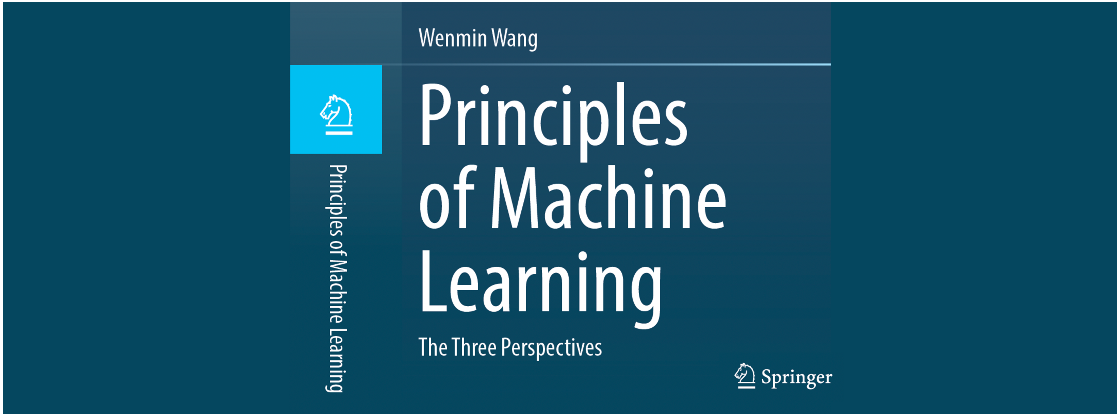


https://link.springer.com/book/10.1007/978-981-97-5333-8

Part II Frameworks

6 Symbolic Framework

6 Symbolic Framework

About the Symbolic Framework

- Symbols: refer to all advanced symbols invented and understood by humans, including signs, numbers, characters, and words.
- Symbolic systems: the systems composed of symbols, logical relationships between symbols, and their processing methods.
  - Various natural languages have their symbolic systems, such as English, Chinese, and Spanish.
  - Many disciplines have their symbolic systems, such as mathematics, physics, chemistry, logic, and computer science.
- Symbolic framework: another theoretical framework of machine learning based on symbolic systems.

Symbolism

- Symbolism is also known as logicism, the school of psychology, or computerism.
- Symbolism is based on the principles of physical symbology and the principle of limited rationality.
- It views human thinking as a combination of symbols, uses symbols and their relationships to represent information and knowledge, and completes human cognition and learning through the storage, extraction, reasoning, and transformation of symbols.
- Symbolists invent specific algorithms to process these symbols, solve problems, and derive new knowledge.

Symbolic Learning Theory

What is Symbolic Learning Theory

"A theory that attempts to explain how imagery works in performance enhancement. It suggests that imagery develops and enhances a coding system that creates a mental blueprint of what has to be done to complete an action."

https://dictionary.apa.org/symbolic-learning-theory

Physical Symbol System

What is Physical Symbol System

"A physical symbol system consists of a set of entities, called symbols, which are physical patterns that can occur as components of another type of entity called an expression (or symbol structure). Thus, a symbol structure is composed of several instances (or tokens) of symbols related in some physical way (such as one token being next to another). ... ... A physical symbol system is a machine that produces through time an evolving collection of symbol structures. Such a system exists in a world of objects wider than just these symbolic expressions themselves."

https://dl.acm.org/doi/pdf/10.1145/360018.360022

GOFAI

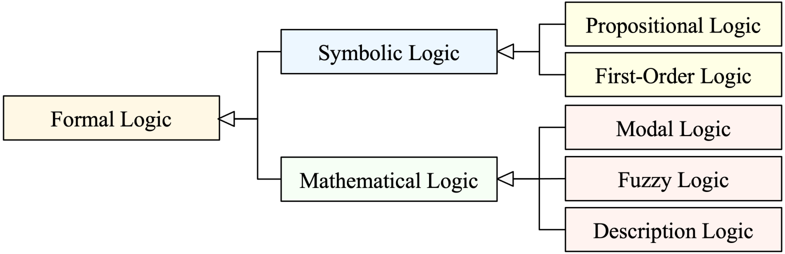
- GOFAI (good old-fashioned artificial intelligence) was appeared in 1985.
- The period was a time of rapid development of computer, which provided good power for statistical learning and multi-layer neural networks.
- It was also from this period that symbolic AI, which do not require high computational power, gradually declined.
- One more reason is that symbolic AI can only handle problems that can be represented as physical symbols, while vision, hearing, and speech are difficult to represent with symbols.

