

# 基于相似阈值学习的台风路径分类方法

赵丹枫<sup>1</sup>, 许强<sup>1</sup>, 黄洲<sup>1</sup>, 黄冬梅<sup>1</sup>, 苏诚<sup>2</sup>

(1.上海海洋大学信息学院,上海 201306; 2.自然资源部东海局预报中心,上海 200136)

**摘要:**台风路径分类是分析台风特征的重要途径,同时也是判断台风影响区域和范围的重要方法。目前台风的分类研究主要采用主观识别和  $K$ -means 聚类等方法,这些方法存在传统聚类算法产生的随机性问题和依赖人工经验设置模型参数的智能化问题,以及针对特定台风研究而导致的局限性问题。针对以上问题,本文提出一种基于相似阈值学习的台风路径分类方法,首先,提出基于密度质心的混合聚类算法,实现了热点区域挖掘,减少了传统方法的随机性;其次,提出基于 fastDTW 的自适应模体处理算法,实现了更为智能化的路径分类;最后,以登陆湛江的台风路径为例,展示并验证了混合聚类方法的分类效果。

**关键词:**台风灾害;相似路径;路径分类;模体挖掘

中图分类号:X16 文献标识码:A 文章编号:1007-6336(2022)04-0644-09

## A typhoon tracks classification method based on similarity threshold learning

ZHAO Dan-feng<sup>1</sup>, XU Qiang<sup>1</sup>, HUANG Zhou<sup>1</sup>, HUANG Dong-mei<sup>1</sup>, SU Cheng<sup>2</sup>

(1.Information College, Shanghai Ocean University, Shanghai 201306, China; 2.Forecasting Center, East China Sea Bureau of Ministry of Natural Resources, Shanghai 200136, China)

**Abstract:** The classification of typhoon tracks is an important way to analyze the characteristics of typhoons, and it is also an important method to judge the areas and ranges affected by typhoons. The main methods such as subjective identification and  $K$ -means clustering are applied to the research of typhoon classification. These methods have some problems such as randomness caused by traditional clustering algorithms, also rely on artificial experience to set model parameters and are not universal due to specific typhoon research. Aiming at the above problems, this paper proposes a typhoon tracks classification method based on similarity threshold learning. Firstly, a hybrid clustering algorithm based on density center of mass is proposed to realize hot-spot mining and reduce the randomness of traditional methods. Secondly, an adaptive motif processing algorithm based on fastDTW algorithms is proposed to realize more intelligent tracks classification. Finally, the typhoon track landing in Zhanjiang is taken as an example to demonstrate and verify the classification effect of this method.

**Key words:** typhoon disasters; similar tracks; classification of tracks; motif discovery

中国位于西北太平洋和南海沿岸,海岸线长约 1.8 万公里,地理位置十分特殊。据统计,平均每年约 20 个台风(包括热带低压)可以影响中

国沿海,这些台风带来的影响给社会经济造成了极大的损失<sup>[1]</sup>。为减少台风的负面影响,包括中国在内的世界各国都加大了对台风的监测与研

收稿日期:2020-03-22, 修订日期:2020-07-20

基金项目:国家自然科学基金项目(42106190);上海市科委地方能力建设项目(20050501900)

作者简介:赵丹枫(1982-),女,黑龙江哈尔滨人,讲师,主要研究方向为海洋大数据技术, E-mail: dfzhao@shou.edu.cn

通讯作者:黄冬梅,研究方向海洋大数据, E-mail: dmhuang@shou.edu.cn

究,其主要热点集中于路径预报、强度预报、风雨预报以及风暴潮预报<sup>[2]</sup>。

在台风路径的预报中,如何分类引起学者的广泛关注<sup>[3-4]</sup>,即根据台风路径的空间特征<sup>[5]</sup>(路径形状、位置等)进行分类,并对分类处理后的台风路径在生命周期、能量、活动季节和登陆概率等方面进行分析。因此,路径分类问题对获取台风特征和预测台风的登陆区域具有重要意义。

传统的台风路径分类,主要采取主观识别法<sup>[2]</sup>和客观分析法<sup>[6-8]</sup>。主观识别法主要依靠人工经验,需要研究者有大量的实践经历,但容易产生对专家经验的依赖,并且分类不够准确、稳定;客观分析法中,基于 *K-means* 聚类算法<sup>[9]</sup>的台风路径分类方法具有简单、通用、计算复杂度低、对数据的输入顺序不敏感等优势,受到广泛关注,但也面临算法本身的随机性和不能很好地实现智能化的问题。

Elsner 通过 *K-means* 聚类算法,仅利用强度最大时的经纬度信息建立参数<sup>[6]</sup>,将北大西洋台风简单分类为直行路径与北折弯曲路径,此方法没有充分利用各个时次的位置信息。Nakamura 在 Elsner 基础上,进一步以每个定位时次的风速为权重,并将各台风路径的质心和方差作为路径形状和位置特征量,最后将台风分为六类<sup>[7]</sup>。Camargo 等<sup>[8]</sup>指出,传统 *K-means* 聚类算法对台风时间长度的敏感度较低,无法区分不同时间长度的路径,故提出有限混合模型 (finite mixture model, FMM) 聚类算法。有限混合法将台风路径上各点的经度、纬度时间序列作为独立变量,通过二次多项式回归函数对西北太平洋台风进行分类,因为计算过程复杂,不能将路径间的相似度数值化,并且由于聚类方法本身的特性,故存在一定的随机性。Kowaleski 等<sup>[10]</sup>在前人的基础上提出了回归混合模型,对全球 ECMWF、NCEP、CMA 和 JMA 综合预报系统的台风桑迪轨迹和气旋相空间预报进行了聚类。回归混合模型通过贝叶斯信息准则、聚类分配强度和均方预测误差选择最优模型规格,能较好地实现分类,但此模型仅限于特定台风,存在一定的局限性。目前,在台风路径分类的研究中,聚类算法

