

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/221907669>

Introduction to Machine Learning

Chapter · February 2010

DOI: 10.5772/9394 · Source: InTech

CITATIONS

43

READS

874

1 author:



Taiwo Ayodele

University of Portsmouth

27 PUBLICATIONS 147 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Towards e-learning security: A machine learning approach [View project](#)



Security review of email summarization systems [View project](#)

Introduction to Machine Learning

Taiwo Oladipupo Ayodele
University of Portsmouth
United Kingdom

1. Introduction

In present times, giving a computer to carry out any task requires a set of specific instructions or the implementation of an algorithm that defines the rules that need to be followed. The present day computer system has no ability to learn from past experiences and hence cannot readily improve on the basis of past mistakes. So, giving a computer or instructing a computer controlled programme to perform a task requires one to define a complete and correct algorithm for task and then programme the algorithm into the computer. Such activities involve tedious and time consuming effort by specially trained teacher or person. Jaime et al (Jaime G. Carbonell, 1983) also explained that the present day computer systems cannot truly learn to perform a task through examples or through previous solved task and they cannot improve on the basis of past mistakes or acquire new abilities by observing and imitating experts. Machine Learning research endeavours to open the possibility of instruction the computer in such a new way and thereby promise to ease the burden of hand writing programmes and growing problems of complex information that get complicated in the computer.

When approaching a task-oriented acquisition task, one must be aware that the resultant computer system must interact with human and therefore should closely match human abilities. So, learning machine or programme on the other hand will have to interact with computer users who make use of them and consequently the concept and skills they acquire- if not necessarily their internal mechanism must be understandable to humans. Also Alpaydin (Alpaydin, 2004) stated that with advances in computer technology, we currently have the ability to store and process large amount of data, as well as access it from physically distant locations over computer network. Most data acquisition devices are digital now and record reliable data. For example, a supermarket chain that has hundreds of stores all over the country selling thousands of goods to millions of customers. The point of sale terminals record the details of each transaction: date, customer identification code, goods bought and their amount, total money spent and so forth, This typically amounts to gigabytes of data every day. This store data becomes useful only when it is analysed and tuned into information that can be used or be predicted.

We do not know exactly which people are likely to buy a particular product or which author to suggest to people who enjoy reading Hemingway. If we knew, we would not need any analysis of the data; we would just go ahead and write down code. But because we do not, we can only collect data and hope to extract the answers to these and similar question from

data. We can construct a good and useful approximation. That approximation may not explain everything, but may still be able to account for some part of data. We believe that identifying the complete process may not be possible, we can still detect certain patterns or regularities. This is the niche of machine learning. Such patterns may help us understand the process, or we can use those patterns to make predictions: Assuming that the future, at least the near future, will not be much different from the past when the sample data was collected, the future predictions can be expected to be right.

Machine learning is not just a database problem, it is a part of artificial intelligence. To be intelligent, a system that is in a changing environment should have the ability to learn. If the system can learn and adapt to such changes, the system designer need not foresee and provide solutions for all possible situations. Machine learning also help us find solutions to many problems in vision, speech recognition and robotics. Lets take the example of recognising of faces: This is a task we do effortlessly; we recognise family members and friends by looking their faces or from their photographs, despite differences in pose, lighting, hair, style and so forth. But we do consciously and are able to explain how we do it. Because we are not able to explain our expertise, we cannot write the computer program. At the same time, we know that a face image is not just a random collection of pixel: a face has structure, it is symmetric. There are the eyes, the nose, the mouth, located in certain places on the face. Each person's face is a pattern that composed of a particular combination of these. By analysing sample face images of person, a learning program captures the pattern specific to that person and then recognises by checking for the pattern in a given image. This is one example of pattern recognition.

