Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Master Thesis Presentation

“Decentralized Information Gathering with Continuous and very large Discrete Observation Spaces”

Caus Danu

July 8th 2020

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Planning

Cambridge Dictionary:

1. “The act of **deciding** how to **do** something”

2. “The process of planning activities or events in an organized way so that they are **successful** or **happen on time**”

The game of Go. Courtesy of [2]

Chess-playing computer. Courtesy of [1]

Notice the game boards. No sophisticated perception ability is required from a computer other than understanding a uniform grid of cells. Hence, the computer cares almost exclusively about optimal action sequences given the time constraints.

Perception

Courtesy of [3]

Cambridge Dictionary:

1. “An awareness of things through the physical senses, **especially sight**”
2. “A belief or opinion, **often held by many people** and based on how things **seem**”

Goal:

Find a ***Policy***, i.e. a **mapping** between the perceived **observations** and appropriate **actions** to take, such that the task outcome is optimized with respect to time and utility.

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Joint Perception and Planning

An example of "Visual Search". Courtesy of [4]

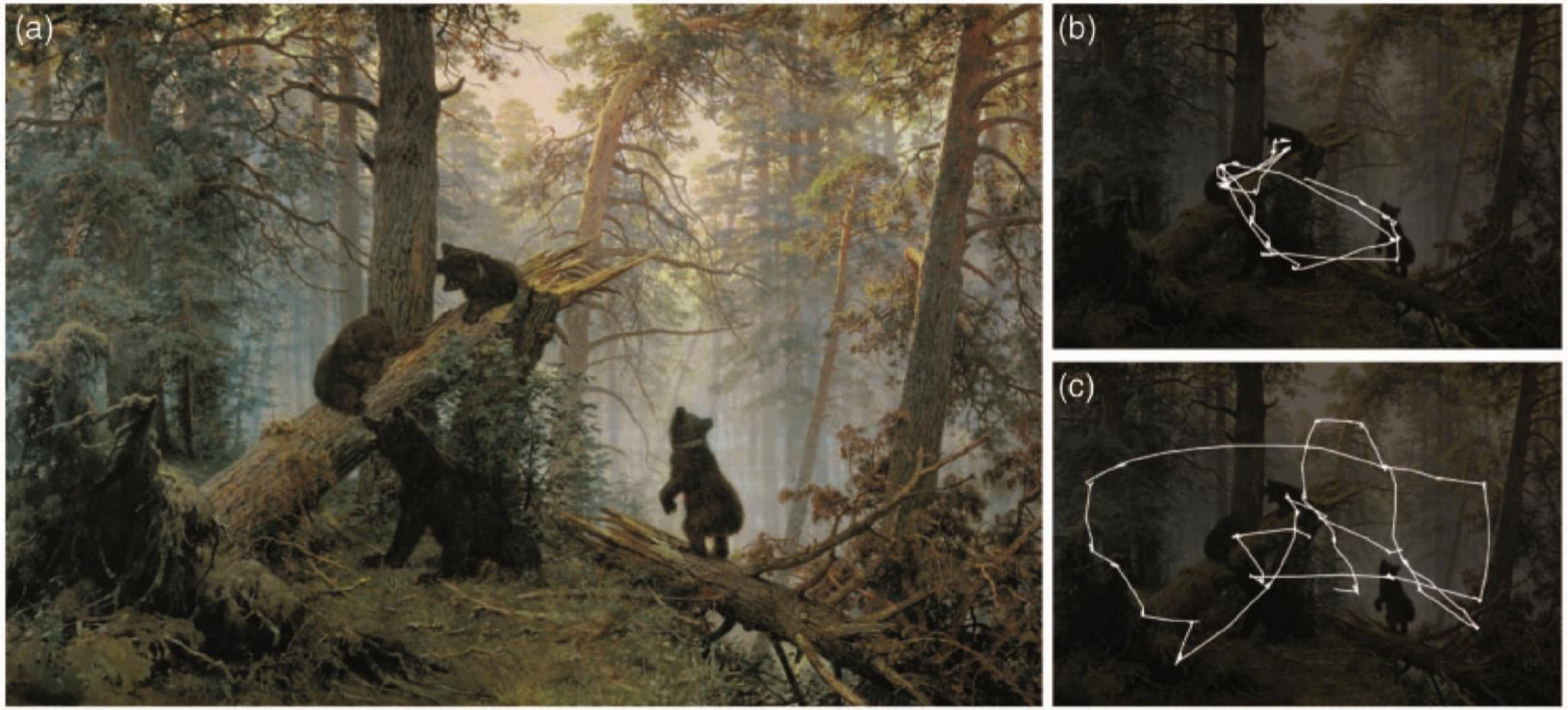


Figure 1. Eye movements across *Morning in the Pine Forest* (1889) by the Russian artists Ivan Shishkin (who drew the scenery) and Konstantin Savitsky (who drew the bears). (a) Original stimulus. (b) An observer's scan path across the image while finding the bear with the lightest fur. (c) The same observer's eye movements during free exploration of the scene.

