

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

長友 健太郎[†] 西村 竜一[†] 小松 久美子^{††}
 黒田 由香^{†††} 李 晃伸[†]
 猿渡 洋[†] 鹿野 清宏[†]

高精度な言語モデルの融合手法として、相補的バックオフアルゴリズムに基づく融合アルゴリズムを提案するとともに、それを用いた言語モデルの融合ツールを構築した。N-gram 言語モデルは、学習元のコーパスの話題や知識、語調や発話様式などの特徴を反映する。そのため、タスクごとの特徴を反映した複数の言語モデルを融合することで、より多様な入力に対処できるモデルを構築できる。この言語モデルの融合において、既存の融合手法では、モデルの持つ特性が損なわれるためタスクに対する特徴がばやけてしまう。また、従来手法である学習元コーパスの単純な結合および再学習による融合を行うためには、学習元のコーパス自体が必要になる。これに対して、他方のモデルには現れない未観測 N-gram の生起確率を他方のモデルから相互に推定する高精度な相補的バックオフアルゴリズムを提案する。さらに本手法を用いて、学習元コーパスが不要で利便性の高い言語モデル融合ツールを構築した。実際に医療相談、グルメ・レシピ検索および新聞記事の各タスクの言語モデルを融合し、それらを評価した結果、各モデルの特性をなるべく保存しながら、コーパス結合モデルと比較しても精度が劣化しないモデルを得ることができた。

Complemental Back-off Algorithm for Merging Language Models

KENTARO NAGATOMO,[†] RYUICHI NISIMURA,[†] KUMIKO KOMATSU,^{††}
 YUKA KURODA,^{†††} AKINOBU LEE,[†] HIROSHI SARUWATARI[†]
 and KIYOHITO SHIKANO[†]

A new complemental back-off algorithm for merging two N-gram language models is proposed. By merging several topic-dependent or style-dependent models, we can construct a general model that covers wider range of topics easily. However, a conventional method that simply concatenates the training corpora or interpolating each probabilities often levels off the task-dependent characteristics in each language models, and weaken the linguistic constraint in total. We propose a new back-off scheme that assigns the unseen N-gram probabilities according to the probabilities of the another model. It can assign more reliable probabilities to the unseen N-grams, and no original corpora is needed for the merging. We implemented a command tool that realizes this method, and evaluated it on three recognition tasks (medical consulting, food recipe query and newspaper article). The results reveal that our merged model can keep the same accuracy of each original one.

1. はじめに

N-gram 言語モデルは大語彙連続音声認識システム

における言語モデルとして広く用いられている。この単純で扱いやすいモデルの性能は学習テキスト(コーパス)の質と量に依存するため、より高精度なモデルの構築には大量の整ったコーパスが必要になる。

我々が収集するコーパスは、話題、語尾様式(である、ですますなど)、発話様式などの特徴を持っており、そのコーパスから構築される言語モデルもその特徴を引き継ぐ。音声認識システムでは入力音声と言語モデルの特徴に近いほど高い認識率を得ることができる。この性質を利用して、システムが扱う入力を特定タスクや語調に限定すれば、高い認識性能を持つシステムを構築することができる。しかし、特定の発話の

[†] 奈良先端科学技術大学院大学情報科学研究科
 Graduate School of Information Science, Nara Institute
 of Science and Technology

^{††} 財団法人イメージ情報科学研究所
 Laboratories of Image Information Science and Tech-
 nology

^{†††} TIS 株式会社
 TIS Inc.
 現在、日本電気株式会社
 Presently with NEC Corporation

みを対象としたシステムは受理される発話パターンが少なくなり、ユーザにとって不便なものとなる。

例として音声で操作するウェブブラウザシステムを考える。このシステムが扱う発話は、(a) ブラウザそのものの操作、(b) ウェブにおいて一般的な用語、(c) ユーザの個人的興味の対象に特有の言葉、のように多様な範囲に及ぶ。(b) や (c) はさらにさまざまなドメインを内包する。ユーザはその広範なドメインのすべての入力音声認識されることを期待するだろう。このように音声認識システムではいくつかの異なる特徴を持つタスクの音声の認識が要求される。

もう1つの例として音声ワープロを考える。今日の音声ワープロは、新聞記事の読み上げなどでは高い認識率を示すが、会話音声などのくだけた文については不十分である¹⁾。これは音声ワープロの言語モデルが書き言葉らしい特徴を持つのに対し、会話音声の話言葉の特徴を持つためである。

より実用的な認識システムを実現するためには、このように広いドメインの知識や発話現象をカバーする汎用な言語モデルが不可欠である。しかし、すべての発話現象をカバーした大量のテキストを収集するのは不可能である。また、多様なコーパスを整った形で収集するには、限られたドメインのコーパスを整備するよりも多大な労力が必要となる。さらに、新しい固有名詞の出現や単語の用法の変化など、話し言葉は時間とともに変化しており、そのような時事的な言語特徴の変化まで完全にカバーするコーパスを収集するのは困難である。

そこで異なる特徴を持った複数の言語モデルの融合技術が求められている。モデルの融合によって、さまざまな異なる特徴を持つ言語モデルを集め、これらの特徴を任意に組み合わせることで、言語モデルの特徴を制御することが可能になる。

最も単純な融合手法は、それぞれの学習元コーパスをつなぎ合わせて再学習する方法(コーパス結合)である。より扱いやすい方法としては、各モデルにおける N-gram 確率を重み付きで内挿する方法が研究されており、特に少量コーパスによる学習データ不足の補間や言語モデルのタスク適応などに用いられている^{2),3)}。