Machine learning is programming computers to optimise a performance criterion using example data or past experience. We have a model defined up to some parameters, and learning is the execution of a computer program to optimise the parameter of the model using the training data or past experience. The model may be *predictive* to make predictions in the future, or *descriptive* to gain knowledge from data, or both. Machine learning uses the theory of statistics in building mathematical models, because the core task is making inference from sample. The role of learning is twofold: First, in training, we need efficient algorithms to solve the optimised problem, as well as to store and process the massive amount of data we generally have. Second, once a model is learned, its representation and algorithmic solution for inference needs to be efficient as well. In certain applications, the efficiency of the learning or inference algorithm, namely, its space and time complexity may be as important as its predictive accuracy.

1.1 History of Machine Learning

Over the years, Jaime et al (Jaime G. Carbonell, 1983) elaborated that research in machine learning has been pursued with varying degrees of intensity, using different approaches and placing emphasis on different, aspects and goals. Within the relatively short history of this discipline, one may distinguish three major periods, each centred on a different concept:

- neural modelling and decision-theoretic techniques
- symbolic concept-oriented learning
- knowledge-intensive approaches combining various learning strategies

1.1.1 The Neural Modelling (Self Organised System)

The distinguishing feature of the first concept was the interest in building general purpose learning systems that start with little or no initial structure or task-oriented knowledge. The major thrust of research based on this approach involved constructing a variety of neural model-based machines, with random or partially random initial structure. These systems were generally referred to as neural networks or self-organizing systems. Learning in such systems consisted of incremental changes in the probabilities that neuron-like elements (typically threshold logic units) would transmit a signal. Due to the early computer technology, most of the research under this neural network model was either theoretical or involved the construction of special purpose experimental hardware systems, such as perceptrons (Forsyth, 1990), (Ryszard S. Michalski, 1955), (Rosenblatt, 1958) pandemonium (Selfridge, 1959), and (Widrow, 2007). The groundwork for this paradigm was laid in the forties by Rashevsky in the area of mathematical biophysics (Rashevsky, 1948), and by McCulloch (McCulloch, 1943), who discovered the applicability of symbolic logic to modelling nervous system activities. Among the large number of research efforts in this area, one may mention many works such as (Rosenblatt, 1958), (Block H, 1961), (Ashby, 1960), (Widrow, 2007). Related research involved the simulation of evolutionary processes, that through random mutation and “natural” selection might create a system capable of some intelligent behaviour (for example, (Friedberg, 1958), (Holland, 1975)). Experience in the above areas spawned the new discipline of pattern recognition and led to the development of a decision-theoretic approach to machine learning. In this approach, learning is equated with the acquisition of linear, polynomial, or related discriminant functions from a given set of training examples. Examples include, (Nilsson, 1982). One of the best known successful learning systems utilizing such techniques (as well as some original new ideas involving non-linear transformations) was Samuel’s checkers program, (Ryszard S. Michalski J. G., 1955). Through repeated training, this program acquired master-level performance somewhat, different, but closely related, techniques utilized methods of statistical decision theory for learning pattern recognition rules.

1.1.2 The Symbolic Concept Acquisition Paradigm

A second major paradigm started to emerge in the early sixties stemming from the work of psychologist and early AI researchers on models of human learning by Hunt (Hunt, 1966). The paradigm utilized logic or graph structure representations rather than numerical or statistical methods. Systems learned symbolic descriptions representing higher level knowledge and made strong structural assumptions about the concepts to be acquired. Examples of work in this paradigm include research on human concept acquisition (Hunt, 1966) and various applied pattern recognition systems. Some researchers constructed task-oriented specialized systems that, would acquire knowledge in the context of a practical problem. Ryszard (Ryszard S. Michalski J. G., 1955), learning system was an influential development in this paradigm. In parallel with (Winslett, 1983) work, different approaches to learning structural concepts from examples emerged, including a family of logic-based inductive learning programs.

