



(12)发明专利

(10)授权公告号 CN 104464280 B

(45)授权公告日 2017.05.03

(21)申请号 201410453712.9

G06F 17/30(2006.01)

(22)申请日 2014.09.05

审查员 熊健

(65)同一申请的已公布的文献号

申请公布号 CN 104464280 A

(43)申请公布日 2015.03.25

(73)专利权人 广州市香港科大霍英东研究院

地址 511458 广东省广州市南沙区南沙资讯科技园软件楼N301室

(72)发明人 丁焯 高民 谭浩宇 倪明选

(74)专利代理机构 广州三环专利代理有限公司

44202

代理人 梁顺宜 郝传鑫

(51)Int.Cl.

G08G 1/01(2006.01)

G08G 1/065(2006.01)

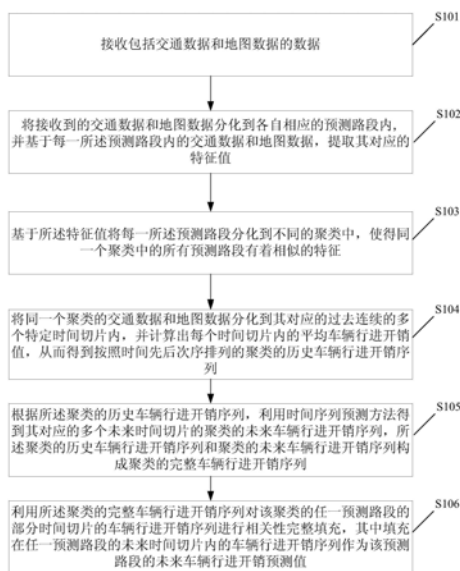
权利要求书4页 说明书13页 附图5页

(54)发明名称

车辆行进开销的预测方法及系统

(57)摘要

本发明公开了一种车辆行进开销的预测方法,包括步骤:将接收到的交通数据和地图数据分化到各自相应的预测路段内并提取其对应的特征值;基于特征值将每一预测路段分化到不同的聚类中;通过计算并预测获得每个聚类的完整车辆行进开销序列;利用聚类的完整车辆行进开销序列对该聚类的任一预测路段的部分时间切片的车辆行进开销序列进行相关性完整填充,其中填充在任一预测路段的未来时间切片内的车辆行进开销序列作为该预测路段的未来车辆行进开销预测值。本发明还公开了一种车辆行进开销的预测系统。



1. 一种车辆行进开销的预测方法,其特征在于,包括步骤:

A、将接收到的交通数据和地图数据分化到各自相应的预测路段内,并基于每一所述预测路段内的交通数据和地图数据,提取其对应的特征值;

B、基于所述特征值将每一所述预测路段分化到不同的聚类中,使得同一个聚类中的所有预测路段有着相似的特征;

C、将同一个聚类的交通数据和地图数据分化到其对应的过去连续的多个特定时间切片内,并计算出每个时间切片内的平均车辆行进开销值,从而得到按照时间先后次序排列的聚类的历史车辆行进开销序列;所述车辆行进开销值是指车辆行进速度的平均值;

D、根据所述聚类的历史车辆行进开销序列,利用时间序列预测方法得到其对应的多个未来时间切片的聚类的未来车辆行进开销序列,所述聚类的历史车辆行进开销序列和聚类的未来车辆行进开销序列构成聚类的完整车辆行进开销序列;

E、利用所述聚类的完整车辆行进开销序列对该聚类的任一预测路段的部分时间切片的车辆行进开销序列进行相关性完整填充,其中填充在任一预测路段的未来时间切片内的车辆行进开销序列作为该预测路段的未来车辆行进开销预测值;所述相关性完整填充是指将所述聚类的完整车辆行进开销序列中的对应数值填补到预测路段的部分时间切片的车辆行进开销序列的缺失值处以完成缺失值填充。

2. 如权利要求1所述的车辆行进开销的预测方法,其特征在于,每一所述预测路段可以由连续的多个属于此路段的时空分割路段表示,所述时空分割路段定义为给定时间段下的特定长度路段,且所述给定时间段与所述特定时间切片一致。

3. 如权利要求1或2所述的车辆行进开销的预测方法,其特征在于,所述特征值包括静态特征值和动态特征值,所述静态特征值包括但不限于时空分割路段类型、时空分割路段长度、时空分割路段的近邻时空分割路段平均类型、时空分割路段的近邻时空分割路段平均长度、时空分割路段的近邻时空分割路段类型方差、时空分割路段的近邻时空分割路段长度方差以及时空分割路段的近邻时空分割路段数量;所述动态特征值包括但不限于时空分割路段的近邻时空分割路段第*i*个时间段平均行进速度以及时空分割路段的近邻时空分割路段第*i*个时间段行进速度方差;其中,*i*=0时表示同一时间段内,*i*=1与*i*=-1分别表示下一个时间段内和上一个时间段内,依此类推。

4. 如权利要求1所述的车辆行进开销的预测方法,其特征在于,所述步骤B具体包括:

B1、通过公式(1)将每一所述预测路段的特征值进行归一化处理,从而减少各个特征值的大小差距:

$$F_{norm} = \frac{F_{origin}}{F_{max}} \quad \text{公式(1)}$$

其中, F_{origin} 是每个时空分割路段的初始特征值, F_{max} 是此类特征的最大值, F_{norm} 为归一化后的特征值,因此归一化后的特征值均处于零到一之间;

B2、基于经过归一化处理后的特征值并通过K-Means的聚类方法将每一所述预测路段分化到不同的聚类中,从而使得同一个聚类中的所有预测路段有着相似的特征,具体包括步骤:

B21、初始化定义K个聚类中心;

B22、根据公式(2)中定义的预测路段与聚类中心的距离,将各个预测路段归类到其距离最近的聚类中心处:

$$Dis_{ac} = \sqrt{\sum_{i=1}^n (f_i^a - f_i^c)^2} \quad \text{公式(2)}$$

其中,Dis_{ac}表示任一预测路段到任一聚类中心的距离,f_i^a代表该预测路段的第i个特征值,而f_i^c代表该聚类中心的第i个特征值,n表示特征值的数量;

B23、根据每个聚类中的预测路段的平均特征值更新该聚类的中心,直至预测路段的平均特征值与聚类中心完全重合无法更新为止。

5.如权利要求1所述的车辆行进开销的预测方法,其特征在于,所述步骤E具体包括步骤:

E1、定义所述聚类的完整车辆行进开销序列与该聚类中的任一预测路段的部分时间切片的车辆行进开销序列的距离如公式(4)所示:

$$Distance(CP,R) = \frac{\sum_{R.tc} (R.tc - CP.correspondingTc)^2}{N} \quad \text{公式(4)}$$

其中,CP表示聚类的完整车辆行进开销序列,R表示任一预测路段的部分时间切片的车辆行进开销序列,R.tc表示该预测路段已拥有的部分时间切片的车辆行进开销值,CP.correspondingTc表示聚类完整车辆行进开销时间序列中对应于R.tc同一时间片段中的车辆行进开销值,N是不为0的自然数;

E2、将聚类的完整车辆行进开销序列的数据整体放大/缩小α倍,从而使得公式(4)中的距离Distance(CP,R)最小;其中,α为正实数;

E3、将缩放后的所述聚类的完整车辆行进开销序列中的对应数值填补到所述预测路段的部分时间切片的车辆行进开销序列的缺失值处,填充完整后形成预测路段的车辆行进开销序列,并使该预测路段的车辆行进开销序列与所述聚类的完整车辆行进开销序列的相关性最高;

E4、对于填充完整后的所述预测路段的车辆行进开销序列,提取填充在该预测路段的未来时间切片内的车辆行进开销序列作为该预测路段的未来车辆行进开销预测值。

6.一种车辆行进开销的预测系统,其特征在于,包括:

预测路段分化单元,用于将接收到的交通数据和地图数据分化到各自相应的预测路段内;

预测路段特征值提取单元,基于每一所述预测路段内的交通数据和地图数据,提取其对应的特征值;

预测路段聚类单元,基于所述特征值将每一所述预测路段分化到不同的聚类中,使得同一个聚类中的所有预测路段有着相似的特征;

聚类的历史车辆行进开销序列计算单元,用于将同一个聚类的交通数据和地图数据分化到其对应的过去连续的多个特定时间切片内,并计算出每个时间切片内的平均车辆行进开销值,从而得到按照时间先后次序排列的聚类的历史车辆行进开销序列;所述车辆行进开销值是指车辆行进速度的平均值;

聚类的未来车辆行进开销序列预测单元,用于根据所述聚类的历史车辆行进开销序列,利用时间序列预测方法得到其对应的多个未来时间切片的聚类的未来车辆行进开销序列,所述聚类的历史车辆行进开销序列和聚类的未来车辆行进开销序列构成聚类的完整车辆行进开销序列;以及

