[CP-04-004] Artificial Intelligence Tools and Platforms for GIS

Abstract

Artificial intelligence is the study of intelligence agents as demonstrated by machines. It is an interdisciplinary field involving computer science as well as, various kinds of engineering and science, for example, robotics, bio-medical engineering, that accentuates automation of human acts and intelligence through machines. AI represents state-of-the-art use of machines to bring about algorithmic computation and understanding of tasks that include learning, problem solving, mapping, perception, and reasoning. Given the data and a description of its properties and relations between objects of interest, AI methods can perform the aforementioned tasks. Widely applied AI capabilities, e.g. learning, are now achievable at large scale through machine learning (ML), large volumes of data and specialized computational machines. ML encompasses learning without any kind of supervision (unsupervised learning) and learning with full supervision (supervised learning). Widely applied supervised learning techniques include deep learning and other machine learning methods that require less data than deep learning e.g. support vector machines, random forests. Unsupervised learning examples include dictionary learning, independent component analysis, and autoencoders. For application tasks with less labeled data, both supervised and unsupervised techniques can be adapted in a semi-supervised manner to produce accurate models and to increase the size of the labeled training data.

Keywords: artificial intelligence (AI), deep learning, machine learning

Author & citation

Lunga, D. (2019). Artificial Intelligence Tools and Platforms for GIS. The Geographic Information Science & Technology Body of Knowledge (4th Quarter 2019 Edition), John P. Wilson (ed.). DOI: 10.22224/gistbok/2019.4.16

This Topic is also available in the following editions:

DiBiase, D., DeMers, M., Johnson, A., Kemp, K., Luck, A. T., Plewe, B., and Wentz, E. (2006). Computational Intelligence. The Geographic Information Science & Technology Body of Knowledge. Washington, DC: Association of American Geographers.

This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

Explanation

1. Definitions

2. Introduction to Artificial Intelligence
3. Data Mining
4. Machine Learning
5. Artificial Neural Networks
6. Deep Learning
7. Reinforcement Learning
8. Hypothesis Space
9. Transfer Learning
10. Machine Learning in Specialized Geospatial Functions

1. Definitions

A **basin of attraction** defines the set of starting states that ultimately results in finding a local optimum.

Data mining (also known as knowledge discovery in data) is the process of discovering and extracting useable patterns from large data sets.

Deep learning (DL) is a subfield of machine learning that is inspired by artificial neural networks to approximate a mapping function between input and output values. It structures algorithms in layers to create a hierarchical structure of nonlinear transformations that enable learning the best function approximation.

A **feasible solution** is defined as the set of values for the decision variables that satisfies all constraints in an optimization problem.

A **globally optimal solution** is a feasible solution with an objective value that is better than all other feasible solutions for the model.

Hypothesis space in learning problems defines a basin of attraction including set of starting states that ultimately result in finding a local optimum for estimated models. It is characterized by properties including the dimensionality, representational capacity as well as local optima and basin of attraction.

Learning is defined as a process by which a machine finds application task specific patterns from data. The process can be achieved as a supervised, semi-supervised or unsupervised task.

Local optimal is a solution for which no better feasible solutions can be found in the immediate neighborhood of the given solution.

Machine learning refers to a set of computational or algorithmic steps that characterize a mathematical formulation to allow learning from data.

An **optimization problem** is defined as finding the best solution from all feasible solutions.

Reinforcement learning (RL) is a branch of artificial intelligence that is autonomous and adapts a self-teaching framework - which is both systematic and dynamic - via the notion of learning through actions.

Training is the process by which a machine learns to find correlations between inputs and expected outputs.

2. Introduction to Artificial Intelligence

The volumes of geospatial data continue to grow in tremendous amounts while positively influencing new developments in hardware and software technologies that includes artificial intelligence and its impact on societal problems. Big geospatial data and the recent advances in cost affordable hardware computing resources are both behind the surge in machine learning, deep learning, and reinforcement learning - all subfields of artificial intelligence. In this brief note, we present an overview of select artificial intelligence methods e.g. machine learning, deep learning, reinforcement learning and transfer learning methods. We describe example learning hypothesis spaces that includes a search for optimality solutions, and we identify current artificial intelligence tools that maybe useful for geographic information sciences and technologies (GIS&T). Research and development efforts continue to demonstrate that integrating data extracted from geographical information systems with artificial intelligence has potential to advance society's understanding of the world around. A few examples where artificial intelligence is expanding the performance of specialized geospatial analysis functions are also presented.

