

## セグメント高速探索法とその評価

岡本 知子†, 杉山 雅英†

†会津大学 大学院 コンピュータ理工学研究科

〒965-8580 会津若松市一箕町

E-mail: †{m5081130, sugiyama}@u-aizu.ac.jp

### あらまし

蓄積信号系列から目的信号系列のセグメント探索問題は信号を特徴ベクトル時系列を出現確率ベクトルに変換することでベクトル探索問題となる。これまでに Active 探索法の高速化の手法として区分化クラスタリング中心判定法 (DSCC) を提案しその有効性を示した。区分化クラスタリング中心判定法は区分化中心判定法 (M1)、クラスター中心判定法 (M2)、距離ブルーニング法 (M3) から構成され、これらで判定ができない区間に対して Active 探索を適用する。本報告ではクラスター中心判定法と距離ブルーニング法の探索における距離計算回数、処理時間の比較を行なう。また区分化中心判定法の高速化を検討しその有効性を示す。

キーワード セグメント探索, クラスタリング, 出現確率, VQ

## An Efficient Segment Searching and its Experimental Evaluation

Tomoko OKAMOTO†, Masahide SUGIYAMA†

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

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

E-mail: †{m5081130, sugiyama}@u-aizu.ac.jp

### Abstract

The problem of searching query signal from stored signal is formulated as a vector search problem where a series of feature vectors extracted from signal is converted to an output probability function. As an efficient searching algorithm we have proposed DSCC (Determination using Segmentation-Clustering-Centers) method and have shown its effectiveness. DSCC is composed three methods; M1 (Determination using Segmentation-Centers), M2 (Determination using Clustering-Centers) and M3 (Distance Pruning). When a current vector can not be determined using these three methods, AS (Active Search) is applied. This report describes the difference between M2 and M3 methods. Furthermore, the improvement of M1 and its effectiveness is described.

**Keyword** Segment Searching, Clustering, Output Probability, VQ

### 1 まえがき

時空間系列において指定の系列を効率的に探索するためにさまざまな手法が提案されてきた。[1, 2] では指定系列長ごとに出現確率に変換し時系列におけるベクトル間の変化の上限を利用した Active 探索法が提案されその有効性が実験的に示されている。Active 探索法を高速化するために区分化やクラスタリングで枝刈りを行なう手法も提案されている [3, 4]。また [5] では時系列を歪み基準

で区分化クラスタリングし距離行列を用いることにより探索を効率化する手法を述べ、探索実験によりその有効性を示した。本報告では、前処理として既に報告されている最大  $l_1$  歪み区分化法 [5] により区分化処理を行ない、クラスター中心判定法と距離ブルーニング法の前処理及び探索における処理時間や距離計算回数の比較を行なう。この二つの手法は区分化までを同様な処理を行ない、前者はクラスタリングによる中心判定を利用し、後者は距離行列を用いて距離計算の削減に重点を置いた枝刈りを

行なう。この報告ではこれまでに提案したクラスター中心判定法及び距離ブルーニング法について述べ、距離ブルーニング法の改良法を提案する。そして実際に1時間のラジオ番組データを用いた探索実験による評価結果について述べる。またその実験結果に基づき高速化が可能であると考えられる区分化中心判定法 (M1) に対する高速化を提案し評価結果を述べる。

## 2 提案手法

### 2.1 準備

**出現確率ベクトルの性質** 時系列の音声の特徴ベクトル  $\mathbf{x}_t (t = 0, 1, \dots, T-1)$  を  $M$  個の VQ 符号帳で量子化して得られる符号ベクトルを  $\mathbf{r}_t$  とする。 $\mathbf{r}_t$  を式 (1) を用いて時間窓  $L$  平均化することにより  $M$  次元の出現確率ベクトル  $\mathbf{p}_t$  に変換する。

$$\mathbf{p}_t = \frac{1}{L} \sum_{l=0}^{L-1} \mathbf{r}_{t+l} \quad (1)$$

このとき時系列に並んだ出現確率ベクトルの任意の二点  $\mathbf{p}_t, \mathbf{p}_{t+n}$  の  $l_1$  距離は不等式 (2) を満たす。ただし、 $n$  は整数値である。

