IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727
PP 80-83
www.iosrjournals.org

Machine Learning: Challenges, Opportunities and Applications

Anand Khandare[1], Saurabh Indoria[2], Kalindi Awasthi[3], Shubham Chhaparia[4].
[1](Computer Engineering, Thakur College of Engineering and Technology/ Mumbai University, India)
[2](Computer Engineering, Thakur College of Engineering and Technology / Mumbai University, India)
[3](Computer Engineering, Thakur College of Engineering and Technology / Mumbai University, India)
[4](Computer Engineering Thakur College of Engineering and Technology / Mumbai University, India)

Abstract: *With the enhancement in the technology sector, Machine Learning is evolving on its own, due to key role it can play in providing potential solutions to the various sectors like data mining, image processing, language processing . In this paper, we discuss the various domains wherein the machine learning provides a broader scope for efficient solutions. This paper has four sections (i) Introduction (ii) Data sets (iii) Machine learning models (iv) Challenges (v) scope (vi)applications of machine learning.*
Keywords: *Dataset, Machine Learning, training.*

I. Introduction

Machine Learning majorly provides computers with the ability to learn on itself without explicitly being programmed to do so. It basically focuses on development of programs which can adapt to new set of data. Machine learning supports that kind of data analysis which learns from previous datasets and pattern due to which it claims to replace data analytics and predictions carried out manually.

II. Data sets

Machine learning since basically deals with learning from previous data hence it needs a good training set to work properly for its learning. These datasets act as benchmarks from which machine learning nets are trained.

Machine learning works with two data sets training and test. Running training set teaches to weigh various features through neural networks and assigning them accordingly the coefficients for minimizing errors in the results.

The second set known as test set functions as approval which is used after the training and optimization of your data. The neural net is tested against this final random sampling.

III. Machine learning models

There are six machine learning models which comes from different algorithm approaches and performs accordingly under different datasets

3.1 Decision tree based method

The basic idea is to recursively divide the training set into buckets of homogeneous members wherein the members are represented as leaf node of tree. The advantage of tree is the flexibility in terms of input and output variables. The level of decision nodes indicates the degree of influences of other input variables.

3.2 Linear regression methods

The basic assumption is output variable can be expressed in terms of linear combination of set of input variable : $y=w1x1+w2x2+\ldots\ldots$.

The whole objective is to learn the weights by minimizing the error function lost. The strength of this model is the high performance in the scoring and learning.

3.3 Neural network

It can be considered to be a multiple layer perceptron wherein each is a logistic unit with multiple binary input and single output. It therefore enables neural network to study non-linear relationships between input and output.

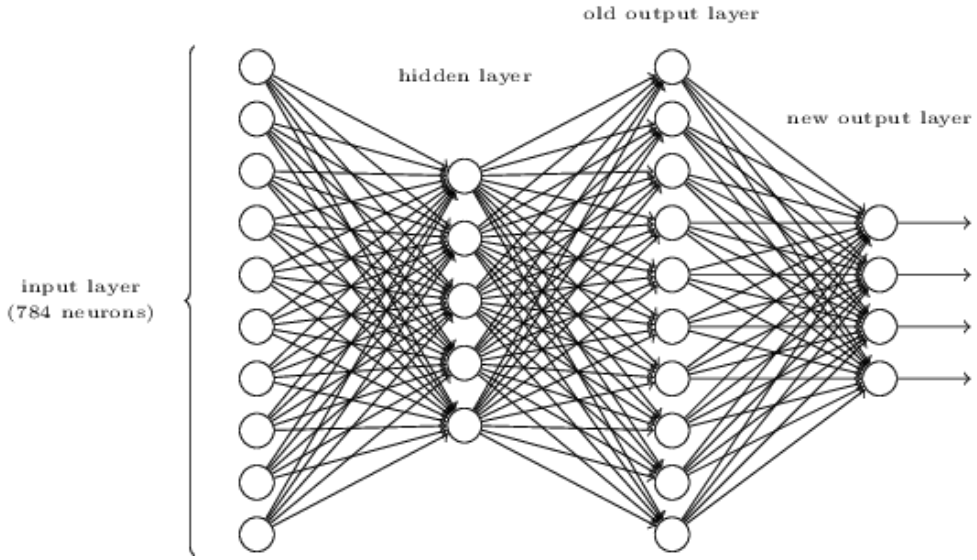

Figure 1. A deep neural network.

3.4 Bayesian Network

It's a dependency graph where every node points to binary variable and edge to dependency relationship. The learning is the probability of all incoming edges at each node and then compute probability of output variable from the observed input variable.

3.5 Support vector machine

It takes numeric input and gives binary output. It finds linear plane with maximum margin to separate two class of output [1].

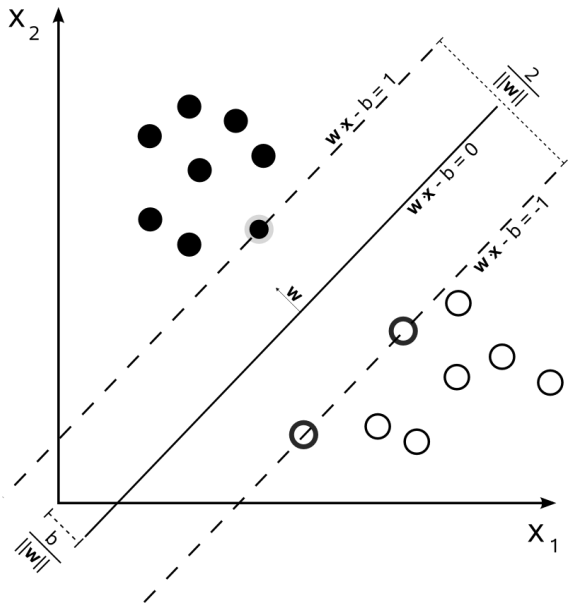

Figure 2. Classification using SVM.

