原著論文

パートナーロボットの環境知覚に基づく 発話内容の学習†

久保田 直行*・矢口 愛子*

国内における少子高齢社会の到来により、高齢者や子どもの見守りのため、家庭用パートナーロボットの開発への期待が高まってきている。しかし現状のロボットはシナリオに依存することが多く、自然なコミュニケーションを実現できてはいない。本研究の目的は、環境と発話内容の関係を学習することで、人間とロボットの間での自然なコミュニケーションを実現することである。本論文ではロボットが環境と発話内容の関係を学習するためにスパイキングニューラルネットワークを用いた手法を提案した。提案手法を用い、コミュニケーションを通じて、ロボットが日用品の用途を学習する実験と、時間と発話の関係を学習する実験を行った。実験の結果、ロボットが人間との相互作用によって認知環境に基づいた発話内容の更新を行えるようになる、ということを示した。

キーワード：パートナーロボット、環境知覚、自然なコミュニケーション、学習

1. はじめに

国内における認知症患者は主に65歳以上の高齢者であり、その数は2025年には323万人にのぼると推測されている[1]。認知症の改善に努めることは現在非常に意義を持つといえる。現在国内において認知症の基準は介護保険法により「脳血管疾患、アルツハイマー病その他の要因に基づく脳の器質的な変化により日常生活に支障が生じる程度にまで記憶機能及び認知機能が低下した状態」と定義されている。また近年診断基準として注目されつつあるDSM-IV[2]に基づく、記憶障害および失行・失認・失語・実行機能障害のうちの一つが認められ、かつ社会生活に支障をきたし、脳や全身的な原因疾患をもち、意識障害を伴わない場合を認知症と認めるとされている。すなわち認知症の進行抑制・改善にあって我々が目指すべきところは、患者の身体的な認知機能の回復に加え、その社会的な回復も含まなくてはならないといえる。そして認知症という用語で一括りされている症状の違いを把握し、一人一人の状態にあわせてリハビリテーションを行うことが重要であると考えられる[3]。しかし現在、国内では少子高齢社会の到来により、若年人口の減少から介護者不足が深刻な問題となっている。また、高齢者の問題だけではなく、共働きの家庭における子ども

の見守りへのニーズも共に高まっており、家庭内における保護者の代替となる存在が求められている。そこで近年、技術の進展に伴い、パートナーロボットへの期待が高まってきた。

しかしここで、従来のパートナーロボットの大きな問題点として、事前設計による行動決定や、シナリオベースの会話しか行うことができないという点があげられている。その問題の解消のために現在では、学習機能を備えたロボットの開発が行われるようになってきた。しかし、ロボットの行動学習にその主眼は置かれてきたため、コミュニケーション学習についてはほとんど議論がなされてこなかった。そこで本研究ではこの点に着目し、パートナーロボットのコミュニケーション学習を行うことを提案する。パートナーロボットがコミュニケーション学習を行うと、コミュニケーション相手の生活に沿った学習が行われるため、ユーザの習慣を学習することが可能となる。人間同士のコミュニケーションの場合、互いの習慣というものを背景知識として共有することにより、相手に合わせた円滑なやり取りを行うことが期待できる。そのことは初対面の相手よりも馴染みの相手の方が円滑にコミュニケーションできることからわかる。

本研究では、人間とパートナーロボットの自然なコミュニケーションの実現を目的とする。本論文では、パートナーロボットがコミュニケーション学習を通じて人間の習慣を習得することで、従来のシナリオベースのコミュニケーションでは実現できなかった、柔軟な発話選択を行えるようになることを示す。

† Learning of Utterance Contents for A Partner Robot Based on the Environmental Perception

Naoyuki KUBOTA and Aiko YAGUCHI

* 首都大学東京システムデザイン研究科
Graduate School of System Design Tokyo Metropolitan University

2. ロボットが可能な支援

2.1 ロボットの有用性の検討

現状ではコミュニケーションロボットとして ifbot や ApriPoco があげられる[4]。ifbot は表情と会話能力を売りにしたロボットで、計算やなどなどなどの脳を鍛える会話や感情表現できる顔を持っており、認知症予防のための機能を持つ。現在は老人ホームへの導入や、個人での購入もされている。一方の ApriPoco は、人間の音声指示を記憶し、赤外線リモコン操作を行ってくれるロボットである。音声による学習機能を備えた画期的なロボットであり、家電や AV 機器の操作という特定条件下においては優秀なインタフェースの提案を行っている。2つのロボットで大きく異なるのは学習能力の有無で、ApriPoco は学習機能を持つが、機器操作に特化しているという点で本研究とは異なる。

