

DOI:10.19651/j.cnki.emt.2210865

基于改进动态时间规整算法的终端波形比对方法^{*}

杨雄¹ 郭佳豪¹ 方鑫¹ 张旭辉²

(1. 国网江苏省电力公司电力科学研究院 南京 211103; 2. 南京工程学院 南京 211167)

摘要: 为深入挖掘配电终端录波性能分析,提出了一种用于一二次深度融合设备检测终端录波平台的改进型动态时间规整算法。通过分帧及加窗完成对波形的预处理,计算源信号波形与受噪声干扰的终端录波波形的短时能熵比;将两组波形的短时能熵比序列作为输入测试向量,利用DTW对两个波形的能熵比序列进行路径规划并求解相似度;计算两组短时能熵比序列的公共子串长度定义优化匹配系数,修正波形相似度。实验仿真和实测数据分析结果表明,结合能熵比与公共子串的DTW算法提升了算法的计算效率和准确率。实验证明该方法能为配电终端的录波性能评估提供数据支撑。

关键词: 改进型DTW;波形相似度;配电终端;公共子串;短时能熵比

中图分类号: TM743 **文献标识码:** A **国家标准学科分类代码:** 510.4030

Method based on optimized dynamic time warping algorithm for waveform comparison in terminal

Yang Xiong¹ Guo Jiahao¹ Fang Xin¹ Zhang Xuhui²

(1. State Grid Jiangsu Electric Power Company Electric Power Research Institute, Nanjing 211103, China;

2. Nanjing Institute of Technology, Nanjing 211167, China)

Abstract: Aiming at deeply exploring the analysis of the recording performance of the distribution terminal, this paper proposes an improved dynamic time warping algorithm for detecting the terminal recording platform of the primary and secondary deep fusion equipment. Complete the waveform preprocessing by framing and windowing, and calculate the short-term energy entropy ratio of the source signal waveform and the terminal recorded waveform affected by noise interference; The short-time energy entropy ratio sequences of the two waveforms were used as the input test vector, and the path planning and similarity calculation are carried out by using DTW for the two waveforms. And by computing two groups of short-time energy entropy than public substring sequence length defined optimization matching waveform similarity coefficient of correction. Experimental simulation and measured data analysis results show that the DTW algorithm combining energy-entropy ratio and common substring can accurately improve the computational efficiency and accuracy of the algorithm. Experimental results show that this method are conducive for the evaluation of recording performance of power distribution terminals.

Keywords: improved DTW; waveform similarity; power distribution terminal; common substring; short-term energy entropy ratio

0 引言

配电终端作为配电网中的重要感知单元,覆盖率随配电网建设的发展不断增长^[1],且大多安装在室外,易受外部环境干扰。其录波功能在系统发生故障时,自动准确地记录故障前、后过程中存在的各种电气量变化,所录波形是故障分析的重要依据,因此有必要对终端波形进行检测。时间

序列数据中,异常数据往往携带着大量有效信息,本文在此基础上对终端所录波形与功率源标准波形进行比对,以波形相似度为指标对终端的录波性能进行分析。

随着产品的复杂化,在频域进行信号分析更加全面便捷,可以及时发现^[2],相似性衡量论,是用来对比某些形态、图像或各种数据之间相似关系的理论^[3-5]。在音频对比、图像处理等领域有着重要的应用价值^[6-9]。相似度算法

收稿日期:2022-07-28

* 基金项目:国网江苏省电力有限公司科技项目(J2021016)资助

不断丰富,选择合适的算法对不同场景下的相似度判定有着重要意义^[10-11]。动态时间规整(dynamic time warping, DTW)算法是一种经典优化算法,通过使用符合特定要求的时间规整函数定义测试模块与参照模块之间的时刻相对关联,计算两模块匹配时的最小累计距离^[12-14]。

目前,各领域科研技术人员对 DTW 算法开展大量研究。冉启胜等^[15]引入加权函数对 DTW 算法改进以抑制边缘效应,但该方法只针对首端和末端优化,不适用于终端波形的全局比对。李正欣等^[16]提出了分段线性拟合的动态时间弯曲相似性度量,使用 DTW 对多维分段拟合后的时间序列进行度量,在数据量大的序列上表现出较好的效果,但实验结果受参数的选择影响较大。王语涵等^[17]通过计算短时自相关函数降低了 DTW 的复杂度,但是该方法不能有效体现时间序列中的突变事件。

