

电子装置、鼻咽癌筛查分析方法和计算机可读存储介质

本申请基于巴黎公约申明享有2018年4月9日递交的申请号为CN 2018103116296、名称为“电子装置、鼻咽癌筛查分析方法和计算机可读存储介
5 质”中国专利申请的优先权，该中国专利申请的整体内容以参考的方式结合在本申请中。

技术领域

本申请涉及大数据分析应用技术领域，特别涉及一种电子装置、鼻咽癌
筛查分析方法和计算机可读存储介质。

10

背景技术

鼻咽癌是一种高发恶性肿瘤，发病率为耳鼻咽喉恶性肿瘤之首，病因复
杂。传统医学上对鼻咽癌的发病原因尚不明确，现有的研究大多来自于临床
观察（与遗传、环境、病毒有关）。而临床观察获取个人信息有限，同时，现
15 有的研究通常依赖于研究人员专业医学知识和个人经验，因此，研究的准确
性无法满足要求，无法形成客观准确的、指标明确的发病原因筛查体系。

因此，如何实现对高危鼻咽癌医学判断对象客观准确的、高效的筛查，
进而实现及早预防或治疗，已经成为一个亟待解决的技术问题。

发明内容

本申请提供一种电子装置、鼻咽癌筛查分析方法，旨在实现客观准确的、
高效的筛查出高危鼻咽癌医学判断对象，从而使客户及早进行预防或治疗。

本申请第一方面提供一种电子装置，包括存储器和处理器，所述存储器
上存储有可在所述处理器上运行的鼻咽癌筛查分析系统，所述鼻咽癌筛查分
25 析系统被所述处理器执行时实现如下步骤：

在收到一个待筛查客户的鼻咽癌筛查请求后，从所述鼻咽癌筛查请求中
获取该待筛查客户的特征标签；

若从所述鼻咽癌筛查请求中获取特征标签失败，则从所述鼻咽癌筛查请
求中获取该待筛查客户的客户属性数据；

30 从多个预先确定的业务服务器提取出与该待筛查客户的客户属性数据对
应的各种特征数据；

按照预设的特征标签提取规则，对提取出的各种特征数据进行特征标签
分析，以分析出该待筛查客户的特征标签；

将该待筛查客户的特征标签代入预先训练好的分析模型中，以分析出该待筛查客户是否是高危鼻咽癌医学判断对象。

优选地，所述特征标签包括基本特征信息、偏好习惯信息、行为信息及/或社会关系信息。

5 本申请第二方面提供一种鼻咽癌筛查分析方法，该方法包括步骤：

在收到一个待筛查客户的鼻咽癌筛查请求后，从所述鼻咽癌筛查请求中获取该待筛查客户的特征标签；

若从所述鼻咽癌筛查请求中获取特征标签失败，则从所述鼻咽癌筛查请求中获取该待筛查客户的客户属性数据；

10 从多个预先确定的业务服务器提取出与该待筛查客户的客户属性数据对应的各种特征数据；

按照预设的特征标签提取规则，对提取出的各种特征数据进行特征标签分析，以分析出该待筛查客户的特征标签；

15 将该待筛查客户的特征标签代入预先训练好的分析模型中，以分析出该待筛查客户是否是高危鼻咽癌医学判断对象。

本申请第三方面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鼻咽癌筛查分析系统，所述鼻咽癌筛查分析系统可被至少一个处理器执行，以使所述至少一个处理器执行如下步骤：

20 在收到一个待筛查客户的鼻咽癌筛查请求后，从所述鼻咽癌筛查请求中获取该待筛查客户的特征标签；

若从所述鼻咽癌筛查请求中获取特征标签失败，则从所述鼻咽癌筛查请求中获取该待筛查客户的客户属性数据；

从多个预先确定的业务服务器提取出与该待筛查客户的客户属性数据对应的各种特征数据；

25 按照预设的特征标签提取规则，对提取出的各种特征数据进行特征标签分析，以分析出该待筛查客户的特征标签；

将该待筛查客户的特征标签代入预先训练好的分析模型中，以分析出该待筛查客户是否是高危鼻咽癌医学判断对象。

30 本申请技术方案，在接收到待筛查客户的鼻咽癌筛查请求后，根据该鼻咽癌筛查请求去获取该待筛查客户的特征标签，具体为先直接从鼻咽癌筛查请求中获取该待筛查客户的特征标签，在获取失败时，则根据该鼻咽癌筛查请求中的客户属性信息，去从各个预先确定的业务服务器中获取该客户属性

信息对应的各种特征数据，再按照预设的提取规则从这些特征数据中提取出该待筛查客户的特征标签；在获得特征标签后，将获得的特征标签输入预先训练好的分析模型中，从而分析得出该待筛查客户是否为高危鼻咽癌医学判断对象。本申请通过将分别代表待筛查客户的各种特征数据的特征标签，输入预先训练好的分析模型中进行分析，以分析出待筛查客户是否为高危鼻咽癌医学判断对象，实现客户准确的、高效的筛查出高危鼻咽癌医学判断对象，让客户可以及早的进行预防或治疗。

附图说明

- 10 图1为本申请鼻咽癌筛查分析方法一实施例的流程示意图；
图2为本申请鼻咽癌筛查分析方法中分析模型的训练流程示意图；
图3为本申请鼻咽癌筛查分析系统一实施例的运行环境示意图；
图4为本申请鼻咽癌筛查分析系统一实施例的程序模块图。