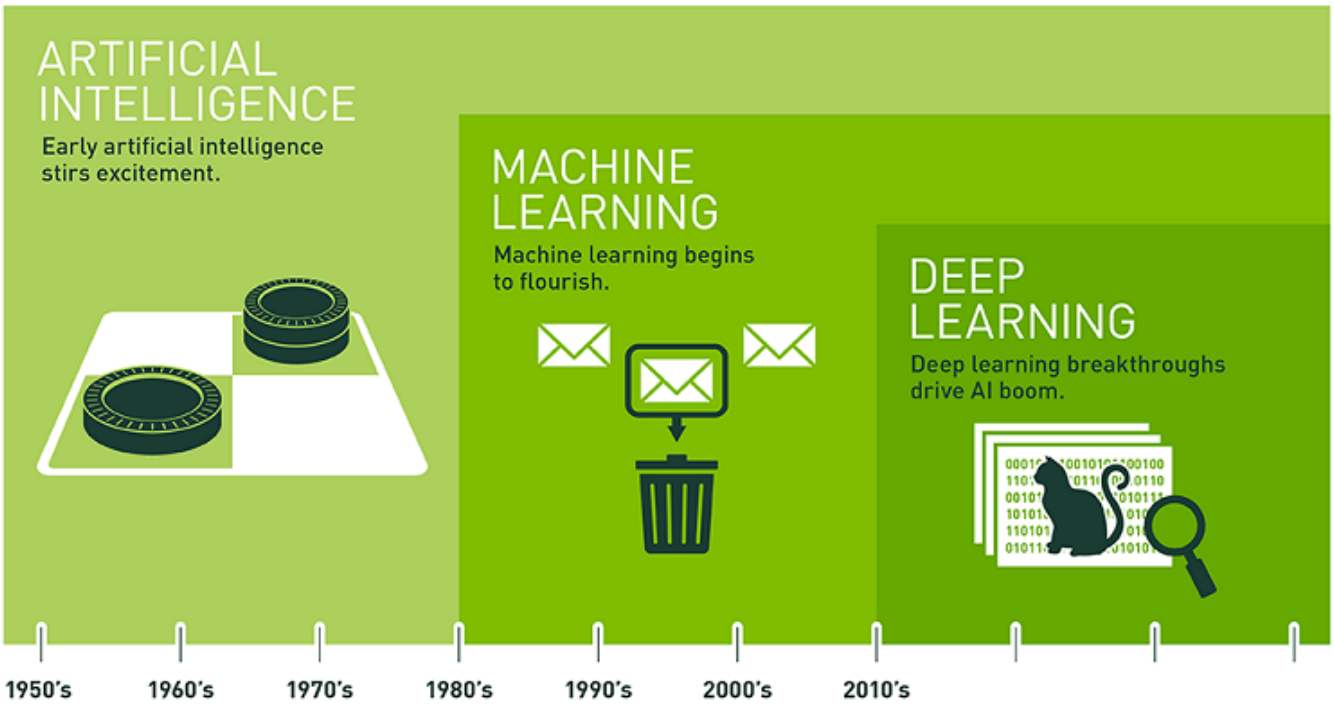


Figure 1. The timeline for artificial intelligence. Source: NVIDIA.

Artificial intelligence (AI) is an interdisciplinary field involving computer science, as well as, various kinds of engineering and science, e.g. robotics, bio-medical engineering, that accentuates automation of human acts and intelligence through machines. For over two

decades, AI's successes emerged in applications that required an overlap between low and high-level cognitive functions as captured in various techniques including reinforcement learning, and machine learning (Duch and Mandziuk 2007; Russell and Norvig 2010. Figure 1 shows the timelines of AI evolution from the mid 1950s. Back then, AI focused on problem solving and symbolic methods. Perceptual reasoning tasks emerged in the 1960s when the Defense Advanced Research Projects Agency (DARPA) took interest in AI. In the 1970s and 1980s, DARPA went on to apply AI on the problem of optical character recognition using early architectural designs of neural networks (Le Cun et al. 1990). The mid 1990s saw a rise in the emergence of machine learning methods however, their performance and scalability were constrained by hardware and lack of large data. From the late 2000s till present, advances in computing hardware technologies (e.g. introduction of graphic processing units to computing) and the availability of large volumes of data have enabled most of the present-day machine learning breakthroughs (Krizhevsky, Sutskever, and Hinton 2012). We are continuing to see AI algorithms enable computers to learn from experience by repetitively processing large volumes of data, adjusting to changes from inputs, and recognizing patterns of interest. To contrast the current wave of success with early work that motivated AI to solve problems and understanding of symbolic knowledge (Russell and Norvig 2010), research findings from the past decade indicate that complex tasks that are typically associated with cognitive abilities can, to a reasonable degree, be captured and reproduced by fitting functions to data (Darwiche 2018). Fitting complex functions to data has improved recently due to the availability of large volumes of data, recent technological gains in hardware computing, and continued advances in statistical methods. The geospatial community is equally seeing a growing number of practical applications which correspond to functions that are simple enough to enable new learning algorithms that can be evaluated efficiently. This includes very scalable methods for detecting object patterns from imagery datasets. Breathtaking success in geospatial application problem include predicting severe weather patterns detected at exascale (Kurth et al. 2018) to enabling self-driving cars Fridman et al. (2018). The following section briefly explains few branches of AI that are continuing to enable such incredible performance to complement and augment cognitive abilities.

3. Data Mining

Data mining, also known as knowledge discovery in data, is the process of discovering and extracting useable patterns from large data sets (Tan et al. 2018). Data is processed in batches using various machine learning techniques including recommendation algorithms, association, clustering, and numeric predictions etc. In practice, data mining augments traditional applications that query the database to collate, summarize, and analyze its contents by infusing sophisticated mathematical algorithms e.g. from artificial intelligence and machine learning, to find previously unknown or undiscovered correlations among different and large datasets (Tan et al. 2018)**.** Key features of data mining include automatic pattern predictions based on trend and behavior analysis, mapping between input and outputs based on likely outcomes, creation of decision-oriented information, extraction of information from large data sets and databases for analysis, and clustering based on finding and visually documented groups of facts not previously known. In GIS applications, the same functions in data mining are adapted, however, on spatial data with the objective to discover and extract correlations in data while incorporating geography - hence giving credence to the name spatial data mining (Li, Wang, and Li 2016).

