

(19)中华人民共和国国家知识产权局



(12)发明专利申请

(10)申请公布号 CN 107247961 A

(43)申请公布日 2017.10.13

(21)申请号 201710324975.3

(22)申请日 2017.05.10

(71)申请人 西安交通大学

地址 710049 陕西省西安市碑林区咸宁西路28号

(72)发明人 曲桦 张艳鹏 刘军 赵季红

(74)专利代理机构 西安通大专利代理有限责任公司 61200

代理人 范巍

(51)Int.Cl.

G06K 9/62(2006.01)

G06N 7/02(2006.01)

权利要求书2页 说明书9页 附图5页

(54)发明名称

一种应用模糊轨迹序列的轨迹预测方法

(57)摘要

本发明提出一种应用模糊轨迹序列的轨迹预测方法,通过引入模糊轨迹序列模型,对用户的移动轨迹进行模糊化处理,并通过模糊时间序列的预测方法,对用户的运动轨迹进行预测。方法中对用户的轨迹进行了模糊化处理,设计了网格化的模糊化方法,由此实现了模糊轨迹序列模型。通过引入离群点处理的机制,本方法可以处理历史信息不足时的轨迹预测问题,通过引入离群点检测机制以及离群点预测模型,以降低离群点对算法整体性能的影响。

1. 一种应用模糊轨迹序列的轨迹预测方法,其特征在于:该预测方法包括以下步骤:

1) 针对任意一待预测用户,确定用户轨迹训练样本集的论域,所述用户轨迹训练样本为所述论域范围内若干个用户的轨迹序列,将用户轨迹训练样本集中的用户轨迹序列以截取的方式重构为若干个包含k个时刻用户位置坐标的高维轨迹序列,根据论域的网格化划分方案将所述高维轨迹序列模糊化为高维模糊轨迹序列,且满足使相同或相似的高维轨迹序列在模糊化后仍为相同或相似的高维模糊轨迹序列,得到用于高维模糊轨迹序列模型训练的集合;

2) 在每一个高维模糊轨迹序列后加入下一刻对应用户的位置坐标,得到高维模糊轨迹序列模型的训练集,对高维模糊轨迹序列模型的训练集应用聚类方法获得簇,并由簇中心构建相应的模糊逻辑规则,依据模糊逻辑规则对由用户轨迹训练样本集中的用户轨迹序列重构得到的高维轨迹序列进行划分:对于任意一个高维轨迹序列,若该高维轨迹序列不满足模糊逻辑规则,则将该高维轨迹序列加入离群点样本训练集合,利用离群点样本训练得到离群点预测模型;

3) 在所述论域范围内,将某一个待预测用户截止当前时刻的顺序采集的k个位置坐标转变为相应的高维轨迹序列,得到待预测样本,对待预测样本根据所述模糊逻辑规则进行划分,若待预测样本为满足模糊逻辑规则的一般样本,通过对簇中心最后S维分别求加权平均数获得用户轨迹预测结果,所述S的取值与用户位置坐标维数相同;若待预测样本为不满足模糊逻辑规则的离群点样本,则利用离群点预测模型进行轨迹预测。

2. 根据权利要求1所述一种应用模糊轨迹序列的轨迹预测方法,其特征在于:所述步骤1)具体包括以下步骤:

1.1) 计算所述轨迹训练样本集中用户坐标经、纬度的最值,则用户活动范围定义为经、纬度的最值所围成的矩形区域,根据该矩形区域确定用户轨迹训练样本集中用户轨迹序列的论域;

1.2) 将所述论域划分为网格状区域,每个网格状区域对应不同的模糊集;将所述轨迹训练样本集中全部用户轨迹序列转化为高维轨迹序列;将所有高维轨迹序列中坐标模糊化为对应的模糊集,形成高维模糊轨迹序列集合;

1.3) 对具有某一相同形式的高维模糊轨迹序列,计算这些高维模糊轨迹序列对应的模糊化前的高维轨迹序列中下一刻用户坐标的方差;将计算得到的所有不同形式的高维模糊轨迹序列对应的方差求和,作为对论域划分方案的评价指标;

1.4) 以最小化所述评价指标为目标,利用启发式优化算法找出尽可能好的论域划分方案,作为最终的论域划分方案;

1.5) 利用最终的论域划分方案,按照步骤1.2)生成用于高维模糊轨迹序列模型训练的集合。

3. 根据权利要求2所述一种应用模糊轨迹序列的轨迹预测方法,其特征在于:所述步骤1.2)中,对论域进行网格状划分过程中,分别在经、纬度方向上选取划分点,并根据划分点对论域进行划分,在选择划分点时,对划分点的数目通过设置上限和下限进行限制。

4. 根据权利要求3所述一种应用模糊轨迹序列的轨迹预测方法,其特征在于:选择等分方式对论域进行划分,在经、纬度方向上的划分点的数目均以指数形式增加,对由此形成的每个划分方案进行评价,选择使得评价指标最小的划分方案作为参考方案,计算上述每个

划分方案的评价指标与参考方案评价指标之间差的绝对值作为对应划分方案的相对评价指标；在上述所有划分方案中，选择相对评价指标大于设定阈值且划分点比参考方案划分点数目少的方案中，划分点数目最多的方案的划分点数目作为划分点数目的下限，选择相对评价指标大于设定阈值且划分点比参考方案划分点数目多的方案中，划分点数目最少的方案的划分点数目作为划分点数目的上限。

5. 根据权利要求1所述一种应用模糊轨迹序列的轨迹预测方法，其特征在于：所述步骤2)中，模糊逻辑规则的构建具体包括以下步骤：

2.1) 对步骤1)中生成的用于高维模糊轨迹序列模型训练的集合中每个高维模糊轨迹序列，在序列后新增一维，值为该序列对应的用户轨迹序列中下一刻用户的2维平面位置坐标，生成用于聚类的高维模糊轨迹序列训练集；

2.2) 对步骤2.1)中生成的用于聚类的高维模糊轨迹序列训练集，应用自适应高斯核的模糊C均值聚类算法生成高维轨迹序列簇，并计算簇中心作为基于一般样本的轨迹预测模型的模糊逻辑规则。

6. 根据权利要求5所述一种应用模糊轨迹序列的轨迹预测方法，其特征在于：所述步骤2.2)中，自适应高斯核的模糊C均值聚类算法在完成聚类算法的每一次迭代后，依据现有样本相对于簇中心的分布情况，重新计算簇的核宽度并更新。

7. 根据权利要求1所述一种应用模糊轨迹序列的轨迹预测方法，其特征在于：所述模糊逻辑规则定义为：假设 $x(t)$ 是由 $F(t-k), F(t-k+1), \dots, F(t-1)$ 决定的，且 $F(t)$ 代表t时刻用户坐标对应的模糊集， $x(t)$ 为t时刻的用户坐标，那么就存在模糊逻辑规则 $F(t-k), F(t-k+1), \dots, F(t-1) \rightarrow x(t)$ 。

