



## 基于统计聚类方法的儿童下肢肌电信号周期识别

闫成起<sup>1</sup>, 赵利华<sup>2</sup>, 陈梦婕<sup>2</sup>, 周 军<sup>1</sup>

(1. 上海交通大学 电子信息与电气工程学院 电子系, 上海 200240; 2. 上海交通大学医学院附属儿童医院, 上海 200062)

**摘要:** 为运用肌电信号分析髋脱位儿童和正常儿童的差异, 提出一种基于统计的聚类方法, 识别步态中下肢肌电信号的周期起始时刻。使用非参数贝叶斯模型将肌电信号序列聚类为状态序列, 并通过  $k$  均值聚类算法将该状态序列标记为肌肉活跃和不活跃两种状态, 将肌肉活跃状态的起始时刻作为肌电信号周期的起始位置, 并且利用窗函数方法提高预测准确性。实验结果表明, 该方法对于预测正常儿童周期起始位置的识别误差较小, 平均值为 2.15%, 并且在 5% 的置信度水平下与 SampEN、SNEO 和 IP 等检测算法相比具有较高的预测准确率。

**关键词:** 肌电信号; 周期识别; 统计聚类方法; 非参数贝叶斯模型;  $k$ -means 算法; 滑动窗

开放科学(资源服务)标志码(OSID):



**中文引用格式:** 闫成起, 赵利华, 陈梦婕, 等. 基于统计聚类方法的儿童下肢肌电信号周期识别[J]. 计算机工程, 2021, 47(5): 273-276, 284.

**英文引用格式:** YAN Chengqi, ZHAO Lihua, CHEN Mengjie, et al. Period identification for electromyography signals of children's lower limb based on statistical clustering method[J]. Computer Engineering, 2021, 47(5): 273-276, 284.

## Period Identification for Electromyography Signals of Children's Lower Limb Based on Statistical Clustering Method

YAN Chengqi<sup>1</sup>, ZHAO Lihua<sup>2</sup>, CHEN Mengjie<sup>2</sup>, ZHOU Jun<sup>1</sup>

(1. Department of Electronics, College of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China; 2. Children's Hospital Affiliated to Medical College of Shanghai Jiao Tong University, Shanghai 200062, China)

**[Abstract]** To promote the application of Electromyography (EMG) signals in the analysis of differences between normal children and children with hip dislocation, this paper proposes a method based on statistical clustering for detecting the starting point of the period of EMG signals from lower limb muscles of walking children. The method employs the nonparametric Bayesian model to cluster EMG signal sequences as pattern sequences, which are subsequently marked with tags of active state and inactive state by using the  $k$ -means algorithm. The starting point of the active state of muscle activities is taken as the starting point of a period of EMG signals, and the window function method is used to improve the prediction accuracy. Experimental results show that the average recognition error of this method is as small as 2.15%, and is significantly different from that of the other detection algorithms, including SampEN, SNEO and IP when the confidence level is 5%.

**[Key words]** Electromyography (EMG) signal; periodic identification; statistical clustering method; nonparametric Bayesian model;  $k$ -means algorithm; sliding window

DOI: 10.19678/j.issn.1000-3428.0057923

### 0 概述

肌电 (Electromyography, EMG) 信号的周期起始点检测是指从采集的肌电信号序列中判断出每一步的起始时刻, 它是肌电信号应用于智能假肢控制、康复训练和诊断评估的重要条件。人们在行走时, 双

腿交替迈动表现出明显的周期性, 因此借助人体步行运动的生理学特点, 研究人员采用加速度传感器、压力板以及三维步态采集系统判断周期的起始位置<sup>[1-3]</sup>, 利采用左腿胫骨前肌 (Left leg Tibialis Anterior, LTA) 的表面肌电信号 (Surface Electromyography, SEMG) 分析周期起始位置。从单条腿获取的 SEMG

