Grenze International Journal of Engineering and Technology, July Issue

An Approach to Supervised Machine Learning: A survey

Padma Charan Sahu[1], Sanjay Kumar Rauta[2] and Dr. Ratnakar Dash[3]
[1]GIET UNIVERSITY (Ph.D Scholar) Department of ECE, Gunupur, India
Email: kit.padma@gmail.com
[2-3]VIPMS Department of Computer Science, Berhampur, India, NIT Rourkela, India
Email: rout143sanjaya@gmail.com

***Abstract*—The main objective of machine learning is for analysing by using computers for solving a given problem using past experience. Machine Learning can be applied as affiliation examination through supervised learning, Semi supervised learning and Reinforcement Learning yet right now will concentrate on quality and shortcoming of supervised learning characterization calculations. The objective of supervised learning is to assemble a brief model of the circulation of class regarding corresponding parameters. The subsequent classifier is then used to allot class marks to the testing cases where the estimations of the corresponding parameters are known, yet the estimation of the class name is obscure. We are hopeful that this examination will assist new analysts with guiding new research regions and to analyze the adequacy and impuissance of regulated learning calculations.**

***Index Terms*— Supervised Machine Learning, Semi Supervised learning, Distribution Point.**

I. INTRODUCTION

After the “AI winter” of the 80s and the 90s, interest in the application of data-driven Artificial Intelligence (AI) techniques has been steadily increasing in a number of engineering fields, including speech and image analysis [1] and communications [2]. Unlike the logic-based expert systems that were dominant in the earlier work on AI (see, e.g., [3]), the renewed confidance in data driven methods is motivated by the successes of pattern recognition tools based on machine learning. These tools rely on decades-old algorithms, such as back propagation [4], the Expectation Maximization (EM) algorithm [5], and Q-learning [6], with a number of modern algorithmic advances, including novel regularization techniques and adaptive learning rate schedules [7].Their success is built on the unprecedented availability of data and computing resources in many engineering domains. While the new wave of promises and breakthroughs around machine learning arguably falls short, at least for now, of the requirements that drove early AI research [3], [8], learning algorithms have proven to be useful in a number of important applications – and more is certainly on the way.

This paper provides a very brief introduction to key concepts in supervised machine learning. Unlike other review papers such as [9]–[11], the presentation aims at highlighting conditions under which the use of machine learning is justified in engineering problems, as well as specific classes of learning algorithms that

Grenze ID: 01.GIJET.6.2.504_1

are suitable for their solution. The presentation is organized around the description of general technical concepts, for which an overview of applications to communication networks is subsequently provided. These applications are chosen to exemplify general design criteria and tools and not to offer a comprehensive review of the state of the art and of the historical progression of advances on the topic.

We proceed in this section by addressing the question "What is supervised machine learning?", by providing a taxonomy of supervised machine learning methods.

Machine Learning (ML) can be considered as a subfield of Artificial Intelligence since those algorithms can be seen as building blocks to make computers learn to behave more intelligently by somehow generalizing rather that just storing and retrieving data items like a database system and other applications would do. Machine learning has got its inspiration from a variety of academic disciplines, including computer science, statistics, biology, and psychology. The core function of Machine learning attempts is to tell computers how to automatically find a good predictor based on past experiences and this job is done by good classifier. Classification is the process of using a model to predict unknown values (output variables), using a number of known values (input variables). The classification process is performed on data set D which holds following objects:

- Set size → A = $\{A_1\ A_2........\ A_n\}$ where $A_1\ A_2........\ A_n$ denotes the number of attributes or the size of the set A.
- Class label→ C: Target attributes; C = $\{c_1\ c_2..........\ c_n\}$ where $c_1\ c_2..........\ c_n$ is the number of classes.

Given a data set D, the objective of Machine Learning is to produce a prediction/classification function to relate values of attributes in A and classes in C. Data mining is one of the most tools of machine learning among the number of different applications by improving the efficiency of systems and the designs of machines [1].

In machine learning algorithms every instance of particular dataset is represented by using the same set of features. The nature of these features could be continuous, categorical or binary. If instances are given with known labels (i.e. the corresponding correct outputs) then the learning scheme is known as supervised, while in unsupervised learning approach the instances are unlabeled. Through applying these unsupervised (clustering) algorithms, researchers are optimistic to discover unknown, but useful, classes of items [3].

Another kind of machine learning is reinforcement learning. Here the training information provided to the learning system by the environment (i.e. external trainer) is in the form of a scalar reinforcement signal that constitutes a measure of how well the system operates. The learner is not told which action has to take, as in most forms of machine learning, but instead must discover which actions yield the most reward by trying them [1]. A number of Machine Learning applications involve tasks that can be set up as supervised. The below figure depicts the general classification architecture.