本文提出了一种公共子串与能熵比相结合的 DTW 算法,分析一二次融合设备终端采样波形准确率。首先,针对计算复杂度高及特征事件不够突出的问题,提出对波形进行预处理的方法,该方法对源信号波形与受噪声干扰的波形进行分帧,依次计算两组波形的短时能熵比作为输入;然后,针对波形比对中的“病态匹配”,提出了修正相似度的方法,通过计算基于最长公共子串长度的惩罚系数达到目的;最后,验证及应用了改进的 DTW 算法,利用 Matlab 对终端所录波形进行比对,实验表明该算法有效提高了波形相似度的计算效率和准确率。

1 理论依据

1.1 DTW 算法理论

两个波形序列 X 和 Y , 长度是 $|X|$ 和 $|Y|$, 对其进行相似度比较。求一条曲线 $W = w_1, w_2, \dots, w_k$ 为规整路径, 曲线长度大于两序列的长度且小于两序列长度之和。 W 从 $w_1 = (1, 1)$ 起始, 一直延伸至 $w_k = (|X|, |Y|)$, i 和 j 值单调递增以确保曲线不交叉, 即:

$$w = (i, j), w_k = (i', j'), i \leq i' \leq i + 1, j \leq j' \leq j + 1 \quad (1)$$

规整代价最小的路径为:

$$DTW = \left\{ \sqrt{\sum_{k=1}^K w_k / K} \right\} \quad (2)$$

获得的归整路线是相对距离最短的归整路线:

$$D(i, j) = Dist(i, j) + \min[D(i-1, j), D(i, j-1), D(i-1, j-1)] \quad (3)$$

路径从左下角采用步进模式一直匹配至右上角, 即扭曲曲线^[18], 扭曲曲线保证匹配路径过程中经过的元素值总和最小。DTW 距离计算过程中, 波形的单位点复制达到局部范围内的调整, 在此基础上对齐匹配^[19-21]。

规划路径满足以下约束:

$$g(i, j) = \min \left\{ \begin{array}{l} g(i-1, j) + d(i, j) \\ g(i-1, j-1) + 2d(i, j) \\ g(i, j-1) + d(i, j) \end{array} \right\} \quad (4)$$

式中: $g(i, j)$ 表示两个波形序列的距离不断叠加得到最小值 $dist(X, Y)$, 再把相似距离归一化后通过 $\omega(X, Y) = \frac{1}{1 + dist'(X, Y)}$ 转化为相似度, $dist'(X, Y)$ 为归一化的距离。

1.2 波形预处理

终端录波采样频率高, 采样点多, 加上 DTW 算法匹配路径过程复杂, 导致计算效率低下, 而短时能熵比作为一种时域分析算法, 在不丢失关键信息的基础上不仅可以通过分帧, 加窗提高算法的计算效率, 而且其能显著提高信噪比的特点更可以凸显波形信号中的突变事件^[22]。因此, 采用短时能熵比对波形信号预处理。

计算过程如下:

1) 设长度为 N 的波形序列为 $x(n)$, $n = 1, 2, \dots, N$, 对波形序列进行除直流分量、幅值归一化处理。

2) 选取合适的频率及帧移, 对初始波形序列进行分帧及加窗, 常见的加窗函数如下:

$$\text{矩形窗: } \omega(n) = 1 \quad (5)$$

$$\text{海宁窗: } \omega(n) = 0.5 - (1 - \cos(2\pi n / (L - 1))) \quad (6)$$

$$\text{汉明窗: } \omega(n) = 0.54 - 0.46 \cos(2\pi n / (L - 1)) \quad (7)$$

其中, L 为窗的长度, 且 $0 \leq n \leq L - 1$ 。

将 $\omega(n)$ 分帧, 得到第 i 帧波形信号为 $y_i(n)$:

$$y_i(n) = \omega(n)x((i-1)i_{nc} + n) \quad (8)$$

$$1 \leq n \leq \omega_{len}, 1 \leq i \leq f_n$$

式中: $y_i(n)$ 表示加窗范围内的波形信号一帧的数值; ω_{len} 表示帧长, i_{nc} 表示帧移长度; f_n 表示总帧数。

3) 对分帧后的信号 $y_i(n)$ 进行 FFT 变换, 得到第 k 条谱线频率分量 f_k 的能量谱 $Y_i(k)$, 信号每一段频率分量对应的谱概率密度为 $p_i(k)$:

$$p_i(k) = Y_i(k) / \sum_{i=0}^{N/2} Y_i(k) \quad (9)$$

式中: $p_i(k)$ 表示第 i 帧第 k 个频率分量 f_k 对应的概率密度。

4) 第 i 帧的谱熵为 H_i :

$$H_i(k) = - \sum_{i=0}^{N/2} p_i(k) \ln p_i(k) \quad (10)$$

第 i 帧的能量为 E_i :

$$E_i(k) = \lg(1 + \sum_{n=1}^N x_i(n)/a) \quad (11)$$

式中: a 为常数, 根据不同的波形能量变化程度调节。

5) 得到能熵比的值 R_i , 完成预处理:

$$R_i = \sqrt{1 + |E_i/H_i|} \quad (12)$$

依次将 2 组波形的短时能熵比输入 DTW, 得到对应的动态路径及距离 $dist_m(X, Y)$ 。

1.3 相似度修正方法

在基于 DTW 算法的相似度匹配运算中, 发现在步进运算流程中, 路径走向只能保持短暂的单调特性, 易产生波

峰波谷走向不同的“病态匹配”。考虑到录波过程电压幅值会存在突变的问题,有必要对波形相位进行局部范围内的调整,解决“病态匹配”的负面影响。定义一种惩罚系数对波形的距离调节,用惩罚系数乘以距离求得新的距离。具体方法如下:

1)计算最大标准差 sd_{max} 。首先计算波形序列的平均值 x ,则标准差 sd 的计算,如式(13)~(14)所示。

$$sd_x = \sqrt{\frac{\sum_{i=1}^n (x_i - x)^2}{n}} \quad (13)$$

同样:

$$sd_y = \sqrt{\frac{\sum_{i=1}^n (y_i - y)^2}{n}} \quad (14)$$

式中: n 代表波形序列 X 的元素数量,最大标准差取两个时间序列中标准差较大的一个,即

$$sd_{max} = \max(sd_x, sd_y) \quad (15)$$

2)定义偏移范围。 X 与 Y 为数值序列,序列长度分别为 a, b 。把最大标准差设定为合适的偏移范围,即两个元素都在范围内时,将这两个元素纳入公共子串。

定义矩阵 $dq[i][j](0 \leq i < a), (0 \leq j < b)$:

$$dq[i][j] = \begin{cases} 0, & X[i] - Y[j] \geq sd_{max} \\ 1, & X[i] - Y[j] < sd_{max} \text{ 且 } j = 0 \\ dq[i-1][j-1] + 1, & X[i] - Y[j] < sd_{max} \text{ 且 } i, j > 0 \end{cases} \quad (16)$$

用动态规划求解矩阵中的路径,路径中的元素在小于偏移容忍时不断叠加,得到的数值即为公共子串的长度 1。

3)定义惩罚系数 α ,如式(17):

$$\alpha = 1 - \frac{l}{\min(\text{len}(X), \text{len}(Y))} \quad (17)$$

4)按式(18)、(19)进行距离算法优化。

$$dist_n(X, Y) = \alpha \times dist_m(X, Y) \quad (18)$$

$$\omega'(X, Y) = \frac{1}{dist'_n(X, Y)} \quad (19)$$

$dist'_n(X, Y)$ 为归一化的优化后距离,在改进的 DTW 算法下,波形匹配得以区域化,缩减了公共部分的相似距离,有效解决“病态匹配”对整体精度的影响。

1.4 算法流程

综上,提出一种结合能熵比与公共子串的 DTW 优化方法,算法流程如图 1 所示。

具体步骤:

1)选取合适的窗函数及帧长计算两组波形的能熵比序列,完成对波形序列的预处理。

2)将得到的两组能熵比序列输入 DTW,构造矩阵,得到两组波形的相似度。

3)计算能熵比序列的公共子串,得到优化匹配系数,对波形相似度修正并输出结果。

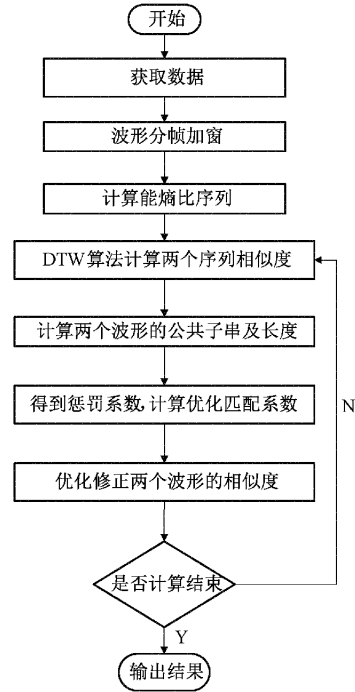


图 1 算法流程

2 实验验证

2.1 波形比对平台

以搭建的一二次深度融合设备检测平台为研究对象,其录波系统如图 2 所示。

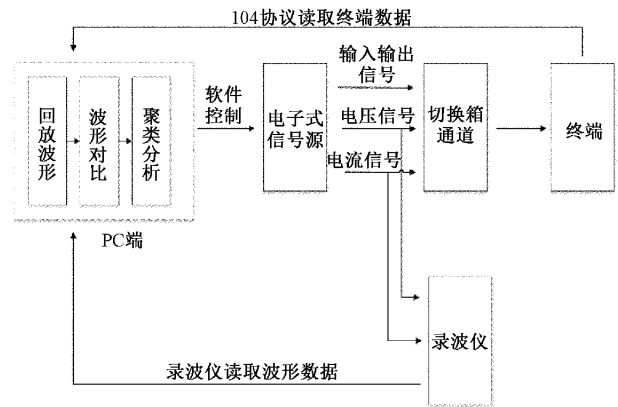


图 2 终端录波系统

PC 机通过协议控制电子式信号源,电子式信号源输出波形给终端,PC 机通过 104 协议读取波形数据,采样速率设置为 200 kS/s。PC 机设置录波仪触发方式后,录波仪自动采集电子式信号源输出的波形。

PC 机读取录波仪波形数据和终端波形数据后生成波形时间序列,将功率源波形与终端采样波形进行对比,得到两波形的相似度,采集大量不同状态下的样本,对样本数据进行聚类,得出相似度标准,将相似度分为相似,疑似相似及不相似,对应终端的 3 种录波性能状态。在之后判断终

端录波性能时,对采集后的波形进行相似度分析后对样本数据库,判断终端的实时录波性能状态,本文主要涉及其中的波形比对算法。

为验证算法有效性,选取录波平台实测终端波形,对录波平台的录波仪采样波形在 PC 端进行回放得到源信号波形如图 3 所示,采样中期存在电压突变。对录波平台中终端采样得到的受噪声干扰的波形在 PC 端进行回放得到受噪声干扰的终端波形如图 4 所示,突变事件被噪声覆盖。选取前 5 后 9 共 14 个周期作为采样周期,通过对比不同噪声强度下的波形来验证算法的可行性。

