

## 類似セグメント探索 RDDS 法の評価

渡辺善之, 谷本有紀, 酒井章裕, 杉山雅英†

†会津大学 コンピュータ理工学部

〒965-8580 会津若松市一箕町

E-Mail: †sugiyama@u-aizu.ac.jp

あらまし 複数の時系列に含まれる高速類似セグメント探索 RDDS 法に関して以下の項目に関して実験によりその有効性を述べる。1. クエリ探索の高速化, 2. 2時系列探索の高速化, 3. 3時系列探索の実装と評価。

キーワード: Active 探索, 類似セグメント探索法, 分割統治法

## Evaluation of Similar Segment Search Algorithm in Multiple Time Series

Y.Watanabe, Y.Tanimoto, A.Sakai, M.Sugiyama†

†School of Computer Science and Engineering, The Univ. of Aizu

Ikki-machi, Aizu-Wakamatsu, Fukushima, 965-8580 Japan

E-Mail: †sugiyama@u-aizu.ac.jp

**Abstract** On an efficient similar segment search algorithm, RDDS (Recursive Diamond Division Search), this report describes the following three investigation results; 1. An efficient query search algorithm in one time series, 2. Improvement of similar segment search algorithm in two time series, 3. Implementation and evaluation of similar search algorithm in three time series.

**Keyword:** Active searching, similar segment search algorithm, Divide-and-Conquer algorithm

### 1 まえがき

音声やビデオ等の時系列に含まれるクエリ探索の高速手法について柏野等は Active 探索法 (以下では AS 法と略する) を提案しその有効性を示した [1]。AS 法の基本原理はある時刻で計算した類似度とその近傍での類似度の値との差の上限を用いて類似度計算をスキップ (省略) することができる性質を利用することである。特徴量を有限個のシンボルで表現することで特徴量の時系列をシンボルのヒストグラムで表現し正規化ヒストグラム間の重なり度を類似度として時刻軸での近傍での類似度の差の上限を導き出した。一方, 我々は AS 法の導出で用いられる類似度を距離に置き換えることで近傍での類似度の差の上限の導出を距離の三角不等式に帰着できることを示し, AS 法の原理の単純化及び一般化を行った [2]。

2つの時系列に含まれる類似セグメント探索問題に対して西村等 [3],[4] は AS 法のスキップ幅が 2次元空間において菱形になることに着目しスキップ領域を 2次元空間において接続することで AS 法を 2次元に拡張した RIFAS 法

(Reference Interval Free Active Search) を導出し AS 法を反復して適用する手法に比べて 20-40 倍程度の高速化を実現した。例えば 3次元空間における正八面体の接続は容易ではないので RIFAS 法は 3つ以上の時系列に含まれる類似セグメント探索への拡張は難しい。柏野等 [5] も西村等と同一の手法で AS 法を反復する方法に比べて 10 倍程度高速化を報告している。

我々は複数の時系列に共通に含まれる類似セグメント探索問題を定式化し高次元時刻空間においてスキップ半径の評価式を導出し探索問題の解法として RDDS 法を提案して AS 法及び RIFAS 法との比較実験によりその有効性を明らかにした [6]。

本報告では高速類似セグメント探索 RDDS 法に関して以下の検討結果を述べる。1. クエリ探索の高速化, 2. 2時系列探索の高速化, 3. 3時系列探索の実装と評価。報告は以下のように構成されている。2節で AS 法と RIFAS 法を, 3節で RDDS 法を概説し, 4節で RDDS の検討結果について述べる。

## 2 AS 法と RIFAS 法

### 2.1 出現確率の時系列の性質

音声やビデオなどの特徴ベクトルの時系列  $v_t$  ( $t = 0, \dots, T-1$ ) を  $U_1 = \{x = (x_m) \in R^M \mid x_m \geq 0, \|x\|_1 = 1\}$  のベクトル  $r_t$  (例えばベクトル量子化による出現符号ベクトル) に変換する。ここで  $\|x\|_1$  は  $l_1$  ノルムであり、 $L$  時刻分の  $r_t$  の相加平均を  $p_t$  とする。  $U_1$  は凸である [7] ので  $p_t \in U_1$  となる。この時系列を  $P = (p_t)$  と表わす。任意の時系列  $P$  の任意の時刻のベクトル  $\forall p_t \in P$  及び任意のベクトル  $q \in U_1$ 、整数  $n$  及び  $l_p$  距離 ( $p \geq 1$ ) に対して式 (1) が成り立つ [8]。

$$|d_p(p_{t+n}, q) - d_p(p_t, q)| \leq \frac{2^{\frac{1}{p}} |n|}{L}. \quad (1)$$

