

(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

(43) 国际公布日
2020 年 2 月 20 日 (20.02.2020)



(10) 国际公布号
WO 2020/034593 A1

(51) 国际专利分类号:
G06Q 10/06 (2012.01)

(21) 国际申请号: PCT/CN2019/073294

(22) 国际申请日: 2019 年 1 月 27 日 (27.01.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810918521.3 2018 年 8 月 13 日 (13.08.2018) 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: 金戈(JIN, Ge); 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)。 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: 北京中强智尚知识产权代理有限公司 (BEIJING STRONG WISDOM INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市朝阳区北苑东路 19 号院 3 号楼 2501 室, Beijing 100012 (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,

(54) Title: METHOD AND APPARATUS FOR PROCESSING MISSING FEATURE IN CROWD PERFORMANCE FEATURE PREDICTION

(54) 发明名称: 人群绩效特征预测中的缺失特征处理方法及装置

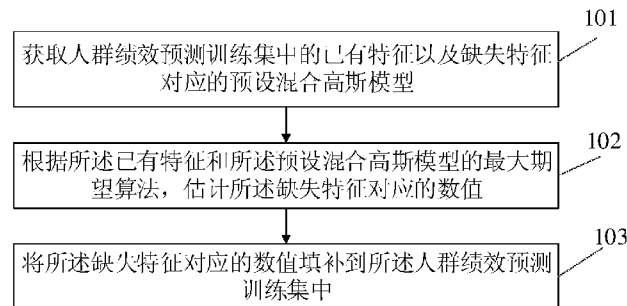


图 1

- 101 Acquire existing features in a crowd performance prediction training set, and a preset mixture Gaussian model corresponding to a missing feature
- 102 According to the existing features and an expected maximum algorithm for the preset mixture Gaussian model, estimate a numerical value corresponding to the missing feature
- 103 Fill in the numerical value corresponding to the missing feature in the crowd performance prediction training set

(57) Abstract: Disclosed are a method and apparatus for processing a missing feature in a crowd performance prediction, wherein same relate to the technical field of information and mainly aim to avoid deviations in the filling-in of a missing feature and avoid deviations in the association between a prediction result of a trained prediction model and a corresponding feature, thereby being capable of improving the prediction precision of the trained prediction model. The method comprises: acquiring existing features in a crowd performance prediction training set, and a preset mixture Gaussian model corresponding to a missing feature, wherein the preset mixture Gaussian model is composed of multivariate Gaussian distributions corresponding to the missing feature; according to the existing features and an expected maximum algorithm for the preset mixture Gaussian model, estimating the missing feature; and filling in the



WO 2020/034593 A1

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)。

根据细则4.17的声明:

- 关于申请人有权申请并被授予专利(细则4.17(ii))
- 发明人资格(细则4.17(iv))

本国际公布:

- 包括国际检索报告(条约第21条(3))。

missing feature in the crowd performance prediction training set. The present application is suitable for processing a missing feature in a crowd performance prediction.

(57) 摘要: 本申请公开了一种人群绩效预测中的缺失特征处理方法及装置, 涉及信息技术领域, 主要目的在于能够避免缺失特征的填补存在偏差, 避免经过训练得预测模型的预测结果与相应特征间的关联存在偏差, 从而能够训练得预测模型的预测精度。所述方法包括: 获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型, 所述预设混合高斯模型由所述缺失特征对应的多元高斯分布组成; 根据所述已有特征和所述预设混合高斯模型的最大期望算法, 估计所述缺失特征; 将所述缺失特征填补到所述人群绩效预测训练集中。本申请适用于人群绩效预测中的缺失特征的处理。

人群绩效特征预测中的缺失特征处理方法及装置

技术领域

本申请要求与2018年8月13日提交中国专利局、申请号为2018109185213、申请名称为“人群绩效特征预测中的缺失特征处理方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在申请中。

背景技术

近年来，很多行业开始重视人群绩效，尤其是开始重视识别人群绩效等级，通过识别出目标人群并对目标人群进行奖励，能够大大促进企业整体效益的提升。目前通常可以通过训练人群绩效预测模型预测人群绩效等级，在训练预测模型时，训练集的数据量越大训练得到的预测模型预测精度越高。然而，在实际应用中，能够用来训练人群绩效等级预测模型的完整训练集较少，大部分情况下训练集会存在缺失特征。因此，为了提升预测模型的预测精度，通常需要填补或者补全缺失特征。

目前，通常采用传统填补方法，如特殊值插补、热卡插补等方式填补人群绩效预测训练集中的缺失特征。然而，上述填补方式的填补效果受限于特征分布状态，且缺失特征通常为非随机缺失特征、非随机缺失特征的缺失与否与特征本身存在关联。例如，在收入调查中，受访人群中高收入人群或者低收入人群都不会填写具体收入，因此收入水平的缺失与收入水平自身存在关联。若通过上述方式填补缺失特征，会造成缺失特征的填补存在偏差，从而造成，进而造成训练得预测模型的预测精度较低。

发明内容

本申请提供了一种人群绩效特征预测中的缺失特征处理方法及装置，主要在于能够避免缺失特征的填补存在偏差，避免经过训练得预测模型的预测结果与相应特征间的关联存在偏差，从而能够训练得预测模型的预测精度。

根据本申请的第一个方面，提供一种人群绩效预测中的缺失特征处理方法，包括：

获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型，所述预设混合高斯模型由所述缺失特征对应的多元高斯分布组成；

根据所述已有特征和所述预设混合高斯模型的最大期望算法，估计所述缺失特征对应的数值；

将所述缺失特征对应的数值填补到所述人群绩效预测训练集中。

根据本申请的第二个方面，提供一种人群绩效预测中的缺失特征处理装置，包括

获取单元，用于获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型，所述预设混合高斯模型由所述缺失特征对应的多元高斯分布组成；

估计单元，用于根据所述已有特征和所述预设混合高斯模型的最大期望算法，估计所述缺失特征对应的数值；

填补单元，用于将所述缺失特征对应的数值填补到所述人群绩效预测训练集中。

根据本申请的第三个方面，提供一种计算机非易失性可读存储介质，其上存储有计算机可读指令，该计算机可读指令被处理器执行时实现以下步骤：

获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型，所述预设混合高斯模型由所述缺失特征对应的多元高斯分布组成；

根据所述已有特征和所述预设混合高斯模型的最大期望算法，估计所述缺失特征对应的数值；

将所述缺失特征对应的数值填补到所述人群绩效预测训练集中。

根据本申请的第四个方面，提供一种计算机设备，包括存储器、处理器及存储在存储器上并可在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以下步骤：

获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型，所述预设混合高斯模型由所述缺失特征对应的多元高斯分布组成；

