

話題と話者に関するPLSAに基づく言語モデル適応

秋田 祐哉^{†,††} 河原 達也^{†,††}

† 京都大学 情報学研究科
〒606-8501 京都市左京区吉田本町
†† 科学技術振興機構 さきがけ研究21

あらまし 複数の話者と話題からなる討論音声の認識のための、統計的言語モデルの話題・話者適応を提案する。話題をカバーするコーパスと話者性をカバーするコーパスからそれぞれ構築した言語モデルを混合することで、ベースライン言語モデルを構築する。それぞれのコーパスを用いて、話題と話者に関するPLSA (Probabilistic Latent Semantic Analysis) を行い、初期の(話者ごとの)音声認識結果にこのPLSAを適用することで、当該の話題と話者に適応したunigram確率を求める。この確率によるスケールングをベースライン言語モデルに行ってそれぞれの適応を実現する。実際の討論音声を用いて評価を行い、それぞれの適応によりテストセットパープレキシティが削減され、あわせて平均6.8%の改善が得られた。

キーワード 言語モデル, 話題適応, 話者適応, PLSA, 音声認識

Language Model Adaptation based on PLSA on Topics and Speakers

Yuya AKITA^{†,††} and Tatsuya KAWAHARA^{†,††}

† School of Informatics, Kyoto University, Kyoto 606-8501, JAPAN
†† PRESTO, Japan Science and Technology Agency

Abstract We address an adaptation method of statistical language models to topics and speaker characteristics. A baseline language model is composed of a topic-oriented model and a speaker-oriented model, which are trained from different corpora covering various topics and speakers, respectively. Then, PLSA (Probabilistic Latent Semantic Analysis) is performed on the same corpora and initial ASR result to provide unigram probabilities conditioned on input speech. Finally, the baseline model is adapted by scaling N-gram probabilities with these unigram probabilities. Experimental evaluation on real discussions showed that both of topic and speaker adaptation improved test-set perplexity, and total improvement of 6.8% was obtained.

Key words Language model, Topic adaptation, Speaker adaptation, Probabilistic latent semantic analysis, speech recognition

1. はじめに

近年の情報通信技術の発展と普及によって、講義や講演、討論などの記録として音声を集積し、ネットワーク等を通じて公開する音声アーカイブが現実的なものとなってきている。音声アーカイブは、従来のテキスト主体の記録と比較して客観性や臨場感などの点で優れている。さらにインデックスや要約といった高次の情報を付加して提供することで利便性が高まり、記録としての価値が飛躍的に向上する。

このような音声アーカイブの構築において、音声認識の果たす役割は重要である。音声の書き起こし作業は人的・時間的コストが大きく、自動化が求められている。また、インデキシン

グや要約は書き起こしたテキストに大きく依存することもあり、より高い音声認識精度が望まれる。これに関して、長時間の自発発話(いわゆる「話し言葉」)を対象とした音声認識の研究が近年さかんに行われている。

音声認識において精度を高めるためには、認識対象の音声にマッチした大規模コーパスを用いて言語モデルを学習する必要がある。例えば「日本語話し言葉コーパス」[1]では、学会講演等の音声の収集と、それを用いた音声認識の研究が行われ、70%を超える単語認識精度が達成されている[2]。しかし、多くのタスクでは学習するための十分なデータが得られず、認識対象音声の特徴を部分的にカバーするコーパスを組み合わせるモデルを適応することが一般的に行われている。

話し言葉音声には、音声により伝達される具体的な情報のカテゴリである話題に関連した表現と、話し言葉特有のフィルターや文末表現が含まれる。これまでの研究では、話題に対する適応の観点から言語モデルの混合や補間、クラス N-gram といった手法が提案されており、一定の成果を収めている [3]~[5]。しかし、討論のような複数話者環境においては、話者により話し言葉表現に偏りが出る場合があり、全話者に共通した平均的な言語モデルでは対処が難しいことが考えられる。この観点からの言語モデルの適応についてはあまり研究がなされていない。

