

識別的誤り訂正学習における対立単語列と素性の選定

大庭 隆伸[†] 堀 貴明[†] 中村 篤[†]

[†] 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
〒 619-0237 京都府相楽郡精華町光台 2-4
E-mail: †{oba, hori, ats}@cslab.kecl.ntt.co.jp

あらまし 本稿では音声認識の誤り訂正モデルに着目する。このモデルは音声認識器の出力する複数の認識結果候補から、それらの言語的素性に基づいて、より誤りの少ない単語列を選択するために用いられる。一般に、言語的素性に単語 N-gram を採用し、それにより表現された正解パターンと誤りパターンの識別学習を通してモデルのパラメータ推定が行われる。しかし従来の学習法では、認識スコアの高い一位候補 (もしくは上位候補) を対立単語列とするため、下位候補に出現する多様かつ潜在的な誤りパターンをモデルに反映できず、また単語 N-gram 素性のデータスパース性により、学習データに出現しない多くのパターンをカバーできない等の問題があった。そこで本稿では、日本語話し言葉コーパスを用いた実験により、WER 基準での対立単語列選定、および素性分類別の誤り訂正モデル学習を行い、効果を検証する。それにより WER の高い単語列との識別がモデル性能に支配的であること、品詞 N-gram 等のスパース性を軽減可能な素性を採用することで、評価データの違いへの頑健性が向上することを示す。
キーワード 誤り訂正, 識別学習, 単語誤り率, 対立単語列, 素性選定

The Effect of Competitor and Feature Selection in Discriminative Training of Error Corrective Models

Takanobu OBA[†], Takaaki HORI[†], and Atsushi NAKAMURA[†]

[†] NTT Communication Science Laboratories, NTT Corporation.
2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0237 Japan
E-mail: †{oba, hori, ats}@cslab.kecl.ntt.co.jp

Abstract We focus on error corrective models for ASR, which select a more accurate word sequence among multiple word sequences produced by a speech recognizer. In such approaches, the corrective model is usually trained discriminatively using hypothesis/reference pairs, typically using word N-gram features to improve discrimination of the reference word sequence from hypothesis word sequences with high recognition scores. However, it is also important for error correction to consider various error patterns. This can be efficiently achieved through the use of hypothesis word sequences with high word error rates (WERs). The various error patterns can be expressed using features that alleviate the data sparseness problem. In this paper, we evaluate the impact of training using competitors with various WERs, as well as the impact of different features, on the corrective model's performance. Our experiments using the Corpus of Spontaneous Japanese show that an accurate and compact model can be generated via training to discriminate the reference from the single worst competing word sequence (with the highest WER), and that models using features that alleviate the data sparseness problem, such as Part-of-Speech N-gram features, achieve robust error correction on evaluation sets.

Key words error correction, discriminative training, word error rate, competitor, feature selection

1. はじめに

現在の最も一般的な音声認識技術では、入力音声に対する各単語列仮説の妥当性を音響モデルおよび言語モデルを用いて評

価し、最も妥当な単語列を認識結果として選択する。これに対し、他の知識源を利用し、より高精度に認識を行う手法が数多く提案されている。その一つに音声認識の誤り傾向を知識源とする誤り訂正アプローチがある。この誤り訂正は、音声認識器

の出力する複数の単語列仮説を、事前に誤り傾向を学習したモデルで再評価することにより実現される [1]- [7].

通常、誤り訂正モデルは単語列仮説から抽出される言語的な素性ベクトルを利用し、それにより表現される誤りパターンと正解パターンの組を用いて識別的に学習される。これに関し、従来は文字通り誤りを訂正するため、音声認識結果 (一位候補もしくは上位候補) に含まれる誤りパターンの訂正を重視した学習が行われてきた。すなわち、音声認識スコア順で上位に位置する単語列を対立単語列とした識別学習が行われてきた [1], [2], [5].

