

逻辑回归模型的可视化解释方法及装置

技术领域

本公开涉及计算机技术领域，尤其涉及一种逻辑回归模型的解释方法及装置。

5 背景技术

随着大数据和人工智能的发展，机器学习技术得到越来越广泛的应用。在机器学习中，可以通过大量的历史样例数据对模型进行训练，从而训练出的模型能够针对新的样例数据给出预测结果，例如，模型可以是基于逻辑回归算法的逻辑回归模型。

- 10 现有技术中，大多会提供一些指标来评价训练出的逻辑回归模型的预测效果，比如 AUC（Area under curve，曲线下面积）、准确度、精确度等。但是，对于模型的细节无法得知，模型缺少解释性，从而不利于用户理解模型。

- 对于逻辑回归模型，目前常见的解释方案是评分卡。在评分卡中，将每一个特征名（例如，职业）下所有特征（工人、技术员、服务员、管理员等）及其对应的权重值以表格或卡片的方式展示。但是这种方式在特征维度很高的情况下，表格或者卡片会非常庞大，不利于用户查看。特别是，在特征维度达到千万维时，是不太可能以这种方式直接输出给用户使用的。评分卡方式解释模型虽然细节暴露得很全面，但是用户难以在整体上把握模型，用户并不能从评分卡中直观快速的得出哪些特征名有较高的区分度。

- 20 另外，用户虽然可对逻辑回归模型的输出参数进行一些统计分析，但是这些分析需要熟练掌握机器学习技术和相关的知识积累才可以完成，并且这种分析偏向于实验研究性质，对于模型解释而言，并没有形成一个体系化的展示方案。

发明内容

- 25 有鉴于此，本公开实施例提供一种逻辑回归模型的可视化解释方法及装置，能够帮助用户直观快速地理解逻辑回归模型。

- 第一方面，本公开实施例提供一种逻辑回归模型的可视化解释方法，包括：接收逻辑回归模型的解释请求；根据所述解释请求，获取逻辑回归模型的模型参数，所述模型参数包括所述逻辑回归模型中的各特征及各特征的权重值；对
30 获取的所述模型参数中的各特征按所属特征名进行聚合；针对每一特征名进行特征统计以获得各特征名的特征统计信息，其中，所述特征统计信息指示同一

特征名下各特征的权重值的分布信息和/或同一特征名下各特征的维度信息；通过图形化界面来展示所述特征名及对应的所述特征统计信息。

第二方面，本公开的实施例还提供一种逻辑回归模型的可视化解释装置，包括：接收单元，用于接收逻辑回归模型的解释请求；获取单元，用于根据所述解释请求，获取逻辑回归模型的模型参数，所述模型参数包括所述逻辑回归模型中的各特征及各特征的权重值；聚合单元，用于对获取的所述模型参数中的各特征按所属特征名进行聚合；统计单元，用于针对每一特征名进行特征统计以获得各特征名的特征统计信息，其中，所述特征统计信息指示同一特征名下各特征的权重值的分布信息和/或同一特征名下各特征的维度信息；展示单元，
5 用于通过图形化界面来展示所述特征名及对应的所述特征统计信息。

第三方面，本公开的实施例还提供一种电子设备，所述电子设备包括：壳体、处理器、存储器、电路板和电源电路，其中，电路板安置在壳体围成的空间内部，处理器和存储器设置在电路板上；电源电路，用于为上述电子设备的各个电路或器件供电；存储器用于存储可执行程序代码；处理器通过读取存储器中存储的可执行程序代码来运行与可执行程序代码对应的程序，用于执行本公开的实施例提供的任一种逻辑回归模型的可视化解释方法。
15

第四方面，本公开的实施例还提供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一个或者多个程序，所述一个或者多个程序可被一个或者多个处理器执行，以实现本公开的实施例提供的任一种逻辑回归模型的可视化解释方法。
20

本公开的实施例提供的逻辑回归模型的可视化解释方法及装置，能够以图形化的界面有效地展示逻辑回归模型中特征名的特征统计信息，从而帮助用户直观快速地了解特征对模型预测的影响，更好地理解逻辑回归模型。

25 附图说明

为了更清楚地说明本公开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公开的一些实施例，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附图。

