



(12) 发明专利

(10) 授权公告号 CN 109920249 B

(45) 授权公告日 2021.09.10

(21) 申请号 201910185609.3

审查员 双珍珍

(22) 申请日 2019.03.12

(65) 同一申请的已公布的文献号
申请公布号 CN 109920249 A

(43) 申请公布日 2019.06.21

(73) 专利权人 长沙理工大学
地址 410114 湖南省长沙市(雨花区)万家
丽南路2段960号

(72) 发明人 陈曦 何宇明 李捷 彭朔

(74) 专利代理机构 北京旭路知识产权代理有限公司 11567
代理人 姚自奇 莫舒颖

(51) Int. Cl.
G08G 1/01 (2006.01)
G06N 20/00 (2019.01)

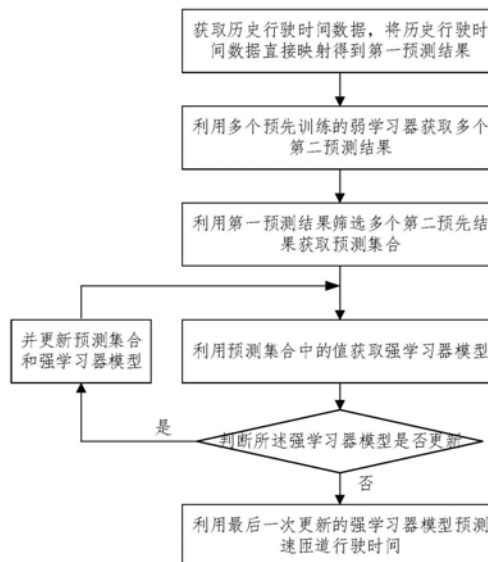
权利要求书2页 说明书4页 附图3页

(54) 发明名称

一种基于多模型融合的高速匝道行驶时间预测方法

(57) 摘要

本发明涉及一种基于多模型融合的高速匝道行驶时间预测方法;S1、获取历史行驶时间数据,将历史行驶时间数据直接映射得到第一预测结果;S2、利用多个预先训练的弱学习器获取多个第二预测结果;S3、利用第一预测结果筛选多个第二预先结果获取预测集合;S4、利用预测集合中的MAPE值获取强学习器模型并更新预测集合和强学习器模型;S5、重复执行上述步骤S4;直至强学习器模型不再更新,利用最后一次更新的强学习器模型作为高速匝道行驶时间的预测模型;本发明方法能够对高速匝道的行驶时间进行预测,且得到的强学习器模型比任何单一模型的预测精度都高,得到的强学习器模型具有较佳的鲁棒性和泛化能力。



1. 一种基于多模型融合的高速匝道行驶时间预测方法,其特征在于,包括:

S1、获取历史行驶时间数据,将所述历史行驶时间数据直接映射得到第一预测结果SWL₀;

S2、利用多个预先训练的弱学习器获取多个第二预测结果SWL_i;

S3、利用所述第一预测结果SWL₀筛选所述多个第二预测结果SWL_i获取包含多个弱学习器模型的预测集合SWL_j;

S4、利用预测集合SWL_j中平均绝对误差百分比误差最小的两个弱学习器模型获取强学习器模型SLM,将预测集合SWL_j中所述选取的两个弱学习器模型剔除,并将强学习器模型SLM作为新的弱学习模型增加至预测集合SWL_j,更新强学习器模型和预测集合SWL_j;

S5、重复执行上述步骤S4直至强学习器模型SLM不再更新,利用最后一次更新的强学习器模型SLM作为高速匝道行驶时间的预测模型。

2. 如权利要求1所述的方法,其特征在于,在步骤S2和S3之间还包括:利用公式一计算所述第一预测结果的平均绝对百分比误差MAPE₀和每一所述第二预测结果的平均绝对百分比误差MAPE_i;

$$\text{公式一: } MAPE = \frac{1}{R} \sum_{r=1}^R \frac{1}{T} \sum_{t=1}^T \left| \frac{d_{rt} - p_{rt}}{d_{rt}} \right|$$

