Artificial Intelligence Fundamentals

Federico Matteoni

A.A. 2021/22

Index

0.1 Introduction

Prof.s: Maria Simi, Vincenzo Lomonaco
AI is taking over the world. Formalizing common sense is a lot more difficult. We can formalize knowledge in very specific and small domains. But is deep learning the final solution to AI? "It will transform many industries, but it's not magic. Almost all of AI's recent progress is based on one type of AI, in which some input is used to quickly generate simple response." (*Andrew Ng*)
This AI can do supervised learning, but requires huge amount of data (tens of thousands of pictures to build a photo tagger, for example). The rule of thumb of Ng is: if a person can do a mental task with less than one second of thought, we can automate it using AI either now or in the near future.
The challenges are:

> Software is not a problem, the community is open and the software can be replicated the software can be replicated
>
> Data is exceedingly difficult to get access to. Data is the defensible barrier for many businesses
>
> Talent, because downloading and applying open-source software to your data won't work. AI needs to be customized to context and data, that's why there's a war for the scarce AI talent that can do this work.
>
> Computational resources are also very important.

Deep Learning Is only one approach inside the much wider field of ML and ML is only one approach in the wider field of AI. Book: *Thinking Fast and Slow*, Kahneman. Two systems: system 1 does perceptual tasks, simple computations, system 2 instead does complex computation, recalling from memory... this is a distinction in our brains.

Machine Learning Is AI all about machine learning? Possible arguments against ML are:

> Explanation and accountability: ML systems are not (yet?) able to justify in human terms their results. For some applications this is essential: knowledge must be meaningful to humans to be able to generate explanations? Some regulations requires the right to an explanation in decision-making, and seek to prevent discrimination based on race, opinions, sex... (see GDPR)
>
> ML systems learn what's in the data, **without understanding what's true or false, real or imaginary, fair or unfair**. It is possible to develop unfair, bad models. People are generally more critical about information.

Building AI systems is a goal far from being solved, still quite challenging. Complex AI systems requires the combination of several techniques and approaches, not only ML.

AI Fundamentals Is mostly about reasoning and *slow thinking*. Different approaches, "good old-fashioned artificial intelligence" or "symbolic AI": teaching about the foundations of the discipline, now 60 years old.

Symbolic AI High-level human readable representations of problems, the general paradigm of searching for a solution, knowledge representation and reasoning, planning. Dominant paradigm from the mid 1950s until late 1980s. Central to the building of AI systems is the physical symbol systems hypothesis (PSSH), formulated by Newell and Simon (*Computer Science as Empirical Inquiry: Symbols and Search*)
The approach is based on the assumption that many aspects of intelligence can be achieved by the manipulation of symbols (the PSSH): *a physical symbol system has the necessary and sufficient means for general intelligent action.* Human thinking is a king of symbol manipulation system (so a symbol system is **necessary** for intelligence), and machine can be intelligent (a symbol system is **sufficient** for intelligence). This cannot be prove, we can only collect empirical evidence: observation and experiments on human behavior in tasks requiring intelligence, and solving tasks of increasing complexity.

Strong and Weak AI The Chinese room argument, by John Searle, introduced the following distinction: strong ai relies on the *strong* assumption that human intelligence can be reproduced in all its aspects (general AI), including adaptivity, learning, consciousness..., while weak AI is the simulation of human-like behavior, without effetctive thinking or understanding, no claim that it works like the human mind. Dominant approach today, fragmented AI.
One strong argument against strong AI is the lack of needs by the systems: biological need, safety, relationships, self esteem, self-actualization (Maslow's hierarchy of needs).
What stands in the way of all-powerful AI is not a lack of smarts: it's that computers can't have needs, cravings or desires.

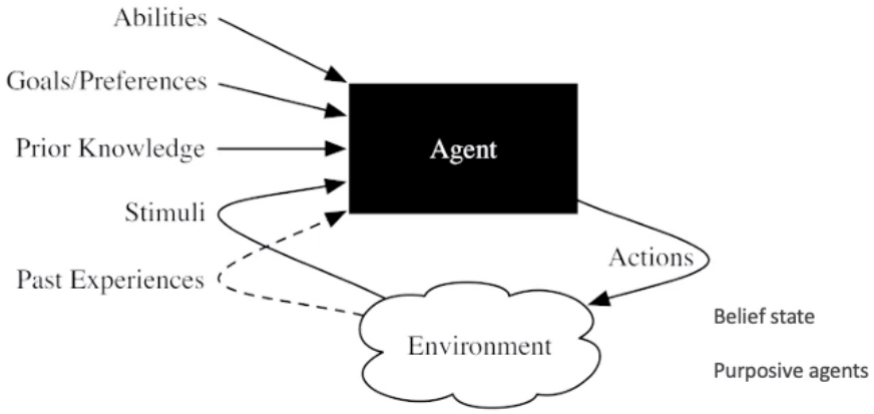
AI is the enterprise of building intelligent computational agents

Agents An agent is something that acts in an environment. We are interested in what an agetn does, that is how it acts. We judge and agent by its action. An agent acts intelligently when: what it does is appropriate given the circumnstances and its goals, it is flexible to changing envoronments and changing goals, learns from experience, makes appropriate choices given its perceptual and computational limitations.
Computational agent is an agent whose decsiions about its actions can be explained in terms of computation and implkemented on a physical device.

