

硕 士 学 位 论 文

中期检查报告

**交通事故预测的**

**机器学习方法研究与应用**

作 者 姓 名 任红雷

作 者 学 号 SY1521103

学 科 专 业 软 件 工 程

专 业 方 向 软件工程与管理

学院指导教师 宋友 教授

培 养 院 系 软件学院

论文开题时间 **2017年 04月**

中期报告提交时间 **2018年 01月**

**目 录**

[1 课题简介 1](#_Toc502735105)

[1.1 课题背景 1](#_Toc502735106)

[1.2 主要研究内容和目标 2](#_Toc502735107)

[2 论文工作进展情况 3](#_Toc502735108)

[2.1 开题报告工作计划 3](#_Toc502735111)

[2.2 实际进展情况 3](#_Toc502735112)

[3 论文工作成果介绍 3](#_Toc502735113)

[3.1 课题所实施的解决方案介绍 3](#_Toc502735117)

[3.2 开题报告中所列关键问题的解决情况 9](#_Toc502735118)

[3.3 创新性的方法、技术、成果 12](#_Toc502735119)

[4 论文后期工作及进度安排 13](#_Toc502735120)

[5 尚存的问题及措施 13](#_Toc502735121)

[5.1 论文后期工作存在的困难和问题 13](#_Toc502735124)

[5.2 准备采取的措施 13](#_Toc502735125)

[5.3 如期完成全部论文工作的可能性 13](#_Toc502735126)

[**参考文献** 14](#_Toc502735127)

1 课题简介

1.1 课题背景

随着现代交通的发展和城市化趋势，交通领域的问题愈发凸显，在一些大城市如：北京，上海，广州等，交通事故，交通拥堵越来越多，据统计2016年中国交通事故约50万起，因交通事故而死亡的人数约25万人[1]。此外，国务院在第十二个五年规划纲要中明确提出要将交通事故致死率减少36%[2]。这些证明了交通安全问题已变成一个不可忽视的问题，为了减少交通事故以及相关的生命和财产损失，对交通事故准确的实时预测研究和应用显得很有必要。

同时随着大数据时代的到来，传统的交通研究也开始从基于统计规律的研究转变为基于交通大数据的研究。越来越多的异质类型的数据如交通流探测器产生的：浮动车GPS数据、线圈车流量数据，地铁、公交的刷卡数据，基于互联网和社交媒体产生的LBS数据。这些不同的数据源每天产生着大量的数据，为研究和和解决交通问题提供了无限的可能。而近年来兴起的机器学习和深度学习技术，也为我们分析和解决交通问题提供了新思路和有效的手段。因此，采用基于城市交通大数据的深度学习模型的实时城市交通事故预测研究显得很有必要。

从学术研究领域看，本研究课题属于城市计算领域下的智能交通子领域。城市计算是一个新兴的交叉学科领域，计算机学科与传统的交通、城市规划、能源、环境、经济和社会学等多个领域在城市空间的交汇而不断发展壮大[3, 4]。由于交通事故造成了大量的生命和财产损失，对交通事故的实时预测研究，将可以使得交通管理者们实时调整安全资源的配置、实时发布道路安全信息、使得驾驶员和行人可以选择安全性更高的道路和区域出行，避免高风险的事故区域，从而改善道路的安全性，避免生命和财产损失。

目前许多交通事故方面的研究都偏重于利用传统的统计方法对影响交通事故严重性的因素进行分析，对于实际指导和解决交通事故问题的意义不大。另外一些研究偏重于粗粒度的交通事故热点区域的聚类和发现，由于其热点区域的颗粒度较大，并且采用的统计方法较为简单，因此无法复杂的数据中学习在高维条件空间下（如区域、天气、节假日、功能区、道路类型等）的规律。另外，不少基于深度学习的交通流预测领域的研究也对基于深度学习的城市交通事故预测的研究有着相当重要的方法指导作用。目前许多考虑空间和时间特征的深度学习交通流预测模型已经被提出，而对于实时交通预测目前的研究重点还是统计方法。尽管已有Chen等提出了一些实时的深度学习城市交通事故预测模型[5]，但是其考虑的还十分不全面，忽略了空间、时间、天气等对交通事故影响很大的因素。

综上所述，本课题针对实时城市交通事故预测问题，本研究旨在探索可以较准确的预测事故的机器学习模型，同时尽可能的考虑这些重要的因素，并给出它们的重要性排序。

1.2 主要研究内容和目标

1．设计一种具有较高准确率和召回率的交通事故预测的机器学习方法

利用相关性分析找出了交通事故时空模式，以及相关影响因素。并在此基础上，设计了一种能够较精准的预测城市区域交通事故的机器学习方法。

2．基于预测性能对影响交通事故的因素进行了分析和排序

单个的交通事故较为随机，通过统计方法分析出影响交通事故的重要因素，对改善城市交通管理策略、以及进行更好的事故预测至关重要。基于格兰杰因果系数，通过比较不同特征和因素组合的情况下的预测性能，对影响交通事故的主要因素进行了重要性排序。

3．设计异质交通大数据的收集和存储策略

根据不同类型的交通数据（出租车轨迹数据、交通事故数据、天气、空气质量等）的特点和数据量，设计了较为通用和高效的收集和存储策略。

4．原型系统的开发与验证

通过建立原型系统，使用前述所提出的交通事故预测方法、框架、存储方案，基于 B/S(浏览器/服务器)架构，设计了一种较为通用的、具有良好交互性的，实时交通事故可视化、及事故预测平台，方便交通管理者和决策者进行实时的交通决策。本课题从系统功能和运行效率两方面验证工作的有效性，为一般性的交通事故预测平台提供参考依据。

2 论文工作进展情况



2.1 开题报告工作计划

表 1 论文工作计划

|  |  |
| --- | --- |
| 文献调研，确定研究课题，撰写论文开题报告 | 2017/03-2017/04 |
| 通过文献研究法对影响交通事故的因素进行确定，并获取相关数据 | 2017/04-2017/06 |
| 通过文献研究，初步确定若干机器学习模型 | 2017/06-2017/07 |
| 准备数据、对深度学习模型进行训练，并与其他模型进行比较 | 2017/07-2017/09 |
| 撰写论文中期报告 | 2017/09-2017/12 |
| 研究影响模型的关键因素；开发可视化的事故预测平台 | 2017/09-2017/12 |
| 撰写毕业论文 | 2018/01-2018/03 |

2.2 实际进展情况

表 2 实际进展情况

|  |  |  |
| --- | --- | --- |
| 文献调研，确定研究课题，撰写论文开题报告 | 2017/03-2017/04 | 按计划完成 |
| 通过文献研究法对影响交通事故的因素进行确定，并获取相关数据 | 2017/04-2017/06 | 按计划完成 |
| 通过文献研究，初步确定若干机器学习模型 | 2017/06-2017/07 | 按计划完成 |
| 准备数据、对深度学习模型进行训练，并与其他模型进行比较 | 2017/07-2017/09 | 按计划完成 |
| 撰写论文中期报告 | 2017/09-2017/12 | 按计划完成 |
| 研究影响模型的关键因素；开发可视化的事故预测平台 | 2017/09-2017/12 | 按计划完成 |
| 撰写毕业论文 | 2018/01-2018/03 | 进行中 |

