



(12) 发明专利申请

(10) 申请公布号 CN 111899511 A

(43) 申请公布日 2020. 11. 06

(21) 申请号 202010768642.1

(22) 申请日 2020.08.03

(71) 申请人 西南交通大学

地址 610031 四川省成都市二环路北一段

(72) 发明人 叶彭姚 杜孟秋 张南 肖炎

邓艺峰

(74) 专利代理机构 北京金智普华知识产权代理有限公司 11401

代理人 杨采良

(51) Int. Cl.

G08G 1/01 (2006.01)

G08G 1/123 (2006.01)

G06N 3/02 (2006.01)

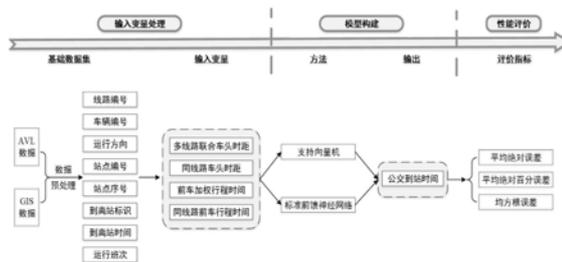
权利要求书4页 说明书21页 附图5页

(54) 发明名称

一种共线线路AVL数据的公交车到站时间预测方法

(57) 摘要

本发明属于城市智能公共交通信息处理技术领域,公开了一种共线线路数据的公交车到站时间预测方法,采集智能公交运营基础数据并建立数据库;对采集到的数据进行预处理;搭建预测模型框架;基于搭建的预测模型框架,确定预测模型结构,构建公交车到站时间预测模型;利用构建得到的公交车到站时间预测模型进行公交车到站时间预测。本发明提出了一种基于共线线路运营数据的到站时间预测方法,通过引入其他共线线路的运营信息,以细化车辆间的车头时距间隔,提高历史数据的时效性和可靠性,同时也能将线路车辆间的相互影响纳入考虑,减小了干扰因素对公交到站时间预测带来的影响。



CN 111899511 A

1. 一种共线线路数据的公交车到站时间预测方法,其特征在于,所述共线线路数据的公交车到站时间预测方法包括:

采集智能公交运营基础数据并建立数据库;

对采集到的数据进行预处理;搭建预测模型框架;

基于搭建的预测模型框架,确定预测模型结构,构建公交车到站时间预测模型;

利用构建得到的公交车到站时间预测模型进行公交车到站时间预测。

2. 如权利要求1所述共线线路数据的公交车到站时间预测方法,其特征在于,所述公交运营基础数据包括:公交车自动定位数据即AVL数据或者GPS数据和公交线路站点地理信息数据即GIS数据。

3. 如权利要求1所述共线线路数据的公交车到站时间预测方法,其特征在于,所述预处理方法包括:

(1) AVL数据与GIS数据关联:利用AVL数据、GIS数据共有的字段ROUTEID和STATIONNUM进行关联,给AVL数据匹配每条线路的站点名称、站点序号、运营方向信息;

(2) 公交运行班次划分:

(2.1) 将原始数据按日期、线路号、车辆编号和记录时间进行排序,依次比较下一行数据和本行数据的日期、线路号、车辆编号和运行方向字段,当其中任一字段的值发生改变,则运行班次-temp字段值增加1;

(2.2) 基于第一次班次划分的结果再划分一次运营班次,根据运营班次-temp字段和站点序号对数据进行排序,比较下一行数据和本行数据的运营班次-temp字段和站点序号字段,当运营班次-temp字段发生改变或下一行数据的站点序号小于本行数据,则运营班次字段值增加1;将得到的运营班次字段将作为最终的公交运营班次值存储于原数据中;

(3) AVL数据处理:

(3.1) 重复数据处理:同一班次到离站数据在首末站的重复数据,保留时间最晚的一条数据作为首站点的出站记录,保留时间最早的一条数据作为末站点的到站记录;同一班次到离站数据在中间站点的重复数据,保留同一站点中时间最早和时间最晚的两条数据,分别作为本站的到站和离站记录;

(3.2) 异常数据处理:根据异常数据的特点,以运行班次为单位,按进站时间的先后对数据排序,若出现站点次序倒置或站点编号顺序错乱的数据,则标记并删除该数据;

(3.3) 缺失数据处理:提取共线运营的不同线路经过的相同站点的到离站数据,再根据各班次总的到离站记录的数量和共线运营的不同线路经过的相同站点的数量来判定该班次数据是否存在缺失;当各班次总的到站或离站记录数小于选取的公交通道站点数,标记并删除该班次的到离站数据。

4. 如权利要求1所述共线线路数据的公交车到站时间预测方法,其特征在于,所述预测模型框架搭建方法包括:通过模型假设、参数设置、车头时距变量处理和行程时间变量处理进行预测模型构建;

1) 确定预测目标:

所述预测目标为基于线路 j 和 j' ($j \neq j' \in J$) 的车辆运行信息预测线路 j ($j \in J$) 的第 i 辆公交在站点 $k+\mu$ 处的到站时间,即预测线路 j ($j \in J$) 的第 i 辆公交在站点 k 和站点 $k+\mu$ 之间的行程时间;公式如下:

$$t_{a,k+\mu}^{i,j} = t_{a,k}^{i,j} + \hat{T}_{k,k+\mu}^i;$$

其中, $t_{a,k+\mu}^{i,j}$ 表示第*i*辆公交在站点*k*+ μ 处的到达时间; $t_{a,k}^{i,j}$ 表示第*i*辆公交在站点*k*处的到达时间; $\hat{T}_{k,k+\mu}^i$ 表示第*i*辆公交在站点*k*和站点*k*+ μ 之间的行程时间预测值;

2) 进行模型假设:

所述模型假设包括:

- a) 在公交通道中, 超车行为不允许发生;
- b) 在不同线路的所有车辆中, 驾驶员状况均相同;
- c) 公交车辆之间及车辆和调度中心之间能够实现实时信息的交互;

3) 确定车头时距变量:

选择路段起点处的车头时距作为预测模型的输入变量之一, 当目标车辆*i*到达站点*k*时, 已有*n*辆公交经过了站点*k*+ μ , *p*为仍在公交通道上行驶的车辆数, 则两类车头时距变量的计算公式如下:

$$H_k^{i,J} = t_{a,k}^{i,j} - t_{a,k}^{i-p-1,J}$$

$$H_k^{i,j} = t_{a,k}^{i,j} - t_{a,k}^{i-p-n,j}$$

式中: $H_k^{i,J}$ 表示目标车辆*i*和前车*i*-*p*-1在站点*k*处的联合车头时距, 其中前车*i*-*p*-1属于线路集合*J*中的任意线路; $H_k^{i,j}$ 表示目标车辆*i*和前车*i*-*p*-*n*在站点*k*处的车头时距, 其中前车*i*-*p*-*n*和目标车辆*i*为同一线路; $t_{a,k}^{i,j}$ 为目标车辆*i*在站点*k*处的到达时间; $t_{a,k}^{i-p-1,J}$ 为公交*i*-*p*-1在站点*k*处的到达时间; $t_{a,k}^{i-p-n,j}$ 为公交*i*-*p*-*n*在站点*k*处的到达时间;

4) 确定行程时间变量:

利用车头时距的倒数对已经通过的前几辆车的行程时间进行加权, 加权平均行程时间的计算公式如下:

$$\bar{T}_{k,k+\mu}^i = \sum_{m=1}^{\sigma} \frac{1/t_{a,k}^{i,j} - t_{a,k}^{i-p-m,J}}{\Gamma} \times T_{k,k+\mu}^{i-p-m,J}$$

$$\Gamma = \sum_{m=1}^{\sigma} \frac{1}{t_{a,k}^{i,j} - t_{a,k}^{i-p-m,J}}$$

$$T_{k,k+\mu}^{i-p-m,J} = t_{a,k+\mu}^{i-p-m,J} - t_{a,k}^{i-p-m,j}, m = 1, \dots, \sigma$$

式中: $\bar{T}_{k,k+\mu}^i$ 表示线路集合*J*中任意线路前 σ 辆车的加权平均行程时间; Γ 表示权重和; σ 为加权时选取的前车数量, 该参数在模型预测阶段需先优化; $T_{k,k+\mu}^{i-p-m,J}$ 表示公交*i*-*p*-*m*在站点*k*和站点*k*+ μ 之间的行程时间; $t_{a,k+\mu}^{i-p-m,j}$ 表示公交*i*-*p*-*m*在站点*k*+ μ 处的到达时间;

$t_{a,k}^{i-p-m,j}$ 表示公交*i-p-m*在站点*k*处的到达时间;

5) 选取和目标车辆线路相同的前车的行程时间作为预测模型的输入变量,计算公式如下:

$$T_{k,k+\mu}^{i-p-n,j} = t_{a,k+\mu}^{i-p-n,j} - t_{a,k}^{i-p-n,j} ;$$

式中: $T_{k,k+\mu}^{i-p-n,j}$ 表示与公交*i*线路号相同的车辆*i-p-n*在站点*k*和站点*k+μ*之间的行程时间; $t_{a,k+\mu}^{i-p-n,j}$ 和 $t_{a,k}^{i-p-n,j}$ 分别表示公交*i-p-n*在站点*k+μ*和站点*k*处的到达时间;

6) 得到基于共线线路的公交行程时间预测模型为:

$$\hat{T}_{k,k+\mu}^i = f(H_k^{i,j}, H_k^{i,j}, \bar{T}_{k,k+\mu}^i, T_{k,k+\mu}^{i-p-n,j}) ;$$

式中: $\hat{T}_{k,k+\mu}^i$ 表示公交*i*在站点*k*和站点*k+μ*之间的行程时间预测值; $f(\cdot)$ 表示输入变量 $H_k^{i,j'}$ 、 $H_k^{i,j}$ 、 $\bar{T}_{k,k+\mu}^i$ 、 $T_{k,k+\mu}^{i-p-n,j}$ 和输出变量 $\hat{T}_{k,k+\mu}^i$ 之间待训练的函数关系。

5. 如权利要求1所述共线线路数据的公交车到站时间预测方法,其特征在于,所述公交车到站时间预测模型构建方法包括:

采用支持向量回归模型进行公交到站时间预测;选取径向基函数作为所用支持向量机模型的核函数,采用网格搜索的方法获取模型参数最优值;

采用神经网络模型进行公交到站时间预测;选取标准的三层前馈神经网络用于公交到站时间的预测,采用量化共轭梯度算法对神经网络模型训练;并进行敏感性分析获取参数的最优取值;

选取平均绝对误差、平均绝对百分误差和均方根误差三个指标来评价预测上述两种模型的优劣程度;评价指标的计算公式如下:

$$MAE = \frac{\sum_{i=1}^N |T_{k,k+\mu}^i - \hat{T}_{k,k+\mu}^i|}{N} ;$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|T_{k,k+\mu}^i - \hat{T}_{k,k+\mu}^i|}{T_{k,k+\mu}^i} \times 100\% ;$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (T_{k,k+\mu}^i - \hat{T}_{k,k+\mu}^i)^2}{N}} ;$$

式中: $T_{k,k+\mu}^i$ 表示公交*i*在站点*k*和站点*k+μ*之间的实际行程时间; N 为观测到的预测样本数量。

6. 一种计算机设备,其特征在于,所述计算机设备包括存储器和处理器,所述存储器存储有计算机程序,所述计算机程序被所述处理器执行时,使得所述处理器执行如下步骤:

采集智能公交运营基础数据并建立数据库；
对采集到的数据进行预处理；搭建预测模型框架；
基于搭建的预测模型框架，确定预测模型结构，构建公交车到站时间预测模型；
利用构建得到的公交车到站时间预测模型进行公交车到站时间预测。

7. 一种计算机可读存储介质，存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行如下步骤：

采集智能公交运营基础数据并建立数据库；
对采集到的数据进行预处理；搭建预测模型框架；
基于搭建的预测模型框架，确定预测模型结构，构建公交车到站时间预测模型；
利用构建得到的公交车到站时间预测模型进行公交车到站时间预测。

