Overview and Application of Machine Learning

Nishant Jethwa[1], Parth Joshi[2], Rohit Pandey[3]

[1,2,3]Student, Department of Computer Engineering, NMIMS Mukesh Patel School of Technology Management and Engineering, Mumbai, India

Abstract: **With all the technological advancement, the one that stands out is the field of Artificial Learning and Machine Learning, reason being machines can reduce the work of human but cannot make it smart i.e. it always requires timely intervention of humans to validate or operate the machines so that is why ML and AI comes into the picture it not helps the machines to learn the task itself but also takes off the burden from humans to manually do the task. So, this paper gives an overview on what is machine learning is and different types of algorithms it consists of and few of the applications of machine learning and algorithms used in those applications.**

Keywords: **NLP, supervised learning, unsupervised learning, Computer Vision, E-mail Filtering, Recommendation System.**

1. Introduction

The term Machine Learning was coined by Arthur Samuel in 1959, an American pioneer in the field of computer gaming and artificial intelligence and stated that "it gives computers the ability to learn without being explicitly programmed". And in 1997, Tom Mitchell gave a "well-posed" mathematical and relational definition that "A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

Suppose that you decide to check out that offer for a vacation. You browse through the travel agency website and search for a hotel. When you look at a specific hotel, just below the hotel description there is a section titled "You might also like these hotels". This is a common use case of Machine Learning called "Recommendation Engine". Again, many data points were used to train a model in order to predict what will be the best hotels to show you under that section, based on a lot of information they already know about you.

This paper talks about what Machine Learning is actually about, and it also covers the various learning approaches for machine learning like supervised, unsupervised, semi-supervised and reinforcement learning. This paper also talks about has some applications of machine learning. Every paper on overview an application of machine learning covers and focuses on single application but this paper will give user a brief about different applications with a table which gives the list of algorithms used in each of the application.

2. Structure of the papers

Section III gives the reader a brief idea about what Machine learning is and how it works. Section IV covers the different types of machine learning algorithms. Section V different applications of machine learning in various fields with a table which gives the list of algorithms used in each of the application.

3. Machine learning

A. What is Machine Learning

Machine Learning is a sub-area in Artificial Intelligence, whereby the term refers to the ability of IT systems to independently find solutions to problems by recognizing patterns in dataset. Machine Learning enables IT systems to recognize patterns on the basis of existing algorithms and data sets and to develop adequate solution concepts. Therefore, in Machine Learning, artificial knowledge is generated on the basis of experience. To enable the software to independently generate solutions, the required algorithms and data must be fed into the systems in advance and the respective analysis rules for the recognition of patterns in the data stock must be defined. Once these two steps have been completed, the system can perform the following tasks by Machine Learning:

- Finding, extracting and summarizing relevant data
- Making predictions based on the analysis data
- Calculating probabilities for specific results
- Adapting to certain developments autonomously
- Optimizing processes based on recognized patterns

B. How does Machine Learning work?

Machine Learning works in a similar way to human learning. For example, if a child is shown some pictures of a specific objects, they can learn to identify and differentiate between them. Machine Learning works in the same way, through data input and certain commands, the computer is enabled to "learn" to identify certain objects (persons, objects, etc.) and to distinguish between them. For this purpose, the software is supplied with data and trained. For instance, the programmer can tell the system that a particular object is a table (="table ") and another object is not a table (="not table"). The software receives continuous feedback from the programmer. These feedback signals are used by the algorithm to adapt and optimize the model. With each new data set fed into the system, the model is further optimized so that it can clearly distinguish between "humans" and "non-humans" in the end.

International Journal of Research in Engineering, Science and Management
Volume-2, Issue-11, November-2019
www.ijresm.com | ISSN (Online): 2581-5792

507

4. Types of learning in machine learning

Supervised Learning: Supervised learning is when the model is getting trained on a labelled dataset. Labelled dataset is one which have both input and output parameters. In this type of learning both training and validation datasets are labelled.

Types of Supervised Learning:

- *Classification:* It is a Supervised Learning task where output is having defined labels (discrete value). For example, in above Figure A, Output – Purchased has defined labels i.e. 0 or 1; 1 means the customer will purchase and 0 means that customer won't purchase. The goal here is to predict discrete values belonging to a particular class and evaluate on the basis of accuracy. It can be either binary or multi class classification. In binary classification, model predicts either 0 or 1; yes, or no but in case of multi class classification, model predicts more than one class.
  Example: Gmail classifies mails in more than one classes like social, promotions, updates, forum.
- *Regression:* It is a Supervised Learning task where output is having continuous value.
  Example in above Figure B, Output – Wind Speed is not having any discrete value but is continuous in the particular range. The goal here is to predict a value as much closer to actual output value as our model can and then evaluation is done by calculating error value. The smaller the error the greater the accuracy of our regression model.