### 2.2 Active 探索法とそのスキップ幅

時系列に含まれるクエリ探索問題は以下で定式化される。

時系列に含まれるクエリ探索問題

時系列  $P$  に対して  $q \in U_1$  をクエリ、 $\theta \geq 0$  を探索閾値とし、以下を満たす時刻  $t$  を見つけること。

$$d_p(p_t, q) \leq \theta. \quad (2)$$

幾何学的に言えば  $q$  を中心とする半径  $\theta$  の球 (探索球と呼ぶ) に属する  $p_t$  の時刻を求める問題と言える。全探索の距離計算回数は時系列の長さ  $T$  に等しい。従ってこの高速化問題は距離計算回数を削減することで実現される<sup>1</sup>。不等式 (1) から時刻  $t$  での距離と探索閾値で定まる以下の式 (3) で定まる値

$$R = \frac{L|\theta - d_p(p_t, q)|}{2^{\frac{1}{p}}}, \quad (3)$$

に対して  $\forall n < R$  に対して以下が成り立つ。

$$\begin{cases} d_p(p_t, q) > \theta \implies d_p(p_{t+n}, q) > \theta \\ d_p(p_t, q) \leq \theta \implies d_p(p_{t+n}, q) \leq \theta. \end{cases}$$

ここで  $R$  をスキップ幅と呼ぶ。AS 法の動作原理を図 1 に示す。時刻  $t$  での  $p_t$  とクエリ  $q$  との距離  $d_p(p_t, q)$  を計算し、 $p_t$  が探索球の内側 (外側) であれば  $n < R$  を満たす時刻までは  $p_{t+n}$  が探索球の内側 (外側) にあることになるのでその距離  $d_p(p_{t+n}, q)$  を計算する必要がなく、距離計算をスキップできることになる。次に距離を計算する時刻は  $t + R$  となる。

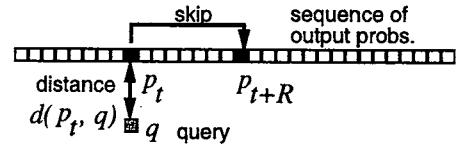


図 1: AS 法のスキップ幅

### 2.3 RIFAS 法

RIFAS 法は 2 つの時系列に含まれる類似部分を高速探索する手法である。AS 法のスキップ幅が 2 次元空間において菱形になることに着目しスキップ領域を 2 次元空間において接続することで AS 法を 2 次元に拡張する方法である。2 時系列に含まれる類似セグメント探索問題は以下のように定式化される。

2 時系列に含まれる類似セグメント探索問題

二つの時系列  $P = (p_t), Q = (q_s)$  に対して以下を満たす時刻の組  $(t, s)$  を見つけること。

$$d_p(p_t, q_s) \leq \theta. \quad (4)$$

図 2 に示すように時刻  $s = 0$  とし、 $q_s$  を AS 法におけるクエリとして  $d(p_t, q_s) \leq \theta$  を満たす  $t$  を求める問題を考え、 $t$  軸に AS 法を適用して式 (3) のスキップ幅  $R$  を求めることで高速化可能である。これを AS 反復法と呼ぶことにする。この時、 $p_t$  をクエリと見なすと  $s$  軸に関してもスキップ幅  $R$  だけ  $s$  軸方向にもスキップ可能である。従って  $(t, s)$  平面でスキップ可能な領域は  $(t, s)$  を中心とする菱形となる。<sup>2</sup>  $t$  軸に沿って  $t'$  にスキップする毎に菱形のスキップ領域が定まるのでそれら隣り合う菱形の交点の高さの最小値を求めることで  $s$  軸方向での次に求めるべき  $s'$  までスキップすることができる。