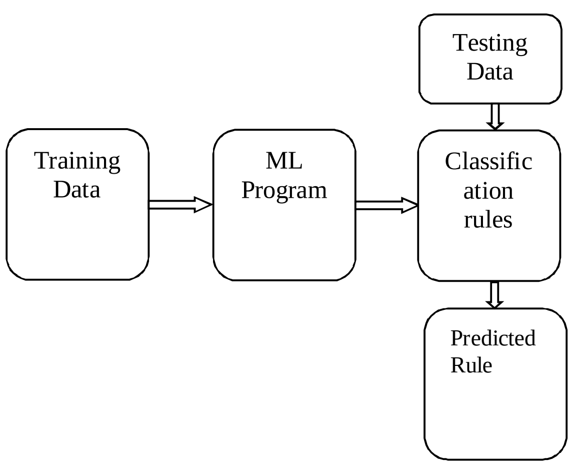
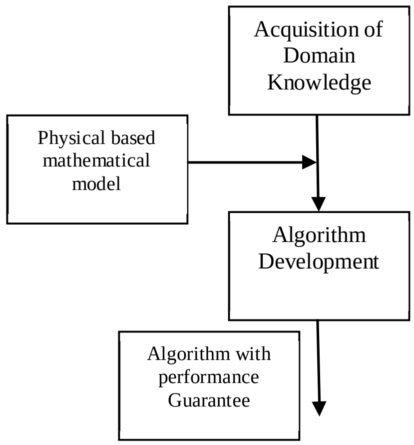


Fig 1 Classification Architecture

The problem of interest in studying the ML technique basically belongs to some mathematical model. Based on this model, an algorithm is introduced which guarantees that the given model is accurate representation of reality.



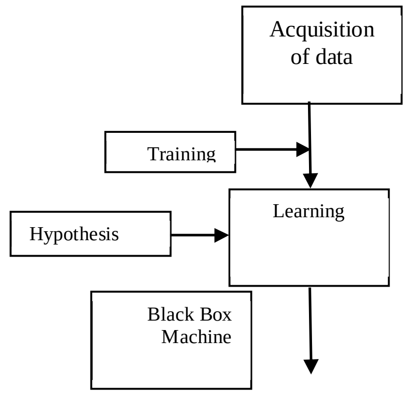


Fig 2.a. Conventional engineering design flow   Fig 2.b. Baseline machine learning methodology

In contrast, in its most basic form, the machine learning approach substitutes the step of acquiring domain knowledge with the potentially easier task of collecting a sufficiently large number of examples of desired behaviour for the algorithm of interest. These examples constitute the training set. As seen in Fig. 2(b), the examples in the training set are fed to a learning algorithm to produce a trained "machine" that carries out the desired task. Learning is made possible by the choice of a set of possible "machines", also known as the hypothesis class, from which the learning algorithm makes a selection during training. An example of an hypothesis class is given by a neural network architecture with learnable synaptic weights. Learning algorithms are generally based on the optimization of a performance criterion that measures how well the selected "machine" matches the available data. For the problem of designing a channel decoder, a machine learning approach can hence operate even in the absence of a well-established channel model. It is in fact enough to have a sufficiently large number of examples of received signals – the inputs to the decoding machine – and transmitted messages – the desired outputs of the decoding machine – to be used for the training of a given class of decoding functions [13].

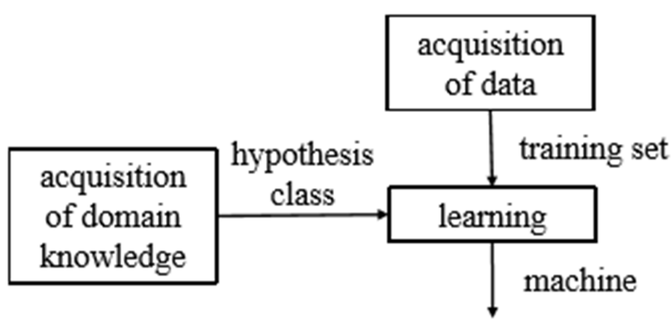
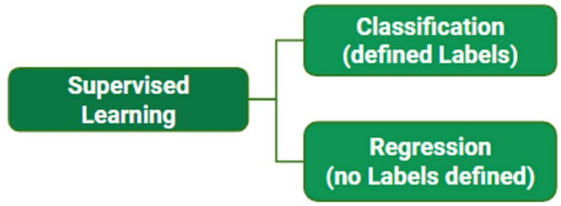


