

非負値行列因子分解に基づき動的適応した n -gram 言語モデルによるパープレキシティ削減効果の分析

藤田 悠哉^{1,a)} 奥 貴裕¹ 小林 彰夫¹ 佐藤 庄衛¹

概要: 情報番組のリスピーク音声の認識率改善のための、言語モデル適応手法について検討している。情報番組では複数の曖昧かつ細かい話題が遷移することから、確率的潜在話題解析 (pLSA) を拡張して、 K 単語連鎖生起確率分布に対して非負値行列因子分解 (NMF) を用いて潜在話題解析を行い、認識結果テキストを利用して言語モデルを動的に適応する手法を提案している。本報告では、パープレキシティによる評価実験を行い、提案法の効果を分析した。 $K = 2$ として動的適応した 2-gram 言語モデルで算出したパープレキシティを、ユニグラム・スケーリングにより適応した 2-gram 言語モデルで算出した値と比較した。実験結果から、評価テキストの種類により効果に差があるが、一定の条件下ではユニグラム・スケーリングと同等以上の性能が得られることが分かった。また、オラクルとの比較により、提案法は基底数の増加と共に過適応に陥る場合があるが、適応に用いる単語履歴長を長くすることで過適応回避可能なことがわかった。

キーワード: 確率的潜在話題解析, 動的適応, 言語モデル, 話題適応, 非負値行列因子分解

Analysis of dynamically adapted n -gram language models based on NMF

FUJITA YUYA^{1,a)} OKU TAKAHIRO¹ KOBAYASHI AKIO¹ SATO SHOEI¹

Abstract: For the purpose of improving WERs of re-spoken speech of a talk and variety show of TV, we have been studying a method of language model adaptation. Inside a program we are considering, there are many ambiguous topics and they transit as the program goes. In order to tackle such situation, we are proposing a method that extended pLSA by dealing probabilities of K -consecutive words generation with non-negative matrix factorization (NMF). Using past recognition results as adaptation text, we adapt the language model dynamically aiming to capture the transition of ambiguous topic within a program. In this paper, we analyzed experimental results that simulated speech recognition by testset perplexity of evaluation text. We set $K = 2$, and adapted 2-gram language model by past H -words history before the evaluating sentence, then calculated perplexity. The results showed that the effectiveness of proposed method varied depending on the type of evaluation text, but under some conditions proposed method yield equal or superior results compared to unigram scaling method. And by comparing the oracle, when we set a large value for the number of basis vector, proposed method tend to suffer overfitting. But by extending the length of word history used for adaptation, overfitting could be eased.

Keywords: Probabilistic latent semantic analysis, PLSA, Dynamic adaptation, Language model, Topic adaptation, Non-negative matrix factorization

1. はじめに

1.1 研究の背景

NHK では、聴覚障害者の方などに対する情報保証手段

¹ NHK 放送技術研究所
NHK Science and Technology Research Laboratories
^{a)} fujita.y-gc@nhk.or.jp

の1つとして、字幕放送拡充の取り組みを進めている。特に生放送番組に対してリアルタイムに字幕を付与する生字幕制作の方法の1つとして、音声認識技術を用いた手法の研究開発をNHK放送技術研究所で行っている[1]。既にスポーツ番組などでは、字幕専用のアナウンサー(字幕キャスター)が復唱した音声を認識するリスピーク方式で生字幕制作が行われている[2]。また、一部のニュース番組では、番組の音声を直接認識するダイレクト方式[3]と、リスピーク方式を組み合わせたハイブリッド方式[4]による字幕付与が行われている。なお、いずれの方式も認識誤りが放送されることのないよう、人手による修正を行っている。

リスピーク方式では、十分な認識率を得るために、放送前に放送内容にマッチした言語モデルを構築する必要がある。現状は現場の運用者が経験的に対応しており、字幕付与可能な番組を限定する1つの要因となっている。従って、放送前に入手可能な番組構成表の様な情報源から、自動的に放送内容にマッチした言語モデルを構築する事が望まれている。また、1つの番組内で多様な話題が出現する情報番組(ワイドショー)などにおいては話題に応じて言語モデルを動的に適応することで認識率の向上が期待される。本研究では、この言語モデルの動的適応手法について取り上げる。

適応する話題の例として、NHKで放送されている情報番組“あさいち”の番組ホームページ[5]に掲載されている放送予定をFig.1に示す。この中でも例えば9月27日のメインテーマ“JAPA”なび美人になるなら庄内へ行こう!の中では、庄内地方の水、野菜、温泉などが美肌効果と関連して取り上げられている。この様な番組における話題とは、ニュース番組における政治、経済の様に明確で大まかなものではなく、曖昧かつ細かいものと考えられる。曖昧な話題に言語モデルを適応するには、pLSA[6]の様な潜在話題に基づく手法が適していると考えられる。pLSAでは、あるテキストのユニグラム確率分布が話題に依存するという仮定のもと、そのユニグラム分布を複数のユニグラム基底分布の線形和で表現する。具体的には、話題に依存する様な潜在変数を z とおき、あるテキスト d のユニグラム分布 $p(w|d)$ を次式の様に近似するモデルをEMアルゴリズムで学習する。