6 Symbolic Framework

Types of Formal Logic

Formal logic abstracts logic into a symbolic language, giving logic certain mathematical properties and allowing mathematical operations.

Formal logic can be divided into symbolic logic and mathematical logic.



Symbolic Logic

Symbolic logic, also known as classical logic.

Be a development of formal logic employing a special symbols.

Be capable of manipulation in accordance with precise rules.

Consisting of *propositional logic* and *first-order logic*.

Category	Symbol	Brief Explanation
Connectives	$\neg$	Logical negation
	$\wedge$	Logical conjunction
	$\vee$	Logical disjunction
	$\Rightarrow$	Implication
	$\Leftrightarrow$	Material equivalence
	$\models$	Entails
	$\not\models$	Not valid
Quantifiers	$\forall$	Universal quantifier
	$\exists$	Existential quantifier
Equality	$=$	Equal to

Propositional Logic

Propositional logic is a formal system for dealing with statements that have a value of true or false, and for constructing rules for proving theorems.

$$
\begin{array}{rcl}
\textit{Sentence} & \rightarrow & \textit{AtomicSentence} \mid \textit{ComplexSentence} \\
\textit{AtomicSentence} & \rightarrow & \textit{True} \mid \textit{False} \mid P \mid Q \mid R \mid \ldots \\
\textit{ComplexSentence} & \rightarrow & < \textit{Sentence} > \mid [\ \textit{Sentence}\ ] \\
& \mid & \neg\ \textit{Sentence} \\
& \mid & \textit{Sentence} \land \textit{Sentence} \\
& \mid & \textit{Sentence} \lor \textit{Sentence} \\
& \mid & \textit{Sentence} \Rightarrow \textit{Sentence} \\
& \mid & \textit{Sentence} \Leftrightarrow \textit{Sentence} \\
\textit{OPERATOR PRECEDENCE} & : & \neg, \land, \lor, \Rightarrow, \Leftrightarrow
\end{array}
$$

First-Order Logic

First-order logic is a formal system built on top of propositional logic.

$$
\begin{array}{rcl}
\mathit{Sentence} & \rightarrow & \mathit{AtomicSentence} \mid \mathit{ComplexSentence} \\
\mathit{AtomicSentence} & \rightarrow & \mathit{Predicate} \mid \mathit{Predicate}(\mathit{Term}, \ldots) \mid \mathit{Term} = \mathit{Term} \\
\mathit{ComplexSentence} & \rightarrow & < \mathit{Sentence} > \mid [\ \mathit{Sentence}\ ] \mid \neg\ \mathit{Sentence} \\
 & \mid & \mathit{Sentence} \land \mathit{Sentence} \mid \mathit{Sentence} \lor \mathit{Sentence} \\
 & \mid & \mathit{Sentence} \Rightarrow \mathit{Sentence} \mid \mathit{Sentence} \Leftrightarrow \mathit{Sentence} \\
 & \mid & \mathit{Quantifier}\ \mathit{Variable}, \ldots\ \mathit{Sentence} \\
\mathit{Term} & \rightarrow & \mathit{Function}(\mathit{Term}, \ldots) \mid \mathit{Constant} \mid \mathit{Variable} \\
\mathit{Quantifier} & \rightarrow & \forall \mid \exists \\
\mathit{Constant} & \rightarrow & A \mid X_1 \mid \mathit{John} \mid \ldots \\
\mathit{Variable} & \rightarrow & a \mid x \mid s \mid \ldots \\
\mathit{Predicate} & \rightarrow & \mathit{True} \mid \mathit{False} \mid \mathit{After} \mid \mathit{Loves} \mid \mathit{Raining} \mid \ldots \\
\mathit{Function} & \rightarrow & \mathit{Mother} \mid \mathit{LeftLeg} \mid \ldots \\
\mathit{OPERATOR\ PRECEDENCE} & : & \neg, =, \land, \lor, \Rightarrow, \Leftrightarrow
\end{array}
$$

