RE-004

話題依存言語モデル構築のためのLSAと単語発音情報を用いた語彙推定
Vocabulary Estimation Using LSA and Word Pronunciation for Topic-dependent Language Modeling

藤原 裕幸 † 西崎 博光 ‡ 関口 芳廣 ‡
Hiroyuki Fujiwara Hiromitsu Nishizaki Yoshihiro Sekiguchi

1. はじめに

近年の音声認識技術の発達により，ニュース等の原稿読み上げ（朗読）音声は高精度に認識することができ，自動字幕付与などに応用されている．最近ではこの音声認識技術を講義・講演の書き起こしや会議の議事録自動作成等に利用するための研究が進められている．

しかし，講義・講演や会議等では，言い淀み・言い直しが多く含まれる話し言葉が使われることが多く，現在の技術では精度良く認識することができないのが現状である．その大きな理由として，事前に認識対象音声の話題を特定できないこと，特定できたとしても話題に適した言語モデルを構築することがコスト等の面で困難であること，認識対象の語彙集合を適切に選択できないこと，が挙げられる．

認識対象音声の話題に合った言語モデルを作成するためには，事前に話題を調査し，その話題のテキストを大量に収集しなければならない．ある一定量以上の認識対象の話題にあった学習コーパスが存在しなければ，精度良いモデルを作成することができず，音声認識率も向上しない．

ニュース音声認識の場合は，例え話題が異なっていてもその時事性や同報性を利用することで，Web上のニュース記事からドメインの合った大量の学習コーパスを用いて適切な言語モデルを学習できる [1]．筆者らが研究対象にしている日本語の講義音声認識では，話し言葉という点でドメインが適合している「日本語話し言葉コーパス」(Corpus of Spontaneous Japanese，以下"CSJ"と記す)[2] により学習した言語モデルを使用する方法がある．しかし，これらは話し言葉という点だけは同じであるが，講義中で使われる語彙や言い回しはCSJのものとは全く異なるものである．

我々は，以前，Webを用いた言語モデルの話題適応化についての報告を行った [3]．この報告では，講義音声を認識する際，講義の話題に合ったWebページを利用することで，その講義音声に適した言語モデル・認識辞書語彙の学習を行い，その有効性を確かめている．また，単純な処理ながらモデル化される語彙を推定する手法の検討も行っており，語彙推定の可能性を見出した．

音声認識精度は言語モデルのパープレキシティや未知語率に大きく影響される．語彙サイズが異なる2つの言語モデルがあった場合，同じ未知語率，同じテストセットパープレキシティであるならば，サイズが小さい言語モデルを利用する方が音声認識率は良くなるはずである．また，モデルがコンパクトである方が，認識時間も短縮され，メモリ制限の問題がシビアな組込みシステムなどにも導入しやすい．

そこで，本研究では，話題適応化によって言語モデル(認識辞書)の語彙数を増加させてモデルのパラメータを増加させるのではなく，よりコンパクトな（パラメータが少ない）話題適応化言語モデルを構築するための語彙推定手法を提案する．Webドキュメントで話題適応化した言語モデルに，提案する語彙推定を適用することで，話題に適応し，かつコンパクトな言語モデルが構築できる．語彙推定には，潜在的意味解析 (Latent Semantic Analysis, LSA) と単語の発音情報を利用する．

認識実験の結果，言語モデルの適応化および語彙推定手法により，講義音声の単語正解率，正解精度をそれぞれ最大で5.2%，7.2%，キーワードになりやすい名詞のみに着目した名詞正解率を最大で14.1%改善することができた．また，語彙推定処理により言語モデルが小さくなり，認識処理時間の短縮も実現できている．

2. 関連研究

言語モデルの適応化を目的とした研究は多く存在する．ここでは講義・講演等の話し言葉音声認識を対象とした言語モデル適応化について紹介する．

講義認識用の言語モデルにそれ以前の講義音声の書き起こしデータを用いる方法がある [4]．これは，講義の連続性(多くの大学で1科目あたり15コマの授業が連続的に開講されている)を利用し，以前の講義音声の書き起こしを用いて適応化を行っている．しかし，講義の書き起こしの作成は非常に高コストであり，現実問題として書き起こしを用意することは難しい．

そこで，授業で用いる教科書や講義で使用したパワーポイント等の電子スライド情報を利用する方法が提案されている [5] [6]．

MITの研究グループは [5]，講義で使われたテキストやSwitchboardコーパスを用いて言語モデルを適応化することで，講義音声の認識率改善を実現している．日本では，根本ら [6] が講義で使用されたスライドを用いて言語モデルを適応化する手法を提案し，その有効性を示している．

しかし，これらの手法はスライドを利用している講義

†山梨大学大学院医学工学総合教育部，Educational Interdisciplinary Graduate School of Medicine and Engineering
‡山梨大学大学院医学工学総合研究部，Research Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi

音声を認識する場合のみに利用できる．現在でも講師の多くは黒板を用いた講義を実施しており，この場合は当然スライド情報を用いることができない．そこで，小暮らは [7]，大学では学生向け (電子) シラバスが用意されていることに着目し，これを利用することで言語モデル適応化を行うためのドキュメントを収集する方法を提案している．この方法では，講義の前に言語モデルを適応化することができるので，リアルタイムで講義音声を比較的精度よく認識することが可能となる．リアルタイムで認識する際は，話題に適応化された言語モデルを利用し，かつ言語モデルがコンパクトである方が認識処理速度も高速になる．本研究の手法が目指す言語モデルは，このような使い方を想定している．

