社団法人 情報処理学会 研究報告
IPSJ SIG Technical Reports

2007－ＳＬＰ－69（９）
2007／12／20

音声認識における確率モデルの重み係数の自動推定

江森 正† 大西 祥史† 篠田 浩一††

† 日本電気株式会社 〒211-0001 川崎市中原区下沼部1753
†† 東京工業大学 〒152-8552 東京都目黒区大岡山2-12-1
E-mail: †{t-emori,y-onishi}@bp.jp.nec.com, ††shinoda@cs.titech.ac.jp

あらまし 音声認識における確率モデルのスケーリング係数を効率的に推定する新しい手法を提案する。音声認識システムの多くは、音響モデルと言語モデルで構成されており、それらの値のバランスを取るためにスケーリング係数が導入されている。従来、スケーリング係数の値は事前にテストデータを用い、認識実験を行うことによるグリッドサーチで調整されていた。提案法では、スケーリング係数を対数線形モデルの重みパラメータとし、最適値を事後確率最大化基準 (*maximum a posteriori probability*) を基に勾配法を用いて推定を行う。事後確率はワードラティスを用いて計算を行った。ワードラティスを用いることによる推定値の初期値依存を避けるため繰り返し過程を導入する。繰り返し過程では、ワードラティスの生成と勾配法によるスケーリング係数値の推定が交互に繰り返される。実験の結果、提案手法により推定されたスケーリング係数の値が初期値に依存せずほぼ同じ値に推定されることを確認した。また、提案手法により推定された値を用いた場合の単語正解精度と、グリッドサーチで最適化された値を用いた単語正解精度の差は0.1%であることを確認した。

キーワード 音声認識、スケーリング係数、対数線形モデル、ワードラティス、単語ラティス

Efficient estimation method of scaling factors among probabilistic models in speech recognition

Tadashi EMORI†, Yoshifumi ONISHI†, and Koichi SHINODA††

† NEC Corporation, 1753, Shimonumabe, Nakahara-ku, Kawasaki, 211-8666, Japan
†† Tokyo Institute of Technology 2-12-1, Ookayama, Meguro-ku, Tokyo, 152-8552, Japan
E-mail: †{t-emori,y-onishi}@bp.jp.nec.com, ††shinoda@cs.titech.ac.jp

Abstract We propose a new efficient method for estimating scaling factors among probabilistic models in speech recognition. Most speech recognition systems consist of more than one model, include an acoustic and a language model, and require scaling factors to balance probabilities among them. The scaling factors are conventionally optimized in preliminary recognition tests using data for development. In our proposed method, the scaling factors are regarded as parameters of a log-linear model, and they are estimated using a gradient-ascent method based on the *maximum a posteriori* probability criterion. Posterior probability is computed using word-lattices generated by a speech recognizer. We employ an iteration technique which repeats a word-lattice-generation/scaling-factor-estimation process, and the resulting scaling factor estimation is robust with respect to the changes in initial values. In an experimental evaluation of our method by LVCSR using Japanese dialogue speech data, estimated scaling factors were nearly identical to optimal values obtained in a greedy grid search. We have also confirmed that estimated scaling factors changed little with variations in initial values.

Key words speech recognition, scaling factor, log-linear model, word lattice

1. はじめに

ベイズの枠組みによる統計的な音声認識は、音響モデルと言語モデルで構成されている。この枠組みにおける音声認識は、音響時系列 o が得られたときに、事後確率 $p(w|o)$ が最大となる単語列 $\hat{w}$ を得ることを目的とする。

$$\hat{w} = \underset{w}{\operatorname{argmax}}\, p(w|o) = \underset{w}{\operatorname{argmax}}\, \frac{p(o|w)p(w)}{p(o)} \qquad (1)$$

ここで、$p(o|w)$ は単語列 w が得られたときに観測時系列 o が生成される確率で、隠れマルコフモデル (HMM) で表される。$p(w)$ は単語列 w の観測される確率で、n グラムモデルで表される。音声認識では各モデルの値のダイナミックレンジのバランスを取るためのスケーリング係数や、単語の挿入誤りを減らすための単語挿入ペナルティーが用いられる [1]。このとき、式 (1) は次のように再定義される。

$$\hat{w} \simeq \underset{w}{\operatorname{argmax}} \frac{p(o|w)p(w)^{\kappa_1} e^{\kappa_2 N}}{p(o)} \tag{2}$$