30 图 1 为本公开的实施例提供的逻辑回归模型的可视化解释方法的一种流程图；

图 2 为本公开的实施例中对特征名及特征统计信息的一种概括信息展示示意图；

图 3 为图 2 所示的概括信息展示示意图中关于某个特征名的详细信息展示示意图；

5 图 4 为本公开的实施例提供的逻辑回归模型的可视化解释装置的一种结构示意图；

图 5 为本公开的实施例提供的电子设备的一种结构示意图。

具体实施方式

10 下面结合附图对本公开实施例进行详细描述。在此需要说明的是，在本公开中出现的“并且/或者”、“和/或”均表示包含三种并列的情况。例如“包括 A 和/或 B”表示如下三种并列的情况：（1）包括 A；（2）包括 B；（3）包括 A 和 B。又例如“执行步骤一并且/或者步骤二”表示如下三种并列的情况：（1）执行步骤一；（2）执行步骤二；（3）执行步骤一和步骤二。

15 应当明确，所描述的实施例仅仅是本公开一部分实施例，而不是全部的实施例。基于本公开中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本公开保护的范围。

第一方面，本公开实施例提供一种逻辑回归模型的可视化解释方法，能够以图形化的界面有效地展示逻辑回归模型中特征名的特征统计信息，从而帮助用户直观快速地了解特征对模型预测的影响，更好地理解逻辑回归模型。

20 本公开的实施例提供的逻辑回归模型的可视化解释方法可由机器学习平台等软件来实现，为此，可在模型被训练出的同时予以可视化解释，即，在模型训练完毕后自动触发对该模型进行可视化解释；或者，也可对事先已经训练好的任意模型进行可视化解释。作为示例，可根据用户在前端浏览器的操作来相应地在后端进行一系列处理，从而根据用户的指示对模型进行可视化解释。作为示例，可由用户在前端浏览器输入关于可解释方式的一系列配置，从而后端自动地根据所述配置来进行操作。本公开的实施例对此不做限定。

图 1 为本公开的实施例提供的逻辑回归模型的可视化解释方法的一种流程图，如图 1 所示，本实施例提供的逻辑回归模型的可视化解释方法可以包括：

30 S11，接收逻辑回归模型的解释请求。

本步骤中，无论是在新的逻辑回归模型被训练出的同时自动触发对该模型

的可视化解释，还是事后对已经训练好的逻辑回归模型进行可视化解释，都会首先接收对逻辑回归模型的解释请求，这里，作为示例，所述解释请求可以是来自用户的手动指示，也可以是由于模型训练完毕而自动触发的请求等等。

S12，根据所述解释请求，获取逻辑回归模型的模型参数，所述模型参数包
5 括所述逻辑回归模型中的各特征及各特征的权重值。

本步骤中，可以根据接收到的解释请求获取需要解释的逻辑回归模型的模型参数。本领域技术人员应知晓，逻辑回归模型可看做一系列特征及其相应的权重，相应地，所述模型参数可包括逻辑回归模型的各个特征以及每个特征在模型中对应的权重值。

10 作为示例，在机器学习模型的训练过程中，特征一般可表示为经过哈希变换后的值，相应地，系统中的特征原文数据可还包括特征不经过哈希变换之前的原始值与特征经过哈希变换之后的哈希值之间的映射关系。为此，在步骤 S12 中，可根据特征原文数据而获得特征的原始值，并进一步获得各特征的权重值。

S13，对获取的所述模型参数中的各特征按所属特征名进行聚合。

15 可选的，特征名可对应于原始数据表的一个字段或几个字段经过特征工程后得到的结果，其旨在描述训练样本某一方面的性状。机器学习样本的特征根据特征处理方法有离散特征和连续特征之分。对于离散特征，每个特征名对应于一组特征，每个特征指示所属特征名的一种取值情况，并对应一个权重值。对于连续特征，特征名下只有一个特征及其对应的权重，且该特征的取值是连
20 续变化的。为了描述方便，下文主要以离散特征的处理为例进行说明。

由于逻辑回归模型的特征规模巨大（例如，常常达到千万级特征维度），为了更直观地展示不同的特征对模型预测结果的不同影响，本步骤中，可以将模型参数中的所有特征分别按照其所属的特征名进行聚合，进一步地，本公开的示例性实施例旨在展示同一特征名下各特征的权重值之间的差异程度，通过
25 这种差异程度来体现相应特征名的分类作用。

