様式 C-19

科学研究費補助金研究成果報告書

平成２３年　５月２０日現在

機関番号：３２６８９
研究種目：若手研究（B）
研究期間：2009～2010
課題番号：21700205
研究課題名（和文）モデル構造の逐次最適化機能を有するオンライン適応型パターン認識に関する研究
研究課題名（英文）A study on online adaptive pattern recognition with sequential optimization of model structures

研究代表者
小川 哲司（OGAWA TETSUJI）
早稲田大学・高等研究所・助教
研究者番号：70386598

研究成果の概要（和文）：パターン認識システムの精度とシステムを使用する環境の変動に対する頑健性を効率的に向上させるために，データの性質に応じて，認識システムに用いる確率モデルの構造と分布パラメータを適応的に最適化する方式を開発した．さらにこの枠組みを，音声情報を用いた話者認識や画像情報を用いた顔認識システムに適用することを試みた．

研究成果の概要（英文）：I developed a method of adaptively optimizing both the structure and parameters of statistical models used in pattern recognition systems to effectively improve robustness of those systems to environmental changes. In addition, I attempted to apply this framework to speaker recognition systems using speech information and face recognition systems using image information.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
21 年度	1,500,000	450,000	1,950,000
22 年度	1,100,000	330,000	1,430,000
年度			
年度			
年度			
総　計	2,600,000	780,000	3,380,000

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：　パターン認識，話者認識，顔認識

１．研究開始当初の背景

音声認識，顔・ジェスチャ認識，文字認識などでは，パターン認識技術が利用される．申請者は，パターン認識の精度と信頼性の向上を目的に，パターン認識に用いる確率モデルの精密化（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方式（平成 19 年度～20 年度，科研費若手研究（B)）について検討を行い，それらが音声認識や性別・年齢推定の性能向上に寄与することを確認した．これらは，主に確率分布におけるパラメータの静的な最適化について検討したものであり，観測データに応じた適応的な最適化や，混合正規分布（Gaussian Mixture Model; GMM）における混合数のような「確率モデルの構造」の最適化が検討課題として残っていた．一方，パターン認識技術を利用したアプリケーションは，環境の変動に対処するため，確率モデルの学習に利用するデータ（学習データ）を逐次取得し，モデルを適応的に最適化する「日々成長するシステム」であることが望ましい．実際，環境変動の影響により，過去に得られたデータとは性質が大きく異なるデータ（既存モデルに対する「はずれ値」）が学習データとして度々得られる．このような場合に，構造を固定したままモデルの最適化を試みたところ，確率分布の分散が不当に増大し，その結果として認識性能の大幅な劣化を招くことを確認した．つまり，このような「はずれ値」をモデル内部で適切に表現するためには，モデル構造を観測データに応じて適切に調整しながら，パラメータを最適化する必要がある．

２．研究の目的

環境の変動に頑健なパターン認識システムを構築し，音声や顔画像情報に基づく人物認識（話者認識）における有効性を明らかにする．そのために，以下の 4 項目について検討を行う．なお本研究では，混合正規分布（Gaussian mixture model; GMM）のような生成的なアプローチと，サポートベクタマシン（SVM）をはじめとするカーネル法のような識別的なアプローチの双方について検討を行う．

(1) GMM を対象とし，データが得られる度にモデルの構造（GMM の混合数）と分布パラメータを同時に最適化する枠組みを構築し，モデル構造と分布パラメータを同時に最適化することの有効性を確認する．

(2) GMM を対象とし，モデルのパラメータのみならず，話者クラス数をも推定可能な話者モデリングの方式を開発する．これにより，教師なしの枠組みで話者をモデル化することを目指す．

(3) 話者認識においてカーネル法に基づく識別的なアプローチは，GMM よりも良好な性能を与えることが知られている．そこで，生成的アプローチにおけるモデル構造に相当するカーネル関数の種類やそのチューニングパラメータを，データの性質に応じて厳密に決定することなく良好な性能が得られるマルチカーネル学習（multiple kernel learning; MKL）を基礎として，環境変動の影響を受けにくい話者モデリングの方式を開発することを試みる．

(4) 以上の方式を，音声を用いた話者認識システムと顔画像を用いた人物認識システムに適用し評価を行うことで，データの性質の変動に対する提案方式の頑健性を明らかにする．

３．研究の方法

(1) 変分ベイズ学習によるモデル構造と分布パラメータの最適化

逐次入力される学習データの性質に応じてGMMの混合数を適応的に制御しながら分布パラメータの最適化を行う方式について検討を行った．ここでは，教師あり学習の枠組みのもと，変分ベイズ学習に基づくアルゴリズムの定式化と実装を行った．この枠組みを顔画像による人物認識システムに適用することで，基本的な知見を得た．

