

# 无人值守传感器目标分类方法

永 胜

(内蒙古自治区财政厅, 呼和浩特 010098)

**摘要:** 地面传感器阵列对目标的分类能力是影响其使用效能的关键因素之一; 传感器阵列获取的观测数据受外部环境的影响, 当外部各种物理场特征随时间发生改变时, 采用固定的分类标准进行判断会降低传感器阵列对目标分类识别结果的可靠性; 提出基于动态数据驱动的传感器阵列目标分类方法, 对时变的外部环境特征加以提取作为目标识别计算的反馈控制量, 构建环境改进模型作为决策用有限状态自动识别机, 为其中每个状态训练相应的分类方法, 实现对分类准则的自适应选择, 提高算法性能。

**关键词:** 动态环境建模; 数据融合; 动态数据驱动

## Target Classification Method for Unattended Sensors

Yong Sheng

(Finance Department, Inner Mongolia Autonomous Region, Hohhot 010098, China)

**Abstract:** The classification ability of the ground sensor array is one of the key factors that affect its use efficiency. Sensor array to obtain the observation data are influenced by the external environment, when an external characteristics change over time, various physical fields using fixed classification standard to judge will reduce the credibility of the sensor array of target classification recognition results. Proposed sensor array object classification method based on dynamic data driven, the external environment of time-varying feature extraction for target recognition calculation amount of feedback control, build environment improved model as decision with finite state recognition machine automatically, for each of these state training corresponding classification method, the adaptive selection of classification standards, improve the algorithm performance.

**Keywords:** dynamic context modeling; data fusion; DDD

### 0 引言

地面传感器阵列通过获取各种物理场特征信息实现对所监控区域内目标的探测与识别。在不确定性和动态中检测运动目标, 环境对情报、监视和侦察系统至关重要。为此, 大型分布式传感器网络通常是有吸引力的, 因为它们可以覆盖范围很广的区域, 运行成本适中。每个传感器都在网络中具备通信和计算能力以及具备集体智慧的行为。这种分布式传感器网络的成本效益非常关键, 保持可靠的传感性能同时限制需要进行可靠决策所需的通信量至关重要。单个传感器信息处理能力通常受诸多方面因素制约, 需要通过网络实现数据交互与融合。目前, 常用的数据融合方法很少考虑对传感器工作环境变化的主动响应<sup>[1]</sup>, 这与传感器工作环境的日趋复杂构成了一对矛盾, 特别是通过炮射<sup>[2]</sup>或无人机等方式布撒的传感器, 分布的地域较广, 面临的工作环境更加多样, 给目标识别与分类带来了极大的挑战。为了准确识别目标并消除环境干扰影响。目前的目标检测方法可分为三类: 时域方法、频域方法以及时频域方法。一般来说, 由于干扰噪声的存在时域分析无法做到非常准确的检测目标, 因此相关研究集中在频率域或时间频率目标检测和分类领域特别是基于小波变换的方法, 焦点集中在去噪和时频局部化特性, 达到了良好的分析效果。对于环境因素影响

的处理, 文献 [3] 提出了一种基于模型的传感器阵列分布方法, 从传感器布设位置入手解决问题, 该方法适合于人工布设传感器的场景, 但很难应用于机载布撒的传感器阵列。文献 [4] 采用目标分类与关联跟踪相结合的方法实现对环境影响的处理, 其本质也是一种基于模型的方法, 该方法能够处理传感器目标分类算法对环境噪声的适应性, 但对于回声等因素尚无法处理。与基于模型的方法相比, 动态数据驱动方法并不明确依赖于基于物理原理的科学模型。如果在不同的操作条件下有足够的培训数据, 数据驱动方法明显比基于模型的方法在计算执行时间和内存需求方面有优势。本文提出一种基于动态数据驱动的目标分类算法。对时变的外部环境特征加以提取作为目标识别计算的反馈控制量, 采用环境改进模型作为决策用有限状态自动识别机, 为其中每个状态训练相应的分类识别方法, 实现识别算法对环境影响的自适应过程, 从而提高分类识别的准确度。

