Machine Learning: Trends, Perspectives, and Prospects

[1]Sujeet Kumar , [2]Pradeep Kumar Sharma

[1]Assistant Professor
Katihar Engineering College, Katihar
[2]Assistant Professor,
Katihar Engineering College, Katihar

Abstract: The challenge of how to create computers that automatically get better through experience is addressed by machine learning. Situated in the nexus of computer science and statistics, as well as the foundation of artificial intelligence and data science, it is one of the rapidly expanding technical topics of today. Current advances in machine learning have been fueled by the creation of novel learning theories and algorithms as well as by the constant proliferation of low-cost computing and internet data. In science, technology, and business, data-intensive machine learning techniques are being adopted, which is resulting in a rise in the use of evidence in decision-making in a variety of fields, including marketing, manufacturing, healthcare, education, financial modeling, and law enforcement.
Keywords: Machine Learning, Artificial Intelligence, Models, Process of ML

Introduction

The field of machine learning is concerned with two linked questions: How can one build computer systems that do things on their own? Enhance with time and experience? Additionally which basic laws of statistics, computation, and information theory apply to all learning systems, including those in computers, people, and organizations? The study of machine learning is crucial for answering these fundamental problems in science and engineering as well as for the incredibly useful computer software that it has generated and implemented in numerous fields.

Over the past 20 years, machine learning has advanced significantly from being a lab curiosity to a useful technology with broad commercial use. Machine learning has become the go-to technique in artificial intelligence (AI) for creating useful software for tasks like computer vision, audio recognition, natural language processing, robot control, and other applications.

Nowadays, a lot of AI system engineers understand that, depending on the application, it can sometimes be far simpler to train a system by providing it with instances of desired input-output behavior than to manually programme it by predicting the correct response for every conceivable input. Machine learning has also had a significant impact on computer science and a variety of other industries that deal with data-intensive problems, like consumer services, the identification of problems in intricate systems, and the management of supply chains. Similar wide-ranging effects have been seen in other empirical sciences, such as biology, cosmology, and social science, as machine-learning techniques have been created to analyze large-scale experimental data in unique ways. A list of some recent uses for machine learning can be found in Fig. 1.

The challenge of increasing a performance metric during task execution through some kind of training experience is known as a learning problem. For instance, one objective in learning to identify credit card fraud is to categories each credit card transaction as either "fraud" or "not fraud." The accuracy of this fraud classifier may be the performance indicator that needs to be improved, and the training set might be a set of past credit card transactions that have all been flagged as fraudulent or not in hindsight. Alternatively, an alternative performance metric might be defined, one that penalizes more when "fraud" is mistakenly labeled as "not fraud" than when "not fraud" is mistakenly labeled as "fraud." Another way to define a different kind of training experience would be to include instances of labeled and unlabeled credit card transactions.

To address the broad range of data and problem types seen in many machine-learning challenges, a variety of machine-learning algorithms have been created (1, 2). From a conceptual standpoint, machine-learning algorithms can be understood as sifting through a vast array of potential programmes in order to choose one that maximizes the performance metric, with the use of training data.

The representation of candidate programmes (such as decision trees, mathematical functions, and general programming languages) and the methods used to search through this space of programmes (such as evolutionary search methods and optimization algorithms with well-understood convergence guarantees) are two aspects of machine-learning algorithms that differ greatly from one another. Here, we concentrate on strategies that have shown to be especially effective thus far.

Several algorithms concentrate on function approximation problems, in which the learning problem is to increase the accuracy of a function that is embodied in an input transaction (e.g., given an input transaction, output a "fraud" or "not fraud" label). Experience is comprised of a sample of known input-output pairs of the function. The function can be expressed directly in certain situations as a parameterized functional form; implicitly, it can be found by a factorization, optimization, or search procedure, Process or one that is based on simulation. The function typically depends on parameters or other adjustable degrees of freedom, even when it is implicit. And training is the process of determining these parameters' values in order to maximize the performance metric.

Theoretically characterizing the skills of certain learning algorithms and the intrinsic complexity of any given learning problem is a crucial scientific and practical goal, regardless of the learning algorithm: To what extent is the algorithm able to learn from a given kind and quantity of training data? To what extent is the algorithm resistant to errors in the training data or in the modeling assumptions? Is it possible to create a successful solution for a learning problem given a certain amount of training data, or is the learning problem inherently unsolvable? These theoretical explanations of machine-learning algorithms and issues usually draw on the well-known frameworks of computational complexity theory and statistical decision theory. The goal is to simultaneously characterize the sample complexity (how much data are required to learn accurately) and the computational complexity (how much computation is required), and to specify how these depend on features of the learning algorithm, such as the representation it uses for what it learns. In fact, attempts to theoretically characterize machine-learning algorithms have led to blends of statistical and computational theory (3–6). Optimization theory is a particular type of computational analysis that has shown to be especially helpful recently. Its upper and lower bounds on the rates of convergence of optimization procedures combine well with the formulation of machine-learning problems as the optimization of a performance metric (7, 8). Machine learning as a topic of research is at the intersection of computer science, statistics, and several other disciplines that focus on inference and decision-making under uncertainty, as well as automatic improvement over time. The study of human learning psychology, evolutionary biology, adaptive control theory, educational methods, neuroscience, organizational behavior, and economics are examples of related fields.

