

Applicability of A.I. in Big Data

Aditya Anand
Student, Department of Computer Applications
Maharishi Markandeshwar University,
Mullana
Contact No. +91- 8409422218
E-Mail: anand.aditya.anand@gmail.com

Abstract: Big Data and Artificial Intelligence are the most field in today's era. AI adds an intelligence layer to big data to tackle complex analytical tasks much faster than humans. Artificial Intelligence (AI) is a broad field of study focused on using computers to do things that require human level of intelligence. It has been around since 1950s, mainly for playing games. But it had no direct fruitful practical application due to hardware limitations. It was in 1980s, when boom in microprocessor industry has changed the way we use computer in our daily lives. But even then, the use of computers was limited to multi-national corporate high-level employee office desktop. Then Artificial Intelligence was used mostly for filtering spam mails, segregating important mails and in some scientific and defence projects. These artificial intelligences were mostly hard coded. i.e. the intelligence was merely based on simple programming of if-else and loops.

Poonam Bhatia Anand
Faculty, Department of Computer Applications
Maharishi Markandeshwar University, Mullana
Contact No.- +91-9996401586
E-Mail: Poonam.anand@mmumullana.org

With the ever-growing use of computer in every sector, the enormous amount of data is producing every second. The source of data are machines, social media, mobile devices etc. Huge amount of data is generated from these sources which are either in unstructured or structured forms. This Data is termed as Big Data. Most of data generated from various devices are in unstructured format which is difficult to handle by traditional methods and techniques. In this paper, the concept of digital data and various forms of data is reviewed.

AI can be used as key concept to reveal the value of big Data. The concept of Big Data and Artificial Intelligence is reviewed in this paper. The methods and techniques of AI for handling big data are elaborated. The different ways of capturing, structuring and analysing big data using AI are discussed in this paper. The applicability of AI in handling Big Data is explained with examples.

Keywords: Digital Data, Unstructured Data, Semi-Structured data, Natural Language Processing, Machine Learning

Artificial Intelligence

Artificial intelligence is considered as artificial entity, which can do work which require human level or equivalent to human level of intelligence. Though there are various domains of Artificial intelligence, it's application has found its way to work in fields of economics, medicine, defence and even into something as simple as sorting pictures and music as per our individual like.

Evolutionary mechanism

When a baby is born, he is completely unaware of its immediate environment. He learns about its surrounding by perceiving through its sense organ. It has no explicit teacher, nature is its teacher. His interaction with surrounding generates lots of information about its causes and effect, about consequences of action and what about to do to achieve goals. Interaction with outer world and experience is biggest teacher in human life. With regular interaction, any human can achieve perfection and mastery in its work. Like driving, typing, writing etc. A human brain learns by comparing the punishment and achievement obtained by its action of work. [1]

Inspired by human brain, the concept of neural network (Deep learning, subdomain of machine learning) came up. Just like human brain which learns by information gained by its sense organ and comparing the result with predicted result or punishment/achievement,

an artificial intelligence learns by analysing raw digital data and comparing the result with predefined parameters. Once the machine has learned it can process task and keep evolving with data feedback.

Training Vs Programming

In last 3 decades computers have greatly influenced the way we live and work. Their usage has increased tremendously, from scientific supercomputers to personal mobile devices. With such increase in computational power and digital data collection, it is very hard to code an Artificial intelligence system which can work on such diverse and huge amount of digital data. When a computer is presented with a picture, it doesn't see an image, it sees pixels, which are numbers to represent some values on pre-set scale. Everything else, we want them to know, we would have them to describe them in terms of pixel.

Machine learning takes a different approach. Rather than providing instructions (program), computer is provided with few thousand examples (Training). These examples are well labelled. The system learns the difference between them and creates a dictionary of its own. Now, the computer has built its own program to identify new group of digital data to segregate them for post processing. [2]

The Big Bang in Artificial Intelligence

In the current decade, the computers have completely changed the way we live our lives. Marketing, socializing, work, education, communication, business, commerce, defence and even politics are now done on the digital frontier. A digital grid to which we all are part of.

DATA:

This is the era of digital data. Every day, we humans combined generate around ~2.5 quintillion bytes of digital data. At individual level known as 'digital footprint'. [3]

Every industry of every sector, with the help of A.I manipulate this dataset to target their customer very specifically. Crowd sourced datasets are built through creative approaches for example, Facebook asking users to "tag" friends in their photos to create labelled facial recognition datasets. More complex datasets are generated manually by experts - e.g. asking radiologists to label specific parts of the heart.

