Neural networks through the lens of evolutionary dynamics

Baciu, Dan C.

DOI
10.1016/j.biosystems.2024.105386

Publication date
2025

Document Version
Final published version

Published in
BioSystems

BioSystems 248 (2025) 105386

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Neural networks through the lens of evolutionary dynamics

Dan C. Baciu [a,b,c,*]

[a] Interpretation Laboratory, UCSB, USA
[b] Interpretation Laboratory, ASB, Switzerland
[c] TU Delft, Netherlands

ARTICLE INFO

Keywords:
Evolutionary dynamics
Artificial neural networks
Artificial intelligence

ABSTRACT

This article revisits Artificial Neural Networks (NNs) through the lens of Evolutionary Dynamics. The two most important features of NNs are shown to reflect the two most general processes of Evolutionary Dynamics. This overlap may serve as a new and powerful connection between NNs and Evolutionary Dynamics, which encompasses a body of knowledge that has been built over multiple centuries and has been expanded to inspire applications across a vast range of disciplines. Consequently, NNs should also be applicable across the same range of disciplines—that is, much more broadly than initially envisioned. The article concludes by opening questions about NN dynamics, based on the new connection to Evolutionary Dynamics.

1. Introduction

In the study of Artificial Intelligence, Neural Networks engineers have largely followed an empirical approach. They have been creative in devising new Neural Networks architectures and selecting the most powerful ones through empirical testing. In this technological evolution, the development of theory has often taken a backseat to the pursuit of empirical successes. High-impact journals such as *Nature Machine Intelligence*, *IEEE Transactions on Neural Networks and Learning Systems*, *Neural Networks*, and the *Journal of Machine Learning Research* often rejected theoretical articles, favoring instead contributions that achieved new state-of-the-art results. The disadvantage of this approach was the ensuing lack of clarity about how new developments were linked to fields such as Evolutionary Dynamics, for example. The advantage, however, was empirical success. Multiple significant empirical breakthroughs have brought the development of increasingly powerful technologies. This article highlights two generations of Neural Networks (NNs), specifically: the initial generation and Transformers.

The initial generation of NNs was developed significantly in the 20th century. This generation became particularly powerful with the introduction of neuron layers arranged as sequential dependency models. Examples are autoregressive NNs and Recurrent Neural Networks (Hopfield, 1982; Elman, 1990; Rumelhart, Hinton and Williams, 1986; Hochreiter and Schmidhuber, 1997; Jordan, 1986). Transformers, on the other hand, are primarily a product of the 21st century. This second generation of networks stands out due to the development of "self-attention mechanisms" (Bahdanau et al., 2014; Xu et al., 2015). Many researchers consider the publication of the article "Attention is All You Need" (Vaswani et al., 2017) a pivotal moment.

Given that the breakthroughs of NNs have been empirically driven, the theory behind these advancements has had to be repeatedly adjusted. For example, the theory behind self-attention mechanisms was initially inspired by research in natural language and human cognition. However, the field of application self-attention mechanisms is much broader, highlighting the need for connections to more general theoretical frameworks.

The present article revisits neuron layers and self-attention mechanisms through the lens of Evolutionary Dynamics. From the mathematical description of neuron layers, I recover the quasispecies equation—a general equation of evolution used to describe variation-selection processes. This equation has proven valuable in studying evolution and creativity across a broad range of disciplines (Domingo and Schuster, 2016; Singh et al., 2023; Baciu, 2023). Furthermore, I show that self-attention mechanisms are a form of frequency-dependent selection, which is also the basis of game theory (Hofbauer and Sigmund, 1998; Baciu, 2023). Logistic growth and Lotka-Volterra equations are a foundational example of frequency-dependent selection (Baciu, 2023).

Both variation-selection processes and frequency-dependent selection have been described before NNs were invented. They build on over a century of research and have been instrumental in studying creativity, play, and diversification across numerous fields (Lotka, 1910; Ross,

* Freudenreichstr. 22, 3047 Bremgarten, Switzerland.
E-mail address: baciu@ucsb.edu.

https://doi.org/10.1016/j.biosystems.2024.105386
Received 31 October 2024; Received in revised form 27 December 2024; Accepted 27 December 2024
Available online 1 January 2025

1911; Baciu, 2023). Thus, this article sets the stage for linking Neural Networks to a truly broad and long-established body of knowledge.

