



(12) 发明专利申请

(10) 申请公布号 CN 114049144 A

(43) 申请公布日 2022. 02. 15

(21) 申请号 202111281896.1

G06N 3/08 (2006.01)

(22) 申请日 2021.11.01

(71) 申请人 江苏科技大学

地址 212008 江苏省镇江市京口区梦溪路2号

(72) 发明人 熊罗庚 郑尚 邹海涛 于化龙 高尚

(74) 专利代理机构 南京苏高专利商标事务所 (普通合伙) 32204

代理人 曹坤

(51) Int. Cl.

G06Q 30/02 (2012.01)

G06Q 50/28 (2012.01)

G06K 9/62 (2022.01)

G06N 3/04 (2006.01)

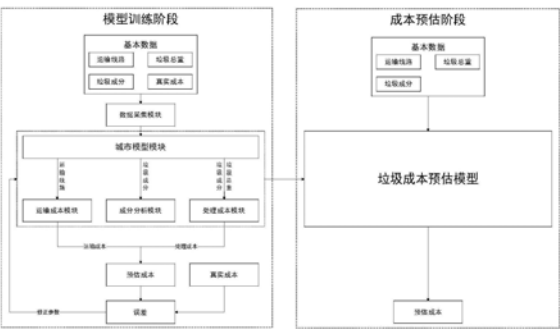
权利要求书1页 说明书4页 附图1页

(54) 发明名称

一种建立多方法决策的垃圾成本预估数字孪生系统的方法

(57) 摘要

本发明公开了一种建立多方法决策的垃圾成本预估数字孪生系统的方法。属于城市垃圾处理成本估算领域,具体步骤:通过样本采集模块收集城市当地的数据采集库中关于垃圾处理的历史样本;后通过城市模型模块建立城市模型存放并分析历史样本,建立垃圾成本预估所必要的前置条件;将历史样本赋予垃圾成本预估所必要的前置条件后,将其作为数据集,对运输成本模块、成分分析模块及处理成本模块进行训练,训练完成后即获得垃圾成本预估模型,即建立多方法决策的垃圾成本预估数字孪生系统。本发明根据数据库中的历史信息来预测未知数据,预测出的成本具有一定的可信度,预测速度较快,并且可以存储至移动端及固定端设备中,方便决策者随时随地进行预测。



1. 一种建立多方法决策的垃圾成本预估数字孪生系统的方法,其特征在于,

所述垃圾成本预估数字孪生系统包括样本采集模块、城市模型模块、运输成本模块、成分分析模块及处理成本模块;其具体操作步骤如下所述:

(1)、通过样本采集模块连接到城市当地的数据采集库,收集数据库中关于垃圾处理的历史样本;

(2)、根据收集的历史样本,通过城市模型模块建立城市模型用于存放并分析历史样本,并建立垃圾成本预估所必要的前置条件;

(3)、将历史样本赋予垃圾成本预估所必要的前置条件后,将其作为数据集,从而对运输成本模块、成分分析模块及处理成本模块进行训练,训练完成后即获得垃圾成本预估模型,即建立多方法决策的垃圾成本预估数字孪生系统。

2. 根据权利要求1所述的一种建立多方法决策的垃圾成本预估数字孪生系统的方法,其特征在于,在步骤(1)中,所述通过样本采集模块收集的历史样本包括垃圾运输路线、垃圾的总重量、垃圾的主要组成成分和处理垃圾所花费的真实成本。

3. 根据权利要求1所述的一种建立多方法决策的垃圾成本预估数字孪生系统的方法,其特征在于,在步骤(2)中,所述垃圾成本预估所必要的前置条件包括但不限于城市道路索引及垃圾处理的真实成本等级。

4. 根据权利要求1所述的一种建立多方法决策的垃圾成本预估数字孪生系统的方法,其特征在于,在步骤(3)中,所述运输成本模块,由Word2Vec嵌入模型与长短时神经网络组成。

