数理解析研究所講究録
229巻 1975年 89-98

DPマッチング法とその応用

日本電気株式会社中央研究所　迫江博昭

1. まえがき

音声パタンは代表的な時系列パタンであってその時間軸正規化の問題に対しては古くから種々の試みがなされている。ここで述べるDPマッチングはAnalysis by Synthesisの考えに立脚するものである。すなわち，時間軸変動を非線形変換で近似してモデル化し，モデル化された変動のもとで入力パタンと標準パタンの間の最大一致をとることによって時間軸変動の影響を除去する。実際の計算の段階でダイナミックプログラミングが有効に利用されるのでこの方法をDPマッチング法と称している。

以下ではDPマッチングの考え方とその実行アルゴリズムを述べ，次いで単語音声認識，連続単語音声認識，識別関数学習，手書き文字認識などへの適用・実験例を示している。

2. DPマッチング

2.1 時間軸変動のモデルと時間正規化類似度

音声パタンをベクトル値を取る時間関数として

$$A = \mathbb{A}(t) \quad ; \quad 0 \leqq t \leqq T_A \qquad (1)$$

で示す。時間軸の変動を時間軸 t の非線形変換

$$t = u(s) \quad ; \quad 0 \leqq s \leqq T \qquad (2)$$

によって近似する。ただし,

$$u(s) \in \mathbb{U} \qquad (3)$$

ここに, $\mathbb{U}$ は変換されたパタン $A' = \mathbb{A}(u(s))$ が A と同じ単語クラスに属するような $u(s)$ の集合を規定する。同様に別の音声パタン

$$B = \mathbb{B}(\tau) \quad ; \quad 0 \leqq \tau \leqq T_B \qquad (4)$$

を考え, (2)(3) 式に対応して次の(5)(6) 式を考える。

$$\tau = v(s) \quad ; \quad 0 \leqq s \leqq T \qquad (5)$$

$$v(s) \in \mathbb{V} \qquad (6)$$

A と B の時間軸正規化類似度(以下類似度と略称)を

$$S(A, B) = \frac{1}{T} \min_{u(s),\, v(s)} \left[\int_0^T \| \mathbb{A}(u(s)) - \mathbb{B}(v(s)) \| \, ds \right] \qquad (7)$$

と定義する。この尺度は $A' = \mathbb{A}(u(s))$ の集合 $\{A'\}$ と $B' = \mathbb{B}(v(s))$ の集合 $\{B'\}$ の間の距離となっており (Fig. 1), 次の性質をもつ。

(i) $S(A, B) \geqq 0$ (ii) $S(A, A) = 0$

(iii) $S(A, B) = S(\tilde{A}, \tilde{B})$ if $\tilde{A} \in \{A'\}$ $\tilde{B} \in \{B'\}$

このように時間軸変動に対して安定な性質は音声パタンのマッチング尺度として理想的であると言えるが，以上の定義をそのまま実行することは困難で，いくつかの近似が必要である。

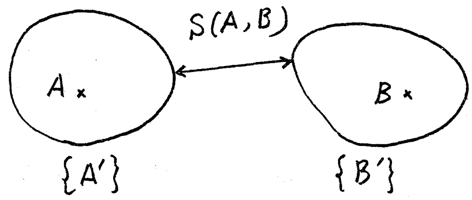


Fig. 1 S(A,B)の定義

2.2 類似度の近似式（対称形）

(3)(6)式の条件は具体的には，音声パタンの各音素の継続長やスペクトラムの変化速度を限定するなどの意味を持つが，この条件を完全に実現するには音素レベルの認識が完了していなくてはならず実際的でない。よって次の条件で(3)式を近似する（(6)式の$v(s)$に対しても同様）

$$\left.\begin{array}{ll} \text{(i) } u(s) \text{ は連続} & \text{(ii) } u(s) \text{ は単調増加} \\ \text{(iii) } u(s) \sim s & \text{すなわち } u(s) \text{ の値は } s \text{ の近傍にある} \end{array}\right\} \quad (8)$$

