

# エージェント間の対話における非論理バイアスの有効性

作元 佑輔<sup>†</sup> 田口 亮<sup>†</sup> 篠原 修二<sup>†</sup> 入部 百合絵<sup>‡</sup> 桂田 浩一<sup>†</sup> 新田 恒雄<sup>†</sup>

<sup>†</sup>豊橋技術科学大学 大学院工学研究科

<sup>‡</sup>豊橋技術科学大学 情報メディア基盤センター

## 1. はじめに

発達心理学や比較心理学の分野において、ヒトの推論過程は純論理的な演繹過程ではないことが知られている。代表的な例として、相互排他性バイアスというものがある。このバイアスは、「pならばq」を知っている(信じて)時、そのことから「pでないならばqではない」も信じてしまう傾向性と捉えられる[1]。しかし、このような推論は、論理的に正しいとはいえない。一方、比較心理学の分野では、対称性や刺激等価性の成立に関する研究を通して、ヒトは「pならばq」から「qならばp」を推論する傾向性(対称性バイアス)をもつことが知られている[2]。これは論理的に間違った推論であるにもかかわらず、年齢や能力などとは無関係に多くのヒトで容易に成立する。

本研究の目的は、これらの傾向性を利用することで、エージェントの語彙学習を効率化することである。本稿では、バイアスを持つエージェントモデルと持たないエージェントモデルを構築し、両者に語彙学習課題を与えることで性能を評価する。

## 2. 語彙学習課題

まず、対象集合を  $O = \{o_1, \dots, o_i, \dots, o_n\}$ 、ラベル集合を  $L = \{l_1, \dots, l_j, \dots, l_k\}$  とする。本研究では簡単のため、対象は色属性と形属性のみを持つこととし、各対象には、その色と形を表す二つのラベルが割り当てられているものとする。

エージェントは、対象集合からランダムに選ばれた対象  $o_i$  を提示され、対象  $o_i$  に割り当てられた色と形を表すラベルのうちのいずれかを教示される。例えば白くて丸い対象を見せられ、それを「丸」と呼ぶことを教えられる。エージェントはこのような経験を繰り返すことによって

- 「ラベル  $l_k$  は対象  $o_i$  である」という言明に対する確信度 ( $B(o_i | l_k)$ ) と表記する
- 「対象  $o_i$  はラベル  $l_k$  である」という言明に対する確信度 ( $B(l_k | o_i)$ ) と表記する

という二つの確信度を形成していく。エージェントはこれらの確信度に基づいて、「何?」(What 質問と呼ぶ)と「どれ?」(Which 質問と呼ぶ)の二つの質問に答えることが求められる。

What 質問は、ある対象  $o_i$  を見せ、「 $o_i$  は何?」と質問することである。エージェントは、対象  $o_i$  に対して最も確信度  $B(l | o_i)$  の高いラベルを回答する。Which 質問は、あるラベル  $l_k$  を用いて「ラベル  $l_k$  はどれ?」と質問することである。エージェントは、 $l_k$  に対して最も確信度  $B(o_i | l_k)$  の高い対象を選択する。

## 3. エージェントモデル

### 3.1. 条件付確率モデル

バイアスを持たないモデルとして、以下に示すように、確信度として条件付確率を用いるエージェントを考える。

$$B(l_k | o_i) = \frac{C(o_i, l_k)}{\sum_i C(o_i, l_k)} \quad (1)$$

$$B(o_i | l_k) = \frac{C(o_i, l_k)}{\sum_k C(o_i, l_k)} \quad (2)$$

ここで、 $C(o_i, l_k)$  は  $o_i$  と  $l_k$  の共起頻度を表す。

### 3.2. 対称性バイアスモデル

バイアスを持つエージェントとして、篠原ら[3]は対称性バイアスモデルを提案している。このモデルでは、式(1)(2)で求めた確信度を、式(3)(4)を満たすように修正する。

