What Is Machine Learning?

1

Issam El Naqa and Martin J. Murphy

Abstract

Machine learning is an evolving branch of computational algorithms that are designed to emulate human intelligence by learning from the surrounding environment. They are considered the working horse in the new era of the so-called big data. Techniques based on machine learning have been applied successfully in diverse fields ranging from pattern recognition, computer vision, spacecraft engineering, finance, entertainment, and computational biology to biomedical and medical applications. More than half of the patients with cancer receive ionizing radiation (radiotherapy) as part of their treatment, and it is the main treatment modality at advanced stages of local disease. Radiotherapy involves a large set of processes that not only span the period from consultation to treatment but also extend beyond that to ensure that the patients have received the prescribed radiation dose and are responding well. The degrees of the complexity of these processes can vary and may involve several stages of sophisticated human-machine interactions and decision making, which would naturally invite the use of machine learning algorithms into optimizing and automating these processes including but not limited to radiation physics quality assurance, contouring and treatment planning, image-guided radiotherapy, respiratory motion management, treatment response modeling, and outcomes prediction. The ability of machine learning algorithms to learn from current context and generalize into unseen tasks would allow improvements in both the safety and efficacy of radiotherapy practice leading to better outcomes.

I. El Naqa (✉)
Department of Oncology, McGill University, Montreal, QC, Canada

Department of Radiation Oncology, University of Michigan, Ann Arbor, USA
e-mail: issam.elnaqa@mcgill.ca; ielnaqa@med.umich.edu

M.J. Murphy
Department of Radiation Oncology, Virginia Commonwealth University, Richmond, VA, USA
e-mail: MMurphy@mcvh-vcu.edu

I. El Naqa et al. (eds.), *Machine Learning in Radiation Oncology: Theory and Applications*, DOI 10.1007/978-3-319-18305-3_1

1.1 Overview

A machine learning algorithm is a computational process that uses input data to achieve a desired task without being literally programmed (i.e., "hard coded") to produce a particular outcome. These algorithms are in a sense "soft coded" in that they automatically alter or adapt their architecture through repetition (i.e., experience) so that they become better and better at achieving the desired task. The process of adaptation is called training, in which samples of input data are provided along with desired outcomes. The algorithm then optimally configures itself so that it can not only produce the desired outcome when presented with the training inputs, but can generalize to produce the desired outcome from new, previously unseen data. This training is the "learning" part of machine learning. The training does not have to be limited to an initial adaptation during a finite interval. As with humans, a good algorithm can practice "lifelong" learning as it processes new data and learns from its mistakes.

There are many ways that a computational algorithm can adapt itself in response to training. The input data can be selected and weighted to provide the most decisive outcomes. The algorithm can have variable numerical parameters that are adjusted through iterative optimization. It can have a network of possible computational pathways that it arranges for optimal results. It can determine probability distributions from the input data and use them to predict outcomes.

The ideal of machine learning is to emulate the way that human beings (and other sentient creatures) learn to process sensory (input) signals in order to accomplish a goal. This goal could be a task in pattern recognition, in which the learner wants to distinguish apples from oranges. Every apple and orange is unique, but we are still able (usually) to tell one from the other. Rather than hard code a machine with many, many exact representations of apples and oranges, it can be programmed to learn to distinguish them through repeated experience with actual apples and oranges. This is a good example of *supervised learning*, in which each training example of input data (color, shape, odor, etc.) is paired with its known classification label (apple or orange). It allows the learner to deal with similarities and differences when the objects to be classified have many variable properties within their own classes but still have fundamental qualities that identify them. Most importantly, the successful learner should be able to recognize an apple or an orange that it has never seen before.

A second type of machine learning is the so-called *unsupervised algorithm*. This might have the objective of trying to throw a dart at a bull's-eye. The device (or human) has a variety of degrees of freedom in the mechanism that controls the path of the dart. Rather than try to exactly program the kinematics a priori, the learner practices throwing the dart. For each trial, the kinematic degrees of freedom are adjusted so that the dart gets closer and closer to the bull's-eye. This is unsupervised in the sense that the training doesn't associate a particular kinematic input configuration with a particular outcome. The algorithm finds its own way from the training input data. Ideally, the trained dart thrower will be able to adjust the learned kinematics to accommodate, for instance, a change in the position of the target.

