

# 基于时空权重相关性的交通流大数据预测方法

李欣<sup>1,2,†</sup> 罗庆<sup>1,2</sup> 孟德友<sup>1,2</sup>

1. 河南财经政法大学中原经济区“三化”协调发展河南省协同创新中心, 郑州 450046;  
2. 河南财经政法大学资源与环境学院, 郑州 450046; † E-mail: lixin992319@163.com

**摘要** 将分布式增量大数据聚合方法与交通流数据清洗规则相结合, 可以为交通流预测分析提供更准确可靠的数据源。通过交通流在路网中的相关性分析, 使用多阶路口转弯率构建空间权重矩阵, 完成对 STARIMA 交通流预测模型的改进。实验结果表明, 该方法可以在工作效率及准确程度上满足交通流大数据预测的需求, 为交通诱导信息发布提供依据。

**关键词** 交通流; 大数据; 分布式增量; 路网相关性; STARIMA

**中图分类号** P91

## Traffic Flow-Big Data Forecasting Method Based on Spatial-Temporal Weight Correlation

LI Xin<sup>1,2,†</sup>, LUO Qing<sup>1,2</sup>, MENG Deyou<sup>1,2</sup>

1. Collaborative Innovation Center of Three-Aspect Coordination of Central Plain Economic Region, Henan University of Economics and Law, Zhengzhou 450046; 2. College of Resource and Environment, Henan University of Economics and Law, Zhengzhou 450046;  
† E-mail: lixin992319@163.com

**Abstract** A distributed incremental aggregation method combined with traffic flow data cleansing rules is proposed, and it can provide more accurate and reliable data for traffic flow forecast analysis. Through the correlation analysis of traffic flow in road network, the authors used the multi-allocation of turning rate in the intersection to build the spatial weight matrix, and improved the STARIMA traffic flow forecasting model. The experiment result proves that this method can meet the needs of traffic flow big-data forecasting in the efficiency and accuracy, and provide the basis for the traffic routing information.

**Key words** traffic flow; big-data, distributed incremental; road network correlation; STARIMA

近年来, 不断增长的汽车保有量与道路设施不足的矛盾已成为中国城市化进程中一个日益突出的问题。利用各种传感器提供的交通流数据建立智能化交通管理平台, 使用数据挖掘技术进行交通预测和诱导是解决城市交通问题的有效方法。交通流数据具有海量(volume)、多元(variety)和高速更新(velocity)等特点<sup>[1]</sup>, 加上车辆出行的随机性和复杂性, 必须将大数据技术与数据挖掘技术结合起来, 才能做出精准的交通流预测。

本文将交通流数据清洗规则与大数据聚合流程

相结合, 获取有效的预测数据源; 利用路网相关性, 设计和改进交通流预测模型时空权重矩阵; 使用河南智能交通综合管理平台交通流数据, 进行理论验证分析。结果表明, 该模型预测结果比已有成果更准确, 可以为实际应用提供参考。

## 1 交通流预测研究现状

大数据条件下, 首先需要利用数据聚合技术为交通流预测提供数据基础。2010年, Bose等<sup>[2]</sup>提出一种增量并行数据挖掘方法。2012年, Aghabozorgi

等<sup>[3]</sup>使用增量和降维方法处理用户数据。2013 年, Laptev 等<sup>[4]</sup>通过样本抽样方法减少运算数据量。Zhang 等<sup>[5]</sup>提出基于分布内存的方法, 提高了基于 MapReduce 数据挖掘算法的性能。以上研究的主要思路是依靠抽样、降维、分布、增量方法或较昂贵的分布式硬件来完成大数据的获取和管理。

在交通流预测方面, 已经取得的研究成果主要包括历史平均模型<sup>[6]</sup>、状态空间模型<sup>[7]</sup>、时间序列模型<sup>[8]</sup>以及神经网络模型<sup>[9-10]</sup>等。Yue<sup>[11]</sup>提出基于互相关函数时空依赖性分析的单点交通流预测时空 Kalman 滤波模型。Kamarianakis 等<sup>[12]</sup>使用时空自回归移动平均模型<sup>[13]</sup>(space-time autoregressive integrated moving average, STARIMA)进行交通流预测, 体现了路网相关性在预测中的作用。清华大学、北京交通大学和上海海事大学的学者结合实际应用, 对 STARIMA 模型做了进一步改进<sup>[14-18]</sup>。

以上研究中限制条件较多, 或仅考虑单点交通流预测, 或对时空相关性提出较多限制, 对交通流真实规律的体现不足。本文首先使用分布式增量数据聚合方法, 完成交通流大数据的清洗和统计提取, 从而保证预测分析具备可靠的数据基础; 然后通过

