Nesta...

MACHINES THAT LEARN IN THE WILD

Machine learning capabilities, limitations and implications

July 2015

INTRODUCTION

Like many other areas of Artificial Intelligence, the technical capabilities of machine learning approaches are regularly oversold and this hype overshadows the real advances. Machine learning algorithms have become an increasingly important part of our lives. They are integral to all sorts of applications from the speech recognition technology in Siri to Google's search engine. Unfortunately machine learning systems are often more noticeable in our lives because of failures rather than successes. We come face-to-face with the limitations of auto-text recognition daily, while spam filtering algorithms quietly remove mass mail from our inboxes completely unnoticed. Improvements to machine learning algorithms are allowing us to do more sophisticated computational tasks. But it is often unclear exactly what these tools can do, their limitations and the implications of their use - especially in such a fast moving field.

ABOUT THIS REPORT

This short report comes out of a workshop exploring the capabilities and limitations of machine learning algorithms. Rather than a complete resource looking at the specific capabilities of different algorithms, this report is an introduction to some of the current capabilities and limitations in the field. It includes some areas where machine learning approaches are employed effectively and the challenges when using the technology.

AUTHOR

Harry Armstrong, Technology Futures Researcher at Nesta.
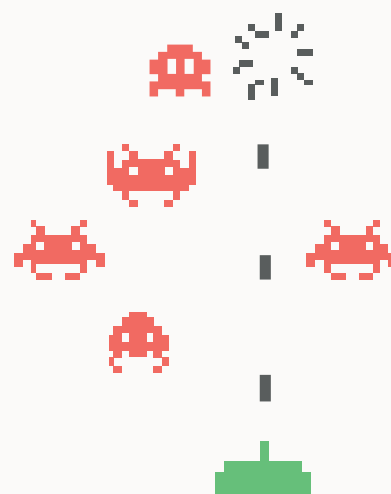
ACKNOWLEDGEMENTS

Nesta...

Nesta is an innovation charity with a mission to help people and organisations bring great ideas to life.

We are dedicated to supporting ideas that can help improve all our lives, with activities ranging from early–stage investment to in–depth research and practical programmes.

www.nesta.org.uk   **©Nesta 2015**

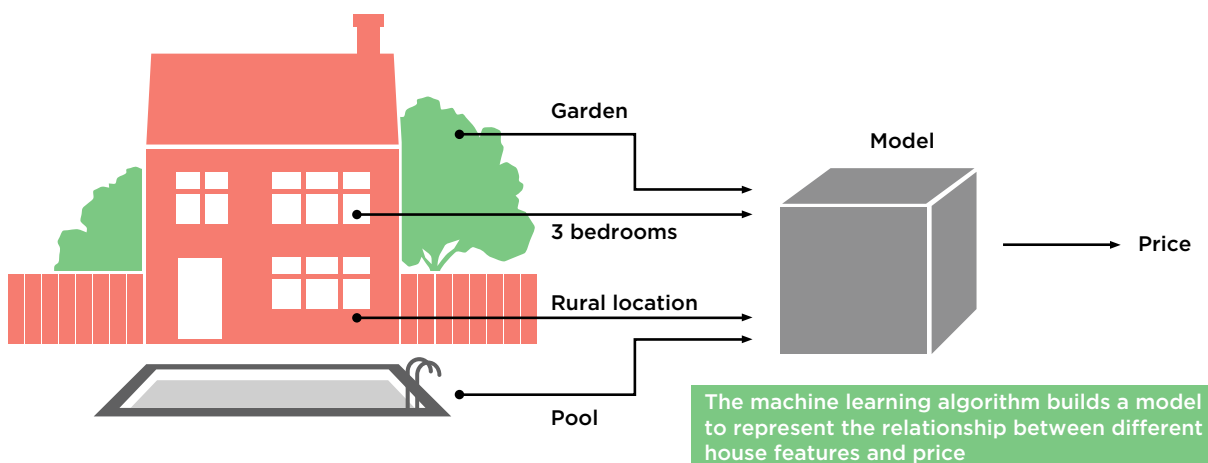# MACHINES THAT LEARN IN THE WILD

CONTENTS

THE PROBLEM

These techniques can be applied to a vast array of problems and when used effectively produce powerful insights. Unfortunately, machine learning tools are often treated like a silver bullet, a magical tool able to solve an organisation's problems. Machine learning is not magic, just a set of algorithms and data processing. Overreliance on the technology or a misunderstanding of its abilities could have serious consequences as these techniques become more widely adopted. These computational systems are primarily designed as tools to help people get better at what they do, not as a tool to replace people. An active collaboration between technology and informed human counterparts will ensure the best outcomes.

WHAT IS MACHINE LEARNING?

Machine learning algorithms 'learn' to predict outputs based on previous examples of relationships between input data and outputs (called training data). A model of the relationship between inputs and outputs is gradually improved by testing its predictions and correcting when wrong. Machine learning is a set of computerised techniques for recognising patterns in data. It is a way of generating something like the 'line of best fit'. It's useful to automate this process when the data has many features and is very complex.

For example, lots of factors affect house prices. There is no simple relationship between, say, house size and price. A computer can crunch through details of thousands or millions of houses for sale to model the relationship between different factors and price.

