論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Leveraging Better Training Targets for Deep Neural Network Acoustic Models in Speech Recognition
著者(和文)	PriceRyanWilliam
Author(English)	Ryan Price
出典(和文)	学位:博士（学術）, 学位授与機関:東京工業大学, 報告番号:甲第10361号, 授与年月日:2016年9月20日, 学位の種別:課程博士, 審査員:篠田 浩一,徳永 健伸,秋山 泰,村田 剛志,藤井 敦
Citation(English)	Degree:, Conferring organization: Tokyo Institute of Technology, Report number:甲第10361号, Conferred date:2016/9/20, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	審査の要旨
Type(English)	Exam Summary

（博士課程）

論文審査の要旨及び審査員

報告番号	甲第　　号		学位申請者氏名		Ryan William PRICE	
論文審査 審査員		氏　名	職　名		氏　名	職　名
	主査	篠田　浩一	教授	審査員	藤井　敦	准教授
	審査員	徳永　健伸	教授			
		秋山　泰	教授			
		村田　剛志	准教授			

論文審査の要旨（2000字程度）

この論文は、"Leveraging Better Training Targets for Deep Neural Network Acoustic Models in Speech Recognition"と題し、英文5章から成っている。

第1章「Introduction」では、研究の背景について述べた上で本論文の構成を示している。まず、近年の音声認識においては、Deep Neural Network (DNN)を用いた音響モデリングが主流であり、多くの応用場面で顕著に認識性能を向上させていることを述べている。そして、本論文では、DNNを用いた音響モデリングの実時間で応答する組み込み型の音声認識への応用と、教師ラベル付きの学習データが極めて少量である場面への適応の2つの課題に焦点を当てることを述べている。

第2章「Deep Neural Network Acoustic Modeling」では、まず、従来のガウス混合分布を出力分布とした隠れマルコフモデル(HMM)を用いる音響モデリングについて概観している。その上で、多層のフィードフォワード型ニューラルネットワークと、それとHMMを組み合わせたハイブリッド音響モデリングについて説明している。さらに、そのパラメータ学習における評価基準としてクロスエントロピーが主に用いられていることを説明し、そこでの教師ラベルは0か1の2値、すなわちhard targetであることを述べている。

第3章「Soft Target Training for Fast and Accurate Acoustic Models」では、組み込み型の音声認識への応用のためのDNNの小規模化、及び、入力特徴量正規化の実時間処理が課題であることを述べている。その上で、soft target trainingの概念を説明し、その従来研究を紹介している。Soft target trainingでは0か1の2値を教師ラベル(hard target)として学習された大規模なDNNの出力(0から1の間の実数)を教師ラベル(soft target)として用いて、より小さな規模のDNNを学習する。そして、上述の2つの課題を解決するために、正規化後の特徴量を用いて学習された大規模DNNの出力を教師とし、音声特徴量を正規化せずにそのまま入力として用いるsoft target trainingの手法を提案している。音声認識性能を評価した結果、提案法で学習された小規模DNNは、正規化された特徴量を用いて学習された大規模DNNに比べ、パラメータ数が約1/8であり、認識誤りを8.2%～11.2%削減したことを述べている。さらに、音声信号の有無を識別する音声区間検出のタスクに対しても提案法を適用し、その効果を確認したことを述べている。

第4章「Speaker Adaptation of DNN Using a Hierarchy of Output Layers」では、まず、音声認識においては少量の話者の発声を用いた話者適応手法が有効であること、DNNを用いた音声認識向けの話者適応手法はいまだ十分な性能をもつものがなく、その性能向上が課題であることを述べている。そのために、音素環境に依存したtriphoneの状態に対応する出力層をもつDNNの出力層の上に更に音素環境に依存しないmonophoneの状態に対応する出力層を載せ、その全体を少量の話者の発声を用いて学習する、すなわち階層的な教師ラベルを用いる適応手法を提案している。話者の発声に出現しないtriphone状態の出力も学習することが可能である。評価の結果、話者の1分間の音声を学習に使用することで、文誤り率を2.1%、文字誤り率を5.0%削減したことを述べている。

第5章「Conclusions and Future Work」では本論文で得られた成果をまとめ、将来の研究における課題について述べている。

以上で述べたように、本論文では、DNNを用いた音声認識の実用化に必要な、高性能かつ小規模DNNの構築方法と、少量データで性能を向上させる話者適応化法を提案し、その有効性を評価実験により確認している。本論文で得られた成果は、音声認識処理の発展に貢献するとともに、機械学習に基づくパターン認識手法全般において学術上寄与するところが大きい。よって本論文は博士（学術）の学位論文として十分価値があるものと認める。