3 论文工作成果介绍



3.1 课题所实施的解决方案介绍

3.1.1 总述

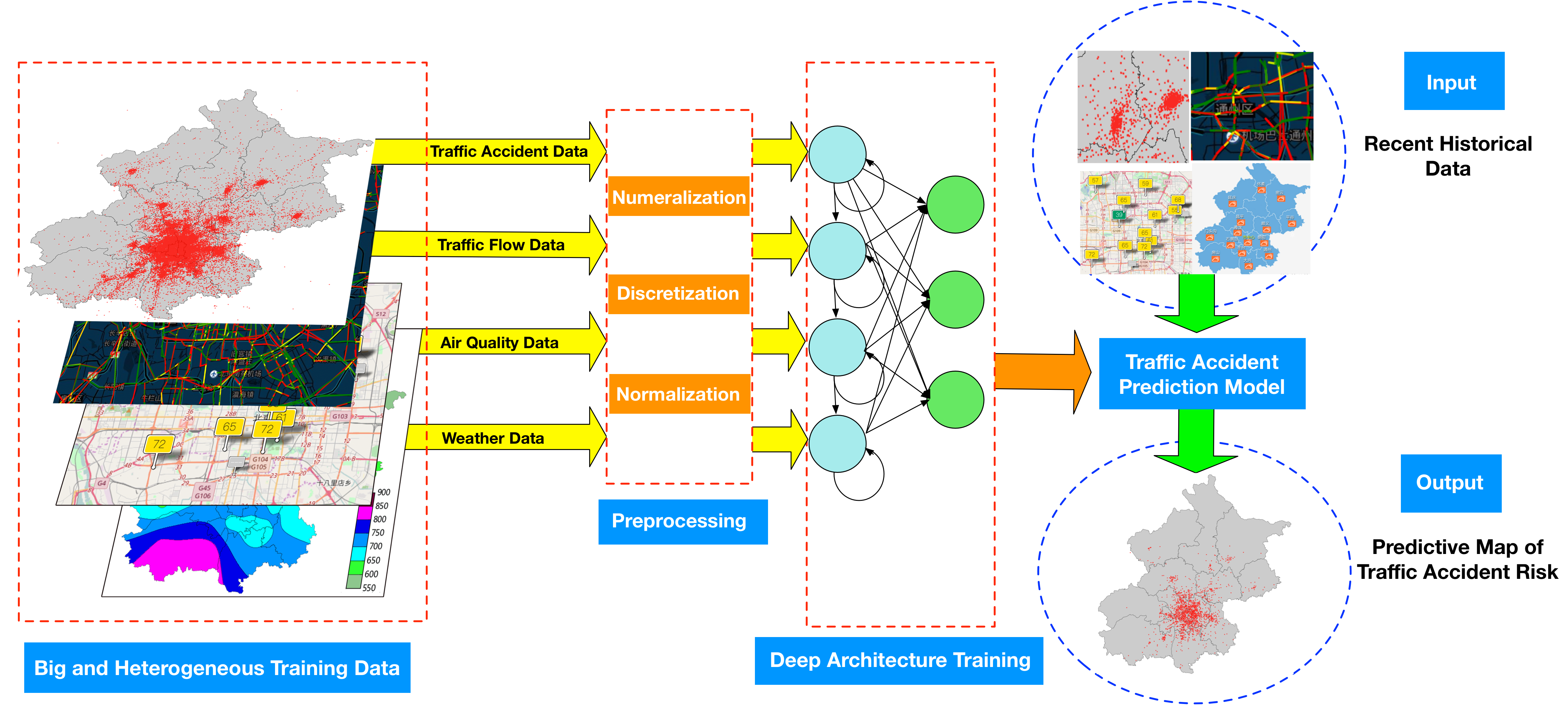
在大数据和深度学习的帮助下，尽管高性能的 实时交通流预测已经使得人们可以通过选择相对不拥堵的路线，来避免交通堵塞。而另外一个重要的问题，交通事故，尽管已经造成了巨大的经济损失和人员伤亡， 却一直没能被很好的理解和解决。基于上述的考虑，本课题收集了与交通事故相关的异质大数据并且构建了一个基于深度循环神经网络的交通事故预测模型。此模型集成了交通事故、交通流、天气条件和天气质量的历史短期和周期性的特征，以便于模型能更有效地学习交通事故的规律。

3.1.2 研究思路

研究的思路是，首先通过相关性等分析和找出与事故可能相关的因素，例如交通流等。接着，通过学习和文献调研，选取若干基线模型作为准确率判断的基础。之后根据事故发生的规律和特点，设计合适的机器学习模型来对交通事故进行预测，并将此模型与基线模型的预测准确率进行对比，不断的改进模型的准确率。

3.1.3 解决方案介绍

本课题的主要目标是进行较为准确的城市交通事故预测。为了达到此目的，本课题设计了如图1 所示的机器学习预测方法架构。



**图1 交通事故预测方法的流程图**

如图1所示，首先，我们收集了与交通事故相关的异质大数据。它包括了交通事故、交通流、天气和空气质量的数据。其次，为了便于处理我们采用了数值化、离散化和归一化的方法对数据进行了预处理，并且将预处理后的数据输入到了深度模型中以便于训练。经过对数据的训练之后，我们将近期的历史数据输入至训练好的模型。之后，我们就可以从模型的输出获得实时的交通事故预测地图。

1. 数据收集解决方案

为了准确的预测交通事故，我们收集了与交通事故相关的交通流、天气、空气质量的异质大数据，它们具体如下:

• 交通事故数据

我们收集了北京从2016年1月-2017年 2月一共948,367条交通事故记录数，每条事故包括事故发生时间、事故的经度和纬度三个属性。这些交通事故都是发生在车与车之间的，因此行人移动的特征与事故关系不大。