其中,R表示测试数据的路线数量,T表示测试数据的时间段数量,d_{rt}表示r路线t时间段的平均行驶时间,p_{rt}表示表示r路线t时间段的预测行驶时间。

3. 如权利要求2所述的方法,其特征在于,在S3中,包含多个弱学习器模型的预测集合SWL_j中的每一个弱学习器模型的平均绝对百分比误差MAPE_j均小于MAPE₀。

4. 如权利要求3所述的方法,其特征在于,在步骤S4还包括:利用公式二获取预测集合SWL_j,并根据公式三更新预测集合SWL_j;

公式二:WL_j=MAPE_j,j=1,2,...,J;

公式三:SWL_j=(j,WL_j),j=1,2,...,J;

式中J表示模型中用到的机器学习模型数量,j表示对应的弱学习器,MAPE_j表示第j个模型的平均绝对误差百分比误差。

5. 如权利要求4所述的方法,其特征在于,在步骤S4中利用两个弱学习器模型构建强学习器模型SLM的方法包括:

A1、以预设迭代速率v调整每个弱学习器的权重比,得到多种权重比的强学习器模型;

A2、获取每一所述权重比的强学习器模型的平均绝对误差百分比误差,将平均绝对误差百分比误差最小时对应的强学习器模型作为最终的强学习器模型SLM。

6. 如权利要求5所述的方法,其特征在于,

所述多个预先训练的弱学习器模型包括XGBoost、LightGBM、SVM、线性回归和KNN中的至少两种。

7. 如权利要求1-6任一所述的方法,其特征在于,步骤S1之前包括利用下述步骤获取多个预先训练的弱学习器:

B1、针对训练样本集利用自助采样法获取多个样本采样集;

B2、针对每一样本采样集进行数据处理获取适用于每一学习模型的训练样本;

B3、利用所述训练样本训练对应的学习模型并得到多个预先训练的弱学习器。

一种基于多模型融合的高速匝道行驶时间预测方法

技术领域

[0001] 本发明涉及机器学习技术领域,尤其涉及一种基于多模型融合的高速匝道行驶时间预测方法。

背景技术

[0002] 路段的行驶时间能够直观的反映道路拥挤程度。匝道行驶时间作为匝道控制策略优劣的重要评价指标,对其进行合理的预测,能够为匝道控制提供科学的数据支持。行驶时间的预测目前有很多种,比如参数模型有人工神经网络模型和卡尔曼滤波模型,他们都需要大量的标定参数。还有非参数模型,比如历史平均方法、KNN算法和非参数回归模型,这类算法应用就较为广泛,因为它们机制比较简单,且不需要设定大量参数。匝道的行驶时间属于短时行程,对于短时行程时间预测,已经有许多基于数据驱动的方法,其中KNN就比较常见,有许多人就在KNN的基础上进行一些改进来进行预测,比如Zhang L等人的一种改进的K近邻模型用于短期交通流预测,刘伟铭等人的基于PSO-LSSVM的高速公路短时行程时间预测。当然也有其他模型来进行短时行程时间的预测,比如基于SVM的城市快速路行程时间预测研究。

发明内容

[0003] (一)要解决的技术问题

[0004] 针对现有单一预测模型参数设置复杂,且预测效果稳定性差和泛化能力不强的问题,本发明提供一种基于多模型融合的高速匝道行驶时间预测方法。

[0005] (二)技术方案

[0006] 为了达到上述目的,本发明采用的主要技术方案包括:

[0007] S1、获取历史行驶时间数据,将所述历史行驶时间数据直接映射得到第一预测结果 SWL_0 ;

[0008] S2、利用多个预先训练的弱学习器获取多个第二预测结果 SWL_i ;

[0009] S3、利用所述第一预测结果 SWL_0 筛选所述多个第二预先结果 SWL_i 获取包含多个弱学习器模型的预测集合 SWL_j ;

