

プロセスデータを対象とした類似波形の高速検索法

A Method for Quickly Searching Similar Waveform Patterns in Historical Process Data

黒田 知宏^{*1}

Tomohiro Kuroda

大谷 哲也^{*2}

Tetsuya Ohtani

和田 英彦^{*3}

Hidehiko Wada

プラントのオペレータは過去の経験に基づきプラントの操作をすることが少なくない。そのため、蓄積された大量のデータから、現在のプラントデータに類似している過去のデータを簡単に検索することができると、オペレータの判断の助けになる。このような状況においては、大量のデータの中から類似する波形をいかに効率的に見つけることができるかが問われている。

本稿では、現在のプロセスデータの波形、あるいは、ヒストリカルデータのある区間の波形に注目し、注目している波形と類似の動きをしている波形を、蓄積されたデータの中から高速に検索する技術を紹介する。

Plant operators often rely on their experience when operating plants. However, they could make more appropriate decisions by quickly referring to plant data similar to the current condition among a huge archive of process data. In such cases, efficient referencing of data is crucial.

This paper introduces a technique to satisfy this need by focusing on the waveform patterns in current data and a certain period of historical data. This method can quickly identify and retrieve waveform patterns that match the target ones.

1. はじめに

石油、化学、鉄鋼、食品、薬品などのプラントにおいて、操業オペレータは過去の経験に基づき運転することが多い。経験に基づく運転を行う際に、蓄積された大量のプロセスデータから、現在のプラントの状況と関連するデータを検索することができると、その結果を参考にして効率の良いプラント運転を実現できる。例えば、連続生産プロセスにおいて性状の異なる銘柄へ生産を切り替える際に、過去に起こった同様の銘柄変更時の運転データを参考に運転するケースが考えられる。また、過去に発生した異常時のプロセスデータと現在のプロセスデータを比較することで、プラントの異常を事前に検知し、事故を未然に防ぐことが可能となる。これらを実現するためには、蓄積された大量の過去の運転データから、現在のプラントの状態と類似している状態を見つけ出す技術が必要となる。

類似している状態を見つけ出す方法としては、時系列に並んだプロセスデータ（以下、トレンドデータとする）の特定の区間を波形とみなし、注目している区間のトレ

ンドデータの波形と蓄積されたトレンドデータの波形の類似度合いを測る方法がある。

波形間の類似度合いは、トレンドデータ間のユークリッド距離により定量化できる⁽¹⁾。しかし、単純なユークリッド距離では、波形としては似ていても、平均値が異なる、あるいは位相がずれている場合にユークリッド距離が大きくなってしまい、類似性を反映しないケースが起こる。

実際にプラントにおいては、同じ製品を製造していても、季節によって測定値の平均値が変化することや、タンク内の物質の差によって反応の時間遅れが変わり、プロセス値の変化のタイミングや時定数が変わることが起こり得る。そのため、比較しているトレンドデータの波形が完全に一致することはなく、測定値軸方向と時間軸方向に対してある程度のずれを許容した波形間の類似度合いを定量化する必要がある。

位相のずれや時間軸方向の伸縮を許容するための方法として、Dynamic Time Warping⁽²⁾（以下、DTWとする）がある。DTWは異なるデータ数の2つの波形を比較することができ、時間軸方向のずれを許容するための方法として提案されている⁽³⁾。しかしながら、DTWでは、トレンド波形の一部を自由に伸縮するために全体の形が類似しない場合でも波形間の距離を小さく見積もることがある（図1）。図1では、横軸が時間方向を、縦軸がセンサの測定値方向を表している。図1の点線の波形は、上昇し始めるまでが実線の波形よりも長くなっている。しか

*1 マーケティング本部 イノベーションセンター
インキュベーション部

*2 ソリューションサービス事業本部 高度ソリューション事業部
高度ソリューションセンター RTO 部

*3 マーケティング本部 事業開発センター 技術戦略室

し、上がってから下がるという変化部分の形は似ているので、DTW では波形間の距離を小さく見積もってしまう。このような波形を類似していると判断してしまうのは、プロセスの挙動が似ていると判断する基準としては不適切であり、今回の目的には使えない。プロセスのより本質的な挙動を識別するための波形の類似性尺度が必要である。

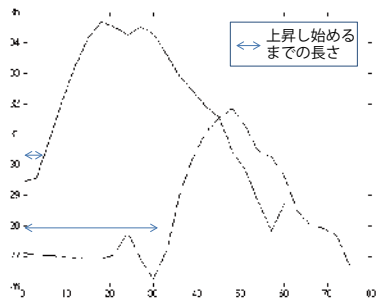


図 1 DTW を用いて類似波形を検索した時の例

本稿では、プラントの状態を識別するために有効となるトレンドデータの波形の類似度を算出する方法と、現在注目している波形（以下、参照波形とする）と類似の動きをしている波形を、蓄積されたデータの中から高速に検索する技術について述べる。

