Restricted Boltzmann Machineを用いた多感覚情動コミュニケーション

Multimodal Emotion Communication by using Restricted Boltzmann Machines

堀井隆斗　長井志江　浅田稔
Takato Horii　Yukie Nagai　Minoru Asada

大阪大学大学院工学研究科
Osaka University Graduate School of Engineering

It is assumed that emotion recognition and generation systems share the same internal model. People interpret emotional expressions of a partner based on their internal emotion model, whereas the model enables people to express their emotional states. Despite the close link between generation and recognition of emotions, there has been little computational models to take the link into account. This paper presents a multimodal Restricted Boltzmann machines (RBMs), which is able to both generate and recognize emotional states. By hierarchically integrating RBMs for multiple sensory modalities (e.g., vision, audio, and tactile), our model represents emotional states as the activations of the units at the higher layer. Our experiments demonstrate that the model can generate emotional expressions similar to those presented by an interaction partner like mirroring. Additionally, the model can better infer the paters' emotion from deficient modality inputs through an imaginary reconstruction of its own emotional expression. We discuss the advantage of our emotion recognition-generation model in relation to the mirror neuron system.

1. はじめに

情動は意思決定やコミュニケーションなど，人の認知機能や他者との相互作用において様々な影響を与える．特にコミュニケーションにおいては，他者の情動状態を表情や発話情報から推定し，その情動状態に応答する形で表情や身体運動を用いて多感覚的に自身の情動状態を表出することが重要になる．近年では，人工物にこれらの能力を付加する研究が盛んであり，人-ロボット相互作用の分野においても人の情動状態の推定やロボットの情動表出に関する研究が数多く行われている [1,2].

Kishi et al. [1] は，快度，活動度と強度の 3 つの特徴量によって構成されるメンタルモデルを人型ロボットに適用することで，ロボットが環境情報に応じて情動表出を調整可能であることを実験により示した．用いられたメンタルモデルは外界からの刺激に応じて各特徴量の値を増減させることで，ロボットの情動状態を連続的な情報として表現する．しかし彼らの研究ではメンタルモデルは連続的な情動変化を表現可能なものの，表情表出は予め作りこまれた 6 つの表情セットから選択されるため，認識した情動状態と情動表出の関係性を柔軟に対応させることができず，情動状態の強度や混合を表現することはできなかった．またその他の研究においても，情動表出と認識を同一のモデルで扱った研究は少なく，それぞれのシステムが独立に開発されていた．

心理学の分野においては，情動状態や表情の連続性を表現するモデルは古くから提案されている [3–5]．Russell [3, 4] は，人の情動状態や表情の関係性が快-不快軸と覚醒-睡眠軸を特徴量とする 1 つの低次元空間において表現されることを示している．また他者の行動理解や状態推定にはミラーニューロンシステムやメンタライジングシステムと呼ばれる脳機能が重要な役割を果たしていることが知られている [6]．ミラーニューロンシステムやメンタライジングシステムは，自己の経験に基づく感覚-運動表現を用いて他者の行為やそれに伴う他者の内的状態の変化を推定している．このような脳内での情報表現や，同一特徴空間上における情動認識と表出の連続的な関係性をさまざまな感覚情報に対して獲得することが，人-ロボット相互作用においても重要であると考えられる．

そこで本研究では人の多感覚情動信号を利用することで情動状態間の関係を学習し，情動状態に対応する多感覚信号を生成することのできるシステムを提案する．本稿では確率的ニューラルネットワークの一種である restricted Boltzmann machine(RBM) [7] を用いて人の情動表出を学習した際の情動表出の模倣能力と，多感覚入力信号の一部が観測できない場合の状態推定と未観測情報の補完能力について，人-ロボット相互作用場面を想定した計算機実験を行ったので，その結果を報告する．

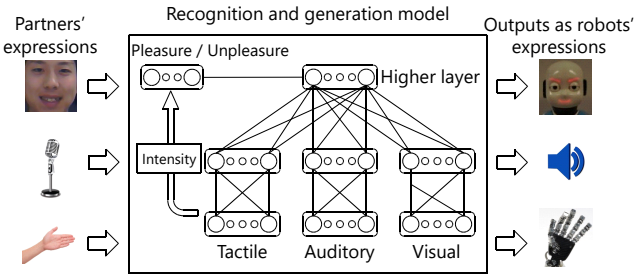

図 1: 提案モデルと想定する相互作用実験

2. 提案モデル

図 1 に提案モデルと実験設定の概略を示す．本稿では対面形式の相互作用を想定し，視覚，聴覚，触覚の 3 つの感覚入力が得られる状況を想定する．提案モデルは各感覚入力信号を抽象化する感覚モジュール (図 1 提案モデル中の下部) と，それら多感覚信号を統合するための統合モジュール (図 1 提案モデ