一方で，授業シラバスのような事前情報が利用できない状況を考えた場合，Web を利用することが有効である [8]．梶原らは [8]，Web ドキュメントを用いた講演音声認識のための反復適応化手法を提案している．

また，確率的 LSA(Probabilistic LSA) を用いた言語モデル適応の研究も行われている．秋田ら [9] は，音声認識結果に PLSA を施し，話題に適応した単語の unigram 確率を求めることで，この確率に基づくスケーリングにより言語モデルを適応化している．栗山ら [10] は，PLSA により語彙を話題語，汎用語といったクラスに分割し，クラスごとに別々に話題や文型 (発話スタイル) を適応することで，従来の PLSA モデルに比べてパープレキシティが大きく下がったことを報告している．

これらのように，様々な適応化手法が提案されているが，これらは適応化によりモデルサイズが増加する．そこで本研究の目的は，話題に適応化できておりかつコンパクトな言語モデルの構築を目指している．コンパクトな言語モデルを構築するための手法は，踊堂ら [11] や Stolck[12] が提案している．これらの手法は，エントロピー等の指標により N-gram パラメータ数の削減を図っている．

これらの手法と異なり，提案手法は，言語モデルに登録する語彙を推定し，不要な語彙の登録を制限する．これによりモデルのパラメータ数を大幅に削減し，コンパクトな適応化言語モデルの構築を目指す．

3. Web を利用した言語モデル適応化

本研究では，言語モデル適応化に徳田ら [3] の Web を用いた適応化手法を採用する．本節では，その手法について簡単に説明する．

3.1 適応化概要

Web ドキュメントを利用した言語モデル適応化処理の概要を図 1 に示す．処理の大まかな流れは以下の通りである．

step(1): まず，**LVCSR-1** において，認識対象音声を CSJ 講演集合から学習した言語モデル (これを”CSJ 基本言語モデル”と記す)[4] を用いて認識する．これをここでは”**基本認識結果**”と呼ぶ．

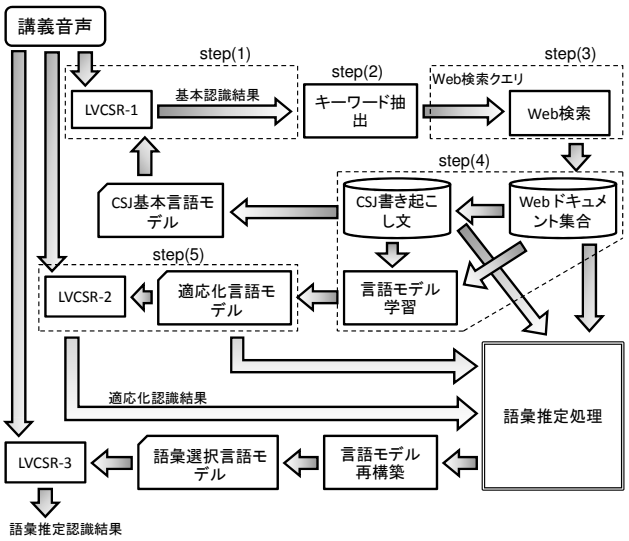


図 1: 言語モデルの適応化処理の概要

step(2): 基本認識結果から対象音声の話題を適切に表すキーワードを抽出し，Web 検索クエリを構成する．

step(3): Web 検索クエリを用いて Web ドキュメントを収集し，不要なタグ等を除去する．また，Microsoft 社のパワーポイント形式や PDF 形式の場合はテキストデータに変換して利用する．

step(4): 収集した Web ドキュメントと CSJ 講演集合を用いて言語モデルを学習 (適応化) する．学習されたモデルを**適応化言語モデル**と記す．

step(5): LVCSR-2 において，認識対象の講義音声を適応化言語モデルを用いて認識する．この出力を**適応化認識結果**と呼ぶ．

なお，図 1 には，認識に不必要な語彙を削除するための語彙推定処理が含まれているが，本節では，**LVCSR-2**”結果が得られるまでの処理，すなわち，Web ドキュメントによる言語モデル適応化とそれを用いた音声認識実験についてのみ述べるに留める．語彙推定処理についての詳細と音声認識実験に関しては第 4. 節で述べる．

3.2 Web 検索用クエリの構成

Web 検索用のキーワード[5] には認識対象の講義名と図 1 の仮認識結果から抽出した単語を用いる．本研究ではキーワードには名詞のみを採用する．

- **講義名の名詞**
  講義名に含まれる名詞でおおよその話題を特定する．

[4] CSJ に出現している単語や言い回しは，ある程度広い話題や発話スタイルををカバーしていると仮定している．

[5] 本研究では，単語 (複合語)1 つ 1 つをキーワード，検索エンジンに入力するキーワードの集合をクエリと呼ぶ．

表 1: 名詞 bigram の例 (科目「プログラミング入門」)

bigram	頻度	N/N_{max}
ポイント␣クラス	10	1.00
プロダクト␣クラス	6	0.6
単語␣クラス	5	0.5
明示␣的	4	0.4
関連␣性	2	0.2
間接␣的	2	0.2
…	…	…

