

話題の連続/不連続変化を考慮したトピックモデルに基づく 音声認識

佐古 淳[†] 有木 康雄[†] 岩田 具治^{††} 渡部 晋治^{††} 堀 貴明^{††}

[†] 神戸大学大学院自然科学研究科

〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

^{††} 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: [†]sakoats@me.cs.scitec.kobe-u.ac.jp, ^{††}ariki@kobe-u.ac.jp,

^{†††}{iwata,watanabe,horii}@cslab.kecl.ntt.co.jp

あらまし 本稿では、話題の連続/不連続変化を考慮したトピックモデルを提案し、MIT 講義音声コーパスを用いて行った評価実験について述べる。実環境においては、話者交代や話し方、話題の変化などにより音響的・言語的特徴が時々刻々と変化する。このような変化に対応するため、発話単位で環境適応を行う逐次追従型音声認識が注目されている。本研究では、言語的な変化に着目し、話題の連続的な変化を考慮するモデルとして Online Topic Model を用いた言語モデル適応、及び話題の不連続な変化を考慮するモデルとして Topic HMM を用いた言語モデル適応を提案する。評価実験により、これらのトピックモデルを用いることで単語誤り率 (WER) を改善出来ることを確認した。また、刻々と変化する話題を追跡することで、認識結果全体を用いた言語モデルのバッチ適応よりも良好な結果を得ることが出来た。

キーワード 逐次追従型音声認識, 言語モデル適応, トピックモデル

Speech Recognition by Topic Models with Continuous/Discontinuous Topic Changes

Atsushi SAKO[†], Yasuo ARIKI[†], Tomoharu IWATA^{††}, Shinji WATANABE^{††}, and

Takaaki HORII^{††}

[†] Graduate School of Science and Technology, Kobe University

1-1 Rokkodaicho, Nada-ku, Kobe, Hyogo, 657-8501 Japan

^{††} NTT Communication Science Laboratories, NTT Corporation

2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237 Japan

E-mail: [†]sakoats@me.cs.scitec.kobe-u.ac.jp, ^{††}ariki@kobe-u.ac.jp,

^{†††}{iwata,watanabe,horii}@cslab.kecl.ntt.co.jp

Abstract In this paper, we propose topic models with continuous/discontinuous topic changes, and describe experiments using MIT Open Course Ware corpus. In a real environment, acoustic and language features vary momentarily depending on speakers, speaking styles or topic changes. To accommodate these changes, speech recognition with incremental tracking of changing environments has attracted attention. We propose a language model adaptation technique by Online Topic Model for continuous topic changes, and a technique by Topic HMM for discontinuous topic changes. The experimental results showed the improvements of Word Error Rate with these topic models. Moreover, the proposed methods outperformed the batch adaptation of language model using whole speech recognition results by tracking temporal changes of topics.

Key words Incremental tracking of changing environments, Language model adaptation, Topic model

1. はじめに

近年、一般のエンドユーザーが視聴可能なマルチメディア・コンテンツが増大している。マルチメディア・コンテンツ作成コストが低下し、多くのコンテンツが作成される一方、インターネットを通じた大容量データの配信・受信のためのインフラも整備されつつある。このように大量のコンテンツが配信される状況においては、ユーザーが自力で欲しいコンテンツを探し当てることは困難である。ユーザーが求めるデータを検索するための技術が必要となる。しかし、多くのマルチメディア・データは映像と音声のみで構成されテキスト情報を含んでいないため、一般的な Web 検索エンジンでは検索が困難な場合が多い。検索のためのテキスト情報を得るため、音声認識技術を用いて音声データをテキストに変換する手法が注目されている。具体的には、WWW 上のポッドキャストを対象として PodCastle [1] や、MIT の講義映像/音声を対象として MIT Lecture Browser [2]、音声認識と統計翻訳を組み合わせた多言語動画検索/閲覧システム [3] などがある。

