

# **Fundamentals of Artificial Intelligence and Knowledge Representation**

Last update: 13 October 2023

Academic Year 2023 – 2024  
Alma Mater Studiorum · University of Bologna

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	AI systems classification . . . . .	1
1.1.1	Intelligence classification . . . . .	1
1.1.2	Capability classification . . . . .	1
1.1.3	AI approaches . . . . .	1
1.2	Symbolic AI . . . . .	1
1.3	Machine learning . . . . .	2
1.3.1	Training approach . . . . .	2
1.3.2	Tasks . . . . .	2
1.3.3	Neural networks . . . . .	2
1.4	Automated planning . . . . .	3
1.5	Swarm intelligence . . . . .	3
1.6	Decision support systems . . . . .	3
<b>2</b>	<b>Search problems</b>	<b>5</b>
2.1	Search strategies . . . . .	5
2.1.1	Search tree . . . . .	5
2.1.2	Strategies . . . . .	5
2.1.3	Evaluation . . . . .	6
2.2	Non-informed search . . . . .	6
2.2.1	Breadth-first search (BFS) . . . . .	6
2.2.2	Uniform-cost search . . . . .	6
2.2.3	Depth-first search (DFS) . . . . .	7
2.2.4	Depth-limited search . . . . .	7
2.2.5	Iterative deepening . . . . .	7
2.3	Informed search . . . . .	8
2.3.1	Best-first search . . . . .	8
2.4	Graph search . . . . .	10
2.4.1	A* with graphs . . . . .	10

# 1 Introduction

## 1.1 AI systems classification

### 1.1.1 Intelligence classification

Intelligence is defined as the ability to perceive or infer information and to retain the knowledge for future use.

**Weak AI** aims to build a system that acts as an intelligent system. Weak AI

**Strong AI** aims to build a system that is actually intelligent. Strong AI

### 1.1.2 Capability classification

**General AI** systems able to solve any generalized task. General AI

**Narrow AI** systems able to solve a particular task. Narrow AI

### 1.1.3 AI approaches

**Symbolic AI (top-down)** Symbolic representation of knowledge, understandable by humans. Symbolic AI

**Connectionist approach (bottom up)** Neural networks. Knowledge is encoded and not understandable by humans. Connectionist approach

## 1.2 Symbolic AI

**Deductive reasoning** Conclude something given some premises (general to specific). It is unable to produce new knowledge. Deductive reasoning

**Example.** "All men are mortal" and "Socrates is a man"  $\rightarrow$  "Socrates is mortal"

**Inductive reasoning** A conclusion is derived from an observation (specific to general). Produces new knowledge, but correctness is not guaranteed. Inductive reasoning

**Example.** "Several birds fly"  $\rightarrow$  "All birds fly"

**Abduction reasoning** An explanation of the conclusion is found from known premises. Differently from inductive reasoning, it does not search for a general rule. Produces new knowledge, but correctness is not guaranteed. Abduction reasoning

**Example.** "Socrates is dead" (conclusion) and "All men are mortal" (knowledge)  $\rightarrow$  "Socrates is a man"

**Reasoning by analogy** Principle of similarity (e.g. k-nearest-neighbor algorithm). Reasoning by analogy

**Example.** "Socrates loves philosophy" and Socrates resembles John  $\rightarrow$  "John loves philosophy"

**Constraint reasoning and optimization** Constraints, probability, statistics. Constraint reasoning

## 1.3 Machine learning

### 1.3.1 Training approach

<b>Supervised learning</b> Trained on labeled data (ground truth is known). Suitable for classification and regression tasks.	Supervised learning
<b>Unsupervised learning</b> Trained on unlabeled data (the system makes its own discoveries). Suitable for clustering and data mining.	Unsupervised learning
<b>Semi-supervised learning</b> The system is first trained to synthesize data in an unsupervised manner, followed by a supervised phase.	Semi-supervised learning
<b>Reinforcement learning</b> An agent learns by simulating actions in an environment with rewards and punishments depending on its choices.	Reinforcement learning

### 1.3.2 Tasks

<b>Classification</b> Supervised task that, given the input variables $X$ and the output (discrete) categories $Y$ , aims to approximate a mapping function $f : X \rightarrow Y$ .	Classification
<b>Regression</b> Supervised task that, given the input variables $X$ and the output (continuous) variables $Y$ , aims to approximate a mapping function $f : X \rightarrow Y$ .	Regression
<b>Clustering</b> Unsupervised task that aims to organize objects into groups.	Clustering