本研究では、討論音声を対象に、話題に加えて話者に対しても言語モデルを適応させることを検討する。適応の手段として確率的潜在意味解析 (PLSA, Probabilistic Latent Semantic Analysis) に基づく N-gram 確率の unigram スケーリングを利用し、話題と話者性のそれぞれをカバーするコーパスに基づく unigram スケーリングを組み合わせて適応を実現する。適応言語モデルに対し実際の討論音声を用いて評価を行い、それぞれの適応がテストセットパフォーマンスの削減に加算的に機能することを示す。

2. 話題と話者の適応に関する検討

2.1 適応における目的と課題

討論音声では、討論が扱うテーマに関連した単語・句が頻出する。テーマは 1 回の討論を通じてほぼ変化がなく、また話者によるばらつきも小さい。話題適応は、言語モデルにおけるこのような単語 (を含む N-gram) の出現確率を高め、認識対象音声に対する予測能力を高めることを目的とする。

このような、一貫したテーマに関する長時間の音声に対する話題適応は、講義・講演を中心に研究が行われている。たとえば、さまざまな話題を含む大量のテキストから構築した基本モデルに少量の話題関連テキストから構築したモデルを補間・混合する手法 [2], [3] や、クラス N-gram において話題関連テキストの単語出現確率をクラス内確率に反映させる手法 [4]、音声認識結果によるモデルを混合する手法 [2] などがある。我々も以前、潜在意味解析 (LSA, Latent Semantic Analysis) [6], [7] に基づくテキスト選択による話題適応を行い、テストセットパフォーマンスを約 6%削減している [8]。

一方、討論は複数の参加者が発話し、それぞれの話者ごとに発話スタイルにばらつきが見られる。具体的には、フィルターの多寡や分布傾向、特有の口癖などである。また、司会やパネリストなどの話者の立場によって発話の目的が異なることから、質問など特定の表現形式が多用されることがある。話者適応は、このような話し言葉表現に対する言語モデルの予測能力を話者ごとに向上させることが目的となる。

話者ごとの話し言葉表現の傾向を扱った研究としては、ホテル業務における対話を対象として、話者の立場による文型の違いを考慮した、クラス N-gram による適応が提案されている [9]。このようなタスクでは定型的な発話が多く見られるが、一般の話し言葉表現はより多様で、文型を分類することも難しい。

また、単純なコーパス混合やモデルの混合手法の適用を考えたい場合、有効なモデルを構築するためには、まず話し言葉表現

の傾向が類似したテキストを話者ごとに選択する必要がある。頑健なパラメータの推定のためには一定量のテキストを確保する必要があるが、選択されたテキストは話者の傾向に必ずしも全面的にマッチするわけではなく、結果として特徴が平均化されて効果が失われるおそれがある。本研究においても [8] と同様の手法で実際に話者適応を試みたが、選択されたテキストによる言語モデルは話者による傾向を表現できておらず、適応の効果はほとんど得られなかった。

2.2 PLSA を用いた適応

これらの問題に対処するため、本研究では PLSA [10] に着目した。PLSA は、コーパスに含まれる文書ごとの単語頻度とともに、単語の生起確率で文書の特徴づける枠組みである。この枠組みを利用して、ある未知のテキストについてそれに適応した N-gram 確率を直接推定することができる。LSA では単語頻度空間における部分空間を利用した文書間類似度は求められるが、確率は求まらない。したがって [8] ではこれを用いてテキスト選択を行った上で言語モデル混合により適応を行っている。これに対し PLSA は確率空間における部分空間であり、直接的に確率を表現していることから、このようなテキスト選択が不要となる。なお、PLSA の話題適応に関する有効性は [11] でも報告されている。

PLSA における N-gram 出現確率は、unigram を例にとると、単語 w に対して (1) 式のように定式化される。

$$P(w|d) = \sum_{j=1}^N P(w|t_j)P(t_j|d) \quad (1)$$

ここで d はコーパス中の 1 つの文書である。 t_j は非観測の潜在変数 (latent variable) であり、その総数 N は部分空間の基底数に相当する。