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Decentralized POMDP

At time t the environment has a state s_t

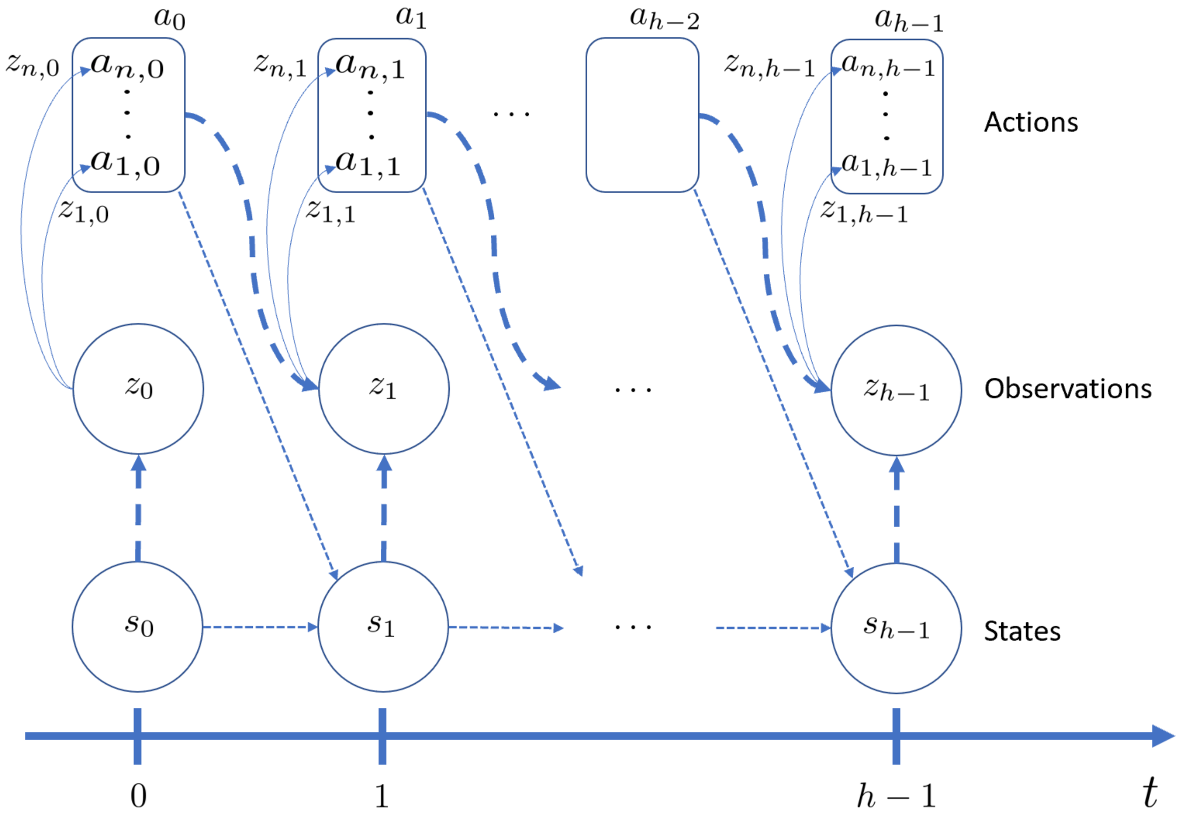
Each agent takes a local action, which is a component of a joint actions vector a_t

The environment transitions to the next state s_{t+1} according to the transition model $P(s_{t+1}|s_t, a_t)$

The environment then emits a joint observation according to the observation model $P(z_{t+1}|a_t, s_{t+1})$

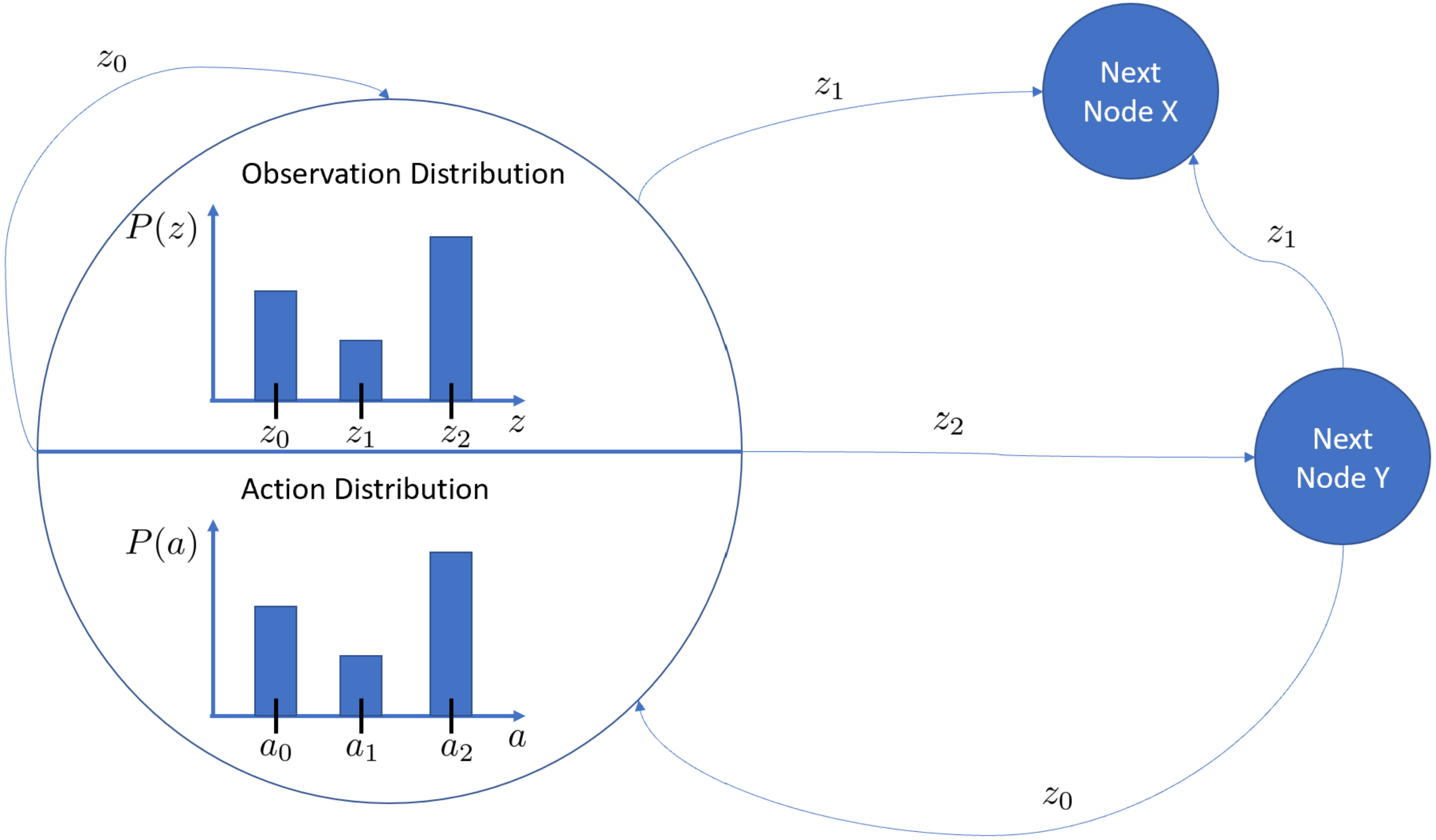
Each agent has access to its own local observation

All agents receive a common reward $R(a_t, s_t)$



Dec-POMDP Framework. Inspired by and adapted from [5]

G-DICE



Graph data structure with each node containing action and observation distributions

