说 明 书

**一种基于聚类的区域出行意愿强度分析方法及装置**

**技术领域**

本发明属于信息技术邻域，尤其涉及一种基于聚类的区域出行意愿强度分析方法。

**背景技术**

随着新能源汽车行业的快速发展，充电桩数量有限的瓶颈逐渐显现出来，同时由于充电桩的位置明显不合理而导致其使用率偏低。

为使电动汽车充电桩得到合理利用，目前主要基于在已有充电桩的基础上，采用0-1规划及覆盖的数学模型对充电桩的位置进行分配，采取动态调整与静态增设的方法，使得目标函数（即对规划后的电动汽车充电桩的利用率）最大。

然而，我们应该更多关注充电桩建站地点在市民出行意愿网络中的重要性，这样的充电桩建设才有意义。

**发明内容**

本发明的目的在于针对现有技术的不足，提供一种基于聚类的区域出行意愿强度分析方法,综合考虑一个区域在各个时间段市民出行意愿中的重要性。

市民出行的轨迹记录字段包括：身份ID、出发地点经纬度、目的地点经纬度、出发时间、到达时间，记录数有n条，则出行轨迹记录出发时间集为sttime[0..n-1]，出发地点经纬度集为location[0..n-1]＝{(x0,y0),(x1,y1),…,(xn-1,yn-1)}。

以出发时间为维度，聚类分析方法如下：

1. 对出行记录的出发时间转化为十进制，如时间11:30则为11.5，选取出行轨迹记录出发时间的m个为初始中心，记为a[0]= sttime[0],a[1]= sttime[1],..,a[m-1]= sttime[m-1](m<n)；
2. 与对于sttime[0]，…，sttime[n-1], 分别与a[0],a[1]…a[m-1]比较，假定与a[i]差值最小，就标记为i;
3. 对于所有标记为i的点，重新计算a[i]={所有标记为i的sttime[j]之和}／标记为i的个数；
4. 重复2,3,直到所有a[i]值的变化小于给定阈值。

同时以出发地点为维度，聚类分析方法如下：

1. 选取出出行轨迹记录出发地点的k个为初始中心，记为b[0]= location[0],b[1]= location[1],…,b[k-1]= location[k-1](k<n)；
2. 与对于location[0]，…，location[n-1], 分别与b[0],b[1],…,b[k-1]比较，假定与b[i]距离最小，就标记为i;
3. 对于所有标记为i的点，重新计算b[i]={所有标记为i的location[j]之和}／标记为i的个数；
4. 重复2,3,直到所有b[i]值的变化小于给定阈值。

定义某区域时间权值TimeWeight＝所在区域类的出行时间总和／总出行时间总和

定义某区域权值ZoneWeight=所在区域类的记录数／总记录数

基于以上的聚类分析方法，并定义某地区p的出行意愿强度函数如下：

TravelStrength=ZoneWeight＊TimeWeight

如图1所示，本发明提供的区域出行意愿强度分析装置如下：

该装置部署在互联网环境下，通过数据采集器自行爬取互联网中的出行数据，并通过数据分拣器进行数据清洗，清洗后的出行数据存入海量数据存储器。聚类分析器基于清洗的数据，以出发时间和出发区域为维度，通过聚类分析，并结合出行意愿强度函数，最终输出各个地区的出行意愿强度列表。

（1）数据采集器：该部件基于爬虫技术，采集公交、出租车、自行车等出行数据，支持分布式部署。

（2）数据分拣器：该部件主要进行清洗数据，并结构化存储。

（3）聚类分析器：该部件基于聚类算法，对出行数据进行时空网络分析。

（4）出行意愿强度分析器：该部件基于聚类分析结构，计算各个区域的出行意愿强度。

**具体实施例**

本发明提供的一种基于聚类的区域出行意愿强度分析方法：

（1）基于爬虫技术，采集公交、出租车、自行车等出行数据， 本发明提供的架构支持多节点方式进行爬虫。

（2）分拣采集的出行数据：清洗掉爬虫的脏数据，并结构化放入到海量数据存储库。

每一笔清洗后的出行数据结构如下：

|  |  |
| --- | --- |
| 数据项 | 释义 |
| ID | 身份ID |
| StartLongitude | 出发地点经度 |
| StartLatitude | 出发地点纬度 |
| DestLongitude | 目的地点经度 |
| DestLatitude | 目的地点纬度 |
| StartTime | 出发时间 |
| DestTime | 到达时间 |

（3）基于发明内容提及的聚类分析采集的出行数据，以出发地点为维度，定义k＝15，则区域权值分为15类，每一类包含相对应的区域ID列表。

（4）计算各个区域的出行意愿强度，数据结构定义如下：

|  |  |
| --- | --- |
| 数据项 | 释义 |
| ZoneID | 区域ID |
| ZoneWeight | 区域权值 |
| TimeWeight | 时间权值 |
| TravelStrength | 出行意愿强度 |
| NextNode | 指向下一结点 |

本发明的有益效果是：

1、出行意愿强度分析装置：本装置提供出行意愿强度分析全流程架构，解决了市民出行强度分析的实时性以及准确性。

2、聚类分析方法：本发明给出了一种聚类分析方法，对市民出行轨迹进行时空分析，同时提供了市民出行意愿强度计算方法，最终为充电桩选址部署提供参考依据。

　　说 明 书 附 图

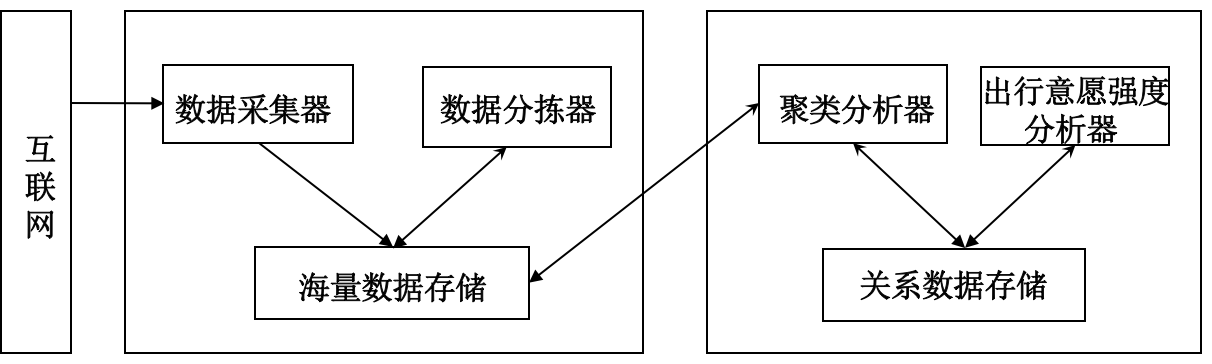


图1 区域出行意愿强度分析装置