

# 基于时空聚类的在制品流转状态分析

王益聪<sup>1</sup>,郭宇<sup>1</sup>,黄少华<sup>1</sup>,张蓉<sup>1</sup>,冯上海<sup>2</sup>

(1. 南京航空航天大学 机电学院,江苏 南京 210016;

2. 中航工业江西洪都航空工业集团有限公司,江西 南昌 330024)

**摘要:**针对离散制造车间制造过程监控的需求,提出一种基于时空聚类的在制品流转状态分析方法。通过改进的动态密度网格聚类方法对车间在制品历史位置数据进行分析处理,获取在制品时空轨迹模型;并通过计算实时数据与模型的改进 Hausdorff 距离来判断在制品流转状态。以某机加工车间为例,验证该方法的可行性和有效性。

**关键词:**离散制造;位置数据;聚类;Hausdorff 距离;状态分析

**中图分类号:**TP391 **文献标志码:**A **文章编号:**1671-5276(2020)02-0012-05

## Analysis of Work-in-Process Logistics State Based on Spatiotemporal Clustering

WANG Yicong<sup>1</sup>, GUO Yu<sup>1</sup>, HUANG Shaohua<sup>1</sup>, ZHANG Rong<sup>1</sup>, FENG Shanghai<sup>2</sup>

(1. College of Mechanical and Electrical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China; 2. AVIC Jiangxi Hongdu Aviation Industry Group Company Ltd., Nanchang 330024, China)

**Abstract:** Aiming at the demands for manufacturing process monitoring in the discrete manufacturing workshops, an analysis method of Work-in-Process logistics state based on spatiotemporal clustering is proposed. The historical location data collected in the manufacturing process is processed by the modified grid-based clustering method, and the spatio-temporal model of Work-in-Process is established, then the real-time data and the modified Hausdorff distance of the model are calculated for judging the logistics state of Work-in-Process. The machining workshop is taken for example, and the feasibility and effectiveness of this method are verified.

**Keywords:** discrete manufacturing; location data; clustering; Hausdorff distance; state analysis

## 0 引言

离散制造车间中,制造过程根据产品加工工艺分为多道工序,在制品需要在多个工位间流转<sup>[1]</sup>。通过车间实时定位系统,能够有效获取在制品实时位置信息。对在制品位置数据的挖掘能够提取在制品在车间流转的频繁时空路径,通过计算在制品流转过程中实时数据和挖掘的频繁路径的偏差,判断在制品流转状态是否发生异常。

随着无线通信技术的发展,室内定位技术在车间的应用越来越成熟。文献[2]提出了一种新的基于射频识别(radio frequency identification, RFID)和 WiFi 技术的实时定位系统,实现仓库的自动化、数字化和智能化管理,大大提高了仓库运行效率,降低了仓库管理成本;文献[3]研究了实时定位技术对生产调度的影响,工艺员可以通过定位系统提供的位置数据流实时跟踪在制品的生产状态,进而掌握并合理安排每个工位的加工时间,实现生产资源的动态调度和规划;文献[4]将基于 RFID 技术的实时定位系统应用到半导体制造车间中,研究了在车间生产流程中

通过实时定位技术获取的不同生产过程对象位置信息的价值,表明实时定位的应用提供了更好的可视化水平以及更加高效的生产效率。现代观测技术、计算机网络和地理信息系统的快速发展,生成大量时空数据,对时空数据分析方法的研究日益增多。文献[5]提出了基于时空对象的聚类方法,有助于全面分析时空对象空间位置、属性特征及其变化特点,为多粒度时空对象分析提供思路;文献[6]利用个体出行的 GPS 轨迹数据,在 DBSCAN 的基础上提出一种新的时空聚类算法,以时空邻近条件定义簇间距离,识别 GPS 轨迹中的停驻点;文献[7]对船载 AIS 数据展开时空聚类分析,在 DBSCAN 算法基础上提出船载 AIS 数据时空聚类算法,发现隐含的时空模式,为船舶交通管理提供了一种新途径。

