F-007

画像・音声刺激を用いた対話的逐次学習から獲得された知識の捨象による言語シンボル概念獲得モデル

The Concept of Linguistic symbol acquisition model by abstraction of knowledge acquired from interactive sequential learning using image and sound

椎野友博 † 荒井秀一 †
Tomohiro Shiino Shuichi Arai

1. まえがき

現在までに人工知能の分野では, 柔軟な言語理解を行う対話システムの実現や人間の知能の解明を目的として, 計算機やロボットに人間の言語や, その概念を獲得させようとする研究が数多く行われてきた [1][2]. このような言語獲得に関する研究に対して認知心理学者の Steven Pinker は, 言語とは人間が知覚し記憶した内容から獲得されるものであるにも関わらず認知心理学, 心理学等の知見が取り入れられていないことを指摘し, 認知心理学の立場から言語獲得モデルが満たすべき 6 つの条件を提示している [3]. 我々はこれまでに, Steven Pinker が提示した 6 つの条件を充足する言語シンボル概念獲得モデルを提案してきた [4]. しかし, 従来の枠組みでは, 概念の中でも, 外延, つまりカテゴリしか獲得できておらず, 言葉の意味内容である概念の内包までは獲得できていなかった. 概念の内包は, 獲得した多くの記憶の中から偶然的な性質を捨て去る (捨象) ことによって得られるものである. そこで, 本稿では, 本モデルが獲得した記憶から捨象を行うことで概念の内包を獲得する手法と, その際に必要となる尺度として画像特徴の安定性と特異性を提案する.

2. 言語シンボル概念獲得モデルの概要

本章では, 我々が提案してきた枠組みの概要を説明すると共に, Steven Pinker の 6 つの条件のうち, 特に多くの研究が満たせていない Input Condition, Cognitive Condition, Learnability Condition の 3 つを本枠組みがどのように充足しているのかを示す.

2.1.Input Condition の充足

Input Condition は, ”人間が環境から得られない刺激や刺激の量を必要としないこと” という条件である. 人間の言語, およびその概念は, 人間が独自の環世界 [5] の中で獲得するものであるため, 環世界に含まれない刺激を用いてはならない. そこで, 我々は, 本モデルが人間の環世界に反する刺激を授受することを防ぐために, 人間が環境から受容することができる刺激のみを伝達する”場”の定義した. そして, 人間や本モデルが授受する刺激を, この”場”を介したもののみに限定することで, 人間の環世界を担保した.

2.2.Cognitive Condition の充足

Cognitive Condition は, ”人間の知覚, 記憶等の認知能力に反しないこと” という条件である. 人間は, 受容した刺激をそのまま記憶するのではなく, 情報量を削減し, 抽象的な情報へと変換する”抽象化”を行い, 知覚, 記憶している. この条件を充足するためには, 本モデルが扱う画像・音声刺激の抽象化が, 人間の視覚・聴覚刺激の抽象化に則していなければならない. さらに, 人間は逐次的に記憶を蓄積し, 自らの世界を広げ, より深い知識へと変化させることができることから, 本モデルの記憶も, 逐次的に記憶を蓄え, 強化することができる必要がある.

2.2.1. 画像刺激の抽象化

人間は視覚刺激を受容すると, 視覚刺激に含まれる物体の輪郭線が抽出され, 輪郭線によって構成される領域の形状が知覚される. さらに人間は, 物体の概形だけでなく, 物体に含まれる部分領域間の位置関係や部分領域の形状等を取り込み知覚している. そこで本枠組みでは, 本モデルが受容した画像刺激から抽出された線分の集合, 大まかな概形, 親から教示された各部分領域の位置関係, そして, 画像領域の画像特徴量を抽出し, 図 1 のように 4 つの階層に分けて記憶する. これにより, 画像刺激に含まれる画像部分領域の全体部分関係を逐次的に学習し, 知識として蓄えることができる.

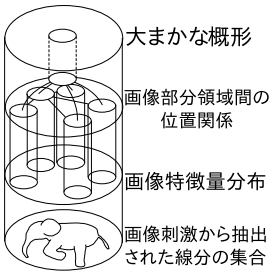

図 1: 画像記憶として保持する 4 つの内容

2.2.2. 音声刺激の抽象化

大人は, 母語の音韻カテゴリ間の違いについては敏感であるが, 母語に含まれない音韻カテゴリ間の違いには鈍感であると言われている [6]. 生後まもない乳児にも, 音声をカテゴリに分けて知覚する能力は備わっているが, 母語の音韻カテゴリは, 様々な聴覚刺激を経験していく中で, 獲得していくものと考えられている [7]. そこで, 本モデルでは, このような人間の知覚特性を状態数可変の HMM を用いてモデル化し, 学習初期の状態から音声刺激を経験していく中で, 徐々に状態数を減らし, 音韻カテゴリの知覚が行えるようにした.