Computer designed to learn:

Software

Artificial neural networks (deep learning), inspired by the human brain, are structured to understand complex concepts. Furthermore, just like human child brain, their strength lies in their ability to learn from data and feedback. The "deep" in deep learning refers to many layers of artificial neurons, each of which contribute to the network's performance.

Hardware

The computer that powers this kind of artificial intelligence is intensive, but not complex. Processing such huge datasets through CPU are often work intensive. Since CPUs are made to perform various kinds of process. Custom made GPUs (parallel processing Graphical Processing Unit) with help of special programming library (CUDA, OpenCL) can perform this task in ~1/10th of resource that a CPU will take. Furthermore, such single hardware setup can have multiple GPUs to perform complex calculations while a low budget single CPU is enough to instruct them.

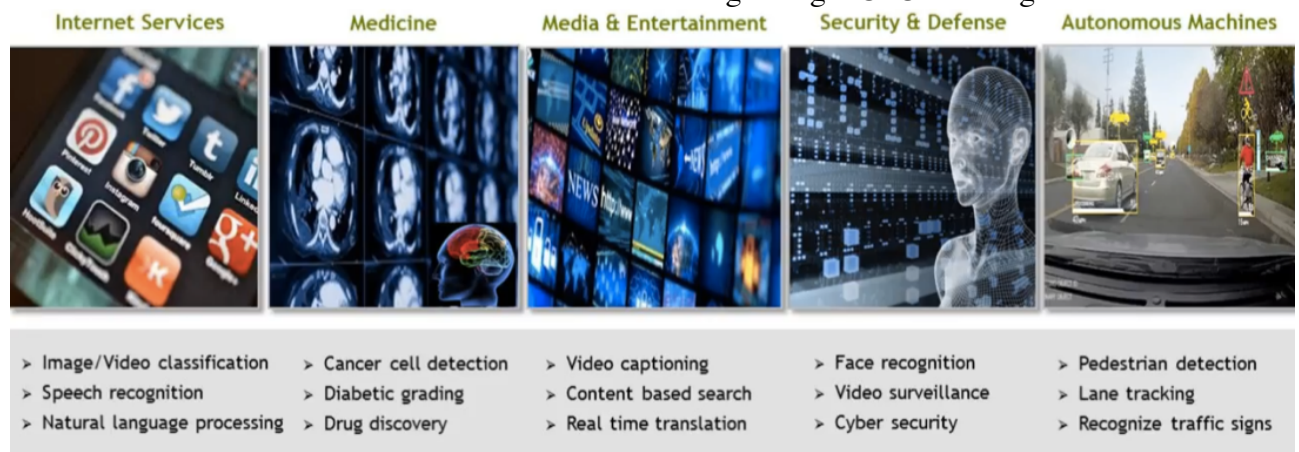


Fig [1]: - various sector where A.I can be utilized

Scope of A.I for developers

As discussed above, artificial intelligence has become a buzz word for everyone. Boom of digital data and its industry, and their penetration into every other sector has opened an ocean of opportunity for Computer science engineers. But with the growth of A.I, the job of IT and marketing persons will go scarce. A computer engineer with knowledge in programming and in development of A.I and big data can easily grasp a desired work place in upcoming technological scenario.

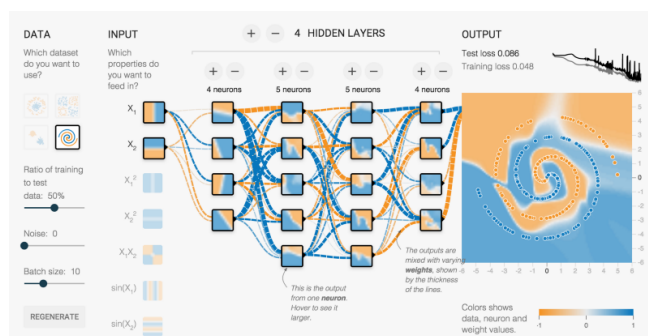
When it comes to development, artificial intelligence is often referred to as a field for total nerds and geniuses. But with the evolution of various frameworks and prebuilt libraries, it has become a friendlier IT field.