While there has been more communication between these domains over the last 10 years, we have only begun to explore the possible synergies and the variety of formalisms and experimental techniques that are employed in these many fields to research systems that get better with time.

Factors advancing machine learning:

The ability of networked and mobile computer systems to collect and transfer enormous volumes of data has grown rapidly over the past 10 years; this phenomenon is frequently referred to as "Big Data." Scientists and engineers who gather these kinds of data have frequently looked to machine learning to find answers to the challenge of extracting meaningful information, forecasts, and choices from these kinds of data sets. The sheer volume of data necessitates the development of scalable processes that combine statistical and computational methods aspects, but the problem is not just in the volume of contemporary data sets but also in the fact that a large portion of them are granular and customized. Transportable Large volumes of data on specific people can be collected via devices and embedded computing, and machine learning algorithms can use this information to tailor their services to the requirements and situations of each unique user. Additionally, by connecting these customized services, a broader service that leverages the quantity and variety of data from several users while still being tailored to their specific wants and situations can be created. Many sectors of business, research, and government have examples of this tendency towards collecting and analyzing vast amounts of data to boost productivity and services. Historical crime data is used to help assign local police to specific locations at specific times; historical traffic data is used to improve traffic control and reduce congestion; large experimental data sets are captured and curate to accelerate progress in biology, astronomy, neuroscience, and other data intensive empirical sciences. Historical medical records are used to determine which patients will respond best to which treatments. It seems like we are only beginning a long-term trend towards more evidence-based, data-intensive decision-making in many facets of government, business, and science.

