Evolved Artificial Intelligence for Stochastic Clustering Unsupervised Learning

(Patent Pending)

Randal Allen

Lone Star Analysis

Orlando, FL

rallen@lone-star.com

ABSTRACT

Using mainstream Artificial Intelligence/Machine Learning (AI/ML) algorithms, entire classes of practical use cases are doomed to be computationally intractable or require many multiple universe-lifetimes of data collection. I/ITSEC paper 19149 introduced an innovative architectural approach to AI where neural sigmoid functions are replaced by mathematical models of arbitrary complexity, thus collapsing net sizes and depth, and ultimately reducing computational and data requirements. Without the constraints imposed by neural assumptions, mathematical models may be nonlinear and/or discontinuous and may be guided by human knowledge of the system. This patent pending/novel approach, along with advanced optimization methods presented in I/ITSEC paper 19109, forms the basis a new family of Evolved AI solutions.

During a January 2020 National Defense interview, NDIA's Senior Fellow for AI expressed how algorithms and framework have evolved beyond supervised learning into unsupervised and reinforcement learning. Having presented the algorithms and laid the framework for Evolved AI, the focus of this paper shifts to applications of this emerging technology to stochastic clustering.

The paper first describes how the laptop-executable approach combines elements of both hard and soft clustering without the need for cleaning/scaling data, nor the need for training data. Unlike k-Means and k-Medoids, the cluster number (k) is not needed a priori, as with Hierarchical Clustering. By leveraging Fuzzy c-Means and Gaussian Mixture Models, data points may belong to more than one cluster having different sizes and correlations. Overall, cluster number is adaptively determined from the distribution of resultant cluster permutations. The paper then presents and discusses an example in the context of pulse spectrum analysis where preliminary work in applying multi-dimensional stochastic clustering has proven successful. Upon summarizing results, the paper concludes by recommending applications of this emerging technology to Training and Education, for example measuring pilot training exercise data and clustering results in terms of ideal execution.

ABOUT THE AUTHOR

Randal Allen is the Chief Scientist of Lone Star Analysis. He is responsible for applied research and technology development across a wide range of M&S disciplines and manages intellectual property. He was recently issued a predictive/prescriptive analytics patent and has multiple AI patents pending. He maintains a CMSP with NTSA. He has published and presented over 15 technical papers and is co-author of the textbook, "Simulation of Dynamic Systems with MATLAB and Simulink." He is an Associate Fellow of AIAA and is a certified space professional. He holds a Ph.D. in Mechanical Engineering (University of Central Florida), an Engineer's Degree in Aeronautical and Astronautical Engineering (Stanford University), an M.S. in Applied Mathematics and a B.S. in Engineering Physics (University of Illinois, Urbana-Champaign). He serves as an Adjunct Professor/Faculty Advisor in the MAE department at UCF where he has taught over 20 aerospace-related courses.

Evolved Artificial Intelligence for Stochastic Clustering Unsupervised Learning

(Patent Pending)

Randal Allen

Lone Star Analysis

Orlando, FL

rallen@lone-star.com

MOTIVATION

Motivation for this paper is driven by the need for an *augmented* approach to deep learning neural nets in response to their many criticisms. Recapitulation of the criticisms are summarized below:

- They're <u>shallow</u>: Little innate knowledge is required on behalf of the designer where architectural selection becomes an exercise in numerical investigation. The depth of deep learning is composed of an arbitrary number of nodes and hidden layers, this includes the large-scale repetition of similar, simple nodes. In the end, the designer is often unable to explain why one architecture is used over another.
- They're <u>greedy</u> and <u>brittle</u>: Big data is required to train and test deep learning neural nets which often break when presented with new data. The need for large datasets is driven by the memorization technique behind deep learning. After the neural net is trained to memorize a dataset, it often breaks when attempting to apply it to a generalized set of data outside the training space. Furthermore, the order in which the training data is presented impacts performance.
- They're <u>opaque</u>: There is a lack of transparency due to the difficulty in understanding the meaning between the inputs, interconnections, and outputs. This opaqueness is what labels them as *black boxes*. Deep learning neural nets essentially fit data to a curve, but nobody knows the parameters associated with the equation. The U.S. Military wants its autonomous machines to explain themselves (Knight, 2017) as well as wanting to understand any assumptions, limitations, and associated errors.

