

# 画像・音声による刺激の提示同時性を利用した 言語シンボル概念獲得モデル\*

5X-2

黒田 和史 伊東 洋一郎 荒井 秀一 石川 知雄†  
武蔵工業大学‡

## 1 はじめに

近年、計算機上に言語概念を構築する研究が多数行われている。これらの多くはシンボルの意味や関係を前提知識として与え、その上で概念の獲得を行うシンボリックな処理を行っている。

一般的に概念とは経験される多くの事物に共通の内容を抽出して抽象化を行い、個々の事物にのみ属する偶然的な性質を捨て、捨象することである。ここで言語シンボルについて考えてみると、言語シンボルは概念を指し示すラベルであり、概念を象徴化したものであるといえる。このことから言語シンボルは概念に先立って形成される物ではなく、後天的に獲得される物であるため、言語概念を獲得する際に言語シンボルを用いる事はシンボルの変換を行っているにすぎず、言語概念を獲得しているとは言えない。

そこで本研究では、シンボリックな処理を行わずに、外部からの刺激である、画像・音声刺激とその提示同時性により言語シンボル概念を獲得するモデルを提案する。

## 2 言語シンボル概念

本研究において獲得を目的としている言語シンボル概念について、図 1 の概念図を元に説明する。

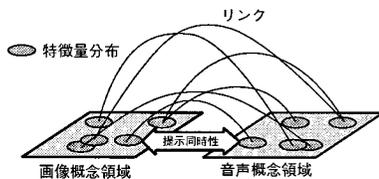


図 1: 提示同時性による言語シンボル概念の表現

まず、外部から入力された画像・音声刺激を各々抽象化し、特微量の分布で表現する。この画像・音声それぞれの特微量分布を本研究では画像概念、音声概念と呼ぶことにする。ここで画像および音声という複数のメディアを扱っているのは単一のメディアからだけでは一つの概念に特定できないからである。たとえば、同意義語のように単一の音声情報に対して複数の概

念が存在する場合である。このような場合には複数のメディアからの情報を関連づけることによって、一つ概念に特定できると考えられる。

そこで本研究では、同時に与えられた刺激には何らかの関係があるという提示同時性を利用することにより画像概念と音声概念との間にリンクを張ることにし、そのリンクも含めた画像・音声概念をまとめて言語シンボル概念と定義する。

また、知識獲得という分野においては、知識の獲得手法、表現方法はもちろん、獲得した知識をどのように利用できるかという、知識の再利用性も重要である。本研究においても、獲得した知識により入力刺激を認識して判断し、逐次学習を行いたいので、知識の再利用が可能でなければならない。また、完全に同一でなくとも類似した刺激を同一な刺激であると見做せる表現形態でなければならない。このような条件を満たす知識の表現形態として、音声認識の分野で広く用いられている HMM (Hidden Markov Model) がある。これを画像・音声それぞれに適応し、概念を表現することとする [1]。

### 2.1 画像概念

人間は何らかの物体を目にしたとき、大雑把な全体の印象情報 (概形) の認識をし、明確に認識できない場合には徐々に細部の詳細な部分情報を用いた認識へとトップダウン的な処理を行っている。本研究でも画像概念を階層的な表現形態を用いて記述し、画像全体の比較から部分へ、すなわち概略から詳細へ順次比較を行なえるようにした。

また各階層は以下のように定義した。

**印象情報** 物体の輪郭から、輪郭上の位置とその位置での角度の差分を表現する偏角関数から HMM を作成し、物体全体の概形を表現する。

**構造情報** 物体を部分領域に分割し、部分領域間の関係によって構造を表現する。部分領域は領域の分割線を教示する事によって学習し、複数の分割線を教示する事によって詳細な知識の記述を可能とする。

**形状特徴情報** 部分領域毎に形状的な特微量 (面積・骨格長・幅・周囲長・最外殻距離・部分領域の偏角関数から作成した HMM) を用いて物体の詳細な形状を表現する。

**セグメント画像** 入力画像をセグメント化する事によって作成し、上記の各情報の生成に用いる。

\*Knowledge Acquisition Model of Linguistic Symbol Using Synchronization of Image and Voice Stimulus