Fig.3. Machine learning methodology that integrates domain knowledge during model selection

Most current machine learning applications fall in the supervised learning category, and hence aim at learning an existing pattern between inputs and outputs. Supervised learning is relatively well-understood at a theoretical level [14], [15], and it benefits from well established algorithmic tools.. Nevertheless, it arguably poses a more fundamental practical problem in that it directly tackles the challenge of learning by direct observation without any form of explicit feedback. This paper only covers supervised learning. Reinforcement learning requires a different analytical framework grounded in Markov Decision Processes

II. Issues of Supervised Learning Algorithms



In supervised learning the first step is dealing with dataset. In order to perform a better training on data set an appropriate expert could suggest better selection of features. If concerned expert is not in reach, then the other approach is “brute-force”, which means measuring everything available in the hope that the right (informative, relevant) features can be isolated. However, a dataset collected by the “brute-force” method is not directly suitable for induction. Ultimately, in most cases it contains noise and missing feature values, and therefore requires significant pre-processing [1]. In the next step, data preparation and data preprocessing is a key function of researcher in Supervised Machine Learning (SML). A number of techniques have been introduced by different researchers to deal with missing data issue. Hodge & Austin [4] have conducted a survey of contemporary techniques for outlier (noise) detection. Karanjit & Shuchita [5] have also discussed different outlier detection methods which are being used in different machine learning. H. Jair [6] has done comparison on 6 different outlier detection methods by performing experiment on benchmark datasets and a synthetic astronomical domain. The major issues are

A. Computational learning theory

The goal of the supervised learning algorithm is to optimize some measure of performance such as minimizing the number of mistakes made on new samples. In addition to performance bounds, computational learning theory studies the time complexity and feasibility of learning. In computational learning theory, a computation is considered feasible if it can be done in polynomial time. There are two kinds of time complexity results:

- Positive results – Showing that a certain class of functions is learnable in polynomial time.
- Negative results – Showing that certain classes cannot be learned in polynomial time.

Negative results often rely on commonly believed, but yet unproven assumptions, such as:

- Computational complexity – $P \neq NP$ (the P versus NP problem);
- Cryptographic – One-way functions exist.

There are several different approaches to computational learning theory based on making different assumptions about the inference principles used to generalize from limited data. This includes different definitions of probability (see frequency probability, Bayesian probability) and different assumptions on the generation of samples. The different approaches include Exact learning, Probably approximately correct learning (PAC learning), VC theory, Bayesian inference, Algorithmic learning theory

B. Inductive bias

The inductive bias of a learning algorithm is the set of assumptions that the learner uses to predict outputs given inputs that it has not encountered.

C. Over fitting (machine learning)

In statistics, over fitting is "the production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably". An over fitted model is a statistical model that contains more parameters than can be justified by the data. The essence of over fitting is to have unknowingly extracted some of the residual variation (i.e. the noise) as if that variation represented underlying model structure.

Under fitting occurs when a statistical model cannot adequately capture the underlying structure of the data. An under fitted model is a model where some parameters or terms that would appear in a correctly specified model are missing. Under fitting would occur, for example, when fitting a linear model to non-linear data. Such a model will tend to have poor predictive performance.

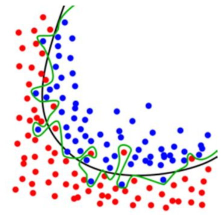


Fig.5.The green line represents an over fitted model and the black line represents a regularized model. While the green line best follows the training data

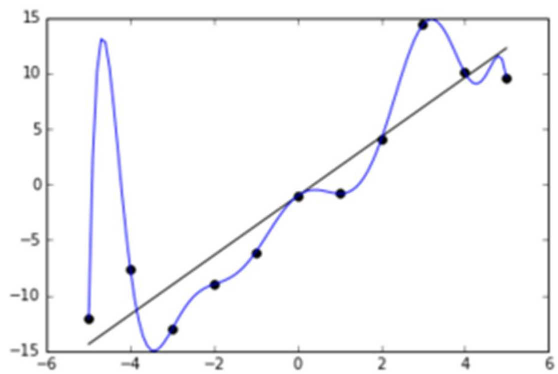


Fig 6.Noisy data is fitted to a linear function and a polynomial function

D. Algorithm Selection

The selection of algorithm for achieving good results is an important step. The algorithm evaluation is mostly judge by prediction accuracy. The classifier's (Algorithm) evaluation is most often based on prediction accuracy and it can be measured by given below formula

$$Accuracy = \frac{Number\ of\ correct\ classification}{Total\ number\ of\ test\ cases}$$