### 1 识别算法框架

#### 1.1 定义

本节首先给出与后续算法相关的一些数学定义和表达式。主要是关于有限状态自动识别机和概率有限状态机的定义。有限状态自动识别机 (DFSA) 是一个三元组:

$$G = (\sum, Q, \delta) \tag{1}$$

其中:  $\sum$  是非空有限集, 称为符号集, 满足  $|\sum| < \infty$ ;  $Q$  是非空有限集, 称为状态集, 满足  $|Q| < \infty$ ;  $\delta: Q \times \sum \rightarrow Q$  是状

收稿日期: 2017-08-15; 修回日期: 2017-10-18。

作者简介: 永胜 (1977-), 男, 高级工程师, 博士, 主要从事电子技术方向的研究。

态转移映射。

概率有限状态机 (PFSA) 是基于一个 DFSA 构建的, 形如:

$$K = (G, \pi) = \left( \sum, Q, \delta, \pi \right) \quad (2)$$

其中,  $\pi: Q \times \Sigma \rightarrow [0, 1]$  为概率映射函数满足  $\sum_{\sigma \in \Sigma} \pi(q, \sigma) = 1, \forall q \in Q$ 。

如果  $Q_G$  是已知环境的集合,  $\Psi$  是特征的集合,  $C$  为事件分类的有限集。则函数:

$$A: Q_G \times \Psi \rightarrow C \quad (3)$$

为基于环境的事件分类器, 满足:  $\forall q \in Q_G, \forall P \in \Psi, A(q, P) \in C$ 。

### 1.2 目标分类算法框架

无人值守的地面传感器工作在复杂的环境条件之中, 其观察数据会随这些因素改变。因此, 目标分类算法及其训练需要一个模型来捕捉外部因素。在某地理区域内布置传感器阵列, 阵列中传感器之间具有通信能力。每个传感器都包括不同的传感器模块, 如被动红外, 声学 and 地震模块等<sup>[6-8]</sup>, 即传感器阵列包含一些独立的传感器, 每个独立的传感器可感知一种或几种物理场信息。每一个传感器模块都有一个有限的处理和通信能力, 能够完成目标检测和分类任务, 并将结果进行彼此沟通。令  $c$  为待探测的事件分类数量, 例如: 如车辆 (C1), 动物 (C2), 人员 (C3), 则  $c=3$ , 且定义  $C_0$  为上述事件均未出现的状态。动态数据驱动算法执行的通用框架如图 1 所示。

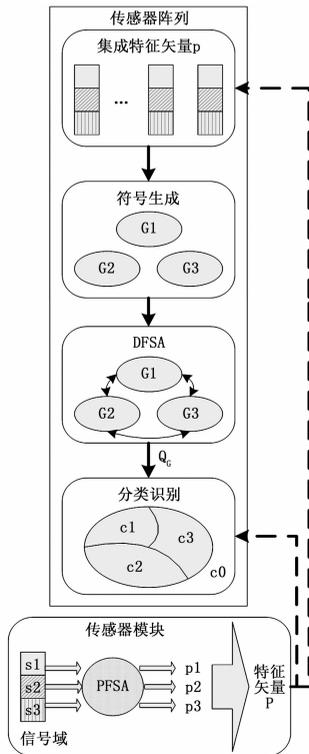


图 1 算法框架示意图

算法的最底层, 采用 SDF 法, 通过来自传感器的信号域数据 ( $S_1, \dots, S_m$ ) 用来计算各个状态的概率矢量 ( $p_1, \dots, p_m$ )。并将结果合并为一个特征矢量 ( $P$ )。对于每个独立的传感器来说, 都存在一个  $P$  矢量, 如果传感器节点数量为  $n$ , 则

可以获得矢量  $P_1, \dots, P_n$ , 环境特征不能从单个传感器数据中独立获得, 需要进行特征层面的数据融合。待探测事件的集合  $\{C_i, i=1, \dots, m\}$  的具备空间维度上的一些特征<sup>[5]</sup>, 采用下式对所获得的特征矢量进行处理:

$$P^{(i)} = [p_1^{(i)} \ p_2^{(i)} \ \dots \ p_m^{(i)}] \quad (4)$$

$$\tilde{P} = \begin{bmatrix} \sum_{i=1}^n \alpha_i \otimes p_1^{(i)} \\ \vdots \\ \sum_{i=1}^n \alpha_i \otimes p_m^{(i)} \end{bmatrix} \quad (5)$$