例如，在本公开的一个实施例中，逻辑回归模型的模型参数中可包括特征：职业为蓝领工人、职业为技术人员、职业为服务人员、职业为退休人员、职业为管理人员，这些特征可聚合到相应的特征名“职业”之下。又例如，逻辑回归模型的模型参数中可包括特征：年龄为 20~30 岁、年龄为 30~40 岁、年龄为
30 40~50 岁、年龄为 50~60 岁、年龄为 60 以上等，这些特征可聚合到相应的特征名“年龄”之下。

S14, 针对每一特征名进行特征统计以获得各特征名的特征统计信息, 其中, 所述特征统计信息指示同一特征名下各特征的权重值的分布信息和/或同一特征名下各特征的维度信息。

本步骤中, 可对每一个特征名下所包括的特征进行统计, 以便得到逻辑回归模型的各特征名的特征统计信息。其中, 同一特征名下各特征的权重值的分布信息可以指示同一特征名下的各特征的权重值(全部权重值或非零权重值)是如何分布的, 例如权重值主要分布在哪些范围, 各特征的权重值分布的比较分散还是比较集中, 关键取值点位于哪里, 分散的是否均匀等等。同一特征名下各特征的维度信息可以指示每个特征名下的全部特征的维度信息, 或者每个特征名下的某类特征(例如, 权重值不为零的有效特征)的维度信息。维度信息可以让用户了解特征名下的有效特征维度, 从而帮助用户了解该模型上线性能。

S15, 通过图形化界面来展示所述特征名及对应的所述特征统计信息。

本步骤中, 可经由图形化界面来展示各特征名及其特征统计信息。这里, 作为示例, 可经由图表、图形、文字等形式对各特征名及其相关的特征统计信息进行展示。此外, 用户还可通过设置在图形化界面上的控件来进行交互操作, 相应地, 可根据用户的交互操作来进一步调整展示内容。

本公开的实施例提供的逻辑回归模型的可视化解释方法, 能够将逻辑回归模型的模型参数按照各特征所属特征名进行聚合, 并针对每一特征名进行特征统计来获得各特征名的特征统计信息, 然后通过图形化界面来展示特征名及对应的特征统计信息。由于特征统计信息指示出了同一特征名下各特征的权重值的分布信息和/或同一特征名下各特征的维度信息, 因此能够帮助用户直观快速地了解特征对模型预测的影响, 从而更好地理解逻辑回归模型。

可选的, 在步骤 S14 中, 各特征的维度信息可以指示以下项之中的一项或多项: 同一特征名下各权重值非零的特征的绝对数量相对于所述同一特征名下的特征总数的占比、同一特征名下所有特征的维度数或所述维度数相对于所述逻辑回归模型的特征总维度数的占比、同一特征名下各权重值非零的特征的维度数或所述维度数相对于所述逻辑回归模型的所有权重值非零的特征总维度数的占比。

其中, 权重值非零的特征又称为有效特征, 相应的, 同一特征名下各权重值非零的特征的绝对数量相对于所述同一特征名下的特征总数的占比又称为有

效特征占比。例如，如果一个特征名下包括 10 个特征，其中有 3 个特征对应的权重值为 0，其余 7 个特征对应的权重值不为 0，则可展示该特征名的有效特征占比为 70%。

同一特征名下所有特征的维度数或所述维度数相对于所述逻辑回归模型的特征总维度数的占比，即一个特征名下包括多少特征或这些特征在逻辑回归模型所包括的所有特征的总维度数中占据多少比例。例如，如果一个逻辑回归模型的所有特征名下的所有特征的总维度数为 100，其中一个特征名 A 下的所有特征的维度数为 20，则可显示特征名 A 下的所有特征的维度数为 20，或者，可显示所述维度数相对于所述逻辑回归模型的特征总维度数的占比为 20%。

同一特征名下各权重值非零的特征的维度数或所述维度数相对于所述逻辑回归模型的所有权重值非零的特征总维度数的占比，即一个特征名下包括多少权重值非零的特征或这些权重值非零的特征在逻辑回归模型所包括的所有权重值非零的特征的总维度数中占据多少比例。例如，如果一个逻辑回归模型的所有特征名下的所有特征的总维度数为 100，其中权重值非零的特征的总维度数为 80，其中一个特征名 B 下的权重值非零的特征的维度数为 10，则可展示特征名 B 下各权重值非零的特征的维度数为 10，或者，可显示所述维度数相对于所述逻辑回归模型的所有权重值非零的特征总维度数的占比为 $10/80 \times 100\% = 12.5\%$ 。

