

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 16 日現在

機関番号：32689

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25330207

研究課題名(和文) システム協調型適応学習に基づくパターン認識システムの全体最適化に関する研究

研究課題名(英文) A study on total optimization of multiple pattern recognition systems using cooperative and 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

交付決定額(研究期間全体)：(直接経費) 3,700,000円

研究成果の概要(和文)：複数のパターン認識システムを協調的に最適化しながら、システム全体を効率的かつ自動で成長させる方式について検討を行った。特に、システムの利用者や使用環境の違いにより生じるデータの変動に対してシステムを適応的に改善するための要素技術として、データの性質の違いに頑健なクラスタリング技術および、複数システムを協調的に用いて認識を行うマルチストリームパターン認識フレームワークの開発を行った。

研究成果の概要(英文)：Attempts have been made to cooperatively optimize multiple pattern recognition systems, developing a total system efficiently and automatically. Specifically, the clustering technique that is robust against the environmental changes and multistream pattern recognition framework, which cooperatively exploits information yielded from multiple systems, have been developed as the fundamental technologies for adaptively refining the systems to cope with the changes in characteristics of data (e.g., users and surrounding environments of the system).

研究分野：知覚情報処理

キーワード：パターン認識 性能予測 深層学習 全体最適化

1. 研究開始当初の背景

パターン認識を用いたアプリケーションを実環境において頑健に動作させるためには、使用者や使用環境の違いにより生じる多様なデータ変動に対して適応的に成長するシステムの実現と、膨大なコストを要する教師ラベル付与の低減、効率化が必須である。申請者は、「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))、「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」(平成 23 年度～24 年度、科研費若手研究(B))を通じ、主に音声認識システムを対象として、教師ラベルの作成を極力排しながら、ユーザやアプリケーションを制限せずに効率的に認識システムを改善する(人手を掛けずにシステムを成長させる)方式について検討を行ってきた。一方で、「リズムのある会話を可能とするコミュニケーションロボットに関する研究」(平成 20 年度～22 年度、科研費基盤研究(B)・研究代表者:小林哲則)に連携研究者として参加した際、ヒューマンインタフェースは複数のパターン認識システムを複雑に統合した枠組みであり、個々のシステム(例えば、音声認識システム)のみを適切に改善できても、その効果をシステム全体に適切に反映する枠組みがなければ、システム全体として見た場合改善の効果は限定的であることも痛感した。以上の経験から、実環境で頑健に動作するパターン認識アプリケーションを実現するためには、異なる属性や複数のパターン認識システム間の協調的な全体最適化(個々のシステムの改善がその他のシステムや統合されたシステムに効果的に反映され、統合システムの改善が個々のシステムの改善に効果的に反映される方式)に基づくアプローチでシステムを構築することが重要であると確信するに至った。提案アプローチを採用することで、一部の認識システムにおいて教師ラベル情報が得られればシステム全体を効果的に改善できることから、教師ラベル付与の低減、効率化も期待できる。

2. 研究の目的

複数のパターン認識システムを協調的に最適化しながら、システム全体を効率的かつ自動で成長させる枠組みについて検討を行う。特に、システムの利用者や使用環境の違いにより生じるデータの変動に対して、システムを適応的に改善するための要素技術を開発する。

3. 研究の方法

複数のパターン認識システムの協調的な最適化を実現するための要素技術として、主に以下の項目について検討を行った。

(1) ニューラルネットワークに基づく識別

器の不確かさを推定する技術の開発:

深層ニューラルネットワーク(DNN)に基づくパターン認識システムが出力する認識結果の不確かさを推定する技術について検討を行った。この技術は、パターン認識システムにメタ認知機能(すなわち、「わかっているかわからないかわかる」機能)を導入することに相当し、複数のパターン認識システムを協調的に利用する際の重要な要素技術となる(つまり、信頼性の高いシステムの結果を信頼し、信頼性の低いシステムの結果は信頼せずに用いる)。本研究では、深層自己符号化器の復元誤差に基づく不確かさの推定方式に焦点を当てて検討を行い、(2)で述べるマルチストリーム型パターン認識システムにおける入力データに応じたシステムの適応的な選択や、雑音抑圧のためのフィルタの推定などに用いた。

本研究では、DNNに基づく識別器により事後確率を生成し、事後確率ベクトルを用いて構築した深層の自己符号化器(図1)の誤差を用いて識別器の不確かさを予測することを試みた。自己符号化器の復元誤差は、学習データ分布に従うデータを入力した場合に小さくなるため、学習データと評価データのミスマッチの度合いを測定可能である。したがって、事後確率ベクトルを入力した際に得られる復元誤差は、学習データに対する入力データの事後確率(認識結果)の不確かさを表す。