$$p(w|d) \approx \sum_z p(w|z)p(z|d) \quad (1)$$

ある潜在変数 z に1つの話題が相当する様な制約は陽にないないので、 $p(z|d)$ の値により柔軟な話題の表現が可能となっている。言語モデル適応に用いる際は、式(1)の基底分布 $p(w|z)$ の張る空間に適応テキスト d' の分布 $p(w|d')$ を射影し、適応後のユニグラム分布 $p_{\text{PLSA}}(w|d')$ を得る。そしてベースとなる n -gram言語モデル $p_{\text{Base}}(\cdot)$ を次式に従いスケールアップする。

| 放送予定 PROGRAM | |
|-----------------|--|
| 9月26日(水) | |
| メインテーマ | 私がオチたワケ ~妻から夫への暴力~ |
| ピカピカ☆日本 | 道の駅 みま |
| 解決!ゴハン | 鶏のクリーム煮カレー風味 |
| 9月27日(木) | |
| メインテーマ | “JAPA”なび美人になるなら、庄内へ行こう! |
| ピカピカ☆日本 | 愛南町深浦市場の市場食堂 |
| 解決!ゴハン | 教えて!ポルチーニ活用術 |
| 9月28日(金) | |
| メインテーマ | プレミアムトーク 堤幸彦 |
| 特選!エンタ | イケメン!男性ヴォーカルグループ |
| あさいちガーデン | 解決!ガーデニング 教えて矢澤先生 |
| 10月1日(月) | |
| メインテーマ | 足腰を若く!40代からの“ロコモ”ケア |
| ピカピカ☆日本 | 秋津野ガルテン (あきずのがるてん) |
| 解決!ゴハン | 中国風ちまき弁当 |
| 10月2日(火) | |
| メインテーマ | スゴ技Q 体脂肪 撃退!おいしさ UP! えのきたけ 徹底使いこなし術 |
| ピカピカ☆日本 | 道の駅SanPin中津 |
| 解決!ゴハン | スピード混ぜごはん弁当 |

図1 NHKの情報番組“あさいち”の放送予定の例
 Fig. 1 Example of scheduled topic of a variety program broadcasted by NHK

$$p_{\text{adapt}}(w|w_{i-n+1}^{i-1}) \propto \frac{p_{\text{PLSA}}(w|d')}{p_{\text{Base}}(w)} p_{\text{Base}}(w|w_{i-n+1}^{i-1}) \quad (2)$$

適応テキスト d' には、過去の音声認識結果[7]や、講義スライドのテキストを用いる[8]試みが行われている。Fig.1の様な番組の音声認識の場合、事前に入手できる番組構成表を適応テキストとして、例えばその日の“メインテーマ”用のモデル、“解決!ゴハン”用のモデル、といった様に話題別の言語モデルを構築し、切り替えて使う事が考えられる[9]。また、情報番組の性質として、ある程度の時間同じ話題について話すことが多いため、過去の音声認識結果による適応も行い、局所的な話題に動的適応することで、認識率の向上が期待される。

1.2 提案法の狙い

一方、我々は K 単語連鎖生起確率分布が話題に依存すると仮定したモデルに基づき、非負値行列因子分解(Non-negative matrix factorization, NMF)[10]で潜在話題解析を行う手法を提案している[11],[12]。NMFは音楽情報処理[13]やドキュメントクラスタリング[14]にも応用され、

近年注目を集めているアルゴリズムである. pLSA と NMF はある条件下では等価であることが指摘されており [15], 提案法は pLSA を K 単語連鎖生起確率に拡張したものと捉えられることから, pLSA とのアナロジーで考えることができる.

提案法では単語出現確率に代わり K 単語連鎖生起確率を扱うため, 細かい詳細な話題を捉えられるのではないかと考えた. Web 上のテキストや放送済字幕など大量の学習テキストから先述の様な細かい話題を捉え, 適応できると期待した. そして, 単語連鎖数 K を, 適応する n -gram 言語モデルの次数 n より大きくすれば, スケーリングすることなくダイレクトに n -gram 確率が得られる. しかし, 一般的に単語連鎖生起確率を扱うモデルは変数が膨大となり, 計算量が問題となる. 我々はリアルタイム音声認識を想定しており, 実時間で動作する事が必須要件となる. これに対し, NMF は単純な行列演算のみから成るアルゴリズムのため, GPGPU 等による高速化が容易である. 加えて, NMF は行列演算という単純さを保ったまま各種拡張が可能であることも利点と考える. 例えば [16] では divergence を一般化した β -divergence を距離尺度に用いた場合に, あらゆる β で収束が保証された NMF アルゴリズムが提案されている.

なお, NMF による言語モデル適応化手法は [17] が存在するが, 学習テキスト中の単語間共起回数の期待値の行列を NMF で分解して適応に用いており, 本手法とは異なる.

1.3 本報告の内容

これまでの実験結果 [12] から, 提案法により適応した言語モデルによるテストセット・パープレキシティの値が適応無しの場合のベースライン言語モデルによる値より悪化する場面があった. その原因として, 1) 学習テキストの不足による基底の学習不足, 2) 適応に用いる単語履歴による過適応, の2つが考えられた. 本報告では, 1) に対して学習データを増量し, 2) に対しては適応に用いる単語履歴長を長くとする実験を行った. さらに, 評価テキストも含めて適応(オラクル)した実験も行い比較した.

以下, 2章で提案法の定式化を述べ, 3章で実験について記述する. そして4章で考察を述べたのち, 5章でまとめる.

