



# 基于多重运动特征的轨迹相似性度量模型

朱进<sup>1</sup> 胡斌<sup>2,3</sup> 邵华<sup>4</sup>

1 苏州科技大学环境科学与工程学院,江苏 苏州,215009

2 南京师范大学虚拟地理环境教育部重点实验室,江苏 南京,210023

3 江苏省地理信息资源开发与利用协同创新中心,江苏 南京,210023

4 南京工业大学测绘学院,江苏 南京,210009

**摘要:**针对目前只能对单一运动特征(速度、方向等)进行轨迹相似性分析的不足,提出了基于多重运动特征的轨迹相似性度量,该度量对于分析和理解移动对象的运动行为和规律具有重要意义。将其应用于基于多重运动特征的运动序列模式发现。该相似性度量借鉴数据立方体的思想,将多重运动特征时间序列进行量化和符号化表示,在多重运动特征值域空间中计算两字符间的距离作为字符间替换代价,最终以加权编辑距离作为相似性度量。将该相似性度量与谱聚类方法相结合进行运动序列模式发现。实验以飓风数据为例,通过气象文献中飓风的发生与运动规律验证了模型的有效性。

**关键词:**运动特征;轨迹相似性;加权编辑距离;谱聚类;运动序列模式

中图法分类号:P208

文献标志码:A

移动对象的内在特征及其与外部环境的交互关系可通过其运动特征如速度、加速度、方向等反映。例如,大西洋飓风的速度从其发生到消亡一般经历由加速到减速的过程,飓风在登陆后速度通常减弱;低纬度飓风受东北信风的影响,一般向西行进,而高纬度飓风受副热带高压附近反气旋的影响,通常向西北和北行进<sup>[1]</sup>。通过移动对象轨迹计算出的运动特征可以揭示移动对象的内在特征,进而发现某些潜在知识。

当前,关于轨迹运动特征的研究主要集中在交通模式检测中<sup>[2-4]</sup>,而利用轨迹运动特征进行相似性分析的研究则相对分散。文献[5]构建了隐含个体多重运动特征的轨迹几何形态相似性度量。文献[6]基于欧氏距离,提出了多个运动特征间的轨迹相似性度量,但该方法要求轨迹具有相同的长度及采样间隔。文献[7]提出了轨迹运动特征相似性度量(normalized weighted edit distance, NWED),NWED 不要求轨迹具有相同的长度,但只能对单一运动特征进行相似性分析。而移动对象的多个运动特征有时会表现出一定的

相关性和依赖性,例如野生动物在速度较慢而曲率较大时可能在四处寻找食物,信鸽在即将飞回家的时候速度会突然增加且方向基本不变<sup>[8]</sup>。可见发现多个运动特征间的相关性对于分析和理解移动对象(如野生动物、飓风)的运动行为和特点具有重要意义。目前对不同长度的轨迹进行多重运动特征相似性分析的度量较为缺乏。

本文针对上述问题,提出多重运动特征编辑距离(multiple movement features edit distance, MMFED),将其作为基于多重运动特征的轨迹相似性度量,并利用MMFED结合谱聚类<sup>[9]</sup>发现多重运动特征序列模式。MMFED利用数据立方体<sup>[10]</sup>的思想,对多重运动特征进行量化,每一个立方体单元表示一定区间范围内的多重运动特征,并对其进行符号化表示,相似性度量利用加权编辑距离来表示,其中字符间的替换代价用多重运动特征值域空间中的欧氏距离来计算。本文利用飓风数据对比了MMFED与NWED,验证了MMFED的合理性,并通过飓风发生与运动的已知规律验证了本方法的有效性。

# 1 多重运动特征轨道相似性度量

MMFED是基于编辑距离构建的,编辑距离各个字符之间的替换代价通常为1,它利用数据立方体思想对多重运动特征建模,采用立方体中的欧氏距离作为字符间的替换代价,是一种加权编辑距离。

## 1.1 构建原理与流程

MMFED构建流程如图1所示,共分为3个步骤。

1)利用移动对象的轨迹(图1(a)),计算出移动对象每个采样点或相邻采样点间轨迹段的运动特征。轨迹可表示为点序列  $P_i \in \{P_1, P_2, \dots, P_n\}$ ,每个点  $P_i$  包含位置和时间数据,即  $P_i = (x_i, y_i, t_i)$ ,利用式(1)~(5),可以计算出速度  $v_i$ 、加速度  $a_i$  和曲率  $s_i$ ,其中  $d(P_i, P_{i+1})$  为  $P_i$  和  $P_{i+1}$  之间的欧氏距离,曲率  $s_i$  揭示的是路径的弯曲度。

