第 45 卷第 5 期
2013 年 10 月

南 京 航 空 航 天 大 学 学 报
Journal of Nanjing University of Aeronautics & Astronautics

Vol. 45 No. 5
Oct. 2013

一种基于层次聚类的机场噪声数据挖掘方法

徐 涛[1,2] 谢继文[1] 杨国庆[2]

(1. 中国民航大学计算机科学与技术学院，天津，300300；
2. 中国民航信息技术科研基地，天津，300300)

摘要：针对机场噪声数据集特征，提出一种基于代表点的快速层次聚类算法，该算法在传统凝聚层次聚类算法的基础上，借助聚类代表点法和二分法策略进行改进。同时，提出一种聚类代表点和聚类算法相似性定义相结合的聚类结果评价方法，并采用其对聚类结果进行评价。实验结果表明，该算法不仅运行效率高，而且能够较准确地发现特定类型飞行事件的噪声分布模式，利用该分布模式能够较准确地预测特定类型飞行事件的噪声分布状况。

关键词：数据挖掘；机场噪声预测；代表点；快速层次聚类算法；聚类结果评价

中图分类号：TP18 **文献标志码**：A **文章编号**：1005-2615(2013)05-0715-07

Airport Noise Data Mining Method Based on Hierarchical Clustering

Xu Tao[1,2], *Xie Jiwen*[1], *Yang Guoqing*[2]

(1. College of Computer Science and Technology, Civil Aviation University of China, Tianjin, 300300, China;
2. Information Technology Research Base of Civil Aviation Administration of China, Tianjin, 300300, China)

Abstract: According to the characteristics of airport noise data, a fast hierarchical clustering algorithm based on representative point is presented. This algorithm improves the traditional condensed hierarchical clustering algorithm by using clustering representative point method and dichotomy strategy. Meanwhile, a clustering result evaluation method which combines the clustering representative point and the definition of similarity in clustering algorithm is proposed. The experimental results show that the proposed algorithm not only has high performance, but also can discover the noise distribution model of a specific type of flight event correctly. The method can accuratly predict the noise distribution model of these flight events.

Key words: data mining; prediction of airport noise; representative point; fast hierarchical clustering algorithm; clustering result evaluation

近年来，随着民航业的飞速发展，全国各地正在不断改建、扩建、新增大批机场，机场吞吐能力及占地规模都相应地扩大。与此同时，机场周边地区的城镇化进程也随之加快，机场用地与城镇用地越来越靠近，由机场噪声影响所引起的矛盾、纠纷也越来越多[1]。因此，科学地预测机场周围噪声分布状况，对合理规划机场周围用地具有重要意义。

目前，国内对机场噪声的预测理论和方法尚缺少深入细致的研究，大都沿用国外的理论和方法。国外很早就开展了该领域的研究，并建立了一系列预测模型[2]。然而利用这些预测模型预测机场噪声时，需要根据特定的预测模型计算机场周围每个

基金项目：国家自然科学基金重点(61139002)资助项目；国家高技术研究发展计划("八六三"计划)(2012AA063301)资助项目；中国民用航空局科技项目(MHRD201006，MHRD201101)资助项目；中央高校基本科研业务费专项资金(3122013P013)资助项目。

收稿日期：2012-11-24；**修订日期**：2013-09-08

通信作者：徐涛，男，教授，博士生导师，E-mail：txu@cauc.edu.cn。

位置点的噪声值,使得时间代价较大。针对上述问题,可以采用数据挖掘的方法挖掘机场历史噪声数据集以发现特定类型飞行事件的噪声分布模式及代表性的位置点,然后通过计算代表性位置点的噪声值来预测其他位置点的噪声值,从而缩短机场噪声预测的时间。