2. NMF による言語モデル適応の定式化

我々が提案している NMF に基づく言語モデル適応手法 [11], [12] について説明する. この手法では, 学習テキスト毎の K 単語連鎖生起確率ベクトルを並べた行列を NMF で基底分解する. そして, その基底が張る部分空間に適応テキストの K 単語連鎖生起確率分布を射影したときの K 単語連鎖生起確率から適応化言語モデルを構築する.

2.1 基底系の学習

語彙サイズを S とおき, j 番目の語彙の単語を w_j と表

す. ある K 単語連鎖を $t_i = w_j w_k \cdots w_l$ とおく (i は K 単語連鎖のインデックスを示す). n 番目のテキストを d_n とすると, d_n における K 単語連鎖 t_i の出現確率は $p(t_i|d_n)$ と表す事ができる. この $p(t_i|d_n)$ を並べたベクトル

$$\mathbf{v}_n = [p(t_1|d_n), \dots, p(t_M|d_n)]^T \quad (3)$$

をテキスト d_n の特徴ベクトルとする. ここで M は学習テキスト中に出現する異なる t の数で, $M < S^K$ である. これをテキスト数 D 本並べた確率行列

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_D \end{bmatrix} \quad (4)$$

を NMF で基底分解する. 具体的には

$$\mathbf{V} \simeq \mathbf{W}\mathbf{H} \quad (5)$$

となる様な非負値の基底行列 \mathbf{W} と係数行列 \mathbf{H} を, 次式で表される KL-Divergence が最小となる様に反復的に求める.

$$D(\mathbf{V} \parallel \mathbf{W}\mathbf{H}) = \sum_{i,j} \left(\mathbf{V}_{ij} \log \frac{\mathbf{V}_{ij}}{(\mathbf{W}\mathbf{H})_{ij}} - \mathbf{V}_{ij} + (\mathbf{W}\mathbf{H})_{ij} \right) \quad (6)$$

ここで \mathbf{V}_{ij} は, 行列 \mathbf{V} の i 行 j 列目の要素である. これは再急降下法などで解くことも可能だが, NMF では次式の乗算更新アルゴリズムを用いる. なお, この更新式の下, 式 (6) は増加しないことが保証されている [18].

$$\begin{aligned} H_{a\mu} &\leftarrow H_{a\mu} \frac{\sum_i W_{ia} V_{i\mu} / (\mathbf{W}\mathbf{H})_{i\mu}}{\sum_k W_{ka}}, \\ W_{ia} &\leftarrow W_{ia} \frac{\sum_\mu H_{a\mu} V_{i\mu} / (\mathbf{W}\mathbf{H})_{i\mu}}{\sum_\nu H_{a\nu}} \end{aligned} \quad (7)$$

分解する因子の数(基底数)を R とし, 基底 \mathbf{W} と係数 \mathbf{H} を書き下すと, 分解の過程は次の様に表せる.

$$\mathbf{V} \simeq \mathbf{W}\mathbf{H} \quad (8)$$

$$= \begin{bmatrix} \mathbf{b}_1 & \cdots & \mathbf{b}_R \end{bmatrix} \times \begin{bmatrix} \mathbf{h}_1 & \cdots & \mathbf{h}_D \end{bmatrix} \quad (9)$$

ここで \mathbf{b}_r は, t の基底分布であり, 基底のインデックスを r とおくと

$$\mathbf{b}_r = [p(t_1|r), \dots, p(t_M|r)]^T \quad (10)$$

と書ける. \mathbf{h}_n は, 特徴ベクトル \mathbf{v}_n を表現するための, 基底に対する重みベクトルである. $R < D$ とすることで, \mathbf{W} の各列ベクトルの確率分布は, 次元圧縮した基底ベクトルと見なす事ができる.

NMF の特徴として, 基底 \mathbf{W} および係数 \mathbf{H} がスパースになるよう学習が進む [10]. この性質により, 特定の話題で特徴的に現れる t の分布が各基底に現れる事が期待される.

なお, 式 (7) の乗算更新アルゴリズムは, 行列の積および行列の要素毎の積, 除算のみで構成されているため, GPGPU 等による高速化が可能である [19]

2.2 適応テキストによる言語モデル適応

学習済の基底 \mathbf{W} が張る部分空間へ、適応テキストの K 単語連鎖生起確率分布を射影する。適応テキストを d' とおき、 d' の特徴ベクトル \mathbf{v}' を

$$\mathbf{v}' = [p(t_1|d'), \dots, p(t_M|d')]^T \quad (11)$$

とおくと、適応の手順は以下の様に定式化できる。

$$\mathbf{h}' = \underset{\mathbf{h}}{\operatorname{argmin}} D(\mathbf{v}' \parallel \mathbf{W}\mathbf{h}) \quad (12)$$

$D(\cdot \parallel \cdot)$ は2つの確率分布間の類似度を測る尺度で、今回は KL-divergence を用いた。これは最小二乗法などで最適解を求めることもできるが、 $\mathbf{W}\mathbf{h}'$ が非負値となる様、式 (7) の乗算更新アルゴリズムを、 \mathbf{W} を更新せずに適用することで求める。すると、 $\mathbf{W}\mathbf{h}'$ が適応後の K 単語組確率ベクトル

$$\mathbf{W}\mathbf{h}' = [p(t_1|\mathbf{h}'), \dots, p(t_M|\mathbf{h}')]^T \quad (13)$$

となる。 $t_i = w_j w_k \dots w_l$ であるので、このベクトルの各要素は K 単語連鎖の同時確率値とみなすことができる。このベクトルから適応化言語モデルを構築する。

