

様 式 C - 1 9、F - 1 9 - 1、Z - 1 9 （共通）

科学研究費助成事業

研究成果報告書



令和 元 年 6 月 2 4 日現在

機関番号：3 2 6 8 9

研究種目：若手研究(B)

研究期間：2017 ~ 2018

課題番号：1 7 K 1 2 7 1 8

研究課題名（和文）セグメントクラスタリングに基づく未知ドメインデータからの知識獲得法の提案

研究課題名（英文）Knowledge acquisition from unknown domain data with segmental clustering

研究代表者

俵 直弘（TAWARA, Naohiro）

早稲田大学・理工学術院・講師（任期付）

研究者番号：5 0 7 2 6 2 5 5

交付決定額（研究期間全体）：（直接経費） 3,000,000 円

研究成果の概要（和文）：系列データから特徴的なパターンを発見し分類することで、事前知識が存在しない未知のドメインデータから、そのドメイン特有のクラス単位を自動的に獲得する枠組みの検討を行った。特に、複数の側面を持つ系列データから所望のクラスを発見するための要素技術として、クラス獲得のためのクラスタリング技術、目的のクラスに関し識別的な特徴を選択的に抽出するための特徴抽出法、そして既知ドメイン知識を組み合わせて未知ドメインを表現するための枠組みを提案しその有効性を示した。

研究成果の学術的意義や社会的意義

本研究成果により、まだ有効な手法が確立されていない未知ドメインからの知識獲得や既知ドメイン知識から未知ドメイン知識を得るための新たな指針を与えることができた。また、本研究で提案した複数の側面をもつ観測データから特定の側面の情報を選択的に取り出すことができるディスエンタングリング特徴抽出アプローチは、本研究で想定した未知ドメインデータのみならず、より汎用的なパターン認識問題への適用が期待される。

研究成果の概要（英文）：Attempts have been made to discover domain-specific classes from an unknown domain. Specifically, the clustering technique to discover class, feature extraction technique to obtain class discriminative feature, and expressing unknown domain by combining known domain knowledge have been developed as the fundamental technologies for obtaining desired class from sequence data observed in the target domain.

研究分野：知覚情報処理

キーワード：パターン認識 クラスタリング 深層学習 ドメイン適応 低資源言語音声認識 音響モデル 言語モデル

1. 研究開始当初の背景

音声信号のような明確な境界区切りが存在しない系列データを対象としたパターン認識では、認識単位となるクラスの事前設計が重要となる。例えば音声認識では、音声学や言語学の知見に基づき音声ラベルを設計し、得られた音声ラベル系列と音響系列との対応関係を学習することで、音声信号からテキストへの変換を実現する。一方、例えば文字を持たない言語のように、区切りに関し十分な知見が存在しない未知のドメインでは、このようなラベル系列と観測系列のペアに基づくアプローチの適用が困難だった。そこで本研究では、未知ドメインの系列データ集合から、特有の部分系列（セグメント）を発見し、そのドメイン特有のクラス単位を自動的に獲得する枠組みの実現を目指す。申請者は先行研究で、可変長セグメントを表現するための確率的生成モデルを導入することで、従来は扱いが困難であった長さが変動するセグメントを対象としたクラスタリング手法を提案した。その研究の中で、柔軟なクラス獲得を実現するためにはセグメント内の多様な変動構造を考慮した表現を扱える枠組みが不可欠であるという確信に至った。同時に、本モデルは各クラスで同一の複雑さで表現されるため、クラス毎に複雑さが変化するデータに対し性能が悪化する問題が未解決だった。このため所望のクラスの違いに起因する変動とそれ以外の違いに起因する変動とを適切に弁別できる識別的な特徴抽出を行う枠組みも必要であるという結論に至った。

2. 研究の目的

系列データから特徴的なパターン（セグメント）を発見し分類する枠組みを構築することで、事前知識が存在しない未知のドメインで収集したデータからそのドメイン特有のクラス単位を自動的に獲得できる枠組みの検討を行う。特に、データから各クラスの複雑さを推定することによるクラス獲得法と、獲得した認識クラスに特有な識別的な特徴を抽出できる特徴抽出法に焦点を当て要素技術を開発する。

