

## 文脈適応による複数 $N$ -gram の動的補間を用いた言語モデル

高橋 力矢<sup>†</sup> 峯松 信明<sup>†</sup> 広瀬 啓吉<sup>††</sup>

<sup>†</sup> 東京大学大学院情報理工学系研究科

<sup>††</sup> 東京大学大学院新領域創成科学研究科

E-mail: †{rikiya,mine,hirose}@gavo.t.u-tokyo.ac.jp

あらまし 本稿では、複数の  $N$ -gram を文脈適応により動的補間する言語モデルを提案する。提案モデル内の個々のサブ  $N$ -gram はそれぞれ異なる特定トピックの言語モデルである方が、全体性能が高くなる。このような特定トピックの言語モデルの構築にはテキストのクラスタリングが必要であるが、我々はクラスタリングを用いないモデル学習アルゴリズムを提案する。提案アルゴリズムは、文単位でトピックへの所属確率を与えることで、構築された言語モデルによる学習データのエンтроピーを最小化する。その結果、タグによる記事分類を基にした補間モデルよりも提案手法で学習した言語モデルの方が性能が高いことが示された。またオープンデータに対して、提案モデルはベースラインの  $N$ -gram よりも unigram 補間モデルで 12.7%、bigram 補間モデルで 4.3%、補正パープレキシティが低下した。

## A Language Model with Context Adaptability by Adaptive Interpolation of Sub $N$ -gram Models

Rikiya TAKAHASHI<sup>†</sup>, Nobuaki MINEMATSU<sup>†</sup>, and Keikichi HIROSE<sup>††</sup>

<sup>†</sup> Graduate School of Information Science and Technology, University of Tokyo

<sup>††</sup> Graduate School of Frontier Sciences, University of Tokyo

E-mail: †{rikiya,mine,hirose}@gavo.t.u-tokyo.ac.jp

**Abstract** This paper proposes a new language model (LM) implemented by adaptively controlling a mixture of sub  $N$ -gram models. The total performance of the proposed LM is higher when each sub  $N$ -gram model is topic-specific. Such topic-specific LMs generally need a text-clustering method but we propose a training algorithm without using a text-clustering method. The proposed training algorithm gives the topic-belonging-probabilities to each sentence in the training text, and minimizes the entropy of the training data calculated by the built LM. In our experiment, the LMs built by the proposed training algorithm are better performed than the LMs built by manually using the topic tag. The adjusted perplexity of the open data, by the proposed LMs, is lower than that by the traditional  $N$ -gram models. The decrease score is 12.7% in the unigram interpolation, and 4.3% in the bigram interpolation.

### 1. はじめに

本稿では、複数の  $N$ -gram を直前の文脈により動的補間する言語モデルを提案する。提案モデルは Probabilistic Latent Semantic Analysis [4] (以下 PLSA と略す) 言

語モデルを高次の  $N$ -gram について拡張したものである。PLSA ではパラメータ学習の際にテキストのクラスタリングを使用することでサブ  $N$ -gram のドメインを決定している。一方本稿では、クラスタリングを用いずに学習データから直接パラメータを推定する手法を提案する。

PLSA モデルは履歴単語列から各モデルの補間比を推定する。認識時の実際の使用を想定するとこの履歴は一文程度の長さである方がモデルとして意味がある。そのため学習時も文単位で補間比を変動させてパラメータを決定する。

## 2. 先行研究

大語彙音声認識で使われる言語モデルの主流は  $N$ -gram である。 $N$ -gram( $N = 2, 3$ ) の欠点として、離れた単語間の関連性を扱っていないこと、そのため単純な学習だけでは話題に応じた言語モデルが構築しにくいことがある。離れた単語間の依存関係を扱うには  $N$  を増加させれば良いが、単に  $N$  を増大させてもスパースネスの関係で可用性のない  $N$ -gram 対ばかり増えてしまう。

