2F3-1in

画像・音声刺激による対話的逐次学習を用いた言語シンボル概念獲得モデル

The Concept of Linguistic Symbol acquisition model using interactive online learning by Image and Sound stimulus

椎野 友博　荒井 秀一
Tomohiro Shiino　Shuichi Arai

東京都市大学 知識情報工学科
Intelligent Information Technology, The Tokyo City University

Abstruct In the field of AI, many studies which pursue a mechanism of language acquisition by modeling an intelligence of human have been done. However, these studies have some unnatural point from the viewpoint of developmental psychology. Also, they couldn't model a meaning understanding of language by human since they didn't acquire a concept of linguistic symbol. In this study, we propose a model of a concept acquisition of linguistic symbol through dialogue between humans and computers using a theory of other academic.

1. はじめに

現在までに, 人間の乳幼児がどのように言語を獲得しているのかを明らかにするために, 人工知能の分野では, 乳幼児の言語獲得能力に対する仮説をモデル化した言語獲得モデルが数多く提案されてきた [安藤 13][新田 13].

ロボティクスの分野では, 言語獲得モデルをロボットに実装し, その振る舞いを観察することで乳幼児の言語獲得能力に構成論的にアプローチする研究が行われている [安藤 13]. 安藤, 中村らは, カメラやマイク, 触覚センサから入力された物体の画像, 音声, 圧力情報を用いて物体のカテゴライズを行い, 単語音声と物体のカテゴリ間を対応付けることで音声刺激が指示する概念の獲得を行っている.

しかし, この言語獲得モデルには, 音声刺激が指示する概念の獲得が充分でない. 一般に, 概念とは, "内包"と"外延"の 2 つの要素からなり, "内包"は, 経験される数多くの事例の中から, 共通の性質を抜き出し, それ以外を捨象する事で獲得される概念の内容, "外延" は, 同一本質を持つ一定範囲の事物を指している [広辞苑 第四版]. 安藤, 中村らの枠組みでは, "外延"の獲得は行えても, その概念の内容となる"内包"の獲得が行えていない.

乳幼児の言語獲得能力をモデル化し, 計算機が人間の乳幼児と同じ過程を経て言語を獲得する事は, シンボルグラウンディング問題を解決するためにも必要である [今井 03]. 人間とエージェント間で自然な対話を実現するために, 新田, 小玉らは, 言語獲得期の乳幼児が持つとされる生得的な学習バイアスを適用した概念獲得モデルを提案した [新田 13]. エージェントは教示者である人間から物体の名前や形状を指示する語の教示を受け, 逐次的に語意の獲得を行う. しかし, 抽出する画像特徴が人間の知覚に基づいていないのに加えて, 音声刺激として, キーボードから入力した文字列を用いている点で, 人間の乳幼児の言語獲得過程から離れてしまっている

このように, これまでに提案されてきた言語獲得モデルでは音声刺激が指示する概念の"内包"までを獲得できておらず, 人間の乳幼児の学習モデルとして不自然な点が数多く含まれていた. そこで我々は, 乳幼児の学習モデルとして不自然ではない, 音声刺激が指示する概念の獲得を行うことができる言語獲得モデルを提案してきた [長島 04].

連絡先: 椎野 友博, 東京都市大学 情報工学専攻 知識情報処理研究室

本稿では, 認知心理学者の Steven Pinker が提示した言語獲得モデルが満たすべき 6 つの条件を挙げて, 本枠組みがそれらを充足している事を示すことで, モデルが妥当であることを示す. そして, 概念の"内包"となる, 事物の共通の性質が, 特徴量分布の固有値・固有ベクトルによって表現されている事を示す.

2. モデルの説明

言語獲得モデルは, 人間の乳幼児が行っているものとして不自然でない事や, 人間の知覚に則っている事が求められる. 認知心理学者の Steven Pinker はこのような言語獲得モデルが満たさなくてはならない条件として以下の 6 つを提示した [Pinker 79].

