ロボットの語意学習のための主観的整合性に基づくマルチモーダルカテゴリゼーション

Multimodal Categorization Based on Subjective Consistency for Robot Learning Word-Meaning from Human

笹本 勇輝

Yuki Sasamoto

大阪大学 大学院工学研究科

Graduate School of Engineering, Osaka University

yuki.sasamoto@ams.eng.osaka-u.ac.jp

吉川 雄一郎

Yuichiro Yoshikawa

大阪大学 大学院基礎工学研究科

Graduate School of Engineering Science, Osaka University

yoshikawa@irl.sys.es.osaka-u.ac.jp

浅田 稔

Minoru Asada

大阪大学 大学院工学研究科

Graduate School of Engineering, Osaka University

asada@ams.eng.osaka-u.ac.jp, <http://www.er.ams.eng.osaka-u.ac.jp/asadalab/index.html>

keywords: multimodal categorization, word-meaning, subjective consistency

Summary

This paper proposes a novel method of multimodal categorization for learning word-meaning. It is assumed that the multimodal observations do not necessarily capture matched object unlike previous work, but parts of them are still often matched with each other. Subjective consistency is introduced, which measures to what extent probability of object category in one modality is judged to be close to those in different modalities and expected to be utilized for finding the matched parts in multimodal observations. We apply this idea to extend the previous method called Multimodal Latent Dirichlet Allocation for coping with the above assumption. Experimental results both with real and artificial data show the efficiency of the proposed method for multimodal categorization using multimodal data involving unmatched observations which are considered to be normal in more realistic situation of learning word-meaning through interaction with humans.

1. はじめに

近年のロボット技術の発展に伴い、人と同じ生活環境で活躍するロボットの実現に期待がよせられている。これに対し、ロボットは人の振舞を観察し、それに示されている人の意図を理解できる必要があり、人の発話の意味（それが示す事物）の特定、すなわち語意の理解は、その典型的な課題であると考えられる。ユーザの発話について学習する場合、対象となる語意は、対面するユーザや環境に存在する事物に応じて異なることが予想されるため、ロボットはユーザと相互作用し、自律的に語意学習できることが望ましい。また実環境においては、常にユーザとロボットが注意を共有しているとは限らない。すなわち、ユーザが発する言葉が必ずしもロボットが注目している事物を表すとは限らない。従って、ロボットは、ユーザとの相互作用から得られたデータすべてを単純に

学習に利用するのではなく、語意の学習に必要なデータを取捨選択できることが望ましい。本研究では、そのような必ずしもユーザとロボットが同じ事物に注目しているとは限らない状況におけるロボットの語意学習の実現を目指す。

ロボットがユーザとの相互作用を通じて語意を学習するための手法は、これまでいくつか提案されている。Gorin et al. は、ユーザの発話による指令を可能にすることを目的として、ユーザの発話とシステムの所望の動作の対応を示したデータベースから、ユーザの発話を分類する手法を提案している [Gorin 91, Gorin 94]。しかしこれらの研究では、システムの所望の動作はあらかじめカテゴリ化されていた。ユーザの発話や指示内容は、対面するユーザや環境に存在する事物に応じて異なることが予想されるため、それを事前に仮定できない状況も考慮すべ

きである．

これに対し，Roy and Pentland や Yu は，ユーザの発話と，ユーザの発話が見る物体の関係を学習する課題において，発話とその発話に対応する物体についての観測情報のデータベースから，この両方ともをカテゴリ化する問題を扱った [Roy 02, Yu 05]．また，田口らや 中村らは，さらに考慮するモダリティの数を増やすことで，高精度な，あるいは人と同様のカテゴリ化ができる可能性を示している [田口 10, 中村 10]．しかし，これらの研究では，ユーザがロボットが注目している物体についての発話を行うなど，マルチモーダルな観測情報はモダリティ間で対応するものを捉えていることが想定されており，この想定が崩れる際に，どのようにシステムが対応すべきかには焦点はあてられていなかった．

一方，ユーザとロボットが必ずしも同じ事物に注目するとは限らない状況において，ロボットが人の発話音声とその発話に対して自身がとるべき行動の対応関係を学習する手法が提案されている．しかしながら，ロボットがとった行動に対して，予め設定した目標へ到達したかどうか [小野 09]，あるいはユーザからの報酬 [中谷 11] などの外的な評価が与えられることが想定されている．ロボットが，物体の特徴抽出器などロボットの特定の行動とは直接的な関係がみえにくいものについて学習する場合，ユーザがそれを適切に評価できるかは不明である．そのため，外的な評価に頼らない，自律的な学習機構の開発が望まれる．

そこで本研究では，ユーザとロボットが必ずしも同じ事物に注目しているとは限らない状況においてロボットが自律的に語意を学習するための手法を提案する．ここでの問題は，ロボットの観測に，注目している事物とは対応しないデータが含まれることである．例えば，人がロボットが見ている物体の特定に失敗する場合，ロボットが見ている物体（ロボットの注目画像特徴）と聴取した人の発話音声（人の音声特徴）は，それぞれ別の事物を表す．しかし，その場合でも，人が見ている物体（人の注目画像特徴）と発話音声は同じ事物を表す，すなわち，部分的にはある事物に対応するデータが含まれることがある．本研究では，マルチモーダルな観測のモダリティ間の整合性を評価することで注目すべきモダリティを決定する主観的整合性を導入し，部分的な対応データを利用したカテゴリ化を実現できることを示す．以下では，まず，2 章において，本研究で想定するユーザとロボットが必ずしも同じ事物に注目しているとは限らない状況における語意学習について説明し，3 章でこれに関する従来手法 [中村 10] を述べた上で，これを拡張する形で提案手法のアイデアを説明し，実装方法について述べる．4 章では，提案手法の有効性を評価するために実施した，実データを用いた実験と計算機シミュレーションの結果を示し，最後に，本研究の位置づけと今後の展望について述べる．

図 1 人から語意を学習するロボットが置かれている環境

2. 語意学習状況

人（教示者）とロボット（学習者）と物体が存在する環境を考える（図 1）．ロボットは，環境中のいずれかの物体をランダムに選択し，それを注視する．ロボットの語意学習の教示者となる人は，ロボットと同様にいずれかの物体を注視し，同時に語を発声する．このような状況において，ロボットは，ロボットの注視先の画像特徴ベクトル w^{v_s} ，人の注視先の画像特徴ベクトル w^{v_o} 及び人が発する語の音声特徴ベクトル w^s を観測できるとし，本研究では，これら 3 種の特徴ベクトルの対応関係を学習することを語意学習の課題とする．

また本研究では，人とロボットが同じ事物に注目しているとは限らないことを仮定し，人がロボットの見ている物体と同じ物体に注目する確率（対応率） P_c を導入する．図 1 (a) に示すように人とロボットが常に同じものに注目する場合，すなわち $P_c = 1$ の場合，各観測 w^{v_s} ，