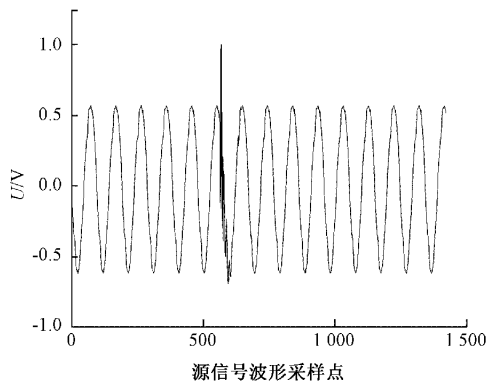


图 3 源信号的电压波形图

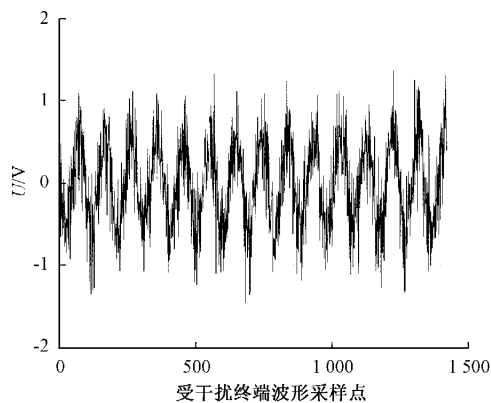


图 4 受噪声干扰的终端波形

2.2 结合公共子串与能熵比的 DTW 算法的应用

针对计算复杂度高的问题,计算波形序列的短时能熵比,对图 3 和 4 的采样波形进行加窗分帧,计算两个波形的短时能熵比,提取其特征,分别得到源信号波形能熵比序列如图 5 所示,受干扰波形能熵比如图 6 所示。

帧长是短时能熵比分辨率的决定因素。当信号采样频率不变时,帧长越小,短时能熵比的时间分辨率越高,但帧长过短不利于发挥短时能熵比的优势来提高信噪比。因此,帧长的选择应考虑信号的采样频率、时间分辨率要求和信噪比要求,以选择合适的值。通过观察采集的波形信号,可以发现 30 ms 内的振动信号相对稳定。当帧长较小时,短时能熵比的平滑效果不明显。帧长越大,平滑效果越好,

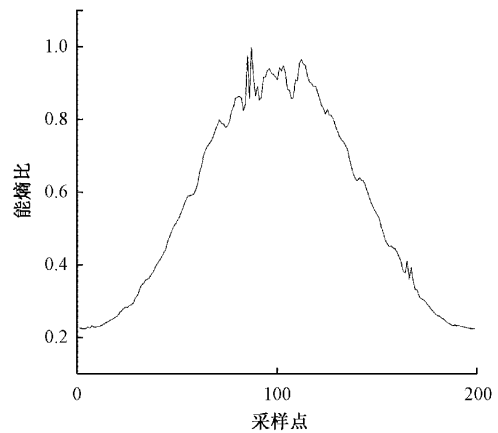


图 5 源信号波形能熵比

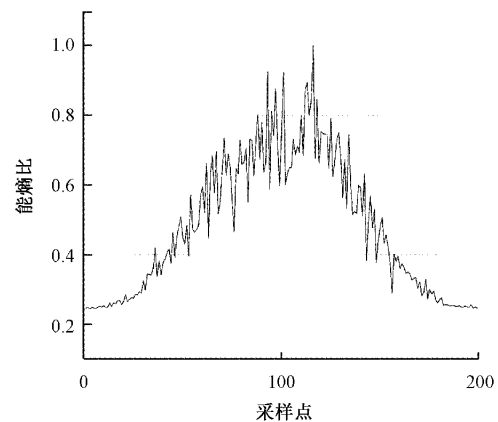


图 6 受干扰波形能熵比

这有利于 DTW 算法进行路径规划。将帧长设置为 25 ms 可以很好地保留特征事件。设置帧长 ω_{len} 为 25 ms, 帧帧 $i_{nc}=20$ 。

加窗后的每一小段为一个新的采样单位,发现图 4 中的波形在加入噪声干扰后失去了冲击事件的特征,而在图 6 中,短时能熵比序列仍能在采样中期突出波形中的冲击事件。

将图 5 和 6 得到的能熵比序列作为输入测试向量,利用 DTW 对两个能熵比序列进行扭曲及路径规划,得到源信号波形与受噪声干扰的终端波形经短时能熵比处理后在 DTW 算法下的点位匹配如图 7 所示,图 3 和 4 的两个波形作为测试向量输入 DTW 得到的动态路径规划图如图 8 所示,这两个波形的能熵比序列的动态路径规划图如图 9 所示。

