说明书摘要

本发明公开了一种芯片辅助设计方法，包括：获取原始芯片数据并进行预处理以得到预处理后的芯片数据；基于复式加权结构进行特征提取，即，基于芯片经验对预处理后的芯片数据进行探索性分析、基于统计检验对预处理后的芯片数据进行过滤法分析、基于模型方向对预处理后的芯片数据进行嵌入法分析，以分别得到第一、第二和第三特征化数据；对第一、第二和第三特征化数据进行加权处理并进行分析以得到特征化数据集；选择不同的算法基于特征化数据集建立不同的预测模型，并对预测模型进行训练和评估以得到最优预测模型；基于最优预测模型对芯片数据的数据特征进行分析以得到数据特征的重要性排序，基于重要性排序并且通过最优预测模型进行芯片辅助设计。

摘要附图

获取原始芯片数据，并对原始芯片数据进行预处理以得到预处理后的芯片数据

基于芯片经验对预处理后的芯片数据进行探索性分析以得到第一特征化数据，基于统计检验对预处理后的芯片数据进行过滤法分析以得到第二特征化数据，基于模型方向对预处理后的芯片数据进行嵌入法分析以得到第三特征化数据

对第一特征化数据、第二特征化数据和第三特征化数据进行加权处理并进行分析以得到特征化数据集

选择不同的算法基于特征化数据集建立不同的预测模型，并对预测模型进行训练和评估以得到最优预测模型

基于最优预测模型对芯片数据的数据特征进行分析以得到数据特征的重要性排序，基于所述重要性排序并且通过最优预测模型进行芯片辅助设计

001

002

003

004

005

权利要求书

1. 一种芯片辅助设计方法，其特征在于，包括以下步骤：

获取原始芯片数据，并对所述原始芯片数据进行预处理以得到预处理后的芯片数据；

基于芯片经验对所述预处理后的芯片数据进行探索性分析以得到第一特征化数据，基于统计检验对所述预处理后的芯片数据进行过滤法分析以得到第二特征化数据，基于模型方向对所述预处理后的芯片数据进行嵌入法分析以得到第三特征化数据；

对所述第一特征化数据、所述第二特征化数据和所述第三特征化数据进行加权处理并进行分析以得到特征化数据集；

选择不同的算法基于所述特征化数据集建立不同的预测模型，并对所述预测模型进行训练和评估以得到最优预测模型；以及

基于所述最优预测模型对芯片数据的数据特征进行分析以得到数据特征的重要性排序，基于所述重要性排序并且通过所述最优预测模型进行芯片辅助设计。

2. 根据权利要求1所述的芯片辅助设计方法，其特征在于，获取原始芯片数据包括：

通过EDA工具获取原始芯片数据；或

通过测试数据获取原始芯片数据；或

通过服务器数据获取原始芯片数据。

3. 根据权利要求1所述的芯片辅助设计方法，其特征在于，对所述原始芯片数据进行预处理以得到预处理后的芯片数据包括：

对所述原始芯片数据进行缺失值处理；和/或

对所述原始芯片数据进行重复值处理；和/或

对所述原始芯片数据进行异常值处理；和/或

对所述原始芯片数据进行字符数据编码处理；和/或

对所述原始芯片数据进行无量纲化处理。

4. 根据权利要求1所述的芯片辅助设计方法，其特征在于，基于芯片经验对所述预处理后的芯片数据进行探索性分析以得到第一特征化数据包括：

通过直方图分析离散芯片特征的趋势变化；

通过核密度曲线分析连续芯片特征的趋势变化；

通过计算交叉表分析芯片特征相较于标签的趋势情况；

通过分组箱线图分析搭配特征相较于标签的趋势情况；

通过相关系数和热力图分析特征与特征、特征与标签之间的相关性；

通过主成分分析法分析创造特征。

5. 根据权利要求1所述的芯片辅助设计方法，其特征在于，基于统计检验对所述预处理后的芯片数据进行过滤法分析以得到第二特征化数据包括：

基于方差过滤和/或相关性过滤对所述预处理后的芯片数据进行过滤以得到第二特征化数据。

6. 根据权利要求1所述的芯片辅助设计方法，其特征在于，对所述第一特征化数据、所述第二特征化数据和所述第三特征化数据进行加权处理并进行分析以得到特征化数据集包括：

