**基于时间约束和轨迹线划分的改进型Hausdorff距离的三维时空轨迹相似度量**

**邓赓，吴博，苏以鹏，王胜，张玉潇**

（中南财经政法大学信息与安全工程学院，湖北 武汉430073）

**摘 要：**随着道路拥堵和私家车数量剧增等问题日益严峻，蓬勃发展的网约车拼车已成为更多人的选择。为了提供更高质量的拼车服务，轨迹匹配问题将是研究的热点。但传统的基于Hausdorff距离的轨迹匹配在使用中暴露出了两方面的问题：①只考虑了路径上点的坐标信息，忽略了用户的等待时间；②直接计算整段路径的Hausdorff距离，无法体现特殊路段对匹配度量的影响。本文针对上述问题提出两点改进：①提出带有时间属性的改进的Hausdorff距离计算方法；②借助微元的思想，利用转向点轨迹线分割原路径，将匹配度量细化到每个子路段。本文根据Google地图的数据来源，利用爬虫获取到武汉市行政路线图，随机选取670条路线作为实验数据，实验结果表明：相较于传统算法MCT，改进算法的轨迹匹配中成功匹配的轨迹更少，Hausdorff值偏大，预计等待时间更小，这说明加入时间约束后，判断轨迹是否匹配的条件更严格，匹配的质量更高。

**关键词：**时间约束；轨迹线划分；拼车；Hausdorff算法

**Three Dimensional Spatiotemporal Orbit Similarity Measure Of Modified Hausdorff Distance Based on Time Constraint and Trajectory Line Division**

**DENG Geng, WU Bo, SU Yi-peng, WANG Sheng ,ZHANG Yu-xiao,**

(School of Information and Security Engineering, Zhongnan University of Economics and Law, Wuhan 430073,China )

**Abstract:** With more private car and road congestion problem increasingly grim，the carpool which is vigorously developing has become the choose of more and more people. In order to provide higher quality carpool service, trajectory matching problem will be the study hotspot. But the traditional path matching based on Hausdorff distance has exposed two aspects of the problem in application: ①The coordinate information of the point on the path is considered, but the waiting time of the user is ignored; ②The direct calculation of the Hausdorff distance of the entire path can not reflect the impact of special sections on the matching metric. In this paper, two improvements are proposed for the above mentioned problems: ①Propose an improved Hausdorff distance calculation method with time attribute ; ②With the help of Micro thought, using turning point trajectory to segment original path, the matching measure will be subdivided into each sub section. In this paper, according to the data source of Google map, the crawler is used to get to Wuhan city administration route map,670 routes were randomly selected as the experimental data. Experimental results show that: Compared to the traditional algorithm MCT, during the trajectory matching based on improved algorithm, the success of the match is less, the Hausdorff value is larger, and the waiting time is shorter,which indicates that when the time constraint is added, the condition for judging whether the track is matched is more strict, and the quality of the match is higher.

**Keywords:** Time constraint; trace line division; carpooling; Hausdorff algorithm

**1 引言**

如今，随着私家车的兴起，其所造成的道路拥堵和尾气污染等问题日益严峻。与此同时，网约车拼车以其方便乘客出行和充分利用私家车资源的特点，受到用户的欢迎。网约车拼车是一种类型的拼车，也即具有相同方向相似线路的乘客乘坐同一辆网约车。网约车拼车利用先进的GPS定位技术通过互联网平台手机软件连接司机和乘客，采用同时搭载多个顺路乘客的方式，提高了载客效率。网约车具有出租车相当的移动性，同时也具有很多优点。①网约车连接私家车司机和乘客，使得用户出行更加方便也使更多闲置的资源得到了有效利用。②拼车时乘客只需支付较少的费用，降低乘客的出行成本；同时司机可以开一次车载多个客，增加司机的收入。更高的乘客率、更少的空车率和更高的车辆使用效率有利于缓解城市交通拥堵。③随着网约车的合法化和规范化，乘客在拼车时能够得到有效的法律保障，降低了出行的安全风险。总之，网约车拼车体现了分享经济的一种出行方式，它有利于提高交通资源的利用，对于缓解城市交通拥堵，减少环境污染，加速城市化进程，具有积极的意义[1]。因此，网约车拼车作为一种提高交通资源利用率的出行方式，得到社会的广泛关注，也使得拼车问题成为新的研究热点[2]。