### 2.1 文脈による動的適応を用いた言語モデルの種類

前述した事項を踏まえ、 $N$ -gram を基本としつつ離れた単語間の依存関係をうまく圧縮して取り入れた言語モデルがいくつか提案されてきた。それらは次の二つに大別される。両者とも  $N$  より長い履歴に対して言語モデルを適応させる点では同一である。

$N$  より離れた単語対の依存関係を見るモデル キャッシュモデル [1]、トリガーモデル [2] がこれに該当する。 $N$  に関係なく、履歴内の単語と関連があると考えられる単語の出現確率を上昇させる。また動的適応モデルではないが、離れた単語間の関連性を扱った言語モデルは適応先タスクからある程度離れたタスクに対しても有効であることが報告されている [9]。

topic detection を基にするモデル あらかじめ複数の  $N$ -gram モデルを準備し、認識時の文脈を見てそれらのサブモデルを適宜使い分ける。その代表的なものが PLSA 言語モデルである。

### 2.2 文脈適応モデルの精度

文脈に対する動的適応モデルの一つの問題点は、ターゲット語彙によってパープレキシティの低下度にばらつきがあることである。トリガーモデルでは、実際には依存関係が最も強いのはその単語自身、次いで語幹を等しくする単語対、と報告されている [2]。topic detection を基にするモデルの場合、言語体系や補間比の推定方法によりパープレキシティの低下度も変わる。Broadcast News (BN94) コーパスを用いた英語 bigram で全体語彙に対して 10%、ターゲット語彙に対して 33% 低下すると報告されている [7]。毎日新聞を用いた研究では、ベースラインの trigram から語彙全体では 5% 程度、ターゲット語彙では 25% 程度パープレキシティが低下する [8]。

実際には、ある単語と関連のある単語群を広く予測する言語モデルの方が汎化能力が高いと考えられる。そこで本稿でも topic detection 型の言語モデルを掘り下げる。

## 2.3 PLSA 言語モデル

unigram を補間する PLSA 言語モデルはベースラインの unigram モデルに比べて、性能が大きく向上する。三品らによれば、毎日新聞一年分を用いた unigram 補間モデルはベイズ適応を用いるとベースラインの unigram に比べて 50% 以上パープレキシティが低下する (混合数 100 以上の場合) [4] [10]。

unigram を動的補間する場合は、「その話題によって出現しやすい単語」を予測している。一方、trigram を動的補間する場合は「その話題によって出現しやすい言語表現、言い回し」を予測すると考えられる。trigram は単語の連鎖に関する確率モデルである。単独の単語よりも、単語の連鎖の方が、ドメインによって違いが大きいと考えられる。そこで PLSA 言語モデルは高次  $N$ -gram モデルの補間においてより役立つのではないかと考えられる。(注1)

## 3. 言語モデリング

前述した背景を踏まえ、PLSA 言語モデルを高次の  $N$ -gram モデルに拡張する。式 (1) は提案モデルによる単語の出現確率を表す。 $m$  はサブ  $N$ -gram モデルに関する混合数、 $L$  は履歴の長さ。

$$\begin{aligned} P(w_n | w_{n-L}^{n-1}) &= \sum_{i=1}^m P(w_n | w_{n-N+1}^{n-1}, i) P(i | w_{n-L}^{n-1}) \\ &= \sum_{i=1}^m \theta_{n,i} \lambda_{n,i} \end{aligned} \quad (1)$$

それぞれの  $N$ -gram モデルにおける単語生起確率  $P(w_n | w_{n-N+1}^{n-1}, i)$  を  $\theta_{n,i}$  と表し、認識時の各時刻で、それぞれのモデルが選択される確率  $P(i | w_{n-L}^{n-1})$  を  $\lambda_{n,i}$  と表す。毎日新聞の場合一文あたり 20 形態素程度であることを考え、 $L$  の値を 10 ~ 20 程度にすると文の中でも動的適応の効果がある。