利用统计分析软件对加权处理后的所述第一特征化数据、所述第二特征化数据和所述第三特征化数据进行分析。

7. 根据权利要求1所述的芯片辅助设计方法，其特征在于，选择不同的算法基于所述特征化数据集建立不同的预测模型包括：

基于所述特征化数据集建立芯片品质预测模型；和/或

基于所述特征化数据集建立芯片性能评分模型。

8. 根据权利要求1所述的芯片辅助设计方法，其特征在于，对所述预测模型进行训练和评估以得到最优预测模型包括：

通过K次交叉验证、学习曲线、网格搜索中的至少一种对所述预测模型进行训练。

9. 根据权利要求7所述的芯片辅助设计方法，其特征在于，对所述预测模型进行训练和评估以得到最优预测模型包括：

通过混淆矩阵、准确率、精确度、召回率、f1分数、roc曲线中的至少一种对所述芯片品质预测模型进行评估。

10. 根据权利要求7所述的芯片辅助设计方法，其特征在于，对所述预测模型进行训练和评估以得到最优预测模型包括：

通过均方误差对所述芯片性能评分模型进行评估。

11. 根据权利要求1所述的芯片辅助设计方法，其特征在于，进一步包括以下步骤：

将所述重要性排序作为反馈信息进行反馈，以对芯片设计进行调整。

12. 根据权利要求11所述的芯片辅助设计方法，其特征在于，还包括：

根据所述反馈信息设计新制芯片，

获取所述新制芯片的芯片数据，并基于所述最优预测模型对所述新制芯片的芯片数据进行分析。

13. 一种芯片辅助设计装置，其特征在于，包括：

第一模块，配置用于获取原始芯片数据，并对所述原始芯片数据进行预处理以得到预处理后的芯片数据；

第二模块，配置用于基于芯片经验对所述预处理后的芯片数据进行探索性分析以得到第一特征化数据，基于统计检验对所述预处理后的芯片数据进行过滤法分析以得到第二特征化数据，基于模型方向对所述预处理后的芯片数据进行嵌入法分析以得到第三特征化数据；

第三模块，配置用于对所述第一特征化数据、所述第二特征化数据和所述第三特征化数据进行加权处理并进行分析以得到特征化数据集；

第四模块，配置用于选择不同的算法基于所述特征化数据集建立不同的预测模型，并对所述预测模型进行训练和评估以得到最优预测模型；以及

第五模块，配置用于基于所述最优预测模型对芯片数据的数据特征进行分析以得到数据特征的重要性排序，基于所述重要性排序并且通过所述最优预测模型进行芯片辅助设计。

14. 一种计算机设备，其特征在于，包括：

至少一个处理器；以及

存储器，所述存储器存储有可在所述处理器上运行的计算机指令，所述指令由所述处理器执行时实现权利要求1-12任意一项所述方法的步骤。

15. 一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现权利要求1-12任意一项所述方法的步骤。

说明书

一种芯片辅助设计方法、装置、设备及可读介质

技术领域

本发明涉及芯片设计技术领域，尤其涉及一种芯片辅助设计方法、装置、设备及可读介质。

背景技术

芯片一般是指集成电路的载体，其设计分为前端设计（逻辑设计）和后端设计（物理设计），前端设计涉及功能设计，后端设计涉及工艺有关的设计，具体包括RTL编写、功能验证、逻辑综合、形式验证、DFT（Design for Testability，可测试性设计）、布局布线、Sign Off、版图验证等多个流程，使其具备制造意义的芯片。

随着人工智能的发展，工业界正在经历着一场变革，朝着更加智能化的发展方向迈进。据统计，每一款芯片设计生成的数据量平均超过500TB，这对于数据科学来说是一笔不菲的财富，特别是开发具有不同功耗、性能和温度的芯片时，具有极高的利用价值。一般来说，芯片设计工程师面临大量的设计数据和可变性，往往只是利用主观经验浅度参考，导致过度设计或欠设计芯片，造成高昂的决策成本。或者说，怎样利用芯片数据，设计出在性能、功耗、面积等方面表现最佳，去实现更高性能的芯片，机器学习技术是极佳的领域。

目前在芯片设计领域，机器学习的主流发展形式是伴随EDA（Electronic design automation，电子设计自动化）工具，EDA厂商将机器学习模块嵌入到EDA工具中，利用内部的芯片设计数据支持训练模型，智能化分析决策，生成高性能的预期芯片，代表厂商有Cadence、Solido等。在晶圆厂中，机器学习技术被应用在芯片的计量和检测中，以查明芯片中的缺陷。然而，该技术的具体实现被禁止访问，其技术逻辑在内部实现，细节设计在机器学习实验室中完成，所以无法给予特定评价，目前存在的问题包括：芯片数据特征的理解处于较低水平，导致芯片数据的有效利用率不足；技术实现角度固化，并不是自动利用数据，而是把实验室训练好的模型安装进EDA，这便会导致功能的灵活度较差，而且约束较多，不利于发展。

发明内容

有鉴于此，本发明实施例的目的在于提出一种芯片辅助设计方法、装置、设备及可读介质，通过特征工程上设计复式加权结构，能够增强对芯片数据特征的理解程度，提高芯片数据的有效利用率，输出更可靠的结果，基于机器学习的芯片辅助设计思路，灵活性更加，约束较少，且成本较低，技术角度能够适用多变的应用场景。