> **Scientific Goal**: undestrand the principles that make intelligent behavior possibile in natural or artificial systems
>
> **Engineering Goal**: design and synthesis of userful, intelligente artifacts, agents that are useful in many applications

Artificial Intelligence Artificial intelligent is not the opposit of real intelligence. Intelligence cannot be *fake*: in an artificial agent behaves intelligently, it is intelligent. It is only the external behavior that defines intelligence, according to the **Turing Test** (weak AI). So **artificial intelligence is real intelligence created artificially**.
More updated test: Winograd schemas.
Human intelligence: biology (surviving various habitats), culture (language, tools, concepts, wisdom passed from parents and teachers to children) and life-long learning experience (learning throughout life). Another form is social intelligence, exhibited by communities and organizations.

So agents are situated in environments, inputs are abilities, goals, prior knowledge, stimuli and past experiences, and outputs actions which affect the environment.

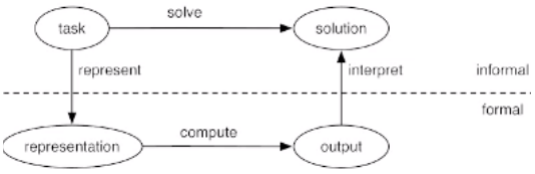


Design process

> design time computation, that goes into the design
>
> offline computation, that the agent can do before acting in the world (ex: specializing the model)
>
> online computation, done by the agent that is acting

Designing an intelligent agent that can adapt to complex environments and changing goals is a major challenge. Two strategies: simplify environments and build strong reasoning systems for these simple environments, or build simple agents for natural/complex environments simplifying the task.
Steps in the design process:

> define the task in natural language, what need to be computed
>
> define what is a solution and its quality: optimal, satisfying, aproximately optimal, probable...
>
> formal representation for the task, choosing how to represent knowledge for the task, including representations suitable for learning.
>
> compute an output
>
> interpret output as solution



Levels of abstraction A model of the world is a syumbolic rtepresentation of the beliefs of the agents. In is necessarily an abstraction: more abstract representations are simpler and human-readable but they may not be effective enough. Low level descriptions are more detailed and accurate but more complex too. Multiple level of abstractions are possibile (hierchical design). Two levels always present in the design: knowledge level (what the agent knows and its goals, not in terms of how we represent) and the symbol level (internal representation and reasoning system). **Modularity** extent to which a system/task can be decomposed

flat: not modular

modular: interacting modules that can be understood on their own

hierarchical: modules are decomposed into simpler modules

Planning horizon how far ahead in time the agent plans

non planning agent

finite horizon planner: looks for a fixed amount of stages, greedy if only one step ahead

indefinite horizon planner: finite but not predetermined number of stages

infinite horizon planner: keeps planning forever (ex: stabilization module of a legged robot)

Representation concerns how the state of the world is described

Atomic states

feature-based representation: set of propositions that are true or false (PROP, CSP, most ML)

individuals and relations, or relational representation

Computational limits that determines whether an agent has

perfect rationality, reasons about the best actions without constraints

bounded rationality, decides on the best action that it can find given its limits

An anytime algorithm is an algorithm where the solution improves with time.
Learning dimensions determines whether

knowledge is given in advance, or

knowledge is learned (from data or past experience)

Learning typically means finding the best model th **Uncertainty**, which can be

in sensing (fully/partially observable states)

about the effects of the actions (deterministic/stochastic)

Preference dimension which considers thetehr the agent has

goal (achievemtn goal a proposition true in a final state, or mainentance goal, proposition true in all psobbile states)

complex preferences, involving trade-offs among the desiderability of various...

Number of agents

single agent reasoning

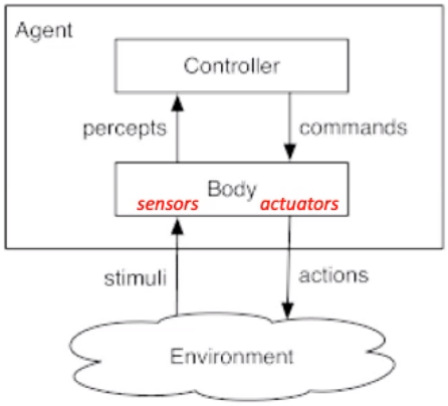
multi agent resoning

Interaction considers wheter the agent does:

offline reasoning or

online reasoning

Agent Architectures Agent interacts with an environment, receives informations with sensors and acts in the world with actuators. Robot: physical body. Program: software agent, digital environment.
Agent is made of body and controller, which receives percepts from the body and sends command to the body. A body includes sensors that converts stimuli into percepts and actautors that convert commands into actions.



