

DNNの低ランク近似と識別学習の組み合わせ法

Combination of Low-rank Approximation and Discriminative Training for DNN

太刀岡勇気 Yuuki Tachioka[†] 渡部晋治 Shinji Watanabe^{††} Jonathan Le Roux^{††} John Hershey^{††}

[†]三菱電機(株) 情報技術総合研究所 Information Technology R&D Center, Mitsubishi Electric Corporation

^{††}Mitsubishi Electric Research Laboratories

1 はじめに

音声認識の分野において、Deep neural network(DNN)は、従来のガウス混合分布に基づくモデルの性能を多くの場合上回るが、パラメータ数も多くなる。Xueらは、特異値分解(SVD)をDNNに適用し、パラメータ数を削減する方法を提案した[1]。一方で、系列識別学習により、クロスエントロピー(CE)基準で学習したDNNから性能が向上する[2]。ただし、低ランクモデルにおける識別学習の有効性とモデル縮減と識別学習の適用順が明らかでないため、本報では効果的な適用順を調査する。

2 低ランク近似によるDNNモデルサイズの縮減

DNNのパラメータ数を削減するために、SVDを用いた層*l*における重み行列 $A_{m \times n}^l$ を分解する[1]。

$$A_{m \times n}^l = U_{m \times k} [\text{diag}(\sigma_1, \dots, \sigma_n)] V_{k \times n}$$

σ は特異値($\sigma_1 \geq \dots \geq \sigma_n$)。行列 U と V は、直交正規化された列ベクトルを持つ。 k 番目までの特異値とそれに対応する特異値ベクトルにより低ランク近似する。

$$A_{m \times n}^l \approx U_{m \times k} [\text{diag}(\sigma_1, \dots, \sigma_k)] V_{k \times n} = A_{m \times k}^{l+\frac{1}{2}} A_{k \times n}^l$$

これにより図1のように、モデル縮減できる。

3 系列の識別学習とSVDの併用

系列の識別学習(ここでは相互情報量最大化(MMI))により、CEモデルから性能が向上する[2]。図2には、識別的に学習された低ランクモデルを構築するための3つの手順を示す。全手順に共通で、初期モデルはCE学習された完全モデルである。第1の手法(赤線)は、SVDとファインチューニング(FT)を初期モデルに適用し、その後CE低ランクモデルに対して識別学習を行う。第2の手法(緑線)は、MMI完全モデルにSVDとFTを適用する。第3の手法(青線)は、第2の手法によって得られたMMI低ランクモデルに、さらに識別学習を行う。

4 騒音下音声認識実験

非定常騒音環境下の5千語彙の音声認識タスクである、第2回CHiMEチャレンジトラック2(開発・評価セット)で単語誤り率(WER)の観点から評価した。騒音抑圧音声の12次元のMFCCとその動的特徴量を連結後に、特徴量変換(線形判別分析と最尤線形変換)および適応を行った。DNNの隠れ層数は3である。完全モデル(331中間ノード)から、SVDにより200ノードに削減し、パラメータ数は2.85Mから1.91Mに減少した。

図2には、上段に開発セットに対するWER、下段に評価セットに対するWERを示す。識別学習により、完

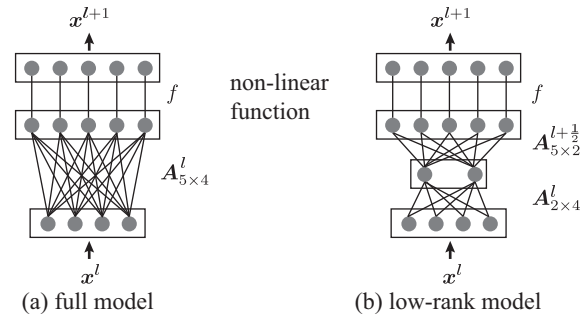


図1 Reducing DNN model parameters via low-rank factorization, from (a) $5 \times 4 = 20$ to (b) $5 \times 2 + 2 \times 4 = 18$.

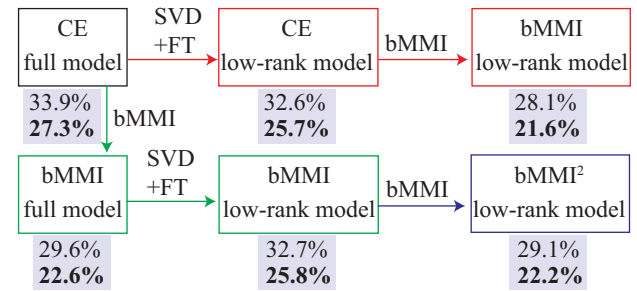


図2 Three approaches to generate SVD model with fine tuning (FT) and bMMI. The upper rows show the WERs for the development set of the second CHiME challenge, whereas the lower rows show those for the evaluation set.

全モデルの性能が顕著に向上した。赤線で示す手法1では、CE低ランクモデルにおいて識別学習が完全モデルよりも有効であった。緑線で示す手法2では、bMMI完全モデルを低ランク近似すると性能が低下した。青線で示す手法3では、bMMI低ランクモデルに識別学習を再び行うことで性能は向上したが、手法1よりも低い性能となった。よって、学習の手間が最も少なく、性能も最高である手法1が最良と結論付けられる。開発セットと評価セットで傾向は同じであった。

5 おわりに

DNNモデルのパラメータを削減する際には、モデル縮減後に識別学習を行うと最も効果的である。

参考文献

[1] J. Xue *et al.* "Restructuring of deep neural network acoustic models with singular value decomposition," INTERSPEECH, 2013.
 [2] K. Veselý *et al.* "Sequence-discriminative training of deep neural networks," INTERSPEECH, 2013.