交通路段时空相关性分析, 将  $n$  阶相关性空间权重矩阵与 STARIMA 模型相结合, 在此基础上构建预测模型; 最后按照时间周期进行交通流预测。

## 2 分布式增量交通流大数据聚合方法

### 2.1 交通流大数据聚合流程

本文研究的分布式增量交通流大数据聚合方法, 主要原理是将交通流数据管理节点分为中心节点和分布节点。分布节点在每一时间周期内采集和清洗各种传感器获取的交通流数据, 进行统计整理并传送到中心节点。中心节点完成所有分布节点统计数据的存储和聚合后, 针对历史全集和增量阶段交通流数据进行基于时空权重相关性的交通流预测。图 1 展示分布式增量交通流大数据聚合原理。

数据聚合过程可以划分为多个时间周期。第一阶段是系统初次运行阶段, 称做历史全集数据聚合阶段, 需要对所有分布节点的历史全集数据执行聚合运算, 数据经过清洗整合到中心节点后, 利用预测算法进行分析; 第二阶段是按照时间间隔划分的周期增量数据聚合阶段, 对系统运行新一时间周期内产生的增量数据, 执行聚合运算, 随着增量数据

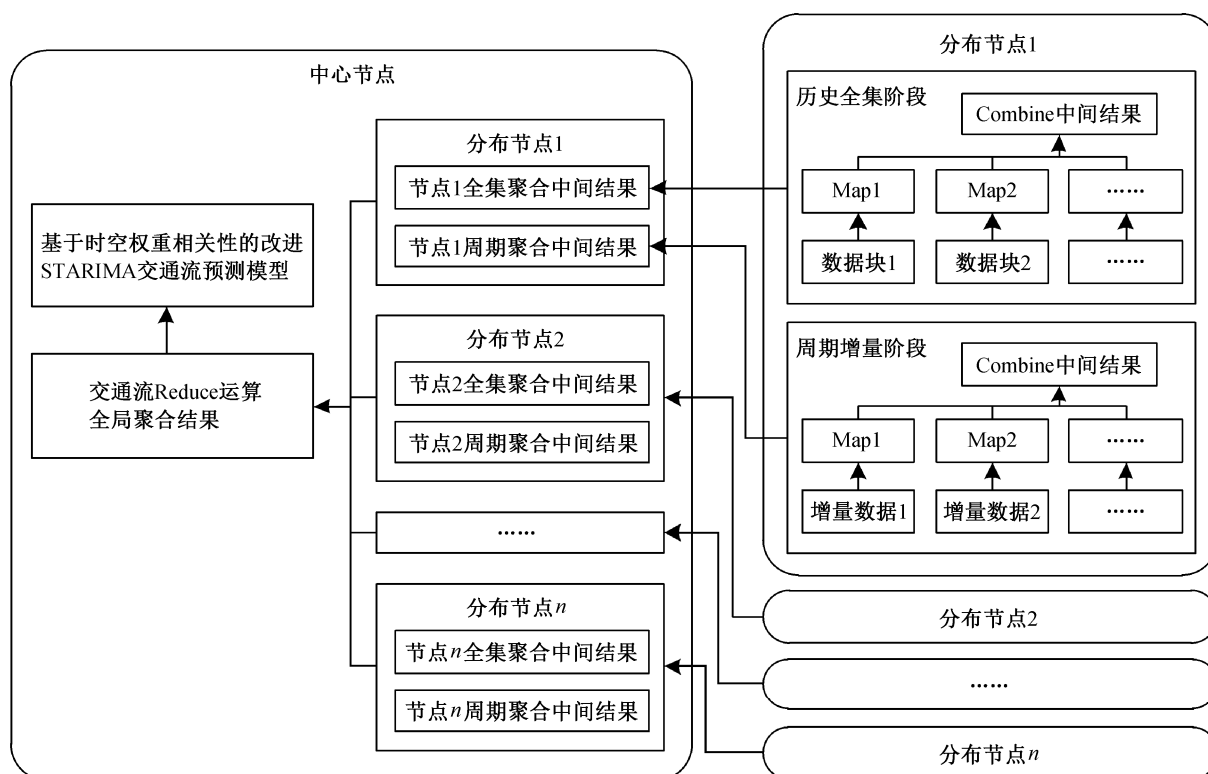


图 1 分布式增量交通流大数据聚合原理

Fig. 1 Distributed incremental aggregation principle of traffic flow big-data

的累积,可以对之前的预测结果进行修正。聚合和预测过程基于 MapReduce 模型运行,分布节点主要完成数据分块,并通过 Map 运算执行数据清洗算法,在遍历数据同时,Combine 运算可以完成路段交通流统计处理,将其作为中间结果传送到中心节点,由中心节点的 Reduce 运算进行中间统计结果合并,并最终由基于时空权重相关性的交通流预测模型生成预测结果。