根据所述已有特征和所述预设混合高斯模型的最大期望算法，估计所述缺失特征对应的数值；

将所述缺失特征对应的数值填补到所述人群绩效预测训练集中。

本申请提供的一种人群绩效特征预测中的缺失特征处理方法及装置，与目前通常采用传统填补方法，如特殊值插补、热卡插补等方式填补人群绩效预测训练集中的缺失特征相比，本申请能够获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型，所述预设混合高斯模型由所述缺失特征对应的多元高斯分布组成；能够根据所述已有特征和所述预设混合高斯模型的最大期望算法，估计所述缺失特征对应的数值。与此同时，能够将估计后的缺失特征对应的数值填补到所述人群绩效预测训练集中，从而能够实现基于缺失特征对应的预设混合高斯模型填补缺失数据，由于所述预设混合高斯模型是由所述缺失特征对应的多元高斯分布组成，因此能够保证填补后的缺失特征反映其分布状态且与本身相关，能够避免缺失特征的填补存在偏差，能够避免经过训练得预测模型的预测结果与相应特征间的关联存在偏差，进而能够提升训练得预测模型的预测精度。

附图说明

此处所说明的附图用来提供对本申请的进一步理解，构成本申请的一部分，本申请的示意性实施例及其说明用于解释本申请，并不构成对本申请的不当限定。在附图中：

图 1 示出了本申请实施例提供的一种人群绩效预测中的缺失特征处理方法流程图；

图 2 示出了本申请实施例提供的另一种人群绩效预测中的缺失特征处理方法流程图；

图 3 示出了本申请实施例提供的一种人群绩效预测中的缺失特征处理装置的结构示意图；

图 4 示出了本申请实施例提供的另一种人群绩效预测中的缺失特征处理装置的结构示意图；

图 5 示出了本申请实施例提供的一种计算机设备的实体结构示意图。

具体实施方式

下文中将参考附图并结合实施例来详细说明本申请。需要说明的是，在不冲突的情况下，本申请中的实施例及实施例中的特征可以相互组合。

如背景技术，目前，通常采用传统填补方法，如特殊值插补、热卡插补等方式填补人群绩效预测训练集中的缺失特征。然而，上述填补方式的填补效果受限于特征分布状态，且缺失特征通常为非随机缺失特征、非随机缺失特征的缺失与否与特征本身存在关联。例如，在收入调查中，受访人群中高收入人群或者低收入人群都不会填写具体收入，因此收入水平的缺失与收入水平自身存在关联。若通过上述方式填补缺失特征，会造成缺失特征的填补存在偏差，从而造成，进而造成训练得预测模型的预测精度较低。

为了解决上述问题，本申请实施例提供了一种人群绩效预测中的缺失特征处理方法，如图 1 所示，所述方法包括：

101、获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型。

其中，所述预设混合高斯模型可以由所述缺失特征对应的多元高斯分布组成。所述训练集可以包括人群绩效特征和人群绩效等级，本申请实施例中的缺失特征可以为人群绩效特征中的特征。所述人群绩效特征可以包括但不限于：月均学习课程数、工作地址经纬度、单月内互联网交易产品数、半年内互联网交易等级。例如，有 100 条训练样本，其中有 40 条训练样本中没有半年内互联网交易等级，则可以确认 40 条训练样本中的“半年内互联网交易等级”为缺失特征。已有特征可以为其中 60 条训练样本中的“月均学习课程数、

工作地址经纬度、单月内互联网交易产品数、半年内互联网交易等级”和其中“40条训练样本中的“月均学习课程数、工作地址经纬度、单月内互联网交易产品数”。

需要说明的是，所述预设混合高斯模型的概率分布可以如下方式表示：

$$p(x) = \sum_{i=1}^m \pi_i N(x_j | \mu_i, \Sigma_i)$$

若本申请实施例中的缺失特征为“半年内互联网交易等级”，则所述多元高斯分布可以为多个半年内互联网交易等级的类别，根据上述公式，半年内互联网交易等级的类别具体可以为 m 个。

102、根据所述已有特征和所述预设混合高斯模型的最大期望算法，估计所述缺失特征对应的数值。

其中，所述最大期望算法可以为一种迭代算法，可以包括极大似然估计计算和期望计算，所述极大似然估计计算和所述期望计算可以为迭代进行的。本申请实施例可以通过已有特征和极大似然估计计算迭代计算所述预设混合高斯模型的各元高斯分布的分布参数，然后通过期望计算和迭代计算出的分布参数，计算所述缺失特征属于各元高斯分布的概率，从而可以估计所述缺失特征所属的高斯分布，确定所述缺失特征对应的数值。

例如，若100条训练样本中有60条训练样本的特征完整，40条训练样本存在特征缺失，则可以根据60条训练样本中的已有特征和预设混合高斯模型估计出40条训练样本中的缺失特征，估计出缺失特征对应的数值后，通过将缺失特征对应的数值填补到所述人群绩效预测训练集，即可以得到特征完整的100条训练样本。

103、将所述缺失特征对应的数值填补到所述人群绩效预测训练集中。

需要说明的是，可以根据已有特征的位置将所述缺失特征对应的数值填补到所述人群绩效预测训练集的相应位置上。例如，训练样本3为（月均学习课程数80，工作地址经纬度（123.435，41.819），单月内互联网交易产品100，），训练样本3中缺失特征为“半年内互联网交易等级”，在估计出训练样本3中的“半年内互联网交易等级”对应的数值为半年内互联网交易等级2时，可以根据“月均学习课程数80，工作地址经纬度（123.435，41.819），单月内互联网交易产品100”，将所述“半年内互联网交易等级2”填补到训练样本3中，得到（月均学习课程数80，工作地址经纬度（123.435，41.819），单月内互联网交易产品100，半年内互联网交易等级2）。

本申请实施例提供的一种人群绩效特征预测中的缺失特征处理方法，与目前通常采用传统填补方法，如特殊值插补、热卡插补等方式填补人群绩效预测训练集中的缺失特征相比，本申请实施例能够获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型，所述预设混合高斯模型由所述缺失特征对应的多元高斯分布组成；能够根据

所述已有特征和所述预设混合高斯模型的最大期望算法，估计所述缺失特征。与此同时，能够将估计后的缺失特征填补到所述人群绩效预测训练集中，从而能够实现基于缺失特征对应的预设混合高斯模型填补缺失数据，由于所述预设混合高斯模型是由所述缺失特征对应的多元高斯分布组成，因此能够保证填补后的缺失特征反映其分布状态且与本身相关，能够避免缺失特征的填补存在偏差，能够避免经过训练得预测模型的预测结果与相应特征间的关联存在偏差，进而能够提升训练得预测模型的预测精度。

进一步的，为了更好的说明上述人群绩效预测中的缺失特征的过程，作为对上述实施例的细化和扩展，本申请实施例提供了另一种人群绩效预测中的缺失特征处理方法，如图2所示，所述方法包括：

