

時系列データの周波数に基づく類似度について

小山 克正* 宝珍 輝尚† 中西 秀哉‡ 小嶋 護‡

* 大阪府立大学 大学院理学系研究科 情報数理科学専攻
〒 599-8531 堺市学園町 1-1

† 京都工芸繊維大学 大学院工芸科学研究科 情報工学部門
〒 606-8585 京都市左京区松ヶ崎御所海道町

‡ 核融合科学研究所
〒 509-5292 岐阜県土岐市下石町 3 2 2-6

本論文では、科学実験で得られるような動きの激しい時系列データを対象に、その類似性について検討する。まず、時系列データの周波数成分の大きさをもとにした非類似度を提案する。動きの激しい時系列データの非類似度としては、一般に用いられるユークリッド距離より提案した非類似度の方が良好であることを実験により示す。また、提案した非類似度の改善として、周波数に高低に合わせた非類似度や包絡線を考慮した非類似度について検討する。

Similarity Based on Spectrum of Time Series

Katsumasa KOYAMA* Teruhisa HOCHIN† Hideya NAKANISHI‡ Mamoru KOJIMA‡

* Osaka Prefecture University
1-1, Gakuen-cho, Sakai, Osaka, 599-8531 Japan

† Kyoto Institute of Technology
Goshokaido-cho, Matsugasaki, Sakyo-ku, Kyoto-shi, Kyoto 606-8585 Japan

‡ National Institute for Fusion Science
322-6, Oroshi-cho, Toki, Gifu, 509-5292 Japan

This paper studies on the dissimilarity of waveforms. The dissimilarity based on the distance in the frequency domain is proposed for the correct similar search of the waveforms having major power at high frequency. The correctness of the proposed distance is evaluated through the metrics used in evaluating that of the information retrieval, i.e. precision and recall. The experimental result shows that the proposed distance works better than the Euclidean one, which is the traditional dissimilarity, in retrieving waveforms having major power at high frequency. Additionally, this paper improves the proposed method by considering the characteristics in recognizing waveforms and the envelopes of waveforms.

1 はじめに

計算機の技術の急速な進展に伴い、文章のみならず画像や音楽、動画といった様々なデータを計算機で扱えるようになった。しかし、膨大な量のデータの中から自分の望むデータを的確に、かつ、迅速に見つけ出すことは困難である。このようなものの中に、株価の変動や気温の変化、科学実験の測定データなど時系列の形式をとるものがある。核融合科学研究所の核融合実験で測定される様々な実験データも時系列データである [1]。核融合

科学研究所では、これまでの実験データは一度解析されると、その後はただ単に保存されているだけで、再度解析されることほとんどない。何か興味深いデータに対し、それと類似しているデータを過去のデータの中から探し出したいという欲求があるが、現状では困難である。

膨大な量の時系列データの中から、ある時系列データと類似の時系列データを検索するという時系列データの類似検索問題について様々な研究がなされている [1-22]。単純には、時系列をユークリッド空間の1点とみなしてそのユークリッド距離を非類似度として検索を行うことが考

えられる。しかし、時系列をユークリッド空間の1点とみなすと、ユークリッド空間の次元は非常に高いものとなる。多次元空間中のオブジェクトを高速に求めるための多次元インデックス構造は、10次元程度が実用の限界であることが実験的に明らかになっており [4]、何らかの手段で次元を削減する方法が研究されてきている。Agrawalらは離散フーリエ変換 (DFT) を用いて次元数を減少させる方法を提案した [2]。また Rafiei らは DFT の性質を利用し Agrawal らの方法を改良した [3]。その他に、時系列の波形をパルス波で近似しその面積を扱う APCA (Adaptive Piece-wise Constant Approximation) [4] や離散ウェーブレット変換 [5, 14] を用いた次元数の減少による高速化の研究がなされている。ユークリッド距離とは異なる類似度として、時系列を複数の小区間に分割し、それぞれを指標となる時系列で近似する MVQ (Multiresolution Vector Quantized) 近似では、全体に含まれる指標の時系列の比率をもとに類似度を計算している [6]。しかし、多くの研究では、株価の変動や売り上げの推移のような動きの緩やかな時系列データを対象としており、科学実験の測定データなどに多く見られる動きの激しいデータには適応できないと思われる。

そこで本論文では、核融合実験で多く見られる短い時間で激しく変化する時系列データを対象とし、その的確な類似検索の実現を目的として、その類似度について検討する。まず動きの激しい時系列データの非類似度として、周波数成分をもとにした非類似度を提案する。そして、動きの激しい時系列データに対しての良好性を実験的に評価する。この結果、時間軸上のユークリッド距離よりも人間の感覚に近い判断ができることを明らかにする。また、これを周波数の高低についての波形の認知の差と高周波波形の包絡線を考慮することにより改良した結果についても述べる。