8. 一种实施权利要求1~5任意一项所述共线线路数据的公交车到站时间预测方法的城市的公共交通体系。

一种共线线路AVL数据的公交车到站时间预测方法

技术领域

[0001] 本发明属于城市智能公共交通信息处理技术领域,尤其涉及一种共线线路AVL数据的公交车到站时间预测方法。

背景技术

[0002] 目前,交通问题已经随着社会和经济的发展逐渐成为大中型城市中普遍存在的问题,极大的影响着居民的出行效率,城市的交通问题是由道路交通条件的发展和日益增长的出行需求之间的矛盾造成的,主要体现在交通拥堵、事故频发、污染严重以及通勤距离和时间增加等方面。城市公共交通系统具有节能环保、高效集约、实惠经济等特点,能够为城市居民提供多样化的出行选择,还可以降低环境污染、节约土地资源、减少交通事故。因此,优先发展公共交通、改善公交服务水平、提高公交吸引力是解决上述交通问题的首选途径之一。

[0003] 许多城市的轨道交通尚未形成体系,而在轨道交通已建成网络的特大级城市中,由于其人口密度巨大,轨道交通也远不能满足城市居民日常出行需求,未来相当长的时间内,常规公交系统仍然是承担城市公共交通客流需求的主要交通方式。然而,常规公交系统的发展还面临诸多问题,如调度方式落后、运营效率低下、智能化系统化不足等。这些问题严重影响了公交客运的服务质量,降低了常规公交的竞争力,直接反映为各大城市较低的常规公交分担率。

[0004] 大量的研究表明,由于受到道路交通阻抗的外在影响,乘客敏感性转移,对于车内旅行时间的敏感性小于站点等待时间。准确的到站时间预测可以使乘客根据自身出行需求决定出行时间,以减少乘客在站点的等待时间。此外,精确的公交到站时间预测还是公交出行路径诱导、公交运行动态调度管理等的重要前提。但是,由于公交车运行过程中受到各种因素的影响,产生的干扰使得公交车到站时间预测的误差达到200%以上。这些因素主要包括以下三类:第一,国内城市路网结构先天不足,公交线路多集中规划于主次干道上,导致线路的重复系数较高,且同一个站点有多条线路的公交车停靠,公交车在站间运行和站点停靠时线路间的相互影响均较大。第二,较高的发车频率和线路重复系数导致的公交车辆行驶间隔不均匀,会使得历史数据时效性较差,造成单条线路到站时间预测结果出现较大偏差。第三,异质交通流、乘客出行需求时空分布不均匀、交叉口延误、交通事故等带来的干扰,都会导致公交行程时间出现较大波动,从而影响公交车到站时间的预测精度。

[0005] 通过上述分析,现有技术存在的问题及缺陷为:

[0006] 现有预测方法预测结果偏差大,预测精度不高。

[0007] 现有的公交到站时间预测方法仅考虑相同线路的运行信息,而实际运营中多条线路服务于同一通道,在预测中没有充分利用其它线路的运行信息;

[0008] 现有的预测模型仅考虑相同线路的运行信息,到站时间预测结果准确度低于多线路模型。

[0009] 解决以上问题及缺陷的难度为:

[0010] 在数据预处理流程中,所有存在缺失记录的班次全部删除,没有采取合适的算法进行数据补齐。因此,不能有效评估数据缺失情况对预测结果的影响。

[0011] 虽然本发明提出的基于共线线路的公交到站时间预测模型较之常用的单线预测模型来说精度有所提高,但是,公交车辆在运营过程中会受到很多因素的干扰,如道路条件、交通状况、客流强度等等,这也导致公交到站时间预测的精度难以进一步提升。

[0012] 解决以上问题及缺陷的意义为:

[0013] (1) 本发明提出一种多条线路共线运行情况下的公交到站时间预测方法,从公共交通的角度,利用共线线路的公交运营数据作为输入变量进行公交到站时间的预测,以提高预测结果的精度和可靠性。

[0014] (2) 将该方法预测结果与典型的单条线路到站时间预测方法所得结果进行比较,横向对比验证该方法的有效性。此外,研究还对支持向量机模型、前馈神经网络模型的性能进行了评价和比较,为研究人员和实践者提供有价值的结论。

[0015] (3) 从乘客出行的角度,精确实时的公交到站时间预测有助于乘客在出行前安排出行的交通方式和路线、准确估计出行时间,合理分配出行,以减少乘客在站的候车时间、缓解焦虑情绪并降低站台拥挤程度。

发明内容

[0016] 针对现有技术存在的问题,本发明提供了一种共线线路数据的公交车到站时间预测方法。

[0017] 本发明是这样实现的,一种共线线路数据的公交车到站时间预测方法,包括:

[0018] 步骤一,采集智能公交运营基础数据并建立数据库;

[0019] 步骤二,对采集到的数据进行预处理;搭建预测模型框架;

[0020] 步骤三,基于搭建的预测模型框架,确定预测模型结构,构建公交车到站时间预测模型;

[0021] 步骤四,利用构建得到的公交车到站时间预测模型进行公交车到站时间预测。

[0022] 进一步,步骤一中,主要是把采集到的数据统一导入到数据库中,然后进行下面的数据预处理;“建立数据库”是将数据导入数据库,得到含有基础数据的一个数据库。

[0023] 所述公交运营基础数据包括:公交车自动定位数据即AVL数据或者GPS数据和公交线路站点地理信息数据即GIS数据。

[0024] 进一步,步骤二中,所述预处理方法包括:

[0025] (1) AVL数据与GIS数据关联:利用AVL数据、GIS数据共有的字段ROUTEID和STATIONNUM进行关联,给AVL数据匹配每条线路的站点名称、站点序号、运营方向信息;

[0026] (2) 公交运行班次划分:

[0027] (2.1) 将原始数据按日期、线路号、车辆编号和记录时间进行排序,依次比较下一行数据和本行数据的日期、线路号、车辆编号和运行方向字段,当其中任一字段的值发生改变,则运行班次-temp字段值增加1;

[0028] (2.2) 基于第一次班次划分的结果再划分一次运营班次,根据运营班次-temp字段和站点序号对数据进行排序,比较下一行数据和本行数据的运营班次-temp字段和站点序号字段,当运营班次-temp字段发生改变或下一行数据的站点序号小于本行数据,则运营班

次字段值增加1;将得到的运营班次字段将作为最终的公交运营班次值存储于原数据中;

[0029] (3) AVL数据处理:

[0030] (3.1) 重复数据处理:同一班次到离站数据在首末站的重复数据,保留时间最晚的一条数据作为首站点的出站记录,保留时间最早的一条数据作为末站点的到站记录;同一班次到离站数据在中间站点的重复数据,保留同一站点中时间最早和时间最晚的两条数据,分别作为本站的到站和离站记录;

[0031] (3.2) 异常数据处理:根据异常数据的特点,以运行班次为单位,按进站时间的先后对数据排序,若出现站点次序倒置或站点编号顺序错乱的数据,则标记并删除该数据;

[0032] (3.3) 缺失数据处理:提取共线运营的不同线路经过的相同站点的到离站数据,再根据各班次总的到离站记录的数量和共线运营的不同线路经过的相同站点的数量来判定该班次数据是否存在缺失;当各班次总的到站或离站记录数小于选取的公交通道站点数,标记并删除该班次的到离站数据。

[0033] 进一步,步骤二中,所述预测模型框架搭建方法包括:通过模型假设、参数设置、车头时距变量处理和行程时间变量处理进行输入预测模型构建;

[0034] 1) 确定预测目标:

[0035] 所述预测目标为基于线路 j 和 j' ($j \neq j' \in J$) 的车辆运行信息预测线路 j ($j \in J$) 的第 i 辆公交在站点 $k+\mu$ 处的到站时间,即预测线路 j ($j \in J$) 的第 i 辆公交在站点 k 和站点 $k+\mu$ 之间的行程时间;公式如下:

$$[0036] \quad t_{a,k+\mu}^{i,j} = t_{a,k}^{i,j} + \hat{T}_{k,k+\mu}^i;$$

[0037] 其中, $t_{a,k+\mu}^{i,j}$ 表示第 i 辆公交在站点 $k+\mu$ 处的到达时间; $t_{a,k}^{i,j}$ 表示第 i 辆公交在站点 k 处的到达时间; $\hat{T}_{k,k+\mu}^i$ 表示第 i 辆公交在站点 k 和站点 $k+\mu$ 之间的行程时间预测值;

[0038] 2) 进行模型假设;

[0039] 所述模型假设包括:

[0040] a) 在公交通道中,超车行为不允许发生;

[0041] b) 在不同线路的所有车辆中,驾驶员状况均相同;

[0042] c) 公交车辆之间及车辆和调度中心之间能够实现实时信息的交互;

[0043] 3) 确定车头时距变量:

[0044] 选择路段起点处的车头时距作为预测模型的输入变量之一,当目标车辆 i 到达站点 k 时,已有 n 辆公交经过了站点 $k+\mu$, p 为仍在公交通道上行驶的车辆数,则两类车头时距变量的计算公式如下:

$$[0045] \quad H_k^{i,j} = t_{a,k}^{i,j} - t_{a,k}^{i-p-1,j}$$

$$[0046] \quad H_k^{i,j} = t_{a,k}^{i,j} - t_{a,k}^{i-p-n,j}$$

[0047] 式中: $H_k^{i,j}$ 表示目标车辆 i 和前车 $i-p-1$ 在站点 k 处的联合车头时距,其中前车 $i-p-1$ 属于线路集合 J 中的任意线路; $H_k^{i,j}$ 表示目标车辆 i 和前车 $i-p-n$ 在站点 k 处的车头时距,其中前车 $i-p-n$ 和目标车辆 i 为同一线路; $t_{a,k}^{i,j}$ 为目标车辆 i 在站点 k 处的到达时间; $t_{a,k}^{i-p-1,j}$ 为公交

$i-p-1$ 在站点 k 处的到达时间; $t_{a,k}^{i-p-n,j}$ 为公交 $i-p-n$ 在站点 k 处的到达时间;

[0048] 4) 确定行程时间变量:

[0049] 利用车头时距的倒数对已经通过的前几辆车的行程时间进行加权,加权平均行程时间的计算公式如下:

$$[0050] \quad \bar{T}_{k,k+\mu}^i = \sum_{m=1}^{\sigma} \frac{1}{t_{a,k}^{i,j} - t_{a,k}^{i-p-m,j}} \times T_{k,k+\mu}^{i-p-m,j}$$

$$[0051] \quad \Gamma = \sum_{m=1}^{\sigma} \frac{1}{t_{a,k}^{i,j} - t_{a,k}^{i-p-m,j}}$$

$$[0052] \quad T_{k,k+\mu}^{i-p-m,j} = t_{a,k+\mu}^{i-p-m,j} - t_{a,k}^{i-p-m,j}, m = 1, \dots, \sigma$$

[0053] 式中: $\bar{T}_{k,k+\mu}^i$ 表示线路集合 J 中任意线路前 σ 辆车的加权平均行程时间; Γ 表示权重和; σ 为加权时选取的前车数量,该参数在模型预测阶段需先优化; $T_{k,k+\mu}^{i-p-m,j}$ 表示公交 $i-p-m$ 在站点 k 和站点 $k+\mu$ 之间的行程时间; $t_{a,k+\mu}^{i-p-m,j}$ 表示公交 $i-p-m$ 在站点 $k+\mu$ 处的到达时间; $t_{a,k}^{i-p-m,j}$ 表示公交 $i-p-m$ 在站点 k 处的到达时间;

[0054] 5) 选取和目标车辆线路相同的前车的行程时间作为预测模型的输入变量,计算公式如下:

$$[0055] \quad T_{k,k+\mu}^{i-p-n,j} = t_{a,k+\mu}^{i-p-n,j} - t_{a,k}^{i-p-n,j};$$