预测路段的行进开销预测单元,利用所述聚类的完整车辆行进开销序列对该聚类的任一预测路段的部分时间切片的车辆行进开销序列进行相关性完整填充,其中填充在任一预测路段的未来时间切片内的车辆行进开销序列作为该预测路段的未来车辆行进开销预测值;所述相关性完整填充是指将所述聚类的完整车辆行进开销序列中的对应数值填补到预测路段的部分时间切片的车辆行进开销序列的缺失值处以完成缺失值填充。

7.如权利要求6所述的车辆行进开销的预测系统,其特征在于,每一所述预测路段可以由连续的多个属于此路段的时空分割路段表示,所述时空分割路段定义为给定时间段下的特定长度路段,且所述给定时间段与所述特定时间切片一致。

8.如权利要求6所述的车辆行进开销的预测系统,其特征在于,所述特征值包括静态特征值和动态特征值,所述静态特征值包括但不限于时空分割路段类型、时空分割路段长度、时空分割路段的近邻时空分割路段平均类型、时空分割路段的近邻时空分割路段平均长度、时空分割路段的近邻时空分割路段类型方差、时空分割路段的近邻时空分割路段长度方差以及时空分割路段的近邻时空分割路段数量;所述动态特征值包括但不限于时空分割路段的近邻时空分割路段第*i*个时间段平均行进速度以及时空分割路段的近邻时空分割路段第*i*个时间段行进速度方差;其中,*i*=0时表示同一时间段内,*i*=1与*i*=-1分别表示下一个时间段内和上一个时间段内,依此类推。

9.如权利要求6所述的车辆行进开销的预测系统,其特征在于,所述预测路段聚类单元具体包括:

归一化模块,用于通过公式(1)将每一所述预测路段的特征值进行归一化处理,从而减少各个特征值的大小差距:

$$F_{norm} = \frac{F_{origin}}{F_{max}} \quad \text{公式(1)}$$

其中, F_{origin} 是每个预测路段的初始特征值, F_{max} 是此类特征的最大值, F_{norm} 为归一化后的特征值,因此归一化后的特征值均处于零到一之间;

K-Means聚类模块,基于经过归一化处理后的特征值并通过K-Means的聚类方法将每一所述预测路段分化到不同的聚类中,从而使得同一个聚类中的所有预测路段有着相似的特征;

所述K-Means聚类模块通过以下步骤将每一所述预测路段分化到不同的聚类中:

(1)、初始化定义K个聚类中心;

(2)、根据公式(2)中定义的预测路段与聚类中心的距离,将各个预测路段归类到其距离最近的聚类中心处:

$$Dis_{ac} = \sqrt{\sum_{i=1}^n (f_i^a - f_i^c)^2} \quad \text{公式(2)}$$

其中, Dis_{ac} 表示任一预测路段到任一聚类中心的距离, f_i^a 代表该预测路段的第*i*个特征值, 而 f_i^c 代表该聚类中心的第*i*个特征值, n 表示特征值的数量;

(3)、根据每个聚类中的预测路段的平均特征值更新该聚类的中心, 直至预测路段的平均特征值与聚类中心完全重合无法更新为止。

10. 如权利要求6所述的车辆行进开销的预测系统, 其特征在于, 所述预测路段的行进开销预测单元通过以下步骤计算预测路段的未来车辆行进开销预测值:

(1) 定义所述聚类的完整车辆行进开销序列与该聚类中的任一预测路段的部分时间切片的车辆行进开销序列的距离如公式(4)所示:

$$Distance(CP, R) = \frac{\sum_{R.tc} (R.tc - CP.correspondingTc)^2}{N} \quad \text{公式(4)}$$

其中, CP表示聚类的完整车辆行进开销序列, R表示任一预测路段的部分时间切片车辆行进开销序列, R.tc表示该预测路段已拥有的部分时间切片的车辆行进开销值, CP.correspondingTc表示聚类完整车辆行进开销时间序列中对应于R.tc同一时间片段中的车辆行进开销值, N是不为0的自然数;

(2) 将聚类的完整车辆行进开销序列的数据整体放大/缩小 α 倍, 从而使得公式(4)中的距离Distance(CP, R)最小; 其中, α 为正实数;

(3) 将缩放后的所述聚类的完整车辆行进开销序列中的对应数值填补到所述预测路段的部分时间切片的车辆行进开销序列的缺失值处, 填充完整后形成预测路段的车辆行进开销序列, 并使该预测路段的车辆行进开销序列与所述聚类的完整车辆行进开销序列的相关性最高;

(4) 对于填充完整后的所述预测路段的车辆行进开销序列, 提取填充在该预测路段的未来时间切片内的车辆行进开销序列作为该预测路段的未来车辆行进开销预测值。

车辆行进开销的预测方法及系统

技术领域

[0001] 本发明涉及交通数据处理领域,尤其涉及一种车辆行进开销的预测方法及系统。

背景技术

[0002] 随着技术的发展,越来越多的交通数据得以被人们所收集并利用。例如,在GPS设备的普及下,大都市出租车辆均装备了GPS设备。通过这些设备以及其他一些相关车辆信息采集设备,出租车可以不断地将自己的位置信息,速度信息以及载客与否信息传送给数据中心。于是,数据中心便可以收集海量的交通数据,从而进行一系列的数据挖掘与分析任务。由于海量交通数据富含的丰富价值,基于交通数据的挖掘工作也变成了时下下一个热活跃的热点研究领域。交通数据相关的研究工作主要包含两个目的。其一是通过对海量交通数据的分析,来发掘城市的一些动态特征。这方面的研究包括研究城市规划合理程度,分析城市不同区域的功能模块等等。而第二个目的,则是通过利用海量交通数据,从而为用户提供与交通相关的服务。例如,根据历史交通车辆的行进轨迹对用户做最优路线推荐,以及通过出租车行驶历史数据提供优化的打车推荐服务,帮助用户更加容易打到出租车,也使得出租车更容易找到用户等等。

[0003] 基于交通数据的研究,一方面,由于大城市出租车数量众多且它们均会频繁发送车辆信息,数据中心会收集到海量数据,对挖掘提出了数据规模上的挑战。而另一方面,由于出租车仅仅占大城市所有车辆中的一个微小的部分,对于这类数据开展的数据挖掘任务同时又面临着数据稀缺的挑战。然而,在众多对于交通数据进行的数据挖掘任务中,对于道路行进开销的数据挖掘任务尤为重要,因为它扮演着对于其他数据挖掘任务提供基础的角色。比如知道各个道路的行进开销有助于城市建设规划,同时,也方便于系统提供合理路线给用户而完成旅行路线推荐的服务。由此,此项对道路行进开销估计与预测的工作有着强烈的动机。

[0004] 道路行进开销的数据挖掘任务,具体来说,就是通过采集交通数据,将这些交通数据归类到相应的道路上,并通过这些交通数据来计算不同时间段上这条道路的车辆行进单位距离所需要的时间耗费。所以,道路行进开销的数据挖掘也可以理解成对于道路期望的行进速度的计算,因为一旦我们可以掌握道路期望的行进速度,此道路上车辆行进单位距离所需要耗费的时间也可以自然快捷地计算出来。需要指出的是,不同车辆在同一条道路同一个时间段的行进开销有所不同,这跟驾驶员的驾驶习惯,当时道路上的具体道路交通状况有关。这里所说的道路行进开销的计算是相对于此道路在这一特定时间段的车辆行进开销期望而言。这个行进开销值可以对这条路段这个时间的交通状况提供一个合理的参考。

[0005] 这项工作中设计到的道路行进开销的数据挖掘任务涉及了两方面。第一方面,是对于道路行进开销的估计。这一部分,主要着重于时间点的当下,由于出租车数据的稀疏特性,部分路段在特定时间段并没有出租车覆盖,也没有相应的交通数据,因此其道路行进开销并不能通过简单根据其上车辆行进速度计算而得到。然而,对于收集到交通数据的路段

和未收集到交通数据的路段,我们都需要对其当前的车辆行进开销进行估计。第二方面,是对于道路行进开销的预测。这一部分,主要着重于时间点的未来,我们根据当前收集到的交通数据,对于路段在未来时间的车辆行进开销进行预测。这一部分的工作相对第一部分更为重要,因为对于一些行进路径推荐的服务,了解未来的道路行进开销对于优化路径的提出有着重要的意义。

[0006] 在现有技术中,关于道路车辆行进开销的估计与道路车辆行进开销的预测,均有相关工作展开。其中,关于道路车辆行进开销的预测的工作,在本发明之前也有进行。在现有技术中,对于一个路段,收集其不同时间段中的车辆行进开销,并将这一系列的车辆行进开销作为时间序列,运用时间序列预测的方法对未来的车辆行进开销作预测。另一项工作,假设同一路段每天同一时间的车辆行进开销相同,并且通过计算历史数据每天同一时间的车辆行进开销的平均值对未来某天同一时间的车辆行进开销作预测。这些先前工作均忽略了出租车数据的稀疏特性。它们只能对于有交通数据的路段进行未来的车辆行进开销预测,而对于大部分没有被出租车覆盖到的路段或者某些时间段缺乏数据的路段,均无法进行准确有效的未来车辆行进开销预测。

