単語間の関連性を利用した音声認識用言語モデルのドメイン適応

広瀬啓吉[†] 峯松信明[††] 森谷高明[†]

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ドメインであるため，必ずしもそのような所望のドメインについて大量のテキストを収集できるとは限らない．この問題に対処する方法として，目的のドメイン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ドメイン適応が広く研究されている．本論文では，主にドメイン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．筆者らの提案する単語間の関連性を用いたドメイン適応モデルは，比較的長いスパンにおける単語と単語（列）の共起関係を利用し，目的ドメインにおいて潜在的に生起する可能性の高い単語列の出現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ドメイン適応と比較してテキストの量によらず性能向上が見られた．また，提案する共起関係モデル化手法はテキストの表面的な言語形式への依存性を比較的抑える形態で実装されており，その結果，評価ドメインと適応用学習ドメイン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．

Adaptive Training of Language Models with Inter-word Co-occurrence for Speech Recognition

Keikichi Hirose,[†] Nobuaki Minematsu[††] and Takaaki Moriya[†]

In large vocabulary speech recognition, statistical language modeling usually requires a huge amount of text corpus, which is usually difficult to arrange for a specific domain on which speech recognition is conducted. One probable solution of this problem is domain adaptation, where MAP (Maximum A Posteriori) adaptation is successfully used. In this paper, we investigate the language model adaptation using inter-word co-occurrence information through a series of experiments. This paradigm introduces correlation between words of a long distance additionally into the language model, enabling the language probabilities to 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 was observed compared to the conventional MAP adaptation. It should be especially noted that the proposed method can decrease the perplexity even when the MAP adaptation does not work well, i.e., adaptation of original language models to a target domain by using text close to but strictly out of the target domain. This is considered to be because the proposed method captures the word correlation with limited respect of the surface of the text.

1. はじめに

自動音声認識を目的とした言語モデリング手法として，N-gram をベースとした統計的モデリング手法が広く利用されている[1)]．この場合，モデルの学習には大量のテキストコーパスが必要となるため，その現実的な解として，新聞記事コーパスを用いて言語モデルを構築することが一般的に行われている（以下，ドメイン独立言語モデルと呼ぶ）．一方，自動音声認識の実用的なアプリケーションを考えた場合，認識対象はある特定のドメインに関する場合がほとんどである．数年分の新聞記事コーパスから作成された独立言語モデルが用意されていたとしても，対象とするドメインが変わるとその性能が著しく劣化することは周知の事実であり，これを回避するために，少量の対象ドメインテキストデータを用いて独立言語モデルを適応する手法が近年検討されている．

言語モデルのドメイン適応手法としては，最大エントロピー法[2)]に基づくものや最大事後確率（Maximum A Posteriori; MAP）推定法[3),4)]によるものなどが知

† 東京大学大学院新領域創成科学研究科
Graduate School of Frontier Sciences, The University of Tokyo
†† 東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo

られているのに対し，ドメイン適応の際に，対象ドメインテキストのほかに独立言語モデル構築時に利用した新聞テキスト☆が利用できる場合は，N-gram カウントに基づく MAP 推定法が簡便な方法である[5),6)]．

MAP 推定によるドメイン適応は，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[6)]．このような欠点に対処する方法として，筆者らにより，単語間の関連性☆☆を言語モデルのドメイン適応に利用する手法が提案されている[7),11)]．本手法は，任意の単語間の関連性を定義したうえであらかじめ計算しておき，ドメイン適応の際に，対象ドメインの適応データ中の全単語を trigger word として N-gram カウント数を増減させ，言語モデルの補正（適応）を行う手法である．本論文では，この手法を拡張し，より詳細な実験的検討によってその効果を評価する．単語間の関連性を利用する手法としては trigger モデルが提案され[2),8),9)]，その効果も報告されている．また，trigger モデルを用いたドメイン適応についてもすでに報告されている[2)]．本手法も単語間の関連性を考慮するという立場は trigger モデルと同じである．文献 2) では単語履歴中の単語を trigger word として扱っているが，提案手法では，適応テキスト中の単語を trigger word として扱うなど種々の相違点があり，これらについては 3.2 節で述べる．

実用上は，なるべく広範囲なドメインに対して適応された言語モデルが必要となる場合や，目的ドメインのテキストが収集できないため類似のドメインで補わなければならないような場合も考えられる．本論文で考察する単語間の関連性は，テキストに直接現れる表現形態への依存性を抑え，より潜在的な関連性にも着眼する形で定義されており，その結果，提案手法による適応言語モデルが有効に働くドメイン領域が，従来法よりも広範囲になることが期待される．本論文では，ドメイン適応における「頑健性」に対しても提案手法が優れた性能を示すことを実験的に検証する．

