

# 北京数据集介绍

19373750 杨直凡

处理数据的代码在[https://github.com/yyf-buaa/BJ\\_Dataset/blob/master/main.py](https://github.com/yyf-buaa/BJ_Dataset/blob/master/main.py).

处理为原子文件格式后的bj\_dataset分为三部分，bj\_dataset.geo,bj\_dataset.rel,bj\_dataset.dyna

## bj\_dataset.geo

此文件存储了**所有层面**的交通实体，包括区域，路段，POI

字段	含义
geo_id	统一的编号，从0开始
type	形状类型，区域为Polygon，路段为LineString，POI为Point
coordinates	描述 geo 实体的位置信息，采用 <a href="#">Geojson</a> 的 coordinates 格式
function	区域的功能（从0开始编号的数字），路段统一为-1（原始数据中没有功能），POI功能（以P开始的字符串），用于下游任务
traffic_type	描述实体的类型，['region','road','poi']

## bj\_dataset.rel

此文件存储了不同层次交通实体之间的映射关系，包括  
road2region,region2road,poi2region,region2poi,poi2road,road2poi

字段	含义
rel_id	统一的编号，从0开始
type	均为'geo'
origin_id	关系起点方的 ID，为geo表中实体的id
destination_id	关系终点方的 ID，为geo表中实体的id
rel_type	关系的类型 ['road2region', 'region2road', 'poi2region', 'region2poi', 'poi2road', 'road2poi']

## bj\_dataset.dyna

此文件存储了**路段**轨迹数据

**此文件的一行只存储一个点，因此一条轨迹需要若干条连续的dyna记录来存储**

字段	含义
dyna_id	统一的编号，从0开始
type	均为'trajectory'

字段	含义
time	时间信息
entity_id	轨迹的拥有者id
traj_id	同一个人的轨迹id（同一个人可能会有多条轨迹）
geo_id	一个road的id，标记轨迹上的一个点
total_traj_id	总的轨迹id（从0开始）