A Survey: Systematic Machine Learning Approach through Python Programming

*Mr.Piyush Anjankar, Mr.Ashish Palandurkar[1], Mr.Jagdish Pimple[2], Mr.Shankar Gadhve [3]

*, 1, 2, 3 Assistant Professor, Dept. of Information Technology, Nagpur Institute of Technology, Nagpur

*piyush.anjankar@gmail.com, [1]ash.pal1988@gmail.com, [2]pimplejagdish@gmail.com, [3]shankargadhve87@gmail.com

Abstract

This paper mainly emphasizes on a survey on building a Machine learning systems using python as a programming language. The paper provides various types of machine learning techniques with its applications, as machine learning is one of the advanced topics of discussion in computer science, it should be well understood before the development. This paper gives the various machine learning libraries freely available for the development of a machine learning system using python such as Pandas, NumPy, SciPy etc. This all libraries are available freely under the BSD license for development and utilization of various modules of these libraries

Keywords: *Machine Learning (ML); Supervised; NumPy; SciPy; Learning etc.*

1. Introduction

Machine Learning recently gained its popularity after new advancement occurred in deep learning area [1], Machine learning is a part of Artificial Intelligence in computer science which gives computer systems an capability to learn without being explicitly programmed.[2] Machine learning techniques are using it every day to day life, such as in social media sites like facebook, twitter etc. Machine learning allows the computer system to learn for examples and experiences, by this it makes computer system more intelligent and performance oriented. The past research in the area of machine learning proved that it increased the system performance is better than the traditional system.

Machine learning models can carry out very complex and composite processes by learning through data when a system is in the learning phase, instead of following a default configured a set of rules such as in conventional system likes to do. Growing of data fives ease machine learning systems to be learned on a large and serious pool of data, while increasing of the processing power of computer system has also increased the significant capability of these machine learning systems. [2]

Machine learning systems are used to teach machines through the data, that how to manage and handle the data more precisely and effectively to obtain solutions. With a large amount of data set, sometimes it is not to understand the pattern between the data, also to extort relative and required information from the various features and attribute of dataset [3], in such situations most of the time the data handling required to imply machine learning method comes useful for computations, and generally machine learning with the data science or big data will be used in such scenarios.

The development of the machine learning model includes different stages, such as collection of data, organizing of data, selecting of best model, learning from data, estimation, guessing *etc*. The machine learning system is something which provides a predictive output on supplying of input from a dataset based on criteria learned and

various conditions learn by the model, at the time of learning phase of the system. For the learning of system required some accurate data to be learned from, the organization of data is also very importation, as a power of data pre-processed will be making a crucial and important decision on the system quality. Once the system learns, a system is all set for the estimation of new unknown inputs, to be guessed correctly as per the training provided. The precise use of model assessment, model selection, and algorithm diversity technique is extremely essential in view of machine learning study and research also in many built-up settings for an enhanced outcome from the system. [5]

1.1 Python Introduction

Python is basically a general-purpose programming language, but in recent year it gets tremendous demand for various purpose like web application development, data analysis, natural language processing, Image Processing, Machine Learning and so on, due to its ease and power of use in domain-specific scripting languages such as HTML5, MATLAB or R programming Language, As a concern of Machine Learning Python provides libraries for data loading, visualization, statistics, natural language processing, image processing, and more. This vast toolbox provides data scientists with a large array of general- and special-purpose functionality. One of the main advantages of using Python is the ability to interact directly with the code, using a terminal or other tools like the Jupyter Notebook. [10] Moreover the advantages of python to be used for Machine Learning are-

- Less Code
- Prebuilt Library
- Platform Agnostic/Flexibility with built-in API *etc.*

2. History of Machine Learning

The growth of Machine Learning was from the start of the heart of the computer's first generation, it was in the 1940s when the first by hand operated computer system, ENIAC, was invented, and from the beginning of this idea was, to make a intelligent machine to replica and duplicate human thoughts as well as learning like a humans. *Frank Rosenblatt* given the concept of *Perceptron* which was extremely undemanding classifier during this year's but when it was integrated with large numbers, in a group as a network, it became a dominant [11].

A rising importance on the logical, knowledge-based come within reach to cause a split between AI and machine learning, by the years in the 80s, expert systems had come to ruled AI, and statistics be out of favor, thankful to statistics, machine learning becomes extremely prominent in the 1990s. The juncture of statistics and computer science gave a new beginning to the predictive approach in the field of AI. This shifts the field in the way towards data-driven approach. Now, having huge scale and scatters data available, scientists started to make more intelligent systems that were so clever to investigate itself and be trained from huge and scatter data.

We can believe that the year during the 90s as one of the fascination years of machine learning community for development and understanding various concepts of Machine Learning. During this decade there were considerable aids to the field of machine learning. This massive innovation also contributed a lot to scientific development in AI particular. This development is also continuous in the 21st century as we are finding and using several stunning systematic aids to an area of AI such as the model of Deep Learning.