以下，2 章で従来の MAP 推定法によるドメイン適応について言及した後，3 章で本論文で提案する，単語間の関連性をドメイン適応に利用する手法を説明するとともに，先行研究との相違点や本手法の特徴について述べる．次に 4 章，5 章で，本手法が適応テキスト量や，適応ドメインと評価ドメインの差異に対し頑健であることを，補正パープレキシティ[10)]に基づく評価実験を通して検証する．6 章では，本手法を連続音声認識実験によって評価し，最後に 7 章で本論文をまとめる．

☆ 厳密には新聞テキストにおける N-gram カウント値．
☆☆ 実装上は，単語と単語列間の関連性となる．

2. MAP 推定による言語モデルのドメイン適応

本論文では，言語モデルのドメイン適応のベース手法として，MAP 推定により少量の目的ドメインのテキストと多量の独立テキストを混合する方法を用いる[6)]．以下，特定のドメインに依存しない大量テキストを独立テキスト，目的ドメインの少量テキストを適応テキスト，評価に用いる特定ドメインのテキストを評価テキストと呼ぶことにする．また，独立テキストのみによって構築された言語モデルを独立モデル，適応テキスト（および独立テキスト）を用いて特定ドメインへ適応されたモデルを適応モデルと呼ぶ．

y を観測データ，θ を推測データ（モデルパラメータ）とすると，ベイズの定理より，

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \tag{1}$$

となる．MAP 推定では，事後確率 $p(\theta|y)$ を最大化することで θ を推測する．θ に関する $p(\theta|y)$ の最大化を考える場合，$p(y)$ は一定であるから，結局式 (1) 右辺の分子の最大化に帰着される．すなわち，

$$\begin{aligned}\hat{\theta} &= \arg\max_{\theta} p(\theta|y) \\ &= \arg\max_{\theta} p(y|\theta)p(\theta)\end{aligned} \tag{2}$$

となる．N-gram 言語モデルの推定の場合，観測データ y は単語の生起であり，推測データ θ は N-gram 言語モデルパラメータ（N-gram 確率）となる．$p(\theta)$ は独立テキストから計算されるモデルパラメータの事前分布であり，$p(y|\theta)$ は適応テキストにて観測された y に対する尤度分布である．尤度分布関数としては，通常の N-gram モデルパラメータの最尤推定と同様，複数回試行のベルヌイ試行を与えることが多い．一方，事前分布 $p(\theta)$ は，事後分布と自然共役となるように分布を与えることが多く，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 ω を乗じたものととらえることができ，以下の式を得る．

$$p(y|\theta) \propto \theta^{\omega N^A_{hw}}(1-\theta)^{\omega(N^A_h - N^A_{hw})} \tag{3}$$

$$p(\theta) \propto \theta^{N^I_{hw}}(1-\theta)^{N^I_h - N^I_{hw}} \tag{4}$$

ただし w，h は単語および単語履歴であり，N^I_x，N^A_x 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) を用いて式 (2) を解くと，

$$p(w|h) = \arg\max_{\theta} p(y|\theta)p(\theta) \tag{5}$$

$$= \frac{N^I_{hw} + \omega N^A_{hw}}{\sum_w N^I_{hw} + \omega N^A_{hw}} \tag{6}$$

が導かれる．

3. 単語間の関連性を利用した言語モデル

3.1 単語関連性の定義とドメイン適応への応用

本節では，前節の MAP 推定法によるドメイン適応に，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．手順は以下のとおりである．

(1) ドメインに関係なく現れる単語のリスト（以下，ドメイン独立語リストと呼ぶ）を作る．これは，助詞・助動詞など，ドメインに共通して出現すると考えられる単語のセットであり，それらは適応処理において無視されることとなる．

(2) 独立テキスト T^I を，$T^I = T^I_1 T^I_2 \cdots T^I_{n_I}$ に分割する．ただし分割された各 T^I_k は，各々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．筆者らの先行研究[11]によれば，記事，段落，文の 3 段階の分割粒度について実験的に検討したところ，文分割が最も良い性能を示したので，ここでも，文単位で分割することで T^I_k を定義した．

(3) 言語モデルの各語彙（単語）に対して，T^I_k 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）が同時に出現するときに 1 を返し，そうでないときに 0 を返す関数を $q^k_{v[hw]}$ とし，k に関して和をとり正規化する（**図 1** 参照）．なお，v と hw の出現順序に関する情報，v と hw 間のテキスト上の距離に関する情報，T^I_k 中に観測される v や hw の個数（複数回観測される場合もある）に関する情報は無視する形となっている．すなわち，w と hw 間の共起関係の定量化としては，比較的粗い実装を行っているといえる．

$$q^I_{v[hw]} = \frac{1}{n_I}\sum_{k=1}^{n_I} q^k_{v[hw]} \tag{7}$$
$$(0 \le q^I_{v[hw]} \le 1)$$