8. 根据权利要求1所述一种应用模糊轨迹序列的轨迹预测方法，其特征在于：计算高维轨迹序列与所有簇间的隶属度，若所有隶属度均小于设定阈值 η ，则不满足模糊逻辑规则；若存在大于等于设定阈值 η 的隶属度，则满足模糊逻辑规则。

9. 根据权利要求1所述一种应用模糊轨迹序列的轨迹预测方法，其特征在于：所述离群点预测模型的建立方法包括以下步骤：将离群点样本对应的用户轨迹序列转化为间隔时间的位移序列以及位移方向的序列，然后采用Levenberg-Marquardt算法分别构建关于下一个间隔时间的位移以及位移方向的预测模型；Levenberg-Marquardt算法需要通过多次迭代来避免陷入局部极小，每次迭代均在设定范围内随机生成初始权向量，并选择最优的结果；Levenberg-Marquardt算法的初始权向量设定时，每一维度的值大于前一维度。

10. 根据权利要求1所述一种应用模糊轨迹序列的轨迹预测方法，其特征在于：所述步骤3)中，对簇中心最后S维求加权平均数时，若待预测样本与簇间隶属度大于等于阈值 η ，则权值为该隶属度；否则，权值为0。

一种应用模糊轨迹序列的轨迹预测方法

技术领域

[0001] 本发明涉及用户轨迹序列预测问题,特别涉及一种引入了模糊轨迹序列模型的轨迹预测方法。

背景技术

[0002] 1、轨迹预测

[0003] 随着定位技术的发展及广泛应用,基于位置的服务(Location-based Service, LBS)逐渐成为生活中不可获缺的部分。通过分析用户的轨迹信息,挖掘其中隐含的用户信息来完善用户的服务体验,成为数据挖掘的一个重要领域。通过用户的轨迹信息对用户的轨迹进行预测,对导航服务、交通管理以及基于位置的广告投放等应用至关重要,成为了轨迹挖掘的一个研究热点。

[0004] 2、轨迹预测的常见方法

[0005] 轨迹预测问题主要分为长期预测以及短期预测:长期预测适用于时间跨度长、采样间隔大的轨迹样本预测问题,用户的移动目标对用户轨迹起主要决定作用,主要通过历史用户轨迹的挖掘,发掘用户相似轨迹或规律用于样本的预测;短期预测则适用于时间跨度短、采样间隔小的轨迹样本预测问题,用户的移动习惯对用户轨迹起主要决定作用,主要通过对短期用户行为模式进行分析建模完成预测。

[0006] 轨迹预测的方法可以分为两类,分别是统计类方法以及机器学习类方法。统计类方法主要通过数学领域的统计模型以及概率模型对用户轨迹进行建模,建立的模型简单,算法运行效率高,但因此难以有效的处理现实中复杂的数据集;机器学习类方法通过应用机器学习领域的预测技术,挖掘轨迹信息建立用户轨迹模型进行预测,能够建立起复杂的用户轨迹模型,但算法效率低,算法运行时间长。

[0007] 现有的轨迹预测方法都立足于所得到的用户轨迹是大致精确的这一前提,不论是统计学的预测方法还是机器学习的方法,均对样本的数据不做过多的预处理而直接使用,而在实际生活中通过定位服务所得到的轨迹序列均包含误差,这种误差影响了现有方法的预测精度。

发明内容

[0008] 本发明的目的在于针对不精确、存在偏移的用户轨迹序列的预测问题,提出一种应用模糊轨迹序列的轨迹预测方法,该方法通过引入模糊轨迹序列对用户的轨迹进行预测,使得预测方法能够处理不精确的、存在坐标偏移的轨迹序列,能够更广泛的应用于现实生活中的轨迹预测。

[0009] 为了达到上述目的,本发明采用了以下技术方案:

[0010] 1)针对任意一待预测用户,确定用户轨迹训练样本集的论域,所述用户轨迹训练样本为所述论域范围内若干个用户的轨迹序列,将用户轨迹训练样本集中的用户轨迹序列以截取的方式重构为若干个包含k个时刻用户位置坐标的高维轨迹序列,根据论域的网格

化划分方案将所述高维轨迹序列模糊化为高维模糊轨迹序列,且满足使相同或相似的高维轨迹序列在模糊化后仍为相同或相似的高维模糊轨迹序列,得到用于高维模糊轨迹序列模型训练的集合;

[0011] 2) 在每一个高维模糊轨迹序列后加入下一刻对应用户的位置坐标,得到高维模糊轨迹序列模型的训练集,对高维模糊轨迹序列模型的训练集应用聚类方法获得簇,并由簇中心构建相应的模糊逻辑规则,依据模糊逻辑规则对由用户轨迹训练样本集中的用户轨迹序列重构得到的高维轨迹序列进行划分,对于任意一个高维轨迹序列:若该高维轨迹序列不满足模糊逻辑规则,则将该高维轨迹序列加入离群点样本训练集合,利用离群点样本训练得到离群点预测模型;

[0012] 3) 在所述论域范围内,将某一个待预测用户截止当前时刻的顺序采集的k个位置坐标转变为相应的高维轨迹序列,得到待预测样本,对待预测样本根据所述模糊逻辑规则进行划分,若待预测样本为满足模糊逻辑规则的一般样本,通过对簇中心最后S维分别求加权平均数获得用户轨迹预测结果,所述S的取值与用户位置坐标维数相同;若待预测样本为不满足模糊逻辑规则的离群点样本,则利用离群点预测模型进行轨迹预测。

[0013] 所述步骤1)具体包括以下步骤:

[0014] 1.1) 计算所述轨迹训练样本集中用户坐标经、纬度的最值,则用户活动范围定义为经、纬度的最值所围成的矩形区域,根据该矩形区域确定用户轨迹训练样本集中用户轨迹序列的论域;

[0015] 1.2) 将所述论域划分为网格状区域,每个网格状区域对应不同的模糊集;将所述轨迹训练样本集中全部用户轨迹序列转化为高维轨迹序列;将所有高维轨迹序列中坐标模糊化为对应的模糊集,形成高维模糊轨迹序列集合;

[0016] 1.3) 对具有某一相同形式的高维模糊轨迹序列,计算这些高维模糊轨迹序列对应的模糊化前的高维轨迹序列中下一刻用户坐标的方差;将计算得到的所有不同形式的高维模糊轨迹序列对应的方差求和,作为对论域划分方案的评价指标;

[0017] 1.4) 以最小化所述评价指标为目标,利用启发式优化算法找出尽可能好的论域划分方案,作为最终的论域划分方案;

[0018] 1.5) 利用最终的论域划分方案,按照步骤1.2)生成用于高维模糊轨迹序列模型训练的集合。

[0019] 所述步骤1.2)中,对论域进行网格状划分过程中,分别在经、纬度方向上选取划分点,并根据划分点对论域进行划分,在选择划分点时,对划分点的数目通过设置上限和下限进行限制。

