音声言語情報処理 35－9
（2001．2．3）

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＊

＊ 奈良先端科学技術大学院大学 情報科学研究科
＊＊ イメージ情報科学研究所　＊＊＊ TIS 株式会社

あらまし　N-gram 言語モデルは学習元のコーパスの話題や知識に依存するが、異なるタスクについて構築された複数の N-gram 言語モデルの融合を行なうことで、広いタスクをカバーする言語モデルを構築できる。この際、ある言語モデルでは観測されるが他のモデルでは観測されないような未観測 N-gram が存在する。これらの頻度をその N-gram を含む他方のモデルから推定することで、各モデルのタスク依存性をなるべく保存しつつ、どのタスクについても高い精度を示すモデルの融合が可能である。本報告では、この相補的バックオフアルゴリズムを実装した言語モデル融合ツールを構築し、これを評価した。医療相談及びグルメ&レシピ検索の二つのタスクにおいて、このツールを用いて両タスクごとに構築した言語モデルを融合させた結果、いずれのタスクに対しても精度の劣化しないモデルを得ることができた。

キーワード N-gram 言語モデル，言語モデルの融合, 相補的バックオフ

Complemental Backoff Algorithm for Merging Language Models

Kentaro NAGATOMO*　Ryuichi NISIMURA*
Kumiko KOMATSU　Yuka KURODA*****
Akinobu LEE*　Hiroshi SARUWATARI*　Kiyohiro SHIKANO*

*** Graduate School of Information Science, Nara Institute of Science and Technology**
**** Laboratories of Image Information Science and Technology　*** TIS Corp.**

Abstract　An N-gram Language Model depends on topics and knowledges of learning corpus. By merging language models depending on different tasks, we can get a task-dependent model that covers wider topic. At this case, we will face the incoherency of N-gram : some N-gram entries observed in a model may not appear in others. By estimating the frequency of such unobserved N-gram from other models which contain the N-gram, task dependency of each model can be preserved and we can make a model which covers both tasks concurrently. We call this estimation alogrtithm as "complemental back-off". In this report, we develop and evaluate a tool implementing this algorithm. Results show that our merged model shows less degradation for two individual tasks : medical consulting and groumet & recipe.

Key words　N-gram language model, language model merging, complemental backoff

1　はじめに

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 N-gram 言語モデルが今日広く用いられている。このモデルは単純で扱いやすいという性質を持つが、その性能は基となるコーパスの質と量に依存し、大量の整ったテキストデータが必要とされる。

我々が収集できるコーパスは何らかのタスクに依存しており、そこから構築される言語モデルもまたそのタスク依存性を引き継ぐ。この性質を利用して、特定タスクに限定することで高い認識率を得るシステムを構築することができる。さらに、言語モデルのタスク依存性を自由に制御できれば、音声認識を用いたアプリケーションにとって非常

に有用である。このようなタスク依存性の制御を目的として、異なるタスクに依存する複数の言語モデルを融合する方法が研究がされている。

もっとも単純な融合手法はコーパスを直接結合することであるが、融合のために常に元のコーパスを保持する必要があり現実的とは言えない。コーパスではなく頻度情報のみを保持する方法もある。多くのツール [1][2] ではこちらの方法がとられているが、この場合に問題となるのは言語モデル間の N-gram エントリの不整合である。

あるモデルで観測された N-gram が他のモデルでも観測されるという保証はない。言語モデルの融合は基本的にすべての N-gram の重み付き和をとるから、このような不整合が生じないようにするか、あるいは何らかの方法で不整合を起こす N-gram に出現頻度を与える必要がある。

広く使われている SLM Toolkit[1] の場合、言語モデルの融合は mergeidngram というコマンドで行なうことができるが、このコマンドが処理できるのは語彙が同じモデルに限られる。すなわち、N-gram の不整合が起こらないようあらかじめ調整しておいたモデルでなければ融合できない。実用上の利便性を考えるなら、どのような N-gram 言語モデルでも、モデルそのものだけを用いて融合できることが望ましい。

本研究では、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。互いに未観測な N-gram を相補的に推定することでより良い融合を行なう。また融合に際してコーパスや頻度情報ファイルは必要なく、言語モデルファイルのみからの融合が可能である。

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した。このツールを用いることで、非常に簡便に精度の高い言語モデルの融合を行なうことが可能となる。

2 言語モデルの融合

本研究でいう言語モデルの融合とは、各モデルに含まれる各々の N-gram の出現頻度の重み付き和をとることを言う。以降では融合元の言語モデルを二つに限定する。基になるコーパスを F、G と表し、それぞれから構築された二つの言語モデルを L_f、L_g とすると、ある N 単語の組 w_{i-N+1}^{i} のコーパス F における出現頻度 $C_f(w_{i-N+1}^{i})$、コーパス G における出現頻度 $C_g(w_{i-N+1}^{i})$ から、融合後のモデルにおける w_{i-N+1}^{i} の出現頻度 $C(w_{i-N+1}^{i})$ 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 λ_f および λ_g は任意の重み係数である。

