6N-7

音声テンプレート統合法

菅原一秀
日本アイ・ビー・エム株式会社　サイエンス・インスティチュート

1．はじめに

テンプレートを用いる音声認識ではその良否が直接認識結果に反映される。テンプレートの最も簡単な作成法は、ある一つの発声をもってテンプレートとする方法である。しかしこの方法は一つの発声のみに頼るので偏りが避けられないし、発声の変動が大きくなったり、テンプレートを適応化させようとした時にはうまくいかない。これらの問題点を解決するためにテンプレートを複数個使う方式（マルチ・テンプレート法）があるが、与えられた学習用発声全部をそのままテンプレートとして持つのでは記憶域及び計算量が大きくなってしまい、得策ではない。そこで、与えられたテンプレートから一つあるいは少数のテンプレートを作りだす方法が要求される。本稿ではその方法の一つとして、「重み付きＤＰマッチング」を利用した新しいテンプレート作成法（AWDP法:Average by Weighted DP）を示す。

2．従来技術の問題点とそれらの解決法

いくつかのテンプレートをＤＰマッチングにより統合する代表的な方法として次の二つがある。(1)どれか一つのテンプレートを選びこれを軸にして他のテンプレートを最適パスを介しこれに投影し、平均化などの操作により新しいテンプレートを得る。(PDP法：Projection by DP) (2)最適パス上で二つのテンプレートのそれぞれの重みに基づき内分して新しいテンプレートを得る。(新美[2] ADP法:Average by DP)

(1)のPDP法には「どのテンプレートを軸として選ぶか」、「特徴量は平均化されるが、時間長は軸として選ばれたもののが使われる」等テンプレートの扱いが不均等であるという問題点がある。また、(2)のADP法にはテンプレートが３個以上のときそれらを組合せる順序により結果が異なってくるという問題がある。

これらを解決するためには、現在二つのテンプレートに対して行われているＤＰマッチングを任意個数Ｎ個のテンプレートに対して行えるよう拡張することが考えられる。しかし、この場合計算量はＮに対し指数関数的に増大し少量のデータに対しても実行が困難になる。

ADP法が通常のＤＰで最適パスを求めた後、二つのテンプレートに対する重みで内分するのに対し、本稿で提案する AWDP法 はＤＰマッチングの段階で重みを考慮して最適パスを求めるので、各テンプレートの重みをより良く反映できることが期待される。

3．テンプレート統合法

以下に説明する１～３の手順による。

1．　重み付きＤＰマッチング
二つのテンプレートを w:(1-w) (0≦w≦1) の重みでマッチングする

2．　最適パスの追跡
１．のマッチングで得られた最適パスは正規化時間に対応しており、これをたどって新しいテンプレートを得る

3．　複数テンプレートの統合
上記１，２の方法を組合せて、Ｎ個（２≦Ｎ）のテンプレートから一つのテンプレートを作る

3．1　重み付きＤＰマッチング

統合すべきテンプレートを

$A : a_1 \ldots a_m$

$B : b_1 \ldots b_n$

とする。ＤＰマッチングを図１の様な荷重で行う。つまり、下の漸化式により 累積距離 g を計算する。

$$g(1,1) = d(a_1,b_1)$$

$$g(i,j) = \min \begin{cases} g(i-1,j)+w*d(a_i,b_j) \\ g(i,j-1)+(1-w)*d(a_i,b_j) \\ g(i-1,j-1)+d(a_i,b_j) \end{cases}$$

$(1 \leq i \leq m,\ 1 \leq j \leq n)$

ここで d は局所距離

また、パスの長さを
やはり図１の様に定義する。

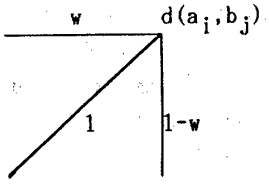

図 1．荷重とパス長

Speech Template Merging Method
Kazuhide SUGAWARA
Science Institute, IBM Japan, Ltd.

