内部

**清华大学工业工程系**

负责人：姜 海

成 员：王正礼、赵天成

2016年11月

**多数据源的路况分析研究**





摘 要

本项目的目标是在多数据源的背景下，基于现有的RTMS数据、浮动车GPS的速度数据、从第三方（高德地图和百度地图）获取的路况数据以及道路交通事故数据开展路况分析的研究，完成了对RTMS数据进行数据质量分析和评估、对交通事故数据进行分类、检验不同数据源的速度数据的一致性、建立不同流量模式的特征库、量化交通事故对交通拥堵的时空影响。

我们通过分析RTMS数据的时间和空间完整度、数据重复情况、流量数据是否满足“守恒”等发现RTMS数据的数据存在不少问题；通过提取关键词，对122交管平台和103.9交通广播台的交通事故数据进行分类；通过分析RTMS速度数据和基于GPS的路段速度数据的一致性、第三方地图软件实时拥堵程度数据的相关性以及高德地图实时拥堵程度和GPS路段速度的一致性对不同数据源的速度数据进行了一致性的检验；通过分析不同路段的流量特征，对不同路段的宏观流量模型参数（流量-速度-密度图）进行分类，建立了不同流量模式的特征库；通过建立数学模型，研究了交通事故引发的交通拥堵在时间维度和空间维度上是如何传播的，并量化了事故引发的拥堵时长。

目 录

[1. 项目目标 1](#_Toc466580073)

[2. 数据来源和说明 2](#_Toc466580074)

[2.1 RTMS数据 2](#_Toc466580075)

[ RTMS数据的RepeatStatus和ErrorStatus字段有较多异常。 4](#_Toc466580076)

[ RTMS数据的记录条数低于理论值 4](#_Toc466580077)

[ RTMS数据的其他问题 5](#_Toc466580078)

[2.2 2014年5月的RTMS数据说明 8](#_Toc466580079)

[2.3 基于GPS的路段速度数据 9](#_Toc466580080)

[2.4 第三方地图软件的实时拥堵程度数据 11](#_Toc466580081)

[2.5 交通事故数据 11](#_Toc466580082)

[ 122交管平台事故数据 12](#_Toc466580083)

[ 103.9交通广播台事件数据 13](#_Toc466580084)

[3. 检验数据的一致性 16](#_Toc466580085)

[3.1 RTMS速度数据和基于GPS的路段速度数据的一致性 16](#_Toc466580086)

[3.2 第三方地图软件实时拥堵程度数据的相关性 18](#_Toc466580087)

[3.3 高德地图实时拥堵程度和GPS路段速度的一致性 19](#_Toc466580088)

[4. RTMS数据流量模式特征分析 20](#_Toc466580089)

[5. 交通事故引发的交通拥堵在时空上传播规律的研究 22](#_Toc466580090)

[5.1 模型描述 23](#_Toc466580091)

[5.2 案例研究 26](#_Toc466580092)

[ 仿真数据 26](#_Toc466580093)

[ 真实数据 28](#_Toc466580094)

[6. 参考文献 36](#_Toc466580095)

# **项目目标**

本项目的目标是在多数据源的背景下，基于现有的RTMS（Remote Traffic Microwave Sensor）速度和流量数据、浮动车GPS的速度数据、从第三方（高德地图和百度地图）获取的路况数据以及道路交通事故数据（源自103.9交通广播台和122交管平台），开展路况分析的研究，具体如下：

* 对RTMS数据进行数据质量分析和评估。通过分析RTMS数据，包括数据的时间和空间完整度、数据重复情况、流量数据是否满足“守恒”，为我们进行数据融合提供判断和取舍的依据；
* 对交通事故数据进行分类。通过分析122交管平台和103.9交通广播台的交通事故数据，提取关键词，对交通事故数据进行分类；
* 检验不同数据源的速度数据的一致性。基于现有的RTMS速度数据、浮动车GPS的速度数据和从第三方（高德地图和百度地图）获取的路况数据，运用统计学中的检验方法，判断这三种不同来源的速度数据是否存在显著差异，为BRTC独立发布路况数据提供同行业的横向对比结果；
* 建立不同流量模式的特征库。通过分析不同路段的流量特征，对不同路段的宏观流量模型参数（流量-速度-密度图）进行分类，进而建立不同流量模式的特征库；
* 量化交通事故对交通拥堵的时空影响。结合交通事故数据和RTMS数据，建立优化模型，分析交通事故对交通拥堵在时间维度和空间维度上的传播规律。

# **数据来源和说明**

在这一部分我们首先对各个数据源的数据类别、时间和空间范围进行说明，然后结合数据样例描述数据的具体内容，接下来通过分析数据的时间和空间的完整度、重复情况、逻辑错误，来评价各个数据源的数据质量。

截至2016/08/19，我们得到的数据如表 1所示。从表中可以看出我们从BTRC获得的数据主要包括交通状况数据（速度和流量）和交通事故数据两部分。交通状况数据主要包括RTMS数据和基于GPS的速度数据，交通事故数据主要包括122交管平台收到的报警数据和103.9交通台收到的事件数据。此外，我们还抓取了百度地图和高德地图以道路不同颜色表示的道路拥堵状况的数据。下面，我们将对每种类型的数据进行详细说明。

表 1 各数据源数据情况

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **数据类别** | **来源** | **时间范围** | **区域范围** | **样例** |
| 速度 | RTMS | 2016.04.27—2016.08.01 | 全城 | 表 2 |
| 基于GPS的速度数据 | 2016.04.01—2016.04.29  2016.05.12—2016.05.31  2016.06.01—2016.06.30  2016.07.01—2016.07.29 | 表 3 |
| 流量 | RTMS | 2016.04.27—2016.08.01 | 全城 | 表 2 |
| 交通事故 | 122交管平台 | 2015.11.11—2015.11.30  2015.12.03/11/15/16/26/27  2016.01—07 | 全城 | 表 4 |
| 103.9交通广播台 | 2016.04.13—2016.06.15 | 表 5 |

# 

## RTMS数据

RTMS数据的来源为北京全城6651个RTMS设备在4-7月记录的断面速度和流量信息。该数据是RTMS设备每2分钟收集一次，每一条记录都包含以下8个字段：

* ID: 每个RTMS设备的唯一编号；
* ReportTime: 数据收集时间，即RTMS设备收集该条记录的时间；
* Volume: 在2分钟内RTMS设备探测到的经过该断面的车辆数，单位为辆；
* Speed: 在2分钟内RTMS设备探测到的经过该断面的所有车辆速度的平均值，单位为km/h；
* Time\_Occp: 时间占有率，即在2分钟内有车经过断面的时间占2分钟的比例，为百分数；
* RepeatStatus: 表示数据是否重复。其中，1表示所有字段重复（保留第一条标01，删除的标1），2表示时间地点重复，数据不重复（通过取平均值的方式新生成一条标02，删除的标2）；
* ErrorStatus: 表示数据是否存在逻辑错误，其中0表示正常，31表示速度大于150，32表示占有率为0，流量不为0或相反，33表示单车道流量大于100，34表示流量为0，占有率为0，速度大于0；
* StoreTime: 数据入库时间，即数据被录入数据库的时间。

RTMS数据的数据样例如表 2所示， 表 2中第一条记录表示的是编号为HI9281c的RTMS设备，在2016/5/5 6:20--2016/5/5 6:22的时间间隔内，探测到通过断面的车辆为26辆，平均速度为66km/h，在这两分钟内，有34%的时间RTMS设备探测到有车辆通过该断面，而且数据没有重复或逻辑错误，该条记录于2016/5/5 6:31入库。因为RTMS设备位置匹配存在问题，且RTMS设备的坐标信息不完整，我们目前只能利用1538个（总共有6651个）RTMS设备的数据完成后续的分析。

表 2 RTMS设备数据样例

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **ID** | **ReportTime** | **Volume** | **Speed** | **Time\_Occp** | **RepeatStatus** | **ErrorStatus** | **StoreTime** |
| HI9281c | 2016/5/5 6:22 | 26 | 66 | 34 | 0 | 0 | 2016/5/5 6:31 |
| HI9281c | 2016/5/5 22:12 | 136 | 86 | 46 | 0 | 0 | 2016/5/5 22:25 |
| HI9281c | 2016/5/5 23:12 | 113 | 78 | 31 | 2 | 0 | 2016/5/5 23:25 |
| HI9281c | 2016/5/5 10:46 | 79 | 67 | 35 | 0 | 0 | 2016/5/5 11:04 |
| HI9281c | 2016/5/5 10:50 | 74 | 60 | 8 | 0 | 0 | 2016/5/5 11:04 |
| HI9281c | 2016/5/5 10:56 | 58 | 69 | 23 | 0 | 0 | 2016/5/5 11:04 |
| HI9281c | 2016/5/5 10:58 | 73 | 78 | 34 | 0 | 0 | 2016/5/5 11:04 |
| HI9281c | 2016/5/5 18:36 | 86 | 82 | 20 | 2 | 0 | 2016/5/5 18:48 |
| HI9281c | 2016/5/5 18:36 | 92 | 89 | 47 | 2 | 0 | 2016/5/5 18:37 |
| HI9281c | 2016/5/5 18:32 | 102 | 73 | 5 | 2 | 0 | 2016/5/5 18:37 |

关于RTMS数据，主要存在以下几个问题:

### RTMS数据的RepeatStatus和ErrorStatus字段有较多异常。

所有RTMS设备共回传约1.8亿条记录。其中，RepeatStatus和ErrorStatus字段显示，有大量记录存在重复，即仅有ID字段和ReportTime字段重复或所有字段重复。同时还有部分记录存在逻辑错误。去除重复和错误数据后，剩余的记录条数约为4千万条，正确且不重复数据的仅占总数据的22%左右。不同RTMS设备的数据正确且不重复的情况也不相同。在全部6651个RTMS设备中，有61%的RTMS设备收集的数据正确且不重复的比例在1%以下。有78%的RTMS设备（5193个）收集的数据正确率在70%以下。各个RTMS设备传回数据正确且不重复比例的直方图如图 1所示。图中，横坐标是正确且不重复数据的比例区间，每个区间的长度是0.02，左侧纵坐标是正确且不重复的数据比例在该区间内的RTMS设备的频数，图中红线为累计频率曲线。如，从第一条数据我们可以看出，有约4600个RTMS设备的数据正确且不重复的比例在0-0.02之间，占总RTMS设备数量的70%。从图中可以看出有很大一部分RTMS设备的数据质量存在严重问题，正确且不重复的比例非常低。同时，也有部分RTMS设备数据正确且不重复的比例较高，能达到80%以上。