### 1.3.3 Neural networks

A neuron (**perceptron**) computes a weighted sum of its inputs and passes the result to an activation function to produce the output. Perceptron

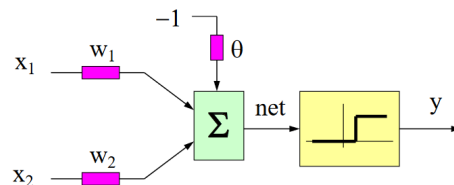


Figure 1.1: Representation of an artificial neuron

A **feed-forward neural network** is composed of multiple layers of neurons, each connected to the next one. The first layer is the input layer, while the last is the output layer. Intermediate layers are hidden layers. Feed-forward neural network

The expressivity of a neural networks increases when more neurons are used:

**Single perceptron** Able to compute a linear separation.

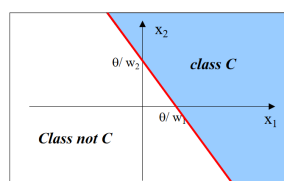


Figure 1.2: Separation performed by one perceptron

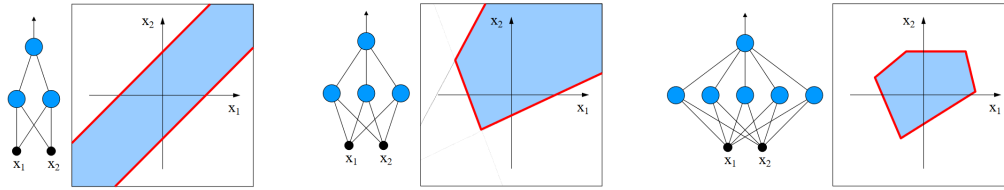


Figure 1.3: Separation performed by a three-layer network

**Three-layer network** Able to separate a convex region ( $n_{\text{edges}} \leq n_{\text{hidden neurons}}$ )

**Four-layer network** Able to separate regions of arbitrary shape.

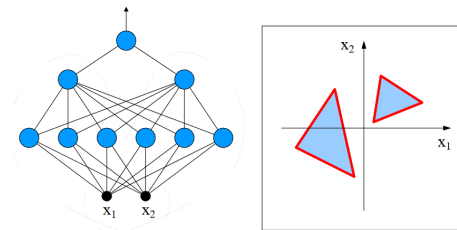


Figure 1.4: Separation performed by a four-layer network

**Theorem 1.3.1** (Universal approximation theorem). A feed-forward network with one hidden layer and a finite number of neurons is able to approximate any continuous function with desired accuracy.

Universal approximation theorem

**Deep learning** Neural network with a large number of layers and neurons. The learning process is hierarchical: the network exploits simple features in the first layers and synthesis more complex concepts while advancing through the layers.

Deep learning

## 1.4 Automated planning

Given an initial state, a set of actions and a goal, **automated planning** aims to find a partially or totally ordered sequence of actions to achieve a goal.

Automated planning

An **automated planner** is an agent that operates in a given domain described by:

- Representation of the initial state
- Representation of a goal
- Formal description of the possible actions (preconditions and effects)

## 1.5 Swarm intelligence

Decentralized and self-organized systems that result in emergent behaviors.

Swarm intelligence

## 1.6 Decision support systems

**Knowledge based system** Use knowledge (and data) to support human decisions. Bottlenecked by knowledge acquisition.

Knowledge based system

*Not required for the exam*

Different levels of decision support exist:

<b>Descriptive analytics</b>	Data are used to describe the system (e.g. dashboards, reports, ...). Human intervention is required.	Descriptive analytics
<b>Diagnostic analytics</b>	Data are used to understand causes (e.g. fault diagnosis) Decisions are made by humans.	Diagnostic analytics
<b>Predictive analytics</b>	Data are used to predict future evolutions of the system. Uses machine learning models or simulators (digital twins)	Predictive analytics
<b>Prescriptive analytics</b>	Make decisions by finding the preferred scenario. Uses optimization systems, combinatorial solvers or logical solvers.	Prescriptive analytics