図2 特徴ベクトルの計算過程の一例

w^{v_o} , w^s は同じ物体に対応する特徴となる．これまでの研究では，このように同時に観測された特徴をカテゴリ化することで特定の物体と特徴の対応関係，すなわち，語意が学習されていた．しかし実際には，人がロボットの注目物体の特定に失敗したり，人がロボットの注目物体以外（例えば，ロボットの顔）を見ながら物体の名称を発したり，あるいは，ロボットと同じ物体を見てはいるがその物体の名称以外を発したりする場合が起こる，すなわち $P_c < 1$ となる場合が起こると考えられる．そのような場合， w^{v_s} , w^{v_o} , w^s はそれぞれ異なる物体に対応する特徴となる（図 1 (b)）．そのため，同時に観測された特徴を単純にカテゴリ化するだけでは，実際には物体とは対応しない特徴もその物体に対応する特徴として捉えてしまう可能性がある．

一方，このような場合でも，観測される複数のモダリティの一部は同じものを捉えたものである場合があると考えられる．例えば，人の音声はロボットの見ているものとは対応しないが，人自身の見ているものとは対応する場合や、人がロボットの方を見ながら，ロボットの見ているものに関する語を発する場合などである．このようなとき，観測の一部，ここでは w^{v_o} と w^s あるいは w^{v_s} と w^s ，は同じ物体に対応する特徴であるため，その部分的な対応関係の学習だけでも進められることが望ましい．しかし，観測された特徴のうち，どのモダリティの特徴が対応しているものであるかは状況に依存し，事前には規定できないため，観測に応じて適応的に，その対応をロボット自身が判断することが必要であると考えられる．本研究では，そのような部分的な対応をロボットが主観的に抽出するための主観的整合性を導入し，それに基づいた語意学習手法を提案する．

3. 主観的整合性に基づくマルチモーダルカテゴリゼーション

本研究では，マルチモーダルカテゴリゼーションの従来手法である Multimodal LDA [中村 10] を拡張する形で，観測されるマルチモーダル情報が部分的にしか対応しな

(a) 従来手法

(b) 提案手法

図3 ロボットの語意学習のための (a) 従来手法 [中村 10]，および (b) 提案手法のグラフィカルモデル．ただし，(a) の表記は提案手法との比較のため原著から変更している．

いことがありうる状況に対処するためのアイデアである主観的整合性に基づいたカテゴリ化を導入し，その実装について述べる．

3.1 基本的な学習手法: Multimodal LDA

Multimodal LDA [中村 10] は，元々は文書分類の手法として提案された LDA [Griffiths 04] をマルチモーダル情報を扱うように拡張した手法であり，マルチモーダル情報をそれらの潜在変数で対応づけ，その対応に基づいて各モダリティのカテゴリを形成するものである．

Multimodal LDA におけるパラメータの推定方法は以下の手順によって実現される．ロボットが複数のモダリティにより M 回の観測をすることを考え，一度の観測でモダリティ m において特徴ベクトル w^m が観測されるとする．これは，モダリティ m における N^m 個の各注目領域の情報を量子化した特徴を並べた N^m 次元ベクトルである（図 2）．ここでモダリティ m の各注目領域の情報が量子化される特徴の種類は W^m である．

Multimodal LDA では，このような特徴を，図 3 (a) のグラフィカルモデルに示す生成過程によって得られるも

のと見なす．ここで， θ は潜在変数の多項分布パラメータ， β^m はモダリティ m の特徴の多項分布パラメータ， α ， π^m は，それらのディリクレ事前分布パラメータ（ハイパーパラメータ）である．このグラフィカルモデルにおいて，潜在変数 z_{mij} は， θ ， β^m を周辺化した，

$$p(z_{mij} = k | z^{-mij}, w^m, \alpha, \pi^m) \propto \left(\frac{N_{kj}^{-mij} + \alpha}{N_j^{-mij} + K\alpha} \right) \frac{N_{ml(i)k}^{-mij} + \pi^m}{N_{mk}^{-mij} + W^m \pi^m} \quad (1)$$

に従ったギブスサンプリングのアルゴリズムにより推定できることが示されている [中村 10]．ここで， $i \in \{1, \dots, N^m\}$ は，量子化した特徴ベクトルのインデックスであり，注目領域に対応する． $j \in \{1, \dots, M\}$ は観測回， z_{mij} は j 番目の観測におけるモダリティ m の特徴ベクトルにおける i 番目のインデックスの特徴を生成した分布を示す潜在変数（語に対応するカテゴリ）である． K は潜在変数の数（学習するカテゴリ数）， $l(i) \in \{1, \dots, W^m\}$ は特徴ベクトルにおける i 番目のインデックスの特徴の量子 ID である． N_* は， M 回の観測を通じて，添字 $*$ で特定される特徴が観測された回数を表す．また，上付きの添字 $-mij$ は観測された全ての特徴を要素とする集合 U から j 番目の観測におけるモダリティ m の特徴ベクトルにおける i 番目のインデックスの特徴を除いた集合 U^{-mij} についての量を表す指標である．従って具体的には， N_{kj}^{-mij} は U^{-mij} の集合に含まれるもののうち， j 番目の観測で得られる特徴の潜在変数が k であった回数であり，同様に N_{mk}^{-mij} はモダリティ m の特徴の潜在変数が k であった回数， $N_{ml(i)k}^{-mij}$ はモダリティ m の $l(i)$ の潜在変数が k であった回数である．

式 (1) の右辺乗算の第一項はおよそ N_j^{-mij} と N_{kj}^{-mij} の比によって増減する項であるため全体として，同じ観測回の（共起する）別のモダリティの特徴の潜在変数も考慮して現在の特徴の潜在変数を k とするべき程度を表しているともみなすことができる．右辺乗算第二項はおよそ N_{mk}^{-mij} と $N_{ml(i)k}^{-mij}$ の比によって増減する項であり，全体として，現在のモダリティの $l(i)$ が，別の観測回や同じ観測回の特徴ベクトルにおいて現在注目しているものとは別のインデックスの特徴も考慮したときに，潜在変数 k についてどれほど特長的に観測される特徴であるかを表しているとみなすことができる．すなわち，式 (1) の右辺は，現在の観測回の注目モダリティの特徴ベクトルにおいて注目しているインデックスの特徴の潜在変数を「すべてのモダリティが示す潜在変数との共起性」（右辺乗算第一項）と「注目している特徴の潜在変数における特有性」（右辺乗算第二項）に基づいてサンプリングすることを意味する．

式 (1) 中の N_* ははじめ一様乱数からサンプリングされた z_{mij} によって初期化しておき，特定の mij において，式 (1) に従ったサンプリング結果に基づき更新して

く．具体的には，特定の mij において k とサンプリングされた場合，

$$N_{mijk} = N_{mijk} + 1 \quad (2)$$

と加算し，各 N_* の添字が特定する範囲に含まれている場合，これに応じて更新する．式 (2) より更新した N_* を用いて式 (1) によるサンプリングを繰り返していくことで， z_{mij} が収束し， N_* の収束値 $\bar{N}_*$ が得られる．これを用いて，最終的なパラメータの推定値が

$$\hat{\theta}_{kj} = \frac{\bar{N}_{kj} + \alpha}{N_j + K\alpha}, \quad \hat{\beta}_{w^m k}^m = \frac{\bar{N}_{mw^m k} + \pi^m}{\bar{N}_{mk} + W^m \pi^m} \quad (3)$$