RIFAS 法は AS 法をクエリ軸方向に拡張する手法であるので処理は逐次的であり、菱形のスキップ領域の中心から見て正方向の上半面、即ち菱形の 4 分の 1 しか利用していない。また菱形の接続で得られるスキップ幅の最小値しかクエリ方向でスキップできない。

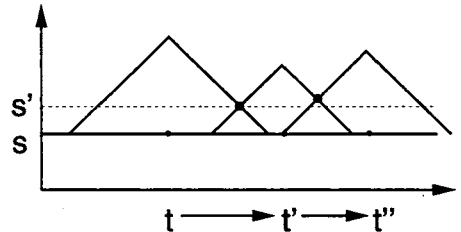


図 2: RIFAS 法とスキップ領域の接続

<sup>1</sup>距離に基づいて導出法を述べるが類似度に基づいた導出も同様である。

<sup>2</sup>この導出については一般化した形で 3 節での述べる。

### 3 複数時系列中の類似セグメント探索

#### 3.1 探索問題の定式化とスキップ半径

複数 ( $I$  個) の時系列  $P^{(i)}$  ( $i \in [1, I]$ ) の時刻を  $t_i$  で、 $i, j$  番目の時系列の時刻  $t_i, t_j$  のベクトル  $p_{t_i}^{(i)}, p_{t_j}^{(j)}$  間の距離  $d_p(p_{t_i}^{(i)}, p_{t_j}^{(j)})$  を  $d(t_i, t_j)$  と表わす。時刻ベクトル  $t = (t_1, t_2, \dots, t_I)$  における時系列の距離を式 (5) で定義する。

$$d(t) = \sum_{i < j} d(t_i, t_j). \quad (5)$$

$d$  の対称性から  $i < j$  の時刻の組とし、組数は  ${}_I C_2 = I(I-1)/2$  で  $d(t_i, t_j) \leq 2^{\frac{1}{p}}$  なので  $0 \leq d(t) \leq I(I-1)2^{\frac{1}{p}-1}$  となる。類似セグメント探索問題を以下で定式化する。

##### 複数時系列類似セグメント探索問題

$I$  個の時系列  $P^{(i)}$  に対して閾値  $\theta (\geq 0)$  以下となる時刻ベクトル  $t$  を見つけること。

$$d(t) \leq \theta. \quad (6)$$

式 (6) を満たす時刻ベクトル  $t$  が  $I$  個の時系列に含まれる類似セグメントの時刻を示す。  $d(t) = 0$  であれば  $p_{t_i}^{(i)} = p_{t_j}^{(j)}$  ( $\forall i, j$ ) となる。全ての時刻の組  $t$  に対して距離  $d(t)$  を求めれば式 (6) を満たす時刻を検出することができるがその距離計算回数は時系列の長さ ( $T_i$ ) の積  $T = \prod_i T_i$  となり膨大な数となる。従って探索の高速化問題は距離計算回数の削減することである。

$d(t)$  の値域から  $\theta \leq I(I-1)2^{\frac{1}{p}-1}$  である。式 (1) に対応する不等式 (7) が成り立つ。

##### 性質 1 時刻

$$|d(t+n) - d(t)| \leq \frac{2^{\frac{1}{p}}(I-1)}{L} \|n\|_1, \quad (7)$$

ここで  $\|n\|_1 = \sum_i |n_i|$  である。

証明を [9] に示した。式 (7) は式 (1) の多次元時刻への拡張である。  $L_i = L$  に限定せず式 (7) の一般化が可能であるが、簡単化して  $L_i = L$  に限定して述べる。

##### 性質 2 時刻ベクトル $n$ が以下を満たすとす。

$$\|n\|_1 < \frac{L}{2^{\frac{1}{p}}(I-1)} |\theta - d(t)|. \quad (8)$$

この時、  $d(t) > \theta$  であれば  $d(t+n) > \theta$  であり、  $d(t) < \theta$  であれば  $d(t+n) < \theta$  となる。

