入了对数据的非线性核变换采用传统线性算法的方法，并且在这里被定义为“直接核方法”。这种直接核方法的简洁和优势在于：该问题的非线性特征被记录在该核中，并且对于所应用的算法而言是透明的。

[0067] 人们还可以在神经网络类型的流图中表示该核变换，并且第一隐层现在会产生核变换数据，而第一层的权重正是训练数据的描述符。第二层包含可以采用数值方法来计算的权重，譬如采用核岭回归方法（参见图 9）。当使用径向基函数核时，这类神经网络看上去往往与径向基函数神经网络 [17、18] 非常相似，只是第二层的权重计算结果不同。

[0068] 通过使所述核居中来处理偏移值也是重要的。参见通用的预测方程：

[0069]

$$\hat{\vec{y}}_n = X_{nm}\vec{w}_m \qquad (7)$$

[0070] 其中把权向量$\vec{W}_m$应用于数据矩阵 X_{nm} 以便达到（预测）输出$\hat{\vec{y}}_n$，不存在恒定的偏置项。结果，对于被居中的数据而言，该偏移项（“偏离”）始终是零，无须被显式包含在内。包含该偏离值的更加通用的预测模型不是应用公式 7，而是被写成：

[0071]

$$\hat{\vec{y}}_n = X_{nm}\vec{w}_m + b \qquad (8)$$

[0072] 其中 b 是偏离项。因为我们实施了首先通过马氏缩放来使数据居中的操作，所以此偏离项是零，可以忽略。

[0073] 在处理核时，由于需要某些类型的偏离值，所以情况更加复杂。这里我们将只给出一个技巧，在实际情况中很适用的技巧，要了解说明其原因的更多细节，可以参见文献 [11、12、16、19]。即使应用核变换之前对数据进行了马氏缩放，所述核仍需要某些类型的居中以便能够忽略预测模型中的偏离项。用于使核居中的直截了当的方式是从每一列训练数据核中减去平均值，并且当使测试核居中时，存储此平均值供以后调用。用于使该核居中的第二个步骤是再次审查最新获得的垂直居中的核，这次是逐行进行，并且从每一水平行中减去行平均值。

[0074] 对测试数据的核需要按照类似过程以一致的方式加以居中。在该情况下，所存储的训练数据核的列平均值将被用于测试数据核的垂直居中。然后这种垂直居中的测试核再被水平居中，即，为每一行计算垂直居中测试核的平均值，并且该垂直居中测验核的每个水平元由该元减去该行平均值来替代。

[0075] 如上所述，这个用于使核居中的算法的优点在于：它同样适用于矩形数据核。图 10 中画出了对所述训练数据、有效数据以及测试数据的流程图，用于预处理所述数据、对该数据应用核变换、然后使该核居中。

[0076] 对无监督以及有监督学习方法都进行了研究。对于无监督学习来说，由于经常把 SOM 应用于新颖性检测和自动聚类，所以使用了直接核 (DK)-SOM。所使用的 DK-SOM 具有展开边缘的 9×18 六边形网络。对于有监督学习来说，使用了四种基于核的回归算法：对从复数数据空间提取相关参数有效的经典支持向量机；由 Rosipal（罗西帕尔）[10] 提出的核部分最小二乘 K-PLS；直接核部分最小二乘 (DK-PLS) 以及最小二乘支持向量机（即 LS-SVM，亦称核岭回归）。

[0077] 支持向量机或者 SVM 由于它们的效率、模型灵活性、预测功效以及理论上的透明度已经被经证明是强有力的机器学习工具 [2、11、15]。SVM 的非线性属性可以仅仅归因于核变换，而诸如自组织映射或者 SOM[9] 之类的其它方法从本质上讲就是非线性的，因为它们包含各种基于邻域的运算。与 SVM 不同，SOM 的主要用途常常是作为可视化工具 [4] 来在某种二维映射上揭示高维数数据的潜在相似性 / 聚类结构，而不是用于回归或者分类预测。

[0078] 使用了内部开发的、用于分析的 Analyze/StripMiner 软件包 [14]，但也使用了适用于所述 SVM 模型的 SVMLib[1]。DK-SOM、SVM、DK-PLS 以及 LS-SVM 中的参数值已在测试之前使用所述训练集进行了优化。结果与已训练专家实现的分类质量相似，并且对于所有已测试方法均类似，即便这些方法使用了不同的数据预处理。这一点非常重要，因为它表明在任何测试方法中不存在过度训练。DK-PLS、SVMLib 以及 LS-SVM 之间的吻合程度特别好，对于这些数据而言，这些方法之间没有明显差异。图 2 和 3 中显示了所述结果。图 2 列出了对

于阳性和阴性情况正确分类的图形数目和错误的数目。图 3 提供了预测质量的附加测度。图 8 中显示了进一步的结果。在图 8 中，RMSE 表示均方根误差（越小越好），而 CC[%] 指的是正确分类情况的百分比。DK-PLS 方法获得了最优的结果，它也表现出了最强的鲁棒性。这已胜过组合的三个标准测试 (ECG、ECHO 以及 Troponin I) 的预测精度，对于这些患者来讲，这个精度是 58%。

[0079] 调整之后，SVM 的 Parzen 窗口宽度 σ 被选择为 10。在 SVMLib 中，调整参数 C 如 [10] 中所建议的那样被设置为 $1/\lambda$。基于其它应用 [14] 和缩放实验的经验，对于 n 个数据核，根据如下公式来确定岭参数 λ：

[0080] $$\lambda = \min\left\{1; \left(\frac{n}{1500}\right)^{\frac{3}{2}}\right\} \qquad (9)$$

