Reviewing the Current State of Machine Learning for Artificial Intelligence with Regards to the use of Contextual Information.

Martin W Kinch
University of Greenwich
Chatham Ave
ME4 4TB
M.W.Kinch@greenwich.ac.uk

Wim J.C Melis
University of Greenwich
Chatham Ave
ME4 4TB
Wim.J.C.Melis@greenwich.ac.uk

Simeon Keates
University of Greenwich
Chatham Ave
ME4 4TB
S.Keates@greenwich.ac.uk

ABSTRACT
This paper will consider the current state of Machine Learning for Artificial Intelligence, more specifically for applications, such as: Speech Recognition, Game Playing and Image Processing. The artificial world tends to make limited use of context in comparison to what currently happens in human life, while it would benefit from improvements in this area. Additionally, the process of transferring knowledge between application domains is another important area where artificial system can improve. Using context and transferability would have several potential benefits, such as: better ability to function in multiple problem domains, improved understanding of human interaction and stronger grasping of current and potential future situations. While these items are all quite usual to us humans, it is particularly challenging to integrate them into artificial systems, as will be shown within this review. The limitations of our current systems with regards to these topics and the achievable improvements, if these items would be addressed, will also be covered. It is expected that by utilising transferability and/or context, many algorithms in the artificial intelligence field will be able to expand their functionality considerably and should provide for more general purpose learning algorithms.

Keywords
Machine Learning, Artificial Intelligence, Transfer Learning, Contextual Information Processing.

1. INTRODUCTION
In this paper the current state of Machine Learning (ML) and Artificial Intelligence (AI) will be reviewed, with a specific focus towards the concepts of Contextual Information Processing (CIP) and Transfer Learning (TL). This paper will also aim to provide a motivation as to why there is a need for more work in the areas of TL and CIP.

The paper starts with a more general review, followed by a section that covers current literature with a focus on CIP and TL. Further on an opinion is provided about possible future implementations on how to improve current artificially intelligent systems, after which the paper will conclude with a final discussion.

2. CURRENT RESEARCH
Research in ML and AI fields has become increasingly more popular. Much of the work in these fields has been centered around Artificial Neural Networks (ANN), a field closely linked with its neuroscientific counterpart that studies neural networks, which aim to create artificial equivalents based on the findings of neuroscience. Due to advances in the thinking around how these networks function, many have taken to adopting them in their work with AI. Much of the current work with regards to ANNs is often found following the footsteps of previous research. This is partly related to the fact that considering the huge variety across individual's human brain, one tends to gather as much data as possible and then rely on statistical significance to draw conclusions.

2.1 Artificial Neural Networks
ANNs use neural networking logic to provide powerful ML solutions to problems. One such problem can be seen in the case of a recent spate of development around game playing systems. Although not relevant to all types of learning, game playing applications still offer a way to prove learning concepts that can then be expanded over time towards other learning situations.

The most notable game playing ANN would be AlphaGo [16], a system developed to take part in the game of GO (an ancient Chinese board game more complex then chess). AlphaGo proved itself highly effective by defeating several high profile players in the game. AlphaGo was designed as a neural network, which applies tree search as a method of rooting out potentially advantageous moves. Much of the underlying success with this approach is the consideration to remove a vast number of highly unlikely moves, long before processing occurs, with the aim to limit the amount of processing needed for each move. The system is trained on data mainly comprising of beneficial moves, followed by a reinforcement process that adapts the neural network to promote positive play styles.

There are of course limitations to the way that AlphaGo functions, since there is still a high level of processing needed before a successful move can be chosen, which raises a debate about how "intelligent" the overall implementation is, although there is much to be said about the savings in processing over competing algorithms. Yet there is still a long way to go before this could be considered a fully "intelligent" system.

Other game playing AI's have also moved into the scene with the following examples coming from the same research institution. For example [9] shows a system that uses image processing and reinforcement learning applied to a neural network to play classic Atari games. The system works by using the pixel data from the output of the game, and more specifically taking note of the score. The inputs to the game will then be adjusted in relation to the output on the screen, and actions that promote a score increase will instigate a positive reward to the system. Currently, the system has been compared against human players and has shown results better than humans in certain specific cases. The system has meanwhile been expanded to play over forty-nine different Atari games [10].

Evidently the above example has shown that some of the work in the field of ANNs makes use of reinforcement learning. In most of these cases the neural network is paired with a Reinforcement Learning (RL) algorithm to provide feedback during learning.