のように算出される．

以上の手順により，式 (1) の右辺乗算第一項の働きにより，人とロボットが常に同じ物体に注目する場合，その共起性を手がかりに，各モダリティの特徴ベクトルにおける各インデックスの特徴が示す語のカテゴリ $z = k$ を推定することができ，正しい対応関係を示すパラメータの学習ができる．しかしながら，人とロボットが必ずしも同じ物体に注目していない場合，すなわち，いずれかのモダリティに観測される特徴が他のモダリティに観測される特徴が示す語カテゴリと一致しない分布から生成される場合，共起性が高く評価されないために，正しい語カテゴリのサンプリングへのバイアスが期待できない．

3.2 主観的整合性に基づく拡張

人とロボットが必ずしも同じ物体に注目していない場合の Multimodal LDA の問題に対し，本節では，マルチモーダル情報の部分的な対応関係をロボットが主観的に抽出し，そのような場合でも，正しい語カテゴリのサンプリングを可能にする語意学習手法を提案する．

これまでの手法では，観測回毎に，一つのカテゴリ分布パラメータ θ を求めていた．しかしこれでは，マルチモーダルデータの一部が他と対応しないことを表現できない．そこで，本研究では，観測したモダリティ毎に潜在変数の分布パラメータ $\theta^{v_s}, \theta^{v_o}, \theta^s$ を考慮したグラフィカルモデルを考える (図 3 (b))．ここで，3.1 節と同様に， j 番目の観測のモダリティ m の特徴ベクトルにおける i 番目のインデックスの特徴を生成した分布を示す潜在変数 z_{mij} を

$$p(z_{mij} = k | z^{-mij}, w^m, \alpha, \pi^m) \propto \left(\sum_{r \in \{v_s, v_o, s\}} \lambda_{mij}^r \frac{N_{rkj}^{-mij} + \alpha^r}{N_{rj}^{-mij} + K\alpha^r} \right) \frac{N_{mk}^{-mij} + \pi^m}{N_{mk}^{-mij} + W^m \pi^m} \quad (4)$$

に従ってサンプリングすることを考える．ここで， λ_{mij}^r は， z_{mij} が，別のモダリティ r のデータを観測したことで得られる潜在変数の分布からどの程度生成されるかの度合い（以後，主観的整合性と呼ぶ）を表している．マルチモーダル情報の対応学習の問題の従来研究において，主観的整合性に基づいて各モダリティのデータを重

みづけることで、他のモダリティのデータとは対応しない整合性の低いデータを排除して対応学習できることが示されている [笹本 13]。本研究では、このアイデアを応用し、式 (4) のようにモダリティ毎に示される潜在変数の分布を主観的整合性によって重みづけることを考える。主観的整合性は、以下のように定義する。

$$\lambda_{mij}^r = \frac{\kappa_{mij}^r}{\sum_{o \in \{v_s, v_o, s\}} \kappa_{mij}^o}. \quad (5)$$

ここで、 κ^r は、モダリティ r が示す潜在変数の分布が、他のモダリティが示す分布とどの程度一致しているかを表し、以下の式で表す。

$$\kappa_{mij}^r = \exp \left(- \sum_{\substack{q \in \{v_s, v_o, s\} \\ q \neq r}} D(P(z_{mij}|\tilde{\theta}_j^r), P(z_{mij}|\tilde{\theta}_j^q)) / \sigma \right), \quad (6)$$

ただし、 $P(z_{mij}|\tilde{\theta}_j^{m'})$ は、 j 番目の観測において、モダリティ m' のデータを観測したことで得られる潜在変数の分布に相当し、パラメータは、

$$\tilde{\theta}_{kj}^{m'} = \frac{N_{m'kj}^{-mij} + \alpha^{m'}}{N_{m'j}^{-mij} + K\alpha^{m'}} \quad (7)$$

と計算される。また、 $D(P(z_{mij}|\tilde{\theta}_j^{m'}), P(z_{mij}|\tilde{\theta}_j^{m''}))$ は、 $P(z_{mij}|\tilde{\theta}_j^{m'}), P(z_{mij}|\tilde{\theta}_j^{m''})$ の分布間の距離を表し、 σ はその距離に対する逆感度パラメータである。これらの分布は、 z_{mij} に関する離散分布であるので、分布間の距離は、以下のように Intersection で評価する。

$$D(P(z_{mij}|\tilde{\theta}_j^{m'}), P(z_{mij}|\tilde{\theta}_j^{m''})) = \sum_{k=1}^K \min(\tilde{\theta}_{kj}^{m'}, \tilde{\theta}_{kj}^{m''}). \quad (8)$$

式 (4) は、先行研究におけるサンプリングの式 (1) と比べて、右辺乗算第一項が異なることがわかる。式 (4) の右辺乗算第一項は、式 (1) のように、単純にすべてのモダリティが示す潜在変数との共起性を評価するのではなく、よく対応する、すなわち整合性のとれるモダリティが示す潜在変数との共起性を評価するように変更されたものと見なすことができる。これにより、あるモダリティのデータが示す潜在変数が、他のモダリティが示すものと一致していない場合でも、一致しているモダリティの潜在変数を重視し、それが示す語カテゴリへバイアスしてサンプリングできると考えられる。

N_* は、従来研究と同様に更新していくが、その更新量は、現在の mij における潜在変数の分布が、他のモダリティが示す潜在変数の分布とどの程度一致しているかに応じて変更するよう、以下のように計算する。

$$N_{mijk} = N_{mijk} + \kappa_{mij}^m. \quad (9)$$

最終的なパラメータの推定値は、 N_* の収束値を用いて、

$$\hat{\theta}_{kj}^m = \frac{\bar{N}_{mkj} + \alpha^m}{\bar{N}_{mj} + K\alpha^m}, \quad \hat{\beta}_{w^mk}^m = \frac{\bar{N}_{mw^mk} + \pi^m}{\bar{N}_{mk} + W^m\pi^m} \quad (10)$$

と算出される。

図 4 実験 1 における環境

以上のように、整合性の高いモダリティとの共起性が手がかりに語カテゴリを推定することで、あるモダリティで観測された特徴が示す語のカテゴリが、他のモダリティで観測された特徴が示す語のカテゴリと一致しない場合でも、共起性が高く評価されない問題を抑制できると考えられる。また、その整合性に応じて、 N_* の更新量を決定することで、注目している特徴が、サンプリングされた語カテゴリが示す特徴をどの程度捉えたものとみなすかを調整できる。これにより、ある特徴が示す語と他のモダリティによってバイアスされてサンプリングされた語が異なる場合には、その特徴をその語カテゴリの特徴としないようにカテゴリ化できると考えられる。

4. 実験

提案手法の有効性を確かめるために、実ロボット及び計算機シミュレーションを用いて以下の二つの実験を実施した。

実験 1: 実データを用いた検証 提案手法が実環境において有効であるかを検証する。具体的には、まず、人とロボットが常に同じ物体に注目する状況での人の教示を記録し、そのデータの組み合わせを対応率にもとづき非対応データと入れ替える。これにより、仮想的に人とロボットが同じ物体に注目しない状況を作り、ロボットが提案手法によって正しく語意を獲得できるかを検証する。

