

文章编号: 1002-0268 (2006) 04-0115-05

基于环形线圈检测器采集信息的交通 状态分类方法应用研究

皮晓亮¹, 王 正¹, 韩 皓¹, 孙 亚²

(1. 上海海事大学 交通运输学院, 上海 200135; 2. 同济大学 ITS 研究中心, 上海 200092)

摘要: 在智能交通系统中, 交通状态判别算法通常被用来进行道路环境中实时交通状态的判断。这些算法将外场设备采集到的实时交通流数据与既有的交通状态分类标准特征作比较, 来识别交通系统运行的状态。应用聚类分析方法, 结合数据预备技术和交通工程技术, 对环形线圈监测系统采集的交通流基础特征数据进行挖掘, 实现了一种交通状态分类方法, 并对交通管理控制系统中实时交通状态的判断识别提供可靠的参照标准。

关键词: 聚类分析; 交通状态; 环形线圈检测; 数据挖掘

中图分类号: U491.1⁺15

文献标识码: A

Application Research of Traffic State Classification Method Based on Collected Information from Loop Detector

PI Xiao-liang¹, WANG Zheng¹, HAN Hao¹, SUN Ya²

(1. Shanghai Maritime University, Shanghai 200135, China;

2. ITS Research Center of Traffic Engineering Department, Tongji University, Shanghai 200092, China)

Abstract: In an intelligent transportation management system, every kind of algorithms are usually used to recognize real-time traffic state. Traffic flow data will be often compared with a given traffic state classification standard, thus the real-time traffic state will be detected. With the clustering analysis method, data preparation technology and traffic engineering technology, this article focuses on application research of traffic state classification method based on collected information from loop detector system. This traffic state classification method provides a nice reference to real time traffic state recognition of the transportation management and control.

Key words: Traffic state; Loop detector system; Data mining; Clustering analysis method

0 引言

在智能交通管理系统中, 各种交通状态判别算法通常被用来进行道路环境中实时交通状态的判断。应用这些算法, 可以对道路基础信息采集系统采集到的各类交通数据进行分析, 根据他们的变化趋势得出目前交通系统的运行状况。通常来说, 一个交通状态判别系统包括数据获取、数据准备、数据规约、数据转

换、数据比较、实时状态判断等过程。在这个过程中, 系统将采集到的基础数据经过一定处理后与一个既定的交通状态判断标准进行比较, 从而判别出目前交通系统处于何种状态。这个判别的结果将提供交通系统的管理者和决策者, 他们针对不同的情况作出相应的交通控制、管理和诱导措施。

从 20 世纪 70 年代早期, 关于道路实时交通状态判别算法的研究逐渐开展起来。这些算法的研究包括

收稿日期: 2005-03-25

基金项目: 国家重点基础研究发展规划 973 资助项目 (TG1998030408)

作者简介: 皮晓亮 (1982-), 女, 河南南阳人, 工学硕士, 现执教于上海海事大学交通运输学院。 (xipi@cet.shmtu.edu.cn)

©1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

比较算法、时间序列分析、McMaster 算法、人工智能方法、Macroscopic 算法、参数标定方法等。

这些研究大部分都属于时空透视法的范畴。在这些交通状态判断方法中,分别应用到了不同的交通状态分类结果做为算法的比较标准。交通状态分类过程是基于采集到的海量交通流基本特征参数。由于交通系统是一个随时间不断变化的时变系统,交通流的基本特征不仅每时每刻都在发生着量变,而且在一个相当长的时期内,不断积累的量变会带来整个交通系统交通流特征一个整体的质变。在交通系统中,交通流特征在短时期内发生的小范围、暂时性的波动不会影响到对系统交通状态分类的描述,但是随着系统缓慢的质变发生后,原有的交通状态描述标准就会与现在的实际运行状况发生较大偏移,如果这时仍然用采集到的实时基础数据与这个错误的状态分类标准做比较,得出的判别结果必定是存在问题的。因此不管使用何种交通状态判别算法来识别实时道路的交通状态,都需要对系统的交通状态分类标准进行定期的修订,从而保证该状态分类标准与实际情况的吻合,为交通状态的判断提供可靠的比较依据。