201、获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型。

其中，所述预设混合高斯模型可以由所述缺失特征对应的多元高斯分布组成。在本申请实施例中，所述已有特征可以以特征向量的形式存在，若“月均学习课程数、工作地址经纬度、单月内互联网交易产品数”三个维度的特征，唯一标定一个半年内互联网交易等级类别，则可以将“月均学习课程数、工作地址经纬度、单月内互联网交易产品数”以特征向量的形式表示。

对于本申请实施例，为了得到所述预设混合高斯模型，所述方法还可以包括：确定所述缺失特征对应的多元高斯分布；根据所述多元高斯分布构建所述预设混合高斯模型。

例如，所述缺失特征为“半年内互联网交易等级”，确定“半年内互联网交易等级”对应的多个类别，每个类别可以对应一元高斯分布，可以将“月均学习课程数、工作地址经纬度、单月内互联网交易产品数”作为观测样本向量，即可以以所述观测样本向量观测缺失特征的多元高斯分布，将半年内互联网交易等级划分为多个类别，然后根据观测样本向量所属类别的权重、对应的均值向量、协方差矩阵，构建所述半年内互联网交易等级对应的预设混合高斯模型。

202、根据所述已有特征和所述预设混合高斯模型的最大期望算法的极大似然估计计算，估计所述预设混合高斯模型的各元高斯分布的初始分布参数。

其中，所述分布参数可以包括混合系数、均值和协方差。若所述已有特征以特征向量形式存在，所述分布参数可以包括混合系数、均值向量和协方差矩阵，所述混合系数可以为属于对应高斯分布的样本数与总样本数的比值。所述训练集可以包括特征完整的第一训练集和含缺失特征的第二训练集，所述已有特征包括所述第一训练集中第一已有特征、第二已有特征和所述第二训练集中的第三已有特征，所述第一已有特征与所述第三已有特征相对应，所述第二已有特征与所述缺失特征相对应，所述步骤202具体可以包括：根据所

述第一已有特征、所述第二已有特征和所述极大似然估计计算，估计所述预设混合高斯模型的各元高斯分布的初始混合系数、初始均值及初始协方差。

需要说明的是，极大似然估计计算可以包括：

$$\text{计算第 } i \text{ 元高斯分布的混合系数: } \pi_i = \frac{\sum_{j=1}^l \gamma_{ij}}{l}$$

$$\text{计算第 } i \text{ 元高斯分布的均值: } \mu_i = \frac{1}{\sum_{j=1}^l \gamma_{ij}} \sum_{j=1}^l \gamma_{ij} x_j$$

$$\text{计算第 } i \text{ 元高斯分布的协方差: } \sum_i = \frac{1}{\sum_{j=1}^l \gamma_{ij}} \sum_{j=1}^l \gamma_{ij} (x_j - \mu_i)(x_j - \mu_i)^T$$

γ_{ij} 可以表示为样本 x_j 属于第 i 元高斯分布的概率。

例如，训练集为 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l), x_{l+1}, x_{l+2}, \dots, x_n\}$ ，特征完整的第一训练集可以为 $D_1 = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$ ，含缺失特征的第二训练集可以为 $D_2 = \{x_{l+1}, x_{l+2}, \dots, x_n\}$ ，所述第一训练集中第一已有特征可以为 $x_j, j=1, \dots, l$ ，第二已有特征可以为 $y_i, i=1, \dots, l$ ，所述第二训练集中的第三已有特征可以为 $x_j, j=l+1, \dots, n$ ，所述第二训练集中的缺失特征可以为 $y_i, i=l+1, \dots, n$ ；具体 x_j 可以对应（月均学习课程数、工作地址经纬度、单月内互联网交易产品数）， y_i 可以对应 x_j 属于第 i 类的“半年内互联网交易等级”的概率 γ_{ij} ， x_j 所属的类的概率值为 1，其余类为 0。因此，可以将第一已有特征“ x_j ”和第二已有特征“ γ_{ij} ”分别代入极大似然估计计算，计算各元高斯分布的初始混合系数、初始均值及初始协方差。

203、根据所述初始分布参数及所述最大期望算法的期望计算，初步估计所述缺失特征所属的高斯分布。

对于本申请实施例，与所述步骤 202 相对应的，所述步骤 203 具体可以包括：所述初始混合系数、初始均值、初始协方差、所述第三已有特征及所述期望计算，初步估计所述缺失特征所属的高斯分布。具体地，可以根据初始混合系数、初始均值、初始协方差、所述第三已有特征及所述期望计算，初步估计所述缺失特征属于各元高斯分布的概率；根据所述属于各元高斯分布的概率，初步估计所述缺失特征所属的高斯分布。

需要说明的是，期望计算可以包括：

$$\gamma_{ij} = \frac{\pi_i N(x_j | \mu_i, \sum_i)}{\sum_{i=1}^m \pi_i N(x_j | \mu_i, \sum_i)}$$

其中, m 可以为高斯分布的总元数。

例如, 接着步骤 202 所述的例子, 在计算出初始参数 π_i 、 μ_i 、 $\sum_i$ 后, 可以将 x_j , $j=1, \dots, n$ 、 π_i 、 μ_i 、 $\sum_i$, 分别代入到上述公式, 分别计算“半年内互联网交易等级”属于各元高斯分布的概率, 即属于各个“半年内互联网交易等级”类别的概率 γ_{ij} ; 可以将对应概率最高的高斯分布确定为“半年内互联网交易等级”的高斯分布。

204、根据所述已有特征、所述所属的高斯分布和所述极大似然估计计算, 迭代更新所述各元高斯分布的分布参数, 并迭代估计所述缺失特征所属的高斯分布。

需要说明的是, 根据步骤 204 能够实现根据所有特征迭代更新所述各元高斯分布的分布参数, 即极大似然估计计算中的上限值由 l 变为 n :

$$\text{计算第 } i \text{ 元高斯分布的混合系数: } \pi_i = \frac{\sum_{j=1}^n \gamma_{ij}}{n}$$

$$\text{计算第 } i \text{ 元高斯分布的均值: } \mu_i = \frac{1}{\sum_{j=1}^n \gamma_{ij}} \sum_{j=1}^n \gamma_{ij} x_j$$

$$\text{计算第 } i \text{ 元高斯分布的协方差: } \sum_i = \frac{1}{\sum_{j=1}^n \gamma_{ij}} \sum_{j=1}^n \gamma_{ij} (x_j - \mu_i)(x_j - \mu_i)^T$$

205、当迭代更新的分布参数收敛时, 获取根据收敛的分布参数和所述期望计算估计的高斯分布, 并根据所述估计的高斯分布估计所述缺失特征对应的数值。

对于本申请实施例, 为了确定更新的分布参数是否收敛, 所述方法还可以包括: 计算前后两次迭代更新的分布参数的参数差值; 若所述参数差值小于预设阈值, 则确定更新的分布参数收敛。