4. Machine Learning

Machine learning (ML) refers to a set of computational or algorithmic steps that characterizes a mathematical formulation to allow learning from data (Goodfellow, Bengio, and Courville 2016; Skymind 2018; Pang-Ning et al. 2018). “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” (Mitchell 1997). As shown in Figure 2, ML techniques are commonly subdivided into two broad categories: **supervised** and **unsupervised** learning methods (Goodfellow, Bengio, and Courville 2016; Castrounis 2016; Mitchell 1997).

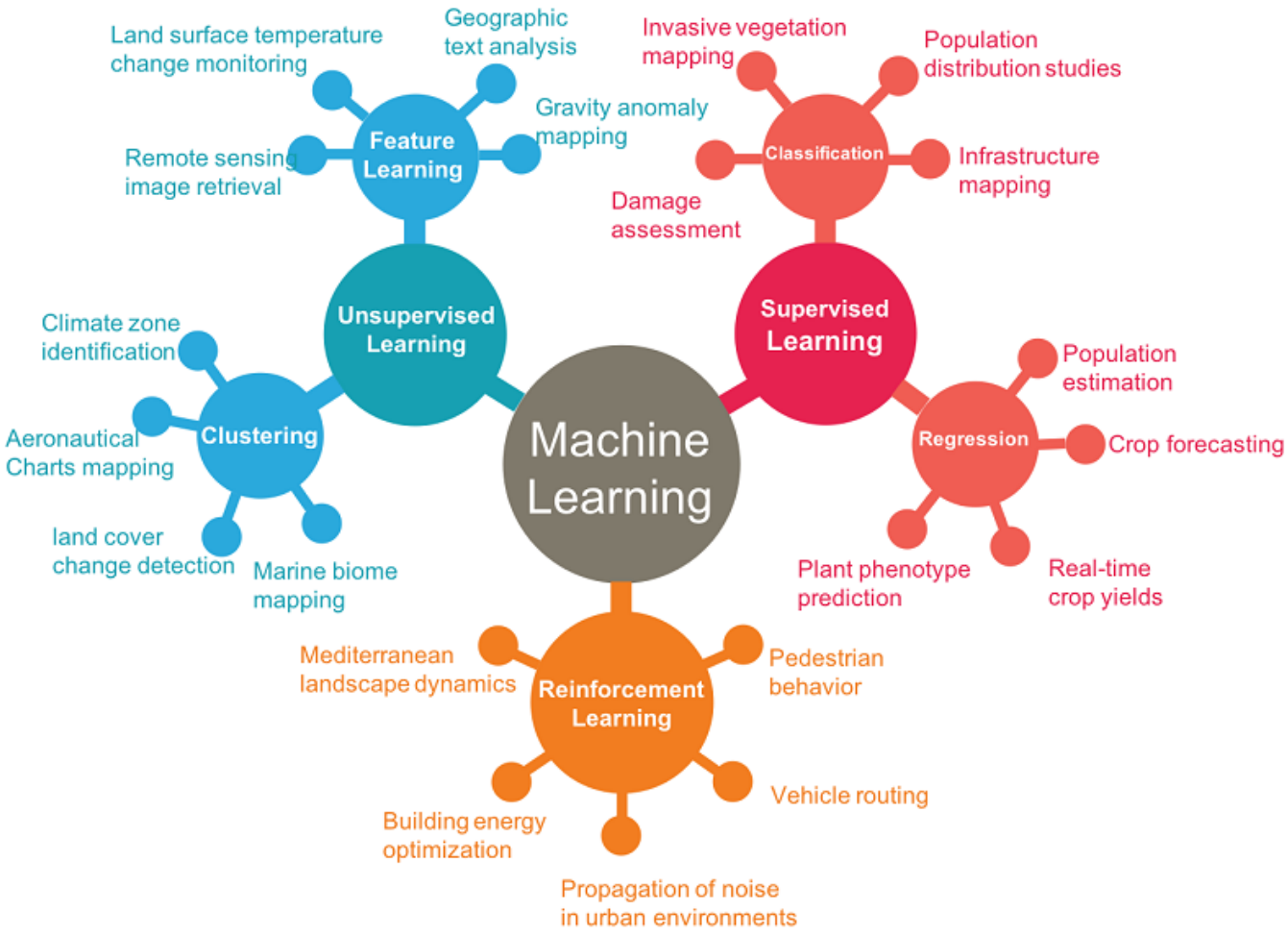


Figure 2. Example GIS applications and their potential formulation for automated learning as an ML problems. Source: author.

4.1 **Supervised learning** algorithms experience a dataset containing features, but each input example is also associated with a label or target output. The methods are usually defined from three views of (1) parametric, (2) non-parametric, and (3) those that are in-between parametric and non-parametric techniques. Parametric methods assume fitting data to a defined complex function for use in prediction tasks while non-parametric methods do not. Supervised learning is the common and most studied type of learning

category because it is easier to train a machine to learn with labeled data than with un-labeled data. Depending on the application task, supervised learning can be used to solve either regression or classification problems. Satellite imagery offers one example dataset from the geospatial community that can be used to illustrate regression or classification tasks. For tasks that require prediction of discrete values such as detecting and classifying a set of objects from satellite imagery into different classes, the tasks are solved as a classification problem. To predict continuous values, such as the count of objects in an image the problem is solved as a regression task.