[0020] 选择等分方式对论域进行划分,在经、纬度方向上的划分点的数目均以指数形式增加,对由此形成的每个划分方案进行评价,选择使得评价指标最小的划分方案作为参考方案,计算上述每个划分方案的评价指标与参考方案评价指标之间差的绝对值作为对应划分方案的相对评价指标;在上述所有划分方案中,选择相对评价指标大于设定阈值且划分点比参考方案划分点数目少的方案中,划分点数目最多的方案的划分点数目作为划分点数目的下限,选择相对评价指标大于设定阈值且划分点比参考方案划分点数目多的方案中,划分点数目最少的方案的划分点数目作为划分点数目的上限。

[0021] 所述步骤2)中,模糊逻辑规则的构建具体包括以下步骤:

[0022] 2.1) 对步骤1) 中生成的用于高维模糊轨迹序列模型训练的集合中每个高维模糊轨迹序列, 在序列后新增一维, 值为该序列对应的用户轨迹序列中下一刻用户的2维平面位置坐标, 生成用于聚类的高维模糊轨迹序列训练集;

[0023] 2.2) 对步骤2.1) 中生成的用于聚类的高维模糊轨迹序列训练集, 应用自适应高斯核的模糊C均值聚类算法生成高维轨迹序列簇, 并计算簇中心作为基于一般样本的轨迹预测模型的模糊逻辑规则。

[0024] 所述步骤2.2) 中, 自适应高斯核的模糊C均值聚类算法在完成聚类算法的每一次迭代后, 依据现有样本相对于簇中心的分布情况, 重新计算簇的核宽度并更新。

[0025] 所述模糊逻辑规则定义为: 假设 $x(t)$ 是由 $F(t-k), F(t-k+1), \dots, F(t-1)$ 决定的, 且 $F(t)$ 代表 t 时刻用户坐标对应的模糊集, $x(t)$ 为 t 时刻的用户坐标, 那么就存在模糊逻辑规则 $F(t-k), F(t-k+1), \dots, F(t-1) \rightarrow x(t)$ 。

[0026] 计算高维轨迹序列与所有簇间的隶属度, 若所有隶属度均小于设定阈值 η , 则该样本不满足模糊逻辑规则; 若存在大于等于设定阈值 η 的隶属度, 则样本满足模糊逻辑规则。

[0027] 所述离群点预测模型的建立方法包括以下步骤: 将离群点样本对应的用户轨迹序列转化为间隔时间的位移序列以及位移方向的序列, 然后采用Levenberg-Marquardt算法分别构建关于下一个间隔时间的位移以及位移方向的预测模型; Levenberg-Marquardt算法需要通过多次迭代来避免陷入局部极小, 每次迭代均在设定范围内随机生成初始权向量, 并选择最优的结果; Levenberg-Marquardt算法的初始权向量设定时, 每一维度的值大于前一维度。

[0028] 所述步骤3) 中, 对簇中心最后S维求加权平均数时, 若待预测样本与簇间隶属度大于等于阈值 η , 则权值为该隶属度; 否则, 权值为0。

[0029] 本发明具有以下有益效果:

[0030] 本发明通过引入模糊轨迹序列模型, 并通过模糊时间序列的预测方法, 对用户的运动轨迹进行预测。方法中对用户的移动轨迹进行了模糊化处理, 设计了网格化的模糊化方法, 由此实现了模糊轨迹序列模型。通过引入离群点处理的机制, 本方法可以处理历史信息不足时的轨迹预测问题, 并引入离群点检测机制以及离群点预测模型, 以降低离群点对算法整体性能的影响。本发明中通过引入模糊轨迹序列模型对用户轨迹序列进行建模, 使得模型可以处理不精确的、存在坐标偏移的轨迹序列, 为处理现实生活中定位服务存在的偏差的问题提供了新的解决思路。

[0031] 进一步的, 本发明采用离群点判定机制及离群点预测模型, 可以识别训练样本不足导致的离群点这一特殊情况, 同时通过对预测模型的调整, 改善对离群点样本的预测精度, 提升算法的鲁棒性以及预测的平均精度。

[0032] 进一步的, 本发明中采用的聚类方法区别于传统的核版本的模糊C均值聚类, 采用动态调整的核宽度策略, 同时引入核宽度估算机制, 使得聚类方法能自适应于不同尺寸的簇, 提升了预测算法的鲁棒性。

附图说明

[0033] 图1为应用模糊轨迹序列的轨迹预测方法流程图;

[0034] 图2为对用户轨迹论域的划分方案及模糊集示意图;

- [0035] 图3为论域划分方案搜索流程图；
- [0036] 图4为应用模糊轨迹序列的轨迹预测方法中模糊逻辑规则构建流程图；
- [0037] 图5为应用离群点预测模型的轨迹序列表征方式转换示意图；
- [0038] 图6为应用模糊轨迹序列的轨迹预测方法对南加州展会数据集的仿真结果。

具体实施方式

[0039] 为了使本发明的内容、效果以及优点更加清楚明白，下面结合附图和实施例对本发明进行详细描述。

[0040] 本发明是基于模糊逻辑的预测方法，通过对轨迹数据进行模糊化预处理，使得预测方法可以处理获得的不准确的、有偏差的轨迹预测问题。本发明设计了一种自适应的高斯核模糊C均值聚类方法，通过引入核宽度的调整算法，使聚类方法能够更适用于更为广泛的各种情况。同时本发明采用离群点模型应对训练样本稀疏引发的问题。图1中展示了本发明的策略流程，通过计算待预测样本与每个提取到的模糊逻辑规则对应的簇间隶属度，判定样本是否为离群点，对一般性样本采用加权平均数进行预测，而对离群点样本采用线性模型预测。该方法优势在于能够应对现实生活中轨迹预测问题，可以容忍数据采集中出现的偏差，同时也适用于训练样本数量不足、覆盖不全面的情况。

- [0041] (一) 定义用户轨迹论域划分方案及轨迹序列的模糊化处理

[0042] 用户轨迹的论域定义了用户的活动范围以及用户可能到达的区域，为了使得定义的论域能够覆盖用户所有的轨迹，计算用户经度坐标的最值 x_{min} 、 x_{max} 以及纬度坐标的最值 y_{min} 、 y_{max} ，参见图2，用户的活动范围可以用计算得到的经纬度坐标最值围成的矩形区域 $[(x_{min}, y_{min}), (x_{min}, y_{max}), (x_{max}, y_{min}), (x_{max}, y_{max})]$ 来表示，考虑到用户可能到达的区域，将用户的活动范围扩大为 $[(x'_{min}, y'_{min}), (x'_{min}, y'_{max}), (x'_{max}, y'_{min}), (x'_{max}, y'_{max})]$ ，其中：

$$[0043] \Delta x = 0.1 \cdot (x_{max} - x_{min})$$

$$[0044] \Delta y = 0.1 \cdot (y_{max} - y_{min})$$

$$[0045] x'_{min} = x_{min} - \Delta x$$

$$[0046] x'_{max} = x_{max} + \Delta x$$

$$[0047] y'_{min} = y_{min} - \Delta y$$

$$[0048] y'_{max} = y_{max} + \Delta y$$

[0049] 该区域 $[(x'_{min}, y'_{min}), (x'_{min}, y'_{max}), (x'_{max}, y'_{min}), (x'_{max}, y'_{max})]$ 定义为用户轨迹的论域。