The article concludes with a question that I would like to open. In describing cultural change through variation-selection processes and frequency-dependent selection, I have observed a type of behavior that is common and that the quasispecies and Lotka-Volterra equations can describe very effectively (Baciu, 2018). However, I am uncertain whether this particular behavior has yet been leveraged to enhance the dynamics of Neural Networks.

2. Neuron layers and variation-selection processes

Any Artificial Neural Network takes input data and transforms it into output data, which can replace the original input data in computer memory. Thus, any Neural Network can be described as a function that maps one set of values onto another. This leads to the very generic equation, Equation (1).

$$\boldsymbol{X}_{t+1} = \phi(\boldsymbol{X}_t). \quad (1)$$

Here, ϕ is a yet-to-be-determined function that transforms the input data $\mathbf{X_t}$ into the output data $\mathbf{X_{t+1}}$. Input and output data can be stored in any digital format, such as numbers, vectors, matrices, or tensors. Considering that the output data at time *t+1* replaces the input data at time *t*, the equation becomes a recurrence equation as stated above.

With this setup, our Neural Network is a first sequential dependency model. Data passes through it repeatedly, with outputs becoming new inputs. This idea can be expanded and refined, as shown in Equation (2).

$$\boldsymbol{X}_{t+1} = \phi(\boldsymbol{X}_{t-n:t}). \quad (2)$$

Here, *n* time steps are considered as input data, while the output data is only the next time step. This is the case in autoregressive models such as ChatGPT, where a sequence of *n* words is used to predict the next one.

Equations (1) and (2) are still very generic. They say little about the function ϕ, which remains entirely undetermined in the equations. To be able to infer exact parameters, Neural Networks proceed by decomposing this unknown function ϕ into multiple nested functions. In most Neural Networks architectures, each of these nested functions represents a layer of the Neural Network. For example, a Neural Network with two layers breaks down the function $\phi(x)$ into $f_2(f_1(x))$, where the two nested functions correspond to the network's two consecutive layers. Some researchers like imagining the nested functions as distinct time steps. This is entirely unnecessary, as the splitting into multiple inner and outer functions is merely a way to handle the larger (composite) function. Thus, the sub-functions can be seen as nested functions that work together, at once. Let us proceed with this, simpler interpretation.

In the rest of this article, nested functions will be dealt with individually, keeping in mind that they are later combined into composite functions to obtain the desired complex behavior. This systematic layer-by-layer evaluation ensures simply that we have a clear understanding that refects each mathematical component accurately.

There are three types of layers: neuron layers, activation functions, and transformer layers. Each of these layers corresponds to a different type of function. Let us begin with neuron layers.

Each neuron layer consists of neurons connected in such a way that every neuron in a layer *y* is connected to all neurons in the previous layer *x*. This architecture can be mathematically described as Equation (3).

$$y_j = \sum_{i=1}^{n} w_{ij} x_i + b. \quad (3)$$

Here, y_j represents a neuron in the new layer *y*, while x_i is a neuron in the previous layer *x*, which consists of *n* neurons. The coefficient w_{ij} is the weight of the connection from x_i to y_j. The constant *b* allows for additional adjustments to the model and is often referred to as bias in the literature (Goodfellow et al., 2016).

Considering that the Neural Network is a sequential dependency model, the outputs from the model become new inputs. This means that the values of the neurons in layer *y* are re-inserted into the neurons on the layer *x*. If the NN has only one layer, this is immediately the case. If it has multiple layers, it means that there are additional processes at work, which we can deal with separately. Consequently, we arrive at Equation (4).

$$x_{i,t+1} = \sum_{i=1}^{n} w_{ij} x_{i,t} + b. \quad (4)$$

Here, the layers *x* and *y* each have *n* neurons, and the layer *y,* now denoted as $x_{i,t+1}$, is fed back into the equation to account for the sequential dependency feature of the Neural Network, as in Equation (1).

From Equation (4), it is evident that the basic Evolutionary Dynamics equation of variation-selection processes can be recovered easily. Variation-selection processes can be formulated with the quasispecies equation, which in its simplest form is written as Equation (5).

$$x_{i,t+1} = \sum_{i=1}^{n} w_{ij} x_{i,t}. \quad (5)$$

In the standard notation of the quasispecies equation, the weights, here denoted as w_{ij}, are referred to as mutation and selection rates and are denoted as $r_i\, q_{ij}$ (Nowak, 2006). Mathematically, these values perform the same role, serving as coefficients in Evolutionary Dynamics as well as NNs. The mathematical function of the "weights" or "mutation and selection rates"—however one wishes to call them—is to specify how often x_i is transformed into x_j. Technically speaking, these values determine the level of creativity in the system, as they govern the transformation of one variable into another.