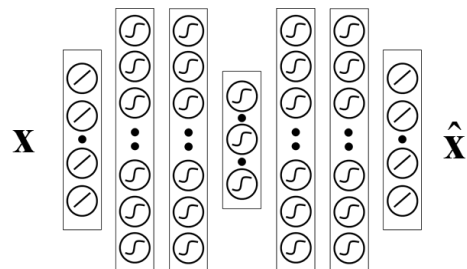


図1: 深層自己符号化器。7層のニューラルネットワークにより構成される。中間にボトルネック層を有する。

(2) マルチストリーム型パターン認識による環境変動に頑健な教師なし適応技術の開発:

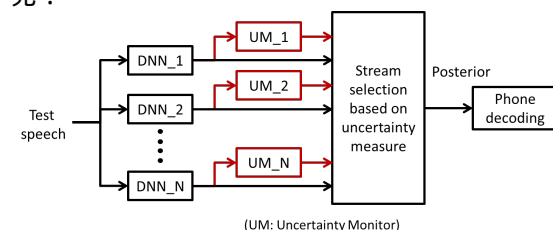


図2: ストリーム選択に基づくパターン認識システムの構成(音声認識の例)

異なる現象を扱うパターン認識システムを深層学習により多数構築しておき、最適なシステムを入力データの性質に応じて選択

的に用いることで、環境の変化や未知環境に対して頑健に高い性能を与える認識方式を開発し、音声認識に焦点を当て評価を行った(図2)。異なる雑音環境ごとに構築した複数のシステムを(1)で開発した技術により選択することで、入力データに最も性質が類似する雑音環境に対するシステムを適切に選択できるかどうか、また、音声認識において一般的に用いられているマルチスタイル学習(様々な雑音下音声を用いた単一システムの構築)の性能を上回るかどうかについて検証を行った。さらに、異なる音響特徴ごとに構築したシステムを選択的に用いることを試み、雑音や残響と言ったより広範な音響環境の違いに対しても頑健に高い性能を与えるか検証を行った。

(3)環境変動に頑健なクラスタリング技術の開発:

(1)および(2)に関する検討を通じ、類似する性質のデータに対して認識システムを構築し、それらを選択的に用いることの有効性が見て取れた。このとき、データの類似性を推定する技術(クラスタリング)が重要となる。本研究では、時系列データに適した特徴ベクトル(i-vector)を用いて計算した類似度行列に対してスペクトラルクラスタリングを適用することで、雑音等の環境変動に頑健なクラスタリング方式を提案し、発話者のクラスタリングにおいて有効性を検証した。前述の通り、このクラスタリングの結果は、複数システムを協調的に最適化する際に利用可能である。

4. 研究成果

(1)ニューラルネットワークに基づく識別器の不確かさを推定する技術の開発:

DNNに基づく認識器の出力(事後確率)の不確かさ、つまり入力データに対する認識器の信頼性を自己符号化器から得られる復元誤差を用いて予測する方式の有効性を調査した。当該技術については、2014年7月にプラハで開催されたサマワークショップ(2014 Frederick Jelinek Memorial Workshop)の音声認識グループ(ASR Machines That Knows When They Do Not Know)にて海外の研究者らとともに集中的に研究を行った。本ワークショップでは、それまで検討を行ってきたDNN出力の時間的な変動量(M measureと呼ぶ)に基づく方式の拡張や自己符号化器の復元誤差を用いる方法の提案を行い、事後確率のエントロピーなど従来の認識システムの信頼性推定技術に対する有効性を明らかにした。

【主な業績:雑誌論文10, 17】

また、ここで開発した技術は、産業機器の故障予兆検知に応用していく予定である。

(2)マルチストリーム型パターン認識による環境変動に頑健な教師なし適応技術の開

発:

DNNの出力(事後確率)の不確かさを自己符号化器から得られる復元誤差を用いて予測し、マルチストリーム音声認識におけるストリーム選択に利用した。

11種類の雑音条件(クリーン, 9種類の雑音, それらの混合)各々に対して構築したストリーム(DNNに基づく識別器と自己符号化器に基づく不確かさの推定器)のうち、自己符号化器の復元誤差が最小となるストリームを発話ごとに選択し連続音素認識を行った。評価は学習済みストリームに含まれる9種類の雑音(Seen noise)および含まれない6種類の雑音(Unseen noise)の下で発話された音声を用いて行った。

