

単語間の関連性を利用した音声認識用言語モデル

森谷 高明[†]

広瀬 啓吉[†]

峯松 信明[‡]

moriya@sail.t.u-tokyo.ac.jp hirose@gavo.t.u-tokyo.ac.jp mine@gavo.t.u-tokyo.ac.jp

[†] 東京大学大学院新領域創成科学研究科

[‡] 東京大学大学院情報理工学系研究科

〒 113-8656 東京都文京区本郷 7-3-1

あらまし 一般に、大語彙音声認識システムにおける統計的言語モデルの構築には大量の学習用コーパスが必要である。しかし音声認識の対象は多くの場合ある特定の内容のタスクであるため、必ずしもそのような所望のタスクについて大量のテキストを収集できるとは限らない。そのためこのような問題に対処する方法として、目的のタスクに関する少量のテキストを用いて言語モデルを補正するタスク適応が研究されている。本研究では、主にタスク適応の性能向上を念頭に、単語間の関連性という要素を言語モデルに組み込み、実験的検討を行った。我々の提案する単語間の関連性を用いたタスク適応モデルは、比較的長いスパンで単語生起関係を捉え、目的タスクにおいて潜在的に生起する可能性の高い単語列の生成確率を補正するモデルである。この手法を導入した結果、MAP 推定を用いた従来のタスク適応と比較して、テキストの量によらず性能向上がみられた。また新たな知見として、評価タスクと適応学習タスクが一致しないような、従来の MAP 推定ではあまり適応効果が見られない場合においても、本手法が有効であることが確認された。

キーワード 単語間の関連性, 音声認識, 言語モデル, タスク適応, MAP 推定, 頑健性

Language Modeling with Characteristics of Distant Word Co-occurrence for Speech Recognition

Takaaki MORIYA[†]

Keikichi HIROSE[†]

Nobuaki MINEMATSU[‡]

moriya@sail.t.u-tokyo.ac.jp hirose@gavo.t.u-tokyo.ac.jp mine@gavo.t.u-tokyo.ac.jp

[†] Graduate School of Frontier Sciences, University of Tokyo,

[‡] Graduate School of Information Science and Technology, University of Tokyo,

7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, JAPAN

Abstract In a large vocabulary speech recognition system, statistical language modeling usually requires a huge amount of text corpus. But it is difficult to collect corpus of the sufficient size because recognition task is often specific to a domain. One strategy to solve this problem is task adaptation, where MAP adaptation is widely and successfully used. In this paper, we experimentally investigate language model adaptation with co-occurrence of distant words. This paradigm introduces correlations between distant words additionally into the language model. By using this method, the language probabilities can be modified according to words, which are not actually seen but can be potentially found in a given context. Experiments showed that higher reduction rate of perplexity is observed in comparison with MAP estimation. It should be especially noted that the proposed method can decrease the perplexity even when the MAP estimation increases the perplexity, i.e., adaptation of original language models to a target task by using text data close to but strictly out of the target task.

Keywords co-occurrence of distant words, speech recognition, language model, task adaptation, MAP estimation, robustness

2 MAP 推定による言語モデルのタスク適応

1 はじめに

音声認識の対象はある特定の内容に関するタスクである場合が多い。高精度の音声認識用統計的言語モデルの構築には大量の学習用コーパスが不可欠であるが、実際には特定のタスクについて言語モデル作成に十分な量のテキストが収集できるとは限らない。このような問題に対処する方法として、近年言語モデルのタスク適応が検討されている。

言語モデルのタスク適応の方法としては、最大エントロピー法^[1]に基づくものや最大事後確率(Maximum A Posteriori; MAP)推定^[2, 3]によるものなどが知られているが、タスク適応の際に、目的タスクのテキストの他に新聞などタスクに依存しない大量のテキストが利用できる場合は、MAP 推定が簡便な方法である。

MAP 推定によるタスク適応は、従来の N-gram 言語モデルの作成手順をそのまま流用することができ簡便である反面、最適重みや適応効果がテキストの種類や量に大きく依存する。このような欠点に対処する方法として、本研究では単語間の関連性を言語モデルのタスク適応に利用する手法を提案する^[4]。本手法は、任意の単語間の関連性を定義しあらかじめ計算しておき、タスク適応の際に目的タスクのテキストの単語情報を加味した上で単語間の関連性を利用し、言語モデルの補正を行う方法である。このようにいわば間接的な単語頻度情報を取り入れることによって、テキスト量に依存するという MAP 推定の欠点を補うことを図る。