• 交通流数据

我们收集了2016年8月北京的出租车每隔5分钟上报一次的GPS位置和行驶平均速度数据。

• 天气数据

我们收集了北京2016年1月-2017年2月每天的天气描述数据。如：晴、小雨、小雪等描述。

• 空气质量数据

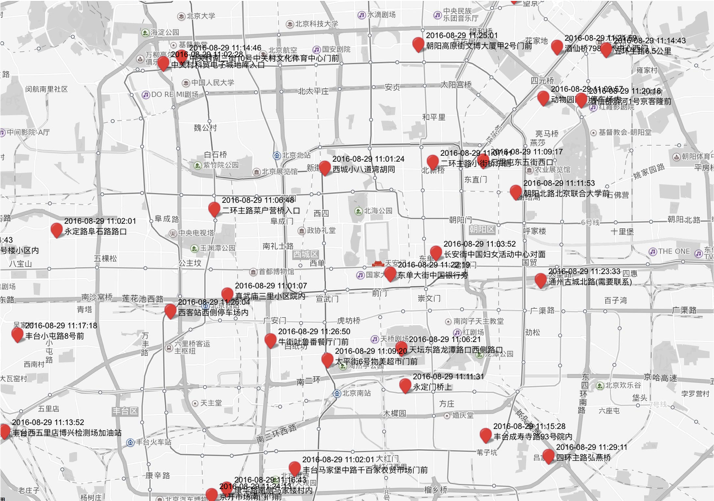
我们收集了与天气数据对应时间段，每天的PM2.5指数数据。

1. 数据预处理方案

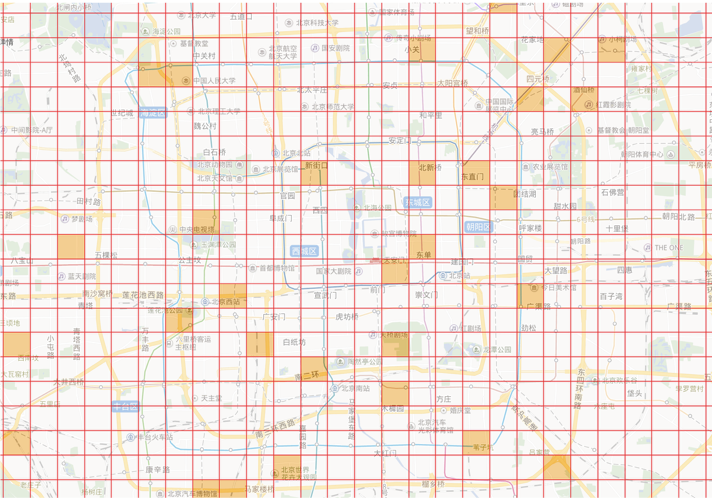
在我们构建深度线性模型并预测交通事故之前，一个合适的数据结构对于构建准确的和高性能的模型至关重要。首先我们对原始数据进行了一定的预处理:例如数值化，离散化，标准化等等。

对于交通事故数据，我们首先将数据根据时空信息进行了离散化。对于时间维度，我们选择60分钟作为时间间隔。对于空间维度，我们将事故的位置映射到1000米\*1000米的网格中。在我们的实验中，对于北京的数据而言，每个网格的纬度跨度为∆dlat = 0.0084，经度 跨度为∆dlng = 0.012。原始事故数据如图2所示，我们将交通事故数据进行了空间离散化，如图3所示。

将数据按照空间和时间离散化后，我们将原本连续的交通事故时空信息转换成了一个带有区域索引和时间索引的离散矩阵形式，记为S。对于特定区域r和特定时间段t，发生的事故总量可以被表示为矩阵S的一个元素，即Sr,t。为了表述方便，下文我们将矩阵的元素Sr,t记为Sr(t)。



**图2 2016年8月29日11：00-11：30的交通事故原始数据可视化**



**图3进行空间离散化之后的事故可视化**

对于交通流数据，我们的原始数据是每隔5分钟出租车上报的GPS位置和行驶速度数据。首先我们按照与交通事故的位置数据同样的离散化方法将交通流数据的离散化。这样我们得到的交通流矩阵的每个元素为在特定区域r和特定时间段t内的上报的出租车GPS位置和行驶 速度集合。之后，我们对每个矩阵元素的行驶速度集合分别进行平均从而得到一个新的平均行驶速度矩阵V， 其中每个矩阵元素Vr (t)为在特定区域r和特定时间段t内经过的出租车的平均速度。

因为节假日的交通流状况和交通事故的模式和平时具 有很大的不同，除了与交通事故矩阵和交通流矩阵，我 们还构建了矩阵H来标注每个交通事故矩阵元素Sr(t)对应的时间t是否处于节日或者假日。我们是根据中国法 定放假的节日来定义节假日的。具体而言，包括：元旦，春节，清明节，五一劳动节，端午节，国庆节，中秋节。另外，由于一天中不同的时间段的交通事故模式，相差很大，如高峰时间较正常时间事故要多很多，我们根据工作时间模式和中国人的生活方式，按照如下方式对事故矩阵元素Sr (t)所在的时间段t进行标注，形 成了与交通事故矩阵对应的时间段矩阵G: 00:00– 06:59 (午夜)、 07:00–08:59 (早高峰)、 09:00– 11:59 (早上工作时间)、 12:00–13:59(中午休息时间)、 14:00-16:59 (下午工作时间)、 17:00–19:59 (晚高峰)和 20:00–23:59 (夜间)。

对于天气数据，我们收集到的原始数据是每天的天气 描述，但是这种文字描述很难输入到计算机中，并且对于建立预测模型是没有太大帮助的。因此，必须将文字 转化成计算机容易输入和处理的数值形式。我们主观地 根据天气对交通影响的严重性，将天气描述转化成了表 1对应的天气严重性数值。如果描述为混合天气，如小雨转晴等，其天气严重性为小雨和晴对应严重性的平 均。在本例中，天气严重性为 (2 + 0)/2 = 1。对应交 通事故矩阵，我们构建了与之每个元素对应的天气严重性矩阵W 。举例而言，元素Wr (t)代表区域r和时间t对 应的天气严重性。