A third type of machine learning is *semi-supervised learning*, where part of the data is labeled and other parts are unlabeled. In such a scenario, the labeled part can be used to aid the learning of the unlabeled part. This kind of scenario lends itself to most processes in nature and more closely emulates how humans develop their skills.

There are two particularly important advantages to a successful algorithm. First, it can substitute for laborious and repetitive human effort. Second, and more significantly, it can potentially learn more complicated and subtle patterns in the input data than the average human observer is able to do. Both of these advantages are important to radiation therapy. For example, the daily contouring of tumors and organs at risk during treatment planning is a time-consuming process of pattern recognition that is based on the observer's familiarity and experience with the appearance of anatomy in diagnostic images. That familiarity, though, has its limits, and consequently, there are uncertainty and interobserver variability in the resulting contours. It is possible that an algorithm for contouring can pick up subtleties of texture or shape in one image or simultaneously incorporate data from multiple sources or blend the experience of numerous observers and thus reduce the uncertainty in the contour.

More than half of the patients with cancer receive ionizing radiation (radiotherapy) as part of their treatment, and it is the main treatment modality at advanced stages of disease. Radiotherapy involves a large set of processes that not only span the period from consultation to treatment but also extend beyond, to ensure that the patients have received the prescribed radiation dose and are responding well. The complexity of these processes can vary and may involve several stages of sophisticated human-machine interactions and decision making, which would naturally invite the use of machine learning algorithms to optimize and automate these processes, including but not limited to radiation physics quality assurance, contouring and treatment planning, image-guided radiotherapy, respiratory motion management, treatment response modeling, and outcomes prediction.

1.2 Background

Machine learning is the technology of developing computer algorithms that are able to emulate human intelligence. It draws on ideas from different disciplines such as artificial intelligence, probability and statistics, computer science, information theory, psychology, control theory, and philosophy [1–3]. This technology has been applied in such diverse fields as pattern recognition [3], computer vision [4], spacecraft engineering [5], finance [6], entertainment [7, 8], ecology [9], computational biology [10, 11], and biomedical and medical applications [12, 13]. The most important property of these algorithms is their distinctive ability to learn the surrounding environment from input data with or without a teacher [1, 2].

Historically, the inception of machine learning can be traced to the seventeenth century and the development of machines that can emulate human ability to add and subtract by Pascal and Leibniz [14]. In modern history, Arthur Samuel from IBM coined the term "machine learning" and demonstrated that computers could be programmed to learn to play checkers [15]. This was followed by the development of the perceptron by Rosenblatt as one of the early neural network architectures in 1958 [16]. However, early enthusiasm about the perceptron was dampened by the observation made by Minsky that the perceptron classification ability is limited to linearly separable problems and not common nonlinear problems such as a simple XOR logic [17]. A breakthrough was achieved in 1975 by the development of the multilayer perceptron (MLP) by Werbos [18]. This was followed by the development of decision trees by Quinlan in 1986 [19] and support vector machines by Cortes and Vapnik [20]. Ensemble machine learning algorithms, which combine multiple learners, were subsequently proposed, including Adaboost [21] and random forests [22]. More recently, distributed multilayered learning algorithms have emerged under the notion of deep learning [23]. These algorithms are able to learn good representations of the data that make it easier to extract useful information when building classifiers or other predictors [24].

1.3 Machine Learning Definition

The field of machine learning has received several formal definitions in the literature. Arthur Samuel in his seminal work defined machine learning as "a field of study that gives computers the ability to learn without being explicitly programmed" [15]. Using a computer science lexicon, Tom Mitchell presented it as "A computer program is said to learn from experience (*E*) with respect to some class of tasks (*T*) and performance measure (*P*), if its performance at tasks in *T*, as measured by *P*, improves with experience *E*" [1]. Ethem Alpaydin in his textbook defined machine learning as the field of "Programming computers to optimize a performance criterion using example data or past experience" [2]. These various definitions share the notion of coaching computers to intelligently perform tasks beyond traditional number crunching by learning the surrounding environment through repeated examples.

1.4 Learning from Data

The ability to learn through input from the surrounding environment, whether it is playing checkers or chess games, or recognizing written patterns, or solving the daunting problems in radiation oncology, is the main key to developing a successful machine learning application. Learning is defined in this context as estimating dependencies from data [25].