本文试图利用环形线圈检测系统采集到的基础交通流特征数据,使用聚类分析的方法,实现一种交通状态判断标准的制定和修正。

1 基于聚类分析的交通状态量化方法

聚类方法是适合交通状态数据分类的一种数据挖掘方法。基于聚类分析的交通状态量化方法是根据交通检测器所采集道路上交通流的基础数据为基础,通过对这些数据的分析处理,运用交通工程理论和数据挖掘等技术,从它们的基础交通数据中找出这些道路各自的特点,最后由聚类分析方法,根据每种交通状态其中的样本{流量,速度,占有率}具有一定的相似性和不同状态之间的样本具有相异性,得到适合不同道路的合理交通状态分类。

1.1 聚类分析

所谓聚类分析就是依据交通流数据样本间关联的度量标准将其自动分成几个交通流数据群组。且使同一群组内的样本相似,而属于不同群组的样本相异。K-Means Cluster 聚类分析方法对于大规模基础交通特征数据集的应用占有优势。它包括硬 K-均值算法(Hcm)和模糊 K-均值算法(Fcm)。其中 Hcm 算法是 Fcm 算法的一个特例。

本文对交通流基本参数聚类的目标函数为

$$J_m(U, V) = \sum_{i=1}^n \sum_{j=1}^c u_{ij}^m \|x_i - v_j\|^2$$

其约束条件为

$$\sum_{j=1}^c u_{ij} = 1; u_{ij} \geq 0; c \geq j \geq 1; n \geq i \geq 1$$

式中, n 为环形线圈检测器采集的交通信息集经过数据规约后得到 3 个特征{流量,速度,占有率}的样本个数; c 为聚类中心数,即为 4 种(顺畅交通流,稳定交通流,拥挤交通流,堵塞交通流); u_{ij} 为矩阵的第 i 行第 j 列,代表第 j 个样本对第 i 个聚类中心的隶属程度; $V = \{v_1, v_2, v_3, v_4\}$ 为聚类中心矩阵,代表 4 种交通状态; $\|x_i - v_j\|$ 表示第 j 组数据对于第 i 类聚类中心的距离; m 为模糊指数,目的是加强 x_i 属于各类的从属程度的对比度。

1.2 聚类分析算法的基本步骤

本文使用的 Fcm 算法,基本步骤为:

(1) 随机选择一个二个类的初始划分矩阵 U , 它可以是硬划分,也可以是软划分。然后计算这 4 个类的重心。

(2) 通过把交通流数据样本分配给与其重心距离最近的类生成一个新的分区。

(3) 用类的重心来计算新类的中心距离。

(4) 重复步骤 2 和 3 直到求出目标函数的最优解。

1.3 聚类分析算法的特征数据样本

本文使用的交通流基础数据的特征数据样本如图 1~图 4 所示。

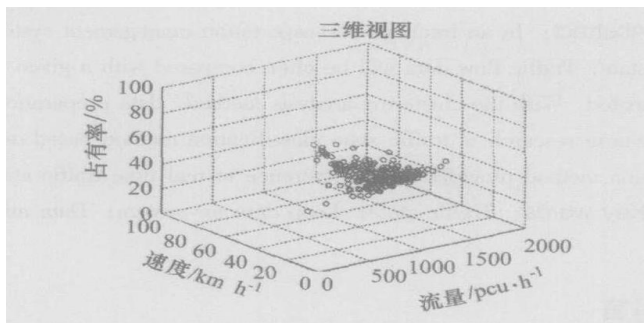


图 1 堵塞流数据样本

Fig 1 Data sample of block flow

2 交通状态量化聚类分析方法实现

基于聚类分析的交通状态量化方法对所采集到的交通基础数据有着较高的要求,所采集的数据要包括所有交通特征流的信息,所以在进行数据分析前所需要的原始数据容量高达 G 级。而且使用的数据库一般为 Oracle、Sybase 等等巨型数据库。基于聚类分析的交通状态量化的一般流程如图 5 所示。