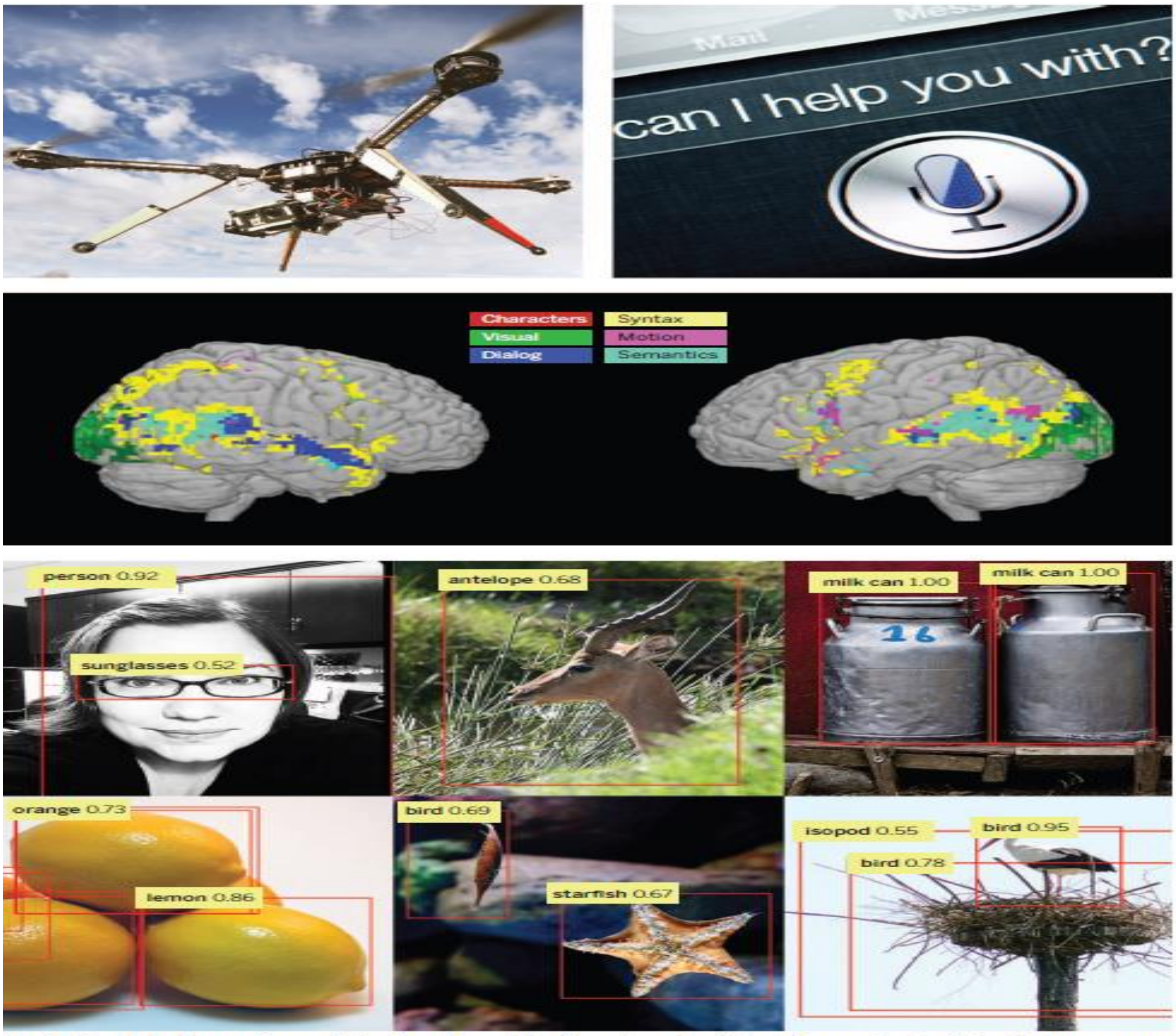

Fig.1 *Machine learning applications Many fields of science and technology are being significantly impacted by machine learning; recent examples of successful applications include speech and natural language processing (top left), robotics and autonomous vehicle control (top left), and (bottom), computer vision applications (middle), and brain research (right). [The panel in the middle is taken from (29). R. Girshick annotated the images in the bottom panel with object identification; the images are from the ImageNet collection.]*

Large-scale data is becoming more and more important in all spheres of human Endeavour, which has put additional demands on the underlying machine learning algorithms. Large data sets necessitate algorithms that are computationally tractable, highly personal data demands algorithms that minimize privacy consequences, and the availability of vast amounts of unlabeled data makes it difficult to create learning algorithms that can leverage it. The ensuing sections examine the impact of these needs on the latest advancements in machine-learning algorithms, theory, and application.

Fundamental techniques and current advancements: Supervised learning techniques are the most used machine-learning approaches (1). Supervised learning systems, such as email spam classifiers, facial recognition systems over images, and patient diagnostic systems, all represent the function approximation problem previously discussed. In these systems, the training data is a set of (x, y) pairs, and the objective is to generate a prediction y* in response to a query x*. The inputs could be simpler objects like texts, photos, graphs, or DNA sequences, or they could be more complicated objects like classical vectors. In a similar vein, other varieties of output y have been investigated.

Many studies have been conducted on problems such as multiclass classification (where y takes on one of K labels), multilevel classification (where y is labeled simultaneously by several of the K labels), ranking problems (where y provides a partial order on some set), and general structured prediction problems (where y is a combinatorial object such as a graph, whose components may be required to satisfy some set of constraints). A great deal of progress has been made by concentrating on the simple binary classification problem, in which y takes on one of two values (for example, "spam" or "not spam"). Part-of-speech tagging, where the objective is to concurrently label every word in an input phrase x as being a noun, verb, or other part of speech, is an example of the latter problem. Real-valued components of y or a combination of discrete and real-valued components are also included in supervised learning scenarios.

The majority of the time, supervised learning systems use a learnt mapping f (x) to generate an output y (or a probability distribution over y given x) for every input x. There are numerous methods for mapping functions, such as logistic regression, decision forests, decision trees, support vector machines, neural networks, kernel machines, and Bayesian classifiers (1). To estimate these various mapping types, a range of learning algorithms has been presented. Additionally, generic methods like boosting and multiple kernel learning combine the results of numerous learning algorithms.

Methods for learning f from data frequently draw on concepts from numerical analysis or optimization theory; advances are driven by the unique nature of machine learning issues, such as the fact that the objective function or function to be integrated is frequently the sum over a large number of components. The variety of learning algorithms and architectures reflects the variety of application requirements; different architectures capture various mathematical structures, provide varying trade-offs between computational complexity, data volume, and performance, and offer varying degrees of amenability to post-hoc visualization and explanation. Deep networks, which are multilayer networks of threshold units, each of which computes some basic parameterized function of its inputs, are one significant area of recent advancements in supervised learning (9, 10). Gradient-based optimization techniques are used by deep learning systems to modify the parameters of such a multilayered network in response to faults at the output. By utilizing contemporary parallel computing architectures, like graphics processing units initially designed for video games, deep learning systems with billions of parameters have been constructed. These systems can be trained using the vast arrays of images, videos, and speech samples that are accessible on the Internet. Large-scale deep learning systems have produced significant increases in performance over earlier methods in computer vision (11) and speech recognition (12). These systems have had a significant impact in these fields in recent years.