5. 根据权利要求1所述的一种建立多方法决策的垃圾成本预估数字孪生系统的方法,其特征在于,在步骤(3)中,所述成分分析模块由一个随机森林组成。

6. 根据权利要求1所述的一种建立多方法决策的垃圾成本预估数字孪生系统的方法,其特征在于,在步骤(3)中,所述处理成本模块,由数个多项式回归函数组成。

一种建立多方法决策的垃圾成本预估数字孪生系统的方法

技术领域

[0001] 本发明属于城市垃圾处理成本估算领域,涉及一种建立多方法决策的垃圾成本预估数字孪生系统的方法。

背景技术

[0002] 垃圾是人类日常生活和工作中产生的废弃物,具有排出量大,成分复杂,有污染性,可回收等特点。如果不能妥善处理垃圾,则会带来污染环境,浪费资源等后果,进而对人们的生活造成重大影响。而垃圾分类处理可以有效处理以上问题,例如,干垃圾用于焚烧发电,湿垃圾用于堆肥,可回收垃圾被回收后二次利用,这样不仅解决了垃圾污染问题,对社会也具有一定价值。因此,垃圾的分类与处理是当今社会必须面临的重大任务和挑战。

[0003] 然而,部分地区的垃圾处理系统信息化程度较低,既无法让决策者快速掌握城市垃圾的具体情况,也无法做到通过数据建立城市模型,从而预测处理垃圾所需要的成本。并且,垃圾分类处理往往伴随着高额的成本,由于影响因素过多,导致无法准确预估处理成本,因此在开展垃圾分类处理的工作在部分地区较为困难。本专利提出一种多方法决策的垃圾成本预估数字孪生系统,通过从基层采集的数据构建模型,使用LSTM计算垃圾运输成本;使用随机森林判别垃圾处理的大致成本等级;使用线性决策函数对垃圾处理成本进行精确预估。该系统训练完成后可存储于移动端与固定端设备中,方便决策者浏览信息与进行成本预估。本专利通过处理垃圾成本的预估,在一定程度上辅助决策者规划城市环保建设。

[0004] 长短期记忆网络LSTM(Long Short-Term Memory)是一种循环神经网络,适合于处理和预测时间序列中间隔和延迟非常长的重要事件。LSTM在各领域都有不错的表现,例如机器翻译,语音识别,文本分类等。

[0005] 专利CN202110691506.1基于LSTM开发出了一种焚烧垃圾的仿真算法,该专利对焚烧空间的参数进行了详细的规划,通过对垃圾焚烧的图像进行分析,得到焚烧的确切效果。该专利的缺点是未能对焚烧的成本进行预估,无法为决策者提供所需成本方面的建议;因此,如何解决上述问题就值得技术人员思考了。

发明内容

[0006] 发明目的:本发明目的是提供了一种建立多方法决策的垃圾成本预估数字孪生系统的方法,采用由LSTM构成的运输成本模块,由模糊系统构成的成分分析模块,由线性回归决策函数构成的处理成本模块共同预测垃圾处理所需的成本。

[0007] 技术方案:本发明所述的一种建立多方法决策的垃圾成本预估数字孪生系统的方法;

[0008] 所述垃圾成本预估数字孪生系统包括样本采集模块、城市模型模块、运输成本模块、成分分析模块及处理成本模块;其具体操作步骤如下所述:

[0009] (1)、通过样本采集模块连接到城市当地的数据采集库,收集数据库中关于垃圾处

理的历史样本；

[0010] (2)、根据收集的历史样本,通过城市模型模块建立城市模型用于存放并分析历史样本,并建立垃圾成本预估所必要的前置条件；

[0011] (3)、将历史样本赋予垃圾成本预估所必要的前置条件后,将其作为数据集,从而对运输成本模块、成分分析模块及处理成本模块进行训练,训练完成后即获得垃圾成本预估模型,即建立多方法决策的垃圾成本预估数字孪生系统。

[0012] 进一步的,在步骤(1)中,所述通过样本采集模块收集的历史样本包括垃圾运输路线、垃圾的总重量、垃圾的主要组成成分和处理垃圾所花费的真实成本。

[0013] 进一步的,在步骤(2)中,所述垃圾成本预估所必要的前置条件包括但不限于城市道路索引及垃圾处理的真实成本等级。

[0014] 进一步的,在步骤(3)中,所述运输成本模块,由Word2Vec嵌入模型与长短时神经网络(LSTM)组成。

[0015] 进一步的,在步骤(3)中,所述成分分析模块由一个随机森林组成。

[0016] 进一步的,在步骤(3)中,所述处理成本模块,由数个多项式回归函数组成。

[0017] 进一步的,所述垃圾成本预估数字孪生系统训练完成后,可存储至包括但不限于手机,电脑,平板等设备中,方便使用者随时随地进行信息的查看与预测。

[0018] 有益效果:本发明与现有技术相比,本发明根据数据库中的历史信息来预测未知数据,预测出的成本具有一定的可信度,预测速度较快,并且可以存储至移动端,固定端设备中,方便决策者随时随地进行预测。本发明的技术方案填补了垃圾处理成本预估方面的空白。