1.1.3 The Modern Knowledge-Intensive Paradigm

The third paradigm represented the most recent period of research starting in the mid-seventies. Researchers have broadened their interest beyond learning isolated concepts from examples, and have begun investigating a wide spectrum of learning methods, most based upon knowledge-rich systems specifically, this paradigm can be characterizing by several new trends, including:

1. **Knowledge-Intensive Approaches:** Researchers are strongly emphasizing the use of task-oriented knowledge and the constraints it provides in guiding the learning process. One lesson from the failures of earlier knowledge and poor learning systems is that to acquire and to acquire new knowledge a system must already possess a great deal of initial knowledge.
2. **Exploration of alternative methods of learning:** In addition to the earlier research emphasis on learning from examples, researchers are now investigating a wider variety of learning methods such as learning from instruction, (e.g. (Mostow, 1983), learning by analogy and discovery of concepts and classifications (R. S. Michalski, 1983).

In contrast to previous efforts, a number of current systems are incorporating abilities to generate and select tasks and also incorporate heuristics to control their focus of attention by generating learning tasks, proposing experiments to gather training data, and choosing concepts to acquire (e.g., Mitchell et al (Mitchell, 2006).

1.2. Importance of Machine Learning

These are benefits of machine learning and these are why research in machine learning is now what could not be avoided or neglected. Using machine learning techniques make life easier for computer users. These are the importance of machine learning. They are:

- Some tasks cannot be defined well except by example; that is we might be able to specify input and output pairs but not a concise relationship between inputs and desired outputs. We would like machines to be able to adjust their internal structure to produce correct outputs for a large number of sample inputs and thus suitably constrain their input and output function to approximate the relationship implicit in the examples.
- It is possible that hidden among large piles of data are important relationships and correlations. Machine learning methods can often be used to extract these relationships (data mining).
- Human designers often produce machines that do not work as well as desired in the environments in which they are used. In fact, certain characteristics of the working environment might not be completely known at design time. Machine learning methods can be used for on the job improvement of existing machine designs.

- The amount of knowledge available about certain tasks might be too large for explicit encoding by humans. Machines that learn this knowledge gradually might be able to capture more of it than humans would want to write down.
- Environments change over time. Machines that can adapt to a changing environment would reduce the need for constant redesign. New knowledge about tasks is constantly being discovered by humans. Vocabulary changes. There is a constant stream of new events in the world. Continuing redesign of AI systems to conform to new knowledge is impractical. But machine learning methods might be able to track much of it.

1.3 Machine Learning Varieties

Research in machine learning is now converging from several sources and from artificial intelligent field. These different traditions each bring different methods and different vocabulary which are now being assimilated into a more united discipline. Here is a brief listing of some of the separate disciplines that have contributed to machine learning (Nilsson, 1982).

- **Statistics:** A long-standing problem in statistics is how best to use samples drawn from unknown probability distributions to help decide from which distribution some new sample is drawn. A related problem is how to estimate the value of an unknown function at a new point given the values of this function at a set of sample points. Statistical methods for dealing with these problems can be considered instances of machine learning because the decision and estimation rules depend on a corpus of samples drawn from the problem environment. We will explore some of the statistical methods later in the book. Details about the statistical theory underlying these methods can be found in Orlitsky (Orlitsky, Santhanam, Viswanathan, & Zhang, 2005).
- **Brian Models:** Non linear elements with weighted inputs have been suggested as simple models of biological neurons. Networks of these elements have been studied by several researchers including (Rajesh P. N. Rao, 2002). Brain modelers are interested in how closely these networks approximate the learning phenomena of living brain. We shall see that several important machine learning techniques are based on networks of nonlinear elements often called neural networks. Work inspired by this school is some times called connectionism, brain-style computation or sub-symbolic processing.
- **Adaptive Control Theory:** Control theorists study the problem of controlling a process having unknown parameters which must be estimated during operation. Often, the parameters change during operation and the control process must track these changes. Some aspects of controlling a robot based on sensory inputs represent instances of this sort of problem.