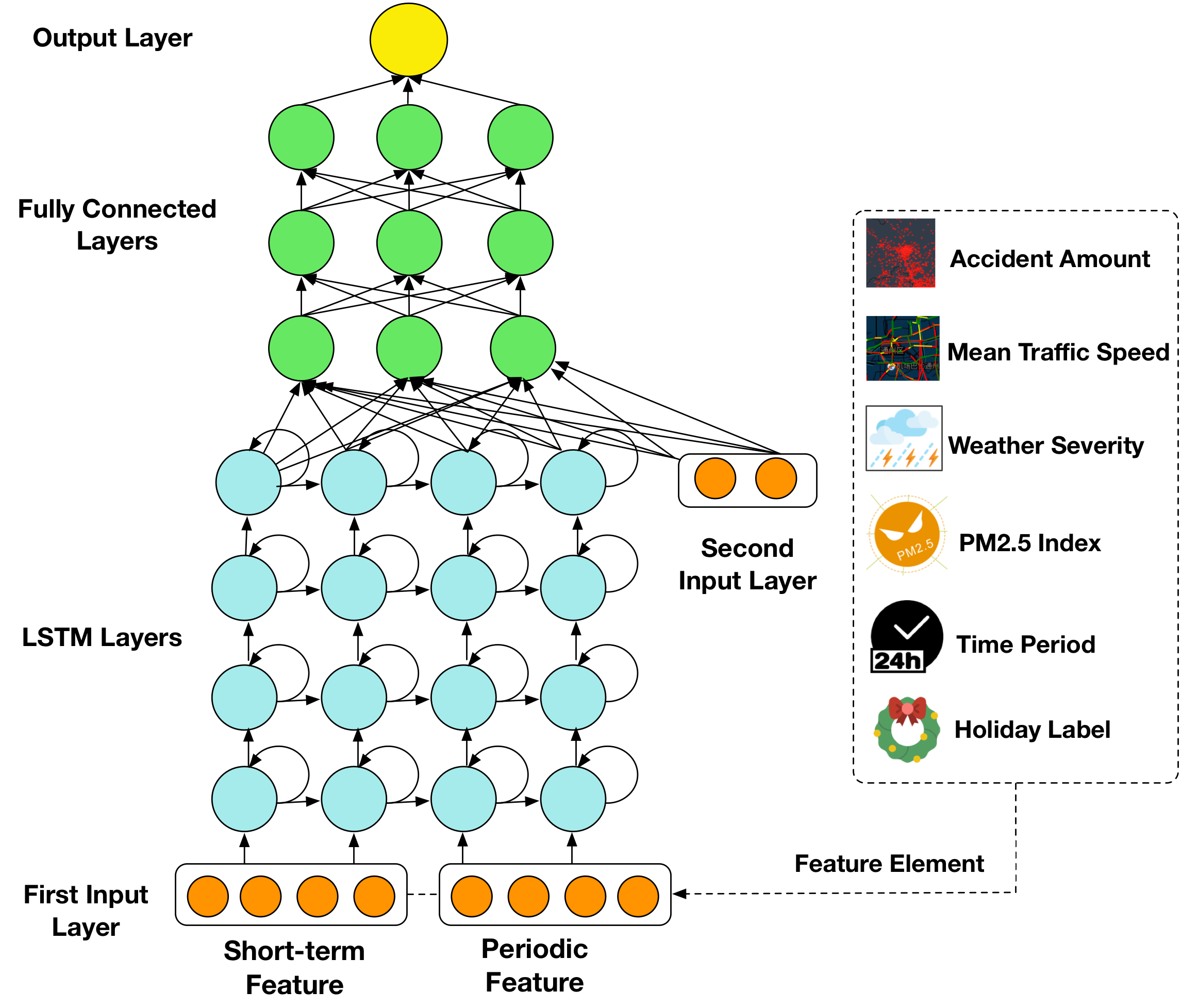
对于空气质量数据，我们提供将其除以最大PM2.5指数(500)来对数据进行归一化。即归一化后的PM2.5指数=PM2.5指数/500.0。类似于天气严重性矩阵W，我们利用归一化的数据，构建了与交通事故矩阵对应的空气质量矩阵A。举例而言，元素Ar(t)代表在区域r和时间段t内的标准化空气质量指数。经过预处理后，我们得到了6个矩阵，它们是:交通事故矩阵S和与之元素对应的交通流矩阵V，节假日矩阵H，时间段矩阵G，天气严重性矩阵W，空气质量矩阵A。我们将这些矩阵叠加到一起，形成一个新的矩阵I，对于特定区域r和特定时间t，Ir(t)=[Sr(t), Vr(t), Hr(t), Gr(t), Wr(t), Ar(t)]T。深度学习模型的输入序列元素就是矩阵的每个元素Ir(t)。输出就是该区域，该时段预测发生事故的数量。

**表 1 天气描述与其对应的天气严重性**

|  |  |
| --- | --- |
| 天气描述 | 天气严重性 |
| 晴, 浮尘, 阴, 多云 | 0 |
| 霾，雾 | 1 |
| 小雨，阵雨 | 2 |
| 中雨, 小雪 | 3 |
| 大雨, 雨夹雪, 暴雨 | 4 |
| 中雪、大雪 | 5 |

