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

2008－SLP－74 (44)
2008 / 12 / 10

単語誤り最小化に基づく識別的リスコアリングによる音声認識

小林 彰夫† 奥 貴裕† 本間 真一† 佐藤 庄衛†
今井 亨† 都木 徹†

† NHK 放送技術研究所
〒157-8510 東京都世田谷区砧 1-10-11
E-mail: †{kobayashi.a-fs,oku.t-le,homma.s-fc,sato.s-gu,imai.t-mq,takagi.t-fo}@nhk.or.jp

あらまし 本報告では，ニュース音声認識における単語誤りの傾向を反映したリスコアリング手法を提案する．提案法では，リスコアリングの際，音声認識の単語仮説の誤り傾向に応じて，仮説にペナルティを与える．単語仮説のペナルティは，言語的な文脈により活性化する素性関数とその重みにより定義される．素性関数の重みを求めるため，単語誤り最小化に基づく学習法を提案し，学習データ中の単語誤りを削減するような目的関数を用いて学習を行う．さらに，ニュース音声認識をターゲットとした時期依存適応学習を導入し，話題の時間的な関連性を用いて認識率の改善を図る．ニュース音声を用いたリスコアリング実験の結果，提案法は単語誤り率 7.4%となり，trigram によるラティスリスコアリングに比べて 6.3%の単語誤り削減率が得られた．

キーワード ラティスリスコアリング，識別学習，単語誤り最小化，言語モデル適応化

Discriminative Rescoring Based on Minimization of Word Errors for Speech Recognition

Akio KOBAYASHI†, Takahiro OKU†, Shinichi HOMMA†, Shoei SATO†,
Toru IMAI†, and Tohru TAKAGI†

† NHK Science & Technical Research Laboratories
1-10-11 Kinuta, Setagaya-ku, Tokyo, 157-8510 Japan
E-mail: †{kobayashi.a-fs,oku.t-le,homma.s-fc,sato.s-gu,imai.t-mq,takagi.t-fo}@nhk.or.jp

Abstract This paper describes a novel method of rescoring that reflects tendencies of errors in word hypotheses in speech recognition for transcribing broadcast news. The proposed rescoring assigns penalties to sentence hypotheses according to the recognition error tendencies in the training lattices themselves using a set of weighting factors for feature functions activated by a variety of linguistic contexts. We introduced new techniques to obtain the factors and it is based on the minimization of word errors, which explicitly reduces expected word errors. Moreover, we proposed a new time-dependent-adaptive training scheme, which features similarities among temporal correlated articles of broadcast news. The results of transcribing Japanese broadcast news achieved a word error rate (WER) of 7.4%, which was a 6.3% reduction relative to conventional trigram lattice rescoring.

Key words lattice rescoring, discriminative training, word error minimization, lanugage model adaptation

1. はじめに

コーパスに基づく音声言語処理は，数々のアプリケーションで成果を挙げている．例えば，大語彙音声認識システムはテレビ放送のクローズドキャプション (字幕) の作成のためにすでに利用されている [1]．このようなアプリケーションでは，統計的音響・言語モデルが重要な役割を果たしているが，いわゆる読み上げニュースを除けば，満足できるような認識率が得られているとは言いがたく，屋外など雑音環境下での発話や，対談などのややくだけた発話の認識率改善が求められている．

言語的な観点からみた場合，認識率劣化の原因は，実際の発話と言語モデル学習に用いられるコーパスとの統計的な分布の不一致が考えられる．例えば，我々の用いている学習コーパスの大部分は，いわゆる書きことばで占められている．一方，現

実の発話にみられるような口語的な表現は，コーパス内では低頻度であり，認識誤りを起こしやすい．音声認識結果から収集した単語仮説の誤り傾向が利用できるのであれば，誤りの可能性の高い単語仮説にペナルティを，そして誤りの可能性の低い仮説に報償を与えることで，認識率の改善が期待できる．

単語仮説の誤り傾向は，学習データ中における，仮説の正解/誤りのパターンの分布である．したがって，仮説の正誤のパターンを識別的に学習し，誤り傾向を反映したペナルティを得れば，認識率を改善できると考えられる．従来，単語仮説の誤り傾向を識別的に学習する方法として，文献[2]〜[7]などが行われてきた．これらの手法は，仮説の誤り傾向の学習に関して，正解単語列のスコアを最大化するという手法で共通しており，いずれも単語誤り率の削減に一定の効果がある．しかし，単語誤りを大きく削減するには，正解単語列のスコアを最大化する基準を用いて誤り傾向を学習するのではなく，単語誤りの数を直接的に削減するような基準を用いて学習する方が効果的ではないかと考えられる．