本論文では、まず、2で、時系列に関する定義、ユークリッド距離、ならびに、離散フーリエ変換について述べる。次に、3でユークリッド距離に代わる非類似度を提案し、その良好性を評価する。その結果をもとに、4及び5で提案した非類似度を改良し、それを実験により評価する。最後に、6でまとめる。

2 時系列について

2.1 時系列の定義

まず、時系列の定義から始める。

定義 1 時系列 x は、実数の系列である

$$x = [x_t] = [x_0, x_1, \dots, x_{n-1}] \quad (1)$$

ただし、 n は十分大きい

このとき、 n は「時系列の大きさ」または「時系列の長さ」という。

2.2 ユークリッド距離について

ユークリッド距離は、長さ n の時系列を n 次元ユークリッド空間の点とみなし、その距離をいう。ここでは、以下のように定義する。

定義 2 時系列 $x = [x_t], y = [y_t]$ の距離 $D(x, y)$ は

$$D(x, y) = \sqrt{\sum_{t=0}^{n-1} \{(x_t - \mu_x) - (y_t - \mu_y)\}^2} \quad (2)$$

ただし、 μ_x, μ_y はそれぞれ x, y の平均値

この式 (2) は厳密にはユークリッド距離とは異なり、各時系列の平均値を差し引いている。これは時系列データが計測器の仕様・状態や保存形式の関係で、データ全体になんらかの修正が加えられている場合が考えられるからである。しかし、本質的な部分では同じであると考えている。

2.3 離散フーリエ変換 (DFT) について

離散フーリエ変換 (DFT) について説明する。DFT とは系列の時間領域から周波数領域への変換で、以下のように定義される。

定義 3 時系列 $x = [x_t] = [x_0, x_1, \dots, x_{n-1}]$ の DFT は、以下のように定義される複素数 X_f の系列で $X = [X_f] = [X_0, X_1, \dots, X_{n-1}]$ の形で定義される。また、以降 $[x_t]$ の DFT が $[X_f]$ のとき $[x_t] \iff [X_f]$ と表すこととする。

$$X_f = \sum_{t=0}^{n-1} x_t e^{\frac{2\pi i f t}{n}} \quad (3)$$

ただし、 i は $\sqrt{-1}$

また、次の操作で逆変換可能で元の時系列 $[x_t]$ に復元できる。

$$x_t = \sum_{f=0}^{n-1} X_f e^{\frac{-2\pi i f t}{n}} \quad (4)$$

留意すべき DFT の性質として、時間時間領域でのシフトは位相差にのみ影響するという、ならびに、対称性を持つことがある。

[シフト]

$$[x_{t-t_0}] \iff [X_f \cdot e^{i f t_0}] \quad (5)$$

[対称性]

$$X_f = X_{n-f} \quad (6)$$

また、時系列の長さ n が 2 のべき乗 (2^k) のとき、 $O(n \log(n))$ の時間計算量で計算可能な DFT の高速アルゴリズムとして高速フーリエ変換 (FFT) が知られている。

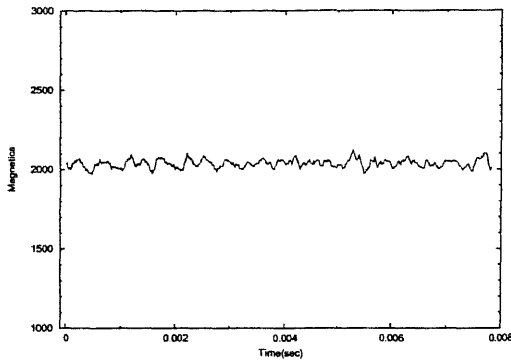


図 1: 検索キー A の波形

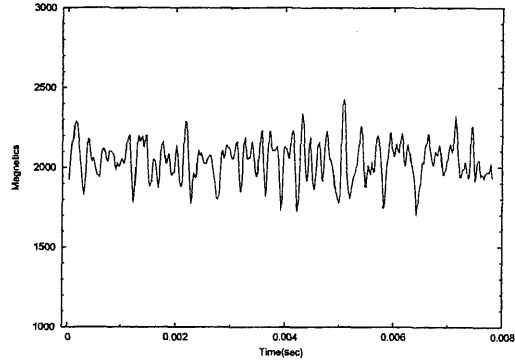


図 2: 検索キー B の波形

3 周波数に基づく非類似度と評価

多くの研究では、ユークリッド距離を時系列が類似性を表す非類似度としていて、この値がある閾値 ϵ よりも小さい場合に、つまり、 $D(x, y) < \epsilon$ のとき、2つの時系列が類似しているとしている。しかし、このユークリッド距離は時系列全体に及ぶような比較的緩やかに変化する時系列に対しては、感覚的にもうまく非類似度を計算できると思われるが、短時間に激しく変化するような動きの激しい時系列に対してもうまく非類似度を計算できるかは疑問である。