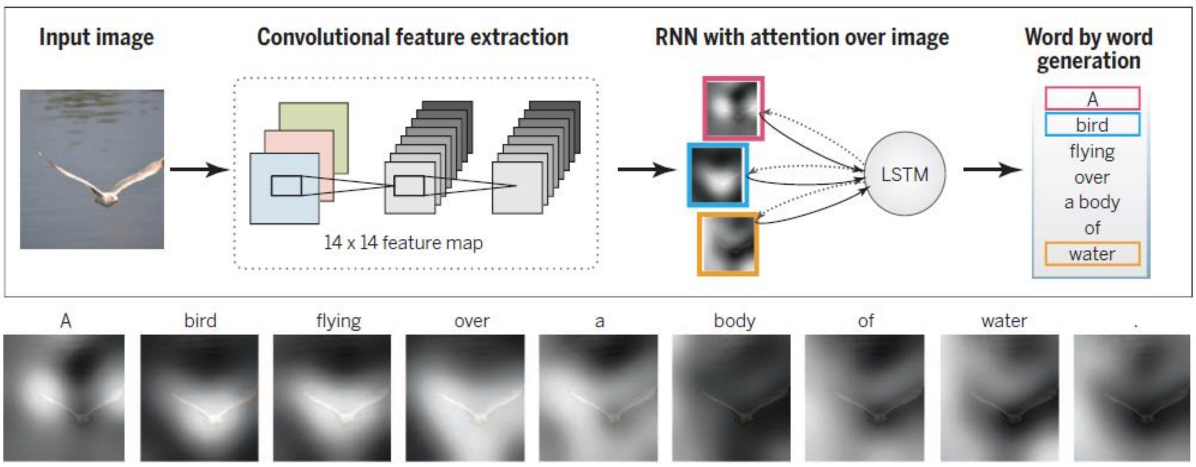

Fig. 2. *Deep networks automatically generate text labels for photos. A recurrent neural network trained to produce a written caption (top) uses the output of a convolution neural network taught to understand photos.*

The sequence at the bottom illustrates how the network focused word-by-word on various areas of the input image to create the caption, word by word. [Reproduced from (30) with permission]

One way to think of deep networks' core layers is as learnt representations of the incoming data. Deep learning algorithms that find useful representations of the input without the need for labeled training data have also been developed, even though supervised learning techniques have accounted for a large portion of the field's practical success in deep learning (13). In machine-learning research, the overall issue is known as unsupervised learning, or the second paradigm (2).

In general, unsupervised learning entails the examination of unlabeled data while making assumptions about the data's structural characteristics, such as algebraic, combinatorial, or probabilistic aspects.

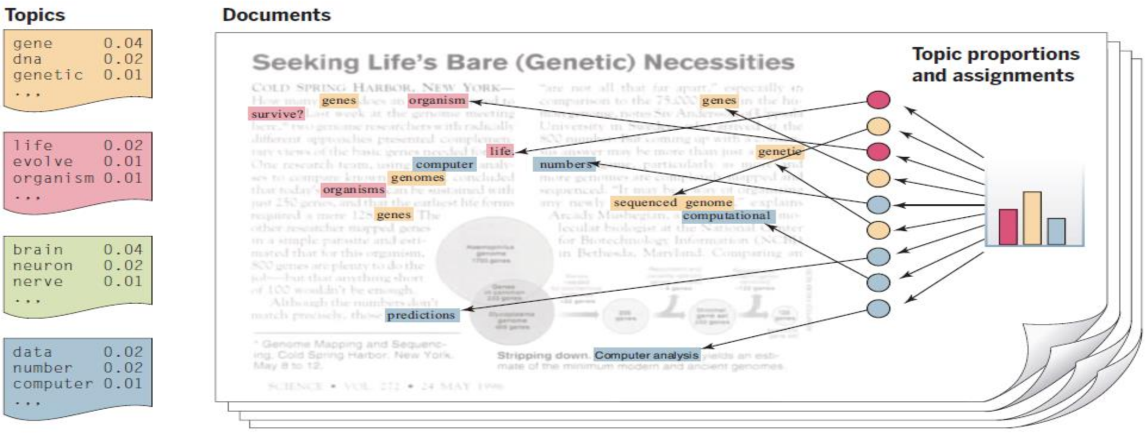

Figure 3. *Topic models in A document is seen as a collection of words when using the topic modeling analysis methodology, and each word is thought to have been produced by an underlying set of subjects (shown by the colors in the figure). Each document is characterized by a probability distribution over subjects (histogram), which are topics and probability distributions across words (leftmost column). These allocations are deduced from examining a set of documents and may be used to categorize, index, and condense the text of those documents. [From(31). Association for Computing Machinery, Inc. all rights reserved. Reprinted with consent]*

Assuming, for instance, that the data are on a low-dimensional manifold, one can seek to clearly identify that manifold from the data. Principal components analysis, manifold learning, factor analysis, random projections, and auto encoders are a few examples of dimension reduction techniques that make various specific assumptions about the underlying manifold, such as that it is a smooth nonlinear manifold, a collection of sub manifolds, or a linear subspace. The topic modeling framework shown in Fig. 3 is another illustration of dimension reduction.