Top AI Frameworks

TensorFlow:

Developed by Google, TensorFlow is an open-source software. It is a highly scalable AI framework with an architecture that it can run almost on any kind of device. Let it be a CPU or GPU powered desktop, server, cloud computer or even a mobile device. This framework is available in python and can also be integrated with C++ program. [4]

Fig [2]: - brief insight into google tensor flow



CNTK:

Developed by Microsoft to counter Google's TensorFlow, CNTK is an abbreviation for 'Computational Network Toolkit'. It provides the user with various libraries to enhance learning algorithms and model descriptions. [5]



Fig [3]: - Microsoft cognitive toolkit

Caffe:

Caffe is a fast, powerful, efficient and open source framework developed in c++. It is developed by Berkeley AI Research (BAIR) and by community contributors. Yangqing Jia created the project during his PhD at UC Berkeley. Caffe already powers academic research projects, start-ups prototypes, and even large-scale industrial applications in vision, speech, and multimedia. It is easy to learn and well supported by online development community. [6]

Applicability

Business Intelligence & Analytics

Studying market trend and future demands often requires analysis of large sets of digital data. To remain ahead in ever increasing competition, time is a very crucial parameter. Manual or tool-based analysis of huge sets of digital data requires a significant amount of time. Furthermore, it very difficult to monitor and analyse live data feed.

Introducing of A.I in business can help a business to tap the power of automated digital modelling and predictive analysis. Integrating products with its users on social media generates lots of review. These reviews can be analysed with the help of sentimental analysis (natural language processing) to generate the overall emotional insight of users on the product. Emotional shift (like or dislike) over a feature can be reviewed to improve the next iteration of the product. Moreover, the emotional insight from the products of other business organisation can be utilised in the development of a new product.

Election Campaign

Big data analysis can be used in election campaign to micro target a certain set of voters. Combining these small set of voters can play a very crucial role in manipulating the national/local election.

Door to door step campaigners collect the data and feed them to a bigger digital database (such as local issues, email id, phone number, address and often campaigners try to convince voters to join camping through social media or telephone). These databases are aligned with social media digital data generated through data mining. With the help of Artificial Intelligence (natural language processing (sentimental analysis), predictive analysis) the

political parties can deduce the demands (from the upcoming government), emotional shift and needs of the voters. The report generated is very specific to demographic distribution (based on cast, religion, work groups, profession and even gender). The sorted digital data is sent back to lowest hierarchy of campaigners to convince back the voters to vote for them. Furthermore, these analysed data are exploited by leaders during election campaign speech, held locally.

Defence and National security

Terrorism and corruption are one of the most immanent threat to a nation. To prevent such threats to a nation, the national security teams need to keep updating the intelligence. Since gathering intelligence is mostly the work of surveillance through various sources. Interpreting, analysing and segregating important intel requires human level of intelligence. Sadly, the time taken to gather the raw digital data and segregating intel from this huge dataset requires relevant amount of time. The time bound is very important thing in national security.

As a response to this, an Artificial Super Intelligence (ASI) can be built which has right to read every email and monitor phone call, bank transaction, surveillance feed and every possible digital data with their associated user id. The objective of the ASI should be to predict and prevent threats to a nation.

Such kind of surveillance can be questioned for violating and abusing privacy of the citizens on mass scale. The ASI should only output the well-established national identity number (such as Adhaar or passport number). It should be upon the national security agencies to investigate and identify the threat to the nation.

Since the ASI is a non-living entity, and outputs an id number only, it doesn't violate any privacy. To prevent the ASI catastrophe the system should have very restricted and contained interaction with the outer world.

Conclusion:

Artificial intelligence has been there for more than 5 decades but was heavily restricted on its application due to hardware and software limitations. Most of the advancement has taken place in present decade. Reduction in price of hardware and contribution to the software by pool of developers, scientists and mathematicians around the globe has catapult the application of Artificial intelligence.

Use of Artificial intelligence to analyse huge sets of digital data can benefit an organisation, government or even to an individual by optimising the data sets as per use. This ever-increasing demand of A.I will firmly drive the IT industry job towards itself. It is very important for the young generation of developers to learn and understand this domain of computer science.

In recent times, use of artificial intelligence has proved itself to be very effective. Start-up business who utilised AI to study the insight of customers and other prevailing products, have benefited a lot. Election campaign has become more and more aggressive and people demand specific. Artificial intelligence can be utilised to by government to gather intelligence and analyse them in real time. Deployment of artificial super intelligence can play a very crucial role removing immediate threats to a nation.

References

- [1]. Bellman, Richard Ernest. *"The Theory of Dynamic Programming."* Santa Monica, CA: RAND Corporation, 1954
- [2]. Richard S. Sutton, Andrew G. Barto *"Introduction to Reinforcement Learning"* MIT Press Cambridge, MA, USA ©1998 ISBN:0262193981
- [3]. <https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN>
- [4]. Source: - [tensorflow.org](https://www.tensorflow.org)
- [5]. Source: - github.com/Microsoft/CNTK
- [6]. Source: - <http://caffe.berkeleyvision.org>