証明を [9] に示した。式 (8) で等号を含めると  $d(t+n) \leq \theta$  等となる。多次元時刻  $t$  とすることで式 (3) で定義したスキップ幅は時刻領域におけるノルムで記述される。式 (8) の右辺をスキップ半径  $R$  と呼ぶことにする。式 (8)

を満たす時刻ベクトル  $n$  の集合は  $l_1$  ノルムの球で  $I$  次元空間の超菱形領域 (1 次元では区間, 2 次元では菱形, 3 次元では正八面体) となる。  $\theta, d(t)$  の値域からスキップ半径は式 (9) を満たす。

$$R \leq \frac{IL}{2}. \quad (9)$$

#### 3.2 スキップ半径と探索

探索時は時刻  $t$  での距離  $d(t)$  を計算し式 (6) を判定する。  $d(t) \leq \theta$  であれば  $W (< R)$  を半径とする  $l_1$  球  $B_1(t, W)$  内の全ての時刻  $t+n$  が式 (6) を満たす。一方、  $d(t) > \theta$  であれば球  $B_1(t, W)$  の内部の全ての時刻は条件を満たさず枝刈りできる。  $W$  が大きければ 1 回の距離計算で多くの時刻の組に対して検出、枝刈りが判定可能である。  $d(t) - \theta$  が 0 に近づく、即ち検出と枝刈り領域の境界に  $t$  が近づくとき  $R$  は小さくなり判定の効率が低下する。

#### 3.3 再帰的菱形分割探索法 (RDDS)

類似セグメント探索問題の解法として再帰的菱形分割探索法 (RDDS: Recursive Diamond Division Search) を提案した。図 3 に示すように時刻領域  $T = [0, T_1] \times [0, T_2] \times \dots \times [0, T_I]$  を同一の半径  $W$  の  $l_1$  球<sup>3</sup>で被覆する。

$$T \subset \cup_i B_1(t_i, W). \quad (10)$$

球の中心  $t_i$  での距離  $d(t_i)$  を求めれば式 (8) から導かれる式 (11) を用いて図 3 に示すように球  $B_1(t_i, W)$  が検出領域、枝刈り領域、判定不可能のいずれかになる。判定不可能の場合には球を分割し距離を再度求め判定を行う。

$$\begin{cases} d(t) > \theta + \frac{2^{\frac{1}{p}}(I-1)W}{L} & \text{(枝刈り)} \\ d(t) < \theta - \frac{2^{\frac{1}{p}}(I-1)W}{L} & \text{(検出)} \\ d(t) \text{ 上を満たさない} & \text{球を再帰分割} \end{cases} \quad (11)$$

高次元で式 (10) を満たす  $l_1$  球は  $l_\infty$  球 (超立方体) と以下の包含関係 (12) を利用して構成する。

$$B_\infty(t_i, r) \subset B_1(t_i, Ir). \quad (12)$$

時刻領域  $T$  を被覆する半径  $r = W/I$  の  $l_\infty$  球の中心  $t_i$  を定める。この時  $T \subset \cup_i B_\infty(t_i, W/I) \subset \cup_i B_1(t_i, W)$  を満たす。半径  $W = (IL)/2$  に設定すると  $l_\infty$  球の半径は次元によらず式 (13) で定まる。従って超立方体 ( $l_\infty$  球) の一辺の大きさは  $L$  となる。

$$r = W/I = L/2. \quad (13)$$

<sup>3</sup>  $I$  次元空間の  $l_1$  球は正  $2^I$  面体で頂点数  $2I$ , 正  $I$  面体  $2^I$  個であり、  $l_\infty$  球は正  $2I$  面体で頂点数  $2^I$ , 正  $2I-2$  面体  $2I$  個で構成され、互いに双対である [10]。