しかし、単純な N-gram 確率の融合では、言語モデル間の未観測な N-gram エントリの不整合の問題が生じる。あるタスクにとって特徴的な N-gram は、他方にとっては未観測であることが多い。特に、固有名詞などの一方にしか現れない語は他方にとっては未知語であり、それらの N-gram 確率も他方にとって未観

測である。これら未観測の N-gram に対して、通常のバックオフ手法を用いて (N-1)-gram から推定することは可能であるが、他タスクに特有の単語に対してそのモデル内の情報のみから正しい確率を割り当てることは難しい。このような信頼性の低い確率に基づいて融合を行うことは、融合後のモデルの精度低下を引き起こす。また、そのようにして融合相手のモデルにおいて推定された未観測 N-gram 確率に基づいて融合を行うことで、融合前のモデルが持つ N-gram 確率の分布が平坦化されてしまい、結果として、タスクに依存したモデルの特徴が薄らいでしまう。

本研究では、言語モデルの融合において不整合を起こす N-gram の出現頻度を相補的なバックオフを用いて推定する手法を提案する。互いの言語モデルに未観測な N-gram を融合相手のモデルから相補的に推定することで、タスクごとの特徴を反映しつつ、より高精度な融合を行うことができる。

さらに、このアルゴリズムを実装した言語モデル融合ツールを構築する。コーパス結合は、巨大な学習元コーパスを保持し、それを元にモデルの再学習をする必要があるため、融合処理が複雑なものになり利便性に欠ける。また、既存の学習ツール^{4),5)} では、各モデルの学習過程で得られた頻度情報を用いて結合・再学習を行う工夫がなされているが、モデル自身以外の頻度情報が必要なのは好ましくない。また、実際のツールでは、両者の語彙が完全に同一であることを前提としており、コーパスを直接保持する場合と同様な制約を受ける。本研究で提案する相補的バックオフを用いた融合ツールでは、元コーパスや頻度情報ファイルを用いずに容易に高精度な言語モデルを融合することが可能となる。

以下、2章で言語モデルの融合について、3章で未観測 N-gram の確率の推定方法について述べる。4章で提案手法である相補的バックオフに基づく N-gram の融合について述べる。提案手法を実装した融合ツールの評価を5章で述べ、6章で本論文のまとめを行う。

2. 言語モデルの融合

言語モデルの融合は、各モデルに含まれるそれぞれの N-gram の出現頻度の重み付き和をとることで行われる。以下の議論では、簡単のため、融合する言語モデルを2つに限定する。基になるコーパスをそれぞれ F, G と表し、それぞれから構築された2つの言語モデルを L_f, L_g とする。このとき、ある N 単語の組 w_{i-N+1}^i の融合後のモデルにおける出現頻度 $C(w_{i-N+1}^i)$ は、

$$C(w_{i-N+1}^i) = \lambda_f C_f(w_{i-N+1}^i) + \lambda_g C_g(w_{i-N+1}^i) \quad (1)$$

となる。ここで $C_f(w_{i-N+1}^i)$ はコーパス F における出現頻度、 $C_g(w_{i-N+1}^i)$ はコーパス G における出現頻度である。また、 λ_f および λ_g は、 $\lambda_f + \lambda_g = 1$ の重み係数である。

N-gram 言語モデルでは、ある N 単語の組 w_{i-N+1}^i において w^i の出現する条件付き確率 $P(w^i | w_{i-N+1}^{i-1})$ をコーパスに出現する N-gram の頻度 $C(w_{i-N+1}^i)$ を用いて次式のように求める。

$$P(w^i | w_{i-N+1}^{i-1}) = \frac{C(w_{i-N+1}^i)}{C(w_{i-N+1}^{i-1})} \quad (2)$$

これより、融合後のモデルにおける出現確率は

$$\begin{aligned} P(w^i | w_{i-N+1}^{i-1}) &= \frac{C(w_{i-N+1}^i)}{C(w_{i-N+1}^{i-1})} \\ &= \frac{\lambda_f C_f(w_{i-N+1}^i) + \lambda_g C_g(w_{i-N+1}^i)}{\lambda_f C_f(w_{i-N+1}^{i-1}) + \lambda_g C_g(w_{i-N+1}^{i-1})} \end{aligned} \quad (3)$$

となる。

通常、言語モデルには出現確率とバックオフのための情報のみが保持され、頻度情報は含まれない。この場合でも、コーパスに含まれる総単語数 $C(*)$ が分かれば、 $C(w_{i-N+1}^i)$ は、

$$\begin{aligned} C(w_{i-N+1}^i) &= C(*) P(w^{i-N+1}) \dots \\ &P(w^{i-1} | w_{i-N+1}^{i-2}) P(w^i | w_{i-N+1}^{i-1}) \end{aligned} \quad (4)$$

として計算できる。式 (3) および式 (4) から、融合においては各コーパスの大きさ $C_f(*)$ 、 $C_g(*)$ が必要となるが、実際にはそれらの比を考慮した重み λ_f および λ_g を与えればよい。

3. 未観測 N-gram 確率の推定

言語モデルにおいて、学習時に現れなかった N-gram を未観測 N-gram と呼ぶ。この未観測 N-gram の出現頻度または出現確率の推定について述べる。

未観測の N-gram について正しい出現確率を推定することは難しい。しかし、コーパスで観測されない未観測 N-gram の出現確率の合計は、経験的または discounting method などの統計的手法によって推定される。すなわち、言語モデルの学習時にすべての確率値を出現 N-gram に割り振らずに、あらかじめ確率値を割り引いておき、その確率値を未観測 N-gram 集合の確率の推定値として割り当てる。

