Brief Review of Machine Learning

Industry Trends and Needs

Machine learning (ML) seeks to answer many fundamental questions of the laws that encompass learning systems in individual human/animal biology, societies, and computers. The most important question is how to construct a system that improves its own rules/boundaries/parameters through experience. It is also a practical field where the software produced is applicable amongst many industries: automobile, healthcare, surveillance, natural language processing, mechanical/chemical engineering industries, etc. Currently, machine learning's effects are felt the most in fields with data-intensive issues such as business logistics chains, problem-diagnosis in complex systems, and consumer-driven services such as advertisement and companies with high volumes of personal financial transactions. There are also interesting applications in computational sciences, such as analyzing high volume experimental data (in biology or physics fields).

Current Solutions

Generally, algorithms in ML typically have three similar sections that feed into each other. Firstly, there is the experience section, which is the training dataset. Without data that represents the experience for an algorithm to learn from, to date, it has been impossible to create a machine learning system. Then, there is the algorithm or program itself. Lastly, there are the performance metrics which are chosen according to the experience and subject field. The program, guided by experience, seeks to optimize its performance. ML algorithms vary according to the way they represent data and the way they search through data and/or performance metrics; for example, decision trees, mathematical function approximations, functions derived through searching data are all various approaches to ML algorithms. Overall, explicitly or implicitly, an algorithm has some form of tunable degrees of freedom which are trained with the available and optimized with respect to the performance metrics.

ML algorithms are divided into three broad types: supervised learning, unsupervised learning and reinforcement learning. Supervised learning algorithms are designed around the fact that the training data is in the form of some (x1, x2, …, xn, y1, y2, …, yn) associations with the goal of the algorithm being to produce a predictive y values in response to x values [1]. There's been a lot of progress with simple binary classification problems where there is only one type of output (y value) expected. These types of algorithms generally have some form of a mapping, a lot of times with probability distribution over y given x. The mapping takes the form of decision trees, decision trees, decision forests, logistic regression, support vector machines, neural networks, kernel machines, and Bayesian classifiers. This diversity exists because of all the different fields and industries where ML is applied. Between these types of algorithms, there are tradeoffs between complexity, performance, and available data.

Deep learning falls under supervised learning, and it is a relatively recent area of development in ML. It has multiple layers of threshold units that compute parameterized functions, which together form a network [2]. Deep learning optimizes its algorithms through gradients formed from errors in the outputs due to the predicted parameters. This type of learning has been possible because of parallel computing, specifically graphics processing units (GPUs), which were initially developed for video gaming. We are now able to utilize billions of parameters in the deep neural network trained on hundreds of gigabytes of data [3]. This has been majorly helpful in the computer vision and natural language processing fields.

This review has focused on ML algorithms that rely on labeled training data. The other type of ML is unsupervised learning, with unlabeled training data. This is possible through the current scientific knowledge base on algebraic, combinatorial, or probabilistic associations between features or properties of the data being considered [4]. This type of learning includes dimension reduction methods such as principal components analysis, manifold learning, factor analysis, random projections, autoencoders and more [5]. One needs functions to define criteria for dimensional reduction, and there are techniques to optimize these criteria. Then, clustering is an important aspect of unsupervised learning. It solves the problem of finding a way to partition the training data and create the rules for predicting test data.

The next type of ML algorithm is reinforcement learning algorithms. It is based on decades of neuroscience and psychology research into reinforcement learning in animals (sometimes unethical human studies as well) [6]. The algorithms utilize control-theory research such as policy iteration, value iteration, rollouts, and variance reduction. The best-known formulation in ML is the Markov decision process [7]. The learning task is not to learn the outcomes, but rather to learn the policy or control strategy or action for an agent in an unknown, dynamic environment. The policy is trained to choose actions for various states, and the goal is to maximize reward over a given time period. More generally, reward is associated with a series of actions rather than just one action. This helps the machine learn consequences of individual actions as well as sequences of actions.

Critical Analysis of Solutions

There are multiple items that affect the accuracy, reliability, stability, and useability of ML algorithms. This section aims to briefly review many of these items. First, it is important to consider the data itself. The methodology used to obtain it and its size are important factors. Depending on the size of the training dataset, the algorithm's capacity to predict accurately from unseen data might change [8]. Smaller training datasets but excellent predictive accuracy might imply that the algorithm was partially already trained on a separate different dataset (sometimes this dataset could be from a completely different industry). Next, based on the skew of the training dataset, the algorithm might result in consistently more false positives or false negatives. All data acquisition methodologies have inclusion and exclusion criteria at various levels in the data acquisition process. With these criteria, the dataset should end up trying to represent the maximal spectrum of ranges in data values. If imbalances in classes or categories are inevitable, there are techniques to mitigate their effects. One can balance the classes through under sampling or taking away instances of the overrepresented class. The other way is to add copies of the underrepresented class through another ML model with noise or variations added into the data [9]. There is also the method of assigning higher weights to underrepresented classes during training the algorithm. Another question to ask is about the gold standard of the data obtained. If understanding the data requires expert knowledge, such as in healthcare or chemical industry instances, it is important for the data to have been curated and labeled by the experts in the field. Datasets are also sometimes subject to pre-processing which may remove pertinent data from the dataset. So, it is important to be aware of any processing done on datasets found online. Sometimes, there are glaringly obvious pieces of missing data in datasets [10]. Being aware of all these dataset variations is important to critically comment on the ML algorithm being discussed. Additionally, all these considerations contribute to the generalizability of any ML algorithm's performance in the real world.

