様式Ｆ－１９

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成２５年　５月２６日現在

機関番号：３２６８９
研究種目：若手研究（B）
研究期間：2011 ～ 2012
課題番号：23700218
研究課題名（和文）　クラスタリングと教師なし適応学習に基づく時系列パターン認識システムの効率的な改善

研究課題名（英文）　Effective improvement of time-series pattern recognition systems using clustering and unsupervised adaptive training

研究代表者
小川　哲司（OGAWA　TETSUJI）
早稲田大学・理工学術院・准教授
研究者番号：70386598

研究成果の概要（和文）：音声データの構造化・検索支援のための基幹技術として，音声データを発話者や雑音といった音環境ごとにクラスタリングする技術の開発と，音声認識システムを教師なしの枠組みで適応的に最適化するための要素技術の開発を行った．

研究成果の概要（英文）：I developed technologies for clustering speech data into acoustic attributes such as speakers and types of noise and technologies for adaptively optimizing speech recognition systems in unsupervised ways. The developed technologies would be essential for constructing a system structuring speech data and a speech retrieval system.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
交付決定額	2,700,000	810,000	3,510,000

研究分野：総合領域
科研費の分科・細目：情報学・知能情報処理・知能ロボティクス
キーワード：クラスタリング・ベイズ学習・教師なし学習・音環境理解・パターン認識

１．研究開始当初の背景

研究代表者は，「状態と出力に相互依存関係を有する確率モデルの構造最適化と頑健性強化に関する研究」（平成 17 年度～18 年度，科研費若手研究（B）），「相補的な識別器の生成とその統合に基づくパターン認識に関する研究」（平成 19 年度～20 年度，科研費若手研究（B））を通じ，精度と頑健性を兼ね備えたパターン認識システムの構築法について検討を行い，これらが音声認識，ジェスチャ認識，性別・年齢推定の性能向上に寄与することを明らかにした．さらに，「モデル構造の逐次最適化機能を有するオンライン教師なし適応型パターン認識システム」（平成 21 年度～22 年度，科研費若手研究（B））において，確率モデルの精度と自由度を観測データに応じて適切に調整可能な枠組みについて検討を行った．しかし，研究代表者がこれまでに検討を行ってきたパターン認識システムの構築方式は，教師ラベルが付与されたデータを用いて学習を行う教師あり学習を主な対象としてきた．データに付与されたラベルの精度はそのラベル情報を用いて構築された認識システム（確率モデル）の性能に直結する．しかし，実生活で得られる知覚情報は極めて多様であり，その多様性を網羅するようあらゆるデータに対して教師ラベルを付与することは現実的ではなく，これがパターン認識アプリケーションの実用化，普及を妨げる要因となっている．以上の経験から，実環境におけるデータの多様性と膨大なコストを要する教師ラベルの付与の問題を同時に解決するために，人手を掛けずに日々成長するシステム，つまり教師ラベルの作成を極力排したパターン認識シ

ステムの構成法を確立することが必須であると確信するに至った.

２．研究の目的

教師なしクラスタリングに基づく環境適応型学習方式を開発することで,「人手を掛けずにシステムを日々成長させる」ための基幹システムを構築する．そのために，ノンパラメトリックベイズモデリングに基づき，音声データを類似した発話者，周辺雑音といった音環境ごとにクラスタリングする方式を開発する．さらに，得られた音環境クラスタ情報に基づいて適応学習を行う（音環境クラスタごとに確率モデルを時々刻々最適化する）ことで，音声認識システムに用いる確率モデルを教師なしの枠組みで効率的に改善する方式を開発する.

３．研究の方法

音声認識システムを教師なしの枠組みで効率的に改善するための基幹技術として，以下の異なるアプローチについて実装・評価を行った.

(1) シンプルな構成の音声認識システムを音環境変動に追従させることで，システムを適応的に最適化するアプローチ.
(2) 音声認識システムをあえて冗長に構成しておき，音環境の変動をモニタリングしながらシステムの性能を劣化させる要因をそぎ落とすことで，システムを適応的に最適化するアプローチ.

