Wenmin Wang

Principles of Machine Learning

The Three Perspectives

Springer

Principles of Machine Learning

Part II Frameworks

7 Behavioral Framework

Contents

About the Behavioral Framework

- The behavioral framework is also one of theoretical frameworks for machine learning, especially learning from the environment.
- The framework rooted in behaviorism doctrine and behavioral psychology.
- Behavioral learning theory, including respondent conditioning and operant conditioning, proposed the concept of reinforcement learning.
- Two types of decision models: Bayesian decision models and Markov decision models.
- Three types of behavior decisions: single-stage decision, multi-stage decision, as well as sequential decision.
- The sequential decision is made during interactive learning with the environment.

Behaviorism Doctrine

Behaviorism doctrine is a systematic theory for studying the behavior of intelligent agents, focusing on observable, quantifiable behavior, and then conducting psychological research and analysis.

The main point is that psychology should study observable, measurable psychological behavior, and verify and explain it through experiments.

The intelligent activities and behavioral characteristics of humans are formed in the process of self-adaptation and self-optimization in a given environment, learning through interaction with the rewards and punishments of environmental feedback.

Behaviorism views the behavior of intelligent agents as the external manifestation of their intelligent activities. If a machine's behavior appears to be intelligent, it should be considered an intelligent machine.

Behavioral Psychology

From 1920s to mid-1950s, behavioral psychology gradually became mainstream of psychology. The popularity lies in establishing psychology as an objective, measurable science.

The three representative branches of behavioral psychology:

- Methodological behaviorism: Only the dynamic behavior of agents can be objectively observed. Although thoughts are still recognized as existing, but not considered part of behavioral science.
- Psychological behaviorism: Humans not only follow the learning principles of animals but also the unique learning principles of humans.
- Purposive behaviorism: It combines objective research on behavior and also considers the purpose of behavior.

Behavioral Learning

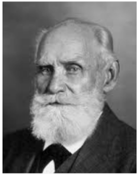
Behavioral learning refers to all behaviors being learned through interaction with the environment.

Learning is the result of responding to external events, and innate factors have little influence on behavioral learning.

The study of behavioral learning focuses on the observable external behavior of intelligent agents, because observable external behavior can be objectively measured and studied systematically.

The view of behavioral learning believes that the behavioral learning has no essential difference between humans and animals. Therefore, research and experiments on behavioral learning can be conducted not only on humans but also on animals.

Respondent Conditioning



Respondent conditioning, also known as classical conditioning, originated from the conditioned response (CR) proposed by Russian physiologist Ivan Pavlov in 1897.

While studying the eating process of dogs, he noticed that some dogs would salivate before the food was served.

He found through experiments that the bell ringing before the food was served was the cause of the dog's salivation, because at this time the dog could not smell any food.

This behavior can be considered a phenomenon, indicating that the dog has learned that food will be delivered with the bell.

This is a process where a psychological response (hearing the bell) leads to a physiological response (salivation).

The Law of Effect

Edward Thorndike was an American psychologist.

He created a box, known as Thorndike's puzzle box, to observe the behavior of cats in 1898. The cats could escape from the box through simple actions, such as pulling a rope or pushing a rod.

In 1905, Thorndike proposed the law of effect as below:

> "Responses that produce a satisfying effect in a particular situation become more likely to occur again in that situation, and responses that produce a dissatisfying effect become less likely to occur again in that situation."

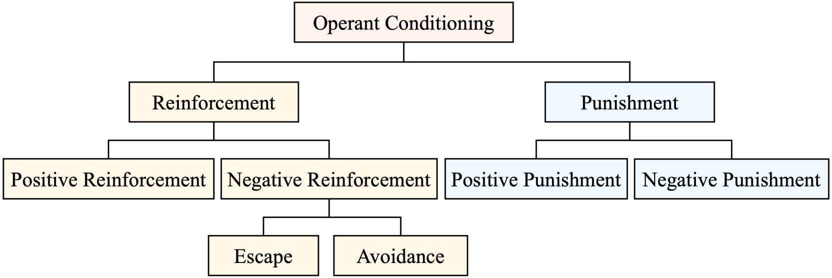
The two terms "satisfying" and "dissatisfying" later evolved into "reinforcement" and "punishment", leading to the birth of the theory of operant conditioning.