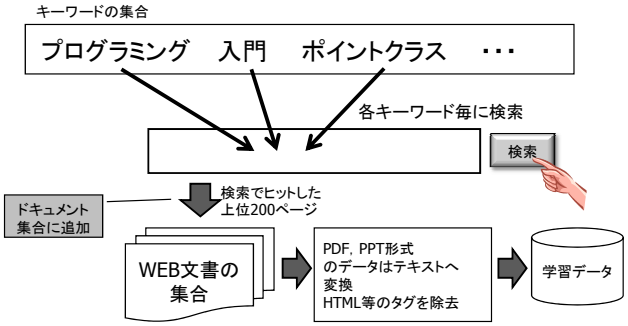


図 2: Web ドキュメント収集

例えば「プログラミング入門」という講義であれば「プログラミング」「入門」をキーワードとして利用する．

- **名詞 bigram**
  仮認識結果から抽出した 2 個組み名詞を用いる．例えば「ポイント」＋「クラス」⇒「ポイントクラス」．

名詞 bigram の抽出方法について説明する．認識結果に出現しているある bigram の頻度を N とする．すべての bigram の中で最大頻度のものを N_{max} とすると，$N/N_{max} >= 0.30$ の条件を満たす全ての bigram を検索用キーワードとして抽出する．表 1 に名詞 bigram の例を示す．表 1 では意図的に 2 単語を分けて掲載しているが，実際に Web 検索で用いる場合は 1 単語として扱っている．

抽出した検索用キーワード群から検索用クエリを構成し Web ドキュメントを収集するが，これには次の方法を用いた．名詞 bigram に基づく 1 つのキーワードを 1 クエリとして Web 検索する．もしキーワードが K 個あった場合，K 回の Web 検索を行い，各キーワードに対して最大 200 ページ収集する (図 2 参照)．検索エンジンには，”Yahoo! 検索 Web API”を利用した．

3.3 適応化処理

言語モデルの適応化には，Palmkit [6] 付属のツールを利用した．これは，2 つの言語モデルに含まれる単語 N-gram の出現確率を重み付きで補間するのではなく，図 3 に示すように 2 つの学習テキストから求めたそれぞれの N-gram 頻度を重み付きで混合することにより適応化を行う．

[6] *http://palmkit.sourceforge.net/*

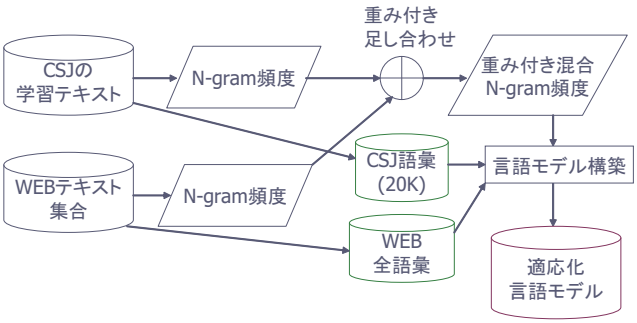


図 3: 言語モデルの適応化処理

表 2: 認識対象音声の仕様

科目名	発話時間 [分]	単語数 [個]*
プログラミング入門	57.6	13,486
線形代数学 I	28.8	5,856
数値計算法	20.8	5,049
物理学実験	24.1	5,383

全員男性講師．
音声はすべてピンマイク（SONY ECM-T145）で収録．
* 講義全体の音声は，文献 [13] の方法で小さい発話単位に自動分割されている．

今回の実験では，CSJ から学習した基本言語モデルに対し，インターネット上から収集した Web ドキュメントを混合し適応化する．この混合比にはいろいろなパターンが考えられるが，本研究では単純に (CSJ:Web)=1:1 とした．

3.4 音声認識実験

3.4.1 実験条件

適応化した言語モデルの効果を調査するため講義音声に対する認識実験を行った．今回の実験では，純粋に言語モデルの適応化効果のみを調査するため，音響モデルの話者適応化は行っていない．

使用する音声特徴パラメータは，16bit 量子化，16kHz サンプリングされた音声に 25ms 窓長のハミング窓を掛け，12 次元の MFCC, Δ MFCC, Δ Δ MFCC, および 1 次元の Δ pow, Δ Δ pow の 38 次元の特徴を抽出したものである．音響モデルには CSJ に収録されている 797 講演の男性話者から学習した総状態数約 3,000 の 32 混合 triphone を用いた．

適応化前の元の言語モデルは CSJ に含まれる 3,300 講演 (学会講演・模擬講演・読み上げ・対話などすべてを含む，約 123M バイト) から学習した語彙数 20,000 の単語 trigram である．

