University of London
Computing and Information Systems/Creative Computing
CO3310 Artificial Intelligence
Coursework assignment 2 2020–2021

Question 1 - Machine Learning

a) **[12 marks]**

I. Supervised learning : A category within the Artificial Intelligence (AI) subfield of Machine Learning (ML). A way that a system can 'learn' to improve from past experience. Specifically, to learn to understand how objects can be categorized from 1 of 3 primary ML categories of feedback; 'examples'. A system receives a set of <data:label> pairs as input and from these examples 'learns' how to assign the correct label to new input data without a label. The system is supervised with the complete input set to 'learn' the concept, so that it can apply it. Supervised learning can be divided in two main categories, 'classifications' (described above) and 'regression' where the 'concept' is 'continuous' such as predicting weather from historical data.

Application example : A system is input data of bioacoustic audio clips and labels that identify the audio clips as the calls of a species of wildlife (a Howler monkey). The system 'learns' the 'concept' from the examples and can then identify that new 'unlablebeld' audio clips are the calls of Howler monkeys (or not) and apply the Howler monkey classification or label to the audio clip.

II. Unsupervised learning : The 2nd of the 3 primary categories of ML. Unlike supervised learning, <data:label> pairs are not presented before the data is input. Data is input, the system does not attempt to 'learn' a concept to apply labels to data but attempts to identify "useful properties" (G.D.Konidaris, 2013, p.49) about all the input data, such as how many distinct objects are in the data.

Application example : Bioacoustic audio data clips are fed into the system and it is identified that the audio clip's, as a set, contain 13 distinct species of monkey calls (Howler monkeys, Spider monkeys, Wooly monkeys …) that can be identified because they sound unique from one another, without a 'learning' input set.

III. Reinforcement learning : The 3rd of the 3 ML categories. The feedback received takes the form of a 'reward' for a sequence of actions the system performs in an environment with the agent's goal to receive the maximum value of rewards for sequences of actions over a span of time. The 'reward' is returned for the sequence's performance (hint's may be given) but unlike supervised learning, actions are not paired with 'correct' labels so that the 'sequence' is evaluated. Reinforcement learning does not learn a 'concept' but a 'control policy' to improve performance of actions in an environment.

Application example : ML reinforced learning applied to the game of chess can 'reward' a sequence of actions that improve the position of the system in the game that will make the system win or more likely to win. Each piece can be given a value and when a sequence of actions results in the opponent's pieces being taken that outweighs the pieces the opponent took the reward would be positive, with a reward for checkmate being very positive. Negative results are returned for less favourable positions and very negative for losing. The system can learn which sequences have the greatest 'reward' that correlates to winning the game or a winning position.

b) **[14 marks]**

I. Information : In this context, the term can be used to define a measure of what is contained in a "set of examples'' (G.D.Konidaris, 2013, p.52) used as the input to learn a 'concept' in supervised ML and used to evaluate the usefulness of attributes to split decision trees.
Expressible in fractions of 'positive' and 'negative' examples, such as there are 10 <data:label> pairs of bioacoustic audio clips that identify Howler monkeys and 10 that do not in a total set of 20.
The fractions can be calculated by formula to show the measure of the 'information' contained in that set.
For example, 10/ 20 positive examples, and 10/ 20 negative examples of 20 examples would = 1 (formula not shown, described below).
Expressing the measurement of 'Information' in a set in this way can be used to quantify what measurement of information in the set is needed to apply labels to future unlabeled input data and a measurement of what a subset contains.
An information set with all the same labels would have an 'information' measurement of 0. 0 'information' measure means a label decision can be made with certainty.
This concept is useful in ML to measure sets and subsets to make comparisons to find lower 'information measurement' sets to create sets that are closer to containing only the same labels, or no 'information'.