評価対象となるシステムは以下の5通りである。1) Multistyle: クリーン音声と9種類の雑音重畳音声を混合して構築したシステム, 2) Matched (condition): 入力発話と同一種類の雑音重畳音声で構築されたストリームを手動で選択して用いたシステム(正解ストリームに相当), 3) UM_ENTROPY (従来法): DNN出力のエントロピーを最小にするストリームを選択して用いるシステム, 4) UM_M (従来法): DNN出力のフレーム間距離の平均値(M measure)を最大にするストリームを選択して用いるシステム, 5) UM_AE_LDA (提案法): DNN出力の復元誤差を最小にするストリームを選択して用いるシステム(UM_AE_LDA-top2は復元誤差が小さい2つのストリームを統合して用いたシステム)。

図3より、想定した入力(Seen)であれば、提案するストリーム選択に基づくシステム(UM_AE_LDA)は、入力と同じ特性を持つストリーム(最適なストリームを手動で選択した場合)と同等の性能を、またマルチスタイル学習により構築したシステムよりも良好な性能を与えた。また、図4より、提案システム(UM_AE_LDA)は想定外の入力(Unseen)であれば、マルチスタイル学習によるシステムと同等の性能を与えた。

【主な業績:雑誌論文6, 9, 31】

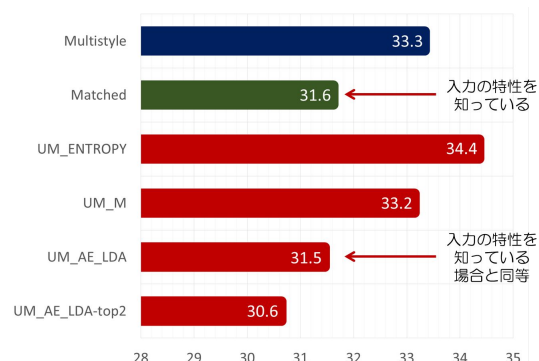


図3: システムの不確かさに基づくストリーム選択の効果。システムが想定した雑音環境下(seen noise)での発話音声(400発話)に対する音素誤り率。

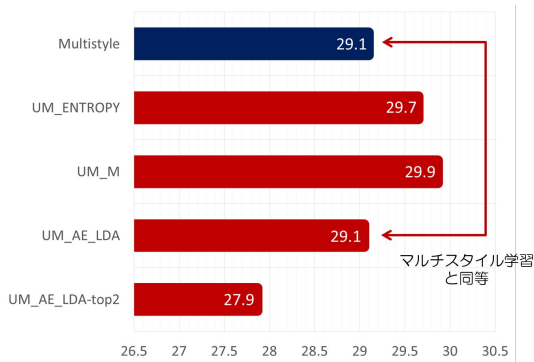


図4：システムの不確かさに基づくストリーム選択の効果．システムが想定していない雑音環境下（unseen noise）での発話音声（400発話）に対する評価．

（3）環境変動に頑健なクラスタリング技術の開発：

i-vector に基づく発話表現とスペクトラルクラスタリングを組み合わせたクラスタリング法において、学習データの性質と評価データの性質にミスマッチが生じる場合の影響を調査した．

図5より、従来技術であるk平均法によるクラスタリングでは音響環境（言語・背景雑音等）が一致しないときに性能が著しく低下するのに対し、スペクトラルクラスタリングでは性能の低下がごくわずかにとどまることを実証し、実データに対して極めて有効なクラスタリング法であることを明らかにした．

【主な業績：雑誌論文7，16，30】

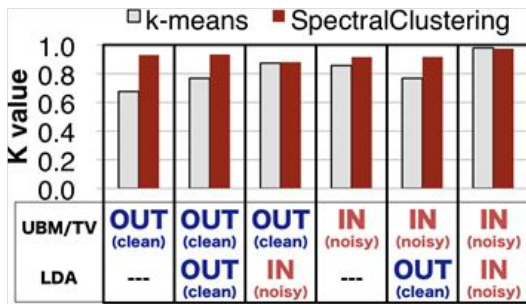


図5：k平均法とスペクトラルクラスタリングを用いた場合の話者クラスタリング精度．話者モデルのパラメータ（UBM，TV行列，LDA行列）の学習データがインドメイン（評価データと同一の性質）の場合とアウトドメイン（評価データと異なる性質）の場合について評価を行った．スペクトラルクラスタリングは、学習データと評価データの性質が異なる場合であっても常に良好な性能を与えることが明らかになった．

