Introduction to AI and Autonomous Systems Exercises

Fredrik Heintz, Linkoping University ¨

November 16, 2020

1. What is your preferred definition of artificial intelligence and why?

My preferred definition of artificial intelligence (AI) is the one proposed by the European Commission and publicised in 2018:

> *"Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.*
>
> *AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications)" (European Comission 2019).*

I prefer this definition as it in a concise way describes important aspects of what defines AI systems such as goal-directed interaction with the environment and autonomy. In addition it also provides concrete examples of what technologies may constitute AI-systems.

2. What is an intelligent agent? What are important types of intelligent agents? Give some examples of intelligent agents. Can all aspects of AI be encapsulated in the concept of intelligent agents? Are all intelligent systems also autonomous systems? Motivate.

An agent perceives its environment through sensors and uses the sensory information to perform actions upon its environment through actuators. (Russell and Norvig 2003) describe different classes of intelligent agents based on their level of intelligence and capability:

1. **simple reflex agents**:
    - Can react to stimuli in the environment but has no internal state (memory of previous stimuli, states or performed actions)
    - example: robotic lawn mower operating in infinite loop
2. **model-based reflex agents**
    - Has a limited internal state. The action at time t+1 is dependent on the state and sensory input at time t.
    - Example: a self-steering system that relies on previous information

that is not deductible simply through current perceptory information.

3. **goal-based agents**
    - Has a more developed internal state. Can reason, deduce and plan actions to act towards a goal.
    - Example: robot that uses a search strategy to reach a destination.
4. **utility-based agents**
    - Can perform more advanced reasoning while assessing tradeoffs and conflicting goals.
    - example: route recommendation system that assesses different routes and takes into consideration where traffic jams are likely to occur.
5. **learning agents**
    - Can learn through interaction with the environment using feedback to alter the actions taken by actuators. Learning allows the agent to operation in unknown environments and improve over time.
    - Autonomous lawn mower that creates an optimal cutting strategy of a garden through trial and error.

I would argue that artificial intelligence can be encapsulated in the concept of intelligent agents. An implementation of artificial intelligence is often a system of functionalities, decision modules, perceptors and actuators interacting through interfaces. Such autonomous systems are also considered intelligent agents. However, not all intelligent agents can be considered artificial intelligence. Using the EU-definition of artificial intelligence, which describes goal-orientation, autonomy and interaction with the environment, level 1 and level 2 of Russel and Norvig's classification of intelligent agents can't be considered artificial intelligence.

3. Search is a key part of AI. Why? What are some common search strategies and what are their strengths and weaknesses?

Search is a common problem-solving method for intelligent agents that aim to reach a desired goal state. A search problem consists of a state space (set of possible states), a start space and a desired end state. The solution to a search problem is a plan, a sequence of actions that transforms the start state to the goal state. This structured way of defining and solving a problem suits well to machine implementation in the form of artificial intelligence.

Two main classes of search strategy classes are uninformed search (blind search) and informed search (heuristic search).

<u>Uninformed search</u>

Uninformed search algorithms are flexible in the sense that they are agnostic to the problem that is being solved. They can easily be applied to new domains and are efficient when applied to many different problems. Not being specific to the problem domain can however also lead to inefficiencies in terms of memory and time.

Informed search

Informed search methods use domain knowledge to guide the search process to make it more efficient. Heuristic methods are suitable for tougher problems that could not be solved by uninformed search methods in an efficient manner. Disadvantages with informed search methods include that they are not guaranteed to find a solution for all problems and that they are prone to combinatorial explosions.

4. What is heuristic search? Why are they important? Give some examples of heuristics and heuristic search methods.

Heuristic search methods (informed search) use outside information to guide the search process by ranking alternatives using a heuristic function at non-goal states by how promising they seem. Heuristic search methods when applied correctly can be more efficient than classic methods. Finding an approximate solution is often sufficient for many applications. Heuristic methods trade completeness, optimality or accuracy for computational speed.

Heuristic search methods include greedy search and A* search.

The travelling salesman problem (TSP) is a famous NP-hard problem in computer science which asks the question: "Given a list of cities and the distances between each pair of cities, what is the shortest possible route that visits each city exactly once and returns to the origin city?". A heuristic for TSP is using a greedy method which repeatedly selects the shortest edge between cities (Nilsson 2003). As with other heuristics this method is not guaranteed to find the optimal solution.

5. What is the physical-symbol system hypothesis? Do you believe in it? Why/Why not?

The hypothesis states that: “A physical symbol system has the necessary and sufficient means for general intelligent action” (Simon and Herbert 1976).

Induced by Newell and Simon's hypothesis is the conclusion that human thinking, capable of producing intelligent action, is an example of a physical symbol system. The hypothesis also implies that machines can be intelligent because a symbol system is sufficient for intelligence.

While I on a theoretical level agree with the hypothesis, actually determining the symbols for specific functions is an impossible task. Probability based connectionist approaches show more promise in many fields of artificial intelligence.

6. What is inductive, deductive and abductive reasoning? Give some examples of each.

Deductive reasoning uses one or more premises to reach a certain logical conclusion.

Example: if x=3 and y=2 then 3x+y=11

Inductive reasoning uses one or more premises that provide evidence towards a likely conclusion.

Example: 95% of students graduate high school. Adam goes to high school. It is therefore likely that Adam will graduate high school.

Abductive reasoning like inductive reasoning determines the most likely conclusion from a set of premises. Abductive reasoning is more similar to guesswork and is more typically applied when formulating a hypothesis for further testing rather than immediately drawing conclusions based on the data currently at hand.

Example: Almost all Swedish students graduate from high school. Adam has graduated from high school. Therefore Adam is Swedish.