基于上述目的，本发明实施例的一方面提供了一种芯片辅助设计方法，包括以下步骤：获取原始芯片数据，并对所述原始芯片数据进行预处理以得到预处理后的芯片数据；基于芯片经验对所述预处理后的芯片数据进行探索性分析以得到第一特征化数据，基于统计检验对所述预处理后的芯片数据进行过滤法分析以得到第二特征化数据，基于模型方向对所述预处理后的芯片数据进行嵌入法分析以得到第三特征化数据；对所述第一特征化数据、所述第二特征化数据和所述第三特征化数据进行加权处理并进行分析以得到特征化数据集；选择不同的算法基于所述特征化数据集建立不同的预测模型，并对所述预测模型进行训练和评估以得到最优预测模型；以及基于所述最优预测模型对芯片数据的数据特征进行分析以得到数据特征的重要性排序，基于所述重要性排序并且通过所述最优预测模型进行芯片辅助设计。

在一些实施方式中，获取原始芯片数据包括：通过EDA工具获取原始芯片数据；或通过测试数据获取原始芯片数据；或通过服务器数据获取原始芯片数据。

在一些实施方式中，对所述原始芯片数据进行预处理以得到预处理后的芯片数据包括：对所述原始芯片数据进行缺失值处理；和/或对所述原始芯片数据进行重复值处理；和/或对所述原始芯片数据进行异常值处理；和/或对所述原始芯片数据进行字符数据编码处理；和/或对所述原始芯片数据进行无量纲化处理。

在一些实施方式中，基于芯片经验对所述预处理后的芯片数据进行探索性分析以得到第一特征化数据包括：通过直方图分析离散芯片特征的趋势变化；通过核密度曲线分析连续芯片特征的趋势变化；通过计算交叉表分析芯片特征相较于标签的趋势情况；通过分组箱线图分析搭配特征相较于标签的趋势情况；通过相关系数和热力图分析特征与特征、特征与标签之间的相关性；通过主成分分析法分析创造特征。

在一些实施方式中，基于统计检验对所述预处理后的芯片数据进行过滤法分析以得到第二特征化数据包括：基于方差过滤和/或相关性过滤对所述预处理后的芯片数据进行过滤以得到第二特征化数据。

在一些实施方式中，对所述第一特征化数据、所述第二特征化数据和所述第三特征化数据进行加权处理并进行分析以得到特征化数据集包括：利用统计分析软件对加权处理后的所述第一特征化数据、所述第二特征化数据和所述第三特征化数据进行分析。

在一些实施方式中，选择不同的算法基于所述特征化数据集建立不同的预测模型包括：基于所述特征化数据集建立芯片品质预测模型；和/或基于所述特征化数据集建立芯片性能评分模型。

在一些实施方式中，对所述预测模型进行训练和评估以得到最优预测模型包括：通过K次交叉验证、学习曲线、网格搜索中的至少一种对所述预测模型进行训练。

在一些实施方式中，对所述预测模型进行训练和评估以得到最优预测模型包括：通过混淆矩阵、准确率、精确度、召回率、f1分数、roc曲线中的至少一种对所述芯片品质预测模型进行评估。

在一些实施方式中，对所述预测模型进行训练和评估以得到最优预测模型包括：通过均方误差对所述芯片性能评分模型进行评估。

在一些实施方式中，方法进一步包括以下步骤：将所述重要性排序作为反馈信息进行反馈，以对芯片设计进行调整。

在一些实施方式中，方法还包括：获取所述新制芯片的芯片数据，并基于所述最优预测模型对所述新制芯片的芯片数据进行分析。

本发明实施例的另一方面，还提供了一种芯片辅助设计装置，包括：第一模块，配置用于获取原始芯片数据，并对所述原始芯片数据进行预处理以得到预处理后的芯片数据；第二模块，配置用于基于芯片经验对所述预处理后的芯片数据进行探索性分析以得到第一特征化数据，基于统计检验对所述预处理后的芯片数据进行过滤法分析以得到第二特征化数据，基于模型方向对所述预处理后的芯片数据进行嵌入法分析以得到第三特征化数据；第三模块，配置用于对所述第一特征化数据、所述第二特征化数据和所述第三特征化数据进行加权处理并进行分析以得到特征化数据集；第四模块，配置用于选择不同的算法基于所述特征化数据集建立不同的预测模型，并对所述预测模型进行训练和评估以得到最优预测模型；以及第五模块，配置用于基于所述最优预测模型对芯片数据的数据特征进行分析以得到数据特征的重要性排序，基于所述重要性排序并且通过所述最优预测模型进行芯片辅助设计。

本发明实施例的再一方面，还提供了一种计算机设备，包括：至少一个处理器；以及存储器，存储器存储有可在处理器上运行的计算机指令，指令由处理器执行时实现上述方法的步骤。

本发明实施例的再一方面，还提供了一种计算机可读存储介质，计算机可读存储介质存储有被处理器执行时实现如上方法步骤的计算机程序。

本发明至少具有以下有益技术效果：特征工程上设计复式加权结构，能够增强对芯片数据特征的理解程度，提高芯片数据的有效利用率，输出更可靠的结果，此外，基于机器学习的芯片辅助设计思路，灵活性更加，约束较少，且成本较低，技术角度能够适用多变的应用场景。