具体地, 若分布参数为混合系数、均值、协方差, 可以计算前后两次迭代更新的混合系数的差值; 若所述混合系数差值小于预设混合系数阈值, 则可以确定混合系数收敛。或者计算前后两次迭代更新的均值的差值; 若所述均值差值小于预设均值阈值, 则可以确定均值收敛。或者计算前后两次迭代更新的协方差的差值; 若所述协方差差值小于预设协方差阈值, 则可以确定协方差收敛。即在上述任何一个参数收敛时, 可以停止迭代极大似然估计计算和期望计算, 此时, 根据最后一次迭代估计缺失特征所属的高斯分布, 估计所述缺失特征。所述预设混合系数阈值、所述预设均值阈值和所述预设协方差阈值均可以为根据用户需求设置的, 也可以为根据系统默认模式设置的, 本申请实施例在此不进行限定。

206、将所述缺失特征对应的数值填补到所述人群绩效预测训练集中，并根据填补缺失特征对应的数值后的人群绩效预测训练集训练人群绩效预测模型。

需要说明的是，可以根据已有特征的位置，确定缺失特征对应的数值填补到所述人群绩效预测训练集中的位置，若填补缺失特征对应的数值之前的训练集为 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l), x_{l+1}, x_{l+2}, \dots, x_n\}$ ，则可以分别根据 $x_{l+1}, x_{l+2}, \dots, x_n$ 的位置填补 $y_{l+1}, y_{l+2}, \dots, y_n$ ，最后得到 $y_{l+1}, y_{l+2}, \dots, y_n$ 缺失特征后的训练集可以为 $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l), (x_{l+1}, y_{l+1}), (x_{l+2}, y_{l+2}), \dots, (x_n, y_n)\}$ 。此外，在本申请实施例中，所述人群绩效预测模型可以为确定人群绩效等级的决策树模型或者逻辑回归模型等。具体地，若人群绩效预测模型为决策树模型，可以利用决策树算法对填补缺失特征后的人群绩效预测训练集进行训练，得到决策树模型。若人群绩效预测模型为逻辑回归模型，可以利用逻辑回归模型算法对填补缺失特征后的人群绩效预测训练集进行训练，得到逻辑回归模型模型。

为了更好的理解本申请实施例，提供如下应用场景，包括但不限于此：假设人群绩效特征预测训练集有训练样本 1：（月均学习课程数 100，半年内互联网交易等级 1），训练样本 2（月均学习课程数 50，半年内互联网交易等级 2），训练样本 3（月均学习课程数 60，半年内互联网交易等级 2），训练样本 4（月均学习课程数 80，），训练样本 5（月均学习课程数 70，），则缺失特征为训练样本 4 和训练样本 5 中的“半年内互联网交易等级”，半年内互联网交易等级有 2 个类别，分别为：半年内互联网交易等级 1、半年内互联网交易等级 2。

首先，可以将训练样本 1、训练样本 2、训练样本 3 中的“月均学习课程数”和“半年内互联网交易等级”分别代入到极大似然估计计算中得到初步估计的 π_i 、 μ_i 、 $\sum_i$ ，然后将训练样本 4 和训练样本 5 的“月均学习课程数”分别代入到期望计算中，计算训练样本 4 中的“半年内互联网交易等级”分别属于半年内互联网交易等级 1 的概率，属于半年内互联网交易等级 2 的概率，若属于半年内互联网交易等级 1 的概率大于属于半年内互联网交易等级 2 的概率，则可以确定训练样本 4 中的“半年内互联网交易等级”所属的类别为：半年内互联网交易等级 1。同理地，可以计算训练样本 5 中的“半年内互联网交易等级”所属的类别为：半年内互联网交易等级 1。

然后，可以将训练样本 1、训练样本 2、训练样本 3、训练样本 4 和训练样本 5 的“月均学习课程数”和“半年内互联网交易等级”分别代入到极大似然估计计算中，迭代更新 π_i 、 μ_i 、 $\sum_i$ ，并根据更新的 π_i 、 μ_i 、 $\sum_i$ 和期望计算分别更新训练样本 4 和训练样本 5

中“半年内互联网交易等级”所属的类别，直到 π_i 、 μ_i 、 $\sum_i$ 收敛时，将估计的“半年内互联网交易等级”所属的类别确定为最后估计结果，如最终确定：训练样本4中的“半年内互联网交易等级”所属的类别为：半年内互联网交易等级1；训练样本5中的“半年内互联网交易等级”所属的类别为：半年内互联网交易等级2。

因此，补入缺失特征对应的数值后，得到的训练集可以为：训练样本1：（月均学习课程数100，半年内互联网交易等级1），训练样本2（月均学习课程数50，半年内互联网交易等级2），训练样本3（月均学习课程数60，半年内互联网交易等级2），训练样本4（月均学习课程数80，半年内互联网交易等级1），训练样本5（月均学习课程数70，半年内互联网交易等级2）。

本申请实施例提供的另一种人群绩效特征预测中的缺失特征处理方法，与目前通常采用传统填补方法，如特殊值插补、热卡插补等方式填补人群绩效预测训练集中的缺失特征相比，本申请实施例能够获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型，所述预设混合高斯模型由所述缺失特征对应的多元高斯分布组成；能够根据所述已有特征和所述预设混合高斯模型的最大期望算法，估计所述缺失特征。与此同时，能够将估计后的缺失特征填补到所述人群绩效预测训练集中，从而能够实现基于缺失特征对应的预设混合高斯模型填补缺失数据，由于所述预设混合高斯模型是由所述缺失特征对应的多元高斯分布组成，因此能够保证填补后的缺失特征反映其分布状态且与本身相关，能够避免缺失特征的填补存在偏差，能够避免经过训练得预测模型的预测结果与相应特征间的关联存在偏差，进而能够提升训练得预测模型的预测精度。

进一步地，作为图1的具体实现，本申请实施例提供了一种人群绩效预测中的缺失特征处理装置，如图3所示，所述装置包括：获取单元31、估计单元32和填补单元33。

所述获取单元31，可以用于获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型，所述预设混合高斯模型由所述缺失特征对应的多元高斯分布组成。所述获取单元31是本装置中获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型的主要功能模块。

所述估计单元32，可以用于根据所述已有特征和所述预设混合高斯模型的最大期望算法，估计所述缺失特征对应的数值。所述估计单元32是本装置中根据所述已有特征和所述预设混合高斯模型的最大期望算法，估计所述缺失特征的主要功能模块，也是核心模块。

所述填补单元33，可以用于将所述缺失特征对应的数值填补到所述人群绩效预测训练集中。所述填补单元33是本装置中将所述缺失特征对应的数值填补到所述人群绩效预测训练集中的主要功能模块。

对于本申请实施例，所述最大期望算法可以包括极大似然估计计算和期望计算，所述估计单元 32 可以包括：估计模块 321、更新模块 322 和获取模块 323，如图 4 所示。