本論文では計算的知能に基づきロボット自身が知覚情報を得て情報を分類、記憶することができる。これを記憶の体制化と呼ぶ。ロボットの体制化と人間の体制化をカップリングすると様々な利点が生まれる。コミュニケーションを通じて体制化が進行すると、ロボットと人間との背景知識の共有化が進み、少ない言葉で会話が成立するようになる。するとユーザは「ロボットが自分を理解してくれている」ということを理解できるようになる。子どもや高齢者の見守りをロボットが代替する場合、ユーザ自身に飽きや不信を感じさせないことが望ましい。また保護者や介護者にとっても安心して見守りを任せられる存在でなければならない。ロボットが人間のパートナーとしての役割を果たすために、ユーザー一人一人に対しパーソナライズしていくことは非常に重要である。例えば認知症患者の場合、自分の言葉がうまく伝わらないというストレスを感じやすい立場にあり、その軽減に効果があるということは大きな利点であるといえる。また、対話という形をとるため、患者がリハビリテーションを行っているという意識をせずに訓練が可能であるという利点もあげられる。また人を相手とした場合と異なり、うまく話をできなくてもプレッシャーや恥を感じなくて済むため、気楽にリハビリテーションを継続することが可能であると考えられる。以上のことからロボットによる支援の意義は大きいといえる。本論文では、体制化のためにセンサから得られた知覚情報や記憶した発話情報などの記号情報間の関係性をロボットは学習する。コミュニケーションに応じて関係性が更新されることで、ユーザに合わせた記憶の分類が行われ、ロボットと人間とで対話の基になる背景知識の共有が行われる。

2.2 時間概念の獲得

ロボットによる効果的な介助を行うにあたり、時間の概念を取り入れることは非常に重要である。人間の生活は常に時間と共にあり、特に治療を要する人にとっては規則正しい生活を心がけることは大切で、かつ社会への適合を考える際に正しい時間の感覚を持つことは必須である。認知リハビリテーションにロボットを用いる場合、要介護者の生活のモニタリングを行うことが考えられる。何時にどのような行動をとったのかを単に記録することは容易であるが、その場合記録された膨大なデータを人の手によって解析、分類しなければならず、介護者や治療者側にとって大きな負担になりうる。このときロボット自身が人間と同じように時間の概念を持っていれば、介護者の代替としてその負担を軽減することが可能となる。例えば夜中に大きな物音を立てるのは好ましくないことや、昼間まで眠っているのは規則正しい生活とはいえないことなど、時間に基づいた「常識」を理解できれば、異常行動に対するマーカー付けや適切な指摘、処置が自律的に行える。すべての行動(入力)に対し反応(出力)を記述することは実環境においては非常に困難であるが、時間概念を持ったロボットならばそれに基づく常識や習慣を身につけることが可能となる。そこから行動が出力されるならば、人間にとって共生がしやすく、パートナーとしての信頼関係を築きやすくなる。

リハビリテーションの効果的な訓練時間を提案したり、投薬の管理をする際など、時間の概念を取り入れることでロボットの行動がより柔軟に、有用になる場面は少なくないだろう。本論文では、ロボットの記憶の体制化を時間に即して行うことで、対話の流れやユーザの習慣をロボットが学習するシステムの提案を行う。

以下にロボットが環境と発話内容の関係性を学習するための、スパイキングニューラルネットワークを用いた手法を示す。

3. 対話のための計算的知能

3.1 MOBiMac

本論文では、パートナーロボットとして一体型のパソコンの筐体を用いた MOBiMac を開発した(図1)。このロボットには2つのサーボモータ、7つの超音波センサ、Pan-Tilt CCD カメラを搭載する。センサ情報や CCD カメラから得られた画像の処理を通し、ロボットは、障害物回避や目標追従を行うことができる。また、ロボットが自然な会話に必要な環境情報を取得するために、人間検出、物体認識、顔の向き検出などを行う。人間検出と物体認識は、定常状態遺伝的

図1 パートナーロボット MoBiMac

アルゴリズム (SSGA) を用いたテンプレートマッチングにより抽出する。さらに同時に複数の物体を検出・識別するために k-means により、画像上の物体候補を分類し、分類後のそれぞれの対象形状の識別にスパイキングニューラルネットワーク (SNN) を用いる。それぞれの形状に対応するセンサニューロンが発火すると、スパイク出力情報は相互認知環境に対する発話システムに送られる [5][6]。

コミュニケーション学習は学習対話モードにおいて行われる。知覚された各記号情報間の関係性が更新されると記憶の体制化が進み、人間の習慣や言動の理解が可能となる。それが MOBiMac の発話の選択強度に影響する。発話システムでは各発話の選択強度に応じて会話の流れが形成され、結果としてユーザに合わせた自然なコミュニケーションの方策が選択されるようになる。以下に、発話システムと関係性の学習について詳述する。

3.2 発話システム

画像処理では SNN と SSGA を用いて物体抽出や人間検出を行う [7][8]。MOBiMac はカメラ画像から抽出した色情報や差分画像などにより物体認識やジェスチャー認識などを行う。画像処理結果の知覚情報は、MOBiMac の内部で数値的に処理され、これを入力情報として発話内容の選択強度を計算する。基本的に、値が大きい知覚情報の組合せとこれらの変化によりロボットは環境の状況を知覚していると考えられる。

ここでは家庭内における人間とロボットとの会話コミュニケーションを想定する。後述の本論文における対話実験を例にとると、道具の名称や用途の教示を行う学習対話や、挨拶やスケジュールの確認など簡単なタスクを要求する対話などがある。ロボットからの発話では「言葉を教えて」や「それは何?」といった学習に関するものの他、日常会話では「おはよう」といった挨拶