[0081] 更普遍的是，凭经验，我们已经发现 λ 正比于数据数目 n 的 3/2 次幂。

[0082] 直接核方法 (DK-PLS 及 LS-SVM)、K-PLS 和传统的基于核的 SVM(SVMLib) 之间的一致性表明由此公式产生的岭参数近似最优的选择。

[0083] 现在转而讨论用于评定模型质量的尺度，对于回归问题，获取错误的另一个方式是通过均方根误差指标或者 RMSE，依照如下公式把它定义为均方差的平均值（对于训练集或者测试集）：

[0084] $$RMSE = \sqrt{\frac{1}{n}\sum_i (\hat{y}_i - y_i)^2} \qquad (10)$$

[0085] 虽然均方根误差是用于比较不同预测方法对同一数据的性能的有效方法，但是从 RMSE 取决于数据的响应如何被缩放这个意义上讲，它不是绝对的尺度。为了克服此障碍，还使用了对响应值的缩放和幅值依赖性较小的附加误差测度。用于评定已训练模型质量的第一尺度是 r^2，它被按照如下公式定义为响应的目标值和预测值之间的平方相关系数：

[0086] $$r^2 = \frac{\sum_{i=l}^{n_{train}} (\hat{y}_i - \bar{y})(y_i - \bar{y})}{\sqrt{\sum_{i=l}^{n_{train}} (\hat{y}_i - \bar{y})^2}\sqrt{\sum_{i=l}^{n_{train}} (y_i - \bar{y})^2}} \qquad (11)$$

[0087] 其中 n_{train} 表示训练集中的数据点的数目。r^2 取 0 和 1 之间的值，并且 r^2 的值越高，模型越好。使用 r^2 来评定模型质量的明显缺陷在于：它只表示线性相关，表明如果把 $\hat{y}$ 绘制成 y 的函数，那么预测值在多大程度上排成一行。虽然当 r^2 等于 1 时，人们会期待得到接近完美的模型，但情况并不总是这样。第二种用于评定已训练模型质量的更为有效的尺度是所谓的“压缩 r 平方”，或者 R^2，经常用于化学计量学建模 [6]，其中 R^2 被定义为 [7]：

[0088] $$R^2 = 1 - \frac{\sum_{i=l}^{n_{train}} (y_i - \hat{y}_i)^2}{\sum_{i=l}^{n_{train}} (y_i - \bar{y})^2} \qquad (12)$$

[0089] R^2 被认为是比 r^2 更好的测度，因为它同样考虑到残差。正如 r^2 一样，R^2 的范围在 0 和 1 之间，并且 R^2 的值越高，模型越佳。R^2 的尺度通常小于 r^2。对于较大数据集来说，R^2 势必收敛至 r^2，并且 r^2 和 R^2 之间对于这种数据的比较经常揭示隐藏的偏离值。

[0090] 为了评定有效集和测试集的质量，我们引入了类似的尺度 q^2 和 Q^2，对于测试集内的数据而言，q^2 和 Q^2 分别被定义为 $1-r^2$ 和 $1-R^2$。对于能对测试数据进行理想预测的模型来说，我们往往希望 q^2 和 Q^2 等于零。引入在训练集和测试集之间对称的尺度的理由实际上是为了避免混淆。Q^2 和 q^2 的值应用于有效集或测试集，为了获得良好的预测模型，人们往往希望这些值非常低。R^2 和 r^2 的值应用于训练数据，很容易发现，如果预测值接近实际值，那么它们两个都接近于 1。因此，它们中的任何一个与 1 有显著偏差都表明模型具有很差的预测能力。

[0091] 同所述核方法相比，部分最小二乘那样的线性方法会产生很差的预测模型。对于 K-PLS 和 DK-PLS 来说，选择了 5 个隐藏变量，但是结果并非决定性地取决于隐藏变量数目的正确选择。还尝试了直接核主元分析 (DK-PCA)，这是 K-PCA 的直接核方案 [11-12,16]，但是结果对主元数目的选择更加敏感，而且不如使用其它直接核方法所获得的结果那么好。

[0092] 图 4 中显示了基于小波变换数据和 DK-PLS 的心磁图数据的典型预测结果。从该图可见，在预测值中，总共六个数据点被错误分类（一个健康的或阴性情况，五个缺血情况）。已训练专家也难以根据通过专用方法获得的时变磁场的二维可视显示来正确识别这些病例。

[0093] 对于医疗数据来说，能够在假阴性和假阳性情况之间，或者在灵敏度和特异性之间（它们是与假阳性和假阴性相关的不同尺度）做出权衡通常是非常重要的。在机器学习方法中，这种权衡可以很容易通过改变解释所述分类的阈值来实现。例如，在图 4 中，人们可以不使用零作为判别值，而是将判别阈值移动到更加理想的水平，从而影响假阳性 / 假阴性的比例。

[0094] 判别值的这种改变的所有可能结果的汇总可以用 ROC 曲线显示，图 5 所示为上述情况。ROC 曲线（或者接收器操作员特性）的概念源自二十世纪四十年代用于识别飞行器的雷达设备的早期开发，并且在 [13] 中予以概述。

[0095] 图 6 显示了基于（左）直接核主元分析 (DK-PCA) 和（右）直接核 PLS(DK-PLS) 的 73 个训练数据的投影。患病的情况被表示为实心圆圈。与图 6 左侧显示的 DK-PCA 的结果相比，图 6 的右侧显示了根据 DK-PLS 的前两个元所得到的不同类之间的更清楚的分离和较宽的边缘。最初用明暗交叉线显示在这些图 (pharmaplot) 上的测试数据表明，对两种方法，健康和患病情况之间分离得极好。