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。

適用時にはこのようにして求めた確率値のみを用いるため、通常は言語モデルには頻度情報は含まれず、出現確率のみ保持されている。このような場合でも、コーパス全体の総語彙数 $C(*)$ が分かれば

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

として計算できる。

ここで、融合後の確率計算に必要なのは N-gram の厳密な出現頻度ではなく、二つのモデルでの頻度の比であるから、結局 $C(*)$ は融合重み λ と等価になる。以降では融合重み λ とコーパスの総語彙数 $C(*)$ を区別しない。

3 未観測 N-gram 確率の推定

モデル L_f では観測されるが他方のモデル L_g では未観測であるような N-gram を L_g の未観測 N-gram と名付ける。この未観測 N-gram の出現頻度または出現確率を推定する。

頻度情報がない以上、未観測 N-gram について正しい出現確率を推定することは非常に難しい。しかし、コーパスで観測されない N-gram の出現確率の合計は、経験的または統計的手法[1]によって概算することができる。一般に流通する言語モデルには、このようにしてあらかじめ未観測 N-gram

[1] いわゆる discounting method を指す。

のために間引かれた値が含まれる。この値を未観測確率集合と表記する。

未観測 N-gram の出現確率の合計は、この未観測確率集合よりも小さくなければならない。また、どんなに多くの未観測 N-gram を推定したとしても、やはり未観測確率集合は 0 になってはならない。

あるコンテキスト w_{i-N+1}^{i-1} の未観測確率集合を $P(*|w_{i-N+1}^{i-1})$ と表す。コンテキストに属すすべての N-gram の出現確率の和は 1 であるから、この値は

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

として求めることができる。

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集合に占める割合を $\beta(w_{i-N+1}^{i-1})$ で表す。β の推定は未観測確率集合の再計算を意味する。

さらにある N-gram w_{i-N+1}^{i-1} がそのうちの $\gamma(w^i|w_{i-N+1}^{i-1})$ を占めるならば、この未観測 N-gram の出現確率は

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

で表すことができる。

3.1 未観測 N-gram を無視する場合

最も簡単な推定は、未観測 N-gram の出現確率に 0 を割り当てることである。この場合パラメータ β と γ のいずれかまたは両方が 0 である。

未観測 N-gram の推定を考慮しなかった従来の融合法はこのケースに分類される。

3.2 未観測 N-gram を推定する場合

β の推定は未観測確率集合の推定と同じことなので、よく知られる witten-bell 法や good-turing 推定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尤もらしい値となる。

ところで、未観測 N-gram は、それが実際に観測される他方のモデルのタスク依存性を特徴づけている要素である。融合モデルではそのタスク依存性をそのまま保存することが望ましいので、むしろ他方のモデルにとって尤もらしい値を割り当てる方がよい。

そこで、未観測 N-gram の確率を、融合相手のモデルの N-gram 分布を参考にして推定する。二つのモデルが互いの未観測 N-gram を相互に補完し合うように推定するため、この推定法を相補的バックオフ[2]と呼ぶ。

これに対して従来のバックオフを自己バックオフと呼んで区別する。

相補的バックオフでは、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。言語モデルのタスク依存性がモデルに特有の（つまり他のモデルでは未観測の）N-gram 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、タスク依存性を保存したモデルの融合が期待できる。

4 相補的バックオフの一例

相補的バックオフはスキームであり、具体的にはさまざまな実現法が考えられる。ここでは後に述べる言語モデル融合ツールに実装した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まったく同様である。

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 を、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推定する。

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

4.2 パラメータ γ の推定

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

[2]バックオフという言葉は、狭義では N-gram を (N-1)-gram から推定することを意味するが、広義では単に N-gram を他の情報から推定することを指す。ここでの使い方は厳密には正しくないが、直感的に分かりやすいのでこの呼称を用いる。

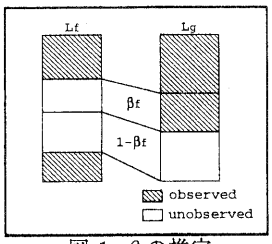

図 1: β の推定

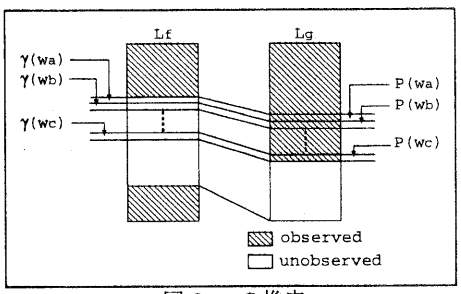

図 2: γ の推定

の総和に占める w_{i-N+1}^{i} の割合に等しいと仮定する。すなわち、

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

5 言語モデル融合ツールの評価

相補的バックオフアルゴリズム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 ARPA 標準フォーマットの言語モデルを任意の重みで融合することができる。元コーパスや頻度ファイルは必要ない。語彙数は任意に制限することができ、元となった二つのモデルの異なり語彙数がこの制限より多い場合は、融合後の unigram の頻度上位が自動的に選ばれる。