Garden

3 bedrooms

Rural location

Pool

Model

Price

The machine learning algorithm builds a model to represent the relationship between different house features and price

Using these algorithms bypasses the need to write specific code to solve each specific problem and each algorithm can be used to solve lots of different problems by adapting the model to fit different data sets. The field of machine learning uses many of the same techniques and ideas as other fields, such as statistics - from which it also borrows methods but uses different nomenclature. There are multiple types of problems that machine learning algorithms can be used to solve. The most common are classification, regression and clustering.

Classification Labelled data is used to train the algorithm so it can guess the label to attach to new unlabelled data. The algorithm is effectively modelling the differences and similarities between groups or classes. Examples include spam email filtering and fraud detection from unusual transactions or payments.

Classification might be used with the house market data to find rural properties, when they haven't been labelled as such. Having two or more of a bundle of features like 'farmland', 'near village' or 'own water supply' may together predict whether a property is rural or not.

Regression Data is given a real value rather than a label. The algorithm must predict values for new data. The price of a stock or a market over time is an example application.

Regression could be used to calculate house prices based on the historical local market data. The analysis would test out several factors including property size, proximity to amenities or whether it has a garden or not. It would then look for a strong relationship between these factors and price variation. A model for new prices could be developed based on these relationships.

Clustering Data is unlabelled but can be divided into groups based on similarity or other measures of structure within the data. The algorithm tries to find the hidden structure of the data.

Clustering could be used to try and discover new determinants of a house's price. Taking a price range, say £250,000 to £350,000, a clustering algorithm can create a map that groups houses together that share similar features. There might be a group that are small but urban. There might be another group that share period features and gardens. By comparing the groups across different price ranges, the analysis would start to show segmentation in the market and how it changes as prices increase.

Supervised, Unsupervised and Reinforcement learning

Machine learning algorithms can be broadly split into categories based on how they learn:

Supervised learning requires a training data set with labelled data, or data with a known output value (e.g. rural/not rural or house price). Classification and regression problems are solved through supervised learning.

Unsupervised learning techniques don't use a training set and find patterns or structure in the data by themselves. Clustering problems can be solved with an unsupervised approach.

Semi-supervised learning uses mainly unlabelled and a small amount of labelled input data. Using a small amount of labelled data can greatly increase the efficiency of unsupervised learning tasks. The model must learn the structure to organise the data as well as make predictions.

Reinforcement learning uses input data from the environment as a stimulus for how the model should react. Feedback is not generated through a training process like supervised learning but as rewards or penalties in the environment. This type of process is used in robot control.

There are a few key limitations of machine learning approaches that impact on their usefulness for certain tasks, as well as their ability to function in real-world environments. Machine learning algorithms function very well on tasks related to familiar data from a training set. Limitations tend to surface when the algorithm tries to incorporate new data. As these systems advance, they are quickly becoming better at categorising familiar data and performing tasks such as image or speech recognition. Improvements in understanding what a picture or sentence means or interpreting it for use in other contexts have been much slower.

Assumptions and inductive bias

Machine learning algorithms will make assumptions about the 'best' function that fits the data. It is possible to find multiple functions that fit with a given training data set. To choose one, the machine learning algorithm will need to make assumptions about what the function being modelled looks like. This inductive bias guides the learning algorithm to one hypothesis or function over another. Assumptions can be about the distribution of the data; many algorithms assume the data lies in a standardised distribution or that some factors are more relevant than others when classifying data. For example, the *minimum description length* assumption is akin to Occam's razor where the simplest solution is considered more likely to be correct. An algorithm that uses this assumption will create a model biased to the simplest function that fits the data.

These assumptions inform the algorithm's ability to cope with new data outside the training set. There is a trade-off between simple models that can generalise well beyond the training data and complex models that better fit the training data. Both simple and complex hypotheses can cause problems when trying to find the best function:

Overfitting The model uses complex hypotheses and focuses on irrelevant factors in the training set limiting the ability to generalise when faced with new data.

Underfitting The model only considers simple hypotheses and therefore excludes the 'real' function.
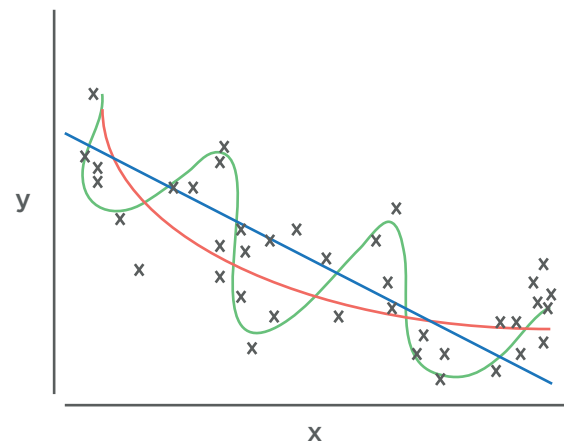
Several different functions might fit the same data set

Assumptions made by the learning system have implications for the model, final outputs, how well the model will cope with new data and how fast the machine learning algorithm can learn.