对比初始路径规划图,在通过加窗减少了采样点的同时突出了录波瞬间的突变事件。从图中发现动态路径规划过程仍存在病态匹配,即在源波形采样点为 100 和受干扰波形采样点为 136 时,存在一点匹配连续多点的情况。对此,结合基于公共子串的惩罚系数对距离修正。

结合公共子串与能熵比的 DTW 算法的运算流程中各参数以及结果如表 1。

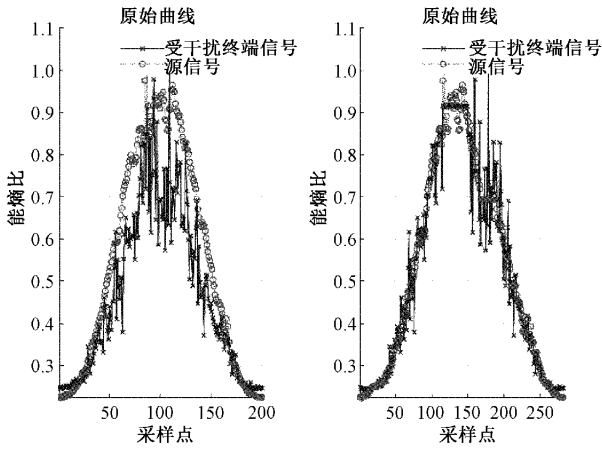


图 7 能熵比序列的点位匹配图

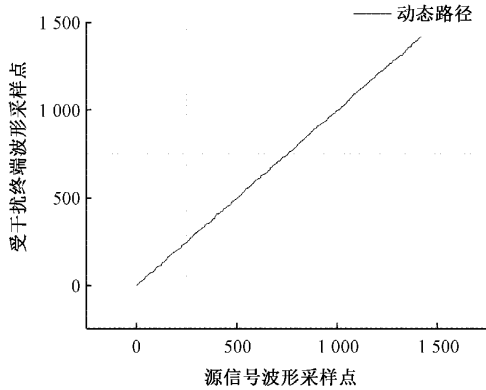


图 8 初始序列的点位匹配图

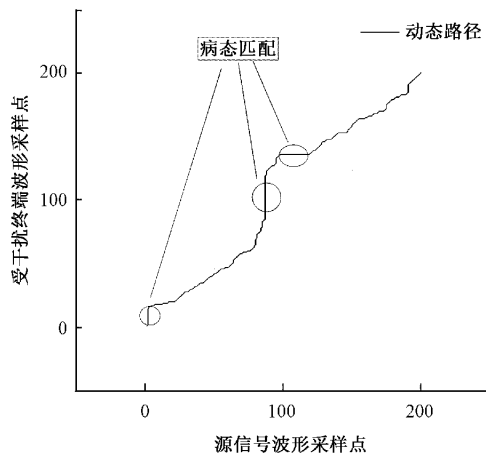


图 9 能熵比序列的路径规划

表 1 结合公共子串与能熵比的 DTW 算法的波形距离

噪声方差	$dist_m$	α	$dist_n$
0.1	9.272	0.570	5.285 1
0.2	10.915	0.535	5.839 7
0.3	10.863	0.560	6.083 2
0.5	12.691	0.570	7.233 9

表中 $dist_m$ 为经能熵比预处理后的相似距离, $dist_n$ 为结合能熵比和公共子串的 DTW 算法下的相似距离。分析距离发现, 结合能熵比与公共子串的 DTW 算法有效提高了波形相似度计算的准确率。随着噪声信号方差增加, 其相似度逐级递减, 距离逐渐增加, 在方差分别为 0.2 和 0.3 时优化前距离分别为 10.915 和 10.863, 对波形序列采用公共子串对其“病态匹配”情况进行修正后的距离为 5.839 7 和 6.083 2, 符合预期。