Learnability Condition
一般の乳幼児が彼らのコミュニティの言語を学ぶのと同様に, そのモデルが自然言語を学習することができる程に強力であること.

Equipotentiality Condition
特定の言語圏でのみ適用可能なものではないこと. 例えば, 最初から英語の文法を仮定してしまっているモデルはこの条件を満たすことができていない.

Time Condition
一般的な乳幼児が取りうる期間内で言語を獲得することができること.

Input Condition
乳幼児が実世界から本来得られるはずのない情報や情報の量を必要としないこと.

Developmental Condition
発達心理学等の研究によって明らかになっている言語獲得期の乳幼児に見られる現象と対応が取れること.

Cognitive Condition
そのモデルによって説明されるメカニズムが, 一般に知られている人間の乳幼児の認知能力に反しないこと.

本節では, 本研究の言語獲得モデルの説明を行うと共に, 上記の Pinker の 6 つの条件を充足している事を示す.

2.1 Input Conditon の充足

言語獲得モデルは, 入力刺激として実世界から乳幼児が得ることが出来る刺激のみを用いなくてはならない. 本モデルではこの条件を満たすために, 図 1 のような実世界上から得ることが出来る音声, 画像刺激のみを伝達する事ができる ”場”を定義した. この”場”に対して, 人間や計算機は接続をし, 刺激の投入と受容を行う. この”場”には, 複数の計算機や人間が接続する事が可能であり, 投入された刺激をその場に接続している全員と共有する役目を負っている.

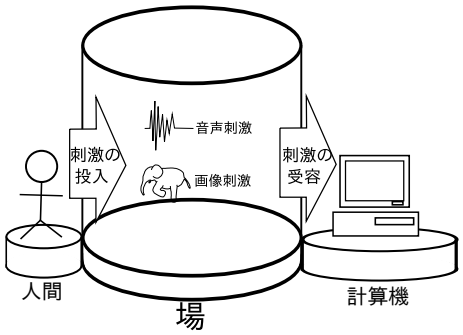

図 1: ”場”のイメージ

2.2 受容した刺激の抽象化と記憶

Cognitive Condition を満たすためには, モデルによって説明されるメカニズムが人間の乳幼児が持っている認知能力に反してはいけない. ここで定義される認知能力とは, 目や耳等の感覚器官から得られた刺激を視覚, 聴覚によって知覚し, 記憶と同定する能力や新たに記憶する能力の事である.

本研究では, 人間が機能として保持している視覚, 聴覚刺激を知覚, 記憶する能力をモデリングし計算機に持たせる事で, Cognitive Condition を満たした. 以下にその詳細を記す.

2.2.1 画像刺激の抽象化と記憶

親が物体を乳幼児に指し示して教示を行う場合, 乳幼児は, 親が指し示した物体を目に写るシーンから抜きだす必要がある. しかし, 今回は, モデルの簡略化のために, 物体は既に分節済みとし, 入力刺激として使用する画像は線画像とした. これは, 親が, 乳幼児に物体が描かれた絵本の中を指差して教示を行っている状況と同じであり, 特異な状況ではないため, ”Equipotentiality Condition”を満たしている.

乳幼児が線画を見たときに受容する視覚刺激である光は, 眼球から網膜に投射され, 電気信号へと変換された後に, 視覚野へと伝播する. 視覚野は, 一次から四次視覚野と呼ばれる領域に分かれ, それぞれで異なった役割を担っている.

本研究では, 視覚刺激が大脳で知覚される過程に則って, 計算機が受容した画像刺激を抽象化し記憶する流れを以下のように定義した.

1:セグメント化

一次, 二次視覚野では, 物体に存在する曲線や直線等の局所的なエッジの抽出が行われている [岡田 01]. 本研究では, 入力画像刺激から得られる輪郭線のエッジを”曲線”, ”直線”,”T 字分岐”の 3 つで近似し抽出する.

