



(12)发明专利申请

(10)申请公布号 CN 110660485 A
(43)申请公布日 2020.01.07

(21)申请号 201910770957.7

(22)申请日 2019.08.20

(71)申请人 南京医渡云医学技术有限公司
地址 210043 江苏省南京市江北新区扬子
科创中心A栋15层
申请人 南京懿医云大数据科技有限公司

(72)发明人 郎超 王尧 刘水清

(74)专利代理机构 北京嘉科知识产权代理事务
所(特殊普通合伙) 11687
代理人 刘力

(51)Int.Cl.
G16H 50/30(2018.01)
G16H 50/20(2018.01)
G16H 10/20(2018.01)

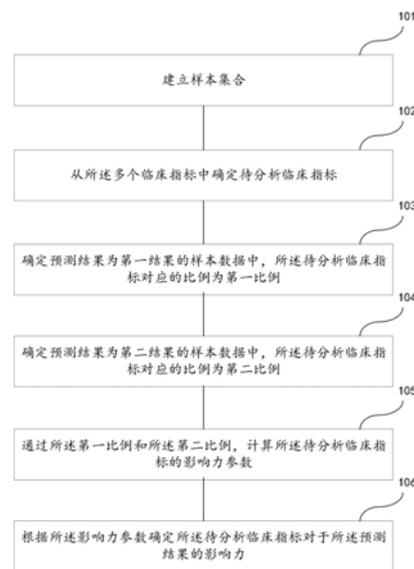
权利要求书2页 说明书7页 附图3页

(54)发明名称

一种获取临床指标的影响力的方法及装置

(57)摘要

本发明公开了一种获取临床指标的影响力的方法及装置,包括:建立样本集合,所述样本集合中包括多个样本数据,所述样本数据包括预测结果和多个临床指标;从所述多个临床指标中确定待分析临床指标;确定预测结果为第一结果的样本数据中,所述待分析临床指标对应的比例为第一比例;确定预测结果为第二结果的样本数据中,所述待分析临床指标对应的比例为第二比例;通过所述第一比例和所述第二比例,计算所述待分析临床指标的影响力参数,并根据所述影响力参数确定所述待分析临床指标对于所述预测结果的影响力;通过统计计算得到影响力参数;从而分析待分析临床指标对于预测结果的关联性和影响力;由此辅助预测模型对于预测结果进行解释。



1. 一种获取临床指标的影响力的方法,其特征在于,包括:
建立样本集合,所述样本集合中包括多个样本数据,所述样本数据包括预测结果和多个临床指标;
从所述多个临床指标中确定待分析临床指标;
确定预测结果为第一结果的样本数据中,所述待分析临床指标对应的比例为第一比例;确定预测结果为第二结果的样本数据中,所述待分析临床指标对应的比例为第二比例;
通过所述第一比例和所述第二比例,计算所述待分析临床指标的影响力参数,并根据所述影响力参数确定所述待分析临床指标对于所述预测结果的影响力。
2. 根据权利要求1所述方法,其特征在于,所述从所述多个临床指标中确定待分析临床指标包括:
将一项特定数值的临床指标确定为所述待分析临床指标。
3. 根据权利要求1所述方法,其特征在于,所述通过所述第一比例和所述第二比例,计算所述待分析临床指标的影响力参数包括:
将所述第一比例除以第二比例的商值作为所述影响力参数。
4. 根据权利要求3所述方法,其特征在于,所述根据所述影响力参数确定所述待分析临床指标对于所述预测结果的影响力包括:
当所述影响力参数大于预设的临界值,则认为所述待分析临床指标与所述预测结果正相关。
5. 根据权利要求3所述方法,其特征在于,所述根据所述影响力参数确定所述待分析临床指标对于所述预测结果的影响力包括:
当所述影响力参数小于预设的临界值,则认为所述待分析临床指标与所述预测结果负相关。
6. 根据权利要求3所述方法,其特征在于,所述根据所述影响力参数确定所述待分析临床指标对于所述预测结果的影响力包括:
当所述影响力参数等于预设的临界值,则认为所述待分析临床指标与所述预测结果无相关性。
7. 根据权利要求1~6任意一项所述方法,其特征在于,还包括:
当所述第一比例或所述第二比例数值为0,则通过所述样本集合中样本数据的正态分布计算所述第一比例或所述第二比例的期望数值;
将所述期望数值作为所述第一比例或所述第二比例的数值。
8. 一种获取临床指标的影响力的装置,其特征在于,包括:
样本集合模型,用于建立样本集合,所述样本集合中包括多个样本数据,所述样本数据包括预测结果和多个临床指标;
指标确定模块,用于从所述多个临床指标中确定待分析临床指标;
比例计算模块,用于确定预测结果为第一结果的样本数据中,所述待分析临床指标对应的比例为第一比例;确定预测结果为第二结果的样本数据中,所述待分析临床指标对应的比例为第二比例;
指标分析模块,用于通过所述第一比例和所述第二比例,计算所述待分析临床指标的影响力参数,并根据所述影响力参数确定所述待分析临床指标对于所述预测结果的影响