- [0050] 参见图3，对用户轨迹的论域进行划分包括以下步骤：

[0051] 1) 采用等时间间隔方式获取用户的位置坐标，构成用户轨迹序列。设定时间间隔 t ，对于用户的一次轨迹采样，每隔时间 t 收集一次位置坐标，并将坐标序列存储作为用户的一条轨迹。对于记录的用户的一条含有 n 个采样点的轨迹，第 i 个样本点对应的用户坐标为 (x_i, y_i) ，其中， x_i 为经度坐标， y_i 为纬度坐标，则可得到：

$$[0052] x_{min} = \min(x_1, x_2, \dots, x_n)$$

$$[0053] x_{max} = \max(x_1, x_2, \dots, x_n)$$

$$[0054] y_{min} = \min(y_1, y_2, \dots, y_n)$$

$$[0055] y_{max} = \max(y_1, y_2, \dots, y_n)$$

[0056] 由此可确定用户轨迹的论域 $[(x'_{\min}, y'_{\min}), (x'_{\min}, y'_{\max}), (x'_{\max}, y'_{\min}), (x'_{\max}, y'_{\max})]$ 。

[0057] 2) 参见图3,对于用户轨迹的论域 $[(x'_{\min}, y'_{\min}), (x'_{\min}, y'_{\max}), (x'_{\max}, y'_{\min}), (x'_{\max}, y'_{\max})]$,分别在经度和纬度范围内选取各自的划分点 $(p_1^x, p_2^x, \dots, p_u^x)$ 和 $(p_1^y, p_2^y, \dots, p_v^y)$,由这些划分点确定论域划分的网格。每个网格对应一个模糊集 S_i $(i=1, 2, \dots, (u+1) \cdot (v+1))$,这些模糊集构成一种用户轨迹论域划分方案。

[0058] 3) 对用户的一条轨迹序列 $T = [(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$,对每个时刻,将该时刻用户位置坐标与前 $k-1$ 个连续时刻的位置坐标合成一个 k 维轨迹序列,将一条轨迹序列转化为 $n-k+1$ 条 k 维轨迹序列:

[0059] $T_1 = [(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)]$

[0060] $T_2 = [(x_2, y_2), (x_3, y_3), \dots, (x_{k+1}, y_{k+1})]$

[0061] ...

[0062] $T_{n-k+1} = [(x_{n-k+1}, y_{n-k+1}), (x_{n-k+2}, y_{n-k+2}), \dots, (x_n, y_n)]$

[0063] 4) 对得到的每条 k 维轨迹序列 $T_i = [(x_i, y_i), (x_{i+1}, y_{i+1}), \dots, (x_{i+k-1}, y_{i+k-1})]$,将每个轨迹中每个坐标点替换为所在网格区域对应的模糊集,得到 k 维模糊轨迹序列:

[0064] $T_i^f = [S_{f(i)}, S_{f(i+1)}, \dots, S_{f(i+k-1)}]$

[0065] 其中, $f(t)$ 代表 t 时刻用户坐标所在的模糊集下标, $t=i, i+1, \dots, i+k-1$ 。

[0066] 5) 对所得到的划分方案进行评估,并通过利用启发式优化算法寻找到一个可行划分方案。

[0067] 6) 根据得到的可行的论域划分方案,执行步骤3) 和步骤4) 生成 k 维模糊轨迹序列,这些序列构成了用于高维模糊轨迹序列模型训练的集合。

[0068] 在步骤2) 中,划分点的数目需要满足设置的上、下限,划分点数目上、下限的确定方法为:利用等分的方法,依次在论域的经度和纬度范围内选择 $1, 10, \dots, 10^b$ 个等分划分点;并分别根据步骤2) 中描述生成等分用户轨迹论域划分方案,其中 b 取值为使得等分后的每个区间长度小于数据集的分辨率的最小值;针对每个等分用户轨迹论域划分方案,执行步骤3) 和步骤4),并对这些等分用户轨迹论域划分方案进行评价,并给出评价指标值;假设包含 10^1 个等份划分点的论域划分方案的评价指标值为 $e_1, l=0, 1, 2, \dots, b$,且所有方案中评价指标的最小值为 e_k :则对于 $l < k$,若满足 $e_l > (1 + \varphi) \cdot e_k$,且对于任意整数 p $(1 < p \leq k)$, $e_p \leq (1 + \varphi) \cdot e_k$,则 10^1 为划分点数目的下限;而对于 $l > k$,若满足 $e_l > (1 + \varphi) \cdot e_k$,且对于任意整数 p $(k \leq p < l)$, $e_p \leq (1 + \varphi) \cdot e_k$,则 10^1 为划分点数目的上限; φ 的取值范围为 $(0, 1]$,优选为0.2。

[0069] (二) 用户轨迹论域划分方案的评价标准及可行方案的搜索方法

[0070] 良好的用户论域划分方案,应满足使相同或相似的 k 维用户轨迹序列,模糊化后仍为相同或相似的模糊轨迹序列。因而对一种用户轨迹论域的划分方案评价过程包括以下步骤:

[0071] 1) 针对(一)中步骤6) 得到的用于高维模糊轨迹序列模型训练的集合,将其中形式相同的 k 维模糊轨迹序列归入同一个集合;

[0072] 2) 对其中所有具有和 T_i^f 相同形式的 k 维模糊轨迹序列所构成的集合 $S_{T_i^f} = \{T_j^f \mid T_j^f = T_i^f\}$, 计算集合中所有序列对应的模糊化前的高维 (k 维) 轨迹序列的下一刻用户位置的方差, 即计算集合 $\{(x_{j+k}, y_{j+k}) \mid T_j^f \in S_{T_i^f}\}$ 的方差, 作为 T_i^f 形式对应的方差;

[0073] 3) 针对用于高维模糊轨迹序列模型训练的集合中所有不同的 k 维模糊轨迹序列形式, 利用步骤 2) 计算该形式对应的方差, 并求得所有的方差之和; 当得到的方差之和越小, 则证明划分方案性能越好。

[0074] 参见图 3, 采用启发式优化算法搜索得到最佳划分方案, 为了建立问题的解空间, 将用户轨迹的经纬度范围分别等分为由 u' 、 v' 个划分点隔开的小区域, 其中, $u' >> u$ 且 $v' >> v$ 。则问题转化为分别在 u' 个经度划分点和 v' 个纬度划分点中分别选取 u 、 v 个划分点的问题, 通过启发式算法搜索该问题的最优解, 并作为最终采用的划分方案。