ここで、N は w に含まれる単語数、$e^{\kappa_2 N}$ は単語挿入ペナルティー κ_1, κ_2 はスケーリング係数を表す。これらのスケーリング係数はベイズの枠組みでは最適化することが不可能なため、事前にテストデータを用いた認識テストを行うことで最適化が行われてきた [2]。

近年、統計機械翻訳等の研究分野において、複数の確率モデル対するスケーリング係数の最適化を行うため対数線形モデルが用いられてきた [3] [4]。これらの研究と同様、スケーリング係数 κ_1, κ_2 を対数線形モデルの重みパラメータとすることで、式 (2) を対数線形モデルとして定式化が可能である。すなわち、対数線形モデルを用いたアプローチは、音声認識におけるスケーリング係数の最適化に適用可能である。Beyerlein [5] は、対数線形モデルを用いる代わりに、N ベストリストを用いた "平滑化誤り数基準 (smoothed error count measure)" を最小化する識別的な手法を用いることでスケーリング係数の最適化を行っている。このような識別的な手法では、推定される値は N ベストリストの内容により変化する。また、N ベストリストの内容は、用いられるスケーリング係数の値に依存する。すなわち、N ベストリストの内容と推定されるスケーリング係数の値は相互依存の関係にあるため、どのような場合にも最適なスケーリング係数の値が推定される保証はない。

提案法では、スケーリング係数を対数線形モデルの重みパラメータと見なし、最適値を最大事後確率基準 (*maximum a posteriori probability*) を基にした勾配法を用いて推定する。事後確率の計算にはワードラティスを用いる。ワードラティスには多くの異なる単語列が含まれており、N ベストリストを用いるよりも頑健な推定が可能である。一方で、含まれる単語列数は理想的な場合と比較して依然少ないことから、スケーリング係数の推定値がその内容によって変化する。この問題に対処するため、ワードラティスの生成と勾配法によるスケーリング係数の推定を交互に繰り返す「繰り返し過程」を導入する。

以下、提案手法で用いられる対数線形モデルとその重みパラメータの最適化について、第 2 章で述べる。対数線形モデルの音声認識への適応と、音声認識に適用された対数線形モデルを用いたスケーリング係数の推定手法を 3 章にて説明する。スケーリング係数の初期値依存を避けるための学習手順である繰り返し過程について第 4 章で述べる。提案法を確認するための実験結果を第 5 章に示す。

2. 対数線形モデル

2.1 定　　義

対数線形モデルを用いて、観測系列 x が得られたときのラベル列 y の事後確率は次のように表される。

$$p_{\Lambda}(y|x) = \frac{\exp\left(\sum_{k=1}^{K} \lambda_k F_k(x, y)\right)}{Z_{\Lambda}(x)} \tag{3}$$

ここで、K は素性関数の数であり、$F_k(x, y)$ は k 番目の素性関数、λ_k は k 番目の素性関数の重みパラメータである。$Z_{\Lambda}(x)$ は全ラベル列を考慮したときに、確率の和が 1 になるようにするための正規化項 (分配関数) であり、次のように表される。

$$Z_{\Lambda}(x) = \sum_{y' \in \mathcal{Y}} \exp\left(\sum_{k=1}^{K} \lambda_k F_k(x, y')\right)$$