然而在实际的拼车过程中，乘客和司机经常会面临两个问题：一方面乘客们往往需要额外绕行一段距离和耗费相应的时间来实现拼车，甚至有时同时参与拼车的两名乘客的目的地方向截然相反；另一方面司机在面对众多的拼车请求时，难以快速选择一项对于司机、乘客来说最优的方案[3]。由于没有准确有效的拼车系统支撑，造成包括乘客和司机在内的整个拼车过程不能高效运行，这对于乘客和司机来说都是不必要的损失[4]。为此我们查阅了大量关于拼车的研究，发现对于拼车模型的设计有很多种，根据实验结果发现其中较为合理有效的一种是基于Hausdorff距离的拼车模型[5]。经过研究发现：Hausdorff距离用于计算两个点集合之间的距离，因此可以被用来计算两条轨迹间的距离。然而现实中两条乘客的行车路径是有序且有时间限制的，Hausdorff距离只能解决无序点集合之间的距离。为此，我们在模拟乘客行车路径时候采取了转向点的概念，通过对已提取的路径转向点加上时间限制将问题转化成了三维坐标下的有序点集的Hausdorff距离。最终实验的模拟结果通其他相关研究结果对比证实：本文针对拼车问题提出的基于时间约束和轨迹线划分的改进型Hausdorff距离的拼车方法更加准确有效。

本文将在2中介绍相关研究，在3中介绍模型，其中3.1介绍数据来源、3.2介绍数据预处理、3.3中介绍轨迹相似度匹配，在4中介绍实验部分，其中4.1介绍实验数据、4.2介绍实验流程、4.3介绍实验结果及分析。

**2 相关研究**

拼车问题的关键在于多个乘客出行路径匹配的问题。静态拼车问题一般被建模为VRPPDTW 问题(vehicle routing problem with pickup and deliveries and time windows)，用启发方式进行分配[6]。由于缺少有效的信息技术和通讯技术支持，拼车请求匹配列表的自动生成问题遇到瓶颈。随着智能交通系统ITS、地理信息系统GIS、GPS定位技术以及移动通讯技术的发展，无规律的个人的动态拼车请求可以在需要的时候及时提出[7]。调度中心可以通过ITS技术得到司机的位置和运动轨迹，并且能实时获取移动用户的位置和需求。文献[8]基于任务调度中的sufferage算法原理来解决动态拼车调度问题，但是这种方法仅考虑了最小执行时间MET(Minimum Execution Time)并没有考虑多起点到多终点(“many-to-many”)路径之间的拟合度。

为了解决多起点到多终点问题邵增珍等[9]提出匹配度聚类调整车辆路线以将更多乘客划分到某辆车上，然后利用先验聚类对车辆内部乘客上下车顺序排序，最后采用迁入迁出策略进一步优化拼车方案。由于Hausdorff距离可以用来计算两个点集之间的距离，因此它被许多学者用来计算两条路径间的距离，肖强等人[10]利用Hausdorff距离将文献[6]中的路径匹配度量化。Atev等人[11]提出一种基于Hausdorff距离的双谱聚类算法并成功地将其运用到车辆轨迹分析。文献[12]针对文献[8]提出基于改进的Hausdorff距离的分配算法，利用基于匹配度的聚类方法的得到最最优匹配的拼车需求。文献[13]提出一种基于时间约束的Hausdorff距离的时空轨迹相似度量，利用滑动窗口(sliding window)对文献[12]中的算法进行了改进。