The following examples will deviate from the game playing side of ANNs towards a broader overview. A more general ANN can be found in [7], which as a paper considers neural networks from an optimisation point of view. While bringing some good ideas of "whitening" the neural network to provide enhanced training, there was a distinct lack of encouraging results, which were rather modest, but showed good room for further improvement by the authors. The key consideration with this paper is that like many other in this area, the focus lies on pure mathematical tweaking to achieve better results. Which while beneficial, bodes the question why one does not attempt to set back and challenge the grounds that others have founded, especially considering that these optimisations only provide for modest improvements.

The above clearly shows that while promising, ANNs have a long way to go before they reach human levels of ability, at least outside the scope of very narrowly defined situations. However, there may be a point where ANNs begin to slow down in their advancement, mainly due to the limitations of exploring only a limited subsection of the human brain and learning process.

Nevertheless, for those further interested in the fields of ANNs would do well to look into, On Intelligence [8], a book which comprises a considerable amount of theory and forward thinking with the subject that belies a lot of the rational around ANNs and their biological counterparts. Those interested in more general knowledge about the subject of ANNs would also do well to look to [1], an article that describes some of the features of popular algorithm types and offers a wide range of reference material to support further reading.

2.2 Reinforcement Learning

Reinforcement Learning is something that was covered briefly in the previous section on ANNs. RL algorithms have a variety of uses and are not necessarily tied to their ANN counterparts, virtually any algorithm that must make decisions, particularly about its own learning, would benefit from this type of learning.

Bringing the discussion back to [9] and [10], it is worth noting that both iterations of their presented algorithms used RL to improve the performance of their neural network. It is important to note however, that the RL algorithms generating the improvement to the neural network are not completely restricted to this application. Although neural networks generally require some type of reinforcement to function effectively, much of the implementation may have been able to function purely on the reinforcement of a feature matching algorithm. Without the need for such an implementation, a short cut could well reduce performance cost as well as simplify the algorithm's design.

With reinforcement learning, there is considerable use for it to function independently and alongside existing methods, and it may be worth considering that at times the impact of an ANN may in fact be more of a complication than an assistance.

2.3 Evolutionary Learning

Another field in ML is Evolutionary Machine Learning (EML), which closely links to Darwinian style evolution theory. This presents itself often in that manner, as presented in [3], which details work using digital evolution to solve communication issues. The presented principle is to provide a situation where successful and efficient delivery from the entry to the exit node of the communication pathway are rewarded. The constructed pathway responds to changes in the exit node location. The main strength of this work lies in the limited memory capacity of the delivery agents. However, the main drawback of this approach is when a new generation is "born" and some of the old generation "die", message sections can get lost. While this normally results a small loss, it can create an impact in the application's overall viability.

A similar example can be seen in [4], where part of the processing of images was performed by a genetic EML algorithm. The system was designed for detecting breast cancer from mammographic images. From the results, it seems that the genetic variant fluctuates between being better or worse than other competing algorithms. However, it should be noted that accuracy was improved by imparting some transfer learning to the genetic EML. It would perhaps be interesting to see how other types of EML would perform like in similar test conditions.

The area of EML clearly has some benefits but the consideration here is how to prevent data loss via multiple generations of the EML agents. Mostly this is fixed with transfer learning, but it may be better handled during the creation of new generations and a possibly closer linking of the two may be beneficial.

2.4 Speech Recognition

The field of Speech Recognition (SR) has been a recent battleground with new products hitting shelves from multiple different retailers. However, many problems associated with SR tools are still consistently present in new algorithms and products. An example that showcases the difficulties as well as the advances is [6]. While their algorithm is functional, the used benchmarking algorithm showed an 8% word error rate, while their system attained 14.1% in the same Google based test. However, given that the system provided no need for a dictionary and attempted direct translation from utterance to letters, much can be said for the system. A similar system to the above [2], also received accuracy rates within the bounds of the 8% to 14.1% presented in [6], possibly showing a pattern regarding accuracy of speech recognisers.

Additionally, there is the issue posed that an error rate of even 8% is quite significant as overall this means that about one out of ten words will be incorrect, which is more than enough to get most people frustrated when using this system.