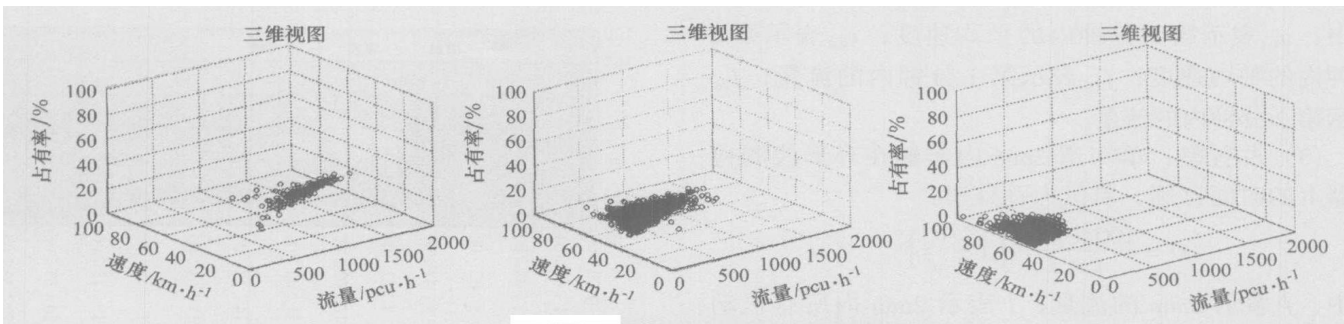


图 2 拥挤流数据样本

图 3 稳定流数据样本

图 4 通畅流数据样本

Fig 2 Data sample of crowd flow

Fig 3 Data sample of stable flow

Fig 4 Data sample of free flow

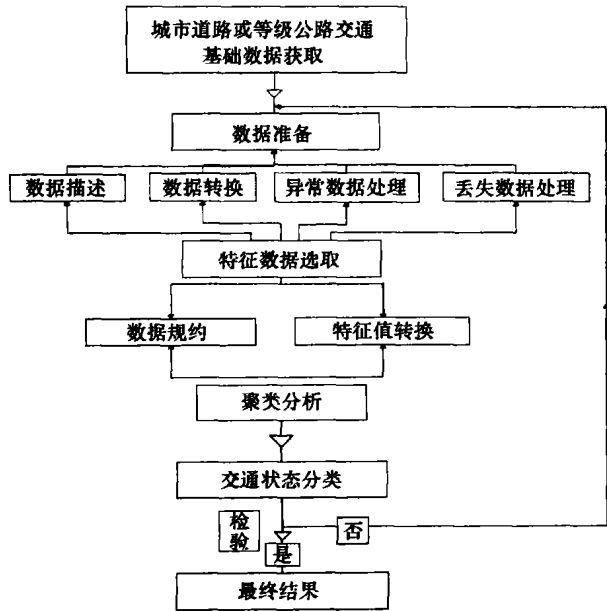


图 5 交通状态量化流程

Fig 5 Process of classify traffic state

2 1 交通流原始数据获取

交通基础数据信息可从相应的交通控制管理中心获得。为了获取特定的历史数据, 使用了特定的结构查询语言 (SQL) 进行访问 Oracle 等大型数据库, 输出的结果保存成 Excel 文件格式, 以便统计分析和处理。

获得的交通基础数据信息为评价范围内路段的环形线圈检测器所检测的数据。由于线圈检测技术发展已经很成熟并且性价比高, 一般情况下, 该检测器是目前交通控制中心应用最广的交通数据检测器。获得交通基础数据的时间范围可根据具体情况而定, 本文使用了近期内 1 个月的历史数据, 以便对目前的交通状态做准确的量化判断。

交通监控中心所存储的环形线圈检测器采集的交通信息为结构化数据, 因为它由定义好的具有数字值或文字数字值的字段组成。这些结构化数据通常以表格的形式来描述的。表格的列是存储在表格中交通数