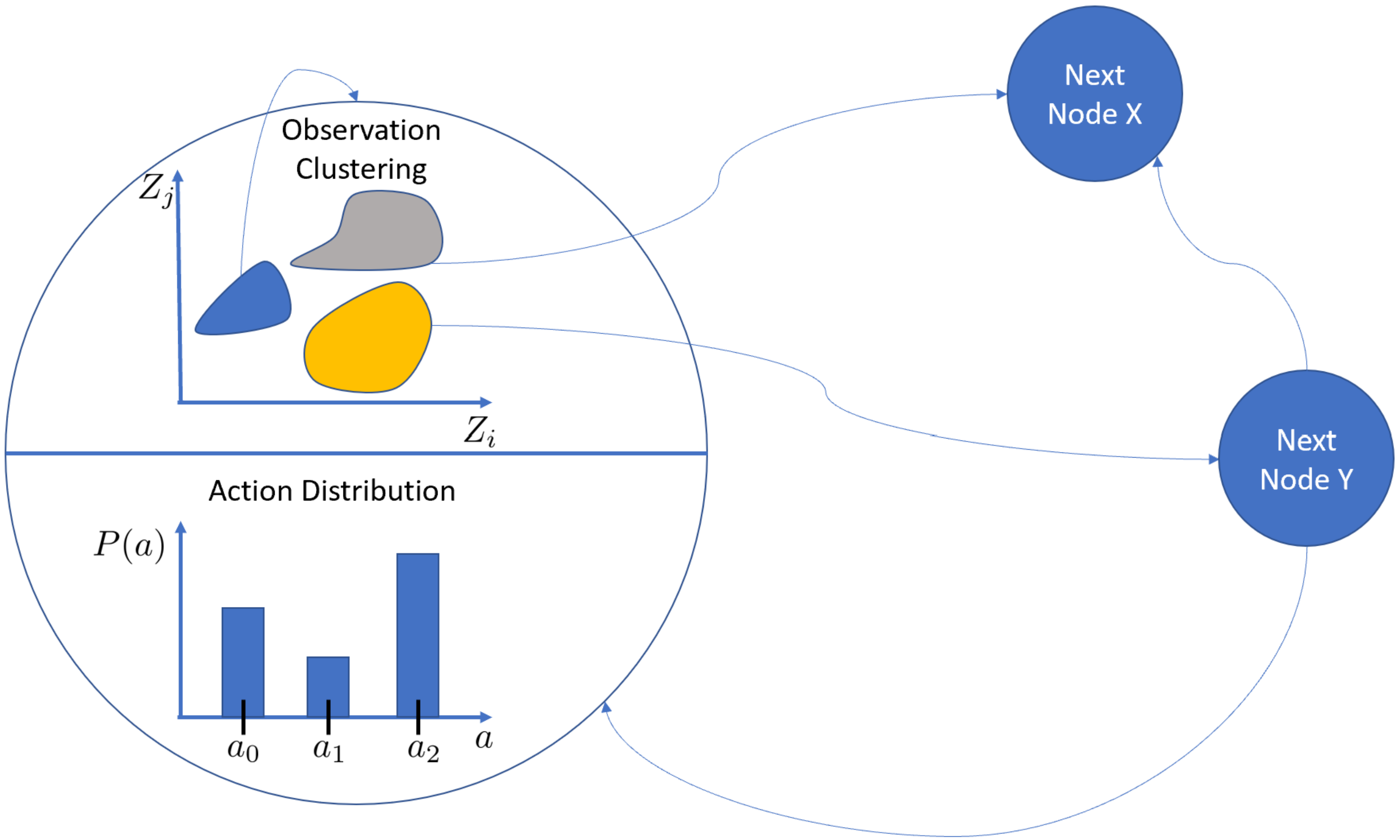
Peculiarities of G-DICE:

- **Observations are discrete**
- Actions are discrete

G-DICE Gist:

- Sample an action in a current node and carry it out
- Receive an observation at the next time step
- Based on the current node and the observation, transition to the next node

Continuous Observation G-DICE



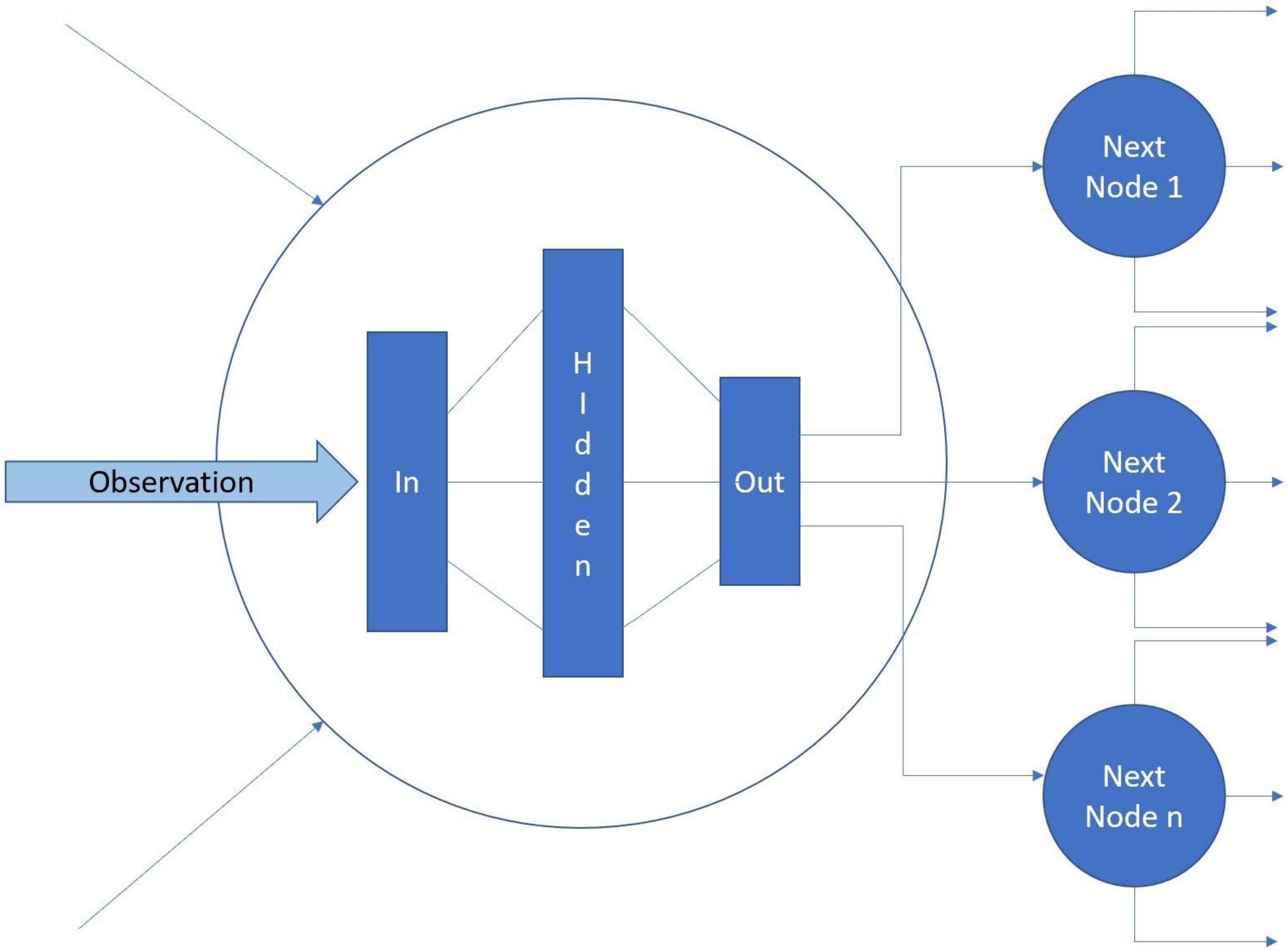
Recent methods from other works:

- Radial Basis functions
- Beta functions
- Dirichlet functions

Interesting to research:
Deep-learning within graph nodes

Cluster continuous observations using machine learning techniques

New proposal: COGNet-DICE



Leverage deep learning and embed neural networks in graph nodes

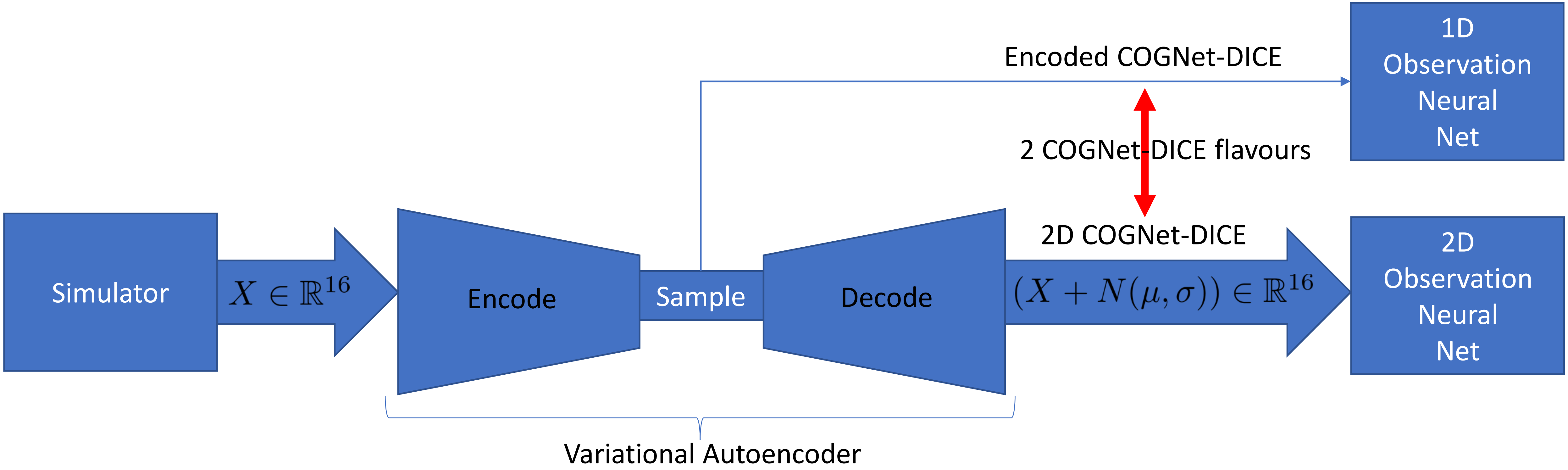
Methodology details:

- Thresholding
- Avoidance of policy degeneration through “single gene” mutation
- Entropy Injection for exploration purposes

Not used but very desirable:

- Experience replay

Scaling to higher dimensions



Reasons and Expectations:

1. Lower training space and time complexity => Efficiency
2. Robustness to noise in the online test application phase => Transfer Learning

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Research Questions

1. Can a graph-based direct cross-entropy algorithm be combined with neural networks and successfully trained to directly couple the tasks of perception and planning ?

2. If yes, how does such an algorithm perform in the case of continuous observations and states ? In particular: is such an algorithm useful for source localization problems with continuous observations and states ?

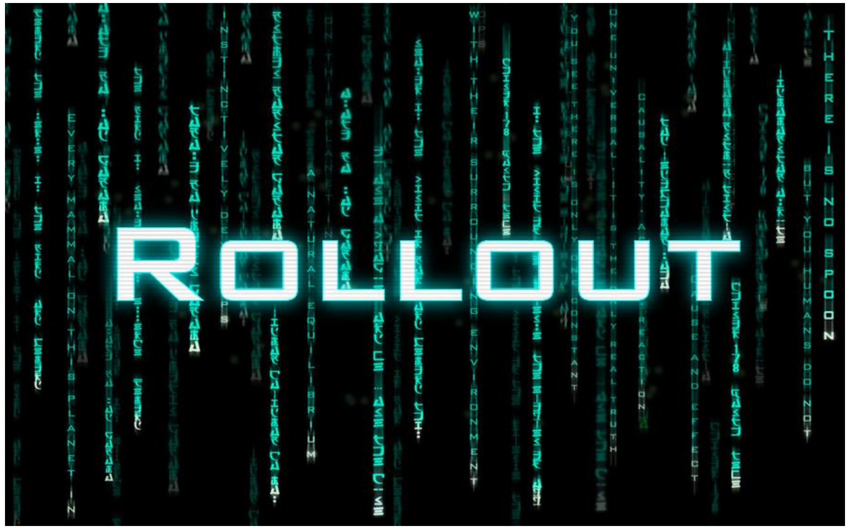
3. And lastly, does this algorithm scale to very high-dimensional discrete inputs, such as images, when we use a lossy compression scheme derived with the help of a Variational Autoencoder ?