获得了特征名及其对应的特征统计信息后，在本公开的一个实施例中，可以将所述各特征名及对应的特征统计信息以图表形式进行展示，从而将逻辑回归模型的特征名及其特征统计信息按照某种排列直观地呈现给用户，使得即使特征维度很高，依然不降低用户对模型的理解。而且，作为示例，通过将所有特征名的统计信息在一张图表中展示，也可以让用户更加直观的对模型整体有大致把握。

可选的，将各特征名及对应的特征统计信息以图表形式进行展示可以包括：通过箱线图来分别表示同一特征名下各特征的非零权重值和/或全部权重值的分布信息，其中，所述箱线图包括以下至少一项：最小值、第一四分位数、中位数、第三四分位数和最大值；将所述各特征名及对应的箱线图和维度信息展示在图表中。

下面首先对箱线图的绘制进行简要说明。

可选的，在本公开的一个实施例中，可以将同一特征名下各特征的全部权重值或非零权重值按照权重值的大小进行升序排列或降序排列；以排序后的首

位权重值、末位权重值、和/或预设分位数的权重值为关键点绘制所述特征名的箱线图。其中，预设分位数可以根据需要进行指定。例如，在本公开的一个实施例中，预设分位数可以包括第一四分位数、中位数、以及第三四分位数，则在绘制箱线图时，可以将每个特征名下的非零权重按照从小到大排列，统计权重最小值，1/4 分位数，中位数，3/4 分位数，最大值，以这五个点为基础绘制箱线图。当然在本公开的其他实施例中，这几个分位数也可以替换成其他的分位数，并且预设分位数的数量也可以进一步增加，比如可以增加 10%和 90%分位数。

箱线图中展示了同一特征名下特征的权重值的分布情况，为了方便对分布状况进行观察，可选的，在各个特征名的箱线图之间还可以标出权重值为 0 的参考线，通过观察每个箱线图相对 0 参考线的偏移状况确定该特征名对模型预测偏向于正向影响还是负向影响。

绘制完特征名的箱线图后，即可将所述各特征名及对应的箱线图和维度信息展示在图表中。例如，可按照特征名、箱线图和维度信息的顺序来进行展示。本公开的一个实施例中，展示各特征名及对应的箱线图和维度信息的图表可如图 2 所示。

从图 2 最上方的统计数字可以看出，该逻辑回归模型中，共有特征名 52 个，所有特征名下包括的所有特征的总维度数为 37699046。这两个统计值可以使用户对模型大小有整体把握。图 2 所示的表格分四列，其中，左起第一列为特征名，第二列表示非零权重分布箱线图，第三列表示有效特征占比，即同一特征名下有效特征的占比，第四列表示特征维度相对于总特征维度的占比（其中，这里的特征维度可针对全部特征或非零特征，而百分比数据右侧的数字可指示特征维度的绝对数值），可以看出，第二列展示了特征的分布信息，第三列和第四列分别从不同角度展示了特征的维度信息。为了方便用户对箱线图进行查看和对比，根据本公开的示例性实施例，图表一侧可设置有关于箱线图的缩放条，例如，可在图标下方设置有标有权重值刻度的缩放条，这样在一些特征名的箱线图过于集中的情况下，可以对箱线图进行放大显示。

以倒数第二个特征名 FCTR1 为例进行说明。在箱线图中，图下方的缩放条可以指示权重值范围（例如，-1.5 到 1），与坐标指示相对应，FCTR1 对应的箱线图中，最左侧短竖线对应的是权重最小值，最右侧短竖线对应的是权重最大值，中间的竖线对应的是中位数，箱体框的左边界对应的是 1/4 分位数，箱体框

的右边界对应的是 3/4 分位数，与坐标 0 值对齐的纵轴线是权重为 0 的参考线。

从图 2 的箱线图中可以了解特征名 FCTR1 下特征的大致分布状况，并且由 FCTR1 的箱线图相对 0 参考线的偏移情况可以看出，FCTR1 的大部分特征对模型预测偏向于负向影响，只有一少部分的特征对模型预测偏向于正向影响。