同时，在被给出坐标的1538个RTMS设备中，数据正确且不重复的比例分布如图 2所示，图中横纵坐标的含义与图 1完全相同。在这一部分RTMS设备中，仅有30%的RTMS设备（461个）的数据正确且不重复的比例在70%以下，52%的RTMS设备（813个）的数据正确且不重复的比例在80%以下。这一部分RTMS设备的数据质量相对较好，可以用于后续的分析。

根据RTMS设备的坐标信息，画出RTMS设备的数据正确且不重复的比例与其空间分布如图 3所示。图中，每个数据点代表了一个RTMS设备，数据点的颜色代表该RTMS设备记录的数据中正确且不重复的比例（0-1之间）。从图中我们可以大致看出数据质量较高的RTMS设备在空间上的分布情况。可以看到，数据质量较差的RTMS设备主要分布在北京市二环以内的地区。而二环及以外的主要道路和快速路上的RTMS设备数据质量相对较好。

### RTMS数据的记录条数低于理论值

在删除重复记录和错误记录之后，对剩余记录的条数进行分析。绝大多数RTMS设备数据记录的时间范围在2016/4/25-2016/8/2之间。在该时间范围内，正常情况下每个RTMS设备每两分钟回传一条记录，即每天应收集 24×30 = 720 条记录。而根据总记录条数看，绝大多数RTMS设备的记录收集频率低于这一正常值。造成记录条数低于理论值的主要原因包括RTMS设备收集数据频率低于正常频率和其在较长一段时间内完全没有收集数据。我们以西北五环上编号为HI9290a的RTMS设备为例，统计该RTMS设备每个小时收集的记录条数并画出直方图，正常情况下应为每小时30条，如图 4所示。在该图中，横坐标代表时间，区间长度为一小时，纵坐标代表数据条数。可以看出，在大部分时间内，该RTMS设备都以小于2分钟每条的频率收集数据，但在2016/6/29到2016/7/18之间完全没有收集数据。该RTMS设备的数据情况与绝大多数RTMS设备的数据完整情况相似。

正常情况下，在2016/4/25-2016/8/2之间，按照每两分钟计算，每个RTMS设备应该收集71280条记录，而实际上大量RTMS设备回传数据的条数远远小于这个值。对所有提供坐标信息的RTMS设备收集记录条数占理论记录条数的比例进行统计，并画出所有RTMS设备记录条数占理论记录条数的比例的直方图，如图 5所示，此图的格式与图 1相同，唯一的区别是直方图的横坐标区间长度为0.01。从图中我们可以看出收集记录最多的RTMS设备收集的记录条数也仅仅占理论记录条数的60%，有大量的RTMS设备的这一比例在30%-40%之间。

我们对所有提供了坐标信息的RTMS设备，也画出了其记录条数占理论条数比例的空间分布，如图 6所示，从图中我们可以看出，大部分RTMS设备的数据完整比例在35%左右，少部分数据完整比例较低的RTMS设备集中在二环以内。

由于绝大多数RTMS设备收集数据的频率都小于理论值2分钟每条，因此我们考虑用一个更长的时间间隔来分析记录条数和理论记录条数的关系。例如，对上文中同一个RTMS设备，我们以每5分钟为一个时间区间，只考虑这个时间区间内是否有数据记录，并记录一小时内有数据记录的时间区间的个数。当数据完整时，一小时内的12个时间区间应该都有数据记录，即每天应收集 24×12 = 288 条记录。类似的，我们画出其每小时有记录的小区间的个数的直方图，如图 7所示。从图中我们可以看出，当时间间隔增大时，在大多数时间间隔内都有数据记录。如果以该时间间隔内的所有记录的平均值作为该时间间隔的记录，那么记录条数占理论条数的比例会得到提高。因此我们在后续的分析中，可以考虑采用更大的时间间隔进行分析。

### RTMS数据的其他问题

在去除重复和错误记录之后，很多RTMS设备的数据分布仍然不佳。同样，以编号HI9290a的RTMS设备为例，其全部记录中的速度数据分布直方图如图 8所示。从该分布直方图中我们可以看出，其速度数据记录几乎在60km/h-90km/h之间均匀分布。造成这一问题的原因很可能是RTMS设备自身的问题。在进行后续分析时，我们应该注意尽量避免使用这一类RTMS设备的数据。

另外，RTMS设备提供的流量数据也存在流量不守恒的问题。即上游路段的流量和下游路段的流量之间差距较大的情况。造成这一现象的原因可能是有关RTMS设备收集的数据出现问题。以路段粗细表示该路段日均流量的大小，画出各路段的流量分布图如图 9所示，从图中我们可以看出，一些路段存在流量不守恒的问题，即上游路段和下游路段在图中的粗细差别较大。在对数据进行分析时，我们需要谨慎对待出现不守恒问题的路段，根据实际情况确定造成这一现象的原因。

此外，RTMS数据还存在同一RTMS设备连续数周乃至数月收集的速度、流量值都相同，导致该RTMS设备记录中绝大多数数据都记录了同一个（速度，流量）值的现象。这种情况表示RTMS设备出现故障，因此这些RTMS设备的数据在后续的分析中不宜被使用。为了直观展示出现这种现象的RTMS设备的比例和其位置分布，将各个RTMS设备的数据中，（速度，流量）值的众数（即出现次数最多的记录值）占该RTMS设备的总数据条数的比例画在地图上，如图 10所示。从图中可以看出，二环以内的几乎所有RTMS设备都出现了这一问题，而在其他地区也有部分RTMS设备出现了问题。这也代表了原始数据的质量较差。在后续的分析中，我们应注意尽量避免使用这些RTMS设备的数据。

|  |  |
| --- | --- |
| 图 1 RTMS设备数据正确且不重复比例分布 | 图 2 有坐标信息的RTMS设备数据正确比例分布 |
| 图 3 RTMS设备数据正确且不重复比例空间分布 | 图 4 RTMS设备每小时记录条数（以HI9293d为例） |
| 图 5 RTMS设备数据完整比例分布 | 图 6 RTMS设备数据完整比例空间分布 |
| 图 7 RTMS设备每小时记录条数（HI9293d） | 图 8 RTMS速度数据分布（HI9290a） |
| 图 9 北京市各路段日均流量分布图 | 图 10 RTMS设备众数占总数据条数比例 |

## 2014年5月的RTMS数据说明

这部分数据的来源同样是北京市的RTMS设备，但是时间范围为2014/05/01-2014/05/31，并且记录的RTMS设备数量较少，仅有1522个，其中1031个有坐标信息。其记录的信息与上节中的数据基本相同，包括断面速度，断面流量等，也是每两分钟收集一次。这部分数据与上节中RTMS数据的最大区别在于已经去除了重复的数据和存在逻辑错误的数据，因此在原始数据中没有Repeat\_status和error\_status字段，数据样例如下：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| device\_id | report\_date | report\_hour | minute | volume | speed | time\_occp |
| HI9290a | 2014/5/1 | 0 | 4 | 128 | 74.5 | 32.5 |
| HI9290a | 2014/5/1 | 0 | 10 | 15 | 74.5 | 29.5 |
| HI9290a | 2014/5/1 | 0 | 36 | 72 | 81.5 | 29.5 |
| HI9290a | 2014/5/1 | 0 | 28 | 37 | 86 | 17 |
| HI9290a | 2014/5/1 | 0 | 16 | 28 | 62 | 24 |
| HI9290a | 2014/5/1 | 0 | 22 | 49 | 60 | 36 |
| HI9290a | 2014/5/1 | 0 | 48 | 50 | 78 | 14 |
| HI9290a | 2014/5/1 | 0 | 54 | 16 | 78 | 38 |
| HI9290a | 2014/5/1 | 0 | 14 | 140 | 67 | 21 |

因为这部分数据没有Repeat\_status和error\_status字段，经检验也不存在相关的数据重复现象和逻辑错误，因此我们不需要对这部分内容进行检验。接下来我们对这部分数据的数据条数与理论条数的比较等方面来检验这部分数据的质量。

在记录时间范围内，正常情况下每个RTMS设备每两分钟回传一条记录，即每天应收集 24×30 = 720 条记录。而根据总记录条数看，绝大多数RTMS设备的记录收集频率仍然略低于这一正常值，但相对上节中的数据，2014/05的RTMS数据条数更多，完整程度较好。对所有RTMS设备收集记录条数占理论记录条数的比例进行统计，并画出所有RTMS设备记录条数占理论记录条数的比例的直方图，如图 11所示，此图的格式与图 1相同。其记录条数占理论条数的比例的空间分布如图 12所示。从这两幅图中我们可以看出，大部分的RTMS设备的数据完整比例在80%左右，完整度比上节中的数据高。

另外，2014/05的RTMS数据同样存在流量不守恒的问题，上游路段和下游路段的流量之间差距较大。同样的，我们以路段粗细表示该路段在2014/05的日均流量大小，画出各路段的流量分布图如所示。

|  |  |
| --- | --- |
| 图 11 RTMS数据完整比例分布（2014/05） | 图 12 RTMS数据完整比例空间分布（2014/05） |