3. 研究の方法

未知ドメインにおけるクラス獲得を実現するための要素技術として、以下の項目について検討を行った。

(1) 深層ニューラルネットワークに基づく認識クラス獲得と識別的特徴抽出法の開発

クラス単位が事前に得られない未知ドメインで収集したデータから、何らかの意味のあるクラス単位を抽出するためには、頻出するセグメントを自動的に発見し獲得する技術が必要となる。そこで、本研究では、ディリクレ過程混合モデルと呼ばれるクラスタ数を自動決定できる確率過程を用いて対象ドメインデータをクラスタリングすることで、対象ドメイン特有のクラスを獲得する枠組みの実現を試みた。図1に本枠組みを未知言語の音声信号に適用し得られる音素クラス獲得のイメージ図を示す。このとき、獲得したクラスの認識に不要な変動成分（例えば話者の違い）が本手法に与える影響を調査し、未知ドメインデータからクラス獲得するための課題を検証した。さらにこれら識別に悪影響を与える変動成分を抑圧することで、目的の情報のみを選択的に抽出できる特徴抽出法を深層ニューラルネットワークにより実現することを試みた。

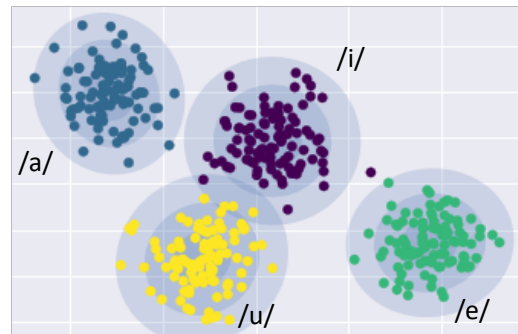


図 1. ディリクレ過程混合モデルによるクラス獲得のイメージ図. 各プロットが入力系列、薄い楕円が獲得したクラスを示す。

(2) 複数の認識クラスを分離抽出（ディスエンタングリング）する識別的表現学習法の開発

(1)では獲得したクラスの識別を容易にする特徴量を得るために、獲得したクラス情報以外に起因する変動成分を抑圧するDNNを導入することで目的クラスに関し識別的な特徴を得た。しかし用途によっては不要な情報を抑圧するのではなく、それぞれ別の特徴量として分離・抽出する枠組みが求められる場合がある。本研究ではこの分離を行う枠組みをディスエンタングリングと呼び、(1)で得られた枠組みを拡張することで、複数の異なるクラス単位に対する識別的な特徴量をそれぞれ分離抽出する新たなDNNの実現を試みた。

(3) ドメイン属性による未知ドメイン表現法の開発

(1), (2)に関する検討を通じ、未知ドメインのデータに対しパターン認識を適用する際に、学習時に利用できる既知ドメインの知識を適切に利用することで未知ドメインにおけるパターン認識の性能を改善できることが確認できた。そこで、既知ドメインの情報を有効利用するための新たな枠組みとして、既知ドメインを複数の要素（ドメイン属性）に分解し、その組み合わせにより未知ドメインを表現する枠組みの開発を行った。

具体的には、文書からの言語モデル推定を対象として、図2に示すように対象文書のドメインは話し方（style）と話題（Topic）からなるドメイン属性に分解できると仮定し、これらドメイン属性ごとに専用の言語モデル（Expert 言語モデル）をそれぞれ作成した。未知ドメインデータに適用するときはこれら Expert 言語モデルを動的に組み合わせる枠組みを導入することで、未知ドメインの文書に対しても頑健に機能する言語モデルの実現を試みた。