また実用上は、なるべく広範なタスクに対し適応化された言語モデルが要求される場合や、目的タスクのテキストが収集できないため類似のタスクで補わなければならないような場合も考えられる。本稿では、このようにタスクに対する「頑健性」が要求される場合にも、テキストに直接出現しなかった単語を扱える本手法が有効であることを示す。

本稿では、第 2 節にて従来の MAP 推定法によるタスク適応について言及した後、第 3 節にて単語間の関連性をタスク適応に利用する方法を紹介し、いくつかの性質について述べる。そして第 4~6 節にて、本手法が学習テキスト量や評価タスクに対し頑健であることを示すため、補正パープレキシティ^[5]による評価実験を行う。また第 7 節において、本手法のパラメータの性質について言及する。最後に第 8 節において、本手法を音声認識実験によって評価する。

本研究では、言語モデルのタスク適応のベース手法として、MAP 推定により少量の目的タスクのテキストと大量のテキストを混合する方法を用いる^[6]。以下、特定のタスクに依存しない大量テキストを独立テキスト、目的タスクの少量テキストを適応テキスト、評価に用いる特定のタスクのテキストを評価用テキストと呼ぶことにする。

\mathbf{y} を観測データ、 θ を推測データとすると、ベイズの定理より、

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})} \propto p(\mathbf{y}|\theta)p(\theta) \quad (1)$$

MAP 推定では、次式のように θ を推測する。

$$\hat{\theta} = \arg \max_{\theta} p(\theta|\mathbf{y}) = \arg \max_{\theta} p(\mathbf{y}|\theta)p(\theta) \quad (2)$$

言語モデルの推定の場合、観測データ \mathbf{y} は単語 w の出現回数 (N_w) に相当し、推測データ θ は履歴単語列 h から次の単語 w へ遷移する確率 $P(w|h)$ に相当する¹。

MAP 推定を N-gram 言語モデルのタスク適応に利用する場合、 $p(\mathbf{y}|\theta)$ は適応テキストにおける単語遷移確率 $P(w|h)$ の尤度に相当する。尤度関数としては、通常の N-gram モデルの最尤推定のときと同様、複数回試行のペルヌイ試行を与えることが多い。一方、事前分布 $p(\theta)$ は独立テキストにおける単語の遷移確率 $P(w|h)$ に相当し、多くの場合事後分布と自然共役となるように分布を与える。事前分布としてベータ分布やディリクレ分布を与えた場合、適応テキストにおける単語の出現回数に重み W_N を乗じたものと捉えることができる。すなわち

$$p(N|P) \propto P^{W_N N_{hw}^A} (1-P)^{W_N (N_h^A - N_{hw}^A)} \quad (3)$$

$$p(P) \propto P^{N_{hw}^I} (1-P)^{N_h^I - N_{hw}^I} \quad (4)$$

ただし N_x^I , N_x^A は独立テキスト、適応テキストにおける単語 (列) x の出現回数である。式 (3)(4) から、MAP 推定により、

$$P(w|h) = \arg \max_P p(N|P)p(P) \quad (5)$$

を求める。これを解くと、

$$P(w|h) = \frac{N_{hw}^I + W_N \cdot N_{hw}^A}{\sum_w (N_{hw}^I + W_N \cdot N_{hw}^A)} \quad (6)$$

という式が導かれる。

¹ 本稿では、単語 w の直前から $(N-1)$ 単語前までの単語列を、履歴単語列 h と表記する。

3 単語間の関連性を利用した言語モデル

3.1 タスク適応への応用

本節では、前節の MAP 推定法によるタスク適応に、単語間の関連性という新たな要素を組み入れることを試みる。手順は以下の通りである。

1. タスクに関係なく現れる単語のリスト (タスク独立語リスト) を作る。これは、助詞・助動詞など、タスクに共通して出現する単語のセットである。
2. 独立テキスト T^I を、 $T^I = T_1^I T_2^I \dots T_{n_I}^I$ に分割する。ただし分割された各 T_k^I は、それぞれある特定の話題について述べている文の集合となるようにする。
3. 言語モデルの語彙に含まれている単語に関して、 T_k^I 内に単語 v と任意の単語列 hw (bigram なら 2 つ, trigram なら 3 つの単語の並び。 h は w に対する履歴単語列) が同時に出現するとき 1 を返し、そうでないときに 0 を返す関数を $q_{v[hw]}^k$ とし²、 k に関して和をとる。(図 1 参照)

