

# 基于 L-Kshape-HACA 的空战态势分割聚类

张 鹏<sup>1</sup>, 黄长强<sup>1</sup>, 魏政磊<sup>1</sup>, 周 欢<sup>1</sup>, 王永乾<sup>2</sup>

(1. 空军工程大学航空工程学院, 西安, 710038; 2. 95596 部队, 河南商丘, 476000)

**摘要** 在近距空战中, 实时获取可靠的空战态势信息对于决策指引是非常重要的。针对近距空战态势变化剧烈以及评估参数多维耦合的问题, 提出了一种基于 L-Kshape-HACA 的多元空战态势分割聚类方法。以分层时序聚类分析为框架, 利用拉普拉斯中心性方法确定聚类数目, 同时采用 Kshape 对多元时间序列进行聚类分析, 解决了多维参数下的态势信息提取问题。利用 12 组近距空战数据进行测试, 并与 14 种聚类算法进行比较, 结果表明 L-Kshape-HACA 在聚类中心确定和态势分割准确性上与实际空战态势更加符合。

**关键词** 无人自主空战; 层次聚类分析; 多元态势提取; 拉普拉斯中心性方法

**DOI** 10.3969/j.issn.1009-3516.2021.03.003

中图分类号 V279 文献标志码 A 文章编号 1009-3516(2021)03-0015-08

## Air Combat Situation Segmentation Clustering Based on L-Kshape-HACA

ZHANG Peng<sup>1</sup>, HUANG Changqiang<sup>1</sup>, WEI Zhenglei<sup>1</sup>, ZHOU Huan<sup>1</sup>, WANG Yongqian<sup>2</sup>

(1. Aeronautical Engineering College, Air Force Engineering University, Xi'an 710038, China;  
2. Unit 95596, Shangqiu 476000, Henan, China)

**Abstract** In close air combat, it is very important to get a piece of reliable air combat situation information in time for decision-making guidance. Aimed at the problems that there are drastic changes in the situation of close air combat and multi-dimensional coupling of evaluation parameters, a multi-element air combat situation segmentation clustering method based on L-Kshape-HACA is proposed. Taking the hierarchical time series cluster analysis as a framework, the number of clusters is determined by using the Laplace centrality method, and the cluster analysis on multivariate time series is made by using Kshape, solving the problem of situation information extraction under multi-dimensional parameters. The test is performed by using the 12 sets of close air combat data, and the 14 clustering algorithms are compared. The results show that L-Kshape-HACA is more consistent with the actual air combat situation in the accuracy of cluster center determination and situation segmentation.

**Key words** close air combat; hierarchical cluster analysis; multivariate situation extraction; laplace centrality method

当前军事装备正向着无人化、智能化的方向发展<sup>[1]</sup>, 制约无人作战飞机作战模式由对地侦查、打击向对空作战转变的主要因素在于自主决策能力不

足, 过分依赖地面站远程操控, 无法满足高动态空战的实时性需求。基于当前空战数据实时提取敌我态势信息来指导无人平台做出自主决策是无人自主空

收稿日期: 2021-01-21

基金项目: 陕西省自然科学基金(2020JQ-481)

作者简介: 张 鹏(1996—), 男, 重庆人, 硕士生, 研究方向: 无人自主空战。E-mail: 2191289994@qq.com

引用格式: 张鹏, 黄长强, 魏政磊, 等. 基于 L-Kshape-HACA 的空战态势分割聚类[J]. 空军工程大学学报(自然科学版), 2021, 22(3): 15-22. ZHANG Peng, HUANG Changqiang, WEI Zhenglei, et al. Air Combat Situation Segmentation Clustering Based on L-Kshape-HACA[J]. Journal of Air Force Engineering University (Natural Science Edition), 2021, 22(3): 15-22.

战的关键技术。空战态势知识提取可以转换为近距空战态势模式的发现,其本质是多元时间序列的分割聚类问题。多元时间序列的聚类算法包括两种思路:一种是通过对多元时间序列进行压缩降维,再利用传统的点聚类方法进行聚类<sup>[2]</sup>;另外一种是结合多元时间序列相似性度量方法改进传统的点聚类方法<sup>[3-4]</sup>。前者又分为基于特征的聚类和基于模型的聚类,后者为基于初始序列的聚类。基于特征和模型的聚类会忽略原始时间序列的局部特征。基于原始时间序列的聚类一般不对时间序列进行压缩,而是直接进行聚类,该方法随着时间序列增长,其计算代价越大<sup>[5]</sup>。同时目前的空战态势知识提取文献大多数以空战态势信息到固定态势类别的映射居多,不能体现聚类中心的客观性和态势变化的连续性。

针对以上问题,本文提出了一种基于拉普拉斯中心性和 Kshape 的分层聚类时序分析方法,解决了多维参数下的态势信息提取问题。

## 1 基于拉普拉斯中心性的聚类数目确定

目前的分割方法主要是基于特征点的分割和基于监督学习的分割点检测<sup>[6-7]</sup>。然而实际的时序单元特征无法具体描述或者表示,因此采用基于无监督学习的分割方法是非常有必要的<sup>[8]</sup>。考虑到传统聚类算法以及基于峰值密度算法(density peak clustering,DPC)<sup>[9]</sup>的聚类个数无法确定的缺点,本文采用拉普拉斯中心峰值聚类方法(Laplacian peaks clustering algorithm,LPC)<sup>[10]</sup>,将原始数据转换到加权完全图中,通过构造决策图可以选取聚类个数以及聚类中心;LPC 在整个聚类过程中不需要任何经验参数,因此 LPC 也是一种无参数聚类算法。

### 1.1 相关参数定义

敌我双方的态势信息主要取决于空战中双方的相对运动关系,如表 1 所示。

表 1 态势描述参数

态势特征参数	取值范围
敌我双方相对距离 $R/\text{km}$	$[0, 20]$
敌我双方相对高度 $\Delta h/\text{m}$	$[0, 15\,000]$
我方 UCAV 速度 $v_u/(\text{m/s})$	$[82, 590]$
我方 UCAV 相对敌方的方位角 $\varphi_u/\text{rad}$	$[0, \pi]$
我机进入角 $q_u/\text{rad}$	$[0, \pi]$
敌机速度 $v_e/(\text{m/s})$	$[82, 590]$
敌机相对我机的方位角 $\varphi_e/\text{rad}$	$[0, \pi]$
敌机进入角 $q_e/\text{rad}$	$[0, \pi]$