ある言語モデルが与えられたとき、そのコンテキスト w_{i-N+1}^{i-1} の未観測確率値を $P(* | w_{i-N+1}^{i-1})$ と表す。

あるコンテキストに属するすべての N-gram の出現確率の和は 1 であるから、この値は

$$\begin{aligned} P(* | w_{i-N+1}^{i-1}) &= 1 - \sum_{C(w_{i-N+1}^i) > 0} P(w^i | w_{i-N+1}^{i-1}) \end{aligned} \quad (5)$$

として算出できる。

ある 2 つのモデルの融合を考えると、一方のモデルで観測されているが他方では未観測である N-gram、および双方で未観測である N-gram を考慮する必要がある。融合時には、前者の互いに未観測な N-gram については確率の推定を行いその推定値を元に融合を行うが、後者は融合時には未知であるので、融合後のモデルにおいて確率を推定することとなる。このため、融合時に互いに未観測 N-gram の確率を推定する際には、融合後に未観測な N-gram 集合に対して割り当てる確率値を残しておく必要がある。すなわち、融合において推定した確率の合計が、上記の未観測確率値 $P(* | w_{i-N+1}^{i-1})$ よりも小さくなければならない。この推定するすべての未観測 N-gram の確率の総和が、もとの未観測確率値に占める割合を $\beta(w_{i-N+1}^{i-1})$ で表す。 β の推定は未観測確率集合の再計算を意味する。

未観測 N-gram の確率 $\hat{P}(w^i | w_{i-N+1}^{i-1})$ を求める際には、この $\beta(w_{i-N+1}^{i-1}) P(* | w_{i-N+1}^{i-1})$ を統計情報に従って分配する。その分配率を $\gamma(w^i | w_{i-N+1}^{i-1})$ とすると、この未観測 N-gram の出現確率は次式で表される。

$$\begin{aligned} \hat{P}(w^i | w_{i-N+1}^{i-1}) &= \gamma(w^i | w_{i-N+1}^{i-1}) \beta(w_{i-N+1}^{i-1}) P(* | w_{i-N+1}^{i-1}) \end{aligned} \quad (6)$$

未観測 N-gram 確率の推定は、この 2 つのパラメータ β と γ をいかに設定するかという問題に置き換えられる。通常の単一コーパス内での未観測 N-gram 推定では、 β を求めることは未観測確率集合を推定することと同じになるので、よく知られる Witten-Bell 法⁶⁾ や Good-Turing 推定⁷⁾ などを用いることができる。また、 γ の推定には、(N-1)-gram や近似 N-gram からのバックオフ推定が利用できる。

これらの推定法は、いずれも未観測 N-gram を含むモデル自身による推定法である。そのためこれらの推定によって得られた確率値は、そのモデルから見たときにもっともらしい値となる。

4. 相補的バックオフに基づく N-gram 融合

あるタスクでのみ観測される N-gram は、そのタスクの言語的特徴をよく表す要素であるといえる。すなわち、ある N-gram が一方のタスクでよく現れるに

もかわらず別のタスクでは未観測であるということは、その N-gram がタスクの知識を反映する重要な要素であると考えられる。従来の融合では、この未観測な N-gram の相互関係について考慮しておらず、それぞれのモデルが持つ特徴的な N-gram 確率の分布からなる特性が平坦化され、モデルの特徴を決定づける要素が互いに打ち消しあう結果となっていた。このことは、たとえば少量の書き起こしコーパスから作ったモデルと新聞記事から作ったモデルを融合するといったような、特性の偏ったモデルを他のモデルと融合して強化する用途では有効であるが、複数のタスクを同時に認識するといった、それぞれのモデルの特性を残しつつ両タスクに高い精度を得たい場合には問題となる。これまでの言語モデル融合に関する研究では、各言語モデルでの N-gram 確率を重み付きで内挿する方法が主に研究されており²⁾、このような相互に未観測 N-gram の確率については考慮されていない。

本研究では、未観測 N-gram の推定に際して、その N-gram が実際に観測された他方のモデルでの N-gram 分布を参考にする推定法を提案する。この手法では、未観測 N-gram 確率の推定値は融合相手のモデルのものと同じ傾向を持つ。言語モデルの特徴がモデルに特有の N-gram 確率の分布であるモデルの特性に現れているとするならば、このアルゴリズムを用いることで、モデルの特徴を保存したモデルの融合が期待できる。この推定法では、2 つのモデルが互いの未観測 N-gram を相互に補完し合うと考えることができるので、本手法を相補的バックオフと名付ける。

相補的バックオフはスキームであり、具体的な実現方法はさまざまなものが考えられる。本論文では後述する言語モデル融合ツールの実装に利用するアルゴリズムについて述べる。便宜上、言語モデル L_f のあるコンテキスト w_{i-N+1}^{i-1} に属する未観測 N-gram の推定に限定して述べるが、そのほかの場合でも同様である。

4.1 パラメータ β の推定

パラメータ β の推定法を図 1 に示す。言語モデル L_f の未観測確率集合 $P_f(*|w_{i-N+1}^{i-1})$ に占める未観測 N-gram の確率の総和の大きさ β_f は、モデル L_g の当該 N-gram の確率の総和とモデル L_g の未観測確率集合の比で、以下の式を用いて推定できる。

$$\beta_f(w_{i-N+1}^{i-1}) = \frac{\sum_{C_g(w_{i-N+1}^i) > 0} P_g(w^i | w_{i-N+1}^{i-1})}{\sum_{C_g(w_{i-N+1}^i) > 0} P_g(w^i | w_{i-N+1}^{i-1}) + P_g(* | w_{i-N+1}^{i-1})} \tag{7}$$

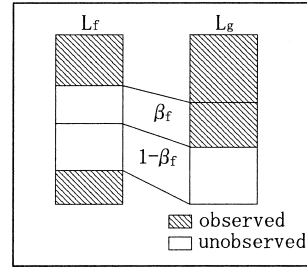


図 1 β の推定
Fig. 1 Estimation of β .

