## 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

隐空间模型假设一个图的节点可以嵌入表示到一个可以描述节点隐层属性的隐层空间，并认为相似节点在隐层属性上相似度高。例如，在社交网络的社团发现[1]、话题发现[2]等问题中，认为在某些隐形属性上（如社团、话题等）距离近的用户更容易聚集为一个社团，或者更容易形成一个话题。在道路网络中，假设路网拥有很多隐层属性，且隐层属性间可能会相互影响，同时这些隐层属性会影响路网的流量信息，从而便可以根据路网的历史流量数据从而推算出未来路网的流量数据[3]。

隐空间模型实现原理是通过最小化实际值和根据隐空间属性表示得到的估计值间的差别，常用的有KL散度、最小平方误差等指标。具体实现方法有很多种，最常用的是非负矩阵分解(Non-negative Matrix Factorization)。非负矩阵分解是将非负矩阵分解成两个非负矩阵和,其中

非负矩阵分解可以通过乘子迭代法迭代优化目标函数，从而得到局部最优解。

### 2因子分解隐空间模型

### 3因子分解隐空间模型

### LSM-RN模型

#### LSM 基础模型

一个拥有个节点的道路网络在时间的流量数据可表示成

对路网中的每一个顶点，假设其具有某些隐空间特征，设顶点在隐空间上的表示为,其中,为隐空间属性个数。不同的隐空间属性会影响道路的流量情况，例如交叉路口比非交叉路口流量大，城内道路比城际道路流量大等。举例说明如下：

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 交叉路口 | 城市内 | 居民区 | 商业区 |  |
|  | 0.9 | 0.1 | 0.8 | 0.2 |  |

所有节点的隐空间表示组合一起可以构成隐空间矩阵 。

起始点不同的隐空间属性值会对流量值产生影响，例如两个交叉口间的路段相较其他普通路段会更大。这种隐空间属性之间的影响可以通过矩阵来描述，其中。节点和节点之间的流量可以看作隐空间属性的线性加和，如下：

路网中所有节点在隐空间上的映射表示可以表示成矩阵，我们希望通过隐空间表示矩阵和隐空间属性矩阵还原路网矩阵G，求解和的过程可以看作非负矩阵分解的过程，目标函数定义如下：

$$
arg\min\_{U\ge0,B\ge0}J=\Arrowvert{G-UBU^T}\Arrowvert\_F^2
$$

在实际的路网数据中，道路网络是非常稀疏的，节点的平均度数非常小，且可能会出现数据丢失的情况，路网矩阵流量数据为0的路段有可能是两节点之间并无连接，也有可能是数据的丢失，但两种情况是无法区分的。因此为了克服矩阵的稀疏性，只考虑观测到的有值的部分，即流量值的元素。定义指示矩阵，其中

目标函数变为

$$
arg\min\_{U\ge0,B\ge0}J=\Arrowvert{Y\odot{G}-UBU^T}\Arrowvert\_F^2
$$

其中，表示矩阵的哈达马积，满足。

实际路网数据中可能存在数据丢失，路网中相邻路段间的流量不可能发生跳变，因此我们可以合理假设相邻路段间的流量变化是平滑的，可以对目标函数加入惩罚因子如下：

$$
penalty=\frac{1}{2}\Sigma\_{i,j=1}^{n}\Arrowvert{u\_{i}}-u\_{j}\Arrowvert^2W\_{ij}
$$

对该式子进行化简如下：

$$
\begin{array}{ll}
panalty&=\frac{1}{2}\Sigma\_{i,j=1}^{n}\Arrowvert{u\_{i}}-u\_{j}\Arrowvert^2W\_{ij}\\
& =\Sigma\_{i}^{n}u\_{i}^{t}u\_{i}D\_{ii}-\Sigma\_{i,j=1}^nu\_{i}^Tu\_{j}W\_{ij}\\
& =Tr(U^TDU)-Tr(U^TWU)\\
& =Tr(U^TLU)
\end{array}
$$

其中矩阵为图的拉普拉斯矩阵

### 数据集说明

[1] WANG F, LI T, WANG X, 等. Community discovery using nonnegative matrix factorization[J]. Data Mining and Knowledge Discovery, Springer, 2011, 22(3): 493–521.

[2] SAHA A, SINDHWANI V. Learning evolving and emerging topics in social media: a dynamic nmf approach with temporal regularization[C]//Proceedings of the fifth ACM international conference on Web search and data mining. ACM, 2012: 693–702.

[3] DENG D, SHAHABI C, DEMIRYUREK U, 等. Latent space model for road networks to predict time-varying traffic[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2016: 1525–1534.