假设以我机质心为坐标原点  $O$ , $X$  轴沿 UCAV

纵轴指向机头方向, $Y$  轴垂直于 UCAV 对称面并值向右, $Z$  轴位于 UCAV 对称面并垂直于机体纵轴向下,我方 UCAV 与敌机位置坐标为  $\mathbf{P}_u = [x_u, y_u, z_u]$  和  $\mathbf{P}_e = [x_e, y_e, z_e]$ ,则双机相对态势定义见图 1。

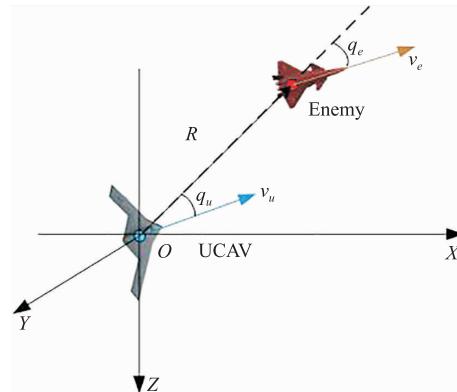


图 1 空战几何关系

假设原始未分类的空战态势描述参数数据集为  $\mathbf{X}_s = \{\mathbf{X}_{s,1}, \mathbf{X}_{s,2}, \dots, \mathbf{X}_{s,t_n}\}$ , 其中  $\mathbf{X}_{s,i} = [R_i, \Delta h_i, v_{u,i}, \varphi_{u,i}, q_{u,i}, v_{e,i}, \varphi_{e,i}, q_{e,i}]$ ,  $i=1, 2, \dots, t_n$ ,  $t_n$  为获取的空战数据末端时间点。为了挖掘聚类中心, 定义 2 个主要指数: 第  $i$  个数据点的拉普拉斯中心性  $c_i$  和最小距离值  $\delta_i$ 。其次, 将数据集转换为加权完全图  $\mathbf{G} = (\mathbf{X}_s, \mathbf{E}, \mathbf{W})$ , 将每个数据点看作加权网络  $\mathbf{E}$  的节点,  $\mathbf{W}$  表示边权值的集合。具体定义如下:

**定义 1** 相异度矩阵  $\mathbf{W}$ 。 $w_{i,j}$  是数据点  $\mathbf{X}_{s,i}$  与  $\mathbf{X}_{s,j}$  之间的边权值, 用于衡量多元时间序列相似性。由于原始态势参数序列不等长且具有高动态性, 本文采用动态时间弯曲距离法(dynamic time warping, DTW)<sup>[11]</sup> 进行计算。

$$\mathbf{W}(\mathbf{G}) = \begin{bmatrix} 0 & w_{1,2} & \cdots & w_{1,t_n} \\ w_{2,1} & 0 & \cdots & w_{2,t_n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{t_n,1} & w_{t_n,2} & \cdots & 0 \end{bmatrix} \quad (1)$$

**定义 2** 和对角矩阵  $\mathbf{Z}$ :

$$\mathbf{Z}(\mathbf{G}) = \text{diag}(z_1, z_2, \dots, z_n) \quad (2)$$

式中:  $z_i = \sum_{j=1}^{t_n} w_{i,j}$ 。

**定义 3** 拉普拉斯矩阵  $\mathbf{L}$ :

$$\mathbf{L}(\mathbf{G}) = \mathbf{Z}(\mathbf{G}) - \mathbf{W}(\mathbf{G}) \quad (3)$$

**定义 4** 拉普拉斯能量  $E_L(\mathbf{G})$ :

$$E_L(\mathbf{G}) = \sum_{i=1}^{t_n} \lambda_i^2 \quad (4)$$

式中:  $\lambda_1, \lambda_2, \dots, \lambda_{t_n}$  为拉普拉斯矩阵  $\mathbf{L}(\mathbf{G})$  的特征值。

**定义 5** 拉普拉斯中心性  $c_i$ 。在加权完全图  $\mathbf{G}$  中, 删除节点  $\mathbf{X}_{s,i}$  以及与其相连的边, 得到的加权完全图为  $\mathbf{G}_i$ 。节点  $\mathbf{X}_{s,i}$  的拉普拉斯中心性  $c_i$  可以定

义为:

$$c_i = \frac{E_L(\mathbf{G}) - E_L(\mathbf{G}_i)}{E_L(\mathbf{G})} \quad (5)$$

式中: $E_L(\mathbf{G}_i)$ 表示删除节点 $\mathbf{X}_{s,i}$ 之后的加权完全图的拉普拉斯能量,拉普拉斯中心性的意义在于删除当前节点后的拉普拉斯能量相对下降。当前节点拉普拉斯中心性越高,则该节点重要程度越高<sup>[11]</sup>。

**定义 6** 最小距离值 $\delta_i$ 。对于当前节点 $\mathbf{X}_{s,i}$ ,通过比较拉普拉斯中心性 $c_i$ ,可以得到最小距离值 $\delta_i$ ,其具体计算如下:

$$\delta_i = \begin{cases} \max \{w_{i,j} \mid j=1, 2, \dots, t_n\}, & \text{if } c_i = \max c \\ \min_{j: c_j > c_i} w_{i,j}, & \text{otherwise} \end{cases} \quad (6)$$

## 1.2 基于正态分布“ $3\sigma$ ”准则的聚类个数提取策略

根据上述定义的分析,具有较高最小距离值和较高拉普拉斯中心性的数据节点可以作为聚类中心,由此可以确定聚类个数。利用最小距离值 $\delta$ 和拉普拉斯中心性 $c$ 构造决策图<sup>[12]</sup>。为了自动完成这一过程,本文采用了基于正态分布“ $3\sigma$ ”准则的聚类个数提取策略。为了从数据节点中自动提取聚类个数,引出另外一个决策量 $\gamma$ ,其定义如下:

$$\gamma_i = c_i \cdot \delta_i \quad (7)$$

为了方便提取聚类中心与聚类个数,对 $\gamma$ 值进行放大:

$$E_{\gamma,i} = \sum_{j \in X_s \setminus \{i\}} \sqrt{(\gamma_i - \gamma_j)^2} \quad (8)$$

