Outline

• Motivation: affordances
  – Modeling
  – Learning

• Using the model
  – Imitation games
  – Task learning
  – Mirror an canonical neurons

• Extensions: multimodal perception (words)

• Discussion
Outline

• **Motivation: affordances**
  - Modeling
  - Learning

• Using the model
  - Imitation games
  - Task learning
  - Mirror an canonical neurons

• Extensions: multimodal perception (words)

• Discussion
What are affordances?

Affordances Definition

“The Affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill” (Gibson, 1986)
What are affordances?
Robotic Affordances

- Relations between actions, objects and effects
- Involves action and perception
- **Bayesian networks to model relations**
- Learn through experimentation or observation

![Diagram of robotic affordances]

Mathematical model: BN

- We use **Bayesian Networks** to represent affordances

- Nodes are:
  - Actions and action parameters
  - Object properties
  - Resulting effects

- BNs provide a unified and sound probabilistic framework for learning and using affordances
Learning Bayesian Networks

- Use a set of acquired data $D$ to learn the network

Prior knowledge

Structure Learning

Parameter Learning

$G^* = \arg\max_G p(G \mid D) = \arg\max_G \eta p(D \mid G)p(G)$

- Likelihood of the data given the model

$$p(D \mid G) = p(X^{1:N} \mid G) = \prod_i p(x_i^{1:N} \mid x_i^{1:N}_{P(a(x_i))})$$

- Prior over models

$$p(G)$$

- Interventional data
Object and effects description

- Percepts of object features and effects are clusterized in an unsupervised manner.
- The categories form the space of objects.

Object features and effects are described using unsupervised learned clusters from:

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Action</td>
<td>grasp, tap, touch</td>
</tr>
<tr>
<td>H</td>
<td>Height</td>
<td>discretized in 10 values</td>
</tr>
<tr>
<td>C</td>
<td>Color</td>
<td>green1, green2, yellow, blue ball, box</td>
</tr>
<tr>
<td>Sh</td>
<td>Shape</td>
<td>small, medium, big</td>
</tr>
<tr>
<td>S</td>
<td>Size</td>
<td>small, medium, big</td>
</tr>
<tr>
<td>V</td>
<td>Object velocity</td>
<td>small, medium, big</td>
</tr>
<tr>
<td>HV</td>
<td>Hand velocity</td>
<td>small, medium, big</td>
</tr>
<tr>
<td>Di</td>
<td>Object-hand velocity</td>
<td>small, medium, big</td>
</tr>
<tr>
<td>Ct</td>
<td>Contact duration</td>
<td>none, short, long</td>
</tr>
</tbody>
</table>
Learning affordances: Baltazar

Learned object affordances

- 300 trials using random exploration, i.e. object action pair
- The structure network learnt by the MC3 algorithm (Markov Chain Monte Carlo Model Composition algorithm)

Not all network structures explain the data equally well.
Robotic Affordances

Affordances as the link between sensory-motor coordination and social interaction

Outline

• Motivation: affordances
  – Modeling
  – Learning

• Using the model
  – **Imitation games**
  – Task learning
  – Mirror and canonical neurons

• Extensions: multimodal perception (words)

• Discussion
From SMM to social interaction

- Affordances can answer the following questions:
  - Core capabilities for social interaction
  - Emulation: achieve the same effect

<table>
<thead>
<tr>
<th>inputs</th>
<th>outputs</th>
<th>function</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (O, A) )</td>
<td>( E )</td>
<td>Predict Effect</td>
</tr>
<tr>
<td>( (O, E) )</td>
<td>( A )</td>
<td>Recognize action &amp; Planning</td>
</tr>
<tr>
<td>( (A, E) )</td>
<td>( O )</td>
<td>Object recognition &amp; selection</td>
</tr>
</tbody>
</table>

Simple emulation

- One step Bayesian decision
  \[
  \langle a^*, o^* \rangle = \operatorname*{arg\,max}_{a \in A, o \in O} \mathbb{E} [ r(a^d, f^d, e^d, a, f^o, e^o) ]
  \]

- It is possible to define different simple reward functions to obtain different behaviors:
  - Matching an effect:
    \[
    r(e^d) = \begin{cases} 
    1, & \text{if } E^i = \hat{e}^d \\
    0, & \text{otherwise} 
    \end{cases}
    \]
  - Matching effect and Object:
    \[
    r(e^d, f^d, f^i) = \begin{cases} 
    1, & \text{if } E^i = \hat{e}^d \land F^i = \hat{f}^d \\
    0, & \text{otherwise} 
    \end{cases}
    \]
Imitation games

Objective: select **action and object** to obtain the same effect on **a similar object**

Demonstration (grasp on small box)

Which action gives the same effect?