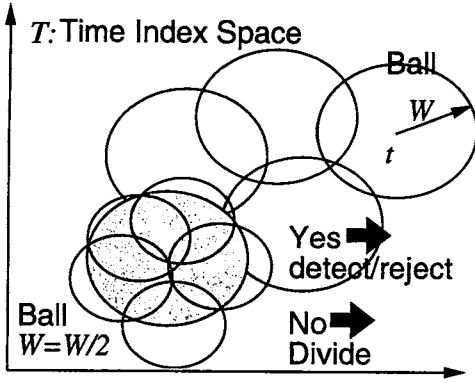


図 3: 球による時刻空間の被覆と球の分割

## 4 RDDS の評価

### 4.1 評価実験条件

評価実験に CampusWave データベース [11] を用いた。これは会津若松市内の FM 局の音楽リクエスト番組であり、2名の女性パーソナリティの対話音声、リクエスト曲、CM 音声などを含んでいる。音声長は各々約 1 時間であり、セグメント長を 10 秒 ( $L = 625$ )、 $l_p$  距離は  $p = 1$  とした。VQ 符号帳サイズを  $M = 32$  とし第 1 回データから LBG 法で作成した。

### 4.2 RDDS の時系列クエリ探索の高速化

RDDS において式 (11) の判定の際に計算したスキップ領域情報を効果的に利用できていない：問題 (a)：式 (11) の第 1 もしくは第 2 の条件を満たす場合、スキップできる領域が菱形より大きくても、その固定の菱形の分しか判定できない。問題 (b)：逆にスキップ領域が菱形よりほんの少し小さい場合には第 1、第 2 の条件を満たさないことになるのでスキップ領域があるにもかかわらず再帰分割が行われることになる。

これは均一の大きさの菱形で時刻空間を分割被覆することに起因している。そこで計算したスキップ半径を用いてスキップ可能領域を除外して再帰分割する RDDS-lr 法を提案する。図 4 に RDDS-lr 法の動作の概念を示す。

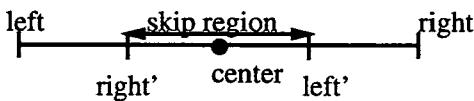


図 4: RDDS-lr の探索手順

探索範囲  $[left, right]$  の中心  $center = (left + right) / 2$  でクエリとの距離  $d = d(p_{center}, q)$  を計算しスキップ半

径  $R = \frac{L}{2} |d - \theta|$  を求める。この領域はスキップ (枝刈/検索) 可能であるので未処理の  $[left, right]$  及び  $[left', right']$  の二つの範囲に対して再帰的に探索を呼び出す。ここで以下のように計算される。

$$\begin{cases} right' = center - R - 1 \\ left' = center + R + 1. \end{cases} \quad (14)$$

データベースをブロック  $[left, right]$  に等分割し、 $W_B = (right - left + 1) / 2$  をブロック幅と呼ぶことにする。ここでブロック数は  $T / (2 * W_B + 1)$  である。ブロック幅を大きくするとブロック数は単調減少するが、探索処理はデータのバッファリングが必要となりリアルタイム処理に不向きになる。

CampusWave データベースの第 1 回のデータを用いてその中の 33 個所のセグメント (コマーシャル音声部分) をクエリとするクエリ探索の評価実験を行った。探索閾値  $\theta$  とブロック幅  $W_B$  を変化させた時の RDDS-lr と AS、RDDS の距離計算回数を図 5 に示す。各々距離計算回数は 713, 645, 520 程度となる。RDDS では半径  $W$  を適切に設定する必要があったが、RDDS-lr ではある程度の大きさに  $W_B$  を設定すれば良好な結果が得られることが分かる。小さな  $W_B$  では、RDDS 同様、RDDS-lr でもブロック数が増加するので距離計算回数が増加する。RDDS、RDDS-lr 双方ともにある値 ( $W_1 \sim 250$ ) までは計算回数が減少する。 $W_1$  はおよそ平均スキップ幅 (300 程度) に対応している。これは  $W_B = W_1$  の時、ほとんどのブロックで 1 回の距離計算でスキップ判定が行われるからである。一方  $W_B > W_1$  とするとブロック数は減少するものの多くのブロックでスキップ判定が終了せず再帰分割による距離計算が必要となり距離計算回数は増加してしまう。さらに  $W_B > 2000$  を越えるとブロック数の減少効果が大きく距離計算回数は 520-530 とほぼ一定の値になる。2 次元平面を可変な菱形で埋めることが難しい。RDDS-lr の多次元への拡張は今後の検討課題である。

