UNIVERSITY OF WESTMINSTER

INFORMATICS INSTITUTE OF TECHNOLOGY

UNIVERSITY OF WESTMINSTER

TRENDS IN COMPUTER SCIENCE

4COSC008C

Machine Learning

Overview of Machine Learning. Describe and compare two different machine learning techniques.

- Mohamed Akram Gazali
- UOW Number - W1956088
- IIT Number - 20221038

GROUP MEMBERS

1. Sakith Dissanayaka - W1956131
2. Ahamed Amhar - W1956142
3. Hansith Fonseka - W1956144

Contents

INTRODUCTION

Machine learning (ML) is a component of artificial intelligence that gives computers the ability to learn by watching examples (data), making predictive inferences based on that data, and then using those predictions to make new decisions without human input or guidance. According to Tom Mitchell "A computer program is said to learn from **experience E** with respect to some **task T** and some performance **measure P**, if its performance on T, as measured by P, improves with experience E". This means that a machine can be said to learn if it can gain experience through performing a certain task and enhance its performance in carrying out related activities in the future. (Chandramouli, Dutt, and Kumar D, 2018)

The four types of machine learning techniques include (Sarker, 2021):

1. Supervised Learning
2. Unsupervised Learning
3. Reinforcement Learning
4. Semi-supervised Learning

In this report, a brief overview of ML and a comparison between supervised and unsupervised learning have been made.

SUPERVISED LEARNING

What is supervised learning?

The training stage of a machine learning system is referred to as supervised learning. The machine learning algorithm is given sets of **labeled** data during the supervised training phase. Each set contains the desired (right) outcome as well as the inputs the algorithm should utilize to make predictions. (Mishra, 2020)

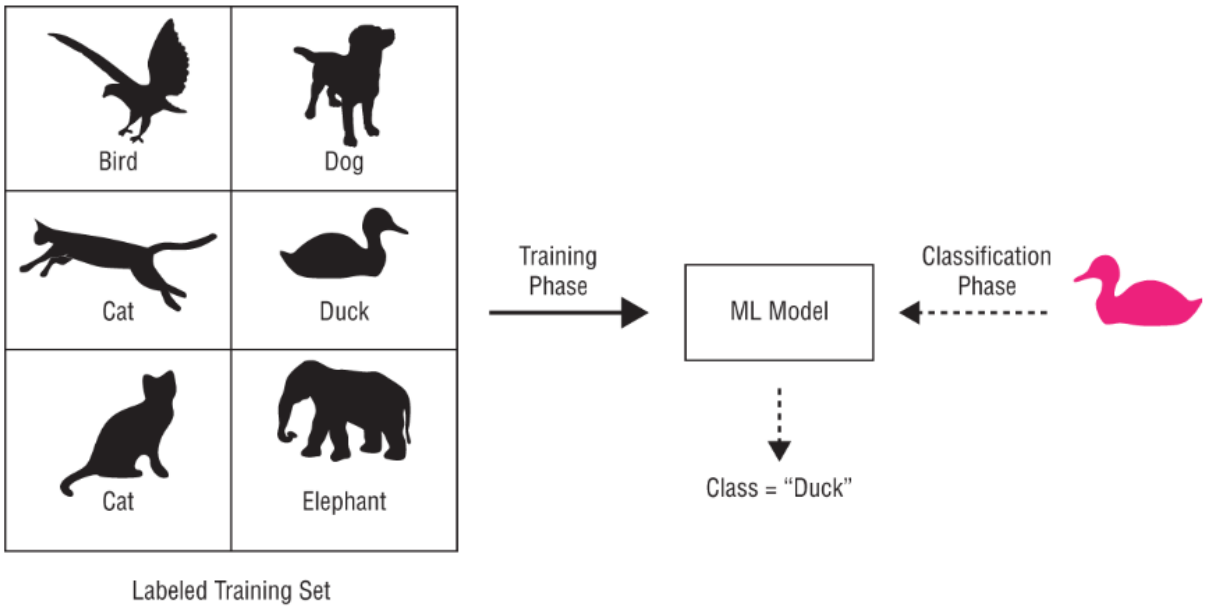

Figure 1: Labeled training set (Mishra, 2020)

How does the algorithm work?