$$Q_{v[hw]}^I = \frac{1}{n_I} \sum_{k=1}^{n_I} q_{v[hw]}^k \quad (7)$$

また $q_{v[hw]}$ を適応テキストにおいても計算しこれを $q_{v[hw]}^A$ とする。 $q_{v[hw]}^I$ 、 $q_{v[hw]}^A$ は単語 v と単語列 hw との間の関連性の程度を示しており、値が大きいほど両者の関連性が高いことを表している。なお、各 T_k^I 内で $q_{v[hw]}^k$ を計算するのは、異なる話題間での単語の関連性を計算することを防ぐためである。

4. 適応テキスト T^A 内で、タスク独立語リストに含まれない全ての単語 v^A について³、 $q_{v^A[hw]}$ の和を計算する。すなわち、

$$Q_{hw}^X = \frac{1}{\sum_{v^A} C(v^A)} \sum_{v^A \in T^A} C(v^A) q_{v^A[hw]}^X \quad (X=I, A) \quad (8)$$

ただし、 $C(v^A)$ は T^A 中に単語 v^A が現れる回数とする。この Q_{hw} は、適応テキスト T^A 内で hw に関連する単語がどれくらい多く存在するかを表している。(図 2 参照)

² このほかに、単語間の距離を定義し連続値を返す関数なども考えられるが、本稿ではこのような最も基本的な 2 値関数を用いる。

³ hw にはタスク独立語リストに登録されている語が含まれていてもよい。

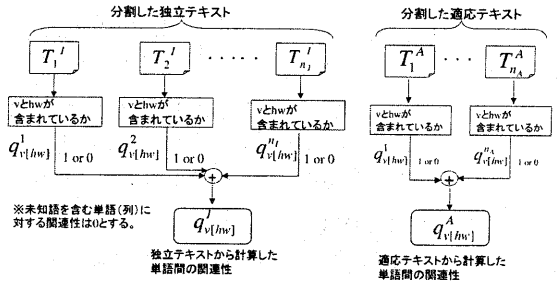


図 1. 単語間の関連性の計算方法 (概念図)

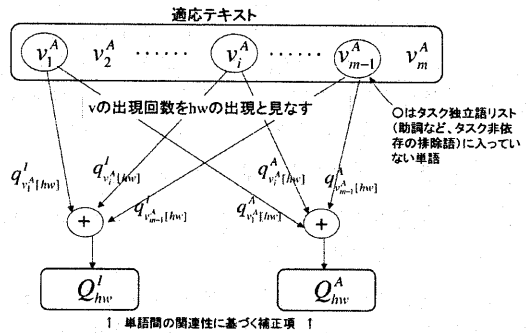


図 2. 単語間の関連性に基づく補正項の算出 (概念図)

5. Q_{hw}^I と Q_{hw}^A の重み付き和をとる。

$$Q_{hw} = (1 - \lambda) Q_{hw}^I + \lambda Q_{hw}^A \quad (9)$$

独立テキストと適応テキストのタスクが離れている場合、従来の単語間関連性の計算方法 [7] では性能向上が難しい。したがってこのように適応テキストからも計算した単語間関連性の効果を取り入れることによって、適応テキストにのみ存在する単語の生起確率が上昇することが期待される。

λ は Q_{hw}^I と Q_{hw}^A の相互関係を表しており、最適な性能 (最小パープレキシティ) を与える λ が小さいほど、適応テキストから関連性を計算する必要性が低く、独立テキストから計算した単語間の関連性だけで十分な性能が確保できるということの意味する。

6. MAP 推定によるタスク適応の式 (6) に、(9) 式の単語間の関連性に基づく補正項 Q_{hw} を組み

込み、単語遷移確率を次式で計算する。

$$P(w|h) = \frac{N_{hw}^I + W_N \cdot N_{hw}^A + \alpha Q_{hw}}{\sum_w (N_{hw}^I + W_N \cdot N_{hw}^A + \alpha Q_{hw})} \quad (10)$$