そこで，本稿では，単語誤り最小化に基づく識別的リスコアリング手法の提案を行う．提案法は，学習ラティスの単語誤りの期待値を最小化する基準を用いて，単語仮説の誤り傾向を学習する．仮説の誤り傾向は，単語 n-gram などの言語的な文脈により活性化する素性関数とその重みにより表現され，リスコアリング時にペナルティとして利用される．

また，本研究では，ニュース音声認識をターゲットとしていることから，ニュースを対象とした識別的リスコアリングの学習方法として，時期依存適応による学習法を提案する．

これまで，言語モデルの適応化では，ニュースの話題の時間的な関連性を利用している[8]．時間的な関連性とは，例えば，夜10時のニュースの話題は，直近の夜7時のニュースの話題と関連があり，共通の単語や句の出現が期待されるということである．字幕制作のための言語モデルでは，このような時間的な話題の関連性を利用した適応化により，単語誤り率を削減している．

一方，時間的関連性にしたがって共通の単語や句が出現するのであれば，両者の誤りの傾向も類似するのではないかと考えられる．夜7時のニュースの音声認識結果から得られた認識誤りの傾向を利用すれば，夜10時のニュースの認識誤りを削減できるのではないかと考えられる．また，話題の時間的関連性を利用して単語誤りを削減できるのであれば，字幕制作システム[9]への活用も可能である．文献[9]のシステムでは，音声認識を用いて字幕を制作しており，音声認識結果と正解単語列(字幕)が得られる．これらを用いて，直前に放送されたニュースの誤り傾向を学習すれば，認識誤りの削減が期待できる．

そこで，単語誤り最小化に基づく素性関数の重みの学習法に，話題の時間的関連性を利用した時期依存適応化学習方法を導入し，単語誤り率の削減について検討を行うこととする．

2. 単語誤り最小化に基づく識別的リスコアリング

2.1 識別的リスコアリング

音声認識の第1パスの出力として，ラティス $\mathcal{L}$ が得られたとする．第2パスのラティスリスコアリングでは，対数音響スコア f_0，対数言語スコア f_1，それぞれの重みを λ_0, λ_1 として，

$$g(\boldsymbol{w}|\boldsymbol{x}) = \lambda_0 f_0(\boldsymbol{x}|\boldsymbol{w}) + \lambda_1 f_1(\boldsymbol{w}) \qquad (1)$$

を入力音声 $\boldsymbol{x}$, 文仮説 $\boldsymbol{w}$ に対する識別関数とする．最良仮説 $\boldsymbol{w}^*$ は，

$$\boldsymbol{w}^* = \arg\max_{\boldsymbol{w}\in\mathcal{L}} g(\boldsymbol{w}|\boldsymbol{x}) \qquad (2)$$

により求める．

ここで，単語仮説の誤りを反映したリスコアリングができるように，式(1)を書き直し，

$$\hat{g}(\boldsymbol{w}|\boldsymbol{x}) = \lambda_0 f_0(\boldsymbol{x}|\boldsymbol{w}) + \lambda_1 f_1(\boldsymbol{w}) + \sum_{i=2}^{I} \lambda_i f_i(\boldsymbol{w}) \qquad (3)$$

とする．上式の f_i $(i = 2, \ldots, I)$ は素性関数，λ_i は該当する素性重みである．識別的リスコアリングでは，単語仮説の誤り傾向を反映するようなペナルティを素性関数とその重みで表現して，認識率の改善を行う．

2.2 素性関数

式(3)の素性関数は，文仮説の文脈(単語や単語クラスのn項組)により活性化する関数である．例えば，単語trigram素性は

$$f_{i=h_1(u_1,u_2,u_3)}(\boldsymbol{w}) = c_{u_1,u_2,u_3}(\boldsymbol{w}) \qquad (4)$$

のように，文仮説 $\boldsymbol{w}$ に含まれる単語3項組 (u_1, u_2, u_3) の数 c_{u_1,u_2,u_3} を返す関数として定められる．ただし，$h_1(u_1, u_2, u_3)$ は単語3項組に対して，該当する素性関数の番号を返すハッシュ関数である．