$$B(l_k | o_i) = \frac{B(o_i | l_k)}{\sum_k B(o_i | l_k)} \quad (3)$$

$$B(o_i | l_k) = \frac{B(l_k | o_i)}{\sum_i B(l_k | o_i)} \quad (4)$$

このような修正を行うことで、 $B(o_i | l_k)$  と  $B(l_k | o_i)$  の間には、

$$B(o_i | l_k) = \frac{|O|}{|L|^2} B(l_k | o_i) \quad (5)$$

のような比例関係が成立するようになる。すなわちエージェントは、「 $o_i$  は  $l_k$  である ( $l_k$  は  $o_i$  である)」と信じるならば、「 $l_k$  は  $o_i$  である ( $o_i$  は  $l_k$  である)」と信じる傾向性を持つことになる。

## 4. 実験

### 4.1. パラメータ設定

今回実施した数値実験では、色属性数を 2 (白, 黒)、形属性数を 3 (丸, 三角, 四角) とした。また対象数は、色属性数 × 形属性数 = 6 である。ラベル数  $K$  は、色属性数 + 形属性数 = 5 である。実験では毎回一つの対象をランダムに選び、その対象に割り当てられたラベルを教示する。そして毎回の教示の後、What 質問と Which 質問を行う。

### 4.2. 結果と考察

図 1 に、What 質問に対する正解率の時間発展を示す。ここで、各対象には色と形を表す二つのラベルが割り当てられているため、正解も二つある。よって、二つのラベルの内いずれかを回答できれば正解とした。教示をしていない時の正解率が約 40% であるが、これは、ラベル数 5 の内、二つのラベルが正解であるので、 $2/5=0.4$  となる。学習のはやきに関しては、対称性バイアスモデルが、条件付確率モデルよりもはやいことがわかる。その理由とは、図 3, 4 を使って説明する。

図 3, 4 は、教示した回数 3 回目 (3 つのオブジェクトについて 3 つのラベルが教示された。それぞれ、▲には「三角」、●には「黒」、■には「四角」である。) の条件付確率モデルと、対称性バイアスモデルの確信度  $B(l | o)$  の一例である。見方は、ある対象  $o_i$  についてラベル方向に走査する。そして、もっとも高い確信度に対応するラベルを学習者エージェントの回答とする。これは What 質問の出力に利用される確信度である。条件付確率モデルでは、▲は「三角」、●は「黒」、■は「四角」と回答するので教示された対象  $o_i$  に関しては全て正解である。しかし、教示されていない対象  $\Delta$ ,  $\circ$ ,  $\square$  については、確信度が全て同じ値であり、ラベル  $l_k$  がランダムに選ばれるため、正解率は  $2/5$  となる。よって、図 3 における What 質問に対する正解率は、70% となる。一方、対称性バイアスモデルでは、▲は「三角」、●は「黒」、■は「四角」と回答する。これらは全て正解となる。教示をしていない対象  $o_i$  については、 $\Delta$  は「白」、 $\circ$  は「丸」、と回答して正解する。ところが、 $\square$  は「丸」と回答して不正解となる。よって、図 4 における What 質問に対する正解率は、約 83% となる。

図 2 に、Which 質問に対する正解率の時間発展を示す。両モデルとも教示をしていない時の正解率が約 41% であった。これは、形ラベルが正解する確率と色ラベルが正解する確率の和なので、 $1/6 + 1/4 \approx 0.41$  となる。学習のはやきに関しては、対称性バイアスモデルが、条件付確率モデルよりもはやいことがわかる。その理由については、図 5 を使って説明する。

図 5 は、教示した回数 3 回目 (3 つのオブジェクトについて 3 つのラベルが教示された。それぞれ、▲には「三角」、●には「黒」、■には「四角」である。) の条件付確率モデルの確信度  $B(o | l)$  の一例である。見方は、あるラベル  $l_k$  について対象方向に走査する。そして、もっとも高い確信度の対象を学習者エージェントの回答とする。これは Which の出力に利用される確信度である。

条件付確率モデルでは、「黒」は●、「三角」は▲、「四角」は■、と回答するので教示されたラベル  $l_k$  に関しては全て正解となる。一方教示されないラベル「白」「丸」については、確信度が全て同じ値であるので、正解率はそれぞれ  $1/2$ ,  $1/3$  となる (色属性数は 2, 形属性数は 3 であるため)。よって、図 5 における Which 質問に対する正解率は、約 63.8% となる。

一方、対称性バイアスモデルでは、図 4 と同様の確信度であるので、Which 質問に対する正解率は、それぞれ黒は  $1/3$ 、三角は  $3/4$ 、四角は 100%、白は 100%、丸は 100% だった。よって、

\* Effectiveness of illogical biases at Agent-Agent Interaction.