所述估计模块 321，可以用于根据所述已有特征和所述极大似然估计计算，估计所述预设混合高斯模型的各元高斯分布的初始分布参数。

所述估计模块 321，还可以用于根据所述初始分布参数及所述期望计算，初步估计所述缺失特征所属的高斯分布。

所述更新模块 322，可以用于根据所述已有特征、所述所属的高斯分布和所述极大似然估计计算，迭代更新所述各元高斯分布的分布参数，并迭代更新所述缺失特征所属的高斯分布。

所述获取模块 323，可以用于当迭代更新的分布参数收敛时，获取根据收敛的分布参数和所述期望计算估计的高斯分布。

所述估计模块 321，还可以用于根据最后估计的高斯分布估计所述缺失特征对应的数值。

在具体应用场景中，所述估计模块 321，具体可以用于根据所述初始分布参数及所述期望计算，初步估计所述缺失特征属于各元高斯分布的概率；并根据所述属于各元高斯分布的概率，初步估计所述缺失特征所属的高斯分布。

需要说明的是，所述训练集包括特征完整的第一训练集和含缺失特征的第二训练集，所述已有特征包括所述第一训练集中第一已有特征、第二已有特征和所述第二训练集中的第三已有特征，所述第一已有特征与所述第三已有特征相对应，所述第二已有特征与所述缺失特征相对应；所述估计模块 321，具体可以用于根据所述第一已有特征、所述第二已有特征和所述极大似然估计计算，估计所述预设混合高斯模型的各元高斯分布的初始混合系数、初始均值及初始协方差。

所述估计模块 321，具体还可以用于所述初始混合系数、初始均值、初始协方差、所述第三已有特征及所述期望计算，初步估计所述缺失特征所属的高斯分布。

对于本申请实施例，为了确定更新的分布参数是否收敛，所述估计单元 32 还可以包括：计算模块 324 和确定模块 325。

所述计算模块 324，可以用于计算前后两次迭代更新的分布参数的参数差值。

所述确定模块 325，可以用于若所述参数差值小于预设阈值，则确定更新的分布参数收敛。

对于本申请实施例，为了获取缺失特征对应的预设混合高斯模型，所述装置还可以包括：确定单元 34 和构建单元 35。

所述确定单元 34, 可以用于确定所述缺失特征对应的多元高斯分布。所述确定单元是本装置中确定所述缺失特征对应的多元高斯分布的主要功能模块。

所述构建单元 35, 可以用于根据所述多元高斯分布构建所述预设混合高斯模型。所述构建单元 35 是本装置中根据所述多元高斯分布构建所述预设混合高斯模型是的主要功能模块。

此外, 为了得到人群绩效预测模型, 所述装置还可以包括: 训练单元 36。

所述训练单元 36, 可以用于根据填补缺失特征后的人群绩效预测训练集训练人群绩效预测模型。所述训练单元 36 是本装置中根据填补缺失特征后的人群绩效预测训练集训练人群绩效预测模型的主要功能模块。

需要说明的是, 本申请实施例提供的一种人群绩效预测中的缺失特征处理装置所涉及各功能模块的其他相应描述, 可以参考图 1 所示方法的对应描述, 在此不再赘述。

基于上述如图 1 所示方法, 相应的, 本申请实施例还提供了一种计算机非易失性可读存储介质, 其上存储有计算机可读指令, 该计算机可读指令被处理器执行时实现以下步骤: 获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型, 所述预设混合高斯模型由所述缺失特征对应的多元高斯分布组成; 根据所述已有特征和所述预设混合高斯模型的最大期望算法, 估计所述缺失特征对应的数值; 将所述缺失特征对应的数值填补到所述人群绩效预测训练集中。

基于上述如图 1 所示方法和如图 3 所示装置的实施例, 本申请实施例还提供了一种计算机设备的实体结构图, 如图 5 所示, 该计算机设备包括: 处理器 41、存储器 42、及存储在存储器 42 上并可在处理器上运行的计算机可读指令, 其中存储器 42 和处理器 41 均设置在总线 43 上所述处理器 41 执行所述计算机可读指令时实现以下步骤: 获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型, 所述预设混合高斯模型由所述缺失特征对应的多元高斯分布组成; 根据所述已有特征和所述预设混合高斯模型的最大期望算法, 估计所述缺失特征对应的数值; 将所述缺失特征对应的数值填补到所述人群绩效预测训练集中。

通过本申请的技术方案, 能够获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型, 所述预设混合高斯模型由所述缺失特征对应的多元高斯分布组成; 能够根据所述已有特征和所述预设混合高斯模型的最大期望算法, 估计所述缺失特征对应的数值。与此同时, 能够将估计后的缺失特征对应的数值填补到所述人群绩效预测训练集中, 从而能够实现基于缺失特征对应的预设混合高斯模型填补缺失数据, 由于所述预设混合高斯模型是由所述缺失特征对应的多元高斯分布组成, 因此能够保证填补后的缺失

特征反映其分布状态且与本身相关，能够避免缺失特征的填补存在偏差，能够避免经过训练得预测模型的预测结果与相应特征间的关联存在偏差，进而能够提升训练得预测模型的预测精度。

显然，本领域的技术人员应该明白，上述的本申请的各模块或各步骤可以用通用的计算装置来实现，它们可以集中在单个的计算装置上，或者分布在多个计算装置所组成的网络上，可选地，它们可以用计算装置可执行的计算机可读指令代码来实现，从而，可以将它们存储在存储装置中由计算装置来执行，并且在某些情况下，可以以不同于此处的顺序执行所示出或描述的步骤，或者将它们分别制作成各个集成电路模块，或者将它们中的多个模块或步骤制作成单个集成电路模块来实现。这样，本申请不限制于任何特定的硬件和软件结合。

以上所述仅为本申请的优选实施例而已，并不用于限制本申请，对于本领域的技术人员来说，本申请可以有各种更改和变化。凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包括在本申请的保护范围之内。

权 利 要 求

1、一种人群绩效预测中的缺失特征处理方法，其特征在于，包括：

获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型，所述预设混合高斯模型由所述缺失特征对应的多元高斯分布组成；

根据所述已有特征和所述预设混合高斯模型的最大期望算法，估计所述缺失特征对应的数值；

将所述缺失特征对应的数值填补到所述人群绩效预测训练集中。

2、根据权利要求1所述的方法，其特征在于，所述最大期望算法包括极大似然估计计算和期望计算，根据所述已有特征和所述预设混合高斯模型的最大期望算法，估计所述缺失特征对应的数值，包括：

