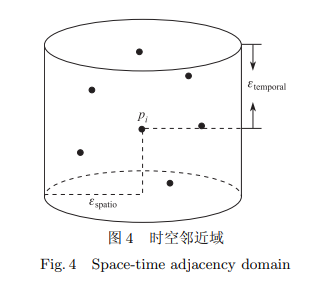
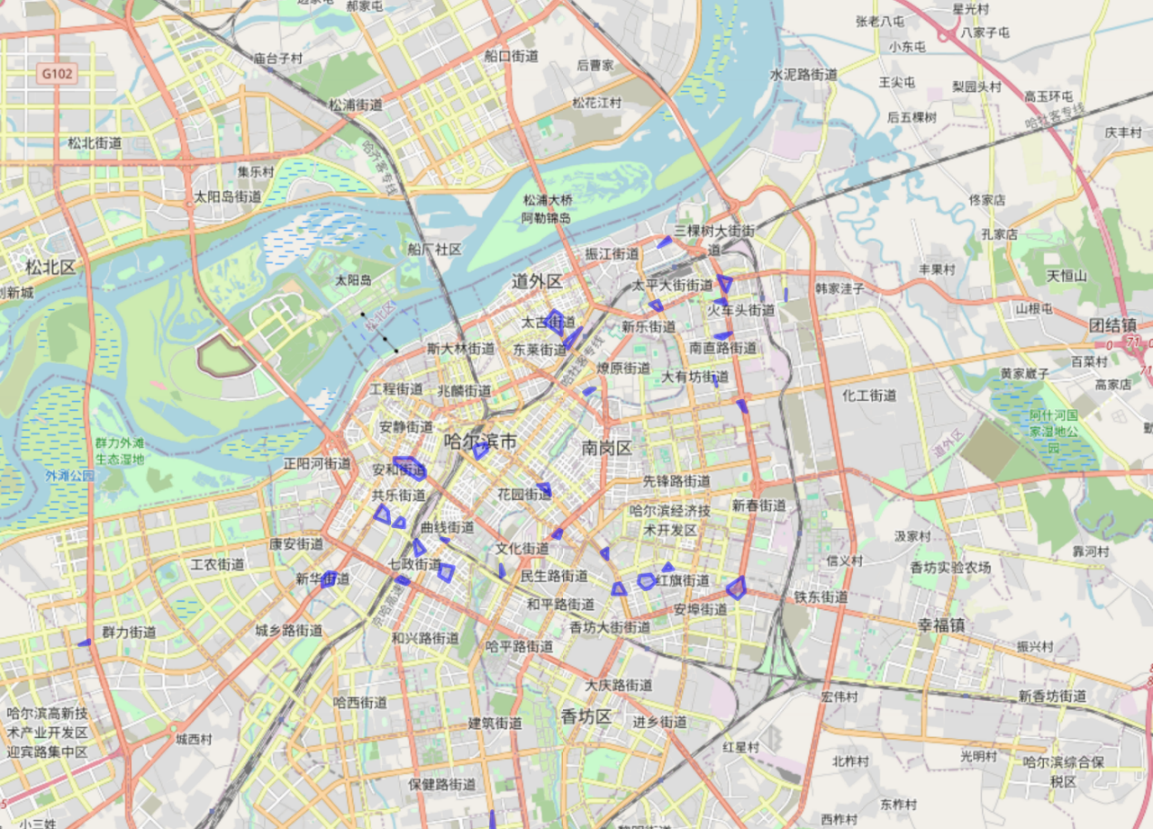
摘要：为降低城市出租车的空载率，缓解路网交通拥堵压力，重需设计有效的出租车载客热门区域推荐方法。针对传统的出租车相关推荐方法忽略实际路况导致推荐精度较低的现状，提出了一个两阶段的载客热门区域实时推荐算法。首先，离线挖掘阶段，基于出租车历史轨迹数据集提取基于时段属性的载客热门区域。随后，在线推荐阶段，根据出租车请求位置及时间，结合实时路况，对载客热门区域进行评测排序。