通过分析发现,实时定位在车间主要应用于制造要素的可视化跟踪和监控,而缺少对位置数据和制造过程之间潜在关系的挖掘。本文以室内定位技术为基础,有效获取离散制造车间在制品实时位置数据,实时计算车间现场在制品流转状态偏差,及时发现异常,为制造过程动态优化提供依据。

**基金项目:**国家自然科学基金资助项目(51575274);“十三五”国防基础科研项目(JCKY2016605B006);“十三五”国防基础科研项目(JCKY2016203B083)

**第一作者简介:**王益聪(1993—),男,江苏连云港人,硕士研究生,研究方向为制造物联网技术。

## 1 在制品流转状态分析

在制品的时空数据是定位系统产生的一连串包含时空信息的位置点组成的集合  $P = \{p_1, p_2, \dots, p_n\}$ , 每个点包含时空坐标  $p_i = (x_i, y_i, t_i)$ 。通过对在制品时空数据的聚类分析, 可以发现一些隐含的流转模式和规律, 获取在制品流转过程时空模型, 然后通过计算在制品实时位置数据与聚类结果的相似度, 来分析在制品流转状态, 其中在制品流转状态集合为 {正常, 预警, 异常}。技术路线如图 1 所示。

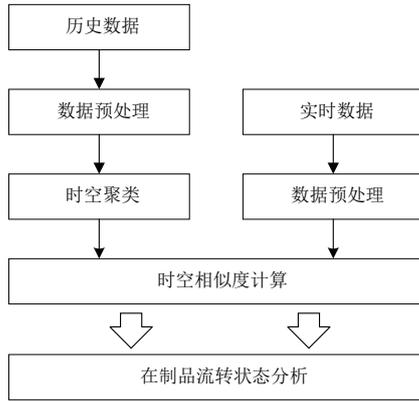


图 1 在制品流转状态分析技术路线

## 2 在制品流转数据时空聚类方法

### 2.1 在制品时空轨迹定义

在制品时空轨迹是在空间维度和时间维度上的运动轨迹, 表达了在制品空间位置随时间的变化。时空轨迹由一系列时空记录组成, 每条记录包括在制品的位置信息和记录时间。

实时定位系统产生的位置数据记录主要包括移动目标  $o$ 、位置数据  $(x, y)$  和时间  $t$ 。对于在制品  $o_i$ , 将其时空轨迹定义为  $T_i = \{(x_1, y_1, t_1), (x_2, y_2, t_2), \dots, (x_n, y_n, t_n)\}$ , 如图 2 所示。

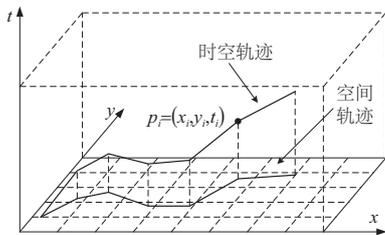


图 2 在制品时空轨迹

在制品时空轨迹包含了丰富的生产过程信息。仅考虑空间轨迹, 可以获取在制品在车间哪些区域、工位间流转, 即在制品工艺路线; 结合时间信息, 能够反映在制品在何时到达哪个工位, 加工了多长时间等。

### 2.2 基于网格的在制品时空轨迹聚类算法

网格聚类算法的原理是将数据空间划分为网格单元,

将数据集映射到网格单元中。这种方法的优点是处理速度很快, 其处理时间独立于数据对象数目, 只与划分的单元数目有关, 是空间数据处理中常用的将数据离散化的方法<sup>[8]</sup>。针对离散制造车间位置数据的特点, 采用网格聚类算法对在制品时空数据进行处理。

传统意义上的轨迹通过位置点序列来表达, 这给存储和处理带来了困难, 而网格聚类将轨迹的位置点映射到划分的网格单元中, 极大减小了存储空间和处理的复杂度。如图 3(a) 所示, 轨迹由一系列带有  $x, y$  坐标的特征点表达, 即  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ 。而在网格表达形式中, 同一条轨迹表达如图 3(b) 所示, 将轨迹按经过的网格单元表达, 用网格单元的行列索引号来表示, 即  $\{(3, 1), (3, 2), (3, 3), \dots, (6, 8), (6, 9)\}$ 。网格表达形式具有数据结构简单、定位存取性能好的特点。时空轨迹则由三维空间网格单元序列来表示。