2.2.3. 刺激による記憶の逐次的強化

本モデルは, 画像・音声刺激を受容すると, その刺激をこれまでに獲得してきた画像・音声記憶を用いてトップダウンに同定する. 刺激がどの記憶にも当てはまらないと判断された場合は, 未知の刺激として新たに記憶を作成し, 刺激がある記憶に同定されると, 受容した刺激を用いて記憶の強化が行われる. これによって, 本モデルは未知の刺激から世界を広げ, 深い知識へ変化させていくことができる.

2.3.Learnability Condition の充足

Learnability Condition は, ”一般的な全ての子供が言語を獲得しているのと同様に, そのモデルのメカニズムが, 自然言語を獲得することができるほどに強力であること”という条件である. 本モデルでは, 画像・音声刺激を受容したモデルが逐次的に記憶を広め, 深くしていくことができるため強力なモデルである. しかし, 記憶として獲得できているものは, 概念の中でも外延のみであり, 概念の内包までが獲得できていない. そこで本稿では, 本モデルが獲得した広く浅い知識から, もう一段階深い知識である概念の内包の獲得手法を提案する.

2.3.1. 捨象による概念の内包の獲得

概念の内包は, 経験された多くの事物の中から共通の性質を抜き出し (抽象), 個々の事物にのみ属する偶然的な性質を捨て去ること (捨象) によって獲得されるものである [8] . しかし, 捨象を行うに当たって, 画像・音声記憶単体では, どの記憶から, 何を共通の性質として取り出せばいいのかが判断できない. そこで, 本研究では人間が同時に受容した刺激間を関連づけて記憶する心理現象 (提示同時性) に倣い, 画像・音声刺激を本モデルが同時に受容した場合, それらが抽象化されて得られた記憶間にリンクを張る. これによって, 同じ

†東京都市大学大学院 工学研究科
Graduate School of Engineering, Tokyo City University

音声記憶に結びついている画像記憶がわかるため, 同じ音声記憶に結びついていることを手がかりとして, 図 2 のように, 複数の画像記憶が保持している画像特徴量分布, 画像部分領域の位置関係を 1 つにまとめ上げることができる. 本研究で

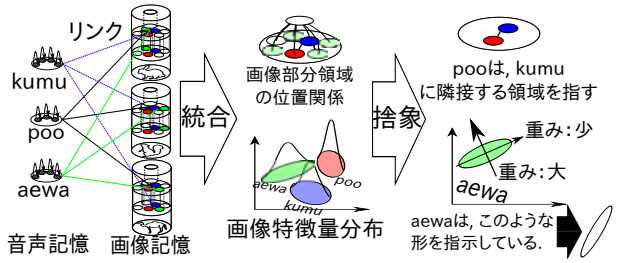

図 2: 画像記憶の統合と捨象

は, 統合後の画像記憶を捨象し共通の性質を抜き出すことで概念の内包が獲得できると考えた. 具体的には, 図 2 のように, 統合後の画像部分領域の位置関係から, 共通の性質として, 最も隣接包含確率が高いものを抜き出すことで, 「/wawae/が/kumu/に隣接する領域を指示する言語シンボルである」という概念の内包が獲得することができ, 統合正規化後の画像特徴量分布から, 固有値が小さい固有ベクトル方向に大きな重みを付けることで「/aewa/がどのような傾向を持った画像特徴を指示する言語シンボルであるのか」という画像特徴に関する概念の内包が獲得できる. しかし, 捨象を行うためには, 画像記憶の統合に用いた音声記憶が, 画像部分領域の位置関係と画像特徴量分布のどちらを指示するものであるのかが判断できなければならない. 音声記憶が画像部分領域の位置関係を指示しているか否かはその音声記憶が画像刺激中の一定の部分領域に結び付いた回数を数えることで判断できるが, 音声記憶が画像部分領域の画像特徴を指示しているか否かは, 単純な数え上げでは判断することができない. そこで, 本稿では, ある音声記憶がどれだけ強く画像部分領域の画像特徴を指示しているのかを評価する尺度を提案する.

3. 安定性と特異性の提案

ある音声記憶がどれだけ強く画像特徴を指示しているのかは, 画像特徴量分布に存在する 2 点の性質によって判断できると考えられる.