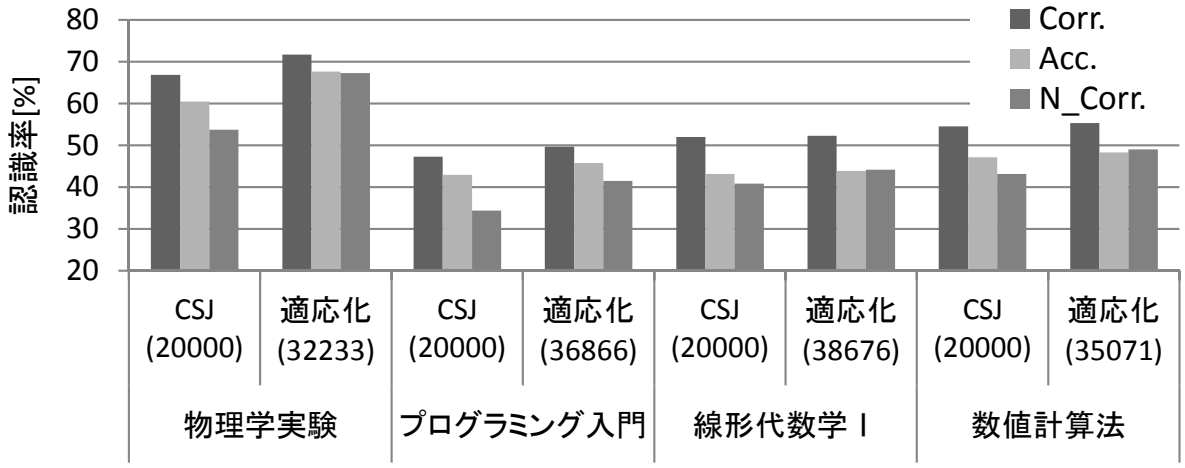


図 4: 適応化言語モデルを用いた場合と用いない場合の音声認識率の比較 (カッコ内数値は語彙サイズ)

音声認識デコーダには大語彙連続音声認識システム Julius-4.1.2[7] を用いた．

認識対象の音声には，山梨大学工学部コンピュータ・メディア工学科コンピュータサイエンスコースで開講された 4 講義を用いた．これらの講義に対する詳細な情報を表 2 に示す．表 2 に示してある通り，講義音声の録音にはピンマイクを用いているため [14]，ハンドマイクやヘッドセットマイクを用いて収録した音声よりも比較的認識が難しい．また，黒板の講義 (線形代数学 I，数値計算法) ではノイズが多く含まれており，認識精度を下げる要因になっている．

3.4.2 音声認識実験結果

各講義毎の音声認識実験結果を図 4 に示す．図 4 において”**CSJ**”は適応化前の言語モデルを使用したときの認識率，”**適応化**”が適応化した言語モデルを利用したときの認識率である．カッコ内の数値は認識辞書の語彙数である．評価には，単語正解率 (Corr.)，単語正解精度 (Acc.)，名詞正解率 (N_Corr.) の 3 つの指標を用いる．要約やインデキシングなど講義コンテンツ処理をする場合には，内容語，特に専門用語を含む名詞の認識が重要であるため，本研究では名詞正解率を評価指標の一つとして採用している．

また，適応化を行った場合と行わない場合の言語モデルごとのテストセットパープレキシティ[8] および未知語率を表 3 に示す．

図 4 より，話題適応化を行うことで，全ての講義音声において単語正解率，単語正解精度，名詞正解精度が改善されていることが分かる．講義によっては最大で単語正解率が 4.9%，単語正解精度 7.2%，名詞正解率が 13.6%改善されている．また，表 3 に示してある通り，話題適応化を行うことで未知語率やテストセットパープレキシティも大きく改善されている．

表 3: 話題適応化を行った場合と行わない場合の未知語率 (OOV [%]) と補正パープレキシティ (APP) の比較

科目名		言語モデル	
		CSJ	適応化
物理学実験	OOV	7.03	1.99
	APP	331.7	190.8
プログラミング入門	OOV	7.20	2.32
	APP	491.1	334.2
線形代数学 I	OOV	4.33	1.78
	APP	276.9	198.8
数値計算法	OOV	3.84	0.76
	APP	262.6	179.0

このことから，Web ドキュメントを用いた言語モデルの話題適応化は，講義音声認識率を改善する上で効果があることが分かった．しかし，図 4 に示してあるように，語彙数がベースラインに比べて大変大きくなっている．

そこで，次節では，よりコンパクトな言語モデル構築のための LSA と単語発音情報を利用した語彙推定手法を提案する．

4. 語彙推定

本研究ではモデルサイズを小さくする試みを行っているが，語彙推定により，認識性能を改善することもできる．本節では，モデルサイズ削減による認識率と認識速度の改善ついて説明する．

[7] *http://julius.sourceforge.jp/*

[8] 通常のパープレキシティを未知語数で補正した補正パープレキシティとなっている [15].