## 2.2 交通流大数据清洗规则

交通流数据清洗是大数据聚合的关键步骤,其作用在于可以对冗余、无效、噪声和丢失等类型的错误数据进行剔除,为数据挖掘提供更加准确可靠的数据源,提高预测结果的准确性和有效性。目前,交通流数据清洗规则的研究已经取得一些成果<sup>[19-22]</sup>,但对于多源异构的交通流大数据,清洗规则还没有统一的规范。

本文利用王晓原等<sup>[22]</sup>的高维交通流孤立点检测算法,结合阈值理论,制定清洗规则和清洗步骤,如图 2 所示。经过实验,利用分布节点的 Map 运算执行数据清洗算法,可以将 90%的错误、冗余数据纠正或丢弃,数据质量得到较大提高,以此为数据源的交通流预测结果将更加准确。由于数据清洗过程中对每一条交通流数据进行遍历和清洗分析,

所以在数据遍历过程中即可完成周期路段流量统计,统计结果可作为分布节点中间结果传送到中心节点,最终完成合并和交通流预测。

## 3 基于时空权重相关性的交通流预测模型

### 3.1 交通流在路网中的相关性分析

城市道路交通网络的拓扑结构错综复杂,但任意两条路段的邻接关系只存在两种情况:1)二者邻接,某个路口即为二者邻接点;2)二者非邻接,此时必须经过一条或多条中间路段,才能实现二者的联通。

首先分析较为简单的邻接路段,交通流从上游路段经过路口到达下游路段时存在重分配问题,例如车辆在典型的十字路口存在直行、左转和右转 3 种情况。因此,上游交通流的重分配关系即为邻接路段交通流之间的关系。图 3 为路口上、下游关系示意图,表 1 列出下游路段对应上游路段的编号。

图 3 和表 1 明确了路网中路段的上、下游关系,因此针对第  $k$  个交通流数据采集周期 $[t_k, t_{k+1}]$ ,某条上游路段  $l_i$  与下游路段  $l_j$  之间路口转弯率即为二者的交通流重分配关系,该分配关系体现了路段间的时空相关性,其量化表达式为

$$r_{ij}(k)=\theta_{ij}(k-1), \quad (1)$$

其中,  $r_{ij}$  表示路段  $l_i$  与路段  $l_j$  的时空相关性,  $\theta_{ij}$  表示从上游路段  $l_i$  分配至下游路段  $l_j$  的交通流量。

对于非邻接路段,交通流需要经过多个路口到达指定下游路段,因此会经历多次重分配,即为上

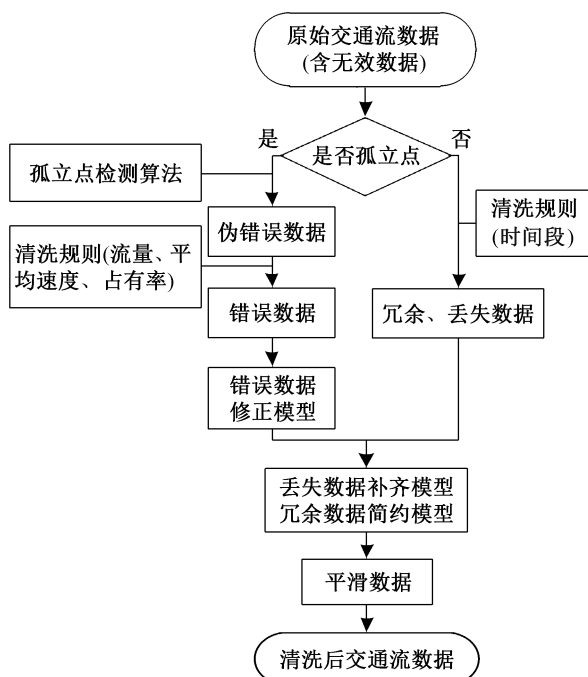


图 2 交通流数据清洗步骤

Fig. 2 Traffic flow data cleaning steps

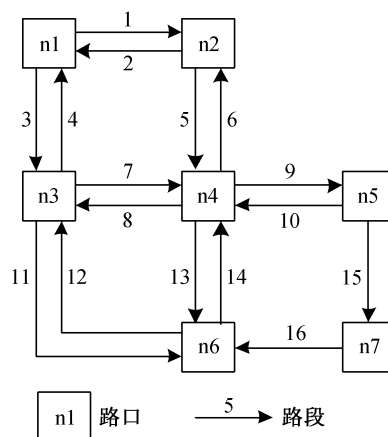


图 3 交通路口上、下游交通流示意图

Fig. 3 Traffic flow upstream and downstream in the traffic intersection

表 1 图 3 中 n4 路口上、下游路段关系  
Table 1 Upstream and downstream relationship  
in intersection n4