15 具体实施方式

以下结合附图对本申请的原理和特征进行描述，所举实例只用于解释本申请，并非用于限定本申请的范围。

本申请提出一种鼻咽癌筛查分析方法。

如图1所示，图1为本申请鼻咽癌筛查分析方法一实施例的流程示意图。

- 20 本实施例中，该鼻咽癌筛查分析方法包括：

步骤S10，在收到一个待筛查客户的鼻咽癌筛查请求后，从所述鼻咽癌筛查请求中获取该待筛查客户的特征标签；

- 25 鼻咽癌筛查请求中包含待筛查客户的客户属性数据（例如，证件号码，或者，姓名和证件号码）及/或待筛查客户的特征标签；其中，特征标签包括基本特征信息（例如，所属城市属于中国西北地区、所属城市饮食习惯偏辛辣、所属城市PM2.5（细颗粒物）经常超标、所属城市恶劣天气偏多、从事有害气体岗位等）、偏好习惯信息（例如，偏好晚睡、偏好烟酒、偏好烧烤食物、偏好户外运动等）、行为信息（例如，游戏次数多、加班次数多、点外卖的次数多、就医次数多等）及/或社会关系信息（例如，未婚、独居、不经常与朋友联系等）。筛查服务器在收到一个待筛查客户的鼻咽癌筛查请求后，先看该请求中是否包含了该客户的特征标签，即从中去获取该客户的特征标签。

30 步骤S20，若从所述鼻咽癌筛查请求中获取特征标签失败，则从所述鼻咽癌筛查请求中获取该待筛查客户的客户属性数据；

如果该鼻咽癌筛查请求中没有包含该客户的特征标签，则从该鼻咽癌筛查请求中获取不到特征标签（即获取特征标签失败），那么，则从该鼻咽癌筛查请求中获取该待筛查客户的客户属性数据，以根据该带筛查客户的客户属性数据去获得该待筛查客户的特征标签。

- 5 步骤 S30，从多个预先确定的业务服务器提取出与该待筛查客户的客户属性数据对应的各种特征数据；

10 筛查服务器与多个预先确定的业务服务器（例如，银行服务器、医疗服务器、保险服务器、即时通讯服务器、游戏服务器、天气服务器、外卖服务器及/或简历服务器等）之间通讯；在获取到该待筛查客户的客户属性数据后，
15 筛查服务器从多个预先确定的业务服务器提取出与该待筛查客户的客户属性数据对应的各种特征数据（例如，银行贷款额度及还款情况信息、门诊病历信息“例如，预设时间内的看病次数、所患的疾病种类、每次患病的持续时间等”、保险信息“例如，所处行业，性别、年龄、婚姻状况、职业等”、即时通讯工具账号的使用信息“例如，通讯工具每天登陆时间信息、每天在线时长等信息”等等、游戏信息“例如，每天游戏登陆时间信息、每天游戏在线时长等信息”、天气信息“例如，最近三年内，PM2.5（细颗粒物）严重超标的天数”、外卖点餐信息“例如，每天点外卖的时间信息、每天所点外卖的外卖类型等”、求职简历上填写的信息“例如，兴趣爱好、性格、工作经历等信息”）。

- 20 步骤 S40，按照预设的特征标签提取规则，对提取出的各种特征数据进行特征标签分析，以分析出该待筛查客户的特征标签；

 筛查服务器中预先设置了特征标签的提取规则，在提取出该待筛查客户对应的各种特征数据后，利用该提取规则，对提取出的各种特征数据进行特征标签分析，从而分析得到该待筛查客户的特征标签。

- 25 步骤 S50，将该待筛查客户的特征标签代入预先训练好的分析模型中，以分析出该待筛查客户是否是高危鼻咽癌医学判断对象。

 筛查服务器中具有预先训练好的分析模型，即该分析模型是经过了大量已知客户数据（包括客户是否患鼻咽癌与客户对应的特征标签）的训练；筛查服务器在得到了当前待筛查客户的特征标签后，将得到的特征标签输入到
30 该预先训练好的分析模型中，分析模型根据输入的各个特征标签，分析出该待筛查客户是否为高危鼻咽癌医学判断对象。

 本实施例技术方案，在接收到待筛查客户的鼻咽癌筛查请求后，根据该鼻咽癌筛查请求去获取该待筛查客户的特征标签，具体为先直接从鼻咽癌筛

查请求中获取该待筛查客户的特征标签，在获取失败时，则根据该鼻咽癌筛查请求中的客户属性信息，去从各个预先确定的业务服务器中获取该客户属性信息对应的各种特征数据，再按照预设的提取规则从这些特征数据中提取出该待筛查客户的特征标签；在获得特征标签后，将获得的特征标签输入预先训练好的分析模型中，从而分析得出该待筛查客户是否为高危鼻咽癌医学判断对象。本实施例通过将分别代表待筛查客户的各种特征数据的特征标签，输入预先训练好的分析模型中进行分析，以分析出待筛查客户是否为高危鼻咽癌医学判断对象，实现客户准确的、高效的筛查出高危鼻咽癌医学判断对象，让客户可以及早的进行预防或治疗。

10

如图 2 所示，图 2 为本申请鼻咽癌筛查分析方法中分析模型的训练流程示意图。

本实施例中，所述分析模型为多元线性回归模型，所述分析模型的训练过程包括：

15 步骤 E1，根据预先确定的客户属性数据，从多个预先确定的业务服务器提取出第一预设数量的客户的各种特征数据；

客户属性数据例如为证件号码，或姓名和证件号码；筛查服务器与多个预先确定的业务服务器（例如，银行服务器、医疗服务器、保险服务器、即时通讯服务器、游戏服务器、天气服务器、外卖服务器及/或简历服务器等）之间通讯，各个预先确定的业务服务器中具有客户属性数据与特征数据的映射关系；筛查服务器先选取第一预设数量（例如 50 万）的客户，然后，对于每一个选取的客户，分别根据该客户的客户属性数据去预先确定的各个业务服务器中，分别提取出其客户属性数据对应的特征数据，从而得到选取的第一预设数量的客户的各种特征数据。