(2) 教師なし話者モデリングと話者クラスタリングへの応用

教師なしで話者情報をモデル化する枠組みとして，データの性質に応じて話者クラス数を推定しながら，分布パラメータを同時に最適化する方式を，ノンパラメトリックベイズモデリングの枠組みを用いて開発した．提案方式を音声情報を用いた話者クラスタリングに適用し，従来のベイズ情報量基準（BIC）に基づく凝集的クラスタリングと比較を行った．

(3) 発話スタイル変動に頑健なカーネル法に基づく話者モデリング

データのクラス内変動に対して頑健なマルチカーネルの構築方式である，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（MKL based on Conditional Entropy Minimization; MCEM）を話者モデリングに適用した．音声を用いた話者照合実験を行い，提案方式と従来のマージン最大化に基づく MKL（RMKL）を実験的に比較した．このとき，異なる発話スタイルの音声を用いて学習を行うことで，データの性質の変動に対する提案方式の頑健性についても明らかにした．

４．研究成果

＜研究の主な成果＞

(1) 変分ベイズ学習によるモデル構造と分布パラメータの最適化

学習データを逐次取得する度に，既存モデルの学習に用いたデータと逐次入力されるデータの全てを用いて（バッチ型），GMM の混合数と分布パラメータ（平均・分散・重み）を同時に最適化する枠組みを変分ベイズ法に基づき実装した．さらに，GMM を用いた人物認識／話者認識のための識別器として，変分ベイズ予測に基づく識別器および最尤識別に基づく識別器を構築した．以上の評価を，顔画像を用いた人物認識において行った．撮影環境が異なるデータが日々蓄えられる（学習データとして追加される）度に，GMM の分布パラメータと混合数を同時に最適化する実験を行ったところ，以下の結果を得た．

① 撮影環境が異なるデータが学習データとして追加されることで，最適なモデル構造（混合数）が変化する．

② モデルパラメータのみならずモデル構造も同時に最適化することで，構造を最適化しない最尤法を使った枠組みの誤りを削減した．

(2) 教師なし話者モデリングと話者クラスタリングへの応用

逐次入力されるデータの性質に応じて話者クラス数と分布パラメータを同時に最適化する枠組みを，ノンパラメトリックベイズモデリングの枠組みを用いて実装した．ここでは，発話を単位としたディリクレ過程混合モデル（Utterance-Oriented Dirichlet Process Mixture Model; UO-DPMM）を定式化し，観測データに応じて話者クラス数が適切に決まる（無限の話者数を扱える）柔軟な枠組みで話者のモデル化を実現した．話者数を与えることなくデータと話者の対応関係が得られることから，本方式により，教師なしの枠組みで話者のモデル化が実現されたと言える．

本方式を音声による話者クラスタリング問題に適用し，BIC に基づく従来方式と比較を行ったところ，提案方式は従来方式よりも高速かつ高精度な話者クラスタリングを実現できることが明らかになった．さらに，提案方式はチューニングパラメータの値の変動に対しても頑健に高い性能を与えることを示した．これは，実用上重要な性質である．

(3) 発話スタイル変動に頑健なカーネル法に基づく話者モデリング

発話した時期の違いや発話スタイルの変動に頑健でチューニングが容易な話者認識システムを構築した．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（MCEM）を用いて話者識別器を構築することで，データは特徴空間においてクラスごとに密集し，クラス間では散らばる．この性質により，話者内のデータ変動に頑健な認識を可能とした．40 名の話者照合実験の結果を表１に示す．本実験により以下の知見が得られた．

① 提案した MCEM に基づく話者照合システムは、従来のマージン最大化に基づくシステム（RMKL）に対して良好な性能を与えた．

② MCEM に基づくシステムは、RMKL に基づくシステムと比較して発話スタイルの変動に対して頑健な性能を与えた．

表 1：RMKL に基づく SVM（従来方式）および MCEM に基づく SVM（提案方式）を用いた話者照合システムの性能比較

Talking style		RMKL	MCEM	
Training data	Test data	EER (%)	EER (%)	Improve. (%)
NC	NC	6.3	5.9	6.2
pinLC75	pinLC75	8.0	7.7	4.2
Average		7.1	6.8	5.2
NC+pinLC75	NC	9.3	8.1	12.0
	carLC60	10.0	9.5	4.7
	pinLC60	9.8	8.8	10.3
	pinLC70	8.4	7.3	14.1
	pinLC75	8.7	8.3	5.4
	depLC60	9.3	8.1	12.9
	depLC70	9.3	8.5	8.1
	depLC75	9.8	9.0	8.4
Average		9.3	8.4	9.5