$$d_i = d(P_i, P_{i+1}) \quad (1)$$

$$\Delta t_i = t_{i+1} - t_i \quad (2)$$

$$v_i = d_i / \Delta t_i \quad (3)$$

$$a_i = (v_{i+1} - v_i) / \Delta t_i \quad (4)$$

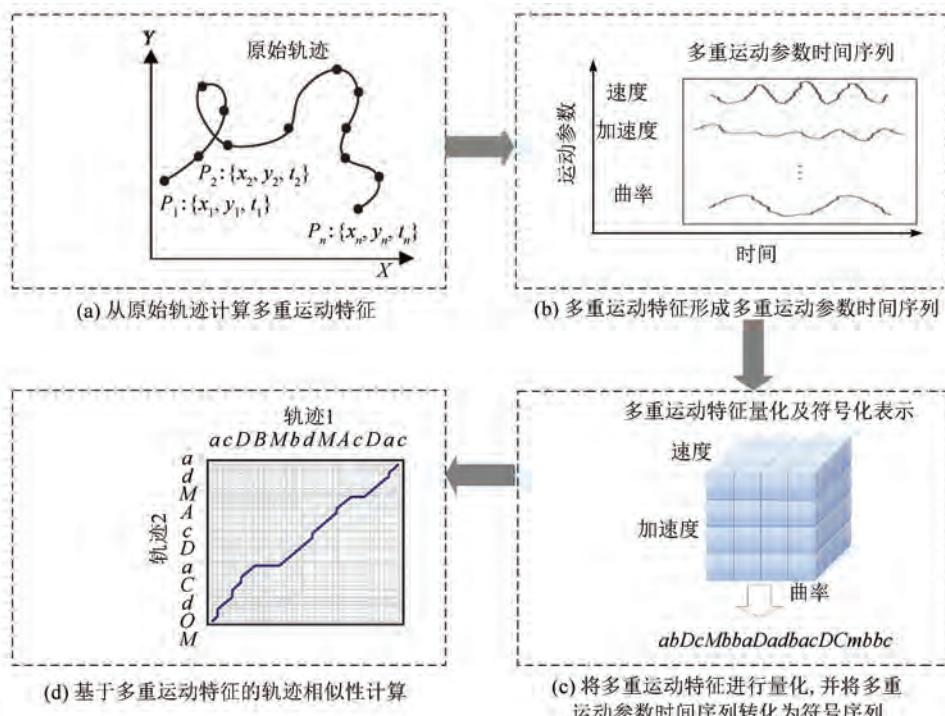


图1 基于多重运动特征的轨迹相似性度量

Fig.1 Trajectory Similarity Measure Based on Multiple Movement Features

## 1.2 运动特征量化及符号化表示

MMFED在NWED的基础之上,从单一运

$$s_i = (d(P_{i-1}, P_i) + d(P_i, P_{i+1})) / d(P_{i-1}, P_{i+1}) \quad (5)$$

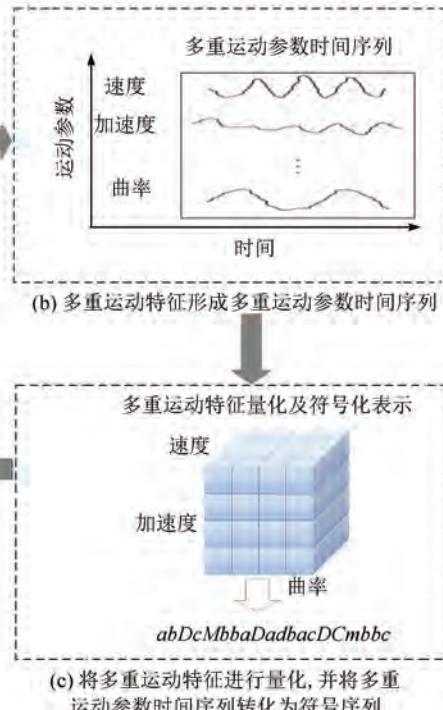
方向是连续采样点之间的移动方向,它用方向和基本方向(例如东)之间的夹角来表示;转角可通过计算连续方向之间的差异而得到,若两个连续方向为  $\alpha$  和  $\beta$ ,则转角可利用方向统计学<sup>[11]</sup>中两个角度之间的距离公式来计算:

$$t_a = 1 - \cos(\alpha - \beta) \quad (6)$$

提取运动特征后,可将轨迹表示为多重运动特征时间序列(图1(b)):  $(f_{t_1}^1, f_{t_1}^2, \dots, f_{t_1}^m), \dots, (f_{t_n}^1, f_{t_n}^2, \dots, f_{t_n}^m)$ 。其中,  $f$  表示运动特征;  $m$  ( $m \geq 2$ ) 为多重运动特征的个数; 元组  $(f_{t_i}^1, f_{t_i}^2, \dots, f_{t_i}^m)$  表示时刻  $t_i$  时的多重运动特征。

2)将多重运动特征表示为数据立方体,其中数据立方体的每个维度对应多重运动特征的一个特征,这样每个立方体单元对应一定区间范围内的多重运动特征,若每个立方体单元用不同的字符来表示,则立方体单元(字符)之间的欧氏距离代表编辑距离中替换字符的代价(图1(c))。

3)对符号化的轨迹多重运动特征序列,基于加权编辑距离计算轨迹相似性度量(图1(d))。  
§ 1.2 和 § 1.3 分别阐述步骤2)和步骤3)。



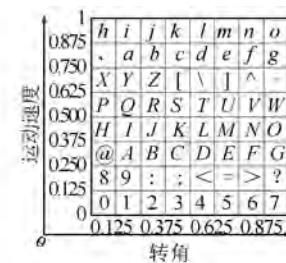
动特征扩展到多重运动特征。NWED方法利用改进的 Levenshtein 编辑距离<sup>[12]</sup>作为相似性度