数据挖掘[3]是从大量历史数据中提取可信、新颖、有效并能被人们理解的模式,进而发现隐含的、有意义的知识。

聚类是数据挖掘的一个重要研究分支[4]。根据聚类过程所采用的集聚规则,聚类算法大致分为基于划分的聚类算法[5]、基于层次的聚类算法[6]、基于密度的聚类算法[7]及其他聚类算法等,没有任何一种聚类算法可以普遍适用于多种多样的数据集。基于划分的聚类算法的优点是时间复杂度较低,但该方法对聚类数目和聚类中心的选取非常敏感;基于层次的聚类算法只需定义合并或分裂准则,就能合理而有效地进行聚类,其缺点是时间复杂度较大;基于密度的聚类方法适用于具有间隙性、分布不均匀的数据集。上述聚类算法都有一定的自身局限性,导致其不能直接应用于海量机场噪声数据集的聚类分析。因此,本文提出一种基于代表点的快速层次聚类算法,该算法不仅克服了传统层次聚类算法效率低的缺点,而且能够较准确地发现特定类型的飞行事件的噪声分布模式。

1 基本概念及理论

1.1 机场噪声数据的特征

机场噪声[8]是指飞机在起飞、降落和滑行过程中产生的各种噪声。

飞行事件是指同一时刻发生的一架或多架飞机飞行的事件。

机场噪声数据是指飞行事件发生时产生的一个噪声文件,即飞行事件发生时,机场周围每个位置点监测到的噪声值所形成的一个噪声文件。

机场噪声数据集是指若干个飞行事件产生的噪声文件的集合。

机场噪声数据集主要有以下特征:

(1) 数据量大

体现在数据对象(机场周围位置点)数目 n 庞大以及维度(单次飞行事件次数)m 高。

(2) 空间邻域相关性

机场噪声数据中,相邻位置点的噪声值具有高度相关性,即任意一次飞行事件下,每个位置点的噪声值与其周围邻域其他位置点的噪声值相同或相近。

1.2 聚类基本概念及凝聚层次聚类算法

1.2.1 聚类基本概念

聚类[9]指按照给定的相似性定义将物理或抽象对象的集合分成若干个簇,使得同一簇内的对象尽可能相似,不同簇的对象尽可能相异。

为方便机场噪声数据挖掘问题的表述,作如下定义。

聚类对象:机场周围的每一个位置点;

对象属性:每次飞行事件下该位置点的噪声值;

聚类对象间的距离:设位置点 $\boldsymbol{Q}_i(q_{i1},q_{i2},\cdots,q_{im})(i=1,2,\cdots,n)$ 是第 i 个聚类对象,其中 $q_{ik}(k=1,2,\cdots,m)$ 是位置点 $\boldsymbol{Q}_i$ 在第 k 次飞行事件下的噪声值,$d(i,j)$ 表示位置点 $\boldsymbol{Q}_i$ 与位置点 $\boldsymbol{Q}_j$ 的距离,且

$$d(i,j)=\left(\frac{1}{m}\left(\sum_{k=1}^{m}(q_{ik}-q_{jk})^2\right)\right)^{\frac{1}{2}} \tag{1}$$

簇:任意一次飞行事件下其噪声值均十分接近的那些位置点的集合;

簇中心:设 $\boldsymbol{V}_i(v_{i1},v_{i2},\cdots,v_{im})$ 为第 i 个簇的簇中心,其中 $v_{ij}(j=1,2,\cdots,m)$ 为 $\boldsymbol{V}_i$ 的第 j 维(m 为维度),且

$$v_{ij}=\frac{\sum_{k=1}^{n}q_{kj}\times\phi(k,i)}{g_i} \tag{2}$$

式中:n 表示机场周围位置点数目;g_i 为第 i 个簇的对象数目;ϕ 为对象隶属函数,如果第 k 个对象隶属于第 i 个簇,则 $\phi(k,i)$ 为 1,否则为 0,即

$$\phi(k,i)=\begin{cases}1 & \boldsymbol{Q}_k\in C_i\\ 0 & \boldsymbol{Q}_k\notin C_i\end{cases} \tag{3}$$

其中,C_i 表示第 i 个簇。

代表点:用来描述机场周围噪声分布状况的少量位置点。

1.2.2 凝聚层次聚类算法

凝聚层次聚类算法[10]采用自底向上的合并策略:首先设定每一个对象为一个簇;然后进行迭代循环,每次循环寻找距离最小的两个簇,根据这两个簇的距离是否小于给定的聚类约束条件,确定是否将这两个簇进行合并。