Deep learning and artificial neural networks

Some of the most promising machine learning algorithms use a deep learning architecture, referring to the many hidden layers in the models that are used. These techniques don't start with many assumptions, and so are not as liable to inductive biases. There is a different kind of issue; it is difficult to extract or interpret how the model is working.

History of the artificial neural network
Artificial neural networks have been around for a while. The original deep learning brain from the 1960s was called the Perceptron. Perceptrons take input signals and produce a single binary output. The simplest, single-layer Perceptron makes an assumption that each input independently affects final classification and interactions between inputs are not possible. It has serious limitations in its ability to perform some basic logic tasks. This resulted in the field losing the majority of its support in the 1970s. Advances came in the 1980s with the introduction of more complex networks, which were able to overcome the past problems but severe limitations still existed. These were called artificial neural networks as they were designed to reflect neuronal processing in the brain. They learnt slowly, inefficiently and were unable to learn basic concepts like the past tense of regular verbs. Despite another crash in
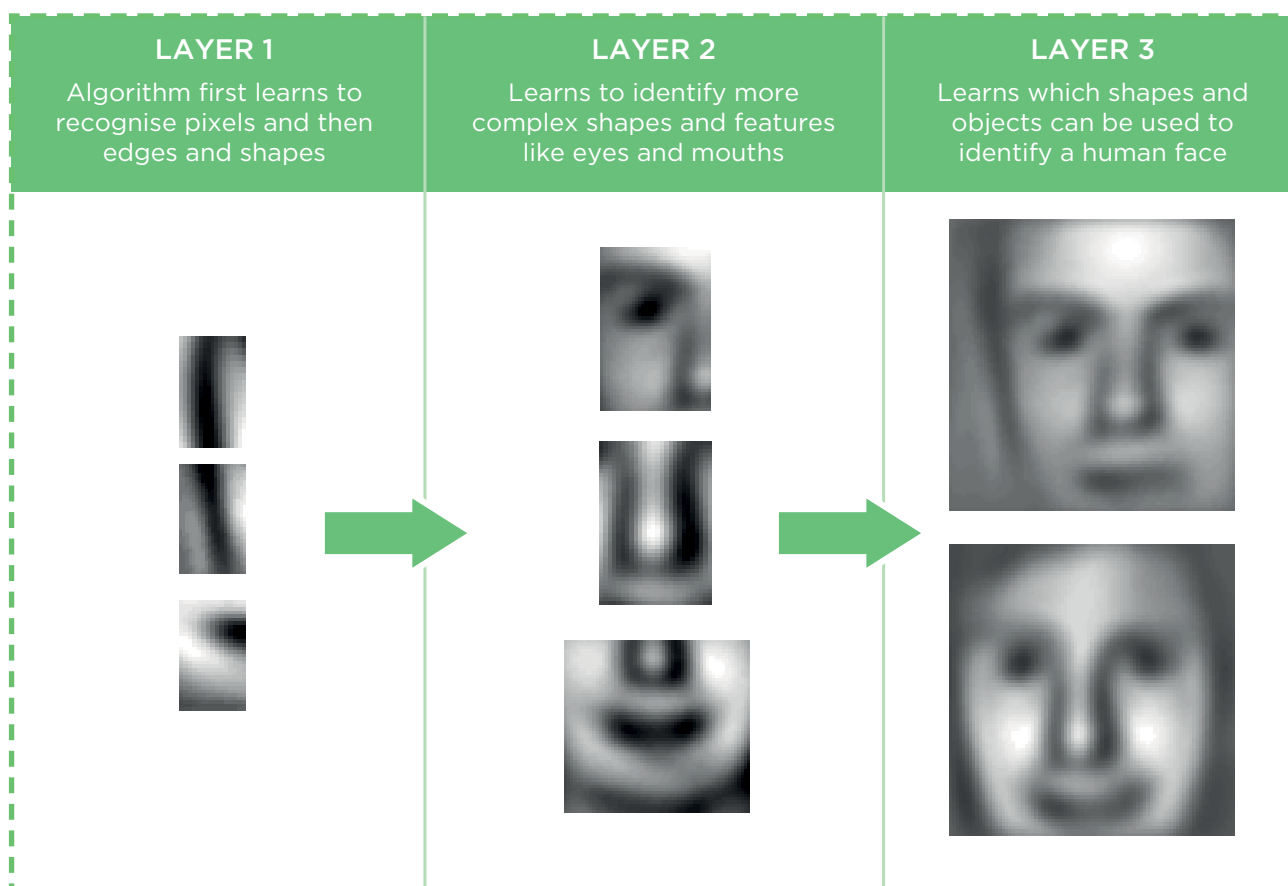
popularity, researchers such as Geoff Hinton[1] still focused on improving these techniques. Another breakthrough came in 2006 with the introduction of deep learning or deep neural networks. These techniques exploit the abundance of cheap computation available to build much larger more complex neural networks and often use large datasets where little of the data is labelled. Many machine learning 'breakthroughs' in recent years are the development of existing techniques combined with greater, cheaper computing power.