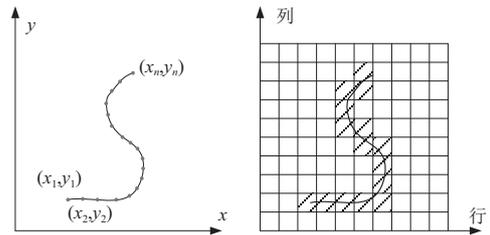


图 3 轨迹的两种表达形式

图 3 轨迹的两种表达形式

网格聚类算法核心思想如下: 将数据集的每一个维度划分成网格单元, 扫描所有网格, 如果一个网格单元中包含的数据点数超过了给定的密度阈值, 则称该单元是密集的<sup>[9]</sup>。这些连通的密集单元的最大集合就是簇的定义。算法的相关定义如下:

**定义 1 (网格单元)** 对于在制品时空数据集  $P = \{p_1, p_2, \dots, p_n\}$ , 通过设置网格步长  $\lambda_i$ , 对车间物理空间和时间维度进行划分, 将数据的每一维均匀划分为  $k_i$  部分, 整个车间时空域就被划分为有限个类矩形单元, 每一个类矩形单元可以描述为  $G_i = (g_i^x, g_i^y, g_i^t)$ , 称为网格单元, 其中  $1 \leq g_i \leq k_i$ , 为网格单元各个维度的索引。

**定义 2 (网格单元密度)** 当 1 条数据  $p_i \in G_j$ , 即  $(x_i, y_i, t_i) \in G_j$ , 则称该条数据属于网格单元  $G_j$ 。网格单元  $G_j$  中包含的数据总数称为  $G_j$  的网格单元密度  $\text{density}(j)$ 。

**定义 3 (密集网格单元、稀疏网格单元)** 设置密度阈值  $\tau$ , 对于 1 个网格单元  $G_j$ , 当  $\text{density}(j) > \tau$  时, 称该网格单元为密集网格单元, 否则为稀疏网格单元。

**定义 4 (邻接网格单元)** 对于 2 个网格单元  $G_i = (g_i^x, g_i^y, g_i^t)$  和  $G_j = (g_j^x, g_j^y, g_j^t)$ , 当满足式 (1) 时, 称 2 个网格单元是邻接的。

$$\begin{cases} |g_i^l - g_j^l| = 1 & r = x, y, t \\ g_i^l = g_j^l & l = x, y, t \text{ 且 } l \neq r \end{cases} \quad (1)$$

**定义 5 (网格单元密度相连)** 对于 2 个邻接网格单元  $G_i = (g_i^x, g_i^y, g_i^t)$  和  $G_j = (g_j^x, g_j^y, g_j^t)$ , 当  $G_i$  和  $G_j$  都是密集网格单元时, 称  $G_i$  和  $G_j$  密度相连。

定义6(簇) 数据空间中最大密度相连的网格单元的集合称为1个簇 $C_i$ 。

## 2.3 算法流程

网格聚类算法主要是对车间采集的在制品历史数据进行分析处理,得到的结果作为在制品流转状态实时分析的依据。具体流程如图4所示。

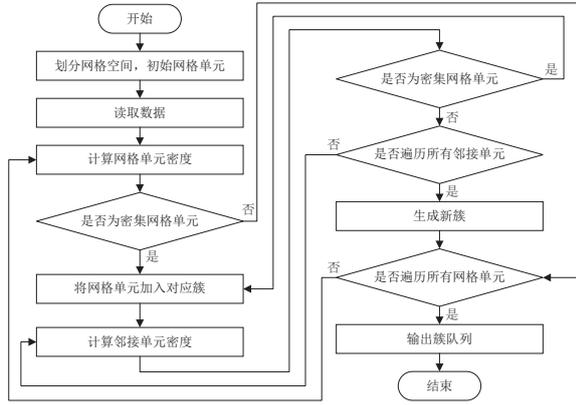


图4 算法流程图

算法流程详细步骤如下:

步骤1:根据车间布局和生产过程的特点,确定数据空间每一维的网格划分参数 $\lambda_x, \lambda_y, \lambda_t$ ,初始化网格单元状态。

步骤2:读取数据,更新网格单元状态。

步骤3:根据网格单元密度判断是否为密集网格单元,若是密集网格单元,则跳至步骤4,否则跳至步骤7。

步骤4:将密集网格单元加入相应的簇中。

步骤5:判断密集网格单元的邻接网格是否密集,若是,则跳至步骤4,若否,跳至步骤6。

步骤6:判断邻接单元是否遍历完成,若否,跳至步骤5,若遍历完成则生成1个簇。

步骤7:判断网格单元是否遍历完成,遍历完成则输出簇队列,否则跳至步骤3。

算法输出的结果是一系列簇 $C_1, C_2, \dots, C_n$ ,每个簇为1个网格序列,对在制品的时空轨迹。

## 3 基于改进 Hausdorff 距离的时空相似度计算

### 3.1 改进 Hausdorff 距离

在制品流转状态异常检测是对制造过程中实时采集的数据进行处理,通过计算实时流转轨迹和聚类结果的相似度来分析在制品的流转状态。时空轨迹相似度计算主要依赖于轨迹之间距离的定义以及轨迹之间的匹配程度。Hausdorff 距离是描述两点集之间相似程度的一种度量方法。给定两点集 $A = \{a_1, a_2, \dots\}$ ,  $B = \{b_1, b_2, \dots\}$ ,两个点集之间的 Hausdorff 距离为:

$$H(A, B) = \max [h(A, B), h(B, A)] \quad (2)$$

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (3)$$

$$h(B, A) = \max_{b \in B} \min_{a \in A} \|b - a\| \quad (4)$$

其中: $\|a - b\|$ 是点 $a$ 和 $b$ 之间的距离度量; $H(A, B)$ 称为双向 Hausdorff 距离; $h(A, B)$ 称为点集 $A$ 到点集 $B$ 的单向 Hausdorff 距离; $h(B, A)$ 称为点集 $B$ 到点集 $A$ 的单向 Hausdorff 距离。双向 Hausdorff 距离是2个单向 Hausdorff 距离中的较大者,它度量了2个点集之间的最大不匹配程度<sup>[10]</sup>。

传统的 Hausdorff 距离中 $\|a - b\|$ 多采用欧式距离,而对于时空轨迹点之间的距离需要同时考虑时间和空间,因此给出时空距离的定义。

给定1条轨迹 $T' = \{p'_1, p'_2, \dots, p'_i\}$ 和1条待匹配轨迹 $T = \{p_1, p_2, \dots, p_i\}$ ,其中 $p_i = (x_i, y_i, t_i)$ , $T$ 上一点 $p_i$ 到轨迹 $T'$ 的空间距离、时间距离和时空距离定义如下:轨迹 $T$ 上任意一点 $p_i$ 到轨迹 $T'$ 的空间距离 $d_s(p_i, T') = \min_{p'_j \in T'} \|p_i - p'_j\|$ ,其中 $\|p_i - p'_j\| = \sqrt{(x_i - x'_j)^2 + (y_i - y'_j)^2}$ 。对于 $T$ 中任意一点 $p_i$ ,在 $T'$ 中有空间距离最近的点 $p'_j$ ,则时间距离 $d_t(p_i, T') = |t_i - t'_j + \theta|$ ,其中 $\theta$ 是为了避免 $d_t = 0$ 设置的一个时间常数。将空间距离和时间距离归一化后得到:

$$D_s(p_i, T') = \frac{1}{1 + e^{-d_s(p_i, T')}} \quad (5)$$

$$D_t(p_i, T') = \frac{1}{t_n - t_1} \times d_t(p_i, T') \quad (6)$$

结合时间距离和空间距离,定义时空距离为:

$$D_{ST}(p_i, T') = D_s(p_i, T') \times D_t(p_i, T') \quad (7)$$

在制品时空轨迹相似度的计算主要是计算实时轨迹和聚类结果之间的 Hausdorff 距离,因此只计算实时数据到聚类结果的单向 Hausdorff 距离,最后改进的 Hausdorff 距离为:

$$H_m(T, T') = \max_{p_i \in T} D_{ST}(p_i, T') \quad (8)$$

由上面的距离定义可知,改进的 Hausdorff 距离取值范围在0~1之间,且 $H_m(T, T')$ 越小,轨迹 $T$ 和 $T'$ 的时空相似度越高。 $H_m(T, T') \leq H_1$ 时,在制品流转为正常状态; $H_1 < H_m(T, T') \leq H_2$ 时,为预警状态; $H_m(T, T') > H_2$ 时,为异常状态。其中, $H_1, H_2$ 为根据在制品实际流转情况得到的状态变化的 Hausdorff 距离临界值, $0 < H_1 < H_2 < 1$ 。

### 3.2 基于时间约束的时空相似度计算

已有的 Hausdorff 距离计算相似度考察的是两个无序的集合,而时空轨迹数据是有序的。因此采用基于时间约束的时空相似度计算方法。

在离散制造车间生产过程中,采集的位置数据是一系列大量、连续到达的流数据。在实时的时空相似度计算过程中,如果每到来一条新的数据就计算相似度,会造成算法复杂度的增加,而且短时少量的数据对于在制品流转轨迹的影响不明显。因此,设定一个时间间隔 $t_{gap}$ ,采用滑动时间窗模式更新相似度计算结果,即计算 $t_{now} - t_{gap}$ 时刻到当前时刻 $t_{now}$ 时间段内的实时轨迹 $T$ 与聚类获取的轨迹 $T'$ 中 $t_{now} - 2t_{gap}$ 时刻到 $t_{now} + t_{gap}$ 时间段内子轨迹的改进 Hausdorff 距离,如图5所示。

实际计算中,轨迹被转化成网格序列,对于轨迹 $T$ 中每

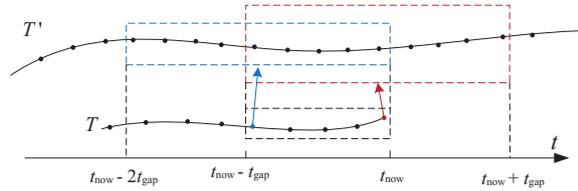


图5 滑动时间窗模式

一个网格单元  $G_i = (g_i^x, g_i^y, g_i^t)$ , 在轨迹  $T$  中都有对应时间的网格单元  $G_j = (g_j^x, g_j^y, g_j^t)$ 。计算轨迹  $T$  中每个网格单元  $G_i$  到轨迹  $T$  中对应网格单元  $G_j$  前后  $2t_{gap}$  时间段内子轨迹  $T'$  的时空距离  $D_{ST}(G_i, T')$ , 然后按式(8)计算 Hausdorff 距离  $H_m(T, T')$  作为当前时刻的轨迹时空相似度。