式中: $E_{\gamma,i}$ 为节点 $\mathbf{X}_{s,i}$ 由决策量 $\gamma_i$ 放大后的值。

假设数据节点的扩大后决策值 $E_\gamma$ 服从正态分布,根据正态分布的“ $3\sigma$ ”准则,将 $(\mu - 3\sigma, \mu + 3\sigma)$ 区域看作该随机变量实际可能的取值区间。由于潜在的聚类中心的决策值 $E_\gamma$ 远大于非聚类中心的 $E_\gamma$ 值,可以得到 $E_\gamma$ 值大于 $\mu + 3\sigma$ 的数据节点。图 2 给出基于拉普拉斯中心性的决策图,其中红色的数据节点为 $E_\gamma$ 值大于 $\mu + 3\sigma$ 的数据节点,即聚类中心,因此聚类个数为 3。

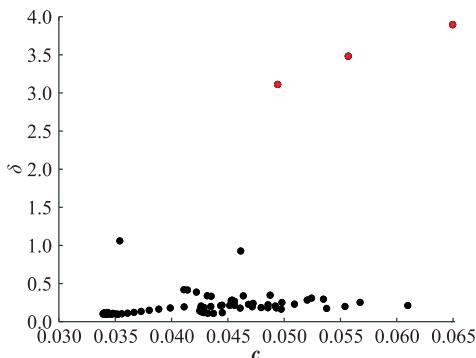


图 2 某空战态势描述参数决策图

## 2 基于拉普拉斯中心性的聚类中心确定

时间序列的聚类算法的提出主要来源于 2 种途径<sup>[13]</sup>:第 1 种是替代相似性度量方法,属于基于原始数据的聚类方法;第 2 种是将时间序列转换为静态数据,包括基于特征和基于模型的聚类方法。基于特征和基于模型的聚类方法是对原始时间序列进行特征提取或者模型表示,不能体现出时序信息。本文选择基于原始数据的时间序列聚类方法,其主要依赖于矩阵的分解。

针对以上问题的分析,本文提出了一种基于原始数据的多元时间序列聚类方法,即 Kshape。首先,提出一种基于互相关度量的形状距离度量方法;其次,基于形状距离度量方法,设计一种计算时间序列聚类中心的计算方法;最后,根据相似性度量方法与聚类中心计算方法,给出了基于 Kshape 的多元时间序列聚类算法具体步骤。

### 2.1 时间形状相似性度量方法

为了提高时间序列相似性度量计算效率,本文设计了一种互相关度量方法。

对于序列 $\mathbf{X}_s = \{X_{s,1}, X_{s,2}, \dots, X_{s,n}\}$ 与 $\mathbf{Y}_s = \{Y_{s,1}, Y_{s,2}, \dots, Y_{s,m}\}$ ,为了实现位移不变性,固定时间序列 $\mathbf{Y}_s$ 不动,滑动 $\mathbf{X}_s$ ,互相关度量就是计算滑动过程中对齐的元素之间的内积。 $\mathbf{X}_s$ 时间序列的移位序列计算表达式为:

$$\mathbf{X}_{s(\kappa)} = \begin{cases} (\overbrace{0, \dots, 0}^{\lfloor \frac{n}{2} \rfloor}, x_{s,1}, x_{s,2}, \dots, x_{s,n-\kappa}), & \kappa \geq 0 \\ (x_{s,1-\kappa}, \dots, x_{s,n-1}, x_{s,n}, \underbrace{0, \dots, 0}_{\lceil \frac{n}{2} \rceil}), & \kappa < 0 \end{cases} \quad (9)$$

式中: $\mathbf{X}_{s(\kappa)}$ 表示移位 $\kappa$ 之后的时间序列。一般情况下, $\kappa \in (-n, n)$ 或者 $\kappa \in [-n+1, n-1]$ 。假设 $n=3$ ,那么其移位时间序列的个数为 $2n-1=5$ ,为了方便理解,给出示意图见图 3。移位序列中, $\mathbf{X}_{s(0)}$ 就是原始时间序列,其他移位序列的某些元素需要填充为 0。

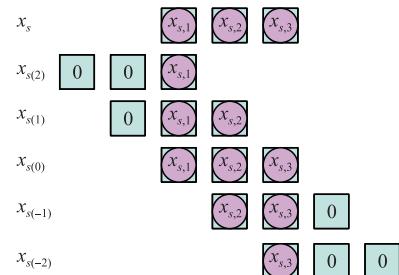


图 3 移位时间序列

根据移位序列 $\mathbf{X}_{s(\kappa)}$ ,可以得到互相关序列 $\mathbf{CC}_\omega(\mathbf{X}_s, \mathbf{Y}_s) = (c_1, c_2, \dots, c_\omega)$ ,其长度为 $2n-1$ ,与

移位序列的个数相等。互相关序列可以定义如下:

$$\mathbf{CC}_\omega(\mathbf{X}_s, \mathbf{Y}_s) = \mathbf{R}_{\omega-n}(\mathbf{X}_s, \mathbf{Y}_s) \quad (10)$$

式中:  $\omega \in \{1, 2, \dots, 2n-1\}$ 。同时,  $\mathbf{R}_{\omega-n}(\mathbf{X}_s, \mathbf{Y}_s)$  的计算公式如下:

$$\mathbf{R}_k(\mathbf{X}_s, \mathbf{Y}_s) = \begin{cases} \sum_{l=1}^{n-k} x_{s,l+k} y_{s,l}, & k \geq 0 \\ \mathbf{R}_{-k}(\mathbf{X}_s, \mathbf{Y}_s), & k < 0 \end{cases} \quad (11)$$