補足： 前項の説明で $w=p/(p+q)$, $1-w=q/(p+q)$ （p, q は自然数）に選ぶと、上の定義は p+q 次元での、次のような制限付のマッチングを縮退させたものと考えることができる。

1. 1～p次元のテンプレートはAのp個のコピー
2. p＋1～p＋q次元のテンプレートはBのq個のコピー
3. パスの荷重はその両端の市街地距離
4. マッチングは $\{(x_1,\ldots,x_{p+q}) \mid x_1=\ldots=x_p=X,\ x_{p+1}=\ldots=x_{p+q}=Y\}$ で定まる空間で行われる。

このマッチングを2次元に縮退させる方法は明らか。パスの重みを正規化することにより図1．の定義を得る。台形型ＤＰマッチング(大河内[1])を用いればより性質の良い最適パス(正規化時間)が得られる。

3．2 最適パスの追跡

3.1 のマッチングで得た最適パスの長さは $w*m+(1-w)*n$ である。これを原点から出発してたどり、パスの長さの累積が整数 k になる点 p_k $(1 \leq k \leq w*m+(1-w)*n)$ で新しい特徴量 c_k を求める。例えば、p_kに一番近い格子点上でＡとＢの $(1-w):w$ の内分をとることが考えられる。

3．3 複数テンプレートの統合

3.1, 3.2 の方法を使い、二個のテンプレートＡ，Ｂから、それらを $w:(1-w)$ の割合で反映した新しいテンプレートＣができる。Ｎ個のテンプレートが与えられた時にはこの方法を組合せて、一つのテンプレートを作ることができる。そのとき、中間的にできるテンプレートに対してはそれを作るために使用したテンプレート数に比例した重みを掛ける。

5．実験

本稿で説明したAWDP法と従来から提案されていたPDP法及び ADP法による単語音声認識の比較実験を行った。表1．に2発声(発声1,2)、及び3発声(発声1,2,3)からこれらの方法により作成したテンプレートによる単語認識実験の結果を示す。（2発声のときは ADP法と AWDP法は同じ）話者Ａ，Ｂとも、発声 1~5 は同時に収録し、それ以外は 1~2 週間をおいて収録した。認識対象とした発声はテンプレートの作成に使用されなかったもの全部である。

語彙 ： 約 6,000 語から選んだ類似 150 単語
話者 ： 成人男性 2 名
発声 ： 話者Ａ 7回
　　話者Ｂ 6回

基になるテンプレート数（Ｎ） ： 2 及び 3
重み（ｗ） ： 1／Ｎ
ＤＰパス： 対称型(図1．)

表1．誤認識率（%）

	方式		話者 A	話者 B
2発声	単一	1	2.8	1.5
		2	1.87	1.3
	PDP	1→2	1.87	0.5
		2→1	0.8	1.2
	AWDP		1.1	0.67
3発声	ADP		0.8	0.67
	AWDP		0.77	0.72

5．考察

統合テンプレートを使った場合は一つのテンプレートを使った場合に比べて、話者Ａ，Ｂともに誤認識がかなり減少した。このことは本稿で述べたテンプレートの統合方式の有効性を示している。なお、他のＤＰマッチング方式（「補間型」や「スタガード・アレイ方式」）においてパスに重みを付けたものについても実験を行い、上の実験と同様の結果を得ている。

PDP法では、その非対称性により、投影の方向により結果が異なってくることが上の実験においても示されている。ADP法及びAWDP法によればこのような事は少なくなり、（Ｎ＝2の時はまったくない）更に、上の実験では認識率でみてもPDP法によるものより概ね良い結果を得ている。この実験ではADP法とAWDP法はほぼ同等の結果を示していた。不特定話者のためのテンプレート作成時等、統合すべきテンプレートが多くなった場合に両方法間にどのような差が出てくるか今後の検討課題である。 また、両方法ともテンプレートの数が3以上になると組合せる順序により得られるテンプレートが異なる点は解決されていないので、この点も検討を要する。

［参考文献］

[1] 大河内正明，"台形型ＤＰマッチングに関する考察"，音声研究会資料 S81-64 (1981-12)

[2] Niimi, Y., "A Method for Forming Universal Reference Patterns in an Isolated Word Recognition System," Proc. of the 4-th Int. Joint Conf. on Pattern Recognition, 1978, pp. 1022-1024

[3] 新美康永，"音声認識"，共立出版，東京，1979

[4] Niimi, Y. and Kobayashi, Y., "Synthesis of Speaker-Adaptive Word Templates By Concatenation of the Monosyllabic Sounds," Proc. of the ICASSP 86, Tokyo, 1986, pp. 2651-2654