$$d_1(\mathbf{p}_t, \mathbf{p}_{t+n}) \leq \frac{2|n|}{L} \quad (2)$$

**探索問題** 我々が検討する探索問題は式 (3) で定義される。ここで、 $\mathbf{P} = (\mathbf{p}_t)$  を蓄積ベクトル、 $\mathbf{q}$  を目的ベクトル、 $\theta$  を探索閾値とする。これは、図1に示すように目的ベクトル  $\mathbf{q}$  を中心とし半径  $\theta$  とする探索球に属する点を全て検出することである。この問題を効率良く解決する手法を以下に述べる。

$$d_1(\mathbf{p}_t, \mathbf{q}) \leq \theta \quad (3)$$

ここで、 $\mathbf{P} = (\mathbf{p}_t)$  を蓄積ベクトルの時系列、 $\mathbf{q}$  を目的ベクトル、 $\theta$  を探索閾値とする。これは、図1に示すように目的ベクトル  $\mathbf{q}$  を中心とし半径  $\theta$  とする探索球に属する点  $\mathbf{p}_t$  を全て検出することになる。この問題を効率良く解決する手法を以下に述べる。

これまでに提案した高速セグメント探索法は時系列に沿って区分化し判定する区分化中心判定法 (M1)、区分化中心を歪みを用いてクラスタリングし、クラスター中心で判定を行なうクラスター中心判定法 (M2)、距離行列を利用した距離ブルーニング法 (M3) から成り立っている。

#### M1: 区分化中心判定法

小球に区分化した中心  $\mathbf{c}_i$  を用いて判定

#### M2: クラスター中心判定法

$\mathbf{c}_i$  をクラスタリングしその代表ベクトル  $\mathbf{b}_j$  を用いて判定

#### M3: 距離ブルーニング法

$\mathbf{c}_i$  もしくは  $(\mathbf{b}_j)$  の距離行列を作成し相互の距離関係を用いて判定

これらの手法で判定できない場合には Active 探索法を用いる。以下で各手法を述べる。

## 2.2 区分化中心判定法 (M1)

**区分化処理** クラスター中心判定法と距離ブルーニング法に共通する処理として区分化処理がある。区分化手法については最大  $l_1$  歪み区分化法、平均  $l_2$  歪み区分化法及び一定数区分化法を提案した [5]。本報告では、処理時間はかかるが区分化半径が指定以下の大きさとなる最大  $l_1$  歪み区分化法を用いる。

**区分化中心判定法** 区分化中心判定法は、探索球と小球の包含関係を利用して小球に区分化された時系列を判定することにより計算量を削減することを目的としている。 $\mathbf{P} = (\mathbf{p}_t)$  の区分化処理で得られた小球の中心を  $\mathbf{c}_i (i = 0, 1, \dots, I-1)$  とする。また最大  $l_1$  歪み区分化法ではそれぞれの区間の半径は区分化閾値  $\delta$  以下となる。式 (4), (5), (6) では区分化した小球半径が必ずしも  $\delta$  と一致しない場合を含めてそれぞれの小球半径を  $\delta_i$  として記述する。図1は区分化中心判定法の概念を表している。式 (4) の右辺は  $\mathbf{c}_i$  を中心とする半径  $\delta_i$  の球に  $\mathbf{p}_t$  が含まれる場合に成り立つ不等式であり、これを用いて式 (5), (6) が導かれる。式 (5) は小球の中心  $\mathbf{c}_i$  とクエリ  $\mathbf{q}$  との距離が  $\theta + \delta_i$  より大きければその小球に属する  $\mathbf{p}_t$  全てが探索球の外側にある判定でき、同様に式 (6) は小球の中心  $\mathbf{c}_i$  とクエリ  $\mathbf{q}$  との距離が  $\theta - \delta_i$  以下であればその小球に属する  $\mathbf{p}_t$  全てが探索球の内側にあると判定できることを意味している。

$$|d_1(\mathbf{p}_t, \mathbf{q}) - d_1(\mathbf{c}_i, \mathbf{q})| \leq d_1(\mathbf{p}_t, \mathbf{c}_i) \leq \delta_i \quad (4)$$

$$d_1(\mathbf{c}_i, \mathbf{q}) > \theta + \delta_i \rightarrow d_1(\mathbf{p}_t, \mathbf{q}) > \theta \quad (5)$$

$$d_1(\mathbf{c}_i, \mathbf{q}) \leq \theta - \delta_i \rightarrow d_1(\mathbf{p}_t, \mathbf{q}) \leq \theta \quad (6)$$