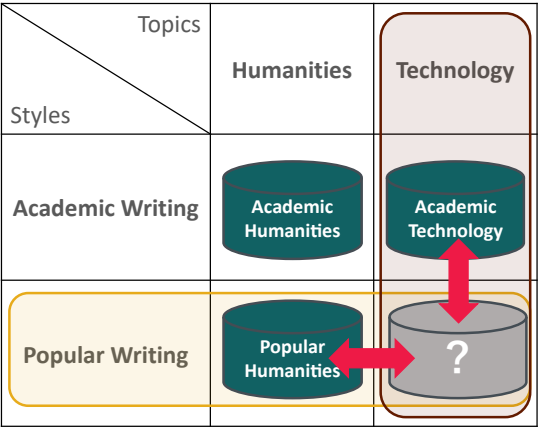


図 2. ドメイン属性による未知ドメイン表現。ドメインをTopics とStyles の2種類のドメイン属性に分解し、未知ドメインをこれらドメイン属性の組合せで表現する。

4. 研究成果
＜研究の主な成果＞

(1) 深層ニューラルネットワークに基づく認識クラス獲得と識別的特徴抽出法の開発

未知ドメインの系列データからクラス獲得を行う典型的な応用例として、未知言語の音声からその言語の音素体系を獲得するゼロ資源言語音声認識問題を取り上げた。

本研究では、まず、未知言語の音声データに対し、提案するディリクレ過程混合モデル(DPMM)に基づくクラス単位獲得法を適用することで、対象言語特有の音素クラス体系が得られることを示した。しかし、ここで得られたクラスは、音韻クラスのみならず話者の違いに起因するクラスも独立した単位として得られていることを明らかにした。そこで、獲得した音素単位から話者の違いに起因するクラスを取り除くための枠組みとして、図3に示すDNNを新たに提案した。このDNNは、DPMMで獲得した音素クラス情報を用いて、音声データからこれら獲得音素クラスの事後確率を推定する特徴抽出器を構築することで、対象言語特有の音素識別に適した特徴抽出を行う。さらに、ここで得られた特徴量から話者推定を行うDNNを構築し、この話者識別DNNを騙すように特徴抽出器を最適化する敵対的学習の枠組みを導入することで、音素識別はできるが話者識別は行えない特徴量を得る。この枠組みにより話者性のみを除去した特徴抽出を実現した。表1に3言語に対し音素識別実験を行った結果を示す。提案手法で得られた識別的特徴量を用いることで、音響特徴量をそのまま用いた場合や既存言語の音素体系を用いた得られた特徴量を用いた場合よりも高い精度が得られることを示した。

【主な業績：雑誌論文2, 3, 6, 13, 16, 17】

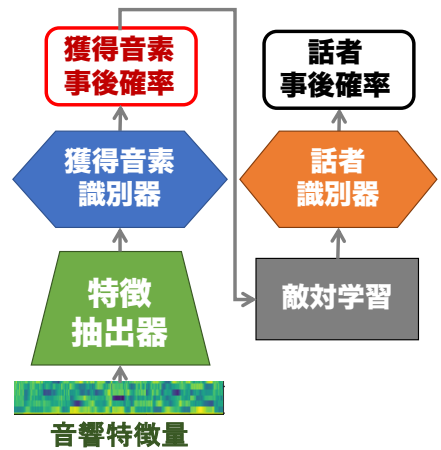


図 3. 獲得したクラスから話者変動を除去する DNN. 敵対学習は後段の話者識別器の性能を下げるように特徴抽出器を最適化する。

表 1. 未知言語の音素照合エラー率 [%]
(ただし、英語、フランス語、ツォンガ語のいずれも未知言語とみなした)。従来法では既知言語の音素体系得られた特徴量を用いた。

Feature	English	French	Xitsonga
音響特徴量	10.8	14.8	25.7
DANN-BNF (従来法)	8.7	12.2	19.7
DANN-post (提案法)	8.2	11.4	12.5

(2) 複数の認識クラスを分離抽出（ディスエンタングリング）する識別的特徴抽出法の開発

(1)で得られたDNNを拡張し、複数の異なるクラス分類に起因する変動をもつデータから各クラス特有の特徴を分離して抽出するDNNを提案した（図4）。このDNNは特徴抽出器と各々のク