実験 2: 人工データを用いた詳細な検証 提案手法の効果を詳細に検証するために、人工データを用いた計算機シミュレーションを実施する。具体的には、3つのモダリティの内、いずれか1つのモダリティのデータが他の2つのモダリティのデータと正しく対応しない場合 (実験 2-1)、特定のモダリティのみ他のモダリティと対応しない場合 (実験 2-2) に提案手法により対応しないデータがどのように抑制できるようになり、どのように対応関係が獲得されていくかを検証する。

図5 実験1における学習対象及び非学習対象(観測データとして用いる画像の一例))

4.1 実験1: 実データを用いた検証

§1 基本設定

図4のように、ロボットと人の間に複数の物体がランダムに配置されており、ロボットはこのうち一つをランダムに選択して注視するとする。そして人は、一定の対応率 P_c でロボットの注視物体の語を正しく教示する。すなわち確率 P_c で、ロボットと同じ物体を注視すると同時に、この物体の名称を発し、確率 $1 - P_c$ で、そうでない非教示的な振る舞い(注視および発話)をするとする。ロボットはこのいずれの場合においても、ロボットの注目先の画像特徴ベクトル w^{v_s} 、人の注目先の画像特徴ベクトル w^{v_o} 、および人が発した音声特徴ベクトル w^s を観測する。これら一連の観測をまとめて1ステップと呼び、 M ステップ繰り返して得られたデータを用いて語意学習を行う。

正しい教示がなされる場合、 w^{v_s} 、 w^{v_o} 、 w^s は対応する物体に関する特徴を含むデータの組となるため、対応データと呼ぶ。これに対し、人が非教示的に振る舞う場合に得られるものを非対応データと呼ぶ。本研究では、非対応データとして、いずれか二つのモダリティの特徴は対応しているが、残りの一つのみがこれとは異なるものを含むものとし、どの二つのモダリティが対応するものとなるかはランダムに決定されるとする。これは、

- (1) ロボットの注目先の画像特徴が対応しない場合
人は自分の見ているものの名称を発しているが、ロボットの注目物体と異なる物体を見ている
 - (2) 人が発した音声特徴が対応しない場合
ロボットと人が見ているものは同じであるが、人がこの物体の名称以外の語を発する
 - (3) 人の注目先の画像特徴が対応しない場合
人はロボットが見ているものの名称を発するが、人がその物体を見ていない
- という3つ場合が等確率に起こることを仮定したことに相当する。なお、これら以外に、
- (4) どのモダリティの特徴も対応しない場合
人はロボットが見ているものには注目しておらず、また自分の見ているものの名称以外の語を発する。

という場合も考えられる。しかし、本研究では、マルチモーダルな観測に含まれてしまう他とは対応しないモダリティのデータをいかに排除した学習が可能であるか、に焦点をあてるため、この(4)は想定しないこととした。

対応データ、非対応データの一例として、人の注目先の画像を図5に示す。本実験では、図5に示すような6つの子供の玩具を学習対象とし、各玩具について20回観測しておくことで得た各モダリティのデータを再サンプルすることを各ステップにおける観測とみなす。これらのうち、いずれかのモダリティの観測を対応しないものに入れ替えたものが非対応データとして扱われる。例えば、図5の物体1を観測とする場合、非対応データとして入れ替えられる候補となる観測データは、物体2や物体3など学習対象の内の別の物体に関するものに加え、ロボットの外見などの非学習対象を捉えた観測である。学習対象でないものの観測は、モダリティごとにそれぞれ20回ずつ行った。モデルのハイパーパラメータは、経験的に、 $\alpha^m = 1.0$ 、 $\pi^m = 1.0$ ($m \in \{v_s, v_o, s\}$)、 $\sigma = 0.5$ とした。

§2 特徴ベクトル

ロボットが観測する画像、音声特徴として、中村ら[中村10]と同様に、各モダリティのセンサ情報を量子化した特徴ベクトルを用いる。

ロボットの注目先の画像特徴ベクトル w^{v_s} は、ロボットの頭部のUSBカメラから取得した画像の物体領域(ROI)において10ピクセル間隔で観測される、合計4個のマルチスケール(半径4,8,12,16ピクセルの領域)の各領域におけるDSIFT特徴量[Vedaldi08]を500個のコードブックベクトルを用いて量子化した特徴を並べたベクトルとする。一般に、カメラ画像から物体を認識する場合には、オクルージョンの問題があるが、ここでは、カメラ画像には一つの物体のみ含まれることとした。

人の注目先の特徴量ベクトル w^{v_o} の取得には、人の注目点の画像情報を得る必要がある。これに関して、人の顔画像から推定する方法等が提案されている(例えば、[小野07])が、ここでは対応しないデータを排除した学習に主眼を置くため、この問題は扱わず、人の頭部にUSBカメラを設置し取得した画像を人の注目画像として用いる。 w^{v_o} は、このようにして得た人の注目画像に、 w^{v_s} の取得と同様の処理を施して取得する。

音声特徴ベクトル w^s は、ロボットの頭部側面にあるマイクから取得した音声から20ms毎の各領域における12次元のMFCC特徴量[Boersma14]を50個のコードブックベクトルを用いて量子化した特徴を並べたベクトルとする。

§3 評価方法

主観的整合性を用いないで学習した場合(式(1)(2)に基づいた学習手法: 従来手法)と、主観的整合性を用いて学習した場合(式(4)(9)に基づいた学習手法: 提案手法)の学習後の分類性能を比較する。分類性能は、学習後

図6 対応率 P_c に応じた学習後の F 値の 10 回の平均値と標準偏差

に、物体毎に新たに 10 回づつ観測して得られたデータ、 w_{test}^{vs} , w_{test}^{vo} , w_{test}^{s} を、学習時に推定したパラメータを用いて、

$$\begin{aligned}\hat{z} &= \arg \max_z P(z | w_{test}^{vs}, w_{test}^{vo}, w_{test}^s) \\ &= \arg \max_z \int P(z | \theta) P(\theta | w_{test}^{vs}, w_{test}^{vo}, w_{test}^s) d\theta \quad (11)\end{aligned}$$

のように分類した結果が、正しく対応する語意であるかどうかで評価する。ただし、主観的整合性を用いる場合は、 $\theta = \{\theta^{vs}, \theta^{vo}, \theta^s\}$ である。また、 $P(\theta | w_{test}^{vs}, w_{test}^{vo}, w_{test}^s)$ は、式 (1)(提案手法の場合は、式 (4)) を、学習時の収束値 $\bar{N}_{mw^m k}$ を初期値として用いて再計算することで求める。

分類した結果が正しいかどうかは、F 値で評価した。F 値は、適合率 P と再現率 R の調和平均として、以下のように計算される。

$$F = \frac{2 * P * R}{P + R}, \quad (12)$$

ただし、モデルが獲得した語のカテゴリが、実際のどの語に対応するかは、モデル自体からは判断できないので、ここでは、獲得したカテゴリ k' を実際の語 k とおき、式 (12) を計算する処理を、すべての $k, k' \in \{1, \dots, K\}$ の組み合わせについて行い、最も F 値の高かった組み合わせを正しい分類とした。適合率 P と再現率 R は、以下のように計算される。

$$P = \sum_{k=1}^K \frac{a_k}{a_k + c_k}, \quad R = \sum_{k=1}^K \frac{a_k}{a_k + b_k}.$$