In an MIT Technology Review article (Hao 2020), two AI experts (Dr. Gary Marcus and Dr. Danny Lange) weigh-in on the future of AI. Dr. Marcus argues in alignment with the criticisms above and claims deep learning shortcomings are inherent to its technique. Dr. Lange, while taking the opposite side of the debate, still concedes (i) training data needs to be constantly tailored for a specific application and (ii) the AI solution needs to be fed enough data and have the right architecture. Quantified, he claims 10 to 20 million images are far from enough data. Thus, Dr. Lange also acknowledges shallow, greedy, and brittle criticisms.

All this criticism culminates into trust issues with deep learning neural net architectures. Just changing our attitude about AI is not going to resolve the issue. Instead, action needs to be taken to resolve these criticisms.

ROBUST / EVOLVED AI

(Marcus 2020) classifies deep learning as *narrow intelligence* in the context of closed-ended domains, e.g., games, object classification, and speech recognition. Where there is success, the results are also spotty and pointillistic, with unreliability leading to untrustworthiness. For example, figure 1 (left) shows a sample of how an object in a noncanonical orientation and context fools many current object classification systems (Alcorn et al, 2018). The rolled-over school bus was classified as a snowplow. Figure 1 (right) shows a sample of how an adversary inserted material to fool a large-scale language model (Jia Liang, 2017). From the QuAD dataset, the BiDAF Ensemble model originally gets the answer correct but is fooled by the addition of an adversarial distracting sentence, "Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV."

(Marcus 2020) goes on to identify requirements necessary for moving toward what he calls Robust AI – a hybrid, knowledge-driven, cognitive model-based approach. As each of the requirements is examined in more detail, the author will reintroduce the postulated approach referred to as Evolved AI (Allen, 2019).

Article: Super Bowl 50
Paragraph: "*Peyton Manning became the first quarterback ever to lead two different teams to multiple Super Bowls. He is also the oldest quarterback ever to play in a Super Bowl at age 39. The past record was held by John Elway, who led the Broncos to victory in Super Bowl XXXIII at age 38 and is currently Denver's Executive Vice President of Football Operations and General Manager. Quarterback Jeff Dean had jersey number 37 in Champ Bowl XXXIV.*"
Question: "*What is the name of the quarterback who was 38 in Super Bowl XXXIII?*"
Original Prediction: John Elway
Prediction under adversary: Jeff Dean

Figure 1 - Failures of Image and Speech Recognition

Reassuringly, both AI experts (Dr. Marcus and Dr. Lange) agreed on the need for a hybrid/multi-model approach (Hao 2020). A hybrid approach incorporates symbolic algorithms into the model architecture. Algorithms like these are prevalent in everyday computer operating systems and applications we fully trust. Evolved AI architectures may be based on mathematical models, physics-based models, deep learning neural net models – the point here is to develop a hybrid architecture leveraging the benefits each model has to offer. Google Search is an example of a hybrid of symbolic manipulation with highly optimized deep learning – not quite robust AI, but nevertheless, a step in that direction. Evolved AI also allows for architectures of arbitrary complexity and diversity, which begins to mimic organic intelligence with specialized processes. Just as our brains have specialized centers for speaking, seeing, reading, and recognizing faces, it makes sense that an AI architecture which has specialized capabilities would be more powerful and reliable than one which was… …shallow.

Robust AI requires various forms of structured, abstract knowledge in a causal nature. Rather than starting the learning process from scratch, the learning systems should be prepopulated with knowledge, a priori. Evolved AI may call upon subject matter experts to provide what would be considered innate knowledge for the system. For example, think of the first day school where the professor gives a quiz to baseline student knowledge. If the students have a great deal of innate knowledge, it makes the new learning process easier. Furthermore, depending on the learning domain, this innate knowledge could be spatial, physical, psychological, temporal, and/or causal. At the very least, the knowledge should be based on common sense.