とする。ただし α は Q_{hw} の重み係数である。即ち、式(8)の単語 v^A の存在によって単語列 hw があたかも出現したかのように扱う訳である。

3.2 本手法の特徴

提案する手法の大きな特徴は、適応テキストに出現した単語の頻度情報を元に、あらかじめ独立テキストと適応テキストで計算しておいた単語間の関連性をモデルに取り込むところにある。しかも単語間の関連性は、bigram や trigram よりもより広いスパンで単語の頻度情報を捉えることができる。したがって、適応テキストにはたまたま出現しなかったが、評価用テキストには潜在的に出現する可能性のある単語の情報も扱うことができる。ゆえに従来のMAP 推定法によるタスク適応の場合と比べて、

- 適応用テキストの量が少ない場合
- 適応用テキストと評価用テキストのタスクが一致しない場合

のようなケースに対し特に有効であると考えられる。前者は、タスク適応に十分な量の適応用データの確保が難しいという根本的な問題を克服するものである。後者は、少量の適応テキストでより広範囲のタスクに適応化できるということであり、認識対象タスクに対するロバスト性を強化することを意味する。

4 実験条件

4.1 テキストの準備

我々の先の報告^[7]では、独立テキストを新聞、適応テキスト・評価用テキストを特定の新聞記事とした場合について、テキストの分割方法や量、タスク独立語リストの設定などの基礎的な性質について報告を行った。今回は、独立テキストを新聞とし、適応テキストのタスクが独立テキストからより離れている場合について検証を行うこととした。この際、

- 新聞の表現と完全には一致し得ないが、なるべく会話文に近い平易な表現の適応テキストであること
- 独立テキストから目的タスクを推定可能であるようにするため、未知語率をできる限り抑えること

- * 可能な限り、固有名詞等以外は新聞の語彙でカバーできる適応テキストであること

表 1. テキスト概要 (文数・総形態素数・異なり形態素数)

テキスト	タスク	文数	総形態素数	異なり形態素数
独立	新聞記事 2 年分	2438662	58290111	200380
適応	ピーターパン	133	2156	639
		526	8931	1738
		699	11766	2201
評価用 A	ピーターパン	107	1709	570
評価用 B	マッチ売り	96	1719	495

- * カバー率を高めるため、独立テキストと適応テキストを重み付き混合した後、語彙を 2 万語に制限する

などの点を考慮し、J.Barrie の小説 (物語)『ピーターパン』日本語訳を適応テキスト、および評価用テキストとして用意した。尚、適応テキストの量による性能変化を検証するため、適応テキストは 133 文、526 文、699 文の 3 通りを用意した。また、適応テキスト『ピーターパン』と同一ではないが近いタスクの評価用テキストとして、H.Andersen の童話『マッチ売りの少女』日本語訳も用意した⁴。尚、すべての実験において、適応テキストと評価用テキストに重複する文のないようにした。独立テキストとしては、毎日新聞 CD-ROM95~96 年度版を用いた。言語モデル作成の際、テキストの整形は IPA の言語モデルの作成方針に準じ、音声情報として不要な形態素を排除した。形態素解析には『茶釜』ver2.0^[8]を用いた。以上、実験に用いたテキストの概要を表 1 に示す。

4.2 語彙および言語モデル

言語モデルの作成には The CMU-Cambridge Statistical Language Modeling Toolkit^[9] を使用し、2 万語 bigram を構築した。言語モデルの語彙は、独立テキストと適応テキストの単語出現頻度を 1 対 1 で混合し上位 2 万語を抽出したものをを用いた。本来、語彙を作成する段階でも N-gram 頻度計数を行うときと同様に適応テキストの単語出現頻度に重み定数を乗じることが可能である。しかし、語彙の制限方法には様々な方法が考えられ、最適な語彙を事前に設定することは容易ではないことや、語彙設定による性能変化を排除するという方針などから、今回は語彙を固定して実験を行った。なお 1 対 1 混合では事実上適応テキストの単語頻度を無視していることを意味するが、4.1 節の要件から、語彙の設定はタスク適応に十分なものであると考えられる。実際、各評価用テキストに対する未知語率は 8%程度であった。

⁴ このほかに童話『裸の王様』日本語訳も用意したが、実験結果は『マッチ売りの少女』の場合とほぼ同じであった。

4.3 その他の条件設定

タスク独立語リストは、毎日新聞 CD-ROM 95～98 年度版から相互情報量をもとに抽出した 8000 語を用いた [7, 10]。3.1 節手順 2 のテキスト分割方法は、予備実験から最適な性能を与えると予想される文単位分割とした。