2 基于代表点的快速层次聚类算法

机场噪声数据集的高维特性使得聚类数目和聚类中心的确定变得更加困难,因此基于划分的聚类算法无法直接应用于机场噪声数据集的聚类分

析。

机场噪声数据的空间邻域相关性特征使得机场噪声数据具有连续性、无间隙性等特性，因此无法采用基于密度的聚类方法对其进行聚类。

挖掘机场历史噪声数据集旨在发现满足最大类内距小于某一阈值这一约束条件的簇，而凝聚层次聚类算法适合于解决该约束条件下的聚类问题。由于机场周围的位置点数目较多，而凝聚层次聚类算法的时间复杂度较大，使得其无法有效地解决机场噪声数据挖掘问题。针对上述问题，本文提出一种基于代表点的快速层次聚类算法，该算法在传统凝聚层次聚类算法的基础上，借助聚类代表点法和二分法策略进行改进。将该算法应用于海量机场噪声数据集时，具体流程如图 1 所示。

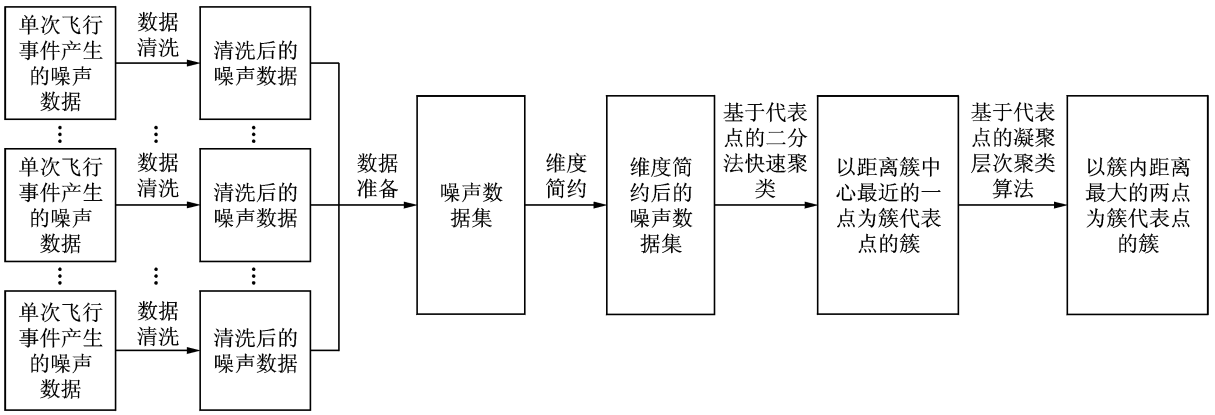

图 1　利用海量机场噪声数据挖掘机场周围噪声分布状况流程图

2.1　数据预处理

2.1.1　对单次飞行事件所产生的机场噪声数据进行清洗

现实世界中的数据通常是有噪声、不完全、不一致的，机场噪声数据也不例外。例如高温、大风、下雨天气致使噪声监测设备不能稳定工作或噪声传输设备出现故障，从而使得获得的噪声数据存在缺值、含有脏数据等问题。因此，在利用数据挖掘方法对其进行挖掘之前有必要进行数据清洗。

针对单次飞行事件产生的噪声数据进行清洗主要包括填补缺失数据、修正脏数据。由于机场噪声数据具有邻域相关性特征，因此可以采用邻域均值法[11]填充缺失数据、修正脏数据。

2.1.2　数据准备

以机场周围的位置点为对象，不同飞行事件下该位置点的噪声值为属性，建立噪声数据存储矩阵 $\boldsymbol{Z}_{n\times m}$，并将海量机场历史噪声数据集导入该矩阵，其中 n 为机场周围位置点数目，m 为飞行事件次数。

2.1.3　维度约简

由于机场噪声数据集的维度较高，有必要采用矩阵分解法对其进行维度约简。常用的矩阵分解法包括非负矩阵分解法、QR 分解法和奇异值分解法等。本文采用奇异值矩阵分解法对噪声数据集 $\boldsymbol{Z}_{n\times m}$ 进行维度简约。