連絡先: 堀井隆斗，大阪大学工学研究科，〒 565-0871 大阪府吹田市山田丘 2-1 U1W-618，06-6879-4724，takato.horii@ams.eng.osaka-u.ac.jp

ル中の上部) からなり，それぞれのモジュールは RBM [7] によって構成される．感覚モジュールでは連続値のセンサ信号を扱うために，Cho et al. [8] によって提案された可視層ノードの分散が学習可能な Gaussian-Bernoulli RBM を利用する．視覚，聴覚，触覚の各感覚モジュールは対応する感覚様式の信号を独立に学習する．また，それぞれの感覚モジュールの出力信号は統合モジュールの入力値として扱われ，統合モジュールの隠れ層の表現を学習するために利用される [9]．統合モジュールにおける隠れ層の発火確率は，次式の条件付確率を用いて表現される．

$$p(h_s = 1|\boldsymbol{h^v}, \boldsymbol{h^a}, \boldsymbol{h^t}) = \sigma\Big(\sum_i h_i^v w_{is} + \sum_j h_j^a w_{js} + \sum_k h_k^t w_{ks} + b_s\Big) \quad (1)$$

ここで，$\boldsymbol{h^v}, \boldsymbol{h^a}, \boldsymbol{h^t}$ はそれぞれ視覚，聴覚，触覚の感覚モジュールの出力，w は各ノード間の結合加重，b_s は統合モジュールの隠れ層ノード h_s におけるバイアスを表す．提案モデルは各感覚モジュールと統合モジュールの学習を行うことによって，統合モジュールの隠れ層ノードの活動として，他感覚入力信号からそこに内在する情動情報を抽出することができる．

また本モデルは RBM が生成モデルであることを用いて，統合モジュールの隠れ層の信号から各感覚モジュールの可視層の信号を生成可能である．さらにこの特徴を利用して，入力信号が欠損している感覚モジュールにおいても，他の感覚信号から連想される統合モジュールの出力からその信号を再構成することが可能となる．例として視覚感覚モジュールの入力信号が欠損している場合を想定すると，聴覚と触覚の信号を用いて以降の一連のサンプリングを行うことで対応する視覚入力信号の生成が可能である．

$$h_i^v \sim p(h_i^v = 1|\boldsymbol{0}) \quad (2)$$

$$h_j^a \sim p(h_j^a = 1|\boldsymbol{v^a}) \quad (3)$$

$$h_k^t \sim p(h_k^t = 1|\boldsymbol{v^t}) \quad (4)$$

$$h_s \sim p(h_s = 1|\boldsymbol{h^v}, \boldsymbol{h^a}, \boldsymbol{h^t}) \quad (5)$$

$$\hat{h_i^v} \sim p(\hat{h_i^v} = 1|\boldsymbol{h}) \quad (6)$$

$$\hat{\boldsymbol{v^v}} \sim p(\hat{\boldsymbol{v^v}} = \boldsymbol{v}|\hat{\boldsymbol{h^v}}) \quad (7)$$

ここで $\boldsymbol{v^a}, \boldsymbol{v^t}$ は聴覚と触覚の入力信号，$\hat{\boldsymbol{v^v}}$ と $\hat{\boldsymbol{h^v}}$ はサンプリングによって生成された視覚の感覚モジュール信号である．式 (2) - (4) と式 (5) は，それぞれ感覚モジュールと統合モジュールの RBM における順方向のサンプリングを，式 (6) と式 (7) は逆方向のサンプリングを表す．提案モデルはこのようにして多感覚信号からの情動認識と，情動状態に対応する多感覚信号の生成を行う．

3. 実験

3.1 実験設定

提案モデルの有用性を確認するために，人の多感覚情動信号を用いてモデルの学習と，人-ロボット相互作用場面を想定した 2 つの計算機実験を行った．計算機実験では相互作用場面を想定し，ロボットシステムを想定したカメラとマイク，触覚センサに対して人が多感覚情動信号を表出した際のデータを収集し利用する．提案モデルの学習と評価実験には，文献 [10] と同様のデータセットを用いた．このデータセットでは，基本 6 情動 (喜び，驚き，怒り，嫌悪，悲しみ，恐れ) と中性の状態時に人間が生成した顔表情，発話，接触が，視覚，聴覚，触覚の 3 つの感覚様式の信号として記録されている．視覚情報では 900 次元のグレースケール画像を主成分分析し，固有値の大きなものから選択した 20 個の主成分を特徴量として扱う．聴覚は発話音声データを 10 区間に分節化したものから強度と基本周波数を算出し特徴量とする．触覚特徴量は，データセットに含まれる触覚センサ信号から接触時間，接触領域，最大接触力とその接触時間，変位速度，周波数強度 ($1-60$, $61-100$, $101-200$Hz の 3 種類) とセンサ信号の振動回数である．