両アプローチともに，教師ラベルを陽に用いることなくパターン認識システムを環境変動に対して最適化する方式である．各アプローチの研究方法は以下の通りである.

(1) 環境変動に頑健な音環境クラスタリングと確率モデルの適応学習

シンプルな音声認識システムを音環境変動に追従させるアプローチにおいて，性能の鍵を握る音環境のクラスタリングに焦点を当てて検討を行った．本検討項目では，性質の異なる２つの方式について検討を行い，話者クラスタリングにより実証実験を行った．まず，①クラスタ数未知の問題を解くのに適したノンパラメトリックベイズモデリングの枠組みを積極的に活用したセグメント単位ディレクレ過程混合モデル（Segment-Oriented Dirichlet Process Mixture Model; SO-DPMM）を考案した．この方式は，データ数やデータの変動に対して頑健に高いクラスタリング精度の達成が期待できる．その一方，大規模なデータに対して計算量の増加が無視できず，また短い発話や複数人が同時に発話した場合に所望の性能が得られない可能性があった．そこで，これらの問題を低減可能な方式として，②i-vector と呼ばれる高精度な話者表現と非負値行列分解を用いた新たなクラスタリング方式を考案した.

このときに得られるクラスタリング結果に基づいて，音声認識に用いる確率モデル（音響モデル）を教師なしの枠組みで改善する方式について検討を行い，プロトタイプシステムの開発を行った.

(2) データドリブン型マルチストリーム音声認識システムの開発

冗長に構成された音声認識システムから性能劣化要因を時々刻々そぎ落とすアプローチとして，多数のシステムに対する音声認識性能を予測しながら，良好な性能を与える音声認識システムを逐次選択・統合することで，雑音の変動に対して頑健に高い性能を与える音声認識方式について検討を行った.

本検討項目では，①冗長な音声認識システムの構成法，②認識システムの性能を予測する性能モニタリング技術，③性能モニタリングの結果に基づき音声認識システムを最適化するための，システム選択・統合法に焦点を当てて検討を行った.

４．研究成果

＜研究の主な成果＞

(1)環境変動に頑健な音環境クラスタリング

データに応じてクラスタ数とモデルパラメータを同時に最適化可能な話者モデリングとして，セグメント単位ディレクレ過程混合モデル（SO-DPMM）を開発し，従来方式に対する有効性を示した．表 1 は，英語音声（TIMIT データベース），日本語音声（日本語話し言葉コーパス；CSJ）を用いた話者クラスタリング実験により，SO-DPMM の有効性を示した例である.

表１：様々な話者数・発話数のデータに対する話者クラスタリング精度．AHC-BIC は従来の話者クラスタリング方式．SO-DPMM が提案方式である．評価尺度である K 値は値が高い程性能が良いことを表す.

	Method	K value
英語音声① (24 話者, 192 発話)	AHC-BIC SO-DPMM	0.778 0.816
英語音声② (144 話者,1152 発話)	AHC-BIC SO-DPMM	0.516 0.680
日本語音声① (5 話者,599 発話)	AHC-BIC SO-DPMM	0.596 0.913
日本語音声② (20 話者,1290 発話)	AHC-BIC SO-DPMM	0.705 0.773
日本語音声③ (20 話者,4642 発話)	AHC-BIC SO-DPMM	0.239 0.782

この結果から，話者数やデータ量の変動に対して SO-DPMM は頑健に高い性能を与えることが明らかになった．

本検討項目においては，SO-DPMM の最適化方式についても詳細な検討を行い，データの量や質（背景雑音等）の変動に対して頑健に高い性能を与えるサンプリング方式を開発し，従来の変分ベイズ法に基づく最適化と比較して良好な性能を達成した．

また，発話内に複数話者が混在する場合に対して頑健な方式として，i-vector を話者表現として用いた非負値行列分解に基づく話者クラスタリングを開発し，データ量の変動に対して頑健な性能を達成した．

