

COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

ROBOTIC MANIPULATION TASK SOLVING USING
COGNITIVELY-INSPIRED CAUSAL LEARNING
BACHELOR THESIS

2024
MIROSLAV CIBULA

COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

ROBOTIC MANIPULATION TASK SOLVING USING
COGNITIVELY-INSPIRED CAUSAL LEARNING
BACHELOR THESIS

Study Programme: Applied Computer Science
Field of Study: Computer Science
Department: Department of Applied Informatics
Supervisor: prof. Ing. Igor Farkaš, Dr.
Consultant: Mgr. Michal Vavrečka, PhD.

Bratislava, 2024
Miroslav Cibula



ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Miroslav Cibula
Študijný program: aplikovaná informatika (Jednoodborové štúdium, bakalársky I. st., denná forma)
Študijný odbor: informatika
Typ záverečnej práce: bakalárska
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Robotic manipulation task solving using cognitively-inspired causal learning
Riešenie robotických manipulačných úloh pomocou kognitívne inšpirovaného kauzálneho učenia

Anotácia: Pozorovanie a učenie sa kauzálnych vzťahov v danom prostredí je dôležitý prvok kognície ľudí a vyšších živočíchov. Prítomnosť kauzálneho modelu sveta umožňuje agentovi predikovať efekt jeho akcií na prostredie. Tieto informácie dokáže následne v kombinácii s mentálnou simuláciou využiť na kreatívne a flexibilné plánovanie a univerzálne riešenie úloh v známom prostredí.

Cieľ:

1. Navrhnuť systém umožňujúci robotickému ramenu (agentovi) učiť sa kauzalitu manipuláciou s objektmi, reprezentovať a uchovávať tieto informácie, a aplikovať ich pri riešení úloh v známom prostredí.
2. Implementovať simulované randomizovateľné prostredie pre tréning a inferenciu systému a otestovať systém na sade experimentov v ňom.
3. Zanalyzovať výsledky z experimentov a efektívnosť systému.

Literatúra: Hellström, T. (2021). The relevance of causation in robotics: A review, categorization, and analysis. *Paladyn, Journal of Behavioral Robotics*, 12(1), 238–255. doi:10.1515/pjbr-2021-0017
Lee, T.E. et al. (2021). Causal reasoning in simulation for structure and transfer learning of robot manipulation policies. *IEEE International Conference on Robotics and Automation (ICRA)*. doi:10.1109/icra48506.2021.9561439
Vavrečka, M., Sokovnin, N., Mejdrechová, M., & Šejnová, G. (2021). MyGym: Modular Toolkit for Visuomotor Robotic tasks. *IEEE 33rd Int. Conf. on Tools with AI (ICTAI)*, 279–283. doi:10.1109/ictai52525.2021.00046

Vedúci: prof. Ing. Igor Farkaš, Dr.
Konzultant: Mgr. Michal Vavrečka, PhD.
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry: doc. RNDr. Tatiana Jajcayová, PhD.

Dátum zadania: 05.10.2023

Dátum schválenia: 10.10.2023

doc. RNDr. Damas Gruska, PhD.
garant študijného programu



Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

.....
študent

.....
vedúci práce



THESIS ASSIGNMENT

Name and Surname: Miroslav Cibula
Study programme: Applied Computer Science (Single degree study, bachelor I. deg., full time form)
Field of Study: Computer Science
Type of Thesis: Bachelor's thesis
Language of Thesis: English
Secondary language: Slovak

Title: Robotic manipulation task solving using cognitively-inspired causal learning

Annotation: Observing and learning causal relations in a given environment is an essential element of cognition in humans and high animals. A causal model of the world allows an agent to predict the effect of its actions on the environment. When combined with mental simulation, such information can be utilized for creative, flexible, and universal task-solving in a known environment.

Aim:

1. To design a system enabling a robotic arm (agent) to learn causality by manipulating objects, to represent and store this information, and to apply it in solving tasks in known environment.
2. To implement a simulated randomizable environment for the system training and inference and to test the system on a set of experiments in it.
3. To analyze results from the experiments and the overall system efficiency.