$\lambda_{n,i}$  は学習時に決定されるパラメータではなく、履歴  $w_{n-L}^{n-1}$  により適応・認識時に推定されるパラメータである。 $\lambda_{n,i}$  の推定はモデルパラメータ  $\{\theta_{n-k,i}; 1 \leq k \leq L, 1 \leq i \leq m\}$  を用いて行う。故に求めるべきパラメータ集合  $\Theta$  は、 $\Theta = \{\theta_{n,i}; 1 \leq n \leq C, 1 \leq i \leq m\}$  である。 $C$  は学習テキストにおける単語の総出現回数である。

以上のモデル化をした場合、学習テキストの平均エントロピーは式 (2) で表される。

(注1): パープレキシティの低下度は unigram の場合に及ばないが、精度向上する  $N$ -gram 対の数は比較的多いと考えられる。

$$Q(\Theta) = \frac{1}{C} \sum_{n=1}^C \log \left\{ \sum_{i=1}^m \theta_{n,i} \lambda_{n,i} \right\} \quad (2)$$

#### 4. 学習アルゴリズム

モデル学習には EM アルゴリズムを用いる。提案モデルが潜在変数  $i$  を含むためである。変分ベイズ学習を用いる方法もあるが、今回は実験していない。データ量が十分ある時には、最尤法と変分ベイズ法による結果にさほど違いがないためである。

##### 4.1 EM アルゴリズム

式 (2) における  $Q(\Theta)$  の単調増加を保証する EM アルゴリズムとして、図 1 に示した手順を用いる。

単語  $w_n (1 \leq n \leq C)$  に対し、 $w_n$  の所属する一文を  $s(n)$  とする。

$s(n)$  に所属する単語を  $w_{n'} (p \leq n' \leq q)$  とする。つまり  $s(n) = w_p^q$  である。

**E-step:**

$n$  が決定された場合のモデル  $i$  の支持確率  $P(i|n)$  を計算する。 $P(i|n)$  は  $s(n)$  が同一の文ならば同一の値である。

$$P(i|n) = P(i|s(n)) = \frac{\sum_{n'=p}^q \theta_{n',i} \lambda_{n',i}}{\sum_{i=1}^m \sum_{n'=p}^q \theta_{n',i} \lambda_{n',i}} \quad (3)$$

**M-step:**

それぞれの  $n$  に対し、パラメータ  $\theta_{n,i}, \lambda_{n,i}$  を求める。 $R_n$  は  $\forall n' \in R_n, w_{n'-N+1}^n \equiv w_{n'-N+1}^n$  となる  $n'$  の空間。 $W_n$  は  $w_{n'-N+1}^n$  と同一の単語列に対して、その後出現する異なり単語の空間。

$$\theta_{n,i} = \frac{\sum_{n' \in R_n} P(i|n')}{\sum_{w_{n'} \in W_n} \sum_{n' \in R_n} P(i|n')} \quad (4)$$

$$\begin{aligned} \lambda_{n,i} &= \frac{1}{L} \sum_{k=1}^L P(i, w_{n-k-N+1}^{n-k} | w_{n-k}) \\ &= \frac{1}{L} \sum_{k=1}^L \frac{\theta_{n-k,i} P(i|n)}{\sum_{i=1}^m \theta_{n-k,i} P(i|n)} \end{aligned} \quad (5)$$

図 1 サブ  $N$ -gram 学習のための EM アルゴリズム

図 1 に示した手順では、一文ごとに補間比  $P(i|s(n)) \equiv P(i|n)$  を学習している。実際には学習時に各単語レベルで補間比  $\lambda_{n,i}$  を割り振ることもできる。しかしこの場合、それぞれの時刻において、ある単語はモデル 1 にカウントされ、その次の単語はモデル 2 にカウントされるといった事態が起きる。これを繰り返すと品質の悪いモデルが構築されてしまう。そこで、補間比の変動を許す最小のフレームとして文を採択した。言語モデルの構築には未知語や存在しない  $N$ -gram 対へ確率値を割り振るため、スムージングを行う。今回の実験では Witten-Bell スムージングを用いる。図 1 に示した手順で構築された