的随机性、人工经验和参数问题的智能化以及研究特定台风而产生的局限性等问题依然存在。

模体挖掘源于数据挖掘思想。模体是指在一组序列中频繁出现且最具有代表性的子序列,该方法具有不需要先验干预和数据自动选择的优势<sup>[11-14]</sup>。通过模体挖掘方法,本文从一个全新的角度研究和改进台风路径的分类方法。现有的模体挖掘方法依然需要人工设定参数,不能实现自主挖掘模体<sup>[15-16]</sup>。

本文针对以上问题,提出基于阈值学习的台风路径分类方法。首先,改进传统的 *K-means* 聚类算法,提出基于密度质心的混合聚类算法,实现了数据的预处理,取得比传统聚类算法更好的效果;其次,提出了基于 fastDTW 的自适应模体方法,通过自主阈值学习解决目前存在的依赖先验知识和人工设定参数等问题。

## 1 材料与方法

### 1.1 数据

本文数据来自中国气象台公布的 1949—2018 年台风数据<sup>[17]</sup>,此数据为 CMA 台风最佳路径数据集,包含 1949—2018 年西北太平洋(含南海、赤道以北,东经 180°以西)海域台风每 6 h 的经纬度和强度(图 1)。经过整理,本实验以 TC 编号、每一时刻的经纬度信息、台风最大压强和台风最大风速等作为实验数据。

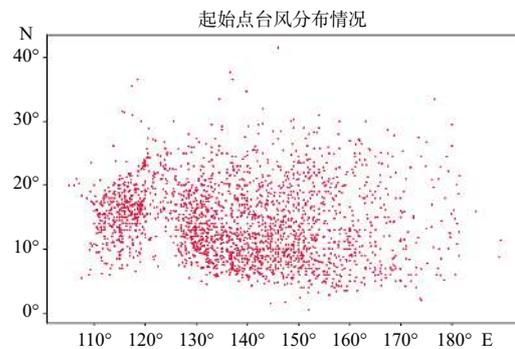


图 1 西北太平洋台风起始点分布情况

Fig. 1 Distribution of typhoon genesis points over Northwest Pacific

### 1.2 定义

定义 1: 设台风路径的观测点  $P=(\text{Lon}, \text{Lat})$ , 其中 Lon 为经度、Lat 为纬度,若  $P$  为起始

点或登陆点,则称  $P$  为关键节点;若  $P$  为起始点则记  $P^*$ 、为登陆点则记  $P^\#$ 。

定义 2: 设存在一个集合  $T=\{P_1, P_2, \dots, P_m\}$ ,若  $T$  中  $P_1$  至  $P_m$  按时间顺序排列,且代表一条台风的生成到结束所经过的所有观测点,则称  $T$  为台风路径,其中  $m$  表示此路径观测点的个数。

定义 3: 设一个集合  $D=\{P_1, P_2, \dots, P_n\}$ ,其中  $P_1, P_2, \dots, P_n$  为关键节点, $n$  表示集合  $D$  中关键节点的个数。若集合  $D$  中元素全为  $P^*$ ,则称为起始热点区域,记为  $D^*$ ;若元素全为  $P^\#$ ,则称为登陆热点区域,记为  $D^\#$ 。

定义 4: 设一个关键节点  $P$ 、常数  $\text{MinPts}$  以及半径  $R$ ,若在以  $P$  为圆心、 $R$  为半径的范围中,节点数量  $\text{Num}(P, R) > \text{MinPts}$ ,则称  $P$  为核心对象,记为  $x$ 。

定义 5: 设存在矩阵  $A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1i} \\ a_{21} & a_{22} & \dots & a_{2i} \\ \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{ij} \end{pmatrix}$ ,

若  $a_{ij}$  为台风  $T_i$  与  $T_j$  两条路径之间的相似值,则称  $A$  为相似矩阵。

### 1.3 基于混合聚类算法的热点区域挖掘

$K$ -means 算法在聚类过程中,存在初始中心点的选择具有较高的随机性问题。如果初始中心选择质量较低,则聚类容易陷入局部最佳解而

忽略全局<sup>[9]</sup>,同时也会带来聚类结果的波动。传统  $K$ -means 算法的随机性将影响台风核心对象的选取,从而影响热点区域挖掘的稳定性,并降低路径分类的精确度。因此,本文针对初始点选择的随机性问题,提出了一种稳定性与代表性俱佳的混合聚类算法,以实现更准确的热点区域挖掘和台风路径分类。