ここで、$\mathcal{Y}$ は全ての可能なラベル列とする。

2.2 推　　定

本節にて、対数線形モデルの重みパラメータの推定問題について考える。重みパラメータのセット $\mathbf{\Lambda} = (\lambda_1, \cdots, \lambda_k)$ は、学習の基準として事後確率最大化基準を基に学習データセット $(x_r, y_r); r = 1, \cdots, R$ を用いて推定される。これは、全学習データに対する事後確率の対数値の和を $L(\mathbf{\Lambda})$ とすると、次式のように表される。

$$\begin{aligned}
\hat{\mathbf{\Lambda}} &= \underset{\mathbf{\Lambda}}{\operatorname{argmax}}\, L(\mathbf{\Lambda}) \\
L(\mathbf{\Lambda}) &= \sum_{r=1}^{R} \log p_{\mathbf{\Lambda}}(y_r|x_r) \\
&= \sum_{r=1}^{R} \left(\sum_{k=1}^{K} \lambda_k F_k(x_r, y_r) - \log Z_{\mathbf{\Lambda}}(x_r) \right)
\end{aligned} \tag{4}$$

このとき、$L(\mathbf{\Lambda})$ を最大化する重みパラメータの値を求める解析的な方法は無いため、一般には GIS や勾配法のような繰り返し法がしばしば用いられる [6] [7]。勾配法は、式 (4) を最大にするパラメータ λ_k を計算する方法であり、推定のための更新式は次のように表される。

$$\lambda_k^{t+1} = \lambda_k^t + \eta\, \partial_{\lambda_k} L(\mathbf{\Lambda})|_{\mathbf{\Lambda}=\mathbf{\Lambda}^t} \tag{5}$$

ここで、t は繰り返し数、λ_k^t は t 回目のパラメータ値、η は学習定数である。$\partial_{\lambda_k} L(\mathbf{\Lambda})|_{\mathbf{\Lambda}=\mathbf{\Lambda}^t}$ は $L(\mathbf{\Lambda})$ の λ_k に関する勾配で、次の式で計算される。

$$\nabla_{\lambda_k} L(\mathbf{\Lambda})|_{\mathbf{\Lambda}=\mathbf{\Lambda}^t} = S^k - E^k_{\mathbf{\Lambda}^t} \tag{6}$$

ここで、S^k は k 番目の素性関数に関する全学習データの和、$E^k_{\mathbf{\Lambda}^t}$ は学習データ全てについて k 番目の素性関数の期待値の和である。S^k と $E^k_{\mathbf{\Lambda}^t}$ はそれぞれ次のように表される。

$$S^k = \sum_{r=1}^{R} F_k(x_r, y_r) \tag{7}$$

$$\begin{aligned}
E^k_{\mathbf{\Lambda}^t} &= \sum_{r=1}^{R} E_{\mathbf{\Lambda}^t}[F_k(x_r, y)] \\
&= \sum_{r=1}^{R} \sum_{y' \in \mathcal{Y}} p_{\mathbf{\Lambda}^t}(y'|x_r) F_k(x_r, y')
\end{aligned} \tag{8}$$

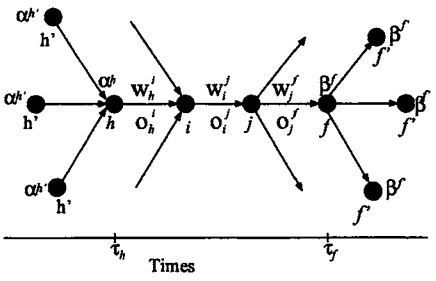

図 1 サンプルワードラティス。点 h', h, i, j, f, f' はノードを表す。τ_h と τ_f はそれぞれノード h と f に対応する時刻を表す。o_i^j と w_i^j はそれぞれノード i とノード j 間のアークに対応する観測量と単語を表す。α^h と β^f はそれぞれ、ノード h の前向きスコア、ノード f の後ろ向きスコアを表す。

$E_{\mathbf{\Lambda}^t}[F_k(x_r, y)]$ は分布関数 $p_{\mathbf{\Lambda}^t}(y|x_r)$ を用いた素性の期待値である。全ての重み係数の値 $\lambda_1, \cdots, \lambda_K$ は式 (9) で表される収束条件を満足するまで繰り返される。

$$\frac{L(\mathbf{\Lambda}^{t+1}) - L(\mathbf{\Lambda}^t)}{L(\mathbf{\Lambda}^t)} \leq D \tag{9}$$