音声認識を行う際、認識の対象となる音声データ全てに対し、単一の音響モデル・言語モデルを用いる場合が多い。しかしながら、実環境においては、話者交代や外部ノイズ、話し方の変化、話題の変化などにより音響的特徴・言語的特徴ともに時々刻々と変化していくものと考えられる。このような変化に対応するため、音響モデルにおいては、発話単位で環境適応を行う逐次追従型音声認識 [4] が提案されている。これに対し本稿では、言語モデルにおける逐次追従型音声認識について述べる。言語モデル適応手法としては、日々追加される単語や文書を反映して言語モデルを更新する手法 [5] や、出現した単語に応じてその後の単語出現確率を変化させるトリガモデル [6]、キャッシュモデル [7] などが提案されている。

本研究では、言語的特徴の時変性を扱うために、不連続な話題の変化・連続的な話題の変化、双方に着目した。ここで不連続な話題変化は、スポーツから政治といった話題の変化を対象にしており、あらかじめ用意された複数の話題を用いて、話題から話題への状態遷移モデルとして記述される。また連続的な変化は、発話を重ねるにつれ話題が徐々に変化していく状況を対象にしており、単一の話題が過去の話題に補充される形で滑らかに変化する時間発展モデルとして記述される。具体的手法として、Probabilistic Latent Semantic Analysis (PLSA) に基づく Topic HMM [8]、及び Online Topic Model (OTM) [9] をそれぞれ不連続変化を追跡するモデル、連続的な変化を追跡するモデルとして用いた。

話題の不連続な変化を追跡する方法として、HMM の出力確率を N -gram 確率とすることで各状態にサブタスクに対応した言語モデルを獲得する手法 [10] や、文脈自

由文法を HMM によって制御する手法 [11] などが提案されている。これに対し Topic HMM では、単語の直接的な出力確率を制御するのではなく、PLSA の話題分布を HMM によって制御する。

一方、連続的に変化を追跡する方法として、カルマンフィルタが一般に用いられるが、そこでは出力が連続変数であることを仮定しているため、単語のような離散変数には適さない。離散変数の場合にカルマンフィルタを変分ベイズ法を用いて解く方法も提案されている [12] が、オンライン学習ではないため、変化追跡には適さない。Online Topic Model (OTM) では、認識結果から逐次的に単語出現確率を予測する。

本研究では、上記のような話題の連続/不連続に基づくトピックモデルを逐次追従型音声認識に用い、マサチューセッツ工科大学 (MIT) の講義音声コーパスを用いて評価した。以下、次節において話題の連続/不連続を考慮したトピックモデルについて述べ、3. 節において評価実験について述べる。

2. 話題の連続/不連続を考慮したトピックモデル

本節では、話題の連続/不連続な変化を考慮したトピックモデルについて述べる。話題の変化を考慮したトピックモデルは、単語 w 、時刻 t (もしくはある時刻における文書 t)、及び潜在トピック z を導入することにより、

$$P(w|t) = \sum_z P(w|z,t)P(z|t) \quad (1)$$

のように定式化される。ここで、時刻 t の変化に伴って変化するパラメータのモデル化として不連続にパラメータを変化させるモデル化、及び連続的に変化させるモデル化の 2 通りが考えられる。不連続なモデルは、状態遷移モデルとして記述され、時刻 t において所属する状態が変化したときに初めてパラメータが変化するモデルである。また、連続的なモデルは、ディリクレ分布を用いて時刻 t から時刻 $t+1$ へのパラメータの変化を直接的に記述するモデルである。このモデルでは、時刻が変化する度に連続的にパラメータが変化する。

以下、まず最初に不連続な変化を考慮したトピックモデルとして PLSA に基づく Topic HMM について述べ、その後、連続的な変化を考慮した Online Topic Model について述べる。

