

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(10) 国际公布号

WO 2019/218482 A1

(43) 国际公布日

2019 年 11 月 21 日 (21.11.2019)

(51) 国际专利分类号:

G06K 9/62 (2006.01)

(21) 国际申请号:

PCT/CN2018/097561

(22) 国际申请日:

2018 年 7 月 27 日 (27.07.2018)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

201810455659.4 2018年5月14日 (14.05.2018) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];

中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(72) 发明人: 卢少烽(LU, Shaofeng); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。洪博然(HONG, Boran); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。徐亮(XU, Liang); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。阮晓雯(RUAN, Xiaowen); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。肖京(XIAO, Jing); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳中一专利商标事务所(SHENZHEN ZHONGYI PATENT AND TRADEMARK OFFICE);

(54) Title: BIG DATA-BASED POPULATION SCREENING METHOD AND APPARATUS, TERMINAL DEVICE AND READ-ABLE STORAGE MEDIUM

(54) 发明名称: 基于大数据的人群筛选方法、装置、终端设备及可读存储介质

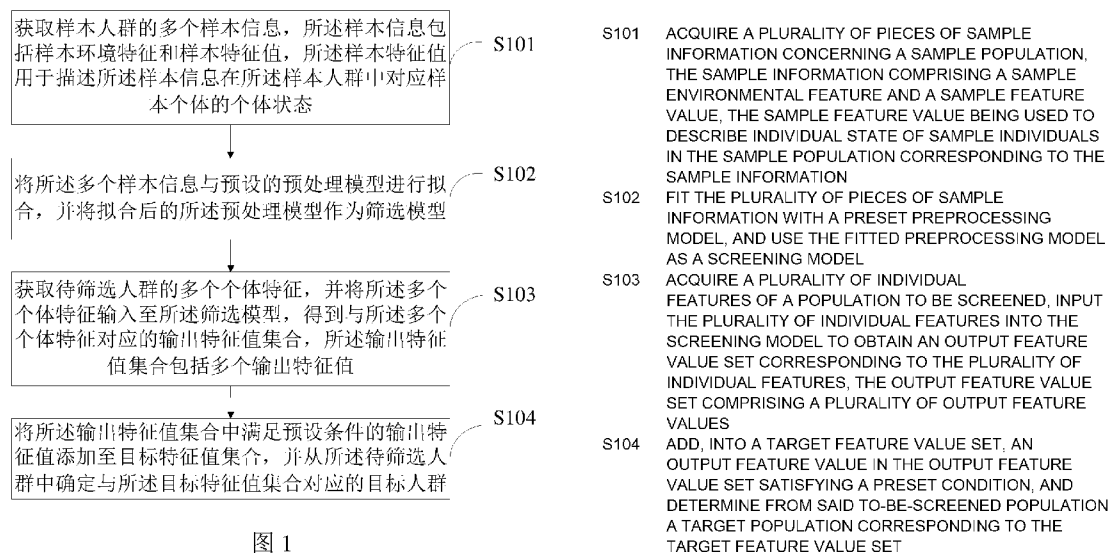


图 1

(57) Abstract: A big data-based population screening method, a terminal device and a computer readable storage medium. The method comprises: acquiring a plurality of pieces of sample information concerning a sample population, the sample information comprising a sample environmental feature and a sample feature value; fitting the plurality of pieces of sample information with a preset preprocessing model, and using the fitted preprocessing model as a screening model (S102); acquiring a plurality of individual features of a population to be screened, inputting the plurality of individual features into the screening model to obtain an output feature value set corresponding

中国广东省深圳市福田区深南中路
1014 号老特区报社四楼 (5 号信箱),
Guangdong 518028 (CN)。

- (81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。
- (84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告 (条约第21条(3))。

to the plurality of individual features, the output feature value set comprising a plurality of output feature values (S103); adding, into a target feature value set, an output feature value in the output feature value set satisfying a preset condition, and determining from said to-be-screened population a target population corresponding to the target feature value set (S104). Said method implements screening, according to a plurality of features, on a population to be screened, improving the accuracy of population screening.

(57) 摘要: 基于大数据的人群筛选方法、终端设备及计算机可读存储介质, 包括: 获取样本人群的多个样本信息, 样本信息包括样本环境特征和样本特征值; 将多个样本信息与预设的预处理模型进行拟合, 并将拟合后的所述预处理模型作为筛选模型 (S102); 获取待筛选人群的多个个体特征, 并将所述多个个体特征输入至所述筛选模型, 得到与所述多个个体特征对应的输出特征值集合, 所述输出特征值集合包括多个输出特征值 (S103); 将所述输出特征值集合中满足预设条件的输出特征值添加至目标特征值集合, 并从所述待筛选人群中确定与所述目标特征值集合对应的目标人群 (S104)。以上方法实现了根据多个特征对待筛选人群进行筛选, 提升了人群筛选的准确率。

基于大数据的人群筛选方法、装置、终端设备及可读存储介质

本申请要求于 2018 年 05 月 14 日提交中国专利局、申请号为 201810455659.4、发明名称为“基于大数据的人群筛选方法及终端设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

5 技术领域

本申请属于数据处理技术领域，尤其涉及基于大数据的人群筛选方法、装置、终端设备及计算机可读存储介质。

背景技术

10 在现实生活中，往往存在从基数庞大的人群中筛选出部分目标人群的需求，筛选一般以某个特征为依据，比如说以某个县所有人为基础，筛选出年龄超过五十岁的目标人群。而统计学是关于认识客观现象总体数量特征和数量关系的科学，在需要筛选出因某个状态改变而产生特殊性的目标人群时，需要运用统计学，查找出与该状态改变关联的相关特征，从而根据相关特征筛选出目标人群。

15 但是，在现有的人群筛选方法中，根据统计学能够查找出的相关特征较少，从而根据相关特征筛选出的人群为目标人群的准确度较低。比如在对患有慢阻肺的人群进行筛选时，主要根据已患病人群的统计结果，如果统计出某一个区域附近患有慢阻肺的人数较多，或位于某一个年龄区间的患有慢阻肺的人数较多，则将该区域或该年龄区间作为相关特征筛选出目标人群，但实际上慢阻肺的相关特征较多，故实际的目标人群应当不限于该区域或该年龄区间进行筛选。综上，现有的人群筛选方法可依据的相关特征少，筛选准确度低。

20 技术问题

有鉴于此，本申请实施例提供了基于大数据的人群筛选方法、装置、终端设备及计算机可读存储介质，以解决现有技术中进行人群筛选时可依据的相关特征少，筛选的准确度低的问题。