量。Levenshtein 编辑距离是将一个字符串转变为另一个字符串所需的最少操作数目, 操作包括插入、删除和替换, 每种操作代价都为 1。NWED 利用加权编辑距离来计算, 加权编辑距离对不同的操作其代价也不同, 例如应用于拼写检查时, 对容易打错的字符对(如 *a* 容易打成 *e*), 加权编辑距离对它们的替换代价赋予更大的权重。类似于 Levenshtein 距离, NWED 对于插入和删除的操作代价也为 1, 但替换代价根据不同运动参数类别的相似程度赋予不同的权重, 从 0(类别相同没有代价)变化到 1(最大代价)。NWED 考虑了运动参数时间序列的振幅和频率, 将运动参数分为 9 个类别, 每个类别对应一个字符, 替换代价表示为一个  $9 \times 9$  的代价矩阵。为控制字符数目, MMFED 不考虑运动参数时间序列的频率, 仅考虑振幅, 将每个运动参数时间序列按振幅划分为 8 个等级, 每个时刻的多重运动特征对应多重运动特征空间中的一个字符, 字符的插入和删除代价为 1, 替换代价为字符在多重运动特征值域空间中的欧氏距离。

MMFED 先将多重运动特征序列转换为字符串序列, 即将  $(f_{t_1}^1, f_{t_1}^2, \dots, f_{t_1}^m), \dots, (f_{t_n}^1, f_{t_n}^2, \dots, f_{t_n}^m)$  转换为序列  $(c_{t_1}, c_{t_2}, \dots, c_{t_n})$ , 其中  $c_{t_i}$  为时刻  $t_i$  时多重运动特征  $(f_{t_1}^1, f_{t_1}^2, \dots, f_{t_1}^m)$  对应的字符。此处将运动特征分为方向特征以及其他特征(速度、加速度、曲率、转角)。对其他特征, 将特征值归一化到区间  $[0, 1]$ ; 对方向特征, 不作归一化处理。之所以将特征分为方向特征以及其他特征, 是由于计算方向之间的距离需利用式(6), 而不能像一般数值特征那样处理。对相应的特征进行归一化处理后, 借鉴数据立方体的思想, 将每个特征的值域划分为 8 个等长的区间, 这样, 多重运动特征值域空间形成  $8^m$  个立方体单元, 每个立方体对应一个字符。

以最简单的双重运动特征为例说明多重运动特征的量化及符号化过程。如图 2 所示, 对于不含方向的双重运动特征, 将区间  $[0, 1]$  划分为 8 个等长的区间:  $[0, 0.125], [0.125, 0.25], \dots, [0.875, 1]$ 。图 2(a) 中, 速度和转角双重运动特征值域空间形成  $8 \times 8$  的网格, 每个网格单元对应一定区间范围内的速度和转角, 也对应一个字符, 图 2 中用 ASCII 码从 0 到 *o* 之间连续的 64 个字符表示每个网格单元。例如速度、转角对  $(0.64, 0.33)$  对应字符 *E*。图 2(b) 中, 将区间  $[0, 2\pi]$  划分为 8 个等长的区间:  $[0, \pi/4], [\pi/4, \pi/2], \dots, [7\pi/4, 2\pi]$ 。例如速度、方向对  $(0.68, 3\pi/5)$  对应

字符 *Z*。三重运动特征依次类推, 三重运动特征值域空间形成  $8 \times 8 \times 8(512)$  的立方体, 如图 3 所示(图中省略符号)。



(a) 速度-转角



(b) 速度-运动

图 2 双重运动特征的量化及符号化

Fig.2 The Quantification and Symbolization of Duplex Movement Features

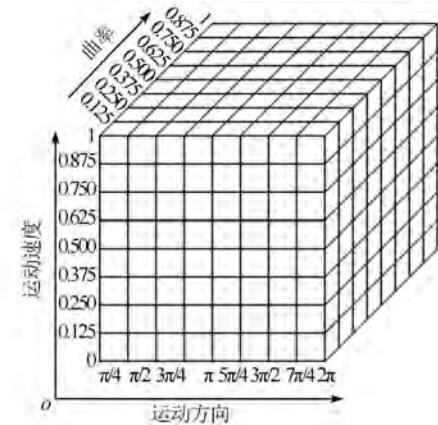


图 3 三重运动特征速度-方向-曲率的量化及符号化

Fig.3 The Quantification and Symbolization of Triple Movement Features Velocity-Direction-Sinuosity

量化、符号化多重运动特征之后, 字符在运动特征值域空间(立方体)中的距离即为编辑距离计算时字符替换的代价。若  $T_i$  为序列  $T$  的第  $i$  个字符,  $P_j$  为序列  $P$  的第  $j$  个字符, 将  $T_i$  替换为  $P_j$  (或  $P_j$  替换为  $T_i$ ) 的代价计算过程如下。首先计算  $T_i$  在立方体中每个维度中的序号  $d_{T_i}^1, d_{T_i}^2, \dots, d_{T_i}^m$  (在双重运动特征值域空间中即为行、列号, 为简洁起见, 下文记为  $d_T^1, d_T^2, \dots, d_T^m$ )。类似地, 得到  $P_j$  在立方体中每个维度中的序号  $d_P^1, d_P^2, \dots, d_P^m$ 。每个维度上的 8 个值域区间用序号 0~7 表示, 即  $[0, 0.125]$  对应 0,  $[0.875, 1]$