There is the ML model or algorithm itself to consider. Conventional ones usually need knowledge of the dataset (or expert knowledge if the field is niche) to predefine some features that help the algorithm learn. This is a labor intensive and precision-based task. The models learned this way are the most open to interpretation and analysis [2]. Contrastingly, some types of ML algorithms rely on neural networks that are essentially black boxes. Deep learning relies heavily on convolutional neural networks (CNN) which emulate connectivity patterns like those in a general visual cortex found in nature. These networks are very high performing and require less expert knowledge in feature development [11]. However, the tradeoff is lower interpretability and robustness. A black box means that we, as humans, cannot tell how the algorithm took precise steps to make its prediction. The deep learning algorithm is trained on massive datasets, and it extracts parameters into an increasingly complex array of weighted connections that is difficult for human experts to understand. Usually, interpretability and performance are inversely proportional in ML.

Another aspect to consider are the performance metrics being utilized in the algorithm and how they are optimized. The specific output could be a classification or probability, an action, or some other output [12]. Each type of output has different performance metrics; especially when differentiating between simple prediction and causal inference. Adjacent metrics such as sensitivity and specificity are also important for the researchers to consider. ML algorithms make errors in output that are difficult to foresee; and these metrics could help make decisions on the risks of using the algorithms. A related challenge is overfitting of model parameters in a such a way that result in excellent metrics. External testing helps visualize this error and let us know when an algorithm is too good to be true. It is similarly important for the algorithm and its results to be repeatable and reproduceable. Repeatability is when some minor differences in data values should give similar predictions or causal inferences and does so. Reproducibility is when similar datasets derived from different hardware with different protocols result in the same predictions and casual inferences. Deploying large scale has been the goal for ML algorithms but making a fair assessment of these risks is paramount before doing so.

Proposal of an Interesting and Emerging Approach to ML

Interpretable Machine Learning is an emerging subfield in ML. There are two main approaches: model-based and post hoc interpretability. Model-based interpretability has not been able to achieve the level of accuracy and performance that black box models have achieved. If it does, because it would be more descriptive and understandable, it would be preferable. Post hoc interpretation is iterating through the parameters discovered by the model and analyzing their appropriateness.

My Master's project is on model-based Interpretable ML. Current research in this area has been into networks that have the following incorporated concepts: case-based reasoning, grammatical sentence structure rules, rules to decompose images into sections, and more [13]. One example of case-based reasoning is developing the network to dissect an image into predefined, prototypical parts, and then, combining the various prototypes to make a final categorical classification. Researchers demonstrated that this interpretable network could classify images with comparable accuracy as any number of non-interpretable, complex deep learning networks. Another technique is to use a probabilistic classifier with three types of 2 reasoning: positive, negative and indefinite. This provides a clear decision process the neural network took in regard to each image and each classification category. However, for both techniques, the challenges are in the increase the performance time (reduced efficiency) of the models. In contrast, there is a recent concept whitening technique introduced in [14] that preserves both the accuracy and performance efficiency of the ML models. I find this claim very interesting. My project is to reproduce it, apply it to a novel, more complex CNN, and in a medical use case scenario.

References

1. T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer, New York, 2011)
2. Schmidhuber J., Deep learning in neural networks: An overview. Neural Netw. 61, 85–117 (2015). 10.1016/j.neunet.2014.09.003
3. Krizhevsky A., Sutskever I., Hinton G., Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 25, 1097–1105 (2015)
4. S. Sra, S. Nowozin, S. Wright, Optimization for Machine Learning (MIT Press, Cambridge, MA, 2011).
5. Hinton G. E., Salakhutdinov R. R., Reducing the dimensionality of data with neural networks. Science 313, 504–507 (2006). 10.1126/science.1127647
6. Schultz W., Dayan P., Montague P. R., A neural substrate of prediction and reward. Science 275, 1593–1599 (1997). 10.1126/science.275.5306.1593
7. R. S. Sutton, A. G. Barto, Reinforcement Learning: An Introduction (MIT Press, Cambridge, MA, 1998).
8. Obermeyer Z, Emanuel EJ. Predicting the future-Big data, machine learning, and clinical medicine. N Engl J Med. 2016;375:1216–9.
9. Faes L, Liu X, Wagner SK, Fu DJ, Balaskas K, Sim DA, et al. A clinician's guide to artificial intelligence: How to critically appraise machine learning studies. Transl Vis Sci Technol. 2020;9:7.
10. Liu X, Rivera SC, Moher D, Calvert MJ, Denniston AK SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI Extension? BMJ. 2020;370:m3164. doi: 10.1136/bmj.m3164.
11. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell. 2019;1:206–15.
12. Tatman R, VanderPlas J, Dane S. 2nd Reproducibility in Machine Learning Workshop at ICML 2018. Stockholm, Sweden: A practical taxonomy of reproducibility for machine learning research
13. C. Chen, O. Li, D. Tao, A. Barnett, C. Rudin, and J. K. Su, “This looks like that: Deep learning for interpretable image recognition,” in Advances in Neural Information Processing Systems, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alch´e-Buc, E. Fox, and R. Garnett, Eds., vol. 32. Curran Associates, Inc., 2019.
14. Z. Chen, Y. Bei, and C. Rudin, “Concept whitening for interpretable image recognition,” Nature Machine Intelligence, Dec 2020. [Online]. Available: https://www.nature.com/articles/s42256-020-00265-z