Research Topics in Computer Science

A brief introduction to Machine Learning, the IoT, and Big Data

John Hornik
Research in Computer Science
Assignment 2
Lewis University
Romeoville, IL, USA
johnphornik@lewisu.edu

***Abstract*—This paper presents three major research areas in computer science: Machine Learning, the Internet of Things (IOT), and Big Data. All three of these areas are explored in an introductory manner in order to give the reader a better understanding of what they entail and how they are related.**

Keywords—Machine Learning; Big Data; IOT

I. Introduction

In recent years, technology has brought great advancements to many facets of everyday life. From the cars we drive to the homes we live in and the watches we wear on our wrists. Electronic devices are getting smarter and more efficient. There's the machine learning algorithms that help devices learn on their own as well as the actual interconnected devices that these algorithms are implemented into. From there, there are terabytes or even petabytes of data that come from these machines which companies use to help better their organizations so that they can help in advancing these technologies even more. In the following sections, there will be a brief overview of these concepts within the fields of machine learning, the internet of things, and big data.

II. Machine Learning

A. Introduction to ML

According to Arthur Samuel, a pioneer in the fields of computer science and artificial intelligence, machine learning is the science of getting computers to learn without being explicitly programmed [1]. In a sense, it is giving them the ability to almost think for themselves. This technology allows machines to perform many tasks that previously could only be performed by human beings. In many cases, they can execute these tasks better, faster, and more intelligently than humans can. Recent advancements in computing power have allowed machine learning to be applied to more complex problems. The exponential increase in data has also allowed scientists and engineers to apply machine learning to even more domains than ever before [2]. Some applications for machine learning include face recognition, recommendation systems, defect detection, robot navigation, and video games.

B. How it works

The broad idea behind machine learning is that the "machine" or program learns by studying a training set of examples. This then allows it to apply what it has learned to new data which it has not encountered yet. In general, it implements three strategies; supervised learning, unsupervised learning, and reinforcement learning [2].

C. Supervised Learning

In supervised learning, the training set contains data and the correct output of the task with that data. This can be compared to a math class where the teacher gives a student some math problems with their worked-out solutions. The student can then solve similar problems using the work and solutions given. This type of learning utilizes classification algorithms which take as input a dataset and the class of each piece of data which, in turn, helps the computer learn how to classify new data. Logic regression, classification trees, random forests, and artificial neural networks (ANNs) are all examples of classification algorithms. Supervised learning also employs regression algorithms which predict the value of an entity's attributes. Linear regression, decision trees, Bayesian networks, fuzzy classification, and ANNs are all examples of regression algorithms [2].

D. Unsupervised Learning

In unsupervised learning, the training set contains data but no solutions. Using the previous example, this can be compared to the teacher giving the student some math problems and some work but the student has to come to the solutions on his or her own. This type of learning utilizes clustering algorithms that take as input a dataset which can cover many dimensions. These algorithms can then partition this dataset into clusters satisfying certain criteria. One of the more popular of these methods is k-means clustering. This method partitions n observations into k clusters where each observation belongs to the cluster with the nearest mean. Hierarchical clustering and Gaussian mixture models are both examples of clustering algorithms. Unsupervised learning also implements dimensionality reduction algorithms which take the dataset and project it to fewer dimensions. This gives a clearer view of the data's fundamental aspects. Principal component analysis, random projection, and tensor reduction are examples of dimensionality reduction [2].

E. Reinforcement Learning

Reinforcement learning is the problem of getting an agent to act in the world so as to maximize its rewards [3]. This can be exemplified with a chess game. If reinforcement learning were to be implemented, the machine would not know which moves

will win the game, however, it would know whether a game is won or lost. Utilizing this, it can then play many games and slowly begin favoring the moves that produce winning combinations. Fundamentally, in reinforcement learning, an agent receives an observation which includes a reward. It then chooses an action which causes the environment to move to another state. As it progresses, it attempts to collect as many rewards as possible. All of its actions are compared to the actions of an agent that performs optimally. This, in turn, gives it an ability to reason long term consequences as well. This can be related to someone who spends money to fix their car before the issue creates bigger and more expensive problems in the long run. Even though there is an initial high cost, it can prevent significantly higher costs in the long term. Robot control, telecommunications, and gaming are some fields which utilize reinforcement learning [3].

F. Machine Learning Tools

Among all of the tools that a programmer can have, R and Python are the two most utilized in machine learning. R is more popular among those using machine learning for statistical approaches due to the fact that it contains many libraries that relate to statistical applications. Python is more popular among those using machine learning in a computer science-based approach. This is also due to the fact that its libraries focus on numerical and scientific computing. Both of these languages, however, work well with machine learning as long as the dataset being worked on is small enough to fit in the computers main memory. In cases where it does not, a distributed platform such as Hadoop or Spark must be utilized. Spark, for example, is sort of a scripting environment in which the user can interact with it using more known languages such as R, Python, or Java [2].