发明内容

[0007] 本发明的目的提供一种车辆行进开销的预测方法及系统,克服现有技术中没有解决的数据稀疏问题,从而可以完整地对于任意路段都可以完成未来车辆行进开销的预测。

[0008] 为实现上述目的,本发明提供了一种车辆行进开销的预测方法,包括步骤:

[0009] A、将接收到的交通数据和地图数据分化到各自相应的预测路段内,并基于每一所述预测路段内的交通数据和地图数据,提取其对应的特征值;

[0010] B、基于所述特征值将每一所述预测路段分化到不同的聚类中,使得同一个聚类中的所有预测路段有着相似的特征;

[0011] C、将同一个聚类的交通数据和地图数据分化到其对应的过去连续的多个特定时间切片内,并计算出每个时间切片内的平均车辆行进开销值,从而得到按照时间先后次序排列的聚类的历史车辆行进开销序列;

[0012] D、根据所述聚类的历史车辆行进开销序列,利用时间序列预测方法得到其对应的多个未来时间切片的聚类的未来车辆行进开销序列,所述聚类的历史车辆行进开销序列和聚类的未来车辆行进开销序列构成聚类的完整车辆行进开销序列;

[0013] E、利用所述聚类的完整车辆行进开销序列对该聚类的任一预测路段的部分时间切片的车辆行进开销序列进行相关性完整填充,其中填充在任一预测路段的未来时间切片内的车辆行进开销序列作为该预测路段的未来车辆行进开销预测值。

[0014] 本发明提供了一种车辆行进开销的预测系统,包括:

[0015] 预测路段分化单元,用于将接收到的交通数据和地图数据分化到各自相应的预测路段内;

[0016] 预测路段特征值提取单元,基于每一所述预测路段内的交通数据和地图数据,提取其对应的特征值;

[0017] 预测路段聚类单元,基于所述特征值将每一所述预测路段分化到不同的聚类中,使得同一个聚类中的所有预测路段有着相似的特征;

[0018] 聚类的历史车辆行进开销序列计算单元,用于将同一个聚类的交通数据和地图数据分化到其对应的过去连续的多个特定时间切片内,并计算出每个时间切片内的平均车辆行进开销值,从而得到按照时间先后次序排列的聚类的历史车辆行进开销序列;

[0019] 聚类的未来车辆行进开销序列预测单元,用于根据所述聚类的历史车辆行进开销序列,利用时间序列预测方法得到其对应的多个未来时间切片的聚类的未来车辆行进开销序列,所述聚类的历史车辆行进开销序列和聚类的未来车辆行进开销序列构成聚类的完整车辆行进开销序列;以及

[0020] 预测路段的行进开销预测单元,利用所述聚类的完整车辆行进开销序列对该聚类的任一预测路段的部分时间切片的车辆行进开销序列进行相关性完整填充,其中填充在任一预测路段的未来时间切片内的车辆行进开销序列作为该预测路段的未来车辆行进开销预测值。

[0021] 与现有技术相比,本发明公开的车辆行进开销的预测方法及系统具有如下有益效果:通过各个交通车辆信息点的具体信息来划预测路段,并寻找分类相似的预测路段,然后基于同一个聚类中的所有预测路段有着相似的特征的原则,通过计算聚类的完整车辆行进开销序列模式后来对该聚类的任一预测路段的部分时间切片的车辆行进开销序列进行相关性完整填充,从而使填充在任一预测路段的未来时间切片内的车辆行进开销序列作为该预测路段的未来车辆行进开销预测值。因此,本发明通过找寻相似路段,并且认为相似路段的车辆行进开销的变化模式相似来进行车辆行进开销的预测。通过这一方法,可以克服先前工作没有解决的数据稀疏问题,从而可以完整地对于任意路段都可以完成未来车辆行进开销的预测。

附图说明

[0022] 图1是本发明实施例1中一种车辆行进开销的预测方法的流程示意图。

[0023] 图2是图1所示的一种车辆行进开销的预测方法中的步骤S103的具体流程示意图。

[0024] 图3是图1所示的一种车辆行进开销的预测方法中的步骤S106的具体流程示意图。

[0025] 图4是本发明实施例2中一种车辆行进开销的预测系统的结构示意图。

[0026] 图5是图4所示的一种车辆行进开销的预测系统的预测路段聚类单元的结构示意图。

具体实施方式

[0027] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0028] 实施例1

[0029] 参见图1,是本发明实施例提供的一种车辆行进开销的预测方法的流程示意图,包括步骤:

[0030] S101、接收包括交通数据和地图数据的数据;

[0031] 其中,所述交通数据至少包括车辆不同时间点的当前时间信息、位置经度信息和

行进速度信息,所述地图数据包括路段的起始终止经纬度信息。

[0032] S102、将接收到的交通数据和地图数据分化到各自相应的预测路段内,并基于每一所述预测路段内的交通数据和地图数据,提取其对应的特征值;

[0033] S103、基于所述特征值将每一所述预测路段分化到不同的聚类中,使得同一个聚类中的所有预测路段有着相似的特征;

[0034] S104、将同一个聚类的交通数据和地图数据分化到其对应的过去连续的多个特定时间切片内,并计算出每个时间切片内的平均车辆行进开销值,从而得到按照时间先后次序排列的聚类的历史车辆行进开销序列;

[0035] S105、根据所述聚类的历史车辆行进开销序列,利用时间序列预测方法得到其对应的多个未来时间切片的聚类的未来车辆行进开销序列,所述聚类的历史车辆行进开销序列和聚类的未来车辆行进开销序列构成聚类的完整车辆行进开销序列;

[0036] S106、利用所述聚类的完整车辆行进开销序列对该聚类的任一预测路段的部分时间切片的车辆行进开销序列进行相关性完整填充,其中填充在任一预测路段的未来时间切片内的车辆行进开销序列作为该预测路段的未来车辆行进开销预测值。

[0037] 在步骤S102中,基于步骤S101输入的交通数据与地图数据进行路段的时空分割,同时由于交通数据中有车辆的位置信息和当前时间信息,所以各个数据点可以根据此类时空信息分化到其应该属于的相关预测路段中。

[0038] 其中,每一所述预测路段可以由连续的多个属于此路段的时空分割路段表示,所述时空分割路段定义为给定时间段下的特定长度路段,且所述给定时间段与所述特定时间切片一致。优选的,所述预测路段与所述特定长度路段的路段长度相同。

[0039] “时空分割路段”即是在给定时间段下的特定长度路段,在本实施例中,这里的时间段长度默认为三十分钟。定义在这个短暂时间段中一个路段的车辆行进开销保持不变或者变化微小可以忽略。

[0040] 由于定义的时空分割路段是将一个路段在时间轴上按照三十分钟为一个时间片进行切片,因此,本实施例的预测路段可以由连续的四十八个属于此路段的时空分割路段表示,这样可以表现出完整一天的预测路段。于是,预测路段的特征可以由所有完整一天的时空分割路段的特征并集来表示。

[0041] 在步骤S102中,本实施例为每一个预测路段中的每一个时空分割路段提取特征值以对其进行描述。在实施例中,每一个时空分割路段提取的特征值包括为静态特征和动态特征两类。其中,静态特征主要用于描述时空分割路段及其附近时空分割路段的路网的结构,这些特征在长期是相对稳定不变的特征,被归为静态特征一类。而动态特征旨在提取该时空分割路段的动态信息,主要包括该时空分割路段及其近邻时空分割路段的车辆平均行进速度的相关信息。该类信息相关于实时的车辆行进情况,会不断地发生变化,被归为动态特征一类。对于每一时空分割路段提取的完整特征其及相关描述总结与表1中:

[0042]

特征	描述
T _{rid}	时空分割路段类型
L _{rid}	时空分割路段长度
E _t (rid)	时空分割路段的近邻路段平均类型

$E_l(\text{rid})$	时空分割路段的近邻路段平均长度
$V_t(\text{rid})$	时空分割路段的近邻路段类型方差
$V_l(\text{rid})$	时空分割路段的近邻路段长度方差
N_{rid}	时空分割路段的近邻路段数量
$E_i(\text{sid})$	时空分割路段的近邻时空分割路段第 <i>i</i> 个时间段平均行进速度
$V_i(\text{sid})$	时空分割路段的近邻时空分割路段第 <i>i</i> 个时间段行进速度方差