[0010] S4、利用预测集合 SWL_j 中平均绝对误差百分比误差最小的两个弱学习器模型获取强学习器模型SLM,将预测集合 SWL_j 中所述选取的两个弱学习器模型剔除,并将强学习器模型SLM作为新的弱学习模型增加至预测集合 SWL ,更新强学习器模型和预测集合 SWL_j ;

[0011] S5、重复执行上述步骤S4直至强学习器模型SLM不再更新,利用最后一次更新的强学习器模型SLM作为高速匝道行驶时间的预测模型。

[0012] 可选地,在步骤S2和S3之间还包括:利用公式一计算所述第一预测结果的平均绝对百分比误差 $MAPE_0$ 和每一所述第二预测结果的平均绝对百分比误差 $MAPE_i$;

[0013] 公式一:
$$MAPE = \frac{1}{R} \sum_{r=1}^R \frac{1}{T} \sum_{t=1}^T \left| \frac{d_{rt} - p_{rt}}{d_{rt}} \right|$$

[0014] 其中,R表示测试数据的路线数量,T表示测试数据的时间段数量, d_{rt} 表示r路线t时间段的平均行驶时间, p_{rt} 表示表示r路线t时间段的预测行驶时间。

[0015] 可选地,在S3中,包含多个弱学习器模型的预测集合 SWL_j 中的每一个弱学习器模型的平均绝对百分比误差 $MAPE_j$ 均小于 $MAPE_0$ 。

[0016] 可选地,在步骤S4还包括:利用公式二获取预测集合 SWL_j ,并根据公式三更新预测集合 SWL_j ;

[0017] 公式二: $WL_j(i=1,2,\dots,J)=MAPE_j$;

[0018] 公式三: $SWL_j=(j,WL_j),j=1,2,\dots,J$;

[0019] 式中J表示模型中用到的机器学习模型数量,j表示对应的弱学习器, $MAPE_j$ 表示第j个模型的平均绝对误差百分比误差。

[0020] 可选地,在步骤S4中利用两个弱学习器模型构建强学习器模型SLM的方法包括:

[0021] A1、以预设迭代速率 v 调整每个弱学习器的权重比,得到多种权重比的强学习器模型;

[0022] A2、获取每一所述权重比的强学习器模型的平均绝对误差百分比误差,将平均绝对误差百分比误差最小时对应的强学习器模型作为最终的强学习器模型SLM。

[0023] 可选地,所述多个预先训练的弱学习器模型包括XGBoost、LightGBM、SVM、线性回归和KNN中的至少两种。

[0024] 可选地,步骤S1之前包括利用下述步骤获取多个预先训练的弱学习器:

[0025] B1、针对训练样本集利用自助采样法获取多个样本采样集;

[0026] B2、针对每一样本采样集进行数据处理获取适用于每一学习模型的训练样本;

[0027] B3、利用所述训练样本训练对应的学习模型并得到多个预先训练的弱学习器。

[0028] (三)有益效果

[0029] 本发明的有益效果是:本发明方法为高速匝道的形式时间预测提供了一种新思路,且获得的强学习器模型泛化能力强,且针对不同类型的数据预测结果较为稳定,具有较佳的鲁棒性;此本方法通过多次调整和融合机器学习模型的权重比使得最终获取的模型具有更好的准确性,为实际高速匝道行驶时间的预测提供了准确的预测数据,具有较佳的实用性。

附图说明

[0030] 图1为本发明一实施例提供能的一种基于多模型融合的高速匝道行驶时间预测方法流程图;

[0031] 图2为本发明一实施例提供能的融合策略示意图;

[0032] 图3为本发明一实施例提供的模型选择时各模型效果图;

[0033] 图4为本发明一实施例提供的进行融合时模型比例选择图;

[0034] 图5为本发明一实施例提供的应用融合模型的测试效果图。

具体实施方式

[0035] 为了更好的解释本发明,以便于理解,下面结合附图,通过具体实施方式,对本发明作详细描述。

[0036] 实施例

[0037] 如图1所示,本实施例提供了一种基于多模型融合的高速匝道行驶时间预测方法,具体包括:

[0038] 如图2所示,利用下述步骤获取多个预先训练的弱学习器:

[0039] B1、针对训练样本集利用自助采样法获取多个样本采样集;

[0040] B2、针对每一样本采样集进行数据处理获取适用于每一学习模型的训练样本;

[0041] B3、利用所述训练样本训练对应的学习模型并得到多个预先训练的弱学习器。

[0042] 举例来说,本实施例选取XGBoost(以下简称Xgb)、LightGBM、SVM、线性回归(Linear regression)和KNN作为多个预先训练的弱学习器模型包括。

[0043] S1、获取历史行驶时间数据,将所述历史行驶时间数据直接映射得到第一预测结果 SWL_0 ;

[0044] 举例来说,直接映射(Direct mapping)是将历史数据直接作为预测数据,例如即直接用6点到8点的路口A到收费站1的平均旅行时间作为8点值10点钟路口A到收费站1的平均旅行时间。

[0045] S2、利用多个预先训练的弱学习器获取多个第二预测结果 SWL_i ;优选地,在步骤S2和S3之间还包括:利用公式1计算所述第一预测结果的平均绝对百分比误差 $MAPE_0$ 和每一所述第二预测结果的平均绝对百分比误差 $MAPE_i$;

[0046] 公式1:
$$MAPE = \frac{1}{R} \sum_{r=1}^R \frac{1}{T} \sum_{t=1}^T \left| \frac{d_{rt} - p_{rt}}{d_{rt}} \right|$$

[0047] 其中,R表示测试数据的路线数量,T表示测试数据的时间段数量, d_{rt} 表示r路线t时间段的平均行驶时间, p_{rt} 表示表示r路线t时间段的预测行驶时间。S3、利用所述第一预测结果 SWL_0 筛选所述多个第二预先结果 SWL_i 获取包含多个弱学习器模型的预测集合 SWL_j ;

[0048] 具体地,举例来说,S3中,包含多个弱学习器模型的预测集合 SWL_j 中的每一个弱学习器模型的平均绝对百分比误差 $MAPE_j$ 均小于 $MAPE_0$ 。

[0049] 举例来说,将直接映射得到的预测结果作为比较基石,仅在弱学习器的测结果的MAPE值小于直接映射得到的预测结果的MAPE值时才保留该模型即当单一模型比这个结果好是时,再考虑是否将那个模型纳入模型融合;如图3所示,仅SVM学习器和XGBoost学习器的MAPE小于直接映射得到的预测结果的MAPE值。

[0050] S4、利用预测集合 SWL_j 中平均绝对误差百分比误差最小的两个弱学习器模型获取强学习器模型SLM,将预测集合 SWL_j 中所述选取的两个弱学习器模型剔除,并将强学习器模型SLM作为新的弱学习模型增加至预测集合 SWL ,更新强学习器模型和预测集合 SWL_j ;

[0051] 具体地举例来说,利用公式2获取预测集合 SWL_j ,并根据公式3更新预测集合 SWL_j ;

[0052] 公式2: $WL_j (i = 1, 2, \dots, J) = MAPE_j$;

[0053] 公式3: $SWL_j = (j, WL_j), j = 1, 2, \dots, J$;

[0054] 式中J表示模型中用到的机器学习模型数量,j表示对应的弱学习器, $MAPE_j$ 表示第j个模型的平均绝对误差百分比误差。

[0055] 举例来说,在本实施例中首先以SVM学习器和XGBoost学习器预测强学习器模型,并在将该强学习器模型作为新的弱学习器器与剩余的弱学习中MAPE值最小的弱学习进行再次融合更新强学习器模型。

[0056] 在步骤S4中利用两个弱学习器模型构建强学习器模型SLM的方法包括：

[0057] A1、以预设迭代速率 v 调整每个弱学习器的权重比，得到多种权重比的强学习器模型；

[0058] A2、获取每一所述权重比的强学习器模型的平均绝对误差百分比误差，将平均绝对误差百分比误差最小时对应的强学习器模型作为最终的强学习器模型。具体地，如图4在本实施例中，以0.1为迭代速率，调整了XGBoost模型和SVM模型的权重比；