7. Dealing with uncertainty is a key problem in AI. Why? What are some approaches?

The information provided to an artificial intelligence system from the outside environment is often messy, noisy and unpredictable. There can also be a lack of data causing decisions to have to be made on incomplete information. Probability based AI implementations are required to function in this unpredictable environment, where different courses of actions are evaluated and the action with the most favourable expected outcome is selected.

Quantifying the uncertainty can play an important role in dealing with uncertainty. Also understanding which variables contribute to the uncertainty or expected outcome is valuable in uncertain environments. Bayesian machine learning is a field within machine learning which aims to provide insight into how models are operating and why they reach certain conclusions or recommendations.

A method for building AI implementations that are apt at handling uncertainty is to train the AI system on noisy data. Another method for handling uncertainty is to design the AI-system to be cautious in new, uncertain situations and to favour risk minimizing courses of action.

8. What is machine learning? What are the main types of machine learning approaches? Give some examples of when each approach is suitable.

Machine learning are computer algorithms that can automatically learn and improve automatically through experience. Machine learning algorithms are not explicitly programmed and instead use training data that represent the problem domain to make predictions or decisions on new unseen data.

There are two main classes of machine learning, supervised and unsupervised learning algorithms. The main difference is that that supervised algorithms use labelled data for training. The aim of supervised learning is to learn a function that, given training data and desired outputs, best approximates the relationship between input and output. Supervised learning is applied typically for classification when inputs are mapped to a set of output labels or regression, when input is mapped to a continuous output value.

Unsupervised learning does not have/require labeled data and its goal is to deduce structure or patterns with a given set of data. A common supervised learning application is clustering.

Both for supervised and unsupervised machine learning problems there are a host of different algorithms such as linear and logistic regression, tree based models, naive bayes and neural networks. Factors that determine suitability include the complexity of the problem, amount of data available and the need for explainability.

9. What is a GAN, how do they work in principle (no technical details necessary) and why are they interesting?

Generative adversarial networks (GANs) can be used to generate new data points from a modelled distribution of the training data set. It is an unsupervised method using deep learning architectures, including convolutional neural networks. GAN's consist of two main sub-models: a generator model that generates new data examples and a discriminator which attempts to classify data examples as either real (actual data examples) or generated data examples. The two models are trained simultaneously until the discriminator is correct in half of all cases, indicating that the generator model produces realistic data examples.

GANs were first proposed by Ian Goodfellow et al in 2014. It is a rapidly evolving field with displaying promise in a multiple of domains including image generation and translation.

10. What is game theory? Give some examples of interesting problems that can be addressed by game theory.

Game theory attempts to mathematically model the interaction among rational decision-makers. In the field of artificial intelligence, game theory can inform the actions of intelligent agents and the expected behaviours of other intelligent agents in multi-agent systems.

Game theory has been applied to a vast number of domains including economics, political governance, biology and psychology. Examples of applications include modelling price formation in an auction setting or the fair division of resources e.g. in an inheritance dispute.

11. Many reasoning and learning approaches translate problems into optimization

problems. Give some examples and discuss pros and cons of this.

Essentially all supervised machine learning algorithms boil down to solving an optimization problem. Translating problems into mathematical optimization problems creates well defined problems with clear evaluation metrics that are well suited for machine computation. The optimized metric can be dependent on the domain or the type of problems (e.g. accuracy for classification or mean squared error for regression). At the same time it is possible to separate the definition of the problem from the methods used for solving it. The same learning algorithms can thus be applied to a multitude of domains.

Using a training set that is not representative of the actual data is a common error in learning algorithms. The ai system has then been optimized on non-representative data. Choosing a representative data set is not always trivial. A common mistake is to not use training samples that are as messy or noisy as actual data.

Optimizing on an evaluation metric without reviewing what lies behind the ai systems reasoning can have precarious consequences. This is especially relevant for black blox learning algorithms where there is a lack of explainability for why the AI arrives at a specific decision. In 2014, Amazon scrapped a recruiting tool that showed bias against women. Because the roles that were being recruited for were male dominated the algorithm taught itself to prefer male applicants and downgraded applications that referred to words such as “women’s” (Dastin 2018).

12. What are the most interesting and important open problems in AI according to you?

One of the most important problems is how to build AI systems that take decisions or produce recommendations that humans can trust and depend on. This is especially considering advancement in blackbox deep learning models that can perform advanced functions but can’t explain what reasoning lies behind its decisions.

Another problem that can be linked to the issue of trust towards AI systems is the lack of knowledge about AI within the general public and many businesses. By learning the basics about how AI systems function, this increased understanding can lead to increased trust. Within business and organizations, understanding the limitations and potential of AI can lead to more widespread implementation that help people.

Data is a very important factor in AI. Large informative datasets are necessary to build advanced intelligence. Combining this fact with privacy and integrity concerns regarding the collection of personal data is a key challenge for many AI applications.

Bibliography

Dastin, Jeffrey. 2018. “Amazon scraps secret AI recruiting tool that showed bias against women.” *Reuters*, October 11, 2018. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G.

European Comission. 2019. “A definition of Artificial Intelligence.” European Commision. https://ec.europa.eu/digital-single-market/en/news/definition-artificial-intelligence-main-capabilities-and-scientific-disciplines.

Nilsson, Christian. 2003. “Heuristics for the Traveling Salesman Problem.” http://160592857366.free.fr/joe/ebooks/ShareData/Heuristics%20for%20the%20Traveling%20Salesman%20Problem%20By%20Christian%20Nillson.pdf.

Russell, Stuart J., and Peter Norvig. 2003. *Artificial Intelligence: A Modern Approach (2nd ed.)*.

Simon, Newell, and Simon A. Herbert. 1976. “Computer Science as Empirical Inquiry: Symbols and Search.” *Commun. ACM*, 113-126.