

視認性を考慮した周波数に基づく時系列データの類似度

Similarity of Time Series Based on Their Spectrum Considering Human Recognition

小山 克正¹⁾ 宝珍 輝尚²⁾ 中西 秀哉³⁾ 小嶋 護³⁾

Katsumasa Koyama¹⁾ Teruhisa Hochin²⁾ Hideya Nakanishi³⁾ Mamoru Kojima³⁾

1) 大阪府立大学 大学院理学系研究科 情報数理科学専攻
Osaka Prefecture University

2) 京都工芸繊維大学 大学院工芸科学研究科 情報工学部門
Kyoto Institute of Technology
hochin@kit.ac.jp

3) 核融合科学研究所
National Institute for Fusion Science

本論文では、科学実験で得られる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、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周波数成分の大きさをもとにした非類似度（スペクトル距離）の方が良好であることを再現率と適合率を用いた実験的評価により示す。また、周波数の高低による視認性の違いや外形線を考慮することにより、スペクトル距離の非類似度としての特性が改善できることを実験的評価により示す。

キーワード：時系列，類似度，周波数，視認性

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．興味深いデータに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要求があるが，現状では困難である．

膨大な量の時系列データの中から，ある時系列データに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は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を用いて次元数を減少させる方法を提案している²⁾．また，Rafieiらは離散フーリエ変換の性質を利用しAgrawalらの方法を改良している³⁾．その他に，時系列データを階段状データに近似しその面積を扱うAPCA(Adaptive Piece-wise Constant Approximation)⁴⁾や離散ウェレット変換を用いた次元数の減少による高速化の研究^{5),14)}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⁾．

ここで，時系列データが類似しているとは，前述のように， n 個の点からなる時系列を n 次元空間の一点とし，そのユークリッド距離の小さいものがより類似しているとするのが一般的である．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，このような時系列データに対してはユークリッド距離は良い非類似度であると言えるであろう．すなわち，2つの時系列データのユークリッド距離は，2つの時系列データに対して人間がとらえる非類似度を良く表す測度となっているということである．では，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ついてはどうであろうか．この時系列の例を図1~図3に示す．明らかに，図3の時系列は図2の時系列より図1の時系列と類似している．しかし，図1の時系列と図2の時系

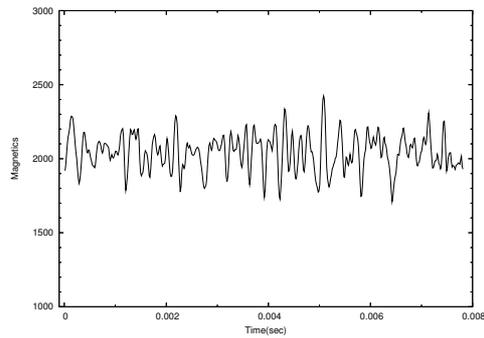


図 1: 検索キー A の波形

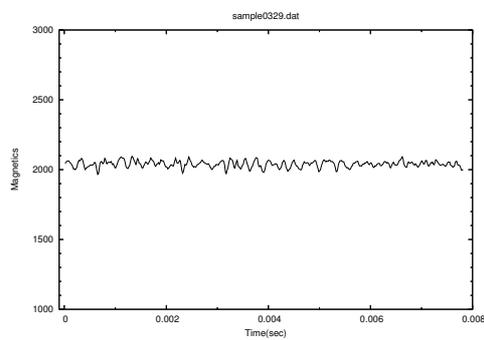


図 2: キー A と似ていない波形

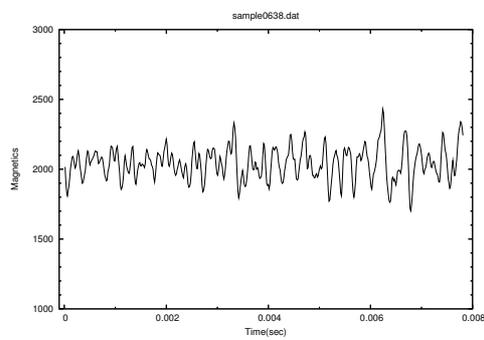


図 3: キー A と似ている波形

列のユークリッド距離は 2648.4 であるが、図 1 の時系列と図 3 の時系列のユークリッド距離は 3723.5 であり、後者は前者の約 1.4 倍となっている。このように動きの激しい時系列に対しては、時間軸上のユークリッド距離は非類似度としてはうまく働かない。すなわち、動きの激しいデータについては、2 つの時系列データの時間軸上のユークリッド距離は 2 つの時系列データに対して人間がとらえる非類似度を的確に表す測度となっていない