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33件）

- 1) 小川哲司, Mallidi Harish, Vesely Karel,

Hermansky Hynek, “ニューラルネットワークに基づく識別器の不確かさの推定とマルチストリーム音声認識への適用,” 音講論集, pp.67-70, March 2016. (査読無)

- 2) 小川哲司, 小林哲則, 新田恒雄, “高次相関を考慮した音響特徴量のDNNに基づく音声認識での利用,” 音講論集, pp.161-162, March 2016. (査読無)
- 3) 大町基, 小川哲司, 小林哲則, 藤枝大, 片桐一浩, “連想記憶に基づく線形分離行列推定を用いたタンデム接続型音源分離,” 音講論集, pp.21-24, March 2016. (査読無)
- 4) 佐藤洋輔, 小川哲司, 堀内靖雄, 黒岩眞吾, “深層学習を用いた出現音素の偏りに頑健な話者照合手法,” 電子情報通信学会総合大会, March 2016. (査読無)
- 5) 浅見太一, 大谷大和, 小川哲司, 木下慶介, 倉田岳人, 齋藤大輔, 塩田さやか, 太刀岡勇気, 中村静, 増村亮, 渡部晋治, “国際会議 INTERSPEECH2015 参加報告,” 情処研報, Feb. 2016. (査読無)
- 6) Sri Harish Mallidi, Tetsuji Ogawa, Hynek Hermansky, “Uncertainty estimation of DNN classifiers,” Proc. ASRU2015, pp. 283-288, Dec. 2015. (査読有)
- 7) Naohiro Tawara, Tetsuji Ogawa, Shinji Watanabe, Atsushi Nakamura, Tetsunori Kobayashi, “A sampling-based speaker clustering using utterance-oriented Dirichlet process mixture model and its evaluation on large-scale data,” APSIPA Trans. Signal & Info. Process., vol.4, e16 doi:10.1017/ATSIP.2015.19, Oct. 2015. (査読有)
- 8) Tetsuji Ogawa, Kenshiro Ueda, Kouichi Katsurada, Tetsunori Kobayashi, Tsuneko Nitta, “Bilinear map of filter-bank outputs for DNN-based speech recognition,” Proc. INTERSPEECH2015, pp.16-20, Sept. 2015. (査読有)
- 9) Sri Harish Mallidi, Tetsuji Ogawa, Karel Vesely, Phani Nidadavolu, Hynek Hermansky, “Autoencoder based multi-stream combination for noise robust speech recognition,” Proc. INTERSPEECH2015, pp.3551-3555, Sept. 2015. (査読有)
- 10) Motoi Omachi, Tetsuji Ogawa, Tetsunori Kobayashi, Masaru Fujieda, Kazuhiro Katagiri, “Blind source separation using associative memory model and linear separation filter,” Proc. EUSIPCO2015, pp.1103-1107, Sept. 2015. (査読有)
- 11) 森岡幹, 俵直弘, 小川哲司, 岩田具治, 小川厚徳, 堀貴明, 小林哲則, “複数の文脈長を考慮したリカレントニューラルネットワークに基づく言語モデル,” 音講論集, pp.17-20, Sept. 2015. (査読無)
- 12) 大町基, 小川哲司, 小林哲則, 藤枝大, 片桐一浩, “連想記憶に基づくブラインド