Entropy : Is the key quantity in information theory (Shannon and Weaver 1949). It is a measure of how uncertain a random variable is. Entropy is decreased when data can increase the certainty calculation of that variable. For example a variable that is certain has an entropy of zero. A fair coin has an uncertainty for each side landing of 50/ 50 which is described as having '1 bit of entropy', '0 and 1' 'heads and tails', a 4 sided dice would have 2 bits of entropy (Norvig, 2010, p.703). Entropy is measured in base 2 or bits.
In ML it can be used as a measure of how uncertain a system is about sets of data and variables and can be used to measure the potential increase in certainty of different groupings of the data.
In ML classification when a system can measure a variable or a set of data having entropy 0 the system can label with certainty.

Information gain : Defines the potential decrease of entropy or the increase in certainty for sets of data. In supervised ML classification decision trees 'information gain' is used to evaluate how 'useful' a variable (or 'attribute') from the input dataset, to 'learn' the 'concept', can be for a split in decision tree to calculate classifications or apply labels.

Information gain is a heuristic.
Attributes that split the tree to sets with low 'information' mean the sets are close to all containing the same label. As the 'information' measurement of a set with all the same labels = 0, attribute 'test nodes' in the tree that will eventually split the tree to subsets with 'information' 0 will result in a classification or label decision being made with certainty or no entropy.
The 'information gain' of splitting at each 'attribute' can be calculated, (with the below formula) compared, and the attribute with the greatest 'information gain' selected for the node. This results in an efficient arrangement of the tree and subsets with the least 'information' measure that are closer to certainty for a decision and labeling.

II. $I(A) = -f_{A+} \log_2 f_{A+} - f_{A-} \log_2 f_{A-}$

As described in the definition of 'Information' above. This formula shows how the 'information' measure can be expressed or calculated.

fA+	=	the fraction of 'positive' <data:label> pair examples in a set.
fA-	=	the fraction of 'negative' examples.
Log2(x)		is the inverse of exponential (x^2). It will find the exponent of 2 to equal x. E.g. log2(16) = 4, as 2^4 = 16.
-		negation

In a set (with 'positive' and 'negative' examples) the formula takes the probability of 'positive' examples and multiplies by the log2( ) of the same probability. Base 2 is used as 'entropy', described earlier is measured in base 2/ bits. The product for each 'case' ('positive' and 'negative') are then 'negated' as log2( ) will give a negative exponent for fractions. E.g. log2(½) = -1 and 'negated' = 1. Giving the 'information' of the set from it's containing 'positive' and 'negative' probabilities in a base 2, bit measurement used in entropy.

With the example of the coin from the definition of 'entropy', the 'Information' or uncertainty contained in the set is 1, a 'bit' of 'information', as a coin has a 50% probability of it's outcomes. It's input data set contains this 'information'.

I(coinSet) = - ½ log2(½) - ½ log2(½) = 1

In comparison to calculate the 'information' of data pairs all with the same label :

I(A) = - 1 log2(1) = 0 as there is no uncertainty and no 'information' contained in the set.

$$Gain(B) = I(A) - \sum_{i=1}^{v} f_{B_i} I(B_i)$$

This formula is used to evaluate the 'information gain' potential of an attribute that can be used to split a decision tree, as described earlier.

Gain(B)	=	Information gain of attribute B.
v	=	the potential values of the attribute B.
Σ	=	Sigma notation for the sum total of a series of calculations. Starting from 'i' to 'v'. If Σ n (i=1 and v=3) = 1+2+3 = 6.
i	=	the first potential value of attribute B (in the fair sided coin toss 'Heads' with v as the only other possibility 'tails').
Bi	=	the set when attribute B is of value i.
A	=	The complete input set of examples.
fBi	=	The fraction of the examples in A (that are also in the Bi set). The probability.

The 'information' measurement calculated with the above formula for Bi, I(Bi), is multiplied by fBi (the probability of Bi from the examples in both Bi and A), and the summation of this for each potential value 'i' of attribute B is subtracted from the constant 'information' measurement of the whole set of examples 'A'. The summation of multiplying the I(Bi) by the probability of each Bi gives the average to be subtracted from the I(A) = Information Gain (Bi) for the whole attribute.