- **Psychological Models:** Psychologists have studied the performance of humans in various learning tasks. An early example is the EPAM network for storing and retrieving one member of a pair of words when given another (Friedberg, 1958). Related work led to a number of early decision tree, (Hunt, 1966) and semantic network, (Anderson, 1995) methods. More recent work of this sort has been influenced by activities in artificial intelligence which we will be presenting. Some of the work in reinforcement learning can be traced to efforts to model how reward stimuli influence the learning of goal seeking behaviour in animals, (Richard S. Sutton, 1998). Reinforcement learning is an important theme in machine learning research.
- **Artificial Intelligence** From the beginning, AI research has been concerned with machine learning. Samuel developed a prominent early program that learned parameters of a function for evaluating board positions in the game of checkers. AI researchers have also explored the role of analogies in learning and how future actions and decisions can be based on previous exemplary cases. Recent work has been directed at discovering rules for expert systems using decision tree methods and inductive logic programming. Another theme has been saving and generalizing the results of problem solving using explanation based learning, (Mooney, 2000) ,(Y. Chali, 2009).
- **Evolutionary Models**
In nature, not only do individual animals learn to perform better, but species evolve to be better fit in their individual niches. Since the distinction between evolving and learning can be blurred in computer systems, techniques that model certain aspects of biological evolution have been proposed as learning methods to improve the performance of computer programs. Genetic algorithms and genetic programming (Oltan, 2005) are the most prominent computational techniques for evolution.

2. References

- Allix, N. M. (2003, April). Epistemology And Knowledge Management Concepts And Practices. *Journal of Knowledge Management Practice* .
- Alpaydin, E. (2004). *Introduction to Machine Learning*. Massachusetts, USA: MIT Press.
- Anderson, J. R. (1995). *Learning and Memory*. Wiley, New York, USA.
- Anil Mathur, G. P. (1999). Socialization influences on preparation for later life. *Journal of Marketing Practice: Applied Marketing Science* , 5 (6,7,8), 163 - 176.
- Ashby, W. R. (1960). *Design of a Brain, The Origin of Adaptive Behaviour*. John Wiley and Son.
- Batista, G. &. (2003). An Analysis of Four Missing Data Treatment Methods for Supervised Learning. *Applied Artificial Intelligence* , 17, 519-533.
- Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford, England: Oxford University Press.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*. New York, New York: Springer Science and Business Media.
- Block H, D. (1961). The Perceptron: A Model of Brian Functioning. 34 (1), 123-135.

- Carling, A. (1992). *Introducing Neural Networks*. Wilmslow, UK: Sigma Press.
- D. Michie, D. J. (1994). *Machine Learning, Neural and Statistical Classification*. Prentice Hall Inc.
- Fausett, L. (19994). *Fundamentals of Neural Networks*. New York: Prentice Hall.
- Forsyth, R. S. (1990). The strange story of the Perceptron. *Artificial Intelligence Review* , 4 (2), 147-155.
- Friedberg, R. M. (1958). A learning machine: Part, 1. *IBM Journal* , 2-13.
- Ghahramani, Z. (2008). Unsupervised learning algorithms are designed to extract structure from data. 178, pp. 1-8. IOS Press.
- Gillies, D. (1996). *Artificial Intelligence and Scientific Method*. OUP Oxford.
- Haykin, S. (19994). *Neural Networks: A Comprehensive Foundation*. New York: Macmillan Publishing.
- Hodge, V. A. (2004). A Survey of Outlier Detection Methodologies. *Artificial Intelligence Review* , 22 (2), 85-126.
- Holland, J. (1980). Adaptive Algorithms for Discovering and Using General Patterns in Growing Knowledge Bases Policy Analysis and Information Systems. 4 (3).
- Hunt, E. B. (1966). Experiment in Induction.
- Ian H. Witten, E. F. (2005). *Data Mining Practical Machine Learning and Techniques* (Second edition ed.). Morgan Kaufmann.
- Jaime G. Carbonell, R. S. (1983). Machine Learning: A Historical and Methodological Analysis. *Association for the Advancement of Artificial Intelligence* , 4 (3), 1-10.
- Kohonen, T. (1997). Self-Organizing Maps.
- Luis Gonz, I. A. (2005). Unified dual for bi-class SVM approaches. *Pattern Recognition* , 38 (10), 1772-1774.
- McCulloch, W. S. (1943). A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophysics* , 115-133.
- Mitchell, T. M. (2006). *The Discipline of Machine Learning*. Machine Learning Department technical report CMU-ML-06-108, Carnegie Mellon University.
- Mooney, R. J. (2000). Learning Language in Logic. In L. N. Science, *Learning for Semantic Interpretation: Scaling Up without Dumbing Down* (pp. 219-234). Springer Berlin / Heidelberg.
- Mostow, D. (1983). *Transforming declarative advice into effective procedures: a heuristic search example* In I? . S. Michalski,. Tioga Press.
- Nilsson, N. J. (1982). *Principles of Artificial Intelligence (Symbolic Computation / Artificial Intelligence)*. Springer.
- Oltean, M. (2005). Evolving Evolutionary Algorithms Using Linear Genetic Programming. 13 (3), 387 - 410 .
- Orlitsky, A., Santhanam, N., Viswanathan, K., & Zhang, J. (2005). Convergence of profile based estimators. *Proceedings of International Symposium on Information Theory. Proceedings. International Symposium on*, pp. 1843 - 1847. Adelaide, Australia: IEEE.
- Patterson, D. (19996). *Artificial Neural Networks*. Singapore: Prentice Hall.
- R. S. Michalski, T. J. (1983). *Learning from Observation: Conceptual Clustering*. TIOGA Publishing Co.
- Rajesh P. N. Rao, B. A. (2002). *Probabilistic Models of the Brain*. MIT Press.
- Rashevsky, N. (1948). *Mathematical Biophysics:Physico-Mathematical Foundations of Biology*. Chicago: Univ. of Chicago Press.
- Richard O. Duda, P. E. (2000). *Pattern Classification* (2nd Edition ed.).