[0059] S5、重复执行上述步骤S4直至强学习器模型SLM不再更新，利用最后一次更新的强学习器模型SLM作为高速匝道行驶时间的预测模型。

[0060] 举例来说，在本实施例中，XGBoost和SVM对高速匝道行驶时间进行融合预测，并且XGBoost所占比为7，SVM所占比为3，既最后用XGBoost模型预测的结果乘以0.7加上SVM预测结果乘以0.3为最终预测结果；通过多次调整和融合机器学习模型的权重比使得最终获取的模型具有更好的准确性，为实际高速匝道行驶时间的预测提供了准确的预测数据，具有较佳的实用性。

[0061] 仿真实例

[0062] 如图5所示，在本实施例中选用上述实施例构建的强学习器模型和前述的多个弱学习器分别对测试数据进行预测，并得到了每种方法的预测结果，通过结果可以看出针对测试数据和训练数据，SVM学习器和XGBoost表现出的预测效果不同，在不同的数据情况下，单一模型最优的可能不是同一个，但是融合后的模型比单一模型效果要好，总结起来就是模型融合能够增强预测模型的泛化能力，也能一定程度上提升预测的精度。

[0063] 最后应说明的是：以上所述的各实施例仅用于说明本发明的技术方案，而非对其限制；尽管参照前述实施例对本发明进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述实施例所记载的技术方案进行修改，或者对其中部分或全部技术特征进行等同替换；而这些修改或替换，并不使相应技术方案的本质脱离本发明各实施例技术方案的范围。