其中:  $i$  表示单独获取某种物理场的传感器,  $m$  为一个传感器节点上上述传感器的总数,  $\alpha_i$  是权值,  $n$  为传感器节点的总数。

上述两式中涉及的矢量运算定义如下:

$$p_i = [p_{j1}, p_{j2}, \dots, p_{jd}] \forall j \in \{1, 2, 3\} \quad (6)$$

$$p_1 \oplus p_2 = p_3$$

$$p_{3i} = \frac{p_{1i} p_{2i}}{\sum_{j=1}^d p_{1j} p_{2j}} \forall i \in \{1, \dots, d\}$$

$$p_1 \otimes \alpha_1 = p_2$$

$$p_{2i} = \frac{p_{1i}^{\alpha_1}}{\sum_{j=1}^d p_{1j}^{\alpha_1}} \forall i \in \{1, \dots, d\} \quad (7)$$

考虑矢量加法和标量乘法的闭合性, 将其变为 Hilbert 空间中的符号表示法, 详见文献 [6]。

SDF 处理算法可获得特征空间中的一个数据点, 在时间上重复该算法, 在算法的上层, 采用 K 均值聚类方法, 引入非监督学习机制, 将相同的特征, 即特征空间中相同的数据点加以合并, 获得特征数据集如:  $G_1, G_2, \dots, G_k$ ,  $K$  的选择由模型复杂性及模型误差共同决定。这样, 特征集合可以借助符合  $G_k$  来表示, 即获得了公式 (1) 中的  $\Sigma$ 。通过 D-Markov 方法获得 DFSA 模型, 该模型即为环境进化或环境改变模型。模型中每个状态都是 D 符号的链, 也可以认为是瞬时或短周期符号的集成, 这样就获得了关于环境的符号表示。DFSA 中的各个状态, 称为环境状态。一旦出现新的特征矢量, DFSA 过程对环境状态将做出更新, 从而使得对目标的分类能够应用最新的状态反馈信息, 适应环境的短时间和长期变化。

## 2 基于符号化动态滤波的目标特征提取

符号化动态滤波 (Symbolic Dynamic Filtering, SDF) 用于从单个传感器数据中提取特征参数, 它的理论尚未完全成熟, 目前普遍将动态滤波处理成一个双时间尺度问题, 在短周期上, 系统对数据的获取和处理模型可认为是时不变的, 而采用长周期模型来描述系统中的缓变特征。其基本步骤为:

第一步, 在长周期上采集传感器数据的时间序列  $\{q\} \in Q, Q \subseteq n \times N, n$  为时间序列的维数,  $N$  是每个时间序列的点数;

第二步, 将状态空间划分为  $\{B_0, \dots, B_{(|\Sigma|-1)}\}$ , 每个元素用一个符号  $\sigma_i \in \Sigma$  表示, 范数形式任取, 将时间序列映射到  $\sigma_i$ , 即  $\{q\}$  转换为符号序列  $\{s\}$ ;

第三步, 使用训练数据, 构建序列  $\{s\}$  以获取 PFSA, 使用测试数据, 通过上述 PFSA 识别函数  $\pi$ , 此处的 PFSA 是 D-Markov 机 [3]。使用上述  $\pi$  和  $\delta$ , 可获得转移概率矩阵  $\Pi = |Q| \times |Q|$ , 其元素含义为从状态  $q_i$  到状态  $q_j$  的转移概

率。将转移概率矩阵进行上三角变换，其变换矢量  $p$  用来描述时间序列的特征。

以震动传感器为例，传感器测量的直接结果是一个时间序列。采用 SDF 方法时，时间序列首先被转换至小波域，小波系数是在不同的时间变换和尺度下产生的，基于单个信号的时频特性选择小波基函数与小波尺度。获得小波表面轮廓后，将其进行划分，将小波系数的像素值在最大值和最小值之间分成不同的相互排斥的区域，选择不同的符号来标识区域的分区，即每个区域都被贴上标签，这样就得到了符号集，如果一个值位于上述特定的区域，则它被编码为与该区域相关的符号。该过程的本质是一种基于原始样本数据的特征提取，其关键在于选取能够描述数据核心结构的特征，被选特征具备使得同类样本在特征空间中呈现距离值最小化的特点，而特征矩阵就是对这一结果的数学描述。

