

学位論文審査結果の要旨

専攻主任



博士（工学） 申請者氏名 丹羽 孔明

審査委員

主査 教授 川上 敬

副査 教授 和嶋 雅幸

副査 教授 木下 正博

副査 教授 三田村 保

機械学習のための深層学習による非線形特徴変換

従来、高度な認識や制御といった知的な工学システムを構築するための主要な機構である機械学習システムにおいて、学習器に渡す入力データとして何らかの特徴ベクトルの設計および抽出をあらかじめ行わなければならないという問題があった。本論文ではこの特徴情報ベクトルの設計・抽出を深層学習の仕組みを用いて自動化し、その結果を Actor-Critic 型の強化学習システムへの入力として与えるような、知的な認識および知的な制御を実現可能とするシステムのフレームワークを提案している。さらにこのフレームワークに音響信号情報を入力データとした、いくつかのケースに対して適した詳細モデルを提案し、実験によりその有効性を検証しているものである。すなわち、深層学習による概念学習を行うコンポーネントを非線形特徴変換機構として有し、音響信号より記号情報を介さずに適切なシステムの出力を決定するモデルを構築している。このモデルは音楽や音声・音響に関するデータを記号化せず、時系列に並んだ周波数成分、あるいは信号波形のフレームとして表現された音響信号を概念学習させ、そこから強化学習のメカニズムにより適切な出力を学習する。このモデルを工学システムやロボットが感知する音響信号からの応答や動作決定、および新しい楽曲信号の生成などに応用し、各応用例に適切な多層ニューラルネットワークの詳細モデルを構築し検証を行った結果を考察するもので全6章から構成されている。

第1章は序論であり、今日における機械学習の現状、および、音響信号とロボティクスや情報処理の関係、音響信号を対象とした機械学習についての背景を述べ、本研究の目的および意義を述べている。

第2章では、論文の主要な課題である特徴情報の設計問題を解決できる機械学習システムとして提案する概念モデルについて述べている。具体的には深層学習により音響信号の概念学習を行うコンポーネントを前段に配置し、生の音響信号から記号情報を介さずにシステムの出力を学習するモデルについて、強化学習の枠組みを用いたフレームワークを提示している。このときに、音響信号を学習するコンポーネントが考慮すべき課題として、音響信号の特徴抽出と時間依存性についても詳述している。

第3章では、第2章で提案した概念モデルの詳細モデルの一つとして、時系列データである時間領域の音響信号波形ベースの記憶・復元・予測を行うモデルを深層学習によって構築できることを示してい

る。具体的には制約ボルツマンマシン(RBM)と Conditional RBM の多層ニューラルネットワークを用いて、音響信号の概念学習を行うコンポーネントを構築し、実音楽の音響信号を訓練データとして、時系列に並んだ信号波形のフレームについて、その予測と復元を対象課題としてモデルの検証を行った。学習後のモデルから得られた音響信号データについて、被験者による聴取実験と周波数分析による分析の結果、訓練に用いた信号波形は比較的忠実に再現・予測され、聴取時には訓練データの音楽が忠実に聞き取れることを確認している。実験の結果から本章で提案した詳細モデルは、時系列データである信号波形のフレームから特徴量の自動獲得と時間依存性を考慮した学習が可能であることを示している。

第4章では第3章で検証した詳細モデルの性能をさらに向上させるため、人間を含めた生物が持っている聴覚フィルタバンクによる周波数分析モジュールを深層学習に組み合わせた詳細モデルの構築について示している。具体的には入力となる音響信号をまず聴覚フィルタバンクで周波数分析し、その結果を Auto Encoder とリカレントニューラルネットワークの一種である Echo State Network を用いた多層ニューラルネットワークに受け渡すことで非線形特徴量変換を行うシステムを構築している。本章での対象問題として、IEEE の D-CASE challenge に含まれるオフィスで発生する音響イベントの検出が本システムで可能かについて検証している。この問題は雑音を含む実環境下での多クラス分類課題であり、様々な実用的問題に応用可能なものである。実験の結果、本システムの音響イベント検出性能は challenge に提示されている Baseline モデルの精度を超え、F 値で 30.17% の分類精度を示した。以上により、本章で提案した詳細モデルは、聴覚フィルタバンク分析による時系列な周波数成分から特徴量の自動獲得と時間依存性を考慮した学習が可能であることを示し、雑音環境下においても音響信号の認識が可能であることを示している。

第5章は人間の作曲活動をモデル化した自動作曲システムの構築について述べている。第2章で示した概念モデルにおける Actor-Critic 型の強化学習コンポーネントにより音データを出力するシステムとして、Actor では音楽における音高・音価の組をノード、各ノード間を遷移確率を持った有向のリンクで接続したネットワークとし、ある状態での遷移確率に従ってネットワーク上のノードの音高・音価を順次出力するモデルを提案している。この強化学習コンポーネントを正しく学習するには出力となるパターンの良さを表す報酬というスカラー量が必要となるため、この報酬をシステム自身の出力とモデルからの予測との誤差による報酬と、ドーパミン神経系の働きを模擬する報酬の2モデルを提案し、それぞれ検証を行っている。これらは時系列性を表現するために Elman net 型のリカレントニューラルネットワークを用いてモデルを構築し、予測されたデータは訓練用データとよく合致する結果が得られ、実験の結果から音楽的経験を背景とした予測処理を表現する報酬のモデルが構築されている可能性を示した。またドーパミン神経系の報酬モデルでは、体感に対応する模擬データを定義することで、報酬の関数として機能することを示している。

第6章では本論文の結論として提案した様々なモデルに対する評価実験結果について要約して述べるとともに、本研究で得られた知見を総括している。

以上を要するに、機械学習が様々な領域で高い学習性能を発揮するためには、入力データの適切な特徴量設計・変換が人手により実行されることがこれまでは必須であったが、本論文で提案されたシステム構造を用いることで、より学習に向けた非線形な特徴変換が自律的に行えることを提示し、特に音響信号を入力とする多様な機械学習システムにおいて高い学習性能を持つような詳細構造を明らかにした点で知能情報工学に資するところ大である。

よって筆者は博士(工学)の学位を授与される資格があるものと認める。