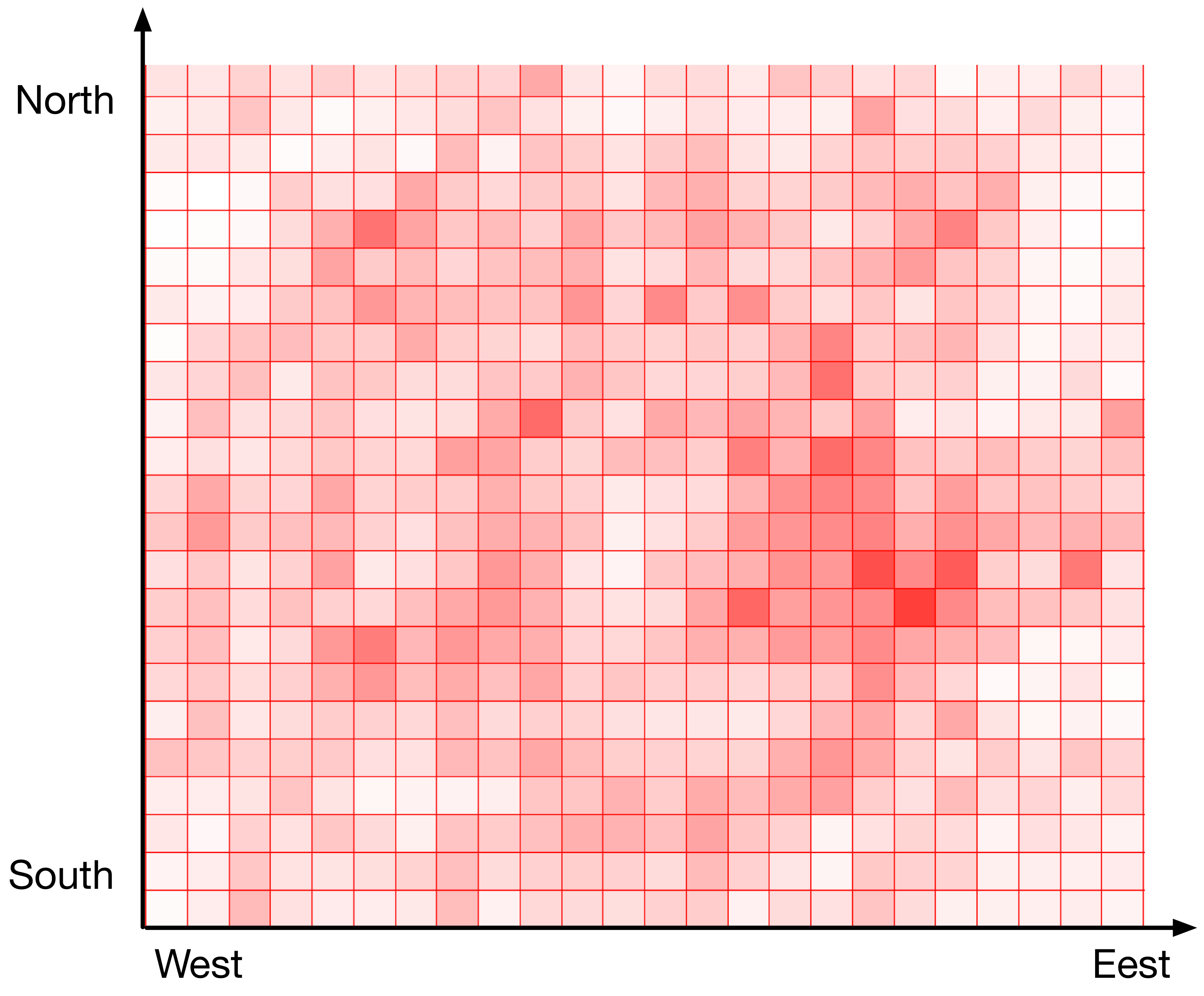
3. 深度学习模型

本课题提出的基于LSTM的交通事故预测的深度学习模型如图3所示。图中左边部分显示了神经网络的结构。包含了2个输入层，4个LSTM层，3个全连接层和1个输出层。



**图3 基于LSTM的交通事故预测方法的深度模型**

第一输入层由短期特征、周期特征组成，每个特征包含若干特征元素，如图3右半部分所示，每一个特征元素都是一个包含预处理后的数据的向量，这些预处理后的数据包括: 交通事故数量、平均交通速度，天气严重性，归一化后的PM2.5指数，时间段和节假日标注。第二输入层包括该数据的经纬度原始值，这是为了将事故的空间模式进行编码，让模型进行学习，其深层次原因是因为我们基于对事故空间发生频率的统计。发现不同区域, 事故的发生频率差异非常大。如图4所示，我们统计了2016年全年的交通事故的空间分布，每个格为1km\*1km区域。红色程度越深，代表发生事故越频繁，图中红色最深的地方在国贸附近。



**图4 2016年全年的交通事故的空间分布图**

3.2 开题报告中所列关键问题的解决情况

在开题报告中，提出了以下关键问题：

1. 应该考虑哪些影响交通事故的因素加入到深度学习模型。

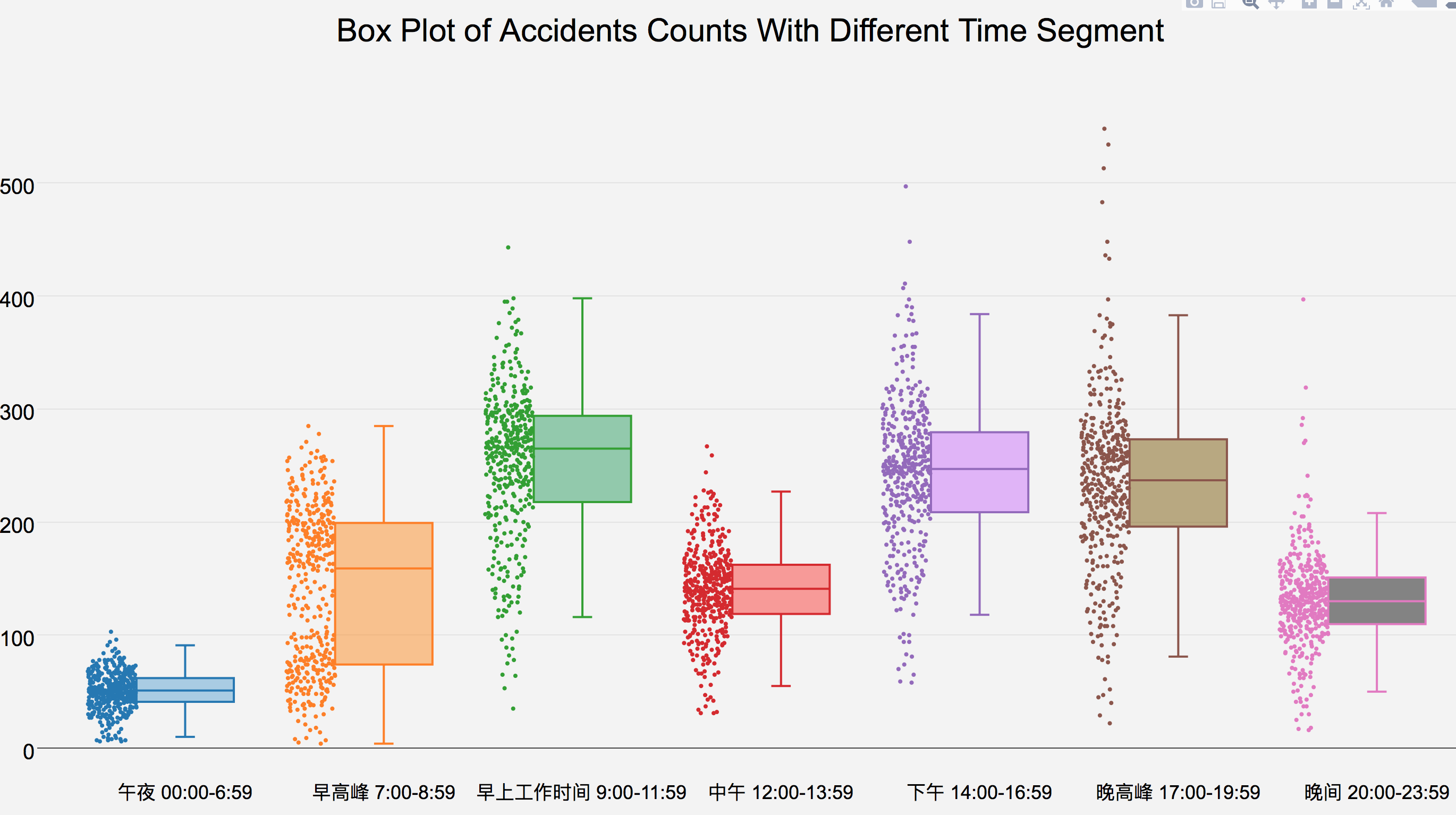
解决情况：通过相关性分析，我们选取了与交通事故发生规律最相关的若干因素加入到了机器学习模型中，它们包括：交通流（道路平均车速）、交通事故的周期特征、事故的位置信息、所在的一天当中的时间段、天气、PM2.5指数。

具体而言，因为交通流（道路平均车速）与车流量、道路通行能力大致呈反比关系，平均车速越高，通行越顺畅，车流量整体越小。而从众多文献的报道中发现，车流量与事故发生的概率密切相关，因此将交通流加入到模型考虑的因素中。

此外，我们发现交通事故在空间上有很大的分布不平衡，部分区域事故频率远比其他区域高如图4所示。其次，我们发现交通事故有很大的时间周期性，事故发生的概率整体上有周期性的涨落，如图5所示。横轴为不同时间间隔的长度，每条黑线代表12小时的倍数时间点。纵轴代表事故的空间距离，单位为千米。颜色的深浅代表相关性的高低。我们可以看到，事故有一定的时间周期的相关性，周期大约为24小时。而24小时前后附近的若干小时，也有一定的相关性。从空间上看，事故大约跟周围4公里内的区域有较弱的相关性。