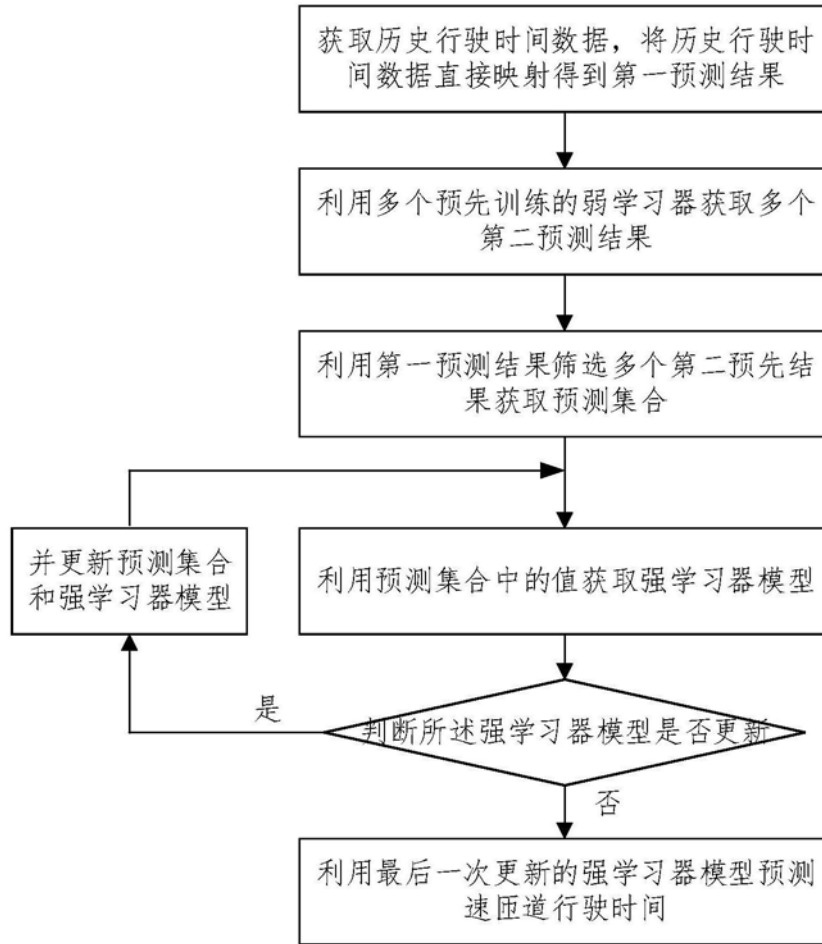


图1

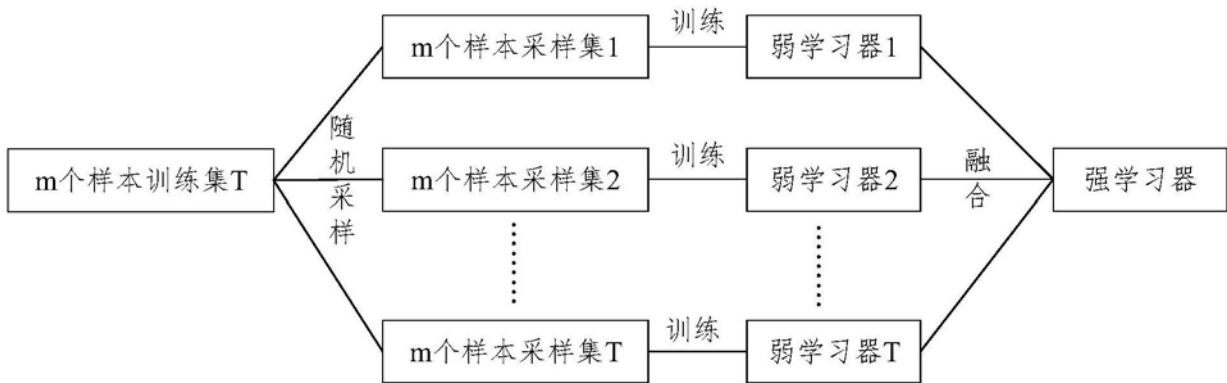


图2

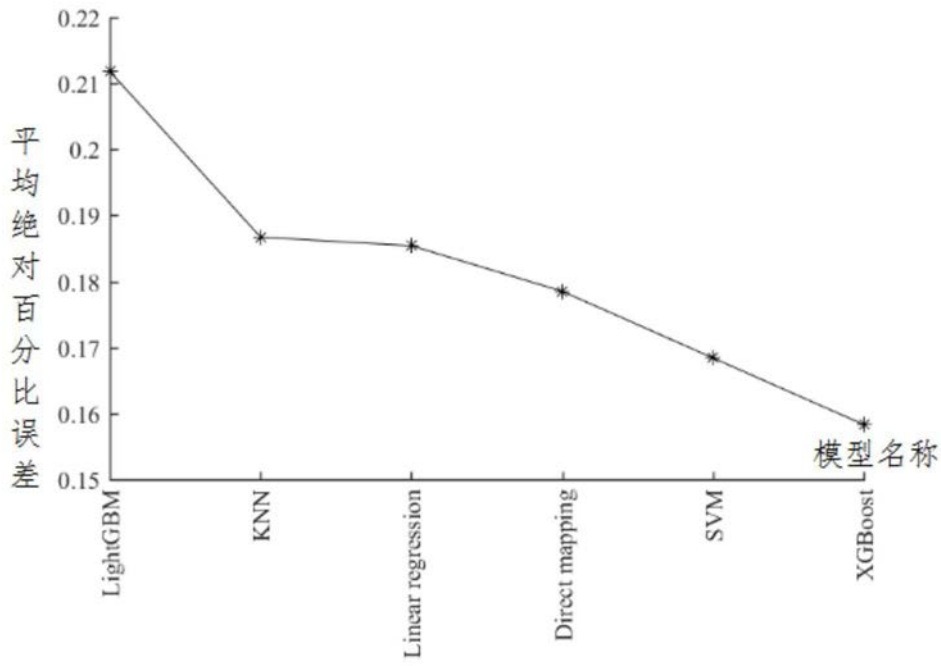