对应7。字符对应的每个维度中的序号反映字符在立方体中的位置,可用于计算字符间的距离。维度*i*上序号间的差值 $|d_T^i - d_P^i|$ 可作为该维度上的距离。由于每个维度上的区间用0~7表示,因此序号间差值的最大值为7,为使每个维度上的距离变换到区间[0, 1]之间,以 $|d_T^i - d_P^i|/7$ 作为维度*i*上的距离。在立方体中,利用欧氏距离来计算两个字符间的距离。由于每个维度上的最大距离为1, *m*个维度最大的欧氏距离则为 $\sqrt{m}$ ,将欧氏距离再除以 $\sqrt{m}$ ,使其变换为区间[0, 1]之间,因此,字符间距离*d'*的计算公式为:

$$d' = \sqrt{\frac{1}{m} \sum_{i=1}^m \left( \frac{d_T^i - d_P^i}{7} \right)^2} \quad (7)$$

$$d' = \sqrt{\frac{1}{m} \left( \sum_{i=1}^m \left( \frac{d_T^i - d_P^i}{7} \right)^2 + \frac{1}{4} \left( 1 - \cos\left(\frac{\pi}{4}(d_T^m - d_P^m)\right) \right)^2 \right)} \quad (8)$$

总之,若多重特征中不含方向特征,利用式(7)计算字符间的替换代价,否则利用式(8)计算。

### 1.3 多重运动特征相似性度量构建

两字符序列*T*和*P*(长度分别为*n*和*m*)的MMFED利用动态规划思想<sup>[13]</sup>计算,主要利用式(10),逐步填满一个*n*×*m*的相似性矩阵*M*(10),*sub\_cost(T<sub>i</sub>, P<sub>j</sub>)*为两个字符间的替换代价,min函数为取3个元素中的最小值。元素*M(n, m)*是加权编辑距离的结果,最后为了消除不同长度的轨迹对结果的影响,对加权编辑距离使用式(11)标准化,文献[14]证明了最终结果MMFED是度量距离,可用索引等技术加速计算。

$$M_{T,P} = C_{0\dots n, 0\dots m} \quad (9)$$

$$C_{i,j} =$$

$$\begin{cases} j, & \text{当 } i = 0 \\ i, & \text{当 } j = 0 \\ C_{i-1, j-1}, & \text{当 } i, j > 0 \text{ 且 } T_i = P_j \\ sub\_cost(T_i, P_j) + min(C_{i-1, j-1}, C_{i-1, j}, C_{i, j-1}), & \text{其他} \end{cases} \quad (10)$$

$$MMFED_{T,P} = \frac{2 \times C_{n,m}}{n+m+C_{n,m}} \quad (11)$$

### 1.4 相似性度量应用

多重运动特征时间序列反映了轨迹运动特征的演变趋势,MMFED反映了运动特征演变趋势的相似性,可将MMFED与聚类方法结合来发现具有相似运动特征演变趋势的轨迹簇,即相似运动特征序列模式的轨迹。

理论上可以使用任何标准的聚类算法与MMFED相结合对多重运动特征时间序列进行聚类。但传统的聚类方法例如*k*-means等,通常

式(7)即为*T<sub>i</sub>*、*P<sub>j</sub>*间的替换代价*sub\_cost(T<sub>i</sub>, P<sub>j</sub>)*。

式(7)是不含方向特征的字符间替换代价,对于方向特征,两个方向 $\alpha$ 和 $\beta$ 间的距离可用式(6)计算。为表述方便,假设方向特征在第*m*个维度上,则在方向维度上,维度*d<sub>T</sub><sup>m</sup>*和*d<sub>P</sub><sup>m</sup>*间的距离为 $1 - \cos\left(\frac{\pi}{4}(d_T^m - d_P^m)\right)$ (每个区间的大小为 $\pi/4$ ),由于该距离在区间[0, 2]之间,为将其变换到区间[0, 1]之间,将该距离再除以2。因此,含有方向特征的字符间距离(替换代价)的计算公式为:

假设样本空间为凸球状。当样本空间形状不为凸时,算法容易陷入局部最优。近年来,在机器学习领域出现了一种新的聚类分析算法,即建立在谱图划分理论基础上的谱聚类算法(spectral clustering)<sup>[15]</sup>。与传统聚类方法相比,谱聚类算法可以识别非凸球状的聚簇,避免高维特征向量的奇异性问题,实现起来方便、简单,性能较高。因此,本研究采用经典的自调整谱聚类算法(self-tuning spectral clustering)<sup>[16]</sup>作为谱聚类的实现。

## 2 实验评估与验证

实验运行环境为CPU Intel® Core™ i5-2400 3.10 GHz 4核,内存4 GB,操作系统Windows® 8.1 64位,开发环境微软Visual C++ 2010。实验采用1900~2014年的北大西洋飓风数据集为实验数据(<http://www.nhc.noaa.gov/data/#hurdat>),数据集共有1 230条飓风轨迹,相邻采样点时间间隔为6 h。实验进行双重、三重运动特征序列模式发现,并通过已知的飓风规律来验证本方法的有效性。

### 2.1 双重运动特征序列模式发现

为更好地对飓风行为进行预测,将飓风分为热带飓风与非热带飓风两类<sup>[1]</sup>。热带飓风在低纬度地区形成,一般向西行进,转弯后向北或者东北方向前进;而非热带飓风在高纬地区形成,一般向东北方向行进。区分这两类飓风的重要因素是飓风形成的时间和纬度。