Y.Sakumoto<sup>†</sup>, R.Taguchi<sup>†</sup>, S.Shinohara<sup>†</sup>, Y.Iribe<sup>‡</sup>, K.Katurada<sup>†</sup>, T.Nitta<sup>†</sup>

<sup>†</sup>Graduate School of Engineering, Toyohashi Univ.of Tech.

<sup>‡</sup>Information and Media Center, Toyohashi Univ.of Tech.

Which 質問に対する正解率は、約 81.6%となる。

では、What 質問、Which 質問ともに対称性バイアスモデルの学習がはやいのはなぜだろうか。対称性バイアスは、教示した内容(対象  $o_i$  はラベル  $l_k$  である)だけでなく、教示されていない内容(ラベル  $l_k$  は対象  $o_i$  である)も学習してしまう特性を持つ。図 3 より条件付確率モデルは、教示していない対象  $o_i$  の確信度に変化はなかった。しかし、図 4 では、教示していない対象  $o_i$  の確信度も上昇していた。これは、対称性バイアスの効果によるものである。以上のことから、対称性バイアスモデルの方が条件付確率モデルよりも性能が良いといえる。

### 5. まとめと今後の課題

今回は、「条件付確率モデル(バイアスなし)」、「対称性バイアスモデル(バイアスあり)」という 2 体のエージェントモデルを構築し、それらの性能を評価した。その結果、語彙学習課題において、対称性バイアスモデルが条件付確率モデルよりも性能が良いことがわかった。しかし、この結果は、色属性数 2、形属性数 3 のもとで得られた結果である。この属性数を増やした場合、正解率の上昇の仕方にどのような変化が生じるか確認する必要がある。さらに、人間を相手に同様の実験を行い比較検討する必要がある。これらについては、今後の課題としたい。

### 参考文献

- [1] E. M. Markman, J. L. Wasow, M. B. Hansen: "Use of the mutual exclusivity assumption by young word learners", *Cognitive Psychology*, 47, pp. 241-275 (2003).
- [2] 山崎, 岡ノ谷, 入来: "動物の論理を比較する: 意味の獲得へと導くメカニズム", *信学技報*, NC2005-93 (2006-1), pp. 19-24 (2006).
- [3] 篠原 修二, 田口 亮, 橋本 敬, 桂田 浩一, 新田 恒雄, "語彙学習エージェントにおけるバイアスの自律調整について", *Vol. 22, No. 2*, 掲載予定 (2007) .

