INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 1)
Available online at www.ijariit.com

Research on Machine Learning

ABSTRACT

In machine learning, a computer first learns to perform a task by studying a training set of examples. The computer then performs the same task with data it hasn't encountered before. This article presents a brief overview of machine-learning technologies. The article below contains an introduction about the machine learning, difference between machine learning and artificial intelligence and also seven steps of machine learning to better understand on how the machine learning works

Keywords: *Machine Learning, Artificial intelligence, Deep learning, Data mining, Reinforcement Learning, Neural Network, Natural Language Processing, Transfer Learning.*

1. INTRODUCTION

Machine learning is a field of computer science that gives computers the ability to learn without being explicitly programmed and by making the machine to learn from past experience A standard definition of machine learning is given by Tom Mitchell in 1997, A computer program is said to learn from experience 'E' with respect to some class of task 'T' and performance measure 'P', if its performance at tasks in 'T', as measured by 'P', improves with experience 'E'. Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses on prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with data mining, where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning. Machine learning can also be unsupervised and be used to learn and establish baseline behavioral profiles for various entities and then used to find meaningful anomalies.

Machine Learning (ML) is coming into its own, with a growing recognition that ML can play a key role in a wide range of critical applications, such as data mining, natural language processing, image recognition, and expert systems. ML provides potential solutions in all these domains and more and is set to be a pillar of our future civilization.

What is the Difference between Artificial Intelligence and Machine Learning?

Artificial Intelligence (AI) and Machine Learning (ML) are two very hot buzzwords right now, and often seem to be used interchangeably. They are not quite the same thing, but the perception that they are can sometimes lead to some confusion. So I thought it would be worth writing a piece to explain the difference.

Both terms crop up very frequently when the topic is Big Data, analytics, and the broader waves of technological change which are sweeping through our world.

In short, the best answer is that: Artificial Intelligence is the broader concept of machines being able to carry out tasks in a way that we would consider "smart" and, Machine Learning is a current application of AI based around the idea that we should really just be able to give machines access to data and let them learn for themselves.

Early Days

Artificial Intelligence has been around for a long time – the Greek myths contain stories of mechanical men designed to mimic our own behavior. Very early European computers were conceived as "logical machines" and by reproducing capabilities such as basic arithmetic and memory, engineers saw their job, fundamentally, as attempting to create mechanical brains.

As technology, and, importantly, our understanding of how our minds work, has progressed, our concept of what constitutes AI has changed. Rather than increasingly complex calculations, work in the field of AI concentrated on mimicking human decision making processes and carrying out tasks in ever more human ways.
Artificial Intelligence – devices designed to act intelligently – are often classified into one of two fundamental groups – applied or general. Applied AI is far more common – systems designed to intelligently trade stocks and shares, or manoeuvre an autonomous vehicle would fall into this category.

Generalized AIs – systems or devices which can, in theory, handle any task – are less common, but this is where some of the most exciting advancements are happening today. It is also the area that has led to the development of Machine Learning. Often referred to as a subset of AI, it's really more accurate to think of it as the current state-of-the-art.
Two important breakthroughs led to the emergence of Machine Learning as the vehicle which is driving AI development forward with the speed it currently has.

One of these was the realization – credited to Arthur Samuel in 1959 – that rather than teaching computers everything they need to know about the world and how to carry out tasks, it might be possible to teach them to learn for themselves.
The second, more recently, was the emergence of the internet, and the huge increase in the amount of digital information being generated, stored and made available for analysis.Once these innovations were in place, engineers realized that rather than teaching computers and machines how to do everything, it would be far more efficient to code them to think like human beings, and then plug them into the internet to give them access to all of the information in the world.

Neural Networks

The development of neural networks has been key to teaching computers to think and understand the world in the way we do while retaining the innate advantages they hold over us such as speed, accuracy, and lack of bias.
A Neural Network is a computer system designed to work by classifying information in the same way a human brain does. It can be taught to recognize, for example, images, and classify them according to elements they contain.

Essentially it works on a system of probability – based on data fed to it, it is able to make statements, decisions or predictions with a degree of certainty. The addition of a feedback loop enables "learning" – by sensing or being told whether its decisions are right or wrong, it modifies the approach it takes in the future. Machine Learning applications can read the text and work out whether the person who wrote it is making a complaint or offering congratulations. They can also listen to a piece of music, decide whether it is likely to make someone happy or sad, and find other pieces of music to match the mood. In some cases, they can even compose their own music expressing the same themes, or which they know is likely to be appreciated by the admirers of the original piece.
These are all possibilities offered by systems based around ML and neural networks. Thanks in no small part to science fiction, the idea has also emerged that we should be able to communicate and interact with electronic devices and digital information, as naturally as we would with another human being. To this end, another field of AI – Natural Language Processing (NLP) – has become a source of hugely exciting innovation in recent years, and one which is heavily reliant on ML.