2:領域形成

視覚野で抽出された物体のエッジは, より高次の視覚野へ伝播すると統合され輪郭線を形成する. しかし, これによって形成される輪郭線は, 物体の全体を指すものだけであるとは限らない. 人間は物体全体だけではなく, 物体に属する部分領域の形状と位置関係を脳内に取り入れている. これと同様に, 本研究では, セグメント化によって抽出されたセグメント群を統合し, 親から教示された, ”足”,”胴体”,”頭”等のパーツ毎に部分領域を形成する.

3:画像特徴抽出

輪郭線が形成されると, 四次視覚野に隣接している下側頭葉と呼ばれる領域で形状の知覚が行われる [土井 04]. 同様に, 本研究では, 形成された各部分領域から”長さ”,”幅”,”周囲長”,”面積”,”最外郭距離”の 5 次元の特徴量を抽出する.

4:記憶

視覚刺激の抽象化が完了すると, 下側頭葉内で記憶が行われる. ここで記憶された内容は, 次回以降物体を見た時の解釈に利用される [岩井 91]. 本研究では, 画像記憶の抽象化によって得られた記憶を図 2 のように 4 階層のシリンダでモデル化した.

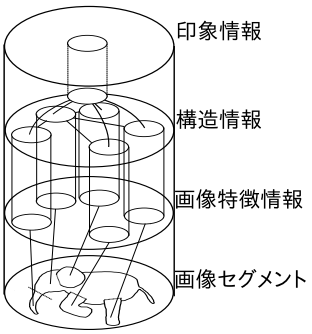

図 2: シリンダで表現される画像記憶

シリンダの各階層の詳細を以下に記す.

印象情報

一次視覚野で物体の輪郭線の傾きの変化が捉えられるのと同様に入力画像刺激の輪郭線の傾きの変化から偏角関数を求め, HMM を作成して記憶する.

構造情報

親から教示された部分領域間の隣接・包含関係を記憶する.

画像特徴情報

画像の部分領域から得られた 5 次元の特徴量を用いて 5 次元の分布を形成し, 保持する.

画像セグメント

入力画像刺激のセグメント化によって得られたセグメント群を記憶する.

2.2.2 音声刺激の抽象化と記憶

人間は聴覚刺激から抽出される情報の中でも特に音の「大きさ」,「高さ」,「スペクトル構造」, 「時間情報」に敏感であることが知られている. 本研究では, これら 4 つの特徴量を, 「16 次元の LPC ケプストラム」, 「パワー」, 「ピッチ周波数」,「有声/無声音の度合い」として合計 20 次元の特徴量ベクトルを抽出し, HMM を作成して記憶を行う. 音声刺激から抽出される特徴量は言語圏に依存したものではないため, ”Equipotentiality Condition” を満たしている.

2.3 記憶の強化

外界から受容した刺激が脳内で知覚されると記憶との照らし合わせが行われる. 刺激が記憶と同定されると, 記憶の強化が行われ, 経験として蓄積されていく.

これと同様に, 本研究では, 計算機が受容した画像・音声刺激が一つの記憶に同定された場合に記憶の強化学習を行う.

2.4 画像記憶と音声記憶の結びつけ

人間が同時に受容した視覚と聴覚刺激は, 脳内で結びつけられた状態で記憶される [Barbara 98]. 本研究でも同様に, "場" に対して画像と音声刺激が同時に投入され, 計算機が同時に受容した場合, 両メディアの刺激の抽象化によって形成された記憶の間にリンクを張る. これにより, 片方のメディアの刺激が知覚され, 記憶と同定された時に, もう一方のメディアの記憶が想起される "認識" を行う事が出来るようになる.

以上のように人間が生得的に機能として持っている刺激の抽象化能力と, それらを結びつける能力のみを持たせた非常にシンプルなモデルであるため, 複雑な処理を必要とせず, "Time Condition" が満たせる.

