

Saliencyに基づく能動性を利用した語彙獲得

Lexical acquisition using the active nature based on saliency

菊池 匡晃 (阪大) 正 荻野 正樹 (阪大) 正 浅田 稔 (阪大, 阪大 FRC)

Masaaki KIKUCHI, Osaka University, 2-1, Yamadaoka, Suita, Osaka

Masaki OGINO, Osaka University

Minoru ASADA, HANDAI Frontier Research Center, Osaka University

This paper proposes a lexical acquisition model which makes use of saliency to associate visual features of observed objects with the labels that is uttered by a caregiver. A robot changes its attention and learning rate based on saliency. Simulation experiments show that the learning model with saliency effectively associate the given labels with the observed features. Moreover, in the experiment with a real humanoid robot, the visual features are represented with self organizing maps which adaptively represents the shape of observed objects independent of the viewpoints.

Key Words: lexicon acquisition, saliency, humanoid

1 はじめに

近年、ロボットは実社会において人間と共に活動することが期待されており、人間との言語コミュニケーションは最重要課題のひとつである。人間と言語コミュニケーションするためには語彙を共有する必要があるが、語彙情報をあらかじめ全てロボットに教えておくことは事実上不可能である。そのため、ロボットが生活する間に得た世界の情報をもとに自ら語彙を獲得することが求められる。

従来の語彙獲得ロボットはそのほとんどが受動的に観測した視覚的特徴量とラベルの同時生起性に基づく統計的手法によるものであった^{1, 2, 3, 4)}。しかし、このような統計処理では語彙の獲得までに多くの事例を必要とし、実環境下では現実的ではない。1つの事例に対する漠然とした指さしから、発話された言葉の意味を推測することは論理的に不可能であるにも関わらず、人間の幼児はほとんどの場合、即時に言葉に意味を付加することができる。この現象は発達心理学の分野では様々な制約でラベル付けする対象を絞ることによって起こると考えられており、Markman⁵⁾は事物全体原理、相互排他原理などによる制約を、Landau et. al.⁶⁾は形状類似バイアスによる制約を提案している。

人間の幼児が受動的に得られる情報のみで語彙獲得をしているとは考えにくく、ロボット自身の能動性が語彙獲得において重要であると考えられる。そこで本研究ではロボット自身が感じる特徴量の顕著性 (saliency) に基づいて能動的に注視する対象を選択し、探索し、さらに特徴量と単語のマッピング時に saliency をバイアスとして制約をかけることで効率よく語彙を獲得していくシステムを提案する。提案するシステムの有効性をシミュレーション実験により検証する。またヒューマノイドロボットに適用して実環境下においても効率よく語彙が獲得されることを示す。本稿では提案するシステムをシミュレーション上で実装し、その有用性を検証した。また実環境下での語彙獲得を考えると、見る方向によって物体の形状が異なって見えるため、正しくラベル付けができないという問題が起こる。そこで人間の視覚野をモデル化した特徴抽出器をロボットが学習によって獲得していく、視点によらないロバストな物体認識システムを提案する。この認識システムを用いてヒューマノイドロボットに提案

する語彙獲得システムを実装し、実環境下で視点の違いによらず正しい語彙が獲得されることを示した。

2 Saliency を利用した語彙獲得

2.1 タスクと仮定

ロボットが環境中に存在する対象物の特徴に関する語彙を、養育者とのコミュニケーションを通して獲得していくタスクを取り扱う。ロボットは対象物について以下の2種類の情報を得ることができる。

- 特徴量：ロボットがある1つの対象物を注視すると、色や形状など複数の属性に対応する特徴量がセンサ入力として得られる。ただし属性と特徴量との関係はロボットにとって未知であり、属性のカテゴリの識別はできない。
- ラベル：養育者が教える対象物の特徴を説明する単語をラベルと呼ぶ。全てのラベルは単一の属性をさし、包含関係はない。

2.2 Saliency に基づく能動性

ロボットは以下のような2つの saliency に基づく性質を持つ。

1. 注視対象物の能動的選択
視野内に見えている複数の対象物の中から最も saliency の高い特徴を持つものを選択して注視する。
2. *Saliency* による学習バイアス
特徴量のラベルへのマッピングには Hebb 学習を用いる。結合の更新の際に saliency によってバイアスをかけ、saliency の高い特徴ほど強く結合される。