[0056] 式中: $T_{k,k+\mu}^{i-p-n,j}$ 表示与公交 i 线路号相同的车辆 $i-p-n$ 在站点 k 和站点 $k+\mu$ 之间的行程时间; $t_{a,k+\mu}^{i-p-n,j}$ 和 $t_{a,k}^{i-p-n,j}$ 分别表示公交 $i-p-n$ 在站点 $k+\mu$ 和站点 k 处的到达时间;

[0057] 6) 得到基于共线线路的公交行程时间预测模型为:

$$[0058] \quad \hat{T}_{k,k+\mu}^i = f(H_k^{i,j}, H_k^{i,j}, \bar{T}_{k,k+\mu}^i, T_{k,k+\mu}^{i-p-n,j});$$

[0059] 式中: $\hat{T}_{k,k+\mu}^i$ 表示公交 i 在站点 k 和站点 $k+\mu$ 之间的行程时间预测值; $f(\cdot)$ 表示输入变量 $H_k^{i,j}$ 、 $H_k^{i,j}$ 、 $\bar{T}_{k,k+\mu}^i$ 、 $T_{k,k+\mu}^{i-p-n,j}$ 和输出变量 $\hat{T}_{k,k+\mu}^i$ 之间待训练的函数关系。

[0060] 进一步,步骤三中,所述公交车到站时间预测模型构建方法包括:

[0061] 首先,采用支持向量回归模型进行公交到站时间预测。其中,选取径向基函数作为所用支持向量机模型的核函数,采用网格搜索的方法获取模型参数最优值。

[0062] 再者,采用神经网络模型进行公交到站时间预测。本发明选取标准的三层前馈神经网络用于公交到站时间的预测,考虑到算法参数数量和收敛速度,量化共轭梯度算法被用作模型的训练算法。针对不同的案例样本,预测模型需要选取不同的参数取值才能使预测结果最优,本发明对预测模型进行敏感性分析以获取参数的最优取值。

[0063] 最后,选取平均绝对误差、平均绝对百分误差和均方根误差三个指标来评价预测模型的优劣程度;评价指标的计算公式如下:

$$[0064] \quad MAE = \frac{\sum_{i=1}^N |T_{k,k+\mu}^i - \hat{T}_{k,k+\mu}^i|}{N}$$

$$[0065] \quad MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|T_{k,k+\mu}^i - \hat{T}_{k,k+\mu}^i|}{T_{k,k+\mu}^i} \times 100\%$$

$$[0066] \quad RMSE = \sqrt{\frac{\sum_{i=1}^N (T_{k,k+\mu}^i - \hat{T}_{k,k+\mu}^i)^2}{N}}$$

[0067] 式中： $T_{k,k+\mu}^i$ 表示公交*i*在站点*k*和站点*k+μ*之间的实际行程时间；*N*为观测到的预测样本数量。

[0068] 本发明的另一目的在于提供一种计算机设备，所述计算机设备包括存储器和处理器，所述存储器存储有计算机程序，所述计算机程序被所述处理器执行时，使得所述处理器执行如下步骤：

[0069] 采集智能公交运营基础数据并建立数据库；

[0070] 对采集到的数据进行预处理；搭建预测模型框架；

[0071] 基于搭建的预测模型框架，确定预测模型结构，构建公交车到站时间预测模型；

[0072] 利用构建得到的公交车到站时间预测模型进行公交车到站时间预测。

[0073] 本发明的另一目的在于提供一种计算机可读存储介质，存储有计算机程序，所述计算机程序被处理器执行时，使得所述处理器执行如下步骤：

[0074] 采集智能公交运营基础数据并建立数据库；

[0075] 对采集到的数据进行预处理；搭建预测模型框架；

[0076] 基于搭建的预测模型框架，确定预测模型结构，构建公交车到站时间预测模型；

[0077] 利用构建得到的公交车到站时间预测模型进行公交车到站时间预测。

[0078] 本发明的另一目的在于提供一种实施所述共线线路数据的公交车到站时间预测方法的城市的公共交通体系。

[0079] 结合上述的所有技术方案，本发明所具备的优点及积极效果为：本发明提出了一种基于共线线路运营数据的到站时间预测方法，通过引入其他线路的运营信息，以细化车辆间的车头时距间隔，提高历史数据的时效性和可靠性，同时也能将线路车辆间的相互影响纳入考虑，减小了干扰因素对公交到站时间预测带来的影响。

[0080] 本发明基于公交AVL数据提出了一套完整的数据预处理流程，可操作性强可直接应用于其他城市AVL数据处理。本发明提出一种多条线路共线运行情况下的公交到站时间预测方法，从公共交通的角度，利用共线线路的公交运营数据作为输入变量进行公交到站时间的预测，以提高预测结果的精度和可靠性。

[0081] 本发明的预测结果与典型的单条线路到站时间预测方法所得结果进行比较，横向对比验证该方法的有效性。此外，研究还对支持向量机模型、前馈神经网络模型的性能进行了评价和比较，为研究人员和实践者提供有价值的结论。

[0082] 对比的技术效果或者实验效果有：

[0083] 基于支持向量机的多线模型和单线模型预测结果对比

通道	运行方向	模型	早高峰			晚高峰		
			MAE(s)	MAPE (%)	RMSE (s)	MAE(s)	MAPE (%)	RMSE (s)
[0084] 人民南路	南向	多线模型	120.31	14.28%	172.9	130.96	14.04%	173.94
		单线模型	134.69	15.61%	183.8	137.07	14.61%	179.09
	北向	多线模型	55.59	7.93%	74.55	110.26	11.96%	132.36
		单线模型	67.81	9.48%	85.53	117.42	12.75%	141.41
蜀都大道	南向	多线模型	47.52	13.76%	62.6	60.98	16.77%	73.13
		单线模型	52.02	14.78%	68.42	62.4	17.04%	74.55
	北向	多线模型	42.09	11.94%	54.46	61.32	15.85%	81.56
		单线模型	45.57	13.00%	57.15	65.07	16.86%	84.3

[0085] 基于神经网络的多线模型和单线模型预测结果对比

通道	运行方向	模型	早高峰			晚高峰		
			MAE(s)	MAPE (%)	RMSE (s)	MAE(s)	MAPE (%)	RMSE (s)
[0086] 人民南路	南向	多线模型	101.7	11.89%	147.69	130.82	13.78%	170.06
		单线模型	103.48	12.31%	148.94	141.47	14.92%	186.75
	北向	多线模型	56.5	8.05%	76.19	116.45	12.53%	143.83
		单线模型	87.34	12.26%	120.19	120.29	12.94%	145.64
蜀都大道	南向	多线模型	54.32	15.73%	70.72	60.29	16.94%	76.32
		单线模型	59.97	17.06%	77.46	69.04	18.61%	81.6
	北向	多线模型	43.86	12.61%	57.41	61.34	15.76%	83.46
		单线模型	47.51	13.48%	56.9	65.69	16.73%	87.28

[0087] 这两个表分别是利用支持向量机和神经网络两种方法进行到站预测的结果；

[0088] 其中每个表中均有两种模型进行对比(多线、单线模型的不同之处在于输入变量的不同),此处多线模型即为本文的改进之处,以往的预测模型输入变量中仅考虑了相同线路的信息(即单线模型),通过评价指标可以得出,相对于单线模型来说多线模型的预测准确度更高,预测误差小。

[0089] 本发明两种方法(支持向量机、神经网络)的结果均表明多线模型的预测效果优于单线模型。

[0090] 本发明提出了基于共线线路AVL数据的公交车到站时间预测方法的一整套流程,包括数据预处理、模型构建、实例验证;而现有技术中着重强调模型的构建过程这一个方面,即主要讨论在已有预测方法中加入其他共线线路信息能够提高预测准确度。

[0091] 本发明同时利用支持向量机和机器学习两种算法进行多线模型的验证(现有技术已公开发表文献仅采取支持向量机一种算法),结果表明两种算法的多线模型预测结果均优于单线模型,这在一定程度上可以证明该模型具有一定的普遍适用性,即考虑其他共线线路信息的多线模型预测结果优于单线模型。

附图说明

[0092] 为了更清楚地说明本申请实施例的技术方案,下面将对本申请实施例中所需要使用的附图做简单的介绍,显而易见地,下面所描述的附图仅仅是本申请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下还可以根据这些附图获得其他的附图。

- [0093] 图1是本发明实施例提供的共线线路AVL数据的公交车到站时间预测方法流程图。
- [0094] 图2是本发明实施例提供的线路共线示意图。
- [0095] 图3是本发明实施例提供的第一次划分运行班次流程图。
- [0096] 图4是本发明实施例提供的第二次划分运行班次流程图。
- [0097] 图5是本发明实施例提供的基于共线线路的公交到站时间预测模型框架示意图。
- [0098] 图6是本发明实施例提供的公交通道内多条线路车辆的时空轨迹示意图。
- [0099] 图7是本发明实施例提供的多线路到站序列示意图。
- [0100] 图8是本发明实施例提供的支持向量机模型结构示意图。
- [0101] 图9是本发明实施例提供的M-P模型结构示意图。
- [0102] 图10是本发明实施例提供的具有单层隐藏层的感知神经网络模型结构示意图。

具体实施方式

[0103] 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。

[0104] 针对现有技术存在的问题,本发明提供了一种共线线路AVL数据的公交车到站时间预测方法,下面结合附图对本发明作详细的描述。

[0105] 如图1所示,本发明实施例提供的共线线路AVL数据的公交车到站时间预测方法包括以下步骤:

[0106] S101,采集智能公交运营基础数据并建立数据库;

[0107] S102,对采集到的数据进行预处理;搭建预测模型框架;

[0108] S103,基于搭建的预测模型框架,确定预测模型结构,构建公交车到站时间预测模型;

[0109] S104,利用构建得到的公交车到站时间预测模型进行公交车到站时间预测。

[0110] 步骤S101中,本发明实施例提供的公交运营基础数据包括:公交车自动定位数据即AVL数据或者GPS数据和公交线路站点地理信息数据即GIS数据。

[0111] 步骤S102中,本发明实施例提供的预处理方法包括:

[0112] (1) AVL数据与GIS数据关联:利用AVL数据、GIS数据共有的字段ROUTEID和STATIONNUM进行关联,给AVL数据匹配每条线路的站点名称、站点序号、运营方向信息;

[0113] (2) 公交运行班次划分:

[0114] (2.1) 将原始数据按日期、线路号、车辆编号和记录时间进行排序,依次比较下一行数据和本行数据的日期、线路号、车辆编号和运行方向字段,当其中任一字段的值发生改变,则运行班次-temp字段值增加1;

[0115] (2.2) 基于第一次班次划分的结果再划分一次运营班次,根据运营班次-temp字段和站点序号对数据进行排序,比较下一行数据和本行数据的运营班次-temp字段和站点序号字段,当运营班次-temp字段发生改变或下一行数据的站点序号小于本行数据,则运营班次字段值增加1;将得到的运营班次字段将作为最终的公交运营班次值存储于原数据中;

[0116] (3) AVL数据处理:

[0117] (3.1) 重复数据处理:同一班次到离站数据在首末站的重复数据,保留时间最晚的

一条数据作为首站点的出站记录,保留时间最早的一条数据作为末站点的到站记录;同一班次到离站数据在中间站点的重复数据,保留同一站点中时间最早和时间最晚的两条数据,分别作为本站的到站和离站记录;

[0118] (3.2) 异常数据处理:根据异常数据的特点,以运行班次为单位,按进站时间的先后对数据排序,若出现站点次序倒置或站点编号顺序错乱的数据,则标记并删除该数据;

[0119] (3.3) 缺失数据处理:提取共线运营的不同线路经过的相同站点的到离站数据,再根据各班次总的到离站记录的数量和共线运营的不同线路经过的相同站点的数量来判定该班次数据是否存在缺失;当各班次总的到站或离站记录数小于选取的公交通道站点数,标记并删除该班次的到离站数据。