## 2.3 クラスター中心判定法 (M2)

区分化処理で得られた区分化中心  $\mathbf{c}_i$  を区分化閾値  $\delta$  より大きいクラスタリング閾値  $\delta^*$  でクラスタリングする。クラスター作成の手順は [1] と同じである。区分化中心判定法と同様に探索球とクラスターとの位置関係で判定を行うことにより計算量を削減する。ここで述べる手法は一定数区分化法で区分化した区分化中心判定法 (M1) と共に用いることで、[?] で提案された局所的、大局的クラスタリングを用いた探索とほぼ等価である。

**クラスター中心判定法** 小球の中心 (代表ベクトル) の集合  $C = \{\mathbf{c}_0, \mathbf{c}_1, \dots, \mathbf{c}_{I-1}\}$  は距離の近いベクトルを含む

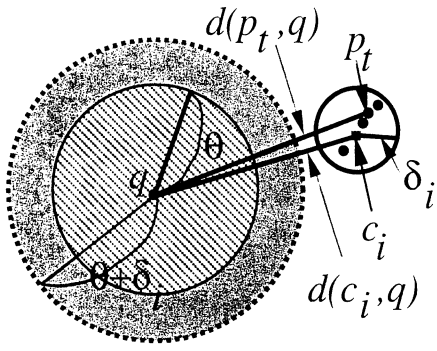


図 1: 区分化中心判定法の概念

可能性があるため  $C$  をクラスタリングしてよりコンパクトな集合  $B = \{b_0, \dots, b_{J-1}\}$  を作成する。ここで  $I > J$  である。またクラスタリング処理は時系列の出現順序に関係なく歪みが閾値  $\delta^*$  以下になるもの全てを同じクラスターに属させることにする。クラスターの中心ベクトル  $b_j$  とそのクラスター  $\wp_j$  内のベクトル  $c_i$  に対して式 (7) が成り立つとする。ここで  $\delta_i$  は小球の半径、 $\delta_j^*$  はクラスターの半径である式 (8) を用いて導かれる。式 (9) から、クラスターの中心  $b_j$  とクエリ  $q$  との距離が  $\theta + \delta_i + \delta_j^*$  より大きい場合はそのクラスターに属する全ての小球の中心  $c_i$  が探索球の外側と判定できる。

$$\forall c_i \in \wp_j, d_1(c_i, b_j) \leq \delta_j^* \quad (7)$$

$$|d_1(c_i, q) - d_1(b_j, q)| \leq d_1(c_i, b_j) \leq \delta_j^* \quad (8)$$

$$d_1(b_j, q) > \theta + \delta_i + \delta_j^* \quad (9)$$

同様に、式 (10) よりクラスターの中心  $b_j$  とクエリ  $q$  との距離が  $\theta - \delta_i - \delta_j^*$  以下の場合はそのクラスターに属する全ての小球の中心が探索球の内側と判定できる。

$$d_1(b_j, q) \leq \theta - \delta_i - \delta_j^* \quad (10)$$

**探索アルゴリズム** クラスター中心判定法 (M2) を利用した探索の手順について述べる。区分化の中心  $c_i$  をクラスタリングし、クラスターの中心  $B = \{b_j\}$  を得る。図 2 にクラスター中心判定法を用いたアルゴリズムの概念を示す。まずクラスターの中心  $b_j (j = 0, 1, \dots, J-1)$  とクエリ  $q$  との距離を計算し式 (9)、(10) で判定し、式 (9)、(10) が成り立たなかったクラスターに属する区分化中心  $c_i$  とクエリ  $q$  との距離を計算し、区分化中心判定法 (M1) を適用する。それでも判定できない小球に対し Active 探索法を適用する。