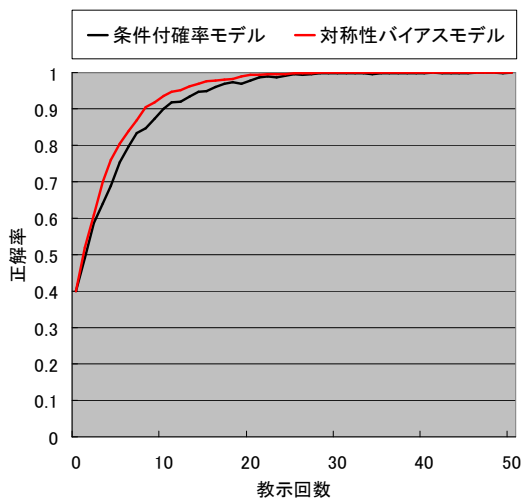


図 1 What 質問した時の正解率

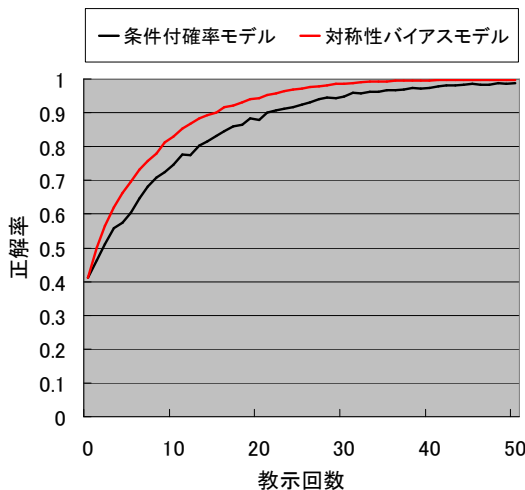


図 2 Which 質問した時の正解率

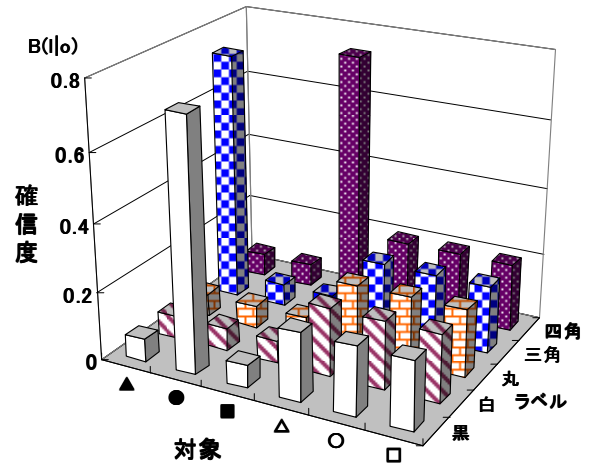


図 3 条件付確率モデルの確信度  $B(l|o)$

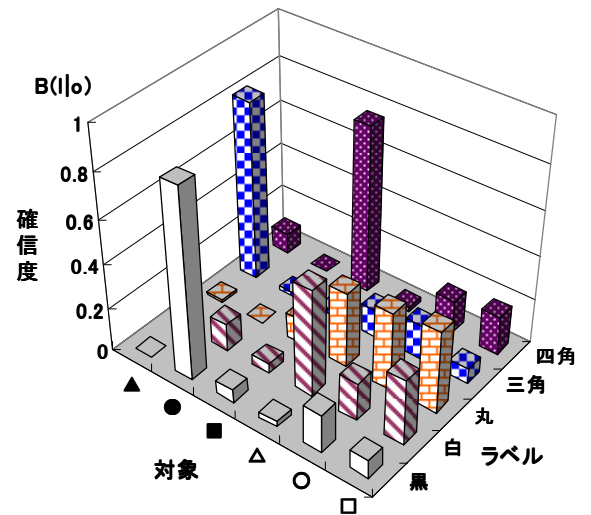


図 4 対称性バイアスモデルの確信度  $B(l|o)$

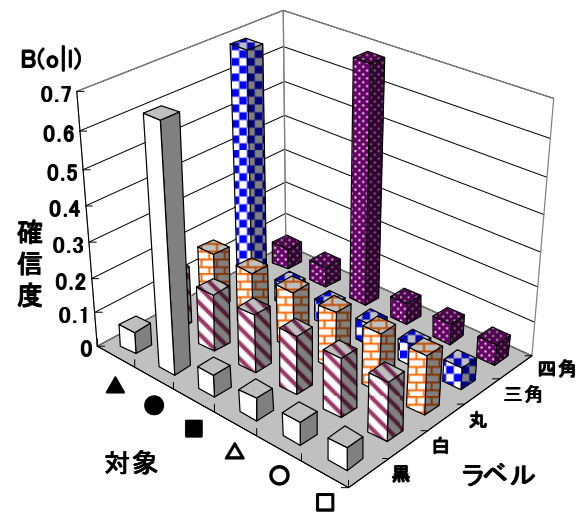


図 5 条件付確率モデルの確信度  $B(o|l)$