5 実験 A：適応テキストと評価用テキストが同じタスクである場合

本節では、適応テキストと評価用テキストが同一タスク（ピーターパン）である場合について、本手法のタスク適応性能を検証する。まず 5.1 節で MAP 推定の重み定数の最適値を求めた上で、5.2 節で単語間の関連性を利用したタスク適応の性能を検証する。

5.1 MAP 推定の重み定数の決定

適応テキストと評価用テキストがともに『ピーターパン』の場合について、まず (6) 式の MAP 推定の重み W_N の最適値を求めた。 W_N と補正パープレキシティの関係を図 3 に示す。この結果から、適応テキストの文数が多いほど MAP 推定の効果が高く（最大 53% 減）、 W_N の最適値は適応テキストの文数が 133 文、526 文、699 文の場合、それぞれ 3000、2500、2500 であった。尚、この種の実験には莫大な計算労力を要するため最適重み定数の推定は容易な作業ではないが、この結果を見る限り、重みの値はおおよそ独立テキストと適応テキストの文数の比から見当をつけても問題ないものと思われる。

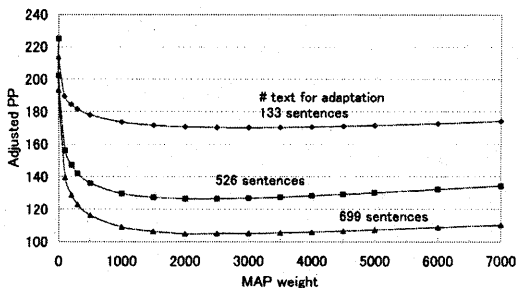


図 3. MAP 推定の重み W_N と補正パープレキシティ (評価用テキスト A)

5.2 単語間の関連性の利用

次に、 W_N を前節の実験より求めた最適値に設定し、単語間の関連性を用いたタスク適応の評価を行った。 λ および α を変化させた場合のグラフを図

4～6 に示す。 $\alpha = 0$ の時は、従来の MAP 推定から作成される言語モデルに相当する。このように、 α を変化させることによって補正パープレキシティが減少しており、適応テキストと独立テキストのタスクが比較的離れている場合においても、本手法の有効性を確認することができた。

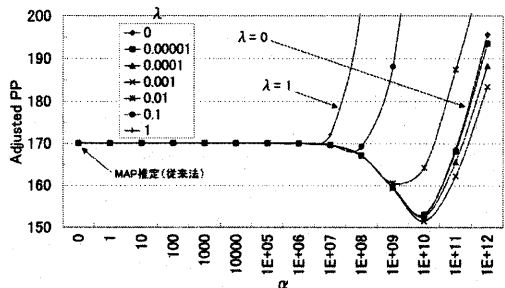


図 4. 適応テキスト 133 文, $W_N = 3000$ (評価用テキスト A)

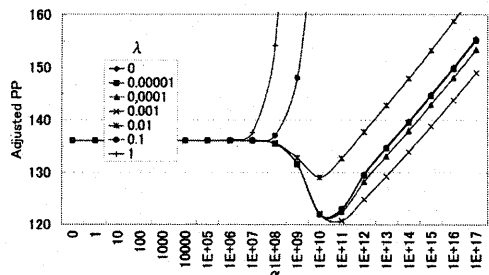


図 5. 適応テキスト 526 文, $W_N = 2500$ (評価用テキスト A)

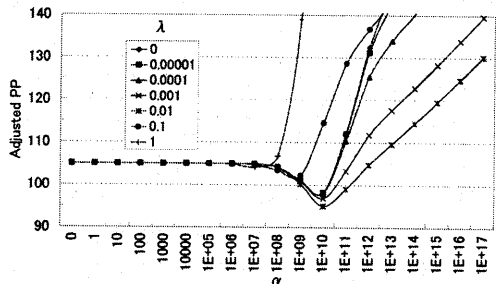


図 6. 適応テキスト 699 文, $W_N = 2500$ (評価用テキスト A)

適応テキストの文数によらず、補正パープレキシティは従来の MAP 推定法の場合から最大 10% 程度減少していることから、本手法によって、適応テキストが少ないほど適応効果が劣化する MAP 推定の欠点を補うことができる。補正パープレキシティの