附图说明

为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本发明的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的实施例。

图1为本发明提供的芯片辅助设计方法的实施例的示意图；

图2为本发明提供的芯片辅助设计装置的实施例的示意图；

图3为本发明提供的计算机设备的实施例的示意图；

图4为本发明提供的计算机可读存储介质的实施例的示意图。

具体实施方式

为使本发明的目的、技术方案和优点更加清楚明白，以下结合具体实施例，并参照附图，对本发明实施例进一步详细说明。

需要说明的是，本发明实施例中所有使用“第一”和“第二”的表述均是为了区分两个相同名称非相同的实体或者非相同的参量，可见“第一”“第二”仅为了表述的方便，不应理解为对本发明实施例的限定，后续实施例对此不再一一说明。

基于上述目的，本发明实施例的第一个方面，提出了芯片辅助设计方法的实施例。图1示出的是本发明提供的芯片辅助设计方法的实施例的示意图。如图1所示，本发明实施例的芯片辅助设计方法包括如下步骤：

001、获取原始芯片数据，并对原始芯片数据进行预处理以得到预处理后的芯片数据；

002、基于芯片经验对预处理后的芯片数据进行探索性分析以得到第一特征化数据，基于统计检验对预处理后的芯片数据进行过滤法分析以得到第二特征化数据，基于模型方向对预处理后的芯片数据进行嵌入法分析以得到第三特征化数据；

003、对第一特征化数据、第二特征化数据和第三特征化数据进行加权处理并进行分析以得到特征化数据集；

004、选择不同的算法基于特征化数据集建立不同的预测模型，并对预测模型进行训练和评估以得到最优预测模型；以及

005、基于最优预测模型对芯片数据的数据特征进行分析以得到数据特征的重要性排序，基于重要性排序并且通过所述最优预测模型进行芯片辅助设计。

在本实施例中，在特征工程中，通过复式加权结构，探索性分析、过滤法、嵌入法，再通过SPSS（Statistical Product Service Solutions，统计产品与服务解决方案）专业统计分析的方式可以解决芯片数据特征理解度过低的困境，提高芯片数据的有效利用率。开放式利用芯片数据更具发展，技术角度分支多，灵活性高，且约束较少，无论是对于芯片的品质预测，或是对于芯片的性能评分，我们都给出了自己的解决方案。基于机器学习的芯片应用不仅仅局限于此，可能涉及到更复杂的应用场景，这取决于项目的需求。

在本发明的一些实施例中，获取原始芯片数据包括：通过EDA工具获取原始芯片数据；或通过测试数据获取原始芯片数据；或通过服务器数据获取原始芯片数据。

在本实施例中，假定芯片数据完整存在于数据库中并标记；芯片数据可来源于EDA工具中，如仿真数据、变化性度量，或来源于芯片的测试数据、或来源于服务器数据等，包括但不限于电流、电压、温度、频率、数据流量、工艺库参数、IP（Internet Protocol，互联网协议）选型等。

收集的数据包括：数据库的结构化数据、文件日志的半结构化数据、音频、视频数据的非结构化数据。其中文件日志通过Flume（一种分布式的海量日志采集、聚合和传输的系统）收集，音频、视频通过Kafka（一种开源流处理平台）消息队列收集。

在本发明的一些实施例中，对原始芯片数据进行预处理以得到预处理后的芯片数据包括：对原始芯片数据进行缺失值处理；和/或对原始芯片数据进行重复值处理；和/或对原始芯片数据进行异常值处理；和/或对原始芯片数据进行字符数据编码处理；和/或对原始芯片数据进行无量纲化处理。

在本实施例中，对原始复杂的芯片数据清洗，使其规范并可以使用数学语言描述，预处理模块涉及缺失值、重复值、异常值、字符数据编码、无量纲化等处理方案。

在本实施例中，对于缺失值处理包括：直接删除；0填充，利用0填充缺失值；均值填充，利用该特征下的均值填充缺失值；预测填充，利用其余特征指标预测缺失值。

在本实施例中，对于重复值处理包括：直接删除去重。

在本实施例中，对于异常值处理包括：既定规则，按照物理规律，删除超异常的items（项）；3σ原则，按照统计学经验，若芯片数据服从正态分布，数值分布在（μ-3σ,μ+3σ)中的概率为0.9974，即 P(|x-μ|>3σ)≤0.003，距离在均值μ的3σ外的芯片数据可认定为异常值(统计上，样本量大于30可假定数据服从正态分布，即使非正态分布，也基本满足3σ原则)。

在本实施例中，对原始芯片数据进行字符数据编码处理包括：直接编码，适用于分类变量；独热编码，适用于名义变量；分箱编码，根据需要，可基于距离分箱、分位数分箱、熵值分箱等；word embedding（词嵌入），利用神经网络编码。