しかし、誤り訂正の精度向上には、より多様な誤りパターンの獲得を考慮した学習が有効であると考えられる。そのための手法として、単純に、正解単語列を下位候補を含めた全単語列から識別するように学習を行う手法 [3], [4] が考えられる。しかしながら、音声認識器の出力する複数の単語列仮説において、単語誤り率 (WER) の高い単語列に含まれる誤りが、WER の低い単語列に含まれる誤りの多くを内包する性質を考えれば、このような手法では、各単語列から同様の誤りパターンを表現した素性、つまり効果の重複する素性が数多く抽出され、不必要にサイズの大きなモデルが生成されてしまう。

これに対し、WER の高い単語列を重視した識別学習では、より WER の低い他の単語列の識別を兼ねた学習が可能であり、その上で多様な誤りパターンの獲得が実現されるものと期待できる。結果として、高精度でかつコンパクトな誤り訂正モデルを生成可能であると考えられる。

一方、誤り訂正に使用する素性として、これまで主に単語 N-gram が用いられてきた。しかし、多様な誤りパターンの獲得という意味では、単語 N-gram 素性のみの使用では必ずしも十分と言えない。学習データ中に出現した誤認識を含む単語 N-gram パターンは、全素性空間において非常にスパースに分布する。そのため、未知の様々な音声データに出現しうる誤りパターンを広くカバーすることは難しく、データスパース性の軽減が重要となる。そこで本稿では、文字 N-gram、品詞 N-gram 等のより抽象化された、次元数の小さい素性を使用し、その効果を検証する。このような素性は単語 N-gram に比べて密に分布するため、限られたサンプルで広範な領域をカバーできる。そのため学習と評価データ間の言語的差異が大きい場合でも、高い効果が期待できる。

本稿では日本語話し言葉コーパス (CSJ) [8] を用い、WER 基準の対立単語列選定、および素性分類の誤り訂正モデルの性能を実験的に評価する。まず誤り訂正モデルの識別的学習において、正解単語列と WER の高い単語列の識別が、モデルの高精度化に支配的に作用することを示す。また、対立単語列を WER の高い単語列のみに限定することで、コンパクトなモデルを生成する。加えて、学習と評価データの差異に応じて、効果的に作用する素性分類が変化することを示す。品詞 N-gram は単語 N-gram に比べその差異に頑健に機能し、また、様々な分類された素性を同時に利用することで、精度、頑健性がともに向上する。

- 1: Initialization: $\alpha = 0$ and $\lambda = \lambda'$
- 2: Algorithm:
- 3: For $t = 1 \dots T, i = 1 \dots M$
- 4: $\mathbf{w}^* = \arg \max_{\mathbf{w} \in \text{GEN}(\mathbf{a}_i)} \{\lambda \text{RecScore}(\mathbf{w}, \mathbf{a}_i) + \alpha \cdot \Phi(\mathbf{w})\}$
- 5: $\alpha = \alpha + \Phi(\mathbf{w}_i^{\text{ref}}) - \Phi(\mathbf{w}^*)$

図 1 パーセプトロンアルゴリズム

Fig. 1 Perceptron algorithm.

2. 誤り訂正

2.1 リランキングに基づく誤り訂正

まず本稿における誤り訂正の具体的な手順を述べる。

音響入力 \mathbf{a} に対する音声認識の仮説集合を $\text{GEN}(\mathbf{a})$ とする。 $\text{GEN}(\mathbf{a})$ は、一般に単語 N-best リストやラティスである。また、ある単語列 $\mathbf{w} \in \text{GEN}(\mathbf{a})$ に対する音声認識スコアを $\text{RecScore}(\mathbf{w}, \mathbf{a})$ で表記する。

このとき、一般に誤り訂正は次式に基づき行われる。

$$\mathbf{w}^* = \arg \max_{\mathbf{w} \in \text{GEN}(\mathbf{a})} \{\lambda \text{RecScore}(\mathbf{w}, \mathbf{a}) + \alpha \cdot \Phi(\mathbf{w})\} \quad (1)$$