[0043] 表1

[0044] 表1中,前七个特征用于描述路网信息,为静态特征。最后两个特征用于描述动态车辆行进情况,属于动态特征。在地图数据中,路段均会根据其道路等级进行编号标示,由此数字编号可以得到路段的类型信息。对于 $E_i(\text{sid})$ 与 $V_i(\text{sid})$, $i=0$ 使得两个特征分别表示此时空分割路段的近邻路段在同一时间片内的平均车辆行进速度与车辆行进速度方差。而 $i=1$ 与 $i=-1$ 分别表示下一个时间片内与上一个时间片内相关的特征值,依此类推。

[0045] 提取每个预测路段的特征值后,则需要基于所述特征值将每一所述预测路段分化到不同的聚类中。

[0046] 具体的,结合图2,步骤S103的预测路段聚类过程具体包括:

[0047] S1031、将每一所述预测路段的特征值进行归一化处理,从而减少各个特征值的大小差距;

[0048] S1032、基于经过归一化处理后的特征值并通过K-Means的聚类方法将每一所述预测路段分化到不同的聚类中,从而使得同一个聚类中的所有预测路段有着相似的特征。

[0049] 其中,在步骤S1041中,通过公式(1)对所述每一预测路段的特征值进行归一化处理:

$$[0050] \quad F_{\text{norm}} = \frac{F_{\text{origin}}}{F_{\text{max}}} \quad \text{公式(1)}$$

[0051] 其中, F_{origin} 是每个预测路段的初始特征值, F_{max} 是此类特征的最大值, F_{norm} 为归一化后的特征值,因此归一化后的特征值均处于零到一之间。经过归一化处理,可以使得各个特征值的大小差距较归一化之前变小,进而使得此后在计算特征相似度时各个特征值的贡献会近似相同,从而避免因为个别特征值的较大贡献使得特征相似度计算失真。

[0052] 对每个预测路段的特征值进行归一化处理,通过步骤S1042所示的K-Means聚类方法进行预测路段聚类,具体包括过程:

[0053] (1)、初始化定义K个聚类中心;

[0054] (2)、根据公式(2)中定义的预测路段与聚类中心的距离,将各个预测路段归类到其距离最近的聚类中心处:

$$[0055] \quad Dis_{ac} = \sqrt{\sum_{i=1}^n (f_i^a - f_i^c)^2} \quad \text{公式(2)}$$

[0056] 其中, Dis_{ac} 表示任一预测路段到任一聚类中心的距离, f_i^a 代表该预测路段的第*i*个特征值,而 f_i^c 代表该聚类中心的第*i*个特征值;

[0057] 由此特征值相似度计算公式(2)可以看出,如果各个特征值的有效范围差距较大,则数值大的特征值可能会较大程度地定义特征值相似性而较小的特征值的相似性会被模糊,使得整个样本相似度计算失真。由此体现出公式(1)中特征值归一化处理的必要性。这使得各个特征值在计算样本相似度时会有着相似的贡献,使得样本相似度可以真实地反映所有特征值的相似度。

[0058] (3)、根据每个聚类中的预测路段的平均特征值更新该聚类的中心,直至预测路段的平均特征值与聚类中心完全重合无法更新为止。

[0059] 在该步骤中,根据每个聚类中的预测路段的平均特征值以更新该聚类的中心。如此上述将预测路段归类并更新聚类中心的过程会不断进行直到收敛,即预测路段的平均特征值与聚类中心完全重合无法更新为不同的聚类中心。这样的聚类过程,可以根据各个预测路段的特征值将它们分化到不同的聚类中,使得同一个聚类中的预测路段有着相似的特征。在本发明中,定义在这样的每一个聚类中,特征相似的预测路段会有着相似的性质,从而有着相似的车辆行进开销随时间变化模式。本发明正是根据这个前提从而克服交通数据稀疏的特性,完成对预测路段进行车辆行进开销预测。

[0060] 此后,在完成所有预测路段的聚类后,则进行“聚类的车辆行进开销模式预测”,即步骤S104~S105。这里的车辆行进开销模式预测针对每一个不同的聚类展开。对于每一个聚类,相应的会有大量交通数据归属于此聚类中的各个预测路段中。在步骤S104中,首先,本实施例以三十分钟为一个单位将时间轴切片,并且将所有的该聚类中的交通数据分化到其对应的时间切片中,计算出每个时间切片中的平均车辆行进开销值(即每个聚类的每个时间切片内的所有预测路段的平均车辆行进开销值,该平均车辆行进开销值可以通过计算每个聚类的每个时间切片内的所有预测路段的车辆行进速度的平均值获得)。然后,将这些平均车辆行进开销值按照时间先后次序排列从而形成一个时间序列数据。这个时间序列反映了过去时间中,这个路段聚类中车辆行进开销的平均变化模式。由于认为此路段聚类中各路段特征相似,其车辆行进开销变化模式也相似,因此,得出的此聚类的平均车辆行进开销变化模式的时间序列数据可以代表整个聚类中路段的模式变化情况。从而,此时间序列数据可以作为此聚类中各路段车辆行进开销预测的模板。然而,这里的时间序列数据仅仅为已观察到的历史数值(即聚类的历史车辆行进开销序列),本实施例还需要通过时间序列预测方法来预测这个时间序列数据在未来时间内的值,即步骤S105。

[0061] 在步骤S105中,本实施例采用差分自回归移动平均模型(ARIMA)进行时间序列数据的预测。ARIMA时间序列预测模型是一个经典的著名时间序列预测方法。这个预测模型中ARIMA(p,d,q)称为差分自回归移动平均模型,其中,AR是自回归,p为自回归项;MA为移动平均,q为移动平均项数,d为时间序列成为平稳时所做的差分次数。所谓ARIMA模型,是指将非平稳时间序列转化为平稳时间序列,然后将因变量仅对它的滞后值以及随机误差项的现值和滞后值进行回归所建立的模型。ARIMA模型根据原序列是否平稳以及回归中所含部分的不同,包括移动平均过程(MA)、自回归过程(AR)、自回归移动平均过程(ARMA)以及ARIMA过程。此预测模型的基本思想为将预测对象随时间推移而形成的数据序列视为一个随机序列,用数学模型来近似拟合这个时间序列数据。这个模型一旦被识别后就可以从时间序列的过去值及现在值来预测未来值。因此,将步骤S104所得的聚类的历史车辆行进开销序列这个时间序列数据输入ARIMA预测模型中,即可根据已有时间序列数据得到未来若干连续

时间片内的车辆行进开销预测值(即聚类的未来车辆行进开销序列)。所述聚类的历史车辆行进开销序列和聚类的未来车辆行进开销序列构成聚类的完整车辆行进开销序列。

[0062] 至此,对于所有的聚类,通过步骤S104~S105可以得到每个聚类中的完整车辆行进开销平均变化模式(即聚类的完整车辆行进开销序列),包括历史数据中观测到的变化模式(聚类的历史车辆行进开销序列)与通过时间序列预测模型得出的预测变化模式(聚类的未来车辆行进开销序列)。这个完整的车辆行进开销平均变化模式可以作为这个聚类中的所有预测路段的车辆开销预测的模板,帮助不同预测路段完成车辆行进开销预测,即进行步骤S106。

[0063] 在步骤S106中,由于通过步骤S105已经得到每个聚类的完整车辆行进开销平均变化模式。同时,对于此聚类中的每一个预测路段,通过计算交通数据可以得到该预测路段部分时间切片内的车辆行进开销值(即每个时间切片内的车辆行进速度的平均值)。至此,拥有两个时间序列数据,即通过步骤S105获得的(1)聚类的完整车辆行进开销序列和(2)预测路段的部分时间切片的车辆行进开销序列。前者是完整的时间序列数据,而后者由于交通数据的稀疏特性,只有少部分的观测值,而大部分的时间片内的车辆行进开销值,包括未来时间片内的车辆行进开销值均为缺失。根据同一个聚类中预测路段的车辆行进开销变化模式相似的原则,所述两个时间序列数据的相关度应该较高。据此,可以填充预测路段缺失的车辆行进开销值,包括此预测路段未来时间片内的车辆行进开销值,以使得此预测路段的车辆行进开销变化模式的时间序列与此预测路段所在的聚类的完整的车辆行进开销平均变化模式有着尽可能高的相关性。

[0064] 结合图3,步骤S106具体包括步骤:

[0065] S1061、定义所述聚类的完整车辆行进开销序列与该聚类中的任一预测路段的部分时间切片的车辆行进开销序列的距离如公式(4)所示:

$$[0066] \quad Distance(CP, R) = \frac{\sum_{R.tc} (R.tc - CP.correspondingTc)^2}{N} \quad \text{公式(4)}$$

[0067] 其中,CP表示聚类的完整车辆行进开销序列,R表示任一预测路段的部分时间切片的车辆行进开销序列,R.tc表示该预测路段已拥有的部分时间切片的车辆行进开销值,CP.correspondingTc表示聚类完整车辆行进开销时间序列中对应于R.tc同一时间片段中的车辆行进开销值;