根据所述已有特征和所述极大似然估计计算，估计所述预设混合高斯模型的各元高斯分布的初始分布参数；

根据所述初始分布参数及所述期望计算，初步估计所述缺失特征所属的高斯分布；

根据所述已有特征、所述所属的高斯分布和所述极大似然估计计算，迭代更新所述各元高斯分布的分布参数，并迭代更新所述缺失特征所属的高斯分布；

当迭代更新的分布参数收敛时，获取根据收敛的分布参数和所述期望计算估计的高斯分布，并根据最后估计的高斯分布估计所述缺失特征对应的数值。

3、根据权利要求2所述的方法，其特征在于，所述根据所述初始分布参数及所述期望计算，初步估计所述缺失特征所属的高斯分布，包括：

根据所述初始分布参数及所述期望计算，初步估计所述缺失特征属于各元高斯分布的概率；

根据所述属于各元高斯分布的概率，初步估计所述缺失特征所属的高斯分布。

4、根据权利要求2所述的方法，其特征在于，所述训练集包括特征完整的第一训练集和含缺失特征的第二训练集，所述已有特征包括所述第一训练集中第一已有特征、第二已有特征和所述第二训练集中的第三已有特征，所述第一已有特征与所述第三已有特征相对应，所述第二已有特征与所述缺失特征相对应；所述根据所述已有特征和所述极大似然估计计算，估计所述预设混合高斯模型的各

元高斯分布的初始分布参数, 包括:

根据所述第一已有特征、所述第二已有特征和所述极大似然估计计算, 估计所述预设混合高斯模型的各元高斯分布的初始混合系数、初始均值及初始协方差;

所述根据所述初始分布参数及所述期望计算, 初步估计所述缺失特征所属的高斯分布, 包括:

根据所述初始混合系数、初始均值、初始协方差、所述第三已有特征及所述期望计算, 初步估计所述缺失特征所属的高斯分布。

5、根据权利要求2所述的方法, 其特征在于, 所述当迭代更新的分布参数收敛时, 获取根据收敛的分布参数和所述期望计算估计的高斯分布之前, 所述方法还包括:

计算前后两次迭代更新的分布参数的参数差值;

若所述参数差值小于预设阈值, 则确定更新的分布参数收敛。

6、根据权利要求1所述的方法, 其特征在于, 所述获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型之前, 所述方法还包括:

确定所述缺失特征对应的多元高斯分布;

根据所述多元高斯分布构建所述预设混合高斯模型。

7、根据权利要求1所述的方法, 其特征在于, 所述方法将所述缺失特征填补到所述人群绩效预测训练集中之后, 所述方法还包括:

根据填补缺失特征对应的数值后的人群绩效预测训练集训练人群绩效预测模型。

8、一种人群绩效预测中的缺失特征处理装置, 其特征不在于, 包括:

获取单元, 用于获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型, 所述预设混合高斯模型由所述缺失特征对应的多元高斯分布组成;

估计单元, 用于根据所述已有特征和所述预设混合高斯模型的最大期望算法, 估计所述缺失特征对应的数值;

填补单元, 用于将所述缺失特征对应的数值填补到所述人群绩效预测训练集中。

9、根据权利要求8所述的装置, 其特征不在于, 所述最大期望算法包括极大

似然估计计算和期望计算, 所述估计单元包括:

估计模块, 用于根据所述已有特征和所述极大似然估计计算, 估计所述预设混合高斯模型的各元高斯分布的初始分布参数;

所述估计模块, 还用于根据所述初始分布参数及所述期望计算, 初步估计所述缺失特征所属的高斯分布;

更新模块, 用于根据所述已有特征、所述所属的高斯分布和所述极大似然估计计算, 迭代更新所述各元高斯分布的分布参数, 并迭代更新所述缺失特征所属的高斯分布;

获取模块, 用于当迭代更新的分布参数收敛时, 获取根据收敛的分布参数和所述期望计算估计的高斯分布;

所述估计模块, 还可以用于根据最后估计的高斯分布估计所述缺失特征对应的数值。

10、根据权利要求 9 所述的装置, 其特征在于, 所述估计模块, 具体用于根据所述初始分布参数及所述期望计算, 初步估计所述缺失特征属于各元高斯分布的概率; 并根据所述属于各元高斯分布的概率, 初步估计所述缺失特征所属的高斯分布。

11、根据权利要求 9 所述的装置, 其特征在于, 所述训练集包括特征完整的第一训练集和含缺失特征的第二训练集, 所述已有特征包括所述第一训练集中第一已有特征、第二已有特征和所述第二训练集中的第三已有特征, 所述第一已有特征与所述第三已有特征相对应, 所述第二已有特征与所述缺失特征相对应; 所述估计模块, 具体用于根据所述第一已有特征、所述第二已有特征和所述极大似然估计计算, 估计所述预设混合高斯模型的各元高斯分布的初始混合系数、初始均值及初始协方差。所述估计模块, 具体还用于根据所述初始混合系数、初始均值、初始协方差、所述第三已有特征及所述期望计算, 初步估计所述缺失特征所属的高斯分布。

12、根据权利要求 9 所述的装置, 其特征在于, 所述估计单元包括:

计算模块, 用于计算前后两次迭代更新的分布参数的参数差值;

确定模块, 用于若所述参数差值小于预设阈值, 则确定更新的分布参数收敛。

13、根据权利要求 8 所述的装置, 其特征在于, 所述装置还包括:

确定单元，用于确定所述缺失特征对应的多元高斯分布；

构建单元，用于根据所述多元高斯分布构建所述预设混合高斯模型。

14、根据权利要求 13 所述的装置，其特征在于，所述装置还包括：

训练单元，用于根据填补缺失特征对应的数值后的人群绩效预测训练集训练人群绩效预测模型。

15、一种计算机非易失性可读存储介质，其上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处理器执行时实现人群绩效预测中的缺失特征处理方法，包括：

获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型，所述预设混合高斯模型由所述缺失特征对应的多元高斯分布组成；根据所述已有特征和所述预设混合高斯模型的最大期望算法，估计所述缺失特征对应的数值；将所述缺失特征对应的数值填补到所述人群绩效预测训练集中。

16、根据权利要求 15 所述的计算机非易失性可读存储介质，其特征在于，所述最大期望算法包括极大似然估计计算和期望计算，所述计算机可读指令被处理器执行时实现所述根据所述已有特征和所述预设混合高斯模型的最大期望算法，估计所述缺失特征对应的数值，包括：根据所述已有特征和所述极大似然估计计算，估计所述预设混合高斯模型的各元高斯分布的初始分布参数；根据所述初始分布参数及所述期望计算，初步估计所述缺失特征所属的高斯分布；根据所述已有特征、所述所属的高斯分布和所述极大似然估计计算，迭代更新所述各元高斯分布的分布参数，并迭代更新所述缺失特征所属的高斯分布；当迭代更新的分布参数收敛时，获取根据收敛的分布参数和所述期望计算估计的高斯分布，并根据最后估计的高斯分布估计所述缺失特征对应的数值。

17、根据权利要求 15 所述的计算机非易失性可读存储介质，其特征在于，所述处理器执行所述计算机可读指令时实现所述方法还包括：根据填补缺失特征对应的数值后的人群绩效预测训练集训练人群绩效预测模型。

18、一种计算机设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机可读指令，其特征在于，所述处理器执行所述计算机可读指令时实现人群绩效预测中的缺失特征处理方法，包括：