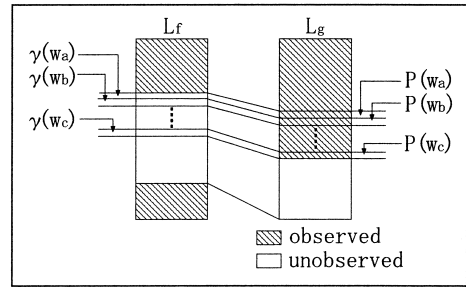


図 2 γ の推定
Fig. 2 Estimation of γ .

ただし、このとき、 $C_f(w_{i-N+1}^i) = 0$ である。

4.2 パラメータ γ の推定

相補的バックオフにおけるパラメータ γ の推定法を図 2 に示す。言語モデル L_f のすべての未観測 N-gram の確率の総和に占めるある未観測 N-gram w_{i-N+1}^i の出現確率の割合は、モデル L_g における当該 N-gram の出現確率の総和に占める w_{i-N+1}^i の出現確率の割合に等しいと仮定する。すなわち、

$$\gamma_f(w^i | w_{i-N+1}^{i-1}) = \frac{P_g(w^i | w_{i-N+1}^{i-1})}{\sum_{C_f(w_{i-N+1}^i) = 0} P_g(w^i | w_{i-N+1}^{i-1})} \tag{8}$$

となる。

4.3 融合確率の計算手順

ある N-gram w_{i-N+1}^i の融合確率の計算の手順は以下ようになる。

- (1) モデル L_f で観測されず、モデル L_g でのみ観測される N-gram をすべて調べ、推定パラメータ $\beta_f(w_{i-N+1}^{i-1})$ を求める。
- (2) 同様に、推定パラメータ $\beta_g(w_{i-N+1}^{i-1})$ を求める。
- (3) モデル L_f における確率 $P_f(w^i | w_{i-N+1}^{i-1})$ を求める。もし N-gram が L_f にとって未知ならば、推定パラメータ $\gamma_f(w^i | w_{i-N+1}^{i-1})$ を求め、式 (6) より推定確率 $\hat{P}_f(w^i | w_{i-N+1}^{i-1})$ を求める。

- (4) モデル L_g における確率 $P_g(w^i | w_{i-N+1}^{i-1})$ を同様に求める。もし N-gram が L_g において未知ならば、推定パラメータ $\gamma_g(w^i | w_{i-N+1}^{i-1})$ を求め、式 (6) より推定確率 $\hat{P}_g(w^i | w_{i-N+1}^{i-1})$ を求める。
- (5) 両モデルにおける確率あるいは推定確率を用いて、融合後の確率 $P(w^i | w_{i-N+1}^{i-1})$ を式 (4) および式 (3) より求める。
- (6) 手順 (3) から手順 (5) までを、コンテキストに属するすべての観測 N-gram について繰り返す。なお、いずれのモデルでも未観測な N-gram については、融合後のモデルも通常のバックオフを用いて出現確率値を計算することになる。ただし、1-gram の場合は、融合後のすべての 1-gram 確率値の総和をとり、これを 1 から引いて未知語クラスの確率値に割り当てる。

5. 言語モデル融合ツールの開発および評価

提案手法を実装した言語モデル融合ツールを構築した。このツールは 2 つの言語モデルを重みを付けて融合することができる。特徴としては、相補的バックオフアルゴリズムにより高精度な融合が行えることや既存のツールのように学習元コーパスや頻度ファイルを用意する必要がないことがあげられる。また、語彙に関して、元の 2 つのモデルで語彙が共通である必要はない。任意の ARPA 標準フォーマットの言語モデルファイルを簡単に融合することができる。融合後の語彙は元の 2 つのモデルの和となるが、融合後の 1-gram の頻度上位で語彙数を制限することもできる。

このツールを用いて異なるタスクに依存した 2 つの言語モデルを融合させ、相補的バックオフアルゴリズムの評価実験を行った。実験の詳細を以下に示す。

5.1 実験条件

実験に用いた音声認識対象タスクは、医療相談、グルメ・レシピ検索および新聞記事の 3 つである。言語モデルの学習に用いたテキストコーパスの諸元を表 1 に示す。これらのコーパスは、新聞記事タスクを除いてウェブの掲示板などから人手で収集したものである^{8),9)}。また、新聞記事タスクのためのテキストコーパスとしては、新聞記事 1 年分のテキストを用いた。

これらタスクに依存した言語モデルとして、前述のコーパスから学習したバックオフ 3-gram 言語モデル (医療相談モデル、グルメ・レシピモデルおよび新聞記事モデル) を作成した。学習に使用した語彙数は、学習元コーパスの出現頻度が高いものから上位 20,000 語であり、この場合の学習元コーパスに出現する全単語に対する単語力バレージは、医療相談モデ

表 1 学習用テキストコーパスの緒元
Table 1 Specification of training corpora.