During training, each set is iterated through while the algorithm is given inputs and its output is compared to the desired outcome. The algorithm's parameters are changed to bring the output closer to (or equal to) the desired output using the difference between the actual output and the desired output. The important factor is the **quantity and quality** of data that dictate the model's accuracy. The algorithm can predict outcomes on data that it has never seen before after it has mastered making predictions on the training set. (Mishra, 2020)

UNSUPERVISED LEARNING

What is unsupervised learning?

Using machine learning algorithms, unsupervised learning analyzes and organizes **unlabeled datasets**. These algorithms identify data clusters or hidden patterns without the aid of a human. The three common unsupervised learning algorithms are **clustering**, **association**, and **dimensionality reduction**. (Mohammed, Mohammed. and Khan, 2017)

How does the algorithm work?

K-means clustering

Data points are divided into K groups using the K-means clustering technique, which determines the number of clusters depending on the distance from the centroid of each group. The clustering of data into categories will depend on which data points are closest to a specific centroid. Smaller K values indicate bigger groups and less granularity, whereas greater K values indicate smaller groupings and higher granularity. Market segmentation, document clustering, picture segmentation, and image compression all often employ K-means clustering. (Mahesh, 2019)

The below diagram will help to understand the algorithm.

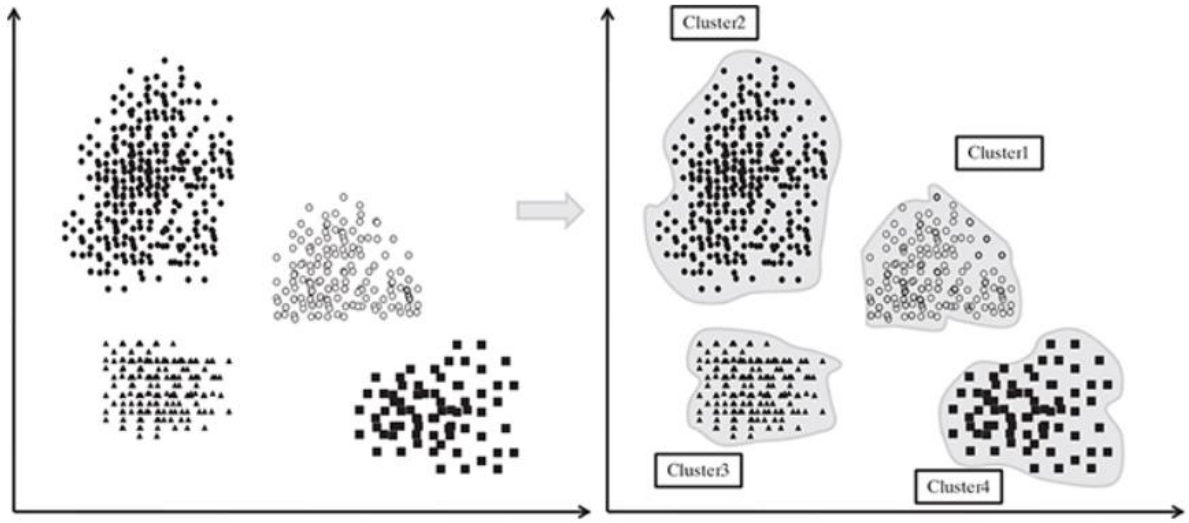

Figure 2: K-means algorithm. An example of conducting two rounds of the k-means algorithm Each training example is allocated to the nearest cluster centroid in each cycle. (Chandramouli, Dutt, and Kumar D, 2018)

COMPARISON BETWEEN SUPERVISED AND UNSUPERVISED LEARNING

Labeled vs unlabeled data

The usage of labeled datasets is the key differentiator between the two methods. In simple terms, supervised learning uses labeled input and output data, while an unsupervised learning algorithm does not. The labels for the input and output variables need to be accurate, and training supervised learning models might take a while. Unsupervised learning techniques, on the other hand, might produce radically inaccurate results unless there is human validation of the output variables. (Brink, 2017)

