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 Informatics
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

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

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. Navrhnúť 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

52373788

Comenius University Bratislava
Faculty of Mathematics, Physics and Informatics

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

List of Figures

List of Tables

List of Abbreviations

List of Symbols

x	scalar
$\boldsymbol{v}$	vector
$\boldsymbol{M}$	matrix
$\boldsymbol{u} \subseteq \boldsymbol{v}$	subvector of vector $\boldsymbol{v}$
$\boldsymbol{u} \subset \boldsymbol{v}$	proper subvector of vector $\boldsymbol{v}$
$\boldsymbol{v} \setminus \boldsymbol{u}$	vector $\boldsymbol{v}$ without its subvector $\boldsymbol{u}$
$\boldsymbol{v} + \boldsymbol{u}$	vector concatenation
$\mathcal{P}(A)$	power set of set A
$\lvert A \rvert$	cardinality of set A
$\mathbb{E}[\cdot]$	expected value

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 & 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

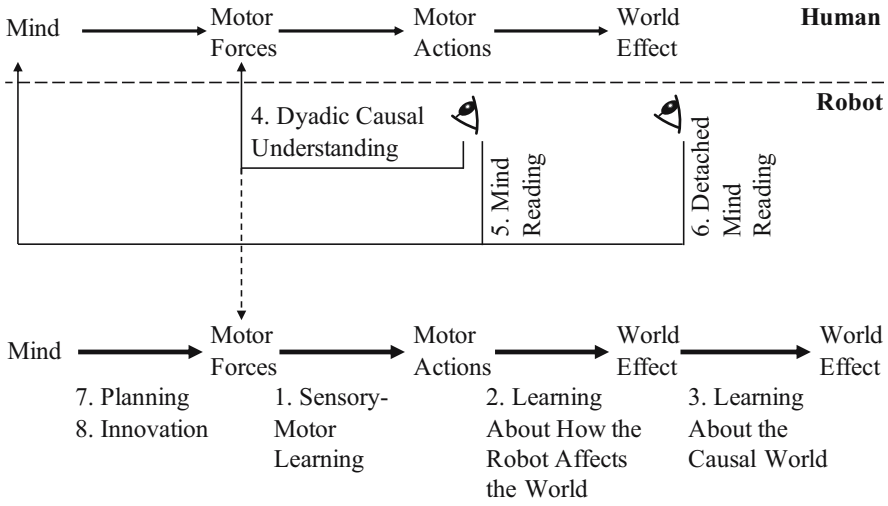

Figure 1.1: Diagram of robot causal cognition categorization. Bold arrows refer to learning causal relations, while thin solid arrows refer to "inference of causes related to an interacting human". Adjacent categories are required to facilitate the agent's understanding of respective relationships (Hellström,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

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 & 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 & Kawato, 1998).

While the FM (Dearden & 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 & Peters, 2011).

Forward and inverse models in robotics take inspiration from the internal model principle of control theory (Francis & Wonham, 1976) modelling physiological internal models (Miall & Wolpert, 1996; Sperry, 1950; von Holst & Mittelstaedt, 1950; Wolpert & Flanagan, 2001). 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) (Minsky & Papert, 2017; Rosenblatt, 1958; Rosenblatt et al., 1962) 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 $d_1 = \dim(\boldsymbol{x})$ and $d_L = \dim(\boldsymbol{y})$ where $\boldsymbol{x}$ and $\boldsymbol{y}$ 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

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

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

$$h^{(l)}(\boldsymbol{v}) \triangleq \varphi_l \left(W^{(l)} \boldsymbol{v} \right) \tag{1.2}$$

where $W^{(l)}$ is the weight matrix of the l-th layer, the approximation of sought function f can be defined as

$$\hat{f}(\boldsymbol{x}) \triangleq \left(h^{(L)} \circ h^{(L-1)} \circ \cdots \circ h^{(2)} \circ h^{(1)} \right) (\boldsymbol{x}). \tag{1.3}$$

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

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

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

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

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 $\boldsymbol{x}$ 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 $\boldsymbol{x}$ for the model f:

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

where $|S|$ and $|F|$ denote number of features in the coalition and the total number of feature, respectively. $\Delta_i(S, \boldsymbol{x})$ denotes the marginal contribution of feature i to coalition S defined as

$$\Delta_i(S, \boldsymbol{x}) = f_{S \cup \{i\}}\left(\boldsymbol{x}_{S \cup \{i\}}\right) - f_S\left(\boldsymbol{x}_S\right) \tag{1.7}$$

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

SHAP Methods

As computing Shapley values is generally NP-hard (Matsui & Matsui, 2001), various methods for estimating them have been developed, with the most popular being SHAP (Lundberg & 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

$$\text{MDP}\colon\ (\mathcal{S}, \mathcal{A}, T, R, \gamma)\,, \tag{1.8}$$

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

$$P\left(\boldsymbol{s}(t+1) \mid \boldsymbol{s}(t), \boldsymbol{a}(t)\right) \equiv T\left(\boldsymbol{s}(t), \boldsymbol{a}(t), \boldsymbol{s}(t+1)\right), \tag{1.9}$$

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

The process of taking action in a particular state produces a trajectory

$$\tau = [\boldsymbol{s}(0), \boldsymbol{a}(0), r(0), \boldsymbol{s}(1), \boldsymbol{a}(1), r(1), \ldots] \tag{1.10}$$

which can be evaluated by computing its discounted cumulative return

$$G = \sum_{t=0}^{\infty} \gamma^t r(t) \tag{1.11}$$

where $0 \leq \gamma \leq 1$ is a discount factor. The goal of RL algorithms is to learn optimal policy π^* such that

$$\pi^* = \underset{\pi}{\operatorname{argmax}}\, V^{\pi}\left[\boldsymbol{s}(0)\right] \tag{1.12}$$

where V^{π} denotes a state-value function defined as

$$V^{\pi}(\boldsymbol{s}(0)) = \mathbb{E}\left[G \mid \boldsymbol{s}(0), \pi\right] \tag{1.13}$$

providing the expected cumulative return of a trajectory produced by an agent starting in state $\boldsymbol{s}(0)$ and taking action $\boldsymbol{a}(t) \sim \pi(\boldsymbol{s}(t))$ with stochastic policy $\pi(\boldsymbol{s}(t)) \equiv P(\boldsymbol{a}(t) \mid \boldsymbol{s}(t))$.

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 (T. E. Lee et al., 2023; Stocking et al., 2022), 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 (T. E. 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 (T. E. 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.

Furthermore, 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 (Mandlekar et al., 2021; Zare et al., 2023), 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

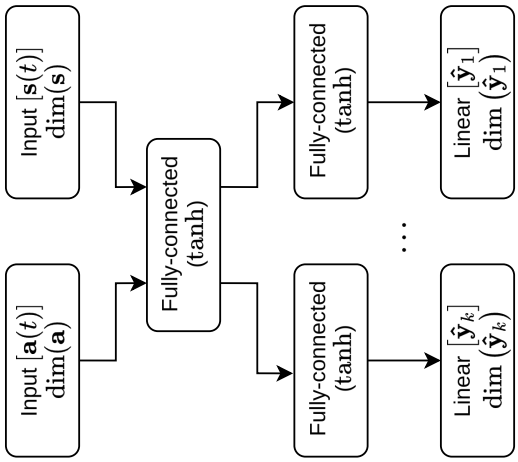

Figure 4.1: General forward model architecture.

4.3 Knowledge Extraction

4.4 Planning

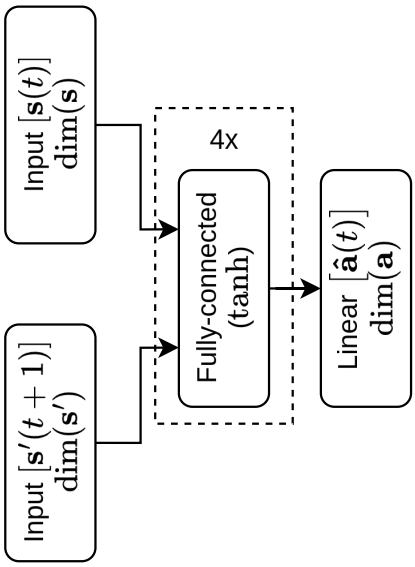

Figure 4.2: General monolithic inverse model architecture.