实验对速度-方向双重运动特征利用谱聚类进行聚类分析,为了检测聚类算法是否能将飓风

分为热带飓风与非热带飓风, 在谱聚类的  $k$ -means 聚类阶段, 设置聚簇数目为 2, 得到如图 4 所示的聚簇 1 和聚簇 2。从图 4 中可以看到, 聚簇 1 中轨迹的起源点大多低于  $20^{\circ}\text{N}$ , 且行进方向

大多先向西前进, 再转向北或东北方向前进; 聚簇 2 中轨迹的起源点大多高于  $20^{\circ}\text{N}$ , 且行进方向大多直接向东北方向前进。

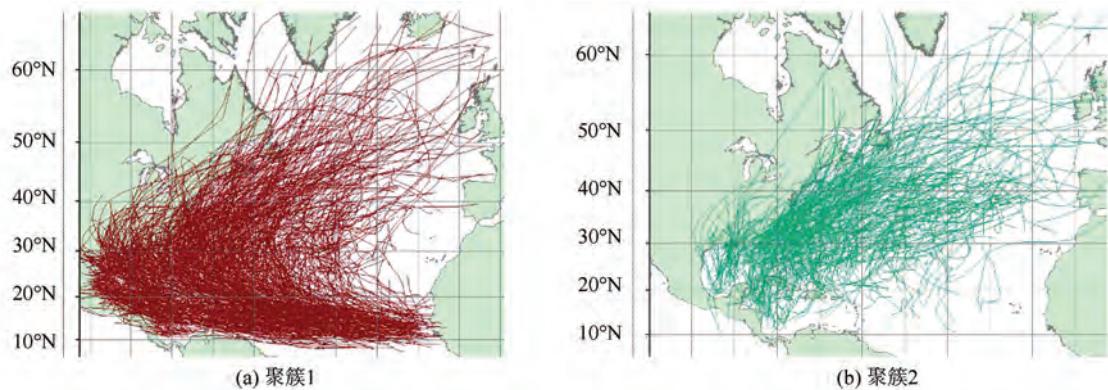


图 4 速度-方向运动特征聚类分析的聚簇

Fig.4 The Clusters of Velocity-Direction Clustering Analysis

图 5(a)和 5(b)分别为聚簇 1 和聚簇 2 的速度随采样点序号(即时间顺序)演变的模式。由于本方法将每个维度的运动特征划分为 8 个等级, 图 5 中对 8 个等级分别赋予渐进变化的不同颜色。图 5(a)、5(b)中分别显示了每个聚簇 300 条轨迹的速度模式。可以看出, 两类聚簇的速度模式区别较大, 聚簇 1 在轨迹前半阶段速度较大, 有较多绿色; 而聚簇 2 前半阶段速度较小, 大多为褐色。

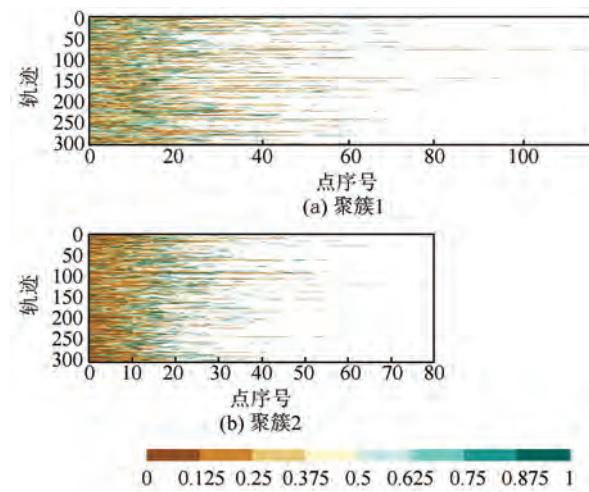


图 5 聚簇的速度特征序列模式

Fig.5 The Velocity Sequential Pattern of Clusters

图 6(a)和图 6(b)分别是聚簇 1 和聚簇 2 的方向随采样点序号演变的模式。类似于图 5 中速度特征, 对方向特征的 8 个等级, 图 6 中也对这 8 个等级分别赋予渐进变化的不同颜色, 同样显示了每个聚簇 300 条轨迹的方向变化模式。可以看出, 两类聚簇的方向模式区别十分明显, 聚簇 1 在

轨迹前半阶段方向大多集中在区间  $[3\pi/4, \pi]$ , 而聚簇 2 前半阶段方向大多集中在区间  $[0, \pi/2]$ , 这与图 4 显示的情况是一致的。