获取人群绩效预测训练集中的已有特征以及缺失特征对应的预设混合高斯模型，所述预设混合高斯模型由所述缺失特征对应的多元高斯分布组成；根据所

述已有特征和所述预设混合高斯模型的最大期望算法,估计所述缺失特征对应的数值;将所述缺失特征对应的数值填补到所述人群绩效预测训练集中。

19、根据权利要求 18 所述的计算机设备,其特征在于,所述最大期望算法包括极大似然估计计算和期望计算,所述计算机可读指令被处理器执行时实现所述根据所述已有特征和所述预设混合高斯模型的最大期望算法,估计所述缺失特征对应的数值,包括:根据所述已有特征和所述极大似然估计计算,估计所述预设混合高斯模型的各元高斯分布的初始分布参数;根据所述初始分布参数及所述期望计算,初步估计所述缺失特征所属的高斯分布;根据所述已有特征、所述所属的高斯分布和所述极大似然估计计算,迭代更新所述各元高斯分布的分布参数,并迭代更新所述缺失特征所属的高斯分布;当迭代更新的分布参数收敛时,获取根据收敛的分布参数和所述期望计算估计的高斯分布,并根据最后估计的高斯分布估计所述缺失特征对应的数值。

20、根据权利要求 18 所述的计算机设备,其特征在于,所述处理器执行所述计算机可读指令时实现所述方法还包括:根据填补缺失特征对应的数值后的人群绩效预测训练集训练人群绩效预测模型。