Capabilities and limitations: lessons from deep learning

The important innovation in deep learning is a system that learns categories incrementally through it's hidden layer architecture, defining low-level categories like letters before moving on to higher level categories such as words. In the example of image recognition this means identifying light/dark areas before categorising lines and then shapes to allow face recognition. Each neuron or node in the network represents one aspect of the whole and together they provide a full representation of the image. Each node or hidden layer is given a weight that represents the strength of its relationship with the output and as the model develops the weights are adjusted.

| LAYER 1	LAYER 2	LAYER 3
Algorithm first learns to recognise pixels and then edges and shapes	Learns to identify more complex shapes and features like eyes and mouths	Learns which shapes and objects can be used to identify a human face

This distributed representation is very powerful when it comes to many different problems such as language processing, visual object recognition, speech recognition, natural language processing, automatic image captioning and prediction models for molecule interactions. Deep learning is particularly effective on large data sets, where the likely patterns in the data are unknown. It can be extremely effective at unsupervised learning, but is limited when the number of possible models is large. It doesn't do well with smaller data sets. Deep learning is outperformed by other algorithms when some prior information about patterns in the data is known.

Using the Atari computer games as a test bed,[2] Google DeepMind has continued to develop the abilities of its deep learning system. DeepMind do two things that make their work particularly promising for large scale applications. First, they use back propagation. This means improving the model by looking at how well the output data can predict 'backwards' onto the input data, using a gradient-descent model to minimise the error in this backwards prediction. Second, they have a model that records a history of previous input and output data. It changes the algorithm based on all of the data available rather than just the most recent data in the stream.
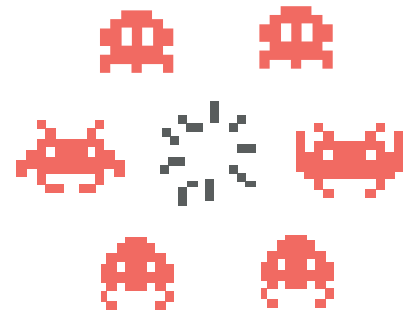
But there are still fundamental limitations to what it is capable of doing. The games that it failed to achieve human like performance required longer term planning or more sophisticated pathfinding. It is also unable to incorporate transfer learning, prior knowledge about the world, so has to learn each game from scratch even if it has played it in the past. Solutions are being developed to overcome some of these fundamental limitations but it does illustrate we are still a long way away from any kind of general intelligence.

The Google system that learnt to recognise cats from millions of internet images[3] works about 70 per cent better than the previous state of the art. This leap in ability illustrates the rate of progress in the field. This system is still only able to recognise objects in the images on which it was trained at an accuracy of 15.8 per cent. But if the pace of improvement continues, this will improve rapidly. Though an algorithm may be able to learn to recognise cats, it still does not understand what a cat is or what the images mean. Developing 'understanding' and common sense reasoning are two of the biggest challenges for machine learning systems and are developing at a slower pace than more technical abilities like image recognition.

Deep learning suffers from a lack of theory around model creation. This can make the process feel like a black box requiring empirical confirmation of the output, unlike other algorithms which provide interpretable models. Given a dataset and network design, there could be two neural networks with different weights but the same result. However, it is possible to interrogate the hidden layers to see what the machine is doing. This has been done for Google cat computer vision paper.

APPLICATIONS

M ost algorithms currently sit in the background of our lives, ploughing through large amounts of data to perform specific tasks like removing spam from our inbox or detecting credit card fraud. The broader applications of these techniques are being explored in areas like medicine and public policy where arguably some of the biggest potential benefits lie.

GAMES

For a long time games have been used as a way of testing machine systems and showing off capabilities. Games can be effective test beds as a less noisy more controlled simulation of real world environments and problems. Famously in 2011 IBM's Watson beat two previous winners in a game of Jeopardy!. Gary Kasparov was beaten at chess in 1997 chess by DeepBlue and today's programs routinely beat chess grandmasters. Within a closed game environment with a specific task they are extremely effective but those same systems are not able to perform other tasks, even very simple things. This is where DeepMind's work is particularly exciting. It was able to adapt to different games with a diverse range of rewards and inputs in many different Atari games. Importantly the system was able to work out what state it was in and its relationship with its environment without prior knowledge – a difficult problem and one that is particularly important in robotics.

Games are not only a way to train or demonstrate abilities. It is also an effective platform to test computational creativity as these are deeply complex creative tasks of which creating art or music are only subproblems. At the Creative Computing Department Goldsmiths College, Michael Cook is exploring the ability of his learning system ANGELINA[4] to understand and devise games that humans would enjoy playing. Initially his work focused on developing simple 2D games which included acquirable skills through powerups to help players complete levels. Later Mike gave ANGELINA the ability to read news articles, search the web and then create games based on current events. The system has been adapted to act more independently and even test new game mechanics simply by looking at code. More recently ANGELINA has been competing directly with humans to create 3D games autonomously.[5] This kind of creative ability may become more valuable as machine learning systems become a bigger part of our lives.

