2008－SLP－72 (9)
2008／7／18

識別的言語モデルの可能性

大庭 隆伸

日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
〒 619-0237 京都府相楽郡精華町光台 2-4

oba@cslab.kecl.ntt.co.jp

識別学習手法に基づく言語のモデリングは，音声認識を含む音声言語処理の進展に大きく貢献した技術と言える．本稿では，対立単語列との識別によりもたらされる効果を述べ，音声認識における識別的言語モデルの研究について紹介する．そして，識別的言語モデルに音響的特徴を取り込む枠組みについて議論する．

Discriminative Language Models — Introduction and Future Prospect —

Takanobu Oba

NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation
2-4 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0237, Japan

Language modeling has a major impact for spoken language processing including speech recognition. Specifically, discriminative approaches are currently attracting a lot of attention. We discuss the development of language models based on discriminative approach in spoken language processing through reviewing recent works with discriminative language modeling.

1 はじめに

音声言語処理の技術躍進に期待が集まる中，その中心技術のひとつである言語モデルの重要性は益々高まっている．

n-gram 言語モデルは，単語や品詞といった極めて基本的な言語的情報を対象としたモデリング手法である．マルコフ性を仮定することで，ある位置での単語の出現確率を，直前の数単語のみを利用して推定する．その簡略さと精度の高さから極めて広範に利用されている．

n-gram 言語モデルに続く言語モデルに関する多くの研究では，言語構造や上位概念，もしくは話題の特定に利用できる情報を導入して，モデル性能を向上させるアプローチを取ることが主流である．代表的な研究として，修辞関係を利用した依存言語モデル [1, 2]，話題などを潜在的パラメータとして導入した PLSA[3, 4, 5] や LDA[6] などがある．また，対話用途などでタスクを限定している環境では意味構造を利用した言語モデルや，講演内容の書き起しではスライド上の情報を利用した手法も提案されている [7, 8]．

一方，近年，識別的アプローチによるモデリングが注目され，その有効性が確認されている．その要因のひとつは，対立単語列を用いた学習にある．適切に対立単語列を設定することで，従来の言語モデルでは扱うことの難しかった情報を取り込むことができる．その対立単語列に内在する情報は，豊富な素性の導入により，多角的な観点から抽出される．

本稿では，まず識別的言語モデルについて概観した後，対立単語列との識別による効果について，音声認識における誤り訂正での例を交えて説明する．そして，3 章にて，識別的言語モデルの可能性に関する議論として，音響的情報を積極的に利用した識別的言語モデルを考え，それと音声認識との関係について言及する．

2 識別的言語モデル

2.1 識別モデルの導入

単語列を $\mathbf{w}$，i 番目の単語と単語 n-gram をそれぞれ w_i，w_{i-N+1}^{i} と表すと，従来の確率的単語 n-gram 言語モデルにおいて，生成確率を最大にする単語列の探索は次のように定式化される．

$$\mathbf{w}' = \arg\max_{\mathbf{w}} \sum_{i} \log P(\mathrm{w}_i | \mathrm{w}_{i-N+1}^{i-1}) \qquad (1)$$

この探索はViterbiアルゴリズムを用いて効率的に実行できる．

識別モデルを導入した場合も，この探索アルゴリズムを同様に使用することができる．このことは識別的単語モデルの導入を促進する一要因として挙げられる．識別的言語モデルにおける探索を定式化すると以下のようになる．

$$\mathbf{w}' = \arg\max_{\mathbf{w}} \sum_i \{\sum_k \lambda_k \phi_k(\mathrm{w}_{i-N+1}^i)\} \qquad (2)$$

$\phi_k(\mathrm{w}_{i-N+1}^i)$ は素性関数であり，w_{i-N+1}^i における番号 k に対応する情報の有無 (1 or 0) を表す二値関数である．λ_k は，その重みパラメータである．これを学習により推定する．

今，ϕ_k を単語 n-gram の有無に限定すると，特定の k' のみで $\phi_{k'}(\mathrm{w}_{i-N+1}^i) = 1$ となる．結果的に $\sum_k \lambda_k \phi_k(\mathrm{w}_{i-N+1}^i) = \lambda_{k'}$ というスコアを得る．つまり，単語列のスコアは各単語 n-gram に対応するスコアの和となり，その意味において式 (1) と同様である．