## 2.4 距離ブルーニング法

区分化処理を終えて得た小球の中心ベクトルの集合  $C = \{c_i\}$  の相互の距離を求めた  $I \times I$  の距離行列を作成する。

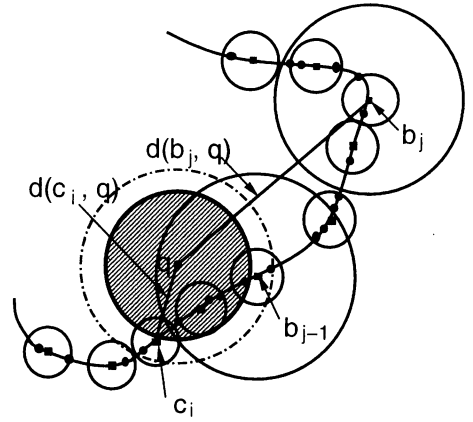


図 2: クラスター中心判定法

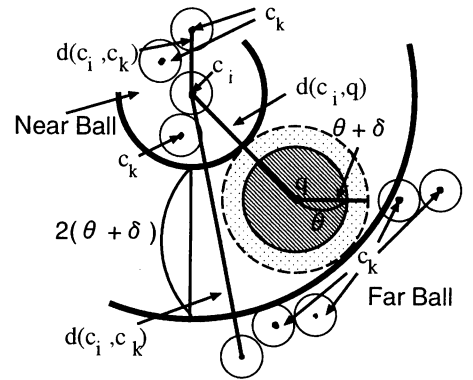


図 3: 距離ブルーニング法

ある小球  $c_i$  と探索セグメント  $q$  の距離  $d_1(c_i, q)$  を求める。式 (11) 又は式 (12) を満たす任意の  $c_k$  は探索球の外側にあるので枝刈りができる。その概念を図 3 に示す。式 (11) では、従来の距離ブルーニング法 [7] と等しく、 $c_i$  を中心、 $d_1(c_i, q) - (\theta + \delta)$  を半径とした球 (図 3 の Near Ball) の内側は除外でき、さらに式 (12) では  $c_i$  を中心、 $d_1(c_i, q) + (\theta + \delta)$  を半径とする球 (図 3 の Far Ball) の外側は全て除外できることになる。この場合、 $2(\theta + \delta)$  の幅で環状に  $c_k$  が残ることになる。距離行列を事前に作成しているため、 $d_1(c_i, c_k)$  はメモリを参照するのみとなる。式 (11)、(12) は、AESA (Approximating and Eliminating Search Algorithm) [9] と呼ばれ、画像の探索に用いられていた手法である。

$$d_1(c_k, q) > d_1(c_i, q) - d_1(c_i, c_k) > \theta + \delta_i \quad (11)$$

$$d_1(c_k, q) > d_1(c_i, c_k) - d_1(c_i, q) > \theta + \delta_i \quad (12)$$

一方、探索球の外側と同様に探索球の内側に含まれる  $c_k$  に対してもブルーニングが可能である。式 (13) を満たす

任意の  $c_k$  を探索球の内側にあると判定できる。

$$d_1(c_k, q) \leq d_1(c_i, q) + d_1(c_i, c_k) \leq \theta + \delta_i \quad (13)$$

**比較判定改良** 本報告では [10] の結果に基づいて比較判定の改良法を提案する。これは従来は式 (11)~(13) の結果に従ってそれぞれの点  $c_i$  が探索球内側か外側かの判定結果を変数に格納していたのを式 (14) の計算に置き換え直接変数に格納することにする。

$$\frac{d_1(c_i, q) - d_1(c_i, c_k)}{\theta - \delta} \quad (14)$$

式 (14) では式 (11) を満たす点は +1 以上の整数値を、式 (12) を満たす点には -1 以下の整数値を、式 (11), (12) を共に満たさない点には 0 を格納する。そのため変数に 0 が格納されてる点に対して判定を行えば良いことになる。式 (13) の探索球の内側に関しては  $\theta = 0.1$  のときに 33 探索平均で 0.879 回行われているだけであるため処理時間に大きな影響はないので式 (13) は省くことにする。

### 3 評価実験

#### 3.1 実験条件

表 1 に実験条件をまとめた。評価用データに CampusWave データベース [8] 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 音声などを含んでおり、音声長は約 1 時間である。音声分析条件は、音声特徴量として LPC ケプストラム係数と対数パワー (計 17 次元)、標本化周波数 16kHz、LPC 分析次数 14 次、分析窓長 256 点 (16ms)、フレーム更新周期 256 点 (16ms)、窓関数はハミング窓、高域強調  $(1 - 0.97z^{-1})$  を用いた。また、出現確率計算のための VQ 符号帳は LBG 法を用いて作成し、サイズを 32 とした。探索セグメントの継続長を 10 秒とし蓄積データの中で 1 度だけ出現するコマースル音声部分から 33 箇所とした。実験には Pentium4(2.66GHz)、メモリ 512GB で OS Vine Linux 3.1 を載せたコンピュータを用いた。gcc 3.3.2 では 03 の最適化オプションを用いた。