APPLICATIONS

DRIVERLESS CARS

Machine learning algorithms are an integral part of driverless cars and will have an increasingly important role in their operational ability. These learning systems are broadly used for tasks such as image recognition or scheduling but learning in noisy real–world environments is difficult. Trying to record, evaluate and combine lots of different types of data is very challenging. John Leonard at MIT used a dashboard camera to illustrate how many events experienced everyday are unusual by an algorithm's standards[6] and would be difficult for an autonomous car to cope with. The solution so far has been to try and capture as many of these rare events as possible through extensive testing on the road. Using this data, algorithms are used to devise responses which are then tested in simulations. It is hoped that eventually the software will be as safe as a human but there will always be very rare events that get missed.

The Google driverless car relies on the Light Detection And Ranging (LiDAR) system to get accurate information on the car's surroundings which is used by the integrated learning systems to understand the environment and decide on an action. LiDAR cannot see colour or make out signs so machine vision systems are still needed to identify traffic light colour or road signs. Other groups are working on cars that only rely on computer vision. While the hardware is very cheap, the machine learning systems which are the basis for computer vision are not yet good enough. Advances to image recognition algorithms still have a way to go before they'll be able to perform these tasks to the required standard. Safety is the key concern and any system must be able to very accurately understand its environment. In the future when combined with developments in transfer learning these cars could overcome many of the current challenges. However, there may always be a need for a human backup in the system to cope when the machine cannot, especially as the very rare events cannot be planned for through current methods even with extensive testing. If this is the case, many of the potential benefits of autonomous cars will never be realised. If we do decide to trust machines to act autonomously then we need clear regulation on who is responsible if something should go wrong and who provides what kind of oversight.
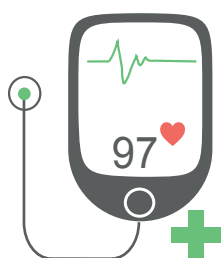
MEDICINE, HEALTHCARE AND MACHINE LEARNING

IBM's Watson is now being employed to sort through, store and analyse massive amounts of medical information. The use of these advanced systems to find new correlations and make predictions will be extremely useful. Some have billed it as potentially the world's best diagnostician. But this ignores the actual challenges and areas where machine learning approaches are having the biggest impact.

When it comes to diagnosis or decision making, machine learning algorithms are not a good replacement for clinicians - at least in most situations. A good diagnosis must take into account structured (e.g. diagnosis codes, medications), unstructured data (clinical notes), image data (X-rays) and even subtle visual cues from the patient (do they look ill, how did they answer family history questions) in a very short time frame. This presents a number of challenges which machine learning techniques are not good at dealing with such as assimilating heterogeneous data sets.

Machine diagnosis through sight and sound

Machine learning techniques have a definite clinical value in relatively simpler problems. Algorithm based analysis of imaging has become a tremendously successful diagnosis tool in recent years. With the growing reliance on radiologic imaging for patient care and improvements to image quality and complexity it is becoming harder for radiologists to interpret the vast amount of imaging information for each patient. Here machine learning capabilities are employed to provide an effective way to automate the analysis and diagnosis of medical images.[7] Open platforms provide a service where radiological images can be uploaded and analysed to find features without a doctor to supervise or direct it. These are extremely accurate methods and provide an automatic way to generalise knowledge gained from training data to unknown test data. These systems are being used in hospitals across the US, Brazil, China and the UK.
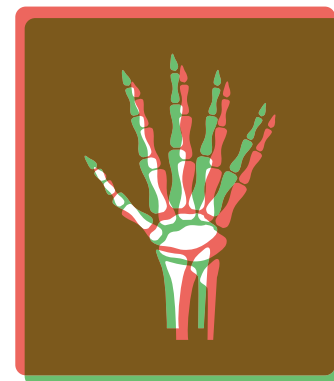
The computational pathologist

The C-path (computational pathologist) system is an excellent example of the benefits and capabilities of using a machine learning system for image based diagnosis. Breast cancer prognosis historically has been mainly based on the identification and quantification of three specific features of the cancer. These are judged by eye through a microscope and used to estimate patient survival rate. Researchers at Stanford used a machine learning based model to create the C-Path program which is able to measure a much richer set of features (6,642 features) in the breast cancer and surrounding tissue.[8] Not only did the system perform significantly better than humans in analysing and evaluating the images but also identified unknown features which were more accurate predictors of prognosis. These newly identified features will also benefit the screening of precancerous tissue and the same technique could be applied to better predicting and evaluating the effect of different cancer treatments.

These learning algorithms have also been successfully applied to picking up on small differences in speech like tremors in someone's voice. Max Little[9] started to apply these learning systems to speech data during his PhD in mathematics and soon realised the potential for diagnosing disease such as Parkinson's. He went on to run a research project where anyone could ring a phone number and record a series of 'aaah' sounds. They are then asked for information: including, whether they have Parkinson's disease, gender and age. The information is used to train the algorithms on the differences between healthy controls and Parkinson's sufferers. This data created without the biases of a laboratory recording situations - nervousness, discomfort - improved the accuracy of vocal Parkinson's tests to 97 per cent.

Personalised medicine and consumer technology

Machine learning techniques will be instrumental in providing personalised medicine and finding value in the emerging large medical and personal data sets from genomics and consumer health technologies. As with Stanford's C-path tool, learning systems will take on the dual role of diagnostician and researcher, finding previously unknown correlations. Research has highlighted the potential of data mining to infer clinically relevant models from molecular data[10] - providing decision-support for genomic medicine.