図2 発話の流れ

や「今日は何某の予定だったね」といった発話が想定される。図2にロボットと人間の会話の流れを示す。人間の発話 m, l はロボットが認識した発話インデックス、 i, j, k はロボットが発話する発話インデックスを表す。ロボットが発話 k を発話後、発話 m を認識すると、直前の発話 k, m の情報に基づき発話 j が選択・発話される。この様に会話が進む中で、続けて人間の発話 l を認識したとき、次にロボットがどの発話を選択するかは、各発話の選択強度に依存する。ロボットの K 個の知覚情報を用いて、発話 i の選択強度 s_i^U は以下のように計算される [9]。

$$s_i^U = v_j^R \cdot s_{j,i}^R + v_l^H \cdot s_{l,i}^H + s_i^S + y_i^U \quad (1)$$

$$y_i^U = \sum_{k=1}^K (c_{k,i} \cdot x_k) \quad (2)$$

$s_{j,i}^R$ はロボットが発話 j を発話したあとの発話 i の選択強度である。 $s_{l,i}^H$ はロボットが人間の発話 m を認識したあとの人間の発話 l の選択強度である。 v_j^R, v_l^H は発話毎の有効度を表すパラメータである。日常的に頻出する発話ほど有効度は高く、その値は事前に適宜設定しておく。 s_i^S は抑制要素となっている。また、 $c_{k,i}$ は i 番目と k 番目の画像などから得られる知覚情報 (x_k) に対するつながりの強さである。 y_i^U は知覚情報に基づく発話システムへの入力値である。 L を発話総数とすると、発話 i の選択確率は次のボルツマン選択によって計算される。

$$s_i^P = \frac{\exp(s_i^U / \tau^U)}{\sum_{j=1}^L \exp(s_j^U / \tau^U)} \quad (3)$$

τ^U は温度と呼ばれる定数のパラメータであり、この値は小さい値に設定している。温度が低いと、ロボットの発話は、高い選択価値に合わせて限定的に選択される。さらに、発話選択時に、同じ発話内容を連続して選択してしまうことを防ぐために抑制要素 s_i^S が負の大きな値に設定される。

$$s_i^s = \begin{cases} -D & \text{if selected} \\ \tau^s \cdot s_i^s & \text{otherwise} \end{cases} \quad (4)$$

($D > 0$)

τ^s は割引率である．このように発話 i は過去の発話 l, j によって決定され，その発話 j も発話 m, k によって決定されている．過去の発話の情報はこうして連続することで会話の流れを形成する．そのため会話初期には発話システムは機能しない．

ロボットと人間との自然なコミュニケーションが成立するためには，ロボットが人間の言動の意味を理解する必要がある．そのためには周辺環境から知覚された対象物体形状，色，発話のような記号情報の意味づけを行わねばならない[6][10]．コミュニケーションを介して得られた記号情報の学習は発話単語に対応するニューロンと，ジェスチャー，色，および形のボタンに対応するニューロンの間の結合学習で行われる．

3.3 関係性の学習

人間と相互作用しながらロボットが効率よく認知発達するためには，パタンの記号化・分節化が重要であるとともに，記号間，あるいは，分節間の関係性の学習が重要になる．

図3は学習のシナリオをフローチャートで表示したものである．MOBiMacは3つの対話モードを持つ．1つ目は自由な会話コミュニケーションを行う通常対話モード，2つ目は事前に用意されたシナリオベースの対話を行うシナリオ対話モード，3つ目は関係性の学習を行う学習対話モードである．モードの切り替えはキーワードなどのキューとなる音声や画像の入力により行われる．関連性理論では，知覚主体が認知している環境を認知環境と呼ぶ[11]．図4のように，発話，対象物，ジェスチャーの3つのつながりを更新することでロボットの認知環境が整えられる．学習が進むと記憶が体制化され，知覚した情報から関連性の高い発話や，他の知覚情報を想起することが可能になる．

ここでは，ロボットの認知発達をパターン間や記号間の関連性の想起が洗練されることとし，SNNを用いて図4のような様々な情報間の関係性を相関学習する手法を提案する．構造化学習の概念に従い，ヘブ則による学習では，物体認識の精度にあわせて学習係数が変化する．すなわち，分節化や記号化の精度が高いものほど，関係性の学習が加速する．

基本的に，SNNはHodgkin-Huxley方程式に従うが，本論文では，計算コスト削減のために次のように単純化する．各ニューロンは各記号情報に対応している．階層は2階層となっており，1層目のニューロン

図3 学習のためのシナリオ

図4 関係性の学習

i における内部状態 $h_i(t)$ は次のように計算される．

$$h_i(t) = \gamma^{syn} \cdot h_i(t-1) + h_i^{ref}(t) + q_i(t) \quad (5)$$

γ^{syn} は減衰率， h_i^{ref} はニューロンの不応期， $q_i(t)$ は外部からの入力電位である．ニューロンの内部状態が既定した閾値を超えた場合，以下のようなパルスとして出力される．

$$p_i(t) = \begin{cases} 1 & \text{if } h_i(t) \geq \theta_i \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