1. 安定性 : 統合後の画像特徴量分布の最小固有値の大きさ
2. 特異性 : 統合後の画像特徴量分布の重心から知識全体の重心までの距離

具体的に 1 と 2 について説明する.

1. 最小固有値が小さければ小さいほど, 統合後の画像特徴量分布の一部が安定しており, 共通の画像特徴を指示していることを示しているため, 音声記憶が画像特徴を指示しているのか否かを判断することができると考えられる.

2. 画像特徴を指示する言語シンボルは, 経験してきた事物全体からみて特異な画像特徴に結びつく. そこで, 統合後の画像特徴量分布の重心から, 知識全体の重心までのマハラノビス距離が, 画像特徴の特異さを表す尺度として利用できると考えられる.

4. 実験・結果

3 章で提案した特異性と安定性の有用性を実験によって示す. 実験に用いる画像刺激として, カテゴリ数 6, 音声刺激のカテゴリ数 10 事例, 画像刺激数を各カテゴリ 15 事例用意し, 図 3 のように教示した. その結果得られた最小固有値とマハラノビス距離を 2 次元平面上にプロットしたところ図 4 のようになった. 図 4 を見ると赤い点線で 2 つのクラスタに分類することができる. 赤い点線よりも上に存在するものが画像特徴を指示していると判断されるものであり, それぞれ/aewa/(長い),/iki/(小さい),/haiki/(細い),/poepoe/(丸い),/kumu/(胴体) の 5 つであった. それ以外の言語シンボルは/poo/(頭),/ihu/(鼻),/huelo/(尻尾),/pepeiao/(耳),/wawae/(足) の 5 つであった. /kumu/(胴体) が画像特徴を指示する言語シンボルで

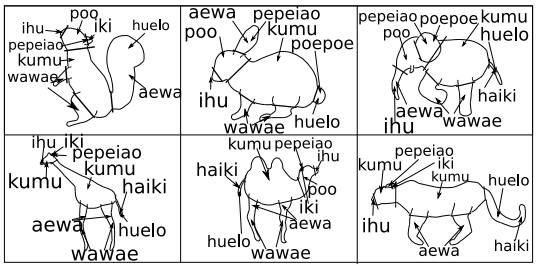

図 3: 実験に用いた画像刺激と音声刺激

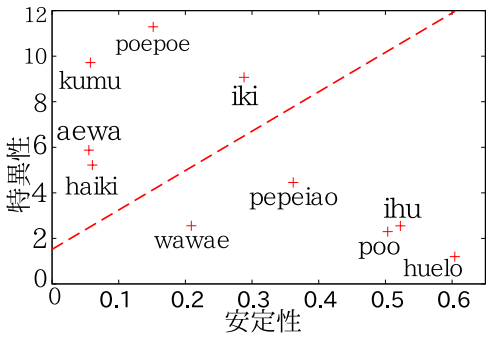

図 4: 実験結果

あると判断されたのは実験に用いた画像刺激の部分領域がすべて同じ画像特徴を保持していたためであると考えられる. これらの結果より安定性と特異性の有用性が示された.

5. まとめ

本稿では, 本モデルが捨象による概念の内包を獲得することを目指し, 捨象をする際に必要となる音声記憶が画像特徴をどれだけ強く指示しているのかを評価する尺度として安定性と特異性を提案し, 実験によってその有用性を示した. 今後は, 提案した 2 つの尺度を具体的に本モデルにどのように組み込むのかを検討する.

参考文献

[1] 安藤 義記, 中村 友昭, 荒木 孝弥, 長井 隆行, "階層マルチモーダルカテゴリゼーションによる多様な概念と語意の学習, "人工知能学会全国大会論文集 2013.

[2] 新田 恒雄, 小玉 智志, 田口 亮, 木村 優志, 入部 百合絵, 桂田 浩一, "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.

[3] steven pinker, "Formal models of language learning, "Cognition, pp.217-283 1979.

[4] 長島 徹, "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, pp.193-198 2004.

[5] 佐藤 恵子, "ユクスキュルの環世界説と進化論, 総合教育センター紀要, no.27, pp.1-15 2007.

[6] Werker,J.F & Tees, R.C "Cross Language speech perception:Evidence for perceptual reorganization during the first year of life. Behavior and Development" 7,pp. 49-63 1984

[7] Peter D. EIMAS "Auditory and linguistic processing of cues for place of articulation by infants" "Perception & Psychophysics" pp 513-521 1974

[8] 広辞苑 第四版 岩波書店