Lecture #10 – Modeling the Physical World

CMP722
ADVANCED COMPUTER VISION

Aykut Erdem // Hacettepe University // Spring 2019

Video: The Jenga-playing robot (MIT)
Previously on CMP722

• graph structured data
• graph neural nets (GNNs)
• GNNs for “classical network problems”
Lecture overview

• physical scene understanding
• intuitive physics
• interaction networks
• relation networks
• visual interaction networks
• learning physics engines via graph networks

• Disclaimer: Much of the material and slides for this lecture were borrowed from
  —Peter Battaglia’s slides on “Structure in physical intelligence”
How do you understand a scene?
How do you understand a scene?

1. Parse it into physical objects and relations
2. Reason about the objects and their interactions

"Precarious"

Attached?    Fall?

Support
“Infinite use of finite means”
- von Humboldt, on the productivity of language

"Precarious"
Kenneth Craik, “The Nature of Explanation”, 1943:

"If the organism carries a 'small-scale model' of external reality and of its own possible actions within its head, it is able to try out various alternatives, conclude which is the best of them, react to future situations before they arise, utilize the knowledge of past events in dealing with the present and future, and in every way to react in a much fuller, safer, and more competent manner to the emergencies which face it." (pg 61)

"This concept of 'thinghood' is of fundamental importance for any theory of thought." (pg 77)
Claim: Human intelligence is structured

Founded on objects, relations, reasoning

- Objects and relations reflect decisions made by evolution, experience, and task demands about how to represent the world in an efficient and useful way
- Structure in our core cognitive knowledge evident very early in infancy (Spelke)
- Model-building over recognizing patterns (Tenenbaum)
- Combinatorial generalization via compositionality ("infinite use of finite means")
What is the mechanism of human intuitive physics?

Intuitive Physics Engine: the "physics engine in the head"

Battaglia, Hamrick, Tenenbaum, 2013, PNAS
Experiments: What will happen? Why?

Will it fall?

In which direction?

Different masses

Comiples scenes

Infer the mass

Predict fluids

Battaglia et al., 2013

Hamrick et al., 2016

Bates et al., 2015, 2018
Message from cognition

Humans use richly structured representations of objects and relations to reason about, and interact with, their everyday environment.

What insights does humans’ structured intelligence offer AI?
We need better object- and relation-centric models in AI

A graph is a natural way to represent entities and their relations:
• “Nodes” correspond to entities, objects, events, etc.
• “Edges” correspond to their relations, interactions, transitions, etc.
• Inferences about entities and relations respect the graphical structure.

Graphs can capture data from many complex systems:

• Physical systems
• Scene graphs
• Social networks
• Linguistic structure
• Programs

• Search trees
• Communication networks
• Transportation networks
• Chemical structure
• Phylogenetic trees
Intuitive physics as reasoning about graphs
Intuitive physics as reasoning about graphs
Interaction Network

Strong relational inductive bias: Deep learning architecture which operates on graphs

Related to the broad family of "Graph Neural Networks" (Scarselli et al, 2009; Li et al, 2015) and "Message-Passing Neural Networks" (Gilmer et al., 2017). Chang et al. (2016) also proposed a similar version in parallel.
Interaction Network

Relational reasoning
Compute interaction

Object reasoning
Apply object dynamics

Relation model
Shared MLP ($f$) applied per-relation

Object model
Shared MLP ($g$) applied per-object

Relations ($E$)

Objects ($V$)

Per-object external signals ($U$)

Per-object predictions

$v_1 \quad e_1 \quad v_3$
$v_2 \quad e_2$
$u_1 \quad u_2 \quad u_3$

$f(v_1, v_2, e_1) \rightarrow \tilde{e}_1$
$f(v_3, v_2, e_2) \rightarrow \tilde{e}_2$
$0 \rightarrow \tilde{v}_1$
$\tilde{e}_1 + \tilde{e}_2 \rightarrow \tilde{v}_2$
$0 \rightarrow \tilde{v}_3$

$g(v_1, \tilde{v}_1, u_1) \rightarrow v'_1$
$g(v_2, \tilde{v}_2, u_2) \rightarrow v'_2$
$g(v_3, \tilde{v}_3, u_3) \rightarrow v'_3$

Battaglia et al., 2016, NeurIPS
Interaction Network

Can learn a general-purpose physics engine, simulating future states from initial ones

- **n-body**
  - Gravitational forces

- **Balls**
  - Rigid collisions between walls and balls

- **String**
  - Springs and rigid collisions

Battaglia et al., 2016, NeurIPS
1000-step rollouts from 1-step supervised training

n-body

Ground truth

Model

Balls

Strings

Battaglia et al., 2016, NeurIPS
Zero-shot generalization to larger systems

Ground truth

Model

Battaglia et al., 2016, NeurIPS
Interaction Network for system-level predictions

A "global model" can be added, which aggregates the per-object outputs to make predictions.

Can be trained to predict potential energy of a system, outperforming MLP baselines.

Battaglia et al., 2016, NeurIPS
Relation Network

Remove “object model” and predict global outputs only using “relation model”’s output.