附图说明

[0019] 图1是本发明的操作流程；

[0020] 图2是本发明实施例的对比示意图。

具体实施方式

[0021] 以下结合附图和具体实施例,对本发明做出进一步说明。

[0022] 如图1所述,本发明所述的一种建立多方法决策的垃圾成本预估数字孪生系统的方法,其特征在于,

[0023] 所述垃圾成本预估数字孪生系统包括样本采集模块、城市模型模块、运输成本模块、成分分析模块及处理成本模块;其具体操作步骤如下所述:

[0024] (1)、通过样本采集模块连接到城市当地的数据采集库(城市当地的垃圾处理厂或相关部门),收集数据库中关于垃圾处理的历史样本;

[0025] 所述通过样本采集模块收集的历史样本具体包括垃圾运输路线、垃圾的总重量、垃圾的主要组成成分和处理垃圾所花费的真实成本;本案中垃圾的组成大致分为:可回收垃圾,干垃圾及湿垃圾三种成分;

[0026] (2)、根据收集的历史样本,通过城市模型模块建立城市模型用于存放并分析历史样本,并建立垃圾成本预估所必要的前置条件;

[0027] 所述垃圾成本预估所必要的前置条件包括但不限于城市道路索引及垃圾处理

的真实成本等级等;

[0028] 具体的,

[0029] 一、道路索引:城市中每条道路的唯一标识符,由与道路数量相等的维度的独热向量组成(由长度与道路总数相等的独热向量组成);如:城市共有X条街道,则有X个向量与之一一对应,这些向量分别为:10000...,01000...,00100...,00010...,00001...,以此类推,每个向量的长度均为X;

[0030] 二、成本等级:将历史样本以真实成本升序排列,并等分为三部分。其中,真实成本最低的部分划分为成本等级A,真实成本居中的部分划分为成本等级B,真实成本最高的部分划分为成本等级C;

[0031] (3)、开始系统训练:将历史样本赋予垃圾成本预估所必要的前置条件后,将其作为数据集,从而对运输成本模块、成分分析模块及处理成本模块进行训练,训练完成后即获得垃圾成本预估模型,即建立多方法决策的垃圾成本预估数字孪生系统;

[0032] 具体的,对于每一批次运输的垃圾:

[0033] 所述运输成本模块,由Word2Vec嵌入模型与长短时神经网络(LSTM)组成;

[0034] 在训练运输成本模块中,首先输入运输路线的向量,随后送入Word2Vec模型中进行嵌入,建立道路之间的联系,最后再进入LSTM网络得到预估的运输成本;其中部分参数如下:

[0035] Word2Vec向量长度:300;

[0036] LSTM神经元个数:256;

[0037] IterMax:50;

[0038] Batch Size:32;

[0039] L2:0.01;

[0040] Learning Rate:0.001;

[0041] 所述成分分析模块由一个随机森林组成;

[0042] 在训练成分分析模块中,接收垃圾的主要组成成分与成本等级,使用随机森林自动建立垃圾成分与成本等级之间的依赖关系;该任务为典型的多分类任务,分类特征为可回收物占比K,干垃圾占比G与湿垃圾占比S,数据标签为成本等级A,B,C;

[0043] 一、设置用于生成决策树的特征池;垃圾的主要组成成分包括:可回收物占比K(废纸,塑料,玻璃,金属,布料等),干垃圾占比G(砖瓦陶瓷、渣土、卫生间废纸、纸巾等),湿垃圾占比S(剩菜剩饭、骨头、菜根菜叶、果皮等)。因此,对于可回收物占比K,干垃圾占比G与湿垃圾占比S,按每5%进行一次划分,即0-5%,5%-10%,10%-15%...90%-95%,95%-100%,共分为20个段落,将K,G与S的段落送入特征池,则特征池内共60个特征;