基金项目: 上海交通大学“科技创新专项资金”(YG2017MS33)。

作者简介: 闫成起 (1994—), 男, 硕士研究生, 主研方向为统计机器学习; 赵利华、陈梦婕, 医师; 周 军, 副教授。

收稿日期: 2020-03-31 修回日期: 2020-05-28 E-mail: yan01121123@sjtu.edu.cn

信号来看, LTA 从上一次活动开始到再一次活动持续的时间定义为一个跨步周期, 每个跨步周期可看作是一个步态活动段<sup>[4-5]</sup>。步态活动段是神经系统、骨关节系统和相关肌肉协同调节的结果。因此, 研究步态活动段对脑瘫病情诊断和髋关节脱位患者在术后康复训练中具有重要的指导意义和评估价值<sup>[6-7]</sup>。

由于采集方法、系统以及实验环境的差异, SEMG 信号的周期起始点检测方法的评价指标并不统一, 文献[8]提出的基于奇异谱的变化点分析方法取得了较好的效果, 但是当信号受到噪声干扰时, 该方法检测误差较大。近年来, 随着非线性方法的研究, 一些能够描述肌电非线性特征的周期起始点检测算法应运而生<sup>[9-11]</sup>。最常用的有综合轮廓算法<sup>[12]</sup> (Integrated Profile, IP)、基于样本熵 (Sample-Entropy, SampEN) 的活动段检测算法<sup>[13-14]</sup> 和基于非线性能量算子 (Smoothed Nonlinear Energy Operator, SNEO) 的峰值检测算法<sup>[15-16]</sup>。IP 算法的起始点检测使用极值算法, 在轮廓特征不规则的条件下该算法效果不佳。SampEN 算法和 SNEO 算法的起始点检测效果依赖阈值的选取, 而肌电信号的非平稳性会影响阈值的选取, 使得检测效果不够稳定。

本文提出一种基于统计聚类算法的步态周期起始点检测方法。使用 BP-AR-HMM (Beta Process Autogressive Hidden Markov Model)<sup>[17-19]</sup> 挖掘步态过程的状态序列, 通过多个 SEMG 信号的差异性标注对应肌肉在步态中的状态, 使用  $k$  均值聚类算法<sup>[20]</sup> 识别肌肉活动段, 并且结合滑动窗投票 (Smooth) 方法实现周期起始位置的标注, 定量分析算法预测的准确性。

### 1 基于聚类算法的周期识别

步态周期起始位置检测算法流程如图 1 所示。首先对 SEMG 信号进行降噪及标准化处理, 减少噪声对模型性能的影响, 然后将多个 SEMG 信号序列共同作为 BP-AR-HMM 模型的输入, 标注出 SEMG 信号对应的 LTA 肌肉在正常行走中的状态变化过程, 模型输出为肌肉的状态序列和对应状态的方差参数, 但是模型输出的肌肉状态较多而且很多持续时间太短, 不利于分析和周期起始位置的确定, 而在步态中肌肉对应的状态是缓慢变化的, 代表相邻状态方差比较接近, 因此利用  $k$ -means 方法聚类状态对应的方差参数, 使肌肉状态分为活跃状态和不活跃状态, 将肌肉活跃状态的起始位置作为步态周期的初始值。SEMG 信号是一类非平稳信号, 会存在很短的肌肉活动状态, 进而影响模型对起始位置的检测效果, 所以通过滑动窗方法平滑这些状态来提高模型性能。

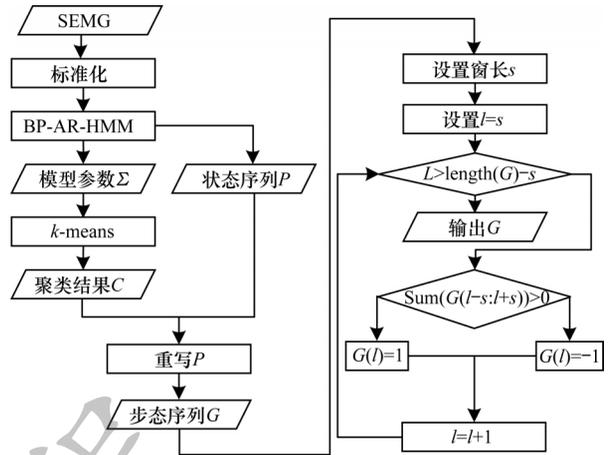


图 1 基于聚类方法的步态周期起始点检测算法流程  
Fig.1 Procedure of gait cycle starting point detection algorithm based on clustering method

