



(12)发明专利

(10)授权公告号 CN 108710905 B

(45)授权公告日 2020.07.10

(21)申请号 201810441771.2

CN 106548257 A,2017.03.29,

(22)申请日 2018.05.10

CN 107832908 A,2018.03.23,

(65)同一申请的已公布的文献号

CN 102385724 A,2012.03.21,

申请公布号 CN 108710905 A

CN 105976049 A,2016.09.28,

(43)申请公布日 2018.10.26

CN 101894185 A,2010.11.24,

(73)专利权人 华中科技大学

CN 105825045 A,2016.08.03,

地址 430074 湖北省武汉市洪山区珞喻路

US 2007244589 A1,2007.10.18,

1037号

CN 103440368 A,2013.12.11,

专利权人 珠海格力电器股份有限公司

杨仕美等.基于LSSVM和信息熵的航材备件组合预测方法.《火力与指挥控制》.2012,第37卷(第9期), (续)

(72)发明人 王天江 易全政 李绍斌 陈彦宇

审查员 徐葛森

谭泽汉

(74)专利代理机构 华中科技大学专利中心

42201

代理人 李智 曹葆青

(51)Int.Cl.

G06K 9/62(2006.01)

(56)对比文件

CN 101086770 A,2007.12.12,

权利要求书2页 说明书8页 附图1页

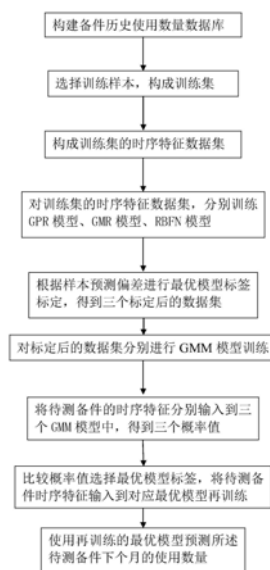
(54)发明名称

一种基于多模型联合的备件数量预测方法和系统

有效的提高预测精度。

(57)摘要

本发明公开了一种基于多模型联合的备件数量预测方法和系统,该方法包括:构建备件历史使用数量数据库,选择训练集,对每个训练样本构建时序特征,对训练集分别训练GPR模型、GMR模型、RBFN模型,根据样本预测偏差对训练样本进行最优模型标签标定,对标定后的数据集分别进行GMM模型训练,将待测样本的时序特征输入到不同GMM模型中,得到三个概率值,比较概率值选择最优模型标签,将待测样本的时序特征输入到对应的最优模型进行再训练,使用再训练的最优模型预测待测样本下个月的使用数量。本发明通过时序特征提高数据鲁棒性,并提出基于GMR模型、RBFN模型的备件预测方法,其实现简单,通过从多个模型中选出最优预测模型,能够



CN 108710905 B

[接上页]

**(56)对比文件**

Zhixiong Li et al.. "A novel prediction model for aircraft spare part intermittent demand in aviation transportation logistics using multi-components accumulation and high resolution analysis".《Journal of AEROSPACE ENGINEERING》.2015,

庞强 等. "基于高斯混合模型与主元分析的多模型切换方法".《化工学报》.2013,第64卷(第8期),

张伟 等. "基于实时学习的高斯过程回归多模型融合建模".《信息与控制》.2015,第44卷(第4期),

熊君星 等. "基于BP神经网络的备件ABC分类模型".《机械设计与制造》.2008,(第2期),

1. 一种基于多模型联合的备件数量预测方法,其特征在于,该方法包括以下步骤:

(1) 构建备件历史使用数量数据库,该数据库包括多种备件的历史使用数量,每种备件对应一个样本;

(2) 按照备件种类从所述数据库中选择训练样本,构成训练集;

(3) 对所述训练集中的每个训练样本,基于所述训练样本除最后一个月所有历史使用数量构建时序特征,构成训练集的时序特征数据集;

(4) 对训练集的时序特征数据集,分别训练预测模型高斯过程回归GPR模型、混合高斯回归GMR模型、径向基神经网络回归RBFN模型;

(5) 对所述训练集中的每个训练样本,用训练好的三个预测模型分别预测所述训练样本最后一个月使用数量,并根据样本预测偏差对所述训练样本进行最优模型标签标定,得到三个标定后的数据集,其中,每个标定后的数据集中的样本的最优模型标签相同;

(6) 对标定后的三个数据集分别进行GMM模型训练,得到三个不同的GMM模型;

(7) 对待测样本,基于所述待测样本的所有历史使用数量构建时序特征,将该时序特征输入到三个不同的GMM模型中,得到三个概率值,其反映了所述待测样本使用该GMM对应的预测模型预测效果最佳的概率;

(8) 比较三个概率值选择最优模型标签,将所述待测样本的时序特征输入到对应的最优模型进行再训练拟合,得到再训练的最优模型;