3.1 非類似度

時系列に DFT を行うことで、時系列の各周波数毎に「振幅の大きさ(周波数成分)」及び「位相」の情報が得られる。しかし、短時間に激しく変化するような時系列では、この「位相」は人間の目にはほとんどわからないと思われる。

そこで、周波数成分にのみ着目し、以下に示す非類似度を提案する。

定義 4 時系列 $x = [x_t], y = [y_t]$ の DFT を $[X_f], [Y_f]$ としたとき x と y の非類似度 $D'(x, y)$ を以下で定義する。

$$D'(x, y) = \sqrt{\sum_{f=1}^{\frac{n}{2}} (|X_f| - |Y_f|)^2} \quad (7)$$

ただし、 $|X_f|$ は複素数 X_f の絶対値

式(7)で $f=0$ を除くのは、式(2)で各時系列の平均値を差し引いていることと対応している。また、和が $f = \frac{n}{2}$ までなのは式(6)の対称性のためである。

3.2 評価実験

核融合科学研究所の核融合実験で実際に得られたデータを用いて、時系列を類似性を計算する非類似度として、ユークリッド距離と上で定義した非類似度のどちらが良好かを検証するための実験を行う。

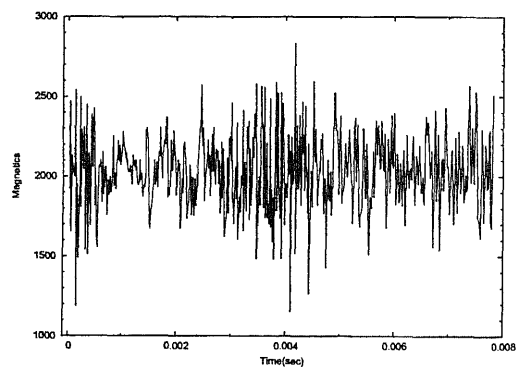


図 3: 検索キー C の波形

3.2.1 実験データ

実験には、核融合研究所の実験で発生する磁場揺動を測定したデータを用いた。過去の測定データの中から、1000個のデータをランダムに選出した。この計測データは、1波形あたり約13万点で構成されていて非常に長い時系列データのため、今回はその一部(512点)分だけを使用した。

3.2.2 実験方法

まず、対象となるデータの中から検索キーをいくつか選出した。検索キーの波形を図1~図3に示す。検索キーA(図1)は変化の幅の小さいもの、検索キーB(図2)は比較的標準的なもの、検索キーC(図3)は変化の間隔が狭く激しく変化するものである。

次に、検索キーの波形と検索対象すべての波形を、著者の一人が実際に肉眼で見比べ、類似していると判断した波形を各検索キーに対する正解集合とした。ここで、正解集合として、それぞれの検索キーに対し「よく似ているもの」、および「似ているもの」の2種類を用意した。ここでの「よく似ているもの」は「似ているもの」に含まれる。