(1) 式における確率は EM アルゴリズムによって推定される [10]。まず、学習コーパスに含まれる文書群 $\{d_i\}$ ($i = 1, \dots, M$) の各文書における単語 w_k ($k = 1, \dots, K$) の頻度 $n(w_k, d_i)$ から、EM アルゴリズムによって $P(w_k|t_j)$ および $P(t_j|d_i)$ 、すなわち部分空間の基底と部分空間における座標を推定する。このとき、 r 番目の EM 計算における推定式は (2)・(3)・(4) 式のようなになる。

E-step

$$P^{(r)}(t_j|w_k, d_i) = \frac{P^{(r-1)}(w_k|t_j)P^{(r-1)}(t_j|d_i)}{\sum_{z=1}^N P^{(r-1)}(w_k|t_z)P^{(r-1)}(t_z|d_i)} \quad (2)$$

M-step

$$P^{(r)}(w_k|t_j) = \frac{\sum_{y=1}^M n(w_k, d_y)P^{(r)}(t_j|w_k, d_y)}{\sum_{z=1}^K \sum_{y=1}^M n(w_z, d_y)P^{(r)}(t_j|w_z, d_y)} \quad (3)$$

$$P^{(r)}(t_j|d_i) = \frac{\sum_{z=1}^K n(w_z, d_i)P^{(r)}(t_j|w_z, d_i)}{\sum_{z=1}^N \sum_{z=1}^K n(w_z, d_i)P^{(r)}(t_z|w_z, d_i)} \quad (4)$$

次に、 $P(w|t_j)$ が文書に依存しないことからこれを固定し、(2)・(4) 式を用いて適応用文書 d' に対する $P(t_j|d')$ を同様に推定する。これを用いて適応後の unigram 確率 $P(w|d')$ を求める。

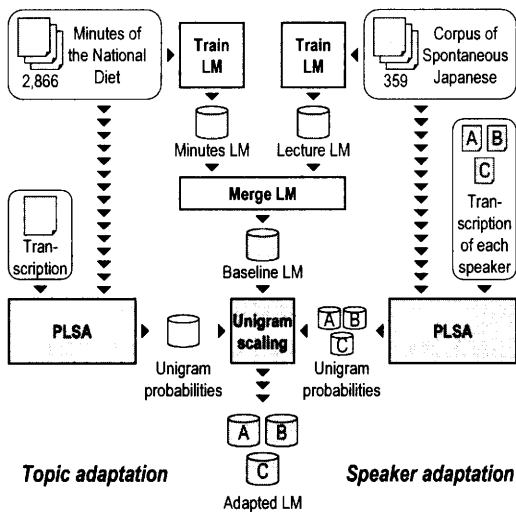


図1 言語モデル適応の処理の流れ

Fig. 1 Process flow of proposed adaptation method

この手法は莫大な計算量を必要とするため、bigram や trigram に適用することは現実的ではない。したがって、このように推定した unigram 確率を用いて、既存の言語モデルの N-gram 確率を操作する unigram スケーリング [12] が提案されている。unigram スケーリングでは、言語モデル中の unigram 確率と PLSA で推定した unigram 確率との比を N-gram 確率に乗ずることで適応が行われる。たとえば trigram の場合、単語列 $w_{i-2}w_{i-1}w_i$ に対する適応後の trigram 確率 $P'(w_i|w_{i-2}w_{i-1})$ は、もとの trigram 確率 $P(w_i|w_{i-2}w_{i-1})$ に対して (5) 式のように定式化される。

$$P'(w_i|w_{i-2}w_{i-1}) \propto \frac{P(w_i|d)}{P(w_i)} P(w_i|w_{i-2}w_{i-1}) \quad (5)$$

$P(w_i)$ はもとの言語モデルにおける単語 w_i の unigram 確率である。本研究でも学習コーパスからあらかじめ trigram モデルを学習しておき、unigram スケーリングを適用することとする。