θ_i は発火の閾値である．ニューロンが発火した場合，下記のように不応期を考慮して，内部状態が抑制される．

$$h_i^{ref}(t) = \begin{cases} \gamma^{ref} \cdot h_i^{ref}(t-1) - R & \text{if } p_i(t-1) = 1 \\ \gamma^{ref} \cdot h_i^{ref}(t-1) & \text{otherwise} \end{cases} \quad (7)$$

($R > 0$)

γ^{ref} は減衰率を表す．2層目のパルス出力と内部状態は，1層目の N 個のニューロンからの出力に従い，次のように計算される．

$$P_j(t) = \begin{cases} 1 & \text{if } H_j(t) \geq \theta_j \\ 0 & \text{otherwise} \end{cases}$$

$$H_j(t) = \gamma^{syn} \cdot H_j(t-1) + \sum_{i=1, i \neq j}^N w_{i,j} \cdot h_i^E(t-1) \quad (8)$$

$$h_i^E(t) = \sum_{k=1}^T \eta^k \cdot p_i(t-k)$$

$H_j(t)$ は2層目のニューロン j の内部状態、 $w_{i,j}$ はニューロン i, j 間の結合強度、 $h_i^E(t)$ は興奮性後シナプス電位 (EPSP)、 η^k は時間減衰を表す。2層目の不応期に関する抑制は1層目と同様に計算される。また、結合強度は以下のヘブ則を用いた学習に基づき更新される。

$$w_{i,j}^N \leftarrow \gamma^{wgt} \cdot w_{i,j}^N - \xi^{wgt} \cdot p_i(t-1) \cdot P_j(t) \quad (9)$$

ここで、 γ^{wgt} は減衰率、 ξ^{wgt} を学習率とする。例えば、学習の結果、ある発話と知覚情報の結合強度が強化されると、同様の知覚をした際に(2)式に従いその発話の選択強度が強化される。

4. 実験

4.1 道具の用途の学習

4.1.1 実験の概要と目的

ロボットに日用品の使い方を教えるという対話実験を行う。この実験ではコミュニケーション学習により、ユーザの使い慣れた言葉をロボットが認識できるようになることを目指す。ここでは日用品としてコップ、湯のみ、グラス、積み木の4種類を用意し、ロボットに用途と任意の名称を教示する(図5)。学習対話モードを利用して言語情報ではなく、知覚情報から発話の想起を行うことができれば、関係性の学習により体制化が進行する様子を確認できる。体制化が進行すると、未知の物体を知覚した際に既知の物体の知識を頼りに類推を行うことができる。手順(1)では未学習の状態から1物体を学習させ、未知物体を見せて様子を確認する。手順(2)ではさらにもう1物体を学習させて同様に未知物体を見せる。手順(3)も同様に行う。一連の試行から関係性の学習により認知発達の過

図5 日用品の用途学習の様子

程を再現する。ここでは関係性の学習という例題に特化するため、教示する名称以外の言語情報は事前に準備しておく。日常会話では少なくとも数百~1,000個程度の語彙が必要である[12]。しかし一つの状況に特化した場合、使用される語彙も限定されることから、予備実験の結果に基づき今回は催促の「言葉を教えて」や「飲む」など用途を表す言葉、知覚結果を伝えるための「(名称)かな?」など21語を事前に準備した。

4.1.2 実験の流れ

実験は以下の手順(1)~(3)の3ステップで実行した。まず手順(1)では、未学習のMOBiMacを通常対話モードに設定し、赤いコップを見せた。その後、MOBiMacを学習対話モードに設定した。「言葉を教えて」というMOBiMacからの催促があるので、用意した赤色の湯飲みを認識させながら音声で「ゆのみ」という名称をMOBiMacに記憶させた。MOBiMacは名称を覚えると何に使う道具なのかを尋ねてくるので、「のむ」ときに使用する物であることを学習させた。その後、もう一度通常対話モードに設定して赤いコップを見せた。次に手順(2)では、再び学習対話モードに切り替え赤色の積み木を認識させながら「つみき」という名称と、「あそぶ」ときに使用するものであることを学習させた。その後もう一度通常対話モードに設定して赤いコップを見せた。最後に手順(3)では、再び学習対話モードに切り替え赤色のグラスを認識させながら今度は「ぐらす」という名称と、「のむ」ときに使用する物であることを学習させた。その後再び通常対話モードに設定し赤いコップを見せた。