[0075] (三) 应用用户 k 维模糊轨迹序列构建 k 维模糊逻辑规则

[0076] k 维模糊逻辑规则定义为: 假设 $x(t)$ 是由 $F(t-k), F(t-k+1), \dots, F(t-1)$ 决定的, 且 $F(t)$ 代表 t 时刻用户坐标对应的模糊集, $x(t)$ 为 t 时刻的用户坐标, 那么就存在模糊逻辑规则 $F(t-k), F(t-k+1), \dots, F(t-1) \rightarrow x(t)$ 。

[0077] k 维模糊逻辑规则构建方法包括以下步骤:

[0078] 1) 对所有的 k 维用户轨迹序列 $\{T_i \mid i=1, 2, \dots, n-k+1\}$, 应用所选的论域划分方式, 进行模糊化处理, 得到 k 维模糊轨迹序列集 $T^f = \{T_i^f \mid i=1, 2, \dots, n-k+1\}$;

[0079] 2) 对任意一条 k 维模糊轨迹序列 $T_i^f \in T^f$, 将该 k 维模糊轨迹序列后加入下一刻用户位置 (x_{j+k}, y_{j+k}) , 构成 k+2 维模糊轨迹序列:

[0080] $T'_j = [S_f(j), S_f(j+1), \dots, S_f(j+k-1), x_{j+k}, y_{j+k}]$

[0081] 将所有 k 维模糊轨迹序列转化为这样的 k+2 维模糊轨迹序列并构成集合。

[0082] 3) 对所有的 k+2 维模糊轨迹序列, 应用自适应高斯核的模糊 C 均值聚类算法, 获得聚类结果, 得到的每个簇中心即为一条 k+2 维模糊逻辑规则。

[0083] 由于模糊 C 均值聚类算法容易陷入局部极小, 所以最终得到的结果受初始给定的簇中心影响, 为了消除这种影响而得到性能较好的 k+2 维模糊逻辑规则, 需要迭代该过程, 每次均随机给予初始聚类中心, 并最终选择性能最好的结果, 并由此最终构建得到 k+2 维模糊逻辑规则; 对得到的聚类结果进行性能评价的方法为, 利用聚类结果构建的 k+2 维模糊逻辑规则对训练集中的样本进行预测, 并依据预测结果的均方根误差来判定性能, 均方根误差越小, 代表所得到的模型性能越好。

[0084] (四) 自适应高斯核模糊 C 均值聚类算法流程

[0085] 参见图 4, 自适应高斯核模糊 C 均值聚类算法通过引入高斯核函数作为隶属度函数, 一方面减小了离群点样本对簇本身的影响, 另一方面提升了簇的鲁棒性; 通过自适应的核宽度设置, 使得到的簇能够更好的适应不同的样本分布。算法包括以下步骤:

[0086] 1) 迭代 k+2 次, 第 h ($0 < h \leq k$) 次迭代中, 在用户轨迹论域范围内随机生成 r_h 个值, 后两次迭代中, 分别在用户轨迹论域的经度和纬度范围内生成 r_{k+1} 和 r_{k+2} 个值, 并组合成

$$r = \prod_{h=1}^{k+2} r_h \text{ 个簇中心; 对第 } i \text{ 个簇中心, 为其核宽度赋一个初始值 } \sigma_i^0.$$

[0087] 2) 通过模糊C均值聚类方法迭代过程迭代一次,迭代过程中,采用簇对应核宽度下的高斯核函数作为该簇的隶属度函数;在第h次迭代过程中,k+2维模糊轨迹序列T'j对簇中心Ci对应的簇的隶属度可由下式计算:

$$[0088] \text{mem}(C_i, T'_j) = \exp\left(-\frac{\|T'_j - C_i\|}{2(\sigma_{C_i}^{h+1})^2}\right)$$

[0089] 其中 $\|T'_j - C_i\|$ 表示 T'_j 与 C_i 间的欧式距离;

[0090] 3) 利用Parzen窗估计法,对每个得到的簇,根据样本关于簇中心的分布情况,计算新的核宽度并更新;

[0091] 4) 迭代直至聚类结果收敛,并给出包含k+2维模糊轨迹序列的簇中心的聚类结果。

[0092] 由于模糊轨迹序列中包含模糊集,在计算模糊化后的样本对簇的隶属度以及样本间距离时,定义在计算含有模糊轨迹序列的欧式距离时,将模糊集用模糊集对应的矩形区域的中心坐标值来代替。

[0093] (五) 离群点样本判定及离群点预测模型构建

[0094] k维模糊逻辑规则构建完成后,对训练集中样本 T_i ,计算样本对得到的簇中心 C_j 对应的簇的隶属度,隶属度函数采用高斯核函数 $\text{mem}_k(T_i, C_j)$,核宽度为聚类结果给出的簇中心 C_j 的核宽度,由于 T_i 为k维向量,因而 $\text{mem}_k(T_i, C_j)$ 的计算过程中仅考虑 C_j 的前k维数据;对任意样本 T_i ,若存在簇中心 C_j ,使得 $\text{mem}_k(T_i, C_j) \geq \eta$,则表明样本可以被所构建的k维模糊逻辑规则所表征;若对于任意簇中心 C_j , $\text{mem}_k(T_i, C_j) < \eta$,则样本不被所构建的k维模糊逻辑规则所表征,样本不适合于一般样本预测模型,将样本加入离群点样本集合。

[0095] 对由训练集中得到的离群点样本集合,囿于样本数目不足,因此采用简单的线性模型进行学习训练,使用Levenberg-Marquardt算法得到离群点的预测模型;利用Levenberg-Marquardt算法构建模型时,包括以下步骤:

[0096] 1) 参见图5,将用户轨迹序列 $T = [(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$ 转化为位移与位移方向表征的序列:

$$[0097] T^p = [(\rho_2, \theta_2), (\rho_3, \theta_3), \dots, (\rho_n, \theta_n)]$$

[0098] 其中, (ρ_i, θ_i) 为 (x_i, y_i) 关于以 (x_{i-1}, y_{i-1}) 为极点以x轴正方向为极轴的极坐标系表示, $i=2, 3, \dots, n$;

[0099] 2) 对离群点样本集合中任意样本 T'_i ,生成离群点位移预测模型训练样本 $[\rho_{i+1}, \rho_{i+2}, \dots, \rho_{i+k}]$ 以及位移方向预测模型训练样本 $[\theta_{i+1}, \theta_{i+2}, \dots, \theta_{i+k}]$,并分别构成位移预测模型训练集和位移方向预测模型训练集;