ラスの識別器から構成される．特徴抽出器の目的は，音声から音素特徴量と話者特徴量を同時に抽出することで，このとき，各特徴量が各々のクラス識別器で識別しやすくなるように特徴抽出器を最適化することで，各クラスに関する識別的特徴量を得る．このとき，各々の特徴量をもう一方の識別器に入力した場合にはその識別性能が下がるような排他的な制約を与えることで，互いに排他的な特徴量を得ることができることを示した．表2に示すように，各特徴量が互いに排他的になる制約を与えることで各々の識別問題において性能が改善されていることから，提案法により分離性能が改善されたことがわかる．

【主な業績：雑誌論文9，15】

(3) ドメイン属性による未知ドメイン表現法の開発

提案するドメイン属性を用いて構築した言語モデルを用いて，単語系列から次の単語を推定した．文書のドメインとして，その文書の話題（科学，人文など）とスタイル（学術文書，大衆文学）からなる2種類のドメイン属性を定義し，未知ドメインはこれらドメイン属性の組み合わせで表現できるとした．

提案するアプローチに基づき構築した言語モデルの構造を図5に示す．ドメイン属性毎にLong short-term memory (LSTM) による言語モデルを構築し，Mixerが各ドメイン属性の言語モデルの重みを自動決定する．この構造により，未知ドメインの文書に対しても，単語予測に有効な既知ドメイン属性を動的に選択できる枠組みを提案した．表3に示すように提案したドメイン属性を考慮することで未知ドメインに対する言語モデルの性能が改善されることを示した．

【主な業績：雑誌論文7，10，14，21】

5. 主な発表論文等

〔雑誌論文〕（計23件）

- (1) **Naohiro Tawara**, Tetsunori Kobayashi, Tetsuji Ogawa, “Multi-channel speech enhancement using time-domain convolutional denoising autoencoder,” *Proc. Interspeech2019*, Sept. 2019 (to appear). (査読有)
- (2) Yosuke Higuchi, **Naohiro Tawara**, Tetsunori Kobayashi, Tetsuji Ogawa, “Adversarial Training of DPGMM-based Feature Extractor for Zero-Resource Languages” *Proc. Interspeech2019*, Sept. 2019 (to appear). (査読有)
- (3) 樋口陽祐, **俵直弘**, 小林哲則, 小川哲司, “敵対的訓練を用いたゼロリソース言語のためのDPGMMに基づく話者の違いに不変な特徴抽出,” *情報処理学会研究報告*, July 2019. (to appear) (査読無)
- (4) **Naohiro Tawara**, Hikari Tanabe, Tetsunori Kobayashi, Masaru Fujieda, Kazuhiro Katagiri, Takashi Yazu, Tetsuji Ogawa, “Postfiltering Using an Adversarial Denoising Autoencoder with Noise-aware Training” *Proc. ICASSP2019*, pp. 3282–3286, May 2019 [doi: 10.1109/ICASSP.2019.8682684]. (査読有)