5 在图 2 的第三列展示的维度信息中，分别列出了有效特征占比的具体数值及柱状图。具体的，有效特征占比可以用来描述一个特征名的稠密程度，有效特征占比越低表示该特征名越稀疏，有效特征占比越高表示特征名越稠密。特征名过于稀疏可能会引起模型欠拟合，过于稠密可能会引起过拟合，并且会增大模型尺寸，影响实时预测的资源消耗。由于有效特征占比受逻辑回归中损失函数正则项系数的影响，用户了解有效特征占比就可以更具体地了解损失函数正则项系数对模型的影响。从图 2 中可以看出，特征名 FCTR1 中有效特征占比为 35.71%。

在图 2 的最右侧一列展示的维度信息中，分别列出了每个特征名下的特征维度数量及其在总特征维度数量中的占比。例如，特征名 FCTR1 下的特征的维度为 70，在总维度数 37699046 中，占比近似为 0。这些信息可以从一定程度上帮助用户了解数据分布。

结合每个特征名的非零权重分布的箱线图、有效特征占比、特征维度占比等统计信息，可以帮助用户调整参数优化模型。例如，当权重分布箱线图过于集中，有效特征占比过低，而且该特征名的特征维度很高时，如果计算资源有限，则可以舍弃该特征，减小特征样本尺寸，以节约训练模型的计算资源和时间。

可选的，在将各特征名及对应的箱线图和维度信息展示在图表中时，可以根据不同的指标来对各特征名及相应的特征统计信息进行排序显示。这些指标本身可以是展示的内容，也可以不显性地展示，或者可在与用户之间的交互过程中得以展示。

例如，在本公开的一个实施例中，可以基于同一特征名下各特征的权重值的方差（方差越大，表示对应的特征名区分样本的能力越强，方差越小，则表明该特征名区分样本的能力越差）、标准差或平均偏差的升序或降序，在图表中排列所述各特征名及对应的箱线图和维度信息，在该例子中，方差、标准差或平均偏差本身可显示在图形界面中，也可不显示在图形界面中；也可以基于特征名的生成顺序，在图表中排列所述各特征名及对应的箱线图和维度信息，

也就是说，基于特征工程过程中特征名被生成的原始顺序来进行展示；或者，基于同一特征名下各权重值非零的特征的绝对数量相对于所述同一特征名下的特征总数的占比的升序或降序，在图表中排列所述各特征名及对应的箱线图和维度信息；还可以基于同一特征名下所有特征的维度数或所述维度数相对于所述逻辑回归模型的特征总维度数的占比的升序或降序，在图表中排列所述各特征名及对应的箱线图和维度信息，或者基于同一特征名下各权重值非零的特征的维度数或所述维度数相对于所述逻辑回归模型的所有权重值非零的特征总维度数的占比的升序或降序，在图表中排列所述各特征名及对应的箱线图和维度信息。

需要说明的是，虽然以图 2 为例说明了特征名及其对应的特征统计信息的展示，但本公开不限于此。在本公开的其他实施例中，展示出的特征统计信息也可以只包括分布信息或维度信息，在展示相关信息时，也可以以其他方式或角度进行展示，本公开的实施例对此不做限定。

根据本公开的示例性实施例，在基于特征名的特征统计信息对逻辑回归模型进行可视化解释的基础上，还可以进一步展示出与逻辑回归模型相关的更多细节信息。

举例而言，在本公开的一个实施例中，本公开的实施例提供的逻辑回归模型的可视化解释方法还可以包括：检测鼠标在各特征名中的任一特征名的箱线图上的悬停操作；在检测到悬停操作的情况下，弹出关于所述任一特征名的特征提示框，其中，所述特征提示框中显示有所述任一特征名下的提示特征及其权重值和/或所述任一特征名下各特征的权重值统计值，其中，所述提示特征包括具有以下至少一项权重值的特征：最小值、第一四分位数、中位数、第三四分位数和最大值；所述权重值统计值包括以下至少一项：均值、方差、标准差和平均偏差。

作为示例，所述特征提示框可显示在悬停操作所涉及的箱线图周围。

通过上述方式，既可帮助用户更准确的查看特征名的关键权重值，同时还能使用户了解权重值对应的具体特征，从而有助于用户验证业务直觉或者从中提取业务知识，方便用户进一步理解逻辑回归模型。