[0100] 3) 采用Levenberg-Marquardt算法构建预测模型,即分别在离群点位移预测模型训练集以及位移方向预测模型训练集上进行训练,得到位移预测模型和位移方向预测模型;两个预测模型得到的预测结果可以运算出下一刻用户位置的预测值。

[0101] 由于Levenberg-Marquardt算法会陷入局部极小,因而需要随机给出符合条件的多个初始参数,随机生成的初始参数需要符合的条件为:对于生成的初始参数,变量的每一维度的值必须大于前一维度的值。迭代找到使得Levenberg-Marquardt算法自身目标函数取得最小值的初始参数,用来构建离群点预测模型。

[0102] (六) 利用模糊轨迹序列模型预测用户下一刻位置

[0103] 对用户轨迹序列进行预测,即是根据用户前k时刻用户位置坐标,预测下一刻用户的位置。则可得到待预测k维用户轨迹序列 $T_p^k = [(x_1^p, y_1^p), (x_2^p, y_2^p), \dots, (x_k^p, y_k^p)]$ 。对得到的k维模糊逻辑规则对应的簇中心 C_i ,计算待预测样本得到的k维轨迹序列与每个簇间隶属度 $mem_k(T_p, C_i)$,若存在i,使得 $mem_k(T_p, C_i) \geq n$,则为一般样本,则对各个簇中心最后2维分别求加权平均数,得到用户下一刻位置,其中,若待预测样本与簇间隶属度大于等于阈值n,则权值为该隶属度;否则,权值为0:

$$[0104] (\sum_{i=1}^r mem_k(T_p, C_i) \cdot C_i(k+1), \sum_{i=1}^r mem_k(T_p, C_i) \cdot C_i(k+2))$$

[0105] 其中,r表示簇中心的个数, $C_i(k+1)$ 、 $C_i(k+2)$ 表示簇中心 C_i 的第k+1、k+2维值;若不存在i,使得 $mem_k(T_p, C_i) \geq n$,待预测样本视为离群点样本, T_p 转化为k-1维的用户位移序列以及位移角度序列,根据Levenberg-Marquardt算法得到的参数分别求得用户下一刻的位移 ρ_p 以及位移角度 θ_p ,用户下一刻位置计算为:

$$[0106] (\rho_p \cdot \cos \theta_p + x_k^p, \rho_p \cdot \sin \theta_p + y_k^p)$$

[0107] x_k^p 、 y_k^p 为第k时刻用户位置坐标;

[0108] (七) 仿真实验

[0109] 参见图6,对于用户的轨迹数据,采用奥兰多迪士尼、纽约以及南加州展会上采集到的用户轨迹数据集,数据采样时间间隔为30秒一次,3个数据集分别包含41条、39条、19条用户轨迹,分别包含44386个、39554个、5861个用户坐标,用户坐标以某固定点为原点的直角坐标系表示,单位为英尺。

[0110] 算法分别在奥兰多、纽约以及南加州展会上的3个用户轨迹数据集上运行,对于每个数据集上,选取其中每条轨迹前90%的用户轨迹序列作为学习算法的训练集,选取剩余的10%的用户轨迹作为算法验证的测试集,算法在每个数据集间独立运行。

[0111] 仿真参数设计如表1所示:

[0112] 表1. 仿真参数

| 仿真参数 | 参数值配置 |
|----------------------------|-------|
| 多维用户轨迹序列维度 | 5 |
| 初始核宽度 | 1 |
| 离群点判定阈值 | 0.6 |
| [0113] 候选经度划分点间隔 | 1m |
| 候选纬度划分点间隔 | 1m |
| 聚类算法最大迭代次数 | 200 次 |
| Levenberg-Marquardt 算法迭代次数 | 100 次 |

[0114] 本发明应用模糊轨迹序列模型的轨迹预测方法,例如,在奥兰多用户数据集上,对测试集样本的预测均方根误差为13.2746英尺、13.4603英尺、13.9036英尺。

[0115] 经过多次仿真实验,用户轨迹序列维度k的取值范围为 $3 \leq k \leq 10$, $k < 3$ 时,预测误差

较大,当 $k>10$ 时,运算复杂度较高,且会引起过拟合问题;离群点判定阈值 η 的取值范围为 $0.1\leq\eta\leq0.6$,阈值过小则会无法有效识别离群点,阈值过大则会降低模型的适用范围。

[0116] 本发明通过模糊时间序列的预测方法,对用户的运动轨迹进行预测,方法中对用户的轨迹进行了模糊化处理,设计了网格化的模糊化方法,由此实现了模糊轨迹序列模型,使得预测方法能够处理现实中获取的不精确的、有偏差的用户轨迹信息。通过引入离群点处理的机制,本方法可以处理历史信息不足时的轨迹预测问题,并引入离群点检测机制以及离群点预测模型,以降低离群点对算法整体性能的影响,提升了算法的总体性能,改善了算法的鲁棒性。