技术方案

25 本申请实施例的第一方面提供了一种基于大数据的人群筛选方法，包括：

获取样本人群的多个样本信息，所述样本信息包括样本环境特征和样本特征值，所述样本特征值用于描述所述样本信息在所述样本人群中对样本个体的个体状态；

将所述多个样本信息与预设的预处理模型进行拟合，并将拟合后的所述预处理模型作为筛选模型；

获取待筛选人群的多个个体特征，并将所述多个个体特征输入至所述筛选模型，得到与所述多个个体特征对应的输出特征值集合，所述输出特征值集合包括多个输出特征值；

将所述输出特征值集合中满足预设条件的输出特征值添加至目标特征值集合，并从所述待筛选人群中确定与所述目标特征值集合对应的目标人群。

5 本申请实施例的第二方面提供了一种基于大数据的人群筛选装置，可以包括用于实现上述基于大数据的人群筛选方法的步骤的单元。

本申请实施例的第三方面提供了一种终端设备，包括存储器以及处理器，所述存储器中存储有可在所述处理器上运行的计算机可读指令，所述处理器执行所述计算机可读指令时实现上述基于大数据的人群筛选方法的步骤。

10 本申请实施例的第四方面提供了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可读指令，所述计算机可读指令被处理器执行时实现上述基于大数据的人群筛选方法的步骤。

有益效果

15 本申请实施例通过获取样本人群的多个样本信息，其中，每个样本信息包括样本环境特征和样本特征值，样本特征值用于描述样本特征值所属的样本信息对应的样本个体的个体状态，获取完毕后，将多个样本信息与预设的预处理模型进行拟合，并将拟合完成的预处理模型作为筛选模型进行输出，然后通过筛选模型进行对待筛选人群的筛选，获取待筛选人群的多个个体特征，将多个个体特征输入至筛选模型，得到经筛选模型计算后输出的与多个个体特征对应的输出特征值集合，特征值集合包括多个输出特征值，最后将输出特征值集合中满
20 足预设条件的输出特征值添加至目标特征值集合，从待筛选人群中确定出与目标特征值集合对应的目标人群，本申请实施例通过基于多个样本信息训练出筛选模型，从而可根据多个特征对待筛选人群进行筛选，提升了人群筛选的准确率。

附图说明

图 1 是本申请实施例一中基于大数据的人群筛选方法的实现流程图；
25 图 2 是本申请实施例二中基于大数据的人群筛选方法的实现流程图；
图 3 是本申请实施例三中基于大数据的人群筛选方法的实现流程图；
图 4 是本申请实施例四中基于大数据的人群筛选方法的实现流程图；
图 5 是本申请实施例五中基于大数据的人群筛选方法的实现流程图；
图 6 是本申请实施例六中基于大数据的人群筛选装置的结构框图；
30 图 7 是本申请实施例七中终端设备的示意图。

本发明的实施方式

为了对本申请的技术特征、目的和效果有更加清楚的理解，现对照附图详细说明本申请的具体实施方式。

请参阅图 1，图 1 是本申请实施例提供的一种基于大数据的人群筛选方法的实现流程图。

5 如图 1 所示，该人群筛选方法包括以下步骤：

S101：获取样本人群的多个样本信息，所述样本信息包括样本环境特征和样本特征值，所述样本特征值用于描述所述样本信息在所述样本人群中对应样本个体的个体状态。

在本申请实施例中，为了筛选出需求的目标人群，首先获取样本人群中多个样本个体的多个样本信息，其中一个样本信息对应一个样本个体。样本人群根据样本条件确定，样本条件包括样本来源和样本选择因素等，样本来源决定样本人群选取的来源地，比如样本来源为
10 某个市的个体档案，样本选择因素决定从来源地选择的构成样本人群的样本个体，比如样本选择因素为年龄在五十岁以上的男性，故样本来源和样本选择因素组成的样本条件即是在某个市的个体档案中筛选出年龄在五十岁以上的男性的多个样本个体组成样本人群。当然，样本条件并不限于上述例子，即可根据实际应用场景进行确定。为了提升对目标人群筛选的准确
15 度，故在样本人群对应的样本条件确定后，可设置从样本人群中获取的样本信息的数量，使样本信息的数量处于预设的数量级，数量级可根据准确度要求自由设置。比如数量级可以为一千个。样本人群中的样本信息包括样本环境特征和样本特征值，其中，样本环境特征用于指示该样本信息对应样本个体的环境状态，可能包含一个或多个子特征，比如样本环境特征可以包括样本个体的年龄、家庭饮水类型、婚姻状况、职业、血压状况、呼吸频率、糖尿病状况、饮酒状况和吸烟状况等，同样地，样本环境特征包含的子特征也可以根据实际应用场景进行确定。样本信息中的样本特征值用于指示与样本信息对应的样本个体的个体状态，其中，个体状态与本申请实施例中筛选目标人群的目的相关联，举例来说，若需要筛选出受
20 焦虑症影响的目标人群，则个体状态与个体是否患有焦虑症相关，则可将样本特征值的取值范围限定为 0 和 1 两个整数，样本特征值为 1 指示对应的样本个体已受焦虑症影响，样本特征值为 0 指示对应的样本个体未受焦虑症影响。
25

优选地，从健康档案或特定数据库获取多个血液信息样本。通常来说，健康档案或特点数据库如医院数据库中存放有不同种类的样本个体，存储信息的数量级高，并且样本环境特征包含的特征较为齐全，样本特征值已预先确定，并且准确度较高，故可直接从健康档案或特定数据库中确定样本人群，再基于样本人群进行多个样本信息的获取。

30 S102：将所述多个样本信息与预设的预处理模型进行拟合，并将拟合后的所述预处理模

型作为筛选模型。

由于样本信息下的样本环境特征可能包含多个子特征，而个体状态与对应的样本个体的多个子特征有关，即多个子特征会对样本特征值的生成产生影响。举例来说，若某样本个体代表的一个成年男性处于已婚状态，育有三个子女，职业变更频率高，并且处于饮酒和吸烟的状态，则该样本个体患焦虑症的概率，即样本特征值为 1 的概率较大。在上述例子中，样本个体的样本环境特征包括的婚姻状态、子女个数、职业、饮酒状况和吸烟状况与该样本个体的样本特征值相关联。但是，在普遍情况下，无法获知个体状态受多个子特征影响的变化规律，故在本申请实施例，将样本环境特征和样本特征值已确定的多个样本信息与预设的预处理模型进行拟合，在拟合过程中不断训练预处理模型，最后将拟合完成的预处理模型输出为筛选模型。

S103: 获取待筛选人群的多个个体特征，并将所述多个个体特征输入至所述筛选模型，得到与所述多个个体特征对应的输出特征值集合，所述输出特征值集合包括多个输出特征值。

由多个样本信息生成筛选模型后，对待筛选人群进行分析，首先，获取待筛选人群的多个个体特征，个体特征与待筛选人群中的待筛选个体对应。一般地，个体特征与样本个体的样本环境特征的格式相同，即包含子特征的种类相同。可选地，在获取到个体特征后，对个体特征进行处理，仅保留个体特征中与样本环境特征相同种类子特征。比如个体特征包括姓名、年龄、性别、饮酒状况和吸烟状况，而样本环境特征包括年龄、饮酒状况和吸烟状况，则在将个体特征输入至筛选模型之前，先对个体特征进行处理，只获取个体特征中种类为年龄、饮酒状况和吸烟状况的子特征，降低了后续计算的复杂度。另外，若处理后的个体特征中某个子特征的值为空，则取多个样本个体的样本环境特征中该子特征的平均值或预设的值作为个体特征中该子特征的值，防止后续计算出错，提升了计算的稳定性。