提案手法の概要を Fig. 1 に、学習の流れを Fig. 2 に示す。

2.3 Saliency の定式化

頻繁に観測される特徴には慣れ、saliency を感じなくなることであらわす項 S_1 と、ラベルを知っている特徴には saliency を感じなくなることであらわす項 S_2 の積によって saliency はあらわされる。特徴量 i に対する saliency S_i は以下のように定義される。

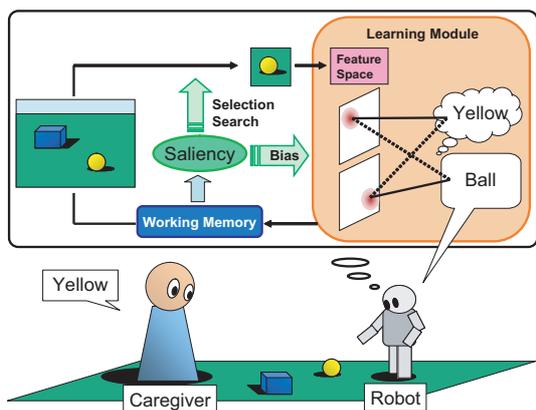


Fig.1 An overview of the system utilizing saliency for lexicon acquisition

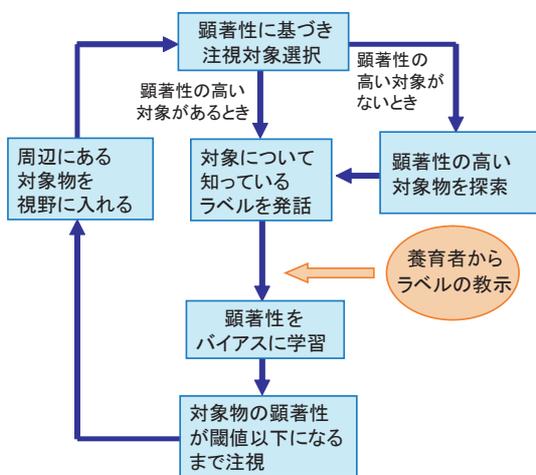


Fig.2 Flow of learning

$$S_1^i(t) = S_1^i(t-1) + \Delta S_1^i(t-1) \quad (1)$$

$$\Delta S_1^i(t) = \frac{\alpha(1 - S_1^i(t)) - \beta S_1^i(t) I^i(t)}{\tau} \quad (2)$$

$$S_2^i = 1 - \frac{1}{1 + e^{-\gamma(w-\theta)}} \quad (3)$$

$$S^i = (S_1^i + c_1)(S_2^i + c_2) \quad (4)$$

ただし $\alpha, \beta, \gamma, \tau, \theta, c_1, c_2$ は定数である。

2.4 養育者とのコミュニケーション

ロボットは注視する対象物を決めると、対象物を指差してラベルの教示を求める。またその対象物について知っているラベルがあるときには、全て発言して知識を養育者に伝える。養育者はロボットが発言したラベルの正誤によって、以下のようにラベルを教示する。

- 発言が正しい、または発言なしのとき
発言されていないラベルのうち1つを教示。
- 発言が誤りのとき
発言されたラベルが誤りであると教え、同じ属性の正しいラベルを教示。
- 全ての属性について発言が正しいとき
何も教示しない。

なお、養育者とロボットの共同注意は必ず成立するものとしている。

3 シミュレーション実験

3.1 実験設定

ロボットは環境中に存在する対象物の特徴をあらゆるラベルを学習していく。シミュレーションでは色、形状だけでなく大きさ、重さ、硬さの合計5種類の属性についての特徴量をロボットは観測することができる。対象物は色40通り、形80通り、大きさ、重さ、硬さはそれぞれ8通りのバリエーションを持ち、ロボットは各特徴量に対応する合計144のラベルを学習する。学習の進度に応じて対象物が増加していく環境下において、能動的選択および学習バイアスの有無による比較実験を行った。