Bot sensors and Agents act in time. T is a set of time points, with start at 0, totally ordered, discrete and each t has a next time $t+1$.
Percept trace/stream: function of time into percepts (past, present, future)
Command trace: a function of time into commands (past, present, future)
History at time t: percepts up to t and commands up to $t-1$

Causal Trasduction Function from history to commands. Transduction comes from *finite state transducers*, where both new states and commands are emitted. "Causal" because only previous and current percepts and previous commands can be considered. A controller ideally implements a causal transduction.
But complete history is usually unavailable, only the memory of it. The belief state of an agent at time t is all the information that the agent remembers from the previous times. The behavior of an agent can be described by two functions:

> **Belief State function** $remember : S \times P \to S$ with S being the set of belief states and P the set of percepts
>
> **Command function** $command : S \times P \to C$ with C being the set of commands.

The controller implements both, an approximation of the causal transduction.

Problem Solving as search The dominant approach to AI is formulating a task as a search in a state space. The paradigm is as follows:

> Define a goal (a set of states, a boolean test function...)
>
> Formulate the task as a search problem: define a representation for states and define legal actions and transition functions
>
> Find a solution (a sequence of actions) by means of a search process
>
> Execute the plan

This is a basic technique in AI: search happens inside the agent, it's the planning stage before acting. It's different from searching the world, when an agent may have to act in the world and interleave an action with planning.
Search is a general paradigm, underlying much of the artificial intelligence field. An agent is usually given only a description of what it should achieve, not an algorithm to solve it. The only possibility is to search for a solution. Searching can be computationally very hard (NP-Complete).
Humans are able to solve specific instances by using their knowledge about the problem. This extra knowledge is called **heuristic knowledge**.

Assumptions in classic problem solving Problem solving agents are goal driven agents, that work under simplified assumptions made in the design process.

> States are treated as black boxes: we only need to know the heuristic value and whether they are a goal by applying the boolean goal function. The internal structure doesn't matter from the point of view of search algorithms. **Atomic representations**.

The agent has **perfect knowledge** of the state (full accessibility), no uncertainty in sensors.

Actions are **deterministic**, so that the agent know the consequences of its actions.

The state space is generated incrementally: can be infinite, so may not fit in memory.

Problem formulation A problem is defined formally by five components:

Initial state

Possible actions in state s, $Actions(s)$

Transition model: a function $Result : State \times Action \to State$
$Result(s, a) = s'$, a **successor state**

Goal states are defined by a boolean function
$Goal\text{-}Test(s) \to \{true, false\}$

Path-cost function, that assigns a numeric cost to each path. The sum of the cost of the actions on the path $c(s, a, s')$

Graphs for searching A (directed) graph consists of a set N of nodes and a set A of arcs, which are ordered pairs of nodes. Node n_2 is a neighbor/successor of n_1 if $\exists\,(n_1, n_2) \in A$, and a path is a sequence of nodes $(n_0, \ldots, n_k)$ such that $(n_{i-1}, n_i) \in A$ with length k.
The cost of the path is the sum of the costs of its arcs $cost((n_0, \ldots, n_k)) = \sum_{i=1}^{k} cost((n_{i-1}, n_i))$.
A **solution** is a path from a start node to a goal node and an **optimal solution** is one with minimum cost.

Search algorithms A problem is given as input to a search algorithm. A solution to a problem is a path (actions sequence) that leads from the initial state to a goal state.
Solution quality is measured by the path cost function and an optimal solution has the lowest path cost among all solutions. Different strategies (algorithms) for searching the state space may be characterized by:

Their time and space complexity, completeness, optimality...

Uninformed search methods vs informed/heuristic search methods, which use an heuristic evaluation function of the nodes

Direction of search (forward or backwards)