设 $\boldsymbol{Z}_{n\times m}=\boldsymbol{U}_{n\times n}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}}_{m\times m}$，$\boldsymbol{Z}_{n\times m}$ 的奇异值为 $\sigma_i(i=1,2,\cdots,k)$，$k=\mathrm{rank}(\boldsymbol{Z}_{n\times m})$。首先取 σ_i 的前 t 项，使得 $\dfrac{\sum_{i=1}^{t}\sigma_i}{\sum_{j=1}^{k}\sigma_j}\geqslant\delta$（$\delta$ 为阈值，一般取 85%）；然后选择 $\boldsymbol{U}_{n\times n}$ 的前 t 列（即 $\boldsymbol{U}_{n\times t}$），$\boldsymbol{D}_{m\times m}$ 的前 t 行、前 t 列（即 $\boldsymbol{D}_{t\times t}$），$\boldsymbol{V}_{m\times m}$ 的前 t 行、前 t 列（即 $\boldsymbol{V}_{t\times t}$）；最后由 $\boldsymbol{U}_{n\times t}\boldsymbol{D}_{t\times t}\boldsymbol{V}^{\mathrm{T}}_{t\times t}$ 得到维度简约后的矩阵 $\boldsymbol{W}_{n\times t}$。

通过上述步骤，噪声数据集 $\boldsymbol{Z}_{n\times m}$ 由 m 维降为 t 维。

2.2　聚类算法

传统凝聚层次聚类算法效率低下的主要原因如下：

（1）每次循环只选择距离最小的两个簇进行合并；

（2）计算两个簇间的距离时，需要计算簇间任意两个对象之间的距离。

针对上述问题，本文采取如下策略对传统凝聚层次聚类算法进行改进：

（1）机场噪声数据具有邻域相关性等特征，因此在聚类初始阶段的每次循环中，可以同时合并距离较小的许多簇；

（2）通过簇代表点简化簇间距离的计算方式，将两个簇间的距离转换为两个簇的代表点之间的最大距离，即

$$d(C_i,C_k)=\max_{x\in O_i,y\in O_k} d(x,y) \tag{4}$$

式中：C_i 和 C_k 分别表示第 i 和第 k 个簇；O_i 和 O_k 分别表示第 i 和第 k 个簇的代表点集合；x 和 y 分别表示第 i 和第 k 个簇的代表点；$d(x,y)$表示代表点 x 与代表点 y 的距离。

基于上述改进策略，本文提出的基于代表点的快速层次聚类算法分为以下两个阶段：

第一阶段采用基于代表点的二分法进行快速聚类。具体步骤如下：

(1) 令 $s=0$，把每一个位置点看作一个簇，同时也是簇的代表点；

(2) 选取 $n/2^{(s+1)}$ 个簇作为基准簇，根据式(4)计算其余 $n/2^{(s+1)}$ 个簇到每个基准簇的距离，分别把它们划分到与其距离最近的基准簇中；

(3) 根据式(2)重新计算每个簇的簇中心，并选择其作为簇的代表点，$s++$，若 $s<\alpha$，则转回到(2)；

(4) 遍历每一个簇，寻找与簇中心距离最近的一个位置点，将其作为新的簇代表点。

上述步骤中，s 为当前迭代次数，α 为二分法总的迭代次数。

通过上述快速聚类，可以得到以距簇中心最近的一点为簇代表点的簇。同时，可以分析得到基于代表点的二分法快速聚类算法的时间复杂度为 $O(n^2m)$。

第二阶段采用基于代表点的凝聚层次聚类算法进行聚类，具体步骤如下：

(1) 根据第一阶段所得到的簇代表点，利用公式(4)计算任意两个簇间的距离，然后将其存入邻接矩阵 $\boldsymbol{D}$；

(2) 遍历邻接矩阵 $\boldsymbol{D}$，寻找距离最小的两个簇 C_i 和 $C_j(i<j)$，它们之间的距离为 d；

(3)若 d 大于阈值 λ，则算法结束，否则将簇 i 与簇 j 合并为一个新的簇 r，选择使 $d(C_i,C_j)$最大的两个代表点作为新簇 r 的代表点，同时更新邻接矩阵 $\boldsymbol{D}$，根据公式

$$d(C_k,C_r)=\max(d(C_k,C_i),d(C_k,C_j)) \tag{5}$$

计算新簇 r 与任意簇 k 之间的距离 $d(C_k,C_r)$，并将其存入邻接矩阵 Dis，转到步骤(2)。

通过上述聚类，可以得到以簇内距离最大两点作为簇代表点的若干簇。同时，可以分析得到基于代表点的传统层次聚类算法的时间复杂度为 $O\left(\frac{1}{2^{3\alpha}}n^3+\frac{1}{2^{2\alpha}}n^2m\right)$。