適応する N -gram 言語モデルの次数が $N \leq K$ の時は、 $\mathbf{W}\mathbf{h}'$ から適当な計算により N 単語組のカウント値に変換し、一般の N -gram 言語モデル作成ツール (CMU-SLM Toolkit など) に与えることで適応化言語モデルが得られる。また、条件付き確率の定義 $p(x|y) = p(x, y)/p(y)$ に従って $\mathbf{W}\mathbf{h}'$ から直接 n -gram 確率 ($n \leq N$) を計算することもできる。

一方、 $N > K$ の時は、[6] におけるユニグラム・スケーリングをバイグラム以上の高次の場合にも一般化し、適応なしの N -gram 言語モデルの確率値を $\mathbf{W}\mathbf{h}'$ に応じてスケーリングし、適応化言語モデルを得る。

2.3 音声認識結果を用いた動的適応化

適応テキスト d' はタスクによって次の様なものが考えられる。例えば講義・講演の音声認識などでは、事前に得られるスライドや要旨といった認識対象とマッチしたテキストを d' として言語モデルを適応し、音声認識に用いる [8]。一方、認識対象とマッチしたテキストが事前に入手できない場合、過去の音声認識結果を d' として言語モデルを適応し、次の発話から利用する方法もある [7]。

情報番組の音声認識をタスクとする本報告では、ある話題が一定時間継続した後、遷移していくと想定されるため、発話直前の認識結果の話題に適応することで局所的な話題適応を行う。具体的には、評価する発話より前の認識結果の H 単語列を適応テキスト d' として言語モデルを適応する。ここで、ある K 単語連鎖 t が学習テキスト中に出現した回数 (カウント値) を $C(t)$ とおくと、 $t \in H : C(t) = 0$ なる t の確率は適応に用いることができないため、特徴ベク

トルを次式により正規化する。

$$\mathbf{v}' \leftarrow \frac{1}{\sum_{t \in H: C(t) > 0} p(t|d')} \mathbf{v}' \quad (14)$$

また、適応に用いられなかった t の確率値の和 $1 - \sum_{t \in H: C(t) > 0} p(t|d')$ を別の確率分布で置き換えて補間する。その様な方法の1つとして、 \mathbf{v}' に学習テキスト全体の t の確率分布を適宜補間する。具体的には、学習テキスト全体の特徴ベクトル

$$\mathbf{v}_B = [p(t_1), \dots, p(t_M)]^T \quad (15)$$

を用いて、 \mathbf{v}' を

$$\mathbf{v}' \leftarrow \lambda \mathbf{v}' + (1 - \lambda) \mathbf{v}_B \quad (16)$$

と置き換えて、式 (12) に基づき適応化言語モデルを得る。ここで、 λ の値は、

$$\lambda = \frac{\sum_{t \in H: C(t) > 0} p(t|d')}{\sum_{t \in H: C(t) > 0} p(t|d')} \quad (17)$$

とした。学習データに現れない t に相当する確率は基底 \mathbf{W} で表現できないので、学習データ全体の確率で補うという考え方である。

3. パープレキシティによる評価実験

3.1 実験条件

情報番組の音声認識を想定して、3種類の料理番組の放送済字幕を学習・評価テキストとし、テストセット・パープレキシティによる評価実験を行った。評価する発話より前の H 単語列を d' として適応した言語モデルでパープレキシティを算出した。 H 等のパラメータを Table 1 に示す。学習・評価テキストは、NHK の放送済字幕テキストから、電子番組表 (EPG) の分類区分により「グルメ・料理」に該当する 1322 番組 (3498k 単語)、10 番組 (17.4k 単語) を利用した。番組の種類は、料理解説 (Cooking (Long))、短編の料理解説 (Cooking (Short))、料理情報番組 (Cooking show) の3種類とした。それぞれの詳細を Table 2, Table 3 に示す。なお、評価テキストは [12] で用いたものと同じである。

この条件下では、それぞれのテキスト種類に対し、話題に相当する様な基底が学習され、適応に用いられることによりパープレキシティが削減されると期待した。例えば、Cooking(Long) の番組内容は、1回の放送で1つの食材を用いた料理の解説や、1つの料理の解説を行うことが多いため、食材または料理に依存する単語連鎖生起確率分布が基底に学習されると期待した。一方、Cooking(Short) も番組内容は Cooking(Long) と同様だが、放送時間が短く単語数が少ない。また、一人のキャラクターが寸劇の様に料理解説を行うため、特徴的な言い回しがある。この言い回しに相当する基底が学習されることを期待した。Cooking(Show)

表 1 実験パラメータ

Table 1 Parameter values for experiment

| | |
|---|--------------|
| Order of language model (N) | 2 |
| Order of adaptation (K) | 2 |
| History length used for adaptaion (H) | 60, 200, 400 |
| Smoothing method | Back-off |
| Discount method | Witten-Bell |

表 2 学習テキストの詳細

Table 2 Detail of training text

| Type of program | #Programs | #Words |
|-----------------|-----------|--------|
| Cooking (Long) | 735 | 2502k |
| Cooking (Short) | 404 | 212.3k |
| Cooking show | 183 | 783.2k |
| Total | 1322 | 3498k |

表 3 評価テキストの詳細

Table 3 Detail of evaluation text

| Type of program | #Programs | #Words | #OOVs |
|-----------------|-----------|--------|-------|
| Cooking (Long) | 5 | 13.5k | 22 |
| Cooking (Short) | 4 | 1.7k | 7 |
| Cooking show | 1 | 2.2k | 51 |
| Total | 10 | 17.4k | 80 |