If 'information gain' is calculated for all attributes the attribute with the greatest 'information gain' can be selected to split a decision tree into efficient, low 'information' subsets or to reach a 'certain', no entropy set (all the same labels, no 'information') and closer towards a decision for labeling, classification.

'Information gain' expresses the resulting 'information' contained in subsets of splitting the tree in this way, the attribute with the greatest 'information gain' will be chosen to split the tree, remember we need to reach a 0 'information' set for a decision.

III.

The information of drawing a single card from a 52 card deck.
The probability of picking each card is 1 in 52.
If the input <data:label> set contains 52 pairs and 1 is labeled 'correct' and the other 51 are labeled 'incorrect' to represent a card being selected from the deck.
1/ 52 represents the 'positive' events.
51/ 52 represents the 'negative' events.

I(event) = - 1/52 log2(1/52) - 51/52 log2(51/52) = 0.1370994789

c) **[24 marks]**

The answer (on the following 3 pages) calculates the ‘information gain’ (with the above 2 formula) for each attribute (celebrity, genre, xmas) by the sigma summation of multiplying the fraction of examples in an attribute’s potential value set that are also in the set of total examples, by the ‘Information’ measure of that potential attribute value set calculated with the fractions of ‘positive’ and ‘negative’ labeled examples in that set, and finally subtracting from the constant ‘information’ measure of the total example set.

By selecting the attribute with the greatest ‘information gain’ (Genre) to be a split and the first test in the tree we group the subsets to have more efficient lower ‘information’ measures and more certainty until there is no entropy in a subset and a decision of the label can be made with certainty as we need ‘information’ to measure 0 (when all examples in that set are the same).

1

- Information gain for each feature.

- $I(A) = f_{A+} = \frac{7}{12}$ 'positive label examples' 'Best sellers'.

  $f_{A-} = \frac{5}{12}$ 'negative label examples'

  $= -\frac{7}{12}\log_2\left(\frac{7}{12}\right) - \frac{5}{12}\log_2\left(\frac{5}{12}\right)$

  $= 0.9798687567$

Attributes { Celebrity, Genre, Xmas }.

(1st) Attribute 'Celebrity', potential values { yes, no }.

- B_i / B'yes' $= \{1, 3, 4, 8, 9, 10\}$, B_i / B'no' $= \{2, 5, 6, 7, 11, 12\}$

  f_{B_i} / f_B'yes' $= \frac{6}{12}$, f_{B_i} / f_B'no' $= \frac{6}{12}$

- $I(B\text{'yes'}) = f_{A+} = \frac{4}{6}$

  $f_{A-} = \frac{2}{6}$

  $= -\frac{4}{6}\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\log_2\left(\frac{2}{6}\right)$

  $= 0.9182958341$

  $I(B\text{'yes'})\,\frac{6}{12} = 0.4591479170$

- $I(B\text{'no'}) = f_{A+} = \frac{3}{6}$

  $f_{A-} = \frac{3}{6}$

  $= -\frac{3}{6}\log_2\left(\frac{3}{6}\right) - \frac{3}{6}\log_2\left(\frac{3}{6}\right)$

  $= 1$

  $I(B\text{'no'})\,\frac{6}{12} = 0.5$

Σ

$$\begin{array}{r} 0.459147917 \\ +\ 0.5 \\ \hline 0.959147917 \end{array}$$

- $Gain(B) = 0.9798687567 - 0.959147917$

$Gain(\text{Celebrity}) = 0.0207208397$

2nd Attribute 'Genre' { 'fiction', 'non-fiction', 'How-to' } (2)

- [B_i] B'fiction' = { 1, 3, 5, 6, 9, 10, 12 }

  B'non-fiction' = { 2, 4, 8, }

  B'How-to' = { 7, 11 }

- [fB_i] fB'fiction' = $\frac{7}{12}$

  fB'non-fiction' = $\frac{3}{12}$

  fB'How-to' = $\frac{2}{12}$

- I(B'fiction') $fA_+ = \frac{5}{7}$

  $fA_- = \frac{2}{7}$

  $= -\frac{5}{7} \log_2\left(\frac{5}{7}\right) - \frac{2}{7} \log_2\left(\frac{2}{7}\right)$

