HMM-Based Semantic Learning for a Mobile Robot

Kevin Squire

Language Acquisition and Robotics Group
University of Illinois at Urbana-Champaign

Adviser: Stephen E. Levinson
Language Learning
Language Learning ... by Robot!
Philosophy of Language Acquisition

Fundamental Ideas:

- The Language Engine is primarily semantic, not syntactic.
- There is no such thing as a disembodied mind.
  - Language and meaning is acquired through interaction with the real world.
  - Sensory-motor function is essential for human-like cognition.
- Mental processes are largely based on associative memory and learning.
Ph.D. Background and Research

1. Infrastructure Development
   - Hardware
   - Software

2. Research
   - Semantic Learning
   - HMM Cascade Model
1. Infrastructure Development
   - Hardware
   - Software
2. Research
   - Semantic Learning
   - HMM Cascade Model
Robot Hardware
Robot Hardware

Modifications:

- Added
  - cameras
  - microphones
  - on-board computer
  - wireless transmitter

- Miscellaneous structural changes
- Replaced power supply, rewired to supply power to all components
- Installed Linux
Robot Hardware

Modifications:

- Added
  - cameras
  - microphones
  - on-board computer
  - wireless transmitter

- Miscellaneous structural changes
  - Replaced power supply, rewired to supply power to all components

- Installed Linux
Robot Hardware

Modifications:

- Added
  - cameras
  - microphones
  - on-board computer
  - wireless transmitter
- Miscellaneous structural changes
- Replaced power supply, rewired to supply power to all components
- Installed Linux
Distributed Communications
Distributed Computing Framework

Illy (robot)

- Audio Source (Sound Card)
- Audio Server (Sink)
- Sound Source Location (Sink)

Hal (workstation)

- Audio Source (Remote)
- Audio Server (Sink)
- Audio Ring Buffer
- Speech Recognition (Sink)
Distributed Computing Framework

Allowed the integration of:

- Sound source localization (D. Li)
- Vision based navigation & learning (W. Zhu)
- Speech recognition (Q. Liu/R.S. Lin)
- Simple working memory (K. Squire)

Next steps:

- Centralized controller (M. McClain)
- **Semantic learning (K. Squire)**
Ph.D. Background and Research

1. Infrastructure Development
   - Hardware
   - Software

2. Research
   - Semantic Learning
   - HMM Cascade Model
Ph.D. Background and Research

1. Infrastructure Development
   - Hardware
   - Software

2. Research
   - Semantic Learning
   - HMM Cascade Model
Cognitive Framework
Semantic Memory

- Concept Model
  - Visual Model
  - Auditory Model
- Semantic Memory
  - Sensory Inputs
  - to Working Memory
- Procedural Memory
- Semantic Memory
- Episodic Memory
- Associative (Long term) Memory
Concept: abstract symbol associated with symbolic representations in the various senses
Semantic Memory

Semantic Memory

Concept Model

Visual Model

Auditory Model

Sensory Inputs

to Working Memory
Cascade of HMMs

Semantic Memory

Concept Model

Visual Model

Auditory Model

Sensory Inputs

to Working Memory

\[ \begin{align*}
S &= \begin{cases} 
1 & \text{if } x_n \text{ is visual input} \\
2 & \text{if } x_n \text{ is auditory input}
\end{cases} \\
\hat{y}_n &= \langle x_{n}^{\text{vis}}, x_{n}^{\text{aud}} \rangle
\end{align*} \]
Hidden Markov Models (HMMs)
Hidden Markov Models (HMMs)
Hidden Markov Models (HMMs)
Hidden Markov Models (HMMs)
Maximum-Likelihood Estimation (1)

Traditional methods (Baum-Welch reestimation):

- Let $p_n(y_1, \ldots, y_n; \varphi)$ be the likelihood of observations $(y_1, \ldots, y_n)$ given HMM $\varphi$.

- Maximize $p_n(y_1, \ldots, y_n; \varphi)$ by solving

  $$\nabla \varphi p_n(y_1, \ldots, y_n; \varphi) = 0.$$  

- Implemented as an Expectation-Maximization (EM) procedure.

- Requires all of $(y_1, \ldots, y_n)$ be available.
Maximum-Likelihood Estimation (2)

Recursive (stochastic gradient) procedure (LeGland and Mevel):

- Rewrite \( p_n(y_1, \ldots, y_n; \varphi) \) as a sum

\[
\log p_n(y_1, \ldots, y_n; \varphi) = \sum_{k=1}^{n} \log b(y_k; \varphi)' u_k(\varphi)
\]

where

\[
b_i(y_k; \varphi) = p(y_k|x_k = i)
\]

\[
u_{ki}(\varphi) = \Pr(x_k = i|y_1, \ldots, y_{k-1}).
\]

- Update parameters \( \varphi \) at time \( n + 1 \) with

\[
\varphi = \varphi + \varepsilon \left( \nabla \varphi \log b(y_{n+1}; \varphi)' u_{n+1}(\varphi) \right)
\]
Cascade of HMMs
Semantic Learning Simulation

