

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

Efficient Learning of Word Meanings by Agents Using Biases Observed in Language Development of Children

田口 亮
Ryo Taguchi

豊橋技術科学大学 大学院工学研究科
Graduate School of Engineering, Toyohashi University of Technology
taguchi@vox.tutkie.tut.ac.jp, <http://www.vox.tutkie.tut.ac.jp/~taguchi/>

木村 優志
Masashi Kimura

(同 上)
kimura@vox.tutkie.tut.ac.jp, <http://www.vox.tutkie.tut.ac.jp/~kimura/>

小玉 智志^{*1}
Satoshi Kodama

(同 上)

篠原 修二
Shuji Shinohara

(同 上)
shinohara@vox.tutkie.tut.ac.jp

入部 百合絵
Yurie Iribe

豊橋技術科学大学 情報メディア基盤センター
Graduate School of Engineering, Toyohashi University of Technology
iribe@imc.tut.ac.jp, <http://www.imc.tut.ac.jp/~iribe/>

桂田 浩一
Kouichi Katsurada

豊橋技術科学大学 大学院工学研究科
Graduate School of Engineering, Toyohashi University of Technology
katsurada@tutkie.tut.ac.jp, <http://www.vox.tutkie.tut.ac.jp/~katsurada/>

新田 恒雄
Tsuneo Nitta

(同 上)
nitta@tutkie.tut.ac.jp, <http://www.vox.tutkie.tut.ac.jp/~nitta>

keywords: symbol grounding, word learning, mutual exclusivity bias, shape bias

Summary

Recently, studies on learning of word meanings by agents have begun. In these studies, a human shows objects to an agent and utters words such as “red” or “box”. The agent finds out object’s feature represented by each spoken word. In our method, firstly, the agent learns probability distribution $p(x)$ and conditional probability distribution $p(x|w)$, where x is an object feature and w is a word. If a word w does not represent a feature x , $p(x)$ and $p(x|w)$ will be almost same distribution because x is independent of w . This fact enables the agent to use distance between $p(x)$ and $p(x|w)$ when inferring which feature the word represents. Previous works also employ similar stochastic approaches to detect the feature. However, such approaches need a lot of examples to learn correct distributions.

To overcome this problem, we apply two types of biases: shape bias and mutual exclusivity bias that are observed in children’s language development. When a child hears a novel word about an object, he/she often applies the word to other objects similar in shape. This tendency is called the shape bias. This bias works effectively for learning word meanings, because not a few words in the real world do not represent its color and material but represent its shape. In order to implement this bias into the agent, we formulate it as the variable that reduces the distances of non-shape features. Therefore, when a novel word is taught, the agent with the bias decides that the word represents shape feature and it can apply the word to other objects similar in shape. On the other hand, if a child already knows some words about an object, he/she often seeks the meaning of a novel word outside the meanings of known words. This tendency is called the mutual exclusivity bias. This bias is formulated as the variable that reduces the distances associated with features represented by known words. For example, if the agent already knows a word representing a shape, it decides that other words do not represent the same shape by reducing the distance associated with the shape. Experimental results show that proposed method with biases can acquire word meanings more efficiently than the traditional stochastic only approach.

1. ま え が き

IT 技術の進展と普及に伴うサービスの多様化とシステムの複雑化から、自然対話によるインタフェース技術への社会的要請が芽生えつつある。現在の人間-機械システムでは一般に、対話に利用する背景知識を、辞書や対話シナリオの形で与える必要がある。しかし、こうしたアプローチでは実世界の事物・事象に関する膨大な知識を開発者が記述しなければならないため、対話に使用する知識をシステム自らが学習・獲得していく機構の実現が望まれる [赤穂 97]。

近年、上記の課題に対して、「人間-エージェント-実世界の事物・事象」の三項関係を基に、ラベルの意味（語意）をエージェント自身の感覚・運動系の信号と対応付けて獲得させる、新しい語意獲得研究が行われ始めている [中川 95, 赤穂 97, 金 00, Roy 02, Iwahashi 03]。これらの研究では、新たな知識が獲得できるというだけでなく、語意をセンサ信号の特徴パターンとして学習するため、実世界に接地した知識を獲得することができる。

本論文では、画像オブジェクトをエージェントに提示し、そのオブジェクトに関連したラベルを教示し、語意を獲得させる課題を扱う。教示するラベルは、オブジェクトが持つ属性（色や形など）を説明したラベル（「赤」や「丸」など）である。ただし、そのラベルがどの属性を説明しているかは与えない。従ってエージェントは、そのラベルが対象とする属性を自ら導出しなければならない。この課題は、多くの先行研究で扱われている基本的な課題の一つであるが、その効率性については十分な議論がなされていない。従来研究では、ラベルが対象とする属性の導出を確率計算にのみ頼るため、語意の獲得に多くの事例を必要とした。

現実世界で語意を論理的に推論するためには膨大な事例が必要となることは、哲学者 Quine によっても指摘されている [Quine 60]。「それにも関わらず人間の子供はどうしてラベルを数回聞いただけでその意味がわかるのか？」という問いに対して、発達心理学の分野では「幼児は意味解釈の範囲を制限するようなバイアスを持っている」という仮説が提案されている [今井 97]。

我々は効率的な語意学習を実現するには、確率計算だけではなく、幼児に見られるようなバイアスを利用する必要があると考えている。本論文では、幼児に見られるバイアスのうち、相互排他性バイアス [Markman 89] と形状類似バイアス [Landau 88] をエージェントに実装する手法を提案する。

2. 問 題 設 定

前章で述べた問題を以下のように定式化する。

オブジェクトは色や形など $|I|$ 種類の属性集合 I を持っており、教示されるラベルはそのオブジェクトが持つ一つ以上の属性を表すものとする。ラベルが表す属性を対象属性と呼ぶ。属性の数 $|I|$ は既知とするが、各ラベルの対象属性は未知とする。

4 章で詳しく述べるように、実験では $I = \{\text{形状, 色相, 明度}\}$ の 3 属性を考える。例えば、色相を対象属性とするラベルは「赤」や「青」などに相当し、形状と色相を対象属性とするラベルは、赤く丸いオブジェクトを「梅干」と呼ぶことに相当する。

対象属性の組み合わせは $2^{|I|} - 1$ 通り存在するため、一つのオブジェクトに対し、最大で $2^{|I|} - 1$ 種類のラベルを教示することができる。ただし、一度に教示できるラベルはその中の一つとする。

与えられた学習サンプルを元に、新たに提示されたオブジェクトに対して、その属性を説明する $2^{|I|} - 1$ 種類のラベル集合を出力することが目的である。

先行研究ではラベルを音声で教示するが、本論文では認識誤りなどによる実験の煩雑化を避けるため、キーボードにより入力する。従って、先行研究で扱っている音声の切り出しや、音声間のマッチングの問題は対象としない。

