Henry Suarez

23965259

Hsaurez000@citymail.cuny.edu

Supervised learning: Evolution, Ethical Concerns, Advancements, and Impacts.

Abstract

Supervised learning is a crucial category in machine learning and has undergone significant evolution since its inception. It's important to understand the beginnings of supervised learning and how it's been growing since it was introduced in the 1950's. Following the stages of supervised learning we should analyze the impacts that supervised learning can have on all industries and what technologies will come from it. An important aspect of supervised learning to consider is bias mitigation and ethical consideration. On how these could impact models which could lead to impacting people in the future. To get ahead of these, we must have ethical frameworks and learn how to continue to develop them as the technology itself involves. Finally, we go into discussing the newest emerging trends in supervised learning, discussing what certain researchers are working on and how they can change industries.

Introduction

Supervised learning, also described as supervised machine learning, is a category of AI/machine learning. It's described as an algorithm that is trained based on its input data with expected/labeled output data. In other words, the algorithm learns based on input data and output data. Then eventually after enough training the algorithm should be able to make predictions based on any input data given. Now that machine learning and AI have resurged again being a "gold rush" in technology, everyone is trying to find a niche, business, opportunity…etc. Supervised learning has also developed greatly from when it first started and has impacted many areas. But can there be any possible harmful effects of this growth? One main area of concern is bias and fairness in algorithmic decision-making. As AI/ML grows into areas like healthcare, computer vision, education, and law. It's increasingly important that we try to reduce the amount of bias and unfairness that could be potentially present in algorithmic systems.

Understanding supervised learning

Before we can dive deeper into the ethical implications of supervised learning, we have to get an understanding of what is supervised learning and how it works. Supervised learning is a branch/training method of machine learning, where algorithms are trained on labeled data to produce outputs. This involves training the algorithm with provided datasets and then a testing phase where the model attempts predictions on new, unseen data. In simpler terms, we train a model with label input and output. Then eventually it should be able to "predict" an output based on how much it's been trained.

Evolution of supervised learning

The start of Supervised learning can be accredited to multiple pioneers in the 1950's. Two significant contributors were Arthur Samuel and Frank Rosenblatt. Arthur Samuel developed a program for playing checkers and "coined the term machine learning". His design had a scoring function that used the position of pieces on the board and attempted to measure the chance of either side winning [4]. Frank Rosenblatt developed the perceptron laying the groundwork for supervised learning [4]. Notably, the creation of the perception marked a leap forward establishing the principles of supervised learning. However, at the time these developments seemed promising but produced "broken expectations" as it could not recognize many kinds of patterns. This eventually stalled and slowed research until a resurgence during the 1990's [4]. Regardless, these significant contributions worked as a springboard paving the way for future evolution.

Advancements

Over time supervised learning techniques expanded in complexity. Decision tree algorithms emerged and allowed for effective classification tasks. The pioneer of this section was J.R Quinlan with his research paper "Induction of Decision Trees" [11]. From here newer techniques moved away from simple linear regressions to more sophisticated models. The next step was using polynomial regress which allowed for fitting more complex curves to data if the data was nonlinear [8]. Going further, Bayesian methods started to join the mix and introduced the concept of priors, enabling models to use existing information to make better predictions. Supervised learning would continue to branch out from all these newly developed techniques and go on to discover new areas in machine learning.

Impacts

Now supervised learning techniques can be found in almost all industries from education, military, finance, healthcare… etc. In healthcare machine learning has presented "numerous opportunities for improving patient outcomes and reducing healthcare costs… including personalized treatment plans, disease diagnosis and detection" [6]. As well as changing the financial services sector by "enabling more accurate predations and better risk management". One of the newest areas of impact are the developments made in transportation and autonomous vehicles. Using supervised learning to predict the maintenance of vehicles and to power autonomous vehicles. It's clear that supervised learning is going to continue to make significant contributions in all sections of industry directly impacting everyday life.

Bias in Supervised Learning

Bias in algorithmic systems refers to skewed or prejudiced decision-making within AI models. However, these biases are not inherent to the technology itself, but a reflection of the biases presented in the data used. Recognizing these biases is fundamental in understanding how they can lead to ethical and social implications. There are multiple types of bias in supervised learning. Sample bias and measurement bias are two examples. Sample bias "occurs when the training data is not representative of the population under study" [1]. This usually means that the dataset that is used doesn't correctly represent all groups of people. This is a common problem in computer vision with face detection as some models struggle with detecting faces with darker skin tones. Another bias is measurement bias which occurs when "data collected for training differs from the data collected during production" [5]. An example of this would be if data is collected from a specific camera and then you attempt to run a trained algorithm on another camera, it won't work as well.

Strategies for Bias Detection and Mitigation

Detecting and mitigating bias in algorithmic systems involves employing various strategies. Implementing bias audits and checks helps identify errors in training data. One way to detect bias is "programmers normally examine the set of outputs that the algorithm produces to check for anomalous results" [7]. This method involves comparing the results and investigating the outcomes if it seems a certain group or result seems unfairly rated. Another method for detecting bias is called "Learning Fair Representation which transforms the training data by finding a latent representation that encodes the data, minimizing the information loss of non-sensitive attributes" [2]. Another strategy that can be used is after the model has been trained to correct the classifiers, "These methods adapt a previously trained classification model to obtain a fairer run… adjust any learned predictor to remove discrimination according to the equalized odds and equality of opportunity constraints" [2]. There are tens of other strategies that could be employed but that would be a research paper on its own.