- 音源分離のエコーキャンセリングへの応用,” 音講論集, pp.593-596, Sept. 2015. (査読無)
- 13) 俵直弘, **小川哲司**, 小林哲則, “スペクトラルクラスタリングに基づく話者クラスタリングのための因子分析法の効果の検証,” 音講論集, pp.173-174, Sept. 2015. (査読無)
- 14) Kesaaki Minemura, **Tetsuji Ogawa**, Tetsunori Kobayashi, “Feature extraction for rotary-machine acoustic diagnostics focused on period,” Proc. INTERNOISE2015, Aug. 2015. (査読有)
- 15) 岡本拓磨, **小川哲司**, 落合翼, 柏木陽佑, 亀岡弘和, 木下慶介, 郡山知樹, 齋藤大輔, 篠崎隆宏, 高木信二, 滝口哲也, 太刀岡勇氣, 俵直弘, 橋本佳, 藤本雅清, 松田繁樹, 三村正人, 吉岡拓也, 渡部晋治, “国際会議 ICASSP2015 参加報告,” 情処研報, July 2015. (査読無)
- 16) Naohiro Tawara, **Tetsuji Ogawa**, Tetsunori Kobayashi, “A comparative study of spectral clustering for i-vector-based speaker clustering under noisy conditions,” Proc. ICASSP2015, pp.2041-2045, April 2015. (査読有)
- 17) Hynek Hermansky, Lukas Burget, Jordan Cohen, Emmanuel Dupoux Naomi Feldman, John Godfrey, Sanjeev Khudanpur, Matthew Maciejewski, Sri Harish Mallidi, Anjali Menon, **Tetsuji Ogawa**, Vijayaditya Peddinti, Richard Rose, Richard Stern, Matthew Wiesner, Karel Vesely, “Towards machines that know when they do not know: Summary of work done at 2014 Frederick Jelinek Memorial workshop,” Proc. ICASSP2015, pp.5009-5013, April 2015. (査読有)
- 18) 上田賢次郎, **小川哲司**, 小林哲則, 桂田浩一, 新田恒雄, “テンソル積による基底変換に基づく音声認識に関する研究,” 音講論集, pp.7-10, March 2015. (査読無)
- 19) 大町基, **小川哲司**, 小林哲則, 藤枝大, 片桐一浩, “連想記憶と線形分離フィルタを用いたブラインド音源分離,” 情処研報, Feb. 2015. (査読無)
- 20) 俵直弘, **小川哲司**, 小林哲則, “i-vectorを用いたスペクトラルクラスタリングによる雑音環境下話者クラスタリング,” 情処研報, Feb. 2015. (査読無)
- 21) 浅見太一, 岩野公司, **小川哲司**, 駒谷和範, 齋藤大輔, 篠田浩一, 太刀岡勇氣, 東中竜一郎, 福田隆, 増村亮, 渡部晋治, “国際会議 INTERSPEECH2014, SLT2014 参加報告,” 情処研報, Feb. 2015. (査読無)
- 22) Yuichi Kubota, Motoi Omachi, **Tetsuji Ogawa**, Tetsunori Kobayashi, and Tsuneo Nitta, “Effect of frequency weighting on MLP-based speaker canonicalization,” Proc. Interspeech2014, pp.2987-2991, Sept. 2014. (査読有)
- 23) 俵直弘, **小川哲司**, 小林哲則, “スペクトラルクラスタリングに基づく話者クラスタリング,” 音講論集, pp.95-98, Sept. 2014. (査読無)
- 24) 久保田雄一, 大町基, **小川哲司**, 小林哲則, 新田恒雄, “MLPを用いた話者正準化に基づく音声認識の検討,” 情処研報, July 2014. (査読無)
- 25) **小川哲司**, 塩田さやか, “i-vectorを用いた話者認識,” 日本音響学会誌, vol.70, no.6, pp.332-339, June 2014. (査読無)
- 26) 久保田雄一, 大町基, **小川哲司**, 小林哲則, 新田恒雄, “標準話者母音スペクトルへの変換に基づく話者正準化,” 音講論集, pp.77-78, March 2014. (査読無)
- 27) Walaa Gouda, Walid Gomaa, and **Tetsuji Ogawa**, “Vision based SLAM for humanoid robots: A survey,” Proc. JEC-ECC2013, pp.170-175, Dec. 2013. (査読有)
- 28) Hideitsu Hino and **Tetsuji Ogawa**, “Integration of MKL-based and i-vector-based speaker verification by short utterances,” Proc. ACPR2013, pp.562-566, Nov. 2013. (査読有)
- 29) 福地佑介, 俵直弘, **小川哲司**, 小林哲則, “因子分析モデルに基づく話者照合の環境変動に対する頑健性の調査,” 音講論集, pp.75-78, Sept. 2013. (査読無)
- 30) Naohiro Tawara, **Tetsuji Ogawa**, Shinji Watanabe, Atsushi Nakamura, and Tetsunori Kobayashi, “Blocked Gibbs sampling based multi-scale mixture model for speaker clustering on noisy data,” Proc. MLSP2013, Sept. 2013. (査読有)
- 31) **Tetsuji Ogawa**, Feipeng Li, and Hynek Hermansky, “Stream selection and integration in multistream ASR using GMM-based performance monitoring,” Proc. Interspeech2013, pp.3332-3336, Aug. 2013. (査読有)
- 32) **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, pp.349-356, July 2013. (査読無)
- 33) 俵直弘, **小川哲司**, 渡部晋治, 中村篤, 小林哲則, “効率的なサンプリング手法を用いた話者モデリング,” 情処研報, July 2013. (査読無)

〔図書〕(計 1件)

- 1) **小川哲司**, “話者ダイアライゼーション,” 音響キーワードブック, 音響学会(編), コロナ社, pp.434-435, 2016. (総ページ数: 482ページ)

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