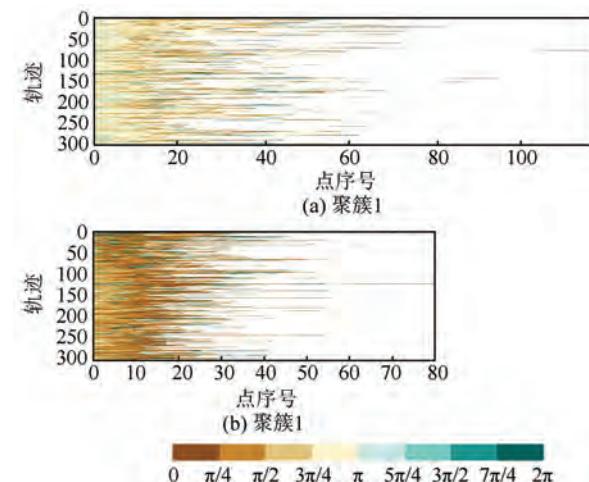


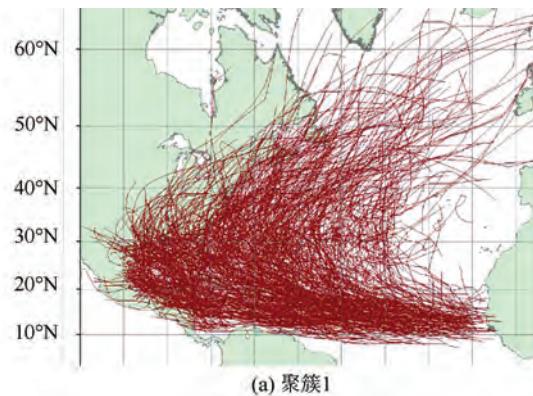
图 6 聚簇的方向特征序列模式

Fig.6 The Direction Sequential Pattern of Clusters

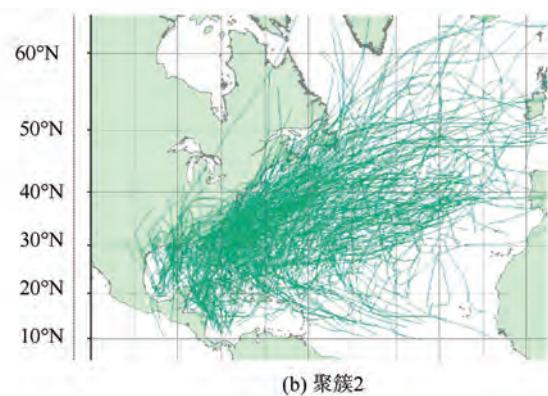
## 2.2 三重运动特征序列模式发现

在 § 2.1 对速度-方向特征进行分析的基础之上, 实验对速度-加速度-方向三重运动特征进行运动模式发现, 使用谱聚类进行聚类分析, 设置聚簇数目为 2, 得到如图 7 所示的聚簇 1 和聚簇 2。从图 7 可以看到, 聚簇 1 中轨迹的起源点大多低于  $20^{\circ}\text{N}$ , 行进方向大多先向西前进, 再转向北或东北方向前进; 聚簇 2 中轨迹的起源点大多高于  $20^{\circ}\text{N}$ , 且行进方向大多直接向东北方向前进。

图 8(a)和图 8(b)分别是聚簇 1 和聚簇 2 各 300 条轨迹的速度随采样点序号演变的模式。从图 8 中可以看到, 两类聚簇的速度模式区别较大,



(a) 聚簇1



(b) 聚簇2

图7 速度-加速度-方向运动特征聚类分析的聚簇

Fig.7 The Clusters of Velocity-Acceleration-Direction Clustering Analysis

聚簇1在轨迹前半阶段速度较大,有较多绿色;而聚簇2前半阶段速度较小,大多为褐色。但图8聚簇之间的区别不如图5聚簇之间的区别明显,这是由于该相似性度量还包含了加速度的缘故。

图9(a)和图9(b)分别是聚簇1和聚簇2各300条轨迹的加速度随采样点序号演变的模式。从图9中可以看到,两类聚簇的加速度模式区别不明显,聚簇2的加速度似乎比聚簇1的加速度稍大,绿色稍多。

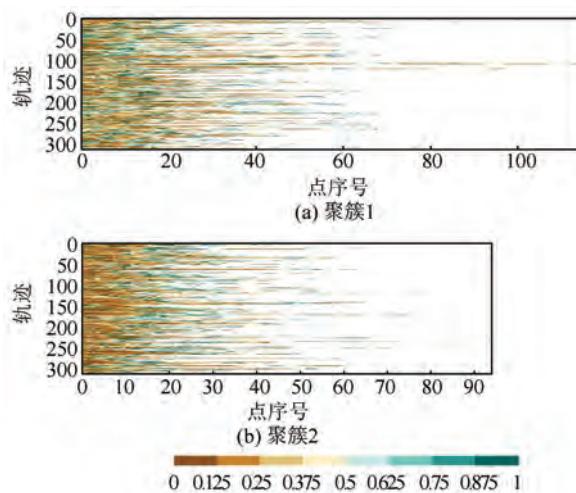


图8 聚簇的速度特征序列模式

Fig.8 The Velocity Sequential Pattern of Clusters

聚簇1和聚簇2的方向特征序列模式类似§2.1的速度-方向双重特征序列模式,这里不再赘述。

### 2.3 运动序列模式发现效果分析

为了验证运动序列模式发现的效果,通过实验检验发现的具有相似运动模式的聚簇是否与气象文献中飓风发生与运动的规律一致。飓风的季节和起源点的纬度是影响飓风动态行为的两个重要因素<sup>[1]</sup>,起源点在19°N以南为低纬,以北为高纬;且起源于相似季节(夏季、秋季)并具有相似运