[0120] 步骤S102中,本发明实施例提供的预测模型框架搭建方法包括:通过模型假设、参数设置、车头时距变量处理和行程时间变量处理进行输入预测模型构建;

[0121] 1) 确定预测目标:

[0122] 所述预测目标为基于线路j和j' ($J \neq j' \in J$) 的车辆运行信息预测线路j ($j \in J$) 的第i辆公交在站点k+ μ 处的到站时间,即预测线路j ($j \in J$) 的第i辆公交在站点k和站点k+ μ 之间的行程时间;公式如下:

$$[0123] \quad t_{a,k+\mu}^{i,j} = t_{a,k}^{i,j} + \hat{T}_{k,k+\mu}^i;$$

[0124] 其中, $t_{a,k+\mu}^{i,j}$ 表示第i辆公交在站点k+ μ 处的到达时间; $t_{a,k}^{i,j}$ 表示第i辆公交在站点k处的到达时间; $\hat{T}_{k,k+\mu}^i$ 表示第i辆公交在站点k和站点k+ μ 之间的行程时间预测值;

[0125] 2) 进行模型假设;

[0126] 所述模型假设包括:

[0127] a) 在公交通道中,超车行为不允许发生;

[0128] b) 在不同线路的所有车辆中,驾驶员状况均相同;

[0129] c) 公交车辆之间及车辆和调度中心之间能够实现实时信息的交互;

[0130] 3) 确定车头时距变量:

[0131] 选择路段起点处的车头时距作为预测模型的输入变量之一,当目标车辆i到达站点k时,已有n辆公交经过了站点k+ μ ,p为仍在公交通道上行驶的车辆数,则两类车头时距变量的计算公式如下:

$$[0132] \quad H_k^{i,J} = t_{a,k}^{i,j} - t_{a,k}^{i-p-1,J}$$

$$[0133] \quad H_k^{i,j} = t_{a,k}^{i,j} - t_{a,k}^{i-p-n,j}$$

[0134] 式中: $H_k^{i,J}$ 表示目标车辆i和前车i-p-1在站点k处的联合车头时距,其中前车i-p-1属于线路集合J中的任意线路; $H_k^{i,j}$ 表示目标车辆i和前车i-p-n在站点k处的车头时距,其中前车i-p-n和目标车辆i为同一线路; $t_{a,k}^{i,j}$ 为目标车辆i在站点k处的到达时间; $t_{a,k}^{i-p-1,J}$ 为公交i-p-1在站点k处的到达时间; $t_{a,k}^{i-p-n,j}$ 为公交i-p-n在站点k处的到达时间;

[0135] 4) 确定行程时间变量:

[0136] 利用车头时距的倒数对已经通过的前几辆车的行程时间进行加权,加权平均行程时间的计算公式如下:

$$[0137] \quad \bar{T}_{k,k+\mu}^i = \sum_{m=1}^{\sigma} \frac{1/t_{a,k}^{i,j} - t_{a,k}^{i-p-m,j}}{\Gamma} \times T_{k,k+\mu}^{i-p-m,j}$$

$$[0138] \quad \Gamma = \sum_{m=1}^{\sigma} \frac{1}{t_{a,k}^{i,j} - t_{a,k}^{i-p-m,j}}$$

$$[0139] \quad T_{k,k+\mu}^{i-p-m,j} = t_{a,k+\mu}^{i-p-m,j} - t_{a,k}^{i-p-m,j}, m=1, \dots, \sigma$$

[0140] 式中： $\bar{T}_{k,k+\mu}^i$ 表示线路集合J中任意线路前 σ 辆车的加权平均行程时间； Γ 表示权重和； σ 为加权时选取的前车数量，该参数在模型预测阶段需先优化； $T_{k,k+\mu}^{i-p-m,j}$ 表示公交*i-p-m*在站点k和站点k+ μ 之间的行程时间； $t_{a,k+\mu}^{i-p-m,j}$ 表示公交*i-p-m*在站点k+ μ 处的到达时间； $t_{a,k}^{i-p-m,j}$ 表示公交*i-p-m*在站点k处的到达时间；

[0141] 5) 选取和目标车辆线路相同的前车的行程时间作为预测模型的输入变量，计算公式如下：

$$[0142] \quad T_{k,k+\mu}^{i-p-n,j} = t_{a,k+\mu}^{i-p-n,j} - t_{a,k}^{i-p-n,j} ;$$

[0143] 式中： $T_{k,k+\mu}^{i-p-n,j}$ 表示与公交*i*线路号相同的车辆*i-p-n*在站点k和站点k+ μ 之间的行程时间； $t_{a,k+\mu}^{i-p-n,j}$ 和 $t_{a,k}^{i-p-n,j}$ 分别表示公交*i-p-n*在站点k+ μ 和站点k处的到达时间；

[0144] 6) 得到基于共线线路的公交行程时间预测模型为：

$$[0145] \quad \hat{T}_{k,k+\mu}^i = f(H_k^{i,j}, H_k^{i,j}, \bar{T}_{k,k+\mu}^i, T_{k,k+\mu}^{i-p-n,j}) ;$$

[0146] 式中： $\hat{T}_{k,k+\mu}^i$ 表示公交*i*在站点k和站点k+ μ 之间的行程时间预测值； $f(\cdot)$ 表示输入变量 $H_k^{i,j}$ 、 $H_k^{i,j}$ 、 $\bar{T}_{k,k+\mu}^i$ 、 $T_{k,k+\mu}^{i-p-n,j}$ 和输出变量 $\hat{T}_{k,k+\mu}^i$ 之间待训练的函数关系。

[0147] 步骤S103中，本发明实施例提供的公交车到站时间预测模型构建方法包括：

[0148] 首先，采用支持向量回归模型进行公交到站时间预测。其中，选取径向基函数作为所用支持向量机模型的核函数，本发明采用网格搜索的方法获取模型参数的最优值。

[0149] 再者，采用神经网络模型进行公交到站时间预测。本发明选取标准的三层前馈神经网络用于公交到站时间的预测，考虑到算法参数数量和收敛速度，量化共轭梯度算法被用作模型的训练算法。针对不同的案例样本，预测模型需要选取不同的参数取值才能使预测结果最优，本发明对预测模型进行敏感性分析以获取参数的最优取值。

[0150] 最后，选取平均绝对误差、平均绝对百分误差和均方根误差三个指标来评价预测模型的优劣程度；评价指标的计算公式如下：

$$[0151] \quad MAE = \frac{\sum_{i=1}^N |T_{k,k+\mu}^i - \hat{T}_{k,k+\mu}^i|}{N}$$

$$[0152] \quad MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|T_{k,k+\mu}^i - \hat{T}_{k,k+\mu}^i|}{T_{k,k+\mu}^i} \times 100\%$$

$$[0153] \quad RMSE = \sqrt{\frac{\sum_{i=1}^N (T_{k,k+\mu}^i - \hat{T}_{k,k+\mu}^i)^2}{N}}$$

[0154] 式中： $T_{k,k+\mu}^i$ 表示公交*i*在站点*k*和站点*k+μ*之间的实际行程时间；*N*为观测到的预测样本数量。

[0155] 本发明涉及的模型如下：

[0156] (1) 支持向量机

[0157] 支持向量机是一类按监督学习方式对数据进行二元分类的广义线性分类器，分类的方式是找到一个最优的超平面把两个类别分开。作为常见的核学习方法之一，SVM在解决复杂的小样本非线性问题和识别高维模式方面的应用较为广泛。该算法具有两个关键特性：其一是支持向量机对过拟合问题具有较强的抵抗能力，在解决各种时间序列预测问题时具有较好的泛化性能，这是因为SVM算法是基于结构风险最小化原理，能够使泛化误差的上界最小；其二是SVM算法的解是唯一且全局最优的，因SVM模型的训练等价于求解线性约束二次规划问题。

[0158] 将SVM由分类问题推广至回归问题即为支持向量回归，分类问题中的超平面决策边界就是回归问题的模型，其一般形式为：

$$[0159] \quad f(x) = \omega \cdot \varphi(x) + b$$

[0160] 式中 $\varphi(x)$ 为将输入向量*x*映射至高维特征空间的非线性转换函数。系数 ω 和*b*是由最小化正则风险函数估计得到，该风险函数的一般形式如下：

$$[0161] \quad \frac{1}{2} \|\omega\|^2 + C \frac{1}{l} \sum_{i=1}^l L_{\varepsilon}(y_i, f(x_i))$$

[0162] 式中 $\|\omega\|^2$ 为正则项，可用来控制函数容量。第二项为经验误差，由 ε -不敏感损失函数衡量，该损失函数的一般形式如下：

$$[0163] \quad L_{\varepsilon}(y_i, f(x_i)) = \begin{cases} |y_i - f(x_i)| - \varepsilon, & |y_i - f(x_i)| \geq \varepsilon \\ 0, & \text{otherwise} \end{cases}$$

[0164] 该函数定义了一个 ε 区间，若预测值在该区间内，则损失为0，反之损失则为预测值和 ε 之间的差值。*C*为正则化常数，*C*值越大，经验风险相对于正则化项越重要。*C*和 ε 均为使用者自定义的参数。

[0165] 最小化是一个标准的最优化问题，该问题可通过引入拉格朗日函数解决。定义如下的拉格朗日函数：

$$[0166] \quad L = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) - \sum_{i=1}^l (\eta_i \xi_i + \eta_i^* \xi_i^*) - \sum_{i=1}^l \alpha_i (\varepsilon + \xi_i - y_i + \omega \cdot \varphi(x) + b) - \sum_{i=1}^l \alpha_i^* (\varepsilon + \xi_i^* - y_i + \omega \cdot \varphi(x) + b)$$

[0167] 根据上式分别对 ω 、*b*、 ξ_i^* 求偏微分并令它们等于0，得到方程后将其代入下式，得

到如下凸二次规划的对偶问题：

$$\begin{aligned}
 \max W(\alpha_i, \alpha_i^*) &= \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) - \\
 [0168] \quad & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) (\varphi(x_i) \cdot \varphi(x_j)) \\
 \text{s.t.} \quad & \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases}
 \end{aligned}$$

[0169] 根据原始条件权重向量等于训练数据的线性组合可得：

$$[0170] \quad f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \varphi(x_i) \cdot \varphi(x) + b$$

[0171] 引入核函数 $K(x_i, y_i)$ ，代入式(3-14)可得：

$$[0172] \quad f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b$$

[0173] 求解该函数用以得到最优超平面，即支持向量回归模型。其中，函数 $K(x_i, y_i)$ 的值等于向量 x_i 和 y_i 在高维特征空间的映射 $\varphi(x_i)$ 和 $\varphi(y_i)$ 的内积，即：

$$[0174] \quad K(x_i, y_i) = \varphi(x_i) \cdot \varphi(y_i)$$

[0175] 常见的核函数主要有线性核函数、多项式核函数和径向基函数(RBF)核函数等，考虑到核函数的参数数量，本发明选取径向基函数作为所用支持向量机模型的核函数。因此，基于支持向量机的多条线路共线运营下的公交到站时间预测模型结构如图8所示。

[0176] (2) 神经网络

[0177] 神经网络是受生物学中神经网络的启发，模拟人脑的神经系统对复杂信息的处理机制的一种模糊计算系统，具有大规模并行、分布式处理、自组织、自学习等优点。该系统是建立在一组被称为人工神经元的连接单元或节点上的，这些单元或节点可以向其他神经元传递信号，神经元接收来自其它神经元传递过来的输入信号，这些信号的表达，通常通过神经元之间连接的权重(weight)大小来表示，权重决定了一个节点对另一个节点的影响强度。每个神经元的输出由输入之和的非线性函数来计算，并与当前神经元的阈值进行比较，只有当输出超过该阈值时才向外传送。已有研究提出了较多的神经元模型，其中提出最早且影响最大的是M-P模型，它是前馈神经网络模型的基础。M-P模型结构示意图如图9所示，其输出的通用表达式如下式所示。

$$[0178] \quad y = f\left(\sum_{i=1}^n w_i y_i - \theta\right)$$

[0179] 其中 w_i 为来自第 i 个神经元的连接权重。 x_i 为来自第 i 个神经元的输入。 θ 为当前神经元单元的阈值。 n 为输入信号数目。 $f(\cdot)$ 为输出变换函数，通常采用S形函数。