在本公开的另一个实施例中，本公开的实施例提供的逻辑回归模型的可视化解释方法还可以包括：接收对各特征名中的任一特征名的详情查看指令；根据接收到的所述详情查看指令，展示所述任一特征名下至少一部分特征及其权

重值。

具体的，当用户想在如图 2 显示的各特征名及其对应的特征统计信息中，查看其中一个特征名的一些具体特征时，可以使用详情查看指令。详情查看指令的具体形式可以多种多样，例如可以用鼠标点击对应的特征名或悬停在所述特征名上等方式来选择期望查看详情的特征。在接收到详情查看指令后，可以通过切换窗口或弹出窗口等方式展示出该特征名下的一部分或全部特征及其权重值等详情信息。

举例而言，当详情查看指令指示需要查看图 2 的第四个特征名“FTotalClicks”时，可以显示出特征名“FTotalClicks”下的一部分或全部特征，具体可如图 3 所示。可选的，特征的具体展示数量可以根据特征数量或用户需要来确定。例如，当该特征名下的特征数量较少时，可以将这些特征全部显示出，当该特征名下的特征数量较多时，可以只选择其中一部分来显示。本实施例中，考虑到有些特征名下的特征维度非常高，为了体现信息展示的优先及高效性，可以展示出权重值排在最前面/最后面的若干特征，从而使用户能够精细地了解到特征对模型预测的贡献。例如，用户可手动选择权重值大小位于前 500 或权重值大小位于后 500 的特征。

如图 3 所示，图 3 的左半部分展示了各特征名及对应的箱线图，右半部分详细展示了被选中的特征名“FTotalClicks”下特征的详细信息。由图 3 可知，“FTotalClicks”下仅包括 13 个特征，由于用户选择展示出权重值前 500 的特征，因此所有 13 个特征按照权重值降序排列进行了展示，其中第一页（即当前页）展示了 10 个特征，剩下的 3 个特征在第二页（未示出）展示。

进一步的，为了便于用户快速获取到所关注的信息，在本公开的一个实施例中，在对各特征名及其对应的特征统计信息进行图形化的展示时，还提供了搜索功能。可选的，该搜索功能可以包括对特征名的搜索和/或对特征的搜索。对于特征名的搜索，本公开的实施例提供的逻辑回归模型的可视化解释方法可以包括：接收特征名搜索指令；根据所述特征名搜索指令在特征名之中搜索目标特征名；以及展示搜索到的目标特征名及其对应的特征统计信息。

具体实施时，用户可以在搜索框中输入需要查找的目标特征名，系统可以在逻辑回归模型的所有特征名中对目标特征名进行搜索，当搜索到目标特征名后，可以将目标特征名及其对应的各种特征统计信息（例如特征的箱线图和维度信息等）通过图表的形式展示出来，进一步地，可显示目标特征名的特征详

情、关键特征、特征权重值统计指标（例如，方差、标准差和平均偏差等）等。

同样的，为了方便用户在特征名所包括的众多特征中查找到某个所关心的特征，本公开的实施例提供的逻辑回归模型的可视化解释方法还可以包括：接收特征搜索指令；根据所述特征搜索指令在特征之中搜索目标特征；以及展示
5 搜索到的目标特征及对应的权重值。

需要说明的是，上述特征名搜索以及特征搜索的搜索策略可以采用模糊搜索或者精确搜索，可以在总体情况展示页面、详细信息展示页面或其他任意显示界面下进行，并不受当前页面展示内容的限制。例如，在对特征名及特征统计信息进行总体情况的展示时，不但可以搜索某个感兴趣的特征名，也可以直接搜索某个感兴趣的特征。
10 接搜索某个感兴趣的特征。

第二方面，本公开的实施例还提供一种逻辑回归模型的可视化解释装置，能够以图形化的界面展示逻辑回归模型中的各特征名及其特征统计信息，从而帮助用户直观快速地了解特征对模型预测的影响，更好地理解逻辑回归模型。

如图 4 所示，本公开的实施例提供的逻辑回归模型的可视化解释装置可以
15 包括：接收单元 31，用于接收逻辑回归模型的解释请求；获取单元 32，用于根据所述解释请求，获取逻辑回归模型的模型参数，所述模型参数包括所述逻辑回归模型中的各特征及各特征的权重值；聚合单元 33，用于对获取的所述模型参数中的各特征按所属特征名进行聚合；统计单元 34，用于针对每一特征名进行特征统计以获得各特征名的特征统计信息，其中，所述特征统计信息指示同
20 一特征名下各特征的权重值的分布信息和/或同一特征名下各特征的维度信息；展示单元 35，用于通过图形化界面来展示所述特征名及对应的所述特征统计信息。