は、料理の解説に加えて、例えば料理人をスタジオに招いてのトークや、料理にちなんだゲスト出演者によるトークなどが含まれている。従って、口語調の言い回しに相当する様な基底が学習される事を期待した。

なお、提案法では K 単語組を扱うので、話題に特徴的な n -グラム確率 ($n \leq K$) を表現できる。従って、学習テキストと評価テキストの分布傾向がマッチしている場合は、ユニグラム・スケーリングによる適応化言語モデルよりもパープレキシティが改善することが期待される。

本実験では、(a) 適応無しの言語モデル (Baseline), (b) Baseline をユニグラム・スケーリング [6] (適応アルゴリズムは NMF を利用) で適応した言語モデル (Unigram scaling), (c) 提案法で適応した言語モデル (Proposed) によるパープレキシティを比較した。加えて、評価テキストを含めて適応した言語モデルによるパープレキシティをオラクルと見立て、比較した。それぞれ (b') ユニグラム・スケーリングによる適応のオラクル (Unigram scaling (Oracle)), (c') 提案法による適応のオラクル (Proposed (Oracle)) と表記する。

3.2 実験結果

Fig. 2, Fig. 3, Fig. 4 に実験結果を、Fig. 5, Fig. 6, Fig. 7 にオラクルを示す。

Fig. 2 から、Cooking (Short) の評価テキストに対しては、履歴長 H が 200, 400 の場合は基底数に関わらず、提案法

が最良の値となった。また、基底数が少ないほど良い値であった。基底数を少なくできる事は計算コストの面からも好ましいことである。一方、履歴長 H が 60 の場合、基底数 12 以上ではユニグラム・スケーリングより悪い値となった。提案法は $K = 2$ 単語連鎖生起確率を元に適応するため、適応テキストにはある程度の単語数が必要と考えられる。

Fig. 3 から、Cooking (Long) の評価テキストに対しては、履歴長 H が 200, 400 かつ基底数が小さい場合、ユニグラム・スケーリングと提案法の差は僅かであった。

また、Fig. 4 から、Cooking show の評価テキストに対しては、提案法は効果が得られなかった。この評価テキストは他の評価テキストに比べて未知語率が一桁ほど高く、従ってバイグラムヒット率が低いため、 $K = 2$ 単語連鎖確率による適応には不向きと考えられる。

なお、評価テキストに関わらず、提案法は基底数の増加と共に悪化する傾向が見られた。悪化の程度は、履歴長 H が短いほど大きい。これは、ユニグラム・スケーリングでは特徴量ベクトルが語彙サイズ S の長さであるのに対し、提案法は K 単語連鎖の異なり数 S^K (実際は学習データに出現したものの数 $M < S^K$) の長さであり、特徴空間が広大となるため、履歴単語列に対して過適応となっている可能性がある。

4. 考察

実験結果から、提案法がユニグラム・スケーリングと同等以上の効果を発揮する必要条件は、

- (1) 履歴長 H を 200 以上にする
- (2) 基底数を 12 以下とする
- (3) ヒット率が十分な評価テキストであること

がわかった。基底数が小さく済むことは計算コストの観点から好ましいが、最適な基底数は学習テキストと評価テキストに依存すると考えられるため、基底数設定のための基準が必要となる。ヒット率の問題は、あらかじめ十分な学習テキストを用意することで回避できると考える。実際の情報番組へ適用する場合には、大量の学習テキストと番組構成表テキストを用いて、その日の放送内容の大まかな話題ごとに適応した言語モデルを元に動的適応することを想定しているので、あらかじめヒット率を十分保つことができると考える。

また、提案法は基底数の増加と共に悪化する傾向が見られた。オラクルでは Fig. 5, Fig. 6, Fig. 7 が示す様に基底数の増加と共に改善が確認できることから、過適応に陥っていることが分かった。過適応を回避できれば、基底数に関わらず提案法がユニグラム・スケーリングと同等以上の性能を発揮すると期待できる。