[0096] 图 7 表示环绕方式中在六边形网格上的基于直接核 SOM 的典型 9×18 自组织映射。所述环绕方式指的是左右边界（以及顶部和底部边界）互相渗透，并且所述映射是环状投影的展开。黑色六边形表明患病的情况，明亮六边形表明健康的情况。全彩色的六边形表明训练数据的位置，白色和黑色阴影编号是健康和患病测试情况的图形识别符。大部分错误分类实际上出现在映射的边界区域。该映射中的单元通过半监督学习而被染色，就是说，包含 36×32 或者 1152 个特征的每一数据向量通过表示色彩的附加场来补充。对于权重向量，数据向量中的彩色元依照类似方式被更新，但是不使用它们来计算用于确定获胜单元的距离尺度。常规 SOM 实现方式的结果映射与直接核 DK-SOM 获得的映射非常相似。在 128 兆奔腾 III 计算机上生成 DK-SOM 的执行时间是 28 秒，而生成普通 SOM 需要 960 秒，这是因为在对数据进行核变换之后，数据大小从原始的 1152 有效地降至 73（训练数据的数目）。SOM 和 DK-SOM 的精细调整在监控模式下采用学习向量的量化来完成 [9]。虽然基于

SOM和DK-SOM的结果仍然非常好，但是它们不如利用其它基于核的方法(SVMLib、LS-SVM和K-PLS)所获得的结果那样好。

[0097] 特征选择

[0098] 上一部分中介绍的结果是采用所有1152(36×32)个描述符获得的。如果人们可以精确识别用于进行优良二元分类所需的最重要信息在时间或小波信号的什么位置，或者是对每位患者在不同位置测量的36个心磁图信号中的哪些信号，那么这往往可以给所属领域的专家提供更有价值的信息。这种信息可以通过特征选择来推导。

[0099] 特征选择，即对数据向量的最重要的输入参数的识别，可以依照两种不同的方式来进行：过滤方式和环绕方式。通常，这两个方法互相独立地被采用；然而，在这份公开内容的范围内及其关联的权利要求书中，它们也可以结合使用。

[0100] 在过滤方式中，基于事先规定的并且通常是无监督的程序来删除特征。这种程序的一个例子可以是删除包含四个σ离群值的描述符列，在化学计量学的PLS应用中经常发生这种情况。另外常见的是，在过滤模式中删除“表亲(cousin)”描述符，即与其它描述符的相关性达95％以上的特征。根据建模方法，惯用的方法是删除表亲描述符并且只保留以下的描述符：(i)与响应变量具有最高相关性的描述符，或者(ii)对解释所述模型的领域专家具有最清楚的域透明度的描述符。

[0101] 特征选择的第二种方式是基于环绕方式。人们希望只保留获得良好预测模型所必需的最相关的特征。建模质量通常在适当选择最优特征子集之后得以改善。确定特征的正确子集可以基于不同的概念来进行，而且由此生成的特征子集经常取决于建模方法。环绕方式中的特征选择通常采用训练集和有效集来进行，而且该有效集被用来确认所述模型没有因选择一组虚假描述符而过度训练。特征选择的两个通常可用的方法是基于遗传算法和灵敏度分析的使用。

[0102] 利用遗传算法方法的思想是要能够从训练集内获得最优特征子集，而且对于有效集也表现出良好的性能。

[0103] 灵敏度分析的概念[8]采用了特征的凸显性，即，一旦已经构造了预测模型，就对每一个描述符的平均值使用所述模型，并且所述描述符在最小值和最大值之间一次一个地加以调节。描述符的灵敏度是指预测响应中的变化。它的前提是，当描述符的灵敏度很低时，它大概不是构造良好模型的主要描述符。在一个迭代步骤期间可以删除几个最不灵敏的特征，灵敏度分析过程被多次重复，直到留下某个近似最优的特征集。遗传算法方法和灵敏度分析方法都是真正的软计算方法，并且要求一些启发和经验。两个方案的优点在于，遗传算法和灵敏度方法是通用的方法，它们不依赖于特定的建模方法。

[0104] 有关机器学习的进一步评述

[0105] 我们在此不是回顾所有可资利用的机器学习技术，而是首先讨论为什么我们不简单地使用支持向量机(SVM)这个同时适用于线性和非线性问题的最新解决方案。科学地讲，如上所述，我们的目标是找到对于给定问题能够执行得同样好的一类技术以便确保获得稳定的解。在此类技术中，最优模型是最易于调整并且执行最迅速的一个。把这些模型以SVM作为标准进行比较有助于验证任何最新开发的技术的性能。

[0106] 关于有监督学习，我们在此就有监督学习中的所谓机器学习悖论给予简短说明，这是开发了大量模型以便走出困境的原因。

[0107] 通常把数据矩阵表示为 X_{Nm}，把响应向量表示为 $\vec{y}_N$，假定在数据集内存在 N 个数据点和 m 个描述特征。我们想要通过归纳法按照如下方式从 X_{Nm} 推断出 $\vec{y}_N$（记为 $X_{Nm} \Rightarrow \vec{y}_N$），即我们的推理模型由 n 个训练数据点导出，而且对采样以外的数据（即 N-n 个有效数据以及测试数据点）工作得很好。换言之，我们旨在构造如下类型的线性预测模型：

[0108] $$\hat{\vec{y}}_n = X_{nm}\vec{w}_m \qquad (13)$$

[0109] 此公式假定一个必须在先前步骤中确定的已知权重向量 $\vec{w}_m$，在最优情况下，实际学习满足公式

[0110] $$X_{nm}\vec{w}_m = \vec{y}_n \qquad (14)$$

[0111] 这里，X_{nm} 是训练数据，并且 $\vec{y}_n$ 表示该已知解（“标记”）。

[0112] 应当注意，所述数据矩阵通常是不对称的。如果是这种情况，那么采用数据矩阵的逆就能直截了当地得到一个答案。因此，我们将采用伪逆变换，这通常不会产生 y 的精确预测值，但是将从最小二乘意义上依照最优方式来预测 y。以下举例说明对权重向量的伪逆解：