ここで、 a_k は、モデルがテストデータをカテゴリ k と分類した回数を表し、 c_k は、カテゴリ k と分類されたデータの内、実際には対応する語 k 以外のデータである場合の数であり、 b_k は、実際の語は k のデータであるが、それをカテゴリ k 以外に分類した場合の数である。

§4 実験結果

対応率 P_c を 0.0, 0.1, 0.2, 0.5, 0.6, 0.9, 1.0, それぞれの場合で、一つの物体につき 60 回観測したデータを用いて、すなわち総ステップ数 $M = 360$ とした語意学習のシミュレーションを実施した。図 6 は、 P_c に基づい

図7 実験2で用いる人工データの分布

て対応データ・非対応データの組み合わせを変えて 10 回学習させた際の、学習後の分類性能の平均値と標準偏差を示したものである。

図 6 より、全ての対応率の条件において、提案手法 (白丸) の方が主観的整合性を用いない従来手法 (黒丸) よりも高い分類性能を示していることが分かる。

全ての観測が対応するものである場合 ($P_c = 1.0$)、従来手法 (黒丸) でも、高い分類性能が得られているが、対応率が下がると ($P_c < 1$)、分類性能が低下している。これに対して、提案手法を用いた場合、中程度の対応率 ($P_c = 0.6, 0.5$) まで、対応率が 1.0 であるときと同程度の分類性能が維持されていることが分かる。また、対応率がかなり低くなった場合 ($P_c = 0.2, 0.1, 0.0$) でも、対応率が中程度である場合に比べて分類性能は低下するものの、従来手法に比べて高い分類性能が得られている。すなわち、提案手法を用いることで、対応率に対して頑健に学習が可能であると言える。

本実験では、対応率が 0.5 まで低下しても、完全に教示する状況 ($P_c = 1.0$) と同程度の分類性能が得られている。この状況は、観測データのうち半分は非対応データである場合、すなわち、人が 2 回に 1 回は完全には教示しない状況に相当しており、提案手法により、人がロボットに語意を教示する際の注意に関する制限を緩和可能であることが示唆される。

4.2 実験 2: 人工データを用いた詳細な検証

次に、提案手法による学習の効果を詳細に分析するために、人工データセットを用いた検証を行った。実験 2-1 として、あるモダリティのデータが他のモダリティのデータと正しく対応しない場合に、提案手法により対応しないデータを用いた学習が抑制される過程をより詳細に分析する。そのために、すべてのモダリティの特徴の種類を $W^m = 6$ と仮定し、実験 1 と同じ設定の語意獲得シミュレーションを実施する。さらに、実験 2-2 として、特定のモダリティのデータのみ他のモダリティのデータと対応しない場合での検証を行う。そのために、注目率 P_a を導入し、それに応じてモダリティ v_s のみに非対応データが観測されるとした、語意獲得シミュレーションを実

図8 対応率 P_c に応じた学習後の分類性能の100回の平均値と標準偏差

施する．

§1 実験2-1(モダリティ間で正しく対応するデータが得られない場合): 基本設定, 特徴ベクトルおよび評価方法

実験1と同様に, 3つのモダリティの内, いずれか1つのモダリティのデータが他の2つのモダリティのデータと正しく対応しない状況を想定する．ただし, 対応率 P_c の確率で, 3つすべてのモダリティで同じカテゴリのデータ(以後, 全対応データと呼ぶ)が観測され, $1 - P_c$ で, いずれか2つのモダリティのみ同じカテゴリのデータが観測され, 残り1つのモダリティには別のカテゴリのデータが観測される(以後, 部分対応データと呼ぶ)．

簡単のため, カテゴリ数 K を6, 特徴の種類数 W^m をすべてのモダリティで同じ6とする．特徴は, 各カテゴリに対応する K 個の分布から生成されるとし, 1ステップにおいてその分布から生成された100点の特徴を特徴ベクトルとする．理想的な観測状況においては, 1ステップで観測される特徴ベクトルは各カテゴリにおいて排他的であるとした．すなわち, 特徴は, 図7左のような, 一次元の排他的な分布に従うと仮定する．ただし, 観測ノイズを考慮し, それぞれ中心位置の異なる正規ノイズを加えた分布から生成されるものとする(図7右)．

1ステップ毎に, ロボットが K 個のうちのいずれかのカテゴリに注目するか決定される．そして対応率に応じて, 人がいずれのカテゴリの物体を注視するのか, またどのカテゴリの物体の名称を発するのか, がランダムに選択される．各モダリティでは, これらのカテゴリに対応する観測特徴量の分布から生成されたデータが観測される．一つの物体について100ステップ観測するものとし, 語意学習の総ステップ数 M は600とする．モデルのハイパーパラメータは, 実験1と同じ, $\alpha^m = 1.0$, $\pi^m = 1.0 (m \in \{v_s, v_o, s\})$, $\sigma = 0.5$ とした．

提案手法の有効性は, 前節と同様の方法で評価する．ただしテストデータとして, カテゴリ毎に, 各モダリティの特徴ベクトルが従う分布から新たにデータを100個ずつ生成したものをを用いる．

(a) 主観的整合性を用いないで学習した場合

(b) 主観的整合性を用いて学習した場合

図9 $P_c = 1.0$ の場合の形成したカテゴリ毎の特徴量分布の一例

§2 実験結果(2-1)

対応率 P_c を0.0から1.0まで0.1ずつ変化させ, 600ステップの語意学習シミュレーションを100回実施した．図8は, P_c 毎の, 学習後の分類性能の100回の平均値と標準偏差を示している．図8より, $P_c = 1.0$ のとき, すなわち, 全対応データのみが観測される場合では, 提案手法と従来手法の両方で, 高い分類性能が得られていることがわかる．しかしながら, 対応率が低くなるにつれ, 従来手法では $P_c = 0.8$ 付近から分類性能が低くなっているのに対して, 提案手法では $P_c = 0.3$ 付近まで高い分類性能が維持されており, 提案手法の有効性が確認できる．なおこの実験で, $P_c = 0.0$ であっても比較的高い分類性能が得られているのは, $P_c = 0.0$ であっても, 毎回, いずれか二つのモダリティでは対応するデータが観測されるためであり, このような場合でも, 対応しないモダリティのデータを排除することにある程度成功していることを示している．

図9は, $P_c = 1.0$ の場合のモダリティ v_s について推定された特徴量の分布 $\hat{\beta}_k^{v_s}$ の一例を示している．従来手法では, カテゴリの形成において, 裾野の広い分布が推定されている(図9(a))のに対し, 提案手法では, 裾野の狭い分布が推定されている(図9(b))ことが分かる．これは, 従来手法では, 観測されるノイズをそのまま含めた形で, 特徴量分布を推定しているのに対し, 提案手法では,

(a) 主観的整合性を用いないで学習した場合

(b) 主観的整合性を用いて学習した場合

図 10 $P_c = 0.5$ で (a) 主観的整合性を用いないで学習した場合及び (b) 主観的整合性を用いて学習した場合にサンプリングしたカテゴリの遷移

そのようなノイズを排除した真の分布 (図 7 左) に近い分布を推定できていることを示している。これは提案手法により、観測データに一樣に含まれるノイズが、特定のカテゴリとは対応しないデータ、すなわち非対応データとして扱われ、抑制された結果であると考えられる。