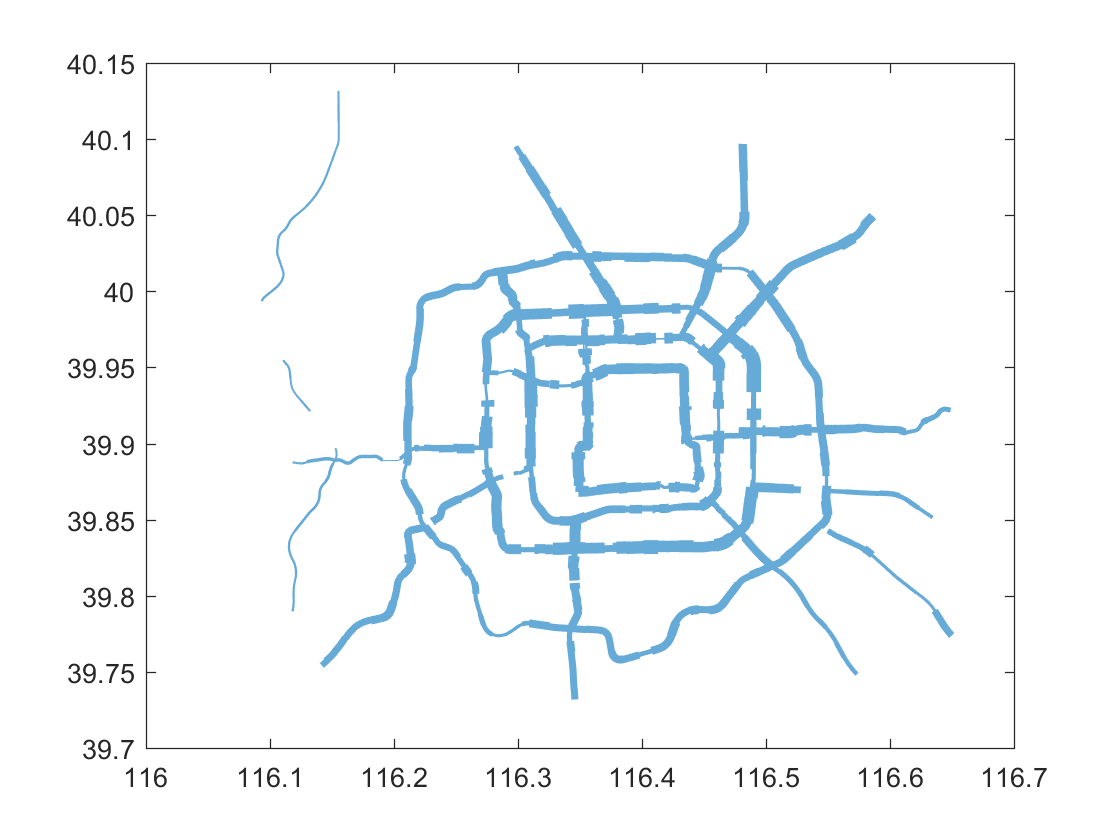


图 13 北京市各路段日均流量分布图（2014/05）

## 基于GPS的路段速度数据

基于GPS的路段速度数据是通过北京全城的浮动车实时GPS数据计算得到，记录的时间范围是2016/04/01-2016/07/29，相邻记录的时间间隔为15分钟，小于RTMS设备数据的2分钟。每条记录包含以下5个字段，数据样例如表 3所示：

* Date: 该条记录的日期；
* Time: 该条记录的时间；
* Linkid: 该条记录的路段编号（与之前的RTMS设备编号有对应关系）；
* Speed: 该时间该路段的速度值，用多辆浮动车GPS数据计算得到；
* Sample\_number: 该条记录用于计算速度所用的浮动车数量。

表 3 BTRC行程速度数据样例

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date** | **Time** | **Linkid** | **Speed** | **Sample\_number** |
| 2010-04-01 | 07:45 | 184568 | 9.4 | 1 |
| 2010-04-01 | 07:45 | 184569 | 9.4 | 1 |
| 2010-04-01 | 07:45 | 184573 | 9.4 | 1 |
| 2010-04-01 | 07:45 | 184574 | 9.4 | 1 |
| 2010-04-01 | 07:45 | 184578 | 9.4 | 1 |
| 2010-04-01 | 07:45 | 184589 | 9.4 | 1 |
| 2010-04-01 | 07:45 | 184712 | 9.4 | 1 |
| 2010-04-01 | 07:45 | 184785 | 9.4 | 1 |
| 2010-04-01 | 07:45 | 184786 | 9.4 | 1 |
| 2010-04-01 | 07:45 | 185422 | 9.4 | 1 |

相比RTMS数据而言，GPS路段速度数据重复比例很低，在1%以下。同时，其记录条数与理论记录条数也符合的较好。以与检测器HI9290a对应的快速路路段（西北五环处）的速度记录为例，其速度分布图如图 14所示。图中横坐标是速度区间，每个区间的长度为2km/h，纵坐标是该速度区间内的记录的频数。从图中可以看出，速度分布在80km/h处有单峰，这符合快速路上的车速分布的实际情况。



图 14 基于GPS的路段速度分布（以HI9290a为例）

## 第三方地图软件的实时拥堵程度数据

第三方地图软件，如百度地图和高德地图，采用不同的颜色来表示道路的实时路况。百度地图和高德地图均采用四个颜色等级代表不同的道路的拥堵等级，分别是：绿色--畅通、橙色或黄色--缓行、红色--拥堵和深红色--非常拥堵。虽然分类较为粗糙，但是仍然有很大的参考价值。我们通过网页截图来抓取该数据，并通过图像处理来获得信息。抓取数据时间范围为2016/04/01到2016/10/01，抓取频率为每五分钟一次。在抓取过程中由于网络连接等问题，存在抓取失败的现象，在这段时间里的总体抓取成功比例约在75%左右。抓取的道路范围包括北京市全部快速路和北京市三环苏州桥，蓟门桥和四环万泉河桥，学院桥之间的矩形区域的所有有路况记录的道路。

通过抓取截图并对图像进行处理，我们可以大致了解一个时间点上一段道路的拥堵情况，可以用于与RTMS速度数据和GPS路段速度数据进行比较，以验证数据的一致性。高德地图和百度地图均通过固定的颜色来表达道路拥堵状况，所以，我们通过像素颜色识别来确定该截图中所显示的道路拥堵情况。经过测试，我们的颜色识别结果较为准确。例如，图 15为高德地图的一张路况截图，从图中可以看到道路的拥堵程度用红黄绿等颜色表示。百度地图的路况截图与其相似。

我们可以基于三种不同的数据源来检验不同来源的速度数据的是否存在显著差异，判断数据的可信度，并为我们开展基于速度数据的研究工作提供参考信息。

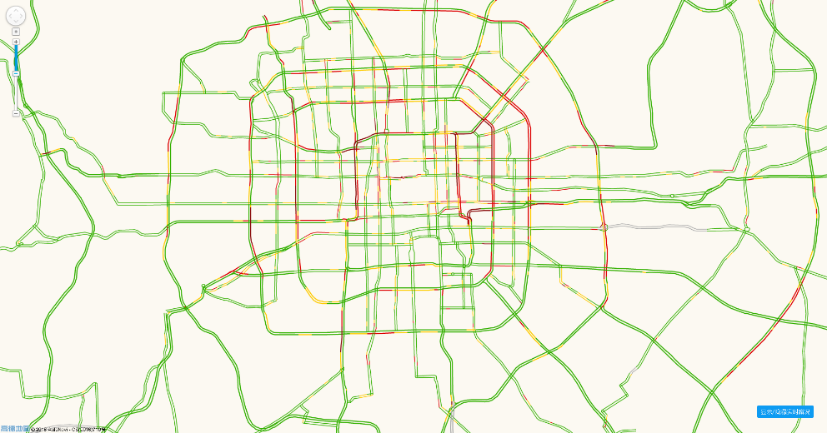


图 15 高德路况截图样例

## 交通事故数据

交通事故数据的数据源有两个，分别是122交管平台和103.9交通广播台。122交管平台的交通事故数据包含的是所有交通事故的报警记录。103.9交通广播台的交通事件数据记录的是广播台听众报告的事件信息，包括事故，拥堵，管制等事件，但事故数据的记录没有122交管平台完整。

在对每种类型的事件数据进行简单描述后，针对122交管平台的事故数据，我们对其事故类型、伤亡情况、肇事车型进行统计，以初步了解交通事件数据并对比两种数据源记录的数据间的差异。值得一提的是，为了对比两种数据源所记录的数据的差异，我们在统计时，使用数据的时间跨度是一样的，都是从2016/4/13到2016/6/15。

### 122交管平台事故数据

122交管平台记录的交通事故数据数据包含以下5个字段，数据样例如表 4所示：

* 案件编号：每个案件的唯一编号（重复报警的事故编号相同）；
* 报警时间：122交管平台收到报警的时间；
* 地点：事故发生地点的描述；
* 案件内容：事故具体内容，包含事故类型，涉事车辆，处理结果等；
* 车型：涉事车辆的车型。

表 4 交警122交管平台事故数据样例

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **案件编号** | **报警时间** | **地点** | **案件内容** | **车型** |
| 20160519006624 | 2016/5/19 16:00 | 京港澳高速主路六里桥客运站前 | 速腾与凌志刮撞,报警人称无人伤,已告知标划现场挪车，请与报警人联系核实处理。 | 小型汽车号牌 |

122交管平台的事故数据中案件内容的记录比较统一，主要描述事故类型（主要为刮撞、追尾和撞固定物3种）、伤亡情况和对事故的处理，数据记录中还包含对肇事车型的记录。在2016/4/13到2016/6/15期间，122交管平台记录的事故数据共有175482条。据此，我们统计了如图 16所示的案件内容中各事故类型所占比例和如图 17所示的各伤亡类型所占比例，以及如图 18所示的肇事车辆中的各车型所占比例。

|  |  |
| --- | --- |
| 图 16案件内容中各事故类型所占比例 | 图 17案件内容中各伤亡类型所占比例 |

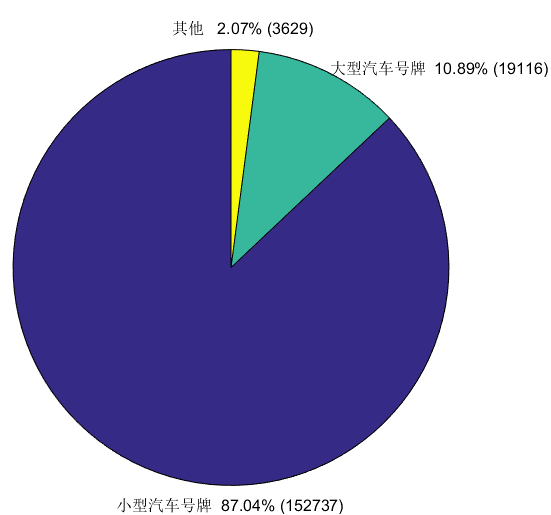


图 18肇事车辆中的各车型所占比例

从图14中可以看出，在各事件类型中，刮撞、追尾、撞固定物、其他事故类型的事故数量分别为148682、13574、6952、6274，占比分别为84.73%、7.74%、3.96%、3.58%。从图15中可以看出，在各伤亡类型中，无人伤、有人伤、死亡的事故数量分别为14874、28592、16，占比分别为83.70%、16.29%、0.01%。从图15中可以看出，在各肇事车辆类型中，小型汽车号牌、大型汽车号牌、其他的数量分别为152737、19116、3629，占比分别为87.04%、10.89%、2.07%。

