

1L-02 相補的バックオフを用いた言語モデル融合アルゴリズム

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

1 はじめに

音声認識のための N-gram 言語モデルは、学習元のテキストの話題や知識、語彙を反映するが、一方で異なるタスクの発話には脆弱である。複数のタスク依存言語モデルを融合することで、タスクごとの依存性を制御することが可能である [1]。

融合の際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。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。

本研究では、融合時に不整合を起こす 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精度の高い融合を試みる。

2 言語モデルの融合

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 の出現確率 $P(w^i|w_{i-N+1}^{i-1})$ は次式のように表される。

$$P(w^i|w_{i-N+1}^{i-1}) = \frac{\lambda_f C_f(w_{i-N+1}^i) + \lambda_g C_g(w_{i-N+1}^i)}{\lambda_f C_f(w_{i-N+1}^{i-1}) + \lambda_g C_g(w_{i-N+1}^{i-1})}$$

3 未観測 N-gram 確率の相補的推定

一方のモデル 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。コーパスで観測されないあらゆる N-gram の出現確率の合計を $P(*|w_{i-N+1}^{i-1})$ 、これに占める推定するすべての未観測 N-gram の確率の総和の割合を $\beta(w_{i-N+1}^{i-1})$ 、さらにある未観測 N-gram w_{i-N+1}^{i-1} がそのうちの $\gamma(w^i|w_{i-N+1}^{i-1})$ を占めるとすると、この未観測 N-gram の出現確率は

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

で表すことができる。

未観測 N-gram はタスク依存性を特徴づけている要素である。そこで、未観測 N-gram の確率を融合相手のモデルの当該 N-gram 分布を参考にして推定することで、それぞれのモデルの各タスクへの依存性を

保持しながら両者をカバーするモデルが作成できる。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と呼ぶ。ここでは β と γ を以下のように推定する。

パラメータ β の推定 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。

パラメータ γ の推定 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 の総和に占める w_{i-N+1}^i の割合に等しいと仮定し、その割合をそのまま用いる。

4 実験的評価

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 [2] を作成し、実験による評価を行った。コーパスは医療相談およびグルメレシピ検索の二つのタスクのものをを用いた。

実験に用いたモデルは、各々のタスクに依存したモデル（医療相談、グルメレシピ）、相補的バックオフによりこの二つを融合したモデル（融合/CB：語彙制限あり、融合/CBUL：語彙制限なし）、未観測 N-gram 確率を 0 として融合したモデル（融合/NB：語彙制限あり、融合/NBUL：語彙制限なし）、そして二つのコーパスを単純につないで構築したモデル（コーパス結合）の計 7 つである。語彙数 20000 語（語彙制限のないモデルでは 28483 語）の単語トライグラムで、融合モデルについては、重みが大いほどグルメレシピモデルに近付く。

評価文として医療相談 150 文とグルメレシピ 200 文の計 350 文を用いた。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会話調の文章である。

カバレッジ、パープレキシティ テストセットに対するカバレッジおよびパープレキシティを図 1 および 2 に示す。

語彙数を制限した融合モデルでは融合後の unigram 頻度上位を選んでいるため、重みによってカバレッジが変化する。これをコーパス結合モデルと比較したところ、同等から 0.3% 程度低い値を示した。語彙数を制限しないモデルでは、コーパス結合に勝る値を得られた。いずれの場合も、融合/CB モデルと融合/NB モデルでカバレッジの違いは全くない。

Merging Language Models with Complemental Backoff Algorithm.