下游路段编号	对应上游路段编号
6	7, 10, 14
8	5, 10, 14
9	5, 7, 14
13	5, 7, 10

游路段到下游路段的多阶重分配, 分配过程与驾驶员的路径选择偏好相关。由于路径选择问题的影响因素较多, 本文仅考虑正常状态下多数驾驶员的路线选择, 即最短路径。此时, 某条上游路段到指定下游路段仅有唯一路径, 交通流多阶重分配可在此路径中体现。

非邻接路段中, 设上游路段  $l_i$  与下游路段  $l_j$  的最短路径中的路口数量为  $n$ , 则在第  $k$  个交通流数据采集周期  $[t_k, t_{k+1}]$  内, 在  $n$  个路口转弯率多阶重分配即为二者的交通流重分配关系, 其量化表达式为

$$r_{ij}^n(k) = \theta_{n-1,j}(k-1)\theta_{i1}(k-n) \prod_{p=1}^{n-2} \theta_{l_{p'}l_{p+1}}(k-n-p)。(2)$$

从式(1)和(2)发现, 下游路段交通流量是多阶上游路段累积分配的结果。交通流分配阶数的增加意味着经过路口数量的增加, 而距离下游路段越远的上游路段, 与其时空相关性也越小, 分配至下游路段的交通流也越少。经过实验发现, 上、下游路段距离在 2 阶以内时, 其时空相关性较大, 3 阶以上的上、下游路段时空相关性较小, 可以忽略不计。

## 3.2 基于时空权重相关性的改进 STARIMA 模型

### 3.2.1 改进 STARIMA 模型空间权重矩阵

根据路网中交通流相关性分析, 可以发现上、下游路段的时空相关性随阶数增大而减小。因此, 可以依据此规律进行时空自回归移动平均模型 STARIMA 中的空间权重矩阵设计, 通过模型的应用进行交通流预测。

时空自回归移动平均模型 STARIMA 由 Pfeifer 等<sup>[23-25]</sup>提出, 基本公式为

$$Z(t) = \sum_{k=1}^p \sum_{h=0}^{m_k} \phi_{kh} \mathbf{W}^{(h)} Z(t-k) - \sum_{l=1}^q \sum_{h=0}^{n_l} \theta_{lh} \mathbf{W}^{(h)} \varepsilon(t-l) + \varepsilon(t), \quad (3)$$

$k$  为时间延迟,  $h$  为空间间隔,  $p$  为时间自回归延迟,  $m_k$  为第  $k$  个时间自回归项的空间间隔,  $\phi_{kh}$  为时间延迟为  $k$  且空间间隔为  $h$  的自回归参数,  $q$  为移动时间平均延迟,  $n_l$  为第  $l$  个时间移动平均项的空间间隔,  $\theta_{lh}$  为时间延迟为  $l$  且空间间隔为  $h$  的移动平均参数,  $\varepsilon(t)$  为随机误差,  $\mathbf{W}^{(h)}$  为  $h$  阶空间权重矩阵。

式(3)中,  $\mathbf{W}^{(h)}$  为一个  $N \times N$  阶矩阵, 但矩阵元素的取值并没有严格的要求, 使用者需要根据实际问题来确定矩阵元素的含义和取值。

针对交通流预测问题, 该矩阵可作为路段间的空间权重矩阵进行元素赋值, 其代表的实际意义即为路段间的时空相关性。考虑到文献[23]中对权重矩阵的 3 种限制条件, 结合前面对交通流在路网中的相关性分析, 本文提出的空间权重矩阵元素计算公式为

$$\begin{cases} w_{ij}^{(l)} = r_{ij}^l(k), & l = 1, \\ w_{ij}^{(l)} = r_{ij}^l(k) / \sum_{i=1}^N r_{ji}^l(k), & l > 1. \end{cases} \quad (4)$$

式(4)不但满足文献[23]的 3 种限制条件, 还在上、下游路段拓扑关系的基础上, 体现了交通流的多阶分配相关性。因此, 该权重矩阵可作为新的确定时空相关性的方法, 在交通流预测分析中进行应用。

### 3.2.2 模型应用步骤

基于时空权重相关性的改进 STARIMA 交通流预测模型, 其应用流程由以下 6 个步骤组成。

1) 路段拓扑关系抽象化。由城市路网构建拓扑关系, 并对其进行抽象化, 得到的网状结构中需要包含交通流方向数据、路段长度和路段等级等数据, 并可准确地表示路段上、下游关系。

2) 确定空间权重矩阵元素。基于路段抽象化网状结构, 使用文献[26]中的动态转弯率预测模型, 估计交通流重分配比例, 再利用式(4)构建一阶和二阶空间权重矩阵。时空相关性较小的三阶以上权重矩阵忽略不计。