(9) 使用再训练的最优模型预测所述待测样本下个月的使用数量。

2. 如权利要求1所述的备件数量预测方法,其特征在于,步骤(2)还包括:选择完训练样本后,数据库中剩余的样本均为测试样本,构成测试集,测试样本用于优化预测模型,使用希尔系数或均方差进行预测效果的度量。

3. 如权利要求1所述的备件数量预测方法,其特征在于,构建时序特征的过程如下:若待预测的是某备件第N年第M月的备件数量,则24维时序特征的后12维依次取该备件第N-1年第M月到第N年第M月的前一个月的使用数量;根据该备件第N-1年第M月之前的历史使用数量计算每个月的月均值 $avr'_1, avr'_2, \dots, avr'_{12}$ , 24维时序特征的前12维从 $avr'_M$ 开始依次循环取值 $avr'_1, avr'_2, \dots, avr'_{12}, M=1, 2, \dots, 12$ 。

4. 如权利要求3所述的备件数量预测方法,其特征在于,月均值计算公式为
$$avr'_i = \frac{\text{第}N-1\text{年第}M\text{月之前、所有年份的第}i\text{月历史使用数量之和}}{\text{第}N-1\text{年第}M\text{月之前有第}i\text{月的年数}}, i=1, 2, \dots, 12$$
。

5. 如权利要求1所述的备件数量预测方法,其特征在于,进行最优模型标签标定时,标定0标识针对当前样本高斯过程回归模型预测效果最佳,标定1标识混合高斯回归模型预测效果最佳,标定2标识径向基神经网络回归模型预测效果最佳;若存在多种模型样本预测偏差是最小,按照GPR>GMR>RBFN的优先级顺序进行标签标定。

6. 一种基于多模型联合的备件数量预测系统,其特征在于,该系统包括:

备件历史使用数量数据库,该数据库包括多种备件的历史使用数量,每种备件对应一个样本;

预测模型训练模块,该模块按照备件种类从所述数据库中选择训练样本,构成训练集;对所述训练集中的每个训练样本,基于所述训练样本除最后一个月所有历史使用数量构建时序特征,构成训练集的时序特征数据集;对训练集的时序特征数据集,分别训练预测模

型高斯过程回归GPR模型、混合高斯回归GMR模型、径向基神经网络回归RBFN模型；

标定模块,该模块对所述训练集中的每个训练样本,用训练好的三个预测模型分别预测所述训练样本最后一个月的使用数量,并根据样本预测偏差对所述训练样本进行最优模型标签标定,得到三个标定后的数据集,其中,每个标定后的数据集中的样本的最优模型标签相同;

GMM模型训练模块,该模块对标定后的三个数据集分别进行GMM模型训练,得到三个不同的GMM模型;

预测模块,该模块对待测样本,基于所述待测样本的所有历史使用数量构建时序特征,将该时序特征输入到三个不同的GMM模型中,得到三个概率值,其反映了所述待测样本使用该GMM对应的预测模型预测效果最佳的概率;比较三个概率值选择最优模型标签,将所述待测样本的时序特征输入到对应的最优模型进行再训练拟合,得到再训练的最优模型;使用再训练的最优模型预测所述待测样本下个月的使用数量。

7.如权利要求6所述的备件数量预测系统,其特征在于,所述系统还包括预测模型的优化模块,该模块在预测模型训练模块选择完训练样本后,数据库中剩余的样本均为测试样本,构成测试集,测试样本用于优化预测模型,使用希尔系数或均方差进行预测效果的度量。

8.如权利要求6所述的备件数量预测系统,其特征在于,构建时序特征的过程如下:若待预测的是某备件第N年第M月的备件数量,则24维时序特征的后12维依次取该备件第N-1年第M月到第N年第M月的前一个月的使用数量;根据该备件第N-1年第M月之前的历史使用数量计算每个月的月均值 $avr'_1, avr'_2, \dots, avr'_{12}$ ,24维时序特征的前12维从 $avr'_M$ 开始依次循环取值 $avr'_1, avr'_2, \dots, avr'_{12}, M=1, 2, \dots, 12$ 。

9.如权利要求8所述的备件数量预测系统,其特征在于,月均值计算公式为
$$avr'_i = \frac{\text{第}N-1\text{年第}M\text{月之前、所有年份的第}i\text{月历史使用数量之和}}{\text{第}N-1\text{年第}M\text{月之前有第}i\text{月的年数}}, i=1, 2, \dots, 12。$$

10.如权利要求6所述的备件数量预测系统,其特征在于,进行最优模型标签标定时,标定0标识针对当前样本高斯过程回归模型预测效果最佳,标定1标识混合高斯回归模型预测效果最佳,标定2标识径向基神经网络回归模型预测效果最佳;若存在多种模型样本预测偏差是最小,按照GPR>GMR>RBFN的优先级顺序进行标签标定。