First-Order Logic

First-order logic can use quantified variables of non-logical objects and statements containing variables.

$$
\begin{aligned}
&\mathit{Mary} \in \mathit{FemalePersons}\\
&\mathit{John} \in \mathit{MalePersons}\\
&\mathit{SisterOf}(\mathit{Mary}, \mathit{John})\\
&\mathit{FemalePersons} \subset \mathit{Persons}\\
&\mathit{MalePersons} \subset \mathit{Persons}\\
&\forall x,\, x \in \mathit{Persons} \;\Rightarrow\; [\forall y,\, \mathit{HasMother}(x, y) \;\Rightarrow\; y \in \mathit{FemalePersons}]
\end{aligned}
$$

Examples of first-order predicate logic expressions.

Propositional Logic vs. First-Order Logic

Formal Language	Logic Scope	Logical Value
Propositional Logic	Facts	True/False/Unknown
First-Order Logic	Facts, objects, relations	True/False/Unknown

- Propositional logic is also known as zeroth-order logic, a term that facilitates comparison with first-order logic.

  It uses connectives only.
- First-order logic is also known as predicate logic, or first-order predicate calculus.

  It uses not only connectives, but also quantifiers, equality, and predicates.

Modal Logic

Modal Logic can be defined as $\mathcal{M} = \langle W, R, v \rangle$, where W is a set of possible worlds, R is a binary relation called accessibility on W, and v is a valuation function to assign a truth value to each pair of an atomic formula and a world.

$$\textit{if } v(w, P) \textit{ then } w \vDash P$$

$$w \vDash \neg P \textit{ if and only if } w \nvDash P$$

$$w \vDash (P \land Q) \textit{ if and only if } w \vDash P \textit{ and } w \vDash Q$$

$$w \vDash \Box P \textit{ if and only if for every element } u \textit{ of } W, \textit{ if } wRu \textit{ then } u \vDash P$$

$$u \vDash \Diamond P \textit{ if and only if for some element } u \textit{ of } W, \textit{ it holds that } wRu \textit{ and } u \vDash P$$

$$\vDash P \textit{ if and only if } w* \vDash P$$

Combined modal logic with neural networks, connectionist modal logic was proposed.

Fuzzy Logic

Fuzzy logic is a form of many-valued logic that is formed by assigning truth values to propositions.

Its truth value is [0, 1], where 0 represents “totally false”, 1 represents “totally true”, otherwise represents “partially true”.

Fuzzy logic are mathematical methods for representing fuzzy and imprecise information, to recognize, characterize, operate, interpret, and utilize.

Fuzzy logic and fuzzy sets have been widely used in machine learning, known as fuzzy machine learning, such as:

fuzzy clustering, fuzzy neural networks, and fuzzy deep neural networks.

Description Logic

Description logic is a set of formalized knowledge representation languages.

It is a decidable fragment of first-order logic, and a subset of first-order logic.

Description Logic	First-Order Logic
Individual	Constant
Concept	Unary predicate
Role	Binary predicate

Symbol	Description	Example
$\top$	Universal concept	$\top$
$\bot$	Empty concept	$\bot$
$\sqcap$	Concept conjunction	$C \sqcap D$
$\sqcup$	Concept disjunction	$C \sqcup D$
$\neg$	Concept Negation	$\neg C$
$\forall$	Universal Quantifier	$\forall R, C$
$\exists$	Existential Quantifier	$\exists R, C$
$\sqsubseteq$	Concept Inclusion	$C \sqsubseteq D$
$\equiv$	Concept Equivalence	$C \equiv D$
$\doteq$	Concept Definition	$C \doteq D$
$:$	Concept Assertion	$a : C$
$:$	Role Assertion	$(a, b) : R$