3. Types of Machine Learning

There some variation of how to define the types of Machine Learning but commonly they can be divided into categories according to their purpose and the main categories are the following:

1. Supervised method of learning
2. Unsupervised method of Learning
3. Semi-supervised method of Learning
4. Reinforcement method of Learning

3.1 Supervised Learning

The ultimate goal of the supervised, the machine learning algorithm is to build a hypothesis, h that makes predictions based on evidence in the presence of uncertainty. As adaptive algorithms identify patterns in data, a computer "learns" from the observations of given input data. When exposed to more observations which were not in the learning process of hypothesis, the computer improves its predictive performance. Mathematically supervised learning can be view as mapping function of the input variable to output variable as-

Let say X be input associated with Output Y, during training phase of Leaning of system, then the function Y= f (X), The objective is to fairly accurate the map function f (X) so that when we have original input combinations of data (X) from the dataset that you can estimate the output variables (Y) for that input data.[14]

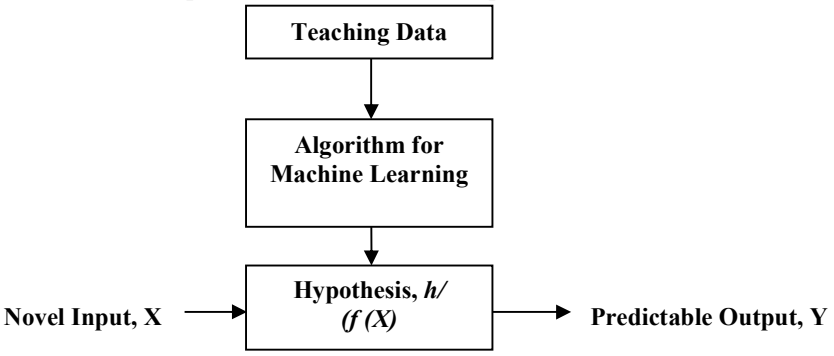

Figure 1: Process for Supervised Learning

It is called supervised learning because the process building of hypothesis or mapping function of learning from the training dataset can be the contemplation of as an instructor supervising the learning process of mapping. As we know the exact output, the Machine Learning system persistently gives a predictive projection on the training data and is been correct by the instructor time by time. Learning going to stops when the Machine Learning algorithm reaches to a satisfactory stage of performance. Supervised learning algorithms roughly divided into-

- Regression
- Classification

3.2 Unsupervised Learning

Unsupervised Learning is a class of Machine Learning techniques to find the patterns in data. The data given to the unsupervised algorithm are not labeled, which

means only the input variables(X) are given with no corresponding output variables. In unsupervised learning, the algorithms are left to themselves to discover interesting structures in the data. [12]

Unsupervised machine learning is great machine learning technique in which you only have input data (X) and no associated equivalent production variables with the applied input combination [11].Algorithms are left to their own devices to find out and to represent the appealing structure in the data for estimations. Unsupervised learning problems algorithms roughly divided into-

- Problems of Clustering's
- Problems of Associations

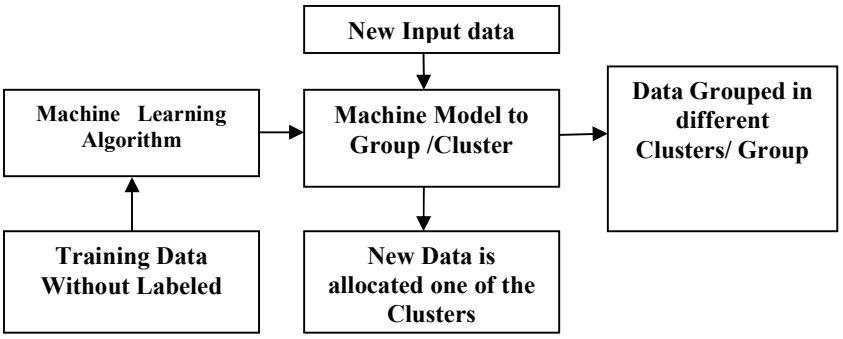

Figure 2: Un-Supervised Learning Process

3.3 Semi-supervised Learning

It is a special form of Learning, Traditional learning use only labeled data (feature / label pairs) for training purpose. Labeled instances, however, are often difficult, expensive, or time-consuming to obtain, as they require the efforts of experienced human annotators. Meanwhile, unlabeled data may be relatively easy to collect, but there have been few ways to use them. Semi-supervised learning addresses this problem by using a large amount of unlabeled data, together with the labeled data, to build better Learning Mechanism. Because semi-supervised learning requires less human effort and gives higher accuracy, it is of great interest both in theory and in practice. [6]

Many real-life problems of today's world fall into this machine learning area. This in view of the fact that it can be costly or prolonged in terms of time to label each data item as it needs access to experts from the domain of particular domain area. Whereas unlabeled data set is inexpensive and simple to assemble and store.