I(B'fiction') fB'fiction' = 0.8631205686 · $\frac{7}{12}$ = 0.5034869983

- I(B'non-fiction') = $fA_+ = \frac{1}{3}$

  $fA_- = \frac{2}{3}$

  $= -\frac{1}{3} \log_2\left(\frac{1}{3}\right) - \frac{2}{3} \log_2\left(\frac{2}{3}\right)$

I(B'non-fiction') fB'non-f' = 0.9182958341 · $\frac{3}{12}$ = 0.2295739585

- I(B'HowTo') $fA_+ = \frac{1}{2}$

  $fA_- = \frac{1}{2}$

  $= -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2}\left(\log_2 \frac{1}{2}\right)$

I(B'HowTo') fB'HowTo' = 1 · $\frac{2}{12}$

Σ

0.5034869983
0.2295739585
+ = 0.1666̇

0.8997276335

- Gain(Genre) = I(A) − 0.8997276235

0.9798627567 − 0.8997276335 =

Gain(Genre) = 0.0801411332

3rd Attribute 'Xmas' (3)

- [Bi] B'yes' = {3, 5, 6, 9}

  B'no' = {1, 2, 4, 7, 8, 10, 11, 12}

- [fBi] $fB'yes' = \frac{4}{12}$

  $fB'no' = \frac{8}{12}$

- $I(B'yes')$ $fA+ = \frac{3}{4}$

  $fA- = \frac{1}{4}$

  $= -\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right)$

$I(B'yes')fB'yes' = 0.8112781245 \cdot \frac{4}{12} = 0.2704260415$

$I(B'no')$ $fA+ = \frac{4}{8}$

$fA- = \frac{4}{8}$

$= -\frac{4}{8}\log_2\left(\frac{4}{8}\right) - \frac{4}{8}\log_2\left(\frac{4}{8}\right)$

- $I(B'no')fB'no' = 1 \cdot \frac{8}{12} = 0.666^{\circ}$

- Gain (xmas) = I(A) − 0.9370927075

- 0.9798687567 − 0.9370927075

Σ

$$
\begin{array}{r}
0.2704260415 \\
+\ 0.666^{\circ} \\
\hline
0.9370927075 \\
\hline\hline
\end{array}
$$

GAIN (xmas) = 0.0427760492

- Gain (celebrity) = 0.02
- Gain (Genre) = 0.08
- Gain (xmas) = 0.04

largest gain, first node "Genre"

Question 2 - The future of AI ?

a) **[15 marks]**

I. **Cognitive model :** An AI's understanding from the input of a problem or an environment that can be formed or updated by how that input information is processed or perceived by the AI. The AI can then base intelligence decisions and actions from the understanding and construction of the cognitive model. Considered "Central to how an organism [or AI, in this case] views the world (Gallistel, 1990) such as humans. In the context of the paper they are presented as a central component for the solution of a more robust AI and closer to human intelligence.

II. **Symbol manipulation :** A symbol can be any encoding to represent information, such as binary arranged to represent, for example, a base 10 number. And manipulation, such as addition and comparison of those encodings are used to form the basis of the principles used in programming languages or a system to represent at a higher level of abstraction to define variables, instances, bindings and operations. In the context of the paper symbol-manipulation is presented as an improvement to be incorporated into the architecture of neural networks and deep machine learning to represent properties and changes over time and rationalization that typically are not.

III. **Hybrid architecture :** Hybrid architecture, a combination of approaches and concepts. In the context of the paper, by combining symbol-manipulation techniques with the current trend in AI of deep learning large scale knowledge bases, a system can improve its performance in a similar way that humans learn, for example, by a combination (or hybrid) of both causal and abstract understanding. The hybrid architecture of prior knowledge and more flexible reasoning is expected to go beyond current AI for it to perform more robustly such as identifying outlier cases correctly.