Literature: Hellström, T. (2021). The relevance of causation in robotics: A review, categorization, and analysis. *Paladyn, Journal of Behavioral Robotics*, 12(1), 238–255. doi:10.1515/pjbr-2021-0017
Lee, T.E. et al. (2021). Causal reasoning in simulation for structure and transfer learning of robot manipulation policies. *IEEE International Conference on Robotics and Automation (ICRA)*. doi:10.1109/icra48506.2021.9561439
Vavrečka, M., Sokovnin, N., Mejdrechová, M., & Šejnová, G. (2021). MyGym: Modular Toolkit for Visuomotor Robotic tasks. *IEEE 33rd Int. Conf. on Tools with AI (ICTAI)*, 279–283. doi:10.1109/ictai52525.2021.00046

Supervisor: prof. Ing. Igor Farkaš, Dr.
Consultant: Mgr. Michal Vavrečka, PhD.
Department: FMFI.KAI - Department of Applied Informatics
Head of department: doc. RNDr. Tatiana Jajcayová, PhD.

Assigned: 05.10.2023

Approved: 10.10.2023
doc. RNDr. Damas Gruska, PhD.
Guarantor of Study Programme

Student

Supervisor

Acknowledgments: Tu môžete poďakovať školiteľovi, prípadne ďalším osobám, ktoré vám s prácou nejako pomohli, poradili, poskytli dáta a podobne.

Abstrakt

Slovenský abstrakt v rozsahu 100-500 slov, jeden odstavec. Abstrakt stručne sumarizuje výsledky práce. Mal by byť pochopiteľný pre bežného informatika. Nemal by teda využívať skratky, termíny alebo označenie zavedené v práci, okrem tých, ktoré sú všeobecne známe.

Kľúčové slová: jedno, druhé, tretie (prípadne štvrté, piate)

Abstract

Abstract in the English language (translation of the abstract in the Slovak language).

Keywords:

Contents

Introduction	1
1 Preliminaries	3
1.1 Causal Learning	3
1.2 Forward and Inverse Models	5
1.3 Model Analysis	7
1.4 Sequence Modelling	8
1.5 Reinforcement Learning	8
2 Related Work	9
2.1 Causal Learning in Robotic Applications	9
2.2 Planning as a Sequence Modelling Problem	10
3 Aims and Task Formulation	11
4 Methods	13
4.1 Synthetic Data Generation	13
4.2 Forward and Inverse Models	13
4.3 Knowledge Extraction	13
4.4 Planning	13
5 Experiments and Results	15
5.1 Learning Kinematics	15
5.2 Simple Intuitive Physics	15
5.3 Task Solving	15
Conclusion	17

List of Figures

1.1	Diagram of robot causal cognition categorization.	4
-----	---	---

List of Tables

List of Abbreviations

C1	robot causal cognition category 1	4
C2	robot causal cognition category 2	4
FM	forward model	5
HIM	hindsight information matching	10
IM	inverse model	5
LSTM	long short-term memory	8
MDP	Markov decision process	8
MLP	multilayer perceptron	6
RCSL	return-conditioned supervised learning	10
RL	reinforcement learning	8
RNN	recurrent neural network	8
RvS	reinforcement learning via supervised learning . .	10
XAI	explainable artificial intelligence	7

List of Symbols

x	scalar
\mathbf{v}	vector
\mathbf{M}	matrix
$\mathbf{u} \subseteq \mathbf{v}$	subvector of vector \mathbf{v}
$\mathbf{u} \subset \mathbf{v}$	proper subvector of vector \mathbf{v}
$\mathbf{v} \setminus \mathbf{u}$	vector \mathbf{v} without its subvector \mathbf{u}
$\mathbf{v} \# \mathbf{u}$	vector concatenation
$\mathcal{P}(A)$	power set of set A

Introduction

Observing and learning causal relations in a given environment is an essential element of cognition in humans and other higher animals. Thanks to this ability, agents can assemble their intuitive knowledge (such as intuitive physics and psychology) about the world in which they operate from multiple observations and use them to further predict the environment's behaviour, mainly in response to their actions. Such ability is principal to common sense understanding – a concept mastered even by young children while having proven to be highly inapprehensible for artificial intelligence.

In this thesis, we were inspired by causal learning and other mechanisms observed in human cognition, leveraging them in the construction and study of a system solving robotic manipulation tasks in a simulated environment. Specifically, we use forward and inverse models to learn the effects of actions performed by an agent (simulated robotic arm). These actions are a product of simple motor babbling or other strategies allowing the agent to interact with the environment and observe its behaviour.