Rollouts

1. Run many simulations to collect data:

   - Current Node
   - Sampled Action
   - Observation
   - Next Node

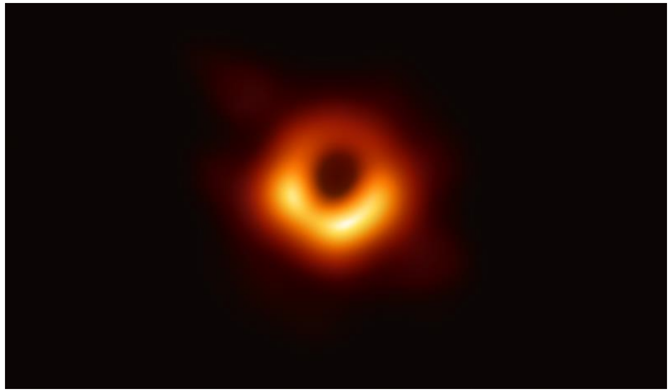
2. Apply “cross-entropy method” to make sense out of the collected data



Rollout <=> Simulation

Cross-Entropy

Entropy measures uncertainty



Courtesy of [6]

Cross-Entropy Learning Method:

1. Sample N_{total} policy samples using simulation rollouts

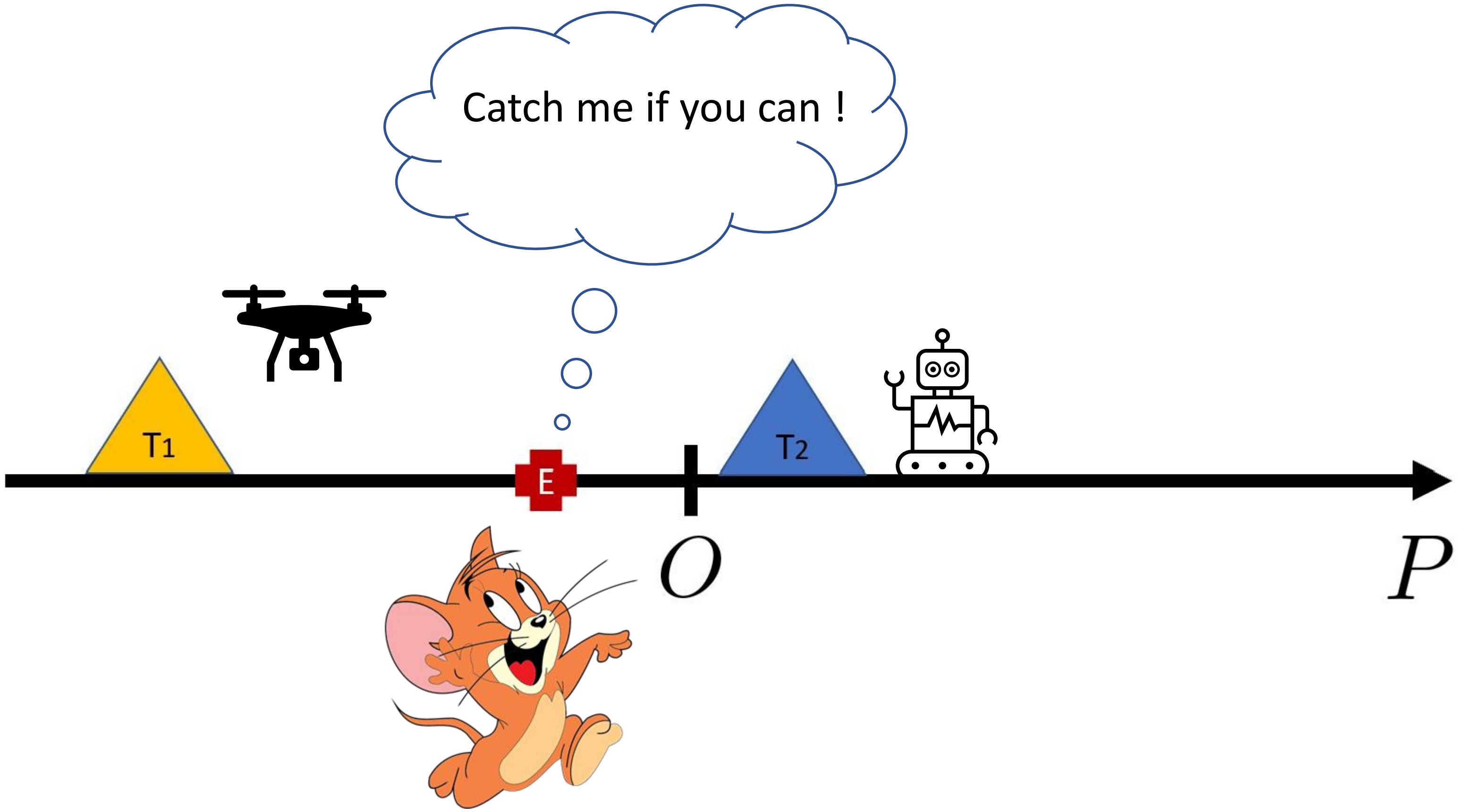
2. Estimate model using only a subset of the best samples N_{best} that result in high reward

Intuition:

By filtering the best samples, we put emphasis only on the **useful information** that we want to extract from the world