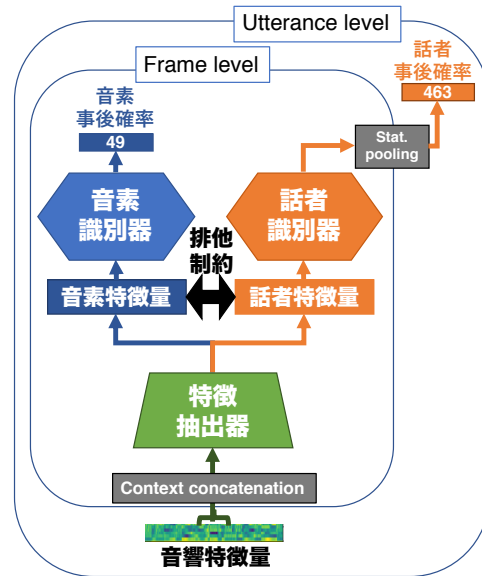


図4. 特定の情報を分離抽出するDNN（音声から話者と音韻を分離する例）

表2. 音韻(Phone)性と話者(Speaker)性を分離するDNNから得られた特徴量をそれぞれ音素・話者照合精度[%]で評価．

特徴量	音素照合精度	話者照合精度
音響特徴量	70.9	58.7
分離制約なし	83.2	74.7
分離制約有り	84.0	75.9

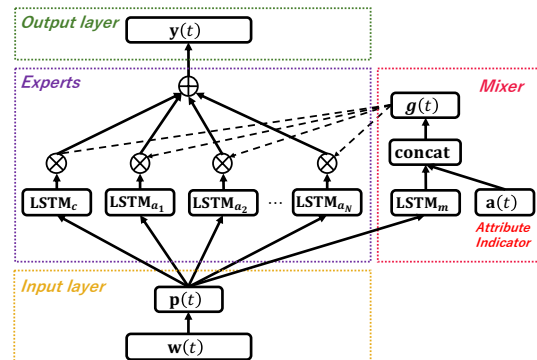


図5. ドメイン属性に基づく未知ドメインモデル推定（言語モデルの例）. Expertsの各ブロックが各ドメイン属性で学習したモデルを示し，未知ドメインの入力系列 $w(t)$ に対しMixerが各ドメイン属性の重みを推定し，選択的に利用して次の系列 $y(t)$ を予測する．

表3. ドメイン属性の導入による効果．未知ドメインのドキュメントに対する各言語モデルのテストセットパープレキシティ．

モデル	Academic Writing		Popular Writing	
	Humanities	Technology	Humanities	Technology
(1) ベースライン言語モデル	265.61	343.25	260.81	357.37
(2) ドメイン単位で学習	266.94	342.57	259.05	362.60
(3) ドメイン属性単位で学習	263.58	338.49	254.75	353.21