Further, as a by-product of this approach, we hypothesize that these models trained on a sufficient amount of observations contain knowledge about the environment and the task the agent was performing. We argue that this knowledge is similar to intuitive theories assembled by humans from causal experience collected since an early age. As such knowledge can be helpful in the analysis of the environment, the task, and their properties, we explore methods for extracting this information by analyzing the trained forward model using explainable artificial intelligence methods.

Finally, we propose a system for solving more complex robotic manipulation tasks. We use sequence modelling for preliminary trajectory generation, subsequently post-processed with the aid of the trained forward and inverse models. We were inspired by imitation learning, utilizing it for the sequence modelling optimization within supervised learning paradigm instead of in robotics more commonly used reinforcement learning approach.

[Short overview of the following chapters]

Chapter 1

Preliminaries

In this chapter, serving as a theoretical overview, we summarize methods, approaches and technologies used in the thesis. Specifically, we provide an overview of causality and causal learning from both human cognition and machine intelligence point of view as we use causal learning as a central concept in our proposed methods (Chapter 4).

We further describe the biological and robotic backgrounds of forward and inverse models used as facilitators for causal learning as well as artificial neural networks used for their implementation. In addition, as we leverage the models' analysis, we summarize the principles behind the family of analysis methods used.

Lastly, we describe the principal components of sequence modelling and reinforcement learning used for the proposed planning method.

1.1 Causal Learning

Causal learning refers to capturing and learning causal relationships from observations of the behaviour of an environment in which the agent (e.g., human or robot) operates. This ability allows agents to form intuitive theories and use them to predict the environment's behaviour in response to their actions (Gerstenberg and Tenenbaum, 2017) establishing common sense understanding including the knowledge of intuitive physics and psychology (Lake et al., 2016).

Human Causal Cognition

Causal cognition has been studied extensively on both human and machine intelligence levels. Regarding human cognition, Gärdenfors and Lombard (2018; 2017) propose a causal cognition evolution model categorizing levels of causal understanding varying in complexity. This model's grades range from understanding the perceived effects of the agent's motor actions to understanding interactions between entities of

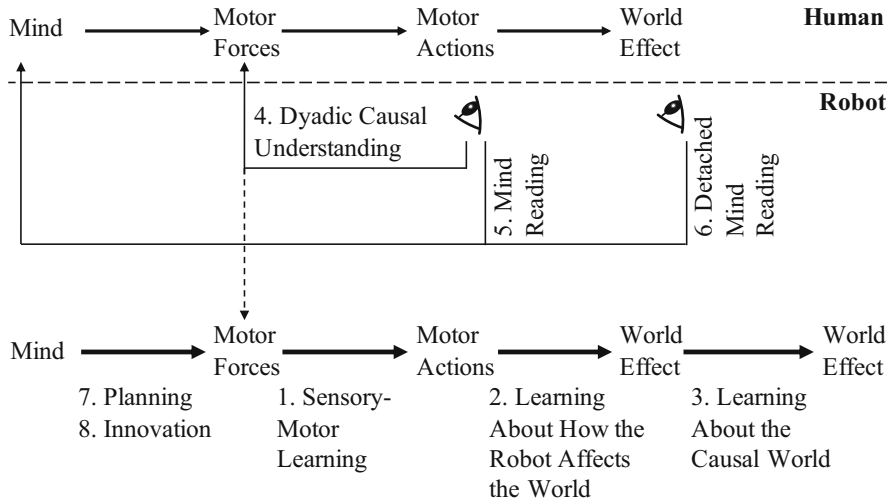


Figure 1.1: Diagram of robot causal cognition categorization. Bold arrows refer to learning causal relations, while thin solid arrows refer to “inference of causes related to an interacting human”. Adjacent categories are required to facilitate the agent’s understanding of respective relationships (Hellström, 2021).

the environment and the ability to extrapolate from this knowledge.

Causality in Robotics

Regarding machine intelligence, Lake et al. (2014; 2016) argue that causality might be one of three central “ingredients” needed to replicate rapid learning akin to human learning. The argument supports the current effort to transfer causal cognition to robotics, involving embodied agents interacting with the world. Analogically to the model of the evolution of human causal cognition mentioned above, Hellström (2021) proposes a categorization of robot causal cognition ranging in difficulty from simple sensory-motor learning to the ability to plan and beyond. These grades are divided into three groups: learning causal relations, inferring the causes related to an interacting human, and robot deciding how to act (Figure 1.1).