以上から、ヒット率を十分保つような工夫と過適応を回避することにより、提案法は安定してユニグラム・スケーリングより良い結果が得られると期待される。

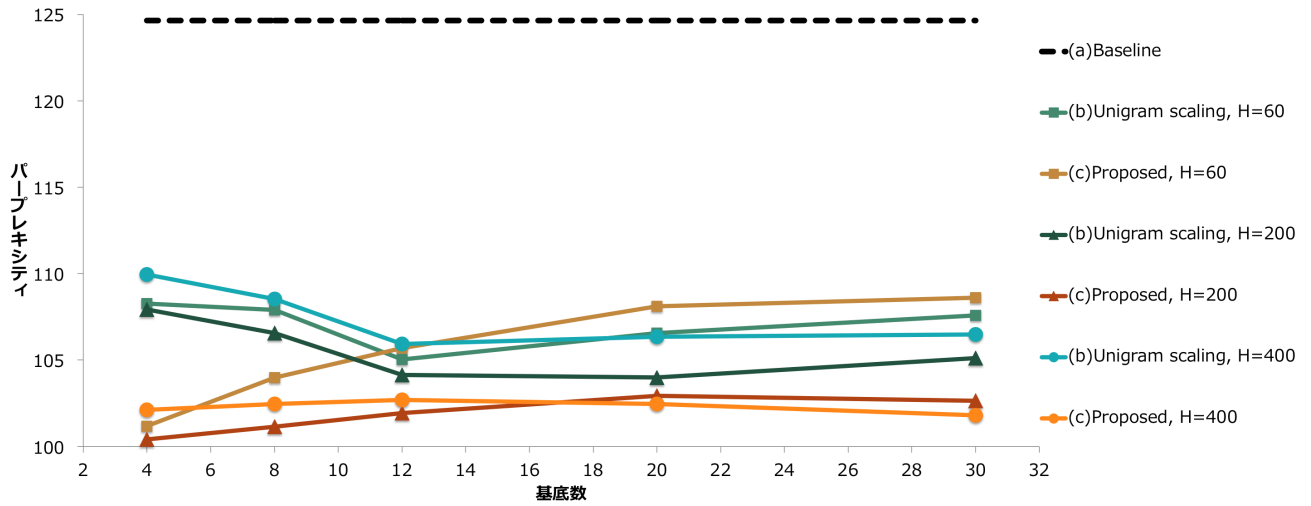


図 2 Cooking (Short) の評価テキストに対するパープレキシティ
 Fig. 2 Perplexity results for the text type of "Cooking (Short)"

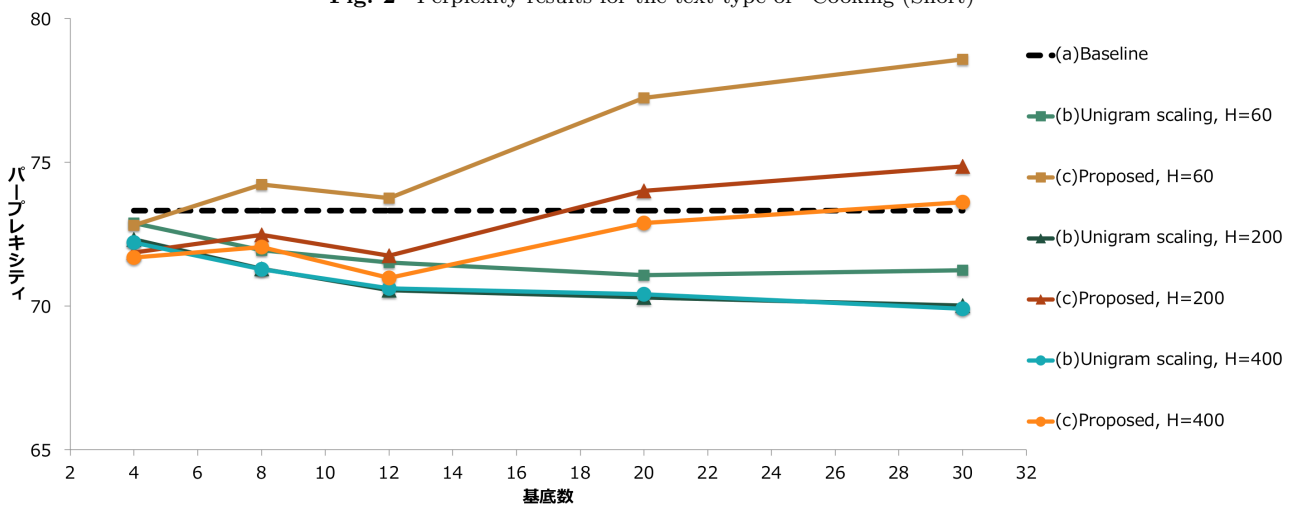


図 3 Cooking (Long) の評価テキストに対するパープレキシティ
 Fig. 3 Perplexity results for the text type of "Cooking (Long)"

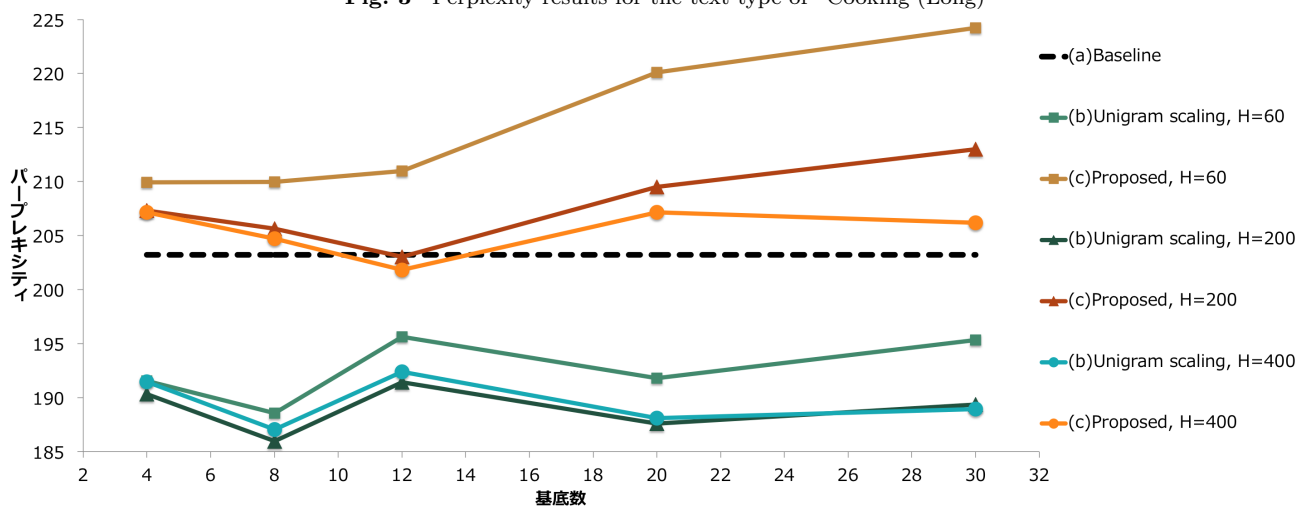


図 4 Cooking Show の評価テキストに対するパープレキシティ
 Fig. 4 Perplexity results for the text type of "Cooking Show"