Description Logic

$$
\begin{aligned}
Coure &\sqsubseteq \forall enrolls.Student \sqcap \geq 5\ entrolls \sqcap \leq 50\ enrolls \sqcap \\
&\quad \forall taughtby.(Professor \sqcup GradStudent) \sqcap 1\ taughtby \\
AdvCourse &\sqsubseteq Course \sqcap \forall enrolls.(GradStudent \sqcup \neg UndStudent) \sqcap \leq 20\ enrolls \\
BasCourse &\sqsubseteq Course \sqcap \forall taughtby.Professor \\
GradStudent &\sqsubseteq Student \sqcap \forall degree.String \sqcap \geq 1\ dgree \\
UndStudent &\sqsubseteq Student
\end{aligned}
$$

Description logic is used in AI to describe and interpret concepts known as terminological knowledge.

It provides a logical formalization method for ontology and the semantic web.

Some ones are in research combining description logic with neural symbols.

6 Symbolic Framework

Logical Learning Theory

- Logical learning theory was formally proposed by Joseph Rychlak in 1981.
- Core concepts are logical form and its argumentation. The former refers to manifestation of logic, and the later is carried out by logical reasoning.
- Human thinking can be divided into concrete and abstract thinking.
- Logical learning is most suitable for abstract thinking, but machine learning belong to concrete thinking.
- The view that the cognitive process is a logical process is worth pondering and exploring for ones engaged in AI, especially ML.

Logical Learning Modes

- Deductive Reasoning — Deductive Learning
- Inductive Reasoning — Inductive Learning
- Abductive Reasoning — Abductive Learning

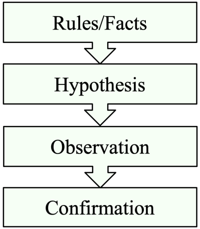
Deductive Learning

Deductive learning originates from deductive reasoning.

Deductive reasoning is to draw individual conclusion from known general premise, and deductive learning is to propose hypothesis from facts/rules, and confirm it by observation.

Three types of deductive reasoning.

	Modus ponens	Modus tollens	Syllogism
Major premise	$P \rightarrow Q$	$P \rightarrow Q$	$P \rightarrow Q$
Minor premise	P	$\neg Q$	$Q \rightarrow R$
Conclusion	Q	$\neg P$	$P \rightarrow R$

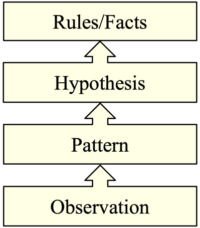


Example: All humans are mortal; Socrates is human; therefore Socrates will die.

Inductive Learning

Inductive learning originates from inductive reasoning.

Inductive reasoning is from individual instance to derive general conclusion, and inductive learning is from observation, to form hypothesis and use it to learn general facts/rules.



Two major types of inductive reasoning.

	Premise	Conclusion
Generalization	Proportion Q of the sample has attribute a.	So, the proportion Q of the population has attribute a.
Prediction	Proportion Q of observed members of group G have had attribute a.	So, there is a probability corresponding to Q that other members of group G will have attribute a when next observed.

Example: All swans we have seen are white, therefore all swans are white.

Abductive Learning

Abductive learning originates from abductive reasoning.

Abductive reasoning starts from observed facts and seeks their most likely premises, and abductive learning is to start from observed facts, seek optimal hypotheses, and generate new knowledge.

Abductive reasoning.

Abduction	Reasoning Rule
Fact	Q
Hypothesis	$P \to Q$
Premise	P

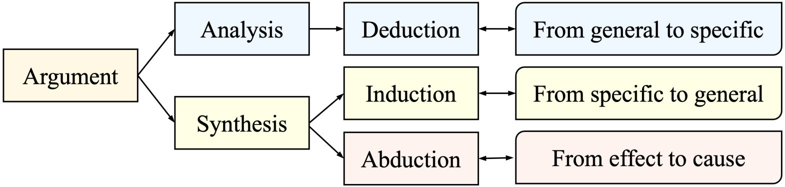
- The surprising fact, C, is observed.
- But if A were true, C would be a matter of course.
- Hence, there is reason to suspect that A is true.