IV. **Robust A.I :** Robust AI can perform in a variety, or change, of environments and can be applied to a variety of tasks or tasks that change. The opposite of a non-transferable pointalistic task specific AI that will not perform if the task or problem changes and will fail in outlier cases. In the context of the paper, the hybrid approach described above is conjectured to make a more robust AI that is applicable to changing tasks, more open environments and will be an improvement on the current robustness of AI with a more general intelligence and particularly with performance to outlying tasks.

b) **[15 marks]**

I. **Machine Learning models :**

The limitations of current models are summarised as "data hungry, shallow, brittle, and limited in their ability to generalize" (Marcus, p.4) and that current models are limited by the absence of cognitive models and particular kinds of reasoning and flexibility as they attempt to calculate closed-ended solutions to open-ended environments and tasks, and are easily tripped up by slight changes in tasks or parameters.

Examples:

- An unnamed "pointillistic" visual system that can recognise school buses but can not identify a school bus if it has crashed and it is on it’s side.

- The BiDAF text analysis model can correctly answer questions relating to a text, but is confused by the addition of non related text being inserted in the input data text, such as from a text about the most recent super bowl with extra text from another superbowl and the question “who was the quarterback ?” being answered with “the quarterback of the past superbowl” not the current quarterback from the focus of the text.

- The GPT-2 system that can return an apparent intelligent response to text fragments will also return sentences that are unbelievable or incorrect, not representational of robust intelligence. For example, "There are six frogs on a log. two leave, but three join” GPT-2 returns the answer “The number of frogs on the log is now seventeen".
  The system's intelligence limitation is considered to be 'approximate' knowledge, like winning a game of scrabble in french without knowing french only memorisation of the token value of the tiles.

- Current models knowledge bases that ‘memorize’ rather than reason need data to match the complexity of the environment. Data needed for very complex environments, such as the world necessitates exponential levels of data and highlights the limitation, amongst others, of the wide breadth of the approach of large scaling hardcoded data driven knowledge base systems (Marcus, 2020).

- DaAvila, Garcez, Lamb and Gabbay (2009) identified the limitations of knowledge representation and inference in many unnamed models in an analysis study and concluded that their limitations could be surpassed with a hybrid architecture approach of mixing symbol-manipulation with neural networks into an improved hybrid ‘neurosymbolic’ model.

- Large investment of resources, personnel, money and computation in current machine learning models has not returned a satisfactory measure payoff of robust intelligence at a desired level. A significant limitation.

- Hardware utilization is expected and required to increase with "larger clusters of GPUs and TPU's" (Marcus, p.6) with their own limitations of organisation and manageability when using increased hardware in architectures to improve current machine learning models.

II. Formal, symbolic knowledge bases such as CYC :

Symbolic knowledge bases and CYC have a sophisticated semantic reasoning that can infer from facts to logical connections to return responses with an 'understanding' that is replicant of a learnt knowledge over something that is written down, like riding a bike over reading a bike manual. They share some of the same limitations identified for other models of being 'data hungry'. CYC's reasoning is dependent on fixed language and limited by its perceptual inputs.

Examples:

- CYC, given a plot summary, time points and some truthful statements about the play 'Romeo and Juliet' can produce a rich intelligent response when queried about the play but the measure of 'intelligence' of its response is reliant and heavily dependant on the prior structuring of the input data to be optimally compatible, (i.e in a 'formal logic structure') and it's performance greatly affected without.

- CYC and formal, symbolic knowledge bases have the same limitation that the investment of time (decades), effort, money, research, personnel and computational investment has only resulted in a modest measure of robust intelligence. This classical approach to representation of abstract knowledge has been "brutally hard work" (Marcus, p.24). CYC is the largest project of effort to attempt this classical approach in the history of AI and is largely considered as a failed project.

- CYC Requires "manual hard-wiring of each fact" (Marcus, p.27) and this approach is considered unrealistic in its breadth with the same limitation mentioned earlier to meet exponential open-world and less fixed problems. They have limited reasoning.