Following the definition of a criterion function that incorporates these presumptions, optimization or sampling algorithms are created to maximize the criterion. These algorithms frequently make use of general statistical concepts like maximum likelihood, the technique of moments, or Bayesian integration.

Another way to look about clustering is as the problem of partitioning the observed data (and a rule for predicting future data) when there aren't any labels explicitly saying which division is intended. Numerous clustering techniques have been created, all of which are predicated on particular notions about what constitutes a "cluster." Computational complexity is a major problem in both dimension reduction and clustering because the idea is to take advantage of the extra huge data sets that become accessible when supervised labels are not used.

Reinforcement learning is a third major paradigm in machine learning (14, 15). In this instance, the data found in the training set falls somewhere between supervised and unsupervised learning. The training data in reinforcement learning are assumed to provide merely a hint as to whether an action is proper or not; if an action is incorrect, there remains the difficulty of identifying the correct action. This is in contrast to training examples that show the correct output for a given input. More generally, it is presumed that reward signals in the context of input sequences refer to the entire sequence; credit or blame for individual actions in the sequence is not explicitly given. In fact, while bandit problems—simplified versions of reinforcement learning—are studied, these problems typically involve a general control-theoretic setting in which the learning task is to learn a control strategy (a "policy") for an agent acting in an unknown dynamical environment. The learned strategy is then trained to choose actions for any given state with the goal of maximizing its expected reward over time. Over time, there has been a growing connection between control theory and operations research, thanks to the development of formulations like partially observed Markov decision processes and Markov decision processes (15, 16). The principles of policy iteration, value iteration, rollouts, and variance reduction are commonly used in reinforcement-learning algorithms, with modifications made to meet the unique requirements of machine learning (e.g., large-scale problems, few assumptions about the unknown dynamical environment, and the use of supervised learning architectures to represent policies). The use of reinforcement learning algorithms to predict the response of dopaminergic neurons in monkeys learning to associate a stimulus light with a subsequent sugar reward is one prominent example of the strong connections between decades of work on learning in psychology and neuroscience (17).

Despite the fact that these three learning paradigms aid in concept organization, a lot of recent research incorporates combinations of these categories. For instance, discriminative training combines structures created for unsupervised learning with optimization formulations that employ labels; while semi supervised learning uses unlabeled data to supplement labeled data in a supervised learning setting. The process of using training data to choose from a family of models as well as to fit a model is known as model selection. It is important to note that training data do not always clearly suggest Choosing a model results in the application of Bayesian optimization techniques and algorithms designed for bandit situations. Engaged education

Emerges when the student is given the option to select data points and ask the instructor for specific information, like the label of an example that isn't otherwise labeled. The goal of causal modeling is to identify which variables causally influence other variables, as opposed to just finding predictive relationships between them (for example, a high white blood cell count can predict the presence of an infection, but the infection is what causes the high white blood cell count). Across all of these paradigms, a variety of factors affect how learning algorithms are designed, such as whether data are available in batches or arrive sequentially over time, how data have been sampled, the need for users to be able to understand the learned models, and robustness problems that occur when data deviate from pre-existing modeling assumptions.

New developments:

Because machine learning is still relatively new, it is growing quickly, frequently through the creation of novel formalizations for problems in the area that are motivated by real-world applications. (The creation of recommendation systems, as shown in Fig. 4, is one example.) Concern for the environment in which machine-learning algorithms function is one of the main trends propelling this expansion. While a classical machine-learning system consisted of a single programme running on a single machine, it is now common for machine-learning systems to be deployed in architectures that include thousands or tens of thousands of processors, meaning that communication constraints and issues of parallelism and distributed processing take centre stage. This means that the word "environment" in this context also refers to the computing architecture. In fact, as Fig. 5 illustrates, machine-learning systems are becoming more and more like intricate software suites that operate on massively parallel and distributed computing platforms, offering data analysts a variety of services and methods.

The term "environment" can also refer to the data source, which can include a group of individuals who might be concerned about their privacy or ownership, an analyst or decision-maker who might have specific needs from a machine-learning system (like that the output be able to be visualized), or the social, legal, or political context in which the system is being implemented.

Other agents or machine learning systems may also be a part of the environment, and the entire group of systems may be antagonistic or cooperative. In general, environments give a learning algorithm access to a variety of resources while imposing limitations on those resources. Researchers in machine learning are formalizing these relationships more and more in an effort to create algorithms that can be proven to work well in a variety of settings and that explicitly let users express and manage resource trade-offs.