Example: The theory of biological evolution, proposed by Darwin.

Deductive vs. Inductive vs. Abductive

Intrinsic connection between argumentation methods and reasoning patterns.

Argumentation is a series of arguments, presented as argument outlines, used to provide reasons for them to accept its conclusion.



The intrinsic connection between argumentation and reasoning modes.

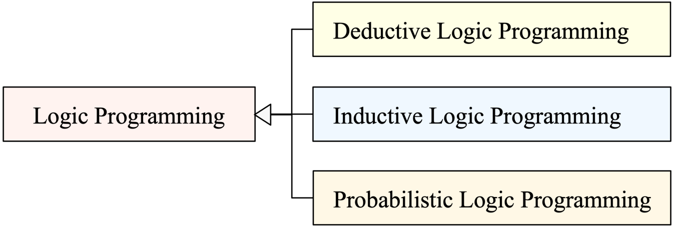
Deductive vs. Inductive vs. Abductive

Deduction	Rule	All beans in this bag are white	Specific conclusion (always true)
	Case	These beans are from this bag	
	Result	These beans are white	
Induction	Case	These beans are from this bag	General conclusion (maybe true)
	Result	These beans are white	
	Rule	All beans in this bag are white	
Abduction	Rule	All beans in this bag are white	Best explanation (maybe true)
	Result	These beans are white	
	Case	These beans are from this bag	

Examples and characteristics of the three major reasoning modes.

Logic Programming Methods

Logic programming (LP) is a programming paradigm based on formal logic.



LP $=$ Logic $+$ Computational Control.

Deductive Logic Programming

The representative of deductive logic programming (DLP) is Prolog (programming in logic), the program is a set of facts and rules.

A rule is represented as $H :- B_1, \ldots, B_n$, means $H \Leftarrow B_1 \wedge \cdots \wedge B_n$.

```
likes(bill, car).
animal(X) :- cat(X).
bird(X)   :- animal(X), has(X, feather).
```

Its reasoning adopts forward and backward chaining, initiated by “query”.

$$\text{DLP} = \text{Logic} + \text{Computational Control} + \text{Facts and Rules}.$$

Inductive Logic Programming

Inductive logic programming (ILP) is a supervised method that uses logic programming to represent background knowledge, samples, and hypotheses.

ILP can include additional information in the learning problem.

Proposition	Expression
Necessity	$B \not\models E^+$
Sufficiency	$B \wedge h \models E^+$
Weak consistency	$B \wedge h \not\models$ false
Strong consistency	$B \wedge h \wedge E^- \not\models$ false

Let B be background knowledge and serve as a logical inference, E^+ and E^- be positive and negative examples. A correct hypothesis h is a logical proposition in the table.

ILP $=$ Logic $+$ Computational Control $+$ Statistics.

Probabilistic Logic Programming

Probabilistic logic programming (PLP) combines programming with probability.

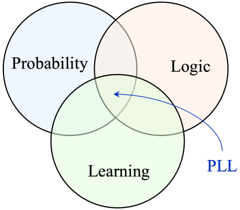
$$\text{PLP} = \text{Logic} + \text{Computational Control} + \text{Probability}$$

ProbLog (Probabilistic Prolog) is formed by extending probability on Prolog, in which a fact is defined as:

$$F = \{p_1 :: f_1, p_2 :: f_2, \ldots, p_n :: f_n\}.$$

For a fact set $F' \subseteq F$, its probability distribution is:

$$P_F(F') = \prod_{f_i \in F'} p_i \prod_{f_i \in F \setminus F'} (1 - p_i).$$



PLL: Probabilistic logic learning

6 Symbolic Framework

About the Rule Learning

Rule learning

- Also known as rule-based learning, or rule-based machine learning.
- Including association rule learning, decision trees, and random forests.

Association Rule Learning

- Also known as association analysis, used to discover strong rules between variables in large transaction datasets.
- New rules will be generated as the analysis increases.
- Its goal is to simulate the feature extraction of human brain, and improve the ability to abstract its associations from new unclassified data.