2. 類似波形検索法

蓄積されたトレンドデータから、参照波形と類似の動きをしている波形、つまり類似度合いが高い波形を検索する方法について説明する。2.1 節で類似度合いの算出方法について、2.2 節で高速検索法について説明する。

2.1 類似度算出法

参照波形と類似している波形を検索するために、波形間の距離と類似度を定義する。

2.1.1 波形間の距離

本稿では、波形間の距離としてユークリッド距離を用いる。しかしながら、プロセスにおいては、季節による測定値の変化、タンク内容量の差による応答時間や遅れの変化といったことが起こる。そこで、これらプロセス特有の挙動を考慮して、ある時刻の波形と参照波形の距離を計算する際に、測定値方向のオフセット、時間軸方向のシフトおよび伸縮を考慮する。これらオフセット、シフトおよび伸縮を考慮した時の距離の最小値を求め、この距離を参照波形と検索対象波形の2つの波形間の距離とする。オフセット、シフト、伸縮の内容を次に示す。

- オフセット：測定値方向の値のズレを許容する。具体的には、参照波形と検索対象波形の平均値が同じになるように補正する。
- シフト：時間方向の変化タイミングのズレを許容する。

具体的には、ある時刻の波形と参照波形の距離を計算する際に、その時刻の前後数点を始点とした波形との距離も算出し、一番距離が小さくなる時の距離を求める。

- 伸縮：時間方向の変化のズレを許容する。具体的には、ある時刻の波形と参照波形の距離を計算する際に、始点と終点が揃うように対象としている時刻の波形を時間軸方向に伸縮させる。そして、伸縮させた波形と参照波形の距離を計算し、一番距離が小さくなる時の距離を求める。

2.1.2 類似度

波形間の距離で評価する場合、距離が0の場合に同じ波形と判断され、距離が大きい場合に異なる波形と判断される。この場合、どれだけ距離が小さければ十分似ているかを判断し辛い。そのため、算出した波形間の距離を正規化し、類似度を0～100とした。類似度が最大値（今回の場合、100）の場合に同じ波形、類似度が小さくなればなるほど形が異なる波形と判断される。

2.2 高速検索法

トレンドデータの全てのデータに対して参照波形との類似度を計算すると、類似度が大きい波形を確実に見つけることができるが、データ量が多くなれば検索に時間がかかる。そこで、検索時間を考慮した高速検索法を提案する。高速検索法では、検索対象となるデータからプロセスで起こり得る典型的な変化を表した波形（以下、波形パターンとする）に類似している波形のリスト（以下、インデックスとする）をあらかじめ作成しておき、作成したインデックスから類似波形を検索する手法である。インデックスに登録された波形に対してのみ類似度を計算するため、類似度を計算する対象となるデータを少なくすることができ、検索時間を大幅に短縮することができる。本検索法の手順の概要を2.2.1 節に示す。

2.2.1 アルゴリズム概要

ここでは、高速検索法の手順を述べる。手順(2)～(4)に関しては、2.2.2 節～2.2.4 節で詳細を説明する。

- (1) 前処理：トレンドデータに対して、異常値除去、平滑化、圧縮を行う。
 - 異常値除去：データ全体の平均値や各データの一つ前のデータから離れすぎている場合に、一つ前の時刻の値に修正する。
 - 平滑化：指数加重移動平均を計算する。
 - 圧縮：n 点のデータを平均し、1 点のデータとする。例えば、1 秒周期のデータを1 分周期に圧縮する。
- (2) インデックスの作成
- (3) 参照波形と類似している波形パターンの決定
- (4) インデックス上の波形との類似度算出
- (5) ソート：類似度を計算した波形を、類似度の高い順にソートする。

2.2.2 インデックスの作成

波形パターンとトレンドデータの全データとの類似度を算出することで、トレンドデータの中から波形パターンに類似している波形を事前に登録する。その結果、各波形パターンに類似している、つまり類似度の高い上位 m 件のトレンドデータの波形を各波形パターンのインデックスに登録する。波形パターンの例を図 2 に示す。直線的に上昇し続ける波形、途中で上がり方が急になる波形などである。波形パターンはあらかじめ決めておく。