在本实施例中，对原始芯片数据进行无量纲化处理包括：归一化，公式如下：

，一般值域缩放至[0,1]；

标准化公式如下：

。

芯片数据经过预处理之后，可理解转化为高质量数据，有利于提高结果的准确性。

在本发明的一些实施例中，基于芯片经验对预处理后的芯片数据进行探索性分析以得到第一特征化数据包括：通过直方图分析离散芯片特征的趋势变化；通过核密度曲线分析连续芯片特征的趋势变化；通过计算交叉表分析芯片特征相较于标签的趋势情况；通过分组箱线图分析搭配特征相较于标签的趋势情况；通过相关系数和热力图分析特征与特征、特征与标签之间的相关性；通过主成分分析法分析创造特征。

在本实施例中，包括但不限于以下分析方法，可以充分借鉴《统计学基础》：

直方图，分析离散芯片特征的趋势变化；

核密度曲线，分析连续芯片特征的趋势变化；

计算交叉表，分析芯片特征相较于标签的趋势情况；

分组箱线图，分析搭配特征相较于标签的趋势情况；

相关系数和热力图，分析特征与特征，特征与标签之间的相关性；

创造特征，主成分分析法、科学经验等；

芯片设计工程师根据工程经验、探索性分析结果等，选择出主要芯片数据特征，获取特征化数据T1。

在本发明的一些实施例中，基于统计检验对预处理后的芯片数据进行过滤法分析以得到第二特征化数据包括：基于方差过滤和/或相关性过滤对预处理后的芯片数据进行过滤以得到第二特征化数据。

在本发明的一些实施例中，对所述第一特征化数据、所述第二特征化数据和所述第三特征化数据进行加权处理并进行分析以得到特征化数据集包括：利用统计分析软件对加权处理后的所述第一特征化数据、所述第二特征化数据和所述第三特征化数据进行分析。

在本实施例中，过滤法主要考虑方差、相关性等进行特征筛取，可设定阈值控制芯片特征的留余量。具体包括：

方差过率，过滤掉方差低于阈值Threshold的芯片特征，一般认为此类特征无意义；

相关性过率，过滤掉与标签相关性较低的芯片特征，一般认为此类特征对标签影响很小，包括：卡方检验，考虑芯片特征与标签之间的独立性； F检验，考虑芯片特征与标签间的线性关系；互信息法，考虑数芯片征与标签间的线性关系与非线性关系；

获取过滤法后的特征化数据T2。

在本发明的一些实施例中，基于模型方向对预处理后的芯片数据进行嵌入法分析以得到第三特征化数据。利用模型本身挑选芯片特征，工业界中常用的特征选择方式，可以有效提高模型得分。嵌入法对模型有限制条件，一般来讲，模型应具有coef\_属性、feature\_importances\_属性、L1或L2惩罚项等其中之一。

嵌入法将所有芯片特征导入到模型(如随机森林)中，经过训练调优后，利用模型属性feature\_importances\_评估特征的重要性，设定阈值，得到数据特征子集，即获取特征化数据T3。

在本实施例中，将三个角度获取到的数据特征按照出现次数进行统计，并赋予1，2，3的权重，将加权后的数据导入统计分析软件进行分析。在优选的实施例中，将加权后的数据导入SPSS(如果内存不够，可以采样)专业分析软件，做系统的统计分析，如方差齐性检验、相关性分析、主成分分析、因子分析、回归分析、聚类、方差扩大因子法（去除多重共线性）等，得到最终的特征化数据集T。利用SPSS分析的目的是为了得到关于芯片数据的最终的特征化数据集T，最终的特征化数据集T是通过加权芯片的特征化数据T1、T2、T3并经分析得到的，属于数据的再利用，例如以下所述的在相关性分析和主成分分析中使用的情况。

在本实施例中，以相关性分析为例，将高权重的关联特征次数累加，收集排名前n的弱数据特征，加工处理，划到最终集；

在本实施例中，以主成分分析为例，主要将原始高维空间数据投影到低维数据空间，可将权重3的数据特征值域增至3倍，权重2的数据特征值域增值2倍，降维操作时投影会向该类数据特征偏斜，提高关注度；

在本发明的一些实施例中，选择不同的算法基于特征化数据集建立不同的预测模型包括：基于特征化数据集建立芯片品质预测模型；和/或基于特征化数据集建立芯片性能评分模型。对所述预测模型进行训练和评估以得到最优预测模型包括：通过K次交叉验证、学习曲线、网格搜索中的至少一种对所述预测模型进行训练。

在本实施例中，算法建模是模型选择、训练优化的过程。基于特征化数据T建立芯片品质预测模型①、基于特征化数据T建立芯片性能预测模型②；

在本实施例中，芯片品质预测模型①属于分类问题，假定只判断一块芯片的异常与否，即二分类模型，考虑逻辑回归、随机森林、支持向量机、朴素贝叶斯等分类算法；假如采用随机森林，树模型的Bagging集成模型，其优点是可处理任意数据类型，擅长处理高维数据，模型泛化能力强。具体而言，把输入的芯片数据设为x，芯片是否异常设定为目标，异常芯片，正常芯片。二分类模型，f作为随机森林函数，能够把芯片数据x映射到目标中，f的确定是以熵增来确定的。