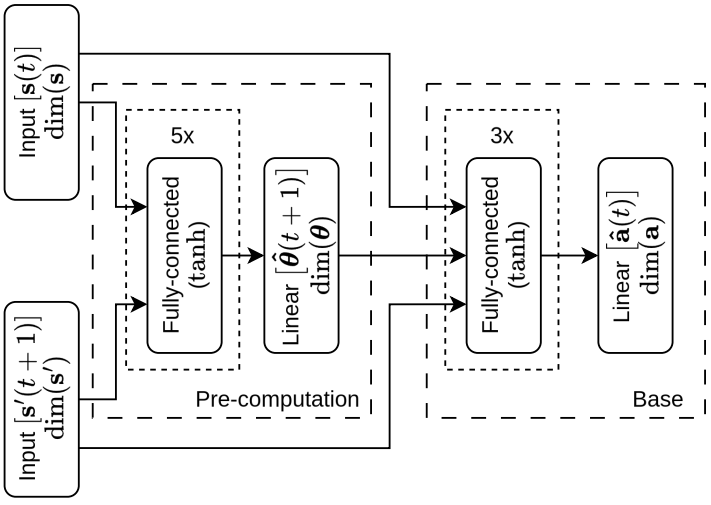

Figure 4.3: Inverse model architecture with $\boldsymbol{\theta}(t+1)$ pre-computation pre-network.

Chapter 5

Experiments and Results

5.1 Learning Kinematics

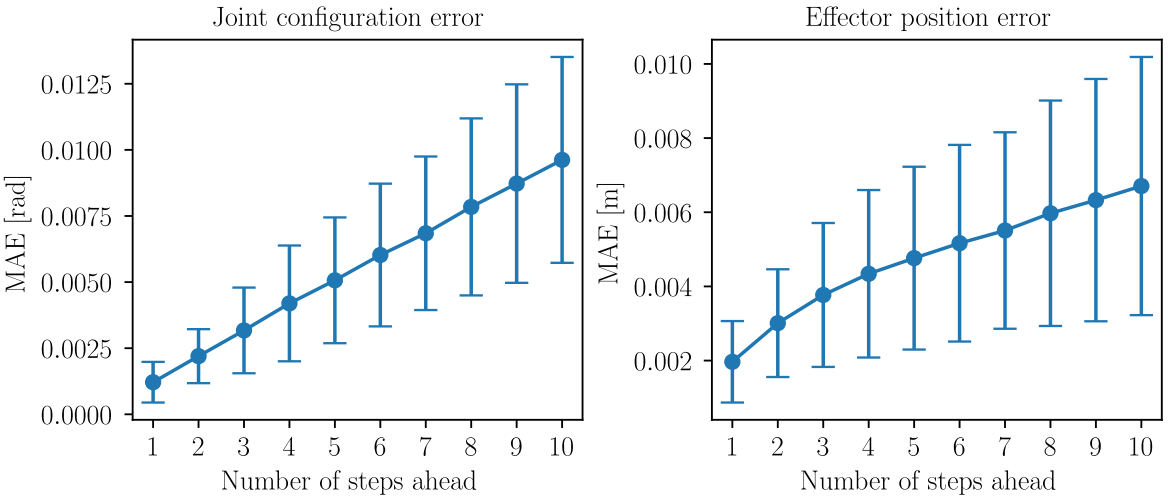

Figure 5.1: Error of the forward model during mental simulation 10 steps ahead.

