Machine Learning and its application: A Succinct Study

[1]Amit Kumar Sharma, [2]Tirth Raj Kumar, [3]Vikash Kumar Soni

[1,2,3]Department of IT, C.V Raman College of Engineering, Bhubaneswar
amitsharmanhes@gmail.com[1], stirthraj.12345@gmail.com[2], vikashkrsoni007@gmail.com[3]

Abstract— Machine learning is an art of computer programming that gives the computer the ability to learn from itself by the help of algorithm given by the users. It is based on the predictive statistical algorithm such as KNN and regression. .If you want a machine learning to solve some problem you must first train the machine using training data. Training data are past example showing the machine how to solve particular problem

The primary aim is to allow the computers learn from experience. Machine learning is about making computer to act without explicit programming them .It figure out how to perform task based on generalizing from data or example and they can learn to improve themselves from experiences. It is used by the business industry, science, government to solve important task.

***Index Terms*— Machine Learning, Explicit Programmed, Predictive, algorithm.**

I. INTRODUCTION

Suppose its Saturday night at home and you are just ordering a pizza and then placed your smartphone down on the table. Within a second your friend message you, who want to hang out at your home tonight. But your friend does not have a car. If your friends is to come over, then you will need to drive over to his house to collect him .This normally would not be a problem, as you could wait for a pizza, have a few slices, and then leave to collect your friend from his house. But tonight that would mean missing the start of an important sports game televised live on television. One the match begin, you won't be leaving the living room for anyone. Thus, the only available option is to collect your friend first and attempt to return back in time to receive the pizza.

You need to make a quick decision. Do I have enough time to pick up my friend and get home before the pizza arrives?

The pizza is estimated to arrive within 30 minutes, and if you leave now, then you should have time to collect your friend and be back home within 30 minutes. As you know the route to your friend's house, you can predict the time needs. Now Let's assume that you have ordered pizza previous occasions and the delivery time exceeded the estimated arrival time more than 10 minutes then there is approximately 50 percent chances that you will have time to collect your friend and return in time. Yours confidence level is now greater than 70 percent. Based on previous experience, your predication rate of 50% does not meet your confidence threshold and you should stay at home and figure out an excuse to decline yours friend's invitation. Basing one's decision on existing data is known as the empirical method.

Predictive models in machine learning consists of at least two variable. One variable is the result you wish to predict, known as the dependent variable(y). In this example the dependent variable is whether the pizza delivery will be greater or equal to 10 minutes late. The second variable is independent variable(x), which again predicts whether the pizza will be late but in relation to an independent event. In our scenario "day of the week" could be an independent variable. Machine learning then tells how x(the independent variable)affects y(the dependent variable).

For decades' Machine operated only by responding to direct user commands. In other words' Computer were designed to perform set tasks in response to pre-programmed commands. Now, Computer don't strictly need to receive an 'input command' to perform tasks but rather 'input data'. Specifically, the machine creates a predictive model based on previous experience captured in the data. From the input of data, the machine is then able to formulate decision on how, where and when to perform a certain action. In this case of our earlier example, the model was predicting whether the pizza delivery would be late based on existing data.

II. MANY DOMAINS AND APPLICATION

i. Medicine:

- Diagnose a disease

➢ Inputs are as Symptoms , lab measurements , test results , DNA tests etc

➢ Outputs are as either one of the above-mentioned or "none of the above-mentioned"

- Data mining historical medical records to learn which future patients will respond best to which treatments.

ii. Vision:

- Say what objects appear in an image
- Convert hand-written digits to character 0..9
- Detect where object appear in an image.

iii. Robot Control:

- Design autonomous mobile robots that learn to navigate from their own experiences.

iv. Financial:

- Predicts if a stock will rise or fall in the next few milliseconds
- Predict if a user will click on an advertisement or not in order to decide which advertisement to show.

III. TYPES OF MACHINE LEARNING

A. Supervised Learning

The only difference between supervised and unsupervised learning lies in the type of training data you use. In supervised learning you provide labelled training data. This means that each example you feed into your algorithm is classified into a recognizable data class or type. If this is the first time reading about labelled training data, it may appear confusing. Do not panic, an example will make everything clear once again:

Think of yourself trying to learn a new language, let's say Italian; your teacher gives you training data in form of vocabulary. This data is labelled because each new word you learn is classified into recognizable terms. For instance, your teacher tells you "ciao"means "hello"in English, she tells you "mangiare" means "eating", etc…Each example of a new Italian word has been associated into an English word you can understand. These are the fundamentals of supervised learning and machines work the exact same way.

B. Unsupervised Learning

As humans, unsupervised learning is a trickier concept to understand because we do it subconsciously. You can think of it as learning through observation.