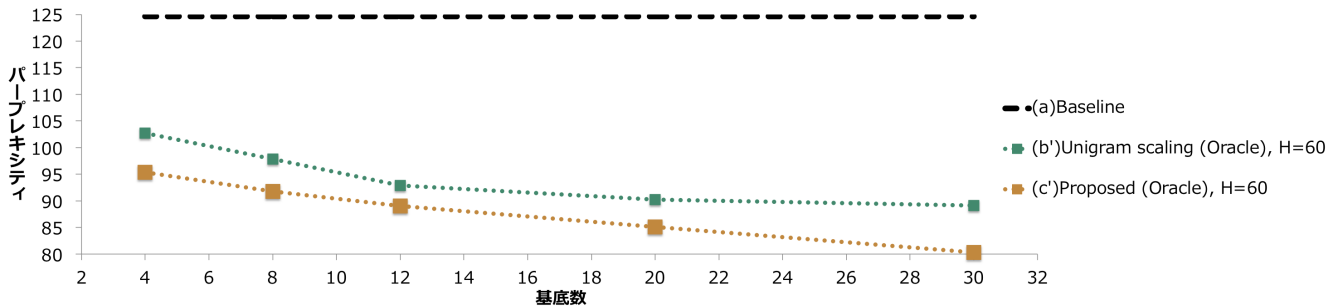


図 5 Cooking (Short) の評価テキストに対するパープレキシティのオラクル
 Fig. 5 Perplexity results for the text type of “Cooking (Short)” (Oracle)

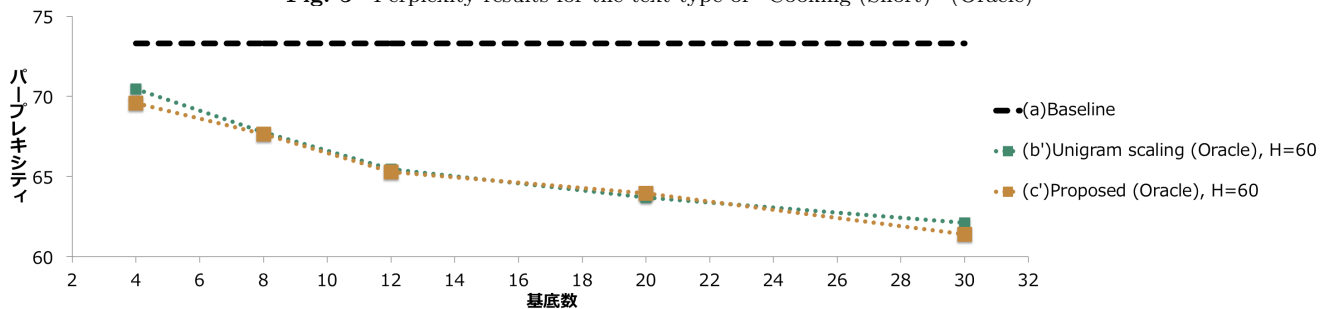


図 6 Cooking (Long) の評価テキストに対するパープレキシティのオラクル
 Fig. 6 Perplexity results for the text type of “Cooking (Long)” (Oracle)

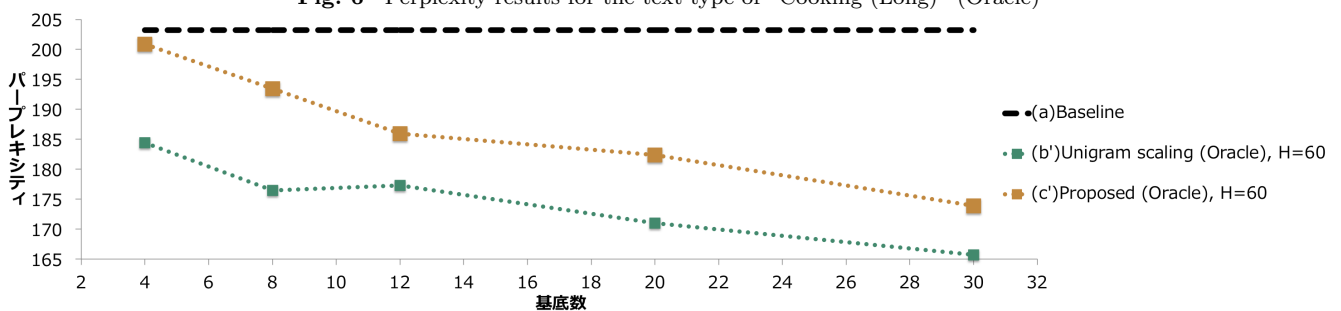


図 7 Cooking Show の評価テキストに対するパープレキシティのオラクル
 Fig. 7 Perplexity results for the text type of “Cooking Show” (Oracle)