确定多个个体特征后，将多个个体特征输入至筛选模型，每个个体特征在经筛选模型计算后可得到对应的输出特征值，输出特征值与样本信息的样本特征值的格式相同，但并不限定于样本特征值的取值范围。获取多个个体特征对应的多个输出特征值输出结果，生成输出特征值集合。

S104: 将所述输出特征值集合中满足预设条件的输出特征值添加至目标特征值集合，并从所述待筛选人群中确定与所述目标特征值集合对应的目标人群。

输出特征值集合确定后，从输出特征值集合中获取满足预设条件的输出特征值，并将该输出特征值添加至目标特征值集合。在本申请实施例，预设条件可设定将输出特征值集合中处于预设比例的输出特征值添加至目标特征值集合，比如将输出特征值集合中的多个输出特征值进行大小排序，并在排序完成后，将比例处于前 10% 的输出特征值添加进目标特征值

集合。此外，预设条件还可根据特征阈值进行设定，特征阈值的确定方法在后文进行具体阐述。由于输出特征值对应待筛选人群中的待筛选个体，故根据目标特征值集合可确定出待筛选人群中的目标人群，完成对待筛选人群的筛选。

通过图 1 所示实施例可知，在本申请实施例中，首先获取样本人群中多个样本个体对应的多个样本信息，其中，样本信息包括样本环境特征和样本特征值，样本特征值用于描述样本个体的个体状态，将多个样本信息与预设的预处理模型进行拟合，并将拟合完成的预处理模型输出为筛选模型，进行对待筛选人群的筛选，获取待筛选人群多个待筛选个体的多个个体特征，将多个个体特征输入至筛选模型，得到由多个输出特征值组合成的输出特征值集合，最后将输出特征值集合中满足预设条件的输出特征值添加至目标特征值集合，从待筛选人群中确定与目标特征值集合对应的多个待筛选个体，作为目标人群，实现了自动筛选，并且提升了对目标人群筛选的准确度。

请参阅图 2，图 2 是本申请实施例二提供的一种基于大数据的人群筛选方法的实现流程图。相对于图 1 对应的实施例，本实施例对 S102 进行细化后得到 S201~S202，详述如下：

S201：将所述多个样本信息输入至所述预处理模型，以训练所述预处理模型，其中，将所述样本信息的样本环境特征作为所述预处理模型的输入参数，将所述样本信息的样本特征值作为所述预处理模型的参照参数。

在将多个样本信息输入至预处理模型时，将样本信息中的样本环境特征作为预处理模型的输入参数，将样本信息中的样本特征值作为预处理模型的参照参数，具体地，基于多个样本信息构建样本信息集，为 $(Characters_{environ1}, Value_{environ1}), (Characters_{environ2}, Value_{environ2}), \dots, (Characters_{environn}, Value_{environn})$ ，其中， $Characters_{enviroin}$ 代表第 i 个样本信息的样本环境特征，在其下可能包括多个子特征， $Value_{enviroin}$ 代表第 i 个样本信息的样本特征值， n 代表从样本人群获取的多个样本信息的总个数。将样本信息集输入至预处理模型，以训练预处理模型，本申请实施例中，预处理模型对输入参数的计算公式为：

$$Value_{enviroin}^{\wedge} = \sum_{k=1}^K f_k(Characters_{enviroin})$$

在上述公式中， $Value_{enviroin}^{\wedge}$ 代表对输入参数为 $Characters_{enviroin}$ 的预测值，即是将 $Characters_{enviroin}$ 作为输入参数输入至预处理模型后，预处理模型计算后的输出结果。公式中的 $f()$ 指示一个存在于函数空间的函数，函数空间指的是从一个集合到另一个集合的给定种类的函数的集合，即 $f()$ 函数最初处于未知状态， K 则表示预处理模型中存在 K 个上述

的 $f()$ 函数, 需要将所有的 $f()$ 函数计算出的结果累加后, 才能得到最终的预测值。

在实际对预处理模型的训练过程, 依赖于上述公式, 在本申请实施例, 采用次序学习的方法对 $f()$ 函数进行学习, 以使最终得到的 K 个 $f()$ 函数最大限度地符合多个样本信息中的数据。举例来说, 在输入参数为 $Characters_{enviro ni}$ 的基础上, 进行 t 轮的预测值预测, 并在进行第 t 轮的预测值预测时, 保留第 $t-1$ 轮的预测值预测结果, 即依照次序训练预处理模型, 使得预测值 ($Value_{enviro ni}^{\wedge}$) 与真实值 ($Value_{enviro ni}$) 之间的差距逐渐减小, 具体见下:

$$Value_{enviro ni}^{\wedge (0)} = 0$$

$$Value_{enviro ni}^{\wedge (1)} = f_1(Characters_{enviro ni}) = Value_{enviro ni}^{\wedge (0)} + f_1(Characters_{enviro ni})$$

$$\begin{aligned} Value_{enviro ni}^{\wedge (2)} &= f_1(Characters_{enviro ni}) + f_2(Characters_{enviro ni}) \\ &= Value_{enviro ni}^{\wedge (1)} + f_2(Characters_{enviro ni}) \end{aligned}$$

.....

$$\begin{aligned} Value_{enviro ni}^{\wedge (t)} &= \sum_{k=1}^t f_k(Characters_{enviro ni}) \\ &= Value_{enviro ni}^{\wedge (t-1)} + f_t(Characters_{enviro ni}) \end{aligned}$$

上述公式中的 $Value_{enviro ni}^{\wedge (t)}$ 是在给出输入参数为 $Characters_{enviro ni}$ 的基础上, 进行第 t 轮的预测后的预测值。为了确定在次序学习过程中所需求的 $f()$ 函数, 使其尽量贴近于样本信息集, 故构建优化函数, 具体公式见下:

$$Opt^{(t)} = \sum_{i=1}^n l(Value_{enviro ni}, Value_{enviro ni}^{\wedge (t-1)} + f_t(Characters_{enviro ni})) + \Omega(f_t) + D$$

在上述公式中, $Value_{enviro ni}$ 是样本信息集中与输入参数 $Characters_{enviro ni}$ 对应的参照参数, 即是样本信息中的样本特征值, 在本申请实施例, 作为优化函数的参数。优化函数公式中的 $\Omega(f_t)$ 为正则项, D 为常数项, 其中, 正则项控制优化函数的训练程度, 防止样本信息集和预处理模型过拟合; 常数项为一个常量, 设置常数项是为了限制优化函数的数值范围。值得

一提的是, $\sum_{i=1}^n l(Value_{enviro ni}, Value_{enviro ni}^{\wedge (t-1)} + f_t(Characters_{enviro ni}))$ 为误差函数, 对优化函数进行优化的过程, 即是确定合适的 $f()$ 函数使得上述误差函数的值尽量减小的过程。