通过上述分析可知，本文提出的基于代表点的快速层次聚类算法的时间复杂度为 $O\left(n^2m+\frac{1}{2^{3\alpha}}n^3+\frac{1}{2^{2\alpha}}n^2m\right)$，而传统层次聚类算法的时间复杂度为 $O(n^3m)$。因此，本文提出的基于代表点的快速层次聚类算法的效率明显优于传统层次聚类算法，但这并没有付出额外的空间代价，其空间复杂度仍为 $O(n^2)$。

2.3 聚类结果评价

聚类是一种无监督的学习方法，事先没有任何先验知识，因此需要一定的措施或方法对聚类结果进行有效性验证及评价[12]。通常采用聚类评价指标对聚类结果进行评价，而聚类算法与聚类评价指标所采用的相似性定义往往并不一致，从而导致在某种意义下聚类评价指标的无用性。

可解释性与可描述性是聚类结果评价的一个重要依据[13]，因此在评价聚类结果时应该首先对聚类结果作出相应的描述与解释，再根据描述信息作出评价，而聚类代表点法[14]是一种经典的聚类描述方法。鉴于此，作者提出一种聚类代表点和聚类算法相似性定义相结合的聚类结果评价方法，该方法根据聚类结果是否与外部数据相匹配对聚类结果进行评价，因此能够对聚类结果作出更加科学合理的评价。其评价步骤如下：

(1) 选择合适的簇代表点，使其尽可能精确地描述簇信息；

(2) 根据簇代表点及聚类算法的相似性定义度量聚类结果与外部数据的匹配程度；

(3) 根据匹配程度对聚类结果进行评价。

将基于代表点的快速层次聚类算法应用于机场噪声数据集，并采用聚类代表点和聚类算法相似性定义相结合的聚类结果评价方法对其聚类结果进行评价时，需将特定类型的飞行事件的噪声数据集分为两组：采用基于代表点的快速层次聚类算法对第一组数据进行聚类，用以发现该飞行事件的噪声分布模式。假设通过聚类得到簇 $C_i(i=1,2,\cdots,N)$，P_{i1} 和 P_{i2} 为 C_i 的两个代表点。根据簇代表点和聚类算法的相似性定义度量聚类结果与第二组数据的匹配程度，进而分析利用该模式进行噪声预测的可行性与有效性，其评价流程如下：

(1) 令 $j=1$。

(2) 令 $\varepsilon_{ji}=0$，$i=1$。ε_{ji} 表示利用第 j 个噪声数据文件进行度量时，第 i 个簇中不匹配位置点的数目。

(3) t_{i1}，t_{i2} 分别为第 i 个簇的代表点 P_{i1}，P_{i2} 在第 j 个噪声数据文件中的噪声值。max 为 t_{i1}，t_{i2} 中

的较大值，min 为 t_{i1}，t_{i2} 中的较小值，$l=\max-\min$，若 $l<\lambda$，则 $\max=\max+(\lambda-l)/2$，$\min=\min-(\lambda-l)/2$；否则，$\max=\max-(l-\lambda)/2$，$\min=\min+(l-\lambda)/2$。令 $k=1$。

(4) h_{ik} 为第 i 个簇中的第 k 个位置点 P_{ik} 在第 j 个噪声数据文件中的噪声值，若 $\min\leqslant h_{ik}\leqslant\max$，则该位置点与聚类结果相匹配；否则，该位置点与聚类结果不匹配，$\varepsilon_{ji}++$。