(2)データドリブン型マルチストリーム音声認識システムの開発

多数の多層パーセプトロンから得られる音素事後確率の推定値をモニタリングしながら信頼性の高い多層パーセプトロンを逐次選択することで，雑音に頑健なマルチストリーム音声認識を実現した．2 段階のストリーム形成処理により得られた 381 ストリームから，GMM の尤度に基づいて信頼性の高いストリームを選択し，その出力である音素事後確率を用いて HMM-ANN 音素認識を行った．図 1 に結果の一例を示す．地下鉄騒音下での連続音素認識実験において，提案するマルチストリーム音声認識システム（PM_GMM）は，従来のマルチストリームシステム（21B），一般的な（シングルストリーム）音声認識システム（Single）の性能を大幅に改善した．また，提案システムは，性能予測のために必要な時間間隔が短い場合であっても，良好な性能を得た．

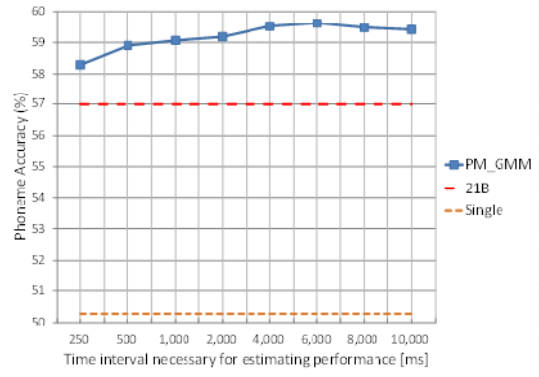

図１：性能予測に必要な時間長に対する適応的なマルチストリーム音声認識システムの音素正解精度．騒音は地下鉄騒音で SNR は 15dB．Single は一般的なシングルストリーム音声認識システム，21B は従来のマルチストリーム音声認識システム，PM_GMM は提案する性能モニタリング機能を有するマルチストリーム音声認識システム．

＜研究成果の国内外における位置づけと今後の展望＞

本課題で検討を行った枠組みは，システムの学習に用いるラベル情報の精度を教師なしで得られる限界のレベルにまで高めることを目指したものであり，教師なし学習の最高峰と位置づけられる．さらに本方式は，従来のパターン認識アプリケーションにおいて導入されている適応学習方式，つまりアプリケーションに閉じた形でシステムを改善するのではなく，様々なアプリケーションで入力されたデータに対して統一的に音環境クラスタリングを行うことで，アプリケーションやユーザの壁を取り払ってシステムを改善することを可能にするという特徴を持つ．つまり，様々なアプリケーションで入力されたデータに対して統一的にクラスタリングを行うことで，類似した（話者の）データが存在すれば，異なるアプリケーションや異なる話者のデータであっても効率的に流用してシステムを改善することが可能となる．これは，クラスタリングに基づく教師なし適応学習によって初めて実現できる枠組みである．また，データドリブン型マルチストリーム音声認識は，予め音声認識システムを冗長に生成しておき，性能を予測しながら時々刻々システムを最適化するというもので，クラスタ情報すら必要としない．これらの方式は，ラベル情報が付与されていない膨大なメディアデータの活用を大いに促進させるとともに，メディア情報の構造化の基幹技術となる可能性を持つ．さらに，これら性質の異なるアプローチを統合することで，更なる性能の向上も期待できる．

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 18 件）

① Tetsuji Ogawa, Feipeng Li, Hynek Hermansky, ``Stream selection and integration in multistream ASR using GMM-based performance monitoring,'' Proc. INTERSPEECH2013, Aug. 2013 (to appear).（査読有）

② 小川哲司, 松井知子, ``話者認識で用いる機械学習，” 日本音響学会誌，vol.69, no.7, July 2013 (to appear).（査読無）

③ 小川哲司, Li Feipeng, Hermansky Hynek, ``性能モニタリングに基づく多層パーセプトロンの適応的選択による雑音に頑健なマルチストリーム音声認識，” 音講論集，pp.167-170, March 2013.（査読無）

④ 網野加苗，石原俊一，小川哲司，長内隆，

黒岩眞吾, 越仲孝文, 篠田浩一, 柘植覚, 西田昌史, 松井知子, 王龍標, ``話者認識技術の現状と課題, '' 信学技法(SP), vol.112, no.450, pp.63-70, Feb. 2013. (査読無)