Global vs local search methods

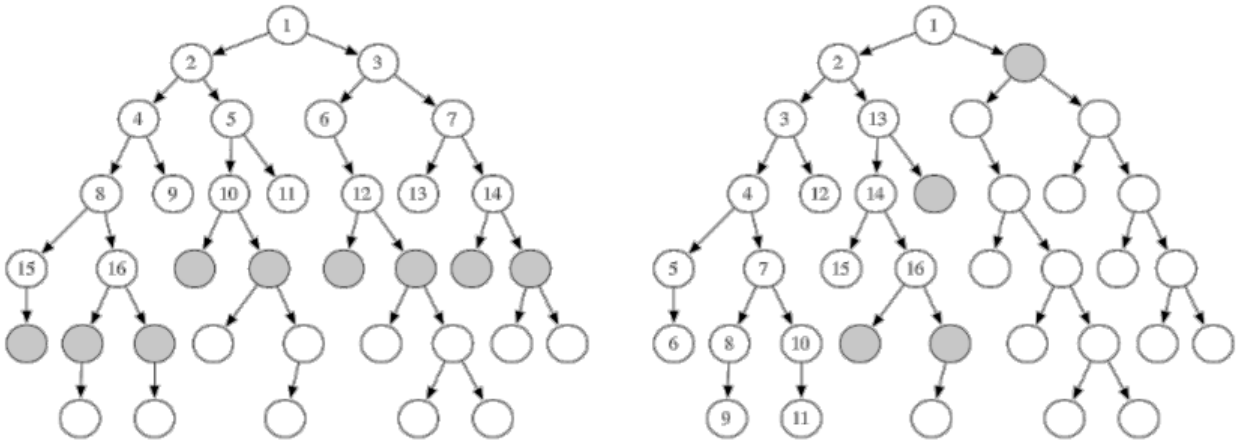
Generic search algorithm

```
input:
a graph
a set of nodes
boolean function goal(n) that tests if n is a goal node

frontier := {s | s is a start node}
while frontier is not empty:
select and remove path (n0, ..., nk) from frontier
if goal(nk):
return (n0, ..., nk)
for each neighbor n of nk
add (n0, ..., nk, n) to frontier
end while
return fail
```

Other algorithms With b max number of successors, d depth of solution and m max distance of solution.

Breadth and Depth search Respectively:



Breadth search: complete, optimal, time $O(b^d)$, space $O(b^d)$
Depth search: not complete, time $O(b^m)$, space $O(bm)$

Depth bounded search: supposes to know the distance of the solution, performs depth-first up to a limit without giving up completeness

Iterative deepening: tries depth limit 1, then 2, then 3 and so on, freeing memory from one iteration to the next.

Uniform cost search: at each stage, selects a path on the frontier with lowest cost.
The frontier is priority queue ordered by path cost, so the first path to goal is the least-cost path. When arc costs are equal is equivalent to breadth-first search.
This strategy is complete, provided that the branching factor is finite and there is some $\epsilon > 0$ such that all the costs are $> \epsilon$. It's also optimal, since it guarantees that the paths with lower costs are found first.

Heuristic search The idea is to not ignoring the goal when selecting the paths. Often there's extra knowledge that can be used to guide the search: **heuristics**, provided by an heuristic function $h : N \to R \Rightarrow h(n)$ is the estimate of the cost of the shortest path from node n to the goal node.
h needs to be efficiently to compute. An **admissible heuristic** $h^*(n)$ is a non-negative heuristic function that **underestimates** the minimum cost of a path from a node n to a goal: $\forall\, n \;\; h(n) \leq h^*(n)$

Best first search selects the most promising node on the frontier according to the heuristic function.

A^* **search** With an heuristic function in the form $f(n) = g(n) + h(n)$ with:

$g(n)$ being the cost of path leading to n (so the previous path up until n, $cost(n)$)

$h(n)$ is an admissible heuristic (so, $h(n) \geq 0$)

Then $f(n)$ estimates the total path cost of going from a start node to a goal via n. The special cases are $h = 0$ (lowest cost search) and $g = 0$ (greedy best first).
Properties of A^*:

Complete

Always finds an optimal solution, if the branching factor is finite and arc costs are bounded above 0 (which means that $\exists\, \epsilon > 0 \mid$ arc costs are $> \epsilon$)

Optimiziations are possibile when searching graphs

The operate some sort of graph pruning:

Cycle pruning: doesn't add nodes to the frontier with states already encountered along the path (easy)

Multiple-path pruning: maintains an explored set of nodes that are at the end of paths that have been expanded. When an n is selected, if its state is already in the explored set, it's discarded.

Memory requirement is exponential ($O(b^d)$). Can be mitigated in some ways:

IDA^*: performs repeated depth-bounded searches with value of $f(n)$ used as bound

Recursive best-first, similar to branch & bound

SMA^* (simplified memory-bounded A^*)

Beam search, keeps in frontier only the best k paths, with k being the beam width (gives up optimality)

Consistent heuristics An heuristic that statisfies the monotone restriction guarantees consistency $h(n) \leq cost(n, n') + h(n')$
Consistency $\Rightarrow$ admissibility. With the monotone restriction, the f-values of the paths selected from the frontier are monotonically non-decreasing.

Features Often better to describe states in terms of features: **factored representation**, more natural and efficient than explicitly enumerating states. Often, features are not independent and there are constraints that specify legal combinations of assignments. We can exploits these constraints to solve tasks.
Constraint satisfaction is about generating assignments that satisfy a set of hard constraints and how to optimize a collection of soft constraints (preferences).