(5) $k++$，若 $k<g_i$，则跳转回(4)。

(6) $i++$，若 $i<N$，则跳转回(3)。

(7) 计算第 j 个噪声文件的匹配度 β_j，$\beta_j=\left(1-\sum_{i=1}^{N}\varepsilon_{ji}\right)\Big/n\times100\%$。

(8) $j++$，若 $j<=\theta$，则跳转回(2)。

(9) 计算聚类结果与外部数据的匹配程度 γ，$\gamma=\sum_{i=1}^{\theta}\beta_i\Big/\theta$，进而对聚类结果作出评价。其中，$\theta$ 表示第二组数据的文件数目，i 表示簇编号，j 表示噪声数据文件编号，k 表示簇内位置点的编号。

3　实验结果与分析

3.1　实验数据的构建

单次飞行事件产生的噪声与多种因素相关，主要包括飞行航迹、机型、机场地理环境和气象条件[15]。其中，飞行航迹和机型是主要影响因素，飞行航迹决定着噪声的分布模式，机型决定着噪声的大小；其他因素为次要影响因素。

挖掘机场历史噪声数据旨在发现特定类型的飞行事件的噪声分布模式，进而借助其进行噪声预测。一般来说，分布模式的泛化能力越强，预测准确度就越低。针对机场噪声数据挖掘问题，特定类型的飞行事件中的已知噪声影响因素与未知噪声影响因素之间的重要性比值越大，挖掘得到的分布模式的预测准确度越高；反之，预测准确度越低。由于机场噪声预测关注预测的准确度，而飞行航迹在噪声影响因素中所占的重要性比重较大，因此着重挖掘具有相同飞行航迹的飞行事件的历史噪声数据具有重要意义。为此，本文从国内某大型机场的历史噪声数据集中截取 6 个月的实测数据，并将其构建为如下 3 组数据进行试验。

Data-Set-1：同一机型在某一航迹下飞行 100 次的噪声数据文件。(已知噪声影响因素：飞行航迹、机型、机场地理环境；未知噪声影响因素：气象条件)。

Data-Set-2：不同机型在某一航迹下飞行 100 次的噪声数据文件(已知噪声影响因素：飞行航迹、机场地理环境；未知噪声影响因素：机型、气象条件)。

Data-Set-3：不同机型组合同时在多条特定航迹下飞行 100 次的噪声数据文件(已知噪声影响因素：飞行航迹组合、机场地理环境；未知噪声影响因素：机型组合、气象条件)。

3.2　实验及分析

本文针对任意一组噪声数据进行实验时，都是从其 100 个噪声文件中随机选取 90 个噪声文件用以发现其噪声分布模式，其余 10 个噪声文件作为外部数据用以评价聚类结果。

实验 1　在 3 组噪声数据上进行实验，用以发现阈值 λ 与簇数目 N 之间的关系，从而确定一个合适的 λ，使得聚类结果较有意义且簇数目较小。一般来说，1，2，3 级声级计的允许误差分别为 0.7，1 和 1.5 dB；而绘制噪声等值线时，等值线之间的间隔一般设置为 5 或 10 dB。因此，令 λ 分别取 0.7，1，1.5，3，5，10，实验结果如图 2 所示(维度简约时，δ 设置为 85%；采用基于代表点的二分法快速聚类时，迭代次数 α 设置为 3 次)。

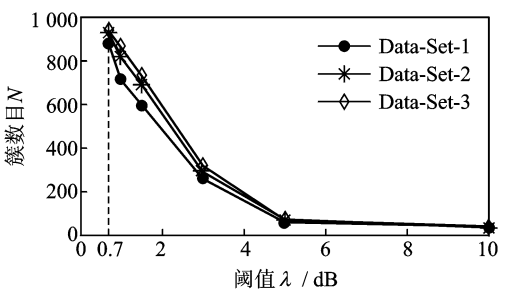

图 2　阈值 λ 与簇数目 N 之间的关系

由图 2 可以看出，随着 λ 的增大，聚类所得到的簇数目逐渐减少。当 $\lambda\in[0,5]$ 时，随着 λ 的增大，簇数目减少较明显；当 $\lambda>5$ 时，随着 λ 的增大，簇数目缓慢减少；而 $\lambda=5$ 时，簇数目比较适中。此外，进行噪声预测时，通常不关心某一具体位置点的噪声值大小，而是关心某一区域受噪声影响的程度，一般认为噪声达 50 dB 就会影响人们的正常生活，每增加 5 dB，影响程度加重一级。因此，λ 取 5 dB较合适，不仅聚类结果较有意义，且簇数目较小。