基于混合聚类算法的热点区域挖掘方法是一种基于密度质心的混合聚类算法。混合聚类算法相比传统  $K$ -means 算法的随机性问题,具有明显的优势。混合聚类有两部分:(1)运用 DBSCAN 算法<sup>[18]</sup>密度思想,筛选出  $n$  个初始聚类中心  $\{C_1, C_2, C_3, \dots, C_n\}$ ;(2)使用  $K$ -means 算法,对初始聚类中心进行聚类(如图 2 所示)。首先,输入所有台风路径的起始点以及登陆点,并形成对应的两类节点集合  $D^*$  和  $D^\#$ ;其次,算法通过轮廓系数<sup>[19]</sup>来评价聚类效果,并筛选出密度参数( $R$ 、 $\text{MinPts}$ )以及起始点集合  $D^*$  作为输入,找出核心对象集合  $X\{x_1, x_2, \dots, x_n\}$ ;再次,依次以某一核心对象  $P_i$  为出发点,找出由其密度可达的样本生成的聚类簇中心  $\{C_1, C_2, \dots, C_k\}$ ;最后,将已生成的  $K$  个聚类簇为单位,输入  $K$ -means 算法中,输出各簇类的质心,并以输出各簇类质心为圆点绘制相切圆,最后确定热点区域  $D$ 。



图 2 基于混合聚类算法的热点区域挖掘方法流程

Fig. 2 Flow chart of hot spot mining method based on hybrid clustering algorithm

### 1.4 基于 fastDTW 的自适应模体处理

#### 1.4.1 相似矩阵建立

相似矩阵的建立,采用 fastDTW 方法(fast dynamic time warping, fastDTW)<sup>[20]</sup>对台风路径进行相似度计算,并将所得结果以矩阵的结构进行存储。fastDTW 方法先引入两条长度分别为  $|X|$  和  $|Y|$  的台风路径  $T_1$ 、 $T_2$ ,再利用距离公式  $\text{Dist}(T_1, T_2)$  进行计算:

$$\text{Dist}(T_1, T_2) = \sum_{k=1}^{k=K} \text{Dist}(T_{k1}, T_{k2}) \quad (1)$$

如图 3 和公式(1)所示,将两路径  $T_1$ 、 $T_2$  平均划分为  $K$  段,并结合公式计算出  $T_1$ 、 $T_2$  两台风

路径之间的相似度  $\text{Dist}(T_1, T_2)$ 。最后,将所有满足热点区域的台风路径依次进行相似计算,并存入相似矩阵  $A$  中:

$$a_{ij} = \begin{cases} \text{Dist}(T_i, T_j) & i \neq j \\ 0 & i = j \end{cases} \quad (2)$$

若有  $i$  条路径,则将相似矩阵的  $a_{ij}$  为:

$$a_{ij} = \begin{pmatrix} 0 & \text{Dist}(T_1, T_2) & \dots & \text{Dist}(T_1, T_j) \\ \text{Dist}(T_2, T_1) & 0 & \dots & \text{Dist}(T_2, T_j) \\ \dots & \dots & \dots & \dots \\ \text{Dist}(T_i, T_1) & \text{Dist}(T_i, T_2) & \dots & 0 \end{pmatrix}$$

相似矩阵的  $a_{ij}$  是对应两条台风路径的相似度  $\text{Dist}(T_i, T_j)$ 。同时,相似矩阵  $A$  为方阵,且根据

公式(2)所示, 对角线相似值均为 0。

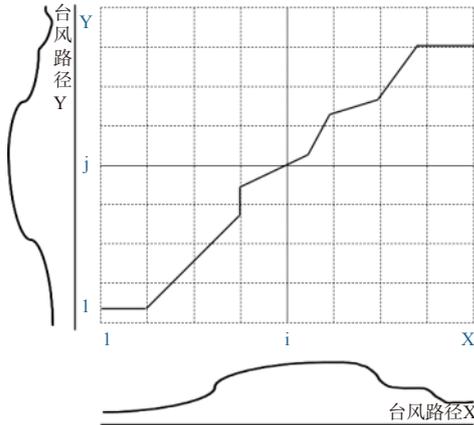


图 3 fastDTW 计算两条台风路径的相似度

Fig. 3 fastDTW calculate the similarity between the two typhoon tracks

### 1.4.2 路径模体选择

建立相似矩阵, 不仅可以直观表示台风路径之间的相似关系, 也有助于路径模体的选择。路径模体选择, 包括对矩阵的相似值进行排序, 计算均值以及中位数, 并计算  $TM$  指数(track motif index), 最终输出路径模体。

$$M_i = \frac{\sum_{k=1}^{k=j} Dist(T_i, T_k)}{Num}$$

$$m_i = \begin{cases} V_{(Num+1)/2} & Num = 2k, k \in N^* \\ \frac{V_{(Num/2)} + V_{(Num/2)+1}}{2} & Num = 2k + 1, k \in N^* \end{cases} \quad (3)$$