3. 獲得した概念を用いた表象

前節までの内容で, 画像刺激と音声刺激の抽象化と記憶の形成, 強化, リンクの学習までが行う事が出来る. 異種メディア間にリンクを張ることで, 同じ音声記憶と結びついている画像記憶から共通の性質を抜き出すことが出来る. 本節では, 本言語獲得モデルによって行わせる, 概念を用いた表象の定義を行う.

3.1 表象を行わせる必要性について

ユクスキュルが唱えた "環世界説" で示される通り, 人間を含め, 全ての生き物は, 感覚器官によって知覚される "知覚世界" と, 声や手によって作用する事が出来る "作用世界" が連環しあう, 完結した 1 つの "環世界" の中で生活している [佐藤 07]. そのため, 教示者である親は, "環世界" を無視して, 乳幼児の知識体系を直接覗き見ることも, 獲得された概念を見ることも出来ない.

実世界上の親は, 乳幼児がある事物に関する概念を獲得し, 理解することが出来たという事を乳幼児が発した声や手ぶり身振りで判断するしかない. これと同様に, 本枠組みでも, 概念の獲得が行えたという判断を, 計算機が教示者の問いかけに対して表象した内容を用いて行う. 逆にいえば, 表象が出来なければ概念が獲得出来たとは言えない.

教示を行った事物の概念が学習者に獲得された事を教示者が確認できるパターンは数多く存在しているが, それらは大まかに, 以下の 3 つのパターンに分類されると考えられる.

パターン 1 大人が聞いて, "長い" と感じる物体を見た時に学習者が "nagai" と音声刺激を発する.

パターン 2 大人が聞いて, "足" だと感じる部位を見た時に学習者が "ashi" と音声刺激を発する.

パターン 3 大人が聞いて, "長い形状" を指していると感じる音声刺激を聞いた時に, 学習者が, 大人が見て "長い" と感じる絵を描く.

パターン 4 大人が聞いて, "足" を指していると感じる音声刺激を聞いた時に, 学習者が, 大人が見て "足" だと感じる絵を描く.

これら 4 つのパターンで示されるような表象が行えた時, 計算機は, 概念を獲得することが出来たと言える. 本稿では, パターン 1 と 2 のように未知の画像刺激に対して, 大人が納得するような表象を行う事に焦点を当てる. パターン 2 の推論は, 構造情報の隣接包含関係によって容易に行えるが, パターン 1 のように, 未知の画像刺激に対して, その画像刺激の形状を指示するような音声刺激を発声させる研究はまだ行われていない. そこで, 形状を指示する音声刺激を未知の物体の形状に対して, 音声刺激を発するために, モデルが持たなくてはならない機能を定義する.

3.2 音声刺激が指示する形状概念の獲得

本モデルでは, 画像刺激を画像記憶と同定する際に, 画像記憶として保持している各部分領域の特徴量分布の重心から画像刺激から抽出された画像特徴量までのマハラノビス距離を算出し, 入力画像刺激と画像記憶との適合度として用いている. 3. 節の例 3 のように, 未知の物体の形状について, "nagai","hosoi","ookii" 等の音声刺激を発声するためには, 音声記憶と画像記憶間に張られたリンクによって, 同一の音声記憶に結びついている複数の画像特徴量分布を統合, 共通の性質以外を捨象し, その音声記憶が指示する形状がどのような傾向を持つものなのかを推定出来なければならない.

本研究では, 音声記憶が指示する形状の共通の性質が統合後の画像特徴量分布が持つ, 固有ベクトルと特徴量分布の重心位置によって, 表現されると考えた.