Output and human intervention

In supervised learning, predictions are iteratively made on the data, and adjustments are made for the proper response. This process allows the algorithm to "learn" from the training dataset. Unsupervised learning models are more likely to be inaccurate than supervised learning models, while supervised learning techniques need human interaction up front to identify the data correctly. In contrast, unsupervised learning models operate independently to identify the underlying structure of unlabeled data. Consider the fact that algorithms still need some human involvement for output variable validation. (Rhys, 2020)

Goals

An unsupervised learning algorithm's goal is to extract insights from enormous volumes of new data. The machine learning algorithm itself determines what is distinctive or intriguing about the dataset. (Brink, 2017)

Application

Spam identification, sentiment analysis, weather forecasting, and price forecasts are just a few applications where supervised learning models dominate. Unsupervised learning, on the other hand, is ideal for, recommendation engines, customer segmentation, anomaly detection, and medical imaging. (Delua, 2021)

CRITICAL EVALUATION

Machine learning can uncover patterns in massive amounts of data that humans cannot see or locate. Supervised learning allows us to classify large amounts of data, yet the outcomes are very accurate and reliable. Unsupervised learning, in comparison, can process enormous amounts of data instantly. However, there is a lack of openness regarding the data clustering and a greater chance of unreliable outcomes.

Therefore, another type of learning such as semi-supervised learning can help in this situation. A good middle ground is semi-supervised learning, which makes use of a training dataset that contains both labeled and unlabeled data. It's especially helpful when there is a lot of data and it's challenging to extract the important elements from it.

CONCLUSION

In this report, I have briefly described the overview of machine learning and its different types of techniques and explained how machines are programmed to learn by itself. The two fundamental concepts of supervised and unsupervised learning have been discussed. These complex learning algorithms must then be taught using real-world data and information about the target application that has been gathered before the system can support intelligent decision-making.

Furthermore, a comprehensive comparison between the two techniques has been made to understand the differences and similarities. Finally, I have critically evaluated the two techniques. It is expected that it will give insight to the readers to make informed decisions when choosing between the two types.

In a competition in March of last year, the top Go player in the world, 18-time champion Lee Se-dol, was defeated by the artificial intelligence (AI) computer program AlphaGo from Google (Menlo Park, California), winning 4 of 5 games. In conclusion, machine learning is a field that is here to stay and grow exponentially over the next decades to come.

REFERENCES

1. Chandramouli, Dutt and Kumar D. (2018). Machine Learning. India: Pearson Education India. Available from https://tinyurl.com/yzjn44t8 [Accessed 1 November 2022].
2. Mishra, A. (2020). Machine learning for iOS developers, 1st edition. Hoboken, New Jersey: Wiley. Available from https://tinyurl.com/yhv6u7p3 [Accessed 1 November 2022].
3. Mohammed, M., Mohammed. and Khan, M.B. (2017). *Machine learning : algorithms and applications*, 1st edition. Boca Raton: CRC Press. Available from https://doi.org/10.1201/9781315371658 [Accessed 1 November 2022].
4. Brink, H., Richards, J.W. and Fetherolf, M. (2017). *Realworld machine learning*, 1st edition. Shelter Island, New York: Manning Publications. . Available fromb https://tinyurl.com/yc2anv3p [Accessed 30 October 2022].
5. Delua, J. (2021). Supervised vs. unsupervised learning: What's the difference? *IBM*. Available from https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning [Accessed 1 November 2022].
6. Mahesh, B. (2019). Machine Learning Algorithms -A Review. *Machine Learning Algorithms - A Review*, 9 (1). Available from https://doi.org/10.21275/ART20203995 [Accessed 2 November 2022].
7. Rhys, H. (2020). *Machine Learning with R, the tidyverse, and mlr*, 1st edition. Shelter Island, New York: Manning. Available from https://tinyurl.com/4bbwrc98 [Accessed 1 November 2022].
8. Sarker, I.H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*, 2 (3). Available from https://doi.org/10.1007/s42979-021-00592-x [Accessed 2 November 2022].