上式は適応テキスト（T^A_k）に対しても算出可能であり，これを $q^A_{v[hw]}$ とする．$q^I_{v[hw]}$，$q^A_{v[hw]}$ 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^X_k 内のみで $q^k_{v[hw]}$ 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．

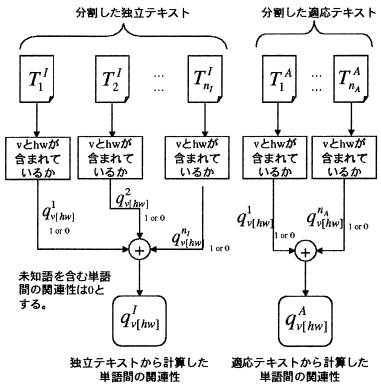

図 1 単語間の関連性の計算方法
Fig. 1 Calculation of the inter-word correlation.

(4) 適応テキスト T^A 内で，ドメイン独立語リストに含まれていないすべての単語 v^A について☆，$q^X_{v^A[hw]}$ の和を計算する．すなわち，

$$Q^X_{hw} = \frac{\sum_{v^A \in T^A} C(v^A) q^X_{v^A[hw]}}{\sum_{v^A \in T^A} C(v^A)} \tag{8}$$
$$(0 \le Q^X_{hw} \le 1)$$

ただし，$C(v^A)$ は T^A 中に単語 v^A が出現する回数とする．この Q^X_{hw} は，適応テキスト T^A 中の単語 v^A を trigger word として，単語列 hw がどのくらい想起されるか，に相当する量を表している（**図 2** 参照）．

(5) Q^I_{hw} と Q^A_{hw} の重み付き和をとる．

$$Q_{hw} = (1-\lambda)Q^I_{hw} + \lambda Q^A_{hw} \tag{9}$$
$$(0 \le Q_{hw} \le 1)$$

独立テキストと適応テキストのドメインが離れている場合，筆者らの一部が提案した独立テキストのみから単語間の関連性を計算する手法[11]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を補正することが可能となる．λ は Q^I_{hw} と Q^A_{hw} 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

☆ hw にはドメイン独立語リストに登録されいる語が含まれていてもよい．

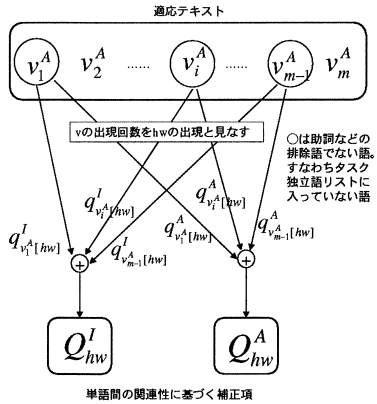

図 2 単語間の関連性に基づく補正項の算出
Fig. 2 Adjustment of N-gram count with the inter-word correlation.

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．

(6) MAP 推定によるドメイン適応の式 (6) に，式 (9) の単語間の関連性に基づく補正項 Q_{hw} を組み込み，適応単語遷移確率を次式で計算する．

$$P(w|h) = \frac{N^I_{hw} + \omega N^A_{hw} + \alpha Q_{hw}}{\sum_w N^I_{hw} + \omega N^A_{hw} + \alpha Q_{hw}} \tag{10}$$

ただし α は Q_{hw} の重み係数である．すなわち，式 (8) の単語 v^A の存在によって（v^A を trigger word として）単語列 hw があたかも出現したかのように取り扱うわけである．

3.2 本手法の特徴と先行研究との差異

提案手法の大きな特徴は，あらかじめ（独立テキストと適応テキストを用いて）計算しておいた単語間関連性と適応テキスト中の単語を参照し，MAP 推定時の N-gram カウント値を補正するところにある．しかも単語間の関連性は，bigram や trigram よりも広いスパンでの単語間の共起情報を扱っている．さらに，対象とする単語（と単語列）の出現順序，v と hw 間のテキスト上での距離，T^I_k 中に観測される v や hw の個数に関する情報は無視するなど，共起関係の定量化としては比較的粗い実装を行っている．また，v と hw の共起のみに着眼しているため，当然，v と hw 間の単語列（v と hw は含まない）を無視する形となる．この粗い共起関係のモデル化は,「適応テキストに直接現れていないが，潜在的に現れるであろう情報」を組み込むことを意図している．