3) 时间序列平稳化。使用文献[10]中的序列图差分方法, 将非平稳的实际交通流时间序列平稳化。

4) 确定模型的阶数和参数。利用时空自相关函数<sup>[27]</sup>和时空偏相关函数<sup>[28]</sup>确定 STARIMA 平均阶数, 同时使用预测值与实际值的残差平方来估计参数。

5) 模型校验和诊断。检查参数统计显著性, 同时判断预测值与实际值误差序列能否满足随机误差条件, 如果未满足校验条件则返回上一步, 重新确



定阶数和参数。

6) 交通流预测分析。模型阶数和参数符合要求后, 将交通流历史全集数据和周期增量数据导入模型, 即可进行预测分析。

## 4 应用实例分析

### 4.1 数据来源及实验环境

本文实验数据由智能交通综合管理平台提供。该平台包括交通视频监控、交通信息采集和评估、交通信息诱导和发布以及车辆智能监测记录等系统, 目前已在河南省多个地市实现部分应用。

郑州市范围内的交通流数据采集传感器主要包括地磁检测器、微波检测器和视频检测器, 三类传感器数据融合浮动车 GPS 数据, 可以提供丰富准确的交通流量信息。平台数据总量已达到 160 多亿条, 日均增量约 2000 万条。本文选取 2015 年 11 月 9 日至 11 月 22 日共计 14 天的数据进行实验。

实验环境包括: 一台中心节点服务器, 用于进行数据整合和交通流预测分析; 4 台分布节点服务器, 用于进行周期增量交通流数据聚合和清洗; 服务器配置均为 Intel5620 2.4 GHz, 6 核, 4 GB 内存, 2 TB 硬盘。

### 4.2 交通流大数据分布式增量聚合实验

实验中设定 15 分钟为一个数据增量周期。以数据日均增量 2000 万条计算, 每一周期内的数据增量约为 20 万条。由于交通流数据存在不同时段内分布不均匀的特点, 因此一个高峰时段内的数据量可能达到平均量的 2 倍, 约 40 万条。由于智能交通综合管理平台是不间断运行的, 交通流数据也在持续增长, 因此必须在一个数据增量周期内完成快速聚合处理, 才能满足交通流预测的需求。

针对交通流大数据聚合处理, 本文使用以下两种方法进行实验, 并对实验结果进行比较。

1) 基于 MPI 的数据聚合。MPI 是一种基于消息传递的并行计算程序设计模式。实验中利用文献[29]的 MPI 方法完成交通流大数据聚合, 在中心节

点服务器设置一个主进程, 在 4 台分布节点服务器设置 4 个从进程, 由从进程并行完成数据遍历清洗以及路段交通流统计, 将中间结果传送到主进程, 然后由主进程完成分布节点的数据融合, 利用本文模型完成预测。

2) 基于分布式增量 MapReduce 的数据聚合。分布节点的 4 台服务器首先对数据集进行平均分块, 共配置 48 个 Map 运算, 每个 Map 运算对应一个数据块, 由 Map 运算完成该数据块的遍历和清洗运算, 每个分布节点配置一个 Combine 运算, 负责路段交通流统计数据处理, 并将其作为中间结果传送到中心节点, 在中心节点使用 Reduce 运算完成中间统计数据融合, 并执行预测分析计算。

实验中比较了两种方法在 2 万、5 万、10 万、20 万、50 万和 100 万条数据条件下的运行效率, 结果如图 4 和表 2 所示。

通过效率对比可以看出, 数据量小于 50 万条时, MPI 方法所耗时间较少, 这是由于 MapReduce 方法所包含的应用架构逻辑较多, 该架构本身耗费的时间在数据处理总时间中所占比例较大。当数据量逐渐增大时, 数据处理耗时所占比例越来越大, 应用架构耗时所占比例不断减小, 两种方法所耗时间逐渐接近, 当数据量增加至 50 万条时, Map-

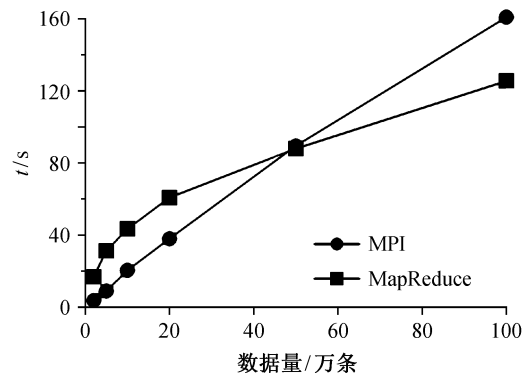


图 4 两种算法不同数据量时间对比  
Fig. 4 Cost of time of two algorithms for different amount of data