3. 提 案 手 法

3.1 確率による語意学習法

エージェントはオブジェクトの画像から、属性毎に特徴抽出を行う。具体的な特徴抽出方法については、4 章の実験条件で述べる。ここでは、一般的な議論をするため、抽出された特徴を

$$\mathbf{X} = \{\mathbf{x}_i \mid i \in I\} \quad (1)$$

と表す。ただし、 i は属性、 \mathbf{x}_i は属性 i の特徴（ベクトル）を表す。

オブジェクトの属性を表現するラベルを出力するためには、各ラベルの対象属性を正しく判定すると共に、「どの色が青と呼べるのか？」といった各ラベルが表す特徴の範囲を学習しなければならない。この特徴の範囲は、ラベル w が与えられた場合の特徴 \mathbf{X} の確率分布 $p(\mathbf{X} \mid w)$ として学習できる [中川 95, Roy 02, 金 00, Iwahashi 03]。

ここで属性が互いに独立であることを仮定すると、

$$p(\mathbf{X}) = \prod_i p(\mathbf{x}_i) \quad (2)$$

$$p(\mathbf{X} \mid w) = \prod_i p(\mathbf{x}_i \mid w) \quad (3)$$

となる。従って、特徴 \mathbf{X} が与えられた場合の、ラベル w の確率は

$$\begin{aligned} p(w \mid \mathbf{X}) &= \frac{p(w)p(\mathbf{X} \mid w)}{p(\mathbf{X})} \\ &= p(w) \prod_i \frac{p(\mathbf{x}_i \mid w)}{p(\mathbf{x}_i)} \end{aligned} \quad (4)$$

†1 現在、日立情報システムズ

となる．式 (4) は全ての属性に着目して，ラベル w の出力確率を計算することを意味している．しかし，例えば，「青」というラベルの場合，オブジェクトの色相属性だけに着目して使用されるため，形状属性とは独立に計算できる．そこで，式 (4) を次のように変形する．

$$p(w | \mathbf{X}) = p(w) \prod_{i \in \mathbf{A}_w} \frac{p(\mathbf{x}_i | w)}{p(\mathbf{x}_i)} \quad (5)$$

ここで， \mathbf{A}_w は，ラベル w の対象属性の集合 ($\mathbf{A}_w \subseteq \mathbf{I}$ and $\mathbf{A}_w \neq \phi$) である．なお， $p(w)$ はラベル w の生起確率であるが，本論文では，教示回数に依存せずにラベルを出力させるため， $p(w)$ はラベルによらず一定と仮定する．

任意の属性集合 \mathbf{A} ($\mathbf{A} \subseteq \mathbf{I}$ and $\mathbf{A} \neq \phi$) を表現するのに最も適したラベル $\tilde{w}_\mathbf{A}$ は，次の式で導く．

$$\tilde{w}_\mathbf{A} = \arg \max_{w \in \mathbf{W}_\mathbf{A}} p(w | \mathbf{X}) \quad (6)$$

ただし， $\mathbf{W}_\mathbf{A}$ は \mathbf{A} を対象属性集合とするラベルの集合 ($\mathbf{W}_\mathbf{A} = \{w | \mathbf{A}_w = \mathbf{A}\}$) である．エージェントは，属性の組み合わせ毎に，最も尤度の大きいラベルを出力する．例えば，形状を対象属性とするラベルが 7 種類ある場合には，エージェントは提示されたオブジェクトに対し，形状を表現するのに最も適したラベルをその 7 種類の中から選択する．

ここで問題となるのは， $p(\mathbf{x}_i | w)$ および $p(\mathbf{x}_i)$ の学習方法と， \mathbf{A}_w の導出方法である．続く 3.1.1 節で特徴の範囲を学習する方法について説明し，3.1.2 節で対象属性の判定方法について述べる．

§ 1 $p(\mathbf{x}_i | w)$ と $p(\mathbf{x}_i)$ の学習

$p(\mathbf{x}_i | w)$ と $p(\mathbf{x}_i)$ のモデルには様々なものが考えられるが，ここでは，混合正規分布でモデル化することを考える．混合正規分布は，有限個の正規分布の重みつき線形和として表現される確率密度関数で，複雑な特徴量に対しても比較的柔軟に対応できる．

本研究では，混合正規分布のパラメータ（各正規分布の平均と分散，重み，および混合数）を Online-EM 法 [石井 98] を用いて推定する．Online-EM 法は，混合正規分布の学習で一般的に用いられる EM 法を逐次処理ができるように拡張した手法であり，本実験のように学習効率を問題とする場合には都合が良い．

§ 2 対象属性の判定

特徴 \mathbf{x}_i がラベル w に独立ならば，条件付確率の定義から， $p(\mathbf{x}_i | w) = p(\mathbf{x}_i)$ が成り立つ．例えば「青」というラベルの場合，ある特定の色相（青い色）のオブジェクトに対して教示されるため，図 1 に示すように， $p(\mathbf{x}_{\text{色相}})$ に比べ， $p(\mathbf{x}_{\text{色相}} | \text{青})$ は偏った分布となる．一方で，青い色であれば，オブジェクトの形に関係なく「青」と言えるため， $p(\mathbf{x}_{\text{形状}})$ と $p(\mathbf{x}_{\text{形状}} | \text{青})$ は等しくなる．従って， $p(\mathbf{x}_i | w) \neq p(\mathbf{x}_i)$ の場合に，属性 i はラベル w の対象属性であると判定することができる．

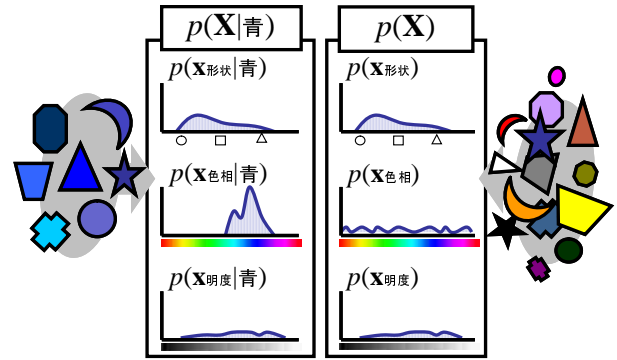


図 1 確率分布の学習例

$p(\mathbf{x}_i | w)$ をオンラインで学習することを考えると， $p(\mathbf{x}_i | w)$ と $p(\mathbf{x}_i)$ が完全に一致することはない．そこで， $p(\mathbf{x}_i | w)$ と $p(\mathbf{x}_i)$ の KL-divergence を計算し，閾値よりも大きい場合に対象属性と判定する*1．KL-divergence は確率分布間の情報損失を表しており，その値が大きいほど分布が異なっていることを意味する．