## 一种基于多模型联合的备件数量预测方法和系统

### 技术领域

[0001] 本发明属于人工智能领域,更具体地,涉及一种基于多模型联合的备件数量预测方法和系统。

### 背景技术

[0002] 备件是设备备用的零件,缺少对备件的采购和储备数量进行合理的规划,可能会对企业业务及发展造成很大的困扰乃至损失。备件储备数量过多,将会占用企业的大量周转资金与储存仓库空间,不利于企业的长期发展;备件储备数量过少,将会造成维修或生产业务无法及时完成,造成企业经济损失。

[0003] 现有的备件数量预测方法中主要包括两类,一类是基于ARMA等统计学模型,一类是基于SVM、GPR等机器学习模型。但这些预测方法存在一些缺陷:(1)不同备件之间使用规律不具有相似性,比如某种零件容易发生周期性故障而有些零件基本上不发生故障,因此选定的模型不适用于所有零件;(2)在预测模型选择上依赖于人工经验,而非备件样本本身反映出的规律;(3)备件生产过程中可能会对工艺进行调整,备件采购过程中可能会对来源进行调整,上述调整将导致备件前后不一样,旧的模型无法适用于新的备件。

### 发明内容

[0004] 针对现有技术的上述缺陷,本发明提供了一种基于多模型联合的备件数量预测方法和系统,其目的在于,提供一种高鲁棒性、适应备件数量变换规律的预测方法。

[0005] 为实现上述目的,一方面,本发明提供一种基于多模型联合的备件数量预测方法,该方法包括以下步骤:

[0006] (1) 构建备件历史使用数量数据库,该数据库包括多种备件的历史使用数量,每种备件对应一个样本;

[0007] (2) 按照备件种类从所述数据库中选择训练样本,构成训练集;

[0008] (3) 对所述训练集中的每个训练样本,基于所述训练样本除最后一个月的所有历史使用数量构建时序特征,构成训练集的时序特征数据集;

[0009] (4) 对训练集的时序特征数据集,分别训练预测模型高斯过程回归GPR模型、混合高斯回归GMR模型、径向基神经网络回归RBFN模型;

[0010] (5) 对所述训练集中的每个训练样本,用训练好的三个预测模型分别预测所述训练样本最后一个月使用数量,并根据样本预测偏差对所述训练样本进行最优模型标签标定,得到三个标定后的数据集,其中,每个标定后的数据集中的样本的最优模型标签相同;

[0011] (6) 对标定后的三个数据集分别进行GMM模型训练,得到三个不同的GMM模型;

[0012] (7) 对待测样本,基于所述待测样本的所有历史使用数量构建时序特征,将该时序特征输入到三个不同的GMM模型中,得到三个概率值,其反映了所述待测样本使用该GMM对应的预测模型预测效果最佳的概率;

[0013] (8) 比较三个概率值选择最优模型标签,将所述待测样本的时序特征输入到对应

的最优模型进行再训练拟合,得到再训练的最优模型;

[0014] (9) 使用再训练的最优模型预测所述待测样本下个月的使用数量。

[0015] 优选地,步骤(2)还包括:选择完训练样本后,数据库中剩余的样本均为测试样本,构成测试集,测试样本用于优化预测模型,可使用希尔系数或均根方差进行预测效果的度量。

[0016] 优选地,构建时序特征的过程如下:若待预测的是某备件第N年第M月的备件数量,则24维时序特征的后12维依次取该备件第N-1年第M月到第N年第M月的前一个月的使用数量;根据该备件第N-1年第M月之前的历史使用数量计算每个月的月均值 $avr'_1, avr'_2, \dots, avr'_12$ , 24维时序特征的前12维从 $avr'_M$ 开始依次循环取值 $avr'_1, avr'_2, \dots, avr'_12, M=1, 2, \dots, 12$ 。

[0017] 优选地,月均值计算公式为

$$[0018] \quad avr'_i = \frac{\text{第N-1年第M月之前、所有年份的第i月历史使用数量之和}}{\text{第N-1年第M月之前有第i月的年数}}, i = 1, 2, \dots, 12$$

[0019] 优选地,进行最优模型标签标定时,标定0标识针对当前样本高斯过程回归模型预测效果最佳,标定1标识混合高斯回归模型预测效果最佳,标定2标识径向基神经网络回归模型预测效果最佳;若存在多种模型样本预测偏差是最小,按照GPR>GMR>RBFN的优先级顺序进行标签标定。

[0020] 另一方面,本发明提供一种基于多模型联合的备件数量预测系统,该系统包括:

[0021] 备件历史使用数量数据库,该数据库包括多种备件的历史使用数量,每种备件对应一个样本;

[0022] 预测模型训练模块,该模块按照备件种类从所述数据库中选择训练样本,构成训练集;对所述训练集中的每个训练样本,基于所述训练样本除最后一个月所有历史使用数量构建时序特征,构成训练集的时序特征数据集;对训练集的时序特征数据集,分别训练预测模型高斯过程回归GPR模型、混合高斯回归GMR模型、径向基神经网络回归RBFN模型;