Kevin Squire—Lincoln Laboratory Interview, 6 June 2005 – p.22
Simulation Results

<table>
<thead>
<tr>
<th></th>
<th>Maximum Likelihood Classification</th>
<th>Viterbi Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\phi}^a$</td>
<td>90.1%(3.7%)</td>
<td>89.9%(4.2%)</td>
</tr>
<tr>
<td>$\hat{\phi}^v$</td>
<td>97.9%(1.6%)</td>
<td>99.1%(2.4%)</td>
</tr>
<tr>
<td>$\hat{\phi}^c$</td>
<td>98.4%(1.1%)</td>
<td>98.8%(1.1%)</td>
</tr>
</tbody>
</table>

Average classification accuracy for the learned models over 50 runs. The number in parenthesis is the standard deviation.
Robot Implementation

The robot should:

- Recognize visual inputs
- Recognize auditory inputs
- Learn concepts
Cascade of HMMs as an Associative Memory

Cascade Model:
Cascade of HMMs as an Associative Memory

Auditory-only Classification:
Cascade of HMMs as an Associative Memory

Visual-only Classification:
Cascade of HMMs as an Associative Memory

Audio-Visual Learning:

\[
\phi \xrightarrow{\text{vis}} x_n \xrightarrow{\text{con}} y_n \xrightarrow{\text{aud}} x_n
\]

\[
\phi \xrightarrow{\text{vis}} \hat{x}_n \xrightarrow{\text{con}} y_n = \{x_n^{\text{vis}}, x_n^{\text{aud}}\} \xrightarrow{\text{aud}} \hat{x}_n
\]

\[
\phi \xrightarrow{\text{vis}} \hat{x}_n \xrightarrow{\text{aud}} y_n
\]
## Robot Demonstration

<table>
<thead>
<tr>
<th>Visual Objects</th>
<th>Words</th>
<th>Concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Cat" /></td>
<td>cat</td>
<td>cat</td>
</tr>
<tr>
<td><img src="image2" alt="Dog" /></td>
<td>dog</td>
<td>dog</td>
</tr>
<tr>
<td><img src="image3" alt="Red Ball" /></td>
<td>red ball</td>
<td>red ball</td>
</tr>
<tr>
<td><img src="image4" alt="Green Ball" /></td>
<td>green ball</td>
<td>green ball</td>
</tr>
<tr>
<td><img src="image5" alt="Ball" /></td>
<td>ball</td>
<td>ball</td>
</tr>
<tr>
<td><img src="image6" alt="Animal" /></td>
<td>animal</td>
<td>animal</td>
</tr>
</tbody>
</table>
Robotic Controller

1–Explore
- Explore
- Speech: "illy"/ turn toward sound
- Silence (timeout)
- Run into object
- Lost object
- Object far/ approach object
- Learn object
- Object near
- Speech/ repeat & learn
- Silence (timeout)
- Interact
- Found nothing or wrong object/ say name
- Found desired object/ say name
- Search for object
- Hear known object/ search for object
- Unknown speech/ beep
- 6–Interact
- 7–Search
- 5–Play 2
- 4–Play 1
- 3–Learn Name
- 2–Found Object
- 1–Explore
Video
Conclusion

1. Built a platform upon which to conduct language acquisition research.

2. Proposed a general model of semantic concept learning.

3. Successfully implemented this model in a real robot using a cascade of HMMs.
Future Directions/Interests

1. Grow/shrink models (combine/split states) as appropriate (e.g., use minimum description length (MDL) or related measures).

2. Apply associative learning to spatial/action information, other modalities.

3. Study sentence comprehension (e.g., combine syntactic and semantic information).

4. Incorporate reinforcement into current unsupervised training scheme.
Questions?
Robot Visual HMM

- Initialization: K-means on labeled data
- Online Learning with RMLE
- Features:
  - Color histogram
  - Moment
  - Height/width ratio
- Fixed number of classes
Robot Auditory Model

- 3 State HMM
  - States: Silence, Voiced, Unvoiced
  - Features: 3 log-area ratios + log-energy + voicing
  - Online update with RMLE possible
- “Word Recognizer”
  - Features: histogram of audio states + word length
  - Can distinguish some words
  - Initial training offline with RMLE
Robot Concept HMM

- 6 State Discrete HMM
- Observations: states of Visual and Auditory models
- Initialized offline
- Online update with RMLE
RMLE Convergence (1)

Conditions for RMLE Convergence:

- The transition probability matrix $A(\varphi^*)$ is aperiodic and irreducible.

- The mapping $\varphi \rightarrow A(\varphi)$ is twice differentiable with bounded first and second derivatives and Lipschitz continuous second derivative. The mapping $\varphi \rightarrow b(y_k; \varphi)$ is three times differentiable, and the function $b(y_k; \theta)$ is continuous on $\mathbb{R}$ for every $\theta \in \Theta$. Alternately, for $y_k$ drawn from a finite alphabet, the mapping $\varphi \rightarrow b(y_k; \varphi)$ is twice differentiable with bounded first and second derivatives and Lipschitz continuous second derivative.
Conditions for RMLE Convergence:

- Under $P_{\varphi^*}$, the extended Markov chain

$$\{X_n, Y_n, u_n(\varphi), w_n(\varphi)\}$$

is geometrically ergodic (see LeGland and Mevel, '96).

Because of this geometric ergodicity, the initial values of $u_0(\varphi)$ and $w_0(\varphi)$ are forgotten exponentially fast.