25 步骤 E2，按照预设的特征标签提取规则，对提取的各个客户的各种特征数据进行分析，以确定各个客户的特征标签；

筛查服务器中预先设置了特征标签的提取规则，在提取出第一预设数量的客户的各种特征数据后，根据该预先设置的提取规则，分别对每个客户的各种特征数据进行特征标签分析，最终分析得到每一个客户的特征标签。

30 步骤 E3，根据预先确定的鼻咽癌与客户属性数据的映射关系，确定所述第一预设数量的客户中患有鼻咽癌的异常客户，确定各个异常客户对应的特征标签，将各个正常客户对应的特征标签和各个异常客户对应的特征标签作为预设模型的训练样本；

筛查服务器的客户数据库中记录了每个客户的患病记录，筛查服务器根据客户的客户属性数据通过查找客户数据库就可确定该第一预设数量的客户中，哪些为患有鼻咽癌的异常客户，哪些为未患鼻咽癌的正常客户；在确定第一预设数量的客户中的异常客户和正常客户后，将每一个正常客户对应的特征标签作为一个训练样本，每一个异常客户对应的特征标签也作为一个训练样本。

步骤 E4，将所述训练样本分为第一百分比的训练集和第二百分比的验证集，所述第一百分比和第二百分比之和小于或者等于 100%；

将所有的训练样本的第一百分比作为训练集，以及所有的训练样本的第二百分比作为验证集，例如，第一百分比为 65%，第二百分比为 35%。当然，训练集和验证集中均包括部分正常客户对应的训练样本和部分异常客户对应的训练样本。

步骤 E5，利用训练集中的各个正常客户的特征标签和各个异常客户的特征标签对所述分析模型进行训练，并在训练完成后利用验证集中的各个正常客户的特征标签和各个异常客户的特征标签对训练的所述分析模型的准确率进行验证；

先利用训练集训练分析模型，在训练完成后，再用验证集验证分析模型的分析结果的准确率。

步骤 E6，若准确率大于预设阈值，则模型训练结束；

筛查服务器中预先设置了准确率阈值（即所述预设阈值，例如 98%），如果根据验证集确定分析模型的分析结果的准确率大于该准确率阈值，则说明分析模型的训练达到了预期的标准，筛查服务器则确认分析模型的训练结束。

步骤 E7，若准确率小于或者等于预设阈值，则执行上述步骤 E1、E2、E3 以增加训练样本的数量，并基于增加后的训练样本重新执行上述步骤 E4 和 E5。

如果根据验证集确定分析模型的分析结果的准确率小于或等于该准确率阈值，则说明分析模型的训练还没有达到预期的标准，筛查服务器此时，则继续执行步骤 E1、E2、E3，并增加第一预设数量的值（例如每次增大 5 万），并基于增加后的训练样本重新执行上述步骤 E4 和 E5，直至达到步骤 E6 的要求。

本申请鼻咽癌筛查分析方法中所述预设的特征标签提取规则为：

对于连续数值的各种特征数据种类设置对应的标签阈值；

例如，所属城市 PM2.5（细颗粒物）最近三年每年超标的天数对应的标签阈值可以为 60 天，当所属城市 PM2.5 最近三年每年超标的天数大于对应的标签阈值“60 天”时，代表所属城市 PM2.5 经常超标；所属城市最近三年每年蓝色以上预警等级的天数对应的标签阈值可以为 55 天，当所属城市最近三年每年蓝色以上预警等级的天数大于对应的标签阈值“55 天”时，代表所属城市恶劣天气偏多；最近一年晚于 23:00 睡觉的天数对应的标签阈值可以是 100 天，当最近一年晚于 23:00 睡觉的天数大于对应的标签阈值“100 天”时，代表偏好晚睡；最近一年点烧烤类的外卖次数对应的标签阈值可以是 80 次，当最近一年点烧烤类的外卖次数大于对应的标签阈值“80 次”时，代表偏好烧烤食物；最近一年就医次数对应的标签阈值可以是 30 次，当最近一年就医次数大于对应的标签阈值“30 次”时，代表就医次数多。

对于非为连续数值的各种特征数据种类设置对应的标签范围；

例如，所属城市对应的标签范围包括：西北地区城市集合、华北地区城市集合、华中地区城市集合、华南地区城市集合等等，当客户所属城市属于所述西北地区城市集合中的城市时，代表该客户所属城市对应的标签信息是“属于中国西北地区”；所属城市饮食习惯对应的标签范围包括：偏辛辣城市集合、偏油腻城市集合、偏清淡城市集合、偏甜/咸城市集合等，当客户所属城市属于所述偏辛辣城市集合中的城市时，代表该客户所属城市对应的标签信息是“饮食习惯偏辛辣”；从事岗位对应的标签范围包括：有害气体岗位集合、非易产生挥发性有害气体的岗位集合、易产生挥发性有害气体的岗位集合、无害气体岗位集合等，当客户从事岗位属于有害气体岗位集合中的岗位，代表该客户所从事行业对应的标签信息是“从事有害气体岗位”。

根据连续数值的各种特征数据种类与标签阈值的映射关系，确定出各个客户的各种连续数值的特征数据对应的标签信息，及根据非连续数值的各种特征数据种类与标签范围的映射关系，确定出各个客户的各种非连续数值的特征数据对应的标签信息。

此外，本申请还提出一种鼻咽癌筛查分析系统。

请参阅图 3，是本申请鼻咽癌筛查分析系统 10 较佳实施例的运行环境示意图。

在本实施例中，鼻咽癌筛查分析系统 10 安装并运行于电子装置 1 中。电子装置 1 可以是桌上型计算机、笔记本、掌上电脑及服务器等计算设备。该电子装置 1 可包括，但不仅限于，存储器 11、处理器 12 及显示器 13。图 3