### 103.9交通广播台事件数据

103.9交通广播台记录的交通事件数据包含以下3个字段，数据样例如表 5所示：

* 事件编号：事件的唯一编号；
* 时间：交通台收到事件报警的时间；
* 事件内容：事件发生地点和事件类型，包括交通事故，和交通拥堵、交通管制等。

表 5 交通台103.9事件数据样例

|  |  |  |
| --- | --- | --- |
| **事件编号** | **时间** | **事件内容** |
| 1291946 | 2016/7/17 11:05 | 动物园门前车多，造成西外大街西直门桥至动物园门前辅路、西北二环外环以及北展桥上南向北，排队缓慢。车辆可以绕行车公庄大街一线通行。 |
| 1291949 | 2016/7/17 11:25 | 北京西站北广场车多，造成莲花池西路西向东拥堵，队尾排过金家村桥。 |
| 1291952 | 2016/7/17 13:51 | 北京西站北广场持续车多，莲花池西路岳家楼桥至莲花桥西向东、西三环紫竹桥至莲花桥北向南，拥堵排队。 |
| 1291955 | 2016/7/17 14:08 | 北四环学院桥，东向西方向，外侧车道事故 |
| 1291958 | 2016/7/17 14:55 | 东三环双井桥，南向北方向，外侧车道事故 |
| 1291961 | 2016/7/17 14:58 | 东三环农展桥，北向南方向，外侧车道事故 |
| 1291964 | 2016/7/17 15:06 | 北五环主路来广营桥，西向东中间车道事故 |
| 1291973 | 2016/7/17 17:55 | 平安大街厂桥路口至南锣鼓巷路口，东西双方向车多拥堵；地安门外大街、德胜门内大街的南北双方向，也是拥堵排队。 |

103.9交通广播台记录的交通数据的事件内容的格式比较多样。事件内容中有的事件播报包含方向，有的事件播报不含方向,时间类型除了交通事故，还包含交通拥堵、交通管制等。在2016/4/13到2016/6/15期间，103.9交通广播台记录的事件数据共有1957条。在对事件内容进行统计时，我们采取如表2所示的统计字段。根据这些统计字段，我们统计了如图 19所示的事件内容中各事件类型所占比例和如17所示的事件内容中有无方向所占的比例。

|  |  |
| --- | --- |
| 图 19 事件内容中各事件类型所占比例 | 图 20事件内容中有、无方向所占比例 |

表 6 交通广播台103.9事件数据的统计字段

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **类别** | | **字段** | **逻辑关系** | **数量关系** |
| 有/无方向 | 有方向 | ‘向’，‘进京’，‘出京’，‘离京’ | 或 | 无方向=总数-有方向 |
| 无方向 | --- | --- |
| 事故/拥堵/事故+拥堵/管制/其他 | 事故 | ‘事故’‘刮撞’，‘追尾’ | 或 | 不包含（事故+拥堵）的事件 |
| 拥堵 | ‘缓慢’，‘拥堵’，‘车多’，‘车流量大’ | 或 | --- |
| 交通管制 | ‘管理’， ‘管制’， ‘封闭’ | 或 | 不包含（交通管制+拥堵）的事件 |
| 事故+拥堵 | ‘事故’+‘缓慢’， ‘事故’+‘拥堵’ | 或 | --- |
| 交通管制+拥堵 | ‘管理’+‘缓慢’，‘管理’+‘拥堵’，  ‘管理’+‘车多’，‘管理’+‘车流量大’，  ‘管制’+‘缓慢’，‘管制’+‘拥堵’，  ‘管制’+‘车多’，‘管制’+‘车流量大’ | 或 | --- |
| 其他 | --- | --- | 其他=总数-事故-拥堵-（拥堵+事故） |

从图17中可以看出，在各事件类型中，事故、拥堵、事故+拥堵、交通管制、交通管制+拥堵、其他的事件数量分别为947、376、214、159、9、252，占比分别为48.39%、19.21%、10.94%、8.12%、0.46%、12.88%。从图18中可以看出，有、无方向的事件数量分别为1836、121，占比分别为93.82%、6.18%。

# **检验数据的一致性**

在上一节中，我们详细描述了RTMS数据、基于GPS的路段速度数据、第三方地图软件的实时拥堵程度的数据和交通事件数据，并简单分析了这些数据的数据质量。在这一节中，我们将使用配对t检验和回归等方法，对不同数据源的速度数据进行一致性检验。

## RTMS速度数据和基于GPS的路段速度数据的一致性

由于计算GPS路段速度时采用的路段划分与RTMS设备的路段划分不同，我们只使用848个RTMS设备的速度数据以及与之对应的路段的GPS路段速度数据来检验这两种不同数据源的速度数据的一致性。每一次检验的两组样本分别是一个RTMS设备的速度数据和对应的路段的GPS行程速度数据，这些速度数据按照在同一个时间间隔进行配对。

为了更加直观的显示这两个数据源的速度数据关系，我们用部分RTMS设备的速度数据和与之的对应路段的GPS速度数据画散点图，如图 22所示。图中以RTMS速度数据为横坐标，以GPS路段速度数据为纵坐标，坐标的单位都是km/h。同时画出直线y=x作为参考。从图中可以看出，两个数据源数据的相关性不够强，同时，RTMS设备数据的分布在70km/h-90km/h之间分布特别集中，而GPS路段速度数据则相对比较分散。

我们认为，在不失去真实性的前提下，可以假设两组数据之差服从独立的正态分布，即满足进行配对t检验的假设条件。所以，我们对每个RTMS设备以及与之对应的路段的GPS速度数据，分别进行配对t检验来检验这两个数据源的速度数据的一致性。每次配对t检验的数据量都在1000条以上，80%的配对t检验的数据量在10000条以上。

在置信度为0.05的情况下，有94%的路段的配对t检验显示RTMS设备速度和对应路段的GPS行程速度存在显著差异。从前文给出的GPS行程速度数据分布和RTMS设备速度数据分布直方图中我们也可以看出，两者的分布相差较大，因此得到这一结果也不出人意料。进一步地，我们希望分析这一差异的来源是RTMS设备的系统性误差，还是RTMS设备工作不正常导致的数据异常。我们采用的方法是对两组数据进行回归。如果回归结果较好，且不同RTMS设备的速度数据和与之对应的GPS行程速度数据的回归系数比较接近，则能够说明配对t检验显示的显著差异主要来源为一个数据源的系统误差；如果回归结果也较差，那么数据不一致的现象极有可能是一个或两个数据源出现了问题。

我们分别对每个RTMS设备的速度数据和对应路段的GPS行程速度数据进行回归分析，并得到每次回归的R2值，采用的数据与配对t检验采用的数据相同。所有回归的R2值的分布直方图如图 21所示。图中横坐标为R2值的区间，每个区间长度为0.04；纵坐标为R2在这一区间内的回归的比例，比如从最左边的一条数据我们可以看出，有75%左右的RTMS设备与对应路段的回归的R2在0-0.04之间。从图中我们可以看出绝大多数回归的结果都非常不好，可以认为两组数据基本不存在线性关系。那么，这两组数据不一致的原因更有可能是因为数据出现问题。

我们还尝试通过对同一个路段RTMS速度数据和GPS速度数据的24小时变化来分析其相关关系。我们所用的数据是同一个RTMS探测器速度数据及其对应路段的GPS速度数据在每天各个时间段的平均值，这里以15分钟为一个时间段，所以每个路段有24×4=96个RTMS速度数据和同样数量的GPS速度数据。这一检验的结果也显示，大部分RTMS速度数据和对应的GPS速度数据相关性不强。各个RTMS设备的速度数据与对应的GPS速度数据的相关系数按照RTMS设备的位置画出，见图 23。

|  |  |
| --- | --- |
| 图 21 所有回归 R2值分布直方图 | 图 22 GPS速度与RTMS设备速度散点图 |
| 图 23 RTMS速度数据和GPS速度数据的相关系数分布 | |

## 第三方地图软件实时拥堵程度数据的相关性

我们可以通过类似的方法来检验第三方地图软件，如高德地图和百度地图的实时拥堵程度数据的一致性。由于两个地图软件对拥堵程度的划分都是按照路段通行速度进行的，而且分级相对比较均匀，我们使用数字1-4分别量化畅通，缓行，拥堵，非常拥堵四个拥堵程度等级，并定义路段的拥堵程度为该路段在截图中对应的所有像素的拥堵程度等级的平均值。

检验中采用数据是 2016/07/01到2016/07/31期间百度地图和高德地图的实时拥堵程度数据 ，为保证在不同拥堵程度下都有足够的数据量，所选路段是北京的一些常见拥堵路段，共有26个路段。

为了直观的展示两者的关系，下面我们画出北京西三环路段不同时刻两个地图软件报告的拥堵程度散点图，图中横坐标为高德地图报告的该路段拥堵程度，纵坐标为百度地图报告的该路段同一时刻的拥堵程度。从图 25中可以看出，西三环路段百度地图和高德地图报告的拥堵程度相关性较强。同时，我们发现，很多时刻百度地图报告道路畅通，而高德地图报告存在一定程度的缓行或拥堵。通过观察路况截图和其他路段回归结果的散点图，可以发现这一现象在不同路段的数据中普遍存在。