在本实施例中，芯片性能评分模型②属于回归问题，即对一块芯片的性能给出预测得分，考虑线性回归模型等。假如采用线性回归模型，基本的数学模型，其优点是简单且可解释性强；具体而言，把输入的芯片数据设为x，芯片的性能得分设为目标，的大小代表芯片的性能，越大代表芯片的性能越优，越小代表芯片的性能越差；线性回归模型，g作为线性回归函数，能够把芯片数据x映射到目标中，g确定是对芯片数据的拟合，最小二乘法即可。

芯片品质预测模型①或芯片性能预测模型②，都需要训练过程，可选择K次交叉验证、学习曲线、网格搜索等技巧，提高模型的得分，具体包括：

K次交叉验证，将数据集分成K份，依次将其中K-1份作为训练集，剩下1份作为测试集，重复K次，观察训练结果；

学习曲线，分析模型拟合情况；

网格搜索，通过穷举参数、组合参数进行交叉验证优化模型的一种方法。

在通过上述方法对模型进行训练后，在本发明的一些实施例中，可以通过混淆矩阵、准确率、精确度、召回率、f1分数、roc曲线中的至少一种对芯片品质预测模型进行评估，或者通过均方误差对芯片性能评分模型进行评估。完成评估后，将对模型进行分析，详情将在后续具体实施例中进一步说明。

在完成分析后，在本发明的一些实施例中，方法，进一步包括以下步骤：将所述重要性排序作为反馈信息进行反馈，以对芯片设计进行调整。可以将重要性排序反馈给芯片设计工程师，芯片设计工程师在收到该反馈信息后可以根据不同情况对芯片进行调整。在一个实施例中，芯片设计工程师可以根据所述反馈信息设计新制芯片，获取新制芯片数据，并基于最优预测模型对新制芯片的芯片数据进行分析。例如，当芯片设计工程师利用反馈信息设计了一款新制芯片，可将新制芯片的数据导入到模型中，进行数据仿真验证。

下面根据具体实施例进一步阐述本发明的具体实施方式。针对现有背景技术存在的不足，本发明提供基于机器学习的开放式芯片辅助设计方法，特别地，特征工程上设计复式加权结构，能够增强对芯片数据特征的理解程度，提高芯片数据的有效利用率，输出更可靠的结果。此外，基于机器学习的开放式芯片辅助设计思路，灵活性更佳，约束较少，且成本较低，技术角度能够适用多变的应用场景。具体如下：

基于机器学习的开放式芯片辅助设计方法，包括芯片设计数据、数据预处理模块、特征工程模块、算法建模模块、评估模块、分析等。

芯片数据，假定完整存在于数据库中并标记。数据来源于EDA工具，如仿真数据、变化性度量，或来源于芯片的测试数据、或来源于服务器数据等；

数据预处理模块，对原始复杂的芯片数据清洗，使其规范并可以使用数学语言描述。预处理模块涉及缺失值、重复值、异常值、字符数据编码、无量纲化等处理方案；

特征工程模块，设计复式加权结构创新，抽取、创造具有价值的芯片数据，为后续算法建模提供特征化数据。以三个角度出发，开展数据特征工程，以芯片经验方向做探索性分析获取特征化数据T1，以统计检验方向做过滤法分析获取特征化数据T2，以模型方向做嵌入法分析获取特征化数据T3，将三个角度得到的数据特征按照出现次数统计，并赋予1，2，3的权重，将加权后的数据导入SPSS（如果内存不够，可以采样）专业分析软件，做系统的统计分析，如方差齐性检验、主成分分析、因子分析、回归分析、聚类、方差扩大因子法（去除多重共线性）等，得到完善的特征化数据T；

算法建模模块，基于特征化数据T建立芯片品质预测模型①或基于特征化数据T建立芯片性能预测模型②。芯片品质预测模型①，属于分类问题，假定只判断一块芯片的异常与否，即二分类模型，考虑逻辑回归、随机森林、支持向量机、朴素贝叶斯等分类算法；芯片性能预测模型②，属于回归问题，即对一块芯片的性能给出预测得分，考虑线性回归模型等。无论芯片品质预测模型①或芯片性能预测模型②，都需要训练过程，可选择K次交叉验证、学习曲线、网格搜索等技巧，提高模型的得分；

评估模块，评估芯片数据在不同算法下的预测效果，基于评估结果选择最优的算法。评估的指标取决于模型的种类，对于芯片品质预测模型①，可以采用混淆矩阵(confusion\_matrix)、准确率、精确度、召回率、f1 score、roc曲线(Receiver Operating Characteristic Curve)等衡量模型效果；对于芯片性能预测模型②，可以采用均方误差MSE等衡量模型的效果；具体包括以下分类评价指标：