2.3 算法性能比较

路径规划过程中存在的“病态匹配”经公共子串使波形相位适度迁移, 实现区域走向相似性的特性得以修正。以“病态匹配”的比例作为评价算法好坏的标准, 通过数值来验证算法的准确率:

$$rate = \frac{lnc - (ff - xx)}{lnc} \times 100\% \quad (20)$$

式中: lnc 为匹配点总长度, ff 为原始“病态匹配”点长度, xx 为“病态匹配”点在公共子串算法下修正的长度。

针对传统的 DTW 算法、基于能熵比的 DTW 和结合能熵比与公共子串的 DTW 算法在不同白噪声环境下进行比较, 得到的准确率结果如表 2 所示。

表 2 不同算法的准确率比较

噪声方差	不同算法的准确率/%		
	DTW	基于能熵比的 DTW	改进型 DTW
0.1	80.9	70.2	95.4
0.2	77.5	72.7	98.2
0.3	80.6	71.2	97.2
0.5	80.0	70.0	95.8

对比不同算法的准确率发现, 传统 DTW 算法准确率在 80% 左右, 能熵比处理后的 DTW 算法准确率为 70% 左右, 结合能熵比与公共子串的 DTW 算法的准确率可以达到 95%。说明本文提出的能熵比与公共子串有效契合, 优化后的 DTW 算法很大程度解决了“病态匹配”。

在 Matlab 平台进行波形比对, 记录不同采样点数量在不同算法下计算的响应时间, 传统 DTW 算法及结合能熵比与公共子串的 DTW 算法随采样点数量增加, 响应时间对比如图 10 所示。

由于数据经预处理后响应时间相差小, 主要时间差别在计算能熵比环节, 所以整体性能随节点数增加, 响应时间变化不大, 优势更明显。说明通过短时能熵比对波形进行预处理有效提高了计算效率。

计算 500 组不同噪声程度的终端波形与源波形的距离, 以距离为参数, 利用 K-means 对其进行聚类。样本集聚类结果如图 11 所示, 归一化后的聚类中心为 0.306 9 和 0.610 5, 经式(19)换算得到各等级相似度范围, 相似度大于 76.52% 时, 波形判定为相似, 相似度小于 62.09% 时, 波

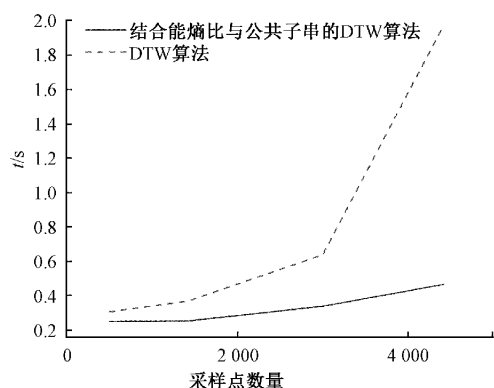


图 10 优化前后算法的响应时间

形判定为不相似,相似度处于 62.09%~76.52%之间时,判定为疑似相似。

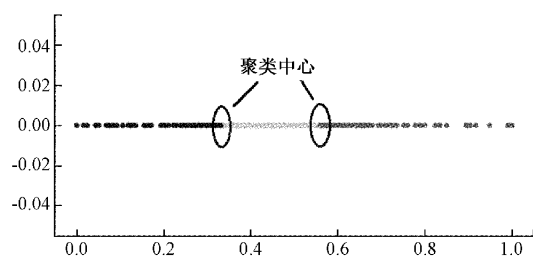


图 11 K-means 聚类结果

3 结 论

电压电流波形的相似性挖掘,对于电气科学研究领域有着重要意义。文章对终端波形在不同噪声下的时间序列进行相似性数据分析,优化后达到以下效果:

利用短时能熵比对波形数据进行预处理,很好地突出了波形中的冲击事件,并进行特征提取;降低输入特征的复杂度,提升了在波形数据处理中的相似性匹配运算效率。通过采用最长公共子串的匹配方法,对 DTW 优化距离进行计算,克服了周期性相位不匹配的距离计算中的“病态匹配”问题,提高了算法的准确率。

该方法为终端录波性能分析提供数据支撑和决策依据,对之后通过海量数据探究不同干扰情况下波形相似度,实际应用中,终端的波形不仅受噪声影响,后期需针对多种影响因素构造评价体系。