これらの条件のもとにDPを適用することによって(7)式の最小問題は一応計算可能であるが，$u(s)$と$v(s)$に関する2次元の最小問題となり所用計算量は必ずしも少なくない。よって次のような近似で1次元の問題に縮小する。

対称形の近似 $$s = (u(s) + v(s))/2 = (t + \tau)/2 \quad (9)$$

これより $$ds = (dt + d\tau)/2 \quad (10)$$

となり類似度S(A，B)は次のように近似される。

$$S_1(A,B) = \frac{1}{(T_A+T_B)} \min_{w(t)} \left[\int_{\tau=w(t)} \| A(t) - B(\tau) \| (dt + d\tau) \right] \quad (11)$$

ここに, $\quad w(t) = u(v^{-1}(t)) \quad ; \quad 0 \leq t \leq T_A \quad (12)$

2.3 DPによる計算 (1) (2)

音声パタンA, Bを時間標本化して

$$A = a_1, a_2 \cdots, a_i, \cdots a_I \; ; \; B = b_1, b_2, \cdots, b_j, \cdots b_J \quad (13)$$

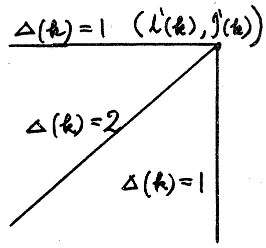
と示す．i と j の組合せの点 (i, j) を $(i+j)$ に関して正順にとり，番号 $k = 1 \sim K \; (K = I + J)$ を付して $(i(k), j(k))$ で示す．ただし，

$$i(1) = 1 \quad j(1) = 1 \quad , \quad i(K) = I \quad j(K) = J \quad (14)$$

とし，かつ隣り合う点の間には (8) の条件 (i)(ii) を考慮して

$$\left.\begin{aligned} & i(k) = i(k-1) \quad , \quad j(k) = j(k-1) + 1 \\ \text{or} \quad & i(k) = i(k-1) + 1 \quad , \quad j(k) = j(k-1) \\ \text{or} \quad & i(k) = i(k-1) + 1 \quad , \quad j(k) = j(k-1) + 1 \end{aligned}\right\} \quad (15)$$

の3種の関係のみを許す。



いま $\quad d(i, j) = \| a_i - b_j \| \quad (16)$

とし，(10) 式に対応して

$$\Delta(k) = (i(k) - i(k-1)) + (j(k) - j(k-1)) \quad (17)$$

とおくと (11) 式は

$$S_1(A,B) = \frac{1}{(I+J)} \min_{i(k), j(k)} \left[\sum_{k=1}^{K} d(i(k), j(k)) \cdot \Delta(k) \right] \quad (18)$$

となり，次のようなDP法の手続きで計算される。

初期条件

$$\left.\begin{aligned} g(0,0) &= 0 \\ g(i,0) &= \infty \quad (i \neq 0) \\ g(0,j) &= \infty \quad (j \neq 0) \end{aligned}\right\} \tag{19}$$

漸化式

$$g(i,j) = \min\begin{bmatrix} d(i,j) + g(i-1,j) \\ d(i,j) + g(i,j-1) \\ 2d(i,j) + g(i-1,j-1) \end{bmatrix} \tag{20}$$

$$(1 \leqq i \leqq I \quad , \quad 1 \leqq j \leqq J)$$

制約条件（整合窓の条件）

$$j - r \leqq i \leqq j + r \tag{21}$$

類似度

$$S(A,B) = \frac{g(I,J)}{(I+J)} \tag{22}$$

(21) 式の整合窓の条件は (8) の(iii) の条件を近似する。

2．4 非対称形の近似[2]

(9) 式のかわりに

$$s = u(s) = t \tag{23}$$

とする。このとき (17) 式は

$$S_2(A,B) = \frac{1}{T_A} \min_{u(t)} \left[\int_0^T \| A(t) - B(\tau) \| \, dt \right]_{\tau = u(t)} \tag{24}$$