The most common behavior of this equation is straightforward to predict. When the equation is expressed in matrix notation, the eigenvector corresponding to the largest eigenvalue eventually predominates. In genetic systems, this behavior is found in competitive exclusion: the fastest-growing quasispecies comes to predominate, excluding all others. As a result, the system reaches fixation. A similar type of fixation is observed in Neural Networks that process text. Neural Networks with this architecture often drift toward a single topic, which then reaches fixation. When predicting the next word in a sequence, the model tends to get stuck predicting words related only one topic. This type of fixation was easy to observe in NNs utilized on smartphones in the early 2020s. The NNs were implemented to suggest the next word to people typing text messages. If one repeatedly took the next suggestion, without actually typing anything by oneself, the result was a nonsensical text composed of few and similar words. For example the text "I'm teaching a course on AI, what topics" was completed word by word to the nonsensical "I'm teaching a course on AI, what topics *you are available to confirm your eta schedule with sunshine and your availability for your availability for your availability and availability for your availability for tomorrow* …"

The reason for this fixation behavior is that the equation is linear. Therefore, to improve the model, nonlinearity must be introduced. This requires us moving from variation-selection processes to frequency-dependent selection, which is the nonlinear Evolutionary Dynamics counterpart. The move towards nonlinearity can be achieved in two different ways. The simpler method is to introduce a nonlinear component without altering much of the equations. This approach is frequently applied to the quasispecies equation as follows (Equation (6)).

$$x_{i,t+1} = \sum_{i=1}^{n} w_{ij} x_{i,t} - \phi(\boldsymbol{x}_t). \quad (6)$$

Here, $\mathbf{x}$ is the vector containing all values of x_i and the Greek letter ϕ denotes a nonlinear function of $\mathbf{x}$. The role of $\phi(\mathbf{x})$ is to prevent exponential growth when the largest eigenvalue exceeds one, which is the most common outcome (Nowak, 2006). The history of this term in the equation goes back to the mid-19th century, when Quetelet and Verhulst adjusted the Malthusian population growth model as follows (Equation (7)).

$$x_{t+1} = wx_t - \phi(x_t). \tag{7}$$

The first two terms of the equation, $x_{t+1} = wx_t$, are the Malthusian population growth model, which is a linear equation describing exponential population growth. It can be interpreted as a special case of the quasispecies equation in which there is only one variable. As in Equation (5), w is the rate of change, and x is a scalar value, representing the variable being transformed. So far, the only distinction to Equation (5) is that the input and output data are limited to a single scalar value that changes over time.

Quetelet and Verhulst modified the Malthusian growth model by introducing the component $-\phi(x_t)$. This component performs the same function as later in the quasispecies equation: it is used to prevent exponential growth (Verhulst, 1838; Quetelet, 1842).

After introducing this component, consisting of an unknown function designed to limit population growth, Verhulst sought to identify a function that could be fitted empirically to population growth data. This approach allowed him to set $-\phi(x_t)$ as $-wx_t^2$, as follows (Equation 8):

$$x_{t+1} = w_1 x_t - w_2 x_t^2. \tag{8}$$

The resulting function is a sigmoid growth function, also known as an s-curve (Bejan and Lorente, 2012). The function grows exponentially at first, but the negative $-w_2x_t^2$ eventually increases much faster than w_1x_t. As a result, the growth is curbed, and a characteristic s-shape emerges.

This type of nonlinear behavior, which has entered the quasispecies equation through population studies, has a matching counterpart in NNs. S-curves are commonly used as activation layers that follow the neuron layers of NNs (Goodfellow et al., 2016; LeCun et al., 1998). Thus, it can be concluded that even the additional nonlinear term $-\phi(x)$ in the quasispecies equation has corresponding counterparts in NNs. The quasispecies equation can be recovered from neuron layers as well as neuron layers with sigmoid activation layers.

Strictly speaking, the nonlinear component $\phi(x)$ that has just been introduced is a first example of frequency-dependent selection, which brings us directly to the next section, where frequency-dependent selection is discussed more broadly.