Example of Supervised Learning Algorithms:

- Linear Regression
- Nearest Neighbor
- Guassian Naive Bayes
- Decision Trees
- Support Vector Machine (SVM)
- Random Forest

Unsupervised Learning: It's a type of learning where we don't give target to our model while training i.e. training model has only input parameter values. The model by itself has to find which way it can learn.

Types of Unsupervised Learning,

- *Clustering:* Broadly this technique is applied to group data based on different patterns, our machine model finds. For example, in above figure we are not given output parameter value, so this technique will be used to group clients based on the input parameters provided by our data.
- *Association:* This technique is a rule based ML technique which finds out some very useful relations between parameters of a large data set. For e.g. shopping stores use algorithms based on this technique to find out relationship between sale of one product w.r.t to others sale based on customer behavior. Once trained well, such models can be used to increase their sales by planning different offers.

Example of Unsupervised Learning Algorithms:

- K-Means Clustering
- DBSCAN – Density-Based Spatial Clustering of Applications with Noise
- BIRCH – Balanced Iterative Reducing and Clustering using Hierarchies
- Hierarchical Clustering

Semi-supervised Learning: As the name suggests, its working lies between Supervised and Unsupervised techniques. We use these techniques when we are dealing with a data which is a little bit labelled and rest large portion of it is unlabeled. We can use unsupervised technique to predict labels and then feed these labels to supervised techniques. This technique is mostly applicable in case of image data-sets where usually all images are not labelled.

Reinforcement Learning: In this technique, model keeps on increasing its performance using a Reward Feedback to learn the behavior or pattern. These algorithms are specific to a particular problem e.g. Google Self Driving car, AlphaGo where a bot competes with human and even itself to getting better and better performer of Go Game. Each time we feed in data, they learn and add the data to its knowledge that is training data. So, more it learns the better it gets trained and hence experienced.

- Agents observe input.
- Agent performs an action by making some decisions.
- After its performance, agent receives reward and accordingly reinforce and the model stores in state-action pair of information.

Some algorithms:

- Temporal Difference (TD)
- Q-Learning
- Deep Adversarial Networks

5. Applications of machine learning

A. Computer Vision

Computer vision is the study of enabling computers to understand and interpret visual information in a way similar to a human eye. Computer vision is closely linked with machine learning, as the computer must interpret what it sees, and then perform appropriate analysis or act accordingly. Computer vision tasks include methods for acquiring, processing, analysing and understanding digital images. These tasks are the most basic part of computer vision which are used to perform other bigger tasks like converting a 2D image into a 3D image, image classification, object localisation, object detection, image segmentation etc. During the past ten years, computer vision has grown from a research area to a widely accepted technology, capable of providing a dramatic increase in productivity and improving living standards. A computer vision system can be used to perform operations on videos or images and the system uses various algorithms.

International Journal of Research in Engineering, Science and Management
Volume-2, Issue-11, November-2019
www.ijresm.com | ISSN (Online): 2581-5792

508

B. E-Mail Filtering

Email filtering is the processing of email to organize it according to specified criteria. The term can apply to the intervention of human intelligence, but most often refers to the automatic processing of incoming messages with anti-spam techniques - to outgoing emails as well as those being received. In e-mail filtering task some features could be the bag of words or the subject line analysis. Thus, the input to e-mail classification task can be viewed as a two-dimensional matrix, whose axes are the messages and the features. E-mail classification tasks are often divided into several sub-tasks. First, Data collection and representation are mostly problem specific (i.e. e-mail messages), second, e-mail feature selection and feature reduction attempt to reduce the dimensionality (i.e. the number of features) for the remaining steps of the task. Finally, the e-mail classification phase of the process finds the actual mapping between training. Algorithms/Methods like Naïve Bayes classifier method, K-nearest neighbor classifier method, Artificial Neural Networks classifier method, Support Vector Machines classifier method etc.

C. Natural Language Processing (NLP)

Natural Language Processing is the technology used to aid computers to understand the human's natural language it can also be framed as ability of a computer program to understand human language as it is spoken. NLP is a component of artificial intelligence. The ultimate objective of NLP is to read, decipher, understand, and make sense of the human languages in a manner that is valuable. Syntactic and Semantic analysis are the two main techniques used to complete natural language processing. A typical interaction between humans and machines using Natural Language Processing could go as follows:

- A human talks to the machine
- The machine captures the audio.
- Calculating Audio to text conversion takes place
- Processing of the text's data
- Data to audio conversion takes place
- The machine responds to the human by playing the audio file.

Natural Language Processing is the driving force behind the following common applications:

- Language translation applications such as Google Translate
- Word Processors such as Microsoft Word and Grammarly that employ NLP to check grammatical accuracy of texts.
- Interactive Voice Response (IVR) applications used in call centers to respond to certain users' requests.
- Personal assistant applications such as OK Google, Siri, Cortana, and Alexa.

NLP entails applying algorithms to identify and extract the natural language rules such that the unstructured language data is converted into a form that computers can understand. When the text has been provided, the computer will utilize algorithms to extract meaning associated with every sentence and collect the essential data from them. Sometimes, the computer may fail to understand the meaning of a sentence well, leading to obscure results.