仅示出了具有组件 11-13 的电子装置 1，但是应理解的是，并不要求实施所有示出的组件，可以替代的实施更多或者更少的组件。

存储器 11 在一些实施例中可以是电子装置 1 的内部存储单元，例如该电子装置 1 的硬盘或内存。存储器 11 在另一些实施例中也可以是电子装置 1 的外部存储设备，例如电子装置 1 上配备的插接式硬盘，智能存储卡（Smart Media Card, SMC），安全数字（Secure Digital, SD）卡，闪存卡（Flash Card）等。进一步地，存储器 11 还可以既包括电子装置 1 的内部存储单元也包括外部存储设备。存储器 11 用于存储安装于电子装置 1 的应用软件及各类数据，例如鼻咽癌筛查分析系统 10 的程序代码等。存储器 11 还可以用于暂时地存储已经输出或者将要输出的数据。

处理器 12 在一些实施例中可以是一中央处理器（Central Processing Unit, CPU），微处理器或其他数据处理芯片，用于运行存储器 11 中存储的程序代码或处理数据，例如执行鼻咽癌筛查分析系统 10 等。

显示器 13 在一些实施例中可以是 LED 显示器、液晶显示器、触控式液晶显示器以及 OLED（Organic Light-Emitting Diode，有机发光二极管）触摸器等。显示器 13 用于显示在电子装置 1 中处理的信息以及用于显示可视化的用户界面。电子装置 1 的部件 11-13 通过系统总线相互通信。

请参阅图 4，是本申请鼻咽癌筛查分析系统 10 较佳实施例的程序模块图。在本实施例中，鼻咽癌筛查分析系统 10 可以被分割成一个或多个模块，一个或者多个模块被存储于存储器 11 中，并由一个或多个处理器（本实施例为处理器 12）所执行，以完成本申请。例如，在图 4 中，鼻咽癌筛查分析系统 10 可以被分割成第一获取模块 101、第二获取模块 102、提取模块 103、第一分析模块 104 及第二分析模块 105。本申请所称的模块是指能够完成特定功能的一系列计算机程序指令段，比程序更适合于描述鼻咽癌筛查分析系统 10 在电子装置 1 中的执行过程，其中：

第一获取模块 101，用于在收到一个待筛查客户的鼻咽癌筛查请求后，从所述鼻咽癌筛查请求中获取该待筛查客户的特征标签；

鼻咽癌筛查请求中包含待筛查客户的客户属性数据（例如，证件号码，或者，姓名和证件号码）及/或待筛查客户的特征标签；其中，特征标签包括基本特征信息（例如，所属城市属于中国西北地区、所属城市饮食习惯偏辛辣、所属城市 PM2.5（细颗粒物）经常超标、所属城市恶劣天气偏多、从事有害气体岗位等）、偏好习惯信息（例如，偏好晚睡、偏好烟酒、偏好烧烤食物、偏好户外运动等）、行为信息（例如，游戏次数多、加班次数多、点外卖

的次数多、就医次数多等)及/或社会关系信息(例如,未婚、独居、不经常与朋友联系等)。筛查服务器在收到一个待筛查客户的鼻咽癌筛查请求后,先看该请求中是否包含了该客户的特征标签,即从中去获取该客户的特征标签。

第二获取模块 102,用于在从所述鼻咽癌筛查请求中获取特征标签失败后,从所述鼻咽癌筛查请求中获取该待筛查客户的客户属性数据;

如果该鼻咽癌筛查请求中没有包含该客户的特征标签,则从该鼻咽癌筛查请求中获取不到特征标签(即获取特征标签失败),那么,则从该鼻咽癌筛查请求中获取该待筛查客户的客户属性数据,以根据该带筛查客户的客户属性数据去获得该待筛查客户的特征标签。

提取模块 103,用于从多个预先确定的业务服务器提取出与该待筛查客户的客户属性数据对应的各种特征数据;

筛查服务器与多个预先确定的业务服务器(例如,银行服务器、医疗服务器、保险服务器、即时通讯服务器、游戏服务器、天气服务器、外卖服务器及/或简历服务器等)之间通讯;在获取到该待筛查客户的客户属性数据后,筛查服务器从多个预先确定的业务服务器提取出与该待筛查客户的客户属性数据对应的各种特征数据(例如,银行贷款额度及还款情况信息、门诊病历信息“例如,预设时间内的看病次数、所患的疾病种类、每次患病的持续时间等”、保险信息“例如,所处行业,性别、年龄、婚姻状况、职业等”、即时通讯工具账号的使用信息“例如,通讯工具每天登陆时间信息、每天在线时长等信息”等等、游戏信息“例如,每天游戏登陆时间信息、每天游戏在线时长等信息”、天气信息“例如,最近三年内,PM2.5(细颗粒物)严重超标的天数”、外卖点餐信息“例如,每天点外卖的时间信息、每天所点外卖的外卖类型等”、求职简历上填写的信息“例如,兴趣爱好、性格、工作经历等信息”。

第一分析模块 104,用于按照预设的特征标签提取规则,对提取出的各种特征数据进行特征标签分析,以分析出该待筛查客户的特征标签;

筛查服务器中预先设置了特征标签的提取规则,在提取出该待筛查客户对应的各种特征数据后,利用该提取规则,对提取出的各种特征数据进行特征标签分析,从而分析得到该待筛查客户的特征标签。

