

文章编号:1007-2780(2020)11-1204-06

基于 Hadoop 的交通视频大数据监控方案

李晓蕾*

(宁波财经学院,浙江 宁波 315175)

摘要:为了解决海量交通视频数据的监控和分析问题,本文对 Hadoop 大数据背景下的交通视频监控技术进行了深入研究,提出了基于交通视频数据的异常检测算法的设计方案,实现了交通数据的实时更新和异常分析,同时针对海量交通监控视频,设计了基于 Hadoop 组件 MapReduce 的并行实现算法,并通过浙江省某市的实际交通数据验证算法的有效性和准确性。经过实验证明,本文算法可以有效计算出交通拥堵情况和异常情况,相对于传统方案,本文方案可以聚焦 10 min 范围内的时间粒度对交通情况进行实时分析,相对于传统的分布式计算模型,本文的方案 10 min 延迟可以控制在 2.1 s,比传统方案延迟降低了 81%,基本满足交通视频监控的实时和细颗粒度等要求。

关键词:并行计算;海量数据分析;分布式计算;异常堵点检测

中图分类号:TP274 **文献标识码:**A **doi:**10.37188/YJYXS20203511.1204

Big video monitoring scheme of traffic video based on Hadoop

LI Xiao-lei*

(Ningbo University of Finance and Economics, Ningbo 315175 China)

Abstract: In order to solve the problem of monitoring and analyzing massive traffic video data, the in-depth research on traffic video surveillance technology in the context of hadoop big data is conducted, and a design scheme of anomaly jam detection algorithm is proposed based on traffic video data to realize traffic real-time data update and anomaly analysis. At the same time, for the massive traffic monitoring video, a parallel implementation algorithm is designed based on Hadoop component MapReduce. Finally, the effectiveness and accuracy of the algorithm is verified by actual traffic data of a city in Zhejiang Province. The algorithm in this paper can effectively calculate the traffic congestion and abnormal conditions. Compared with the traditional scheme, this scheme can focus on the time granularity in the range of 10 min to analyze the traffic situation in real time. Compared with the traditional distributed computing model, the 10 minute delay of this scheme can be controlled at 2.1 s, which is 81% lower than the traditional scheme, which basically meets the real-time, fine-grained requirements for traffic video surveillance requirements.

Key words: parallel computing; massive data analysis; distributed computing; anomaly blocking point detection

收稿日期:2019-12-17;修订日期:2020-03-04.

基金项目:浙江省基础公益研究项目(No.LQ20F020025)

Supported by Zhejiang Basic Public Welfare Research Project(No. LQ20F020025)

* 通信联系人, E-mail:lixiaolei1291129@163.com

1 引言

在交通道路维护中,实时交通监控是一项复杂而富有挑战的任务。传统的交通监控系统的监控方案从高清摄像机中获取或存储数字视频到计算机中,其流媒体数据通过 mysql 等数据库技术保存数据,而交警则通常采用手动记录行驶的车辆,通过路边运动摄像机对车辆进行计数的方法来进行人工监控。如果视频分析没有自动完成,则需要大量的人工操作,并需要多方协同来手动分析视频中的场景。这些方法的人力成本和时间成本都相对昂贵。特别是交通事件的发生最常见的问题是车辆的颜色、尺寸和类型,这些信息用人力往往很难准确地捕捉到,并且时常存在误差。基于视频流的分析可以辅助交警或者交通管理人员对交通现场进行快速处理,可以极大提升交通道路的安全性,因此视频流分析技术一大重要应用场景便是自动化安全和监视系统。执法机构、交通监控工作站和组织都可以将视频分析技术用于公共安全。与手动应用程序相比,自动视频分析需要较少的成本和人力来分析视频流,因此得到了学术界和工业界的青睐。许乐等研究者用自动化的视频分析研究不同自动化系统视频图格式,实现格式之间的转化,减轻了数据维护人员的工作量^[1]。视频分析还可以帮助检测人脸、车辆颜色和车牌等物体,完成自动对象检测、目标分类等复杂的计算机任务^[2]。其中对象检测是交通监控系统 and 执法机构应用中最常见的技术之一,最常见的算法是模板匹配^[3],使用高斯混合模型(GMM)进行背景分离^[4]和级联分类器^[5]。模板匹配仅适用于模板中定义的对象,但是这类方法计算量巨大,不适用于新对象时常出现或者消失的场景,因此不适用于交通视频。背景帧提取方法是一种检测在背景中移动的对象并分离出来的方案,但是背景分离和帧差异也面临计算成本高昂的问题,通常不适用于实时处理。江鹏宇用级联分类器 AdaBoost 算法对运动目标区域进行行人检测,减小视频中背景的干扰,加快检测速度,并利用卡尔曼滤波算法和匈牙利最优匹配算法对视频监控中的行人进行跟踪^[6]。近年来 Hadoop 快速发展,为大规模数据处理提供了高效的解决方案。毕莎莎针对实时云转码系统中已转码的视