ここで，$\mathrm{f}_k(\mathbf{w}) = \{\sum_i \phi_k(\mathrm{w}_{i-N+1}^i)\}$ と表記する．これは $\mathbf{w}$ 中における k に対応する素性の頻度である．このとき式 (2) は

$$\mathbf{w}' = \arg\max_{\mathbf{w}} \sum_{k=1}^{K} \lambda_k \mathrm{f}_k(\mathbf{w}) = \arg\max_{\mathbf{w}} \boldsymbol{\lambda} \cdot \mathbf{f}(\mathbf{w}) \quad (3)$$

となる．$\cdot$ は内積演算である．この式は，K 種の観点から単語列 $\mathbf{w}$ を評価した結果，最大スコアを与える単語列を探索する問題と解釈できる．

識別的言語モデルの学習では，これら K 種の観点に基づき，正解単語列が対立単語列から識別されるように，パラメータ $\boldsymbol{\lambda}$ を決定する．パラメータの推定には，averaged-perceptron[9]，Conditional Random Fields (CRFs)[10]，boosting[11]，SVMs[12] など，様々な手法を利用できる．

素性関数には二値関数以外を定義することもでき，柔軟な設計が可能である．多様かつ豊富な素性を用いることで，多角的な観点から単語列の識別を行うことができる．

2.2 対立単語列を用いた学習

識別的言語モデルは，正解単語系列，対立単語系列の各々に内在する情報を利用して，系列間の識別を高めることでその効果を発揮している．

可能な単語列の総数は無限であるので，実用上，対立単語列を有限個に選定する必要がある．ここで選定された単語列に内在する情報しか利用できないため，選定作業は慎重に行われる必要がある．

一般的に用いられる方法のひとつは，デコーダの出力単語列を使用することである．これは音声認識

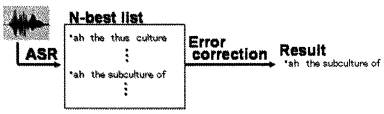

図 1: 誤り訂正．音声認識結果である単語 N-best リストの各単語列に，識別的言語モデルを適用（Error correction に相当）し，誤りの少ない単語列を選択．

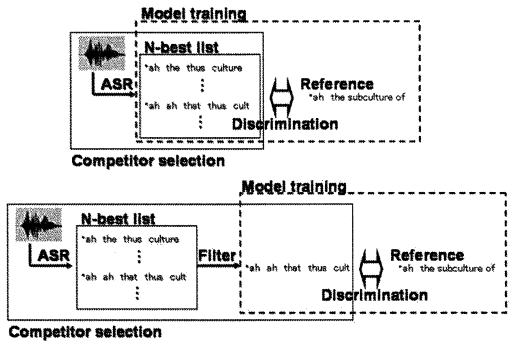

図 2: 誤り訂正モデルの学習．上段は音声認識単語列の集合である N-best リストを対立単語列とし学習を行う．下段では，N-best リストからさらに単語列を厳選し学習を行う．

や機械翻訳の分野で広く使用される．デコーダの出力単語列は，デコーダの処理パターンと入力に相関があるため，識別モデルの学習は間接的にこれらの情報を扱うことになる．

対立単語列に内在する情報は，素性により表現することになる．多様かつ豊富に素性を導入することで，多角的な観点から，その情報を扱うことができる．

これらの効果について例示するため，次節では，音声認識における識別的言語モデルに関する研究を幾つか紹介する．

2.3 音声認識における誤り訂正

Roark らは，音声認識システムが出力する複数の認識単語列 (単語ラティスや単語 N-best リスト) に識別的言語モデルを適用することで，認識精度の向上に成功している [13, 14, 15]．この手法は誤りを含む認識結果から正解単語列を識別する枠組みとなっていることから，誤り訂正法とも呼ばれる．

図 1 には，その手順を示している．まず音声認識システムを用いて複数単語列を出力する．これを $\mathcal{W}$ と表す．その後，$\mathcal{W}$ 中の各単語列に識別的言語モデルを適用し，より誤りの少ない単語列を次式に