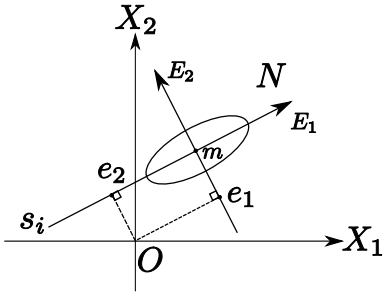

図 3: 空間 s_i 内における原点 O までのマハラノビス距離

二次元の座標系 N 内の, ある音声記憶に結びついている 5 次元の画像特徴量分布を統合した分布 k を作成後, 全音声記憶について画像特徴量分布を統合し, それらの特徴量分布の平均的な共分散行列を用いて, 分布の正規化を行う. 正規化を行った分布に対して主成分分析を行い, 得られた 5 本の固有ベクトルで, 図 3 で示される新たな座標系 S_i を定義する. この時, 座標系 N から見た原点 O^N の座標は $(0,0)$, 座標系 S_i から見た原点 O^{S_i} は, (e_1, e_2) となる. 点 m から座標系 S_i における全世界の重心 O^{S_i} までのマハラノビス距離は, 式 1 で定義される.

$$D_{O^{S_i}} = \sqrt{(O^{S_i} - m)\Sigma^{-1}(O^{S_i} - m)} \quad (1)$$

$$= \sqrt{(O^{S_i})\Sigma^{-1}(O^{S_i})} \quad (2)$$

Σ^{-1} は分布 k の各固有ベクトルの固有値 λ_i を用いて式 3 のように定義される.

$$\Sigma^{-1} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{\lambda_1} & 0 \\ 0 & \frac{1}{\lambda_2} \end{pmatrix} \quad (3)$$

式 3 を式 1 に代入して整理すると式 4 のようになる.

$$D_{O^{S_i}} = \sqrt{\frac{e_1^2}{\lambda_1} + \frac{e_2^2}{\lambda_2}} \quad (4)$$

この式 4 から, マハラノビス距離とは分布 k が持つ固有ベクトル上の分散に相当する固有値 λ_i が大きくなればなるほど, マハラノビス距離 $D_{O^{S_i}}$ に及ぼす影響が小さくなっていくような距離尺度である事が分かる. つまり, 分布 k の分散の大きい固有ベクトル方向に弱い重みが与えられることにより, 自動的に捨象が行われていることを示している.

3.3 実験

音声刺激に結びつけられた画像特徴分布を統合し, 各固有ベクトルがどれだけマハラノビス距離にどれだけ影響を与えているのかを式 5 で定義される貢献度で評価した.

$$C^{S_i} = \frac{\frac{e_i^2}{\lambda_j}}{\sum_{j=0}^{n} \frac{e_j^2}{\lambda_j}} \tag{5}$$

マハラノビス距離が, 特徴量分布の固有値の大きい方向に軽い重みをつけた距離尺度であるならば, 特定の”形状”指示する音声刺激に結びつけられた分布の各主成分の最低貢献度は, ”名前”を指示するものにくらべて, 低くなるはずである. そこで, 親が子供に対して絵本に描かれている動物の線画を指差しながら教示を行っている状況を想定し, 教示に用いる画像刺激として”ゾウ”,”ウサギ”,”リス”,”ラクダ”の画像各 30 枚を用意し, 図 4 のように教示を行った.

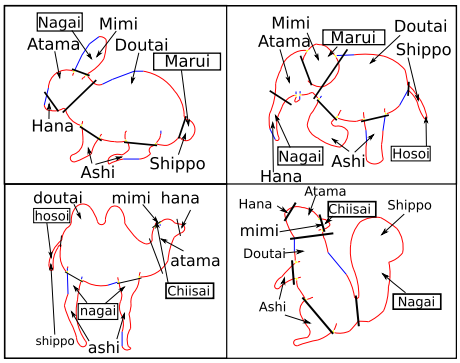

図 4: 実験に用いた画像刺激と音声刺激の組み合わせ

これらの教示後, 式 5 を用いて, 各主成分の貢献度を求めた.