实验 2　在 Data-Set-2 上进行实验，用以发现二分法迭代次数与匹配度的关系及二分法迭代次数与运行时间的关系。通过确定合适的迭代次数，使得匹配度与运行时间达到均衡。令 α 分别取 1，2，3，4，5，6，7，二分法迭代次数与运行时间及匹配

度的关系分别如图3,4所示(维度简约时,δ设置为85%;阈值λ设置为5 dB)。

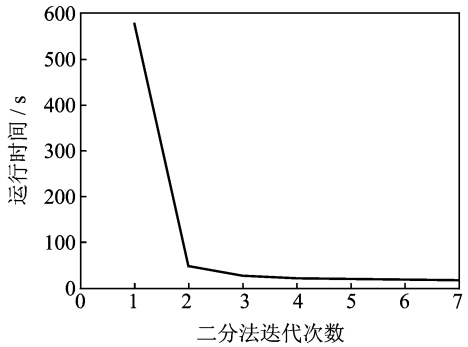

图3 二分法迭代次数与运行时间之间的关系

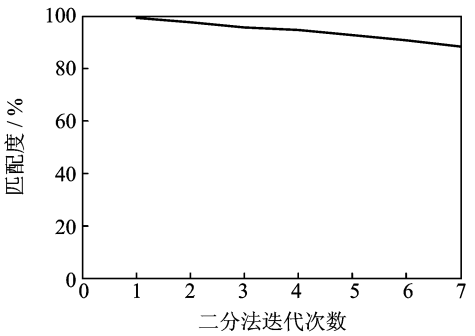

图4 二分法迭代次数与匹配度之间的关系

由图3,4可以看出,随着α逐渐增大,运行时间逐渐降低,匹配度也逐渐降低。当α由0变为3的过程中,运行时间急剧降低,而匹配度的降低却极为缓慢;而当α继续增大时,运行时间基本不变,匹配度却持续降低。然而,α为3、允许误差λ为5 dB时,匹配度高达95.8%。因此,α取3较为合适。

实验3 在三组噪声数据上进行实验,用以验证本文提出的基于代表点的快速层次聚类算法的有效性,实验结果如表1所示(维度简约时,δ设置为85%;采用基于代表点的二分法快速聚类时,迭代次数α设置为3次;阈值λ设置为5 dB)。

表1 3组噪声数据的匹配度实验结果 %

数据分组	最高匹配度	最低匹配度	平均匹配度
Data-Set-1	98.2	95.4	96.9
Data-Set-2	97.3	94.2	95.7
Data-Set-3	96.5	93.7	94.8

由表1可知,针对3组噪声数据进行实验时,匹配度都较高。一方面表明本文提出的基于代表点的快速层次聚类算法能够较准确地发现特定类型的飞行事件的噪声分布模式;另一方面也说明了利用机场历史噪声数据集挖掘机场周围噪声分布模式,进而进行噪声预测具有一定的科学性及合理性。

此外,Data-Set-1的匹配度高于Data-Set-2的匹配度,并且Data-Set-2的匹配度高于Data-Set-3的匹配度,这说明分布模式的泛化能力影响着预测精度。

4 结束语

本文针对机场噪声数据集特征,提出一种基于代表点的快速层次聚类算法,该算法是在传统凝聚层次聚类算法的基础上,借助聚类代表点法和二分法策略进行改进得到。最后,采用国内某大型机场历史噪声数据集对其进行验证,通过理论及实验分析可知,该算法能够较准确地发现机场周围噪声分布模式,同时利用该模式能够较准确地预测特定类型的飞行事件的噪声分布状况。

参考文献:

[1] 夏梓耀,黄锡生.中国机场噪声污染防治立法问题研究[J].北京航空航天大学学报:社会科学版,2011,24(4):38-45.
Xia Ziyao, Huang Xisheng. A study on the legislation issues of airport noise abatement in China[J]. Journal of Beijing University of Aeronautics and Astronautics: Social Sciences Edition, 2011, 24(4): 38-45.