3. Self-attention mechanisms and frequency-dependent selection

The architecture of Neural Networks introduces an additional way to connect neurons. In "self-attention mechanisms," the neurons in each layer x are directly connected not only to neurons in other layers but also to one another within the same layer. Many engineers believe that this type of architecture has driven a significant breakthrough, transforming Neural Networks from a relatively obscure technology into one that spearheads the explosion of Artificial Intelligence applications today (Uszkoreit, 2017). The architecture of these layers can be mathematically described as Equation (9).

$$y_i = x_i f_i(\mathbf{x}). \tag{9}$$

Here, x_i is a scalar value stored in layer x, while $f_i(\mathbf{x})$ is a function of all scalar values stored in layer x. The layer y follows layer x with the same number of neurons.

While the mathematical description provided here is not the one typically used in textbooks, it captures the essential requirements of all variants of self-attention and multi-headed attention. Nonlinearity is introduced directly into the equations, with a multiplication between x_i and $f_i(\mathbf{x})$, were $f_i(\mathbf{x})$ is a value computed as a function of some or all variables within the same NN layer. This architecture, which connects variables within the "same layer," inspires the term "self" in self-attention (Vaswani et al., 2017). By denoting this architecture as an autoregressive architecture, I set y as x_{t+1}, as in Equation (2). Thus, the following equation is obtained (Equation 10).

$$x_{t+1} = x_i f_i(\mathbf{x}_t). \tag{10}$$

Here, x_i represents each value as it changes, passing through the system.

It is possible to also consider multiple time steps, as in Equation (2). This makes for Equation (11):

$$x_{t+1} = x_i f_i(\mathbf{x}_{t-n:t}). \tag{11}$$

Here, time is considered in a more complex fashion, with the time step t in equation (10) expanded to a range of n steps, represented as $t - n{:}\ t$ in equation (11). All of these time steps become part of the same NN layer.

One example of how this architecture is used can be found in neural networks that predict the next word in a sentence, such as ChatGPT, once again. In this case, the architecture predicts the meaning of the next word based on a context of n preceeding words. Typically, all pairwise relationships between the n words are analyzed, which feeds forward to adjust the numerical representation of each individual word for further processing. In this context, the function $f_i(\mathbf{x}_{t-1:t})$ contains operations such as $x_{i,t} \times x_{i,t-1}$. Following the above arrangement, these values are now on the same layer, and as in the case of Verhulst's logistic growth model, the multiplication between variables on the same layer makes the equations nonlinear.

Evidently, Equations (10) and (11) are a case of frequency-dependent selection. They are equivalent to generalized Lotka-Volterra equations, also known as the replicator equation (Hofbauer and Sigmund, 1998). Thus, it is shown that self-attention layers can be described as a form of frequency-dependent selection.

Admittedly, frequency-dependent selection allows for somewhat more flexibility, as the function $f(x)$ can be any function, whereas only a limited range of possible functions has yet been tested in Transformer layers. Typically, self-attention mechanisms contain multiplications of the type $x_{i,t} \times x_{i,t-1}$. Frequency-dependent selection has more flexibility, allowing more generally for any $x_i \times x_j$ on the same layer. Nevertheless, it is remarkable to notice that even frequency-dependent selection often limits computation to pairwise multiplications, although in principle, it would be possible to also compute multiplications between three variables. This provides an additional similarity of how nonlinearity is implemented in practice.

In general, the main idea of frequency-dependent selection is to introduce nonlinearity through a multiplication between variables on the same layer or processing step. This is also the case for self-attention in Transformer layers and it therefore provides an excellent theoretical connection between NNs and Evolutionary Dynamics. This connection should inspire engineers to explore new architectures. The **Discussion** and **An open question** sections propose one such possibility. Frequency-dependent selection knows of an exciting range of slightly different nonlinear equations.

4. Discussion: in search of an all-encompassing description of the world

Perhaps it seems counterintuitive that Neural Networks can be as useful as they are. At their core, they are simply mathematical functions. Yet, they can be used, for example, to translate texts, translate text into computer code, or generate new texts. Why is something like this even possible with just functions?

The idea that functions can be used to study almost everything that we encounter has long predated the invention of NNs. In the mid 19th century, Gottfried Semper, a student of Carl Friedrich Gauss's, proposed utilizing an equation equivalent to Equation (1), applying it to the study of art and cultures.