3. 対数線形モデルを用いた音声認識

3.1 素性関数

本節では対数線形モデルを音声認識に適応するため、素性関数および重みパラメータの定義について述べる。対数線形モデルを (2) 式で表される音声認識に適用する場合、重みパラメータや素性関数は次のように定義される。

$$F_1(\boldsymbol{o}, \boldsymbol{w}) = \log p(\boldsymbol{o}|\boldsymbol{w}) \tag{10}$$

$$F_2(\boldsymbol{o}, \boldsymbol{w}) = \log p(\boldsymbol{w}) \tag{11}$$

$$F_3(\boldsymbol{o}, \boldsymbol{w}) = N \tag{12}$$

$$\lambda_1 = 1, \lambda_2 = \kappa_1, \lambda_3 = \kappa_2, \tag{13}$$

式 (3) のラベル系列 $\boldsymbol{y}$ は単語系列 $\boldsymbol{w} = (w_1, \cdots, w_N)$、観測時系列 $\boldsymbol{x}$ はそれぞれの単語に対応する音響観測時系列 $\boldsymbol{o} = (o_1, \cdots, o_N)$ に置き換えられている。ここで o_i は単語 w_i に対応する音響観測時系列、N は単語列 $\boldsymbol{w}$ に含まれる単語数を表す。$F_1(\boldsymbol{o}, \boldsymbol{w})$ と $F_2(\boldsymbol{o}, \boldsymbol{w})$ はそれぞれ音響確率と言語確率の対数値を表し、$F_3(\boldsymbol{o}, \boldsymbol{w})$ 単語挿入ペナルティーの対数値を表す。重みパラメータの定義として、λ_1 は 1 に固定し、λ_2 は言語モデルに対するスケーリング係数、λ_3 は単語挿入ペナルティーに対するスケーリング係数とする。これらは式 (2) ではそれぞれ κ_1 と κ_2 として表されている。ここで、式 (10) から式 (13) を式 (3) へ代入することで式 (2) が得られる。

音響モデルとして HMM、言語モデルとしてトライグラムモデルを用いた場合、これらの素性関数は次のように表される。

$$F_1(\boldsymbol{w}, \boldsymbol{o}) = \sum_{i=1}^{N} \log p_h(o_i|w_i) \tag{14}$$

$$F_2(\boldsymbol{w}, \boldsymbol{o}) = \sum_{i=1}^{N} \log p_n(w_i|w_{i-1}, w_{i-2}) \tag{15}$$

ここで、$p_h(o_i|w_i)$ は単語 w_i が得られたときに o_i が出現する確率を表し、o_i の出現確率は w_i にのみ依存すると仮定している。$p_n(w_i|w_{i-1}, w_{i-2})$ はトライグラムモデルであり、2 単語 w_{i-1}, w_{i-2} が得られたときの単語 w_i の出現確率である。

3.2 ワードラティスを用いた推定

式 (8) における期待値 $E_{\mathbf{\Lambda}^t}[F_k(\boldsymbol{w}, \boldsymbol{o})]$ は、理想的には単語列の組み合わせ全てを用いて計算される。しかし、それは実際の音声認識では不可能であるため、我々は認識によって得られるワードラティスに含まれる単語列を用いた計算方法を採用する [10]。ワードラティスは多くの単語列の組み合わせをコンパクトに表現できるため、N ベストリストを用いるよりも効率的に多くの単語列の仮説を扱うことが可能である。

図 1 にサンプルのワードラティスを挙げる。図 1 における太い点 h', h, i, j, f, f' はノードを表す。ここで、ワードラティスにおける始端と終端のノードをそれぞれ B,E と表す。単語 w_i^j はノード i,j 間のアークに相当する。o_i^j は、ノード h,f の間に相当する音響観測列とする。ここで、ノード h からノード f までの 3 単語連鎖を、ω_h^f と定義する。これらの定義を用いると、$E_{\mathbf{\Lambda}^t}[F_k(\boldsymbol{w}, \boldsymbol{o})]$ は次のように計算できる。

$$E_{\mathbf{\Lambda}^t}[F_k(\boldsymbol{w}, \boldsymbol{o})] = \sum_{\omega_h^f \in \Omega} \gamma_{\mathbf{\Lambda}^t}(\omega_h^f) f_k(o_i^j, \omega_h^f) \tag{16}$$