通过计算式(10)中的  $\mathbf{CC}_\omega(\mathbf{X}_s, \mathbf{Y}_s)$  中最大的互相关序列值, 得到对应的  $\omega$  值。基于  $\omega$  值, 可以得到  $\kappa = \omega - n$ , 对应的移位序列  $\mathbf{X}_{s(\kappa)}$  是最优的匹配序列。由于问题的不同, 则采用的标准化方法不同。本文选取了文献[15]中的系数标准化方法, 即  $NCC_c$ , 具体定义如下:

$$NCC_c(\mathbf{X}_s, \mathbf{Y}_s) = \frac{\mathbf{CC}_\omega(\mathbf{X}_s, \mathbf{Y}_s)}{\sqrt{\mathbf{R}_0(\mathbf{X}_s, \mathbf{X}_s)} \cdot \sqrt{\mathbf{R}_0(\mathbf{Y}_s, \mathbf{Y}_s)}} \quad (12)$$

式中:  $NCC_c(\mathbf{X}_s, \mathbf{Y}_s)$  的值域为  $[-1, 1]$ 。 $NCC_c(\mathbf{X}_s, \mathbf{Y}_s) = 1$  说明两个时间序列相似,  $NCC_c(\mathbf{X}_s, \mathbf{Y}_s) = -1$  说明两个时间序列负相关。

为了设计一种基于形状距离的度量方法, 将  $NCC_c(\mathbf{X}_s, \mathbf{Y}_s)$  取值映射到  $[0, 2]$ , 具体计算公式如下:

$$SBD(\mathbf{X}_s, \mathbf{Y}_s) = 1 - \max_{\omega} \frac{\mathbf{CC}_\omega(\mathbf{X}_s, \mathbf{Y}_s)}{\sqrt{\mathbf{R}_0(\mathbf{X}_s, \mathbf{X}_s)} \cdot \sqrt{\mathbf{R}_0(\mathbf{Y}_s, \mathbf{Y}_s)}} \quad (13)$$

为了提高基于形状距离(shape-based distance, SBD)的相似性度量计算过程, 采用  $\mathbf{X}_s$  与  $\mathbf{Y}_s$  离散傅里叶变换内积的离散傅里叶逆变换计算<sup>[14]</sup>。具体的计算公式如下:

$$\mathbf{CC}(\mathbf{X}_s, \mathbf{Y}_s) = F^{-1}\{(\mathbf{X}_s)(\mathbf{Y}_s)\} \quad (14)$$

$$F(\mathbf{x}_k) = \sum_{r=0}^{|x|-1} \mathbf{x}_r e^{-2jk\pi/|x|} \quad (15)$$

$$F^{-1}(\mathbf{x}_r) = \sum_{k=0}^{|x|-1} (\mathbf{x}_k) e^{2jr\pi/|x|} \quad (16)$$

式中:  $F$  为傅里叶变换,  $F^{-1}$  为傅里叶逆变换;

$j = \sqrt{-1}$ ;  $k = 0, 1, \dots, |x| - 1$ ;  $r = 0, 1, \dots, |x| - 1$ 。

## 2.2 聚类中心提取策略

根据基于 SBD 的相似性度量方法, 本文将聚类中心计算作为一个优化问题, 其目标是寻找当前聚类中心到所属其他时间序列之间的距离平方和最小化, 从而得到最优的聚类中心  $\mu_k^*$ , 具体表示如下:

$$\begin{aligned} \mu_k^* &= \arg \max_{\mu_k} \sum_{\mathbf{x}_{s,i} \in C_k} NCC_c(\mathbf{X}_{s,i}, \mathbf{u}_k)^2 = \\ &\arg \max_{\mu_k} \sum_{\mathbf{x}_{s,i} \in C_k} \left( \max_{\omega} \frac{\mathbf{CC}_\omega(\mathbf{X}_{s,i}, \mathbf{u}_k)}{\sqrt{\mathbf{R}_0(\mathbf{X}_{s,i}, \mathbf{X}_{s,i})} \cdot \sqrt{\mathbf{R}_0(\mathbf{u}_k, \mathbf{u}_k)}} \right)^2 = \\ &\arg \max_{\mu_k} \sum_{\mathbf{x}_{s,j} \in C_k} \left( \sum_{l \in [1, n]} x_{s,l} \mu_{k,l} \right)^2 = \\ &\arg \max_{\mu_k} \sum_{\mathbf{x}_{s,j} \in C_k} (\mathbf{X}_{s,j}^\top \mu_k)^2 = \end{aligned}$$

$$\arg \max_{\mu_k} \mu_k^\top \cdot \sum_{\mathbf{x}_{s,j} \in C_k} (\mathbf{X}_{s,j} \mathbf{X}_{s,j}^\top) \mu_k \quad (17)$$

为了向中心调整和保证聚类中心标准化, 设置  $\mathbf{u}_k = \mathbf{u}_k Q$ , 其中  $Q = I - \frac{\mathbf{O}}{n}$ ,  $I$  为指示矩阵,  $\mathbf{O}$  为元素全为 1 的矩阵。将标准化的  $\mathbf{u}_k$  与  $\mathbf{S} = \sum_{\mathbf{x}_{s,i} \in C_k} (\mathbf{X}_{s,i} \mathbf{X}_{s,i}^\top)$

代入到式(17)中, 得到:

$$\mu_k^* = \arg \max_{\mu_k} \frac{\mu_k^\top Q^\top S Q \mu_k}{\mu_k^\top \cdot \mu_k} = \arg \max_{\mu_k} \frac{\mu_k^\top M \mu_k}{\mu_k^\top \cdot \mu_k} \quad (18)$$

式中:  $M = Q^\top S Q$ 。在文献[14]中, 最优中心  $\mu_k^*$  的求解是实对称矩阵  $M$  的最大特征值对应的特征向量。

## 2.3 基于 Kshape 的多元时间序列聚类

基于 SBD 度量方法与聚类中心提取策略的 Kshape 聚类算法类似于 k-means, 可以更新得到聚类划分较好的聚类中心, 且可以有效地处理时间序列的缩放、平移、转移不变性。Kshape 聚类过程是一个迭代寻优的过程, 主要的步骤为:①时间序列样本分配过程: 根据聚类中心, 分配时间序列  $\mathbf{X}_s$  到距离最近聚类中心; ②聚类中心更新过程。依据上一迭代中聚类关系的变化, 重新计算聚类中心。Kshape 算法重复步骤①与步骤②, 直到聚类关系不再发生变化(各类别样本到聚类中心的 SBD 距离平方和不再发生变化)或者达到最大迭代次数。