4.1.3 実験結果

手順(1)では、赤いコップを見せた際には、MOBiMacは色と形状の認識はしたが(図6)、発話をしなかった。

図6 画像認識

上左：現画像、上右：差分画像、下左：色／形状認識、下右：人間検出、下図のバーはセンサニューロンのスパイク出力を発火後時間減衰される数値情報で表示したもの。

図7 学習結果

湯飲みの学習を教えた後2度目に確認のためにコップを見せるとMOBiMacは「それは湯飲みだったね」、「飲むためのものでしょ」と覚えた言葉をした。

手順(2)で、積み木を学習させてから赤いコップを見せると、MOBiMacは「湯飲みかな？積み木かな？共通点がないな、謎の物体だ」と話した。

手順(3)で、グラスの学習を終えた後3度目にコップを見せると「湯飲みかな？積み木かな？グラスかな？飲むためのものに違いない」と発話した。

学習結果を図7に示す。RCG_Wordが記憶した各物体の名称、Attributeが属性、Internal stateが内部状態を表す。名称は手前から「湯飲み」、「積み木」、「グラス」を表している。属性は左から「赤色」、「四角形」、残り四つが用途でそれぞれ「飲む」、「食べる」、「書く」、「遊ぶ」に対応している。内部状態は色と形状に関しては0～1の連続値で表され、用途に関しては1か0の2値で表されている。例えば湯飲みに注目すると、「赤色」、「四角」、「飲む」の項目との関係性が学習できたことがわかる。

4.1.4 考察

はじめ、手順(1)で赤いコップを見せた時にMOBiMacは発話を行わなかった。図6下部のスパイク出力のグラフを見ると、赤色を検出したことを表す「Rd」や四角を表す「Sq」などが発火していることがわかる。この時点ではMOBiMacにはコップについての知識が体制化されていないため、画像の情報が何を表しているのか理解できず、発話に繋がらなかったと考えられる。

次に湯飲みを学習させたあと再びコップを見せた。このときMOBiMacは目の前に見えているコップの特徴と、過去に学習した湯飲みの特徴との比較を行い、類似度が十分に高かったためコップを既知である湯飲みだと認識したのだといえる。

手順(2)では湯飲みと積み木の見た目上の区別がつかない事をうかがわせる発言が見られた。図7を

見ると、湯飲みと積み木で四角形との関係性が異なることが示されているが、本実験においてはMOBiMacが色情報の差異でしか画像の分類をしていないことに起因している。形状を分類に利用していないのは、現状のMOBiMacでは色と、画像平面上の形状の情報だけで物体を記憶しているために、照明や物体の凹凸、背景に影響を受けやすく、特に形状の差異の認識精度が十分に信頼できないためである。図7より色情報を比較すると、どちらの物体でもほぼ等しい値を示しており、どちらかの物体であると判断することは難しいとわかる。その結果MOBiMacは既に体制化されている知識の中から湯飲みと積み木の用途に着目し、共通点からどのような物体であるかの推論を行ったと考えられる。湯飲みと積み木の見た目の区別はつかないが、その用途が異なる物であるという知識は、コミュニケーション学習を通じ得られたことが発話からも示された。

また手順(3)でグラスを追加学習させてからコップを見せた際には先ほどの手順(2)と同様に既知の物の中から類似度の高い物を選択し、それらの共通点から目の前の物体が何であるかの推論を行っている。今度は湯飲みとグラスが「飲む」もので、積み木が「遊ぶ」ものであることから「飲む」用途に用いる物体であるとの可能性が高いと判断していることがわかる。

この実験で、MOBiMacはコミュニケーション学習を通じてユーザの任意の言葉により道具の用途を学習し、特定の日用品に関する知識を体制化できた。このことから、事前設計では困難であったユーザの使い慣れた言葉への対応が可能となった。

4.2 時間概念獲得のモデル実験

4.2.1 実験の概要と目的

典型的なコミュニケーション学習の例としてMOBiMacに朝、挨拶をし、現在時刻とその日の予定を確認する対話シナリオを用意する。実際の生活時間を基準に実験を行うと、実験自体が大規模になり過ぎ、結果の取得までに長い時間を要する。本論文では実生活における時間軸よりも大幅に短いスパンでの概念形成を試み、以て実生活時間への拡張時の可能性を示唆することを目的とする。対話の流れを記憶することで体制化が進行し円滑に対話が進む様子を示す。モデルの簡略化のためユーザの使用する語彙は最低限に抑え、挨拶「おはよう」、呼びかけ「MOBiMac」、時刻を問う「今何時?」、予定を問う「今日の予定は?」の4つを用いる。対話においてMOBiMacは人間が発話した言葉、発話した時刻、その言葉から誘発されるタスクを記憶する。本実験では10分以内に続けて発話された言葉同

士の関係性を強化していくこととする。ただし簡略化のために時間減衰を省く。以下に、ある発話がされたとき、関係性が閾値を超えている言葉があった場合に、続けて入力されるであろう言葉を先読みし、MOBiMacが関連したタスクを遂行する様子を示す。

4.2.2 実験の流れ

対話の流れを(1)挨拶から時刻を問う、(2)呼びかけから予定を問う、(3)挨拶から予定を問う、(4)挨拶から呼びかけ、の4種類用意し、60分間の関係性の変化を記録した。以後、先に入力した発話を発話1、続けて入力したものを発話2と呼ぶこととする。また、実験開始直後から学習が進んでいく間 MOBiMac の反応の変化を同時に記録していった。

4.2.3 実験結果

発話間の関係性の学習により体制化の進行する様子を図8に示す。図8(a)のグラフの縦軸は発話間の関係性の強化された回数を「強化段階」として表したものである。図8(b)～(d)は体制化の進行の様子を可視