このような単語間の関連性に着眼した言語モデリング手法として trigger モデルをあげることができる．trigger モデルは,「単語間の関連性を使い，単語履歴中の離れた単語からの制約をも考慮して次単語を予測する手法」である．一方，提案した手法は,「あくまでも N-gram 言語モデルの枠組みにおいて，離れた単語も考慮して定義される単語間の関連性を事前に（適応時に）導入した」という点で大きな相違があるが,「単語間の関連性を使う」という立場から見れば両者は類似した情報を活用している．両者において使われる単語間関連性の相違点としては，提案手法では，単語間の関連性（厳密には単語と単語列間の関連性）の定義において，対象とする単語間の距離，出現順序，着目するコンテキスト中の出現回数などを無視し，いうなれば粗い定量化が行われている．さらに，trigger モデルでは，単語履歴中の離れた単語を trigger word として利用することを前提としているが，提案手法では trigger word を単語履歴ではなく，適応テキスト全体に求めている点も差異の 1 つである．trigger モデルと N-gram を線形補間して定義される言語モデルは，N-gram を「さらに離れた単語履歴情報に対して適応した」モデルと考えることができる．先行研究における trigger モデルのドメイン適応[2]は上記の考えを直接利用したものであり,「trigger word が単語履歴中に，すなわち適応対象ドメインテキストに存在する」という事実に対してドメイン適応と位置づけている．以上のように提案手法は，従来提案された trigger モデル，および，trigger モデルによるドメイン適応と明確な差異を持つ手法として位置づけられる．

なお，trigger モデルを議論する場合，単語と単語との間の関連性をモデル化することが多いが，提案手法では単語と単語列との間の関連性を定義している．trigger モデルでも単語列を考慮することは可能であり，逆に，提案手法でも単語と単語に限定して関連性を議論することもできる．後者の場合，単語 v^A を trigger word として想起される単語 w をモデル化することとなり，最終的に 式 (10) における補正項が Q_w となり，これは，N-gram カウントを unigram ベースの特徴量で補正することと同値である．

3.3 期待される効果

上記した提案手法の特徴をまとめると，1）単語間の関連性および適応テキスト出現単語を参照し，想起され易い単語列 hw のカウント情報を補正する．2）テキスト情報を簡略化して用いることで，単語間の関連性を粗くモデル化している，などがあげられる．これらの特徴に直接対応して予想される効果としては，a）適応テキストにはたまたま出現しなかったが，評

表 1 実験で使用したテキストデータ概要
Table 1 Text data used in the experiments.

テキスト	ドメイン	文数	総形態素数	異形態素数
独立	新聞記事 2 年分	2,438,662	58,290,111	200,380
		133	2,156	639
適応	ピーターパン	526	8,931	1,738
		699	11,766	2,201
評価 A	ピーターパン	107	1,709	570
評価 B	マッチ売り	96	1,719	495

価テキストには潜在的に出現する可能性のある単語列の情報を扱うことができる（適応テキスト量の実質的増量効果），b）特定ドメインに対して適応した場合でも，そのドメイン周辺のドメインには有効に寄与することができる（ドメイン適応におけるの頑健性向上）があげられる．以下では，主にこの 2 つの効果について実験的に検証する．

4. 実験条件

各種実験について詳説する前に，共通して利用された実験条件についてまとめる．

4.1 テキストの準備

筆者らの一部によって行われた先行研究[11]では，独立テキストを新聞，適応テキスト・評価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，ドメイン独立語リストの設定などの基礎的な性質について報告が行われている．今回は，独立テキストを新聞とし，適応テキストのドメイン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．この場合，

- 新聞の表現と完全には一致しえないが，なるべく会話文に近い平易な表現で構成されたテキストを適応テキストとして選択する．
- 言語モデルのタスク適応は本来，N-gram 確率の補正と未知語に対する処理の両者を扱う必要があるが，ここでは前者が検討対象であるため，未知語率をできる限り抑えられるテキストを適応（および評価）テキストとして選択する．
  - 可能な限り，固有名詞など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．
  - 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，および評価テキストとして用意した．なお，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ドメインの評価テキストとして，H. Andersen の童話『マッチ売りの少女』日本語訳も用意した☆．なお，すべての実験において，適応テキストと評価テキストには，重複する文がないようにした．独立テキストとしては，毎日新聞 CD-ROM 95〜96 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『茶筌』ver2.0 [12]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 [13]を用い（Good-Turing discounting を使用），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，今回は上記語彙に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 節の要件に対しては，上記の語彙の設定はドメイン適応に十分なものであると考えられる．実際，各評価テキストに対する未知語率は 8%程度であった．

4.3 その他の設定条件

ドメイン独立語リストは，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[11),14)]．3.1 節，手順 (2) のテキスト分割方法は，予備実験から準最適な性能を与えると予想される文単位分割とした．

☆ 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．

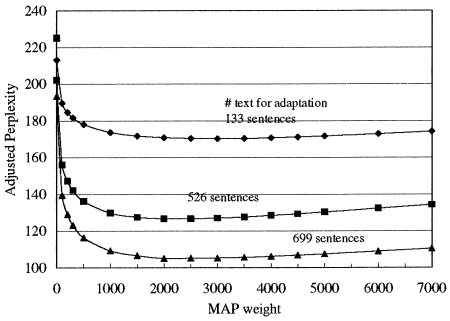

図 3 MAP 推定の重み ω と補正パープレキシティ（評価テキスト A）

Fig. 3 Adjusted perplexity as a function of MAP weighting factor ω (test text A).