## 2 Search problems

### 2.1 Search strategies

<b>Solution space</b>	Set of all the possible sequences of actions an agent may apply. Some of these lead to a solution.	Solution space
<b>Search algorithm</b>	Takes a problem as input and returns a sequence of actions that solves the problem (if exists).	Search algorithm

#### 2.1.1 Search tree

<b>Expansion</b>	Starting from a state, apply a successor function and generate a new state.	Expansion
<b>Search strategy</b>	Choose which state to expand. Usually is implemented using a fringe that decides which is the next node to expand.	Search strategy
<b>Search tree</b>	Tree structure to represent the expansion of all states starting from a root (i.e. the representation of the solution space).  Nodes are states and branches are actions. A leaf can be a state to expand, a solution or a dead-end. Algorithm 1 describes a generic tree search algorithm.	Search tree

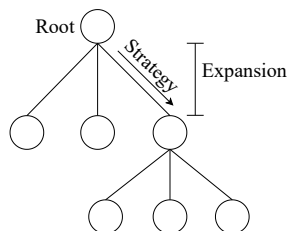


Figure 2.1: Search tree

Each node contains:

- The state
- The parent node
- The action that led to this node
- The depth of the node
- The cost of the path from the root to this node

#### 2.1.2 Strategies

<b>Non-informed strategy</b>	Domain knowledge not available. Usually does an exhaustive search.	Non-informed strategy
<b>Informed strategy</b>	Use domain knowledge by using heuristics.	Informed strategy

---

**Algorithm 1** Tree search

---

```
def treeSearch(problem, fringe):
    fringe.push(problem.initial_state)
    # Get a node in the fringe and expand it if it is not a solution
    while fringe.notEmpty():
        node = fringe.pop()
        if problem.isGoal(node.state):
            return node.solution
        fringe.pushAll(expand(node, problem))
    return FAILURE

def expand(node, problem):
    successors = set()
    # List all neighboring nodes
    for action, result in problem.successor(node.state):
        s = new Node(
            parent=node, action=action, state=result, depth=node.depth+1,
            cost=node.cost + problem.pathCost(node, s, action)
        )
        successors.add(s)
    return successors
```

---

### 2.1.3 Evaluation

**Completeness** if the strategy is guaranteed to find a solution (when exists).

Completeness

**Time complexity** time needed to complete the search.

Time complexity

**Space complexity** memory needed to complete the search.

Space complexity

**Optimality** if the strategy finds the best solution (when more solutions are possible).

Optimality

## 2.2 Non-informed search

### 2.2.1 Breadth-first search (BFS)

Always expands the less deep node. The fringe is implemented as a queue (FIFO).

Breadth-first search

<b>Completeness</b>	Yes
<b>Optimality</b>	Only with uniform cost (i.e. all edges have same cost)
<b>Time and space complexity</b>	$O(b^d)$ , where the solution depth is $d$ and the branching factor is $b$ (i.e. each non-leaf node has $b$ children)

The exponential space complexity makes BFS impractical for large problems.

### 2.2.2 Uniform-cost search

Same as BFS, but always expands the node with the lowest cumulative cost.

Uniform-cost search

<b>Completeness</b>	Yes
<b>Optimality</b>	Yes
<b>Time and space complexity</b>	$O(b^d)$ , with solution depth $d$ and branching factor $b$



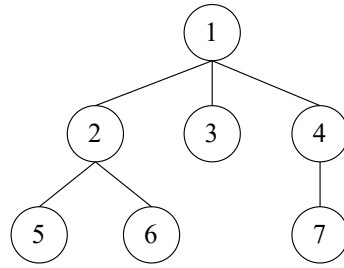


Figure 2.2: BFS visit order

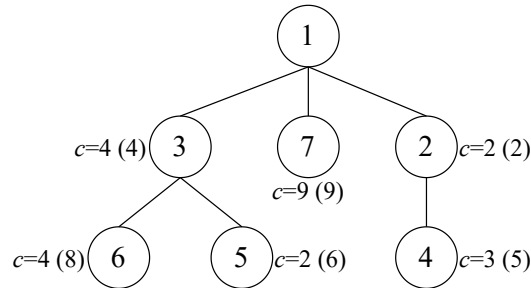


Figure 2.3: Uniform-cost search visit order.  $(n)$  is the cumulative cost

### 2.2.3 Depth-first search (DFS)

Always expands the deepest node. The fringe is implemented as a stack (LIFO).