离线阶段：通过从历史出租车轨迹数据集中提取出载客状态由空载状态跳变为载客状

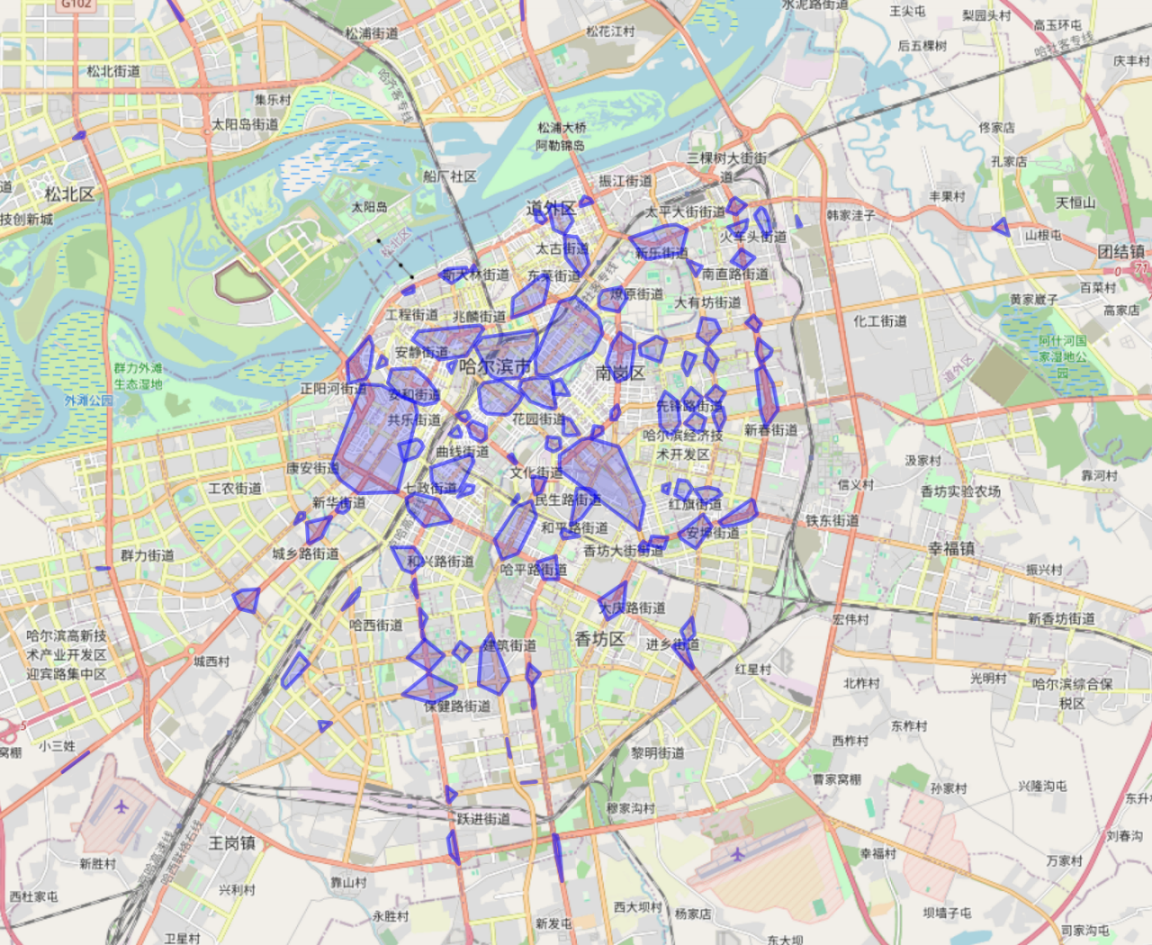
态的GPS采样点，将此点作为出租车载客点。基于上文中对载客热门区域的定义和分析，本文采用改进的DBSCAN聚类算法对提取出的载客点进行时空聚类。该方法将传统的DBSCAN聚类算法在时空域上进行扩展,将载客点分布于时空立体空间中进行聚类,通过两个参数temporal 和 spatio分别作为时间维度和空间维度的度量半径,如图所示，不仅考虑了空间属性的相似性,同时也考虑了时间维度的相似性.采用时空密度作为实体空间相似性的度量标准,将时空簇视为一系列由低密度区域(噪声)分割开的高密度连通区域.由于改进的DBSCAN聚类算法在聚类过程中考虑了时间维度上的相似性，使得聚类结果为在某一连续时段内数据点密度较大的的时空簇，即位于同一个簇中的载客点不仅在地理位置上邻近，在时间维度上也是较邻近的，最终达到在某一连续时间段和密度较大的聚类目标。



基于改进的DBSCAN聚类算法，将提取出的载客点按其时间戳分为24个时间片进行时空聚类，生成各时间片中时空密度较高的出租车载客点集合，其聚类结果即时空簇便为某个连续时段内载客事件发生较为频繁的区域。采用 Graham扫描法，发现各时空簇在空间属性上的凸包(Convex Hull)，以凸多边形表示载客热门区域的儿何属性，实现对时空簇到地理几何数据的转化，同时计算出各时空簇的簇心，代表该载客热门区域对出租车司机进行推荐，最后将该时空簇所处于的时间片作为其时段属性。以下就是不同时段的热门区域。



（上午9点的热门区域）



（下午3点的热门区域）

在线阶段：结合实时路况，为出租车推荐最优载客区。