As mentioned previously, current deep learning approaches are trying to memorize what is presented to them, hence the need for large datasets. Was it more beneficial for the student to memorize the answers to a test or learn critical thinking to solve problems? Robust AI requires the ability to draw upon inferences through reasoning. Naturally, the ability to inference is based on probability and Evolved AI operates on a stochastic engine. Additionally, the nonlinear, nonconvex optimization engine (Allen, 2019) provides the ability to solve problems with discontinuities. As such, Boolean logic gates may be incorporated into hybrid architectures providing the basis for fundamental reasoning and decision-making. In fact, Evolved AI has been successfully applied to reinforcement learning and results will be presented in a forthcoming paper.

The seminal paper (Allen, 2019) presented a detailed application to supervised learning. In this paper, the flexibility of Evolved AI is extended by application to unsupervised learning. However, in the spirit of Robust AI, the underlying algorithm is a hybrid architecture comprising the benefits of various unsupervised learning approaches as well as statistical inference to determine centroidal location.

UNSUPERVISED LEARNING / STOCHASTIC CLUSTERING

During a January 2020 National Defense interview, NDIA's Senior Fellow for AI expressed how algorithms and framework have *evolved* beyond supervised learning into unsupervised and reinforcement learning. In direct response, this paper focuses on an unsupervised learning application within the Evolved AI context. As mentioned previously, a forthcoming paper will respond with a reinforcement learning application, thus demonstrating the flexibility and applicability of the Evolved AI to solve a wide range of machine learning problem types. Before describing unsupervised learning in the context of Evolved AI, machine learning methods are summarized.

In general, there are three types of machine learning methods, selection and implementation of which depends on how data is being manipulated – they are supervised, unsupervised, and reinforcement learning. In the case of supervised learning, a model is trained on known input and known output data to predict future outputs. There are two subsets to supervised learning: *regression* techniques for continuous response prediction and *classification* techniques for discrete response prediction. In the case of unsupervised learning, clustering techniques are used to identify patterns in the input data. There are two subsets of unsupervised learning: *hard clustering* where each data point belongs to only one cluster and *soft clustering* where each data point can belong to more than one cluster. Finally, with reinforcement learning, a model is trained on successive iterations of Markov decision-making, where rewards are accumulated based on the results of each decision. A machine learning practitioner will recognize there are many methods to solve these problems, each having their own set of implementation requirements.

Since the focus of this paper is on unsupervised learning, details of hard clustering and soft clustering algorithms are examined. Non-neural net hard clustering algorithms include k-Means, k-Medoids, and Hierarchical Clustering; soft clustering algorithms include Fuzzy c-Means and Gaussian Mixture Models, summarized in Table 1.

Table 1 - Unsupervised Learning Methods and Algorithms

Hard Clustering	Soft Clustering
K-Means	Fuzzy c-Means
K-Medoids	Gaussian Mixture Models
Hierarchical Clustering	

Hard Clustering

The *k-Means* algorithm partitions data into k mutually exclusive clusters. Whether or not a data point fits within a cluster is determined by its distance relative to the cluster's centroid (which doesn't have to coincide with a data point from the dataset). The *k-Medoids* algorithm behaves like the k-Means algorithm except the centroid must coincide with a data point from the dataset. These algorithms excel when the number of clusters is known and when a large dataset needs to be clustered quickly. Clearly the limitation is having to know the number of clusters a priori. *Hierarchical Clustering*, which excels when you don't know how many clusters there are, analyzes similarities between data point pairs and groups them into a binary hierarchical tree.

Soft Clustering

The *Fuzzy c-Means* algorithm flexibly allows data points to overlap with more than one cluster. Still, the number of clusters must be known ahead of time. The *Gaussian Mixture Model* algorithm also allows data points to overlap, but the data points are generated from multivariate normal (Gaussian) probability distributions.

Evolved AI (Stochastic Clustering)

Evolved AI's unsupervised learning algorithm is a hybrid architecture comprising the benefits of hard and soft clustering leveraging statistical inference to determine cluster number as well as centroid location. Unlike k-Means and k-Medoids, the cluster number (k) is not needed a priori – like Hierarchical Clustering which also does not depend on knowing the cluster number ahead of time. By leveraging Fuzzy c-Means and Gaussian Mixture Models, data points may belong to more than one cluster having different sizes and correlations. Overall, cluster number is adaptively determined from the distribution of resultant cluster permutations.

