Artificial Intelligence

An introduction

Maik Kschischo

Institute for Computer Science
University of Koblenz

Table of Contents

Literature

Recommendations

1. Russel, P. and Norvig, S,.Artificial Intelligence: A Modern Approach., 4th ed., Pearson. 2021.
2. Aggarwal, C.C. Artificial Intelligence: A textbook., Springer Nature Switzerland AG 2021. https://doi.org/10.1007/978-3-030-72357-6
3. Ertel, W. Introduction to Artificial Intelligence., 2nd ed., Springer Nature Switzerland AG 2017. https://doi.org/10.1007/978-3-319-58487-4

Assignments

- Oral or written exam, depending in the number of students

Recommendations

- Follow the lectures!
- Read!!
- Try to solve the problems!!!

Chap. 1 Overview and history of AI

What is this course about?

- Introduction to the deductive method (GOFAI: good old fashioned AI method)
- More recent probabilistic graphical models
- Bridging the gap between inductive AI and Machine Learning (ML)
- For ML and Deep Learning, see the specialized courses

Aims:

- Learn about the merits and limitations of deductive approaches
- Start thinking, how the GOFAI can be best combined with inductive ML methods
- Learn about Bayesian Networks
- Recap on probability (useful for ML as well)

Precursors

Myth, legend and fiction

Egyptian, greek and jewish mytholgy (Talos, Golem)

Figure: The death of Talos depicted on a 5th century BC krater now in the Jatta National Archaeological Museum in Ruvo di Puglia.

Early Artificial Intelligence (1943-1952)

- 1943: McCulloch and Walter Pits propose a model of artificial neurons.
- 1949: Donald Hebb demonstrates an updating rule for modifying the connection strength between neurons (now called Hebbian learning).
- 1950: Alan Turing proposes a test for a machine's ability to exhibit intelligent behavior equivalent to human intelligence (Turing test).

Birth of Artificial Intelligence (1952-1956)

- 1952: An Allen Newell and Herbert A. Simon create "Logic Theorist". This program proves 38 of 52 Mathematics theorems, and finds new and more elegant proofs for some theorems.
- 1956: At the Dartmouth Conference John McCarthy calls the new field 'Artificial Intelligence'

Enthusiastic phase (1956-1974)

- The researchers emphasize developing algorithms which can solve mathematical problems.
- 1966: Joseph Weizenbaum create sthe first chatbot named as ELIZA.
- 1973: WABOT-1, the first intelligent humanoid robot is built in Japan.
- 1966: Failure of machine translation.
- 1969: Criticism of perceptrons (early, single-layer artificial neural networks) by Minsky and Papert.
- 1971–75: DARPA's frustration with the Speech Understanding Research program at Carnegie Mellon University.

First AI winter (1974-1980)

- Almost no funding, because of overselling during the enthusiastic phase.

Boom

1980–1987

- **Expert Systems**: A program that answers questions or solves problems about a specific domain of knowledge, using logical rules that are derived from the knowledge of experts.
- **Cyc:** Assemble a comprehensive ontology and knowledge base that spans the basic concepts and rules about how the world works (attack the **commonsense knowledge problem**).
- **Revival of neural networks:**
  - Hopfield Networks, which provide a model of human memory (John Hopfield, 1982).
  - Backpropagation algorithm (Paul Verbos, David Rumelhart, 1985).
  - Applications in optical character recognition and speech recognition.

Bust: second AI winter

1987–1993

- Expert systems turned out too difficult to maintain.
- Cheap PCs from IBM and Apple became more powerful than specialized AI machines.
- Sharp cuts in both academic and commercial research funding.

AI 1993-2011 I

Moores law: Speed and memory capacity of computers doubles every two years

- 1997: Deep Blue became the first computer chess-playing system to beat a world chess champion, Garry Kasparov.
- 2005: Stanford robot won the DARPA Grand Challenge by driving autonomously for 131 miles along an unrehearsed desert trail.
- 2011; IBM's question answering system, Watson, defeated the two greatest Jeopardy! quiz show champions.

Intelligent agents: a system that perceives its environment and takes actions which maximize its chances of success

- Influence of **decision theory** to AI (Pearl, Kaelbling, Newell).
- Probabilistic reasoning: Judea Pearl.

AI 1993-2011 II

- New tools from probability theory came to AI: Bayesian networks, hidden Markov models, information theory, stochastic modeling and classical optimization

Deep Learning and Big Data

2011-2020

Breakthroughs in Neural Networks:

- Hinton, Bengio, Le Cun and others realized that deeper networks learn representations of data and can avoid overfitting, even if shallow networks can represent the same functions in theory (Universal Approximation Theorem).
- Specialized architectures revolutionized image analysis (convolutional networks, resnets).
- Deep learning systems achieved enourmous succes in games like GO.
- Computation on GPU accelerators.