式中:  $Num$  为参与相似矩阵运算的台风路径总条数;  $|V_1, V_2, \dots, V_{Num}|$  是排序后的矩阵第  $i$  行的相似值,  $N^*$  是正整数。则第  $i$  行的  $TM$  指数:

$$TM = \frac{M_i + m_i}{\sum_{k=1}^{k=i} M_k + m_k} \quad (4)$$

式中:  $TM$  指数越小, 则可认定在台风路径集合中, 第  $i$  条路径出现更为频繁、更具代表性, 即路径模体。

### 1.4.3 相似分布计算

找出路径集合中段路径模体, 是解决台风路

径分类较为依赖先验知识的关键一步, 同时是建立模体资料库的关键一步。本文以路径模体为标准, 重新构建一个相似链表  $L$ , 通过计算台风路径的相似阈值, 初步实现台风路径的分类。首先, 将链表分为  $1, 2, \dots, N$  段, 代入计算  $SVB$  指数(similar value distribution,  $SVB$ ):

$$SVB = \frac{num_a + num_b + num_c}{\sum_{k=1}^{k=N} num_k} \quad (5)$$

式中:  $a, b, c$  为对应相似阈值段;  $num_a, num_b, num_c$  分别表示根据相似阈值所归属于对应段的台风路径数量。若  $SVB$  指数大于 80%, 则输出  $a, b, c$  阈值段以及  $N$  值。此时, 链表划分为  $N$  段所包含的台风路径构成此模体路径下的模体总库,  $a, b, c$  所包含的台风路径构成 3 个模体资料子库。

### 1.4.4 基于 fastDTW 的自适应模体处理算法

基于 fastDTW 的自适应模体处理算法, 相对现有的台风分类方法, 重点解决了先验知识的问题, 同时也解决了机器学习等传统模型参数过多的问题。模体处理通过无人干预、在数据中自主学习相似阈值的方法, 找出台风中的路径模体, 实现了台风路径分类智能化, 为提高台风路径分类的精确度打下基础。

如图 4 所示, 通过 fastDTW 算法将满足区域条件的台风集合 ( $D$ ) 中的各路径依次进行比较, 计算其相似性, 并保存至相似矩阵; 其次, 输出对应的相似矩阵  $A_{ij}$ ; 再次, 计算对应矩阵每行的均值、中位数,  $TM$  指数, 输出路径模体; 最后, 构建相似链表  $L$ , 计算  $SVB$  指数, 输出相似值段  $N$ , 以及子库所需的  $a, b, c$  3 个阈值段。

## 2 结果与讨论

### 2.1 结果

#### 2.1.1 热点区域处理结果

通过基于混合聚类算法的热点区域挖掘方法, 本实验以近 70 年所有台风路径为例, 提取台



图 4 基于 fastDTW 的自适应模体处理算法流程图

Fig. 4 Flow chart of adaptive motif processing algorithm based on fastDTW

风路径的起始点、登陆点,并利用基于密度质心的混合聚类算法,将其分别进行聚类。如图 5 所示,实验输出登陆区域共 5 个,其中黄色圆圈则表示热点登陆区域质心。同时,本次实验主要以

登陆中国的台风作为研究样本,结合输出登陆区域结果,将登陆区域确定为湛江区域、汕尾区域、台东区域、温州区域以及黄海区域共 5 个登陆区域。

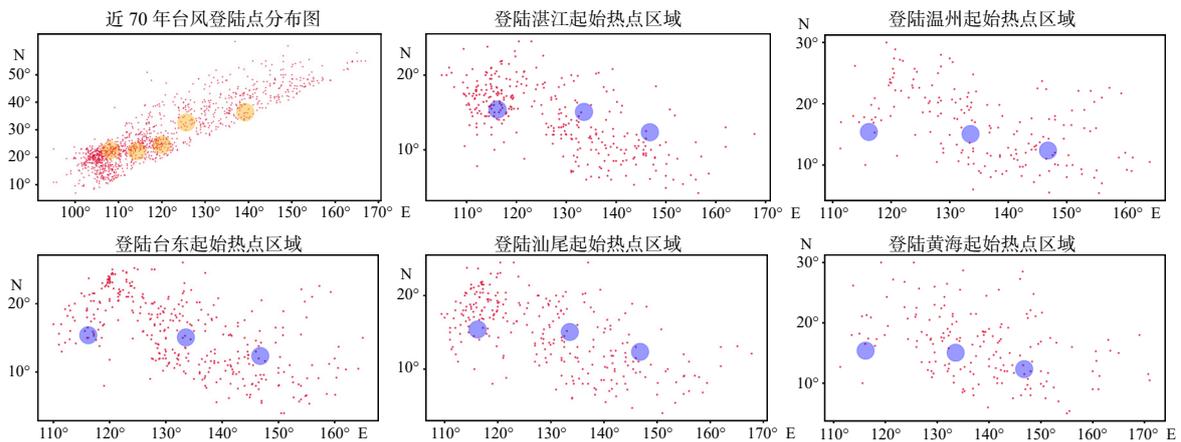


图 5 台风登陆的起始热点区域

Fig. 5 The starting hot spot area of typhoon landing

轮廓系数<sup>[19]</sup>为聚类结果质量的评价标准,用于计算聚类簇间的相似度,其取值范围为 $[-1, 1]$ 。若轮廓系数越接近 1,则表示聚类效果越好。本实验将起始区域以登陆区域结果为依据,共分成 5 组数据集,输入参数( $R, MinPts$ ),并通过轮廓系数进行结果分析。输出误差分析结果如表 1 所示,聚类结果如图 5 所示。