$p(\mathbf{x}_i | w)$ と $p(\mathbf{x}_i)$ の KL-divergence は

$$\begin{aligned} KL(p(\mathbf{x}_i | w) \| p(\mathbf{x}_i)) \\ = \int p(\mathbf{x}_i | w) \ln \frac{p(\mathbf{x}_i | w)}{p(\mathbf{x}_i)} \end{aligned} \quad (7)$$

となる．KL-divergence は非対称であるため，

$$\begin{aligned} KL_2(p(\mathbf{x}_i | w) \| p(\mathbf{x}_i)) \\ = KL(p(\mathbf{x}_i | w) \| p(\mathbf{x}_i)) KL(p(\mathbf{x}_i) \| p(\mathbf{x}_i | w)) \end{aligned} \quad (8)$$

として対称化する．対称化した KL-divergence は $0 \sim \infty$ の値を取るが，扱い易いようにシグモイド関数を使って $0 \sim 1$ に正規化する．

$$\begin{aligned} D(i, w) &= \frac{1}{1.0 + e^{-\alpha_i(d - \beta_i)}} \\ d &= KL_2(p(\mathbf{x}_i | w) \| p(\mathbf{x}_i)) \end{aligned} \quad (9)$$

この結果から確信度 $Conf(i, w)$ を計算する．確信度 $Conf(i, w)$ は属性 i がラベル w の対象属性か否かを判断する指標であり，次式により計算する．

$$Conf(i, w) = B(i, w) D(i, w) \quad (10)$$

ただし， $B(i, w)$ は 3.2 節で説明するバイアスである．

ラベル w の対象属性の集合 \mathbf{A}_w は次式で表される．

$$\mathbf{A}_w = \{i | Conf(i, w) > 0.5\} \quad (11)$$

$p(\mathbf{x}_i)$ が事前に学習されると仮定すると，各ラベルの学習初期においては，全ての属性において $D(i, w)$ が大きな値となるため， $\mathbf{A}_w = \mathbf{I}$ となる．そして， $p(\mathbf{x}_i | w)$ の学習が進むにつれて，無関係な属性が正しく判定できるようになり，正しい \mathbf{A}_w が導出される．

*1 [Roy 02] では， $KL_2(p(\mathbf{x}_i | w) \| p(\mathbf{x}_i | \bar{w}))$ から対象属性の判定を行っている（ \bar{w} は w の否定）．我々の手法はこれと同様の結果が得られるが，我々が用いている $p(\mathbf{x}_i)$ はラベルに独立なため，[Roy 02] の手法と比べ計算量やメモリ量が少なくなる．また， $p(\mathbf{x}_i)$ は語意学習に先立って学習することができる．

3.2 バイアスの定式化と学習機構への組み込み

前節では、学習した確率分布を利用した属性判定の手法を説明した。この手法では、無関係な属性を判定するためには、確率分布 $p(x_i | w)$ が $p(x_i)$ に近づくまで学習されなければならない。

語意を論理的な方法で学習するためには膨大な事例が必要となることは、哲学者 Quine によっても指摘されている [Quine 60]。これは、現実世界では語意として解釈可能な仮説が無数に存在することに起因する。しかし一方で、幼児は少ない学習回数で正確にラベルの意味を同定できることが知られている [Carey 78]。現在、その理由について、仮説空間を狭める役割を果たすバイアスを、幼児が持っているとして仮定する考えが提唱されている [今井 97]。例えば幼児は、名前の知っているオブジェクトと未知のオブジェクトがある状況で新奇なラベルを聞くと、そのラベルは未知のオブジェクトに付けられた名前であると解釈する。さらに、未知のオブジェクトがない場合には、そのオブジェクトに関する未知の属性（色や材質など）や部分（耳や取っ手など）の名前であると解釈する。この傾向性は相互排他性バイアスと呼ばれる [Markman 89]。また、Landau らは、オブジェクトの形状、サイズ、触感をコントロールした実験から、幼児はサイズや触感には着目せず、形状の類似したオブジェクトにラベルの適用範囲を拡張することを示し、この傾向性を形状類似バイアスと名づけた [Landau 88]。

これらのバイアスは、人間の言語の規則性や、学習した知識を利用することに相当するため、実環境で語意を学習するエージェントにおいても有効性が期待できる。本論文では、これら二つのバイアスをエージェントの学習機構に組み込み、語意獲得の効率を改善することを目指す。

本論文では相互排他性バイアス、形状類似バイアスの二つを、確信度 $Conf(i, w)$ を抑制するバイアス $B(i, w)$ ($0 \leq B(i, w) \leq 1$) として次式で定式化する。

$$B(i, w) = \min\{E(i, w), S(i, w)\} \quad (12)$$

ここで、 $E(i, w)$ は相互排他性バイアス ($0 \leq E(i, w) \leq 1$)、 $S(i, w)$ は形状類似バイアス ($0 \leq S(i, w) \leq 1$) である。例えば、 $B(i, w)$ が低いと $Conf(i, w)$ は抑制され、その結果、属性 i は対象属性とみなされ難くなる。以下 $E(i, w)$ および $S(i, w)$ の算出方法を述べる。

§1 相互排他性バイアス

本論文では、複数のオブジェクトを一度に提示する実験は扱わないため、未知のオブジェクトがない場合の相互排他性バイアス、すなわち、新奇なラベルを未知の属性ラベルと解釈する傾向性をエージェントに与える。

エージェントは、オブジェクトが提示されると、それまでの学習結果を利用してラベル集合 W' を出力する。 W' が対象とする属性の集合を

$$A_{W'} = \sum_{w' \in W'} A_{w'} \quad (13)$$

とすると、 W' が対象としていない属性集合 ($I - A_{W'}$) が、未知の属性集合となる。従って、新奇なラベル w が教示された場合に、 W' が対象とする属性 $i' \in A_{W'}$ の確信度 $Conf(i', w)$ を抑制することで、前述のような傾向性が再現できる。例えば、 $W' = \{w'\}$ であり、 w' が形状を対象属性とするラベルである場合には、 $Conf(\text{形状}, w)$ が 0 になるように抑制すれば、 w は形状以外の属性を対象としたラベルであると判定される。具体的な相互排他性バイアス $E(i, w)$ の計算方法を次式に示す。

$$E(i, w) = \begin{cases} 1, & \Theta_i \leq 0.5 \\ b_w, & \Theta_i > 0.5 \end{cases} \quad (14)$$

$$\Theta_i = \max_{w' \in W'} \left[\frac{Conf(i, w')}{1.0 + e^{-\lambda\{t(w') - \tau\}}} \right] \quad (15)$$

ただし、 b_w はバイアスの強さ (3.2.3 節参照)、 λ と τ はシグモイド関数のパラメータ、 $t(w')$ は w' の学習回数を表す。相互排他性バイアスが正しく機能するためには、 w' が十分に学習されていない場合、 w' の学習回数が少ない場合には、バイアスの計算に用いられないようシグモイド関数によって調整している。