第二分析模块 105,用于将该待筛查客户的特征标签代入预先训练好的分析模型中,以分析出该待筛查客户是否是高危鼻咽癌医学判断对象。

筛查服务器中具有预先训练好的分析模型,即该分析模型是经过了大量已知客户数据(包括客户是否患鼻咽癌与客户对应的特征标签)的训练;筛

查服务器在得到了当前待筛查客户的特征标签后，将得到的特征标签输入到该预先训练好的分析模型中，分析模型根据输入的各个特征标签，分析出该待筛查客户是否为高危鼻咽癌医学判断对象。

本实施例技术方案，在接收到待筛查客户的鼻咽癌筛查请求后，根据该
5 鼻咽癌筛查请求去获取该待筛查客户的特征标签，具体为先直接从鼻咽癌筛
查请求中获取该待筛查客户的特征标签，在获取失败时，则根据该鼻咽癌筛
查请求中的客户属性信息，去从各个预先确定的业务服务器中获取该客户属
性信息对应的各种特征数据，再按照预设的提取规则从这些特征数据中提取
10 出该待筛查客户的特征标签；在获得特征标签后，将获得的特征标签输入预
先训练好的分析模型中，从而分析得出该待筛查客户是否为高危鼻咽癌医学
判断对象。本实施例通过将分别代表待筛查客户的各种特征数据的特征标签，
输入预先训练好的分析模型中进行分析，以分析出待筛查客户是否为高危鼻
咽癌医学判断对象，实现客户准确的、高效的筛查出高危鼻咽癌医学判断对
象，让客户可以及早的进行预防或治疗。

15

本实施例的鼻咽癌筛查分析系统中，所述分析模型为多元线性回归模型，
该分析模型的训练过程为（可参照图2）：

步骤 E1，根据预先确定的客户属性数据，从多个预先确定的业务服务器
提取出第一预设数量的客户的各种特征数据；

20 客户属性数据例如为证件号码，或姓名和证件号码；筛查服务器与多个
预先确定的业务服务器（例如，银行服务器、医疗服务器、保险服务器、即
时通讯服务器、游戏服务器、天气服务器、外卖服务器及/或简历服务器等）
之间通讯，各个预先确定的业务服务器中具有客户属性数据与特征数据的映
射关系；筛查服务器先选取第一预设数量（例如 50 万）的客户，然后，对于
25 每一个选取的客户，分别根据该客户的客户属性数据去预先确定的各个业务
服务器中，分别提取出其客户属性数据对应的特征数据，从而得到选取的第
一预设数量的客户的各种特征数据。

步骤 E2，按照预设的特征标签提取规则，对提取的各个客户的各种特征
数据进行分析，以确定各个客户的特征标签；

30 筛查服务器中预先设置了特征标签的提取规则，在提取出第一预设数量
的客户的各种特征数据后，根据该预先设置的提取规则，分别对每个客户的
的各种特征数据进行特征标签分析，最终分析得到每一个客户的特征标签。

步骤 E3，根据预先确定的鼻咽癌与客户属性数据的映射关系，确定所述

第一预设数量的客户中患有鼻咽癌的异常客户，确定各个异常客户对应的特征标签，将各个正常客户对应的特征标签和各个异常客户对应的特征标签作为预设模型的训练样本；

5 筛查服务器的客户数据库中记录了每个客户的患病记录，筛查服务器根据客户的客户属性数据通过查找客户数据库就可确定该第一预设数量的客户中，哪些为患有鼻咽癌的异常客户，哪些为未患鼻咽癌的正常客户；在确定第一预设数量的客户中的异常客户和正常客户后，将每一个正常客户对应的特征标签作为一个训练样本，每一个异常客户对应的特征标签也作为一个训练样本。

10 步骤 E4，将所述训练样本分为第一百分比的训练集和第二百分比的验证集，所述第一百分比和第二百分比之和小于或者等于 100%；

将所有的训练样本的第一百分比作为训练集，以及所有的训练样本的第二百分比作为验证集，例如，第一百分比为 65%，第二百分比为 35%。当然，训练集和验证集中均包括部分正常客户对应的训练样本和部分异常客户对应的训练样本。

15 步骤 E5，利用训练集中的各个正常客户的特征标签和各个异常客户的特征标签对所述分析模型进行训练，并在训练完成后利用验证集中的各个正常客户的特征标签和各个异常客户的特征标签对训练的所述分析模型的准确率进行验证；

20 先利用训练集训练分析模型，在训练完成后，再用验证集验证分析模型的分析结果的准确率。

步骤 E6，若准确率大于预设阈值，则模型训练结束；

25 筛查服务器中预先设置了准确率阈值（即所述预设阈值，例如 98%），如果根据验证集确定分析模型的分析结果的准确率大于该准确率阈值，则说明分析模型的训练达到了预期的标准，筛查服务器则确认分析模型的训练结束。

步骤 E7，若准确率小于或者等于预设阈值，则执行上述步骤 E1、E2、E3 以增加训练样本的数量，并基于增加后的训练样本重新执行上述步骤 E4 和 E5。

30 如果根据验证集确定分析模型的分析结果的准确率小于或等于该准确率阈值，则说明分析模型的训练还没有达到预期的标准，筛查服务器此时，则继续执行步骤 E1、E2、E3，并增加第一预设数量的值（例如每次增大 5 万），并基于增加后的训练样本重新执行上述步骤 E4 和 E5，直至达到步骤 E6 的要求。