基于以上登陆区域,再次筛选对应登陆区域的台风起始点,再通过混合聚类算法自动筛选

出 3 个起始热点区域。综上,本次实验形成两类热点区域,完成数据预处理过程。

同时,通过与  $K$ -means 算法进行对比实验,进一步证明了混合聚类算法更为稳定、更具代表性。对比实验共进行 4 次,并以两种聚类算法输出的区域质心为中心,分别计算满足条件的台风轨迹数目。通过图 6 可知,混合聚类算法相对于  $K$ -means 算法,在输出结果上更为稳定,能更好地解决聚类算法的随机性问题。

表 1 聚类评价结果对比分析

Tab.1 Comparative analysis of cluster evaluation results

算法	$R=2$	$R=3$	$R=4$	$R=5$	$R=6$
$K$ -means算法	0.65433	0.5124	0.49177	0.4701	0.40854
DBSCAN算法	0.01261	0.35646	0.23986	0.27828	0.31626
混合聚类算法	0.57232	0.72362	0.62431	0.5923	0.55673

### 2.1.2 基于 fastDTW 的自适应模体处理结果

路径模体的选择,是自适应模体处理过程的关键步骤。首先,针对预处理后的路径数据,建立对应的相似矩阵,通过  $TM$  指数,选择出此路径集合中的路径模体。本实验以 1949—2018 年登陆湛江的台风数据来进行自适应模体处理,通过建立相似矩阵,实现了路径模体的选择。

以湛江为例,本文绘制了近 70 年登陆湛江的台风轨迹路径集合和路径模体(图 7)。经过计算,其  $TM$  指数分别为 0.003522、0.007527、0.017643,即为路径模体。同时,将 3 条路径分别进行相似分布计算,最终得到  $N$  为 11,将长度为 500 的相似链表分为 11 段,3 个阈值段  $a$ 、 $b$ 、 $c$  分别为 0~150、0~150、50~200,  $SVB$  指数值分别

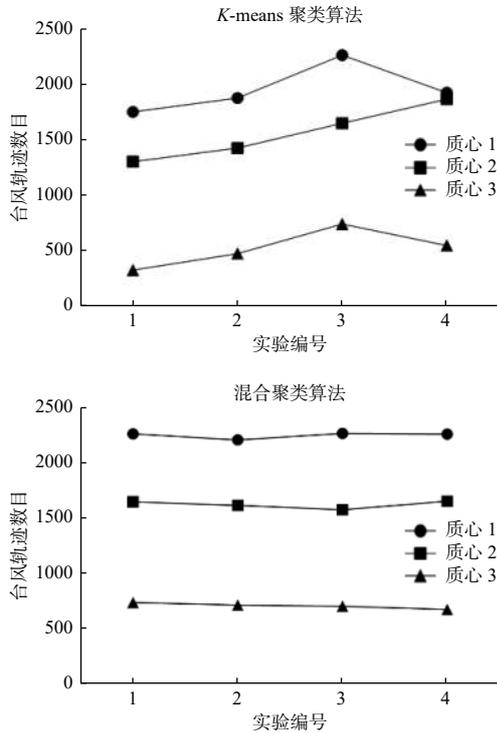


图 6 算法稳定性对比实验结果

Fig. 6 Algorithm performance comparison results

为 95.92%、89.16%、89.47%。

采用上述方法, 本文对汕尾、温州、台东以及黄海进行了相同的计算, 得到路径模体如图 8。

将本文的实验结果与传统 K-means 聚类方法分类的结果比较可知, 基于 fastDTW 的自适应

模体处理方法可以将起始南海区域的台风进一步划分成 5 类, 具有一定的创新性和代表性。本文的算法将传统的西移、转向台风路径从不同的起始区域和登陆区域进行细化。

### 2.2 讨论

#### 2.2.1 台风路径可视化和数据库领域应用

2.1.2 节路径模体以及相似阈值学习的结果, 可以用于模体资料库的构建。如图 9 所示, 以相似段值  $N=11$  为子类, 进一步数值化路径相似分布, 构建模体资料库对应的子资料库。

图 9 分别绘制了南海区域登陆湛江的 3 个台风模体子资料库, 其相似分布分别为 0 ~ 50、50 ~ 100、100 ~ 150; 在菲律宾东部区域登陆湛江的 3 个台风模体子资料库, 分别为 0 ~ 50、50 ~ 100、100 ~ 150; 在群岛区域登陆湛江的 3 个台风模体子资料库, 分别为 50 ~ 100、100 ~ 150、150 ~ 200。

由图 9 可知, 历史台风路径的相似阈值越接近, 则地理位置和路径形状都越接近, 这充分证明了基于相似阈值学习的台风路径分类方法, 能很好地应用于未来模体资料库的建立。因此, 该方法对台风路径的可视化和台风数据库的建设有一定的指导意义。