このツールを用いて相補的バックオフアルゴリズムの評価実験を行った。

タスクは医療相談およびグルメレシピ検索の二つを用意した。実験で用いたコーパスの諸元を表 1 に示す。これらのコーパスは Web の掲示板等から人手で収集したものである。[3]

各タスクごとに構築したタスク依存モデル（医療相談モデルおよびグルメ&レシピモデル）、および以下の四種類の融合モデルを両タスクで比較、評価する。

表 1: 実験に用いた学習コーパス

	医療相談	グルメレシピ
タスク	医療相談	レストラン案内、料理レシピ
語彙数	43217	81402
文章数	476411	1011912

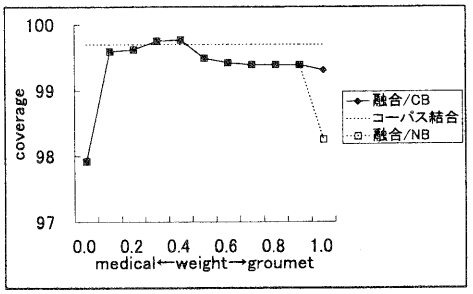

図 3: カバレージ（グルメレシピ＋医療相談）

- 融合/CB モデル：相補的バックオフを施した融合モデル。重みが大きいほど医療相談モデルに近づく。
- 融合/NB モデル：未観測 N-gram の確率を推定せず 0 とした融合モデル。重みが大きいほど医療相談モデルに近づく。
- 融合/UL モデル：融合/CB モデルで語彙数の制限を掛けないモデル。
- コーパス結合モデル：元コーパスを単純につないで構築したモデル。

いずれのモデルも単語 2-gram/3-gram である。語彙数は融合/UL モデル以外は 20000 語、融合/UL モデルのみ 28483 語である。

評価文として医療相談 150 文、グルメレシピ 200 文、およびそれら二つを結合した 350 文の三つを用いた。この評価文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文章である。

5.1 カバレージ、パープレキシティ

グルメレシピ＋医療相談に対する各モデルのカバレージおよびテストセットパープレキシティをグラフ 3、4 に示す。またすべての評価文に対する結果を表 2 に示す。

融合重みが異なる場合、同じ語彙であっても各 unigram の確率値が異なる。語彙は unigram の確率値によって制限されるため、重みが変わるとカバレージも変化する。

表 2: カバレージおよびパープレキシティ

	医療相談		グルメレシピ		医療相談+グルメレシピ	
	cov	pp	cov	pp	cov	pp
医療相談	99.73	26.38	96.81	89.37	99.92	55.81
グルメレシピ	95.23	29.92	99.88	29.40	95.23	29.57
コーパス結合	99.46	25.66	99.94	30.20	99.70	28.38
融合/CB(weight=0.7)	98.66	25.06	99.84	30.97	99.39	28.58
融合/NB(weight=0.7)	98.66	25.50	99.84	31.26	99.39	28.94
融合/UL(weight=0.7)	99.80	26.99	99.92	31.53	99.87	29.71

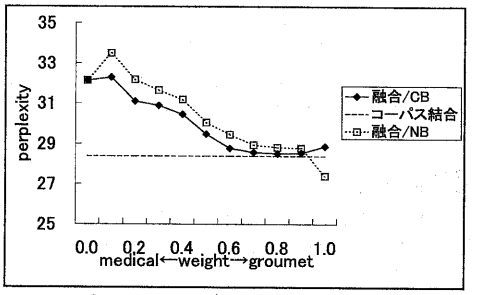

図 4: パープレキシティ（グルメレシピ＋医療相談）

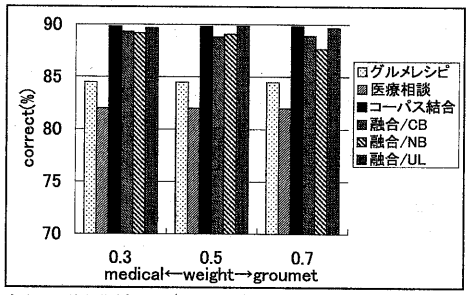

図 5: 認識結果（グルメレシピ＋医療相談）

融合モデルのカバレージはコーパス結合モデルのそれと同様か多少落ちる程度の値を示す。特に語彙数に制限を掛けないモデルでは、コーパス結合に勝る値を得られた。

また、カバレージについては相補的バックオフを施したモデルとそうでないモデルでほとんど違いがない。unigram についてはどちらもほぼ同じエントリを持つので当然と言える。重みが 0.0、1.0 の際に融合/NB モデルで急激に低下するのは、一方モデルの語彙が完全に無視されてしまうためで、このように極端な重みの場合には融合/CB モデルが優れている。