APPLICATIONS

Many technology companies such as Apple and Google are investing heavily in the analytics that sit behind personalised medicine and consumer technology. Max Little and the Parkinson's research group were part of one of the first projects to collect research data using Apple's iPhone Research Kit. The mPower app (Mobile Parkinson Observatory for Worldwide, Evidenced-based Research) offers iPhone users a series of tests for Parkinson's including finger dexterity and mobility, as well as recording vocal input.

The development and use of these learning systems could be an important step towards providing large–scale personalised medical treatments. These tools will function in partnerships with clinicians. It is important they are designed with this in mind, particularly as historically clinicians use discrete consultations to make decisions; they have not been exposed to the continuous flow of metrics now possible.

MACHINE LEARNING IN GOVERNMENT

Machine learning is currently exploited by a handful of people in government such as the Government Digital Services (GDS) who are using it to predict page views to do anomaly detection[11] or the HMRC who are using clustering techniques to segment VAT customers. Uptake is still limited and there is a great deal of untapped potential. GDS have so far focused on demonstrating the capabilities of machine learning algorithms on a number of products and prototype services. One of the first steps to increasing exploitation of these learning systems is to develop a 'data first' mind-set at a much earlier stage in the policy process. The UK Government has the opportunity to be a leader in this area and benefit from the rich academic and commercial expertise across the country.

The Office for National Statistics (ONS) Innovation Labs

In 2013 the ONS created its Innovation Labs[12] as a resource for learning, research and innovation in big data and machine learning tools. The Labs uses open source technologies not available on the standard ONS secure network overcoming the existing infrastructure barriers. ONS are exploring a number of applications such as the usefulness of Twitter geo-location data to inform student movement after graduation within England and Wales, which has been difficult with existing data. It is this kind of exploration which is needed to find the value in using these techniques, build the internal expertise but also the understanding of potential pitfalls which are discussed in the next section.

READ BEFORE USE: IMPORTANT CONSIDERATIONS WHEN USING THESE SYSTEMS

There are many potential benefits to using machine learning techniques and plenty of future applications but there are also a number of pitfalls and challenges which will impact if and how widely the benefits are felt. These are not technical challenges but the way in which the learning algorithm is deployed and the ultimate goal of its use.

Understanding assumptions for better outcomes

Even with good quality, unbiased data the conclusions drawn by different machine learning approaches can be widely different. Variations in approaches to generating a model can be enough to determine whether a significant correlation is seen or not. This variation arises in part because of the different assumptions algorithms make but no matter what machine learning algorithm is used, analysis of data will always require some difficult subjective decisions to be made about the best models or variables to use. In a closed system like the Atari games where the parameters are fairly simple and the goal is clear, this is not so much of a problem. When it comes to solving complex problems in the real world, this is a big challenge to which there are not technical solutions. There have been some interesting attempts to explore this problem and find ways around it such as the idea of 'crowdsourcing analytics':

Crowdsourcing analytics[13] - questioning methodologies and finding the best solution
When presented with the same data and question (are football referees more likely to give dark skinned players red cards and is this relationship affected by their country of origin?), 29 teams devised their own approach to model the problem. Analytic approaches varied widely across the teams resulting in highly varied effect sizes and a moderate consensus. This was followed by a feedback stage where teams discussed and questioned each other's approaches. Even after this process there was variation in the outputs (20 teams found a significant positive effect while nine teams found a non-significant relationship) but there was a stronger consensus on a positive association.

Using this methodology many approaches can be trialled by different groups and the results brought together and compared. This is potentially a very powerful concept as it encourages the evaluation of different methodologies, the assumptions of those systems and the subjective decisions made in the course of trying to model the problem. Used in the right way this process could create a much deeper understanding of the capabilities and limitations of machine learning approaches. Any system which forces those modelling the data to question their models and assumptions will produce better informed outcomes.

Machines as collaborators

Machine learning algorithms are rarely set up to give a reason for a particular decision or output. This perception of machine learning as an opaque decision–making tool instils a level of mistrust in its outputs. For the likes of physicians or policymakers it is important to have clear justifications for a decision, it is not good enough to rely on the supposed quality of the algorithm. This is particularly important when systems may be prone to errors or the decisions behind the choice of model is not known. People understandably place more trust in humans than machines but this reluctance to trust these learning systems is a big challenge in realising their full potential. A better understanding of how these systems actually operate will alleviate some of these trust barriers but improvements to the design of machine-human interfaces are also need. A machine learning algorithm could fairly easily provide justifications for its decisions. We need to start thinking about computers as better collaborators and new models of interaction. These are tools designed to be used by people and this will require a partnership model, one which is flexible to the needs of the operator and allows the system to work as a collaborator or autonomously when required.