据的特征, 表格的行则是特殊实体的特征值。

如表 1 所示, 这里仅给出了交通监控中心所存储的环形线圈检测器采集的交通信息的部分字段。环形线圈检测器采集到的交通基础信息: 交通量、车速、占有率等, 以上信息还以分车型 (大、中、小) 统计的, 并且同时还包含了一些其它必要的信息, 如日期、采集设备 ID、设备工作状态等等。

表 1 交通监控中心环形线圈检测器采集的交通信息 (部分字段)

Tab 1 Traffic data from loop detector system of traffic control center (part of the data)

FDT_RECORD DATETIME	FSTR_SDEVICE ID	FSTR_LOOP STATUS1	FINT_LANE FLOW1-V1	FINT_LANE FLOW1	FINT_LANE RATIO1	...
1 04-9-8 0 01	A20SVD10090	正常	3	3	1	...
2 04-9-8 0 02	A20SVD10090	正常	1	1	0	...
3 04-9-8 0 03	A20SVD10090	正常	1	1	0	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮

2 2 交通流基础数据准备

聚类分析方法对要进行处理交通流原始数据有着一定的要求。

1. 交通基础数据对象为单车道内的流量、速度、占有率。

2. 交通基础数据所统计的时间间隔应为 2min 左右 (该文以 2min 为例说明)。时间间隔太短无法准确的把握交通状态的特征, 而且状态特征还易于走向极端现象。所以这里所处理数据时间间隔为 2min。那么 3 大交通参数也要做如下相应的变化:

(1) 车流量: 单车道 2min 内为各车型的车辆数目乘上相应的车型比重后之和。计算方法为: 第 1 分钟内车流量与第 2 分钟内车流量之和, 单位为 pcu/h。

(2) 平均速度: 单车道 2min 内所有机动车的平均速度。计算方法为: 每分钟内车平均速度乘以相应的流量之后的和与 2min 内流量的比值, 单位为 m/h。

$$V = \frac{v_{1st} \times f_{1st} + v_{2st} \times f_{2st}}{f_{1st} + f_{2st}}$$

式中, v_{1st} 表示第 1 分钟内的平均速度; v_{2st} 表示第 2 分钟内的平均速度; f_{1st} 表示第 1 分钟内的流量; f_{2st} 表示第 2 分钟内的流量。

(3) 占有率: 单车道 2min 内车辆在环形线圈检测器上的时间比例, 单位为百分率。

$$O = \frac{100 \times f \times l}{t \times v} \left[1 + \frac{\sigma^2}{v^2} \right]$$

其中, f 表示 2min 的流量; l 表示 2min 内所有机动车辆的车身长度; v 表示 2min 内的平均速度; t 表示时间间隔, 这里为 2min。

对于环形线圈检测器采集的交通信息集, 具有完整数据案例可能性很小。这是因为由于环形线圈检测器处于非正常工作状态、通讯网络中断等原因情况下, 都会导致数据丢失。

本文中所用的数据挖掘方法对数据丢失处理采用去除所有丢失样本, 因为所得交通基础数据容量极大, 所以缺少少量的样本不会对算法产生影响。

另外对于数据样本中的异常敏感数据也需进行处理。所谓异常、敏感数据就是这些样本和其余样本数据有很大不同或不一致。环形线圈检测器采集的交通信息集的异常敏感数据可能是由测量误差造成的, 也可能是数据固有的可变性的结果。

由于数据挖掘聚类方法对异常点具有敏感性, 所以有必要在数据预处理阶段把他们清除。异常数据的处理方法分为两阶段: 异常敏感点的检测和异常敏感点的清除。

本文使用交通工程的相关理论完成异常敏感点的检测, 异常敏感点的清除基本做法就是把异常敏感点当作丢失点进行处理。

由于聚类分析是对所有的特征点进行分析, 因此必须找出这些特征点的共同点和不同点, 才能得到合理的结果。所以该方法要求交通基础数据中, 所有的交通状态的特征点均要出现。