Operant Conditioning

Operant conditioning was proposed by Burrhus Skinner, America psychologist and behaviorist.

He made an operant conditioning chamber (Skinner box) in 1930.

Skinner box was an enclosed apparatus containing a bar that an animal subject can manipulate in order to obtain reinforcement.



Operant Conditioning

Reinforcement	*Increase behavior*	Positive Reinforcement		*Add appetitive stimulus following correct behavior*
		Negative Reinforcement	Escape	*Remove noxious stimulus following correct behavior*
			Avoidance	*Behavior avoids noxious stimulus*
Punishment	*Decrease behavior*	Positive Punishment		*Add noxious stimulus following behavior*
		Negative Punishment		*Remove appetitive stimulus following behavior*

The experimental results of operant conditioning showed:
some experiments increase the likelihood of behavior, i.e., *reinforcement*,
else decrease the likelihood of behavior, i.e., *punishment*.

Reinforcement Schedules

In operant conditioning, the timing and frequency of reinforcement for a certain behavior can have a significant impact on the intensity and ratio of its response.

Four types of intermittent reinforcement schedules.

Schedule	Description	Result
Fixed Interval	Reinforcement is delivered at predictable time intervals (e.g., after 5, 10, 15, 20 minutes)	Moderate response rate with significant pause
Fixed Ratio	Reinforcement is delivered at predictable response numbers (e.g., after 2, 4, 6, 8 responses)	High response rate with pause
Variable Interval	Reinforcement is delivered at unpredictable time intervals (e.g., after 5, 7, 10, 20 minutes)	Moderate yet steady response rate
Variable Ratio	Reinforcement is delivered at unpredictable response numbers (e.g., after 1, 4, 5, 9 responses)	High and steady response rate

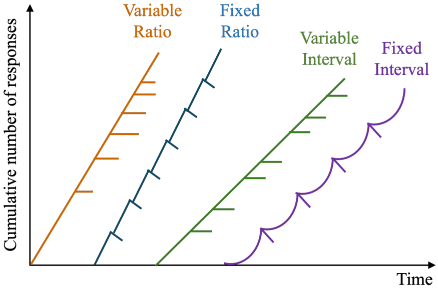
Reinforcement Schedules

(1) Variable ratio schedule:
be unpredictable and produces a high and stable response rate, with almost no pause after reinforcement.

(2) Fixed ratio schedule:
be predictable and produces a high response rate, with a pause after reinforcement.

(3) Variable interval schedule:
be unpredictable and produces a moderate, stable response rate.

(4) Fixed interval schedule:
producing a fan-shaped response pattern, reflecting a significant pause after reinforcement.



Intermittent reinforcement schedules and their response patterns.

About the Behavioral Decision Theory

- Behavioral decision theory examines the psychological processes of judgment, decision, and behavior of intelligent agents.
- The behavioral decision of an intelligent agent begins with the observation of the environment, makes judgments and decisions based on the feedback from the environment, and then takes appropriate actions.
- This section first introduces decision theory, and then discusses the theory of behavioral decision on this basis.

Decision Theory

- Decision is the action and process of making a choice among several possible options. In the field of psychology, decision is considered a cognitive process.
- Decision theory studies how to make the best or nearly optimal decisions, which combines psychology, statistics, philosophy, and mathematics to analyze the decision process.
- Decision theory provides a useful theoretical framework for assessing the merits of a specific action plan. There are two types: decisions under certainty and decisions under uncertainty.
- Utility theory is the root of decision theory, used to analyze the utility of agent's behavior in a given state. Expected utility is the average value of all possible outcomes of the behavior.