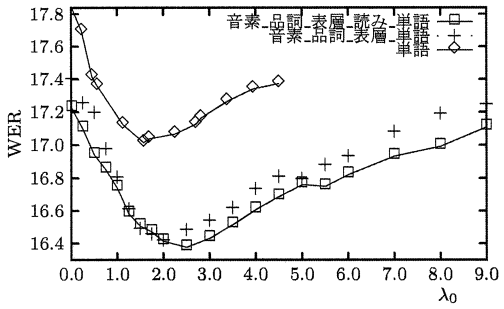

図 3: 音声誤り訂正モデルの素性別精度.

基づき探索する.

$$\mathbf{w}' = \arg\max_{\mathbf{w}\in\mathcal{W}}\{\lambda_0 \log P(\mathbf{o}|\mathbf{w})P(\mathbf{w}) + \boldsymbol{\lambda}\cdot\mathbf{f}(\mathbf{w})\} \quad (4)$$

つまり，音声認識スコアを，識別的言語モデルのスコアで補正することで，リランキングを行っている．λ_0 はスケーリングパラメータである．

これまでに，CRFsやSVMsなど，種々のパラメータ推定法が使用され，比較されている [16, 17, 18, 19] が，同じ素性を利用した場合，概ね同等の精度が得られている．図 2 上段は，その学習に関わる手続きを表している．対立単語列を複数の認識単語列とし，正解単語列との識別によりパラメータの推定を行う．

著者らは，対立単語列の単語誤り率 (WER) と誤り訂正モデルの精度の関係を分析し，誤りを多く含む単語列との識別が，高精度なモデル推定に効果的に作用することを実験的に示している [20, 21]．この中で，対立単語列を，単語 N-best リストの中で最も誤りを多く含む単語列のみに限定した場合について評価を行っている．具体的な手順を図 2 下段に示す．まず，音声認識システムを用いて単語 N-best リストを生成する．次に，その中から，図中 Filter において，WER の最も高い単語列を抽出し，それのみを対立単語列とする．この結果，図中上段における結果と比較して，同等もしくはそれ以上に高精度なモデルの生成を実現している．これは対立単語列の選定の重要性を示す結果である．

識別的言語モデルを音声認識において使用し始めた当初は，単語 n-gram 頻度のみが素性として用いられていた．しかし，品詞 n-gram 頻度素性の有効性が報告されており [22]，著者らも，学習データと言語的差異の大きなデータに対しても頑健に動作することを確認している．さらに，単語の読みや音素等の n-gram 頻度も同時利用することで，高精度な誤り訂正モデルを生成できることを報告している [23]．図 3 はその結果である．最終的に誤り訂正モデル適用前の WER= 17.8% を，16.4% まで低下させている．

結果から考察すると，単語より抽象度の高い品詞を素性に用いることで，多くの誤りパターンを表現し，単語や品詞では扱うことの難しい，音響モデルの誤りに由来する認識誤りを，音素素性の導入により取り込むことができたものと考えられる．既存の素性では扱うことの難しい情報を取り扱うことのできる素性を追加していくことで，それに沿って，より高精度なモデルを作成することができる．

3 識別的言語モデルの拡張と音声認識

既に述べた通り，識別モデルにおいては柔軟な素性設計が許されており，より音響的な性質を反映した音素単位等の素性を単語単位の素性と併せて利用することで，音声認識の高精度化を図ることも出来る．ここでは，さらに一歩踏み込んで，音響的な素性を直接的に取り扱う形態について考察する．

まず，素性ベクトルとして $\mathbf{f}(\mathbf{w})$ の代わりに，入力音声 $\mathbf{o}$ から得られる音響的素性も含む $\mathbf{f}(\mathbf{o},\mathbf{w})$ を導入し，これまでの式 (3) に対応する探索問題

$$\mathbf{w}' = \arg\max_{\mathbf{w}} \sum_{k=0}^{K} \lambda_k \mathrm{f}_k(\mathbf{o},\mathbf{w}) = \arg\max_{\mathbf{w}} \boldsymbol{\lambda}^{+}\cdot\mathbf{f}(\mathbf{o},\mathbf{w}) \quad (5)$$