Association Rule Learning

Definition: (Association Rule)

Association rule (AR) can be defined as a 3-tuple, AR $= \langle I, T, R \rangle$, where $I = \{i_1, i_2, \ldots, i_m\}$ is an itemset, $\forall i_i \in \{1, 0\}$; $T = \{t_1, t_2, \ldots, t_n\}$ is a transaction dataset, $\forall t_i \subseteq I$; $R : X \Rightarrow Y$ is a rule, $X, Y \subseteq I$, and $X \cap Y = \emptyset$.

Trans	butter	bread	egg	milk	beer	diaper
1	0	1	1	1	0	0
2	1	0	1	1	0	0
3	1	1	0	1	0	0
4	0	0	0	0	1	1

A supermarket's database T contains 6 items and 4 transactions.

Instance rules derived from the T:

$\{$*butter, egg, bread*$\} \Rightarrow \{$*milk*$\}$.
$\{$*diaper*$\} \Rightarrow \{$*beer*$\}$.

Association Rule Discovery

It is to select useful rules from all possible association rules.

Support: the support of antecedent and the support of association rule.

$$\text{supp}\left(X\right) = \frac{\left|t \in T; X \subseteq t\right|}{\left|T\right|} = P\left(X\right).$$

$$\text{supp}\left(X \Rightarrow Y\right) = \text{supp}\left(X \cup Y\right) = P\left(X \cap Y\right).$$

Confidence: only measures for association rules.

$$\text{conf}\left(X \Rightarrow Y\right) = \frac{\text{supp}\left(X \Rightarrow Y\right)}{\text{supp}\left(X\right)} = \frac{\text{supp}\left(X \cup Y\right)}{\text{supp}\left(X\right)} = \frac{P\left(X \cap Y\right)}{P\left(X\right)} = P\left(Y|X\right).$$

Association Rule Synthesis

The rule synthesis is usually decomposed into the following two subtasks:

- Frequent itemset synthesis, that is, discover frequent itemset from many transactions in the transaction dataset.
- Strong rules synthesis, extract strong rules based on the discovered frequent itemset.

The computational cost required to synthesize frequent itemset is much greater than the computational overhead required to synthesize strong rules.

Apriori algorithm is a famous method reducing the amount of computation:

https://en.wikipedia.org/wiki/Apriori_algorithm

6 Symbolic Framework

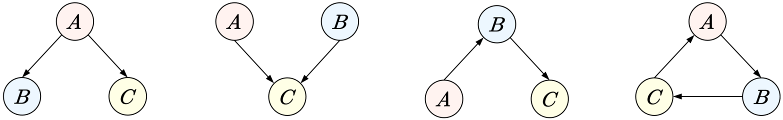
About the Causal Learning

There are several terms related to causal learning: causality, cause and effect, and causal relation.

> "Causality refers to causal relations, i.e. the relations between causes and effects."

We will discuss on causal relation, causal inference, and causal learning that are closely related to machine learning.

Causal Relations



(a) (b) (c) (d)

(a) Common-cause: cause A leads to effects B and C respectively.

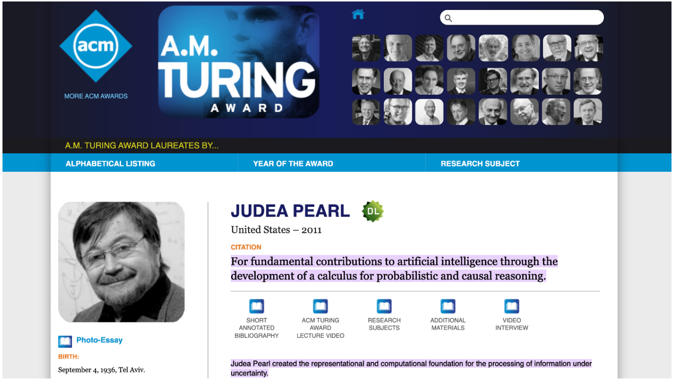
(b) Common-effect: causes A and B both lead to the same effect C.

(c) Causal chains: cause A leads to effect B, and B in turn causes effect C.

(d) Causal homeostasis: A, B, and C are interrelated.