なお、 $\Phi(\mathbf{w})$ は素性ベクトル、 α はその重みベクトルであり、誤り訂正のモデルパラメータとなる。また、 λ は認識スコア重み定数、 \cdot は内積演算を表す。

原理的には、素性ベクトル Φ には音響的特性を表現する素性を導入可能であるが、ここでは言語的特性を表現する素性に限定している。具体的には N-gram 頻度を採用する。例えば

$$\Phi_j(\mathbf{w}) = \text{Count of word bi-gram "の 次" in } \mathbf{w}$$

といったものを使用する。単語 N-gram に限らず、音素記号、文字、品詞等の N-gram の出現頻度を使用することも可能である。

2.2 誤り訂正モデルの学習

本稿ではパラメータ α の推定にパーセプトロンアルゴリズムを採用する。高速な学習が可能であり、実装も容易である。局所最適解へ収束する危険性を持つが、他の手法とほぼ同程度の誤り訂正モデルを生成可能であることが経験的に知られている。

アルゴリズムを図 1 に示す。図中の i は音声データのインデックス、 M はその総数である。 T は再帰回数の上限値を表す。

まず学習に先立ち、音声認識器を用いて各音声データの複数単語列集合 $\text{GEN}(\mathbf{a}_i)$ と正解単語列 $\mathbf{w}_i^{\text{ref}}$ を用意する。そして、図中 1 行目のようにパラメータ α が零ベクトル $\mathbf{0}$ に、認識スコア重みが任意の定数 λ' に設定される。 α は 5 行目の式により更新される。更新に関与する単語列 \mathbf{w}^* は、4 行目の式、すなわち式 (1) により選ばれる。暫定的な α の下でリランキングを行い、その最上位単語列をパラメータ更新に使用する。この処理を全音声データに対し再帰的に T 回適用することで学習を行う。また、 $\mathbf{w}_i^{\text{ref}}$ を、 $\text{GEN}(\mathbf{a}_i)$ 内の最も WER の低い単語列 (オラクル単語列) で代用することにより、更に高精度な誤り訂正モデルが生成可能であることが経験的に知られており [1]、本稿でもこれを使用する。

最後に、次式により平均パラメータを定義する。

$$\alpha_{ave} = \frac{1}{TM} \sum_t \sum_i \alpha_t^i \quad (2)$$

α_t^i は t 回目の \mathbf{a}_i に対する更新後のパラメータ α である。未知の評価データに対し誤り訂正式 (1) により適用する場合、平均パラメータを用いることで認識精度が向上することが知られている。

3. WER 基準による対立単語列の選定

本稿では、以下の要領でパラメータ推定に関する対立単語列の限定を行う。

今、複数単語列集合として N-best リストを考える。そしてリスト中の各単語列を WER 順に昇順で並び替え、その順位を上付き文字で表現する。つまり \mathbf{w}^k は WER 順で第 k 位の単語列を表す。 \mathbf{w}^1 は最も WER の低い単語列、すなわちオラクル単語列である。

ここで N-best リストを構成する単語列集合 $\text{GEN}(\mathbf{a})$ の部分集合 $\text{GEN}(\mathbf{a}, x, y) = \{\mathbf{w}^1\} \cup \{\mathbf{w}^k | x \leq k \leq y, 2 \leq x \leq y \leq N\}$ を導入する。 $\text{GEN}(\mathbf{a}, 2, N)$ は N-best リスト中の全単語列、 $\text{GEN}(\mathbf{a}, 2, 2)$ はオラクルとその次に WER の低い単語列からなる集合 $\{\mathbf{w}^1, \mathbf{w}^2\}$ である。学習時に $\text{GEN}(\mathbf{a})$ に代え、 $\text{GEN}(\mathbf{a}, x, y)$ を用いることで、パラメータ更新に関わる単語列は WER の順位により制限を与える。