### 3 特征层面数据融合

经过 SDF 处理传感器数据而获得的符号系统是一个概率有限状态自动机，为了从传感器数组中推断出环境特征，需要将 SDF 获得的特征向量进行加权平均，权重的选择依据传感器工作性能先验知识来确定，如果传感器性能良好，则每个传感器的权重一致，如果预先知道某传感器性能退化，则降低其获得的权重。

对于加权处理后的数据，应用于基于 K 均值的非监督学习技术对其进行分类。K 均值聚类用于将分组 G 的元素分成  $k$  个互斥簇。在初始启动集群质心的位置之后，根据预定义的距离分配每个  $G_i$  到最近的群集测量。集群中心通过最小化每个  $G_i$  与其相应的聚类中心之间的距离度量的总和来更新。该集群分配和距离最小化过程是迭代重复的，直到算法收敛，G 的要素没有进一步的变化集群。集群质心的位置被随机初始化以避免局部最小。并行使用重复与随机初始启动以确保算法返回距离度量的最小总和。该 k 均值聚类的输出是重心的位置，可用于设置分类目标的门槛。由于很难预先知道哪些条件会显著影响信号的生成聚合特性，即使所有条件都已知，也很难确定数据中集群，即分类的数量。可以用残差来估计建模误差，通过选择一个参数权衡建模误差和模型的复杂性。这样，一旦获得了数据集中的集群数量，K 均值聚类就会给出 K 不相交的数据集。集群标签被分配到该集合中的每个数据点，在后续环境模型中作为分类符号使用。

### 4 环境模型

要获得环境及其变化的模型，需要借助于传感器之间的网络通信功能，将不同传感器的数据进行基于时间和空间的聚合。如假设一组训练数据来自分布于不同地质特征环境中的震动传感器，一部分传感器布置于水泥地面，另一部分则布置于土质地面。这样，经过网络的数据交换，位于网络中的每个传感器都会获得这个环境信息，从而在目标分类过程中利用与环境信息有关的判断准则进行判断，达到处理环境变化的效果。环境的变化在模型中表示为状态转移函数，通过 D-Markov 方法获得，D 的层数表示状态数的复杂性，即模型深度，更大的深度提供了更复杂的模型，能够提高分类精度，但网络中的通信量增加，必须进行权衡，选择合适的深度以匹配

网络的其他性能。D-Markov 机的构造是基于依据相对重要性进行状态分割产生不同长度的符号块，进而从符号块中引入历史数据完成状态合并，在执行任何状态合并之前，SDF 获得的字符都可以作为状态来处理。

### 5 仿真分析

采用仿真数据对上述算法进行验证与分析，通过对数据叠加参数不同的高斯白噪声模拟不同的环境特征。仿真数据的类型分为三种，分别表示动物、人和车辆；此外，另采用一组数据模拟没有任何目标的情形。图 2 中给出了叠加到数据中的模拟环境特征带来的影响，对于仿真的同一组具有车辆信息的时间序列数据，添加不同的环境特征，其结果如图 2 所示，图中上面的两个子图表示时域数据的区别，下面两个子图代表特征矢量的区别，从中可以看出，对于同一目标，在不同环境下，其特征矢量发生了显著变化。

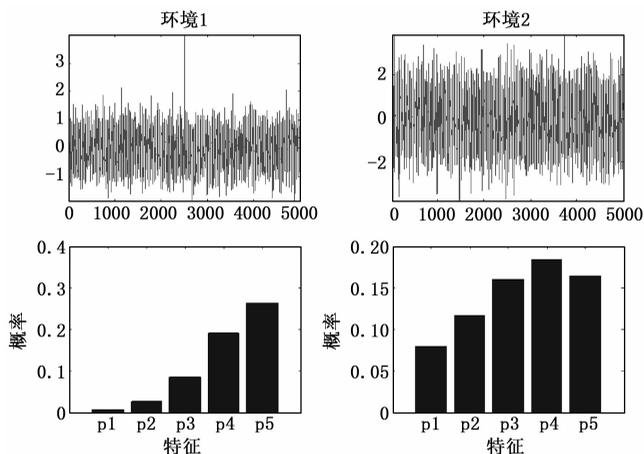


图 2 不同环境特征对于目标数据的影响

在获得了仿真的目标数据，并对其添加代表不同环境的噪声后，对所产生的数据进行采样，采样所使用的频率根据计算需求和数据类型确定。将采样后的数据分为两组，其一用于对计算模型的训练，其二用于对算法的验证。用于模型训练的数据将获得上文算法中的特征矢量，从而得到完备的目标分类模型。进一步地，将用于验证的数据代入，验证算法的有效性。