4.2 **Unsupervised learning** algorithms experience a dataset containing many features, then learn useful properties of the structure of this dataset. The methods are driven by the need to discover natural patterns from the data without knowledge of the underlying labeling information. There are no correct answers given to the machine during learning, natural patterns in data hopefully do guide the machine to learn to detect key patterns and to group data by such patterns i.e. unsupervised learning are machines trying to learn “on their own”, without help from labeled data. Unsupervised learning tasks can be solved either as clustering or association problems depending on the application. As a clustering problem, unsupervised learning can be performed by a set of machine learning techniques that learns to discover structures in data and learn to solve the problem by searching for similarities using natural patterns from data. If there is a common cluster or group, the algorithm would then categorize them in a certain form. An example of this could be seeking to group images with similar content to enable content-based retrieval from large databases. As an association problem, unsupervised learning tries to solve this problem by trying discover important features while understanding the rules and meaning behind different groups. Finding a relationship between customer purchases in a certain geography or demographic is a common example of an association problem in online retail recommendation platforms.

5. Artificial Neural Networks

While a comprehensive summary of machine learning is beyond the scope of this entry, some basic context is still crucial for understanding the emergence of crucial building blocks to few modern machine learning techniques. In particular, artificial neural networks dates back to the 1940s-1960s where it was known as cybernetics, it was known as connectionism in the 1980s-1990s and its most recent resurgence under the name deep learning in 2006 (Goodfellow, Bengio, and Courville 2016). As a computational model for capturing biological learning, artificial neural network (ANNs) came to define the current basic building blocks for modern deep learning. ANNs models are viewed as engineered systems inspired by biological brain to provide what can be seen as proof by example that intelligence can be learned. In the 1940s, ANNs were due to **simple linear models** that took a set of inputs values and associated them with an output value. The model would be defined by learning a set of **weights** and the linear model would be used in recognize different categories e.g. recognizing two different categories by checking if the **linear function** is positive or negative. For the model to make correct predictions, the weights needed to be adapted manually. In the 1950s, the perceptron (Rosenblatt 1958) became the first algorithm that could learn model weights that mapped inputs to corresponding classes. The training algorithm used to adapt model weights was a special case of the **stochastic gradient descent**. Most neural networks building block as used today’s deep

learning methods are based on model neuron called the rectified liner unit using neuroscience as an inspiration (Goodfellow, Bengio, and Courville 2016).

6. Deep Learning

Deep learning (DL) is a recent artificial neural network-based technology that is widely applied in supervised and unsupervised learning tasks to define machine learning algorithms that are autonomous and inspired by the self-teaching structure and function of artificial neural networks. The concept of DL is often referred to as deep neural networks, which refers to the many layers of data transformation involved. Whereas a neural network may be designed with a single layer for data transformation (Mitchell 1997; Collobert & Bengio 2004; Russell and Norvig 2010), a deep neural network often has more than two layers. The layers are organized in a hierarchy with each layer nonlinearly mapping information signals toward more abstract representation (Goodfellow et al. 2016). The training process often entails the use of millions of records of existing data to train algorithms to find patterns with predictions on new data made using estimated models. In several applications it has been demonstrated that with more data and computation the predictive performance of DL methods is scalable at rates that surpasses previous machine learning methods (e.g. shallow based learning methods including support vector machines - see Figure 3.) (Coates 2018; Collobert and Bengio 2004). As an example, one trains a DL algorithm to recognize buildings from a satellite image. This is achieved providing by many examples images whose pixels are labeled to denote either having a building or not building. Learning is achieved as an optimization search for optimal model weights (or parameters) that help distinguish patterns by discriminating image features (for example edges, shapes, colors, spatial context). Learned features are used during model inference stage to then classify image pixels on whether they contain building or not.

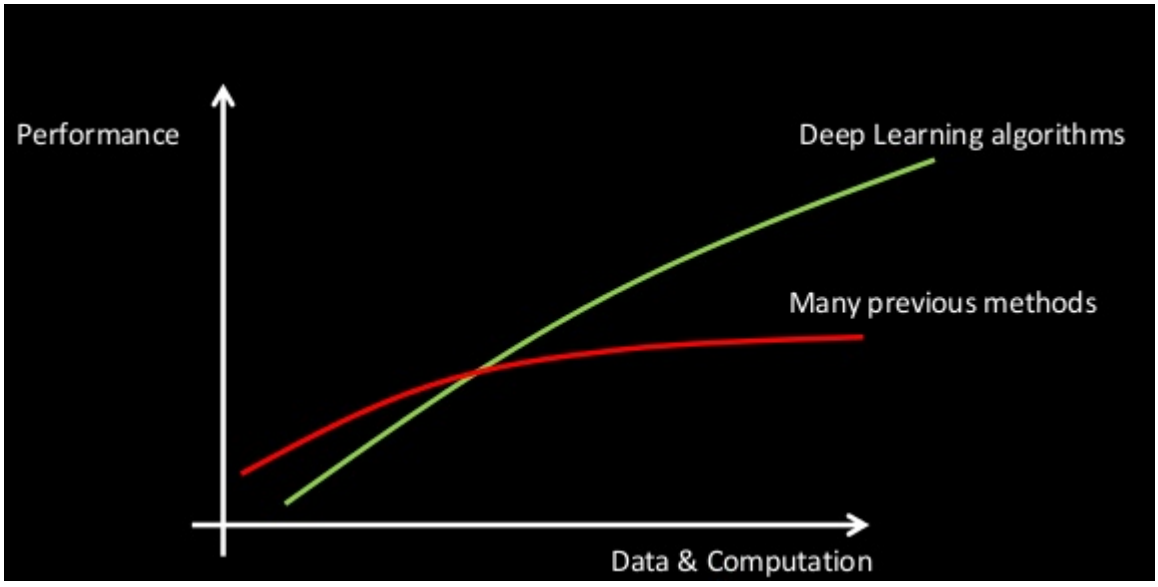


Figure 3. Scability of deep learning algorithms. Source: after Coates (2018).