#### 2.2.2 台风灾害分析领域应用

通过九类路径模体资料库压强变化的分析,

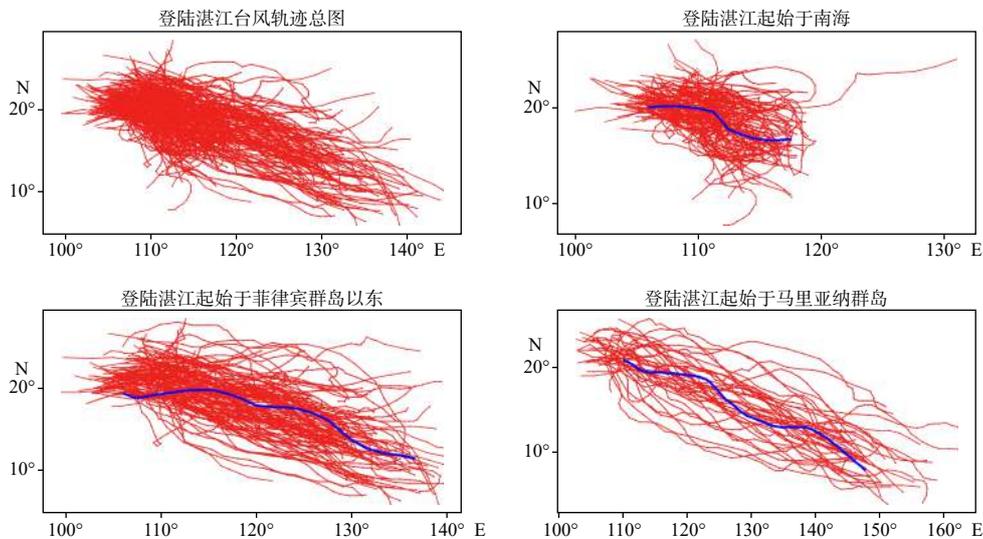


图 7 登陆湛江的台风轨迹和路径模体

Fig. 7 Typhoon track and path motif landing in Zhanjiang

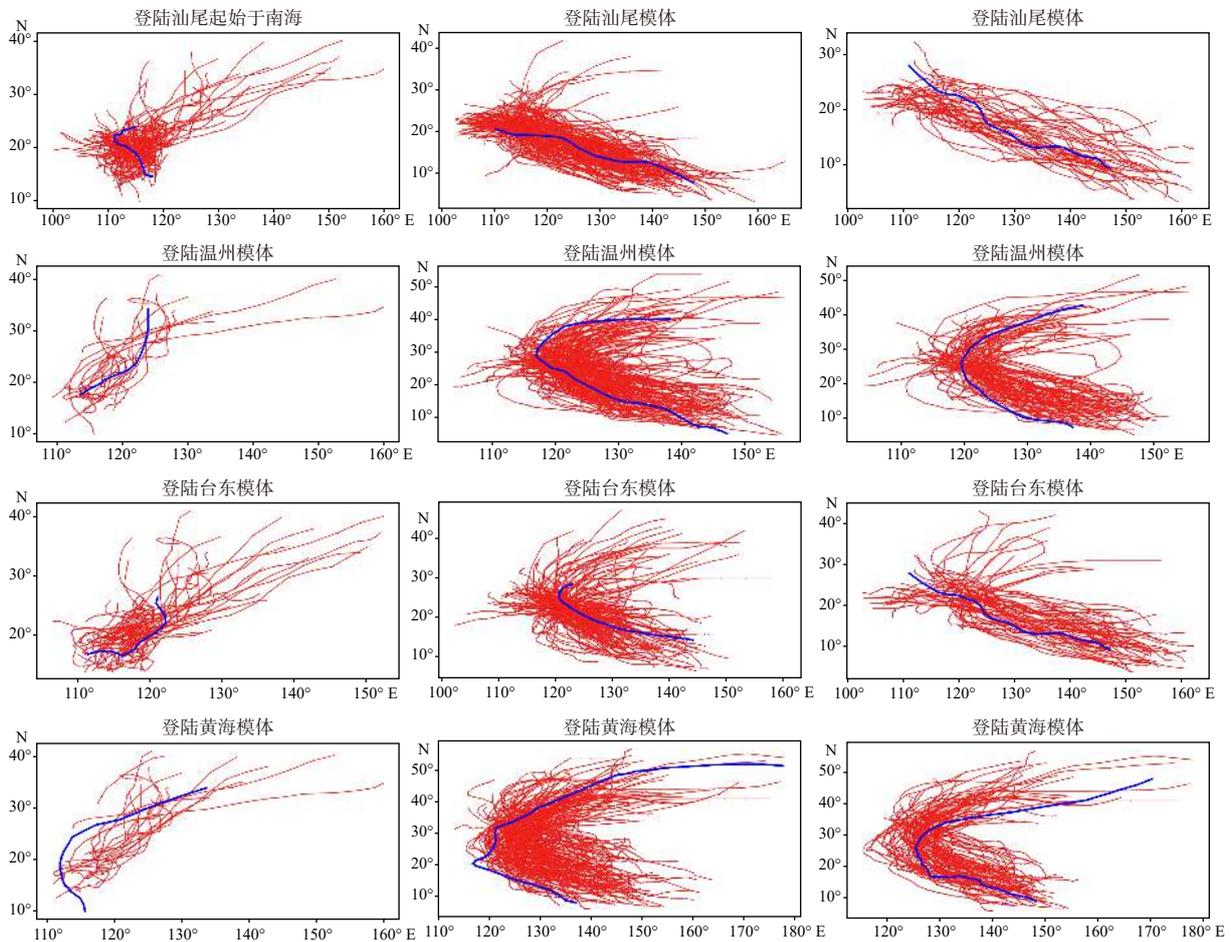


图 8 登陆路径模体

Fig. 8 Landing path motif

进一步证实了基于相似阈值学习的台风路径分类算法在台风灾害分析领域的应用价值。

图 10 从左至右分别表示起始南海、菲律宾以东、各群岛区域并登陆湛江的台风(图 7)子库的压强变化与分布情况。由图 10 可知,起始南海的路径模体资料库中 3 类子库的压强分布以及变化趋势基本一致。其中,子库一台风的生命周期较短,普遍为 96 h,且于 48 h 时压强达到最小值;子库二台风的生命周期适中,普遍为 150 h,且于 90 h 时压强达到最小值;子库三台风的生命周期较长,普遍为 210 h,且于 132 h 时压强达到最小值。