2.1 PLSA-based Topic HMM

Probabilistic Latent Semantic Analysis (PLSA) [13] では、ある文書 t における、ある単語 w の出現確率 $P(w|t)$ (式 1) を、潜在変数 z を用いて、

$$P(w|t) = \sum_z P(w|z)P(z|t) \quad (2)$$

のように表現する。潜在変数 z をトピックととらえることで、 $P(w|z)$ は、あるトピックから単語 w が出現する

Ergodic HMM

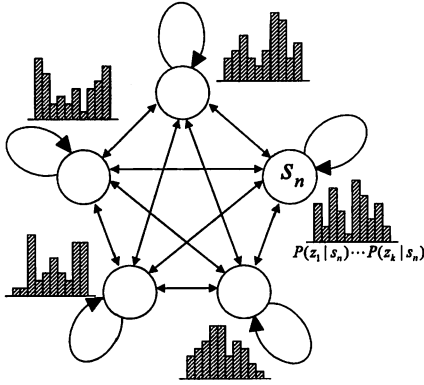


図1 Topic HMM.

確率であり文書 t には依存しない。 $P(z|t)$ は、文書 t の持つトピック分布となる。これらのパラメータはEMアルゴリズムによって学習される。

Topic HMM は、HMM の各状態に典型的なトピック分布を持ち、また状態間に移移確率を持つようなモデルである(図1)。HMM には、あらゆる状態遷移を許すErgodic HMM を用いる。PLSA によって求められる文書のトピック分布 $P(z|t)$ を特徴量として学習を行うことで上記のようなモデルを得る。ただし、ここでいう“文書”は、時間的変化が観測出来るような数発話の集合を想定しており、新聞の記事1つのようないわゆる“文書”とは異なっている。しかしながら、トピックモデルの慣習に従い文書と表現する。また、これらの“文書”の出現順序は保たれている必要がある。

次に、学習の具体的手順について述べる。まず、学習コーパス中の文書全体を用いてPLSAを行い、各文書 t のトピック分布 $P(z|t)$ を得る。次に、各文書のトピック分布を以下のようにベクトル化し、特徴量とする。

$$\mathbf{x}_t = \begin{pmatrix} P(z_1|t) \\ \vdots \\ P(z_K|t) \end{pmatrix}. \quad (3)$$

ここで、 K は潜在トピック数である。こうして求めた特徴量 \mathbf{x}_t を連続的に観測される特徴量としてHMMの学習を行う。HMMの学習後、各状態の平均ベクトル μ_s を、その状態における典型的トピック分布と考え用いる。

次に、Topic HMM を用いた言語モデル適応について述べる。Topic HMM の各状態には、典型的トピック分布が存在するだけで、 N -gram 確率が存在するわけではない。本研究では、PLSAに基づくUnigram Rescalingを用いることでTopic HMM からの言語モデル適応を行う。Topic HMM の状態 s における N -gram 確率は、

$$P(w_i|w_{i-1}, w_{i-2}, s) \propto \frac{P(w_i|s)}{P(w_i)} P(w_i|w_{i-1}, w_{i-2}),$$

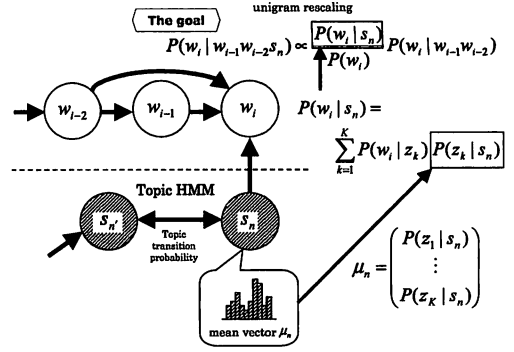


図2 Topic HMM を用いた言語モデル適応.