[0113] $$(X_{mn}^T X_{nm})\vec{w}_m = X_{mn}^T \vec{y}_n$$

[0114] $$(X_{mn}^T X_{nm})^{-1}(X_{mn}^T X_{nm})\vec{w}_m = (X_{mn}^T X_{nm})^{-1} X_{mn}^T \vec{y}_n$$

[0115] $$\vec{w}_m = (X_{mn}^T X_{nm})^{-1} X_{mn}^T \vec{y}_n$$

[0116] $$\vec{w}_m = (K_F)_{mm}^{-1} X_{mn}^T \vec{y}_n$$

[0117] $K_F = X^T_{mn}X_{nm}$ 是所谓的“特征核矩阵”，并且是机器学习悖论的理由：学习只是因为特征中的冗余而出现——但是，K_F 是病态的（降秩的）。正如先前表明的那样，存在多种方式来解决所述悖论：

[0118] 1、通过采用主元（计算特征核的特征向量）来固定 KF 的秩降 [18]

[0119] 2、通过调整：使用 $K_F+\lambda I$ 来代替 K_F（岭回归）[17,20-23]

[0120] 3、通过局部学习

[0121] 我们使用了四个基于核的回归算法：对从复数数据空间中提取相关参数有效的经典支持向量机 [2、1215]；由 Rosipal 提出的核部分最小二乘 K-PLS[10]；直接核部分最小二乘 (DK-PLS) 以及最小二乘支持向量机（即 LS-SVM，亦称核岭回归 [24-28]）。另外，我们测试了直接核主元分析 (DK-PCA)。

[0122] 部分最小二乘 (PLS) 是 QSAR 和化学度量学中的标准分析方法之一 [29]。核 PLS(K-PLS) 是最近开发的 PLS 非线性方案，由 Rosipal 和 Trejo（特里乔）提出 [10]。K-PLS 与 SVM 功能相当，但是 SVM 不同，结果变得更加稳定。K-PLS 目前被用来预测与人血清蛋白的结合亲和力。

[0123] 在作为此公开内容的基础的工作中，我们将 K-PLS 改进为 DK-PLS，并且利用了在 Analyze/Stripminer 程序中为 K-PLS、DK-PLS、DK-PCA 和 LS-SVM 开发代码的早期经验 [14]。K-PLS 和 DK-PLS 之间的差异在于在 K 方法中使用特征（数据）核矩阵，而在 DK 方法中，此矩阵由（非线性）核变换矩阵来替代。对于计算矩阵求逆来说，我们应用了在 Analyze/Stripminer 程序中实现的 Møller（穆勒）的比例共轭梯度算法 [30]。

[0124] 现在讨论无监督学习，我们注意到所述 SOM[9、17、31-36] 是由 Kohonen（科荷南）

开发的无监督学习神经网络。所述 SOM 是基于竞争学习的迭代方法。它提供从高维输入数据空间到低维输出映象的映射，这通常是一维或者二维映射，参见图 11。元（或者数据点）被载入输入层，所述 SOM 则使用竞争学习算法加以训练 [4]。所述权重依照如下公式来更新：

[0125] $$\vec{w}_m^{new} = (1-\alpha)\vec{w}_m^{old} + \alpha\vec{x}_m$$

[0126] 其中 α 是学习速率参数。由于学习的结果，所述输入数据将被映射到“获胜的”神经元。由于这个过程，所述 SOM 经常被用于降维和聚类。此外，SOM 的显著的特征是：它保留输入数据从高维输入空间到输出映象上的拓扑结构的方式是输入数据之间的相对距离或多或少得到了保留 [38]。在输入空间中位置彼此靠近的输入数据点被映射到输出映象上的邻近神经元。基于 SOM 的观察法是数据探索的通用工具。它们被用于数据的聚类、相关检测和投影 [4、39]。

[0127] 传统的 SOM 是一种基于高维输入数据到低维输出映象上的投影的方法。这里公开的是一种新的基于核的 SOM。所述核 SOM 现在根据数据的核表达式来加以训练，而不是根据原始数据来训练。这里使用核变换数据不足以“揭示”数据中的非线性，这是因为 SOM 原本是非线性的，但是因为该核具有更少的有效特征，所以增加了（学习）速度。

[0128] 总而言之，我们已经使用并且开发了图 12 所示的一组机器学习工具。

[0129] 结束语

[0130] MCG 数据的二元分类代表某种具有挑战性的问题，但是如果要使 MCG 在临床实践中成功，那么它的求解是非常重要的。把现有机器学习技术（诸如 SOM 和 SVM）应用于 MCG 数据将产生 74％的预测精度。首先把数据变换到小波域并对小波系数应用核变换，甚至单独应用核变换而不用小波变换，可以取得非常显著的改进。这样做使预测精度增加至 83％。

[0131] Rosipal[10] 提出的核 PLS(K-PLS)、直接核 PLS(DK-PLS)、支持向量机 (SVMLib) 和最小二乘 SVM(LS-SVM) 之间的结果的一致性通常非常好。在这种情况下，DK-PLS 具有优越的性能，而基于核的方法之间的差异不是很显著。这种极高的一致性显示出了直接核方法的鲁棒性。只有当通过公式 (1) 选择的岭参数接近最优时，才可以实现它。在支持向量机中，这种选择还定义了调整参数 C，并且 C 被取为 $1/\lambda$。