また、バイアスを適用する機会を増やすため、初めてラベルが教示された時だけでなく、教示されたラベル w が出力ラベル集合 W' に含まれない場合に、式 (14) によりバイアスの再計算を行う。

§2 形状類似バイアス

形状類似バイアスを本実験の枠組みで言い換えれば、新奇なラベルの対象属性を形状とみなすことに相当する。そこで、

$$S(i, w) = \begin{cases} 1, & i = \text{形状} \\ b_w, & i \neq \text{形状} \end{cases} \quad (16)$$

を与えることで形状以外の属性を抑制させる。ただし、すでに相互排他性バイアスにより、形状属性が抑制されている場合には、その働きを妨げないように全ての属性 i に対して $S(i, w) = 1$ とする。

§3 バイアスの抑制

バイアスは教示手順次第では語意の獲得効率を低下させることがある。例えば、常に形状類似バイアスが適用されると、色や材質に関するラベルを正しく学習することができない。人間の場合、形状類似バイアスが強く現れるのは幼児期だけであり、加齢に伴ってその傾向性は弱まっていくことが報告されている。

そこで本手法では、ラベルの学習回数に応じてバイアスを抑制することを考える。具体的には以下の式によってバイアスの強さを調整する。

$$b_w = 1 - \gamma^{t(w)} \quad (17)$$

ただし、 γ は減衰率 ($0 \leq \gamma \leq 1$)、 $t(w)$ はラベル w の学習回数である。

また、バイアスによって対象属性 A_w が空集合になった場合には、全ての属性 i に対してバイアスを $B(i, w) = 1$

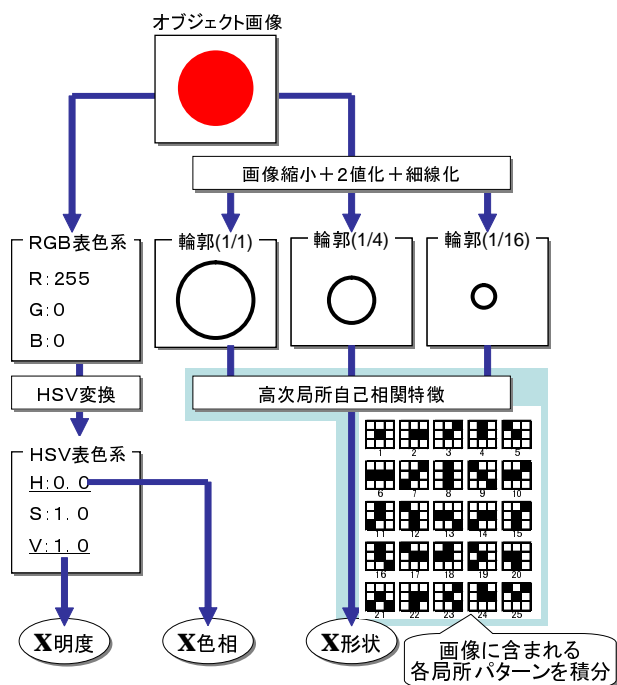


図 2 特徴抽出の概要

にリセットし無効化する．これは「対象属性を持たないラベルは無い」という仮定を与えることを意味する．

4. 語意獲得実験

4.1 実験条件

§1 画像オブジェクトと特徴抽出

エージェントは与えられた画像オブジェクトから形状、色相、明度の 3 属性 ($I = \{ \text{形状}, \text{色相}, \text{明度} \}$) の特徴を抽出する．図 2 に特徴抽出の概要を示す．

形状属性の特徴抽出では、まず画像を 2 値化した後、細線化することでオブジェクトの輪郭を抽出する．そして、この輪郭画像から 25 次元 \times 3 層 (1 層は元画像、残りの 2 層は元画像を 1/4, 1/16 に縮小した画像) の高次局所自己相関特徴 [栗田 97] を求める．この特徴は局所パターンに対応する画素値 (2 値画像の場合 0 または 1) の積を、画像全体で合計することによって得られる特徴量である．各次元は、0 ~ 画素数の値を取るが、各層の 1 次元目が画素数に相当するため、この値で、2 ~ 25 次元目の特徴を 0 ~ 1 に正規化した．従って、形状の特徴 $x_{\text{形状}}$ は 72 次元 (24 次元 \times 3 層) となる．

また、画像中で最も頻度の高かった RGB 値を HSV 座標系に変換し、変換後の H 成分を色相の特徴 $x_{\text{色相}}$ 、V 成分を明度の特徴 $x_{\text{明度}}$ 、として利用する．ただし、H 成分は 0 ~ 2π の値を取るが、これを 0 ~ 1 に正規化した．また、白、黒、灰色といった無彩色の場合、色相が求まらないため、0 ~ 1 のランダムな値を $x_{\text{色相}}$ に代入した．HSV 表色系へと変換することで、RGB 表色系よりも人間の語彙体系との親和性の高い特徴量となる．

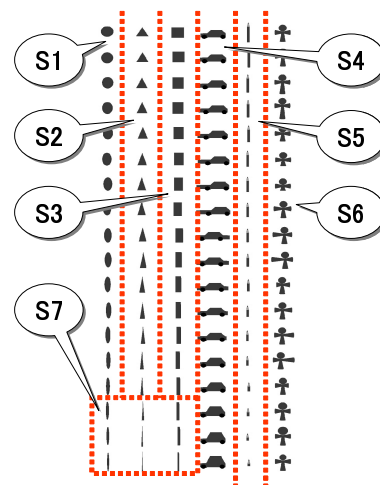


図 3 オブジェクトの形状パターンと対応するラベル

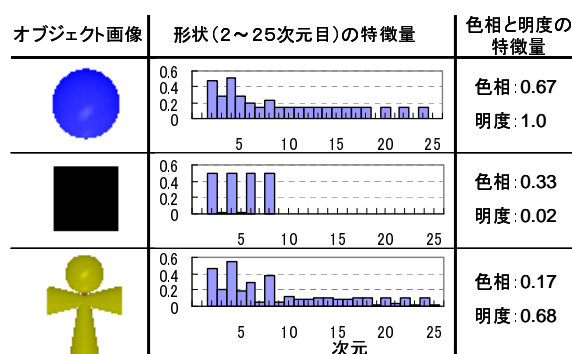


図 4 画像と抽出される特徴量の例

本実験で使用した画像オブジェクトは、3DCG のモデルをレンダリングし、フルカラーのビットマップ形式で保存したものを使用した．モデルの形状は、図 3 に示す 108 パターンを用意した．また、色相および明度は 0 ~ 1 の値域を等分割して 100 パターンを用意した．画像オブジェクトはそれらを独立に組み合わせて生成するため、計 1,080,000 種類 (形状 108 \times 色相 100 \times 明度 100) となる．