5. おわりに

我々が提案している、 K 単語連鎖生起確率に NMF で潜在話題解析を行い、言語モデルを適応する手法について、認識結果テキストを用いた動的適応を想定した、パープレキシティによる評価実験を行った。基底数を小さくとり、適応に用いる履歴長を長くすることで、ヒット率が十分高い評価テキストに対し、提案法はユニグラム・スケーリングによる適応と同等以上の性能を示すことが確認できた。ヒット率が低い評価テキストにおいては、ユニグラム・スケーリングより性能が悪化するが、学習テキストを大量に収集し、あらかじめ十分なヒット率を保つことで回避できると考える。また、オラクルとの比較により、提案法は基底数の増加とともに過適応となる場合があることが確認された。これは履歴長を長くすることである程度緩和できることが分かった。

今後、ヒット率を十分保つための学習データ収集法の検討と、過適応の根本的な解決手段の検討を進めて行きたい。

参考文献

- [1] 今井亨, 奥貴裕, 小林彰夫: 音声認識によるリアルタイム字幕放送の進展, 情報処理学会研究報告. SLP, 音声言語情報処理, Vol. 2011, No. 4, pp. 1-6 (2011-10-21).
- [2] 松井淳, 本間真一, 小早川健, 尾上和穂, 佐藤庄衛, 今井亨, 安藤彰男: 言い換えを利用したリスピーク方式によるスポーツ中継のリアルタイム字幕制作, 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理, Vol. 87, No. 2, pp. 427-435 (2004-02-01).
- [3] 安藤彰男, 今井亨, 小林彰夫, 本間真一, 後藤淳, 清山信正, 三島剛, 小早川健, 佐藤庄衛, 尾上和穂, 世木寛之, 今井篤, 松井淳, 中村章, 田中英輝, 都木徹, 宮坂栄一, 磯野春雄: 音声認識を利用した放送用ニュース字幕制作システム, 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理, Vol. 84, No. 6, pp. 877-887 (2001-06-01).
- [4] 本間真一, 小林彰夫, 奥貴裕, 佐藤庄衛, 今井亨, 都木徹: ダイレクト方式とリスピーク方式の音声認識を併用した

- リアルタイム字幕制作システム, 映像情報メディア学会誌, Vol. 63, No. 3, pp. 331–338 (2009-03-01).
- [5] NHK, : NHK あさイチ 放送予定, <http://www.nhk.or.jp/asaichi/weekly/index.html> (2012).
- [6] Gildea, D. and Hofmann, T.: Topic-Based Language Models Using EM, in *Proceedings of EUROSPEECH 99*, pp. 2167–2170 (1999).
- [7] 中里理恵, 貞光九月, 富山良介, 山本幹雄, 板橋秀一: 生成文書モデルを用いた文書読み上げ音声認識, 情報処理学会研究報告. SLP, 音声言語情報処理, Vol. 2005, No. 69, pp. 19–24 (2005-07-15).
- [8] Kawahara, T., Nemoto, Y. and Akita, Y.: Automatic lecture transcription by exploiting presentation slide information for language model adaptation, in *Proceedings of 2008 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4929–4932 (2008).
- [9] 本間真一, 小林彰夫, 奥貴裕, 今井亨: 情報番組向け話題スイッチング言語モデルの検討, 日本音響学会研究発表会講演論文集, No. 3-P-9 (2011 秋).
- [10] Lee, D. D. and Seung, H. S.: Learning the parts of objects by non-negative matrix factorization, *Nature*, Vol. 401, pp. 788 – 791 (1999).
- [11] 藤田悠哉, 奥貴裕, 小林彰夫, 今井亨: 非負値行列因子分解に基づく話題性を反映した言語モデルの構築, 日本音響学会 研究発表会 講演論文集, No. 3-P-11 (2012 春).
- [12] 藤田悠哉, 奥貴裕, 小林彰夫, 佐藤庄衛: n-gram 言語モデルの非負値行列因子分解に基づく動的適応化, 日本音響学会 研究発表会講演論文集, No. 1-1-17 (2012 秋).
- [13] 中野允裕, 北野佑, ルルージョナトン, 亀岡弘和, 小野順貴, 嵯峨山茂樹: 可変基底 NMF に基づく音楽音響信号の解析, 情報処理学会研究報告. SIGMUS, 音楽情報科学, Vol. 2010, No. 10, pp. 1–6 (2010-02-08).
- [14] Shahnaz, F., Berry, M. W., Pauca, V. and Plemmons, R. J.: Document clustering using nonnegative matrix factorization, *Information Processing and Management*, Vol. 42, No. 2, pp. 373 – 386 (2006).
- [15] Gaussier, E. and Goutte, C.: Relation between PLSA and NMF and implications, in *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 601–602 (2005).
- [16] Nakano, M., Kameoka, H., Le Roux, J., Kitano, Y., Ono, N. and Sagayama, S.: Convergence-guaranteed multiplicative algorithms for nonnegative matrix factorization with β -divergence, in *2010 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, pp. 283 –288 (2010).
- [17] Novak, M. and Mammone, R.: Use of non-negative matrix factorization for language model adaptation in a lecture transcription task, in *Proceesings of 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Vol. 1, pp. 541 –544 vol.1 (2001).
- [18] Lee, D. D. and Seung, H. S.: Algorithms for Non-negative Matrix Factorization, in *Proceedings of the Conference on Advances in Neural Information Processing Systems (NIPS)*, Vol. 13, pp. 556–562 (2001).
- [19] Battenberg, E. and Wessel, D.: Accelerating nonnegative matrix factorization for audio source separation on multi-core and many-core architectures, in *10th International Society for Music Information Retrieval Conference (ISMIR)* (2009).