[0132] 获得的结果对于医学界来说是十分有意义的。对于检测由冠状动脉血管造影术定义的缺血而言，DK-PLS 被用来达到 92％的灵敏度和 75％的特异性。应当注意的是，MCG 是一种纯功能性的工具，对于心脏电生理学中的异常十分灵敏，因此，可只诊断疾病的结果。然而，金标 (gold standard，冠状动脉血管造影术）是一种纯解剖学工具，并且诊断缺血性心脏病的一个原因。由于 MCG 检测金标无法看见的异常，所以它将始终生成“假阳性”，这就解释了为什么在这种应用中特异性比较低。

[0133] 应注意的是，这里是把核变换作为独立预处理阶段的数据变换加以应用的。数据实际上由非线性数据核替代，并且后来应用传统的线性预测模型。对数据的非线性核变换采用传统线性算法的方法定义了这里所谓的“直接核方法”。这种直接核方法的简洁性和优点在于：问题的非线性方面被记录在所述核中，并且对于所应用的算法而言是透明的。

[0134] 虽然本文论述的核在本质上是高斯型曲线，但是这只是示例性的，而不是限制性的。例如，不作为限制，还可以使用所谓的仿样核，并且将其视为本公开内容的范围之内。

[0135] 虽然已经举例说明并且描述了本发明的某些首选的特征，但是本领域技术人员将

会想出多种修改、改变和替换。因此应当理解，所附权利要求书旨在覆盖属于本发明真实精神之内的所有这些修改和变化。

[0136] 参考文献

[0137] References

[0138] [1]C.-C.Chang and C.-J.Lin, LibSVM, OSU, see http://www.csie.ntu.edu.tw/-cilin/libsvmSVMLib.

[0139] [2]N.Cristianini and J.Shawe-Taylor[2000]Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge University Press.

[0140] [3]I.Daubechies[1992], Ten Lectures on Wavelets, Siam, Philadelphia, PA.

[0141] [4]G.Deboeck and T.Kohonen(Eds.)[1998]Visual Explorations in Finance with Self-Organizing Maps. Springer.

[0142] [5]V.Froelicher, K.Shetler, and E.Ashley[2002] "Better Decisions through Science :Exercise Testing Scores."Progress in Cardiovascular Diseases, Vol.44(5), pp.385-414.

[0143] [6]A.Golbraikh and A.Tropsha[2002]"Beware of q21"Journal of Molecular Graphics and Modelling, Vol 20, pp.269-276.

[0144] [7]R.A.Johnson and D.W.Wichem[2000]Applied Multivariate Statistical Analysis, 2 ed., Prentice Hall.

[0145] [8]R.H.Kewley, and M.J.Embrechts[2000] "Data Strip Mining for the Virtual Design of Pharmaceuticals with Neural Networks," IEEE Transactions on Neural Networks, Vol.11(3), pp.668-679.

[0146] [9]T.Kohonen[1997]Self-Organizing Maps, 2nd Edition, Springer.

[0147] [10]R.Rosipal and L.J.Trejo[2001] "Kernel Partial Least Squares Regression in Reproducing Kernel Hilbert Spaces," Journal of Machine Learning Research, Vol.2, pp.97-128.

[0148] [11]B. Schölkopf and A.J.Smola[2002]Learning with Kernels, MIT Press.

[0149] [12]B. Schölkopf, A.Smola, and K-R Muller[1998]"Nonlinear Component Analysis as a Kemel Eigenvalue Problem," Neural Computuation, Vol.10, 129-1319, 1998.

[0150] [13]J.A.Swets, R.M.Dawes, and J.Monahan[2000] "Better Decisions through Science," Scientific American, pp.2-87.

[0151] [14]The Analyze/StripMiner, the description and the code are available at http:/www.drugmining.com.

[0152] [15]V.Vapnik[1998]Statistical Learning Theory, John Wiley&Sons.

[0153] [16]W.Wu, D.L.Massarat and S.de Jong[1997]"The Kemel PCA Algorithm for Wide Data.Part II :Fast Cross-Validation and Application in Classification of NIR Data," Chemometrics and Intelligent Laboratory Systems, Vol.37, pp.271-280.

[0154] [17]A.E.Hoerl, and R.W.Kennard[1970]"Ridge Regression :Biased Estimation for Non-Orthogonal Problems," Technometrics, Vol.12, pp.9-82.

[0155] [18]J.Principe, N.R.Euliano, and W.C.Lefebre[2000]Neural and Adaptive

Systems :Fundamentcls through Simulations, John Wiley&Sons, Inc.

[0156] [19]W. Wu. D. L. Massarat and S. de Jong[1997] "The Kernel PCA Algorithm for Wide Data. Part I :Theory and Algorithms," Chemometrics and Intelligent Laboratory Systems, Vol. 36, pp. 165-172.

[0157] [20]Vladimir Cherkassky and Filip Mulier[1998]Learning from Data. Concepts, Theory. and Methods, JohnWiley&Sons, Inc.

[0158] [21]S. Haykin[1999]Neural Networks :A Comprehensive Foundation(2nd Ed.), Prentice Hall.

[0159] [22]A. N. Tikhonov[1963] "On Solving Incorrectly Posed Problems and Method of Regularization," Doklady Akademii Nauk USSR, Vol. 151, pp. 501-504.

[0160] [23]A. N. Tikhonov and V. Y. Arsenin[1977]Solutions of ill-Posed Problems, W. H. Winston. Washigton D. C.

[0161] [24]Evgeniou, T., Pontil, and M. Poggio, T. [2000] "Statistical Learning Theory :A Primer," International Journal of Computer Vision, Vol. 38(1), pp. 9-13.

[0162] [25]T. Evgeniou, M. Pontil, and T. Poggio[2000]"Regularization Networks and Support Vector Machines," in Advances in Large Margin Classifiers, MIT Press.