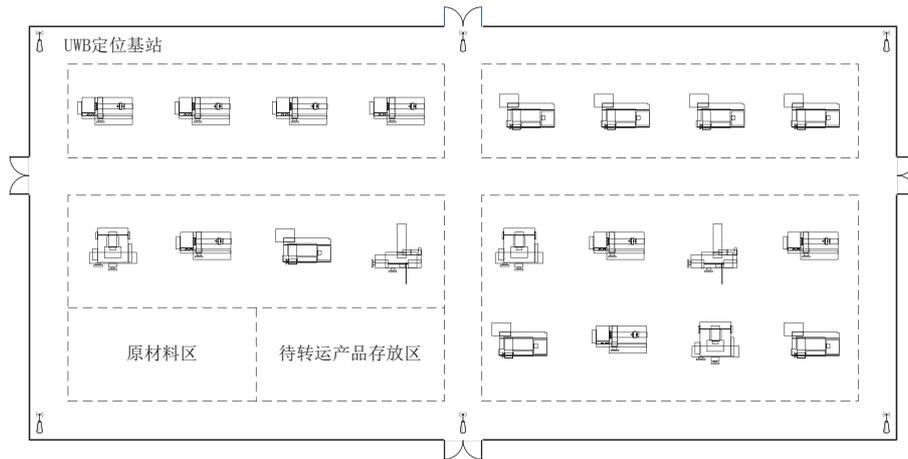


图6 车间布局

车间正常生产过程中采集到的每条位置数据包括: ID、采集时间和坐标信息, 数据采集频率为 4 Hz。由于原始采集的数据中时间记录为绝对时间, 无法直接对时间进行划分。在对原始数据进行聚类之前, 先进行预处理: 将在制品离开原材料区的时刻记为  $t_0$ , 则处理后的数据为  $(ID, x, y, t-t_0)$ 。

车间采用的 UWB 定位系统动态定位误差 90% 的情况下  $< 30\text{ cm}$ , 约 99% 的情况下定位误差  $< 60\text{ cm}$ 。同时考虑加工区、缓冲区和通道等区域的尺寸, 避免同一网格跨越不同功能区域, 设置车间空间网格划分粒度为  $0.5\text{ m} \times 0.5\text{ m}$ 。为保证一定的实时性以及尽可能降低算法复杂度, 设置时间维度网格步长  $\lambda = 1\text{ min}$ 。网格聚类的关键参数密度阈值  $\tau$  对聚类结果影响较大, 实验结果表明,  $\tau$  值偏大会导致本属于同一个簇的网格被分到两个簇, 容易丢失在制品流转过程信息;  $\tau$  值偏小会将不同簇的网格聚到一起, 同时会受到异常数据和稀疏数据的影响, 导致聚类结果失真, 同时增加了算法的时间复杂度。对于不同的参数  $\tau$  取值, 算法的正确率和运行时间如图 7 所示。

从图中可以看出, 在密度阈值设置为 80 和 90 时, 算法正确率为 100%, 而算法运行时间随着密度阈值的增大而减小。综合考虑算法正确率和运行时间, 设置密度阈值  $\tau = 90$ , 对在制品历史轨迹数据进行聚类。部分聚类结果如表 1 所示。

## 4 案例分析

以某机加车间为例, 车间部署了 UWB 室内定位系统, 通过定位标签实时获取在制品等制造要素的位置信息, 在 IntelliJ IDEA 平台进行了实验。车间尺寸为  $92\text{ m} \times 44\text{ m}$ , 共有 1 个原材料区、1 个待转运产品存放区和 20 个加工工位。原材料区和待转运产品存放区尺寸约为  $20\text{ m} \times 10\text{ m}$ , 各工位尺寸约为  $10\text{ m} \times 10\text{ m}$ , 其中每个工位包含缓冲区(约  $10\text{ m} \times 2\text{ m}$ )和加工区(约  $10\text{ m} \times 8\text{ m}$ )。物料从原材料区粘贴定位标签, 并配送到各加工工位, 加工完成的产品运送到待转运产品存放区。车间布局如图 6 所示。

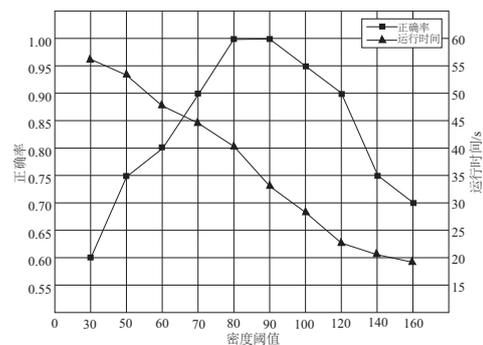


图7 不同参数下算法正确率和运行时间

表1 算法结果

序号	网格
0	(15, 30, 1)
1	(76, 42, 2)
2	(162, 46, 43)
3	(168, 24, 89)
4	(106, 74, 134)
5	(80, 74, 191)
6	(72, 22, 229)