[0068] S1062、将聚类的完整车辆行进开销序列的数据整体放大/缩小 α 倍,从而使得公式(4)中的距离Distance(CP,R)最小;其中, α 为正实数;

[0069] S1063、将缩放后的所述聚类的完整车辆行进开销序列中的对应数值填补到所述预测路段的部分时间切片的车辆行进开销序列的缺失值处,填充完整后形成预测路段的车辆行进开销序列,并使该预测路段的车辆行进开销序列与所述聚类的完整车辆行进开销序列的相关性最高;

[0070] S1064、对于填充完整后的所述预测路段的车辆行进开销序列,提取填充在该预测路段的未来时间切片内的车辆行进开销序列作为该预测路段的未来车辆行进开销预测值。

[0071] 为了通过填充路段缺失的车辆行进开销值以使得两个时间序列(即聚类的完整车辆行进开销序列和预测路段的部分时间切片的车辆行进开销序列)有尽可能高的相关性,本实施例首先定义预测路段的部分时间切片的车辆行进开销序列与聚类的完整车辆行进

开销序列的距离如公式(4)所示。然后,将聚类的完整车辆行进开销序列数据整体地放大/缩小 α 倍,其中 α 为正实数。根据公式4计算缩放后的聚类的完整车辆行进开销序列与预测路段的部分时间切片的车辆行进开销序列的距离,从而得出最合适的 α 值使得两时间序列的距离Distance(CP,R)最小。此时,将缩放后的聚类的完整车辆行进开销序列中的对应数值填补到预测路段的部分时间切片的车辆行进开销序列的缺失值处以完成缺失值填充。填充之后,两时间序列可以保证差异最小,也满足了最初规定的同一路段聚类中路段车辆行进开销变化模式相似的出发点。当完成预测路段的部分时间切片的车辆行进开销序列的缺失值填充的时候,我们选取填充在该预测路段的未来时间片内的车辆行进开销值作为该预测路段的未来车辆行进开销预测值。同理,可以对所有聚类中的所有预测路段完成未来车辆行进开销预测。

[0072] 至此,完成了预测路段的车辆行进开销预测的解决方案。整个解决方案开始于对各个预测路段进行特征提取,之后基于相似预测路段性质相似,有相似的车辆行进开销模式的思想来克服交通数据稀疏的特性,从而完成对所有预测路段进行车辆行进开销的预测。

[0073] 可见,在本实施例中,通过各个交通车辆信息点的具体信息来划预测路段,并寻找分类相似的预测路段,然后基于同一个聚类中的所有预测路段有着相似的特征的原则,通过计算聚类的完整车辆行进开销序列模式后来对该聚类的任一预测路段的部分时间切片的车辆行进开销序列进行相关性完整填充,从而使填充在任一预测路段的未来时间切片内的车辆行进开销序列作为该预测路段的未来车辆行进开销预测值。因此,本发明通过找寻相似路段,并且认为相似路段的车辆行进开销的变化模式相似来进行车辆行进开销的预测。通过这一方法,可以克服先前工作没有解决的数据稀疏问题,从而可以完整地对于任意路段都可以完成未来车辆行进开销的预测。

[0074] 实施例2

[0075] 参见图4,是本发明实施例提供的一种车辆行进开销的预测系统的结构示意图,包括:

[0076] 数据输入单元201,用于接收包括交通数据和地图数据的数据;

[0077] 其中,所述交通数据至少包括车辆不同时间点的当前时间信息、位置经度信息和行进速度信息,所述地图数据包括路段的起始终止经纬度信息。

[0078] 预测路段分化单元202,用于将接收到的交通数据和地图数据分化到各自相应的预测路段内;

[0079] 预测路段特征值提取单元203,基于每一所述预测路段内的交通数据和地图数据,提取其对应的特征值;

[0080] 预测路段聚类单元204,基于所述特征值将每一所述预测路段分化到不同的聚类中,使得同一个聚类中的所有预测路段有着相似的特征;

[0081] 聚类的历史车辆行进开销序列计算单元205,将同一个聚类的交通数据和地图数据分化到其对应的过去连续的多个特定时间切片内,并计算出每个时间切片内的平均车辆行进开销值,从而得到按照时间先后次序排列的聚类的历史车辆行进开销序列;

[0082] 聚类的未来车辆行进开销序列预测单元206,利用时间序列预测方法得到其对应的多个未来时间切片的聚类的未来车辆行进开销序列,所述聚类的历史车辆行进开销序列

和聚类的未来车辆行进开销序列构成聚类的完整车辆行进开销序列；

[0083] 预测路段的行进开销预测单元207,利用所述聚类的完整车辆行进开销序列对该聚类的任一预测路段的部分时间切片的车辆行进开销序列进行相关性完整填充,其中填充在任一预测路段的未来时间切片内的车辆行进开销序列作为该预测路段的未来车辆行进开销预测值。

[0084] 在预测路段分化单元202中,基于数据输入单元201输入的交通数据与地图数据进行路段的时空分割,同时由于交通数据中有车辆的位置信息和当前时间信息,所以各个数据点可以根据此类时空信息分化到其应该属于的相关预测路段中。

[0085] 其中,每一所述预测路段可以由连续的多个属于此路段的时空分割路段表示,所述时空分割路段定义为给定时间段下的特定长度路段,且所述给定时间段与所述特定时间切片一致。优选的,所述预测路段与所述特定长度路段的路段长度相同。

[0086] “时空分割路段”即是在给定时间段下的特定长度路段,在本实施例中,这里的时间段长度默认为三十分钟。定义在这个短暂时间段中一个路段的车辆行进开销保持不变或者变化微小可以忽略。

[0087] 由于定义的时空分割路段是将一个路段在时间轴上按照三十分钟为一个时间片进行切片,因此,本实施例的预测路段可以由连续的四十八个属于此路段的时空分割路段表示,这样可以表现出完整一天的预测路段。于是,预测路段的特征可以由所有完整一天的时空分割路段的特征并集来表示。

[0088] 在预测路段特征值提取单元203中,本实施例为每一个预测路段中的每一个时空分割路段提取特征值以对其进行描述。在实施例中,每一个时空分割路段提取的特征值包括为静态特征和动态特征两类。其中,静态特征主要用于描述时空分割路段及其附近时空分割路段的路网的结构,这些特征在长期是相对稳定不变的特征,被归为静态特征一类。而动态特征旨在提取该时空分割路段的动态信息,主要包括该时空分割路段及其近邻时空分割路段的车辆平均行进速度的相关信息。该类信息相关于实时的车辆行进情况,会不断地发生变化,被归为动态特征一类。对于每一时空分割路段提取的完整特征其及相关描述总结与表1中:

[0089]

特征	描述
T_{rid}	时空分割路段类型
L_{rid}	时空分割路段长度
$E_t(rid)$	时空分割路段的近邻路段平均类型
$E_l(rid)$	时空分割路段的近邻路段平均长度
$V_t(rid)$	时空分割路段的近邻路段类型方差
$V_l(rid)$	时空分割路段的近邻路段长度方差
N_{rid}	时空分割路段的近邻路段数量
$E_i(sid)$	时空分割路段的近邻时空分割路段第i个时间段平均行进速度
$V_i(sid)$	时空分割路段的近邻时空分割路段第i个时间段行进速度方差

[0090] 表1

[0091] 表1中,前七个特征用于描述路网信息,为静态特征。最后两个特征用于描述动态

车辆行进情况,属于动态特征。在地图数据中,路段均会根据其道路等级进行编号标示,由此数字编号可以得到路段的类型信息。对于 $E_i(sid)$ 与 $V_i(sid)$, $i=0$ 使得两个特征分别表示此时空分割路段的近邻路段在同一时间片内的平均车辆行进速度与车辆行进速度方差。而 $i=1$ 与 $i=-1$ 分别表示下一个时间片内与上一个时间片内相关的特征值,依此类推。

[0092] 提取每个预测路段的特征值后,则需要基于所述特征值将每一所述预测路段分化到不同的聚类中。

[0093] 具体的,结合图5,预测路段聚类单元204具体包括:

[0094] 归一化模块2041,用于将每一所述预测路段的特征值进行归一化处理,从而减少各个特征值的大小差距;

[0095] K-Means聚类模块2042,基于经过归一化处理后的特征值并通过K-Means的聚类方法将每一所述预测路段分化到不同的聚类中,从而使得同一个聚类中的所有预测路段有着相似的特征。

[0096] 其中,在归一化模块2041中,通过公式(1)对所述每一预测路段的特征值进行归一化处理:

$$[0097] \quad F_{norm} = \frac{F_{origin}}{F_{max}} \text{ 公式(1)}$$

[0098] 其中, F_{origin} 是每个预测路段的初始特征值, F_{max} 是此类特征的最大值, F_{norm} 为归一化后的特征值,因此归一化后的特征值均处于零到一之间。经过归一化处理,可以使得各个特征值的大小差距较归一化之前变小,进而使得此后在计算特征相似度时各个特征值的贡献会近似相同,从而避免因为个别特征值的较大贡献使得特征相似度计算失真。