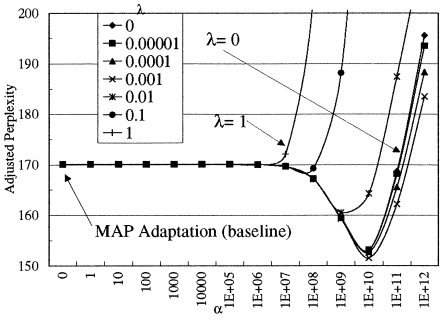

図 4 各種パラメータ値における補正パープレキシティ（適応テキスト 133 文，$\omega = 3000$，評価テキスト A）

Fig. 4 Reduction of adjusted perplexity as a function of α for various λ values (#adaptation sentences = 133, $\omega = 3000$, test text A).

5. 単語間の関連性を考慮したドメイン適応

5.1 適応ドメインと評価ドメインが同一の場合における評価実験

本節では，適応テキストと評価テキストが同一ドメイン（ピーターパン）である場合について，本手法のドメイン適応性能を検証する．この場合，パープレキシティの削減率のほかに，適応テキスト量の実質的増量効果の大きさが検証対象となる．まず 5.1.1 項で MAP 推定の重み最適値を求めたうえで，5.1.2 項で単語間の関連性を利用したドメイン適応の性能を検証する．

5.1.1 MAP 推定の重み定数の決定

適応テキストと評価テキストがともに『ピーターパン』の場合について，まず式 (6) の MAP 推定の重み ω の最適値を求めた．ω 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），ω 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 3,000，2,500，2,500 であった．なお，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．

5.1.2 単語間の関連性の利用とその効果

次に，ω 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ドメイン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**，**図 5**，**図 6** に示す．$\alpha = 0$ のときは，従来の最適化 MAP 推定から作成される言語モデルに相当する．いずれの図においても，$(\alpha, \lambda) = (10.0^{10}, 0.001)$ 付近で補正パープレキシティの最大削減率が実現されている．3.3 節で述べたように，本提案手法の効果の 1 つとして「適応テキスト量の実質的増量効果」が期待されている．一方図 3 において，3 種類の適応テキスト量におけるパープレキシティ削減の様子が示されている．この図より「図 4〜図 6 で示されたパープレキシティ削減が，何文程度の適応テキストの増量に対応するのか」を粗

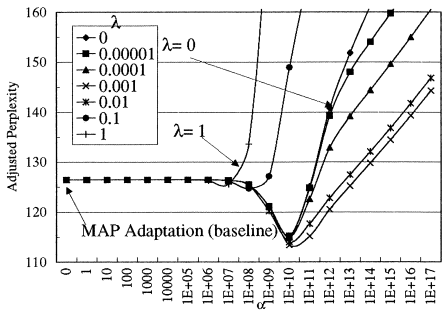

図 5 各種パラメータ値における補正パープレキシティ（適応テキスト 526 文，$\omega = 2500$，評価テキスト A）

Fig. 5 Reduction of adjusted perplexity as a function of α for various λ values (#adaptation sentences = 526, $\omega = 2500$, test text A).

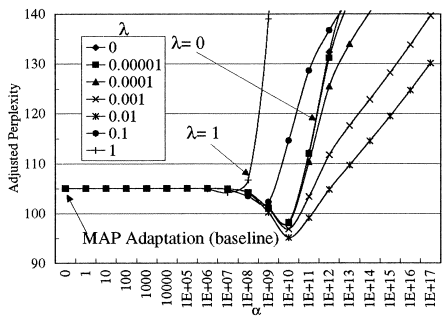

図 6 各種パラメータ値における補正パープレキシティ（適応テキスト 699 文，$\omega = 2500$，評価テキスト A）

Fig. 6 Reduction of adjusted perplexity as a function of α for various λ values (#adaptation sentences = 699, $\omega = 2500$, test text A).

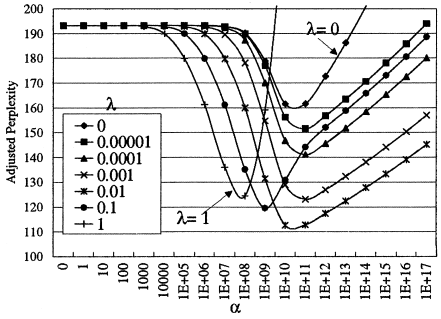

図 7 各種パラメータ値における補正パープレキシティ（適応テキスト 699 文，$\omega = 1$，評価テキスト A）

Fig. 7 Reduction of adjusted perplexity as a function of α for various λ values (#adaptation sentences = 699, $\omega = 1$, test text A).