そして、それぞれの検索キーについて、それぞれの非類似度に基づいて検索を行い、再現率と適合率を用いて

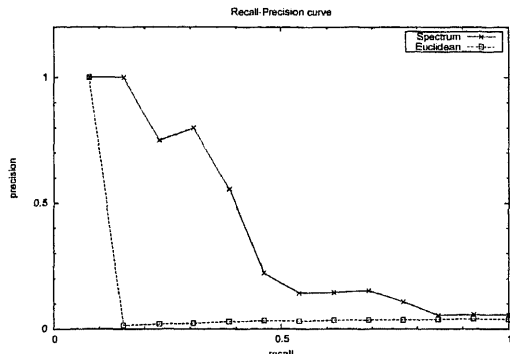


図 4: キー A に「似ているもの」の検索結果

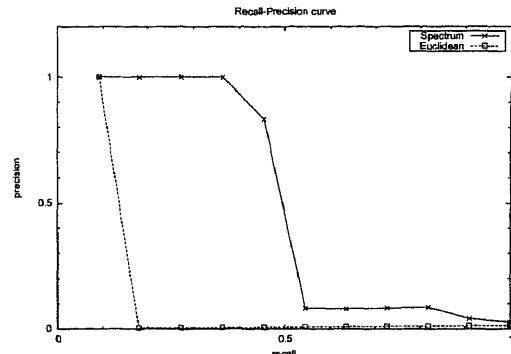


図 7: キー B に「よく似ているもの」の検索結果

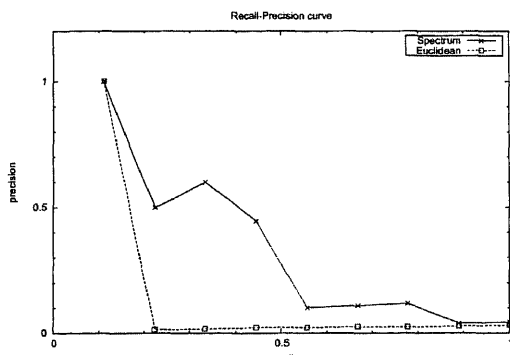


図 5: キー A に「よく似ているもの」の検索結果

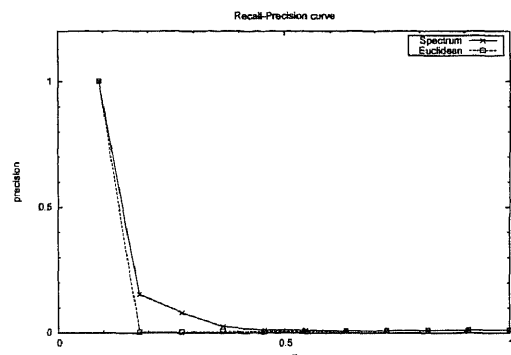


図 8: キー C に「似ているもの」の検索結果

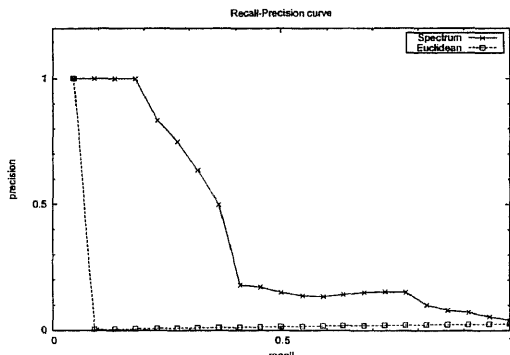


図 6: キー B に「似ているもの」の検索結果

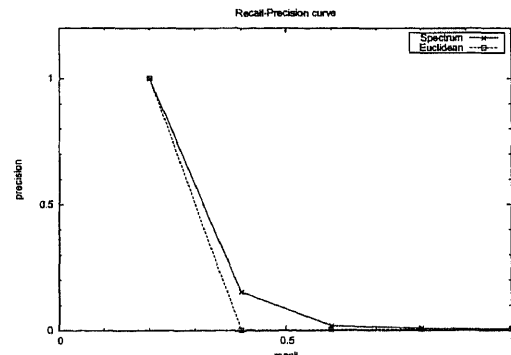


図 9: キー C に「よく似ているもの」の検索結果

評価した。再現率は式 (8) で、適合率は式 (9) で計算する。

$$\text{再現率} = \frac{\text{検索された正解集合の要素}}{\text{正解集合の全要素数}} \quad (8)$$

$$\text{適合率} = \frac{\text{検索された正解集合の要素}}{\text{検索した総数}} \quad (9)$$

3.2.3 結果と考察

それぞれの検索キーに対する再現率・適合率曲線を図 4~9 に示す。

全体を通して、ユークリッド距離より提案した非類似度の方が精度が良いと言える。今回扱ったような短時間

に激しく変動するような時系列データでは、位相を考慮しない方が人間の感覚に近いと考えられる。

また、ユークリッド距離では精度が著しく低くなってしまっている。ユークリッド距離では、系列が少しずれただけでも大きな距離となってしまう、その結果、人間の眼で類似していると判断されるものよりも、振幅がほぼ 0 のものの方が距離が小さくなり、検索結果の上位に出現することが大きな原因であると考えられる。