[2] Dikshit P, Crossley W. Airport noise model suitable for fleet-level studies[C]//9th AIAA Aviation Technology, Integration and Operations Conference, Aircraft Noise and Emissions Reduction Symposium. Hilton Head, South Carolina, USA: AIAA, 2009: 138-146.

[3] Koperski K, Han J, Adhikary J. Mining knowledge in geographical data[J]. Transaction on Knowledge and Data Engineering, 1993, 23(10): 903-913.

[4] 孙爽,章勇. 一种基于语义相似度的文本聚类算法[J]. 南京航空航天大学学报,2006,38(6):712-716.
Sun Shuang, Zhang Yong. Clustering method based on semantic similarity[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2006, 38(6): 712-716.

[5] 牛新征,佘堃.面向大规模数据的快速并行聚类划分算法研究[J].计算机科学,2012,39(1):134-151.
Niu Xinzheng, She Kun. Study of fast parallel clustering partition algorithm for large data sets[J]. Com-

puter Science, 2012, 39(1): 134-151.

[6] 倪维健，黄亚楼，李飞，等. 一种基于加权多代表点的层次聚类算法[J]. 计算机科学, 2005, 32(5): 150-154.
Ni Weijian, Huang Yalou, Li Fei, et al. An agglomerative hierarchical clustering algorithm based on weighted representative points[J]. Computer Science, 2005, 32(5): 150-154.

[7] 江敏，皮德常，孙兰. 一种多约束的密度聚类算法的研究[J]. 计算机科学, 2011, 38(10): 143-146.
Jiang Min, Pi Dechang, Sun Lan. Research on density clustering algorithm with a multiple constraints [J]. Computer Science, 2011, 38(10): 143-146.

[8] Jandl B, Rokitansky C. Prediction of noise exposure levels using simulated flight trajectories[C]// 30th Digital Avionics Systems Conference. Seattle, Washington, USA: AIAA, 2011: 1-16.

[9] 孙吉贵，刘杰，赵连宇. 聚类算法研究[J]. 软件学报, 2008, 19(1): 48-61.
Sun Jigui, Liu Jie, Zhao Lianyu. Clustering algorithms research[J]. Journal of Software, 2008, 19 (1): 48-61.

[10] 赵玉艳，郭景峰，郑丽珍，等. 一种改进的 BIRCH 分层聚类算法[J]. 计算机科学, 2008, 35(3): 180-183.
Zhao Yuyan, Guo Jingfeng, Zheng Lizhen, et al. Improved BIRCH hierarchical clustering algorithm [J]. Computer Science, 2008, 35(3): 180-183.

[11] 韩家炜. 数据挖掘概念与技术[M]. 北京：机械工业出版社, 2001.
Han Jiawei. Data mining concepts and techniques [M]. Beijing: Machinery Industry Press, 2001.

[12] Shtern M, Tzerpor V. Refining clustering evaluation using structure indicators[C]// International Conference on Software Maintenance. Edmonton, Alberta, Canada: ICSM, 2009: 297-305.

[13] 吕宗磊，王建东，李莹，等. 一种基于模态逻辑的聚类结果评价方法[J]. 计算机研究与发展, 2008, 45(9): 1477-1485.
Lü Zonglei, Wang Jiandong, Li Ying, et al. An index of cluster validity based on modal logic[J]. Journal of Computer Research and Development, 2008, 45(9): 1477-1485.

[14] 李晓翠，孟凡荣，周勇. 一种基于代表点的快速聚类算法[J]. 南京大学学报：自然科学版, 2012, 48(4): 504-512.
Li Xiaocui, Meng Fanrong, Zhou Yong. The fast clustering algorithm based representative points[J]. Journal of Nanjing University: Natural Sciences, 2012, 48(4): 504-512.

[15] Reza S M, Teh Y W, Lahsasna A. Applied data mining approach in ubiquitous world of air transportation[C]// 4th International Conference on Computer Sciences and Convergence Information Technology. Seoul, Korea: ICCIT, 2009: 1218-1222.