単語n-gram素性では，学習データのスパースネスが問題となる．そこで，スパースネスを回避するために，単語クラス素性を導入する．単語クラスn-gram素性は，言語モデルの学習コーパスを用いたクラスタリング結果を用いて定義する．ただし，クラスタリングアルゴリズムは文献[10]に従い，本稿の実験で用いる60kの語彙を1,000クラスに分類し，クラスタリング結果に基づいてクラスbigram，trigram素性を構成した．

クラス素性と同様，意味クラス素性を導入する．本稿の実験では語彙サイズ60kの単語エントリに対して，分類語彙表[11]にしたがって27kのエントリに838のクラスを割り当て，意味クラスbigram，trigram素性を構成した．

なお，単語クラス素性，意味クラス素性は，単語からクラスへマッピングする関数を $s(\cdot)$ として，

$$f_{i=h_2(s(u_1),s(u_2),u_3)}(\boldsymbol{w}) = c_{s(u_1),s(u_2),u_3}(\boldsymbol{w}) \qquad (5)$$

のように定義する．h_2 は h_1 と同様のハッシュ関数である．

2.3 単語誤り最小化学習

本節では，式 (3) のパラメータ $\Lambda = \{\lambda_2, \ldots, \lambda_I\}$ を求めるために，単語誤り最小化に基づく学習方法を提案する．

発話 $\boldsymbol{x}_m$ $(m = 1, \ldots, M)$ に対し，ラティス $\mathcal{L}_m$ および正解単語列 $\boldsymbol{w}_{m,0}$ が学習データ D として与えられたとする．ただし，正解単語列 $\boldsymbol{w}_{m,0}$ はラティス $\mathcal{L}_m$ に重畳されている．ラティス $\mathcal{L}_m$ の n 番目の文仮説を $\boldsymbol{w}_{m,n}$ $(n = 1, \ldots, N)$ として，素性関数の重みを求める目的関数を次のように定める．

$$\mathbf{L}(\Lambda) = \sum_{m=1}^{M} \frac{\sum_{\boldsymbol{w}_{m,n} \in \mathcal{L}_m} \{Acc(\boldsymbol{w}_{m,n}) \exp(\hat{g}(\boldsymbol{w}_{m,n}|\boldsymbol{x}_m))\}}{\sum_{\boldsymbol{w}_{m,k} \in \mathcal{L}_m} \exp(\hat{g}(\boldsymbol{w}_{m,k}|\boldsymbol{x}_m))} \tag{6}$$

上式で，$\hat{g}$ は式 (3) の識別関数，$Acc(\boldsymbol{w}_{m,n})$ は正解単語数 (挿入誤りを引いた正解単語の数) であり，目的関数 $\mathbf{L}(\Lambda)$ は，学習ラティスの正解単語数の期待値として定められる．ラティス中のすべての仮説について，正解数の期待値を求めるのは計算コストがかかるため，Povey [12] の文献にならって，ラティス $\mathcal{L}_m$ の 2 つのノード t', t を結ぶエッジ $e_{t't}$ $(t' < t)$ について，正解数 $acc(e_{t't})$ $(-1 \leqq acc(e_{t't}) \leqq 1)$ を以下のように近似する．

$$acc(e_{t't}) = \begin{cases} -1 + 2l & \text{if same word} \\ -1 + l & \text{if different word} \end{cases} \tag{7}$$

ここで，l は $e_{t't}$ 上の単語仮説と，正解単語列 $\boldsymbol{w}_{m,0}$ の該当する単語との間でオーバーラップするフレーム数の比で定義される．学習ラティス $\mathcal{L}_m$ の正解数の期待値は，ラティスの各ノードをトポロジカルに巡回することで求められる．例えば、ラティス $\mathcal{L}_m$ の始端からノード t に至る前向きの正解数の期待値を $\alpha_{acc}(t)$ とすると，

$$\alpha_{acc}(t) = \frac{\sum_{t':e_{t't} \in \mathcal{L}_m} \{\alpha_{acc}(t') + acc(e_{t't})\} \alpha(t') \times s(e_{t't})}{\sum_{t':e_{t't} \in \mathcal{L}_m} \alpha(t') \times s(e_{t't})} \tag{8}$$