[0044] 二、从数据集中有放回地抽取训练集;共抽取100个训练集,每个训练集的大小是数据集的50%;

[0045] 三、从特征池中每次随机挑选50%的特征建立决策树;每次树进行分裂时,选择信息增益最大的特征进行分裂;其中信息增益的公式为:

[0046]
$$Gain = - \sum_{i=1}^n p_i \log_2(p_i)$$

[0047] 其中, p_i 表示树分裂时每个分支的概率大小;共建立100棵决策树形成随机森林,

使用上述的训练集一对一地训练决策树；

[0048] 所述处理成本模块，由数个多项式回归函数组成；

[0049] 在训练处理成本模块中，在各个成本等级下建立各自的关于垃圾重量与处理成本的多项式回归函数；对于每一个成本等级：输入该等级下包含的所有数据的垃圾总重与真实成本，建立平面坐标系，确定垃圾总重与真实成本的函数；函数方程式为：

$$[0050] \quad \hat{y}_i = ax_i^2 + bx_i + c$$

$$[0051] \quad s.t. \min(\sum_{i=1}^n (y_i - \hat{y}_i)^2)$$

[0052] 其中， x_i 表示输入的垃圾总重， $\hat{y}_i$ 表示成本预测值， y_i 表示真实成本； a, b, c 表示系统需要学习的参数，均为实数；处理成本模块所得预测值 $\hat{y}_i$ 即为预估处理成本。

[0053] 将预估运输成本和预估处理成本相加后得到预估成本。预估成本与真实成本之间存在误差，使用该误差反向调节模块中的参数。训练次数达到IterMax后结束训练，得到垃圾处理成本预估模型。

[0054] 进一步的，对于一个新的垃圾处理样本，将数据输入至垃圾成本预估模型后，输出模型对该次垃圾处理所预估的成本需求。

[0055] 进一步的，所述垃圾成本预估数字孪生系统训练完成后，可存储至包括但不限于手机，电脑，平板等设备中，方便使用者随时随地进行信息的查看与预测。

[0056] 为验证本方法的性能，在公开数据集上进行留一交叉验证法，使用平均绝对误差(MAE)作为性能的度量值，误差值越小，则模型的性能越好；误差的计算公式如下：

$$[0057] \quad MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

[0058] 其中， $\hat{y}_i$ 为成本预测值， y_i 为真实成本。

[0059] 为验证本方法的性能的优越性，使用相同数据集对包含本方法在内的数个预测方法进行性能检验；首先，随机抽取数据集中70%的数据样本进行模型训练，训练完成后，使用剩余30%的数据样本进行性能验证；图2所示为各方法进行性能验证时的平均绝对误差，其中线性回归模型的误差值高达70.56%，说明该模型不适合进行垃圾成本预测；逻辑回归模型与多项式回归模型的误差值分别为62.59%与60.68%，其误差值仍然偏高；而本文方法的误差值仅为49.35%，远低于其他三种方法，说明该方法相较于前三种方法更适合用于垃圾成本预测。

[0060] 以上仅是本发明的优选实施方式，本发明的保护范围并不局限于上述实施例，凡属于本发明思路下的技术方案均属于本发明的保护范围。应当指出，对于本技术领域的普通技术人员来说，在不脱离本发明原理前提下的若干改进和润饰，应视为本发明的保护范围。