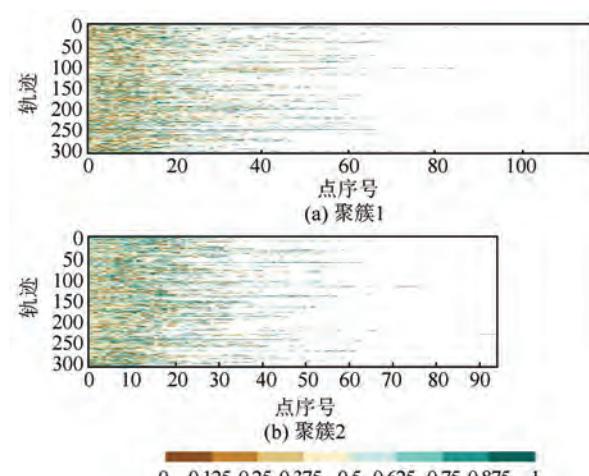


图9 聚簇的加速度特征序列模式

Fig.9 The Acceleration Sequential Pattern of Clusters

动特点的飓风起源点在空间上通常接近。

实验利用曼-惠特尼U检验(Mann-Whitney U test,又称曼-惠特尼秩和检验)来检验聚簇是否发源于不同的位置以及不同的月份。曼-惠特尼U检验是一种非参数检验方法,它假设两样本分别来自除了总体均值以外完全相同的两个总体,其目的是检验两个总体的均值是否有显著差异。实验对双重运动特征(速度-v-d、速度-v-t<sub>a</sub>、速度-v-s)聚类,三重运动特征(速度-加速度-方向v-a-d、速度-转角-方向v-t<sub>a</sub>-d)聚类,并与单一运动特征(速度v、方向d、转角t<sub>a</sub>)即NWED聚类进行对比。检验的属性包括起源点经度、起源点纬度、低纬或高纬(19°N以南为低纬,以北为高纬)、起源点在80°W东侧或西侧(衡量与美国海岸线的距离)、季节(5~8月为夏季,9~12月为秋季)以及月份。实验结果见表1。

从表1中可以看出,双重和三重运动特征聚类产生的两类聚簇在飓风起源点来自于低纬或高纬地区这一属性上是统计显著的( $p = 0.000$ )。

§ 2.1 和 § 2.2 的轨迹空间形态也显示了这一点。聚簇 1 大部分来源于 19°N 以南, 聚簇 2 大部分来源于 19°N 以北。而在利用 NWED 相似性度量得到的聚簇中, 只有利用方向特征得到的聚簇在低纬和高纬上统计显著( $p = 0.000$ ), 利用速度( $p = 0.310$ )和转角( $p = 0.576$ )特征得到的聚簇在低纬和高纬上统计不显著。类似地, 双重和三重运动

特征聚类产生的两类聚簇在飓风起源点来自于 80°W 东侧或西侧这一属性上是统计显著的( $p = 0.000$ ), 而在利用 NWED 相似性度量得到的聚簇中, 只有利用方向特征得到的聚簇统计显著( $p = 0.005$ ), 利用速度和转角特征得到的聚簇统计不显著( $p$  分别为 0.557 和 0.834)。

表 1 利用曼-惠特尼 U 检验得到的  $p$  值

Tab.1 The  $p$  Value of Mann-Whitney U Test

运动特征	特征	起源点	起源点	低纬 /	起源点在	季节	时间粒度单位
		纬度	经度	高纬	80°W 东或西	(夏/秋)	
双重	$v-d$	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.008</b>	<b>0.006</b>
	$v-t_a$	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	0.311	0.468
	$v-s$	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.014</b>	0.059
三重	$v-a-d$	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.027</b>	<b>0.008</b>
	$v-t_a-d$	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.001</b>	<b>0.003</b>
单一	$v$	<b>0.016</b>	0.611	0.310	0.557	0.094	0.124
	$d$	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.005</b>	0.354	0.126
	$t_a$	0.545	0.101	0.576	0.834	0.680	0.500

注: $p$  值小于 0.05 用粗体表示

从表 1 中还可以看出, 双重和三重运动特征聚类产生的两类聚簇在季节和月份属性上同样是统计显著的, 除了双重的速度-转角特征( $p$  分别为 0.311 和 0.468)以及速度-曲率特征( $p$  在月份上为 0.059), 其他包含速度和方向的双重、三重特征  $p$  值都小于 0.05。相反, 单一运动特征聚类产生的两类聚簇在季节和月份属性上统计不显著( $p$  值都大于 0.05)。以上结果说明, 利用多重运动特征, 有时能更好地发现或者揭示移动对象运动的规律和行为。

### 3 结语

本文提出了基于多重运动特征的轨迹相似性度量 MMFED, 并结合谱聚类将其用于运动序列模式发现中。MMFED 扩展了 NWED 相似性度量, 利用数据立方体的思想对多重运动特征进行量化及符号化表示, 对符号化后的运动特征序列, 利用加权编辑距离计算相似性度量, 其中字符替换代价使用多重运动特征域空间中的欧氏距离来计算。实验利用大西洋飓风数据集验证了 MMFED 和运动模式发现方法的有效性。

需要指出的是, 虽然本方法可以对双重以上的轨迹运动特征进行相似性分析, 但是如果运动特征数目过高, 相似性度量中包含的语义信息会过于含糊和复杂, 导致相似性分析和聚类分析的结果比较复杂。因此, 应根据问题领域和需求, 选