In this work, we focus on low-level causality regarding two categories: sensorimotor self-learning (abbr. C1) and learning the consequences of agent’s own actions on objects in the environment (abbr. C2).

Causality in Machine Learning

Besides applications in robotics, causality is also studied as part of the machine learning research as Zhang et al. (2017) and Zhu et al. (2020) argue that causality understanding can be beneficial toward building more robust models with common sense.

Causal learning is predominantly part of the symbolic paradigm (Kotseruba and

Tsotsos, 2018) as it mainly operates with symbolic representations on different levels (Schölkopf, 2022). This is in contrast with the sub-symbolic models and systems (Rosenblatt, 1958) generally following the parallel distributed processing paradigm (McClelland et al., 1987, 1988) inspired by low-level brain mechanisms and neural structures.

For a further comprehensive overview of causal cognition in humans, in robots, and causality research in machine learning, see (Gärdenfors and Lombard, 2018), (Hellström, 2021), and (Schölkopf, 2022; Zhang et al., 2017), respectively.

1.2 Forward and Inverse Models

Causal and especially sensorimotor knowledge produced by causal learning performed by a robotic system, solving C1 and C2 tasks in our case, can be represented by a pair of complementary internal models: the forward model (abbr. FM) and the inverse model (abbr. IM) (Wolpert and Kawato, 1998).

While the FM (Dearden and Demiris, 2005) unambiguously predicts perceivable consequences of the agent’s actions, the IM predicts actions needed to reach the desired state from the initial state. In contrast to the FM, the IM is mathematically ill-defined in general, as the IM also models inverse kinematics, which is ill-posed in redundant robots (Nguyen-Tuong and Peters, 2011).

Forward and inverse models in robotics take inspiration from the internal model principle of control theory (Francis and Wonham, 1976) modelling physiological internal models (Wolpert and Flanagan, 2001; Miall and Wolpert, 1996; Sperry, 1950; von Holst and Mittelstaedt, 1950). It is generally acknowledged (Dogge et al., 2019) that humans use a forward internal model to predict the outcomes of their motor actions. Dogge et al. (2019) describe “[physiological] forward models [...] as simulations of the motor system that use a copy of the motor command, known as an efference copy [...], to predict the sensory consequences of the action in question (known as corollary discharge)”.

It should be noted that while in this thesis we use forward models to predict environment-related outcomes beyond body-related outcomes, Dogge et al. (2019) argue that involvement of biological motor-based forward models to such extent is “limited and hitherto unjustified”.

Artificial Neural Networks

As both models represent functions, modelling (learning) can be performed by artificial neural networks as universal function approximators (Hornik et al., 1989). For the forward and inverse modelling in this work, we specifically use multilayer perceptrons (abbr. MLP) (Rosenblatt, 1958; Rosenblatt et al., 1962; Minsky and Papert, 2017) as universal regressors.

The topology of MLP models consists of L layers of neural units with layers $l = 1$, $l = L$ and $1 < l < L$ defined as an input layer, output layer and hidden layers, respectively. Each layer is composed of d_l neural units, with $d_1 = \dim(\mathbf{x})$ and $d_L = \dim(\mathbf{y})$ where \mathbf{x} and \mathbf{y} are real input and output vectors of a function f being approximated. MLP is fully connected, meaning that each unit i of each layer l except the output layer is connected with every neuron j of the subsequent layer $l + 1$ using oriented synapse with assigned weight $w_{ij}^{(l)} \in \mathbb{R}$. Then, activation of i -th neuron in l -th layer can be computed as

$$h_i^{(l)} = \varphi_l \left(\sum_{j=1}^{d_{l+1}} w_{ij}^{(l)} h_j^{(l-1)} \right), \quad (1.1)$$

where φ_l denotes activation function of the l -th layer. Additionally, $\mathbf{h}^{(L)} \equiv \hat{\mathbf{y}}$ where $\hat{\mathbf{y}}$ is a predicted output. To reformulate, considering layer a function

$$h^{(l)}(\mathbf{v}) \triangleq \left[\varphi_l \left(\sum_{j=1}^{d_{l+1}} w_{ij}^{(l)} v_j \right) \right]_{1 \leq i \leq d_l}, \quad (1.2)$$

the approximation of sought function f can be defined as

$$\hat{f}(\mathbf{x}) \triangleq (h^{(L)} \circ h^{(L-1)} \circ \dots \circ h^{(2)} \circ h^{(1)})(\mathbf{x}). \quad (1.3)$$