ということである．

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，その類似性について検討する．ここで，類似性を判定する方法は，時系列を時間軸上で考える方法^{4),6)}と時系列を周波数領域に変換して考える方法^{2),3),5)}に大別できる．本論文では，多少の周波数の違いに左右されにくくできるのではないかと考えられる周波数領域での非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が良好であることを実験的に明らかにする．さらに，周波数の高低による波形の視認性の違いと高周波波形の外形線を考慮することにより改良した非類似度を提案する．そして，この非類似度の良好性を実験的に明らかにする．

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，離散フーリエ変換，ならびに，周波数成分をもとにした非類似度について述べる．次に，3で周波数成分をもとにした非類似度の良好性を評価する．さらに，4及び5で周波数成分をもとにした非類似度の改良を提案し，その良好性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 を「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 $D(x, y)$

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

ただし, μ_x, μ_y はそれぞれ x, y の平均値 .

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あるからである . しかし, 本質的には同じである .

2.3 離散フーリエ変換 (DFT) について

離散フーリエ変換 (Discrete Fourier Transform; 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こと (式 (5)), ならびに, 対称性を持つこと (式 (6)) がある .

[シフト]

$$[x_{t-t_0}] \iff [X_f \cdot e^{it_0}] \quad (5)$$

[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が知られている .

2.4 スペクトル距離

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)」および「位相」の情報が得られる。周波数成分にのみ着目し、以下に示す非類似度をスペクトル距離と呼ぶ。

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 $f = n/2$ までなのは式 (6) で示した対称性のためである。

3 スペクトル距離の良好性

ここでは、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、非類似度としてのスペクトル距離の良好性を実験的に評価する。

3.1 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 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 A (図 1) は比較的標準的なもの、検索キー B (図 4) は変化の幅の小さいもの、検索キー C (図 5) は変化の間隔が狭く激しく変化するものである。

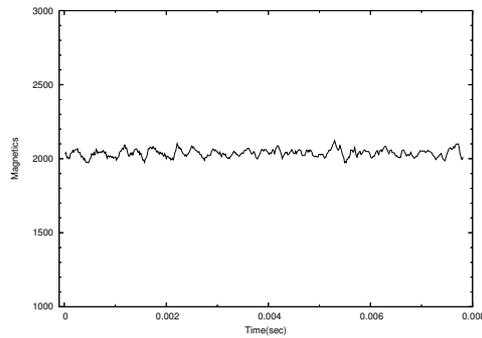


図 4: 検索キー B の波形

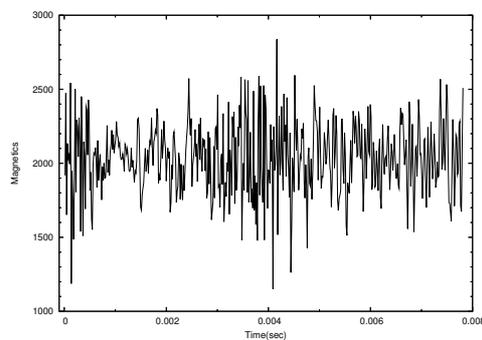


図 5: 検索キー C の波形

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、ユークリッド距離とスペクトル距離それぞれに基づいて検索を行い、再現率 (recall) と適合率 (precision) を用いて評価した。ここで、再現率 = 検索された正解集合の要素の数 / 正解集合の全要素数、適合率 = 検索された正解集合の要素の数 / 検索した総数により計算する。適合率が1ということは、不正解が検索結果に含まれていないということである。また、再現率が1ということは、全ての正解が検索されたということである。正解集合の全要素数 (全ての正解の数) を n とすると、1件目から n 件目までに不正解が現れないのが検索において最も望ましいことであり、この場合、横軸に再現率 (recall) をとり縦軸に適合率 (precision) をとって再現率・適合率曲線を表すと、再現率によらず適合率が常に1の直線になる。これが最も望ましいことを表す。