Tagging Simulation

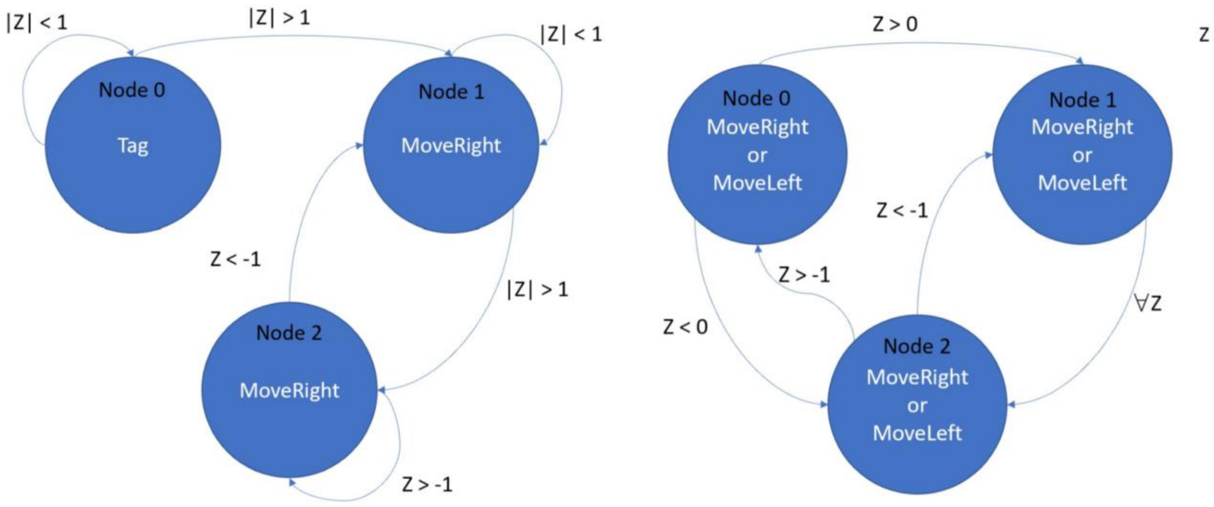
Intention: Emulate a **dynamic** source localization problem in a 1D world



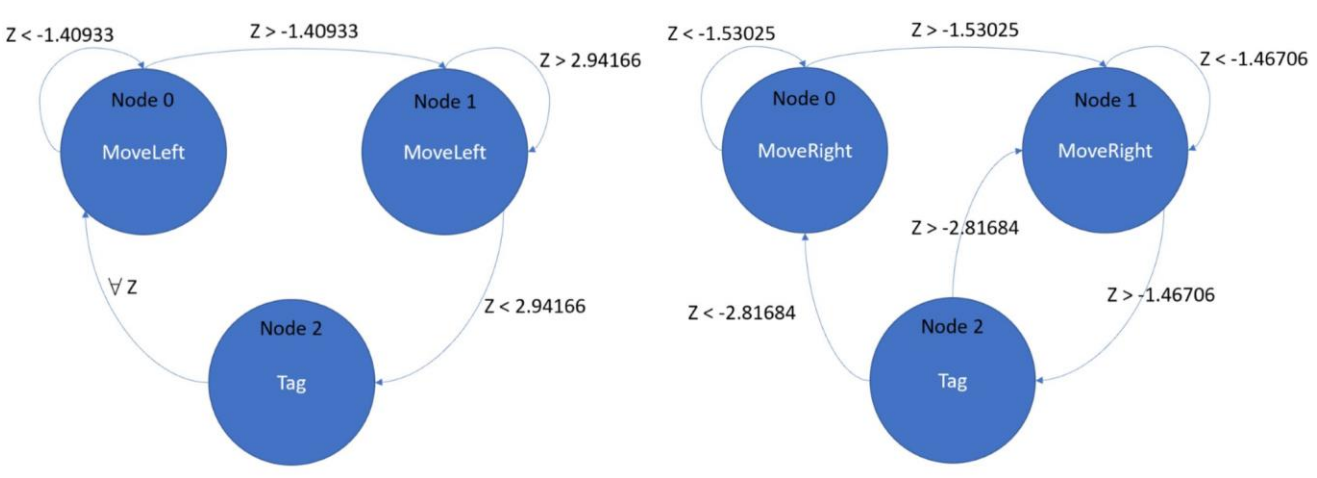
UHH Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

COGNet-DICE vs COG-DICE

N_{nodes}	COGNet-DICE policy value	COG-DICE policy value
3	1.594	1.540
6	11.103	1.794



(a) Agent 0 (b) Agent 1

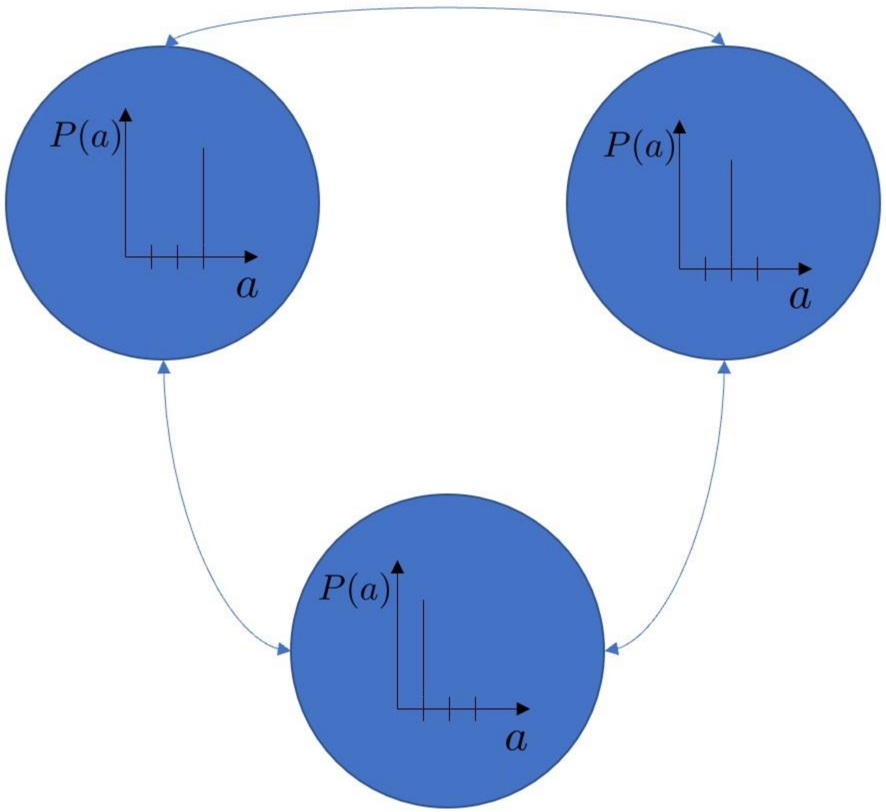


(a) Agent 0 (b) Agent 1

COGNet-DICE COG-DICE

Separate training of perception only

Method



Results

N_{nodes}	Policy value
3	consistently greater than 2
6	consistently greater than 8