本申请鼻咽癌筛查分析系统中所述预设的特征标签提取规则为：

对于连续数值的各种特征数据种类设置对应的标签阈值；

例如，所属城市 PM2.5（细颗粒物）最近三年每年超标的天数对应的的标签阈值可以为 60 天，当所属城市 PM2.5 最近三年每年超标的天数大于对应的标签阈值“60 天”时，代表所属城市 PM2.5 经常超标；所属城市最近三年每年蓝色以上预警等级的天数对应的的标签阈值可以为 55 天，当所属城市最近三年每年蓝色以上预警等级的天数大于对应的的标签阈值“55 天”时，代表所属城市恶劣天气偏多；最近一年晚于 23:00 睡觉的天数对应的的标签阈值可以是 100 天，当最近一年晚于 23:00 睡觉的天数大于对应的的标签阈值“100 天”时，代表偏好晚睡；最近一年点烧烤类的外卖次数对应的的标签阈值可以是 80 次，当最近一年点烧烤类的外卖次数大于对应的的标签阈值“80 次”时，代表偏好烧烤食物；最近一年就医次数对应的的标签阈值可以是 30 次，当最近一年就医次数大于对应的的标签阈值“30 次”时，代表就医次数多。

对于非为连续数值的各种特征数据种类设置对应的标签范围；

例如，所属城市对应的标签范围包括：西北地区城市集合、华北地区城市集合、华中地区城市集合、华南地区城市集合等等，当客户所属城市属于所述西北地区城市集合中的城市时，代表该客户所属城市对应的标签信息是“属于中国西北地区”；所属城市饮食习惯对应的标签范围包括：偏辛辣城市集合、偏油腻城市集合、偏清淡城市集合、偏甜/咸城市集合等，当客户所属城市属于所述偏辛辣城市集合中的城市时，代表该客户所属城市对应的标签信息是“饮食习惯偏辛辣”；从事岗位对应的标签范围包括：有害气体岗位集合、非易产生挥发性有害气体的岗位集合、易产生挥发性有害气体的岗位集合、无害气体岗位集合等，当客户从事岗位属于有害气体岗位集合中的岗位，代表该客户所从事行业对应的标签信息是“从事有害气体岗位”。

根据连续数值的各种特征数据种类与标签阈值的映射关系，确定出各个客户的各种连续数值的特征数据对应的标签信息，及根据非连续数值的各种特征数据种类与标签范围的映射关系，确定出各个客户的各种非连续数值的特征数据对应的标签信息。

30

进一步地，本申请还提出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鼻咽癌筛查分析系统，所述鼻咽癌筛查分析系统可被至少一个处理器执行，以使所述至少一个处理器执行上述任一实施例中的鼻咽癌筛查

分析方法。

5 以上所述仅为本申请的优选实施例，并非因此限制本申请的专利范围，凡是在本申请的发明构思下，利用本申请说明书及附图内容所作的等效结构变换，或直接/间接运用在其他相关的技术领域均包括在本申请的专利保护范围内。

权利要求书

1、一种电子装置，其特征在于，所述电子装置包括存储器和处理器，所述存储器上存储有可在所述处理器上运行的鼻咽癌筛查分析系统，所述鼻咽癌筛查分析系统被所述处理器执行时实现如下步骤：

5 在收到一个待筛查客户的鼻咽癌筛查请求后，从所述鼻咽癌筛查请求中获取该待筛查客户的特征标签；

若从所述鼻咽癌筛查请求中获取特征标签失败，则从所述鼻咽癌筛查请求中获取该待筛查客户的客户属性数据；

10 从多个预先确定的业务服务器提取出与该待筛查客户的客户属性数据对应的各种特征数据；

按照预设的特征标签提取规则，对提取出的各种特征数据进行特征标签分析，以分析出该待筛查客户的特征标签；

将该待筛查客户的特征标签代入预先训练好的分析模型中，以分析出该待筛查客户是否是高危鼻咽癌医学判断对象。

15 2、如权利要求1所述的电子装置，其特征在于，所述预设的特征标签提取规则为：

对于连续数值的各种特征数据种类设置对应的标签阈值；

对于非为连续数值的各种特征数据种类设置对应的标签范围；

20 根据连续数值的各种特征数据种类与标签阈值的映射关系，确定出各个客户的各种连续数值的特征数据对应的标签信息，及根据非连续数值的各种特征数据种类与标签范围的映射关系，确定出各个客户的各种非连续数值的特征数据对应的标签信息。

3、如权利要求1所述的电子装置，其特征在于，所述分析模型为多元线性回归模型，所述分析模型的训练步骤包括：

25 E1、根据预先确定的客户属性数据，从多个预先确定的业务服务器提取出第一预设数量的客户的各种特征数据；

E2、按照预设的特征标签提取规则，对提取的各个客户的各种特征数据进行分析，以确定各个客户的特征标签；

30 E3、根据预先确定的鼻咽癌与客户属性数据的映射关系，确定所述第一预设数量的客户中患有鼻咽癌的异常客户，确定各个异常客户对应的特征标签，将各个正常客户对应的特征标签和各个异常客户对应的特征标签作为预设模型的训练样本；

E4、将所述训练样本分为第一百分比的训练集和第二百分比的验证集，

所述第一百分比和第二百分比之和小于或者等于 100%;

E5、利用训练集中的各个正常客户的特征标签和各个异常客户的特征标签对所述分析模型进行训练,并在训练完成后利用验证集中的各个正常客户的特征标签和各个异常客户的特征标签对训练的所述分析模型的准确率进行验证;

E6、若准确率大于预设阈值,则模型训练结束;

E7、若准确率小于或者等于预设阈值,则执行上述步骤 E1、E2、E3 以增加训练样本的数量,并基于增加后的训练样本重新执行上述步骤 E4 和 E5。

4、如权利要求 3 所述的电子装置,其特征不在于,所述预设的特征标签提取规则为:

对于连续数值的各种特征数据种类设置对应的标签阈值;

对于非为连续数值的各种特征数据种类设置对应的标签范围;

根据连续数值的各种特征数据种类与标签阈值的映射关系,确定出各个客户的各种连续数值的特征数据对应的标签信息,及根据非连续数值的各种特征数据种类与标签范围的映射关系,确定出各个客户的各种非连续数值的特征数据对应的标签信息。