The fields of data mining and machine learning are intertwined. Data mining utilizes machine learning algorithms to interrogate large databases and discover hidden knowledge in the data, while many machine learning algorithms employ data mining methods to preprocess the data before learning the desired tasks [26]. However, it should be noted that machine learning is not limited to solving database-like problems but also extends into solving complex artificial intelligence challenges by learning and adapting to a dynamically changing situation, as is encountered in a busy radiation oncology practice, for instance.

Machine learning has both engineering science aspects such as data structures, algorithms, probability and statistics, and information and control theory and social science aspects by drawing in ideas from psychology and philosophy.

1.5 Overview of Machine Learning Approaches

Machine learning can be divided according to the nature of the data labeling into supervised, unsupervised, and semi-supervised as shown in Fig. 1.1. Supervised learning is used to estimate an unknown (input, output) mapping from known (input, output) samples, where the output is labeled (e.g., classification and regression). In unsupervised learning, only input samples are given to the learning system (e.g., clustering and estimation of probability density function). Semi-supervised learning is a combination of both supervised and unsupervised where part of the data is partially labeled and the labeled part is used to infer the unlabeled portion (e.g., text/image retrieval systems).

From a concept learning perspective, machine learning can be categorized into transductive and inductive learning [27]. Transductive learning involves the inference from specific training cases to specific testing cases using discrete labels as in clustering or using continuous labels as in manifold learning. On the other hand, inductive learning aims to predict outputs from inputs that the learner has not encountered before. Along these lines, Mitchell argues for the necessity of an

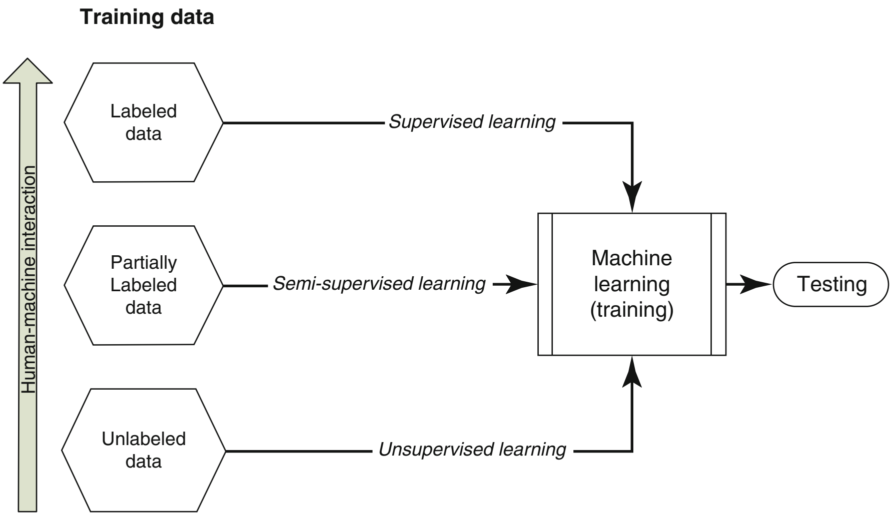

Fig. 1.1 Categories of machine learning algorithms according to training data nature

inductive bias in the training process to allow for a machine learning algorithm to generalize beyond unseen observation [28].

From a probabilistic perspective, machine learning algorithms can be divided into discriminant or generative models. A discriminant model measures the conditional probability of an output given typically deterministic inputs, such as neural networks or a support vector machine. A generative model is fully probabilistic whether it is using a graph modeling technique such as Bayesian networks, or not as in the case of naïve Bayes.

Another interesting class of machine learning algorithms that attempts to control learning by accommodating a feedback system is reinforcement learning, in which an agent attempts to take a sequence of actions that may maximize a cumulative reward such as winning a game of checkers, for instance [29]. This kind of approach is particularly useful for online learning applications.

1.6 Application in Biomedicine

Machine learning algorithms have witnessed increased use in biomedicine, starting naturally in neuroscience and cognitive psychology through the seminal work of Donald Hebb in his 1949 book [30] developing the principles of associative or Hebbian learning as a mechanism of neuron adaptation and the work of Frank Rosenblatt developing the perceptron in 1958 as an intelligent agent [16]. More recently, machine learning algorithms have been widely applied in breast cancer detection and diagnosis [31–33]. Reviews of the application of machine learning in biomedicine and medicine can be found in [12, 13].