As input data for the function, Semper envisioned using data about culture and availability of materials (clay, metal, stone, etc.). As output data, he hoped to receive designs that historically emerged in response to the conditions specified by the input data. Considering that he input data would have been formatted as a text describing present conditions, and the output data would have been descriptions of artworks emerging in these conditions, the function could have worked exactly like a NN operating on text.

Of course, in the 19th century, there weren't digitized datasets for Gottfried Semper to process and there was no practicable way to fit such an unknown, multivariate function. Semper's idea was nevertheless groundbreaking and entirely correct. Today, Neural Networks prove that it works. The innovation of Neural Network was to envision how to break down an unknown, multivariate function into nested functions, and fit these functions to data. Semper's formula was elegant and correct, although not completely practical in absence of data and learning algorithms. The NN approach is extremely practical. Although it is not always elegant and resourceful, it solves what Semper could not do: fitting the function to data.

Neural Networks have become operational by splitting Semper's unknown function ϕ into three different types of nested functions.

1. Neuron layers corresponding to variation-selection processes,
2. Activation layers, which introduce a first, basic version of frequency-dependent selection, and
3. Attention mechanisms holding a more general implementation of frequency-dependent selection.

Given that the internal architecture of Neural Networks, which makes them operational, makes them excellently match up with the field of Evolutionary Dynamics, the question emerges, how such a matchup emerged. Is the distinction between neuron layers and attention mechanisms and that between variation-selection processes and frequency-dependent selection perhaps something truly fundamental, permeating all science? This question can be answered in the affirmative.

Variation-selection processes can be brought down to linear equations, and they can be used to describe any type of creativity. Frequency-dependent selection requires nonlinearity, and it can be used to describe any kind of playful behavior. The applications of the mathematics go beyond not only Neural Networks but also beyond Evolutionary Dynamics. The history of formulating the models is also older than both the creation of the first Neural Networks and the first formulation of Lotka-Volterra equations. It is a history that begins in Antiquity.

The Roman philosopher and poet Lucretius already described the idea of creativity through variation and selection in the first century BC (Lucretius, Greenblatt, 2011). In addition, he was aware of its limitations. He wrote that a world conceived through variation and selection alone would end up behaving like raindrops falling linearly from the sky. He foresaw the need of deviation from linearity in what he termed "tilting" or "swerving." Thus, linearity and nonlinearity have been around for a very long time.

In modern times, a fascinating, sparkling description of variation-selection processes can be found in the work of Alfred Russel Wallace, specifically the article that, according to some historians, he has written after a life-threatening fever, before sending it to Charles Darwin, who used it to inspire himself for his own essay about natural selection (Wallace, 1858; Darwin and Wallace, 1858). Darwin and Wallace's work has later attracted countless scientists, although it initially only contained verbal explanations, while lacking a correct mathematical description of variation-selection processes.

The mathematical description of variation-selection processes has come to us somewhat later. It was advanced by Eigen and Schuster, and it has led to new perspectives in multiple fields, but in particular in virus dynamics and vaccines (Singh et al., 2023; Domingo and Schuster, 2016; Nowak, 2006; Nowak and May, 2000). In addition, variation-selection processes can be used to describe creativity in any kind of physical, biological, social, or cultural system (Baciu, 2023). Thus, neuron layers and variation-selection processes reflect a system's creativity.

Frequency-dependent selection is not less important. It is a necessary counterpart to variation-selection processes because it introduces nonlinearity. The range of applications is equally stunning (Page and Nowak, 2002). Frequency-dependent selection returns in the basic equations of ecology, virology, game theory, and is also a necessary ingredient of chaos theory (Vandermeer and Goldberg, 2013; Nowak and May, 2000, Hofbauer and Sigmund, 1998). Like variation-selection processes, frequency-dependent selection also works to describe processes in an immensely broad array of applications. In the broadest sense, it can describe diversification and interplay in any kind of physical, biological, social, or cultural system (Baciu, 2023).

The distinction between the two processes is important to understand. It is the same distinction as that between linearity and nonlinearity and between additions and multiplication. Variation-selection processes are expressed with linear operations and sums. Between variables, there are only additions signs. On the other hand, frequency-dependent selection is described with multiplications between variables. This makes the distinction between these two processes the same as that between additions and multiplications in mathematics and Or and AND operations in logic. It is a fundamental distinction that permeates all human thinking (Baciu, 2023).