3. 提案手法

PLSA と unigram スケーリングにより言語モデルを適応するプロセスを図 1 に示す。まず、話題と話者のそれぞれをカバーするコーパスを用いて独立に言語モデルを構築し、これらを混合してベースライン言語モデルとする。一方、それぞれのコーパスに PLSA を行い、ベースライン言語モデルによる音声認識結果をこの部分空間に写像することにより、対象討論の話題・話者に適応した unigram 確率をそれぞれ求める。これらの unigram 確率を用いてベースライン言語モデルにスケールリングを行い、適応言語モデルとする。以下では、まず対象とする討論について、次いで適応の各ステップについて述べる。

3.1 評価用討論音声

本研究では、NHK のテレビ討論番組『日曜討論』を評価用音声として利用する。『日曜討論』は、政治・経済・外交などの分野における時事問題を対象に、政治家や学者、評論家などが

表 1 各討論の話者数と発話数

Table 1 Number of speakers and utterances in test-set discussions

ID	0624	0805	0819	0902	0916
話者数	5	5	5	8	6
発話数	534	665	609	541	612
ID	1118	1125	1209	1216	0113
話者数	8	5	5	5	5
発話数	474	371	613	559	524

表 2 言語モデルの仕様

Table 2 Specification of language models

モデル	国会	講演	ベースライン
学習データ	衆議院会議録 (1999-2002)	日本語話し言葉 コーパス (CSJ) (模擬講演のみ)	—
総単語数	70M	2.9M	—
異なり単語数	72K	37K	—
文書数	2,866	359	—
語彙サイズ	30K	9.5K	32K
平均 PP	180.73	123.46	102.77
平均 OOV	4.24%	7.65%	1.98%

5-8 名程度参加し議論するものである。討論のテーマは毎回異なり、おおむね異なる人物が参加している。番組は毎回 1 時間であり、2001 年 6 月から 2002 年 1 月までの間に放送された中から 10 回分を用いた。

討論音声は 400ms の無音により区切って発話とする。1 討論あたりの平均発話数は 550 である。各討論の話者数と発話数を表 1 に示す。合計の話者数は 57 である。

3.2 学習コーパスとベースライン言語モデル

話題と話者の特徴をカバーするために、2 つのコーパスを利用して言語モデルを構築する。それぞれのコーパスと、それをもとに構築した言語モデルの仕様を表 2 に示す。

国会モデルは話題をカバーするために用い、4 年分の衆議院の全会議録（本会議・各委員会・党首討論など）を利用して構築した。会議録は会議の種類と日付によって分かれており、合計で 2,866 である。審議の内容によって各会議録（文書）のテキストサイズは大きく異なる。本研究ではこの違いについて補正や重み付けは行っていない。これらの会議録では発言の内容が忠実に書き起こされているが、フィラーの除去や文末表現・口語的表現の簡単な修正が行われている。

講演モデルは話者性をカバーするために用いられる。学習には日本語話し言葉コーパス (CSJ) に含まれる模擬講演のみを利用する。講演の総数は 1,245 であるが、同一人物が趣味や旅行など一定の与えられた話題について講演を行っている。話題が共通しているので、PLSA により話者性が主に抽出されると考えた。話者数は 359 であり、話者毎に講演の書き起こしをまとめたものを PLSA における文書の単位とする。

ベースライン言語モデルは、これら 2 つの言語モデルを相補的バックオフを用いた手法 [13] により重み付け混合したものである。混合比は予備実験によりあらかじめ国会 0.6・講演 0.4

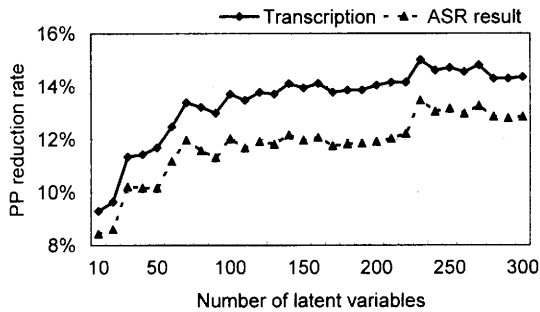


図2 国会モデルにおける潜在変数のPP削減率への影響

Fig.2 Effect of number of latent variables on PP reduction rate for the minutes model