タスク名	医療相談	グルメ・レシピ	新聞記事 1 年分
異なり単語数	42,055	79,741	142,338
総単語数	3,606,701	9,719,540	24,684,738
総文数	319,498	805,908	906,106

ル 99.19%、グルメ・レシピモデル 98.42%、新聞記事モデル 96.54% である。具体的なバックオフ N-gram 言語モデルの作成手順は、「日本語ディクテーション基本ソフトウェア (99 年度版)」¹⁰⁾ のものに準ずる。ディスカウントには Witten-Bell 法⁶⁾ を用いた。

次に、以下の融合手法を用いて、「医療相談+グルメ・レシピ」および「新聞記事+グルメ・レシピ」で融合されたバックオフ 3-gram 言語モデルをそれぞれ作成した。

- コーパス結合 (corpus)
- 相補的バックオフなしモデル融合 (no-CB)
- 相補的バックオフありモデル融合 (CB)

“corpus” は、従来手法である学習元コーパスを単純につないで作成した学習元コーパスから学習をやり直して構築したモデルである。“no-CB” は、相補的バックオフを行わずに、今回、実装した融合ツールにより 2 つの言語モデルどうしをモデル融合した言語モデルである。この場合、融合前に未観測であった N-gram の頻度は相補的に推定せず 0 として融合を行っており、コーパス結合モデル (corpus) よりも性能が落ちることが予測される。“CB” では、本論文で提案する手法である相補的バックオフを行いながら no-CB と同様にツールを用いたモデルの直接融合を行った。

なお、以下の実験内で使用する融合重み (weight) は、「医療相談+グルメ・レシピ」の融合および「新聞記事+グルメ・レシピ」の融合のすべての場合において、グルメ・レシピモデルの重み係数 λ を表している。つまり、融合重みが大きいほどグルメ・レシピタスクに適合したモデルが生成されることになる。

上記の融合モデルの融合後の語彙数は、同じ語彙数での性能を比較するため、2 つのモデルの融合後の単語頻度の上位から 20,000 語に制限した。さらに、この語彙数の制限を外した場合についても同様に評価を行った。“unlimited-no-CB” および “unlimited-CB” は、上記の no-CB および CB のモデル融合後の語彙数の制限を行わないモデルである。なお、ツールを用いて直接モデル融合した場合の融合後の異なり語彙数は、「医療相談+グルメ・レシピ」の融合で 28,483 語、「新聞記事+グルメ・レシピ」の融合で 29,951 語であった。そこで、これらと同じ語彙数のコーパス

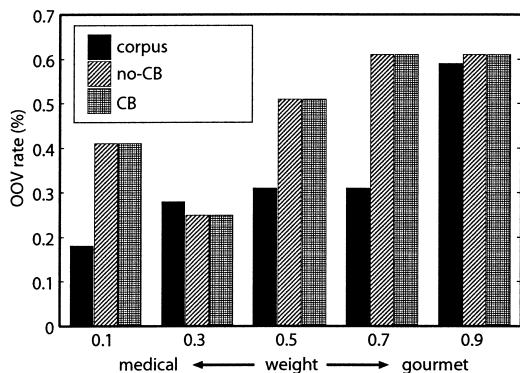


図3 未知語率(医療相談+グルメ・レシピ, 語彙数 20k)
Fig. 3 OOV rate (medical + gourmet, 20k words).

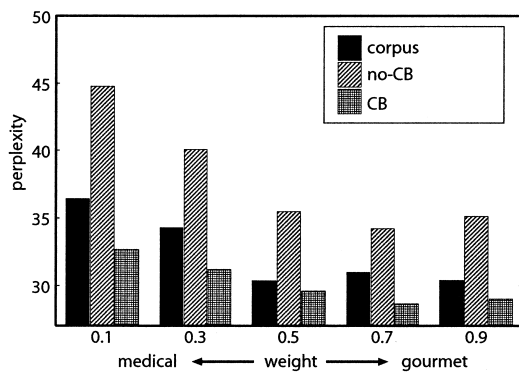


図5 パープレキシティ(医療相談+グルメ・レシピ, 語彙数 20k)
Fig. 5 Perplexity (medical + gourmet, 20k words).

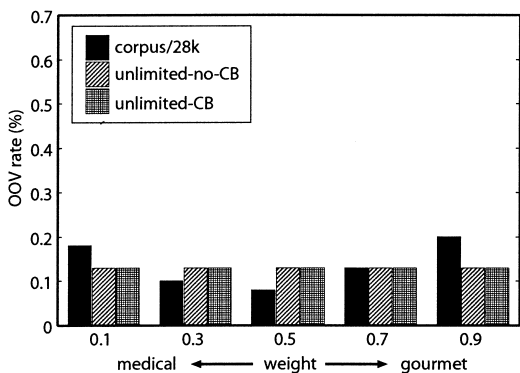


図4 未知語率(医療相談+グルメ・レシピ, 語彙数 28k)
Fig. 4 OOV rate (medical + gourmet, 28k words).

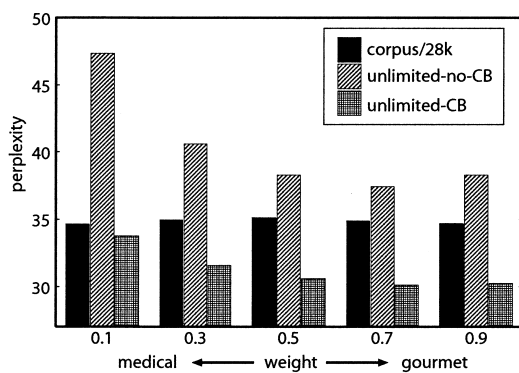


図6 パープレキシティ(医療相談+グルメ・レシピ, 語彙数 28k)
Fig. 6 Perplexity (medical + gourmet, 28k words).