ここで、Ω はワードラティス含まれる全ての 3 単語連鎖である。$\gamma_{\mathbf{\Lambda}^t}(\omega_h^f)$ は単語連鎖 ω_h^f の占有確率である。関数 $f_k(o_i^j, \omega_h^f)$ は、次のように定義される。

$$f_1(o_i^j, \omega_h^f) = \log p(o_i^j|w_i^j)$$

$$f_2(o_i^j, \omega_h^f) = \log p(w_j^f|w_h^i, w_i^j)$$

$$f_3(o_i^j, \omega_h^f) = 1$$

ワードラティス上の単語の占有確率を計算するための方法が提案されている [8] [9]。これらの方法を用いて、図 1 上のノード h と f の間の単語連鎖の占有確率は、次のように表される。

$$\gamma_{\mathbf{\Lambda}^t}(\omega_h^f) = \frac{\alpha_{\mathbf{\Lambda}^t}^h \beta_{\mathbf{\Lambda}^t}^f \exp\left(\sum_{k=1}^{K} \lambda_k^t f_k(o_i^j, \omega_h^f)\right)}{Z_{\mathbf{\Lambda}^t}(\boldsymbol{o})} \tag{17}$$

ここで、$\alpha_{\mathbf{\Lambda}^t}^h$ はノード h における前向きスコアで、$\beta_{\mathbf{\Lambda}^t}^f$ はノード f における後ろ向きスコアである。前向きスコアは、ワードラティスの先頭から再帰的に計算される。ノード h の前向きスコアは次のようになる。

$$\alpha_{\mathbf{\Lambda}^t}^h = \sum_{h'} \sum_{h''} \alpha_{\mathbf{\Lambda}^t}^{h'} \exp\left(\sum_{k}^{K} \lambda_k^t f_k(o_h^i, \omega_{h''}^i)\right)$$

h'' はノード h' に接続されるアークの始端ノードを表す。この式は、ノード h に接続する全てのアークの前向きスコアを全て足し合わせることで前向きスコアが計算できることを表す。後ろ向きスコアは同様の手順でワードラティスの最後尾のノード E から計算される。前向きスコアと後ろ向きスコアを用いて、式 (17) 中の分配関数 $Z_{\mathbf{\Lambda}}(\boldsymbol{o})$ は計算することができ、次のように表される。

$$Z_{\mathbf{\Lambda}^t}(\boldsymbol{o}) = \alpha_{\mathbf{\Lambda}^t}^E = \beta_{\mathbf{\Lambda}^t}^B \tag{18}$$

$L(\Lambda)$ の勾配は式 (16) で得られた期待値を用いることで計算され、式 (5) に適用することでみパラメータは更新される。

4. 繰り返し過程の導入

前節で述べたように、ワードラティスには多くの異なる単語列が含まれており N ベストリストを用いるよりも頑健な推定が可能である。一方、それら異なり単語列の数は理想的な場合と比較して依然少ないため、スケーリング係数の推定値はワードラティスに含まれる単語列の内容によって変化する。また、ワードラティスは生成されるときに用いられたスケーリング係数の値によりその内容が変化する。すなわち、ワードラティスの内容とスケーリング係数の値には相互依存関係があることを意味する。この問題に対処するため、我々はワードラティスの生成とスケーリング係数の推定を交互に行う「繰り返し過程」を導入する。この繰り返し過程は、初期値による推定値の変動に対し頑健に推定できることが期待できる。アルゴリズムの概要を次に示す。

Step 0　スケーリング係数の値を任意の値に設定する。
Step 1　2 章で記述された式 (7) の $\mathbf{S}^k$ を計算する。
Step 2　学習データの全ての発声についてワードラティスを生成する。
Step 3　式 (5) と収束条件の式 (9) を用いてスケーリング係数を推定する。
Step 4　もしスケーリング係数が予め定められた条件に収束した場合、この過程を終える。
収束していない場合は Step2 へ戻る。