化したもので、強化段階を線分の太さで表している。実験開始後13分で対話の流れ(1)の関係性が2段階強化されているが、挨拶のやり取りの後、時刻を問うまでは MOBiMac は待機をしているままであった。その後19分経過した際再び対話の流れ(1)の関係性が2段階強化されると、MOBiMac はその後挨拶がされるだけで問わずとも現在時刻を告げるタスクを遂行するようになった。また、27分、30分、38分経過後にそれぞれ(4)、(2)、(3)の流れにおいて発話間の関係性が強化されたが、その後呼びかけを入力した後は何らかの発話が入力されるまで待機をし、挨拶を入力した場合では時刻を告げるタスクが遂行された。さらに、45分、47分、52分、56分経過後には対話の流れ(4)の関係性が順次強化されていったが、MOBiMac は55分までは挨拶の後には時刻を告げるタスクを出力し、56分経過後からは挨拶の後に呼びかけに応じる出力を行うようになった。

4.2.4 考察

結果から、発話間の選択強度が3～4段階強化されたときにMOBiMacは発話間に習慣性を見出し先読みを開始するようになったことがわかる。47分～55分の間には対話の流れ(1)と同様に(4)の関係性も十分に高くなっていたが56分経過後に初めて、挨拶の交わされた後に行われる先読みのタスクが時刻の提示から呼びかけに応じるものに変化した。これは関係性の学習において、MOBiMac が過去の習慣を重視し、新規の習慣に馴染むまでにより多くの関係性の強化を要したと見なすことができる。

今回はモデルの簡略化のために関係性の減衰する機会を設けなかったが、ロボットが誤って時間の概念を体制化してしまった際に、叱咤することで関係性を減少させることや、時間減衰の項を取り入れることで、さらに知的な振る舞いが可能になる。習慣づけも忌避も、同様に自然に行えることはコミュニケーション学習の利点である。また、実際には本実験で行ったような10分～1時間程度のスパンにするのではなく、例えば朝・昼・夜に時間帯を分割し数日間から1ヶ月分程度のログを保存できるように拡張すれば、実生活に適った時間間隔でのユーザの習慣を記憶することができる。その際、学習が進み記号間の結合強度が一定以上のものだけに探索範囲を絞ることや、状況を同時に学習しその状況依存的にクラスタリングを行うことで、処理の効率化を行っていく[13]。

このようにロボットが習慣を記憶し、推測を行って行動するようになると、より少ない言葉のやりとりで意思の疎通を図ることができるようになるため、人間

図8 関係性の学習の時間推移

との協調や信頼関係を築きやすくなる。これは先述した人間とロボットの体制化のカップリングの話と繋がるが、「自分を理解してくれている」とユーザに思わせるのに大きな効果が現れる。ロボットが習慣付いて、黙っていても行動を起こすようになると気づけば、ユーザ側からも自然と習慣を守ろうとする動きが誘発できる可能性がある。パートナーロボットを利用するという意識をせずとも、コミュニケーションをとることが生活リズムを整え健全な活動を行う契機となる。

ロボットを用いると一方的な見守りではなく、双方向の相互作用を通じることで、ロボットと人間が対等な関係を築くことができる。要介護者がユーザとなる場合など、人間の介護士に対しては抱いてしまう劣等感を解消でき、ユーザの能動的な活動が促進されるのである。

5. まとめ

本論文では、ロボットが人間との自然なコミュニケーションを行うことを目的として、SNNを用いて環境と発話内容の関係性を学習する手法を提案し、パートナーロボット MOBiMac に実装した。実験を通じ、MOBiMac はコミュニケーション学習により、ユーザに合わせた記憶の体制化を行った。結果から、パートナーロボットがコミュニケーション学習を通じて人間の習慣を習得することで、従来のシナリオベースのコミュニケーションでは実現し得なかった、より自然なコミュニケーションを行うことができた。

コミュニケーション学習を行うということは、ロボットが自身の経験に即した記憶の体制化を行うということであり、ユーザ個人の習慣とロボットの習慣がマッチしていく。コミュニケーションをとればとるほど、ユーザに合った自然なコミュニケーションが可能になるという良い循環がおこる。発展すれば秘密の共有や流行語の利用も可能になると考えられ、ロボットがパートナーとして生活にとけ込むために非常に有効な学習方法である。

今後は、コミュニケーション学習を洗練させ、どの概念を抽出すれば効率よく知識を体制化できるのかということを調査し、ロボットが自律的にその判断を行い認知発達する様な方法論を模索していく。また、パートナーロボットの臨床的な活用については、認知リハビリテーションへの適用を視野に入れ、今回実験で取り扱った、道具用途の教示方法を拡張するなどコンテンツの充実をしていく。また、画像認識の精度を向上させることは勿論、使用法やその他の詳しい性質なども音声で教えられるようにしていく。そして本論文では赤い物体のみで実験を行ったが、MOBiMacは