力。

9. 一种可读介质,包括执行指令,当电子设备的处理器执行所述执行指令时,所述电子设备执行如权利要求1至7中任一所述的方法。

10. 一种电子设备,包括处理器以及存储有执行指令的存储器,当所述处理器执行所述存储器存储的所述执行指令时,所述处理器执行如权利要求1至7中任一所述的方法。

一种获取临床指标的影响力的方法及装置

技术领域

[0001] 本发明涉及计算机技术领域,尤其涉及一种获取临床指标的影响力的方法及装置。

背景技术

[0002] 在医学上,临床数据的重要性不言而喻。患者的大部分情况都可能通过临床数据直接或间接的体现。一些情况下,多种临床数据结合,可能会隐含的表达出患者的某些病症,或者某些潜在的健康风险。对于这种情况,很难通过人工的数据分析发现。

[0003] 基于当前的人工智能、机器学习、大数据分析等前沿技术建立预测模型,并对临床数据进行分析,能够相比从前更早的发现疾病或潜在患病风险,十分有利于治疗和康复。可以说,预测模型在医学领域中的应用,有着非常重要的医学价值。

[0004] 但是基于某些机器学习算法建立的预测模型,如支持向量机算法(Support Vector Machine,简称SVM)或朴素贝叶斯算法(Naive Bayes)的预测模型,在得到预测结果的同时,不能提供对于预测结果的解释,无法分析输入的各种临床指标对于预测结果的影响力和关联性。

发明内容

[0005] 本发明提供一种获取临床指标的影响力的方法及装置,通过统计计算待分析临床指标与预测结果的比例关系,得到影响力参数,从而实现了对于临床指标影响力的分析。

[0006] 第一方面,本发明提供了一种获取临床指标的影响力的方法,包括:

[0007] 建立样本集合,所述样本集合中包括多个样本数据,所述样本数据包括预测结果和多个临床指标;

[0008] 从所述多个临床指标中确定待分析临床指标;

[0009] 确定预测结果为第一结果的样本数据中,所述待分析临床指标对应的比例为第一比例;确定预测结果为第二结果的样本数据中,所述待分析临床指标对应的比例为第二比例;

[0010] 通过所述第一比例和所述第二比例,计算所述待分析临床指标的影响力参数,并根据所述影响力参数确定所述待分析临床指标对于所述预测结果的影响力。

[0011] 优选地,所述从所述多个临床指标中确定待分析临床指标包括:

[0012] 将一项特定数值的临床指标确定为待分析临床指标。

[0013] 优选地,所述通过所述第一比例和所述第二比例,计算所述待分析临床指标的影响力参数包括:

[0014] 将所述第一比例除以第二比例的商值作为所述影响力参数。

[0015] 优选地,所述根据所述影响力参数确定所述待分析临床指标对于所述预测结果的影响力包括:

[0016] 当所述影响力参数大于预设的临界值,则认为所述待分析临床指标与所述预测结

果正相关。

[0017] 优选地,所述根据所述影响力参数确定所述待分析临床指标对于所述预测结果的影响力包括:

[0018] 当所述影响力参数小于预设的临界值,则认为所述待分析临床指标与所述预测结果负相关。

[0019] 优选地,所述根据所述影响力参数确定所述待分析临床指标对于所述预测结果的影响力包括:

[0020] 当所述影响力参数等于预设的临界值,则认为所述待分析临床指标与所述预测结果无相关性。

[0021] 优选地,还包括:

[0022] 当所述第一比例或所述第二比例数值为0,则通过所述样本集合中样本数据的正态分布计算所述第一比例或所述第二比例的期望数值。

[0023] 第二方面,本发明提供了一种获取临床指标的影响力的装置,包括:

[0024] 样本集合模型,用于建立样本集合,所述样本集合中包括多个样本数据,所述样本数据包括预测结果和多个临床指标;

[0025] 指标确定模块,用于从所述多个临床指标中确定待分析临床指标;

[0026] 比例计算模块,用于确定预测结果为第一结果的样本数据中,所述待分析临床指标对应的比例为第一比例;确定预测结果为第二结果的样本数据中,所述待分析临床指标对应的比例为第二比例;

[0027] 指标分析模块,用于通过所述第一比例和所述第二比例,计算所述待分析临床指标的影响力参数,并根据所述影响力参数确定所述待分析临床指标对于所述预测结果的影响力。

[0028] 第三方面,本发明提供了一种可读介质,包括执行指令,当电子设备的处理器执行所述执行指令时,所述电子设备执行如第一方面中任一所述的方法。

[0029] 第四方面,本发明提供了一种电子设备,包括处理器以及存储有执行指令的存储器,当所述处理器执行所述存储器存储的所述执行指令时,所述处理器执行如第一方面中任一所述的方法。

[0030] 本发明提供了一种获取临床指标的影响力的方法及装置,通过统计计算待分析临床指标与预测结果的比例关系,得到影响力参数;从而通过影响力参数分析待分析临床指标对于预测结果的关联性和影响力;由此能够通过所述影响力分析,辅助某些预测模型对于预测结果进行解释,提升了预测结果的医学分析价值。

[0031] 上述的非惯用的优选方式所具有的进一步效果将在下文中结合具体实施方式加以说明。

附图说明

[0032] 为了更清楚地说明本发明实施例或现有的技术方案,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中记载的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

- [0033] 图1为本发明一实施例提供的一种获取临床指标的影响力的方法的流程示意图；
- [0034] 图2为本发明一实施例提供的另一种获取临床指标的影响力的方法的流程示意图；
- [0035] 图3为本发明一实施例提供的一种获取临床指标的影响力的装置的结构示意图；
- [0036] 图4为本发明一实施例提供的一种电子设备的结构示意图。

具体实施方式

[0037] 为使本发明的目的、技术方案和优点更加清楚，下面将结合具体实施例及相应的附图对本发明的技术方案进行清楚、完整地描述。显然，所描述的实施例仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0038] 前述已知，基于某些机器学习算法建立的预测模型属于黑盒模型。也就是在得到预测结果的同时，不能提供对于预测结果的解释，所以无法分析输入的各种临床指标对于预测结果的影响力和关联性。这导致了预测结果难以有效的用于进一步的医学分析。

[0039] 针对这种情况，本发明提供了一种获取临床指标的影响力的方法及装置，通过统计计算待分析临床指标与预测结果的比例关系，得到影响力参数，从而实现了对于临床指标影响力的分析。

[0040] 参见图1所示，为本发明所述获取临床指标的影响力的方法的具体实施例。本实施例中，所述方法包括以下步骤：

[0041] 步骤101、建立样本集合。

[0042] 本实施例中，将针对样本集合进行临床指标影响力的分析，以确定特定的待分析临床指标对于预测结果的影响力和关联性。所述样本集合中包括多个样本数据。每个样本数据又可以包括一个预测结果和多个临床指标。

[0043] 具体来说，每个样本数据可以是一个患者对应的临床数据，该临床数据中包括血压、血糖、心率、体重、血脂等多项临床指标以及临床指标的数值。并且还包括一个针对该患者的临床数据得到的预测结果。预测结果例如“是否存在心脏病风险”。

[0044] 本实施例中，所述预测结果一般是二元结果。即预测结果只存在“是”或“否”两种情况。在医学上，亦通常称之为“阳性”或“阴性”两种情况。在本实施例中，则可将“阳性”称为第一结果；将“阴性”称为第二结果。

[0045] 步骤102、从所述多个临床指标中确定待分析临床指标。

[0046] 本实施例中，所述待分析临床指标具体为，一项特定数值的临床指标。例如，可确定数值为(150, 100) mmHg的血压指标作为待分析临床指标，并在后续步骤中判断其对于预测结果“是否存在心脏病风险”的影响力。

[0047] 步骤103、确定预测结果为第一结果的样本数据中，所述待分析临床指标对应的比例为第一比例。

[0048] 从样本集合中分类出全部预测结果为“第一结果”，即“存在心脏病风险”的样本数据。并计算这部分样本数据中，血压指标数值为(150, 100) mmHg的样本数据所占的比例，并将这一比例作为第一比例。