$P_c = 0.5$ のときの、ギブスサンプリング時にサンプリングされた潜在変数 (カテゴリ) の遷移を図 10 に示す。図 10 は、各ステップにおいて得られた観測特徴の下での、100 回のギブスサンプリングを経て得られた潜在変数の適切さの遷移を表す。図中の白抜き逆三角形は、各ステップでギブスサンプリングの収束結果として、正しいカテゴリがサンプリングできた割合を、白抜き四角形は、対応しないカテゴリをサンプリングした割合について、100 回実施したシミュレーションにおける平均と分散を表す。サンプリングされた潜在変数が、正しいカテゴリを表すものであるかどうかは、学習終了時に、4.1.3 節の分類性能の評価方法に基づいて、潜在変数 k' と真のカテゴリ k の対応を特定し、それを遡って適用することで得た。従来手法 (主観的整合性を用いないで学習) の場合 (図 10 (a))、正しいカテゴリのサンプリングが学習途中から、ある程度できるようになっていくが、対応しないカテゴリのサンプリングは学習終了時まで、排除できていないことが分かる。一方、提案手法 (主観的整合

図 11 対応率 P_c に応じた最終的な主観的整合性

性を用いた学習) の場合 (図 10 (b)) では、より高い頻度で、正しいカテゴリをサンプリングできるようになっていっていることが分かる。語意学習は基本的に、観測特徴をサンプリングされた潜在変数に割り付ける形で進むため、提案手法の狙い通り、他のモダリティとは対応しない、すなわち整合性の低いデータが、語意学習に反映されないようになっていくことで、高い分類性能の維持が実現されていると言える。

異なる対応率の条件においても、提案手法により図 10 に示したようなサンプリングのバイアスが可能であるかを評価するため、対応率が 1 以外の条件において、学習終了時の部分対応データに対する主観的整合性 λ (式 (5)) の平均値と標準偏差を図 11 に示す。図中の白抜き四角形は、部分対応データの中で対応するモダリティのデータ (以後、対応情報と呼ぶ) に対する主観的整合性を表し、白抜きの逆三角形は、対応しないモダリティのデータ (以後、非対応情報と呼ぶ) に対するものを表す。図 11 より、対応率が低くなると ($P_c < 0.3$)、その差は小さくなっていくものの、非対応情報に対する主観的整合性は、対応情報に対する主観的整合性に比べ低い値となっていることが分かる。すなわち、提案手法により整合性の高い情報が学習により貢献し、主観的整合性で適切に非対応情報を抑制できていることを示している。

§3 実験 2-2 (特定のモダリティのみ他のモダリティと対応しない場合): 基本設定、特徴ベクトルおよび評価方法

3 つのモダリティの内、特定の 1 つのモダリティのみにおいて、他の 2 つのモダリティと異なるカテゴリの観測特徴量が生成することがありうるとする。ここで想定しているのは、人がロボットに対して教示を行う、あるいはロボットの語意学習にとって正しい教示データとなる振舞をする (人の注視先と発話した名称は対応する) が、ロボットが何に注目しているかにはあまり関心がない状況である。すなわち、人が示す特徴が、ロボットが注目する特徴と必ずしも一致しないような状況であり、4.1.1 節で説明した非対応データの想定のうち (1) が特に起こり易いと想定できる場合である。そこで、注目率と呼ぶ

(a) モダリティ v_o と s のデータを用いた場合

(b) モダリティ v_s のデータを用いた場合

図 12 対応率 P_c に応じた学習後の分類性能の 100 回の平均値と標準偏差

確率 P_a を導入し、確率 P_a で全てのモダリティの観測が同一のカテゴリの特徴量分布から生成され、確率 $1 - P_a$ で、モダリティ v_o と s の観測は同じカテゴリの特徴量分布から生成されるが、 v_s の観測は異なるカテゴリから生成されるとする。提案手法は、このように部分対応データが偏る場合においても有効に作用し、非対応データの影響を低減した語意学習が可能であると考えられる。

特徴ベクトル、その他各種パラメータは実験 2-1 と同じものを用い、提案手法の有効性も同様の方法で評価する。

§4 実験結果 (2-2)

注目率 P_a を 0.0 から 1.0 まで 0.1 ずつ変化させ、600 ステップの語意学習シミュレーションを 10 回実施した。非対応データが観測されるモダリティとそれ以外のモダリティの学習を比較するために、テストデータに対して、学習後にモダリティ v_o と s を用いて分類した場合 (図 12 (a))、と v_s を用いて分類した場合 (図 12 (b)) の分類性能を見た。モダリティ v_o と s を用いた分類においては、従来手法 (図 12 (a) 黒丸) では、注目率が低くなるにつれ分類性能が低下しているのに対し、提案手法 (図 12 (a) 白丸) では、高い分類性能が維持できていることが分かる。また、図 12 (b) から、モダリティ v_s を用いた分類においても、従来手法 (黒丸) に比べ、提案手法 (白丸) が、注目率が 0.5 付近以上のある程度高い状況では、高い分類性能を維持できていることが分かる。しか

図 13 $P_c = 0.5$ の場合のモダリティ毎のデータに対する主観的整合性の遷移

し、注目率 P_a が 0.5 付近以下に低くなると、提案手法を用いても、モダリティ v_s の分類性能は低下していく。これは本実験設定では、モダリティ v_s に他のモダリティ v_o と s と対応するデータが観測される確率は $1 - P_a$ であるため、提案手法により、モダリティ v_s のデータが、学習からほとんど排除されてしまうことによるものと考えられる。

$P_a = 0.5$ のときの、ギブスサンプリング時に提案手法によって計算された各モダリティの主観的整合性の遷移を図 13 に示す。図中の白抜き逆三角形と白抜き菱形はそれぞれモダリティ v_o と s についての主観的整合性であり、ほぼ重なったまま学習初期に上昇したのち、その値を維持し、0.4 弱に収束していることが分かる。これに対し、図中の白抜き四角形はモダリティ v_s についての主観的整合性であり、学習初期から下がり始め、0.2 強程度に収束していることが分かる。このように提案手法では、ある特定のモダリティのみが他のモダリティと対応する確率が低い場合においても、対応しないモダリティの観測を学習に反映させる程度を主観的整合性に基づいて低減させることで、部分的に対応する観測についての学習を阻害せず、データ中の対応するモダリティの観測を効率的に利用した学習が可能になっていると考えられる。

5. 議 論

5.1 ロボットの語彙学習手法としての位置づけ

ロボットの言語学習における課題として、岩橋は、記号接地、パースペクティブ、信念共有、同時協調的行為、発話交代の 5 つをあげている [岩橋 12]。本研究で扱った、人とロボットが必ずしも同じ物体に注目するとは限らない状況におけるロボットの語意学習は、どのように言葉と事物を対応付けるかの記号接地と、どのように他者と共通の知識・信念を持つかの信念共有の課題に関連すると考えられる。

記号接地に関して、これまでいくつか研究されており、1 章で述べたマルチモーダル情報の共起確率に基づいた手法や [Roy 02, Yu 05, 田口 10, 中村 10]、事前に想定さ

れたバイアスを利用する手法 [田口 07, 菊池 08, 中野 09], が提案されている. これに対して本研究は, マルチモーダル情報を一つの語意に接地している点や主観の整合性という一種のバイアスを学習に利用している点で類似している. しかし, 本研究では, 問題として必ずしも正しく対応するマルチモーダル情報が共起して観測されとは限らない状況を取り扱う点, また, 形状類似性や顕著性などの単一のモダリティで規定されるバイアスではなく, 整合性という複数のモダリティで規定されるバイアスを導入している点で, 異なると考えられる.