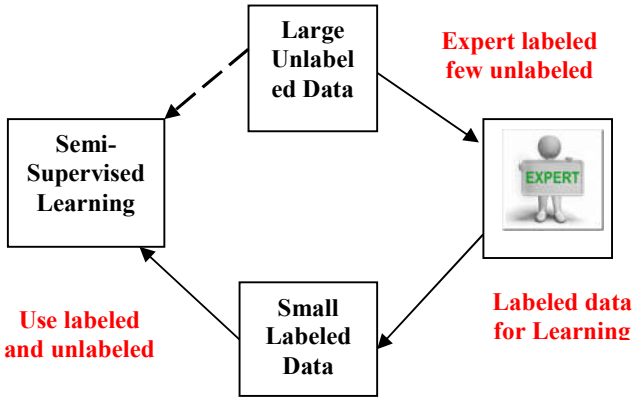

Figure 3: Semi-Supervised Learning Process

3.4 Reinforcement Learning

Reinforcement learning (RL) is floating to transform the field of Machine Learning (ML) and represents a footstep on the road to building self-directed and independent systems with a high-level perceptive and understanding towards the real visual world[7]. Reinforcement learning is most of the time used for training of machine learning model to build a succession of the decision made on the basis of training. The agent in RL learns to attain objective in a vague, potentially highly compound and complex surroundings. In this, a system faces a game-like circumstance. The system employs a to try and guess method to approach towards the solution to the problem. To get the model to do as per user requirement, the model gets either rewards or penalty for the events is performed. Its objective is to capitalize maximum total reward.

Even though the user places the reward return strategy–that are the set of laws of the game–he gives the representation no hints for how to solve the game. By to maximum benefit the power to search and a lot of trials, this is currently the most efficient way to indication towards machine's ingenuity [14].

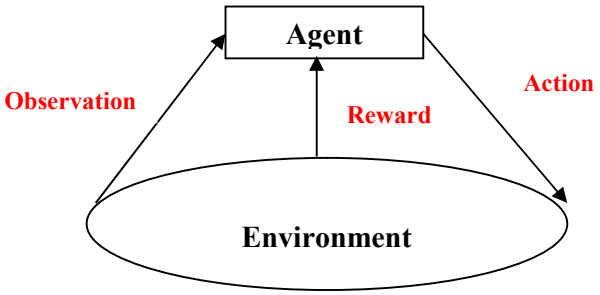

Figure 4: Reinforcement Learning Process

4. Application of Machine Learning

As on present, the field of machine learning is well structured about three primary kinds of research:

(1) task-oriented studies: - The implementation and analysis of learning systems to get better enhance performance in a predetermined set of tasks also known as the engineering approach.

(2) Cognitive simulation: - The exploration and computer simulation of human learning processes.

(3) Theoretical analysis: - The theoretical investigation of the gap between possible learning methods and algorithms independent of the application domain [9].

Now a day's Machine learning(ML) on application domains is widely used in different areas such as Image processing, computer vision, bioinformatics, game playing, agriculture, intrusion detection, Information retrieval, healthcare, marketing, malware detection, image de-convolution, Business Data Analytics (BDA) and so on[8].

5. Machine Learning with Python

To Start with programming in python, we all need to install python either version 2.7.x or python 3.5.x higher, the installation of python is simple code from Command window on Microsoft Windows whereas python is already by default available with Linux and Apple Macintosh System.

The stand along python environment wills not going to help in the programming of machine language, but with this, we required various standard libraries, which will be downloaded individually or as complete distributions. We will recommend you to go with standard free distributions available freely like CPython, ActivePython, Anaconda, WinPython, Python(x, y), Enthought Canopy etc. these distributions are already standard packages for various purpose such as Machine Learning with the standard library for Machine Learning. The basic Libraries required for Machine Learning are-

- NumPy
- SciPy
- Pandas
- Sci-Kit-Learn
- MatPlotLib

5.1 NumPy

NumPy is the fundamental package for scientific computing with Python. It contains among other things:

- a powerful N-dimensional array object
- sophisticated (broadcasting) functions
- tools for integrating C/C++ and Fortran code
- useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of database

5.2 SciPy

SciPy is a set of open source (BSD licensed) scientific and numerical tools for Python. It currently supports special functions, integration, ordinary differential equation (ODE) solvers, gradient optimization, parallel programming tools, an expression-to-C++ compiler for fast execution, and others. A good rule of thumb is that if it's covered in a general textbook on numerical computing (for example, the well-known Numerical Recipes series); it's probably implemented in scipy.