Figure 2 (left) shows a simplistic artificial neural network architecture used in deep learning where additional hidden layers have been added providing depth. In practice, deep learning networks may have tens of hidden layers or more. It is assumed the reader has a general understanding of how deep learning neural networks operate.

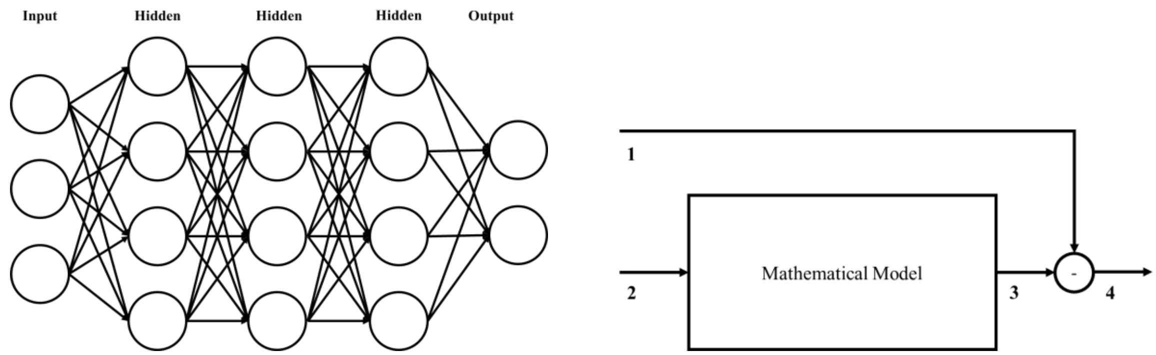

Figure 2 - Artificial Neural Network for Deep Learning (left), Evolved AI's Architecture (right)

Figure 2 (right) shows the architecture of Evolved AI which functions as follows: signal #2 is sent to the mathematical model yielding output signal #3. The error signal #4, which is the difference between the feedforward signal #1 and the output signal #3, is minimized. Minimization of error signal #4 is achieved through optimization techniques by adjusting parameters within the mathematical model. Unsupervised learning (stochastic clustering) problems are solved by connecting the known input to both signal #1 and signal #2. By minimizing the error, signal #3 will match signal #1 and thus, characterize the mathematical model. For example, if the input signals (#1 and #2) happen to be a data point with a value of four and the mathematical model is the function ax+b, then the system will force the parameters (a,b) = (1,0), (3/4,1), (1/2,2), etc. As new data is presented to the model, e.g., five to yield (a,b) = (1,0), 4/5,1), (3/5,2), etc. parameters (a,b) will take on various pairs tending to a pair with a higher probability of occurrence, in this case (1,0).

The mathematical model may be generic or specific, depending on the application. If available, a practitioner with area expertise should incorporate a priori knowledge into the design of the mathematical model. The philosophical argument is not one of nature versus nurture, but nature *and* nurture, where leveraging a priori knowledge complements the learning process. In the next section, Evolved AI's hybrid architecture is applied to pulse spectrum analysis where the mathematical model characterizes radio frequency (RF) signals by center frequency (FC), pulse width (PW), and bandwidth (BW). But, lets look at the mathematical model in some detail first.