通过本文的路径模体资料库可以发现,各子资料库的压强分布以及变化趋势具有明显的规律性,能很好地体现基于阈值学习的台风路径分

类方法的正确性。本实验虽然没有直接以压强等相关物理量参与分类,但是分类后的台风路径中,其物理量却有相当明显的差异。

### 3 结论

(1)通过热点区域处理,确定了台风路径起始、登陆的热点区域,解决了传统聚类的随机性问题,实现了更加精确与直观的聚类效果,并与现有研究进行对比,证明了此方法的正确性以及实用性。

(2)以登陆湛江区域的台风为例,通过 fastDTW 自适应模体处理方法挖掘出  $TM$  指数分别为 0.003522、0.007527、0.017643 的台风路径模体,计算后的  $SVB$  指数分别为 95.92%、89.16%、89.47%,通过模体资料库分析可知,混

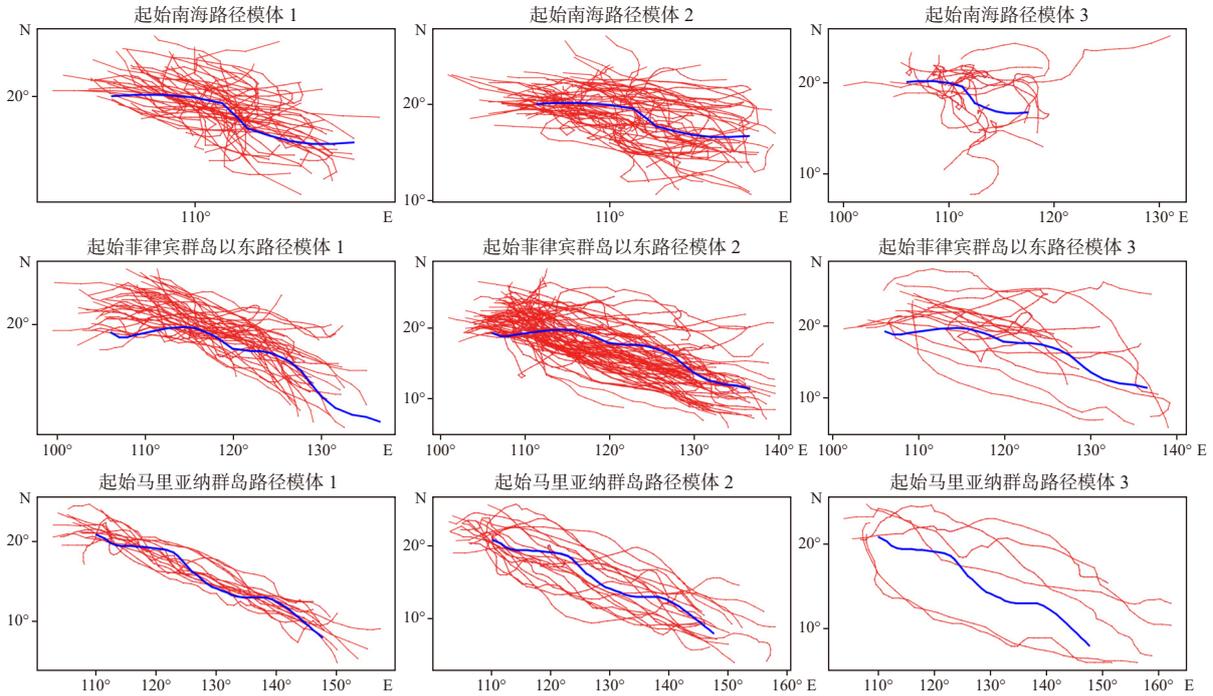


图 9 以登陆湛江的路径模体为标准, 根据相似阈值划分出的台风路径模体子资料库

Fig. 9 Taking the landing path motif of Zhanjiang as the standard, the sub database of typhoon path motifs were classified according to the similarity threshold