Skills and understanding

If these techniques are to be utilised effectively we need to cultivate informed workers at all levels who can correctly deploy and interrogate the outcomes of algorithmic processes. The 'Crowdsourcing Analytics' experiment highlights the importance of having people who are able to understand the machine learning process and will not simply take its conclusions as correct. In many cases this will require a different kind of data scientist, one that does not have the core technical ability to write code but enough of a general understanding of what can and cannot be achieved using machine learning approaches to effectively evaluate its outputs. This 'type II' data scientist[14] does not need an in depth understanding of the code but might lead a team containing data scientists and needs to be able to translate between the business or policy problem and the technical environment. Without some understanding of what these learning systems can and can't do there is the potential for a lot of poor quality problem solving and the outcomes on society could be very negative. There are examples of courses trying to fill this gap, like the MSc course at Sheffield,[15] targeted at non-data scientists that aims to teach students fundamental data science principles and its application within organisations to support data-driven approaches to problem solving.

Coding errors and quality assurance

Algorithm programming errors are not uncommon. In the majority of cases these bugs are only a nuisance and can be fixed easily after identification. The more we rely on learning systems for important tasks such as driving cars or medical diagnosis the outcomes of these errors might be far more serious. Progress is being made in the methods of software behaviour 'verification' but it is particularly challenging to guarantee that a system built automatically via machine learning methods will behave properly in unpredictable real–world environments. This is an important area for regulation and work will need to be done to decide on what quality standards there should be if any and how they can be tested. Learning systems designed to test other learning systems and find weaknesses could become a useful technical tool in guaranteeing quality. Algorithms generated to master video games have regularly revealed weaknesses in the game's software architecture.

Speed of change and regulation

Maintaining an adequate understanding of capabilities and limitations in the long run will be challenging as the field of machine learning is advancing quickly. Ensuring regulation is able to stay up to date with these advances while not standing in the way of progress is only going to become more difficult. Both of these challenges are closely linked and require the co-ordinated efforts of business, academia, government and broader society to overcome them. Some effort has been made to assess the possible regulation of high frequency trading as part of the UK Government's Foresight work.[16] This report only briefly touched on the use of learning systems but highlighted some of the complexity and difficulties around regulation. Though there is currently little regulation governing this space, there is a positive move in both the public and private sectors to set up ethical frameworks for best practice as a first step. A great deal of thought and consultation has gone into developing these frameworks but continued re-assessment will be needed as the field progresses and capabilities improve. One key area is to ensure the benefits from using these analytical methods are available for everyone, especially when so much technical development is happening in the private sector.

OPPORTUNITIES AND CHALLENGES FOR THE FUTURE

These techniques are very powerful tools and will become increasingly important. Despite their abilities these systems still need human collaboration. This requires people with an adequate level of skill and understanding to effectively evaluate the outputs of the machine learning algorithm. To prevent these systems being perceived as a black box, better user orientated design and an ability to easily interrogate the model behind the outputs could be hugely beneficial. As adoption of these tools grows and rapid advancements open up new abilities ensuring there is a broad understanding of what these systems can and can't do will become increasingly important.

The UK government has an opportunity to become a world leader in applying these techniques effectively to public policy and make use of the extensive expertise which exists here in the UK. Three things will need to happen for this opportunity to be realised:

1. Promote spaces like the ONS' Innovation lab within government departments to allow experimentation, skills development and ability to explore potential issues.

2. Bring more skilled data scientists into government but equally upskill the existing analytical streams of the civil service. Help foster a better, broader understanding of capabilities and limitations to managers and policymakers who will have to translate between the technical and policy worlds.

3. Introduce and trial methods like 'Crowdsourcing Analytics' to expose the importance of subjective decisions in the analytical approach.

A deeper understanding of how to effectively use these systems for policy would be a hugely valuable asset, particularly for recognising where regulation might be needed and what it should look like. The first steps will be to try and understand where regulation might be important, how the implementation of these techniques may impact society and how to deal with the rapid growth in capabilities. For driverless cars or high-frequency trading many of the potential pitfalls and dangers of using machine learning algorithms are well known, but creating frameworks to mitigate these challenges is not straightforward. It will require a great deal of thought and testing. More broadly there may be common issues such as verification of software behaviour or responsibility that will need to be regulated but for many areas the potential importance of regulation is unclear. More focussed discussions between academia, industry and government paired with more speculative foresight exercises would be an effective first step in addressing these challenges.

ENDNOTES

1. See: http://www.cs.toronto.edu/~hinton/

2. Mnih, V. et al., (2015) Human-level control through deep reinforcement learning. 'Nature.' 518, 529-533. February 2015.

3. Le, Q. et al., (2012) 'Building High-level Features Using Large Scale Unsupervised Learning.' ICML 2012: 29th International Conference on Machine Learning, June 2012.

4. See: http://www.gamesbyangelina.org

5. Cook, M. and Colton, S. (2014) 'Ludus Ex Machina: Building A 3D Game Designer That Competes Alongside Humans.' See: http://www.gamesbyangelina.org/papers/iccc2014.pdf

6. See: http://bits.blogs.nytimes.com/2014/05/29/police-bicyclists-and-pedestrians-the-real-challenges-for-self-driving-cars/?_r=1

7. Wang, S. and Summers, R.M. (2012) Machine Learning and Radiology. 'Medical Image Analysis.' 16(5):933-51.

8. Beck, A.H., Sangoi, A.R., Leung, S., Marinelli, R.J., Nielsen, T.O., van de Vijver, M.J., West, R.B., van de Rijn, M. and Koller, D. (2011) Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. 'Science Translational Medicine.' November. 9;3(108).