値が $\lambda = 1$ のときには減少せず $\lambda < 1$ の場合に減少しているのは、MAP 推定の重み W_N の影響による。すなわち重み W_N を設定することで適応テキストに直接出現した単語の頻度情報を MAP 推定を用いて取り入れる一方、 $\lambda < 1$ の条件下で適応テキストのみならず独立テキストの単語間の関連性を組み込み、独立テキスト側から“予想”した単語情報を間接的に取り入れていることに他ならない。

次に参考として、適応テキスト 699 文のもとで $W_N = 1$ とし MAP 推定の効力を弱めた場合の実験結果を図 7 に示す。 α を変化させ単語間の関連性を利用することで補正パープレキシティがほぼ半減していることがわかる。また $W_N = 2500$ or 3000 の場合と異なり $\lambda = 1$ (適応テキストの関連性のみ利用)でも補正パープレキシティが減少していることから、MAP 推定の重みの設定が適切でない場合に、独立テキストと適応テキストでの単語間の関連性を利用し間接的に単語頻度を補正することによって、適応テキストを直接的に扱う MAP 推定による適応効果の補償ができるということがわかった。

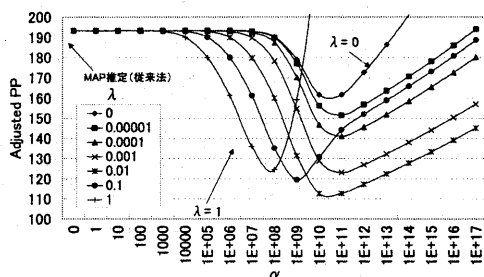


図 7. 適応テキスト 699 文, $W_N = 1$ (評価用テキスト A)

6 実験 B: 適応テキストと評価用テキストが異なるタスクである場合

本節では、適応テキストと評価用テキストのタスク内容にギャップがある場合について、評価用テキストのタスクに対する本手法の頑健性を検証する。まず 6.1 節で MAP 推定の重み定数の最適値を求めた上で、6.2 節で単語間の関連性を利用したタスク適応の性能を検証する。

6.1 MAP 推定の重み定数の決定

適応テキスト『ピーターパン』で適応化を施し、評価用テキスト B『マッチ売りの少女』を評価した場合について、まず(6)式の MAP 推定の重み W_N の最適値を求めた。 W_N と補正パープレキシティの

関係を図 8 に示す。 W_N の最適値は 1000(適応テキストが 526 文の場合)であった。図 3 と比較して、MAP 推定による補正パープレキシティの減少量が小さい(最大 12%減)ことがわかる。しかも、適応テキストが多いと最小パープレキシティの値は小さいが最適重みの範囲が狭くなる傾向が見られ、逆に適応テキストの文数が少ないときは最小パープレキシティの値は大きい最適重みの範囲は図 3 の場合とほぼ同じであった。換言すれば、適応テキストと評価用テキストが違うタスクの場合は、適応テキストの量と W_N の関係があまり単純ではないため、パラメータの設定は簡単ではないようである。

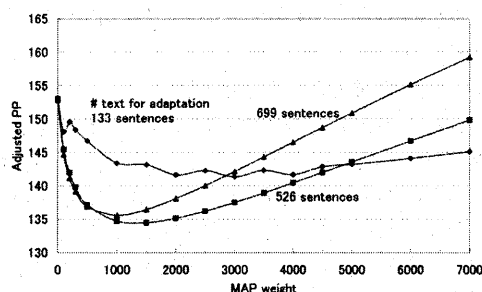


図 8. MAP 推定の重み W_N と補正パープレキシティ(評価用テキスト B)

6.2 単語間の関連性の利用

ここでは、適応テキストが 526 文で、 $W_N = 1000$ の場合と $W_N = 2500$ の場合の実験を行った。前者は 6.1 節の結果から MAP 推定の重みを適切に設定した場合であり、後者は 5.1 節の実験結果を適用し、重み設定が最適でないときを想定した場合である。それぞれ、 λ および α を変化させた場合のグラフを図 9, 10 に示す。このように、 α を変化させ単語間の関連性を取り入れたタスク適応を利用すると、 W_N の重み設定の仕方によらず、補正パープレキシティの値が減少することが確認された。MAP 推定による適応の場合と違って、5.2 節と同様補正パープレキシティが最大 10%程度減少しており、本手法が評価用テキストのタスク内容に対して一定の頑健性を有していることが明らかになった。これは、本手法が学習データで直接出現しなかった単語列の情報も扱っているため、適応テキストのタスクが評価用テキストのそれと多少異なっている、タスクがある程度近ければ評価用テキストに出現する単語を“予想”できるということを意味している。

