

制限ボルツマンマシンによる 自由エネルギー最小化に基づく能動的知覚

Active perception based on free-energy minimization on restricted Boltzmann machines

堀井隆斗 *1
Takato Horii

長井隆行 *1*2
Takayuki Nagai

*1電気通信大学

The University of Electro-Communications

*2大阪大学

Osaka University

A robot, which would like to respond quickly in the world, should select more informative signals to estimate the cause of its sensation (e.g., a state of the environment, a category of a handling object, an emotional state of interaction partner, etc.). This paper proposes an active perception framework that selects the robot's action to perceive critical sensory signals based on a free-energy minimization in an energy-based model. We employed a restricted Boltzmann machine as a fundamental component for an estimation network of the cause of sensations. Our framework demonstrated better performance for the attention control in emotional human-robot interaction than other methods.

1. はじめに

多様な情報にあふれる実世界で活動するロボットにとって、環境や自身の状態を認識するために複数の感覚情報を処理することが重要である。Nakamura と Nagai [Nakamura 17] は、質感や色、聴覚、触覚、単語の感覚様式に基づいて、ロボットが多感覚情報を処理することで知覚物体の関係性を構造化する手法を提案している。また Horii ら [Horii 18] は、視覚、聴覚、触覚の感覚様式に注目し、それらの感覚情報を統合することで感情発達を再現する計算論モデルを提案している。このように複数の感覚情報を統合・処理することで物体のカテゴリや感情といった概念構造を獲得することができる。そしてこのような概念構造から環境や自身の状態を推定することで、他者との相互作用が可能となる。

概念構造を獲得したロボットは、一部の感覚情報からでも概念情報を予測することができる。しかし、本来複数の感覚情報を処理することで獲得された概念情報を、一部の感覚情報から予測するとは難しい。例えば不透明なペットボトルが見えた時にそれがゴミであるか否か、電話先の友人が大きな声を発した時に喜んでいのか怒っているのかを判断するためには、追加の感覚情報が必要であることは想像にたやすい。このような場面で概念情報（ここでは物体のカテゴリや他者の感情状態）の推定に効果的な感覚情報を選択・知覚することを能動的知覚と呼ぶ。Taniguchi ら [Taniguchi 18] は、ロボットのための能動的知覚として、現在推定している概念情報と予測される未観測感覚情報との関係性を表す相互情報量を推定し、その相互情報量が最大となる感覚情報を選択する手法を提案している。

一方で、近年の神経科学研究において、人の能動的知覚（推論）を説明する基本原理としての自由エネルギー原理 [Friston 10] が注目されている。これは人の脳が入力される感覚信号を予測する内部モデルを構成し、その予測と実際の信号間の誤差を最小化するように認識や行動を実行するというものである。本研究ではこの自由エネルギー原理に基づいた能動的知覚のための計算論モデルと手法を提案する。具体的には深層学習モデルの1つである Restricted Boltzmann Machine (RBM) [Hinton 10] を用いた多感覚情報統合モデルにおいて、

連絡先: 堀井隆斗, 電気通信大学, 東京都調布ヶ丘 1-5-1,
takato@uec.ac.jp

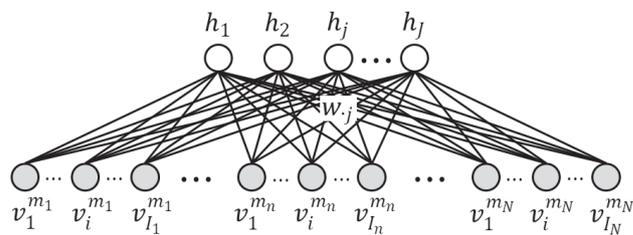


図 1: RBM による多感覚統合モデル

観測情報から概念情報を介して未観測情報を予測し、その中で自由エネルギーが最も小さい（つまり最も予測誤差が小さい）感覚情報を選択する。本稿では提案する能動知覚手法と Taniguchi ら [Taniguchi 18] の手法を、人とロボットの感情コミュニケーション場面を想定した場面において評価した結果について示す。