ただし、WER と音声認識スコアの負の相関に注意する必要がある。 $\text{GEN}(\mathbf{a}, x, y)$ を用いて図 1 に示される学習を適用する際、対立単語列 \mathbf{w}^* は 4 行目の式に従い選択される。つまり、 λ 値が大きいくほど、大きな音声認識スコア $\text{RecScore}(\mathbf{w}, \mathbf{a}_i)$ を持つ単語列が \mathbf{w}^* として選択される割合が増す。大きな音声認識スコアを持つ単語列の WER は低い値を持つ傾向が強いため、結果として WER の低い単語列が \mathbf{w}^* として選択される。これでは、WER 基準での単語列の限定の効果が損なわれてしまう。そこで、本稿では $\lambda = 0$ に設定する。これは学習時のみ適用される事項である。生成したモデルを誤り訂正に使用する際は正の値を用いる。曖昧性の解消のため、以降それぞれにおける λ を、 λ^{train} , λ^{test} と表記する。

要約すると、学習時は $\text{GEN}(\mathbf{a}, x, y)$ および $\lambda^{\text{train}} = 0$ によりパラメータを推定し、評価時は $\text{GEN}(\mathbf{a})$ および $\lambda^{\text{test}} > 0$ により式 (1) を適用し誤り訂正を実現する。

次にモデルサイズについて記述する。モデルサイズはパラメータ α の要素数に一致する。したがって、実装上は、学習データ $\text{GEN}(\mathbf{a}, x, y)$ から抽出可能な素性の種類の総数分の領域が計算機上に確保されればよい。しかし、事実上はパラメータ更新式 (図 1, 5 行目) 中の $\Phi(\mathbf{w}_i^{\text{ref}}) - \Phi(\mathbf{w}^*)$ において、非零を持つ要素のみが必要となる。そのため、実際には、該当要素が新規に出現した場合のみ計算機上にその要素分の領域が確保される。コンパクトで、かつ高精度なモデル生成には、誤り訂正に効果的な素性が効率的に獲得される単語列を、対立単語列 \mathbf{w}^* として選ぶことが重要である。そして、その上で、学習中の再帰処理において、訂正効果の重複する類似の素性が抽出される単語列の選択を抑えることも重要となる。

表 1 素性分類

Table 1 Feature classes.

分類名	例
単語	戻る+モデル+動詞/ラ行五段/連体形
表層	戻る
読み	モデル
品詞	動詞
品詞細分類	ラ行五段
文字	戻
読み表記文字	モ
音素	m

表 2 学習セットおよび評価セットのデータサイズ

Table 2 Data size of training and evaluation sets.

	講演種別	講演数	発話数	パープレキシティ
学習セット	学会講演	150	25,130	68.7
評価セット 1	学会講演	10	1,293	76.1
評価セット 2	学会講演	10	1,156	74.4
評価セット 3	模擬講演	10	717	142.3

4. 素性の分類

誤り訂正において広く使用される単語 N-gram 素性以外に、本稿では表 1 に分類した単位の N-gram を採用し、比較する。

本稿における単語は、CSJ 規定の短単位に一致し、表層、読み、品詞等の情報の組により与えられる。文字素性は、単語に比べ短い時間セグメントに出現する認識のパターンを表現するものと考えられる。品詞素性は、単語と同じ時間セグメントを有するが、その種類の総数が少なくデータスパース問題の影響は小さい。音素素性は、極めて短い時間セグメントの認識パターンを表現し、その種類の総数も少ない。どの素性についても、一般に、素性の抽出幅、すなわちコンテキストを広範に設定することで、多彩な認識誤りのパターンを捉えることができる。しかしそれはデータスパース問題とのトレードオフ関係にあり、両者の調整が必要となる。