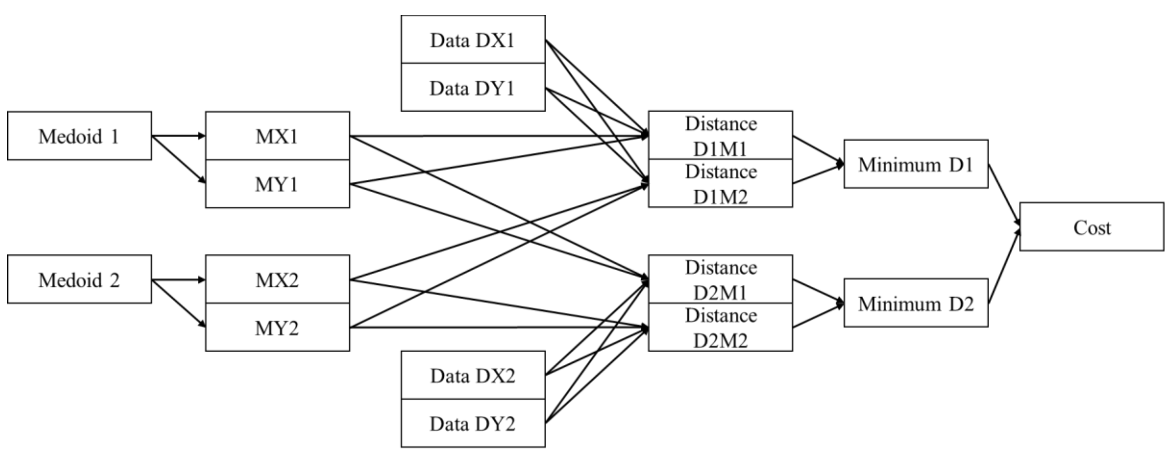

Figure 3 -An oversimplified example of k-Medoids for two clusters and two sets of data

Figure 3 shows an oversimplified mathematical model for k-medoid clustering. Two (two-dimensional) medoids and two data points are used to describe the model. The minimum distance between each data point [(DX1,DY1) and (DX2,DY2)] is computed for each medoid [(MX1,MY1) and (MX2,MY2)] to associate each data point with a medoid. The minimum distances (for each medoid) are summed as the cost function with the goal of minimizing this cost. Optimization techniques automatically assign data points to their closest medoid. With this basic understanding, the model used in the next section is expanded to include three-dimensional data (X=center frequency, Y=pulse width, and Z=bandwidth), nine data sets [(DX1,DY1,DZ1) through (DX9,DY9,DZ9)], and a randomly-selected number of medoid candidates.

PULSE SPECTRUM ANALYSIS

Often radar signals have a series of pulses with constant pulse repetition frequency (PRF), constant pulse width (PW), and constant pulse height. As such, basic radar signals may be identified by these parameters.

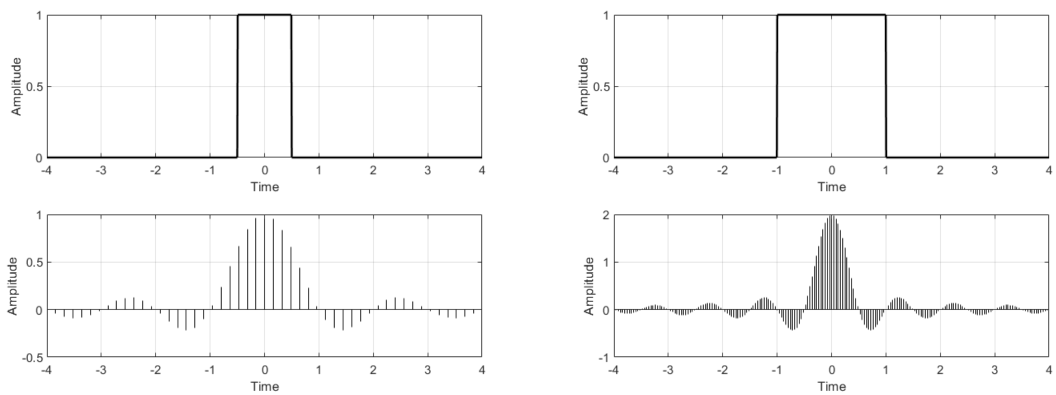

Figure 4 - Fourier Transform of Rectangular Pulses

Figure 4 shows Fourier transforms of two different rectangular pulse trains. To the left is a set of plots for a unity pulse width (above) and corresponding "sinc" envelope (below). To the right is a set of plots showing the impact of doubling the pulse width and lowering the pulse repetition frequency (above) where the sidelobes are narrowed and the spectral density increases (below). Note the main lobe width is 2/PW as may be seen from where each side of the main lobe crosses zero. In summary, narrower pulse width corresponds with wider sidelobes and higher pulse repetition frequency corresponds with lower spectral density. Finally, a "chirp" radar signal has a varying center frequency. With this introduction to pulse spectrum analysis, hopefully it is clear why pulse width, pulse repetition frequency, and center frequency are important parameters. Recall, pulse width and center frequency are two of the parameters being sought after by Evolved AI's unsupervised learning (stochastic clustering) mathematical model.