7. Reinforcement Learning

Although not yet widely adapted in many geospatial applications, reinforcement learning (RL) is a growing branch of artificial intelligence that is autonomous and adapts a self-teaching framework based on learning through actions. RL performs actions within an environment to maximize a reward function along a particular dimension and consequently trigger learning through trial and error via goal-oriented algorithms. Using the concepts of agents, environments, states, actions and rewards, RL methods differ from other machine learning techniques in the sense that learning is incremental and there is no static training data collected ahead of the learning. Instead RL agents interact with environments, generate data, or passively wait for new data and decide how to act to perform a given task. Because of this, the algorithms are known as learning dynamically from the environments. RL encompasses a level of learning more general than that of supervised or unsupervised approaches. By learning from experience, rather than from training data, RL algorithms are expected to outperform other methods in more ambiguous, real life environments while choosing from an arbitrary number of possible actions (Sharma 2018).

8. Hypothesis Space

Given a set of input observations (e.g. image pixels) an ML algorithm may seek to learn a function that is able to predict an output value (e.g. class category). The problem of learning the corresponding model is then reduced to a search for one optimal hypothesis that explains the relation between inputs and expected output values. The hypothesis space can be characterized by properties including the dimensionality, representational capacity as well as local optima (Dulek 2013). Learning complexities are introduced via the curse of dimensionality as the effective size of the hypothesis grows exponentially with the number of dimensions (Bellman 2003; Blumer et al. 1989).

The expressive power, richness, or flexibility of a space of functions that can be learned by an algorithm defines the representational capacity of the hypothesis space. Searching through all possible hypotheses from both continuous and discrete spaces can be infeasible. Rather, learning methods use heuristics to traverse their local Euclidean search space (Mitchell 1997). As illustrated in Figure 4, gradient descent-based learning methods initialize from a given hypothesis and traverse its space for a better solution until no improvements can be found (Mitchell 1982, 1997). The common assumption is that the best hypothesis is within some epsilon distance or neighborhood of other good hypotheses. The notion of local optima introduces hypotheses that are good but may not be near to the globally best solution. As shown in Figure 4, it is common for gradient based descent methods to be trapped in local optima, find no better solutions nearby, and return it as the answer, even though a global optima may exist elsewhere. Which local optimum is found does depend on the initialization point and hyperparameter settings of the learning algorithm.