然而，上述研究均是利用Hausdorff距离的算法解决路径匹配问题(Path matching)，然而在结合Hausdorff距离和拼车问题时仍面临以下挑战：

①Hausdorff距离所定义的转向点(Direction point)序列是无序的，然而乘客乘车路径的转向点序列是有序的，如何将转向点序列有序化有待进一步研究；②乘客乘车路径的时间轨迹和空间轨迹是一一对应的，因此Hausdorff距离的时间约束如何定义和分析成为难点；③目前广泛应用的Hausdorff距离拼车算法在路径匹配上误差较大，有待进一步提高。

针对上述情况，为了有序化转向点序列，合理引入时间约束，并且尽可能提高成功率，本文提出一种基于时间约束的有序Hausdorff距离的三维时空轨迹相似度量的拼车算法，通过转向点轨迹线动态分析两条轨迹各子轨迹之间的Hausdorff距离，来判断目标轨迹的拟合程度，根据所有待选目标轨迹的拟合程度排名选出最优匹配方案。

**3 相似轨迹匹配模型**

**3.1 预处理**

由于整个拼车车辆轨迹计算量大且计算过程繁琐复杂，不利于拼车路径匹配分析与计算，于是利用转向点思想，及通过GPS卫星定位系统获得车辆具体位置时，当车辆行驶角度发生改变，计算相邻点的方向角度差 ，若或或，则判定该点为转向点。通过提取行驶轨迹中的所有转向点及初始点和终点，根据时间升序排列形成转向点序列，从而提取、重构、简化车辆轨迹路径，压缩为带时间节点的转向点序列，任意两转向点之间形成的子轨迹方向不变，因此通过提取两个转向点即可还原轨迹路径的时间特征与方向特征，这大大减少了计算量的同时为之后的拼车路径相似度匹配做好良好的数据准备。

**3.2 轨迹相似度匹配**

**3.2.1 Hausdorff距离**

传统Hausdorff距离用于衡量两个无序集合之间的相似程度，它也可以作为两个集合之间距离的一种定义形式：假设有两组集合 和 则这两个集合的Hausdorff距离可以定义为：   
 (1)

其中，,,,使得，即：   
 (2)

(3)

其中，为点集A与点集B之间的距离范式，即对B集合中的每个点到距离此点最近的A集合中的点的距离，进行排序选取距离中的最大值作为。

双向的Hausdorff距离 ，是单向Hausdorff距离 和 的较大者，它度量了不同的两个点集之间的最大不匹配度。

**3.2.2改进的Hausdorff距离**

目前的有关利用Hausdorff距离来进行路径匹配的研究，大都着眼于匹配双方的相对距离，而并没有涉及时间概念。所以当引入时间序列之后，原有的研究方法便不太合适。从定义上来说，Hausdorff距离所针对的对象本是无序的点集，但是在引入时间这一参数之后，点集必然在时间轴具有先后顺序，因此对象由无序的点集转化为了有序的点集。

为了解决有序性和时间相关性的问题，本文提出一种结合时间轴的三维Hausdorff距离计算方法来解决该问题。根据Hausdorff距离定义，须计算路径A 上的任意一轨迹点 P 到路径B最短距离，现采用三维坐标的概念来表示，首先建立经度，纬度和时间轴的三维坐标系，其次将路径刻画于坐标系上。其中L轴为经度，K轴为维度，t轴为时间轴，引入轨迹点的概念，用坐标（L，K，t）来进行表示，此时，路径由线段转化为点的集合来表示，如图1所示。