5.2 Simple Intuitive Physics

5.3 Task Solving

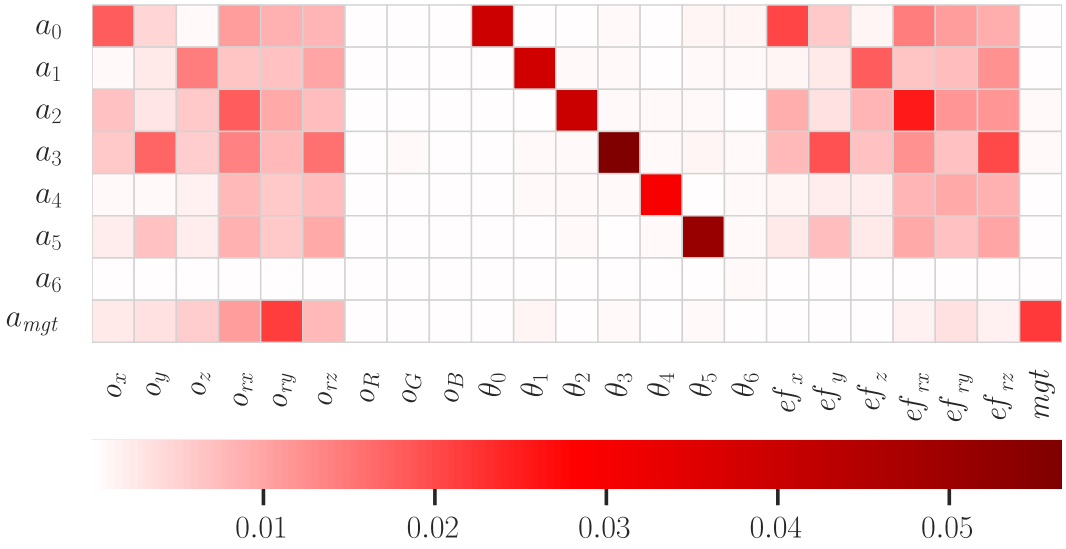

Figure 5.2: Contribution heat map generated by Deep SHAP method on the forward model showing magnitude of contribution of specific actions to output features.