在本申请实施例中，为了在计算层面上方便对优化函数进行优化，对上述的

$\sum_{i=1}^n l(Value_{enviro\text{ni}}, Value_{enviro\text{ni}}^{\wedge(t-1)} + f_t(Characters_{enviro\text{ni}}))$ 进行展开，并定义：

$$g_i = \partial_{Value_{enviro\text{ni}}^{\wedge(t-1)}} l(Value_{enviro\text{ni}}, Value_{enviro\text{ni}}^{\wedge(t-1)})$$

$$h_i = \partial^2_{Value_{enviro\text{ni}}^{\wedge(t-1)}} l(Value_{enviro\text{ni}}, Value_{enviro\text{ni}}^{\wedge(t-1)})$$

5 展开后的优化函数为：

$$Opt^{(t)} = \sum_{i=1}^n [l(Value_{enviro\text{ni}}, Value_{enviro\text{ni}}^{\wedge(t-1)}) + g_i f_t(Characters_{enviro\text{ni}}) + \frac{1}{2} h_i f_t^2(Characters_{enviro\text{ni}})] + \Omega(f_t) + D$$

由于常数项实质并不影响优化函数的优化过程，故提取出展开后的优化函数中的常数项，可生成展开后的优化函数在第 t 轮的化简函数，公式如下：

$$Simp^{(t)} = \sum_{i=1}^n [g_i f_t(Characters_{enviro\text{ni}}) + \frac{1}{2} h_i f_t^2(Characters_{enviro\text{ni}})] + \Omega(f_t)$$

10 在最终的化简函数中，化简函数得到的输出值依赖于 g_i 和 h_i 的值，故能够很快确定合适的 f （）函数，提升了训练的简便性，在本申请实施例中，通过上述的次序学习的方法以及优化函数（化简函数亦可）以训练预处理模型。

S202：将训练后的所述预处理模型输出为所述筛选模型。

当样本信息集中所有的输入参数和参照参数全部输入预处理模型，并且预处理模型训练完成后，将训练完成的预处理模型作为筛选模型（主要是训练完成的 f （）函数）进行输出。当需要进行待筛选人群的筛选时，将待筛选个体的个体特征 $Characters_{enviro\text{nx}}$ 输入筛选模型，

即可通过筛选模型中优化后的计算公式 $Symptom_{valuex}^{\wedge} = \sum_{k=1}^K f_k(Blood_{valuex})$ ，经过计算得到预测值 $Symptom_{valuex}^{\wedge}$ 。

20 通过图 2 所示实施例可知，在本申请实施例中，将样本信息中的样本环境特征作为预处理模型的输入参数，将样本信息的样本特征值作为预处理模型的参照参数，从而将多个样本信息输入至预处理模型，以训练预处理模型，最后将训练完成的预处理模型作为筛选模型进

行输出，提升了筛选模型与多个样本信息的贴合度，并且提升了通过筛选模型进行人群筛选的准确性。

请参阅图 3，图 3 是本申请实施例三提供的一种基于大数据的人群筛选方法的实现流程图。相对于图 1 对应的实施例，本实施例对 S104 细化后得到 S301~S302，详述如下：

5 S301：获取特征阈值，所述特征阈值用于判断所述输出特征值是否满足所述预设条件。

在本申请实施例中，筛选模型确定后，需要获取待筛选人群的多个个体特征，并自动将多个个体特征输入至筛选模型，得到包含多个输出特征值的输出特征值集合。为了获得输出特征值集合中满足预设条件的输出特征值，首先获取特征阈值，满足预设条件的输出特征值即为大于或等于特征阈值的输出特征值。

10 请参阅图 4，图 4 是本申请实施例四提供的一种基于大数据的人群筛选方法的实现流程图。相对于图 3 对应的实施例，本实施例在样本特征值为第一特征值或第二特征值的基础上，对 S301 细化后得到 S401~S402，详述如下：

S401：对所述多个样本信息进行大数据分析，确定样本特征值取值为所述第一特征值的所述样本信息在所述多个样本信息中占有的数量比例。

15 对于不同的应用场景，样本特征值的取值范围可能会出现不同，在本申请实施例中，以样本特征值的取值为第一特征值或第二特征值，并且第一特征值大于第二特征值的情况进行说明。举例来说，样本特征值取值为第一特征值表示对应的样本个体受到焦虑症影响，样本特征值取值为第二特征值表示对应的样本个体未受到焦虑症影响，而筛选目的是从待筛选人群中筛选出受到焦虑症影响的目标人群，则首先对多个样本信息进行大数据分析，提取出样
20 本特征值取值为第一特征值的第一样本数量，从而得到第一样本数量在所有样本信息中占有的数量比例。

S402：根据所述数量比例、所述第一特征值和所述第二特征值计算出所述特征阈值。

由于样本特征值的取值范围限定于第一特征值和第二特征值，故计算第一特征值和第二特征值之间的差值，并将第一特征值减去该差值与数量比例的乘积，得到特征阈值。举例来说，若第一特征值为 2，第一数量比例为 30%，第二特征值为 1，第一数量比例为 70%，则
25 差值为 $2-1=1$ ，特征阈值为 $2-1 \times 30\%=1.7$ 。当然，上述计算方法仅适用于样本特征值的取值为第一特征值或第二特征值，且第一特征值大于第二特征值的情况，对于其他可能的样本特征值可在上述计算方法的基础上进行延伸，本申请实施例不再进行赘述。

请参阅图 5，图 5 是本申请实施例五提供的一种基于大数据的人群筛选方法的实现流程图。相对于图 3 对应的实施例，本实施例对 S301 细化后得到 S501~S503，详述如下：

S501：将所述多个样本信息的样本环境特征输入至所述筛选模型，并获取所述筛选模型

输出的与所述多个样本信息的样本环境特征对应的多个结果特征值。

在本申请实施例中，当筛选模型生成后，将多个样本信息输入至筛选模型。具体将多个样本信息的样本环境特征作为输入参数输入至筛选模型，但是不将多个样本信息的样本特征值作为筛选模型的参照参数，而是通过筛选模型的计算公式直接对多个样本信息的样本环境特征进行计算得到多个输出参数，即多个与样本环境特征对应的结果特征值。一般来说，样本信息经筛选模型计算后的结果特征值与样本信息原有的样本特征值存在差异。

S502：对所述多个结果特征值进行排序，生成结果特征值序列。

获取到多个结果特征值后，对多个结果特征值按照数值大小进行排序，生成数值序列，数值序列的列首即为数值最大的结果特征值。值得一提的是，若多个结果特征值中的第一结果特征值与第二结果特征值相同，则按照第一结果特征值和第二结果特征值对应的样本信息的输入次序生成特征值序列。具体地，设置次序写入机制，当某一个结果特征值需要写入特征值序列时，判断特征值序列中是否存在与该结果特征值相同的已有结果特征值，若不存在与该结果特征值相同的已有结果特征值，则按照该结果特征值的数值大小查找出特征值序列中最小的大于该结果特征值的已有结果特征值，并在该已有结果特征值后的位置写入该结果特征值；若存在与该结果特征值相同的已有结果特征值，则在该已有结果特征值后的位置写入该结果特征值，若存在多个与该结果特征值相同的已有结果特征值，则查找多个已有结果特征值末尾的已有结果特征值，并在该已有结果特征值后的位置写入该结果特征值。

