Mason Cushing

CS4375.004

Professor Karen Mazidi

January 28, 2023

Posted at https://github.com/mta825/Cushing_Portfolio_CS4375

Overview of ML

Machine learning is training computers to find patterns in data and using them to accomplish some informational goal. Its goal is to inform the user about the information they are giving it based on its attributes and its goal. An example of machine learning is giving an algorithm a set of pictures and telling it what is in these pictures (perhaps half have a cat and half have a dog). If done well, the algorithm when given a new image it has never seen before with a cat or a dog in it, it will be able to identify which it is.

Data, pattern recognition, and accuracy are vital to how machine learning works. In order to train a machine learning algorithm, massive amounts of data is required. Data is the very lifeblood on which machine learning acts and has its purpose. Pattern recognition and accuracy are part of the end goals of machine learning. In our example above, recognizing a pattern in the images allows the algorithm to correctly distinguish between cats and dogs, and if the algorithm does a good job, it has a high accuracy. How high that accuracy is depends on the data (specifically the quantity and quality of that data) given to the algorithm to train it to accomplish its purpose.

Artificial intelligence is closely related to machine learning, and those who don't understand the somewhat minute differences between the two often use the words interchangeably. Artificial

intelligence refers more to the autonomous agents that are created to accomplish some purpose. This would include entities like Stockfish, where it is using an algorithm (in this case searching for the best move along the game's search tree) to accomplish some goal for which it is being rewarded based on the creator's objective. Machine learning refers more to the algorithms behind some of these agents. An example of machine learning might be identifying which person is in a given image based on a surveillance camera and knowledge of who is being looked at. The algorithm then collects its own rules that it uses to accomplish its task when given a data point outside of its training data. It is in this way that the machine "learns", and that is what distinguishes artificial intelligence and machine learning.

Some additional examples of what machine learning follow. In recent news, the company OpenAI has produced ChatGPT based on their GPT (Generative Pre-trained Transformer) algorithm. This algorithm was trained on the massive amount of data available on the internet to predict which word comes next in a sequence of words. It then uses this machine learning algorithm to generate writings that often sound very humanlike on many topics when given a prompt.

Another example of machine learning is on many modern phones. Facial recognition is an application machine learning has been absolutely vital to. A machine learning algorithm, when given enough images (to which companies like Apple have a lot of access to) and a label of whether or not a person is in that image, can eventually start to accurately identify whether a person is in the image. In addition, these algorithms have improved to be able to identify who that person is (given a picture of that person beforehand of course).

An observation in machine learning is one data point. It might be only an image and any label data that is attached to an image that would be relevant to an algorithm. For example, it may be just the observation that a cat is in the picture, and then a picture that has a cat. A feature is

another piece of that observation. For our small example of the cat picture, that picture has a feature, and that feature is that it has a cat in it. Quantitative data is data that consists of numbers, measurements, and other strictly numerical data. For example, the encoded image is quantitative data (as the colors of the pixels are encoded in numbers). Qualitative data is data that consists of attributes, labels, and any other descriptor that doesn't have an objective numerical representation. For example, our label that tells our algorithm that the picture has a cat in it is qualitative data (even though that too would be encoded to the algorithm numerically). These terms and the concepts they represent are very important to machine learning, and they refer to core parts of data that machine learning processes.

I personally am interested in machine learning because of its relation to some of my most inspiring projects. My calculus professor recommended I watch the documentary on AlphaGo in my freshman year of high school, and I did so. I found it immensely interesting, and as I had already been coding in Python for some time before I watched this, I immediately began creating my own agent to play a different game (specifically Tic-Tac-Toe) from the ground up. After some considerable struggle and a week of algorithmic training (my algorithm was extremely inefficient as I was lacking in a number of concepts that would have greatly optimized the code) I managed to complete this agent to a reasonable level (that being that it was capable of forcing a draw in every scenario I was able to put it in). I am interested in machine learning because I want to be able to create projects like this for more complex applications like image classification, art generation, or music generation.