CSP **Constraint Satisfaction Problem**, formal definition. A Constraint Satisfaction Problem $CSP = \langle X, D, C\rangle$ consists of three components:

A finite set of **variables**, $X = \{x_1, \ldots, x_n\}$

A **finite domain** for each variable, $D = \{D_1, \ldots, D_n\}$ with each $D_i = \{v_1, \ldots, v_k\}$ containing values assignable to x_i.
Dom is a function that maps every variable in X to a set of objects of arbitrary type. $Dom(x) = D_x$

A **set of constraints** that restrict the values the variables can simultaneously take, C

Task: assign a value from the associated domain to each variable satisfying all the constraints. **NP-hard** in worst cases, but general heuristics exist and structures can be exploited for efficiency.

A **(partial) assignment** of values to a set of variables (**compound label**) is a set of pairs $A = \{\langle x_i, v_i\rangle, \ldots\}$ with $v_i \in D_{x_i}$. A **complete assignment** is an assignment to all the variables of the problem. Can be projected to a smaller partial assignment by restricting the variables to a subset (projection, with the following notation: $\pi_{x_1,\ldots,x_k} A$, with π being the projection operator of relational algebra)
Each constant in C can be represented as a pair $\langle$ scope, rel$\rangle$: scope is a tuple of variables participating in the constraint, and rel is a relation that defines the allowable combinations of values for those variables, taken from the respective domains. The relation can be represented as: an explicit list of all tuples of values that satisfy the constraint (explicit relation), or an implicit relation (an object that supports two operations: testing if a tuple is a member of the relation and enumerating the members of the relation).
We also use $C_{x_1,\ldots,x_k} =$ rel to denote a constraint with scope $= x_1, \ldots, x_k$, so the constraint $C = \langle(x_1, \ldots, x_k)$, rel$\rangle$

CSP solution To solve a CSP problem seen as a search problem, we need to define a state space and the notion of a solution.

State: assignment of values to some or all the variables.
Partial if values are assigned only to some of the variables, complete if every variable is assigned.

Solution: a complete and consistent assignment.
An assignment is consistent if it satisfies all the constraints: $Satisfies(\{\langle x_1, v_1\rangle, \ldots, \langle x_k, v_k\rangle\}, C_{x_1,\ldots,x_k})$ for any constraint in C

Problem characteristics

Number of solutions required (one or all)

Problem size (number of variables and constraints)

Type of variables and constraints

Structure of the constraint graph

Tightness of the problems (measured in terms of the solution tuples over the number of all distinct compund labels of all variables)

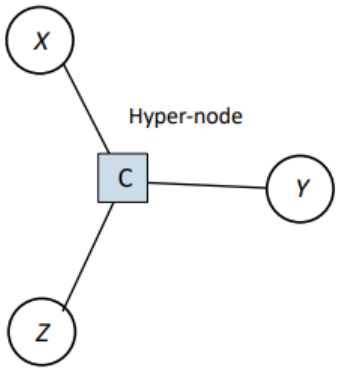
Quality of solutions

Partial solutions

CSP solving techniques Problem reduction techniques/inference/constraint propagation: techniques for transforming CSP into an equivalent problem easier to solve or recognizable as insoluble.
Searching efficiently: heuristics, intelligent backtracking...
Exploiting the structure of the problem: independent sub-problems, tree structured constraint, tere decomp, exploiting symmetry.

Constraint hyper-graphs Binary CSP = CSP with unary and binary constraints only. May be represented as an undirected graph (V, E): nodes corresponds to variables V and edges corresponds to binary constraints between variables ($E = V \times V$). Edges are undirected arcs (can be seen as pair of arcs).
Node x is adiacent to node y is $(x, y) \in E$. A graph is connected if there's a path among any two nodes.

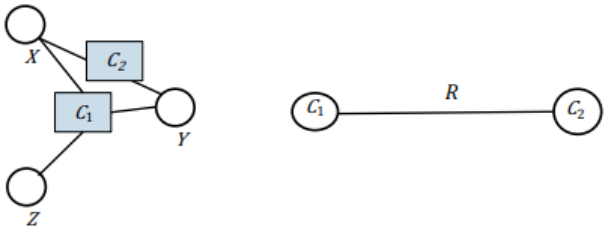
In general, every CSP is associated with a constraint hyper-graph, a generalization of graphs: an hyper-node may connect more than two nodes. The constraint hyper-graph of a CSP $\langle X, D, C \rangle$ is a hyper-graph in which each node represent a variable in C and each hyper-node represents a higher order constraint in C



Dual Graph transformation Alternate way to convert a n-ary CSP to a binary one:

1. Create a new graph in which there is one variable for each constraint in the original graph
2. If two constraints share variables, they are connected by an arc corresponding to the constraint that the shared variables receive the same value

Example: $Dom(x) = Dom(y) = Dom(z) = \{1, 2, 3\}$ with $C_1 = \{(x, y, z), x + y = z\} = \{(1, 2, 3), (2, 1, 3), (1, 1, 2)\}$ and $C_2 = \{(x, y), x < y\} = \{(1, 2), (1, 3), (2, 3)\}$. This will become $Dom(C_1) = \{(1, 2, 3), (2, 1, 3), (1, 1, 2)\}, Dom(C_2) = \{(1, 2), (1, 3), (2, 3)\}$ and $R_{x,y}$ = constraint that x and y will receive the same values



Related concepts

Problem reduction techniques: techniques for transforming CSP into an equivalent problem easier to solve or recognizable as insoluble

Enforcing local consistency: the process of enforcing local consistency properties in a constraint graph causes inconsistent values to be eliminated. Different types of local consistency properties have been studied.