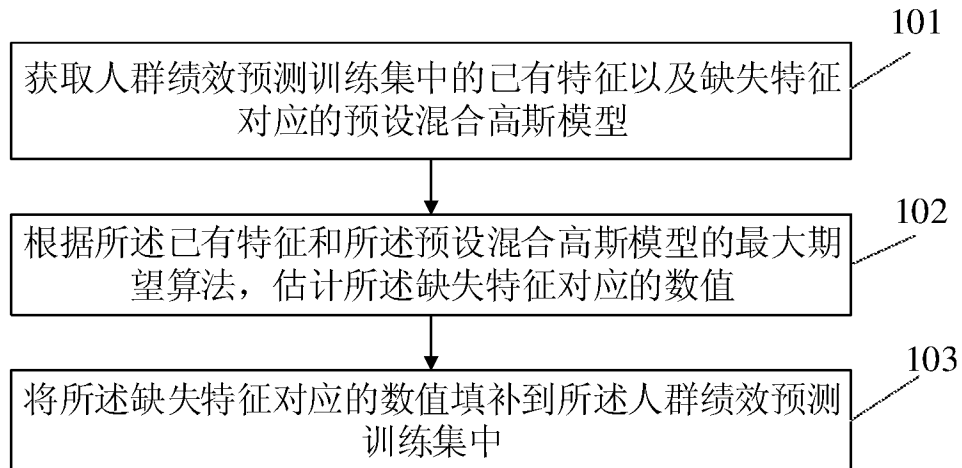


图 1

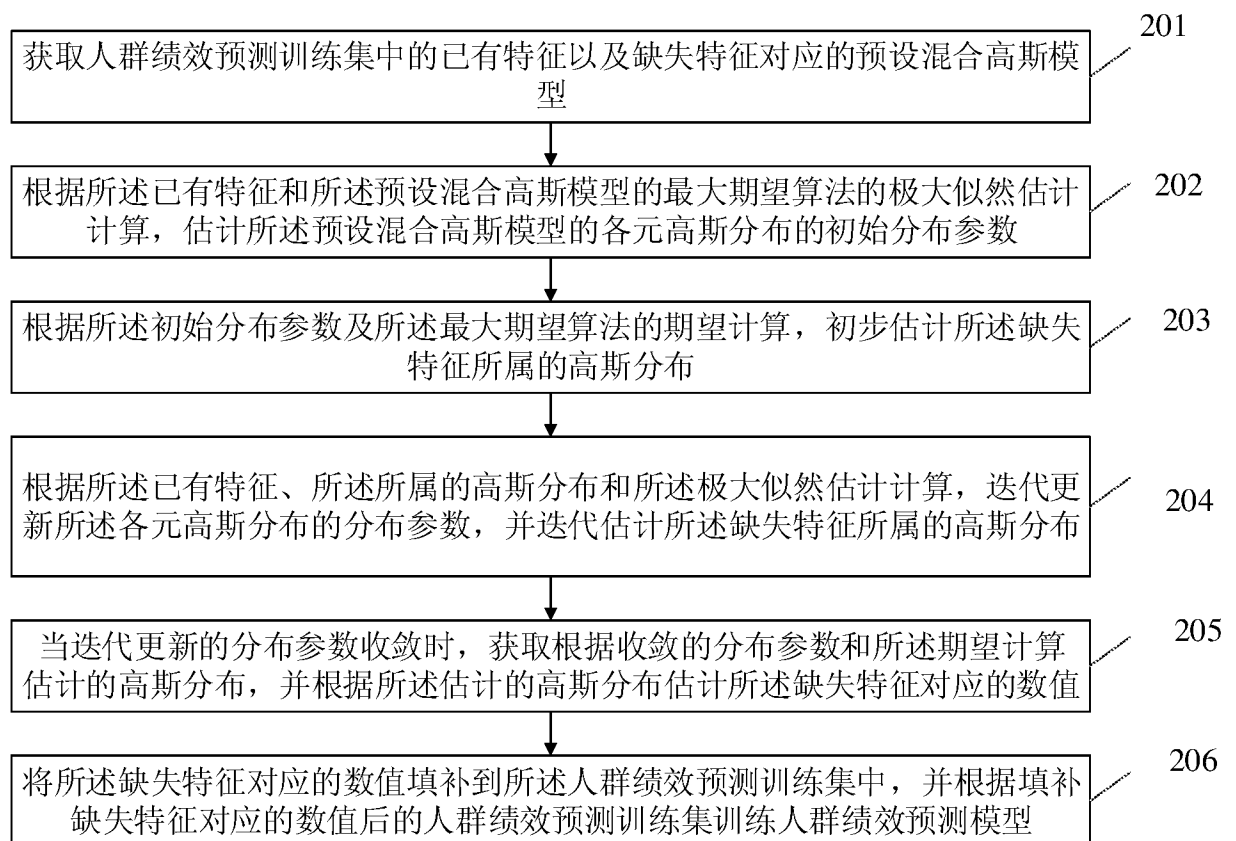


图 2

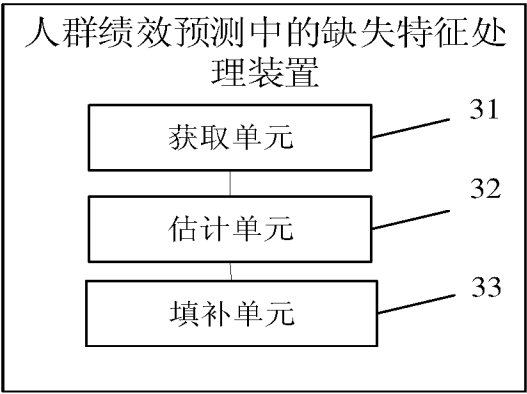


图 3

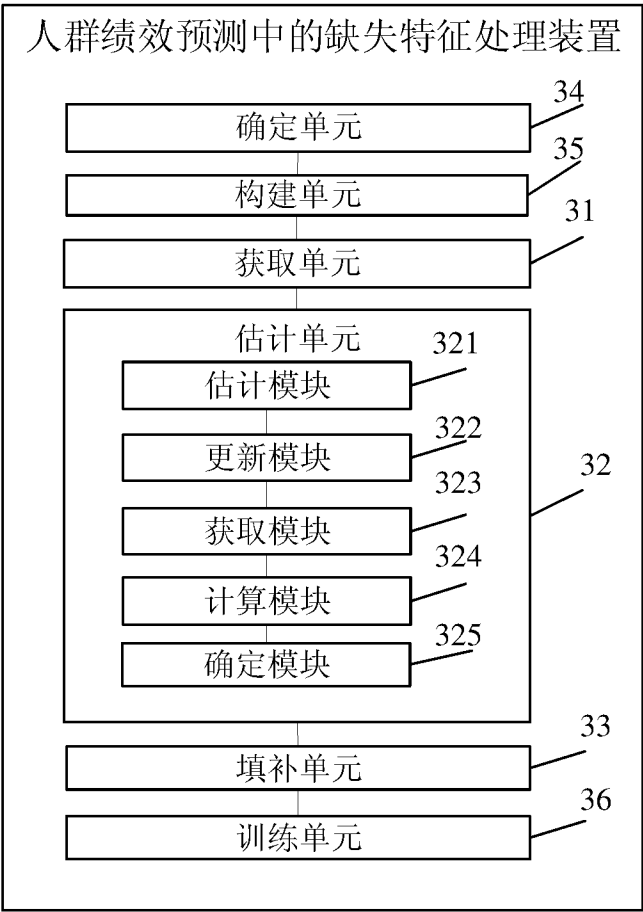


图 4

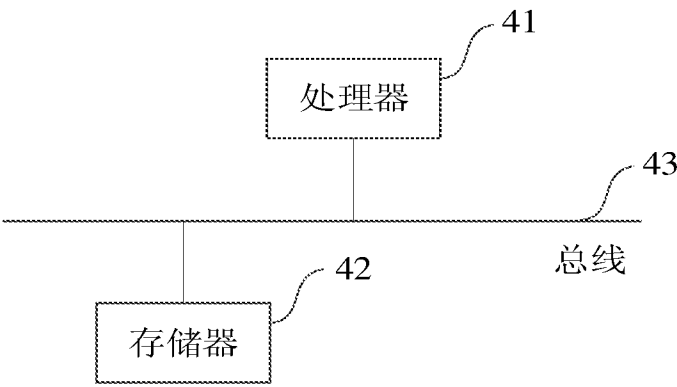


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/073294

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 10/06(2012.01)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)

G06Q

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)

CNABS, CNTXT, CNKI, DWPI, USTXT, EPTXT, WOTXT: 绩效, 预测, 机器学习, 分类, 训练, 特征, 缺失, 填充, 填补, 模型, 高斯, 最大期望算法, EM算法, performance, predict, machine learning, classify, train, feature, missing, absent, fill, model, Gaussian, expectation maximization, EM algorithm.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 101493886 A (WUHAN LANDING MEDICAL HIGH-TECH CO., LTD.) 29 July 2009 (2009-07-29) description, p. 4, paragraphs 1-7	1-20
Y	CN 104573685 A (CENTRAL SOUTH UNIVERSITY) 29 April 2015 (2015-04-29) claim 9	1-20
Y	US 2016180234 A1 (INSIDESALES. COM, INC.) 23 June 2016 (2016-06-23) claim 1	1-20
A	CN 105989843 A (ZTE CORPORATION) 05 October 2016 (2016-10-05) entire document	1-20
A	CN 107842713 A (NORTHEASTERN UNIVERSITY) 27 March 2018 (2018-03-27) entire document	1-20
A	CN 107193876 A (MERIT DATA CO., LTD.) 22 September 2017 (2017-09-22) 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

09 April 2019

Date of mailing of the international search report

30 April 2019

Name and mailing address of the ISA/CN

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

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/073294

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	101493886	A	29 July 2009	CN	101493886	B	23 May 2012
CN	104573685	A	29 April 2015	CN	104573685	B	21 November 2017
US	2016180234	A1	23 June 2016	None			
CN	105989843	A	05 October 2016	WO	2016119501	A1	04 August 2016
CN	107842713	A	27 March 2018	None			
CN	107193876	A	22 September 2017	None			

国际检索报告

国际申请号

PCT/CN2019/073294

A. 主题的分类

G06Q 10/06 (2012.01) 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)

G06Q

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))

CNABS, CNTXT, CNKI, DWPI, USTXT, EPTXT, WOTXT: 绩效, 预测, 机器学习, 分类, 训练, 特征, 缺失, 填充, 填补, 模型, 高斯, 最大期望算法, EM 算法, performance, predict, machine learning, classify, train, feature, missing, absent, fill, model, Gaussian, expectation maximization, EM algorithm.

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
Y	CN 101493886 A (武汉兰丁医学高科技有限公司) 2009年 7月 29日 (2009 - 07 - 29) 说明书第4页第1-7段	1-20
Y	CN 104573685 A (中南大学) 2015年 4月 29日 (2015 - 04 - 29) 权利要求9	1-20
Y	US 2016180234 A1 (INSIDESALES. COM, INC.) 2016年 6月 23日 (2016 - 06 - 23) 权利要求1	1-20
A	CN 105989843 A (中兴通讯股份有限公司) 2016年 10月 5日 (2016 - 10 - 05) 全文	1-20
A	CN 107842713 A (东北大学) 2018年 3月 27日 (2018 - 03 - 27) 全文	1-20
A	CN 107193876 A (美林数据技术股份有限公司) 2017年 9月 22日 (2017 - 09 - 22) 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 4月 9日

国际检索报告邮寄日期

2019年 4月 30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

受权官员

张会

传真号 (86-10) 62019451

电话号码 62412328

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/073294

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	101493886	A	2009年 7月 29日	CN	101493886	B	2012年 5月 23日
CN	104573685	A	2015年 4月 29日	CN	104573685	B	2017年 11月 21日
US	2016180234	A1	2016年 6月 23日	无			
CN	105989843	A	2016年 10月 5日	WO	2016119501	A1	2016年 8月 4日
CN	107842713	A	2018年 3月 27日	无			
CN	107193876	A	2017年 9月 22日	无			

表 PCT/ISA/210 (同族专利附件) (2015年1月)