- (5) 田辺ひかり, 俵直弘, 小林哲則, 藤枝大, 片桐一浩, 矢頭隆, 小川哲司, “敵対的デノイジングオートエンコーダを用いた拡散性雑音除去,” *電子情報通信学会技術研究報告 (SP)*, SP2018-87, pp.155-160, March 2019. (査読無)
- (6) 樋口陽祐, 俵直弘, 小川哲司, 小林哲則, “ゼロリソース言語音声認識のための発話者の違いに頑健な特徴抽出,” *日本音響学会講演論文集*, pp.923-924, March 2019. (査読無)
- (7) 芦川博人, 森岡幹, 俵直弘, 小川厚徳, 岩田具治, 小川哲司, 小林哲則, “ドメイン属性情報を用いた RNN 言語モデルのドメイン汎化,” *日本音響学会講演論文集*, pp.927-930, March 2019. (査読無)
- (8) 俵直弘, 田辺ひかり, 小林哲則, 藤枝大, 片桐一浩, 矢頭隆, 小川哲司, “Noise-aware 学習を用いた敵対的デノイジングオートエンコーダによるポストフィルタリング,” *日本音響学会講演論文集*, pp.159-162, March 2019. (査読無)
- (9) 俵直弘, 小林哲則, 小川哲司, “話者・音韻特徴抽出のためのディスエンタングリングニューラルネットワークの実現にむけて,” *日本音響学会講演論文集*, pp.1003-1004, March 2019. (査読無)
- (10) 芦川博人, 森岡幹, 俵直弘, 小川厚徳, 岩田具治, 小川哲司, 小林哲則, “RNN 言語モデルのためのドメイン属性情報を用いたゼロショット学習,” *情報処理学会研究報告*, vol.2018-SLP-125, no.13, pp.1-6, Dec. 2018. (査読無)
- (11) 俵直弘, 小林哲則, 藤枝大, 片桐一浩, 矢頭隆, 小川哲司, “敵対的デノイジングオートエンコーダによる非線形ひずみ除去フィルタリング,” *日本音響学会講演論文集*, pp.159-160, Sept. 2018. (査読無)
- (12) Naohiro Tawara, Tetsunori Kobayashi, Masaru Fujieda, Kazuhiro Katagiri, Takashi Yazu, Tetsuji Ogawa, “Adversarial autoencoder for reducing nonlinear distortion” *Proc. APSIPA2018*, pp.1669-1673, Nov. 2018. [doi: 10.23919/APSIPA.2018.8659540] (査読有)
- (13) Taira Tsuchiya, Naohiro Tawara, Tetsunori Kobayashi, Tetsuji Ogawa, “Speaker invariant feature extraction for zero-resource languages with adversarial training,” *Proc. ICASSP2018*, pp.2381-2385, April 2018. [doi: 10.1109/ICASSP.2018.8461648] (査読有)
- (14) Tsuyoshi Morioka, Naohiro Tawara, Tetsuji Ogawa, Atsunori Ogawa, Tomoharu Iwata, Tetsunori Kobayashi, “Language model domain adaptation via recurrent neural network with domain-shared and domain-specific representations,” *Proc. ICASSP2018*, pp.6084-6088, April 2018. [doi: 10.1109/ICASSP.2018.8462631] (査読有)
- (15) 俵直弘, 土屋平, 小川哲司, 小林哲則, “敵対的学習に基づく話者特徴抽出,” *日本音響学会講演論文集*, pp.141-144, March 2018. (査読無)
- (16) 島田拓也, 俵直弘, 小川哲司, 小林哲則, “話者正規化における言語非依存性とゼロリソース音声認識における効果,” *日本音響学会講演論文集*, pp.109-112, March 2018. (査読無)
- (17) 土屋平, 俵直弘, 小川哲司, 小林哲則, “敵対的学習を用いた話者の違いに頑健な特徴抽出とゼロリソース音素識別による評価,” *日本音響学会講演論文集*, pp.9-12, March 2018. (査読無)
- (18) 高木信二, 倉田岳人, 郡山知樹, 塩田さやか, 鈴木雅之, 玉森聡, 俵直弘, 中鹿亘, 福田隆, 増村亮, 森勢将雅, 山岸順一, 山本克彦, “国際会議 Interspeech2017 報告,” *研究報告音声言語情報処理 (SLP)*, 2018-SLP-120(14), pp.1-9 (2018-02-13), 2188-8663, Feb. 2018. (査読無)
- (19) Hiroto Ashikawa, Naohiro Tawara, Atsunori Ogawa, Tomoharu Iwata, Tetsunori Kobayashi, Tetsuji Ogawa, “Exploiting end of sentences and speaker alternations in recurrent neural network-based language modeling for multiparty conversations,” *Proc. APSIPA2017*, Dec. 2017. [doi: 10.1109/APSIPA.2017.8282217] (査読有)
- (20) 芦川博人, 俵直弘, 小川厚徳, 岩田具治, 小林哲則, 小川哲司, “複数人対話を対象とした RNN 言語モデルにおける発話終端情報利用の有効性,” *日本音響学会講演論文集*, Sept. 2017. (査読無)
- (21) 森岡幹, 俵直弘, 小川哲司, 小川厚徳, 岩田具治, 小林哲則, “ドメイン依存・非依存の内部表現を有する再帰型ニューラルネットワーク言語モデル,” *日本音響学会講演論文集*, pp.27-30, Sept. 2017. (査読無)
- (22) 菊池康太郎, 俵直弘, 小林哲則, “辞書定義文を用いたゼロショット一般物体認識,” *画像の認識・理解シンポジウム (MIRU)*, Aug. 2017. (査読無)
- (23) 浅見太一, 大谷大和, 岡本拓磨, 小川哲司, 落合翼, 亀岡弘和, 駒谷和範, 高木信二, 高道慎之介, 俵直弘, 南條浩輝, 橋本佳, 福田隆, 増村亮, 松田繁樹, 李晃伸, 渡部晋治, “国際会議 ICASSP2017 報告,” *情報処理学会研究報告*, SLP-2017-117(3), pp.1-8, July 2017. (査読無)

〔学会発表〕（計 1 件）

- (1) Kotaro Kikuchi, **Naohiro Tawara**, Tetsunori Kobayashi, Yoshihiko Hayashi, “Word Vector Augmentation by its Definition for Zero-shot Image Classification,” *CVPR Language and Vision workshop*, July 2017.

6. 研究組織

(1) 研究代表者

俵 直弘 (TAWARA, Naohiro)

早稲田大学・理工学術院・講師（任期付）

研究者番号：50726255