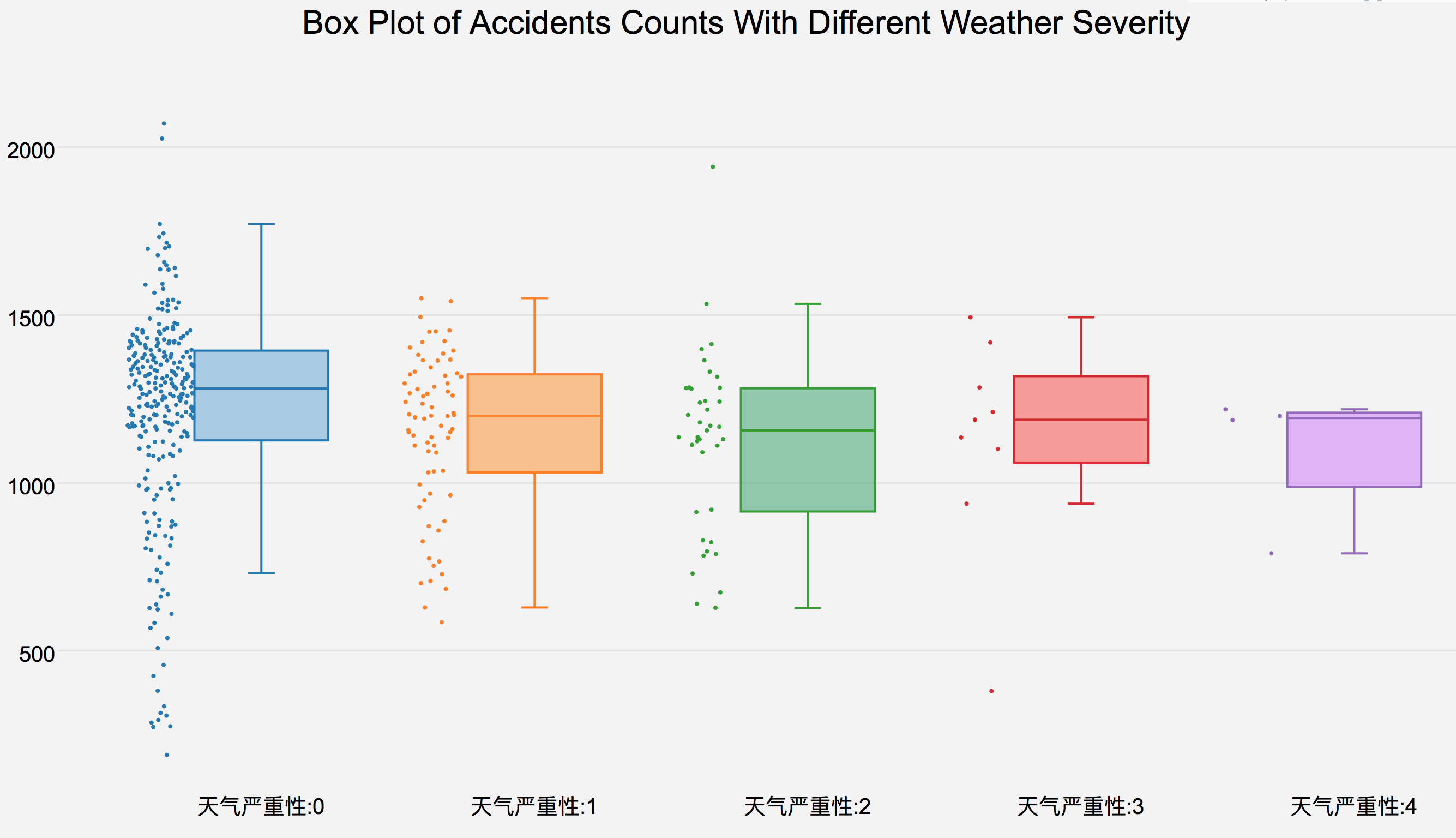
**图5 交通事故的时空相关性的等高线图**



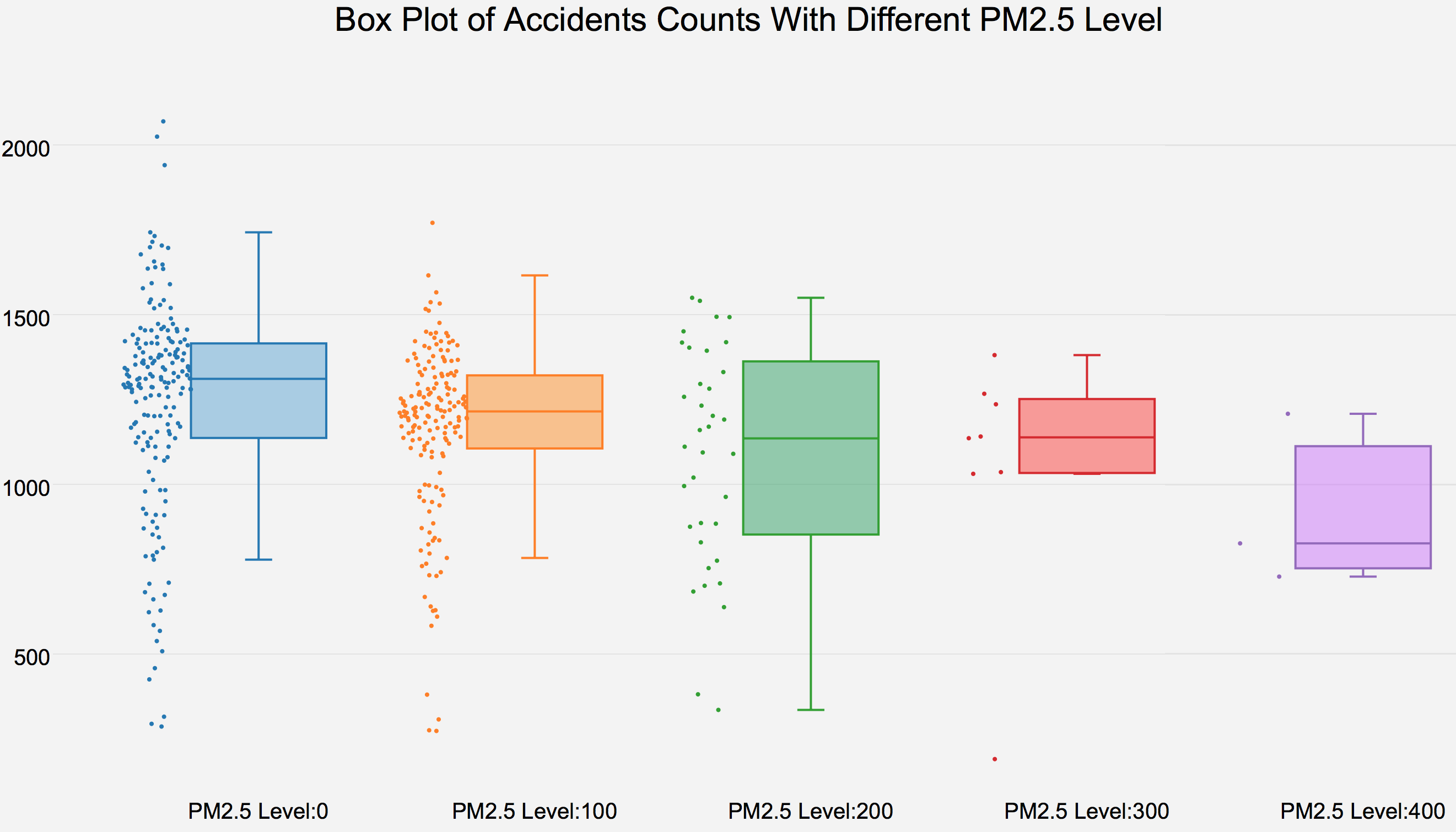
**图6 交通事故不同时间段内的事故量分布箱线图**

如图6所示，我们统计了2016年每天不同时间段内的事故分布量分布。可以看出，不同时间段内的事故量差异很大，比如早晚高峰事故量较多，中午和晚间事故量较少，凌晨最少。因此，加入事故的时间段信息，会对预测事故有一定的帮助。

如图7、图8所示，我们统计了事故对于不同天气严重程度，空气质量指数水平的分布情况。发现这两个因素确实会稍微影响事故在一天发生量的多少。总体而言，天气程度越严重，如大雪、大雨等严重天气，一天中的交通事故量总体越少。同样，PM2.5指数越高，即雾霾等空气污染越严重，事故量总体越少。这两者可能是因为在天气或者空气污染严重的时候，道路上整体的车流量较少导致的。因为这两者与交通事故也有弱相关，我们也将天气严重程度，空气污染指数(PM 2.5指数)，加入到模型考虑的因素中。



**图7 交通事故在不同天气严重程度情况下的分布箱线图**



**图8 交通事故在不同PM2.5水平下的分布箱线图**

2. 对于影响事故的异质数据（交通事故、交通流、天气等），分别采用哪种深度学习策略

解决方案：经过相关性分析后的结果显示，无论是交通事故本身，还是交通流都有明显的时间和空间的相关性、以及区域分布的偏好性。其他弱相关因素包括时间段、空气质量、天气情况等。因此考虑将随着时间变化的交通流、交通事故、天气严重性等数据，按照相关的时间周期组成短期时间特征和周期时间特征。