## 3 分层聚类时序分析

从一组空战数据中寻找态势时序单元(未知)是一个多元时间序列分割聚类问题。时间序列的分割聚类是一种非监督学习问题, 主要应用于计算机视觉与计算机图形学<sup>[15]</sup>。空战态势描述参数具有时序性, 也具有空间性, 可以分解为多种态势时序单元, 目前面临的问题主要体现在:①态势时序单元组合呈指数增长;②每个相同场景的态势描述参数时间序列不一样, 具有时间的伸缩性;③态势时序单元划分粒度可变性。

针对以上问题, 本文提出了一种分层的聚类时序分析方法(HACA)。该方法结合聚类算法与分割算法, 提出坐标下降法在计算分割段以及其聚类过程中选择求解。

### 3.1 聚类时序分析方法

假设给定多元时间序列  $\mathbf{X}_s = \{\mathbf{X}_{s,1}, \mathbf{X}_{s,2}, \dots, \mathbf{X}_{s,n}\}$ , 其中  $\mathbf{X}_s \in R^{d \times n}$ ,  $n$  为时间序列的时间长度,  $d=8$  表示态势描述参数的维度。将  $\mathbf{X}_s$  分割为  $m_c$  个态势时序单元, 其中每个单元对应  $k$  聚类类别中的某一类。假设第  $i$  个分割态势时序单元  $\mathbf{Y}_i = \mathbf{X}_s$ ,

$[s_i, s_{i+1}) = [x_{s_i}, \dots, x_{s_{i+1}-1} \in R^{d \times n}]$ , 该态势时序单元的开始时刻为  $s_i$ , 结束时刻为  $s_{i+1}-1$ , 其长度  $n_i = s_{i+1} - s_i$  必须满足  $n_i \leq n_{\max}$ , 其中  $n_{\max}$  为态势时序最大长度且控制时间序列分割粒度。为了方便表示与计算每个态势时序单元的所属类别, 则定义聚类指示矩阵  $\mathbf{G} \in \{0,1\}^{k \times m_c}$ , 其中聚类指示矩阵元素  $g_{ci} = 1$  表示  $\mathbf{Y}_i$  属于类别  $c$ , 否则  $g_{ci} = 0$ 。结合第 2 节的时间序列聚类算法 Kshape, 基于聚类时序分析的聚类分割方法目标函数为:

$$\begin{aligned} J_{\text{ACA}}(\mathbf{G}, s) &= \sum_{c=1}^k \sum_{i=1}^{m_c} g_{ci} SBD(\mathbf{Y}_i, \mathbf{u}_c) = \\ &SBD([\mathbf{Y}_1, \dots, \mathbf{Y}_{m_c}], \mathbf{U}_c \mathbf{G}) \\ \text{s. t. } &\mathbf{G}^T \mathbf{1}_k = 1_m \text{ and } s_{i+1} - s_i \in [1, n_{\max}] \end{aligned} \quad (19)$$

式中:  $\mathbf{G} \in \{0,1\}^{k \times m_c}$  为聚类类别指示器,  $s \in R^{m_c+1}$  为态势时序单元的分割起始点与末端点时刻,  $\mathbf{Y}_i = \mathbf{X}_{s_i, [s_i, s_{i+1}]}^T$  表示第  $i$  个分割段(态势时序单元),  $\mathbf{U}_c = \{u_1, u_2, \dots, u_k\}$  为聚类中心集合。本文采用的多元时间序列相似性度量方法为基于形状距离(SBD)的度量方法。

式(19)是一个关于  $\mathbf{G}$  和  $s$  整数规划问题。根据上述的定义,  $\mathbf{G}$  表示态势时序单元与聚类中心的指示矩阵,  $s$  表示态势描述参数点与态势时序单元之间的对应关系, 优化关于  $\mathbf{G}$  和  $s$  的聚类分割方法目标函数是一个非线性多项式 NP 难问题。为了解决该问题, 本文采用坐标下降法框架, 利用动态规划策略来优化计算  $s$ , 利用 winner-take-all 策略来优化计算  $\mathbf{G}$ 。每一迭代次数具体优化如下表示:

$$\begin{aligned} \mathbf{G}, s &= \arg \min_{\mathbf{G}, s} J_{\text{ACA}}(\mathbf{G}, s) = \\ &\arg \min_{\mathbf{G}, s} \sum_{c=1}^k \sum_{i=1}^{m_c} g_{ci} SBD(\mathbf{X}_i, \mathbf{u}_c) \end{aligned} \quad (20)$$

式中:  $\mathbf{u}_c$  是类别为  $c$  的聚类中心, 该聚类中心可以由基于 SBD 的聚类中心提取策略, 由上一迭代次数的解  $(\mathbf{G}', s')$  得到。式(20)的计算复杂度随着原始时间序列的增长而增加, 因此, 本文采用了一种基于动态规划的算法去检测可能的态势时序单元。为了进一步提高求解  $(\mathbf{G}, s)$  的效率, 引入一种辅助函数:

$$J(v) = \min_{\mathbf{G}, s} J_{\text{ACA}}(\mathbf{G}, s) |_{\mathbf{X}_{s, [1, v]}} \quad (21)$$

式中:  $\mathbf{X}_{s, [1, v]} = [x_1, x_2, \dots, x_v]$  表示末端时刻为  $v$  的态势时序单元, 通过优化求解  $v$  来替代优化求解  $s$ 。利用动态规划的思想, 可以将式(21)改写为:

$$J(v) = \min_{1 \leq i \leq v} (J(i-1) + \min_{\mathbf{G}, s} J_{\text{ACA}}(\mathbf{G}, s) |_{\mathbf{X}_{s, [1, i]}}) \quad (22)$$

式中: 时间序列  $\mathbf{X}_s$  被划分为  $\mathbf{X}_{s, [1, i-1]}$  与  $\mathbf{X}_{s, [i, v]}$ , 优化公式(22)使其最小化, 得到最优的子序列  $\mathbf{X}_{s, [1, i-1]}$  与  $\mathbf{X}_{s, [i, v]}$ 。尽管分割时间序列  $\mathbf{X}_s$  的可能情况随着