[0163] [26]Poggio, T., and Smale S., [2003]"The Mathematics of Learning :Dealing with Data," To appear in Notices of the AMS, May2003.

[0164] [27]Suykens, J. A. K. and Vandewalle, J. [1999]"Least-Squares Support Vector Machine Classifiiers," Neural Processing letters, Vol. 9(3), pp. 293-300, Vol. 14, pp. 71-84.

[0165] [28]Suykens, J. A. K., van Gestel, T. de Brabanter, J. De Moor, M., and Vandewalle, J. [2003]Least Squares Support Vector Machines, World Scientifiic Pub Co, Singapore.

[0166] [29]Svante Wold, Michael Sjöström, and Lennart Eriksson[2001] "PLS-Regression :a Basic Tool of Chemometrics," Chemometrics and Intelligent Laboratory Systems, 58 :109-130.

[0167] [30] Møller, M. F., [1993] "A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning," Neural Networks, Vol. 6, pp. 525-534.

[0168] [31]H. Ritter, T. Kohonen, "Self-Organizing Semantic Maps," Biological Cybernetics, vol. 61, pp. 241-254, 1989.

[0169] [32]T. Kohonen, Self Organization and Associative Memory, 2nd ed., Springer-Verlag, 1988.

[0170] [33]T. Kohonen, "The Self-Organizing Map," Neurocomputing, 21(1) :1-6, November 1998.

[0171] [34]T. Kohonen, "Thing You Haven't Heard about the Self-Organizing Map," IEEE International Conference on Neural Network, vol. 3, pp. 1147-1156, 1993,

[0172] [35]T. Kohonen, "Generalization of the Self-Organizing Map," International Joint Conference on Neural Networks, vol. 1, pp. 21-26, 1993.

[0173] [36]G.Deboeck and T.Kohonen,Visual Explorations in Finance with Self-Organizing Maps,Springer,2000.

[0174] [37]H.C.Card,G.K.Rosendakl,D.K.Mcneill,and R.D.Mcleod,“Competitive Learning Algorithms and Neurocomputer Architecture,”IEEE Transactions on Computers,vol.47,no.8,pp.847-858,August 1998.

[0175] [38]J.S.Kirk,and J.M.Zurada,“Motivation for Genetically-Trained Topography-Preserving Map,”International Joint Conference on Neural Networks 2002,vol.1,pp.394-399,2002.

[0176] [39]J.Vesanto,J.Himberg,M.Siponen,and A.Ollisimula,“Enhancing SOM Based Dara Visualization,”Proceedings of the International Conference on Soft Computing and Information/Intelligent Systems,pp.64-67,1998.

[0177] [40]G.Stroink,W.Moshage,S.Achenbach:“Cardiomagnetism”.In:Magnetism in Medicine,W. Andrä,H.Nowak,eds.Berlin:Wiley VCH;1998;136-189.

[0178] [41]M. Mäkijärvi, J.Montonen,J.Nenonen:“Clinical application of magnetocardiographic mapping”in:Cardiac mapping M.Shenasa,M.Borgreffe,G.Breithardt,Eds.Mount Kisco,NY:Futura Publishing Co,2003.

[0179] [42]M. Mäkijärvi, K.Brockmeier,U.Leder,et al.:“New trends in clinical magnetocardiography”.In Biomag96:Proc.of the 10th Internat.Conf.on Biomagnetism,Aine C.,et al.,eds.,New York:Springer,2000;410-417.

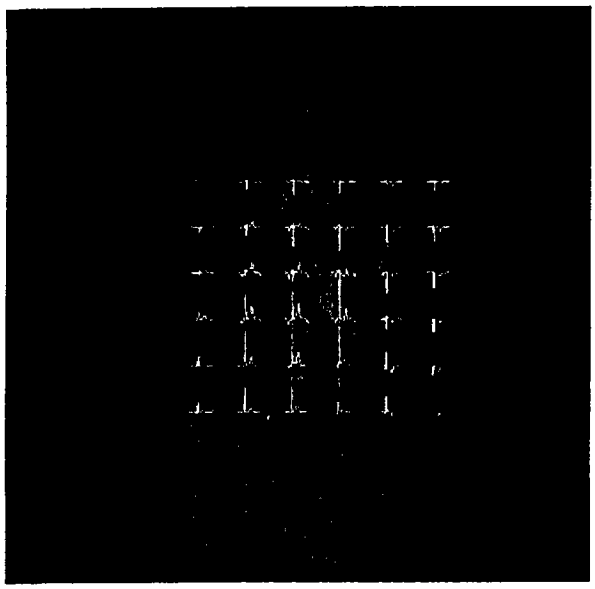


图 1

方法	正确率%	错误率	次数
SVMLib	74	4+5	10
K-PLS	74	4+5	6
DK-PCA	71	7+3	5
PLS	63	2+11	3
K-PLS	**80**	**2+5**	6
DK-PLS	**80**	**2+5**	5
SVMLib	**80**	**2+5**	10
LS-SVM	**80**	**2+5**	**0.5**
SOM	63	3+10	960
DK-SOM	71	5+5	28
DK-SOM	77	3+5	28

图 2

方法	q2	Q2	RMSE
SVMLib	0.767	0.842	0.852
K-PLS	0.779	0.849	0.856
DK-PCA	0.783	0.812	0.87
PLS	0.841	0.142	1.146
K-PLS	0.591	0.694	0.773
DK-PLS	0.591	0.694	0.773
SVMLib	0.591	0.697	0.775
LS-SVM	**0.59**	**0.692**	**0.772**
SOM	0.866	1.304	1.06
DK-SOM	0.855	1.0113	0.934
DK-SOM	0.755	0.859	0.861