と定めた。表2における平均テストセットパープレキシティ (PP) および平均未知語率 (OOV) は、テストセット討論音声の書き起こしを用いて算出したものである。両モデルの混合により、テストセットパープレキシティおよび未知語率を大きく削減することができている。これは、講演モデルが討論に関する話題を含まないことから、話し言葉表現に関して効果があることを示している。

3.3 PLSAによるunigram確率の算出

2.2に述べた手法により、認識結果に適応したunigram確率を求める。衆議院会議録(国会モデル)に対するPLSA(話題適応)では、各討論ごとに認識結果の全文を用い、討論ごとのunigram確率を求める。CSJ(講演モデル)に対するPLSA(話者適応)では、認識結果を話者ごとに分割し、それぞれを用いて話者ごとのunigram確率を求める。分割に必要な話者ラベルは、多数話者モデルに基づく教師なし話者インデキシング手法[14]により付与する。本手法によるテストセット討論音声の話者インデキシング精度は97%である。

PLSAにおいては、EMアルゴリズムが局所最適解に収束する問題を避けるためにannealingを行っている。各コーパスにおけるPLSAでは、衆議院会議録については77回(うちannealed EMが27回、通常のEMが50回)、CSJについては42回(うちannealed EMが27回、通常のEMが15回)の反復計算を行っている。認識結果に対するPLSAでは、CSJに対する場合と同様に42回の反復計算を行う。

適応unigram確率を求めるためには、潜在変数の個数、すなわち部分空間の次元を定める必要がある。本研究では、衆議院会議録・CSJのそれぞれのPLSAについて、国会モデル・講演モデルにunigramスケーリングを行った場合の潜在変数の個数の影響を調べた。参考のため、音声認識結果だけではなく人手による書き起こしも用いてPLSAとunigramスケーリングを行っている。

図2および図3に、潜在変数の個数と、unigramスケーリングの結果削減されたテストセットパープレキシティの割合を示す。図中、“Transcription”が人手による書き起こしの場合、“ASR result”が音声認識結果の場合である。図より、いずれの場合も文書数の10分の1程度で削減率は収束し、それ以

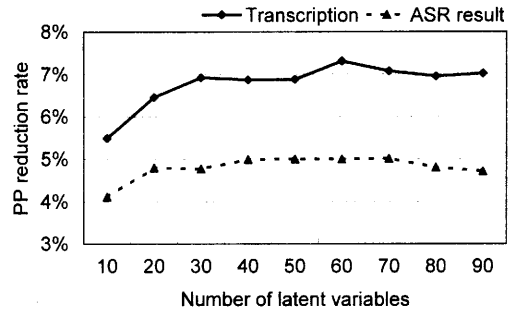


図3 講演モデルにおける潜在変数のPP削減率への影響

Fig.3 Effect of number of latent variables on PP reduction rate for the lecture model

上では削減幅の変化が小さいことがわかる。本研究では最もよい削減率を示した個数を採用することとし、衆議院会議録では230個、CSJでは60個とした。なお、音声認識結果を用いた場合は書き起こしを用いた場合と比較して1-2ポイント改善率が悪いが、これは後述するようにベースラインの音声認識精度が高くないことが原因と考えられる。

3.4 話題と話者の適応の統合

unigramスケーリングは、言語モデルのunigram確率に対する適応unigram確率の比を利用するため、2つのPLSAを行う本手法では何らかの方法で2つの適応unigram確率を統合する必要がある。本研究では、(6)式のようにこれらのunigram確率の重み付き線形和を求め、scalingに用いることとする。

$$P(w|d) = \lambda P_t(w|d) + (1 - \lambda) P_s(w|d) \quad (6)$$

ここで P_t と P_s はそれぞれ衆議院会議録(話題)・CSJ(話者)に関するPLSAで得られたunigram確率である。それぞれの言語モデルは語彙が異なるため、一方に現れない単語に関しては確率を0として扱う。本研究では、予備実験でもっともよい性能を示した混合比 $\lambda = 0.6$ を用いる。これは、ベースライン言語モデルにおける国会・講演の両モデルの混合比と一致している。