Kentaro NAGATOMO*, Ryuichi NISIMURA*,

Kumiko KOMATSU**, Yuka KURODA***,

Akinobu LEE*, Hiroshi SARUWATARI*, Kiyohiro SHIKANO*

* Graduate School of Information Science, NAIST

** Laboratories of Image Information Science and Technology

*** TIS Corp.

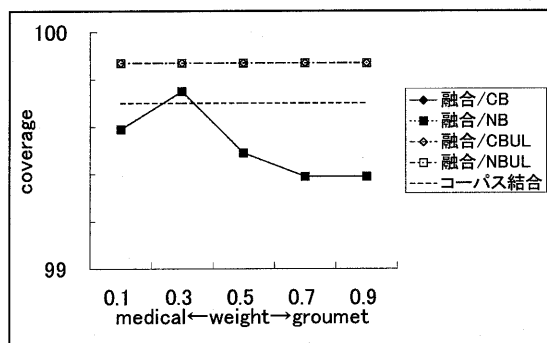


図 1: カバレッジ

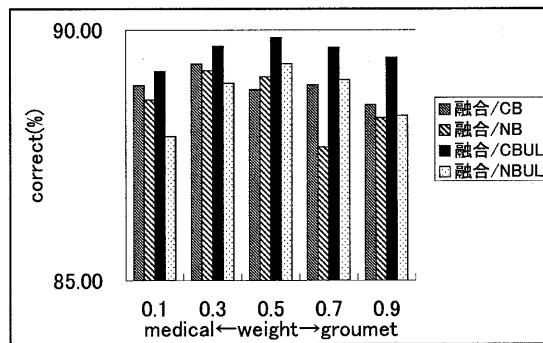


図 3: 認識結果 (融合モデル)

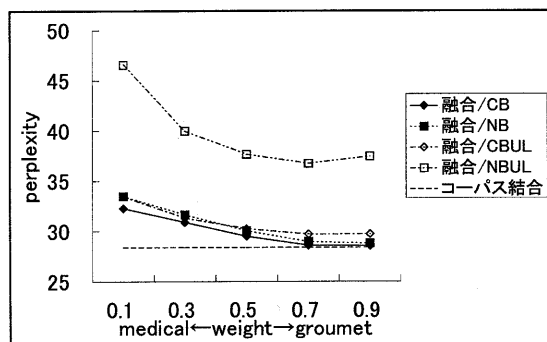


図 2: パープレキシティ

表 1: 各モデルの性能

model	cov(%)	pp	recog(%)
医療相談	97.92	55.81	84.48
グルメレシピ	98.12	29.57	82.00
コーパス結合	99.70	28.38	89.81
融合/CB (weight=0.5)	99.49	29.48	88.81
融合/CBUL(weight=0.5)	99.87	30.25	89.85

パープレキシティについては、融合/CB モデルが融合/NB モデルよりも 0.05~0.5 良い値となった。モデルの重みが評価文と異なるタスクに大きく掛かる場合にその差は大きくなり、語彙数制限を掛けない場合さらに 5~10 程度に広がる。また、ベストな重みを与えた融合/CB モデルでは、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。

音声認識実験 大語彙連続音声認識エンジン Julius[3] 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 [4] (PTM、2000 状態、64 混合、性別依存) を用いた。評価用音声データは 60 歳~80 歳までの話者男女計 101 名による前述の評価文の読み上げ約 7000 文を用いた。

認識結果を表 1 および図 3 に示す。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の向上が認められた。語彙数を制限した場合で 1%程度、制限しない場合では 0.5~0.7%程度向上する。

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%ほどの低下に留まった。語彙数を制限しない融合/CBUL モデルではその差は 0.4%以内に縮まる。

5 おわりに

言語モデルの融合の際に生じる N-gram の不整合を解消する相補的バックオフアルゴリズムを提案し、評価した。その結果、このアルゴリズムは未観測 N-gram の確率を推定ことで高精度な融合ができ、コーパスを直接結合して構築したモデルと同等の性能を持つことが示された。この性質は特に重みを多く掛けたモデル以外の入力に対して顕著となる。

ただし、今回の実験では数字の上で大きな差は見られなかった。これは実験に用いた評価分が簡単であったためにモデルごとの依存性の差が出にくかったものと考えられる。

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。高齢者音声の収集・整備は本プロジェクトの一環として TIS 株式会社によって行われた。ご協力いただいた関係各位に感謝します。

参考文献

- [1] 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
- [2] 長友, 西村, 小松, 黒田, 李, 猿渡, 鹿野: “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,” 情報処理学会研究報告, 2001-SLP-35-9, 2001
- [3] 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
- [4] 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