2. 提案手法

RBM [Hinton 10] による多感覚統合モデルと提案する自由エネルギー最小化に基づく能動的知覚手法について説明する。図 1 に RBM を用いた多感覚統合モデルを示す。ここで $v_i^{m_n} \in \{0, 1\}$ は n 番目の感覚様式 m_n の i 番目の可視層ノードを、 $h_j \in \{0, 1\}$ は j 番目の隠れ層ノードを示す。また $w_{.j}$ は h_j と可視層ノードの結合加重である。このモデルは、学習時に m_N までのすべての感覚様式情報を入力として受け取り、入力情報と隠れ層 \mathbf{h} を介して再構成された情報の誤差を最小化することで、確率分布 $p(\mathbf{v} = \{v^{m_1}, \dots, v^{m_N}\})$ を推定する。また学習によって隠れ層に何らかの特徴表現（例えば物体カテゴリや感情カテゴリなど）が獲得されることが期待される。

学習によって推定される周辺確率 $p(\mathbf{v})$ は次式で表現される。

$$\begin{aligned} p(\mathbf{v}) &= \sum_{\mathbf{h}} p(\mathbf{v}, \mathbf{h}) \\ &= \exp\left(\mathbf{c}^T \mathbf{v} + \sum_{j=1}^J \log(1 + \exp(b_j + \mathbf{W}_{.j} \mathbf{v}))\right) \quad (1) \\ &= \exp(-F(\mathbf{v})) / Z \end{aligned}$$

Algorithm 1 Active perception based on free-energy minimization on RBM

```

add perceived modality  $m_n$  to set  $M_r$ 
for  $n = 0$  to  $N$  do
  if  $m_n \in M_r$  then
     $\mathbf{v}^{m_n} = \mathbf{v}^{m_n}$ 
  else
     $\mathbf{v}^{m_n} = \mathbf{0}$ 
  end if
   $\mathbf{h} \sim p(\mathbf{h}|\mathbf{v}^{m_1}, \dots, \mathbf{v}^{m_n}, \dots, \mathbf{v}^{m_N})$ 
  for  $n = 1$  to  $N$  do
    if  $m_n \notin M_r$  then
       $\mathbf{v}^{m_n'} \sim p(\mathbf{v}^{m_n'}|\mathbf{h})$ 
       $E_n \leftarrow F(\mathbf{v}^{m_n'})$ 
    end if
  end for
end for
 $n = \operatorname{argmin} E_n$ 
 $M_r \leftarrow M_r \cup m_n$ 

```

$F(\mathbf{v})$ は RBM の自由エネルギーである。以上より、ある情報 \mathbf{v} の自由エネルギーが小さいことは、 \mathbf{v} の発生確率 $p(\mathbf{v})$ が高いことと対応する。

次にある部分観測情報 $\mathbf{v}' = \{\mathbf{0}, \dots, \mathbf{v}^{m_n}, \dots, \mathbf{0}\}$ から未観測情報を予測することを考える。それぞれの未観測情報 \mathbf{v}^{m_u} は部分観測情報 \mathbf{v}' から推定された \mathbf{h} を介して再構成できることから、次のように予測される。

$$\begin{aligned} \mathbf{h} &\sim p(\mathbf{h}|\mathbf{v}') \\ \mathbf{v}^{m_u} &\sim p(\mathbf{v}^{m_u}|\mathbf{h}) \end{aligned} \quad (2)$$

ここで、提案手法は予測された未観測情報と部分観測情報から計算される自由エネルギーが最も小さくなるものを次の知覚情報として選択する。すなわち RBM における自由エネルギー最小化に基づく能動的知覚は、現在の観測情報から推定される概念情報に対して尤もらしく予測される感覚様式の情報を選択することとなる。Algorithm1 に提案手法の詳細を示す。

