Lecture #25: Artificial Intelligence and Machine Learning

CS106E Spring 2018, Payette & Lu

In this lecture, we study Artificial Intelligence and Machine Learning. We start by defining and looking at the history of Artificial Intelligence. We explore the technological advances that allowed the recent renaissance in the field, and then some of the common types of AI systems out there in the wild. We contrast the value of these AI systems with some of the fears (some possibly illogical and some very real). We also examine some of the ethical debates surrounding this topic, debates that are and will consider to be at the forefront of the conversations surrounding AI in the near future.

We then go into a discussion on the technique behind most modern AI systems: Machine Learning. We cover some of the basic machine learning methods, state of the art machine learning models (neural networks) and some of the constraints of machine learning.

What is Artificial Intelligence

- There is not a widely accepted definition of artificial intelligence.
    - My favorite definition is that AI is the ability of a computer to do tasks that generally require human intelligence.
    - Another definition is a computer system that is able to do something that it was not explicitly programmed to do
- In general, most definitions are centered around the idea computers simulate human intelligence
- AI is generally considered to be behavior based rather than implementation based
    - In other words, it doesn't matter what's going on inside the computer, it just matters what the computer is saying/doing
- To understand why the definition of AI is so broad and fluid, we need to look at the history of AI

History of Artificial Intelligence

- The modern history of AI probably begins with Alan Turing in 1936.
    - Some argue that his turing machine, the predecessor to the modern computer, was the first instance of artificial Intelligence
    - Turing also coined the term "Turing Test"
    - The test was simple - can this computer system fool a human into thinking it's a human (not a machine).
    - In his mind, this was the gold standard for "intelligent behavior."
    - The Turing Test is still a popular test today, although many question its validity as it only tests external behavior
- In 1955, John McCarthy and some colleagues coined the term "artificial intelligence"
    - They are credited with creating the field.
- Early AI systems were focused on "general intelligence"

    - In 1966, the ELIZA chatbot was able to pass the Turing Test***
        - While it was able to fool some humans, most were not fooled
    - Systems were narrow in scope, not robust as each behavior had to be programmed explicitly
        - Most would fail miserably outside a very specific use-case
- Funding dropped off in the early 70s as researchers and the government got discouraged, and 1974-1980 became known as the "AI Winter"
- The Japanese Government funded a major AI project in 1980, but it was a total failure and the field was quiet again until 1993.

Modern Artificial Intelligence

- AI field has seen a renaissance in the past few decades
- Computer processors have sped up, memory volume has increased, and data processing algorithms have improved
- The adoption of the GPU for highly optimized ML tasks has made analysis of massive datasets possible
- The widespread adoption of Machine Learning has formed the backbone of almost every Artificial Intelligence system.

Artificial Intelligence Subfields

- AI is **EVERYWHERE**
    - **Machine Translation**
        - Google Translate
    - **Spam Filters**
    - **Digital Personal Assistants**
        - Siri
        - Google Assistant
        - Cortana
        - Alexa
    - **Game players**
        - DeepBlue
        - AlphaGo
        - "The Computer" in video games
    - **Speech Recognition Systems**
        - IBM
        - Dragon
    - **Image Recognitions Systems**
    - **Algorithmic Trading Systems**
        - Black-Scholes Model (Caused crash in 1987)
        - Automated Trading Services
    - **Recommender Systems**
        - Amazon's Suggestions
        - Google Ads

- **Autonomous Vehicles**
    - Self-Driving Cars

The Positives of AI

- AI systems can be helpful
- Digital Personal Assistants facilitate a number of every day tasks
- Spam filters reduce the number of successful phishing attacks
- Machine translation has helped information flow around the world
- Banks use AI systems to identify fraudulent credit card charges

Fear Over AI

- Despite the potential positives of AI, people still fear AI
    - Elon Musk says that AI is humanity's biggest existential threat
    - Most portrayals of General AI in pop culture capitalize on this idea
        - HAL
        - The Hosts (HBO's Westworld)
        - The Matrix
        - Ex Machina
    - In my opinion, we're still a ways away from this
        - Most of the resources are not going to general AI, but rather to domain specific tasks
        - Some people think the more likely danger is something like "The Paper-Clip Maximizer" a simple machine that malfunctions and uses all of earth's resources to make paper clips

Ethical Considerations of AI

- On top of the fear of AI, there are a number of legitimate ethical considerations that still need to be worked out
    - Privacy concerns
        - Alexa recording conversations
    - AI systems can exhibit bias found in the datasets used to train them
        - Racist/Sexist classification systems
    - Self Driving Cars
        - Who is at fault for an accident involving a self-driving car?
    - Algorithmic Trading Failures
        - Who is liable for a rogue trading system?
    - The whole point is that AI can do things that the programmer didn't explicitly program.
        - So what types of decisions do we allow them to make?
            - Trolley problem
            - Weaponry
        - How do we draw these lines?
        - Do we allow systems to update themselves?
    - Does the programmer have a responsibility to limit the intelligence of his/her system?