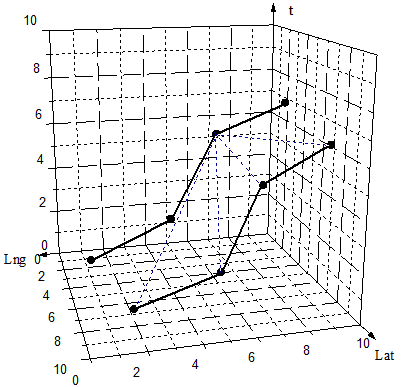


图1：具有时间约束的A、B路径示意图

现将路径A与B带入坐标系中，如图3，取A中的一任意轨迹点a，按照Hausdorff距离的定义，由公式（1），可计算得到Hausdorff距离，但由于引入了时间轴，目前的路径的相似度匹配的计算结果是部分依赖于时间的。同时，时间作为一种变量，其本身性质与经度纬度是不同的，换言之，时间对最后的计算结果的影响是与经纬度无关，具有一定的独立性。因此为了对其进行描述，并将其影响体现在改进的Hausdorff距离中，用T来表示司机到达拼车点的绕路系数[16]。这是一种时间属性的因素，表示司机到达拼车点的预估用时所让顾客付出的代价。

针对于时间因素对Hausdorff距离的影响，我们准备采用分段描述的方式进行精确刻画，此外考虑到时间变量在计算中与其他变量的无关性，将其计算公式独立出来。同时，由于时间与距离的度量不同，不能直接相加，因此借助距离等于速度乘以时间这一基本概念，预设一参数速度V，来对T进行度量上的转化。在V的预设值上，应根据不同实际情况，采用不同的值，本文中根据参考文献[1],代入V的值23.3km/h。因此控制经纬度坐标，代入实验数据，根据所得曲线有：

(4)

综上，提出改进的Hausdorff距离算法来计算引入时间概念的轨迹之间的匹配度。

(5) 其中，,,,使得,即：

(6)

(7)

**3.2.3基于转向点轨迹线的相似轨迹匹配**

针对传统拼车匹配中的时间约束同线路匹配质量问题以及，我们提出一个基于关键点轨迹线的相似轨迹匹配算法来实现更小的轨迹复杂度、更少的绕路时间以及更高的相似路径的匹配方案，具体算法为表1所示。

表1基于关键点轨迹线的相似轨迹匹配算法表

算法 基于关键点轨迹线的相似轨迹匹配

输入：轨迹A的子转向点序列，轨迹B的子转向点序列，轨迹线区间长度L

输出：相似子轨迹的总长度，轨迹间匹配度

(1)：初始化：，

(2)：**while** **do**

(3)： ;

(4)：   
;

(5)： **if**  **then**

(6)： ;

(7)： **Output** ；

(8)： **end if**

(9)： ;

(10)： ;

(11)： **end while**

(12)： ；

(13)： **return**

首先根据本文所提出的基于时间约束的转向点算法提取所有转向点轨迹序列，A的转向点序列，B的转向点序列，由于轨迹按时间先后顺序压缩为转向点序列，而相邻两个转向点之间所形成的线段中的所有点所在的线路都可以通过这两个转向点来表示。其次为了提高Hausdorff距离度量的准确度，本文利用微元思想以及转向点轨迹线来划分转向点序列。对于关键点的选择，为防止转向点之间距离太大对Hausdorff距离度量造成误差，本文在起止转向点之间选取若干点，这些点包括起止转向点在内均间隔。而转向点序列和中的每一个关键点，以该点为中心，延纬度方向做垂线，这些垂线便为轨迹线。