表 1: 未学習入力と再構成データ間の平均二乗誤差

実験条件	視覚信号	聴覚信号	触覚信号
実験 1	0.115	0.080	0.150
実験 2(視覚欠損)	0.119	0.082	0.148
実験 2(聴覚欠損)	0.118	0.083	0.150
実験 2(触覚欠損)	0.118	0.080	0.168

評価実験では，まず未学習データを与えた際の統合モジュールの出力信号をサンプリングする．そしてサンプリングされた出力信号から逆方向サンプリングを行うことによって，各感覚モジュールの信号を再構成し，実際の入力信号との比較を行う．実験 1 では，3 つの感覚様式全てのセンサ信号を用いて感覚モジュールの入力信号を再構成した．これは人-ロボット相互作用実験に応用した場合に，ロボットが人の多感覚情動表出から情動状態を推定し，それに対応した自己の情動表出運動を生成する場面を想定した計算機実験となっている．実験 2 では 3 つの感覚様式のうち 1 つを欠損させ，残り 2 つの感覚信号のみを用いて統合モジュールの出力信号をサンプリングし，そこからすべての感覚モジュールの入力信号を再構成した．これは相互作用中において，ロボットが相手の顔でなく他の物体を見ていたために表情を観測できなかった場合など，人の情動表出信号を全て観測できない状況を想定した計算機実験である．各実験には共通の学習済みモデルを用いており，モデルの各ノード数は可視層 − 隠れ層の順に，視覚の感覚モジュールが $20-10$，聴覚の感覚モジュールが $20-10$，触覚感覚モジュールが $9-10$，統合モジュールが $30-20$ である．また感覚モジュールにおいて情動関連特徴量を抽出するために，fine-tuning [7] 用のノードとして触覚信号の強度から快-不快を表す 2 ノードを統合モジュール上部に付加している．

3.2 実験 1:多感覚信号の学習と再構成

提案モデルに 70 種類の未学習データ (各情動状態につき 10 種類) を入力し，サンプリングした統合モジュールの出力信号を利用して再構成した各感覚信号と入力信号との平均二乗誤差を表 1 に示す．それぞれの値は対応する感覚モジュールのノード平均である．入力信号の値は各感覚モジュールのノード毎に $-1 \sim 1$ の範囲以内に収まるように正規化されており，再構成時の誤差は絶対的に小さいといえる．図 2 に実際に入力した未学習データと，再構成した視覚信号の例を示す．この図より，提案モデルが入力された感覚信号と同様の情動状態に属すると考えられるの感覚信号を，多感覚統合後の抽象情報である統合モジュールの出力信号から，逆方向サンプリングを通じて再構成可能であることがわかる．同様に，他の未学習データにおいても対応する情動状態であると推定される感覚信号が生成されていることから，統合モジュールの隠れ層に，抽象的な情報として情動カテゴリが獲得されていると考えられる．学習データに対する統合モジュールの出力信号を，主成分分析