[0099] 对每个预测路段的特征值进行归一化处理,通过K-Means聚类模块2042并使用K-Means聚类方法进行预测路段聚类,具体包括过程:

[0100] (1)、初始化定义K个聚类中心;

[0101] (2)、根据公式(2)中定义的预测路段与聚类中心的距离,将各个预测路段归类到其距离最近的聚类中心处:

$$[0102] \quad Dis_{ac} = \sqrt{\sum_{i=1}^n (f_i^a - f_i^c)^2} \text{ 公式(2)}$$

[0103] 其中, Dis_{ac} 表示任一预测路段到任一聚类中心的距离, f_i^a 代表该预测路段的第i个特征值,而 f_i^c 代表该聚类中心的第i个特征值;

[0104] 由此特征值相似度计算公式(2)可以看出,如果各个特征值的有效范围差距较大,则数值大的特征值可能会较大程度地定义特征值相似性而较小的特征值的相似性会被模糊,使得整个样本相似度计算失真。由此体现出公式(1)中特征值归一化处理的必要性。这使得各个特征值在计算样本相似度时会有着相似的贡献,使得样本相似度可以真实地反映所有特征值的相似度。

[0105] (3)、根据每个聚类中的预测路段的平均特征值更新该聚类的中心,直至预测路段的平均特征值与聚类中心完全重合无法更新为止。

[0106] 在该步骤中,根据每个聚类中的预测路段的平均特征值更新此聚类的中心。如此上述将预测路段归类并更新聚类中心的过程会不断进行直到收敛,即预测路段的平均特征值与聚类中心完全重合无法更新为不同的聚类中心。这样的聚类过程,可以根据各个预测路段的特征值将它们分化到不同的聚类中,使得同一个聚类中的预测路段有着相似的特征。在本发明中,定义在这样的每一个聚类中,特征相似的预测路段会有着相似的性质,从而有着相似的车辆行进开销随时间变化模式。本发明正是根据这个前提从而克服交通数据稀疏的特性,完成对预测路段进行车辆行进开销预测。

[0107] 此后,在完成所有预测路段的聚类后,则通过聚类的历史车辆行进开销序列计算单元205和聚类的未来车辆行进开销序列预测单元206进行“聚类的车辆行进开销模式预测”。这里的车辆行进开销模式预测针对每一个不同的聚类展开。对于每一个聚类,相应的会有大量交通数据归属于此聚类中的各个预测路段中。在聚类的历史车辆行进开销序列计算单元205中,首先,本实施例按照三十分钟为一个单位将时间轴切片,并且将所有的该聚类中的交通数据分化到其对应的时间切片中,计算出每个时间切片中的平均车辆行进开销值(即每个聚类的每个时间切片内的所有预测路段的平均车辆行进开销值,该平均车辆行进开销值可以通过计算每个聚类的每个时间切片内的所有预测路段的车辆行进速度的平均值获得)。然后,将这些平均车辆行进开销值按照时间先后次序排列从而形成一个时间序列数据。这个时间序列反映了过去时间中,这个路段聚类中车辆行进开销的平均变化模式。由于认为此路段聚类中各路段特征相似,其车辆行进开销变化模式也相似,因此,得出的此聚类的平均车辆行进开销变化模式的时间序列数据可以代表整个聚类中路段的模式变化情况。从而,此时间序列数据可以作为此聚类中各路段车辆行进开销预测的模板。然而,这里的时间序列数据仅仅为已观察到的历史数值(即聚类的历史车辆行进开销序列),本实施例还需要通过聚类的未来车辆行进开销序列预测单元206并通过时间序列预测方法来预测这个时间序列数据在未来时间内的值。

[0108] 在聚类的未来车辆行进开销序列预测单元206中,采用差分自回归移动平均模型(ARIMA)进行时间序列数据的预测。ARIMA时间序列预测模型是一个经典的著名时间序列预测方法。这个预测模型中ARIMA(p,d,q)称为差分自回归移动平均模型,其中,AR是自回归,p为自回归项;MA为移动平均,q为移动平均项数,d为时间序列成为平稳时所做的差分次数。所谓ARIMA模型,是指将非平稳时间序列转化为平稳时间序列,然后将因变量仅对它的滞后值以及随机误差项的现值和滞后值进行回归所建立的模型。ARIMA模型根据原序列是否平稳以及回归中所含部分的不同,包括移动平均过程(MA)、自回归过程(AR)、自回归移动平均过程(ARMA)以及ARIMA过程。此预测模型的基本思想为将预测对象随时间推移而形成的数据序列视为一个随机序列,用数学模型来近似拟合这个时间序列数据。这个模型一旦被识别后就可以从时间序列的过去值及现在值来预测未来值。因此,将步骤S104所得的聚类的历史车辆行进开销序列这个时间序列数据输入ARIMA预测模型中,即可根据已有时间序列数据得到未来若干连续时间片内的车辆行进开销预测值(即聚类的未来车辆行进开销序列)。所述聚类的历史车辆行进开销序列和聚类的未来车辆行进开销序列构成聚类的完整车辆行进开销序列。

[0109] 至此,对于所有的聚类,通过聚类的历史车辆行进开销序列计算单元205和聚类的未来车辆行进开销序列预测单元206可以得到每个聚类中的完整车辆行进开销平均变化模

式(即聚类的完整车辆行进开销序列),包括历史数据中观测到的变化模式(聚类的历史车辆行进开销序列)与通过时间序列预测模型得出的预测变化模式(聚类的未来车辆行进开销序列)。这个完整的车辆行进开销平均变化模式可以作为这个聚类中的所有预测路段的车辆开销预测的模板,帮助不同预测路段完成车辆行进开销预测。

[0110] 在预测路段的行进开销预测单元207中,由于已经得到每个聚类的完整车辆行进开销平均变化模式。同时,对于此聚类中的每一个预测路段,通过计算交通数据可以得到该预测路段部分时间切片内的车辆行进开销值(即每个时间切片内的车辆行进速度的平均值)。至此,拥有两个时间序列数据,即(1)聚类的完整车辆行进开销序列和(2)预测路段的部分时间切片的车辆行进开销序列。前者是完整的时间序列数据,而后者由于交通数据的稀疏特性,只有少部分的观测值,而大部分的时间片内的车辆行进开销值,包括未来时间片内的车辆行进开销值均为缺失。根据同一个聚类中预测路段的车辆行进开销变化模式相似的原则,所述两个时间序列数据的相关度应该较高。据此,可以填充预测路段缺失的车辆行进开销值,包括此预测路段未来时间片内的车辆行进开销值,以使得此预测路段的车辆行进开销变化模式的时间序列与此预测路段所在的聚类的完整的车辆行进开销平均变化模式有着尽可能高的相关性。

[0111] 具体的,预测路段的行进开销预测单元207通过以下具体步骤完成预测路段的车辆行进开销变化模式:

[0112] (1)、定义所述聚类的完整车辆行进开销序列与该聚类中的任一预测路段的部分时间切片的车辆行进开销序列的距离如公式(4)所示:

$$[0113] \quad Distance(CP,R) = \frac{\sum_{R.tc} (R.tc - CP.correspondingTc)^2}{N} \quad \text{公式(4)}$$

[0114] 其中,CP表示聚类的完整车辆行进开销序列,R表示任一预测路段的部分时间切片的车辆行进开销序列,R.tc表示该预测路段已拥有的部分时间切片的车辆行进开销值,CP.correspondingTc表示聚类完整车辆行进开销时间序列中对应于R.tc同一时间片段中的车辆行进开销值;

[0115] (2)、将聚类的完整车辆行进开销序列的数据整体放大/缩小 α 倍,从而使得公式(4)中的距离Distance(CP,R)最小;其中, α 为正实数;

[0116] (3)、将缩放后的所述聚类的完整车辆行进开销序列中的对应数值填补到所述预测路段的部分时间切片的车辆行进开销序列的缺失值处,填充完整后形成预测路段的车辆行进开销序列,并使该预测路段的车辆行进开销序列与所述聚类的完整车辆行进开销序列的相关性最高;

[0117] (4)、对于填充完整后的所述预测路段的车辆行进开销序列,提取填充在该预测路段的未来时间切片内的车辆行进开销序列作为该预测路段的未来车辆行进开销预测值。

[0118] 在预测路段的行进开销预测单元207中,为了通过填充路段缺失的车辆行进开销值以使得两个时间序列(即聚类的完整车辆行进开销序列和预测路段的部分时间切片的车辆行进开销序列)有尽可能高的相关性,本实施例首先定义预测路段的部分时间切片的车辆行进开销序列与聚类的完整车辆行进开销序列的距离如公式(4)所示。然后,将聚类的完整车辆行进开销序列数据整体地放大/缩小 α 倍,其中 α 为正实数。根据公式4计算缩放后的聚类的完整车辆行进开销序列与预测路段的部分时间切片的车辆行进开销序列的距离,从