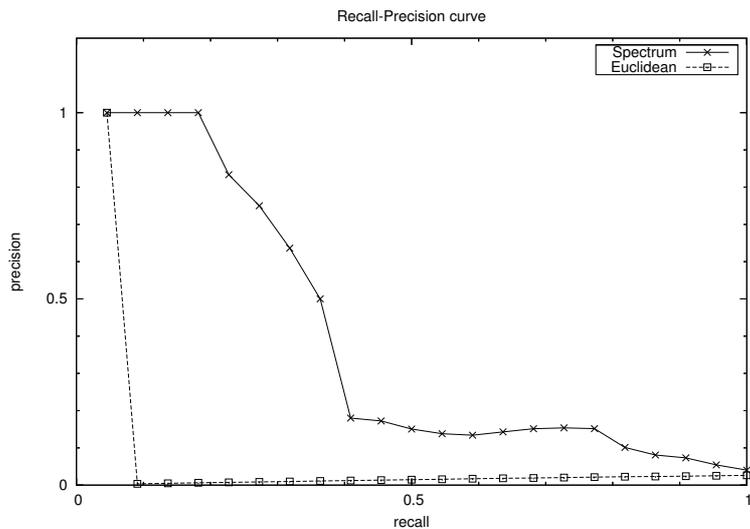


図 6: キー A に「似ているもの」の検索結果

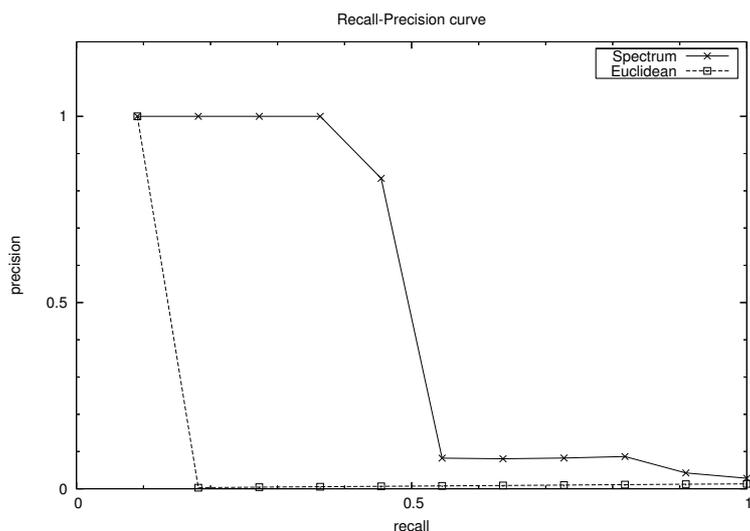


図 7: キー A に「よく似ているもの」の検索結果

しかしながら、実際は、1 件目から n 件目までの間に不正解が現れてしまうので、不正解が現れたところで適合率が低下してしまう。一般には、再現率・適合率曲線は右下がりの曲線となる。2 つの再現率・適合率曲線を比較する場合は、同じ再現率に対して適合率の高いものの方が良い（検索の精度が良い）ので、より右上にある再現率・適合率曲線の方が検索の精度が良いことを表す。