[0023] 标定模块,该模块对所述训练集中的每个训练样本,用训练好的三个预测模型分别预测所述训练样本最后一个月使用数量,并根据样本预测偏差对所述训练样本进行最优模型标签标定,得到三个标定后的数据集,其中,每个标定后的数据集中的样本的最优模型标签相同;

[0024] GMM模型训练模块,该模块对标定后的三个数据集分别进行GMM模型训练,得到三个不同的GMM模型;

[0025] 预测模块,该模块对待测样本,基于所述待测样本的所有历史使用数量构建时序特征,将该时序特征输入到三个不同的GMM模型中,得到三个概率值,其反映了所述待测样本使用该GMM对应的预测模型预测效果最佳的概率;比较三个概率值选择最优模型标签,将所述待测样本的时序特征输入到对应的最优模型进行再训练拟合,得到再训练的最优模型;使用再训练的最优模型预测所述待测样本下个月的使用数量。

[0026] 优选地,所述系统还包括预测模型的优化模块,该模块在预测模型训练模块选择完训练样本后,数据库中剩余的样本均为测试样本,构成测试集,测试样本用于优化预测模型,可使用希尔系数或均根方差进行预测效果的度量。

[0027] 优选地,构建时序特征的过程如下:若待预测的是某备件第N年第M月的备件数

量,则24维时序特征的后12维依次取该备件第N-1年第M月到第N年第M月的前一个月的使用数量;根据该备件第N-1年第M月之前的历史使用数量计算每个月的月均值 $avr'_1,avr'_2,\dots,avr'_12$ ,24维时序特征的前12 维从 $avr'_M$ 开始依次循环取值 $avr'_1,avr'_2,\dots,avr'_12$ , $M=1,2,\dots,12$ 。

[0028] 优选地,月均值计算公式为

$$[0029] \quad avr'_i = \frac{\text{第N-1年第M月之前、所有年份的第i月历史使用数量之和}}{\text{第N-1年第M月之前有第i月的年数}}, i=1,2,\dots,12$$

[0030] 优选地,进行最优模型标签标定时,标定0标识针对当前样本高斯过程回归模型预测效果最佳,标定1标识混合高斯回归模型预测效果最佳,标定2标识径向基神经网络回归模型预测效果最佳;若存在多种模型样本预测偏差是最小,按照GPR>GMR>RBFN的优先级顺序进行标签标定。

[0031] 总体而言,通过本发明所构思的以上技术方案与现有技术相比,具有以下有益效果:

[0032] (1) 本发明通过构建时序特征的数据集,提取反映备件数量变化的平均规律与近一年使用规律作为特征,有效解决了备件前后不一致造成的预测不准确的问题,提取的特征更加具有鲁棒性;

[0033] (2) 本发明提出一种基于GMR模型的备件数量预测方法和一种RBFN 模型的备件数量预测方法,其具有能够处理小样本、非线性、参数自适应获取、实现过程简单的优势;

[0034] (3) 本发明通过对待测样本进行多模型预测,从多个模型中选出最优预测模型,使用最优预测模型来预测未来使用量,能够有效的提高预测精度。

## 附图说明

[0035] 图1为本发明实施例提供的一种基于多模型联合的备件数量预测方法流程图。

## 具体实施方式

[0036] 为了使本发明的目的、技术方案及优点更加清楚明白,以下结合附图及实施例,对本发明进行进一步详细说明。应当理解,此处所描述的具体实施例仅仅用以解释本发明,并不用于限定本发明。

[0037] 本发明的基本思路在于,提出一种基于多模型联合的备件数量预测方法。通过平均年+最近一年的预处理有效的提取历史数据中的表现特征,既保留了历史数据变化情况,又抓住最近一年的数据变化,使得特征更具有鲁棒性以及预测的稳定性。通过多个混合高斯(GMM)模型进行特征分类,选择更加适合的预测方法。对于测试样本使用最优预测方法进行预测,能够显著提高预测的精度与准确率。

[0038] 图1为本发明实施例提供的一种基于多模型联合的备件数量预测方法流程图,如图1所示,该方法包括以下步骤:

[0039] (1) 构建备件历史使用数量数据库,该数据库包括多种备件的历史使用数量,每种备件对应一个样本;

[0040] (2) 按照备件种类从所述数据库中选择训练样本,构成训练集;

[0041] (3) 对所述训练集中的每个训练样本,基于所述训练样本除最后一个月所有历史使用数量构建时序特征,构成训练集的时序特征数据集;

[0042] (4) 对训练集的时序特征数据集,分别训练预测模型高斯过程回归GPR 模型、混合高斯回归GMR模型、径向基神经网络回归RBFN模型;

[0043] (5) 对所述训练集中的每个训练样本,用训练好的三个预测模型分别预测所述训练样本最后一个月使用数量,并根据样本预测偏差对所述训练样本进行最优模型标签标定,得到三个标定后的数据集,其中,每个标定后的数据集中的样本的最优模型标签相同;

[0044] (6) 对标定后的三个数据集分别进行GMM模型训练,得到三个不同的 GMM模型;