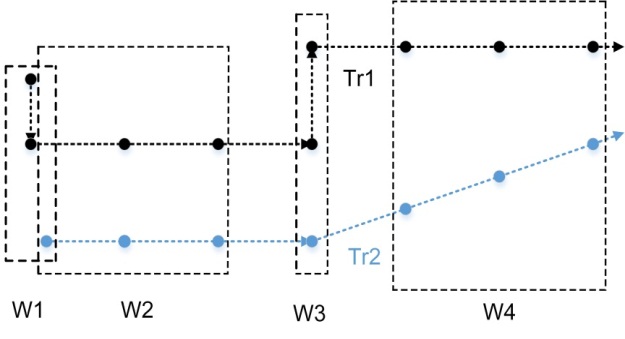


图2：轨迹线划分子轨迹示意图

当轨迹方向从只延经度变化转到只延纬度变化时，该段在轨迹线之间的距离近似等于一个点，而一个点与一条路径显然是不匹配的，证明沿纬度方向轨迹线划分轨迹是合理的。为改进的Hausdorff距离相似性度量临界阈值，用于衡量路径相似匹配的最大容错率。为路径与路径距离更小者的终止点，与分别为路径A、B的下一个轨迹转向点。

于是A，B俩转向点序列被划分为若干转向点子轨迹，通过各轨迹线间形成的转向点子轨迹段来进行子轨迹间的相似度匹配，从而使两条路径中的所有转向点都参与到匹配遍历计算中来，得到所有相似子轨迹的总长度，其所占匹配轨迹中较小长度的比例，与轨迹相似度标准阈值比较从而判断两条轨迹是否相似。

计算满足条件的子轨迹长度和得到相似轨迹之间的匹配度 :

（8）

其中，为相似子轨迹长度和，为A、B两条轨迹长度的更小值，为各子轨迹之间 Hausdorff距离的影响程度。

通过对所有拼车方案中的路径轨迹相似度匹配度排序，所对应的拼车方案k则为实时最优的拼车方案。

**4 实验**

**4.1数据来源**

通过随机选取武汉市行政道路图中民族大道、关山大道、南湖大道以及雄楚大道中的60条行车路线，模拟在同一搜索匹配周期内60名乘客的拼车需求，通过轨迹处理，路径相似度匹配得到最优的拼车方案。

轨迹预处理：首先，我们收集到了武汉市城市道路图，对其中的7千多条道路进行初步帅选。鉴于城市出租车行驶距离一般为2到30公里，所以我们提取了2公里到30公里之间的路径，作为初步筛选的结果。总计有1000多条路线符合要求。接下来提取转向点。根据转向点的取值范围，经过处理后，我们得到整条路线的转向点序列。

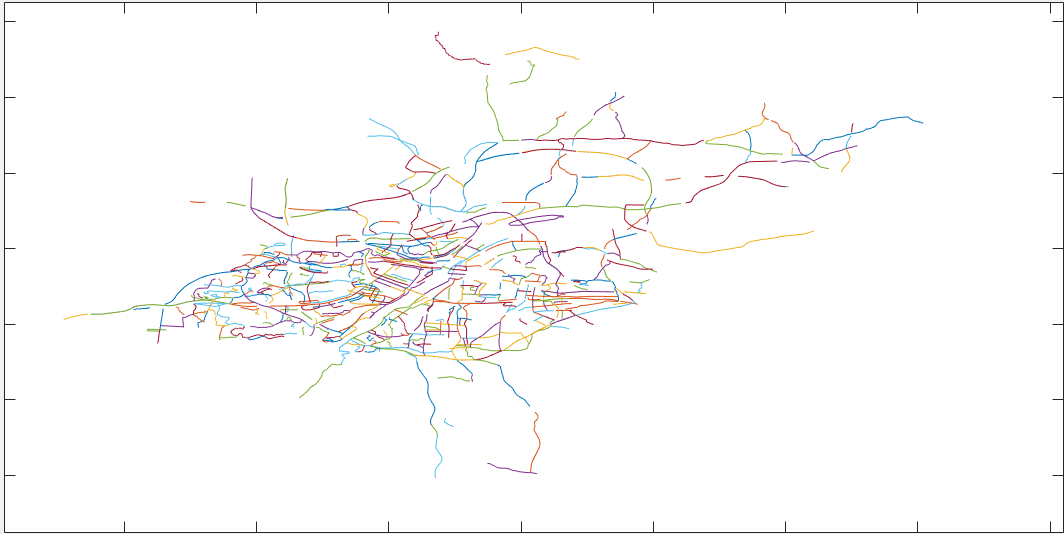


图3为原始的武汉市道路分布图

考虑到过多的转弯点在现实生活中很少出现，对于10个转向点以上的序列，我们予以舍弃。经过最终处理，我们得到了610多条2到10个点的转向点序列，作为我们轨迹预处理的输出结果，是下一步我们实验数据的来源。