图 3

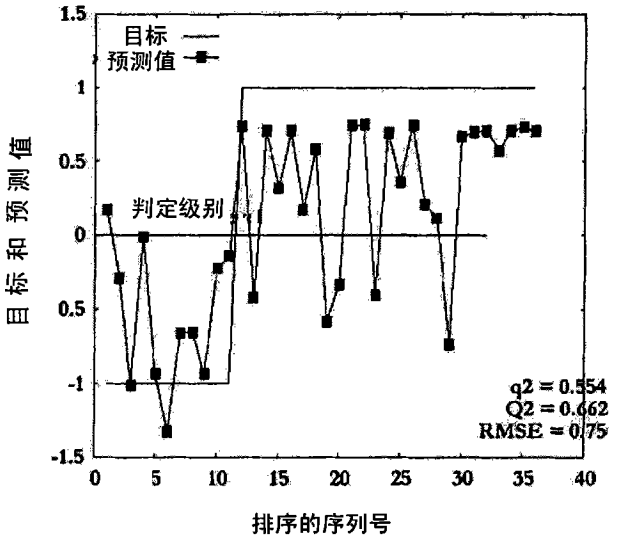


图 4

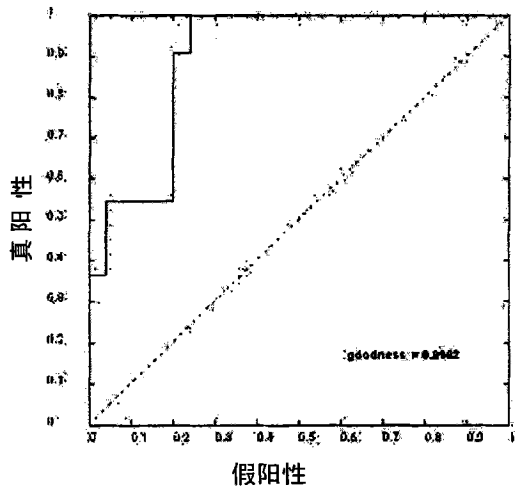


图 5

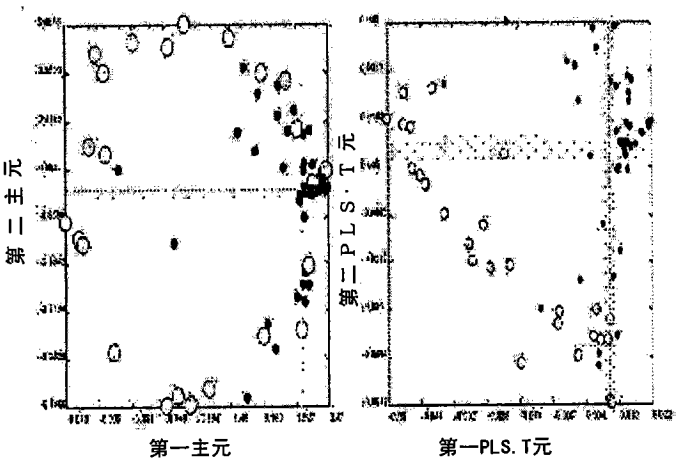


图 6

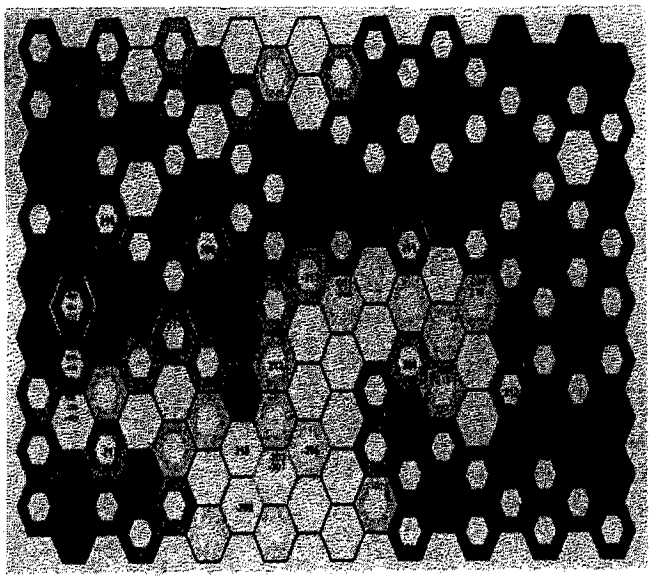


图 7

方法	域	RMSE	CC [%]
SVMLib	时间	0.852	74
K-PLS	时间	0.856	74
DK-PCA	D4－小波	0.87	71
PLS	D4－小波	1.146	63
K-PLS	D4－小波	0.773	80
DK-PLS	D4－小波	**0.75**	**83**
SVMLib	D4－小波	0.775	80
LS-SVM	D4－小波	0.772	80
SOM	D4－小波	1.06	63
DK-SOM	D4－小波	0.861	77