[0180] 神经网络模型由神经元单元聚成的多层构成，其基本的拓扑结构包含接收外部数据的输入层和产生最终结果的输出层，在它们之间是零或更多的隐藏层。模型中不同的层可以对其输入执行不同的转换，而一层的神经元只与紧接前一层和紧接后一层的神经元

相连。根据网络连接的拓扑结构,神经网络模型可以分为反馈网络和前向网络。反馈网络可以用一个无向的完备图表示,该网络允许同一层或前一层神经元之间相互连接,且网络内的神经元单元之间有反馈。前向网络可以用一个有向无环路图表示,该网络中信号从第一层(输入层)传递到最后一层(输出层),中间遍历各层,但网络中不存在反馈。参照前人的研究,本发明选取的神经网络模型为前向网络中具有三层结构的感知机网络模型,该模型除输入层和输出层外,还含有单层隐藏层,其结构示意图如图10所示。

[0181] 神经网络模型常用的训练算法有BP (Backpropagation) 算法、共轭梯度 (Conjugate Gradient) 算法、LM (Levenberg Marquardt) 算法等,考虑到算法参数数量和收敛速度,本发明选取量化共轭梯度 (Scaled Conjugate Gradient) 算法作为模型的训练算法。该算法基于共轭梯度法,算法本身不需要参数,且具有较快的收敛速度。

[0182] (3) 模型评价。

[0183] 预测公交行程时间的算法较多,但预测算法的好坏必须要有一个统一的衡量标准来判断,由于算法的预测值与真实值之间存在误差,因此一个算法的优劣很大程度上取决于预测值与真实值之间的误差大小。本发明选取了平均绝对误差 (MAE)、平均绝对百分误差 (MAPE) 和均方根误差 (RMSE) 三个指标来评估预测方法的优劣程度。其中,平均绝对误差反映了预测值和真实值之间差值的绝对大小,而平均绝对百分误差则反映了误差相对于真实值的大小,这两个指标的值越小,模型的预测精度越高。均方根误差则表明了误差的鲁棒性,该指标的值越小,模型预测的波动性越小,模型越稳健。这三个评价指标的计算公式如下:

$$[0184] \quad MAE = \frac{\sum_{i=1}^N |T_{k,k+\mu}^i - \hat{T}_{k,k+\mu}^i|}{N}$$

$$[0185] \quad MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|T_{k,k+\mu}^i - \hat{T}_{k,k+\mu}^i|}{T_{k,k+\mu}^i} \times 100\%$$

$$[0186] \quad RMSE = \sqrt{\frac{\sum_{i=1}^N (T_{k,k+\mu}^i - \hat{T}_{k,k+\mu}^i)^2}{N}}$$

[0187] 式中: $T_{k,k+\mu}^i$ 表示公交 i 在站点 k 和站点 $k+\mu$ 之间的实际行程时间; N 为观测到的预测样本数量。

[0188] (6) 公交到站时间预测模型实例分析。本步骤主要是选择合适的多线路公交通道数据,根据步骤(2)进行数据处理和步骤(4)进行输入变量的处理和计算和模型参数标定以及步骤(5)进行模型评价。

[0189] 在本发明中,确定公交车到站时间预测模型采用支持向量回归模型,其中,函数 $K(x_i, y_i)$ 的值等于向量 x_i 和 y_i 在高维特征空间的映射 $\varphi(x_i)$ 和 $\varphi(y_i)$ 的内积,即:

$$[0190] \quad K(x_i, y_i) = \varphi(x_i) \cdot \varphi(y_i) \text{。}$$

[0191] 下面结合具体实施例对本发明的技术效果作进一步描述。

[0192] 实施例1:

- [0193] 基于共线线路AVL数据的公交到站时间预测方法,该方法主要分为如下几个步骤:
- [0194] (1) 采集智能公交运营基础数据并建立数据库。该部分数据包括公交车自动定位数据(简称:AVL数据或者GPS数据)和公交线路站点地理信息数据(GIS数据);
- [0195] (2) 根据步骤(1)采集到的数据进行数据预处理流程。其主要包括AVL数据与GIS数据关联,公交运行班次划分,AVL数据处理三个部分;
- [0196] (2.1) AVL数据与GIS数据关联。利用AVL数据、GIS数据共有的字段ROUTEID和STATIONNUM进行关联,给AVL数据匹配每条线路的站点名称、站点序号、运营方向信息;
- [0197] (2.2) 公交运行班次划分。运行班次的划分主要分为以下两步:首先,将原始数据按日期、线路号、车辆编号和记录时间进行排序,依次比较下一行数据和本行数据的日期、线路号、车辆编号和运行方向字段,当其中任一字段的值发生改变,则运行班次-temp字段值增加1。
- [0198] 第一次运行班次的划分完成之后,由于部分班次的首站数据缺失或数据异常现象的存在,会导致部分已划分的班次里实际包含多个班次的数据,因而需要再基于第一次班次划分的结果再划分一次运营班次。根据运营班次-temp字段和站点序号对数据进行排序,比较下一行数据和本行数据的运营班次-temp字段和站点序号字段,当运营班次-temp字段发生改变或下一行数据的站点序号小于本行数据,则运营班次字段值增加1。第二次运行班次划分的流程图如图4所示,此时得到的运营班次字段将作为最终的公交运营班次值存储于原数据中。
- [0199] (2.3) AVL数据处理。在步骤(2.2)运行班次划分之后进行数据处理,主要包括重复数据处理、异常数据处理和缺失数据处理。
- [0200] (2.3.1) 重复数据处理
- [0201] 针对同一班次到离站数据在首末站重复的问题,保留时间最晚的一条数据作为首站点的出站记录,保留时间最早的一条数据作为末站点的到站记录;而对于同一班次到离站数据在中间站点重复的问题,保留同一站点中时间最早和时间最晚的两条数据,分别作为本站的到站和离站记录。
- [0202] (2.3.2) 异常数据处理
- [0203] 根据异常数据的特点,以运行班次为单位,按进站时间的先后对数据排序,若出现站点次序倒置或站点编号顺序错乱的数据,则标记并删除该数据。
- [0204] (2.3.3) 缺失数据处理
- [0205] 首先提取共线运营的不同线路经过的相同站点的到离站数据,再根据各班次总的到离站记录的数量和共线运营的不同线路经过的相同站点的数量来判定该班次数据是否存在缺失。考虑到补齐缺失数据本身存在一定误差,且会对后续的公交到站时间预测结果带来影响,因此当各班次总的到站或离站记录数小于选取的公交通道站点数,则标记并删除该班次的到离站数据。
- [0206] (3) 预测模型框架构建。
- [0207] 基于多条线路共线运营情况下的公交到站时间预测可以这样描述:给定上游某一站点处任意一条公交线路目标车辆的到达时间,考虑共线线路的运营信息,通过预测目标车辆站点间的行程时间来实现其在下流某一站点处到达时间的预测。图5为本技术方案提出的公交到站时间预测模型的逻辑框架,该模型的架构包括输入变量处理、模型建立和性

能评价三个阶段。

[0208] (4) 输入变量处理。该部分主要包括模型假设以及参数设置、车头时距变量处理和行程时间变量处理三个步骤。

[0209] 假设存在一条多条线路共线运行的公交通道,其线路集合为 J 。图6为 J 条线路共 $p+n$ 辆公交在公交通道内的时空轨迹示意图,其中不同线路车辆的运行轨迹以颜色区分。而从时空轨迹示意图中可知,当目标公交到达站点 k 时,已有 n 辆公交经过了站点 $k+\mu$, p 为仍在公交通道上行驶的车辆数。则本研究的目标为基于线路 j 和 j' ($j \neq j' \in J$) 的车辆运行信息预测线路 j ($j \in J$) 的第 i 辆公交在站点 $k+\mu$ 处的到站时间,即预测线路 j ($j \in J$) 的第 i 辆公交在站点 k 和站点 $k+\mu$ 之间的行程时间。

$$[0210] \quad t_{a,k+\mu}^{i,j} = t_{a,k}^{i,j} + \hat{T}_{k,k+\mu}^i \quad (4-1)$$

[0211] 式中: $t_{a,k+\mu}^{i,j}$ 表示第 i 辆公交在站点 $k+\mu$ 处的到达时间; $t_{a,k}^{i,j}$ 表示第 i 辆公交在站点 k 处的到达时间; $\hat{T}_{k,k+\mu}^i$ 表示第 i 辆公交在站点 k 和站点 $k+\mu$ 之间的行程时间预测值。

[0212] 在实际研究中,如图6所示的时空轨迹图可通过交通数据采集技术(如AVL)采集到的数据来构建。而从时空轨迹图中获取的诸如车辆间的车头时距和前车的站间行程时间等有效信息可用于目标车辆行程时间的预测。

[0213] (4.1) 模型假设以及变量解释

[0214] 在所提出的预测方法中,有如下假设:

[0215] a) 在公交通道中,超车行为不允许发生。

[0216] b) 在不同线路的所有车辆中,驾驶员状况均相同。

[0217] c) 公交车辆之间及车辆和调度中心之间能够实现实时信息的交互。

[0218] 该预测方法所需变量及变量含义如表4-1所示。

[0219] 表4-1变量及变量含义

[0220]

变量	描述
$t_{a,k+\mu}^{i,j}$	线路 j 的车辆 i 到达站点 $k+\mu$ 的时间
$H_k^{i,j}$	站点 k 处线路 j 的车辆 i 和线路集合 J 中任意线路的车辆 $i-p-1$ 的车头时距
$H_k^{i,j}$	站点 k 处线路 j 的车辆 i 与其同线路车辆 $i-p-n$ 的车头时距
$t_{a,k}^{i,j}$	线路 j 的车辆 i 到达站点 k 的时间
$t_{a,k}^{i-p-1,j}$	线路集合 J 中任意线路的车辆 $i-p-1$ 到达站点 k 的时间
$t_{a,k}^{i-p-n,j}$	线路 j 的车辆 $i-p-n$ 到达站点 k 的时间
$\bar{T}_{k,k+\mu}^i$	线路集合 J 中任意线路前 σ 辆车的加权平均行程时间
$T_{k,k+\mu}^{i-p-m,j}$	线路集合 J 中任意线路的车辆 $i-p-m$ 在站点 k 和站点 $k+\mu$ 之间的行程时间
$t_{a,k+\mu}^{i-p-m,j}$	线路集合 J 中任意线路的车辆 $i-p-m$ 到达站点 $k+\mu$ 的时间
$t_{a,k}^{i-p-m,j}$	线路集合 J 中任意线路的车辆 $i-p-m$ 到达站点 k 的时间
$T_{k,k+\mu}^{i-p-n,j}$	线路 j 的车辆 $i-p-n$ 在站点 k 和站点 $k+\mu$ 之间的行程时间

	$t_{a,k+\mu}^{i-p-n,j}$	线路 j 的车辆 $i-p-n$ 到达站点 $k+\mu$ 的时间
	$t_{a,k}^{i-p-n,j}$	线路 j 的车辆 $i-p-n$ 到达站点 k 的时间
[0221]	$\hat{T}_{k,k+\mu}^i$	车辆 i 在站点 k 和站点 $k+\mu$ 之间的行程时间预测值
	σ	加权时选取的前车数量
	Γ	权重和