Decision Theory

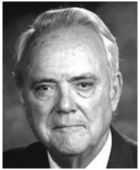
There are three subtypes in decision theory:

(1) Normative decision:
Be a theory about how to make rational and optimal decisions under given hypothesis and constraints. It focuses on the identification of optimal decisions. By designing an ideal decision agent, it can perform calculations perfectly and accurately, and is completely rational in some sense.

(2) Descriptive decision:
Be a theory about how intelligent agents actually make decisions. It focuses on describing the behaviors of decision-makers and usually assumes that these behaviors are generated under certain consistency rules.

(3) Prescriptive decision:
It provides guidelines for decision, so that makes the best possible decisions under uncertain conditions.

Behavioral Decision Theory



"Behavioral decision theory" was published in *Annual Review of Psychology* by Ward Edwards, American psychologist, in 1961.

Research on behavioral decision shows that the decision process of intelligent agents depends on the environment and continuously adjusts its behavior according to the feedback of environment.

Behavioral decision theory is built on decision theory and behavioral science, used to clarify psychological activities related to decision behavior.

Behavioral decision theory is a descriptive decision theory about the judgment, decision, and behavior of intelligent agents.

It integrates the theories of normative decision, descriptive decision, and prescriptive decision, helping decision-makers make better behavioral decisions.

St. Petersburg Paradox

The St. Petersburg paradox was proposed by Swiss mathematician Daniel Bernoulli.

A casino offers a coin-tossing game for a player. Initially, the casino puts 2 dollars of betting money into the pot. Each time the coin shows head the casino will double the bet, when it shows tail the game ends and the player wins all the money in the pot.

Q1: what is the minimum admission fee to ensure the casino does not lose money?

The player has the probability $\frac{1}{2^k}$ to win 2^k dollars for k^{th} game:

$$\mathbb{E}\left[\cdot\right] = \sum_{k=1}^{\infty} \frac{1}{2^k} \cdot 2^k = \sum_{k=1}^{\infty} 1 = \infty.$$

Q2: is any player willing to pay the infinite admission fees to participate in this game?

Decision Process

Definition: (Decision Process)

The decision process (DP) of an intelligent agent can be formalized as a 3-tuple $\mathrm{DP} = \langle S, A, P \rangle$. Where: S and A are both random variables, representing the set of states and the set of actions of the decision process, respectively, and $P(s)$ denotes the probability of the state $s \in S$.

Definition: (Utility)

Given a decision process $\mathrm{DP} = \langle S, A, P \rangle$, the current state $s \in S$, the action to be taken $a \in A$, then its utility U is a function, denoted as $U(s, a)$, its result is a value in the set of real number $\mathbb{R}$, that is $U : S \times A \to \mathbb{R}$.

Expected Utility

Definition: (Expected Utility)

Given a decision process $\mathrm{DP} = \langle S, A, P \rangle$ and utility $U(s, a)$, then its expected utility (EU) is:

$$\mathrm{EU}(S, A) = \mathbb{E}[S, A] = \sum_{s \in S, a \in A} P(s)\, U(s, a)\,.$$

Where, P is the probability of state s, $P(s) \geq 0$ and $\sum P(s) = 1$.

Definition: (Maximum Expected Utility)

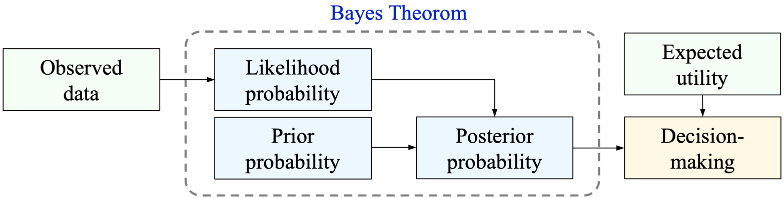
The maximum expected utility (MEU) of the decision process is:

$$\mathrm{MEU}(S, A) = \arg\max_{a \in A} \mathrm{EU}(S, A)\,.$$

Bayesian Decision Theory

Bayesian decision theory (BDT) refers to the decision method based on Bayesian theorem and expected utility.