を考える．具体的な音響的素性のひとつとして，例えば，音響モデルを介して得られる音声認識スコア $\mathrm{f}_0(\mathbf{o},\mathbf{w}) = \log P(\mathbf{o}|\mathbf{w})P(\mathbf{w})$ を想定すると，

$$\begin{aligned}\mathbf{w}' &= \arg\max_{\mathbf{w}}\{\lambda_0\mathrm{f}_0(\mathbf{o},\mathbf{w}) + \sum_{k=1}^{K}\lambda_k\mathrm{f}_k(\mathbf{o},\mathbf{w})\} \\ &= \arg\max_{\mathbf{w}}\{\lambda_0 \log P(\mathbf{o}|\mathbf{w})P(\mathbf{w}) + \boldsymbol{\lambda}\cdot\mathbf{f}(\mathbf{o},\mathbf{w})\}\end{aligned} \quad (6)$$

となり，式 (4) のスケーリングパラメータ λ_0 を他の重みパラメータと一緒に学習によって得ることで，音響的素性を含む識別モデル学習の一種が実現されるということがわかる．このほか $\mathrm{f}(\mathbf{o},\mathbf{w})$ の例として，各単語の信頼度や，単語のセグメント長と音素数の比などが考えられる．

素性 $\mathbf{f}(\mathbf{o},\mathbf{w})$ の導入により，単に音響的情報の利用による効果を得られるというだけでなく，言語的情報まで含めた大局的なパラメータ推定に基づく効果も期待できる．これにより，従来の音声認識で行われてきた，音響・言語モデルの独立な学習と比較して，より高精度なモデル生成が可能になると予想される．

また，$\mathbf{f}(\mathbf{o},\mathbf{w})$ をもとに，次式のような確率により単語列 $\mathbf{w}$ の評価を行うこともできる．

$$P^{+}_{\boldsymbol{\lambda}}(\mathbf{w}|\mathbf{o}) \triangleq \frac{\exp(\boldsymbol{\lambda}\cdot\mathbf{f}(\mathbf{o},\mathbf{w}))}{\sum_{\mathbf{w}}\exp(\boldsymbol{\lambda}\cdot\mathbf{f}(\mathbf{o},\mathbf{w}))} \quad (7)$$

つまり，識別的言語モデルの拡張として音響的素性までも直接取り扱う音声認識の枠組みは，音響的，言語的情報の双方を，素性という柔軟な空間で表現した上で，$\mathbf{o}$ が与えられたときの $\mathbf{w}$ の事後確率 $P^{+}_{\lambda}(\mathbf{w}|\mathbf{o})$ を直接求めようとする問題に相当しているとも解釈できる．

4 まとめ

近年注目を集めている識別的言語モデルについて言及し，音声認識を例に，その使用方法，効果のポイントを説明した．また，識別的言語モデルにおいて音響的情報を扱うように拡張する方法について検討し，音声認識との関係についても議論した．識別的アプローチの登場は我々に様々な恩恵を与えたことは間違いない．更なる発展に向けて，ここでの議論が少しでもヒントになれば幸いである．

参考文献

[1] Chelba, C., Engle, D., Jelinek, F., Jimenez, V. M., Khudanpur, S., Mangu, L., Printz, H., Ristad, E., Rosenfeld, R., Stolcke, A. and Wu, D.: Structure and Performance of a Dependency Language Model, in *Proceedings of Eurospeech*, pp. 2775–2778, Rhodes, Greece (1997).

[2] 森信介, 西村雅史, 伊東伸泰 : 構文構造を反映した確率的言語モデル, 情報処理学会研究報告. 2000-SLP-32-3, No. 64, pp. 13–18 (2000).

[3] Hofmann, T.: Probabilistic Latent Semantic Analysis, in *Proceedings of UAI* (1999).

[4] 秋田祐哉, 河原達也 : 話題と話者に関する PLSA に基づく言語モデル適応, 情報処理学会研究報告. 2003-SLP-49-12, Vol. 103, No. 517, pp. 67–72 (2003).