S503：获取预设的筛选比例，并在所述结果特征值序列中查找出与所述筛选比例对应的所述结果特征值，输出为所述特征阈值。

特征值序列生成后，获取预设的筛选比例，并在特征值序列中查找与该筛选比例对应的筛选位置，将处于该筛选位置的结果特征值输出为特征阈值。举例来说，特征值序列中已写入 300 个结果特征值，而筛选比例为 10%，则筛选位置为第 30 位，则提取特征值序列中位于第 30 位的结果特征值的数值作为特征阈值。可选地，筛选比例可根据大数据分析方法进行确定，例如可统计全国范围的多个全国样本信息，每个全国样本信息包括样本环境特征和样本特征值，首先样本特征值取值为与筛选目的对应的数值的筛选样本数量，从而确定筛选样本数量在所有全国样本信息中占有的比例，将该比例作为筛选比例，通过大数据分析方法可使得生成的特征阈值的普遍适用性较高。当然，筛选比例也可人为预先设定。

参阅图 3，在步骤 S302 中：提取出所述输出特征值集合中大于或等于所述特征阈值的输出特征值，并将提取出的输出特征值添加至所述目标特征值集合。

特征阈值确定后，从输出特征值集合中提取出大于或等于特征阈值的输出特征值，则提取出的输出特征值与筛选的目标个体对应，故将提取出的输出特征值添加至目标特征值集合，

以便后续从待筛选人群中确定与目标特征值对应的目标人群。

通过图 3 所示实施例可知，在本申请实施例中，通过采用不同的方法获取特征阈值，并从输出特征值集合中提取出大于或等于特征阈值的输出特征值，并将提取出的输出特征值添加至目标特征值集合，通过设定特征阈值，提升了目标特征值集合生成的简便性和效率。

5 对应于上文实施例所述的一种基于大数据的人群筛选方法，图 6 示出了本申请实施例提供的一种基于大数据的人群筛选装置的一个结构框图，参照图 6，该人群筛选装置包括：

第一获取单元 61，用于获取样本人群的多个样本信息，所述样本信息包括样本环境特征和样本特征值，所述样本特征值用于描述所述样本信息在所述样本人群中对应样本个体的个体状态；

10 拟合单元 62，用于将所述多个样本信息与预设的预处理模型进行拟合，并将拟合后的所述预处理模型作为筛选模型；

第二获取单元 63，用于获取待筛选人群的多个个体特征，并将所述多个个体特征输入至所述筛选模型，得到与所述多个个体特征对应的输出特征值集合，所述输出特征值集合包括多个输出特征值；

15 目标确定单元 64，用于将所述输出特征值集合中满足预设条件的输出特征值添加至目标特征值集合，并从所述待筛选人群中确定与所述目标特征值集合对应的目标人群。

可选地，所述拟合单元 62，包括：

输入单元，用于将所述多个样本信息输入至所述预处理模型，以训练所述预处理模型，其中，将所述样本信息的样本环境特征作为所述预处理模型的输入参数，将所述样本信息的
20 样本特征值作为所述预处理模型的参照参数；

输出单元，用于将训练后的所述预处理模型输出为所述筛选模型。

可选地，所述目标确定单元 64，包括：

阈值获取单元，用于获取特征阈值，所述特征阈值用于判断所述输出特征值是否满足所述预设条件；

25 提取单元，用于提取出所述输出特征值集合中大于或等于所述特征阈值的输出特征值，并将提取出的输出特征值添加至所述目标特征值集合。

可选地，样本特征值为第一特征值或第二特征值，所述阈值获取单元，包括：

分析单元，用于对所述多个样本信息进行大数据分析，确定样本特征值取值为所述第一特征值的所述样本信息在所述多个样本信息中占有的数量比例；

30 计算单元，用于根据所述数量比例和所述第一特征值和所述第二特征值计算出所述特征阈值。

可选地，所述阈值获取单元，包括：

特征输入单元，用于将所述多个样本信息的样本环境特征输入至所述筛选模型，并获取所述筛选模型输出的与所述多个样本信息的样本环境特征对应的多个结果特征值；

排序单元，用于对所述多个结果特征值进行排序，生成结果特征值序列；

5 阈值输出单元，用于获取预设的筛选比例，并在所述结果特征值序列中查找出与所述筛选比例对应的所述结果特征值，输出为所述特征阈值。

图 7 是本申请实施例提供的终端设备的示意图。如图 7 所示，该实施例的终端设备 7 包括：处理器 70 以及存储器 71，所述存储器 71 中存储有可在所述处理器 70 上运行的计算机可读指令 72，例如基于大数据的人群筛选程序。所述处理器 70 执行所述计算机可读指令 72
10 时实现上述各个基于大数据的人群筛选方法实施例中的步骤，例如图 1 所示的步骤 S101 至 S104。或者，所述处理器 70 执行所述计算机可读指令 72 时实现上述人群筛选装置实施例中各单元的功能，例如图 6 所示单元 61 至 64 的功能。

示例性的，所述计算机可读指令 72 可以被分割成一个或多个模块/单元，所述一个或者多个模块/单元被存储在所述存储器 71 中，并由所述处理器 70 执行，以完成本申请。所述
15 一个或多个模块/单元可以是能够完成特定功能的一系列计算机可读指令段，该指令段用于描述所述计算机可读指令 72 在所述终端设备 7 中的执行过程。例如，所述计算机可读指令 72 可以被分割成第一获取单元、拟合单元、第二获取单元及目标确定单元，各单元具体功能如上所述。

所述终端设备可包括，但不仅限于，处理器 70、存储器 71。本领域技术人员可以理解，
20 图 7 仅仅是终端设备 7 的示例，并不构成对终端设备 7 的限定，可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件，例如所述终端设备还可以包括输入输出设备、网络接入设备、总线等。

所称处理器 70 可以是中央处理单元(Central Processing Unit, CPU)，还可以是其他通用处理器、数字信号处理器 (Digital Signal Processor, DSP)、专用集成电路 (Application Specific
25 Integrated Circuit, ASIC)、现成可编程门阵列 (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、分立硬件组件等。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。

所述存储器 71 可以是所述终端设备 7 的内部存储单元，例如终端设备 7 的硬盘或内存。所述存储器 71 也可以是所述终端设备 7 的外部存储设备，例如所述终端设备 7 上配备的插
30 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，所述存储器 71 还可以既包括所述终端设备 7 的内部存

储单元也包括外部存储设备。所述存储器 71 用于存储所述计算机可读指令以及所述终端设备所需的其它程序和数据。所述存储器 71 还可以用于暂时地存储已经输出或者将要输出的数据。

5 另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。

10 所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用，可以存储在一个计算机可读取存储介质中。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算机设备（可以是个人计算机，服务器，或者网络设备）执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（Read-Only Memory，ROM）、随机存取存储器（Random Access Memory，RAM）、磁碟或者光盘等各种可以存储程序代码的介质。