9. See: http://www.maxlittle.net/publications/index.php

10. Bellazzi, R. and Zupan, B. (2008) Predictive data mining in clinical medicine: Current issues and guidelines. 'International Journal of Medical Informatics.' 77, 2008.

11. See: https://gdsdata.blog.gov.uk/2014/08/15/anomaly-detection-a-machine-learning-approach/

12. See: http://www.ons.gov.uk/ons/.../mwp1-ons-innovation-laboratories.pdf

13. See: https://osf.io/j5v8f/

14. See: http://www.oralytics.com/2013/03/type-i-and-type-ii-data-scientists.html

15. See: http://www.sheffield.ac.uk/is/pgt/courses/data_science

16. Government Office for Science (2012) 'Foresight. Regulatory scrutiny of algorithmic trading systems: an assessment of the feasibility and potential economic impact.' Economic Impact Assessment EIA16.

Nesta...
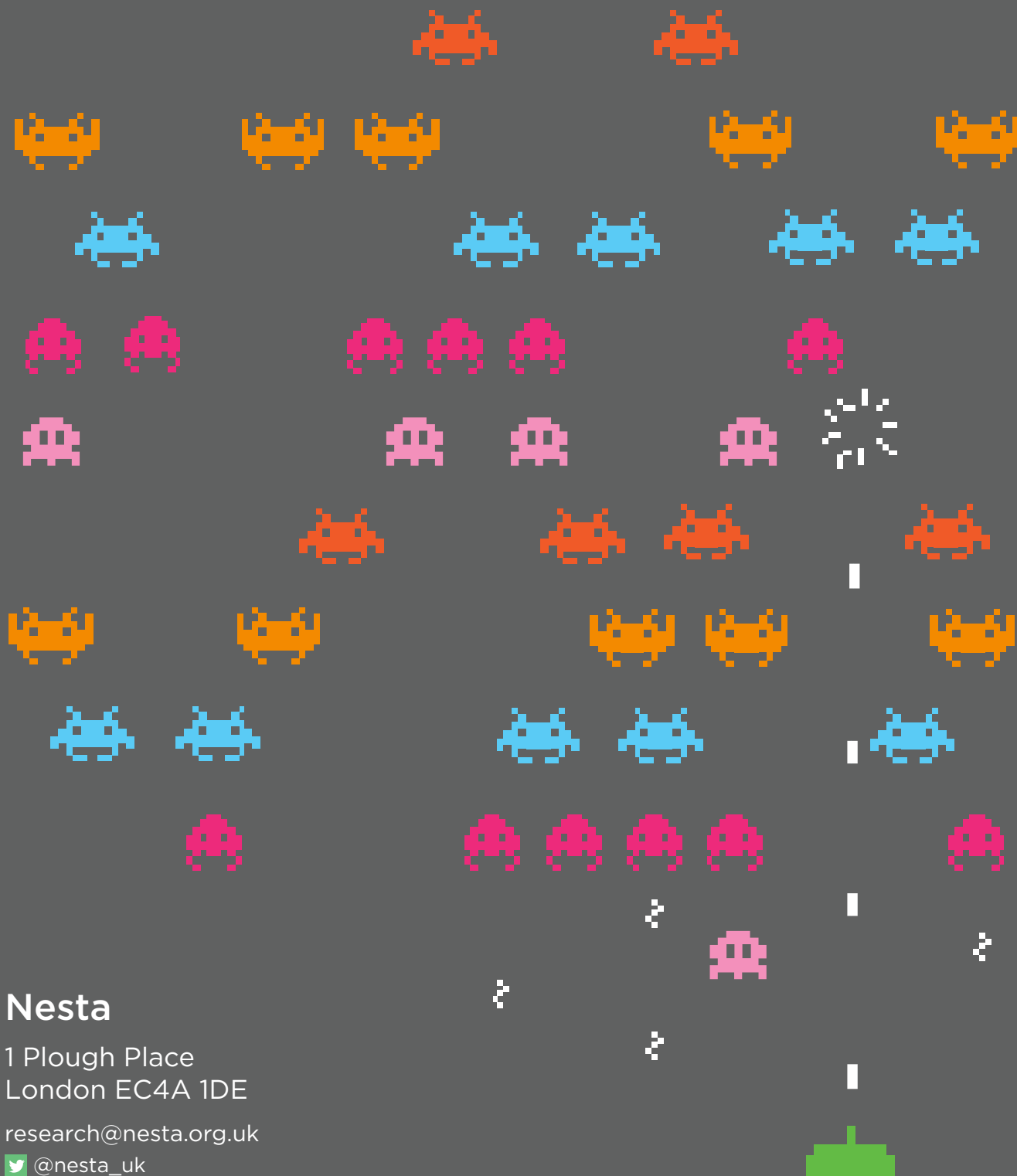
Nesta

1 Plough Place
London EC4A 1DE

research@nesta.org.uk

@nesta_uk

www.facebook.com/nesta.uk

www.nesta.org.uk

July 2015