[5] 栗山直人, 鈴木基之, 伊藤彰則, 牧野正三 : 情報量基準で語彙分割した PLSA 言語モデルによる話題・文型適応, 情報処理学会研究報告. 2006-SLP-64, Vol. 2006, No. 136, pp. 233–238 (2006).

[6] Blei, D. M., Ng, A. Y. and Jordan, M. I.: Latent Dirichlet Allocation, *Journal of Machine Learning Research*, Vol. 3, pp. 993–1022 (2003).

[7] Erdogan, H., Sarikaya, R., Gao, Y. and Picheny, M.: Semantic Structured Language Models, in *Proceedings of the ICSLP*, pp. 933–936 (2002).

[8] Kawahara, T., Nemoto, Y. and Akita, Y.: Automatic Lecture Transcription by Exploiting Presentation Slide Information for Language Model Adaptation, in *Proceedings of ICASSP*, pp. 4929–4932 (2008).

[9] Collins, M.: Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with Perceptron Algorithms, in *Proceedings of EMNLP*, pp. 1–8 (2002).

[10] Lafferty, J., McCallum, A. and Pereira, F.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, in *Proceedings of Machine Learning*, pp. 282–289 (2001).

[11] Freund, Y. and Schapire, R. E.: A Decision-Theoretic Generalization of On-line Learning and An Application to Boosting, *Journal of Computer and System Sciences*, Vol. 55, (1997).

[12] Cortes, C. and Vapnik, V.: Support-Vector Networks, *Machine Learning*, Vol. 20, No. 3, pp. 273–297 (1995).

[13] Roark, B., Saraclar, M. and Collins, M.: Discriminative n-gram language modeling, *Computer Speech and Language*, Vol. 21, No. 2, pp. 373–392 (2007).

[14] Roark, B., Saraclar, M. and Collins, M.: Corrective language modeling for large vocabulary ASR with the perceptron algorithm, in *Proceedings of ICASSP*, Vol. 1, pp. 749–752 (2004).

[15] Roark, B., Saraclar, M., Collins, M. and Johnson, M.: Discriminative Language Modeling with Conditional Random Fields and the Perceptron Algorithm, in *Proceedings of ACL*, pp. 47–54 (2004).

[16] Zhou, Z., Gao, J., Soong, F. K. and Meng, H.: A Comparative Study of Discriminative Methods for Reranking LVCSR N-Best Hypotheses in Domain Adaptation and Generalization, in *Proceedings of ICASSP*, Vol. 1, pp. 141–144 (2006).

[17] Singh-Miller, N. and Collins, M.: Trigger-Based Language Modeling Using A Loss-Sensitive Perceptron Algorithm, in *Proceedings of ICASSP*, pp. 141–144 (2006).

[18] 大庭隆伸, 堀貴明, 中村篤 : 認識誤りに対する各単語 N-gram の関与度を考慮した誤り訂正学習, 春季音響学会講演論文集, pp. 73–74 (2007).

[19] 小林彰夫, 佐藤庄衛, 尾上和穂, 本間真一 and 今井亨, 都木徹 : 単語ラティスの識別的スコアリングによる音声認識, 秋季音響学会講演論文集, pp. 233–234 (2007).

[20] Oba, T., Hori, T. and Nakamura, A.: An Approach to Efficient Generation of High-Accuracy and Compact Error-Corrective Models for Speech Recognition, in *Proceedings on Interspeech2007*, pp. 1753–1756 (2007).

[21] 大庭隆伸, 堀貴明, 中村篤 : 誤り訂正モデルにおける単語誤り率基準での対立仮説選択とその効果, 秋季音響学会講演論文集, pp. 121–122 (2007).

[22] Shafran, I. and Hall, K.: Corrective Models for Speech Recognition of Inflected Languages, in *Proceedings of EMNLP*, pp. 390–398 (2006).

[23] 大庭隆伸, 堀貴明, 中村篤 : 識別的誤り訂正学習における対立単語列と素性の選定, 情報処理学会研究報告, 2007-SLP-69, Vol. 107, No. 405, pp. 235–240 (2007).