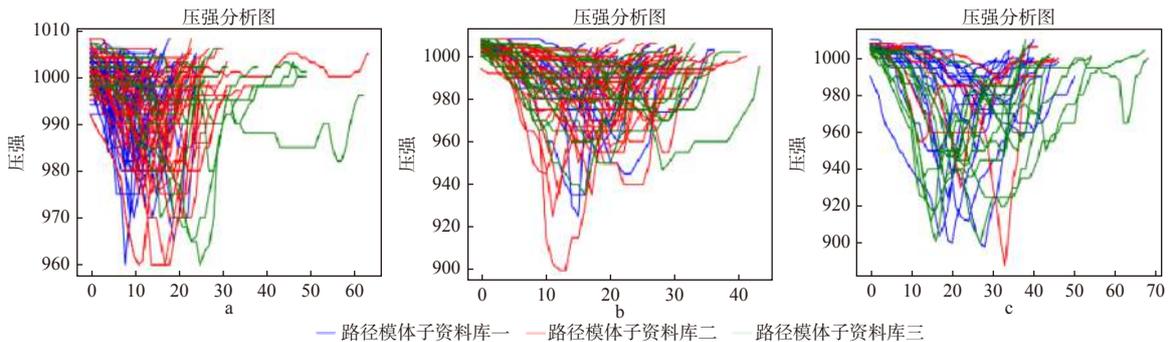


图 10 九类路径模体资料子库中单条台风压强变化趋势以及各子库中路径压强总体分布情况

Fig. 10 The pressure variation trend of a single typhoon in the nine classpath model subdatabase and the overall distribution of the path pressure in each subdatabase

合聚类方法具有一定的应用价值。

(3) 将台风路径进行分类并验证, 结合台风路径可视化以及台风数据库的建设等领域, 实现了台风路径分类的数值化、智能化, 为探索台风路径的可视化与台风登陆区域的灾害分析提供了研究思路。

参考文献:

[1] REN R, YU D Q, WANG L X, et al. Typhoon triggered operation tunnel debris flow disaster in coastal areas of SE

China[J]. *Geomatics, Natural Hazards and Risk*, 2019, 10(1): 562-575.  
 [2] 陈联寿, 端义宏, 宋丽莉, 等. 台风预报及其灾害[M]. 北京: 气象出版社, 2012.  
 [3] MENG Y, LU J, LIAO Q L. Analysis of the characteristics of west-pacific typhoons affecting the sea area around the island of Taiwan[J]. *Journal of Tropical Meteorology*, 2005, 11(2): 222-223.  
 [4] 秦丽娟, 董庆, 薛存金. 西北太平洋热带气旋源地30 a的季节和年代际变化[J]. *海洋环境科学*, 2015, 34(5): 723-728.  
 [5] 王毅, 石汉青, 黄思训. 西北太平洋热带气旋的时空分布

- 特征分析[J]. *自然灾害学报*, 2009, 18(6): 166-174.
- [6] ELSNER J B. Tracking hurricanes[J]. *Bulletin of the American Meteorological Society*, 2003, 84(3): 353-356.
- [7] NAKAMURA J, LALL U, KUSHNIR Y, et al. Classifying north Atlantic tropical cyclone tracks by mass moments[J]. *Journal of Climate*, 2009, 22(20): 5481-5494.
- [8] CAMARGO S J, ROBERTSON A W, GAFFNEY S J, et al. Cluster analysis of typhoon tracks. Part II: Large-scale circulation and ENSO[J]. *Journal of Climate*, 2007, 20(14): 3654-3676.
- [9] 伍育红. 聚类算法综述[J]. *计算机科学*, 2015, 42(S1): 491-499, 524.
- [10] KOWALESKI A M, EVANS J L. Regression mixture model clustering of multimodel ensemble forecasts of hurricane sandy: partition characteristics[J]. *Monthly Weather Review*, 2016, 144(10): 3825-3846.
- [11] LIN J, KEOGH E, WEI L, et al. Experiencing SAX: a novel symbolic representation of time series[J]. *Data Mining and Knowledge Discovery*, 2007, 15(2): 107-144.
- [12] CHIU B, KEOGH E, LONARDI S. Probabilistic discovery of time series motifs[C]//Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Washington: ACM, 2003: 493-498.
- [13] MUEEN A, KEOGH E, ZHU Q, et al. Exact discovery of time series motifs[C]//Proceedings of the 2009 SIAM International Conference on Data Mining. Sparks: Society for Industrial and Applied Mathematics, 2009: 473-484.
- [14] LI Y, LIN J, OATES T. Visualizing variable-length time series motifs[C]//Proceedings of the 2012 SIAM International Conference on Data Mining. Sparks: Society for Industrial and Applied Mathematics, 2012: 895-906.
- [15] GAO Y F, LIN J, RANGWALA H. Iterative grammar-based framework for discovering variable-length time series motifs[C]//Proceedings of 2016 15th IEEE International Conference on Machine Learning and Applications. Anaheim: IEEE, 2016: 111-116.
- [16] YEH C C M, ZHU Y, ULANOVA L, et al. Time series joins, motifs, discords and shapelets: a unifying view that exploits the matrix profile[J]. *Data Mining and Knowledge Discovery*, 2018, 32(1): 83-123.
- [17] YING M, ZHANG W, YU H, et al. An overview of the china meteorological administration tropical cyclone database[J]. *Journal of Atmospheric and Oceanic Technology*, 2014, 31(2): 287-301.
- [18] KUMAR K M, REDDY A R M. A fast DBSCAN clustering algorithm by accelerating neighbor searching using Groups method[J]. *Pattern Recognition*, 2016, 58: 39-48.
- [19] 朱连江, 马炳先, 赵学泉. 基于轮廓系数的聚类有效性分析[J]. *计算机应用*, 2010, 30(S2): 139-141.
- [20] LOHRER J, LIENKAMP M. Building representative velocity profiles using FastDTW and spectral clustering[C]//Proceedings of 2015 14th International Conference on its Telecommunications. Copenhagen: IEEE, 2015: 45-49.