また、各検索キーに対する「よく似ているもの」と「似ているもの」の場合についてあまり大きな違いが見られず、共通して途中で適合率が急に落ち込む現象が見られ

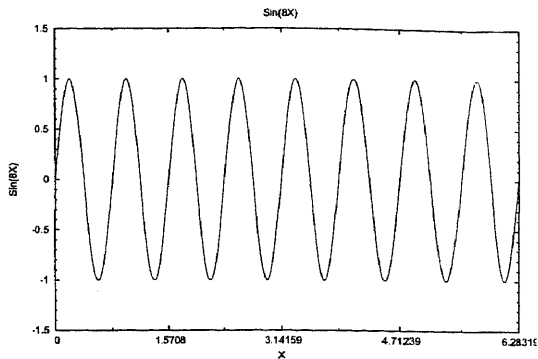


図 10: $\sin(8x)$ の波形

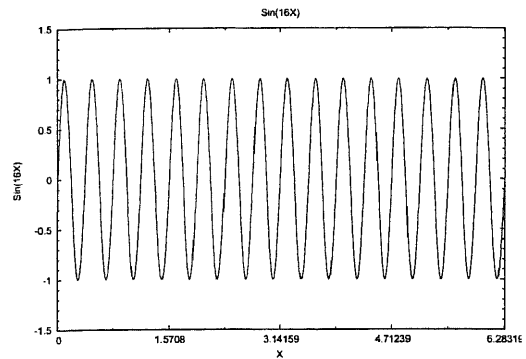


図 12: $\sin(16x)$ の波形

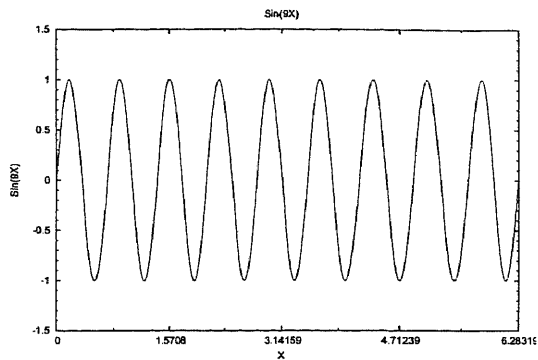


図 11: $\sin(9x)$ の波形

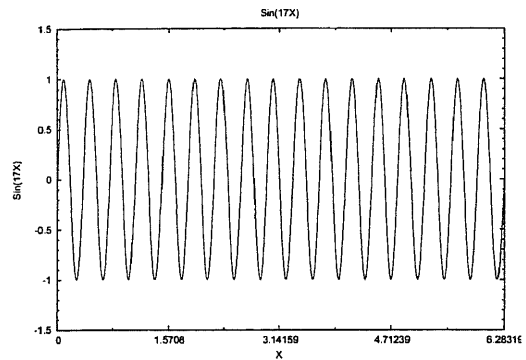


図 13: $\sin(17x)$ の波形

た。これは、スペクトル分布で高い値を持つ周波数が「微妙に異なる」場合でも、人間の眼には類似していると判断されるためと考えられる。そのため、検索キー C(図 3) のように高周波域に大きな周波数成分があるような波形では、提案した非類似度でも精度が低下してしまうと考えられる。

以上をまとめると、激しく変動する時系列データに対しては、提案した非類似度の方がユークリッド距離よりも人間の感覚に近い類似度を与えることがわかった。しかし、変化の間隔が著しく狭く変化の激しいものについては、改善の余地があるようである。

以降、検索キー C に注目し、より人間の感覚に近づくように提案した非類似度の改善を検討する。

4 周波数の高低に基づく非類似度と評価

3.2 において、高周波数域では少々周波数が異なっても人間には区別できないと思われる結果になった。それを考慮して提案した非類似度の改良を考える。まず、どの程度の周波数ならその違いが認識できるかを検証する。その結果を基に周波数の高低に合わせた非類似度を提案する。

4.1 周波数の差について

どの程度の高周波なら周波数の区別がつかなくなるかを検証する。周波数が 1 だけ異なる正弦波 ($\sin(nx)$) を

比較した。 $n = 8, 9$ と $n = 16, 17$ の場合を比較したものが図 10~13 である。 $n = 8, 9$ の違いは何とかわかる程度で、 $n = 16, 17$ ではほとんど区別がつかなかった。周波数をもっと高い場合には、もっと大きな違いでも人の目にはほとんどわからなくなる。

一方、非常に低い周波数では位相の情報も認識できる。つまり、非常に低い周波数ではずれも認識できる。

以上述べてきたように、人間は低い周波数の波形には敏感であるが、周波数の高い波形では鈍感であると考えられる。

4.2 非類似度

周波数の高低による人の認知の違いに着目し、以下のような非類似度を考える。

定義 5 整数 k, l, d 及び列 $\{i_j\}_{j=0}^d$ に対し

$$D''(x, y) = \sqrt{\sum_{f=1}^{k-1} (X_f - Y_f)^2} + \sqrt{\sum_{f=k}^{l-1} (|X_f| - |Y_f|)^2} + \sqrt{\sum_{j=1}^d \left(\sum_{f=i_{j-1}+1}^{i_j} |X_f| - |Y_f| \right)^2} \quad (10)$$

ただし、 $[x_t] \Leftrightarrow [X_f], [y_t] \Leftrightarrow [Y_f]$ で、 $\{i_j\}$ は $l-1 \sim n/2$ を d 個に分割するような数列