$m$  種類の  $N$ -gram モデルに対し、図 2 に示すスムージングを行う。

- 式 (4) において、 $\sum_{n' \in R_n} P(i|n')$  はモデル  $i$  における  $N$ -gram 対の頻度に相当する。 $\sum_{n' \in R_n} P(i|n')$  に最も近い整数を  $C_n$  とする。
- $C_n$  を頻度として構築した  $N$ -gram モデルに通常の Witten-Bell スムージングを適用する。

図 2 構築されたサブ  $N$ -gram のスムージング手法

通常  $P(i|n)$  の収束値は特定の  $i$  に対して 1, その他の  $i$  に対して 0 となる。よって図 2 における、カウントの整数値への丸めは精度に支障をきたさない。

#### 5. 動的適応アルゴリズム

履歴単語列からのドメイン予測には、最尤推定またはベイズ推定を選ぶことが出来る。履歴単語列の長さは短いため、事前分布を仮定したベイズ推定の方が過適応せずに済む。ここでは三品らにより提案された、変分ベイズ学習を基にした推定方法を用いる。詳しくは文献 [10] を参照。

##### 5.1 ベイズ適応

ベイズ推定はパラメータの期待値を用いる。ここでは  $E[\lambda_{n,i}]$  を計算する。式 (6) において、 $\lambda_n = \{\lambda_{n,i}; 1 \leq i \leq m\}$

$$\begin{aligned} P(w_n | w_{n-L}^{n-1}) &= \int P(w_n | w_{n-N+1}^{n-1}, \lambda_n) P(\lambda_n | w_{n-L}^{n-1}) d\lambda_n \\ &= \sum_{i=1}^m \theta_{n,i} E[\lambda_{n,i}]_{P(\lambda_n | w_{n-L}^{n-1})} \end{aligned} \quad (6)$$

$P(\lambda_n | w_{n-L}^{n-1})$  は解析的には導出不可能である。そこで  $P(\lambda_n | w_{n-L}^{n-1})$  を別の関数  $Q(\lambda_n | w_{n-L}^{n-1})$  で近似する。ここでは *Dirichlet* 分布を  $\lambda_n$  の事前分布として導入する。 $\phi_0$  は超パラメータである。

$$P_D(\lambda_n; \phi_0) = \frac{\Gamma(m\phi_0)}{\Gamma(\phi_0)^m} \prod_{i=1}^m \lambda_{n,i}^{\phi_0-1} \quad (7)$$

$$\begin{aligned} Q(\lambda_n | w_{n-L}^{n-1}) &\propto P(w_{n-L}^{n-1} | \lambda_n) P_D(\lambda_n; \phi_0) \\ &\propto \prod_{i=1}^m \theta_{n,i}^{\phi_0 + \overline{N_{n,i}} - 1} \end{aligned} \quad (8)$$

$\overline{N_{n,i}}$  は  $n$  番目の単語が  $i$  番目のモデルから出現した回数。 $z_i^n$  は二値の潜在変数で、“ $n$  番目の単語が  $i$  番目のモデルから出現したかどうか”を 0/1 で表す。式 (9)(10) を繰り返し演算した収束値が  $\overline{N_{n,i}}$  の解を与える。式 (9) で、 $\Psi(x)$  は *digamma* 関数。 $\Psi(x) \equiv \Gamma'(x)/\Gamma(x)$

$$Q(z_i^n = 1 | w_{n-L}^{n-1}) \propto \exp[\Psi(\phi_0 + \overline{N_{n,i}})] \theta_{n,i} \quad (9)$$

$$\overline{N_{n,i}} = \sum_{k=1}^L Q(z_i^k = 1 | w_{n-L}^{k-1}) \quad (10)$$

最終的には、 $E[\lambda_{n,i}]$  は式 (9)(10) における収束値を基に、式 (11) で与えられる。この  $E[\lambda_{n,i}]$  が複数の  $N$ -gram モデルの動的補間比である。

$$E[\lambda_{n,i}] = \frac{\phi_0 + \overline{N_{n,i}}}{\sum_{i=1}^m (\phi_0 + \overline{N_{n,i}})} \quad (11)$$