5. 実験

5.1 実験条件

CSJ の学会・模擬講演データを使用し、種々の誤り訂正モデルを作成し、その精度、モデルサイズを比較した。

まず誤り訂正モデルの学習・評価双方に必要な音声認識の N-best リスト生成について記述する。認識デコーダには、NTT コミュニケーション科学基礎研究所で開発されている SOLON [9] を使用した。音響モデルには評価セットを除く全学会講演から識別的手法 [10] により学習された tri-phone HMM (5000 状態-32 混合) を用いた。言語モデルには学会講演 817 講演分から学習された tri-gram を使用した。語彙サイズは 51,574 である。

N-best リスト生成のため、講演データを発話単位に分割した。学習・評価に使用したデータの講演数、発話数、および上記言語モデルで評価したテストセットパープレキシティは表 2 に示す通りである。

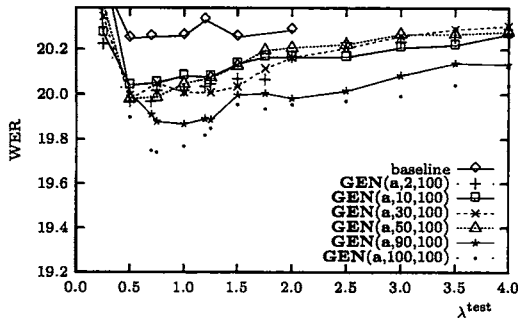


図2 WER 基準での対立単語列選定とその効果
Fig.2 Competitor selection based on WER.

表3 学習時の対立単語列集合とモデルパラメータ数

Table 3 Competitors and parameter size of error corrective models.

GEN(a, x, y)	λ^{train}	パラメータ数
GEN(a, 100, 100)	0.0	1,304,738
GEN(a, 90, 100)	0.0	2,100,474
GEN(a, 50, 100)	0.0	2,302,995
GEN(a, 2, 100)	0.0	2,447,950
GEN(a, 2, 100)	0.5	2,110,782
GEN(a, 2, 100)	1.5	2,234,797

特に評価セット3は模擬講演により構成されており、パープレキシティは他の評価セットに比べ大きい。これは言語モデルおよび誤り訂正モデルに対するタスク外データとして用いるものである。

まず、発話単位の音声データに対し音声認識を適用し、100-best リストを生成した。学習セットの100-best リストを用い、GEN(a, x, y) の x, y 値、素性分類別に誤り訂正モデルの学習を行った。なお、比較のため、100-best リスト中の全仮説、すなわち GEN(a) を用い、 $\lambda^{\text{train}} = \lambda^{\text{test}} > 0$ の制約下で学習・評価を行う baseline 誤り訂正モデルを作成した。

5.2 WER 基準による対立単語列選定の効果

WER 基準で対立単語列を制限した状況下でモデル学習を行い、各モデルの比較を行った。素性には、最も一般的に使用される単語1~N-gram 素性を採用した。Nは、3以上ではほぼ同一の精度を示すが、実験的に最大の誤り削減率を示したN=7における結果を示す。

まず、GEN(a, x, y) について、y = 100 に固定し、x 値を変えて学習された誤り訂正モデルを、評価セット1に対して適用した際の WER を図2に示す。x が大きな値を持つ時、より多くの誤りが訂正されている。低い WER を持つ単語列を学習セットから除外することで、WER の高い単語列との識別が促進され、多量の誤りパターンが効率的に獲得された効果であると考えられる。特に、GEN(a, 100, 100) = $\{w^1, w^{100}\}$ 、すなわちオラクル単語列と WER の最も高い単語列の識別のみで、高精度な誤り訂正モデルが生成可能であるという結果が得られた。

次に各モデルサイズを比較する。表3は学習条件(GEN(a, x, y), λ^{train})別のモデルのパラメータ数である。ここで、パラメータ数とは非零の α_{ave} の要素数である。x ≤ 90 ではパラメータ数はほぼ一定であり、 λ^{train} にもそれほど依存しないことがわかる。一方、x = 100 ではパラメータ数は大幅に減少している。前述の通りこのモデルは最も高精度であり、WER の高い単語列と識別を行うことで、モデル性能に重大な素性が効率的に抽出されていることが分かる。