[0222] (4.2) 车头时距变量

[0223] 在相同路段上运营的不同线路的公交车,其在路段起点处的车头时距越小,那么它们在行驶过程中所经历的外部道路交通环境状况就越相似,最后在到达路段终点处所耗费的行程时间也就越接近。此外,通常情况下,最新的交通数据也会为预测提供更加可靠的交通信息。因此,为了表明前车行程时间信息的及时性和可靠性程度,本发明选择路段起点处的车头时距作为预测模型的输入变量之一。图7为共线线路到达站点 k 时的序列示意图,其中当目标车辆 i 到达站点 k 时,已有 n 辆公交经过了站点 $k+\mu$, p 为仍在公交通道上行驶的车辆数。考虑到前车所属公交线路的不同,两类车头时距变量的计算公式如下:

[0224]
$$H_k^{i,j} = t_{a,k}^{i,j} - t_{a,k}^{i-p-1,j} \quad (4-2)$$

[0225]
$$H_k^{i,j} = t_{a,k}^{i,j} - t_{a,k}^{i-p-n,j} \quad (4-3)$$

[0226] 式中: $H_k^{i,j}$ 表示目标车辆 i 和前车 $i-p-1$ 在站点 k 处的联合车头时距,其中前车 $i-p-1$ 属于线路集合 J 中的任意线路; $H_k^{i,j}$ 表示目标车辆 i 和前车 $i-p-n$ 在站点 k 处的车头时距,其中前车 $i-p-n$ 和目标车辆 i 为同一线路; $t_{a,k}^{i,j}$ 为目标车辆 i 在站点 k 处的到达时间; $t_{a,k}^{i-p-1,j}$ 为公交 $i-p-1$ 在站点 k 处的到达时间; $t_{a,k}^{i-p-n,j}$ 为公交 $i-p-n$ 在站点 k 处的到达时间;

[0227] (4.3) 行程时间变量

[0228] 通过关于公交走廊上相邻前后两辆车的行程时间关联性的研究可知,无论相邻前后两辆车的线路号是否相同,相邻前后两辆车的行程时间均具有较高的相关性,这也证实了前车的行程时间能够用来反映当前的交通状况。因此,在公交行程时间预测中,前车行程时间常常作为预测模型的核心输入变量。此外,通过融入其它线路公交车的行程时间信息,可以细化目标线路公交车的车头时距间隔,更准确和实时的反映道路交通状况的变化。考虑到车头时距越小,目标车辆和前车运营环境相似的概率就会越高,因此,本发明利用车头时距的倒数对已经通过的前几辆车的行程时间进行加权,以提高前车行程时间信息的利用率。加权平均行程时间的计算公式如下:

[0229]
$$\bar{T}_{k,k+\mu}^i = \sum_{m=1}^{\sigma} \frac{1/t_{a,k}^{i,j} - t_{a,k}^{i-p-m,j}}{\Gamma} \times T_{k,k+\mu}^{i-p-m,j} \quad (4-4)$$

[0230]
$$\Gamma = \sum_{m=1}^{\sigma} \frac{1}{t_{a,k}^{i,j} - t_{a,k}^{i-p-m,j}} \quad (4-5)$$

$$[0231] \quad T_{k,k+\mu}^{i-p-m,J} = t_{a,k+\mu}^{i-p-m,J} - t_{a,k}^{i-p-m,j}, m = 1, \dots, \sigma \quad (4-6)$$

[0232] 式中： $\bar{T}_{k,k+\mu}^i$ 表示线路集合J中任意线路前 σ 辆车的加权平均行程时间； Γ 表示权重和； σ 为加权时选取的前车数量，该参数在模型预测阶段需先优化； $T_{k,k+\mu}^{i-p-m,J}$ 表示公交i-p-m在站点k和站点k+ μ 之间的行程时间； $t_{a,k+\mu}^{i-p-m,j}$ 表示公交i-p-m在站点k+ μ 处的到达时间； $t_{a,k}^{i-p-m,j}$ 表示公交i-p-m在站点k处的到达时间。

[0233] 虽然上述共线线路的公交车行程时间信息的融入对预测是必要的，但考虑到不同线路车辆在同一站点处的客流需求强度不同，不同线路在同一公交走廊上的停靠站数不同，因此，为了减小其对预测造成的影响，本发明同时选取了和目标车辆线路相同的前车的行程时间作为预测模型的输入变量，该变量的计算公式如下：

$$[0234] \quad T_{k,k+\mu}^{i-p-n,j} = t_{a,k+\mu}^{i-p-n,j} - t_{a,k}^{i-p-n,j} \quad (4-7)$$

[0235] 式中： $T_{k,k+\mu}^{i-p-n,j}$ 表示与公交i线路号相同的车辆i-p-n在站点k和站点k+ μ 之间的行程时间； $t_{a,k+\mu}^{i-p-n,j}$ 和 $t_{a,k}^{i-p-n,j}$ 分别表示公交i-p-n在站点k+ μ 和站点k处的到达时间。

[0236] 综上所述，基于共线线路的公交行程时间预测方法归纳如下：

$$[0237] \quad \hat{T}_{k,k+\mu}^i = f(H_k^{i,j}, H_{k,k+\mu}^{i,j}, \bar{T}_{k,k+\mu}^i, T_{k,k+\mu}^{i-p-n,j}) \quad (4-8)$$

[0238] 式中： $\hat{T}_{k,k+\mu}^i$ 表示公交i在站点k和站点k+ μ 之间的行程时间预测值； $f(\cdot)$ 表示输入变量 $H_k^{i,j}$ 、 $H_{k,k+\mu}^{i,j}$ 、 $\bar{T}_{k,k+\mu}^i$ 、 $T_{k,k+\mu}^{i-p-n,j}$ 和输出变量 $\hat{T}_{k,k+\mu}^i$ 之间待训练的函数关系。

[0239] 实施例2：

[0240] 利用成都市智能公交系统为例，以数据库为数据处理工具，利用常规公交5个工作日(2019.05.06-2019.05.10)的GPS数据、站点GIS数据来对本发明的具体实施方式进行说明。

[0241] (1) 数据采集。

[0242] 首先采集成都市公交的GPS数据和GIS数据，下表为采集到的数据样本。

[0243] GPS数据样本表

[0244]

字段	字段含义	数据样本
DATATYPE	数据类型	4
ROUTEID	线路编号	16
PRODUCTID	车辆编号	21038004
STATIONNUM	站点编号	30346
LONGITUDE	经度	104.064591
LATITUDE	纬度	30.619281
ACTDATETIME	定位点记录时间	2019-05-08 19:41:23
GPSSPEED	车辆瞬时速度	1.59
ISARRLFT	到离站标识	1
GPSMILE	运营累计里程	38714006

[0245] GIS数据样本表

[0246]	字段	字段含义	数据样本
	LINENO	线路编号	16

[0247]	NAME	站点名称	人南立交桥南站
	STATIONA	站点编号	30346
	STAORDER	站点序号	7
	DIRECTION	运营方向	1
	LOT	站点经度	104.064641
	LAT	站点维度	30.619511

[0248] GPS原始数据中不包含站点名称、站点序号、运营方向,因此需要通过GPS数据和GIS数据相同字段线路编号和站点编号这两个字段进行关联,给GPS数据添加线路运营方向、站点序号以及站点名称字段。由于GPS数据只是记录了车辆的时空信息,为了后续数据的利用需要通过运行班次的划分步骤来获得车辆相对独立的时空轨迹数据。

[0249] (2) 数据预处理。

[0250] 通过公交运行班次的划分以及重复、异常和缺失数据的处理后,即可以得到较为完备的数据。

[0251] 数据预处理后的样表

[0252]	字段	字段含义	数据样本
	DATATYPE	数据类型	4
	ROUTEID	线路编号	16
	PRODUCTID	车辆编号	21038004
	STATIONNUM	站点编号	30346
	LONGITUDE	经度	104.064591
	LATITUDE	纬度	30.619281
	ACTDATETIME	定位点记录时间	2019-05-08 19:41:23
	GPSSPEED	车辆瞬时速度	1.59
	ISARRLFT	到离站标识	1
	GPSMILE	运营累计里程	38714006
	NAME	站点名称	人南立交桥南站
	STAORDER	站点序号	7
	DIRECTION	运营方向	1
	TANGCI	运行班次	38

[0253] (3) 通道实例数据分析

[0254] 成都市公交线路2019年5月6日至10日5个工作日的AVL存档数据被用于本发明模型的实例研究。本节选取成都市人民南路和蜀都大道的部分路段作为实例研究对象,其中,人民南路研究路段南向从华西第二第四医院站起至桐梓林站终,北向从人南立交桥南站起至华西第二第四医院站终;蜀都大道研究路段南向从大慈寺站起至水碾河站终,北向从水碾河站起至蜀都大道红星路口站终。两个路段均设有路侧公交专用道,路段的详细信息如下表所示:

[0255] 实例研究路段相关信息表

路段	方向	经过线路数	站点数	信号交叉口数量	长度(m)	平均站间距(m)
[0256] 人民南路	南向	3	5	6	2500	625
	北向	3	6	7	2400	480
蜀都大道	南向	4	4	4	1400	467
	北向	5	4	4	1300	433

[0257] 考虑到交通状态具有时变特征,将数据根据运营时段的不同进行划分,并针对高峰时段分别建立了早高峰(7:00-9:00)和晚高峰时段(17:00-19:00)的预测模型。其中,前四个工作日的高峰时段数据作为模型的训练数据集,最后一天的高峰时段数据作为模型的测试数据集,模型训练集和测试集的样本量大小如下表所示:

[0258] 研究路段各运营方向高峰时段的模型训练集和测试集样本量

路段	方向	时段	训练集样本量				测试集样本量
			6号	7号	8号	9号	10号
[0259] 人民南路	南向	早高峰	52	53	62	62	55
		晚高峰	48	49	49	48	47
	北向	早高峰	50	50	48	53	49
		晚高峰	47	50	47	54	43
蜀都大道	南向	早高峰	68	73	80	81	75
		晚高峰	95	105	99	101	98
	北向	早高峰	117	139	142	144	140
		晚高峰	117	123	118	125	117

[0260] (3-1) 支持向量机模型预测结果分析

[0261] 支持向量机的核心是核函数,不同的核函数可以构造不同的支持向量机模型。考虑到核函数的参数数量,本发明选取径向基函数作为所用支持向量机模型的核函数。此外,对于不同的案例样本,模型的三个关键参数的选择也会对预测精度产生影响,这三个关键参数分别为支持向量机参数 C 、 ϵ 和加权前车数量 σ 。根据已有的基于支持向量机的行程时间预测的研究,建议 C 和 ϵ 的取值范围分别为 $[2^{-5}, 2^5]$ 和 $[0.1, 0.3]$, σ 的取值范围为 $[2, 10]$ 。为获取参数的最优取值,本发明采用网格搜索的方法在上述范围内检索,最终取得的案例样本的参数最优取值如下表所示。

[0262] 基于支持向量机的各案例模型参数取值一览表

通道	运行方向	早高峰			晚高峰		
		C	ϵ	σ	C	ϵ	σ
[0263] 人民南路	南向	2^4	0.1	4	2^{-2}	0.3	10
	北向	2^1	0.2	10	2^{-3}	0.3	10
蜀都大道	南向	2^2	0.2	4	2^{-5}	0.3	9
	北向	2^{-3}	0.3	10	2^0	0.3	2

[0264] 为了评价共线线路的融合数据对模型预测效果的影响,本发明建立了以与目标车辆隶属于同一线路的前车车头时距和行程时间为输入变量的支持向量机预测模型作为比较,即该基准模型的输入变量为 $H_k^{i,j}$ 和 $T_{k,k+\mu}^{i-p-n,j}$ 。预测模型算法使用R语言编写,并在主频2.8GHz、八核处理器、8.0G内存、操作系统为Win10的计算机上运行,下表为两个基于支持向