[0045] (7) 对待测样本,基于所述待测样本的所有历史使用数量构建时序特征,将该时序特征输入到三个不同的GMM模型中,得到三个概率值,其反映了所述待测样本使用该GMM对应的预测模型预测效果最佳的概率;

[0046] (8) 比较三个概率值选择最优模型标签,将所述待测样本的时序特征输入到对应的最优模型进行再训练拟合,得到再训练的最优模型;

[0047] (9) 使用再训练的最优模型预测所述待测样本下个月的使用数量。

[0048] 以下以空调备件为例,阐述时序特征的构建过程。

[0049] 按配件编码和销售公司对空调备件的历史使用量进行分类和合并,形成不同种类空调备件使用数量的样本记录。计算备件样本记录中不同年份、相同月份的备件使用数量的平均值,作为该备件使用数量各月平均值。如表1所示的某销售公司某备件的历史数据,统计历史月份平均值如表2所示;

[0050] 表1历史数据

年	月	数量	年	月	数量	年	月	数量	
2012	8	6	2014	1	22	2015	5	16	
2012	9	4	2014	2	18	2015	6	22	
2012	11	2	2014	3	18	2015	7	38	
2012	12	6	2014	4	12	2015	8	26	
2013	1	24	2014	5	4	2015	9	26	
2013	2	2	2014	6	38	2015	10	6	
2013	3	4	2014	7	50	2015	12	28	
[0051]	2013	4	6	2014	8	50	2016	1	32
	2013	5	8	2014	9	12	2016	2	6
	2013	6	20	2014	10	26	2016	3	14
	2013	7	46	2014	11	12			
	2013	8	30	2014	12	34			
	2013	9	10	2015	1	32			
	2013	10	2	2015	2	10			
	2013	11	6	2015	3	14			
	2013	12	34	2015	4	8			

[0052] 表2统计历史数据中各月平均值

月份	均值	月份	均值	月份	均值	月份	均值	
[0053]	1	27.5	4	8.67	7	44.67	10	11.33
	2	9.0	5	9.33	8	28.0	11	6.67
	3	12.0	6	26.67	9	13.0	12	25.5

[0054] 对于备件样本记录中没有记录的月份,采用月均值填充。如表1所示,备件样本记



录中缺少2012年10月的记录。如表2所示,10月均值为11.33,因此,备件样本记录应填充 [2012,10,11.33]数据,表示2012年10月用了 11.33个该备件。

[0055] 若待预测的是某备件第N年第M月的备件数量,则24维时序特征的后 12维依次取该备件第N-1年第M月到第N年第M月的前一个月的使用数量;根据该备件第N-1年第M月之前的历史使用数量计算每个月的月均值  $avr'_1,avr'_2,\dots,avr'_{12}$ ,24维时序特征的前12维从  $avr'_M$ 开始依次循环取值  $avr'_1,avr'_2,\dots,avr'_{12},M=1,2,\dots,12$ 。

[0056] 月均值计算公式为

$$[0057] \quad avr'_i = \frac{\text{第N-1年第M月之前、所有年份的第i月历史使用数量之和}}{\text{第N-1年第M月之前有第 i 月的年数}}, i = 1, 2, \dots, 12$$

[0058] 待预测的是某备件2016年3月的使用量。如表3所示,对应的时序特征——时序特征后12维是2015.3的使用总量、2015.4的使用总量,……,2016.2的使用总量;时序特征前12维是2015.3前年月的3月数据平均值,2015.3前年月4月数据平均值,……,2015.3前年月的1月数据平均值;表4为2016年3月的实际使用数量。

[0059] 表3预测2016年3月使用数量时时序特征

	序号	总量	对应于	序号	总量	对应于
	1	11	2015.03 前年月的 3 月数据平均值	13	14	2015.03
	2	9	2015.03 前年月的 4 月数据平均值	14	8	2015.04
	3	6	2015.03 前年月的 5 月数据平均值	15	16	2015.05
	4	29	2015.03 前年月的 6 月数据平均值	16	22	2015.06
	5	48	2015.03 前年月的 7 月数据平均值	17	38	2015.07
[0060]	6	28.67	2015.03 前年月的 8 月数据平均值	18	26	2015.08
	7	8.67	2015.03 前年月的 9 月数据平均值	19	26	2015.09
	8	13.11	2015.03 前年月的 10 月数据平均值	20	6	2015.10
	9	6.67	2015.03 前年月的 11 月数据平均值	21	6.67	2015.11
	10	24.67	2015.03 前年月的 12 月数据平均值	22	28	2015.12
	11	26	2015.03 前年月的 1 月数据平均值	23	32	2016.01
	12	10	2015.03 前年月的 2 月数据平均值	24	6	2016.02

[0061] 表4实际使用数量

序号	总量	对应于
25	14	2016.03

[0063] 例如,2015.03前年月的3月数据平均值计算的是2013.03、2014.03 的平均值,2015.03前年月的1月数据平均值计算的是2013.01、2014.01、2015.01的平均值。

