Knowledge-based Agents

Inference: deriving new sentences from existing ones
Declarative approach to system building: TELLing the agent what it needs to know
Implementation level: data structures inside the KB and algorithms that work on them • **Procedural** approach: encode behaviors directly as program code

Performance measure : Give a value to a task (+1 or -1)
Environment: Grid itself, spawns of things
Actuators: Movement, grab, shoot… etc
Sensors: Breeze near a pit, bump when hitting a wall… etc

Entailment: $\alpha \vDash \beta$, β follows logically from α => derive conclusions: **logical inference**
Properties of inference algorithms: • **Soundness** (or truth preserving): derive only entailed sentences • **Completeness**: derive any sentence that is entailed • If KB is true in the real world, then any sentence α derived from KB by a sound inference procedure is also true in the real world.

WUMPUS World

Observable? Partially/local perception.
Deterministic? Yes for the actions at hand.
Episodic? Sequential, rewards only come after the actions.
Static? Yes, Wumpus doesn't move, nor the pits.
Discrete or **Continuous**? Discrete, the map doesn't get bigger/smaller.
Single-agent? Yes, Wumpus doesn't move, and it is single-player.

Forward Chaining: Data-driven reasoning, Prognostics ⇒ start from the known data, deriving conclusions from incoming percepts. (Start from **P** to achieve **Q**, in **P** ⇒ **Q**)

Backward Chaining: Goal-directed reasoning, Diagnosis ⇒ derive answers to specific goals, starting from a query, instead of the known data (**P**). (Start from **Q** until satisfies **P**, in **P** ⇒ **Q**)

Expert Systems

Advantages: Availability, Reliability, Explainability
Main Tasks: Knowledge acquisition, knowledge representation, reasoning control and explanation

Probability Theory: What exists in the world? Facts; What an agent believes about facts? Degree of belief ∈ [0,1]
Bayes' Rule: $\boldsymbol{P}(\boldsymbol{b} \mid \boldsymbol{a}) = \boldsymbol{P}(\boldsymbol{a} \mid \boldsymbol{b}) * \boldsymbol{P}(\boldsymbol{b}) / \boldsymbol{P}(\boldsymbol{a})$
Naïve Bayes: Assumes variables are conditionally independent, even when they are not.

Machine Learning

Machine Learning: Give ⇒ Data and Output; Receive ⇒ Program.

Artificial Intelligence < Machine Learning < Deep Learning;

Machine learning cycle ⇒ Gather Data > Prepare it > Analyse > Train > Test > Deploy
Supervised: inputs and outputs are given; **Classification, Regression**
Unsupervised: outputs aren't given; **Clustering, Dimensionality Reduction**
Reinforcement: Data given as feedback/rewards/punishment; **Robot, Game AI**

Knowledge Extraction Methodologies:
• **KDD** (Knowledge Discovery in Databases)
• **SEMMA** (Sample, Explore, Modify, Model and Access)
• **CRISP-DM** (Cross-Industry Standard for Data Mining)

Types of attributes:

• Nominal ⇒ ID numbers, eye color, zip codes. **(=, ≠)**
• Ordinal ⇒ rankings (e.g., taste of potato chips on a scale from 1-10), grades, height {tall, medium, short}. **(<,>)**
• Interval ⇒ calendar dates, temperatures in Celsius or Fahrenheit. **(+, -)**
• Ratio ⇒ temperature in Kelvin, length, counts, elapsed time (e.g., time to run a race). **(*, /)**

Discrete Attribute: Binary

Continuous Attribute: numbers (rating 1-100)

Data quality problems: • Noise and outliers • Wrong data • Fake data • Missing values • Duplicate data

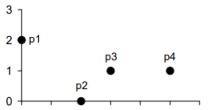
Euclidian Distance: $d(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{k=1}^{n}(x_k - y_k)^2}$

Distance Matrix:

Standardization is necessary, if scales differ

	p1	p2	p3	p4
p1	0	2.828	3.162	5.099
p2	2.828	0	1.414	3.162
p3	3.162	1.414	0	2
p4	5.099	3.162	2	0

Distance Matrix



point	x	y
p1	0	2
p2	2	0
p3	3	1
p4	5	1

Attribute Type	Dissimilarity	Similarity
Nominal	$d = \begin{cases} 0 & \text{if } x = y \\ 1 & \text{if } x \neq y \end{cases}$	$s = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$
Ordinal	$d = \lvert x - y \rvert / (n-1)$ (values mapped to integers 0 to $n-1$, where n is the number of values)	$s = 1 - d$
Interval or Ratio	$d = \lvert x - y \rvert$	$s = -d$, $s = \frac{1}{1+d}$, $s = e^{-d}$, $s = 1 - \frac{d - min\_d}{max\_d - min\_d}$

Data Preprocessing: Aggregation vars• Sampling • Discretization(put values in few categories) and Binarization • Attribute Transformation • Dimensionality Reduction • Feature subset selection • Feature creation;