1.7 Application in Medical Physics and Radiation Oncology

Early applications of machine learning in radiation oncology focused on predicting normal tissue toxicity [34–36], but its application has since branched into almost every part of the field, including tumor response modeling, radiation physics quality assurance, contouring and treatment planning, image-guided radiotherapy, respiratory motion management, as seen from the examples presented in this book.

1.8 Steps to Machine Learning Heaven

For the successful application of machine learning in general and in medical physics and radiation oncology in particular, one first needs to properly characterize the nature of problem, in terms of the input data and the desired outputs. Secondly, despite the robustness of machine learning to noise, a good model cannot substitute for bad data, keeping in mind that models are primarily built on approximations, and it has been stated that "All models are wrong; some models are useful (George

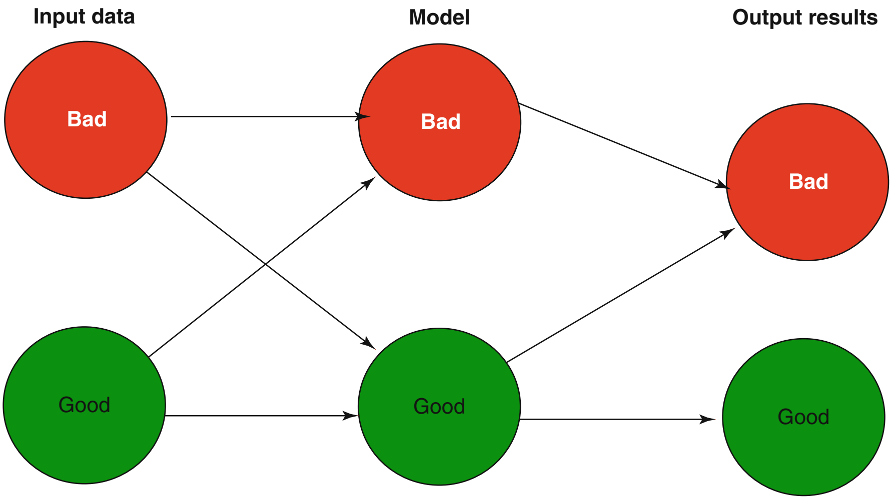

Fig.1.2 GIGO paradigm. Learners cannot be better than the data

Box)." Additionally, this has been stated as the GIGO principle, garbage in garbage out as shown in Fig. 1.2 [37].

Thirdly, the model needs to generalize beyond the observed data into unseen data, as indicated by the inductive bias mentioned earlier. To achieve this goal, the model needs to be kept as simple as possible but not simpler, a property known as parsimony, which follows from Occam's razor that "Among competing hypotheses, the hypothesis with the fewest assumptions should be selected." Analytically, the complexity of a model could be derived using different metrics such as Vapnik–Chervonenkis (VC) dimension discussed in chapter 2. for instance [25]. Finally, a major limitation in the acceptance of machine learning by the larger medical community is the "black box" stigma and the inability to provide an intuitive interpretation of the learned process that could help clinical practitioners better understand their data and trust the model predictions. This is an active and necessary area of research that requires special attention from the machine learning community working in biomedicine.

Conclusions

Machine learning presents computer algorithms that are able to learn from the surrounding environment to optimize the solution for the task at hand. It builds on expertise from diverse fields such as artificial intelligence, probability and statistics, computer science, information theory, and cognitive neuropsychology. Machine learning algorithms can be categorized into different classes according to the nature of the data, the learning process, and the model type. Machine learning has a long history in biomedicine, but its application in medical physics and radiation oncology is in its infancy, with high potential and promising future to improve the safety and efficacy of radiotherapy practice.