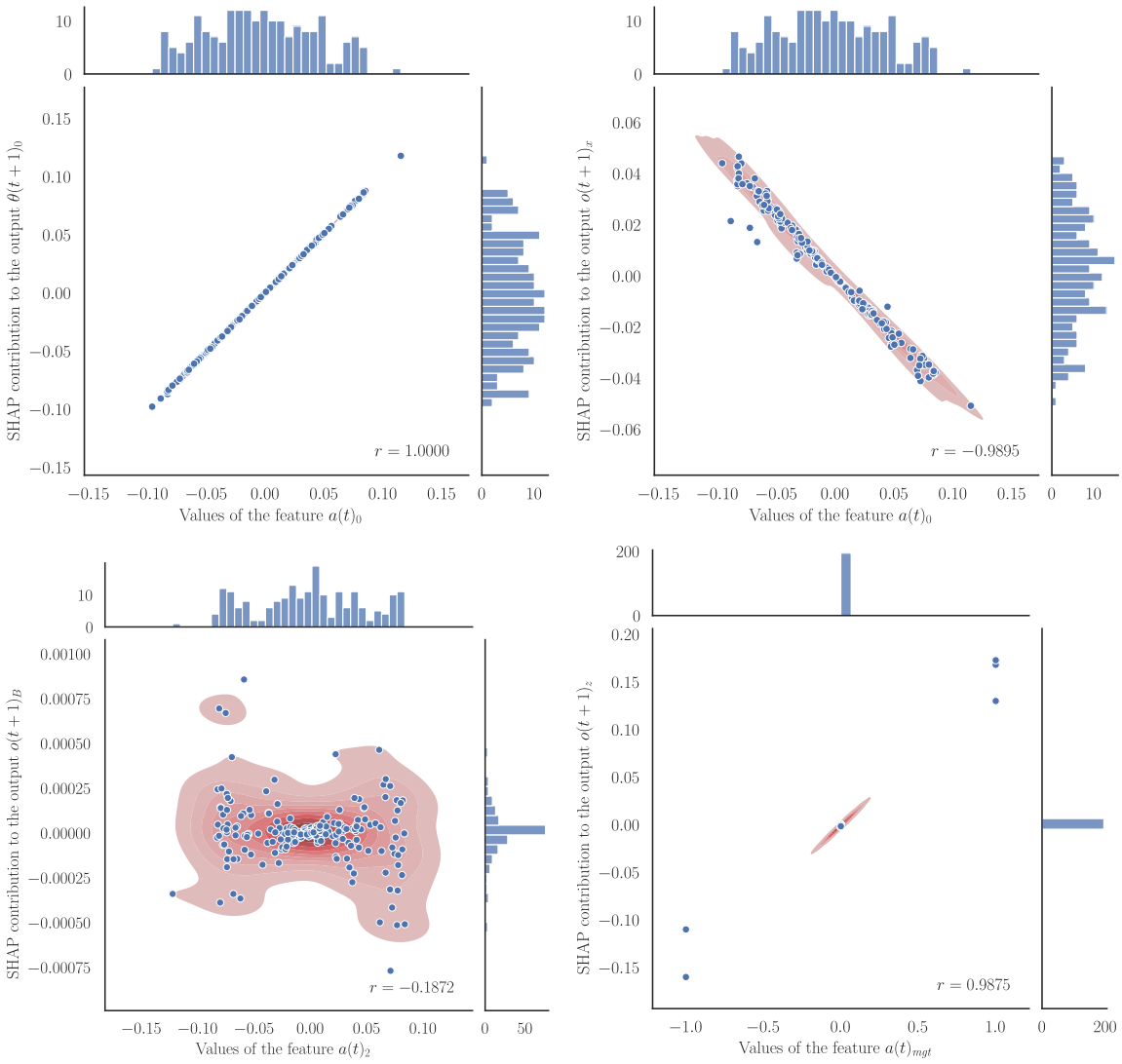

Figure 5.3: A sample of partial dependence plots generated by Deep SHAP method applied to the forward model showing correlation between a value of a specific action component and its contribution to an output variable.

Conclusion

Bibliography

Bellman, R. (1957). A Markovian decision process. *Journal of Mathematics and Mechanics*, *6*(5), 679–684. Retrieved April 30, 2024, from http://www.jstor.org/stable/24900506

Brandfonbrener, D., Bietti, A., Buckman, J., Laroche, R., & Bruna, J. (2022). When does return-conditioned supervised learning work for offline reinforcement learning? In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Eds.), *Advances in neural information processing systems* (pp. 1542–1553, Vol. 35). Curran Associates, Inc.

Chen, L., Lu, K., Rajeswaran, A., Lee, K., Grover, A., Laskin, M., Abbeel, P., Srinivas, A., & Mordatch, I. (2021). Decision Transformer: Reinforcement learning via sequence modeling. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, & J. W. Vaughan (Eds.), *Advances in neural information processing systems* (pp. 15084–15097, Vol. 34). Curran Associates, Inc.

Dearden, A. M., & Demiris, Y. (2005). Learning forward models for robots. *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, 1440–1445.

Diehl, M., & Ramirez-Amaro, K. (2023). A causal-based approach to explain, predict and prevent failures in robotic tasks. *Robotics and Autonomous Systems*, *162*, 104376. https://doi.org/10.1016/j.robot.2023.104376

Dogge, M., Custers, R., & Aarts, H. (2019). Moving forward: On the limits of motor-based forward models. *Trends in Cognitive Sciences*, *23*(9), 743–753. https://doi.org/10.1016/j.tics.2019.06.008

Emmons, S., Eysenbach, B., Kostrikov, I., & Levine, S. (2021). RvS: What is essential for offline RL via supervised learning? https://doi.org/10.48550/ARXIV.2112.10751

Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver; Boyd.

Francis, B. A., & Wonham, W. M. (1976). The internal model principle of control theory. *Automatica*, *12*(5), 457–465. https://doi.org/10.1016/0005-1098(76)90006-6

Furuta, H., Matsuo, Y., & Gu, S. S. (2021). Generalized decision transformer for offline hindsight information matching. https://doi.org/10.48550/ARXIV.2111.10364

Gärdenfors, P., & Lombard, M. (2018). Causal cognition, force dynamics and early hunting technologies. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00087

Gerstenberg, T., & Tenenbaum, J. B. (2017). Intuitive theories. In M. R. Waldmann (Ed.), *The oxford handbook of causal reasoning* (pp. 515–548). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199399550.013.28