複数の色情報を取り扱うことができるので、入力する物体のデータ数を増やしてそれぞれを関係づけ、体制化をより綿密に行っていく。現状では極力ロボットにも形状や色などが区別しやすい物を選んで教えねばならない。将来的には物体の三次元形状から、物体を特徴づける不変の項目を抽出・分類し、新しい物体を記憶する度に更新して画像認識へとフィードバックを行っていく。そして周辺の場所情報を集積した環境地図を構築し、その他の各知覚情報と地図情報を交えた体制化を行うことで、記号間の意味づけをより自然に、的確に行えるように拡張していく。

参考文献

- [1] 厚生労働省, 平成17年版厚生労働白書「地域とともに支えるこれからの社会保障」第2部第2章第3節1介護保障制度の現状, 2005.
- [2] Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Washington, DC, American Psychiatric Association, 1994.
- [3] M. Ylvisaker, E.M.R. Gobble, Community Re-Entry for Head Injured Adults, Butterworth-Heinemann, 1987.
- [4] 財団法人機械産業記念事業財団「少子高齢化における次世代ロボットの導入がもたらす社会経済効果」, 2008.
- [5] 寄田, 矢口, 久保田, 認知症進行抑制のための計算的知能に基づく対話とゲーム, 日本シミュレーション&ゲーミング学会全国大会論文報告集 2008年秋号, pp. 93-98, 2008.
- [6] N.Kubota, Associative Learning for Cognitive Development of Partner Robot through Interaction with People, Proc. of International Multi-Conference on Engineering and Technological Innovation, pp. 24-29, 2008.
- [7] N.Kubota, K.Nishida, Perceptual Control Based on Prediction for Natural Communication of A Partner Robot, IEEE Transactions on Industrial Electronics, Vol.54, No.2, pp.866-877, April, 2007.
- [8] 小嶋, 相沢, 久保田, パートナーロボットの対話のための計算的知能, 日本シミュレーション&ゲーミング学会全国大会論文報告書 2006年秋号, pp.157-160, 2006.
- [9] N.Kubota, J.Pei, A.Yorita, H.Kojima, Human Tracking and Natural Communication of Partner Robots, Proc (CD-ROM) of International Symposium on Advanced Intelligent Systems (ISIS2007), pp.205-210, Sokcho, Korea, September 5-8, 2007.
- [10] N.Kubota, Y.Tomioka, Gesture Recognition for a Partner Robot Based on Computational Intelligence, Proc. of the 2008 IEEE International Conference on Fuzzy Systems (FUZZ 2008) of the 2008 IEEE World Congress on Computational Intelligence (WCCI 2008), pp.1477-1482, Hong Kong, China, June 1-6, 2008.
- [11] D. スペルベル, D. ウィルソン, 訳: 内田聖二ほか, 関連性理論伝達と認知, 研究社出版, 1999.