Constraint propagation/inference: constraints are used to reduce the number of legal values for a variable, which in turn can reduce the legal value for another variable and so on...

Problem reduction Reducing a problem means removing those constraints which appear in no solution tuples. A CSP problem P_1 is reduced to P_2 when P_1 is equivalent to P_2, domains of variables in P_2 are subsets of those in P_1 and the constraints in P_2 are at least as restrictive as those in P_1.
These conditions guarantees that a solution in P_2 is also a solution in P_1.
Problem reduction strategies are of two types: removing redundant values from the domains of the variables or tightening the constraints so that fewer compound labels satisfy them (examples: if $x < y$ and $D_x = \{3, 4, 5\}, D_y = \{1, 2, 4\}$ then those can be reduced to $D_x = \{3\}, D_y = \{4\}$.
Constraints are sets, this means removing redundant compound labels from the set. If the domain of any variable or any constraint is reduced to an empty set, then the problem is **unsolvable**.
Problem reduction is also called consistency checking or maintenance, since it relies on establishing local consistency properties.

Local consistency properties: node consistency, arc consistency, path consistency, k-consistency, forward checking.
All these operations do not change the set of solutions, do not necessarily solve a problem but, used with search, will make the search more efficient by pruning the search tree.

Node/Domain consistency: a node is consistent if all the values in its domain satisfy unary constraints on the associated variable. A constraint network is node-consistent if all the nodes are consistent.
Given a unary constraint on x_i, $C_i = \langle (x_i), R_i \rangle$ then node consistency $D_i \subseteq R_i$ and can be enforced by reducing the domains of the variables $D_i \leftarrow D_i \cap R_i$ (this is the NC-1 algorithm, $O(d \cdot n)$)

Arc consistency: a variable is arc-consistent if every value in its domain satisfies the binary constraints of this variable with other variable. x_i is arc-consistent with respect to x_j if for every value in its domain D_i there is some value in the domain D_j that satisfies the binary constraint on the arc (x_i, x_j)
Example, $X = \{x, y\}$, $D_x = D_y = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ and constraint $\langle (x, y), x = y^2 \rangle$. Considering arc $x \rightarrow y$, it can be made consistent by reducing D_x to $\{0, 1, 4, 9\}$. Considering $y \rightarrow x$, can be made consistent by reducing D_y to $\{0, 1, 2, 3\}$, making the entire edge consistent.

Arc Consistency Algorithm (AC-3) It maintains a queue of arcs to consider, initially all the arcs in CSP. An edge produces two arcs. AC-3 pops off an arc (x_i, x_j) from the queue and makes x_i arc-consistent with respect to x_j:

If this step leaves D_i unchanged, the algorithm just moves on to the next arc

If D_i is made smaller, then we need to add to the queue all arcs (x_k, x_j) where x_k is a neighbor of $x_i \neq x_j$

If D_i becomes empty, then we conclude that the CSP has no solution

When there are no more arcs to consider, we have finished: we are left with a CSP equivalent to the original but smaller.
With a CSP of n variables, each with domain size at most d and c binary constraints (arcs):

Checking consistency of an arc can be done in $O(d^2)$ time

Each arc (x_i, x_j) can be inserted in the queue only d times (because x_i has at most d values to delete)

We have c arcs to consider, so complexity is $O(c \cdot d^3)$, polynomial time

The AC-4 algorithm is an improved version of AC-3, based on the notion of support that doesn't need to consider all the incoming arcs. More information is kept, but complexity is $O(c \cdot d^2)$.

Example $A, B, C \in \{1, 2, 3, 4\}, A < B, A > C$

Queue	Arc	Arc domain
$\{(A, B), (B, A), (A, C), (C, A)\}$		
$\{(B, A), (A, C), (C, A)\}$	(A, B)	$A \in \{1, 2, 3, \not{4}\}$
$\{(A, C), (C, A)\}$	(B, A)	$B \in \{\not{1}, 2, 3, 4\}$
$\{(C, A)\}$	(A, C)	$A \in \{\not{1}, 2, 3\}$
$\{(B, A), (C, A)\}$		
$\{(C, A)\}$	(B, A)	$B \in \{\not{2}, 3, 4\}$
$\{\}$	(C, A)	$C \in \{1, 2, \not{3}, \not{4}\}$

