International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105

JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Volume 9, Issue 8 - August 2021-2022 - Pages 1-5

MACHINE LEARNING AND BIG DATA ANALYTICS APPROACHES FOR IMPROVED PREDICTIVE PERFORMANCE

1.Prof Sangamesh S K, 2 . Prof S B Ullagadi , 3. Prof Pavitra M G
Department of Computer Science and Engineering,
RTE Society's Rural Engineering College Hulkoti

Abstract—: Machine learning is a technology that allows computers to learn directly from examples and experience in the form of data. Traditional approaches to programming rely on hardcoded rules, which set out how to solve a problem, step-by-step. In contrast, machine learning systems are set a task, and given a large amount of data to use as examples of how this task can be achieved or from which to detect patterns. The system then learns how best to achieve the desired output. It can be thought of as narrow AI: machine learning supports intelligent systems, which are able to learn a particular function, given a specific set of data to learn from.

INTRODUCTION

In some specific areas or tasks, machine learning is already able to achieve a higher level of performance than people. For other tasks, human performance remains much better than that of machine learning systems. For example, recent advances in image recognition have made these systems more accurate than ever before. In one image labeling challenge, the accuracy of machine learning has increased from 72% in 2010, to 96% in 2015, surpassing human accuracy at this task. However, human-level performance at visual recognition in more general terms remains considerably higher than these systems can achieve.

While not approaching the human-level intelligence which is usually associated with the term AI, the ability to learn from data increases the number and complexity of functions that machine learning systems can undertake, in comparison to traditional programming methods. Machine learning can carry out tasks of such complexity that the desired outputs could not be specified in programs based on step-by-step processes created by humans. The learning element also creates systems which can be adaptive, and continue to improve the accuracy of their results after they have been deployed.

Machine learning lives at the intersection of computer science, statistics, and data science. It uses elements of each of these fields to process data in a way that can detect and learn from patterns, predict future activity,

Branches of machine learning There are three key branches of machine learning:

• In supervised machine learning, a system is trained with data that has been labeled. The labels categories each data point into one or more groups, such as 'apples' or 'oranges'. The system learns how this data – known as training data – is

structured, and uses this to predict the categories of new – or 'test' – data.

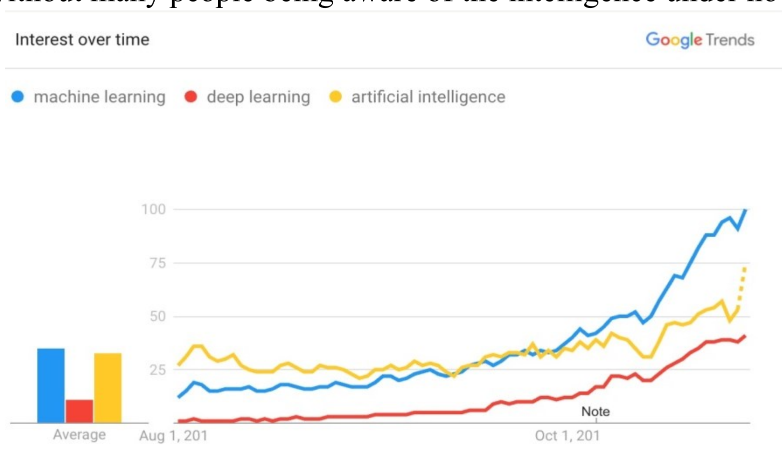
• Unsupervised learning is learning without labels. It aims to detect the characteristics that make data points more or less similar to each other, for example by creating clusters and assigning data to these clusters.

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Volume 9, Issue 8 - August 2021-2022 - Pages 1-5

• Reinforcement learning focuses on learning from experience, and lies between unsupervised and supervised learning. In a typical reinforcement learning setting, an agent interacts with its environment, and is given a reward function that it tries to optimize, for example the system might be rewarded for winning a game. The goal of the agent is to learn the consequences of its decisions, such as which moves were important in winning a game, and to use this learning to find strategies that maximize its rewards.