結合モデル (corpus/28k, corpus/29k) もあわせて作成した。

- コーパス結合, 語彙数 28483 語 (corpus/28k, corpus/29k)
- 相補的バックオフなしモデル融合 + 語彙数無制限 (unlimited-no-CB)
- 相補的バックオフありモデル融合 + 語彙数無制限 (unlimited-CB)

評価文として, 医療相談 150 文, グルメ・レシピ 200 文, 新聞記事 300 文を用意した。医療相談とグルメ・レシピのテストセットは, それぞれのタスクの対話システムでの利用を考慮して人手で作成した丁寧な話し言葉による文章である^{8),9)}。そこで 2 つのテストセットをつなげた 350 文の医療相談+グルメ・レシピ結合テストセットも評価に用いた。また, 新聞記事テストセットは, 前述の学習用コーパス新聞記事 1 年分とは別のテキストから作成したものである。

5.2 未知語率およびパープレキシティ

融合重みが 0.7 のときのそれぞれのモデルの未知語率 (OOV), テストセットパープレキシティ (PP) の

結果を表 2 に示す。また, 医療相談+グルメ・レシピの結合テストセットに対する医療相談とグルメ・レシピの融合モデルの融合重みを変えたときの未知語率を図 3 (語彙数制限あり), 図 4 (語彙数制限なし), テストセットパープレキシティを図 5 (語彙数制限あり), 図 6 (語彙数制限なし) に示す。

未知語率は, モデル融合 (no-CB, CB) とコーパス結合を比較して, 語彙数無制限 (図 3) のときに若干の悪化は見られるが, ほぼすべての融合の場合において同等の性能を示した。パープレキシティについては, ほぼすべての融合重みにおいて提案手法である CB モデルが, 最も良い値を示し, 従来のコーパス結合と同等の精度で融合モデルを構築できることが示された。相補的バックオフを施さずに言語モデルを融合した場合は, パープレキシティが劣化する傾向にあり, 相補的バックオフの有効性が確認できた。

5.3 大語彙連続音声認識実験

次に, 大語彙連続音声認識エンジン Julius¹¹⁾ を用いて音声認識実験による評価を行った。評価用音声データとして 60 歳から 80 歳までの高齢者男性 51 名, 女

表 2 未知語率およびパープレキシティ
Table 2 OOV rate and perplexity.

	医療相談		グルメ・レシピ		新聞記事	
	OOV (%)	PP	OOV (%)	PP	OOV (%)	PP
融合元モデル						
グルメ・レシピ	5.11	29.24	0.12	29.87	11.07	320.87
医療相談	0.27	26.43	3.53	86.71	9.03	366.70
新聞記事	4.84	69.45	2.22	86.50	2.15	56.51
医療相談+グルメ・レシピ融合 (融合重み 0.7)						
corpus	0.54	27.16	0.16	33.56	9.76	534.57
no-CB	1.34	31.25	0.16	36.14	10.61	334.54
CB	1.34	24.89	0.16	31.21	10.61	377.24
corpus/28k	0.20	28.57	0.08	39.35	6.97	726.53
unlimited-no-CB	0.20	38.58	0.08	33.49	6.98	440.17
unlimited-CB	0.20	27.30	0.08	31.90	6.98	488.67
新聞記事+グルメ・レシピ融合 (融合重み 0.7)						
corpus	4.17	34.21	0.21	32.71	2.78	52.46
no-CB	4.44	42.19	0.08	36.05	3.42	99.12
CB	4.44	32.89	0.08	31.36	3.42	60.13
corpus/29k	2.35	42.68	0.08	33.32	1.46	51.42
unlimited-no-CB	3.02	57.34	0.04	36.24	1.79	106.84
unlimited-CB	3.02	39.27	0.04	31.64	1.79	58.78

表 3 単語認識率 (%)
Table 3 Word %correct (%).

	医療相談	グルメ・レシピ	新聞記事
融合元モデル			
医療相談	91.11	75.95	66.65
グルメ・レシピ	79.38	89.79	65.26
新聞記事	79.07	82.64	90.79
医療相談+グルメ・レシピ融合 (融合重み 0.7)			
corpus	91.04	89.58	65.63
no-CB	88.48	88.77	66.36
CB	90.46	89.04	66.29
corpus/28k	91.18	88.95	66.70
unlimited-no-CB	90.36	89.17	67.78
unlimited-CB	91.43	89.45	67.53
新聞記事+グルメ・レシピ融合 (融合重み 0.7)			
corpus	82.17	89.84	89.26
no-CB	80.25	89.85	83.65
CB	80.24	89.18	86.29
corpus/29k	83.02	89.57	90.58
unlimited-no-CB	81.15	90.08	85.65
unlimited-CB	81.86	89.69	88.94

性 50 名の計 101 話者の読み上げ音声を用いた。各話者が前述のテストセットの中から医療相談 30 文、グルメ・レシピ 40 文、新聞記事 100 文を発話したものであり、合計で約 17,000 文分の音声である。音響モデルには、NEDO「シニア支援システムの開発」プロジェクト^{12),13)}によって作成された高齢者向け音響モデル¹⁴⁾(PTM¹⁵⁾、2000 状態、64 混合、性別依存)を用いた。

すべてのテストセットに対して、融合重みが 0.7 のときの単語認識率を表 3 に示す。また、医療相談+グルメ・レシピの結合テストセットに対する医療相談と

グルメ・レシピの融合モデルの融合重みを変えたときの単語認識率を図 7 (語彙数制限あり)、図 8 (語彙数制限なし)に示す。

医療相談+グルメ・レシピの融合の医療相談とグルメ・レシピの結合テストセットに対する実験結果(図 7)では、相補的バックオフなしの言語モデル融合において、コーパス結合モデルに比べて平均して 1.5%程度の認識率の劣化が見られた。これに対して相補的バックオフを施すことで、認識率は最大 1%程度改善され、コーパス結合に迫る性能を示した。さらに語彙数制限を行わないモデルどうしを比較したところ、提案手法

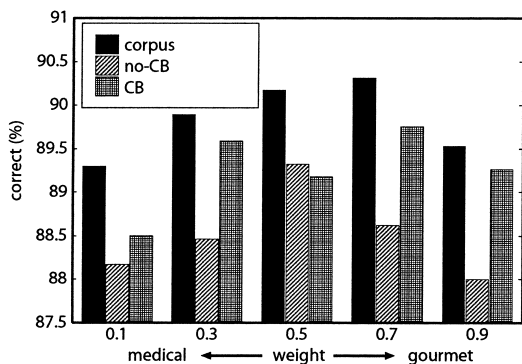


図 7 単語認識率 (医療相談 + グルメ・レシピ, 語彙数 20k)
Fig. 7 Word %correct (medical + gourmet, 20k words).