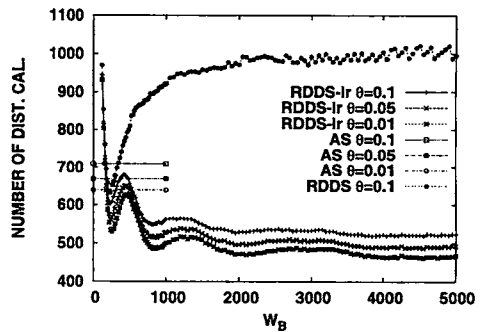


図 5: 探索閾値  $\theta$  とブロック幅  $W_B$  に対する RDDS-lr 法の距離計算回数の変化

### 4.3 2時系列類似セグメント探索の高速化

RDDS法をより高速化するために数学的な距離の公理である三角不等式を用いて距離計算の削減を行う。2次元における $l_1$ 球は菱形であるので図3で述べた $l_1$ 球による時刻空間 $T$ は重複なく効率良く被覆可能である。さらに球の半径を $W$ とすると再分割は球の中心を上下左右に $W/2$ ずらした半径が $W/2$ の4つの球で実現できる[9]。図6に示すように被覆した球の中心 $(iW, jW)$ は $2W$ 毎に等間隔に配列されるので球の中心の $x$ 座標 $iW$  ( $i = 0, 1, \dots$ )は等間隔に配置され、球の中心の $y$ 座標 $jW$  ( $j = 0, 1, 2, \dots$ )を更新する時、 $x$ 座標は繰り返されることになる。従って同一 $y$ 座標に対する球の中心に対応するベクトル間距離を予め計算しておけば球の中心で計算した距離値を同一の $y$ 座標での他の球に対する判定に用いることにより距離計算の削減が可能である[12]。

第1, 第2時系列の $t, s$ 番目の出現確率ベクトルを $p_t^{(1)}, p_s^{(2)}$ で表すことにする。3点 $(2iW, 2jW), (2iW, 2nW), (2nW, 2jW)$ の距離に対して三角不等式

$$d(p_{2iW}^{(1)}, p_{2jW}^{(2)}) - d(p_{2iW}^{(1)}, p_{2nW}^{(1)}) \leq d(p_{2nW}^{(1)}, p_{2jW}^{(2)}),$$

を用いて以下の不等式が成り立つ。

$$\begin{aligned} \frac{L}{2}(d(p_{2iW}^{(1)}, p_{2jW}^{(2)}) - d(p_{2iW}^{(1)}, p_{2nW}^{(1)}) - \theta) \\ \leq \frac{L}{2}(d(p_{2nW}^{(1)}, p_{2jW}^{(2)}) - \theta). \end{aligned} \quad (15)$$

式(11)の枝刈りの条件を $p = 1, I = 2$ として書き換えると $W < \frac{L}{2}(d(t) - \theta)$ となるので、式(15)の左辺が $W$ よりも大であれば、即ち、

$$W < \frac{L}{2}(d(p_{2iW}^{(1)}, p_{2jW}^{(2)}) - d(p_{2iW}^{(1)}, p_{2nW}^{(1)}) - \theta),$$

が成り立てば、式(15)の右辺も $W$ より大となるので式(16)も成り立つことになる。

$$W < \frac{L}{2}(d(p_{2nW}^{(1)}, p_{2jW}^{(2)}) - \theta). \quad (16)$$

即ち点 $(2iW, 2jW)$ での距離値 $d(p_{2iW}^{(1)}, p_{2jW}^{(2)})$ を用いて点 $(2nW, 2jW)$ での枝刈りが可能である。予め計算してある距離行列 $d(p_{2iW}^{(1)}, p_{2nW}^{(1)})$ の値を参照するだけであるので、新たな距離計算を必要としない。

$x$ 座標の偶数の $2iW$ 番目と $2nW$ 番目の球の中心での出現確率ベクトルの相互間の距離行列 $(d(p_{2iW}^{(1)}, p_{2nW}^{(1)}))$ を作成する。奇数番に対しても同様である。 $J = (T - L + 1)/(2W)$ とすると、距離行列の大きさは $(J + 1)^2$ である<sup>4</sup>が距離の対称性から行列作成に必要な距離計算回数は $J(J + 1)/2$ である。