而得出最合适的 α 值使得两时间序列的距离Distance(CP,R)最小。此时,将缩放后的聚类的完整车辆行进开销序列中的对应数值填补到预测路段的部分时间切片的车辆行进开销序列的缺失值处以完成缺失值填充。填充之后,两时间序列可以保证差异最小,也满足了最初规定的同一路段聚类中路段车辆行进开销变化模式相似的出发点。当完成预测路段的部分时间切片的车辆行进开销序列的缺失值填充的时候,我们选取填充在该预测路段的未来时间片内的车辆行进开销值作为该预测路段的未来车辆行进开销预测值。同理,可以对所有聚类中的所有预测路段完成未来车辆行进开销预测。

[0119] 至此,系统完成了预测路段的车辆行进开销预测。整个系统开始于对各个预测路段进行特征提取,之后基于相似预测路段性质相似,有相似的车辆行进开销模式的思想来克服交通数据稀疏的特性,从而完成对所有预测路段进行车辆行进开销的预测。

[0120] 可见,在本实施例中,系统通过各个交通车辆信息点的具体信息来划预测路段,并寻找分类相似的预测路段,然后基于同一个聚类中的所有预测路段有着相似的特征的原则,通过计算聚类的完整车辆行进开销序列模式后来对该聚类的任一预测路段的部分时间切片的车辆行进开销序列进行相关性完整填充,从而使填充在任一预测路段的未来时间切片内的车辆行进开销序列作为该预测路段的未来车辆行进开销预测值。因此,本发明通过找寻相似路段,并且认为相似路段的车辆行进开销的变化模式相似来进行车辆行进开销的预测。从而可以克服先前工作没有解决的数据稀疏问题,从而可以完整地对于任意路段都可以完成未来车辆行进开销的预测。

[0121] 另外,本发明技术方案除了可以应用到上述的车辆行进开销的预测外,还可以用于其它类似路网结构中行进开销的预测,例如网络结构中不同网络路径中数据传输开销的预测。另一方面,本技术方案也可以用于城市中对于路段除车辆行进开销外其他属性值的预测,例如对于路段拥塞状况,车辆吞吐量的预测。

[0122] 以上所述是本发明的优选实施方式,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的前提下,还可以做出若干改进和润饰,这些改进和润饰也视为本发明的保护范围。