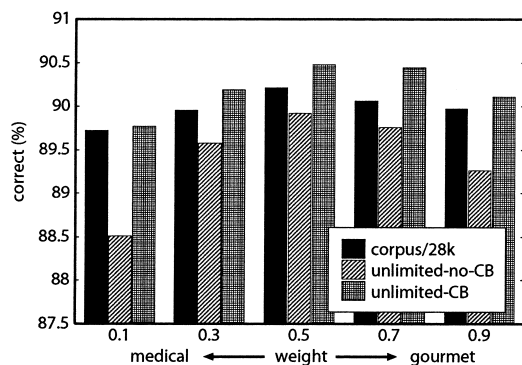


図 8 単語認識率 (医療相談 + グルメ・レシピ, 語彙数 28k)
Fig. 8 Word %correct (medical + gourmet, 28k words).

は、従来手法であるコーパス結合モデルを上回る認識率を示した (図 8)。このことから、提案手法によって、コーパス結合と同等の高精度なモデル融合を学習元コーパスを用いずに行えることが示された。

表 3 の医療相談とグルメ・レシピのそれぞれタスクに対する認識率を見ると、相補的バックオフなしでは融合することでそれぞれのタスクに対して元のモデルから認識精度が劣化した。相補的バックオフを施すことで、融合後も両タスクについて元モデルと同じ精度を保てること示された。新聞記事 + グルメ・レシピの融合モデルの新聞記事タスクに対する認識結果においても同様の傾向が見られた。

しかし、新聞記事 + グルメ・レシピ融合モデルのグルメ・レシピタスクに対する評価では、相補的バックオフによる認識率の劣化が見られる。これは、新聞記事モデルは、グルメ・レシピタスク以外の多くの特性を持ち、その結果、融合されたモデルは、新聞記事モデルのグルメ・レシピタスクに関連する新たな特性を得る以上に、無関係な特性を含んでしまったことが原因であると考えられる。よって、相補的バックオフは、

融合するモデルどうしがある程度似た特性を持つことを前提としたアルゴリズムであり、傾向が異なるモデルの融合により性能の劣化が生じる場合があることが分かる。

6. おわりに

言語モデルの融合において、未観測の N-gram 確率を他方の融合相手の N-gram 確率から推定する相補的バックオフアルゴリズムを提案し、それを実装した N-gram 言語モデル融合ツールを構築した。

言語モデルにおいては、ある話題に関するタスクにおいて現れ、他のタスクのモデルでは現れないような N-gram が、タスクの特徴を表している重要な部分であると考えられる。相補的バックオフを用いることで、融合後もタスクごとの精度をあまり下げることなく、両タスクに対して高い性能を得ることができる。

認識実験の結果、提案手法は相補的バックオフを行わない頻度の重み付き和による融合よりも精度が高く、従来からのコーパスレベルでの融合と同等の精度を持つモデルを作成することが可能であることが示された。本手法を用いた融合ツールは、コーパスや頻度ファイルを用いず完成した言語モデルから高精度な融合モデルを生成することができるので、利便性の面からコーパス結合に比べて優れており、有用であることが分かった。

提案手法である相補的バックオフは、細かく分割されたコーパスから分割に際して欠落した情報を補間する手法であると考えられる。たとえば、新聞記事のように政治、経済、スポーツ、芸能などさまざまなトピックを広くカバーするようなコーパスから学習するモデルのタスクを制御したいとき、コーパスをタスクごとに小さく分割したうえでモデルを作成することが有用であると考えられる。コーパスから言語モデルを構築する際には語彙の制限やカットオフによって何らかの情報が欠落する。コーパスを小さく分割するとその欠落の影響は大きくなる。この欠落した情報を未観測 N-gram ととらえれば、相補的バックオフを用いて分割による情報欠落を抑制することができる。

今後の課題として、3 つ以上の複数の言語モデルについての相補的バックオフアルゴリズムを検討していきたい。融合の優先順位や割合を考慮するために、モデル間の類似度を推し量る尺度の定義が必要になると考えられる。また、品詞などのヒューリスティックを用いて部分的に相補的バックオフを施すアルゴリズムなどについても検討していきたい。

謝辞 本研究は、NEDO(新エネルギー・産業技術総合開発機構)の援助を受けて行われた。認識実験に用いた高齢者音声の収集および整備は、本プロジェクトの一環として TIS 株式会社(株式会社東洋情報システム)によって行われた。ご協力いただいた関係各位に感謝します。

