

状態監視時系列データの類似検索に関する検討

A Method for Similarity-Based Search in Time-Series Sensor Data

藤野 友也† 菅野 幹人† 撫中 達司†
Tomoya Fujino Mikihiro Kanno Tatsuji Munaka

1. はじめに

近年、異常の早期発見を目的として、工場やビルなどで動作中の設備機器の状態を、常時観測する取組みが数多くなされている。観測されたそれらのアナログ値の変化を示す時系列データは一般に膨大であるが、後の活用を想定して蓄積されることが多い。時系列データの活用の際には、過去に蓄積された膨大な時系列データの中から、目的のパターンを抽出する検索を行うことが必須である。しかし、時系列データの検索は、リレーショナルデータベースにおけるレコード検索とは異なり、指定された値の完全一致により検索目的が達成されることは少なく、連続した値の系列・パターンに対して、ある程度の揺らぎを許容して、完全一致ではなく類似性を評価して検索を行う必要がある。

本論文では、大量の時系列データの局所波形の特徴を抽出して類似検索を行う方式を用いて、特徴波形の抜けや強弱に応じた類似度評価を取り入れた方式について報告する。

2. 従来手法と課題

センサ情報など、高周波成分と低周波成分が混在する波形を確認する際の人の判断基準は、滑らかな変化をしている場合は値自身、高周波で振動した波形の場合は高周波部分の周波数とその包絡線である。しかし、通常の時系列類似検索では、そのような周波数の変化に応じた類似性の評価を行うことはなく、蓄積された膨大な時系列データに対して、柔軟な時系列シーケンスの検索を行うと、過大な処理時間を要するという課題がある。

周波数に注目した検索方式としては、フーリエ変換やウェーブレット変換による情報圧縮を行い、圧縮されたデータを検索することで、検索に必要な演算量を減少させる方式が知られている。また時系列類似検索の最近の技術として、時間方向の歪みを吸収して類似判定を行い検索するダイナミックタイムワーピング(Dynamic Time Warping: DTW)方式[1]が知られている。

しかし、ウェーブレット変換を用いた情報圧縮による検索演算量軽減では、高周波成分の包絡線の特徴などに応じた検索ができない。一方、ダイナミックタイムワーピングを用いた手法では、高周波成分を持たない波形については時間方向に柔軟な検索が実現できるが、高周波成分を持つ場合には、周波数を変更する形でマッチングが行われるため、共振周波数など、高周波成分の重要な要素が無視される。またダイナミックタイムワーピングを用いた検索は演算処理が多いという課題があった。

3. 提案手法

3.1 要約類似検索モデル

本論文にて提案する要約類似検索モデルの基本的なアプローチは、以下の通りである。

(1) 時系列データの系列を、変化が緩やかな区間では値、変化が急峻な区間では周波数および包絡線に注目し、それぞれの変曲点で区切った区間を“均一特性区間”とする。

(2) 均一特性区間に対して、高周波成分および低周波成分に関する波形の、周波数や包絡線などの特徴をサマリとして要約した“要約要素”を算出する。

(3) 2つの時系列データの類似性を、それぞれを要約要素に変換した要約要素系列同士を比較して得られる類似性を評価し、評価値が閾値を超えるものを検索結果としてリストアップする。

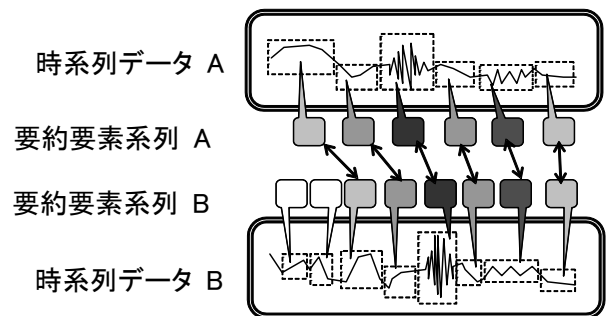


図1 類似性評価方式

時系列データから要約要素系列を前処理として求め、その系列に対して検索を行うことにより、状態監視データでは重視される高周波における周波数の不一致や、低周波における時間の長短の不一致を解消した、柔軟な類似検索が可能となる。

3.2 要約要素

要約要素の内容としては、均一特性区間ごとに、種類項目、変化項目、パラメータ項目および発生時間項目の4種類の項目を設定する。

種類項目が取り得る値は“観測値”、“パルスノイズ”、“周波数”、“包絡線”とし、変化項目が取り得る値は“一定”、“単調変化”、“U字変化”、“ランダム”とする。パラメータには、種類項目と変化項目の内容に応じたパラメータを可変で指定し、発生時間項目は均一特性区間の開始時刻と継続時間を指定する。

例えば、図2に示されるように、時系列データを7つの均一特性区間(A)~(G)とした場合、区間(D)は高周波成分を持つので、種類項目の値が“観測値”、“周波数”、“包絡線”である3つの要約要素を持つ。それぞれの変化項目は、

† 三菱電機株式会社 情報技術総合研究所
Information Technology R&D Center, Mitsubishi Electric Corporation.

順に“一定”, “一定”, “U字変化”である. パラメータ項目には, それぞれの組み合わせで決定される値を持ち, 例えば“包絡線”と“U字変化”の場合は, 区間開始点での観測値を中心とした上下幅の初期値, 包絡線の初期傾き, 包絡線傾きの時間経過による増加量により定義する.