実験に用いた画像オブジェクトの具体例と、抽出される 3 種類の特徴量を図 4 に示す．形状の特徴は、縮小しない画像に対する 2 ~ 25 次元目の特徴を示す (各次元が表す局所パターンは図 2 を参照)．青い丸の形状特徴は多くの局所パターンが均等に現れるのに対して、黒い四角はほぼ直線成分 (2, 4, 6, 8 次元目) のみである．また、黄色い剣玉の形状特徴は、丸と似た特徴を示すが、皿の部分があるために直線成分が丸よりも多く現れている．エージェントはこのように、画像に含まれる局所パターンの割合の違いから形状を識別する．

§2 ラベル

実験 1, 2 では、3 属性のそれぞれを対象にしたラベルを 7 種類、計 21 種類を教示する．形状は、成人男性 12

名からのアンケートをもとに7クラスに分類し, “S1” から “S7” とラベルを付けた. 形状とラベルの対応を図3に示す. 色相および明度は, それぞれの特徴空間を7等分し, “H1” ~ “H7”, “L1” ~ “L7” とラベルを付けた. 実験1と2の違いは, ラベルを教示する順番であるが, 具体的な方法は次節で述べる.

実験3では, 対象属性の数が相互排他性バイアスに与える影響を調べる. 形状類似バイアスの効果は, 形状のラベルとその他のラベルとの比率によっても変化する. ここでは形状類似バイアスに関して, 実験1,2と同程度の効果が得られるようにするため, 色相と明度のラベルの数を減らし比率がほぼ等しくなるように調整する. 具体的には実験1,2と同様の形状ラベル7種類と, 色相と明度のラベルをそれぞれ3種類, および色相と明度の二つを対象属性とするラベル9種類の, 計22種類を教示する. 色相と明度のラベルに関しては特徴空間を3等分し, “H8” ~ “H10”, “L8” ~ “L10” とラベルを付けた. 色相と明度の二つを対象属性とするラベルは, “H8” ~ “H10”, “L8” ~ “L10” の組み合わせ (3^2 通り) を表現する. ラベル名称は “HL1” ~ “HL9” とした.

§3 実験手順

ラベルの学習に先立って, まずエージェントは $p(\mathbf{x}_i)$ を学習する. $p(\mathbf{x}_i)$ は, 1,080,000 通りのオブジェクトからランダムに選択した 5,000 個を用いて学習される.

バイアスの効果は, 学習したラベルの種類や数によって変化する. 例えば, 相互排他性バイアスは既に学習したラベルの知識を用いて, 新奇のラベルの対象属性を推論する. この様子をわかりやすく示すため, 本実験では, 前項で説明したラベルを一斉に教示するのではなく, 学習の序盤, 中盤, 終盤で教示するラベルセットを切り替える. これにより, 例えば学習の中盤では, 序盤に学習したラベルの知識を用いて, 新奇のラベルの対象属性を判定することになる. 具体的には, 実験1では, 21種類のラベルをランダムに3等分し, 7種類ずつのラベルセットを用意する. そして, 1つのラベルセットについて, 150回教示した後, 次のラベルセットに切り替える. 実験2では, 属性毎にラベルセットを分け, 形状, 明度, 色相の順でラベルセットを切り替える. 各ラベルセットの教示回数は150回である. 実験3では, 22種類のラベルをランダムに3分割 (8種類, 7種類, 7種類) し, 200回教示した後にラベルセットを切り替える.

学習効率を評価するため, 一つのラベルを教示した後に100種類の画像オブジェクトをランダムに提示し, ラベル正解精度を確認する. ラベル正解精度は次式で計算する.

$$(C_N - C_S - C_D - C_I) / C_N \quad (18)$$

正解精度は, ラベルのシーケンスを評価する尺度であるが, ここでは, 対象属性をシーケンスと見なして評価する. 具体的には, 正解数 C_N は提示したオブジェクト

表1 実験パラメータ

α 形状	2	β 形状	1
α 色相	0.05	β 色相	15
α 明度	0.05	β 明度	15
λ	0.5	τ	20

表2 実験1で得られた $p(\mathbf{x}_i)$ と $p(\mathbf{x}_i|w)$ の混合数

	$p(\mathbf{x}_{\text{形状}})$	$p(\mathbf{x}_{\text{色相}})$	$p(\mathbf{x}_{\text{明度}})$
ラベルなし	12	170	170
	$p(\mathbf{x}_{\text{形状}} w)$	$p(\mathbf{x}_{\text{色相}} w)$	$p(\mathbf{x}_{\text{明度}} w)$
形状ラベル	1.7	15	15
色相ラベル	3.2	3.6	15
明度ラベル	3.2	15	3.6

に対する本来の正解ラベル数, 置換誤り数 C_S は出力すべき対象属性で異なるラベルが出力された数 (例えば色相を対象に “H1” を出力すべきところを “H2” と出力した場合など), 脱落誤り数 C_D は出力すべき対象属性でラベルが出力されなかった数, 挿入誤り数 C_I は出力すべきでない対象属性で出力されたラベルの数 (例えば全ての属性を対象としてラベルを出力した場合など) とした.

§4 パラメータ

学習に用いたパラメータを表1に示す. バイアスの減衰率 γ (3.2.3節参照) は, 実験1,2では1とし (すなわち減衰なし), 実験3では, 1, 0.9 とした場合を比較する. 全ての条件において, 乱数シードを変更し10回ずつ実験を行う. 次節ではその平均の結果を載せる.

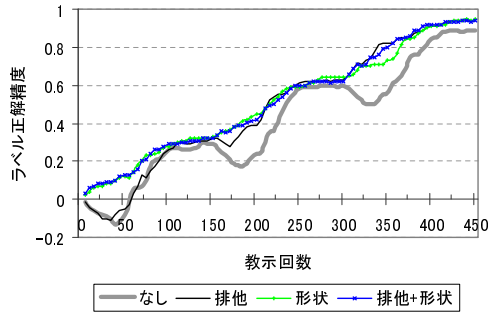
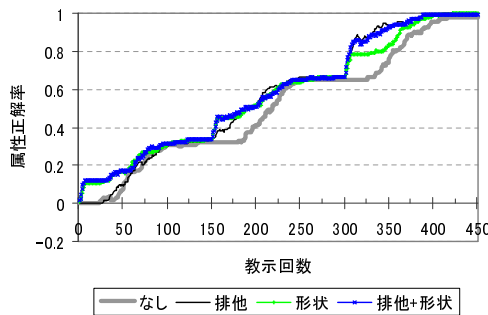
4.2 実験結果

§1 実験1

Online-EM法により学習された $p(\mathbf{x}_i)$ と $p(\mathbf{x}_i|w)$ の混合数を表2に示す. $p(\mathbf{x}_i)$ と $p(\mathbf{x}_i|w)$ の間で混合数が異なるのは, $p(\mathbf{x}_i)$ は5000回学習しているのに対し, $p(\mathbf{x}_i|w)$ は各ラベルにつき約21回程度しか学習されないためである. 形状は色相や明度よりも混合数が少ないこと, 各ラベルの対象属性が非対象属性よりも少ないことは, どちらもその属性に関して教示された特徴の分布が狭いことに起因する. なお, 確率分布の学習はバイアスの有無に関わらず同様に行うため, 混合数に違いは生じない.