表 1 には，様々な発話スタイルの音声入力に対して算出した Equal Error Rate (EER)(%) を示した．ここで，NC は通常発声の音声，LC は雑音下で発話した際に生じるロンバード効果を含んだ音声を意味する．car，pin，dep は各々車内騒音，デパート地下騒音，ピンクノイズを聞きながら発話した音声であることを表し，60，70，75 は発話者に提示される騒音の耳元での音圧（dBA）を表す．RMKL に基づくシステムに対する MCEM に基づくシステムの EER の改善率（Improve）に着目すると，学習データが発話スタイル変

動を含まない場合（NC，pinLC75）と比較して，発話変動を含む場合（NC+pinLC75）の方が EER の改善率が高いことから，MCEM に基づくシステムが発話スタイル変動に対して頑健であることが実験的に明らかになった．

＜研究成果の国内外における位置づけ＞

パターン認識システムの学習データとランタイムにおける入力データの性質の違いに対処するため，逐次的に得られるデータを用いて適応的にシステムを最適化する研究は国内外を問わず盛んに行われている．特に，記述長最小基準やベイズ基準などを情報規範としてモデル構造を選択する方式の有効性が認められている．しかし，音声情報処理など大規模なデータを対象としたシステムでは，この方式は教師ありの枠組みで実現されている例がほとんどである．また，既存の確率モデルと逐次入力される少量データのみを用いた適応的なシステム最適化として，事後確率最大化推定や最尤線形回帰などのモデル適応技術に基づいて，既存モデルのパラメータを更新する方式が一般的に用いられている．しかし，これらはモデル構造を固定した上での枠組みであり，そのままでは既存モデルに対するはずれ値に対処することは困難である．一方，成果(2)で開発した方式は，生成的なアプローチにおいて，モデル構造と分布パラメータのみならず，クラス数をも同時に最適化することを可能とした話者情報の教師なし学習の枠組みであり，他に例を見ない画期的なものと言える．

また，カーネル法に基づく方式は，マージンを最大化するように学習するのが一般的である．これは，データのクラス間での散らばりを最大化するというもので，クラス内での散らばりについては陽には考慮されていない．それに対して成果(3)において開発した MCEM に基づくシステムでは，データのクラス間での散らばりを最大化するのみならず，クラス内での散らばりを最小化しようとするため，話者認識システムにおける性能劣化の要因である，同一話者（同一クラス）内における発話スタイルの変動や発話時期の差の影響を低減することを可能とする．その意味で，成果(3)として得られた話者照合システムは，環境変動に対する頑健性を理論的に備えていると言える．

＜今後の展望＞

観測データに応じた認識システムの動的な最適化は，音声認識，話者認識，顔・ジェスチャ認識などの分野において必須の要素技術である．しかし従来方式は，はずれ値のデータを学習データとして適切に取り扱えない，オンライン処理を行うシステムに不向きである，などの問題がある．一方，本提案方式は，既存モデルに対してはずれ値となるデータをも適切にモデルに反映することを可能とし，アプリケーションを使用する環境の変動に対する頑健性をパターン認識システムに与える．これらの機能を実現することで，パターン認識技術の適用可能性を格段に拡げることが予想される．これらの機能は教師あり学習の枠組みであっても十分に価値があるが，それらを教師なし学習の枠組みに拡張した研究成果(2)を足がかりとし，パターン認識技術の究極の目的とも言える，「人手を掛けずに日々成長するシステム」を構築することが可能になると考えられる．また，本方式を用いた音声情報の構造化（ダイアリゼーション）は，話者のみならず，周辺雑音や性別，話し手の感情といった様々な音環境の構造化をも可能とする．

さらに，本課題の成果である高精度な人物認識システムは，セキュリティシステム，音声対話システム，会議の構造化支援システムといったアプリケーションの機能の拡充，性能向上に寄与すると考えられる．

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４件）

① 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）

② 俵直弘, 渡部晋治, 小川哲司, 小林哲則, ``発話を単位としたディリクレ過程混合モデルに基づく話者クラスタリング，'' 日本音響学会講演論文集，pp.41-44, March 2011.（査読無）

③ 小川哲司，日野英逸，Nima Reyhani，村田昇，小林 哲則，``マルチカーネル学習を用いた話者認識における最適化の検討，'' 情報処理学会研究報告，vol.2010-SLP-84，Dec. 2010.（査読無）

④ 小川哲司，日野英逸，Nima Reyhani，村田昇，小林哲則，``情報論的な最適化に基づくマルチカーネル学習を用いた話者認識，'' 日本音響学会講演論文集，pp.81-84，Sept. 2010.（査読無）

６．研究組織

(1)研究代表者

小川　哲司（OGAWA TETSUJI）

早稲田大学・高等研究所・助教

研究者番号：70386598