Visuomotor Associative Learning under the Predictive Coding Framework: a Neuro-robotics Experiment

Jungsik Hwang\textsuperscript{1,2} and Jun Tani\textsuperscript{2}

\textsuperscript{1}Korea Advanced Institute of Science and Technology, Korea
\textsuperscript{2}Cognitive Neurorobotics Research Unit, Okinawa Institute of Science and Technology, Japan
Research Objectives

Build a **Cognitive Agent** which can

- Develop cognitive functions autonomously

Embodiment

“Learning from sensorimotor experience” acquired from dynamic interaction with the world

Prediction

“Brain = a Prediction Machine”
Learning from Demonstration

Obtaining sensorimotor experience
By showing a robot how to do it.

Then, we make the robot learn from this experience.
How can a robot learn from experience?

Experience consists of

- Visual images (Vision)
- Joint position values (Proprioception)
How can a robot learn from experience?

Dynamic Neural Network

Sensorimotor Experience
Proposed Model

**Predictive** Visuo-Motor Dynamic Neural Network (P-VMDNN)

"Learning to **predict** sensorimotor signals simultaneously in an end-to-end manner"
Proposed Neural Network Model

Visual Pathway

- Predicts pixel-level dynamic visual images
  - P-MSTRNN: Predictive - Multiple Spatio-Temporal Scales RNN

Example of Visual Prediction
• Predicts robot’s action (specified as joint positions)
  – MTRNN: Multiple Timescales RNN

Example of Action Generation
Proposed Neural Network Model

**Lateral Connections** between 2 Pathways

- Bi-directional flow of visuomotor information

---

“Sensorimotor integration is a key part of the “intelligence algorithm” of the neocortex.”

- Jeff Hawkins (2017)
**Proposed Neural Network Model**

**Key Characteristics**

- **Temporal Hierarchy**
  - Imposing different **constraints** on neural activation

\[
u_i(t) = \left(1 - \frac{1}{\tau}\right)u_i(t-1) + \frac{1}{\tau}\sum_j w_{ij}x_j(t)\]

“Emergence of Functional Hierarchy”

- Larger Time Constants \(\rightarrow\) **Slowly-changing Neural Activity**
- Smaller Time Constants \(\rightarrow\) **Fast-changing Neural Activity**
Experiment Setting

- **Task:** Imitating human gestures
  - 9 gestures x 3 human subjects

- **Robot Platform**
  - iCub simulator
  - Vision) 64 x 48 grayscale
  - Action) 10 DoFs
Predictive Visuo-Motor Dynamic Neural Network

Key Features

- Processing of Spatio-Temporal Patterns
- Coupling of Vision & Proprioception
- Mental Simulation
- Prediction Error Minimization
MENTAL SIMULATION
Mental Simulation of Action

• Mental Simulation
  – Ability to imagine probable result of our actions
  – Important in social interaction
  – Needs “What to simulate”

• In Our Experiment
  – Ability to generate visuo-proprioceptive predictions with given intention
    • *Intention: specified as initial states
    • They are learnable parameters.
Mental Simulation of Action

- Mental Simulation in the Proposed Model
  1. Set the “Intention”
     • Specified as the initial states
  2. Generate Output
     • Visual & Proprioceptive predictions
  3. Feed Prediction Output into Input
     • “Closed-loop Generation”
  4. Iterate (2) – (3)
Mental Simulation of Action

- **Mental Simulation in the Proposed Model**

  1. Set the “Intention”
     - Specified as the initial states
  2. **Generate Output**
     - Visual & Proprioceptive predictions
  3. Feed Prediction Output into Input
     - “Closed-loop Generation”
  4. Iterate (2) – (3)
Mental Simulation of Action

- **Mental Simulation in the Proposed Model**
  1. Set the “Intention”
     - Specified as the initial states
  2. Generate Output
     - Visual & Proprioceptive predictions
  3. **Feed Prediction Output into Input**
     - “Closed-loop Generation”
  4. Iterate (2) – (3)

- Without external input from environment
- Only with given intention
Mental Simulation of Action

Mental Simulation in the Proposed Model

1. Set the “Intention”
   • Specified as the initial states
2. Generate Output
   • Visual & Proprioceptive predictions
3. Feed Prediction Output into Input
   • “Closed-loop Generation”
4. Iterate (2) – (3)

• Without external input from environment
• Only with given intention
Result

**Mental Simulation of Action**

- Setting **intention states**
  - at the onset of mental simulation
  - Obtained from training
With given ‘intention’, the model generated coherent visuo-proprioceptive patterns
- Imagination without any input from the external world
Hierarchical representation of visuo-proprioceptive patterns
- **Abstract** information at higher-level: Type of gesture
- **Specific** information at lower-level: Shape of specific human subject

Self-organized **Functional Hierarchy**

Low-level Representation (shape of a specific subject)  
**Initial States obtained from Training**  
High-level Representation (type of the gesture)
PREDICTION ERROR MINIMIZATION
**Prediction Error Minimization**

- Core of “Predictive Coding”
  - Recognizing intention from observation by minimizing prediction error
  - Account for MNS (Mirror Neuron Systems)
    - *Mirror Neurons: Activated while executing & observing an action*