#### 3.2 M2 及び M3 の前処理における評価

M2 及び M3 の探索の前処理として区分化までは両手法とも同一な処理を行なっているが、クラスター中心判定法でのクラスタリングと距離プルーニング法での距離行列の作成では処理時間は異なるため比較を行なう。区分化中心集合を  $C = \{c_i\}$  とすると、クラスター中心判定法のクラスター作成における距離計算回数は  $k$  をクラスター数としたとき、 $kI - (k + 1)$  となる。一方、区分化中心相互の距離行列の作成には  $(I - 1)^2/2$  の距離計算が必要

表 1: 実験条件

標本化周波数	16kHz
窓長	256 pts (16ms)
フレーム更新周期	256 pts (16ms)
窓関数	ハミング窓
高域強調	$(1 - 0.97z^{-1})$
LPC 分析	14
ケプストラム分析	16
音声長	約 1 時間 $I=221493$ フレーム
探索音声	10 s ( $L=625$ フレーム)
探索個数	33 points
VQ 符号長サイズ	$M = 32$

となる。表 5 に探索閾値  $\theta = 0.1$ 、区分化閾値  $\delta = 0.028$  と設定した時の前処理における距離計算回数及びその処理時間を示す。また、このときの区分化中心  $c_i$  の数は約  $I = 10000$  である。

表 2: M2 及び M3 の前処理の比較

手法	$\delta^*$	距離計算回数	処理時間 (sec)
M2	0.48	2070031	0.52
M3	-	約 $((10)^8/2)$	17.140

#### 3.3 探索における評価

探索法を距離計算回数と処理時間について評価した。まず、手法の評価の一つの基準として Active 探索法のみで探索をした結果を表 3 に示す。

表 3: Active 探索のみによる探索結果 (33 個のクエリに対する平均)

探索閾値 $\theta$	距離計算回数	処理時間 (sec)
0.05	665.273	1.148
0.10	704.212	1.218
0.20	794.273	1.372

表 4 は M1+M2+AS のクラスタリング閾値  $\delta^*$  を変化させた時の Active 探索部と M1+M2 の処理時間及び距離計算回数を示す。ここで、探索閾値  $\theta = 0.1$ 、区分化閾値  $\delta = 0.028$  と設定し、距離計算回数、処理時間共に 33 個のクエリに対する平均値を求めている。表 4 より、探索閾値  $\theta = 0.1$  のときに M1+M2+AS の距離計算回数は約 535 回と Active 探索のみの場合の約 704 回より約 1.316 倍の改善が見られる。処理時間に関しては、約 50 倍の改善である。表 5 に M1+M3+AS 手法の Active 探

表 4: クラスタリング閾値を変化させたときの M1+M2+AS 法の探索結果 (33 個のクエリに対する平均) 区分化閾値:  $\delta = 0.028$

クラスタリング 閾値 ( $\delta^*$ )	平均 要素数	処理時間		距離計算回数			
		AS	M1+M2	M1	M2	AS	合計
0.10	2.805	0.0136	0.0133	13.636	3614	15.727	3643.363
0.20	6.167	0.0133	0.0118	22.182	1644	15.727	1681.909
0.48	24.973	0.0142	0.0100	113.455	406	15.727	535.182
0.60	41.215	0.0142	0.0100	393.333	246	15.727	655.060
0.80	102.414	0.0136	0.0112	1761.182	99	15.727	1875.909

表 5: 距離プルーニング法の探索結果 (33 個のクエリに対する平均) 区分化閾値:  $\delta = 0.028$

	処理時間		距離計算回数	M3 プルーニングの判定回数		合計
	AS	M1+M2		Near(式 (11))	Far(式 (12))	
	従来手法 (式 (11) のみ)	0.0155	0.0055	53.909	10099.545	
提案手法 (式 (11), (12))	0.0155	0.0042	28.576	3204.576	6920.303	10124.879
式 (12)	0.0158	0.0176	418.576		9734.879	9734.879