As an illustration of resource limitations, consider a scenario in which a group of people who value their privacy supply the data. The concept of "differential privacy" can be used to formalize privacy. It describes a probabilistic channel between the data and the outside world such that an observer of the channel's output cannot conclusively determine whether or not specific individuals have given data (18). Traditionally, differential privacy has meant ensuring that queries to a privatized database (such as "what is the maximum balance across a set of accounts?") yield results that are almost identical to those given on the no private data.

In the context of machine learning, recent research has brought differential privacy into contact with predictions or other inferential claims (e.g., "given the data I've seen so far, what is the probability that a new transaction is fraudulent?") (19, 20). Users can select a desired level of privacy that considers the types of questions that will be asked of the data and their own personal utility for the answers by putting the overall design of a privacy-enhancing machine-learning system within a decision-theoretic framework. For instance, if a person's genome is being used to determine insurance rates, they might be prepared to disclose the majority of their genome to further study on a hereditary condition, but they might also demand more strict protections.

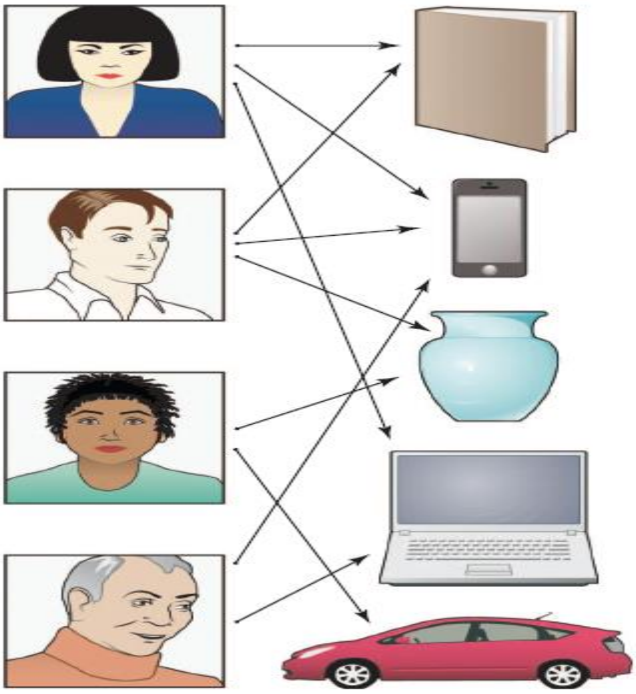

Figure 4: *Systems of recommendations. A machine-learning system that uses data to show connections between a group of users (people, for example) and a group of goods (products, for example) is called a recommendation system. When a user and a product are linked, it indicates that the user has expressed interest in the product in some way, possibly even by making a previous purchase. The challenge for machine learning is to use the data from all users to recommend other products that a particular user would find interesting.*

Another resource that must be controlled within the larger framework of a distributed learning system is communication. For instance, administrative boundaries or the bulk of the data preventing its consolidation at a single location may cause it to be dispersed throughout several physical sites. We could want to impose a bit-rate communication constraint on the machine-learning algorithm in such a scenario. By trading off these quantities against the amount of data, solving the design problem under such a constraint will typically demonstrate how the learning system's performance deteriorates under decreases in communication bandwidth. However, it can also demonstrate how the performance improves as the number of distributed sites (such as machines or processors) increases (21, 22). This field of study seeks to determine lower constraints on feasible performance and the particular algorithms that reach those lower bounds, much like in classical information theory.

Bringing the types of statistical resources explored in machine learning—such as the quantity of data points, the dimension of a parameter, and the complexity of a hypothesis class—into touch with the traditional computational resources of time and space is a primary objective of this overall area of research. A bridge of this kind can be found in the "probably approximately correct" (PAC) learning paradigm, which examines how the addition of a polynomial-time computation restriction affects the relationship between error rates, training data size, and other learning algorithm parameters (3). Recent developments in this field of study include a number of lower bounds that identify fundamental performance gaps that can be filled by using polynomial- and exponential-time methods to solve specific machine-learning problems (such as sparse principal components analysis and sparse regression) (23). However, the trade-offs between time and data at the heart of the issue are not near the polynomial/exponential limit. Algorithms whose time and space needs are linear or sub linear in the issue size (number of data points or number of dimensions) are needed for the enormous data sets that are becoming more and more common. Methods like random projections, algorithm weakening, and sub sampling are the subject of recent research because they allow for scalability without sacrificing statistical control (24, 25). The ultimate goal is to be able to provide machine-learning systems with budgets for time and space in addition to accuracy criteria, and have the system locate an operating point that enables the realization of such needs.

