When Machine Learning Fails: Correlations Expanded

Mona Hedayati
Concordia University, University of Antwerp
Montreal, Canada – Antwerp, Belgium
mona.hedayati@kdg.be

Abstract

This artist talk goes over the making process of my sensor-to-sound performance as it provides an overview of the datasets, I created by logging my somatic markers–biosignals and breathing patterns–tied to my experience of exile. I discuss the machine learning operation involved, its deficiency within the context of my creative project, and the subsequent remedy by integrating human judgement in the loop. Within these dynamics, by way of questioning correlation, a significant operation in data processing tasks, in the final section I offer a critique of ethico-politics of machine learning and its operationalization at scale in social settings.

Keywords

Biosensors; machine learning; immersive sound; sound performance.

Prologue

I installed all the dependencies and launched tensor board for monitoring training. I signed up for a Colab Pro+ account that gave me 500 compute units a month for 50 euros and 70 cents. I have chosen an A100 GPU in my runtime. I mounted my Google Drive where my data was sitting, specified a path to save the checkpoints and ran the training cell. I ended up spending 4 times more than what I expected through Colab's "pay as you go program" and viola, 6 days of training, timing out and resuming training and 3000 checkpoints later I got my brand-new model. Except that a bit later I found that I'm not happy with what the model outputs, but I've heard this adage in data science that you can find in data what you will. I could tune the model and start again: maybe a bit of data augmentation could do it.

Context

For the past 2 years I have been working with physiological sensors to understand their mechanism of quantifying neurophysiological responses to complex stimuli, in my case migration and exile. I have been experimenting with different pipelines, lately involving machine learning, as a way to communicate the ineffability of my lived experience of migration and exile through affective-sensory registers rather than logical sequences of linguistic and visual narration. Sound quickly became my sensory medium as I increasingly recognized its affective properties as a precognitive force with a capacity for experience building and event making. As Steve Goodman declares, sound has an *affectile* tone, given its material vibratory quality and its potential for affection [1]. Primarily, however, I had no intention to shift towards using machine learning for sound generation. But there came my latest experiment: at the onset of the new wave of political oppression and protests in Iran in 2022, I could not help but to follow the news. The only source of news, however, was the cellphone-shot social media feed of the protests captured by protestors and anonymously posted almost hourly. Since I was obsessed with and struck by watching these video documentations, I decided to capture my biosignals using the Empatica E4[1] wearable (fig. 1) while simultaneously audio recording these sessions.

Figure 1. Empatica E4 equipped with 4 sensors (EDA, BVP, thermometer and accelerometer). The green and red light emitted is from the BVP sensor that measures the amount of blood present in the tissue.

My initial intention with the audio recordings was to capture my breathing patterns, but during the process I realized I inevitably capture expressive gestures such as sniffling, gasping, sighing or crying but also my mouse clicks, thumps on the keyboard as well as occasional sound leaked

[1] https://e4.empatica.com/e4-wristband

from the protest videos, often including women shouting given the feminist nature of protests. This unintended level of affective complexity that was captured, opened up the possibility to build further specificity into the acoustic performance given the goal was conveying complexity of my stress response molded by my lived experience. To accomplish this goal, I worked these libraries into sound: the audio recordings into a static piece to be dynamically modulated by the biosignal library. But once I was confronted with these 2 libraries–a sizeable repository of CVS files associated with 4 streams of sensor data and more than 5 hours of audio recordings–I struggled to imagine how to make sense of them. They quickly became just data as I was trained to see them as such: files of audio waveforms and piles of numerical values logged in spreadsheets that I had to "tame" to make sense of.

The Making Process

Consequently, I talked myself into using algorithmic logic. Naively, I wanted the machine to condense hours on end of hard-headed signals into a singular information-rich output. In other words, I was hoping for a complex form of a rule-based system to manage my human emotions. I was under the impression that an objective mediation will render results beyond my messy human logic towards making true sense of data that is based on rules of math and not human desire. What I did not take into account, however, was the reductionist logic couched in the machine operation. As Lev Manovich declares, computers reduce the world into two pragmatic objects, data structures and algorithms in such a way that "any process or task is reduced to an algorithm…and any object in the world is modeled as a data structure, that is, data organized in a particular way for efficient search and retrieval" [2].