三品らは PLSA モデルの学習時のパラメータから超パラメータ  $\phi_0$  を推定している。今回は簡単な洞察と事前実験からの経験則により  $\phi_0 = 0.5$  という値を採用した。式 (9) において、 $\exp[\Psi(\phi_0 + \overline{N_{n,i}})]$  を  $\overline{N_{n,i}}$  で置換すると EM 適応アルゴリズムとなる。 $L$  が十分大きいとき、ベイズ適応と EM 適応の結果は一致すべきである。そこで、

$$\lim_{x \rightarrow \infty} \exp[\Psi(x)] = x - \frac{1}{2} \quad (12)$$

を考慮し、 $\phi_0 = 0.5$  とした。

## 6. 実験

### 6.1 準備

学習データには毎日新聞を用いた。1996 年度版テキストを学習テキスト、1998 年度版テキストを評価テキストとした。語彙は、1991 年度から 1998 年度までの上位 20000 語を採択した。1994 年度版以降のテキストでは、各文に 17 種類の記事分類タグが付随している。このタグをモデル構築と評価テキスト生成の双方に用いた。

#### 6.1.1 モデル構築

図 1 に示した自動学習手順による言語モデルと手動でタグにより分類された 17 種の  $N$ -gram を補間する言語モデルの二種類を構築した。後者はタグによる記事分類をクラスタリングの結果に見立てたものである。

構築したモデルには全て Witten-Bell スムージングをかけた。自動学習した言語モデル内のサブ  $N$ -gram は、図 2 に示した手順を用いてスムージングした。

#### 6.1.2 評価テキスト

モデル評価においては、提案モデルがタスクドメインに追従できているかどうかを調べる必要がある。いつも同じ話題・文脈を保持している評価テキストは適切でない。そこで、話題がリアルタイムに変化するような評価テキストを用意する。記事分類タグに基づくサブテキストに対し、評価テキストを次の手順に従って生成した。

(1) 各サブテキストに対して文の読み取り位置を設ける。

(2) 無作為に一つのサブテキストを選び、連続する  $X$  文を採取する。

(3) (2) の後は文の読み取り位置が  $X$  文進んでいる。これを更に  $Y$  文進める。

(4) 求める文数が得られるまで (2)(3) を繰り返す。

それぞれの年度の新聞に対して、1000 文を生成した。表 1 は用意した評価テキストの種類と属性を示す。(1)(2) の  $X, Y$  は表 1 に従う乱数。

表 1 評価テキストの種類と性質

Name	Property
Slow	$1 \leq X \leq 10, 1 \leq Y \leq 10$
Fast	$1 \leq X \leq 3, 1 \leq Y \leq 10$
VeryFast	$X = 1, 1 \leq Y \leq 100$

また毎日新聞とは別に、ATR503 文も評価テキストとして用意した。ATR503 文は各文の関連性が薄く、文脈が極めて早く変化するテキストと考えられる。また学習テキストから見て ATR503 文は毎日新聞 1998 年度版テキストよりも離れたタスクドメインに所属すると考えられる。

### 6.2 実験結果

まず表 2 に、各評価テキストのベースライン言語モデルによる補正パープレキシティ  $APP_{base}$  及び総単語数  $WC$ 、未知語率  $OOV$  を示す。

表 2 ベースライン言語モデルによる補正パープレキシティ及び総単語数 ( $WC$ ) と未知語率 ( $OOV$ )

text	unigram	bigram	trigram	$WC$	$OOV$
Slow1996	777.3	81.78	28.55	25049	3.76%
Fast1996	794.3	85.75	30.15	24498	4.44%
VeryFast1996	788.3	81.48	28.47	25806	3.63%
Slow1998	759.0	99.79	74.95	25953	3.83%
Fast1998	755.2	98.07	73.09	26953	3.62%
VeryFast1998	749.6	96.77	73.95	26551	3.58%
ATR503	781.4	118.1	99.91	9296	6.57%