#### 1.1 BP-AR-HMM 模型原理

BP-AR-HMM 模型旨在发现不同序列之间的共享模式, 本文应用该方法得到步态中 SEMG 信号对应肌肉的变化规律。概率图模型结构如图 2 所示。

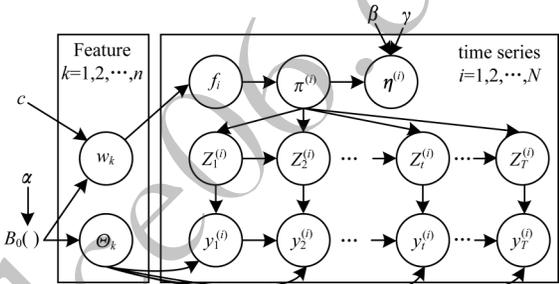


图 2 BP-AR-HMM 模型结构  
Fig.2 Structure of BP-AR-HMM model

模型主要包含 Beta-Bernoulli 过程和 AR-HMM 模型两部分:

1) AR-HMM 模型: 因为步态数据是一维时间序列而且可以很直观地分解成一系列动作, 一般采用 HMM 模型描述肌肉隐状态和步态数据之间的关系。然而, HMM 模型假设在给定隐状态条件下, 步态数据之间是相互独立的, 这种假设不符合步态数据产生的真实过程, 为了能更好地建模步态变化过程, 本文加入了  $r$  阶 AR 模型, 使观测数据同时受隐状态和前  $r$  时刻观测数据的影响, 整个过程如式(1)~式(3)所示:

$$z_t | z_{t-1} \sim \pi_{z_{t-1}} \tag{1}$$

$$y_t | z_t \sim F(\theta_{z_t}) \tag{2}$$

$$y_t = \sum_{v=1}^r A_{v,z_t} y_{t-v} + e_t(z_t) \tag{3}$$

其中,  $e_t(z_t) \sim N(0, \Sigma_{z_t})$ ,  $A_k$  是滞后矩阵,  $F(\cdot)$  是指数分布,  $\pi_k$  是转移分布,  $\theta_k = \{A_k, \Sigma_k\}$  是状态  $k$  的发射参数。

2) Beta-Bernoulli 过程: Beta 过程用来发现步态

序列中可能模式的集合,  $Beta$  过程的随机测度定义为  $B = \sum_k w_k \delta_{\theta_k}$ , 其中  $w_k$  为模式  $\theta_k$  的发生概率。

Bernoulli 过程描述时间序列  $i$  和状态  $k$  的关系  $y_i = \sum_k f_{ik} \delta_{\theta_k}$ ,  $f_{ik}$  表示序列  $i$  是否包含状态  $k$ , 它依赖于  $Beta$  过程的  $w_k$  参数, 即:

$$f_{ik} | w_k \sim \text{Bernoulli}(w_k) \quad (4)$$

### 1.2 k-means 算法和 Smooth 方法

在步态序列中, 假设肌肉在活跃状态下 SEMG 信号的能量不会发生突变。BP-AR-HMM 模型参数  $\Sigma$  为状态方差集合, 通过观察状态方差在步态数据中的变化, 发现当肌肉活动时对应状态集合中的各状态方差相距较近。为了能够更好地识别肌肉活动段, 本文使用  $k$ -means 算法聚类状态方差参数, 利用欧式距离度量两个状态相似程度, 最终将状态集合聚类为两种模式  $C = \{-1, 1\}$ , 其中, 1 代表肌肉活跃状态, -1 代表肌肉不活跃状态, 根据  $C$  与原状态集合的对应关系, 得到步态序列。