图3

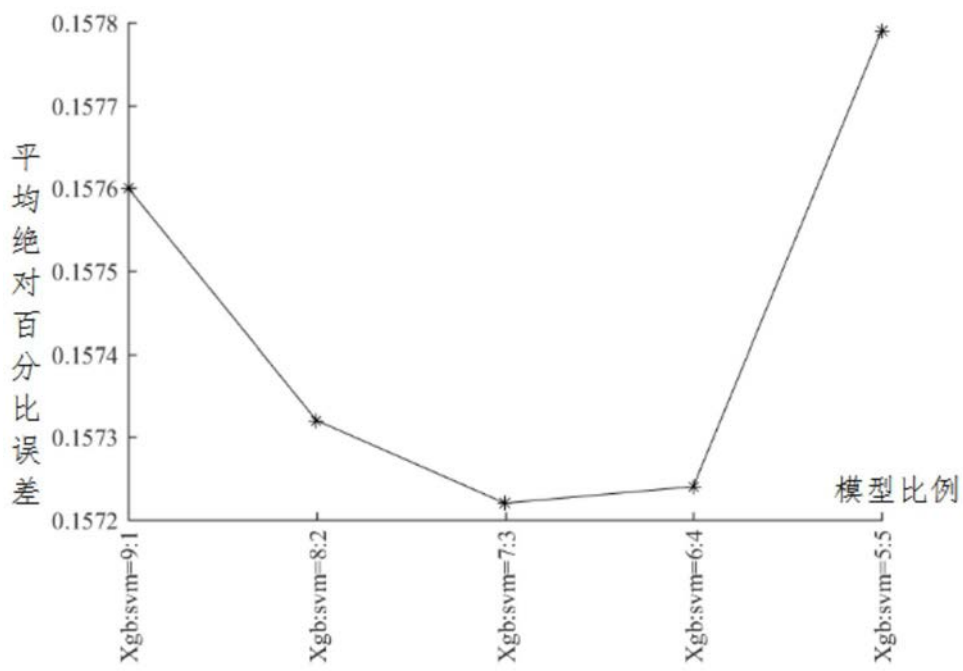


图4

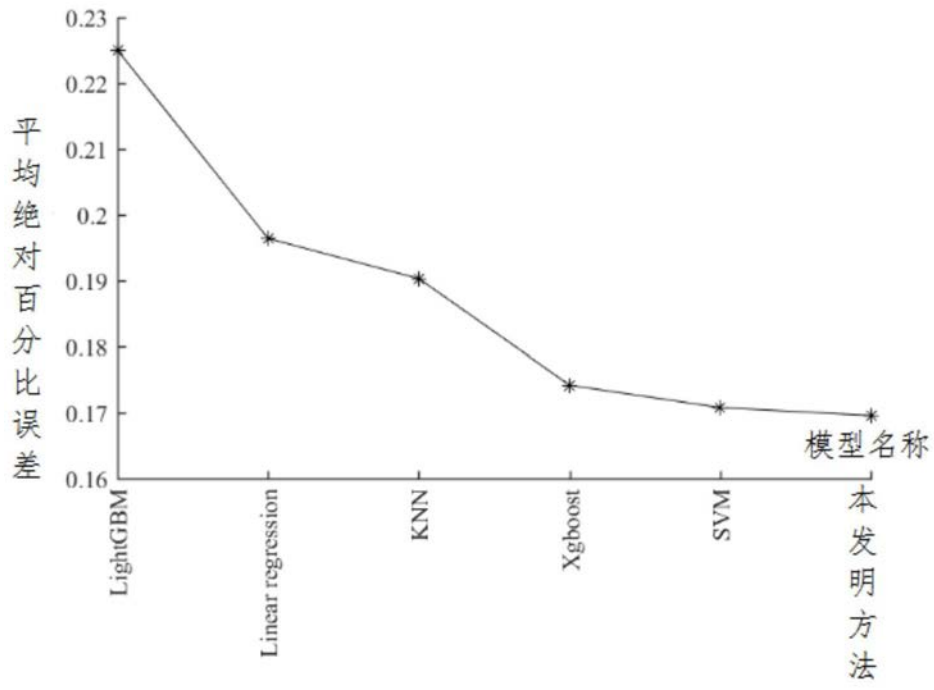


图5