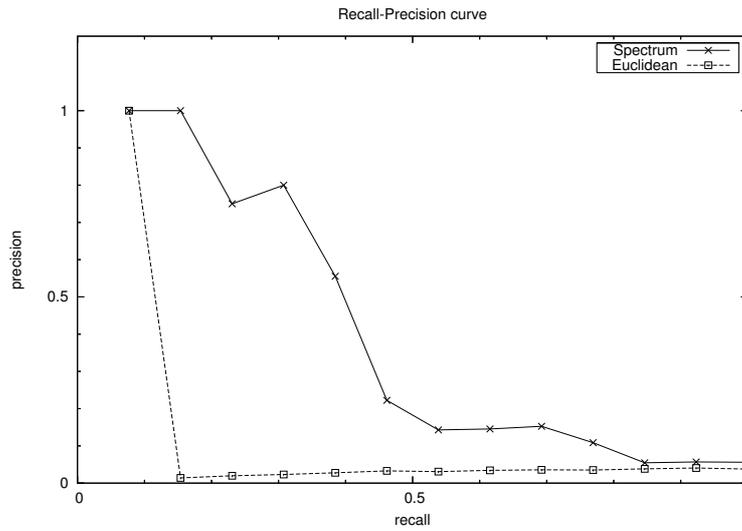


図 8: キー B に「似ているもの」の検索結果

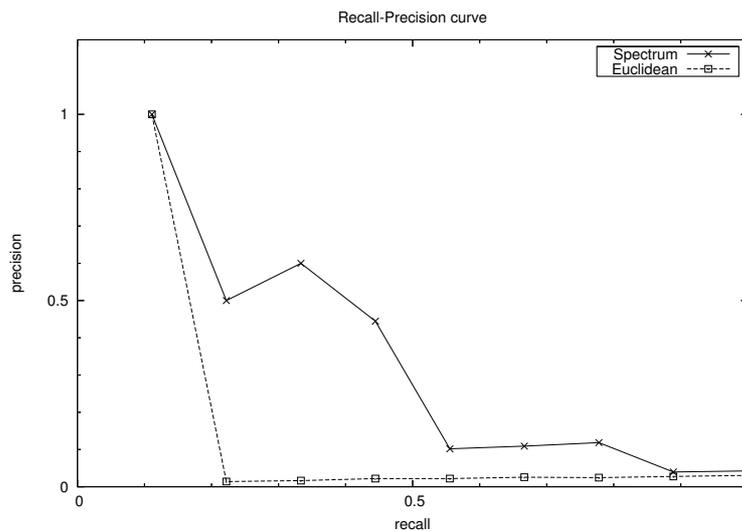


図 9: キー B に「よく似ているもの」の検索結果

3.3 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 6～図 11 に示す．図 6～図 11 では，横軸に再現率 (recall) をとり，縦軸に適合率 (precision) をとっている．

図 6～図 11 のすべてにおいて，スペクトル距離の再現率・適合率曲線の方がユークリッド距離の再現率・適合率曲線よりも右上にあり，全体を通して，ユークリッド距離よりスペクトル距離の方が精度が良いと言える．

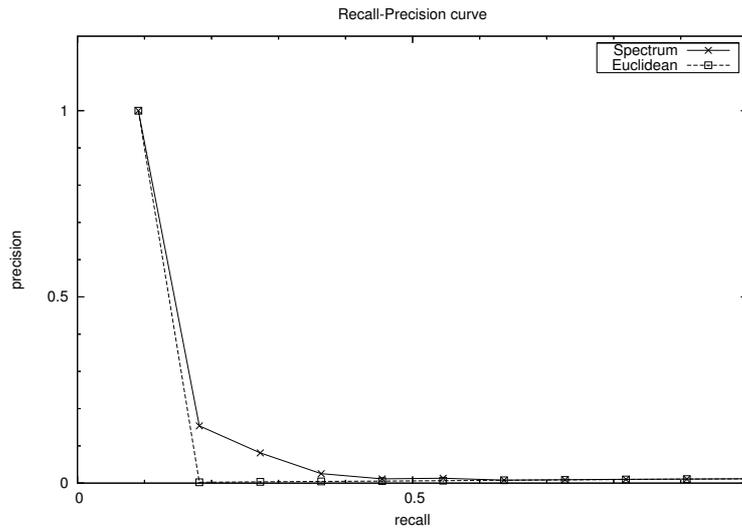


図 10: キー C に「似ているもの」の検索結果

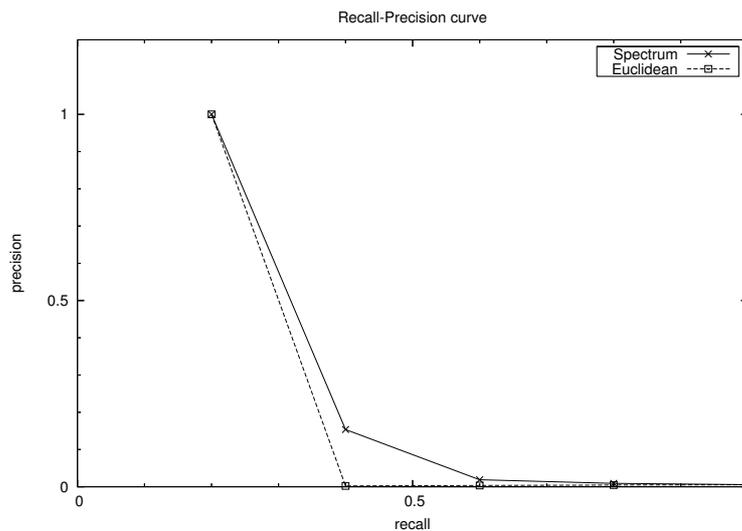


図 11: キー C に「よく似ているもの」の検索結果

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、人間が類似していると判断するものよりも、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。