In order to train the regressor, the model's weights are commonly optimized in a supervised learning scheme where the error of generated predictions $\hat{\mathbf{y}}$ is computed against the ground-truth targets \mathbf{y} using an error function $\mathcal{L}(\hat{\mathbf{y}}, \mathbf{y})$. The computed error is further backpropagated (Rumelhart et al., 1986) through the whole network, with new weights being calculated as

$$w_{ij}^{(l)}(t+1) = w_{ij}^{(l)}(t) + \Delta w_{ij}^{(l)}, \quad (1.4)$$

where $\Delta w_{ij}^{(l)}$ denotes weight adjustment defined as

$$\Delta w_{ij}^{(l)} = -\eta \frac{\partial \mathcal{L}}{\partial w_{ij}^{(l)}} \quad (1.5)$$

with η designating learning rate constant.

1.3 Model Analysis

In this thesis, we analyze neural models using explainable AI (abbr. XAI) methods. Specifically, we study feature importance, evaluating the significance of a specific input feature on the prediction of a specific output feature.

Shapley Values

Feature importance is commonly computed using Shapley values (Shapley, 1953) while interpreting the prediction task for a single data point \mathbf{x} as a cooperative game. Input features are interpreted as players belonging to possible coalitions $S \in \mathcal{P}(F)$, where F is the set of all features. Then, interpreting a model f trained on a set of features as a value function evaluating the worth of coalition, Shapley value ϕ_i defines the marginal contribution of feature i for the input \mathbf{x} for the model f :

$$\phi_i(\mathbf{x}) = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|! (|F| - |S| - 1)!}{|F|!} \Delta_i(S, \mathbf{x}), \quad (1.6)$$

where $\Delta_i(S, \mathbf{x})$ denotes the marginal contribution of feature i to coalition S defined as

$$\Delta_i(S, \mathbf{x}) = f_{S \cup \{i\}}(\mathbf{x}_{S \cup \{i\}}) - f_S(\mathbf{x}_S) \quad (1.7)$$

with f_S and $f_{S \cup \{i\}}$ denoting trained models on the feature subset S and S including the feature i , respectively, and \mathbf{x}_S denoting values of the input features from S .

SHAP Methods

As computing Shapley values is generally NP-hard (Matsui and Matsui, 2001), various methods for estimating them have been developed, with the most popular being SHAP (Lundberg and Lee, 2017), which unifies different additive feature attribution methods. Here, we were experimenting with two variants: KernelSHAP and DeepSHAP.

KernelSHAP is a model-agnostic kernel-based method utilizing the idea of local surrogate models (Ribeiro et al., 2016) to estimate Shapley values. However, since it does not make any assumptions about the analyzed model, it is generally slower than model-specific methods on account of its combinatorial nature. DeepSHAP, on the other hand, is applicable only to neural models as it uses attribution rules of DeepLIFT method (Shrikumar et al., 2017) to propagate SHAP values from the output layer back to the input layer.

SHAP methods are local, providing an explanation for one prediction. However, thanks to their properties, these local explanations can be aggregated across the set of instances, providing global feature importance within the analyzed model.

For further comprehensive review of XAI methods, see (Gilpin et al., 2018).

1.4 Sequence Modelling

[Definition]

[RNN/LSTM + Transformers]

1.5 Reinforcement Learning

Reinforcement learning (abbr. RL) is a machine learning paradigm of algorithms learning by interacting with a given environment. The principal objective of agents controlled by these algorithms is to perform actions maximizing the cumulative reward. Although methods proposed in this work operate beyond the RL paradigm, we reuse some of its principles and nomenclature.

Problems solvable by RL methods are most commonly modelled as Markov decision processes (abbr. MDP) (Bellman, 1957). Discrete-time MDP can be defined as a tuple

$$\text{MDP}: (\mathcal{S}, \mathcal{A}, T, R), \quad (1.8)$$

where \mathcal{S} and \mathcal{A} denote discrete or continuous state and action space, respectively. Then, in a discrete timestep t , an action $\mathbf{a}(t) \in \mathcal{A}$ transitions state $\mathbf{s}(t)$ to $\mathbf{s}(t+1)$, with $\mathbf{s}(t), \mathbf{s}(t+1) \in \mathcal{S}$, with probability

$$P(\mathbf{s}(t+1) | \mathbf{s}(t), \mathbf{a}(t)) \equiv T(\mathbf{s}(t), \mathbf{a}(t), \mathbf{s}(t+1)), \quad (1.9)$$

where T denotes a state transition function. Performing action $\mathbf{a}(t)$ at state $\mathbf{s}(t)$ is evaluated by reward function $R(\mathbf{s}(t), \mathbf{a}(t))$.