表 2 两种算法效率比较  
Table 2 Efficiency of two algorithms

算法	t/s					
	2 万条数据	5 万条数据	10 万条数据	20 万条数据	50 万条数据	100 万条数据
MPI	3.595	9.573	18.365	36.879	91.368	180.774
MapReduce	16.786	31.213	43.357	60.674	87.849	135.407

Reduce 方法所用时间反而少于 MPI 方法。随着数据量继续增加, MapReduce 方法的优势更加显著。经过两种算法效率对比, 二者耗费的时间数量级虽然一样, 但 MapReduce 方法具有更加快速的开发周期和稳定持续的运行效果。因此, 从开发和运行效率上看, 实验使用的交通流大数据分布式增量聚合方法可以满足交通流预测分析的需求。

#### 4.3 基于时空权重相关性的交通流预测实验

交通流数据经过分布节点的聚合和统计后, 会将时间周期内的路段交通流量数据集作为中间结果传送到中心节点, 经过合并, 即可使用本文提出的基于时空权重相关性的改进 STARIMA 模型进行交通流预测。按照本文设计的模型应用流程, 第一步是为交通网构建拓扑, 并将研究区域抽象化。图 5 为郑州市交通道路网, 图 6 是对图 5 中虚线矩形框范围内龙子湖高校园区路网抽象化结果示意图。

本文实验对比了以下两种交通流预测方法。



图 5 郑州市城区道路交通网  
Fig. 5 Zhengzhou city road network

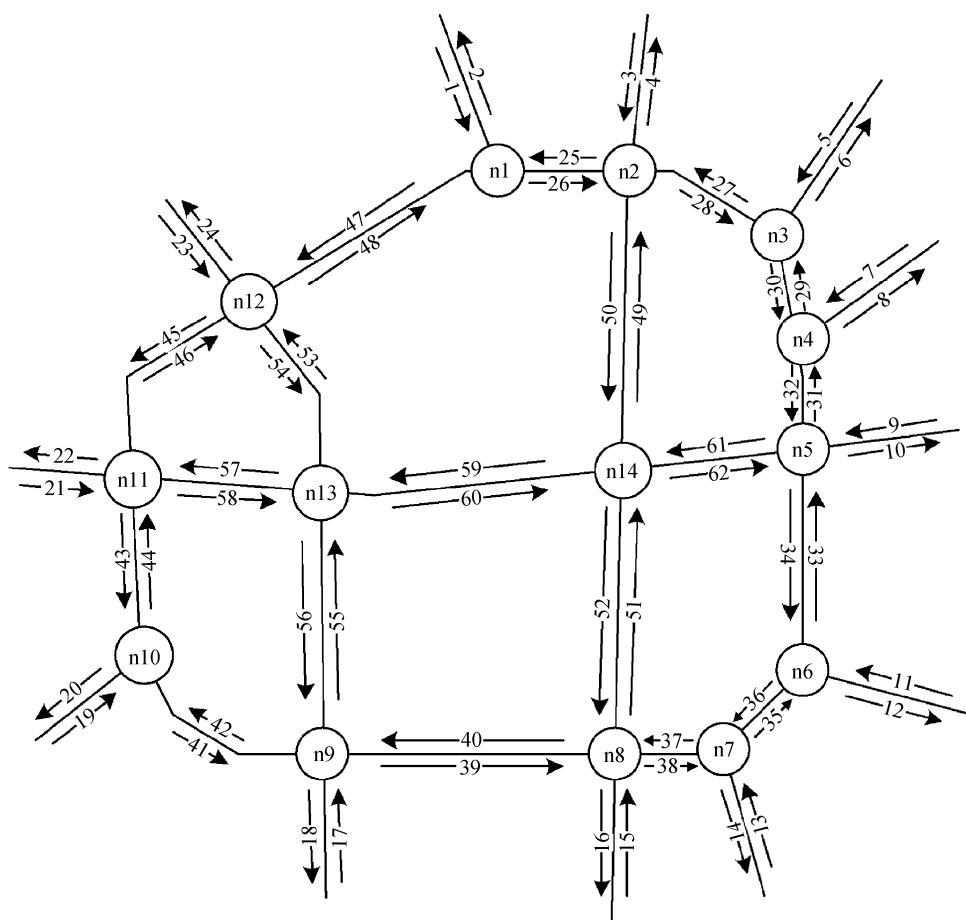


图 6 龙子湖高校园区路网抽象化结果示意图  
Fig. 6 Abstraction of Longzi Lake college area road network

表 3 两种方法预测结果均方误差(MSE)对比  
Table 3 The MSE comparison of two predict methods results