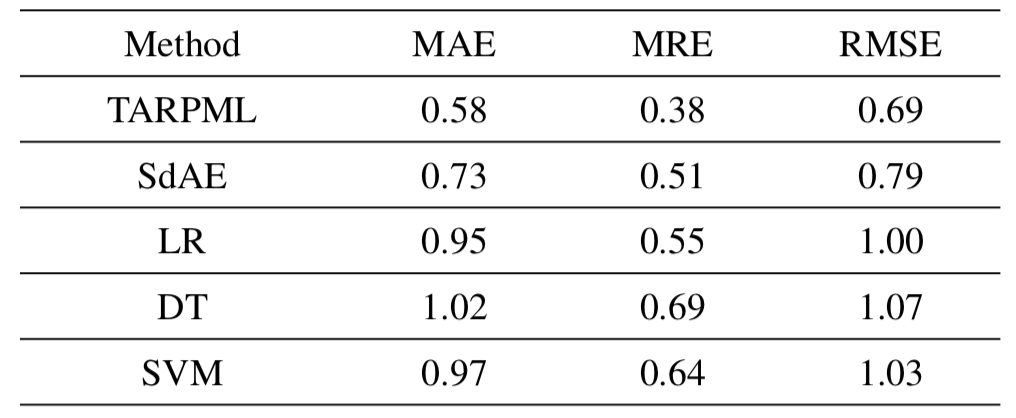
此外，加入不随时间变化的空间信息标志。以便于预测某区域的事故风险时，可以参考以往的事故发生频率。

综上所述，我们构建了基于LSTM的深度学习模型，如图2所示。

3. 如何对不同数据产生的模型结合，以创建预测更加有效的深度学习模型。

解决方案：我们通过将不同数据，交通流、交通事故、天气描述、等，进行预处理（离散化、数值化、归一化）之后，将相同时间的这些数据变成一个向量，以便达到利用这些异质数据的目的。如表2所示，TARPML为我们提出的基于LSTM的深度学习预测方法，SdAE为Chen et al提出的降噪编码机的事故预测方法[5]，LR为逻辑斯蒂回归方法、DT为决策树方法、SVM为支持向量机方法。MAE为平均绝对误差、MRE为平均相对误差、RMSE为均方误差。

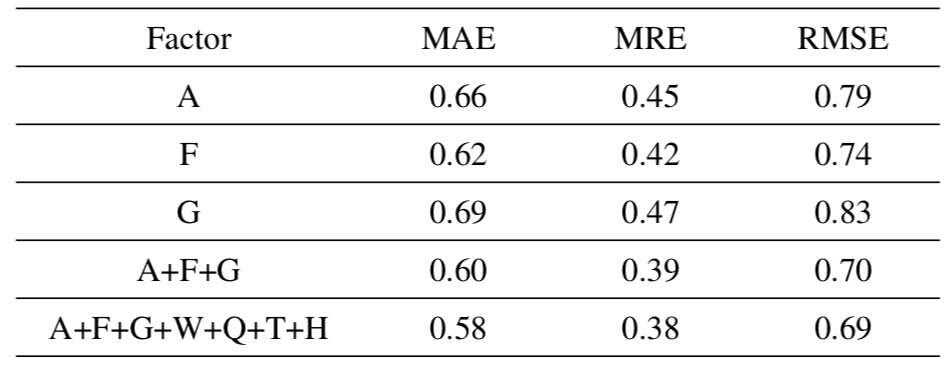
**表2 不同预测方法的预测性能比较**



4. 对于常被人诟病为黑盒子的深度学习模型，如何分析影响其预测准确性的因素，并找出关键因素，以数学形式表达各个因素间影响大小的权关系。

解决方案：对于深度学习模型的黑盒子问题，我们采用了格兰杰因果系数，利用不同因素的预测效果差异来比较因素的重要性。目前，我们通过实验得出的因素对事故预测的重要性排名如下：交通流 > 交通事故 > 空间位置信息 > 空气质量 + 时间段 + 天气严重性。如表3所示为不同因素的预测性能比较，其中，A代表事故因素、F为交通流、G为区域的地理位置信息、W、Q、T、H分别为天气严重性、空气质量指数、时间段标识、是否是节假日。