パープレキシティについては、いずれの評価文においても相補的バックオフを施した方が施さない場合よりも 0.05～0.5%程度良い値を示している。モデルの重みが評価文と異なるタスクに掛かる場合にその影響は大きくなる。また、ベストな重みを与えた場合には、コーパス結合モデルと同等のパープレキシティが得られた。

5.2 認識実験

大語彙連続音声認識エンジン Julius[4] を用いて認識実験を行なった。音響モデルは高齢者向け音響モデル [6]（PTM[5]、2000 状態、64 混合、性別依存）を用いた。

評価用音声データとして 60 歳～80 歳までの高齢者男性 51 名、女性 50 名の計 101 話者を用いた。各話者が 70 文（医療相談 30 文、グルメレシピ 40 文）を発話し、合計で約 7000 文を用いて実験を行なった。

グルメレシピ＋医療相談に対する認識結果を図 5 に、すべての評価セットに対する認識率を表 3 に示す。

認識率においても、基本的にパープレキシティと同様の傾向が見られた。すなわち、相補的バックオフを施したモデルはそうでないモデルよりも認識率が 1%程度向上する。モデルの重みが評価文と異なるタスクに掛かる場合にその影響は大きくなる。

コーパス結合と比較した場合、融合/CB モデルの認識率は 0.5～1.5%ほどの低下で収まっている。語彙数を制限しない融合/UL モデルではその差は 0.4%以内に縮まる。

6 おわりに

言語モデルの融合の際に生じる N-gram の不整合を、相補的バックオフアルゴリズムによって解消するツールを構築し、その性能を評価した。結

表 3: 認識結果

	医療相談	グルメレシピ	医療相談+グルメレシピ
医療相談	89.08	72.86	82.00
グルメレシピ	78.67	88.73	84.48
コーパス結合	89.83	88.67	89.81
融合/CB(weight=0.7)	88.60	87.77	88.91
融合/NB(weight=0.7)	88.61	86.78	87.67
融合/UL(weight=0.7)	89.79	88.28	89.65

果として以下のような特徴があることが分かった。

1. 未観測 N-gram の確率を推定すると、単純に 0 とするよりも高精度な融合ができる。
2. 特に重みを多く掛けたモデル以外の入力に対してその影響は顕著となる。
3. コーパスを直接結合して構築したモデルと同等の認識精度を示す。

相補的バックオフアルゴリズムは、入力タスクが予期できない場面で効果を発揮すると期待できる。異なるタスクの言語モデルを用意しておいて、入力に応じて重みを切替えながら融合することで、複数のタスクの扱うような音声認識システムでの応用が考えられる。また、言語モデルそのものから、コーパスや頻度ファイルからの融合と同等の精度のモデルを得ることができ、利便性に優れている。

相補的バックオフは、タスクごとに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とらえることもできる。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 cutoff 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 と捉えれば、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と考えられる。

ただし、今回の実験では数字の上で大きな差を示すことができなかった。これは実験に用いた評価文がモデルの複雑さに比べてあまりに簡単なものであったことによる。今後の課題として、より適切な評価文を用いた比較実験が挙げられる。

相補的バックオフの具体的な実装については他にもさまざまな手法が考えられる。これらについても検討を進めたい。また、三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も検討していきたい。

また、未観測 N-gram の確率を相補的バックオフと自己バックオフで推定する場合の比較も必要である。

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）によって行なわれた。ご協力いただいた関係各位に感謝します。

参考文献

[1] P.R. Clarkson, R. Rosenfeld: "The CMU-Cambridge Statistical Language Modeling Toolkit v2,"
`http://svr-www.eng.cam.ac.uk/~prc14/toolkit.html`

[2] 伊藤, 好田: "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

[3] 西村, 長友, 小松, 黒田, 李, 猿渡, 鹿野: "Web からの音声認識用言語モデル自動生成ツールの開発," 情報処理学会研究報告,2001-SLP-35-8,2001

[4] 李, 河原, 堂下: "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,J82-D-II No.1,pp.1-9,1999

[5] 李, 河原, 武田, 鹿野: "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,J83-D-II No.12,pp.2517-2525,2000

[6] 馬場, 芳澤, 山田, 李, 鹿野: "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-35-3,2001