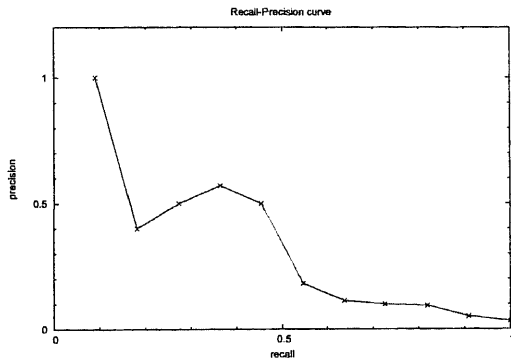


図 14: キー C に「似ているもの」の検索結果

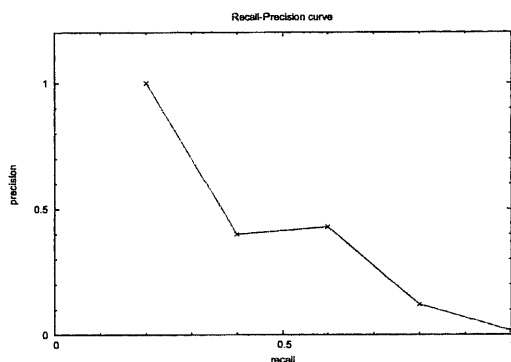


図 15: キー C に「よく似ているもの」の検索結果

ここで、 k, l, d や $\{i_j\}_{j=0}^d$ 分割方法は対象となる時系列によって最良となる条件は異なると考えている。

4.3 評価実験

3.2 で精度の良くなかった検索キー C を用いて、同様の検証を行う。実験データや正解集合は同じもので、上の非類似度に基づく検索を行い再現率・適合率曲線で評価する。なお、式 (10) での $\{i_j\}$ については等分を考える。

式 (12) の k, l, d の値を調節しながら検索キー C で検索を行った。 $k = 1, l = 49, d = 4$ の場合の再現率・適合率曲線は図 14, 15 のようになる。

適切な k, l, d では 4 のときより精度は良くなっている。しかし、「良く似ている」場合の精度はあまり良いとは言えず、人間の感覚とはまだ差があるようである。また、 $k = 1$ のときの精度が良いことを考えると、少なくとも検索キー C に対しては低周波部分はあまり重要でなく、区別できないはずの高周波域に特徴があると予想される。ほかにも、 k, l, d をどのように決めれば良いか、 $\{i_j\}$ の分割はどのようにするのが良いかなど、解決しなければならないことがある。

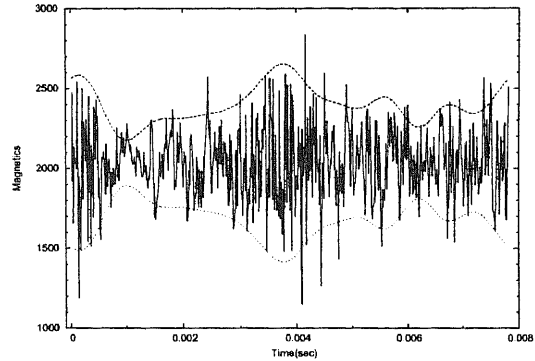


図 16: 検索キー C の包絡線

5 包絡線を考慮した非類似度と評価

検索キー C のように全体を通して上下に激しく震動しているような波形の場合、図 16 のような包絡線が認識でき、これが類似性の判定に影響しているように思われる。そこで、非類似度においてもこの包絡線も考慮することにする。

5.1 包絡線

本来の包絡線は、各点を局所的な最大値（もしくは最小値）でならしたような、あまり高い周波数成分は含まれていないと思われる曲線である。しかし、ここでは以下とする。

定義 6 時系列 $[x_t]$ の包絡線 $[x''_t]$

$$x''_t = \mu_x + |x_t| \quad (11)$$

ただし、 μ_x は $[x_t]$ の平均値

包絡線の形成は高周波成分によるところがほとんどであり、また低周波成分は見た目の情報量が大きいということもあり、包絡線を求める前に低周波成分を除いた方が良く考えられる。

5.2 非類似度

見た目の判断では低周波成分に敏感であるという見解と包絡線は低周波を除いた方が良いという考えから、以下のような非類似度を提案する。

定義 7 時系列 $[x_t], [y_t]$ の非類似度 $D'''(x, y)$

実数 r に対し

$$D'''(x, y) = D_1(x, y) + r \cdot D_2(x, y) \quad (12)$$

整数 k, l, d 及び列 $\{i_j\}_{j=0}^d$ に対し

$$D_1(x, y) = \sqrt{\sum_{f=1}^{k-1} (X_f - Y_f)^2} + \sqrt{\sum_{f=k}^{l-1} (|X_f| - |Y_f|)^2} \quad (13)$$