2.5 Image Processing

Image processing has also seen considerable use of ANNs, indicating a clear trend in AI research. For example [15] details the use of an ANN based system for feature recovery from satellite imagery. ANN is used to provide estimated parameter values for the feature extraction process. Overall, the proposed system functioned well when compared to a competitor, but the results are somewhat marginalised by the limited number of compared systems. However, the use of ANNs to power feature extraction is a nice addition to the system's design.

Another image processing application [5] looks at the imagery in medical diagnosis of retinas. The developed system is designed to notice anomalies in the imagery of the human eye and decide if these anomalies are indicative of a possible disease. The system is designed to locate Regions of Interest (ROI) in the image and using a mix of ANN and clustering to provide the basis for a voting scheme that considers the likelihood of a possible disease in the inspected eye. It is worth noting that this system has very high levels of accuracy, and as such, prompts a suggestion to use this technique in other areas of medical imaging, to see how well the system will perform in these environments.

3. CONTEXT AND TRANSFERABLITY

Context and transferability have been considered in research before, but research into context is partly limited by how it is defined. On the other hand, transferability has seen considerably more interest than context, thanks to e.g. its application in transfer learning. Yet, most of the AI and ML algorithms today generally do not exhibit signs of transferability and as such could be missing out from its benefits considerably.

3.1 Context

Context is something of a problem when it comes to research, not because studying it is particularly difficult, but because it has many different definitions. For instance, the Oxford Dictionary, defines context as: "The circumstances that form the setting for an event, statement, or idea, and in terms of which it can be fully understood" [11], which is a quite broad statement. Previous research into context [14] shows how it was defined in the late 90's, namely as a collection of terms that related more to the physical location and state of the device. As it stands, these collection of terms have provided a considerable amount of benefits to recent smartphone technology. However, it begs the question if there is not more to context than this. Another example, [13] describes context again more strongly as location, or a locational assistance tool. Both papers imply that context is derived mainly with regards to the location or device state, and a quick look at your smartphone device, and potentially its code base would provide considerable support for this.

The question then becomes what would be useful to AI other than the location and device state. A question probably difficult to answer due to the amount of data and its density (the amount of information provided in the data) which becomes more burdensome as quantities increase. However, it is likely that if the data remains limited one should expect few problems when processing context. For the sake of argument, the best definition of context used within this paper is the overall situation and the parts of that situation that are relevant to improve the understanding of a particular event that takes place. For example, if a passerby asks for something from someone, you could well consider that the passerby has asked to borrow a phone or any number of other potential items. If the same situation is analysed with the passerby wearing a balaclava, there is the potential for a completely different understanding of the situation.

Understanding of the situation clearly is important to define what an event generally is, especially when there is a lack of knowledge to fully define that situation. To better define an event's context the situation normally has to be considered in depth more than normally done by most current ML and AI systems. Currently, they are mainly intent to drill the data with statistical analysis, to provide albeit reasonable results. To limit the dependency on data and number crunching, it may be advantageous to try to define as much as possible about the event's context in a manner as simple as possible. This is where situationally dependent systems such as those used by speech recognisers find their ground.

3.2 Transferability

Transferability or Transfer Learning is the process by which data about a particular set of learned skills is used to improve the training of a new application for the same algorithm. There is some work into this subject and most shows significant improvements, but for some reason there is only a limited amount of this technology in real applications.

One example of transfer learning can be found in [12], a system used to analyse lung CT images to identify lesions. The system used a variant of a deep learning algorithm to analyse the images and make decisions. The transfer part of the learning was internal to the same application and used to promote positive confirmations. While this technically falls into the bracket for transfer learning, one should consider that transference is here more of a decision making structure than an actual transfer outcome. Therefore, it would be useful to see if this form of transference could be used innately on an image set from a different disease group to hereby confirm the effectiveness of transferring "knowledge".

Another example can be found in [17] which as a system attempts to use transfer learning to improve feature analysis by using previously learned data for its new learning target. However, rather unfortunately the explanation of the experiments is brief which raises some questions about the proper functioning of the system. On the other hand, the actual description of the transfer learning and its function are very clear.

Considering the above examples one can notice that while work is being established in this area it stills needs much more focus. Much of the difficulty with this field lies in the identification on useful data to transfer, and this may remain a human process until there is a clear indication of what needs to be transferred. However, by referring to the way that humans learn and how they expand their own learning to new situations, there is a considerable amount of information that is beneficial to algorithms, as they learn. It may simply be a case of limiting what is transferred to a small amount of data, that provides marginal, but nonetheless beneficial knowledge transfer.