さらに、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、人間は類似していると判断するためと考えられる。そのため、検索キー C(図 5) 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、スペクトル距離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、スペクトル距離の方がユークリッド距離よりも人間の感覚に近い非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。

そこで、以降では、検索キー C に注目し、より人間の感覚に近づくようにスペクトル距離を改善する。

4 周波数の高低を考慮した非類似度

3.3 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ことを述べた。それを考慮してスペクトル距離を改良する。まず、どの程度の周波数ならその違いが認識できるかについて考察する。次に、その結果をもと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 ($\sin(nx)$) を比較する。 $n = 8, 9$ と $n = 16, 17$ の場合を図 12 ~ 図 15 に示す。 $n = 8$ と $n = 9$ の違いは何とかわかる程度であるが、 $n = 16$ と $n = 17$ ではほとんど区別がつかない。周波数がより高くなると、周波数の差がより大きくても人の目にはその差が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違いも認識できる。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。

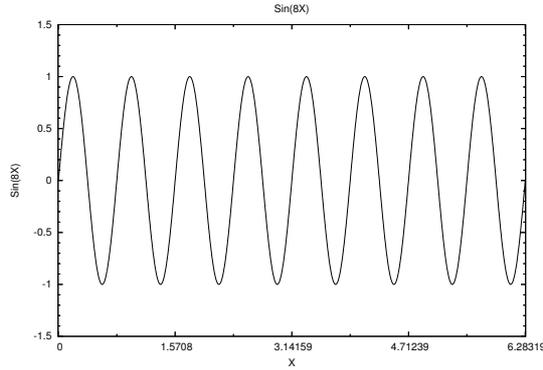


図 12: $\sin(8x)$ の波形

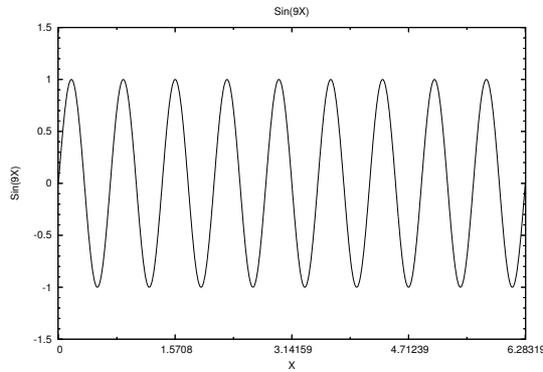


図 13: $\sin(9x)$ の波形

4.2 非類似度

波形の視認性の周波数の高低による違いに着目し、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

$$\begin{aligned}
 D''(x, y) = & \sqrt{\sum_{f=1}^{k-1} |X_f - Y_f|^2} + \sqrt{\sum_{f=k}^{l-1} (|X_f| - |Y_f|)^2} \\
 & + \sqrt{\sum_{j=1}^d \left(\sum_{f=i_{j-1}+1}^{i_j} (|X_f| - |Y_f|) \right)^2} \quad (8)
 \end{aligned}$$

ただし、 $[x_t] \Leftrightarrow [X_f], [y_t] \Leftrightarrow [Y_f]$. また、 $\{i_j\}$ は $l \sim n/2$ を d 個に分割するような数列で、 $i_0 = l - 1, i_d = \frac{n}{2}$.