- CYC has "no perceptual component and lacks [an] adequate natural language front end" (Marcus, p.20) limiting its commercial application without the setting of the right structured data inputs and relying heavily on workarounds and considered very convoluted in its decades long structure and not easy to understand or to work with.

- Formal symbolic knowledge systems are limited to information that is documented or encoded, much information and knowledge, such as common sense, understanding that accumulates over time and emotional intelligence representative of human cognition may not be written down at all. They have a limited utilization of available knowledge and information.

To summarize, the potential 'reasoning' of symbolic knowledge bases, such as CYC are greatly limited by the same 'data hungry' as well as the large scale manual hand coding approach that limits their ability to infer and rationalise beyond convenient structured input data and can be easily tripped up. The paper considers CYCs measure of 'intelligence' impressive but an approach that can only go so far when considering more general intelligence.

c)

I. **[20 marks]**

Marcus conjectures that robust cognitive models are possible with hybrid architecture and the collective combination of the advantages of 'classical large scale machine deep learning knowledge bases' (when on their own have limited, scaling, flexibility and reasoning ability) together with the advantages of the perception, abstract representation and reasoning prowess of symbol-manipulation principles that can represent and reason with abstract and causal knowledge and information that has been historically underutilized.

Marcus states that robust AI needs a rich causal understanding of the world that is not currently possible without a triumvirate.

Marcus outlines the proposed improved AI systems components :
- Hybrid neuro-symbolic architecture
- Partly-innate cognitive frameworks
- Large-scale knowledge bases
- Woven tools for abstract reasoning

To perform more sophisticated processing of observations that are supported by existing understanding to form the cognition model of a more robust AI that does not exist currently when these components are isolated.

Marcus discusses that classical ML models make limited use of information and knowledge and their ‘intelligence’ reliant on a colossal structure and an approach that is mostly utilizing knowledge that is inferred from hardwired "quantified relationships between variables” (Marcus, p.27) that will not produce robust general intelligence despite the resources pumped in to it.

The combination of the impressive, but limited, innovations of ‘large-scale knowledge bases’ such as CYC should not be rejected as failure, although they are concluded to lack the foundational requirement for robust intelligence and can be outperformed in certain tasks by other ML models such as Transformers such as GPT-2 but can be the source to provide the rich prior knowledge for a cognitive model that can.

The classical approach of prior knowledge techniques are identified and understood to have merits and a bottleneck limitation caused from the requirement of prepared structured input may greatly improve with supported restructuring from hybrid architecture and the incorporation of cognitive-manipulation principles to mitigate shortcomings.

To incorporate into the hybrid-architecture, the advanced developments, applications and understanding of the semantic, syntactic and abstract representations and operations used in programming languages (cognitive-manipulation principles) are presented as the reason some models show ‘intelligence’ without being identified or recognised. And are presented as the source to make use of a greater majority of knowledge (than traditionally utilized by deep learning) that can innovate more robust AI.

The combination approach and resulting cognitive model should be able to more gracefully understand exceptions, changes in environments and tasks and perform beyond the limitations of encoded logic systems.

The idea is that it has not been invested in or tried before to the same degree as has been invested in other models and the dominant deep learning approach to AI that Marcus does not expect will go far beyond current limitations.

The combination of the impressive results of data-driven learning techniques to be the 'rich prior knowledge' from the large-scale knowledge bases of multi layered neural networks and the 'sophisticated reasoning techniques' of the ability to represent and update higher abstractions and to reason with casual knowledge of 'cognitive-manipulation principles' will be able to better handle and understand the different forms and structures of input data, utilize more knowledge and information and with less limited reasoning.

And with a hybrid architecture and interwoven framework of components are conjectured to be the solution to a more adequate and automated resilient and robust AI that still makes use of larger and larger data sets and more and more 'compute'.

It seems like a really good idea. Human Intelligence is complex and not fully understood, sticking to a mainstream has never been very innovative. With the stated return of investment of the current state I certainly feel the potential outweighs the risk. It is very well considered.