信念共有に関して, Steels は, 複数のエージェントが自分自身の語意知識に基づいて物体への名前付けをし合うことで他者の知識との違いを知り, それを共通化する手法を提案している [Steels 12]. また, 岩橋は, ユーザとロボットがどの程度信念を共有しているかを表現した確信度関数を導入し, ロボットがそれに基づいて発話することで, ユーザとロボット間で相互適応的に信念を一致させる手法を提案している [岩橋 11]. 本研究で提案した手法でも, 主観の整合性を導入し, 他者の知識との一種の違いを推定している点では同じであるが, 整合性の高い情報を他者と共通化すべき知識として積極的に学習に利用している点が特徴である. しかし, 提案手法は, 上記の学習手法を否定するものではなく, それらの手法とは相補的に機能しうると考えられ, これらと組み合わせた取り組みを行うことは今後の課題である.

5.2 提案手法のスケラビリティ

§1 語意カテゴリ数に関して

本研究で実施した実験では, 語意カテゴリの数は, 予め既知としてた. しかし, ロボットが学習すべき語意の数は, 環境やユーザの要求に応じて異なることが予想されるため, 自動で決定されることが望まれる. これに関して, 階層ディリクレ過程 [Teh 06] を用いたノンパラメトリック化の手法が既に提案されている [中村 13]. 本研究で提案した主観の整合性とこのようなノンパラメトリック化の手法を組み合わせた取り組みを行うことが今後の課題である.

§2 モダリティ数に関して

本研究では, 3 つのマルチモーダル情報から語意を学習することを想定していたが, ロボットが学習すべき語意数の増加に伴って, 必要なモダリティ数も増加すると考えられる. 例えば, 「赤い」という語意の学習には, 形状のみではなく色の情報が必要であり, 「置く」などの動作を表す語意の学習には, 運動の系列情報が必要であると考えられる. 学習に必要なモダリティが増えると, ある語意を表す正しい対応関係はより部分的にしか観測されなくなると予想され, 提案手法はそのような状況において, より有効に働くと考えられる. 今後は, 多くのモダリティを必要とする問題設定において, 提案手法の有効性を詳細に検証していくことが必要である.

§3 学習対象の複雑さに関して

本研究の実験において, 提案手法により, 観測データに一樣に含まれるノイズのような特定のカテゴリとは対応しないデータを抑制してカテゴリを形成できることを確認した. これは提案手法により, 特定の学習対象 (カテゴリ) に対して排他的に対応する特徴の抽出が可能であることを示しており, カテゴリ弁別における有効性が期待できる. しかし学習対象が増える, あるいは個々の学習対象が様々な特徴を持つなどすると, あるカテゴリの特徴が, 多くのカテゴリに含まれない場合でも, 別の特定のカテゴリに含まれてしまうような場合が生じやすくなり, 排他性を重視しすぎることによって, そのような特徴が抽出されにくくなる恐れがある. そのような場合には, 提案モデルにおける整合性に対する逆感度パラメータ σ を調整することを検討すべきである. すなわち σ を少し高い値に設定することで, 観測データに一樣に含まれるノイズのように, 整合性がかなり低い情報の抑制効果は保持したまま, 一部のカテゴリ間のみで共有される特徴のように, ある程度の整合性を示す情報の利用が可能になると考えられる. この学習対象の複雑さに応じて逆感度パラメータを動的に調整する機能については今後の課題である.

§4 対応率に関して

本稿では, あるモダリティのデータが他のモダリティのデータと正しく対応しない状況を, 相互作用を通して一定の対応率でマルチモーダル情報の正しい対応が観測される状況と想定して実験した. しかし, 実環境においてロボットに依るユーザは, ロボットとの相互作用の履歴や自身の状態に応じて教示の仕方を変えうるものであり, その場合対応率は変遷していくと考えられる. そのため, 今後は, 実世界で人とロボットの相互作用を観察する実験を実施し, 実際のユーザがどのように (どのような対応率で) ロボットに回答しうるかを考慮して, 提案手法の有効性を検証する必要がある. しかし, 極端に対応率が低くなってしまった場合などは, 提案手法のみでは学習が困難になる可能性があり, そのような場合には, 逐次的な学習や相互作用の履歴から環境側へ働きかける仕組みが必要になると考えられる. これに対して, 追加的なギブスサンプリング [Banerjee 07] を用いたオンライン化の手法 [Araki 12] や教示者への能動的な発話手法 [中村 09, Sugiura 11] が利用できると考えられ, それら手法と提案手法を組み合わせた手法を開発することが今後の課題の一つである. 本実験では, 人とロボットの注意が共有されないことの影響およびそれに対する提案手法の有効性の検証に主眼を置くため, 音声情報は簡易的に人の発話音声から抽出した MFCC を特徴量として取り扱っていた. これはロボットが発話に変換可能な形式ではないため, このままでは上述のロボットの発話手法とそのまま組み合わせられない. しかしこれは, 音声認識器の認識結果を特徴量とすることで, 対応可能である

と考えられる．文書分類の手法と同様に，複数の単語からなる人の発話の音声認識の結果は，単語のカウントベクトルとして表現し，多項分布によってモデル化できる．提案手法では，入力変数は多項分布に従うと仮定しているため，この拡張に対応可能である．また，音声情報が単語で表現されるため，単語ごとに予めロボットの音声を登録しておくことで，学習結果をそのままロボットの発話に利用することができると考えられる．

6. お わ り に

本研究では，ユーザとロボットが必ずしも同じ事物に注目しているとは限らない状況においてロボットの自律的な語意学習を可能とする主観的整合性に基づいたマルチモーダルカテゴリゼーション手法を提案した．実ロボットを用いた実験結果から，提案手法を用いることで対応率に対して頑健に語意を学習できることを確認した．この結果は，提案手法により，人がロボットに語意を教示する際の制限（常に教示的に振る舞わなければならない等）が緩和可能であることを示唆する．また，人工データをを用いた実験結果から，提案手法を用いることで，対応するデータの主観的整合性を，他のモダリティとは対応しないデータに比べて高く計算することで，すなわち，部分的な対応関係を重視することで，より高い頻度で正しいカテゴリのサンプリングが可能になることによるものであると分析された．

ロボットが，今後より多くの生活場面で活躍していくには，人からの様々な要求に応えられる必要があり，それに応じてロボットの学習対象は多様化していくと考えられる．ロボットがセンシングできるデータが増えるほど，様々な対象を扱えうると期待されるが，学習対象とは関係のないデータも増加すると考えられる．提案手法は，そのような場合でも整合性を指標にしてより関係のあるデータのみに基づいた効率的な学習を可能にすると考えられる．今後は，前節で述べた課題に取り組み手法を精練していくことで，実世界で活躍するロボットの実現に近づくと期待される．