CampusWaveデータベースの第1回、第2回のデータを用いて評価実験を行った。表1にその結果を示す。処

<sup>4</sup>奇数番に対しては $J \times J$ となる。

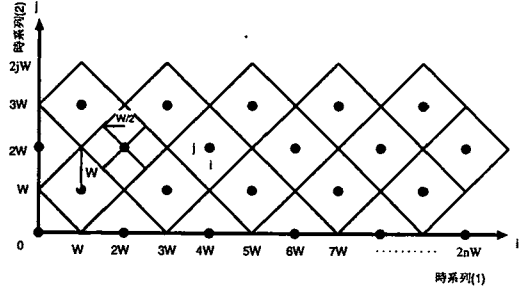


図6: 2次元空間における菱形被覆と距離行列による枝刈

理時間には距離行列計算の時間を含めており、距離行列計算に要する時間を()内の数字で示した。距離計算回数の削減は $W = 256$ の時、111,772 (= 618435 - 506663)回(18.07%)、 $W = 128$ の時、1,146,346 (= 1561240 - 414894)回(73.43%)となる。 $W$ が小さいほど再帰探索を呼び出さずに枝刈判定ができるので距離行列での判定が効果的である。一方、探索処理時間については遅くなっている。その理由として、三角不等式における距離行列の参照回数や比較判定の計算回数が多いことなどが考えられる。比較判定を計算に置き換える実装法による高速化について検討する。

$T = 225000, L = 625, W = 256$ とすると $J = 438$ であるので距離行列に要する計算量は96141(偶数番目)及び95703(奇数番目)となるのでその計算に必要な距離計算の増加は191,844 (= 96141 + 95703)となる。従って1つの時系列を基準として2個以上の時系列に対して類似セグメント探索を行う場合に有効な手法となる。

### 4.4 3時系列類似セグメント探索の評価

3次元空間における $l_1$ 球は正八面体である。それを用いて式(10)を満たすように3次元空間を被覆することは困難である。そこで式(12)で述べたように3次元空間での $l_1$ 球に内接する立方体( $l_\infty$ 球)を利用して3次元空間を被覆する。これは半径 $W$ の正八面体に内接する半径 $W/3$ の立方体である。再帰分割は立方体を8分割して半径 $W/6$ の立方体を作り、それぞれの中心に対して距離を求め式(11)の判定を行う。

3時系列における類似セグメント探索に対してAS法を繰り返し適用するAS反復法や2時系列における類似セグメント探索RDDS-2を用いて候補作成する方法が考えられるが本報告では全探索法と比較することにする。

評価実験にはCampusWaveデータベースの第1回、第2回及び第3回のデータを用いた。1時間の時系列に対して全探索の処理時間が膨大となるため各々の最初の60秒( $T = 3126$ )を用いた。全探索の点の組み合わせ数は

表 1: 従来の RDDS-2 法と提案手法 (RDDS-2+DM) の探索性能の比較.

実験条件		RDDS-2 法		提案手法 (RDDS-2+DM)	
窓長 $W$	探索閾値 $\theta$	距離計算回数	処理時間 (s)	距離計算回数	処理時間 (s)
256	0.1	618435	1.70	506663	1.94(0.05)
256	0.2	794521	1.74	715793	1.89(0.05)
128	0.1	1561240	1.93	414894	2.40(0.18)
128	0.2	1612116	1.94	575058	2.61(0.18)

$T^3 = 3126^3 = 3.05 \times 10^{10}$  となる. なお,  $\theta = 0.3, W = 256, L = 625$ , 再帰呼び出しの最小探索半径を  $W_{\min} = 0$  とした.  $W = 256$  であるので立方体の個数はその 1 辺  $2W/3$  の大きさを用いて  $(T/(2W/3))^3 \sim 18^3 = 5832$  となる.