本公开的实施例提供的逻辑回归模型的可视化解释装置，能够将逻辑回归模型的模型参数按照各特征所属特征名进行聚合，并针对每一特征名进行特征
25 统计来获得各特征名的特征统计信息，然后通过图形化界面来展示特征名及对应的特征统计信息。由于特征统计信息指示出了同一特征名下各特征的权重值的分布信息和/或同一特征名下各特征的维度信息，因此能够帮助用户直观快速地了解特征对模型预测的影响，从而更好地理解逻辑回归模型。这里，组成可视化解释装置的单元可以是执行相应程序步骤的功能模块或器件，例如，所述
30 可视化解释装置可通过前端浏览器和后端服务器来进行操作。

在本公开的一个实施例中，所述维度信息指示以下项之中的至少一项：同

一特征名下各权重值非零的特征的绝对数量相对于所述同一特征名下的特征总数的占比、同一特征名下所有特征的维度数或所述维度数相对于所述逻辑回归模型的特征总维度数的占比、同一特征名下各权重值非零的特征的维度数或所述维度数相对于所述逻辑回归模型的所有权重值非零的特征总维度数的占比。

5 在本公开的一个实施例中，展示单元 35 将所述各特征名及对应的特征统计信息以图表形式进行展示。

在本公开的一个实施例中，展示单元 35 可包括：箱线图绘制模块，用于通过箱线图来分别表示同一特征名下各特征的非零权重值和/或全部权重值的分布信息，其中，所述箱线图包括以下至少一项：最小值、第一四分位数、中位数、
10 第三四分位数和最大值；展示模块，用于将所述各特征名及对应的箱线图和维度信息展示在图表中。

在本公开的一个实施例中，所述展示模块：基于特征名的生成顺序，在图表中排列所述各特征名及对应的箱线图和维度信息；或者基于同一特征名下各特征的权重值的方差、标准差或平均偏差的升序或降序，在图表中排列所述各
15 特征名及对应的箱线图和维度信息；或者，基于同一特征名下各权重值非零的特征的绝对数量相对于所述同一特征名下的特征总数的占比的升序或降序，在图表中排列所述各特征名及对应的箱线图和维度信息；或者，基于同一特征名下所有特征的维度数或所述维度数相对于所述逻辑回归模型的特征总维度数的占比的升序或降序，在图表中排列所述各特征名及对应的箱线图和维度信息；
20 或者，基于同一特征名下各权重值非零的特征的维度数或所述维度数相对于所述逻辑回归模型的所有权重值非零的特征总维度数的占比的升序或降序，在图表中排列所述各特征名及对应的箱线图和维度信息。

进一步的，本公开实施例提供的逻辑回归模型的可视化解释装置还可包括：检测单元，用于检测鼠标在各特征名中的任一特征名的箱线图上的悬停操作；
25 弹出单元，用于在检测到悬停操作的情况下，弹出关于所述任一特征名的特征提示框，其中，所述特征提示框中显示有所述任一特征名下的提示特征及其权重值和/或所述任一特征名下各特征的权重值统计值，其中，所述提示特征包括具有以下至少一项权重值的特征：最小值、第一四分位数、中位数、第三四分位数和最大值；所述权重值统计值包括以下至少一项：均值、方差、标准差和
30 平均偏差。

在本公开的一个实施例中，所述图表一侧设置有关于箱线图的缩放条。

进一步的，本公开实施例提供的逻辑回归模型的可视化解释装置还可包括：查看指令接收单元，用于接收对各特征名中的任一特征名的详情查看指令；详情展示单元，用于根据接收到的所述详情查看指令，展示所述任一特征名下至少一部分特征及其权重值。

- 5 进一步的，本公开实施例提供的逻辑回归模型的可视化解释装置还可包括：搜索指令接收单元，用于接收特征名搜索指令；搜索单元，用于根据所述特征名搜索指令在特征名之中搜索目标特征名；以及搜索展示单元，用于展示搜索到的目标特征名及对应的特征统计信息。