[0064] 测试样本用于优化预测模型时,可使用希尔系数或均根方差等评价标准进行预测效果的度量。

[0065] 步骤(4)具体地包括以下步骤:

[0066] (4.1)对步骤(3)得到的训练集的时序特征数据集,训练高斯过程回归模型(Gaussian Process Regression Model,简称GPR),并进行样本偏差计算,训练过程包括以下子步骤:

[0067] (4.1.1)处理输入格式为: $X = [x_1, x_2, \dots, x_{24}]$ ,输出结果为形式  $Y = [y_1, y_2, \dots, y_{24}]$ ;

[0068] (4.1.2) 定义高斯过程模型核函数使用高斯核函数核,核函数形式为  $k(x, x') = \exp(-\frac{\|x-x'\|^2}{2\delta^2})$ , 其中  $\|x-x'\|^2 = x^T x + (x')^T x' - 2x^T x'$ ,  $\delta$  代表步长; 并假定样本中存在满足高斯分布  $N(0, \sigma^2)$  的噪声, 回归模型在训练时需要训练超参数即为  $\delta$  (步长),  $\sigma^2$  (噪声方差);

[0069] (4.1.3) 计算协方差矩阵, 使用最大似然估计法求解超参数;

[0070] 对于表3中训练数据进行训练, 可得参数训练结果如下所示:  $\delta = 2.4439$ ,  $\sigma = 2.278$ 。

[0071] (4.1.4) 预测时, 根据联合密度后验分布, 在  $x_{N+1}$  处的概率分布满足:  $p(y_{N+1} | Y, X, x_{N+1}, K) = N(YC_N^{-1}k_*, C - k_*^T C_N^{-1}k_*)$ , 其中,  $C_N = k(X, X) + \sigma^2 I_N$ ,  $k_* = k(X, x_{N+1})$ ,  $C = k(x_{N+1}, x_{N+1})$ , 那么  $x_{N+1}$  处值即为  $YC_N^{-1}k_*$ 。

[0072] 将步骤4.1.3训练得到的参数带入模型后, 对于表3预测数据进行预测, 预测结果为18.72, 按照四舍五入取整的方式, 可得2016年3月的预测使用量为19, 样本预测偏差为5。

[0073] (4.2) 对步骤(3)得到的训练集的时序特征数据集, 训练混合高斯回归模型 (Gaussian Mixture Regression Model, 简称GMR), 并进行样本偏差计算, 训练过程包括以下子步骤:

[0074] (4.2.1) 定义  $X = [x_1, x_2, \dots, x_{24}]$ ,  $Y = [y_1, y_2, \dots, y_{24}]$ , 处理模型的输入为  $[X^T, Y^T]$ ;

[0075] (4.2.2) 假定数据的输入和输出的联合概率密度函数满足GMM模型,

$$[0076] \quad f_{x,y}(x, y) = \sum_{i=1}^K \omega_i N(x, y | \mu_i, \Sigma_i) \quad \text{其中, } \sum_{i=1}^K \omega_i = 1, \quad \mu_i = \begin{bmatrix} \mu_{iX} \\ \mu_{iY} \end{bmatrix}, \quad \Sigma_i = \begin{bmatrix} \Sigma_{iXX} & \Sigma_{iXY} \\ \Sigma_{iYX} & \Sigma_{iYY} \end{bmatrix}$$

[0077] 因此模型在训练阶段需要指定高斯分量的个数  $K$ , 需要训练第  $i$  分量的权重系数  $\omega_i$ , 均值矩阵  $\mu_i$ , 方差矩阵  $\Sigma_i$ ;

[0078] (4.2.3) 计算GMM模型的似然函数, 使用EM算法对参数进行训练。

[0079] 对于表3训练数据进行试训练, 选定高斯分量个数为6, 实际模型训练时根据样本特性可选择多个不同的个数进行评估, 训练结果如下所示:

$$[0080] \quad \omega_i = \begin{bmatrix} 0.166 \\ 0.205 \\ 0.135 \\ 0.081 \\ 0.255 \\ 0.158 \end{bmatrix}, \quad \mu_i = \begin{bmatrix} 19.8 & 6.7 \\ 19.8 & 30.0 \\ 13.7 & 14.5 \\ 4.5 & 38.8 \\ 5.2 & 9.1 \\ 10.3 & 25.6 \end{bmatrix}, \quad \Sigma_i = \begin{bmatrix} [13.2 & -2.6] \\ [-2.6 & 0.7] \\ [5.5 & -2.0] \\ [-2.0 & 21.2] \\ [2.6 & 5.9] \\ [5.9 & 14.9] \\ [0.3 & 4.7] \\ [4.7 & 90.1] \\ [10.6 & 0.2] \\ [0.2 & 5.8] \\ [12.0 & -7.7] \\ [-7.7 & 5.5] \end{bmatrix}$$

[0081] (4.2.4) 预测时, 在  $x_{N+1}$  处的后验概率分布推导结果为:

[0082] 
$$\varphi_i = \frac{\omega_i N(x_{N+1} | \mu_{iX}, \Sigma_{iX})}{\sum_{i=1}^K \omega_i N(x_{N+1} | \mu_{iX}, \Sigma_{iX})}, \quad m_i(x) = \mu_{iY} + \Sigma_{iYX} \Sigma_{iXX}^{-1} (x_{n+1} - \mu_{iX})$$

[0083] 最终对于未来的预测值为：
$$y_{N+1} = \sum_{i=1}^K \varphi_i m_i(x)$$

[0084] 将步骤4.2.3训练得到的参数带入模型后,对于表3预测数据进行预测,预测结果为12.75,按照四舍五入取整的方式,可得2016年3月的预测使用量为13,样本预测偏差为1。

[0085] (4.3)对步骤(3)得到的训练集的时序特征数据集,训练径向基神经网络回归模型(Radial Basis Function Network Regression Model,简称RBFN)并进行偏差计算,训练过程包括以下子步骤:

[0086] (4.3.1)处理输入格式为: $X = [x_1, x_2, \dots, x_{24}]$ ,输出结果为形式  $Y = [y_1, y_2, \dots, y_{24}]$ ;

[0087] (4.3.2)设定输出层神经元的输出为  $y(x) = \sum_{i=1}^K \omega_i G(x | \mu_i, \beta_i)$   $G(x | \mu_i, \beta_i)$  是高斯核函数,具体表达式为: $G(x | \mu_i, \beta_i) = \exp(\beta_i ||x - \mu_i||^2)$ ,其中  $\omega_i, \mu_i$  分别代表到输出层权重、每个隐含层神经元的径向基核函数的中心,而  $\beta_i$  等价于  $\beta_i = -\frac{1}{2\sigma_i^2}$   $\sigma_i^2$  是高斯核函数的标准

差,K表示隐含层神经元的个数;

[0088] (4.3.3)训练时,首先使用聚类方法确定中心  $\mu_i$ ,然后使用BP反馈调节方式训练  $\omega_i, \beta_i$ 。

[0089] 对于表3训练数据进行试训练,选定隐含层个数为5,输入层输出层神经元个数均为1,实际模型训练时根据样本特性可选择多个不同的隐含层个数进行评估,训练结果如下所示:

[0090] 
$$\omega_i = \begin{bmatrix} 8.54 \\ 19.86 \\ 34.74 \\ 8.72 \\ -0.74 \end{bmatrix}, \mu_i = \begin{bmatrix} 22.59 \\ 17.18 \\ 4.98 \\ 15.42 \\ 23.40 \end{bmatrix}, \beta_i = \begin{bmatrix} -3.2e-3 \\ -7.2e-1 \\ -1.2 \\ -1.6e-3 \\ -97.9 \end{bmatrix}$$

[0091] (4.3.4)预测时,由步骤(4.3.2)模型带入即可,将步骤4.3.3训练得到的参数带入模型后,对于表3预测数据进行预测,预测结果为15.88,按照四舍五入取整的方式,可得2016年3月的预测使用量为16,样本预测偏差为2。

[0092] 步骤(4)结束后,对所述训练集中的每个训练样本,用训练好的三个预测模型分别预测所述训练样本最后一个月的使用数量,并根据样本预测偏差对所述训练样本进行最优模型标签标定,得到三个标定后的数据集,其中,每个标定后的数据集的样本的最优模标签相同。

[0093] 进行最优模型标签标定时,标定0标识针对当前样本高斯过程回归模型预测效果最佳,标定1标识混合高斯回归模型预测效果最佳,标定2标识径向基神经网络回归模型预

测效果最佳;若存在多种模型样本预测偏差是最小,那么考虑训练所需时间按照GPR>GMR>RBFN的优先级顺序进行标签标定。对于表3的预测情况,因GMR偏差为1,RBFN偏差为2,GPR偏差为 5,因此该样本的最优模型标签标定为1。

[0094] 对标定后的三个数据集分别进行GMM模型训练,得到三个不同的GMM 模型;每个GMM的训练过程包括以下步骤:

[0095] (6.1) 处理输入格式为时序特征 $X = [x_1, x_2, \dots, x_{24}]$ ;

[0096] (6.2) 定义高斯混合模型的密分布,  $p(x) = \sum_{i=1}^K \omega_i N(x | \mu_i, \Sigma_i)$  其中  $\sum_{i=1}^K \omega_i = 1$  该分布

由K个混合成分组成,每个混合成分对应一个高斯分布,其中多元高斯分布概率密度函数为:

$$[0097] \quad N(X | \mu_i, \Sigma_i) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)}$$

[0098] 其中 $\mu_i, \Sigma_i$ 分别为其分布均值向量与协方差矩阵,n是输入向量维度,本实施例中n为24。

[0099] 定义 $\lambda_i = \{\omega_i, \mu_i, \Sigma_i\}$ ,表明对于单个GMM模型,需要学习的参数是 $\lambda_i$ ;