[0049] 步骤104、确定预测结果为第二结果的样本数据中，所述待分析临床指标对应的比

例为第二比例。

[0050] 同理步骤103,本步骤中从样本集合中分类出全部预测结果为“第二结果”,即“不存在心脏病风险”的样本数据。并计算这部分样本数据中,血压指标数值为(150,100) mmHg的样本数据所占的比例,并将这一比例作为第二比例。

[0051] 步骤105、通过所述第一比例和所述第二比例,计算所述待分析临床指标的影响力参数。

[0052] 步骤106、根据所述影响力参数确定所述待分析临床指标对于所述预测结果的影响力。

[0053] 本实施例中,具体可以将所述第一比例除以第二比例的商值作为所述影响力参数。所述影响力参数越大,则说明所述待分析临床指标对于预测结果的影响力越大,即越容易导致“心脏病风险”。

[0054] 通过以上技术方案可知,本实施例存在的有益效果是:通过统计计算待分析临床指标与预测结果的比例关系,得到影响力参数;从而通过影响力参数分析待分析临床指标对于预测结果的关联性和影响力;由此能够通过所述影响力分析,辅助某些预测模型对于预测结果进行解释,提升了预测结果的医学分析价值。

[0055] 图1所示仅为本发明所述方法的基础实施例,在其基础上进行一定的优化和拓展,还能够得到所述方法的其他优选实施例。

[0056] 如图2所示,为本发明所述获取临床指标的影响力的方法的另一个具体实施例。本实施例在前述实施例的基础上,对于影响力参数的计算和分析过程进行了更具体的描述和一定程度的优化。为便于解释说明,本实施例中将结合以下具体场景进行阐述。

[0057] 本实施例中,确定所述待分析临床指标为“年龄39岁”,相应的预测结果同样为“是否存在心脏病风险”。

[0058] 当然应该认为,在其他有关的场景下,本实施例所述方法亦同样适用本实施例中所述方法包括以下步骤:

[0059] 步骤201、建立样本集合。

[0060] 本实施例中,一个样本数据可采用如下数学表示:

[0061] $f = (f_1, f_2, f_3, \dots, f_m)$;其中, f 代表样本数据, $f_1 \sim f_m$ 为样本数据中 m 个临床指标。假设本实施例中涉及的临床指标“年龄”为 f_j ,且 $1 \leq j \leq m$ 。另外,本实施例中将预测结果为“第一结果”表示为 $y=1$,将预测结果为“第二结果”表示为 $y=0$ 。

[0062] 步骤202、从所述多个临床指标中确定待分析临床指标。

[0063] 本实施例中临床指标 f_j 的数值 f_{j0} ;即确定 $f_{j0}=39$ 的临床指标 f_j 作为待分析临床指标。

[0064] 步骤203、确定预测结果为第一结果的样本数据中,所述待分析临床指标对应的比例为第一比例。

[0065] 步骤204、确定预测结果为第二结果的样本数据中,所述待分析临床指标对应的比例为第二比例。

[0066] 本实施例中,所述第一比例表示为 $P(f_j = f_{j0} | y = 1)$;所述第二比例表示为 $P(f_j = f_{j0} | y = 0)$ 。

[0067] 另外在一些情况下,样本集合可能不全面。结合本实施例的场景来说,即恰好不存

在“39岁患者”的样本数据。这将导致第一比例或第二比例数值为0。但实际上这是违反医学常识的,因为显然不可能“心脏病”恰好在39岁不发生。

[0068] 所以为解决这种样本错误,本实施例中可以通过所述样本集合中样本数据的正态分布计算所述第一比例或所述第二比例的期望数值。

[0069] 本实施例中,假设所有第一结果样本数据的“年龄”临床指标符合均值为65岁、标准差为10岁的正态分布;所有第二结果样本数据的“年龄”临床指标符合均值为30岁,标准差为10岁的正态分布。