謝 辞

本研究はJSPS 科研費 24000012 ,24119003 ,24680022 の助成を受けた．

◇ 参 考 文 献 ◇

[Araki 12] Araki, T., Nakamura, T., Nagai, T., Funakoshi, K., Nakano, M., and Iwahashi, N.: Online Object Categorization Using Multimodal Information Autonomously Acquired by a Mobile Robot, *Advanced Robotics*, Vol. 26, No. 17, pp. 1995–2020 (2012)

[Banerjee 07] Banerjee, A. and Basu, S.: Topic Models over Text Streams: A Study of Batch and Online Unsupervised Learning, in *Proceedings of the 2007 SIAM International Conference on Data Mining* (2007)

[Boersma 14] Boersma, P. and Weenink, D.: Praat: doing phonetics by computer [Computer program], <http://www.praat.org/> (2014)

[Gorin 91] Gorin, A., Levinson, S., Gertner, A., and Goldman, E.: Adaptive acquisition of language, *Computer Speech and Language*, Vol. 5, No. 2, pp. 101–132 (1991)

[Gorin 94] Gorin, A., Levinson, S., and Sankar, A.: An experiment in spoken language acquisition, *IEEE Transactions on Speech and Audio Processing*, Vol. 2, No. 1, pp. 224–240 (1994)

[Griffiths 04] Griffiths, T. L. and Steyvers, M.: Finding scientific topics, *Proceedings of the National Academy of Sciences*, Vol. 101, No. 1, pp. 5228–5235 (2004)

[Roy 02] Roy, D. K. and Pentland, A. P.: Learning words from sights and sounds: a computational model, *Cognitive Science*, Vol. 26, No. 1, pp. 113–146 (2002)

[Steels 12] Steels, L.: Grounding Language through Evolutionary Languages Games, *Language Grounding in Robots*, pp. 1–22 (2012)

[Sugiura 11] Sugiura, K., Iwahashi, N., Kawai, H., and Nakamura, S.: Situated spoken dialogue with robots using active learning, *Advanced Robotics*, Vol. 25, No. 17, pp. 2207–2232 (2011)

[Teh 06] Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M.: Hierarchical Dirichlet Processes, *Journal of the American Statistical Association*, Vol. 101, No. 476, pp. 1566–1581 (2006)

[Vedaldi 08] Vedaldi, A. and Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms, <http://www.vlfeat.org/> (2008)

[Yu 05] Yu, C.: The emergence of links between lexical acquisition and object categorization: a computational study, *Connection Science*, Vol. 17, No. 3–4, pp. 381–397 (2005)

[岩橋 11] 岩橋 直人: 人とロボットの言語コミュニケーションにおける間主観性, 人工知能学会誌, Vol. 26, No. 4, pp. 352–359 (2011)

[岩橋 12] 岩橋 直人: ロボットと言語一言語コミュニケーション能力の機械学習, 人工知能学会誌, Vol. 27, No. 6, pp. 563–568 (2012)

[菊池 08] 菊池 匡晃, 荻野 正樹, 浅田 稔: 顕著性に基づくロボットの能動的語彙獲得, 日本ロボット学会誌, Vol. 26, No. 3, pp. 261–270 (2008)

[笹本 13] 笹本 勇輝, 吉川 雄一郎, 浅田 稔: 音声模倣と語彙獲得の共発達のための主観的整合機構に基づく対応学習, 日本ロボット学会誌, Vol. 31, No. 1, pp. 71–82 (2013)

[小野 07] 小野 泰弘, 岡部 孝弘, 佐藤 洋一: 低解像度画像からの視線方向推定: カーネル多重線形モデルによる個人差への対応, 電子情報通信学会論文誌, Vol. J90-D, No. 8, pp. 2212–2222 (2007)

[小野 09] 小野 広司, 左 祥, 伊丹 英樹, 尾関 基行, 岡 夏樹: 最終行動ヒューリスティクスを用いた状況推定による自由発話音声データからの語句意味学習, HAI シンポジウム 2009, pp. 2B–2 (2009)

[中村 09] 中村 慎也, 岩橋 直人, 長井 隆行: 実世界における人とロボットの共有信念の推定に基づいた適応的な発話生成, 知能と情報 (日本知能情報ファジィ学会誌), Vol. 21, No. 5, pp. 663–682 (2009)

[中村 10] 中村 友昭, 長井 隆行, 岩橋 直人: Gibbs Sampling による物体のマルチモーダルカテゴリゼーション, 第 28 回日本ロボット学会学術講演会, pp. 211–3 (2010)

[中村 13] 中村 友昭, 荒木 孝弥, 長井 隆行, 岩橋 直人: 階層ディリクレ過程に基づくロボットによる物体のマルチモーダルカテゴリゼーション, 計測自動制御学会論文集, Vol. 49, No. 4, pp. 469–478 (2013)

[中谷 11] 中谷 仁, 植村 竜也, 荒木 修, 西垣 貴央, 尾関 基行, 岡 夏樹: ロボットの語意獲得のためのユーザの発話分類, 第 25 回人工知能学会全国大会予稿集, pp. 3B1–OS22b–4 (2011)

[中野 09] 中野 史, 吉川 雄一郎, 浅田 稔, 石黒 浩: 相互排他性に基づくマルチモーダル共同注意, 日本ロボット学会誌, Vol. 27, No. 7, pp. 814–822 (2009)

[田口 07] 田口 亮, 木村 優志, 小玉 智志, 篠原 修二, 入部 百合絵, 桂田 浩一, 新田 恒雄: 幼児の学習バイアスを利用したエージェントによる語意学習の効率化, 人工知能学会論文誌, Vol. 22, No. 4, pp. 444–453 (2007)

[田口 10] 田口 亮, 岩橋 直人, 船越 孝太郎, 中野 幹夫, 能勢 隆, 新田 恒雄: 統計的モデル選択に基づいた連続音声からの語意学習, 人工知能学会論文誌, Vol. 25, No. 4, pp. 549–559 (2010)

〔担当委員：山川 宏〕

2014 年 4 月 8 日 受理

著 者 紹 介

笹本 勇輝 (Yuki Sasamoto)

2010 年大阪大学大学院工学研究科知能・機能創成工学専攻博士前期課程修了。同年同大学博士後期課程に入学し現在に至る。認知発達ロボティクスの研究に従事。

吉川 雄一郎 (Yuichiro Yoshikawa)

2005 年大阪大学大学院工学研究科知能・機能創成工学専攻終了。博士 (工学)。同年 ATR 知能ロボティクス研究所研究員。2006 年 JST ERATO 浅田共創知能システムプロジェクト研究員。2010 年より大阪大学大学院基礎工学研究科講師。2013 年同准教授となり現在に至る。認知発達ロボティクスの研究に従事。

浅田 稔 (Minoru Asada) (正会員)

1982 年 大阪大学大学院基礎工学研究科修了工学博士 (大阪大学)。1995 年大阪大学工学部教授, 1997 年同大学院工学研究科知能・機能創成工学専攻教授となり現在に至る。この間 1 年間米国メリーランド大学客員研究員。認知発達ロボティクスを提唱し, その研究に従事。ロボカップ国際委員会プレジデント (2002-2008 年) を経て, 現在, ロボカップフエデレーションジャパンプレジデント。現在, 日本学術振興会科学研究費特別推進研究研究代表 (2012-2017 年)。