3.2 シミュレーション実験の結果と考察

10回試行した結果、Fig. 3に示す平均学習曲線が得られた。横軸が学習ステップ数を、縦軸が正しく獲得された語彙数を示している。ただし養育者がラベルを教示し、ロボットがHebb学習のネットワークを更新するまでを1ステップとした。なお、正しいラベルと特徴量との結合度が $w_{l \rightarrow i} > 0.9$ を満たした場合に語彙を獲得したとみなし、全ての語彙を獲得するまでを1試行とした。また1つの語彙を獲得するまでに要した教示の回数の統計をFig. 4に示す。横軸が語彙を獲得するまでに要した教示回数を、縦軸がその教示回数で語彙を獲得した語数を示している。

提案手法により語彙が非常に効率よく獲得され、単純なHebb学習と比較しておよそ25%の学習ステップ数で全ての語彙を学習することができた。以下では能動的選択および学習バイアスの影響について考察する。能動的選択の影響 学習曲線から能動的選択によって学習の収束が早くなっていることがわかる。冗長な学習を避け、効率よく学習データを収集できているためと考えられる。

学習バイアスの影響 学習曲線から、バイアスがある場合には学習中期から急速に語彙数を伸ばすことがわかる。Fig. 4から学習バイアスがある場合には、語彙獲得までの教示回数のピークが左にシフトしており、少ない教示で効果的に正しいラベルとマッピングができています。特に、能動的選択と共に適用されると全体の7割にあたる約100語を1度の教示で獲得できている。バイアスによってラベル付けする特徴が制約されるため効率がよくなると思われる。

4 実ロボットへの適用

シミュレーションによる実験によって提案システムの語彙獲得タスクにおける有用性が示された。提案システムをヒューマノイドロボットに実装し、実環境下においても語彙獲得が実現されることを示す。実環境下では、同じ対象物でも視点が異なると見え方も異なるという問題が起こる。実環境下での語彙獲得を検証するためにはこの問題を解決する必要がある。ヒトの脳での処理をモデル化した、視点によらない物体認識システムを提案する。この物体認識システムを用いて実ロボットによる語彙獲得タスクを行う。

4.1 視点によらない物体認識

色や形状などの各属性についての特徴量のクラスタリングにはSOMを用いる。ロボットは、それぞれの属性について観測された原特徴データをSOMによってクラスタリングし、原特徴をニューロンの数に量子化する。学習後には各属性の原特徴がそれぞれのSOMに入力され、各ニューロンの活性化度を並べたベクトルを、観測された特徴量としてロボットは認識する (Fig. 5)。