- 进一步的，本公开实施例提供的逻辑回归模型的可视化解释装置还可包括：
10 搜索指令接收单元，用于接收特征搜索指令；搜索单元，用于根据所述特征搜索指令在特征之中搜索目标特征名；以及搜索展示单元，用于展示搜索到的目标特征及对应的权重值。

以上结合图 1 到图 3 示出的具体操作可分别由图 4 的可视化解释装置中的各个单元来执行，这里，对于具体操作细节将不再赘述。

- 15 第三方面，相应的，本公开实施例提供一种电子设备，能够以图形化的界面展示逻辑回归模型中的各特征名及其特征统计信息，从而帮助用户直观快速地了解特征对模型预测的影响，更好地理解逻辑回归模型。

- 如图 5 所示，本公开的实施例提供的一种电子设备，可以包括：壳体 41、处理器 42、存储器 43、电路板 44 和电源电路 45，其中，电路板 44 安置在壳体
20 41 围成的空间内部，处理器 42 和存储器 43 设置在电路板 44 上；电源电路 45，用于为上述电子设备的各个电路或器件供电；存储器 43 用于存储可执行程序代码；处理器 42 通过读取存储器 43 中存储的可执行程序代码来运行与可执行程序代码对应的程序，用于执行前述任一实施例所述的逻辑回归模型的可视化解释方法。

- 25 处理器 42 对上述步骤的具体执行过程以及处理器 42 通过运行可执行程序代码来进一步执行的步骤，可以参见前述实施例的描述，在此不再赘述。

该电子设备以多种形式存在，可具有单机或分布式的运算结构，本公开对此不作限制。

- 第四方面，本公开的实施例还提供一种存储指令的计算机可读存储介质，
30 所述计算机可读存储介质存储有指令（也可以是一个或者多个程序），所述指令可被一个或者多个处理器执行，以实现前述实施例提供的任一种逻辑回归模

型的可视化解释方法，因此也能实现相应的技术效果，前文已经进行了详细说明，此处不再赘述。

第五方面，本公开的实施例还提供一种包括至少一个计算装置和至少一个存储指令的存储装置的系统，所述指令在被所述至少一个计算装置运行时，促使所述至少一个计算装置执行逻辑回归模型的可视化解释的以下步骤：接收逻辑回归模型的解释请求；根据所述解释请求，获取逻辑回归模型的模型参数，所述模型参数包括所述逻辑回归模型中的各特征及各特征的权重值；对获取的所述模型参数中的各特征按所属特征名进行聚合；针对每一特征名进行特征统计以获得各特征名的特征统计信息，其中，所述特征统计信息指示同一特征名下各特征的权重值的分布信息、同一特征名下各特征的维度信息、或同一特征名下各特征的权重值的分布信息和同一特征名下各特征的维度信息；通过图形化界面来展示所述特征名及对应的所述特征统计信息。

这里，所述系统可构成单机计算环境或分布式计算环境，其包括至少一个计算装置和至少一个存储装置，这里，作为示例，计算装置可以是通用或专用的计算机、处理器等，可以是单纯利用软件来执行处理的单元，还可以是软硬件相结合的实体。也就是说，计算装置可实现为计算机、处理器、计算单元（或模块）、客户端、主机、代理装置、服务器等。此外，存储装置可以是物理上的存储设备或逻辑上划分出的存储单元，其可与计算装置在操作上进行耦合，或者可例如通过 I/O 端口、网络连接等互相通信。

需要说明的是，在本文中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

本说明书中的各个实施例均采用相关的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。

尤其，对于装置实施例而言，由于其基本相似于方法实施例，所以描述的比较简单，相关之处参见方法实施例的部分说明即可。

为了描述的方便，描述以上装置是以功能分为各种单元/模块分别描述。当然，在实施本公开时可以把各单元/模块的功能在同一个或多个软件和/或硬件中实现。

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，
5 是可以通过计算机程序来指令相关的硬件来完成，所述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，所述的存储介质可为磁碟、光盘、只读存储记忆体（Read-Only Memory, ROM）或随机存储记忆体（Random Access Memory, RAM）等。

以上所述，仅为本公开的具体实施方式，但本公开的保护范围并不局限于
10 此，任何熟悉本技术领域的技术人员在本公开揭露的技术范围内，可轻易想到的变化或替换，都应涵盖在本公开的保护范围之内。因此，本公开的保护范围应以权利要求的保护范围为准。