表 1: マハラノビス距離における各主成分の貢献度

音声刺激	第 1 主成分	第 2 主成分	第 3 主成分	第 4 主成分	第 5 主成分
ashi	0.085081	0.027104	0.014920	0.386403	0.486491
atama	0.003945	0.045065	0.041008	0.039298	0.870684
doutai	0.00009	0.462234	0.120006	0.245189	0.172475
hana	0.070801	0.000001	0.002204	0.920041	0.006953
mimi	0.021064	0.015366	0.065816	0.083496	0.814258
shippo	0.462971	0.067555	0.263344	0.198678	0.007453
chiisai	0.074370	0.115574	0.044074	0.000022	0.765960
hosoi	0.057161	0.156763	0.000395	0.697779	0.087902
marui	0.000783	0.001745	0.037295	0.301281	0.658896
nagai	0.011496	0.031800	0.040374	0.824162	0.092167

3.4 結果

表 3.3 を見ると, 大人が聞いて形状を指示していると解釈される音声刺激の方が, 名前を指示していると解釈される音声刺激よりも, マハラノビス距離における各主成分の貢献度の最低値が小さくなっているものが多い. 例えば, 音声刺激”chiisai”を見ると, 各主成分の最低貢献度は, 0.000022 なのに対し, ”atama”は, 0.003945 と非常に高い. このように, 主成分分析によって, 明確に次元を削減しなくても, マハラノビス距離による刺激の同定を行っている時点で, 重み付けによる捨象が行えている事が分かる.

しかし, 一方で”doutai”は, 最低貢献度が, 0.000097 と非常に小さな値を取っている. これは, 教示に用いた動物の胴体が, どれも類似した形状をしていたためである. このように, 音声刺激の意味を教示者が意図していないものに捉え, 汎用してしまう現象は言語獲得期の乳幼児にも見られるものであり, 過拡張とよばれている. 実験によって, 親の教示の仕方によって, 音声刺激を本来大人が意図していない意味として捉えていく様子が観測され, Developmental Condition も充足することが出来ている.

4. おわりに

これまでに我々は, 乳幼児がどのように言語を獲得しているのかを明らかにするために, 音声と画像刺激を用いた教示による言語シンボル概念獲得モデルを提案してきた. 本稿では, 人間の乳幼児が実際に行っている学習の姿として不自然ではない言語獲得モデルであることを示すために, 認知心理学者の Steven Pinker が提示した, 言語獲得モデルがみたさなくてはならない, 6 つの条件を, 本言語獲得モデルが充足している事を示した. さらに, 経験される多くの事例の中から共通の性質を抜きだし, 概念の中でも”内包”の獲得を行う手法を提案し, 実験による有用性を示した.

参考文献

[安藤 13] 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.

[広辞苑 第四版] 広辞苑 第四版 岩波書店

[新田 13] 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.

[Pinker 79] Steven Pinker, ”Formal models of language learning, ”Cognition, pp.217-283 1979.

[今井 03] 今井 むつみ, ”言語獲得におけるシンボルグラウンディング. ”人工知能学会誌, pp.580-585 2003.

[長島 04] 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.

[佐藤 07] 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.

[岡田 01] 岡田 真人, ”大脳皮質視覚野の情報表現を眺める, ”統計数理, vol.49, no.1, pp.9-21 2001.

[土井 04] 土井 泰次郎, 藤田 一郎, ”形状知覚と物体認識における側頭葉視覚連合野の役割, ”神経進歩, vol.48, no.4 月, pp.176-184 2004.

[岩井 91] 岩井 栄一, 渡辺 譲二, 阿山 みよし, ”形の認識と下部側頭葉皮質, ”VISION, vol.3 1991.

[Barbara 98] Barbara A.,Morrongiello, ”Crossmodal learning in newborn infants:Inferences about properties of auditory-visual events, ”Infant Behavior and Development, pp.543-553 1998.