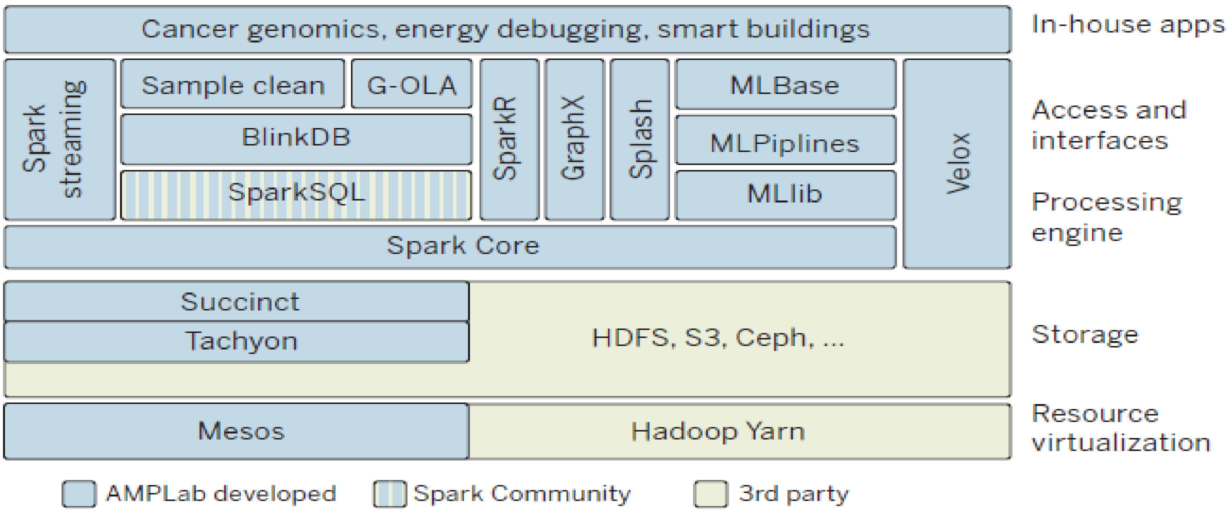

Figure 5: *Stack for data analytics. Distributed and parallel computing platforms serve as the foundation for layered architectures that support scalable machine learning systems. The architecture shown here, an open-source data analysis stack created at the University of California, Berkeley's Algorithms, Machines and People (AMP) Laboratory, consists of three layers: ones that provide distributed storage, data management, and processing; ones that interface with underlying operating systems; and one that offers fundamental machine-learning capabilities like streaming, sub sampling, pipelines, graph processing, and model serving.*

Possibilities and difficulties:

Even with its applications and business triumphs, machine learning is still a relatively new topic with a lot of untapped research potential. Comparing the Kinds of learning we see in naturally occurring systems with the state-of-the-art machine-learning techniques can highlight some of these prospects, Existing systems, including those involving people and other animals, businesses, economy, and biological evolution. For instance, whereas the majority of machine learning While humans clearly acquire a wide range of skills and knowledge through years of varied training experience, both supervised and unsupervised, in a simple-to-more-difficult sequence (e.g., learning to crawl, then walk, then run), algorithms are designed to learn a single specific function or data model from a single data source. This has prompted some researchers to start looking into the issue of how to build computer programmes that are perpetual or never-ending, capable of operating continuously for years and learning thousands of related skills or functions within an overall architecture that enhances the system's capacity to learn one skill based on the knowledge of another (26–28). The concept of mixed-initiative, team-based learning is also implied by the similarity to natural learning systems. For instance, people frequently work in teams to collect and analyze data, in contrast to the current machine learning systems that typically operate in isolation to analyze the given data (e.g., biologists have worked as teams to collect and analyze genomic data, bringing together diverse experiments and perspectives to make progress on this difficult problem). Artificial intelligence techniques that can collaborate with humans to analyze large, complex data sets could combine machine learning's ability to extract minute statistical patterns from large datasets with human analysis's capacity to draw from a variety of prior knowledge to produce new theories and plausible explanations. Many theoretical findings in machine learning are applicable to all learning systems, including natural evolution, animals, organizations, and computer algorithms. As the discipline develops, it's possible that algorithms and theory related to machine learning may offer more and more models for comprehending learning in brain systems, organizations, and biological evolution. Additionally, machine learning may gain from continuing research into these other kinds of learning systems.

Like any potent technology, machine learning poses concerns about what applications society should support and which to prohibit, as previously noted, the drive in recent years to gather new types of personal data due to its financial worth raises clear privacy concerns. Another ethical question that is brought up by the growing value of data is who will control and have access to internet data, as well as who will profit from it. Nowadays, companies gather a lot of data for targeted purposes that increase revenues; there is little to no incentive for sharing this data. Nonetheless, if the public could access the data for free, society might gain a great deal from it, even from already-existing online data.