$$P(w_i|s) = \sum_z P(w_i|z)P(z|s)$$

のように求められる。ここで、 $P(z|s)$ は、HMM の状態 s の平均ベクトル μ_s の要素である。上記をまとめると図2のようになる。

認識は、認識対象の文書系列全体を $\mathbf{D} = (\mathbf{d}_1, \dots, \mathbf{d}_N)$ 、文書を $\mathbf{d}_t = (w_1, \dots, w_{M_t})$ 、文書の状態系列全体を $\mathbf{S} = (s_1, \dots, s_N)$ 、文書毎の観測音声特徴 $\mathbf{O} = (\mathbf{o}_1, \dots, \mathbf{o}_N)$ とすると、

$$(\hat{\mathbf{S}}, \hat{\mathbf{D}}) = \underset{\mathbf{S}, \mathbf{D}}{\operatorname{argmax}} P(\mathbf{S}, \mathbf{D}|\mathbf{O}) \quad (4)$$

のように定式化出来る。ベイズの定理により、以下のように変形する。

$$(\hat{\mathbf{S}}, \hat{\mathbf{D}}) = \underset{\mathbf{S}, \mathbf{D}}{\operatorname{argmax}} P(\mathbf{S}, \mathbf{D})P(\mathbf{O}|\mathbf{D}). \quad (5)$$

ここで、 \mathbf{O} と \mathbf{S} は互いに独立であると仮定した。さらに、状態はひとつ前の状態にのみ依存、文書は状態にのみ依存という仮定をおけば、 $P(\mathbf{S}, \mathbf{D})$ は以下のように展開できる。

$$\begin{aligned} P(\mathbf{S}, \mathbf{D}) &= P(s_1, \dots, s_N, \mathbf{d}_1, \dots, \mathbf{d}_N) \\ &\approx P(s_1)P(\mathbf{d}_1|s_1) \\ &\quad \times \prod_{t=2}^N P(s_t|s_{t-1})P(\mathbf{d}_t|s_t). \end{aligned} \quad (6)$$

また、文書中の単語は過去2つの単語にのみ依存すると仮定すれば、

$$P(\mathbf{d}_t|s_t) = \prod_i P(w_i|w_{i-1}, w_{i-2}, s_t) \quad (7)$$

となる。 $P(s_t|s_{t-1})$ には Topic HMM の状態間遷移確率を用いる。また、 $P(w_i|w_{i-1}, w_{i-2}, s_t)$ は Topic HMM の状態 s_t によって適応された言語モデルを用いる。以上のようにして、Topic HMM を駆動しながら文書毎に適切な言語モデルを用いて音声認識を行う。

2.2 Online Topic Model

Online Topic Model では、時刻 t における単語 w の出現確率を $P(w|t)$ (式 1) を以下のようにパラメトリックに表現する。

$$P(w|t) = \sum_z \phi_{tz} \theta_{tzw}, \quad (8)$$

ここで、 ϕ_{tz} は時刻 t においてトピック z が出現する確率 $P(z|t)$ 、 θ_{tzw} は時刻 t においてトピックが z のとき単語 w が出現する確率 $P(w|z, t)$ を表す。なお、実験では θ_{tzw} として、PLSA で事前に推定した $P(w|z)$ を用いる。

時間変化をモデル化するため、OTM では、トピック出現確率 $\phi_t = (\phi_{tz})_{z=1}^Z$ の事前確率として、平均が ϕ_{t-1} 、分散が $1/\alpha_t$ のディリクレ分布

$$P(\phi_t | \phi_{t-1}, \alpha_t) \propto \prod_z \phi_{tz}^{\alpha_t \phi_{t-1z} - 1z}, \quad (9)$$

を用いる。共役事前分布であるディリクレ分布を用いることにより、パラメータ推定が容易になる。また、平均として一時刻前の平均を用いていることにより、データが入力されなければ一時刻前と同様のトピック出現確率となるため、自然なモデル化であると考えられる。トピック変化に柔軟に対応可能にするため、変化しにくさを表す α_t をデータから推定する。トピック出現確率と同様に、トピック毎の単語出現確率の事前分布として、以下のディリクレ分布

$$P(\theta_{tz} | \theta_{t-1z}, \beta_{tz}) \propto \prod_w \theta_{tzw}^{\beta_{tz} \theta_{t-1zw} - 1zw}, \quad (10)$$