Given that Neural Networks use both of these complementary modes of representation to describe the world, their broad applicability is no surprise. I hope that this perspectives inspires engineers to apply Neural Networks even more broadly: Wherever one can apply additions and multiplications, (or and and or operations if one prefers logic), one can apply Neural Networks!

Initially, the deployment of attention mechanisms was motivated through empirical success, specifically in the task of translating natural languages (Vaswani et al., 2017). New state-of-the-arts were achieved, which motivated further development. The interpretation of the mathematics of "attention" and "self-attention" was provided through theories that applied to natural language.

However, as engineers created larger and larger "Large Language Models", they soon observed that the same Neural Network architecture can be applied across a much broader range of disciplines. This discovery came as a surprise to many, and debates emerged why this was possible at all. Why could one use attention layers for purposes that had little or nothing at all to do with phenomena of attention or natural language? Soon, NNs that contained attention layers were dubbed "Transformers." Perhaps the hope was that the change of name would hide the lack of a comprehensive theory. The public-facing debate over why Transformers can be so broadly applied continues.

With the interpretation of Neural Networks proposed here, the broad applicability of neuron layers and attention mechanisms is not only a possibility but a necessary result. It would be more surprising if an architecture as found in present-day Neural Networks did not apply to the broad range of subjects it does. In addition, the interpretation proposed here links Neural Networks to centuries of scientific thinking across all disciplines. I hope that this link will inspire engineers to broaden the application of NNs even further.

Taken together, I interpret neuron layers in NNs as mathematical models that can describe any kind of creative transformation, and I interpret attention mechanisms as dynamical models that can be broadened to describe any kind of nonlinear interplay between variables. This interpretation provides a way of thinking about Neural Networks that renders justice to their broad applicability and supports future application in the broadest imaginable range of subjects.

5. An open question

In my previous work, I have developed mathematical descriptions of cultural change. These descriptions contain linear as well as nonlinear components, which describe creativity and play, respectively. Creativity comes through as variation-selection processes, while play comes through as frequency-dependent selection (Baciu, 2018, 2015).

In a series of articles, I have tested these mathematical descriptions on massive data. This has been done as part of the project Everything Called Chicago School and What Every1 Says (Baciu, 2015; Baciu, 2018, 2019, 2020, 2023).

In both projects, I used frequency-dependent selection to describe changing fashions that come and go, and sometimes return. In this context, there is an empirical phenomenon that I predicted and observed, related to how fashions spread across large groups of authors and audiences. This phenomenon is relevant here because it is commonly seen in culture, but it may not be a common behavior of Neural Networks.

Here is a description of the cultural phenomenon in question: It begins with a fashion that spreads, growing exponentially in popularity. In response to the spreading fashion, one can commonly observe that boredom and opposition spread as well, as a byproduct. Most fashions that spread also bring a growing opposition. What happens next is that the interplay between fashion and opposition becomes increasingly prevalent in the system. This is a consequence of the fact that both the fashion and the opposition become widespread. Clashes between fashion and opposition must therefore become even more frequent. As such clashes continue, they have a negative effect on the growth of the fashion. In consequence, the fashion begins to disappear, being faced by increasingly fierce opposition or boredom that hinders its growth. Overall a fashion wave is observed in which the fashion becomes 'fashionable', only to go out of fashion again Baciu, 2015, 2018, 2019, 2020, 2023.

This phenomenon can be described with Lotka-Volterra equations as a basic case of frequency-dependent selection. Technically, the above statements, if translated into mathematics, directly provide the necessary equations. I will not list them here, but they are found in my previous work and are a special case of Equation (10). There is also a video that explains the process graphically with watercolors in less than 5 min. The video also explains that, using equivalent equations, the same wave pattern can also be described in ecology, virology, epidemiology, and other fields of science. A link to it is found in Baciu, 2023).

In addition, in human culture, there are also circumstances in which multiple fashions can work together, which can be described with a more complex case of Equation (10) combined with considerations from Equation (5). The overlap of multiple nested processes then leads to larger waves of growth and reform (Baciu, 2015, 2018, 2019, 2020, 2023). Similar phenomena are also observed in virology (Nowak and May, 2000).