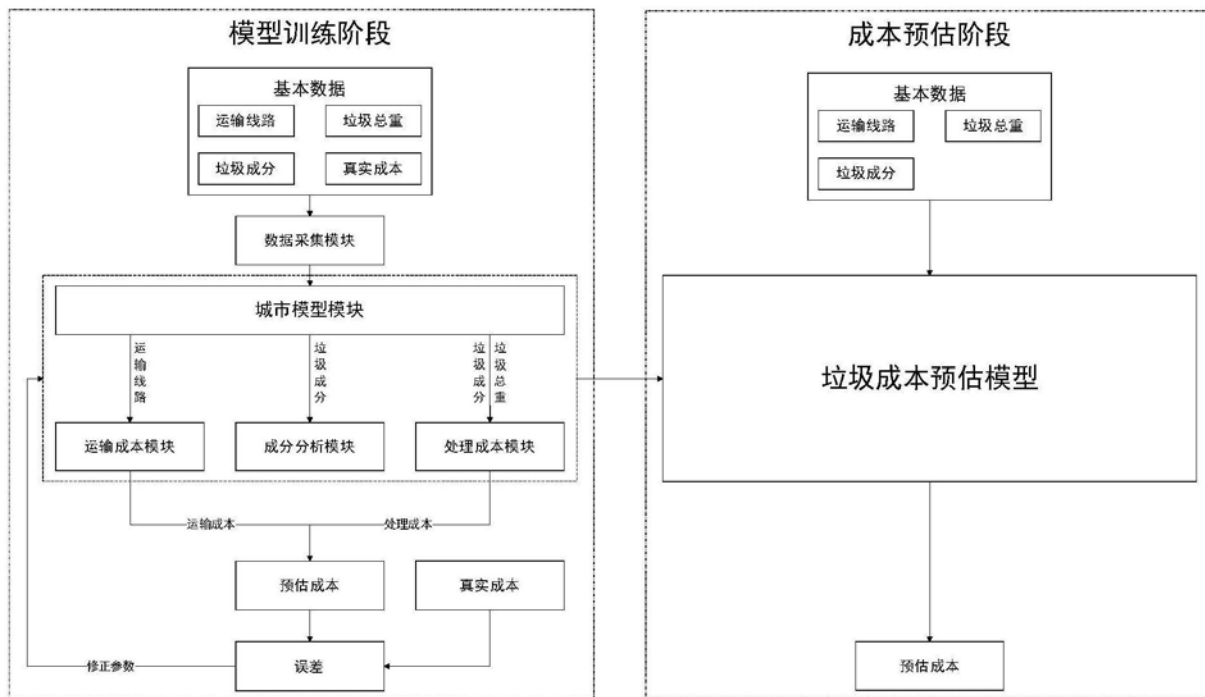


图1

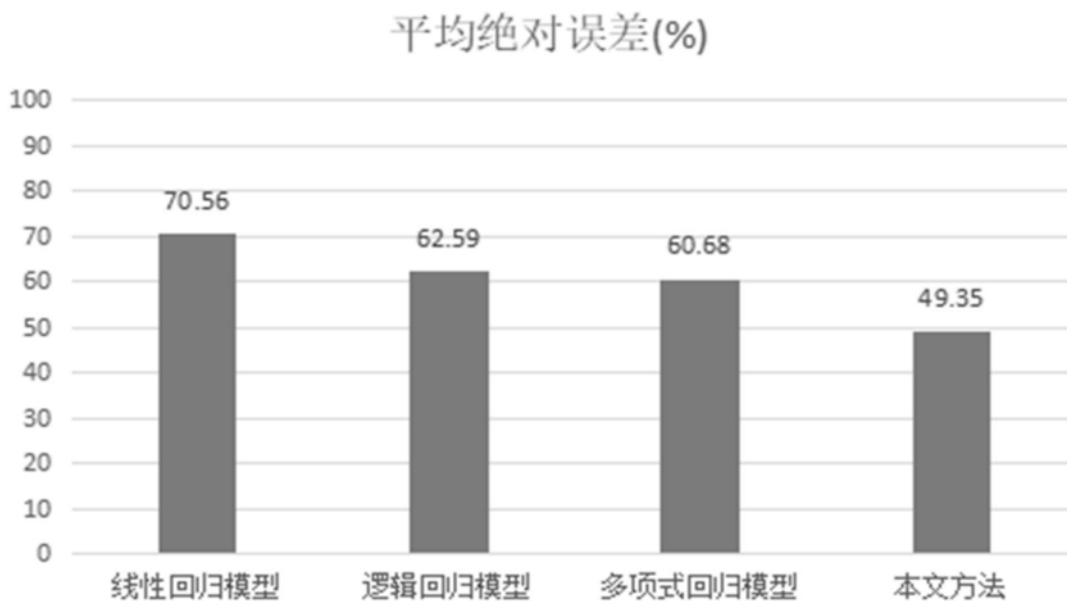


图2