Raposo et al., 2017, ICLR workshop; Santoro et al., 2017, NeurIPS
Relation Networks can infer relations in dot motion

Trained on mass-spring systems

Generalizes to point-light walkers

Santoro et al., 2017, NeurIPS
"Visual interaction network"

An interaction network augmented with a learnable perception system
"Visual interaction network"

Multi-frame encoder (conv net-based)

Interaction network

Watters et al., 2017, NeurIPS
"Visual interaction network"

Can even predict invisible objects, inferred from how they affect visible ones

Watters et al., 2017, NeurIPS
Learning to simulate more complex robotic systems

Alvaro Sanchez-Gonzalez, Nicolas Heess, Tobi Springenberg, Josh Merel, Martin Riedmiller, Raia Hadsell, Peter Battaglia
ICML, 2018
 Systems: "DeepMind Control Suite" (Mujoco) & real JACO

DeepMind Control Suite (Tassa et al., 2018)
Systems: "DeepMind Control Suite" (Mujoco) & real JACO
Kinematic tree of the actuated system as a graph

Representing physical system as a graph:

- Bodies → Nodes
- Joints → Edges
- Global properties

Similar representation to:

- Interaction Networks (Battaglia et al. 2016)
- NerveNet (Wang et al. 2018) (graph-structured policy, rather than model)
Graph Network (GN)  

Battaglia et al., 2018

Graph-to-graph, modular block design

---

Edge update  
Node update  
Global update

---

$g \rightarrow f_g \rightarrow g^*$

$\{n_i\} \rightarrow f_n \rightarrow \{n_i^*\}$

$\{e_j\} \rightarrow f_e \rightarrow \{e_j^*\}$

---

$G \rightarrow GN_1 \rightarrow G' \rightarrow GN_2 \rightarrow G^*$
Forward model: supervised, 1-step training w/ random control inputs

Sanchez-Gonzalez et al., 2018, ICML
Results: Graph Net (GN) vs MLP forward models

More repeated structure:
Better performance over MLP

Better test generalization,
within and outside of the training distribution

Sanchez-Gonzalez et al., 2018, ICML
GN forward model: Multiple systems & zero-shot generalization

**Single model** trained:
- Pendulum, Cartpole, Acrobot, Swimmer6 & Cheetah

**Zero-shot generalization**: Swimmer
- # training links: {3, 4, 5, 6, -8, 9, -, -, ...}
- # testing links: {-, -, -, 7, -, -, 10-14}
GN forward model: Real JACO data

Recurrent graph network

(Real JACO trajectories, rendered using Mujoco)

Prediction Fixed Real JACO
System identification: GN-based inference, under diagnostic control inputs

Unobserved system parameters (e.g. mass, length) are implicitly inferred

Prediction System ID Cartpole
ID phase

Real time

Slowed down 1/5

Sanchez-Gonzalez et al., 2018, ICML
Using learned models for control
Control: Model-based planning

Trajectory optimization: the GN-based forward model is differentiable, so we can backpropagate through it, and find a sequence of actions that maximize reward.

Control Fixed JACO
Imitate full pose (1x)

Target pose
Control trajectory

Sanchez-Gonzalez et al., 2018, ICML
Control: Multiple systems via a single model

Pendulum Balance (3x)

Acrobot Swing up (5x)

Cartpole Balance (3x)

Swimmer6 Move towards target (7x)

Cheetah Move forward (5x)

Sanchez-Gonzalez et al., 2018, ICML
Control: Zero-shot control

Control Fixed Move towards target (5x)

Swimmer3

Swimmer4

Swimmer5

Swimmer6

Swimmer7

Swimmer8

Swimmer9

Swimmer10

Swimmer11

Swimmer12

Swimmer13

Swimmer14

Sanchez-Gonzalez et al., 2018, ICML
Control: Multiple reward functions

Control Fixed Cheetah (k rewards)
Maximize target (3x)

Control Fixed Walker2d (k rewards)
Maximize target (1x)

Horizontal speed
Vertical position
Squared vertical speed
Squared angular speed
Inverse verticality
Feet to head height
Learning to use mental simulation
Learning to use mental simulation
"Imagination-based metacontroller"

"Spaceship task":
• Navigate to your home planet by choosing a force vector
• Challenging because the planets exert gravity

The agent learns 3 components:
1. Action policy (via stochastic value gradients (Heess et al. 2015))
2. GN-based forward model (via supervised 1-step training)
3. Internal strategy for using imagination to test potential actions before selecting one to execute (via REINFORCE)

Hamrick et al., 2017, ICLR
Learning to use mental simulation "Imagination-based planner"

- Red: real actions
- Blue: 1 step of imagination
- Green: 2+ steps of imagination

Pascanu et al., 2017, arXiv
Graph-structured model-free policies
Graph-structured model-free policies for physical construction in humans and AI

The "glue task"

Goal: Glue blocks together to make the tower stable, using the minimum amount of glue.

Jess Hamrick, Kelsey Allen, Victor Bapst, Tina Zhu, Kevin McKee, Josh Tenenbaum, Peter Battaglia
Proc Cog Sci, 2018

Hamrick et al., 2018, Proc Cog Sci
Structured model-free policies for physical construction in humans and AI

Hamrick et al., 2018, Proc Cog Sci
Structured model-free policies for physical construction in humans and AI
Graph-structured representations for model-free RL
Relational deep reinforcement learning

Box-World:
- Acquire gem (white) by opening a sequence of locked boxes
- Model-free (A2C) with self-attention / GN state representation, and message-passing
Conclusions

Human use richly structured generative knowledge

• Combinatorial generalization: “Infinite use of finite means”
• Object- and relation-centric representations
• Structured mental simulation

Graph Networks: strong relational inductive bias

• Naturally support combinatorial generalization via compositional sharing
• Graph-structured representations and policies
• Open-source library: github.com/deepmind/graph_nets (with demos, including physics!)
Reject false choices

• Nature and Nurture
• Structure and Flexibility
• Symbolic and Connectionist
• Hand-engineered and End-to-end

• The “bias versus variance trade-off” is real—however the emphasis shouldn’t be on “versus”, but rather on “trade-off”.

• Biology doesn’t choose between nature versus nurture. It uses nature and nurture jointly, to build wholes which are greater than the sums of their parts.

• There’s great promise in synthesizing new techniques by drawing on the full AI toolkit and marrying the best approaches from today with those which were essential during times when data and computation were at a premium.