There are number of methods which are being used by different researchers to calculate classifier's accuracy. Some researcher's splits the training set in such a way that, two-thirds retain for training and the other third for estimating performance. Cross-Validation (CV) or Rotation Estimation is another approach. CV provides a way to make a better use of the available sample. In k-fold cross-validation scheme, we divide the learning sample into k disjoint subsets of the same size, i.e. $I_{S_1}\ U\ I_{S_2}\ U\ I_{S_3}\ U\ I_{S_4} \ldots\ldots\ldots\ U\ I_{S_K}$ .A model is then inferred by the learning algorithm from each sample ls\ls, i = 1,..,k and its performance is determined on the held out sample is I_{s_i}. Final performance is computed as the average performance over all these models. Notice that when k is equal to the number of objects in the learning sample, this method is called leave-one-out. Typically, smaller values of k (10 or 20) are however preferred for computational reasons [7]. The comparison between supervised ML methods can be done through to perform statistical comparisons of the accuracies of trained classifiers on specific datasets. For doing this we can run two different learning algorithms on samples of training set of size N, estimate the difference in accuracy for each pair of classifiers on a large test set[1]. For classification of data, a good number of techniques have been developed by researchers, such as logical statistics based techniques.

III. Conclusion

In Supervised learning, you train the machine using data which is well "labelled."You want to train a machine which helps you predict how long it will take you to drive home from your workplace is an example of supervised learning. Regression and Classification are two types of supervised machine learning techniques. Supervised learning is a simpler method while unsupervised learning is a complex method. The biggest challenge in supervised learning is that irrelevant input feature present training data could give inaccurate results. The main advantage of supervised learning is that it allows you to collect data or produce a data output from the previous experience. The drawback of this model is that decision boundary might be overstrained if your training set doesn't have examples that you want to have in a class. As a best practice of supervise learning, one can need to decide what kind of data should be used as a training set.

REFERENCES

[1] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath et al., “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,” IEEE Signal processing magazine, vol. 29, no. 6, pp. 82–97, 2012.

[2] M. Ibnkahla, “Applications of neural networks to digital communications–a survey,” Signal processing, vol. 80, no. 7, pp. 1185–1215, 2000.

[3] H. J. Levesque, Common Sense, the Turing Test, and the Quest for Real AI: Reflections on Natural and Artificial Intelligence. MIT Press, 2017.

[4] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning internal representations by error propagation,” California Univ SanDiegoLaJollaInstforCognitiveScience,Tech.Rep.,1985.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” Journal of the royal statistical society. Series B (methodological), pp. 1–38, 1977.

[6] Yen-Liang Chen and Lucas Tzu-Hsuan Hung, “Using decision trees to summarize associative classification rules”, Expert Systems with Applications, Vol. 36, No. 2, Part 1, pp. 2338-2351, 2009.

[7] Samy Bengio, Li Deng, Hugo Larochelle, Honglak Lee, and Ruslan Salakhutdinov, “Guest Editors’ Introduction: Special Section on Learning Deep Architectures”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, No. 8, pp. 1795-1797, 2013.

[8] Qinghua Zheng, Zhaohui Wu, Xiaocheng Cheng, Lu Jiang and Jun Liu, “Learning to crawl deep web”, Information Systems, Vol. 38, No. 6, pp. 801-819, 2013.

[9] Xue-Wen Chen and Xiaotong Lin,” Big Data Deep Learning: Challenges and Perspectives”, IEEE Access Practical Innovations: Open Solutions and Access and IEEE, Vol. 2, pp. 514-525, 2014.

[10] Rajat Raina, Anand Madhavan and Andrew Yg, “Largescale Deep Unsupervised Learning using Graphics Processors”, 26th International Conference on Machine Learning, pp. 609-616, 2009.

[11] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, Advances in Neural Information Processing System, pp. 1106-1114, 2012.

[12] Jeffrey Dean, Greg S. Corrado and Rajat Monga Kai, “Large Scale Distributed Deep Networks”, Advances in Neural Information Processing System, pp. 1232-1240, 2012.

[13] Quoc V. Le, Marc’Aurelio Ranzato, Rajat Monga, Matthieu Devin, Kai Chen, Greg S. Corrado, Jeffrey Dean, and Andrew Y. Ng, “Building High-level Features Using Large Scale Unsupervised Learning”, Proceedings of the 29th International Conference on Machine Learning, 2012.

[14] A. Coats and B. Huval, “Deep Learning with COTS HPS systems”, Journal of Machine Learning Research, Vol. 28, No. 3, pp. 1337-1345, 2013.