

Yidi Liu

Dec 1st, 2023

Intern Technology Solutions Analyst Prompt Report

Machine learning and deep learning, as applications of AI technology, have been a major technology trend in recent years. With the illustrious exploits of AI products including AlphaGo and GPT, this branch of AI technology has shown the world its priceless potential, which grants its important role in any industry that involves data, pattern recognition, and decision-making. This article introduced machine learning techniques and why deep learning is a suitable research point for Cotiviti.

1. Background:

1.1 Machine Learning and Deep Learning

Machine Learning is a concept and a technology that was raised in the 1960s. The mission of machine learning is to “give computers the ability to learn without being explicitly programmed”, defined by Arthur Samuel, the pioneer of machine learning. ^[1] As an important branch of machine learning, neural network was inspired by human neurons. Between its input and output layers, there exist one or more hidden layers to simulate complex information transmission in biological neural networks. The more hidden layers a network has, the deeper it is. When a network is deep enough, we then have a DNN (deep neural network) for deep learning, which is a critical tool for NLP and graph and speech recognition. ^[2,3]

1.2 Mission of Machine Learning

As stated above, the key mission of machine learning models is to learn. Depending on the model selection, a trained model can achieve different tasks, including classifying the class of data points, predicting the value (or class) of unseen data points, inferring what features determine the outcome of data points, clustering similar data points, and detecting anomalies/outliers. ^[4]

2. TPO Specificities and Model Choice:

To decide what model Cotiviti, as a company closely related to healthcare TPO (healthcare Treatment, Payment, and operations) data, should make use of, we need to know what makes the uses of machine learning different in healthcare TPO from other fields. Healthcare-related datasets have many special characteristics: a) they contain sensitive PII (Personally Identifiable Information), b) their volumes are normally huge, c) they can be complex and contain both structured and unstructured data, d) they can be imbalanced, and e) the model produced by them must be explainable. ^[5] These specificities bring challenges for machine learning model selections but also point out the direction where we should go. Actually, if we say we need a model that can easily adapt complex data composed of normal structured data as well as unstructured data like pictures or recordings while maintaining a good performance when being trained with an enormous amount of data, the answer is obvious: Deep Learning. ^[6]

With deep learning's simplicity which removes the need for feature engineering, scalability that enables parallelization on GPUs or TPUs, and versatility in applications like image classification, object and speech recognition (inference), and speech and image synthesis, it, deep learning, with its ability or potential ability to accomplish classification, prediction, inference, clustering, and detecting anomaly, is no doubt the best choice of Cotiviti at this moment. Though points a), d), and e) are still not directly answered by deep learning, they are still solvable. For point a), the practise of introducing HIPAA into model constructions has already been proved possible by providers for example Amazon AWS. ^[7] As for points d) and e), they are the "traditional problems" of machine learning, and they can be solved with techniques including over/undersampling and SHAP.

3. Conclusion:

Overall, considering the role of Cotiviti as a company heavily related to TPO datasets, investment in evaluating deep learning services from providers including AWS for SageMaker, Google for DeepMind, etc.

Bibliography

- [1] Esposito, Mark, et al. "What Is Machine Learning?" The Conversation, 1 Mar. 2023, theconversation.com/what-is-machine-learning-76759.
- [2] Al-Gahmi, Abdulmalek. "Module C: Linear Models and Neural Networks." BMI 6015. 2023.
- [3] "Compare." Amazon, River Transportation Division, Iowa Dept. of Transportation, 1981, aws.amazon.com/compare/the-difference-between-deep-learning-and-neural-networks/?nc1=h_ls.
- [4] "What Is the Difference between Prediction and Inference?" Cross Validated, 1 Jan. 1963, stats.stackexchange.com/questions/244017/what-is-the-difference-between-prediction-and-inference.
- [5] LeSueur, Dan. "5 Reasons Healthcare Data Is Unique and Difficult to Measure." Health Catalyst, 23 Jan. 2023, www.healthcatalyst.com/insights/5-reasons-healthcare-data-is-difficult-to-measure.
- [6] "Trainingoptions." MathWorks, 2023, www.mathworks.com/help/deeplearning/ug/deep-learning-with-big-data.html.
- [7] Engdahl, Sylvia. "Blogs." Amazon, Greenhaven Press/Gale, 2008, aws.amazon.com/blogs/machine-learning/aws-expands-hipaa-eligible-machine-learning-services-for-healthcare-customers/.