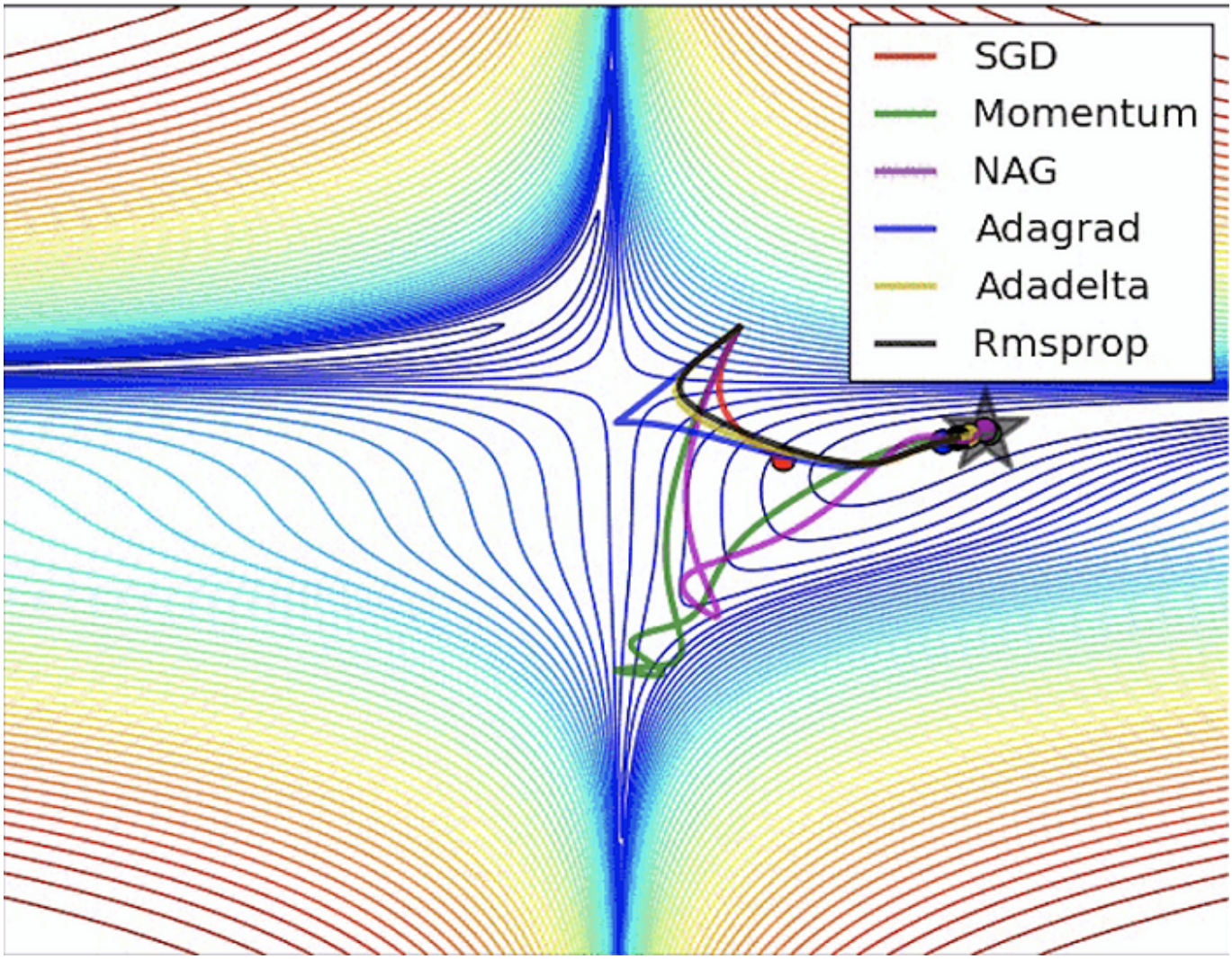


Figure 4. Demonstrating search for optimal values of a given function using various gradient descent based optimization algorithms. The plots shows stochastic gradient descent (SGD), momentum, Nystrov accelerated gradient (NAG). Initially gradient values are large, causing velocity based search techniques shoot off and bounce around - adagrad almost goes unstable. The function to be optimized has its values indicated by contours in different lines from red (highest contours) to blue(regions of lowest values). GIF: http://ruder.io/optimizing-gradient-descent/. Source: Ruder 2016.

To give an example hypothesis, the work of (Darwiche 2018) makes a great distinction on two approaches for solving for the optimal hypothesis in AI problems. Many problems in GIS can be mapped to both approaches i.e. functional-based and a model-based approaches. If we consider the task of object recognition with satellite imagery, the hypothesis space from a functional based perspective entails formulating the task as a function-fitting problem, with image pixels used as function inputs and function outputs corresponding to abstract recognition of object of interest. The functional form can be arbitrary in complexity but of easy to evaluate. Various optimization algorithms can then be applied on the space of functions to search for the optimal values that yields a function for use in predicting what object are contained in new test images. Using model-based approach the recognition tasks could be solved through representation of objects using an ontology. Reasoning through logic and probability becomes the tools to extract knowledge for inference and prediction

on test examples (Russell and Norvig 2010).

9. Transfer Learning

Transfer learning (TL), is not an algorithm nor is it a branch of study but rather a design methodology, within machine learning, for adapting and leveraging models that are estimated from one task for use in a different task. Often generalized feature learning can be achieved from a source domain containing large amounts of labeled data and while the target domain of application may contain less labeled samples. As shown in Figure 5, through the process of fine-tuning models, feature extraction knowledge can then be reused to improve generalization in the target domain. This design methodology is commonly applied in deep learning methods to transfer Task A (with dataset A) knowledge as captured by model weights to Task B (denoted target dataset) (Gerrand et al. 2017). By most arguments this can be viewed as an informed method to initialize weights for tasks where labeled data is scarce.