⑤ Hideitsu Hino and Tetsuji Ogawa, ``An improved entropy-based multiple kernel learning,'' Proc. ICPR2012, pp.1189-1192, Nov. 2012. (査読有)

⑥ Naohiro Tawara, Tetsuji Ogawa, Shinji Watanabe, Atsushi Nakamura, and Tetsunori Kobayashi, ``Fully Bayesian speaker clustering based on hierarchically structured utterance-oriented Dirichlet process mixture model,'' Proc. INTERSPEECH2012, Sept. 2012. (査読有)

⑦ 福地佑介, 俵直弘, 小川哲司, 小林哲則, ``i-vector に基づく発話類似度を用いた非負値行列分解と話者クラスタリングへの適用, '' 情処研報, vol.2012-SLP-92, July 2012. (査読無)

⑧ Naohiro Tawara, Tetsuji Ogawa, Shinji Watanabe, Tetsunori Kobayashi, ``Fully Bayesian inference of multi-mixture Gaussian model and its evaluation using speaker clustering,'' Proc. ICASSP2012, pp.5253-5256, March 2012. (査読有)

⑨ 俵直弘, 小川哲司, 渡部晋治, 中村篤, 小林哲則, ``階層的構造を持つディリクレ過程混合モデルを用いたフルベイズ話者クラスタリング, '' 信学技報(IBISML), vol.111, no.480, pp.21-28, March 2012. (査読無)

⑩ 小川哲司, 小林哲則, ``話者照合における因子分析に基づく特徴抽出に関する評価, '' 音講論集, pp.197-198, March 2012. (査読無)

⑪ 俵直弘, 小川哲司, 渡部晋治, 中村篤, 小林哲則, ``発話単位 DPMM を用いたフルベイズ話者クラスタリングと大規模データによる評価, ''音講論集, pp. 207-210, March 2012. (査読無)

⑫ 俵直弘, 小川哲司, 渡部晋治, 小林哲則, ``階層的発話生成モデルを用いた話者クラスタリングのためのフルベイズモデル推定手法の比較, '' 第 14 回情報論的学習理論ワークショップ(IBIS2011), D-117, Nov. 2011. (査読無)

⑬ 小川哲司, 日野英逸, 村田昇, 小林哲則, ``クラス内変動に頑健なカーネルマシンと話者照合への適用, '' 音講論集, pp.183-186, Sept. 2011. (査読無)

⑭ 俵直弘, 渡部晋治, 小川哲司, 小林哲則, ``多重混合ガウス分布モデルにおけるフルベイズモデル推定手法の検討と話者クラスタリングによる評価, '' 音講論集, pp.175-178, Sept. 2011. (査読無)

⑮ Tetsuji Ogawa, Hideitsu Hino, Noboru Murata, and Tetsunori Kobayashi, ``Speaker verification robust to intra-speaker variation using multiple kernel learning based on conditional entropy minimization,'' Proc. INTERSPEECH2011, pp.2741-2744, Aug. 2011. (査読有)

⑯ Naohiro Tawara, Shinji Watanabe, Tetsuji Ogawa and Tetsunori Kobayashi, ``Speaker clustering based on utterance-oriented Dirichelet process mixture model,'' Proc. INTERSPEECH2011, pp.2905-2908, Aug. 2011. (査読有)

⑰ 小川哲司, 日野英逸, 村田昇, 小林哲則, ``条件付きエントロピー最小化基準に基づくマルチカーネル学習を用いた発話スタイル変動に頑健な話者照合, '' 情処研報, vol.2011-SLP-87, July 2011. (査読無)

⑱ Tetsuji Ogawa, Hideitsu Hino, Nima Reyhani, Noboru Murata, and Tetsunori Kobayashi, ``Speaker recognition using multiple kernel learning based on conditional entropy minimization,'' Proc. ICASSP2011, pp.2204-2207, May 2011. (査読有)

６．研究組織

(1)研究代表者

小川　哲司（OGAWA TETSUJI）
早稲田大学・理工学術院・准教授
研究者番号：70386598