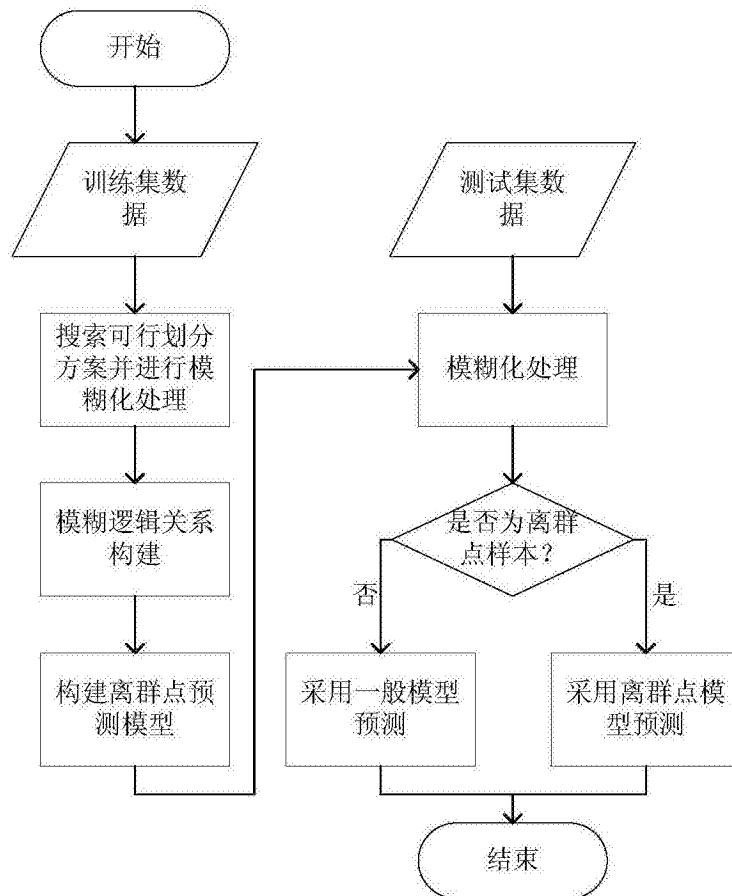


图1

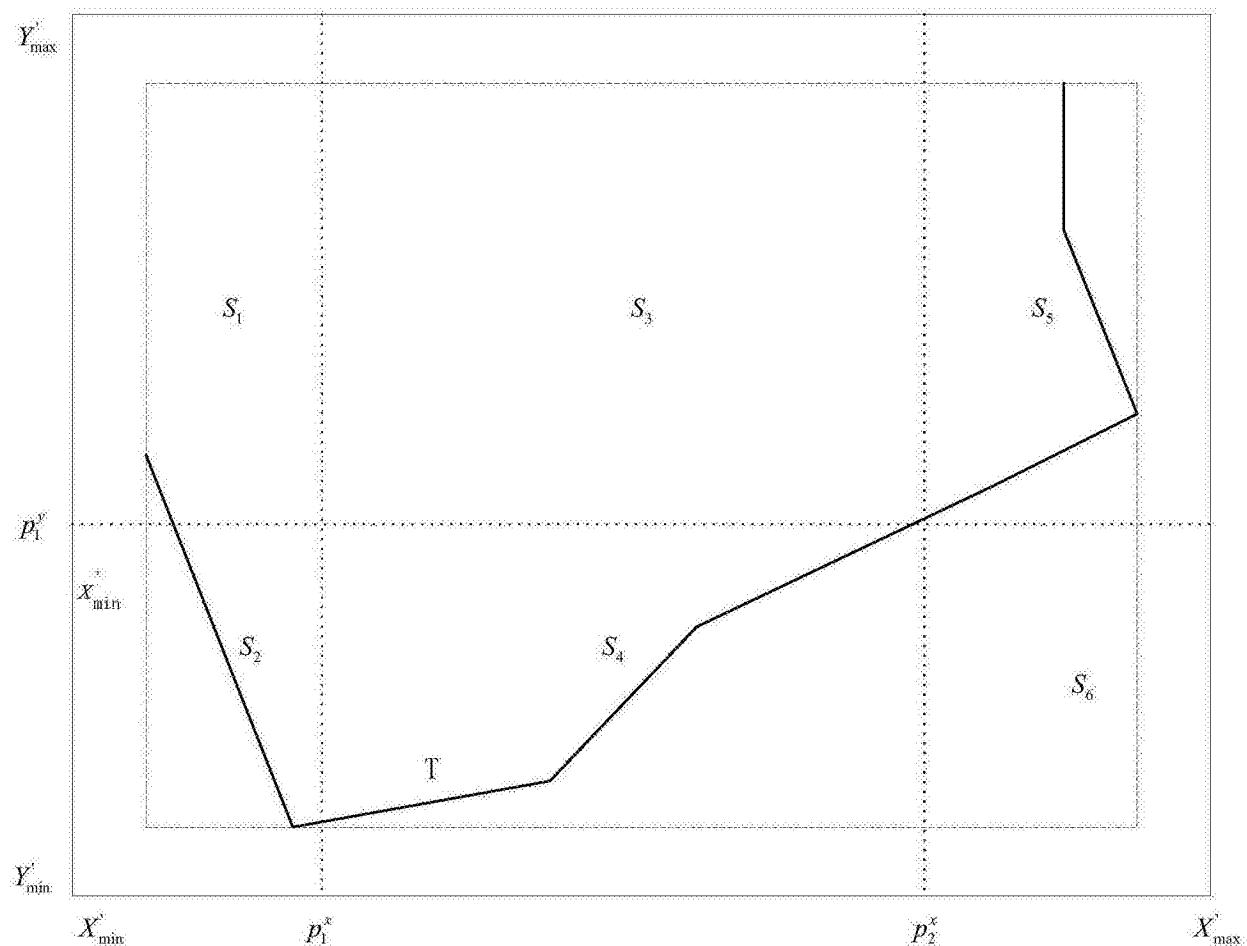


图2

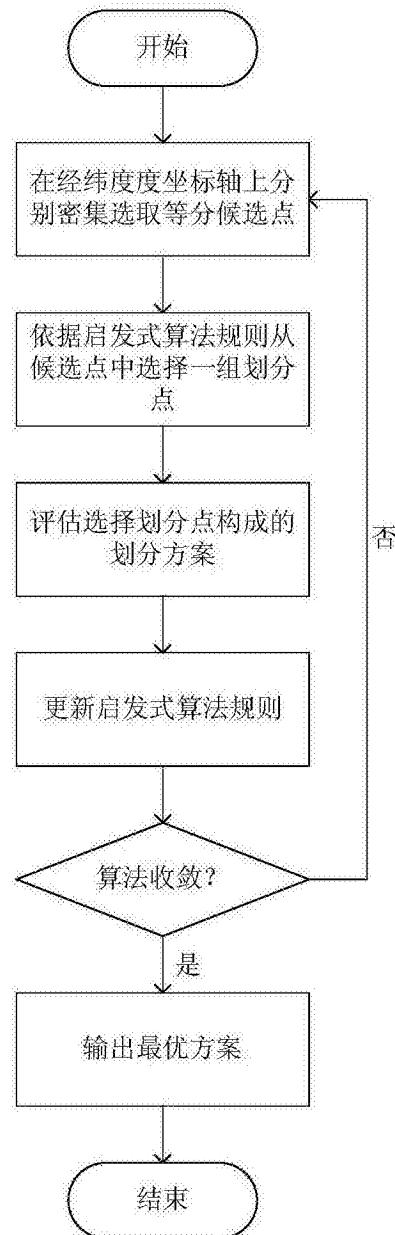


图3