时间序列的时间长度增加  $n$ , 但动态规划方法利用 Bellman 方程解决最优问题仍然是一个有效的方法。

$$\begin{aligned} \mathbf{J}(v) &= \\ &\min_{v=n_{\max} \leq i \leq v} \{ \mathbf{J}(i-1) + \min_g \sum_{c=1}^k g_{ci} SBD(\mathbf{X}_{s, [i, v]}, \mathbf{u}_c) \} \end{aligned} \quad (23)$$

式中:  $\mathbf{u}_c$  为上一迭代次数类别为  $c$  的聚类中心。优化得到的  $i_v^*, g_v^* = \arg \min_{i, g} \mathbf{J}(v)$  分别是最后一个态势时序单元的开始位置与类别。当  $n_{\max} = 1$  时, 态势分割聚类变成了点聚类问题。基于动态规划的分割方法每一次迭代分为两步:

**步骤 1** 前向过程。前向过程主要是从  $v=1$  标置开始扫描时间序列直到  $v=n$ 。对于每个  $v$ , 根据式(23)计算  $\mathbf{J}(v)$ 。对于  $[i-n_{\max}, v-1]$  中每个时间点, 计算态势时序单元  $\mathbf{X}_{s, [i, v]}$  与上一迭代中聚类中心之间的 SBD 距离。在该过程中, 要储存最优的开始位置  $i_v^*$  与该态势时序单元所对应的类别  $g_v^*$  以及其目标函数值  $\mathbf{J}(v)$ 。

**步骤 2** 后向过程。根据**步骤 1**中得到的最优分割点位置与聚类类别, 从右向左依次划分, 通过计算得到  $s$  与  $\mathbf{G}$ 。

重复**步骤 1**与**步骤 2**直到  $J(n)$  收敛不变或者迭代次数达到最大。

### 3.2 分层策略

为了减少聚类时序分析方法(ant colony algorithm, ACA)的计算复杂度, 本文提出了一种基于分层策略的 ACA 方法, 见图 4。该方法将态势描述参数从不同的等级以不同的时间规模进行分割聚类。

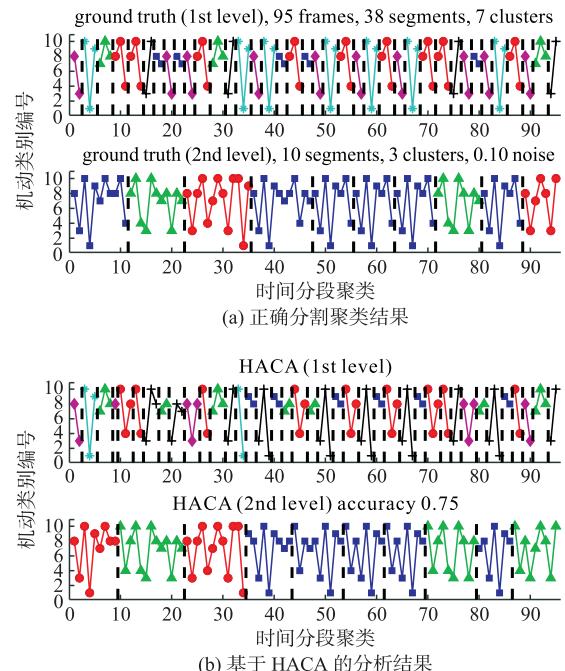


图 4 分层聚类时序分析结果对比

基于分层策略的 ACA 方法(hierarchical ACA, HACA)将分割聚类过程分为两种不同的等级,每一种等级拥有一组时序分割长度方位。设置第 1 层等级的时序分割长度最大值为  $n_{\max}^{(1)}$ ; HACA 设置第 2 层等级的时序分割长度最大值为  $n_{\max}^{(2)}$ ,为了准确提取具有战术意义的态势时序单元,该值相对于  $n_{\max}^{(1)}$  比较大。图 4 对某一个时间序列进行基于 HACA 的分割聚类,得到第 1 层分割聚类结果与第 2 层分割聚类结果,第 2 层分割结果相对来说更加准确。ACA 方法的计算复杂度是  $O(n^2 n_{\max})$ ,HACA 方法的第 2 层计算复杂度是  $O(n^2 n_{\max}^{(2)})$ ,第 1 层计算复杂度是  $O(n^2 n_{\max}^{(1)})$ ,由于  $n_{\max}^{(1)} = n_{\max}$  和  $n_{\max}^{(2)} < n_{\max}$ ,因此,后者计算复杂度较小。

#### 4 基于 L-Kshape-HACA 的分层聚类时序分析仿真

结合基于拉普拉斯中心性的聚类个数确定方法与基于 Kshape 的多元时间序列聚类算法,本文提出了一种基于 L-Kshape 的分层聚类时序分析方法(L-Kshape-HACA),用于解决长时域态势描述参数序列的分割聚类问题。L-Kshape-HACA 分割聚类方法流程图如图 5 所示,其中迭代停止标准均为  $J(n)$  不再发生变化或者达到最大迭代次数。