混淆矩阵(confusion\_matrix)，二分类问题的多维衡量指标体系，1代表异常芯片，0代表正常芯片。在混淆矩阵中，真实值在前，预测值在后，如下表所示：

|  |  |  |  |
| --- | --- | --- | --- |
|  | | 预测值 | |
| 1 | 0 |
| 真实值 | 1 | 11 | 10 |
| 0 | 01 | 00 |

准确率(Accuracy)，所有预测正确的样本占总样本量，具体公式如下：

精确度(Precision)，所有预测为异常类中真正异常类占比，具体公式如下：

召回率(Recall)，所有预测的异常类占总体异常类比，捕捉异常类，具体公式如下：

f1 score，基于Precision和Recall的综合统计量，具体公式如下：

roc曲线，以假正率为横坐标、召回率为纵坐标的一条曲线，一般情况，选择模型的准确率作为评价指标；当面对芯片数据不均衡，会考虑f1 score。

回归评价指标包括均方误差(MSE， Mean Square Error)，具体公式如下：

回归评价指标也可使用均方根误差(RMSE，Root Mean Square Error)、平均绝对误差(MAE，Mean Absolute Error)、判定系数(Coefficient of Determination)等。

分析，通过建立的算法模型探索芯片数据潜在的物理规律。对于芯片品质预测模型①，若采用逻辑回归模型，可以使用系数衡量芯片特征的重要性，即系数的绝对值越大，该芯片特征影响芯片的品质因素越深，如果系数为正，说明该芯片特征具有积极作用。若为负，说明该芯片特征具有消极作用。若采用随机森林模型，可以使用属性feature\_importances\_衡量芯片特征的重要性，其值越大，说明影响因素越大，通过归纳总结可以学到影响芯片品质的特征；对于芯片性能预测模型②，MSE越小意味着模型的准确性越高，该得分可以描述对芯片性能预测的可信度。

当我们分析芯片数据的潜在物理规律后，可以向芯片设计工程师反馈信息，若是品质分类模型，可以根据归纳的芯片特征重要性逐步调整芯片设计细节，如通过改变部分电路结构、工艺参数、布线等，调整该特征或部分特征下的数据，使其由异常芯片转化为正常芯片，提高效率；若是性能预测模型，可以通过数学优化方法，计算芯片输入特征数据使其性能最大化，即芯片的最优设计，便可基于此设计出高性能芯片，即完成基于机器学习的辅助芯片设计。

需要特别指出的是，上述芯片辅助设计方法的各个实施例中的各个步骤均可以相互交叉、替换、增加、删减，因此，这些合理的排列组合变换之于芯片辅助设计方法也应当属于本发明的保护范围，并且不应将本发明的保护范围局限在实施例之上。

基于上述目的，本发明实施例的第二个方面，提出了一种芯片辅助设计装置。图2示出的是本发明提供的芯片辅助设计装置的实施例的示意图。如图2所示，本发明实施例的芯片辅助设计装置包括如下模块：第一模块011，配置用于获取原始芯片数据，并对原始芯片数据进行预处理以得到预处理后的芯片数据；第二模块012，配置用于基于芯片经验对预处理后的芯片数据进行探索性分析以得到第一特征化数据，基于统计检验对预处理后的芯片数据进行过滤法分析以得到第二特征化数据，基于模型方向对预处理后的芯片数据进行嵌入法分析以得到第三特征化数据；第三模块013，配置用于对第一特征化数据、第二特征化数据和第三特征化数据进行加权处理并进行分析以得到特征化数据集；第四模块014，配置用于选择不同的算法基于特征化数据集建立不同的预测模型，并对预测模型进行训练和评估以得到最优预测模型；以及第五模块015，配置用于基于最优预测模型对芯片数据的数据特征进行分析以得到数据特征的重要性排序，基于重要性排序并且通过所述最优预测模型进行芯片辅助设计。

基于上述目的，本发明实施例的第三个方面，提出了一种计算机设备。图3示出的是本发明提供的计算机设备的实施例的示意图。如图3所示，本发明实施例的计算机设备包括如下装置：至少一个处理器021；以及存储器022，存储器022存储有可在处理器上运行的计算机指令023，指令由处理器执行时实现以上方法的步骤。

本发明还提供了一种计算机可读存储介质。图4示出的是本发明提供的计算机可读存储介质的实施例的示意图。如图4所示，计算机可读存储介质031存储有被处理器执行时执行如上方法的计算机程序032。

最后需要说明的是，本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，可以通过计算机程序来指令相关硬件来完成，芯片辅助设计方法的程序可存储于一计算机可读取存储介质中，该程序在执行时，可包括如上述各方法的实施例的流程。其中，程序的存储介质可为磁碟、光盘、只读存储记忆体（ROM）或随机存储记忆体（RAM）等。上述计算机程序的实施例，可以达到与之对应的前述任意方法实施例相同或者相类似的效果。