Breakthroughs in Big Data:

- Computational capacity to process huge amounts of data.
- Storage of huge amounts of data in the Internet.

AI era

2020-present

Attention, transformers and large language models:

- 2017 paper "Attention Is All You Need" by Vaswani et al, overcame problems with recurrent architectures.
- Transformers are based on multi-head attention mechanisms and form the basis of large language models (LLMS).

AI era

2020-present

Attention, transformers and large language models:

- 2017 paper "Attention Is All You Need" by Vaswani et al, overcame problems with recurrent architectures.
- Transformers are based on multi-head attention mechanisms and form the basis of large language models (LLMS).

Where are we going from here?

- Interpretability, explainability or even causality?
- What about deductive reasoning? Can it improve ML?
- Will there be a synthesis of induction and deduction?
- What is general artificial intelligence? How can a computer acquire world knowledge?

What is AI?

or what does it want to be

Aims: Design machines that can

- mimic human behaviour (intelligence),
- make decisions,
- learn from experience,
- reason about facts,
- solve problems.

Behaviour is not explicitely preprogrammed into the system, but the algorithms act in some way flexibly.

Agents in AI

AI studies agents in their environment

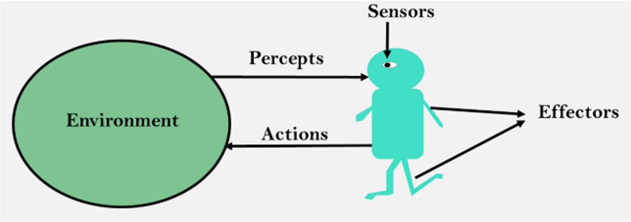

Figure: From https://www.javatpoint.com/agents-in-ai

An agent

- *perceives* its environment through *sensors*
- *acts* upon that environment through *actuators*

Agents

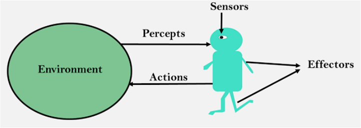

Figure: From https://www.javatpoint.com/agents-in-ai

- **Sensors:** are devices detecting the state of and changes in the environment.
- **Actuators:** are the component of machines that convert energy into motion (e.g. a muscle, an electric motor, gears, etc.).
- **Effectors:** affect the environment (e.g. legs, wheels, arms, fingers, wings, fins, display screen, etc.).

Intelligent (rational) agents

A rational ot intelligent agent can be characterized by the following rules:

- Rule i: AI agents must be able to perceive the environment.
- Rule ii: The observation must be used to make decisions.
- Rule iii: A decision should result in an action.
- Rule iv: The action must be a rational action.

Intelligent (rational) agents

A rational ot intelligent agent can be characterized by the following rules:

Rule i: AI agents must be able to perceive the environment.

Rule ii: The observation must be used to make decisions.

Rule iii: A decision should result in an action.

Rule iv: The action must be a rational action.

Intelligent agent: autonomous entity which acts upon an environment for achieving a *goal*. It can learn from the environment to better achieve the goal.

What is meant with rational?

- there is a performance measure defining the success criterion
- the agent can process prior knowledge in addition to observations
- it can perform a sequence of best possible actions

PEAS

We can group the properties of an inteligent agent under the PEAS representation model:

P: Performance measure

E: Environment

A: Actuators

S: Sensors

Example

Self driving car:

- Performance: safety, time, legal drive, comfort
- Environment: roads, other vehicles, road signs, pedestrian
- Actuators: steering, accelerator, brake, signal, horn
- Camera, GPS, speedometer, odometer, accelerometer, sonar

Turing test

Alan Turing, MIND, 433, VOL. LIX. NO. 236. 1950

A machine passing the Turing test would be considered intelligent in a human like fashion.

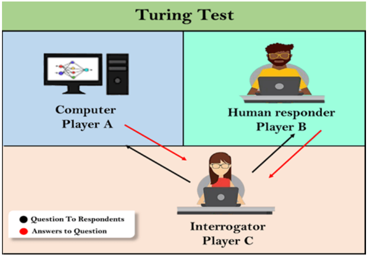

Figure: From https://www.javatpoint.com/agents-in-ai

Imitation game

- A computer (player A) and a human (player B) are placed in two different rooms.
- A human interrogator (player C) addresses each room with questions regarding any topic to which a human should be able to respond.
- If an interrogator would not be able to identify which is a machine and which is human, then the computer passes the test successfully.

The chinese room argument

Searle, John (1980), Behavioral and Brain Sciences, 3 (3): 417–457.

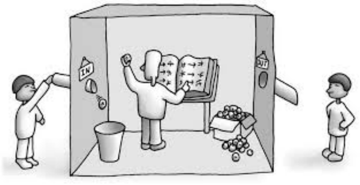

Figure: From Source: Wikicomms

- The person inside the room is provided a list of Chinese characters
- By using an instruction book explaining in detail the rules according to which strings (sequences) of characters may be formed, the person forms sentences.
- To the outside, it appears to be that the person in the room understands and speaks Chinese.
- BUT: The person doesn't understand any Chinese.