となり，対称形の場合とほぼ同様の手続で計算できる。

この場合にはAの時間軸 t を標準として固定してBの時間

軸τを最適に変換し，t軸上で比較していることになる。これに対して対称形ではAとBの時間軸のズレ（t−τ）を最適に補正して(t+τ)軸の上で比較していることになる(Fig.2)。

対称形と非対称形の優劣を一般的に論ずることはむずかしい。しかし対称形では

$$0 \leqq \frac{dt}{ds}, \frac{d\tau}{ds} \leqq 1$$

非対称形では

$$\frac{dt}{ds}=1, \ 0 \leqq \frac{d\tau}{ds} < \infty$$

であることを考えると，対称形の方が両方のパタンを"より均等に比較する"性質をもっていると言える。実際には，AとBの時間軸が元来異なる(たとえば標本化周期が異なる)時は一方の時間軸を基準と考えて非対称形を用い，そうでない時は対称形によるのが適当であろう。

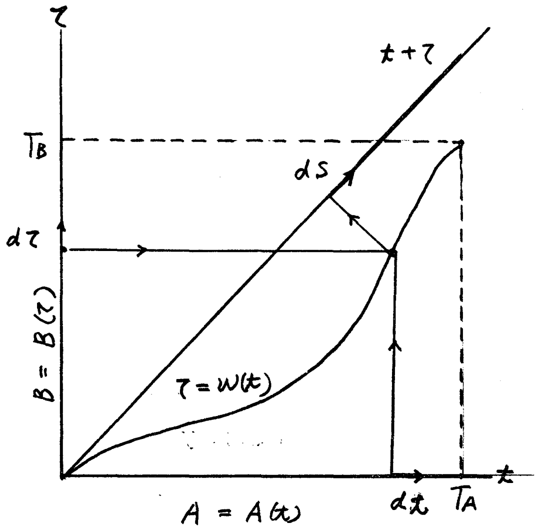


Fig.2 対称形DPマッチング

3. DPマッチングの応用と実験例

3.1 単語音声認識(2)(3)

10チャネルの分析フィルタ出力を20ms周期で標本化したパタンを用い，男性10名，女性5名の合計1500サンプルの数字単語の認識を行なった結果を表1に示す(強制判定)。標準パタンは各人の10数字2回を用いた。この実験では対称形の

類似度による方が非対称形による場合に比して格段に良い結果が得られている.

	男性 10名	女性 5名	合計
対称形	0.1	0.0	0.07
非対称形	0.5	0.0	0.34

表1　数字単語 認識結果　誤り率(%)

男女各2名の地名単語合計1000サンプルを認識した結果98.1% の認識率であった(標準パタンは各単語1個)。極端な時間軸変換を排除してクラス間の分離を改善するために $\tau = w(t)$ の傾斜に制限を導入して,

$$\frac{2}{3} \leqq \frac{dt}{ds}, \frac{d\tau}{ds} \leqq 1 \tag{26}$$

と制限してマッチングを行なうことによって認識率が99.2%に改善された.

3.2 連続単語認識[(1)(4)]