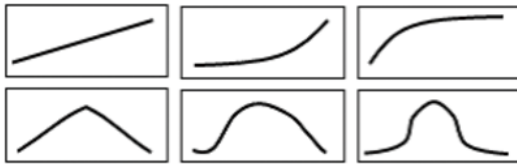


図 2 波形パターンの例

2.2.3 参照波形と類似している波形パターンの決定

参照波形と各波形パターンの類似度を計算する。これにより、どの波形パターンのインデックスを精査すればよいかを決める。類似度上位 n 件の波形パターンを類似パターンとする。

2.2.4 インデックス上の波形との類似度算出

類似パターンのインデックスに登録されている全ての波形と参照波形の類似度を計算する。全データの類似度を算出する場合と比較すると、演算時間が約 N / nm 倍に高速化される。この時、 n は類似パターンの数、 m はインデックスに登録する波形の数、 N は前処理後のトレンドデータのデータ点数である。

3. 評価

開発手法を用いて、プロセスデータの中から形が似ている波形を検索できているかどうか、高速検索法の精度、検索にかかる時間を評価する。

3.1 評価データおよび評価環境

疑似プロセスデータとして、47 日間、温度計を持ち歩き測定した温度データを用いた。このデータは、サンプリング間隔は 10 秒、全データ点数は約 40 万点あり、前処理後のデータ点数は約 2 万点である。全データから一部を抜粋したデータを図 3 に示す。また、評価に用いたパソコンの環境を表 1 に示す。

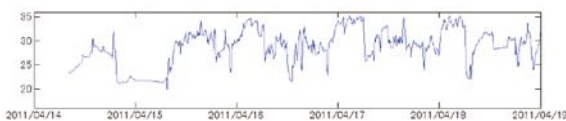


図 3 評価に用いたトレンドデータの一部

表 1 評価環境

CPU	Intel Core i3 3.07 GHz
メモリ	4 GB
ソフトウェア	MATLAB R2010b

3.2 評価方法

3.2.1 形が似ている波形を検索できているかどうかの評価

オフセット、シフト、伸縮を考慮した類似波形検索方法により、プロセスの挙動が参照波形で示しているプロセスの挙動と似ている状態を全データの中から検索できることを確認する。

3.2.2 高速検索の精度の評価

トレンドデータの全データから検索する方法と高速検索法の結果を比較し、同じ検索結果を得られているかどうかを評価する。ここでは、値に変化がある区間のデータから 100 箇所を参照波形として選び、全データ検索の検索結果上位 3 件と高速検索の検索結果上位 3 件を比較し、何 % の検索結果が同じであったかを示す一致率を計算して評価する。全データ検索の検索結果と高速検索の検索結果が完全に一致していれば 100%、全て異なる結果があれば 0% となる。ここでは、全データ検索の検索結果と高速検索の検索結果が共に上位 3 件以内に入っていれば一致しているとした。80% 程度の一致率があれば、似た挙動の例を瞬時に探すような用途に使用できると想定した。

3.2.3 検索時間

3.2.2 節と同様に、データの中から、変化がある区間を参照波形として選び、全データ検索と高速検索法の検索時間を比較する。検索時間が 1 秒以内であれば、オペレータが運転中に使用する際にも問題ないと想定した。

3.3 評価結果および考察

3.3.1 形が似ている波形を検索できているかどうかの評価

全データ検索の結果を図 4 に示す。図 4 左側が元々の測定値で表示した結果を、図 4 右側がオフセット、シフト、伸縮を調整して距離が最小となった状態を表している。点線が参照波形を、実線が検索結果を表している。元々の測定値 (図 4 左側) では波形の長さが異なるため、ユークリッド距離を計算することができない。しかし、提案手法であれば波形を伸縮させるため、距離を計算することができる。図 4 より、参照波形に対してプロセス挙動が似ている波形を検索できていることがわかる。

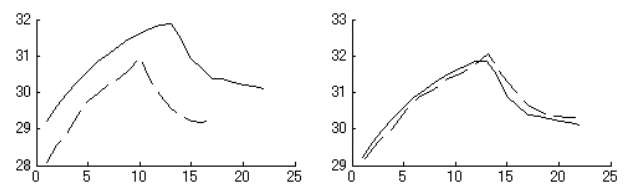


図 4 提案手法を用いて類似波形を検索した時の例

3.3.2 高速検索の精度

全データ検索の検索結果上位3件と高速検索の検索結果上位3件の一致率は約70%であった。このことから、インデックスの改善は必要であるが、似た挙動の例を幾つか見つけたいという用途には利用可能性があることがわかった。