#### 6.2.1 提案学習方法の評価

EM アルゴリズムにより自動分割学習したモデル (以下 Auto と記す) とタグによる手動分割モデル (以下 Tag と記す) の比較を行った。表 2 に示した評価テキストのうち、クローズドデータとして VeryFast1996、オープンデータとして VeryFast1998 を採用した。補間比推定にベイズ推定を用いた場合の、ベースライン bigram に対する補正パープレキシティ  $APP$  の変化を図 3 に示す。Auto モデルの bigram 混合数は 8 とした。

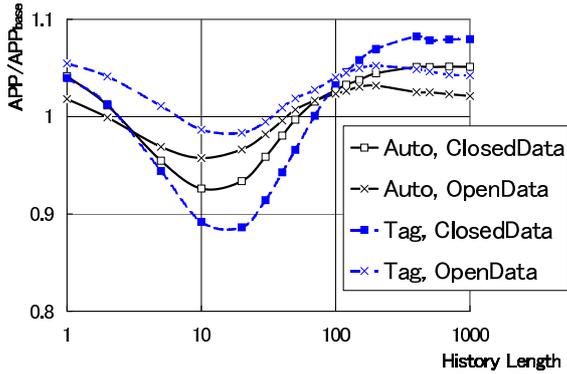


図3 自動分割モデルと手動分割モデルの比較

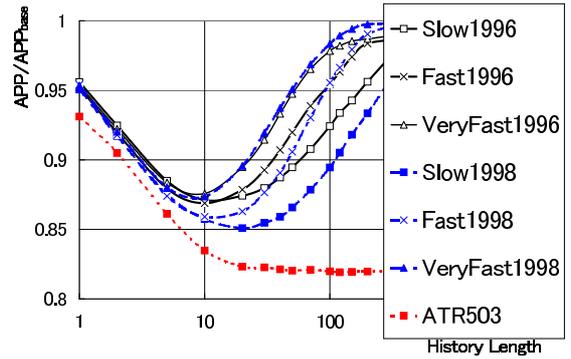
図3によると、手動分割モデルはクローズドデータに対しては自動分割モデルよりも有効であるが、オープンデータに対しては自動分割モデルの方が勝る。この傾向は Slow や Fast の評価テキストでも同じである。クローズドデータに関しては、その生成にタグを用いているので、手動分割モデルと相性が良いと考えられる。

いずれのタスクでも最小 APP を与える履歴の長さ  $L$  は 10~20 単語程度となっている。全体としての適応ではなく、動的適応が効いていることがわかる。

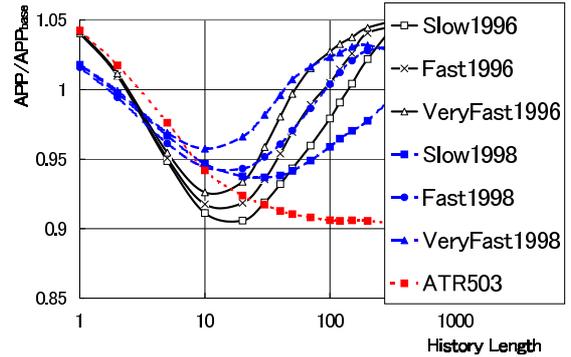
### 6.2.2 $N$ の値による精度の違い

混合数を 8 で固定した場合に、unigram 補間モデル、bigram 補間モデル、trigram 補間モデルがベースラインの  $N$ -gram からどの程度補正パープレキシティが変化するか調べた。図4に表1に示した全ての評価テキストに対する補正パープレキシティの変化率を示す。