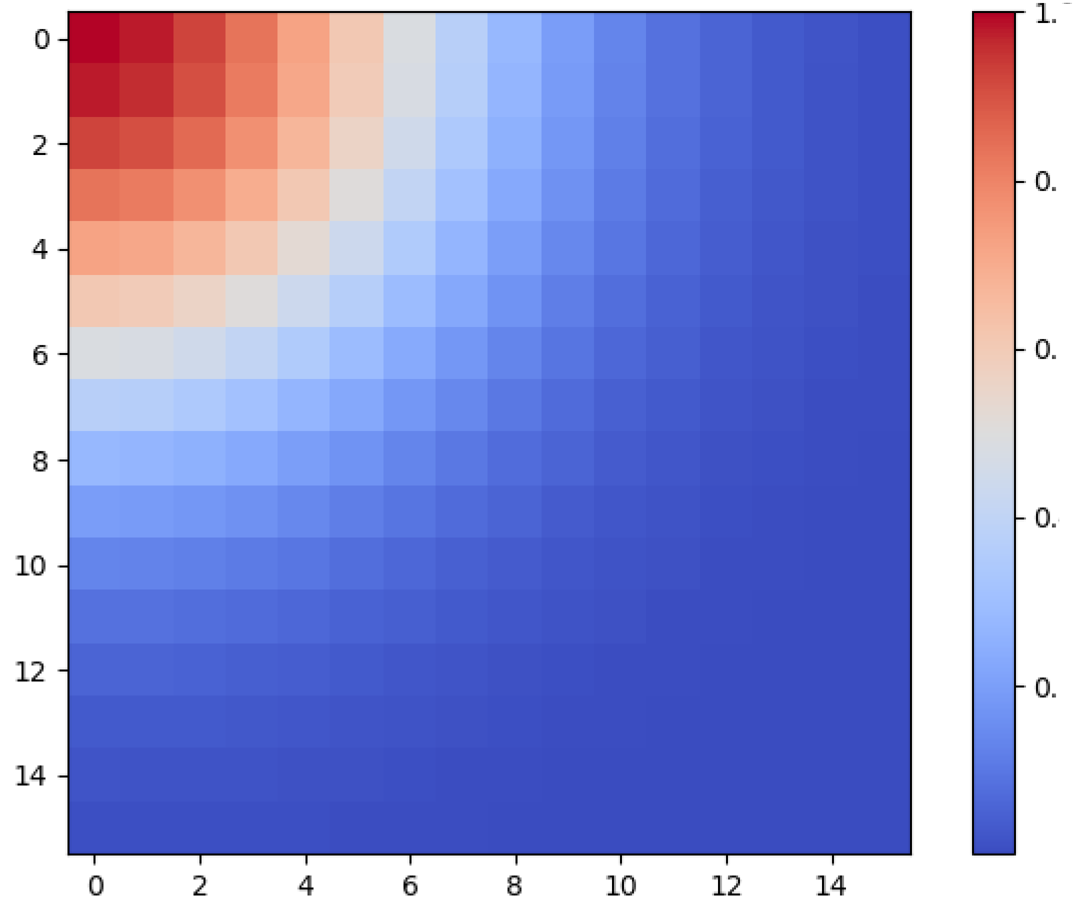
- Good results, but involves injection of human knowledge
- **Defeats the original intention of having a generic system, learning on its own based on mutual feedback from both sub-systems: the planning block and the perception block**

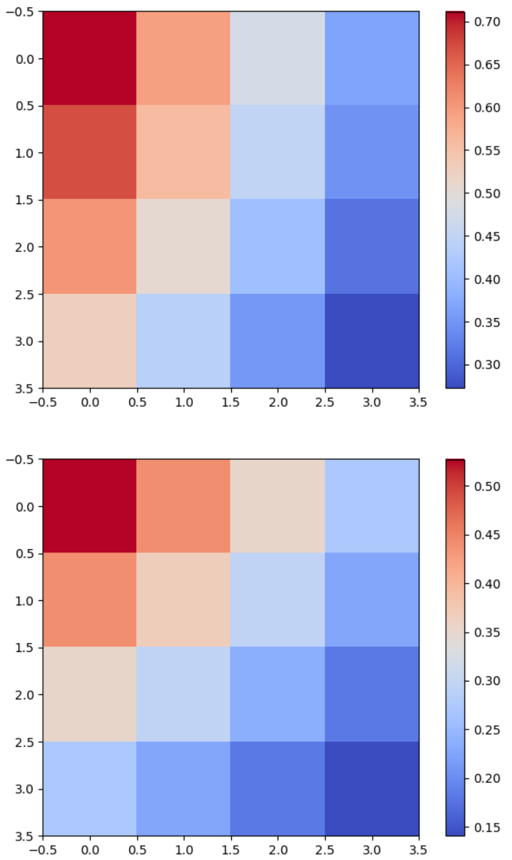
UHH Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Heatmap Simulation