である．ただし，$s(e_{t't})$ は

$$s(e_{t't}) = \exp\{\lambda_0 f_0(e_{t't}) + \lambda_1 f_1(e_{t't}) + \sum_{i=2}^{I} f_i(e_{t't})\} \tag{9}$$

で得られるエッジ $e_{t't}$ におけるスコアを表す．$\mathcal{L}_m$ の終端ノード T_m における正解数の期待値をあらためて $\alpha_{acc}(T_m)$ とおけば，式 (6) は，

$$\mathbf{L}(\Lambda) \approx \sum_{m=1}^{M} \alpha_{acc}(T_m) \tag{10}$$

となる．

重み Λ は，準ニュートン法の 1 つである L-BFGS アルゴリズム [13] により求めることができる．L-BFGS アルゴリズムでは，目的関数の 1 階偏微分が必要となるが，ここでは以下の近似を行う．

学習ラティス $\mathcal{L}_m$ 上のエッジ $e_{t't}$ に対して，ノード t' での前向きの正解単語数の期待値を $\alpha_{acc}(t')$，ラティスの終端からノード t に至る後ろ向きの期待値を $\beta_{acc}(t)$ とすれば，エッジ $e_{t't}$ を通る全仮説の期待値 $\gamma(e_{t't})$ は，

$$\gamma(e_{t't}) = \alpha_{acc}(t') + acc(e_{t't}) + \beta_{acc}(t) \tag{11}$$

である．これを用いれば，目的関数の重み λ_i に関する 1 階偏微分は次のようになる．

$$\frac{\partial \mathbf{L}}{\partial \lambda_i} = \sum_{m=1}^{M} \sum_{e_{t't} \in \mathcal{L}_m} \{\gamma(e_{t't}) - \tilde{\gamma}\} p(e_{t't}) \phi_i(e_{t't}) \tag{12}$$

ただし，$\tilde{\gamma} = \alpha_{acc}(T_m)$，$p(e_{t't})$ はエッジ $e_{t't}$ の事後確率，$\phi_i(e_{t't})$ は，$e_{t't}$ で素性 λ_i が活性化する場合に 1 を返す 2 値関数とする．

単語誤り最小化に基づく素性重みの学習手法は，音響モデルの識別学習 [12], [14] と同様であるが，音響モデルではガウス分布のパラメータを求めるのに対し，本手法では対数線形モデルのパラメータ (式 (3) の λ_i) を求める点が異なる．

2.4 誤り傾向の時期依存適応学習

文献 [8] に示されているように，ニュースには時期依存性がある．例えば，きょうのニュースで使われた単語や句は，明日以降のニュースでも使われやすく，話題どうしに関連性がある．類似した発話どうしでは，同じ単語や文脈で認識誤りが起こりうる．したがって，単語の誤り傾向にも関連性があるのではないかと考えられる．誤り傾向に関連性があれば，過去のニュースから得られた誤り傾向を用いて，将来のニュースにおける認識誤りの削減が期待できる．

そこで，時間的に近接した直近ニュースと評価データとの間に話題関連性があると仮定して，話題関連性を識別的リスコアリングに反映させる手法を提案する．

まず，学習データ D を話題依存性のある直近ニュースの集合 D_2 とそれ以外の長期間ニュース D_1 に分け，素性関数の重みを求めるための損失関数を以下のように再定義する．

$$\mathbf{L}(\Lambda; D) = \mathbf{L}(\Lambda; D_1) + \kappa \mathbf{L}(\Lambda; D_2) \tag{13}$$

κ は，$\mathbf{L}(\Lambda; D_2)$ に対する重みとし，事前に定めた実数値とする．式 (13) に基づいて素性関数の重みを学習した場合，特に重み $\kappa > 1$ であれば，話題関連性があるとみなされた学習データ D_2 の素性関数の重みは，D_1 に含まれる素性に比べて学習されやすくなる．結果として，D_2 に含まれる素性の重みの絶対値は，D_1 の素性に比べて相対的に大きくなる．評価データの誤り傾向が D_2 に近ければ，D_2 に含まれる素性関数の重みによって，単語誤り率の削減が期待できる．