を用いる。

時刻 t におけるトピック出現確率 ϕ_t 、および、パラメータ α_t は、トピック z のギブスサンプリングと α_t に関する最尤推定を繰り返すことにより、得ることができる。ギブスサンプリングでは下式により各単語にトピックを割り当てる。

$$P(z_m = k | \mathbf{W}_t, \mathbf{Z}_{t \setminus m}) \propto \frac{n_{tk \setminus m} + \alpha_t \phi_{tk} \frac{n_{tkw_m \setminus m} + \beta_{tk} \theta_{tkw_m}}{n_{tk \setminus m} + \beta_{tk}}}{n_{t \setminus m} + \alpha_t}$$

ここで、 \mathbf{W}_t は時刻 t における単語集合、 \mathbf{Z}_t は時刻 t におけるトピック集合、 $\mathbf{Z}_{t \setminus m}$ は \mathbf{Z}_t から m 番目の単語を取り除いたときのトピック集合、 n_{tk} はトピック k を割り当てられた時刻 t における単語数、 n_t は時刻 t における単語数、 n_{tkw} はトピック k を割り当てられた時刻 t における単語 w の数、 $n_{\setminus m}$ は m 番目の単語を除いたときの単語数を表す。また、 α_t を下式により更新することにより、尤度を最大化することができる [14]。

$$\alpha_t \leftarrow \alpha_t \frac{\sum_z \phi_{t-1z} (\Psi(n_{tz} + \alpha_t \phi_{t-1z}) - \Psi(\alpha_t \phi_{t-1z}))}{\Psi(n_t + \alpha_t) - \Psi(\alpha_t)}$$

ここで Ψ はディガンマ関数を表す。 β_{tz} も同様に尤度を最大化することにより推定できる。

Graphical model

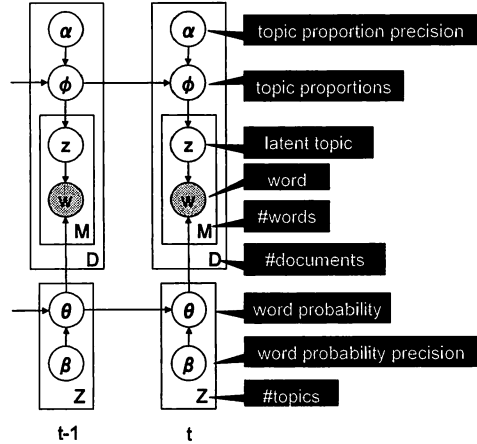


図 3 Online Topic Model のグラフィカルモデルによる表現。

ギブスサンプリングおよび最尤推定を終了条件が満たされるまで繰り返した後、時刻 t におけるトピック出現確率 ϕ_t を下式で計算する。

$$\hat{\phi}_{tz} = \frac{n_{tz} + \alpha_t \phi_{t-1z}}{n_t + \alpha_t}, \quad (11)$$

ここで求めた ϕ_{tz} は、次の時刻のトピック出現確率の事前分布として用いられる。時刻 t のパラメータを推定する際に、時刻 $t-1$ 以前のデータは必要ないため、計算量およびメモリ容量を抑えることができる。Online Topic Model のグラフィカルモデルによる表現を図 3 に示す。認識は、

- 音声認識
- 認識結果から OTM により次の時刻の unigram 確率を推定
- Unigram Rescaling により言語モデル適応
- 適応後の言語モデルを用いて音声認識

を繰り返すことを行う。

3. 評価実験

本節では、MIT 講義音声コーパスを用いた音声認識実験について述べる。実験は、話題の不連続な変化を考慮する Topic HMM を用いた手法、及び話題の連続的な変化を考慮する Online Topic Model を用いた手法それぞれで行った。まず、それぞれに共通する音声認識条件について述べる。

3.1 音声認識条件

学習データは MIT 講義音声コーパスの学習セット 147 講義 (128 時間, 6.2M 単語) を用いた。評価データには 6 講義のうちの 1 講義を用いた (話者クローズのもの)。音響分析条件と HMM の仕様を表 1 に示す。特徴パラ

表 1 音響分析条件と HMM の仕様

Table 1 Condition of acoustic analysis and HMM specification.