量机的到站时间预测模型的预测结果。

[0265] 基于支持向量机的多线模型和单线模型预测结果对比

通道	运行方向	模型	早高峰			晚高峰		
			MAE (s)	MAPE (%)	RMSE (s)	MAE (s)	MAPE (%)	RMSE (s)
[0266] 人民南路	南向	多线模型	120.31	14.28%	172.9	130.96	14.04%	173.94
		单线模型	134.69	15.61%	183.8	137.07	14.61%	179.09

[0267] 蜀都大道	北向	多线模型	55.59	7.93%	74.55	110.26	11.96%	132.36
		单线模型	67.81	9.48%	85.53	117.42	12.75%	141.41
	南向	多线模型	47.52	13.76%	62.6	60.98	16.77%	73.13
		单线模型	52.02	14.78%	68.42	62.4	17.04%	74.55
	北向	多线模型	42.09	11.94%	54.46	61.32	15.85%	81.56
		单线模型	45.57	13.00%	57.15	65.07	16.86%	84.3

[0268] 由上表可知,基于共线线路融合数据的支持向量机模型预测结果均优于基于单线路运营数据的预测模型,且八组案例样本中有六组的MAPE值均小于15%,有一组小于10%,这证实了引入其他线路的运营信息有益于增强模型的预测性能。

[0269] (3-2) 神经网络模型预测结果分析

[0270] 本发明选取标准的三层前馈神经网络用于公交到站时间的预测,考虑到算法参数数量和收敛速度,量化共轭梯度(Scaled Conjugate Gradient)算法被用作模型的训练算法。针对不同的案例样本,预测模型需要选取不同的参数取值才能使预测结果最优。本发明建立的基于共线线路融合数据的神经网络模型含有隐藏层神经元数量、最大迭代次数和加权前车数量三个参数,三个参数的取值范围分别为[1,25]、[100,1000]、[2,10]。本发明对预测模型进行敏感性分析以获取参数的最优取值,最终取得的案例样本的参数最优取值如下表所示。

[0271] 基于神经网络的各案例模型参数取值一览表

通道	运行方向	早高峰			晚高峰		
		S_{ize} e	Max $-it$	σ	S ize	Max $-it$	σ
[0272] 人民南路	南向	3	100	9	5	700	2
	北向	4	600	5	2	200	9
蜀都大道	南向	20	500	10	8	200	10
	北向	4	200	3	14	400	2

[0273] 为了探讨共线线路的融合数据对神经网络模型预测效果的影响,本发明还建立了以与目标车辆线路号相同的前车车头时距和行程时间为输入变量的神经网络预测模型作为比较,即该基准模型的输入变量为 $H_k^{i,j}$ 和 $T_{k,k+\mu}^{i-p-n,j}$ 。预测模型算法使用R语言编写,并在主频2.8GHz、八核处理器、8.0G内存、操作系统为Win10的计算机上运行,下表为两个基于神经网络

络的到站时间预测模型的预测结果。

[0274] 基于神经网络的多线模型和单线模型预测结果对比

通道	运行方向	模型	早高峰			晚高峰		
			MAE(s)	MAPE (%)	RMSE (s)	MAE(s)	MAPE (%)	RMSE (s)
[0275] 人民南路	南向	多线模型	101.7	11.89	147.69	130.82	13.78	170.06
		单线模型	103.48	12.31	148.94	141.47	14.92	186.75
	北向	多线模型	56.5	8.05	76.19	116.45	12.53	143.83

[0276] 蜀都大道	南向	单线模型	87.34	12.26	120.19	120.29	12.94	145.64
	南向	多线模型	54.32	15.73	70.72	60.29	16.94	76.32
		单线模型	59.97	17.06	77.46	69.04	18.61	81.6
	北向	多线模型	43.86	12.61	57.41	61.34	15.76	83.46
单线模型		47.51	13.48	56.9	65.69	16.73	87.28	

[0277] 由上表可知,与支持向量机模型预测结果类似,基于共线线路融合数据的神经网络模型预测结果均优于基于单线路运营数据的预测模型,且八组案例样本中有五组的MAPE值均小于15%,有一组小于10%,这也进一步表明了基于共线线路融合数据的预测模型具有良好的预测性能。

[0278] (3-3) 不同模型预测结果对比

[0279] 尽管已经证实了基于共线线路融合数据的支持向量机模型和神经网络模型均具有良好的到站时间预测能力,但两种模型的预测结果仍然存在差异。下表为两种模型的到站时间预测结果对比,从评价指标值来看,八组案例样本中有三组支持向量机的预测结果优于神经网络,有三组两种模型的预测效果相当。对八组案例样本在两种模型下的到站时间预测值和真实值进行相关性分析,共有四组支持向量机的预测值与真实值之间的相关性系数优于神经网络,其中除人民南路和蜀都大道南向晚高峰的三个案例样本的置信区间为95%之外,其余样本的置信区间均为99%。下表为两种模型下的到站时间预测值和观测值之间的相关性分析结果。

[0280] 两种模型下的到站时间预测结果对比

通道	运行方向	模型	早高峰			晚高峰		
			MAE(s)	MAPE(%)	RMSE(s)	MAE(s)	MAPE(%)	RMSE(s)
[0281] 人民南路	南向	支持向量机	120.31	14.28	172.9	130.96	14.04	173.94
		神经网络	101.7	11.89	147.69	130.82	13.78	170.06
	北向	支持向量机	55.59	7.93	74.55	110.26	11.96	132.36
		神经网络	56.5	8.05	76.19	116.45	12.53	143.83
蜀都大道	南向	支持向量机	47.52	13.76	62.6	60.98	16.77	73.13
		神经网络	54.32	15.73	70.72	60.29	16.94	76.32
	北向	支持向量机	42.09	11.94	54.46	61.32	15.85	81.56
		神经网络	42.84	12.24	54.71	57.6	14.69	79.58

[0282] 两种模型下的到站时间预测值和观测值之间的相关系数和P值一览表

通道	运行方向	模型	早高峰		晚高峰	
			相关系数	P 值	相关系数	P 值
[0283] 人民南路	南向	支持向量机	0.751	0.000	0.342	0.023
		神经网络	0.814	0.000	0.365	0.012
	北向	支持向量机	0.811	0.000	0.348	0.022
		神经网络	0.808	0.000	0.317	0.038
蜀都大道	南向	支持向量机	0.532	0.000	0.256	0.016
		神经网络	0.356	0.002	0.287	0.026
	北向	支持向量机	0.347	0.000	0.223	0.016
		神经网络	0.251	0.003	0.261	0.005

[0284] 通过选取成都市人民南路和蜀都大道的部分路段作为实例研究对象,利用提出的两种预测模型对上述通道上的公交车到站时间进行预测。为了验证预测模型的有效性,使用了与目标车辆线路号相同的前车车头时距和行程时间为输入变量的两种基准模型分别作为比较。预测结果表明,无论是支持向量机还是前馈神经网络,基于共线线路融合数据的模型预测结果均优于基于单线路运营数据的预测模型。此外,本发明还对支持向量机和前馈神经网络模型的预测性能进行了评价和比较,从到站时间预测值和真实值的相关性分析结果来看,支持向量机模型的预测性能优于前馈神经网络模型。

[0285] 本发明通过“基于支持向量机的多线模型和单线模型预测结果对比”和“基于神经网络的多线模型和单线模型预测结果对比”的结果可以表明本发明采取多线模型的预测结果优于单线模型,即本发明采取的多线模型的预测结果在现有的单线模型基础上准确度有所提高。

[0286] 以上所述,仅为本发明的具体实施方式,但本发明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发明揭露的技术范围内,凡在本发明的精神和原则之内所作的任何修改、等同替换和改进等,都应涵盖在本发明的保护范围之内。