通过对单车道 1 天中的交通基础数据分析, 得到两种典型的规律, 即有拥堵状态现象和无拥堵状态现象。特征数据二维视图表示如图 6。

从图 6 中可以看出, 道路上存在两种典型的交通规律: 有拥堵状态现象和无拥堵状态现象。由于数据挖掘方法——聚类分析对数据的要求, 所以在选取特征数据方面, 要求所有的交通状态的特征点均要出现。

本文交通状态数据提取的范围分为以下部分:

占有率在 [90, 100] 之间、[80, 90] 之间、[70, 80] 之间、[60, 70] 之间、[50, 60] 之间、

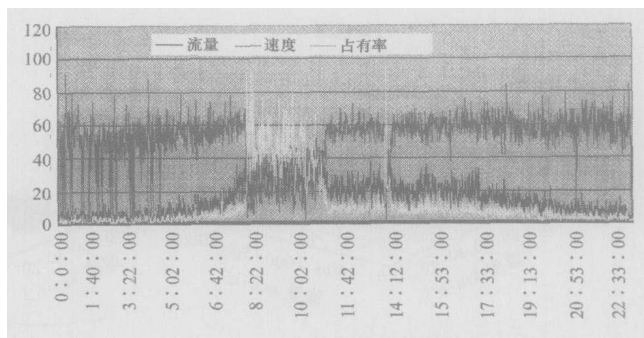


图 6 含有拥堵状态和畅通状态的
3 参数典型分布图

Fig 6 Graph of typical three main parameter including crowd state and stable state

[40, 50] 之间、[30, 40] 之间、[20, 30] 之间、[10, 20] 之间、[0, 10] 之间的, 共有 10 部分的交通基础数据。

交通状态数据提取的要求为: 10 种范围内的数据样本数量相同。交通状态数据处理的方法: 使用统计软件提取和统计。

2.3 原始数据规约

通常许多大型数据集存在数据高维度引起的数据超负荷问题, 这种问题如果得不到适当的数据维度规约, 会使数据挖掘方法不适用或难以得出理想结果。因此通过数据维度规约的工作, 预处理数据集的 3 个主要维度通常以平面文件的形式出现, 即列 (特征), 行 (样本) 和特征的值。环形线圈检测器采集的交通信息原始数据集恰好符合数据挖掘预处理数据集的 3 个主要维度。

2.4 原始数据转换

{流量, 速度, 占有率} 3 个特征值取值范围不同且相差甚远, 例如流量有时可以高达几千 pcu/h, 而占有率相比之下很小。同时在交通基础数据处理过程中发现即使在同一种交通状态下交通量的范围不尽相同, 并且不同路段有时差别很大。

所以本文进行评价的数据集为 {流量, 速度, 占有率}。

在使用数据挖掘方法之前, 应对各特征值之间关系, 权重进行标准化处理。特征值可按比例对应到一个特定的范围, 如 [-1, 1] 或 [0, 1] 等等。本文对流量数据进行了标准化处理, 采用最小-最大标准化方法。

$$c'(i) = (c(i) - \min(c(i))) / (\max(c(i)) - \min(c(i)))$$

该方法可以使流量值在所选区间较好的分布。由于本文中速度和占有率的取值范围相差不大, 其中速

度为 $[0, 150]$, 占有率取值范围为 $[0, 100]$, 所以本挖掘方法不需要再对速度和占有率做数据转换调整。

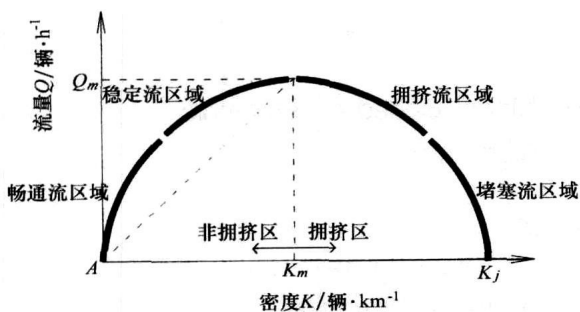
2.5 交通状态聚类分析结果

将以上准备好的数据按照第2节中的聚类分析方法, 对样本进行多次迭代, 得到4种交通状态的聚类中心

$$V = \{v_1, v_2, v_3, v_4\} = \begin{bmatrix} 1 & 260 & 2 & 100 & 1 & 260 & 360 \\ 16 & & 25 & & 55 & & 70 \\ 60 & & 36 & & 11 & & 3 \end{bmatrix}$$