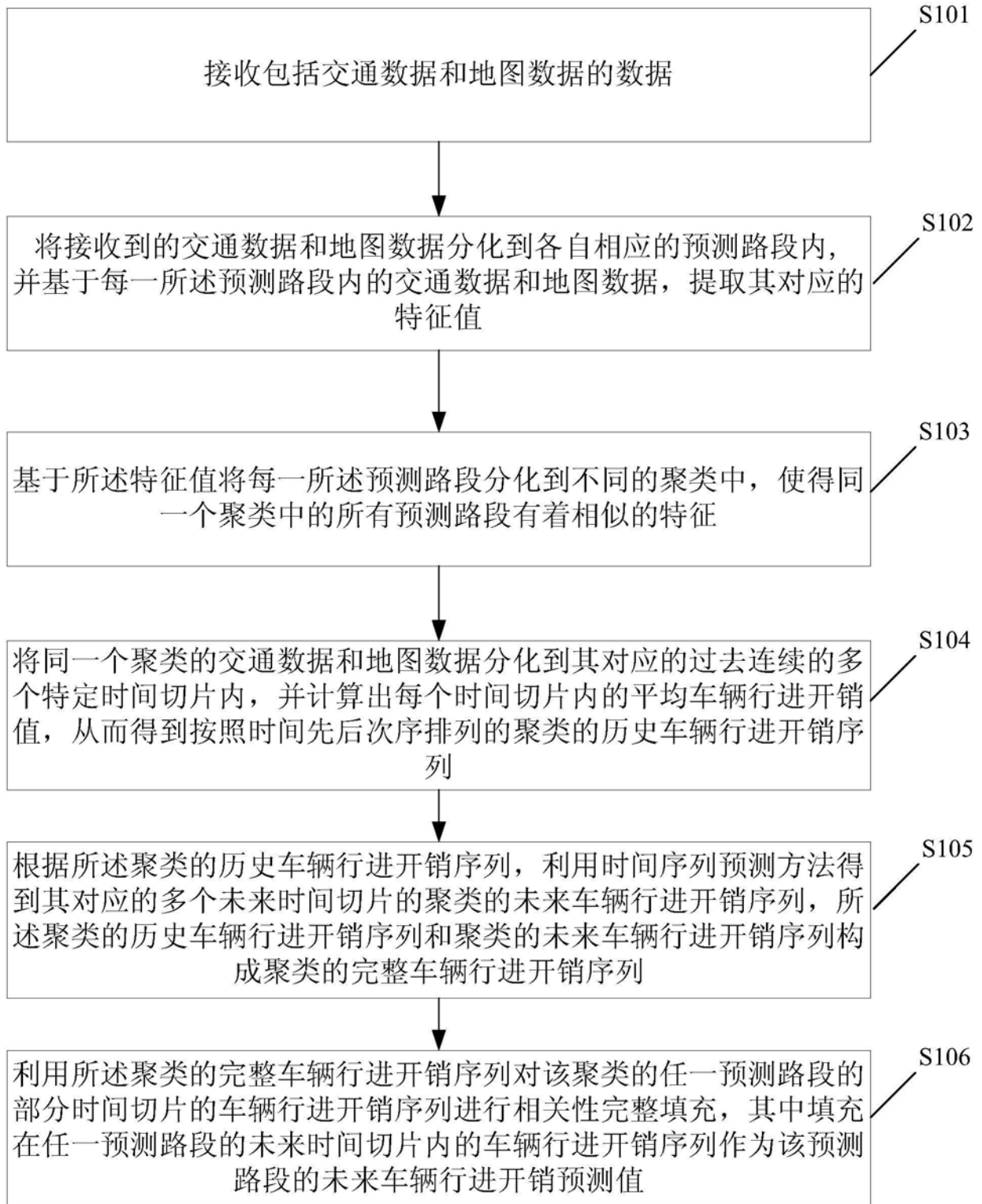


图1

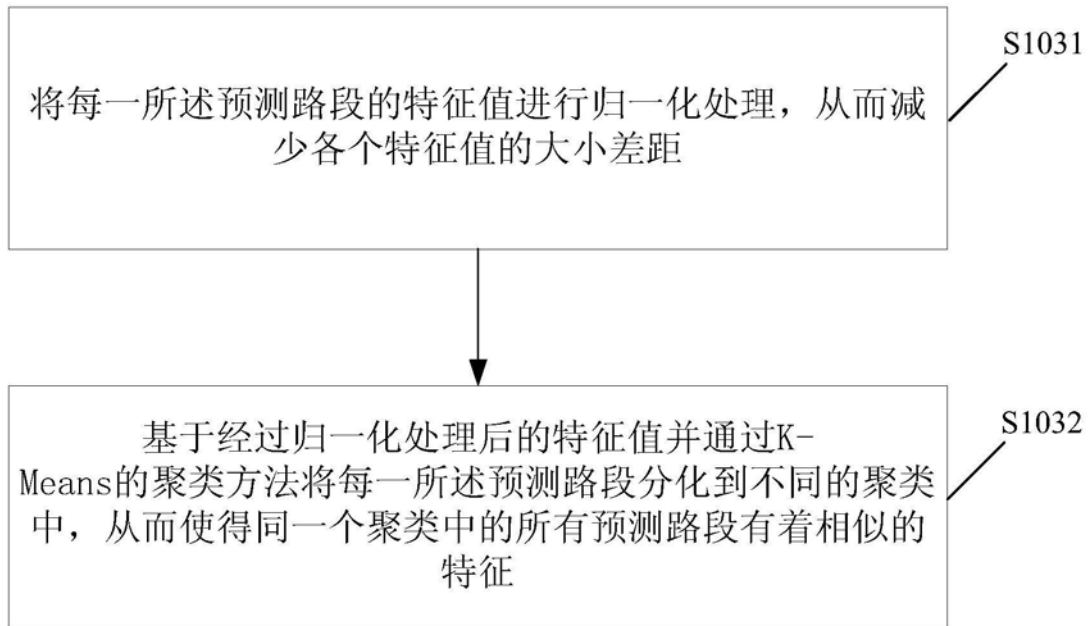


图2

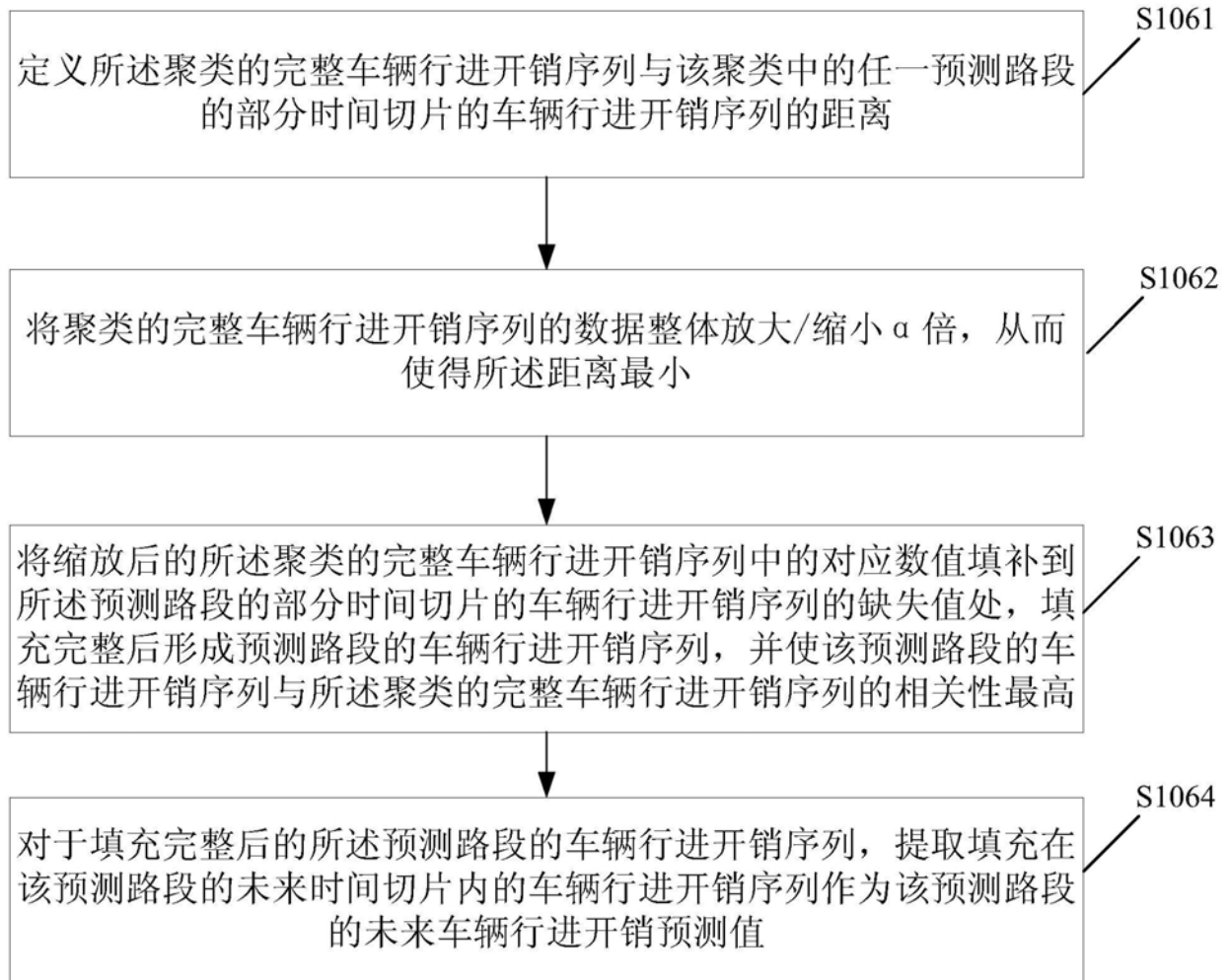


图3

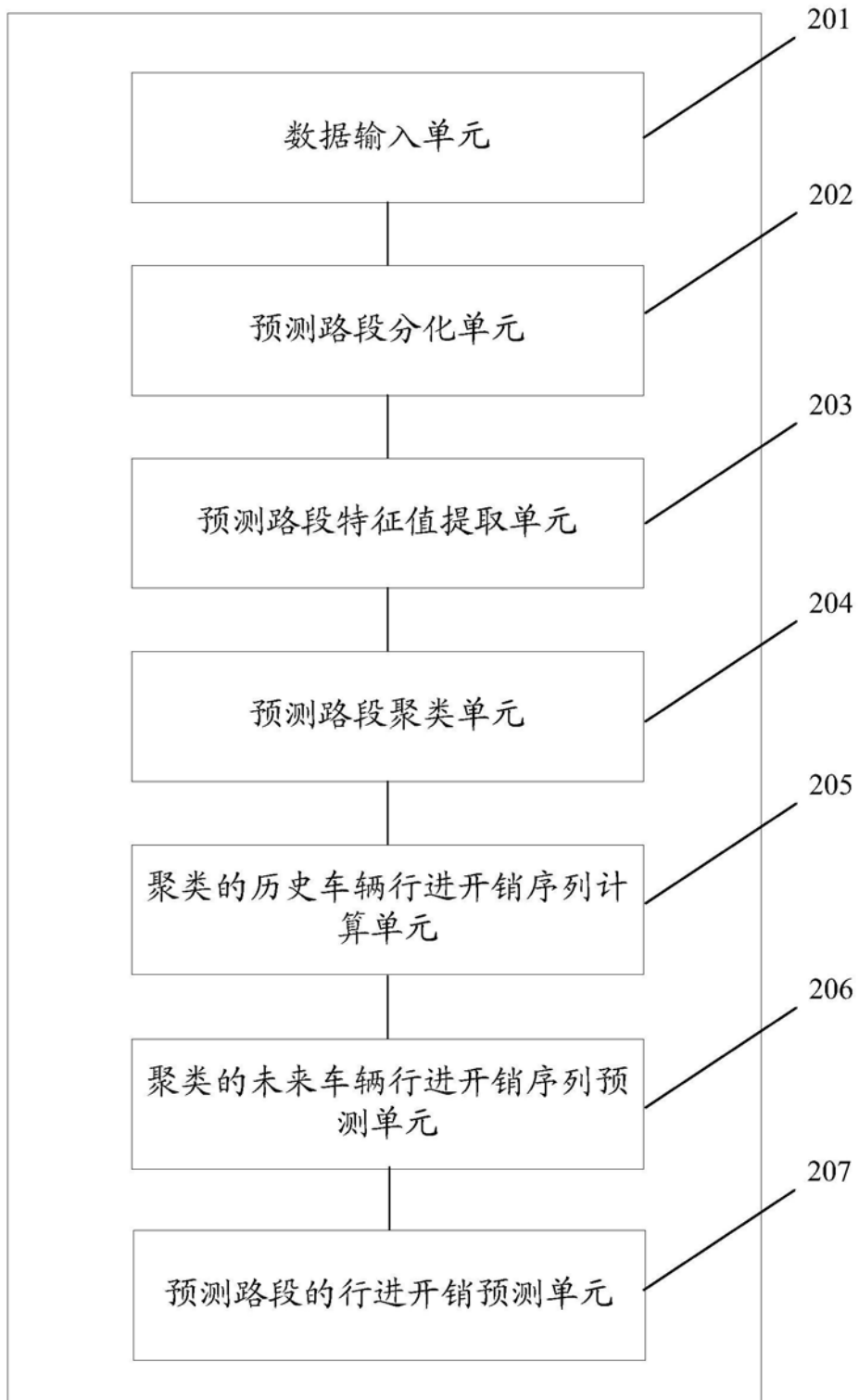


图4

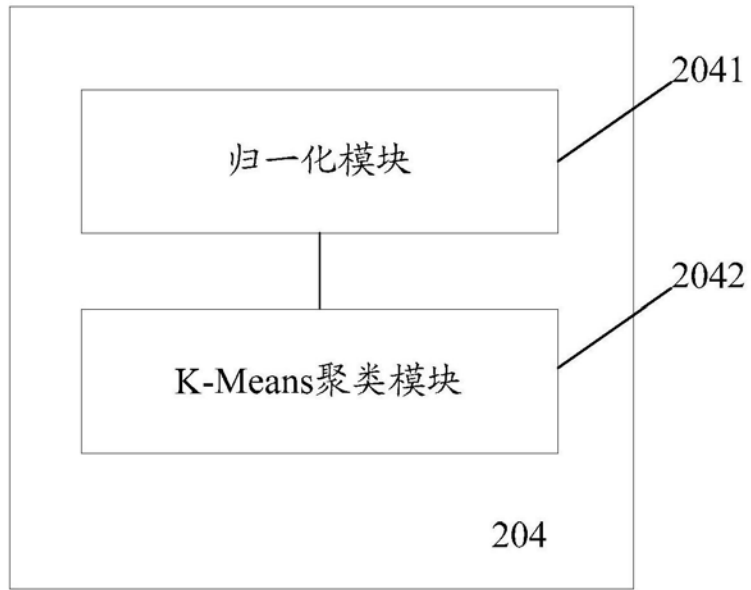


图5