素と M1+M3 の処理時間及び距離計算回数を示す。また、探索閾値  $\theta = 0.1$ 、区分化閾値は  $\delta = 0.028$  である。ここで表 5 は、距離プルーニング法の従来手法と提案手法の違いを示し、従来手法では式 (11) を利用し、提案手法は式 (11), (12) を共に利用し、式 (12) のみを利用してプルーニングした結果を意味する。また、提案手法で式 (11), 式 (12) のどちらの判定が効果的に動作しているかをそれぞれの式で除外されるフレーム数で示した。提案手法の全体の距離計算回数で比較すると探索閾値  $\theta = 0.1$  のときに Active 探索のみの場合、約 704 回が M1+M3+AS の距離計算回数では約 30 回と Active 探索の 23.47 倍の改善となる。処理時間で比較した場合、M1+M3+AS が 0.02 秒と Active 探索の約 61 倍の改善である。式 (11), (12) を比較するとこれらの式を同時に利用する場合 (提案手法) は、式 (12) で除外されるフレーム数の方が多いが、それぞれ単独で用いる場合には式 (11) の方が効率良く動作している。これは、式 (12) は  $c_i$  から離れた領域をプルーニングするため探索球に近づくまでに距離計算を多数行うためであると考えられる。M1 と M3 を比較すると M1+M2+AS の距離計算回数の最適値としてクラスタリング閾値  $\delta^* = 0.48$  のときに距離計算回数が約 535 回に対し、M1+M3+AS は、29.848 回と約 18 倍の改善である。また、Active 探索部の距離計算回数は両手法とも同一であるためその数を除いた回数では約 37 倍の改善である。一方、処理時間での比較を行うと M1+M2+AS は M1+M2 部の処理時間が 0.0100 秒であるのに対し、M1+M3+AS 手法の M1+M3 部の処理時間は 0.0021 秒と 4.619 倍の改善である。

### 3.4 区分化中心判定法 (M1) の高速化

表 4 で示したようにクラスター中心判定法は、クラスタリング閾値  $\delta^*$  が大きくなるほど一つのクラスターに属する要素 (小球) が多くなるため、判定できなかったクラスターに属する点に対して多くの距離計算を行うことになる。一つの小球に  $c_i$  属するフレーム数は平均 24 フレームであり、これは平均 24 フレームに一度距離計算を行うことを意味している。一方、Active 探索法での平均スキップ幅は 309.15 フレームである。従って、小球の中心毎に M1 の判定を行うと Active 探索より効率が悪くなる可能性がある。以下に区分化中心判定の高速化を述べる。

**高速化手法** 提案する高速化手法はフレーム間で行なっていた Active 探索法を小球に応用するものである。最大  $l_1$  歪み区分化法で区分化した小球の中心を  $c_i$  とし、それぞれの区間に属するフレーム数  $f_i$  を保持する。ここで  $c_i$  は  $c_i = p_t$  となる  $p_t$  が存在するので  $c_i$  における時系列の時刻を  $t_i$  とする。任意のフレーム間の距離は式 (2) で与えられスキップ幅の上限値は式 (15) で与えられる。