由于高德地图和百度地图对路段拥堵程度的划分方法不一致，我们不能期望两者在同一时刻同一路段得到的拥堵程度的数值相同，即不能使用配对t检验的方法来检验其拥堵程度的一致性。我们对每条路段两者的拥堵程度数据分别进行回归分析，即每一次回归的两组样本分别是某一路段不同时间的百度地图拥堵程度数据和高德地图同一路段的拥堵程度数据，两组数据按照时间相同进行对应。

对这26个路段分别进行回归分析，每次回归结果的R2值分布如图 24所示。图中横坐标是R2的区间，区间长度为0.075，左侧纵坐标是R2在这一区间内的回归的比例，图中的红线是累计频率曲线。从图中我们可以看出，大部分路段百度和高德地图拥堵程度记录都存在一定的相关性，二者一致性相对较好。观察回归结果不同的路段我们还发现，在道路拥堵程度波动较大的路段，两者路况报告的相关性较好。所以，在后续的分析中，我们可以只采用其中一个来进行分析，这里我们选用高德地图。

|  |  |
| --- | --- |
| 图 24百度高德地图路况回归R2分布 | 图 25 西三环百度和高德地图路况数据散点图 |

## 高德地图实时拥堵程度和GPS路段速度的一致性

下面我们用回归的方法检验高德地图和GPS行程速度数据之间的一致性，即每一次回归的两组样本分别是同一路段不同时间的的GPS行程速度数据和高德地图实时拥堵程度数据，两种数据在时间是一一对应的。回归采用的数据为北京市三环路上的30个路段的高德地图实时拥堵程度数据和GPS行程速度数据，时间范围为2016/07/01到2016/07/31。由于拥堵程度越高意味着速度越低，我们期待两组数据有负相关关系。

为了直观的展示两者的关系，我们选取北三环路安华桥到马甸桥路段，画出其各个时刻高德地图拥堵程度数据和GPS行程速度数据的散点图，以高德地图拥堵程度为横坐标，GPS行程速度为纵坐标（单位km/h），如图 27所示，从图中我们可以观察到两者的负相关趋势，即当拥堵程度增大时，速度趋向于减小，但是这一趋势并不明显。另外，两者不相符的情况，即拥堵程度较大而速度也较大，或者拥堵程度较小而速度较小，都较少出现。

对30个路段分别进行回归，所有回归得到的R2值都较小，均小于0.2，即相关性不强。30个路段回归得到的R2值如图 26所示。图中，横坐标是R2值的区间，每个区间的长度为0.02，左侧的纵坐标是R2值在某一区间内的回归的比例，图中的红线是累计频率曲线。另外，线性回归的一次项系数均小于0，与我们的认知相符。根据以上的分析，我们可以发现高德地图实时拥堵程度的数据和GPS路段速度数据的不存在显著的相关性。

|  |  |
| --- | --- |
| 图 26 高德地图和行程速度数据回归 R2分布 | 图 27 北三环高德地图和行程速度散点图 |

# **RTMS数据流量模式特征分析**

我们希望利用我们希望利用RTMS 设备收集到的速度数据和流量数据，来分析不同路段的流量特征，即路段流量、速度、密度三者的关系。由于需要定位到路段，我们只使用所有给出了坐标信息的1538 个RTMS 设备的数据进行流量特征的分析，并选用2016/04-2016/08的数据以方便RTMS数据和GPS数据进行对比。对每个RTMS探测器的速度和流量数据，我们依次作出如下的8张图来展示其流量特征模式。各张图的内容如下（以编号为HI9282c的设备为例，如图 28所示）：

* 探测器所在的位置。用红点标出该探测器在北京市所处的大致位置；
* 该探测器的速度和流量密度分布。以二维直方图呈现该探测器不同的（速度，流量）组合的条数；
* Q-u图。即速度-流量图，图中纵坐标为流量，单位为辆/2分钟；横坐标为速度，单位km/h；
* Q-k图。即密度-流量图，图中纵坐标为流量，横坐标为密度，通过流量/速度计算得到；
* RTMS速度分布图。显示该探测器的速度数据的分布；
* GPS速度分布图。显示该探测器对应的路段的GPS速度数据的分布；
* RTMS 24小时流量变化图。图中显示了该探测器24小时平均流量的变化。图中折线代表流量，时间间隔是15分钟，黑色柱体代表该15分钟内探测器的数据量
* RTMS和GPS 24小时速度变化图。图中显示了RTMS速度数据和GPS速度数据24小时平均流量的变化，并在图的下方标出RTMS和GPS速度的平均绝对误差，相关系数和回归方程。

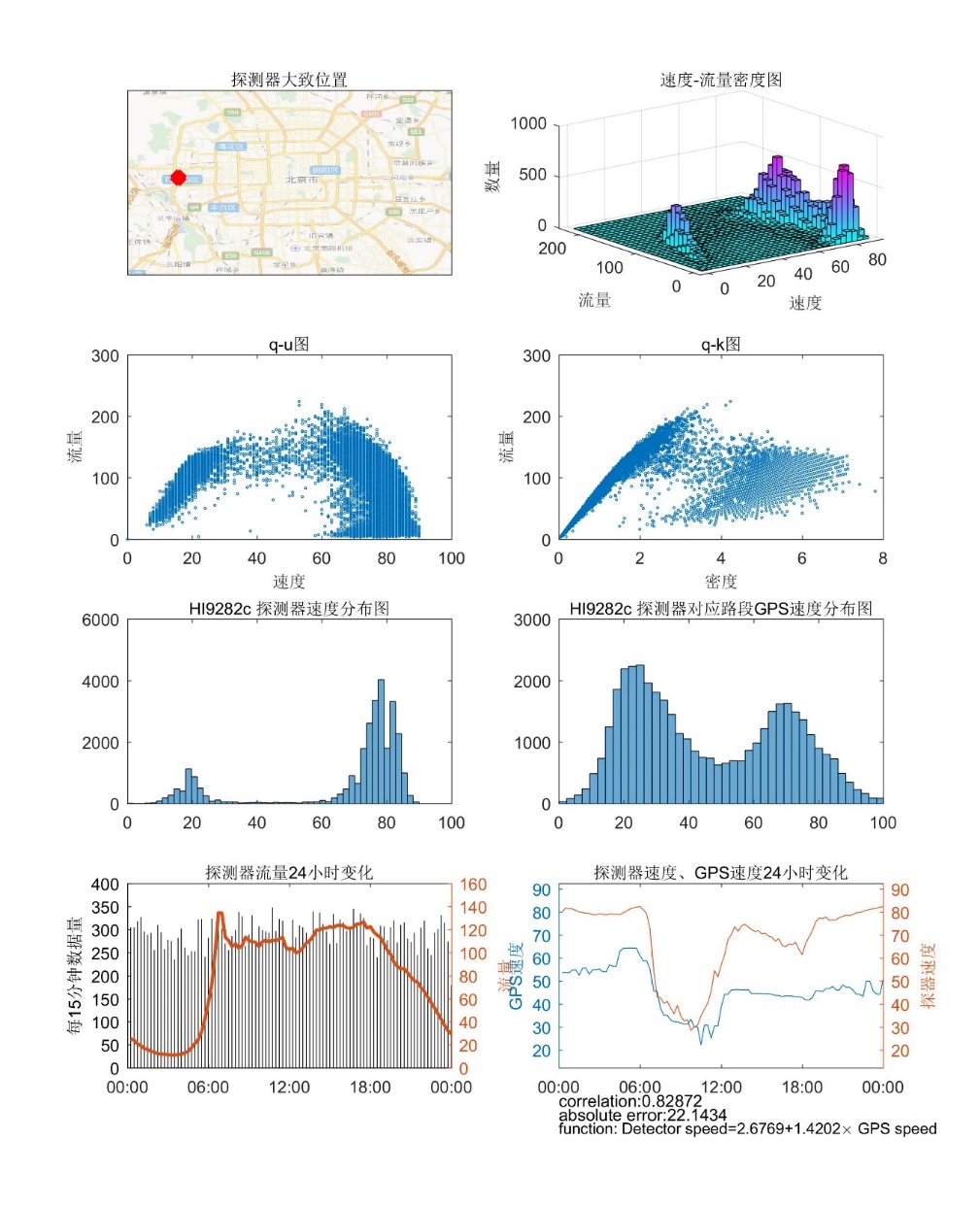


图 28 探测器数据特征图样例

我们对每个RTMS探测器画出这些图片来观察其流量模式的特征，发现除了图 28所展现的正常的数据特征以外，还有很多探测器的数据存在一定问题，主要包括（速度，流量）数据在图中呈现均匀分布、长时间记录同一个（速度，流量）数据和速度数据被截断，即有大量稍小于某一数值的速度数据，但没有大于这一数值的速度数据。三种数据问题的样例分别如图 29、图 30和图 31所示，同时，各种探测器数据特征在空间上的分布如图 32所示。

|  |  |
| --- | --- |
| 图 29探测器数据特征图（数据均匀分布） | 图 30 探测器数据特征图（数据大量重复） |
| 图 31 探测器数据特征图（数据存在截断） | 图 32 各探测器数据特征空间分布图 |

# **交通事故引发的交通拥堵在时空上传播规律的研究**

在这一部分中，我们主要研究交通事故引发的拥堵在时间维度和空间维度上是如何传播的，并量化交通事故引发的拥堵时长。这部分的内容包括模型建立和案例研究，其中，模型建立部分讲述如何建立目标函数和约束方程来确定交通事故引发的拥堵在时间维度和空间维度的传播范围，案例研究部分展示我们根据仿真数据和真实数据所得到的结果。

## 模型描述

交通事故引发的拥堵会随着时间的推移逐渐地向上游路段传播。如图 33所示，如果在路段部分发生了一起交通事故，那么随着时间的推移，路段的上游，即路段，路段，路段会依次受到该事故的影响。