[Image of Predictive Coding Framework]

Predictive Coding Framework
Stefanics, et. al., (2014)
At each time step

1. Generate Visuo-Prop. Predictions
   • From Intention State (Top-Down Process)

2. Compute Prediction Error
   • Difference b/w Predicted & Observed Patterns

3. Backpropagate Prediction Error & Update Intention State
   • Bottom-Up Process

4. Iterates a Few Times

Prediction Error Minimization
At each time step

1. **Generate Visuo-Prop. Predictions**
   - From Intention State (Top-Down Process)

2. **Compute Prediction Error**
   - Difference b/w Predicted & Observed Patterns

3. **Backpropagate Prediction Error & Update Intention State**
   - Bottom-Up Process

4. **Iterates a Few Times**
Prediction Error Minimization

At each time step

1. Generate Visuo-Prop. Predictions
   • From Intention State (Top-Down Process)

2. **Compute Prediction Error**
   • Difference b/w Predicted & Observed Patterns

3. Backpropagate Prediction Error & Update Intention State
   • Bottom-Up Process

4. Iterates a Few Times
At each time step

1. Generate Visuo-Prop. Predictions
   - From Intention State (Top-Down Process)

2. Compute Prediction Error
   - Difference b/w Predicted & Observed Patterns

3. **Backpropagate Prediction Error** & Update Intention State
   - Bottom-Up Process

4. Iterates a Few Times
Prediction Error Minimization

At each time step

1. Generate Visuo-Prop. Predictions
   • From Intention State (Top-Down Process)
2. Compute Prediction Error
   • Difference b/w Predicted & Observed Patterns
3. Backpropagate Prediction Error & Update Intention State
   • Bottom-Up Process
4. Iterates a Few Times
At each time step

1. Generate Visuo-Prop. Predictions
   • From Intention State (Top-Down Process)
2. Compute Prediction Error
   • Difference b/w Predicted & Observed Patterns
3. Backpropagate Prediction Error & Update Intention State
   • Bottom-Up Process
4. Iterates a Few Times

“Perception as an Active Process”

⇔ Solely determined by input
Prediction Error Minimization

Minimizing Visual PE

Minimizing Prop. PE
Prediction Error Minimization

Minimizing Visual PE

- Minimizing the difference b/w
  - Visual Prediction (predicted gesture)
  - Observation (observed gesture)

Prediction Error = Difference (Observed Gesture, Predicted Gesture)
Prediction Error Minimization

Minimizing Visual PE

- Minimizing the difference b/w
  - Visual Prediction (predicted gesture)
  - Observation (observed gesture)
- No External Proprioceptive Signal
  - Robot’s action was generated simultaneously while minimizing Visual PE
Prediction Error Minimization

Minimizing Visual PE

• With Visual PE Minimization
  – Predicted coherent visual & Proprioceptive patterns \(\Rightarrow\) Successful imitation

• Without Visual PE Minimization
  – Did NOT predict Visual & Proprioceptive patterns \(\Rightarrow\) Unsuccessful imitation

Prediction Error
Prediction Error Minimization

Minimizing Prop. PE

• Minimizing the difference b/w
  – Prop. Prediction (Predicted joint position)
  – Observation (Perceived joint position)

\[
\text{Prediction Error} = \text{Minimizing Prop. PE} \\
\text{Difference (Observed Jnt Position, Predicted Jnt Position)}
\]
Prediction Error Minimization

Minimizing Prop. PE

• Minimizing the difference b/w
  – Prop. Prediction (Predicted joint position)
  – Observation (Perceived joint position)

• No External Visual Target Signal
  – Visual Prediction was generated simultaneously while minimizing Prop. PE
Prediction Error Minimization

Minimizing Prop. PE

• With Proprioceptive PE Minimization
  – Successfully minimized Proprioceptive PE
  – Generated corresponding Visual Prediction (imaginary)

• Without Proprioceptive PE Minimization
  – Not able to adapt to incoming Proprioceptive signal
Neural Activation while Minimizing Visual Prediction Error
Neural Activation while Minimizing Visual Prediction Error

Overlapping Trajectories b/w Training & Testing

- Inferring intention latent in observed patterns @ Higher-level
- Recalling the corresponding representations @ Lower-level

⇒ Retrieval of missing sensorimotor signals
Neural Activation while Minimizing Visual Prediction Error

- **MNS-like Behavior** emerged from
  1. Neural connectivity (between two pathways)
  2. Learning sensorimotor experience

**3. Prediction Error Minimization**

- “Predictive Coding Account of MNS” (Kilner, Friston and Frith, 2007)
- “Within predictive coding, recognition of causes is simply the process of jointly minimizing prediction error at all levels of a cortical hierarchy.”
Conclusion

Build a **Cognitive Agent** based on

- **Embodiment**
  
  “Learning from sensorimotor experience”
  
  acquired from dynamic interaction with the world

- **Prediction**
  
  “Brain = a Prediction Machine”

- Complex cognitive behaviors emerged
  - Mental simulation, Intention recognition, MNS-like behavior, etc.
  - From “**Visuo-Motor associative learning under the predictive coding framework**”
Thank you

Please see the following paper for more information.