音響分析	サンプリング周波数	16kHz
	特徴パラメータ	MFCC(39 次元)
	フレーム長	25ms
	フレーム周期	10ms
窓タイプ	ハミング窓	
	音素カテゴリ数	51 (音素 42 種+ノイズ等 9 種)
HMM	状態共有トライフォン	状態数 2193, 混合数 32 (VBEC で自動クラスタリング)
M	状態数	3 状態 3 出力分布

メータは正規化対数パワー+MFCC12 次元とこれらの Δ , $\Delta \Delta$ の計 39 次元である。また、発話単位のケプストラム正規化を行った。音響モデルの状態数/混合数は VBEC を用いて自動的に決定した [15]。また、ベースラインの音響モデル作成後、MLLR による教師なし適応を行った。

辞書は学習コーパスから獲得した。語彙サイズは 70397, 未知語率は 1.6% であった。言語モデルは Kneser-Ney スムージングによる単語トライグラムを用いた。テストセット・パープレキシティは 163.4 であった。

音声認識デコーダーには WFST ベースの SOLON [16] を用いた。また、Confusion Network を用いた Consensus Decoding [17] を同時に用いた場合についても評価を行った。

3.2 文書のサイズ

本研究では、“文書”毎に言語モデルを変化させる。しかしながら、本研究のタスクは自然発話音声であり、文書の単位は必ずしも明確でない。新聞や Web ページのように、1 記事や 1 ページといった単位で文書を考えることは困難である。そこで、本研究では、無音区間で区切った音声を 1 発話とし、8, 16, 32, 64, 128 発話をまとめたものをそれぞれ 1 つの文書として扱った。評価データの発話数は 478 発話であったため、8 発話をまとめた場合の文書数は 60, 以下 16, 32, 64, 128 発話をまとめた場合でそれぞれ 30, 15, 8, 4 文書であった。

3.3 実験結果

PLSA の潜在トピック数は 50 に固定し学習を行った。これにより得られる $P(w|z)$ は、Topic HMM, 及び Online Topic Model の双方で同じものを利用した。また、Topic HMM の状態数も 50 で固定し、学習を行った。Online Topic Model においては、過去 10 文書までの履歴を利用した。Topic HMM, Online Topic Model の双方において、発話数/文書を 8 から 128 まで変化させて実験を行った。

音声認識実験の結果を図 4 に示す。Baseline は、通常の単一言語モデルを用いた場合の認識結果である。結果は、発話数/文書を変化させた際、最も単語誤り率 (WER) が低くなった場合のものである。Consensus Decoding を用

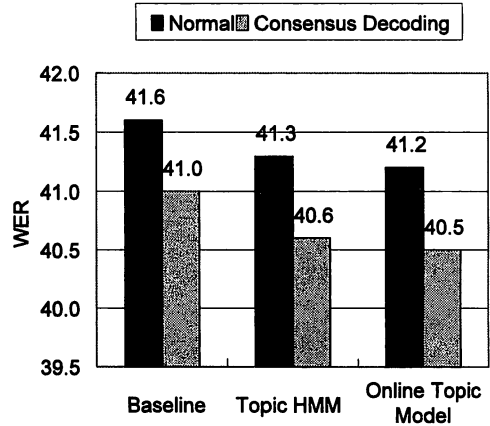


図 4 各種法における実験結果。

いることで Baseline の WER は 41.6 から 41.0 へ改善している。提案手法の Topic HMM を用いることで 40.6, Online Topic Model を用いることで 40.5 へと、さらなる改善が得られた。特に、Online Topic Model については、トリガーモデル・キャッシュモデルに類する効果が得られたと考えられる。提案手法のトピックモデルを用いることで、話題の不連続な変化、連続的な変化のそれぞれに対応出来たものと考えられる。また、連続/不連続のそれぞれで WER 改善の効果が得られていることから、両者を組み合わせることによりさらなる改善が期待される。