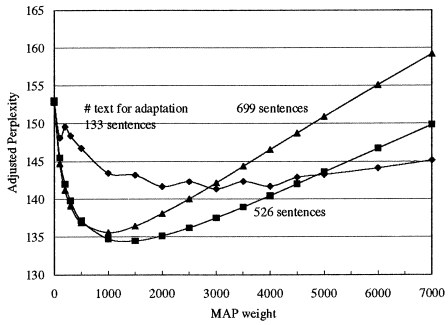

図 8 MAP 推定の重み ω と補正パープレキシティ（評価テキスト B）

Fig. 8 Adjusted perplexity as a function of MAP weighting factor ω (test text B).

く見積もることができる．その結果，図 4〜図 6 に対して各々150，70，40 文程度を新たに追加するのと同等の効果であることが分かった．また，パープレキシティの減少を率という観点から見た場合，適応テキスト量にはよらず，いずれの場合も約 10%の減少率が得られた．

さて，MAP 推定時の重み ω を，$\omega = 1$ とした場合の実験結果を**図 7** に示す（適応テキスト文数は 699）．図 4〜図 6 と同様に，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分かる．しかし ω を最適化した場合と異なり，$\lambda = 1$（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．λ 最適値の ω 依存性は次のように考えることができる．重み ω を適切に設定した場合，適応テキストに直接出現した単語の頻度情報を直接的に MAP 推定を用いて取り入れる一方，$\lambda < 1$ の条件下で補正することで，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．逆に $\omega = 1$，すなわち適応テキスト中の単語頻度情報が実質上組み込まれていない場合，提案手法のよる補正において，適応テキストにおける単語間関連性をより強く利用して単語頻度を間接的に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を行っている．

5.2 適応ドメインと評価ドメインが異なる場合における評価実験

本節では，適応テキストと評価テキストのドメイン内容が異なる場合について，評価テキストのドメインに対する本手法の頑健性を検証する．まず 5.2.1 項で MAP 推定の重み定数の最適値を求めたうえで，5.2.2 項で単語間の関連性を利用したドメイン適応の性能を検証する．

5.2.1 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 式 (6) の MAP 推定の重み ω の最適値を求めた．ω と補正パープレキシティの関係を**図 8** に示す．ω 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）ことが分かる．しかも，適応テキストが多いとパープレキシティの最小値はより小さくなるが，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が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テキストが違うドメインの場合は，適応テキストの量と ω 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．さらに，MAP 推定重み ω の増加にともなって補正パープレキシティは大きく増加していることも分かる．これは，適応テキスト『ピーターパン』と評価テキスト『マッチ売りの少女』間の言語的相違によるものと解釈される☆．本節では，このように評価テキストと適応テキスト間に言語的相違が見られる場合における提案手法の性能を検証する．

5.2.2 単語間の関連性の利用

ここでは，適応テキストが 526 文で，$\omega = 1000$ の場合と $\omega = 2500$ の場合の実験を行った．前者は 5.2.1 項の結果から MAP 推定の重みを適切に設定した場合であり，後者は 5.1.1 項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

☆ なお，評価テキストとして『裸の王様』を用いた場合は，『ピーターパン』による適応モデル（MAP 推定）により補正パープレシティが増加する結果となった．

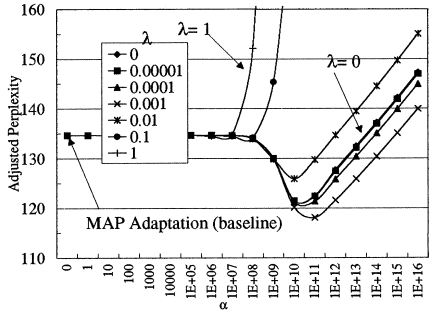

図 9 各種パラメータ値における補正パープレキシティ（適応テキスト 526 文，$\omega = 1000$，評価テキスト B）

Fig. 9 Reduction of adjusted perplexity as a function of α for various λ values (#adaptation sentences = 526, $\omega = 1000$, test text B).

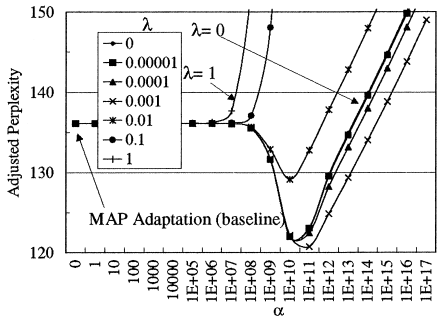

図 10 各種パラメータ値における補正パープレキシティ（適応テキスト 526 文，$\omega = 2500$，評価テキスト B）

Fig. 10 Reduction of adjusted perplexity as a function of α for various λ values (#adaptation sentences = 526, $\omega = 2500$, test text B).