Intention: Emulate a **static** source localization problem in a 2D world

World Scene

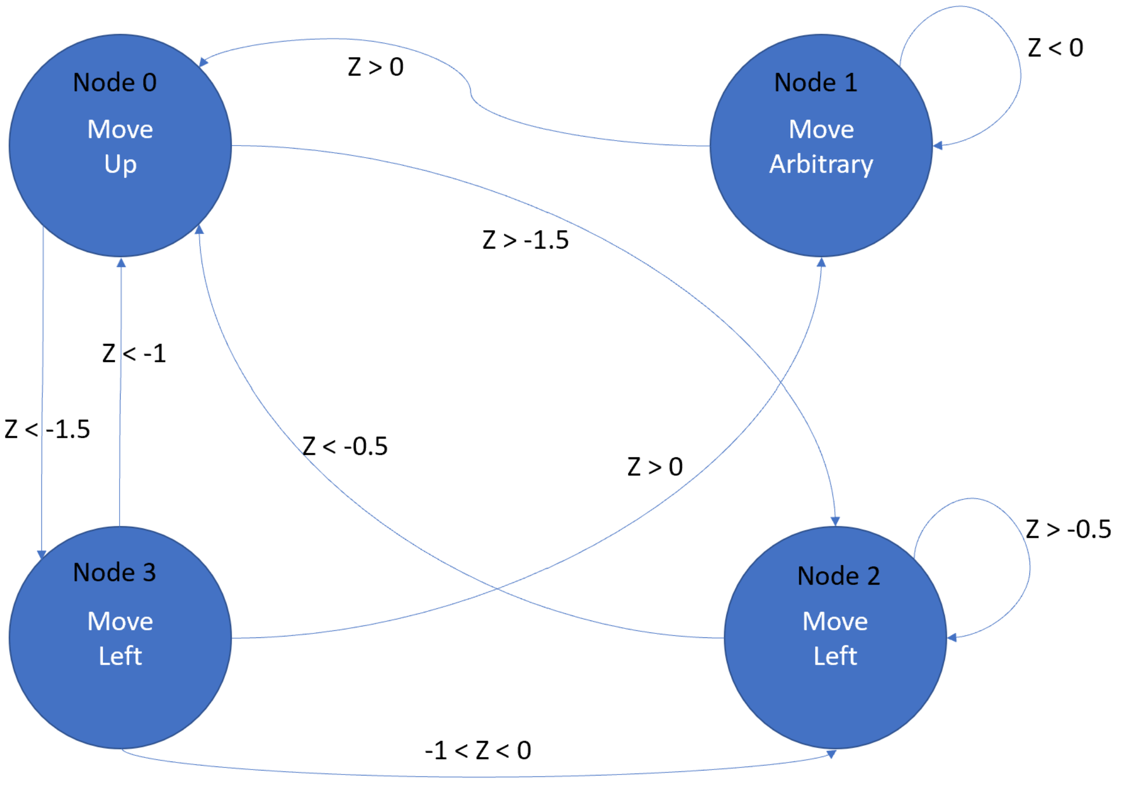




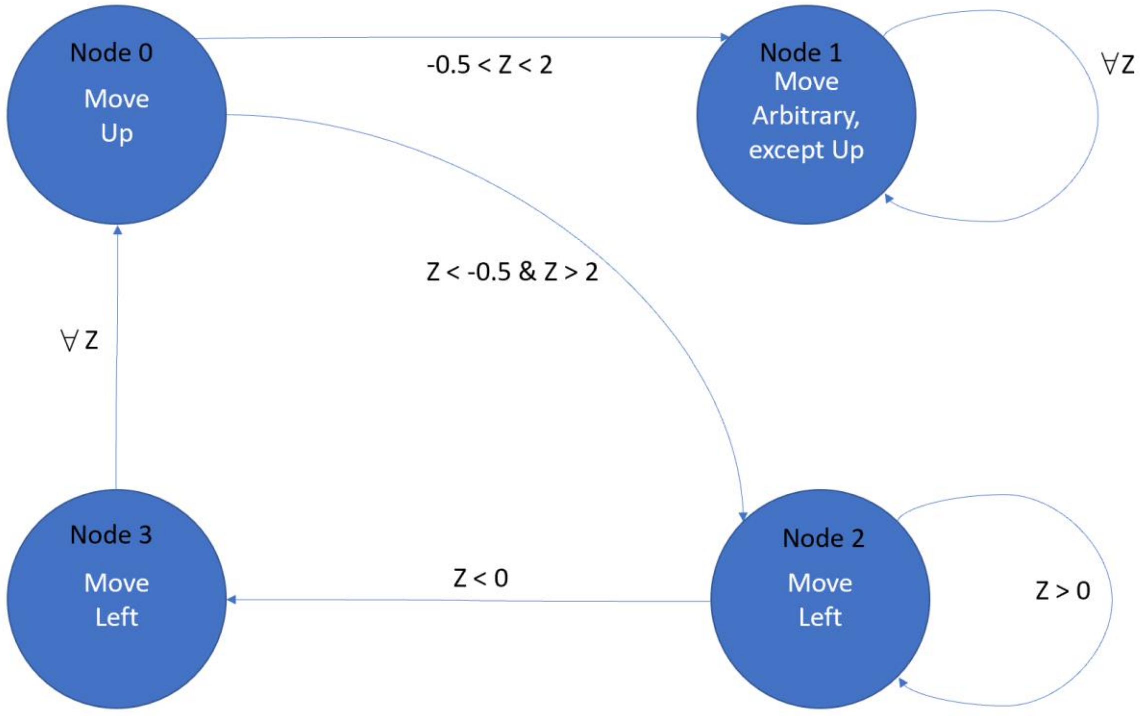
Agent Perception

2D vs Encoded COGNet-DICE

N_{nodes}	2D COGNet-DICE policy value	Encoded COGNet-DICE policy value
4	44.328	43.829
5	45.001	44.346



Encoded COGNet-DICE (Agent 0)



Encoded COGNet-DICE (Agent 1)

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Conclusions

1) Q: Can a graph-based direct cross-entropy algorithm be combined with neural networks and successfully trained to directly couple the tasks of perception and planning ?

A: Yes it can. Separate training also exhibited good results, but defeats the purpose of joint, end-to-end training.

2) Q: If yes, how does such an algorithm perform in the case of continuous observations and states ? In particular: is such an algorithm useful for source localization problems with continuous observations and states ?

A: COGNet-DICE reaches state of the art results with more training time and more collected data points than other methods such as COG-DICE. Also, by using more nodes in a graph, it is easier to reach state of the art results in less time.

3) Q: Does this algorithm scale to very high-dimensional discrete inputs, such as images, when we use a lossy compression scheme derived with the help of a Variational Autoencoder ?

A: Yes, the algorithm does scale to higher dimensions using a Variational Autoencoder, and moreover: the Variational Autoencoder helps coping with noisy observations as well.

Questions

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

References

[1] The game of Chess: https://en.wikipedia.org/wiki/Chess accessed on the 15th of June 2020

[2] The game of Go: https://en.wikipedia.org/wiki/Go_(game) accessed on the 15th of June 2020

[3] Perception from perspective: https://www.braingymmer.com/en/blog/perception/ accessed on the 15th of June 2020

[4] Rolfs, M., 2015. Attention in active vision: A perspective on perceptual continuity across saccades. *Perception*, *44*(8-9), pp.900-919.

[5] Oliehoek, F.A. and Amato, C., 2016. *A concise introduction to decentralized POMDPs* (Vol. 1). Springer International Publishing.

[6] First real photo of a black hole taken in 2019: https://www.nasa.gov/mission_pages/chandra/news/black-hole-image-makes-history accessed on the 16th of June 2020

- https://textpro.me/matrix-style-text-effect-online-884.html was accessed on the 16th of June 2020 to generate matrix style text effects on slide 11
- https://codepen.io/Mamboleoo/pen/obWGYr was accessed on the 15th of June 2020 to generate particle style text effects on slide 12
- Cartoon character Jerry on slide 13 was taken from https://seeklogo.com/vector-logo/75382/jerry accessed on the 15th of June 2020