To demonstrate, let's look at a straightforward example of how society could profit from data that is currently available online by using it to reduce the likelihood that infectious diseases would cause a global pandemic. It is possible to combine location data from online sources (such as cell phone location data, credit card transaction data from retail outlets, and security camera data from public and private buildings) with online medical data (such as emergency room admissions) to create a simple system that would notify people right away if someone they were in close contact with the day before was just admitted to the emergency room with an infectious disease, informing them of the symptoms to watch out for and the precautions they should take. There is a clear conflict and trade-off between individual privacy and public health in this situation, and society as a whole must decide how to resolve it. The main take away from this scenario is that, even if the data are already available online, society does not already have the laws, norms, cultures, or other systems necessary to allow it to use them for its benefit, should that desire arise.

Despite the fact that these are data about each of us, a large portion of them are actually privately held and owned. These kinds of factors imply that machine learning will probably rank among the 21st century's most revolutionary technologies. Even though the future cannot be predicted, society must start thinking about how to optimize its advantages right away.

REFERENCES

[1] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer, New York, 2011).
[2] K. Murphy, Machine Learning: A Probabilistic Perspective (MIT Press, Cambridge, MA, 2012).
[3] L. Valiant, Commun. ACM 27, 1134–1142 (1984).
[4] V. Chandrasekaran, M. I. Jordan, Proc. Natl. Acad. Sci. U.S.A.110, E1181–E1190 (2013).
[5] S. Decatur, O. Goldreich, D. Ron, SIAM J. Comput. 29, 854–879 (2000).
[6] S. Shalev-Shwartz, O. Shamir, E. Tromer, Using more data to speed up training time, Proceedings of the Fifteenth Conference on Artificial Intelligence and Statistics, Canary Islands, Spain, 21 to 23 April, 2012.
[7] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, in Foundations and Trends in Machine Learning 3 (Now Publishers, Boston, 2011), pp. 1–122.
[8] S. Sra, S. Nowozin, S. Wright, Optimization for Machine Learning (MIT Press, Cambridge, MA, 2011).
[9] J. Schmidhuber, Neural Netw. 61, 85–117 (2015).
[10] Y. Bengio, in Foundations and Trends in Machine Learning 2 (Now Publishers, Boston, 2009), pp. 1–127.
[11] A. Krizhevsky, I. Sutskever, G. Hinton, Adv. Neural Inf. Process. Syst. 25, 1097–1105 (2015).

[12] G. Hinton et al., IEEE Signal Process. Mag. 29, 82–97 (2012).
[13] G. E. Hinton, R. R. Salakhutdinov, Science 313, 504–507 (2006).
[14] V. Mnih et al., Nature 518, 529–533 (2015).
[15] R. S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction (MIT Press, Cambridge, MA, 1998).
[16] E. Yaylali, J. S. Ivy, Partially observable MDPs (POMDPs): Introduction and examples. Encyclopedia of Operations Research and Management Science (John Wiley, New York, 2011).
[17] W. Schultz, P. Dayan, P. R. Montague, Science 275, 1593–1599 (1997).
[18] C. Dwork, F. McSherry, K. Nissim, A. Smith, in Proceedings of the Third Theory of Cryptography Conference, New York, 4 to 7 March 2006, pp. 265–284.
[19] A. Blum, K. Ligett, A. Roth, J. ACM 20, (2013).
[20] J. Duchi, M. I. Jordan, J. Wainwright, J. ACM 61, 1–57 (2014).
[21] M.-F. Balcan, A. Blum, S. Fine, Y. Mansour, Distributed learning, communication complexity and privacy. Proceedings of the 29th Conference on Computational Learning Theory, Edinburgh, UK, 26 June to 1 July 2012.
[22] Y. Zhang, J. Duchi, M. Jordan, M. Wainwright, in Advances in Neural Information Processing Systems 26, L. Bottou, C. Burges, Z. Ghahramani, M. Welling, Eds. (Curran Associates, Red Hook, NY, 2014), pp. 1–23.
[23] Q. Berthet, P. Rigollet, Ann. Stat. 41, 1780–1815 (2013).
[24] A. Kleiner, A. Talwalkar, P. Sarkar, M. I. Jordan, J. R. Stat. Soc.,B 76, 795–816 (2014).
[25] M. Mahoney, Found. Trends Machine Learn. 3, 123–224 (2011).
[26] T. Mitchell et al., Proceedings of the Twenty-Ninth Conference on Artificial Intelligence (AAAI-15), 25 to 30 January 2015,Austin, TX.
[27] M. Taylor, P. Stone, J. Mach. Learn. Res. 10, 1633–1685 (2009).
[28] S. Thrun, L. Pratt, Learning To Learn (Kluwer Academic Press,Boston, 1998).
[29] L. Wehbe et al., PLOS ONE 9, e112575 (2014).
[30] K. Xu et al., Proceedings of the 32nd International Conference on Machine Learning, vol. 37, Lille, France, 6 to 11 July 2015, pp. 2048–2057.
[31] D. Blei, Commun. ACM 55, 77–84 (2012).