ぞれ，λ および α を変化させた場合のグラフを**図 9**，**図 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ドメイン適応を利用すると，ω 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．また，5.1.2 項同様，$(\alpha, \lambda) = (10.0^{10}, 0.001)$ 付近で補正パープレキシティの最大削減率が実現され，削減率も約 10%である．以上より，本手法が評価テキストのドメイン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，本手法が適応データには直接出現しなかった単語列の情報を，種々のテキスト情報を簡略化する形で（粗く）定義された関連性に基づいて処理しているため，適応テキストのドメインが評価テキストのそれと多少異なっていても，ドメインがある程度近ければ評価テキストに出現する単語を予想できると考えられる．

5.3 本手法のパラメータについて

本提案手法では単語間の関連性をドメイン適応に利用する場合，新たにパラメータ α と λ を追加しており，この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．本論文においても，様々な α と λ の値における提案手法の性能を実験的に検討しているが，本論文に掲載した実験結果以外にも，独立テキスト，適応テキスト，ドメイン独立語リストなどのサイズを変化させたときの α と λ の最適値の変動の様子を予備実験的に検討している．これらの結果より，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分かっており，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．なお，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，最適な α の自動推定は可能であると考えている☆☆．

一方 λ は主にドメインの種類や MAP 推定の重みなどに依存することが分かっている．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．

6. 音声認識実験による評価

本章では，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 v3.1 [15]を，音響モデルとしては状態数 3,000，混合数 16 の状態共有 triphone を使用した．言語モデルには 2 万語 bigram を用い，1st pass による認識結果を用いて評価した．trigram を用いた 2nd pass の結果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 $\omega = 2500, 3000$ で適用）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（最適パラメータを使用，$\alpha = 10^{10}$，$\lambda = 0.001$，$\omega = 2500, 3000$）の 3 者を比較した．ここで，男性話

☆ なお，MAP 推定のみでは補正パープレキシティが増加した『裸の王様』に対しても，提案手法では補正パープレキシティの削減を実現することができた．

☆☆ 本論文で掲載した種々の実験においても，適応テキストサイズによらず，およそ $\alpha = 1.0^{10}$ が最適値となっている．

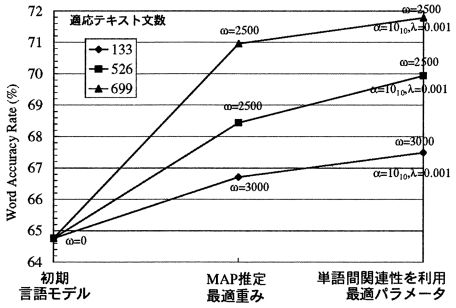

図 11 単語正解精度（評価テキスト A）
Fig. 11 Word accuracy with test text A.

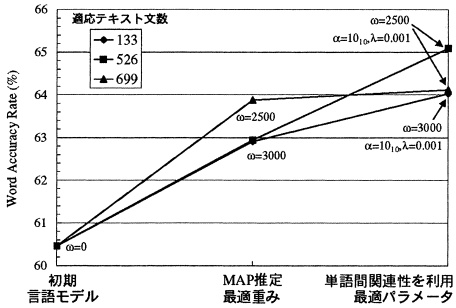

図 12 単語正解精度（評価テキスト B）
Fig. 12 Word accuracy with test text B.

表 2 認識実験による提案手法の評価 [%]（適応テキスト＝526 文）
Table 2 Evaluation of the proposed method through speech recognition experiments [%] (#adaptation sentences = 526).

	評価テキスト A		評価テキスト B	
	MAP	関連性	MAP	関連性
単語正解率	73.7	75.1	67.9	69.8
単語正解精度	68.4	69.9	62.9	65.1
置換誤り率	22.9	21.8	27.3	25.8
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

者 1 名の評価テキスト読み上げ音声について，insertion penalty を −2 に固定したうえで，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誤り削減率が数%上昇することが確認された．

7. まとめと今後の課題

本論文では，単語間の関連性を利用し，N-gram カウントを補正することによって適応テキストの量にかかわらず，安定してドメイン適応の効果を従来法より高める（早める）手法を提案し，実験を通してその効果を実証した（適応テキスト量の実質的増量効果）．また，単語間の関連性を種々のテキスト情報を簡略化する形で（粗く）定義しており，その結果，適応テキストと評価テキストが似ているが異なるドメインに属する場合でも（たとえば目的ドメインのテキストが存在せず，近隣ドメインのテキストのみ存在する場合など），提案手法が有効に寄与することを実験的に検証した（ドメイン適応におけるの頑健性向上）．すなわち提案手法は，対象ドメインの言語表現を従来法と比較してよりコンパクトに表現する一方で，従来法では効果が薄れていた（ときとして効果が観測されない）周辺ドメインに対しても効果的に作用するという，従来の枠組みでは相矛盾する問題として位置づけられていた課題を解決する方法論を提供したことになる．