NLP applications attempt to understand natural human communication, either written or spoken and communicate in return with us using similar, natural language. ML is used here to help machines understand the vast nuances in human language and to learn to respond in a way that a particular audience is likely to comprehend.
Among the different types of ML tasks, a crucial distinction is drawn between supervised and unsupervised learning:

- Supervised machine learning**:** The program is "trained" on a pre-defined set of "training examples", which then facilitate its ability to reach an accurate conclusion when given new data.
- Unsupervised machine learning**:** The program is given a bunch of data and must find patterns and relationships therein.

We will primarily focus on supervised learning here, but the end of the article includes a brief discussion of unsupervised learning with some links for those who are interested in pursuing the topic further.

Supervised Machine Learning

In the majority of supervised learning applications, the ultimate goal is to develop a finely tuned predictor function h(x) (sometimes called the "hypothesis"). "Learning" consists of using sophisticated mathematical algorithms to optimize this function so that, given input data x about a certain domain (say, square footage of a house), it will accurately predict some interesting value h(x) (say, the market price for the said house).
In practice, x almost always represents multiple data points. So, for example, a housing price predictor might take not only square-footage (x1) but also a number of bedrooms (x2), number of bathrooms (x3), number of floors (x4) , year built (x5), zip code (x6), and so forth. Determining which inputs to use is an important part of ML design. However, for the sake of explanation, it is easiest to assume a single input value is used.
So let's say our simple predictor has this form:

$$h(x) = \theta_0 + \theta_1 x$$

Where θ_0 and θ_1 are constants. Our goal is to find the perfect values of θ_0 and θ_1 to make our predictor work as well as possible.

Optimizing the predictor h(x) is done using **training examples**. For each training example, we have an input value x_train , for which a corresponding output, y , is known in advance. For each example, we find the difference between the known, correct value y , and our predicted value h(x_train) . With enough training examples, these differences give us a useful way to measure the "wrongness" of h(x). We can then tweak h(x) by tweaking the values of θ_0 and θ_1 to make it "less wrong". This process is repeated over and over until the system has converged on the best values for θ_0 and θ_1. In this way, the predictor becomes trained and is ready to do some real-world predicting.

The Seven Steps of Machine Learning:

- Gathering the data
- Preparing that data
- Choosing a model
- Training
- Evaluating
- Hyper parameter tuning
- Prediction

Let's pretend that we've been asked to create a system that answers the question of whether a drink is a wine or beer. This question answering system that we build is called a "model", and this model is created via a process called "training". The goal of training is to create an accurate model that answers our questions correctly most of the time. But in order to train a model, we need to collect data to train on. This is where we begin. If you are new to machine learning and want a quick overview first, check out this article before continuing:

Wine or Beer?

Our data will be collected from glasses of wine and beer. There are many aspects of the drinks that we *could* collect data on, everything from the amount of foam, to the shape of the glass.

For our purposes, we'll pick just two simple ones: The color (as a wavelength of light) and the alcohol content (as a percentage). The hope is that we can split our two types of drinks along these two factors alone. We'll call these our "*features"* from now on: color, and alcohol.

The first step to our process will be to run out to the local grocery store and buy up a bunch of different beers and wine, as well as get some equipment to do our measurements—a spectrometer for measuring the color, and a hydrometer to measure the alcohol content. Our grocery store has an electronics hardware section :)

Gathering Data

Once we have our equipment and booze, it's time for our first real step of machine learning: gathering data. This step is very important because the quality and quantity of data that you gather will directly determine how good your predictive model can be. In this case, the data we collect will be the color and the alcohol content of each drink.
This will yield a table of color, alcohol%, and whether it's beer or wine. This will be our training data.

Data Preparation

A few hours of measurements later, we have gathered our training data. Now it's time for the next step of machine learning: Data preparation, where we load our data into a suitable place and prepare it for use in our machine learning training.

We'll first put all our data together, and then randomize the ordering. We don't want the order of our data to affect what we learn since that's not part of determining whether a drink is a beer or wine. In other words, we make a determination of what a drink is, independent of what drink came before or after it.

This is also a good time to do any pertinent visualizations of your data, to help you see if there are any relevant relationships between different variables you can take advantage of, as well as show you if there are any data imbalances. For example, if we collected way more data points about beer than wine, the model we train will be biased toward guessing that virtually everything that it sees is beer since it would be right most of the time. However, in the real-world, the model may see beer and wine an equal amount, which would mean that guessing "beer" would be wrong half the time. We'll also need to split the data into two parts. The first part, used in training our model, will be the majority of the dataset. The second part will be used for evaluating our trained model's performance. We don't want to use the same data that the model was trained on for evaluation since it could then just memorize the "questions", just as you wouldn't use the same questions from your math homework on the exam.

Sometimes the data we collect needs other forms of adjusting and manipulation. Things like de-duping, normalization, error correction, and more. These would all happen at the data preparation step. In our case, we don't have any further data preparation needs, so let's move forward.