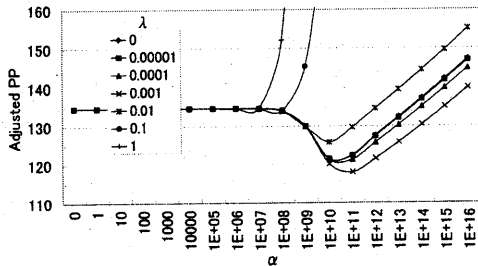


図 9. 適応テキスト 526 文, $W_N = 1000$ (評価用テキスト B)

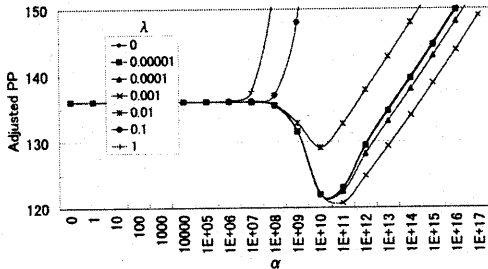


図 10. 適応テキスト 526 文, $W_N = 2500$ (評価用テキスト B)

7 本手法のパラメータについて

単語間の関連性をタスク適応に利用する場合、新たにパラメータ α と λ が追加されるため制御が難しくなるという側面もある。そのため今後はパラメータの自動推定や削減などの検討が必要であると考えられる。数々の実験から、 α は主に学習テキストの量や整形方法の影響を受けることがわかっており、独立テキストが一定ならば α の最適値もほぼ一定になると考えられる。尚、 α の値によってパープレキシティの値は大きく変化するが、グラフの形状はどのような条件でも比較的同じ傾向 (α を横軸・パープレキシティを縦軸にとると下に凸のグラフになる) であるため、 α の自動推定はそう難しくはないものと考えられる。

一方 λ は主にタスクの種類や MAP 推定の重みなどに依存することがわかっている。 λ が 0 に近いほど適応テキストから関連性を計算する必要性が低く、独立テキストから計算した単語間の関連性だけで十分な性能が確保できるということを意味するため、あらかじめ独立テキストから単語間の関連性を計算しデータベース化するなどの方法が可能になる。しかし λ の性質についてはまだ不明な点が多く、 λ の最適値を定量的に求める方法は今後の検討課題で

ある。

8 音声認識実験による評価

本節では、提案手法を音声認識実験によって評価する。認識デコーダに Julius 3.1^[11]、言語モデルには 2 万語 bigram を使い、1st pass による探索結果を用いた、trigram を利用した 2nd pass を用いなかったのは、この実験が適応化言語モデルの比較を目的としているためである。新聞 2 年分のみから作成された初期言語モデル、MAP 推定 (最適重み $W_N = 2500$ or 3000 を適用) による適応化モデル、単語間関連性を利用した適応化モデル (最適パラメータを適用: $\alpha = 10^{10}$, $\lambda = 0.001$, $W_N = 2500$ or 3000) の 3 者を比べた。尚、男性 1 話者について、insertion penalty を -2 に固定した上で、言語重みを 1 から 15 まで変化させその最適値 (=6) で比較を行った。適応テキストと評価用テキストがともに『ピーターパン』の場合と、適応テキストが『ピーターパン』で評価用テキストが『マッチ売りの少女』の場合の単語正解精度の比較を、図 11、図 12 に示す。また、適応テキストの量を 526 文に固定して比較した結果を表 2 に示す。これらの結果、単語正解率や単語正解精度の Error Reduction Rate が数% 上昇することが確認された。

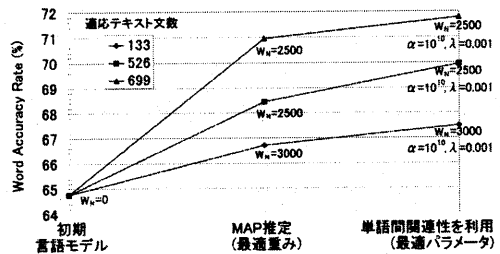


図 11. 単語正解精度 (評価用テキスト A)