次に、発話数/文書を変化させた場合の結果を図 5 及び図 6 に示す。図 5 は Consensus Decoding を用いなかった場合、図 6 は用いた場合の結果である。図中、Baseline は単一言語モデルによる認識結果である。また All は、一度評価データ全てを音声認識し、得られた認識結果全てを用いて Online Topic Model により言語モデル適応を行った場合の結果である。図 5, 図 6 の結果はともに類似した傾向を示している。どちらの場合でも、64 発話/文書の際に最も良い結果を示した。また、提案手法の結果が All の結果よりも WER において良好な結果を示した。認識結果の全てを用いて適応を行うよりも、適切な区切りで話題を逐次追跡し適応を行う方が高精度に認識を行えると考えられる。さらに、All は一度全ての認識を終える必要があるのに対し、提案手法はオンラインでの言語モデル適応・認識が可能であり、この点もアドバンテージであると言える。

4. まとめ

本稿では、話題の連続/不連続な変化を考慮したトピックモデルとして、Online Topic Model (連続), 及び Topic HMM (不連続) について述べた。MIT 講義音声コーパスにおいて評価実験を行い、WER の改善を確認した。特に、評価データ全ての認識結果を用いて言語モデル適応

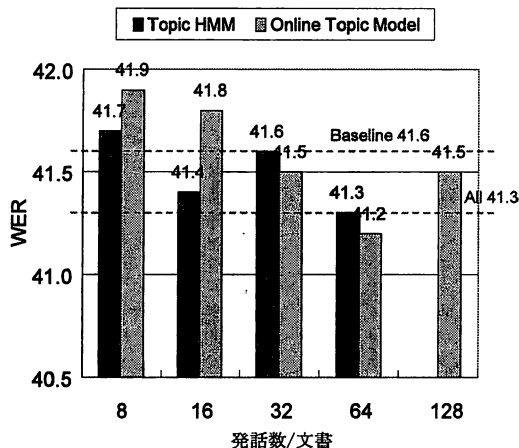


図5 WERと発話数/文書の関係（Consensus Decodingを用いない場合）。

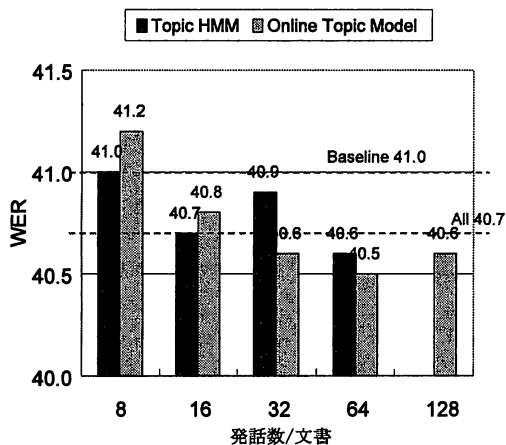


図6 WERと発話数/文書の関係（Consensus Decodingを用いた場合）。

を行う場合に対しても、話題の連続/不連続な変化に逐次追従する提案手法が良好な結果を示した。このことは、適切な話題の区切りで適応を行うことで、より高精度な音声認識が実現出来ることを示していると考えられる。

今後の課題として、音響モデルの逐次適応と併用し評価を行うことや、話題の変化点を検出することで、より適切な単位での言語モデル適応を行うこと、Online Topic Modelにおいて、どの程度トリガー・キャッシュモデルの効果が存在するか調査を行うことがあげられる。また、話題の連続的な変化、不連続な変化の双方を考慮出来るようなモデルについても検討を行いたい。

文 献

[1] M. Goto, J. Ogata, and K. Eto, "PodCastle: A Web 2.0 Approach to Speech Recognition Research," Proceedings of the 8th Annual Conference of the International Speech Communication Association (Inter-

speech 2007), pp.2397-2400, August 2007.