(a) 例 1:怒り情動の表出信号　(b) 例 2:喜び情動の表出信号

図 2: 未学習入力 (左) と再構成された視覚信号 (左) の例

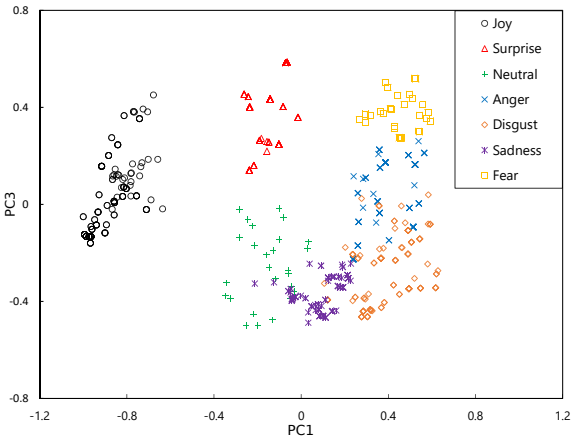

図 3: 統合モジュール出力の第 1-3 主成分空間

(a) ステップ:1　(b) ステップ:100

図 4: 実験 2 において再構成された視覚信号の例

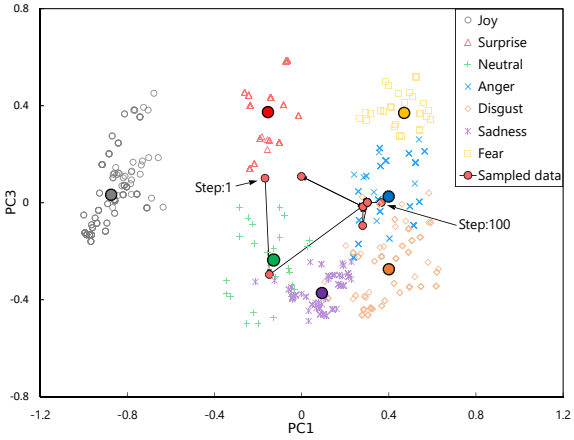

図 5: 10 ステップ毎における統合モジュールの出力の主成分空間上の変化

した結果を図 3 に示す．提案モデルが 7 つの情動状態 (喜び，驚き，怒り，嫌悪，悲しみ，恐れ，中性）のラベルを利用した教師あり学習を行っていないにもかかわらず，統合モジュール出力の主成分空間では情動状態に対応するクラスタが構成されている．以上の結果より，提案モデルは統合モジュールにおいて，入力された多感覚信号の関係を獲得し，さらにその情報を利用することで自ら情動信号を生成可能であることが示された．

3.3　実験 2:欠損感覚信号の再構成

同様の提案モデルに，感覚様式に欠損のある未学習データを入力した際の入力データと再構成データの平均二乗誤差を評価した．利用する未学習データは実験 1 と同じものである．欠損感覚様式は 3 つの中から 1 つが選択され，選択された感覚様式の信号がすべて 0 として扱われる．表 1 にそれぞれの感覚様式を欠損させた際の平均二乗誤差を示す．実験 1 の結果と比較して，入力が欠損している感覚様式の再構成誤差が大きくなっているが，各ノードがとり得る出力の範囲を考慮すると大きな問題でないことがわかる．

次に欠損信号を再構成された信号で置き換え，非欠損感覚信号は未学習データを再度利用することで順方向と逆方向のサンプリングを繰り返す実験を行った．実験における繰り返し回数は 100 回とし，サンプルの採択確率は 1 とした．図 4 に視覚信号を欠損信号としたときの繰り返しサンプリングにおいて，1 回目と 100 回目の再構成信号を示す．なお入力データにおける欠損前の視覚信号は図 2(a) の左側である．また図 5 に，図 3 の主成分空間上に繰り返しサンプリング時の統合モジュール出力の遷移を，10 ステップ毎に描画したものを示す．図中の大きな円は各情動に対応した出力の重心を表している．また統合モジュールの出力と怒り情動クラスタの重心との距離を 10 ステップ毎に算出したものを図 6 に示す．再構成初期においては，提案モデルは聴覚，触覚のみの信号を驚きに近い情動状態に近い信号であると捉えているが，サンプリングを繰り返すことによって次第に怒り情動の分布重心へ推定値が近づいていく様子が見れる．以上の結果より，提案モデルは入力信号の一部が欠損している場合においても，順方向と逆方向のサンプリングを繰り返すことによって入力信号の情動推定結果を更新し，欠損なし信号の再構成と同程度の誤差で欠損信号を生成可能であることが示された．

また結果に見られるような繰り返しサンプリングによるモデル内の状態遷移は，確率的にネットワークの状態を選択する RBM の特徴によるものであり，未観測情報を含むようなデータから状態推定を行う際に利点となる．再構成された視覚信号も，統合モジュールで推定された情動状態の遷移に対応するように信号が順次変化し，次の状態推定に影響を与える．人におけるこのような未観測情報の内部更新を含む逐次的な状態の推定は，ミラーシステムやメンタライジングシステムにて行われていると考えられている [6]．人対人の情動的相互作用においても，観測されなかった他者の情動表出は自身の内部モデルに基づいて推定することが知られており [11]，提案モデルの能力は人-ロボット相互作用実験に応用した場合でも有用であると考えられる．

4.　おわりに

本研究では，確率的ニューラルネットワークの一種である restricted Boltzmann machine を複数用いて感覚信号を抽象化と統合するモデルを提案し，人の多感覚情動信号から情動状態間の関係の学習と，人の情動状態に対応する多感覚信号の生成能力について検証した．検証実験を，提案モデルに視覚，聴