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 . 式 (8) では、周波数を低 ($f = 1 \sim k - 1$) ・中 ($f = k \sim l - 1$) ・高 ($f = l$)

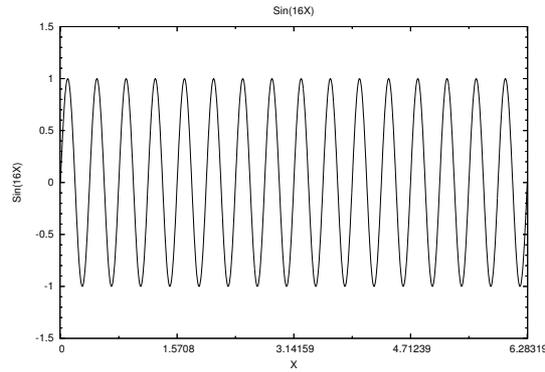


図 14: $\sin(16x)$ の波形

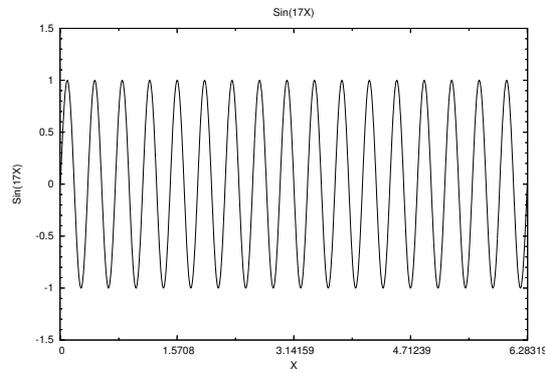


図 15: $\sin(17x)$ の波形

$\sim n/2$) の 3 領域に分け、低周波域では位相まで考慮し、中周波域では周波数成分の大きさのみとし、高周波域ではいくつかの分割ごとに非類似度を求めている

この非類似度 $D''(x, y)$ を漸均スペクトル距離と呼び、以降、2.4 で示した非類似度をこれと区別するため単純スペクトル距離と呼ぶ。

4.3 評価実験

3.3 で精度の良くなかった検索キー C を用いて、3.2 と同様の評価を行う。実験データや正解集合は同じもので、漸均スペクトル距離に基づく検索を行い再現率と適合率で評価する。なお、式 (8) での $\{i_j\}$ については、簡単のため等分とする。

式 (10) 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。最も良い精度が得られた $k = 1, l = 49, d = 4$ の場合の再現率・適合率曲線を図 16 と図 17 に示す。