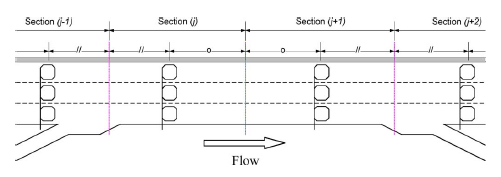


图 33 路段划分和车流方向示意图

通过RTMS设备我们可以得到每个路段在每个时间间隔的速度和流量数据。如果在某个路段发生了一起交通事故，那么我们就可以得到该路段及其上游路段在受到该事故影响下的速度，进而可以得到不受该事故影响下的速度。如表 7所示，假设路段在时间间隔发生了一起交通事故，且该事故在空间维度上的传播范围是从路段到路段1，时间维度上的传播范围是从时间间隔到时间间隔，我们把路段到路段1在时间间隔到时间间隔下的速度定义为受该事故影响下的速度。其中，表示受到路段在时间间隔发生的一起交通事故的影响，路段在时间间隔的受事故影响下的速度。去除受事故影响下的速度后的速度数据就称为不受该事故影响下的速度，相应地，我们可以得到不受事故影响下的速度的均值和标准差。如表 8所示，我们得到了路段在时间间隔发生的一起交通事故的不受事故影响下的速度的均值和方差，其中，表示路段在时间间隔的不受事故影响下的速度的均值和标准差。

表 7 受事故影响下的速度

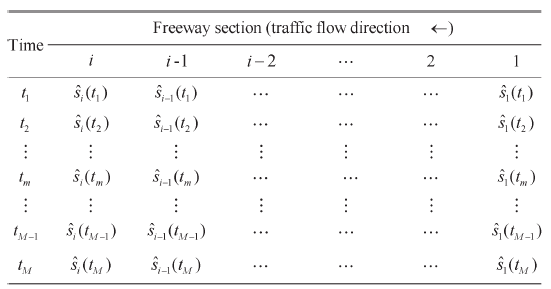
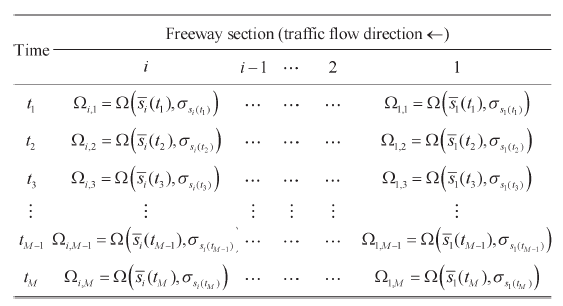


表 8 不受事故影响下的速度的均值和标准差



在得到受事故影响下的速度和不受事故影响下的后，我们根据如下公式（1）来判断路段在时间间隔的速度是否显著地小于该路段在时间间隔的平均速度，其中，为常数值。

（1）

,表示路段在时间间隔的速度显著小于该路段在时间间隔的平均速度，相反，,表示路段j在时间间隔的速度没有显著地小于该路段在时间间隔的平均速度。

如果仅根据来确定交通事故引发的交通拥堵的时空范围，会出现如图 34所示的结果。在图中，黄色的矩形块表示路段在对应的时间间隔受到交通事故的影响，蓝色的矩形块表示路段在对应的时间间隔没有受到交通事故的影响。因为速度存在波动以及一些其他的外部因素，受到事故影响的时空区域零散地分布，不符合交通拥堵在时间维度和空间维度上的传播规律。为了确定交通事故引发的拥堵在时间维度和空间维度上的传播范围，我们建立如下的目标函数和约束方程：

（8）

（2）

（3）

（4）

（5）

（6）

（7）

（7）

其中，交通事故引发的交通拥堵在空间维度上的边界是从事发路段到其上游的个路段，在时间维度上的边界是从事发时间到事发后的M个时间间隔。约束（1）表示路段在时间间隔发生了交通事故，约束（2）~（7）限定了该事故所引发的拥堵随着时间的推移从下游路段逐渐地向上游路段传播，即使得事故引发的拥堵在时间维度和空间维度上满足上游路段受到影响的开始时间和结束时间不早于下游路段，而且拥堵在时间维度和空间维度上的传播是连续的，不存在“空心”的情况。约束（8）表示决策变量是0-1变量，表示路段在时间间隔受到交通事故的影响，表示路段j在时间间隔没有受到交通事故的影响。

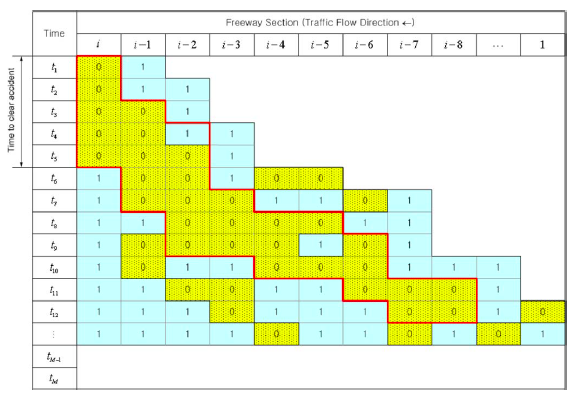


图 34 根据速度所得到的交通事故引发的交通拥堵的时空范围

在确定了事故引发的拥堵的时空范围后，我们可以根据公式（9）来确定该事故引发的拥堵时长

其中，表示路段的长度，表示路段j在时间间隔的流量，表示路段j在时间间隔的受事故影响下的速度，表示路段j在时间间隔的不受事故影响下的速度的均值。

## 案例研究

在本小节中，我们分别根据仿真数据和真实的RTMS数据来做案例研究，以验证我们建立的模型的有效性。

### 仿真数据

首先，我们生成如图 35所示的仅根据速度数据和公式（1）判断所得的交通事故引发的拥堵的时空范围，其中，事故发生在时空区域的左上角，红色矩形块表示路段在对应的时间间隔下的受事故影响下的速度显著小于该路段在相同时间间隔下的不受事故影响下的速度的均值，绿色矩形块表示路段在对应的时间间隔下的受事故影响下的速度并非显著小于该路段在相同时间间隔下的不受事故影响下的速度的均值。从图 35中可以看出，该事故引发的拥堵的时空范围并不符合在时间维度和空间维度上的传播规律。然后，通过我们的模型进行优化，我们得到了如图 36所示的交通事故引发的拥堵的时空范围。图中的结果满足了拥堵在时间维度和空间维度上的传播规律。

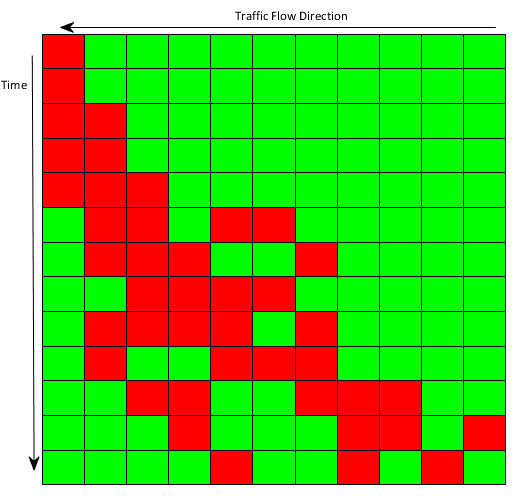


图 35 仅根据速度数据判断所得的交通事故引发的拥堵的时空范围（仿真数据）

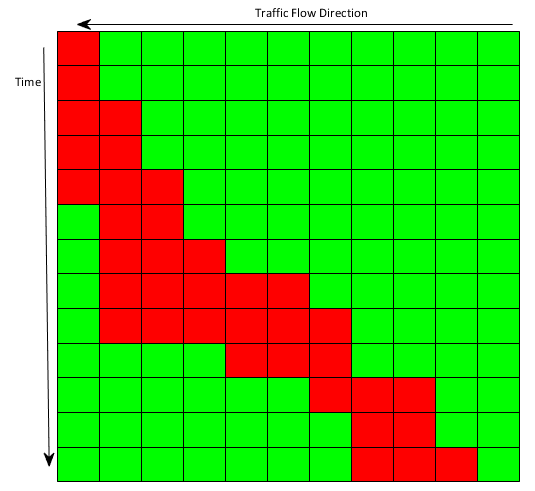


图 36 经过模型优化后所得的交通事故引发的拥堵的时空范围（仿真数据）

通过对比模型优化前后交通事故引发的拥堵的时空范围，我们可以看出，我们所建立的模型能够基于所获得的速度信息，更准确地确定交通事故引发的拥堵是如何随着时间的推移，从下游路段向上游路段传播的。

### 真实数据

我们基于RTMS数据和122交管平台的交通事故数据完成真实的案例研究，以观察交通事故引发的拥堵在时间维度和空间维度上是如何传播的。其中，RTMS数据记录了在路段在相应的时间间隔的速度和流量信息，122交管平台的交通事故数据记录了事故发生的时间、地点、事故类型、伤亡情况以及肇事车辆类型的信息。另外，我们还利用基于GPS的路段速度以及第三方地图软件实时拥堵程度数据进行校验。因为现有RTMS设备的路网存在偏差，我们接下来以我们手动定位的一起交通事故来进行真实的案例研究。

* **案例1**

我们手动定位到的一起交通事故的具体信息如表 9，该事故发生的时间是2016/6/1 19:41，发生的地点是西五环主路香山出口，事故类型是刮撞，伤亡情况是无人员伤亡，肇事车辆类型是大型汽车号牌。该事故在百度地图上的定位如图 37图 1所示。结合如图 38所示的RTMS路网，该事故引发的交通拥堵会沿着图中红色曲线的逆方向向事发路段的上游传播。

表 9 定位到的一起交通事故的信息

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 案件编号 | 报警时间 | 地点 | 案件内容 | 车型 |
| 20160601008532 | 2016/6/1 19:41 | 西五环主路香山出口 | 半挂车与半挂车刮撞，报警人称无人伤，请与报警人联系核实处理。 | 大型汽车号牌 |