For example, a humorous incident occurred in the 1950s during the translation of some words between the English and the Russian languages.

Here is the biblical sentence that required translation:
"The spirit is willing, but the flesh is weak."
Here is the result when the sentence was translated to Russian and back to English:
"The vodka is good, but the meat is rotten."

D. Pattern Recognition

Pattern recognition is the process of recognizing patterns by using machine learning algorithm. Pattern recognition can be defined as the classification of data based on knowledge already gained or on statistical information extracted from patterns and/or their representation. One of the important aspects of the pattern recognition is its application potential.

Examples: Speech recognition, speaker identification, multimedia document recognition (MDR), automatic medical diagnosis.

In a typical pattern recognition application, the raw data is processed and converted into a form that is amenable for a machine to use. Pattern recognition involves classification and cluster of patterns.

- In classification, an appropriate class label is assigned to a pattern based on an abstraction that is generated using a set of training patterns or domain knowledge. Classification is used in supervised learning.
- Clustering generated a partition of the data which helps decision making, the specific decision-making activity of interest to us. Clustering is used in an unsupervised learning.

E. Training and Learning in Pattern Recognition:

Learning is a phenomenon through which a system gets trained and becomes adaptable to give result in an accurate manner. Learning is the most important phase as how well the system performs on the data provided to the system depends on which algorithms used on the data. Entire dataset is divided into two categories, one which is used in training the model i.e. Training set and the other that is used in testing the model after training, i.e. Testing set.

1. *Training set:* Training set is used to build a model. It consists of the set of images which are used to train the system. Training rules and algorithms used give relevant information on how to associate input data with output decision. The system is trained by applying these algorithms on the dataset, all the relevant information is extracted from the data and results are obtained. Generally, 80% of the data of the dataset is taken for training data.

International Journal of Research in Engineering, Science and Management
Volume-2, Issue-11, November-2019
www.ijresm.com | ISSN (Online): 2581-5792

509

Table 1
Algorithms

S. No.	Applications of Machine Learning	
	Name of Application	**Algorithms**
1.	Computer Vision	Naïve Bayes method, K-nearest neighbor method, Artificial Neural Networks classifier method, Support Vector Machines classifier method
2.	Email Filtering	Naïve Bayes method
3.	Narural Language Processing	Linear Regression, Neural Networks, Nearest Neighbor
4.	Recommender System	Linear Regression, k-means clustering
5.	Bioinformetics	Baum Welch **algorithm**. BLAST (biotechnology) Blast2GO. Bowtie (sequence analysis)

2. *Testing set:* Testing data is used to test the system. It is the set of data which is used to verify whether the system is producing the correct output after being trained or not. Generally, 20% of the data of the dataset is used for testing. Testing data is used to measure the accuracy of the system. Example: a system which identifies which category a particular flower belongs to, is able to identify seven category of flowers correctly out of ten and rest others wrong, then the accuracy is 70 %

6. Recommendation System

Recommender systems are one of the most successful and widespread application of machine learning technologies in business. We can apply recommender systems in scenarios where many users interact with many items. We can find large scale recommender systems in retail, video on demand, or music streaming. In order to develop and maintain such systems, a company typically needs a group of expensive data scientist and engineers. That is why even large corporates such as BBC decided to outsource its recommendation services.

A. Algorithms

Machine learning algorithms in recommender systems are typically classified into two categories — content based and collaborative filtering methods although modern recommenders combine both approaches. Content based methods are based on similarity of item attributes and collaborative methods calculate similarity from interactions.

Collaborative methods work with the interaction matrix that can also be called rating matrix in the rare case when users provide explicit rating of items. The task of machine learning is to learn a function that predicts utility of items to each user. Matrix is typically huge, very sparse and most of values are missing. The simplest algorithm computes cosine or correlation similarity of rows (users) or columns (items) and recommends items that k-nearest neighbors enjoyed.

B. Bioinformatics

Bioinformatics is the science of storing, retrieving and analyzing large amounts of biological information. It is a highly interdisciplinary field involving many different types of specialists, including biologists, molecular life scientists, computer scientists and mathematicians.

The table 1, mentions all the algorithms used for all the specific applications covered in the paper.

7. Conclusion

This paper presented an overview on application of machine learning.

References

[1] Charles Feng, Shuning Wu, Ninggwie Liu, "A user Centric Machine Learning Frame work for Cyber Security Operations Center."
[2] Alex Smola and S. V. N. Vishwanatham, "Introduction to Machine Learning".
[3] Fatos T. Yarman Vural, Sanjeev R. Kulkarni, and H. Vincent Poor, "Machine Learning Methods for Attack Detection in the Smart Grid."
[4] V. Rao Vemuri, "Machine Learnin in cyber security."
[5] Barus Ozan Dinçer, Ahmet Sevim, Alaa Eleyan, "Image Processing Based on Anti-Sleep Alarm system for Drowsey Drivers."