表中列出了在制品停留时间较长的网格,结合车间几何空间和语义空间,聚类结果可以解释为在制品的加工路线详细信息,如表2所示。

表2 在制品加工路线

位置	加工时间/min
原材料区	—
工位 4	41
工位 12	46
工位 8	45
工位 16	57
工位 17	48
待转运产品存放区	—

在制品流转轨迹时空相似度计算过程中,设置时间间隔  $t_{\text{gap}} = 5 \text{ min}$ ,即每 5 min 更新一次相似度计算结果。根据车间布局以及在制品加工用时,取在制品流转状态变化临界值  $H_1 = 0.19, H_2 = 0.42$ 。采用某在制品加工过程中采集的数据作为模拟实时数据输入,改进 Hausdorff 距离计算结果如图 8 所示。

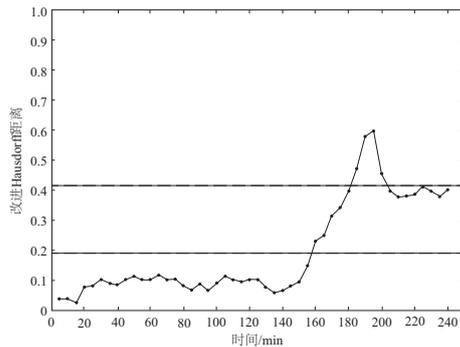


图8 计算结果

从图中的计算结果可以发现,在制品流转状态在 160 min 之前处于正常状态,165 min 时改进 Hausdorff 距离增大,进入预警状态,185 min 时进入异常状态,改进 Hausdorff 距离在 200 min 开始趋于稳定。结合车间实际情况,工位 16 的机床在 165 min 左右时发生故障,导致在制品在工位 16 的缓存区停留时间过长,同时,机床在 195 min 左右恢复正常。由此证明了本文提出方法的有效性。

离散制造车间在制造过程中难免发生异常情况,如设

备故障、物料配送异常、人员离岗等,一旦无法及时发现并处理车间生产异常,将直接影响车间制造活动的正常运行。而通过结合在制品的实时位置数据对在制品流转状态偏差进行实时分析,能够对在制品流转过程中发生的异常进行有效预警,有助于制造过程的正常运行。

## 5 结语

实时定位技术在离散制造车间应用越来越广泛,因此产生了大量的位置数据。本文在离散制造车间实时位置数据的基础上对在制品流转状态分析展开研究。通过位置大数据处理方法建立在制品流转的时空轨迹模型,以此为依据,结合制造过程中实时位置数据计算流转状态偏差。最后,以某机加车间为例进行了验证,结果表明本文提出的方法可以有效分析在制品流转状态。

### 参考文献:

- [1] 蔡浩舒. 基于实时数据的离散制造车间在制品流转状态研究[D]. 南京:南京航空航天大学, 2017.
- [2] BIN D, LI C, DIANLONG C, et al. Application of RTLS in warehouse management based on RFID and Wi-Fi[C]. International Conference on Wireless Communications. IEEE, 2008.
- [3] CHONGWATPOL J, SHARDA R. RFID-enabled track and traceability in job-shop scheduling environment[J]. European Journal of Operational Research, 2013, 227(3): 453-463.
- [4] THIESSE F, FLEISCH E. On the value of location information to lot scheduling in complex manufacturing processes[J]. International Journal of Production Economics, 2008, 112(2): 532-547.
- [5] 龚玺,裴韬,孙嘉,等. 时空轨迹聚类方法研究进展[J]. 地理科学进展, 2011, 30(5): 522-534.
- [6] 周洋,杨超. 基于时空聚类算法的轨迹停驻点识别研究[J]. 交通运输系统工程与信息, 2018, 18(4): 88-95.
- [7] 李永攀,刘正江,郑中义. 基于时空密度的船载 AIS 数据聚类分析方法研究[J]. 重庆交通大学学报(自然科学版), 2018, 37(10): 117-122.
- [8] 窦瑞亭. 基于网格的双向层次聚类算法的研究[D]. 吉林:吉林大学, 2018.
- [9] 王海东. 基于网格方法的数据流聚类算法研究[D]. 西安:西安科技大学, 2012.
- [10] 郑晓璐,潘广贞,杨剑,等. 基于 Hausdorff 距离改进的 ICP 算法[J]. 计算机工程与设计, 2015, 36(9): 2481-2484, 2489.

收稿日期:2019-01-15