5.3 素性分類別の効果

ここでは、素性分類別のモデル性能を評価する。モデル学習は、GEN(a, 100, 100), $\lambda^{\text{train}} = 0.0$ の条件下で行った。素性分類ごとの N-gram コンテキスト長は評価セット1に対する WER を最小にする値に設定した。

図3~5は、各評価セットへ各誤り訂正モデルを適用した際の WER である。図中の、Word, Surf, Read, POS, POS-sub, Char, Read-Char, Phone はそれぞれ単語、表層、読み、品詞、品詞細分類、文字、読み表記文字、音素を表し、付随の数字は対応する素性の N-gram コンテキスト長を表す。例えば、Surf3 は、表層1~3-gram を素性に持つ誤り訂正モデルを表す。

まず、言語モデル学習用および誤り訂正モデル学習用データと同じく学会講演から構成される評価セット1および2に着目する(図3, 4)。極めて短い時間セグメントの認識パターンを表現する音素系列(Phone)において性能がやや劣るものの、どの素性でもほぼ同一の WER が得られている。評価セット1と2の間の比較では、音声認識器の言語モデルにより算出したテストセットパープレキシティのやや大きな評価セット1の方が、若干素性間の精度のばらつきが大きいという結果が得られた。

一方、模擬講演のみで構成される評価セット3では結果は一変する(図5)。各素性分類別の誤り訂正モデルの精度のばらつきは更に拡大し、その中で品詞を素性とする誤り訂正モデルが他のモデルより高い性能を示した。逆に、文字や音素といった時間セグメントの比較的短い変化を表現する素性では、僅かな誤り削減幅に止まっている。

これらの結果は、誤り訂正モデルの学習において、多様な誤りのパターンを獲得することの重要性を示唆するものであると考える。特に、品詞のようにデータスパース問題を緩和する素性では、タスクへの頑健性が向上し、安定的に高い誤り削減率を実現する。

5.4 素性の組合せによる効果

最後に、分類別の各素性を併用した際の誤り訂正モデルを評価する。ただし、異なる分類に属する素性の組合せ素性、例えば、表1において、品詞‘動詞’と読み‘モデル’の組合せからなる‘品詞+モデル’の頻度等の素性は使用していない。

モデル学習は、先と同様に GEN(a, 100, 100), $\lambda^{\text{train}} = 0.0$ の条件下で行った。各素性分類の N-gram コンテキスト長も同様に先の実験に揃えている。

図6~8に、評価セット別の性能を示す。ただし、素性の組合せの総数は多数存在するため、各評価セットに対し最も低い WER を与える5種のモデルに限定している。例えば図中の“Phone4_POS3_Surf3_POS-sub3_Word7”はアンダーパー

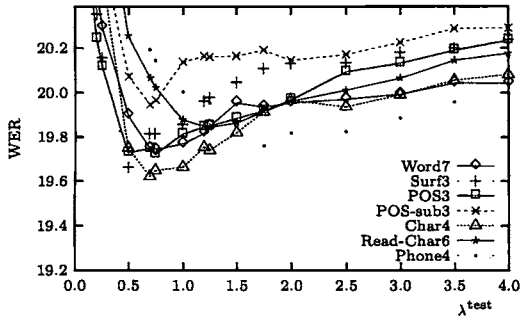


図3 素性分類別性能-評価セット1

Fig. 3 Feature classes and model performance - evaluation set 1.

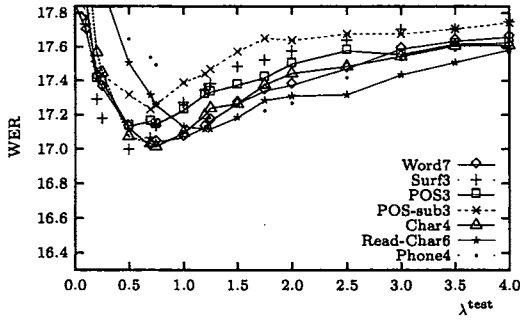


図4 素性分類別性能-評価セット2

Fig. 4 Feature classes and model performance - evaluation set 2.