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*, 80–89. https://doi.org/10.1109/dsaa.2018.00018

Hellström, T. (2021). The relevance of causation in robotics: A review, categorization, and analysis. *Paladyn, Journal of Behavioral Robotics*, *12*(1), 238–255. https://doi.org/10.1515/pjbr-2021-0017

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, *2*(5), 359–366. https://doi.org/10.1016/0893-6080(89)90020-8

Janner, M., Li, Q., & Levine, S. (2021). Offline reinforcement learning as one big sequence modeling problem. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. S. Liang, & J. W. Vaughan (Eds.), *Advances in neural information processing systems* (pp. 1273–1286, Vol. 34). Curran Associates, Inc.

Kotseruba, I., & Tsotsos, J. K. (2018). 40 years of cognitive architectures: Core cognitive abilities and practical applications. *Artificial Intelligence Review*, *53*(1), 17–94. https://doi.org/10.1007/s10462-018-9646-y

Lake, B. M. (2014). *Towards more human-like concept learning in machines: Compositionality, causality, and learning-to-learn* [PhD thesis]. Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/95856

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2016). Building machines that learn and think like people. *Behavioral and Brain Sciences*, *40*. https://doi.org/10.1017/s0140525x16001837

Lee, T. E., Zhao, J. A., Sawhney, A. S., Girdhar, S., & Kroemer, O. (2021). Causal reasoning in simulation for structure and transfer learning of robot manipulation policies. *2021 IEEE International Conference on Robotics and Automation (ICRA)*. https://doi.org/10.1109/icra48506.2021.9561439

Lee, T. E., Vats, S., Girdhar, S., & Kroemer, O. (2023). SCALE: Causal learning and discovery of robot manipulation skills using simulation. In J. Tan, M. Toussaint, & K. Darvish (Eds.), *Proceedings of the 7th conference on robot learning* (pp. 2229–2256, Vol. 229). PMLR.

Lombard, M., & Gärdenfors, P. (2017). Tracking the evolution of causal cognition in humans. *Journal of Anthropological Sciences*, *95*, 219–234. https://doi.org/10.4436/JASS.95006

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, *30*, 4768–4777.

Mandlekar, A., Xu, D., Wong, J., Nasiriany, S., Wang, C., Kulkarni, R., Fei-Fei, L., Savarese, S., Zhu, Y., & Martín-Martín, R. (2021). What matters in learning from offline human demonstrations for robot manipulation. https://doi.org/10.48550/ARXIV.2108.03298

Matsui, Y., & Matsui, T. (2001). NP-completeness for calculating power indices of weighted majority games. *Theoretical Computer Science*, *263*(1–2), 305–310. https://doi.org/10.1016/s0304-3975(00)00251-6

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1988). The appeal of parallel distributed processing. In *Readings in cognitive science* (pp. 52–72). Elsevier. https://doi.org/10.1016/b978-1-4832-1446-7.50010-8

McClelland, J. L., Rumelhart, D. E., & the PDP Research Group. (1987). *Parallel distributed processing: Explorations in the microstructure of cognition, volume 2: Psychological and biological models: Psychological and biological models* (Vol. 2). The MIT Press.

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*, *9*(8), 1265–1279. https://doi.org/10.1016/s0893-6080(96)00035-4

Minsky, M., & Papert, S. A. (2017). *Perceptrons: An introduction to computational geometry*. The MIT Press. https://doi.org/10.7551/mitpress/11301.001.0001

Nguyen-Tuong, D., & Peters, J. (2011). Model learning for robot control: A survey. *Cognitive Processing*, *12*(4), 319–340. https://doi.org/10.1007/s10339-011-0404-1

Oh, J., Guo, Y., Singh, S., & Lee, H. (2018). Self-imitation learning. In J. Dy & A. Krause (Eds.), *Proceedings of the 35th international conference on machine learning* (pp. 3878–3887, Vol. 80). PMLR.