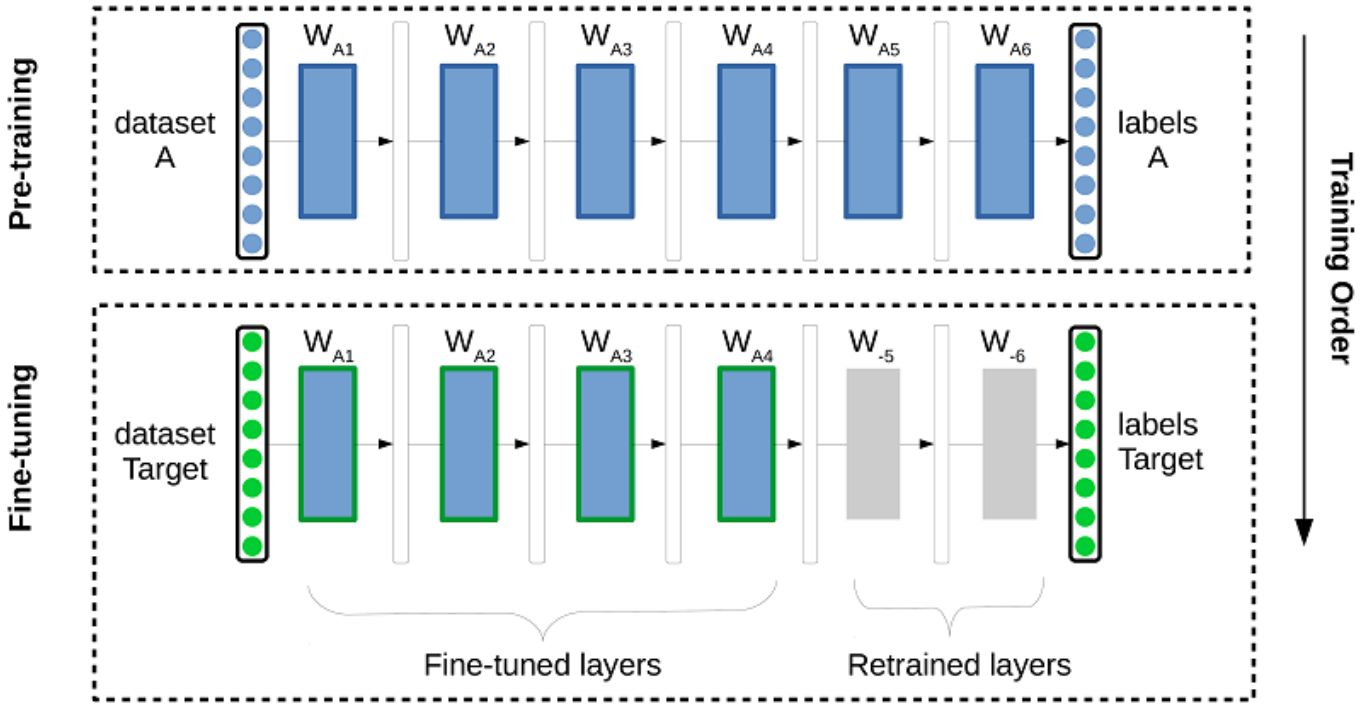


Figure 5. An illustration of transfer learning to a new target dataset with a model pretrained on dataset A. Shown above, transfer learning is performed via fine-tuning (training target model on few iterations with select layers initialized from model trained on dataset A e.g. W_A1,...,W_A4) while layers W_-5 and W-_6 are trained from random initialization. This process transfers knowledge from Task A to Task B, thereby enabling the target task to seek a small labeled target dataset while benefiting from knowledge acquired from dataset A . Source: author, after Gerrand et al. (2017).

10. Machine Learning in Specialized Geospatial Functions

As shown in Figure 2, machine learning techniques are transforming a range of applications including vision-based object detection, autonomous driving, agriculture, food security, infrastructure monitoring and disaster management. For example, in population mapping, the need to understand where people live is fundamental to understanding what people do and what their social needs are with respect to energy security; policy and urban development; resilience; disaster and emergency response; and humanitarian support, as well as understanding the behavioral social dynamics. With large volumes of satellite imagery available and accelerated graphics computing hardware we are seeing increased adaptation of machine learning and AI for global scale mapping solutions e.g. human settlement detection from high resolution satellite imagery (Lunga et al. 2018). Several use cases including support for FEMA disaster response in the western United States to meet specific needs of the intelligence community (Yang et al. 2018) are example use cases that are benefiting from applying AI technologies. Extending the application examples, the work of (Wang et al. 2017) illustrates how the problem of map generalization, scale reduction, and feature symbolization is improved through genetic algorithms designed to incorporate cartographic constraints.

Most geospatial tasks have been conducted manually requiring a large number of human resources. Lack of automation on large tasks is noticeably expensive, prone to infrequent updates and human error. For example, gathering street addresses is a human labor-intensive effort that can now be simplified by AI methods. Recently AI, has demonstrated the potential to automatically create street addresses from satellite images by learning and labeling roads, regions, and address cells (Demir et al. 2018). In another example, planimetric mapping of services is often required by local government tax assessors to create tax assessment databases. As a proxy, swimming pools are added to assessment records because they increase the value of the property. As such finding swimming pools that are not on the updated property list is of great importance to the assessor. Without automation this task is daunting. However, use of GIS and AI tools has demonstrated a reduction in expensive human labor involved in updating the records through field visits of each property (Singh 2018). While pools can increase the value of a property, they can be a cause to worry about disease outbreaks during downturn and slow recovery for real estate economic sectors. Many residential homes can be left with neglected pools that are often a breeding ground for mosquitoes. At a recent user conference, scientists at Esri, demonstrated an integration of ArcGIS software with the latest innovations from AI to perform the detection of swimming pools using aerial imagery. The scientists developed further analytics to identify pools in a state of neglect, which has since been empowering health inspectors to help prevent the spread of vector-borne diseases (Jha and Singh 2018; Rodriguez-Cuenca and Alonso 2014). In environmental epidemiology, exposure modeling is a commonly used approach to conduct exposure assessment to determine the distribution of exposures in study populations. Artificial intelligence coupled with GIS technologies provide important advantages for exposure modeling in environmental epidemiology, including the ability to incorporate large amounts of big spatial and temporal data in a variety of formats; computational efficiency; flexibility in algorithms and workflows to accommodate relevant characteristics of spatial (environmental) processes including spatial nonstationarity; and scalability to model other environmental exposures across different geographic areas (VoPham et al. 2018). In other specialized areas, a wave of new artificial intelligence methods are helping assimilate autonomous vehicles and intelligent transport system by incorporating a great amount of information gathered by traffic cameras and sensors for road mapping (Agachai and Hung 2018). Furthermore, artificial intelligence technology is impacting the discovery of geographic knowledge within

unstructured text data across different languages (Schmidt et al. 2013). There also exist many other applications of AI techniques in geospatial research, such as spatial diffusion prediction in epidemiology, urban expansion analysis, and hyperspectral image analysis (Abdelkader et al. 2017; GIS Resources 2018; Peter 2016).

References

Abdelkader, E. G., David, J. M., Said, E. G., & Joseph, K. (2017). Analysis of urban growth and sprawl from remote sensing data: Case of Fez, Morocco. International Journal of Sustainable Built Environment, 6(1), 160–169.

Agachai, S., & Hung, W. H. (2018). Smarter and more connected: Future intelligent transportation system. IATSS Research, 42(2), 67-71.

Bellman, R. E. (2003). Dynamic Programming. Princeton University Press.

Beni, G. and Wang, J. (1993). Swarm Intelligence in Cellular Robotic Systems. In: Dario, P., Sandini, G., Aebischer, P. (eds) Robots and Biological Systems: Towards a New Bionics?. NATO ASI Series, vol 102. Springer, Berlin, Heidelberg.

Blumer, A., Ehrenfeucht, A., Haussler, D., & Warmuth, M. K. (1989). Learnability and the Vapnik-Chervonenkis Dimension. Journal of the Association for Computing Machinery, 36, 929–965.