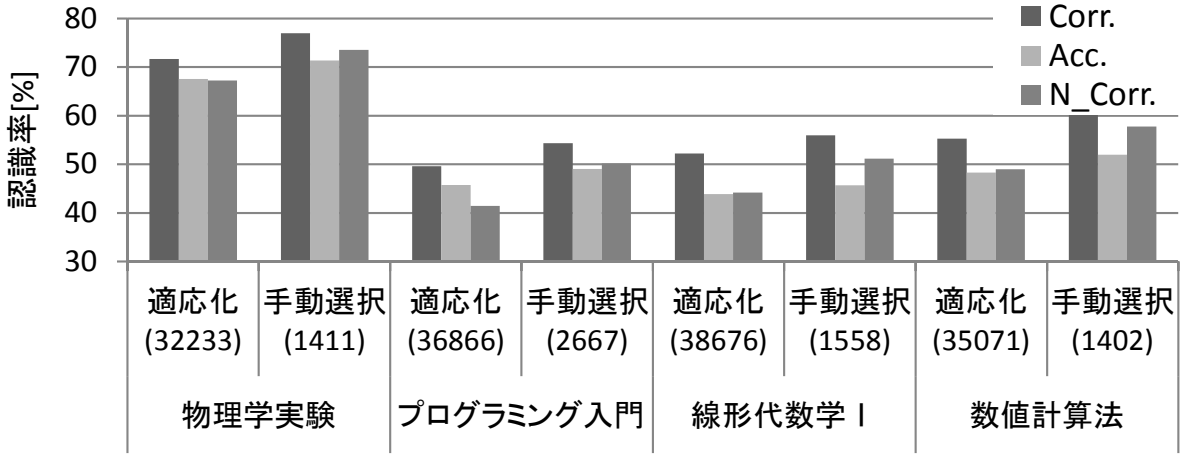


図 5: 語彙選択の有効性を表すグラフ (カッコ内数値は語彙サイズ)

4.1 語彙推定による認識率改善の効果

音声認識性能を向上させるためには，認識辞書に登録する語彙数を増やすことでテストセットに対する未知語率を下げると良い．しかし，辞書の語彙サイズが多い場合と少ない場合とで，テストセットに対して同じ未知語率であるならば，語彙サイズが小さい辞書の方が音声認識率は良くなるはずである．

特に，講義のように特定の話題に関する発話を認識する場合，必要な語彙はある程度の量に限定される[9]．むやみに語彙を増やしても未知語率を下げることは難しく，逆に講義に関係のない余計な語彙を辞書に組み入れることによって認識率の悪化が予想できる．

本研究では，これまでに述べてきた適応化手法により未知語率を下げ，さらに認識に必要な語彙のみを辞書に登録する語彙推定処理を行うことで (つまり，必要な語彙だけを選択し，認識に必要でないと判断した語彙は辞書から除去する)，講義音声の認識率を向上させることを試みた．

まず，手動で認識辞書に登録する語彙を選択した場合，どれだけ認識率が改善するのかを調べた．図 5 に実験結果を示す．**”適応化”**は，Web ドキュメントで適応化を行った結果で，図 4 のそれと同じである．**”手動選択”**が手動で語彙を制限した結果となっている．この実験では，テストセットの各講義音声の書き起こし文に出現した単語以外のすべての単語を認識辞書から削除し，CSJ と Web ドキュメントを用いて言語モデルを再学習している．従って，未知語率は表 3 の適応化と同じである．図中のカッコ内の数値は語彙サイズを表している．

図 5 をみると，すべての講義において単語正解率，単語正解精度，名詞正解率が大きく改善している．このことから，語彙を適切に選択することにより認識率の向上が期待できる．

[9] 90 分の講義で使用される語彙の種類数は 1,000 程度

4.2 LSA

語彙推定には，LSA に基づく単語クラスタリング手法を用いる．LSA は単語のクラスタリングを始め，情報検索やドキュメントクラスタリングなど，幅広い分野で利用され，その有効性が示されている．

LSA はドキュメント群とそれに含まれる単語群において，それらに関連した概念集合を生成することで，その関係を分析する [16]．LSA によってクラスタリングされた単語集合はそれぞれが意味的に似ているという性質を持つ．そこで生成された単語集合を適切に選択して用いることで，話題に沿った単語だけを選択することができるはずである．

4.2.1 単語×ドキュメント行列

LSA では各ドキュメントにおける単語の出現を表した単語×ドキュメント行列が利用される．このような行列のことを共起行列という．各成分の重み付けには $tf \cdot idf$ が用いられるのが一般的である．tf，idf には様々な式が提案されているが，本研究では次の式を用いる [16]．

$$tf_{ij} = log_2(f_{ij} + 1) \quad (1)$$

$$idf_j = 1 + \frac{\sum_{j=1}^{N} \frac{f_{ij}}{F_i} log_2 \frac{f_{ij}}{F_i}}{log_2 n} \quad (2)$$

ここで，n はドキュメント数，f_{ij} は単語 w_i のドキュメント D_j における出現頻度，F_i はドキュメント集合全体を通しての単語 w_i の出現頻度を表す．

これらの方法で重み付けを行い，単語×ドキュメント行列を作成する．

4.2.2 特異値分解

LSA では共起行列に対して特異値分解 (Singular Value Decomposition，SVD) を行い，$m \times n$ 行列を次のよ

うな 3 つの行列 U, S, V^T に分解する (ただし，$l = min(m, n), U : m \times l, S : l \times l, V : n \times l$).

$$X = USV^T \tag{3}$$

ここで U と V^T は直行行列であり，S は対角行列である．行列 U および行列 V^T を左特異ベクトル，右特異ベクトルと呼び，S を行列 X の特異値という．

このとき，式 (3) の結果から単語の相関を求める．単語の相関を求める行列積は次のように展開される．

$$\begin{aligned} XX^T &= (USV^T)(USV^T)^T \\ &= US^2U^T \\ &= (US)(US)^T \end{aligned} \tag{4}$$

このようにして得られたのが特異値分解の結果である．なお，この行列には「上位 r 個の特異値で圧縮した結果は，行列の $rank(r)$ における最小 2 乗誤差になる」という性質がある．そのため，誤差を最小に保ったままの行列の $rank$ をあえて削除することで，より関連の強い単語ベクトルが同一次元に縮退され，類似した値に近似されることが期待できる．