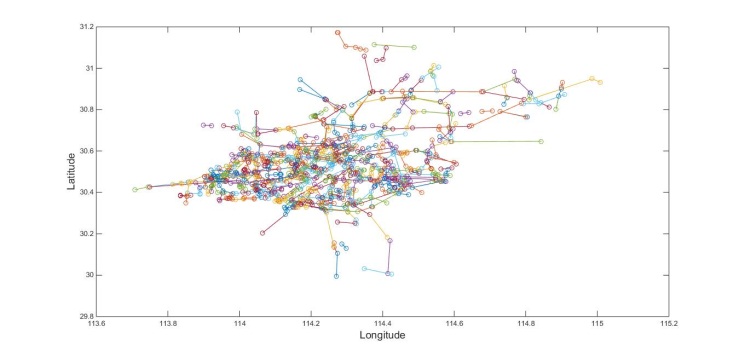


图4为处理后的转向点序列路线图

为了验证转向点选取的合理与必要性，我们将转向点提取前后的路径进行对比，由点集构成的路线处理为转向点序列后，每条路径的平均距离缩短0.3480km，两者平均距离差值的标准差为0.6411km，这说明距离的缩减不至于过于剧烈，且波动范围小。经过多次试验发现，提取转向点后的数据处理运行时间比没有提取转向点的提高72.2%，尽管计算转向点会付出一定代价，但是整体运行时间大幅度缩短保证了拼车匹配调度的实时性，减轻了服务器运算的开销与负担。

**4.2轨迹匹配**

经过初步筛选后，我们对剩余的路线进行轨迹匹配。需要拼车乘客的行驶轨迹转向点序列为集合B，已上车的乘客路线轨迹转向点序列为集合A。将B中的每条路线和A进行匹配，计算带有时间约束的Hausdorff距离，同时计算不带时间约束的Hausdorff距离。最终得到数据图：

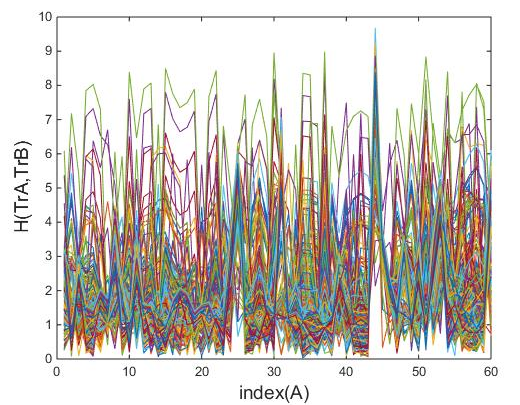


图5为带有时间约束的轨迹匹配图

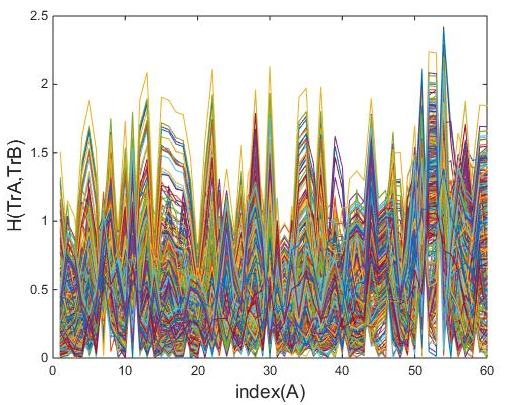


图6为不带时间约束的轨迹匹配图

其中，横坐标表示A路线集合；纵坐标代表B中的某条路线和A中的某条路线匹配算出的Hausdorff距离；一条折线段代表一个B路线。

图5和图6对比可以看出，带有时间约束的Hausdorff距离明显比不带时间约束的Hausdorff距离大，在这表明根据传统的路径匹配算法算出来两条相似路径在我们的算法中很有可能是并不相似的。此外，不同的B路线和同一条A路线匹配的结果差别也更大。在=1的情况下，B路线和A路线匹配的数量分布图如图7所示：