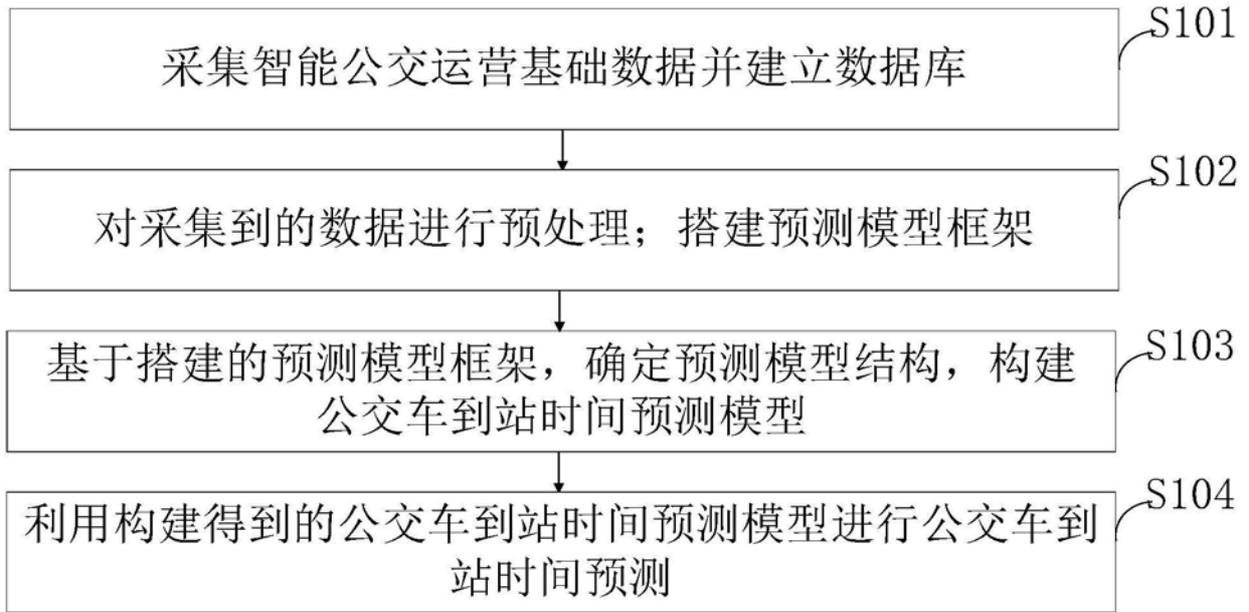


图1



图2

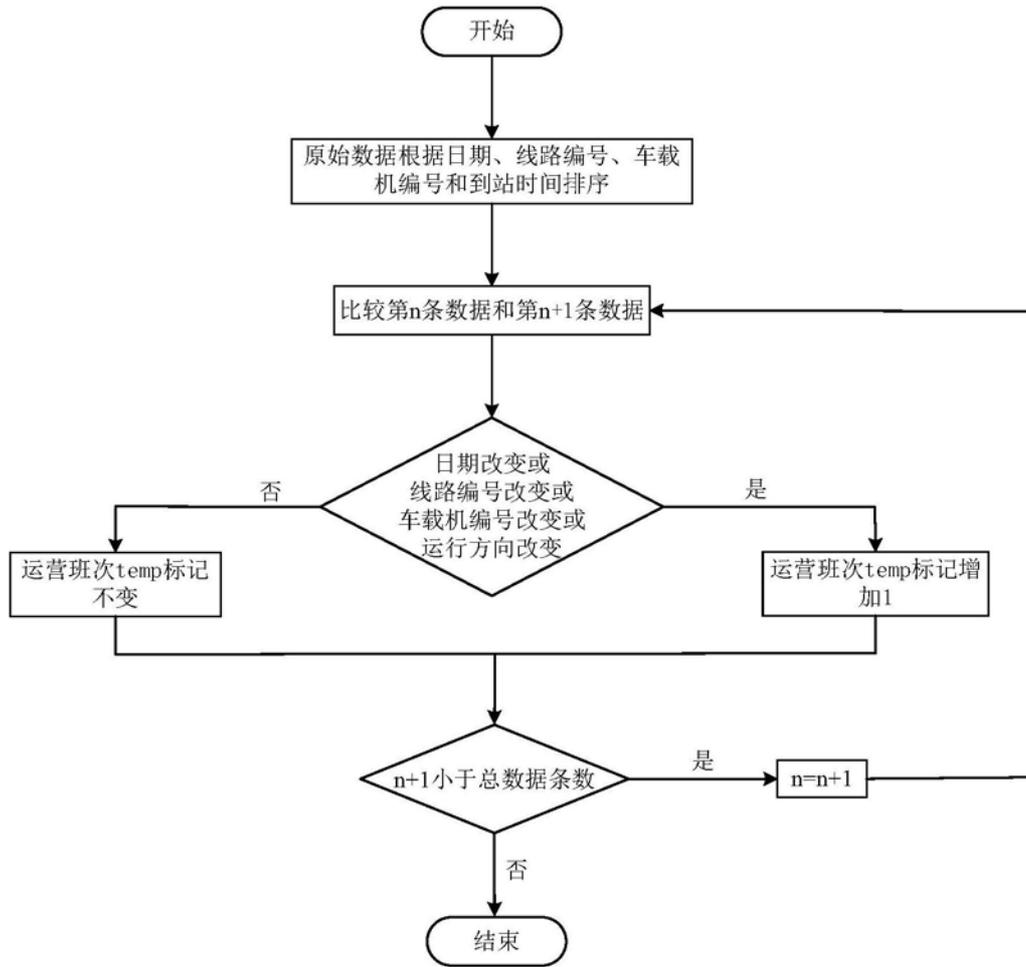


图3

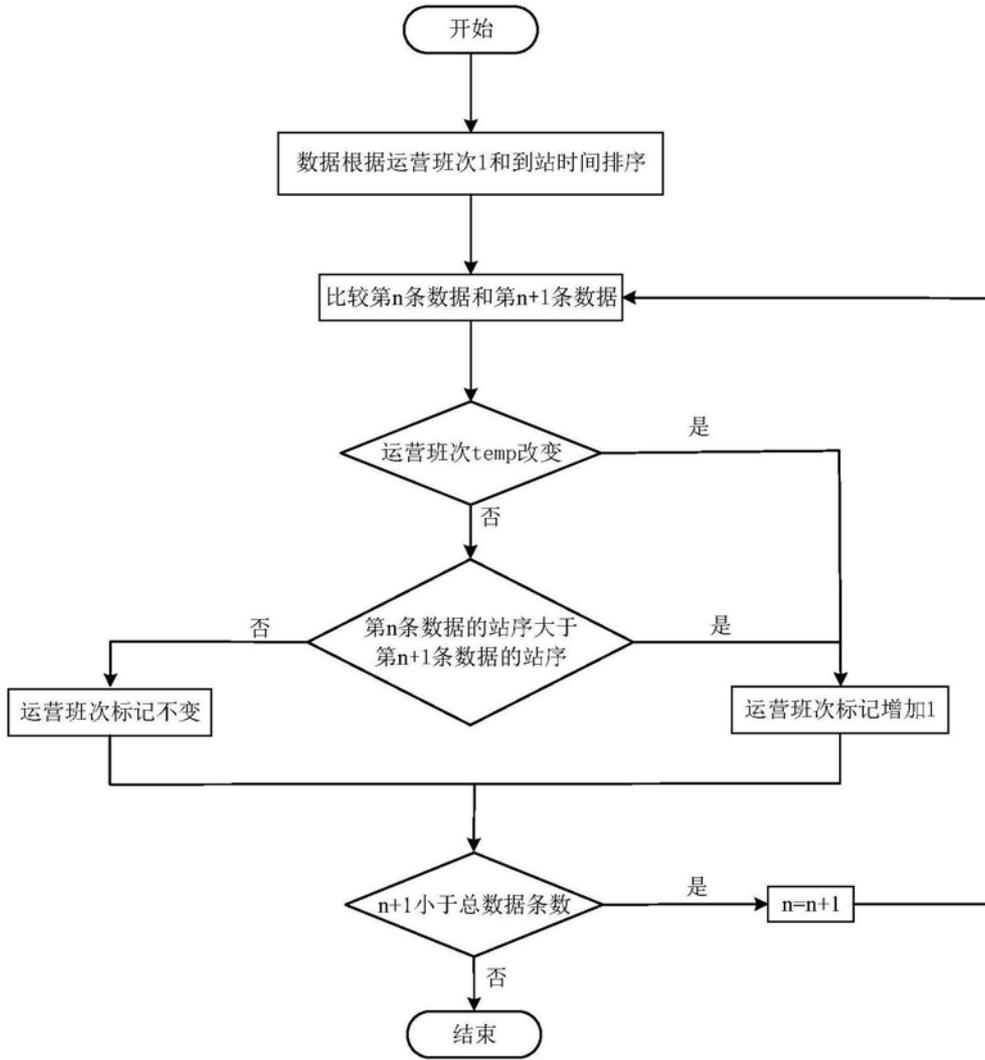


图4

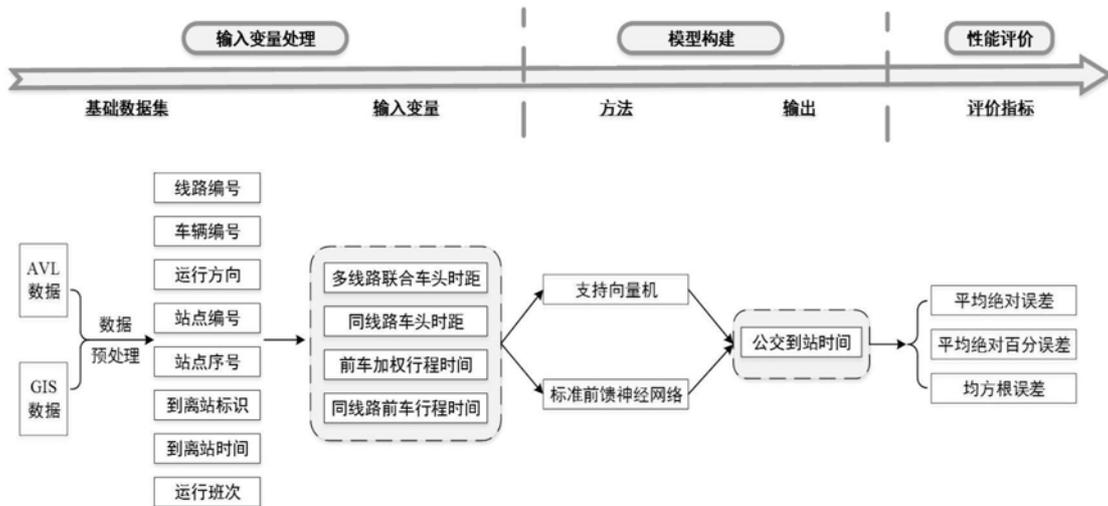


图5

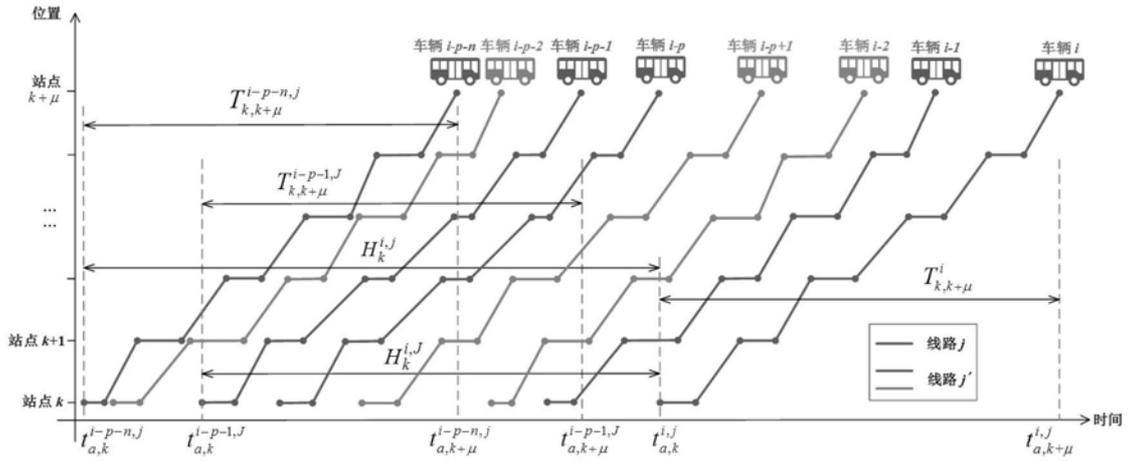


图6

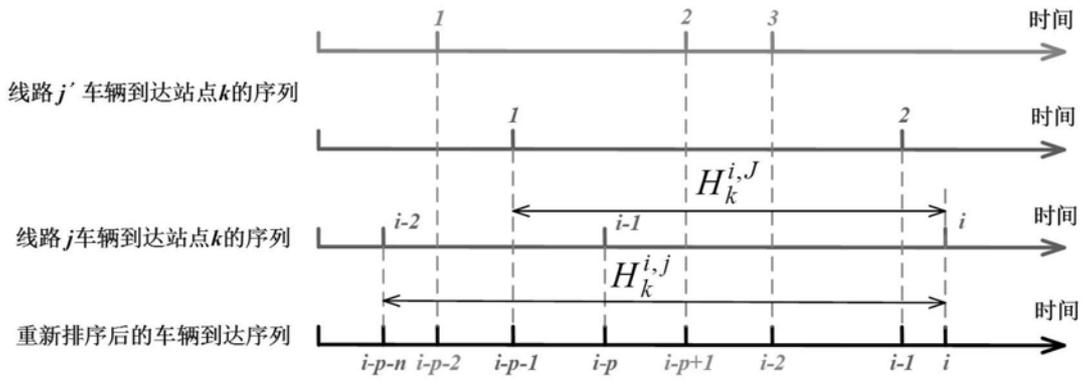


图7

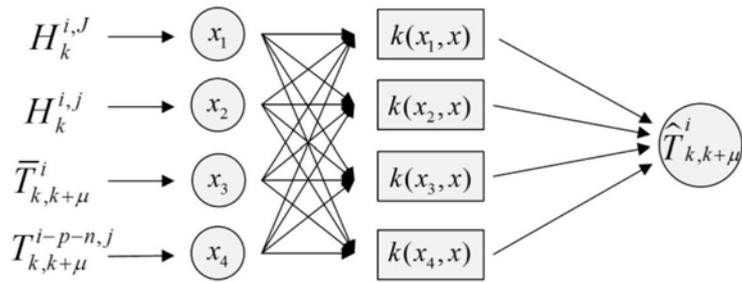


图8

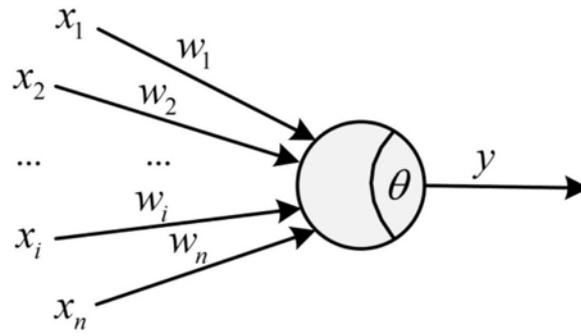


图9

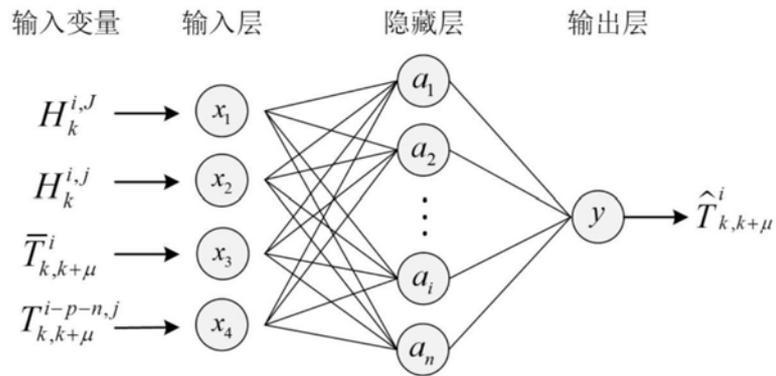


图10