SciPy is a collection of mathematical algorithms and convenience functions built on the Numpy extension of Python. It adds significant power to the interactive Python session by providing the user with high-level commands and classes for manipulating and visualizing data. With SciPy an interactive Python session becomes a data-processing and system-prototyping environment rivaling systems such as MATLAB, IDL, Octave, R-Lab, and SciLab.

The additional benefit of basing SciPy on Python is that this also makes a powerful programming language available for use in developing sophisticated programs and specialized applications.

5.3 Pandas

A panda is a Python package providing fast, flexible, and expressive data structures designed to make working with "relational" or "labeled" data both easy and intuitive. It aims to be the fundamental high-level building block for doing practical, real-world data analysis in Python. Additionally, it has the broader goal of becoming the most powerful and flexible open source data analysis / manipulation tool available in any

language. It is already well on its way toward this goal. A panda is well suited for many different kinds of data:

- Tabular data with heterogeneously-typed columns, as in an SQL table or Excel spreadsheet
- Ordered and unordered (not necessarily fixed-frequency) time series data.
- Arbitrary matrix data (homogeneously typed or heterogeneous) with row and column labels
- Any other form of observational / statistical data sets. The data actually need not to be labeled at all to be placed into a pandas data structure

5.4 Sci-Kit-Learn

Scikit-learn is a Python Package for machine learning algorithms built on above of SciPy and available under the free BSD license. Scikit-learn gives a range of supervised learning and unsupervised learning models through a reliable interface in Python language. This package is built on top of SciPy which need to be pre-installed before using scikit-learn. This package consists of various packages like:

- SciPy
- NumPy
- IPython
- Matplotlib
- Pandas

The basic idea for developing this strong library is to give support and achieves strength for developing machine learning systems. This means a thoughtful hub on concern such as easiness towards usage, code superiority, association, documentation, and performance.

5.5 MatPlotLib

Matplotlib has an extensive codebase that can be daunting to many new users. However, most of Matplotlib can be understood with a fairly simple conceptual framework and knowledge of a few important points.

Plotting requires action on a range of levels, from the most general (e.g., 'contour this 2-D array') to the most specific (e.g., 'color this screen pixel red'). The purpose of a plotting package is to assist you in visualizing your data as easily as possible, with all the necessary control -- that is, by using relatively high-level commands most of the time and still have the ability to use the low-level commands when needed.

6. Conclusion

A Machine learning in need of future, as a machine is expected to work like a Humans, the developing of machine learning algorithm and hypothesis is one of the real developing phases in an era of computer science and statistics. The developing of the machine learning system can be done in various languages, but python is highly preferable language, for development because of its various features as discussed in this paper.

7. References

[1] S. Ozan, "Increasing system performance in machine learning by using multiprocessing," 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1 4.

[2] Diksha Sharma , Neeraj Kumar, “A Review on Machine Learning Algorithms, Tasks and Applications”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 6, Issue 10, October 2017, ISSN: 2278 – 1323
[3] Dey Ayon, “Machine Learning Algorithms: A Review”, International Journal of Computer Science and Information Technologies, Vol. 7 (3), 2016, 1174-1179
[4] Caesar Wu, Rajkumar Buyya, and Kotagiri Ramamohanarao ,Chapter 1 Big Data Analytics = Machine Learning + Cloud Computing
[5] Sebastian Raschka, “Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning”, arXiv preprint arXiv: 1811.12808, 2018/11
[6] "Semi-supervised learning literature survey." Edited Zhu, Xiaojin. Dept. of Computer Science, University of Wisconsin-Madison 2.3 (2006): 4.
[7] K. Arulkumaran, M. P. Deisenroth, M. Brundage and A. A. Bharath, "Deep Reinforcement Learning: A Brief Survey," in IEEE Signal Processing Magazine, vol. 34, no. 6, pp. 26-38, Nov. 2017. doi: 10.1109/MSP.2017.2743240
[8] S. Angra and S. Ahuja, "Machine learning and its applications: A review," 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), Chirala, 2017, pp. 57-60.doi: 10.1109/ICBDACI.2017.8070809
[9] 1 - An Overview Of Machine Learning, Editor(s): Ryszard S. Michalski, Jaime G. Carbonell, Tom M. Mitchell, Machine Learning,Morgan Kaufmann, 1983, Pages 3-23, ISBN 9780080510545
[10] Sarah Guido, Andreas Müller, Editor, Introduction to Machine Learning with Python: A Guide for Data Scientists, O'Reilly Media Release (October 2016)
[11] https://provalisresearch.com/blog/brief-history-machine-learning/
[12] https://towardsdatascience.com/unsupervised-learning-with-python 173c51dc7f03
[13] https://machinelearningmastery.com
[14] https://deepsense.ai/what-is-reinforcement-learning-the-complete-guide/
[15] https://www.scipy.org/docs.html