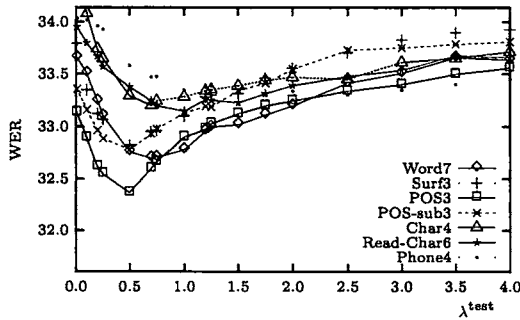


図5 素性分類別性能-評価セット3

Fig. 5 Feature classes and model performance - evaluation set 3.

‘ λ ’で区切られた素性の組から構成されたモデルを表しており、この場合、音素、品詞、表層、品詞細分類、単語の5種の素性分類からなる。

まず、分類別に素性を単独で使用した場合の結果である図3~5と比較する。評価セット2での比較(図4 vs. 7)において、素性の組合せにより大きく精度が向上していることが分かる。各素性の単独使用時には、各モデル間の精度差は殆ど観測されなかったにも関わらず、加算的な精度向上が得られたことは、各素性が誤りのパターンに関する異なる情報を有していることを示している。この結果は他の評価セットでも同様に見られる。特にタスク外データとして用いた評価セット3について

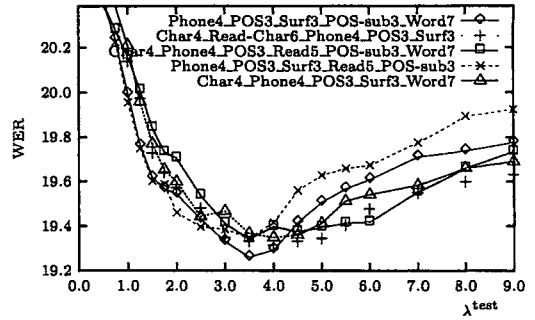


図6 素性組合せ時のモデル性能-評価セット1

Fig. 6 Feature combination and model performance - evaluation set 1.

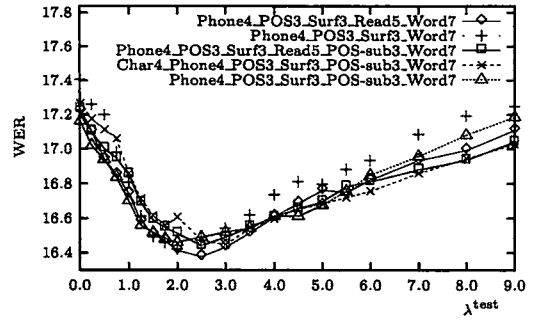


図7 素性組合せ時のモデル性能-評価セット2

Fig. 7 Feature combination and model performance - evaluation set 2.

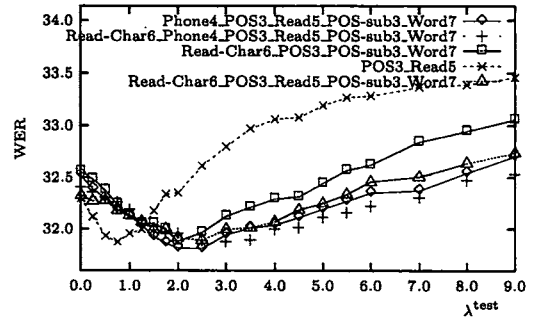


図8 素性組合せ時のモデル性能-評価セット3

Fig. 8 Feature combination and model performance - evaluation set 3.