$$|n| \leq L(d_1(p_t, q) - \theta) / 2 \quad (15)$$

式 (16) は Active 探索のスキップする出現確率ベクトルに対応させたものである。式 (16) の  $n$  はスキップする小球の数を示す。

$$d_1(c_i, c_{i+n}) < \frac{2}{L} |t_i - t_{i+n}| \quad (16)$$

本報告では  $f_i$  を用いて  $n$  を求める方法を用いた。  $c_i$  か

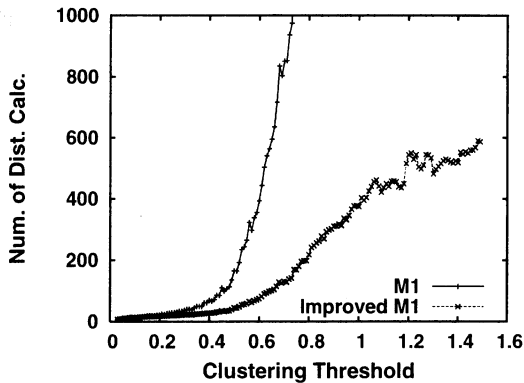


図 4: M1 判定の高速化による距離計算回数とクラスタリング閾値との関係

らのスキップ幅を用いて、

$$\sum_{l=i}^{i+n} f_l < \frac{L}{2}(d_1(c_i, q) - \theta) \quad (17)$$

を満たす最大の  $n$  が区間のスキップ幅になる。

この手法を M1+M2+AS 手法に適用する。まず、クラスター中心判定法で判定できなかったクラスターに属する区間の  $c_i$  とクエリ  $q$  との距離を計算し、最大スキップ幅を式 (15) で求め、その値から式 (17) を満たす最大スキップ小球数  $n$  を求める。それでも判定できない点について Active 探索法を適用する。この手法は M1+M2+AS 手法だけでなく区分化中心判定法 (M1) を用いる手法全てに適用できる。この手法はそれぞれの区間 (小球) にいくつかの点があるかを利用した手法であるため、区分化に最大  $l_1$  歪み区分化法を利用した代わりに一定数区分化法 [5] を用いるとより直接に  $n$  を算出できる。

### 3.5 高速化 M1 における評価実験

提案手法の評価として使用したデータ及び実験条件は前述と同様に設定した。ここで提案手法である M1 部の評価として表 4 の M1 の距離計算回数と M1+M2 の処理時間を比較する。 $\delta^* = 0.48$  のとき提案手法を適用すると距離計算回数が約 535 回から約 457 回に改善される。クラスタリング閾値を大きくすると判定できないクラスターに属する区間が大きくなるため改善の効果は大きい。例えば、 $\delta^* = 0.8$  のとき約 1876 回から約 321 回に改善され、また図 4 に M1 部の距離計算回数を比較した。M1+M2 部の処理時間を比較すると  $\delta^* = 0.8$  のとき約 0.0112 秒が 0.0088 秒と約 1.3 倍改善された。

## 4 むすび

本報告ではクラスター中心判定法 (M2) と距離ブルーニング法の実装とその評価を行った。Active 探索法を除いた距離計算回数で評価すると、M1+M3+AS 手法が M1+M2+AS と比較し 36.786 倍改善された。Active 探索処理時間を除く処理時間で比較すると、4.619 倍の改善であった。また、M1+M2+AS 法における M1 判定の高速化について検討し、距離計算回数改善効果のあることを示した。今後は処理時間の改善や実験データの大容量化及びノイズの加えられたデータに対する探索実験を検討する。

## 謝辞

日頃有益な討論と助言をいただくヒューマンインタフェース学講座の皆様へ感謝します。

## 参考文献

- [1] G.A.Smith, H.Murase, K.Kashino, Quick Audio Retrieval Using ActiveSearch, ICASSP98, Vol.6, p.3777 (1998).
- [2] 柏野, G.A.Smith, 村瀬, ヒストグラム特徴系列に基づく長時間音響信号の高速探索, 音学講論, 2-9-24, pp.561-562 (1998-09).
- [3] 木村, 柏野, 黒住, 村瀬, グローバルな枝刈りを導入した音や映像の高速探索, 信学論, D-II, Vol.J85-D-II, No.10, pp.1552-1562 (Oct. 2002).
- [4] K.Kashino, A.Kimura, T.Kurozumi, A Quick Video Search Method, Proc. of ICPR2004, Vol.3, pp.894-897 (Aug. 2004).
- [5] 杉山, 時系列の歪み一定区分化法とセグメント探索における評価, 音声研資, SP2005-6, pp.1-6 (2005-05).
- [6] 杉山, 幾何学的性質を用いたセグメント高速探索法, 音学講論, 2-1-8 (2005-09).
- [7] 杉山, 出現確率の幾何学的性質を用いたセグメント高速探索法, 信学技報, PRMU2005-22, pp.1-6 (2005-06).
- [8] 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).
- [9] R. Vidal, An Algorithm for Finding Nearest Neighbor in (Approximately) Constant Average Time, Pattern Recognition Letters, No. 4, pp. 145-158, 1986.
- [10] 岡本, 杉山, セグメント高速探索法とその評価, 音学講論, 1-P-5 (2005-09).