本稿では，評価データの直前のニュース番組を適応学習データ D_2 として，時期依存適応学習を行う．

2.5 従 来 法

本稿で提案する手法に対する従来法として，対数回帰に基づく手法 [2], [7] について述べる．

ラティス $\mathcal{L}_m$，正解単語列 $\boldsymbol{w}_{m,0}$ に対して，$q_\Lambda(\boldsymbol{w}_{m,0}|\boldsymbol{x}_m)$ を

$$q_\Lambda(\boldsymbol{w}_{m,0}|\boldsymbol{x}_m) = \frac{\exp \sum_i \lambda_i f_i(\boldsymbol{w}_{m,0})}{\sum_{\boldsymbol{w}_{m,n} \in \mathcal{L}_m} \exp \sum_j \lambda_j f_j(\boldsymbol{w}_{m,n})} \tag{14}$$

のように定める．素性関数の重みを求める目的関数は，正解単語列の対数尤度を最大化する目的関数

表 1　評価データ

	発話数	単語数	パープレキシティ	未知語率 (%)
DevSet	1194	18.5k	16.3	0.06
EvalSet	1096	18.3k	31.8	0.08

表 2　単語誤り最小化学習/時期依存適応学習に用いたデータ

	発話数	単語数	パープレキシティ	未知語率 (%)
学習データ (長期間)	53.1k	1.09M	49.0	0.78
適応学習データ (DevSet)	921	14.1k	20.1	0.17
適応学習データ (EvalSet)	845	13.5k	29.0	0.07

表 3　実験結果 (%)

		DevSet				EvalSet			
		%WER	%Sub	%Del	%Ins	%WER	%Sub	%Del	%Ins
baseline		6.6	3.7	1.1	1.9	7.9	4.9	1.4	1.7
従来法	$\kappa = 1.0$	6.5	3.6	1.2	1.8	7.9	4.9	1.4	1.5
	$\kappa = 70.0$	6.5	3.6	1.1	1.7	7.8	4.8	1.5	1.5
提案法	$\kappa = 1.0$	6.4	3.5	1.2	1.7	7.5	4.8	1.4	1.3
	$\kappa = 30.0$	6.3	3.5	1.1	1.7	7.4	4.7	1.4	1.3

$$\mathbf{L}_c(\Lambda) = \sum_{m=1}^{M} \log q_\Lambda(\boldsymbol{w}_{m,0}|\boldsymbol{x}_m) \qquad (15)$$

として定義される．式 (15) に基づいて，正解単語列のスコアを最大化する学習方法は，条件付き確率場 (Conditional Random Fields) [15] に類似した手法である．目的関数の最大化には，提案法同様 L-BFGS アルゴリズムが使える．ただし，上式の 1 階偏微分は，

$$\frac{\partial \mathbf{L}_c}{\partial \lambda_i} = \sum_{m=1}^{M} \{f_i(\boldsymbol{w}_{m,0}) - \sum_{e_{t't} \in \mathcal{L}_m} f_i(e_{t't}) p(e_{t't})\} \qquad (16)$$

である．

3. 実　　験

3.1 実験条件

評価データは，DevSet(開発データ) と EvalSet(評価データ) の 2 種類を，2004 年 7 月のニュース番組から，それぞれ 7 番組ずづ選んだ．DevSet は，単語誤り最小化学習の繰り返し回数と，時期依存適応学習のパラメータ (式 (13) の κ) を定めるために用いた (表 1)．

誤り最小化学習のための学習データは，2003 年 1 月 5 日から 2004 年 6 月 30 日までの放送ニュースを学習データ (長期間) として選んだ．また，DevSet,EvalSet それぞれについて，時期依存適応学習のために，評価データの直前に放送されたニュースを適応学習データとして選択した (表 2)．

識別的スコアリングの素性関数の重みは，提案法は，式 (6) に基づく目的関数を用いて，また，従来法は式 (15) に基づく目的関数を用いて学習した．L-BFGS に基づく繰り返し学習回数は 40 回を上限とした．

素性関数は単語 bigram,trigram 素性，単語クラス bigram,trigram 素性および意味クラス bigram,trigram 素性を用いた．素性関数は，文献 [16] で示される手法に従って，学習データ (長期間) から，提案法で 534.0k 個，従来法で 165.3k 個を選択した．これらの素性関数に，時期依存適応学習で用いる適応学習データに含まれる素性関数を全て加え，素性関数の重みを学習した．

音声認識は第 1 パスで tree lexicon および bigram でデコードしたのち，trigram ラティスに展開する．第 2 パスはラティス上を識別的リスコアリングによりリスコアリングし仮説を選択する．