- [2] J. Glass, T. J. Hazen, S. Cyphers, I. Malioutov, D. Huynh, and R. Barzilay, "Recent Progress in the MIT Spoken Lecture Processing Project," INTER-SPEECH 2007, pp. 2553-2556 (2007).
- [3] 堀 貴明, 須藤 克仁, 大庭 隆伸, 渡部 晋治, 小川 厚徳, 渡辺 太郎, マクダーモット エリック, 塚田 元, 中村 篤, "「世界メディアブラウザ」 - 音声認識と統計翻訳に基づく多言語動画コンテンツ検索/閲覧システム," 日本音響学会講演論文集 1-1-17 (2008.9).
- [4] 渡部 晋治, 中村 篤, "巨視的な時間発展系に基づく逐次モデル適応-モデルの逐次更新における学習データの発話数に関する考察-", 電子情報通信学会研究報告 2007-SP-130, pp.201-206, 2007-12.
- [5] K. Ohtsuki, L. Nguyen, "Incremental language modeling for broadcast news," Automatic Speech Recognition and Understanding, 2005 IEEE Workshop on Volume, Issue, 27 Nov.-1 Dec. 2005 Page(s): 139 - 144
- [6] D. Beeferman, A. Berger, and J. Lafferty, "Model of Lexical Attraction and Repulsion," In Proceedings of the ACL-EACL'97 Joint Conference, Madrid, 1997.
- [7] D. Beeferman, A. Berger, and J. Lafferty, "Text segmentation using exponential models," In Proceedings of the Second Conference on Empirical Methods in Natural Language Processing, 1997.
- [8] A. Sako, T. Takiguchi, Y. Ariki, "Language Modeling Using PLSA-Based Topic HMM," IEICE TRANSACTIONS on Information and Systems, Vol.E91-D, No.3, pp.522-528, 2008.
- [9] 岩田 具治, 渡部 晋治, 山田 武士, 上田 修功, "トピックモデルに基づくユーザ興味の追跡," 第11回情報論的学習理論ワークショップ (IBIS 2008), 2008.
- [10] T. Nagano, M. Suzuki, A. Ito and S. Makino, "Language Modeling using Stochastic Switching N-gram," in Proceedings of the 18th International Congress on Acoustics, V, pp.3697-3700, 2004.
- [11] T. Kawabata, "Topic focusing mechanism for speech recognition based on probabilistic grammar and topic Markov model," in Proceedings of ICASSP-95, Vol. 1, pp. 317-320, 1995.
- [12] D. M. Blei and J. D. Lafferty. Dynamic topic models. In *ICML '06: Proceedings of the 23rd International Conference on Machine Learning*, pages 113-120, New York, NY, USA, 2006. ACM Press.
- [13] T. Hofmann, "Probabilistic Latent Semantic Analysis," in Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI'99), 1999.
- [14] T. Minka. Estimating a dirichlet distribution. Technical report, M.I.T., 2000.
- [15] S. Watanabe, A. Sako and A. Nakamura, "Automatic Determination of Acoustic Model Topology using Variational Bayesian Estimation and Clustering for Large Vocabulary Continuous Speech Recognition," IEEE transactions on Speech and Audio Processing, vol. 14, issue 3, pp. 855-872, 2006.
- [16] 中村 篤, 大庭 隆伸, 渡部 晋治, 石塚 健太郎, 堀 貴明, シュスターマイク, マクダーモット エリック, 南 泰浩 "音声認識システム SOLOON の日本語話し言葉コーパス (公開版 Ver.1.0) による評価," 第7回音声言語シンポジウム, 電子情報通信学会技術研究報告. SP, 音声, 105(496), pp.7-12, 2005.
- [17] L. Mangu, E. Brill, and A. Stolcke, "Finding consensus in speech recognition: word error minimization and other applications of confusion networks," Computer Speech and Language, 14, pp. 373-400, 2000.