15 以上所述，以上实施例仅用以说明本申请的技术方案，而非对其限制；尽管参照前述实施例对本申请进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围。

权利要求书

1. 一种基于大数据的人群筛选方法，其特征在于，包括：

5 获取样本人群的多个样本信息，所述样本信息包括样本环境特征和样本特征值，所述样本特征值用于描述所述样本信息在所述样本人群中对应样本个体的个体状态；

将所述多个样本信息与预设的预处理模型进行拟合，并将拟合后的所述预处理模型作为筛选模型；

获取待筛选人群的多个个体特征，并将所述多个个体特征输入至所述筛选模型，得到与所述多个个体特征对应的输出特征值集合，所述输出特征值集合包括多个输出特征值；

10 将所述输出特征值集合中满足预设条件的输出特征值添加至目标特征值集合，并从所述待筛选人群中确定与所述目标特征值集合对应的目标人群。

2. 如权利要求 1 所述的人群筛选方法，其特征在于，所述将所述多个样本信息与预设的预处理模型进行拟合，并将拟合后的所述预处理模型作为筛选模型，包括：

15 将所述多个样本信息输入至所述预处理模型，以训练所述预处理模型，其中，将所述样本信息的样本环境特征作为所述预处理模型的输入参数，将所述样本信息的样本特征值作为所述预处理模型的参照参数；

将训练后的所述预处理模型输出为所述筛选模型。

3. 如权利要求 1 所述的人群筛选方法，其特征在于，所述将所述输出特征值集合中满足预设条件的输出特征值添加至目标特征值集合，包括：

20 获取特征阈值，所述特征阈值用于判断所述输出特征值是否满足所述预设条件；

提取出所述输出特征值集合中大于或等于所述特征阈值的输出特征值，并将提取出的输出特征值添加至所述目标特征值集合。

4. 如权利要求 3 所述的人群筛选方法，其特征在于，所述样本特征值为第一特征值或第二特征值，所述获取特征阈值，包括：

25 对所述多个样本信息进行大数据分析，确定样本特征值取值为所述第一特征值的所述样本信息在所述多个样本信息中占有的数量比例；

根据所述数量比例、所述第一特征值和所述第二特征值计算出所述特征阈值。

5. 如权利要求 3 所述的人群筛选方法，其特征在于，所述获取特征阈值，包括：

30 将所述多个样本信息的样本环境特征输入至所述筛选模型，并获取所述筛选模型输出的与所述多个样本信息的样本环境特征对应的多个结果特征值；

对所述多个结果特征值进行排序，生成结果特征值序列；

获取预设的筛选比例，并在所述结果特征值序列中查找出与所述筛选比例对应的所述结果特征值，输出为所述特征阈值。

6. 一种基于大数据的人群筛选装置，其特征在于，包括：

5 第一获取单元，用于获取样本人群的多个样本信息，所述样本信息包括样本环境特征和样本特征值，所述样本特征值用于描述所述样本信息在所述样本人群中对应样本个体的个体状态；

拟合单元，用于将所述多个样本信息与预设的预处理模型进行拟合，并将拟合后的所述预处理模型作为筛选模型；

10 第二获取单元，用于获取待筛选人群的多个个体特征，并将所述多个个体特征输入至所述筛选模型，得到与所述多个个体特征对应的输出特征值集合，所述输出特征值集合包括多个输出特征值；

目标确定单元，用于将所述输出特征值集合中满足预设条件的输出特征值添加至目标特征值集合，并从所述待筛选人群中确定与所述目标特征值集合对应的目标人群。

15 7. 如权利要求6所述的人群筛选装置，其特征在于，所述拟合单元，包括：

输入单元，用于将所述多个样本信息输入至所述预处理模型，以训练所述预处理模型，其中，将所述样本信息的样本环境特征作为所述预处理模型的输入参数，将所述样本信息的样本特征值作为所述预处理模型的参照参数；

输出单元，用于将训练后的所述预处理模型输出为所述筛选模型。

20 8. 如权利要求6所述的人群筛选装置，其特征在于，所述目标确定单元，包括：

阈值获取单元，用于获取特征阈值，所述特征阈值用于判断所述输出特征值是否满足所述预设条件；

提取单元，用于提取出所述输出特征值集合中大于或等于所述特征阈值的输出特征值，并将提取出的输出特征值添加至所述目标特征值集合。

25 9. 如权利要求8所述的人群筛选装置，其特征在于，所述症状特征值为第一特征值或第二特征值，所述阈值获取单元，包括：

分析单元，用于对所述多个样本信息进行大数据分析，确定样本特征值取值为所述第一特征值的所述样本信息在所述多个样本信息中占有的数量比例；

30 计算单元，用于根据所述数量比例和所述第一特征值和所述第二特征值计算出所述特征阈值。

10. 如权利要求8所述的人群筛选装置，其特征在于，所述阈值获取单元，包括：

特征输入单元，用于将所述多个样本信息的样本环境特征输入至所述筛选模型，并获取所述筛选模型输出的与所述多个样本信息的样本环境特征对应的多个结果特征值；

排序单元，用于对所述多个结果特征值进行排序，生成结果特征值序列；

5 阈值输出单元，用于获取预设的筛选比例，并在所述结果特征值序列中查找出与所述筛选比例对应的所述结果特征值，输出为所述特征阈值。

11. 一种终端设备，其特征在于，包括存储器以及处理器，所述存储器中存储有可在所述处理器上运行的计算机可读指令，所述处理器执行所述计算机可读指令时实现如下步骤：

获取样本人群的多个样本信息，所述样本信息包括样本环境特征和样本特征值，所述样本特征值用于描述所述样本信息在所述样本人群中对应样本个体的个体状态；

10 将所述多个样本信息与预设的预处理模型进行拟合，并将拟合后的所述预处理模型作为筛选模型；

获取待筛选人群的多个个体特征，并将所述多个个体特征输入至所述筛选模型，得到与所述多个个体特征对应的输出特征值集合，所述输出特征值集合包括多个输出特征值；

15 将所述输出特征值集合中满足预设条件的输出特征值添加至目标特征值集合，并从所述待筛选人群中确定与所述目标特征值集合对应的目标人群。

12. 根据权利要求 11 所述的终端设备，其特征在于，所述将所述多个样本信息与预设的预处理模型进行拟合，并将拟合后的所述预处理模型作为筛选模型，包括：

20 将所述多个样本信息输入至所述预处理模型，以训练所述预处理模型，其中，将所述样本信息的样本环境特征作为所述预处理模型的输入参数，将所述样本信息的样本特征值作为所述预处理模型的参照参数；