5、如权利要求 1 所述的电子装置,其特征不在于,所述特征标签包括基本特征信息、偏好习惯信息、行为信息及/或社会关系信息。

6、如权利要求 5 所述的电子装置,其特征不在于,所述预设的特征标签提取规则为:

对于连续数值的各种特征数据种类设置对应的标签阈值;

对于非为连续数值的各种特征数据种类设置对应的标签范围;

根据连续数值的各种特征数据种类与标签阈值的映射关系,确定出各个客户的各种连续数值的特征数据对应的标签信息,及根据非连续数值的各种特征数据种类与标签范围的映射关系,确定出各个客户的各种非连续数值的特征数据对应的标签信息。

7、如权利要求 5 所述的电子装置,其特征不在于,所述分析模型为多元线性回归模型,所述分析模型的训练步骤包括:

E1、根据预先确定的客户属性数据,从多个预先确定的业务服务器提取出第一预设数量的客户的各种特征数据;

E2、按照预设的特征标签提取规则,对提取的各个客户的各种特征数据进行分析,以确定各个客户的特征标签;

E3、根据预先确定的鼻咽癌与客户属性数据的映射关系,确定所述第一

预设数量的客户中患有鼻咽癌的异常客户，确定各个异常客户对应的特征标签，将各个正常客户对应的特征标签和各个异常客户对应的特征标签作为预设模型的训练样本；

- 5 E4、将所述训练样本分为第一百分比的训练集和第二百分比的验证集，
所述第一百分比和第二百分比之和小于或者等于 100%；

E5、利用训练集中的各个正常客户的特征标签和各个异常客户的特征标签对所述分析模型进行训练，并在训练完成后利用验证集中的各个正常客户的特征标签和各个异常客户的特征标签对训练的所述分析模型的准确率进行验证；

- 10 E6、若准确率大于预设阈值，则模型训练结束；

E7、若准确率小于或者等于预设阈值，则执行上述步骤 E1、E2、E3 以增加训练样本的数量，并基于增加后的训练样本重新执行上述步骤 E4 和 E5。

8、如权利要求 5 所述的电子装置，其特征不在于，所述预设的特征标签提取规则为：

- 15 对于连续数值的各种特征数据种类设置对应的标签阈值；

对于非为连续数值的各种特征数据种类设置对应的标签范围；

- 20 根据连续数值的各种特征数据种类与标签阈值的映射关系，确定出各个客户的各种连续数值的特征数据对应的标签信息，及根据非连续数值的各种特征数据种类与标签范围的映射关系，确定出各个客户的各种非连续数值的特征数据对应的标签信息。

9、一种鼻咽癌筛查分析方法，其特征在于，该方法包括步骤：

在收到一个待筛查客户的鼻咽癌筛查请求后，从所述鼻咽癌筛查请求中获取该待筛查客户的特征标签；

- 25 若从所述鼻咽癌筛查请求中获取特征标签失败，则从所述鼻咽癌筛查请求中获取该待筛查客户的客户属性数据；

从多个预先确定的业务服务器提取出与该待筛查客户的客户属性数据对应的各种特征数据；

按照预设的特征标签提取规则，对提取出的各种特征数据进行特征标签分析，以分析出该待筛查客户的特征标签；

- 30 将该待筛查客户的特征标签代入预先训练好的分析模型中，以分析出该待筛查客户是否是高危鼻咽癌医学判断对象。

10、如权利要求 9 所述的鼻咽癌筛查分析方法，其特征在于，所述预设的特征标签提取规则为：

对于连续数值的各种特征数据种类设置对应的标签阈值;

对于非为连续数值的各种特征数据种类设置对应的标签范围;

根据连续数值的各种特征数据种类与标签阈值的映射关系, 确定出各个客户的各种连续数值的特征数据对应的标签信息, 及根据非连续数值的各种特征数据种类与标签范围的映射关系, 确定出各个客户的各种非连续数值的特征数据对应的标签信息。

11、如权利要求 9 所述的鼻咽癌筛查分析方法, 其特征在于, 所述分析模型为多元线性回归模型, 所述分析模型的训练步骤包括:

E1、根据预先确定的客户属性数据, 从多个预先确定的业务服务器提取出第一预设数量的客户的各种特征数据;

E2、按照预设的特征标签提取规则, 对提取的各个客户的各种特征数据进行分析, 以确定各个客户的特征标签;

E3、根据预先确定的鼻咽癌与客户属性数据的映射关系, 确定所述第一预设数量的客户中患有鼻咽癌的异常客户, 确定各个异常客户对应的特征标签, 将各个正常客户对应的特征标签和各个异常客户对应的特征标签作为预设模型的训练样本;

E4、将所述训练样本分为第一百分比的训练集和第二百分比的验证集, 所述第一百分比和第二百分比之和小于或者等于 100%;

E5、利用训练集中的各个正常客户的特征标签和各个异常客户的特征标签对所述分析模型进行训练, 并在训练完成后利用验证集中的各个正常客户的特征标签和各个异常客户的特征标签对训练的所述分析模型的准确率进行验证;

E6、若准确率大于预设阈值, 则模型训练结束;

E7、若准确率小于或者等于预设阈值, 则执行上述步骤 E1、E2、E3 以增加训练样本的数量, 并基于增加后的训练样本重新执行上述步骤 E4 和 E5。

12、如权利要求 11 所述的鼻咽癌筛查分析方法, 其特征在于, 所述预设的特征标签提取规则为:

对于连续数值的各种特征数据种类设置对应的标签阈值;

对于非为连续数值的各种特征数据种类设置对应的标签范围;

根据连续数值的各种特征数据种类与标签阈值的映射关系, 确定出各个客户的各种连续数值的特征数据对应的标签信息, 及根据非连续数值的各种特征数据种类与标签范围的映射关系, 确定出各个客户的各种非连续数值的特征数据对应的标签信息。