ラベル正解精度を図5に示す. また, 参考のために, 各条件における属性正解率*2を図6に示す. 両図共に, バイアスを用いない場合 (図中「なし」), 相互排他性バイアスを単独で用いた場合 (図中「排他」), 形状類似バイアスを単独で用いた場合 (図中「形状」), 二つのバイアスを組み合わせた場合 (図中「排他+形状」) を比較して

*2 属性正解率は, エージェントが出力したラベルからではなく, エージェントの内部状態を直接参照し, 全てのラベルに対して対象属性が正解かどうかを判定した.

図 5 実験 1 のラベル正解精度 ($\gamma = 1$)図 6 実験 1 の属性正解率 ($\gamma = 1$)

いる。

450 回の教示が終わった時点での属性正解率は、バイアスを用いた場合には約 100%，用いない場合でも約 98% であり，十分な教示が行われることで，対象属性を正しく判定できるようになることがわかる．また，正解精度は，バイアスを用いた場合には 94%，用いない場合で 89% であった．ラベルセットの切り替えは教示 151 回目，301 回目に行われる．全てのラベルが教示されるのは 301 回目以降になるため，各段階の正解精度には上限があり，序盤の 150 回までは最高で約 33%，151～300 回の間は約 67% である．どの条件においてもラベルセットの切り替えまでの間に，ほぼ上限まで正解精度が上昇した．しかし，バイアスの有無によって上昇の仕方（立ち上がり）が異なる．

バイアスを用いない場合，ラベルセットが切り替わる際に大きく正解精度が低下する．これは，新たに教示されたラベルの対象属性が，形状・色相・明度の全てであると判定され，挿入誤りが増加するためである（3.1.2 節参照）．

一方で，形状類似バイアスを用いた場合，ラベルセットを切り替えても，正解精度は低下は見られなかった．形状類似バイアスは，新たに教示されたラベルの対象属性を形状と判定しやすくする．そのため，各ラベルセットの約 3 分の 1 を占める形状のラベルは，一度の教示で正しく属性判定されるため，図 6 に示すように，属性正解率の立ち上がりが約 11% 上昇する．この時の確信度

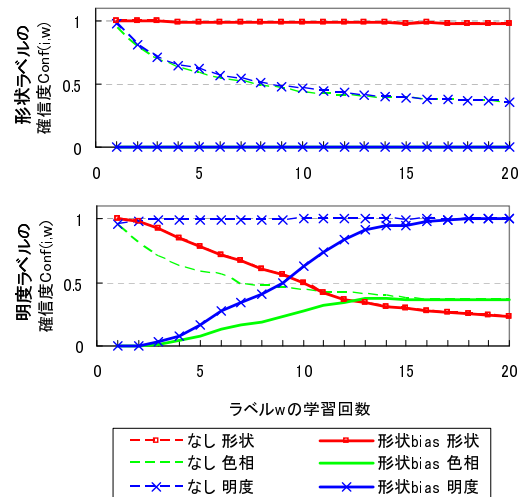


図 7 形状類似バイアスによる確信度の変化

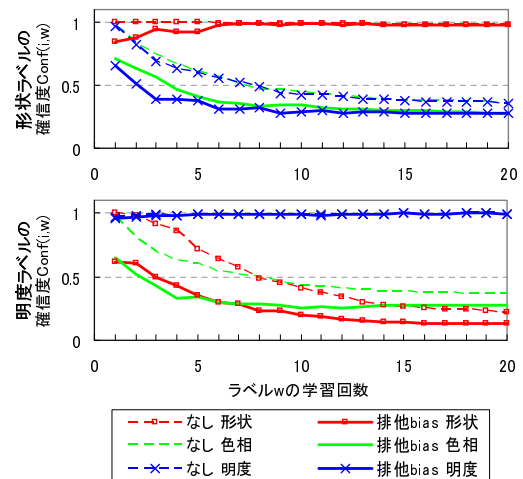


図 8 相互排他性バイアスによる確信度の変化

$Conf(i, w)$ の変化を図 7 に示す．上のグラフが形状のラベル ($w = \text{形状}$) に関する各属性の確信度，下が明度のラベル ($w = \text{明度}$) に関する確信度である．横軸は各ラベルの学習回数を表している．なお，色相のラベルに関しては明度と同様の傾向を持つため省略する． $Conf(i, w)$ が 0.5 よりも大きい時，エージェントは属性 i をラベル w の対象属性と判定する．バイアスを用いない場合，確信度は $p(x_i)$ と $p(x_i|w)$ の距離から計算される．従って，確信度を下げるためには， $p(x_i)$ と同じような分布の形になるまで， $p(x_i|w)$ が学習されなければならない．形状類似バイアスを用いると，図 7 に示すように，形状以外の属性にバイアスが掛かり，確信度の値が低くなる．これによって，エージェントは形状のラベルに関して，一度で正しく属性判定できる．一方で，学習初期においては，色相と明度のラベルも形状と判定される（図 7 下参照）．この誤りは，分布の学習に伴って形状が対象属性から外れる（確信度が 0.5 以下になる）と，対象属性が空

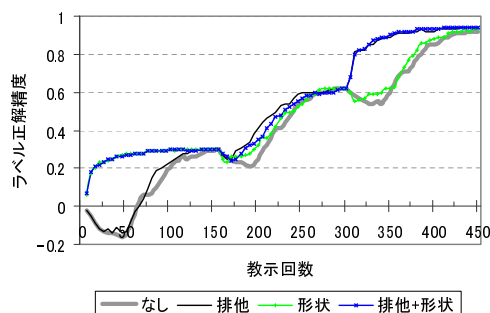


図 9 実験 2 のラベル正解精度 ($\gamma = 1$)

集合となるため、バイアスが無効化 (3・2・3 節参照) され、各ラベルの対象属性が正しく判定されるようになる。図 7 の結果は平均を表しているため、徐々に確信度が上昇しているように見えるが、実際は無効化された瞬間にバイアス無しと同じ値に戻る。