路段 编号	Dynamic STARIMA	路网相关 STARIMA	路段 编号	Dynamic STARIMA	路网相关 STARIMA	路段 编号	Dynamic STARIMA	路网相关 STARIMA
1	8638.27	6643.68	22	3338.29	2839.48	43	3977.67	3058.83
2	7396.01	6032.39	23	6236.77	5064.66	44	4458.98	3847.45
3	6631.31	4983.57	24	7784.39	5338.54	45	8979.37	7432.82
4	8362.58	5986.44	25	2526.67	1438.33	46	8443.69	6234.73
5	2948.21	1863.49	26	3464.32	1974.34	47	2798.55	1846.49
6	4820.22	2799.35	27	8764.39	5890.43	48	3985.42	2275.45
7	9230.23	7390.32	28	6549.48	5438.93	49	7849.43	5893.37
8	7857.46	6074.45	29	12974.45	8865.29	50	6692.38	5624.78
9	15324.01	12984.61	30	13084.44	10474.63	51	6639.32	4542.43
10	11479.46	8753.05	31	5478.34	3740.35	52	5873.57	4147.57
11	3892.27	2775.73	32	4858.34	3275.39	53	3720.32	2475.54
12	4917.48	3295.48	33	7434.95	6578.36	54	3920.28	2143.27
13	3729.33	2903.57	34	7868.23	5920.49	55	2039.58	1343.47
14	3928.23	2638.57	35	6743.23	4839.33	56	3235.62	1634.57
15	4478.19	3902.29	36	7820.33	5923.67	57	4838.88	3822.44
16	6903.36	4632.84	37	8068.35	6488.38	58	5749.23	4727.28
17	10573.48	7296.23	38	7819.28	6367.35	59	6403.45	4884.74
18	9033.54	7018.83	39	3894.22	2057.75	60	6653.63	4954.54
19	4847.34	3892.33	40	4780.44	3628.65	61	7570.49	6343.45
20	3309.21	3087.88	41	7897.36	4929.38	62	8788.48	6932.43
21	5789.22	4274.49	42	8902.34	5563.37			

1) Kamarianakis 等<sup>[12]</sup>的 Dynamic STARIMA 预测模型。将上、下游路段的相关性引入 STARIMA 模型,但只考虑一阶上游路段的影响。

2) 本文基于时空权重相关性的改进 STARIMA 模型。不仅引入路网之间的相关性,而且将  $n$  阶上游路段的影响考虑在内。

以图 6 中龙子湖高校园区路段为例,分别使用两种预测模型,利用前 4 天的交通流数据作为历史数据,对后 10 天的交通流进行预测。以 15 分钟为一个数据增量周期,前 4 天的数据分为 384 个周期,在此基础上预测后 10 天共 960 个周期的交通流量。将预测交通流与系统实际采集的交通流数据对比,计算二者的均方误差,即可判断预测模型的准确度。两种方法预测结果的均方误差见表 3。

从表 3 可以看出,基于时空权重相关性的改进 STARIMA 方法预测结果比 Dynamic STARIMA 方法更准确。原因在于,虽然 Dynamic STARIMA 方法将路段相关性作为交通流预测的重要条件,但仅考虑一阶邻接路段的相关性,而本文提出的方法在此基础上构建了时空权重矩阵,将更符合实际交通流分配规律的二阶以上相关性考虑在内,因此预测

结果更准确。

## 5 总结与展望

本文设计了分布式增量交通流大数据聚合方法。在使用分布式增量方法进行交通流大数据聚合处理时,利用交通流数据清洗规则,将错误、冗余数据纠正或丢弃,在快速聚合数据的同时数据质量得到提高,为交通流预测分析建立了基础。通过对比实验,验证了本文提出的方法在开发和运行效率上能够满足交通流大数据聚合和预测的需要。本文设计的基于时空权重相关性的交通流预测模型以路口转弯率多阶分配为基础构建时空权重矩阵,实现了路段相关性的量化,并将其应用于交通流预测。通过对比实验,预测结果比仅考虑邻接路段相关性的 Dynamic STARIMA 模型更加准确,可以作为交通诱导信息发布的参考依据。

本文的预测模型使用了最短路径作为路段间的唯一路径,但在实际行车时,影响驾驶员选择路线的因素很多,例如最短时间、中途点选择等,因此将路径选择问题与交通流预测模型相结合是下一步工作的研究方向。