在步态过程中, 肌电信号会存在时间很短的抖动状态, 在步态序列中表现为肌肉活跃模式, 会影响周期起始位置标注。通过选取合适大小的滑动窗, 计算窗内每个状态的样本点数量, 选择包含样本点最多的状态为整个滑动窗对应的肌肉状态, 来消除抖动状态的影响。在实验中发现抖动状态持续时间不超过周期时间的 10%, 正常儿童平均周期约为 1.3 s, 所以最终设置窗口大小为 0.13 s。

## 2 实验结果与分析

本文实验共使用 12 名儿童的 SEMG 信号数据, 其中, 5 位髌脱位患者, 7 位正常儿童。根据研究需要, 选取的实验对象符合以下条件: 1) 可以独立行走; 2) 无患有影响下肢运动障碍的疾病。所有数据的使用得到上海市儿童医院的授权和同意, 实验通过伦理委员会审查, 符合自愿原则并且被试者知晓实验内容和步骤。

为得到高质量数据, 实验前对皮肤以及电极贴片都经过标准化处理, 减少皮肤噪声的干扰。在实验过程中, 由家长和主治医生协助儿童完成指定动作。实验动作包括站立、环绕步态分析室以正常速度行走及转弯。使用 Noraxon Gait Analysis System 采集 SEMG 信号数据, 设置采样频率为 1 500 Hz, 多次采集 LTA 的 SEMG 信号后选择较好的三段步态数据, 每段数据长度为 5 个周期。

首先分析单个 SEMG 信号在整个聚类算法中的变化, 如图 3 所示, 纵轴对应序列算法的 3 个步骤, 实线箭头为周期起始位置的真值, 虚线箭头为模型预测的起始位置。BP-AR-HMM 模型将 SEMG 信号标注为一系列状态, 即代表肌肉变化的隐状态序列, 能

够描述步态中 SEMG 信号的变化规律。可以看出有的黑色状态持续时间很短, 而该状态对于分析肌电信号或者周期识别没有太多的帮助, 并且主要目的是通过识别肌肉活动状态来进一步识别周期。通过对比各个状态的方差可以发现, 肌肉活动时各个状态对应的方差较为接近。通过  $k$ -means 算法聚类状态集合的方差参数将方差相近的状态聚为同一模式, 最后得到只有两种模式的步态序列, 结合肌电图发现黑色代表肌肉活动状态, 白色代表不活跃状态, 如图 3 中  $k$ -means 对应的序列可以看出, 会存在时间持续很短的黑色状态, 而该状态会影响到周期识别, 利用滑动窗方法将持续时间较短的黑色状态进一步平滑后得到 Smooth 对应的步态序列, 肌肉活动段的起始位置即为周期的开始时刻(如黑色虚线箭头所示)。

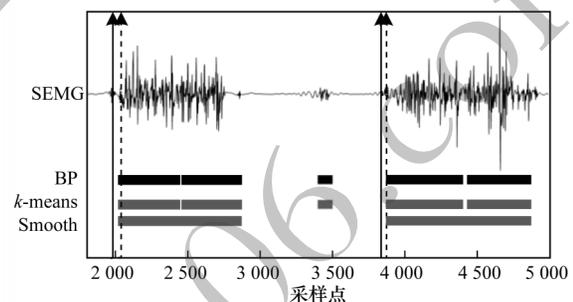


图 3 LTA 信号的分析过程

Fig.3 Analysis process of LTA signal

本文使用预测误差 (error)<sup>[11]</sup> 指标定量分析算法对肌电信号周期的预测准确性, 具体定义如式 (5) 所示:

$$\text{error} = \frac{|\text{true} - \text{predict}|}{\text{period}} \times 100\% \quad (5)$$

其中,  $\text{period}$  为以  $\text{truth}$  为真实起始值对应的周期时间,  $\text{predict}$  为周期起始位置预测值,  $\text{error}$  值越大, 误差越大。

本文研究基于统计聚类方法对周期检测的实用性, 不同方法的预测误差如表 1 所示。其中, 表中数字为 (Mean+Std), Mean 为误差均值, Std 为预测误差集合的标准差, 误差均值反映模型对所有样本预测的准确性, 而标准差代表预测误差在误差均值附近波动的大小, 反映模型预测的稳定性。