4.3 LSA を用いた語彙推定

LSA による語彙推定手法を図 6(上半分) に示す．

LSA による語彙推定では，まず Web ドキュメント集合と CSJ 講演集合のそれぞれに対して LSA を行い，単語をクラスタリングする．クラスタリングされた単語集合から適切な単語集合を選択することで語彙の絞り込みを行う．今回の実験では，特異値分解で得られた特異値を 200 次元に圧縮し，クラスタリング手法には K-means 法を用いる．K-means 法によって分類するクラスタ数は，予備実験で最も高い結果が得られた 8 クラスタとした．

LSA により分類された単語集合の例を図 7 に示す．分類された単語集合の選択方法には，語彙集合と認識結果とのコサイン尺度による類似度計算を行い，その結果の上位 3 集合を採用する．

今回は全ての品詞に対して LSA を施すのではなく，名詞のみに対して LSA を行い語彙を推定する．具体的には，Web で話題適応化した辞書から名詞のみを削除し，推定された名詞群を辞書に新たに追加している．

4.4 発音情報を用いた語彙推定

LSA による語彙推定に対し，先行研究 [3] である程度有効性が確かめられた単語の発音情報を用いる手法も採用する．

図 6(下半分) に発音情報に基づく語彙推定手法を示す．話題適応化した認識辞書中に含まれている単語の音韻列と，認識結果として出力された単語の発音が全く同じ音韻列であった場合，その単語を辞書に登録する．このとき発音の音韻列が一致しなかった単語は辞書から取り除く．

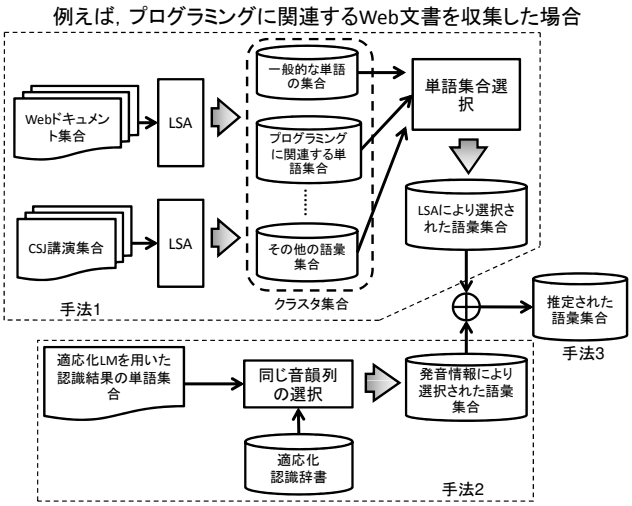


図 6: 語彙推定処理

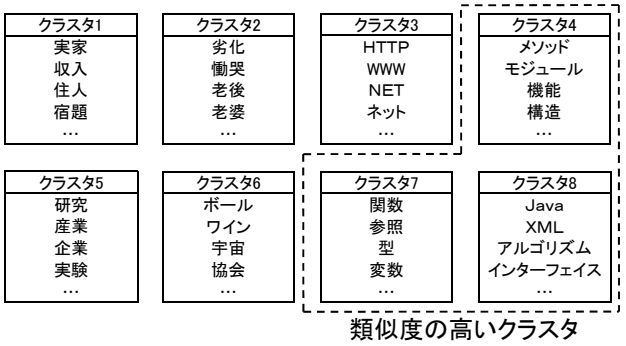


図 7: 分類された単語集合の例 (科目「プログラミング入門」)

同じ発音（音韻）系列の単語を採用することで，誤認識をした場合でも音響的特徴から比較的同じ音韻の単語が出力されると予想できる．

4.5 語彙推定手法による音声認識実験

実験対象とする音声やその分析条件等は，3.4.1 節で述べたものと全く同じである．語彙推定手法を施した言語モデルを用いた音声認識実験では，以下の 3 つの手法を比較する．