AI Frontiers/Current Work/Future Work

- Robotics
    - Natural Movement
    - Environment Understanding
- Natural Language Understanding/Processing
    - Bias Detection
    - Summary Systems
- Content Generation
    - Image and Caption Generation
    - Google's DeepMind Style Transfer

Machine Learning

Machine learning is concerned with algorithms which train a machine learning model to learn how to perform tasks using data rather than hand-coded rules. These tasks often involve classification (i.e. determining what's in a picture), prediction (i.e. which Netflix shows is this user most likely to watch), decision making (i.e. should this autonomous car turn), or data generation (i.e. human speech synthesis).

Machine learning data most frequently takes the form of input-label pairs (*x, y*) where *x* is the input to a machine learning model and *y* is the label or expected output. *x* is typically a multi-dimensional vector. Each element of the input vector is called a *feature*. For example, for an image classification problem, *x* would be an image bitmap with RGB values and *y* the content of the image (i.e. "cat"). In a machine translation problem, *x* might be a sentence in English and *y* a sentence in Spanish.

Data is often split into three partitions: training data, validation/development data, and testing data. Training data is used to train the model, validation data is used to evaluate the performance of the model on previously unseen data for model tuning purposes, and testing data is used for a final performance evaluation with completely new data. In cases where there is not a lot of data, there is usually just a training/testing split, with validation performed via cross-validation on the training data.

<u>Machine Learning Models</u>

At a high level, a machine learning model can be thought of as a parameterized function $\hat{y} = f(x, \theta)$ where *x* is the input data, θ is a set of parameters that varies from model to model, and $\hat{y}$ is the predicted output. For example, in a simple line-fitting model, the parameters would be the slope and intercept of the line. In a neural network, the parameters are all the weights of the network. The goal of machine learning then is to find θ such that $f(x, \theta)$ outputs the desired result *y*. The problem is that θ is often high-dimensional and continuous, which makes exhaustive search hard. Instead, we iteratively calculate θ using optimization techniques to minimize the error between the predicted labels and the true labels.

Classic machine learning models include regression models, support vector machines, and Bayesian models. Choosing a model involves considering a number of trade-offs including running time, amount of data required, and performance of the model. In addition, models often make assumptions about the underlying structure of the data, which can impact the performance of the model if the assumptions are not accurate. For example, the naive Bayes classifier assumes that the input features are independent from each other.

Neural Networks

Neural networks are the sledgehammers of machine learning. They are immensely powerful, but also require a lot of computing power and training data. Neural networks only became feasible in the past few years because of the development of more powerful hardware and the rise of big data. Neural networks are inspired by the firing mechanism of biological neurons in which a neuron only fires after the combination of its inputs reaches some threshold. However, where neurons in a brain are arranged chaotically with connections all over the place, neurons in a neural network are typically arranged in a sequence of layers, where all the neurons in a layer are connected to all the neurons in the next layer. This arrangement is known as a feed-forward network or a fully-connected network. The "firing" of a neuron is calculated by taking a weighted sum over the inputs to the neuron and then applying a nonlinear function to the weighted sum.

Other arrangement of neural networks exist for different applications. Image processing tasks often involve a convolutional network, in which activations are calculated by sliding a n x n convolutional filter across the 2D image. This has the advantage of preserving spatial information, which is often a large part of image processing. Language processing tasks, such as machine translation and natural language understanding, use recurrent neural networks in which the network maintains an internal state that is updated as the network processes an input sequence. This allows the network to preserve temporal information, such as the ordering of words in a sentence.

Constraints of Machine Learning

Machine learning is well-suited for applications in which there is an abundance of representative data, the task has well-defined inputs and outputs, and there is a quantifiable way to determine the error of the model's predictions. If any of these conditions are not met, the performance of the model will suffer.

Representative data means that the data the model is trained and evaluated on is similar to the data that it will be seeing out in the wild. For example, a speech recognition system trained on American English would likely struggle if faced with a heavy Scottish accent. An image classification system trained to recognize objects in photographs would perform very poorly on hand-drawn pictures. Gathering sufficient data is often one of the most time-consuming and expensive parts of machine learning, as a bad dataset will inevitably produce a bad model.

A well-defined task means that the structure of the inputs and outputs are known. For example, in an autonomous driving situation, the inputs are the sensor readings and the outputs are the controls of the vehicle. In an image localization problem, the inputs are images and the outputs are bounding boxes around objects of interest. In contrast, a task such as "Decide governmental policy." is not well-defined, as the space of possible policies to enact is both vague and infinitely large. Another challenge in machine learning is the formalization of general problems into specific well-defined tasks.

Finally, quantifiable error is important because machine learning models are trained via optimization on some error metric. For example, in regression, this error is the difference between the predicted value and the true value. In classification, a variety of error metrics exist, but all of them involve penalizing choosing the wrong class. On the other hand, tasks which involve subjective judgement or creativity are often not quantifiable. For example, given task “Draw a picture,” how does one determine how “picture-y” the output is?