图 37 该起交通事故在百度地图中的定位

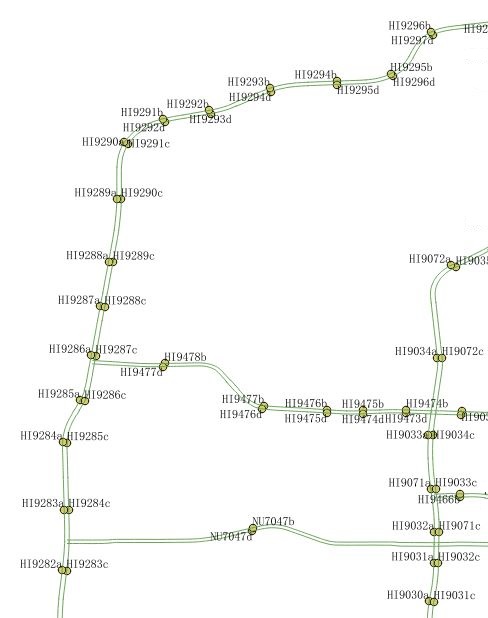


图 38 该起交通事故在RTMS路网下的定位

按照“模型建立”一节中描述的做法，我们在统计了受事故影响下的速度和不受事故影响下的速度的均值和方差后，根据（1）式可以得到如图 39所示的该交通事故引发的拥堵的时空范围（的值为0.2），经过模型优化后的时空范围如图 40所示。和仿真数据的结果一样，我们建立的模型能够更加准确地确定交通事故引发的拥堵的时空范围。

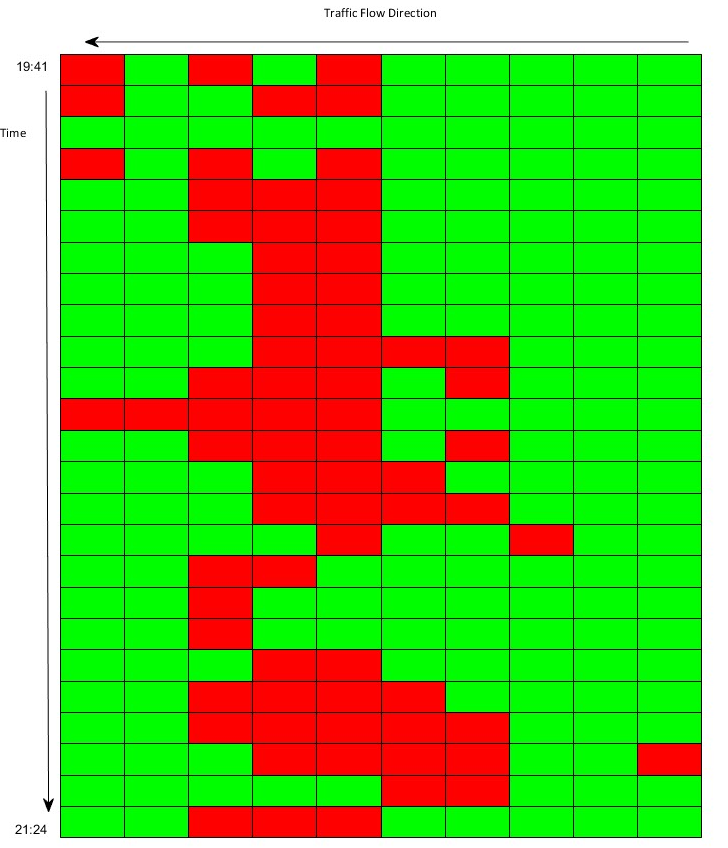


图 39 仅根据速度数据判断所得的交通事故引发的拥堵的时空范围（RTMS数据）

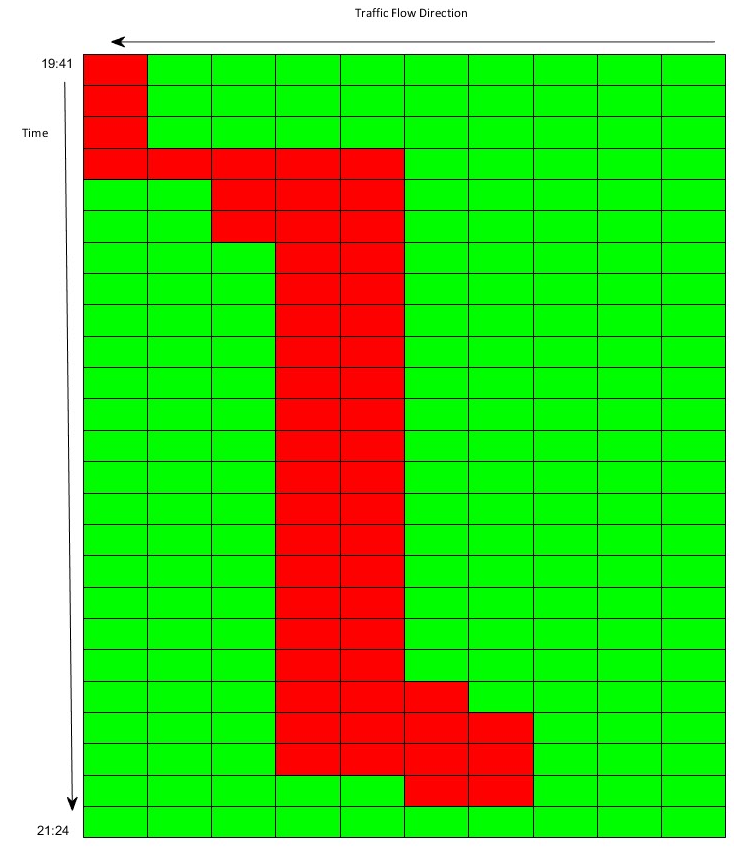


图 40 经过模型优化后所得的交通事故引发的拥堵的时空范围（RTMS数据）

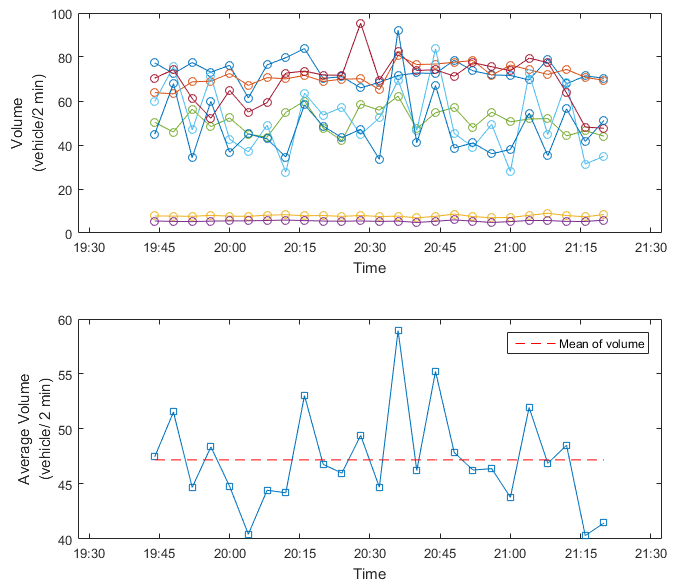


图 41 自事发始到事发后的100分钟内各路段的流量和各时段的平均流量

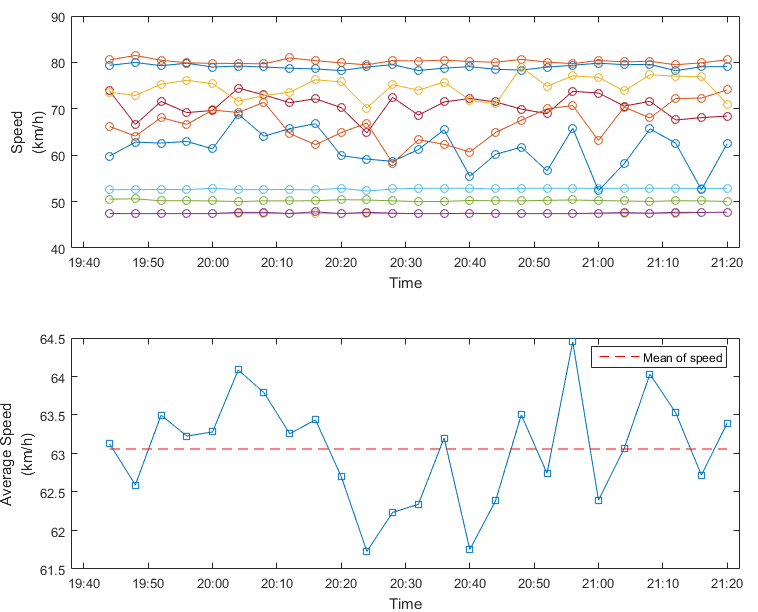


图 42 自事发始到事发后的100分钟内各路段的速度和各时段的平均速度

另外，我们利用公式（9）可以计算得到该事故引发的拥堵时长为252.2,结合如图 41所示的该路段自事发始到事发后的100分钟内流量数据和如图 42所示的相同时段的速度数据，可以计算得到，通过该路段的每辆车因为该事故造成的拥堵约为9.6分钟。

对于该事故，如果利用基于GPS的路段速度数据来做的话，可以得到如图 43所示的该交通事故引发的拥堵的时空范围（的值为0.2），经过模型优化后的时空范围如图 44所示。

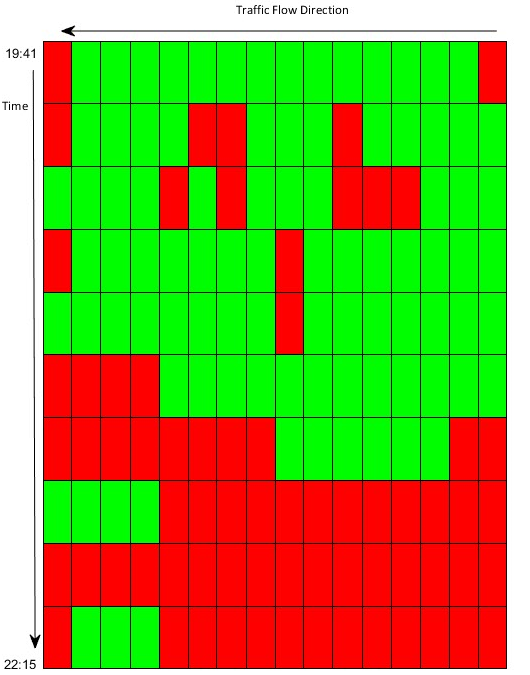


图 43 仅根据速度数据判断所得的交通事故引发的拥堵的时空范围（基于GPS的路段速度数据）

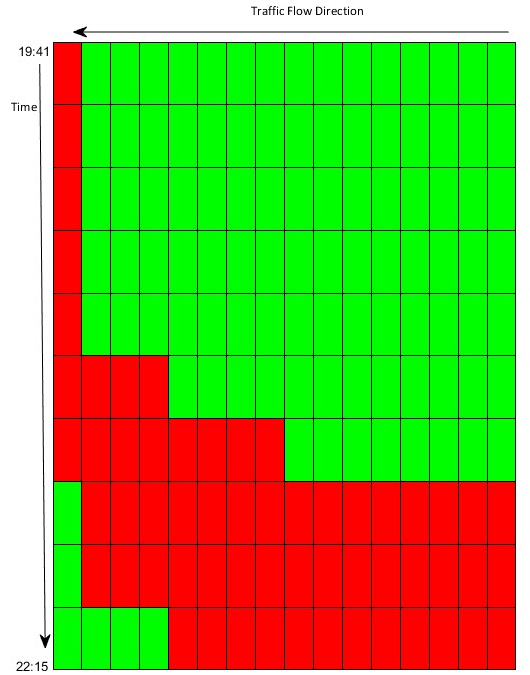


图 44 经过模型优化后所得的交通事故引发的拥堵的时空范围（基于GPS的路段速度数据）

* **案例2**

我们手动定位到的一起交通事故的具体信息如表 10所示，该事故发生的时间是2016/5/14 06:32，发生的地点是北五环主路广顺北大街出口，事故类型是车辆撞固定物，伤亡情况是有人员受伤，肇事车辆类型是小型汽车号牌。该事故在百度地图上的定位如图 45所示。结合如图 46所示的RTMS路网，该事故引发的交通拥堵会沿着图中红色曲线的逆方向向事发路段的上游传播。