3. 実験

提案手法の有用性を評価するために、人とロボットの感情コミュニケーション場面を想定した感情推定のための能動的知覚実験を行った。実験では人の対面多感覚感情コミュニケーションデータである IEMOCAP データセット [Busso 08] から表情、音声、手の動きに関する特徴量を抽出し、それぞれを感覚様式の情報として図 1 の多感覚統合モデルを学習した。学習の詳細な条件に関しては [Horii 16] を参照されたい。また今回の実験では評価値として、部分観測情報とすべての観測情報から推定された隠れ層の発火確率の KL ダイバージェンスを利用した。

図 2 に単一の感覚様式情報を部分観測情報として入力した状態 (baseline) とそれぞれ異なる能動的知覚手法で選択した 1 つの感覚様式情報を付加した状態の評価値を示す。能動的知覚を実行することで状態推定の精度が向上していることが分かる。また無作為抽出 (Random) と Taniguchi ら [Taniguchi 18] の手法 (IG.max) と比較して、提案手法 (FE.min) が高い性能を示すことが確認された。

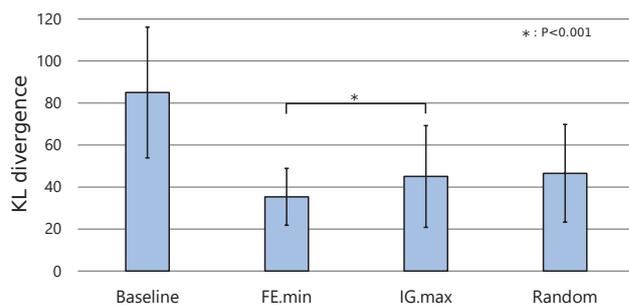


図 2: 各能動的知覚手法による感情状態推定に対する評価値

4. おわりに

人の脳情報処理の基本原則だと考えられている自由エネルギー原理に基づいた能動的知覚手法を提案した。そして、人とロボットの感情コミュニケーションを想定した能動的知覚の実験において情報量を最大化する Taniguchi ら [Taniguchi 18] の手法と情動推定精度を比較し、提案手法が高い性能を発揮することを示した。

今後は提案手法と既存手法の関係性を明らかにするとともに、RBM 以外のエネルギーモデルへの適用を進める。また、ロボットの能動的知覚による状態変化の影響を踏まえた時間方向へ拡張する。

謝辞

本研究は JST CREST(JPMJCR15E3, JPMJCR16E2) の支援を受けた。また研究内容に関して議論した情報通信研究機構長井志江主任研究員に感謝する。

参考文献

- [Busso 08] Busso, C., Bulut, M., Lee, C.-C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J. N., Lee, S., and Narayanan, S. S.: IEMOCAP: Interactive emotional dyadic motion capture database, *Language resources and evaluation*, Vol. 42, No. 4, pp. 335–359 (2008)
- [Friston 10] Friston, K.: The free-energy principle: a unified brain theory?, *Nature reviews neuroscience*, Vol. 11, No. 2, p. 127 (2010)
- [Hinton 10] Hinton, G. E.: A practical guide to training restricted boltzmann machines, Technical report, Department of Computer Science University of Toronto (2010)
- [Horii 16] Horii, T., Nagai, Y., and Asada, M.: Imitation of human expressions based on emotion estimation by mental simulation, *Paladyn, Journal of Behavioral Robotics*, Vol. 7, No. 1 (2016)
- [Horii 18] Horii, T., Nagai, Y., and Asada, M.: Modeling development of multimodal emotion perception guided by tactile dominance and perceptual improvement, *IEEE Transactions on Cognitive and Developmental Systems* (2018)
- [Nakamura 17] Nakamura, T. and Nagai, T.: Ensemble-of-Concept Models for Unsupervised Formation of Multiple Categories, *IEEE Transactions on Cognitive and Developmental Systems* (2017)
- [Taniguchi 18] Taniguchi, T., Yoshino, R., and Takano, T.: Multimodal Hierarchical Dirichlet Process-Based Active Perception by a Robot, *Frontiers in neurorobotics*, Vol. 12, (2018)