なお、処理の順番を逆にして、あらかじめ国会モデルと講演モデルにunigramスケーリングを行った上で言語モデルを混合する手法も考えられるが、混合により適応の効果が薄れる可能性があり、実際、予備実験においても提案手法ほどの性能を得られなかった。

4. 評価実験

提案手法により話題と話者に関して適応させた言語モデルについて、テストセット討論音声により評価を行った。評価の尺度としてはテストセットパープレキシティおよび音声認識における単語認識精度を用いる。

音声認識については、デコーダにJulius 3.4[15]を利用し、逐次デコーディング[16]を適用している。音響モデルには、CSJから学習したPTM triphone HMM(表3)に対して教師なし話者・音素ラベルによるMLLR話者適応を行ったモデル[17]を用いた。

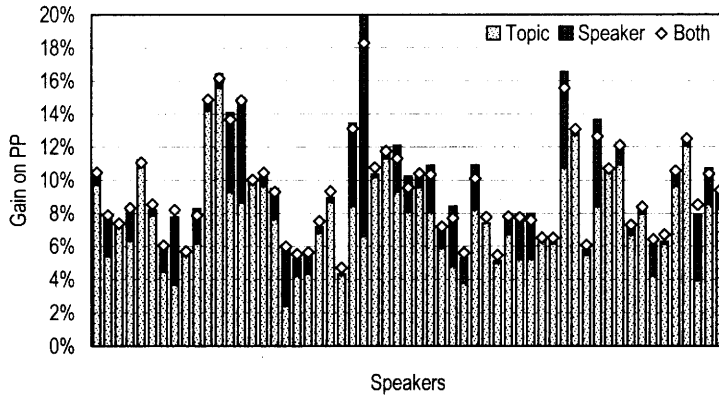


図 4 各話者ごとのパープレキシティの改善（書き起こしによる PLSA の場合）

Fig. 4 Improvement of PP for each speaker (PLSA for manual transcriptions)

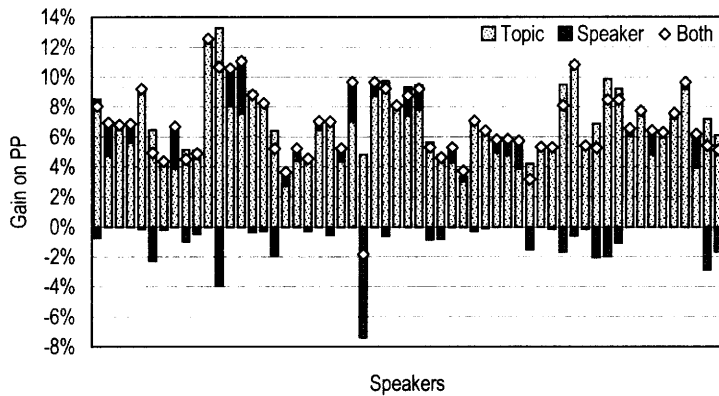


図 5 各話者ごとのパープレキシティの改善（音声認識結果による PLSA の場合）

Fig. 5 Improvement of PP for each speaker (PLSA for ASR results)

表 3 音響モデルの仕様

Table 3 Specification of the acoustic model

学習データ	日本語話し言葉コーパス (CSJ) 60 時間 (男性のみ)
特徴量	MFCC(12) Δ MFCC(12) Δ Energy(1) (計 25 次元)
音素数	43
状態数	3,000
コードブック数	129
混合数	128