- [12] 内田伸子, 言語発達心理学—読む書く話すの発達—, 放送大学教育振興会, 1998.
- [13] N.Kubota, S.Wakisaka, An Emotional Model Based on Location-Dependent Memory for Partner Robots, *Journal of Robotics and Mechatronics*, Vol.21, No.3, pp.317-323, 2009.
- [14] E.J.Gibson, *Principles of Perceptual Learning and Development*, 1969.
- [15] J.Piaget, 芳賀純 翻訳「諸科学と心理学」, 1970.
- [16] 野家啓一「増補 科学の解釈学」, 2007.
- [17] 平田オリザ「演劇入門」, 1998.
- [18] 国立社会保障・人口問題研究所「日本の将来推計人口」(平成18年12月推計)による各年10月1日現在の推計人口([出生中位(死亡中位)]推計値).
- [19] 長島, 荒井, 統計的手法に基づいた画像・音声情報からの概念獲得, 情報処理学会研究報告, ICS, 知能と複雑系, IPSJ SIG Notes, Vol.2004, No.29 (20040315), pp.193-198.
- [20] 林, 三輪, 人間とエージェントが混在する状況におけるコミュニケーションの認知的・感情的心理特性, ヒューマンインターフェース学会論文誌, Vol.10, No.4, pp.445-455, 2008.
- [21] 御領, 菊地, 江草, 新心理学ライブラリ=7 最新 認知心理学への招待—心の働きとしくみを探る—, サイエンス社, 初版 第8刷, 1998.
- [22] 平松琢弥, 豊かなコミュニケーションの創造に向けて, 文学部論叢, 91 巻コミュニケーション情報学篇, pp.63-76, 2006.
- [23] 佐藤, 西田, 市川, 畑村, 溝口, ロボットによる人間の意図の能動的な理解機能, 日本ロボット学会誌, Vol.13, No.4, pp.545-552, 1995.
- [24] 辻幸夫, ことばの認知科学事典, 大修館書店, 2001.
- [25] 瀬名, 橋本, 梅田, 境界知のダイナミズム, 岩波書店, 2006.
- [26] 阿部, 桃内, 金子, 李, 人間の言語情報処理—言語理解の認知科学—, サイエンス社, 1994.
- [27] 牧, 高尾, 平山, 臨床心理学の理論と実践, ミネルヴァ書房, 1990.
- [28] 佐々木和夫, 脳と情動—知・情・意は解明できるか—, 現代思想 2005. vol.33-2, pp.64-69, 特集 脳科学の最前線, 2005.
- [29] 福田敏男, インテリジェントシステム—適応・学習・進化システムと計算機知能—, 昭晃堂, 2000.
- [30] 武藤, 斎藤, ニューラルコンピューティング応用事例ハンドブック, 共立出版, 2001.
- [31] 尾関, 宮田, 青山, 中村, 作業支援システムのための人工エージェントとのインタラクションを援用した物体認識, 画像の認識・理解シンポジウム (MIRU2007) 7月, 2007.
- [32] 尾関, 青山, 中村, 人工エージェントのノンバーバル表現とメタ表示の関係, 第2回デジタルコンテンツシンポジウム講演予稿集.
- [33] 小林宏, 超リアルな受付システムの開発 (ヒューマン・エージェントインタラクション: 擬人化/顔インタフェース再考), 情報処理学会研究報告, SLP, 音声言語情報処理 IPSJ SIG Notes, Vol.2004, No.74 (20040716), p.63.
- [34] 内田, 荒木, 画像に対する発話を対象とした名詞概念獲得システム SINCA, 知能と情報 (日本知能情報フェジ学会誌), Vol.20, No.5, pp.685-695, 2008.
- [35] 小林由紀, ガーデンパス文の理解成績と作動記憶容量との関係 (I), 川村学園女子大学研究紀要, Vol.16, No.1, pp.63-76, 2005.
- [36] 緒方典裕, *Dynamic Constructive Thesaurus*, 言語研究とシソーラス: 第5回国語研究所国際シンポジウム第1専門部会発表論文集, pp.182-189, 1997.
- [37] 横井, 新妻, 佐々木, 橋本, 知能化空間における物と物の関係の抽出, 計測自動制御学会 SI2008 講演会, 2B1 空間知能化とアプリケーション (1), 2B1-2.
- [38] 新妻, 橋本, 空間知のためのモノ情報を考慮した人の活動内容の記述, 計測自動制御学会 SI2008 講演会, 1B4 空間知 (4), 1B4-5.
- [39] 山口, 福里, 鶴飼, 水川, 空間知サービスロボットにおける RT オントロジー構成, 計測自動制御学会 SI2008 講演会, 1B1 空間知 (1), 1B1-1.
- [40] 西本, 中沢, 嵯峨山, 音声対話における擬人化エージェントの利用効果の検討, 情報処理学会研究報告, SLP, 音声言語情報処理, IPSJ SIG Notes Vol.2003, No.75 (20030718), pp.25-30, 2003.
- [41] 三浦, 吉川, 浅田, 音声相互模倣過程を収束に導くマグネット効果, 人工知能学会 第24回 AI チャレンジ研究会, JSAI Technical Report SIG-Challenge-0624, pp.29-24, 2006.
- [42] 志水, 中西, Sensitive Agent: シミュレーションのための人間エージェント間インタラクションの設計, インタラクション 2004, 2004.
- [43] O.Brdiczka, J.L.Crowley, P.Reignier, Learning Situation Models in a Smart Home, *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol.39, No.1, pp.56-63, February, 2009.

(2009年3月6日 受付)
(2009年8月7日 採録)

[問い合わせ先]

〒191-0065 東京都日野市旭が丘6-6
首都大学東京大学院システムデザイン研究科
久保田 直行
TEL: +81-42-585-8441 8441 (内線)
FAX: +81-42-585-8441
E-mail: kubota@tmu.ac.jp

著者紹介

くぼた なおゆき
久保田 直行 [正会員]

1997年名古屋大学大学院工学研究科マイクロシステム工学専攻博士後期課程修了。2005年から首都大学東京システムデザイン学部ヒューマンメカトロニクスシステムコース准教授、現在に至る。2002年科学技術振興機構(JST)戦略的創造研究推進事業「相互作用と賢さ」領域研究員、2006年-2008年同事業発展研究研究員、2007年英国王立協会客員教授、博士(工学)。

やぐち あいこ
矢口 愛子 [非会員]

2009年現在首都大学東京システムデザイン研究科ヒューマンメカトロニクスシステム学域博士前期課程在学。主としてパートナーロボットのコミュニケーションに関する研究に従事。

Learning of Utterance Contents for A Partner Robot Based on the Environmental Perception by

Naoyuki KUBOTA and Aiko YAGUCHI

Abstract :

The expectation to home partner robots has been increasing for the care of aged people and children. However, the communication system of partner robots is based on scenarios designed beforehand, which makes natural communication with people difficult. In this paper, we aim at realizing natural communication between people and robots by the learning of the relationship between perceived information and utterance contents. First, we propose a method based on spiking neural networks for the learning of the relationship. Next, we conducted several experiments on the learning of the relationship between perceived information and time-dependent utterance contents. The experimental results show that the robot can update utterance contents based on the cognitive development according to the interaction with the person.

Keywords : Partner Robots, Environmental Perception, Natural Communication, Learning

Contact Address : **Naoyuki KUBOTA**

Graduate School of System Design Tokyo Metropolitan University
6-6 Asahigaoka, Hino, Tokyo, JAPAN
TEL : +81-42-585-8441 (8441)
FAX : +81-42-585-8441
E-mail : kubota@tmu.ac.jp