

## 論文の内容の要旨

論文題目	深層学習を用いた感情モデルの構築： 感情メカニズムの理解へ向けた構成的アプローチ
学位 申請者	日永田 智絵

感情は人の知能において、非常に重要な要素である。例えば、感情は身体内部の評価と外部刺激の評価に密接な関わりがあり、環境変化に素早く対応するための重要な役割を担っていると言われる。また、人の知能における感情の重要な視点は、意思決定における感情の役割である。さらに、感情の社会的側面も非常に重要である。このように感情は人の知能において切り離すことのできない要素であり、感情のメカニズムが解明されれば、それはすなわち、人の知能の本質的なメカニズム解明に向けた大きな一歩となる。

感情に関して、過去にも多くの研究が行われてきたが、感情全体のメカニズムはまだ完全には明らかになっていない。生理学的研究や神経科学的研究では感情メカニズム解明を目指したモデルや脳部位に関する知見が得られているが、それらは一部分のモデルであったり、各部位に絞ったものが大部分を占めている。各部位に焦点を当てる方法は、その部位を知るために重要であるが、統合的なシステムの中で初めて現れる機能もあると考えられる。そのため、感情の統合的なシステムを構築し、その振る舞いを考察することによって感情メカニズム解明を目指すアプローチも促進されるべきである。

神経科学や心理学では、いくつかの統合的な概念モデルが提案されているものの、それらは大人のデータにより、発達後を切り取って見ている。しかしながら、基本感情にも文化差があることが示唆されているように、感情は初めからできあがっているとは考えにくく、幼児から徐々に感情が発達すると考えるのが自然である。このような考えの一つとして、ブリッジズやルイスなどの感情分化の研究が挙げられる。これらの研究では幼児の振る舞いの観察などにより、カテゴリカルな感情が徐々に創発されていく様子をモデル化している。このようなカテゴリカルな感情は他者から見たときのカテゴリであり、養育者との社会的なやり取りの中で、概念として固定化していくと考えられる。

そこで、本研究では、養育者との社会的なやり取りの中で、感情分化を行うことのできる統合的な感情モデルの構築を目指す。具体的な方法として、既存の統合的

な概念モデルを基に、感情モデルを構築し、それを深層学習で実装した。そのモデルを用いて、養育者とのインタラクションを模したミラーリングタスクを行い、感情分化のシミュレーションを行った。即ち本研究は、感情メカニズムの理解へ向けた構成的アプローチであると捉えることができる。

具体的なモデルは以下のように構成されている。まず、刺激を受け身体的反応を行う生得的な層と、その情報を経験に基づいて抑制・強化する層が存在する。これらの層の出力は内臓などの身体感覚といわれている内受容感覚として知覚される。そして、この内受容感覚と刺激、いわゆる外受容感覚に基づき、未来を予測し行動決定するモデルとなっている。感情はこのモデルのメカニズムの中で創出されるものであるが、行動決定ネットワークの情報が脳皮質に上り、言語と結びつくことによって、怒りや悲しみ、喜びといった状態を感情として認知していると考えられる。このアイデアは感情分化に基づいている。

感情モデルはパートナーロボットの視点からみても、人に共感できるロボットを構築するのに役立つと考えられる。一般に「ロボット」は、心つまり感情の無い存在と考えられている。そうした考えを体現するかのように、ロボット研究では感情について深く扱えずにいた。相手の情動を推定する手法や情動を表出する方法については、従来から研究がなされている。しかしながら、こうした研究の情動は作りこみである。作りこみでは社会的な感情の様な複雑な感情を作ることが困難であり、実際シンプルな基本情動のようなものしか実現できていない。しかし、ロボットが人間社会に受け入れられていくためには、相手の感情を理解・共感し、行動することが必要不可欠であり、シンプルな基本情動のようなものだけでは対応できない。よって感情の計算モデルを構築し、それをロボットに実装することによって、人の心的状態の理解に繋がり、ロボットが人間社会に受け入れられる可能性がある。つまり本研究は、作ることによって感情メカニズムの理解を目指すという構成的アプローチであると同時に、そのモデル自体がパートナーロボットの社会的性能を大きく向上させる可能性を持っている。

提案モデルは、相互に作用する三つのモジュールからなる深層学習を用いて実装した。具体的には、外的刺激を評価する生得的な反応を司るモジュールに Recurrent Attention Model (RAM)、予測を行うモジュールに Convolutional Long Short-Term Memory (LSTM)、行動を強化学習し出力するモジュールに Deep Deterministic Policy Gradient (DDPG) を使い、さらに、平滑化システムや身体内部状態を仮定した関数を全て統合した。本研究では、統合した感情モデルの動作環境として、養育者とのインタラクションを模した表情学習タスクを設計した。このタスクの中で感情モデルがどのような振る舞いをするのかを観察し、考察した。その結果、行動決定のネットワーク中で感情分化を表すような内部状態変化が見られた。

本論文は、次のように構成されている。まず第1章では、本研究の背景について解説し、本研究の目的を述べる。第2章では、感情に関わる研究を紹介する。第3章では感情表出研究の一つとして、著者が行った飛行型ロボットによる情動表現動作を紹介する。第4章では、神経科学及び認知科学の側面から検討した感情モデルを提案し、第5章では深層学習を用いた感情モデルの実装について説明する。第6章で養育者とのインタラクションを模したシミュレーションについて述べ、結果を考察する。最後に第7章で本論文の内容をまとめる。