書き起こしおよび音声認識結果を用いた適応モデルによる話者ごとのテストセットパープレキシティの改善率を、それぞれ図 4・図 5 に示す。比較のため、話題適応のみ、話者適応のみを行った言語モデルによる改善率もあわせて示す。図中、“Topic”と“Speaker”がそれぞれ話題適応のみ・話者適応のみを行った場合の改善率であり、図ではそれぞれの改善率が積算されている。“Both”は両方の適応を行った場合の改善率である。図 4 より、書き起こしを用いた場合は、話題適応・話者適応のいずれ

もおおむね改善を示し、話題適応では平均 7.65%、話者適応では平均 1.79%の改善を得た。話者適応では話者により効果の度合いが異なるが、これはベースライン言語モデルに対するマッチングの度合いに個人差があり、ミスマッチの大きい話者を中心に効果が現れたものと考えられる。両方の適応を行った場合は、多くの話者においてそれぞれの適応の効果が加算的に現れており、平均でも 9.39%と、それぞれの改善率の和に近い数値が得られた。これは、PLSA による unigram 確率の重み付き線形和がそれぞれの効果を減殺することなく機能していることを示している。

一方、図 5 より、音声認識結果を用いた場合には、書き起こしを用いた場合ほどの効果は得られなかった。平均の改善率は、話題適応では 6.59%、話者適応では -0.06%、両方の適応では 6.79%となった。3.3 でも述べたように、これは音声認識誤りに起因するものであるが、誤りの多くが話し言葉表現において発生していることが話者適応の効果に悪影響をもたらしたものと考えられる。

次に、適応言語モデルを用いた音声認識における、討論ごとの単語認識精度を図 6 および図 7 に示す。ベースライン言語モ

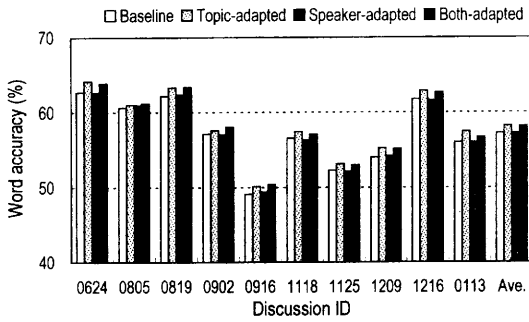


図 6 各討論ごとの単語認識精度 (書き起こしによる PLSA の場合)
Fig. 6 Word accuracy for each discussion (PLSA for manual transcriptions)

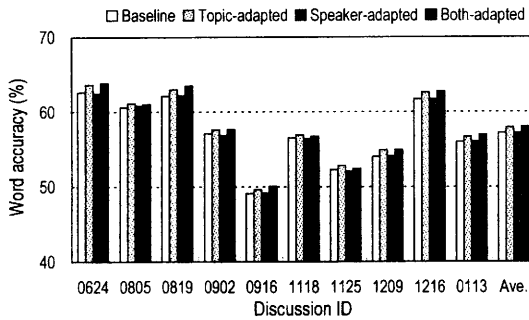


図 7 各討論ごとの単語認識精度 (音声認識結果による PLSA の場合)
Fig. 7 Word accuracy for each discussion (PLSA for ASR results)

デルによる単語認識精度は 57.2%である。これに対し、話題適応モデルは、適応テキストが書き起こしの場合には 1.0%、音声認識結果の場合には 0.6%向上している。話者適応モデルはほとんど改善が得られず、したがって両方の適応を行ったモデルは話題適応モデルとはほぼ同じ精度となっている。テストセットパープレキシティにおける改善が単語認識精度の改善に反映されていないが、この理由として、音響モデルが討論音声とマッチしていないために言語モデルの改善の効果が現れにくいことが考えられる。