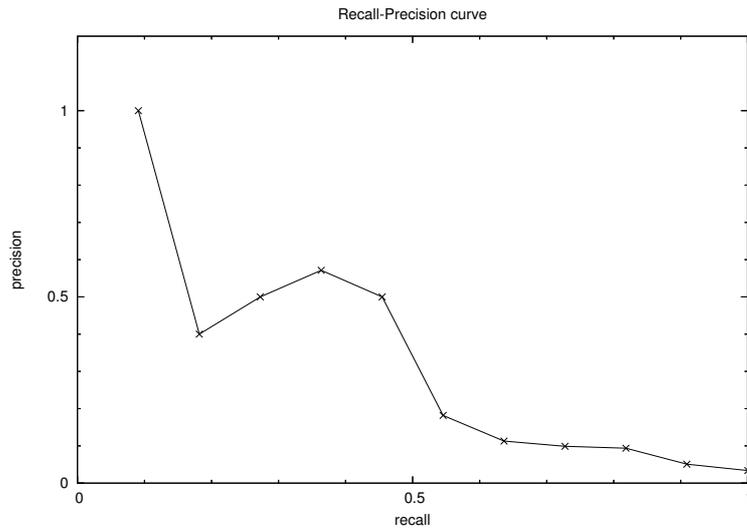


図 16: 漸均スペクトル距離に基づくキー C に「似ているもの」の検索結果

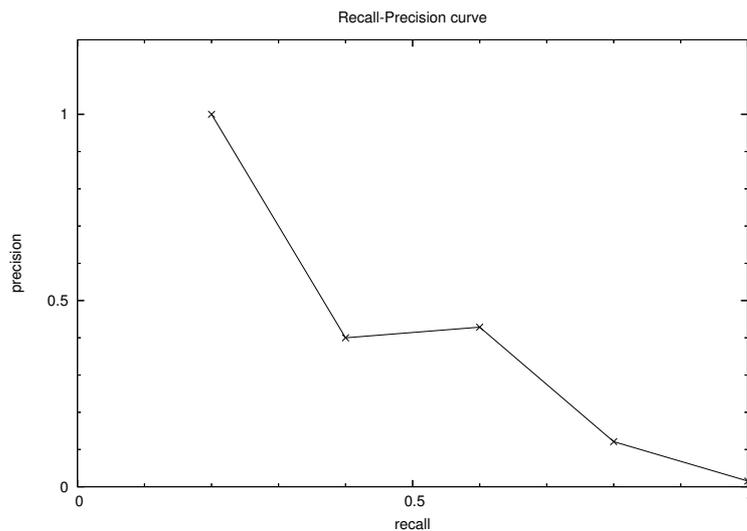


図 17: 漸均スペクトル距離に基づくキー C に「よく似ているもの」の検索結果

適切な k, l, d では 3.3 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原因があると考えられる。なお、 k, l, d の決定法や $\{i_j\}$ の分割法などは今後の課題である。

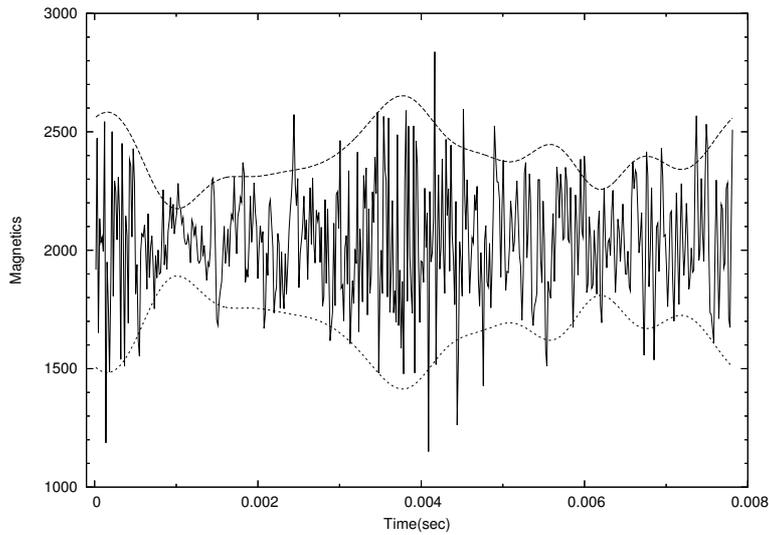


図 18: 検索キー C の包絡線

5 外形線を考慮した非類似度

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 18 に示す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と考えられる．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 外形線

本来の包絡線は，局所的な最大値（もしくは最小値）を結んだような，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．ここでは，簡単のために，包絡線の代わりに以下に示す外形線を考える．

定義 6 時系列 $[x_t]$ の外形線 $[x_t'']$

$$x_t'' = \mu_x + |x_t - \mu_x| \quad (9)$$

ただし， μ_x は $[x_t]$ の平均値

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外形線は高周波成分による影響が大きいことから，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

$$D'''(x, y) = r_1 D_1(x, y) + r_2 D_2(x, y) + r_3 D_3(x, y) \quad (10)$$

整数 k, l, d 及び列 $\{i_j\}_{j=0}^d$ に対し

$$D_1(x, y) = \sqrt{\sum_{f=1}^{k-1} |X_f - Y_f|^2} + \sqrt{\sum_{f=k}^{l-1} (|X_f| - |Y_f|)^2}, \quad (11)$$

$$D_2(x, y) = \sqrt{\sum_{f=0}^{k-1} |X_f'' - Y_f''|^2} + \sqrt{\sum_{f=k}^{l-1} (|X_f''| - |Y_f''|)^2}, \quad (12)$$

$$D_3(x, y) = \sqrt{\sum_{j=1}^d \left(\sum_{f=i_{j-1}+1}^{i_j} (|X_f''| - |Y_f''|) \right)^2}, \quad (13)$$

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外形線で， r_1, r_2, r_3 は $D_1(x, y), D_2(x, y), D_3(x, y)$ の間の重みである． $D_1(x, y)$ は低～中周波の成分での非類似度で，残りの高周波成分から外形線を求め，その外形線の低・中の周波域で非類似度を求めたものが $D_2(x, y)$ で，外形線の高周波域で非類似度を求めたものが $D_3(x, y)$ である．この類似度 $D'''(x, y)$ を漸均・外形スペクトル距離と呼ぶ．