## 論文審査の結果の要旨

学位申請者氏名	日永田 智絵
審査委員主査	中村 友昭
委員	金子 正秀
委員	横井 浩史
委員	内田 雅文
委員	長井 隆行
委員	
委員	

本研究は、養育者との社会的なやり取りの中で、感情分化を行うことのできる統合的な感情モデルの構築を目指したものである。具体的な方法として、既存の統合的な概念モデルを基に、感情モデルを構築し、それを深層学習で実装している。そのモデルを用いて、養育者とのインタラクションを模したミラーリングタスクを行い、感情分化のシミュレーションを行った。即ち本研究は、感情メカニズムの理解へ向けた構成的アプローチである。

まず第1章では、本研究の背景について概説し、本研究の目的を次のように述べている。感情は人の知能において非常に重要な要素であり、環境変化に素早く対応するための重要な役割を担っている。また、意思決定における役割や、社会的側面も非常に重要である。感情は人の知能において切り離すことのできない要素であり、感情のメカニズムが解明されれば、それはすなわち人の知能の本質的なメカニズム解明に向けた大きな一歩となる。

第2章では、感情に関わる研究を紹介している。ここでのサーベイは、感情研究に関して時代的背景や、最新のモデル研究まで多岐に渡っている。例えば、古典的な基本6感情や、円環モデル、ジェームズ・ランゲ説から、近年のカルテット理論や、人-ロボットインタラクションにおける感情デザインの研究まで網羅している。

次に第3章では、感情表出研究の一つとして、申請者が行った飛行型ロボットによる情動表現動作について議論している。これは、2章で述べた人-ロボットインタラクションの文脈で行われた研究であり、飛行型ロボットを使って情動表現がどのように人に伝わるかわを追及した研究である。この研究によって、どのような動きが、どのような感情として受け取られるかについて、重要な示唆が得られたが、一方でロボットが本質的に感情を持たないことの問題が明らかとなった。

第4章では、神経科学及び認知科学の側面から検討した感情モデルを提案している。ここではまず、感情とは何かという定義を行っている。感情は、自分の身体状態に関する信号（内受容）を処理した結果生じるコア・アフェクトが、さらに環境の情報処理や言語と相互作用し概念化したものである。この感情を形成するシステムを、3層の概念モデルとして提案した。

第5章は深層学習を用いた感情モデルの実装について説明している。これは、第4章で提案した感情の概念的な3層モデルを、実際にシミュレーションできる形で実装したものである。まず、第一層は自身の身体状態の評価である内部評価と外からの刺激に対する評価である外部評価の両方に基づいた内受容感覚を生成する。第二層は、実環境に適応し、第一層の生得的反応を修正するシステムである。この記憶ベースの学習は、エージェントが過去に経験し蓄積した情報を用いることによって予測の精度を高める。第三層は、予測と意思決定を行う層であり、Convolutional-LSTM-DDPGを用いて実装している。報酬は、ホメオスタシスの考え方に基づいているが、内受容を一定値に安定化させるのではなく、前の時間の平均的な値を設定値とすることで、ダイナミックに値を安定化させるという考え方を用いている。これは、アロスタシスの考え方に近い。

第6章では養育者とのインタラクションを模したシミュレーションについて述べ、結果を考察している。実験は大まかに4つに分かれている。一つ目の実験では、ソマティック・マーカー仮説の検証として、身体反応のシグナルが本当に意思決定を効率化するかの検証を行った。具体的には、簡単な狩猟タスクにおいて、単純な決まったシグナルを出す第一層を設計し、第三層のConvolutional LSTM DDPGを組み合わせ、第一層の情報があることによって、DDPGがより高い報酬を得るように学習ができるかの検証を行った。これは、三層構造の妥当性を検証する実験である。二つ目の実験では、RAM（第一層）の性能を検証した。第一層は刺激に対して人のような生得的反応を返すと仮定しているので、教師ラベルに対する誤差だけでなく、著者の主観的感覚と子供の傾向を通じて定性的に評価した。三つ目と四つ目の実験は、本研究で主眼の置いている、養育者との社会的なやり取りの中での感情分化をシミュレーションする実験である。三つ目の実験では、RAM（第一層）とConvolutional LSTM DDPG（第三層）を組み合わせ、エージェントの行動とネットワークの内部表現を観察した。第二層はシステムの環境への適応を担当するため、第二実験では第一層と第三層の実装に焦点を当てた。四つ目の実験では、第一層、第二層、第三層を組み合わせ、感情モデル全体を実装し、その動作を検証した。この実験を第二実験と比較することにより、第二層の有意性を調べることが可能となっている。こうした実験の結果として、提案した三層モデルが従来から観察されているような感情分化を再現できることを示した。

最後に第7章で、論文の内容をまとめている。本論文は、人の感情を計算モデルとして実装することで、そのメカニズムに大きく迫った。また、同時に今後の課題について、社会的な高次感情のモデル化や実装の必要性を挙げて説明している。人と人工物が本当の意味で共感しあうためには、今後こうした方向性を追及していくことが重要である。

以上のように本論文は、人の感情メカニズム解明やロボットの知能化に大きく貢献するものである。よって本論文は、博士（工学）の学位論文として十分な価値を有するものと認める。