I am mentioning this example here because I would like to open a question. I am unaware of any Neural Network that is able to detect the dynamic behavior of fashions or is behaving in ways that resemble fashions. The question is: Has anyone trained a network that detects how fashions come, go, and return, as a result of interplay between ideas that spread while also activating opposition? Or has anyone trained a Neural Network that behaves as if it was going through fashions or mood swings, giving output of one kind for some time, only to eventually switch to opposing its own output with new divergent outputs? I know of many NN-based language models that get fixated in topics, but I don't know of any that knows how to change topics as playfully as people do. There's a joke that while it takes only one token to get a model to start conversing with you, it takes at least two tokens to change topic.

Ethics approval and consent to participate

Not applicable

Availability of data and materials

Any data and material used are available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Preprint published 2024.

Abbreviations

NN - Neural Network. ED - Evolutionary Dynamics.

References

Baciu, D.C., 2015. La Tour Eiffel: from technological experiment to icon of Paris. Phoenix 2, pp. 56–61. https://escholarship.org/uc/item/51f3d4bz.

Baciu, D.C., 2018. From everything called Chicago School to the theory of varieties. Dissertation, Illinois Institute of Technology. Retrieved from. http://hdl.handle.net/10560/4376.

Baciu, D.C., 2019. The Chicago School: Large-Scale Dissemination and Reception. In Chicago Schools: Authors, Audiences, and History. Prometheus 2. Published on 2019-11-16, within the section Keynotes.

Baciu, D.C., 2020. Cultural life: theory and empirical testing. Biosystems 197 (104208). https://doi.org/10.1016/j.biosystems.2020.104208.

Baciu, D.C., 2023. Causal models, creativity, and diversity. Humanities and social sciences communications, 10, article 134. https://doi.org/10.1057/s41599-023-01540-1.

Bahdanau, D., Cho, K., Bengio, Y., 2014. Neural Machine Translation by Jointly Learning to Align and Translate.

Bejan, A., Lorente, S., 2012. The S-Curves are Everywhere. J. Inst. Eng. 134 (5), 44–47. https://doi.org/10.1115/1.2012-MAY-5.

Darwin, C., Wallace, A.R., 1858. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. Journal of the proceedings of the linnean society of london. Zoology 3 (9), 45–62.

Domingo, E., Schuster, P., 2016. Quasispecies: from theory to experimental systems. Part of the Book Series: Current Topics in Microbiology and Immunology, Volume 392. Springer.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. MIT Press.Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

Greenblatt, S., 2011. The Swerve: How the World Became Modern. W.W. Norton & Company.

Hofbauer, J., Sigmund, K., 1998. Evolutionary Games and Population Dynamics. Cambridge University Press. https://doi.org/10.1017/CBO9781139173179. Online ISBN: 9781139173179.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324.

Lotka, A.J., 1910. Contribution to the theory of periodic reactions. J. Phys. Chem. 14 (3), 271–274.

Lucretius. (1st Century BCE). De Rerum Natura.

Nowak, M.A., 2006. Evolutionary Dynamics: Exploring the Equations of Life. Harvard University Press.

Nowak, M.A., May, R.M., 2000. Virus Dynamics: Mathematical Principles of Immunology and Virology. Oxford University Press.

Page, K.M., Nowak, M.A., 2002. Unifying evolutionary dynamics. J. Theor. Biol. 219 (1), 93–98. https://doi.org/10.1006/jtbi.2002.3060.

Quetelet, A., 1842. A Treatise on Man and the Development of His Faculties. William and Robert Chambers, Edinburgh.

Ross, R., 1911. The Prevention of Malaria. John Murray.

Singh, K., Mehta, D., Dumka, S., Chauhan, A.S., Kumar, S., 2023. Quasispecies nature of RNA viruses: lessons from the past. Vaccines 11 (2), 308. This article reviews the quasispecies nature of RNA viruses, addressing the theoretical and mathematical origins of quasispecies and their dynamics.

Uszkoreit, J., 2017. Transformer: a novel neural network architecture for language understanding. Google research blog. https://blog.research.google/2017/08/transformer-novel-neural-network.html.

Vandermeer, J., Goldberg, D., 2013. Population Ecology: First Principles. Princeton University Press.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Polosukhin, I., 2017. Attention Is All You Need. In Advances in Neural Information Processing Systems, pp. 5998–6008.

Verhulst, P.F., 1838. Notice sur la loi que la population suit dans son accroissement. Correspondance mathématique et physique 10, 113–121.

Wallace, A.R., 1858. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. Journal of the proceedings of the linnean society of london. Zoology 3 (9), 53–62.

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Bengio, Y., 2015. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. In International Conference on Machine Learning, pp. 2048–2057.