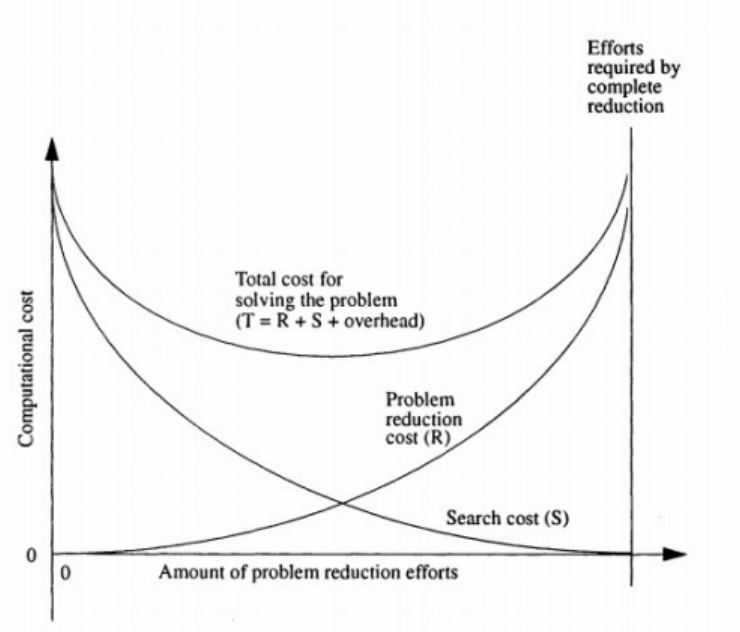
At the end $A \in \{2, 3\}, B \in \{3, 4\}, C \in \{1, 2\}$

Directional Arc Consistency DAC is defined with reference to a total ordering of the variables. A CSP is DAC under an ordering of the variables if and only if for every label $\langle x, a\rangle$ which satisfies the constraints on x there exists a compatible label $\langle y, b\rangle$ for every label y which is after xc according to the ordering.
In the algorithm for establishing DAC (DAC-1) each arc is examined exactly once, by proceeding from last to the first in the ordiering, so the complexity is $O(c \cdot d^2)$. AC cannot always be achieved by running DAC-1 in both directions.

Generalized Arc Consistency GAC, extension of AC-3 to handle n-ary constraints rather than just binary. A variable x_i is GAC with respet to a n-ary constraint if for every value v in the domain of x_i there exists a tuple of values that is a member of the constraint and has its x_i component equal to v.
For example, if $X, y, Z \in \{0, 1, 2, 3\}$ and $X < Y < Z$, to make C consistent we would have to eliminate $2, 3$ from the domain of X because the constraint cannot be satisfied with $X = 2$ or $X = 3$.

Path Consistency Arc consistency tightens down the domains using the arcs (binary constraints). Path consistency is a stronger notion: tightens the binary constraints by using implicit constraints that are inferred by looking at the triples of variables. A path of length 2 between variables x_i, x_j is path-consistent with respect to a third intermediate variable x_m if for every consistent assignment $\{x_i = a, x_j = b\}$ there is an assignment to x_m that satisfies the constraints on (x_i, x_m) and (x_m, x_j). In relational algebra $R_{i,j} \subseteq \pi_{i,j}(R_{i,m} \bowtie D_m \bowtie R_{m,j})$
To achieve path consistency, $R_{i,j} \leftarrow R_{i,j} \cap \pi_{i,j}(R_{i,m} \bowtie D_m \bowtie R_{m,j})$ (algorithm name: PC-2). If all path of length 2 are made consistent, then all paths of any length are consistent.
Called path consistency because you can think of it as a path from x_i to x_j with x_m in the middle.

Combining search and problem reduction Problem reduction techniques are used in combination with search. The more effort one spends on problem reduction, the less effort one needs in searching.



Most problems cannot be resolved by reduction alone, we must search for solutions and combine problem reduction with search. In this context we talk about **constraint propagation** or **inference**. A classical incremental formulation of CSP as a search problem is:

States are partial assignments

Initial state: empty assignment

Goal state: complete assignment that satisfy all constraints

Actions: assign to a specific unassigned variable x_i a value $\in D_i$

Branching factor: d, with d maximum cardinality of the domains. Number of leaves: d^n, with n number of variables. n finite $\Rightarrow$ space graph finite.

CSP as search: simplifications We can exploit **commutativity**. A problem is commutative if the order of application of any given set of action has no effect on the outcome. In this case, the order of the variable assignments does not change the result.
We can consider a single variable for assignment at each step, so the branching factor is d and the number of leaves id d^n. We can also exploit depth limited search: **backtracking search** with depth limit n.
Search strategies:

Generate and Test: generate a full solution and test it, not the best

Anticipated Control: after each assignment we check the constraint, if some is violated we backtrack to previous choices (undoing the assignment)

Backtracking search algorithm

Heuristics and search strategies

`SELECT-UNASSIGNED-VARIABLE`: which variable should be assigned next?