Unsupervised learning is defined by unclassified training data. Essentially, you provide a machine with thousands of problems and their results, but you do not explain how the result is calculated. The algorithm then starts to look for things in common; it tries to identify shared traits and features between the problems and the solutions. With enough data, the algorithm can extracts a pattern and develop a strategy to solve any problem. Again, these concepts are best explained using practical examples:

For instance, let's jump back to how you can learn a new language –Italian. An example of how you can learn Italian using unsupervised learning is by watching movies. If you don't know any Italian and you watch an Italian movie you will not understand anything. Second movie –nothing, third movie –still nothing. After ten movies you might start to recognize common expressions, for instance every time a character says "Arrivederci" you observe someone leave and you conclude this means "goodbye".

If you watched thousands of movies, you would learn how to speak Italian perfectly using this technique, you would simply "pick it up". Clearly, you can't do that because it's impossible; it would take up too much time for you- but not for a computer. You could acquire millions of hours of Italian speech and run them all through a machine learning algorithm. After "listening and learning"to all these hours of Italian, the computer will have the meaning of the language perfectly.

IV. CLASSIFICATION VS REGRESSION

Before diving into the details and applications of each algorithm, I first want to discuss the differences between classification and regression – two key concepts in the field of machine learning.

Classification models are strongly favored over regression models? First, let's distinguish between classification and regression

Classification models: the results of a classification problem will fall into class labels (i.e. results fall into clearly defined boundaries/ categories/ groups). An example of this problem might be predicting the next colour of the roulette table. There are only three clearly-defined options: red, black or green. There is no intermediate answer, there is no half-way output; the answer can only be one of the three options – these are three class labels.

Regression models: the problem result can have a continuous value. An example of this problem is trying to predict the weight of an egg produced by a chicken farm. There are no defined choices or classes for the weight of an egg. An egg can weigh 54.3g, 87.3g or 62.9 – it is a continuous value. Now that I have defined the

terminology in question, we can return to the original question: why are classification models preferred over regression models within the industry? The reason for this bias links to the fundamental purpose of machine learning algorithms: helping users make a decision.

In fact - as machine learning engineers and data analysts –we are looking for well-structured answers that clearly tell us what decision to take. For instance, let's consider the following problem: "should I advertise on a given online platform?"A regression model would might reply "platform viewers are 13% likely to buy your product", whereas a classification model might reply "yes - the viewer's characteristics match your target audience". Which answer is more useful? Which helps us make decisions? Clearly, classification models help us make decisions.

Unfortunately, we live in a disorganized world and clearly-defined structures rarely come by. This means most the problems you will attempt will be regression problems, at least in nature. To solve this issue we tend to convert regression-based problems into classification problems by creating what I refer to as classification buckets. These are artificial boundaries that help data analysts organize and structure natural, messy and unclear data into manageable and organized data that is easy to read, understand and can help us make decisions.

For instance, predicting the conversion rate of a particular customer stream is an obvious regression problem –is it 1%, 5%, 8.3234%? But if you create classification buckets that breakdown the conversion rate into: 0-5% (or "low conversion"), 6-10% (or "medium conversion"), 11-15% (or "high conversion")…you find yourself dealing with a classification problem. The results you obtain are more structured, they are easier to read, interpret and understand. You can present these results to your colleagues or the marketing team and they will help you make real decisions!

V. CONCLUSION

In this paper we discussed about Machine Learning is nowadays essential in the field of computer science. It has domains in nearly every other field of study and is already being implemented in huge industry because Machine Learning can solve problems too difficult or time consuming for humans to implement. To describe machine learning in general terms, a variety of data set are used to learn patterns in data and make accurate predictions based on the patterns it observes.

REFERENCES

[1] http://seventablets.com/blog/what-are-the-pros-and-cons-of-machine-learning/ (dated:16/08/2018)

[2] https://en.wikipedia.org/wiki/Machine_learning (dated:16/08/2018)

[3] https://www.expertsystem.com/machine-learning-definition/ (dated:16/08/2018)

[4] Machine Learning for Absolute Beginners:A Plain English Introduction(First Edition) by Oliver Theobald

[5] Machine learning :Fumdamental Algorithm for Supervised and Unsupervised by Valerio Pellicciari

[6] Brojo Kishore Mishra, Binayak Padhy, Manoj Kumar Das, " Applications of Machine Learning in Web Mining", National Conference on Latest Trends & Innovations in Computing, Electronics & Communications (e-Trendz 2011), AISECT University, Bhopal, 17th – 18th Nov' 2011, ISBN: 978-93-81358-00-9, pp-19-23.

[7] Sushruta Mishra, Pamela Chaudhury, Brojo Kishore Mishra, Hrudaya Kumar Tripathy, "An implementation of Feature ranking using Machine learning techniques for Diabetes disease prediction", Proceeding ICTCS '16 Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies, Article No. 42 , ACM New York, NY, USA ©2016 , ISBN:978-1-4503-3962-9, doi:10.1145/2905055.2905100