[Trajectory]

[Cumulative reward of trajectory + policy + aim of RL]

Chapter 2

Related Work

In this chapter, we provide an overview of existing full or partial solutions alternative to the main contribution points of this thesis: learning causality in robotics (Section 2.1) and planning using sequence modelling aided by causal models (Section 2.2).

2.1 Causal Learning in Robotic Applications

Albeit causal learning (for an overview, see Section 1.1) and causality-based approaches in the context of robotics are presently deemed under-explored (Stocking et al., 2022; Lee et al., 2023), it has been demonstrated that they can be helpful for multiple applications. In many robotics and reinforcement learning applications, causality acts as an attention mechanism revealing relationships between state and action variables regarding a given environment or task. This information is often used to reduce the complexity of either space or identify relevant or important variables.

Our work, especially the knowledge extraction part (Section 4.3), was inspired by CREST (Lee et al., 2021), where authors used causal reasoning in simulation to learn the relevant state space variables for a robot manipulation policy. In their approach, they conduct causal interventions (in a scheme akin to the randomized controlled trials (Fisher, 1925)) to elicit the relationships between action and state variables. This allows them to reduce the complexity of neural network policies using only state variables relevant to the task being solved.

Leveraging the research on CREST, SCALE approach (Lee et al., 2023) for discovering and learning diverse robot skills has been proposed. SCALE uses CREST in a pipeline to identify sets of relevant variables related to individual skills.

A method proposed by Diehl and Ramirez-Amaro (2023) concerns a causal Bayesian network (Pearl, 1985) learning causal relationships between task executions and their consequences. They then utilize this model to allow a robot to “conjecture” whether

and why the action executed in its current state will succeed or fail.

Further, Sontakke et al. (2021) introduce causal curiosity, an intrinsic reward allowing the agent to discover latent causal factors in the dynamics of the environment it operates in. Wang et al. (2022) use the causal dynamics model to remove unnecessary dependencies between the state and action variables and subsequently use the dynamics model to yield state abstractions. Sonar et al. (2021) utilize causality to learn invariant policies.

2.2 Planning as a Sequence Modelling Problem

As stated in Section 1.5, planning is predominantly a reinforcement learning problem. However, as recently demonstrated by Chen et al. (2021) and Janner et al. (2021), planning can also be achieved beyond the RL paradigm. The mentioned approaches use sequence modelling (for an overview, see Section 1.4) in combination with imitation learning to generate trajectories needed to complete a given task. As such, the problem of online RL is being shifted to the domain of supervised learning. Specifically, RL components are often entirely replaced with offline behavioural cloning (Furuta et al., 2021).

Inspired by the prior research, Wen et al. (2022) propose their own Transformer architecture for solving cooperative multi-agent RL problems.

Furuta et al. (2021) further demonstrate that these approaches perform hindsight information matching (abbr. HIM). They define HIM as a method concerning “training policies that can output the rest of trajectory that matches some statistics of future state information” and propose a Generalized Decision Transformer capable of solving any HIM problem.

The paradigm covering these approaches has been coined as return-conditioned supervised learning (abbr. RCSL), whose central idea “is to learn the return-conditional distribution of actions in each state, and then define a policy by sampling from the distribution of actions that receive high return” (Brandfonbrener et al., 2022). A related broader concept has been referred to as reinforcement learning via supervised learning (abbr. RvS) (Emmons et al., 2021).

Planning leveraging the trajectory modelling in a supervised learning scheme inherently requires a training dataset. For this reason, most approaches mentioned above employ imitation learning (Zare et al., 2023; Mandlekar et al., 2021), a process in which an expert demonstrates a desired behaviour and an agent learns by imitation from the collected observations of expert demonstrations. Alternatively, Oh et al. (2018) propose self-imitation learning during which the agent imitates its own past good experiences.