Choosing a Model

The next step in our workflow is choosing a model. There are many models that researchers and data scientists have created over the years. Some are very well suited for image data, others for sequences (like text, or music), some for numerical data, others for text-based data. In our case, since we only have 2 features, color, and alcohol%, we can use a small linear model, which is a fairly simple one that should get the job done.

Training

Now we move onto what is often considered the bulk of machine learning—the training. In this step, we will use our data to incrementally improve our model's ability to predict whether a given drink is wine or beer.

$$\underset{\text{OUTPUT}}{y} = \underset{\text{SLOPE}}{m} * \underset{\text{INPUT}}{x} + \underset{\text{Y-INTERCEPT}}{b}$$

We will do this on a much smaller scale with our drinks. In particular, the formula for a straight line is y=m*x+b, where x is the input, m is the slope of that line, b is the y-intercept, and y is the value of the line at the position x. The values we have available to us for adjusting, or "training", are m and b. There is no other way to affect the position of the line since the only other variables are x, our input, and y, our output.

In some ways, this is similar to someone first learning to drive. At first, they don't know how any of the pedals, knobs, and switches work, or when any of them should be used. However, after lots of practice and correcting for their mistakes, a licensed driver emerges. Moreover, after a year of driving, they've become quite adept. The act of driving and reacting to real-world data has adapted their driving abilities, honing their skills.

$$\text{WEIGHTS} = \begin{bmatrix} m_{1,1} & m_{1,2} \\ m_{2,1} & m_{2,2} \\ m_{3,1} & m_{3,2} \end{bmatrix}$$

$$\text{BIASES} = \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \\ b_{3,1} & b_{3,2} \end{bmatrix}$$

In machine learning, there are many m's since there may be many features. The collection of these m values is usually formed into a matrix, that we will denote W, for the "weights" matrix. Similarly, for b, we arrange them together and call that the biases.

The training process involves initializing some random values for W and b and attempting to predict the output with those values. As you might imagine, it does pretty poorly. But we can compare our model's predictions with the output that it should produce, and adjust the values in W and b such that we will have more correct predictions.

This process then repeats. Each iteration or cycle of updating the weights and biases is called one training "step".

Let's look at what that means in this case, more concretely, for our dataset. When we first start the training, it's like we drew a random line through the data. Then as each step of the training progresses, the line moves, step by step, closer to an ideal separation of the wine and beer.

Evaluation

Once training is complete, it's time to see if the model is any good, using Evaluation. This is where that dataset that we set aside earlier comes into play. Evaluation allows us to test our model against data that has never been used for training. This metric allows us to see how the model might perform against data that it has not yet seen. This is meant to be representative of how the model might perform in the real world.

A good rule of thumb I use for a training-evaluation split somewhere on the order of 80/20 or 70/30. Much of this depends on the size of the original source dataset. If you have a lot of data, perhaps you don't need as big of a fraction for the evaluation dataset.

Parameter Tuning

Once you've done an evaluation, it's possible that you want to see if you can further improve your training in any way. We can do this by tuning our parameters. There were a few parameters we implicitly assumed when we did our training, and now is a good time to go back and test those assumptions and try other values.

One example is how many times we run through the training dataset during training. What I mean by that is we can "show" the model our full dataset multiple times, rather than just once. This can sometimes lead to higher accuracies.

Another parameter is "*learning rate*". This defines how far we shift the line during each step, based on the information from the previous training step. These values all play a role in how accurate our model can become, and how long the training takes.

For more complex models, initial conditions can play a significant role in determining the outcome of training. Differences can be seen depending on whether a model starts off training with values initialized to zeroes versus some distribution of values, which leads to the question of which distribution to use.

The potentially long journey of parameter tuning

As you can see there are many considerations at this phase of training, and it's important that you define what makes a model "good enough", otherwise you might find yourself tweaking parameters for a very long time.

These parameters are typically referred to as "*hyper parameters*". The adjustment, or tuning, of these hyper parameters, remains a bit of an art and is more of an experimental process that heavily depends on the specifics of your dataset, model, and training process.

Once you're happy with your training and hyper parameters, guided by the evaluation step, it's time to finally use your model to do something useful!

Prediction

Machine learning is using data to answer questions. So Prediction, or inference, is the step where we get to answer some questions. This is the point of all this work, where the value of machine learning is realized.

We can finally use our model to predict whether a given drink is wine or beer, given its color and alcohol percentage.

2. CONCLUSION

Machine learning approaches applied in systematic reviews of complex research fields such as quality improvement may assist in the title and abstract inclusion screening process. Machine learning approaches are of particular interest considering steadily increasing search outputs and accessibility of the existing evidence is a particular challenge of the research field quality improvement. Increased reviewer agreement appeared to be associated with improved predictive performance.

3. REFERENCES

[1]. Reference 1 Wikipedia