运用 Hilbert 空间中的符号表示法，确定数据中对应几种目标，图 3 中给出了分组数量与分组评分之间的关系，可以看出，当分组数量取 3 时，对应的评分为最小值，满足符合表示法的规则要求，因此，目标分类为 3，即算法中。

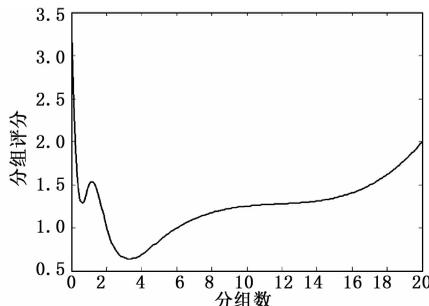


图 3 目标分组数量与分组评分示意图

通过 D-Markov 机获得的 DFSA 模型中每个状态都是 D 符号的链, 即瞬时或短周期符号的集成, 由此获得关于环境的符号表示。符号表示的复杂程度对于识别与分类算法的影响见图 4。

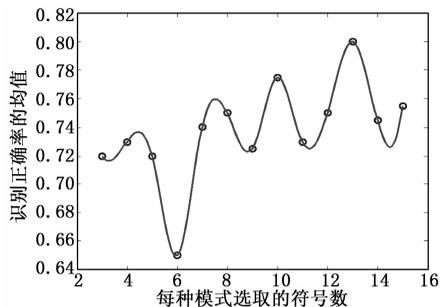


图 4 符号表示的复杂程度对结果的影响

从图 4 中可以看出, 符号表示数目增加, 识别结果总体趋势趋向更高的识别率, 这是因为更多的符号表示数目其物理含义对应的是对环境特征的更具体的分类, 增强了算法的环境适应能力。同时, 也可以看到这个过程中扰动十分明显, 这是由于更细致的分类将引入更多的噪声。因此, 需要根据实际情况选择合适的符号表示数目, 以达到最佳的识别效果。

## 6 小结

文章给出了一种基于动态数据驱动的与环境互动的目标识别分类算法框架。通过一种具有聚合特性的传感器融合技术, 在指定区域的传感器阵列的所有传感节点中, 获取一个符号数据表示传感器数据。利用 K 均值聚类算法获取环境模型, 依赖传感器网络的通信能力在网络中不同区域传感器之间传递环

(上接第 277 页)

身的固有属性, 因此应合理规划电网的结构从而降低系统的脆弱性, 提高电网可靠性水平, 增加系统稳定性能。另外, 在以后的研究工作中, 可更多考虑脆性评估指标的多样性以及故障模式的全面性将更加符合实际电网的运行。

## 参考文献:

[1] Crucittip, Iattorav. Marchiori M. A Topological Analysis of the Italian Electric Power Grid [J]. *Physics A*, 2004, 338: 92-97.

[2] Albert R, Aibert I, Nakarad G L. Structural Vulnerability of the North American Power Grid [J]. *Physical Review E*, 2004, 69 (2): 025103-1-025103-4.

[3] 孟仲伟, 鲁宗相, 宋靖雁. 中美电网的小世界拓扑模型比较分析 [J]. *电力系统自动化*, 2004, 28 (15): 21-24.

[4] 丁明, 韩平平. 加权拓扑模型下的小世界电网脆弱性评估 [J]. *中国电机工程学报*, 2008, 28 (10): 20-24.

[5] Ding Ming, Han Pingping. Reliability assessment to large-scale power grid based on small-world topological model [A]. *Proceedings of the 5th IEEE International Conference on Power System Technology* [C]. Chongqing, China, 2006.