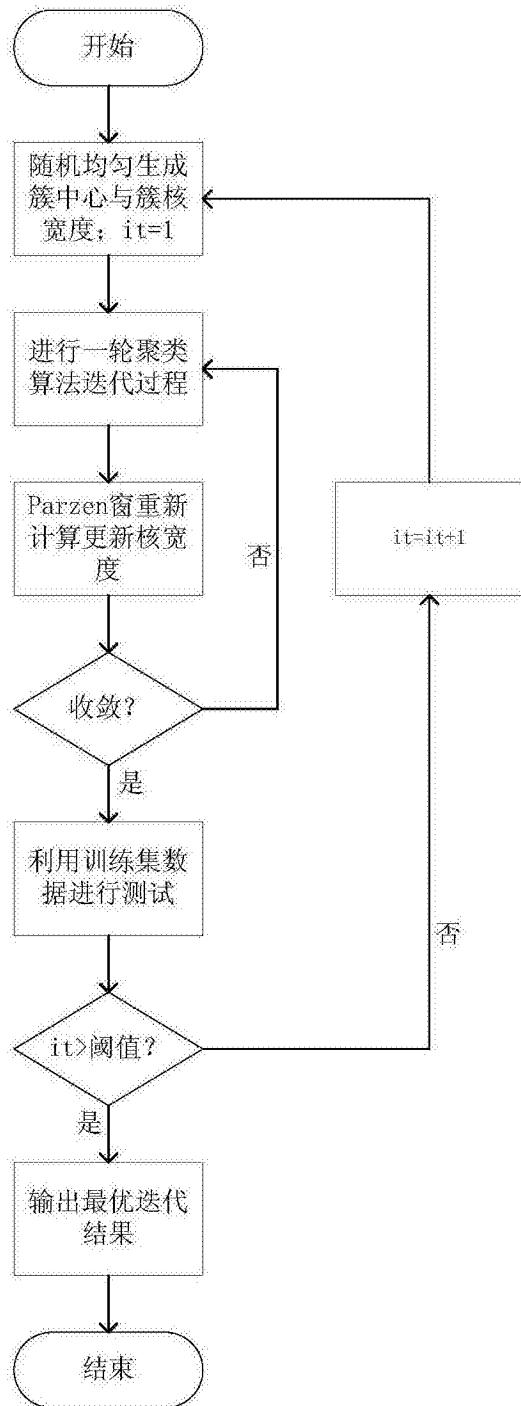


图4

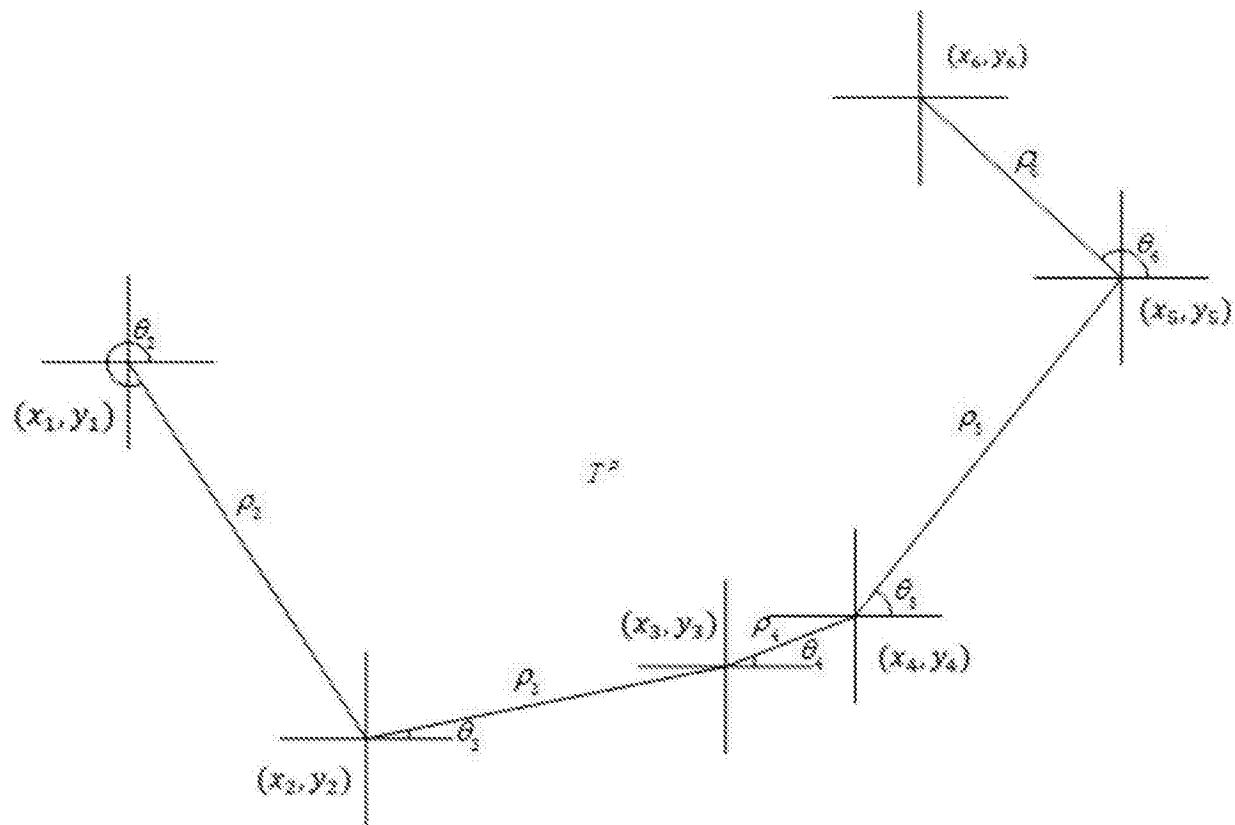


图5

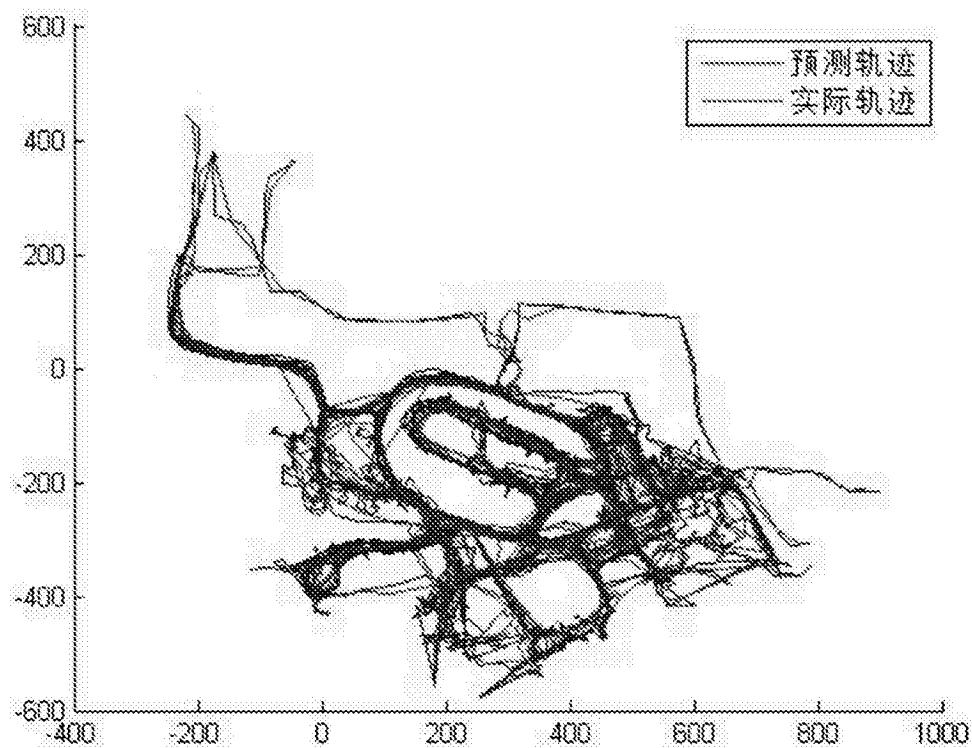


图6