Pearl, J. (1985). Bayesian networks: A model of self-activated memory for evidential reasoning. *Proceedings of the 7th conference of the Cognitive Science Society, University of California, Irvine, CA, USA*, 15–17.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. https://doi.org/10.1145/2939672.2939778

Rosenblatt, F. (1958). The Perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408. https://doi.org/10.1037/h0042519

Rosenblatt, F., et al. (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms* (Vol. 55). Spartan Books.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Schölkopf, B. (2022). Causality for machine learning. In *Probabilistic and causal inference: The works of judea pearl* (1st ed., pp. 765–804). Association for Computing Machinery. https://doi.org/10.1145/3501714.3501755

Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the theory of games ii* (pp. 307–317). Princeton University Press. https://doi.org/10.1515/9781400881970-018

Shrikumar, A., Greenside, P., & Kundaje, A. (2017). Learning important features through propagating activation differences. *Proceedings of the 34th International Conference on Machine Learning*, *70*, 3145–3153.

Sonar, A., Pacelli, V., & Majumdar, A. (2021). Invariant policy optimization: Towards stronger generalization in reinforcement learning. In A. Jadbabaie, J. Lygeros, G. J. Pappas, A., P. Parrilo, B. Recht, C. J. Tomlin, & M. N. Zeilinger (Eds.), *Proceedings of the 3rd conference on learning for dynamics and control* (pp. 21–33, Vol. 144). PMLR.

Sontakke, S. A., Mehrjou, A., Itti, L., & Schölkopf, B. (2021). Causal curiosity: RL agents discovering self-supervised experiments for causal representation learning. In M. Meila & T. Zhang (Eds.), *Proceedings of the 38th international conference on machine learning* (pp. 9848–9858, Vol. 139). 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. https://doi.org/10.1037/h0055479

Stocking, K. C., Gopnik, A., & Tomlin, C. (2022). From robot learning to robot understanding: Leveraging causal graphical models for robotics. In A. Faust, D. Hsu, & G. Neumann (Eds.), *Proceedings of the 5th conference on robot learning* (pp. 1776–1781, Vol. 164). PMLR.

von Holst, E., & Mittelstaedt, H. (1950). Das Reafferenzprinzip: Wechselwirkungen zwischen Zentralnervensystem und Peripherie. *Naturwissenschaften*, *37*(20), 464–476. https://doi.org/10.1007/bf00622503

Wang, Z., Xiao, X., Xu, Z., Zhu, Y., & Stone, P. (2022). Causal dynamics learning for task-independent state abstraction. In K. Chaudhuri, S. Jegelka, L. Song,

C. Szepesvari, G. Niu, & S. Sabato (Eds.), *Proceedings of the 39th international conference on machine learning* (pp. 23151–23180, Vol. 162). PMLR.

Wen, M., Kuba, J., Lin, R., Zhang, W., Wen, Y., Wang, J., & Yang, Y. (2022). Multi-agent reinforcement learning is a sequence modeling problem. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, & A. Oh (Eds.), *Advances in neural information processing systems* (pp. 16509–16521, Vol. 35). Curran Associates, Inc.

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*(7–8), 1317–1329. https://doi.org/10.1016/s0893-6080(98)00066-5

Wolpert, D. M., & Flanagan, J. R. (2001). Motor prediction. *Current Biology*, *11*(18), R729–R732. https://doi.org/10.1016/s0960-9822(01)00432-8

Zare, M., Kebria, P. M., Khosravi, A., & Nahavandi, S. (2023). A survey of imitation learning: Algorithms, recent developments, and challenges. https://doi.org/10.48550/ARXIV.2309.02473

Zhang, K., Schölkopf, B., Spirtes, P., & Glymour, C. (2017). Learning causality and causality-related learning: Some recent progress. *National Science Review*, *5*(1), 26–29. https://doi.org/10.1093/nsr/nwx137

Zhu, Y., Gao, T., Fan, L., Huang, S., Edmonds, M., Liu, H., Gao, F., Zhang, C., Qi, S., Wu, Y. N., Tenenbaum, J. B., & Zhu, S.-C. (2020). Dark, beyond deep: A paradigm shift to cognitive AI with humanlike common sense. *Engineering*, *6*(3), 310–345. https://doi.org/10.1016/j.eng.2020.01.011