各単語クラス n に標準パタン B^n を設定する。入力パタンAは何個かの単語を連続発声したものであるとする。Aの部分パタン A_l を次のように定義する。

$$A_l = a_1\ a_2 \cdots\cdots a_l \tag{27}$$

(18)式を拡張してAの第1単語とBの類似度を

$$S_1(A, B) = \min_{J-r \leq l < J+r} \left[S_1(A_l, B) \right] \tag{28}$$

と定義する。すべての標準パタン B^n について $S_1(A, B^n)$ を求めてそれが最小となる $n = n_1$ を第1単語の認識結果とする。

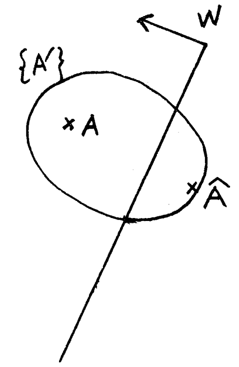
次にB^{n_1}とB^{n}を接続したパタン$B^{n_1}\cdot B^{n}$を標準パタンと考えすべてのnについて$S_1(A, B^{n_1}\cdot B^{n})$を計算し,それが最小となる$n=n_2$を第2単語の認識結果とする。このときDP計算は$B^{n_1}$の部分に関してはすでに終了しているので残りの$B^{n}$の部分について行なうと十分である。第3単語以下に関しても同様である。この手続きでは単語間のセグメンテーションは不用であり,しかも所用計算量はセグメンテーションを行なって認識する場合と同等である。

男性5名の2桁数字合計500個に対する実験結果を表2に示す。標準パタンとしては各人の1桁数字2回を用いている。なお,別の実験では(26)式の傾斜制限を導入することによって誤り率が2.2%から0.5%に低減されるという結果が得られた。

発声者	認識率(%)
A	99.5
B	99.5
C	100
D	99.5
E	99.0
平均	99.6

表2 2桁数字認識結果

3.3 識別関数学習問題への適用[5]



パーセプトロン式の誤訂正過程において,訓練パタンAを時間軸に関して変形して現在の識別関数Wで最も分離しにくい形Âとして用いることによって時間軸変動に対して安定な識別関数が能率良く得られる。パタンの変形は非

対称形のDPマッチングで効率良く処理できる。男性1名の数字音声に関する実験では,通常の学習で91.3%であつたものが,$r=3$の範囲で変形して学習することによつて99.5%に改善された。なお,この方法はLPによる識別関数計算にも適用できる。

3.4 手書き文字認識への適用(Rubber String matching[6])

手書き文字は本質的に線図形であつて線分の系列として近似できる。各線分の方向を固定して線分長を独立に変化させることによつて文字パタンの変動をモデル化する。このモデルによると文字の拡大・縮小の他に回転や字体の本質的な変形をも良好に近似できる。よつて標準パタンを方向が指定される線分の系列で与え,各線分の長さを変数として,2次元メッシュ状の入力パタンとの間に一致度の評価関数を加法的に定義してその最大化をDPによつて行なうことにより,文字パタンの変動に対して安定なマッチングを効率よく行なうことができる。

20名の手書き数字2000サンプルに対する予備的な実験では12個の標準パタンで99.95%正しく認識できるという結果を得ている。

この方法は,元来時間軸の無い2次元パタンに自動的に時間軸を発生し,正規化し,マッチングを行なう手続きになつ

ており，マッチングの結果として入力パタンを線分系列で最適に近似したパタンが得られるという特徴をもっている。

4. むすび

以上，DPマッチングの考え方とその実行アルゴリズム及び応用・実験例を述べた。2.2及び2.3節で導入した幾つかの近似は，実験結果から見て，実用上さしつかえないものと考えられる。実際には所用計算量が問題になるが，NEAC M4/n ミニコンピュータに専用のDPプロセッサを接続したシステムで100語の単語セットに対して実時間認識が可能なシステムをすでに実現している[7]。また文字パタンに対しても十数字の文字セットに対して毎秒100字程度の読取り速度をもつ装置を構成することは容易であると考えている。

謝辞　日頃御指導いただく加藤研究課長に深謝します。また以上の研究のほとんどについて指導いただき，また共同研究者でもあった千葉研究主任に心から謝意を表します。

文献

(1) 迫江・千葉　音響誌　Vol.27 No.9 P.483　(1971)
(2) 迫江・千葉　音響学会 音声研資　S73－22　(1973)
(3) 迫江　音学大会予稿　1－2－15　(1974－6)
(4) 迫江　〃 〃 〃　2－2－19　(1974－10)
(5) 迫江　信学全大予稿　53　(1972)
(6) 迫江　信学研資　PRL－74－20　(1974)
(7) 鶴田・迫江・千葉　信学全大予稿　1596　(1974)