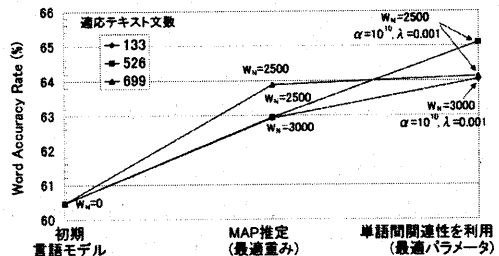


図 12. 単語正解精度 (評価用テキスト B)

表 2. 実験結果：適応テキストを 526 文に固定して比較

	評価用テキスト A		評価用テキスト B	
	MAP	関連性利用	MAP	関連性利用
単語正解率 (%)	73.74	75.05	67.87	69.76
単語正解精度 (%)	68.44	69.94	62.94	65.09
置換誤り率 (%)	22.88	21.77	27.32	25.82
削除誤り率 (%)	3.38	3.19	4.81	4.41
挿入誤り率 (%)	5.31	5.11	4.94	4.68

9 まとめと今後の課題

従来の MAP 推定によるタスク適応は、簡便である反面、最適重みの探索が難しく、適応テキストの量によってタスク適応効果が変動しやすいという欠点を抱えていた。しかし単語間の関連性を利用し間接的に単語遷移確率を補正することにより、適応テキストの量にかかわらず、安定してタスク適応の効果を高めることができた。また、適応テキストと評価用テキストが違うタスクのとき、例えば目的タスクのテキストが十分用意できなかった場合などにおいても、本手法のように適応テキストに直接出現しなかった単語の情報を扱うことのできるタスク適応手法が有効であるという知見が得られた。

すなわち、

- タスク適応の際、目的タスクのテキストの収集が十分にできないとき
- 目的タスクのテキストの代わりに別のタスクのテキストを使って適応化を行わなければならないとき
- MAP 推定の重み定数の最適値が求められないとき

といった場合に単語間の関連性を利用したタスク適応が有効であることが明らかになった。

今後は、7 節で述べたようにパラメータの自動推定や削減などが検討課題である。

謝辞

本研究に対し数々のアドバイスを賜りました佐々木耕樹氏に感謝致します。また、本研究で用いた適応用テキストを提供していただいた青空文庫、Project Sugita Genpaku の皆様に感謝致します。

参考文献

[1] R.Rosenfeld, "A Maximum Entropy Approach to Adaptive Statistical Language Modeling", Computer Speech and Language, Vol.10, No.3, pp.155-186 (1996)

[2] M.Fedelico, "Bayesian Estimation Methods for N-gram Language Model Adaptation", Proc.ICSLP-96, pp.240-243 (1996)

[3] 政瀧 浩和, 匂坂 芳典, 久木 和也, 河原 達也, "最大事後確率推定による N-gram 言語モデルのタスク適応", 電子情報通信学会論文誌 (D-II), Vol.J81-D-II, No.11, pp.2519-2825 (1998)

[4] 森谷 高明, 広瀬 啓吉, 峯松 信明, "単語間の関連性を考慮した言語モデルの適応に関する実験的検討" 日本音響学会春季講演論文集, 2-3-13, pp.79-80 (2001)

[5] J.Ueberla, "Analysing A Simple Language Model - Some General Conclusions for Language Models for Speech Recognition", Computer Speech and Language, Vol.8, No.2, pp.153-176 (1994)

[6] 伊藤 彰則, 好田 正紀, "N-gram 出現回数の混合によるタスク適応の性能解析", 電子情報通信学会論文誌 (D-II), Vol.J83-D-II, No.11, pp.2418-2427 (2000)

[7] K.Sasaki, H.Jiang, K.Hirose, "Rapid Adaptation of N-gram Language Models Using Inter-word Correlation for Speech Recognition", Proc.ICSLP-2000, Vol.4, pp.508-511 (2000)

[8] 松本 裕治 他, "日本語形態素解析システム『茶釜』 version 2.0", (1999)

[9] P.Clarkson, "The CMU-Cambridge Statistical Language Modeling Toolkit v2", (1997)

[10] T.Kawahara, S.Doshita, "Topic Independent Language Model for Key-Phrase Detection and Verification", Proc.ICASSP-1999, pp.685-688 (1999)

[11] 河原 達也 他, "日本語ディクテーション基本ソフトウェア (99 年度版)", 日本音響学会誌, Vol.57, No.3, pp.210-214 (2001)