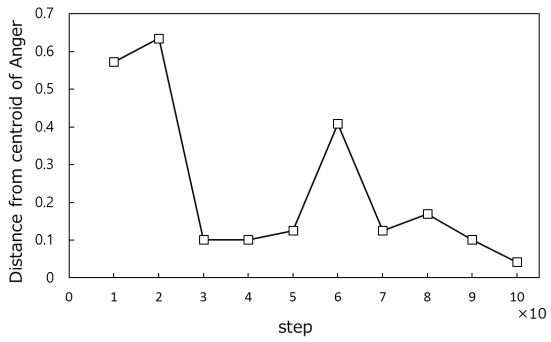

図 6: 10 ステップ毎における統合モジュールの出力と怒り情動クラスタ重心との距離変化

覚，触覚の 3 つの感覚様式から構成される人の多感覚信号を用いて，2 つの条件の下で行った．実験 1 では，人-ロボット相互作用において他者の情動状態を推定し，それに対応し自己の情動表出運動を生成する状況を想定し，未学習データを入力した際の統合モジュールの出力と再構成された各感覚信号を評価した．実験 1 の結果より，提案モデルは多感覚情動信号から情動状態の関係を学習し，未学習データに対しても順方向サンプリングによる情動推定とその情報を利用した逆方向サンプリングによって，入力と同様の情動信号を生成できることが示された．

実験 2 では他者の多感覚情動表出の一部が観測できない状況を想定し，3 つの感覚様式のうち 1 つを欠損させた状況において実験 1 と同様の実験を行った．また順方向と逆方向のサンプリングを繰り返し行うことによって，情動推定と再構成された欠損信号の変化について検証を行った．実験 2 の結果より，1 度のサンプリングでは情動推定と欠損感覚信号生成において誤差が発生するものの，サンプリングを複数回繰り返すことによって欠損を含むデータに対しても提案モデルが頑健に感覚信号を生成できることが示された．これらの結果は人における，自己の脳内表現を用いて他者の行為理解や状態推定やそれに基づく運動生成を行うミラーニューロンシステムやメンタライジングシステムと対応しており，このような能力が人-ロボット相互作用における情動コミュニケーションにおいても有用であることが示された．

今後の課題として，本モデルを実際にロボットに適用することで，実相互作用実験を行うことが必要である．また識別が容易な基本 6 情動だけでなく，より曖昧な情動表出や状態を獲得するために，より大規模な情動表出データセットを利用することも課題である．さらには長期的な実ロボットとの相互作用を通じて内部表象や表出信号がどのように変遷するかについて解析を行いたい．

謝辞

本研究の遂行にあたり，JSPS 科研費 (課題番号 15J00671, 24000012，24119003) の補助を受けた．

参考文献

[1] Tatsuhiro et al. Kishi. Impression survey of the emotion expression humanoid robot with mental model based dynamic emotions. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pp. 1663–1668, 2013.

[2] Angelica Lim and Hiroshi G Okuno. Developing robot emotions through interaction with caregivers. In *IGI Global*, 2014.

[3] J.A. Russell. A circumplex model of affect. *Journal of personality and social psychology*, Vol. 39, No. 6, p. 1161, 1980.

[4] J.A. Russell. Core affect and the psychological construction of emotion. *Psychological review*, Vol. 110, No. 1, p. 145, 2003.

[5] P. Shaver, J. Schwartz, D. Kirson, and C. O'connor. Emotion knowledge: further exploration of a prototype approach. *Journal of personality and social psychology*, Vol. 52, No. 6, p. 1061, 1987.

[6] Frank Van Overwalle and Kris Baetens. Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, Vol. 48, No. 3, pp. 564–584, 2009.

[7] Geoffrey Hinton. A practical guide to training restricted boltzmann machines. *Momentum*, Vol. 9, No. 1, p. 926, 2010.

[8] KyungHyun Cho, Alexander Ilin, and Tapani Raiko. Improved learning of gaussian-bernoulli restricted boltzmann machines. In *Artificial Neural Networks and Machine Learning–ICANN 2011*, pp. 10–17. 2011.

[9] Nitish Srivastava and Ruslan Salakhutdinov. Learning representations for multimodal data with deep belief nets. In *International Conference on Machine Learning Workshop*, 2012.

[10] Takato Horii, Yukie Nagai, and Minoru Asada. Touch and emotion: Modeling of developmental differentiation of emotion lead by tactile dominance. In *Proc. of IEEE Third Joint International Conference on Development and Learning and Epigenetic Robotics*, pp. 1–6, 2013.

[11] Jojanneke ACJ Bastiaansen, Marc Thioux, and Christian Keysers. Evidence for mirror systems in emotions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, Vol. 364, No. 1528, pp. 2391–2404, 2009.