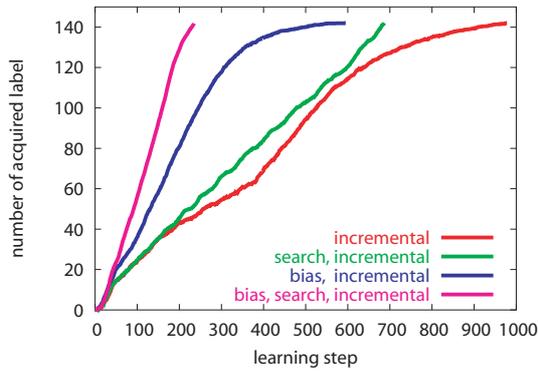


Fig.3 Learning curves

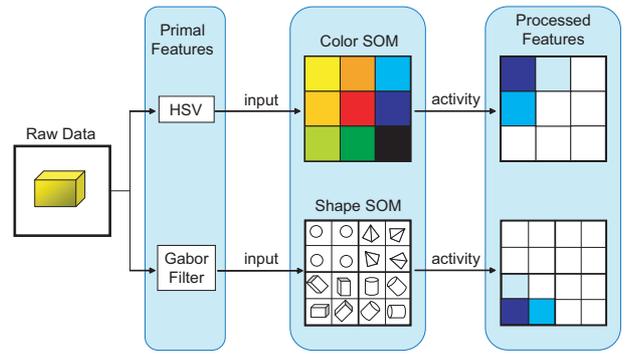


Fig.5 The flow of feature extraction

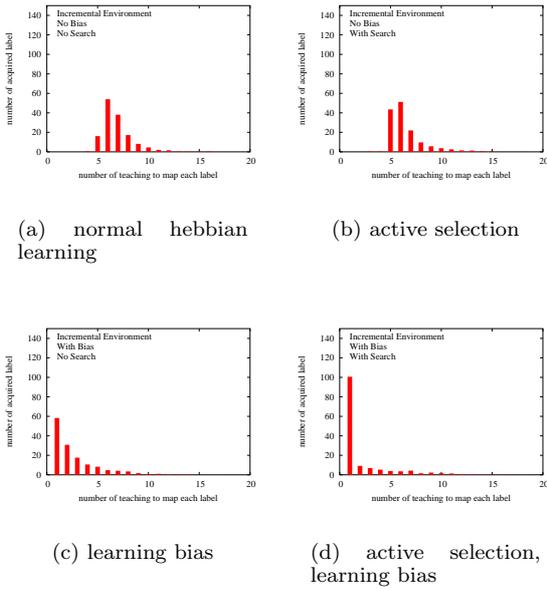


Fig.4 Numbers of teaching upon mapping each label

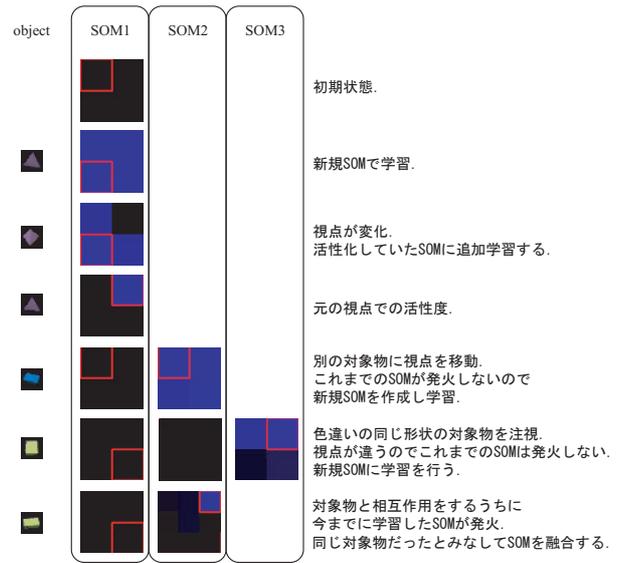


Fig.6 The flow of learning shape SOM

形状における特徴抽出 SOM の学習の流れを Fig. 6 に示す。色における学習も同様に行われる。なお SOM への追加学習や複数の SOM の融合の際には、SOM のニューロンの代表ベクトルと追加データを新たな学習データとして SOM の学習を行う。小さな SOM では学習が収束しない場合があるが、その場合には SOM のニューロンの数を自動的に増やして再学習することで、このような問題を解決する。

4.2 実験設定

実験はカメラを備えたヒューマノイドロボットを用いて行った。実験環境を Fig. 7 に示す。ロボットは首を振って対象物を注視する、歩行して対象物に近づく、キック動作をして対象物と相互作用する、対象物を指差して養育者にラベルの教示を求めるなどの行動をとることができる (Fig. 8)。ロボットは能動的に対象物と相互作用し、さまざまな視点から対象物を観察しながら、視点によらない特徴抽出器を獲得していく。またそれと同時に、養育者とのコミュニケーションを通して対象物に関する語彙を学習する。ロボットは相互作用を終えると、集めたデータを用いて色と形状の SOM を再学習して更新する。その後、対象物を指差して、知っているラベルを発言する。養育者はラベルが誤っていれば訂正し、正しいラベルを教示する。発言内容が正しい、または発言がされなかったときには、発言されていない正しいラベルを教示

する。注視対象についての語彙を教示されると、ロボットは注視対象の saliency が閾値以下になるまで注視した後に、別の対象物を注視する。

4.3 実験結果

実験の結果を Fig. 9 に示す。ロボットは相互作用を通して特徴抽出を学習し、視点の違いによらないロバストな特徴抽出器を獲得できた。また特徴抽出器の学習と並行して saliency を用いた提案システムでラベルを効率よく学習することができた。Fig. 9 を見ると学習初期から効率よく語彙を獲得できているが、200[sec] 以降鈍化していることがわかる。これは同じ箱形の対象物でありながら大きく視点の異なる 2 つの SOM を作成してしまい、これらを融合するための能動的な視点の変化に時間がかかったためであった。ロボットは対象物に歩いて近づく、キックするなどの行動で物体への視点を能動的に変化させるが、視点の大きな変化に時間がかかり、SOM の融合が起こるまで正しく学習できなかったと考えられる。