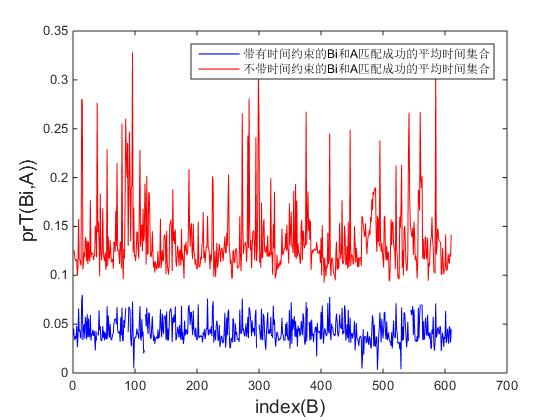


图7路线A、B匹配数量分布图

其中，纵坐标是数量B和A路线集合匹配成功的数量，横坐标是B路线的标号，红色部分是不带时间约束所绘制出来的匹配成功数目曲线，蓝

色部分是带有时间约束所绘制出来的匹配成功数目曲线。可以看出，蓝色曲线整体低于红色曲线，表明带有时间约束匹配成功的数目整体少于不带时间约束匹配成功的数目。

表2

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Similarity\_degree | =1 | | =1.2 | | =1.4 | | =1.6 | | =1.8 | | =2 | |
| Sequece\_B | total\_prT | Success\_rate | total\_prT | Success\_rate | total\_prT | Success\_rate | total\_prT | Success\_rate | total\_prT | Success\_rate | total\_prT | Success\_rate |
| Time\_ constraint | 0.0435h | 24.54% | 0.0526h | 36.55% | 0.0604h | 47.28% | 0.0665h | 56.39% | 0.0719h | 63.17% | 0.0772h | 69.00% |
| Without \_Time\_ constraint | 0.1353h | 89.91% | 0.1375h | 94.80% | 0.1395h | 97.42% | 0.1402h | 98.88% | 0.1405h | 99.63% | 0.1405h | 99.90% |

如表2所示，total\_prT代表所有需要拼车的用户的平均等待时间，Success\_rate代表B中一个需要拼车用户的路线和A中所有路线的匹配成功率，=1代表相似性度量临界阈值，Time\_ constraint为待有时间约束的计算结果，Without \_Time\_ constraint为不带时间约束的计算结果。

在=1且B于A已匹配成功的条件下，带有时间约束的拼车方案的用户平均等待时间明显少于不带时间约束的，虽然匹配成功率较低，但随着的增大，也就是相似性度量临界阈值的放宽，带有时间约束的匹配成功率越来越高。当=3时，整体的成功率达到84.06%。当然，因为放宽了相似性度量临界阈值，所以平均等待时间也增加，但仍然是不带时间约束等待时间的2/3。这表明加入时间约束后，用户的等待拼车的时间大幅度缩短，但匹配成功率仍然维持在较高的水平线上，就提高拼车用户体验而言，这种付出是值得的。同时，跟传统的MCT拼车算法比较，改进的Hausdorff拼车匹配算法无论是成功率还是平均等待时间上都要更优一些。在=3时，改进的Hausdorff拼车匹配算法成功率较MCT提高了14.29%，时间上缩短了0.0405h，整体优于MCT拼车算法。