も同様の結果が得られたことは、素性の併用が他タスクに対する頑健性を向上させるということを示す重要な結果であると言える。

素性分類別には、単語、品詞、音素の組合せ時に高精度な誤り訂正モデルが得られている。単語素性は直接的に誤り単語を検出可能であり、品詞は抽象度の高いシンボル系列でデータスパース問題を軽減し、多様な誤りパターンを表現できる。音素素性は、単独使用時の性能は低いものの、単語や品詞素性に対し極めて短い時間セグメントを表現可能であるため、単語およ

び品詞素性とは異なる情報を内包していると考えられる。その結果、加算的な精度向上に寄与したものと考えられる。

6. ま と め

誤り訂正モデルの識別的学習において、従来、正解と誤りの組の獲得が重視されてきたのに対し、本稿では多様な誤りパターンの獲得に着目した。高精度な誤り訂正モデルの生成には、WERの高い単語列を重視した識別学習が重要であり、特に、対立単語列を、音声認識の複数単語列中の最も WER の高い単語列に限定することで、高精度でコンパクトな誤り訂正モデルを実現した。また、素性には、従来広く用いられてきた単語 N-gram より、品詞 N-gram の方がタスクの違いにも頑健に機能することが明らかになった。これは、抽象度の高い素性の採用により、データスパース問題を軽減し、多様な誤りパターンを表現しうること起因するものと考えられる。加えて、様々な素性の同時利用により各素性が互いに異なる誤りパターンを表現し、更に高精度なモデルが生成可能であることを示した。

文 献

- [1] B. Roark, M. Saraclar and M. Collins, "Corrective Language Modeling for Large Vocabulary ASR with the Perceptron Algorithm," *Proc. ICASSP*, vol. 1, pp. 749-752, 2004.
- [2] B. Roark, M. Saraclar, M. Collins, and M. Johnson, "Discriminative language modeling with conditional random fields and the perceptron algorithm," *Proc. ACL*, pp. 47-54, 2004.
- [3] Z. Zhou, J. Gao, F. K. Soong and H. Meng, "A Comparative Study of Discriminative Methods for Reranking LVCSR N-Best Hypotheses in Domain Adaptation and Generalization," *Proc. ICASSP*, vol. 1, pp. 141-144, 2006.
- [4] 小林彰夫, 佐藤庄衛, 尾上和穂, 本間真一, 今井亨, 都木徹, "単語ラティスの識別的スコアリングによる音声認識," 秋季音響学会講演論文集, pp. 233-234, 2007.
- [5] 大庭隆伸, 堀貴明, 中村篤, "認識誤りに対する各単語 N-gram の関与度を考慮した誤り訂正学習," 春季音響学会講演論文集, pp. 73-74, 2007.
- [6] T. Oba, T. Hori and A. Nakamura, "An Approach to Efficient Generation of High-Accuracy and Compact Error-Corrective Models for Speech Recognition," *Proc. Interspeech*, pp. 1753-1756, 2007.
- [7] 大庭隆伸, 堀貴明, 中村篤, "誤り訂正モデルにおける単語誤り率基準での対立仮説選択とその効果," 春季音響学会講演論文集, pp. 121-122, 2007.
- [8] K. Maekawa, H. Koiso, S. Furui and H. Isahara, "Spontaneous Speech Corpus of Japanese," *Proc. LREC*, pp. 947-952, 2000.
- [9] 中村篤, 他, "音声認識システム SOLON の日本語話し言葉コーパスによる評価 (2006 年版)," 倍学技報 2006-SLP-64, pp. 251-256, 2006.
- [10] E. McDermott, T.J. Hazen, J. Le Roux, A. Nakamura and S. Katagiri, "Discriminative training for large vocabulary speech recognition using Minimum Classification Error," *IEEE Trans. ASLP*, 2007.
- [11] M. Collins, "Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with Perceptron Algorithms," *Proc. EMNLP*, pp. 1-8, 2002.
- [12] I. Shafran and K. Hall, "Corrective Models for Speech Recognition of Inflected Languages," *Proc. EMNLP*, pp. 390-398, 2006.