言語モデルの語彙サイズは 60k 単語とし，150 万文 (42.9M 単語) のニュース原稿，書き起こしから学習した．cut off は bigram に対して 1，trigram に対して 4，スムージングは Good-Turing とした．また，言語モデルは評価データに合わせて，時期依存言語モデルによる適応化を行った [8]．

音響モデルは，118 時間の放送ニュース音声から RASTA [17] を行ったうえで，39 次元 (12 次元 MFCC+対数パワーおよび 1 次・2 次の回帰係数) のパラメータを求めて triphone HMM を学習した．

3.2 実験結果

表 3 にリスコアリング結果を示す．ただし，表の baseline は，trigram ラティスリスコアリングの結果である．DevSet の結果は，提案法では素性重みの繰り返し学習の回数 34 回の点，従来法は 32 回の点を記載している．EvalSet の結果は，DevSet で得られた繰り返し回数および時期依存学習のパラメータを用いたときの単語誤り率である．

表 3 をみると，EvalSet 評価時に，時期依存適応学習を行った提案法 ($\kappa = 30.0$) で，単語誤り率 7.4%となり，baseline, 従来法を含めたすべての手法の中で，単語誤り率が最小となった．提案法の単語誤り率を baseline と比較すると，誤り削減率は 6.3%となり，時期依存適応学習なしの場合 ($\kappa = 1.0$) に対し，単語誤り削減率が 1.4%となった．

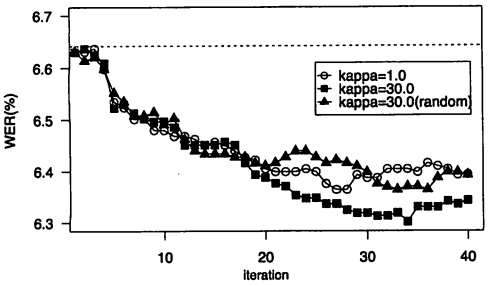

図 1　繰り返し学習と単語誤り率 (提案法)

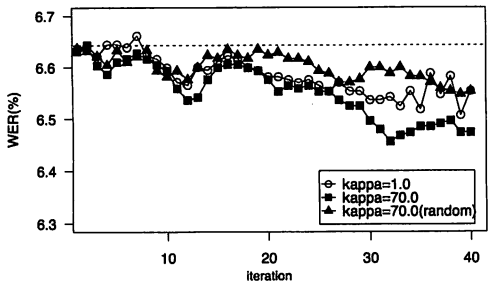

図 2　繰り返し学習と単語誤り率 (従来法)

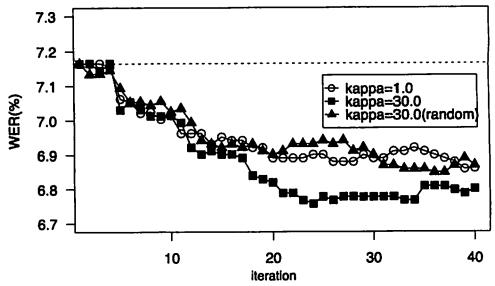

図 3　繰り返し学習と単語誤り率 (提案法, 関連する話題)

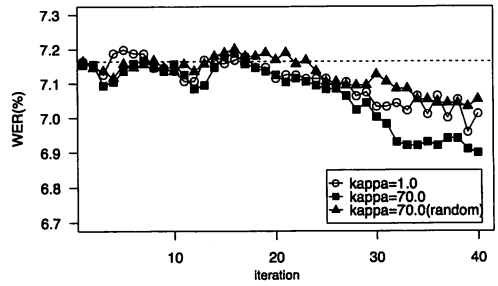

図 4　繰り返し学習と単語誤り率 (従来法, 関連する話題)

4. 考　　察

4.1 学習方法の比較

まず, 提案法と従来法を比較する. 図 1,2 は, 提案法と従来法のそれぞれについて, 素性重みの学習の繰り返し回数と DevSet の単語誤り率の関係を示したものである.

時期依存適応学習なし ($\kappa = 1.0$) のケースで比較すると, 提案法の方が, 繰り返し回数の増加に対する単語誤り率の削減が大きいことがわかる. この理由は, 提案法と従来法の学習方法の違いに起因する. 従来法では, 学習時に正解単語列のスコアを最大化するように素性関数の重みを学習するが, 学習データ中の正解単語列のみのスコアを最大化するため, 特定の素性関数の重みが過大もしくは過小になりやすく, 学習データに強く依存した重みが得られてしまうのではないかと考えられる.