参考文献

- [1] 邢晓敏,徐海瑞,廖孟柯,等. 基于云模型和 D-S 证据理论的配电终端健康状态综合评估方法[J]. 电力系统保护与控制,2021,49(13):72-81.
- [2] 伍弘毅,陈志聪,吴丽君,等. 利用加速度信号时频域特征的枪击识别研究[J]. 电子测量与仪器学报,2022,36(5):180-187.
- [3] 韩瑞琪. 基于网络理论的时间序列分析方法及其应用[D]. 上海:华东理工大学,2020.
- [4] 何珍文,吴冲龙,刘刚,等. 地学时序大数据的相似性度量与索引方法综述[J]. 地质科技通报,2020,39(4):44-50.
- [5] 黄琼桃,刘瑞敏. 云模型的相似性度量综述[J]. 数据通信,2019(6):43-49.
- [6] GAO D, KAO H, WANG B. Misconception of waveform similarity in the identification of repeating earthquakes[J]. Geophysical Research Letters, 2021, 48(13), DOI:10.1029/2021GL092815.
- [7] TSUCHIYAMA A, NAKAJIMA J. Diversity of deep earthquakes with waveform similarity[J]. Physics of the Earth and Planetary Interiors, 2021, DOI:10.1016/j.pepi.2021.106695.
- [8] 余腾,胡伍生,孙小荣,等. 基于非下采样 Contourlet 变换耦合能量相似制约的遥感图像融合算法[J]. 电子测量与仪器学报,2021,35(6):71-78.
- [9] 杜政奇,王敬华,张新慧. 基于参数优化 VMD 和能量相似度的配电网故障区段定位方法[J]. 电子测量技术,2022,45(8):95-101.
- [10] 潘晓,马昂,郭景峰,等. 基于时间序列的轨迹数据相似性度量方法研究及应用综述[J]. 燕山大学学报,2019,43(6):531-545.
- [11] 李国琴,王瑾,谭艳丽,等. 基于图像相似度的多权重图谱 DTI 自动分割算法研究[J]. 电子测量技术,2020,43(6):116-122.
- [12] 程前,李建良. 基于改进动态时间规整的相似性度量及轨迹聚类[J]. 电子测量技术,2021,44(23):1-5.
- [13] 朱田,彭之川,张智腾,等. 一种基于动态时间规整的闭环检测算法[J]. 控制与信息技术,2021(4):83-88.
- [14] 叶科淮,陈志,王仁杰,等. 动态时间规整算法优化[J]. 软件导刊,2021,20(1):132-135.
- [15] 冉启胜,张哲,韩杰祥,等. 基于改进动态时间弯曲距离算法的直流配电网线路纵联保护方案[J]. 电力自动化设备,2022,9(30):1-14.
- [16] 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.
- [17] 王语涵,闫子薇,武天琦,等. 基于 DTW 算法对婴幼儿语音的分析[J]. 齐齐哈尔大学学报(自然科学版),2019,35(4):69-73.
- [18] 金展,康瑞清,王威. 基于地磁序列匹配的大型室内动态定位方法[J]. 仪器仪表学报,2021,42(7):199-207.
- [19] 陈春华,李薇,陈雅莉. 基于优化 DTW 算法的水文要素时间序列数据相似性分析[J]. 水文,2021,41(3):44-48.
- [20] 逮宇佳,曹俊兴,田仁飞,等. 基于动态时间规整 ICA 算法地震随机噪声压制[J]. 石油物探,2018,57(5):

697-704.

- [21] 刘苗苗,周从华,张婷. 基于分段特征及自适应加权的 DTW 相似性度量[J]. 计算机工程, 2021, 47(8): 62-68, 77.
- [22] 万书亭,马晓棣,陈磊,等. 基于振动信号短时熵比与 DTW 的高压断路器状态评估及故障诊断[J]. 高电压技术, 2020, 46(12): 4249-4257.

作者简介

杨雄, 博士, 高级工程师, 主要研究方向为电力系统运行分析与控制。

郭佳豪, 硕士, 助理工程师, 主要研究方向为配电自动化。

方鑫, 硕士, 工程师, 主要研究方向为配网数字化技术。

张旭辉, 硕士在读, 主要研究方向为电力自动化。

E-mail: 1397703135@qq.com