[0100] (6.3) 计算GMM模型的似然函数;

[0101] (6.4) 使用EM算法对参数进行训练,得到 $\lambda_i$ 。

[0102] (6.5) 按照此策略,分别对每一类样本训练其概率密度函数对应的GMM 模型,对于每类的GMM模型,设置高斯分量个数为16。

[0103] 对待测样本,基于所述待测样本的所有历史使用数量构建时序特征,将该时序特征输入到三个不同的GMM模型中,得到三个概率值,其反映了所述待测样本使用该GMM对应的预测模型预测效果最佳的概率;

[0104] 比较三个概率值选择最优模型标签,将所述待测样本的时序特征输入到对应的最优模型进行再训练拟合,得到再训练的最优模型;

[0105] 使用再训练的最优模型预测所述待测样本下个月的使用数量。

[0106] 以上,仅为本申请较佳的具体实施方式,但本申请的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本申请揭露的技术范围内,可轻易想到的变化或替换,都应涵盖在本申请的保护范围之内。因此,本申请的保护范围应该以权利要求的保护范围为准。

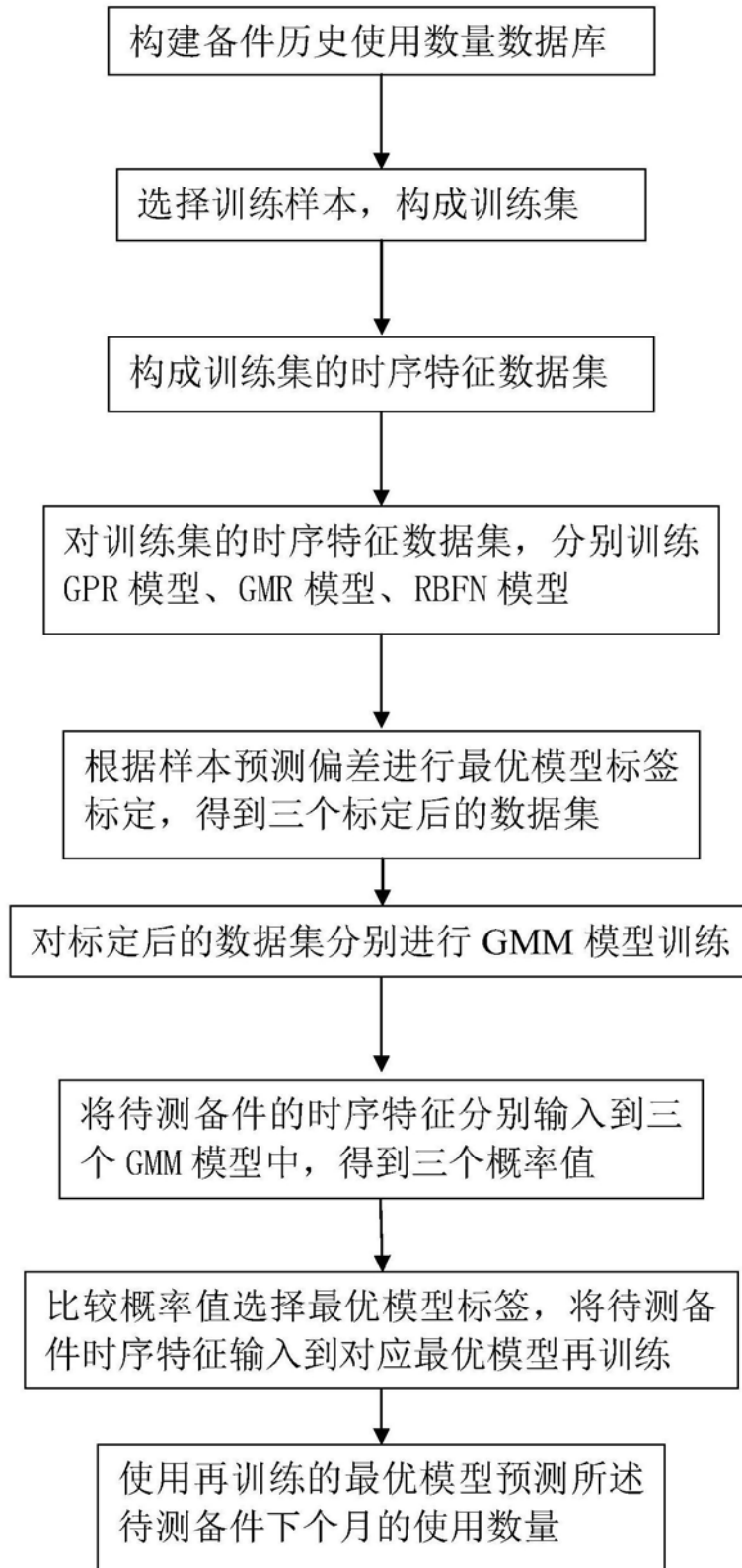


图1