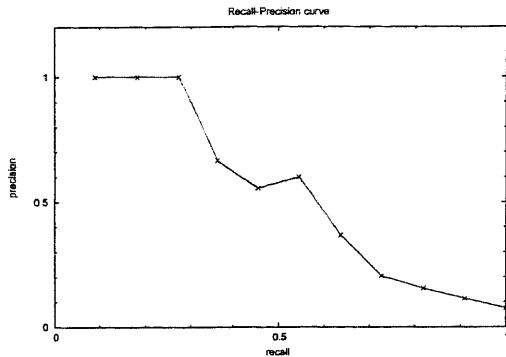


図 17: キー C に「似ているもの」の検索結果

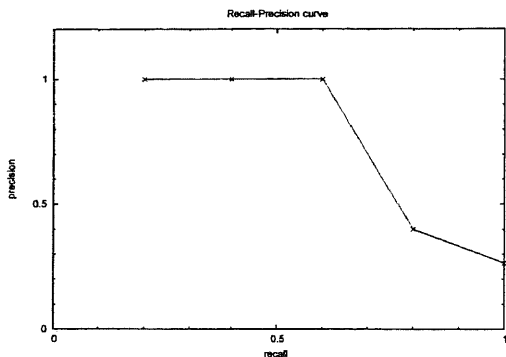


図 18: キー C に「よく似ているもの」の検索結果

$$\begin{aligned}
 D_2(x, y) = & \sqrt{\sum_{f=0}^{k-1} (X''_f - Y''_f)^2} \\
 & + \sqrt{\sum_{f=k}^{l-1} (|X''_f| - |Y''_f|)^2} \\
 & + \sqrt{\sum_{j=1}^d \left(\sum_{f=i_{j-1}+1}^{i_j} |X''_f| - |Y''_f| \right)^2}
 \end{aligned} \tag{14}$$

ただし, $[X''_f] \Leftrightarrow [x''_t], [Y''_f] \Leftrightarrow [y''_t]$ かつ $[x''_t], y''_t$ は $[x_t], [y_t]$ から $f = l$ 以下の低周波成分除いた後の包絡線で, r は $D_1(x, y)$ と $D_2(x, y)$ の間の重みである。

5.3 評価実験

4.3 のときと同様に, 検索キー C を用いて検索を行い, 再現率・適合率曲線で評価する。

式 (12) ~ (14) の条件を調節しながら検索キー C について検索を行った。 $k = 5, l = 57, d = 8, r = 6$ のときの再現率・適合率曲線を図 17, 18 に示す。

条件が適切に設定されれば, かなりの精度が期待できると思われる。さらに, 「よく似ている」の場合も良い結果が得られているので, 人間の判断に近い判定ができて

いると言える。しかし, 4 のときと同様に, k, l, d, r をどのように決めるかや $\{i_j\}$ の分割の方法などの問題が残っている。

6 おわりに

今回は時系列データの中でも, 科学実験の測定データなどに多く見られるような動きの激しい時系列データの類似性について考察した。実験・評価により, 従来用いられてきたユークリッド距離ではうまくいかず, 提案した周波数に基づく非類似度の方が精度が良いこと示した。また, その改善について検討し, 人間の感覚に近い類似性の判定の可能性を示した。ここでは, 周波数の高低についての波形の認知の差と高周波波形の包絡線を考慮した。評価の結果, これらを考慮することで検索精度が向上できることを示した。

今後の課題として, 非類似度を計算するときのパラメータの決定方法についてや他のタイプの時系列データへの応用, 効率的な検索方法の確立などがある。

参考文献

- [1] Nakanishi, H., Hochin, T., Kojima, M. and group, L.: Search and retrieval method of similar plasma waveforms, *Fusion Engineering and Design*, Vol. 71, pp. 189-193 (2004).
- [2] Agrawal, R., Faloutsos, C. and Swami, A. N.: Efficient Similarity Search In Sequence Databases., *Proc. of 1993 Int'l Conf. on Data Organization (FODO'93)*, pp. 69-84 (1993).
- [3] Rafiei, D. and Mendelzon, A. O.: Querying Time Series Data Based on Similarity., *IEEE Trans. Knowl. Data Eng.*, Vol. 12, No. 5, pp. 675-693 (2000).
- [4] Keogh, E. J., Chakrabarti, K., Mehrotra, S. and Pazzani, M. J.: Locally Adaptive Dimensionality Reduction for Indexing Large Time Series Databases., *Proc. of 2001 ACM SIGMOD Int'l Conf. on Management of Data (SIGMOD'01)* (2001).
- [5] Kawagoe, K. and Ueda, T.: A Similarity Search Method of Time Series Data with Combination of Fourier and Wavelet Transforms., *Proc. of 9th Int'l Sympo. on Temporal Representation and Reasoning (TIME'02)*, pp. 86-92 (2002).
- [6] Megalooikonomou, V., Wang, Q., Li, G. and Faloutsos, C.: Multiresolution Symbolic Representation of Time Series., *Pron. of 21st IEEE Int'l Conf. on Data Engineering*, pp. 668-679 (2005).
- [7] Agrawal, R., Lin, K.-I., Sawhney, H. S. and Shim, K.: Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases., *Proc. of 21st Int'l Conf. on Very Large Databases (VLDB'95)*, pp. 490-501 (1995).
- [8] Faloutsos, C., Ranganathan, M. and Manolopoulos, Y.: Fast Subsequence Matching in Time-