- Richard S. Sutton, A. G. (1998). *Reinforcement Learning*. MIT Press.
- Ripley, B. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press.
- Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain . *Psychological Review* , 65 (6), 386-408.
- Russell, S. J. (2003). *Artificial Intelligence: A Modern Approach* (2nd Edition ed.). Upper Saddle River, NJ, NJ, USA: Prentice Hall.
- Ryszard S. Michalski, J. G. (1955). *Machine Learning: An Artificial Intelligence Approach (Volume I)*. Morgan Kaufmann .
- Ryszard S. Michalski, J. G. (1955). *Machine Learning: An Artificial Intelligence Approach*.
- Selfridge, O. G. (1959). Pandemonium: a paradigm for learning. *In The mechanisation of thought processes. H.M.S.O., London. London.*
- Sleeman, D. H. (1983). *Inferring Student Models for Intelligent CAI. Machine Learning*. Tioga Press.
- Tapas Kanungo, D. M. (2002). A local search approximation algorithm for k-means clustering. *Proceedings of the eighteenth annual symposium on Computational geometry* (pp. 10-18). Barcelona, Spain : ACM Press.
- Timothy Jason Shepard, P. J. (1998). Decision Fusion Using a Multi-Linear Classifier . *In Proceedings of the International Conference on Multisource-Multisensor Information Fusion*.
- Tom, M. (1997). *Machibe Learning*. Machine Learning, Tom Mitchell, McGraw Hill, 1997: McGraw Hill.
- Trevor Hastie, R. T. (2001). *The Elements of Statistical Learning*. New york, NY, USA: Springer Science and Business Media.
- Widrow, B. W. (2007). *Adaptive Inverse Control: A Signal Processing Approach*. Wiley-IEEE Press.
- Y. Chali, S. R. (2009). Complex Question Answering: Unsupervised Learning Approaches and Experiments. *Journal of Artificial Intelligent Research* , 1-47.
- Yu, L. L. (2004, October). Efficient feature Selection via Analysis of Relevance and Redundancy. *JMLR* , 1205-1224.
- Zhang, S. Z. (2002). Data Preparation for Data Mining. *Applied Artificial Intelligence*. 17, 375 - 381.