References

1. Mitchell TM. Machine learning. New York: McGraw-Hill; 1997.
2. Alpaydin E. Introduction to machine learning. 3rd ed. Cambridge, MA: The MIT Press; 2014.
3. Bishop CM. Pattern recognition and machine learning. New York: Springer; 2006.
4. Apolloni B. Machine learning and robot perception. Berlin: Springer; 2005.
5. Ao S-I, Rieger BB, Amouzegar MA. Machine learning and systems engineering. Dordrecht/New York: Springer; 2010.
6. Györfi L, Ottucsák G, Walk H. Machine learning for financial engineering. Singapore/London: World Scientific; 2012.
7. Gong Y, Xu W. Machine learning for multimedia content analysis. New York/London: Springer; 2007.
8. Yu J, Tao D. Modern machine learning techniques and their applications in cartoon animation research. 1st ed. Hoboken: Wiley; 2013.
9. Fielding A. Machine learning methods for ecological applications. Boston: Kluwer Academic Publishers; 1999.
10. Mitra S. Introduction to machine learning and bioinformatics. Boca Raton: CRC Press; 2008.
11. Yang ZR. Machine learning approaches to bioinformatics. Hackensack: World Scientific; 2010.
12. Cleophas TJ. Machine learning in medicine. New York: Springer; 2013.
13. Malley JD, Malley KG, Pajevic S. Statistical learning for biomedical data. Cambridge: Cambridge University Press; 2011.
14. Ifrah G. The universal history of computing: from the abacus to the quantum computer. New York: John Wiley; 2001.
15. Samuel AL. Some studies in machine learning using the game of checkers. IBM: J Res Dev. 1959;3:210–29.
16. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. Psychol Rev. 1958;65:386–408.
17. Minsky ML, Papert S. Perceptrons; an introduction to computational geometry. Cambridge, MA: MIT Press; 1969.
18. Werbos PJ. Beyond regression: new tools for prediction and analysis in the behavioral sciences; PhD thesis, Harvard University, 1974.
19. Quinlan JR. Induction of decision trees. Mach Learn. 1986;1:81–106.
20. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20:273–97.
21. Schapire RE. A brief introduction to boosting. In: Proceedings of the 16th international joint conference on artificial intelligence, vol. 2. Stockholm: Morgan Kaufmann Publishers Inc; 1999. p. 1401–6.
22. Breiman L. Random forests. Mach Learn. 2001;45:5–32.
23. Hinton GE. Learning multiple layers of representation. Trends Cogn Sci. 2007;11:428–34.
24. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell. 2013;35:1798–828.
25. Cherkassky VS, Mulier F. Learning from data: concepts, theory, and methods. 2nd ed. Hoboken: IEEE Press/Wiley-Interscience; 2007.
26. Kargupta H. Next generation of data mining. Boca Raton: CRC Press; 2009.
27. Vapnik VN. Statistical learning theory. New York: Wiley; 1998.
28. Mitchell TM. The need for biases in learning generalizations. New Brunswick: Rutgers University; 1980.
29. Sutton RS, Barto AG. Reinforcement learning: an introduction. Cambridge, MA: MIT Press; 1998.
30. Hebb DO. The organization of behavior; a neuropsychological theory. New York: Wiley; 1949.
31. El-Naqa I, Yang Y, Wernick MN, Galatsanos NP, Nishikawa RM. A support vector machine approach for detection of microcalcifications. IEEE Trans Med Imaging. 2002;21: 1552–63.

32. Gurcan MN, Chan HP, Sahiner B, Hadjiiski L, Petrick N, Helvie MA. Optimal neural network architecture selection: improvement in computerized detection of microcalcifications. Acad Radiol. 2002;9:420–9.
33. El-Naqa I, Yang Y, Galatsanos NP, Nishikawa RM, Wernick MN. A similarity learning approach to content-based image retrieval: application to digital mammography. IEEE Trans Med Imaging. 2004;23:1233–44.
34. Gulliford SL, Webb S, Rowbottom CG, Corne DW, Dearnaley DP. Use of artificial neural networks to predict biological outcomes for patients receiving radical radiotherapy of the prostate. Radiother Oncol. 2004;71:3–12.
35. Munley MT, Lo JY, Sibley GS, Bentel GC, Anscher MS, Marks LB. A neural network to predict symptomatic lung injury. Phys Med Biol. 1999;44:2241–9.
36. Su M, Miften M, Whiddon C, Sun X, Light K, Marks L. An artificial neural network for predicting the incidence of radiation pneumonitis. Med Phys. 2005;32:318–25.
37. Tweedie R, Mengersen K, Eccleston J. Garbage in, garbage out: can statisticians quantify the effects of poor data? Chance. 1994;7:20–7.