13、如权利要求 9 所述的鼻咽癌筛查分析方法，其特征在于，所述特征标签包括基本特征信息、偏好习惯信息、行为信息及/或社会关系信息。

14、如权利要求 13 所述的鼻咽癌筛查分析方法，其特征在于，所述预设的特征标签提取规则为：

5 对于连续数值的各种特征数据种类设置对应的标签阈值；

对于非为连续数值的各种特征数据种类设置对应的标签范围；

根据连续数值的各种特征数据种类与标签阈值的映射关系，确定出各个客户的各种连续数值的特征数据对应的标签信息，及根据非连续数值的各种特征数据种类与标签范围的映射关系，确定出各个客户的各种非连续数值的特征数据对应的标签信息。

15、如权利要求 13 所述的鼻咽癌筛查分析方法，其特征在于，所述分析模型为多元线性回归模型，所述分析模型的训练步骤包括：

E1、根据预先确定的客户属性数据，从多个预先确定的业务服务器提取出第一预设数量的客户的各种特征数据；

15 E2、按照预设的特征标签提取规则，对提取的各个客户的各种特征数据进行分析，以确定各个客户的特征标签；

E3、根据预先确定的鼻咽癌与客户属性数据的映射关系，确定所述第一预设数量的客户中患有鼻咽癌的异常客户，确定各个异常客户对应的特征标签，将各个正常客户对应的特征标签和各个异常客户对应的特征标签作为预设模型的训练样本；

20 E4、将所述训练样本分为第一百分比的训练集和第二百分比的验证集，所述第一百分比和第二百分比之和小于或者等于 100%；

E5、利用训练集中的各个正常客户的特征标签和各个异常客户的特征标签对所述分析模型进行训练，并在训练完成后利用验证集中的各个正常客户的特征标签和各个异常客户的特征标签对训练的所述分析模型的准确率进行验证；

E6、若准确率大于预设阈值，则模型训练结束；

E7、若准确率小于或者等于预设阈值，则执行上述步骤 E1、E2、E3 以增加训练样本的数量，并基于增加后的训练样本重新执行上述步骤 E4 和 E5。

30 16、如权利要求 15 所述的鼻咽癌筛查分析方法，其特征在于，所述预设的特征标签提取规则为：

对于连续数值的各种特征数据种类设置对应的标签阈值；

对于非为连续数值的各种特征数据种类设置对应的标签范围；

根据连续数值的各种特征数据种类与标签阈值的映射关系，确定出各个客户的各种连续数值的特征数据对应的标签信息，及根据非连续数值的各种特征数据种类与标签范围的映射关系，确定出各个客户的各种非连续数值的特征数据对应的标签信息。

- 5 17、一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有鼻咽癌筛查分析系统，所述鼻咽癌筛查分析系统可被至少一个处理器执行，以使所述至少一个处理器执行如下步骤：

在收到一个待筛查客户的鼻咽癌筛查请求后，从所述鼻咽癌筛查请求中获取该待筛查客户的特征标签；

- 10 若从所述鼻咽癌筛查请求中获取特征标签失败，则从所述鼻咽癌筛查请求中获取该待筛查客户的客户属性数据；

从多个预先确定的业务服务器提取出与该待筛查客户的客户属性数据对应的各种特征数据；

- 15 按照预设的特征标签提取规则，对提取出的各种特征数据进行特征标签分析，以分析出该待筛查客户的特征标签；

将该待筛查客户的特征标签代入预先训练好的分析模型中，以分析出该待筛查客户是否是高危鼻咽癌医学判断对象。

18、如权利要求 17 所述的计算机可读存储介质，其特征在于，所述特征标签包括基本特征信息、偏好习惯信息、行为信息及/或社会关系信息。

- 20 19、如权利要求 17 所述的计算机可读存储介质，其特征在于，所述分析模型为多元线性回归模型，所述分析模型的训练步骤包括：

E1、根据预先确定的客户属性数据，从多个预先确定的业务服务器提取出第一预设数量的客户的各种特征数据；

- 25 E2、按照预设的特征标签提取规则，对提取的各个客户的各种特征数据进行分析，以确定各个客户的特征标签；

E3、根据预先确定的鼻咽癌与客户属性数据的映射关系，确定所述第一预设数量的客户中患有鼻咽癌的异常客户，确定各个异常客户对应的特征标签，将各个正常客户对应的特征标签和各个异常客户对应的特征标签作为预设模型的训练样本；

- 30 E4、将所述训练样本分为第一百分比的训练集和第二百分比的验证集，所述第一百分比和第二百分比之和小于或者等于 100%；

E5、利用训练集中的各个正常客户的特征标签和各个异常客户的特征标签对所述分析模型进行训练，并在训练完成后利用验证集中的各个正常客户

的特征标签和各个异常客户的特征标签对训练的所述分析模型的准确率进行验证;

E6、若准确率大于预设阈值,则模型训练结束;

5 E7、若准确率小于或者等于预设阈值,则执行上述步骤 E1、E2、E3 以增加训练样本的数量,并基于增加后的训练样本重新执行上述步骤 E4 和 E5。

20、如权利要求 17 所述的计算机可读存储介质,其特征在于,所述预设的特征标签提取规则为:

对于连续数值的各种特征数据种类设置对应的标签阈值;

对于非为连续数值的各种特征数据种类设置对应的标签范围;

10 根据连续数值的各种特征数据种类与标签阈值的映射关系,确定出各个客户的各种连续数值的特征数据对应的标签信息,及根据非连续数值的各种特征数据种类与标签范围的映射关系,确定出各个客户的各种非连续数值的特征数据对应的标签信息。

15