表 10 定位到的一起交通事故的信息

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 案件编号 | 报警时间 | 地点 | 案件内容 | 车型 |
| 20160601008532 | 2016/6/1 19:41 | 西五环主路香山出口 | 半挂车与半挂车刮撞，报警人称无人伤，请与报警人联系核实处理。 | 大型汽车号牌 |



图 45 该起交通事故在百度地图下的定位

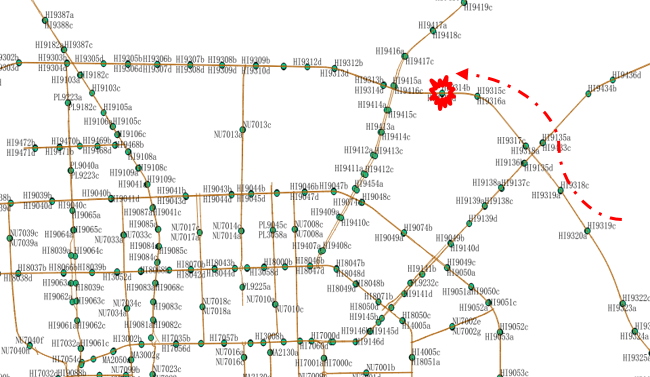


图 46 该起交通事故在RTMS路网下的定位

同样地，结合（1）中的模型和RTMS数据，我们可以得到如图 47所示的该交通事故引发的拥堵的时空范围（的值为0.2），经过模型优化后的时空范围如图 48所示。

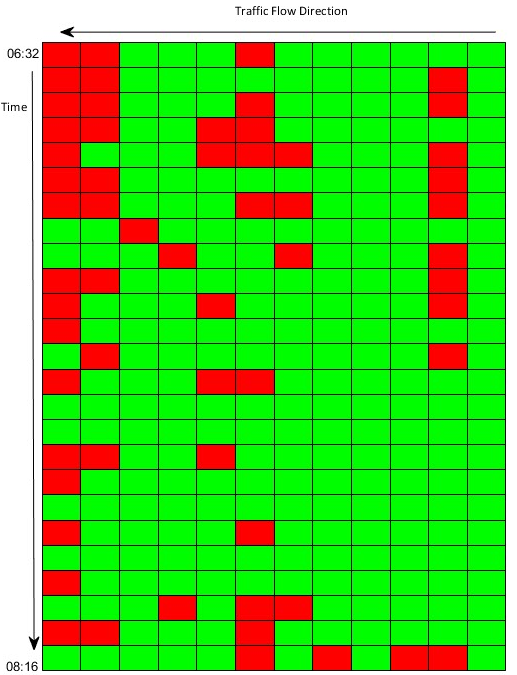


图 47 仅根据速度数据判断所得的交通事故引发的拥堵的时空范围（RTMS数据）

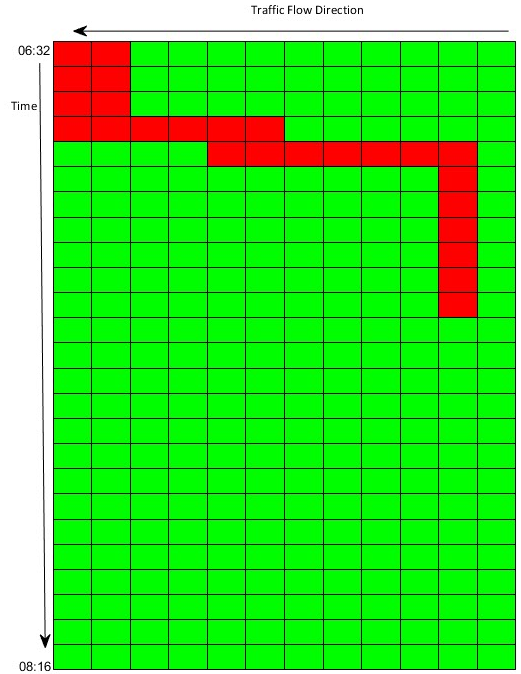


图 48 经过模型优化后所得的交通事故引发的拥堵的时空范围（RTMS数据）

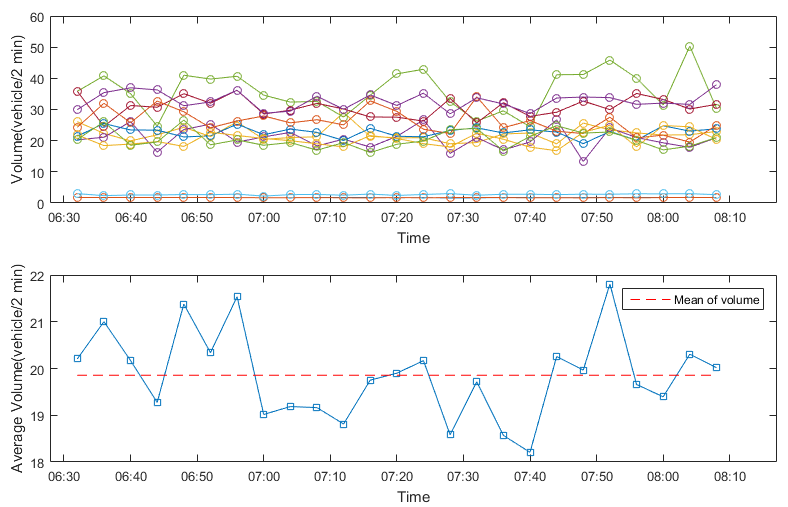


图 49 自事发始到事发后的100分钟内各路段的流量和各时段的平均流量

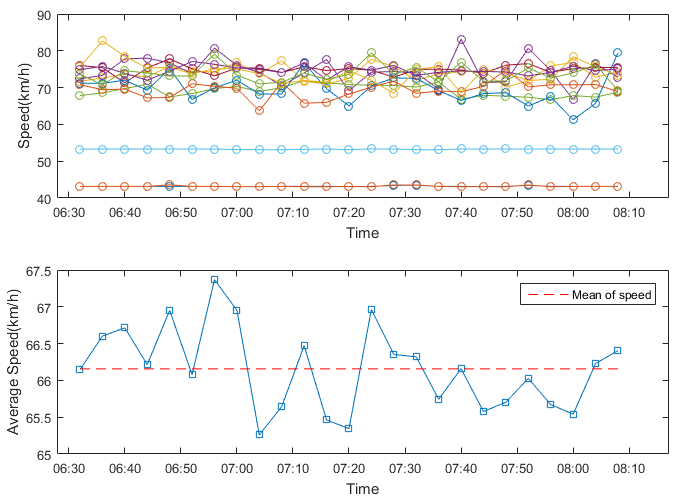


图 50 自事发始到事发后的100分钟内各路段的流量和各时段的平均流量

同样地，我们利用公式（9）可以计算得到该事故引发的拥堵时长为12.6, 结合如图 41所示的该路段自事发始到事发后的100分钟内流量数据和如图 42所示的相同时段的速度数据，可以计算得到，通过该路段的每辆车因为该事故造成的拥堵约为1.3分钟。

结合两个案例中的结果，结合我们采集的第三方地图软件的道路实时拥堵程度的数据，我们可以发现，RTMS速度数据和基于GPS的路段速度数据存在偏差，这体现在从仅根据速度数据判断所得的交通事故引发的拥堵的时空范围，无法观察出是否有事故发生。

# 参考文献

[1] Chung, Y. (2011). Assessment of non-recurrent traffic congestion caused by freeway work zones and its statistical analysis with unobserved heterogeneity. Transport Policy, 18(4), 587-594.

[2] Chung, Y. (2011). Quantification of nonrecurrent congestion delay caused by freeway accidents and analysis of causal factors. Transportation Research Record: Journal of the Transportation Research Board, (2229), 8-18.

[3] Chung, Y. (2012). Assessment of non-recurrent congestion caused by precipitation using archived weather and traffic flow data. Transport Policy, 19(1), 167-173.

[4] Chung, Y., & Recker, W. W. (2012). A methodological approach for estimating temporal and spatial extent of delays caused by freeway accidents. IEEE Transactions on Intelligent Transportation Systems, 13(3), 1454-1461.

[5] Chung, Y., & Recker, W. W. (2013). Spatiotemporal analysis of traffic congestion caused by rubbernecking at freeway accidents. IEEE Transactions on Intelligent Transportation Systems, 14(3), 1416-1422.

[6] Chung, Y. (2013). Identifying primary and secondary crashes from spatiotemporal crash impact analysis. Transportation Research Record: Journal of the Transportation Research Board, (2386), 62-71.

[7] Chung, Y., & Recker, W. W. (2015). Frailty models for the estimation of spatiotemporally maximum congested impact information on freeway accidents. IEEE Transactions on Intelligent Transportation Systems, 16(4), 2104-2112.