[6] 蔡晔, 陈彦如, 曹一家, 等. 基于加权网络结构熵的电网连锁故障研究 [J]. *复杂系统与复杂性科学*, 2013, 10 (1): 54-55.

[7] 和晨萍, 朱耀琴. 恶劣环境下电网脆弱性仿真研究 [J]. *计算机仿真*, 2014, 31 (5): 93-96.

境改变的信息, 以增强每个传感器的目标检测和分类能力。采用仿真数据对所提出的算法进行了仿真, 结果表明该计算框架能够提高传感器阵列对于环境参数变化的适应性, 提高不同环境中目标分类的准确性。

## 参考文献:

[1] Iyengar S G, Varshney P K, Damarla T. On the Detection of Footsteps Based on Acoustic and Seismic Sensing [A]. *Asilomar Conference on* [C]. IEEE, 2008: 2248-2252.

[2] 陈坤培, 王曙光, 宁全利. 一种基于炮射地面震动传感器的目标识别算法研究 [J]. *弹箭与制导学报*, 2011, 31 (2): 185-188.

[3] 王斌, 冯立杰. 阵列式振动传感器环境干扰消除方法研究 [J]. *电子设计工程*, 2012, 20 (18): 155-157.

[4] 张星, 王雪, 刘长. 传感器阵列预测空域多信号分类目标定位跟踪 [J]. *仪器仪表学报*, 2012, 33 (5): 970-975.

[5] Mehmood A, Patel V M, Damarla T. Discrimination of bipeds from quadrupeds using seismic footstep signatures [J]. *IEEE Xplore*, 2013, 88 (60): 6920-6923.

[6] Jin X, Sarkar S, Ray A, et al. Target Detection and Classification Using Seismic and PIR Sensors [J]. *IEEE Sensors Journal*, 2012, 12 (6): 1709-1718.

[7] Mckenna M H, Mckenna J R, Anderson T S. Effects of Local Meteorological Variability on Surface and Subsurface Seismic-Acoustic Signals [A]. *AGU Spring Meeting Abstracts* [C]. 2006.

[8] Wilson D K, Marlin D H, Mackay S. Acoustic/seismic signal propagation and sensor performance modeling [A]. *Proc. of Unattended Ground, Sea, and Air Sensor Technologies and Applications IX* [C]. 2007.

[8] 丁明, 韩平平. 基于小世界拓扑模型的大型电网脆弱性评估算法 [J]. *电力系统自动化*, 2006, 30 (8): 7-10.

[9] 魏震波, 刘俊勇, 朱国俊. 基于可靠性加权拓扑模型下的电网脆弱性评估模型 [J]. *电工技术学报*, 2010, 25 (8): 131-137.

[10] 蔡晔, 曹一家, 李勇. 考虑电压等级和运行状态的电网脆弱线路辨识 [J]. *中国电机工程学报*, 2014, 34 (13): 2124-2129.

[11] 卢锦玲, 朱永利. 基于暂态能量裕度的电力系统脆弱性评估 [J]. *电工技术学报*, 2010, 25 (6): 96-103.

[12] 刘文颖, 等. 计及恶劣天气因素的复杂网络连锁故障事故链模型 [J]. *中国电机工程学报*, 2012, 32 (7): 53-59.

[13] Forrest D W, et al. Proposed Terms for Reporting and Analyzing Out-ages of Electrical Transmission and Distribution Facilities [J]. *IEEE Transactions on Power Apparatus and Systems*, 985, PAS-104 (2): 337-348.

[14] 周彦衡, 吴俊勇, 张广韬, 等. 考虑级联故障的电力系统脆弱性评估 [J]. *电网技术*, 2013, 37 (2): 444-449.

[15] 唐强. 电力系统脆弱性评估方法研究 [D]. 武汉: 华中科技大学, 2012.

[16] 魏震波, 刘俊勇, 朱国俊, 等. 基于可靠性加权拓扑模型下的电网脆弱性评估模型 [J]. *电工技术学报*, 2010, 25 (8): 131-136.

[17] 李勇, 刘俊勇, 刘晓宇, 等. 基于潮流偏的电网连锁故障传播元件的脆弱性评估 [J]. *电力系统自动化*, 2012, 36 (19): 1-6.