## 基于LSM-RN模型的流量预测

路网流量预测是指在给定路网历史流量数据的情况下，预测未来的流量数据。随着科学技术的发展，我们可以容易地获得高精度高频率的手机信令数据，从而获得路网上的流量信息。现有的路网信息预测模型中大多数是通过历史路网中的边信息预测未来路网中的边信息，如自回归滑动平均模型[],支持向量回归模型[]，和高斯过程[], 也有一些模型考虑了路网节点的拓扑关系，将相邻节点的历史信息用于对节点流量的预测，如HMM[], 这些模型将时间信息和空间信息结合，取得了更好的效果。然而一些模型的计算复杂性高，如高斯过程和HMM，预测结果依赖于长时间的线下训练，很难做到实时。基于这些现状，我们采用LSM-RN（Latent Space Modeling for Road Networks）进行预测。

LSM-RN将路网信息嵌入表示到隐空间，并认为在隐空间中，相邻节点的流量信息具有相似性，且节点的相邻时间片上流量信息具有相似性。LSM-RN同时考虑了时间和空间的影响因素，并且可以实现大规模的实时计算。

### 问题定义

道路根据沿路的基站经纬度划分成不同的路段， 如图[]，每个路段对应一个基站，该路段的流量信息通过所对应基站的4G手机信令记录的数据推算得到，如下图中编号为1的路段流量数据由1号基站的数据得到。

经过路段划分，道路网络可以被看作是一个有向图 , 其中是图的节点集，是边集。道路网络中的节点即为不同路段的起点和终点，边即为连接起始点的有向路段，每条边的数据对应该路段上的流量值。因此路网的邻接矩阵的元素对应值代表节点所连接边的权重，在本文中即为流量数据。

流量信息的采样率可以在手机信令数据的采样率基础上调整。如手机心灵数据的采样率为 可以设定流量采样间隔 。根据所设定的时间间隔 span， 我们对每个时间片都可以形成一个路网的快照信息图 。

举例说明，下图中展示了某个时间片下一个有七个节点和10条有向边的路网图，路网

？ 没有形成network, 怎么举例

道路网络的流量动态变化可以表示成，其中可能包含丢失的数据。

根据以上描述，路网流量信息预测可以定义为：

给定一个动态道路网络，预测个时间片以后的路网,其中是预测步长，如若,则我们预测，即下一个时间片的路网信息。

### 隐空间模型

### LSM-RN模型

### 数据集说明

## 相关工作

数据聚类是将无标签的数据集分成个相似集合的过程，其中。聚类方法有很多种[**???**],总体来讲主要可以分为四类：基于划分聚类、 基于分层聚类、基于分布聚类、基于密度聚类等，另外还有一些少数的方法：基于有向图的聚类方法等。基于划分聚类的方法以k-means、k-medoids为代表，先将数据集划分成不同的k簇，然后迭代交换不同类之间的元素从而提升分类效果。k-means算法是将簇内元素的平均值作为簇的中心。而k-medoids算法，例如 PAM，CLARA[],CLARANS[]则是采用簇内的某一个点作为簇的中心，相比k-means，更不容易受到异常点的影响。但划分聚类需要根据先验知识提供簇的数目，值的设定会直接影响聚类结果。分层聚类法通常有从下至上凝聚型聚类和从上至下分裂型聚类，凝聚型聚类初始将每个数据点作为一个类别，如single-link 算法就是将距离最近的两个类别聚成新的一类。分裂型聚类是将初始所有数据看成一个类别，然后不断细分成更小的类别。层次聚类方法的时间复杂度比较高，为,其中N为数据的个数，此外对离异点也比较敏感。BIRCH[**???**] CURE[**???**] 和 C2P[**???**]等算法对传统的分层聚类算法做出了相应的改进。基于分布的聚类假设数据满足某种分布如高斯分布，聚类的目标是寻找最合适的参数，基于分布聚类的最大问题在于过拟合，并且如何选择合适的模型也是比较困难的。基于密度的聚类可以发现空间中的高密度区域并将其与低密度区域分开，DBSCAN[**???**] 和 OPTICS[**???**] 算法是密度聚类的典型代表。这种方法不需要知道任何先验知识，可以通过密度分布寻找到任何形状的聚类类别，并且对异常点不敏感，且不需要先验知识定义聚簇的个数。