Evolved AI Applied to Pulse Spectrum Analysis

Nine datasets totaling over 500,000 points (FC, PW, and BW) serve as the input with the goal of understanding radar signal characteristics. Three candidate medoids (one for each of the parameters FC, PW, and BW) are randomly selected from the nine datasets. For each of the three candidate medoids, their corresponding parameters are identified and used to compute the Euclidean distance between them and the remaining data points. Stochastic optimization is invoked to find the minimum distances between each of the three candidate medoids and the rest of the data points. At the completion of this step, the candidate medoid is no longer a candidate, but the actual medoid for the cluster. The actual medoid may be used to identify the cluster or, as has been done, the centroid of the cluster is calculated – which doesn't differ much only in that the medoid is an actual data point (ala *k-Medoid*) whereas the centroid may not be an actual data point (ala *k-Means*).

Evolved AI's unsupervised learning analysis formed three stochastic clusters as displayed in Table 1. In the table, the first cluster centroid has a center frequency (FC) of 4.64 MHz, a pulse width (PW) of 5.39 μs, and a bandwidth (BW) of 0.38 MHz; the second cluster centroid has FC=6.31 MHz, PW=2.09 μs, and a BW=1.13 MHz; and the third cluster centroid is located at FC=3.18 MHz, PW=2.24 μs, and a BW=1.07 MHz.

Table 1 - Three Cluster Centroids

Cluster	FC (MHz)	PW (μs)	BW (MHz)
1	4.64	5.39	0.38
2	6.31	2.09	1.13
3	3.18	2.24	1.07

Each of the nine datasets associates itself with a cluster centroid as seen in Table 2. In the table, the aggregate parameters of each dataset are reported. For example, the first and sixth datasets associate themselves with the first cluster centroid, i.e., the aggregate center frequency (FC) of 6.05 MHz, a pulse width (PW) of 4.00 μs, and a bandwidth (BW) of 0.40 MHz from the first dataset and the aggregate center frequency (FC) of 3.22 MHz, a pulse width (PW) of 5.98 μs, and a bandwidth (BW) of 0.34 MHz from the sixth dataset associate themselves with the first cluster centroid FC=4.64 MHz, PW=5.39 μs, and a BW=0.38 MHz from Table 1. Likewise, the second, third, fourth, fifth, and seventh datasets associate with the second cluster centroid, and the sixth, eighth, and ninth datasets associate with the third cluster centroid. Interestingly, the sixth dataset associates itself with either the first cluster centroid or the third cluster centroid. By inspection, one can see in the former case, the sixth dataset aligns more with PW and BW; whereas in the latter case the sixth dataset aligns more with FC. This flexibility is attributed to the stochastic nature of the datasets. The physical explanation of this binary association is due to signal type. If the signal is assumed to be a radar chirp, then the center frequency plays a role; otherwise, pulse width and bandwidth dominate.

Table 2 - Cluster Centroids and Datasets

Cluster/Dataset	FC (MHz)	PW (μs)	BW (MHz)
1/1	6.05	4.00	0.40
1/6*	3.22	**5.98**	**0.34**
2/2	6.06	2.49	0.80
2/3	6.47	3.34	0.60
2/4	6.33	1.66	1.20
2/5	6.19	1.99	1.00
2/7	6.48	0.99	2.01
3/6*	**3.22**	5.98	0.34
3/8	3.18	2.98	0.67
3/9	3.15	1.18	1.70

Figure 5 shows a three-dimensional plot of the clusters and their corresponding variance. Cluster centroid #1 from Table 1 is the largest ellipsoid at the base of the plot; cluster centroid #2 is the ellipsoid tilted; and cluster centroid #3 is the ellipsoid located in the bottom-left/near corner.