相互排他性バイアスを用いた場合、学習の序盤においてはバイアスなしと同等であるが、終盤では形状類似バイアスを単独で用いた場合よりも正解精度の立ち上がりが良い (図 5 参照)。この結果から、相互排他性バイアスは、獲得した語意が多いほど効果的であることがわかる。確信度 $Conf(i, w)$ の変化を図 8 に示す。形状、明度どちらのラベルに対しても、対象属性の確信度の値は低くなっている。形状類似バイアスと比較すると、形状ラベルに関しては劣っているが、明度ラベルに関しては早く正しい属性を判定できることがわかる。

二つのバイアスを組み合わせた場合には、学習前半は形状類似バイアスと同等、後半は相互排他性バイアスと同等であり、両バイアスの長所が生かされる結果となった。紙面上の都合により図は省略するが、確信度の変化も図 7 と図 8 の間を取ったような形になる。

§ 2 実験 2

実験 2 の結果を図 9 に示す。この実験では、学習の序盤には形状のラベルだけを教え、中盤では色相だけ、終盤では明度だけを教えた。その結果、実験 1 に比べ、バイアスの有効性が顕著に見られた。

形状類似バイアスを単独で用いた場合、序盤の形状ラベルについては正しく属性判定ができるため、正解精度が大きく向上する。しかし、中盤の色相ラベルや、終盤の明度ラベルも形状ラベルと判定するため、正しい形状ラベルの出力が妨害されることがあった。

一方で相互排他性バイアスを単独で用いた場合は、実験 1 と同様に学習が進むにつれ、正解精度の立ち上がりの良さが向上する。

二つのバイアスを組み合わせた場合には、初めの形状については形状類似バイアスが働き、色相、明度については相互排他性バイアスが働くため、学習の全体を通して正解精度が改善された。

表 3 実験 3 で得られた $p(x_i|w)$ の混合数

	$p(x_{\text{形状}} w)$	$p(x_{\text{色相}} w)$	$p(x_{\text{明度}} w)$
形状ラベル	1.7	19	19
色相ラベル	3.4	8.6	19
明度ラベル	3.4	19	8.6
色相 + 明度	3.4	8.6	8.6

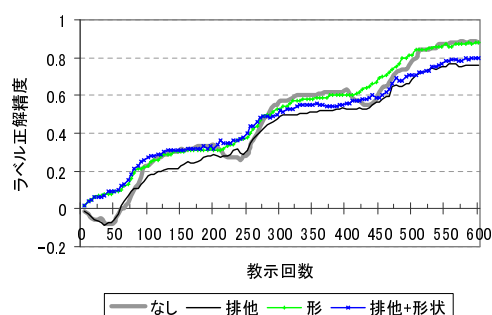


図 10 実験 3 のラベル正解精度 ($\gamma = 1$)

§ 3 実験 3

まず、実験 3 で得られた $p(x_i|w)$ の混合数を表 3 に示す。実験 1 と比べて全体的に数が増えているのは、各ラベルの学習回数が 6 回程度増えているためである。また、色相や明度のラベルに関しては、対象属性の特徴空間が 7 等分から 3 等分へと広がっているため、混合数も増加した。

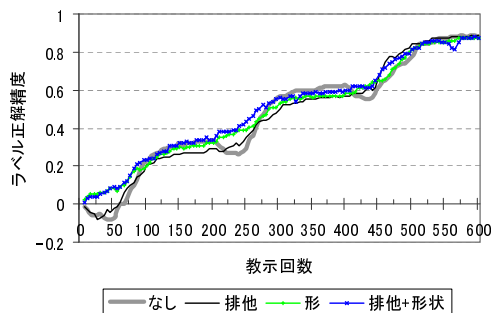
次に、バイアスの減衰率 γ を 1 とした場合の結果を図 10 に示す。形状類似バイアスに関しては実験 1 と同様の有効性が見られたが、相互排他性バイアスを単独、および組み合わせて用いた場合には、ラベル正解精度の低下が見られた。属性正解率の内訳を表 4 に示す。表から、相互排他性バイアスが、複数の属性を対象としたラベルの学習に悪影響を与えることがわかる。

形状類似バイアスが誤って適用された場合、4・2・1 節で説明したように、学習が進むにつれ対象属性が空集合になることで、バイアスが無効化される。しかし、複数の対象属性の一方を、相互排他性バイアスが誤って抑制した場合には、もう一方の対象属性が抑制されずに残っているため、対象属性が空集合になることはなく、バイアスは無効化されない。このことが、表 4 に示した属性正解率の低下に繋がっている。本実験では、色相と明度を対象属性とするラベルを扱っているが、形状を含む複数属性を対象とするラベルを教示する場合には、形状類似バイアスでも同様の問題が生じる。

最後に、 γ を 0.9 とした場合の結果を図 11 に示す。バイアスの強さを学習回数に応じて減衰させることで、相互排他性バイアスによる悪影響がなくなり、実験 1 と同様の利点が見られた。

表 4 実験 3 ($\gamma = 1$) の属性正解率 (教示 596 回から 600 回までの平均)

	形状	色相	明度	色相+明度
なし	1	1	1	1
排除	1	0.97	0.97	0.78
形状	1	1	1	1
排除+形状	0.99	1	1	0.87

図 11 実験 3 のラベル正解精度 ($\gamma = 0.9$)

4.3 考 察

バイアスを用いない場合、確率分布が十分に学習されなければ、正しく対象属性を判定することはできない。しかし、バイアスを用いることで、確率分布の学習が不十分でも正しい対象属性を判定できるようになり、高い正解精度が得られることが示された。すなわち、本手法を用いたエージェントは、少ない学習回数でも正しくラベルを使用できるようになるため、効率的である。

実験 1 では各属性のラベルの数を等しく設定し、その教示の順番もランダムとしているため、形状類似バイアスの代わりに、色相類似バイアスを適用しても、同様の結果が得られる。形状類似バイアスが最も有効に機能するのは、実験 2 のように、学習の初期において、形状のラベルが多く教示される場合である。実世界で人間と協調作業するロボットを想定すると、未知のオブジェクトに対するラベルを学習する機会が多く、また、それらのラベルはオブジェクトの形状を表す場合が多いため、形状類似バイアスの有効性が期待できる。しかし、ロボットのセンサやタスクによっては、他の属性にバイアスを適用する方が有効な場合もあるため、環境に応じて適用するバイアスを調整する機構が望まれる。バイアスの自律的な調整は今後の課題となっている。

また、実験 3 で示したように、複数の属性を対象としたラベルを正しく学習する際には、バイアスを減衰させる必要がある。図 12 に、 γ の値を 1, 0.9, 0.7, 0 とした場合の結果を示す。 γ の値を 1 とするとバイアスは減衰せず、0 とするとバイアスは適用されない。 γ を 0.9 とすると、確率分布から正しい対象属性を判定できるようになる頃にはバイアスが機能しなくなるため、悪影響は見られなくなる。しかし、 γ を 0.7 とすると、バイアスの