将训练后的所述预处理模型输出为所述筛选模型。

13. 根据权利要求 11 所述的终端设备，其特征在于，所述将所述输出特征值集合中满足预设条件的输出特征值添加至目标特征值集合，包括：

获取特征阈值，所述特征阈值用于判断所述输出特征值是否满足所述预设条件；

25 提取出所述输出特征值集合中大于或等于所述特征阈值的输出特征值，并将提取出的输出特征值添加至所述目标特征值集合。

14. 根据权利要求 13 所述的终端设备，其特征在于，所述样本特征值为第一特征值或第二特征值，所述获取特征阈值，包括：

30 对所述多个样本信息进行大数据分析，确定样本特征值取值为所述第一特征值的所述样本信息在所述多个样本信息中占有的数量比例；

根据所述数量比例、所述第一特征值和所述第二特征值计算出所述特征阈值。

15. 根据权利要求 13 所述的终端设备，其特征在于，所述获取特征阈值，包括：

将所述多个样本信息的样本环境特征输入至所述筛选模型，并获取所述筛选模型输出的与所述多个样本信息的样本环境特征对应的多个结果特征值；

对所述多个结果特征值进行排序，生成结果特征值序列；

5 获取预设的筛选比例，并在所述结果特征值序列中查找出与所述筛选比例对应的所述结果特征值，输出为所述特征阈值。

16. 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可读指令，其特征在于，所述计算机可读指令被至少一个处理器执行时实现如下步骤：

10 获取样本人群的多个样本信息，所述样本信息包括样本环境特征和样本特征值，所述样本特征值用于描述所述样本信息在所述样本人群中对应样本个体的个体状态；

将所述多个样本信息与预设的预处理模型进行拟合，并将拟合后的所述预处理模型作为筛选模型；

获取待筛选人群的多个个体特征，并将所述多个个体特征输入至所述筛选模型，得到与所述多个个体特征对应的输出特征值集合，所述输出特征值集合包括多个输出特征值；

15 将所述输出特征值集合中满足预设条件的输出特征值添加至目标特征值集合，并从所述待筛选人群中确定与所述目标特征值集合对应的目标人群。

17. 根据权利要求 16 所述的计算机可读存储介质，其特征在于，所述计算机可读指令被至少一个处理器执行时实现如下步骤：

20 将所述多个样本信息输入至所述预处理模型，以训练所述预处理模型，其中，将所述样本信息的样本环境特征作为所述预处理模型的输入参数，将所述样本信息的样本特征值作为所述预处理模型的参照参数；

将训练后的所述预处理模型输出为所述筛选模型。

18. 根据权利要求 16 所述的计算机可读存储介质，其特征在于，所述计算机可读指令被至少一个处理器执行时实现如下步骤：

25 获取特征阈值，所述特征阈值用于判断所述输出特征值是否满足所述预设条件；

提取出所述输出特征值集合中大于或等于所述特征阈值的输出特征值，并将提取出的输出特征值添加至所述目标特征值集合。

19. 根据权利要求 18 所述的计算机可读存储介质，其特征在于，所述样本特征值为第一特征值或第二特征值，所述计算机可读指令被至少一个处理器执行时实现如下步骤：

30 对所述多个样本信息进行大数据分析，确定样本特征值取值为所述第一特征值的所述样本信息在所述多个样本信息中占有的数量比例；

根据所述数量比例、所述第一特征值和所述第二特征值计算出所述特征阈值。

20. 根据权利要求 18 所述的计算机可读存储介质，其特征在于，所述计算机可读指令被至少一个处理器执行时实现如下步骤：

- 5 将所述多个样本信息的样本环境特征输入至所述筛选模型，并获取所述筛选模型输出的与所述多个样本信息的样本环境特征对应的多个结果特征值；
- 对所述多个结果特征值进行排序，生成结果特征值序列；
- 获取预设的筛选比例，并在所述结果特征值序列中查找出与所述筛选比例对应的所述结果特征值，输出为所述特征阈值。