5 結言

本論文では saliency に基づいて注視対象を選択し、ラベルとのマッピングに saliency をバイアスとしても用いることで効率的な語彙獲得を行うシステムを提案した。シミュレーションによって提案システムの有用性を検証し、非常に効率的に多数の語彙を獲得できることを示した。また提案システムをヒューマノイドロボットに実装し、実

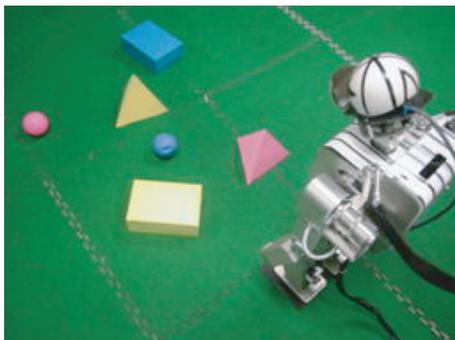


Fig.7 Environmental setting



(a) kicking



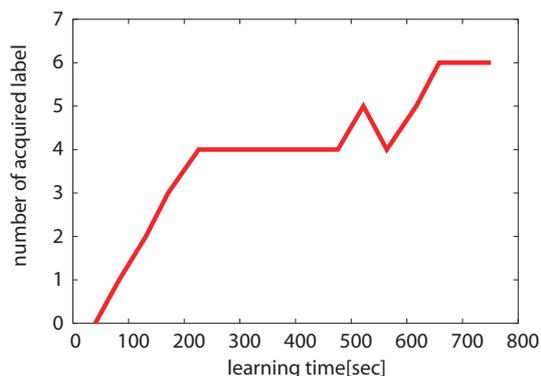
(b) pointing

Fig.8 Examples of robot's behavior

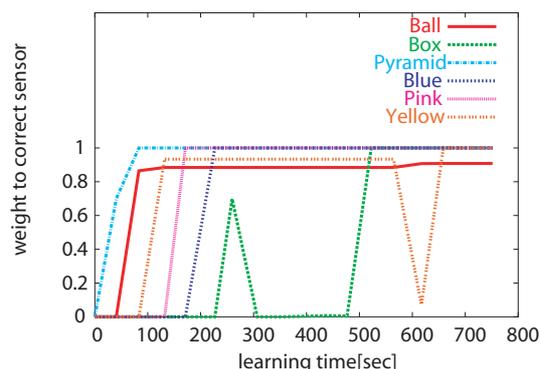
環境における語彙獲得タスクに適用した．提案する特徴抽出器によって，ロボットに対象物の認識を行うことができたため，語彙はシミュレーションと同様に効率よく獲得された．saliency に基づく能動性を対象物との相互作用への動機付けにも利用し，視点によらない物体認識能力の獲得に役立てることができた．このような観点からも，ロボットに能動性を持たせることは重要であると言える．

参考文献

- [1] 赤穂昭太郎, 速水悟, 長谷川修, 吉村隆, 麻生英樹. 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. 80, pp. 1546–1553, 1997.
- [2] 石黒勝彦, 國吉康夫, 大津展之. インターモーダル学習による概念の獲得. 日本機械学会ロボティクス・メカトロニクス講演会'04 講演論文集, 2004.
- [3] Luc Steels and Frederic Kaplan. Aibo's first words. the social learning of language and meaning. *Evolution of Communication*, 2001.



(a) Number of acquired label



(b) Weight changes of each label

Fig.9 Experimental result using real robot

- [4] 岩橋直人. ロボットによる言語獲得: 言語処理の新しいパラダイムを目指して. 人工知能学会誌, Vol. 18, No. 1, pp. 49–58, 2003.
- [5] E. M. Markman. Categorization in children: Problems of induction. *Cambridge, MA: MIT Press, Bradford Books*, 1989.
- [6] B. Landau, L. B. Smith, and S. Jones. The importance of shape in early lexical learning. *Cognitive Development*, Vol. 3, pp. 299–321, 1988.