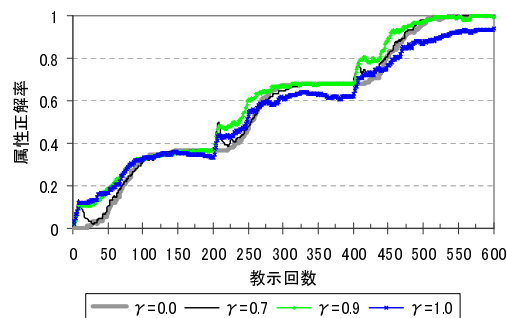


図 12 実験 3 における相互排他性バイアスと形状類似バイアスを組み合わせた場合の属性正解率

減衰速度が早すぎるため、十分な効果が得られない。このように、学習初期におけるバイアスの有効性を維持しつつ、悪影響を排除するためには、確率分布に基づいた属性判定の速度に合わせて γ の値を調整する必要がある。 γ の自律的な調整も今後の課題となっている。

本手法は、3.1 節で述べた通り、属性が互いに独立であることを仮定し、確率計算を近似している。そのため、独立性が成り立たない場合には、学習に悪影響を与えることが想定される。実際の言語環境における属性間の独立性や、それによる悪影響を定量的に評価するため、今後は人間との共同作業を設定するなど、より自然な状況で実験を行っていきたい。

5. む す び

本論文では、エージェントによる語意学習の効率化を目指し、確率分布に基づく属性判定と、相互排他性および形状類似バイアスによる属性判定とを組み合わせた手法を提案した。実験の結果から、バイアスを用いることで、少ない教示でも正しく属性判定できるようになり、高い正解精度が得られることが示された。形状類似バイアスは、学習初期に形状のラベルが多く教示される場合に最も有効に機能し、相互排他性バイアスは、語意知識の増加に伴って効率性が向上した。また、両バイアスは誤って適用される場合があるが、ラベルの学習回数に応じてバイアスの強さを減衰させることで、バイアスの悪影響を排除することができることを示した。

今後は、実際の対話環境により近づけた実験を進めると共に、環境に合わせてバイアスを自律的に調整する機構の実現を目指す。

謝 辞

本研究の一部は文部科学省 21 世紀 COE プログラム「インテリジェントヒューマンセンシング」と、科学技術研究費基礎基盤 (C)18500130 の補助による。

◇ 参 考 文 献 ◇

- [赤穂 97] 赤穂 昭太郎, 長谷川 修, 吉村 隆, 麻生 英樹, 速水 悟: EM 法を用いた複数情報源からの概念獲得, 電子情報通信学会論文誌, Vol. J80-A, No. 9, pp. 1546-1553 (1997)
- [Carey 78] Carey, S. and Bartlett, E.: Acquiring a single new word, *Papers and Reports on Child Language Development*, Vol. 15, pp. 17-29 (1978)
- [今井 97] 今井 むつみ: こたばの学習パラドックス, 共立出版 (1997)
- [Iwahashi 03] Iwahashi, N.: Language acquisition through a human-robot interface by combining speech, visual, and behavioral information, *Information Sciences*, Vol. 156, pp. 109-121 (2003)
- [石井 98] 石井 信, 佐藤 雅昭: オンライン EM アルゴリズムによる動的な関数近似, 電子情報通信学会技術研究報告, NLP97-142, pp. 43-50 (1998)
- [金 00] 金 景柱, 岩橋 直人: 知覚情報の統合に基づく言語音声単位の獲得アルゴリズム, 電子情報通信学会技術研究報告, TL200-21, pp. 9-16 (2000)
- [栗田 97] 栗田 多喜夫, 小林 保雄, 三島 健稔: PARCOR 画像の高次局所自己相関特徴を用いた背景変化および平行移動に強いジェスチャー認識, 電子情報通信学会技術研究報告, PRMU96-213, pp. 159-165 (1997)
- [Landau 88] Landau, B., Smith, L. B., and Jones, S. S.: The importance of shape in early lexical learning, *Cognitive Development*, Vol. 3, pp. 299-321 (1988)
- [Markman 89] Markman, E. M.: *Categorization and naming in children*, MIT Press (1989)
- [中川 95] 中川 聖一, 升方 幹雄: 視聴覚情報の統合化に基づく概念と文法の獲得システム, 人工知能学会誌, Vol. 10, No. 4, pp. 619-627 (1995)
- [Quine 60] Quine, W. V.: *Word and Object*, MIT Press (1960)
- [Roy 02] Roy, D.: Learning visually-grounded words and syntax for a scene description task, *Computer Speech and Language*, Vol. 16, No. 3 (2002)

〔担当委員: 岡 隆一〕

2006 年 12 月 4 日 受理

—— 著 者 紹 介 ——



田口 亮(学生会員)

2004 年豊橋技術科学大学大学院工学研究科修士課程修了。現在, 同大学院博士後期課程に在学中。情報処理学会の会員。人とロボットの言語コミュニケーションに興味を持つ。



木村 優志

2007 年豊橋技術科学大学大学院工学研究科修士課程修了。現在, 同大学院博士後期課程に在学中。



小玉 智志

2005 年豊橋技術科学大学大学院工学研究科修士課程修了。同年, 株式会社日立情報システムズ入社。



篠原 修二

1999 年神戸大学大学院自然科学研究科博士後期課程修了。博士(理学)。ベンチャー企業で, Web アプリケーションやデータベースシステムの開発に携わった後, 2005 ~ 2007 年まで豊橋技術科学大学で教育支援システムの開発を行う。現在, 画像処理の研究に従事。



入部 百合絵(正会員)

1999 年名古屋工業大学工学部生産システム工学科卒業。2001 年名古屋大学大学院人間情報学研究科修士課程修了。2004 年名古屋大学大学院人間情報学研究科博士課程満期退学。現在, 豊橋技術科学大学情報メディア基盤センター助手。博士(学術)。教育支援, ユーザインタフェース等に関する研究に従事。情報処理学会会員。



桂田 浩一(正会員)

1995 年大阪大学基礎工学部情報工学科卒業。2000 年同大学院同大学基礎工学研究科博士後期課程修了。同年豊橋技術科学大学大学院工学研究科助手, 2007 年同講師, 現在に至る。博士(工学)。音声・マルチモーダル対話, 知識処理に関する研究に従事。AAAI, 情報処理学会, ヒューマンインタフェース学会, 日本音響学会, 言語処理学会各会員。



新田 恒雄(正会員)

1969 年東北大学工学部電気工学科卒業(株)東芝総合研究所, 同社マルチメディア技術研究所を経て, 1998 年豊橋技術科学大学大学院工学研究科教授。工学博士。音声認識・合成・文字認識, マルチモーダル対話システム, および概念獲得の研究に従事。IEEE, 電子情報通信学会, 情報処理学会, 日本音響学会各会員。