Castrounis A. (2016). Machine Learning: A Complete and Detailed Overview. Blog posting, KDNuggets.

Coates, A. (2018). AI for 100 Million People with Deep Learning. AI Frontiers.

Collobert, R. & Bengio, S. (2004). Links between perceptrons, MLPs and SVMs. In Proceedings ICML '04: Proceedings of the twenty-first international conference on Machine learning. ACM, New York, NY, USA, 23-.

Darwiche, A. (2018). Human-level intelligence or animal-like abilities? Communications of the ACM 61(10):56-67.

Demir, I., Hughes, F., Raj, A., Dhruv, K., Muddala, S. M., Garg, S., . . . Raskar, R. (2018). Generative street addresses from satellite imagery. ISPRS International Journal of Geo-Information, 7(3).

Duch, W. (2007). What Is Computational Intelligence and Where Is It Going?. In: Duch, W., Mańdziuk, J. (eds) Challenges for Computational Intelligence. Studies in Computational Intelligence, vol 63. Springer, Berlin, Heidelberg.

Duch, W., & Mandziuk, J. (2007). Challenges for Computational Intelligence. Springer.

Fridman, L., Brown, D. E., Glazer, M., Angell, W., Dodd, S., Jenik, B., . . . Reimer, B. (2018, october). Mit autonomous vehicle technology study: Large-scale deep learning based analysis of driver behavior and interaction with automation. Arxiv.

Gerrand, J., Williams, Q., Lunga, D., Pantanowitz, A., Madhi, S. A., & Mahomed, N. (2017). Paediatric frontal chest radiograph screening with fine-tuned convolutional neural networks. . In: Valdés Hernández, M., González-Castro, V. (eds) Medical Image Understanding and Analysis. MIUA 2017. Communications in Computer and Information Science, vol 723. Springer, Cham.

GIS Resources. (2018). Fundamemtals of Hyperspectral Remote Sensing, GIS Resources.

Goodfellow I., Bengio Y., & Courville A. (2016). Deep Learning. Cambridge, MA: The MIT Press.

Jha, D., & Singh, R. (2018). Swimming pool detection and classification using deep learning. Medium.com article.

Karpathy, A. (2015). Convolutional Neural Networks for Visual Recognition. Notes from the Stanford Computer Science class CS231n.

Khandelwal, R. (2019). Overview of different Optimizers for neural networks. Medium.com article.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1 (pp. 1097–1105). USA: Curran Associates Inc.

Kurth, T., Treichler, S., Romero, J., Mudigonda, M., Luehr, N., Phillips, E., . . . Houston, M. (2018, october). Exascale deep learning for climate analytics. SC '18: Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis

LeCun, Y. L., Boser, B., Denker, J. S., Howard, R. E., Habbard, W., Jackel, L. D., & Henderson, D. (1990). Handwritten Digit Recognition with a Back-Propagation Network, in Advances in Neural Information Processing Systems 2: 396–404.

Li, D., Wang S., and Li, D. (2016). Spatial Data Mining: Theory and Application (1st ed.). Springer Publishing Company.

Lunga, D., Yang, H. L., Reith, A., Weaver, J., Yuan, J., & Bhaduri, B. (2018). Domain-Adapted Convolutional Networks for Satellite Image Classification: A Large-Scale Interactive Learning Workflow. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(3), 962–977.

Mitchell, T. M. (1982). Generalization as search. Artificial Intelligence, 18, 203–226.

Mitchell, T. M. (1997). Machine Learning. 1st Edition. McGraw-Hill, Inc.

Pang-Ning, T., Steinbach, M., Karpatne, A. and Kumar, V. (2018). Introduction to Data Mining (2nd Edition). Pearson.

Peter, C. (2016). Spatiotemporal Frameworks for Infectious Disease Diffusion and Epidemiology. Int J Environ Res Public Health, 13(12).

Rodriguez-Cuenca, B., & Alonso, M. C. (2014). Semi-automatic detection of swimming pools from aerial high-resolution images and lidar data. Remote Sensing 6(4): 2628-2646.

Rosenblatt, F. (1958). The Perceptron: A Probabilistic Model for Information Storage and Organization in The Brain. Psychological Review, 65--386

Ruder, S. (2016). An overview of gradient descent optimization algorithms. Ruder.io.

Russell, S. J., & Norvig, P. (2010). Artificial Intelligence: A Modern Approach (3rd ed.). New Jersey: Prentice Hall.

Schmidt, S., Manschitz, S., Rensing, C., & Steinmetz, R. (2013). Extraction of Address Data from Unstructured Text using Free Knowledge Resources. i-Know '13: Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies.

Siddique, N., & Adeli, H. (2013). Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing. John Wiley and Sons.

Singh, R. (2018). How we did it: Integrating ArcGIS and deep learning at UC 2018. ArcGIS Blog.

Skymind. (2018). Machine learning algorithms.

VoPham, T., Hart, J. E., Laden, F., & Chiang, Y.-Y. (2018). Emerging trends in geospatial artificial intelligence (geoai): potential applications for environmental epidemiology. Environmental Health, 17(1), 40.

Wang, L., Guo, Q., Liu, Y., Sun, Y., & Wei, Z. (2017). Contextual Building Selection Based on a Genetic Algorithm in Map Generalization. ISPRS International Journal of Geo-Information, 6(9).

Yang, H. L., Yuan, J., Lunga, D., Laverdiere, M., Rose, A., & Bhaduri, B. (2018). Building Extraction at Scale Using Convolutional Neural Network: Mapping of the United States. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 8, pp. 2600-2614.