表 1 不同检测算法的预测精度比较

Table 1 Comparison of prediction accuracy of different detection algorithms %

组别	聚类算法	SampEN	SNEO	IP
正常儿童	2.15+0.76	3.45+1.45	2.49+0.89	3.12+1.03
髌脱位儿童	5.10+3.49	4.74+3.50	5.31+3.46	4.96+3.39

从表1可以看出,本文方法预测正常儿童周期起始位置误差最小,平均预测误差为2.15%,SampEN算法的误差较高,而且对于正常儿童基于聚类方法的标准差最小,说明算法稳定性较高。不同算法对于髌脱位儿童的周期预测误差均高于正常儿童,因为无论是基于聚类的算法还是本文比较的3种算法都是通过识别肌肉活动状态来识别周期起始位置,而髌关节脱位儿童的步态稳定性较差,所以会造成识别的不准确,同时由于SEMG信号采集过程中噪声的影响,也会增大预测误差。最后通过t检验方法得到的本文方法在正常儿童数据中的预测准确率要明显优于其他3种检测方法( $p < 0.05$ ),而在髌脱位儿童数据中的预测误差没有明显差异。

### 3 结束语

本文提出一种基于聚类算法的运动周期检测方法,应用BP-AR-HMM模型得出步态中SEMG信号对应肌肉的变化规律,并结合滑动窗投票方法实现周期起始位置的标注。实验结果表明,本文方法对正常儿童步态数据的周期起始值预测准确性较高。但该方法对于髌脱位儿童的周期检测误差较大,主要原因为髌脱位儿童的SEMG信号不够稳定,同时信号采集环境和设备都会影响最终的检测效果,且本文只使用了一块肌肉的SEMG信号确定步态周期的起始位置,而步态是由多块肌肉共同作用的结果,使用单一肌肉不能准确地检测步态周期,下一步将尝试使用多块肌肉的SEMG信号来提高检测准确性。