3.6 Nearest neighbor

Here the basic concept is to find n similar data point from the set and use them for interpolation of output value. An example is K-nearest neighbor algorithm used for classification and regression. It is a type of instance-based learning and is one of the simplest machine learning algorithms [2].

IV. Challenges

Although, machine learning techniques outperform human capabilities in certain areas, but there are several challenges which are faced during the development of machine learning models.

4.1 High computation power

Generally, while training neural networks or implementing other models, the amount of processing power required is huge. However, a number of computers can be connected through a network and can be used to share the work, but still, the process is time consuming without high end computers. GPUs are also used to speed up the entire process.

4.2 Huge datasets required

For training any machine learning model, it is always better to have as large dataset as possible. Collecting such datasets is not simple. Also, creating datasets manually requires a huge amount of effort. Eg. IMAGENET project collected a dataset of 15 billion images and categorized them manually which required a help from 50 thousand workers. This dataset was made open source.

4.3 Memory requirement

Loading the datasets and the model into memory is not an easy task. For e.g. The common crawl dataset is 151 Tera bytes in size which includes web crawl data of 1.82 billion pages collected over eight years. Now loading this dataset is not a simple task. Distributed and cloud environments helps in this matter, but that takes a toll on the cost.

V. SCOPE

As there is an increase in the amount of the data nowadays it's giving rise to find efficient ways of handling them. Machine learning provides much efficient ways of data handling i.e. it has major scopes in data mining, information retrieval, design and diagnosis, speech perception and scientific application of all kinds.

Various fields of computer science like image processing, natural language processing, etc. can be combined with Machine Learning models to give unbelievable outputs. Recently, Google created a Machine Learning program that defeated the world champion in Alpha Go, the 19x19 board game with more combinations than the number of atoms in this universe. We can train our machines to do any kind of task making human life easier and probably solving complex problems as well.

Machine learning is considered to be yet in its starting phase, but our future world will be largely contributed by Machine learning where we might even see artificial intelligence almost as capable as human brains.

VI. Applications

Machine learning is used by various companies to provide better user experience, improve their efficiency, growth, etc.

6.1 Google Search Engine

Google's search engine uses machine learning to determine the pattern of searches by the users to give recommendations and display results in more personalized way. It also ranks the pages based on the searches and depending on the location and various other parameters [3][4].

6.2 YouTube

YouTube uses machine learning to provide recommendations about the videos based on the user's watch history and searches [5][6].

6.3 Self driving cars

Tesla motors have launched their self-driving cars which learns from all the cars while users are driving. This trains their neural network and then when the user switches to automatic mode, the machine learning model drives the car successfully. With every minute a car is drove; the network learns more.

6.4 Spam filters

The most common application of Machine learning is the intelligent spam filter for emails. A machine learning model was trained with supervised learning giving a dataset which consisted of spam and normal emails with labels. It was then deployed to filter out the messages which is supposedly spam. With user's feedback, it trains continuously [7].

6.5 Speech Synthesizer and Recognition

Speech synthesizing is too tough to hard-code. Machine learning is used to solve this problem. Along with this, speech recognition is another success in this field. Google's speech recognition is a good example for the same.

6.6 Personal Assistance and Home automation

iPhone's Siri, Microsoft's Cortana, Google's home, Amazon's Alexa, etc. are examples of personal assistance and home automation. These are also programmed through machine learning models. They also include speech synthesizing and recognition modules.

6.7 Image classification

Classifying images based on the content of the images is termed as image classification. Image classifiers can distinguish an image of a cat from an image of a dog.

VII. Conclusion

Looking at the pace at which Machine learning field is advancing, we can conclude that this will be the next computer revolution. There are various challenges in various fields including medical, engineering, arts, etc. which can be possibly solved by machine learning models one day. Already existing models are capable but are heavy in terms of processing. With the advancement in computational capabilities, machine learning will enhance its capabilities.

References

[1]. T. Joachims. Text categorization with support vector machines: Learning with many relevant features. In Proceedings of the *10th European Conference on Machine Learning*, pages 137{142, Heidelberg, Germany, *1998*.

[2]. T.M. Mitchell. Machine learning. McGraw-Hill, 1997.

[3]. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "*Privacy-preserving multi-keyword ranked search over encrypted cloud data*," in *INFOCOM, 2011* Proceedings IEEE. IEEE, 2011, pp. 829–837

[4]. Schafer, J. Ben, Joseph, A. Konstan, John Riedl, "*E-commerce Recommendation Applications*", Data Mining and Knowledge Discover, *2001*, pp.115-153.

[5]. Wikipedia, http://en.wikipedia.org/wiki/tf-idf. N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "*Privacy-preserving multi-keyword ranked search over encrypted cloud data," Parallel and Distributed Systems, IEEE Transactions on, vol. 25, no. 1*, pp. 222–233, 2014.

[6]. S. Qin, R. Menezes and M. Silaghi. *A Recommender System for YouTube Based on its Network of Reviewers*. In *SocialCom '10, 2010.*

[7]. I. Androutsopoulos, G. Paliouras, and E. Michelakis. Learning to filter unsolicited commercial e-mail. Technical report, *National Centre for Scienti_c Research \Demokritos", 2004.*