# Introduction

关于智慧城市的讨论与实践在近些年来越来越多。虽然关于究竟什么是“智慧城市”的定义有诸多版本，但是它们都有一个共同点，就是试图用海量的城市数据和算法去解决老办法无法解决的问题。城市是一个各要素高度耦合的复杂系统，因此想要让城市变得智慧起来，是一个很有挑战的系统性工程。可以预见，需要理清城市这个系统在运行时各个主体是什么，它们如何组成一些城市子系统，这些子系统又怎样相互影响，最终形成城市呈现出来的样子。同时，只是如同解剖一样知道城市如何运作，得到一个超高精度的数字孪生只是一个开始。更重要的是要能够知道当对城市进行一些改动时，整个城市的各方各面会产生什么样的连锁反应。同时做到描述和预测，才能说是形成了一个智慧的闭环。而在这当中，对状态信息的描述是预测效果的前提，而预测的准确率也是衡量描述方法优劣的一个重要参考。

智慧城市所涉及的问题复杂且庞大，所以这篇论文并没有不自量力到要去一举解决这个问题。相反，在遵循描述-预测研究范式的前提下，这篇论文选择了一种局限性视角下的交通状态预测问题作为切入点。这样做的原因有三点。其一，对于像城市这么复杂的研究对象，即便是选择一个较为局限前提展开研究也能在实际当中发挥作用，不是非要等到大一统的模型建立不可。其二，城市的各个方面虽然都跟城市相关，但是其实差异性很大，逐个各自采用不同的方法才是正常的。其三，在有限的时间、资源（特别是数据）和能力下选择一个有希望解决的子集是一个更为务实的选择。

综上所述，这篇论文研究的主要问题是预测街区尺度下的出租车交通状态。

交通是城市能够正常运转最不可或缺的功能之一。交通和物流的高效是城市高效的必要条件，如果交通的运行与管理并不智能，那么城市的智能也难以实现。不同于公共交通系统的中心化特性，私人交通具有去中心化特性。简单来说，公共交通系统的调度一般都由一个调度中心进行控制，停运或者进行改动相对容易是可能的，只不过需要考虑这样做是否恰当。而目前为止，对于私人交通的调度还是一个没有实现的事情。这个局面虽然主要的阻力来自于立法方面（譬如公权力部门是否有权力去收集私人汽车的数据），但是技术上确实并没有很成熟的解决方案。当前的交通管制，更多是以道路区段为单位的临时性管制，以暂时改变环境的特性为手段。虽然在法理层面这种基于道路封闭的交通管制是合理的，但是其在提高城市交通效率这一点上的效果确实不大。一种可能的解决方案是采用基于车辆的动态交通管制，即依据交通状态对整个交通系统内的车辆进行定制化管制，从而提高整体效率（譬如，降低整体通行时间、提高平均速度等）。

若要实现这个目标，那么就需要对城市交通系统内的各种车辆的运行规律有足够的了解。前边提到，虽然对于公共交通的了解和掌控水平较高，但是对于分布式的私人交通的了解由于种种原因并不足够。这时候对于出租车的研究就体现出价值了。出租车在交通系统中的行为模式与私人交通非常接近，但是其又有一定的中心化特征（即，由出租车公司统一运营管理，数据收集和利用更为便利），这就使它成为了绝佳的研究对象。

研究这个问题有两个直接的作用和一个潜在益处。第一，这篇论文的对于出租车交通状况的研究结果可以直接为相应的交通部门提供信息。第二，这是对城市交通状态预测方法的探索与创新，在适当的变换下可以应用于其它场景。潜在益处在于给出了一个将深度学习与传统工程领域的问题结合的研究方法案例，这对于类似的研究可以提供参考。

# Problem Definition

## Core Concepts

如前述，这篇论文研究一种限定前提条件下的交通状态预测问题。这一节的主要工作就是严格地定义本文研究问题相关的概念，formalize将要研究的问题。研究的内容是预测街区尺度下的出租车交通状态，那么首先要定义以下3个核心概念。

**尺度**

所谓尺度，即一种衡量状态的观测精度的概念。一般可以不严谨地用“微观”、“中观”和“宏观”尺度这样暧昧的叙述对尺度进行大致归类。在本研究中，特别指对空间的划分。譬如，街区尺度就是指以街区为空间基本单元，作为trip的起始点或终结点的归类依据。虽然存在trip A 和trip B的起始经纬度坐标a和b不同，但是当且仅当a和b同时落在街区单元X当中时，则在这一尺度下可以认为trip A和trip B的起始点相同。

**状态**

所谓状态，指某种在时间维度呈现变化的事件被观测时所呈现的特征。在本研究中，状态即是在一个时间区间内，用一种数据结构来反映出租车的trip所呈现出的街区间OD总体特征。

**预测**

所谓预测，指一种对未知的未来状态进行推断的行为。在本研究中进行的预测主要是基于过去的状态来推断未来的交通状态。

基于上述三个核心概念，对研究问题进行formalization。

## Formalization

这篇论文研究的问题可以认为是时间序列预测问题，预测的对象是交通状态。交通的本质是时空间的物质移动，交通的state是对这一过程的描述。不直接预测具体的交通流量，而是预测交通的状态。

为了实现这个目标，我们需要使用一种恰当的representation来描述交通的状态，使得如果存在任何feature那么都会被包含在内。如前述，交通状态的定义实际上是把一个连续的车流时空间位移状态离散化的过程。这样做的好处在于可以通过一些先验知识把问题进行简化，同时将交通预测问题formalize为时间序列预测问题，而后者有许多成功的方法。

Assumption 1. 假设城市交通在不同时间点会emerge特定的特征，而不是完全随机的。

假如城市的交通状况是完全随机的，意味着预测是不可能的。因此需要首先验证的是，交通状态存在一些特性，即非随机性。当这一点成立，才可能做预测。

Assumption 2.

## Summary

因此，这篇论文的主要工作就是把交通预测问题形式化为对一个数据结构随着时间变化的推断问题后，再利用相应的模型和算法解决之。（这一部分最后要导出的结论就是，我要用什么方法做本质上是一件什么事的研究，接着自然而然地导向要做相应的文献综述）

# Related Work

前人的工作总是有的。

## 时间序列预测

## 交通预测中的时间序列预测应用

## 循环网络

# State Representation

实验设计与结果

## Data Preparation

## State Data Structure

## Experiment

## Summary

# Exploration of Models

研究结论

## Architecture

## Loss Definition

## Experiments

## Performance

## Summary

# Conclusion

研究结论