[0070] 则所述第一比例的期望值计算如下:

$$[0071] \quad P(f_{j0} = 39|y = 1) = \Phi\left(\frac{39 - 65}{10}\right)$$

[0072] 所述第二比例的期望值计算如下:

$$[0073] \quad P(f_{j0} = 39|y = 0) = \Phi\left(\frac{39 - 30}{10}\right)$$

[0074] 步骤205、将所述第一比例除以第二比例的商值作为所述影响力参数。

[0075] 本实施例中,所述影响力参数表示如下:

$$[0076] \quad S_j = \frac{P(f_j = f_{j0}|y = 1)}{P(f_j = f_{j0}|y = 0)}$$

[0077] 步骤206、根据所述影响力参数确定所述待分析临床指标对于所述预测结果的影响力。

[0078] 本实施例中,所述根据所述影响力参数确定所述待分析临床指标对于所述预测结果的影响力包括:当所述影响力参数大于预设的临界值,则认为所述待分析临床指标与所述预测结果正相关。当所述影响力参数小于预设的临界值,则认为所述待分析临床指标与所述预测结果负相关。当所述影响力参数等于预设的临界值,则认为所述待分析临床指标与所述预测结果无相关性。

[0079] 具体可设定所述临界值为1。当 $S_j > 1$,则表示“年龄39岁”这一临床指标加强了“心脏病风险”。相反当 $S_j < 1$,则表示“年龄39岁”这一临床指标降低了“心脏病风险”。而当 $S_j = 1$,则表示“年龄39岁”这一临床指标与“心脏病风险”没有明显的相关性。

[0080] 通过以上技术方案可知,本实施例在前述实施例的基础上,进一步存在的有益效果是:详细的公开了影响力参数的计算和分析过程,并且进一步的解决的在样本错误情况下的估算方法;使整体技术方案更加完整。

[0081] 如图3所示,为本发明所述获取临床指标的影响力的装置的一个具体实施例。本实施例所述装置,即用于执行图1~2所述方法的实体装置。其技术方案本质上与上述实施例一致,上述实施例中的相应描述同样适用于本实施例中。本实施例中所述装置包括:

[0082] 样本集合模型301,用于建立样本集合,所述样本集合中包括多个样本数据,所述样本数据包括预测结果和多个临床指标。

[0083] 指标确定模块302,用于从所述多个临床指标中确定待分析临床指标。

[0084] 比例计算模块303,用于确定预测结果为第一结果的样本数据中,所述待分析临床

指标对应的比例为第一比例;确定预测结果为第二结果的样本数据中,所述待分析临床指标对应的比例为第二比例。

[0085] 指标分析模块304,用于通过所述第一比例和所述第二比例,计算所述待分析临床指标的影响力参数,并根据所述影响力参数确定所述待分析临床指标对于所述预测结果的影响力。

[0086] 图4是本发明实施例提供的一种电子设备的结构示意图。在硬件层面,该电子设备包括处理器,可选地还包括内部总线、网络接口、存储器。其中,存储器可能包含内存,例如高速随机存取存储器(Random-Access Memory, RAM),也可能还包括非易失性存储器(non-volatile memory),例如至少1个磁盘存储器等。当然,该电子设备还可能包括其他业务所需要的硬件。

[0087] 处理器、网络接口和存储器可以通过内部总线相互连接,该内部总线可以是ISA (Industry Standard Architecture,工业标准体系结构)总线、PCI (Peripheral Component Interconnect,外设部件互连标准)总线或EISA (Extended Industry Standard Architecture,扩展工业标准结构)总线等。所述总线可以分为地址总线、数据总线、控制总线等。为便于表示,图4中仅用一个双向箭头表示,但并不表示仅有一根总线或一种类型的总线。

[0088] 存储器,用于存放执行指令。具体地,执行指令即可被执行的计算机程序。存储器可以包括内存和非易失性存储器,并向处理器提供执行指令和数据。

[0089] 在一种可能实现的方式中,处理器从非易失性存储器中读取对应的执行指令到内存中然后运行,也可从其它设备上获取相应的执行指令,以在逻辑层面上形成获取临床指标的影响力的装置。处理器执行存储器所存放的执行指令,以通过执行的执行指令实现本发明任一实施例中提供的获取临床指标的影响力的方法。

[0090] 上述如本发明图3所示实施例提供的获取临床指标的影响力的装置执行的方法可以应用于处理器中,或者由处理器实现。处理器可能是一种集成电路芯片,具有信号的处理能力。在实现过程中,上述方法的各步骤可以通过处理器中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器可以是通用处理器,包括中央处理器(Central Processing Unit, CPU)、网络处理器(Network Processor, NP)等;还可以是数字信号处理器(Digital Signal Processor, DSP)、专用集成电路(Application Specific Integrated Circuit, ASIC)、现场可编程门阵列(Field-Programmable Gate Array, FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。可以实现或者执行本发明实施例中的公开的各方法、步骤及逻辑框图。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0091] 结合本发明实施例所公开的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件模块组合执行完成。软件模块可以位于随机存储器,闪存、只读存储器,可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器,处理器读取存储器中的信息,结合其硬件完成上述方法的步骤。