参 考 文 献

- 1) 竹澤寿幸: いまこそ話しことば処理技術の研究を, 情報処理, Vol.42, No.2, pp.173-177 (2001).
- 2) Chen, S.F., Seymore, K. and Rosenfeld, R.: Topic Adaptation for Language Modeling using Unnormalized Exponential Models, *Proc. ICASSP98*, Vol.2, pp.681-684 (1998).
- 3) Weintraub, M., Aksu, Y., Dharanipragada, S., Khudanpur, S., Ney, H., Prange, J., Stolcke, A., Jelinek, F. and Shriberg, L.: Fast Training and Portability, 1995 Language Modeling Summer Research Workshop, Technical Reports, Center for Language and Speech Processing, John Hopkins University, Baltimore (1995).
- 4) Clarkson, P. and Rosenfeld, R.: Statistical Language Modeling Using the CMU-Cambridge Toolkit, *Proc. EUROSPEECH '97*, Vol.5, pp.2707-2710 (1997).
- 5) 伊藤彰則, 好田正紀: 単語およびクラス N-gram 作成のためのツールキット, 電子情報通信学会技術研究報告, SP2000-106, pp.67-72 (2000).
- 6) Witten, I.H. and Bell, T.C.: The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression, *IEEE Trans. Information Theory*, Vol.37, No.4, pp.1085-1094 (1991).
- 7) Katz, S.M.: Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer, *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol.35, No.3, pp.400-401 (1987).
- 8) Nisimura, R., Komatsu, K., Kuroda, Y., Nagatomo, K., Lee, A., Saruwatari, H. and Shikano, K.: Automatic N-gram Language Model Creation from Web Resources, *Proc. EUROSPEECH2001*, Vol.3, pp.2127-2130 (2001).
- 9) 小松久美子, 黒田由香, 長友健太郎, 西村竜一, 李 晃伸, 鹿野清宏: 高齢者タスクにおける話し言葉言語モデルの構築, 情報処理学会第 62 回全国大会講演論文集, 1L-1 (2001).
- 10) 河原達也, 李 晃伸, 小林哲則, 武田一哉, 峯松信明, 嵯峨山茂樹, 伊藤克亘, 山本幹雄, 山田篤, 宇津呂武仁, 鹿野清宏: 日本語ディクテーション基本ソフトウェア(99年度版), Vol.57, No.3, pp.210-214 (2001).
- 11) 李 晃伸, 河原達也, 堂下修司: 単語トレリスインデックスを用いた段階的探索による大語彙連続音声認識, 電子情報通信学会論文誌, Vol.J82-D-II, No.1, pp.1-9 (1999).
- 12) 鹿野清宏: 2000 年代に何をすべきか—研究課題と取り組み—NEDO シニア支援システムと音声研究の課題, 情報処理学会研究報告, 2000-SLP-32-5, pp.79-80 (2000).
- 13) 鹿野清宏, 馬場 朗, 芳澤伸一, 山田実一, 西村竜一, 小松久美子, 黒田由香, 李 晃伸: 高齢者音声の認識, 電子情報通信学会技術研究報告, WIT2001-20, pp.25-30 (2001).
- 14) 馬場 朗, 芳澤伸一, 山田実一, 李 晃伸, 鹿野清宏: 高齢者向け音響モデルによる大語彙連続音声認識の評価, 情報処理学会研究報告, 2001-SLP-35-3, pp.13-18 (2001).
- 15) 李 晃伸, 河原達也, 武田一哉, 鹿野清宏: Phonetic Tied-Mixture モデルを用いた大語彙連続音声認識, 電子情報通信学会論文誌, Vol.J83-D-II, No.12, pp.2517-2525 (2000).
(平成 13 年 6 月 8 日受付)
(平成 14 年 7 月 2 日採録)



長友健太郎(正会員)

1997 年大阪府立工業高等専門学校卒業。1999 年神戸大学発達科学部人間環境学科卒業。2001 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。現在、日本電気株式会社マルチメディア研究所で音声認識技術の研究・開発に従事。



西村 竜一(学生会員)

1999 年名古屋大学工学部電気電子情報工学科卒業。2001 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了。現在、同博士後期課程在学中。主に音声認識を基礎とする将来のコミュニケーション技術に関心を持っている。電子情報通信学会, 日本音響学会各学生会員。

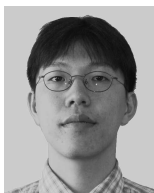


小松久美子

財団法人イメージ情報科学研究所勤務。画像処理や音声認識, メディアコンテンツの研究開発に従事。



黒田 由香
TIS 株式会社勤務。音声認識の研究開発に従事。



李 晃伸 (正会員)
平成 8 年京都大学工学部情報工学科卒業。平成 10 年同大学大学院修士課程修了。平成 12 年同大学院情報学研究科博士課程修了。同年より奈良先端科学技術大学院大学情報科学研究科助手。主として音声認識・理解の研究に従事。博士(情報学)。平成 14 年日本音響学会粟屋潔学術奨励賞受賞。日本音響学会、電子情報通信学会各会員。



猿渡 洋
平成 3 年名古屋大学工学部電気工学科卒業。平成 5 年同大学大学院修士課程修了。平成 12 年同大学院博士課程修了。工学博士。平成 5 年セコム(株)入社。セコム IS 研究所音声情報処理研究室において、超音波アレー信号処理に関する研究に従事。平成 12 年奈良先端科学技術大学院大学助教授。音声信号処理、音響アレー信号処理、ブラインド処理、音場再生等に関する研究に従事。平成 13 年電子情報通信学会論文賞受賞。電子情報通信学会、日本音響学会、IEEE 各会員。



鹿野 清宏 (正会員)
昭和 45 年名古屋大学工学部電気工学科卒業。昭和 47 年同大学大学院修士課程修了。同年電電公社武蔵野電気通信研究所入所。昭和 59~61 年カーネギーメロン大学客員研究員。昭和 61 年~平成 2 年 ATR 自動翻訳電話研究所音声情報処理研究室長。平成 4 年 NTT ヒューマンインタフェース研究所主席研究員。平成 6 年より奈良先端科学技術大学院大学情報科学研究科教授。音情報処理学講座を担当。工学博士。主として音声・音情報処理の研究および研究指導に従事。昭和 50 年電子通信学会米沢賞。平成 3 年 IEEE SP 1990 Senior Award。平成 6 年日本音響学会技術開発賞。平成 12 年情報処理学会山下記念研究賞。平成 13 年 VR 学会論文賞。IEEE, ISCA, 音響学会, 電子情報通信学会, VR 学会各会員。