此外，根据本发明实施例公开的方法还可以被实现为由处理器执行的计算机程序，该计算机程序可以存储在计算机可读存储介质中。在该计算机程序被处理器执行时，执行本发明实施例公开的方法中限定的上述功能。

此外，上述方法步骤以及系统单元也可以利用控制器以及用于存储使得控制器实现上述步骤或单元功能的计算机程序的计算机可读存储介质实现。

本领域技术人员还将明白的是，结合这里的公开所描述的各种示例性逻辑块、模块、电路和算法步骤可以被实现为电子硬件、计算机软件或两者的组合。为了清楚地说明硬件和软件的这种可互换性，已经就各种示意性组件、方块、模块、电路和步骤的功能对其进行了一般性的描述。这种功能是被实现为软件还是被实现为硬件取决于具体应用以及施加给整个系统的设计约束。本领域技术人员可以针对每种具体应用以各种方式来实现的功能，但是这种实现决定不应被解释为导致脱离本发明实施例公开的范围。

在一个或多个示例性设计中，功能可以在硬件、软件、固件或其任意组合中实现。如果在软件中实现，则可以将功能作为一个或多个指令或代码存储在计算机可读介质上或通过计算机可读介质来传送。计算机可读介质包括计算机存储介质和通信介质，该通信介质包括有助于将计算机程序从一个位置传送到另一个位置的任何介质。存储介质可以是能够被通用或专用计算机访问的任何可用介质。作为例子而非限制性的，该计算机可读介质可以包括RAM、ROM、EEPROM、CD-ROM或其它光盘存储设备、磁盘存储设备或其它磁性存储设备，或者是可以用于携带或存储形式为指令或数据结构的所需程序代码并且能够被通用或专用计算机或者通用或专用处理器访问的任何其它介质。此外，任何连接都可以适当地称为计算机可读介质。例如，如果使用同轴线缆、光纤线缆、双绞线、数字用户线路（DSL）或诸如红外线、无线电和微波的无线技术来从网站、服务器或其它远程源发送软件，则上述同轴线缆、光纤线缆、双绞线、DSL或诸如红外线、无线电和微波的无线技术均包括在介质的定义。如这里所使用的，磁盘和光盘包括压缩盘（CD）、激光盘、光盘、数字多功能盘（DVD）、软盘、蓝光盘，其中磁盘通常磁性地再现数据，而光盘利用激光光学地再现数据。上述内容的组合也应当包括在计算机可读介质的范围内。

以上是本发明公开的示例性实施例，但是应当注意，在不背离权利要求限定的本发明实施例公开的范围的前提下，可以进行多种改变和修改。根据这里描述的公开实施例的方法权利要求的功能、步骤和/或动作不需以任何特定顺序执行。此外，尽管本发明实施例公开的元素可以以个体形式描述或要求，但除非明确限制为单数，也可以理解为多个。

应当理解的是，在本文中使用的，除非上下文清楚地支持例外情况，单数形式“一个”旨在也包括复数形式。还应当理解的是，在本文中使用的“和/或”是指包括一个或者一个以上相关联地列出的项目的任意和所有可能组合。

上述本发明实施例公开实施例序号仅仅为了描述，不代表实施例的优劣。

本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也可以通过程序来指令相关的硬件完成，程序可以存储于一种计算机可读存储介质中，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

所属领域的普通技术人员应当理解：以上任何实施例的讨论仅为示例性的，并非旨在暗示本发明实施例公开的范围（包括权利要求）被限于这些例子；在本发明实施例的思路下，以上实施例或者不同实施例中的技术特征之间也可以进行组合，并存在如上的本发明实施例的不同方面的许多其它变化，为了简明它们没有在细节中提供。因此，凡在本发明实施例的精神和原则之内，所做的任何省略、修改、等同替换、改进等，均应包含在本发明实施例的保护范围之内。

说明书附图

获取原始芯片数据，并对原始芯片数据进行预处理以得到预处理后的芯片数据

基于芯片经验对预处理后的芯片数据进行探索性分析以得到第一特征化数据，基于统计检验对预处理后的芯片数据进行过滤法分析以得到第二特征化数据，基于模型方向对预处理后的芯片数据进行嵌入法分析以得到第三特征化数据

对第一特征化数据、第二特征化数据和第三特征化数据进行加权处理并进行分析以得到特征化数据集

选择不同的算法基于特征化数据集建立不同的预测模型，并对预测模型进行训练和评估以得到最优预测模型

基于最优预测模型对芯片数据的数据特征进行分析以得到数据特征的重要性排序，基于所述重要性排序并且通过最优预测模型进行芯片辅助设计

001

002

003

004

005

图1

第一模块

第二模块

第三模块

第四模块

第五模块

011

012

013

014

015

图2

存储器

计算机指令

022

023

处理器

021

图3

可读存储介质

计算机程序

031

032

图4