Strong AI
Searle, John (1980)

- **Strong AI:** "The appropriately programmed computer with the right inputs and outputs would thereby have a mind in exactly the same sense human beings have minds."
- According to Searl, there is a difference between
  - **simulating** a mind and
  - actually **having** a mind.

Strong AI

Searle, John (1980)

- **Strong AI:** "The appropriately programmed computer with the right inputs and outputs would thereby have a mind in exactly the same sense human beings have minds."
- According to Searl, there is a difference between
  - **simulating** a mind and
  - actually **having** a mind.

Is strong AI possible?

Two schools of thought in AI

Deduction and induction

$$\textbf{special} \underset{\text{deduction}}{\overset{\text{induction}}{\rightleftharpoons}} \textbf{general}$$

Induction

"I saw a couple of dogs yesterday. Both had four legs. Therefore, all dogs have four legs."

Two schools of thought in AI

Deduction and induction

$$\textbf{special} \underset{\text{deduction}}{\overset{\text{induction}}{\rightleftharpoons}} \textbf{general}$$

Induction

"I saw a couple of dogs yesterday. Both had four legs. Therefore, all dogs have four legs."

Deduction

"All canine animals have four legs. All dogs are canines. Therefore, dogs have four legs."

Deduction and induction

$$\textbf{special} \underset{\text{deduction}}{\overset{\text{induction}}{\rightleftharpoons}} \textbf{general}$$

Induction

- might make logic mistakes
- inductive **learning**
- statistical approaches, machine learning

Deduction

- mathematically accurate conclusions
- deductive **reasoning**
- logic reasoning and search methods

Deduction and induction

$$\textbf{special} \underset{\text{deduction}}{\overset{\text{induction}}{\rightleftharpoons}} \textbf{general}$$

Induction

- Starts from examples (data)
- A **learning algorithm** is used to derive a model
- Reasoning about unseen examples requires **generalization**
- Approximations and probabilistic predictions

Deduction

- Starts with a **knowledge base** of assertions and hypotheses
- **Logical inferences** allow to reason about unknown facts
- Exact logical conclusions based on the knowledge base

An example

Spam filter for emails

Deductive

- Flag emails from blacklisted senders as spam
- Flag emails containing certain keywords or predefined word patterns as spam
- Uses a knowledge base of senders, keywords and word patterns

Inductive

- Flag spam by comparing email content with that of previous spam/no spam emails
- Uses a data base of emails labelled as spam/no spam and a machine learning algorithm

An example for a deduction

Medical expert system MYCIN (Shortliffe 2002)

- Knowledge base of bacteria and antibiotics, as well as a set of rules indicating their relationship
- Based on a physicians questions, it uses the knowledge base and the rules to make recommendations for specific patients

Advantages

- Recomendations are **explainable**
- Trustworthy, because the available experts knowledge (or hypotheses) about bacteria and antibiotics is represented in the knowledge base

Disadvantages

- Recommendations are limited by the available knowledge
- Will not work on unseen strains of bacteria

Example for an inductive system

OptAB (Wendland et al, 2024)

- Trained on data from patients with sepsis treated with different antibiotics
- Makes predictions about the disease course under different antibiotics treatments

Advantages

- Can recognize patterns in the data which were possibly unknown before
- Can generate new hypotheses
- Tested on unseen patients

Disadvantages

- Explainability is harder
- Mathematical proof that the predictions are always correct, based on available knowledge, is not possible

Summary

Deduction and Induction

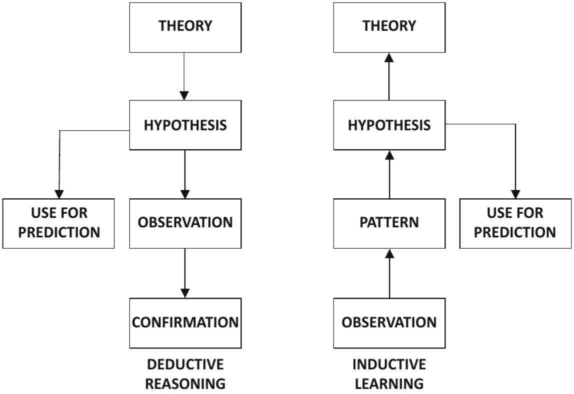

Figure 1.1: The two schools of thought in artificial intelligence

From Aggarwal, Artificial Intelligence, 2021, Fig.1, page 5

- Deductive reasoning: Symbolic AI
- Inductive reasoning: Subsymbolic AI (Machine Learning)

Chap. 2 Propositional logic

Chap. 3 First order logic

Chap. 4 Applications and limitations of logic in AI

Chap. 5 Probability theory and probabilistic logic

Chap. 6 Bayesian Networks

Chap. 7 Further Approaches