3.3.3 検索時間

検索時間を評価した結果、全データ検索による検索時間は約2.0秒、高速検索による検索時間は約0.1秒であった。つまり、演算時間は約20倍高速化された。理論的には、今回は $m=3, n=100, N=20000$ であったので、 $20000/300 \div 67$ 倍に高速されるはずであるが、関数呼び出しなどの処理のオーバーヘッド等により理論値との差が発生したと考えられる。全データ検索の検索時間から、1秒よりは時間がかかってしまうが、データ解析に利用可能な検索時間であることがわかった。

4. 適用事例

ここでは、説明した類似波形検索法を、連続プロセス以外のデータにも適用した事例を紹介する。

4.1 バッチプロセスデータへの適用

バッチプロセスのデータを解析する際に、膨大なトレンドデータから特定のバッチのデータだけを抽出したいという要求がある。この際に、バッチの切り替えを示すようなフラグ信号等があれば、特定のバッチデータを容易に抽出できるが、フラグ信号等がない場合には手作業で抽出する必要があり、この作業は非常に手間がかかる。そこで、類似波形検索法を用いて、指定したバッチデータからその波形と似ている各バッチデータを抽出することができれば、バッチプロセスデータの解析を効率化できる。

あるバッチプロセスの実データに対し、特定の種類のバッチデータを開発手法により抽出できるかどうかを試した。1バッチ分の波形を参照波形として登録しておき、トレンドデータから参照波形と類似している波形を検索したところ、同じ種類のバッチを動作させている時のデータを抽出することができた。検索結果の中には、反応時間が参照波形よりも若干短い場合や長い場合も含まれていた。この結果より、バッチプロセスデータの解析時に本検索法が使用できる可能性があることがわかった。

4.2 プロセスデータ以外への適用

プラントや鉄道、道路、橋梁などの社会インフラにおいて、設備の状態に応じたメンテナンスを行うために、設備の状態を診断したいという要求がある。設備動作時の電流などの波形を見ることで設備の状態診断を行う時に、設備の正常動作または異常動作時の波形をそれぞれ参照波形としてあらかじめ登録しておき、リアルタイムに計測された波形が参照波形に近いかを計算することによって、その設備の状態が正常か異常かを診断することができる。

ある設備の実運転データに対し、正常動作時の波形を参照波形として、正常/異常の判断が可能かどうかを試した。その結果、図5のように正常に動作した時の波形と参照波形との間には非常に高い類似度が得られ、異常があった時の波形と参照波形との間の類似度は、正常時の類似度よりも小さい値となった。この結果より、この設備においては、正常時の波形パターンとの類似度を用いることで、正常/異常の診断を行えることがわかった。

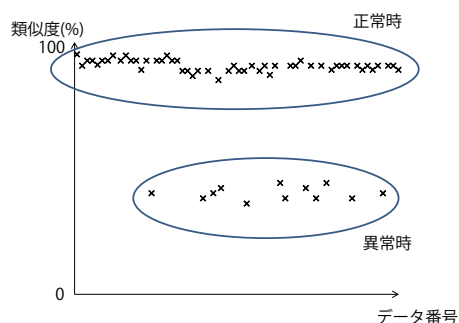


図5 診断結果の例

5. おわりに

プラントのプロセスで起こり得る測定値軸方向および時間軸方向のずれや伸縮を考慮した、波形間の類似度の算出方法を開発した。さらに、参照波形とプロセス挙動が似ている波形を、蓄積されたデータの中から高速に検索する技術を提案した。また、プロセスデータ以外への適用では、設備の状態診断にも適用できる可能性のあることがわかった。

今後の課題としては、高速検索において全データ検索を行った場合との一致率の向上が挙げられる。現在のインデックスは、あらかじめ決めた波形パターンを基にしている。この場合、基本的な波形パターンしか挙げていないため、各プラント固有の動きを網羅することは難しい。また、プラントの設備や運転条件の変更によるプロセス特性の変化に追従できない可能性もある。そこで、実際のプラントに蓄積されたデータを基に波形パターンを最適化するなどの工夫が必要であると考えている。

参考文献

- (1) C. Faloutsos, M. Ranganathan, et al., "Fast Subsequence Matching in Time-Series Databases," SIGMOD '94, 1994, pp. 419-429
- (2) C. Myers, L. Rabiner, "Level Building Dynamic Time Warping Algorithm for Connected Word Recognition," IEEE Transactions on Acoustics Speech and Signal Processing, Vol. 29, No. 2, 1981, pp. 284-297
- (3) E. Keogh, "Exact indexing of dynamic time warping," Journal of Knowledge and Information Systems, Vol. 7, Issue 3, 2005, pp. 358-386

* 本文中で使用されている会社名及び商品名は、各社の登録商標または商標です。