†Kazufumi KURODA, Yoichiro ITO, Shuichi ARAI, Tomoo ISHIKAWA

‡Musashi Institute of Technology

2.2 音声概念

本研究では単語発話音声を入力刺激とし、その抽象的な表現形態の知識である音声概念をHMMによって表現する。具体的には、音声刺激から特徴量(LPCケプストラム10次元・パワー・ピッチ周波数・有声/無声音の度合)を算出し、HMMを作成することで概念を構築する。そして構築された音声概念と刺激とのviterbi演算により比較、識別を行う。

3 言語シンボル概念の獲得

本研究では言語シンボルの意味の獲得を目的としているため、前提知識として言語の情報を利用することはできない。そこで、受容した刺激のみから言語シンボルの意味を獲得している人間の知識獲得過程に着目し、人間が先天的に保持している、物を見分ける、音を聞き分けるといった能力、学習、認識に関する方法論的知識を最低限与えることとした。

これらの前提知識により画像・音声刺激から言語シンボル概念を獲得する過程を図2に示し、これを元に説明する。

(a) 抽象化

同時に提示された画像・音声刺激を各々抽象化し特徴量を算出する。(2.1, 2.2参照)

(b) 画像刺激及び音声刺激の認識

獲得されている各概念と入力刺激との比較を行う。比較において構造的な矛盾や、明らかに形状が異なる場合は、その概念には適合しないものとして除外されるが、構造的にも矛盾がなく、形状も似ているような場合には適合するものとして認識結果の候補となる。

ここで、複数の候補の序列を定める必要がある、そのための尺度が必要になる。本研究ではこの尺度として、受容刺激の特徴量と知識との距離から確信度を生成して、これを適合確率として用いることにした[2]。また、全ての概念と適合しなかった場合には、未知の刺激であると考えて、新規の概念を生成することとした。同様に、比較可能な概念が全く存在しない場合にも新規の概念を生成する

ことで、獲得知識がゼロの状態からの概念獲得が可能とした。

(c) 全体の認識

この時点で画像・音声刺激は各々認識され、その認識結果は確信度によって序列が定まった候補群となっている。そこで、次に画像・音声間のリンク情報を用いて、その組合せを見ながら総合的な判断を行う。ここで、新規の概念同士の組合せにはリンク情報が存在しないため、リンク情報だけを見て判断することは問題がある。そこで、次のような判断の優先順位を考えた。

1. 画像・音声概念共に新規に作成された場合は、全く初めて見聞きした概念であるためこの関係を認識結果とする。
2. 画像・音声概念のいずれかが新規に作成された場合は、初めて見聞きした概念と既知の概念の関係を示すものであるから、新規の概念既知の概念側の最も確信度の高い候補との関係を認識結果とする。
3. 画像・音声各々の確信度が非常に高い場合には、その間にリンクが存在しなくても、既知の概念同士の新たな関係であると考えられるため、これを認識結果とする。
4. それ以外については、リンクの存在する組合せの中で総合的な確信度が最も高い関係を認識結果とする。

(d) 学習

最後に認識結果をもとに、各々の概念に学習を行う。この学習は、刺激そのものの学習と画像・音声間のリンクの学習である。既存のリンクが存在しない場合には新たなリンクを張り、既知のリンクに対しては強化学習を行う。このように、獲得された知識は次の入力刺激の認識に再利用されることになる。

4 おわりに

言語シンボルは抽象的な表現で、多くの情報を含んでいるため複雑な概念を容易に表現できるが、具体的な事象を生成することが困難であった。

それに対し本研究では画像・音声刺激という具体情報の提示同時性を利用し、シンボリックな処理を行うことなく言語シンボル概念を獲得する枠組を提案し、より柔軟な知識獲得を行うことを可能とした。

参考文献

[1] 小松, 荒井: “計算機間対話による物体概念学習のためのHMMを用いた知識表現法”, 信学技報, AI97-19, (1997-07)  
 [2] 小松, 荒井: “エージェントの確信度を用いた画像・音声情報からの物体概念獲得”, 信学技報, AI96-32, (1997-01)

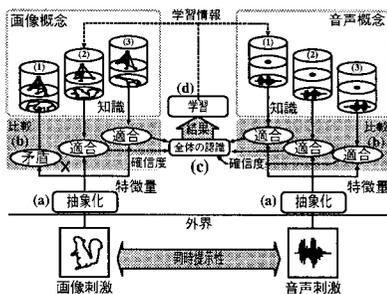


図 2: 刺激の提示同時性を利用した概念の獲得