### 参考文献

- [1] MEN Huichao, WANG Botao. High feasible method of cycle segment analysis for common gait model [J]. Journal of Software, 2019, 30(3): 700-717. (in Chinese) 门慧超,王波涛. 面向通用模型的高可用性步态周期分析方法[J]. 软件学报, 2019, 30(3): 700-717.
- [2] WANG Jing, WU Xiaoming. Gait analysis based on surface electromyography [J]. Journal of Clinical Rehabilitative Tissue Engineering Research, 2012, 16(26): 4834-4840. (in Chinese) 王静,吴效明. 基于表面肌电的步态分析[J]. 中国组织工程研究, 2012, 16(26): 4834-4840.
- [3] ZHANG Ruihong, WANG Rencheng. Analysis and measurement of a lower limb surface EMG during walking [J]. Journal of Tsinghua University (Science and Technology), 2000, 40(8): 73-76. (in Chinese) 张瑞红,王人成. 人体下肢表面肌电信号的检测与分析[J]. 清华大学学报(自然科学版), 2000, 40(8): 73-76.
- [4] ZHAO Mandan, FAN Caizhi, ZHANG Zhenjie, et al. A method for hand movement onset detection from surface EMG signals [J]. Journal of University of Electronic Science and Technology of China, 2016, 45(6): 898-904. (in Chinese) 赵漫丹,范才智,张振杰,等. 基于表面肌电信号进行手部动作发起检测方法[J]. 电子科技大学学报, 2016, 45(6): 898-904.
- [5] GENG Xiaobo, YANG Peng, WANG Xinran, et al. Recognition of walking stance phase and swing phase based on moving window [J]. Journal of Biomedical Engineering, 2014, 31(2): 273-278. (in Chinese) 耿晓博,杨鹏,王欣然,等. 基于移动窗的人体下肢运动支撑期和摆动期的识别[J]. 生物医学工程学杂志, 2014, 31(2): 273-278.
- [6] LEI Min, WANG Zhizhong. The study advances and prospects of processing surface EMG signal in prosthesis control [J]. Chinese Journal of Medical Instrumentation, 2001, 25(3): 156-160. (in Chinese) 雷敏,王志中. 肌电假肢控制中的表面肌电信号的研究进展与展望[J]. 中国医疗器械杂志, 2001, 25(3): 156-160.
- [7] WANG Jian, JIN Dewen. Application of surface electromyography in the field of rehabilitation medicine [J]. Chinese Journal of Rehabilitation Medicine, 2006, 21(1): 6-7. (in Chinese) 王健,金德闻. 康复医学领域的表面肌电应用研究[J]. 中国康复医学杂志, 2006, 21(1): 6-7.
- [8] VAISMAN L, ZARIFFA J, POPOVIC M R. Application of singular spectrum-based change-point analysis to EMG-onset detection [J]. Journal of Electromyography and Kinesiology, 2010, 20(4): 750-760.
- [9] MENG Ming, SHE Qingshan, LUO Zhizeng. The application of HMM in gait recognition using lower limb SEMG [J]. Journal of Huazhong University of Science and Technology Nature Science, 2011, 39(S2): 176-179. (in Chinese) 孟明,余青山,罗志增. HMM在下肢表面肌电信号步态识别中的应用[J]. 华中科技大学学报(自然科学版), 2011, 39(S2): 176-179.
- [10] ZHANG Qizhong, XI Xugang, LUO Zhizeng. A pattern recognition method for surface electromyography based on nonlinear features [J]. Journal of Electronics & Information Technology, 2013, 35(9): 2054-2058.
- [11] YANG Peng, ZHOU Lihong, CHEN Lingling, et al. Walking gait cycle recognition based on peak-valley piecewise integrator algorithm [J]. Computer Engineering, 2011, 37(23): 168-170. (in Chinese) 杨鹏,周丽红,陈玲玲,等. 基于峰-谷分段积分算法的行走步态周期识别[J]. 计算机工程, 2011, 37(23): 168-170.
- [12] SANTELLO M, MCDONAGH M J. The control of timing and amplitude of EMG activity in landing movements in humans [J]. Experimental Physiology, 1998, 83(6): 857-874.
- [13] CHENG Juan, CHEN Xun, PENG Hu. An onset detection method for action surface electromyography based on sample entropy [J]. Acta Electronica Sinica, 2016, 44(2): 479-484. (in Chinese) 成娟,陈勋,彭虎. 基于样本熵的肌电信号起始点检测研究[J]. 电子学报, 2016, 44(2): 479-484.

(上接第 276 页)

- [14] RICHMAN J S, MOORMAN J R. Physiological time-series analysis using approximate entropy and sample entropy [J]. *American Journal of Physiology Heart & Circulatory Physiology*, 2000, 278(6):2039-2049.
- [15] LOU Zhi, YAO Bo, YANG Jihai. Detection of gait activity segment in children with cerebral palsy based on surface electromyography [J]. *Journal of Biomedical Engineering*, 2017, 34(2):342-349. (in Chinese)  
娄智,姚博,杨基海. 基于表面肌电信号的小儿脑瘫步态活动段检测研究[J]. *生物医学工程学杂志*, 2017, 34(2): 342-349.
- [16] MUKHOPADHYAY S, RAY G C. A new interpretation of nonlinear energy operator and its efficacy in spike detection [J]. *IEEE Transactions on Biomedical Engineering*, 1998, 45(2):180-187.
- [17] FOX E B, HUGHES M C, SUDDERTH E B, et al. Joint modeling of multiple time series via the beta process with application to motion capture segmentation [EB/OL]. [2020-02-10]. <https://arxiv.org/abs/1308.4747>.
- [18] JAIN S, NEALY R M. Splitting and merging components of a nonconjugate Dirichlet process mixture model [J]. *Bayesian Analysis*, 2007, 2(3):445-472.
- [19] TU Zhuowen, ZHU Songchuan. Image segmentation by data-driven Markov chain Monte Carlo [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 2002, 24(3): 657-673.
- [20] HUANG Zhexue. Extensions to the k-means algorithm for clustering large data sets with categorical values [J]. *Data Mining and Knowledge Discovery*, 1998, 2(3): 283-304.