一方, 提案法は, 単語誤りを明示的に反映する学習方法で素性の重みを得るため, 誤った単語仮説で活性化する素性関数に対しても, 正解数の期待値に応じた重みが与えられる. このため, 従来法のように, 素性関数の重みが過大または過小に学習されることがなく, 評価データに対して単語誤り率を安定して削減するような重みが得られたのではないかと考えられる.

4.2 時期依存適応学習の効果

提案法, 従来法の時期依存適応学習の効果について述べる.

図 1, 2 に, DevSet について, 時期依存適応学習の重み $\kappa = 30.0$(提案法) としたときの単語誤り率, $\kappa = 70.0$(従来法) としたときの単語誤り率を示す. また, 適応学習データと同数の発話を学習データからランダムに選び, 時期依存適応学習して評価した単語誤り率を同様に示す. 図 1 の提案法による時期依存適応学習の結果では, $\kappa = 30.0$ のとき, 素性重みの学習の繰り返し回数が 34 回の点で単語誤り率が最小となっている. 図 2 の従来法では, 繰り返し回数 32 回の点で単語誤り率が最小となっている. 一方, 提案法, 従来法のどちらも, 発話をランダムに選択して時期依存適応学習した場合, 繰り返し回数が大きくなっても, 単語誤り率の変化は, 時期依存適応学習なし ($\kappa = 1.0$) のケースとほとんど変わらない. したがって, 話題の関連性のある適応学習データを用いた時期依存適応学習は効果があるといえる.

DevSet の 1,194 発話のうち 697 発話 (58.4%) は, 適応学習データに含まれる話題と関連があった. 話題に関連する発話について, 図 1,2 と同様に単語誤り率と繰り返し回数をプロットしたものを図 3,4 に示す. 提案法の結果をみると, $\kappa = 30.0$ のとき, 素性重みの学習の繰り返し回数が 23 回の点で単語誤り率が最小となった. この点での時期依存適応学習なし ($\kappa = 1.0$) の結果に対する単語誤り削減率は 2.1%となった. 学習の繰り返し回数が 40 回に近づくにつれ, 単語誤り率が大きくなっていき, 過学習の様相を示す.

識別的リスコアリングの学習では, 学習データの音響・言語スコアの両方を使って素性重みを求める. したがって, 時期依存適応学習データと, 評価データに含まれる話題が類似してい

たとしても，データ間の話者の違いにより誤り削減の効果が少なくなることが考えられる．実際，時期依存適応学習に用いたデータと，評価データでは話者の重なりはない．話者が同一で，類似した話題 (例えば連続する数日間のニュース) を用いて時期依存適応学習した場合には，より大きな単語誤りの削減効果が得られる可能性がある．話者依存性も考慮した時期依存適応学習に関しては，今後の検討課題としたい．

4.3 誤り削減の効果

最後に，時期依存適応学習がどのような誤りを削減したのかについて述べる．EvalSet について，提案法の $\kappa = 1.0$ と $\kappa = 30.0$ で誤りの内訳を比較したところ，削減された誤りは，助詞「が」「て」「で」や動詞「いる」，接続詞などの機能語であった．例えば，

(誤, $\kappa = 1.0$)　二/か月/足らず/白紙/撤回/した

(正, $\kappa = 30.0$)　二/か月/足らず/**で**/白紙/撤回/した

あるいは，

(誤, $\kappa = 1.0$)　年金/支給/で/**決まっ**/**た**/ミス/**で**

(正, $\kappa = 30.0$)　年金/支給/で/**また**/ミス/**が**

のような誤りが削減された．時期依存適応学習により，誤りに含まれる人名や地名などの固有名詞の誤りも削減されることが期待されたが，誤りは削減されなかった．機能語や頻度の高い一般的な名詞を含む文脈は，適応学習データのみではなく，長期間の学習データにも高い頻度で含まれている．その一方で，固有名詞から成る文脈は，長期間の学習データでの頻度が低い．したがって，時期依存適応学習を行って，素性関数の出現頻度に重みをかけたとしても，他の素性関数に比べて目的関数 (正解数の期待値) を大きく増加させるような重みが得られにくいのではないかと考えられる．