[0092] 本发明实施例还提出了一种可读介质,该可读存储介质存储有执行指令,存储的执行指令被电子设备的处理器执行时,能够使该电子设备执行本发明任一实施例中提供的

获取临床指标的影响力的方法,并具体用于执行如图1或图2所示的方法。

[0093] 前述各个实施例中所述的电子设备可以为计算机。

[0094] 本领域内的技术人员应明白,本发明的实施例可提供为方法或计算机程序产品。因此,本发明可采用完全硬件实施例、完全软件实施例,或软件和硬件相结合的形式。

[0095] 本发明中的各个实施例均采用递进的方式描述,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。

[0096] 还需要说明的是,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、商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、商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、商品或者设备中还存在另外的相同要素。

[0097] 以上所述仅为本发明的实施例而已,并不用于限制本发明。对于本领域技术人员来说,本发明可以有各种更改和变化。凡在本发明的精神和原理之内所作的任何修改、等同替换、改进等,均应包含在本发明的权利要求范围之内。

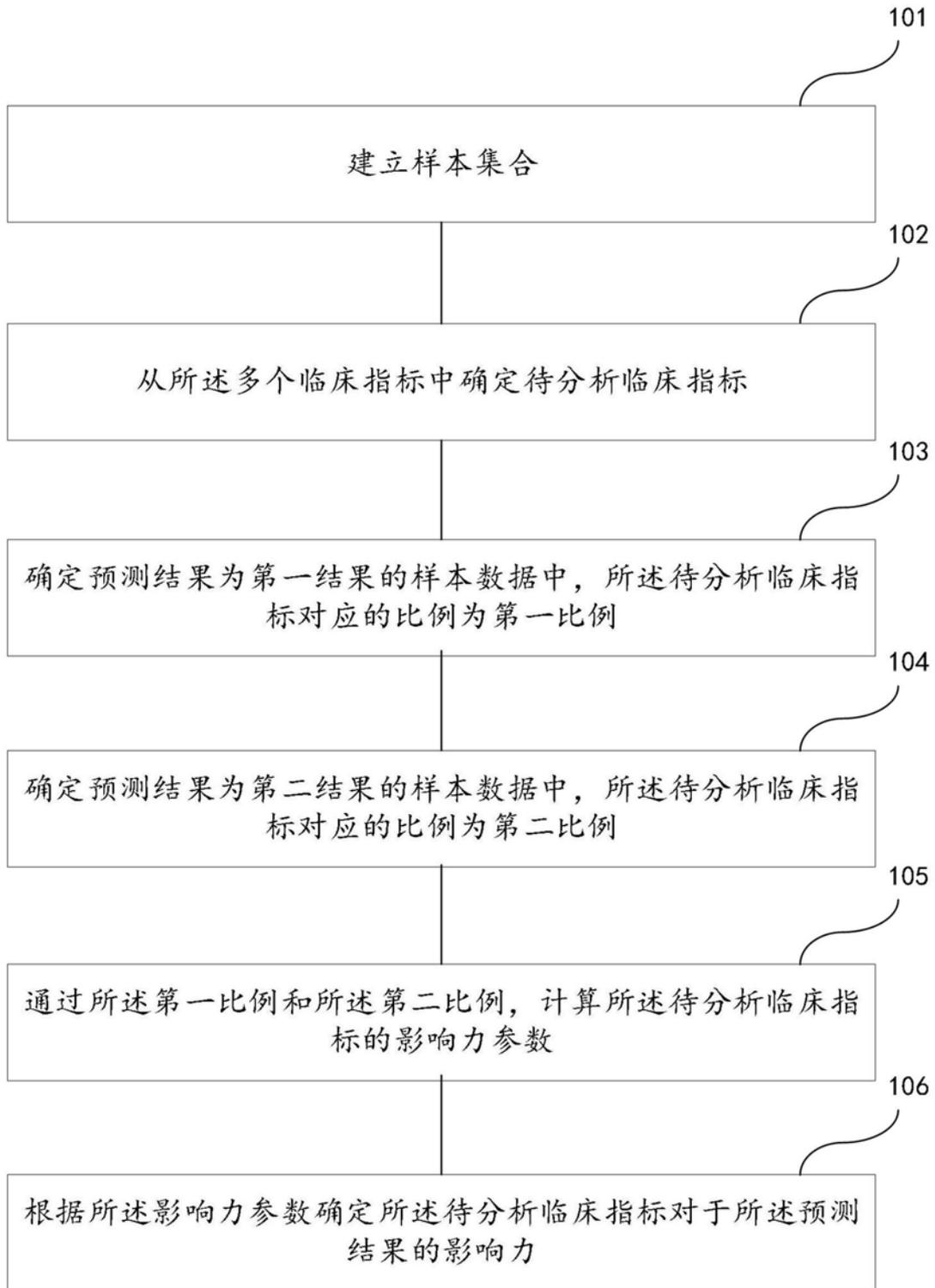


图1

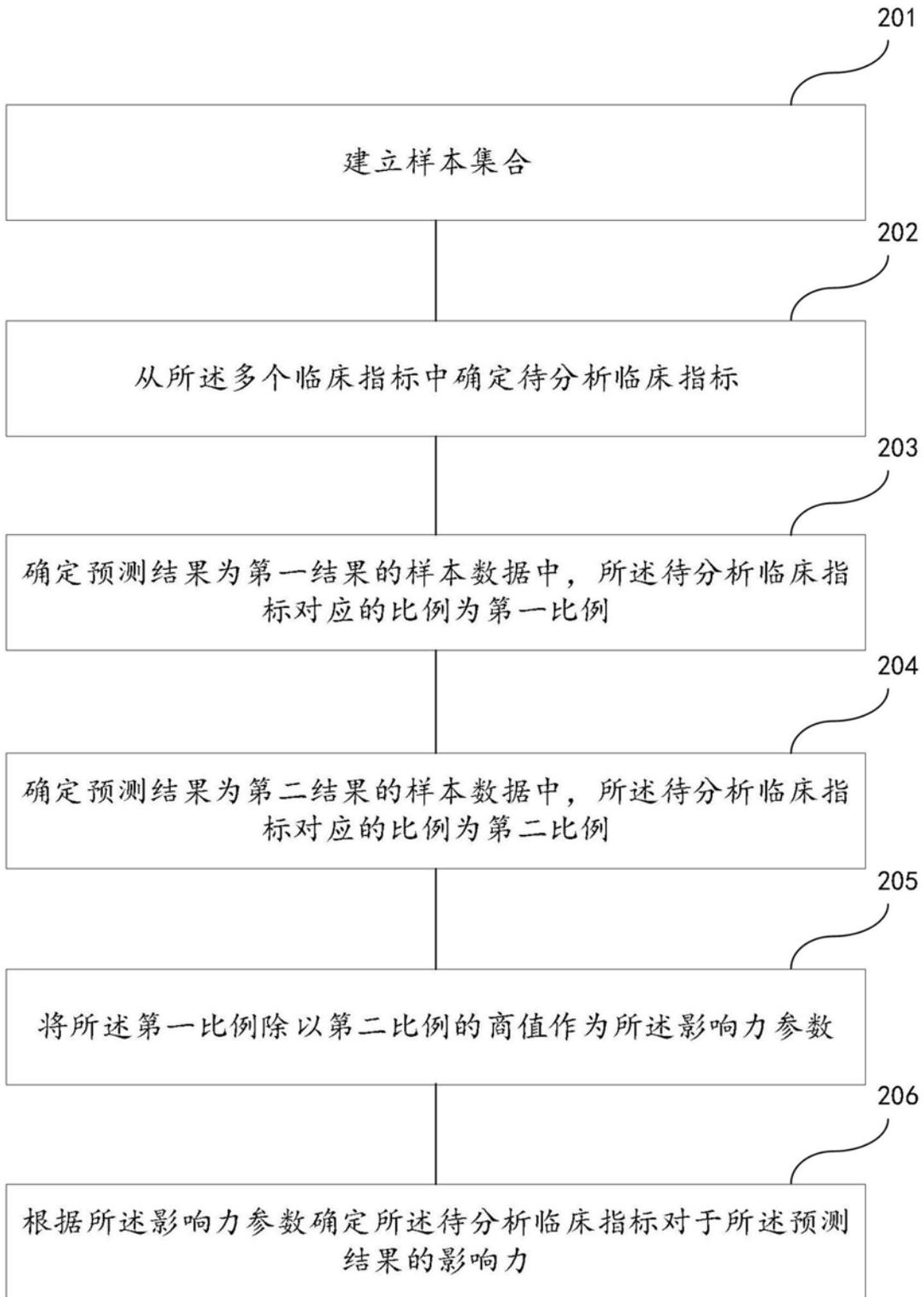


图2

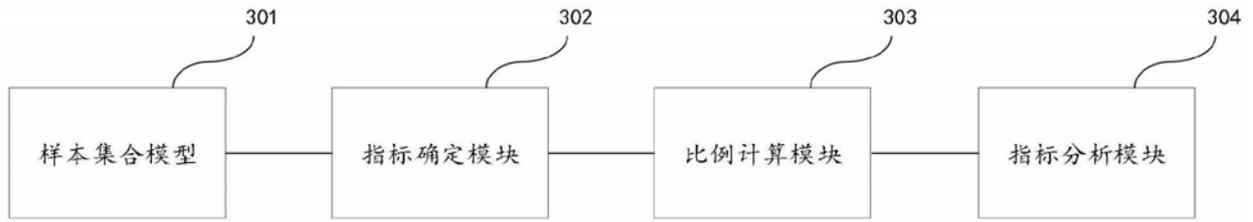


图3

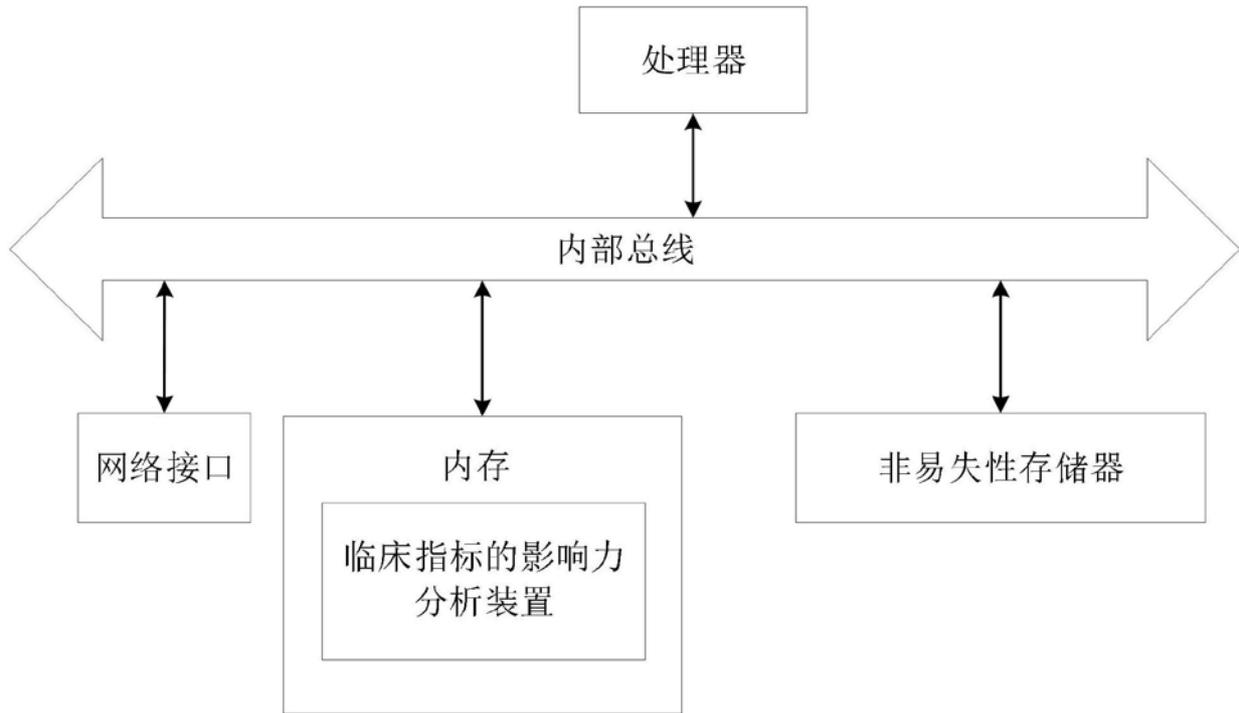


图4