Causal Inference

Judea Pearl is a pioneer and advocate of causal inference.



Judea Pearl: Turing Award for Probabilistic and Causal Reasoning in 2011.

Causal Inference

Level		Activity	Typical Questions	Examples
1	Association $P(y\|x)$	Seeing	• What is? • How would seeing X change my belief in Y?	• What does a symptom tell me about a disease? • What does a survey tell us about the election results?
2	Intervention $P(y\|do(x), z)$	Doing, Intervening	• What if? • What if I do X?	• What if I take aspirin, will my headache be cured? • What if we ban cigarettes?
3	Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	• Why? • Was it X that caused Y? • What if I had acted differently?	• Was it the aspirin that stopped my headache? • Would Kennedy be alive had Oswald not shot him? • What if I had not been smoking the past two years?

Questions at level i ($i = 1, 2, 3$) can be answered only if level j ($j > i$) is available.

Levels of Causal Models

Model	Predict in i.i.d. setting	Predict under distr. shift/intervention	Answer counter-factual questions	Obtain physical insight	Learn from data
Mechanical/physical	yes	yes	yes	yes	?
Structural causal	yes	yes	yes	?	?
Causal graphical	yes	yes	no	?	?
Statistical	yes	no	no	no	yes

- Mechanistic/physical model: providing a comprehensive description of system.
- Structural causal model: abstracted from physical realism, supporting first 3 tasks.
- Causal graphical model: a graphical model, supporting first 2 tasks.
- Statistical model: can predict in i.i.d. environment, and learn from data.

6 Symbolic Framework

About the Neuro-Symbolic Learning

In recent years, hybrid machine learning has attracted the attention of some researchers and become a new research approach.

The neuro-symbolic learning combines neural networks and symbolic methods.

With the rapid development of AI and ML, some very complex tasks, such as Chess, Go, Texas Hold'em, medical diagnosis, and protein structure prediction, artificial intelligence has surpassed human intelligence.

However, there are some simple tasks that even toddlers, chicks, or cubs can complete, but AI is hard to handle.

Moravec's Paradox

The paradox was presented by Hans Moravec in 1988.

"It is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one year old when it comes to perception and mobility."

Moravec thought, the perception and action skills are all biological instincts, which have evolved in the process of natural selection. The older the skill, the longer the evolution time. The evolution time of human abstract thinking is relatively short.

The View on Moravec's Paradox

Human advanced cognitive activities are complex processes, which are often formalized and symbolized for ease of communication and dissemination. So, they are easy for computers to represent and reason.

Chess, Go, and poker are created by humans, and same do AI. So, AI can defeat human intelligence in those games.

However, the perception and action skills rely on intuition and instinct. Intuition is a direct feeling without conscious reasoning; while instinct is the inherent inclination towards a particular complex behaviour.

The feature of intuition and instinct is difficult to formalize and symbolize, so it is not surprising that the effective methods using AI to process intuition and instinct has not yet been found.

Thinking Fast and Slow: System 1 and System 2

System 1 (Fast Thinking)	System 2 (Slow Thinking)
Relying on intuition, memory, and habits, and making judgments or reactions quickly.	Requiring focused attention, deep rational and logical thinking, and making decisions after careful consideration.
Be often unconscious, fast but not pursuing accuracy, but sometimes it is influenced by the rule of thumb that seeing is believing, and the illusions it produces often lead to wrong choices.	Be conscious, relying on long-term learned knowledge and accumulated experience, analyzing and solving problems through attention and logical thinking, and making complex decisions. It is not prone to errors, but it will be influenced by the intuitive judgment of system 1.

Neuro-Symbolic Approaches

The two theoretical frameworks of machine learning:

- The connectionist framework represented by neural networks.
- The symbolic framework with logical reasoning capabilities.

Strength and Weakness of the two frameworks:

- Neural networks: very effective for vision, speech, and natural language, but hard to handle logical reasoning problems.
- Symbolic learning: powerful in logical reasoning, but clumsy in handling vision, and speech.

Neural-symbolic approaches = Neural networks + Symbolic learning.

Thank You