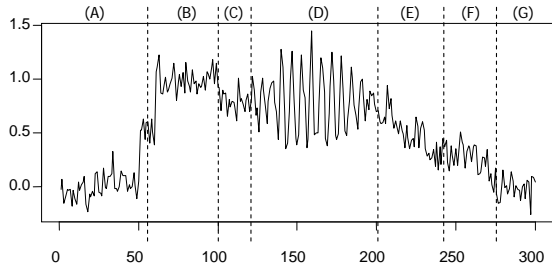


図2 時系列データと均一特性区間

3.3 類似検索

時系列データの類似評価は, 時系列データに対応する要約要素系列を基準として比較して行う. 要約要素系列の比較は, 要約要素のマッチングと, 要約要素同士の類似性判定の方式の2段階により行う.

要約要素系列の要約要素のマッチングは, 動的計画法を用いて行い, 類似性が最も大きいものを採用する.

要約要素の間で類似性の評価をする際, 種類項目の値に応じてそれぞれ, 値域, 周波数, 包絡線, 変化の方式, 継続時間の視点での差異を評価する. 比較する要約要素系列 A, B に対し, 値に関する要約要素を V_A, V_B , 周波数に関する要約要素を H_A, H_B , 包絡線に関する要約要素を E_A, E_B , それぞれの継続時間を T_A, T_B とするとき, 類似性評価値 D は以下の式により算出する.

$$D = \{a \text{ dist}(V_A, V_B) + b \text{ Hzdiff}(H_A, H_B) + c \text{ dist}(E_A, E_B)\} \times d \text{ rate}(T_A, T_B)$$

ここで, a, b, c, d は, 類似検索の各特性の重要性を示す重み係数である. $\text{dist}(A, B)$ は, A, B のパラメータから算出される 2 本の近似曲線間のユークリッド距離を示し, $\text{Hzdiff}(A, B)$ は A, B が示す主要周波数の差を示し, $\text{rate}(A, B)$ は, A, B が示す値(上式では継続時間)の比もしくは比の逆数で 1 以上のものを示す. これらの値は, 元データの個数によらず, 変化の方式とパラメータから定数時間で算出可能である. なおこの例では簡単のために a, b, c, d を定数としているが, 検索する時系列の周波数を引数とする関数としてよい.

状態監視においては, 観測された周波数と固有周波数の突き合わせによる判断や, 緩やかに変化する波形の変動の仕方などに注目することが多い. このため, 高周波成分を持つ場合は, 周波数の重みを大きく取る.

検索する時系列データに高周波成分を含む区間と含まない区間とが混在するケースでは, それぞれに用意された重みパラメータを用いて, 高周波成分では周波数に重みを置き, 低周波成分では値に注目する形式で算出する.

4. 検索例

擬似的に作成した長い時系列データ T (図 3) に対して, 短い時系列データ S (図 4) の類似検索結果を評価するため,

ダイナミックタイムワーピング(以下 DTW 手法)と提案手法による類似性評価を実施した. ここで, パラメータ重み $(a, b, c, d) = (1, 10, 5, 2.5)$ とする.

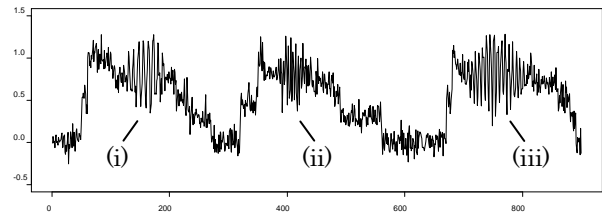


図3 検索先の時系列データ T

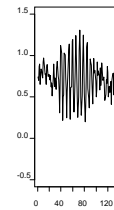


図4 検索する時系列データ S

評価の結果, いずれも図 3 の (i)~(iii) の強い振動部分で距離の極小値をとった. その距離の極小値を表 1 に示す.

表 1 類似性距離の極小値

	DTW 手法	提案手法
(i)	19.87	9.84
(ii)	19.10	22.50
(iii)	19.16	3.33

それぞれの主要周波数は(i)が 0.12Hz, (ii)が 0.23Hz, (iii)が 0.16 Hz であり, S の主要周波数は 0.16Hz である. DTW 手法では, (ii)が最も距離が小さい(類似性が高い)結果を与えるが, 主要周波数の観点では (ii) と最も差がある. 一方, 提案手法の方式では, 主要周波数成分に近い (iii) が最も近いと判定される. 一方, DTW 手法にて最も近いと判定された, 主要周波数が最も差がある (ii) は, 最も距離が大きく, 類似性は低いと判定される.

このように, DTW 手法は, 時間方向の揺らぎを許容する特性のために, 周波数成分を保存した類似評価は困難であるが, 本手法では, 周波数成分の一致性を重視した検索が可能であることを確認した.

5. まとめ

本論文では, 時系列データに対して, 高周波成分と低周波成分とで異なる評価方法で類似判定を行う検索方式を提案した. 本方式を用いることで, 固有周波数や減衰曲線などに重要な意味を持つ電子・機械系の保守事例などの検索の精度を向上させることができる.

今後の課題は, 類似算出時の重みなどのパラメータを, 学習により最適化する方式を検討することである.

参考文献

[1] 櫻井 保志, “ダイナミックタイムワーピング距離に基づくストリーム処理,” DEWS 2007, L6-5.