Depth-first search

<b>Completeness</b>	No (loops)
<b>Optimality</b>	No
<b>Time complexity</b>	$O(b^m)$ , with maximum depth $m$ and branching factor $b$
<b>Space complexity</b>	$O(b \cdot m)$ , with maximum depth $m$ and branching factor $b$

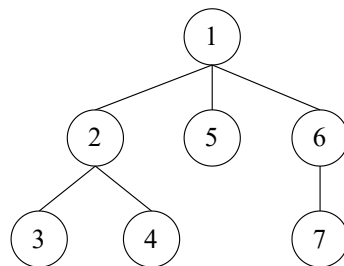


Figure 2.4: DFS visit order

### 2.2.4 Depth-limited search

Same as DFS, but introduces a maximum depth. A node at the maximum depth will not be explored further.

Depth-limited search

This allows to avoid infinite branches (i.e. loops).

### 2.2.5 Iterative deepening

Iterative deepening

Runs a depth-limited search by trying all possible depth limits. It is important to note that each iteration is executed from scratch (i.e. a new execution of depth-limited search).

---

**Algorithm 2** Iterative deepening

---

```
def iterativeDeepening(G):
    for c in range(G.max_depth):
        sol = depthLimitedSearch(G, c)
        if sol is not FAILURE:
            return sol
    return FAILURE
```

---

Both advantages of DFS and BFS are combined.

<b>Completeness</b>	Yes
<b>Optimality</b>	Only with uniform cost
<b>Time complexity</b>	$O(b^d)$ , with solution depth $d$ and branching factor $b$
<b>Space complexity</b>	$O(b \cdot d)$ , with solution depth $d$ and branching factor $b$

## 2.3 Informed search

Informed search uses evaluation functions (heuristics) to reduce the search space and estimate the effort needed to reach the final goal.

Informed search

### 2.3.1 Best-first search

Uses heuristics to compute the desirability of the nodes (i.e. how close they are to the goal). The fringe is ordered according to the estimated scores.

Best-first search

**Greedy search / Hill climbing** The heuristic only evaluates nodes individually and does not consider the path to the root (i.e. expands the node that currently seems closer to the goal).

Greedy search / Hill climbing

<b>Completeness</b>	No (loops)
<b>Optimality</b>	No
<b>Time and space complexity</b>	$O(b^d)$ , with solution depth $d$ and branching factor $b$

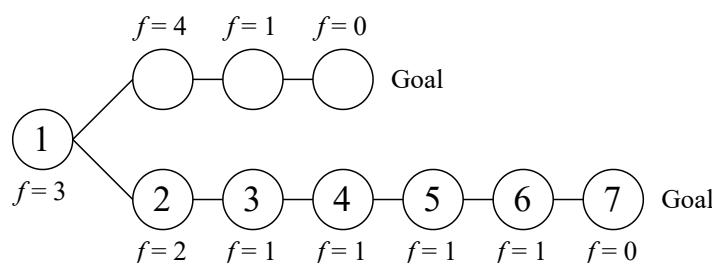


Figure 2.5: Hill climbing visit order

**A\*** The heuristic also considers the cumulative cost needed to reach a node from the root. The score associated to a node  $n$  is:

$$f(n) = g(n) + h'(n)$$

where  $g$  is the depth of the node and  $h'$  is the heuristic that computes the distance to the goal.

**Optimistic/Feasible heuristic** Given  $t(n)$  that computes the true distance of a node  $n$  to the goal. An heuristic  $h'(n)$  is optimistic (i.e. feasible) if:

$$h'(n) \leq t(n)$$

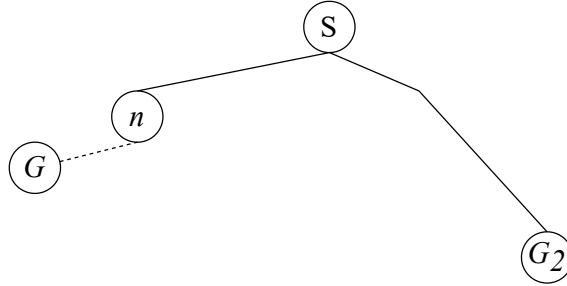
Optimistic/Feasible heuristic

In other words,  $h'$  is optimistic if it always underestimates the distance to the goal.