Subsequently, I trained a model using my datasets as I captured in the prologue using Rave,[2] a VAE (variational autoencoder) algorithm for audio generation. The algorithm overcomes a great number of limitations of the existing ones and outputs 48kHz audio through a two-step training; first a lengthy pre-training process using the VAE and then a GAN (generative adversarial network) architecture that freezes the training step and only increases the dimensionality through multiband decomposition of the raw form [3]. But regardless of all the impressive potential, for my use case the result was underwhelming; banal, toned down, and infused with mediocracy, stripped of what I as a human found critical to make sense of the recordings: all the subtleties, unique and unexpected gestures and surprising elements that made the library complex in its repetitiveness. To remedy the situation, I decided to intersect machine objectivity and human judgment to reach a median: that is the bland machine output, what Hito Steyerl calls "different shades of mean" with what I personally found significant in the library [4].

I worked on this collage in Sound Particles[3] (fig. 2), a sound design software that relies on the concept of particle systems to create large volumes and immersion. To add dynamism to the design using the biosignal library, I used simple statistical analysis methods geared towards signal processing to binarize the values by defining certain thresholds based on the properties of each signal type and its sample rate. This strategy resists conflating complexity through pattern recognition and classification to make sense of data but utilizes the variations in biosignals as a driver for sound. These incoming values were then sent through OSC[4] into Ableton Live[5] that I used for live performance. I then mapped these values into the parameters of different effects and instruments for a constantly evolving result.

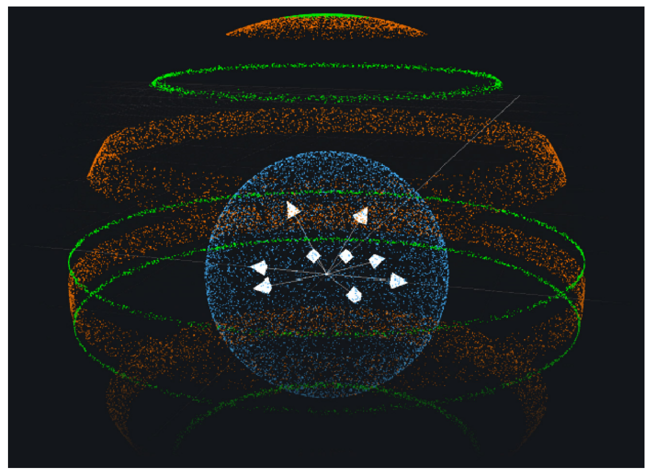

Figure 2. A complex spherical arrangement of sound. Image taken from Sound Particles website. © Sound Particles.

Correlation

Beyond the fact that I have found a strategy out of the predicament of human-machine messy business, through my research I became fascinated with the idea of correlation; the fact that it plays a significant role in data preprocessing tasks like feature engineering and dimensionality reduction for machine learning. Correlation analysis makes sense of data by taming the unruly rows and columns of matrices.

I started drawing likely and unlikely correlations: Akin to early 19th century scientists, I became quite concerned about the objective credibility of my experiment by wondering whether my controlled variables are in fact controlled or not. I started looking for hidden correlations and patterns in my data. Eventually, however, I learned the more I look for hidden correlations the more I find them, but they don't

[2] https://github.com/acids-ircam/RAVE
[3] https://soundparticles.com/
[4] Designed at the Center for New Music & Audio Technologies (CNMAT) by Matt Wright and Adrian Freed in 1977, Open Sound Control (OSC) is an encoding for live data communication specially designed and commonly used for multimedia musical performances but also functional between any hard and software endpoints that target real-time data exchange.
[5] https://www.ableton.com/en/

seem to answer any questions or aid in making any sense out of data.

To stretch the concept of correlation, I then started drawing conceptual and thematic correlations, particularly between the operative logic of totalitarian regimes and that of machine learning and that was when I found a significant correlation thanks to the work of Lorraine Daston on the history of rules in her book *Rules: A Short History of What We Live By* [5]. Daston introduces rulers and rulers: the former being the measuring tool that algorithms go under; from their history of basic arithmetic operations to their 20th century definition, while the latter are those who define and enforce the laws. Totalitarian regimes survive by rules they devise and enforce as laws which she calls *thin rules* or rules that do not leave any leeway for acting differently; they are not made for multiplicity of circumstances and leave no room for adaptation and error. Algorithms, according to her follow the same rule system: they are formulated thinly and are applied rigidly. In other words, they both operate based on the assumption that they will always deal with a stable and predictable world with all probabilities to be foreseen.