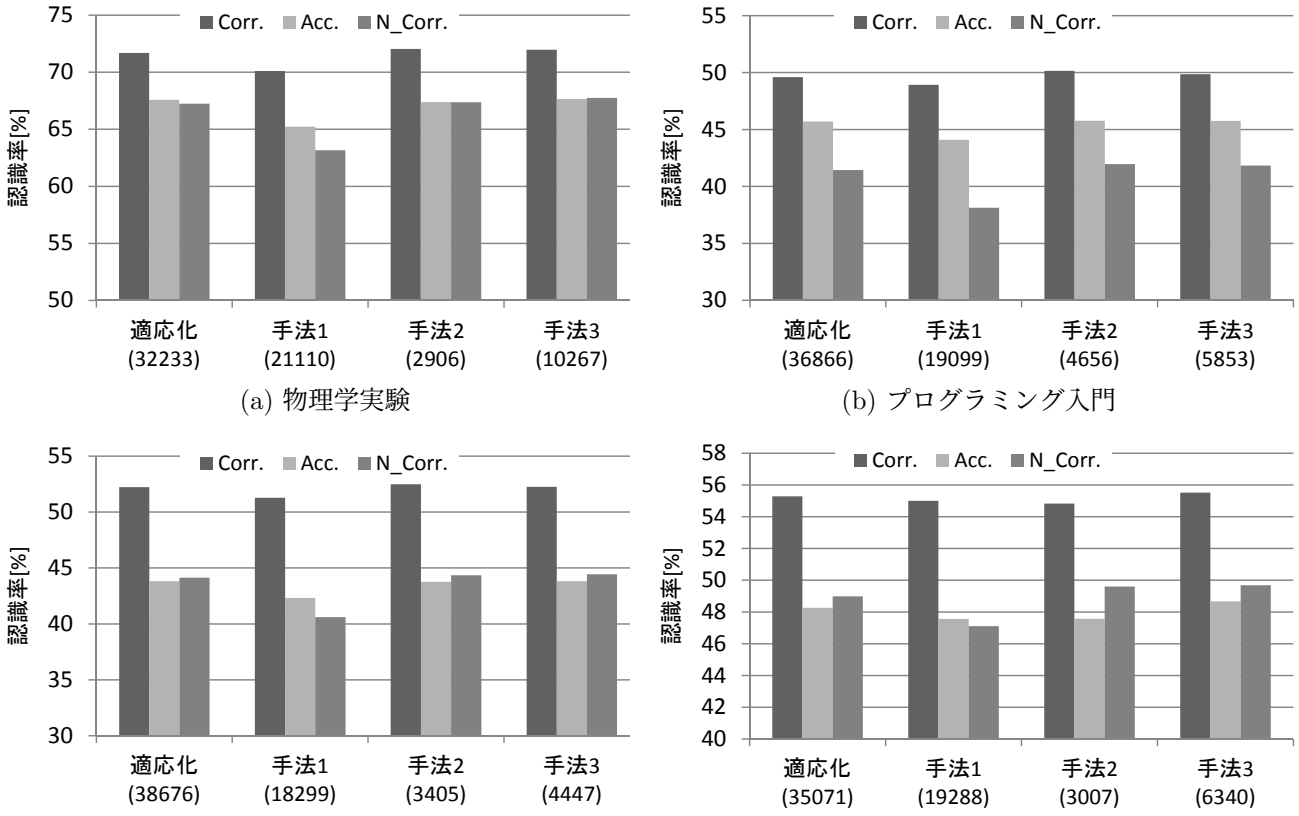
手法 1： LSA のみを用いた場合

手法 2： 単語の発音情報のみを用いた場合

手法 3： LSA(手法 1) と発音情報 (手法 2) の両方を統合した場合

4.6 音声認識実験結果

提案した語彙推定手法により作成した言語モデル・認識辞書を用いて音声認識した結果を図 8 に示す．比較対象として，Web を用いた話題適応化の認識結果も掲載す



(a) 物理学実験　(b) プログラミング入門

(c) 線形代数学 I　(d) 数値計算法

図 8: 語彙選択を行った言語モデルでの音声認識率 (講義別・カッコ内数値は語彙サイズ)

る．また，それぞれの言語モデルにおける未知語率，補正パープレキシティを表 4 に示す．

まず，手法 1 により語彙数を削減することで，単純な適応化よりも言語モデルの補正パープテキシティを下げることができている．しかし，必要な語彙までもが削減されてしまったことから，未知語率が悪化し，結果的に音声認識率が低下した．

手法 2 では，手法 1 に比べて大幅に語彙数を削減できており，補正パープレキシティも下がっている．手法 1 よりも語彙の削減数が多い分未知語率は悪化している．しかし，この大幅な削減が未知語率悪化に伴う認識率低下をカバーする形となっており，認識率は適応化モデルと同等か，講義によっては僅かながら改善している．

手法 1，手法 2 を組み合わせた手法 3 では，当然ながら手法 2 より語彙数は増加し，未知語率は改善している．手法 3 の補正パープレキシティの値は，手法 2 とほぼ同等か若干悪い程度である．そのため，手法 3 では，すべての講義において適応化モデルよりも認識率が改善している．

4.7 語彙推定によるモデルサイズと認識速度の変化

話題適応化モデルに対する語彙推定後の言語モデルサイズと認識時間の削減率を表 5 に示す．適応化モデルに比べて，最大で言語モデルサイズが約 60%，認識時間が約 20%削減され，語彙推定によりコンパクトな話題適応化言語モデルが構築できた．

表 4: 語彙推定を行った時の未知語率 (OOV [%]) と補正パープレキシティ (APP)

科目名		適応化	手法 1	手法 2	手法 3
物理学実験	OOV	1.99	4.01	5.91	4.01
	APP	190.8	175.6	167.4	178.2
プログラミング入門	OOV	2.32	8.28	7.97	7.46
	APP	334.2	265.8	266.2	272.7
線形代数学 I	OOV	1.78	5.81	7.50	6.51
	APP	198.8	182.2	174.1	174.8
数値計算法	OOV	0.76	2.67	5.85	4.17
	APP	179.0	169.4	163.2	164.4

語彙推定処理により，補正パープレキシティを 10%以上削減できたが，認識率の改善は適応化モデルと比べて僅かであった．[10] しかし，コンパクトな言語モデルを構築することができ，より高速な認識処理を行えることが分かった．