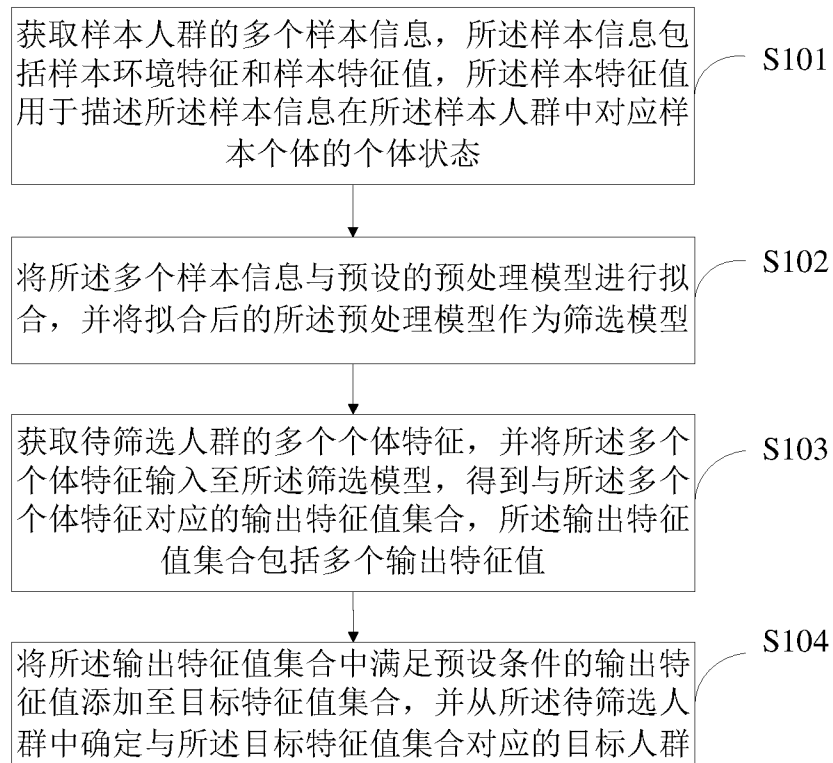


图 1

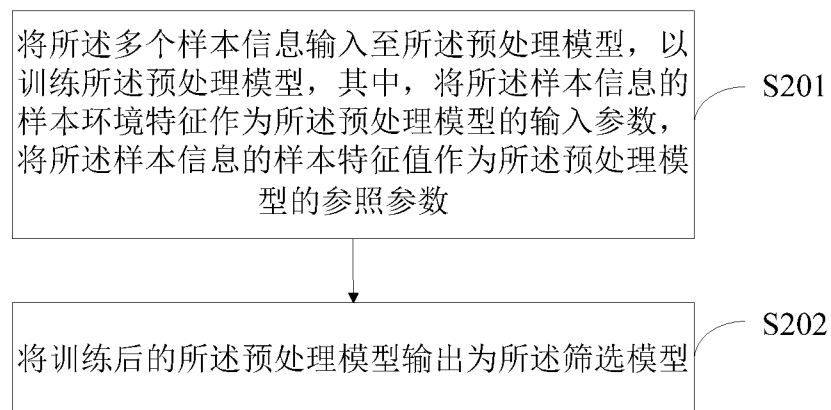


图 2

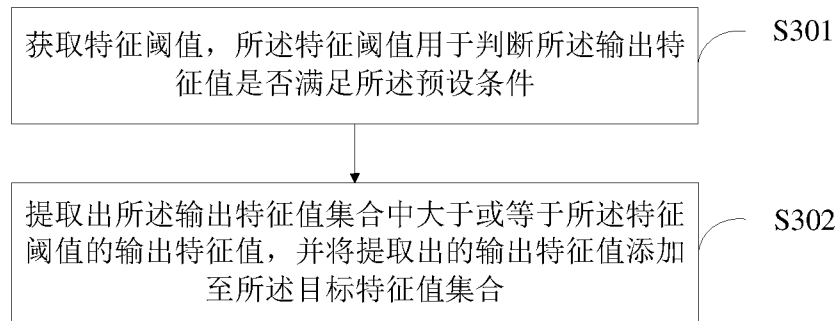


图 3

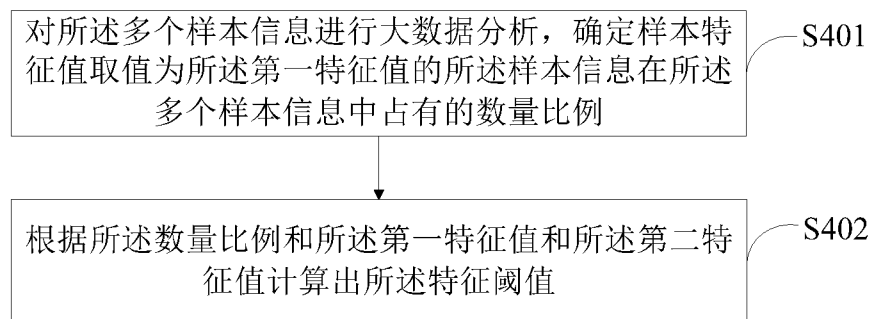


图 4

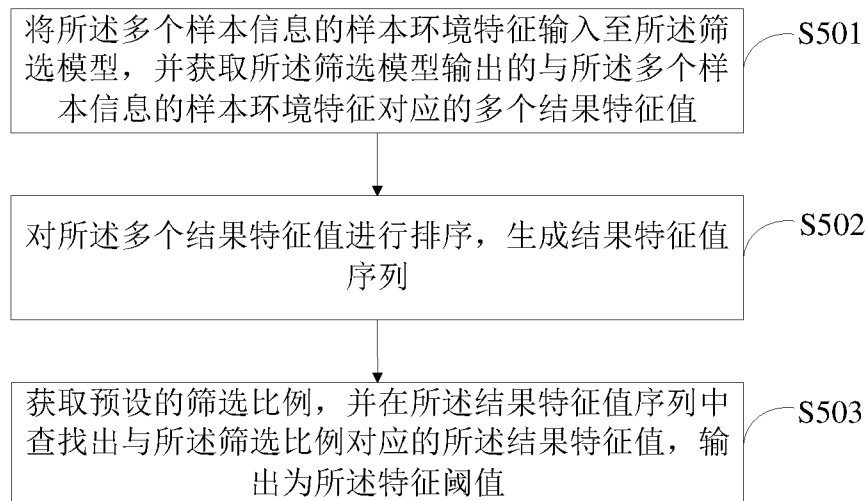


图 5

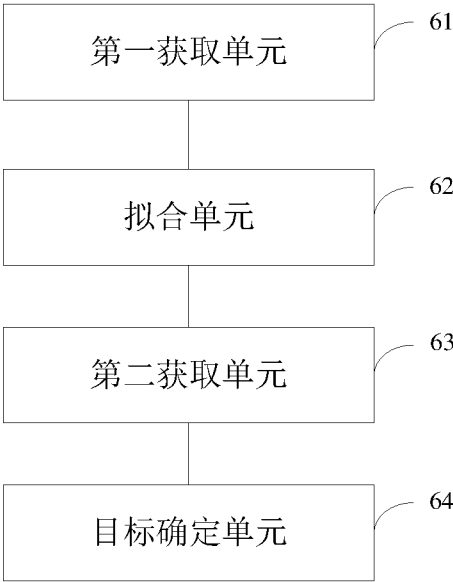


图 6

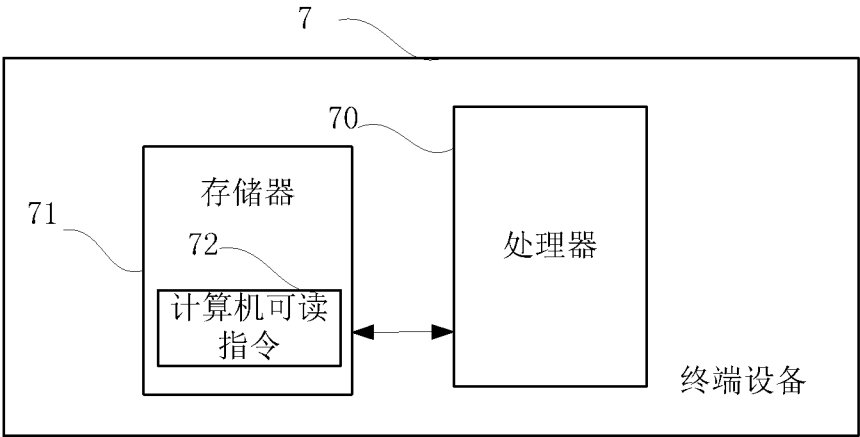


图 7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/097561

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/62(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, EPODOC, CNPAT, CNKI, IEEE: 大数据, 人群, 筛, big data, group, people, filter, determine

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 107895596 A (PING AN TECHNOLOGY SHENZHEN CO., LTD.) 10 April 2018 (2018-04-10) description, paragraphs [0041]-[0067]	1-20
A	CN 106706627 A (WEN, PENG) 24 May 2017 (2017-05-24) entire document	1-20
A	CN 106214120 A (JIN, XIAOLIANG ET AL.) 14 December 2016 (2016-12-14) entire document	1-20

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

23 January 2019

Date of mailing of the international search report

14 February 2019

Name and mailing address of the ISA/CN

National Intellectual Property Administration, PRC (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088
China

Authorized officer

Facsimile No. (86-10)62019451

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/097561

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	107895596	A	10 April 2018	None	
CN	106706627	A	24 May 2017	None	
CN	106214120	A	14 December 2016	None	

国际检索报告

国际申请号

PCT/CN2018/097561

A. 主题的分类

G06K 9/62 (2006.01) i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06K

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

WPI, EPDOC, CNPAT, CNKI, IEEE: 大数据, 人群, 筛, big data, group, people, filter, determine

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 107895596 A (平安科技深圳有限公司) 2018年 4月 10日 (2018 - 04 - 10) 说明书[0041]-[0067]段	1-20
A	CN 106706627 A (温鹏) 2017年 5月 24日 (2017 - 05 - 24) 全文	1-20
A	CN 106214120 A (靳晓亮 等) 2016年 12月 14日 (2016 - 12 - 14) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2019年 1月 23日

国际检索报告邮寄日期

2019年 2月 14日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

受权官员

朱殿尧

电话号码 86-(10)-53961412

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2018/097561

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	107895596	A	2018年 4月 10日	无	
CN	106706627	A	2017年 5月 24日	无	
CN	106214120	A	2016年 12月 14日	无	