频媒体文件重新使用率不高的情况,引入转码缓存机制,采用云存储空间缓存视频,根据视频资源的流行度和请求间隔来实时替换出效能值较小的视频^[7]。刘云恒等学者采用基于 Hadoop 技术的视频大数据处理平台,并采用以 Map-Reduce 算法为基础的人脸检索与识别算法来实现公安视频大数据的智能信息处理^[8]。

2 基于 Hadoop 的交通视频分析技术

2.1 基于 Hadoop 的视频分析

使用基于 Hadoop 的 GPU 集群进行视频流分析可以从不同的资源(如 GPS、社交媒体、物联网、平板电脑、Web 服务和视频)生成大量数据^[9]。在传统的数据库技术中,分析处理和存储视频数据是一项艰巨的任务^[10],为了处理大量视频数据,可以在不同节点之间分配处理。大范围内可能有数百个甚至更多的摄像机在本地服务器上捕获流和过程。视频流分析需要大规模的计算,当前很多企业采用云计算来完成这类海量计算,云计算可提高实时监控交通视频数据的效率。商业供应商还提供视频分析应用程序。一个最常见的做法便是将 Hadoop 用于将视频帧拆分为较小的单元,并以并行模式将它们转换为一系列图像帧。HDFS 通过多个物理磁盘群集存储大量数据。分割成较小单元后的序列文件分布在 HDFS 的节点上,以提高处理速度。原始视频流分为 64 MB(默认)大小的块^[11]。录像模块将视频流传输到监管站,并将视频存储在其本地存储中。视频流是从不同的地方捕获,并由 IP 无线传输模块通过媒体服务器传输到监管站。实况视频流由监管站处理,并存储结果以供进一步分析。有关车速、事故和交通拥堵,以及城市主要站点和线路可以存储在 Hive 里。当道路拥堵发生时,Hive 表可以向旅行者建议替代路线,当检测到道路交通事故时,系统会向附近的医院和公路救援队发送警报。

2.2 算法步骤

对于交通视频数据分析,交通异常的检测是最重要的部分之一,在国内的交通事故中,90%以上交通事故都和道路拥堵有关。因此本文重点研究 Hadoop 下交通拥堵的检测和分级。

本文提出算法的主要核心是根据 Hadoop 中

交通视频得到的路网历史数据,对每个车道的合理速度值进行整理和清晰,根据整个路段中每个时段不同方向上的信息,对每个路段每个时段的拥堵概率进行先验判断。通常来说,每个道路的拥堵状态都具有一定的周期性特点,特别是对于工作日而言,其出行规律具有较高的相似性。因此本文采用将 1 天 24 h 分为若干个时间段,统计各个时间段的拥堵概率。其主要流程如下:

(1)原始数据预处理:由于原始数据在采集过程中可能存在数据丢失、数据监测出错等情况,因此首先用预处理进行修正。将交通视频的流量、速度和交通指挥部门的历史数据进行对比,将超过阈值范围内的流量和速度定义为错误数据进行修正,为了简化处理的复杂度,本文将超过阈值范围内的数据和缺失的统计数据用历史平均数据来代替。

(2)交通拥堵判断:本文将一天分解为 144 个时间段,每隔 10 min 是一个时间单位,在此基础上计算如下指标:

车道占有率:

$$c = 1/T \sum_{t=1}^n t_i, \quad (1)$$

$$o = 1/N \sum_{i=1}^n c_i, \quad (2)$$

式中, N 是一个方向上的车道个数,车道占有率定义为汽车占用道路的时间比, i 代表第 i 个车道, T 代表观测时间。

$$\text{speed}_i = 1/N \sum_{j=1}^n v_j, \quad (3)$$

$$\text{speed} = \frac{1}{V} \sum_{i=1}^n v_i * \text{speed}_i, \quad (4)$$

式中, speed_i 代表 i 车道的速度, v_i 代表 i 车道的流量。将交通部门在某一车道的历史速度求解平均值,如果当前速度 speed_i 低于历史平均值,则认为可能出现了拥堵。

(3)拥堵率计算

某个时间点内拥堵率计算如下:

$$P(h) = \frac{M_h}{N_h}, \quad (5)$$

式中, h 代表对应的时间段 id, M 是 h 时间内通过车速低于历史平均值的次数, N 是统计数据中历史数据个数。

(4)计算路段异常度

本文将路段异常度定义为某天道路拥堵的异

常程度,该指标的数值越高,则路况越拥堵,发生交通状况的可能性也大幅提高,因此需要交警部门进行疏导,从而避免事故的发生。异常度的计算流程如图 1 所示。

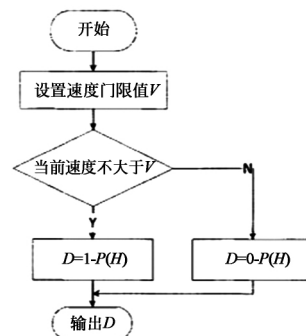


图 1 异常度计算流程

Fig.1 Calculation flow of abnormality degree

从图 1 可以看出,在当前速度大于阈值 V 时, $D=0-P(H)$,即异常度为负值(代表没有交通异常发生),而当当前速度低于阈值 V 时,则异常度为正值。

(5)数据实时更新

随着时间的推移,道路周边基础设施如写字楼的建成、地铁的运营等因素,历史拥堵数据会发生变化,因此需要采用实时更新的方式来保证历史数据的准确性。当然交通视频历史数据的更新本身较为耗费计算资源,为了减少不必要的重复计算,本文提出的分片更新算法如图 2 所示。

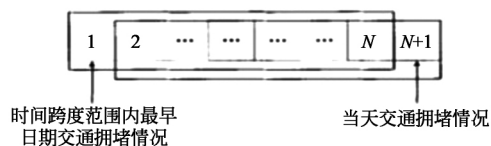


图 2 历史数据更新算法示意图

Fig. 2 Schematic diagram of historical data update algorithm

更新算法的核心是将时间跨度范围内最早日期的拥堵状况与当前的状况进行比较,如果交通情况从拥堵恢复为不拥堵则记为 $M=-1$,否则记为 $M=1$ 。

2.3 基于 Hadoop 的拥堵检测分布式计算

从上述关于异常拥堵的算法中,可以看到由于本文算法对实践切割相对较细,因此要获得全天 144 个时间段的拥堵情况,需要计算车道占有

率、车道交通流量、车道平均速度,其计算复杂度较高,为了提高计算效率,本文采用分布式并行的方案来加快算法的运行时间。本文采用 Hadoop 的 MapReduce 架构,利用 HDFS 将数据冗余备份,由于 MapReduce 的本地化数据技术能尽量减少网络 IO,而且节点可以自适应实现负载均衡,因此本节的重心放在 MapReduce 的计算模型。

对于车道占有率,在 MapReduce 计算的 Map 阶段,我们将数据按行切分,将代表时间段的字段 COLLECT_TIME 划分到不同时间槽中,具体的时间段用 time_id 作为 key,occ 代表拥堵情况。在 Reduce 阶段,对于相同的 key 求解 value 的算术平均值就可以获得平均占有率。

对于车道流量,本文在一个时间段内的某条道路进行计算,设 road_id 为道路的 id,flow 为流量,因此将 road_id 和 time_id 设置为 key,flow 设置为 value,对于相同的 key 求解所有 value 的和就得到了总流量。

对于车辆的平均速度,本文利用同一个道路上不同路段的 way_id 的数据 speed、flow 进行汇集,将 way 设置为 key、speed、flow 设置为 value,在 reduce 阶段,式(3)和(4)可以分别算出车道平均速度和总速度。