择适当数目(尽量不要超过三重)的多重运动特征进行分析。在实际应用中, 可以结合领域知识选择适当数目的多重运动特征进行探索性的相似性和聚类分析, 分析多重运动特征之间的相关性和依赖性, 以及不同运动特征对聚类结果的影响。

### 参 考 文 献

- [1] Elsner J B, Kara A B. Hurricanes of the North Atlantic: Climate and Society [M]. New York: Oxford University Press, 1999: 21-24
- [2] Zheng Y, Liu L, Wang L, et al. Learning Transportation Mode from Raw GPS Data for Geographic Applications on the Web [C]. The 17th International Conference on World Wide Web (WWW'08), Beijing, China, 2008
- [3] Zheng Y, Li Q, Chen Y, et al. Understanding Mobility Based on GPS Data [C]. The 10th International Conference on Ubiquitous Computing (UbiComp'08), Seoul, Korea, 2008
- [4] Zhang Zhihua. Deriving Trip Information from GPS Trajectories[D]. Shanghai: East China Normal University, 2010(张治华. 基于 GPS 轨迹的出行信息提取研究[D]. 上海:华东师范大学, 2010)
- [5] Chen J, Shaw S L, Yu H, et al. Exploratory Data Analysis of Activity Diary Data a Space - Time GIS Approach[J]. Journal of Transport Geography, 2011, 19(3): 394-404
- [6] Dodge S, Weibel R, Laube P. Trajectory Similarity Analysis in Movement Parameter Space [C]. GIS-

- RUK, UK, 2011
- [7] Dodge S, Laube P, Weibel R. Movement Similarity Assessment Using Symbolic Representation of Trajectories[J]. *International Journal of Geographical Information Science*, 2012, 26(9): 1 563-1 588
- [8] Laube P, Dennis T, Forer P, et al. Movement Beyond the Snapshot -Dynamic Analysis of Geospatial Lifelines[J]. *Computers, Environment and Urban Systems*, 2007, 31(5): 481-501
- [9] Li Jingwei. Adaptive Spectral Clustering Based on Shared Nearest Neighbors[D]. Dalian: Dalian University of Technology, 2010(李静伟. 基于共享近邻的自适应谱聚类算法[D]. 大连:大连理工大学, 2010)
- [10] Han J, Kamber M, Pei J. Data Mining Concepts and Techniques [M]. 3rd Edition. Waltham: Elsevier, 2012
- [11] Mardia K V, Jupp P E. Directional Statistics [M]. Chichester UK: John Wiley & Sons, 2000:13-23
- [12] Levenshtein V I. Binary Codes Capable of Correcting Deletions, Insertions, and Reversals[J]. *Soviet Physics Doklady*, 1966, 10(8): 707-710
- [13] Cormen T H, Leiserson C E, Rivest R L, et al. Introduction to Algorithms [M]. 3rd Edition. Cambridge: MIT Press, 2009
- [14] Li Y, Liu B. A Normalized Levenshtein Distance Metric[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(6): 1 091-1 095
- [15] Cai Xiaoyan, Dai Guanzhong, Yang Libin. Survey on Spectral Clustering Algorithms[J]. *Computer Science*, 2008, 35(7):14-18(蔡晓妍, 戴冠中, 杨黎斌. 谱聚类算法综述[J]. 计算机科学, 2008, 35(7):14-18)
- [16] Zelnik-Manor L, Perona P. Self-Tuning Spectral Clustering [J]. *Advances in Neural Information Processing Systems*, 2004: 1 601-1 608

## Trajectory Similarity Measure Based on Multiple Movement Features

ZHU Jin<sup>1</sup> HU Bin<sup>2,3</sup> SHAO Hua<sup>4</sup>

1 School of Environmental Science and Engineering, Suzhou University of Science and Technology, Suzhou 215009, China

2 Key Laboratory for Virtual Geographic Environment, Ministry of Education, Nanjing Normal University, Nanjing 210023, China

3 Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

4 Department of Geomatics Engineering, Nanjing University of Technology, Nanjing 210009, China

**Abstract:** For the shortcoming that existing methods can only measure the trajectory similarity of single movement feature (e.g. velocity, acceleration), the trajectory similarity measure based on multiple movement features is proposed. The measure is significant for analyzing and understanding the movement behaviors and mechanisms of moving objects. The measure borrows the idea of data cube, quantizes and symbolizes the multiple movement feature time series. In multiple movement feature domain space, the Euclidean distances between characters are computed as the substitution costs of weighted edit distance which is computed as the similarity measure. The measure is integrated with the spectral clustering method for movement sequential pattern discovery. Using the hurricane dataset, the known hurricane originating and movement laws in meteorological literatures verify the effectiveness of the measure.

**Key words:** movement features; trajectory similarity measure; weighted edit distance; spectral clustering; movement sequential pattern

**First author:** ZHU Jin, PhD, lecturer, specializes in trajectory data mining. E-mail: 540896749@qq.com

**Corresponding author:** HU Bin, PhD, associate professor. E-mail: hb\_hubin@126.com

**Foundation support:** The National Natural Science Foundation of China, Nos. 41571389, 41501431; Teacher Training Research Funding Project of Suzhou University of Science and Technology, No.331511203; Youth Foundation Project of Suzhou University of Science and Technology, No.341731204.