该矩阵第1行代表流量, 单位为 pcu/h , 第2行代表速度, 单位为 mph , 第3行代表占有率, 单位百分率。

图7为交通流特性的3大基本参数中的密度和速度的关系曲线图。其中在点 (K_m, Q_m) 处, 分为拥挤区域和非拥挤区域。速度和占有率的关系曲线与此也大致相同。



图中: Q_m 代表极大流量, K_m 代表最佳密度, K_j 代表堵塞密度

图7 流量-密度关系曲线

Fig 7 Chart of relationship of density and flow

聚类中心矩阵第1列 $[1\ 260, 16, 60]$, 即流量为 $1\ 260\text{pcu/h}$, 占有率为 60% , 速度为 16mph , 在高占有率和低速度状态下, 根据上图可知, 该向量处于拥挤区的下半区域, 所以 $[1\ 260, 16, 60]$ 代表堵塞交通流区域。

同样可以得出, 第2列 $[2\ 100, 25, 36]$ 代表拥挤交通流聚类中心, 第3列 $[1\ 260, 55, 11]$ 代表稳定交通流聚类中心, 第4列 $[360, 70, 3]$ 代表顺畅交通流聚类中心。

3 聚类分析方法评价及展望

该方法主要以道路上采集的交通流和其他交通基础信息为基础, 其中相应的采集方式为环形检测线圈。

本方法通过对数据的处理和融合, 科学地划分了

交通状态的4个等级: 顺畅流, 稳定流, 拥挤流, 堵塞流; 确定了交通状态的表示形式; 建立了交通状态的三维参数表示; 确定了交通状态的表示形式, 最终建立了交通状态的划分标准, 得到了合理的交通状态分类, 从而为实时道路交通状态的判别提供了可靠的参照比较标准。

该状态分类方法的实施对硬件技术和软件环境也有相应的要求和基本运行条件的约束: 相应道路上采集的基础数据要求包括所有可能的交通特征流的信息, 数据挖掘前所需要的原始数据容量要求高达G级。而且使用的数据库一般为Oracle、Sybase等巨型数据库, 才能达到相应的数据处理速度。

该方法可以为实时交通状态自动识别奠定一定的研究基础: 把采集到的实时交通基础信息与该文的交通状态联系起来, 利用模式识别方法, 把交通采集样本和聚类中心相比较, 通过与交通状态分类矩阵 V 进行运算, 即可判别出相应时刻道路交通数据所代表的交通状态, 如图8所示。



图8 以交通状态分类方法为基础的研究方向

Fig 8 Researches based on traffic state classify method

随着我国大力发展智能交通系统 (ITS), 数据挖掘分析是交通发展的必然方向。其中基础交通信息的采集和融合处理, 成为智能交通系统建设和应用的核心内容, 并逐渐成为城市道路交通管理迫切解决的关键问题。本方法通过先进的计算机技术、数据挖掘技术和交通工程技术对路网的交通量化提供了详细的量化过程, 同时也对智能交通系统其他海量信息挖掘的一般过程提供了一些参考价值。

参考文献:

- [1] Han D H Mannila, P Smith Principles of Data Mining [M]. MIT Press, 2001.
- [2] Chao Chen Detecting Errors and Imputing Missing Data for Single Loop Surveillance Systems [C]. TRB Annual Meeting, 2003.
- [3] 韩悦臻. 城市道路交通状态指标体系及判定方法研究 [D]. 长春: 吉林大学, 2003.
- [4] 楼世博, 孙章. 模糊数学 [M]. 科学出版社, 1983.
- [5] Heribert Kirschfink. Basic Tools for Fuzzy Modeling [M]. Germany, 1998.