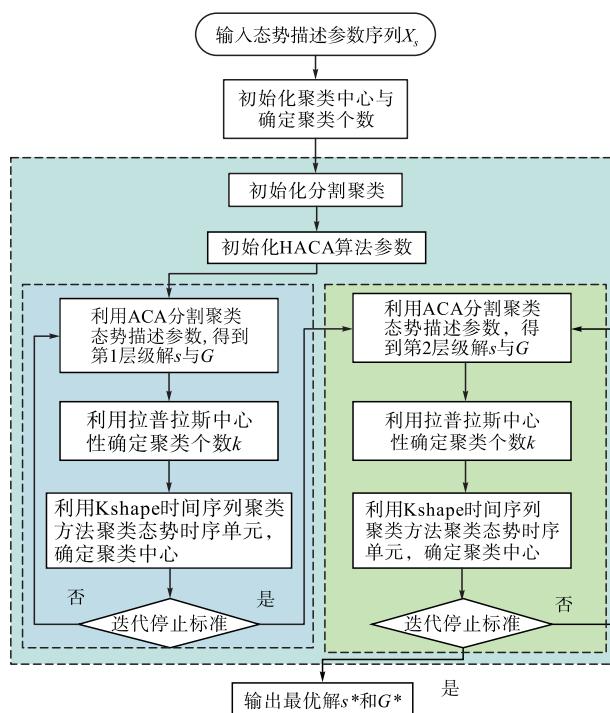


图 5 基于 L-Kshape-HACA 的分层聚类流程图

态势聚类的有效性是实现正确的近距空战决策的数据基础。为了验证 L-Kshape-HACA 分割聚类方法解决近距空战态势描述参数态势时序单元提取

问题的可行性与有效性,利用基于贝叶斯推理的态势评估方法<sup>[16]</sup>、基于多维点数据的聚类算法(聚类数据由多维空间中点数据组成)、基于模型表示的时间序列聚类算法和于多元时间序列数据的分割聚类算法与 L-Kshape-HACA 分割聚类算法进行比较。为了对比基于数据驱动的态势聚类分析与基于推理的态势评估,本仿真实验的对比算法包括文献[16]中基于贝叶斯推理的态势评估方法;为了对比基于时序数据的态势聚类与基于点数据的态势聚类,本仿真实验比较的基于点数据态势聚类算法还包括自组织特征映射神经网络(self-organizing feature map, SOM)<sup>[17]</sup>、层次聚类(hierachical clustering, HC)<sup>[18]</sup>、模糊 C 均值的态势聚类<sup>[19]</sup>、高斯混合的态势聚类<sup>[20]</sup>、k-means、DBSCAN 以及文献[9]中基于密度峰值的态势聚类;为了比较基于分割聚类的时序态势知识提取方法的好坏,本仿真实验比较的分割聚类方法包括 ACA<sup>[8]</sup>、基于谱聚类的分层 ACA(HACA)<sup>[16]</sup>、基于 k-means 的分层 ACA(k-means-HACA)以及基于密度峰值的分层 ACA(DPC-HACA);除此之外,为了测试基于拉普拉斯中心性聚类个数的确定方法是否有效,则需要比较基于 Kshape 的分层 ACA(Kshape-HACA)。本文仿真实验数据来源于某近距空战模拟训练的对抗数据,该实验选取其中的 12 组近距空战对抗数据,其皆为红方(我机)获取胜利的数据,实验仿真环境为 Windows 10,CPU 为 2.80 GHz,8 GB 内存,编程语言 Matlab。

为了找到合适的 L-Kshape-HACA 分割参数,首先从时序分割长度最大值  $n_{\max}^{(1)}$  和  $n_{\max}^{(2)}$  入手,分析了不同的各层级时序分割长度最大值对分层聚类时序方法的影响。在本仿真实验中,选取上述 12 组近距空战对抗数据作为态势描述参数的原始数据,通过特征的选取,得到分割聚类需要的态势描述参数时序数据,见表 2。

表 2 L-Kshape-HACA 不同时序分割最大值设置

算法	$n_{\max}^{(1)}$	$n_{\max}^{(2)}$
L-Kshape-HACA-3-4	3	4
L-Kshape-HACA-3-5	3	5
L-Kshape-HACA-3-6	3	6
L-Kshape-HACA-3-7	3	7
L-Kshape-HACA-4-5	4	5
L-Kshape-HACA-4-6	4	6
L-Kshape-HACA-4-7	4	7
L-Kshape-HACA-4-8	4	8

设定参数下的 L-Kshape-HACA 对 12 组数据进行分割聚类运行 20 次,统计 20 次平均目标函数值与平均时间。不同时序分割长度最大值对 L-

Kshape-HACA 分割聚类算法的影响结果见图 6, 第一等级参数  $n_{\max}^{(1)} = 4$  时的目标函数值优于  $n_{\max}^{(1)} = 3$  的目标函数值, 前者的聚类运行时间少于后者。 $n_{\max}^{(1)} = 4$  时, 第二等级参数  $n_{\max}^{(2)} = 6, 7, 8$  的目标函数值是较好的, 其聚类时间是较少的。由此可见, 第一等级时序分割最大长度越大, 则分割聚类效果越好, 运行时间越短; 第二等级时序分割最大长度越大, 则分割聚类效果越好, 运行时间越短, 但是随着长度达到一定值后, 则效果增加不明显。因此, 本文选取的时序分割长度最大值分别为  $n_{\max}^{(1)} = 4$  和  $n_{\max}^{(2)} = 6$ 。

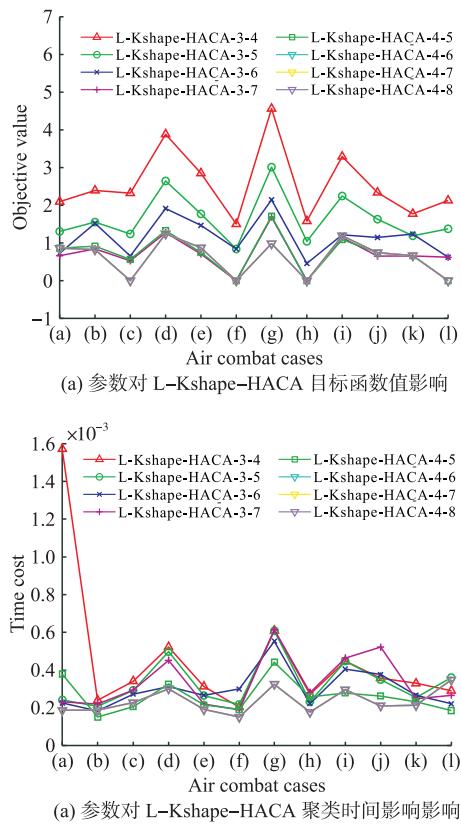


图 6 不同时序分割长度最大值的影响

基于 L-Kshape-HACA 分割聚类方法与比较算法的仿真测试结果如图 7 和图 8 所示。

本文选取了 4 组空战对抗数据进行了对比, 图 7 给出了基于 L-Kshape-HACA 分割聚类算法与其他算法的时间序列分割聚类示意图, 不同的颜色表示不同的类别, 可以看出, 与基于先验知识贝叶斯推理的态势评估结果相比, 基于 L-Kshape-HACA 的态势描述参数序列分割聚类结果是最接近的。图 8 给出了态势描述参数归一化之后的特征空间, 可以看出, 整个态势描述空间可以聚类为 4 类, 态势描述参数序列可以分割为 9 个态势时序单元。为了表达这种抽象的态势时序单元, 结合基于贝叶斯推理的态势评估结果, 根据图 8 所示的对抗轨迹分割聚类结果, 得到以下态势时序单元特征: 1) Cluster #1: 我机其他状态-互不安全的转换; 2) Cluster #2: 我