The reward now also includes information about the object features

---

**Imitation games**

<table>
<thead>
<tr>
<th># 1</th>
<th># 2</th>
<th># 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>demonstration</td>
<td>action selection</td>
<td>imitation</td>
</tr>
<tr>
<td>effect imitation</td>
<td>effect imitation &amp; object selection</td>
<td>effect imitation &amp; object selection</td>
</tr>
<tr>
<td>effect imitation</td>
<td>effect &amp; object imitation</td>
<td>effects &amp; object imitation with the same weighted reward function</td>
</tr>
</tbody>
</table>
Outline

- Motivation: affordances
  - Modeling
  - Learning
- Using the model
  - Imitation games
  - **Task learning**
  - Mirror an canonical neurons
- Extensions: multimodal perception (words)
- Discussion

**Imitation (task)**

Affordance based imitation
- World model (cause-effect)
- Action recognition

Task interpretation (BIRL)
- Bayesian Inverse Reinforcement Learning
  - (sequences of actions, estimating the task reinforcement function.)

Imitation (emulation)
Experiments

Kick the balls out of the table
Touch the large ball
Drop the boxes in the pile

Inaccuarte and incomplete demonstration

**Experiment 2: Inaccurate, incomplete demonstration**
(Demonstrated and learned policies).

<table>
<thead>
<tr>
<th>State</th>
<th>Demo</th>
<th>Learned</th>
</tr>
</thead>
<tbody>
<tr>
<td>(∅, BBall)</td>
<td>-</td>
<td>TouchL</td>
</tr>
<tr>
<td>(∅, Box)</td>
<td>GraspR</td>
<td>TouchL</td>
</tr>
<tr>
<td>(∅, SBall)</td>
<td>TapR</td>
<td>TouchL</td>
</tr>
<tr>
<td>(BBall, ∅)</td>
<td>TouchL</td>
<td>TouchL</td>
</tr>
<tr>
<td>(BBall, Box)</td>
<td>TouchL</td>
<td>TouchL</td>
</tr>
<tr>
<td>(BBall, SBall)</td>
<td>TouchL</td>
<td>TouchL</td>
</tr>
<tr>
<td>(Box, ∅)</td>
<td>GraspL</td>
<td>GraspL</td>
</tr>
<tr>
<td>(Box, BBall)</td>
<td>GraspL</td>
<td>GraspL</td>
</tr>
<tr>
<td>(Box, Box)</td>
<td>GraspL</td>
<td>GraspL</td>
</tr>
<tr>
<td>(Box, SBall)</td>
<td>GraspL</td>
<td>GraspL</td>
</tr>
<tr>
<td>(SBall, ∅)</td>
<td>TapL</td>
<td>TapL</td>
</tr>
<tr>
<td>(SBall, BBall)</td>
<td>TapL</td>
<td>TapL</td>
</tr>
<tr>
<td>(SBall, Box)</td>
<td>TapL</td>
<td>TapL</td>
</tr>
<tr>
<td>(SBall, SBall)</td>
<td>TapL</td>
<td>TapL</td>
</tr>
</tbody>
</table>
Outline

- Motivation: affordances
  - Modeling
  - Learning

- Using the model
  - Imitation games
  - Task learning
    - Mirror an canonical neurons

- Extensions: multimodal perception (words)

- Discussion
Mirror Neurons
[Gallese, Fadiga, Fogassi and Rizzolati, Brain, 1996]

Active during observation of another monkey’s or experimenter’s hands interacting with objects.

Observed & executed actions are the same:

Observed & executed action are NOT the same (tool):

Canonical Neurons (affordances)

Motor neurons
Respond also to the presentation of
- food or
- graspable 3D objects,
- even in the absence of subsequent movement.

Object specific
(size and shape must be congruent with the type of grip coded by the neuron).

Responses for a selective neuron for ring shapes.