**表3 不同因素的预测性能比较**



1. 如何创建一个较为通用、具有良好交互性的在线实时事故预测平台，方便交通管理和决策者使用。

解决方案：通过深入了解事故预测的需求和可能的关注问题，形成较为通用的框架设计。底层上，设计了高效的异质大数据存储策略。基于交通大数据的显示需求，和数据特点，选择合适的可视化框架，并开发了具体的可视化界面。

3.3 创新性的方法、技术、成果

1. 本工作提出了一种较为准确的交通事故预测的深度学习方法，对于交通部门进行警力分配，居民出行提前的事故预防有重要的意义。

2. 通过采用格兰杰因果系数分析，我们间接的对影响事故的若干因素的重要性进行了排名和分析。从而使学术界能够从新的角度理解影响事故的因素。

3. 开发了一个较通用的交通事故显示和预测的可视化平台，以便于交管部门的决策人员使用和管理城市交通。

4 论文后期工作及进度安排

表 3 论文后期工作及进度安排

|  |  |
| --- | --- |
| 继续完善机器学习模型 | 2018/01-2018/04 |
| 毕设论文润色和格式修改 | 2018/01-2018/05 |
| 论文答辩 | 2018/06-2018/06 |

5 尚存的问题及措施



5.1 论文后期工作存在的困难和问题

1. 事故预测方法的准确率目前还不够理想，需要继续完善，探索新的机制，提升预测的准确率。
2. 目前学术论文还在投稿中。因此，成果方面目前空缺。
3. 论文初稿还比较粗糙。

5.2 准备采取的措施

1．尝试不同的思路来改变现有模型，找到一种比较理想、准确的事故预测模型。

2. 争取在答辩之前将论文发表。

3. 丰富和润色论文初稿的内容，严格按照要求修改格式。

5.3 如期完成全部论文工作的可能性

实际研究工作已经基本完成，后续将专注于毕设论文的撰写和模型的改善，应该能如期完成全部论文工作。

**参考文献**

[1]. Violence, W.H.O., I. Prevention and W.H. Organization, Global status report on road safety 2013: supporting a decade of action. 2013: World Health Organization.

[2]. 未知, 中华人民共和国国民经济和社会发展第十二个五年规划纲要. 2011.

[3]. Zheng, Y., et al., Urban computing: concepts, methodologies, and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 2014. 5(3): p. 38.

[4]. 郑宇, 城市计算概述. 武汉大学学报 (信息科学版), 2015. 40(1): 第1-13页.

[5]. Chen, Q., et al. Learning deep representation from big and heterogeneous data for traffic accident inference. in Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. 2016: AAAI Press.

[6]. Yang, C., et al., Geographic information system. 2010, Google Patents

[7]. de Oña, J., R.O. Mujalli and F.J. Calvo, Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. Accident Analysis & Prevention, 2011. 43(1): p. 402-411.

[8]. Savolainen, P.T., et al., The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. Accident Analysis & Prevention, 2011. 43(5): p. 1666-1676.

[9]. Chang, H. and T. Yeh, Risk factors to driver fatalities in single-vehicle crashes: comparisons between non-motorcycle drivers and motorcyclists. Journal of transportation engineering, 2006. 132(3): p. 227-236.

[10]. Malyshkina, N.V. and F.L. Mannering, Empirical assessment of the impact of highway design exceptions on the frequency and severity of vehicle accidents. Accident Analysis & Prevention, 2010. 42(1): p. 131-139.

[11]. Yamamoto, T. and V.N. Shankar, Bivariate ordered-response probit model of driver’s and passenger’s injury severities in collisions with fixed objects. Accident Analysis & Prevention, 2004. 36(5): p. 869-876.

[12]. Huang, H., H.C. Chin and M.M. Haque, Severity of driver injury and vehicle damage in traffic crashes at intersections: a Bayesian hierarchical analysis. Accident Analysis & Prevention, 2008. 40(1): p. 45-54.

[13]. Lee, J. and F. Mannering, Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. Accident Analysis & Prevention, 2002. 34(2): p. 149-161.

[14]. Eluru, N., C.R. Bhat and D.A. Hensher, A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. Accident Analysis & Prevention, 2008. 40(3): p. 1033-1054.

[15]. Mussone, L., A. Ferrari and M. Oneta, An analysis of urban collisions using an artificial intelligence model. Accident Analysis & Prevention, 1999. 31(6): p. 705-718.

[16]. Chang, L. and H. Wang, Analysis of traffic injury severity: An application of non-parametric classification tree techniques. Accident Analysis & Prevention, 2006. 38(5): p. 1019-1027.

[17]. Delen, D., R. Sharda and M. Bessonov, Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. Accident Analysis & Prevention, 2006. 38(3): p. 434-444.

[18]. Xie, Y., D. Lord and Y. Zhang, Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis. Accident Analysis & Prevention, 2007. 39(5): p. 922-933.

[19]. de Oña, J., R.O. Mujalli and F.J. Calvo, Analysis of traffic accident injury severity on Spanish rural highways using Bayesian networks. Accident Analysis & Prevention, 2011. 43(1): p. 402-411.

[20]. 孙铁轩等, 基于 SVM 灵敏度的城市交通事故严重程度影响因素分析. 2014.

[21]. Okabe, A. and I. Yamada, The K‐Function Method on a Network and Its Computational Implementation. Geographical Analysis, 2001. 33(3): p. 271-290.

[22]. Yamada, I. and J. Thill, Comparison of planar and network K-functions in traffic accident analysis. Journal of Transport Geography, 2004. 12(2): p. 149-158.

[23]. Stark, B.L. and D.L. Young, Linear nearest neighbor analysis. American Antiquity, 1981: p. 284-300.

[24]. Okabe, A. and K. Sugihara, Spatial analysis along networks: statistical and computational methods. 2012: John Wiley & Sons.

[25]. Sabel, C.E., et al. Road traffic accident simulation modelling-a kernel estimation approach. in The 17th Annual Colloquium of the Spatial Information Research Centre University of Otago, Dunedin, New Zealand. 2005: Citeseer.

[26]. Steenberghen, T., K. Aerts and I. Thomas, Spatial clustering of events on a network. Journal of Transport Geography, 2010. 18(3): p. 411-418.

[27]. Anderson, T.K., Kernel density estimation and K-means clustering to profile road accident hotspots. Accident Analysis & Prevention, 2009. 41(3): p. 359-364.

[28]. Smeed, R.J., The traffic problem in towns. Town Planning Review, 1964. 35(2): p. 133.

[29]. 韦丽琴等, ARIMA 模型在城市交通事故预测中的应用. 包头医学院学报, 2004. 20(4): 第287-288页.

[30]. 王刚, 道路交通死亡人数的灰色马尔可夫预测. 广东交通职业技术学院学报, 2006. 5(1): 第30-32页.