II.

"If the problem of robust intelligence had already been solved there would be no need for this essay" (Marcus, p.52).

Marcus summaries that a robust Intelligence, and the future of AI, will be innovated by rich cognitive models that will improve the current state of AI's limitations in the :

- The application to wider problems
- Have a more systematic approach
- Be more reliable
- Knowledge 'synthesizable' from more sources.
- Have more flexible reasoning to perform in more complex environments.
- Be able to transfer 'learnt' concepts from one 'context' to another, the same as humans.

Even with increased computation, time and research investment the long duration of the classical approaches to AI without a cognitive approach, has only yielded 'intelligence' that is not robust enough to be applied more generally, to changing tasks, changing environments and the world or space and that are easily tripped by outlier cases and non-pointalistic problems.

"Starting with near blank slates and training them on massive data sets-simply hasn't panned out so far." (Marcus, p.34).

The classical approach and their research investments have not reflected a consideration of the higher-level 'thinking' that is present and involved in replicating or surpassing human cognition and intelligence. More input data is not expected by Marcus to improve Artificial intelligence and that cognitive models can.

Much of knowledge and ‘intelligence’ is derived, or learned over time. The approach of CYC to logically harcode facts and infer has a major disadvantage that 'common sense' intelligence may very well not be documented anywhere.

Cognitive models made from the combination of hybrid architecture between large scale knowledge bases and symbol-manipulation mechanisms for effective reasoning may provide the richness of a cognitive model that can utilize from more knowledge sources and contexts or be closer towards than further slight refinements and slightly more subtle correlations of deep learning neural networks.

Cognitive model intelligence can be more robust by using knowledge that accumulates, changes and updates over time as well as from large data systems and does not have to be encoded or structured fact first.

Much information and knowledge is represented and sourced from different formats that AI without a cognitive model Marcus has shown to have the limitation of not being able to make use of with a robust level of accuracy.

Richer cognitive models move beyond these limitations by being able to identify when facts change or should be interpreted differently. Although under tried this approach is conjectured to be more capable of representing and reasoning with abstract knowledge and casual knowledge particularly with the incorporation of the principles of symbol-manipulation.

As the concept is a relatively unknown or not spoken about and an under pursued area of equal exploration it is presented carefully as a likely concept for innovation and advancement in the field.

Marcus carefully constructs the hybrid solution as a natural step as it he recognises that :

“Systematic ways to induce, represent, and manipulate large databases of structured, abstract knowledge ... are a prerequisite to robust intelligence (Marcus, p.27)

together with :

"Symbol-manipulation, particularly the machinery of operations over variables, offers a natural though incomplete solution to the challenge of extrapolating beyond a training regime” (Marcus, p.23).

Marcus states that without the above hybrid approach and symbol-manipulation too much useful abstract and casual knowledge that can only be manipulated with the principles used in programming languages when supported by mainstream deep learning AI's impressive results are necessary to implement for AI to have the capacity for robust intelligence.

I agree it is likely that the resulting cognitive model will utilize more knowledge, have less limited reasoning and have more 'casual' and 'abstract' 'understanding' that are the causes of the limitations of the current state of AI.

The current return on investment seems to indicate innovation in a different diffraction is reasonable and the conjecture is well thought out but lacks more technical detail that could illuminate the current state more clearly.

Thanks for reading, Mark Start.

References

Shannon and Warren Weaver (1949), The Mathematical Theory of Communication, University of Illinois Press, ISBN 978-0-252-72548-7.

D'Avila Garcez, A. S., Lamb, L. C., & Gabbay, D. M. (2009). Neural-symbolic cognitive reasoning. Springer Science & Business Media.

Gallistel, C.R. (1990). The organization of leanfing. The MIT Press.

G.D Konidaris (2013). Artificial Intelligence. CO3310, University of London Study Guide.

S.Russell and P.Norvig (2010) Artificial Intelligence: A Modern Approach. London: Pearson.

G.Marcus (2020). The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence (essential reading)