One should also aim to combine the values of transferability with work around context, so that situational awareness and its understanding can be transferred through existing algorithms, something that could well improve the speed and depth of ML systems.

4. CONCLUSION

This review has aimed to deliver a cross section of machine learning and artificial intelligence technologies, and provide a commentary on how these may impact development into contextual information processing and further work in transfer learning.

Much of what has been presented works for specific applications, and tends to tie processing down to pure data in order to provide results. By applying at least some elements of transfer learning, many algorithms would improve when used in new environments based on their previous exposure.

By applying context to machine learning algorithms, there is the scope to provide situational data to the systems. However, this extra data could form a double-edged sword, as more data would require a larger memory and processing requirements. However, the extra data should then again aid the system's functionality. Something that would benefit many modern machine learning implementations.

There is certainly much to support for the idea to enhance the use of both context and transferability in new machine learning algorithms.

5. FURTHER WORK

Future work will look to use context and transferability in a variety of means, by applying them to AI and ML algorithms. Currently, the work being undertaken involves a study about context, regarding how the contextual information present in a situation will affect communication. It is expected that following on from the results from this study, it will become possible to apply the conclusions to the development of a new ML algorithm for speech recognition, where later on elements of transferability will be added.

6. REFERENCES

[1] Arel, I. R., "Deep machine learning-a new frontier in artificial intelligence research." *IEEE Computational Intelligence Magazine*, pp. 13-18, 2010.

[2] Bahdanau, D. C., "End-to-end attention-based large vocabulary speech recognition." *IEEE ICASSP*, pp. 4945-4949, 2016.

[3] Beckmann, B. M., "Evolution of cooperative information gathering in self-replicating digital organisms.", *IEEE First International Conference on Self-Adaptive and Self-Organizing Systems*, pp. 65-76, 2007.

[4] Burling-Claridge, F. I., "Evolutionary algorithms for classification of mammographie densities using local binary patterns and statistical features.", *IEEE Congress on Evolutionary Computation (CEC)*, pp. 3847-3854, 2016.

[5] Caramihale, T. P., "Detection of regions of interest in retinal images using artificial neural networks and K-means clustering.", *IEEE International Conference on Applied Electromagnetics and Communications (ICECOM)*, pp. 1-6, 2016.

[6] Chan, W., Jaitly, N., Le, Q., & Vinyals, O., "Listen, attend and spell: a neural network for large vocabulary conversational speech recognition.", *IEEE ICASSP*, pp. 4960-4964, 2016.

[7] Desjardins, G. S., "Natural neural networks.", *Advances in Neural Information Processing Systems*, pp. 2071-2079, 2015.

[8] Hawkins, J. & Blakeslee, S., "On Intelligence.", New York: St. Martin's Griffin, 2004.

[9] Mnih, V. K., "Playing atari with deep reinforcement learning.", *arXiv preprint arXiv*:1312.5602, 2013.

[10] Mnih, V. K., "Human-level control through deep reinforcement learning.", *Nature*, pp. 529-533, 2015.

[11] Oxford University Press, "Definition of Context." *Retrieved from Oxford Dictionaries: https://en.oxforddictionaries.com/definition/context,* Accessed on: 21-04-2017.

[12] Sawada, Y. A., "Transfer learning method using multi-prediction deep Boltzmann machines for a small scale dataset.", *IEEE 14th IAPR International Conference on Machine Vision Applications (MVA)*, pp. 110-113, 2015.

[13] Schilit, B. H., "Context-aware communication.", *IEEE Wireless Communications*, pp. 46-54, 2002.

[14] Schmidt, A. B., "There is more to context than location.", *Computers & Graphics*, pp. 893-901, 1999.

[15] Seetharaman, K. A., "Multi-spectral satellite color image retrieval based on GMRFM and Back-propagation algorithm with Artificial Neural Network.", *IEEE International Symposium In Electronics and Smart Devices (ISESD)*, pp. 278-282, 2016.

[16] Silver, D., Huang, A., Maddison, C. J., et al., "Mastering the game of Go with deep neural networks and tree search.", *Nature*, pp. 484-489, 2016.

[17] Zhang, J., "Deep transfer learning via restricted Boltzmann machine for document classification.", *IEEE 10th International Conference on Machine Learning and Applications and Workshops (ICMLA)*, pp. 323-326, 2011.