5. おわりに

本稿では、討論音声を対象に、話題と話者に関して言語モデルを適応する手法を提案した。話題と話者性のそれぞれをカバーするコーパスをもとに言語モデルを構築し、これらを混合してベースライン言語モデルとする。一方、これらのコーパスとベースライン言語モデルによる音声認識結果を用いて PLSA を行い、音声認識結果に適応した unigram 確率を求め、この確率を用いてベースライン言語モデルの N-gram 確率に unigram スケーリングを行って適応を実現する。この適応手法が話題と話者性のそれぞれにおいてテストセットパープレキシティの削減に有効であることを示し、実際の討論音声を用いた評価では 6.8%のパープレキシティの改善が見られた。

- [1] S. Furui, K. Maekawa, and H. Isahara. Toward the Realization of Spontaneous Speech Recognition - Introduction of a Japanese Priority Program and Preliminary Results -. In *Proc. ICSLP*, 2000.
- [2] 南條浩輝, 河原達也. 講演音声認識のための話題・話者・発音変動に適応した言語モデル. 日本音響学会春季研究発表会講演論文集, 3-4-2, 2003.
- [3] 横山忠介, 篠崎隆宏, 岩野公司, 古井貞照. 言語モデルのバッチ型教師なし適応化法. 情報処理学会研究報告, 2002-SLP-44-31, 2002.
- [4] 森信介, 伊東伸泰, 西村雅史. 単語クラスタリングによる確率的言語モデルの分野適応. 情報処理学会研究報告, 2003-SLP-45-15, 2003.
- [5] 南條浩輝, 加藤一臣, 李見伸, 河原達也. 大規模な日本語話し言葉データベースを用いた講演音声認識. 電子情報通信学会論文誌, Vol. J86-DII, No. 4, pp. 450-459, 2003.
- [6] T. K. Landauer, P. W. Foltz, and D. Laham. An Introduction to Latent Semantic Analysis. *Discourse Processes*, Vol. 25, pp. 259-284, 1998.
- [7] J. R. Bellegarda. Exploiting Latent Semantic Information in Statistical Language Modeling. *Proc. IEEE*, Vol. 88, No. 2, pp. 1279-1296, 2000.
- [8] 秋田祐哉, 河原達也. 討論音声認識のための言語モデルのオフライン話題適応. 日本音響学会秋季研究発表会講演論文集, 2-6-7, 2003.
- [9] 山本博史, 匂坂芳典. 話題と文型の違いを同時に考慮した言語モデル適応. 電子情報通信学会論文誌, Vol. J85-DII, No. 8, pp. 1284-1290, 2002.
- [10] T. Hofmann. Probabilistic Latent Semantic Indexing. In *Proc. SIG-IR*, 1999.
- [11] 三品拓也, 山本幹雄. 確率的 LSA に基づく ngram モデルの変分ベイズ学習を利用した文脈適応化. 情報処理学会研究報告, 2002-SLP-44-30, 2002.
- [12] D. Gildea and T. Hofmann. Topic-based Language Models using EM. In *Proc. Eurospeech*, 1999.
- [13] 長友健太郎, 西村竜一, 小松久美子, 黒田由香, 李見伸, 猿渡洋, 鹿野清宏. 相補的バックオフを用いた言語モデル融合ツールの構築. 情報処理学会論文誌, Vol. 43, No. 9, pp. 2884-2893, 2002.
- [14] Y. Akita and T. Kawahara. Unsupervised Speaker Indexing using Anchor Models and Automatic Transcription of Discussions. In *Proc. Eurospeech*, 2003.
- [15] 河原達也, 住吉貴志, 李見伸, 坂野秀樹, 武田一哉, 三村正人, 伊藤克巨, 伊藤彰則, 鹿野清宏. 連続音声認識コンソーシアム 2002 年度版ソフトウェアの概要. 情報処理学会研究報告, 2003-SLP-48-1, 2003.
- [16] 李見伸, 河原達也, 鹿野清宏. 話し言葉の認識のためのデコーダ Julius の改良. 日本音響学会春季研究発表会講演論文集, 1-3-15, 2001.
- [17] 秋田祐哉, 河原達也. 討論音声認識のための言語モデルと音響モデルの適応の検討. 日本音響学会春季研究発表会講演論文集, 2-4-3, 2003.