**Theorem 2.3.1.** If the heuristic used by A\* is optimistic  $\Rightarrow$  A\* is optimal

*Proof.* Consider a scenario where the queue contains:

- A node  $n$  whose child is the optimal solution
- A sub-optimal solution  $G_2$



We want to prove that A\* will always expand  $n$ .

Given an optimistic heuristic  $f(n) = g(n) + h'(n)$  and the true distance of a node  $n$  to the goal  $t(n)$ , we have that:

$$f(G_2) = g(G_2) + h'(G_2) = g(G_2), \text{ as } G_2 \text{ is a solution: } h'(G_2) = 0$$

$$f(G) = g(G) + h'(G) = g(G), \text{ as } G \text{ is a solution: } h'(G) = 0$$

Moreover,  $g(G_2) > g(G)$  as  $G_2$  is suboptimal. Therefore,  $f(G_2) > f(G)$ .

Furthermore, as  $h'$  is feasible, we have that:

$$\begin{aligned} h'(n) \leq t(n) &\iff g(n) + h'(n) \leq g(n) + t(n) = g(G) = f(G) \\ &\iff f(n) \leq f(G) \end{aligned}$$

In the end, we have that  $f(G_2) > f(G) \geq f(n)$ . So we can conclude that A\* will never expand  $G_2$  as:

$$f(G_2) > f(n)$$

□

<b>Completeness</b>	Yes
<b>Optimality</b>	Only if the heuristic is optimistic
<b>Time and space complexity</b>	$O(b^d)$ , with solution depth $d$ and branching factor $b$

In generally, it is better to use heuristics with large values (i.e. heuristics that don't underestimate too much).

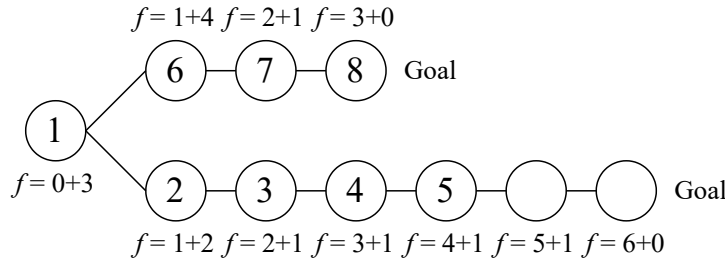


Figure 2.6: A\* visit order

## 2.4 Graph search

Differently from a tree search, searching in a graph requires to keep track of the explored nodes.

Graph search

---

### Algorithm 3 Graph search

---

```
def graphSearch(problem, fringe):
    closed = set()
    fringe.push(problem.initial_state)
    # Get a node in the fringe and
    # expand it if it is not a solution and is not closed
    while fringe.notEmpty():
        node = fringe.pop()
        if problem.isGoal(node.state):
            return node.solution
        if node.state not in closed:
            closed.add(node.state)
            fringe.pushAll(expand(node, problem))
    return FAILURE
```

---

### 2.4.1 A\* with graphs

The algorithm keeps track of closed and open nodes. The heuristic  $g(n)$  evaluates the minimum distance from the root to the node  $n$ .

A\* with graphs

**Consistent heuristic (monotone)** An heuristic is consistent if for each  $n$ , for any successor  $n'$  of  $n$  (i.e. nodes reachable from  $n$  by making an action) holds that:

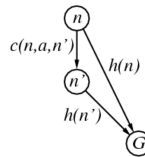
Consistent heuristic (monotone)

$$\begin{cases} h(n) = 0 & \text{if the corresponding status is the goal} \\ h(n) \leq c(n, a, n') + h(n') & \text{otherwise} \end{cases}$$

where  $c(n, a, n')$  is the cost to reach  $n'$  from  $n$  by taking the action  $a$ .

In other words,  $f$  never decreases along a path. In fact:

$$\begin{aligned} f(n') &= g(n') + h(n') \\ &= g(n) + c(n, a, n') + h(n') \\ &\geq g(n) + h(n) \\ &= f(n) \end{aligned}$$



**Theorem 2.4.1.** If  $h$  is a consistent heuristic, A\* on graphs is optimal.