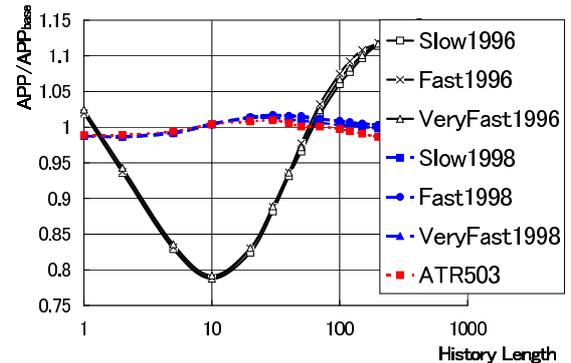
Slow の方が Fast や VeryFast といったデータよりも長い履歴で最適性能を得る。これは予測通りである。unigram 補間モデルや bigram 補間モデルではオープンデータに対して APP が低下しているが、trigram 補間モデルでは低下しなかった。つまり  $N$  の増大に伴い、オープンデータに対する性能が低下している。 $N$  が増えた場合に、提案モデルのような分割を基にした言語モデルは学習量が不足するためと考えられる。これに関しては、今後学習量を増やして様子を見ていく。ただ、trigram に関してクローズドデータの場合、履歴が 10 単語の時点で APP が 20% 低下している。つまり trigram 補間モデルは個々のドメインにおける単語の連鎖の違いについてある程度捕捉している。しかしその単語対はスパースなものなのでオープンデータには効力が低いと考えられる。



(a) unigram 補間モデル



(b) bigram 補間モデル



(c) trigram 補間モデル

図4 8 混合モデルにおける履歴の長さ と APP との関係

また ATR503 文に関しては、unigram 補間モデル、bigram 補間モデルでは  $L$  の増大とともに APP が単調減少している。タスクが学習データから離れた場合、動的適応ではなく、全体的に補間比を安定させての適応になってしまうことが判る。

### 6.2.3 最適混合数

今度は次数と履歴の長さを  $N = 2, L = 10$ (bigram) で固定し、混合数  $m$  の変化による補正パープレキシティの変化を調べ、図5に示した。

クローズドデータに対しては  $m = 8$  が最適で、オープンデータに対しては  $m = 4$  が最適である。ATR503 文の場合は  $m = 2$  の方が良く、学習データから離れるほど学習量の低下に敏感である。

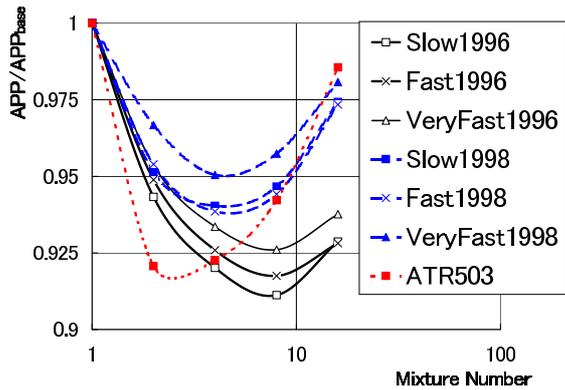


図 5 bigram 補間モデルの混合数と APP の関係

#### 6.2.4 アルゴリズムによる追跡能力の違い

提案モデルは Hofmann らの示した PLSA 言語モデルの学習アルゴリズムとは異なる。そのため最適な適応アルゴリズムも変化する可能性がある。そこで、EM 適応とベイズ適応による性能の違いについて実験した。8 混合 bigram 補間モデルに関して、自動学習モデルにおける EM 適応とベイズ適応の性能を比較したのが図 6 である。