3 实验及结果分析

本文选取浙江省某市的交通视频监控的 200 个采集点的数据进行分析,其交通视频监控设备如图 3 所示。



图 3 交通视频监控画面

Fig.3 Traffic video monitoring screen

标注出该组的道路编号,车道标号 way_id,车道流量为 total_flow,speed 是总速度,occ 是车道占有率,本文的实验流程如下:

(1)数据预处理,如前文所述,由于视频数据在分析过程中,可能存在一定的误判或者丢失,因此为了提高数据的容错能力,本文先进行预处理。其预处理规则见表 1。

表 1 错误数据判别原则

Tab.1 Error data discrimination principle

原理	内容	示例
阈值关系法	最大速度阈值	10 min 内平均速度大于 120 km/h
	最大流量阈值	10 min 内车道流量大于 200 辆
交通流理论	同一数据记录中交通流数据不一致	交通流量和速度的计算结果又明显误差

将上述数据按照“阈值关系理论”进行处理,不符合要求的数据标记为无效数据并进行修正。

将经过预处理的数据按照 144 个 10 min 时间段从 0 点到 24 点的顺序进行排列,按照式(1~5)计算出当前的历史拥堵概率。结果如图 4 所示。从图中可以看出,上午 11:00 到 11:45 拥堵概率接近 45%,下午 18:00 到 19:00 拥堵概率在 70%左右。

本文在选取了 2019 年 8 月 16 日的交通数据进行异常度分析,其结果如图 4 所示。

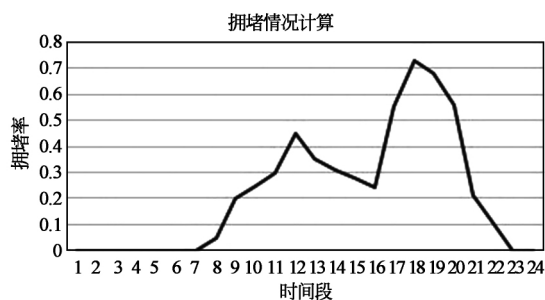


图 4 拥堵情况计算结果

Fig.4 Calculation results of congestion

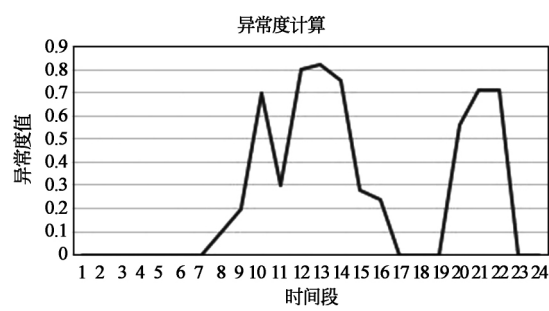


图 5 异常度计算

Fig.5 Calculation of abnormality degree

通过异常度计算(如图 5 所示),可知,异常度的取值在 $[0,1]$ 之间,可以看到 9 : 50,13 : 00,21 : 00 出现了明显的异常,通过调取该市当时的交通监控视频,我们发现这几个时间点都出现了不同程度的交通事故,引起了车辆的缓行,最后导致道路流量产生了异常。

在计算复杂度上,本文采用传统的分布式架构方案进行对比,在 10 台 8 核 16G 的分布式服务器和对等的 Hadoop 集群中,选取 2018 年 8 月 16 日的交通视频进行计算,用计算延迟和处理时间来代表算法结果,统计如表 2。

表 2 计算延迟统计表

Tab.2 Statistics of calculation delay

	10 min 视频计算延迟/s	10 min 视频计算时间/s	30 min 视频计算延迟/s	30 min 视频计算时间/s
传统分布式	10.3	612.5	45	1 845.4
本文方案	2.1	602.1	6.2	1 806.5

可以看到,本文提出的基于 Hadoop 方案的计算延迟可以控制在每 10 min 2.1 s,基本可以满足实时分析的需求,同时可以看到传统分布式算法的计算延迟和计算总时间随着时间的加长有恶化的趋势,但是本文方案可以有效避免这一情况。

4 结 论

本文对 Hadoop 大数据背景下的交通视频监控技术进行了深入研究,并对交通路况异常

拥堵进行了分析,然后阐述了基于交通视频数据的异常堵点检测算法的设计方案,利用拥堵概率、交通异常度等概念为交通状态分析提出了精确指标。针对海量交通监控视频,本文设计了基于 Hadoop 组件 MapReduce 的并行实现算法,最后通过浙江省某市的实际交通数据验证算法的有效性和准确性。经过实验证明,本文算法可以有效计算出交通拥堵情况和异常情况,对大数据背景下海量交通监控视频的分析提供有益参考。