ニュースをターゲットとした音声認識では，言語モデルの適応化が固有名詞の誤りを削減する一方，識別的リスコアリングおよび時期依存適応学習は，言語モデルの適応化では救いきれない単語列の認識誤りを削減する効果があると考えられる．

5. おわりに

ニュース音声認識を対象に，単語誤り最小化に基づく識別的リスコアリング手法と，時期依存学習法について提案した．提案法の特徴は，音声認識の単語仮説の誤り傾向を，学習ラティスの単語誤りが最小となる基準にしたがって学習することと，話題の時間的な関連性に基づく時期依存適応学習を行うことである．この２つの学習によって，評価データに適応して単語誤りを削減するような識別的リスコアリングが可能になる．提案法によるリスコアリングの結果，ニュース音声に対して単語誤り率 7.4%となり，trigram によるラティスリスコアリングに比べて 6.3%の単語誤り削減率となった．

本稿の識別的リスコアリングで利用した素性関数は，n-gram のような文脈に基づいているため，単語仮説の長距離の依存関係をとらえているわけではない．今後は，構文に関する情報などを用いて，仮説間の長距離の依存関係による識別的リスコアリング手法の検討を行う予定である．

文　献

[1] T. Imai, A. Kobayashi, S. Sato, S. Homma, K. Onoe, and T. Kobayakawa, "Speech recognition for subtitling Japanese live broadcasts," Proc. ICA, vol. I, pp. 165–168, 2004.

[2] B. Roark, M. Saraclar, and M. Collins, "Discriminative n-gram language modeling," Computer Speech and Language, vol. 21, pp. 373–392, 2007.

[3] M. Collins, "Discriminative reranking for natural language parsing," Proc. ICML, pp. 175–182, 2000.

[4] M. Collins, "Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms," Proc. EMNLP, vol. 10, pp. 1–8, 2002.

[5] Z. Zhou, J. Gao, F. Soong, and H. Meng, "A comparative study of discriminative methods for reranking LVCSR using n-best hypotheses in domain adaptation and generalization," Proc. ICASSP, vol. I, pp. 141–144, 2006.

[6] B. Roark, M. Saraclar, and M. Collins, "Corrective language modeling for large vocabulary ASR with the perceptron algorithm," Proc. ICASSP, pp. 789–792, 2004.

[7] E. Arisoy, B. Roark, I. Shafran, and M Saraçlar, "Discriminative N-gram Language Modeling for Turkish," Proc. Interspeech 2008, pp. 825–828, 2008.

[8] 小林彰夫, 今井亨, 安藤彰男, 中林克己, "ニュース音声認識のための時期依存言語モデル," 情報学会論文誌,vo.40, no.4, pp. 1421–1429, 1999.

[9] 本間真一, 尾上和穂, 小林彰夫, 佐藤庄衛, 今井亨, 都木徹, "ダイレクト方式とリスピーク方式の音声認識を併用したニュース字幕制作の検討," 秋季音響学会講演論文集, pp. 283–284, 2007.

[10] H. Ney, U. Essen, and R. Kneser, "On structuring probabilistic dependancies in stochastic language modeling," Computer Speech and Language, vol. 8, no. 1, pp. 1–38, 1994.

[11] 独立行政法人国立国語研究所, 分類語彙表, 大日本図書, 2004.

[12] D. Povey, and P. C. Woodland, "Minimum phone error and I-smoothing for improved discriminative training," Proc. ICASSP, vol. I, pp. 105–108, 2002.

[13] D. Liu, and J. Nocedal, "On the limited memory BFGS method for large scale optimization," Math. Programming, vol. 45, no. 3, pp. 503–528, 1989.

[14] W. Macherey, L. Haferkamp, R. Schlüter, and H. Ney, "Investigations on error mimizing training criteria for discriminative training in automatic speech recognition," Proc. Interspeech, pp. 2133–2136, 2005.

[15] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," Proc. ICML, pp. 282–289, 2001.

[16] A. Kobayashi, T. Oku, S. Homma, S. Sato, T. Imai, and T. Takagi, "Discriminative Rescoring Based on Minimization of Word Errors for Transcribing Broadcast News," Proc Interspeech 2008, pp. 1574–1577, 2008.

[17] H. Hermansky, and H. Morgan, "RASTA processing of speech," IEEE Trans. Speech and Audio, vol. 2, pp. 587–589, 1994.