Ethical Frameworks and Regulatory Measures

Ethical considerations and regulations are essential components in ensuring fairness in supervised learning. This area of machine learning is still growing and developing day by day. At first, a rudimentary implementation was suggested by Wallach and Allen [13]. They suggested a top-down and bottom-down approach in which the top-down was "concerned with borrowing moral frameworks from philosophers…" then in the bottom-up approach "the machine learns through manipulation like a child learns while growing up" [14] Another research paper titled "Ethical Framework for Machine Learning" aims to explore a new framework to manage ethical issues. They propose to rate the designer of the ML algorithm on a scale combination of EQ, (emotional quotient), and SQ, (Spiritual Quotient). These two scales have a low and high scale leading to 4 possible types of designers. Low EQ and low SQ result in a mechanical learner, low EQ and high SQ produce an Ethical learner, high EQ Low SQ result in a cognitive learner, and finally a high EQ and high SQ result in an ethics master [14]. As mentioned before these frameworks are all far from perfect and are continuously being worked on.

Emerging Trends

Currently, the field of ML/AI is growing rapidly with new technologies arriving daily and companies adopting these new trends internally and in their products. Some of the newest trends to emerge have been continuous learning and human-in-the-loop learning. Continuous learning is described as “a machine learning approach that enables models to integrate new data with explicit retraining” [3]. The advantage of this new trend allows for “more robust and accurate” models. As well as the “retention of information” and “adaptability”. The areas that can be impacted by these developments are computer vision, cyber security, healthcare, and robotics. Where models need to be continuously updated to ensure that they’re constantly up to date. One of the next trends we’ll discuss is called “Human in the loop”. This method is a way to train models quicker by having a human give feedback to a model as it provides outcomes. The two main advantages of this are that it “improves the accuracy of rare datasets” and “improving safety and precision.” An example of its application would be if you’re looking for information in a “language that is only spoken by a few thousand people the machine learning algorithm may not find any examples to learn from” [12]. Now with HITL, you can correct the algorithm as it runs to ensure that it’s running correctly. These are only a few of the number of trends that have recently emerged, as time goes on, we will see the surge of even newer and more complex ideas.

Conclusion

In summary, supervised learning has gone through many evolutions and changes since its inception. It can reshape the future of various industries. The power of predictive modeling has ability to personalize treatment plans, recognize speech, detect faces, driverless vehicles, improve manufacturing… etc. However, as this sector of machine learning continues to grow, there must be ethical considerations and bias mitigation to ensure fair technology. The future of supervised learning looks expansive and exciting, since we are yet to find out what all its possible implications truly are.

References

[1] Cotton, Richie. “Data Demystified: The Different Types of AI Bias.” *Data Camp*, July 2022, https://www.datacamp.com/blog/data-demystified-the-different-types-of-ai-bias.

[2] Fernandez, Franklin. “Bias Mitigation Strategies and Techniques for Classification Tasks.” *Holistic AI | AI Governance, Risk and Compliance Software*, 8 June 2023, https://www.holisticai.com/blog/bias-mitigation-strategies-techniques-for-classification-tasks.

[3] Ferreiro, Yolanda. “What Is Continuous Learning? Revolutionizing Machine Learning & Adaptability.” *Datacamp* , July 2023, https://www.datacamp.com/blog/what-is-continuous-learning.

[4] Foote, Keith D. “A Brief History of Machine Learning - DATAVERSITY.” *DATAVERSITY*, 3 Dec. 2021, https://www.dataversity.net/a-brief-history-of-machine-learning/.

[5] Tim, Jones. “Machine Learning and Bias.” *IBM Developer*, IBM, 2019, https://developer.ibm.com/articles/machine-learning-and-bias/.

[6] “The Impact of Machine Learning on Society.” *Software House Boring Owl | Warszawa | Web & Mobile*, 19 Apr. 2023, https://boringowl.io/en/blog/the-impact-of-machine-learning-on-the-modern-society.

[7] Lee, Nicol Turner, et al. “Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms | Brookings.” *Brookings*, 22 May 2019, https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/.

[8] “Machine Learning Basics: Polynomial Regression.” *Medium*, 4 June 2020, https://towardsdatascience.com/machine-learning-basics-polynomial-regression-3f9dd30223d1.

[9] Pradhan, Ashirbad. “Decision Tree ID3 Algorithm |Machine Learning | by AshirbadPradhan | Medium.” *Medium*, Medium, 11 June 2023, https://medium.com/@ashirbadpradhan8115/decision-tree-id3-algorithm-machine-learning-4120d8ba013b.

[11] Quinlan, J.R. Induction of decision trees. Mach Learn 1, 81–106 (1986). https://doi.org/10.1007/BF00116251

[12] Wolfewicz, Arne. “Human-in-the-Loop in Machine Learning: What Is It and How Does It Work?” *Levity | No-Code AI Workflow Automation Platform*, 16 Nov. 2022, https://levity.ai/blog/human-in-the-loop.

[13] W. Wallach and C. Allen, Moral Machines: Teaching Robots Right From Wrong, Oxford University Press, 2009.

[14] C. Malhotra, V. Kotwal and S. Dalal, "ETHICAL FRAMEWORK FOR MACHINE LEARNING," 2018 ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K), Santa Fe, Argentina, 2018, pp. 1-8, doi: 10.23919/ITU-WT.2018.8597767