5.3 評価実験

4.3 と同様に，検索キー C を用いて検索を行い，再現率と適合率で評価する．

式 (10) ~ (12) 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．最も良い精度が得られた $k = 7, l = 57, d = 8, r_1 = 1, r_2 = 5, r_3 = 6$ のときの再現率・適合率曲線を図 19 と図 20 に示す．

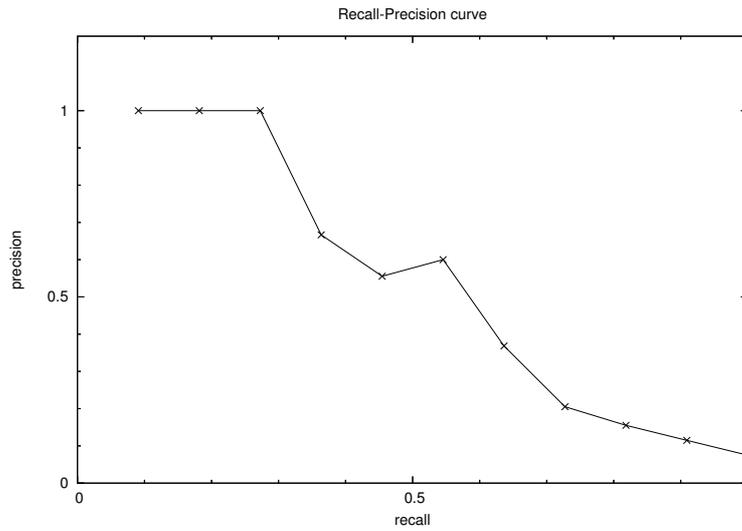


図 19: 漸均・外形スペクトル距離に基づくキー C に「似ているもの」の検索結果

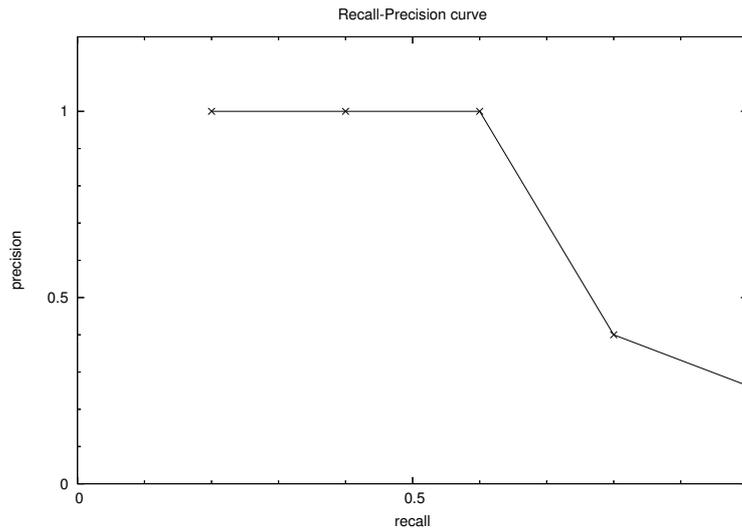


図 20: 漸均・外形スペクトル距離に基づくキー C に「よく似ているもの」の検索結果

図 16, 図 17 と比較して検索精度が改善しており, 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, 漸均スペクトル距離と同様に, k, l, d, r の決定法や $\{i_j\}$ の分割法は今後の課題である.

6 おわりに

本論文では，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動きの激しい時系列データの類似性について検討した．実験的評価により，従来類似性の判定に使用されてきたユークリッド距離よりも，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外形線を考慮した．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．

今後は，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課題である．

謝辞

本研究は，一部，NIFS 共同研究 NIFS06KCHH009 により支援されている．

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] 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).
- [24] 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).
- [25] 小山克正, 宝珍輝尚, 中西秀哉, 小嶋 護: 時系列データの周波数に基づく類似性について, 情報処理学会第 85 回情報学基礎研究会, pp. 53–60 (2006).

This paper studies on the dissimilarity of time series. The spectrum distance is the dissimilarity based on the distance in the frequency domain. The correctness of the spectrum 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 spectrum distance works better than the Euclidean one, which is the traditional dissimilarity, in retrieving waveforms. Additionally, this paper improves the spectrum distance by considering the characteristics in recognizing waveforms and the envelopes of waveforms. This paper shows that these considerations could significantly improve the correctness of the spectrum distance through the experiments.

Keyword: Time series, Similarity, Frequency, Recognition