[10] 言語モデルの適応化も含めた認識の改善は，単語正解率が 5.2%，正解精度が 7.2%，名詞正解率が 14.1%であった．

表 5: 話題適応化モデルに対する語彙推定後の言語モデルサイズと認識時間の削減率 [%]

科目名	言語モデルサイズ	認識時間
物理学実験	47.8	20.8
プログラミング入門	51.9	20.8
線形代数学 I	61.5	22.9
数値計算法	54.3	15.9

5. おわりに

本稿では，よりコンパクトな話題依存言語モデル構築のための語彙推定手法を提案した．

提案手法では，Web ドキュメントを用いて言語モデルを適応化し，LSA と単語発音情報を利用して適応化言語モデルに登録する語彙を推定する．

これにより，話題適応化を行った際に生じる不要語を削除することで，語彙数を最大で 11%に減少することができ，わずかながら音声認識率も向上させることができた．また，提案手法によりコンパクトな話題適応化モデルが構築できたことから，音声認識処理速度を削減し，言語モデルのサイズを大幅に削減できた．

今後は，未知語率を下げずに効率的に語彙を削減する方法を模索していきたい．具体的には，現在は名詞のみについて LSA を利用しているが，他の品詞に対しても LSA を適応することで語彙推定を行うことを考えている．さらに，現在は講義単位での語彙推定を行っているが，講義は一貫して同じ話をしているとは限らない．そこで，文脈毎に語彙推定を行うことで認識率を改善していく方法も検討していく予定である．

参考文献

[1] 伊藤友裕, 西崎博光, 関口芳廣, “Web 上の類似記事を利用した音声文書の認識性能の改善,” 情報処理学会研究報告, 2005-SLP-59(9), pp.49–54, 情報処理学会, 12 2005.

[2] K. Maekawa, “Corpus of spontaneous japanese: Its design and evaluation,” Proc. of the ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition, 2003.

[3] 徳田翔, 西崎博光, 関口芳廣, “講義音声認識のための Web 文書を用いた
言語モデルの適応化と語彙選択,” 第 2 回音声ドキュメント処理ワークショップ講演論文集, pp.97–104, 豊橋技術科学大学メディア科学リサーチセンター, 2008.

[4] 小暮悟, 西崎博光, 土屋雅稔, 中川聖一, “講義コンテンツの収集・分析および講義音声の認識手法に関する検討,” 第 1 回音声ドキュメント処理ワークショップ, pp.1–8, 豊橋技術科学大学メディア科学リサーチセンター, 2 2007.

[5] A. Park, T.J. Hazen, and J.R. Glass, “Automatic processing of audio lectures for information retrieval: Vocabulary selection and language modeling,” Proc. of the ICASSP2005, pp.I–497–500, 2005.

[6] 根本雄介, 秋田祐哉, 河原達也, “講義音声認識のためのスライド情報を用いた言語モデル適応,” 第 1 回音声ドキュメント処理ワークショップ, pp.89–94, 豊橋技術科学大学メディア科学リサーチセンター, 2 2007.

[7] 小暮悟, 西崎博光, 土屋雅稔, 富樫慎吾, 山本一公, 中川聖一, “日本語講義音声コンテンツコーパスの構築と講義音声認識手法の検討,” 第 2 回音声ドキュメント処理ワークショップ, 豊橋技術科学大学メディア科学リサーチセンター, 2 2008.

[8] 梶浦泰智, 鈴木基之, 伊藤彰則, 牧野正三, “WWW を用いた言語モデルの教師なし反復適応法,” 第 1 回音声ドキュメント処理ワークショップ, pp.109–114, 豊橋技術科学大学メディア科学リサーチセンター, 2 2007.

[9] 秋田祐哉 河原達也, “話題と話者に関する PLSA に基づく言語モデル適応化,” 信学技報, SP2003-124, pp.67–72, 電子情報通信学会, 12 2003.

[10] 栗山直人, 鈴木基之, 伊藤彰則, 牧野正三, “情報量基準で語彙分割した PLSA 言語モデルによる話題・文脈適応,” 情報処理学会研究報告, 2006-SLP-64(40), pp.233–238, 情報処理学会, 12 2006.

[11] 踊堂憲道, 伊藤克亘, 鹿野清宏, 中村哲, “N-gram モデルのエントロピーに基づくパラメータ削減に関する検討,” 情報処理学会論文誌,42(2), pp.327–333, 情報処理学会, 2 2001.

[12] A.Stolcke, “Entropy-based pruning of backoff language models,” pp.270–274, Proceedings DARPA Broadcast News Transcription and Understanding Workshop, 1998.

[13] 大津展之, “判別および最小 2 乗法に基づく自動しきい値選定法,” 電子情報通信学会論文誌, no.4, pp.349–356, 1980.

[14] H. Nishizaki, M. Sohmiya, K. Kobayashi, and Y. Sekiguchi, “The effect of filled pauses in a lecture speech on impressive evaluation of listeners,” Proc. of the Interspeech 2007, pp.2673–2676, 2007.

[15] 鹿野清宏, 伊藤克亘, 河原達也, 武田一哉, 山本幹雄, IT Text 音声認識システム, オーム社, 2001.

[16] T.K. Landauer, D.S. McNamara, S. Dennis, and W. Kintsch, Handbook of Latent Semantic Analysis, Lawrence Erlbaum Assoc Inc, 2 2007.