**5 总结**

在拼车问题中，最为关键的是不同的路径的相似度精确计算，然后在进一步的研究中发现，时间代价对最终拼车匹配结果有极其重要的影响。因此我们提出一个将时间纳入计算的改进的Hausdorff距离的路径相似度匹配算法，通过利用路径上转向点上的经纬度和时间属性，来计算改进后的Hausdorff距离，然后通过分段划分的方法，进行路径之间的相似度匹配度计算。最后在实验中，将60条路径的数据代入与计算，通过实验添加时间约束及改进的Hausdorff距离与传统模型进行比较，匹配准确度以及减小时间代价方面有较大提高，验证了本文提出的方法在路径匹配度的计算上具有较好的效果。该领域在未来的进一步发展中，必将更侧重于路径匹配的准确度与及时性。而GPS的定位与道路的刻画方面也将在精确度上提出越来越苛刻的要求，这也是未来研究的重点。

**6 参考文献**

[1] The New York city taxicab fact book [EB /OL].[2015-11-16].http://www.schallerconsult.com/taxi /taxifb.pdf.

[2] Taxi of tomorrow survey[EB/OL].[2015-11-16]. http://www.nyc.gov/html/tlc/downloads/pdf/tot\_survey\_results\_02\_10\_11.pdf.

[3] HUANG Y, POWELL J W. Detecting regions of disequilibrium in taxi services under uncertainty[C] //SIGSPATIAL'12: Proceedings of the 20th Internatio-nal Conference on Advances in Geographic Information Systems. New York: ACM, 2012:139-148.

[4] Jia Yao, Anthony Chen, Seungkyu Ryu, Feng Shi[A general unconstrained optimization formulation for the combined distribution and assignment problem](http://www.cnki.net/kcms/detail/detail.aspx?filename=SJESD02D8B3E1656310BEE37&dbcode=SSJD&v=)[J]. Transportation Research Part B . 2014

[5] Cao Yanyan, Cui Zhiming, Wu Jian, Sun Yong, et al. An improved Hausdorff distance and spectral clustering vehivle trajectory pattern learning approach [J]. Computer Applications and Software . 2012

[6] Bulent catay .A new saving-based ant algorithm for the Vehicle Routing Problem with Simultaneous Pickup and Delivery[J].Expert Systems with Applications, 2010,37:6809-6817.

[7] TANG L L, ZHENG W B, WANG Z Q, et al. Space time analysis on the pick- up and drop off of taxi passengers based on GPS big data[J]. Journal of Geo- Information Science, 2015, 17( 10) : 1179-1186.

[8] FENG Tian. Dynamic Taxipooling Scheduling Algorithm Based on Sufferage [J], Computer Knowledge and Technology, 2011, 28(7):7019-7023

[9] SHAO Zeng-Zhen, WANG Hong-Guo, LIU Hong, et al. Single Carpooling Problem Based on Matching Degree Clustering Algorithm [J]. Journal of Software，2012,23(2):204-212.

[10] XIAO Q, HE R C, ZHANG W, et al. Algorithm research of taxi arpooling based on fuzzy clustering and fuzzy recognition[J]. Joural of Transportation Systems Engineering and Information Technoloy, 2014, 14( 5) : 119-125.

[11] Atev S, Miller G, Papanikolopoulos N P. Clustering of vehicle trajectories[J]. IEEE Trans on Intelligent Transportation Systems, 2010, 11(3): 647-657.

[12] LIU Chun, TAN Mengxi，SHAO Xiongkai, et al. Carpooling algorithm research based on location data mining [J]. Computer Engineering and Applicati- ions, 2015-12-31

[14] Zhang Xiaobin, Yang Dongshan. Hausdorff distance about spatial-temporal trajectory similarity based on time restriction[J].Application Research Of Computers, 2016-08-02

[15] 胡卫明，谢丹，谭铁牛，等．轨迹分布模式学习的层次自组织神经网络方法［J］．计算机学报．2003，26( 4) : 417-426．