When machine learning systems are deployed, there is a key distinction between offline and online learning systems: machine learning in daily life. The term 'machine learning' is not one with high salience for the public; research by the Royal Society and Ipsos MORI showed that only 9% of people recognize it. However, many people are familiar with specific applications of machine learning, and interact with machine learning systems every day. Common applications include commercial recommender systems, virtual personal assistants, image processing, and a range of other systems which are pervasive, without many people being aware of the intelligence under hood.



I.ORIGIN OF MACHINE LEARNING

Despite the recent attention given to, and hype surrounding, machine learning, fundamental ideas in the field are not so new, with early papers being published over sixty years ago.

Within the last decade, even the past five years, the field of machine learning has made revolutionary advances. These advances have been driven in part by the availability of large amounts of data and the accessibility of computing power, but also underpinned by algorithmic advances achieved by revisiting and re-envisioning the simple neural networks put forward in the 1940s and 1950s. Drawing further insights from physiology and neuroscience, artificial neural networks have been created in which hundreds of layers of processing allow systems to perform more complicated tasks. These so-called deep learning techniques have been responsible for some of the more high-profile recent advances in artificial intelligence research, such as the AlphaGo system's victory over Lee Sedol, acknowledged as the strongest human player at the game of Go, in March 2016 .

This recent revolution means that technologies such as voice recognition or image processing, which a few years ago were performing at noticeably below-human levels, can now outperform people at some

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Volume 9, Issue 8 - August 2021-2022 - Pages 1-5

tasks. One way of thinking about many (but not all) applications of machine learning is that the algorithms learn from examples – called training data. The recent success of machine learning in matching or even improving on human behavior for some tasks, including image and speech analysis, is due in no small part to the recent explosion of training data in these areas: the vast number of examples on which to train the algorithms has been a critical part of their improved performance.

Machine learning is therefore both a method that requires data and a tool that enables uses of it; access to data is required to create machine learning methods and train machine learning systems, and these systems can be put to use in making sense of the large, and growing, amount of data available today. In extracting valuable information from data, machine learning can help realize the social and economic benefits expected from so called 'big data'.

In turn, machine learning requires a 'machine friendly' data environment, based on open standards that make using open data easier. This chapter assesses some of the issues around data availability for machine learning.

text" should *not* be selected.

II.MACHINE LEARNING HELPS TO EXTRACT

Ninety percent of the world's data has been created within the last five years76. In this age of 'big data', an increasing volume of information is being collected, from a greater range of sources, and at greater speed than ever before. Image or video uploads to social media, GPS-enabled devices, and other online activities are generating stores of data, as people spend more of their work and leisure time online. This all contributes to the creation of an estimated 2.5 billion gigabytes of data per day77.

These changes to the volume, variety, and velocity78 of data collection have created a potentially rich resource for the digital economy. One estimate suggests that open data could help create $3 trillion of value each year for the global economy. Early indicators of economic growth show promise in this area; digital industries grew 32% faster than the rest of the UK economy from 2010 to 2014, with employment in these sectors growing 2.8 times faster than in other sectors of the economy.

In this new data environment, considerable economic benefits are therefore at stake: data has been described as the 'new oil' for the digital economy; a resource with the potential to support a new industrial revolution.

The nature of this new data environment challenges traditional analytical approaches in data science and statistics, as the complexity of big data can limit the effectiveness of existing methods of analysis. The complexity and scale of data available today therefore demands new tools to create valuable insights. In this context, machine learning has a vital role to play in making sense of large quantities of potentially dynamic data. It can process volumes of data that would be unmanageable for humans, picking out the patterns that subsequently become useful. Machine learning therefore extracts value by deriving new insights from the mass of data, and in turn data is needed to develop machine learning, by training systems to detect patterns or make predictions.