図 7 に 3 つの時系列に対する RDDS 法と全探索の探索結果を示す. 上が全探索, 下が RDDS-3 の結果である. 二つの探索結果を区別するために全探索の結果をずらして表示している. この結果から RDDS-3 は全探索と同一の結果を与えており正しく動作していることが分かる. 表 2 に最小探索半径  $W_{\min}$  を変化させた時の探索処理時間, 距離計算回数及び全探索に対する改善比を示す. ここで全探索の計算時間は 7240 秒, 距離計算回数は約  $3.05^{10}$  である.  $W_{\min} = 0$  の時, 距離計算回数は 3491 倍改善されることが分かる.  $W_{\min}$  が大きくなるに連れて処理時間, 距離計算回数は改善される.  $W_{\min} = 1, 2$  の時の探索結果は  $W_{\min} = 0$  の結果とほぼ同じ結果を与えるが,  $W_{\min} = 4, 8$  と値を大きくすると, 検出結果が異なってくる. 各々 1 時間のデータに対して RDDS-3 を適用した場合の探索処理時間は 1040 秒であり全探索の処理時間は  $7240 * 60^3$  と推定されるので, 改善比は  $1.504 * 10^6 (= 7240 * 60^3 / 1040)$  となり RDDS-3 の有効性が分かる.

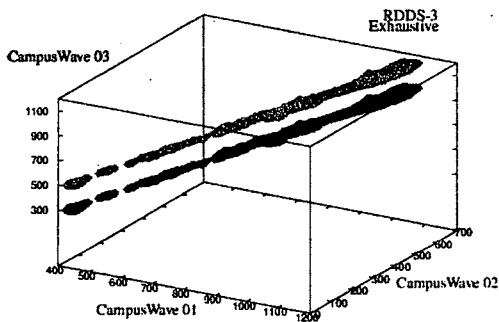


図 7: RDDS-3(下)と全探索(上)との探索結果の比較.

表 2: RDDS-3 の探索処理時間, 距離計算回数と全探索からの改善比.

最小探索半径 ( $W_{\min}$ )	処理時間 (s)	改善比	距離計算回数	改善比
0	3.90	1,856	8,748,795	3,491
1	1.02	7,098	2,539,347	12,029
2	0.25	28,960	599,154	50,983
4	0.09	80,444	167,538	182,328
8	0.05	144,800	57,634	530,015

## 5 むすび

複数時系列中の類似セグメント探索 RDDS 法に関して得られた検討結果を述べその有効性を示した. 今後は  $I = 2$  の記憶容量の削減, 類似画像探索の実装と評価, インタラクティブ探索表示システムの構築, などを検討する.

## 参考文献

- [1] 柏野, G.A.Smith, 村瀬: ヒストグラム特徴列に基づく長時間音響信号の高速探索, 音学論, 2-9-24, pp.561-562 (1998-09).
- [2] 杉山雅英: セグメントの高速探索法, 音声研資, SP98-141, pp.39-45 (1999-02).
- [3] 西村, 他: 時系列パターンの検索手法, 信学技報, PRMU99-125, pp.173-180 (1999-11).
- [4] 西村, 他: アクティブ探索法による時系列データ中の一致区間検出 — 参照区間自由時系列アクティブ探索法 —, 電子情報通信学会論文誌, D-II, Vol. J84-D-II, No.8, pp.1826-1837 (2001-08).
- [5] 柏野, 他: 二つの音響信号に共通に現れる部分信号区間の高速自動抽出, 1-Q-1, pp.133-134 (2000-3).
- [6] 杉山雅英, 複数時系列中の類似セグメント探索法の提案, 音学論, 1-1-9 (2006-03).
- [7] 杉山雅英: セグメント探索のためのノルム及び集合の賭性質, 情報処理学会東北支部第 1 回研究会, 15 (2005-12).
- [8] 杉山雅英, I, 距離を用いた Active 探索法の計算量について, 情報処理学会東北支部第 5 回研究会 (2006-03).
- [9] 杉山雅英, 複数時系列中の類似セグメント探索法の提案と評価, SP2005-196, pp.71-76 (2006-03).
- [10] 数学辞典, 岩波書店.
- [11] 内田, 杉山, CampusWave 音声データベースの作成, 電気関係学会東北支部連合大会, 2A-6 (2000-08).
- [12] R.Vidal: An Algorithm for Finding Nearest Neighbor in (Approximately) Constant Average Time, Pattern Recognition Letters, No.4, pp.145-158 (1986).