参 考 文 献:

[1] 许乐.自动化系统视频图数据格式分析[J]. 科技视界,2016(17):81-82.
XU L. Analysis of video data format of automation system [J]. *Science & Technology Vision*, 2016(17): 81-82. (in Chinese)

[2] 毛建森,屈玉福.基于压缩感知的变尺度目标跟踪技术[J]. 液晶与显示,2016,31(5):497-505.
MAO J S, QU Y F. Tracking of variable scale object based on compressive sensing [J]. *Chinese Journal of Liquid Crystals and Displays*, 2016, 31(5): 497-505. (in Chinese)

[3] 钱方,孙涛,郭劲,等.采用图像特征的激光干扰跟踪效果评估[J]. 液晶与显示,2014,29(5):805-811.
QIAN F, SUN T, GUO J, et al. Assessment of laser-dazzling effects on tracking through image feature measurements [J]. *Chinese Journal of Liquid Crystals and Displays*, 2014, 29(5): 805-811. (in Chinese)

[4] 陈莹,朱明,刘剑,等.高斯混合模型自适应微光图像增强[J]. 液晶与显示,2015(2):300-309.
CHEN Y, ZHU M, LIU J, et al. Automatic low light level image enhancement using Gaussian mixture modeling [J]. *Chinese Journal of Liquid Crystals and Displays*, 2015(2): 300-309. (in Chinese)

- [5] 周晨卉,王生进,丁晓青.基于局部特征级联分类器和模板匹配的行人检测[J]. 中国图象图形学报,2010,15(5): 824-829.
ZHOU C H, WANG S J, DING X Q. Pedestrian detection based on partial feature and model matching [J]. *Journal of Image and Graphics*, 2010, 15(5): 824-829. (in Chinese)
- [6] 江鹏宇,杨耀权,彭璐.一种用于电力监控的行人运动检测与跟踪算法[J]. 电力科学与工程,2019,35(6):31-36.
JIANG P Y, YANG Y Q, PENG B. A pedestrian motion detection and tracking algorithm for electrical power monitoring[J]. *Electric Power Science and Engineering*, 2019, 35(6): 31-36. (in Chinese)
- [7] 毕莎莎,陈清华.基于 Hadoop 视频转码缓存策略的研究[J]. 计算机工程与设计,2015(3):683-686,692.
BI S S, CHEN Q H. Caching strategies based on video transcoding using Hadoop [J]. *Computer Engineering and Design*, 2015(3): 683-686, 692. (in Chinese)
- [8] 刘云恒,刘耀宗.基于 Hadoop 的公安视频大数据的处理方法[J]. 计算机科学,2016,43(S1):448-451,475.
LIU Y H, LIU Y Z. Hadoop-based public security video big data processing method [J]. *Computer Science*, 2016, 43(S1): 448-451, 475. (in Chinese)
- [9] 连彬彬,黄风华,谢小烽,等.基于云计算与 GPU 的大数据挖掘比较[J]. 电子技术与软件工程,2017(13):182-184.
LIAN B B, HUANG F H, XIE X F, et al. Comparison of big data mining based on cloud computing and GPU [J]. *Electronic Technology and Software Engineering*, 2017(13): 182-184. (in Chinese)
- [10] 韩红霞,孙航,张海波.基于 FPGA 的视频与通讯数据融合设计[J]. 液晶与显示,2015(6):987-991.
HAN H X, SUN H, ZHANG H B. Composition design of communication data and video data based on FPGA [J]. *Chinese Journal of Liquid Crystals and Displays*, 2015(6): 987-991. (in Chinese)
- [11] KUMAR C R, SUGUNA S. Visual semantic based 3D video retrieval system using HDFS [J]. *KSII Transactions on Internet and Information Systems*, 2016, 10(8): 3806-3825.

作者简介:



李晓蕾(1981—),女,浙江宁波人,硕士,副教授,2010 年于浙江大学获得硕士学位,主要从事数据挖掘、算法等方面的研究。E-mail:lixiaolei11291129@163.com