Chapter 3

Aims and Task Formulation

Chapter 4

Methods

4.1 Synthetic Data Generation

4.2 Forward and Inverse Models

4.3 Knowledge Extraction

4.4 Planning

Chapter 5

Experiments and Results

5.1 Learning Kinematics

5.2 Simple Intuitive Physics

5.3 Task Solving

Conclusion

Bibliography

- Bellman, R. (1957). A Markovian decision process. *Journal of Mathematics and Mechanics*, 6(5):679–684.
- Brandfonbrener, D., Bietti, A., Buckman, J., Laroche, R., and Bruna, J. (2022). When does return-conditioned supervised learning work for offline reinforcement learning? In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A., editors, *Advances in Neural Information Processing Systems*, volume 35, pages 1542–1553. Curran Associates, Inc.
- Chen, L., Lu, K., Rajeswaran, A., Lee, K., Grover, A., Laskin, M., Abbeel, P., Srinivas, A., and Mordatch, I. (2021). Decision Transformer: Reinforcement learning via sequence modeling. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. S., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 15084–15097. Curran Associates, Inc.
- Dearden, A. M. and Demiris, Y. (2005). Learning forward models for robots. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence, IJ-CAI’05*, page 1440–1445, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Diehl, M. and Ramirez-Amaro, K. (2023). A causal-based approach to explain, predict and prevent failures in robotic tasks. *Robotics and Autonomous Systems*, 162:104376.
- Dogge, M., Custers, R., and Aarts, H. (2019). Moving forward: On the limits of motor-based forward models. *Trends in Cognitive Sciences*, 23(9):743–753.
- Emmons, S., Eysenbach, B., Kostrikov, I., and Levine, S. (2021). RvS: What is essential for offline RL via supervised learning?
- Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh, UK.
- Francis, B. A. and Wonham, W. M. (1976). The internal model principle of control theory. *Automatica*, 12(5):457–465.

- Furuta, H., Matsuo, Y., and Gu, S. S. (2021). Generalized decision transformer for offline hindsight information matching.
- Gerstenberg, T. and Tenenbaum, J. B. (2017). Intuitive theories. In Waldmann, M. R., editor, *The Oxford Handbook of Causal Reasoning*, pages 515–548. Oxford University Press.
- Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., and Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, pages 80–89. IEEE.
- Gärdenfors, P. and Lombard, M. (2018). Causal cognition, force dynamics and early hunting technologies. *Frontiers in Psychology*, 9.
- Hellström, T. (2021). The relevance of causation in robotics: A review, categorization, and analysis. *Paladyn, Journal of Behavioral Robotics*, 12(1):238–255.
- Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366.
- Janner, M., Li, Q., and Levine, S. (2021). Offline reinforcement learning as one big sequence modeling problem. In Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. S., and Vaughan, J. W., editors, *Advances in Neural Information Processing Systems*, volume 34, pages 1273–1286. Curran Associates, Inc.
- Kotseruba, I. and Tsotsos, J. K. (2018). 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artificial Intelligence Review*, 53(1):17–94.
- Lake, B. M. (2014). *Towards more human-like concept learning in machines: Compositionality, causality, and learning-to-learn*. PhD thesis, Massachusetts Institute of Technology.
- Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J. (2016). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40.
- Lee, T. E., Vats, S., Girdhar, S., and Kroemer, O. (2023). SCALE: Causal learning and discovery of robot manipulation skills using simulation. In Tan, J., Toussaint, M., and Darvish, K., editors, *Proceedings of The 7th Conference on Robot Learning*, volume 229 of *Proceedings of Machine Learning Research*, pages 2229–2256. PMLR.

- Lee, T. E., Zhao, J. A., Sawhney, A. S., Girdhar, S., and Kroemer, O. (2021). Causal reasoning in simulation for structure and transfer learning of robot manipulation policies. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE.
- Lombard, M. and Gärdenfors, P. (2017). Tracking the evolution of causal cognition in humans. *Journal of Anthropological Sciences*, 95:219–234.
- Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, volume 30 of *NIPS’17*, page 4768–4777.
- Mandlekar, A., Xu, D., Wong, J., Nasiriany, S., Wang, C., Kulkarni, R., Fei-Fei, L., Savarese, S., Zhu, Y., and Martín-Martín, R. (2021). What matters in learning from offline human demonstrations for robot manipulation.
- Matsui, Y. and Matsui, T. (2001). NP-completeness for calculating power indices of weighted majority games. *Theoretical Computer Science*, 263(1–2):305–310.
- McClelland, J. L., Rumelhart, D. E., and Hinton, G. E. (1988). The appeal of parallel distributed processing. In *Readings in Cognitive Science*, pages 52–72. Elsevier.
- McClelland, J. L., Rumelhart, D. E., and the PDP Research Group (1987). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2: Psychological and Biological Models*, volume 2. The MIT Press.
- Miall, R. C. and Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*, 9(8):1265–1279.
- Minsky, M. and Papert, S. A. (2017). *Perceptrons: An Introduction to Computational Geometry*. The MIT Press, Cambridge, Massachusetts.
- Nguyen-Tuong, D. and Peters, J. (2011). Model learning for robot control: a survey. *Cognitive Processing*, 12(4):319–340.
- Oh, J., Guo, Y., Singh, S., and Lee, H. (2018). Self-imitation learning. In Dy, J. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 3878–3887. PMLR.
- Pearl, J. (1985). Bayesian networks: A model of self-activated memory for evidential reasoning. In *Proceedings of the 7th conference of the Cognitive Science Society, University of California, Irvine, CA, USA*, pages 15–17.

- Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 1135–1144. Association for Computing Machinery.
- Rosenblatt, F. (1958). The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408.
- Rosenblatt, F. et al. (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*, volume 55. Spartan Books, Washington, DC.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.
- Schölkopf, B. (2022). *Causality for Machine Learning*, pages 765–804. Association for Computing Machinery, New York, NY, USA, 1 edition.
- Shapley, L. S. (1953). A value for n-person games. In Kuhn, H. W. and Tucker, A. W., editors, *Contributions to the Theory of Games II*, pages 307–317. Princeton University Press, Princeton.
- Shrikumar, A., Greenside, P., and Kundaje, A. (2017). Learning important features through propagating activation differences. In *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 3145–3153. PMLR.
- Sonar, A., Pacelli, V., and Majumdar, A. (2021). Invariant policy optimization: Towards stronger generalization in reinforcement learning. In Jadbabaie, A., Lygeros, J., Pappas, G. J., Parrilo, P., Recht, B., Tomlin, C. J., and Zeilinger, M. N., editors, *Proceedings of the 3rd Conference on Learning for Dynamics and Control*, volume 144 of *Proceedings of Machine Learning Research*, pages 21–33. PMLR.
- Sontakke, S. A., Mehrjou, A., Itti, L., and Schölkopf, B. (2021). Causal curiosity: RL agents discovering self-supervised experiments for causal representation learning. In Meila, M. and Zhang, T., editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 9848–9858. PMLR.
- Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology*, 43(6):482–489.
- Stocking, K. C., Gopnik, A., and Tomlin, C. (2022). From robot learning to robot understanding: Leveraging causal graphical models for robotics. In Faust, A., Hsu,

- D., and Neumann, G., editors, *Proceedings of the 5th Conference on Robot Learning*, volume 164 of *Proceedings of Machine Learning Research*, pages 1776–1781. PMLR.
- von Holst, E. and Mittelstaedt, H. (1950). Das Reafferenzprinzip: Wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Naturwissenschaften*, 37(20):464–476.
- Wang, Z., Xiao, X., Xu, Z., Zhu, Y., and Stone, P. (2022). Causal dynamics learning for task-independent state abstraction. In Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., and Sabato, S., editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 23151–23180. PMLR.
- Wen, M., Kuba, J., Lin, R., Zhang, W., Wen, Y., Wang, J., and Yang, Y. (2022). Multi-agent reinforcement learning is a sequence modeling problem. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A., editors, *Advances in Neural Information Processing Systems*, volume 35, pages 16509–16521. Curran Associates, Inc.
- Wolpert, D. M. and Flanagan, J. R. (2001). Motor prediction. *Current Biology*, 11(18):R729–R732.
- Wolpert, D. M. and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, 11(7–8):1317–1329.
- Zare, M., Kebria, P. M., Khosravi, A., and Nahavandi, S. (2023). A survey of imitation learning: Algorithms, recent developments, and challenges.
- Zhang, K., Schölkopf, B., Spirtes, P., and Glymour, C. (2017). Learning causality and causality-related learning: Some recent progress. *National Science Review*, 5(1):26–29.
- Zhu, Y., Gao, T., Fan, L., Huang, S., Edmonds, M., Liu, H., Gao, F., Zhang, C., Qi, S., Wu, Y. N., Tenenbaum, J. B., and Zhu, S.-C. (2020). Dark, beyond deep: A paradigm shift to cognitive AI with humanlike common sense. *Engineering*, 6(3):310–345.