III. INTERNET OF THINGS

In today's world, with all of the smartphones, smart homes, smart cameras, smart cars, and everything else "smart", the internet of things (IoT) has risen to become one of the hottest topics in computer science. Most people can think of the internet as the global system of computer networks which are connected using specific protocols. The internet allows us an abundance of information and makes connecting with others as easy as a click-of-a-button on our personal computers. What many people don't yet realize - but encounter on a daily basis - is the growth of physical and portable objects also allowing for an endless abundance of information and make connecting with others as easy as a click-of-a-button. The IoT can be defined as the interconnection via the internet of computing devices embedded in everyday objects, enabling them to send and receive data [4]. Many IoT devices share some common attributes: wired or wireless connectivity, movement of data to the cloud, and applications to configure as well as monitor the device. The following sections will present some examples of common IoT devices that are utilized today as well as security concerns that are commonly associated with these devices.

A. Consumer Electronics

In 2016, Gartner Inc. estimated that there were 6.4 billion IoT products in use [4]. That number has surely grown since then. Wearable IoT devices such as smartwatches, pulse oximeters, and heart-rate monitors are readily available and are convenient in monitoring vital signs of the user. There are devices which are part of the IoT which don't come to mind as quickly as a wearable watch. The ReTiSense Stridalyzer, for example, is an intelligent sole which tracks foot pressure and calculates stress and movement as well as number of steps [4]. Even car makers have jumped on the IoT bandwagon. Bluetooth connectivity allows for merging phone and the cars audio system permitting hands-free calling. Further advances have allowed manufactures to even implement artificial intelligence into their automobiles to make them smarter. In the last year, BMW has partnered with IBM's Watson in order to research AI systems in driver-assist function to push the boundaries of self-driving automobiles. Watson's conversational and machine learning capabilities will allow BMW owners to ask questions and receive answers from the AI. Watson will also have the capability to incorporate traffic updates and weather information to make long trips easier to plan for drivers [6]. One of the most popular applications of IoT are in home automation. Command and control systems such as Amazon Echo or Google Home provide the capability of voice control over other IoT devices such as smart lighting, security systems, thermostats, and TV's [4]. Smart locks, such as Kwikset's Kevo, allow for better security for the home. This system can be controlled using a regular key, a key fob, a smartphone, or a regular internet connection and can help a home owner monitor and track the access of their home.

B. Smart Home

Apart from consumer electronics such as the Amazon Echo, IoT can also be implemented into an entire home rendering it a "smart home". Developments such as this can make a home not only more advanced and efficient but also safer. In IEEE Intelligent Systems [5], a case is brought up where a smart home can be used to house elderly people. This is an application that can potentially save a life. The home is presented as having a monitoring system which implements many IoT devices to measure the physical and cognitive health conditions of the older adults. This system can detect if a person falls and also study behavioral patterns which, in turn, can generate alerts when something seems out of the ordinary. Wearable sensors can be placed on the subjects living within the home. Depth sensors can be placed in rooms for fall monitoring. Wireless motion sensors can monitor daily movements as well. With all of this, if a subject is sensed to be laying on the ground for a certain period of time (when this falls out of usual behavior patterns), an alert may be generated which can signify a possibility that the subject has fallen. As mentioned earlier, such advancements can not only make living in a home easier for elderly adults but can also potentially save a life.

C. IoT Security

One of the most popular discussions related to IoT devices is their security. As these devices continue to grow, their security becomes an immediate concern with users and stakeholders. This inherently becomes a significant topic since companies developing IoT devices need to be sure that their customers and end users trust these devices with their security and privacy. There have, in fact, already been many documented cases of hackers putting systems in jeopardy. In 2014, Proofpoint security firm found that a smart refrigerator sent more than 75,000 spam and phishing e-mails due to malware. In the last year, a hotel in Austria (Seehotel Jagerwirt) had to pay a ransom

to hackers after they hacked into the hotel's computer system and prevented them from creating key-cards for the guests. In 2016, internet infrastructure provider Dyn was hit with an attack which took down websites and even caused outages as well as network congestion across the United States. The main problem with IoT device security is that low-cost devices enter the market with few security protections. This is due to default passwords and the lack of encryption and authentication in order to keep costs low and make the devices easier to use. One way to combat this is to simply apply traditional security fixes that aren't currently being implemented. Device-to-gateway encryption and gateway-to-cloud encryption should be performed. Other techniques such as rate limiting, secure booting, access control, and firewalls can also be implemented in order to make IoT devices safer. Using these techniques coupled with a stronger post-design quality assurance process would surely play a critical role in assuring more secure IoT devices in the future.