生成されたワードラティスに正解の単語列が含まれていない場合、Step3 における推定が破綻する。この問題を避けるため、正解の含まれていないワードラティスに正解単語列を統合する処理を行う。

5. 実　　験

5.1 実験条件

提案手法の有効性を評価するために、大語彙連続音声認識を用いて実験を行った。データベースとして、電話回線を通して収録された日本語対話音声を用いた。この音声データは全て人手によって書き起こされている。HMM の学習には用意された音声データから 208 時間の音声とその書き起こしを用いた。言語モデルの学習には、音響モデルの学習に用いた音声の書き起こしを用い、総単語数は 554k 単語だった。スケーリング係数の推定には、学習データとは別に 2 時間の音声データ (8766 発声) とその書き起こしを用いた。認識率評価には、4 時間の音声データ (28284 単語、40 名) の発声を用いていた。

分析条件はフレーム周期 10mscc、ウィンドウ幅 32mscc で行った。特徴量は、12 次元の MFCC とそれら 1 次と 2 次の導係数、1 次と 2 次のパワーの導係数、調波性特徴量とその 1 次導係数量の合計 40 次元を用いた [11]。HMM は状態数 3000 で、各状態の混合正規分布数を 32 とした。評価はワードラティスの生成とリスコアの 2 パス構成で行い、ワードラティスの生成には言語モデルとして語彙数 31k 単語、83k バイグラムを用いた。また、リスコアには 306k トライグラムの言語モデルを用いた。比較としてグリッドサーチにてスケーリング係数の最適化を行った。グリッドサーチは、言語モデル重み λ_2 と単語挿入ペナルティー重み λ_3 について、それぞれ 4 から 10、−20 から 10 の範囲で 150 点について認識を行った。そのときに得られた単語正解精度の最高値は 65.8%だった。

提案手法の繰り返し過程の初期に対する頑健性を確認のため、3 組の異なる初期値 $(\lambda_2, \lambda_3) = (1, 0), (10, -20), (4, 10)$ を用いてそれぞれ推定を行った。式 (9) における収束条件の閾値は 10^{-4} とした。繰り返し過程の回数 (Step2 と Step3 の回数) は、10 回とした。

5.2 実験結果

図 2 に提案手法を用いて推定の結果を示す。3 組の異なる初期値から推定された値はそれぞれ $(\lambda_2, \lambda_3) = (6.9, -8.3), (6.9, -9.5), (6.8, -9.4)$ だった。これらの値を用いて行った認識実験にて、認識率はいずれの場合においても 65.7%であった。これは、グリッドサーチにより得られた認識率よりも 0.1%程度低いものであった。これは十分な精度で推定ができていると考えられる。また、3 組の異なる初期値を用いて得られた推定値が、ほぼ同じ値になったことから、我々の提案手法が初期値の変化に頑健であることが確認できた。本実験において、2 つのモデルに対するスケーリング係数の推定を行ったが、本提案手法は更にモデルの数を増やした場合でも適用可能である。また、モデルを増やした場合、グリッドサーチではモデル数のべき乗のオーダーで探索数が増えていくが、提案手法は勾配法がベースとなっているため大幅な演算量の増加は無い。このことから、提案手法は従来法と比較して効率的な推定法であるといえる。

6. おわりに

今回、音声認識におけるスケーリング係数の推定方法を提案した。提案手法では、スケーリング係数を対数線形モデルの重みパラメータとみなし、その最適値を最大事後確率基準を用いて推定した。推定には、ワードグラフを用いた勾配法を用いた。勾配法で推定される値がワードグラフの内容に影響を受けることを考慮して、繰り返し過程を導入した結果、初期値に対して頑健に最適化ができることを確認した。実験の結果、我々の手法で得られた認識率と従来方法のグリッドサーチから得られた認識率の差は 0.1%であり、十分な精度で推定できることを確認した。将来の研究として、複数の言語モデルの組み合わせ、それらのスケーリング係数の最適化を行うことで、話題に合わせた言語モデルの選択などについて検討を行う予定である。