Specifically: for uncertain observation data, derive its likelihood, calculate its posterior probability based on the likelihood and prior probability, then calculate its expected utility, and make the best decision according to the principle of maximum expected utility.



Bayesian Decision Theory

Given a decision process $\mathrm{DP} = \langle S, A, P \rangle$, where the state set is $S = \{s_1, s_2, \ldots, s_m\}$, the action set is $A = \{a_1, a_2, \ldots, a_n\}$, and probability $P = \{P(s_1), P(s_2), \ldots, P(s_m)\}$, satisfying $P(s_i) \geq 0$, and $\sum_{i=1}^{m} P(s_i) = 1$.

$$P(s_i|a_j) = \frac{P(a_j|s_i)P(s_i)}{P(a_j)} = \frac{P(a_j|s_i)P(s_i)}{\sum_{s_i} P(a_j, s_i)} = \frac{P(a_j|s_i)P(s_i)}{\sum_{s_i} P(a_j|s_i)P(s_i)}.$$

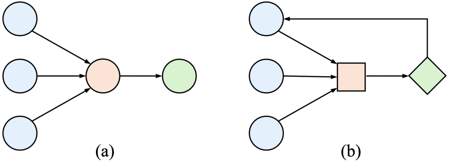
The expected utility of Bayesian decision (BD) is:

$$\begin{aligned}
\mathrm{EU}_{\mathrm{BD}}(S, A) = \mathbb{E}[S, A] &= \sum_{s_i \in S, a_j \in A} P(s_i) P(s_i|a_j) \\
&= \sum_{s_i \in S, a_j \in A} P(s_i) \frac{P(a_j|s_i)P(s_i)}{\sum_{s_i} P(a_j|s_i)P(s_i)}.
\end{aligned}$$

Bayesian Decision Networks

A Bayesian decision network (BDN) is a 4-tuple, $\mathrm{BDN} = \langle \mathrm{BDG}, P_{\mathrm{BDN}}, R, \mathbb{E} \rangle$, where:

- BDG is a directed acyclic Bayesian decision graph, $\mathrm{BDG} = \langle \widetilde{X}, \overrightarrow{E} \rangle$, where: $\widetilde{X} = X \cup D \cup U$, X is the set of random variable nodes, D is the set of decision variable nodes, U is the set of utility variable nodes, and $X \cap D \cap U = \emptyset$; $\overrightarrow{E}$ is a directed edge. BDG satisfies the condition of a directed acyclic graph, $\forall X_i \in \widetilde{X}, \nexists\, (X_i \to X_i)$.
- P_{BDN} is the set of probability distributions of directed graphs, $P_{\mathrm{BDN}} = \{P_{X_i} | X_i \in X\}$.
- R is the set of reward values.
- $\mathbb{E}$ is the expected utility.



(a) Bayesian network, (b) Bayesian decision network.

Markov Decision Process

Definition: (Markov Decision Process)

A Markov decision process (MDP) is a decision process that satisfies the Markov property and can be represented as a 4-tuple, $\text{MDP} = \langle S, A, P, R \rangle$. Where S is a state set, A is an action set, P is a transition probability, and R is a reward function.

The transition probability P can be expressed as:

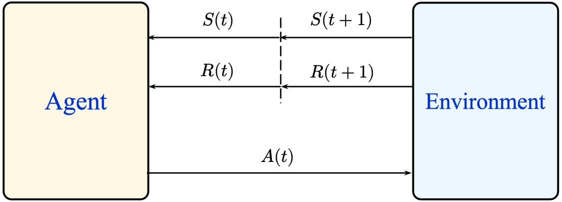
$$P\left(s_{t+1} | s_t, a_t\right) = \mathbb{P}\left[S\left(t+1\right) = s_{t+1} | S\left(t\right) = s_t, A\left(t\right) = a_t\right].$$

The reward function $R : S \times A \to R \in \mathbb{R}$. And given $r_{t+1} \in \mathbb{R}$, then:

$$\begin{aligned} r_{t+1} &= R\left(s_t,\ a_t\right) \\ &= \mathbb{E}\left[R\left(t+1\right) = r_{t+1} \mid S\left(t\right) = s_t,\ A\left(t\right) = a_t,\ S(t+1) = s_{t+1}\right]. \end{aligned}$$

Markov Decision Process

The following is a schematic diagram that illustrated the state, reward, and action between the agent and the environment in the Markov decision process.