IV. Big Data

With the billions of IoT devices that are in use every day as well as other data sources such as transactions, social media, emails, etc., there are bound to be massive amounts of data generated. This data can be used to help organizations expand and make their products even smarter. It can also help companies make better strategic decisions, control their processes, better understand their customers, and reduce costs. These are the concepts behind big data – the extremely large datasets that can be analyzed to reveal patterns and trends relating to human behavior and interactions. In general, big data is believed to involve three dimensions; volume, velocity, and variety [8]. Volume pertains to the amount of data collected by organizations from sources including business transactions, social media, and IoT devices. Velocity is the idea that data streams in at extreme speeds and needs to be processed in a timely manner. Variety pertains to the types of data that are received – structured and unstructured. This section will present some of the benefits that big data introduces as it relates to business.

A. Challeneges

One of the challenges that companies face when dealing with big data is infrastructure due to the fact that there are many different technologies that have become available that contain different usage, licensing, and cost. In order to be successful, a company must choose the correct infrastructure based on its present and future goals. Business Models also pose a difficulty because companies must devise feasible ones in order to determine whether exploitation of big data is worthwhile for their business plan. The final challenge is standardization due to companies requiring standardized data models and architectures in order to support the exchange and processing of data.

B. Potential Analysis of Big Data

As mentioned in IEEE Computing Edge [8], companies these days are tempted to begin buying big data hardware and software infrastructures and collecting whatever data they can retrieve. They then typically end up with massive datasets and pricey infrastructures that don't necessarily provide them with the answers they were initially seeking. In order to avoid situations such as these, an organization needs to first have clearly defined goals. From there, they can then decide on how big data solutions can improve their business plan. This technique is known as potential analysis and can prove to be extremely beneficial for an organization. Potential Analysis is broken up into two main parts: benefit analysis and readiness analysis.

1) Benefit Analysis

Benefit analysis is a technique which initially identifies opportunities and then defines business goals in order to meet those opportunities. For each goal, the organization can decide on big data related strategies which it can implement in order to achieve those goals. Essentially, identifying the information needed in order to implement strategies and its connection with big data sources is the main aspect of benefit analysis.

2) Readiness Analysis

In readiness analysis, an assessment is made of the business's capability in regards to available data, infrastructure, and ability to implement big data strategies.

In general, big data is an impressive tool which companies can use in many ways. They can incorporate it in business models in order to identify past patterns and predict future ones. This can result in better strategies, lower costs, and more advancements in technology.

V. Conclusion

As advancements in technology continue to grow year by year, as do the topics and research relating to them within the realm of computer science. There are many different subject areas that can be covered, however, almost all of them overlap in one way or another. This paper has presented machine learning, the internet of things, and big data. Although these may seem like entirely different areas, they are in fact closely related. Machine learning is a technique that has become extremely popular in recent years due to such things as IoT and big data. Many IoT devices implement machine learning as well as other artificial intelligence techniques as the basis of how they operate. Because these devices have grown so popular and there are so many of them, massive amounts of data are produced as a result and this data can be used productively in order to help organizations continue expanding and creating more advancements in technology for years to come.

References

[1] A. Munoz. *Machine learning and optimization* [Online]. Available: https://www.cims.nyu.edu/~munoz/files/ml_optimization.pdf

[2] P. Louridas and C. Ebert, "Machine learning," IEEE Computing Edge, pp. 8-13, April 2017.

[3] K. Murphy. (1998). *A brief introduction to reinforcement learning* [Online]. Available: http://www.cs.ubc.ca/~murphyk/Bayes/pomdp.html

[4] W. Wong, "IoT for the consumer," Electronic Design, pp. 14-16, November 2016.

[5] T. Banerjee and A. Sheth, "IoT quality control for data and application needs," IEEE Intelligent Systems, vol.32, pp. 68-73, March/April 2017.

[6] F. Guerrini. (2016, Dec 15). *BMW partners with IBM to add watson's cognitive computing capabilites to its cars* [Online]. Available: https://www.forbes.com/sites/federicoguerrini/2016/12/15/bmw-partners-with-ibm-to-add-watsons-cognitive-computing-capabilities-to-its-cars/#787ec7b11a90

[7] B. Albright, "IoT security by design," Digital Engineering, pp. 39-42, April 2017.

[8] J. Heidrich, A. Trendowicz, and C. Ebert, “Exploiting big data’s benefits,” IEEE Computing Edge, pp. 9-14, March 2017.