As noted above, data has been described as the new oil; holding incredible economic potential, but requiring refinement in order to realize this. If not the new oil in and of itself, then data is at least the

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105

JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Volume 9, Issue 8 - August 2021-2022 - Pages 1-5

fuel for machine learning, and a data environment that enables the effective use of data will be key to enabling machine learning to be put to use, and hence to deliver its promised benefits.

III. ALGORITHMS

- The Naïve Bayes Classifier Algorithm.
- K Means Clustering Algorithm.
- Support Vector Machine Algorithm.
- Apriori Algorithm.
- Linear Regression.
- Logistic Regression.
- Artificial Neural Networks.
- Random Forests.

IV.CREATING DATA ENVIRONMENT TO SUPPORT MACHINE LEARNING

Individual, public, private, non-profit, and academic – and in a range of formats. Each of these data sources comes with specific challenges, and the diversity of sources
requires new approaches to managing data in a machine-friendly way. Open data is defined as “data that is published under a license with express permission to re-use, share and modify”. Some of this data is public, but not all public data is open, nor does it need to be. Data can be accessible without being usable, for example owing to practical considerations relating to its quality. Conversely, data which is not open can be made accessible to certain users via framework or access agreements.

Data availability across sectors is important Public sector data can be a key enabler Access to public sector data could catalyze a range of economic activity: the direct value of public sector information to the UK economy has been estimated at £1.8 billion82, with wider social and economic benefits from this totaling up to £6.8 billion83.

There are different sorts of data held in the public sector. Some of this is social, some is not directly related to individuals, while some (for example in the NHS) relates to confidential personal information. Healthcare and related data raises particular issues which are addressed later.

V.CONCLUSION

Machine learning approaches applied in systematic reviews of complex research fields such as quality improvement may assist in the title and abstract inclusion screening process. Machine learning approaches are of particular interest considering steadily increasing search outputs and accessibility of the existing evidence is a particular challenge of the research field quality improvement. Increased reviewer agreement appeared to be associated with improved predictive performance.

REFERENCES

1.Y. Sun et al., "CS2P: Improving Video Bitrate Selection and Adaptation with Data-Driven Throughput Prediction", *Proc. SIGCOMM 2016*, pp. 272-85.

2.B. Mao et al., "Routing or Computing? The Paradigm Shift Towards Intelligent Computer Network Packet Transmission Based on Deep Learning", *IEEE Trans. Computers*, 2017.

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105

JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Volume 9, Issue 8 - August 2021-2022 - Pages 1-5

3.K. Winstein and H. Balakrishnan, "TCP Ex Machina: Computer-Generated Congestion Control", *Proc. ACM SIGCOMM Computer Communication. Rev.*, vol. 43, no. 4, pp. 123-34, 2013.
4.M. Dong et al., "PCC: Re-Architecting Congestion Control for Consistent High Performance", *Proc. NSDI 2015*, pp. 395-408.
5.J. Jiang et al., "CFA: A Practical Prediction System for Video QoE Optimization", *Proc. NSDI 2016*, pp. 137-50.
6.W. Richert, L. P. Coelho, “Building Machine Learning Systems with Python”, Packt Publishing Ltd., ISBN 978-1-78216-140-0 [2]
7.J. M. Keller, M. R. Gray, J. A. Givens Jr., “A Fuzzy KNearest Neighbor Algorithm”, IEEE Transactions on Systems, Man and Cybernetics, Vol. SMC-15, No. 4, August 1985 [3]
8.https://www.geeksforgeeks.org/machine-learning/ [4]]
9.S. Marsland, Machine learning: an algorithmic perspective. CRC press, 2015.
10.M. Bkassiny, Y. Li, and S. K. Jayaweera, “A survey on machine learning techniques in cognitive radios,” IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1136–1159, Oct. 2012.
11.https://en.wikipedia.org/wiki/Instance-based_learning
12.R. S. Sutton, “Introduction: The Challenge of Reinforcement Learning”, Machine Learning Page 225-227, Kluwer Academic Publishers, Boston, 1992
13. P. Harrington, “Machine Learning in action”, Manning Publications Co., Shelter Island, New York, 2012..