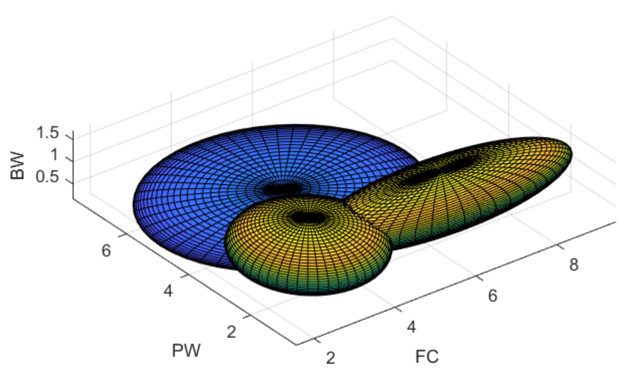

Figure 5 – 3D Plot of Cluster Centroids by FC, PW, and BW

This section concludes by recognizing Evolved AI's hybrid (hard and soft clustering) architecture is flexible to be applied to stochastic unsupervised learning. This hybrid approach is a step in the direction of robust AI.

SUMMARY

Motivation for Robust AI – more than deep learning alone – has been presented. A step in this direction is to explore hybrid architectures and incorporate a priori knowledge. In a previous paper, it was shown how these features were applied to supervised learning. NDIA's Senior Fellow for AI inspired the application of these same features to unsupervised learning. Evolved AI's unsupervised learning algorithm is a hybrid architecture comprising the benefits of hard and soft clustering, leveraging statistical inference, to determine cluster number as well as centroid location.

The system successfully proved its utility for multi-dimensional (center frequency, pulse width, and bandwidth) stochastic clustering when performing pulse spectrum analysis of radar signals. In the final analysis, the three parameters formed themselves into three cluster centroids (Table 1). Given these centroids, the datasets associated themselves with each cluster centroid accordingly (Table 2). Plotting the results (Figure 4) shows a 3D visualization of the stochastic ellipsoids.

The next step (to be done) is to extend the use of this system to determine the pulse repetition frequency (PRF). Referring to the bottom-left plot of Figure 3, the PRF (or pulse repetition interval, PRI) is the time between the vertical spectral lines forming the envelope. The PRF is the primary defining characteristics of a radar system. Long-range signals require lower PRFs (fewer pulses with less energy); while short-range signals require higher PRFs (more pulses with more energy).

Other Applications

While pulse spectrum analysis is interesting and applicable, another area of interest may be training and education. At the January 2020 Air Force Agency for Modeling and Simulation (AFAMS) Training Innovation Workshop, an agenda item covered data science for Pilot Training Next (PTN) efforts. The discussion centered around measuring position, velocity, and orientation data of pilot training exercises and comparing it with ideal execution. Naturally, this is a sensitive topic which invoked concern over the method being used to compute the measurement. Perhaps an objective approach to measuring pilot training exercise data and clustering results in terms of ideal execution might be an additional approach to compliment current methods.

REFERENCES

1. Alcorn, M. A., Li, Q., Gong, Z., Wang, C., Mai, L., Ku, W.-S. et al. (2018). Strike (with) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects. arXiv, 1811.11553v3.
2. Allen, R. (2019). Evolved AI - Interpretable Network Architectures for Machine Learning, I/ITSEC paper 19149.
3. Allen, R. (2019). Adaptive Nonconvex Optimization for AI and Machine Learning, I/ITSEC paper 19109.
4. Hao, K., (2020), A debate Between AI Experts Shows a Battle Over the Technology's Future, MIT Technology Review, https://www.technologyreview.com/s/615416/ai-debate-gary-marcus-danny-lange
5. Knight, W., (2017), The U.S. Military Wants Its Autonomous Machines to Explain Themselves, https://www.technologyreview.com/s/603795/the-us-military-wants-its-autonomous-machines-to-explain-themselves.
6. Marcus, G. and Davis, E., (2019), Rebooting AI: Building Artificial Intelligence We Can Trust, Pantheon.
7. Marcus, G., (2020), The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence, https://arxiv.org/pdf/2002.06177