Coda

I want to predict your political views and tendencies. The hypothesis function makes predictions about you based on your features, as you are broken down into your features. Looks like an 'h' with a parenthesis next to it and an 'x' inside; the x here is you. The body of the hypothesis multiplies each one of your features based on your social media footprint by a coefficient also known as a weight and constantly adjusts these weights so the sum of this equation will be an accurate prediction. I worry here about two major steps: to narrow down your data into the most important features through dimensionality reduction and to come up with weights that multiplied by your features estimate your political views with a low margin of error, the ground truth. The same algebraic equation, essentially a mathematical expression, with some non-linearity thrown in like knobs you can turn is the foundation of deep learning. Deep learning is used as a weapon of truth generation in supposed democracies in predictive policing programs and national security systems for recognition. Over and above that, totalitarian regimes are and will use the mythic power of these knobs to enforce their versions of truth as they have been used so far in psychological profiling and social engineering as well as fake news and disinformation campaigns by state-sponsored cyber armies.

What kind of social order can we maintain using the logics of component analysis, linear and logistic regression? What does it mean to hijack personal data to come up with big data that represents our manipulated psyche, shattered hopes and social dreams? What does it mean then to decontextualize this data through feature extraction to create economies of attention and distraction? What does it mean to rely on the definitive results outputted by models that are trained using such inputs and call the output objective truth? What does it mean to evade responsibility and simply declare the machine responsible when the results are condemned as biased? What does it mean to then hire ghost workers to clean up these so-called "machine biases"?

References

[1] Goodman, Steve. *Sonic Warfare: Sound, Affect, and the Ecology of Fear*. (Cambridge and London: MIT Press, 2012).
[2] Manovich, Lev. *The Language of New Media*. (Cambridge: MIT Press, 2001), 223.
[3] Caillon, Antoine, and Philippe Esling. "RAVE: A Variational Autoencoder for Fast and High-Quality Neural Audio Synthesis," 2021. https://doi.org/10.48550/ARXIV.2111.05011.
[4] Steyerl, Hito. "Mean Images." New Left Review, June 2023. https://newleftreview.org/issues/ii140/articles/hito-steyerl-mean-images.
[5] Daston, Lorraine. *Rules: A Short History of What We Live By*. The Lawrence Stone Lectures. (Princeton: Princeton university press, 2022).

Bibliography

Caillon, Antoine, and Philippe Esling. "RAVE: A Variational Autoencoder for Fast and High-Quality Neural Audio Synthesis," 2021. https://doi.org/10.48550/ARXIV.2111.05011.
Chun, Wendy Hui Kyong. *DISCRIMINATING DATA: Correlation, Neighborhoods, and the New Politics of Recognition*. (Cambridge: MIT Press, 2021).
Daston, Lorraine. *Rules: A Short History of What We Live By*. The Lawrence Stone Lectures. (Princeton (N.J): Princeton university press, 2022).
Finn, Ed. *What Algorithms Want: Imagination in the Age of Computing*. Cambridge, (Cambridge and London: MIT press, 2017).
Goodman, Steve. *Sonic Warfare: Sound, Affect, and the Ecology of Fear*. (Cambridge and London: MIT Press, 2012).
Manovich, Lev. *The Language of New Media*. (Cambridge: MIT Press, 2001).
Steyerl, Hito. "Mean Images." New Left Review, June 2023. https://newleftreview.org/issues/ii140/articles/hito-steyerl-mean-images.

Acknowledgements

This contribution draws on research supported by the Social Sciences and Humanities Research Council of Canada.

Author Biography

Mona Hedayati is an Iranian-Canadian artist and a joint PhD researcher in interdisciplinary humanities at Concordia University, Canada and artistic research at Antwerp Research Institute for the

Arts, Belgium. Her work draws on sound design, computation arts, and sensory studies. She has a BA in translation studies, an MFA in digital media and an advanced master's in social-political art and design. Hedayati's research practice encompasses the interdisciplinary field of science, technology, and society (STS) to bring the social thickness back into the science and technology practices that are conventionally considered immune to social concerns. Given the hybrid nature of her work that hovers across disciplines, Hedayati has presented her scholarly and artistic work internationally across a wide array of platforms as diverse as Kunsthal Extra City, Antwerp; Body Electric Retrospective, Toronto; ISEA, Paris; xCoAx, Weimar; 4s, Honolulu; ARTECH, Faro and Ars Electronica, Linz. As an educator, Hedayati teaches courses on digital culture and critical theory at master's and advanced master's level.