图 8

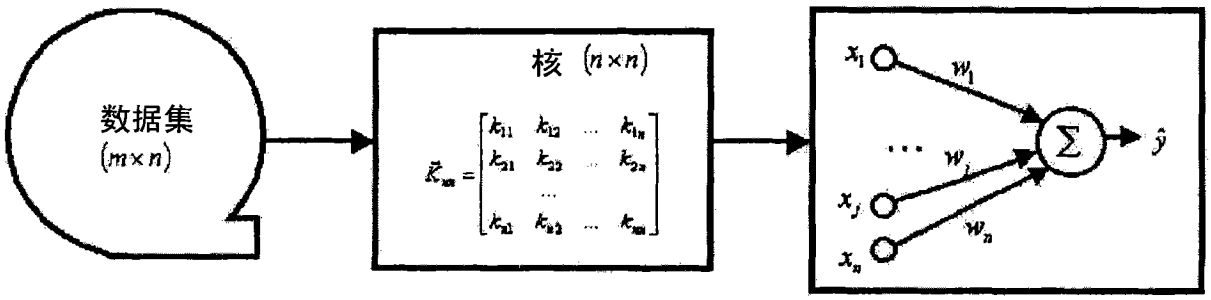


图 9

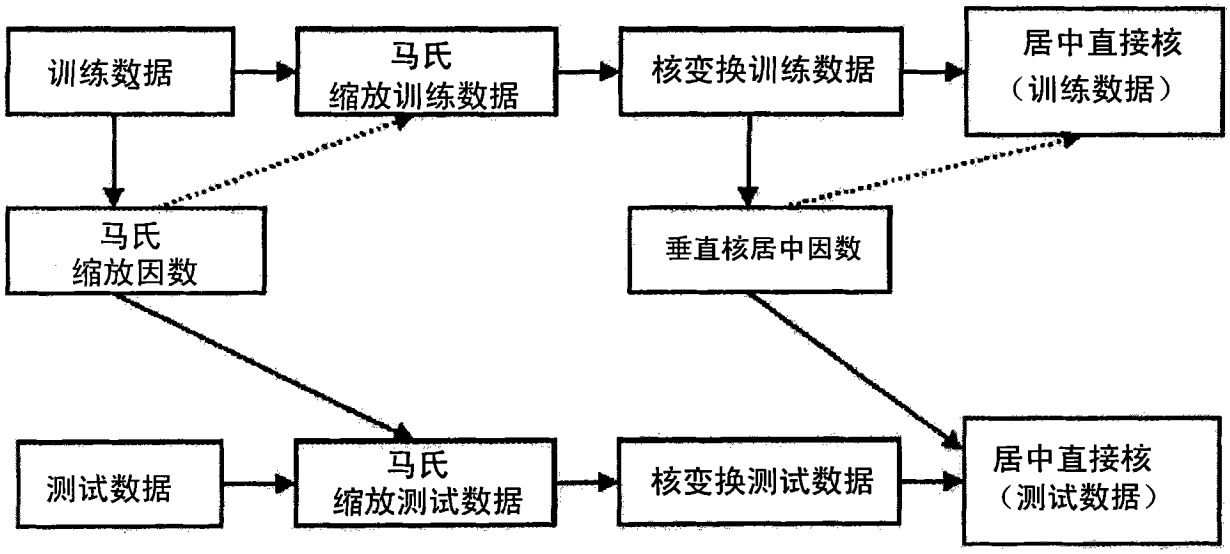

训练数据
马氏
缩放训练数据
核变换训练数据
居中直接核
（训练数据）
马氏
缩放因数
垂直核居中因数
测试数据
马氏
缩放测试数据
核变换测试数据
居中直接核
（测试数据）

图 10

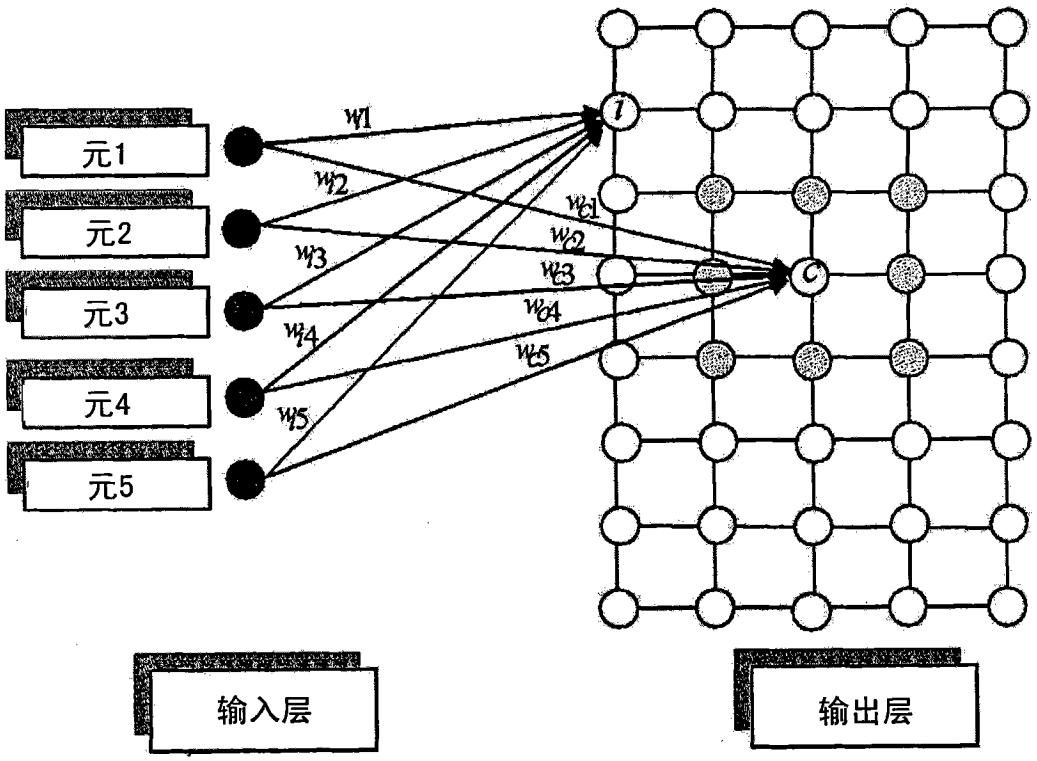

元1
元2
元3
元4
元5
w1
w2
w3
w4
w5
1
C
输入层
输出层

图 11

线性方法	非线性核方法	新开发的方法
PLS	SVMLib	DK-SOM
PCA	SOM	DK-PLS
	LS-SVM	DK-PCA
	K-PLS	

图 12