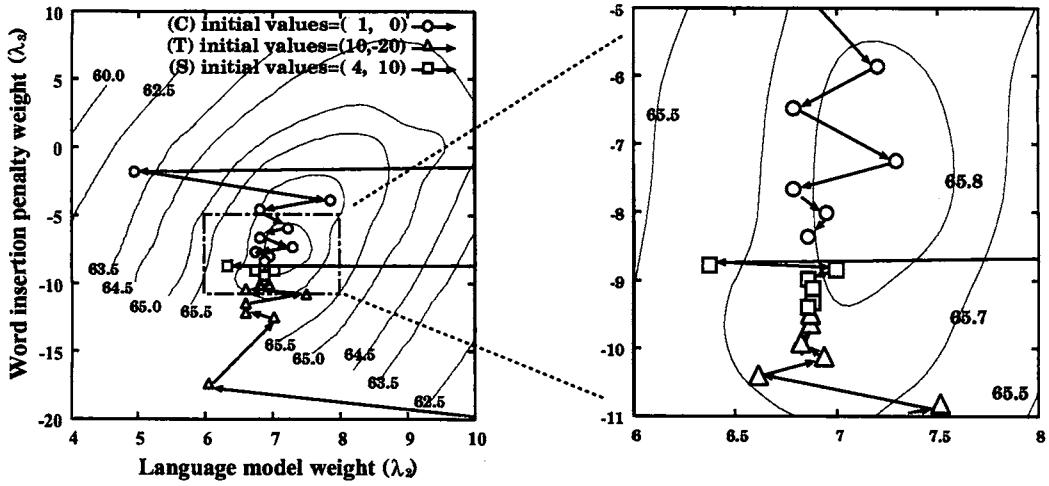

図 2　スケーリング係数の推定結果。それぞれの等高線はグリッドサーチにより得られた同じ認識率を表す。○、△、□は、それぞれ違った初期値を用いて推定されたスケーリング係数の値を示す。それぞれの矢印の先頭の点は矢印の始端の値を用い、勾配法により推定された値である。右の拡大図は、左図にて破線で囲まれた領域を表す。

文　献

[1] L. R. Bahl, R. Bakis, F. Jelinek, and et al, "Language-model/acoustic channel balance mechanism," IBM Technical Disclosure Bulletin, vol. 23(7B), pp.3464-3465, Dec. 1980.

[2] A. Ito, M. Kohda, and S. Makino, "Fast optimization of language model weight and insertion penalty from n-best candidates," Acoustical Science and Technology, Vol. 26(2005), Bo. 4, pp. 384-387, 2005.

[3] F. Josef Och and H. Ney, "Discriminative Training and Maximum Entropy Models for Statistical Machine Translation," Proc. of the 40th Annual Meeting of the Association for Computational Linguistics(ACL), pp. 295-302, July 2002.

[4] R. Zhang, G. Kikui, and H. Yamamoto, "Using Multiple Recognition Hypotheses to Improve Speech Translation," Proc. of the International Workshop on Spoken Language Translation 2005, pp. 40-46, 2005.

[5] P. Beyerlein, "Discriminative Model Combination," Proc. of ICASSP'98, pp. 481-484, 1998.

[6] S. Della Pietra, V. Della Pietra, and J. Lafferty, "Inducing features of random fields," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol 19, pp. 380-393, 1997.

[7] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and et al, "Numerical Recipes in C," Cambridge University Press, 1992.

[8] F. Wessel, R. Schlüter, K. Macherey, and et al, "Confidence Measures for Large Vocabulary Continuous Speech Recognition," IEEE Transactions on Speech and Audio Processing, Vol. 9, No. 3, pp. 288-298, March 2001.

[9] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," Proc. of 18th International Conference on Machine Learning pp. 282-289, 2001.

[10] S. Young and et al, "The HTK Book. Revised for HTK Version 3.2," http://htk.eng.cam.ac.uk/, Dec. 2002.

[11] K. Takagi and T. Watanabe, "Utilization of Spectral Harmonics Structure for Speech Recognition," Proc. of meeting of the acoustical society of Japan, pp. 3-4, September 1997.