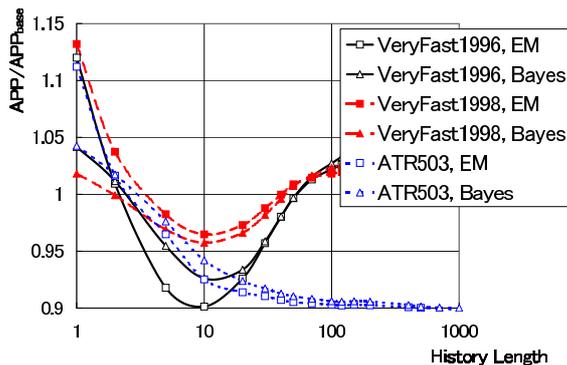


図 6 8 混合 bigram における EM 適応とベイズ適応の比較

履歴を 10 単語としたとき、オープンデータの VeryFast1998 ではベイズ適応が優れるが、クローズドデータ及びオープンデータの ATR503 文では EM 適応の方がやや APP が低い。提案する学習アルゴリズムにより構築されたモデルは、学習時に文毎の所属トピックがばらばらになるため各モデル間の差異が激しい。そこで補間比が変動しやすい EM 適応の方がトータルでは性能が若干高くなる。

しかし三品らも述べているように、EM 適応はベイズ適応に比べて、ベースライン言語モデルに対する性能が文によって勝ち負けが大きく変わる。そのため実際の使用を想定すると、認識時には EM 適応よりもベイズ適応によって補間比を推定する方が望ましいと考えられる。

## 7. ま と め

今回は複数の  $N$ -gram を文脈により動的補間する言語モデルとその学習アルゴリズムを提案した。その結果、話題変化が最も早く条件の厳しいオープンデータにおいても、8 混合 bigram 補間モデルで補正パープレキシティが 4.3%低下した (毎日新聞 1 年分学習の場合)。しかし trigram 補間モデルでは補正パープレキシティが低下しなかった。この原因は学習量の不足にあると考えられる。

今後はより大きなデータでの実験を行うとともに、図 1 に示した文レベルでのモデル学習よりも、より精度の高いモデル構築手法を確立する。具体的には、学習量が少なくとも高次の  $N$ -gram に適用できるようにする。そのために、学習時の補間比決定に関するフレームの長さについて、文以外の最適な単位を調査していく。また提案モデルの認識エンジンへの組み込みを行い、パープレキシティだけでなく、認識率への影響についても調べる。

### 参考文献

- [1] R.Kuhn, R De Mori, "A Cache-based Natural Language Model for Speech Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(6), pp.570-583, 1990.
- [2] R.Rosenfeld, "A Maximum Entropy Approach to Adaptive Statistical Language Modeling", Computer Science Department, Carnegie Mellon University, May 1996.
- [3] R.Rosenfeld, "A Hybrid Approach to Adaptive Statistical Language Modeling", *Proc. ARPA Workshop on Human Language Technology*, pp.76-87, 1994.
- [4] T.Hofmann, "Probabilistic Latent Semantic Indexing", in *Proceedings SIGIE*, pp.50-57, 1999.
- [5] D.Gildea, T.Hofmann, "Topic-Based Language Models Using EM", in *Proceedings EuroSpeech'99*, pp.2167-2170, 1999.
- [6] L.Saul, F.Pereira, "Aggregate and mixed-order Markov models for statistical language processing", in *Proceedings of the 2nd International Conference on Empirical Methods in Natural Language Processing*, 1997.
- [7] R.Florian, D.Yarowsky, "Dynamic Nonlocal Language Modeling via Hierarchical Topic-Based Adaptation", in *Proceedings of ACL'99*, pp.167-174, Jun 1999.
- [8] S.Kurohashi, M.Ori, "Nonlocal Language Modeling based on Context Co-occurrence Vectors", in *Proceedings of VLC'00*, pp.80-86, Oct 2000.
- [9] 広瀬啓吉, 峯松信明, 森谷 高明, "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", *情報処理学会論文誌*, Vol.43, No.07, pp.2065-2074, Jul 2002.
- [10] 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", *情報処理学会音声言語情報処理研究会*, in *Proceedings SIG-SLP*, pp.13-18, Dec 2002.
- [11] 高橋力矢, 峯松信明, 広瀬啓吉, "複数サブモデルの動的補間を用いた言語モデル", *日本音響学会春季講演論文集*, pp.137-138, Mar 2003.