`ORDER-DOMAIN-VALUES`: in which order should the values be tried?

`INFERENCE`: what inference should be performed at each step?
Techniques for **constraint propagation** (local consistency enforcement) can be used

`BACKTRACKING`: where to back up to? When the search ends up in an assignment that violates a constraint, can the search avoid repeating this failure? Forms of **intelligent backtracking**

Choosing the next variable

MRV (minimnum remaining values): variable with fewest "legal" remaining values

Degree heuristic: variable involved in the largest number of constraints

Choosing value

Least constaint variable: prefer the variable rules out the fewest choices

Note that in choosing the variable, a fail-first strategy helps in reducing the amount of search by pruning large parts of the tree earlier. In the choice of value, a fail-last approach works best in CSP where the goal is to find *any* solution. This is not effective if we are looking for all solutions or no solution exists.

Interleaving search and inference One of the simplest form of inference propagation is **forward checking**: efficient constraint propagation, weaker than other forms. Whenever X is assigned, FC process establishes arc consistency of X for the arcs connecting neighbor nodes. For each unassigned Y connected to X, delete from Y's domain any value inconsistent with the value assigned to X.

Constraint learning When the search is at a contradiction, we know that some subset of the conflict set is responsible. Constraint learning is the idea of finding a minimum set of variables from the conflict set that causes the problem: the no-good set. We record the no-good set either by adding a new constraint to CSP or by keeping a separate cache of no-goods. This way we do not repeat the no-good state.

Local search Requires a complete state formulation, keep in memory only current state to improve it iteratively and does not guarantee to find a solution even if it exists (**not complete**).
Used when space too large for systematic search and we need to be very efficient. Also when we need to provide a solution but it's not important to produce solution path. Also when we know in advance that a solution exists.

Local search methods for CSP Complete state formulation: we start with a complete random assignment, and we try to fix it until all the constraints are satisfied.
Local methods are very efficient for large scale problems where the solutions are densely distributed in the space. A basic algorithm is:

```
function Local_search(V, Dom, C) returns a complete & consistent assignment
Inputs: V: a set of variables
Dom: a function such that Dom(x) is the domain of variable x
C: set of constraints to be satisfied
Local: A (complete assignement) an array of values indexed by variables in V

repeat until termination
for each variable x in V do # random initialization or random restart
A[x] := a random value in Dom(x)
while not stop_walk( ) & A is not a satisfying assignment do # local search
Select a variable y and a value w in Dom(y), w != A[y] # a successors
A[y] := w # change a variable
if A is a satisfying assignment then return A # solution found
```

Heuristic Local Search Inject heuristics in the selection of the variable and the value by the means of an evaluation function. **Iterative best improvement**: choose the successor that most improves the current state according to an evaluation function f

Stochastic Local Search Adds randomness, escapes local minima...
Variants

Most improving step

Two stage choice

Any conflict

CSP: Min-Conflict Heuristic

Independent sub-problems: complexity

The structure of problems: tree In a tree-structured graph, two node are connected by only one path: we can choose any variable as root of the tree. Chosen a variable as root, the tree induces a topological sort on the variables: chieldren of a node are listed after their parent

Directional arc consistency

Reducing graphs to trees For example, assigning a value to the node we want to remove and removing inconsistent values for other variables, then solve with tree-csp-solver.
In general not easy.

Cutset conditioning Domain splitting strategy, trying with different assignments: choose subset S of CSP's variabile such that the constraint graph becomes a tree: S is cycle cutset. For each consistent assignmetn to variabiles in S: remove from the domains of remaining vars any vals that are inconsistent, if the remaining CSP has a solution, return it together with the current assignment of S.
Time $O(d^c(n-c)d^2)$ where c is size of cycle cutset and d size of the domain. We have to try each of the d^c coimbinations of values for the cvaraibles in S and for each combination we must solve a tree problem of size $(n-c)$

Tree decomposition The approach consists in a tree decomposition of the contrasint graph into a set of connected sub-roboelkms. Each of them is solved intdepentendly...

every var in the original problem appears in at least one of the sub-problems

if two var connectd by a constraint they must appear together at least in one of the sub problem along with the constraint

if a variabile appears in two subproblems in the tree, it must appear in every subproblem along the path connecting those subproblems

1-2 ensure that all vars and constraints are represented. Condition 3 ensure that any given pvar must have same value in every subproblem.

Solving a decomposed problem

Simmetry important factor for reducing complexity of CSP problems. Value simmetry: the values does not really matter: different colors, but there are 6 equivalent ways of satisfying the constraints. If S is as solution to coloring n var, there are $n!$ solutions.
Simmetry breaking constraints: we impost an arbitrary ordering constraints that requires the values to be in alphabetical order. Breaking value simmetry has proved to be important and effective on many problems.