しかしながら，本提案手法には，以下に示すような課題が残されている．

- 単語間関連性の定義と関連性の利用方法との関係
  本論文では，単語関連性におけるテキスト依存性を抑えることで，適応テキスト量の実質的増量効果，およびドメイン適応におけるの頑健性向上といった効果を示すことができた．この関連性定義とその利用方法との組合せは変更可能であり，たとえば本論文で定義した関連性を用いて先行研究で提案された trigger モデルの適応処理を実装することもできる．比較実験を通して適切な組合せを追求する必要があろう．また本論文では，関連性値を返す関数として（たとえば $q^k_{v[hw]}$ など）単純な 2 値関数を使い，抽象度の高い関連性定義を行っているが，この関連性定義に対しても最適化を検討する必要がある．
- 各種パラメータの最適化，自動設定
  5.3 節で述べたように各種パラメータの自動推定/最適化や削減などについて検討する．
- trigram 化
  本論文では bigram を対象としたが，連続音声認識システムに実際に組み込むには trigram 化することが必須であるため，提案した枠組みにおける trigram への適用を検討する．

謝辞 本研究に対し数々のアドバイスを賜りました Hui Jiang 氏（Lucent Technologies）および佐々木耕樹氏（富士通）に感謝いたします．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 SugitaGenpaku の皆様に感謝いたします．

参 考 文 献

1) 北　研二：統計的言語モデル，東京大学出版会 (1999).
2) Rosenfeld, R.: A maximum entropy approach to adaptive statistical language modeling, *Computer Speech and Language*, Vol.10, No.3, pp.155–186 (1996).
3) Fedelico, M.: Bayesian estimation methods for N-gram language model adaptation, *Proc. ICSLP-96*, pp.240–243 (1996).
4) 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).
5) Matsunaga, S., Yamada, T. and Shikano, K.: Task adaptation in stochastic language models for continuous speech recognition, *Proc. ICASSP'92*, Vol.1, pp.165–168 (1992).
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) Moriya, T., Hirose, K., Minematsu, N. and Jiang, H.: Enhanced MAP adaptation of N-gram language models using indirect correlation of distant words, *CDROM of ASRU'2001* (2001).
8) Rosenfeld, R.: Adaptive statistical language modeling: A Maximum entropy approach, Ph.D. Thesis, School of Computer Science, Carnegie Mellon University.
9) Tillmann, C. and Ney, H.: Selection criteria for word trigger pairs in language modeling, *Grammatical Inference: Learning Syntax from Sentences*, Miclet, L. and de la Higuera, C. (Eds.), pp.95–106, Springer, Lecture Notes in Artificial Intelligence 1147.
10) Ueberla, J.: 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).
11) Sasaki, K., Jiang, H. and Hirose, K.: Rapid adaptation of N-gram language models using inter-word correlation for speech recognition, *Proc. ICSLP-2000*, Vol.4, pp.508–511 (2000).
12) 松本裕治ほか：日本語形態素解析システム『茶筌』version 2.0 (1999).
13) Clarkson, P.: The CMU-Cambridge Statistical Language Modeling Toolkit v2 (1997).
14) Kawahara, T. and Doshita, S.: Topic independent languange model for key-phrase detection and verification, *Proc. ICASSP-1999*, pp.685–688 (1999).
15) 河原達也，李　晃伸，小林哲則，武田一哉，峯松信明，嵯峨山茂樹，伊藤克亘，伊藤彰則，山本幹雄，山田　篤，宇津呂武仁，鹿野清宏：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).

(平成 13 年 11 月 16 日受付)
(平成 14 年 4 月 16 日採録)

広瀬　啓吉（正会員）

昭和 24 年生．昭和 52 年東京大学大学院博士課程修了．工学博士．同年東京大学工学部電気工学科講師．昭和 62 年米国 MIT 客員研究員．平成 6 年東京大学工学部電子工学科教授．平成 8 年同大学大学院工学系研究科電子情報工学専攻教授．平成 11 年より同大学院新領域創成科学研究科基盤情報学専攻教授．音声言語情報処理分野一般についての教育研究開発，特に韻律に着目した研究に従事．IEEE，米国音響学会，ISCA，日本音響学会，電子情報通信学会，人工知能学会，言語処理学会等各会員．

峯松　信明（正会員）

昭和 41 年生．平成 7 年東京大学大学院工学系研究科電子工学専攻博士課程修了．博士（工学）．同年豊橋技術科学大学情報工学系助手．平成 12 年東京大学大学院工学系研究科助教授，平成 13 年同大学院情報理工学系研究科助教授．平成 14 年瑞国 KTH 客員研究員．音声認識，音声分析，音声応用，音声知覚，および音声合成の研究に従事．電子情報通信学会，日本音響学会，日本音声学会，人工知能学会各会員．

森谷　高明（正会員）

昭和 52 年生．平成 13 年東京大学工学部電子情報工学科卒業．現在，同大学大学院新領域創成科学研究科在籍．音声認識用の言語モデル，無線アドホックネットワークに関する研究に従事．日本音響学会会員．