## 参考文献

- [1] 李德仁, 马军, 邵振峰. 论时空大数据及其应用. 卫星应用, 2015(9): 7-11
- [2] Bose J H, Andrzejak A, Hogqvist M. Beyond online aggregation: parallel and incremental data mining with online Map-Reduce // Proc of Workshop on Massive Data Analytics on the Cloud. New York: ACM, 2010: 1-6
- [3] Aghabozorgi S, Saybani M R, Wah T Y. Incremental clustering of time-series by fuzzy clustering. Journal of Information Science and Engineering, 2012, 28(4): 671-688
- [4] Laptev N, Zeng K, Zaniolo C. Very fast estimation for result and accuracy of big data analysis: the EARL system // Proc of ICDE. Piscataway. Brisbane: IEEE, 2013: 1296-1299
- [5] Zhang S B, Han J Z, Liu Z Y, et al. Accelerating mapreduce with distributed memory cache // Proc of ICPADS. Piscataway. Shenzhen: IEEE, 2009: 472-478
- [6] Stephanedes Y J, Michalopoulos P G. Improved estimation of traffic flow for real-time control // Transportation Research Record 795. Washington DC: Transportation Research Board, 1981: 28-39
- [7] Okutani I, Stephanedes Y J. Dynamic prediction of traffic volume through Kalman filtering theory. Transportation Research Part B: Methodological, 1984, 18(1): 1-11
- [8] Ahmaed M S, Cook A R. Analysis of freeway traffic time-series data by using Box-Jenkins technique // Transportation Research Record 722. Washington DC: Transportation Research Board, 1979: 1-9
- [9] Dougherty M S, Cobbett M R. Short-term inter-urban traffic forecasts using neural networks. International Journal of Forecasting, 1997, 13(1): 21-31
- [10] Ledoux C. An urban traffic flow model integrating neural networks. Transportation Research Part C: Emerging Technologies, 1997, 5(5): 287-300
- [11] Yue Yang. Spatial-temporal dependency of traffic flow and its implications for short-term traffic forecasting [D]. Hong Kong: The University of Hong Kong, 2006
- [12] Kamarianakis Y, Prastacos P. Space-time modeling of traffic flow. Computers & Geosciences, 2005, 31: 119-133
- [13] Martin R L, Oeppen J E. The identification of regional forecasting models using space-time correlation functions. Trans Inst Brit Geogr, 1975, 66: 95-118
- [14] 余碧莹, 邵春福. 基于时空模型的道路网交通状态预测 // 第四届中国智能交通年会论文集. 青岛, 2008: 546-551
- [15] Lin Shulan, Huang Hongqiang, Zhu Daqi, et al. The application of space-time ARIMA model On traffic flow forecasting // Proceedings of the 8th International Conference on Machine Learning and Cybernetics. Baoding, 2009: 3408-3412
- [16] Min Xinyu, Hu Jianming, Chen Qi, et al. Short-term traffic flow forecasting of urban network based on dynamic STARIMA model // Proceedings of the 12th International IEEE Conference on Intelligent Transportation Systems. St. Louis, MO, 2009: 461-466
- [17] 瞿莉. 基于动态交通流分配参数的网络交通状态建模与分析[D]. 北京: 清华大学, 2010
- [18] 张和生, 张毅, 胡东成, 等. 区域交通状态分析的时空分层模型. 清华大学学报: 自然科学版, 2007, 47(1): 157-160
- [19] 王晓原, 张敬磊. 交通流信息挖掘的非参数概率变点模型研究. 武汉理工大学学报: 交通科学与工程版, 2010, 34(4): 801-805
- [20] 郭志懋, 周傲英. 数据质量和数据清洗研究综述. 软件学报, 2002, 13(11): 2076-2082
- [21] 张敬磊, 王晓原. 基于非线性组合模型的交通流预测方法. 计算机工程, 2010, 36(5): 202-205
- [22] 王晓原, 张敬磊, 吴芳. 交通流数据清洗规则研究. 计算机工程, 2011, 37(20): 191-193
- [23] Pfeifer P E, Deutsch S J. A three-stage iterative procedure for space-time modeling. Technometrics, 1980, 22(1): 35-47
- [24] Pfeifer P E, Deutsch S J. Identification and interpretation of first-order space-time ARMA models. Technometrics, 1980, 22(3): 397-408
- [25] Pfeifer P E, Deutsch S J. Variance of the sample-time autocorrelation function of contemporaneously correlated variables. SIAM Journal of Applied Mathematics, Series A, 1981, 40(1): 133-136
- [26] Deng Shuo, Hu Jianming, Wang Yin, et al. Urban road network modeling and real-time prediction based on house holder transformation and adjacent vector // Proceedings of the 6th International Symposium on Neural Networks. Berlin: Springer, 2009: 899-908
- [27] Bezdek J C, Pal N R. Some new indexes of cluster validity. IEEE Trans Syst Man Cy, 1998, 28: 301-315
- [28] Kamarianakis Y, Prastacos P. Space-time modeling of traffic flow. Comput Geosci-UK, 2005, 31: 119-133
- [29] 牛新征, 余莹. 面向大规模数据的快速并行聚类划分算法研究. 计算机科学, 2012(1): 134-137