- Series Databases, *Proc. of 1994 ACM SIGMOD Int'l Conf. on Management of Data (SIGMOD'94)*, pp. 419–429 (1994).
- [9] Loh, W.-K., Kim, S.-W. and Whang, K.-Y.: Index Interpolation: An Approach to Subsequence Matching Supporting Normalization Transform in Time-Series Databases., *Proc. of 9th Int'l Conf. on Information and Knowledge Management (CIKM'00)*, pp. 480–487 (2000).
- [10] Kahveci, T. and Singh, A. K.: Optimizing Similarity Search for Arbitrary Length Time Series Queries., *IEEE Trans. Knowl. Data Eng.*, Vol. 16, No. 4, pp. 418–433 (2004).
- [11] Harada, L.: An Efficient Sliding Window Algorithm for Detection of Sequential Pattern., *Proc. of 8th Int'l Conf. on Database Systems for Advanced Applications (DASFAA'03)*, pp. 73–80 (2003).
- [12] Seshadri, P., Livny, M. and Ramakrishnan, R.: Sequence Query Processing., *Proc. of 1994 ACM SIGMOD Int'l Conf. on Management of Data (SIGMOD'94)*, pp. 430–441 (1994).
- [13] Park, S., Lee, D. and Chu, W. W.: Fast Retrieval of Similar Subsequences in Long Sequence Databases, *Proc. of 1999 Workshop on Knowledge and Data Engineering Exchange*, pp. 60–67 (1999).
- [14] Keogh, E. J., Chakrabarti, K., Pazzani, M. J. and Mehrotra, S.: Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases., *Knowl. Inf. Syst.*, Vol. 3, No. 3, pp. 263–286 (2001).
- [15] Li, Q., López, I. F. V. and Moon, B.: Skyline Index for Time Series Data., *IEEE Trans. Knowl. Data Eng.*, Vol. 16, No. 6, pp. 669–684 (2004).
- [16] Hirano, S. and Tsumoto, S.: Mining Similar Temporal Patterns in Long Time-Series Data and Its Application to Medicine., *Proc. of 2002 IEEE Int'l Conf. on Data Mining (ICDM'02)*, pp. 219–226 (2002).
- [17] Kim, Y., Park, Y. and Chun, J.: A Dynamic Indexing Structure for Searching Time-Series Patterns., *Proc. of COMPSAC*, pp. 270–275 (1996).
- [18] Loh, W.-K. and Kim, S.-W.: A subsequence matching algorithm supporting moving average transform of arbitrary order in time-series databases using index interpolation., *Proc. of 12th Australasian Database Conf. (ADC'01)*, pp. 37–44 (2001).
- [19] Shahabi, C., Tian, X. and Zhao, W.: TSA-Tree: A Wavelet-Based Approach to Improve the Efficiency of Multi-Level Surprise and Trend Queries on Time-Series Data., *Proc. of 12th Int'l Conf. on Scientific and Statistical Database Management (SSDBM'00)*, pp. 55–68 (2000).
- [20] Kalpakis, K., Gada, D. and Puttagunta, V.: Distance Measures for Effective Clustering of ARIMA Time-Series., *Proc. of 2001 Int'l Conf. on Data Mining (ICDM'01)*, pp. 273–280 (2001).
- [21] Chan, K.-P. and Fu, A. W.-C.: Efficient Time Series Matching by Wavelets, *Proc. of 15th IEEE Int'l Conf. on Data Engineering (ICDE'99)*, pp. 126–133 (1999).
- [22] Rafiei, D. and Mendelzon, A. O.: Efficient Retrieval of Similar Time Sequences Using DFT., *Proc. of 1998 Int'l Conf. on Data Organization (FODO'98)*, pp. 249–257 (1998).
- [23] Katayama, N. and Satoh, S.: The SR-tree: An Index Structure for High-Dimensional Nearest Neighbor Queries., *1997 ACM SIGMOD Int'l Conf. on Management of Data (SIGMOD'97)*, pp. 369–380 (1997).
- [24] 小山克正, 宝珍輝尚: 時系列データの類似性について., 第11回電子情報通信学会関西支部学生会研究発表講演会講演論文集, pg. 56 (2006).
- [25] 宝珍輝尚, 小山克正, 中西秀哉, 小嶋 護: 周波数に基づく波形の非類似度と類似検索への応用, 情報処理学会第83回情報学基礎研究会, pp. 1–8 (2006).