**Action observation/execution resonance**

**Individual A**

**Individual B**

**Affordances vs mirror/canonical**

<table>
<thead>
<tr>
<th>inputs</th>
<th>outputs</th>
<th>function</th>
</tr>
</thead>
<tbody>
<tr>
<td>((O, A))</td>
<td>(E)</td>
<td>Predict Effect</td>
</tr>
<tr>
<td>((O, E))</td>
<td>(A)</td>
<td>Recognize action &amp; Planning</td>
</tr>
<tr>
<td>((A, E))</td>
<td>(O)</td>
<td>Object recognition &amp; selection</td>
</tr>
</tbody>
</table>
iCub example

Outline

• Motivation: affordances
  – Modeling
  – Learning

• Using the model
  – Imitation games
  – Task learning
  – Mirror an canonical neurons

• Extensions: multimodal perception (words)

• Discussion
The setup

Adding words to the model

- \( p(X) \) represents the world behaviour, i.e. what the robot has learned through experience.
- \( W \) a set of words, description of one experiment,
- Goal: to find mapping between \( W \) and \( X \), achieved by estimating \( p(X,W) \)

\[
p(X,W) = \prod_{\omega_i \in W} p(\omega_i \mid X_{\omega_i}) p(X)
\]

(1)

- \( X_{\omega_i} \) is a subset of nodes \( X \) which are parents of word \( \omega_i \)
- strong assumption is the independence among words, a “bag of words”
- to choose from all models described by (1) we use variation of the simple greedy approach – K2 algorithm.
Results

• Comparing results obtained with ideal speech recognizer (100%) and with the real one (81%), it is visible that mistakes from the real recognizer have only a small influence on the model performance

• model distinguishes non-referential words
• model sensitive to unbalanced training data

Results

• grasp action and its association
Results – Instructing the Robot

Examples of using the Bayesian network to select actions and objects

<table>
<thead>
<tr>
<th>objects on the table</th>
<th>“small grasped”</th>
<th>“moving green”</th>
<th>“ball sliding”</th>
<th>“big rolling”</th>
<th>“has rising”</th>
<th>“sliding small”</th>
<th>“rises yellow”</th>
</tr>
</thead>
<tbody>
<tr>
<td>light green circle big yellow circle medium</td>
<td>-</td>
<td>grasp, p=0.034</td>
<td>-</td>
<td>tap, p=0.227</td>
<td>grasp, p=0.019</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>dark green box small blue box medium</td>
<td>grasp, p=0.122</td>
<td>grasp, p=0.041</td>
<td>-</td>
<td>-</td>
<td>grasp, p=0.037</td>
<td>tap, p=0.25</td>
<td>-</td>
</tr>
<tr>
<td>blue box big</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>tap, p=0.022</td>
<td>grasp, p=0.037</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>dark green circle small</td>
<td>grasp, p=0.127</td>
<td>tap, p=0.127</td>
<td>-</td>
<td>-</td>
<td>grasp, p=0.017</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- incomplete instructions
- both factors taken into consideration: available objects and verbal task assignment

Grasping (where & how)

Figure: (left) predicted grasping probability, (center) variance of parameter $p$, (right) pixels $p > 0.5$ crosses $\times$ failed grasps, plus $+$ successful ones.
Outline

- Motivation: affordances
  - Modeling
  - Learning

- Using the model
  - Imitation games
  - Task learning
  - Mirror and canonical neurons

- Extensions: multimodal perception (words)

Discussion

- Bridging sensorimotor layers and “higher cognitive behavior”
- Relationship to the “mirror” and “canonical” neurons
- Extension to other perceptual modalities (word/meaning)
- Future developments
  - Scale (structural learning with large dimensional graphs)
  - Batch versus incremental
  - Continuous versus discrete variables (parameterized is ok)
  - Application to grasping and handling
Welcome to VisLab – Computer Vision Lab of the Instituto de Sistemas e Robótica (at Instituto Superior Técnico).

The VisLab focuses on the research and development of tools based on computer vision, as well as the application of computer vision in the context of robotics. We are particularly interested in the problems of active vision, visual based control, 3D-reconstruction, motion analysis and segmentation. One characteristic of some of our research work is the multidisciplinary combination between engineering approaches and fields like biology and psychology.

Instituto de Sistemas e Robótica, Instituto Superior Técnico, Torre Norte
Avenida Rovisco Pais, 1, 1049-001 Lisboa Codex, PORTUGAL
Tel.: +351-21-8416050
Fax.: +351-21-8411071

Credits (http://vislab.isr.ist.utl.pt)

Head, face, cover design: Ricardo Beira, M. Praça, L. Vargas
Vision & attention: Alexandre Bernardino, J. Ruesch
Learning, affordances: Manuel Lopes, Luis Montesano
Language acquisition: Jonas Hornstein, Cláudia Soraes
G. Salvi, Verica Krunic – Vision-speech association
Matteo Tajana – model based tracking
Ricardo Nunes – robot life support!

http://vislab.isr.ist.utl.pt
José Santos-Victor – jasv@isr.ist.utl.pt