机中立互为安全-守势-中立相向状态的转换; 3) Cluster #3: 我机攻势-互为安全状态的转换; 4) Cluster #4: 我机互不安全-攻势状态的转换。



图 7 L-Kshape-HACA 与比较算法的分割聚类结果对比

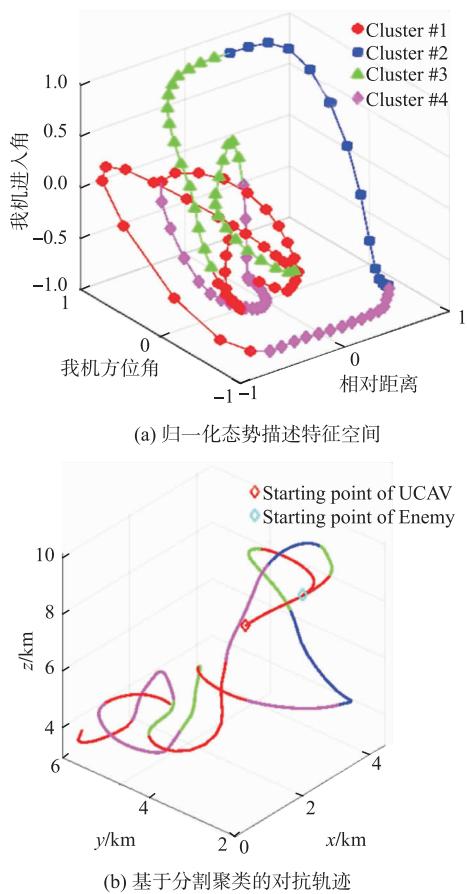


图8 Case(2)空战对抗组的态势时序单元分析

## 5 结语

无人自主空战是未来智能作战的关键技术,本文设计了一种基于数据的无监督态势信息提取新方法,为无人空战决策提供了更加客观可靠的数据基础。本文研究的泛化分析算法会在态势提取过程中忽略飞机独有的战术机动性能,因此,下一步研究必须要针对特定机型的空战数据进行研究,从而提高战术机动态势提取的准确度。

## 参考文献

- [1] 黄长强,唐上钦.从“阿法狗”到“阿法鹰”——论无人作战飞机智能自主空战技术[J].指挥与控制学报,2016,2(3):261-264.
- [2] PAPADIMITRIOUS S, SUN J, FALOUTSOS C. Streaming Pattern Discovery in Multiple Time Series [C]// Proceedings of the 31st International Conference on Very Large Data Bases. Trondlhein: VLDB Endowment, 2005: 697-708.
- [3] 叶燕清.多元时间序列数据挖掘相似性分析方法及应用研究[D].长沙:国防科技大学,2015.
- [4] XIONG Y, YEUNG D T. Time Series Clustering with ARMA Mixtures[J]. Pattern Recognition, 2004, 37

- (8): 1675-1689.
- [5] PIERPAOLO D U, LIVIA D G, MARTA D, et al. Fuzzy Clustering with Spatial-Temporal Information [J]. Spatial Statistics, 2019, 30:71-102.
- [6] 贾镇泽,樊晓光,薛明浩,等.基于机动动作元的敌机战术机动在线识别方法[J].北京理工大学学报,2018,38(8): 820-827.
- [7] 徐西蒙,杨任农,于洋,等.基于运动分解和H-SVM的空战目标机动识别[J].控制与决策,2020,35(5): 1265-1272.
- [8] ZHOU F, TORRE F D L, HODGINS J K. Aligned Cluster Analysis for Temporal Segmentation of Human Motion[C]//2008 8th IEEE International Conference on Automatic Face & Gesture Recognition. [s. l.]: IEEE. 2009: 1-7.
- [9] 张卓然.无人作战飞机自主近距空战关键技术研究[D].西安:空军工程大学,2019.
- [10] YANG X. H, ZHU Q P, HUANG Y J, et al. Parameter-Free Laplacian Centrality Peaks Clustering[J]. Pattern Recognition Letters, 2017, 100: 167-173.
- [11] 邱保志,程柰.基于拉普拉斯中心性和密度峰值的无参数聚类算法[J].计算机应用,2018,38(9): 2511-2514, 2524.
- [12] BERNDT D J, CLJORD J. Using Dynamic Time Warping To Find Patterns in Time Series[C]. Proc of the Workshop on Knowledge Discovery in Databases. Seattle:[s. n.], 1994:229-248.
- [13] 李正欣,郭建胜,毛红保,等.多元时间序列相似性度量方法[J].控制与决策,2017,32(2):368-372.
- [14] JOHN P, LUIS G. K-Shape: Efficient and Accurate Clustering of Time Series [C]// SIGMOD'15, ACM, 2015: 1855-1870.
- [15] FENG Z, FERNANDO D L T, JESSICA K H. Hierarchical Aligned Cluster Analysis for Temporal Clustering of Human Motion[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(3): 582-595.
- [16] HUANG C Q, DONG K S, HUANG H Q, et al. Autonomous Air Combat Maneuver Decision Using Bayesian Inference and Moving Horizon Optimization [J]. Journal of Systems Engineering and Electronics, 2018, 29(1): 86-97.
- [17] 王晓燕. K-均值算法与自组织神经网络算法的改进研究及应用[D].太原:中北大学,2017.
- [18] 黎新华,李俊辉,黎景壮.基于改进DTW-AGNES的网约车需求量时间序列聚类研究[J].重庆交通大学学报(自然科学版),2019,38(8): 13-18.
- [19] 陈梅.面向复杂数据的聚类算法研究[D].兰州:兰州大学,2016.
- [20] 周志华.机器学习[M].北京:清华大学出版社,2016.

(编辑:徐敏)