The Markov decision process is a decision theory framework for sequential decision in discrete-time stochastic control.

It is also the theoretical framework for the reinforcement learning paradigm.

Partially Observable Markov Decision Process

Definition: (Partially Observable Markov Decision Process)

A partially observable Markov decision process (POMDP) can be represented as a 5-tuple, namely: POMDP $= \langle S,\ A, P,\ R, P_o \rangle$, where: P_o is the conditional observation function, and the other elements in the 5-tuple have the same definitions as in the Markov decision process.

Where the conditional observation function P_o can be expressed as:

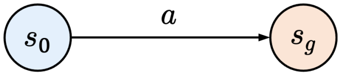
$$P_o\left(o_{t+1}|a_t,\ s_{t+1}\right) = \mathbb{P}\left[O\left(t+1\right) = o_{t+1} \mid A\left(t\right) = a_t, S\left(t+1\right) = s_{t+1}\right].$$

P_o satisfies $\sum_{o \in O} P_o\left(o'|a,\ s'\right) = 1$. At each time period, $P\left(s_{t+1}|s_t, a_t\right)$ causes changes in the environment state, and also through $P_o\left(o_{t+1}|a_t,\ s_{t+1}\right)$ obtains an observation of the environment o_{t+1}.

Single-Stage Decision

Definition: (Single-Stage Decision)

Given a decision process $\mathrm{DP} = \langle S, A, P \rangle$, let the initial state be $s_0 \in S$, and the target state be $s_g \in S$. Then the decision process is called single-stage decision, if taking an action $a \in A$, the initial state s_0 is transformed into the target state s_g with only one stage, that is $P(s_g | s_0, a)$.

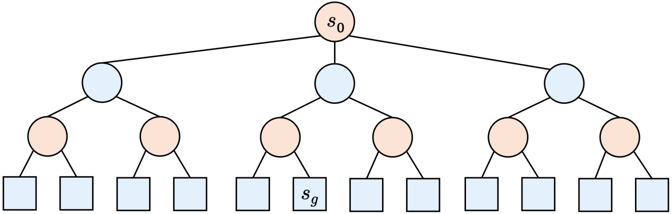


The decision process of a single-stage decision has only one step, also known as a one-off decision.

Multi-Stage Decision

Definition: (Multi-Stage Decision)

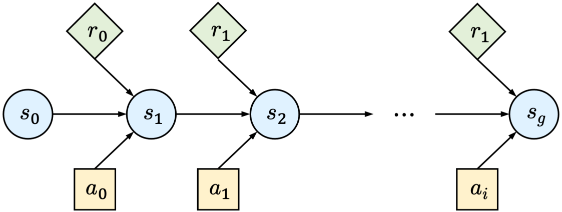
Given a decision process $\mathrm{DP} = \langle S, A, P \rangle$, initial state $s_0 \in S$, the action $a_i \in A$, and the target state $S_g \subset S$. Then the decision process is called multi-stage decision, if its action space can be formed from the initial state to the target state, after several state transitions $P(s_{i+1} | s_i, a_i)$ can reach any target state from s_0 to $s_{i+1} = s_g \in S_g$.



Sequential Decision

Definition: (Sequential Decision)

Given a Markov decision process, $\mathrm{MDP} = \langle S,\ A,\ P,\ R\rangle$, initial state $s_0 \in S$, the action $a_i \in A$ and the reward of the environment $r_i \in R$, then according to each step's state and rewards, and taking its action, that is, $P\left(s_{i+1} | s_i, a_i, r_i\right)$, finally reaching the target state $s_{i+1} = s_g$, known as sequential decision or sequential decision-making.



Thank You